

**UNIVERSIDAD NACIONAL DE SAN ANTONIO
ABAD DEL CUSCO
FACULTAD DE CIENCIAS QUÍMICAS, FÍSICAS Y
MATEMÁTICAS
ESCUELA PROFESIONAL DE MATEMÁTICA CON
MENCIÓN EN ESTADÍSTICA**



TESIS

**SEGMENTACIÓN DE USUARIOS DE TELECOMUNICACIONES EN
CLÚSTERES UTILIZANDO LA ENCUESTA RESIDENCIAL DE
SERVICIOS DE TELECOMUNICACIONES EN EL PERÚ, AÑO 2021**

PRESENTADO POR:

Br. CARMEN KATIUSKA JARA SALAS

**PARA OPTAR AL TÍTULO PROFESIONAL DE
LICENCIADA EN MATEMÁTICA MENCIÓN
ESTADÍSTICA**

ASESORA:

Dra. NELLY MARIA SALAZAR PEÑA.

CUSCO – PERÚ

2024

INFORME DE ORIGINALIDAD

(Aprobado por Resolución Nro.CU-303-2020-UNSAAC)

El que suscribe, Asesor del trabajo de investigación/tesis titulada: SEGMENTACIÓN DE USUARIOS DE TELECOMUNICACIONES EN CLÚSTERES UTILIZANDO LA ENCUESTA RESIDENCIAL DE SERVICIOS DE TELECOMUNICACIONES EN EL PERÚ, AÑO 2021

presentado por: CARMEN KATIUSKA JARA SALAS con DNI Nro.: 48140857 presentado por: con DNI Nro.: para optar el título profesional/grado académico de LICENCIADA EN MATEMÁTICA MENCIÓN ESTADÍSTICA

Informo que el trabajo de investigación ha sido sometido a revisión por DOS veces, mediante el Software Antiplagio, conforme al Art. 6° del **Reglamento para Uso de Sistema Antiplagio de la UNSAAC** y de la evaluación de originalidad se tiene un porcentaje de 5.....%.

Evaluación y acciones del reporte de coincidencia para trabajos de investigación conducentes a grado académico o título profesional, tesis

Porcentaje	Evaluación y Acciones	Marque con una (X)
Del 1 al 10%	No se considera plagio.	X
Del 11 al 30 %	Devolver al usuario para las correcciones.	
Mayor a 31%	El responsable de la revisión del documento emite un informe al inmediato jerárquico, quien a su vez eleva el informe a la autoridad académica para que tome las acciones correspondientes. Sin perjuicio de las sanciones administrativas que correspondan de acuerdo a Ley.	

Por tanto, en mi condición de asesor, firmo el presente informe en señal de conformidad y adjunto la primera página del reporte del Sistema Antiplagio.

Cusco, 17 de SETIEMBRE de 2024.....



Firma

Post firma NELLY MARJA SALAZAR PEÑA

Nro. de DNI 23910370

ORCID del Asesor 0000-0002-7582-9492

Se adjunta:

- 1) Reporte generado por el Sistema Antiplagio.
- 2) Enlace del Reporte Generado por el Sistema Antiplagio: oid: 27259:382204639

NOMBRE DEL TRABAJO

**SEGMENTACIÓN DE USUARIOS DE TELE
COMUNICACIONES EN CLÚSTERES UTIL
IZANDO LA ENCUESTA RESIDENCIAL DE
SE**

AUTOR

CARMEN KATIUSKA JARA SALAS

RECUENTO DE PALABRAS

13319 Words

RECUENTO DE CARACTERES

74579 Characters

RECUENTO DE PÁGINAS

90 Pages

TAMAÑO DEL ARCHIVO

3.0MB

FECHA DE ENTREGA

Sep 16, 2024 3:48 PM GMT-5

FECHA DEL INFORME

Sep 16, 2024 3:50 PM GMT-5**● 5% de similitud general**

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos.

- 4% Base de datos de Internet
- Base de datos de Crossref
- 4% Base de datos de trabajos entregados
- 1% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossref

● Excluir del Reporte de Similitud

- Coincidencia baja (menos de 15 palabras)
- Bloques de texto excluidos manualmente

DEDICATORIA

*A mis abuelos Jorge Mariano Salas Fernández
y Pedro Cirilo Jara Gamarra.*

Carmen Katuska

AGRADECIMIENTO

*A mi alma mater la tricentenario universidad nacional de San Antonio Abad del
Cusco*

A los docentes de la escuela profesional de matemática

*En especial a la Dra. Nelly María Salazar Peña por su asesoría y apoyo
incondicional, para el desarrollo de la presente tesis*

A mis papás y hermanos por su amor y motivación

A Richard Palma y Pamela Gonzales por su motivación a seguir.

Carmen Katuska

PRESENTACIÓN

SEÑOR DECANO DE LA FACULTAD DE CIENCIAS QUÍMICAS, FÍSICAS Y MATEMÁTICAS

SEÑOR DIRECTOR DE LA ESCUELA PROFESIONAL DE MATEMÁTICA

SEÑORES DOCENTES MIEMBROS DEL JURADO

De acuerdo con el reglamento de grados y títulos establecidos por la escuela profesional de matemática mención estadística, presento a vuestra consideración el presente trabajo de tesis titulado SEGMENTACIÓN DE USUARIOS DE TELECOMUNICACIONES EN CLÚSTERES UTILIZANDO LA ENCUESTA RESIDENCIAL DE SERVICIOS DE TELECOMUNICACIONES EN EL PERÚ, AÑO 2021, conducente al título profesional de licenciada en matemática mención estadística.

La investigación tiene la finalidad de proponer la segmentación de usuarios de telecomunicaciones utilizando la técnica de segmentación en dos etapas, esta se diferencia de otras técnicas pues tiene la capacidad de trabajar con variables de naturaleza cualitativa y cuantitativa. La fuente de información fue secundaria utilizando la encuesta residencial de telecomunicaciones ERESTEL del año 2021, a continuación, se presenta una descripción del presente informe dividido en 4 capítulos:

Capítulo I, se presentan los aspectos generales conformado por el planteamiento del problema y descripción del problema. Luego la formulación del problema donde se detalla problema general y problemas específicos, así también los objetivo general y objetivos específicos para culminar este capítulo se presenta la justificación de la investigación y las limitaciones de esta.

Capítulo II, conformado por los antecedentes de la investigación y bases teóricas en la que se presentan los antecedentes de las investigaciones que se tomaron como referencia, así como los autores y teóricas relacionadas a la investigación.

Capítulo III, en la que se presenta la metodología de investigación a través de la identificación de las variables, su operacionalización. También se presenta la población y muestra, así como las técnicas e instrumento de recolección de datos luego se presenta el detalle acerca del procesamiento de datos, así como la confiabilidad y validez.

Capítulo IV, en dicho capítulo se muestran los resultados a nivel descriptivo e inferencial. Para terminar, se presenta la discusión, conclusiones, recomendaciones, referencias y anexos.

Atentamente

Br. Carmen Katuska Jara Salas.

RESUMEN

La investigación titulada “SEGMENTACIÓN DE USUARIOS DE TELECOMUNICACIONES EN CLÚSTERES UTILIZANDO LA ENCUESTA RESIDENCIAL DE SERVICIOS DE TELECOMUNICACIONES EN EL PERÚ, AÑO 2021” se busca como objetivo general segmentar a los usuarios de telecomunicaciones en clústeres utilizando la encuesta residencial de servicios de telecomunicaciones en el Perú, año 2021. Para lo cual se empleó una investigación de alcance descriptivo y explicativo, de enfoque cuantitativo y diseño no experimental. La fuente de información fue secundaria tomando como base de datos la encuesta residencial de servicios de telecomunicaciones ERESTEL para el año 2021, cuya población fueron los usuarios de telecomunicaciones en el Perú siendo el tamaño 47 346 encuestados, luego de realizar la limpieza de los datos atípicos y faltantes, la muestra se calcula en 15 780. Para hallar los resultados se utilizó la técnica de segmentación bietápica o en dos fases luego de hallar esto se concluye que los usuarios de telecomunicaciones en el Perú para el año 2021 se pueden segmentar en 4 clústeres, así mismo se propone 4 perfiles de usuarios con sus respectivas características.

Palabras clave: Clústeres, segmentación, telecomunicaciones, bietápica.

ABSTRACT

The research titled “SEGMENTATION OF TELECOMMUNICATIONS USERS IN CLUSTERS USING THE RESIDENTIAL SURVEY OF TELECOMMUNICATIONS SERVICES IN PERU, YEAR 2021” seeks as a general objective to segment telecommunications users in clusters using the residential survey of telecommunications services in Peru , year 2021. For which a descriptive and explanatory research was used, with a quantitative approach and non-experimental design. The source of information was secondary, taking as a database the ERESTEL residential survey of telecommunications services for the year 2021, whose population was telecommunications users in Peru, with a size of 47,346 respondents, after cleaning the atypical data. and missing, the sample is calculated at 15,780. To find the results, the two-stage or two-phase segmentation technique was used. After finding this, it is concluded that telecommunications users in Peru for the year 2021 can be segmented into 4 clusters. Likewise, 4 user profiles are proposed with their respective characteristics.

Keywords: Clusters, segmentation, telecommunications, two-stage..

ÍNDICE

DEDICATORIA.....	II
AGRADECIMIENTO.....	III
PRESENTACIÓN.....	IV
RESUMEN	VI
ABSTRACT	VII
ÍNDICE.....	VIII
ÍNDICE DE TABLAS	XII
ÍNDICE DE FIGURAS	XIII
CAPÍTULO I	1
ASPECTO GENERALES	1
1.1 Planteamiento del problema.....	1
1.2 Descripción del problema.....	2
1.3 Formulación del problema	3
1.3.1 Problema general	3
1.3.2 Problemas específicos	3
1.4 Objetivos de la investigación.....	4
1.4.1 Objetivo general	4
1.4.2 Objetivos específicos	4
1.5 Formulación de las hipótesis.....	4
1.5.1 Hipótesis general.....	4
1.5.2 Hipótesis específicas	4

1.6 Justificación de la investigación	5
1.6.1 Justificación teórica.....	5
1.6.2 Justificación práctica	5
1.6.3 Justificación de implicancia social.....	5
1.6.4 Justificación metodológica	6
1.7. Limitaciones de la investigación.....	6
1.7.1 Espacial.....	6
1.7.2 Temporal	6
1.7.3 Teórica	6
CAPÍTULO II	7
MARCO TEÓRICO.....	7
2.1 Antecedentes de la investigación	7
2.1.1 Antecedentes internacionales	7
2.1.2 Antecedentes nacionales	8
2.1.3 Antecedentes regionales y locales	11
2.2 Bases teóricas	12
2.2.1 Segmentación de mercados	12
2.2.2 Criterios para la segmentación de clientes	13
2.2.3 Análisis exploratorio de datos (AED)	16
2.2.4 Etapas del análisis exploratorio de datos.....	16
2.2.5 Diagramas de sectores circulares.....	17
2.2.6 Diagrama de barras.....	17

2.2.7 Histograma.....	18
2.2.8 Diagrama de tallo y hojas.....	18
2.2.9 Diagrama de caja y bigotes.....	18
2.2.10 Detección de outliers multivariantes.....	20
2.2.11 Análisis de conglomerados.....	22
2.2.12 Análisis de conglomerados bietapico.....	24
2.2.13 Distancia para variables cualitativas y cuantitativas.....	25
2.2.14 Caracterización de los conglomerados.....	26
2.2.15 Árbol de características de conglomerados (CF TREE).....	27
2.2.16 Algoritmo de conglomerados en 2 etapas.....	28
CAPÍTULO III.....	31
METODOLOGÍA DE LA INVESTIGACIÓN.....	31
3.1 Identificación de las variables de investigación.....	31
3.2 Operacionalización de variables.....	32
3.3 Descripción de la metodología.....	33
3.3.1 Alcance de la investigación.....	33
3.3.2 Enfoque de la investigación.....	33
3.3.3 Diseño de la investigación.....	33
3.4 Población y muestra.....	33
3.4.1 Población.....	33
3.4.2 Muestra.....	34
3.5 Técnica e instrumento de recolección de datos.....	35

3.5.1 Técnica.....	35
3.5.2 Instrumento	35
3.6 Técnicas de procesamiento y análisis de datos	36
3.7 Procesamiento de datos.....	36
3.8 Confiabilidad y validez.....	37
CAPÍTULO IV	39
RESULTADOS DE LA INVESTIGACIÓN.....	39
4.1 Resultados	39
4.1.1 Análisis exploratorio y descriptivo	39
4.1.2 Algoritmo para la obtención de los conglomerados.....	48
4.1.3 Identificación de conglomerados	51
4.1.4 Perfil de usuario de acuerdo con el conglomerado que pertenece	53
DISCUSIÓN	55
CONCLUSIONES.....	57
RECOMENDACIONES	60
REFERENCIAS.....	61
ANEXOS	67

ÍNDICE DE TABLAS

Tabla 1 Operacionalización de variables	32
Tabla 2 Estadísticos descriptivos variables cuantitativas.....	39
Tabla 3 Distancia de Mahalanobis	42
Tabla 4 Frecuencias según sexo	44
Tabla 5 Frecuencias por conglomerado.....	51
Tabla 6 Frecuencias por nivel socio económico por conglomerado.....	52
Tabla 7 Media de pago mensual e ingreso principal por conglomerado	52

ÍNDICE DE FIGURAS

Figura 1	Características de los niveles socioeconómicos en el Perú.	15
Figura 2	Diagrama de cajas y bigotes	19
Figura 3	Esquema gráfico del Conglomerado Bietápico.....	30
Figura 4	Diagrama de caja y bigotes de edad	40
Figura 5	<i>Diagrama de cajas y bigotes ingreso principal mensual</i>	40
Figura 6	<i>Diagrama de cajas y bigotes pago mensual</i>	41
Figura 7	Diagrama de caja distancias de Mahalanobis	42
Figura 8	Diagrama de caja distancia de Mahalanobis valores < 16.2660.....	43
Figura 9	Diagrama de sectores sexo	44
Figura 10	Diagrama de barras nivel de instrucción.....	45
Figura 11	Diagrama de barras nivel socioeconómico	46
Figura 12	Diagrama de sectores área	47
Figura 13	Diagrama de barras tiempo con servicio móvil	47
Figura 14	<i>Resumen del modelo de clasificación</i>	49
Figura 15	Tamaño de los conglomerados	50

CAPÍTULO I

ASPECTO GENERALES

1.1 Planteamiento del problema

El análisis de clúster es un conjunto de técnicas multivariantes que permite clasificar elementos de acuerdo con las similitudes en sus características, formando así grupos en los cuales existe similitudes entre los elementos y diferencias con los elementos de los demás grupos (Martinez, 2024).

A nivel internacional, Palacios y Pastor (2020) mencionan que diversas empresas han utilizado el análisis de clúster para poder formar grupos de acuerdo a las características de sus clientes de esa manera poder brindarles un mejor servicio y fidelizar su marca, es así que en Popayán , Chile se realizó una investigación para segmentar clientes de productos de consumo masivo para brindarles un mejor servicio y fidelizar sus compras por lo que se utilizó técnicas de clúster logrando segmentar 5 tipos de clientes para luego crear estrategias de venta con la información recolectada.

En el Perú, Manrique (2016) propuso la segmentación de clientes de la corporación Lindley en Lima, a través de la técnica denominada clúster bietápico.

El análisis de conglomerados bietápico o en dos fases es un instrumento de exploración que tiene la finalidad de encontrar grupos naturales en un universo de datos, para así producir criterios de información, frecuencia de conglomerados y estadísticas descriptivas, su particularidad respecto a otras técnicas de clasificación es la capacidad de incluir tanto variables de naturaleza cualitativa como cuantitativa, además de grandes volúmenes de datos.

A nivel local, Mamani y Damiano (2014) utilizaron el análisis de clúster para la segmentación de trabajadores de acuerdo con sus características socio demográficas y laborales, para así determinar el nivel de satisfacción laboral y conformar dos clústeres.

En ese entender, una alternativa válida para segmentar clientes de acuerdo con sus características socio económicas es a través de la técnica de segmentación llamada clúster bietápico, por lo que la presente investigación propone la clasificación de clientes en telecomunicaciones utilizando dicha técnica.

1.2 Descripción del problema

La última década, el medio de información más importante respecto a las características de la demanda de servicios de telecomunicación, ha sido la Encuesta Nacional de Hogares (ENAH) que está encargada el Instituto Nacional de Estadística e Informática (INEI) la cual en la sección de transporte y comunicaciones se evalúa y cuantifica el monto gastado en el hogar en los servicios de transporte y comunicaciones. Dicha encuesta no es muy extensiva respecto al tema del gasto, el uso de internet por lo cual era necesario tener una fuente adicional de información.

El sector de telecomunicaciones es dinámico respecto a otras áreas por lo que se tendría una razón más para fomentar la realización de una fuente adicional de información. Esto para poder impulsar la competencia del mercado de telecomunicaciones, calidad de los servicios de estos y su uso de manera frecuente.

A partir del año 2012, cada año se lleva a cabo la encuesta residencial de servicio de telecomunicaciones (ERESTEL), la cual profundiza acerca de varios aspectos respecto a los consumidores de telecomunicaciones. Los aspectos más

resaltantes son: Acceso a telefonía fija y móvil, acceso y uso de internet y características socio económicas, la publicación de OSIPTEL referente a los datos hallados en la encuesta ERESTEL se dan a nivel descriptivo sin explotar más los datos a través de análisis estadístico inferencial, la técnica de clasificación de clúster bietápico es una técnica utilizada en minería de datos por su gran precisión y versatilidad para poder clasificar datos en clúster que tengan naturaleza cuantitativa y cualitativa, esta información puede ser valiosa para las empresas de telecomunicaciones y el gobierno en general pues se otorga un panorama más amplio del comportamiento y grupos que generan los usuarios de acuerdo a sus características de socio económicas y de uso es por ello que el presente estudio tiene como finalidad proponer una clasificación de los usuarios de telecomunicaciones tomando como fuente de información la encuesta ERESTEL del año 2021.

1.3 Formulación del problema

1.3.1 Problema general

¿Cómo es la segmentación de usuarios de telecomunicaciones en clústeres utilizando la encuesta residencial de servicios de telecomunicaciones en el Perú, año 2021?

1.3.2 Problemas específicos

- ¿Cómo son los clústeres en los que se segmenta a los usuarios de telecomunicaciones utilizando la encuesta residencial de servicios de telecomunicaciones en el Perú, año 2021?
- ¿Existen perfiles de usuarios de telecomunicaciones por cada clúster utilizando

encuesta residencial de servicios de telecomunicaciones en el Perú, año 2021?

1.4 Objetivos de la investigación

1.4.1 Objetivo general

Segmentar a los usuarios de telecomunicaciones en clústeres utilizando la encuesta residencial de servicios de telecomunicaciones en el Perú, año 2021.

1.4.2 Objetivos específicos

- Describir las características de los usuarios de telecomunicaciones por cada clúster utilizando la encuesta residencial de servicios de telecomunicaciones en el Perú, año 2021.
- Proponer perfiles del usuario de telecomunicaciones para cada clúster utilizando la encuesta residencial de servicios de telecomunicaciones en el Perú, año 2021.

1.5 Formulación de las hipótesis

1.5.1 Hipótesis general

La segmentación de los usuarios de telecomunicaciones en clústeres utilizando la encuesta residencial de servicios de telecomunicaciones en el Perú, año 2021 es significativa.

1.5.2 Hipótesis específicas

- Las características de los usuarios de telecomunicaciones por cada clúster utilizando encuesta residencial de servicios de telecomunicaciones en el Perú, año 2021 son significativas.

- Existen perfiles de los usuarios de telecomunicaciones por cada clúster utilizando encuesta residencial de servicios de telecomunicaciones en el Perú, año 2021.

1.6 Justificación de la investigación

La justificación de la investigación es la exposición de las razones por las cuales se lleva a cabo el estudio, toda investigación tiene un propósito para que pueda acreditar su realización. A continuación se presentan las justificaciones que respaldan la investigación desarrollada: Justificación teórica

La presente investigación propone la teoría de clustering bietápico aplicado para la segmentación de usuarios de telecomunicaciones, además de poder ser utilizado como precedente para posteriores investigaciones.

1.6.2 Justificación práctica

La propuesta de segmentación de usuarios en telecomunicaciones a través de la técnica cluster bietápico es útil para poder conocer las características de cada conglomerado de usuarios esto puede ser utilizado por diversas empresas dedicadas al rubro de telecomunicaciones para proponer estrategias de fidelización de clientes, así también los resultados hallados brindan información de las diversas características de los usuarios de acuerdo a cada conglomerado que puede ser útil para la propuesta de políticas de estado respecto al rubro de telecomunicaciones.

1.6.3 Justificación de implicancia social

Esta información puede ser útil para que los servicios entregados a los usuarios sean de calidad y pertinentes, ajustados a sus necesidades.

1.6.4 Justificación metodológica

Para la presente investigación se consultaron libros, artículos científicos y publicaciones respecto a la teoría de análisis de cluster específicamente bietapico, durante la investigación se propone la clasificación de clientes de telecomunicaciones utilizando esta técnica la cual fue la más apropiada para alcanzar los objetivos de la investigación.

1.7. Limitaciones de la investigación

1.7.1 Espacial

La investigación se realizó en el territorio peruano.

1.7.2 Temporal

De acuerdo con la ficha técnica de la encuesta residencial de servicios de telecomunicaciones el recojo de datos fue de 20 de octubre del 2021 al 13 de diciembre del 2021 y la presente investigación se realizó entre el año 2023 y 2024.

1.7.3 Teórica

Se tomó la teoría de clúster bietápico y análisis de clúster en general, además de la teoría de segmentación de clientes, esto con la finalidad de darle el respaldo científico teórico a la investigación.

CAPÍTULO II

MARCO TEÓRICO

2.1 Antecedentes de la investigación

2.1.1 Antecedentes internacionales

Rubio y Vilá (2016) en su investigación para la universidad de Barcelona titulado *“El análisis de conglomerados bietápico o en dos fases con spss”* indica que esta técnica de clasificación es una instrumento de exploración que se utiliza para encontrar los grupos formados de manera natural dentro del grupo de datos, generando reglas de discriminación de datos, cantidad de agrupaciones y estudio descriptivo respecto a cada conglomerado hallado y de las principales variables. Este método está conformado por dos etapas, cada una tiene una característica diferente respecto a otros métodos de agrupación como buscar el número óptimo de conglomerados, crear modelos de conglomerados con variables categorías y continuas, además trabajar con gran cantidad de datos.

Espigares y otros (2017) en el estudio titulado *“Uso del método bietápico en el estudio de los procesos de enseñanza y aprendizaje musical a través de moodle”* tuvo por objetivo demostrar la utilidad de las técnicas procedentes del big data para analizar los datos educativos musicales en cursos masivos en línea, como parte de

la muestra en estudio participaron 10 centros educativos de Sevilla en España con 741 estudiantes entre los 12 y 13 años. Concluye que el modelo bietápico delimita el conocimiento musical en alto, medio y bajo, en su mayoría el alumnado se agrupa en un clúster con un nivel de conocimientos musicales agrupados en los clústeres 1 y 2 luego un menor índice de alumnado con bajo nivel de conocimiento musical perteneciente al clúster 3.

González (2021) en su investigación titulada *“Estudio de apreciación del consumidor de plataformas de streaming de video bajo demanda por suscripción”* tuvo por objetivo general analizar la opinión de los consumidores respecto a las plataformas de streaming en Chile mediante métodos de investigación de mercados para definir las características de cada empresa, su valor y emitir juicios de acuerdo a lo hallado, es así que realizaron una encuesta a través de survey monkey, dicha encuesta estuvo conformada por preguntas que incluían los hábitos y comportamientos de los encuestados, la muestra estuvo conformada por 467 personas. La conclusión a la que llegaron fue que los consumidores se dividen en 3 grupos denominados medio, moderado e intensivo.

2.1.2 Antecedentes nacionales

Billadoni (2021) en su investigación titulada *“Clustering de clientes de un grupo de e-marketplace”* buscó obtener clústeres de acuerdo con las principales características de compra de un grupo de clientes de e-marketplace; para ello realizó una investigación aplicada de diseño no experimental transversal y descriptivo. La

población estuvo conformada por los clientes que realizaron compras a través de e-marketplace de los cuales, se seleccionó una muestra a 35 881 clientes que realizaron compras desde marzo a julio del 2021. Para la conformación de los clústeres se utilizó la técnica de clustering bietápico esto debido a la naturaleza de las variables, se formaron dos clústeres de clientes regulares de e-marketplace de agrupación favorable, el primer clúster estuvo conformado por el 31% de clientes y el otro grupo conformado por el 69% restante.

Tonconi (2021) en su investigación titulada *“Caracterización de perfiles en centros de educación técnico-productiva públicos utilizando indicadores de calidad a través del método clúster bietápico”* tuvo como objetivo principal identificar conglomerados de los centros de educación técnico-productiva públicos usando indicadores de condiciones básicas de calidad mediante clúster bietápico , para lo cual utilizó un tipo de investigación explicativo , de diseño cuantitativo y no experimental. La población estuvo conformada por 704 cetpros públicos a nivel nacional, la conclusión a la que arribo es que se puede clasificar en dos conglomerados con cetpros públicos con variables diferenciadoras entre cada grupo.

Ortega (2020) en su investigación titulada *“Clasificación, eficiencia técnica y desempeño económico de los productores de quinua del departamento de Puno”* tuvo por objetivo principal evaluar la clasificación de los productores de quinua en el departamento de Puno con la finalidad de identificar grupo o conglomerados , revisar

su eficiencia técnica y desempeño económico, para lo cual se realizó una investigación de corte transversal ya que la técnica de recolección de datos fue la encuesta a través de un cuestionario de preguntas, además fue del tipo explicativa. La población de estudio estuvo conformada por 67 574 productores de quinua de toda la región Puno, de los cuales se tomó una muestra proporcional de acuerdo con cada región siendo el tamaño de esta de 416. La conclusión a la que llegaron fue a seleccionar a 2 conglomerados el primero que estaba formado por el 79% de los productores y el segundo por los restantes.

Tang y Vargas (2016) propusieron una investigación titulada *“Segmentación de clientes de una tienda de electrodomésticos utilizando el análisis de conglomerados”* el objetivo de aplicar el análisis de conglomerados bietápico a clientes de una tienda retail para luego proponer estrategias de marketing a cada conglomerado. El diseño de la investigación es no experimental de tipo descriptivo teniendo como población 6 284 datos luego de la limpieza de datos se analizó 4980 datos. Concluyendo con el establecimiento de 3 conglomerados, obteniendo un perfil de usuario para cada uno, adicionalmente se propusieron estrategias de marketing para cada perfil de usuario.

Rodriguez y Romero (2021) en su investigación titulada *“Identificación de los perfiles del clientes omnicanal de la categoría moda en las tiendas por departamento ubicadas en Lima Metropolitana, en base a los canales utilizados durante el proceso*

de compra”, buscaron describir los diferentes perfiles de clientes omnicanales de acuerdo con la categoría de moda de las tiendas en Lima esto a través de los medios que utilizan frecuentemente para realizar sus compras. Para esto los autores realizaron una investigación de alcance descriptivo, de enfoque cuantitativo, diseño no experimental, la muestra escogida fue de 250 personas. Concluyen que los clientes omnicanales de las tiendas por departamento en Lima metropolitana se puede dividir en cuatro segmentos definidos como tradicionales, innovadores, entusiastas y funcionales.

2.1.3 Antecedentes regionales y locales

Vera y Soncco (2016) en su investigación *“Algoritmo de optimización basado en el comportamiento social de arañas para clustering”* tuvo como objetivo clasificar a las arañas de acuerdo con su comportamiento por lo que se utilizó un algoritmo de optimización tomando en cuenta la suma de las distancias euclidianas. El algoritmo fue ejecutado reiteradas veces y luego se realizó el cálculo del promedio, mediana, mínimo y máximo valor de los resultados encontrados en conclusión se comprobó la efectividad del algoritmo superando en precisión a las del algoritmo k-means estas fueron corroboradas a través de la prueba de Wilcoxon a las salidas del algoritmo.

Mamani y Damiano (2014), en la investigación realizada *“Satisfacción laboral en el sector construcción mediante análisis factorial y análisis de clúster”* busco clasificar las variables relacionadas a la satisfacción del cliente, las cuales son de

naturaleza multidimensional. Utilizaron el análisis de clúster con la finalidad resumir la gran cantidad de complejas relaciones que se pueden encontrar en dichas variables, luego de este proceso se procedió a determinar los niveles de satisfacción realizando el análisis de acuerdo con las variables sociodemográficas y características laborales. El resultado fue que el nivel de satisfacción laboral de los trabajadores de Cusco se puede clasificar en dos grandes grupos el primer conglomerado conformado por el 51.40% de los encuestados tiene un nivel de satisfacción buena y el 13.5% tienen un nivel de satisfacción alta, siendo estos los dos conglomerados de mayor tamaño.

Pérez (2017) en su investigación titulada “*Clúster ferretero en la ciudad de Urubamba Cusco 2017*” propuso un estudio descriptivo no experimental de enfoque cuantitativo, el cual recopiló información de 20 empresas en la ciudad de Urubamba del sector ferretero con la finalidad de someterlos a un cuestionario respecto para así recabar información de estas y poder generar agrupaciones entre ellas. El resultado al que llegaron fue que el clúster ferretero es deficiente ya que las características como la interdependencia empresarial, flujo de productos y vinculación empresarial son deficientes.

2.2 Bases teóricas

2.2.1 Segmentación de mercados

Valderrey (2011) indica que la segmentación de mercados es un proceso de

clasificar el mercado en conglomerados uniformes de menor medida que contengan características y necesidades parecidas. Este concluye que el mercado está segmentado en grupos homogéneos, esto dentro de cada grupo lo cual es útil para aplicar estrategias de marketing.

Best (2007) propone que la segmentación de mercados forma parte importante de una empresa que se enfoca en satisfacer las necesidades del mercado y así conseguir una mayores ganancias y aumento de la productividad del marketing. Esto conlleva reconocer conjuntos determinados de clientes, con requerimientos y conductas de compra propios, y caracterizarlos demográficamente que los diferencian entre grupos.

Valderrey (2011) menciona que, una adecuada segmentación tiene homogeneidad dentro de sí mismo y diferente entre los segmentos, permanencia de segmentos, los cuales deben de ser determinables y cuantificables, también deben ser asequibles y adaptables, ser grandes para ser rentables. Así también se indica que se puede segmentar a los clientes de acuerdo con las variables demográficas, geográficas, psicográficas y de comportamiento. Cuando se analizan muchas variables es denominada segmentación profunda dado que pueden aportar mayor conocimiento.

2.2.2 Criterios para la segmentación de clientes

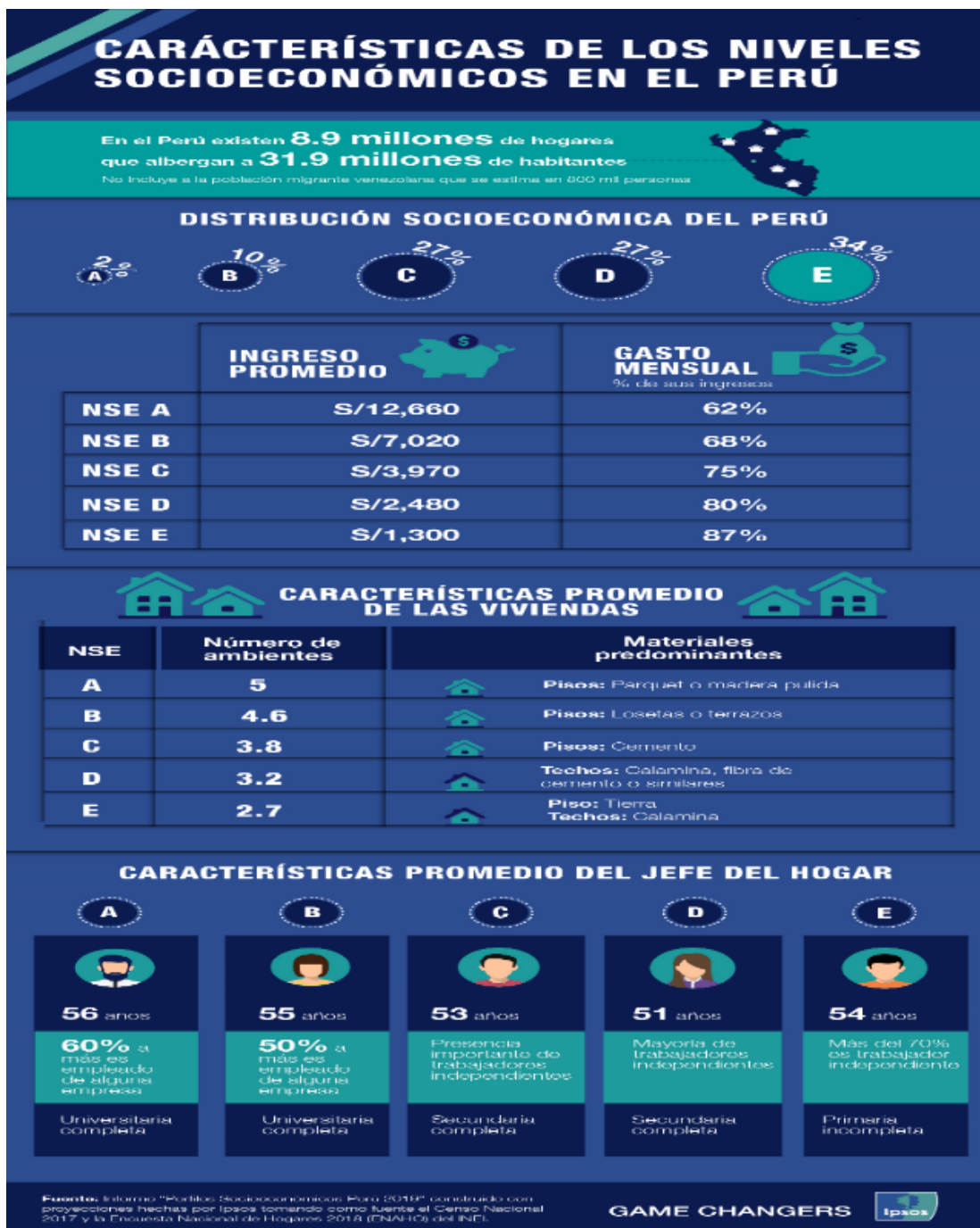
Según Oporto y otros (2018), en el mercado de bienes y servicios de consumo, la segmentación del producto mercado por segmentos homogéneos se puede realizar de diferentes maneras, una de ellas es por las características de los consumidores

dentro de ellas se tienen las generales como características demográficas, sexo, edad, nivel socio económico y las específicas como estructura de consumo, uso del producto, fidelidad a la marca, tipo de compra, gasto mensual en el producto o servicio, años de uso.

Características de los niveles socioeconómicos en el Perú: según Sanchez (2023), se puede definir como la clasificación que ayuda a clasificar a los hogares de acuerdo con las características del jefe de hogar, tenencia de bienes y servicios además de los materiales predominantes de la vivienda siendo 5.

Figura 1

Características de los niveles socioeconómicos en el Perú.



Nota: Tomado del informe "Perfiles socioeconómicos Perú 2019" construido con proyecciones hechas por Ipsos tomando como fuente el censo nacional 2017 y la encuesta nacional de hogares 2018 (ENAH0) del INEI.

2.2.3 Análisis exploratorio de datos (AED)

Esta técnica tiene la finalidad evaluar los datos antes de realizar cualquier análisis estadístico, y así procurar que el investigador tenga un panorama más claro de los datos y su comportamiento respecto a las otras variables.

El AED nos brinda técnicas sencillas para poder estructurar los datos, buscar fallos en la forma y técnica de recolección de estos, además de detectar atípicos y faltantes, por último, confirmar los supuestos implícitos en gran parte de las técnicas multivariantes.

2.2.4 Etapas del análisis exploratorio de datos.

Según Salvador y Gargallo (2003), las etapas del análisis exploratorio de datos son:

1. Acondicionar los datos para manejarlos de manera práctica y sencilla para cualquier técnica estadística.
2. Analizar de manera gráfica el comportamiento de los datos por variable individual a explorar y un resumen descriptivo numérico que conceda cuantificar algunas características gráficas de los mismos.
3. Analizar de manera gráfica las relaciones entre las variables de estudio y un resumen descriptivo numérico que nos permita cuantificar la fuerza de la relación entre las variables.
4. Observamos los requisitos estadísticos previos para algunas técnicas más complejas denominadas supuestos, como normalidad, linealidad y

homocedasticidad.

5. Analizamos datos con comportamiento atípico (outliers) y revisamos su influencia en posteriores análisis estadísticos.
6. Analizamos datos ausentes y revisamos su influencia en posteriores análisis estadísticos. La mejor manera para explicar el comportamiento de una variable es mediante un gráfico. Los gráficos de barras y de sectores especifican la distribución de las variables cualitativas. Respecto a las variables cuantitativas se explican de mejor manera mediante histogramas, diagramas de tallos y hojas además del diagrama de cajas.

2.2.5 Diagramas de sectores circulares.

Córdova (2003) indica que este tipo de diagrama es del tipo circular en el que se divide en sectores cada categoría de la variable cualitativa de acuerdo con la siguiente fórmula: $h_i \times 360^\circ$ donde h_i es la frecuencia relativa de la categoría i de la variable en estudio y el resultado sería el ángulo en el que se divide cada sector.

2.2.6 Diagrama de barras

Utilizada para describir variables del tipo cualitativa, graficamos en barras rectangulares vertical u horizontal cuya altura es proporcional a su frecuencia además las barras se separan unas de otras con espacios.

En el eje horizontal se consideran los valores (cualidades) de la variable cualitativa. Tomar en cuenta que si la variable es del tipo cualitativa ordinal necesariamente se debe considerar el orden en el eje horizontal respecto a la

variable cualitativa nominal puede ir en cualquier orden (Córdova, 2003).

2.2.7 Histograma.

Representa la distribución de variables del tipo cuantitativa, este grafico está conformado por:

- Título, que representa la población o una parte de ella que es la muestra de estudio.
- El eje vertical, en el que se encuentra las frecuencias de las clases.
- El eje horizontal, en que se encuentra las clases o intervalos que representa la variable X (Johnson y Kuby, 2012).

2.2.8 Diagrama de tallo y hojas

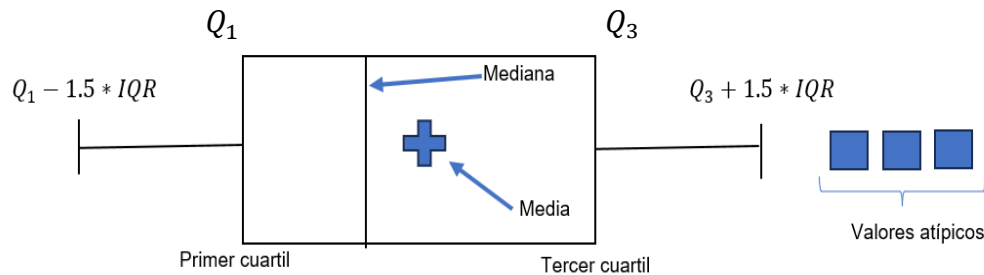
El diagrama de tallos y hojas estructura la data en clases (tallos), para que los datos de cada clase (hojas) se ramifiquen de izquierda a derecha de cada fila. En el gráfico podremos ver el comportamiento de los datos y donde se generan la mayor aglomeración de datos (Levin y otros, 2004).

2.2.9 Diagrama de caja y bigotes

Es una representación gráfica de los datos mediante la recolección de 5 números.

Figura 2

Diagrama de cajas y bigotes



Nota: Tomado de Llinás y Rojas (2006). *Estadística descriptiva y distribuciones de probabilidad*. Barranquilla (Colombia)

La Figura 2, representa el diagrama de tallos y hojas el cual permite realizar un análisis exploratorio de los datos y analizar su simetría, su distribución y los valores extremos o atípicos de una determinada distribución de frecuencias.

El diagrama de caja y bigotes se particiona en cuatro partes que tienen igual frecuencia, una línea vertical divide en dos; la parte central de esta y otras dos áreas denotadas como dos líneas horizontales denominados bigotes, los cuales inician en la parte central de la caja.

La caja central contiene la mitad de los datos, en el interior de la caja central se encuentra la media (promedio) que se denota con un signo positivo (+) y la mediana se representa como una línea vertical en la caja.

En el caso de que la línea no se ubica en la parte central de la caja, no existe simetría de los datos.

Desde el centro vertical de la caja inicia los bigotes a cada lado, tomando en cuenta lo siguiente:

- El bigote izquierdo inicia de izquierda a derecha con el valor $Q_1 - 1.5 RI$ y luego

está el cuartil Q_1 , donde R.I es el rango inter cuartil.

- Respecto al bigote de la derecha tiene dos límites, uno de ellos es Q_3 y el otro límite es $Q_3 + 1.5 RI$, viéndolo de izquierda a derecha.
- Si encontramos datos a la derecha o izquierda de los bigotes se les nombra datos atípicos u outliers.

2.2.10 Detección de outliers multivariantes

Existen diversos métodos para detectar valores atípicos dentro de ellos se encuentra la distancia de Mahalanobis.

Esta técnica estadística calcula la distancia de cada uno de los datos con el centro de masa, y requiere de los parámetros estimados de la distribución multivariante.

Para datos multivariantes que tienen comportamiento normal los datos de la distancia de Mahalanobis tiene aproximadamente una distribución chi – cuadrado cuyos grados de libertad son el número de variables con los que estamos trabajando, por lo tanto, las observaciones con una longitud alta de Mahalanobis corresponderían a valores atípicos.

Distancia de Mahalanobis

$$MSD_i = \sqrt{(x_i - \bar{x})^T \cdot S_n^{-1} (x_i - \bar{x})} \quad (1)$$

donde:

$(x_i - \bar{x})^T$: Matriz transpuesta.

\bar{x} : Media del vector muestral.

S_n : Matriz de covarianza muestral.

Matriz de covarianza muestral

$$S_n = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)(x_i - \bar{x}_n)^T \quad (2)$$

Si la distancia de Mahalanobis es 0 es que el dato se encuentra en el centro de masa y es mayor a cero significa que el dato está alejado del centro de masa.

Los datos con mayores valores tienen menores pesos, por último, se realiza una regresión ponderada para minimizar el efecto de los valores extremos.

Si los datos son multivariantes distribuidos normalmente, los datos de las distancias de Mahalanobis tienen aproximadamente una distribución chi-cuadrado con “p” grados de libertad. Por consiguiente, aquellos datos con distancia de Mahalanobis con valores altos nos muestran a valores atípicos.

Las consecuencias de enmascaramiento y empantamiento tiene un papel importante en la idoneidad de la distancia de Mahalanobis como dictamen de valores atípicos. Las consecuencias de enmascaramiento suelen reducir la distancia de Mahalanobis de un valor atípico.

Quaglino y Merello (2012) mencionan que las consecuencias de enmascaramiento suelen reducir la distancia de Mahalanobis de un valor atípico. Esto suele suceder por ejemplo cuando un grupo reducido de datos atípicos captar a x_n e infla S_n alejado del comportamiento de la mayoría de las observaciones.

Los problemas de enmascaramiento y empantamiento se solucionan utilizando estimaciones robustas, como por ejemplo el estimador M multivariado, el estimador S cuadrático, el estimador de covarianza de mínimo determinante (MCD),

entre otros, los cuales por definición son influidos por los datos atípicos, siendo poco probable que contribuya los parámetros utilizados en el MSD (Muñoz y Amón,2013).

2.2.11 Análisis de conglomerados.

McDaniel y Gates (2005) mencionan que el análisis de conglomerados se puede definir como métodos estadísticos que se utilizan para discriminar objetos o personas que tienen alguna característica en común respecto a ciertas mediciones o variables comunes. La finalidad del análisis de conglomerados es que son mutuamente excluyentes y colectivamente exclusivos para procurar así similitud dentro del grupo y diferencias entre grupos.

Según Pérez (2011), existe técnicas de clasificaciones de conglomerados:

- Técnicas aglomerativas - divisivas: Este método va juntando tanto grupos como elementos de manera consecutiva va agrupando los dos grupos más homogéneos, al punto en el que se define una clasificación determinada, en comparación con el método divisivo inicia con un grupo general que está conformado por todos los individuos, y en cada ciclo anterior a agrupando uno a uno los individuos de los grupos ya formados con anterioridad, así va conformando nuevos grupos.
- Métodos jerárquicos- no jerárquicos: Radica es una sucesión de $g + 1$ clúster: G_0, \dots, G_g en la que G_0 es la división disjunta de todos los elementos y G_g es el conjunto partición. Sucesivamente, la cantidad de divisiones se reduce, en consecuencia, las mismas aumentan en tamaño y son más homogéneas, En contraparte, el método nos jerárquico se conforman grupos homogéneos sin

tener relación de orden o jerarquías entre grupos. Para este método debemos saber inicialmente el número de conglomerados que se necesita.

- Métodos solapados- exclusivos: Este método tolera que un elemento pertenezca a dos grupos a la vez en algún momento de la clasificación. Este método si impide a los elementos forme parte de dos grupos a la vez en el mismo momento.
- Métodos secuenciales- simultáneos: El método secuencial se utiliza el mismo algoritmo de manera recursiva grupo a grupo, entre tanto el método poli técnico ordena los objetos, no necesariamente tienen las mismas características, pero tienen las necesarias como para lograr explicar la similitud entre los miembros dentro del grupo.
- Técnicas directas e iterativas: Esta técnica utiliza algoritmos en los que se elige un elemento y un grupo ya se no se puede mover del mismo, luego los métodos iterativos verifican en cada iteración si la clasificación de uno elemento en un conglomerado es ideal realizando un nuevo reagrupamiento de los elementos si así se requiere.
- Técnicas ponderadas y no ponderadas: La técnica no ponderada proporciona el mismo peso a todas las cualidades (o variables) de los elementos a clasificar, en tanto el método ponderado otorga mayor peso en ciertas características.
- Técnicas adaptativas y no adaptativas: La técnica no adaptativa, el algoritmo seleccionado se encamina a un resultado, la cual tiene un método de conformación de conglomerados es anterior y fijo, luego el método adaptativo es

el que mientras se van formando los conglomerados, adquiere conocimientos de la formación de estos y modifica el criterio de optimización o la medida de homogeneidad a utilizar.

Las técnicas de conglomerados más populares son los secuenciales, aglomerativos, jerárquicos y exclusivos el cual forma el acrónimo en inglés de S.A.H.N. (sequential, agglomerative, hierarchic y nonoverlapping), las cuales son: De unión simple, de distancia máxima, de la media, de la media ponderada, de la mediana, del centroide y de mínima varianza.

2.2.12 Análisis de conglomerados bietapico.

Recibe el nombre dado que es un proceso que se lleva a cabo en dos etapas. Las ventajas de este, es que podemos explorar grupos de conglomerados de un conjunto de datos, que no sería posible de averiguar mediante otro método. Otra de las características de esta técnica es que permite trabajar con variables de naturaleza categórica y continúa asumiendo independencia entre ellas (McDaniel y Gates, 2005).

Además, nos permite elegir instintivamente el número de conglomerados además de la facilidad de trabajar con grandes grupos de datos.

- Como ya mencionamos, el algoritmo trabaja con variables cuantitativas y categóricas.
- Es importante el orden de los datos por lo que se recomienda ordenar de manera aleatoria para evitar el impacto del orden de los datos.

Se da por entendido que las variables siguen una distribución normal y las

categorías siguen una distribución multinomial luego las variables también son independientes. Es importante mencionar que el método de clasificación bietápico es robusto ante los quebrantamientos de los supuestos de distribución e independencia (McDaniel y Gates, 2005).

2.2.13 Distancia para variables cualitativas y cuantitativas.

Salinas (2010) menciona que la medida de la distancia nos ayuda a calcular la homogeneidad entre dos conglomerados. Esto porque el método trabaja según la homogeneidad, por lo que se busca una medida de distancia que nos permite ligar el efecto de las variables cualitativas y cuantitativas

El algoritmo *k-prototype* es un método popular y muy usado para medir la distancia entre atributos mixtos.

El mencionado algoritmo recibe de la influencia de k-medias. Para los datos numéricos ocupamos la distancia euclidiana y para los categóricos se utiliza distancia Binaria, este último es perjudicial para la mencionada medida.

Seguidamente mencionamos otras medidas para atributos mixtos, ocupada para la siguiente investigación:

- Distancia log-verosimilitud: Esta procura la reducción del algoritmo de verosimilitud como consecuencia de vincular a dos grupos conjuntamente. La distancia entre 2 clúster está dada como:

Distancia log-verosimilitud

$$d(i; j) = \xi_i + \xi_j - \xi_{(i,j)} \quad (3)$$

donde:

$$\xi_v = -N_v \left\{ \sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{vk}^2) + \sum_{k=1}^{K^B} \hat{E}_{vk} \right\} \quad (4)$$

Para $v = i, j$

$$\hat{E}_{vk} = - \sum_{l=1}^{L_k} \frac{N_{vkl}}{N_v} \log \frac{N_{vkl}}{N_v} \quad (5)$$

las notaciones son:

N_k : Cantidad de datos en el clúster v .

K^A : Cantidad de variables continuas.

K^B : Cantidad de variables categóricas.

$\hat{\sigma}_k^2$: Varianza estimada de la k -ésima variable continua.

$\hat{\sigma}_{vk}^2$: Varianza estimada de la k -ésima variable continua en el clúster v .

L_k : Cantidad de categorías para la k -ésima variable cualitativa.

N_{vkl} : Cantidad de datos v .

Antes de reseñar el algoritmo, explicaremos dos definiciones importantes para la comprensión de esta. Las definiciones a continuación se fundamentan en la investigación de (Zhang y otros, 1996) del algoritmo BIRCH.

2.2.14 Caracterización de los conglomerados.

Zhang y otros (1996) indican que existe un grupo de datos de N casos con d dimensiones en el clúster denominado $\{\bar{X}_i\}$, donde $i=1,2,3, \dots, N$. El CF se traza como un vector que tiene las siguientes medidas:

Vector compuesto

$$CF = (N, \overline{LS}, SS) \quad (6)$$

donde N es la cantidad de casos en un clúster, \overline{LS} es la suma lineal de los N casos en un clúster y SS es la adición de los cuadrados de los N casos en un clúster.

Una de las propiedades de aditividad, por lo que obedece a lo siguiente:

Propiedades de aditividad

$$CF_1 = (N_1, \overline{LS}_1, SS_1) \quad (7)$$

$$CF_2 = (N_2, \overline{LS}_2, SS_2) \quad (8)$$

Sumando estas dos propiedades se obtiene:

$$CF_1 + CF_2 = (N_1 + N_2, \overline{LS}_1 + \overline{LS}_2, SS_1 + SS_2) \quad (9)$$

De la teoría del CF y su propiedad de aditividad se desglosa que los vectores CF pueden guardar estadísticas de resumen de cada conglomerado, este método nos ayuda para realizar el análisis con grandes grupos de datos, esto porque no asocia o junta caso por caso, lo que hace es juntar las medidas estadísticas de resumen (N, \overline{LS}, SS) de cada grupo.

Con lo que contaríamos con información suficiente para poder agrupar los datos con el algoritmo BIRCH.

2.2.15 Árbol de características de conglomerados (CF TREE).

El CF Tree se constituye con la información dada del CF. Este es un árbol equilibrado que se basa en dos parámetros: B que es un factor de ramificación y un

umbral T.

De primera mano contamos con un nodo sin hojas con la forma: $[CF_i; hijo_i]$, donde $i = 1, 2, 3, \dots, B$; además $hijo_i$ es un indicador para el i -ésimo nodo hijo y CF_i , es el CF del subgrupo denotado por este $hijo_i$. Es por ello que, un nodo sin hojas es un clúster conformado por sus entradas. Un nodo hoja está conformado de L entradas de la siguiente manera CF_i donde $i = 1, 2, 3, \dots, L$. Cada nodo hoja está conformado por dos indicadores o punteros llamados “prev” y “next” utilizados para unir todos los nodos hojas. Un nodo hoja es un clúster conformado por todos los representados por sus ingresos. Aunque los ingresos de los nodos hojas tienen que justificar el parámetro umbral L.

Es importante resaltar que la longitud del árbol es una función T, sucede que, si cuanto mayor es el umbral, cuanto menor es el tamaño.

Nodo a nodo del CF tree muestra la cantidad de grupos que corresponden a esa ramificación y sus respectivos CF.

2.2.16 Algoritmo de conglomerados en 2 etapas

Recibe el nombre debido a que se compone de dos partes:

Paso 1: Recibe el nombre de preagrupamiento dado que en esta etapa se forman los pre - conglomerados. El cual empieza con el desarrollo del CF tree, este asigna la primera observación en un nodo hoja que almacena información sobre la misma. Luego se aumenta cada caso a un nodo hoja existente o se forma uno nuevo, si la longitud de similitud (distancia log verosimilitud) es próxima a un nodo hoja que ya exista se le asocia a él, caso contrario se inicia un nuevo nodo hoja.

Cuando culmina el paso uno, se tendrá construido los pre - clústeres mediante el CF tree, estos guardan las estadísticas de resumen $CF = (N, \overline{L\vec{S}}, SS)$

Paso 2: Luego de constituido el CF tree en el anterior paso, los nodos hojas se juntan usando un algoritmo aglomerativo de clúster. Dicho algoritmo aglomerativo nos proporciona una cantidad de soluciones o modelos. Para seleccionar la opción óptima se puede discriminar las diferentes opciones con el criterio bayesiano de Schwartz / BIC) o el criterio de Información de Akaike (AIC).

El método utilizado es del BIC, el cual arroja instantáneamente el número óptimo de conglomerados.

La relación es la siguiente:

Método BIC

$$BIC(J) = -2 \sum_{j=1}^J \xi_j + m_j \log(N) \quad (10)$$

Luego:

$$m_j = J \left\{ 2K^A + \sum_{k=1}^{K^B} (L_K - 1) \right\} \quad (11)$$

N: Número total de casos.

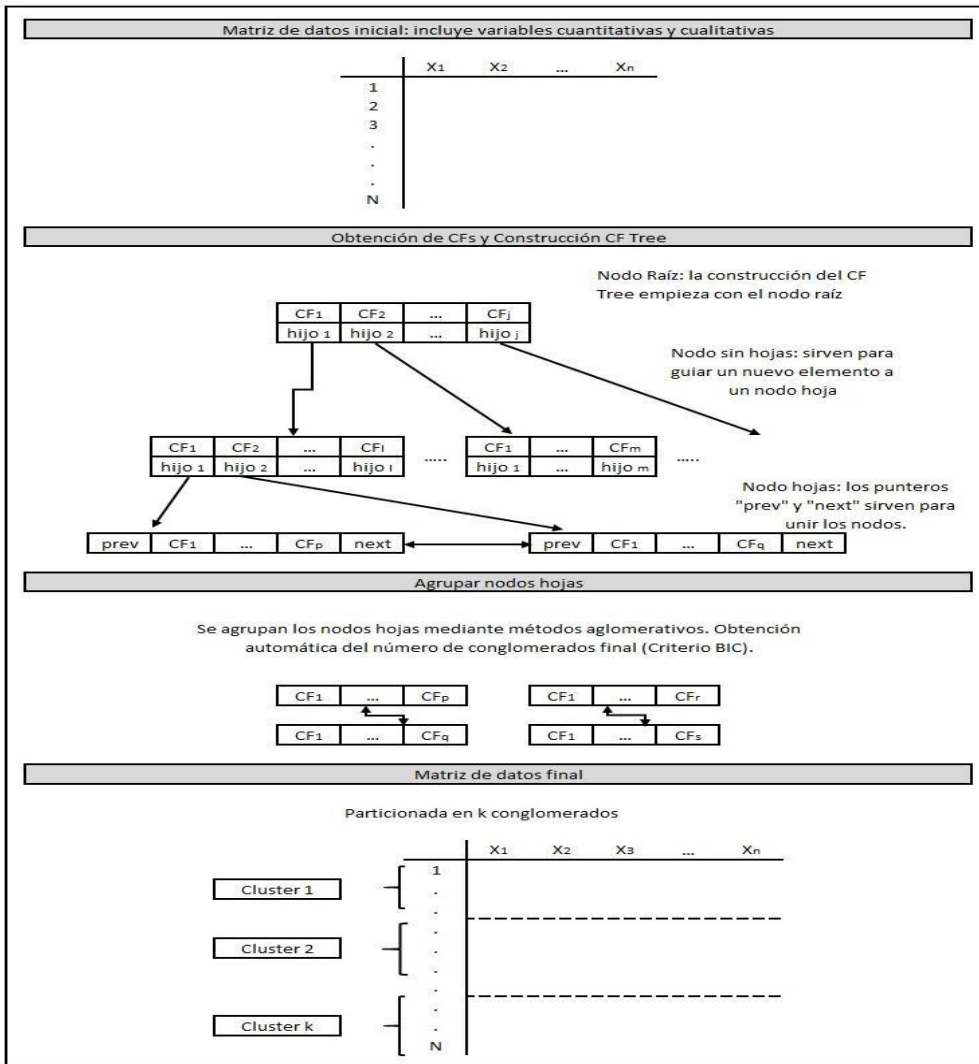
K^A : Cantidad de variables categóricas

L_K : Cantidad de categorías para la k -ésima variable cualitativa.

Es importante mencionar que el presente método para el estudio de conglomerados es muy robusto respecto al quebrantamiento del supuesto de independencia cono de las distribuciones.

Figura 3

Esquema gráfico del Conglomerado Bietápico



Nota: Tomado de Tang, F., y Vargas, C. (2016). Segmentación de clientes de una tienda de electrodomésticos utilizando el análisis de conglomerados. [Tesis de pregrado, Universidad Nacional Agraria La Molina]. <https://hdl.handle.net/20.500.12996/2214>. La Figura representa los pasos seguidos para obtener los conglomerados con los métodos de clúster Bietápico.

CAPÍTULO III

METODOLOGÍA DE LA INVESTIGACIÓN

3.1 Identificación de las variables de investigación

Variable independiente: Características socio económicas del usuario de telecomunicaciones

Dimensiones

- Sexo,
- Edad, Nivel de instrucción
- Ingresos principales
- Nivel socioeconómico
- Área
- Inversión mensual en el servicio móvil,
- Tiempo con el servicio de internet

Variable dependiente: Perfiles de cliente

Dimensiones

- Perfil 1
- Perfil 2
- Perfil 3
- Perfil 4

3.2 Operacionalización de variables.

Tabla 1

Operacionalización de variables

Variables	Definición conceptual	Definición operacional	Indicadores	Escala
Características socio económicas del usuario de telecomunicaciones	Son las características de un individuo que se refieren a las condiciones y variables que lo posición dentro de la sociedad de acuerdo con el aspecto social y económico (Vera y Vera, 2013).	Para la presente investigación se utilizó los indicadores como sexo, edad, nivel de instrucción, ingresos principales, nivel socio económico, área, pago mensual del servicio móvil y tiempo con el servicio de internet como características de los usuarios de telecomunicaciones extraídos de la Encuesta Residencial de servicios de telecomunicaciones (ERESTEL).	Sexo	Nominal
			Edad	De razón
			Nivel de instrucción	Ordinal
			Ingresos principales	Ordinal
			Nivel socio económico	Ordinal
			Área	Nominal
Perfiles de cliente	El perfil de un consumidor o cliente es el conjunto de características innatas a él, que está conformado por la edad, sexo y demás aspectos que lo caractericen (Brito y Vergara, 2022).	De acuerdo con Rodríguez y Romero (2021), los clientes se pueden clasificar en 4 conglomerados de acuerdos a sus características socio económicas y de consumo.	Pago mensual en el servicio móvil	De razón
			Tiempo con el servicio de internet	De razón
			Perfil 1	
			Perfil 2	
			Perfil 3	
			Perfil 4	

Nota: Elaboración propia de la autora, representa las variables en estudio junto a la definición conceptual y operacional, así como sus respectivos indicadores y la escala de estos.

3.3 Descripción de la metodología

3.3.1 Alcance de la investigación

La presente investigación fue de tipo descriptivo y explicativo. De acuerdo con Hernández y otros (2018), los estudios descriptivos tienen como objetivo describir las propiedades, características y perfiles de personas, grupos, comunidades u objetos, es decir solo busca medir o recoger información acerca de las variables de estudio y una investigación explicativa

3.3.2 Enfoque de la investigación

El enfoque de la investigación es cuantitativo, de acuerdo con Hernández y otros (2018) una investigación de cuantitativa hace uso de los datos recabados para su análisis con la finalidad de responder a las preguntas de investigación y probar las hipótesis planteadas, la investigación utilizará las técnicas estadísticas para establecer el comportamiento de una determinada población en estudio.

3.3.3 Diseño de la investigación

Para desarrollar el trabajo de investigación el diseño a utilizar será no experimental, esto debido a que las variables no fueron modificadas durante la investigación. De acuerdo con Hernández y otros (2018) este en este diseño de investigación se observa los sucesos en su forma natural para estudiarlos.

3.4 Población y muestra

3.4.1 Población

La población se entiende como el total de unidades de análisis que conforman el

objeto de estudio y que deben cuantificarse, también puede ser conceptualizado como por la cantidad de elementos que participan de una característica en especial Tamayo (2012).

La población de la presente investigación fueron todos los participantes en la encuesta ERESTEL 2021, de acuerdo la base de datos de personas son un total de 47 346 encuestados.

3.4.2 Muestra

La muestra es la agrupación de operaciones que se hacen con el fin de estudiar la distribución de ciertos caracteres en totalidad de una población, iniciando en la observación de una parte representativa (Tamayo, 2012). El cálculo de la muestra se realizará posterior al análisis exploratorio de datos, proceso que se detalla a continuación:

Población	47 346 encuestas
Mediante el proceso de detección de los datos faltantes y valor atípicos se obtuvo que se tiene:	
Datos perdidos	31 385 encuestas
Datos atípicos (Distancia de Mahalanobis < 16.660)	181 encuestas
Por lo tanto se procede a eliminar	31 566 encuestas

Por tanto, la muestra está conformada por la diferencia entre el total de las encuestas (Población) 47 346 y los datos eliminados por ser datos perdidos y atípicos que sumaron 31 566 dando un total 15 780 encuestas, que conforman así la muestra.

3.5 Técnica e instrumento de recolección de datos

3.5.1 Técnica

La obtención de datos fue de una fuente secundaria, siendo la técnica de recolección a través de la encuesta realizada por OSIPTEL en el año 2021 cuya denominación es encuesta residencial de servicios de telecomunicaciones (ERESTEL) realizado por, la encuesta es procedimiento mediante el cual se obtienen datos que son de interés para realizar una investigación esto a través de la interrogación a los miembros de un grupo, esta técnica es útil cuando se desea uniformizar la técnica de observaciones y aislar ciertos problemas que son materia de estudio (Feria y otros, 2020).

3.5.2 Instrumento

El instrumento de recolección de datos fue el cuestionario está compuesto por alrededor 425 preguntas (ver anexo II), el cuestionario fue desarrollado por el consorcio asociación benéfica Primsa – Analitik investigación y consultoría S.A.C empresa que fue contratada por OSIPTEL para realizar la encuesta ERESTEL 2021.

El cuestionario es un instrumento estandarizado que se usa para la recogida de datos en el trabajo de campo de algunas investigaciones cuantitativas, principalmente las que usan la técnica de la encuesta y es una herramienta que ayuda al investigador a plantear un conjunto de preguntas que recolecten información aplicada a una muestra de personas, usando la técnica cuantitativa para luego realizar la contrastación estadística de algunas relaciones de intereses (Feria y otros, 2020).

3.6 Técnicas de procesamiento y análisis de datos

Para el procesamiento y análisis de los datos de la encuesta ERESTEL 2021 se utilizó estadística descriptiva a través de gráficos como diagrama de cajas y bigotes para observar el comportamiento de los datos cuantitativos y diagrama de sectores para los de naturaleza cualitativa, para la clasificación de datos se utilizó la técnica de clúster bietápico.

3.7 Procesamiento de datos

1. Primero se procedió a obtener la base de datos de la encuesta ERESTEL 2021 publicado por el OSIPTEL, el cuestionario está compuesto por alrededor 425 preguntas, el número no es exacto ya que muchas preguntas se disgregaron en otras debido a su naturaleza de respuesta múltiple.
2. Se analiza los datos que son relevantes para el estudio, esto de acuerdo con las necesidades de la investigación.
3. Se realiza la limpieza de datos eliminando los valores atípicos y los datos perdidos.
4. Se realiza un análisis exploratorio y descriptivo de las variables; para las variables de tipo cualitativa se describe y grafica mediante diagrama de sectores, para las cuantitativas se utiliza los diagramas de cajas y bigotes.

Recibe el nombre de preagrupamiento dado que en esta etapa se forman los pre - conglomerados. El cual empieza con el desarrollo del CF tree, este asigna la primera observación en un nodo hoja que almacena información sobre la misma. Luego se aumenta cada caso a un nodo hoja existente o se forma uno nuevo, si

la longitud de similitud (distancia log verosimilitud) es próxima aun no da hoja que ya exista, caso contrario se inicia un nuevo nodo hoja. Cuando culmina el paso uno, se tendrá construido los pre - clústeres mediante el CF tree, estos guardan las estadísticas de resumen (ecuación 6).

5. Luego de constituido el CF tree en el anterior paso, los nodos hojas se juntan usando un algoritmo aglomerativo de clúster. Dicho algoritmo aglomerativo nos proporciona una cantidad de soluciones o modelos. Para seleccionar la opción óptima se puede discriminar las diferentes opciones con el criterio bayesiano de Schwartz / BIC) o el criterio de Información de Akaike (AIC).

3.8 Confiabilidad y validez

La confiabilidad del instrumento de recolección de datos es el grado en el que su aplicación repetida al mismo sujeto produce los mismos resultados, la validez de contenido del instrumento de recolección de datos es el nivel de representación de cada ítem que manifiesta un dominio específico de contenido (Espinosa y otros, 2021).

De acuerdo con la ficha de la ERESTEL, la investigación fue realizada por el consorcio asociación benéfica Primsa – Analitik investigación y consultoría S.A.C los cuales brindaron las garantías de la encuesta realizada y esta fue aprobada por el OSIPTEL, por lo que la confiabilidad y validez del instrumento de recolección de datos de la encuesta fue proporcionada por dicho consorcio bajo el aval de OSIPTEL.

La información en general brindada por la consultora consorcio asociación benéfica Primsa – Analitik investigación y consultoría S.A.C, se puede descargar a través de la página de web de datos abiertos plataforma del gobierno que de acuerdo

con el marco de gobernanza de datos del estado peruano (Presidencia del Consejo de Ministros, 2024):

“...El propietario original de este activo es el ciudadano, quien cede su información a las organizaciones del gobierno para recibir un valor agregado sobre ellos. Además, este valor depende del beneficio que se obtiene al utilizarlos para la toma de decisiones (del gobierno y del ciudadano)”

CAPÍTULO IV

RESULTADOS DE LA INVESTIGACIÓN

4.1 Resultados

4.1.1 Análisis exploratorio y descriptivo

El análisis de conglomerados es bastante propenso cuando existen valores extremos, por lo que se realiza un análisis del tipo descriptivo y exploratorio, iniciando con las variables cuantitativas para detectar posibles datos atípicos y/o faltantes.

Tabla 2

Estadísticos descriptivos variables cuantitativas

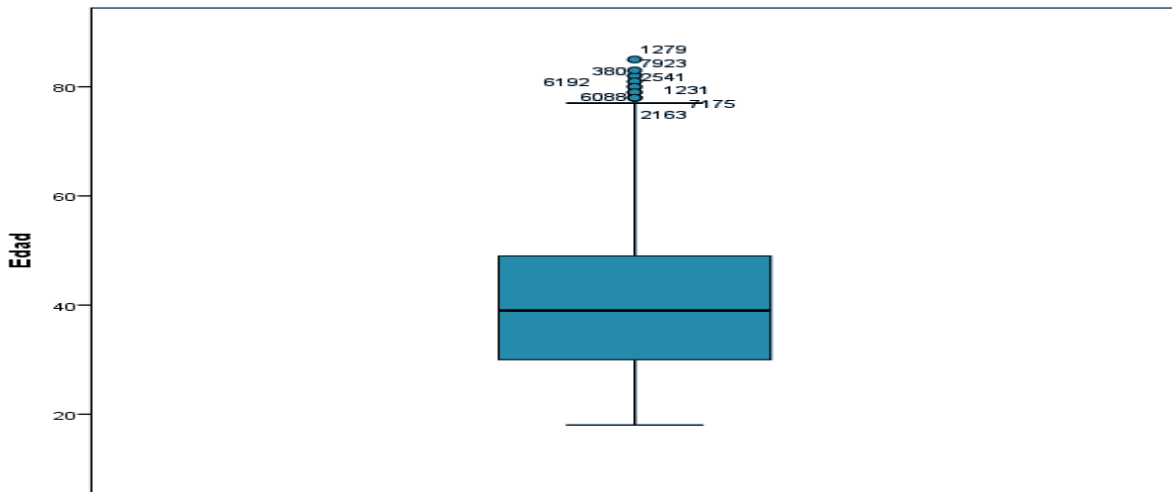
Estadísticos		Edad	Ingreso principal mensual	Pago mensual por telefonía móvil
Media		40.120	1661.360	37.610
Desviación típica		12.340	1 074.812	26.801
Mínimo		18	17	0
Máximo		85	20 000	340
Total	Válidos	15 995	15 995	15 995
	Perdidos	31 385	31 385	31 385

Nota: Datos tomados de la Encuesta residencial de telecomunicaciones (2021).

En la Tabla 2, se observa que el total de datos perdidos asciende a 31 385 datos por lo que se procede a eliminar estos, pues podría sesgar el análisis de clúster bietápico. Luego de ello queda 15 995 encuestas, las cuales se someterán a un análisis gráfico mediante los diagramas cajas y bigotes para luego calcular la distancia de Mahalanobis para cada encuesta, pudiendo obtener así la cantidad de valores atípicos los cuales procederemos también a eliminar.

Figura 4

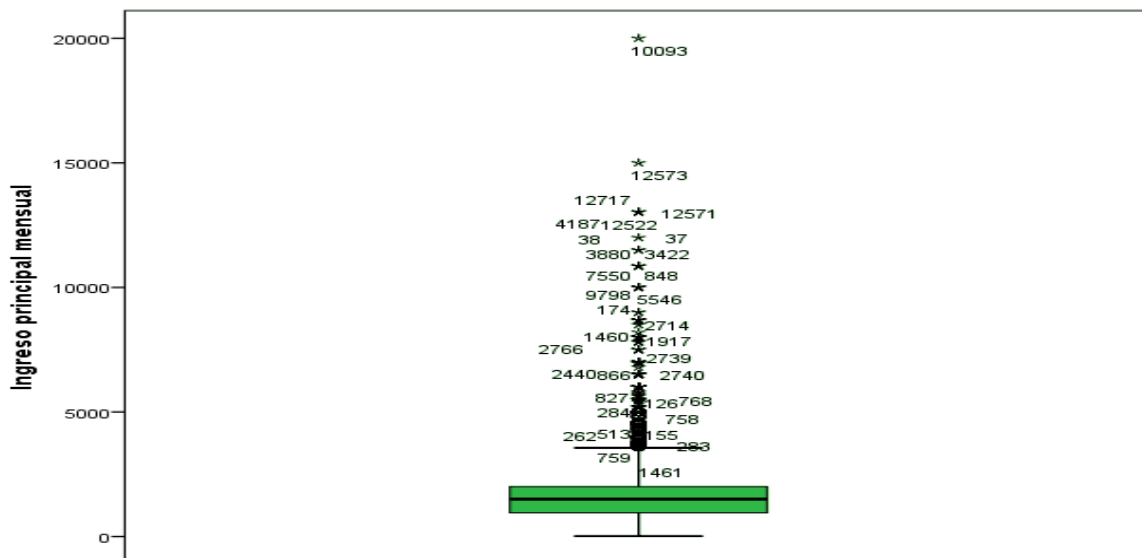
Diagrama de caja y bigotes de edad



Nota: Datos tomados de la encuesta residencial de telecomunicaciones (2021).

Figura 5

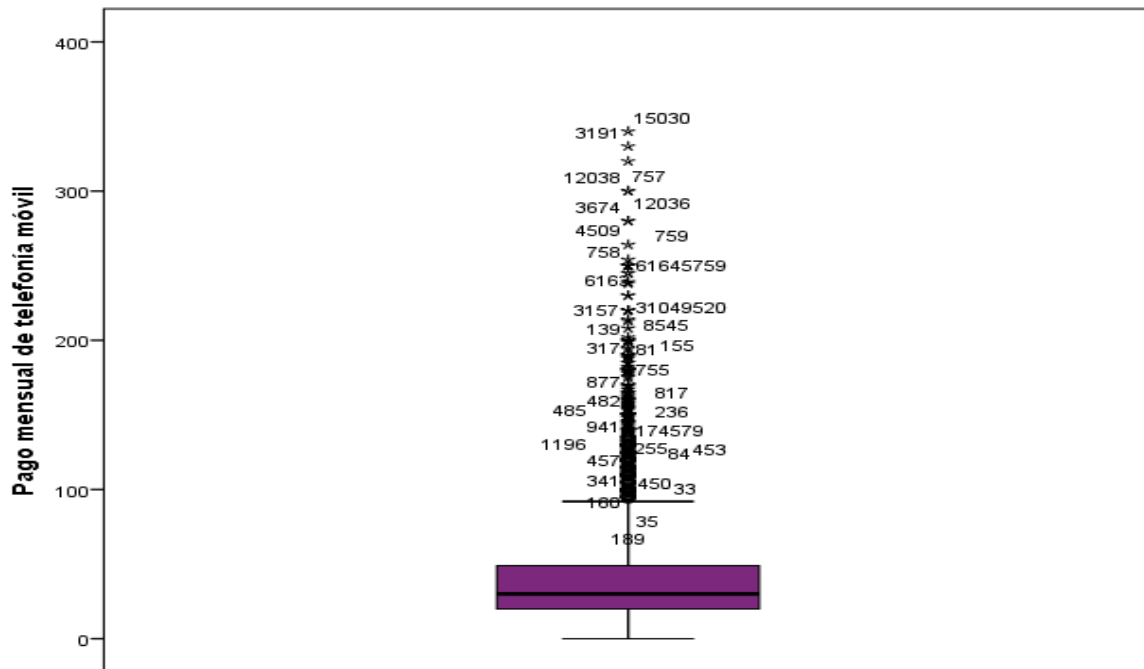
Diagrama de cajas y bigotes ingreso principal mensual



Nota: Datos tomados de la encuesta residencial de telecomunicaciones (2021).

Figura 6

Diagrama de cajas y bigotes pago mensual



Nota: Datos tomados de la encuesta residencial de telecomunicaciones (2021).

De las Figuras 4, 5 y 6 se observa una cantidad importante de datos atípicos, por lo que para poder identificar estos se calculó la distancia de Mahalanobis a través del software spss el cual genera una nueva columna (nueva variable) en la que realiza el cálculo de la distancia de Mahalanobis para cada encuesta. A continuación, se presenta los datos descriptivos de esta nueva variable:

Tabla 3

Distancia de Mahalanobis

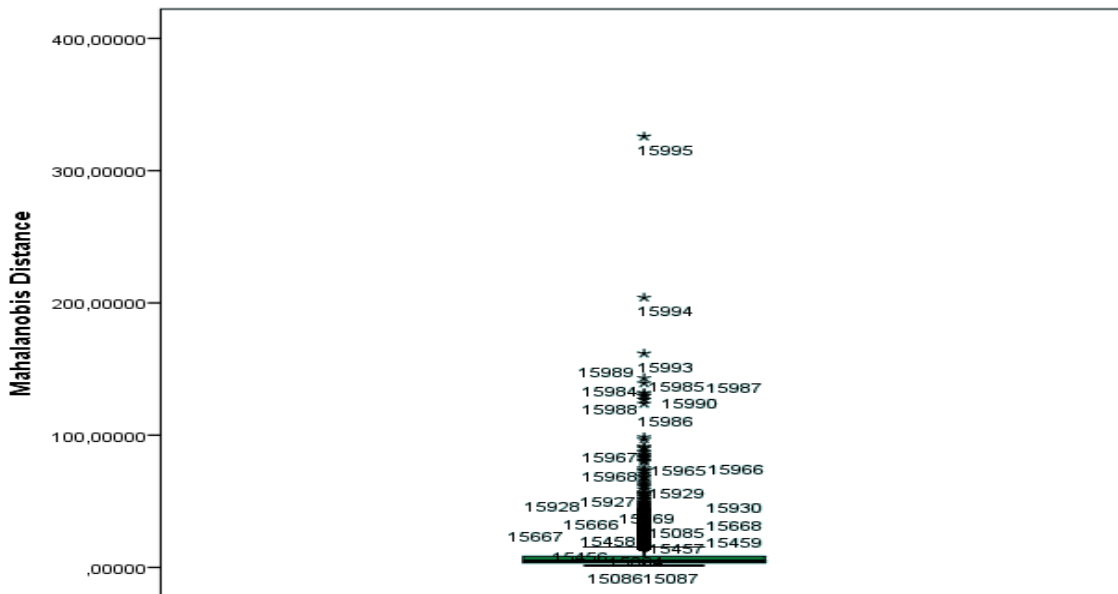
	Frecuencia	Mínimo	Máximo	Media	Desviación típica
Distancia de Mahalanobis	15995	1.41691	325.81633	6.9995624	7.53462152

Nota: Elaboración de la autora.

Según la Tabla 3, el promedio de la variable de Mahalanobis es 6.99, como mínimo 1.41 y como máximo 325.81.

Figura 7

Diagrama de caja distancias de Mahalanobis



Nota: Elaboración propia de la autora.

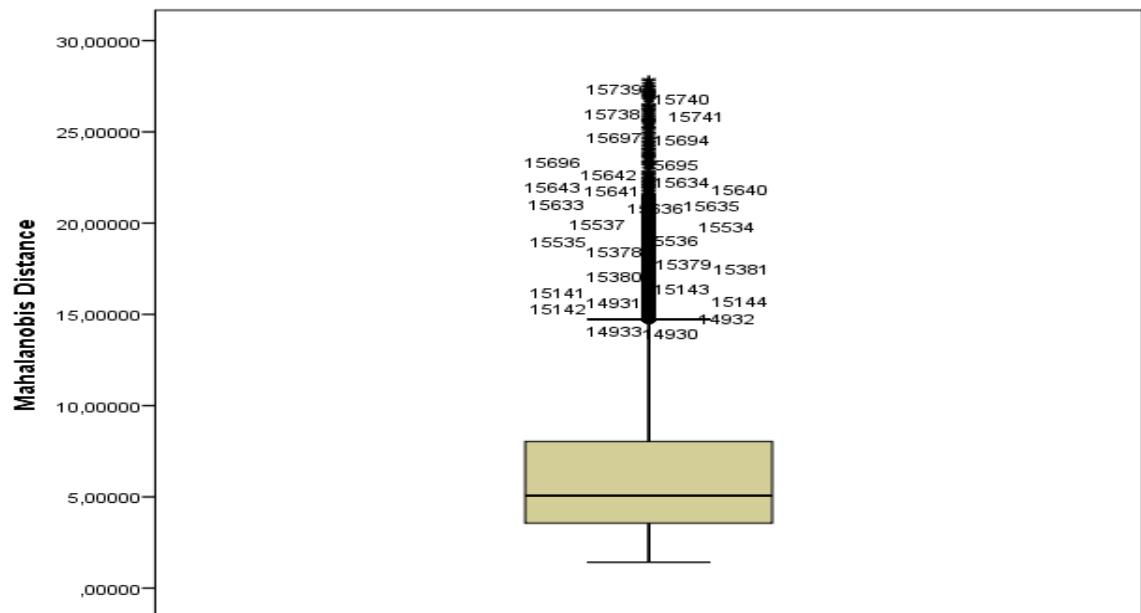
En la Figura 7 se presenta el diagrama de cajas y bigotes en la que se observa gran cantidad de datos atípicos por lo que se procede a calcular la probabilidad

superior de la distribución chi – cuadrado con 3 grados de libertad y los datos atípicos multivariados son los valores que tienen probabilidad menor a 0.001 por lo tanto $D^2 < 16.2660$ (Kline, 2016).

Se procede a eliminar todas las encuestas que tengan valores iguales o superiores a 16.2660. Posterior a la eliminación el diagrama de cajas para la variable distancia de Mahalanobis quedo así:

Figura 8

Diagrama de caja distancia de Mahalanobis valores < 16.2660



Nota: Elaboración propia de la autora.

De la Figura 8, se observa que aún existen datos atípicos, aunque ya se puede trabajar con los ellos pues se eliminaron los valores mayores a 16.2660, quedando 15 780 datos.

A continuación, se presenta el análisis de descriptivo de las variables cualitativas.

Figura 9

Diagrama de sectores sexo



Nota: Elaboración propia de la autora.

Tabla 4

Frecuencias según sexo

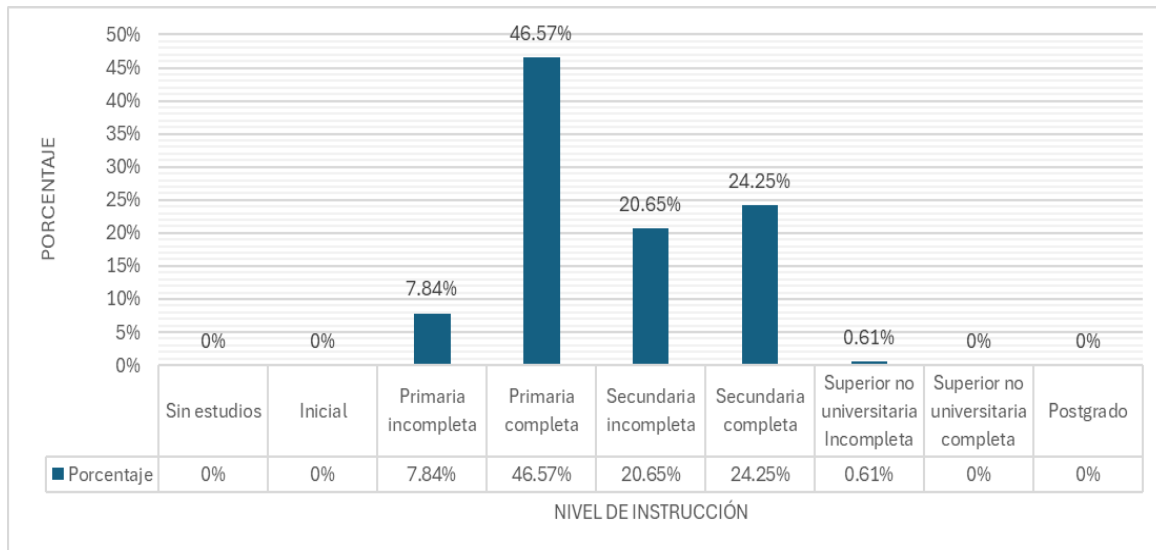
Sexo	Frecuencia	Porcentaje
Mujer	6469	40.99%
Varón	9311	59.01%
Total	15780	100%

Nota: Elaboración propia de la autora.

De la Figura 9 y Tabla 4, se observa que el 59.01% de los encuestados son varones y el 40.99% representan a las mujeres encuestadas, sin datos perdidos en alguna de las dos categorías.

Figura 10

Diagrama de barras nivel de instrucción

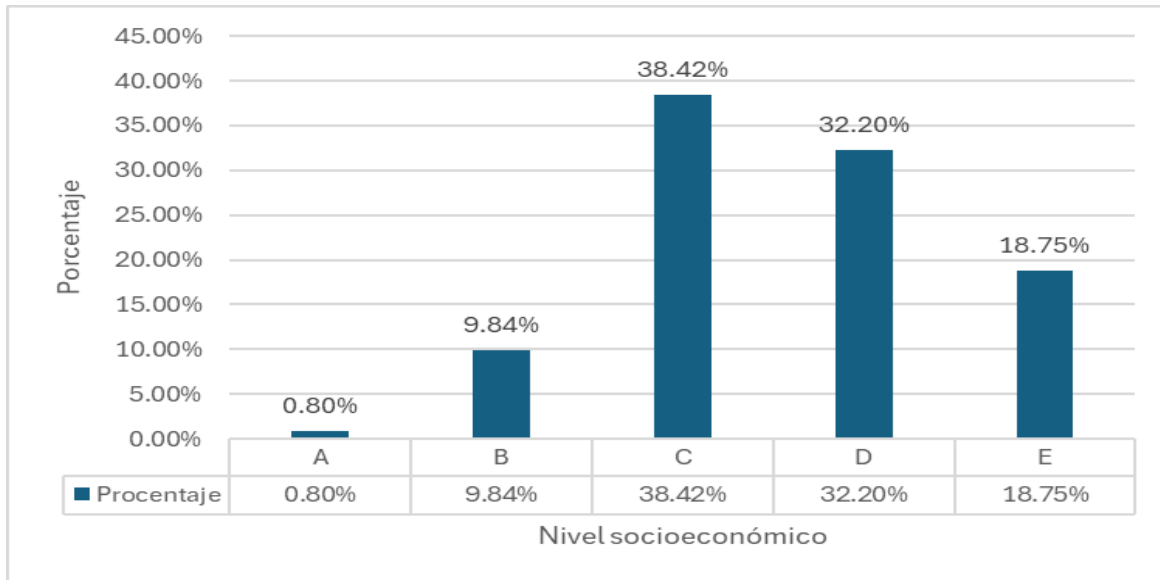


Nota: Elaboración propia de la autora.

En la Figura 10, se observa los porcentajes de los encuestados el 46.57% cuenta con una formación de primaria completa, el 20.65% cuenta con secundaria incompleta y el 24.25% completo la formación secundaria. El 7.84% cuenta con primaria incompleta y solo 0.61% cuenta con una formación universitaria incompleta.

Figura 11

Diagrama de barras nivel socioeconómico



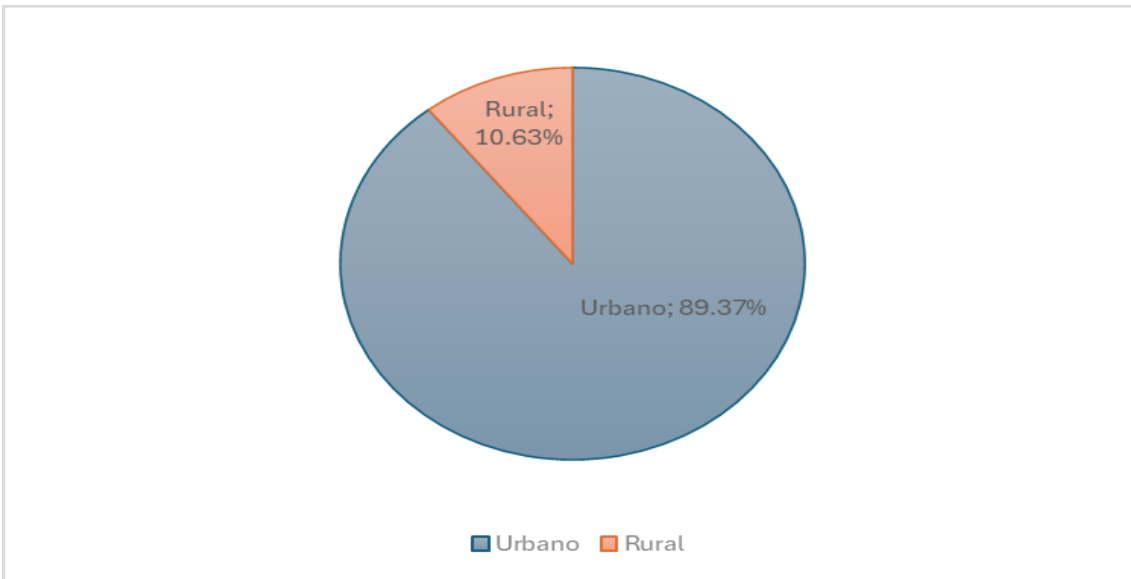
Nota: Elaboración propia de la autora.

De la Figura 11, la mayor cantidad de los encuestados pertenecen al nivel socioeconómico C con un porcentaje del 38.42% próximo al porcentaje de encuestados que pertenecen al estrato D, el 18.75% de los encuestados pertenecen al estrato E, el 9.84% pertenece al estrato B y solo el 0.8% pertenece al estrato A.

De acuerdo con Ipsos (2020), las características de los hogares con nivel socioeconómico del estrato D son familias conforman el 34% de la población y son el mayor número respecto a los demás estratos y solo el 2% de la población peruana conforma el estrato A.

Figura 12

Diagrama de sectores área

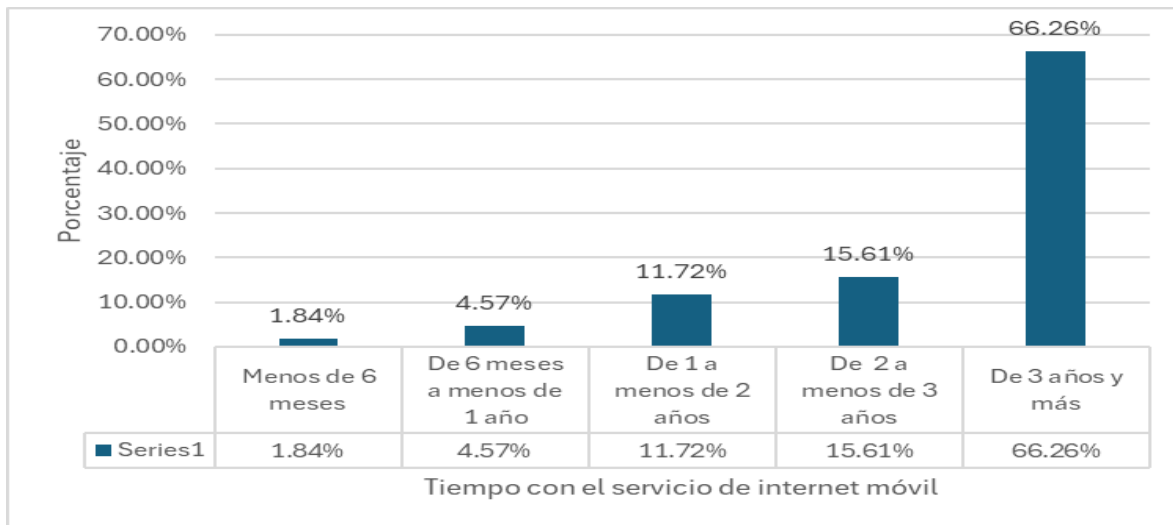


Nota: Elaboración propia de la autora.

La Figura 12, muestra que el 89.37% de los participantes en la encuesta pertenecen al área urbana y solo el 10.63% pertenecen al área rural.

Figura 13

Diagrama de barras tiempo con servicio móvil



Nota: Elaboración propia de la autora.

De la Figura 13, el 66.26% de los encuestados indicaron poseer el servicio de internet móvil entre 3 a más años siendo este el mayor porcentaje y menos del 4% de los encuestados indican poseer este servicio menos de 6 meses.

Luego de la limpieza de los datos se procedió a realizar el análisis de clúster bietápico a continuación, se presenta los resultados obtenidos:

4.1.2 Algoritmo para la obtención de los conglomerados

Se utiliza la técnica de clúster bietápico recibe el nombre debido a que se compone de dos partes:

Paso 1: Recibe el nombre de preagrupamiento dado que en esta etapa se forman los pre - conglomerados. El cual empieza con el desarrollo del CF tree, este asigna la primera observación en un nodo hoja que almacena información sobre la misma. Luego se aumenta cada caso a un nodo hoja existente o se forma uno nuevo, si la longitud de similitud (distancia log verosimilitud) es próxima aun no da hoja que ya exista, caso contrario se inicia un nuevo nodo hoja.

Cuando culmina el paso uno, se tendrá construido los pre - clústeres mediante el CF tree, estos guardan las estadísticas de resumen (ecuación 6).

Paso 2: Luego de constituido el CF tree en el anterior paso, los nodos hojas se juntan usando un algoritmo aglomerativo de clúster. Dicho algoritmo aglomerativo nos proporciona una cantidad de soluciones o modelos. Para seleccionar la opción óptima se puede discriminar las diferentes opciones con el criterio bayesiano de Schwartz / BIC) o el criterio de Información de Akaike (AIC).

El método utilizado es del BIC (ecuación 10), el cual arroja instantáneamente el

número óptimo de conglomerados.

Es importante mencionar que el presente método para el estudio de conglomerados es muy robusto respecto al quebrantamiento del supuesto de independencia cono de las distribuciones. A continuación, se presenta el cálculo óptimo de conglomerados:

Figura 14

Resumen del modelo de clasificación



Nota: Elaboración propia de la autora. La figura muestra el resumen del modelo, detallando que la técnica utilizada fue la técnica de segmentación bietápica, respecto a las entradas se refiere a la cantidad de dimensiones significativas para la clasificación y la cantidad de conglomerados, así también indica la calidad de medida de silueta de cohesión y separación.

La Figura 14, muestra la cantidad de conglomerados que son 4 con una medida de silueta de cohesión y separación calificada como **buena** tomando tres dimensiones de clasificación como significativas que son: Ingreso promedio mensual, pago mensual promedio por el servicio móvil y nivel socioeconómico.

Luego de varias pruebas se concluye que son tres las principales características socio económicas del usuario de telecomunicaciones que son significativas para la

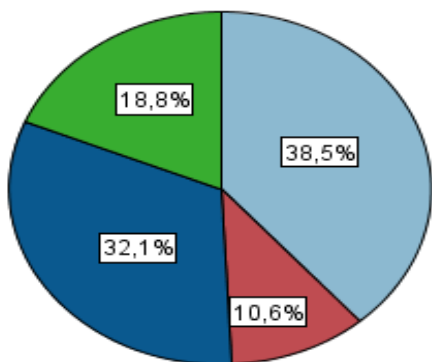
clasificación y que la calidad de los conglomerados sea buena, por lo que en la Figura 15 en el recuadro de entradas se denota un 3 que corresponde a las siguientes dimensiones: Ingreso principal mensual, pago mensual por el servicio móvil y nivel socio económico del hogar.

Respecto a las demás variables como sexo del encuestado, edad del encuestado, área, nivel de instrucción y el tiempo que tiene con el servicio de telefonía móvil no son significativas pues la medida de silueta y cohesión del modelo de clasificación, cuando se consideraron como principales factores de clasificación, indicaron un modelo de clasificación regular a malo, más si serán considerados como campos de evaluación en la descripción de cada perfil.

Figura 15

Tamaño de los conglomerados

Tamaños de conglomerados



Tamaño de conglomerado más pequeño	1678 (10,6%)
Tamaño de conglomerado más grande	6073 (38,5%)
Cociente de tamaños: Conglomerado más grande a conglomerado más pequeño	3,62

Nota: Elaboración propia de la autora. La Figura muestra el tamaño de los conglomerados conformados a través de la técnica de segmentación bietápica, así también se detalla el conglomerados más grande y más pequeño y el cociente de tamaños entre ambos.

Tabla 5

Frecuencias por conglomerado

Conglomerado	Frecuencia absoluta fi	Frecuencia porcentual
1	6073	38.5%
2	1678	10.69%
3	5070	32.1%
4	2959	18.8%
Total	15780	100%

Nota: Elaboración propia de la autora.

En la Tabla 5 se observa que el tamaño del conglomerado más grande está conformado por el 38.5% de los encuestados siendo 6 073 personas, luego el más pequeño conformado por el 10.69% de encuestados siendo en número 1 678 personas.

Luego de la descripción de la clasificación en general, se procede a describir cada grupo de clasificación de acuerdo con el porcentaje que representa del total.

4.1.3 Identificación de conglomerados

Se procede a describir cada conglomerado para así conocer las características de los usuarios en cada clúster con la finalidad de en lo posterior proponer los perfiles.

La importancia de este proceso, recae en que la segmentación de clientes incrementa la efectividad de las campañas de marketing al proporcionar a la empresa información que puede utilizar para alinear su producto con las necesidades y requerimientos de sus clientes lo que contribuye a la fidelización del mismo, como la personalización y atención individualizada de los segmentos para que el cliente se sienta valorado y comprendido, este sentimiento de conexión y atención fortalece la lealtad del cliente (Correa y otros, 2024).

Nivel socio económico

Tabla 6

Frecuencias por nivel socio económico por conglomerado

Conglomerado	A		B		C		D		E	
	fi	%	fi	%	fi	%	fi	%	fi	%
1	0	0.0%	0	0.0%	6062	100.0%	11	.2%	0	0.0%
2	126	100.0%	1552	100.0%	0	0.0%	0	0.0%	0	0.0%
3	0	0.0%	0	0.0%	0	0.0%	5070	99.8%	0	0.0%
4	0	0.0%	0	0.0%	0	0.0%	0	0.0%	2959	100.0%

Nota: Elaboración propia de la autora.

Pago mensual en el servicio móvil e ingreso principal mensual

Tabla 7

Media de pago mensual e ingreso principal por conglomerado

Conglomerado	Pago mensual por servicio móvil		Ingreso principal mensual	
	Media	Desviación típica	Media	Desviación típica
1	41.03	24.613	1805.46	941.145
2	52.24	26.129	2410.27	1147.910
3	31.78	19.224	1428.16	790.092
4	25.96	14.932	1156.95	708.023

Nota: Elaboración propia de la autora.

Clúster 1: Está conformado por individuos que pertenecen al sector C y D en menor medida este último (Tabla 6), que de acuerdo con el INEI realizan un gasto mensual de 75% a 80% de sus ingresos (Figura 1), el pago mensual por el servicio móvil se calcula en 41.03 soles y tienen un ingreso principal mensual promedio de s/ 1805.46 (Tabla 7).

Clúster 2: A este clúster pertenecen individuos que cuyo nivel socio económico está en la clasificación A y B (Tabla 6), aunque en su mayoría lo conforma individuos

del nivel socio económico de clasificación que de acuerdo con el INEI su gasto mensual oscila entre 62 a 68% de sus ingresos (Figura 1). El pago mensual que realizan por el servicio móvil es en promedio s/ 52.24 y el ingreso principal mensual es en promedio s/2 410.7 (Tabla 7).

Clúster 3: El tercer clúster está conformado por individuos que cuyo nivel socio económico está exclusivamente en la clasificación D (Tabla 6), que de acuerdo con el INEI su gasto mensual es del 80% de sus ingresos (Figura 1). El pago mensual que realizan por el servicio móvil es en promedio s/ 31.78 y el ingreso principal mensual es en promedio s/1 428.16 (Tabla 7).

Clúster 4: Por último, el cuarto clúster está conformado por individuos cuyo nivel socio económico está exclusivamente en la clasificación E (Tabla 6), que de acuerdo con el INEI su gasto mensual es del 87% de sus ingresos (Figura 1). El pago mensual que realizan por el servicio móvil es en promedio s/ 25.96 y el ingreso principal mensual es en promedio s/1 156.95 (Tabla 7).

4.1.4 Perfil de usuario de acuerdo con el conglomerado que pertenece

A continuación, se propone el perfil del consumidor para cada conglomerado, así como la denominación de cada conglomerado.

Perfil 1 (clúster 1)

Este individuo reserva un 25% a 30% de sus ingresos posiblemente esto sea para el ahorro o inversión y el pago que realiza por su servicio móvil es el equivalente del 2.27% de su ingreso promedio mensual. El porcentaje de gasto en servicio móvil es el mayor respecto a otros conglomerados.

Perfil 2 (clúster 2)

El individuo que pertenece a este clúster reserva entre el 32% al 38% de su ingreso principal y el gasto que realiza en su servicio móvil es de 2.16% de su ingreso principal.

Perfil 3 (clúster 3)

El individuo que pertenece a esta clasificación gasta entre el 80% de sus ingresos dejando solo el 20% para otras actividades como ahorro o inversión. El gasto en promedio que realiza en el servicio móvil es del 2.22% de su ingreso principal mensual.

Perfil 4 (clúster 4)

El individuo que pertenece a este clúster tiene el mayor porcentaje de gasto de respecto a sus ingresos principales comparado con los demás clústeres y el pago de su servicio corresponde al 2.24% de su ingreso principal mensual. Esto puede deberse a que no cuenta con los servicios de por ejemplo internet en el hogar o en el trabajo por lo que tiene que suplir este servicio con el servicio móvil.

En resumen, el individuo que pertenece al clúster 2 es el que mayor ingreso principal percibe mensualmente y es el que menor gasto realiza en el servicio móvil. El clúster 1 este compuesto por individuos que tienen el mayor gasto mensual en servicio de telefonía. El gasto en servicio móvil en general se encuentra entre el 2.16% al 2.27% del ingreso principal mensual.

DISCUSIÓN

Respecto al objetivo general se halló que los usuarios de telecomunicaciones en el Perú para el año 2021 se pueden segmentar en 4 clústeres, los cuales tiene una medida de silueta de cohesión y separación calificada como buena (Figura 15) tomando tres dimensiones de clasificación como significativas que son: ingreso promedio mensual, pago mensual promedio por el servicio móvil y nivel socioeconómico, en la investigación presentada por Billadoni (2021) clasificaron a los clientes de marketplaces en solo dos clústeres con la diferencia respecto al presente estudio que no especificaron la calidad de la medida de silueta de cohesión y cohesión.

Respecto al objetivo específico uno, se halló que las características por cada clúster es decir la clasificación de usuarios de telecomunicaciones en el Perú para el año 2021 son significativas y presentan características por cada conglomerados lo que proporciona información relevante respecto al comportamiento de los usuarios de telecomunicaciones siendo la técnica de clúster bietápico una técnica de big data altamente fiable que se puede aplicar en diferentes áreas del conocimiento, Billadoni (2021) en su investigación también resalta la gran utilidad de esta técnica para poder clasificar a los clientes en este caso lo hizo con clientes que compran en marketplace, la clasificación realizada por el autor también detallo las características generales de cada conglomerado lo que conlleva al igual que en la presente investigación a proponer perfiles por cada clúster.

Respecto al objetivo específico dos, se propone cuatro perfiles en los cuales se describe las características por cada clúster, esto también lo realizo Tang y Vargas

(2016) segmentando a los clientes de electrodomésticos en conglomerados utilizando también clúster bietápico obteniendo 3 conglomerados los cuales fueron descritos y formaron perfiles por cada grupo, es más esta investigación sirvió de inspiración para poder profundizar en la presente investigación pues la idea inicial era solo proponer la clasificación y posterior a la revisión de la investigación de Tang y Vargas (2016) surgió la idea para poder presentar el perfil de cada integrante por clúster.

CONCLUSIONES

Primero, se concluye que los usuarios de telecomunicaciones en el Perú para el año 2021 se pueden segmentar en 4 clústeres, los cuales tiene una medida de silueta de cohesión y separación calificada como buena (Figura 15) tomando tres dimensiones de clasificación como significativas que son: ingreso promedio mensual, pago mensual promedio por el servicio móvil y nivel socioeconómico.

Segundo, las características por cada clúster de usuarios de telecomunicaciones en el Perú para el año 2021 son significativas pues presentan características diferentes entre clúster y similitud de los individuos que pertenecen al mismo clúster. Es así, que el clúster 1 está conformado por individuos que pertenecen al sector C y D en menor medida este último, que de acuerdo con el INEI realizan un gasto mensual de 75% a 80% de sus ingresos, el pago mensual por el servicio móvil se calcula en 41.03 soles y tienen un ingreso principal mensual promedio de s/ 1 805.46. El clúster 2 pertenecen individuos que cuyo nivel socio económico está en la clasificación A y B, aunque en su mayoría lo conforma individuos del nivel socio económico de clasificación que de acuerdo con el INEI su gasto mensual oscila entre 62 a 68% de sus ingresos, el pago mensual que realizan por el servicio móvil es en promedio s/ 52.24 y el ingreso principal mensual es en promedio s/2 410.7. El clúster 3 está conformado por individuos que cuyo nivel socio económico está exclusivamente en la clasificación D, que de acuerdo con el INEI su gasto

mensual es del 80% de sus ingresos. El pago mensual que realizan por el servicio móvil es en promedio s/ 31.78 y el ingreso principal mensual en promedio s/1 428.16. Por último, el clúster 4 está conformado por individuos cuyo nivel socio económico es exclusivamente la clasificación E, que de acuerdo con el INEI (Figura 1) su gasto mensual es del 87% de sus ingresos. El pago mensual que realizan por el servicio móvil en promedio s/ 25.96 y el ingreso principal mensual es en promedio s/1 156.95.

Tercero, se propone 4 perfiles de usuarios del servicio de telecomunicaciones en el Perú para el año 2021 siendo estos: Perfil 1 (clúster 1), este individuo reserva un 25 a 30% de sus ingresos posiblemente esto sea para el ahorro o inversión y el pago que realiza por su servicio móvil es el equivalente del 2.27% de su ingreso promedio mensual. El porcentaje de gasto en servicio móvil es el mayor respecto a otros conglomerados. Perfil 2 (clúster 2), el individuo que pertenece a este clúster reserva entre el 32 al 38% de su ingreso principal y el gasto que realiza en su servicio móvil es de 2.16% de su ingreso principal. Perfil 3 (clúster 3), el individuo que pertenece a esta clasificación gasta el 80% de sus ingresos dejando solo el 20% para otras actividades como ahorro o inversión y el gasto en promedio que realiza en el servicio móvil es del 2.22% de su ingreso principal mensual. Perfil 4 (clúster 4), el individuo que pertenece a este clúster tiene el mayor porcentaje de gasto de respecto a sus ingresos principales comparado con los demás clústeres y el pago de su servicio corresponde al 2.24% de su ingreso principal mensual. Esto puede deberse a

que no cuenta con los servicios de por ejemplo internet en el hogar o en el trabajo por lo que tiene que suplir este servicio con el servicio móvil.

RECOMENDACIONES

A OSIPTEL, realizar investigaciones respecto al comportamiento del consumidor con la información recabada en la ERESTEL debido a que se posee bastante información que necesita ser analizada mediante la estadística inferencial y no limitarse a solo a nivel descriptivo.

A la comunidad científica, hay que recomendar que se sigan realizando investigación con la información proporcionada por OSIPTEL en especial la encuesta ERESTEL pues la encuesta es relativamente joven a comparación de otras realizadas por el estado pero que brindan información valiosa para ser analizada.

A los usuarios de telecomunicaciones, seguir participando activamente en las encuestas realizadas por el OSIPTEL que, si bien son extensas como la ERESTEL estas son una fuente vital de información para realizar investigaciones como la que se presentada, que en lo posterior pueden ser utilizadas para proponer mejoras en el servicio brindado por las empresas de telecomunicaciones y de igual manera regular sus servicios.

A la comunidad científica, se recomienda continuar con la investigación utilizando como fuente de información la encuesta ERESTEL utilizando técnicas estadísticas de big data, proponiendo una ecuación de regresión en la que se pueda indicar a que clúster pertenece de acuerdo con sus características.

REFERENCIAS

- Best, R. (2007). *Marketing Estratégico* (4ta edición ed.). Pearson Educación.
https://www.academia.edu/36987104/Marketing_Estrat%C3%A9gico_Roger_Best
- Billadoni, M. (2021). *Clustering de clientes de un grupo de e-Marketplaces del Perú*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos].
https://cybertesis.unmsm.edu.pe/bitstream/handle/20.500.12672/17530/Billadoni_vm.pdf?sequence=3
- Brito, A., y Vergara, M. (2022). *Análisis del perfil del cliente para la demanda de productos de pastelería y repostería saludable en la empresa Casa de Tortas Laury's*. [Tesis de maestría, Universidad de Santander].
[https://repositorio.udes.edu.co/server/api/core/bitstreams/2b21e528-d74f-43aa-b8d5-5cf86e6db954/content#:~:text=\(Bonacifa%2C%202021\).,ingreso%2C%20sexo%20y%20estado%20civil.](https://repositorio.udes.edu.co/server/api/core/bitstreams/2b21e528-d74f-43aa-b8d5-5cf86e6db954/content#:~:text=(Bonacifa%2C%202021).,ingreso%2C%20sexo%20y%20estado%20civil.)
- Córdova, M. (2003). *Estadística: Descriptiva e inferencial aplicaciones*. Editorial Moshera EIRL.
- Correa, M., Almeida, B., Espinoza, M., Cebzas, A., y Castillo, K. (2024). *Segmentación Efectiva de Clientes utilizando R: Técnicas para Administración y Marketing Avanzado: Effective Customer Segmentation using R: Techniques for Advanced Marketing and Management*. Editorial Investigativa Latinoamericana (SciELA).
- Espigares, M., Bautista, J., y Duarte, M. (2017). *Uso del método bietápico en el estudio*

- de los procesos de enseñanza y aprendizaje musical a través de Moodle.
Revista de investigación en Ciencias Sociales, Vol. 13(nº 2), pág. 187-200.
<https://doi.org/10.18004/riics.2017.diciembre.187-200>
- Feria, H., Matilla, M., y Mantecon, S. (2020). La entrevista y la encuesta: ¿Métodos o técnicas de indagación empírica? *Revista Didasc@lia*, 63-79.
- Gonzáles, M. (2021). *Estudio de percepción del consumidor de plataformas de streaming de video bajo demanda por suscripción*. [Tesis de pregrado, Universidad Técnica Federico Santa María Departamento de industrias].
<https://repositorio.usm.cl/bitstream/handle/11673/52665/3560903501540UTFSM.pdf?sequence=1>
- Hernández, R., Fernández, C., y Baptista, P. (2018). *Metodología de la investigación* (6° edición ed.). McGraw- Hill Interamericana Editores S.A.
- Ipsos Group S.A. (20 de Febrero de 2020). *Características de los niveles socioeconómicos en el Perú*. <https://www.ipsos.com/es-pe/caracteristicas-de-los-niveles-socioeconomicos-en-el-peru>
- Johnson, R., y Kuby, P. (2012). *Estadística elemental*. Cengage Learning Editores S.A.
- Kline, R. (2016). *Principles and practice of structural equation modeling*. The Guilford Press.
- Levin, R., Rubin, D., Balderas , M., Del Valle, J., y Gómez, S. (2004). *Estadística para administración y economía*. Editorial Pearson Educación.
- Llinás, H., y Rojas, C. (2006). *Estadística descriptiva y distribuciones de probabilidad*. Ediciones Uninorte.

- Mamani, I., y Damiano, D. (2014). *Satisfacción laboral en el sector construcción mediante análisis factorial y análisis de clúster*. [Tesis de pregrado , Universidad Nacional San Antonio Abad del Cusco]. <http://hdl.handle.net/20.500.12918/1450>
- Manrique, C. (2016). *Segmentación de clientes de Corporación Lindley de la región Lima mediante el análisis de Cluster bietápico en octubre 2016*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos].
<https://cybertesis.unmsm.edu.pe/backend/api/core/bitstreams/34c1c356-da7c-4ea3-bada-64e7e430e5aa/content>
- Martinez, J. (06 de julio de 2024). *Análisis de Cluser*.
<https://www.uv.es/ceaces/multivari/cluster/CLUSTER2.htm#:~:text=El%20an%C3%A1lisis%20cluster%20es%20un,la%20clasificaci%C3%B3n%20de%20los%20individuos>.
- McDaniel, C., y Gates, R. (2005). *Investigación de mercados*. Editorial Thomson.
- Muñoz, J., y Amón, U. (2013). Técnicas para detección de outliers multivariantes. *Revista en telecomunicaciones e informática*.
- Oporto, S., Torres, D., y Castillo, M. (2018). *Segmentación de mercado, mercado meta y posicionamiento en el mercado*. [Tesis de pregrado, universidad nacional autónoma de Nicaragua Managua].
<https://repositorio.unan.edu.ni/8579/1/18820.pdf>
- Organización Supervisor de Inversión Privada en Telecomunicaciones [Osiptel]. (2021). *Repositorio institucional - Encuesta Residencial de Servicios de Telecomunicaciones (ERESTEL)*.

<https://repositorio.osiptel.gob.pe/handle/20.500.12630/327>

Ortega, R. (2020). *Clasificación, eficiencia técnica y desempeño económico de los productores de quinua del departamento de Puno*. [Tesis de pregrado, Universidad Nacional Agraria La Molina].

<https://hdl.handle.net/20.500.12996/4492>

Palacios, F., y Pastor, N. (2020). *Segmentación de clientes de una empresa comercializadora de productos de consumo masivo en la ciudad de Popayán soportado en machine learning y análisis RFM (Recency, Frecuency y Money)*. [Tesis de grado, Universidad de Popayán].

<https://fupvirtual.edu.co/repositorio/s/repositorio/item/11653>

Pérez, A. (2017). *Cluster ferretero en la ciudad de Urubamba-Cusco 2017*. [Tesis de grado, Universidad Andina del Cusco].

[https://repositorio.uandina.edu.pe/bitstream/handle/20.500.12557/1652/RESUME N.pdf?sequence=1&isAllowed=y](https://repositorio.uandina.edu.pe/bitstream/handle/20.500.12557/1652/RESUME%20N.pdf?sequence=1&isAllowed=y)

Pérez, C. (2011). *Técnicas de segmentación: Conceptos, herramientas y aplicaciones*. Madrid: Garceta.

Presidencia del Consejo de Ministros. (20 de Julio de 2024). *Plataforma Nacional de Datos Abiertos (PNDA)*.

<https://www.gob.pe/institucion/pcm/campa%C3%B1as/5904-portal-nacional-de-datos-abiertos-pnda>

Quaglino, M., y Merello, J. (2012). *Métodos multivariados en estudios de vulnerabilidad social en la provincia de Santa Fe*. <https://archivo.fcecon.unr.edu.ar/web->

nueva/sites/default/files/u16/Decimocuartas/quaglino_merello_metodo_multivariados_en_estudios.pdf

Rodriguez, A., y Romero, M. (2021). *Identificación de los perfiles del cliente omnicanal de la categoría moda en las tiendas por departamento ubicadas en Lima Metropolitana, en base a los canales utilizados durante el proceso de compra*. [Tesis de pregrado, Pontificia Universidad Católica de Lima].

<https://tesis.pucp.edu.pe/repositorio/handle/20.500.12404/18162>

Rubio, M., y Vilá, R. (2016). El análisis de conglomerados bietápico o en dos fases con SPSS. *REIRE*, 10(1). <https://doi.org/http://doi.org/10.1344/reire2017.10.11017>

Salinas, J. (2010). Segmentación de mercados usando análisis de clúster bietápico. *Anales científicos*.

Salvador, M., y Gargallo, P. (2003). Análisis exploratorio de datos (AED). *5 campus, Estadística*. https://www.yumpu.com/es/document/read/14196219/analisis-exploratorio-de-datos-aed-5campus#google_vignette

Sanchez, L. (14 de noviembre de 2023). *Explorando los Niveles Socioeconómicos en Perú: La nueva fórmula y puntos claves para su interpretación*.

<https://www.ipsos.com/es-pe/explorando-los-niveles-socioeconomicos-en-peru-la-nueva-formula-y-puntos-claves-para-su>

Tang, F., y Vargas, C. (2016). *Segmentación de clientes de una tienda de electrodomésticos utilizando el análisis de conglomerados*. [Tesis de pregrado, Universidad Nacional Agraria La Molina].

<https://hdl.handle.net/20.500.12996/2214>

- Tonconi, C. (2021). *Identificación de perfiles de los centros de educación técnico-productiva públicos usando indicadores de condiciones básicas de calidad mediante clúster bietápico*. [Tesis de pregrado, Universidad Nacional Agraria La Molina Facultad Economía y planificación].
- Valderrey, P. (2011). *Segmentación de mercados* (1era ed. ed.). Ediciones de la U.
- Vera, H., y Soncco, J. (2016). *Algoritmo de optimización basado en el comportamiento social de arañas para Clustering*. [Tesis de pregrado, Universidad Nacional San Antonio Abad del Cusco].
<http://repositorio.unsaac.edu.pe/bitstream/handle/20.500.12918/2447/253T20160077.pdf?sequence=1&isAllowed=y>
- Vera, O., y Vera, F. (2013). Evaluación del nivel socioeconómico: presentación de una escala adaptada en una población. *Revista del cuerpo médico Hospital Nacional Almanzor Aguinada Asenjo*, 6(1), 41- 45.
<https://dialnet.unirioja.es/metricas/documentos/ARTREV/4262712>
- Zhang, T., Ramakrishnan, R., y Livny, M. (1996). *An efficient data clustering method for very large databases*. ACM SIGMOD.

ANEXOS

Anexo I: Matriz de consistencia

TÍTULO: Segmentación de usuarios de telecomunicaciones en clústeres utilizando la encuesta residencial de servicios de telecomunicaciones en el Perú, año 2021					
PROBLEMA GENERAL	OBJETIVO GENERAL	HIPÓTESIS GENERAL	VARIABLES	INDICADORES	METODOLOGÍA
¿Cómo es la segmentación de usuarios de telecomunicaciones en clústeres utilizando la encuesta residencial de servicios de telecomunicaciones en el Perú para el año 2021?	Segmentar a los usuarios de telecomunicaciones en clústeres utilizando la encuesta residencial de servicios de telecomunicaciones en el Perú para el año 2021	La segmentación de los usuarios de telecomunicaciones en clúster utilizando la encuesta residencial de servicios de telecomunicaciones en el Perú para el año 2021 es significativa	Características socio económicas del usuario de telecomunicaciones	Sexo	Alcance: Descriptivo y explicativo Enfoque: Cuantitativo Diseño: No experimental Población: 47 346 participantes de la encuesta residencial de telecomunicaciones ERESTEL 2021. Muestra: 15 780 participantes de la encuesta residencial de telecomunicaciones ERESTEL 2021.
				Edad	
				Nivel de instrucción	
				Ingresos principales	
PROBLEMAS ESPECÍFICOS	OBJETIVOS ESPECÍFICOS	HIPÓTESIS ESPECÍFICAS		Nivel socio económico	
				Área	
				Pago mensual en el servicio móvil	
¿Cómo son los clústeres en los que se segmenta a los usuarios de telecomunicaciones utilizando la encuesta residencial de servicios de telecomunicaciones en el Perú, año 2021?	Describir las características de los usuarios de telecomunicaciones por cada clúster utilizando la encuesta residencial de servicios de telecomunicaciones en el Perú, año 2021.	Las características de los usuarios de telecomunicaciones por cada clúster utilizando encuesta residencial de servicios de telecomunicaciones en el Perú, año 2021 son significativas.		Tiempo con el servicio de internet	
¿Existen perfiles de usuarios de telecomunicaciones por cada clúster utilizando Encuesta residencial de servicios de telecomunicaciones en el Perú para el año 2021?	Proponer perfiles del usuario de telecomunicaciones para cada clúster utilizando Encuesta residencial de servicios de telecomunicaciones en el Perú para el año 2021	Existen perfiles de los usuarios de telecomunicaciones por cada clúster utilizando Encuesta residencial de servicios de telecomunicaciones en el Perú para el año 2021	Perfiles de cliente	Perfil 1	
				Perfil 2	
				Perfil 3	
				Perfil 4	

Anexo II: Encuesta ERESTEL 2021

ENCUESTA RESIDENCIAL DE SERVICIOS DE TELECOMUNICACIONES (ERESTEL) 2021



N° de Cuestionario

ENCUESTADOR (PRESENTACIÓN): Buenos días/ tardes/ noches mi nombre es soy encuestador de PRISMA-ANALITIK en este momento estamos realizando un estudio para el OSIPTEL (el regulador de las telecomunicaciones). Por este motivo, estamos recogiendo información sobre la demanda y uso de los servicios de telecomunicaciones en los hogares del país, y su hogar ha sido elegido al azar para realizar una entrevista, por lo que le pido por favor me brinde su tiempo ya que su participación es muy importante para el desarrollo de las telecomunicaciones en el Perú. Su información permanecerá en absoluta reserva y sólo será utilizada en forma agregada junto al resto de los entrevistados y en ningún caso en forma individual, además de que esta protegida por ley. ¡Muchas gracias!.

A. Ubicación Geográfica				Código	B. Ubicación Muestral				Área			
Departamento					Conglomerado N°				1 Urbana			
Provincia					Zona N° / AER N°				2 Rural			
Distrito					Manzana N°							
Centro Poblado	Nombre				Vivienda N°							
	Categoría				Hogar N°							
C. Dirección de la Vivienda												
Nombre de la Avenida, Calle, Jirón, Pasaje, etc.			Puerta N°	Int.	Piso	Etapa/Sect./Grup.	Mz.	Lote	Km.	Nombres y Apellidos del Informante	Cód.	Teléfono/Celular
Referencia ↳ [De la Sección 1]												
D. Personal de la Encuesta												
Nombre del Encuestador (a)		Fecha 1era. Visita		Fecha 2da. Visita		Fecha 3ra. Visita						
Código del Digitador (a)		Máquina <input type="checkbox"/>		Fecha		Fecha		Tipo de Supervisión		1ra	2da	
								Re entrevista		1	1	
Código del Supervisor (a)		Fecha de Supervisión						Observación		2	2	
								Revisión		3	3	
Observaciones:												

SECCIÓN 14: CARACTERÍSTICAS GENERALES DE LA VIVIENDA Y DEL HOGAR

13 Su hogar tiene:

13.1 Televisor
 1 Sí → ¿Cuántos?
 2 No

(Verificar con P1 de Sección 9, PAGINA 40)

13.1.1 De estos Televisores ¿Cuántos pueden:

- 1 Captar la Señal de TV Digital Terrestre
Sin Usar un Decodificador/
 sintonizador Externo?
- 2 Captar la Señal de TV Digital Terrestre
Usando un Decodificador/
 sintonizador Externo?
- 3 No sabe

	Sí	No	¿Cuántos?
13.2 Radio o Equipo de Sonido	1	2	<input type="text"/>
13.3 Computadora de Escritorio (PC)	1	2	<input type="text"/>
13.4 Laptop, Notebook, Netbook y/o Ultrabook	1	2	<input type="text"/>
13.5 Reproductor de Video DVD o Blue-Ray	1	2	<input type="text"/>
13.6 Tablet	1	2	<input type="text"/>
13.7 Cocina a Gas o Eléctrica	1	2	<input type="text"/>
13.8 Auto, Camioneta	1	2	<input type="text"/>
13.9 Lavadora	1	2	<input type="text"/>
13.10 Microondas	1	2	<input type="text"/>
13.11 Reproductor de MP3 o MP4	1	2	<input type="text"/>

Verificar con:
 Pregunta 2 de Sección 3, página 18
 Pregunta 1 de Sección 7, página 31
 Pregunta 6 de Sección 7, página 31
 Pregunta 10 de Sección 14, página 58

14 ¿Cuál fue el gasto promedio del hogar en ...

14a Los últimos 15 días

- 1 Alimentos y bebidas S/.
- 2 Artículos de limpieza y tocador S/.
- 3 Transporte público S/.
- 4 Combustibles y lubricantes S/.
 (Incluye GLP, GNV para cocinar)

14b En el último mes

- 1 Alquileres S/.
- 2 Agua S/.
- 3 Luz S/.
- 4 Periódicos, revistas S/.
- 5 Telecomunicaciones en el hogar:
 Teléfono (Fijo, Móvil), TV de Paga,
 Internet (Fijo, Móvil) S/.
- 6 Telefonía de Uso Público S/.

14c En los últimos 3 meses (acumulado)

- 1 Vestido y calzado S/.
- 2 Conserv. y reparac. de la vivienda S/.
- 3 Vajillas, cortinas y artículos de cocina S/.
- 4 Esparcimiento y servicio domestico S/.
- 5 Salud S/.

14d En el último año

- 1 Educación S/.
- 2 Pensiones y remesas S/.
- 3 Muebles y artefactos del hogar S/.
- 4 Otro ¿Cuál? S/.

(→ Fin de la Encuesta)

Anexo III Secuencia de las salidas del software SPSS

*Resultado_030723.spv [Documento1] - IBM SPSS Statistics Visor

Archivo Editar Ver Datos Transformar Insertar Formato Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda

FILE='E:\Katiuska\tesis\base datos 2023\datos_030723.sav'.
 DATASET NAME Conjunto_de_datos1 WINDOW=FRONTI.
 FRECUENCIAS VARIABLES=q4
 /ORDER=ANALYSIS.

Resumen del modelo

Algoritmo	Bietápico
Entradas	3
Clústeres	4

Calidad de clúster

Frecuencias

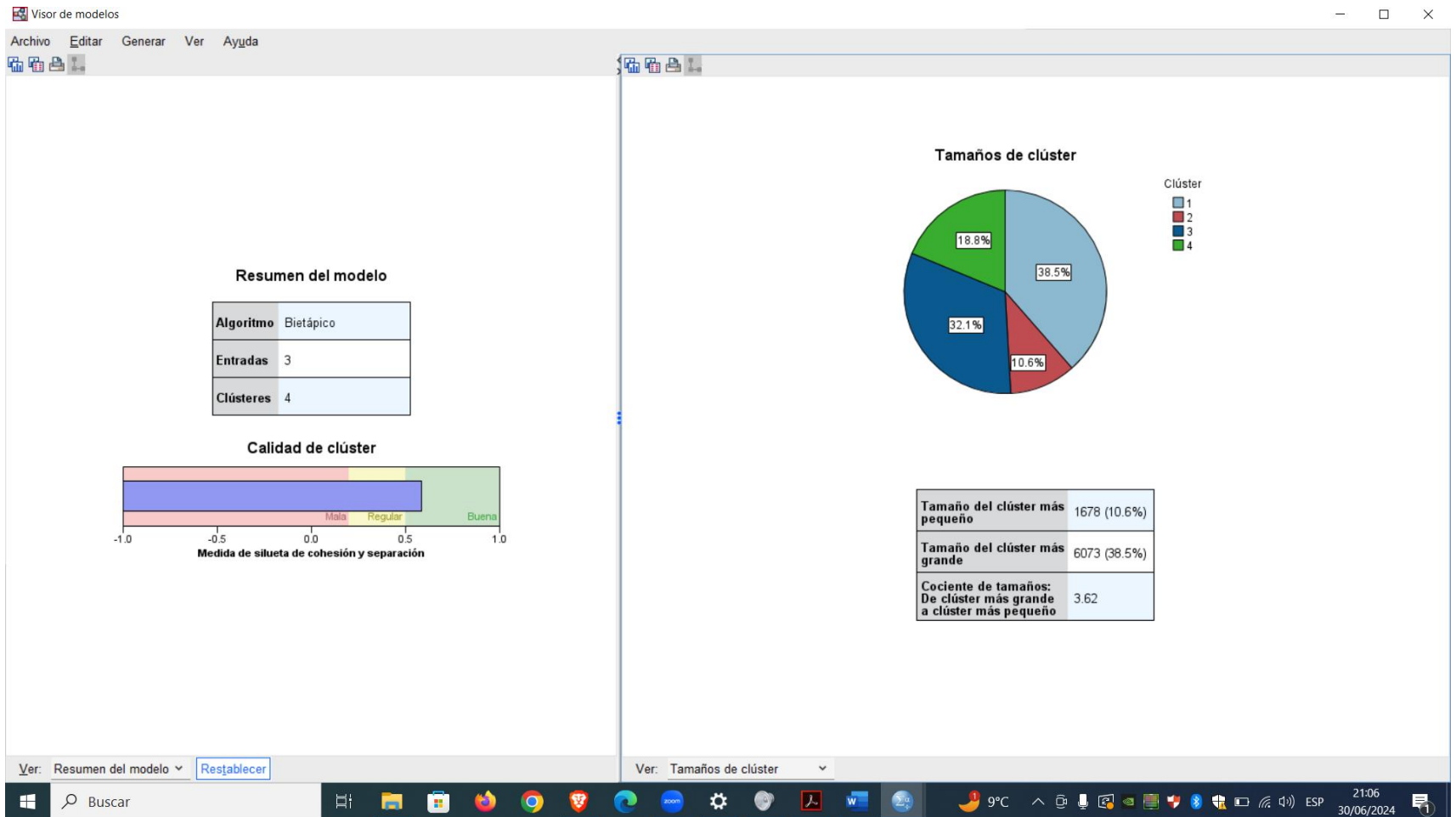
→ [Conjunto_de_datos1] E:\Katiuska\tesis\base datos 2023\datos_030723.sav

Estadísticos

Sexo		
N	Válidos	15780
	Perdidos	0

IBM SPSS Statistics Processor está listo Unicode:ACTIVADO

9°C 21:04 30/06/2024



Visor de modelos

Archivo Editar Generar Ver Ayuda

Resumen del modelo

Algoritmo	Bietápico
Entradas	3
Clústeres	4

Calidad de clúster

Medida de silueta de cohesión y separación

Importancia del predictor

Predictor	Importancia (aprox.)
Área	1.0
Nivel de instrucción	1.0
Ingreso principal mensual	1.0
Pago mensual por servicio móvil	1.0
Nivel socioeconómico del hogar	1.0
S3.1 a. ¿Desde hace cuánto tiempo tiene el servicio de Internet móvil?	0.35
Edad	0.3
Sexo	0.05

Menos importante | Más importante

Ver: Resumen del modelo [Restablecer](#)

Ver: Importancia del predictor

Windows taskbar: Buscar, 9°C, 21:06, 30/06/2024

*Resultado_030723.spv [Documento1] - IBM SPSS Statistics Visor

Archivo Editar Ver Datos Transformar Insertar Formato Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda

Frecuencias

Nivel socioeconómico del hogar

		A		B		C		D		E	
		Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Conglomerado	1	0	0.0%	0	0.0%	6062	100.0%	11	0.2%	0	0.0%
	2	126	100.0%	1552	100.0%	0	0.0%	0	0.0%	0	0.0%
	3	0	0.0%	0	0.0%	0	0.0%	5070	99.8%	0	0.0%
	4	0	0.0%	0	0.0%	0	0.0%	0	0.0%	2959	100.0%
	Combinados	126	100.0%	1552	100.0%	6062	100.0%	5081	100.0%	2959	100.0%

* Tablas personalizadas.
 CTABLES
 /VLABELS VARIABLES=q4 TSC_4012 DISPLAY=LABEL
 /TABLE q4 [COUNT F40.0] BY TSC_4012
 /CATEGORIES VARIABLES=q4 TSC_4012 ORDER=A KEY=VALUE EMPTY=INCLUDE.

Clústeres

Importancia de entrada (predictor)
 1.0 0.8 0.6 0.4 0.2 0.0

Clúster	1	3	4	2
Etiqueta				
Descripción				
Tamaño				

IBM SPSS Statistics Processor está listo Unicode:ACTIVADO H: 391, W: 488 pt. 21:06 30/06/2024

*Resultado_030723.spv [Documento 1] - IBM SPSS Statistics Visor

Archivo Editar Ver Datos Transformar Insertar Formato Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda

IBM SPSS Statistics Processor está listo Unicode:ACTIVADO

	(0073)	(0070)	(2909)	(1076)
Entradas	Ingreso principal mensual 1,805.46	Ingreso principal mensual 1,428.16	Ingreso principal mensual 1,156.95	Ingreso principal mensual 2,410.27
	Nivel socioeconómico del hogar C (99.8%)	Nivel socioeconómico del hogar D (100.0%)	Nivel socioeconómico del hogar E (100.0%)	Nivel socioeconómico del hogar B (92.5%)
	Pago mensual por servicio móvil 41.03	Pago mensual por servicio móvil 31.78	Pago mensual por servicio móvil 25.96	Pago mensual por servicio móvil 52.24
Campos de evaluación	Sexo 0 (56.6%)	Sexo 0 (60.6%)	Sexo 0 (63.7%)	Sexo 0 (54.6%)
	Edad 41.32	Edad 39.16	Edad 37.04	Edad 43.85
	Nivel de instrucción 2 (34.3%)	Nivel de instrucción 1 (62.6%)	Nivel de instrucción 1 (66.9%)	Nivel de instrucción 2 (70.6%)

IBM SPSS Statistics Processor está listo Unicode:ACTIVADO

21:07 30/06/2024

*Resultado_030723.spv [Documento1] - IBM SPSS Statistics Visor

Archivo Editar Ver Datos Transformar Insertar Formato Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda

Edad	Edad	Edad	Edad
41.32	39.16	37.04	43.85
Nivel de instrucción 2 (34.3%)	Nivel de instrucción 1 (62.6%)	Nivel de instrucción 1 (66.9%)	Nivel de instrucción 2 (70.6%)
Área 0 (96.8%)	Área 0 (89.7%)	Área 0 (67.8%)	Área 0 (99.5%)
S3.1a. ¿Desde hace cuánto tiempo tiene el servicio de Internet móvil? 0 (70.6%)	S3.1a. ¿Desde hace cuánto tiempo tiene el servicio de Internet móvil? 0 (62.2%)	S3.1a. ¿Desde hace cuánto tiempo tiene el servicio de Internet móvil? 0 (55.0%)	S3.1a. ¿Desde hace cuánto tiempo tiene el servicio de Internet móvil? 0 (82.5%)

Efectúe una doble pulsación para activar

Tablas personalizadas

[Conjunto_de_datos1] E:\Katiuska\tesis\base datos 2023\datos_030723.sav

Número de conglomerados en dos fases

IBM SPSS Statistics Processor está listo Unicode:ACTIVADO

21:07
30/06/2024

*Resultado_030723.spv [Documento1] - IBM SPSS Statistics Visor

Archivo Editar Ver Datos Transformar Insertar Formato Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda

	Edad	Edad	Edad	Edad
	41.32	39.16	37.04	43.85
	Nivel de instrucción 2 (34.3%)	Nivel de instrucción 1 (62.6%)	Nivel de instrucción 1 (66.9%)	Nivel de instrucción 2 (70.6%)
	Área 0 (96.8%)	Área 0 (89.7%)	Área 0 (67.8%)	Área 0 (99.5%)
	S3.1a. ¿Desde hace cuánto tiempo tiene servicio de Internet móvil? 0 (70.6%)	S3.1a. ¿Desde hace cuánto tiempo tiene servicio de Internet móvil? 0 (62.2%)	S3.1a. ¿Desde hace cuánto tiempo tiene servicio de Internet móvil? 0 (55.0%)	S3.1a. ¿Desde hace cuánto tiempo tiene servicio de Internet móvil? 0 (82.5%)

Efectúe una doble pulsación para activar

Tablas personalizadas

[Conjunto_de_datos1] E:\Katiuska\tesis\base datos 2023\datos_030723.sav

Número de conglomerados en dos fases

IBM SPSS Statistics Processor está listo Unicode ACTIVADO

9°C 21:07 30/06/2024

*Resultado_030723.spv [Documento 1] - IBM SPSS Statistics Visor

Archivo Editar Ver Datos Transformar Insertar Formato Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda

Frecuencias

Nivel socioeconómico del hogar

		A		B		C		D		E	
		Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Conglomerado	1	0	0.0%	0	0.0%	6062	100.0%	11	0.2%	0	0.0%
	2	126	100.0%	1552	100.0%	0	0.0%	0	0.0%	0	0.0%
	3	0	0.0%	0	0.0%	0	0.0%	5070	99.8%	0	0.0%
	4	0	0.0%	0	0.0%	0	0.0%	0	0.0%	2959	100.0%
	Combinados	126	100.0%	1552	100.0%	6062	100.0%	5081	100.0%	2959	100.0%

* Tablas personalizadas.
CTABLES
/VLABELS VARIABLES=q4 TSC_4012 DISPLAY=LABEL
/TABLE q4 [COUNT F40.0] BY TSC_4012
/CATEGORIES VARIABLES=q4 TSC_4012 ORDER=A KEY=VALUE EMPTY=INCLUDE.

Clústeres

Importancia de entrada (predictor)
■ 1.0 ■ 0.8 ■ 0.6 ■ 0.4 ■ 0.2 ■ 0.0

Clúster	1	3	4	2
Etiqueta				
Descripción				
Tamaño				

IBM SPSS Statistics Processor está listo Unicode:ACTIVADO H: 391, W: 488 pt. 21:06 30/06/2024