

Universidad Católica de Santa María
Facultad de Ciencias e Ingenierías Físicas y Formales
Escuela Profesional de Ingeniería de Sistemas



**ANÁLISIS DEL SECTOR AGROPECUARIO A TRAVÉS DE LA CLUSTERIZACIÓN
BASADO EN LA PRODUCTIVIDAD, ECONOMÍA Y PRÁCTICAS DE DESARROLLO
DEL SECTOR PERÚ, 2015 – 2019**

Tesis presentada por el Bachiller:

Barrios Bellido, Nino

para optar el Título Profesional de

**Ingeniero de Sistemas en la
especialidad de Sistemas de
Información**

Asesor:

**Ing. Zúñiga Carnero, Manuel
Mariano**

Arequipa- Perú

2023

UCSM-ERP

UNIVERSIDAD CATÓLICA DE SANTA MARÍA
INGENIERIA DE SISTEMAS
CON ESPECIALIDAD EN SISTEMAS DE INFORMACION
TITULACIÓN CON TESIS
DICTAMEN APROBACIÓN DE BORRADOR

Arequipa, 02 de Enero del 2023

Dictamen: 004203-C-EPIS-2023

Visto el borrador del expediente 004203, presentado por:

2015221871 - BARRIOS BELLIDO NINO

Titulado:

**ANALISIS DEL SECTOR AGROPECUARIO A TRAVÉS DE LA CLUSTERIZACIÓN BASADO EN LA
PRODUCTIVIDAD, ECONOMIA Y PRACTICAS DE DESARROLLO DEL SECTOR. PERU, 2015-2019**

Nuestro dictamen es:

APROBADO

**1635 - SULLA TORRES JOSE ALFREDO
DICTAMINADOR**



**1748 - CALDERON RUIZ GUILLERMO ENRIQUE
DICTAMINADOR**



**2464 - RAMIREZ VALDEZ OSCAR ALBERTO
DICTAMINADOR**



DEDICATORIA

Esta tesis va dedicada a mi mamá Jaqueline, por haberme apoyado a lo largo de toda mi vida, por su constante esfuerzo y sacrificio para brindarme siempre la mejor educación, y a cada miembro de mi familia, hermanos, tíos, primos, abuelos por haber contribuido grandes cosas a mi vida y formación profesional.



AGRADECIMIENTO

Agradezco profundamente a todas las personas que han contribuido en este proceso, brindándome sus conocimientos, motivación, constante apoyo y orientación dedicados a este proyecto.



RESUMEN

Debido a las problemáticas del sector agropecuario relacionados con la rentabilidad, competitividad, comercialización y características del sector agropecuario. Se realizó un estudio con la finalidad de optimizar el desarrollo del sector agropecuario, analizándolo y creando clústeres basados en la productividad, economía y prácticas de desarrollo en el Perú en el periodo del año 2015 al 2019.

Para llevar a cabo esta investigación, que es no experimental de corte longitudinal, con enfoque cuantitativo, se aplicó técnicas de ciencia de datos, para el análisis aplicado a la base de datos que el INEI pone a disposición del público en general, de la cual se extrajo esta información utilizando R Studio, se limpió y se le aplicó el algoritmo de clustering jerárquico aglomerativo, a las variables que son binarias simétricas, empleando Python y librerías de minería de datos como Pandas, Numpy, Matplotlib e Scipy, para posteriormente presentar datos mediante gráficas en Power BI, donde se describe los patrones en común que se comparten entre distritos, y finalmente proponer propuestas para optimizar el desarrollo del sector agropecuario.

Como resultado final se generó una fuente de información para poder describir cada distrito resaltando sus principales características y los patrones en común que comparten con otros departamentos, provincias y distritos, y su evolución a lo largo de los años, esta información servirá de ayuda a las entidades públicas como municipalidades u otras entidades privadas, para poder ver las necesidades y oportunidades de mejorar el sector agropecuario en diversas áreas geográficas del Perú.

Palabras Clave

Clustering jerárquico aglomerativo, variables binarias simétricas, ciencia de datos, machine learning, clústeres, sector agropecuario.

ABSTRACT

Due to the problems of the agricultural sector related to profitability, competitiveness, marketing and characteristics of the agricultural sector. A study was carried out with the purpose of optimizing the development of the agricultural sector, analyzing it and creating Clusters based on productivity, economy and development practices in Peru in the period from 2015 to 2019.

To carry out this research, which is not experimental longitudinally, with a quantitative approach, for this data science techniques were applied, for the analysis applied to the database that the INEI makes available to the general public, which this information was extracted using R Studio, it was cleaned and the agglomerative hierarchical clustering algorithm was applied to the variables that are binary symmetric, using Python and data mining libraries such as Pandas, Numpy, Matplotlib and Scipy, for later present data through graphs in Power BI, where the common patterns shared between districts are described, and finally propose proposals to optimize the development of the agricultural sector.

As a final result, a source of information will be maintained to be able to describe each district, highlighting its main characteristics and the common patterns it shares with other departments, provinces and districts, and its evolution over the years. This information will help the public entities such as municipalities or other private entities, to be able to see the needs and opportunities to improve the agricultural sector in various geographical areas of Peru.

Keywords

Agglomerative hierarchical clustering, symmetric binary variables, data science, machine learning, clusters, agricultural sector.

ÍNDICE

DEDICATORIA	iii
AGRADECIMIENTO	iv
RESUMEN	v
ABSTRACT	vi
ÍNDICE DE FIGURAS	ix
INTRODUCCIÓN	1
CAPÍTULO I	3
PLANTEAMIENTO DE LA INVESTIGACIÓN.....	3
1.1. Objetivos de la investigación	3
1.1.1. General.....	3
1.1.2. Específicos	3
1.2. Alcances y Limitaciones	3
1.3. Tipo y nivel de investigación	3
1.4. Población y muestra o universo	4
1.5. Métodos, Técnicas e Instrumentos de Recolección de Datos	4
1.6. Descripción de la Solución.....	5
1.7. Fundamentos Teóricos	5
1.7.1. Antecedentes del proyecto	5
1.7.2. Bases Teóricas del proyecto.....	9
1.7.3. Técnicas y herramientas.....	10
CAPÍTULO II.....	14
2. DESARROLLO DE LA PROPUESTA.....	14
2.1. Comprensión del área de estudio y fijación de objetivos.....	14
2.2. Recopilación y selección de datos.....	14
2.3. Limpieza y procesamiento de datos	17
2.4. Aplicación de minería de datos	18
2.4.1. Aplicación de algoritmo.....	19
2.4.2. Validación del modelo	22
CAPÍTULO III.....	26
3. RESULTADOS	26
3.1. ANÁLISIS DESCRIPTIVO.....	26

3.1.1.	Módulo 200 AB Destino de los productos cosechados	26
3.1.2.	Módulo 200AB_1 Perdida total de producción	56
3.1.3.	Modulo 400A Producción pecuaria	85
3.1.4.	Módulo 600 Inocuidad agroalimentaria.....	95
3.1.5.	Módulo 700 Servicios de extensión agraria.....	127
3.1.6.	Módulo 900 Servicios financieros	140
3.2.	CONOCIMIENTO PARA LA TOMA DE DECISIONES.....	157
3.2.1.	Destino de los productos cosechados.....	157
3.2.2.	Pérdida total de producción	158
3.2.3.	Producción pecuaria.....	158
3.2.4.	Inocuidad agroalimentaria	158
3.2.5.	Servicios de extensión agraria	159
3.2.6.	Servicios financieros	159
CONCLUSIONES		160
RECOMENDACIONES		161
REFERENCIAS		162
ANEXOS		168

ÍNDICE DE FIGURAS

Figura1 <i>Datos obtenidos de INEI.</i>	11
Figura 2 <i>Módulos almacenados en formato CSV.</i>	12
Figura 3 <i>Modelo propuesto de clustering.</i>	19
Figura 4 <i>Aplicación de algoritmo de clustering.</i>	19
Figura 5 <i>Matriz de distancias.</i>	20
Figura 6 <i>Código que grafica el dendrograma.</i>	21
Figura 7 <i>Matriz de distancias.</i>	22
Figura 8 <i>Características de los clústeres.</i>	23
Figura 9 <i>Matriz de distancias.</i>	24
Figura 10 <i>Índice de cohesión y separación.</i>	25
Figura 11 <i>Módulos clusterizado cargados en Power Bi</i>	25
Figura 12. <i>Características de los clústeres en la selva</i>	28
Figura 13. <i>Nota. Mapa de los clústeres de la selva.</i>	29
Figura 14. <i>Distribución de variables selva 2015-2019</i>	30
Figura 15. <i>Características del departamento de San Martín.</i>	31
Figura 16. <i>Características distrito de Picota</i>	34
Figura 17. <i>Características de los clústeres en la Costa</i>	38
Figura 18. <i>Mapa de los clústeres de la costa.</i>	39
Figura 19. <i>Distribución de variables costa 2015-2019</i>	40
Figura20. <i>Características del departamento de Piura.</i>	41
Figura 21. <i>Características del distrito de Acarí.</i>	44
Figura 22. <i>Características de los clústeres en la Sierra</i>	47
Figura 23. <i>Mapa de los clústeres de la sierra.</i>	48
Figura 24. <i>Distribución de variables sierra 2015-2019</i>	49
Figura 25. <i>Características departamento de Ancash</i>	50
Figura 26. <i>Características del distrito de Huayopata.</i>	53
Figura 27. <i>Características de los clústeres en P224E.</i>	57
Figura 28. <i>Distribución de variables 2016-2019.</i>	58
Figura 29. <i>Principales productos que se perdió su producción.</i>	59
Figura 30. <i>Mapa de los clústeres P224E.</i>	60

Figura 31. <i>Características del departamento de Puno.</i>	61
Figura 32. <i>Características del distrito de Azángaro.</i>	64
Figura 33. <i>Características de los clústeres en P229.</i>	68
Figura 34. <i>Mapa de los clústeres P229.</i>	69
Figura 35. <i>Características del departamento de Piura.</i>	70
Figura 36. <i>Características del distrito de Oxapampa.</i>	73
Figura 37. <i>Características de los clústeres en P22A</i>	76
Figura 38. <i>Mapa de los clústeres P229A.</i>	77
Figura 39. <i>Características del departamento de Puno.</i>	78
Figura 40. <i>Características del distrito de Manu.</i>	81
Figura 41. <i>Características de los clústeres en P403AA.</i>	85
Figura 42. <i>Mapa de los clústeres P403A.</i>	86
Figura 43. <i>Características del departamento de Moquegua.</i>	87
Figura 44. <i>Características del distrito de Río Santiago.</i>	90
Figura 45. <i>Características de los clústeres en P601</i>	97
Figura 46. <i>Mapa de los clústeres P601.</i>	98
Figura 47. <i>Características del departamento de Apurímac.</i>	99
Figura 48. <i>Características del distrito de Huancarani.</i>	102
Figura 49. <i>Características de los clústeres en P602.</i>	106
Figura 50. <i>Mapa de los clústeres P602.</i>	107
Figura 51. <i>Características del departamento de Madre de Dios.</i>	108
Figura 52. <i>Características del distrito de Acombamba.</i>	111
Figura 53. <i>Características de los clústeres en P605.</i>	117
Figura 54. <i>Mapa de los clústeres P605.</i>	118
Figura 55. <i>Características del departamento de Lambayeque.</i>	119
Figura 56. <i>Características del distrito de Moyobamba.</i>	122
Figura 57. <i>Características de los clústeres en P710 año 2015.</i>	127
Figura 58. <i>Mapa de los clústeres P710 año 2015.</i>	128
Figura 59. <i>Características de los clústeres en P710 año 2016.</i>	128
Figura 60. <i>Mapa de los clústeres P710 año 2016.</i>	129
Figura 61. <i>Características de los clústeres en P710 año 2017.</i>	129

Figura 62. <i>Mapa de los clústeres P710 año 2017</i>	130
Figura 63. <i>Características de los clústeres en P710 año 2018</i>	130
Figura 64. <i>Mapa de los clústeres P710 año 2018</i>	131
Figura 65. <i>Características de los clústeres en P710 año 2019</i>	131
Figura 66. <i>Mapa de los clústeres P710 año 2019</i>	132
Figura 67. <i>Características de los clústeres en P713</i>	133
Figura 68. <i>Mapa de los clústeres P713</i>	134
Figura 69. <i>Características departamento de Lima</i>	135
Figura 70. <i>Características del distrito de Tisco</i>	137
Figura 71. <i>Características de los clústeres en P903</i>	140
Figura 72. <i>Mapa de los clústeres P903</i>	141
Figura 73. <i>Características del departamento de Pasco</i>	142
Figura 74. <i>Características del distrito de Paucartambo</i>	144
Figura 75. <i>Características de los clústeres en P904</i>	150
Figura 76. <i>Mapa de los clústeres P904</i>	151
Figura 77. <i>Características del departamento de San Martín</i>	152
Figura 78. <i>Características del distrito de Bajo Biavo</i>	154

INTRODUCCIÓN

El sector agropecuario ha sido uno de los de mayor crecimiento en el Perú en los últimos años, lo que le ha permitido al país fortalecer su posición como exportador regional y mundial de frutas y hortalizas y así posicionarse exitosamente como uno de los más importantes en el mercado mundial (Olivera, 2020). Este crecimiento es muy importante para el motor económico del país ya que genera empleos y oportunidades de inversión (RedAgrícola, 2019).

Las tecnologías de la información, tienen un gran potencial para poder optimizar todas y cada una de las actividades sociales (Bosch, 2012), un estudio realizado por Microsoft y Edelman (2022), la pandemia del Covid-19 ha acelerado exponencialmente la transformación digital del 95% de las pequeñas y medianas empresas de nuestro país. El estudio muestra que las empresas invierten en tecnología, en almacenamiento y computación en la nube. La digitalización progresiva ha aumentado las necesidades de toma de decisiones de las empresas en el campo del análisis de datos, sin embargo, carece de infraestructura para la ciencia de datos.

Al ver esta necesidad de optimizar el sector agropecuario y esta oportunidad en el campo de la ciencia de datos, se puede decir que la industria agropecuaria es un nuevo espacio para tecnologías revolucionarias como la ciencia de datos. Los agricultores que proporcionan insumos esenciales para la industria ahora pueden tomar decisiones basadas en los datos que recopilan de los avances tecnológicos e impactar positivamente en sus procesos de producción, En las unidades agropecuarias, existen diferentes formas de recopilar datos para mejorar la eficiencia en el uso de recursos como agua, suelo, combustible, fertilizantes y productos fitosanitarios (UTEC, 2022), En el sector agropecuario las TIC pueden brindar soluciones para mejorar la eficiencia, productividad y calidad del trabajo de los agricultores y los agentes involucrados en el sector agropecuario (Progis, 2015).

En la presente investigación se tiene por objetivo aplicar la ciencia de datos para poder analizar el sector agropecuario, mediante la aplicación del algoritmo jerárquico aglomerativo a las variables de investigación para generar clústeres, es decir, generar grupos de productores agropecuarios los cuales comparten las mismas características, como aspectos financieros, aspectos de productividad, fenómenos climáticos que influyen en su producción y de prácticas agropecuarias, esta información pretende ser utilizada por agricultores y entidades que están involucradas en el sector como empresas privadas y entidades públicas como municipios

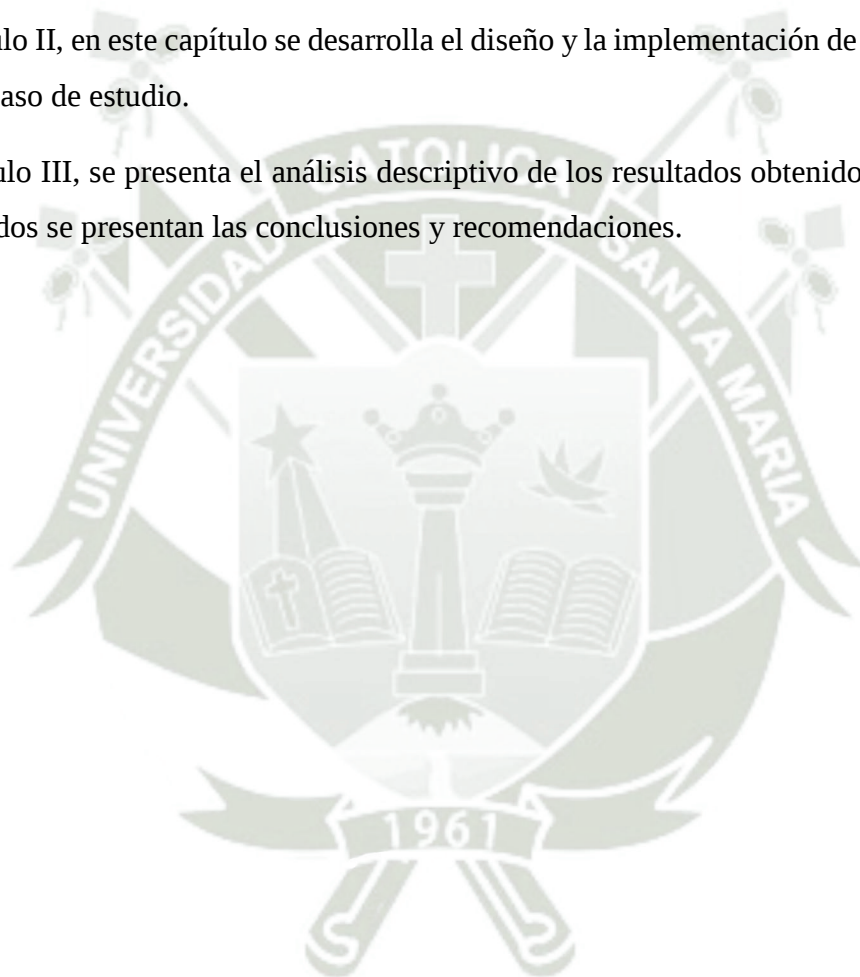
distritales, provinciales y departamentales, para brindarles una fuente de información confiable para que conozcan las características de su población, sus necesidades, preferencias, intereses comunes, entre otros criterios.

Este trabajo de investigación está dividido en cuatro capítulos:

En el capítulo I, presenta el problema a investigar y se describe detalladamente el planteamiento de la investigación, marco teórico y antecedentes del proyecto.

En el capítulo II, en este capítulo se desarrolla el diseño y la implementación de la metodología propuesta, al caso de estudio.

En el capítulo III, se presenta el análisis descriptivo de los resultados obtenidos, y en función a estos resultados se presentan las conclusiones y recomendaciones.



CAPÍTULO I

PLANTEAMIENTO DE LA INVESTIGACIÓN

1.1. Objetivos de la investigación

1.1.1. General

Analizar el sector agropecuario a través de la clusterización basado en la productividad, economía y prácticas de desarrollo del sector en el Perú durante el periodo 2015 al 2019, para conocer las necesidades, preferencias e intereses comunes de los productores agropecuarios.

1.1.2. Específicos

- A. Identificar las bases de datos del sector agropecuario y aplicar la extracción, transformación y carga de datos para los periodos 2015-2019.
- B. Identificar los factores y criterios de segmentación del sector agropecuario peruano (productividad, economía y prácticas de desarrollo).
- C. Implementar el modelo de clustering, para obtener conglomerados de productores agropecuarios que comparten características similares.
- D. Realizar análisis descriptivo de los agrupamientos obtenidos, e interpretarlos para identificar las características en común de los productores.

1.2. Alcances y Limitaciones

La presente investigación tiene los siguientes alcances:

- A. Viabilidad, los datos a utilizar en la presente investigación están limitados a los productores agrícolas y pecuarios, encuestados en el Censo Nacional Agropecuario, que nos provee el INEI.
- B. Lugar, el estudio se realiza a los productores agrícolas y pecuarios del territorio peruano.
- C. Tiempo, se utilizarán los datos comprendidos entre los años 2015-2019.

1.3. Tipo y nivel de investigación

El diseño de la investigación propuesta es no experimental de corte longitudinal, con enfoque cuantitativo, ya que se trabajará con los datos públicos del INEI, los cuales fueron obtenidos mediante la encuesta nacional agropecuaria del año 2014 al 2019, en el cual se observa el

comportamiento natural del sector agropecuario, en base diversas características. En base a estas características e información se evaluará el comportamiento del sector agropecuario a través de cada año, para hacer inferencias respecto a los cambios, sus determinantes y consecuencias, que nos ofrezcan los modelos estadísticos, resultantes del procesamiento de datos.

1.4. Población y muestra o universo

La población de estudio son las unidades agropecuarias, parcelas y productores agropecuarios, obtenidos por la encuesta nacional agropecuaria, que sea aplicó año tras año desde el 2014 al 2019, donde por cada año contiene 23 módulos. La muestra total es de 30 806 unidades agropecuarias, distribuidas en 29 218 unidades agropecuarias de medianos y pequeños productores con los niveles de inferencia nacional y departamental, y 1588 unidades agropecuarias de grandes productores; estos se desarrollaron teniendo en cuenta 23 módulos de investigación, las cuales son:

- Características de la unidad agropecuaria
- Cultivos cosechados en la unidad agropecuaria
- Buenas prácticas agrícolas: condicionadas y no condicionadas
- Producción de la actividad pecuaria
- Buenas practicas pecuarias: condicionadas y no condicionadas
- Inocuidad agroalimentaria
- Servicios de extensión agraria
- Asociatividad
- Servicios financieros

Se considera emplear la técnica de muestreo no probabilístico censal, donde se considera la totalidad de la población de estudio.

1.5. Métodos, Técnicas e Instrumentos de Recolección de Datos

Para llevar a cabo los objetivos de esta investigación se aplica la técnica de observación, utilizando como instrumento la observación documentaria, ya que la investigación se basará en los datos de la Encuesta Nacional Agropecuaria, que dispone de información estadística que permite caracterizar a las pequeñas, medianas y grandes unidades agropecuarias del país, la cual está a la disposición de los investigadores a través del sistema de Microdatos del INEI. En este

repositorio se pueden ver los módulos y la documentación de las encuestas ejecutadas por el INEI, en formatos compatibles con SPSS, Microsoft Excel y Acrobat Reader.

La Encuesta Nacional Agropecuaria fue realizada del 2014 al 2019, en donde cada año contiene 23 módulos, de acuerdo a diversos criterios que fueron tomados en cuenta por los representantes del MINAGRI y MEF. Cada módulo contiene los formularios aplicados durante la encuesta, un diccionario de datos y la información de las encuestas en archivos con extensión sav.

Esta investigación está basada en los aspectos productivos, económicos y de desarrollo, por lo que se solo se utilizarán los módulos referentes a dichos aspectos, que contribuyan y aporten conocimiento a los objetivos de la investigación.

1.6. Descripción de la Solución

Para la realización de la presente investigación se estudiará las variables que influyen en la pérdida de producción, en la venta de cultivos y ganado, prácticas en la producción agropecuaria, aspectos financieros y servicios de extensión agraria.

La información que se utilizará será tomada de la Encuesta Nacional Agropecuaria, que el INEI pone a disposición del público, a través de su sistema de Microdatos, la información será tomada del año 2015 al 2016. Dicha información será procesada aplicando la minería de datos, para crear clústeres de productores agropecuarios que comparten características similares, finalmente se realizará un análisis descriptivo con los clústeres obtenidos.

1.7. Fundamentos Teóricos

1.7.1. Antecedentes del proyecto

Un estudio realizado por Gil-Torres et al. (2019), en Colombia, resalta el desarrollo de las tecnologías y el aumento en la generación de datos, para la generación de información para la toma de decisiones, y la planeación estratégica dentro de las organizaciones, dicho estudio muestra los aportes que nos brindan dichas tecnologías como el análisis de datos geográficos, aplicando algoritmos de minería de datos, donde se descubren patrones inesperados y modelos que a simple vista no se ven, con lo cual se pueden descubrir correlaciones potencialmente útiles para las organizaciones, así mismo el estudio presenta una comparativa de casos resaltando los

principales resultados obtenidos por países que apostaron por el uso de la minería de datos para sus procesos de producción agrícola, optimizando sus resultados y maximizando su producción.

En otro estudio realizado en argentina se evaluó los efectos del clúster en la innovación y la creación de información valiosa, como resultado el estudio demostró que los agrupamientos impactan positivamente en la comunidad en la que se encuentran, favoreciendo la gestión del conocimiento, donde los organismos del estado son muy importantes para su sustentabilidad a través de las políticas públicas hacia la Asociatividad, promoviendo la competitividad y el desarrollo del sector(Arteche et al., 2013).

En un estudio realizado por Wilkinson (2021), el autor busca identificar los principales grupos de delitos de la ciudad de Londres, aplicando la agrupación jerárquica, teniendo en cuenta no solo la cantidad de delitos sino el tipo de delitos, para la aplicación del algoritmo jerárquico aglomerativo, el autor evalúa los resultados que obtiene, interpretando los grupos y asignado etiquetas a cada uno de ellos de acuerdo a sus características más resaltantes, estos resultados deben fáciles de interpretar y etiquetar para poder definir que se tiene un clúster de calidad, de no ser así se debe ajustar el algoritmo, modificando el método de agrupación y la medida de distancia entre clústeres, en la investigación también se grafica los principales crímenes cometidos en un determinada área geográfica, en donde puede determinar que se cumple la primer ley de la geografía la cual establece que “todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas que las cosas distantes”, la cual también se puede demostrar en el desarrollo de esta tesis.

En la investigación de Mamani (2015) se aplica diversos algoritmos de minería de datos basados en clustering y se evalúa su eficiencia para generar conglomerados que ayuden a la interpretación de la información, los cuales permiten evaluar situaciones en una empresa de manera holística y enfocarse en el cumplimiento de los objetivos establecidos en el plan estratégico empresarial. En el trabajo planteado por Obando (2017), se desarrolló una evaluación de los algoritmos de minería de datos más eficientes en un entorno comercial, dando como resultado información detallada acerca de los mejores algoritmos para los procesos comerciales y la ventaja que aporta a la toma de decisiones.

El campo de aplicación para la ciencia de datos y sus técnicas relacionadas, es muy amplio y variado, en el siguiente estudio se utiliza la capacidad de estas tecnologías para poder monitorear

las tendencias y cambios del mercado laboral para que sirva como un componente de ayuda para que las personas que buscan empleo, empleadores, legisladores e inversionistas puedan utilizar, para poder extraer patrones subyacentes de grandes colecciones de datos se utilizó O*NET, el cual es una base de datos pública que contiene los requisitos ocupacionales de los trabajadores, en esta investigación para llevar a cabo sus objetivos sugiere un enfoque de minería de datos donde se desarrolló un modelo de Indexación Semántica Latente (LSI) que cuenta con la capacidad de hacer coincidir los anuncios de trabajo extraídos de la web con los datos que describen las perfiles ocupacionales en O*NET, esta implementación resaltó las tendencias laborales en diferentes industrias y áreas geográficas, identificando los grupos ocupacionales y estudiando los cambios en el contexto laboral en un determinado tiempo, estos descubrimientos brindó lecturas mucho más precisas del mercado laboral en comparación con estadísticas oficiales comunes de empleo, esta investigación aporta un enfoque que se puede aplicar directamente en diversos mercados laborales, sectores comerciales y regiones geográficas (Karakatsanis et al., 2017).

En la investigación titulada “A data science and engineering solution for fast k-means clustering of big data” de Chiang, (2018) propone un enfoque de minería de datos para poder abstraer mercados importantes que van a servir para las estrategias de marketing de Gestión de Relaciones con los Clientes (CRM) en una cafetería en Taiwán, donde se empleó un algoritmo de agrupamiento difuso, reglas de asociación en modo difuso y un análisis de la canasta de mercado para procesar los datos recopilados en un sistema CRM, con la finalidad de identificar a clientes difusos, como resultado de esta investigación se lograron encontrar mercados nítidos y mercados difusos además de reglas de asociación para marketing y patrones de comportamiento en las compras de los clientes, esta contribución proporciona un enfoque de minería de datos en la que integra medidas difusas y algoritmos a priori supervisados para generar las reglas de asociación en modo difuso, finalmente con la información generada y analizada, los investigadores pudieron proponer nuevos servicios para atraer a clientes más jóvenes ofreciendo servicios gratuitos, como internet, información y pedidos a través de teléfonos inteligentes.

En la siguiente investigación se utiliza el método de agrupación de K-Means con densidad poblacional y entropía espacio local en un estudio para poder abordar la problemática de la expansión urbana descontrolada y el peligro que representa para la sostenibilidad de las ciudades, este problema se buscó solucionar mediante la aplicación de un modelo para simplificar la

identificación de patrones de expansión urbana, el cual se basa en interpretaciones de diagramas de patrones clásicos de estructura espacial urbana, finalmente los resultados y las comparaciones de los datos de población abierto y datos de teléfonos móviles, verifican que el agrupamiento es adecuado identificando las unidades de tierra en expansión del área urbana y rural, como resultado final la investigación sugiere un modelo de expansión urbana y su entropía espacial local (Liu et al., 2018).

En el siguiente artículo científico del autor Yu et al. (2020) se dedica a aplicar la adaptive resonance theory (ART) y self-organizing map (SOM) en el aprendizaje automático, en el estudio se demuestra que estos algoritmos no supervisados, determinan mejor los clústeres más óptimos para la agrupación que el K-Means. Esta demostración se dio mediante la aplicación en conjunto de SOM complementado con K-Medias para estudiar cómo funciona el marketing de juegos en línea en Taiwán, finalmente como resultado de esta investigación, después de agrupar los datos, se pudo hacer referencia a las decisiones y responder a preguntas en función de los resultados de agrupación, como, por ejemplo, donde se puede promover un juego nuevo para atraer clientes potenciales, las diferentes edades de los clientes y en que se clasifican, además se pudo aprovechar la información de las preferencias de los clientes para diseñar nuevos juegos basándose en la arquitectura y atributos populares de los juegos.

El rápido crecimiento de empresarial representa un rápido crecimiento de la información, es por eso que en la investigación “A Customer Group Mining Method Based on Cluster Analysis” de (Yogping & Zizhen, 2020) se aprovecha esta información de los datos de marketing para analiza con precisión de datos de los clientes, descubrir las características de consumo de los clientes, y proporcionar diferentes productos a diferentes grupos de clientes, para esto se aplicó un algoritmo de análisis de agrupamiento al proceso de minería de grupos de clientes, que mediante la información clave de los clientes existentes se los clasificó en categorías de clientes y se obtuvo información de clientes de mayor potencial, como resultado de este estudio se demuestra que la segmentación de clientes basado en la agrupación de resultados mejora la eficiencia del análisis de la relación entre grupos de clientes y los productos, identificando factores importantes que afectan la venta de productos y proporcionan una base de apoyo para que las empresas lleven a cabo un servicio de precisión basado en el cliente.

En el siguiente artículo publicado por Miloudi et al. (2021) se presenta un enfoque eficaz para buscar un agrupamiento óptimo en múltiples bases de datos y evaluar la calidad de la agrupación, para esto se desarrolla un algoritmo en clúster basado en gradientes para la minería de datos, el cual lo aplicaron para desarrollar múltiples experimentos en diversas fuentes de datos y como resultado se logró demostrar la superioridad de este algoritmo con respecto a otros algoritmos de clusterización ya existentes, ya que es más óptimo en la ejecución y la precisión.

Según un estudio desarrollado en Sinaloa por los autores Zayas et al. (2017) estudia el aporte al desarrollo mediante la aplicación de clústeres, donde se indica que la aglomeración de un espacio territorial benefician a las empresas para estar una cerca de la otra y de esta manera integrar una gran red de empresas, formando clústeres como una herramienta de gestión para obtener nuevas posibilidades de desarrollo, como resolviendo problemas de baja competitividad e innovación en pequeñas, medianas y grandes empresas de producción agrícola.

Para la llevar a cabo los objetivos de esta investigación se va a tomar en cuenta el artículo Dierckens et al., (2017) ya que se aborda algunas problemáticas de sus algoritmos asociados de K-Means, como los riesgos que se presentan en las selecciones de Centroides al azar, tendencias de generar agrupaciones esféricas y la alta complejidad de tiempo de ejecución en un entorno de Big Data, para esto se desarrolló un algoritmo llamado CluDataSE que aporta a la solución de las problemáticas mencionadas anteriormente y además proporciona una agrupación rápida de k-medias para Big Data, como resultado de esta aplicación CluDataSE demostró una mejora del algoritmo K-Medias, ya que logró resolver las falencias mencionadas, y se logró obtener mejores resultados en los que el K-Means normalmente no podría hacerlo.

1.7.2. Bases Teóricas del proyecto

A. Variables dicotómicas

Es una variable que solo puede tener dos valores, generalmente toma el valor de cero, que se interpreta como ausencia, o toma el valor de uno que se interpreta como presencia, por lo tanto, es una variable que permite conocer la presencia o ausencia de un fenómeno o característica, esta variable además es de tipo categórico, lo cual quiere decir que expresa una cualidad, lo cual permite agrupar las observaciones en categorías(Arias, 2022).

B. Clustering jerárquico

Es un método de minería de datos, que sirve para agrupar datos no supervisados en clústeres, basándose en la distancia entre las observaciones y buscando que los datos dentro de un clúster sean lo más similares entre sí, y se puede representar en un dendrograma en forma de árbol con la finalidad de facilitar la identificación de los agrupamientos, existen dos tipos de clustering jerárquico, cada uno depende de la dirección en la que el algoritmo ejecuta el agrupamiento, estos son:

- **Aglomerativo:** Este tipo de clustering empieza la agrupación analizando cada elemento individual, donde la primera hoja del árbol representa un clúster por separado, que se va fusionando con el clúster más cercano, es decir con el clúster que tiene menor distancia, estas funciones se dan de manera sucesiva produciendo una jerarquía de resultados de clustering, donde al final de los agrupamientos solo queda un único clúster, que aglomera todas las observaciones.
- **Divisivo:** Es lo opuesto a un clúster aglomerativo, en este caso se comienza a la inversa, donde se parte de un único clúster, que aglomera todas las observaciones, y se va dividiendo en clústeres más pequeños.

En el dendrograma se puede ilustrar fácilmente las fusiones o divisiones dependiendo del tipo de clustering, además se puede observar la distancia, todos los niveles de asociación (Malkin, 2018).

1.7.3. Técnicas y herramientas

A. Metodología propuesta

Para el desarrollo de la presente investigación, se consideró como metodología el proceso de descubrimiento de conocimiento en una base de datos o KDD, este proceso consta de 6 etapas, para la transición de datos en conocimiento, estas etapas son, comprensión del área de estudio y fijación de objetivos, recopilación y selección de datos, limpieza y procesamiento de los datos, aplicación de minería de datos, interpretación y análisis de patrones encontrados y finalmente la aplicación del conocimiento para la toma de decisiones.

B. Herramientas a utilizar

En cada etapa del proceso se utilizó herramientas diferentes. A continuación, se describe la herramienta utilizada en todo el ciclo de vida del proyecto.

1) SPSS

La información que el INEI pone a disposición del público está en formato sav, como se muestra en la figura, esta herramienta cuenta con la capacidad de poder realizar cluterización, pero para el desarrollo del proyecto se optó por utilizar Python debido a que a diferencia de SPSS es una herramienta open source, además, debido a la gran cantidad de datos que se utilizaron, se formaron matrices de gran tamaño que son más rápidas de procesar en Python que en SPSS, además de que podemos adaptar nuestro código generado de acuerdo a nuestras necesidades.

Figura1
Datos obtenidos de INEI.

Nro	Año	Periodo	Código Encuesta	Encuesta	Código Módulo	Módulo	Ficha	Descarga
1	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1529	CARACTERÍSTICAS DE LA UNIDAD AGROPECUARIA		
2	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1530	SUPERFICIE COSECHADA, SEMBRADA, PRODUCCIÓN Y DESTINO DE LOS CULTIVOS COSECHADOS		
3	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1531	PRODUCCIÓN DE PASTOS POR CORTE / PASTOREO		
4	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1532	PERDIDA TOTAL DE PRODUCCIÓN O SUPERFICIE SEMBRADA		
5	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1533	PRODUCCIÓN Y DESTINO DE LOS CULTIVOS COSECHADOS		
6	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1534	DERIVADOS AGRÍCOLAS DE LOS CULTIVOS COSECHADOS		
7	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1535	SUBPRODUCTOS AGRÍCOLAS DE LOS CULTIVOS COSECHADOS		
8	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1536	COSTOS DE PRODUCCIÓN DE LOS CULTIVOS COSECHADOS		
9	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1537	BUENAS PRÁCTICAS AGRÍCOLAS		
10	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1538	PRODUCCIÓN PECUARIA		
11	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1539	SUBPRODUCTOS PECUARIOS		
12	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1540	DERIVADOS PECUARIOS		
13	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1541	BUENAS PRÁCTICAS PECUARIAS		
14	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1542	INOCUIDAD AGROALIMENTARIA		
15	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1543	SERVICIOS DE EXTENSIÓN AGRARIA		
16	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1544	ASOCIATIVIDAD		
17	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1545	SERVICIOS FINANCIEROS		
18	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1546	COSTOS DE PRODUCCIÓN DE LA ACTIVIDAD AGROPECUARIA		
19	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1547	CARACTERÍSTICAS DEL PRODUCTOR/A AGROPECUARIO/A Y SU FAMILIA		
20	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1548	CARACTERÍSTICAS DE LA UNIDAD AGROPECUARIA AL DÍA DE LA ENTREVISTA		
21	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1549	CULTIVOS INSTALADOS O SEMBRADOS		
22	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1550	USOS DE LA TIERRA		
23	2019	62	701	ENCUESTA NACIONAL AGROPECUARIA	1551	EXISTENCIA DE GANADO, AVES Y OTROS ANIMALES		

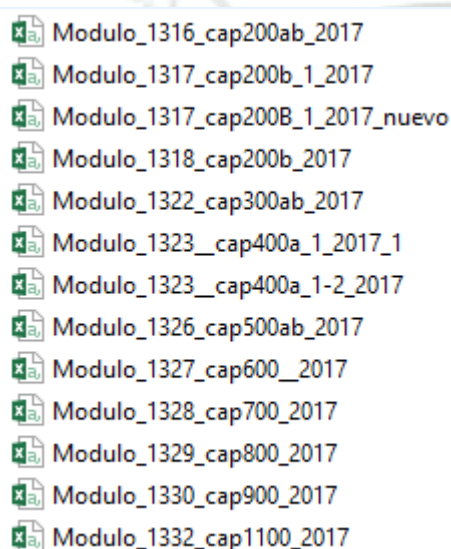
Nota. Datos de la Encuesta Nacional Agropecuaria divididos en módulos, en formato sav, de SPSS.

2) Excel

Con la finalidad de contar con información más fácil de utilizar, se utilizó Excel debido a que su formato es mucho más compatible con varios tipos de softwares, es por eso que los archivos que se utilizaron están en formato xls, y de esta manera sean más fáciles de exportar y utilizar en las siguientes etapas del proceso, donde será necesaria para realizar la transformación y presentación de datos, utilizando las herramientas de Python y Power BI.

Figura 2

Módulos almacenados en formato CSV.



Nota. Módulo transformados y almacenados de formato SAV a CSV.

3) Python

Se utilizó Python para el proceso de transformación, esta herramienta brinda muchas ventajas para realizar minería de datos, ya que, gracias a sus numerosas librerías, facilita la implementación de algoritmos de clustering, y facilita la manipulación de matrices, haciendo más rápido su procesamiento a diferencia de otros softwares y adaptando el código de acuerdo a nuestras necesidades; Para la implementación de algoritmos de clustering se emplearon las siguientes librerías.

- **Pandas:** Pandas es una librería que ayuda a la carga de datos, desde distintas fuentes, como CSV, archivos de texto plano, cargas desde bases de datos SQL, entre otros, la librería Pandas facilita la manipulación de estos datos, mediante el uso de su estructura principal, el DataFrame.
- **NumPy:** Es una librería que da soporte para crear vectores y matrices grandes multidimensionales, además cuenta con una gran colección de funciones, que permiten realizar operaciones matemáticas, complejas sobre arrays multidimensionales, esta librería es ideal para poder implementar clústeres, ya que al implementar algoritmos de clustering se utilizan muchas matrices y se requiere realizar muchas operaciones dentro de estas.
- **Matplotlib:** Es una librería que trabaja en conjunto con NumPy, y que es utilizada para generar y personalizar gráficos a partir de datos contenidos en arrays, en el proyecto se emplearon para poder graficar los dendrogramas y visualizar las características de los clústeres generados.
- **Scipy:** Es una librería que se caracteriza por contener herramientas y algoritmos matemáticos, para optimizar operaciones matemáticas, álgebra lineal, interpolación, funciones espaciales, procesamiento de imágenes y muchas más operaciones utilizadas en la ciencia y la ingeniería, mediante el uso de esta librería se pudo realizar operaciones para poder determinar la matriz de distancias entre los elementos del DataFrame y poder implementar el método de asociación en los clústeres.

4) **Power BI**

Esta herramienta proporciona visualizaciones interactivas y capacidades de inteligencia empresarial, cuenta con una interfaz simple e interactiva que facilita la creación de informes y paneles a los usuarios finales, es por eso que para el presente proyecto se aplica para poder presentar las gráficas y la información, de tal manera que sea más fácil de entender e interpretar, se utilizaron tablas, filtros, gráficos circular, matrices, gráfico de barras, Mapbox visual, entre otras visualizaciones.

CAPÍTULO II

2. DESARROLLO DE LA PROPUESTA

Como se indicó anteriormente, para la implementación de nuestro modelo de clúster, seguiremos los pasos de la metodología KDD (Knowledge Discovery in Databases), donde incluiremos 6 etapas: Comprensión del área de estudio y fijación de objetivos, recopilación y selección de datos, limpieza y procesamiento de los datos, aplicación de minería de datos, interpretación y análisis de patrones encontrados y finalmente la aplicación del conocimiento para la toma de decisiones.

2.1. Comprensión del área de estudio y fijación de objetivos

En esta primera etapa se van a definir los límites y objetivos de lo que pretendemos estudiar, se definirá las principales fuentes de información, y se va dimensionar la cantidad de datos y formatos que se utilizarán a lo largo de la investigación.

El principal objetivo de la investigación es segmentar el sector agropecuario basándonos en la productividad, economía y desarrollo del sector, para la toma de decisiones, esta información generada será de utilidad para los productores agropecuarios, entidades públicas y privadas.

2.2. Recopilación y selección de datos

Los datos recopilados son los que nos proporciona el INEI, de la Encuesta Nacional Agropecuaria aplicada a pequeños y medianos productores agropecuarios, en los años 2015, 2016, 2017, 2018 y 2019, esta encuesta está dividida en módulos, de estos módulos los seleccionados para esta investigación son los siguientes:

A. Módulo 200 AB Destino de los productos cosechados

- **Variable P223:** ¿Cuál fue el destino de venta de su producción?

Este módulo nos va permitir determinar cuáles son los principales destinos de los productos cosechados como por ejemplo el mercado local, el mercado regional, el mercado exterior, la agroindustria, mercados de Lima u otros.

B. Módulo 200AB_1 Pérdida total de producción

- **Variable P224E:** ¿Cuál fue el motivo de la pérdida total de producción o superficie?
- **Variable P229:** ¿Porque considera el cultivo más importante?
- **Variable P229A:** ¿Por qué motivo no vendió su producción?

En este módulo nos permitirá identificar cuáles fueron los motivos que causaron la pérdida total de la producción de los productores, los cuales pueden ser huaycos, inundaciones, plagas, heladas, granizada, sequía u otros, también se creará grupos con los factores por lo que los productores consideran cultivar determinadas cosechas, pueden ser por el buen precio de venta, el cultivo es adaptable a condiciones climáticas, se requiere poca agua para el cultivo, su producto tiene un mercado seguro, tiene un corto periodo vegetativo, lo utiliza para alimentar a sus animales, los utiliza para el autoconsumo u otro factor, otro clúster que se formará en este módulo son los motivos por los cuales los productores no vendieron la totalidad de su producción, estos pueden ser porque lo utilizaron para el autoconsumo, para alimentar a sus animales, porque tuvieron una producción baja, por falta de acceso o medio de transporte, lo utilizaron para semilla u otro factor.

C. Módulo 400A Producción pecuaria

- Variable P403AA: ¿La venta de ganado la realizó en?

Al analizar los datos de este módulo se podrá crear clústeres e identificar los puntos principales donde los productores pecuarios optan por vender su producción, los cuales son: donde los cría, en el mercado local, matadero local, mercado regional, matadero regional, matadero de Lima, Mercado exterior, agroindustria, mercado de Lima u otro.

D. Módulo 600 Inocuidad agroalimentaria

- Variable P601: ¿Qué hace con los residuos de cultivos?
- Variable P602: ¿Qué hace con los residuos de sus animales de crianza?
- Variable P605: ¿Qué entidad del estado le informó?

En este módulo se puede aprovechar la información para determinar problemáticas y oportunidades de desarrollo, en este módulo podremos determinar los grupos de productores y lo

que hacen los residuos de sus cultivos, los quema, lo usan de alimento para sus animales, lo botan, lo utilizan de abono, lo dejan en el campo, u otra actividad, también podremos determinar lo que se hacen con los residuos de sus animales de crianza, lo queman, lo entierran, lo botan, lo usan de abono, lo usan como combustible, lo dejan en el campo, son productores apícolas, no son productores pecuarios o realizan otra actividad, por otra parte también sabremos que entidades del gobierno informan a los productores agrícolas acerca de la inocuidad agroalimentaria, SENASA, Ministerio de Salud, MIDAGRI, Gobierno Regional, Gobierno Local/Municipal, u otra entidad.

E. Módulo 700 Servicios de extensión agraria

- Variable P710: ¿Qué información agropecuaria necesita para desarrollar su actividad?
- Variable P713: ¿A través de qué medios de comunicación tuvo acceso a la información sobre la prevención de los daños causados por los eventos climáticos?

Muchos productores requieren información adicional para realizar sus actividades agropecuarias, en este caso requieren información acerca del clima, manejo de enfermedades y plagas, precios de insumos agropecuarios, precios de venta, técnicas de manejo de cultivo y crianza, técnicas de manejo de semilla, uso de abonos y fertilizantes, técnicas de sistema de riego, análisis de suelo, u otra información, con esta información se crearán clústeres que agrupará productores agropecuarios con requerimiento de información similares, por otra parte también se ve una necesidad por parte de los agricultores de tener información acerca de los fenómenos climáticos para poder prevenir los daños causados por estos, y con la variable P713 se podrán determinar cuáles fueron los principales canales por los que los productores pueden acceder a esta información.

F. Módulo 900 Servicios financieros

- Variable P903: ¿Quién le proporcionó el crédito que obtuvo?
- Variable P904: ¿Utilizó el crédito para?

Finalmente, en este módulo podremos ver los aspectos financieros, donde se crearán grupos de productores de acuerdo a las fuentes de donde obtienen sus créditos para realizar sus actividades, los considerados y más comunes son entidades como AGROBANCO, Caja Municipal, Caja

Rural, Banca Privada, Financiera/EDPYME, ONG, Cooperativa, establecimiento comercial, prestamistas, programas del estado u otro, y para qué utilizan estos préstamos, para comprar insumos agrícolas, pagar mano de obra, asistencia técnica, alquiler de maquinaria, compra de terrenos agrícolas, compra de ganado, u otra actividad.

Esta información está almacenada y dividida en módulos, en formato SAV perteneciente al software estadístico SPSS, y por cuestiones de compatibilidad se decidió transformarlo a formato CSV, ya que es un formato más compatible para trabajar con las herramientas que estamos utilizando como Python y Power Bi.

Al finalizar la identificación de nuestros objetivos, la selección e integración de nuestros datos relevantes para el estudio, tendremos varios dataset de los productores agropecuarios y sus características dependiendo del módulo que se esté trabajando.

2.3. Limpieza y procesamiento de datos

La naturaleza de la encuesta realizada, genera que algunos productores encuestados cumplan cierto requisito para poder responder a una pregunta, es por eso que para ciertas preguntas las respuestas estén en blanco, como por ejemplo existen un porcentaje de productores que solo realizan actividades agrícolas, otros que solo realizan actividades pecuarias y otro grupo que realiza actividades tanto agrícolas como pecuarias, que se les denomina productores agropecuarios, asimismo existen preguntas que están dirigidas sólo uno de estos tres tipos, por lo que los demás omiten esta pregunta y continúan con la siguiente pregunta, para esto se realizó la limpieza de datos, utilizando Python, y se tomó en cuenta sólo los productores que cumplen con el requisito de esta pregunta.

Las preguntas seleccionadas son de tipo binarias o dicotómicas, es decir que solo toman el valor de 0 o 1, en muchos casos la encuesta los considera con 1 si el productor pertenece a una categoría o con un 2 si no pertenece a la categoría, y en este caso para un mejor funcionamiento práctico, a la hora de aplicar el algoritmo, nos conviene trabajar con los valores de 0 que representaría la ausencia de pertenencia a una categoría y con 1 que representaría la presencia del productor en una categoría, para esto simplemente se reemplazó los valores de 2 por 0, y una vez finalizado esto los datos quedarían listos, y son procesado utilizando un DataFrame como una matriz, listos para aplicar el algoritmo de minería de datos.

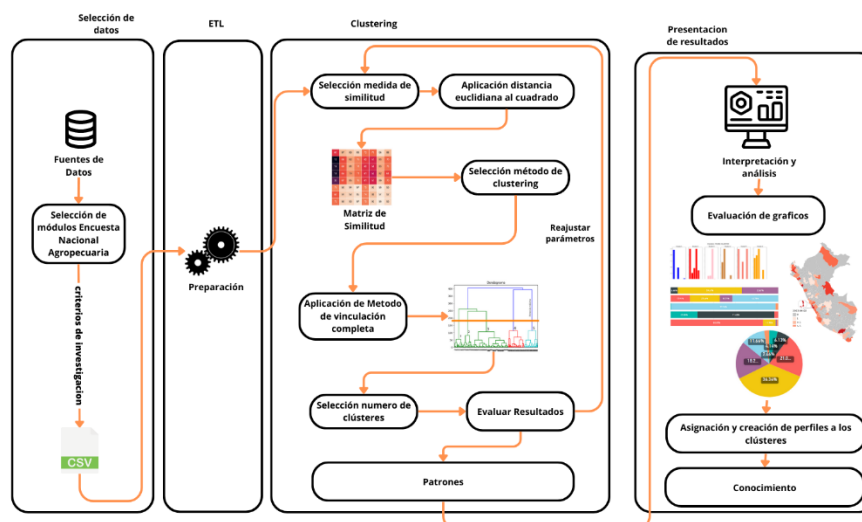
2.4. Aplicación de minería de datos

Para aplicar el algoritmo de clustering se tiene que evaluar el algoritmo que mejor convenga de acuerdo al tipo de variables que estamos trabajando y teniendo en cuenta factores que se utilizan en todos los algoritmos de clustering, los cuales son, la medida de distancias o similitudes y el método de agrupación, por otro lado, también se debe tener en cuenta el tipo de variables con las que se está trabajando, que en este caso son variables binarias o dicotómicas, que expresan cualidades o categorías.

Al evaluar el tipo de variables el algoritmo más indicado a utilizar es algoritmo de clustering jerárquico, ya que este algoritmo trabaja con un árbol jerárquico binario el cual facilita la visualización al crear los conglomerados, después de definir el algoritmo a utilizar se debe elegir una medida de distancia o disimilitud, con esta medida se creará una matriz de distancias, y se procederá al siguiente paso el cual es seleccionar un método de agrupamiento el cual utilizará la matriz de distancias para conformar los grupos, realizando una evaluación se determinó que la mejor opción para el agrupamiento es el método del vecino más lejano, este método realiza el agrupamiento con la finalidad que los elementos que pertenecen a un mismo clúster sean los más parecidos entre sí y los elementos que pertenecen a clústeres diferentes sean los más diferentes, a continuación se describe más detalladamente la implementación del algoritmo, con los ajustes que se realizaron para tener mejores agrupamientos.

Figura 3.

Modelo propuesto de clustering



Nota. Modelo del análisis de clustering para la generación de conglomerados

2.4.1. Aplicación de algoritmo

En las siguientes líneas de código se procesa el DataFrame, el cual contiene la variable P224E que es “Productores que perdieron la totalidad de sus cosechas”, donde en la variable de código “cuadradodist_condensado” almacena la matriz de distancias, utilizando la medida de distancia sqeuclidean (distancia euclidiana al cuadrado),

Figura 4

Aplicación de algoritmo de clustering

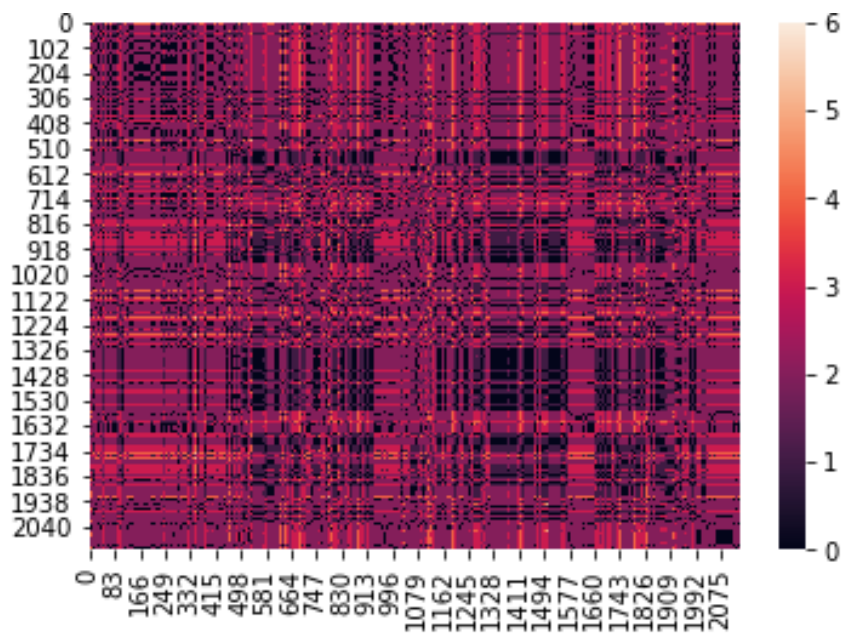
```
df=pd.read_csv("E:/Modulos/2019/701-Modulo1532/X04_Cap200b_1_2019.csv",low_memory=False)
print(df.info())
ext_link=['INDEX','P224E_1','P224E_2','P224E_3','P224E_4','P224E_5','P224E_6','P224E_7']
variable="P224E CLUSTER"
extraccion_df=df[ext_link]
extraccion_df=extraccion_df[df.CODIGO==1]
extraccion_df_final=extraccion_df
extraccion_df_final=extraccion_df_final.drop(['INDEX'], axis = 1)
cuadradodist_condensado=pdist(extraccion_df_final,'sqeuclidean')
x=squareform(cuadradodist_condensado)
plt.figure(figsize = (100,100))
sns.heatmap(x)
```

Nota. Código para generar la matriz de distancias, utilizando la distancia euclidiana al cuadrado.

Este código nos mostrará la matriz de distancias, en la siguiente imagen podemos observar un mapa de calor donde los valores oscuros indican que hay una distancia muy baja y los valores claros tiene una distancia alta.

Figura 5

Matriz de distancias.



Nota. Heatmap con las distancias entre las observaciones.

Utilizando la matriz de distancias, especificamos el método de agrupación que utilizaremos, en la variable “clúster_jerarquico” recibe dos parámetros, el “cuadradocondensado_dist” que es la matriz de distancias y también recibe el parámetro del método que utilizaremos para la agrupación el cual es “complete”.

Figura 6

Código que grafica el dendrograma.

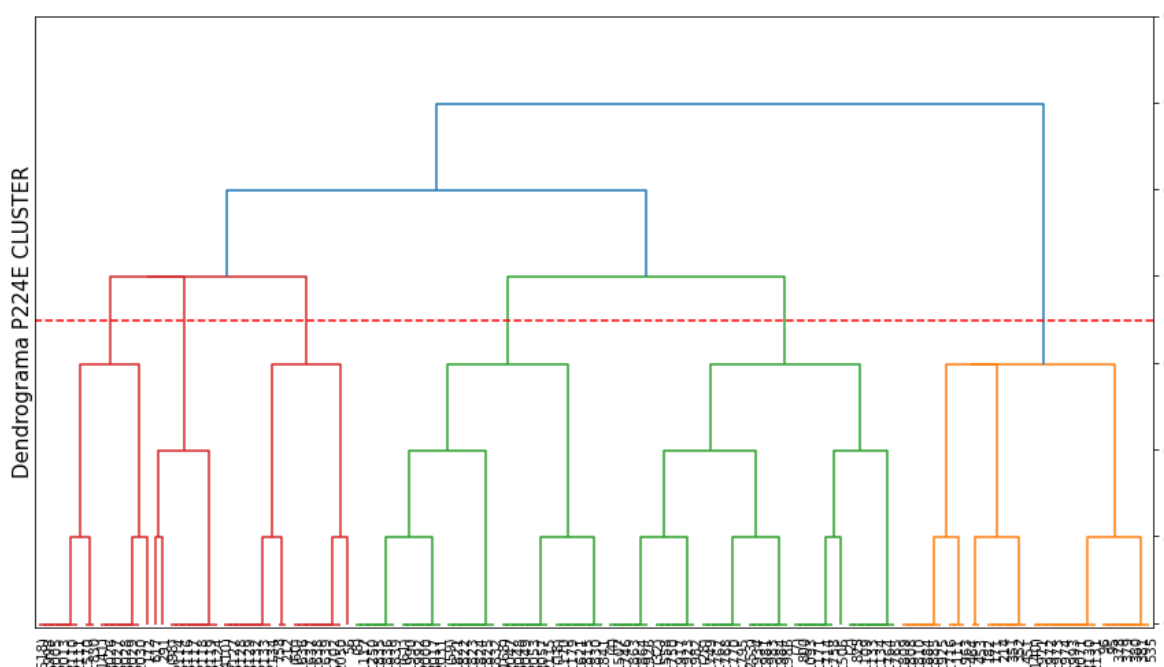
```
fig, axes = plt.subplots(figsize=(8, 15)) #TAMAÑO DEL GRAFICO
cluster_jerarquico = linkage(cuadrado_dist_condensado, 'complete')
dendrogram = sch.dendrogram(cluster_jerarquico, p=10,
                             truncate_mode='level',
                             orientation='right',
                             leaf_font_size=10,
                             leaf_rotation=0,
                             show_contracted=False,
                             above_threshold_color='C0')

n = (np.amax(cuadrado_dist_condensado) + 1)
plt.xlim(-0.05, n)
plt.title("Dendrograma "+variable, fontsize=15)
linea_corte = 3.5
plt.axvline(linea_corte, color="red", linestyle='--')
cluster = fcluster(cluster_jerarquico, t=linea_corte, criterion='distance')
extraccion_df[variable] = cluster
nro_de_clusters = np.max(cluster)
```

Nota. Código que genera el dendrograma utilizando el método de vecino más lejano.

Con estos parámetros configurados graficamos el dendrograma con su línea de corte, en este caso la línea de corte está establecida en 3.5, esta línea de corte cada vez que interseca las líneas verticales determinará el número de clústeres que se están conformando, en la siguiente imagen la línea de corte interseca 6 líneas verticales, lo que quiere decir que se conformaron cuatro clústeres.

Figura 7
Matriz de distancias.



Nota. Dendrograma que grafica las agrupaciones de las observaciones.

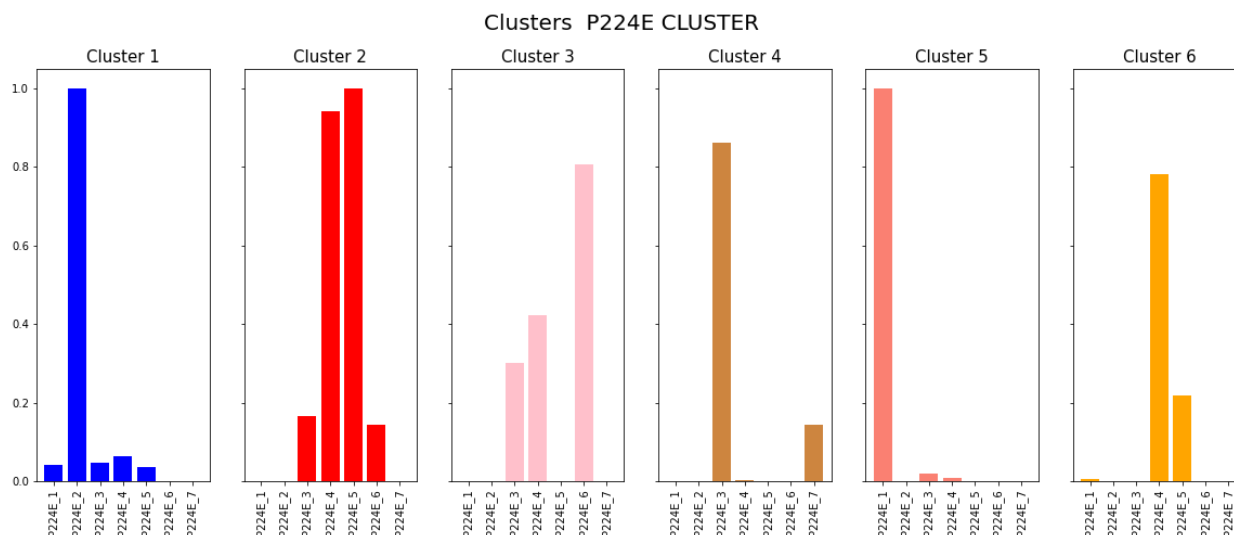
2.4.2. Validación del modelo

Al obtener los grupos se debe evaluar la calidad de los clústeres para garantizar su éxito, el primer paso será observar los agrupamientos, como se puede observar en la siguiente imagen a primera vista podemos observar que cada grupo es diferente entre sí, lo cual nos puede indicar que podemos tener un buen agrupamiento, ayudándonos del dendrograma podemos afirmar que los elementos que pertenecen al mismo clúster se diferencian máximo en tres categorías, ya que nuestra línea de corte está en 3.5, en el siguiente grafico se tiene como se caracteriza la variable P224E, estas variables responde a una pregunta la cual es ¿Cuál fue el motivo por el cual perdió la totalidad de su producción? Y las variables son:

- P224E_1, fue afectada por huaycos, deslizamientos o derrumbes.
- P224E_2, fue afectado por inundaciones.
- P224E_3, fue afectado por plagas y/o enfermedades.
- P224E_4, fue afectado por heladas.
- P224E_5, fue afectado por granizada.

- P224E_6, fue afectado por sequía.
- P224E_7, Otro.

Figura 8
Características de los clústeres.



Nota. Características de los clústeres, resultantes de la agrupación.

Como podemos ver en el clúster 1, se caracteriza por productores que perdieron su producción por inundaciones, el clúster dos se caracteriza por productores que perdieron su producción por heladas y granizadas, el clúster 3 por sequías, plagas y heladas, el clúster 4 afectado por plagas u otra enfermedad, el clúster 5 por huaycos, y el clúster 6 por productores que perdieron la totalidad de sus cosechas por heladas y granizadas, además podemos observar que hay cierta similitud entre el clúster 6 y 2, ya que ambos tienen factores como las heladas y granizadas, pero en este caso debemos centrarnos en cómo se diferencian estos dos clústeres, en el caso del clúster 2, se caracteriza principalmente por la presencia de heladas y granizadas, acompañadas ligeramente de plagas y sequías, y en el clúster 6 únicamente se caracteriza por la presencia de heladas y granizadas.

Otro método de validación es usando la matriz de similitudes, donde se puede inspeccionar visualmente, esta matriz está clasificada por colores a menor distancia tiene color rojo y a mayor distancia tenemos el color azul, los rectángulos enmarcados son los pertenecientes a un solo clúster, los cuales son 6 y podemos ver que la distancia dentro de estos rectángulos es mínima mientras que los que no están enmarcados tienen en su mayoría valores altos.

Figura 9
Matriz de distancias

	distancia euclídea al cuadrado																											
	1:a	2:b	3:c	4:d	5:e	6:f	7:g	8:h	9:i	10:j	11:k	12:l	13:m	14:n	15:o	16:p	17:q	18:r	19:s	20:t	21:u	22:v	23:w	24:x	25:y	26:z	27:ab	28:ac
1:a	0	2	1	1	3	2	3	3	4	3	4	3	3	4	3	4	4	5	2	1	2	3	3	4	2	3	3	4
2:b	2	0	1	3	3	2	3	1	2	3	2	3	3	4	3	4	4	5	2	3	2	1	3	2	2	1	1	2
3:c	1	1	0	2	2	3	4	2	3	4	3	4	4	5	4	5	5	6	3	2	3	2	4	3	3	2	2	3
4:d	1	3	2	0	2	1	2	2	3	2	3	4	4	5	2	3	3	4	3	2	3	4	4	5	3	4	4	5
5:e	3	3	2	2	0	3	4	2	3	4	3	4	4	5	2	3	3	4	5	4	5	4	6	5	3	2	4	3
6:f	2	2	3	1	3	0	1	1	2	1	2	3	3	4	1	2	2	3	2	3	2	3	3	4	2	3	3	4
7:g	3	3	4	2	4	1	0	2	1	2	3	2	4	3	2	1	3	2	3	4	1	2	2	3	3	4	4	5
8:h	3	1	2	2	2	1	2	0	1	2	1	4	4	5	2	3	3	4	3	4	3	2	4	3	3	2	2	3
9:i	4	2	3	3	3	2	1	1	0	3	2	3	5	4	3	2	4	3	4	5	2	1	3	2	4	3	3	4
10:j	3	3	4	2	4	1	2	2	3	0	1	4	2	3	2	3	1	2	1	2	3	4	2	3	3	4	2	3
11:k	4	2	3	3	3	2	3	1	2	1	0	5	3	4	3	4	2	3	2	3	4	3	3	2	4	3	1	2
12:l	3	3	4	4	4	3	2	4	3	4	5	0	2	1	2	1	3	2	3	4	1	2	2	3	1	2	4	3
13:m	3	3	4	4	4	3	4	4	5	2	3	2	0	1	2	3	1	2	1	2	3	4	2	3	1	2	2	1
14:n	4	4	5	5	5	4	3	5	4	3	4	1	1	0	3	2	2	1	2	3	2	3	1	2	2	3	3	2
15:o	3	3	4	2	2	1	2	2	3	2	3	2	2	3	0	1	1	2	3	4	3	4	4	5	1	2	4	3
16:p	4	4	5	3	3	2	1	3	2	3	4	1	3	2	1	0	2	1	4	5	2	3	3	4	2	3	5	4
17:q	4	4	5	3	3	2	3	3	4	1	2	3	1	2	1	2	0	1	2	3	4	5	3	4	2	3	3	2
18:r	5	5	6	4	4	3	2	4	3	2	3	2	2	1	2	1	1	0	3	4	3	4	2	3	3	4	4	3
19:s	2	2	3	3	5	2	3	3	4	1	2	3	1	2	3	4	2	3	0	1	2	3	1	2	2	3	1	2
20:t	1	3	2	2	4	3	4	4	5	2	3	4	2	3	4	5	3	4	1	0	3	4	2	3	3	4	2	3
21:u	2	2	3	3	5	2	1	3	2	3	4	1	3	2	3	2	4	3	2	3	0	1	1	2	2	3	3	4
22:v	3	1	2	4	4	3	2	2	1	4	3	2	4	3	4	3	5	4	3	4	1	0	2	1	3	2	2	3
23:w	3	3	4	4	6	3	2	4	3	2	3	2	2	1	4	3	3	2	1	2	1	2	0	1	3	4	2	3
24:x	4	2	3	5	5	4	3	3	2	3	2	3	3	2	5	4	4	3	2	3	2	1	1	0	4	3	1	2
25:y	2	2	3	3	3	2	3	3	4	3	4	1	1	2	1	2	2	3	2	3	2	3	3	4	0	1	3	2
26:z	3	1	2	4	2	3	4	2	3	4	3	2	2	3	2	3	3	4	3	4	3	2	4	3	1	0	2	1
27:ab	3	1	2	4	4	3	4	2	3	2	1	4	2	3	4	5	3	4	1	2	3	2	2	1	3	2	0	1
28:ac	4	2	3	5	3	4	5	3	4	3	2	3	1	2	3	4	2	3	2	3	4	3	3	2	2	1	1	0

Nota. Matriz de distancias donde se observan 6 clústeres.

Otro método de validación es usando las medias de cohesión y separación, usando la tabla anterior calculamos el promedio de distancias entre los elementos del mismo clúster, esta medida será la cohesión, y también calcularemos el promedio de distancias de los elementos del mismo clúster con los elementos de los demás clústeres, el cual es la separación. Como se ve en la siguiente imagen la distancia de los elementos que comparten el mismo clúster es menor y la distancia de los elementos de un clúster con otro clúster es mayor, por lo que se puede decir que se tiene clústeres aceptables.

Figura 10

Índice de cohesión y separación.

	Cohesión	Separación
Clúster 1	2	3.209
Clúster 2	1.6666667	2.894
Clúster 3	1.7142857	3.136
Clúster 4	1	2.808
Clúster 5	1.3333333	2.917
Clúster 6	1.6666667	2.792

Nota. Matriz de cohesión y separación, indica que es una agrupación aceptable.

Al validar los clústeres se procede a etiquetar las observaciones, y guardar en formato CSV para posteriormente integrarlos en Power Bi para su análisis e interpretación.

Figura 11

Módulos clusterizado cargados en Power Bi

 200AB
 200B_1
 200B_2
 400A_1
 400A_2
 500AB
 600
 700
 800
 900
 1100

Nota. Módulos clusterizado, listos para su visualización utilizando Power Bi.

Dado que la finalidad del proceso KDD, es la generación de conocimiento en el próximo capítulo se realizará la interpretación de los resultados obtenidos, aprovechando las herramientas de visualización de Power Bi.

CAPÍTULO III

3. RESULTADOS

3.1. ANÁLISIS DESCRIPTIVO

Siguiendo con la metodología KDD en este capítulo llevaremos a cabo el análisis y la interpretación de los 6 módulos con 13 variables de investigación del año 2015 al año 2019, descritos en el punto 2.2 y llevaremos a cabo el cumplimiento de nuestros objetivos específicos E y F.

A partir de este análisis se obtuvo resultados para poder describir detalladamente las características y patrones más comunes de los productores agropecuarios distribuidos en todo el Perú, donde también se puede ver los cambios entre los años 2015-2019, en ese sentido se escogerá un departamento sobre el cual haremos un análisis más profundo, el cual servirá como referencia para poder analizar cada departamento que escojamos.

3.1.1. Módulo 200 AB Destino de los productos cosechados

A. Variable P223 ¿Dónde vendió su producción?

Este módulo nos va permitir determinar cuáles son los principales destinos de los productos cosechados como por ejemplo el mercado local, el mercado regional, el mercado exterior, la agroindustria, mercados de lima u otros.

1) Selva

En el periodo 2015-2019 los mercados locales fueron los principales destinos de venta de los pequeños y medianos productores.

En la siguiente gráfica se tiene la descripción de los clústeres clasificados por año, número de clústeres, variables, y la cantidad de productores clasificado por año, por ejemplo, tomando en cuenta el año 2015 podemos observar en la primera fila que pertenece al clúster 4, donde este clúster se caracteriza por tener productores principalmente que venden su producción a mercados locales, mercados regionales y en su minoría la venden al mercado exterior o mercado de Lima, o lo destinan al mercado de Lima, en este clúster se tiene un aproximado de 8919 observaciones en toda la región selva del Perú, y así de esta manera se puede interpretar cada fila de esta tabla. También se puede observar que a lo largo de los 5 años la mayoría de los productores vendieron su producción en el mercado local con una amplia diferencia respecto a otras variables, después la vendieron en el mercado regional, mercado de Lima, mercado exterior, no sabe y una minoría a la agroindustria, como se ve en la siguiente imagen.

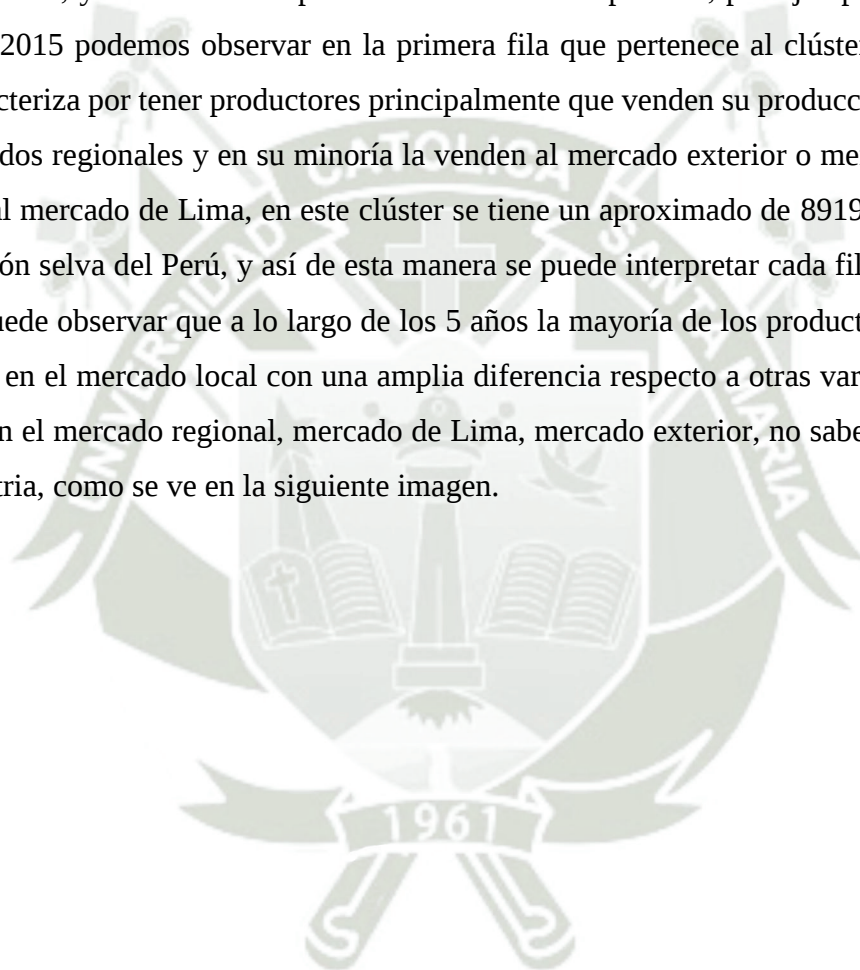


Figura 12.
Características de los clústeres en la selva

AÑO	SELVA_P223	P223_1 Mercado Local	P223_2 Mercado Regional	P223_3 Mercado Exterior	P223_4 Agroindus- tria	P223_5 Mercado Lima	P223_6 No Sabe	Cantidad Observaciones
2015	4	8,919	678	80	7	35	0	8,919
2015	3	0	1,053	0	0	481	672	2,176
2015	1	0	31	379	0	21	0	379
2015	2	0	13	32	183	31	0	183
2016	5	9,232	0	0	26	70	0	9,238
2016	3	737	2,380	0	9	144	0	2,380
2016	2	40	142	819	0	0	0	819
2016	1	0	0	0	0	0	644	644
2016	4	0	1	32	146	473	0	618
2017	1	9,733	0	19	3	0	0	9,733
2017	3	426	1,602	3	0	70	0	1,602
2017	5	0	0	0	0	0	679	679
2017	6	121	0	8	5	600	0	600
2017	4	0	41	430	0	38	0	430
2017	2	11	15	4	69	0	0	69
2018	3	10,495	407	13	0	0	0	10,495
2018	5	12	915	0	1	64	0	915
2018	1	74	0	1	0	680	0	680
2018	4	0	28	453	0	32	0	453
2018	2	10	0	6	229	23	0	229
2019	2	9,399	31	0	27	18	0	9,411
2019	1	51	1,196	443	191	593	0	2,318
2019	3	619	587	59	0	0	0	619
2019	4	0	30	34	7	0	0	34
Total		49,879	9,150	2,815	903	3,373	1,995	63,623

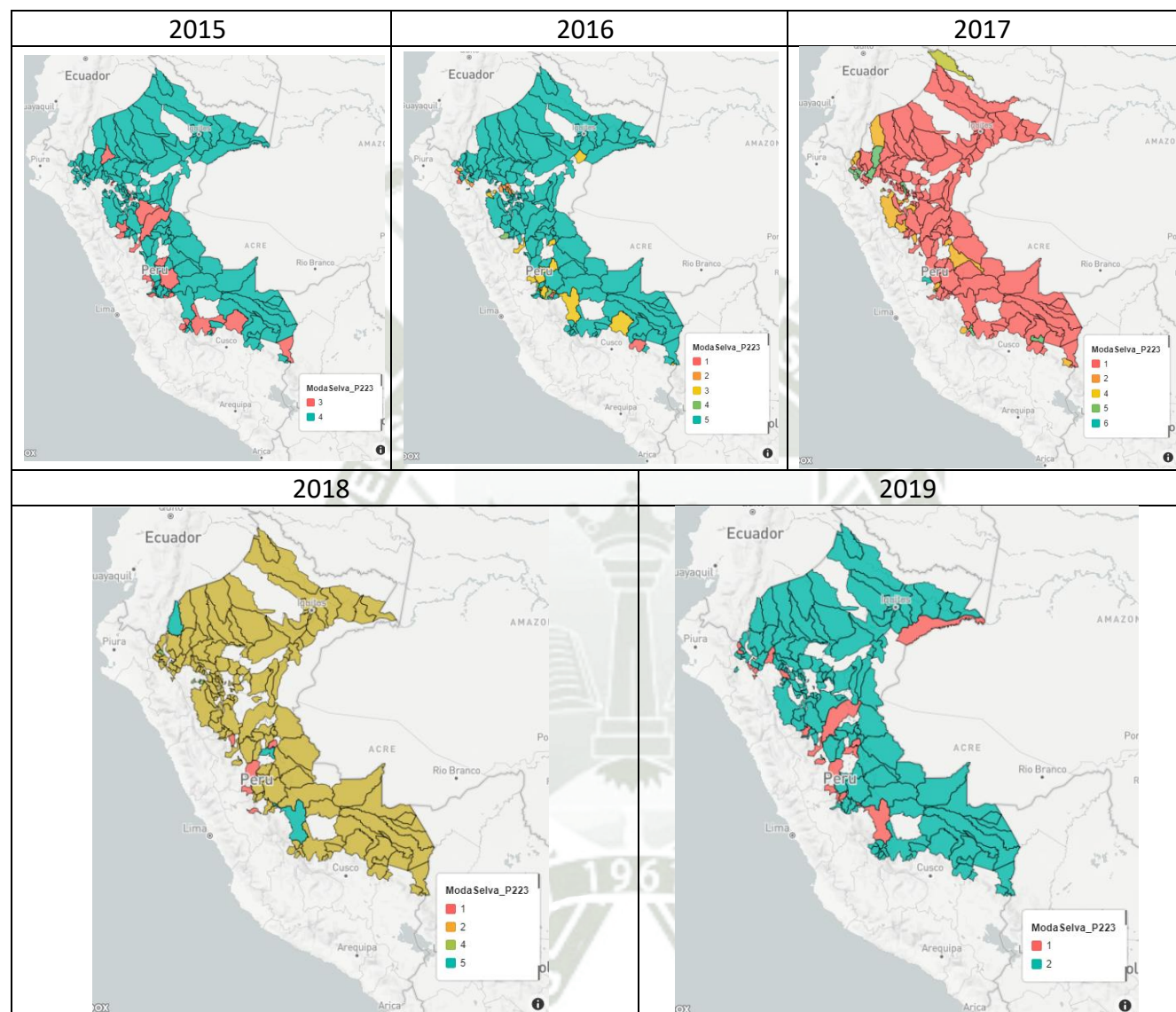
Fuente: Elaboración Propia.

En la siguiente imagen se ve la distribución geográfica de la variable P223 en la selva, a lo largo del año 2015-2019 existe un clúster que predomina en la mayoría de distritos de la selva, este clúster se caracteriza porque sus productores generalmente que ven de su producción únicamente al mercado local, o en el mercado local y a su vez en el mercado regional, exterior o mercado de Lima. El principal destino de la venta de producción es el mercado local es por eso que en los 5 años representa más del 70% de productores como se ve en la figura 6. Analizando este factor podemos interpretar que existe muy poca diversidad en cuanto a la comercialización

de los productos agropecuarios, por parte de los pequeños y medianos productores de la selva del Perú.

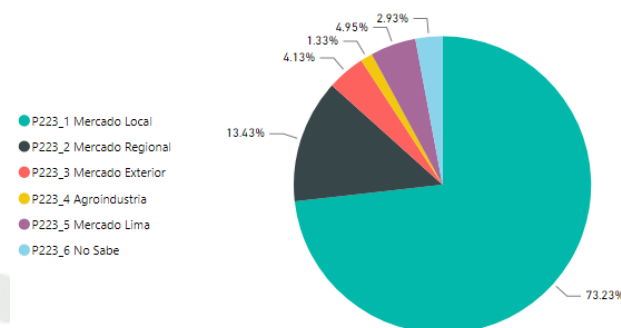
Figura 13.

Nota. Mapa de los clústeres de la selva.



Fuente: Elaboración Propia.

Figura 14.
Distribución de variables selva 2015-2019



Fuente: Elaboración Propia.

Se aplicó el algoritmo de clustering aglomerativo con la medida de distancia euclidiana al cuadrado y el método de agrupación del vecino más lejano, a los datos de los cinco años.

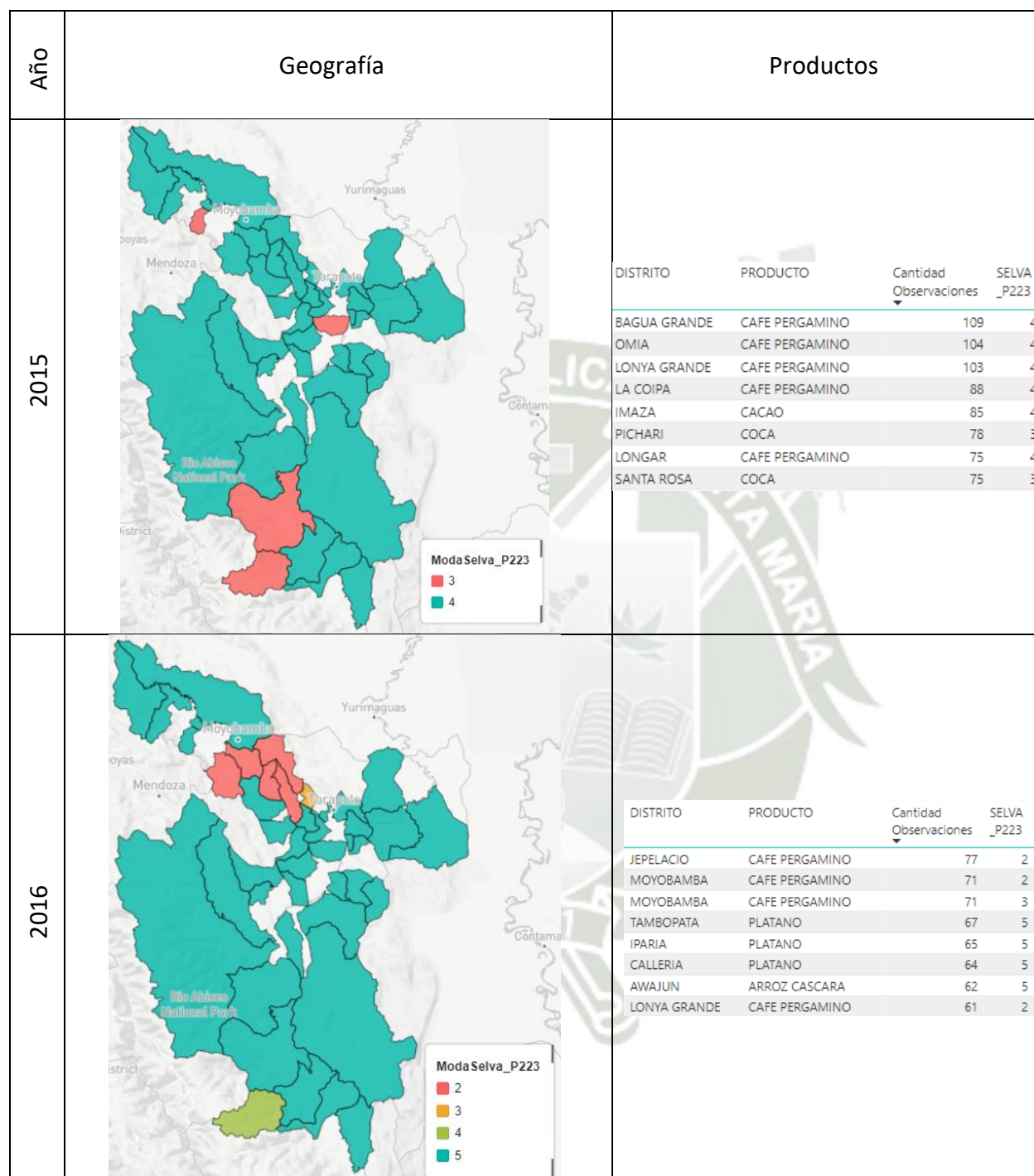
A continuación, se va tomar como ejemplo el departamento de San Martín para describirlo y descubrir los patrones que tiene, y como resultado se obtiene lo siguiente.

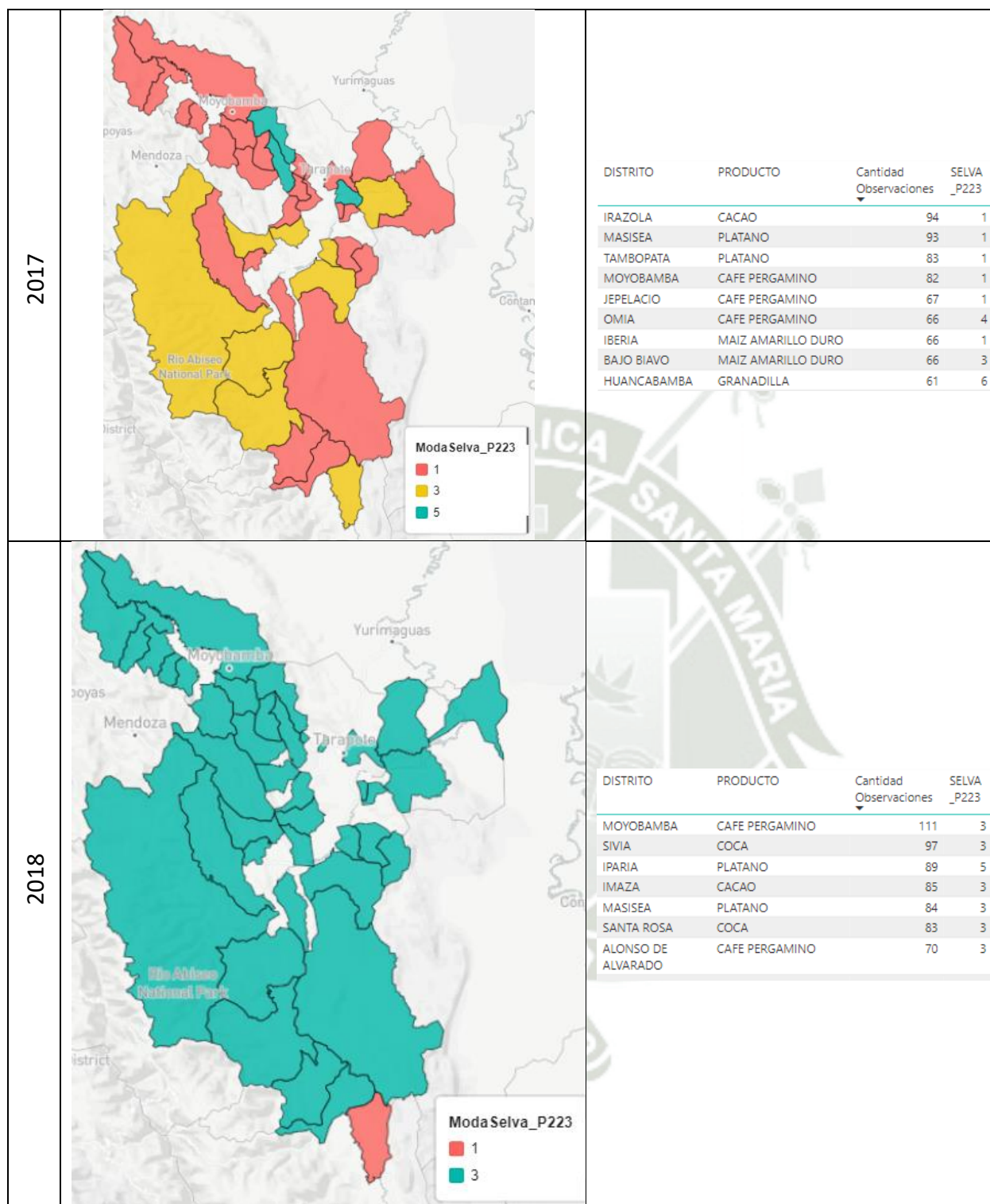
En el departamento de San Martín las observaciones oscilan entre 2953 y 2186, del año 2015 al 2019 el mercado local fue el principal destino de los productos agropecuarios, superando por mucho al mercado regional, que es el segundo con más observaciones, otra variable que tiene un número considerable de productores es el mercado exterior y el mercado de Lima.

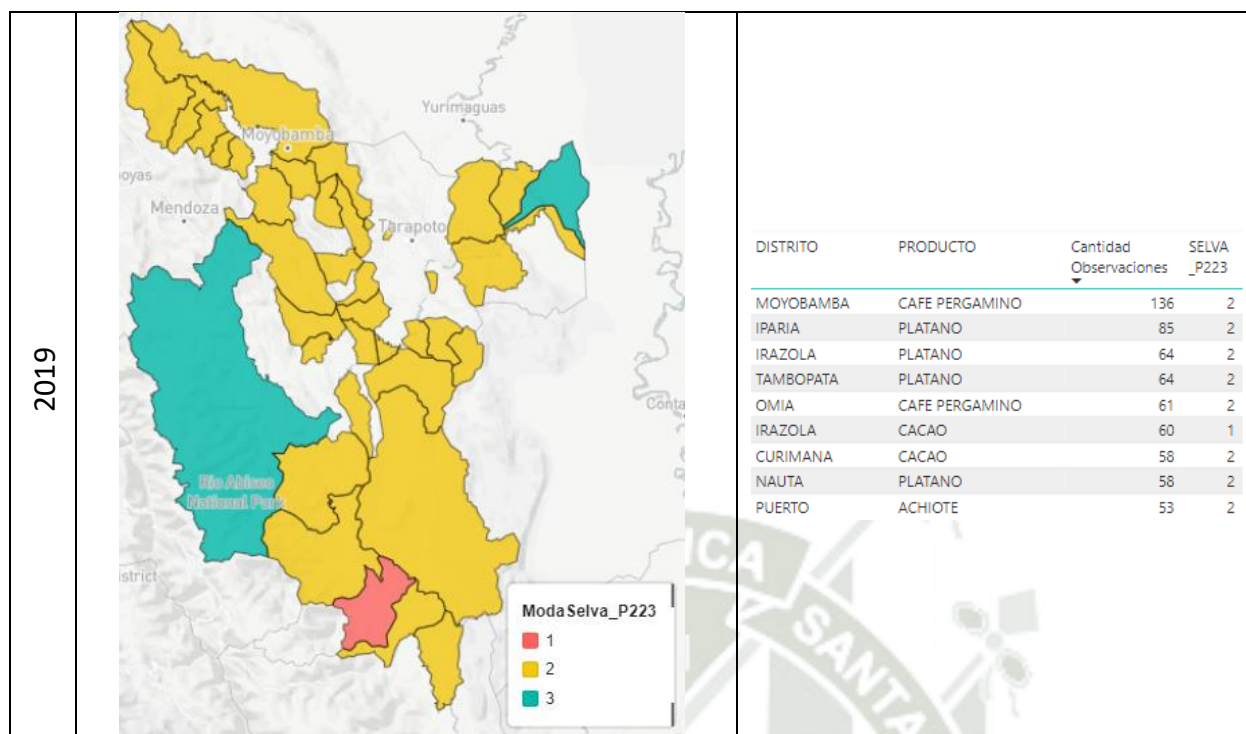
En la figura 7 se observa el mapa de San Martín coloreado utilizando la moda, es decir el clúster que más densidad tiene dentro de cada distrito de San Martín.

Según la nota de prensa del gobierno regional de San Martín, el 16 de febrero de 2022, indica que este departamento es el mayor productor de café a nivel nacional, como se ve en la figura 7, podemos ver el principal distrito productor de cada producto y el clúster al que pertenecen.

Figura 15.
Características del departamento de San Martín



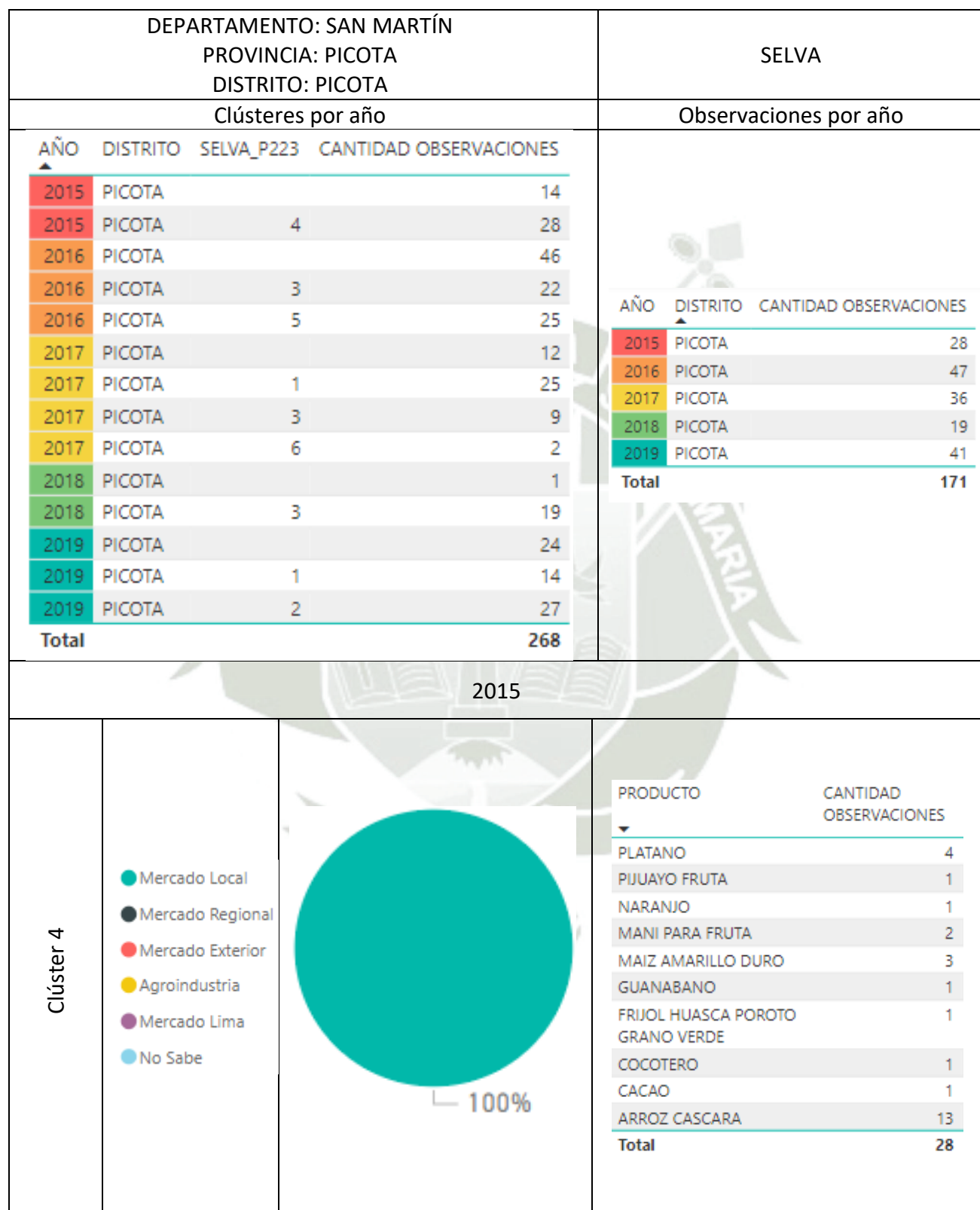




Fuente: Elaboración Propia.

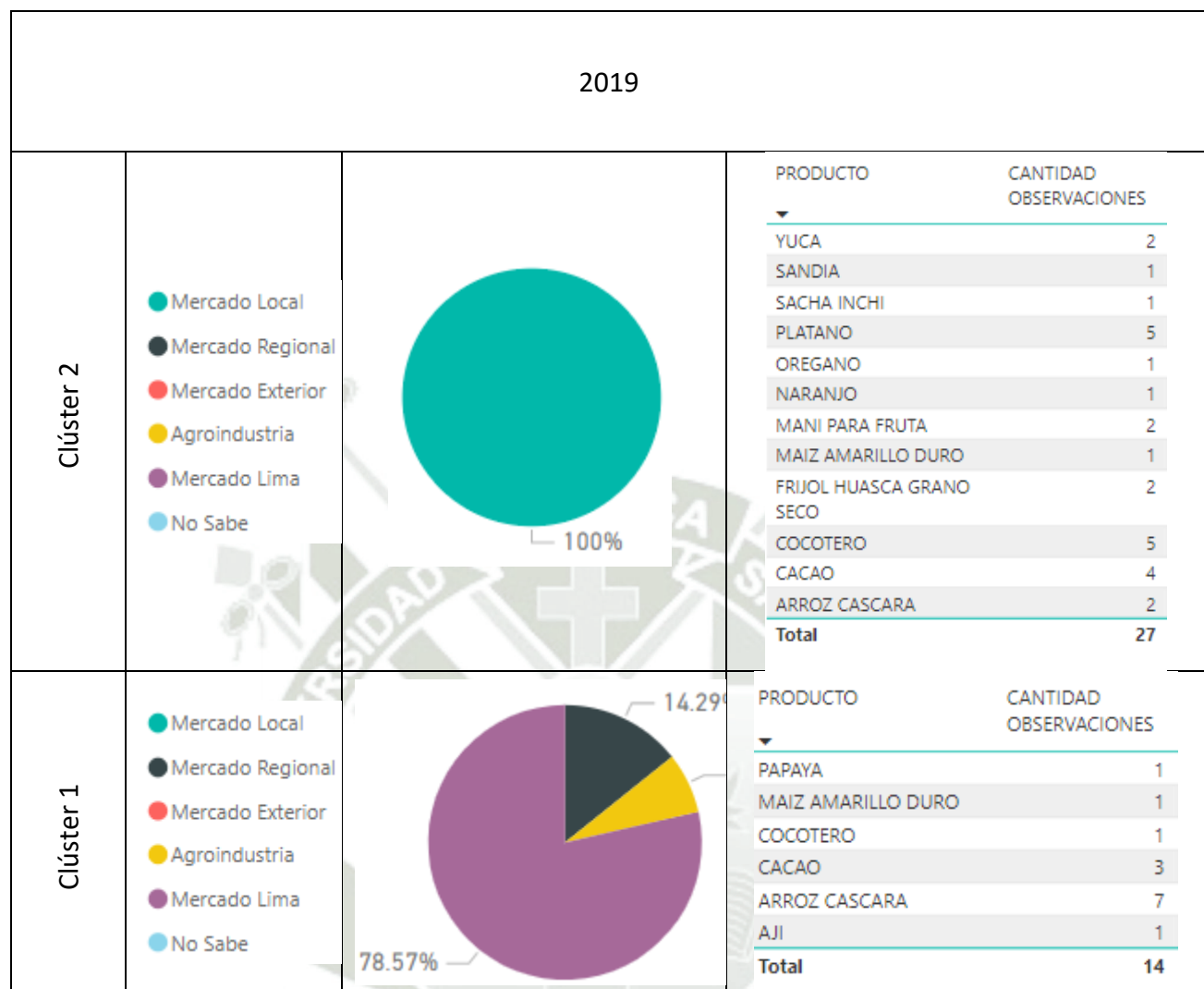
Si observamos en un área geográfica más pequeña como un distrito, en este caso el distrito de Picota, podemos determinar con más grado de detalle sus características, con un total de 268 encuestados y 171 productores clasificados, se puede determinar que en año 2015, la comercialización de sus productos fue únicamente en el mercado local, por lo que en ese año pertenece al clúster 4, en el año 2016, observamos dos clústeres, un clúster de productores que únicamente venden al mercado local, y otro que vende tanto en el mercado local y mercado regional, de igual manera en el año 2017 existen productores que venden al mercado local agrupado al clúster 1, en el clúster 3 agrupados productores que principalmente venden en el mercado regional y/o local, y en este año existe una minoría que vende sus productos al mercado de Lima, en el año 2018 solos e vendió en el mercado local, y por último en el año 2019 la comercialización de productos fue en el mercado regional agrupado en el clúster 2, otro clúster donde se agrupa productores que vendieron al mercado de Lima, al mercado regional y la agroindustria.

Figura 16.
Características distrito de Picota



2016																							
Clúster 5	Mercado Local Mercado Regional Mercado Exterior Agroindustria Mercado Lima No Sabe	100%	<table><tr><th>PRODUCTO</th><th>CANTIDAD OBSERVACIONES</th></tr><tr><td>YUCA</td><td>1</td></tr><tr><td>SANDIA</td><td>1</td></tr><tr><td>PLATANO</td><td>8</td></tr><tr><td>MANI PARA FRUTA</td><td>2</td></tr><tr><td>FRIJOL HUASCA GRANO VERDE</td><td>3</td></tr><tr><td>COCOTERO</td><td>1</td></tr><tr><td>CACAO</td><td>4</td></tr><tr><td>ARROZ CASCARA</td><td>5</td></tr><tr><td>Total</td><td>25</td></tr></table>	PRODUCTO	CANTIDAD OBSERVACIONES	YUCA	1	SANDIA	1	PLATANO	8	MANI PARA FRUTA	2	FRIJOL HUASCA GRANO VERDE	3	COCOTERO	1	CACAO	4	ARROZ CASCARA	5	Total	25
			PRODUCTO	CANTIDAD OBSERVACIONES																			
YUCA	1																						
SANDIA	1																						
PLATANO	8																						
MANI PARA FRUTA	2																						
FRIJOL HUASCA GRANO VERDE	3																						
COCOTERO	1																						
CACAO	4																						
ARROZ CASCARA	5																						
Total	25																						
Clúster 3	Mercado Local Mercado Regional Mercado Exterior Agroindustria Mercado Lima No Sabe	50% 50%	<table><tr><th>PRODUCTO</th><th>CANTIDAD OBSERVACIONES</th></tr><tr><td>YUCA</td><td>2</td></tr><tr><td>PLATANO</td><td>3</td></tr><tr><td>MANI PARA FRUTA</td><td>1</td></tr><tr><td>MAIZ AMARILLO DURO</td><td>1</td></tr><tr><td>FRIJOL HUASCA GRANO VERDE</td><td>1</td></tr><tr><td>CACAO</td><td>3</td></tr><tr><td>ARROZ CASCARA</td><td>11</td></tr><tr><td>Total</td><td>22</td></tr></table>	PRODUCTO	CANTIDAD OBSERVACIONES	YUCA	2	PLATANO	3	MANI PARA FRUTA	1	MAIZ AMARILLO DURO	1	FRIJOL HUASCA GRANO VERDE	1	CACAO	3	ARROZ CASCARA	11	Total	22		
			PRODUCTO	CANTIDAD OBSERVACIONES																			
YUCA	2																						
PLATANO	3																						
MANI PARA FRUTA	1																						
MAIZ AMARILLO DURO	1																						
FRIJOL HUASCA GRANO VERDE	1																						
CACAO	3																						
ARROZ CASCARA	11																						
Total	22																						
2017																							
Clúster 1	Mercado Local Mercado Regional Mercado Exterior Agroindustria Mercado Lima No Sabe	100%	<table><tr><th>PRODUCTO</th><th>CANTIDAD OBSERVACIONES</th></tr><tr><td>YUCA</td><td>1</td></tr><tr><td>SANDIA</td><td>2</td></tr><tr><td>PLATANO</td><td>3</td></tr><tr><td>MANI PARA FRUTA</td><td>3</td></tr><tr><td>MAIZ AMARILLO DURO</td><td>4</td></tr><tr><td>COCOTERO</td><td>5</td></tr><tr><td>CACAO</td><td>3</td></tr><tr><td>ARROZ CASCARA</td><td>4</td></tr><tr><td>Total</td><td>25</td></tr></table>	PRODUCTO	CANTIDAD OBSERVACIONES	YUCA	1	SANDIA	2	PLATANO	3	MANI PARA FRUTA	3	MAIZ AMARILLO DURO	4	COCOTERO	5	CACAO	3	ARROZ CASCARA	4	Total	25
			PRODUCTO	CANTIDAD OBSERVACIONES																			
YUCA	1																						
SANDIA	2																						
PLATANO	3																						
MANI PARA FRUTA	3																						
MAIZ AMARILLO DURO	4																						
COCOTERO	5																						
CACAO	3																						
ARROZ CASCARA	4																						
Total	25																						

Clúster 3	Mercado Local Mercado Regional Mercado Exterior Agroindustria Mercado Lima No Sabe	30.77% 69.23%	PRODUCTO CANTIDAD OBSERVACIONES CACAO1 ARROZ CASCARA8 Total9
	Clúster 6	Mercado Local Mercado Regional Mercado Exterior Agroindustria Mercado Lima No Sabe	100%
2018			
Clúster 3	Mercado Local Mercado Regional Mercado Exterior Agroindustria Mercado Lima No Sabe	100%	PRODUCTO CANTIDAD OBSERVACIONES YUCA1 PLATANO5 MANI PARA FRUTA2 MAIZ AMARILLO DURO2 CATÍ½A DE AZUCAR PARA1 ALCOHOL CACAO5 ARROZ CASCARA3 Total19



Fuente: Elaboración Propia.

2) Costa

En la región de la costa el principal destino de los productos agropecuarios son los mercados locales y regionales, mercado de Lima, mercado exterior, agroindustria y no sabe.

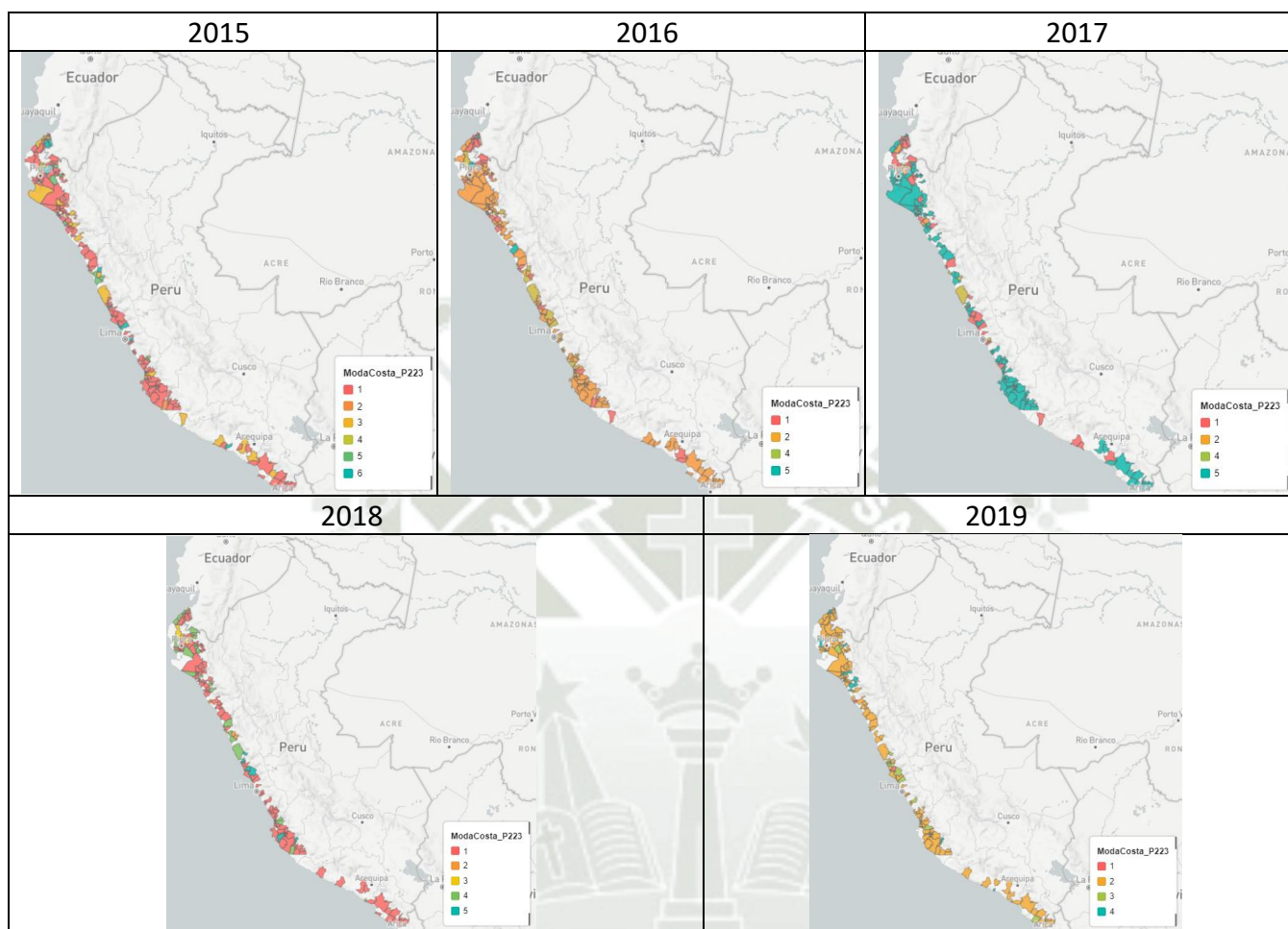
Figura 17.

Características de los clústeres en la Costa

AÑO	COSTA_P223	P223_1 Mercado Local	P223_2 Mercado Regional	P223_3 Mercado Exterior	P223_4 Agroindust ria	P223_5 Mercado Lima	P223_6 No Sabe	Cantidad Observaciones
2015	1	6,316	509	0	0	116	0	6,316
2015	3	19	2,205	4	402	4	0	2,585
2015	6	0	185	73	0	869	0	869
2015	4	20	105	536	28	0	0	536
2015	5	0	0	0	0	0	407	407
2015	2	64	6	64	2	13	0	64
2016	2	6,080	0	0	0	0	322	6,402
2016	1	898	2,764	0	293	1	0	3,049
2016	3	242	245	0	0	1,341	0	1,341
2016	5	6	2	467	33	0	0	467
2016	6	87	115	170	0	4	0	170
2016	4	9	0	54	18	59	0	59
2017	5	6,162	10	1	202	0	0	6,337
2017	1	0	2,090	83	0	1,405	0	3,267
2017	2	890	618	116	0	254	0	890
2017	3	0	0	411	33	0	287	698
2017	4	2	5	0	14	14	0	14
2018	1	7,090	0	0	0	279	0	7,090
2018	4	1,031	2,454	0	18	0	0	2,454
2018	5	31	166	1	9	935	0	935
2018	3	82	75	516	0	52	0	516
2018	2	32	0	53	441	36	0	441
2019	2	7,787	1,041	405	3	0	0	8,115
2019	4	4	1,547	89	9	0	0	1,547
2019	3	225	148	33	1	1,292	0	1,292
2019	1	32	0	12	380	12	0	380
Total		37,109	14,290	3,088	1,886	6,686	1,016	56,241

Fuente: Elaboración Propia.

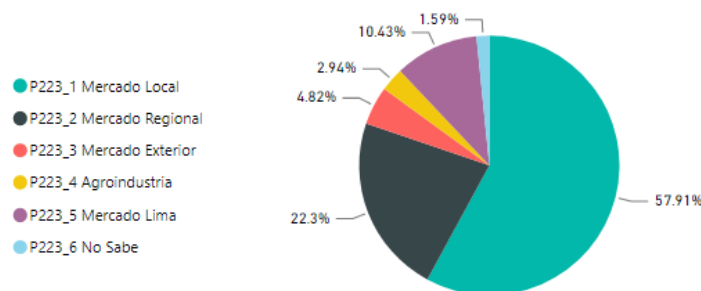
Figura 18.
Mapa de los clústeres de la costa.



Fuente: Elaboración Propia.

En la región de la costa se ve una mínima variación en la que los productores que venden su producción a las agroindustrias supera a los productores que no saben.

Figura 19.
Distribución de variables costa 2015-2019

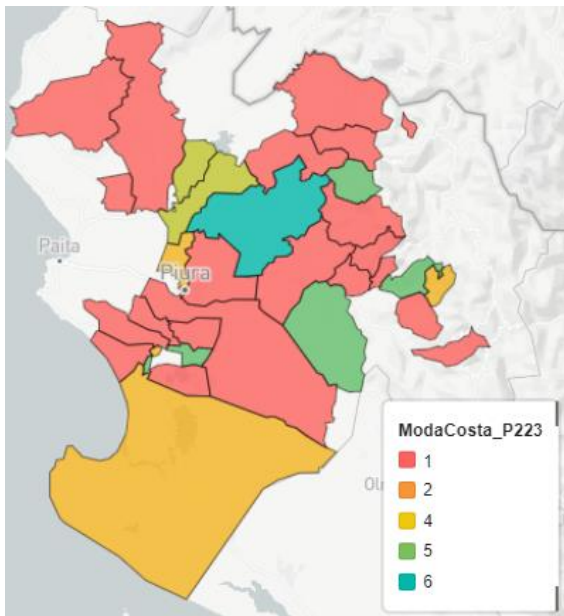
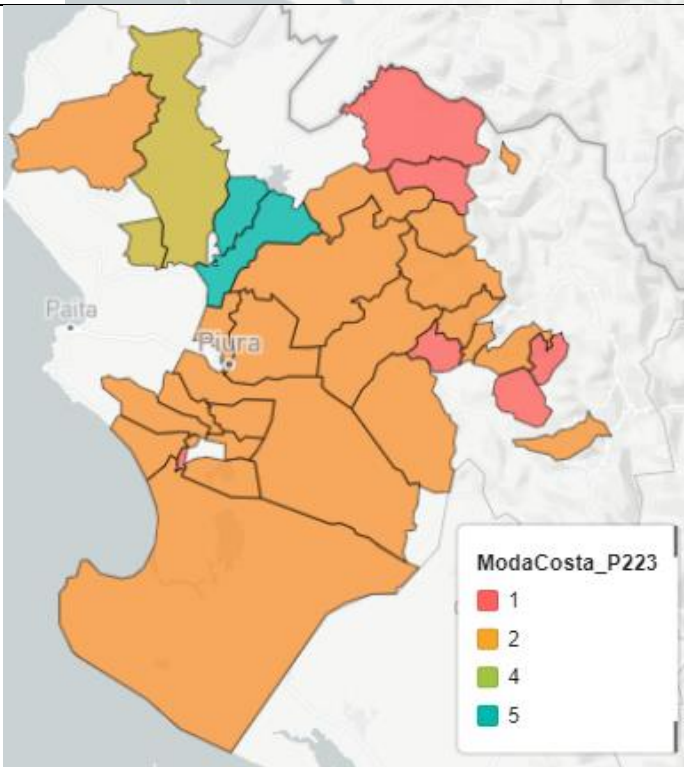


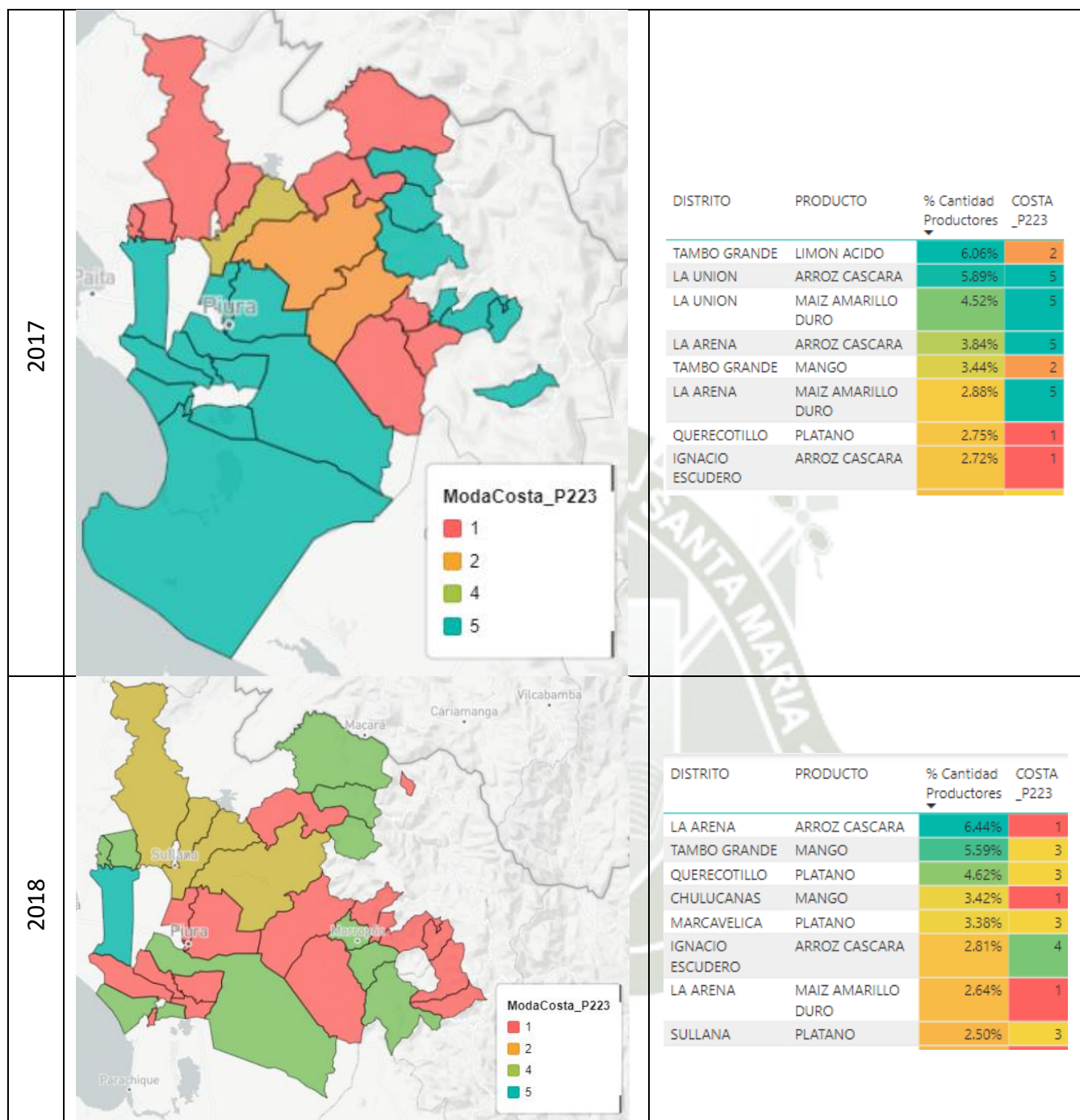
Fuente: Elaboración Propia.

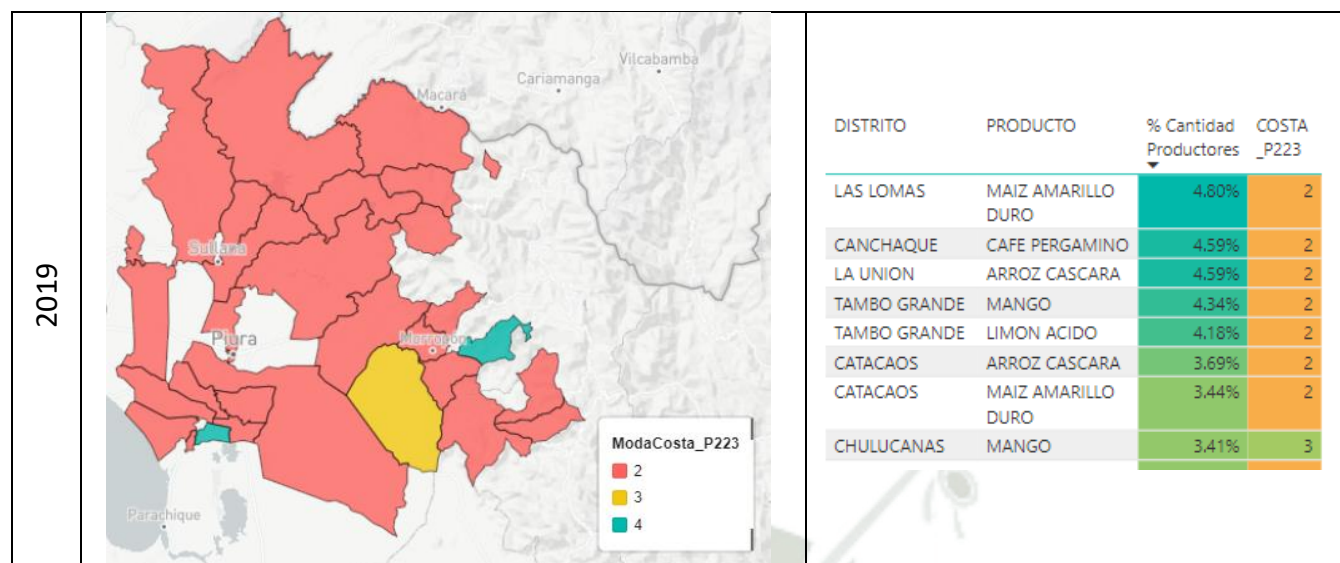
Descripción departamento de Piura, en dicho departamento se ve que existe una gran variedad de Clústeres, debido a que hay más diversidad de comercialización.

En la región de la costa en el departamento de Piura se repite el mismo patrón que en la región San Martín siendo el clúster más común en cada año se caracteriza por productores que destinan la venta de su producción al mercado local u otros mercados, pero en una minoría, el segundo clúster más común cada año se caracteriza por productores que venden su producción en el mercado regional (ferias regionales, centro de acopio regional), pero con la diferencia de los principales productos que fueron vendidos, en esta región los productos que fueron vendidos fueron el arroz cascara, el plátano, limón ácido, mango, maíz amarillo duro, e entre otros.

Figura20.
Características del departamento de Piura.

Año	Geografía	Productos Principales																																								
2015		<table><tr><th>DISTRITO</th><th>PRODUCTO</th><th>% Cantidad Productores</th><th>COSTA _P223</th></tr><tr><td>LA UNION</td><td>ARROZ CASCARA</td><td>6.04%</td><td>1</td></tr><tr><td>SULLANA</td><td>PLATANO</td><td>4.42%</td><td>4</td></tr><tr><td>LA ARENA</td><td>ARROZ CASCARA</td><td>3.68%</td><td>1</td></tr><tr><td>IGNACIO ESCUDERO</td><td>ARROZ CASCARA</td><td>3.50%</td><td>1</td></tr><tr><td>TAMBO GRANDE</td><td>LIMON ACIDO</td><td>3.50%</td><td>6</td></tr><tr><td>QUERECOTILLO</td><td>PLATANO</td><td>3.25%</td><td>4</td></tr><tr><td>CHULUCANAS</td><td>MAIZ AMARILLO DURO</td><td>3.11%</td><td>1</td></tr><tr><td>MARCAVELICA</td><td>PLATANO</td><td>2.57%</td><td>6</td></tr><tr><td>CHULUCANAS</td><td>ARROZ CASCARA</td><td>2.28%</td><td>1</td></tr></table>	DISTRITO	PRODUCTO	% Cantidad Productores	COSTA _P223	LA UNION	ARROZ CASCARA	6.04%	1	SULLANA	PLATANO	4.42%	4	LA ARENA	ARROZ CASCARA	3.68%	1	IGNACIO ESCUDERO	ARROZ CASCARA	3.50%	1	TAMBO GRANDE	LIMON ACIDO	3.50%	6	QUERECOTILLO	PLATANO	3.25%	4	CHULUCANAS	MAIZ AMARILLO DURO	3.11%	1	MARCAVELICA	PLATANO	2.57%	6	CHULUCANAS	ARROZ CASCARA	2.28%	1
DISTRITO	PRODUCTO	% Cantidad Productores	COSTA _P223																																							
LA UNION	ARROZ CASCARA	6.04%	1																																							
SULLANA	PLATANO	4.42%	4																																							
LA ARENA	ARROZ CASCARA	3.68%	1																																							
IGNACIO ESCUDERO	ARROZ CASCARA	3.50%	1																																							
TAMBO GRANDE	LIMON ACIDO	3.50%	6																																							
QUERECOTILLO	PLATANO	3.25%	4																																							
CHULUCANAS	MAIZ AMARILLO DURO	3.11%	1																																							
MARCAVELICA	PLATANO	2.57%	6																																							
CHULUCANAS	ARROZ CASCARA	2.28%	1																																							
2016		<table><tr><th>DISTRITO</th><th>PRODUCTO</th><th>% Cantidad Productores</th><th>COSTA _P223</th></tr><tr><td>LA UNION</td><td>ARROZ CASCARA</td><td>5.86%</td><td>2</td></tr><tr><td>QUERECOTILLO</td><td>PLATANO</td><td>5.69%</td><td>5</td></tr><tr><td>SULLANA</td><td>PLATANO</td><td>4.60%</td><td>5</td></tr><tr><td>LA ARENA</td><td>ARROZ CASCARA</td><td>3.97%</td><td>2</td></tr><tr><td>MARCAVELICA</td><td>ARROZ CASCARA</td><td>3.08%</td><td>3</td></tr><tr><td>LA ARENA</td><td>ARROZ CASCARA</td><td>2.54%</td><td>1</td></tr><tr><td>TAMBO GRANDE</td><td>LIMON ACIDO</td><td>2.39%</td><td>2</td></tr><tr><td>CHULUCANAS</td><td>MANGO</td><td>2.38%</td><td>3</td></tr><tr><td>LA ARENA</td><td>MAIZ AMARILLO DURO</td><td>2.38%</td><td>2</td></tr></table>	DISTRITO	PRODUCTO	% Cantidad Productores	COSTA _P223	LA UNION	ARROZ CASCARA	5.86%	2	QUERECOTILLO	PLATANO	5.69%	5	SULLANA	PLATANO	4.60%	5	LA ARENA	ARROZ CASCARA	3.97%	2	MARCAVELICA	ARROZ CASCARA	3.08%	3	LA ARENA	ARROZ CASCARA	2.54%	1	TAMBO GRANDE	LIMON ACIDO	2.39%	2	CHULUCANAS	MANGO	2.38%	3	LA ARENA	MAIZ AMARILLO DURO	2.38%	2
DISTRITO	PRODUCTO	% Cantidad Productores	COSTA _P223																																							
LA UNION	ARROZ CASCARA	5.86%	2																																							
QUERECOTILLO	PLATANO	5.69%	5																																							
SULLANA	PLATANO	4.60%	5																																							
LA ARENA	ARROZ CASCARA	3.97%	2																																							
MARCAVELICA	ARROZ CASCARA	3.08%	3																																							
LA ARENA	ARROZ CASCARA	2.54%	1																																							
TAMBO GRANDE	LIMON ACIDO	2.39%	2																																							
CHULUCANAS	MANGO	2.38%	3																																							
LA ARENA	MAIZ AMARILLO DURO	2.38%	2																																							

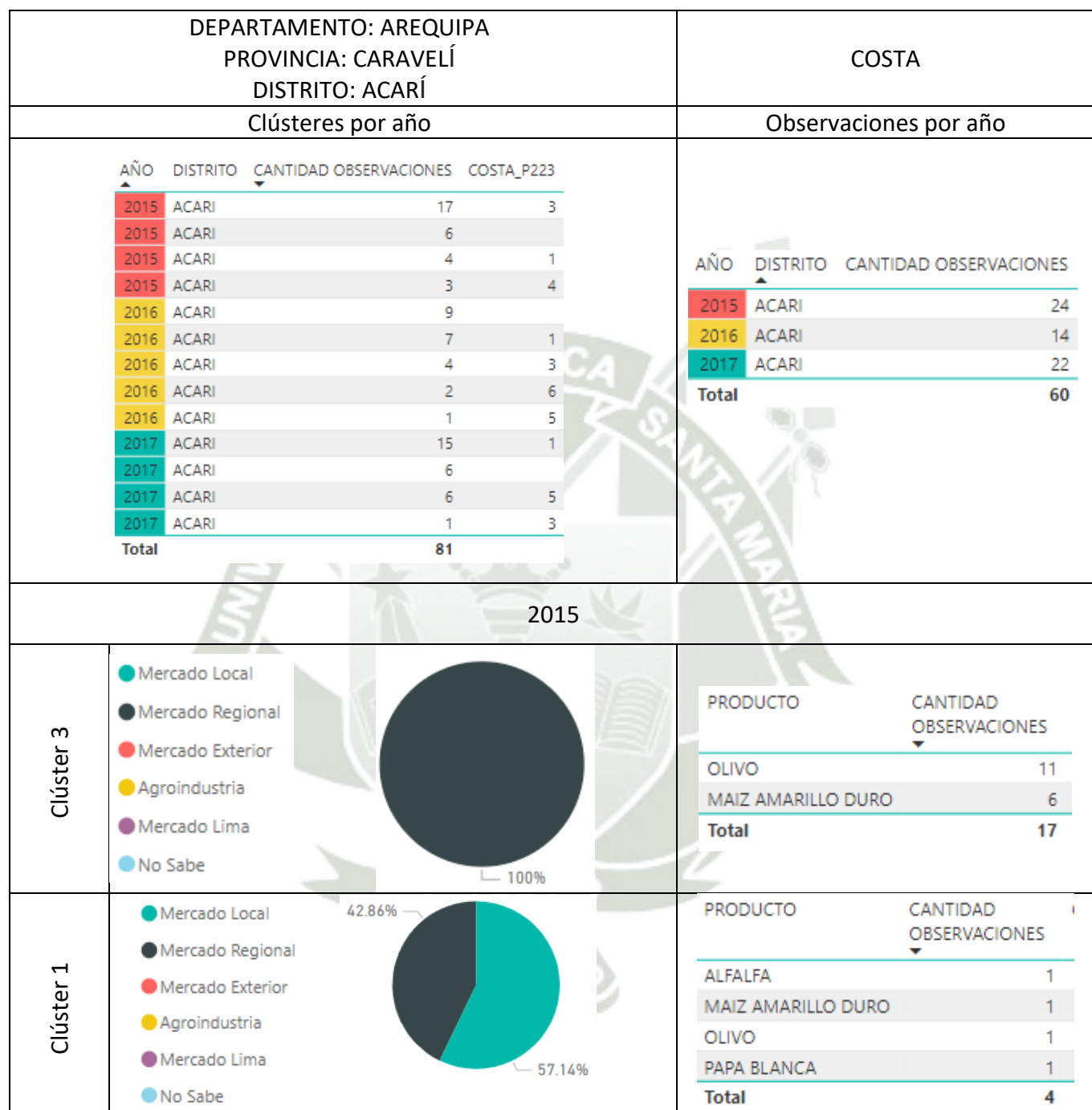




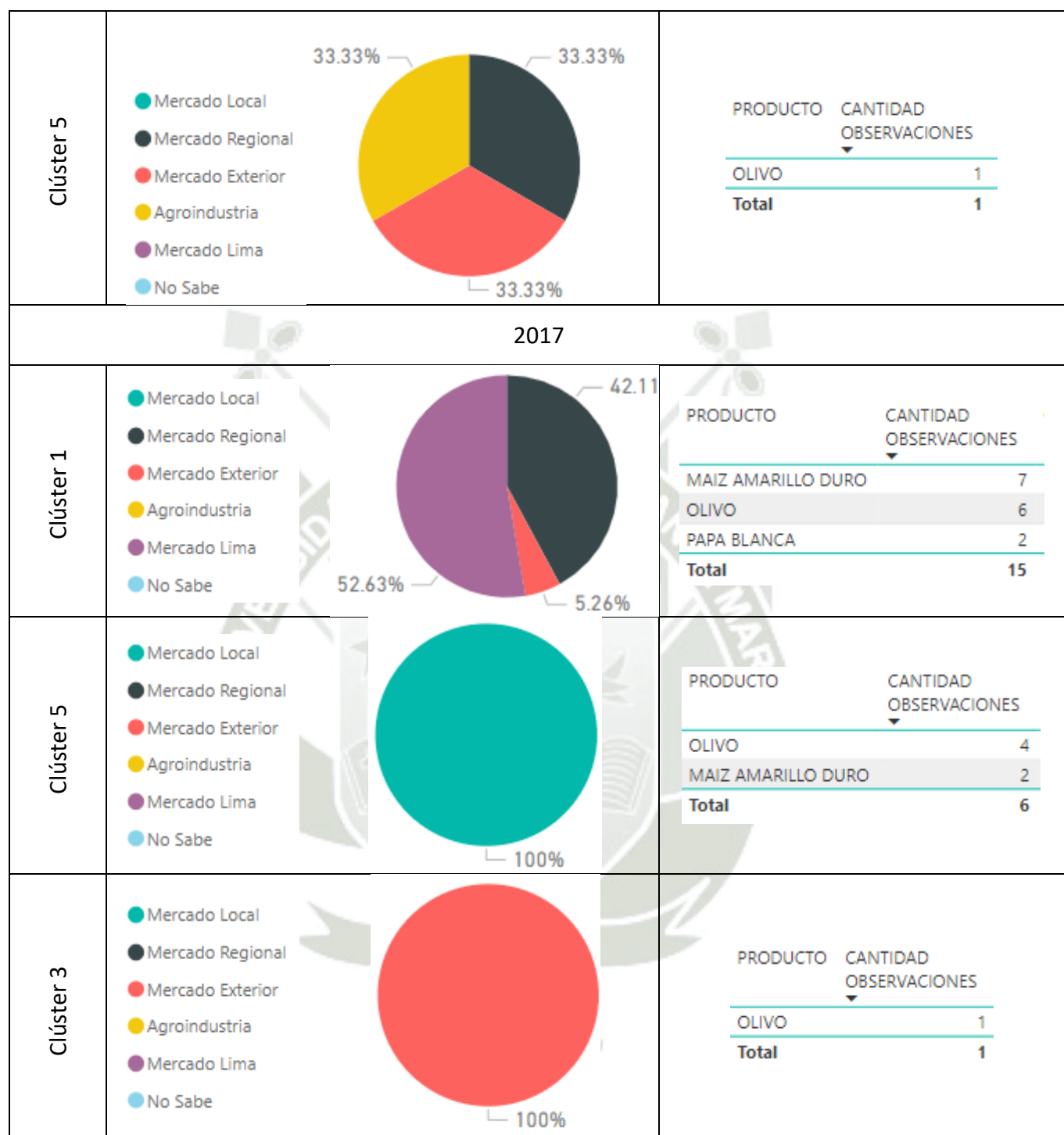
Fuente: Elaboración Propia.

En el distrito de Acarí de un total de 81 observaciones se clasificaron 60, del año 2015 al 2017, en el año 2015 la producción se vendió en el mercado regional, mercado exterior y local, agrupado en los clústeres 3, 1 y 4. Siendo la alfalfa el maíz amarillo el olivo y la papa blanca los productos comercializados, para el año 2016 los productos como olivo, maíz amarillo, papa blanca y cebolla, fueron comercializados en el mercado regional, mercado de Lima, mercado exterior y la agroindustria, en el año 2017 los productos de alfalfa, maíz amarillo, olivo y papa blanca fueron comercializados en el mercado de Lima, local, regional y mercado exterior.

Figura 21.
Características del distrito de Acarí.



Clúster 4	Mercado Local Mercado Regional Mercado Exterior Agroindustria Mercado Lima No Sabe 50% 50%	<table><tr><th>PRODUCTO</th><th>CANTIDAD OBSERVACIONES</th></tr><tr><td>OLIVO</td><td>3</td></tr><tr><td>Total</td><td>3</td></tr></table>	PRODUCTO	CANTIDAD OBSERVACIONES	OLIVO	3	Total	3				
PRODUCTO	CANTIDAD OBSERVACIONES											
OLIVO	3											
Total	3											
2016												
Clúster 1	Mercado Local Mercado Regional Mercado Exterior Agroindustria Mercado Lima No Sabe 12.5% 87.5%	<table><tr><th>PRODUCTO</th><th>CANTIDAD OBSERVACIONES</th></tr><tr><td>OLIVO</td><td>4</td></tr><tr><td>MAIZ AMARILLO DURO</td><td>2</td></tr><tr><td>PAPA BLANCA</td><td>1</td></tr><tr><td>Total</td><td>7</td></tr></table>	PRODUCTO	CANTIDAD OBSERVACIONES	OLIVO	4	MAIZ AMARILLO DURO	2	PAPA BLANCA	1	Total	7
PRODUCTO	CANTIDAD OBSERVACIONES											
OLIVO	4											
MAIZ AMARILLO DURO	2											
PAPA BLANCA	1											
Total	7											
Clúster 3	Mercado Local Mercado Regional Mercado Exterior Agroindustria Mercado Lima No Sabe 20% 80%	<table><tr><th>PRODUCTO</th><th>CANTIDAD OBSERVACIONES</th></tr><tr><td>CEBOLLA</td><td>2</td></tr><tr><td>MAIZ AMARILLO DURO</td><td>1</td></tr><tr><td>PAPA BLANCA</td><td>1</td></tr><tr><td>Total</td><td>4</td></tr></table>	PRODUCTO	CANTIDAD OBSERVACIONES	CEBOLLA	2	MAIZ AMARILLO DURO	1	PAPA BLANCA	1	Total	4
PRODUCTO	CANTIDAD OBSERVACIONES											
CEBOLLA	2											
MAIZ AMARILLO DURO	1											
PAPA BLANCA	1											
Total	4											
Clúster 6	Mercado Local Mercado Regional Mercado Exterior Agroindustria Mercado Lima No Sabe 50% 50%	<table><tr><th>PRODUCTO</th><th>CANTIDAD OBSERVACIONES</th></tr><tr><td>OLIVO</td><td>2</td></tr><tr><td>Total</td><td>2</td></tr></table>	PRODUCTO	CANTIDAD OBSERVACIONES	OLIVO	2	Total	2				
PRODUCTO	CANTIDAD OBSERVACIONES											
OLIVO	2											
Total	2											



Fuente: Elaboración Propia.

3) Sierra

En la región de la sierra se encuentra el mayor número de productores encuestados, cuenta con un total de 119969 encuestados del 2015 al 2019, repitiendo el mismo patrón que en la selva y la costa.

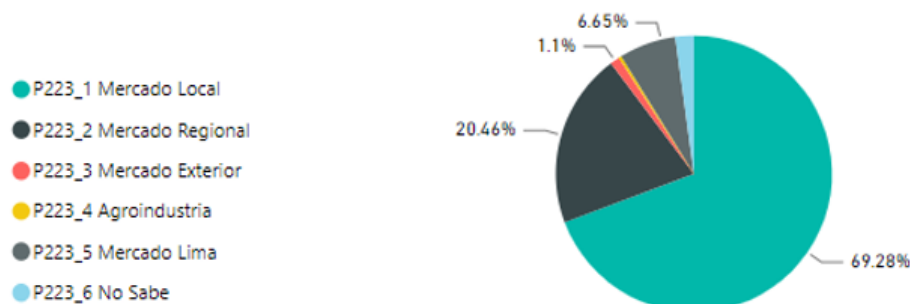
Figura 22.
Características de los clústeres en la Sierra

AÑO	SIERRA_P223	P223_1 Mercado Local	P223_2 Mercado Regional	P223_3 Mercado Exterior	P223_4 Agroindustria	P223_5 Mercado Lima	P223_6 No Sabe	CANTIDAD OBSERVACIONES
2015	3	14,310	0	0	5	0	763	15,073
2015	2	940	5,124	76	0	1	0	5,124
2015	5	151	100	0	0	1,164	0	1,164
2015	1	27	0	320	3	0	0	320
2015	4	0	7	0	46	2	0	46
2016	3	16,621	979	22	0	0	0	16,621
2016	2	0	5,047	43	7	183	917	5,964
2016	1	306	168	12	17	1,761	0	1,761
2016	4	8	0	198	130	0	0	322
2017	5	18,550	5,294	0	0	552	0	22,662
2017	4	1	0	3	12	1,385	0	1,385
2017	1	0	0	0	0	0	1,099	1,099
2017	3	4	17	189	0	1	0	189
2017	2	8	5	0	103	0	0	103
2018	2	18,059	0	0	4	1,650	0	19,440
2018	3	1,275	4,784	0	0	136	0	4,784
2018	1	21	57	258	3	11	0	258
2018	4	10	8	0	61	0	0	61
2019	4	18,360	4,505	1	6	3	0	21,601
2019	3	226	142	0	0	1,663	0	1,663
2019	2	11	19	288	0	17	0	288
2019	1	3	0	1	41	0	0	41
Total		88,891	26,256	1,411	438	8,529	2,779	119,969

Fuente: Elaboración Propia.



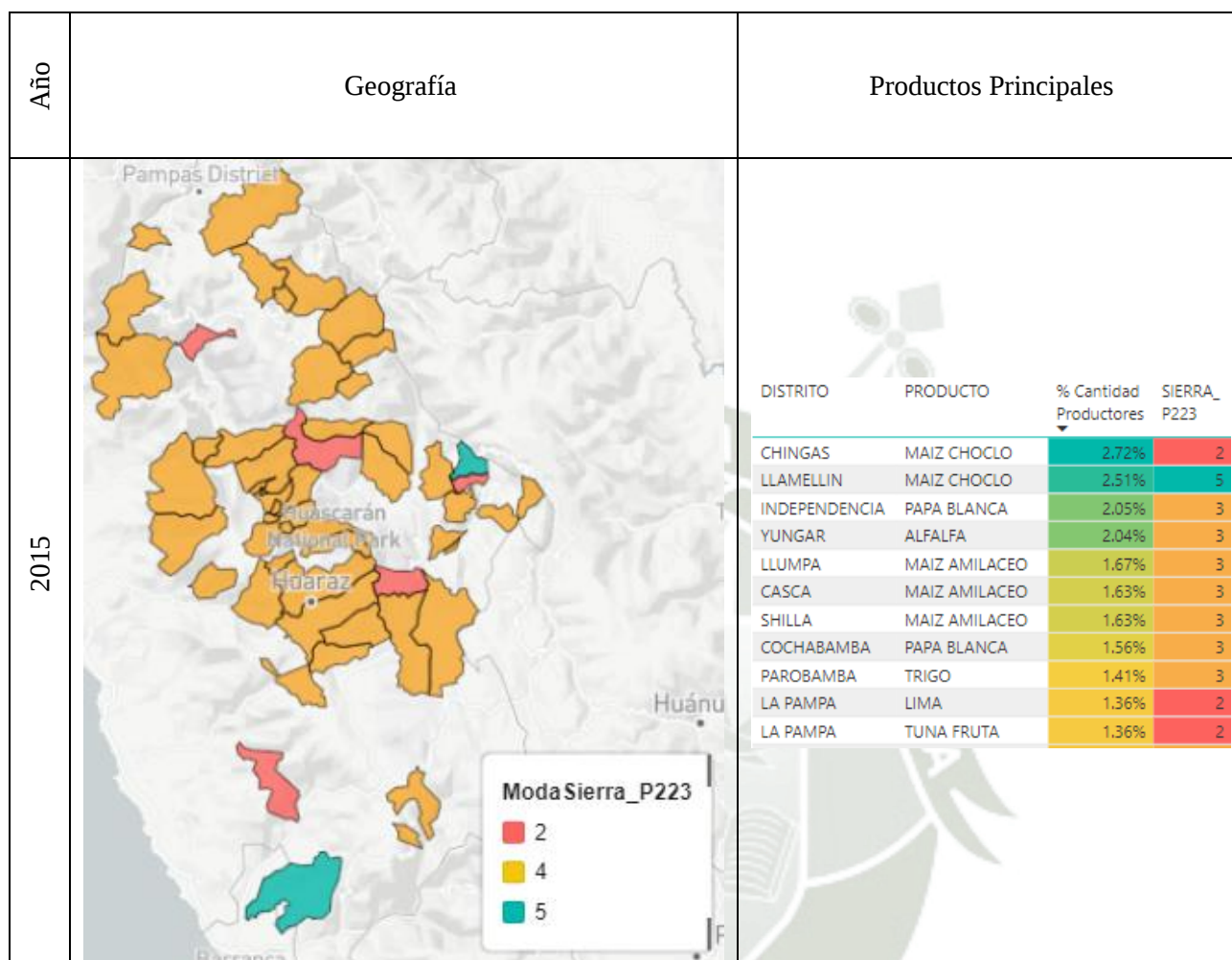
Figura 24.
Distribución de variables sierra 2015-2019

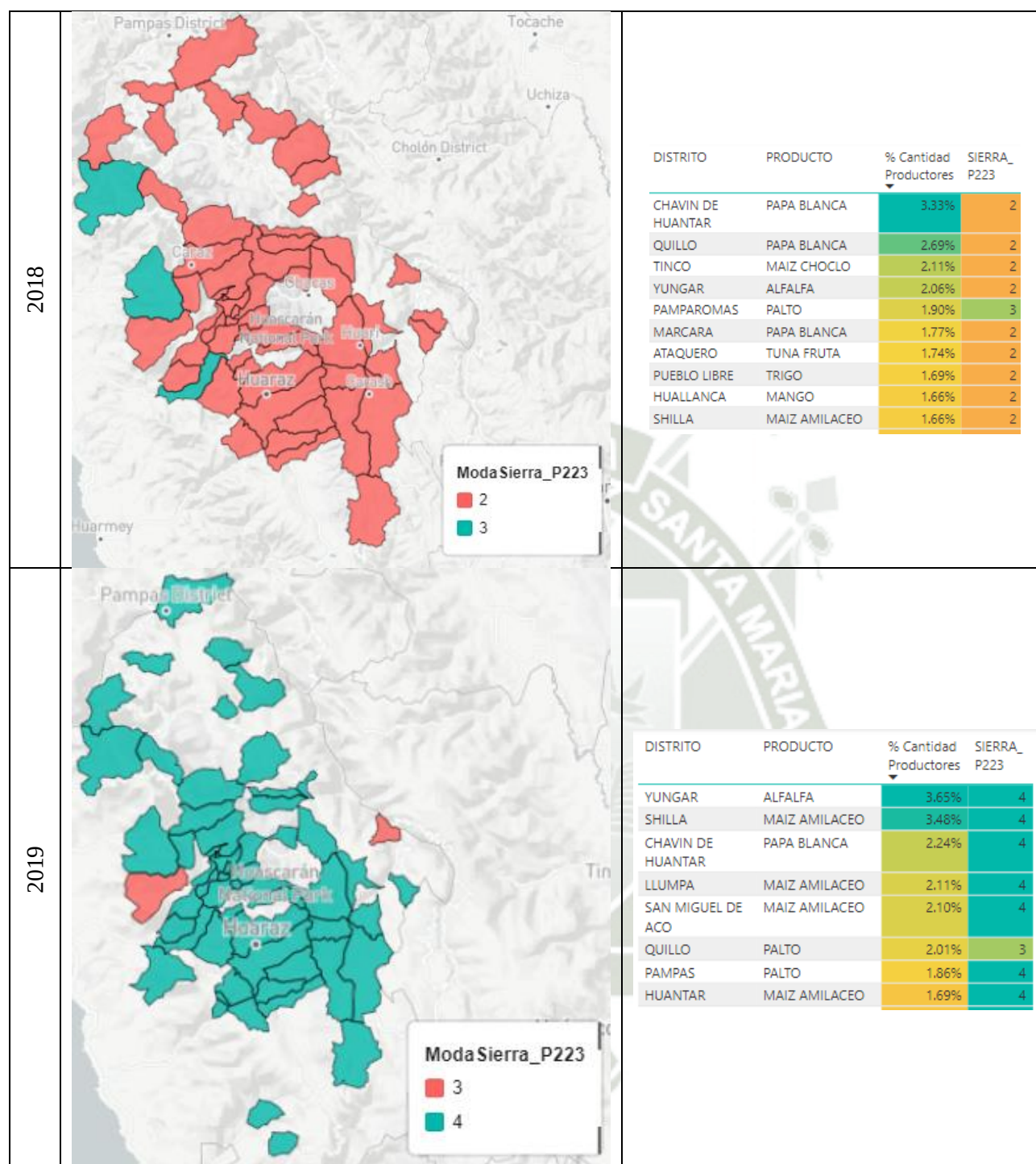


Fuente: Elaboración Propia.

Descripción del departamento de Áncash, con sus principales productos y el clúster al que pertenecen, en este departamento el clúster de productores que venden su producción en el mercado local sigue predominando sobre todos, pero en los años 2017 y 2019 se caracterizó por productores que vendían su producción tanto en el mercado local como en el mercado regional, donde los principales productos fueron el maíz choclo, maíz amiláceo, alfalfa, papa blanca, trigo, entre otros.

Figura 25.
Características departamento de Ancash

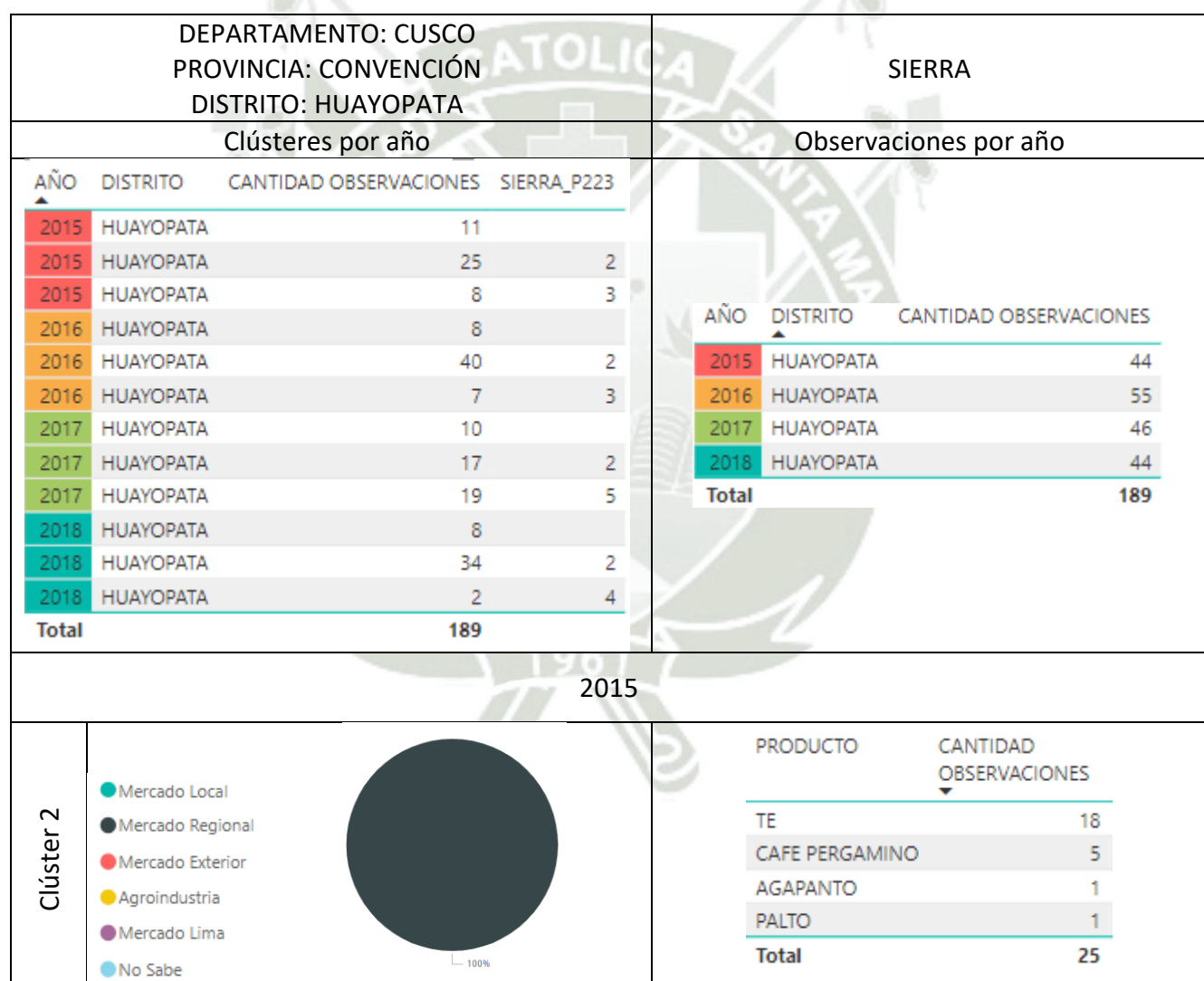


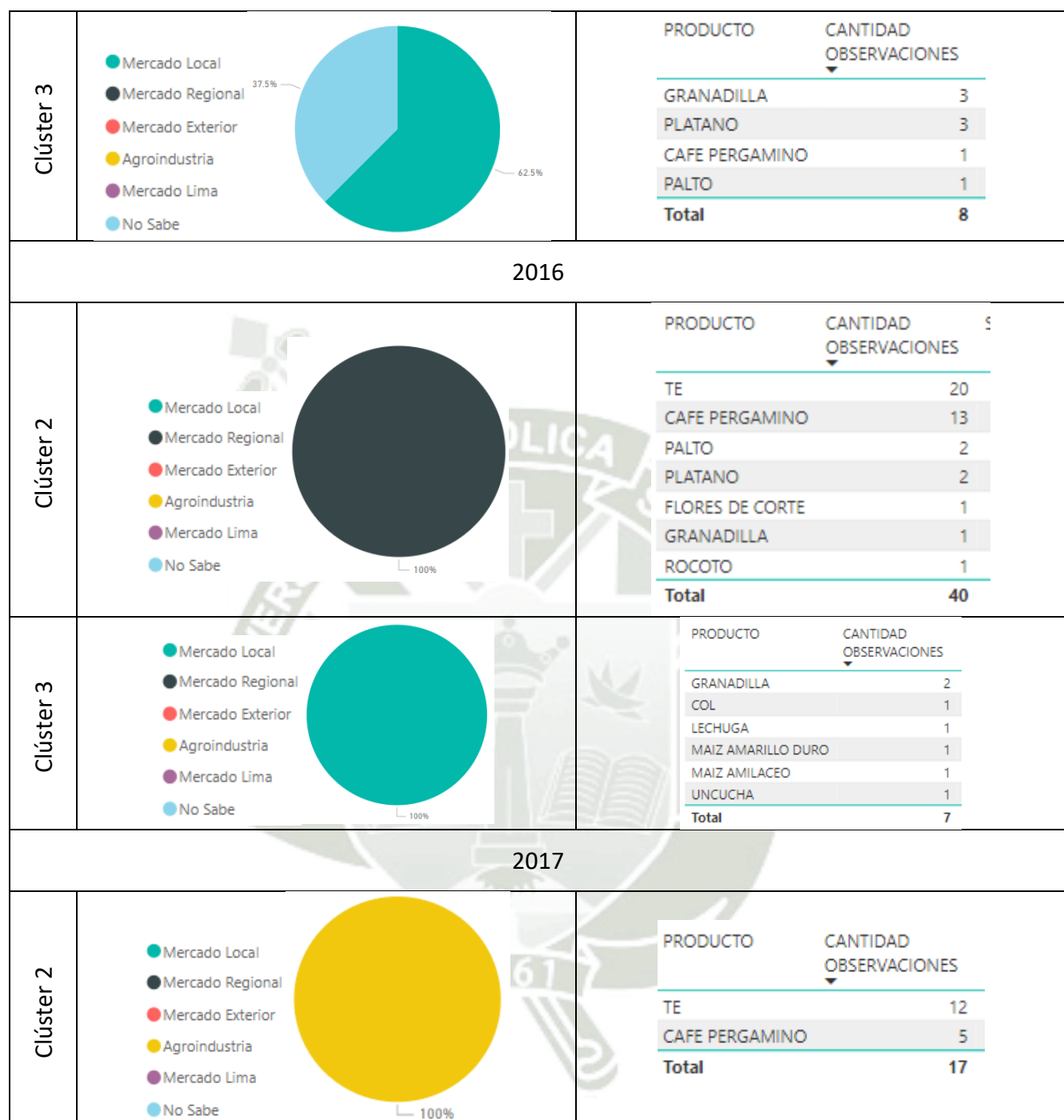


Fuente: Elaboración Propia.

En la figura número 18, se describe la provincia de Huayopata en el departamento de Cusco, en esta figura se describe la cantidad de observaciones por año, la cantidad de clústeres por año y la descripción de cada año, en el año 2015 el distrito de Huayopata perteneció al clúster 2 y 3, ya que el destino de su producción fue únicamente al mercado regional, por lo que se agrupó al

clúster 2, y otra parte de sus productores se caracterizó por vender su producción en el mercado local o no sabe y es por eso que se agrupa al clúster 3, en el año 2016 la venta fue en el mercado regional en la mayoría y en el mercado local, en el año 2017 se puede observar un nuevo destino de venta que es la agroindustria, con productos como el té y el café pergamino en el clúster 2, y productores que vendieron en el mercado local y/o regional, en el año 2018 el principal destino también es el mercado local en el clúster 2, y la agroindustria y/o el mercado local.

Figura 26.
Características del distrito de Huayopata.




Clúster 5	Mercado Local Mercado Regional Mercado Exterior Agroindustria Mercado Lima No Sabe 47.62% 52.38%	<table><thead><tr><th>PRODUCTO</th><th>CANTIDAD OBSERVACIONES</th></tr></thead><tbody><tr><td>GRANADILLA</td><td>5</td></tr><tr><td>CAFE PERGAMINO</td><td>3</td></tr><tr><td>PLATANO</td><td>3</td></tr><tr><td>ARRACACHA</td><td>2</td></tr><tr><td>PALTO</td><td>2</td></tr><tr><td>COL</td><td>1</td></tr><tr><td>FRIJOL BLANCO GRANO SECO</td><td>1</td></tr><tr><td>FRIJOL PANAMITO GRANO SECO</td><td>1</td></tr><tr><td>YACON</td><td>1</td></tr><tr><td>Total</td><td>19</td></tr></tbody></table>	PRODUCTO	CANTIDAD OBSERVACIONES	GRANADILLA	5	CAFE PERGAMINO	3	PLATANO	3	ARRACACHA	2	PALTO	2	COL	1	FRIJOL BLANCO GRANO SECO	1	FRIJOL PANAMITO GRANO SECO	1	YACON	1	Total	19
	PRODUCTO	CANTIDAD OBSERVACIONES																						
GRANADILLA	5																							
CAFE PERGAMINO	3																							
PLATANO	3																							
ARRACACHA	2																							
PALTO	2																							
COL	1																							
FRIJOL BLANCO GRANO SECO	1																							
FRIJOL PANAMITO GRANO SECO	1																							
YACON	1																							
Total	19																							
2018																								
Clúster 2	Mercado Local Mercado Regional Mercado Exterior Agroindustria Mercado Lima No Sabe 100%	<table><thead><tr><th>PRODUCTO</th><th>CANTIDAD OBSERVACIONES</th></tr></thead><tbody><tr><td>TE</td><td>15</td></tr><tr><td>CAFE PERGAMINO</td><td>11</td></tr><tr><td>PLATANO</td><td>3</td></tr><tr><td>AGUAYMANTO</td><td>1</td></tr><tr><td>FLORES DE CORTE</td><td>1</td></tr><tr><td>NARANJO</td><td>1</td></tr><tr><td>PAPA NATIVA</td><td>1</td></tr><tr><td>PASTO SARA SARA</td><td>1</td></tr><tr><td>Total</td><td>34</td></tr></tbody></table>	PRODUCTO	CANTIDAD OBSERVACIONES	TE	15	CAFE PERGAMINO	11	PLATANO	3	AGUAYMANTO	1	FLORES DE CORTE	1	NARANJO	1	PAPA NATIVA	1	PASTO SARA SARA	1	Total	34		
	PRODUCTO	CANTIDAD OBSERVACIONES																						
TE	15																							
CAFE PERGAMINO	11																							
PLATANO	3																							
AGUAYMANTO	1																							
FLORES DE CORTE	1																							
NARANJO	1																							
PAPA NATIVA	1																							
PASTO SARA SARA	1																							
Total	34																							
Clúster 4	Mercado Local Mercado Regional Mercado Exterior Agroindustria Mercado Lima No Sabe 33.33% 66.67%	<table><thead><tr><th>PRODUCTO</th><th>CANTIDAD OBSERVACIONES</th></tr></thead><tbody><tr><td>TE</td><td>2</td></tr><tr><td>Total</td><td>2</td></tr></tbody></table>	PRODUCTO	CANTIDAD OBSERVACIONES	TE	2	Total	2																
	PRODUCTO	CANTIDAD OBSERVACIONES																						
TE	2																							
Total	2																							

Fuente: Elaboración Propia.

4) Comparación Costa, Sierra y Selva

En la selva el principal destino de la producción es el mercado local superando ampliamente a las demás variables, seguido del mercado regional, mercado de Lima, mercado exterior, no sabe y por último la en un porcentaje muy bajo la agroindustria, en la costa el mercado local también es el más grande, pero en comparación con la selva es menor, pero en la venta de producción al mercado regional, mercado de Lima, mercado exterior y agroindustria superan a la selva, por

último en la sierra, el mercado local también es el más común, seguido del mercado regional y mercado de Lima, en la sierra se vende al mercado exterior y agroindustria en poca cantidad respecto a la costa y la selva.

En la costa es donde se encuesta menor cantidad de productores, pero es la región que más diversidad de comercialización tiene, no obstante, sigue siendo una comercialización se solo se centra en pocas opciones para la venta de los productos agrícolas, un caso opuesto es la sierra que es la región de más encuestados donde la comercialización se centraliza en el mercado local y regional, y muy poco en otros destinos de comercialización.

3.1.2. Módulo 200AB_1 Pérdida total de producción

A. Variable P224E ¿Cuál fue el motivo de la pérdida total de producción o superficie?

Este módulo ayuda a determinar cuál es la causa de pérdida de producción la cual afecta directamente la economía de los productores y el desarrollo del sector.

A lo largo de este periodo 2016-2019 las heladas han sido el principal factor de pérdida de producción de cultivos, este fenómeno se asocia principalmente con las granizadas en la sierra centro y sur del país, en las provincias de Huancané, Chucuito, Andahuaylas, entre otros, distritos que muchas veces fueron declaradas en emergencias debido a estos fenómenos.

Otra asociación importante en estos clústeres incluye nuevamente a las heladas, pero esta vez con las plagas, según la Fundación para el Desarrollo Frutícola(FDF) las bajas temperaturas generan micro heridas en las paredes celulares de las plantas dañando el tejido de manera irreversible, dejando vulnerable a los cultivos a la entrada de hongos que podrían generar menor producción durante las siguientes temporadas; otra asociación son las plagas son las sequías, según Seranca una empresa especialista en fumigación y control de plagas, considera a las sequías una aliada de las plagas, ya que, el aumento de plagas es causado por como efectos de la sequía, debido al incremento de temperaturas en la tierra; la ausencia de agua que provoca el calor impide que árboles y plantas puedan nutrirse y realizar correctamente la fotosíntesis, produciendo pérdidas de cultivo, proliferación de plagas y enfermedades, afectando principalmente a los pequeños y medianos productores agrarios.

Figura 27.
Características de los clústeres en P224E.

AÑO	CLUSTER	P_Huaycos	P_Inundaciones	P_Plagas	P_Heladas	P_Granizadas	P_Sequia	P_Otro	Cantidad Observaciones
2016	3	4	0	0	0	0	815	20	834
2016	4	0	0	286	442	187	682	0	705
2016	1	0	0	614	84	9	0	8	614
2016	5	0	53	0	475	268	0	0	568
2016	2	49	140	7	0	0	0	0	189
2017	5	43	819	72	52	0	8	0	819
2017	1	11	1	495	0	0	328	0	761
2017	2	0	6	41	297	346	149	0	426
2017	4	3	3	59	325	1	5	0	325
2017	3	219	0	5	1	0	0	72	291
2018	3	10	25	540	9	0	0	0	540
2018	5	23	133	0	342	0	0	25	501
2018	4	1	8	28	142	204	0	0	204
2018	2	0	0	0	34	22	156	0	156
2018	1	0	0	102	69	1	37	0	102
2019	6	3	0	0	525	147	0	0	672
2019	4	0	0	418	1	0	0	69	485
2019	2	0	0	59	337	358	51	0	358
2019	3	0	0	103	144	0	276	0	342
2019	1	7	172	8	11	6	0	0	172
2019	5	106	0	2	1	0	0	0	106
Total		479	1,360	2,839	3,291	1,549	2,507	194	9,170

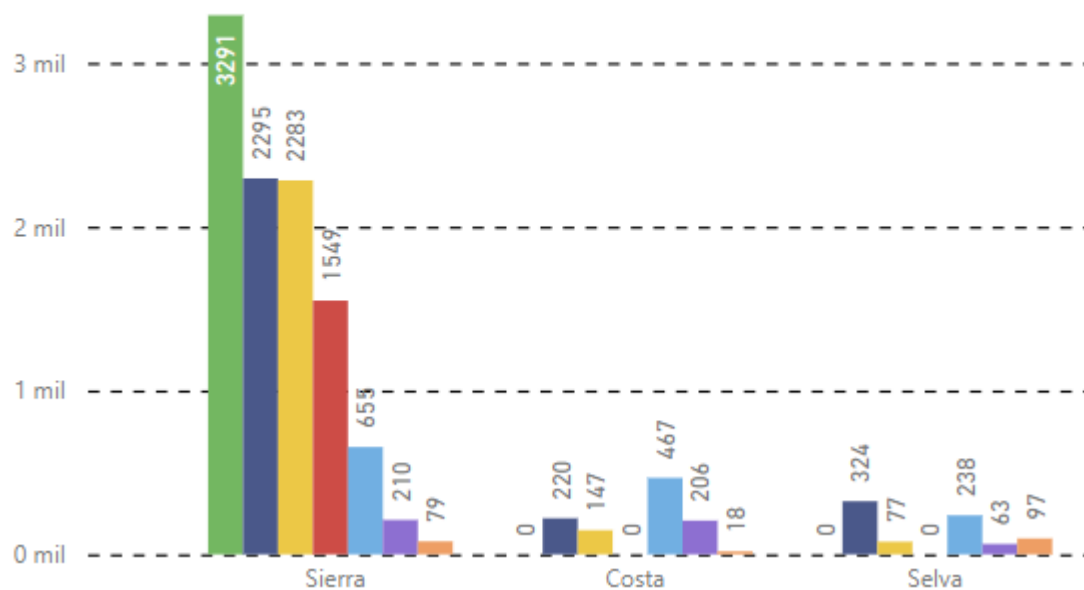
Fuente: Elaboración Propia.

La variable con más observaciones son las heladas, seguido de las plagas y las sequías desde el año 2016 al 2019, por lo tanto, contribuyen en mayor grado a la clasificación de los productores, ya que se encuentra una cantidad considerable en la mayoría de clúster formados.

La geografía en el Perú es muy diversa, en la siguiente imagen se clasifica los desastres por regiones ya que en las tres regiones el clima es muy diverso, en la sierra se ve afectado más por las heladas, plagas, sequías, granizadas, inundaciones, huaycos y otros fenómenos, mientras que, en la región de la costa y selva, también son afectados por estos fenómenos a excepción de las heladas y granizadas que solo se dan en la sierra del Perú.

Figura 28.
Distribución de variables 2016-2019.

P_Heladas P_Plagas P_Sequia P_Granizadas P_Inundaciones P_Huaycos P_Otro


Fuente: Elaboración Propia.

En esta grafica se encuentran los productos que fueron más afectados, como el maíz amiláceo, las habas, la quinua, papa blanca, cebada grano, papa nativa, arveja, maíz choclo, maíz amarillo duro, oca, trigo, olluco, trigo entre otros productos, que son producidos en departamentos de la sierra como Puno, Cusco, Ayacucho, Cajamarca, Junín, Cajamarca, Apurímac, Ancash, entre otros.

En la costa los productos más afectados fueron el maíz amarillo duro, arroz cascara, yuca, camote, frijol de palo, maíz choclo, maíz chala, maíz amiláceo, plátano, algodón, etc. Y en la selva el maíz amarillo duro, yuca, plátano, arroz cascara, sandia, cacao, tomate, etc.

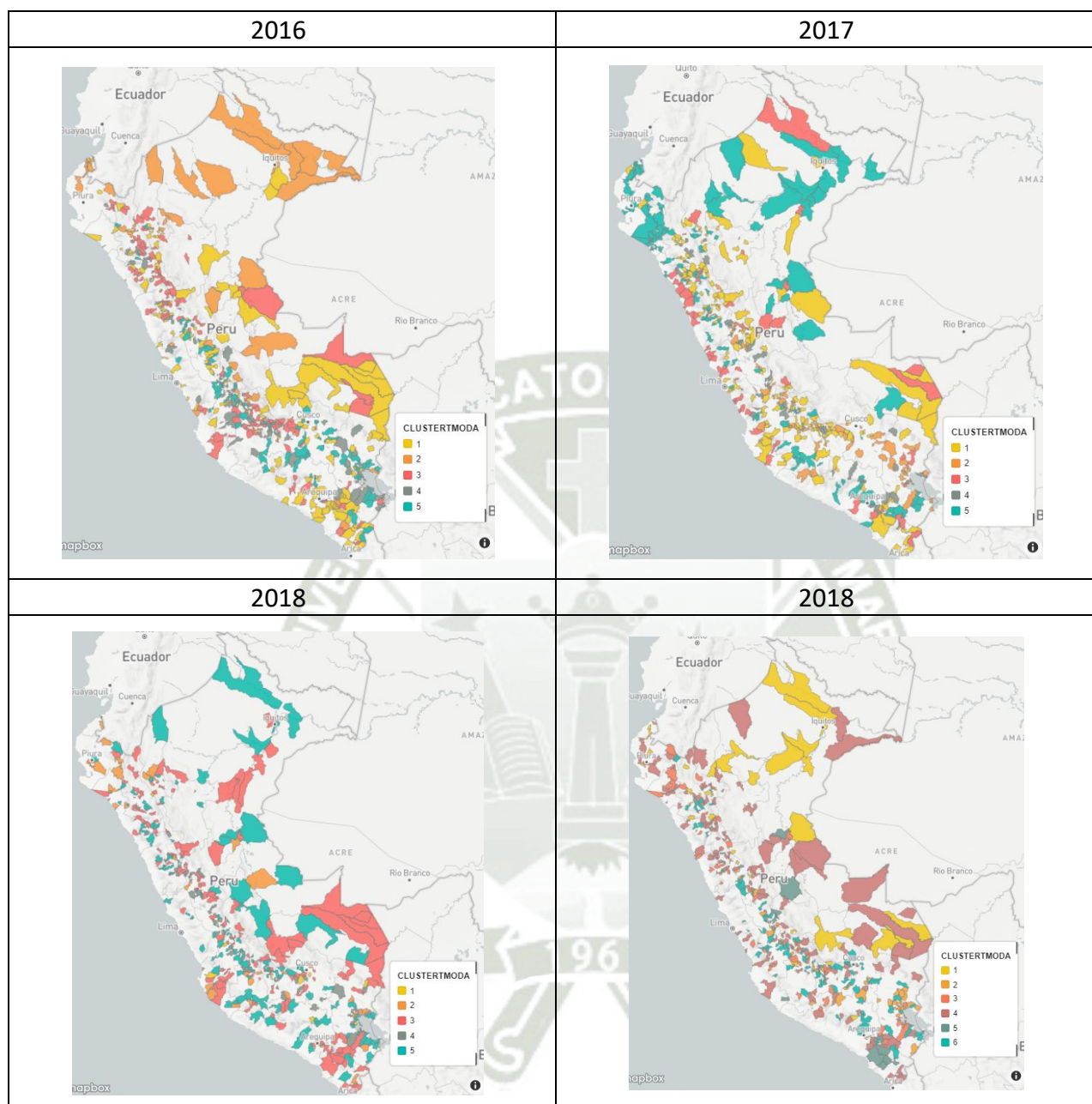
Figura 29.
Principales productos que se perdió su producción.

Cultivos	2016	2017	2018	2019	Total
Otros	5.53%	7.30%	3.60%	3.56%	19.97%
MAIZ AMILACEO	4.71%	5.02%	2.06%	4.78%	16.57%
HABA	6.72%	2.77%	2.94%	3.25%	15.56%
QUINUA	5.21%	1.30%	1.71%	2.78%	11.00%
PAPA BLANCA	1.87%	2.83%	1.81%	2.10%	8.61%
CEBADA GRANO	3.57%	1.14%	1.05%	1.61%	7.34%
PAPA NATIVA	3.92%	0.97%	0.89%	1.16%	6.92%
ARVEJA	2.32%	2.16%	0.83%	1.67%	6.92%
MAIZ CHOCLO	1.10%	1.75%	0.72%	1.05%	4.62%
MAIZ AMARILLO DURO	1.01%	2.31%	0.65%	0.55%	4.51%
OCA	1.44%	0.82%	0.62%	1.05%	3.92%
TRIGO	1.85%	0.57%	0.20%	0.40%	3.02%
OLLUCO	0.85%	0.41%	0.49%	0.81%	2.53%
AVENA FORRAJERA	0.81%	0.23%	0.71%	0.37%	2.12%
YUCA	0.46%	0.67%	0.49%	0.48%	2.11%
TARHUI	1.05%	0.45%	0.20%	0.37%	2.07%
CATZICHUA	0.64%	0.36%	0.50%	0.44%	1.94%
ALFALFA	0.59%	0.23%	0.35%	0.33%	1.50%
MASHUA	0.50%	0.21%	0.24%	0.50%	1.44%
CEBADA FORRAJERA	0.51%	0.12%	0.46%	0.13%	1.22%
FRIJOL BLANCO	0.56%	0.10%	0.14%	0.32%	1.12%
PAPA AMARGA	0.41%	0.30%	0.28%	0.11%	1.11%
PAPA AMARILLA	0.39%	0.38%	0.13%	0.10%	1.00%
AVENA GRANO	0.69%	0.04%	0.07%	0.11%	0.90%
Total	34.58%	26.13%	17.14%	22.80%	100.00%

Fuente: Elaboración Propia.

En la figura 29, se encuentra la distribución geográfica de los clústeres formados para la variable P224E, a lo largo de todo el Perú, existen diversos factores que afectaron la producción agrícola, es por eso que se observa distritos que pertenecen a diferentes clústeres, algunos distritos que están geográficamente cerca pertenecen al mismo clúster, mientras que otros distritos vecinos pertenecen a clústeres diferentes.

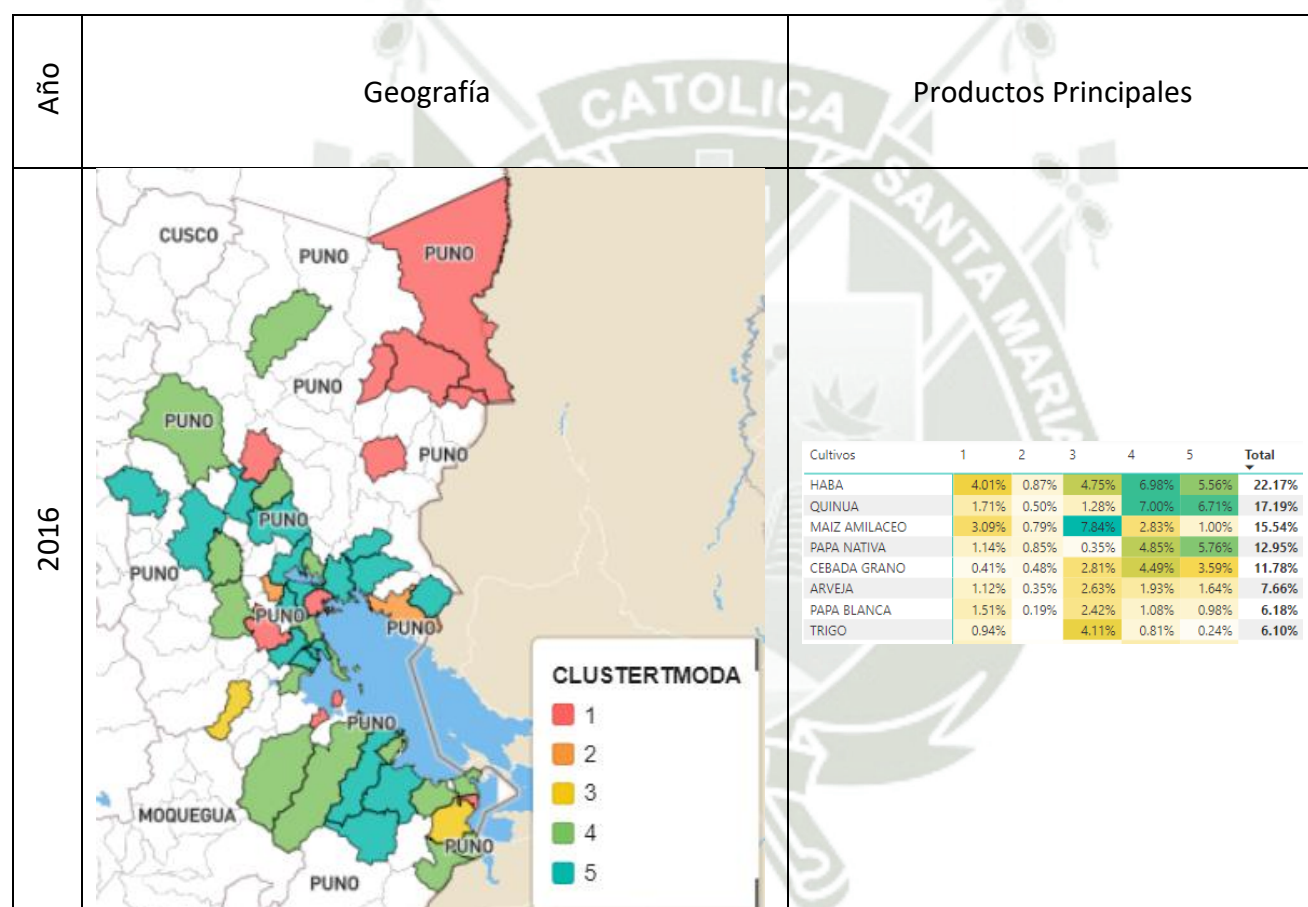
Figura 30.
Mapa de los clústeres P224E.

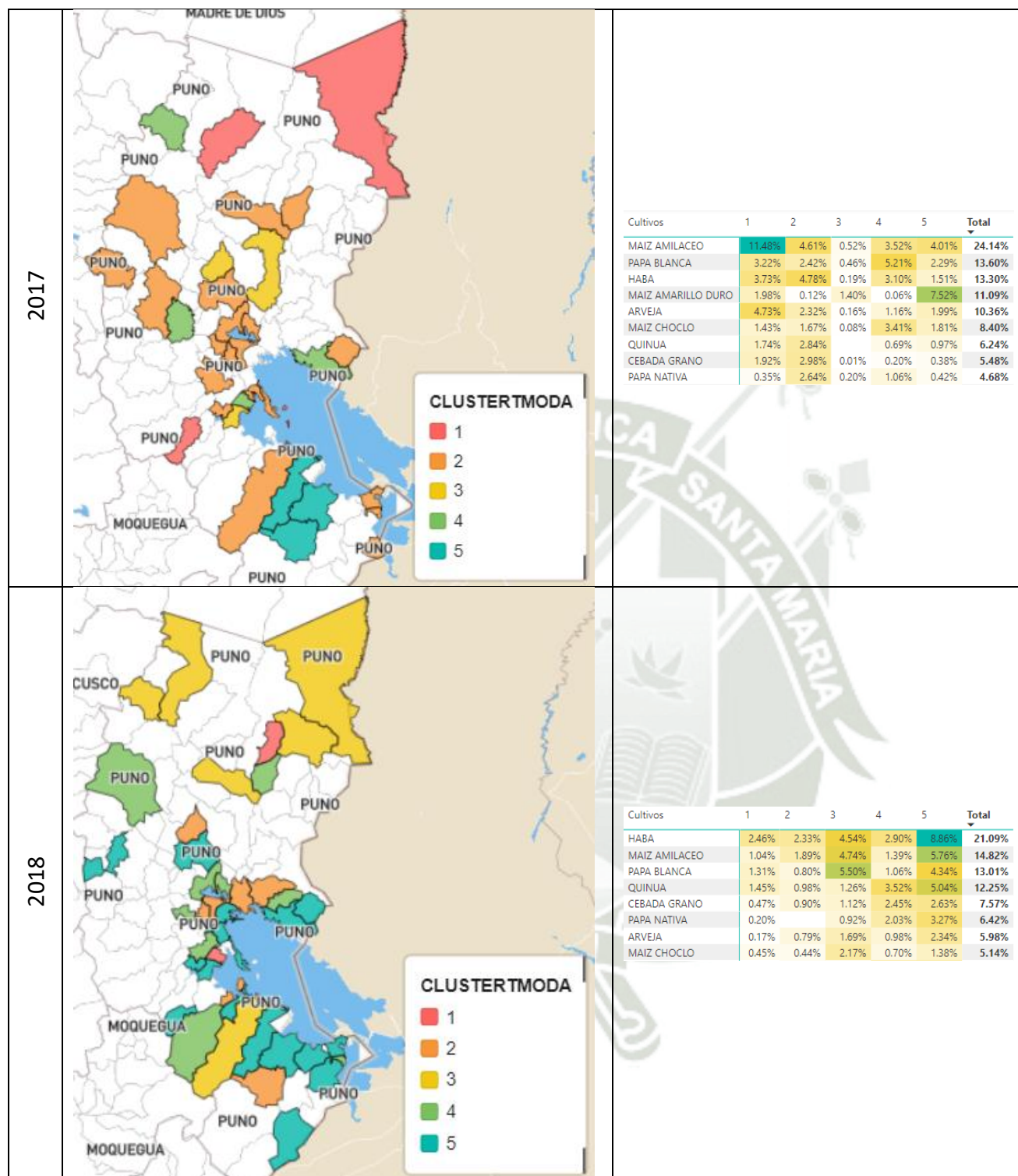


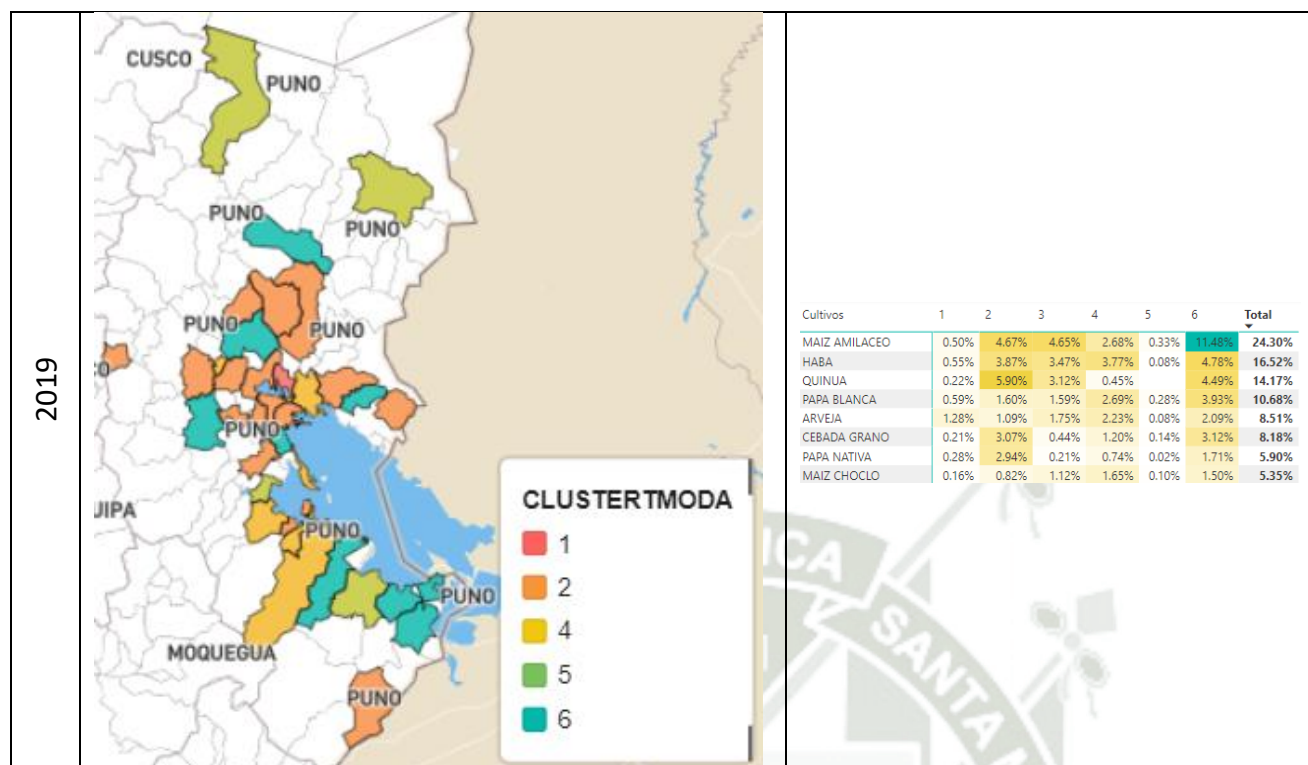
Fuente: Elaboración Propia.

En el departamento de Puno, se clasificaron 13 provincias, en el año 2016 hay 786 observaciones, en el 2017 hay 178, en el año 2018 se tuvieron 337 observaciones y en año 2019 un total de 285 observaciones. Los productos más afectados fueron los que son producidos en la sierra, y una de los departamentos con más observaciones es Puno, como se muestra a continuación en la figura 23, la distribución geográfica de Puno con el clúster que pertenece a cada distrito y los principales productos afectados.

Figura 31.
Características del departamento de Puno.





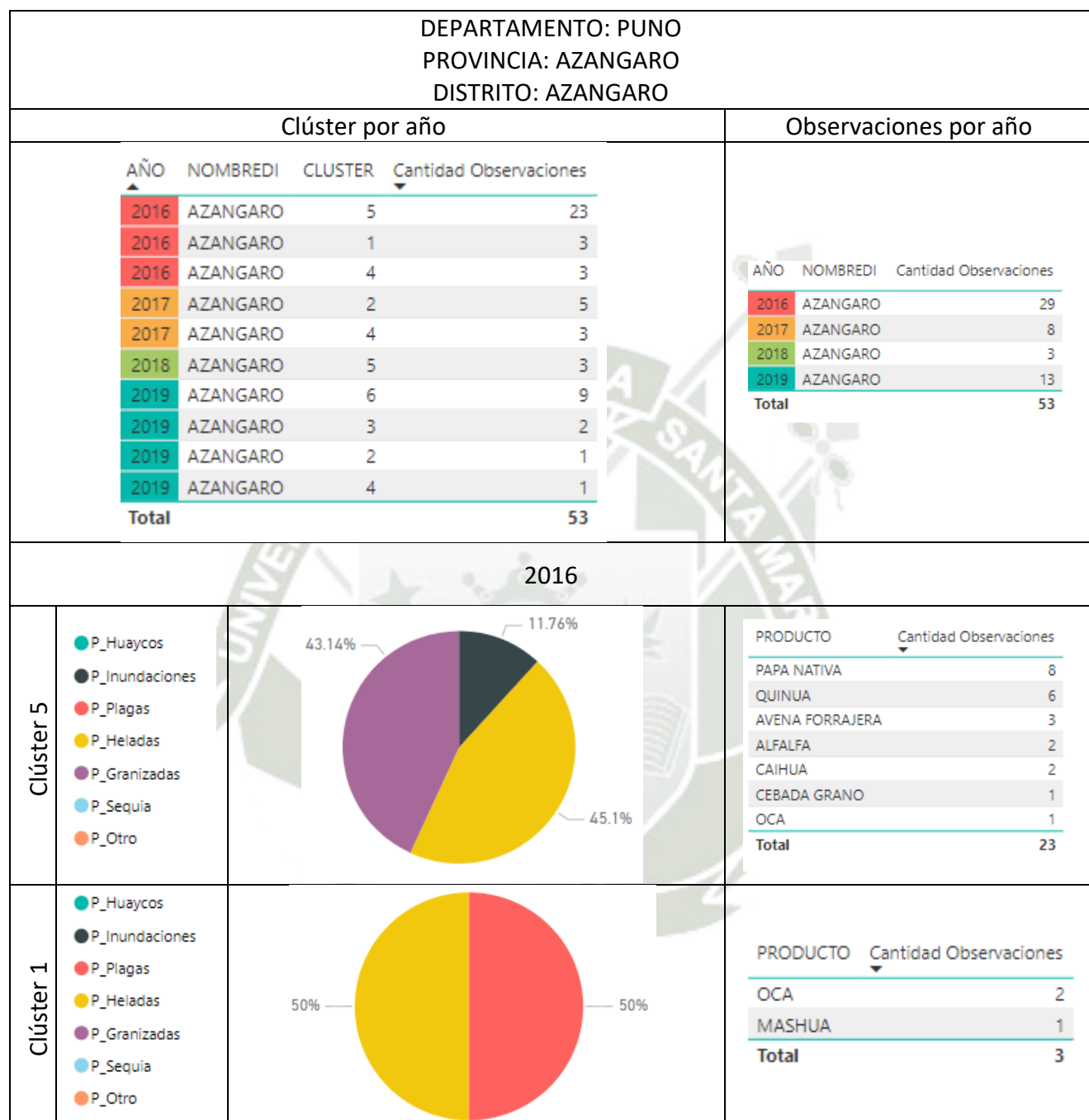


Fuente: Elaboración Propia.

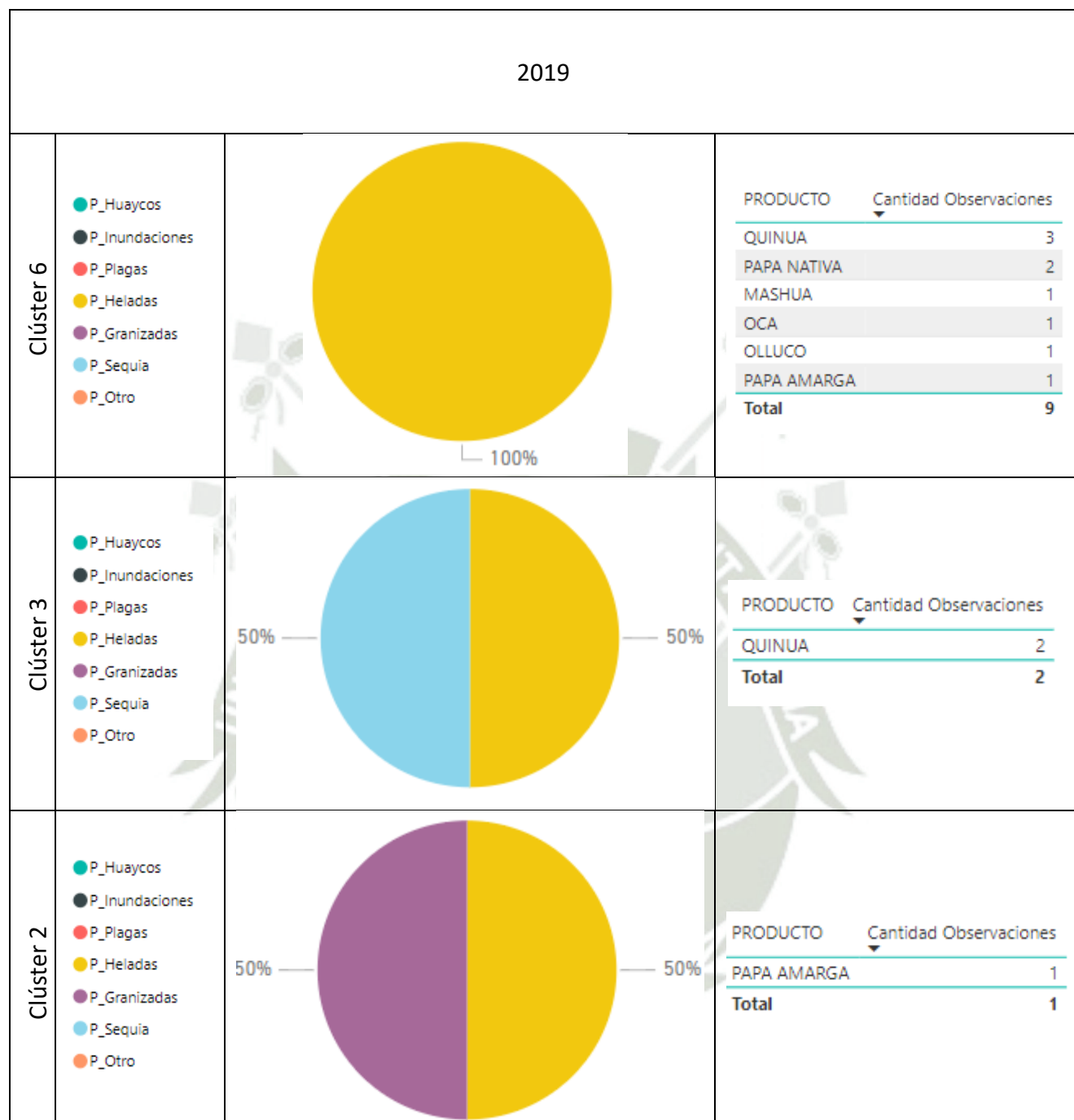
En el año 2016 en el distrito de Azángaro, se perdió producción por heladas, granizadas e inundaciones, se agrupó en el clúster 5, en el clúster 1 están los que perdieron producción por heladas y plagas, el último clúster de este año en Azángaro es el 4 donde están fenómenos como las granizadas las heladas, plagas y sequías. En el año 2017 fueron las heladas, granizadas, sequías y plagas, en el clúster 2, y el clúster 4 que es de heladas plagas y sequía. En el año 2018 solo se tiene un clúster, de inundaciones y heladas. Por último en el año 2019 están únicamente las heladas en el clúster 6, en el clúster 3 están las heladas acompañadas de sequías, en el clúster 4 están otro tipo de motivos de pérdida y por último en el clúster 2 están las heladas acompañadas de granizadas.

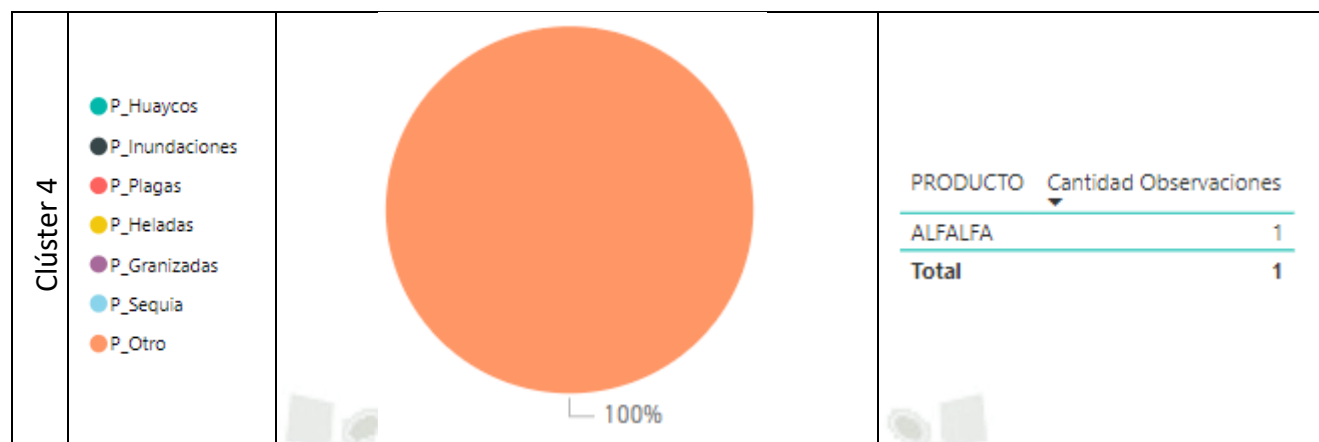
A lo largo de estos años los principales productos perdidos fueron la quinua, papa nativa, alfalfa, caigua, oca, avena forrajera, mashua, entre otros productos.

Figura 32.
Características del distrito de Azángaro.



Clúster 4	P_Huaycos P_Inundaciones P_Plagas P_Heladas P_Granizadas P_Sequia P_Otro	<table><thead><tr><th>PRODUCTO</th><th>Cantidad Observaciones</th></tr></thead><tbody><tr><td>QUINUA</td><td>3</td></tr><tr><td>Total</td><td>3</td></tr></tbody></table>	PRODUCTO	Cantidad Observaciones	QUINUA	3	Total	3				
PRODUCTO	Cantidad Observaciones											
QUINUA	3											
Total	3											
2017												
Clúster 2	P_Huaycos P_Inundaciones P_Plagas P_Heladas P_Granizadas P_Sequia P_Otro	<table><thead><tr><th>PRODUCTO</th><th>Cantidad Observaciones</th></tr></thead><tbody><tr><td>PAPA NATIVA</td><td>2</td></tr><tr><td>QUINUA</td><td>2</td></tr><tr><td>CAIHUA</td><td>1</td></tr><tr><td>Total</td><td>5</td></tr></tbody></table>	PRODUCTO	Cantidad Observaciones	PAPA NATIVA	2	QUINUA	2	CAIHUA	1	Total	5
PRODUCTO	Cantidad Observaciones											
PAPA NATIVA	2											
QUINUA	2											
CAIHUA	1											
Total	5											
Clúster 4	P_Huaycos P_Inundaciones P_Plagas P_Heladas P_Granizadas P_Sequia P_Otro	<table><thead><tr><th>PRODUCTO</th><th>Cantidad Observaciones</th></tr></thead><tbody><tr><td>PAPA NATIVA</td><td>2</td></tr><tr><td>CAIHUA</td><td>1</td></tr><tr><td>Total</td><td>3</td></tr></tbody></table>	PRODUCTO	Cantidad Observaciones	PAPA NATIVA	2	CAIHUA	1	Total	3		
PRODUCTO	Cantidad Observaciones											
PAPA NATIVA	2											
CAIHUA	1											
Total	3											
2018												
Clúster 5	P_Huaycos P_Inundaciones P_Plagas P_Heladas P_Granizadas P_Sequia P_Otro	<table><thead><tr><th>PRODUCTO</th><th>Cantidad Observaciones</th></tr></thead><tbody><tr><td>ALFALFA</td><td>1</td></tr><tr><td>PAPA AMARGA</td><td>1</td></tr><tr><td>PASTO DACTILYS</td><td>1</td></tr><tr><td>Total</td><td>3</td></tr></tbody></table>	PRODUCTO	Cantidad Observaciones	ALFALFA	1	PAPA AMARGA	1	PASTO DACTILYS	1	Total	3
PRODUCTO	Cantidad Observaciones											
ALFALFA	1											
PAPA AMARGA	1											
PASTO DACTILYS	1											
Total	3											





Fuente: Elaboración Propia.

B. Variable P229 ¿Porque considera el cultivo más importante?

Ayuda a determinar cuáles son las características más importantes que un productor toma en cuenta para producir su cultivo más importante

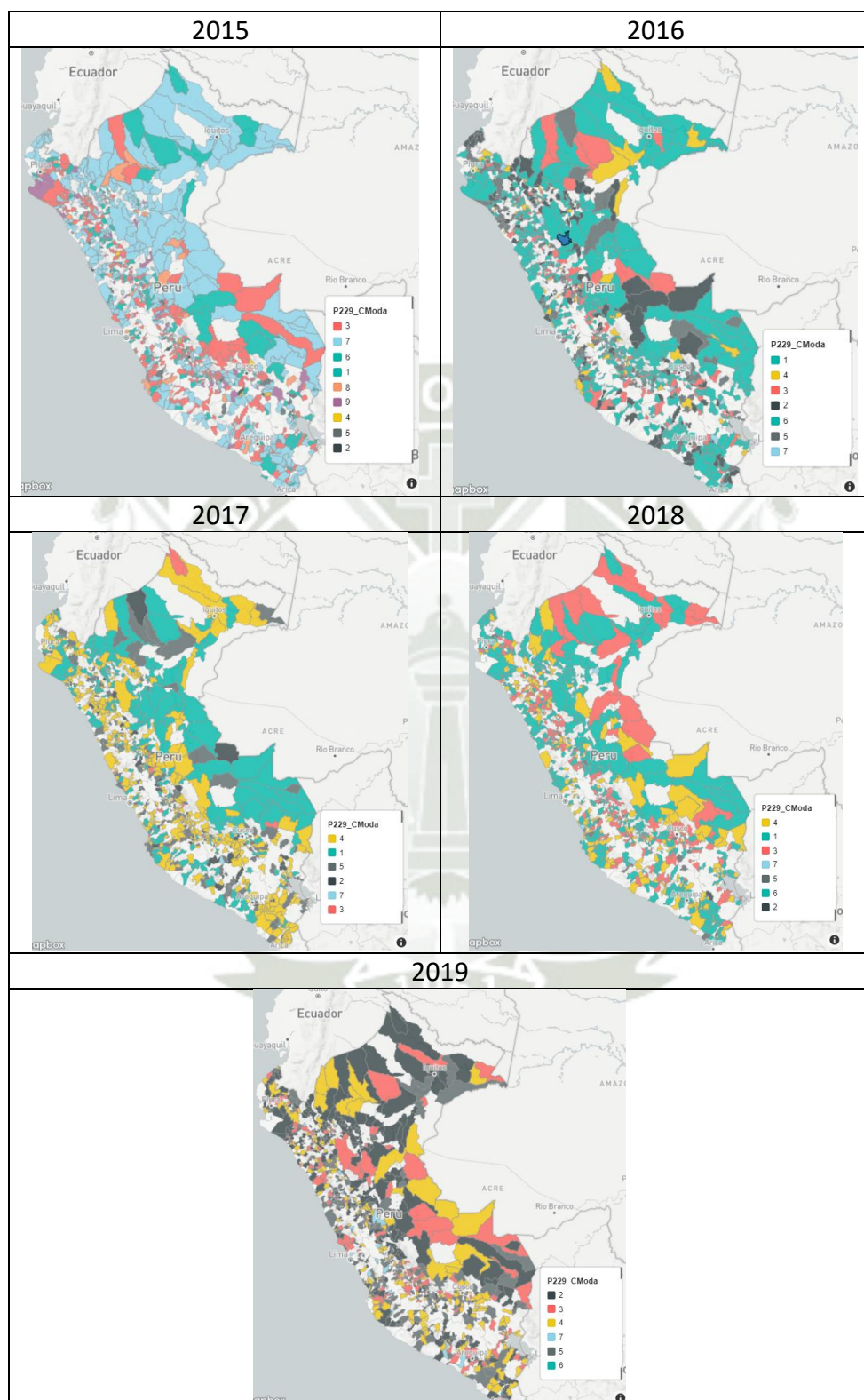
Figura 33.
Características de los clústeres en P229

AÑO	P229 CLUST ER	P_Buen precio de venta	P_Adaptable a condiciones climáticas	P_Requiere poca agua	P_Mercado seguro	P_Corto periodo vegetativo	P_Para alimentar animales	P_Utilizada para autoconsumo	P_Otro	Cantidad Observaciones
2015	7	3,023	420	0	6,105	33	268	558	0	6,520
2015	3	1,019	3,675	0	8	328	1	766	89	3,750
2015	1	2,445	1	163	6	176	0	0	21	2,445
2015	6	70	49	231	692	22	2	1,159	1	1,353
2015	9	34	1,024	7	976	137	98	375	0	1,146
2015	8	168	139	14	504	940	101	25	0	940
2015	5	0	260	440	2	4	109	0	5	531
2015	2	220	226	265	51	84	1	102	0	297
2015	4	36	18	19	1	0	148	148	2	148
2016	1	6,887	1,565	285	3,438	0	92	1,396	2	7,046
2016	2	177	16	89	3,565	23	291	140	18	3,565
2016	5	49	2,105	0	1,513	666	263	1,003	0	2,395
2016	3	54	4	30	666	0	320	1,599	51	1,825
2016	4	572	223	561	415	991	64	17	1	1,452
2016	6	11	1,211	1	6	6	1	16	34	1,232
2016	7	102	85	118	98	279	39	366	0	372
2017	4	6,965	1,667	382	3,555	21	0	1,291	0	6,965
2017	1	200	1,102	2	5,783	559	128	1,080	0	5,783
2017	5	257	2,117	0	121	288	786	860	0	2,702
2017	2	3	257	593	141	46	29	779	87	1,361
2017	3	283	68	16	0	717	0	162	0	717
2017	7	101	43	21	111	20	113	26	0	113
2017	6	0	16	44	46	20	12	37	0	48
2018	1	4,204	16	0	6,408	855	272	0	0	8,601
2018	4	1,357	4,217	352	1,723	0	74	0	0	4,217
2018	3	1,045	1,149	0	1,351	0	330	3,397	0	3,397
2018	5	206	115	590	195	0	39	238	0	590
2018	7	143	311	55	179	523	0	281	0	523
2018	6	17	83	11	136	57	165	116	0	165
2018	2	0	0	0	1	0	1	4	76	76
2019	2	155	1,198	0	6,144	479	0	0	0	6,144
2019	5	3,635	955	0	0	253	0	0	0	3,635
2019	4	180	3,319	680	79	0	0	1	5	3,614
2019	3	3,218	639	203	3,292	0	0	0	0	3,292
2019	7	0	233	30	0	611	0	0	5	611
2019	6	0	7	0	0	1	0	0	138	138
2019	1	18	25	63	30	63	0	0	0	63
Total		36,854	28,558	5,265	47,341	8,202	3,747	15,942	535	87,772

Fuente: Elaboración Propia.

En esta variable existen áreas geográficas definidas, es decir que hay distritos geográficamente cerca que comparten las mismas características.

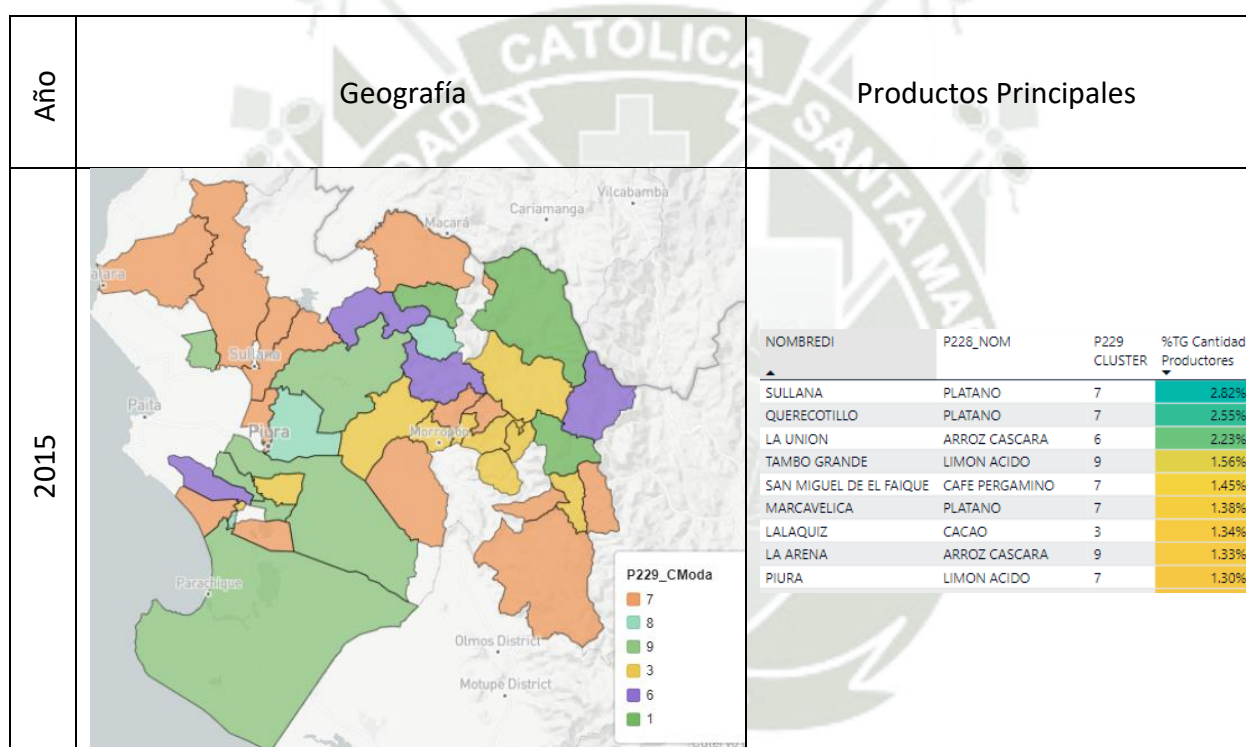
Figura 34.
Mapa de los clústeres P229.

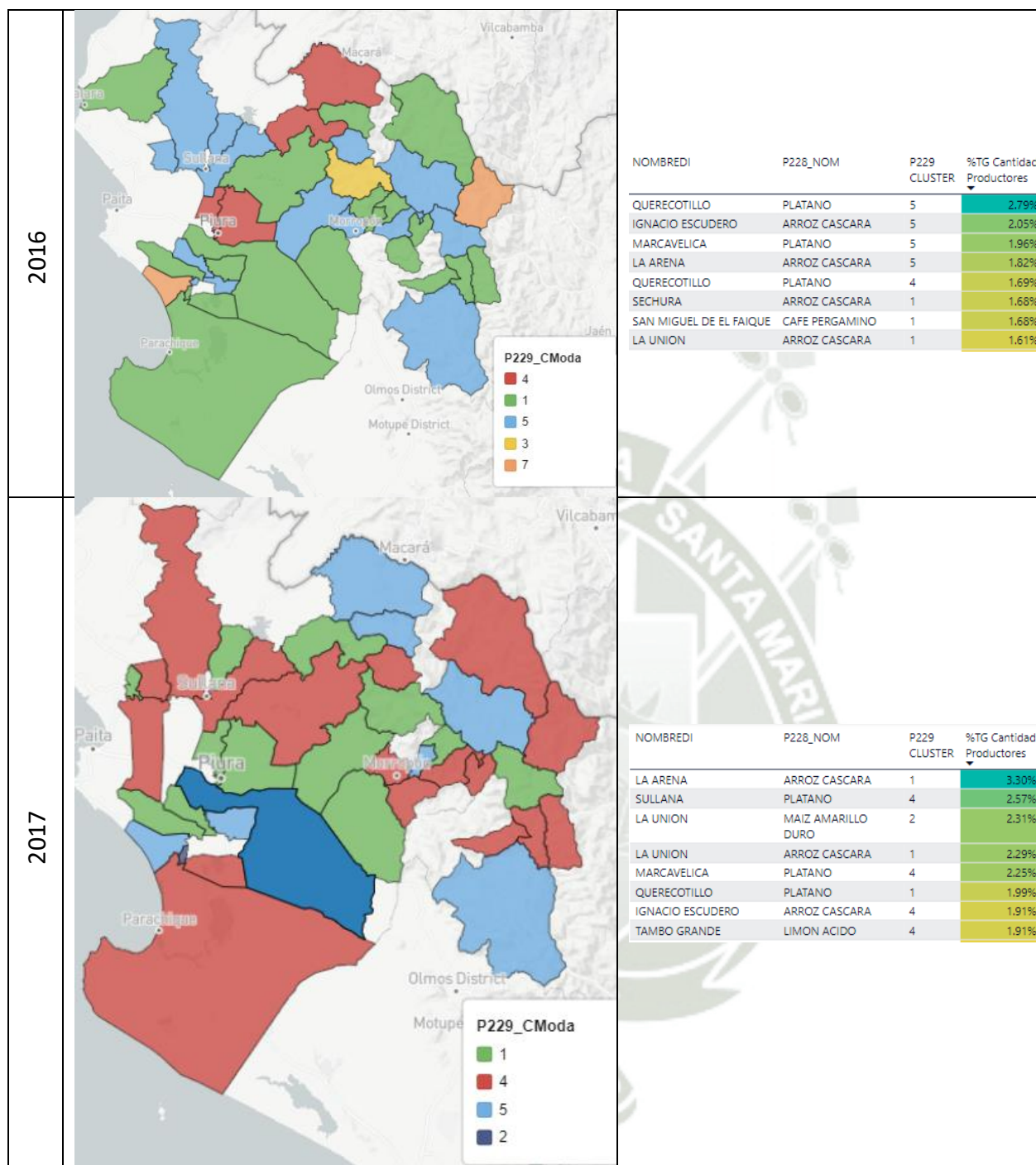


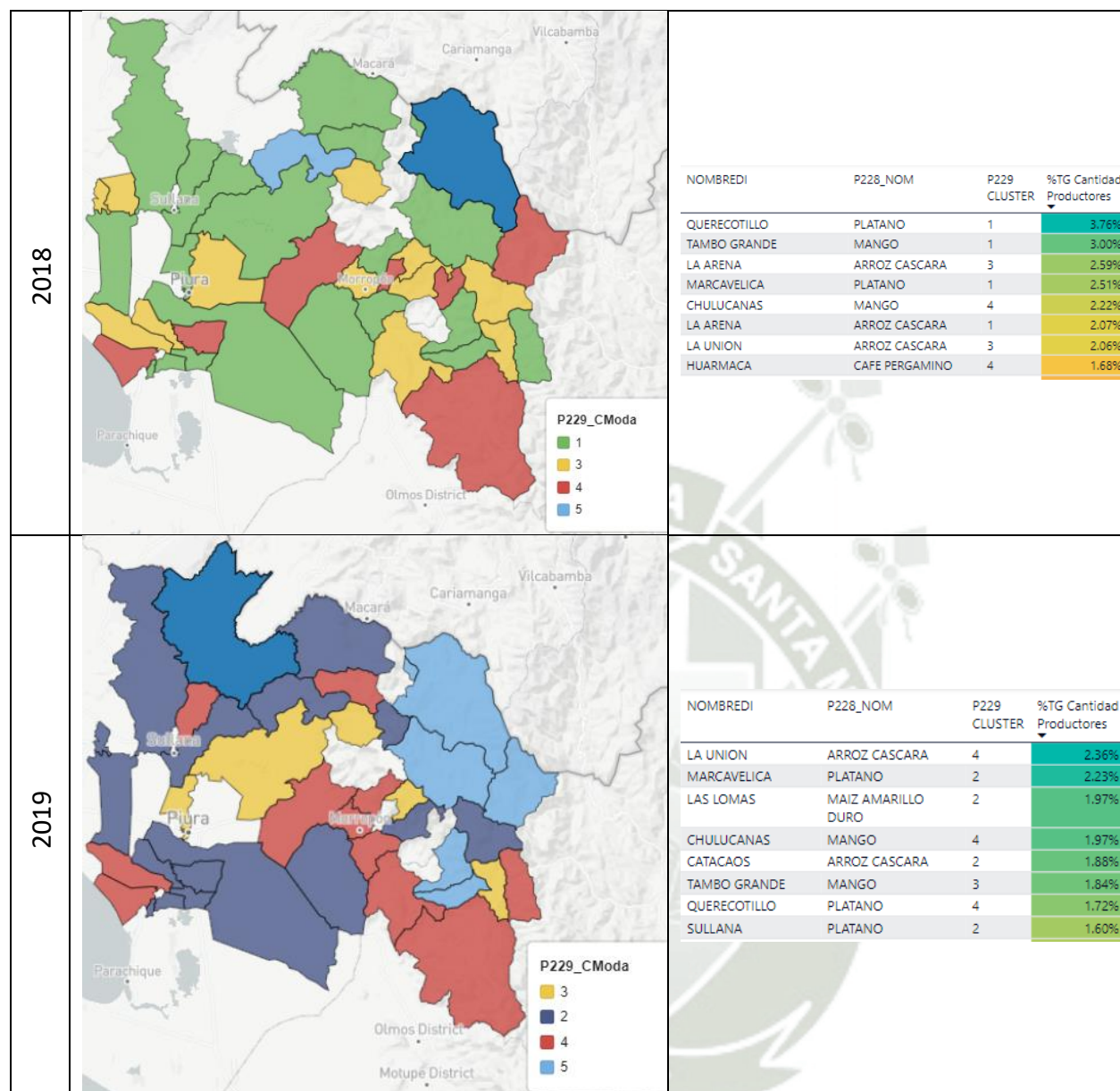
Fuente: Elaboración Propia.

En el Departamento de Piura, existe gran variedad de consideraciones que los productores tienen acerca de su cultivo más importante, la principal es porque consideran que hay un mercado seguro para sus productos, son adaptables a las condiciones climáticas, tienen un buen precio de venta, son utilizadas para el autoconsumo, tienen un corto periodo vegetativo, requieren de poca agua, son utilizadas para alimentar animales u otra razón, siendo los principales productos el arroz cascara, plátano, café, maíz amarillo, limón acido, maíz amiláceo, mango, papa blanca, trigo, papa amarilla, arveja, maní, entre otros.

Figura 35.
Características del departamento de Piura.







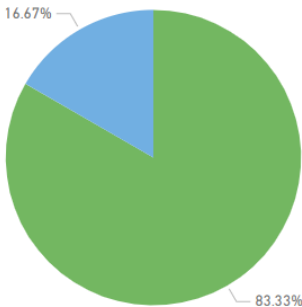
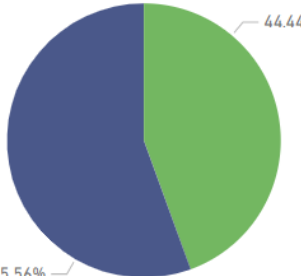
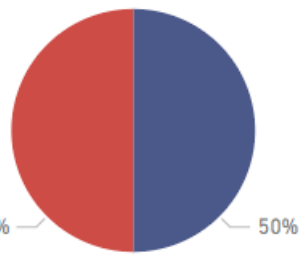
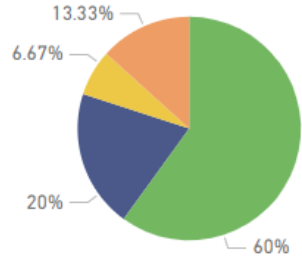
Fuente: Elaboración Propia.

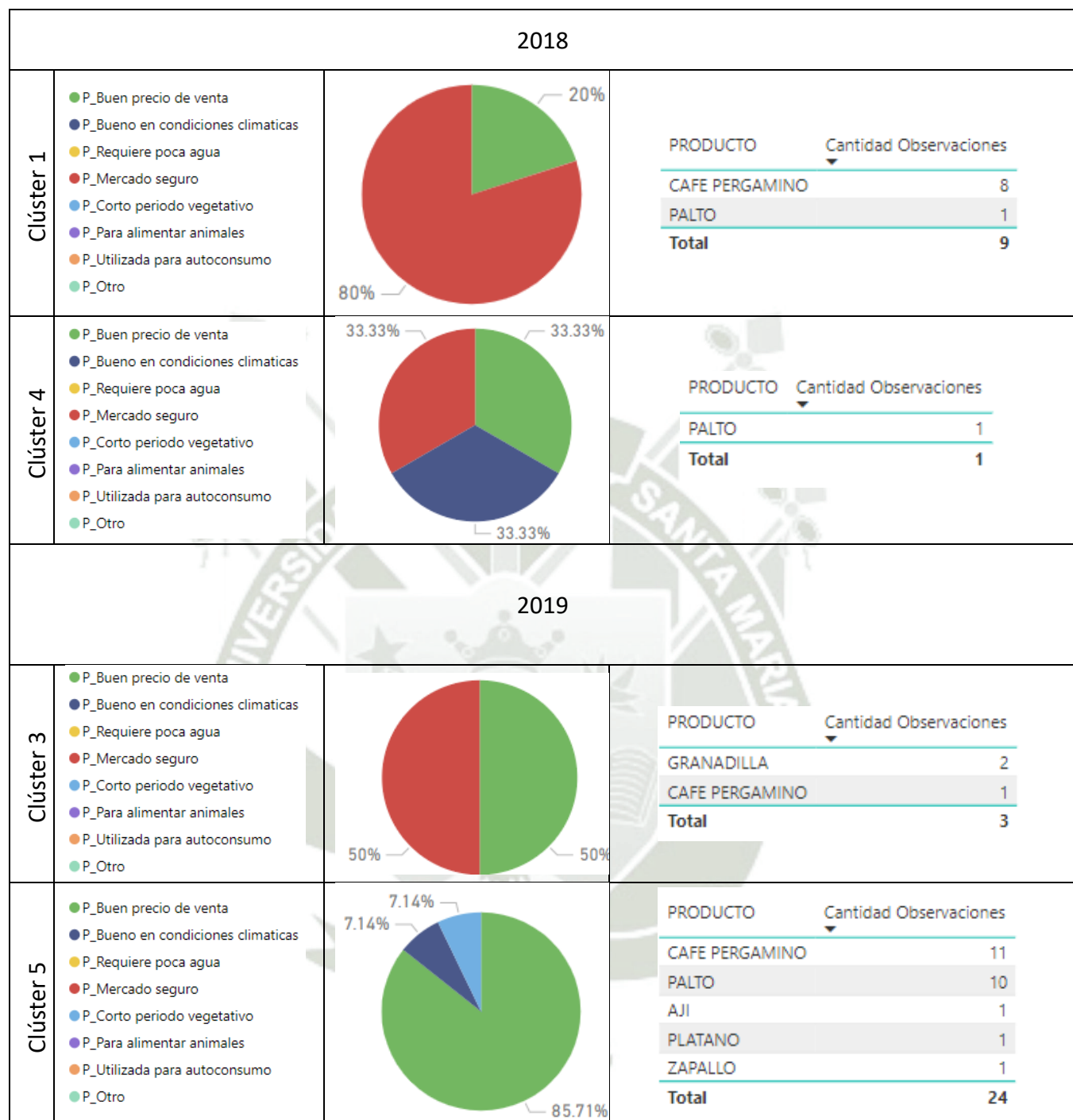
En el distrito de Oxapampa, en el año 2015 hay tres clústeres, el clúster 1 de productores que consideran el buen precio y el buen precio con el corto periodo vegetativo, el clúster 3, productos adaptables a las condiciones climáticas y buen precio de venta, en el clúster 9, productos que se adaptan a las condiciones climáticas y tienen un mercado seguro. En el año 2016 los productores consideraron el buen precio de venta, la adaptación de los productos a las condiciones climáticas, la poca agua y su autoconsumo, agrupado en el clúster 1. En el año 2018, el mercado seguro y el buen precio en el clúster 1, el buen precio de venta adaptable a las condiciones climáticas y

mercado seguro. Por último en el año 2019 consideraron nuevamente el mercado seguro y el buen precio de venta agrupado en el clúster 3, y en el clúster 5 se agrupa los productores que consideraron el buen precio de venta el corto periodo vegetativo y adaptable a las condiciones climáticas. En este distrito del año 2015 al 2019, el café, tiene un buen precio de venta, es adaptable a las condiciones climáticas, y tiene un mercado seguro, la granadilla tiene un buen precio de venta, adaptable a las condiciones climáticas, mercado seguro, corto periodo vegetativo y requiere poca agua, otro producto comercializado es el palto, el cual también tiene un buen precio de venta, adaptable a las condiciones climáticas, y también productos como el rocoto, el ají, zapallo, maíz choclo y plátano son considerados por su buen precio de venta, su adaptación a las condiciones climáticas y su corto periodo vegetativo.

Figura 36.
Características del distrito de Oxapampa.

DEPARTAMENTO: PASCO PROVINCIA: OXAPAMPA DISTRITO: OXAPAMPA			
Clúster por año		Observaciones por año	
AÑO	DISTRITO	P229 CLUSTER	Cantidad Observaciones
2015	OXAPAMPA		2
2015	OXAPAMPA	1	5
2015	OXAPAMPA	3	5
2015	OXAPAMPA	9	2
2016	OXAPAMPA		4
2016	OXAPAMPA	1	9
2018	OXAPAMPA		1
2018	OXAPAMPA	1	9
2018	OXAPAMPA	4	1
2019	OXAPAMPA	3	3
2019	OXAPAMPA	5	24
Total			65
AÑO	DISTRITO	Cantidad Observaciones	
2015	OXAPAMPA	12	
2016	OXAPAMPA	9	
2018	OXAPAMPA	10	
2019	OXAPAMPA	27	
Total		58	

2015															
Clúster 1	- P_Buen precio de venta - P_Bueno en condiciones climaticas - P_Requiere poca agua - P_Mercado seguro - P_Corto periodo vegetativo - P_Para alimentar animales - P_Utilizada para autoconsumo - P_Otro		<table><thead><tr><th>PRODUCTO</th><th>Cantidad Observaciones</th></tr></thead><tbody><tr><td>GRANADILLA</td><td>5</td></tr><tr><td>Total</td><td>5</td></tr></tbody></table>	PRODUCTO	Cantidad Observaciones	GRANADILLA	5	Total	5						
	PRODUCTO	Cantidad Observaciones													
GRANADILLA	5														
Total	5														
Clúster 3	- P_Buen precio de venta - P_Bueno en condiciones climaticas - P_Requiere poca agua - P_Mercado seguro - P_Corto periodo vegetativo - P_Para alimentar animales - P_Utilizada para autoconsumo - P_Otro		<table><thead><tr><th>PRODUCTO</th><th>Cantidad Observaciones</th></tr></thead><tbody><tr><td>GRANADILLA</td><td>2</td></tr><tr><td>CAFE PERGAMINO</td><td>1</td></tr><tr><td>MAIZ CHOCLO</td><td>1</td></tr><tr><td>ROCOTO</td><td>1</td></tr><tr><td>Total</td><td>5</td></tr></tbody></table>	PRODUCTO	Cantidad Observaciones	GRANADILLA	2	CAFE PERGAMINO	1	MAIZ CHOCLO	1	ROCOTO	1	Total	5
	PRODUCTO	Cantidad Observaciones													
GRANADILLA	2														
CAFE PERGAMINO	1														
MAIZ CHOCLO	1														
ROCOTO	1														
Total	5														
Clúster 9	- P_Buen precio de venta - P_Bueno en condiciones climaticas - P_Requiere poca agua - P_Mercado seguro - P_Corto periodo vegetativo - P_Para alimentar animales - P_Utilizada para autoconsumo - P_Otro		<table><thead><tr><th>PRODUCTO</th><th>Cantidad Observaciones</th></tr></thead><tbody><tr><td>GRANADILLA</td><td>2</td></tr><tr><td>Total</td><td>2</td></tr></tbody></table>	PRODUCTO	Cantidad Observaciones	GRANADILLA	2	Total	2						
	PRODUCTO	Cantidad Observaciones													
GRANADILLA	2														
Total	2														
2016															
Clúster 1	- P_Buen precio de venta - P_Bueno en condiciones climaticas - P_Requiere poca agua - P_Mercado seguro - P_Corto periodo vegetativo - P_Para alimentar animales - P_Utilizada para autoconsumo - P_Otro		<table><thead><tr><th>PRODUCTO</th><th>Cantidad Observaciones</th></tr></thead><tbody><tr><td>GRANADILLA</td><td>5</td></tr><tr><td>CAFE PERGAMINO</td><td>2</td></tr><tr><td>ROCOTO</td><td>1</td></tr><tr><td>ZAPALLO</td><td>1</td></tr><tr><td>Total</td><td>9</td></tr></tbody></table>	PRODUCTO	Cantidad Observaciones	GRANADILLA	5	CAFE PERGAMINO	2	ROCOTO	1	ZAPALLO	1	Total	9
	PRODUCTO	Cantidad Observaciones													
GRANADILLA	5														
CAFE PERGAMINO	2														
ROCOTO	1														
ZAPALLO	1														
Total	9														



Fuente: Elaboración Propia.

C. Variable P229A ¿Por qué motivo no vendió su producción?

Esta representa una problemática para la economía de los productores ya que presenta problemas como falta de medios de transporte o derivan sus productos a otras necesidades como alimentar animales o lo usan para semilla.

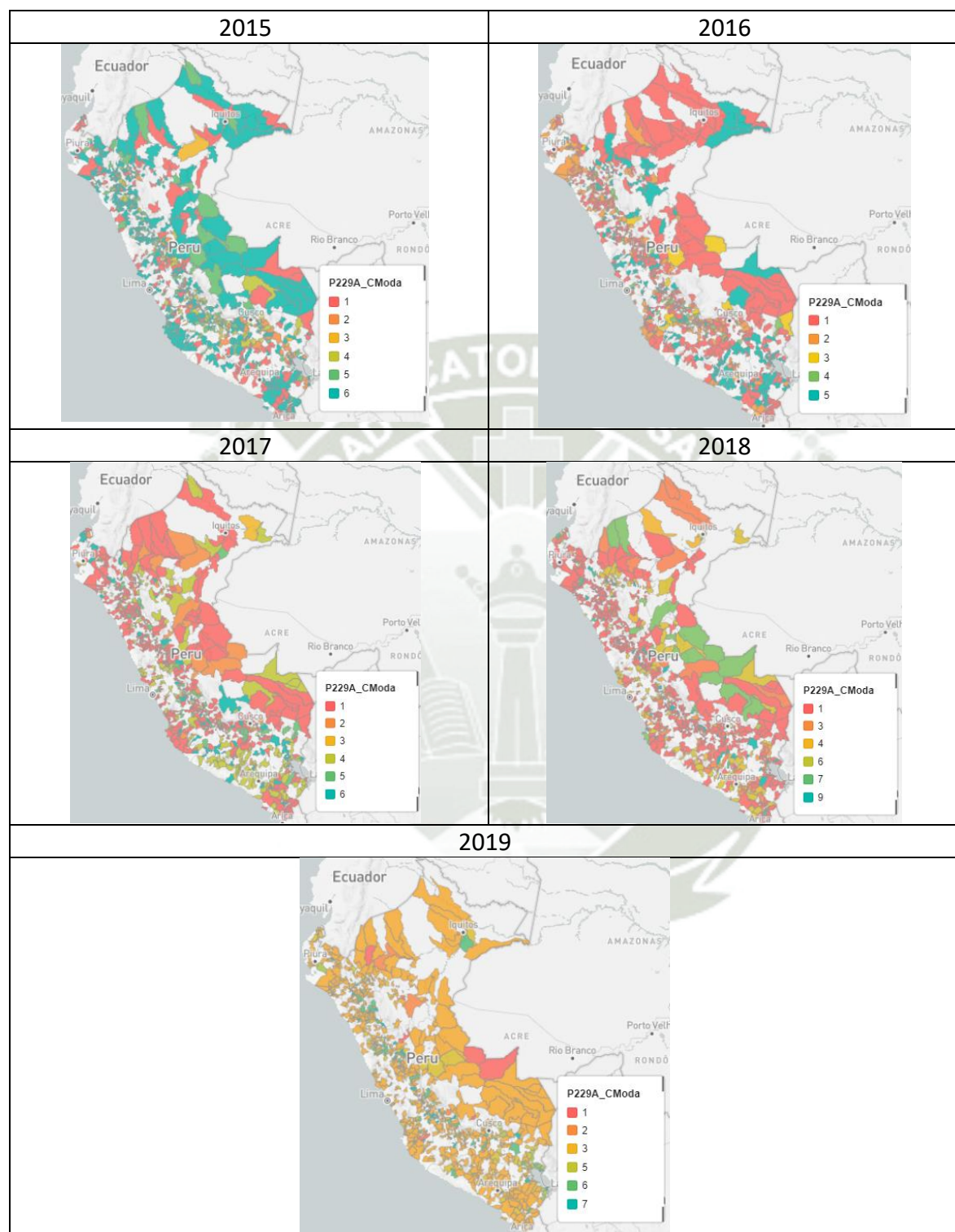
Figura 37.
Características de los clústeres en P22A

AÑO	P229A CLUSTER	PA_Autoconsumo	PA_Alimentar animales	PA_Produccion baja	PA_Falta acceso	PA_Usado para semilla	PA_Otro	Cantidad Observaciones
2015	6	3075	670	0	11	5	65	3,075
2015	1	0	789	317	67	0	175	1,285
2015	5	1228	0	1,090	209	0	22	1,228
2015	4	1110	0	417	68	1,123	0	1,123
2015	2	709	777	489	0	453	1	785
2015	3	91	91	50	91	51	0	91
2016	1	3791	899	0	186	827	21	3,791
2016	2	1551	3	1,765	0	0	90	1,847
2016	5	568	1,369	671	0	201	0	1,373
2016	4	543	10	553	43	555	0	555
2016	3	105	19	121	162	0	0	162
2017	1	3699	0	1,476	0	0	116	3,815
2017	4	975	1,736	0	0	234	0	1,736
2017	5	1049	0	609	0	1,054	0	1,054
2017	6	525	601	757	0	234	0	757
2017	2	309	65	114	309	90	0	309
2017	3	0	9	10	83	2	0	83
2018	1	4502	0	1,443	0	929	41	4,502
2018	2	1026	1,026	275	34	0	0	1,026
2018	5	0	886	27	0	0	32	912
2018	9	252	340	85	0	258	0	340
2018	8	0	0	220	0	9	0	224
2018	3	209	209	199	1	201	13	209
2018	4	55	0	0	5	0	209	209
2018	7	131	6	37	174	0	0	174
2018	6	24	9	12	24	24	0	24
2019	3	3632	1,240	1,612	0	15	0	4,833
2019	4	1140	1,140	0	13	354	0	1,140
2019	6	639	5	0	0	653	1	653
2019	7	535	0	485	21	535	40	535
2019	5	240	227	246	34	237	0	246
2019	2	49	5	33	0	0	224	224
2019	1	112	5	34	153	0	3	153
Total		31874	12,136	13,147	1,688	8,044	1,053	38,473

Fuente: Elaboración Propia.

En el año 2015 para la pregunta P229A, el clúster más común es el 6, de productores que no vendieron su producción ya que la utilizaron para el autoconsumo, alimentar animales, la falta de acceso, la usaron para semilla u otro motivo, en el año 2016 de igual manera, en el año 2017 no se vendió ya que su uso para el autoconsumo, la producción fue baja u otro motivo, para el año 2018, se usó de autoconsumo, la producción fue baja, fue usado para semilla u otro, en el 2019, la causa de no vender fue el autoconsumo, la baja producción, fue usado para alimentar animales o usado de semilla.

Figura 38.
Mapa de los clústeres P229A.

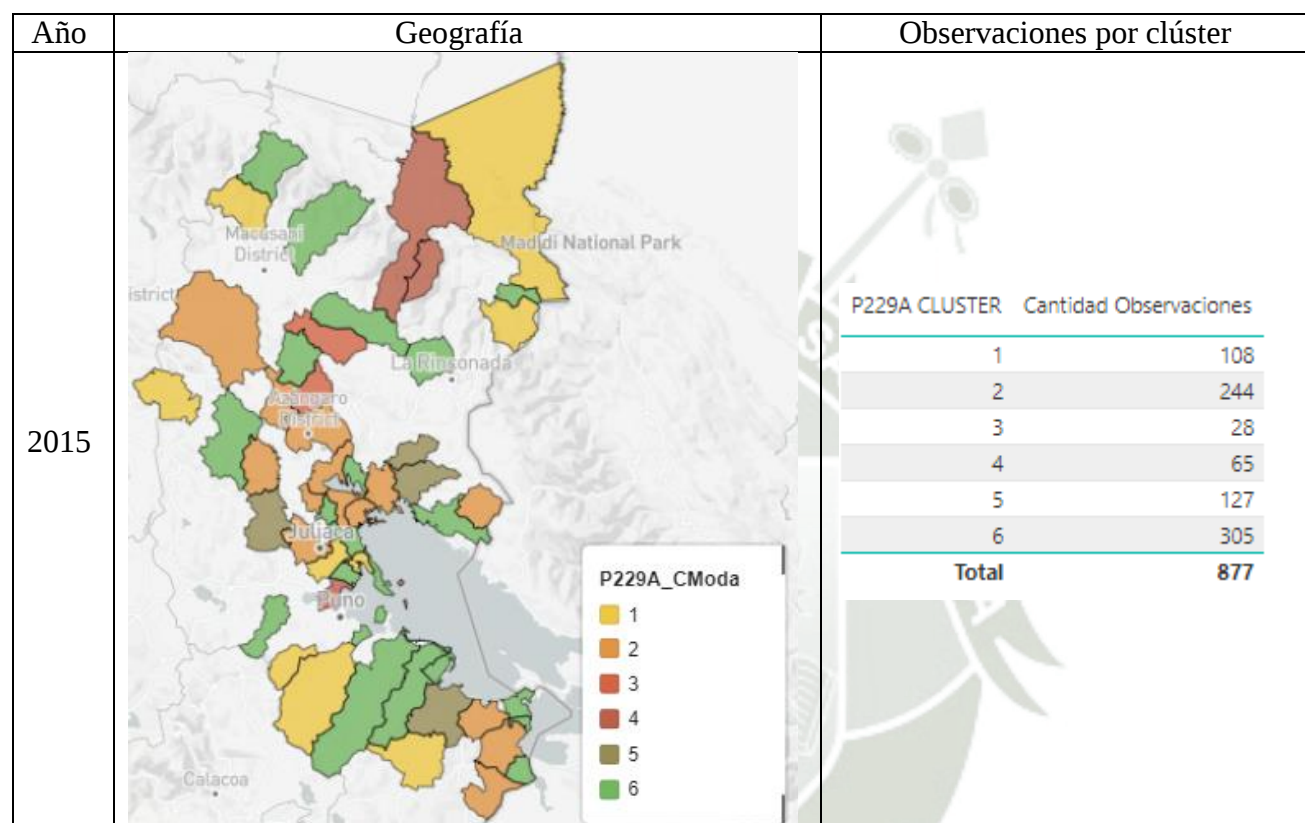


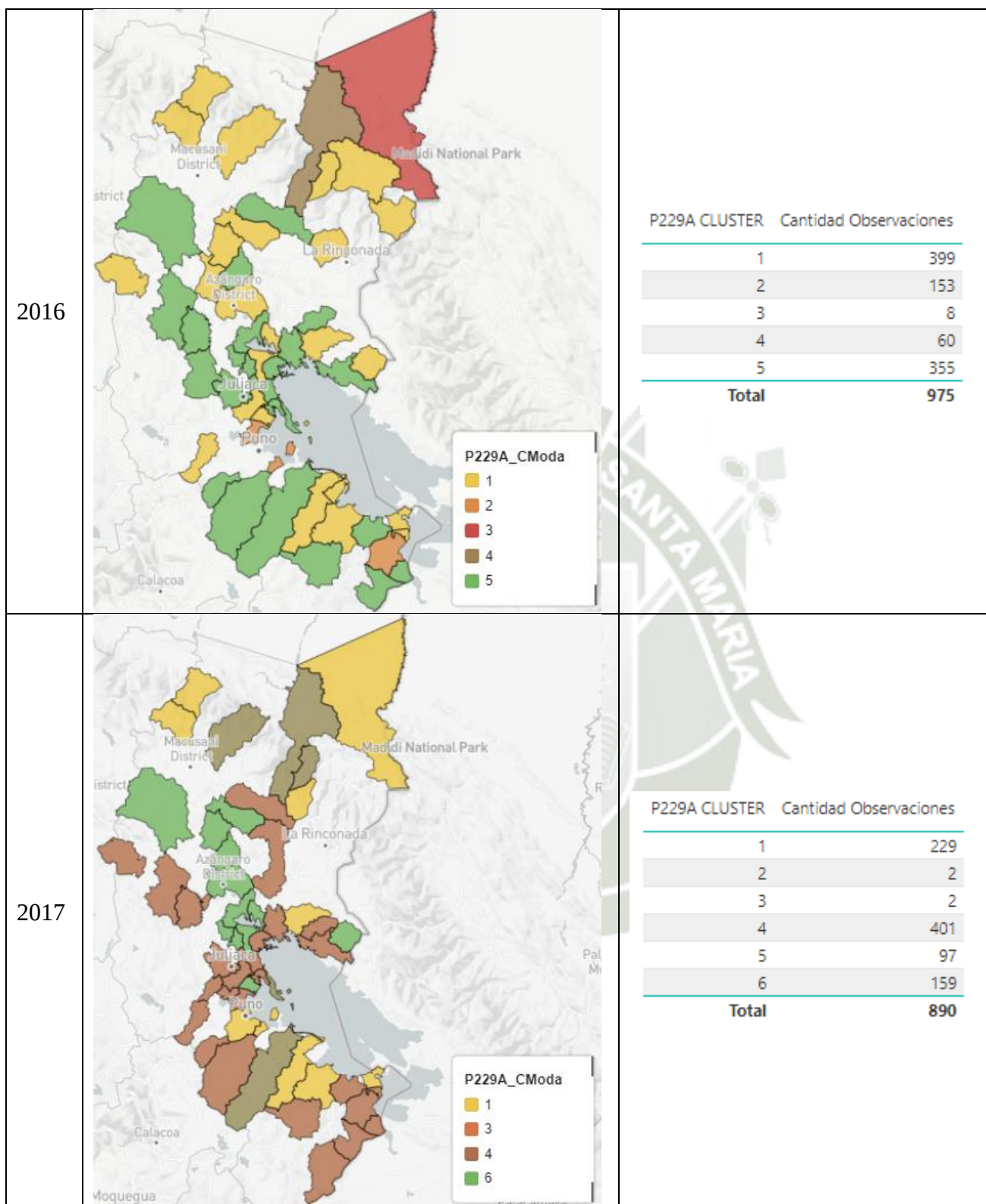
Fuente: Elaboración Propia.

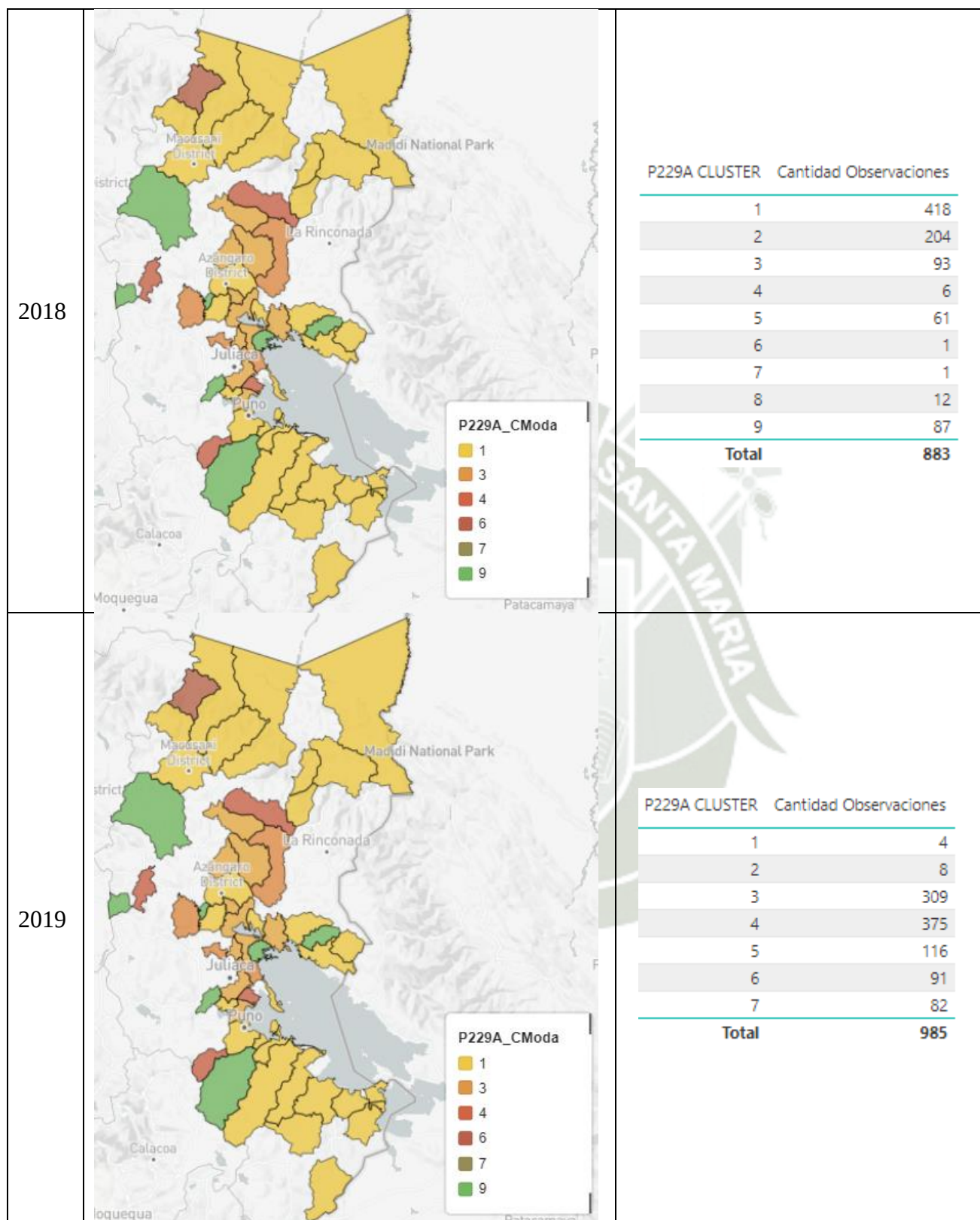
Describiendo el departamento de Puno, del 2015 al 2019, la principal causa de no vender la producción fue debido a que se utilizó para el autoconsumo, seguido de la alimentación de los animales, baja producción, usado para semilla, falta de acceso u otro motivo.

Figura 39.

Características del departamento de Puno.





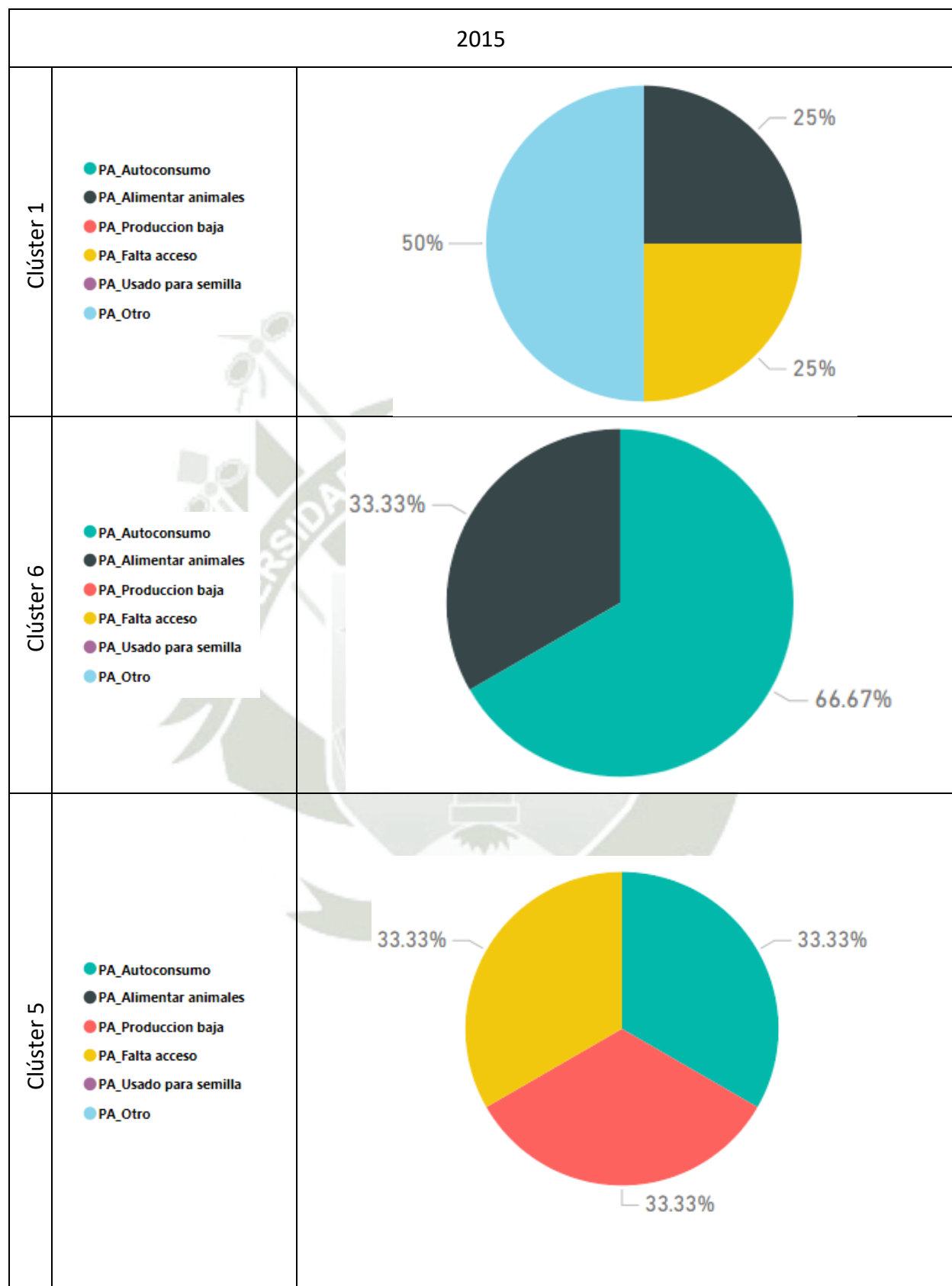


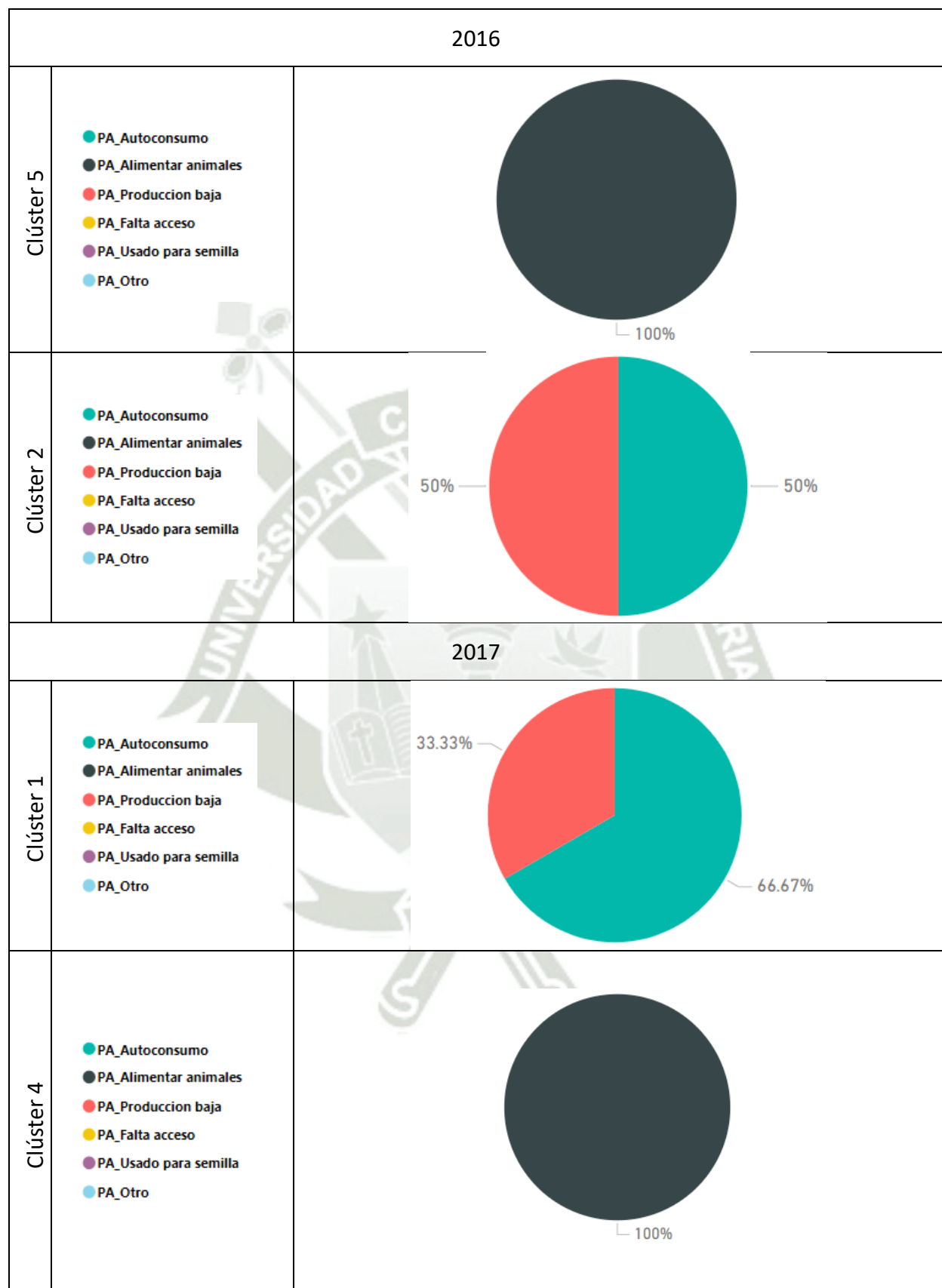
Fuente: Elaboración Propia.

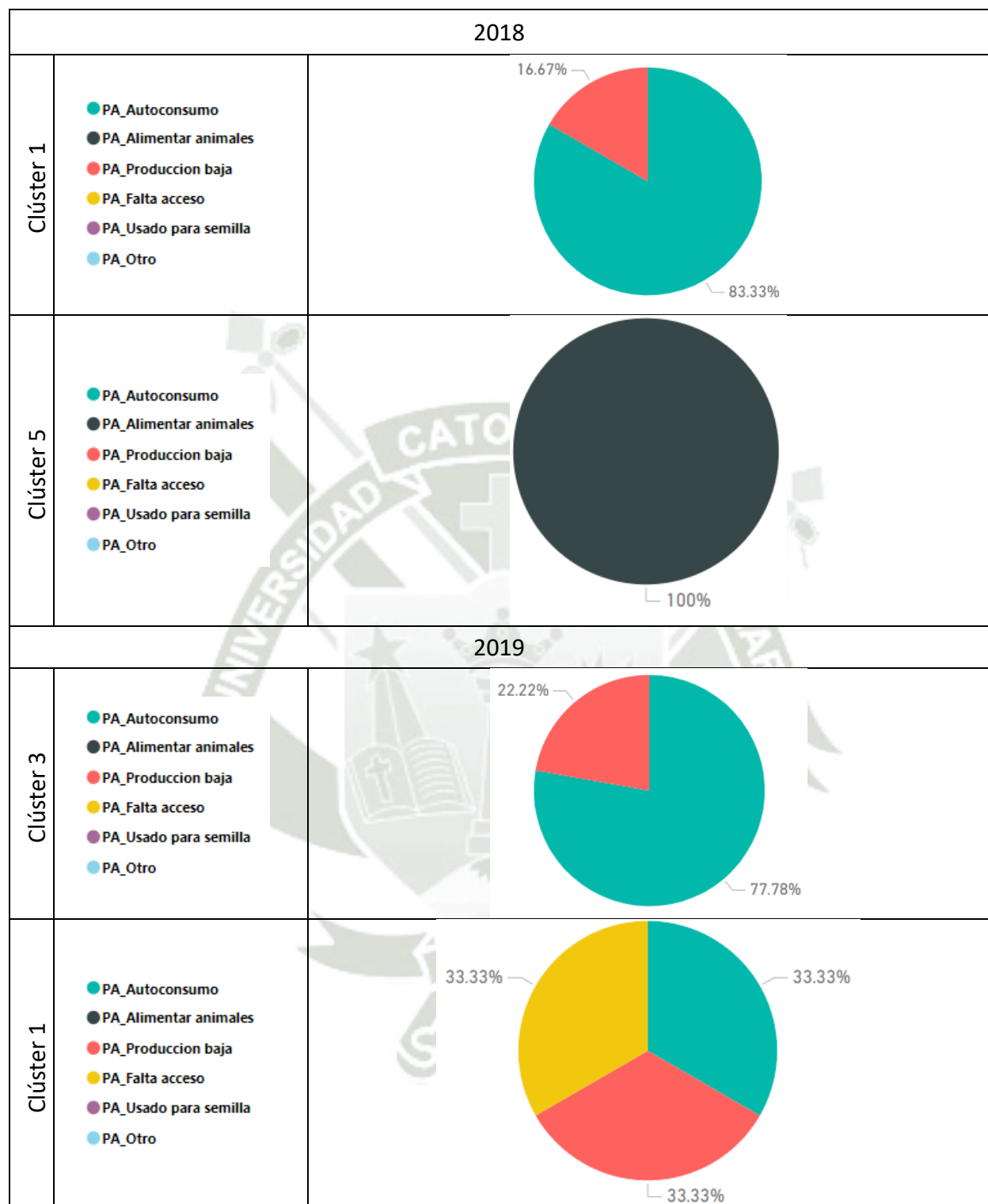
En el distrito de Manu, en Madre de Dios, de un total de 299 observaciones, se clasificaron 32, en el año 2015 en el distrito de Manu no se vendió por falta de acceso, porque se usó para alimentar animales, porque la producción fue baja, fue usado para autoconsumo u otro motivo, en este año se agrupó en el clúster 1, 6 y 5, para el año 2016, fue el autoconsumo, alimento de animales, y la baja producción agrupados en el clúster 5 y 2, en el año siguiente de igual manera pero agrupados en el clúster 1 y 4, en el 2018 igualmente agrupado en el clúster 1 y 5, y para el año 2019 fue por el autoconsumo, la baja producción, y la falta de acceso agrupados en el en el clúster 1 y 3.

Figura 40.
Características del distrito de Manu.

DEPARTAMENTO: MADRE DE DIOS		
PROVINCIA: MANU		
DISTRITO: MANU		
Clústeres por año		Observaciones por año







Fuente: Elaboración Propia.

3.1.3. Módulo 400A Producción pecuaria

A. Variable P403AA: ¿La venta de ganado la realizó en?

Al igual que la variable P223 esta variable ayudará a determinar el destino del ganado, los mercados a los que son vendidos si son locales, regionales o mercados exteriores.

Del 2015 al 2019 la venta del ganado se dio principalmente donde los productores crían sus animales con 85142 observaciones, después en el mercado local con 37261 observaciones, están son las dos principales, después la venden en otros destinos, pero en mucha menor cantidad, estos destinos son el matadero local, el mercado regional, matadero regional, el mercado exterior, mercado de Lima, matadero de Lima agroindustria u otro.

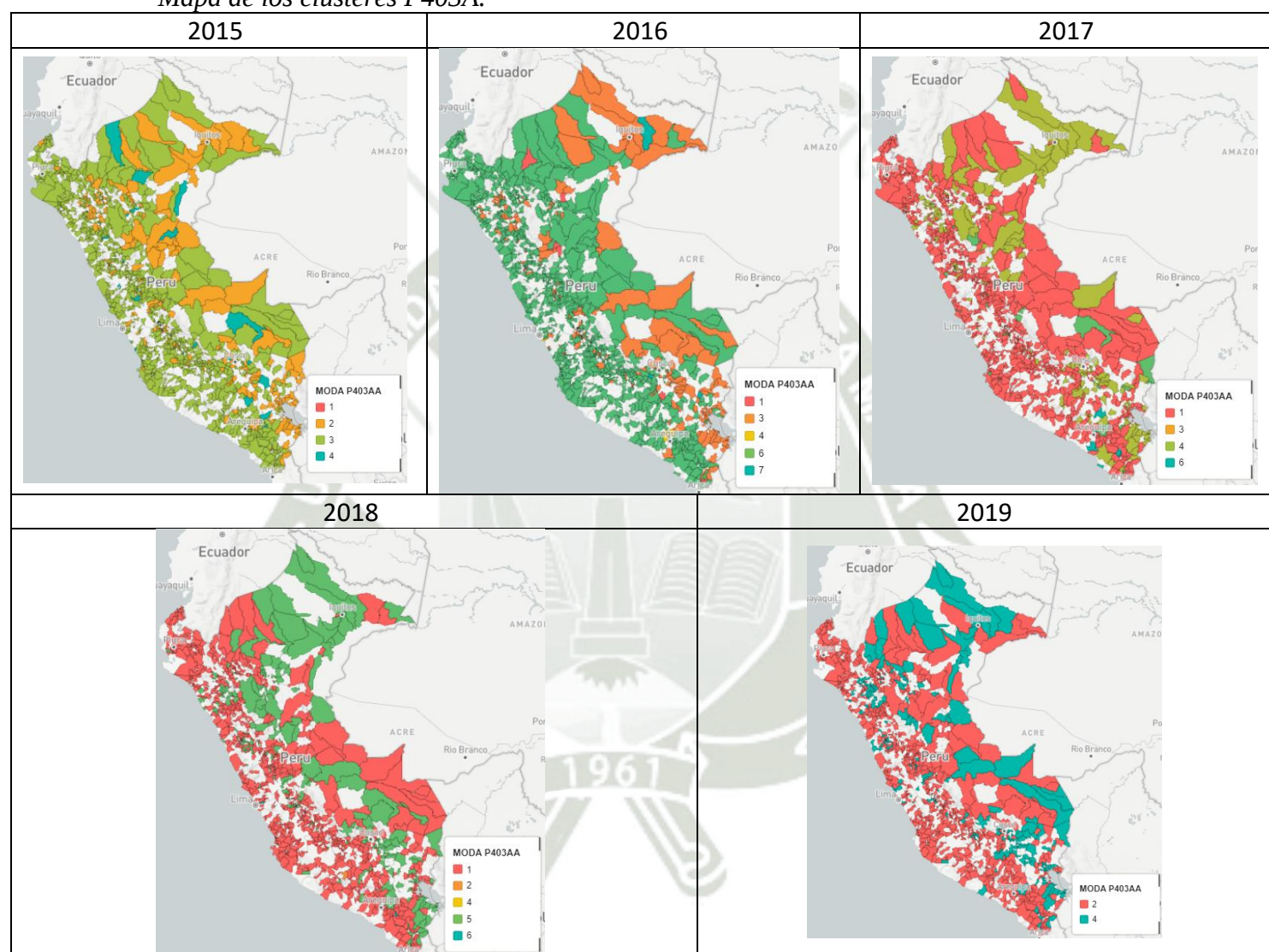
Figura 41.
Características de los clústeres en P403AA.

AÑO	P403AA CLUSTER	P403AA_1 Donde Cria	P403AA_2 Mercado Local	P403AA_3 Matadero Local	P403AA_4 Mercado Regional	P403AA_5 Matadero Regional	P403AA_6 Matadero Lima	P403AA_7 Mercado Exterior	P403AA_8 Agroindustria	P403AA_9 Mercado Lima	P403AA_10 Otro	Cantidad Observaciones
2015	3	14,567	7	1	0	3	0	31	1	0	0	14,592
2015	2	1	5,854	2	0	0	12	0	0	0	0	5,865
2015	4	1,520	1,543	0	202	3	0	0	0	12	0	1,696
2015	1	44	32	283	4	32	1	0	0	0	0	313
2016	6	15,333	0	0	0	0	0	5	9	7	0	15,344
2016	2	3	6,090	0	0	0	0	0	0	16	0	6,102
2016	1	1,608	1,608	0	0	0	0	4	0	0	0	1,608
2016	7	9	1	0	212	40	0	0	0	0	0	251
2016	4	38	32	242	0	0	0	0	0	0	0	242
2016	5	43	42	4	74	0	0	0	0	0	0	74
2016	3	5	2	0	0	0	22	0	0	0	0	22
2017	1	16,146	0	0	0	0	14	0	0	11	0	16,159
2017	4	0	5,338	0	127	0	0	0	0	17	0	5,464
2017	5	1,617	1,590	0	36	0	0	0	0	0	0	1,617
2017	6	37	26	258	3	0	0	0	0	2	0	258
2017	3	1	0	4	0	20	0	0	0	0	0	20
2017	2	1	0	0	0	0	0	7	0	0	1	8
2018	1	15,415	2	0	1	6	3	0	0	0	0	15,415
2018	5	1,812	7,684	0	0	3	0	0	0	0	0	7,684
2018	6	37	29	226	0	0	0	0	0	0	0	226
2018	2	26	20	0	193	0	0	0	0	0	0	193
2018	4	3	0	0	0	20	10	0	1	14	0	45
2018	3	2	1	0	1	0	0	10	0	0	0	10
2018	7	0	1	1	1	1	0	0	0	0	0	1
2019	2	14,990	1	0	161	28	4	0	0	0	0	15,137
2019	4	1,849	7,344	20	27	0	0	0	0	0	0	7,344
2019	1	20	0	182	0	0	0	0	1	5	0	186
2019	3	15	14	0	9	0	0	64	0	0	0	64
Total		85,142	37,261	1,223	1,051	156	66	121	12	84	1	115,940

Fuente: Elaboración Propia.

En el periodo 2015-2019 los Clústeres más comunes son los que predominan en estos 5 años, un clúster es caracterizado por productores que venden el ganado vender el ganado en el lugar donde cría, y el otro grupo de productores que pertenecen a un clúster que se caracterizan por vender en el mercado local y donde los crían.

Figura 42.
Mapa de los clústeres P403A.

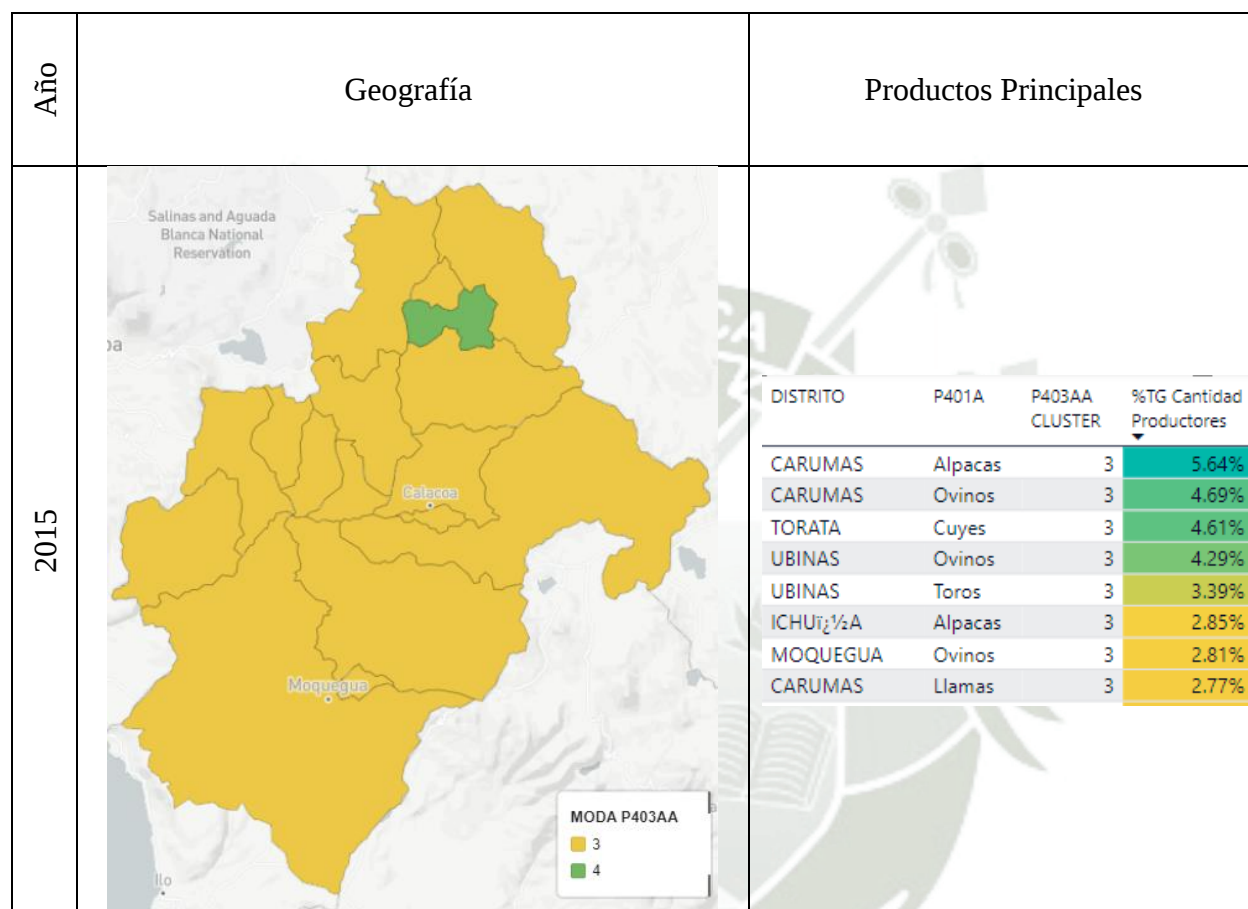


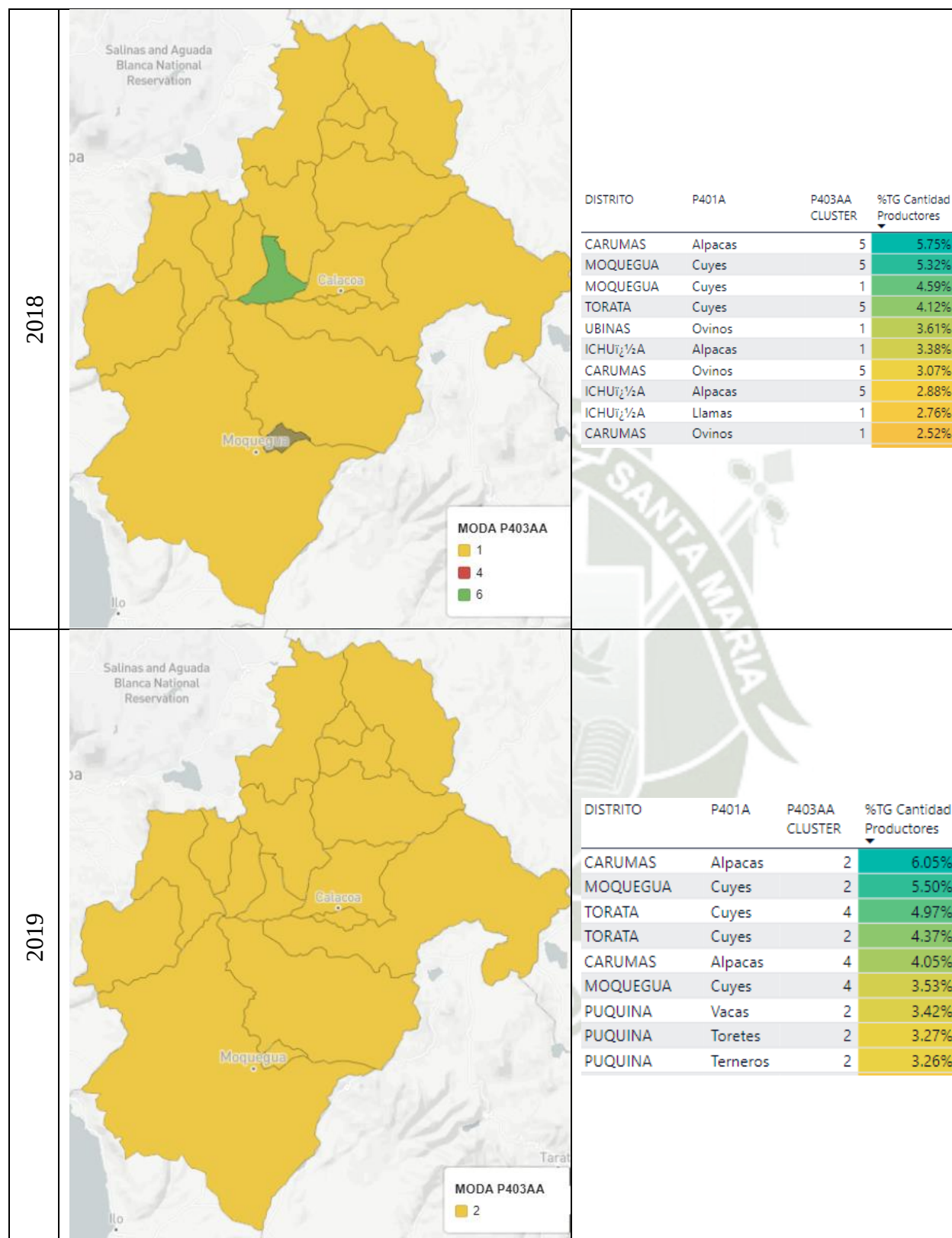
Fuente: Elaboración Propia.

Análisis departamento de Moquegua, en el año 2015 se vendió en el lugar donde se crió, el mercado local, matadero local, mercado regional, matadero regional, agroindustria o mercado de Lima. en los 5 años fueron únicos destinos de venta. Los animales vendidos en estos destinos

fueron principalmente los cuyes, ovinos, alpaca, toretes, toros, porcinos, llamas, gallinas, entre otros.

Figura 43.
Características del departamento de Moquegua.





Fuente: Elaboración Propia.

El Río Santiago Distrito de Amazonas, se tiene 908 observaciones, de las cuales 309 fueron clasificadas, en el año 2015 se vendió en el lugar donde cría, en el mercado exterior agrupado en el clúster 3, en el clúster 4 vendieron donde crían y el ,mercado local, y en el clúster 2 productores que solo vendieron en el mercado local, en el año 2016 se vendió en donde se cría, y el mercado local, agrupados en los clústeres 6, 1 y 2, para el año 2017 en el mercado local, donde se cría, y mercado regional, clústeres 4, 1 y 5, el 2018 donde se cría y el mercado local, clústeres 1 y 5, finalmente el año 2019, agrupa de productores que vendieron en donde se cría, mercado local, matadero local y agroindustria, en los clústeres 1, 2, 3 y 4. Donde los animales comercializados en este distrito son las gallinas, gallos, pollos de engorde, patos, cuyes, porcinos, pavos y toretes. Siendo las gallinas, gallos, pollos de engorde y patos los únicos que solo comercializan en el mercado exterior en el año 2019.

Figura 44.
Características del distrito de Río Santiago.

DEPARTAMENTO: AMAZONAS PROVINCIA: CONDORCANQUI DISTRITO: RIO SANTIAGO																																																																			
Clúster por año		Observaciones por año																																																																	
<table><thead><tr><th>AÑO</th><th>Cantidad Observaciones</th><th>P403AA CLUSTER</th></tr></thead><tbody><tr><td>2015</td><td>61</td><td>3</td></tr><tr><td>2015</td><td>9</td><td>4</td></tr><tr><td>2015</td><td>7</td><td>2</td></tr><tr><td>2016</td><td>33</td><td>6</td></tr><tr><td>2016</td><td>22</td><td>1</td></tr><tr><td>2016</td><td>5</td><td>2</td></tr><tr><td>2017</td><td>15</td><td>4</td></tr><tr><td>2017</td><td>11</td><td>1</td></tr><tr><td>2017</td><td>4</td><td>5</td></tr><tr><td>2018</td><td>30</td><td>1</td></tr><tr><td>2018</td><td>17</td><td>5</td></tr><tr><td>2019</td><td>34</td><td>4</td></tr><tr><td>2019</td><td>30</td><td>2</td></tr><tr><td>2019</td><td>27</td><td>3</td></tr><tr><td>2019</td><td>4</td><td>1</td></tr><tr><td>Total</td><td>309</td><td></td></tr></tbody></table>		AÑO	Cantidad Observaciones	P403AA CLUSTER	2015	61	3	2015	9	4	2015	7	2	2016	33	6	2016	22	1	2016	5	2	2017	15	4	2017	11	1	2017	4	5	2018	30	1	2018	17	5	2019	34	4	2019	30	2	2019	27	3	2019	4	1	Total	309		<table><thead><tr><th>AÑO</th><th>Cantidad Observaciones</th></tr></thead><tbody><tr><td>2015</td><td>135</td></tr><tr><td>2016</td><td>130</td></tr><tr><td>2017</td><td>222</td></tr><tr><td>2018</td><td>242</td></tr><tr><td>2019</td><td>179</td></tr><tr><td>Total</td><td>908</td></tr></tbody></table>	AÑO	Cantidad Observaciones	2015	135	2016	130	2017	222	2018	242	2019	179	Total	908
AÑO	Cantidad Observaciones	P403AA CLUSTER																																																																	
2015	61	3																																																																	
2015	9	4																																																																	
2015	7	2																																																																	
2016	33	6																																																																	
2016	22	1																																																																	
2016	5	2																																																																	
2017	15	4																																																																	
2017	11	1																																																																	
2017	4	5																																																																	
2018	30	1																																																																	
2018	17	5																																																																	
2019	34	4																																																																	
2019	30	2																																																																	
2019	27	3																																																																	
2019	4	1																																																																	
Total	309																																																																		
AÑO	Cantidad Observaciones																																																																		
2015	135																																																																		
2016	130																																																																		
2017	222																																																																		
2018	242																																																																		
2019	179																																																																		
Total	908																																																																		

2015

Clúster 3

- P403AA_1 Donde Cria
- P403AA_2 Mercado Local
- P403AA_3 Matadero Local
- P403AA_4 Mercado Regional
- P403AA_5 Matadero Regional
- P403AA_6 Matadero Lima
- P403AA_7 Mercado Exterior
- P403AA_8 Agroindustria
- P403AA_9 Mercado Lima
- P403AA_10 Otro

Categoría	Porcentaje
P403AA_1 Donde Cria	88.52%
P403AA_7 Mercado Exterior	11.48%

ANIMALES	CANTIDAD OBSERVACIONES
Gallinas	23
Gallos	17
Patos	8
Pollos de engorde	8
Porcinos	2
Cuyes	1
Pavos	1
Toretas	1
Total	61

Clúster 4

- P403AA_1 Donde Cria
- P403AA_2 Mercado Local
- P403AA_3 Matadero Local
- P403AA_4 Mercado Regional
- P403AA_5 Matadero Regional
- P403AA_6 Matadero Lima
- P403AA_7 Mercado Exterior
- P403AA_8 Agroindustria
- P403AA_9 Mercado Lima
- P403AA_10 Otro

Categoría	Porcentaje
P403AA_1 Donde Cria	50%
P403AA_2 Mercado Local	50%

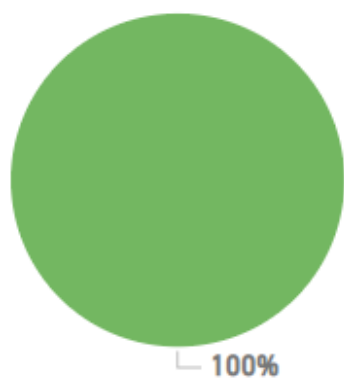
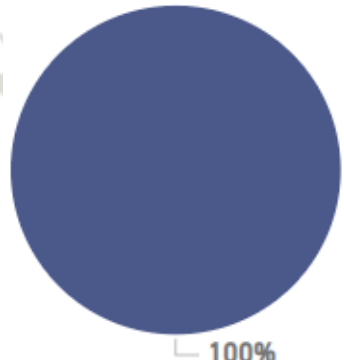
ANIMALES	CANTIDAD OBSERVACIONES
Gallinas	4
Cuyes	2
Gallos	2
Patos	1
Total	9

Clúster 2

- P403AA_1 Donde Cria
- P403AA_2 Mercado Local
- P403AA_3 Matadero Local
- P403AA_4 Mercado Regional
- P403AA_5 Matadero Regional
- P403AA_6 Matadero Lima
- P403AA_7 Mercado Exterior
- P403AA_8 Agroindustria
- P403AA_9 Mercado Lima
- P403AA_10 Otro

Categoría	Porcentaje
P403AA_2 Mercado Local	100%

ANIMALES	CANTIDAD OBSERVACIONES
Gallinas	3
Gallos	3
Pavos	1
Total	7

2016															
Clúster 6	- P403AA_1 Donde Cria - P403AA_2 Mercado Local - P403AA_3 Matadero Local - P403AA_4 Mercado Regional - P403AA_5 Matadero Regional - P403AA_6 Matadero Lima - P403AA_7 Mercado Exterior - P403AA_8 Agroindustria - P403AA_9 Mercado Lima - P403AA_10 Otro		<table><thead><tr><th>ANIMALES</th><th>CANTIDAD OBSERVACIONES</th></tr></thead><tbody><tr><td>Pollos de engorde</td><td>12</td></tr><tr><td>Gallinas</td><td>11</td></tr><tr><td>Gallos</td><td>7</td></tr><tr><td>Patos</td><td>3</td></tr><tr><td>Total</td><td>33</td></tr></tbody></table>	ANIMALES	CANTIDAD OBSERVACIONES	Pollos de engorde	12	Gallinas	11	Gallos	7	Patos	3	Total	33
	ANIMALES	CANTIDAD OBSERVACIONES													
	Pollos de engorde	12													
	Gallinas	11													
Gallos	7														
Patos	3														
Total	33														
Clúster 1	- P403AA_1 Donde Cria - P403AA_2 Mercado Local - P403AA_3 Matadero Local - P403AA_4 Mercado Regional - P403AA_5 Matadero Regional - P403AA_6 Matadero Lima - P403AA_7 Mercado Exterior - P403AA_8 Agroindustria - P403AA_9 Mercado Lima - P403AA_10 Otro		<table><thead><tr><th>ANIMALES</th><th>CANTIDAD OBSERVACIONES</th></tr></thead><tbody><tr><td>Gallinas</td><td>9</td></tr><tr><td>Pollos de engorde</td><td>8</td></tr><tr><td>Gallos</td><td>4</td></tr><tr><td>Patos</td><td>1</td></tr><tr><td>Total</td><td>22</td></tr></tbody></table>	ANIMALES	CANTIDAD OBSERVACIONES	Gallinas	9	Pollos de engorde	8	Gallos	4	Patos	1	Total	22
	ANIMALES	CANTIDAD OBSERVACIONES													
	Gallinas	9													
	Pollos de engorde	8													
Gallos	4														
Patos	1														
Total	22														
Clúster 2	- P403AA_1 Donde Cria - P403AA_2 Mercado Local - P403AA_3 Matadero Local - P403AA_4 Mercado Regional - P403AA_5 Matadero Regional - P403AA_6 Matadero Lima - P403AA_7 Mercado Exterior - P403AA_8 Agroindustria - P403AA_9 Mercado Lima - P403AA_10 Otro		<table><thead><tr><th>ANIMALES</th><th>CANTIDAD OBSERVACIONES</th></tr></thead><tbody><tr><td>Gallinas</td><td>3</td></tr><tr><td>Gallos</td><td>2</td></tr><tr><td>Total</td><td>5</td></tr></tbody></table>	ANIMALES	CANTIDAD OBSERVACIONES	Gallinas	3	Gallos	2	Total	5				
	ANIMALES	CANTIDAD OBSERVACIONES													
	Gallinas	3													
	Gallos	2													
Total	5														

2017

Clúster 4

- P403AA_1 Donde Cria
- P403AA_2 Mercado Local
- P403AA_3 Matadero Local
- P403AA_4 Mercado Regional
- P403AA_5 Matadero Regional
- P403AA_6 Matadero Lima
- P403AA_7 Mercado Exterior
- P403AA_8 Agroindustria
- P403AA_9 Mercado Lima
- P403AA_10 Otro

Categoría	Porcentaje
P403AA_2 Mercado Local	80%
P403AA_4 Mercado Regional	20%

ANIMALES

CANTIDAD
OBSERVACIONES

Gallinas	10
Gallos	3
Pollos de engorde	1
Porcinos	1
Total	15

Clúster 1

- P403AA_1 Donde Cria
- P403AA_2 Mercado Local
- P403AA_3 Matadero Local
- P403AA_4 Mercado Regional
- P403AA_5 Matadero Regional
- P403AA_6 Matadero Lima
- P403AA_7 Mercado Exterior
- P403AA_8 Agroindustria
- P403AA_9 Mercado Lima
- P403AA_10 Otro

Categoría	Porcentaje
P403AA_1 Donde Cria	100%

ANIMALES

CANTIDAD
OBSERVACIONES

Gallinas	6
Gallos	3
Patos	1
Pollos de engorde	1
Total	11

Clúster 5

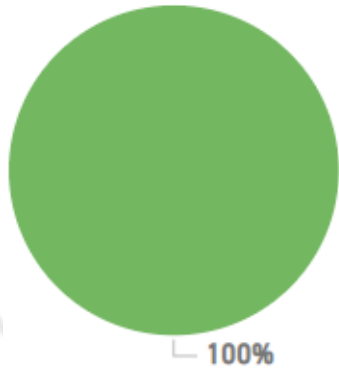
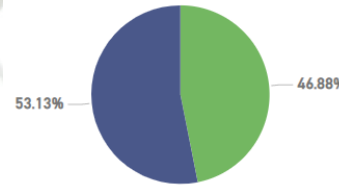
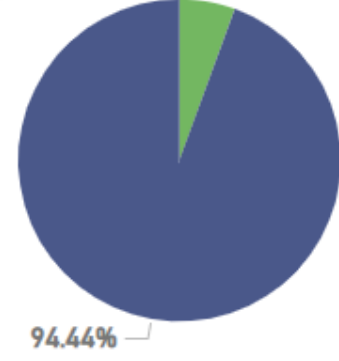
- P403AA_1 Donde Cria
- P403AA_2 Mercado Local
- P403AA_3 Matadero Local
- P403AA_4 Mercado Regional
- P403AA_5 Matadero Regional
- P403AA_6 Matadero Lima
- P403AA_7 Mercado Exterior
- P403AA_8 Agroindustria
- P403AA_9 Mercado Lima
- P403AA_10 Otro

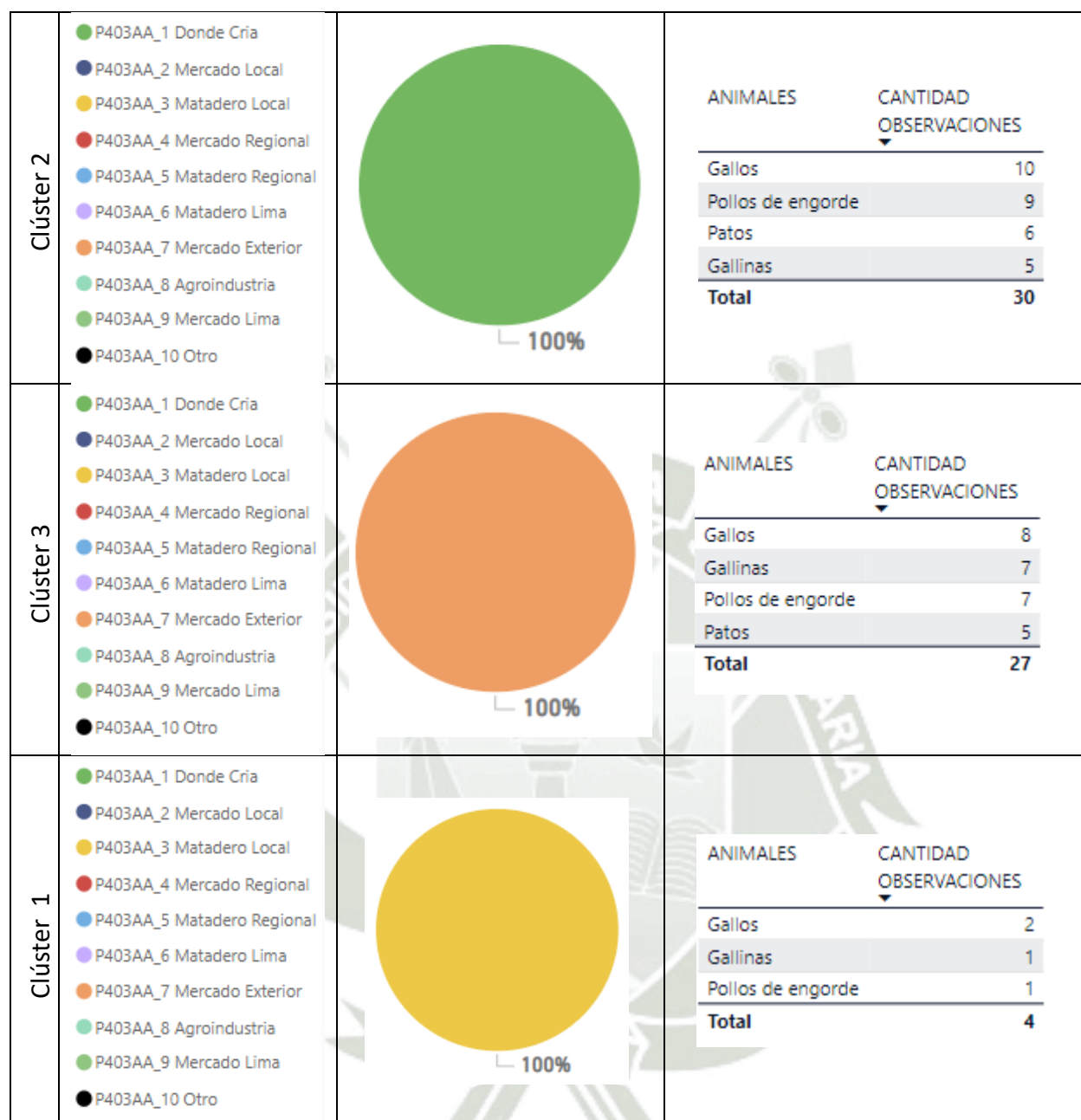
Categoría	Porcentaje
P403AA_1 Donde Cria	50%
P403AA_2 Mercado Local	37.5%
P403AA_4 Mercado Regional	12.5%

ANIMALES

CANTIDAD
OBSERVACIONES

Gallinas	3
Gallos	1
Total	4

2018															
Clúster 1	- P403AA_1 Donde Cria - P403AA_2 Mercado Local - P403AA_3 Matadero Local - P403AA_4 Mercado Regional - P403AA_5 Matadero Regional - P403AA_6 Matadero Lima - P403AA_7 Mercado Exterior - P403AA_8 Agroindustria - P403AA_9 Mercado Lima - P403AA_10 Otro		<table><thead><tr><th>ANIMALES</th><th>CANTIDAD OBSERVACIONES</th></tr></thead><tbody><tr><td>Gallinas</td><td>14</td></tr><tr><td>Gallos</td><td>6</td></tr><tr><td>Patos</td><td>5</td></tr><tr><td>Pollos de engorde</td><td>5</td></tr><tr><td>Total</td><td>30</td></tr></tbody></table>	ANIMALES	CANTIDAD OBSERVACIONES	Gallinas	14	Gallos	6	Patos	5	Pollos de engorde	5	Total	30
	ANIMALES	CANTIDAD OBSERVACIONES													
Gallinas	14														
Gallos	6														
Patos	5														
Pollos de engorde	5														
Total	30														
Clúster 5	- P403AA_1 Donde Cria - P403AA_2 Mercado Local - P403AA_3 Matadero Local - P403AA_4 Mercado Regional - P403AA_5 Matadero Regional - P403AA_6 Matadero Lima - P403AA_7 Mercado Exterior - P403AA_8 Agroindustria - P403AA_9 Mercado Lima - P403AA_10 Otro		<table><thead><tr><th>ANIMALES</th><th>CANTIDAD OBSERVACIONES</th></tr></thead><tbody><tr><td>Gallinas</td><td>7</td></tr><tr><td>Gallos</td><td>5</td></tr><tr><td>Pollos de engorde</td><td>3</td></tr><tr><td>Patos</td><td>2</td></tr><tr><td>Total</td><td>17</td></tr></tbody></table>	ANIMALES	CANTIDAD OBSERVACIONES	Gallinas	7	Gallos	5	Pollos de engorde	3	Patos	2	Total	17
	ANIMALES	CANTIDAD OBSERVACIONES													
Gallinas	7														
Gallos	5														
Pollos de engorde	3														
Patos	2														
Total	17														
2019															
Clúster 4	- P403AA_1 Donde Cria - P403AA_2 Mercado Local - P403AA_3 Matadero Local - P403AA_4 Mercado Regional - P403AA_5 Matadero Regional - P403AA_6 Matadero Lima - P403AA_7 Mercado Exterior - P403AA_8 Agroindustria - P403AA_9 Mercado Lima - P403AA_10 Otro		<table><thead><tr><th>ANIMALES</th><th>CANTIDAD OBSERVACIONES</th></tr></thead><tbody><tr><td>Gallinas</td><td>12</td></tr><tr><td>Gallos</td><td>10</td></tr><tr><td>Pollos de engorde</td><td>10</td></tr><tr><td>Patos</td><td>2</td></tr><tr><td>Total</td><td>34</td></tr></tbody></table>	ANIMALES	CANTIDAD OBSERVACIONES	Gallinas	12	Gallos	10	Pollos de engorde	10	Patos	2	Total	34
	ANIMALES	CANTIDAD OBSERVACIONES													
Gallinas	12														
Gallos	10														
Pollos de engorde	10														
Patos	2														
Total	34														



Fuente: Elaboración Propia.

3.1.4. Módulo 600 Inocuidad agroalimentaria

Los residuos vegetales se pueden aprovechar para muchos fines, pero también generan un impacto ambiental causando infecciones, contaminación, etc. Esta variable nos ayudará a proponer una mejora en el tratamiento de los residuos de los cultivos.

La inocuidad de los alimentos puede definirse como el conjunto de condiciones y medidas necesarias durante la producción, almacenamiento, distribución y preparación de alimentos para asegurar que una vez ingeridos, no representen un riesgo para la salud (MinSalud).

Los alimentos pueden volverse inseguros cuando están contaminados con bacterias, virus, parásitos, químicos o agentes físicos ajenos al producto final. Los síntomas de una intoxicación por alimentos varían de leves a severos, pero comúnmente incluyen náuseas, vómitos, diarrea, calambres estomacales, debilidad, fiebre y escalofríos, en el mejor de los casos. En casos severos, una enfermedad transmitida por los alimentos puede llevar a la hospitalización o incluso a la muerte (Basic Farm, 2020).

A. Variable P601 ¿Qué hace con los residuos de cultivos?

Con esta variable podremos determinar los organismos que aportan información para el cuidado de alimentos y la contaminación de alimentos.

El siguiente gráfico describe cómo se caracteriza cada clúster, cuantos clústeres existen por año y la cantidad de productores que existen por cada clúster, con este gráfico podremos escoger un distrito, ver el clúster al que pertenece y poder determinar sus principales características

Descripción de los clústeres por años y cantidad de productores que tiene cada clúster.

En el Perú del año 2015 al 2019 se tienen 135818 en la variable P601, la mayoría de productores deja sus residuos en el campo, después lo usa de alimento para sus animales, lo usa para abono, lo quema, lo bota o lo usa para otra actividad.

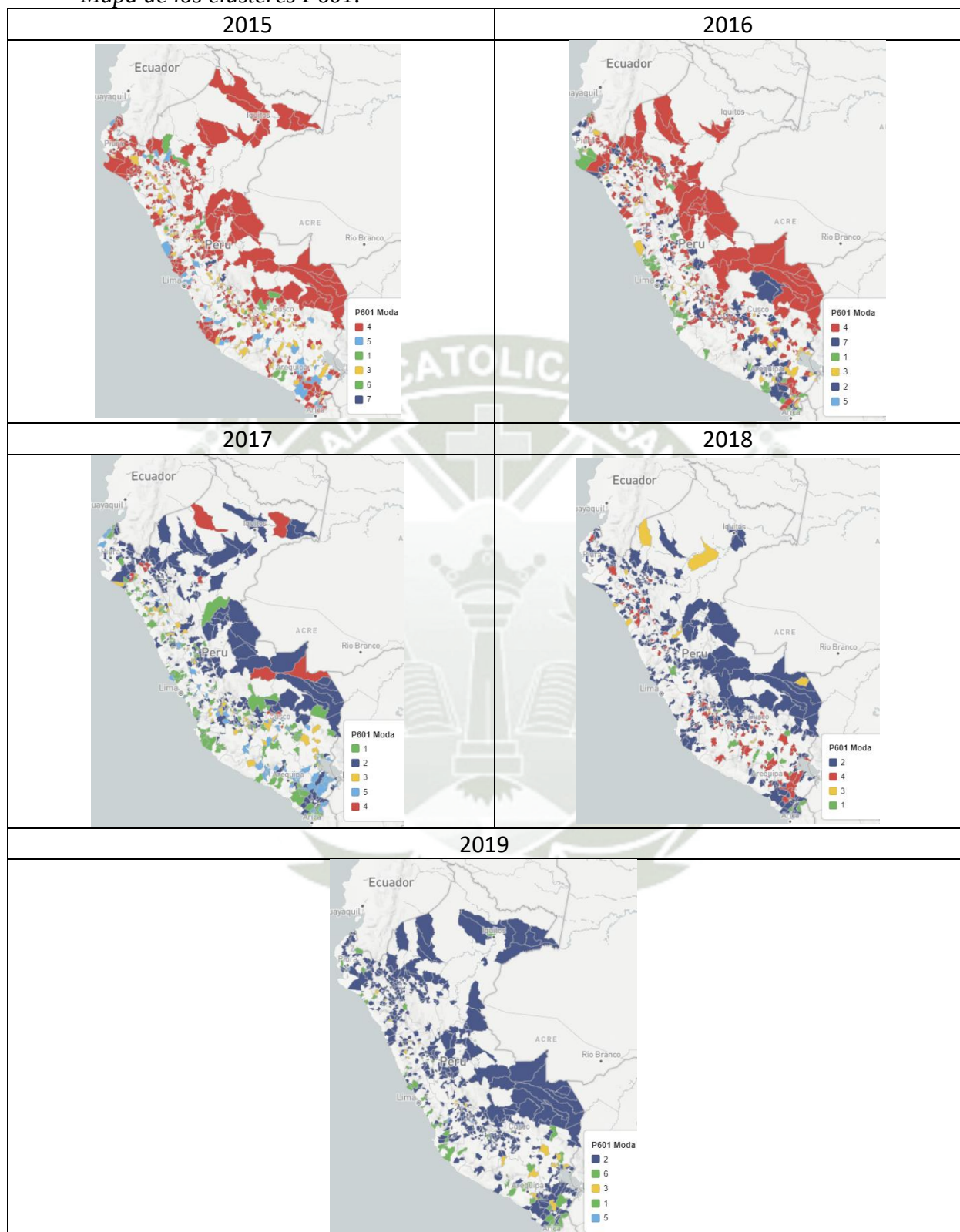
Figura 45.
Características de los clústeres en P601

AÑO	P601 CLUSTER	P601_1 Quema	P601_2 Alimento animales	P601_3 Bota	P601_4 Abono	P601_5 Deja en el campo	P601_6 Otro	P601_7 No es productor agrícola	Cantid Observaciones
2015	4	3,345	1,880	75	147	11,265	0	0	13,645
2015	3	657	4,834	9	1,745	953	11	0	4,834
2015	5	666	0	790	2,299	0	0	1,197	4,286
2015	6	282	0	2	2,608	2,585	50	0	2,608
2015	1	40	182	745	179	535	0	0	745
2015	7	119	37	102	4	158	429	0	521
2015	2	28	40	0	31	42	0	0	210
2016	4	0	2,504	3	38	12,028	338	0	12,028
2016	7	1,159	1,853	57	6,152	3,455	0	0	6,152
2016	1	3,670	0	198	0	1,573	74	0	3,670
2016	3	646	2,940	886	47	968	0	0	3,640
2016	2	0	0	457	0	0	11	1,020	1,477
2016	5	19	79	0	39	11	424	0	424
2016	6	0	0	109	109	43	1	0	109
2017	2	1,512	106	7	22	11,427	369	0	11,491
2017	1	3,018	600	0	5,545	2,703	118	0	7,435
2017	3	753	3,738	92	998	3,373	0	0	3,738
2017	5	0	2,038	145	95	0	405	952	3,482
2017	4	384	0	1,298	119	652	0	0	1,298
2018	2	3,939	0	0	4,377	14,580	365	0	18,375
2018	4	591	6,389	0	1,574	3,114	80	0	6,389
2018	1	1	0	5	0	236	240	931	1,171
2018	3	249	103	1,153	123	535	0	0	1,153
2019	2	2,224	3,302	0	4,104	18,644	22	0	18,644
2019	6	2,252	189	224	1,639	0	0	763	4,206
2019	3	55	2,360	0	629	0	28	0	2,360
2019	1	44	0	438	40	5	264	0	690
2019	5	79	114	679	93	620	0	0	679
2019	4	12	30	0	55	356	358	0	358
Total		25,744	33,318	7,474	32,811	89,861	3,587	4,863	135,818

Fuente: Elaboración Propia.

Para la variable P601 se observa que existen áreas geográficamente cercanas que comparten las mismas características

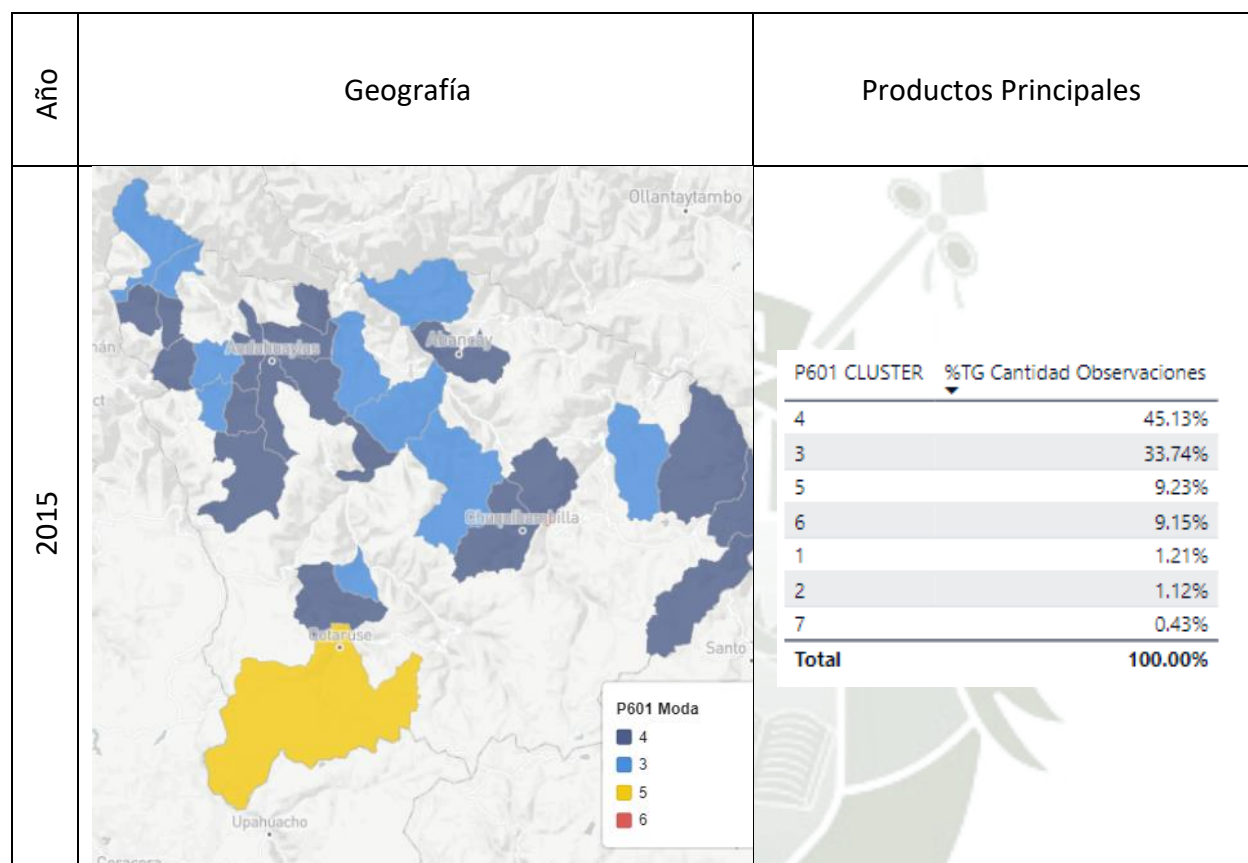
Figura 46.
Mapa de los clústeres P601.

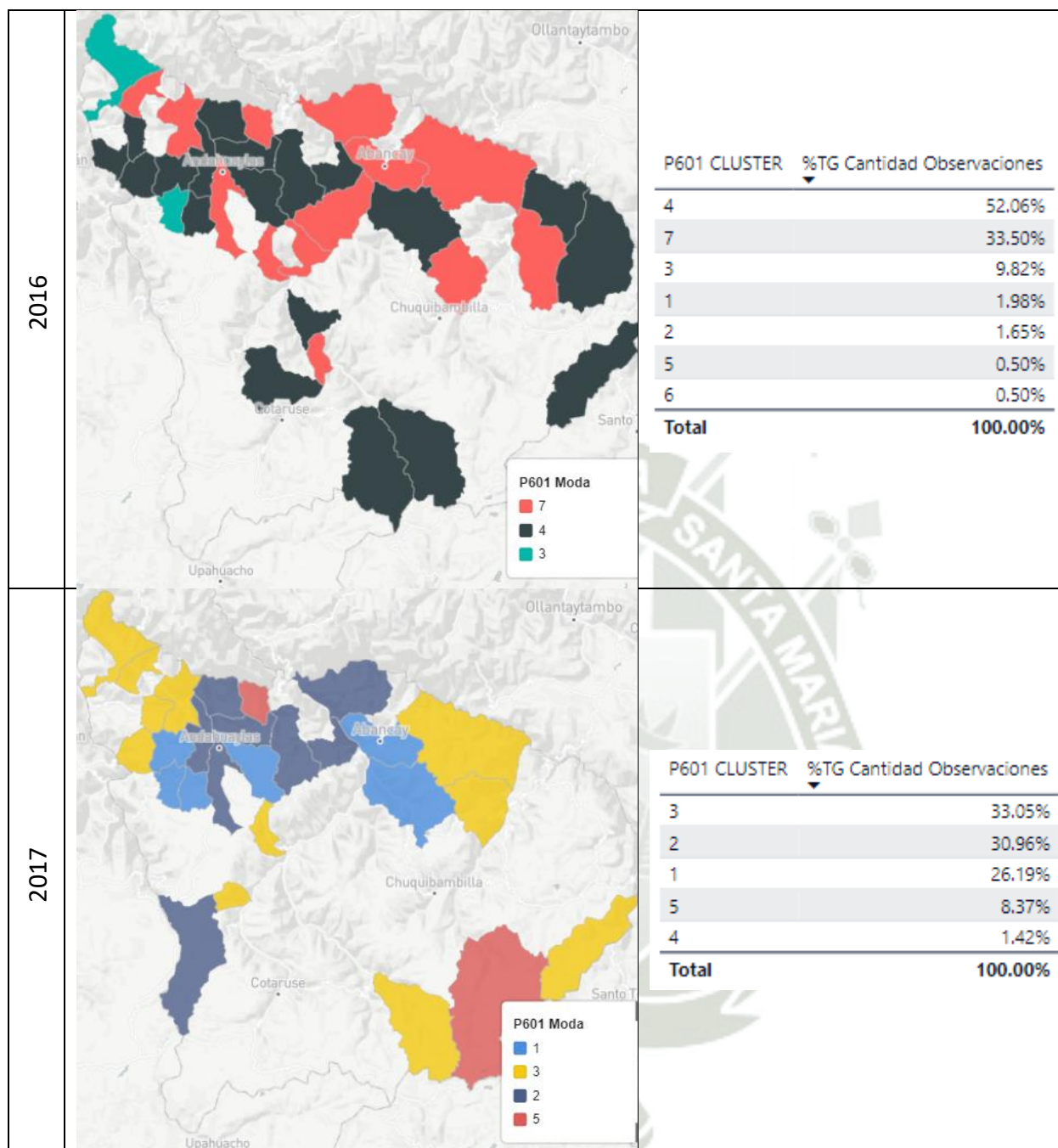


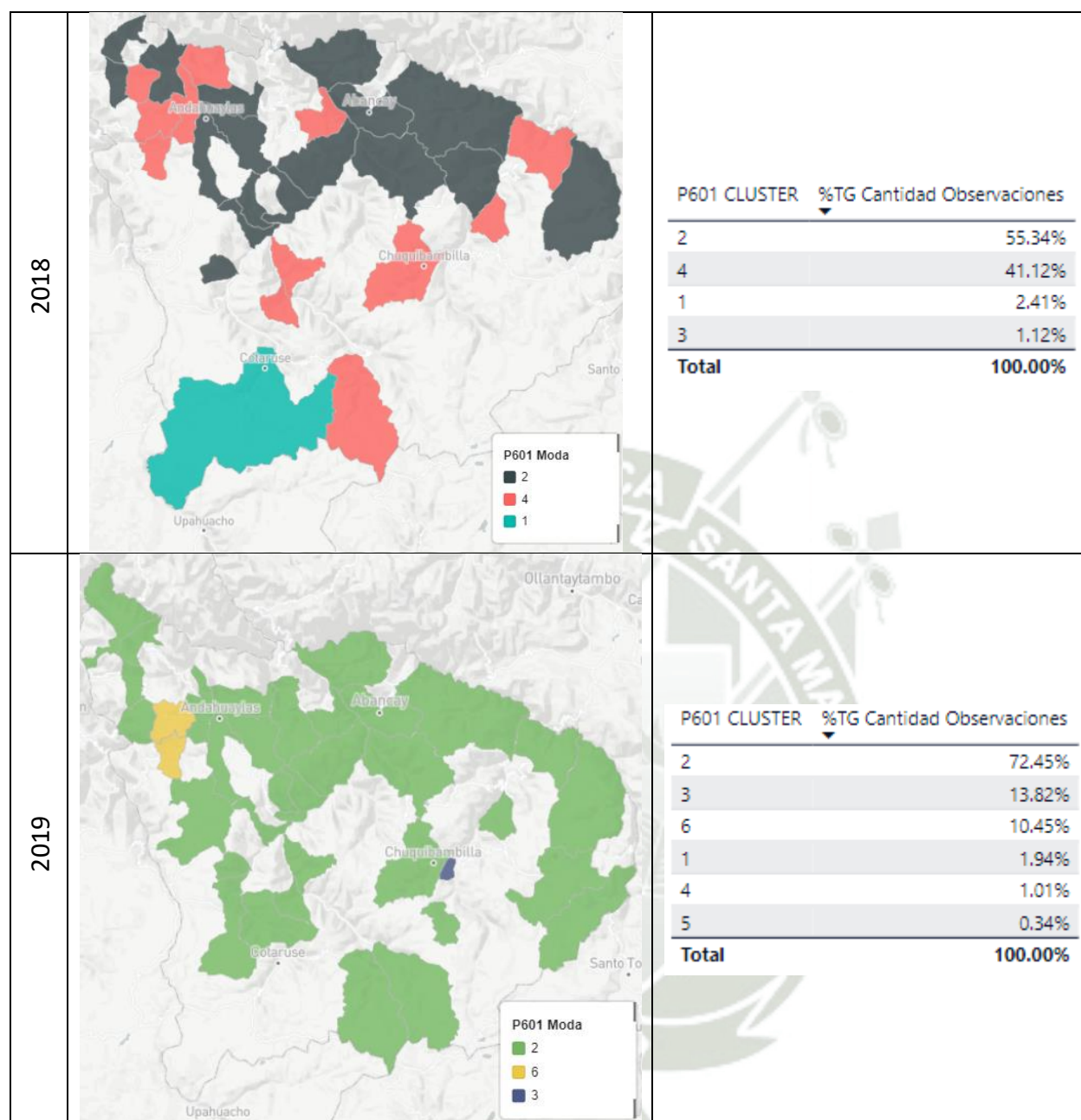
Fuente: Elaboración Propia.

Análisis departamento de Apurímac, en dicho departamento los productores dejan sus residuos en el campo lo usan como alimento de animales, como abono, lo queman u otro.

Figura 47.
Características del departamento de Apurímac.



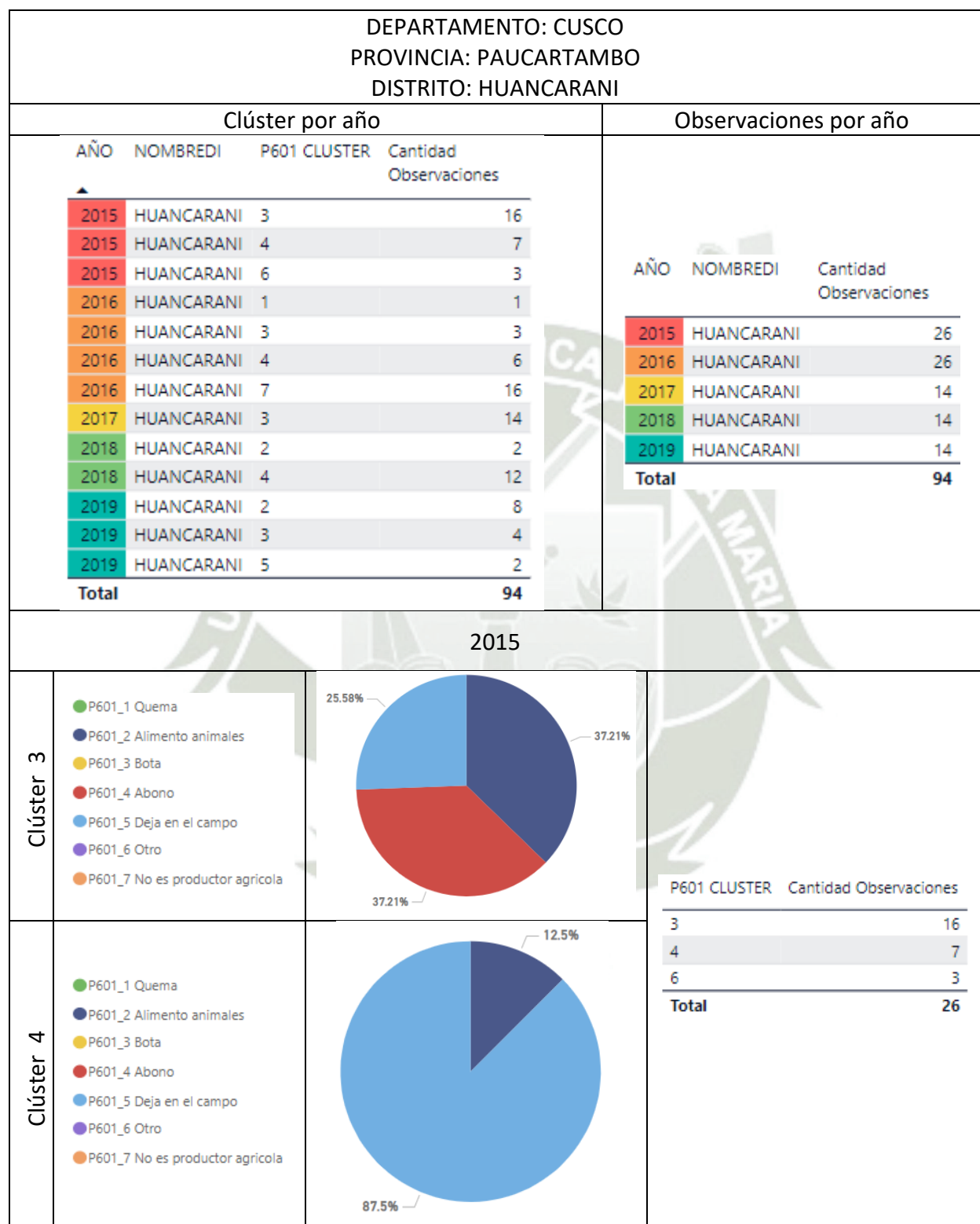


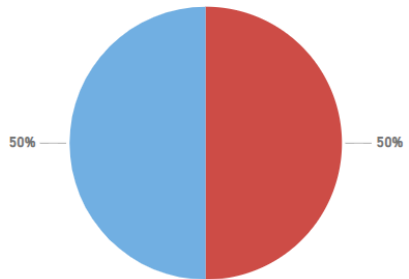
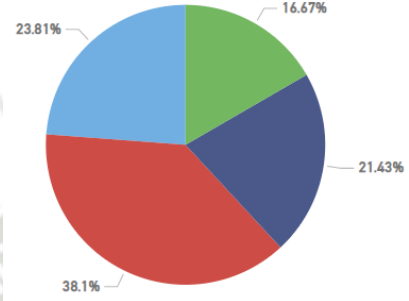
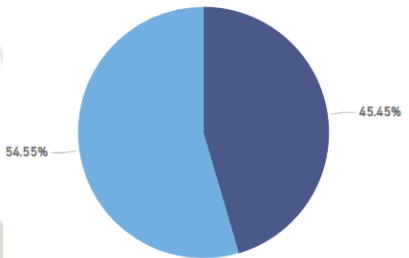
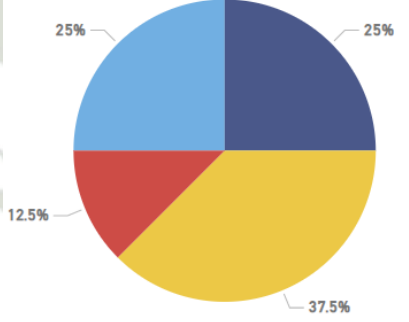
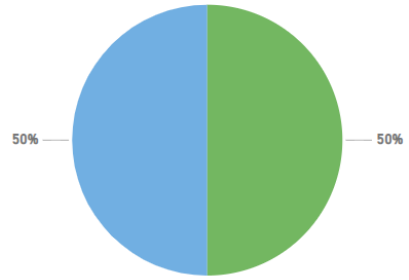


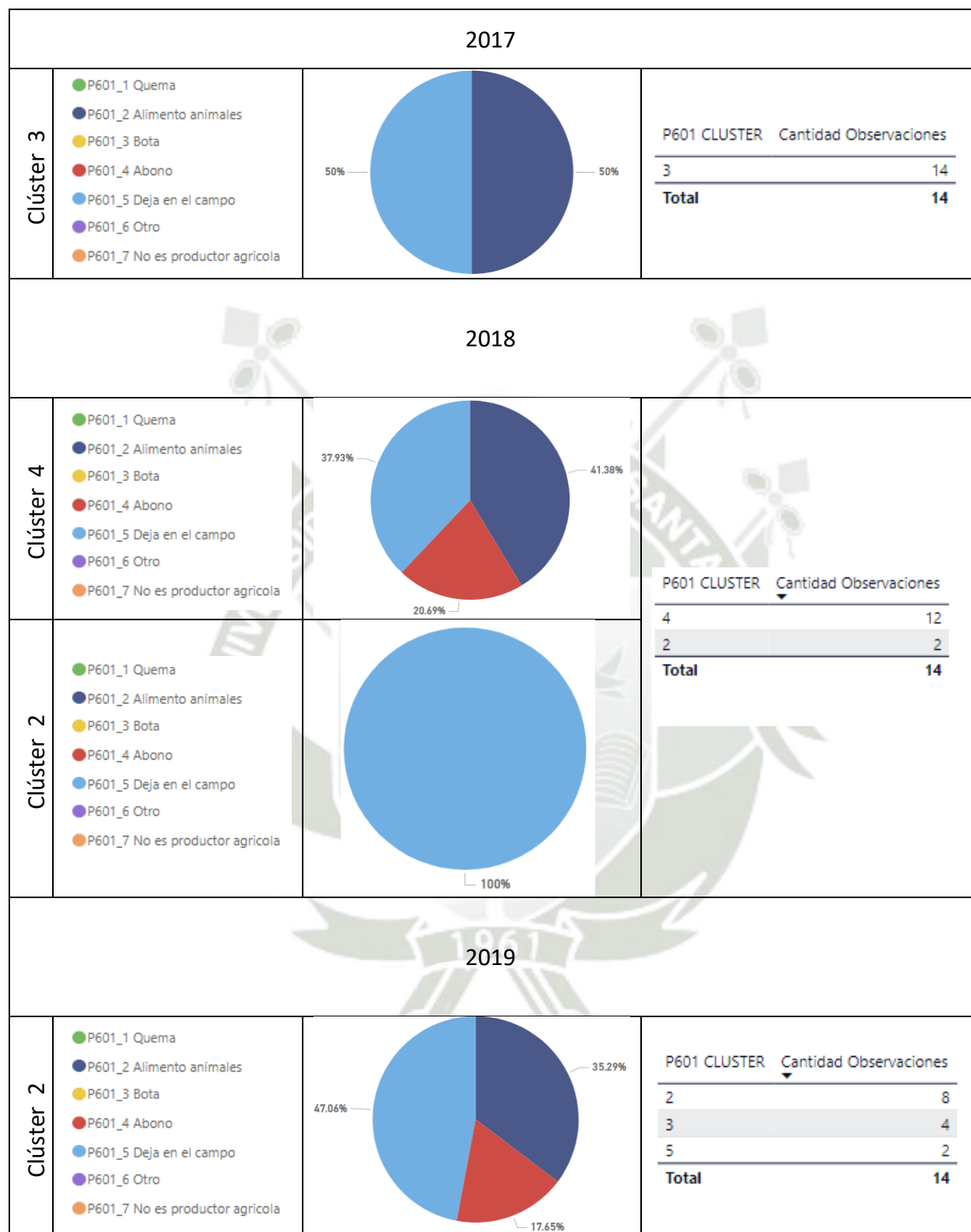
Fuente: Elaboración Propia.

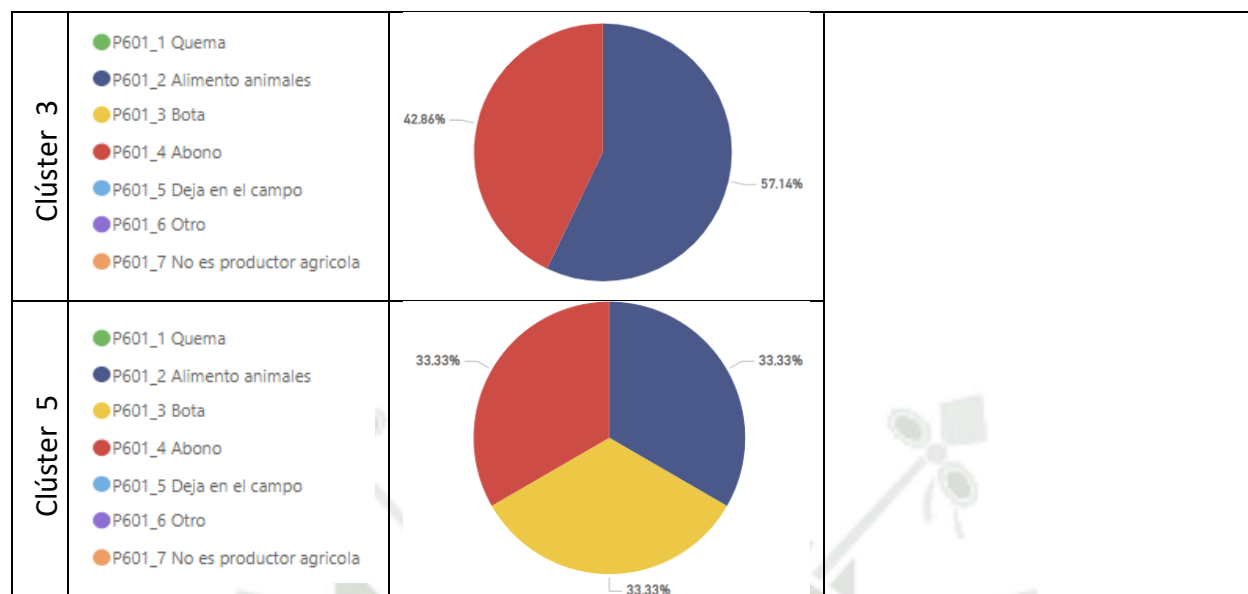
Los productores del distrito de Huancaraní, dejan sus residuos en el campo, lo usan como abono y como alimento para sus animales, y en algunas ocasiones lo botan o lo queman.

Figura 48.
Características del distrito de Huancarani.



Clúster 6	- P601_1 Quema - P601_2 Alimento animales - P601_3 Bota - P601_4 Abono - P601_5 Deja en el campo - P601_6 Otro - P601_7 No es productor agrícola														
2016															
Clúster 7	- P601_1 Quema - P601_2 Alimento animales - P601_3 Bota - P601_4 Abono - P601_5 Deja en el campo - P601_6 Otro - P601_7 No es productor agrícola		 <table><thead><tr><th>P601 CLUSTER</th><th>Cantidad Observaciones</th></tr></thead><tbody><tr><td>7</td><td>16</td></tr><tr><td>4</td><td>6</td></tr><tr><td>3</td><td>3</td></tr><tr><td>1</td><td>1</td></tr><tr><td>Total</td><td>26</td></tr></tbody></table>	P601 CLUSTER	Cantidad Observaciones	7	16	4	6	3	3	1	1	Total	26
P601 CLUSTER	Cantidad Observaciones														
7	16														
4	6														
3	3														
1	1														
Total	26														
Clúster 4	- P601_1 Quema - P601_2 Alimento animales - P601_3 Bota - P601_4 Abono - P601_5 Deja en el campo - P601_6 Otro - P601_7 No es productor agrícola														
Clúster 3	- P601_1 Quema - P601_2 Alimento animales - P601_3 Bota - P601_4 Abono - P601_5 Deja en el campo - P601_6 Otro - P601_7 No es productor agrícola														
Clúster 1	- P601_1 Quema - P601_2 Alimento animales - P601_3 Bota - P601_4 Abono - P601_5 Deja en el campo - P601_6 Otro - P601_7 No es productor agrícola														





Fuente: Elaboración Propia.

B. Variable P602 ¿Qué hace con los residuos de sus animales de crianza?

Al igual que la variable P601, esta variable nos ayudará a generar propuestas de mejora para el tratamiento de residuos de animales de crianza.

Los productores utilizan los residuos provocados por sus animales generalmente para abono, lo dejan en el campo, lo botan, usan de combustible, lo entierra, quema u otro.

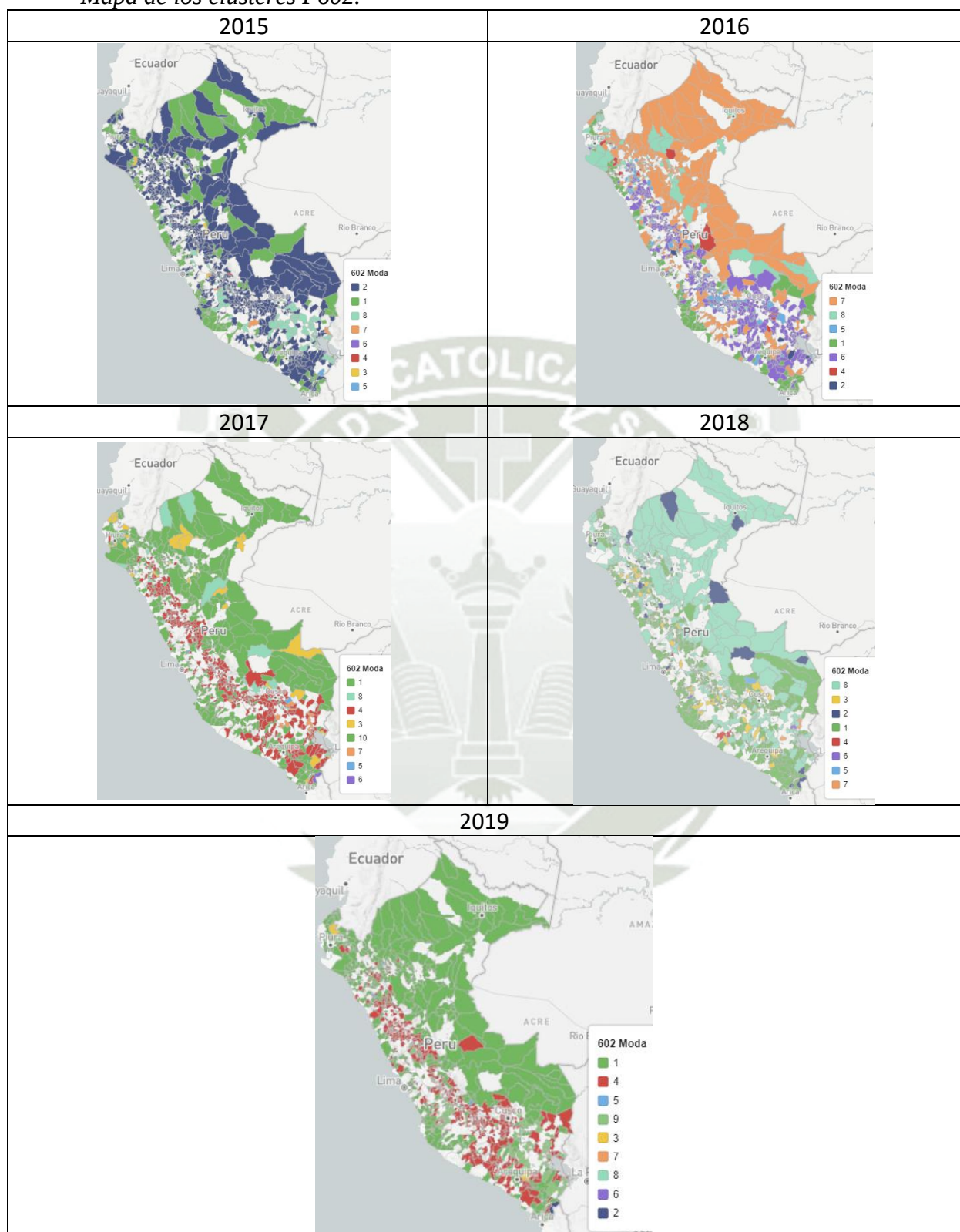
Figura 49.
Características de los clústeres en P602.

AÑO	P602 CLUSTER	P602_1 Quema	P602_2 Entierra	P602_3 Bota	P602_4 Usa de Abono	P602_5 Usa de combustible	P602_6 Deja en el campo	P602_7 Otro	P602_8 Productor apícola	P602_9 No es productor	Cantidad Observaciones
2015	2	0	328	5,907	10,402	0	0	0	0	309	15,371
2015	1	11	10	2	3	13	371	24	6,226	1,696	8,316
2015	8	10	734	395	1,251	1,362	0	0	0	0	2,014
2015	3	450	112	168	0	0	19	0	0	3	450
2015	7	27	10	0	4	309	6	0	0	0	309
2015	6	195	11	8	195	33	0	0	0	0	195
2015	5	0	24	71	19	24	107	0	0	0	107
2015	4	53	0	52	87	0	36	0	0	0	87
2016	6	110	0	387	8,665	1,227	305	0	0	0	8,665
2016	1	127	1	41	13	9	0	201	29	6,100	6,450
2016	7	0	0	900	0	230	6,142	20	0	0	6,142
2016	5	0	370	252	2,691	96	2,430	3	0	0	2,691
2016	8	119	10	2,538	0	0	59	1	0	0	2,586
2016	4	0	703	206	0	152	181	6	0	0	820
2016	2	108	84	0	27	80	37	0	0	0	117
2016	3	8	12	29	4	26	0	0	0	0	29
2017	1	0	478	0	0	339	5,798	1	27	5,943	12,310
2017	4	0	0	0	10,134	722	2,666	164	0	0	10,270
2017	3	0	110	2,351	38	79	11	4	0	0	2,351
2017	8	165	0	1,421	614	1	834	0	0	0	1,421
2017	7	18	1	21	410	410	397	0	0	0	410
2017	5	301	0	0	107	8	94	0	0	0	301
2017	10	0	186	16	186	37	75	0	0	0	186
2017	9	99	99	64	19	0	22	0	0	0	99
2017	6	5	0	27	0	0	38	62	0	0	62
2017	2	0	34	34	0	2	34	1	0	0	34
2018	1	0	352	79	6,513	0	0	6	22	6,062	12,890
2018	8	0	6	678	1,167	1,310	7,395	0	0	0	8,237
2018	3	0	0	4	2,509	21	2,524	201	0	0	2,675
2018	2	193	0	2,109	0	0	0	32	0	0	2,245
2018	6	12	9	396	396	54	0	0	0	0	396
2018	4	159	261	68	107	2	255	1	0	0	318
2018	7	16	14	38	12	200	3	1	0	0	200
2018	5	127	0	0	127	27	38	0	0	0	127
2019	1	168	369	2,015	0	0	5,673	1	16	6,260	14,436
2019	4	0	108	544	9,500	0	2,763	0	0	0	9,500
Total		2,715	4,758	21,805	56,721	8,546	39,943	931	6,320	26,373	135,818

Fuente: Elaboración Propia.

Para la variable P602 podemos ver departamentos cercanos geográficamente, y podemos ver que hay clústeres que predominan en regiones específicas como la selva y la sierra.

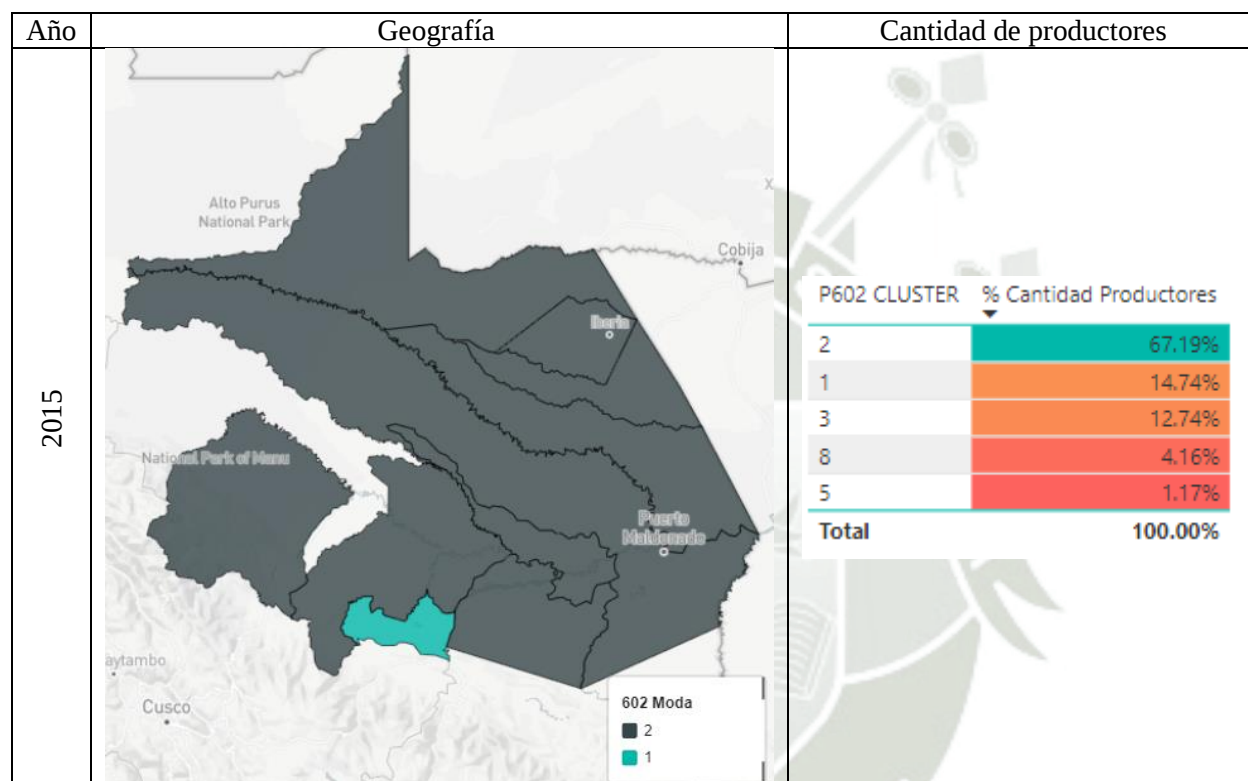
Figura 50.
Mapa de los clústeres P602.

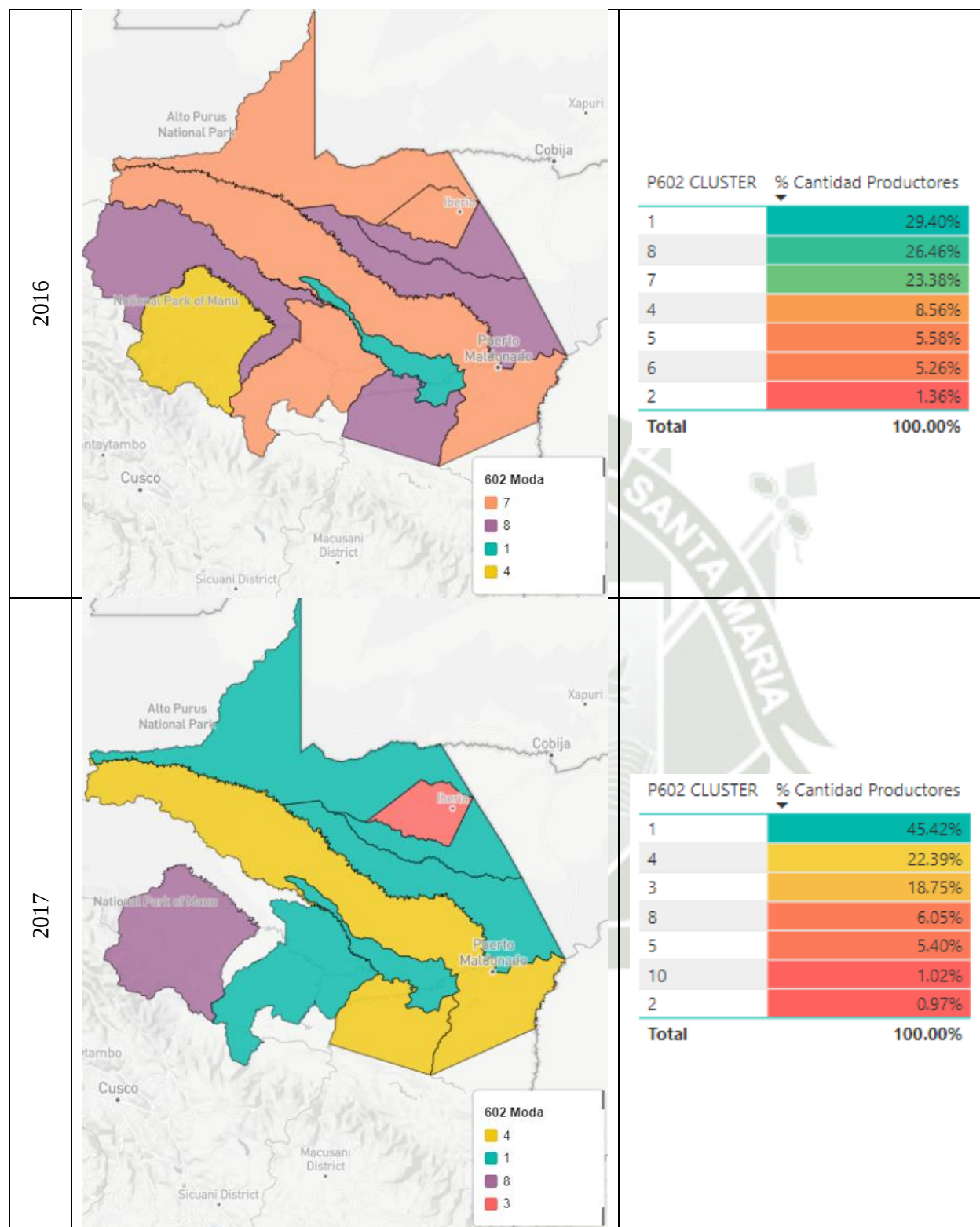


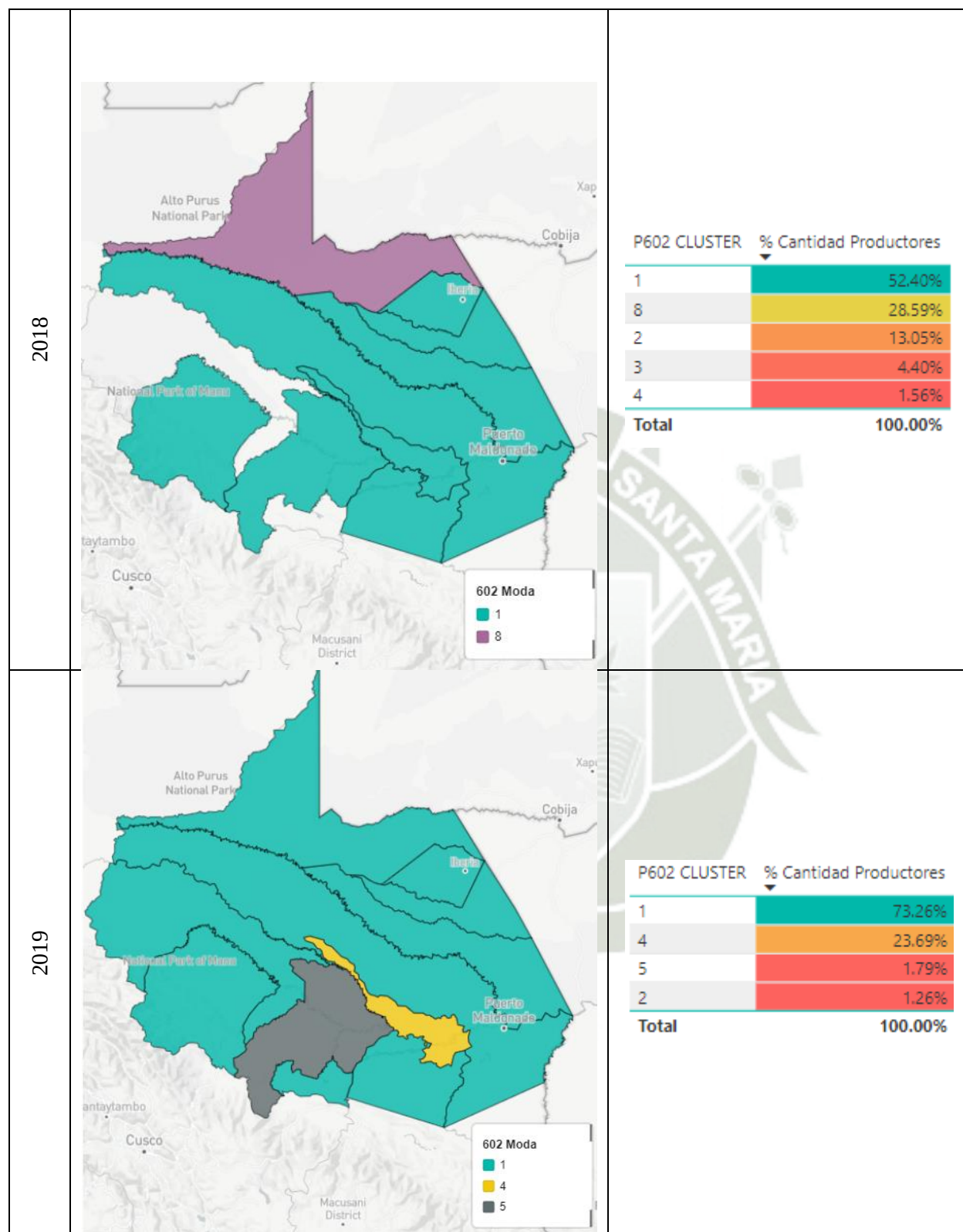
Fuente: Elaboración Propia.

Análisis Madre de Dios, en este departamento los productores realizaron todas las actividades consideradas para esta pregunta, y se clasificaron como se muestra a continuación, en el mapa de Madre de Dios, en este departamento lo productores lo dejan en el campo generalmente, lo botan, lo usan de abono, lo entierra, lo quema u otro.

Figura 51.
Características del departamento de Madre de Dios.







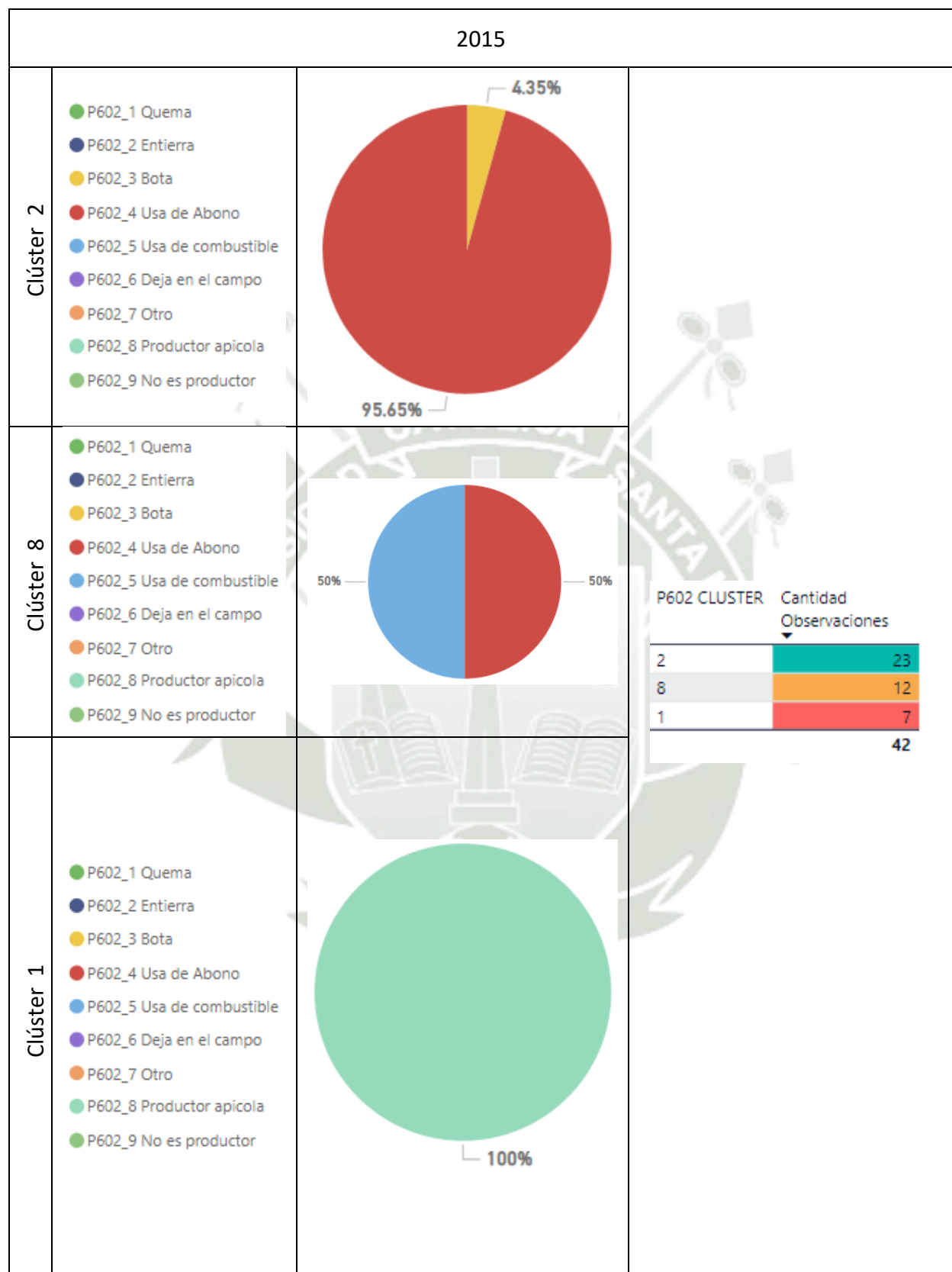
Fuente: Elaboración Propia.

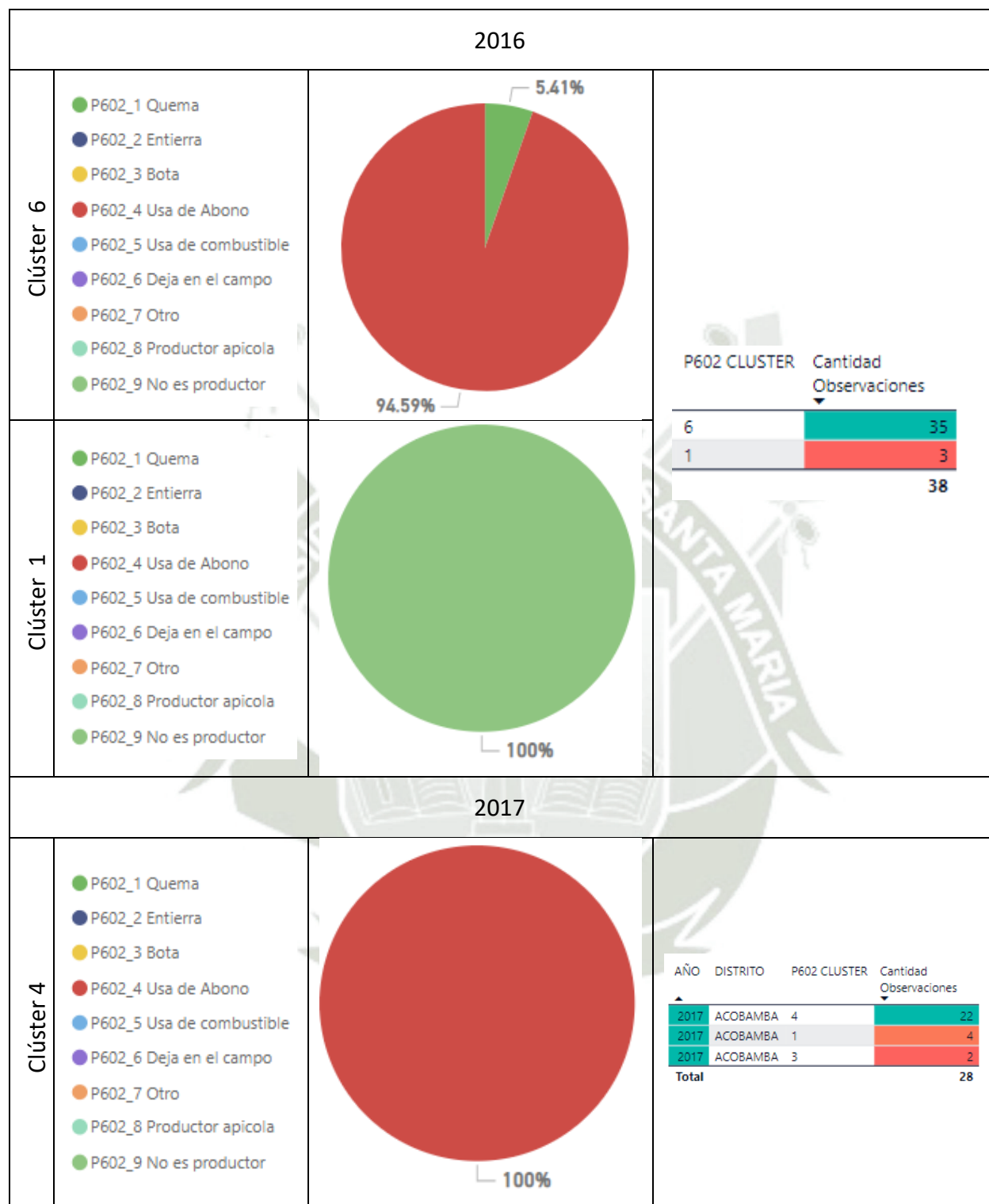
En la siguiente imagen se describe el comportamiento del distrito de Acobamba del año 2015-2019, en este periodo los productores principalmente lo usan como abono, lo dejan en el campo, usan de combustible o son productores apícolas.

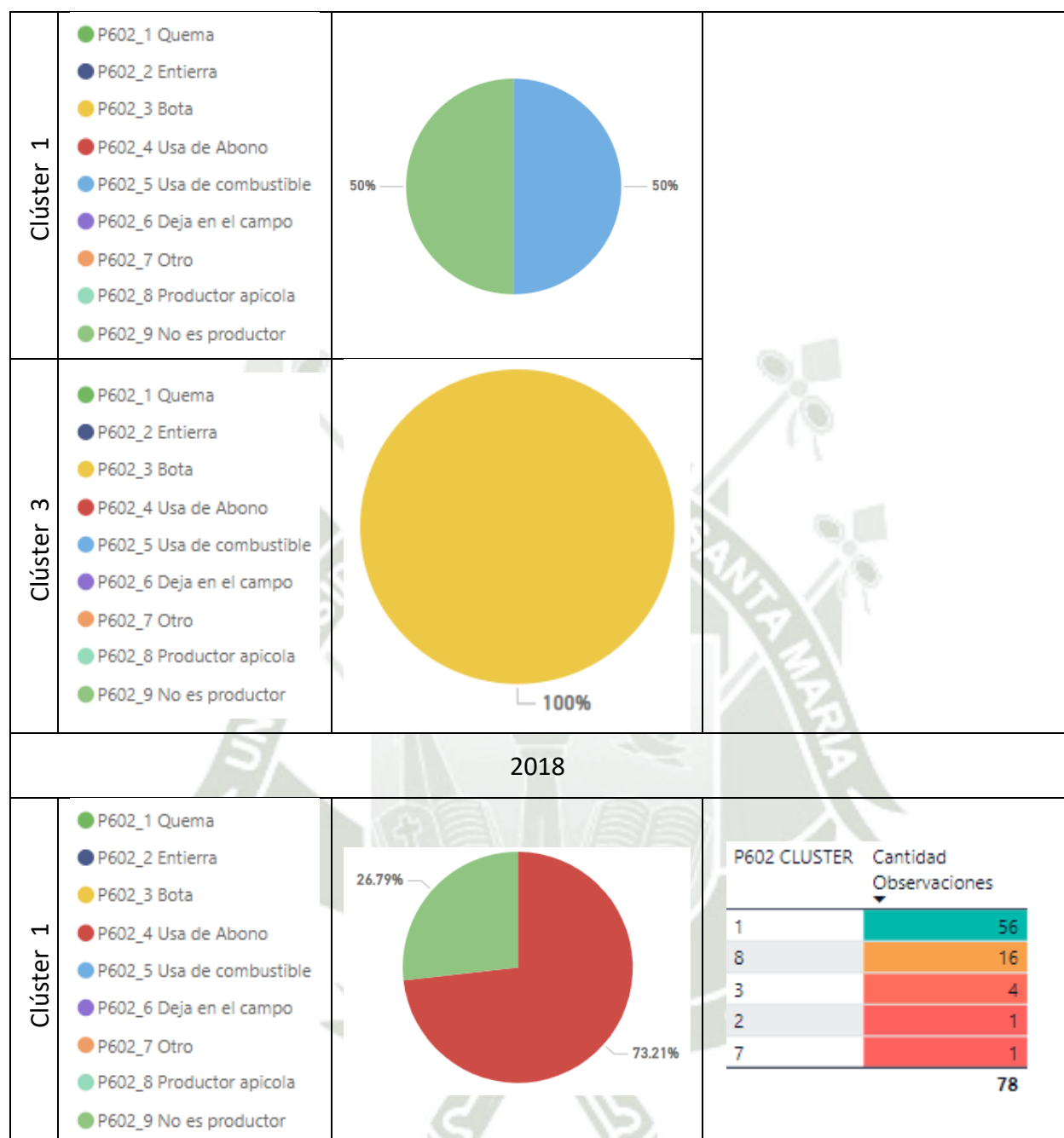
Figura 52.

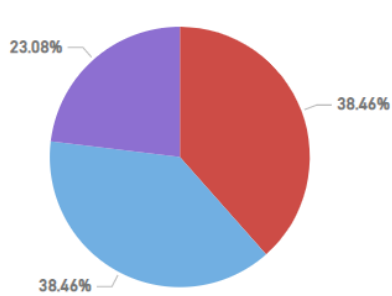
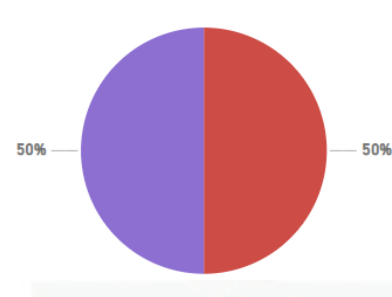
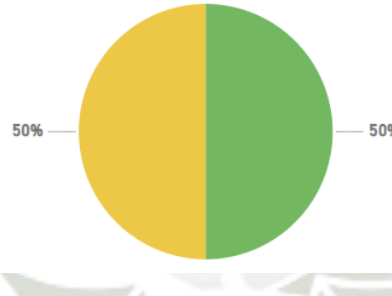
Características del distrito de Acobamba.

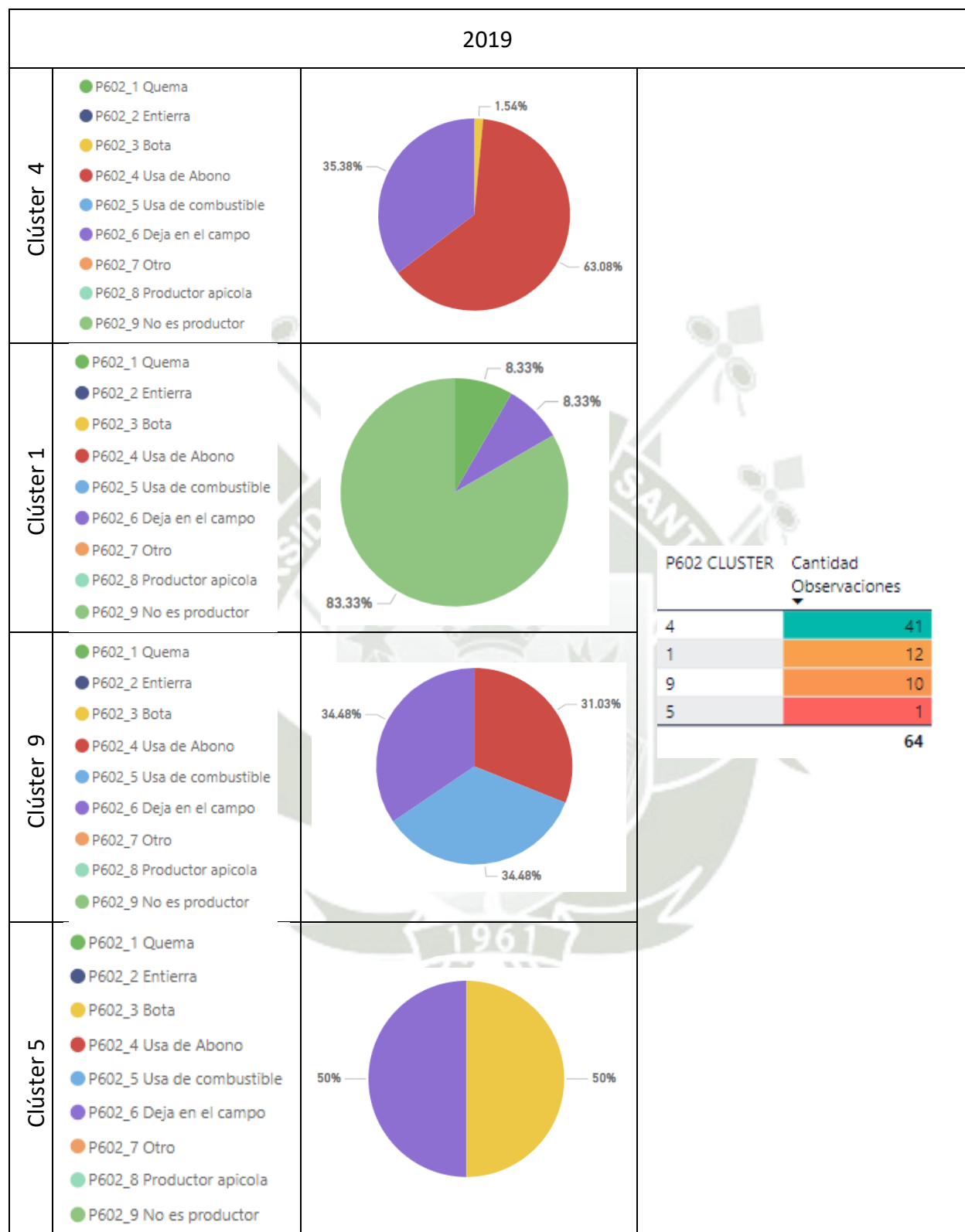
DEPARTAMENTO: HUANCAMELICA						
PROVINCIA: ACOBAMBA						
DISTRITO: ACOBAMBA						
Clúster por año				Observaciones por año		
AÑO	DISTRITO	P602 CLUSTER	Cantidad Observaciones	AÑO	DISTRITO	Cantidad Observaciones
2015	ACOBAMBA	2	23	2015	ACOBAMBA	43
2015	ACOBAMBA	8	12	2016	ACOBAMBA	38
2015	ACOBAMBA	1	7	2017	ACOBAMBA	28
2016	ACOBAMBA	6	35	2018	ACOBAMBA	79
2016	ACOBAMBA	1	3	2019	ACOBAMBA	67
2017	ACOBAMBA	4	22	Total		255
2017	ACOBAMBA	1	4			
2017	ACOBAMBA	3	2			
2018	ACOBAMBA	1	56			
2018	ACOBAMBA	8	16			
2018	ACOBAMBA	3	4			
2018	ACOBAMBA	2	1			
2018	ACOBAMBA	7	1			
2019	ACOBAMBA	4	41			
2019	ACOBAMBA	1	12			
2019	ACOBAMBA	9	10			
2019	ACOBAMBA	5	1			
Total			250			







Clúster 8	- P602_1 Quema - P602_2 Entierra - P602_3 Bota - P602_4 Usa de Abono - P602_5 Usa de combustible - P602_6 Deja en el campo - P602_7 Otro - P602_8 Productor apicola - P602_9 No es productor	 <table><tr><th>Categoría</th><th>Porcentaje</th></tr><tr><td>P602_4 Usa de Abono</td><td>38.46%</td></tr><tr><td>P602_5 Usa de combustible</td><td>38.46%</td></tr><tr><td>P602_6 Deja en el campo</td><td>23.08%</td></tr></table>	Categoría	Porcentaje	P602_4 Usa de Abono	38.46%	P602_5 Usa de combustible	38.46%	P602_6 Deja en el campo	23.08%
Categoría	Porcentaje									
P602_4 Usa de Abono	38.46%									
P602_5 Usa de combustible	38.46%									
P602_6 Deja en el campo	23.08%									
Clúster 3	- P602_1 Quema - P602_2 Entierra - P602_3 Bota - P602_4 Usa de Abono - P602_5 Usa de combustible - P602_6 Deja en el campo - P602_7 Otro - P602_8 Productor apicola - P602_9 No es productor	 <table><tr><th>Categoría</th><th>Porcentaje</th></tr><tr><td>P602_4 Usa de Abono</td><td>50%</td></tr><tr><td>P602_6 Deja en el campo</td><td>50%</td></tr></table>	Categoría	Porcentaje	P602_4 Usa de Abono	50%	P602_6 Deja en el campo	50%		
Categoría	Porcentaje									
P602_4 Usa de Abono	50%									
P602_6 Deja en el campo	50%									
Clúster 2	- P602_1 Quema - P602_2 Entierra - P602_3 Bota - P602_4 Usa de Abono - P602_5 Usa de combustible - P602_6 Deja en el campo - P602_7 Otro - P602_8 Productor apicola - P602_9 No es productor	 <table><tr><th>Categoría</th><th>Porcentaje</th></tr><tr><td>P602_3 Bota</td><td>50%</td></tr><tr><td>P602_1 Quema</td><td>50%</td></tr></table>	Categoría	Porcentaje	P602_3 Bota	50%	P602_1 Quema	50%		
Categoría	Porcentaje									
P602_3 Bota	50%									
P602_1 Quema	50%									
Clúster 7	- P602_1 Quema - P602_2 Entierra - P602_3 Bota - P602_4 Usa de Abono - P602_5 Usa de combustible - P602_6 Deja en el campo - P602_7 Otro - P602_8 Productor apicola - P602_9 No es productor	 <table><tr><th>Categoría</th><th>Porcentaje</th></tr><tr><td>P602_5 Usa de combustible</td><td>100%</td></tr></table>	Categoría	Porcentaje	P602_5 Usa de combustible	100%				
Categoría	Porcentaje									
P602_5 Usa de combustible	100%									



Fuente: Elaboración Propia.

C. Variable P605: ¿Qué entidad del estado le informo?

Con esta variable podremos determinar los organismos que aportan información para el cuidado de alimentos y la contaminación de alimentos.

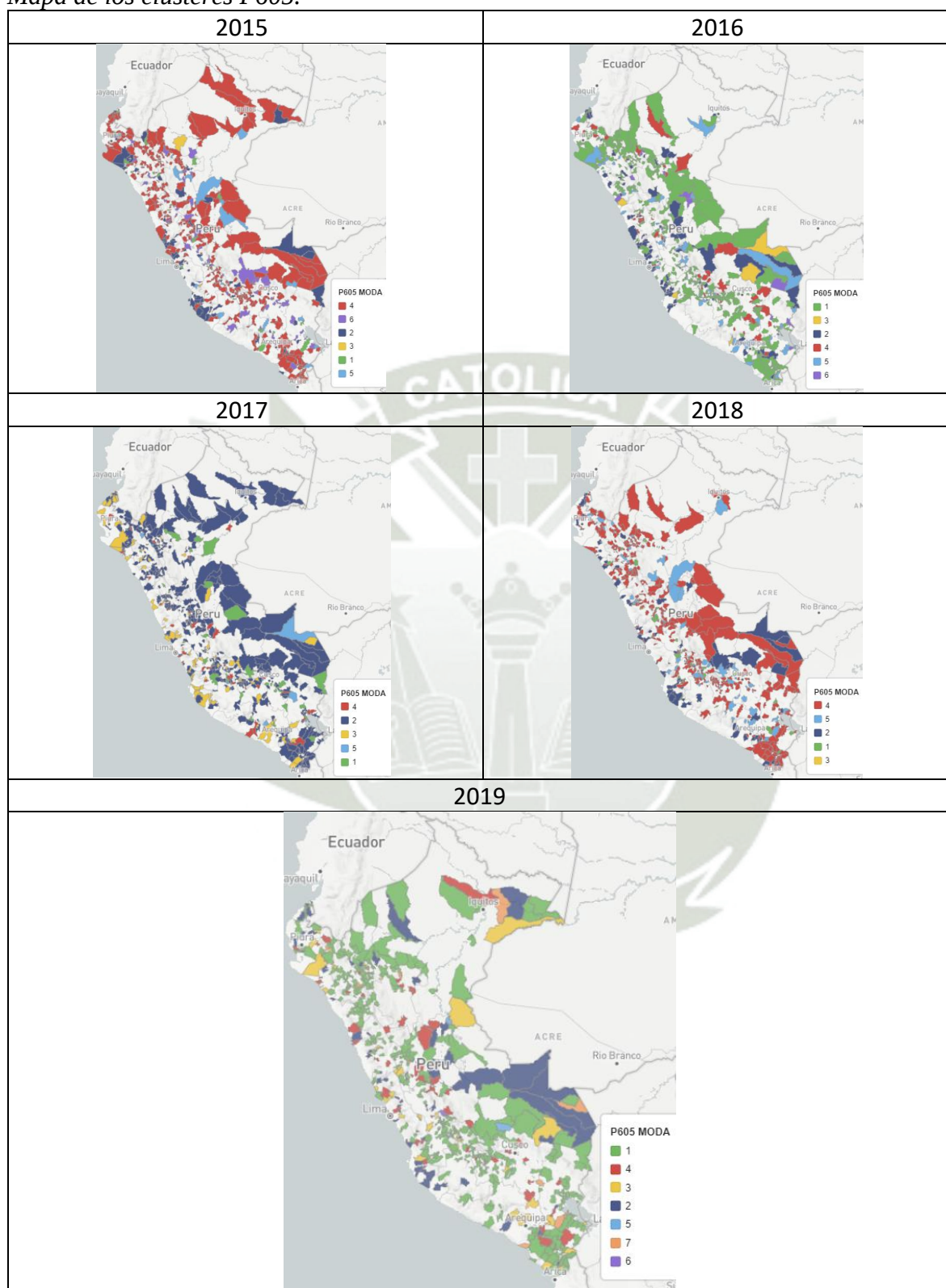
Las entidades que más informaron a los productores acerca de prácticas de inocuidad, de 2015 a 2019 fueron el ministerio de salud, SENASA, gobierno local o municipal, MIDAGRI, otro o el gobierno regional.

Figura 53.
Características de los clústeres en P605.

AÑO	P605 CLUSTER	P605_1 SENASA	P605_2 Ministerio de salud	P605_3 MIDAGRI	P605_4 Gobierno regional	P605_5 Gobierno Local/Municipal	P605_6 Otro	Cantidad Observaciones
2015	4	0	1,366	0	0	0	8	1,366
2015	2	436	62	25	3	0	0	436
2015	1	2	0	104	0	1	69	170
2015	6	0	16	19	0	130	0	145
2015	3	9	44	0	6	52	0	52
2015	5	0	9	3	49	0	0	49
2016	1	0	1,807	33	0	0	18	1,807
2016	2	436	0	0	0	0	59	494
2016	4	3	64	0	3	195	1	195
2016	5	22	15	182	0	8	0	182
2016	3	90	90	0	0	4	0	90
2016	6	5	9	3	42	0	0	42
2017	2	0	1,501	0	33	0	72	1,602
2017	3	411	42	26	0	1	4	411
2017	1	15	0	4	1	174	0	174
2017	4	0	17	164	2	0	2	164
2017	5	2	50	0	0	28	26	51
2018	4	0	1,367	0	30	0	12	1,394
2018	2	296	35	18	2	3	0	296
2018	5	3	0	0	0	106	84	189
2018	1	1	9	136	2	6	3	136
2018	3	0	16	0	4	18	0	18
2019	1	0	1,657	2	0	0	125	1,750
2019	2	235	38	11	0	0	0	235
2019	4	0	32	5	0	177	0	177
2019	3	0	22	125	1	0	0	125
2019	7	3	7	0	37	0	0	37
2019	5	7	0	0	3	9	0	9
2019	6	2	4	4	4	2	0	4
Total		1,978	8,279	864	222	914	483	11,800

Fuente: Elaboración Propia.

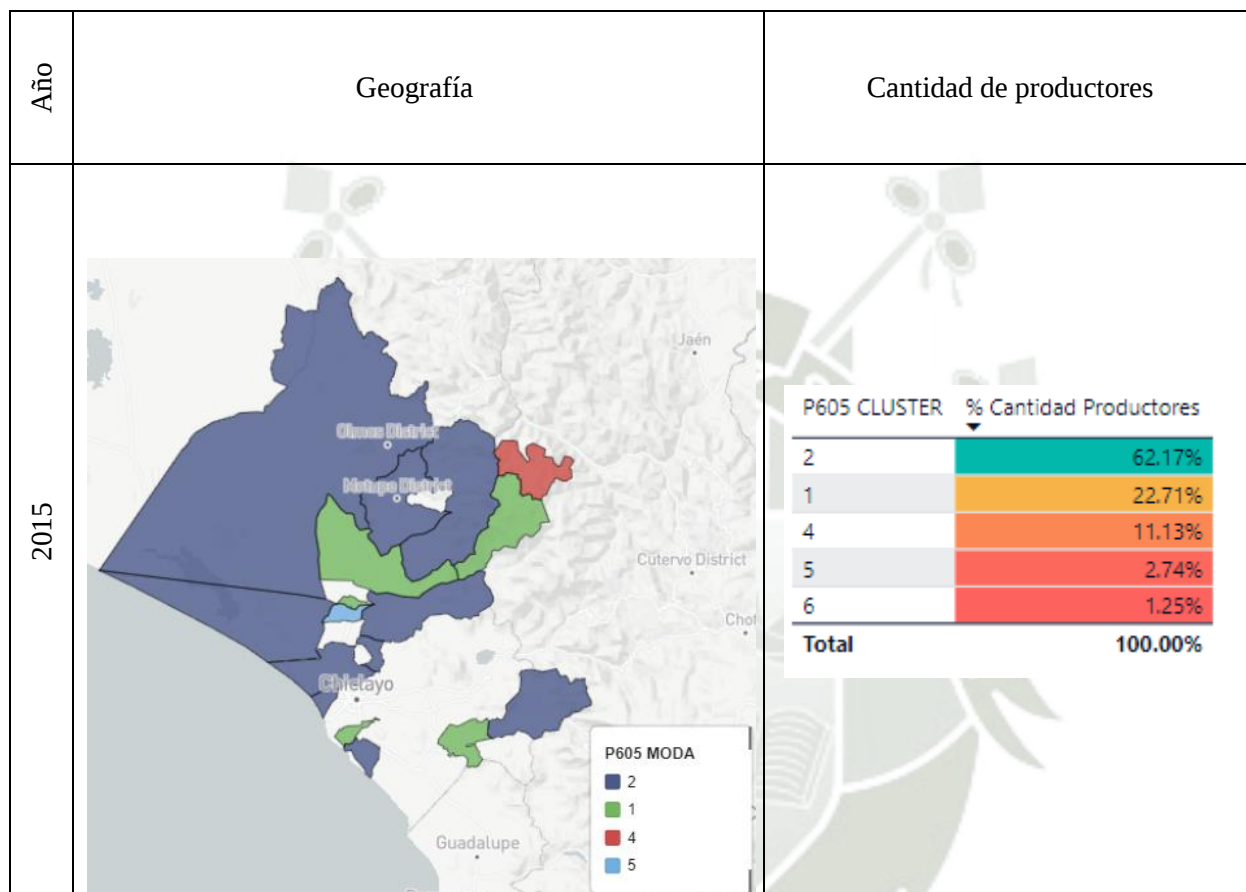
Figura 54.
Mapa de los clústeres P605.

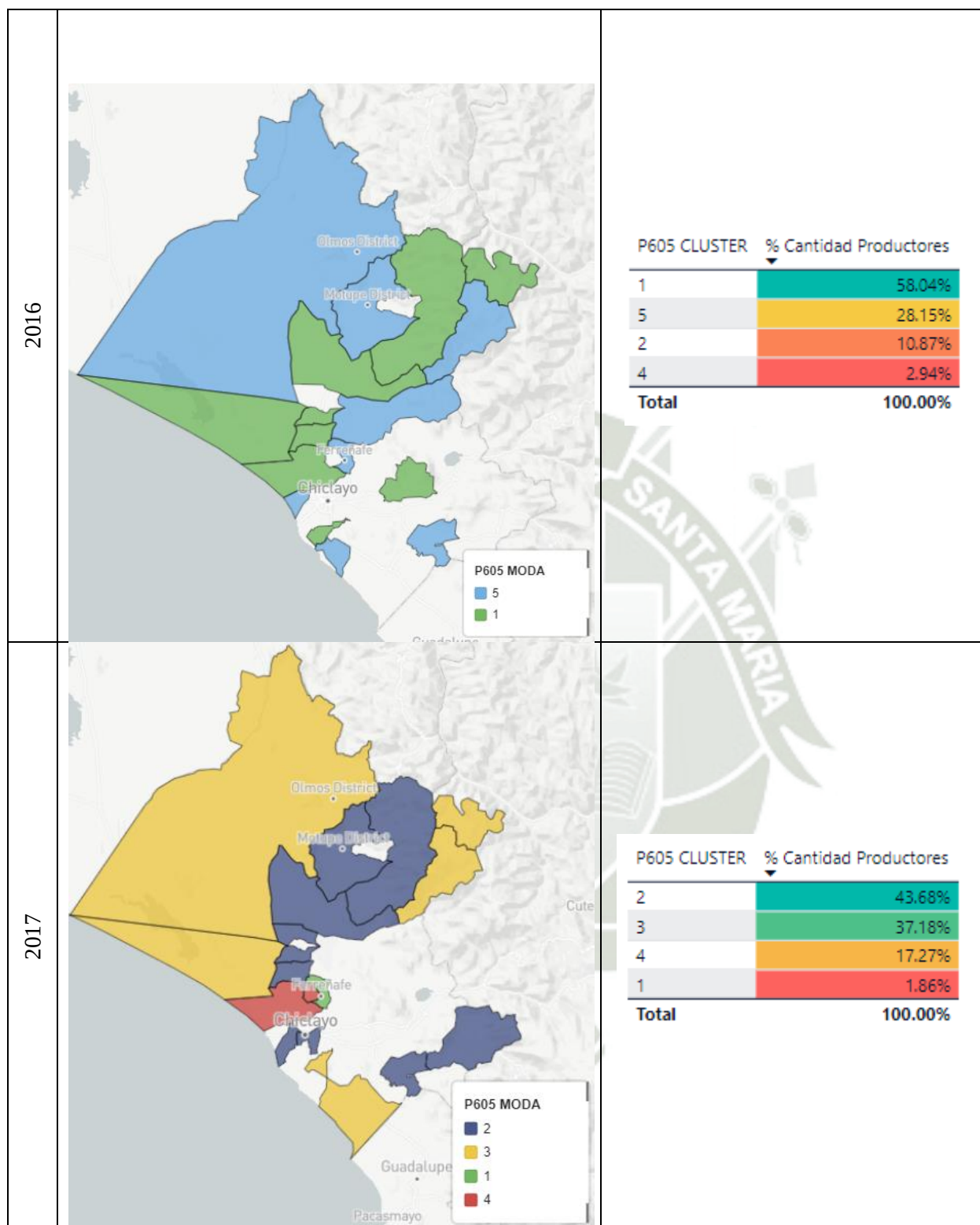


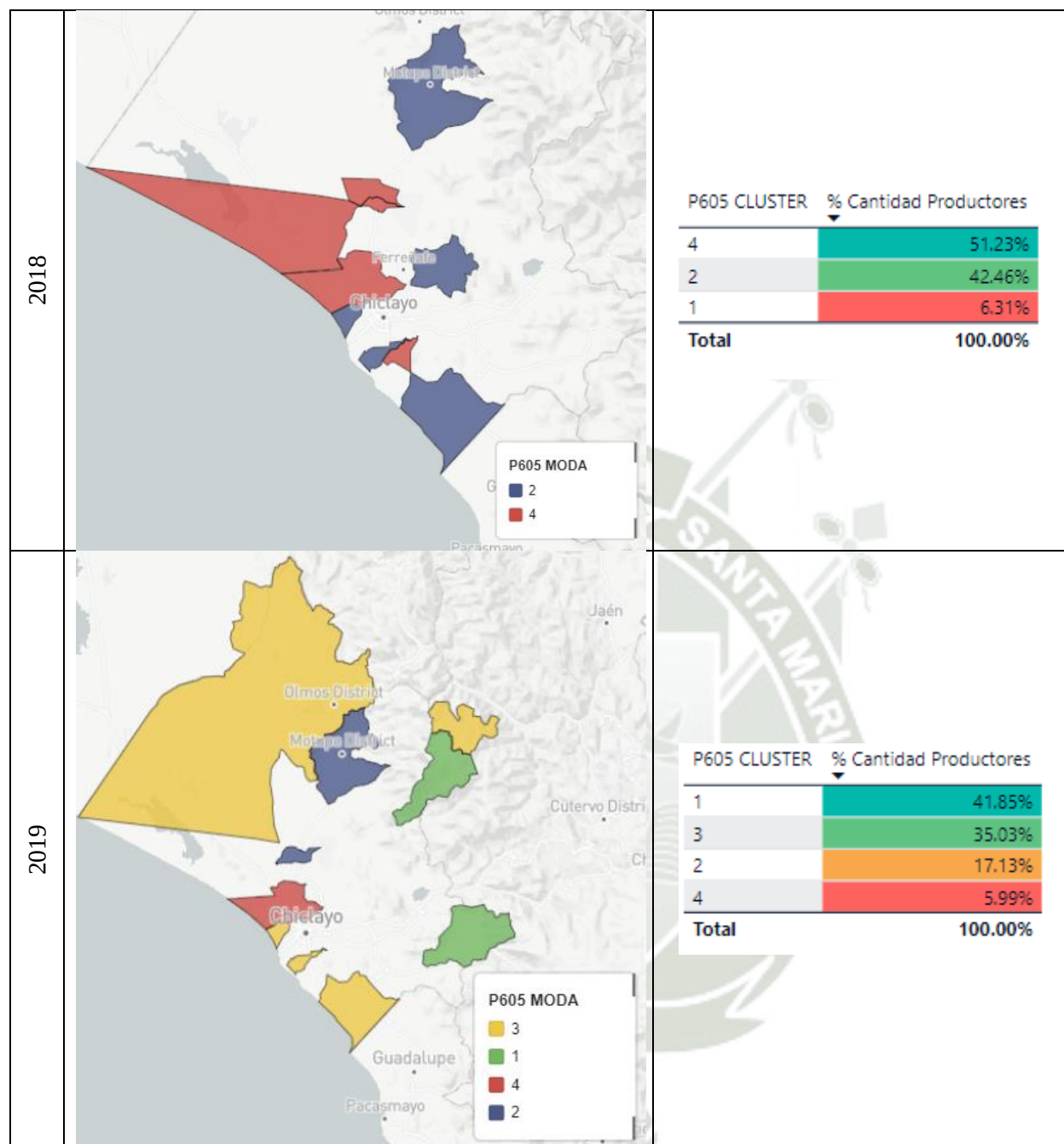
Fuente: Elaboración Propia.

En Lambayeque se recibió información principalmente del Ministerio de Salud, SENASA, MIDAGRI, otro, Gobierno Local Municipal y gobierno regional en ese orden.

Figura 55.
Características del departamento de Lambayeque.





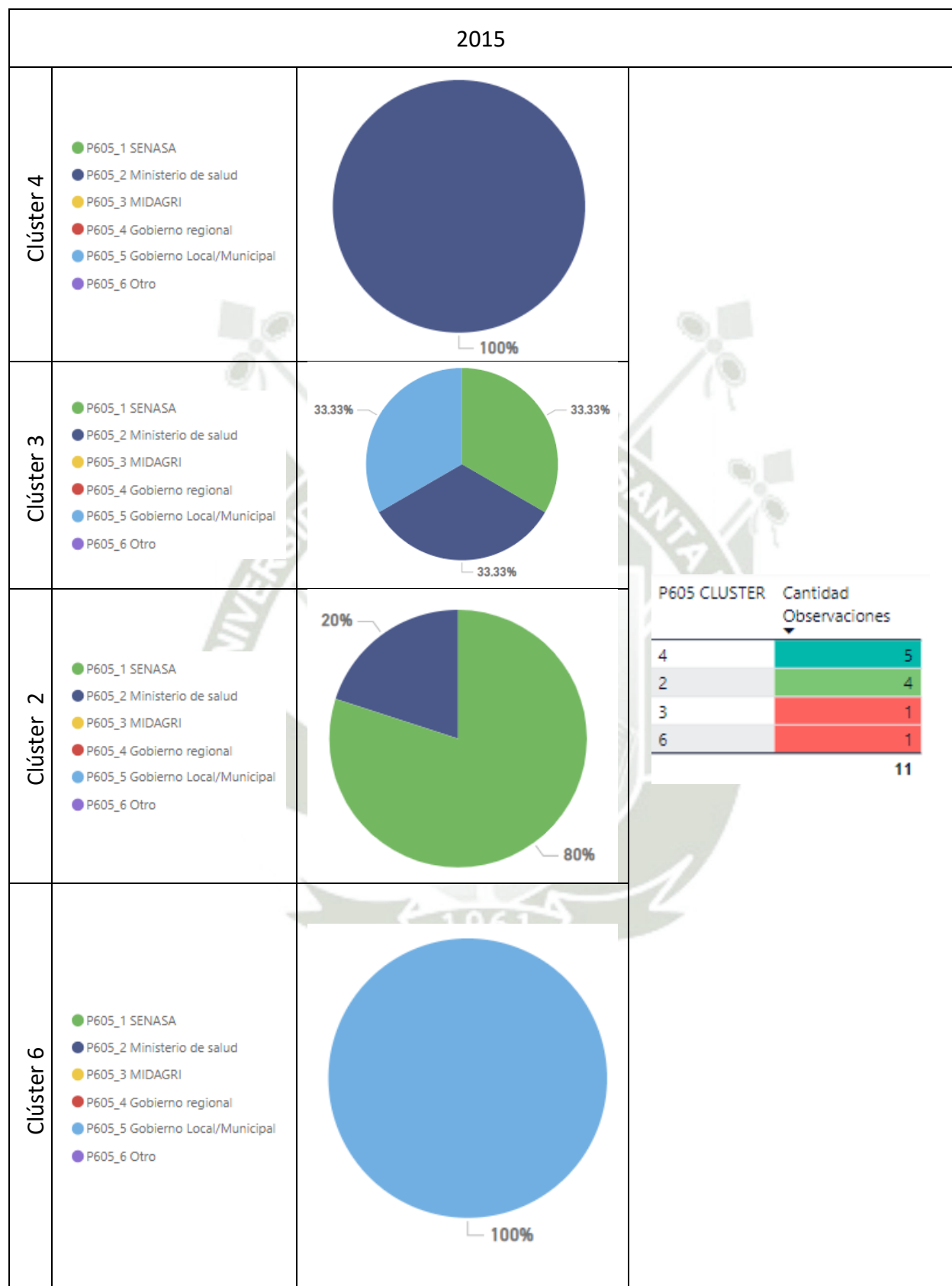


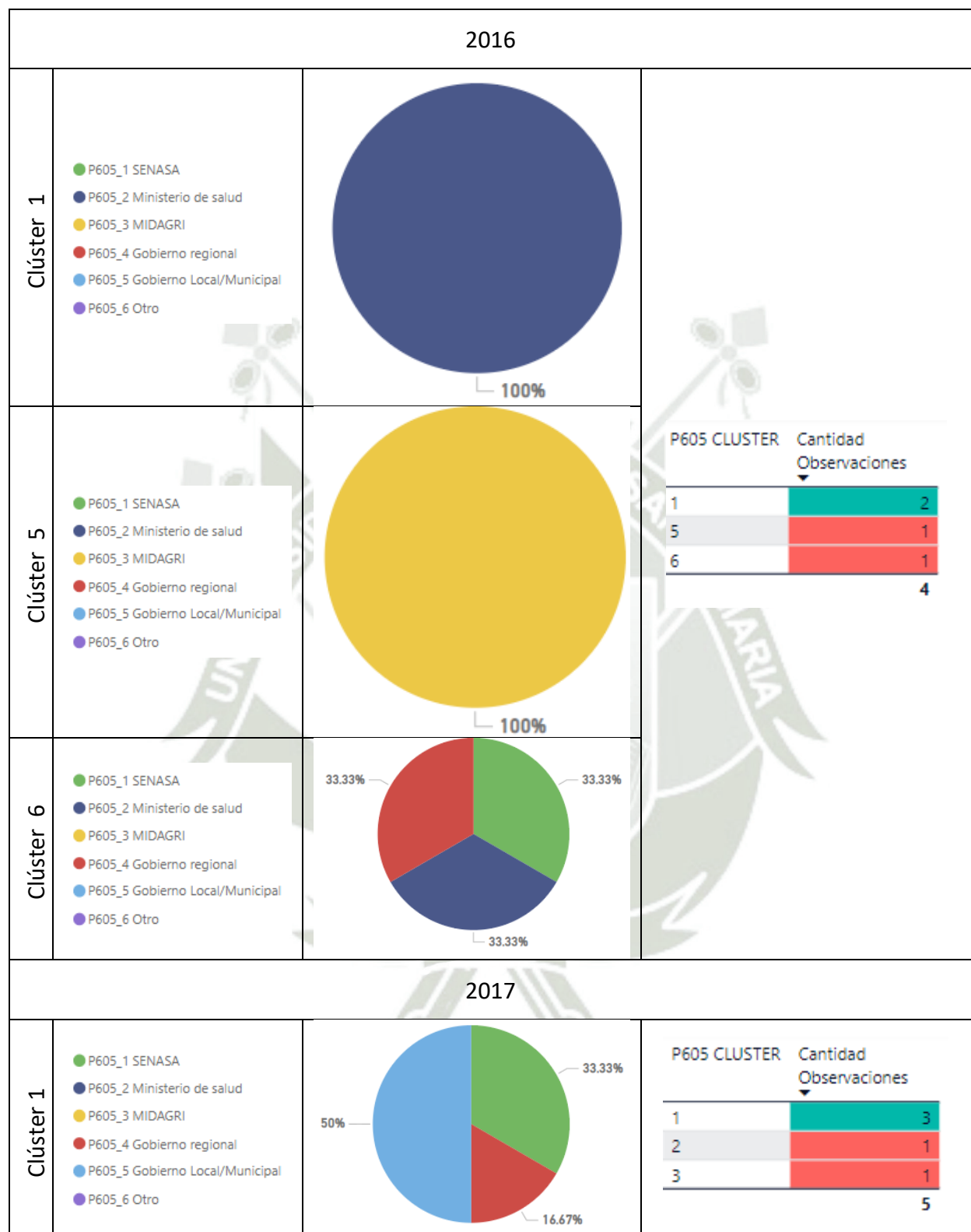
Fuente: Elaboración Propia.

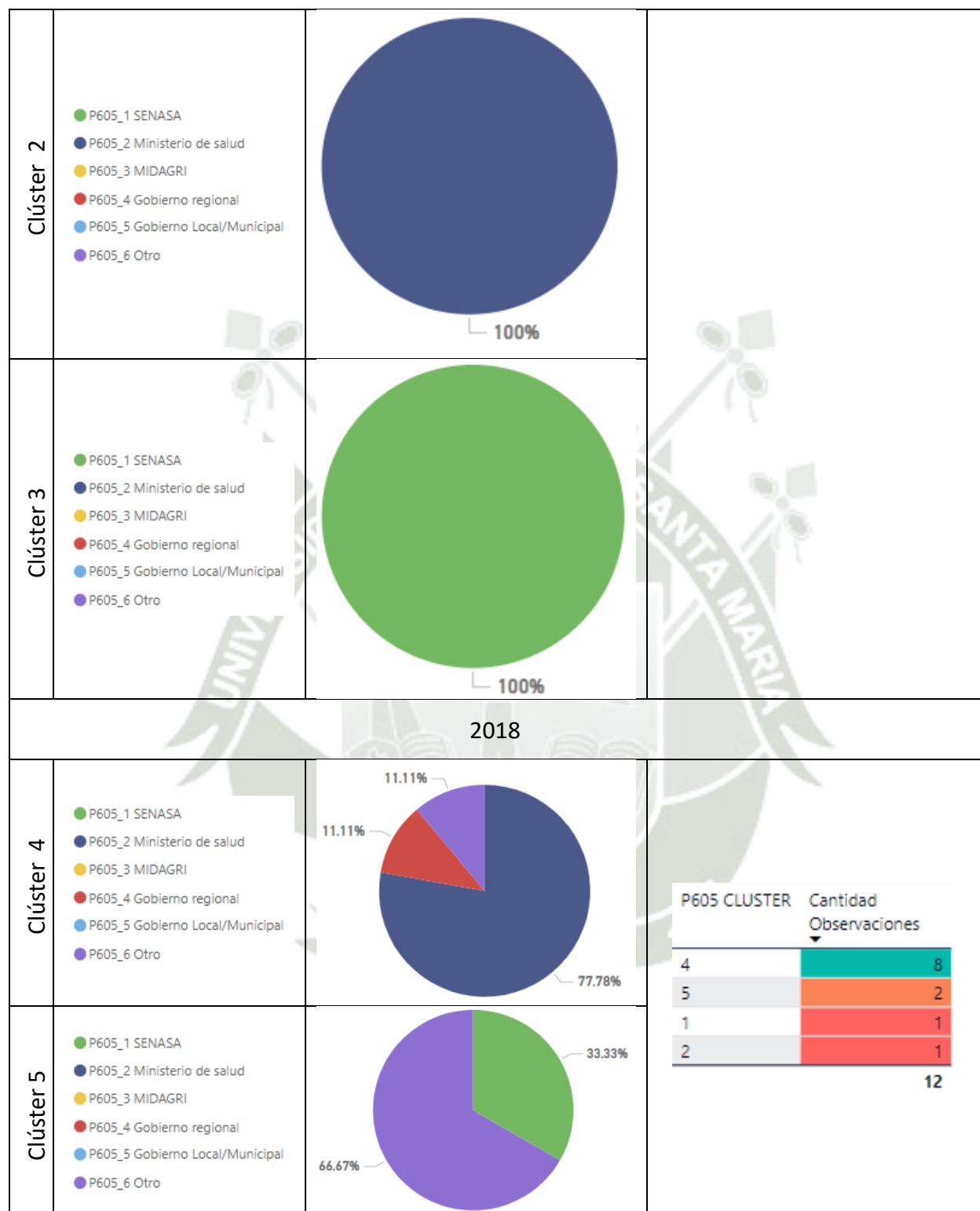
En Moyobamba de un total de 711 observaciones, se clasificaron 38 productores, donde este distrito se caracterizó por que los productores recibieron información principalmente del Ministerio de Salud, SENASA, Gobierno local/Municipal, entre otros, y su clasificación es la siguiente.

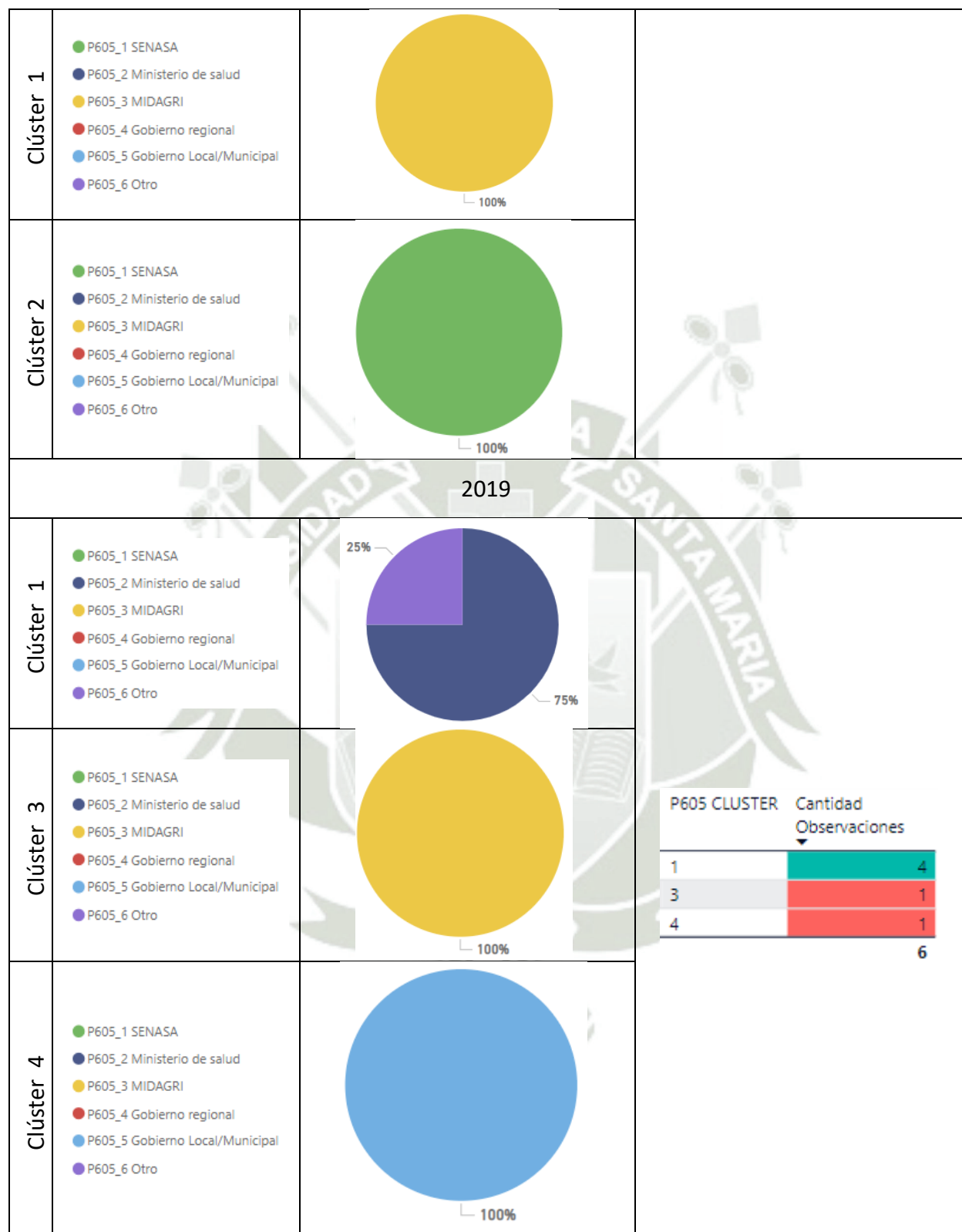
Figura 56.
Características del distrito de Moyobamba.

DEPARTAMENTO: SAN MARTÍN PROVINCIA: MOYOBAMBA DISTRITO: MOYOBAMBA			
Clúster por año		Observaciones por año	
AÑO	DISTRITO	P605 CLUSTER	Cantidad Observaciones
2015	MOYOBAMBA	4	5
2015	MOYOBAMBA	2	4
2015	MOYOBAMBA	3	1
2015	MOYOBAMBA	6	1
2016	MOYOBAMBA	1	2
2016	MOYOBAMBA	5	1
2016	MOYOBAMBA	6	1
2017	MOYOBAMBA	1	3
2017	MOYOBAMBA	2	1
2017	MOYOBAMBA	3	1
2018	MOYOBAMBA	4	8
2018	MOYOBAMBA	5	2
2018	MOYOBAMBA	1	1
2018	MOYOBAMBA	2	1
2019	MOYOBAMBA	1	4
2019	MOYOBAMBA	3	1
2019	MOYOBAMBA	4	1
Total			38
AÑO	DISTRITO	Cantidad Observaciones	
2015	MOYOBAMBA	129	
2016	MOYOBAMBA	137	
2017	MOYOBAMBA	135	
2018	MOYOBAMBA	150	
2019	MOYOBAMBA	160	
Total		711	









Fuente: Elaboración Propia.

3.1.5. Módulo 700 Servicios de extensión agraria

A. Variable P710: ¿Qué información agropecuaria necesita para desarrollar su actividad?

Ayuda a determinar las necesidades de los productores respecto a la información que necesitan para desarrollar su actividad agropecuaria.

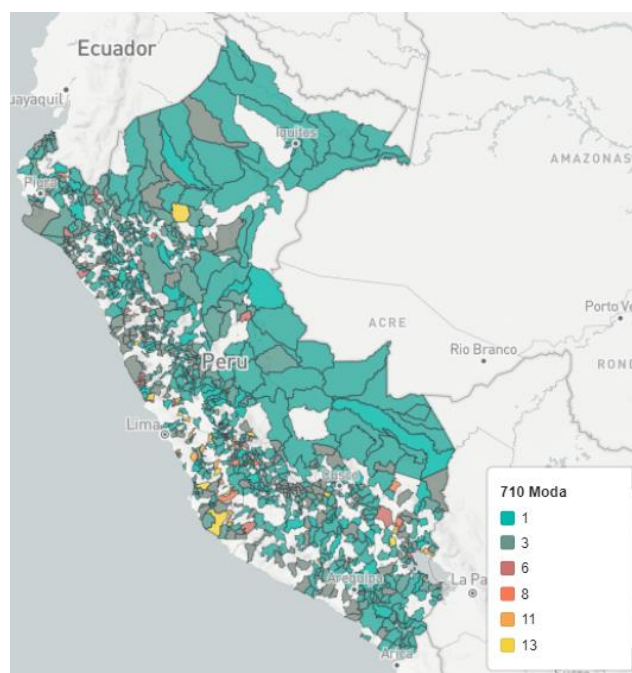
Para el año 2015, los productores requerían información principalmente de técnicas de manejo de cultivos y crianza con 92058 productores, manejo de enfermedades y plagas 70701 productores, uso de abonos y fertilizantes 58054 productores, técnicas de manejo de semillas 29660 productores, con 27408 técnicas y sistema de riego con, entre otros.

Figura 57.
Características de los clústeres en P710 año 2015

AÑO	P710 CLUSTER	P710_1	P710_2	P710_3	P710_4	P710_5	P710_6	P710_7	P710_8	P710_10	Cantidad Observaciones
2015	2	545	2,360	0	0	8,305	1,066	0	672	47	8,305
2015	1	630	14	661	40	2,355	1,198	3,456	522	102	5,169
2015	4	0	4,507	605	0	2,717	2,025	2,880	2,354	19	5,153
2015	3	866	4,099	0	686	2,192	86	3,452	0	8	4,454
2015	6	363	717	517	596	539	666	730	730	1	730
2015	13	710	450	38	29	402	349	544	468	2	710
2015	11	541	501	505	121	578	394	476	170	6	622
2015	8	12	253	183	615	615	16	224	75	3	615
2015	5	93	178	158	373	163	232	373	0	2	373
2015	12	160	208	245	69	0	88	150	74	4	245
2015	9	100	0	88	215	133	187	71	40	1	231
2015	7	3	96	105	110	32	64	9	94	1	132
2015	10	72	29	40	110	37	13	59	110	0	110
Total		4,095	13,412	3,145	2,964	18,068	6,384	12,424	5,309	196	26,849

Fuente: Elaboración Propia.

Figura 58.
Mapa de los clústeres P710 año 2015.



Fuente: Elaboración Propia.

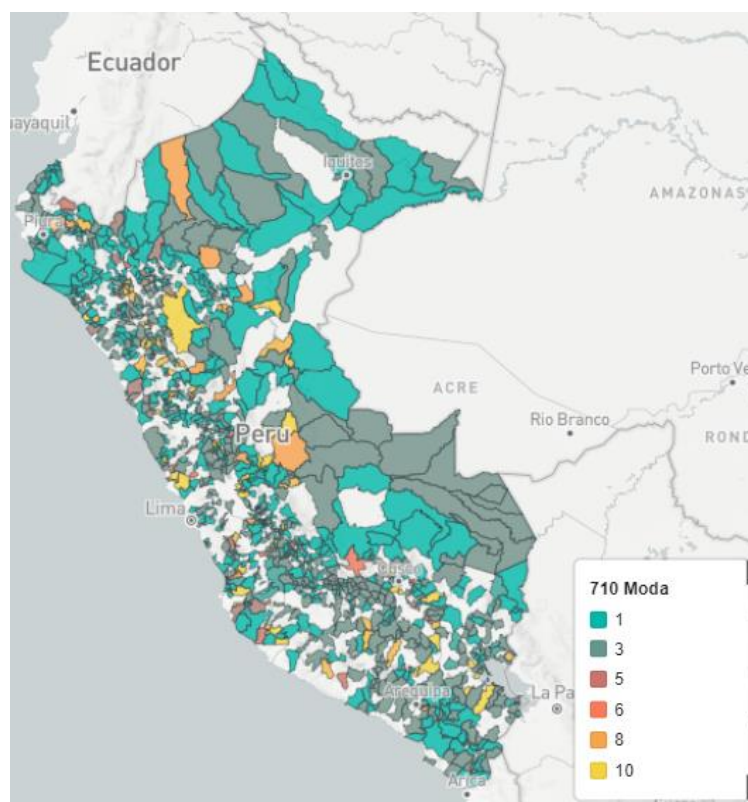
En el año 2016 se repite las mismas necesidades que el año 2015.

Figura 59.
Características de los clústeres en P710 año 2016

AÑO	P710 CLUSTER	P710_1	P710_2	P710_3	P710_4	P710_5	P710_6	P710_7	P710_8	P710_10	Cantidad Observaciones
2016	1	0	9,142	268	372	5,901	1,252	5,282	1,598	45	9,738
2016	3	1,113	74	74	331	8,131	1,002	1,752	1,565	66	8,617
2016	9	0	485	205	25	726	1,812	1,239	599	13	1,812
2016	10	1,195	659	136	5	6	247	327	271	305	1,663
2016	2	1,306	1,557	420	246	1,557	792	1,040	385	10	1,557
2016	4	405	1,238	723	673	1,289	1,287	1,129	1,202	6	1,298
2016	5	47	739	703	302	768	21	370	191	4	855
2016	8	331	378	304	596	37	278	336	20	9	754
2016	6	314	207	479	424	577	6	326	258	0	655
2016	7	177	135	154	174	8	13	394	536	3	551
Total		4,888	14,614	3,466	3,148	19,000	6,710	12,195	6,625	461	27,500

Fuente: Elaboración Propia.

Figura 60.
Mapa de los clústeres P710 año 2016.



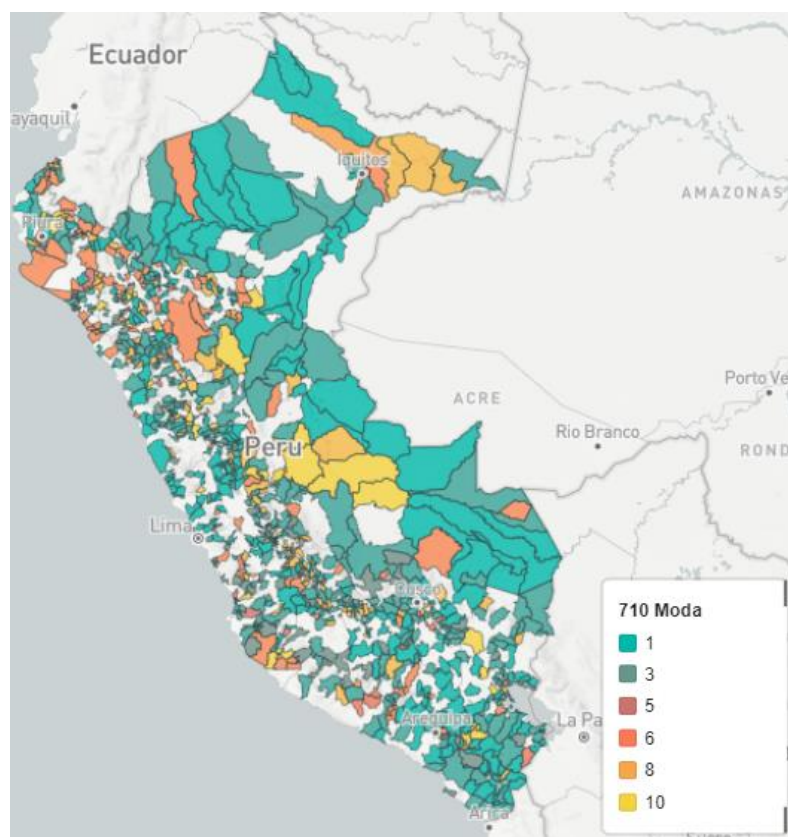
Fuente: Elaboración Propia.

Figura 61.
Características de los clústeres en P710 año 2017

AÑO	P710 CLUSTER	P710_1	P710_2	P710_3	P710_4	P710_5	P710_6	P710_7	P710_8	P710_10	Cantidad Observaciones
2017	2	1,386	6,905	1,052	638	6,636	1,594	3,469	449	54	6,905
2017	1	901	244	286	685	6,263	589	12	28	68	6,490
2017	7	12	2,492	221	283	8	592	1,003	185	566	3,621
2017	9	5	780	479	348	2,550	1,453	3,444	804	22	3,467
2017	3	393	2,027	0	253	1,937	562	1,377	2,263	17	2,263
2017	10	1,329	734	529	279	234	35	1,474	246	12	2,248
2017	6	139	488	315	0	764	54	140	1,298	17	1,298
2017	5	456	156	0	99	291	278	390	218	2	474
2017	8	33	309	461	461	310	214	294	207	4	461
2017	4	194	81	102	103	29	59	119	216	2	216
Total		4,848	14,216	3,445	3,149	19,022	5,430	11,722	5,914	764	27,443

Fuente: Elaboración Propia.

Figura 62.
Mapa de los clústeres P710 año 2017



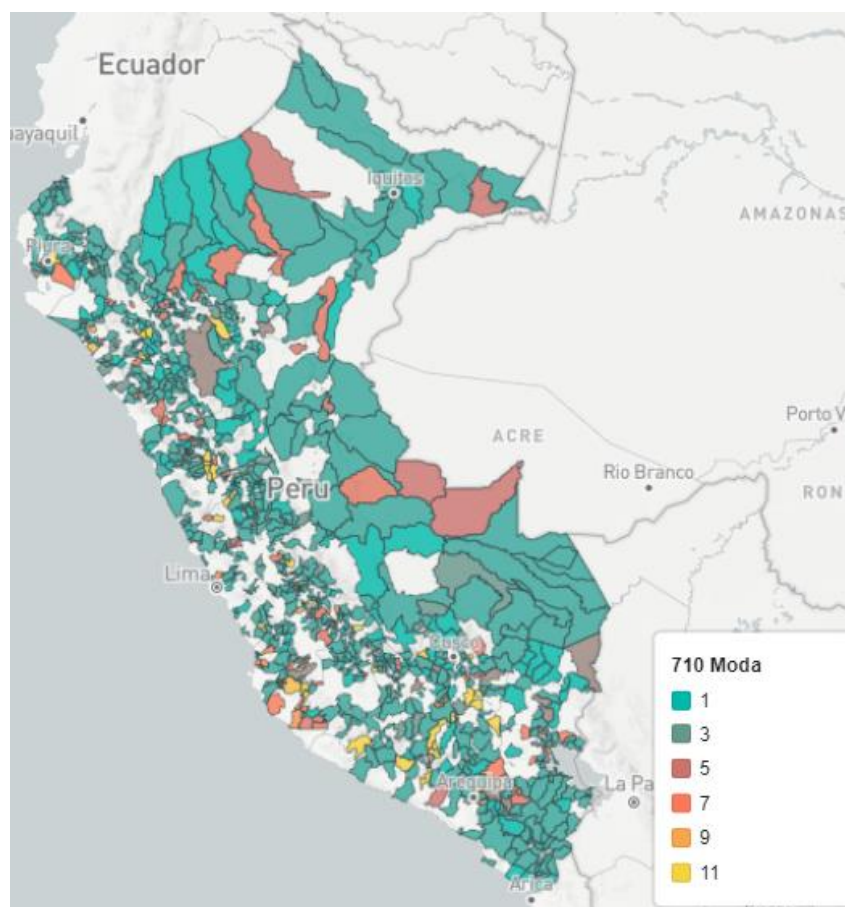
Fuente: Elaboración Propia.

Figura 63.
Características de los clústeres en P710 año 2018

AÑO	P710 CLUSTER	P710_1	P710_2	P710_3	P710_4	P710_5	P710_6	P710_7	P710_8	P710_10	Cantidad Observaciones
2018	2	1,032	2,578	152	625	6,811	972	4,272	606	91	10,619
2018	1	482	5,913	235	697	5,689	0	2,755	572	45	5,913
2018	3	0	2,815	320	55	1,206	552	0	902	38	2,815
2018	6	89	1,244	441	351	1,397	1,990	2,091	496	4	2,091
2018	11	0	61	71	93	667	415	52	1,050	470	1,814
2018	4	857	351	183	2	740	524	224	158	0	959
2018	5	62	131	203	883	436	258	0	0	18	883
2018	7	445	329	864	181	499	7	433	0	2	864
2018	8	431	437	503	565	596	392	496	435	2	596
2018	9	266	213	70	125	16	138	131	289	0	292
2018	10	36	190	242	0	134	107	214	242	0	242
Total		3,700	14,262	3,284	3,577	18,191	5,355	10,668	4,750	670	27,088

Fuente: Elaboración Propia.

Figura 64.
Mapa de los clústeres P710 año 2018



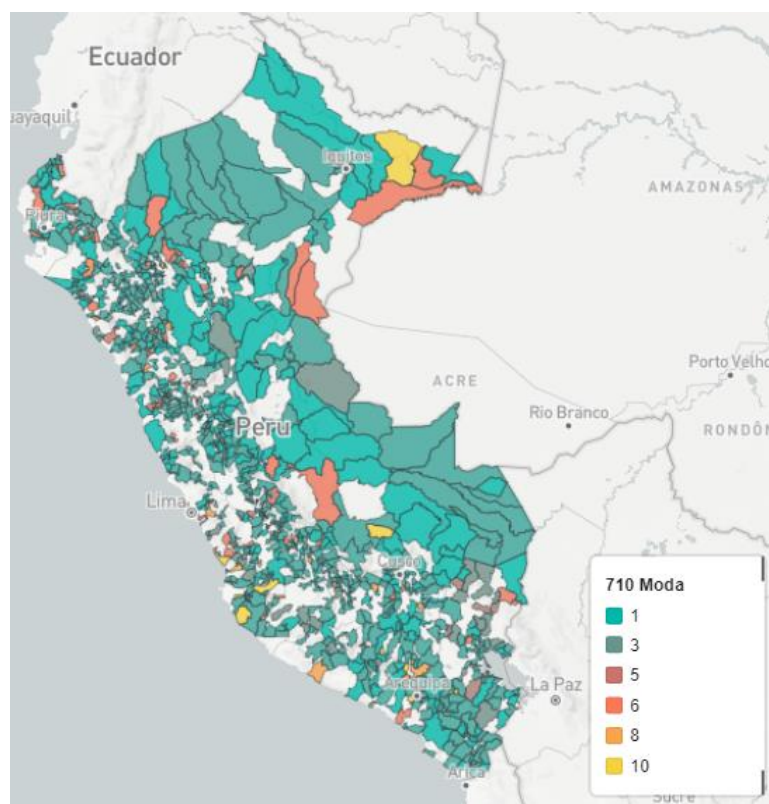
Fuente: Elaboración Propia.

Figura 65.
Características de los clústeres en P710 año 2019

AÑO	P710 CLUSTER	P710_1	P710_2	P710_3	P710_4	P710_5	P710_6	P710_7	P710_8	P710_10	Cantidad Observaciones
2019	2	0	1,524	142	91	7,318	902	3,286	2,607	0	9,607
2019	1	249	7,437	638	518	6,554	2,308	2,834	744	0	8,082
2019	6	417	2,665	416	14	682	1,189	2,520	187	0	2,673
2019	3	1,611	561	325	451	669	129	307	17	621	2,607
2019	5	0	599	380	947	232	171	519	12	0	947
2019	7	126	508	730	77	438	50	124	401	0	903
2019	8	108	82	374	152	840	535	722	370	0	871
2019	10	438	430	364	444	307	399	428	403	0	467
2019	9	74	214	25	454	411	0	0	69	0	454
2019	4	326	177	0	69	326	98	305	0	0	326
Total		3,349	14,197	3,394	3,217	17,777	5,781	11,045	4,810	621	26,937

Fuente: Elaboración Propia.

Figura 66.
Mapa de los clústeres P710 año 2019.



Fuente: Elaboración Propia.

B. Variable P713: ¿A través de qué medios de comunicación tuvo acceso a la información sobre la prevención de los daños causados por los eventos climáticos?

Los eventos climáticos son muy variados en todo el Perú por lo que es importante contar con información para poder contrarrestar los daños causados por el clima, esta variable ayudará a determinar cuál es el mejor medio de comunicación para mantener a los agricultores informados.

Estos fueron los principales medios de comunicación que sirvieron para comunicar a los productores acerca de la prevención, de los daños causados por los eventos climáticos, el principal fue la radio, seguido de televisión, talleres, diarios, folletos, internet, otros y teléfono.

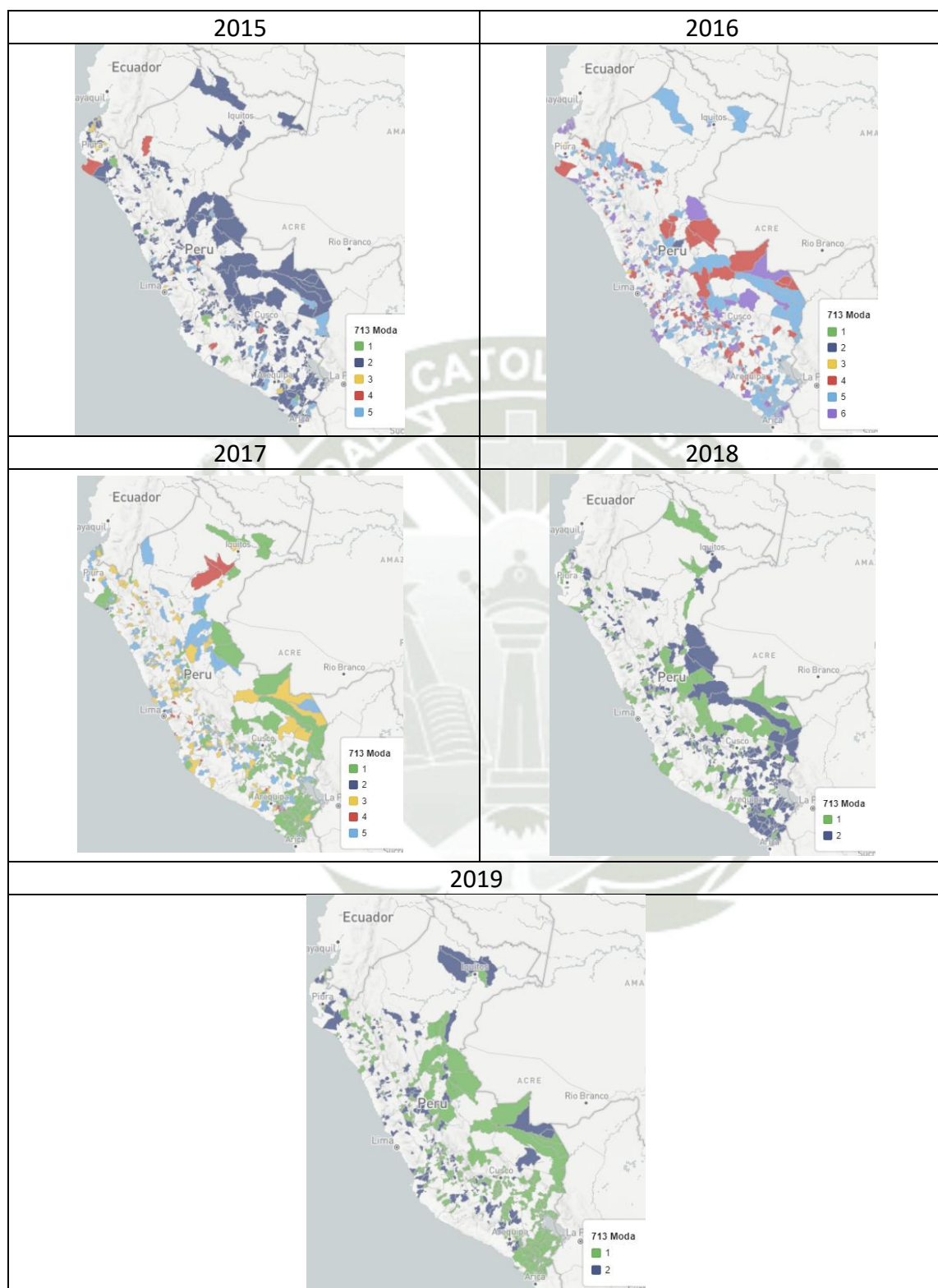
Figura 67.
Características de los clústeres en P713

AÑO	P713 CLUSTER	P713_1 Radio	P713_2 Television	P713_3 Telefono	P713_4 Diarios	P713_5 Folletos	P713_6 Internet	P713_7 Talleres	P713_8 Otros	Cantidad Observaciones
2015	2	817	307	5	0	0	9	251	0	1,147
2015	1	64	0	0	69	4	0	0	50	119
2015	3	112	118	0	115	10	4	6	0	118
2015	5	20	7	0	0	75	2	15	0	75
2015	4	0	4	0	2	2	17	1	0	17
2016	5	1265	0	1	0	0	0	24	120	1,374
2016	6	765	1045	18	143	0	0	0	7	1,084
2016	4	0	9	0	3	18	0	257	0	258
2016	1	84	57	0	25	54	36	1	0	84
2016	2	9	31	0	28	27	25	2	0	77
2016	3	23	23	0	4	3	1	23	0	23
2017	1	1722	130	2	162	0	2	0	0	1,722
2017	5	854	741	0	25	61	0	30	170	963
2017	3	3	270	5	23	0	34	257	7	538
2017	2	144	144	40	99	60	89	14	0	144
2017	4	0	7	0	17	59	0	16	6	69
2018	2	1460	80	1	51	0	70	7	0	1,529
2018	1	317	438	11	0	51	5	229	98	800
2019	1	2383	417	54	25	1	69	13	0	2,391
2019	2	15	166	30	0	48	66	180	68	513
Total		10057	3994	167	791	473	429	1326	526	13,045

Fuente: Elaboración Propia.



Figura 68.
Mapa de los clústeres P713.

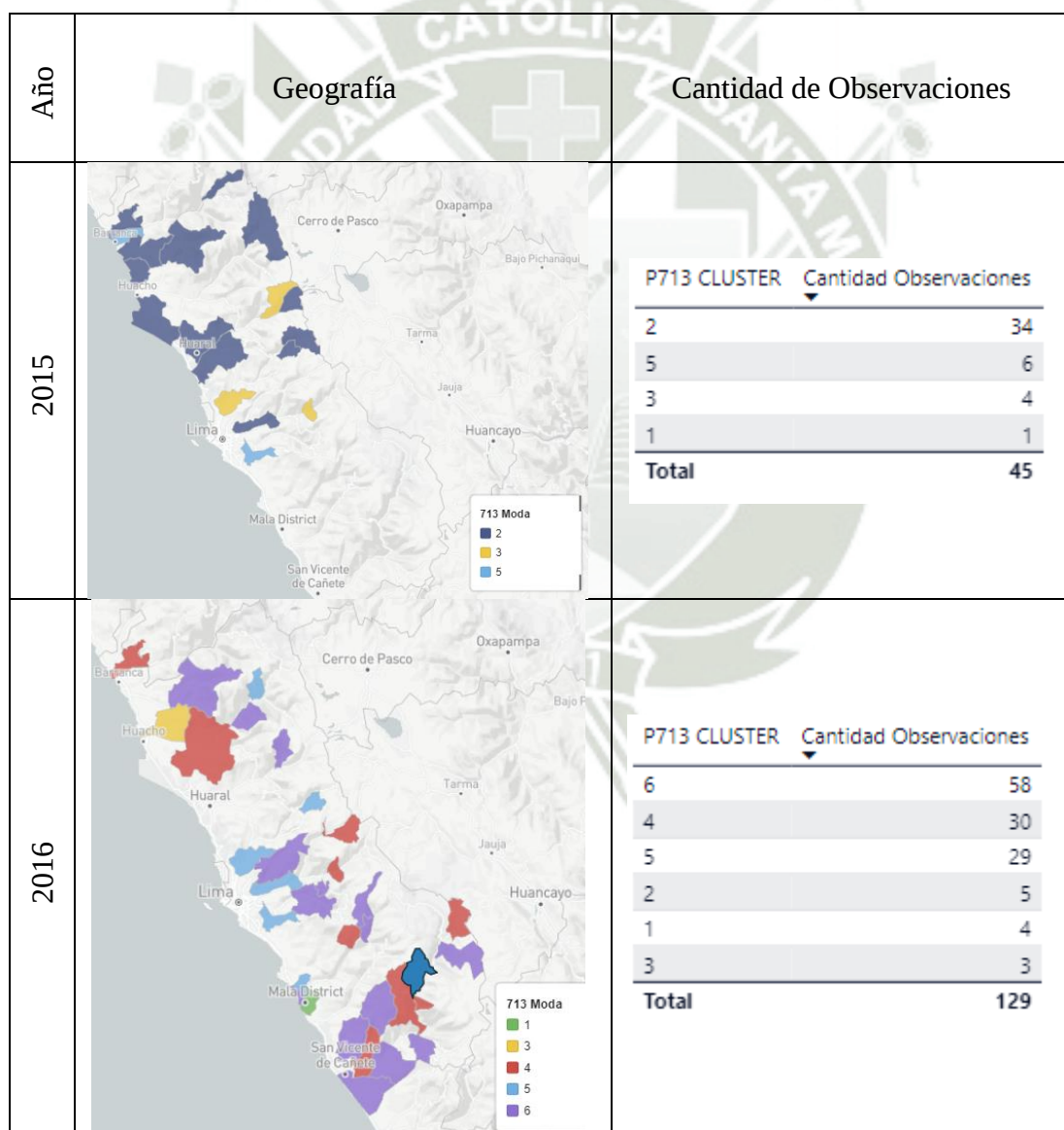


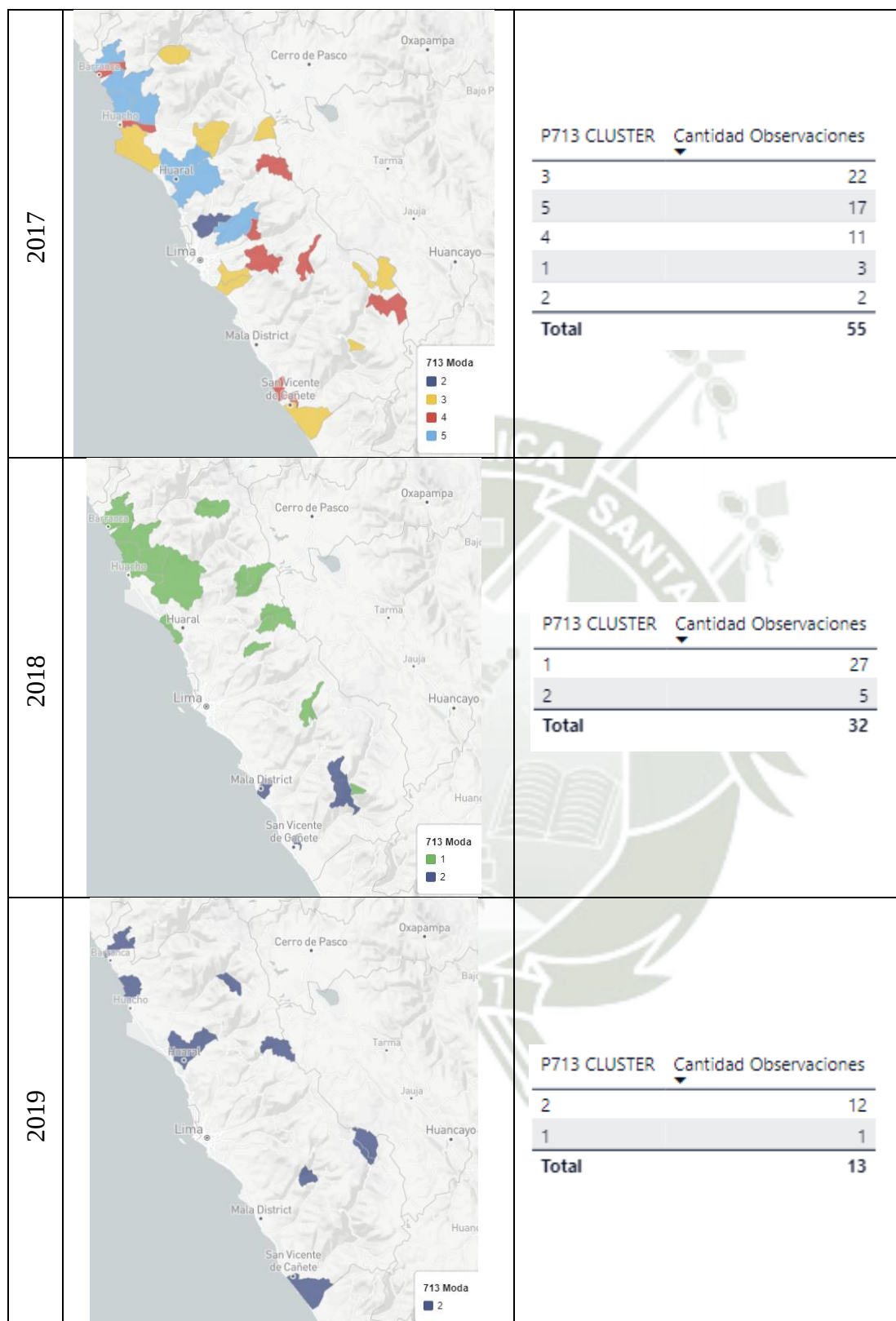
Fuente: Elaboración Propia.

Características del departamento de Lima.

En el año 2015 en el departamento de Lima se tuvo acceso a la información a través de radio, televisión, diarios, folletos internet, talleres, entre otros. Para el año 2016 se tiene mucha más cantidad de observaciones y el acceso a la información fue principalmente a través de televisión, talleres, radio, diarios, folletos, entre otros, en el año 2017 la principal fueron los talleres y folletos, para el año 2018, fueron los talleres, radio y televisión, en el año 2019 los talleres, televisión y folletos.

Figura 69.
Características departamento de Lima

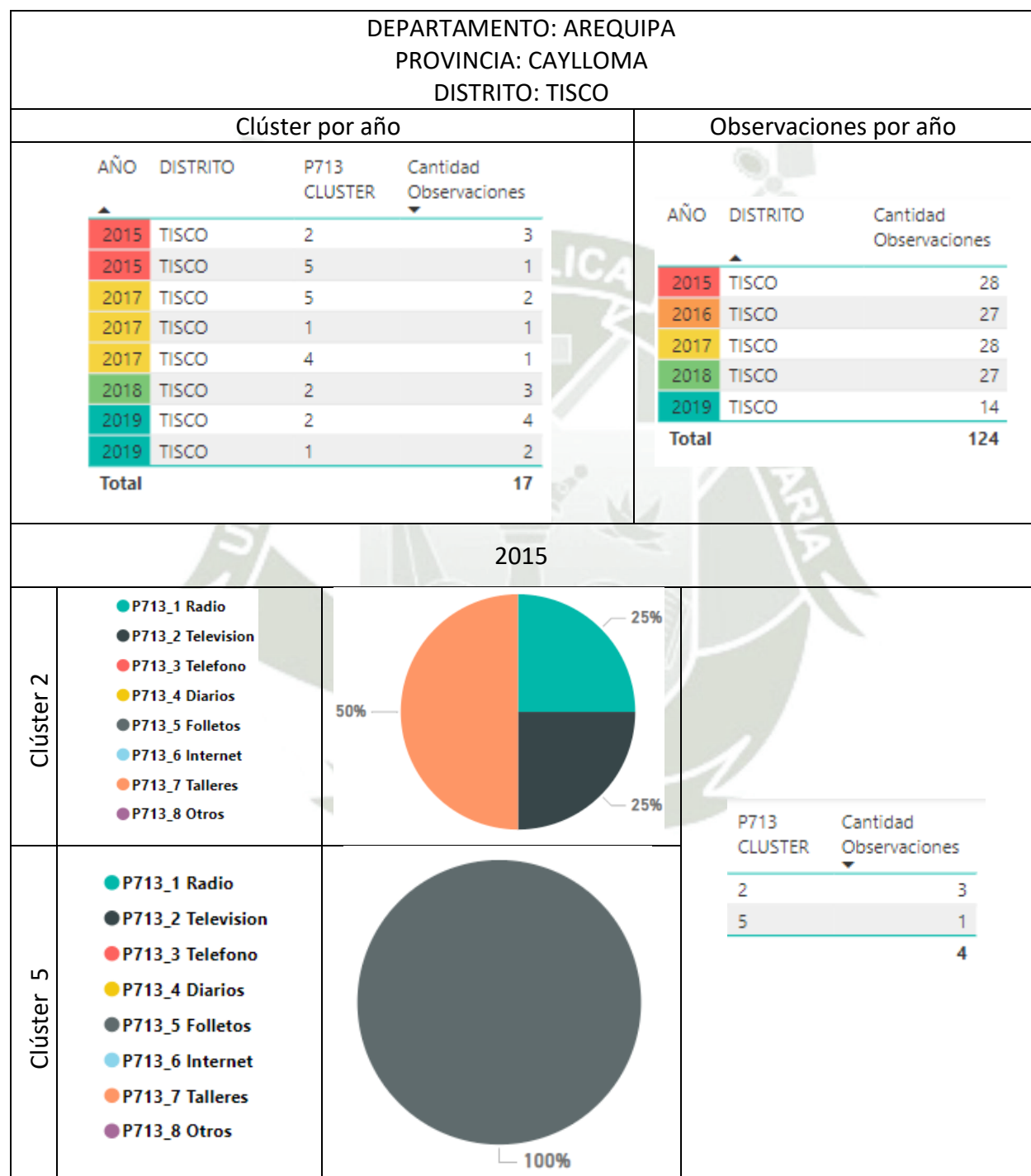


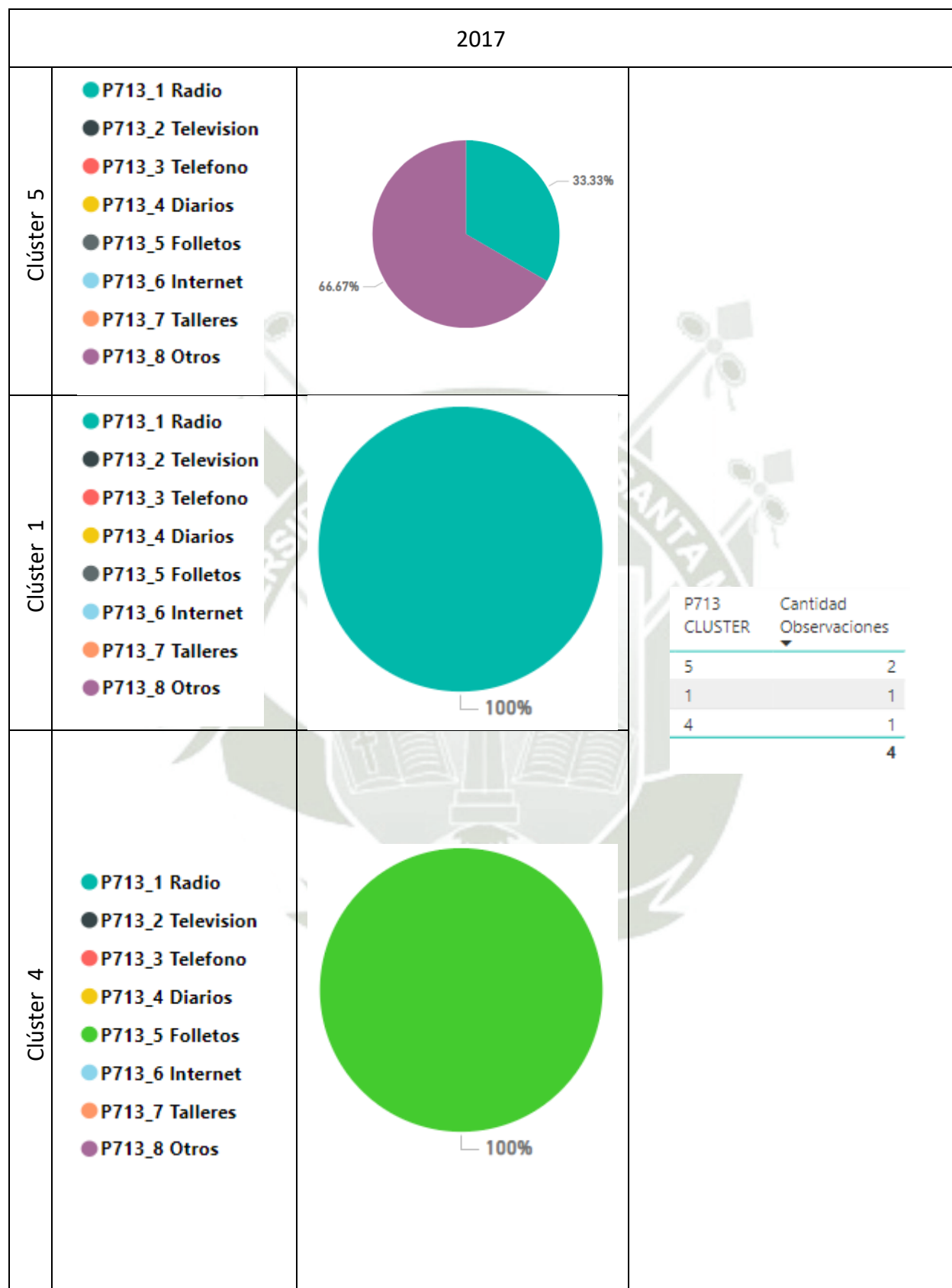


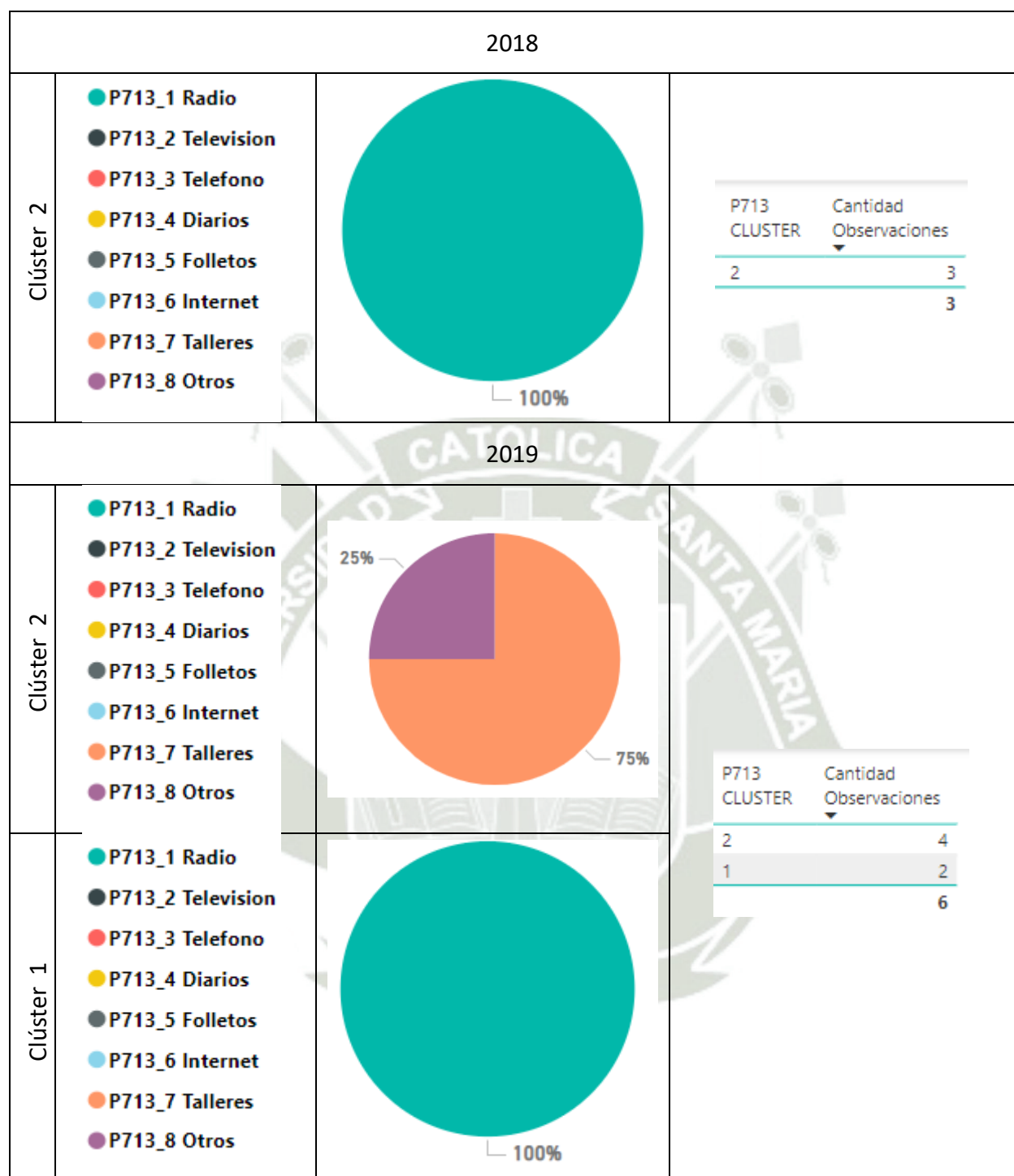
Fuente: Elaboración Propia.

El distrito de Tisco, de un total de 124 observaciones, se clasificaron 17 productores, en este distrito los principales medios de comunicación fueron los talleres, radio, folletos u otros.

Figura 70.
Características del distrito de Tisco.







Fuente: Elaboración Propia.

3.1.6. Módulo 900 Servicios financieros

A. Variable P903: ¿Quién le proporcionó el crédito que obtuvo?

Ayudará a entender el aspecto financiero, y saber la principal fuente de crédito de los productores.

En la siguiente gráfica se encuentra como se caracteriza cada clúster con el número de observaciones que agrupa, y el año al que pertenece, esta pregunta se refiere al organismo que le proporcionó créditos a los productores para realizar sus actividades, donde las principales fueron, Caja municipal, Banca privada, EDPYME, Agrobanco, entre otros.

En el periodo 2015-2019, los productores obtuvieron crédito principalmente de Caja Municipal 5110, Banca Privada 3563, AGROBANCO 2998, ADPYME 2262, Cooperativa 1452, Caja Rural 1314, entre otros.

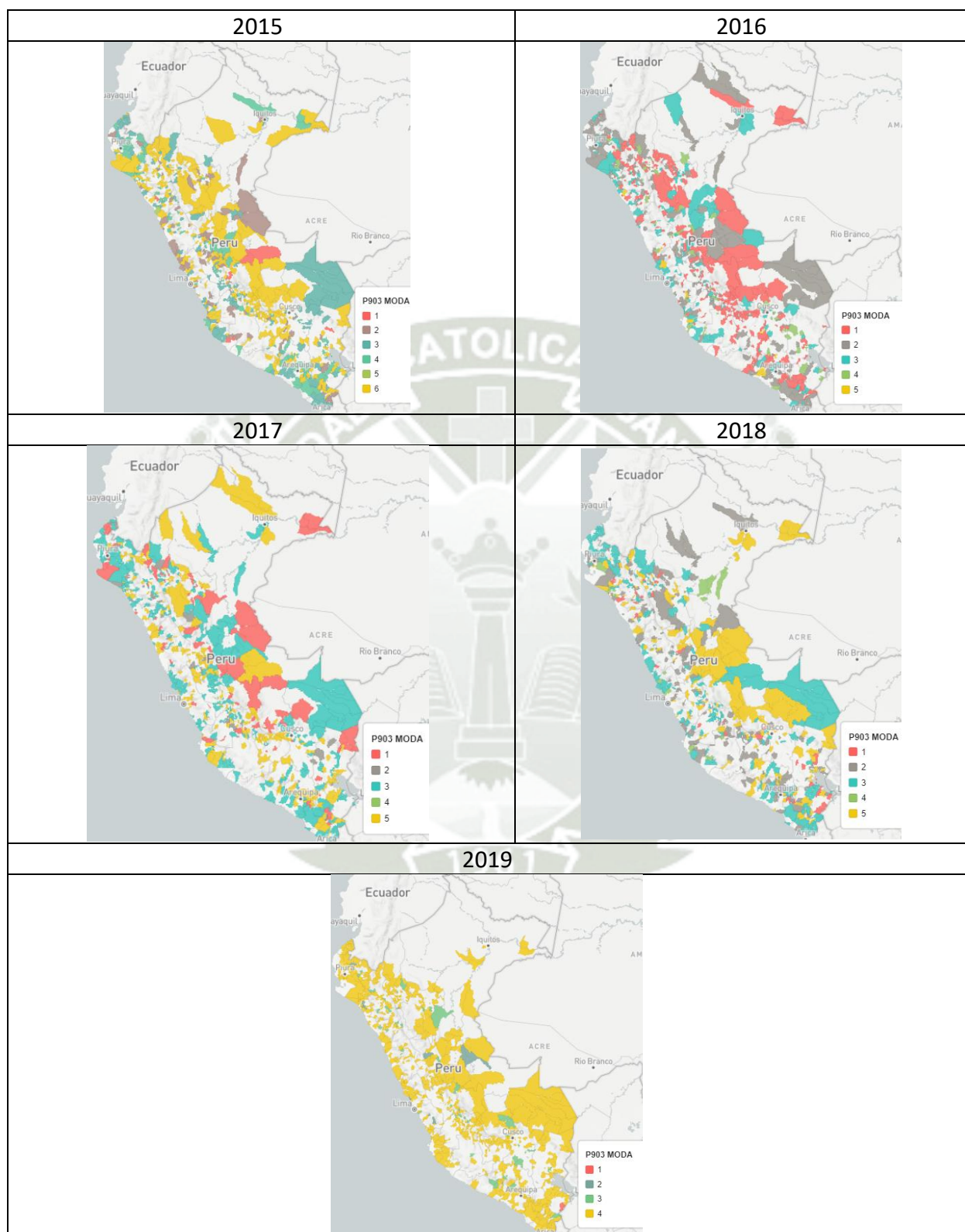
Figura 71.
Características de los clústeres en P903.

AÑO	P903 CLUSTER	P903_1 AGROBANCO	P903_2 Caja Municipal	P903_3 Caja Rural	P903_4 Banca Privada	P903_5 Financiera/ EDPYME	P903_6 ONG	P903_7 Cooperativa	P903_8 Establecimiento Comercial	P903_9 Prestamista	P903_10 Programas del Estado	P903_11 Otro	Cantidad Observaciones
2015	6	545	0	298	1	7	42	310	0	105	1	0	1,281
2015	3	27	933	5	0	0	1	6	0	3	1	0	933
2015	2	9	44	0	541	0	0	3	1	0	11	35	585
2015	4	7	25	0	0	527	0	0	0	1	0	0	527
2015	5	1	2	1	0	0	0	0	35	0	0	0	35
2015	1	1	0	14	15	16	0	0	0	0	0	0	22
2016	1	562	8	5	15	523	37	286	4	0	0	23	1,412
2016	2	0	1,047	0	1	2	0	0	65	121	0	0	1,230
2016	3	0	1	0	726	0	0	7	0	0	8	0	733
2016	4	6	5	244	6	2	0	0	0	0	1	0	244
2016	5	44	109	0	65	30	0	0	1	0	0	0	119
2017	3	1	904	1	35	404	33	0	20	172	0	27	1,552
2017	5	7	9	2	709	2	0	288	0	0	9	0	1,003
2017	1	578	34	0	16	0	0	0	0	0	0	0	578
2017	2	7	7	224	7	5	0	0	0	0	0	0	224
2017	4	7	33	0	11	48	0	0	1	0	1	0	48
2018	3	6	1,007	2	59	6	0	0	4	0	1	0	1,016
2018	2	5	10	0	695	2	0	266	0	2	15	0	961
2018	5	318	33	3	0	350	22	0	0	0	0	0	679
2018	1	0	0	247	7	3	0	4	23	2	0	0	269
2018	4	0	3	0	0	1	0	0	2	132	0	0	132
2019	4	158	839	237	606	3	0	267	13	4	5	0	2,105
2019	3	9	10	16	7	316	23	0	0	0	0	0	349
2019	2	0	47	0	31	15	0	0	0	85	0	0	127
2019	1	0	0	15	10	0	0	15	0	1	0	0	19
Total		2,298	5,110	1,314	3,563	2,262	158	1,452	170	627	53	85	16,183

Fuente: Elaboración Propia.

En el mapa se ve áreas geográficas determinadas donde predomina un clúster, y que en cada departamento también la mayoría comparten características similares.

Figura 72.
Mapa de los clústeres P903.



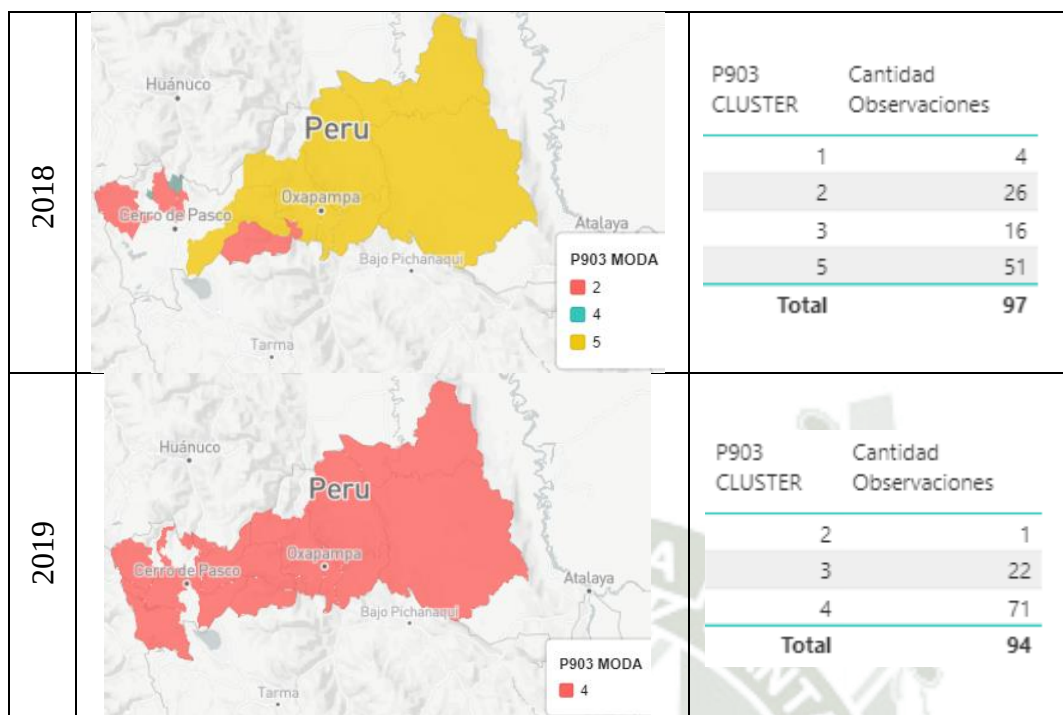
Fuente: Elaboración Propia.

Características del departamento de Pasco, en dicho departamento la principal entidad que brindó crédito a los productores fue, AGROBANCO con 180, Caja municipal con 163, Banca Privada con 155, ADPYME con 117, entre otros.

Figura 73.

Características del departamento de Pasco.

Año	Geografía	Cantidad de Observaciones														
2015		<table><tr><th>P903 CLUSTER</th><th>Cantidad Observaciones</th></tr><tr><td>1</td><td>1</td></tr><tr><td>2</td><td>29</td></tr><tr><td>3</td><td>15</td></tr><tr><td>4</td><td>21</td></tr><tr><td>6</td><td>63</td></tr><tr><td>Total</td><td>129</td></tr></table>	P903 CLUSTER	Cantidad Observaciones	1	1	2	29	3	15	4	21	6	63	Total	129
P903 CLUSTER	Cantidad Observaciones															
1	1															
2	29															
3	15															
4	21															
6	63															
Total	129															
2016		<table><tr><th>P903 CLUSTER</th><th>Cantidad Observaciones</th></tr><tr><td>1</td><td>66</td></tr><tr><td>2</td><td>55</td></tr><tr><td>3</td><td>36</td></tr><tr><td>4</td><td>19</td></tr><tr><td>5</td><td>4</td></tr><tr><td>Total</td><td>180</td></tr></table>	P903 CLUSTER	Cantidad Observaciones	1	66	2	55	3	36	4	19	5	4	Total	180
P903 CLUSTER	Cantidad Observaciones															
1	66															
2	55															
3	36															
4	19															
5	4															
Total	180															
2017		<table><tr><th>P903 CLUSTER</th><th>Cantidad Observaciones</th></tr><tr><td>1</td><td>52</td></tr><tr><td>2</td><td>18</td></tr><tr><td>3</td><td>57</td></tr><tr><td>4</td><td>1</td></tr><tr><td>5</td><td>33</td></tr><tr><td>Total</td><td>161</td></tr></table>	P903 CLUSTER	Cantidad Observaciones	1	52	2	18	3	57	4	1	5	33	Total	161
P903 CLUSTER	Cantidad Observaciones															
1	52															
2	18															
3	57															
4	1															
5	33															
Total	161															



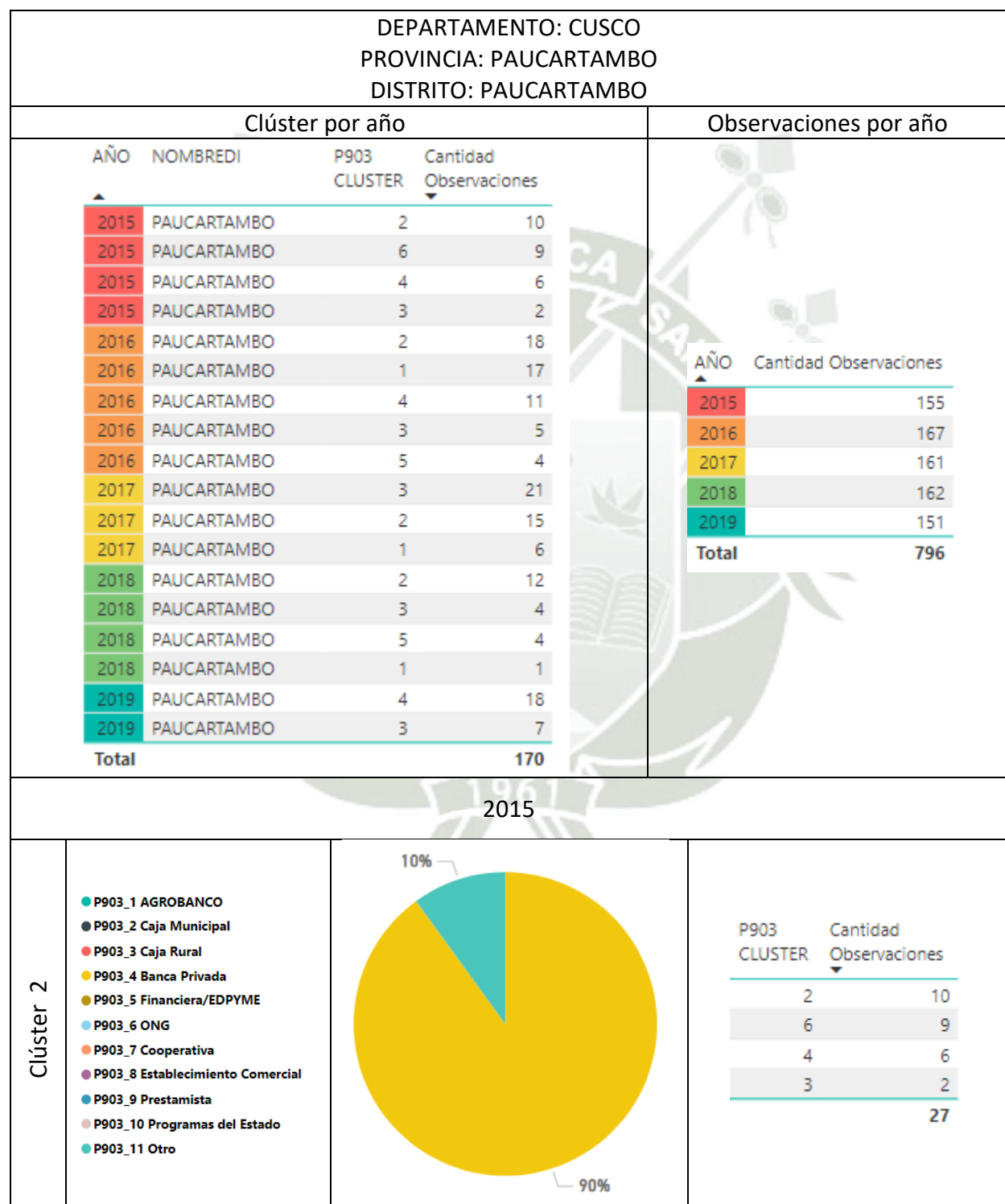
Fuente: Elaboración Propia.

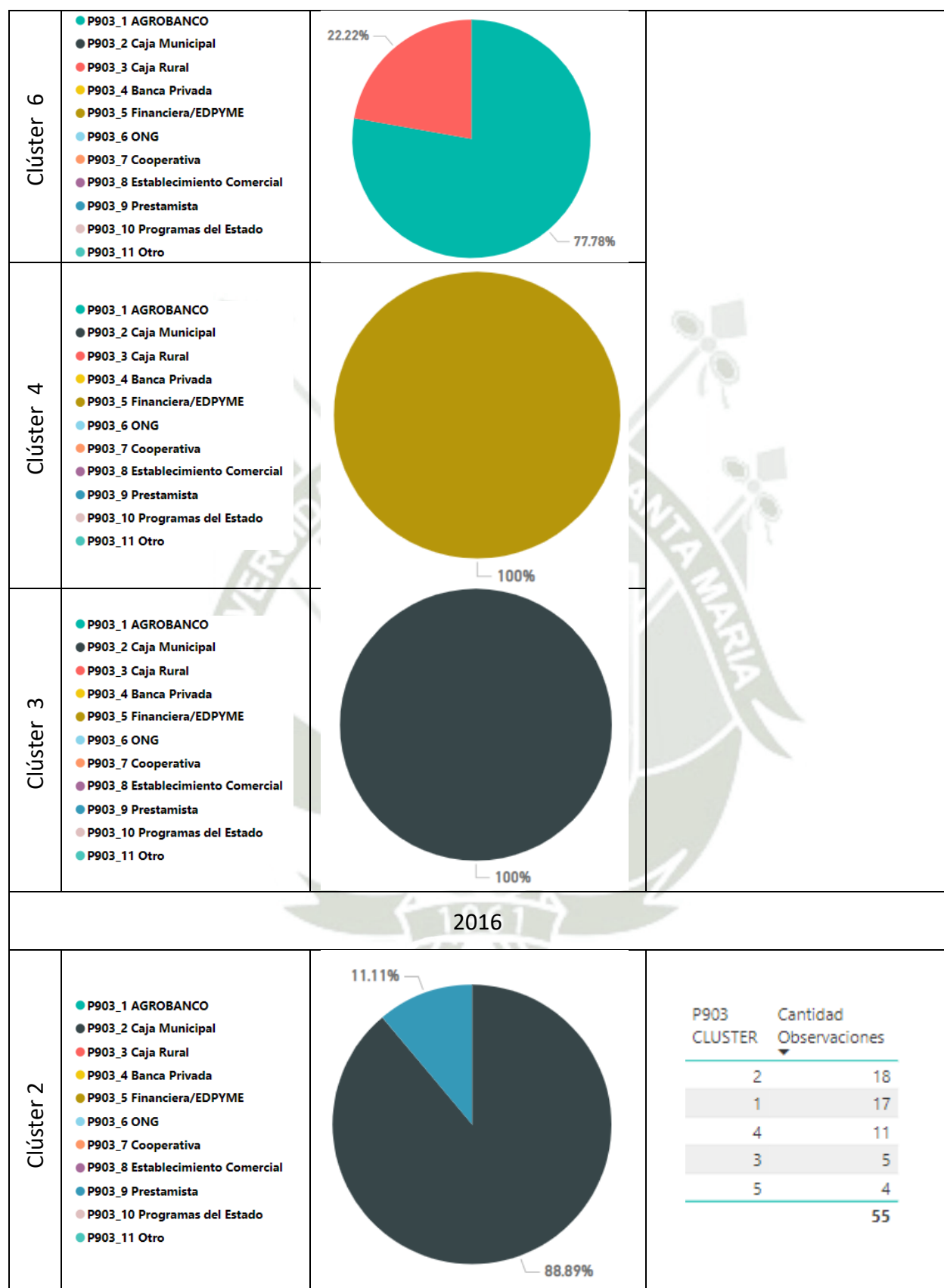


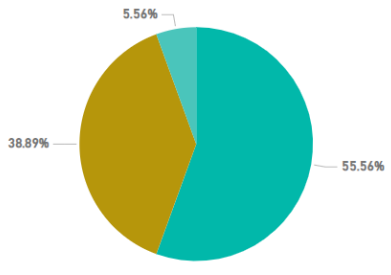
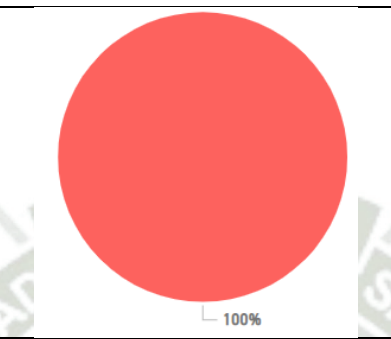
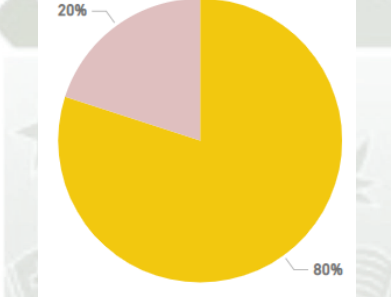
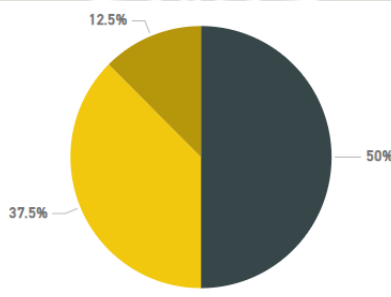
En el distrito de Paucartambo de 796 observaciones, se clasificaron 170, donde los principales fueron Caja Municipal, Banca Privada, Cara Rural, AGROBANCO, EDPYME, entre otros.

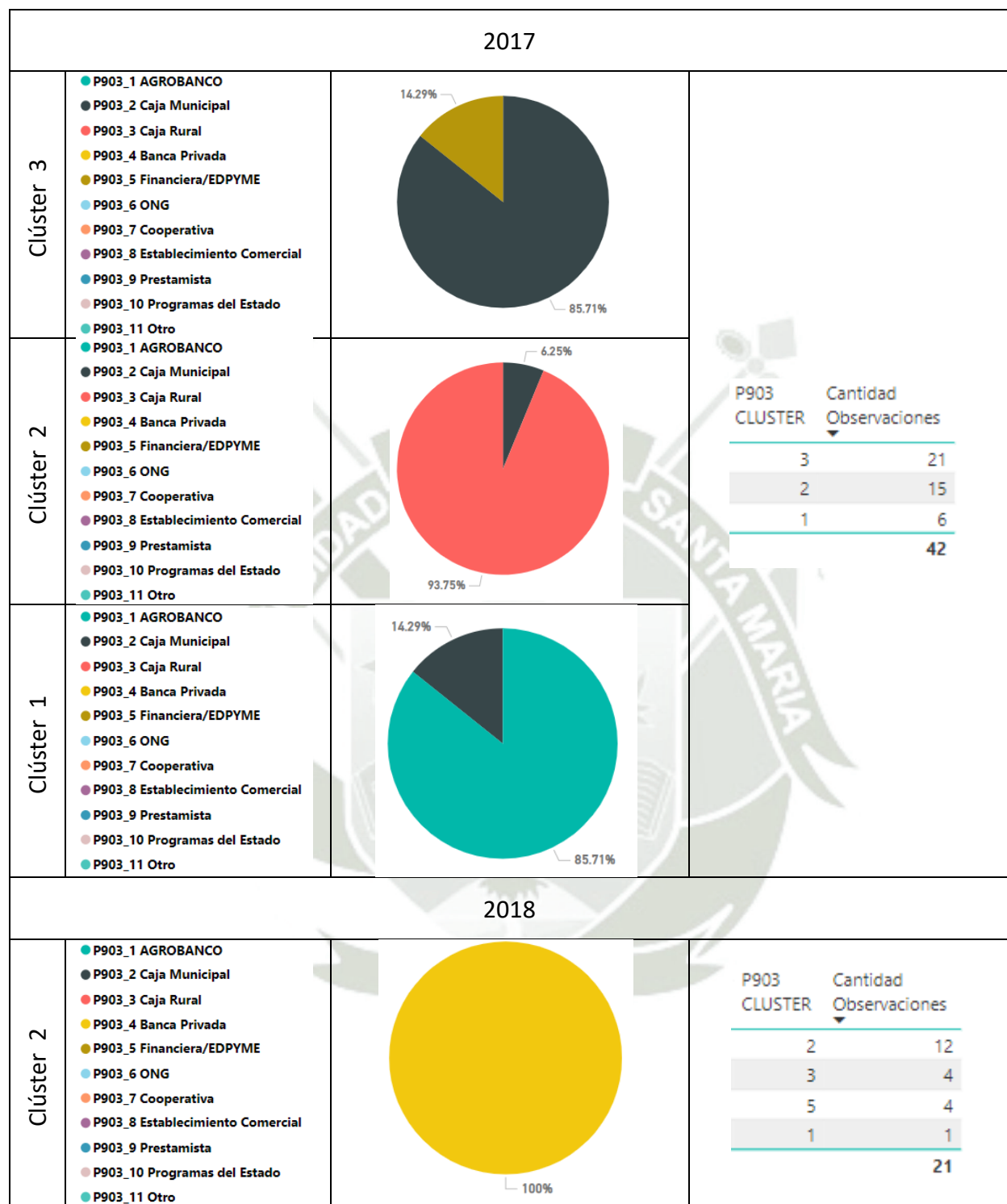
Figura 74.

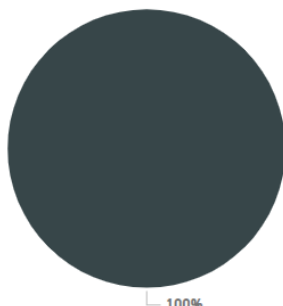
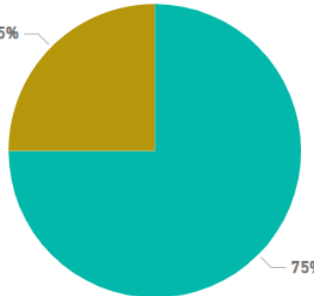
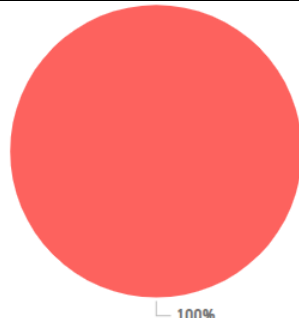
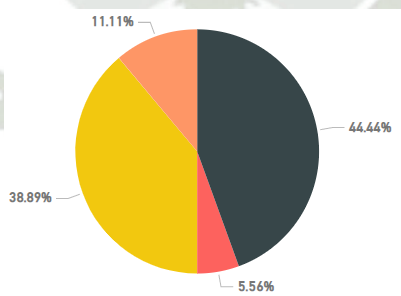
Características del distrito de Paucartambo.

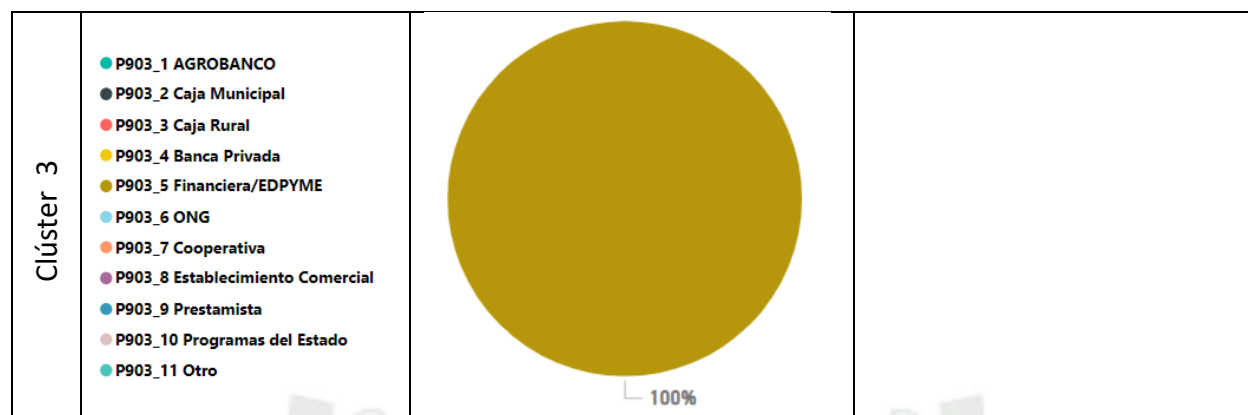




Clúster 1	- P903_1 AGROBANCO - P903_2 Caja Municipal - P903_3 Caja Rural - P903_4 Banca Privada - P903_5 Financiera/EDPYME - P903_6 ONG - P903_7 Cooperativa - P903_8 Establecimiento Comercial - P903_9 Prestamista - P903_10 Programas del Estado - P903_11 Otro		
Clúster 4	- P903_1 AGROBANCO - P903_2 Caja Municipal - P903_3 Caja Rural - P903_4 Banca Privada - P903_5 Financiera/EDPYME - P903_6 ONG - P903_7 Cooperativa - P903_8 Establecimiento Comercial - P903_9 Prestamista - P903_10 Programas del Estado - P903_11 Otro		
Clúster 3	- P903_1 AGROBANCO - P903_2 Caja Municipal - P903_3 Caja Rural - P903_4 Banca Privada - P903_5 Financiera/EDPYME - P903_6 ONG - P903_7 Cooperativa - P903_8 Establecimiento Comercial - P903_9 Prestamista - P903_10 Programas del Estado - P903_11 Otro		
Clúster 5	- P903_1 AGROBANCO - P903_2 Caja Municipal - P903_3 Caja Rural - P903_4 Banca Privada - P903_5 Financiera/EDPYME - P903_6 ONG - P903_7 Cooperativa - P903_8 Establecimiento Comercial - P903_9 Prestamista - P903_10 Programas del Estado - P903_11 Otro		



Clúster 3	- P903_1 AGROBANCO - P903_2 Caja Municipal - P903_3 Caja Rural - P903_4 Banca Privada - P903_5 Financiera/EDPYME - P903_6 ONG - P903_7 Cooperativa - P903_8 Establecimiento Comercial - P903_9 Prestamista - P903_10 Programas del Estado - P903_11 Otro	 100%								
Clúster 5	- P903_1 AGROBANCO - P903_2 Caja Municipal - P903_3 Caja Rural - P903_4 Banca Privada - P903_5 Financiera/EDPYME - P903_6 ONG - P903_7 Cooperativa - P903_8 Establecimiento Comercial - P903_9 Prestamista - P903_10 Programas del Estado - P903_11 Otro	 25% 75%								
Clúster 1	- P903_1 AGROBANCO - P903_2 Caja Municipal - P903_3 Caja Rural - P903_4 Banca Privada - P903_5 Financiera/EDPYME - P903_6 ONG - P903_7 Cooperativa - P903_8 Establecimiento Comercial - P903_9 Prestamista - P903_10 Programas del Estado - P903_11 Otro	 100%								
2019										
Clúster 4	- P903_1 AGROBANCO - P903_2 Caja Municipal - P903_3 Caja Rural - P903_4 Banca Privada - P903_5 Financiera/EDPYME - P903_6 ONG - P903_7 Cooperativa - P903_8 Establecimiento Comercial - P903_9 Prestamista - P903_10 Programas del Estado - P903_11 Otro	 11.11% 44.44% 38.89% 5.56%								
<table><tr><th>P903 CLUSTER</th><th>Cantidad Observaciones</th></tr><tr><td>4</td><td>18</td></tr><tr><td>3</td><td>7</td></tr><tr><td colspan="2">25</td></tr></table>			P903 CLUSTER	Cantidad Observaciones	4	18	3	7	25	
P903 CLUSTER	Cantidad Observaciones									
4	18									
3	7									
25										



Fuente: Elaboración Propia.

B. Variable P904: ¿Utilizó el crédito para?

La finalidad del crédito obtenido, en que lo emplean los productores su crédito obtenido.

Los créditos obtenidos por los productores fueron principalmente para comprar insumos agrícolas, pagar mano de obra, alquilar maquinaria, comprar insumos pecuarios entre otras.

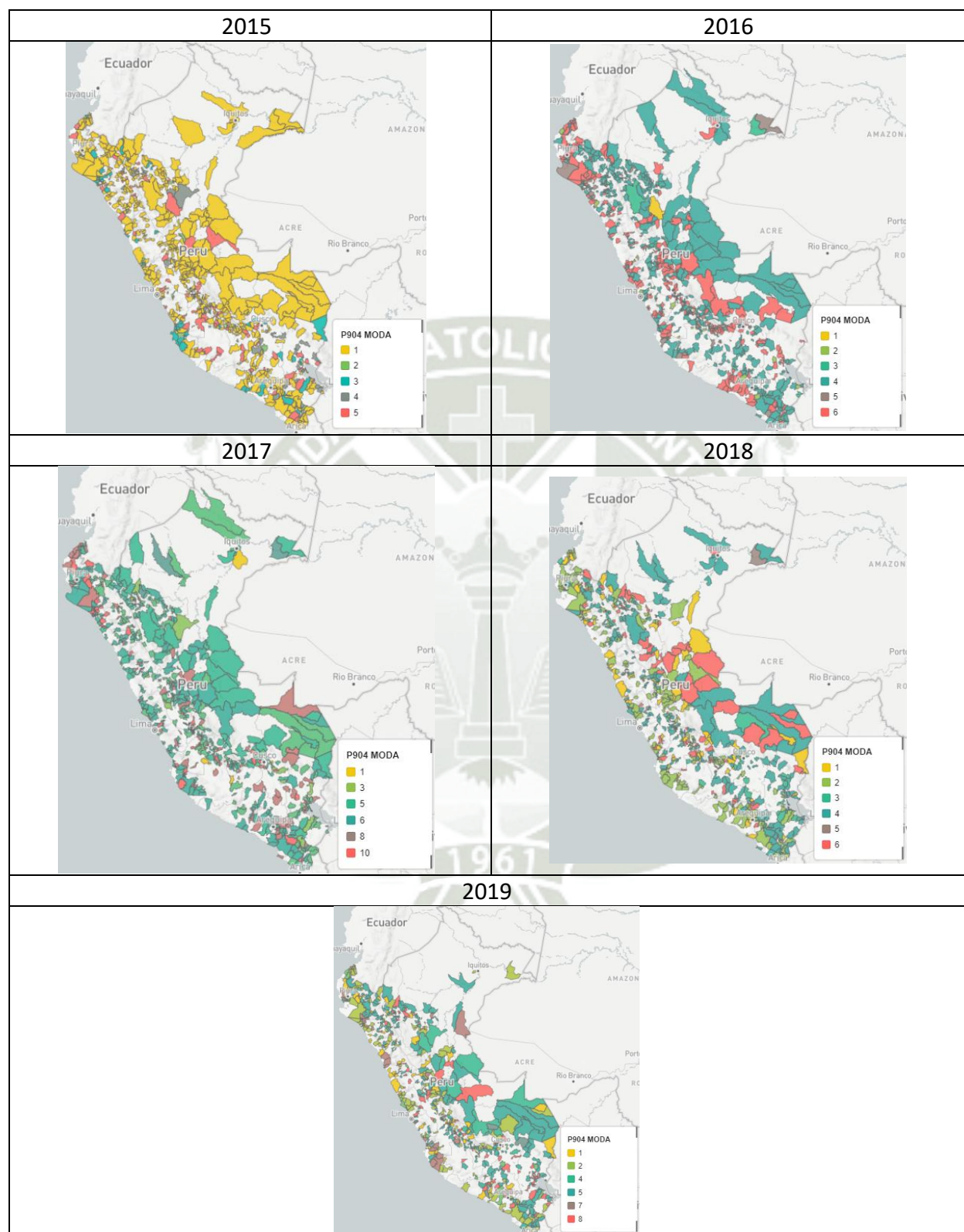
Figura 75.
Características de los clústeres en P904

AÑO	P904 CLUSTER	P904_1	P904_2	P904_3	P904_4	P904_5	P904_6	P904_7	P904_8	P904_9	Cantidad Observaciones
2015	1	1,524	0	742	1	0	0	16	87	657	2,409
2015	3	362	1	318	21	392	0	0	9	0	400
2015	5	109	224	21	1	17	87	1	4	0	310
2015	4	11	21	13	0	9	9	139	1	0	139
2015	2	123	104	103	4	26	25	14	0	0	125
2016	4	679	31	835	1	5	6	148	1	831	1,800
2016	6	1,233	214	120	11	102	0	1	0	0	1,233
2016	5	270	0	297	4	322	1	0	20	0	322
2016	2	0	162	39	1	10	1	5	2	0	162
2016	3	37	8	21	0	10	146	0	1	0	146
2016	1	28	3	23	0	0	0	0	75	0	75
2017	6	996	18	0	0	109	16	2	0	0	996
2017	5	659	2	822	0	3	0	0	46	0	838
2017	4	0	0	0	0	0	0	1	59	581	640
2017	10	339	50	354	0	371	0	0	5	0	371
2017	9	207	330	98	1	0	0	0	5	0	330
2017	7	16	4	20	0	7	96	0	1	0	96
2017	3	17	0	13	0	2	0	88	1	0	88
2017	1	9	31	8	0	2	0	31	0	0	31
2017	8	10	2	10	10	2	0	1	0	0	10
2017	2	2	2	2	0	0	5	5	0	0	5
2018	2	1,196	32	264	19	326	4	3	0	0	1,222
2018	4	68	134	0	5	0	0	0	25	650	812
2018	1	611	64	590	1	0	0	31	0	0	611
2018	6	0	23	196	0	1	96	0	0	0	285
2018	3	5	16	6	0	2	0	100	0	0	100
2018	5	25	6	9	0	0	27	2	0	0	27
2019	2	801	3	11	0	4	19	0	0	0	801
2019	5	0	80	0	2	0	0	0	0	573	653
2019	1	401	1	400	14	6	0	0	1	0	405
2019	7	208	0	175	0	231	0	2	0	0	231
2019	4	0	2	133	0	2	66	0	0	0	197
2019	3	139	141	56	0	27	0	4	1	0	141
2019	8	18	30	25	0	2	3	100	0	0	115
2019	6	27	2	22	0	16	0	0	57	0	57
Total		10,130	1,741	5,746	96	2,006	607	694	401	3,292	16,183

Fuente: Elaboración Propia.

En esta variable existen distritos geográficamente cercanos compartiendo las mismas características.

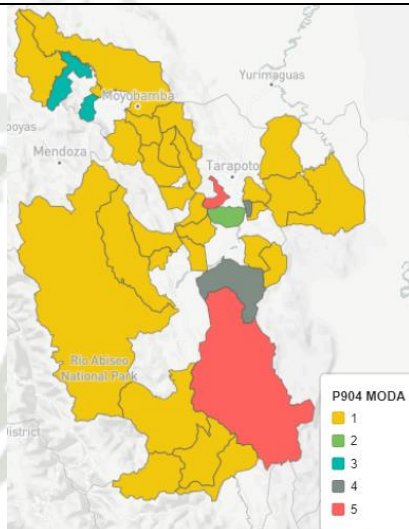
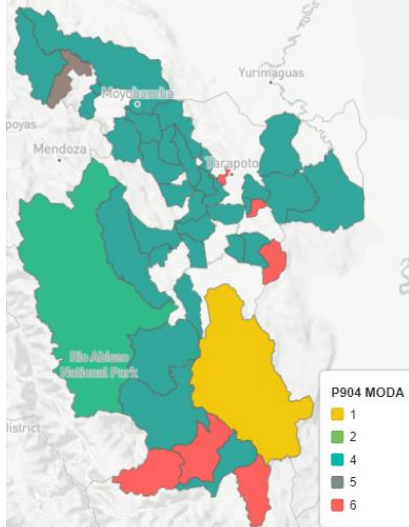
Figura 76.
Mapa de los clústeres P904.

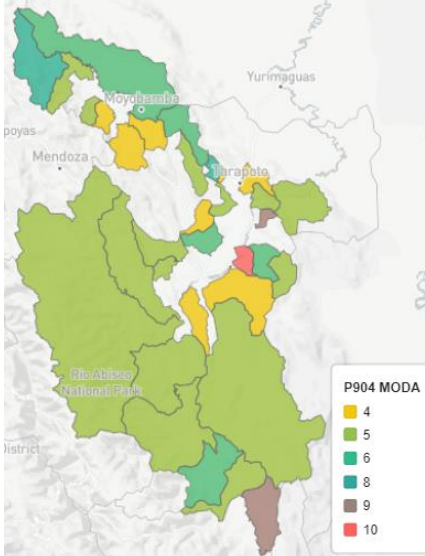
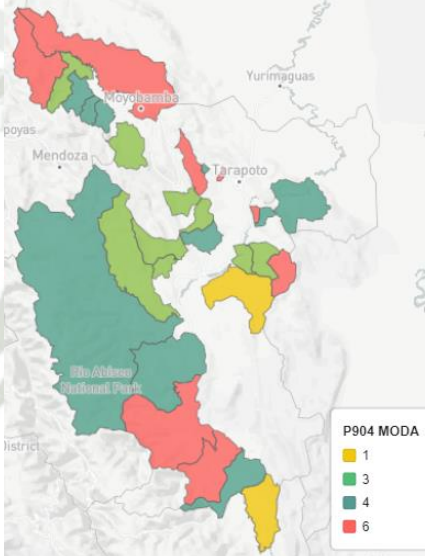
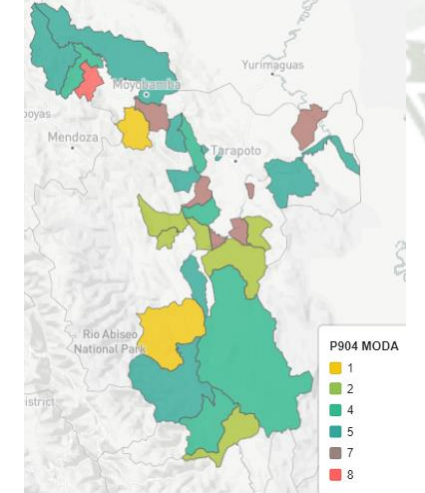


Fuente: Elaboración Propia.

En San Martín, se utilizó el crédito principalmente para otras actividades que no son agrícola ni pecuaria, después se utilizó para comprar insumos agrícolas, pagar mano de obra, compra de terrenos agrícolas, alquilar maquinaria, entre otros. Generalmente la mayor parte de distritos que conforman San Martín pertenecen al mismo clúster.

Figura 77.
Características del departamento de San Martín.

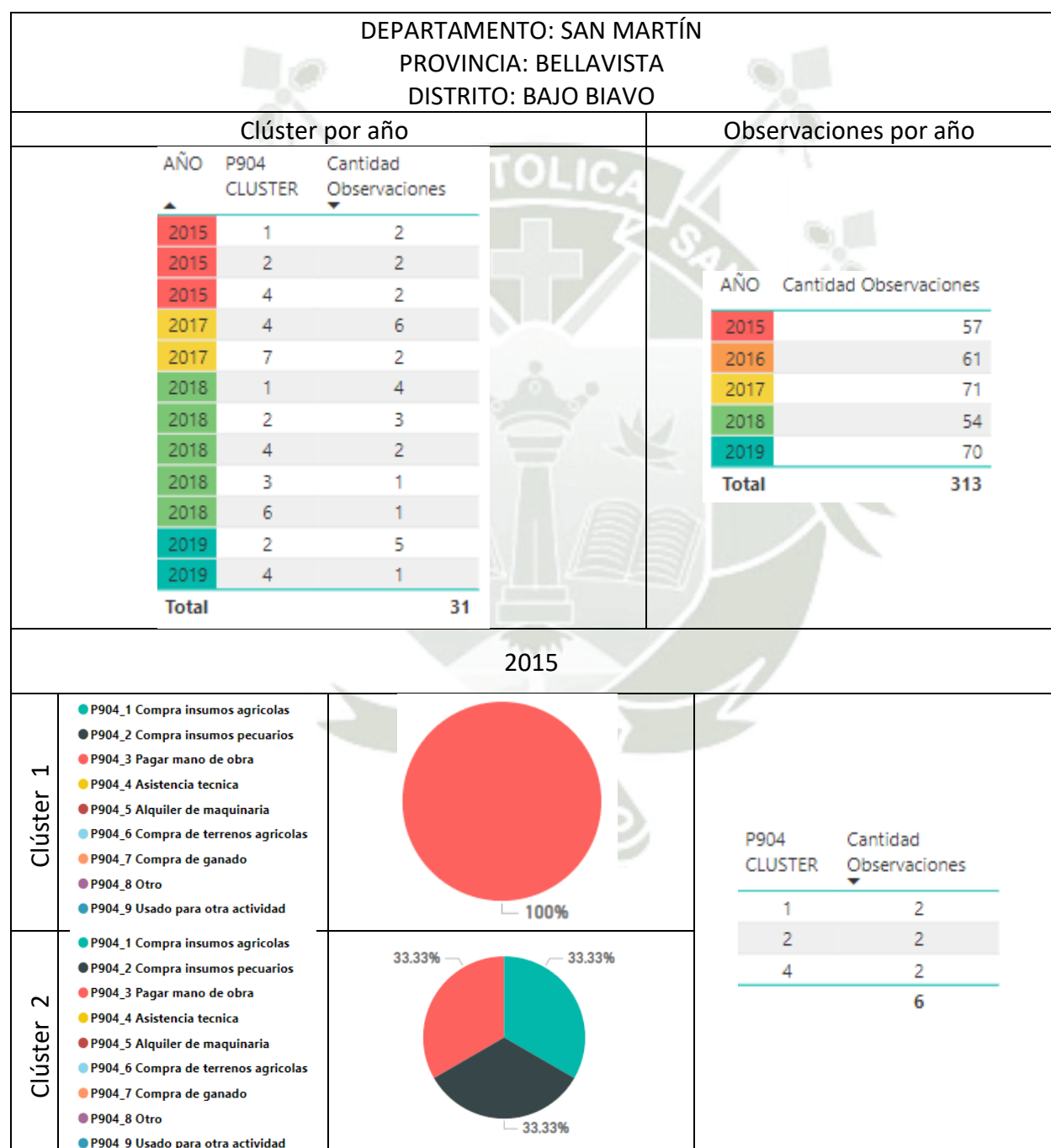
Año	Geografía	Cantidad de Observaciones																
2015		<table><thead><tr><th>P904 CLUSTER</th><th>Cantidad Observaciones</th></tr></thead><tbody><tr><td>1</td><td>107</td></tr><tr><td>3</td><td>12</td></tr><tr><td>5</td><td>12</td></tr><tr><td>2</td><td>9</td></tr><tr><td>4</td><td>9</td></tr><tr><td>Total</td><td>149</td></tr></tbody></table>	P904 CLUSTER	Cantidad Observaciones	1	107	3	12	5	12	2	9	4	9	Total	149		
P904 CLUSTER	Cantidad Observaciones																	
1	107																	
3	12																	
5	12																	
2	9																	
4	9																	
Total	149																	
2016		<table><thead><tr><th>P904 CLUSTER</th><th>Cantidad Observaciones</th></tr></thead><tbody><tr><td>4</td><td>115</td></tr><tr><td>6</td><td>37</td></tr><tr><td>3</td><td>15</td></tr><tr><td>5</td><td>13</td></tr><tr><td>1</td><td>7</td></tr><tr><td>2</td><td>5</td></tr><tr><td>Total</td><td>192</td></tr></tbody></table>	P904 CLUSTER	Cantidad Observaciones	4	115	6	37	3	15	5	13	1	7	2	5	Total	192
P904 CLUSTER	Cantidad Observaciones																	
4	115																	
6	37																	
3	15																	
5	13																	
1	7																	
2	5																	
Total	192																	

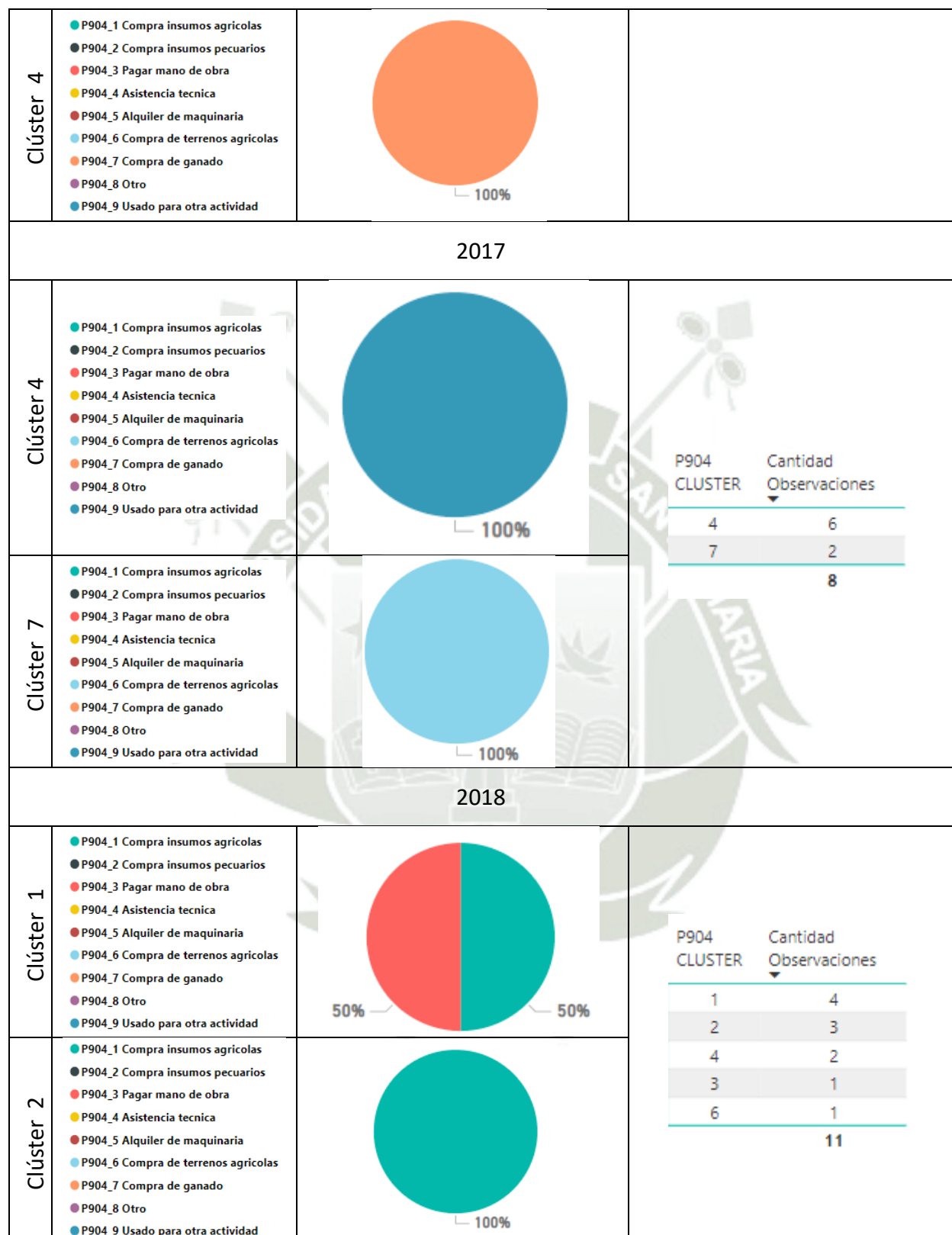
2017		<table><tr><th>P904 CLUSTER</th><th>Cantidad Observaciones</th></tr><tr><td>5</td><td>57</td></tr><tr><td>4</td><td>44</td></tr><tr><td>6</td><td>27</td></tr><tr><td>7</td><td>14</td></tr><tr><td>9</td><td>5</td></tr><tr><td>10</td><td>5</td></tr><tr><td>3</td><td>3</td></tr><tr><td>1</td><td>2</td></tr><tr><td>2</td><td>1</td></tr><tr><td>8</td><td>1</td></tr><tr><td>Total</td><td>159</td></tr></table>	P904 CLUSTER	Cantidad Observaciones	5	57	4	44	6	27	7	14	9	5	10	5	3	3	1	2	2	1	8	1	Total	159
P904 CLUSTER	Cantidad Observaciones																									
5	57																									
4	44																									
6	27																									
7	14																									
9	5																									
10	5																									
3	3																									
1	2																									
2	1																									
8	1																									
Total	159																									
2018		<table><tr><th>P904 CLUSTER</th><th>Cantidad Observaciones</th></tr><tr><td>2</td><td>33</td></tr><tr><td>4</td><td>32</td></tr><tr><td>1</td><td>29</td></tr><tr><td>6</td><td>24</td></tr><tr><td>3</td><td>4</td></tr><tr><td>5</td><td>1</td></tr><tr><td>Total</td><td>123</td></tr></table>	P904 CLUSTER	Cantidad Observaciones	2	33	4	32	1	29	6	24	3	4	5	1	Total	123								
P904 CLUSTER	Cantidad Observaciones																									
2	33																									
4	32																									
1	29																									
6	24																									
3	4																									
5	1																									
Total	123																									
2019		<table><tr><th>P904 CLUSTER</th><th>Cantidad Observaciones</th></tr><tr><td>5</td><td>39</td></tr><tr><td>2</td><td>35</td></tr><tr><td>4</td><td>25</td></tr><tr><td>1</td><td>13</td></tr><tr><td>7</td><td>11</td></tr><tr><td>3</td><td>2</td></tr><tr><td>6</td><td>2</td></tr><tr><td>8</td><td>2</td></tr><tr><td>Total</td><td>129</td></tr></table>	P904 CLUSTER	Cantidad Observaciones	5	39	2	35	4	25	1	13	7	11	3	2	6	2	8	2	Total	129				
P904 CLUSTER	Cantidad Observaciones																									
5	39																									
2	35																									
4	25																									
1	13																									
7	11																									
3	2																									
6	2																									
8	2																									
Total	129																									

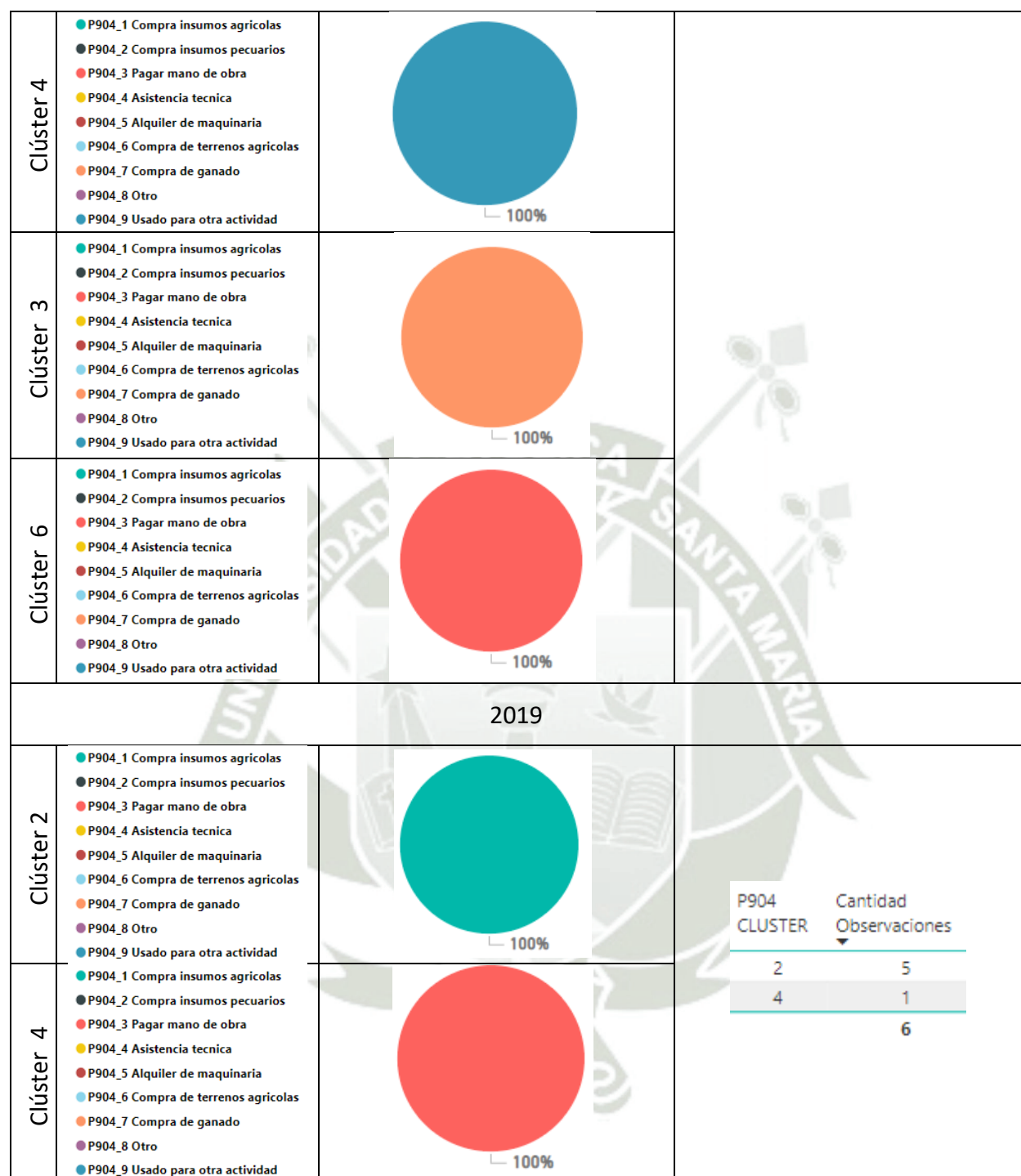
Fuente: Elaboración Propia.

En este distrito de 313 observaciones, se clasificaron 31, del año 2015 al 2019, los productores utilizaron el crédito generalmente para otro tipo de actividades independientes, también para la compra de insumos agrícolas, pagar mano de obra, compra de ganado, compra de terrenos y compra de insumos pecuarios.

Figura 78.
Características del distrito de Bajo Biavo.







Fuente: Elaboración Propia.

3.2. CONOCIMIENTO PARA LA TOMA DE DECISIONES

Al final de este proceso hemos logrado crear grupos de productores que comparten características similares entre sí, con esta información podemos ver conglomerados de agricultores que pueden padecer de similares factores que echan a perder toda la producción, donde son los principales puntos donde se venden los productos, motivos por los cuales se cultiva determinado producto, donde se destina la venta de ganado, lo que se hace con los residuos de ganados, información que los agricultores necesitan para poder producir productos agropecuarios, los principales canales por donde los agricultores reciben información, de donde obtienen sus créditos y para que lo utilizan, esta información se puede aprovechar con muchos fines, como identificar problemas, proponer soluciones a problemas, ver oportunidades de negocio, ofrecer productos y servicios de acuerdo a las necesidades, entre otras.

A continuación, se seleccionarán los clústeres más resaltantes de cada módulo, para poder aprovechar el conocimiento y ver las posibilidades de toma de decisiones con diferentes fines.

3.2.1. Destino de los productos cosechados

En este módulo tenemos clústeres de los agricultores agrupados de acuerdo al lugar donde venden sus productos agropecuarios, en los cinco años existen clústeres similares, los dos más comunes son, el clúster más grande que viene a ser el de productores que únicamente venden en el mercado local, y otro clúster más pequeño de productores que venden tanto en el mercado local como regional, también existen clústeres pero que son más pequeños respecto a los antes mencionados, y representan una minoría que se caracterizan por vender tanto en el mercado exterior y agroindustria, o agroindustrias y mercado de Lima conjuntamente, bajo esta premisa podemos darnos cuenta que los puntos de comercialización no son muy variados.

En ese sentido e identificando los grupos de agricultores y canales de comercialización, es necesario poner atención y enfocarnos en los sistemas de comercialización, de la mano del sector público y colaboración entre los interesados en el sector agropecuario, para permitir que los sectores agrarios tengan una vinculación menos desventajosa con el mercado, ofreciéndole mejores condiciones de comercialización y mejor asociación entre productores agropecuarios, comerciantes mayoristas, minoristas, agroindustriales, entre otros.

3.2.2. Pérdida total de producción

En este módulo veremos las principales causas de pérdida de producción causados por efectos climáticos. Existen varios clústeres, pero uno de los más comunes, es caracterizado por productores que pierden la totalidad de sus cosechas a causa de un conjunto de factores que están asociados entre sí como, heladas, sequías, granizadas y plagas, estos factores están asociados ya que una causa de las heladas y sequías podrían provocar plagas, y generalmente las heladas vienen acompañadas de granizadas.

Teniendo en cuenta estos factores que afectan a cierto segmento de la población del Perú, e identificando geográficamente los distritos afectados, se puede tomar acciones por parte de las entidades públicas como municipios, para contrarrestar los efectos y minimizar las pérdidas de los cultivos de su localidad, creando estrategias de prevención, realizando campañas de capacitación, incentivando métodos para proteger los cultivos del frío como riego por aspersión, acolchados térmicos, cuidado y limpieza de suelos, y también trabajando de manera colaborativa con entidades privadas para el abastecimiento de productos protectores, que se aplican en las plantas antes de la baja de temperaturas, y su efecto protector beneficia en gran medida a resistir las temperaturas de hasta menos 5° C, y su efecto podría durar hasta 6 semanas.

3.2.3. Producción pecuaria

Este módulo es parecido al módulo de destino de productos agropecuarios, de entre todos los clústeres dos son los que tienen mucha más cantidad de productores, el primer clúster es de productores que únicamente venden su ganado en el lugar donde lo crían, y el segundo clúster es el cual donde los productores lo venden tanto en el lugar donde crían y en el mercado local, siendo también un mercado de comercialización poco variado. Viendo esta situación se podría mejorar los canales de comercialización con una mejor red de comercio que integre más a los productores y compradores.

3.2.4. Inocuidad agroalimentaria

Los clústeres en este módulo son acerca de lo que hacen los agricultores con los residuos de las cosechas y animales, tomando como ejemplo un clúster que se caracteriza por productores que queman o botan los desperdicios de sus productos, en dicho clúster podemos identificar problemáticas y oportunidades de desarrollo. En una noticia publicada por la Sociedad Peruana de Derecho Ambiental en 2022, señala que el 95% de incendios forestales, son causados por la

quema indiscriminada de residuos agrícolas, frente a esto las entidades del gobierno pueden tomar decisiones para poder evitar la quema de estos residuos y plantear estrategias de tratamiento de residuos, generando oportunidades de desarrollo, como por ejemplo utilizando estos residuos para la obtención de biocombustibles sólidos, como carbón vegetal, briquetas u otros derivados que pueden ser explotados.

3.2.5. Servicios de extensión agraria

La toma de decisiones en este módulo, pueden beneficiar para ofrecer servicios más especializados de acuerdo a las necesidades de información que necesitan los productores, tanto entidades públicas como empresas privadas podrían aprovechar esta información para ofrecer campañas de capacitación y servicio especializados dirigidos de acuerdo a cómo se caracteriza cada clúster. Y además ofrecer sus servicios teniendo en cuenta cómo se caracterizan los productores de acuerdo a los canales de difusión de información a los que tienen más acceso.

3.2.6. Servicios financieros

Según una publicación de Wiener (2021), señala que la penetración del sistema financiero en la actividad agraria rural ha sido persistentemente baja, el número de productores excluidos pero interesados en obtener crédito y otros servicios, excede largamente a los que accedieron, algo que podemos observar en los clústeres P903, con la información generada en esta investigación se pueden observar los conglomerados de productores que accedieron a créditos financieros, y una gran cantidad de productores que no fueron tomados en cuenta en la creación de estos clústeres, asumiendo que no accedieron a créditos.

La presente información puede ser aprovechada para la toma de decisiones y poder atender eficazmente las necesidades financieras de los agricultores, por otra parte, se puede optimizar esta información para la evaluación de riesgos de los diferentes grupos identificados, otorgando microfinanzas urbanas a los pequeños y medianos productores del sector agrario en el ámbito rural.

CONCLUSIONES

1. Se logró clusterizar el sector agropecuario teniendo en cuenta la productividad, la economía y las prácticas de desarrollo. Como resultado, se crearon perfiles que reflejan las características más representativas de cada distrito. Estos perfiles agrupan aquellos distritos que comparten características semejantes, proporcionando así una visión más clara y específica de la realidad de los territorios agrícolas.
2. Se logró transformar los datos proporcionados por el INEI en conocimiento mediante un proceso de clustering implementado en Python. Posteriormente, se crearon dashboards utilizando Power BI para cada módulo, donde se pueden observar las características de cada distrito y la distribución geográfica de los clústeres en el territorio peruano.
3. Se seleccionó un grupo de módulos, cada uno contiene diferentes variables de investigación pertenecientes al sector agropecuario, de acuerdo a las características de las variables se aplicó un algoritmo de clustering, con su medida de distancia y método de clustering más ideal para la obtención de clústeres.
4. Al haber realizado este proceso de clusterización en todo el país, se obtuvo información de todos los departamentos, provincias y distritos del todo el Perú, esta información puede ser aprovechada por las entidades públicas como municipalidades o privadas como empresas relacionadas al sector, para conocer las necesidades, preferencias e intereses comunes de los productores agropecuarios.
5. A partir del análisis descriptivo, se pudo ver las necesidades de cada distrito y de esta manera proponer mejoras y buscar oportunidades de desarrollo que beneficien a todos los interesados que conforman el sector agropecuario.

RECOMENDACIONES

1. Se debe promover la formación de clústeres regionales, provinciales y distritales para poder atender las necesidades de las provincias de forma directa sin desperdiciar recursos.
2. Al tener información disponible del INEI acerca de diferentes sectores, las entidades públicas y privadas deberían aprovechar mejor dicha información para generar mejoras, aprovechando también las nuevas tecnologías de análisis de datos open source.
3. Aprovechar el potencial de la ciencia de datos, para poder gestionar mejor la información, y de esta manera poder contar con información confiables para la toma de decisiones, optimización del sector agropecuario y la reducción de riesgos.
4. Las entidades públicas del Perú como las municipalidades deberían apostar por la aplicación de las tecnologías de información, para poder atender las necesidades de su población basado en fuentes confiables y reducir riesgos al tomar decisiones.
5. Seguir promoviendo las actividades de difusión de investigaciones estadísticas, para poner a disposición del público general y así poder generar más investigaciones en más sectores, para favorecer a la población peruana.

REFERENCIAS

- Ahumada, E., & Perusquia, J. M. A. (2016). Inteligencia de negocios: Estrategia para el desarrollo de competitividad en empresas de base tecnológica. *Contaduría y Administración*, 61(1), 127–158. <https://doi.org/10.1016/j.cya.2015.09.006>
- Álvarez, A. C. (2012). Aplicación De Técnicas De Minería De Datos Para Mejorar El Proceso De Control De Gestión En Entel. Universidad De Chile Facultad De Ciencias Físicas Y Matemáticas Departamento De Ingeniería Industrial, 1–120. http://repositorio.uchile.cl/bitstream/handle/2250/112065/cf-martinez_ca.pdf?sequence=1
- Basic Farm. (2020). Basic Farm. Obtenido de <https://basicfarm.com/blog/que-es-inocuidad-alimentaria-importancia/>
- Benítez, I. J., & Díez, L. J. (2005). *Técnicas de Agrupamiento para el Análisis de Datos Cuantitativos y Cualitativos* (Issue September). Universidad Politécnica de Valencia.
- Camana, R. G. (2016). Potenciales aplicaciones de la minería de datos en Ecuador. *Potenciales Aplicaciones de La Minería de Datos En Ecuador*, 29(Julio), 170–183. <http://200.10.150.204/index.php/tecnologica/article/view/464/338>
- Capó-Vicedo, J., Expósito-Langa, M., & Masiá-Buades, E. (2007). La importancia de los clústeres para la competitividad de las PYME en una economía global. *Eure*, 33(98), 119–133. <https://doi.org/10.4067/s0250-71612007000100007>
- Cendejas, J. L., Acuña, M. A., Cortés, G., & Bolaños, G. (2016). El uso de modelos y metodologías de minería de datos para la inteligencia de negocios. *Revista de Sistemas Experimentales*, 3(2410–3950), 1–97. http://www.ecorfan.org/spain/researchjournals/Sistemas_Computacionales_y_TICs/vol3num7/Revista_de_Sistemas_Computacionales_y_TIC%60S_V3_N7.pdf
- Chiang, W. Y. (2018). Applying data mining for online CRM marketing strategy: An empirical case of coffee shop industry in Taiwan. *British Food Journal*, 120(3), 665–675. <https://doi.org/10.1108/BFJ-02-2017-0075>
- Choudhary, A. K., Tiwary, M. K., & Harding, J. A. (2006). Minería de datos en la fabricación: una revisión basada en el tipo de conocimiento. *Proceedings of the 4th IEEE International*

Conference on Industrial Informatics, 128(3), 57–71.
<https://core.ac.uk/download/pdf/288384686.pdf>

Cruz, A., Usaquén, S., Vanegas, N., & Lopera, C. (2010). Aplicación de técnicas de clustering para caracterizar proveedores de servicios de mantenimiento. *Revista Salud Pública*, 12(3),

52. <http://www.scielo.org.co/pdf/rsap/v12n3/v12n3a12.pdf>

Delgado, E., Rivera, J., & Hernández, A. (2006). Reducción de dimensiones para clasificación de datos multidimensionales usando medidas de información. *Scientia et Technica*, 3(32), 181–186. <https://doi.org/10.22517/23447214.6235>

Dierckens, K. E., Harrison, A. B., Leung, C. K., & Pind, A. V. (2017). A data science and engineering solution for fast k-means clustering of big data. *Proceedings - 16th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, 11th IEEE International Conference on Big Data Science and Engineering and 14th IEEE International Conference on Embedded Software and Systems*, 925–932. <https://doi.org/10.1109/Trustcom/BigDataSE/ICESS.2017.332>

Duque, N., Hernández, E., Pérez, Á., Arroyave, A., & Espinosa, A. (2016). Modelo para el proceso de extracción, transformación y carga en bodegas de datos. Una aplicación con datos ambientales. *Ciencia e Ingeniería Neogranadina*, 26(2), 95–109. <https://doi.org/10.18359/rcin.1799>

García, J. A. (2016). Líneas de investigación en minería de datos en aplicaciones en ciencia e ingeniería: Estado del arte y perspectivas. *Arxiv, Artificial Intelligence (Cs.AI)*, 1(1), 17. <https://arxiv.org/ftp/arxiv/papers/1609/1609.05401.pdf>

García, J., & Molina, J. M. (2012). *Técnicas de análisis de datos*. http://marmota.dlsi.uji.es/WebBIB/papers/2007/1_Pascual-MIA-2007.pdf

Gironés, J., Casas, J., Minguillón, J., & Caihuelas, R. (2017). Minería de datos: Modelos y Algoritmos. In *Editorial UOC* (Primera ed, Issue 1). https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part

- Gómez, A. A. R., & Bautista, D. W. R. (2010). Inteligencia de negocios: Estado del arte. *Scientia Et Technica*, 1(44), 321–326.
https://www.researchgate.net/publication/277231717_Inteligencia_de_negocios_Estado_del_arte
- Honrado, K. (2019). *Inteligencia de negocio/Conceptualización*. 1–4.
<https://sci2s.ugr.es/sites/default/files/files/Teaching/GraduatesCourses/InteligenciaDeNegocio/Curso18-19/Tema01-Introduccion a la Inteligencia de negocio 2018-19.pdf>
- INEI. (2021). Ficha técnica. *Encuesta Nacional Agropecuaria 2021*.
<http://iinei.inei.gob.pe/iinei/srienaho/Descarga/FichaTecnica/701-Ficha.pdf>
- Janas, T., Kuczera, J., & Chojnacki, T. (1990). Algoritmo de clustering basado en entropía para descubrir grupos en atributos de tipo mixto. In *Chemistry and Physics of Lipids* (Vol. 52, Issue 2, pp. 151–155). [https://doi.org/10.1016/0009-3084\(90\)90158-N](https://doi.org/10.1016/0009-3084(90)90158-N)
- Karakatsanis, I., AlKhader, W., MacCrory, F., Alibasic, A., Omar, M. A., Aung, Z., & Woon, W. L. (2017). Data mining approach to monitoring the requirements of the job market: A case study. *Information Systems*, 65, 1–6. <https://doi.org/10.1016/j.is.2016.10.009>
- Karypis, G., Han, E. H., & Kumar, V. (1999). Chameleon: Hierarchical clustering using dynamic modeling. *Computer*, 32(8), 68–75. <https://doi.org/10.1109/2.781637>
- Lalanne, C. (2016). Statistical Data Analytics. Foundations for Data Mining, Informatics, and Knowledge Discovery. In *Journal of Statistical Software* (Vol. 69, Issue Book Review 3). <https://doi.org/10.18637/jss.v069.b03>
- Liu, L., Peng, Z., Wu, H., Jiao, H., Yu, Y., & Zhao, J. (2018). Fast identification of urban sprawl based on K-means clustering with population density and local spatial entropy. *Sustainability (Switzerland)*, 10(8). <https://doi.org/10.3390/su10082683>
- Loya8, T., & Carden, G. (2018). Business intelligence and analytics. In *Higher Education Strategy and Planning*. <https://doi.org/10.4324/9781315206455-12>

- Mamani, Z. (2015). Aplicación de la minería de datos distribuida usando algoritmo de clustering K-Means para mejorar la calidad de servicios de las organizaciones modernas [Universidad Mayor de San Marcos]. <https://hdl.handle.net/20.500.12672/4472>
- Marín, H. (2014). *Minería de Datos*. Universidad Politécnica de Victoria. <https://www.tamps.cinvestav.mx/~hmarin/Mineria/EC2.pdf>
- Martínez, L. (2011). Minería de datos como hallar una aguja en un pajar. *Ciencia*, 18–28. Michael, B., & Gordon, L. (2011). Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. In *Wiley Publishing* (Second Edi).
- Miloudi, S., Wang, Y., & Ding, W. (2021). A Gradient-Based Clustering for Multi-Database Mining. *IEEE Access*, 9, 11144–11172. <https://doi.org/10.1109/ACCESS.2021.3050404>
- MINAGRI. (2019). *Problemáticas en la agricultura peruana*. Ministerio de Desarrollo Agrario y Riego. <https://www.minagri.gob.pe/portal/22-sector-agrario/vision-general/190-problemas-en-la-agricultura-peruana>
- MinSalud. (s.f.). [minsalud.gov.co](https://www.minsalud.gov.co). Obtenido de <https://www.minsalud.gov.co/salud/Paginas/inocuidad-alimentos.aspx#:~:text=%E2%80%8B%E2%80%8BLa%20inocuidad%20de,un%20riesgo%20para%20la%20salud.&text=consumo>.
- Moine, J. Mi., Haedo, A., & Gordillo, S. (2011). Estudio comparativo de metodologías para minería de datos. *XIII Workshop de Investigadores En Ciencias de La Computación*, 278–281. <http://sedici.unlp.edu.ar/handle/10915/20034>
- Morales, L., Velazco, L., & Perez, S. (2014). Estrategias para la formación de clusters agrícolas en zonas rurales: El caso de la Región San Quintín y su Valle. 20.
- Muñoz, H., Osorio, R. C., & Zúñiga, L. M. (2016). Inteligencia de los negocios. Clave del Éxito en la era de la información. *Clío América*, 10(20), 194. <https://doi.org/10.21676/23897848.1877>
- Nogués, A., & Valladares, J. (2017). Business Intelligence Tools for Small Companies. In *Business Intelligence Tools for Small Companies*. <https://doi.org/10.1007/978-1-4842-2568-4>

- Obando Velásquez, D. A. (2017). Estudio y análisis de entornos comerciales mediante la evaluación, comparación y experimentación de algoritmos de minería de datos. *Universidad Católica de Santa María - UCSM*. <http://renati.sunedu.gob.pe/handle/sunedu/116167>
- Olivera, L. (2020). *Retos y oportunidades de la agroindustria en el Perú*. Retos y Oportunidades de La Agroindustria En El Perú. <https://facultades.usil.edu.pe/ingenieria/retos-y-oportunidades-de-la-agroindustria-en-el-peru/>
- Pascual, D., Pla, F., & Sánchez, S. (2007). Algoritmos de agrupamiento. *Universidad de Oriente*, 164–174. http://marmota.dlsi.uji.es/WebBIB/papers/2007/1_Pascual-MIA-2007.pdf
- RedAgrícola. (2019). *Perú en vías a transformarse en una potencia agroalimentaria - Redagrícola Perú*. RedAgrícola. <https://www.redagricola.com/pe/peru-vias-transformarse-una-potencia-agroalimentaria/>
- Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de datos Conceptos y tendencias. *Revista Iberoamericana de Inteligencia Artificial*, 29, 11–18. <https://www.redalyc.org/pdf/925/92502902.pdf>
- Rodriguez, M. (2006). *Clasificación de series de tiempo por minería de datos* (Vol. 4, Issue 3). <http://marefateadyan.nashriyat.ir/node/150>
- Salvador, C. A. (2018). Descripción del procedimiento metodológico del análisis clúster no jerárquico en el algoritmo CLARANS. *Universidad Agraria La Molina*. Universidad Agraria La Molina.
- Sandoval, L. (2018). Algoritmos de aprendizaje automático para análisis y predicción de datos. *Revista Tecnológica*, 11, 36–40. http://www.redicces.org.sv/jspui/bitstream/10972/3626/1/Art6_RT2018.pdf
- Siegel, E. (2013). Predictive Analytics: The power to predict who will click, buy, lie or die. *Journal of Chemical Information and Modeling*, 01(01), 1689–1699. <https://mirtech.ir/wp-content/uploads/2018/04/Predictive-Analytics.pdf>
- Strickland, J. S. (2012). Predictive Analytics using R. In *International Journal of Computer, Electrical, Automation, Control and Information Engineering* (Vol. 6, Issue 12).

- T. Larose, D., & D. Larose, C. (2015). *Data Mining and Predictive Analytics* (Vol. 4, Issue 3).
<http://marefateadyan.nashriyat.ir/node/150>
- Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*, 4(2), 437. <https://doi.org/10.30865/mib.v4i2.2080>
- Vallejo, D. (2016). *Clustering de Documentos con Restricciones de Tamaño*.
<https://riunet.upv.es/handle/10251/69089>
- Wilford, I. (2006). Minería de datos: Herramienta de apoyo en la selección de equipos de proyectos informáticos. *Ingeniería Industrial*, XXVII (2–3), 7–10.
- Yogping, T., & Zizhen, P. (2020). A Customer Group Mining Method Based on Cluster Analysis. *Lecture Notes in Electrical Engineering*, 590, 390–396.
https://doi.org/https://ezproxy.ucsm.edu.pe:2209/10.1007/978-981-32-9244-4_50
- Yu, S., Yang, M., Wei, L. H., Hu, J. S., Tseng, H. W., & Meen, T. H. (2020). Combination of self-organizing map and k-means methods of clustering for online games marketing, *Sensors and Materials*, 32(8 p2), 2697–2707. <https://doi.org/10.18494/SAM.2020.2800>
- Zamora, T. F. (2018). Aplicación De Técnicas De Minería De Datos Para Pronósticos Del Sector Agrícola. *Pontificia Universidad Católica De Valparaíso*.
- Zayas, B. I., Zayas, B. R. A., & Heiras, P. M. de J. (2017). CLUSTER AGRÍCOLA COMO FACTOR DE DESARROLLO EN LAS EMPRESAS Agricultural cluster development as a factor in companies in the. 533.



ANEXO A. Plan de Tesis

1. PLANTEAMIENTO DE LA INVESTIGACION

1.1. Planteamiento del Problema

El sector agropecuario es uno de los sectores que más crecimiento ha generado en el Perú en los últimos años, lo cual ha generado que nuestro país se consolide como un exportador regional y mundial de frutas y hortalizas, gracias a esto logro posicionarse como un importante participante del mercado global (Olivera, 2020). Para este año 2021, se espera que exportaciones del sector agroindustrial lleguen a US\$ 10, 000 millones, ya que en los últimos años las agroexportaciones registraron un crecimiento promedio del 14.9% anual, según Promperu (RedAgrícola, 2019).

Este crecimiento es muy importante para el motor de la economía del país, ya que genera empleo y oportunidades de inversión, no solo en las agroexportaciones sino también en muchos otros sectores que se ven influenciados directamente por esta (RedAgrícola, 2019).

A pesar del aporte tan importante a la economía que representa el sector agroindustrial, no podemos dejar de tener presente las problemáticas que representan una limitante en el desarrollo de la agricultura en el Perú. Según el ministerio de agricultura y riego la actividad agraria se caracteriza por el desorden en la producción y la disminución de su rentabilidad y competitividad, también los procesos post cosecha y de mercadeo están sumamente desordenados por la falta de una infraestructura vial y la ausencia de un sistema de mercadeo de mayoristas, lo cual afecta en los altos costos de comercialización que afectan a los productores agrarios. Por otro lado, los problemas estructurales que se presentan causados por la geografía agreste y heterogénea, el tema de financiamiento agrario que representa numerosos retos, el acceso a créditos, entre otros (MINAGRI, 2019).

También se debe tener en cuenta el amplio y variado sector agrícola ya que, debido a la alta cantidad de factores y características que influyen el desarrollo de este sector como las unidades agropecuarias, los tipos de cultivos, las buenas prácticas agrícolas, la producción pecuaria, las buenas practicas pecuarias, la inocuidad agroalimentaria, servicios de extensión agraria, servicios financieros, entre otros (INEI, 2019). Debido a esto vemos la necesidad de

afinar la segmentación del sector agrícola, para ya no verlo como una unidad sino como un sector dividido y organizado para enfocarnos eficientemente en las necesidades de cada grupo generado.

Muchas organizaciones están promoviendo la “Iniciativa de Clústeres” como un esfuerzo organizado para incrementar la competitividad mediante un trabajo colaborativo entre los interesados, ya que ofrece numerosas ventajas que facilitan el desarrollo industrial y económico, facilitan la comercialización, proporciona mejoras en los procesos y capacidades de las organizaciones, beneficia la interacción entre productores, proveedores y usuarios, y además facilita la acción colectiva del conjunto de interesados para lograr metas comunes (Morales et al., 2014).

1.2. Objetivos de la Investigación

1.2.1. General

Analizar el sector agropecuario a través de la clusterización basada en la productividad, economía y prácticas de desarrollo del sector en el Perú durante el periodo 2014 al 2019

1.2.2. Específicos

- Identificar las bases de datos del sector agropecuario y aplicar la extracción, transformación y carga de datos para ser utilizados por los algoritmos de minería de datos para los periodos 2014-2019
- Identificar los factores y criterios de segmentación del sector agropecuario peruano basada en la productividad, economía y prácticas de desarrollo, para aplicar algoritmos de agrupamiento.
- Obtener grupos específicos de los principales componentes y características del sector agropecuario, para que pueda ser aprovechado por el sector público y privado.

- Realizar la validación de los resultados obtenidos mediante la aplicación de herramientas como la validación cruzada de minería de datos y métricas de validación tanto externas como internas, para garantizar la aceptabilidad de los agrupamientos.
- Realizar análisis prescriptivos con la información generada, proponiendo estrategias de desarrollo que generen ventajas competitivas en las organizaciones asociadas al sector agropecuario.

1.3. Preguntas de Investigación

A continuación, se definirá una serie de preguntas del problema de estamos estudiando, estas preguntas están formuladas en base a nuestros objetivos.

1.3.1. Pregunta general

¿Cómo será el análisis del sector agropecuario a través de la clusterización basada en la productividad en la productividad, economía y prácticas de desarrollo del sector en el Perú durante el periodo 2014 al 2019?

1.3.2. Preguntas específicas

- ¿Cómo se utilizarán las herramientas para poder recolectar, limpiar y filtrar datos, que servirán de entrada para los algoritmos de minería de datos que se aplicara?
- ¿Cuáles serán los principales factores y criterios de segmentación del sector agropecuario?
- ¿Cuáles serán los principales componentes y características de los grupos específicos se obtendrán?

- ¿Cuál será el resultado final y que conocimiento será generado para realizar análisis prescriptivos, para proponer estrategias de desarrollo a todos los agentes participantes del sector agropecuario?

1.4. Línea y Sub-línea de Investigación a la que corresponde el Problema

Línea Investigación: Inteligencia artificial

Sub-línea de investigación: Reconocimiento de patrones

1.5. Palabras Clave

Clusterización, minería de datos, sector agropecuario, análisis prescriptivos, algoritmos de segmentación.

1.6. Solución Propuesta.

1.6.1. Justificación e Importancia.

El gran crecimiento del sector agropecuario en los últimos años, la evolución de la minería de datos y la información que pone a nuestra disposición el Ministerio de Economía y Finanzas (MEF), el Ministerio de Agricultura y Riego (MINAGRI) en coordinación con el Instituto Nacional de Estadística y Probabilidad (INEI), son los componentes que motivaron esta investigación, mediante un proceso que integre estos componentes se busca aportar conocimiento que sirva para la toma de decisiones para mejorar el desarrollo del sector agropecuario.

Otro componente importante para esta investigación es trabajar políticas de desarrollo productivo desde la lógica de Clústeres, el Ministerio de la Producción está

promoviendo la consolidación de iniciativas de clúster orientado a fortalecer las relaciones entre empresas de una misma zona geográfica y/o cadena de valor a fin de incrementar la productividad y competitividad de las empresas, especialmente Mipymes y contribuir al desarrollo sostenible de las regiones del Perú, por otro lado contribuir a la política de clústeres para el sector público y privado, para ayudar a la generación de información para la medición, monitoreo y evaluación de resultados, útiles en la toma de decisiones para promover políticas publicas orientadas al desarrollo social.

Aprovechando el potencial brindado por la el INEI e implementando una correcta aplicación de técnicas de minería de datos, este estudio pretende generar conocimiento para organizaciones que están involucradas directa e indirectamente con el sector agropecuario, ya que se tendrá información que las organizaciones puedan emplear ofrecer servicios más especializados de acuerdo a los nichos del mercado y en el sector público por los gobiernos locales y regionales del estado para que usando la información de clústeres puedan ofrecer mejores proyectos, estrategias, oportunidades y políticas diseñadas a las necesidades de los agricultores, y así aprovechar mejor los recursos para el desarrollo del sector agropecuario.

1.6.2. Descripción de la Solución

Está demostrado que para muchas empresas la cooperación con otras empresas y agentes que pertenecen a un mismo segmento económico, se constituye una alternativa estratégica que permite aprovechar las ventajas competitivas de todos los interesados, a medida que más agentes participen se puede ir tejiendo un entramado de relaciones que se enlazan hasta crear redes compactas que se desarrollan generalmente en un entorno geográfico concreto limitado (Capó-Vicedo et al., 2007).

Bajo tal premisa, la solución propuesta en este proyecto de tesis es aprovechar la cantidad de datos que provee el INEI de las encuestas nacional del sector agropecuario del año 2014-2019 para poder proponer grupos de clústeres especializados teniendo en cuenta diferentes variables.

Estas conformaciones de clústeres estarán dirigidas a poder formar una gran red de todos los agentes interesados y así poder conformar una gran economía clisterizada, para llevar a cabo este proceso se utilizarán algoritmos de minería de datos, apoyados a por análisis estadísticos e interpretación de resultados para proponer oportunidades de desarrollo en el sector agropecuario.

1.6.3. Aporte

La siguiente investigación presenta un aporte practico que busca proponer grupos específicos de los principales componentes y características del sector agropecuario, para que pueda ser provechado por el sector público y privado, y así poder aportar mediante la realización de análisis prescriptivos para proponer soluciones especializadas a las diversas necesidades que se detectaran, este conocimiento generado brindara la facilidad para generar ventajas competitivas del sector en base a estrategias de desarrollo para todos los agentes participantes del sector.

2. FUNDAMENTOS TEÓRICOS

2.1. Estado del arte

2.1.1. Antecedentes Nacionales

En la investigación de Mamani (2015) se aplica diversos algoritmos de minería de datos basados en clustering y se evalúa su eficiencia para generar conglomerados que ayuden a la interpretación de la información, los cuales permiten evaluar situaciones en una empresa de manera holística y enfocarse en el cumplimiento de los objetivos establecidos en el plan estratégico empresarial.

En el trabajo planteado por Obando (2017), se desarrolló una evaluación de los algoritmos de minería de datos más eficientes en un entorno comercial, dando como resultado información detallada acerca de los mejores algoritmos para los procesos comerciales y la ventaja que aporta a la toma de decisiones.

2.2.2. Antecedentes Internacionales

El campo de aplicación para la ciencia de datos y sus técnicas relacionadas, es muy amplio y variado, en el siguiente estudio se utiliza la capacidad de estas tecnologías para poder monitorear las tendencias y cambios del mercado laboral para que sirva como un componente de ayuda para que las personas que buscan empleo, empleadores, legisladores e inversionistas puedan utilizar, para poder extraer patrones subyacentes de grandes colecciones de datos se utilizó O*NET, el cual es una base de datos publica que contiene los requisitos ocupacionales de los trabajadores, en esta investigación para llevar a cabo sus objetivos sugiere un enfoque de minería de datos donde se desarrolló un modelo de Indexación Semántica Latente(LSI) que cuenta con la capacidad de hacer coincidir los anuncios de trabajo extraídos de la web con los datos que describen las perfiles ocupacionales en O*NET, esta implementación resalto las tendencias laborales en diferentes industrias y áreas geográficas, identificando los grupos ocupacionales y estudiando los cambios en el contexto laboral en un determinado tiempo, estos descubrimientos brindo lecturas mucho más precisas del mercado laboral en comparación con estadísticas oficiales comunes de empleo, esta investigación aporta un enfoque que se puede aplicar directamente en diversos mercados laborales, sectores comerciales y regiones geográficas (Karakatsanis et al., 2017).

En la investigación titulada “A data science and engineering solution for fast k-means clustering of big data” de Chiang, (2018) propone un enfoque de minería de datos para poder abstraer mercados importantes que van a servir para las estrategias de marketing de Gestión de Relaciones con los Clientes (CRM) en una cafetería en Taiwán, donde se empleó un algoritmo de agrupamiento difuso, reglas de asociación en modo difuso y un análisis de la canasta de mercado para procesar los datos recopilados en un sistema CRM, con la finalidad de identificar a clientes difusos, como resultado de esta investigación se lograron encontrar mercados nítidos y mercados difusos además de reglas de asociación para marketing y patrones de comportamiento en las compras de

los clientes, esta contribución proporciona un enfoque de minería de datos en la que integra medidas difusas y algoritmos a priori supervisados para generar las reglas de asociación en modo difuso, finalmente con la información generada y analizada, los investigadores pudieron proponer nuevos servicios para atraer a clientes más jóvenes ofreciendo servicios gratuitos, como internet, información y pedidos a través de teléfonos inteligentes.

En la siguiente investigación se utiliza el método de agrupación de K-Means con densidad poblacional y entropía espacio local en un estudio para poder abordar la problemática de la expansión urbana descontrolada y el peligro que representa para la sostenibilidad de las ciudades, este problema se buscó solucionar mediante la aplicación de un modelo para simplificar la identificación de patrones de expansión urbana, el cual se basa en interpretaciones de diagramas de patrones clásicos de estructura espacial urbana, finalmente los resultados y las comparaciones de los datos de población abierto y datos de teléfonos móviles, verifican que el agrupamiento es adecuado identificando las unidades de tierra en expansión del área urbana y rural, como resultado final la investigación sugiere un modelo de expansión urbana y su entropía espacial local (Liu et al., 2018).

En el siguiente artículo científico del autor Yu et al. (2020) se dedica a aplicar la adaptive resonance theory (ART) y self-organizing map (SOM) en el aprendizaje automático, en el estudio se demuestra que estos algoritmos no supervisados, determinan mejor los clústeres más óptimos para la agrupación que el K-Means. Esta demostración se dio mediante la aplicación en conjunto de SOM complementado con K-Medias para estudiar cómo funciona el marketing de juegos en línea en Taiwán, finalmente como resultado de esta investigación, después de agrupar los datos, se pudo hacer referencia a las decisiones y responder a preguntas en función de los resultados de agrupación, como, por ejemplo, donde se puede promover un juego nuevo para atraer clientes potenciales, las diferentes edades de los clientes y en que se clasifican,

además, se pudo aprovechar la información de las preferencias de los clientes para diseñar nuevos juegos basándose en la arquitectura y atributos populares de los juegos.

El rápido crecimiento de empresarial representa un rápido crecimiento de la información, ese por eso que en la investigación “A Customer Group Mining Method Based on Cluster Analysis” de (Yogping & Zizhen, 2020) se aprovecha esta información de los datos de marketing para analiza con precisión de datos de los clientes, descubrir las características de consumo de los clientes, y proporcionar diferentes productos a diferentes grupos de clientes, para esto se aplicó un algoritmo de análisis de agrupamiento al proceso de minería de grupos de clientes, que mediante la información clave de los clientes existentes se los clasifico en categorías de clientes y se obtuvo información de clientes de mayor potencial, como resultado de este estudio se demuestra que la segmentación de clientes basado en la agrupación de resultados mejora la eficiencia del análisis de la relación entre grupos de clientes y los productos, identificando factores importantes que afectan la venta de productos y proporcionan una base de apoyo para que las empresas lleven a cabo un servicio de precisión basado en el cliente.

En el siguiente artículo publicado por Miloudi et al. (2021) se presenta un enfoque eficaz para buscar un agrupamiento optimo en múltiples bases de datos y evaluar la calidad de la agrupación, para esto se desarrolla un algoritmo en clúster basado en gradientes para la minería de datos, el cual lo aplicaron para desarrollar múltiples experimentos en diversas fuentes de datos y como resultado se logró demostrar la superioridad de este algoritmo con respecto a otros algoritmos de clusterización ya existentes, ya que es más óptimo en la ejecución y la precisión.

Según un estudio desarrollado en Sinaloa por los autores Zayas et al. (2017) estudia el aporte al desarrollo mediante la aplicación de clústeres, donde se indica que la aglomeración de un espacio territorial beneficia a las empresas para estar una cerca de la otra y de esta manera integrar una gran red de empresas, formando clústeres como una herramienta de gestión para obtener nuevas posibilidades de desarrollo, como

resolviendo problemas de baja competitividad e innovación en pequeñas, medianas y grandes empresas de producción agrícola.

Para la llevar a cabo los objetivos de esta investigación se va a tomar en cuenta el artículo Dierckens et al., (2017) ya que se aborda algunas problemáticas de sus algoritmos asociados de K-Means, como los riesgos que se presentan en las selecciones de Centroides al azar, tendencias de generar agrupaciones esféricas y la alta complejidad de tiempo de ejecución en un entorno de Big Data, para esto se desarrolló un algoritmo llamado CluDataSE que aporta a la solución de las problemáticas mencionadas anteriormente y además proporciona una agrupación rápida de k-medias para Big Data, como resultado de esta aplicación CluDataSE demostró una mejora del algoritmo K-Medias, ya que logro resolver las falencias mencionadas, y se logró obtener mejores resultados en los que el K-Means normalmente no podría hacerlo.

2.2. Bases Teóricas de la Investigación

2.2.1. Minería de datos

El concepto de minería de datos que define García y Molina (2012) como un término que engloba un conjunto de componentes resultados de investigación, técnicas y herramientas con el fin de extraer información de utilidad en grandes volúmenes de datos, concretamente la minería de datos debe ser complementada en un proceso que la engloba, y complementa para contribuir directamente al análisis de la información recopilada.

2.2.1.1. Proceso de minería de datos

El proceso de minería de datos consiste extraer información caracterizada por ser útil, comprensible y novedosa, cuyo objetivo principal es descubrir información que esta oculta o implícita, en grandes volúmenes de datos, se diferencia de métodos convencionales ya que, mediante su aplicación, se puede descubrir

información que no es posible descubrir mediante métodos estadísticos convencionales (Moine et al., 2011).

Para tal proceso se necesita transformar información de bajo nivel en conocimiento de alto nivel (Riquelme et al., 2006), para permitir este descubrimiento de patrones se requiere aplicar el proceso Knowledge Discovery in Databases (KDD), estas constan de etapas las cuales consisten en comprender el dominio de la aplicación, extraer la base de datos objetivo, preparar los datos, aplicar algoritmos de minería de datos, interpretación de los descubrimientos y por ultimo aplicar el conocimiento descubierto (Camana, 2016).

El descubrimiento de conocimiento en las bases de datos o más conocido como Knowledge Discovery in Databases (KDD), tiene el mismo objetivo que la minería de datos, el cual es utilizar los datos disponibles en la base de datos y luego procesar los datos para obtener información nueva útil (Utomo y Mesran, 2020), a diferencia que la minería de datos es un paso en el proceso KDD, este incorpora teorías, algoritmos y métodos de la interacción de varios campos de investigación que incluyen tecnología de bases de datos, estadísticas, aprendizaje automático, inteligencia artificial, sistemas basados en el conocimiento y visualización de datos, las cuales cada una proporcionan herramientas específicas que la minería de datos utiliza en varios pasos del proceso KDD (Choudhary et al., 2006)

2.2.1.2. Técnicas de minería de datos

Para poder llevar a cabo la detección de patrones, anomalías o desviaciones, se utilizan diversos algoritmos que se valen de varias técnicas y procedimientos (Martínez, 2011), la tipología de estos algoritmos se clasifica en dos grandes grupos, las técnicas predictivas o supervisadas y las técnicas descriptivas o no supervisadas.

1) Técnicas predictivas supervisadas

Para las técnicas supervisadas que define Gironés et al. (2017), se aplican algoritmos basados en un proceso de aprendizaje en un conjunto de datos, etiquetados previamente, que sirven de entrenamientos, es decir cada ocurrencia del conjunto de datos se debe conocer el valor del atributo o clase, este etiquetado le dará al algoritmo la capacidad para poder deducir una función capaz de predecir el atributo objetivo para un conjunto de datos nuevo. Dentro de este grupo podemos encontrar los principales algoritmos de regresión, clasificación y predicción.

a. Regresión

La analítica predictiva tiene como principal pilar a los modelos de regresión, el enfoque de este modelo radica en establecer una ecuación matemática que servirá como modelo para poder representar cómo interactúan diferentes variables entre sí (Strickland, 2012). Otra definición de (Marín, 2014, diapositiva 24) la considera una tarea que consiste en aprender una función real, que va a asignar a cada instancia un valor real, con el principal objetivo de minimizar el error entre el valor real y el valor que será precedido.

Por ejemplo, si una constructora va a construir y vender departamentos en una determinada zona, y desea determinar el costo adecuado para dichos departamentos, para ello una solución es que la constructora se base en los costos de las viviendas cercanas a la zona, y así desarrollando el modelo se lograría predecir el costo más factible para los nuevos departamentos.

b. Análisis de serie de tiempo

Este tipo de modelo tiene la capacidad de predecir valores futuros que se desconocen, basándose en la historia de una o más variables, es similar a las regresiones, con la diferencia que, en una serie de tiempo, el orden de las observaciones es importante, debido a que se toma en cuenta el orden determinado en el que ocurren (Zamora, 2018). Es decir que la serie de tiempo, es un conjunto de valores sobre algún periodo de tiempo (Rodríguez, 2006).

c. Clasificación

Esta técnica de aprendizaje supervisado es expresada por Sandoval (2018) como una técnica donde se aplican algoritmos para que nos indique a que grupo pertenece un elemento en estudio, estos algoritmos se encargan de encontrar patrones en los datos que le entregamos y los clasifica en grupos. Posteriormente compara los datos resultantes y los ubica en uno de los grupos y de esta manera puede predecir de que se trata. Las variables que se pueden predecir en un conjunto de estados pueden ser Binario:

- Binarios: (Si, No), (Azul, Rojo), (Fuga, No Fuga), etc.
- Múltiple: Comprara (producto1, producto2, producto3, ...), etc.
- Ordenada: Riesgo (Bajo, Medio, Alto), etc.

d. Predicción

Para el especialista en análisis predictivo Siegel (2013) en su *libro Predictive Analytics: The power to predict who will click, buy, lie or die*, “El análisis predictivo consiste en la tecnología que aprende de la

experiencia para predecir el futuro comportamiento de individuos para tomar mejores decisiones” (p 107).

Es un área de la minería de dato que, a partir de la extracción de la información existente, puede predecir tendencias y patrones de comportamiento, sobre cualquier evento desconocido. Este análisis se basa en la identificación de relaciones entre variable de eventos pasados, para que a partir de esta se pueda dichas relaciones y predecir posibles resultados en situaciones futuras. Por ejemplo, en un caso práctico el análisis predictivo no permitirá indicar que clientes son más probables de comprar determinado producto. Esta información, si se utiliza correctamente, permitiría orientar los esfuerzos para ser más productivos para lograr alcanzar los objetivos.

2) Técnicas descriptivas o no supervisadas

Mientras que para las técnicas no supervisadas Gironés et al. (2017) los define como algoritmos que basa el proceso de entrenamiento en un conjunto de datos no etiquetados, es decir que se desconoce el atributo objetivo, por lo tanto, debido a esta condición es usada para la detección de patrones ocultos en grandes cantidades de datos, que permitan agruparlos o segmentarlos.

a. Clustering

En el libro de Michael y Gordon (2011) se definen el clustering como la tarea de segmentar una población variada, en un conglomerado más uniforme. El objetivo considerado en el trabajo de investigación citado por Mamani (2015) es encontrar grupos diferentes unos de otros, particionado un conjunto de datos en colecciones de objetos o instancias, con la finalidad, que en cada partición los objetos sean similares entre si y se diferencien de los objetos que otras particiones contienen.

Por ejemplo, en el estudio de Cruz et al. (2010) se emplea esta técnica para evaluar el desempeño de proveedores en un centro de salud en la ciudad de Bogotá, donde se evaluó a los proveedores de servicio de mantenimiento, para agruparlos según la calidad del desempeño de las actividades que realizan, donde como resultado de esta aplicación se permitió caracterizar a los proveedores de servicio de mantenimiento y caracterizar el inventario de los equipos médicos, demostrando la efectividad de esta técnica se logró identificar a los proveedores más competitivos.

b. Reducción de dimensiones

Para Delgado et al. (2006) la reducción de dimensiones, comúnmente es usada en la etapa de proceso en entrenamiento de sistemas, este consiste en escoger un subconjunto de variables, con el objetivo que, de acuerdo a un criterio de evaluación, las características queden reducidas considerablemente, con la finalidad de identificar el subconjunto que mejor representa el espacio de entrenamiento.

c. Reglas de asociación

Este concepto surgió en los estudios de los patrones de compra en los consumidores en tiendas de comestibles, donde los clientes compran diferentes artículos en diferentes patrones, donde algunos patrones son obvios y otros son sorprendentes (Lalanne, 2016).

d. Esta es una técnica que se utiliza para identificar las posibles relaciones o correlaciones ante distintos sucesos, donde se puede reconocer la ocurrencia de un suceso que puede inducir la aparición de otros, se utiliza principalmente cuando el objetivo es realizar análisis exploratorio. Las asociaciones que se identifican pueden ser utilizadas

para predecir comportamientos, y descubrir correlaciones y co-ocurrencias de eventos (Garcia y Molina, 2012).

2.2.2. Inteligencia de negocios

Muchas organizaciones, soportan gran parte de sus actividades diarias de negocio mediante un sistema de información, mediante las exigencias del negocio el sistema de información va a ser variado, y con el transcurso del tiempo este sistema va a crecer y van a llegar a tener la historia de la organización, almacenados en las bases de datos, este es un recurso que puede ser utilizado para argumentar las decisiones empresariales que se requieran tomar (Gómez y Bautista, 2010).

Para aprovechar este recurso surge la inteligencia de negocios, se denomina así al conjunto de acciones, estrategias y herramientas que tiene por finalidad administrar y crear conocimiento, empleando el análisis de datos que poseen las organizaciones (Ahumada y Perusquia, 2016). Para aportar conocimientos valiosos sobre información operativa identificando problemas y oportunidades del negocio, con ella los involucrados tienen la capacidad de acceder a la información para comprender tendencias, establecer y analizar relaciones en el desempeño del negocio (Muñoz et al., 2016)

2.2.2.1. Aplicaciones de inteligencia de negocios

Algunas de las aplicaciones planteadas por Honrado (2019) son: CRM para permitir el conocimiento y monitoreo constante de sus clientes, servicios públicos, control de riesgo bancario en instituciones financieras, operación y producción para medir el desempeño organizacional, indicadores de control del cuidado de la salud, logística de empresas, organización de expedientes de seguros, análisis de administración de tarjetas de crédito, toma de decisiones en organizaciones de telecomunicaciones, indicadores de inversión en el mercado de valores, indicadores de control para empresas de entretenimiento, entre otras.

2.2.2.2. Proceso ETL

En muchas ocasiones, los datos no pueden ser utilizados de la misma forma en la que se almacena en el sistema, y para poder ser utilizados se necesita que la información pase por algún tipo de transformación que integre la información, se pueda adaptar a las necesidades y se pueda manipular sin afectar la integridad original de la base de datos (Nogués y Valladares, 2017).

En la investigación de Duque et al. (2016) se desarrolla un modelo con la capacidad de realizar filtrados, para detectar, corregir y garantizar la integridad y consistencia de datos. Este proceso ETL consiste en la toma y estandarización de la estructura de datos, esto quiere decir que, debido a la naturaleza de las formas de almacenamiento, como archivos planos y repositorios de datos estructurados, presentan problemas y errores al momento de adquirirlos, es por eso que se necesita este proceso de extracción previo, para pasar a la siguiente fase del proceso.

En la etapa de transformación, a los datos provenientes de la extracción se les aplicara un conjunto de reglas o funciones, para que estén listos para ser cargados al destino final, una funcione importante para esta fase es la limpieza de datos, ya que el éxito de dicho proceso depende mucho de la calidad de datos y los datos objetivos (Loya y Carden, 2018).

Por último, se tiene la fase de carga, donde los datos ingresarán al almacén de datos, estos datos se ingresarán a una tabla de destino llena con los resultados de la transformación, y tomando ciertos criterios dependiendo de los requerimientos organizacionales.

2.2.3. Minería de datos para la inteligencia de negocios

Muchas organizaciones consideran a la información como uno de los activos más importantes que les permite lograr una ventaja competitiva y estratégica, por lo que

consideran identificar esta información de manera estructurada para determinar el valor de los datos y utilizar técnicas de minería de datos (Cendejas et al., 2016).

La minería de datos, se puede aplicar prácticamente en todas las actividades humanas que generen datos, en la siguiente tabla podemos ver los campos y su contribución en las actividades más resaltantes (Riquelme et al., 2006).



Tabla 1

Campos de aplicación de la minería de datos para la inteligencia de negocios

Campo	Aplicación
--------------	-------------------

Banca y Comercio	Segmentación de clientes, Previsión y análisis de ventas, análisis y evaluación de riesgos.
Medicina y Farmacia	Evaluación de la efectividad de los tratamientos, diagnóstico de enfermedades
Seguridad y Detección de Fraude	Reconocimiento facial, identificaciones biométricas, acceso a redes no permitidas, fraudes financieros
Recuperación de información no numérica	Minería de textos, minería web, búsqueda e identificación de imágenes, videos, voz y textos de bases de datos multimedia.
Astronomía	Identificación de nuevas estrellas y galaxias
Minería y Minería	Explotación minera en bases de datos de imágenes satelitales
Agricultura	Segmentación de cultivos y zonas agrícolas, detección de tendencias en el sector agrícola, detección de factores influyentes en el rendimiento de cultivos
Ciencias Ambientales	Identificación de modelos de funcionamiento de ecosistemas naturales y/o artificiales, evaluación, control y gestión de ecosistemas naturales y/o artificiales
Ciencias Sociales	Estudios de los flujos de la opinión pública, planificación de ciudades, evaluación de indicadores sociodemográficos

Adaptado de “Minería de Datos: Conceptos y Tendencias”, por Riquelme, J. C., Ruiz, R., & Gilbert, K. 2006, *Revista Iberoamericana de Inteligencia Artificial*, 29(29), 11–18.

2.2.4. Clusterización

La Clusterización o agrupamiento, es una técnica de la minería de datos que tiene el propósito de identificar y formar de forma automática subgrupos homogéneos, de un conjunto de elementos diverso, según el grado de semejanza entre ellos (Wilford, 2006), mediante la aplicación de los algoritmos de Clusterización se busca maximizar la similitud de los registros dentro del grupo y se minimiza la similitud con los registros fuera del grupo (Larose y Larose, 2015).

2.2.4.1. Medidas de similitud

Para determinar la proximidad entre los datos, aplicando técnicas de clustering, se mide la distancia entre un par de objetos. Las medidas de similitud o proximidad se aplican sobre los objetos, que están compuestos por un conjunto de características. Sea N el número de objetos, cada uno con M características, este conjunto de datos se puede representar en una matriz de tamaño $(N \times M)$ simétrica, llamada matriz de proximidad. Para cada medida de similitud se debe satisfacer las propiedades de simetría, positividad, desigualdad triangular y reflexividad (Álvarez, 2012).

2.2.4.2. Métodos de agrupamiento

Los algoritmos de Clusterización según Benítez y Díez (2005) están basados en la optimización de una función objetivo, que generalmente es la suma ponderada entre las distancias a los centros, este paso consiste en asignar a cada objeto una medida de semejanza al patrón o centroide de cada clúster, para determinar a cuál de los grupos detectados pertenece el objeto.

A continuación, se dan a conocer algunos métodos de agrupamiento más conocidos, y su funcionamiento.

A. Algoritmo K-Means

- Este es un algoritmo explicado por Larose y Larose (2015) es sencillo y eficaz para encontrar Clústeres en los datos, el cual procede de la siguiente manera:
- Paso 1: Determinar en cuanto clústeres K se debe dividir el conjunto de datos.
- Paso 2: Se debe asignar aleatoriamente K registros las cuáles serán las ubicaciones iniciales del centro del grupo.
- Paso 3: Se debe buscar el centro más cercano para cada registro, cada subconjunto de registros tendrá un centro de conglomerados, esto representa una partición del conjunto de datos donde se tiene K conglomerados, C_1 , C_2 , ..., C_K .

- Paso 4: Se debe encontrar el centro del conglomerado para cada uno de los K conglomerados actualizar la ubicación de cada centro al nuevo valor del centroide.
- Paso 5: Por último, se debe repetir el paso 3, 4 y 5 hasta la convergencia o terminación.

B. Algoritmo CLARANS

Este algoritmo de agrupamiento CLARANS (Clustering Large Applications Based on Randomized Search) usa reglas de asociación basados en la cercanía a un centroide, en un conjunto de individuos que poseen características similares. El proceso realizado en este algoritmo consiste en buscar una muestra aleatoriamente en un conjunto de datos, después se obtiene un agrupamiento que al sustituirlo a un solo medoide se denomina el vecino del agrupamiento actual, si el objeto a lo largo del proceso encuentra un mejor vecino, el algoritmo lo mueve al nodo vecino y el proceso se repite, si ya no se encuentran mejores vecinos el agrupamiento más reciente para y se produce un clúster (Salvador, 2018).

C. Algoritmo DBSCAN

Es un algoritmo basado en densidad para realizar la clasificación. El algoritmo consiste en seleccionar un punto P arbitrario, cuando P es un punto central, la construcción de un grupo comienza, y en este se ubican todos los objetos denso-alcanzables desde P, pero cuando P no es un punto central se procede a visitar otro objeto del conjunto de datos, este proceso continuara hasta que todos los objetos hayan sido procesados, aquellos puntos que queden fuera y no pertenezcan ningún grupo se llaman puntos ruido, y los puntos que no son ni ruido ni centrales se les denomina puntos borde. Los puntos centrales y los puntos borde forman parte de la construcción de cada grupo (Pascual et al., 2007).

D. Algoritmo BIRCH

BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) es un algoritmo jerárquico ya que tiene una estructura de árbol, en la que cada nodo del árbol representa un grupo, las hojas grupos de menor tamaño o jerarquía, y un grupo de mayor jerarquía contiene a su subgrupo de hijos. La principal ventaja de BIRCH se da ya que se guarda la información simplificada de los componentes como centroide y radio del grupo que lo componen, es un algoritmo eficiente ya que solo se requiere revisar el conjunto de datos una sola vez (García, 2016).

E. Algoritmo ROCK

El algoritmo ROCK (RObust Clustering using LinKs) emplea ligas y no distancias para medir la similitud y la proximidad entre un par de clústeres a ser mezclados. El algoritmo utiliza el coeficiente Jaccard, el cual mide el grado de similitud entre dos conjuntos. Utiliza el parámetro θ para determinar en cada uno de los objetos, a sus vecinos. Dado un punto P , para determinar si un punto Q es vecino de P el coeficiente de Jaccard $\text{sim}(P, Q)$ debe exceder el valor de θ . Los datos se generan en una matriz de links o ligas $\text{links}(P, Q)$ que contiene el número de vecinos comunes entre los puntos P y Q . El algoritmo utiliza una técnica aglomeraría que en principio todos los datos pertenecen a un clúster diferente y progresivamente se van agrupando (Janas et al., 1990).

F. Algoritmo CHAMELEON

Es un algoritmo de agrupamiento que encuentra grupos en un conjunto de datos utilizando un algoritmo de dos fases, Durante la primera fase, se utiliza un algoritmo de partición de gráficos para agrupar los elementos de datos en una gran cantidad de subgrupos muy pequeños. Para la segunda fase, se emplea un algoritmo de agrupamiento jerárquico aglomerativo con la finalidad de encontrar los grupos genuinos aplicando la combinación repetida

de estos subgrupos (Karypis et al., 1999).

2.2.4.3. Comparación algoritmos de clustering

En la siguiente tabla se puede observar algunas de las características más importantes de los algoritmos mencionados anteriormente complementándolos con otros más.



Tabla 2

Algoritmos de clustering y sus principales características

Nombre	Tipo de Dato	Complejidad	Geometría	Resistencia al Ruido	Parámetros de entrada	Resultados
Algoritmos Particionales						
K-Means	Numérico	$O(n)$	Formas no convexas	No	Número de Clusters	Centroides de los Clusters
M-Modas	Categorico	$O(n)$	Formas no convexas	No	Número de Clusters	Modas de los Clusters
PAM	Numérico	$O(k(n - k)^2)$	Formas no convexas	No	Número de Clusters	Medoides de los Clusters
CLARA	Numérico	$O(k(40 + k)^2 + k(n - k))$	Formas no convexas	No	Número de Clusters	Medoides de los Clusters
CLARANS	Numérico	$O(kn^2)$	Formas no convexas	No	Número de Clusters, Número máximo de vecinos examinados	Medoides de los Clusters
FCM	Numérico	$O(n)$	Formas no convexas	No	Número de Clusters	Centroides de los Clusters, Beliefs
Algoritmos Jerárquicos						
BIRCH	Numérico	$O(n)$	Formas no convexas	Si	Radio de los Clusters, Factor de derivación	CF = (número de puntos en el clúster N, suma lineal de los puntos en el clúster LS, la suma cuadrada de los N datos SS) puntos
CURE	Numérico	$O(n^2 \log n)$	Formas arbitrarias	Si	Número de Clusters, Número de Clusters representativos	Asignación de los datos a los Clusters

ROCK	Catagórico	$O(n^2 + nm_{max} + n^2 \log n), O(n^2, nm_{max})$ Dónde: m_{max} es el número máximo de vecinos para un punto y m_{avg} es el numero promedio de vecinos para un punto	Formas arbitrarias	Si	Número de Clusters	Asignación de los datos a los Clusters
Algoritmos Basados en la Densidad						
DBSCAN	Numérico	$O(n \log n)$	Formas Arbitrarias	Si	Radio del Clúster, Número Mínimo de Objetos	Asignación de los datos a los Clusters
DECLUE	Numérico	$O(n \log n)$	Formas Arbitrarias	Si	Radio del Clúster, Número Mínimo de Objetos	Asignación de los datos a los Clusters
* n : número de puntos en el dataset						
k : número de Clusters						

Adaptado de “Clustering de Documentos con Restricciones de Tamaño” por Vallejo, D. 2016, *Universitat Politècnica de València*.



3. MARCO METODOLOGICO

3.1. Alcances y Limitaciones

Se considera un alcance de investigación explicativo, dado que el investigador, realizará el proceso de clusterización con los datos obtenidos, estableciendo criterios de agrupamiento, identificando cada clúster del modelo obtenido, lo que permita un análisis prescriptivo y ventajas competitivas para el sector.

La principal limitante para realizar esta investigación es el alto consumo de recursos de software al procesamiento de la información, ya que la cantidad de módulos y la información contenida en cada uno requiere de múltiples procesamiento realizados de forma secuencial, el formato en la que se nos entrega la información requiere pasar transformaciones y seguidamente por la aplicación de algoritmos de minería de datos lo cual va a consumir el recurso del tiempo en el procesamiento.

3.2. Tipo y Nivel de la Investigación

El diseño de la investigación propuesta es no experimental de corte longitudinal, con enfoque cuantitativo, ya que se trabajará con los datos públicos del INEI, los cuales fueron obtenido mediante la encuesta nacional agropecuaria del año 2014 al 2019, en el cual se observa el comportamiento natural del sector agropecuario, en base diversas características. En base a estas características e información se evaluará el comportamiento del sector agropecuario a través de cada año, para hacer inferencias respecto a los cambios, sus determinantes y consecuencias, que nos ofrezcan los modelos estadísticos, resultantes del procesamiento de datos.

3.3. Población y Muestra o Universo

La población de estudio son las unidades agropecuarias, parcelas y productores agropecuarios, obtenidos por la encuesta nacional agropecuaria, que sea aplico año tras año desde el 2014 al 2019, donde por cada año contiene 23 módulos.

La muestra total es de 30 806 unidades agropecuarias, distribuidas en 29 218 unidades agropecuarias de medianos y pequeños productores con los niveles de inferencia nacional y departamental, y 1588 unidades agropecuarias de grandes productores; estos se desarrollaron teniendo en cuenta 23 módulos de investigación, las cuales son:

- Características de la unidad agropecuaria
- Cultivos cosechados en la unidad agropecuaria
- Buenas prácticas agrícolas: condicionadas y no condicionadas
- Producción de la actividad pecuaria
- Buenas practicas pecuarias: condicionadas y no condicionadas
- Inocuidad agroalimentaria
- Servicios de extensión agraria
- Asociatividad
- Servicios financieros
- Costos de producción de la actividad agropecuaria
- Características del productor y su familia
- Características de la unidad agropecuaria al día de la entrevista

Se considera emplear la técnica de muestreo no probabilístico censal, donde se considera la totalidad de la población de estudio.

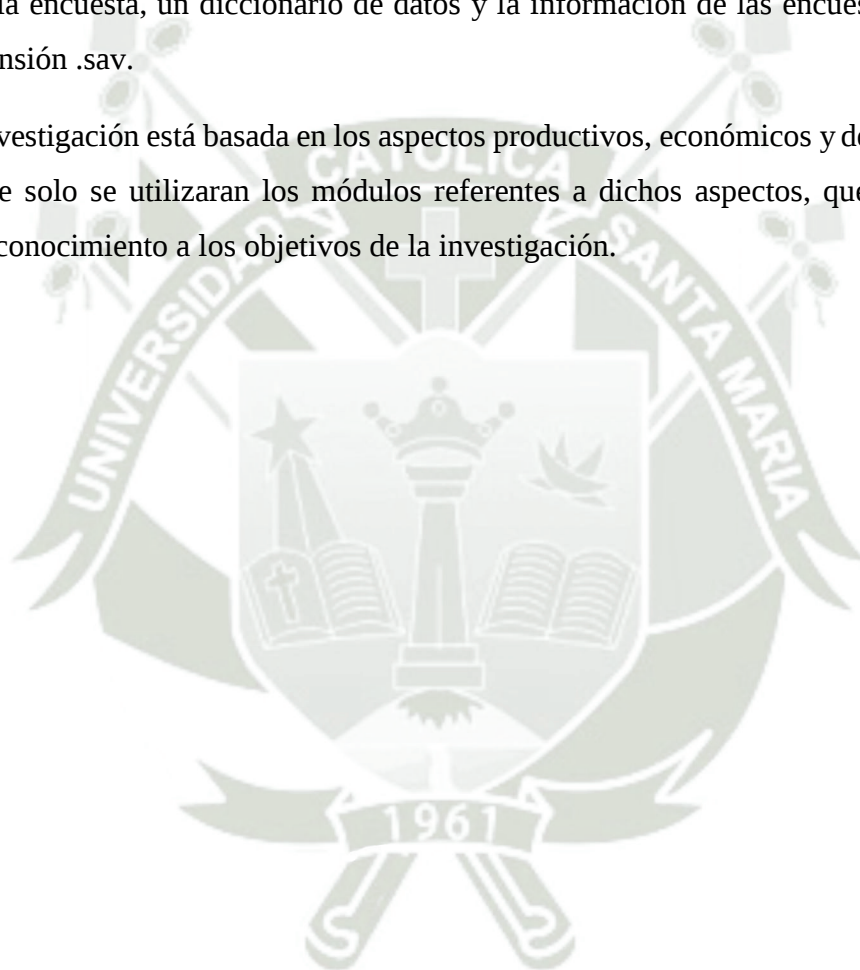
3.4. Métodos, Técnicas e Instrumentos de Recolección de Datos

Para llevar a cabo los objetivos de esta investigación se aplica la técnica de observación, utilizando como instrumento la observación documentaria, ya que la investigación se basará en los datos de la Encuesta Nacional Agropecuaria, que dispone de información estadística que permite caracterizar a las pequeñas, medianas y grandes unidades agropecuarias del país,

la cual está a la disposición de los investigadores a través del sistema de Microdatos del INEI. En este repositorio se pueden ver los módulos y la documentación de las encuestas ejecutadas por el INEI, en formatos compatibles con SPSS, Microsoft Excel y Acrobat Reader.

La Encuesta Nacional Agropecuaria fue realizada del 2014 al 2019, en donde cada año contiene 23 módulos, de acuerdo a diversos criterios que fueron tomados en cuenta por los representantes del MINAGRI y MEF. Cada módulo contiene los formularios aplicados durante la encuesta, un diccionario de datos y la información de las encuestas en archivos con extensión .sav.

Esta investigación está basada en los aspectos productivos, económicos y de desarrollo, por lo que se solo se utilizaran los módulos referentes a dichos aspectos, que contribuyan y aporten conocimiento a los objetivos de la investigación.



3.5. Plan de análisis estadístico de los datos

Para este proyecto se dispone de información que permite caracterizar a las pequeñas, medianas y grandes unidades agropecuarias del país, esta información fue recopilada por el Ministerio de Agricultura y Riego (MINAGRI) y el Ministerio de Economía y Finanzas (MEF) mediante encuestas, donde se realizan entrevistas de forma directa a las unidades agropecuarias durante un periodo determinado de tiempo.

Mediante estas encuestas el Instituto Nacional de Estadística e Informática (INEI) pone a disposición del público en general el sistema de Microdatos, donde encontramos diferentes módulos de la Encuesta Nacional Agropecuaria que contienen diversas tablas con su respectivo diccionario de datos y los cuestionarios aplicados.

La muestra total es de 30 806 unidades agropecuarias distribuidas en diversos módulos, cada módulo cuenta con una o dos tablas estructuradas con gran cantidad de variables resultantes de los cuestionarios aplicados a las unidades agropecuarias, los datos de las tablas están sin procesar, y mediante la aplicación de diversas técnicas se busca generar conocimiento mediante un análisis prescriptivo de las tablas, gráficos y segmentaciones que se generen, y así poder proponer nuevas oportunidades de desarrollo para el sector.

Para llevar a cabo el cumplimiento de los objetivos de esta investigación se aplicará una serie de algoritmos de agrupamiento como K-Means, BDSCAN, BIRCH entre otros, sobre los datos que disponemos; las implementaciones de dichos algoritmos se llevaran a cabo utilizando Python, ya que es un lenguaje de programación que cuenta con librerías que facilitan la implementación de los algoritmos de agrupamiento, además de poder presentar los datos en gráficos interactivos que facilitan su comprensión e interpretación. Para esta implementación se trabajará con puros valores numéricos normalizados y mediante las gráficas que nos proporciona Python mediante sus librerías, se realizará un análisis prescriptivo de los grupos obtenidos, evaluando el comportamiento de las segmentaciones y clasificando la información de acuerdo a criterios especializados que se definirán.

4. PLAN DE TRABAJO

4.1. Recursos

4.1.1. Humanos

Los recursos humanos para el desarrollo de esta investigación en el tesista, el cual llevar a cabo la investigación, con el apoyo y orientación de asesores designados por la escuela.

4.1.2. Materiales

Para la realización de esta investigación se requerirá recursos tecnológicos como: un ordenador conectado a internet, un dispositivo celular, material de oficina, recursos de hardware, material bibliográfico y dispositivos de almacenamiento.

4.1.3. Económicos

Los gastos presupuestados para el desarrollo de la tesis son: servicio de internet, servicio de luz eléctrica, trámites administrativos, gastos de comunicación, mantenimiento de ordenador de escritorio y otros.

4.2. Cronograma

2022																					
Fases de la tesis	Meses	Julio				Agosto				Setiembre				Octubre				Noviembre			
	Semanas	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Formulación de tema																					
Investigación de antecedentes																					
Formulación del problema																					
Marco de referencia																					
Planteamiento operacional																					
Revisión previa de plan de tesis																					
Presentación final de plan de tesis																					
Ajustes del plan de tesis																					
Aplicación de técnicas																					
Recolección de información																					
Procesamiento de datos																					
Conclusiones y recomendaciones																					
Revisión de previa del borrador																					
Presentación de borrador de tesis																					

4.3. Presupuesto

Recursos			
Descripción		Costo por mes	Total
Servicios	Internet	S/. 60.00	S/. 300.00
	Electricidad	S/. 80.00	S/. 400.00
	Comunicación	S/. 40.00	S/. 200.00
Bienes	Material bibliográfico	S/. 15.00	S/. 75.00
	Impresión	S/. 10.00	S/. 50.00
	Materiales de escritorio	S/. 15.00	S/. 75.00
		Total	S/. 1100.00

ANÁLISIS DEL SECTOR AGROPECUARIO A TRAVÉS DE LA CLUSTERIZACIÓN BASADO EN LA PRODUCTIVIDAD, ECONOMÍA Y PRÁCTICAS DE DESARROLLO DEL SECTOR PERÚ, 2015 – 2019

INFORME DE ORIGINALIDAD

2%

INDICE DE SIMILITUD

1%

FUENTES DE INTERNET

0%

PUBLICACIONES

1%

TRABAJOS DEL
ESTUDIANTE

FUENTES PRIMARIAS

1

catalog.ihsn.org

Fuente de Internet

1%

2

sedici.unlp.edu.ar

Fuente de Internet

1%

3

Submitted to Universidad Católica de Santa
María

Trabajo del estudiante

1%

Excluir citas

Apagado

Excluir coincidencias < 1%

Excluir bibliografía

Apagado