



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ciencias Matemáticas

Escuela Profesional de Estadística

Segmentación de clientes en un supermercado según su comportamiento de compras

TRABAJO DE SUFICIENCIA PROFESIONAL

Para optar el Título Profesional de Licenciada en Estadística

AUTOR

Diana Guadalupe SÁNCHEZ RIVERA

ASESOR

Mg. Zoraida Judith HUAMÁN GUTIERREZ

Lima, Perú

2021



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Sánchez, D. (2021). *Segmentación de clientes en un supermercado según su comportamiento de compras*. [Trabajo de suficiencia profesional de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Matemáticas, Escuela Profesional de Estadística]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Diana Guadalupe Sánchez Rivera
Tipo de documento de identidad	DNI
Número de documento de identidad	47160891
URL de ORCID	No aplica
Datos de asesor	
Nombres y apellidos	Zoraida Judith Huamán Gutierrez
Tipo de documento de identidad	DNI
Número de documento de identidad	09890094
URL de ORCID	0000-0003-1862-8077
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Oscar Antonio Robles Villanueva
Tipo de documento	DNI
Número de documento de identidad	32762171
Miembro del jurado 1	
Nombres y apellidos	Ricardo Luis Pomalaya Verastegui
Tipo de documento	DNI
Número de documento de identidad	10460674
Datos de investigación	

Línea de investigación	A.3.2.6. Análisis de Datos y Modelamiento de Problemas de la Sociedad (Empresas, Instituciones, Poblaciones locales, regionales y nacionales).
Grupo de investigación	No aplica
Agencia de financiamiento	Sin financiamiento.
Ubicación geográfica de la investigación	Edificio: Caja Rural de Ahorro y Crédito Cencosud Scotia Perú S.A País: Perú Departamento: Lima Provincia: Lima Distrito: Miraflores Urbanización: Miraflores Calle: Av. Benavides N° 1555 piso 10 Latitud: -12.124029835371605 Longitud: -77.01700780974137
Año o rango de años en que se realizó la investigación	01 de Enero 2021 al 01 Junio 2021
URL de disciplinas OCDE	Estadística, Probabilidades https://purl.org/pe-repo/ocde/ford#1.01.023



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

Universidad del Perú. Decana de América
FACULTAD DE CIENCIAS MATEMÁTICAS
ESCUELA PROFESIONAL DE ESTADÍSTICA

ACTA DE SUSTENTACIÓN DE TRABAJO DE SUFICIENCIA PROFESIONAL EN LA MODALIDAD VIRTUAL PARA OBTENCIÓN DEL TÍTULO PROFESIONAL DE LICENCIADA EN ESTADÍSTICA

En Lima, siendo las 17:30 horas del domingo 03 de octubre del 2021, se reunieron los docentes designados como Miembros del Jurado del Trabajo de Suficiencia Profesional: Dr. Oscar Antonio Robles Villanueva (PRESIDENTE), Mg. Ricardo Luis Pomalaya Verastegui (MIEMBRO) y la Mg. Zoraida Judith Huamán Gutierrez (MIEMBRO ASESOR), para la sustentación del Trabajo de Suficiencia Profesional titulado: “**SEGMENTACIÓN DE CLIENTES EN UN SUPERMERCADO SEGÚN SU COMPORTAMIENTO DE COMPRAS**”, presentado por la señorita **Bachiller Diana Guadalupe Sánchez Rivera**, para optar el Título Profesional de Licenciada en Estadística.

Luego de la exposición del trabajo de suficiencia, el Presidente invitó a la expositora a dar respuesta a las preguntas formuladas.

Realizada la evaluación correspondiente por los miembros del Jurado Evaluador, la expositora mereció la aprobación de **BUENO**, con un calificativo promedio de **DIECISEIS (16)**

A continuación, los miembros del Jurado dan manifiesto que la participante **Bachiller Diana Guadalupe Sánchez Rivera** en vista de haber aprobado la sustentación del Trabajo de Suficiencia Profesional, será propuesta para que se le otorgue el Título Profesional de Licenciada en Estadística.

Siendo las 18:00 horas se levantó la sesión firmando para constancia la presente Acta.

Dr. Oscar Antonio Robles Villanueva
PRESIDENTE

Mg. Ricardo Luis Pomalaya Verastegui
MIEMBRO

Mg. Zoraida Judith Huamán Gutierrez
MIEMBRO ASESOR

La Vicedecana de la Facultad de Ciencias Matemáticas, Mg. Zoraida Judith Huamán Gutiérrez, certifica virtualmente la participación del Jurado Evaluador, el titulado, el acto de instalación y el inicio, desarrollo y término del acto académico de sustentación, dejando constancia en el acta respectiva.

RESUMEN

El presente trabajo tiene como objetivo determinar grupos de clientes en la empresa para aplicar un marketing focalizado y programas de fidelización. Una adecuada segmentación proporciona un mejor entendimiento sobre cada grupo de clientes dando a la empresa una mejor visión para el otorgamiento de sus productos y que generen mayor valor. Para la realización del objetivo, se realizará el algoritmo de clasificación no supervisada de k-means siguiendo los pasos principales como estableciendo el valor de k, asignando los centroides más cercanos a los datos y actualizando hasta que los centroides se establezcan, junto a este método se analizó la técnica de RFM para estudiar el comportamiento del cliente con sus tres variables principales Recencia: Periodo desde la última compra, Frecuencia: Número de compras realizadas en el último periodo y Monetario: dinero gastado durante un cierto periodo. Los resultados nos ayudaran a crear segmentos de clientes efectivos a conocer mejor a los clientes mediante su comportamiento de compras y poder otorgarles mejores beneficios y ofertas de campañas.

ABSTRACT

The objective of this work is to determine customer groups in the company to apply targeted marketing and loyalty programs. Adequate segmentation provides a better understanding of each group of clients, giving the company a better vision for the granting of its products and that they generate greater value. To achieve the objective, the unsupervised k-means classification algorithm will be carried out following the main steps such as demonstrating the value of k, assigning the centroids closest to the data and updating until the centroids are established, together with this method. The RFM technique was analyzed to study customer behavior with its three main variables Recency: Period since the last purchase, Frequency: Number of purchases made in the last period, and Monetary: money spent during a certain period. The results will help us to create effective customer segments to get to know customers better through their purchasing behavior and to be able to grant them better benefits and campaign offers.

ACTA DE SUSTENTACIÓN

Tabla de contenidos

RESUMEN	2
Índice de Figuras	6
Índice de Tablas	vii7
I Introducción	8
II Información del lugar donde se desarrolló la actividad	9
• Institución donde se desarrolló la actividad.....	9
• Periodo de duración de la actividad.....	ιξ
• Finalidad y objetivos de la entidad	ιξ
• Razón social.....	ιξ
• Dirección postal	ιξ
• Correo electrónico del profesional a cargo	ιξ
III Descripción de la actividad	ix
• Organización de la actividad.....	ιξ
• Finalidad y objetivos de la actividad	10
• Problemática	10
• Metodología, Procedimientos	ξιι
• Resultados de la actividad.....	ξιιι
IV Conclusiones	26
V Recomendaciones	27
VI Bibliografía	28
Anexo	299

Índice de Figuras

Figura 1	17
Figura 2	17
Figura 3	17
Figura 4	17
Figura 5	18
Figura 6	18
Figura 7	18
Figura 8	18
Figura 9	19
Figura 10	19
Figura 11	19
Figura 12	19
Figura 13	19
Figura 14	19
Figura 15	20
Figura 16	23

Índice de Tablas

Tabla 1	21
Tabla 2	24
Tabla 3	25
Tabla 4	29
Tabla 5	30
Tabla 6	30

I Introducción

El Sector económico de Retail es uno de los más importantes en Latinoamérica (BBVA Research 2018) que está conformado por entidades especializadas en consumo destacando los Supermercados, tiendas por departamentos y tiendas por especialidad, los cuales en los últimos años tuvo un importante crecimiento generando alta rentabilidad en las empresas por su alto impacto en el producto bruto per cápita e incentivando el consumo interno y puestos de trabajo.

En el Perú en el último año se vio afectado por la crisis sanitaria y económica generado por el COVID 19, entre los sectores más afectados fueron construcción, minería, restaurantes, alojamientos, comercio y consumo lo cual llevo a varias empresas reestructurar sus estrategias de campaña (BCRP, 2020)

La entidad en estudio, nació de una alianza estratégica de dos grupos empresariales nacional e internacional, dedicándose exclusivamente en el financiamiento de compras y productos en efectivo teniendo como propósito bienestar social en el Perú a través de soluciones financieras de consumo enfocados en los atributos de convivencia y accesibilidad. Actualmente las agencias se encuentran en la red de supermercados a nivel nacional y tenemos como producto principal la Tarjeta de crédito, así como seguros y CTS, siendo este segmento tan afectado debido al cierre de los comercios durante el periodo de confinamiento y afectando en el otorgamiento de créditos a personas (Apoyo Asociado, septiembre 2020).

Este trabajo se realizó teniendo en cuenta el siguiente procedimiento, en el Capítulo II se describe el rubro de la entidad donde se aplica el trabajo de suficiencia profesional, se define la finalidad y cuál es el objetivo de entidad como empresa y el periodo en el cual fue desarrollado. En el Capítulo III se describe la aplicación del método de segmentación y el método de RFM, el objetivo por el cual fue utilizado dichos métodos, se explica la problemática de acuerdo antecedentes similares a nuestro trabajo y se muestra los resultados obtenidos. En el Capítulo IV se explica los resultados obtenidos de los métodos aplicados para resolver dicha problemática. En el Capítulo V se menciona las recomendaciones para la implementación de dicho trabajo realizado en la entidad donde se desarrolla la actividad, finalmente en el Capítulo VI citamos la bibliografía utilizada como apoyo para la realización del presente trabajo.

II Información del lugar donde se desarrolló la actividad

- Institución donde se desarrolló la actividad

Entidad Finanzas - Retail

- Periodo de duración de la actividad

11 de enero de 2021 hasta la 11 de julio 2021

- Finalidad y objetivos de la entidad

Finalidad

Ofrecer soluciones financieras de consumo a los clientes del mismo grupo, basándose en los atributos de conveniencia y accesibilidad. Su visión, está direccionada a ser un referente en servicio, innovación y búsqueda de soluciones dentro del sector retail.

Objetivo

Ser un referente de servicio, innovación y búsqueda de soluciones dentro del ámbito del Retail financiero a través del conocimiento del cliente y así generar valor agregado en todas las operaciones de Retail.

- Razón social

Sector Retail financiero

- Dirección postal

Av. Benavides N° 1555 piso 10, Miraflores,

- Correo electrónico del profesional a cargo

d.sanchez1149@gmail.com

III Descripción de la actividad

- Organización de la actividad

Como primer paso se realizó un análisis exploratorio de las variables para tener una mejor vista del comportamiento de cada una, posteriormente se realizó la selección de variables para determinar las características que identifica los grupos, aplicando un enfoque inductivo, cognitivo y criterio experto lo cual refleja la realidad de la empresa, dado que se tienen variables en distintas unidades de medida se llevó a cabo la estandarización y la completitud de los datos.

Como segundo paso se aplicó un método de agrupación interactivo (Jerárquico y No jerárquico) utilizando el método de K mean con el que se van a ir ajustando los grupos asignando un valor de k.

Luego se determinó el número de grupos con una aceptable precisión teniendo como métrica la distancia media entre los puntos de datos de los centroides

Finalmente se define los grupos estratégicos con ayuda de los segmentos encontrados bajo la metodología de RFM, se realiza una comparación entre sus parámetros y la segmentación clúster para luego distribuir las campañas de marketing al usuario final.

- Finalidad y objetivos de la actividad

Finalidad

Determinar grupos de clientes en la empresa para aplicar estrategias focalizadas.

Objetivo

Identificar Segmentos de clientes según estilos de compra

Identificar clientes que generen mayor valor a la empresa

Identificar a los clientes con mayor frecuencia de compras y que dejaron de hacerlo en el último mes de estudio.

Detectar segmentos otorgarles mejores beneficios y fidelizarlo.

- Problemática

Rojas (2020) menciona en su investigación, busca segmentar a los clientes para proporcionar una buena comprensión de la necesidad de cada grupo y ofrecerles el producto que realmente necesiten, en el proceso se utilizó algoritmos de agrupamiento k-means para encontrar las marcas más rentables de la empresa, se halló un modelo con 4 clústeres Muy Alto, Alto, Medio y bajo, donde el clúster Muy Alto agrupa la mayor cantidad de marca de productos (304) pero con un costo medio más bajo (8,664) valor venta medio con un coeficiente de silueta de 0.83 agrupando mucho mejor que los demás modelos planteados, a la vez se identificó que los productos que se venden mas no siempre son los que generan mayor ingreso.

Cuadros, (2017) en su investigación busca construir una relación exitosa entre sus clientes para implementar estrategias de marketing más dirigidas. Dentro de la metodología seguida abarca varias etapas partiendo de la data inicial hasta el agrupamiento de las variables. La inclusión de un nuevo método permite comparar los resultados de segmentación con los parámetros de la metodología de RFM y su valor agregado. Como resultado se obtuvo 5 grupos de clientes siendo el Grupo 1 el que concentra menor número de clientes (22) y mejores calificaciones, pero con una ganancia baja (38%) por ser clientes mayoristas le ofrecen descuentos, en el Grupo 2 se concentra la mayor grupo de clientes (107) con variables muy similares que el Grupo 4 pero con ganancias no tan altas (28%) este grupo se mantiene para retener al cliente con el proveedor y siendo el Grupo 5 y 3 tienen las más bajas calificaciones por ser clientes que no realizan una compra hace más de un año pero que generaron una alta ganancia (59%) encontrando así grupos estratégicos para realizar una campaña dirigida.

Tang, Vargas, (2016). enfoca su tesis en el estudio de clientes que realizaron una compra mediante un crédito en una tienda de retail. Por la naturaleza de la información se encontró 3 conglomerados distinguidos como “Los caza ofertas”, “Los Estacionales” y “Consumidores cautos” obteniendo un equilibrio entre los conglomerados, primer conglomerado 1817 clientes (36.5%), segundo conglomerado con 1309 clientes (27.9%) y el tercer conglomerado 1773 clientes (35.6%) y proponiendo estrategias mix al primer y tercer conglomerado buscando la fidelización e incrementar las colocaciones.

Aerts, (2020) en su tesis de maestría se enfoca en encontrar segmentos de clientes que realizan préstamos directos (P2P) de la Fintech Lending Club realizando un seguimiento a los clientes utilizando K-mean, DBSCAN y modelos de mezcla gaussiana para identificar segmentos de clientes y emplear estrategias. Como resultado se obtuvo que la agrupación por K-means es la más adecuada realizando partición, se divide los clústeres según nivel de ingresos distinguiendo 6 particiones de clúster, clúster 1,2 y 3 representan a clientes con ingresos considerablemente bajo, los clientes del Clúster 4 y 5 tienen ingresos medios mientras que Clúster 6 tienen ingresos altos.

Van, (2017) en su tesis se enfoca en las empresas minoristas los cuales son un foco importante en el mercado dado que definen grupos de tiendas con precios diferenciados. Esta investigación proporciona una solución de agrupación que define agrupaciones de tiendas basadas en el precio. Esta solución de agrupación demuestra potencial con un aumento proyectado en ingresos del 0,36% y un aumento proyectado en las ganancias del 0,76% mediante el uso de datos de uno de los principales cadenas de supermercados de los Países Bajos.

- Metodología, Procedimientos

- **Enfoque**

Cuantitativo

- **Diseño**

El Diseño, No Experimental

- **Nivel**

Explicativo, descriptivo

- **Población de estudio**

La población está conformada por todos los clientes que se encuentran vigentes en el periodo enero 2021 y que realizaron de una a más compra en la tienda.

- **Unidad de análisis**

Cada cliente vigente en el periodo enero 2021.

- **Muestra:**

No presenta muestra, se trabaja con toda la población de clientes.

Variables

Variable Objetivo

Se define como variable objetivo al Comportamiento de compra que realiza un cliente, como el total de compras que realizó el cliente en el mes de enero 2020

Variables Estadísticas

- Mes: mes que se afilio como cliente
- Tipo de cliente: Si el cliente adquirió un tipo de beneficio por alguna deuda obtenida en la entidad
- Línea: monto que se le otorgó al cliente de libre disponibilidad
- Grupo: Agrupación por política de la empresa
- Monto de Compras: Monto total que el cliente utilizo para las compras en el supermercado
- Tipo de actividad: tipo de trabajo del cliente (dependiente o independiente)
- Ingreso: Ingreso bruto del cliente
- Porcentaje de efectividad: Nivel de endeudamiento del cliente
- Puntaje A: puntaje asignado por la empresa 1
- Puntaje B: puntaje asignado por la empresa 2
- Edad: Edad del cliente
- Ultima fecha: fecha de ultima compra
- Género: Sexo del cliente
- Estado civil: Estado civil del cliente

Método

Los métodos para problemas de agrupamiento de clientes caen dentro de los algoritmos de machine learning llamados No Supervisados o No Jerárquicos.

Método No Jerárquico

Las técnicas de agrupación no jerárquica están diseñadas para agrupar elementos, en lugar de variables, en una colección de k agrupaciones, el número de agrupaciones k puede especificarse de antemano o determinarse como parte del procedimiento de agrupación. Debido a que es necesario determinar una matriz de distancias (similitudes), los métodos no jerárquicos se pueden aplicar a conjuntos de datos mucho más grandes que las técnicas jerárquicas. Los métodos no jerárquicos comienzan de una partición inicial de elementos en grupos o de un conjunto inicial de puntos de semilla, que formarán los núcleos de los grupos. Las elecciones para los puntos iniciales deben estar libres de sesgos, comienza con la elección al azar de puntos semilla de entre los elementos o dividir al azar el elemento en grupos iniciales (Johnson, 2019)

Método de k-mean

Esta metodología busca identificar grupos de observaciones particionado en k clúster mediante particiones donde el valor de k es asignado.

- Una consideración es que antes de ejecutar el algoritmo debe definirse el número de clúster cuya denotación es k
- Cada punto representa un clúster conocido como su centroide

Inicialización de los Centroides

En primera instancia se inicializa los centroides asignando k puntos de forma aleatoria de todo el conjunto de datos, es decir para cada observación x_k se calcula el clúster más próximo a su centroide (media de cada una de las observaciones) formando un conjunto de elementos (Jirones, 2017)

$$c = C_1, C_2 \dots C_k$$

Donde cada C_k esta conformado por un conjunto de observaciones $C_k = \{x_{k1}, x_{k2} \dots x_{km}\}$

El conjunto de elementos se desplaza hacia el centroide. Dicha característica de la asignación permite ejecutar el algoritmo múltiples veces con distintos centroides y elegir el mejor resultado.

$$u_k = \frac{\sum_{k=1}^n x_{ki}}{m} \quad m: \text{dimensión}$$

Cálculo de la distancia

Se calcula la suma de las distancias euclídeas al cuadrado entre cada observación (x_i) y el centroide (μ) de su cluster. Esto equivale a la suma de cuadrados internos del clúster.

$$W(C_k) = \sum_{x_i \in C_k} (x_i - u_k)^2$$

Luego se calcula la suma de las distancias euclídeas al cuadrado entre todos los pares de observaciones que forman los clústeres.

$$W(C_k) = \sum_{i=1}^k \sum_{n=1}^m (x_n - u_i)^2$$

x_i observación perteneciente a un clúster

u_k valor medio de los puntos asignados al clúster C_k

Se busca que la suma de cuadrados total dentro del conglomerado mida la bondad del conglomerado lo cual se busca que sea lo más mínimo posible.

El algoritmo se repite con diferentes puntos iniciales y se escoge el resultado con el mínimo valor de la función objetivo.

Criterio para seleccionar el valor de k

La principal característica del método de k-Mean es que se le asigna el número de particiones (k) con lo que se puede probar diferentes valores y seleccionar el más adecuado para el trabajo estudiado.

Uno de los criterios es minimizar la suma de las distancias de cualquier vector a su centroide más cercano buscado que los segmentos sean más concisos. Otro Criterio es maximizar la suma de distancia entre sus centros de cada segmento obteniendo segmentos diferenciados.

Método de RFM

La segmentación por comportamiento es el proceso de "dividir un mercado en grupos de clientes con necesidades similares, y desarrollar programas de marketing que satisfagan esas necesidades" La segmentación conductual se centra en dividir a los clientes por su comportamiento pasado.

La segmentación por comportamiento es una forma de hallar grupos de clientes según las acciones que demuestran en la intención de compra de un producto basado en tres pilares Recencia, Frecuencia y Valor monetario lo cual se define de la siguiente manera:

Recencia: mide cuando fue la última interacción del cliente con la tienda, respecto del momento temporal que nos interese para el estudio; éste podrá ser a final de mes en informes mensuales, a finales de año para memorias anuales o cuando corresponda según la periodicidad que corresponda a cada caso.

Es importante conocer el espacio de tiempo que el usuario hace que no interactúa con nosotros y, a partir de este indicador, podremos definir qué contenido de interés le podemos comunicar.

Frecuencia: Cuantifica el número de veces que el cliente interactúa con la tienda. Cómo en la recencia, deberemos definir qué acciones suman en nuestro indicador y deberán tener coherencia con las acciones anteriormente contempladas.

Valor monetario: Calcula el valor que tiene para nosotros el usuario. Este valor, no necesariamente tiene que ser siempre monetario. Si estamos en un ecommerce, sí lo podrá ser sumando por ejemplo el valor de la cesta promedio o la suma total de gasto por usuario en un tiempo determinado; pero si queremos medir, por ejemplo, el engagement de un usuario, este valor podría ser el número de interacciones que hace con nosotros y que nos nombra en redes sociales por ejemplo, o si estamos en retail, el número de puntos que nos llega a canjear en un espacio de tiempo determinado.

- **Resultados de la actividad**

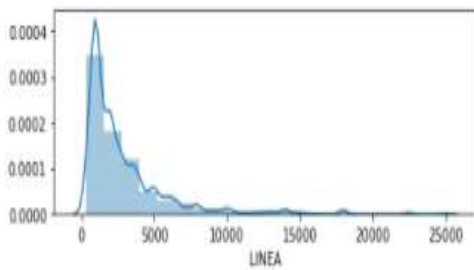
Para la siguiente actividad se aplicó el método de k mean a una base de datos generada por clientes que realizan compras en un supermercado usando la tarjeta de crédito de la entidad asociada.

Recopilación de datos

Para la recopilación o recolección de datos se tomó en cuenta todas las compras que realiza un cliente de diversos productos que se encuentran disponibles en la tienda de supermercado y cuyo registro de compras se almacenan en el sistema interno de información de forma diaria logrando obtener 16 variables para nuestro análisis con lo cual se procedió a realizar un análisis exploratorio de datos.

Figura 1

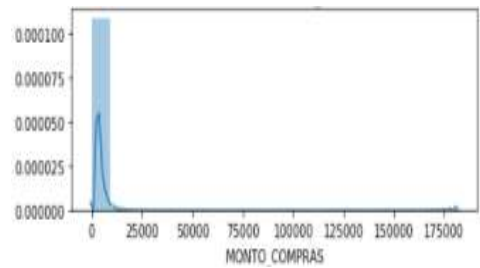
Distribución de Línea utilizada



Fuente: Elaboración propia

Figura 2

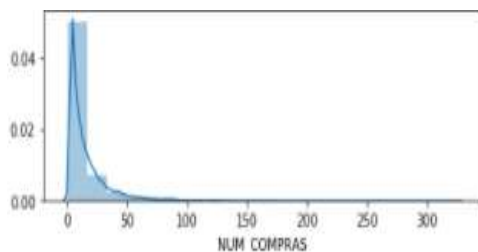
Distribución de Monto de Compras



Fuente: Elaboración propia

Figura 3

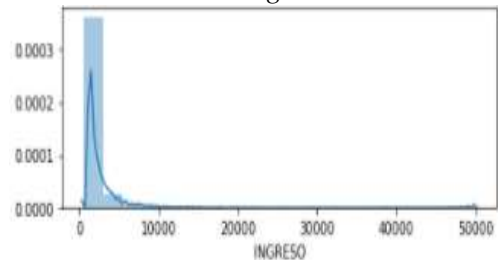
Distribución de Número de Compras



Fuente: Elaboración propia

Figura 4

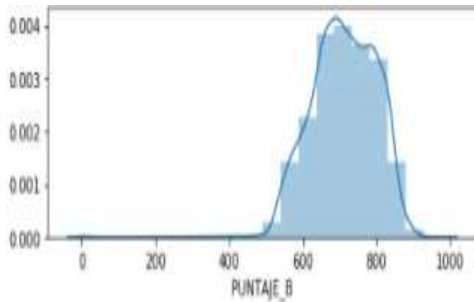
Distribución de Ingreso



Fuente: Elaboración propia

Figura 5

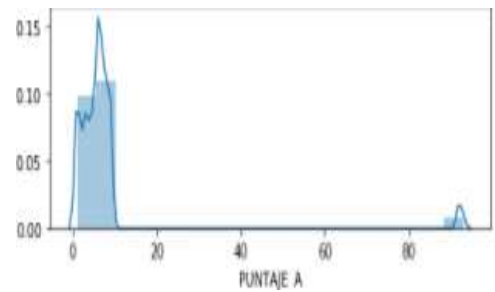
Distribución de Puntaje B



Fuente: Elaboración propia

Figura 6

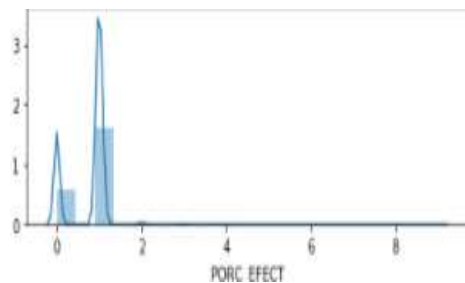
Distribución de Puntaje A



Fuente: Elaboración propia

Figura 7

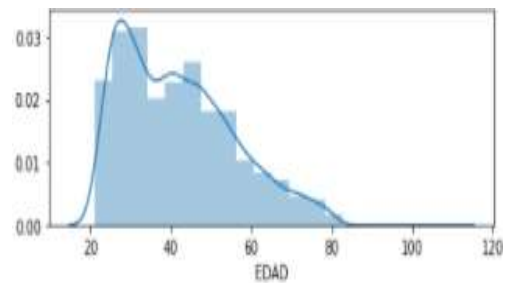
Distribución de % de efectividad



Fuente: Elaboración propia

Figura 8

Distribución de Edad



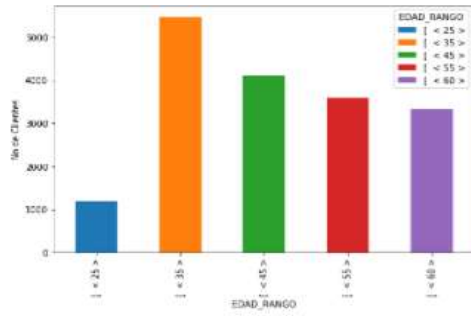
Fuente: Elaboración propia

Se observa las distribuciones de las variables cuantitativas Línea utilizada, Monto de Compras, Número de Compras, Ingreso, Puntaje A, Edad con asimetría a la derecha lo cual y Puntaje B con asimetría a la izquierda los cuales en el capítulo siguiente se explica cómo han sido tratadas para el análisis.

Se realizó análisis exploratorio de las variables cualitativas.

Figura 9

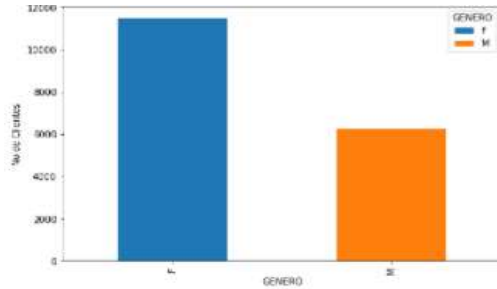
Distribución de Edad



Fuente: Elaboración propia

Figura 10

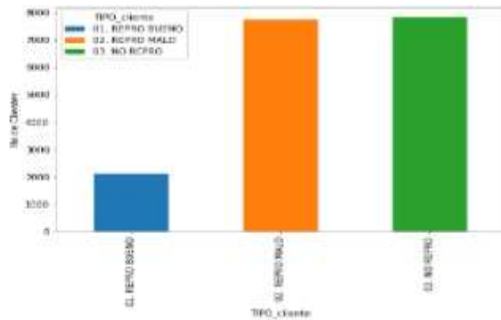
Distribución de Género



Fuente: Elaboración propia

Figura 11

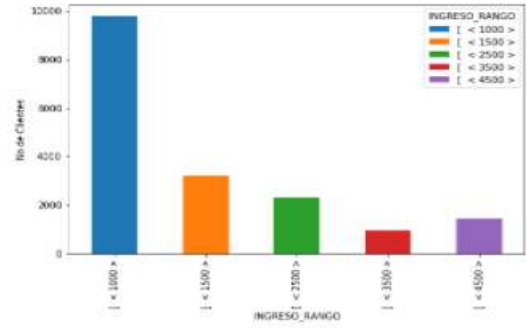
Distribución de Tipo de Cliente



Fuente: Elaboración propia

Figura 12

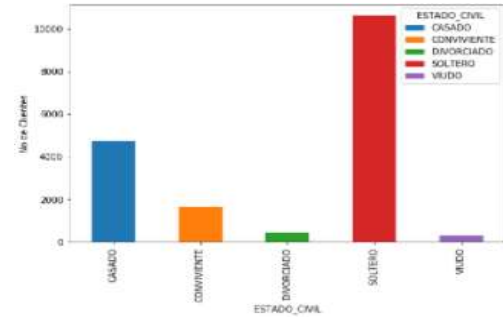
Distribución de Ingreso



Fuente: Elaboración propia

Figura 13

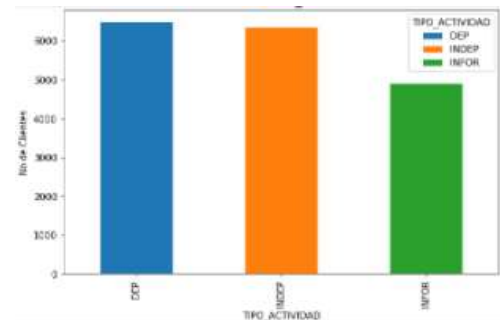
Distribución de Estado Civil



Fuente: Elaboración propia

Figura 14

Distribución de Tipo de Actividad



Fuente: Elaboración propia

Pre procesamiento de los datos

En el pre procesamiento se aplicó técnicas para la limpieza de datos, como los datos faltantes y datos atípicos y la transformación de datos para que finalmente obtener una depuración y comenzar con el análisis.

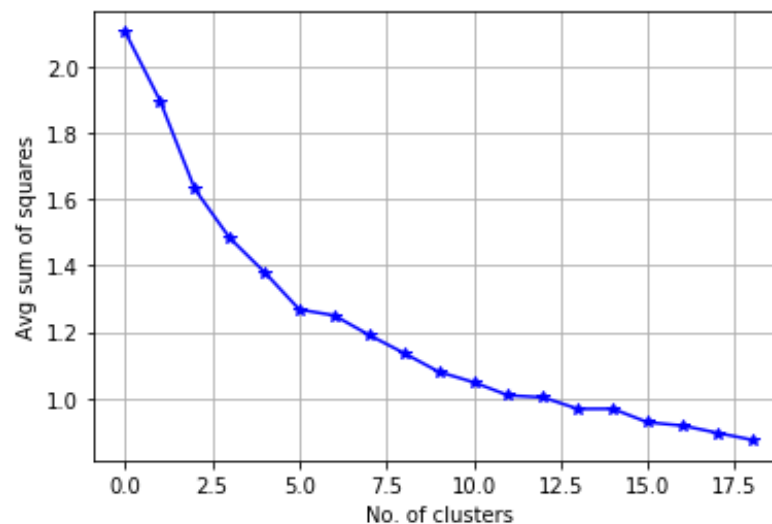
Se procedió con la estandarización de los datos eliminando las unidades de medida y no afecte en el momento del análisis mencionar que en el proceso de transformación solo ingresaron las variables cuantitativas.

Selección del número óptimo para k

De acuerdo a varias inicializaciones del valor de k, para el análisis se decidió tomar el valor de k como k=5, se tomó como referencia el método de codos, en la Figura 15 se observa que en el punto k=5 hay un cambio en la tendencia indicando a la vez el punto óptimo para la definición del valor de k.

Figura 15

Número óptimo de clúster



Fuente: Elaboración propia

Dado que el objetivo es encontrar segmentos de clientes según su comportamiento de compras vamos a elegir el modelo de clusterización que nos de la mejor y más detallada información (en este caso el de k=5).

Tabla 1

Construcción de Segmentos

Clúster	Segmento	Variables de Discriminación	Valor	% Participación
1	Consumidores Promocionales	Línea, N° de compras, Monto de compras, Ingreso, %Efectividad, Puntaje A, Puntaje B, Edad, Línea utilizada, Género, Grupo, Tipo de cliente, Tipo de Actividad	25 - 35	42%
2	Consumidores Casa-Hogar	Línea utilizada, Ingreso, Edad, %Efectividad, Puntaje B, Edad, Línea, Género, Estado civil, Grupo, Tipo de cliente, Tipo de Actividad	55+	28%
3	Consumidores Ahorradores	N.º de compras, Monto de compras, Puntaje B, Edad, Línea utilizada, Género, Estado Civil, Grupo, Tipo de cliente, Tipo de Actividad	25 - 35	16%
4	Consumidores Cautelosos	Puntaje A, Género, %Efectividad, Estado Civil, Tipo de Actividad	55 +	9%
5	Consumidores Estacional	Línea, N.º de compras, Monto de compras, Ingreso, Puntaje A, Puntaje B, Edad, Línea utilizada, Género, Grupo, Tipo de cliente, Tipo de Actividad	45 - 55	5%

Fuente: Elaboración propia

Consumidores Promocionales

Este segmento está conformado por clientes que buscan alguna promoción en el supermercado o algún descuento, del cual el 27% pertenecen al género femenino, 20% son solteras y el 19% trabaja de forma independiente, tienen edad menor a 35 años y cuyo ingreso oscila en el rango de S/1,000 y S/ 1,500

Consumidores Casa-Hogar

Este segmento está conformado por clientes que realizan sus compras cada cierto periodo en especial productos de alimentos, donde el 55% pertenecen al género femenino, 11% son solteras y el 11% trabaja de forma dependiente, tienen edad mayor a 55 años y cuyo ingreso es mayor a S/ 4,500

Consumidores Ahorradores

Este segmento está conformado por clientes que realizan sus compras en el supermercado, pero buscando ahorrar, donde el 60% pertenecen al género femenino, 57% son solteras y el 73% trabaja de forma Dependiente, tienen edad que oscila entre a 25 y 35 años y cuyo ingreso es mayor a S/ 4,500

Consumidores Cautelosos

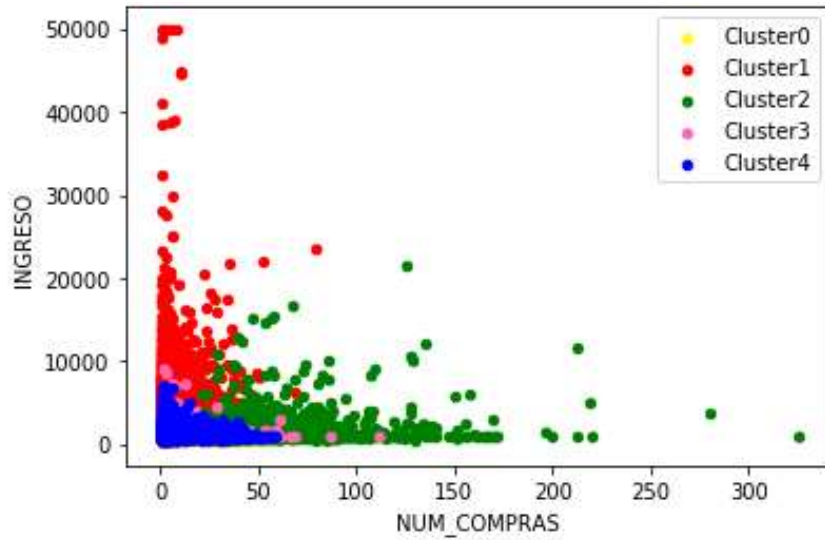
Este segmento está conformado por clientes que tienen un buen comportamiento en el manejo de su tarjeta de crédito a momento de realizar las compras de algún producto en el supermercado, el 90% está conformado por clientes que tienen un trabajo de forma dependiente, el 52% son solteros, el 59% son del género masculino cuyo ingreso oscila en el rango de S/3,500 y S/ 4,500

Consumidores Estacional

Este segmento está conformado por clientes que realizan compras en fechas especiales donde es más transcurrido visitar la tienda, donde el 18% pertenecen al género femenino, 20% son solteras y el 19% trabaja de forma independiente, tienen edad menor a 35 años y cuyo ingreso oscila en el rango de S/1,000 y S/ 3,500

Figura 16

Gráfico de clúster según Numero de compras y nivel de ingreso



Fuente: Elaboración propia

A continuación, se presenta un cuadro resumen de las variables más diferenciadas según cada tipo de clúster o segmento.

Tabla 2*Tipificación según Clúster de las variables cuantitativas*

CLUSTER	LINEA UTILIZADA	NUMERO DE COMPRAS	MONTO DE COMPRAS	INGRESO	% EFECT	PUNTAJE A	PUNTAJE B	EDAD	LINEA	GENERO FEMENINO	GENERO MASCULINO	ESTADO CIVIL	GRUPO	TIPO CLIENTE	TIPO DE ACTIVIDAD
1	*	*	*	*	*	*	*	*	*	*		Soltero	*	*	Independientes
2	*			*	*		*	*	*	*		Casado	*	*	Dependientes
3		*	*							*		Soltero			Dependientes
4					*	*					*	Casado			Independientes
5	*	*	*	*		*	*	*	*	*		Soltero	*	*	Dependientes

Fuente: Elaboración propia

En esta sección se realizó la segmentación bajo el análisis de RFM mediante las variables de Frecuencia, Recencia y Valor Monetario, donde los clientes pueden ser clasificados por el estatus del volumen de sus compras en periodo determinado, por la frecuencia de compra en el supermercado y por su antigüedad desde la última compra realizada.

El cálculo se realizó para cada cliente asignando una puntuación lo cual permite realizar finalmente la segmentación, en la Tabla 3 se observa los 5 segmentos resultantes.

Tabla 3

Segmentación según metodología RFM

Segmentos	Participación	Frecuencia	Monetario	Recencia
1	62%	16%	14%	14
2	21%	23%	27%	14
3	10%	21%	21%	15
4	4%	13%	13%	16
5	4%	25%	26%	16

Fuente: Elaboración propia

Aplicando la metodología RFM busca identificar segmentos a través de las variables Recencia, Frecuencia y valor Monetario en la cual se identificó 5 segmentos:

Clientes Top

Este segmento 5 está conformado por el 4% de los clientes el cual representa el 26% de ventas totales mostrando un segmento más pequeño de clientes potenciales dado que representa la tercera parte de las ventas totales muy similar al segmento 2 (27%) pero a diferencia que está conformado por menos clientes.

Clientes Grandes

Este segmento 4 tiene una participación del 4% de clientes y generan un volumen de ventas del 13% muy similar al segmento 1, de esta forma se identifica un nicho de clientes potenciales dado que la concurrencia a la tienda en promedio lo realiza cada 16 días después de su última compra.

Clientes Medios

Este segmento 3 presenta el 10% de participación de clientes obteniendo el 21% de ventas

Clientes Bajos

el segmento 2 presenta el 21% de participación de clientes registrando el doble de ventas

Clientes Regulares

Estos clientes presentan el mayor porcentaje de participación de clientes (62%) se encuentra en el segmento 1 el cual presenta el 14% de ventas teniendo un promedio de 14 días desde la última compra realizada en el supermercado.

IV Conclusiones

El trabajo concluye los siguiente:

- Se identifico cinco segmentos bajo la metodología de k-mean, pudiendo hallar un perfil de cliente para cada segmento bajo su comportamiento de compras donde el segmento 1 está conformado por el 42% de clientes el segmento 2 por el 28% el segmento 3 por el 16% de clientes el segmento 4 por el 9% de clientes y el segmento 5 por el 8% de clientes.
- A cada segmento hallado se le asignó un nuevo nombre estratégicamente para su identificación de los análisis futuros.
- Bajo el análisis de RFM se identificó que el 5% de clientes generan el 25% de las ventas totales asignando a este grupo de clientes como clientes Top
- Se identifico que los clientes en promedio acuden a realizar una compra al supermercado cada 16 días

V Recomendaciones

De los resultados obtenidos se identificó las siguientes recomendaciones:

- Considerar más variables de estudio que ayuden a dar una mejor descripción de cada segmento y poder detectar un mejor perfil del cliente, una de las posibles variables sería, zona o lugar de la tienda donde realiza la compra, tipos de producto que compra con más frecuencia, medio de pago.
- Comparar la técnica que se aplicó en este caso K-mean con otras técnicas supervisadas o no supervisadas para analizar las diferencias en las segmentaciones que resulta de su aplicación.
- Enriquecer las estrategias de campañas para los segmentos hallados bajo la metodología de k-mean con los segmentos propuestas por el método de RFM.

VI Bibliografía

- Cuadros, A., Gonzales, C., & Jiménez, P. (2017). Analisis Multivariado para segmentación de clientes basado en RFM. *Tecnura*, 44-51.
- Gironés, J., Casas, J., Minguillón, J., & Caihuelas, R. (2017). *Minería de Datos modelos y algoritmos*. Barcelona: Editorial UOC.
- Hair, J., Anderson, R., Tatham, R., & Black, W. (1999). *Análisis Multivariante*. Madrid: Prentice Hall Iberia.
- Jin, X., & Han, J. (2011). *K -Means Clustering*. En: Sammut C., *Enciclopedia de aprendizaje automático de Webb GI (eds)*. Springer. Boston.
- Johnson, R., & Wichern, D. (2007). *Applied Multivariante Statistical Analysis*. New Jersey: Pearson Education.
- Kassambara, A. (s.f.). *Practical Guide To Cluster Analysis in R*. STHDA.
- PERÚ Situación retail moderno. (2018). *Research, BBVA*, 26.
- Rooij, G., & Burg, G. (2017). *Clusterin Stores of Ratailers via Consumer Behavioer*. Rotterdam.

Anexo

Tabla 4

Segmentación vs Estado Civil según metodología RFM

Segmento	Estado Civil	Participación	Frecuencia	Monetario	Recencia
1	CASADO	17%	5%	4%	14
	CONVIVIENTE	6%	2%	1%	14
	DIVORCIADO	1%	0%	0%	13
	SOLTERO	36%	10%	8%	14
	VIUDO	1%	0%	0%	13
2	CASADO	5%	5%	7%	14
	CONVIVIENTE	2%	2%	3%	14
	DIVORCIADO	0%	0%	1%	15
	SOLTERO	13%	15%	17%	14
	VIUDO	0%	0%	0%	14
3	CASADO	3%	5%	6%	14
	CONVIVIENTE	1%	2%	2%	15
	DIVORCIADO	0%	1%	0%	13
	SOLTERO	6%	13%	13%	15
	VIUDO	0%	0%	0%	16
4	CASADO	1%	3%	3%	16
	CONVIVIENTE	0%	1%	1%	17
	DIVORCIADO	0%	0%	0%	16
	SOLTERO	2%	9%	8%	16
	VIUDO	0%	0%	0%	10
5	CASADO	1%	7%	8%	15
	CONVIVIENTE	0%	2%	2%	17
	DIVORCIADO	0%	1%	1%	15
	SOLTERO	2%	15%	14%	16
	VIUDO	0%	0%	0%	14

Fuente: Elaboración propia

Tabla 5*Segmentación vs Género según metodología RFM*

Segmento	Género	Participación	Frecuencia	Monetario	Recencia
1	F	41%	11%	9%	13
	M	21%	5%	5%	14
2	F	13%	15%	16%	14
	M	8%	8%	11%	15
3	F	6%	14%	13%	15
	M	4%	8%	8%	15
4	F	2%	9%	8%	16
	M	1%	5%	4%	16
5	F	2%	15%	14%	16
	M	2%	10%	11%	16

Fuente: Elaboración propia

Tabla 6*Segmentación vs Tipo de Actividad según metodología RFM*

Segmento	Género	Participación	Frecuencia	Monetario	Recencia
1	DEP	23%	6%	5%	14
	INDEP	22%	6%	5%	14
	INFOR	17%	5%	4%	13
2	DEP	7%	8%	10%	14
	INDEP	8%	9%	9%	14
	INFOR	6%	7%	7%	14
3	DEP	3%	7%	8%	14
	INDEP	4%	8%	8%	15
	INFOR	3%	6%	6%	15
4	DEP	1%	5%	5%	16
	INDEP	1%	5%	5%	16
	INFOR	1%	4%	3%	16
5	DEP	2%	11%	11%	16
	INDEP	2%	10%	11%	16
	INFOR	1%	5%	4%	16
	VIUDO	0%	0%	0%	16

Fuente: Elaboración propia