

UNIVERSIDAD PANAMERICANA

FACULTAD DE INGENIERÍA

Con estudios incorporados a la
Secretaría de Educación Pública

**SEGMENTACIÓN DE LOS PUNTOS DE VENTA DE
UNA EMPRESA EN EL SECTOR RETAIL EN
MÉXICO MEDIANTE UN ANÁLISIS DE
APRENDIZAJE MÁQUINA NO SUPERVISADO**

TESIS

QUE PARA OBTENER EL GRADO DE
**MAESTRÍA EN INGENIERÍA Y GESTIÓN DE LA
CADENA DE SUMINISTRO**

P R E S E N T A

DIEGO AGUIRRE DE LA GARZA

ASESOR:

EDGAR AVALOS GAUNA

Resumen

Liverpool es una empresa que ha logrado evolucionar a lo largo de los años. Fundada en 1936 por J. B. Ebrad, hoy cuenta con 123 almacenes distribuidos en las 32 entidades federativas de la república mexicana, logrando posicionarse como una de las cadenas de retail (venta minorista) mexicanas más grandes en nuestro país. La cadena busca constantemente ampliar su oferta de productos mediante la integración de nuevos proveedores que puedan satisfacer las necesidades de sus clientes con buen surtido, precio y calidad. Para esto, es importante contar con procesos adecuados y óptimos que ayuden a mantener la calidad del servicio al cliente. Sin embargo, la metodología actual de agrupación de tiendas carece de rigor sistemático, lo que nos deja un área de oportunidad.

Mediante un análisis máquina no supervisado de los datos, el presente trabajo busca un modelo para mejorar el perfilamiento de las tiendas, que esté más enfocado al cliente, elevando las ventas y las utilidades. Colocar de manera más oportuna la mercancía en cuanto a cantidad y variedad de modelos, reducirá las transferencias de mercancía entre las tiendas y optimizará la rotación de cada una, pues la mercancía no quedará estancada en los centros. Se debe tener en cuenta que durante el año no todos los centros tienen el mismo rendimiento por lo cual deberían de ser considerados de distinta manera a lo largo de los meses.

Actualmente no existe una metodología que tome en cuenta las diferentes variables asociadas a la venta de los artículos y esto afecta al inventario de las tiendas, pues recibirán una variedad y cantidad de productos no óptima. Mediante un análisis máquina no supervisado se podrá observar un mejor perfilamiento de los puntos de venta, pues el algoritmo presenta una recomendación al planeador. El criterio de cada planeador seguirá siendo importante, pues es experto en la materia. Con la implementación del algoritmo, existirá una base que el planeador podrá complementar con su conocimiento de la sección y así determinar si las agrupaciones son correctas o requieren una modificación.

La clusterización de los distintos puntos de venta apoyarán al planeador a colocar de una mejor manera la profundidad y el modelaje. Tener la mercancía adecuada en los centros permitirá que el índice de rotación tenga niveles sanos acordes a su inventario y venta a lo largo del año. Una metodología homologada permitirá que disminuya la cantidad de mercancía que se tenga que solicitar/rebajar, por consiguiente, las utilidades crecerán, pues se evitarán los costos logísticos innecesarios al no tener movimiento durante cierto periodo.

Dedicatoria y agradecimientos

Con este trabajo culminan varios años de arduo trabajo de estudio y me abre el camino para seguir adquiriendo conocimientos en cualquier ámbito al que me quiera enfocar. Estos años me permitieron crecer de manera personal y profesional.

Quiero agradecer, primero que nada, a mis padres, Gerardo Aguirre Cobo y Alicia Verónica de la Garza Gómez, por apoyarme todos los días a concluir este gran logro. Ellos me enseñaron que el trabajo duro y la dedicación hace fuerte a la persona y me guiaron en momentos de felicidad y de tristeza. Hoy les dedico este logro tan importante.

Agradezco el apoyo enorme que me ha brindado el Dr. Edgar Avalos Gauna que durante estos años me ha guiado en cada paso para poder culminar este proyecto.

Agradezco a todos los maestros que me orientaron durante tantos años, al compartir sus conocimientos, a tantos que durante meses me hicieron alguien mejor. Su manera de enseñar me hizo darme cuenta de que cada materia es importante para la formación de un profesional.

Agradezco a todos mis amigos quienes durante años me han apoyado a ser aquella persona que explota al máximo sus conocimientos y personalidades.

Agradezco a todos aquellos con los que compartí tantos años de estudio, risas, conocimiento y mucha felicidad.

Agradezco a la Universidad Panamericana que por muchos años me abrió las puertas convirtiéndose en un segundo hogar. Una institución que me dio todo lo necesario para aprender y ser una mejor persona.

Finalmente, agradezco a Liverpool por permitirme utilizar sus datos para la realización del presente trabajo.

“However difficult life may seem, there is always something you can do and succeed at.”

- Stephen Hawking

RESUMEN.....	2
DEDICATORIA	3
ÍNDICE DE TABLAS Y FIGURAS	6
GLOSARIO DE ABREVIATURAS, SIGLAS Y VARIABLES.....	8
INTRODUCCIÓN.....	9
PLANTEAMIENTO DEL PROBLEMA.....	11
OBJETIVO	11
CAPÍTULO 1. ANTECEDENTES Y ESTADO DE LA CUESTIÓN	13
DIVISIÓN DE LIVERPOOL.....	14
PANDEMIA: SARS-CoV-2.....	17
METODOLOGÍA ACTUAL PARA EL PERFILAMIENTO DE TIENDAS	18
METODOLOGÍA ACTUAL PARA EL PERFILAMIENTO DE TIENDAS	23
MARCO TEÓRICO PROPUESTO.....	23
<i>Ciencia de datos.....</i>	23
<i>Herramienta: Python</i>	24
ESTADO DEL ARTE	24
CAPÍTULO 2. METODOLOGÍA.....	27
LOGÍSTICA.....	27
Control e implementación de la logística global	28
<i>Inteligencia artificial</i>	29
<i>Machine learning</i>	30
Aprendizaje no supervisado.....	31
Reducción de dimensionalidad	32
Técnica de clustering	33
K-MEANS	35
Pruebas de validación.....	36
<i>Análisis exploratorio de los datos (EDA).....</i>	38
<i>Key Performance Indicator (KPI)</i>	40
<i>Datos.....</i>	41
CAPÍTULO 3. RESULTADOS OBTENIDOS.....	44
LIBRERÍAS	44
RESULTADOS EDA	45
<i>Matrices de correlación.....</i>	45
<i>Gráficos de violín.....</i>	47
<i>Mapas coropléticos</i>	50
PCA.....	52
CLUSTERIZACIÓN.....	55
<i>Dendrogramas</i>	55
<i>Método del codo</i>	58
<i>Visualización.....</i>	60
<i>Características del clúster</i>	62
CAPÍTULO 4. ANÁLISIS O DISCUSIÓN DE RESULTADOS	70
GRÁFICAS DE RADAR.....	71
COMPARACIÓN CONTRA METODOLOGÍA ACTUAL	74
CAPÍTULO 5. RESULTADOS Y CONCLUSIONES.....	81
CAPÍTULO 6. SUGERENCIAS PARA TRABAJOS FUTUROS.....	82
REFERENCIAS BIBLIOGRÁFICAS.....	83
ANEXO 01.....	87

ANEXO 02.....	87
ANEXO 03.....	87
ANEXO 04.....	87
GRÁFICAS DE CORRELACIÓN	87
GRÁFICAS DE VIOLÍN	89
MAPAS COROPLÉTICOS	90
PCA.....	92
DENDROGRAMA.....	94
MÉTODO DEL CODO	96
CARACTERÍSTICAS DEL CLUSTER	97
GRÁFICAS DE RADAR	101

Índice de tablas y figuras

Figura 01 - Número de tiendas por estado de la república mexicana.....	10
Figura 02 - Ubicación de tiendas Liverpool en la república mexicana	10
Figura 03 - Estructura de El Puerto de Liverpool.....	13
Figura 04 - Mapa de México con división de zonas de Liverpool.....	14
Figura 05 - Estructura de Liverpool.....	16
Figura 06 - Casos diarios en México por COVID-19.....	17
Tabla 01 - Metodología actual por planner	19
Tabla 02 -Tabla asignación planner Sandalia Liverpool Irapuato	20
Tabla 03 -Tabla asignación planner Tenis Liverpool Lindavista.....	21
Tabla 04 -Tabla asignación planner Tenis Liverpool COLIMA.....	22
Figura 07 - Representación de cardinalidad	35
Figura 08- Método del codo para número óptimo de k.....	37
Figura 09- Método de la silueta para determinar k óptimo	38
Figura 10 - Ejemplo dendrograma.....	40
Tabla 05 - Imagen tomada de la Base de datos	43
Tabla 06 - Librerías usadas.....	45
Figura 11 - Matriz de correlación Tenis 2021.....	46
Figura 12 - Matriz de correlación Sandalia 2021	46
Figura 13- Gráfica de violín Utilidad Tenis 2021.....	48
Figura 14- Gráfica de violín Utilidad Sandalia 2021.....	49
Figura 15 - Mapa coroplético promedio de utilidades por estado.....	50
Figura 16 - Mapa coroplético suma ventas por estado.....	51
Figura 17 - Vistas de PCA Tenis 2021 A) Vista isométrica, B) Vista superior, C) Vista lateral, D) Vista frontal	53
Figura 18 - Vistas de PCA Sandalia 2021 A) Vista isométrica, B) Vista superior, C) Vista lateral, D) Vista frontal	54
Figura 19- Dendrograma Tenis 2021	56
Figura 20- Dendrograma Sandalia 2021	57
Figura 21- Método del codo Tenis 2021.....	58
Figura 22- Método del codo Sandalia 2021	59
Figura 23- Visualización t-SNE, A) Tenis 2019, B) Tenis 2021, C) Sandalia 2019, D) Sandalia 2021	61
Figura 24- Cardinalidad y magnitud clústeres Tenis 2021, A) Cardinalidad, B) Magnitud, C) Cardinalidad vs. magnitud	63
Figura 25- Cardinalidad y magnitud clústeres Sandalia 2021, A) Cardinalidad, B) Magnitud, C) Cardinalidad vs. magnitud	64
Figura 26- Distribuciones de característica por clúster Tenis 2021, A) Gráfica de cajas y bigotes Mes, B) Gráfica de cajas y bigotes Latitud, C) Gráfica de cajas y bigotes Longitud, D) Gráfica de cajas y bigotes mes Ventas, E) Gráfica de cajas y bigotes Utilidad, F) Gráfica de	

cajas y bigotes Descuento, G) Gráfica de cajas y bigotes MSI.....	68
Figura 27- Distribuciones de característica por clúster Sandalia 2021, A) Gráfica de cajas y bigotes Mes, B) Gráfica de cajas y bigotes Latitud, C) Gráfica de cajas y bigotes Longitud, D) Gráfica de cajas y bigotes mes Ventas, E) Gráfica de cajas y bigotes Utilidad, F) Gráfica de cajas y bigotes Descuento, G) Gráfica de cajas y bigotes MSI.....	69
Figura 28- Fases para implementar la inteligencia artificial.....	70
Figura 29- Radar Tennis 2021.....	73
Figura 30- Radar Sandalia 2021	73
Tabla 07- Resultado escenario 1 Tennis.....	75
Tabla 08 - Resultado escenario 2 Tennis.....	76
Tabla 09 - Resultado escenario 3 Tennis.....	77
Tabla 10 - Resultado escenario 1 Sandalia	78
Tabla 11- Resultado escenario 2 Sandalia	79
Tabla 12 - Resultado escenario 3 Sandalia	80
Figura 31- Matriz de correlación Tennis 2019.....	87
Figura 32- Matriz de correlación Sandalia 2019	88
Figura 33- Gráfica de violín Utilidad Tennis 2019.....	89
Figura 34- Gráfica de violín Utilidad Sandalia 2019.....	89
Figura 35- Mapa coroplético suma utilidades por estado.....	90
Figura 36- Mapa coroplético promedio de total descuento por estado	91
Figura 37- Vistas de PCA Tennis 2019 A) Vista isométrica, B) Vista superior, C) Vista lateral, D) Vista frontal	92
Figura 38- Vistas de PCA Sandalia 2019, A) Vista isométrica, B) Vista superior, C) Vista lateral, D) Vista frontal	93
Figura 39- Dendrograma Tennis 2019	94
Figura 40- Dendrograma Sandalia 2019.....	95
Figura 41- Método del codo Tennis 2019.....	96
Figura 42- Método del codo Sandalia 2019	96
Figura 43- Cardinalidad y magnitud clústeres Tennis 2019, A) Cardinalidad, B) Magnitud, C) Cardinalidad vs. magnitud	97
Figura 44- Cardinalidad y magnitud clústeres Sandalia 2019, A) Cardinalidad, B) Magnitud, C) Cardinalidad vs. magnitud	98
Figura 45- Distribuciones de característica por clúster Tennis 2019, A) Gráfica de cajas y bigotes Mes, B) Gráfica de cajas y bigotes Latitud, C) Gráfica de cajas y bigotes Longitud, D) Gráfica de cajas y bigotes mes Ventas, E) Gráfica de cajas y bigotes Utilidad, F) Gráfica de cajas y bigotes Descuento, G) Gráfica de cajas y bigotes MSI.....	99
Figura 46- Distribuciones de característica por clúster Sandalia 2019, A) Gráfica de cajas y bigotes Mes, B) Gráfica de cajas y bigotes Latitud, C) Gráfica de cajas y bigotes Longitud, D) Gráfica de cajas y bigotes mes Ventas, E) Gráfica de cajas y bigotes Utilidad, F) Gráfica de cajas y bigotes Descuento, G) Gráfica de cajas y bigotes MSI.....	100
Figura 47- Radar Tennis 2019.....	101
Figura 48- Radar Sandalia 2019	101

Glosario de abreviaturas, siglas y variables

Símbolo	Definición
D_b	Número de decisiones
D_r	Es la suma de las distancias entre los puntos de un grupo específico
N_b	Número de canastas
T	Cantidad de variables
P	Cantidad de rasgos funcionales
V_b	Clúster independiente de X
W	Punto de inflexión de la curva
k	Número de clústeres
n_r	Número de puntos en el grupo r
r	Rotación del inventario anual
s	Suma de ventas anuales
p	Promedio del inventario de enero a enero
x_{it}	Número de variables en el punto i
x_{jt}	Número de variables en el punto j
x_{ab}	Número de rasgos en el punto a
x_{cb}	Número de rasgos en el punto c

Introducción

Liverpool es una subsidiaria del Puerto de Liverpool, S.A.B. de C.V., el grupo de tiendas departamentales más grande de México. Sus inicios se remontan al año de 1847, cuando Jean Baptiste Ebrard instala en el centro de la Ciudad de México un cajón dedicado a la venta de telas finas. Posteriormente, en 1872, el señor Ebrard comenzó a importar mercancía embarcada desde el puerto de Liverpool, Inglaterra. Liverpool abre las puertas de su primera tienda en el centro de la Ciudad de México en el año de 1936. Hoy cuenta con almacenes en toda la república mexicana y ha logrado posicionarse en el mercado como una compañía reconocida y respetada por sus competidores. Liverpool cuenta actualmente con 123 tiendas en todo México. En la *Figura 01* se puede observar la distribución de tiendas a lo largo de los 32 estados de la república mexicana. Mientras que en la *Figura 02* se puede observar la ubicación de las distintas tiendas de Liverpool alrededor de la república mexicana.

Liverpool se divide en 9 direcciones que a su vez estas se subdividen para satisfacer a cada tipo de cliente. Dichas direcciones pueden o no tener divisiones por mundo para después ser divididas en secciones especializadas para cada tipo de producto (por ejemplo, en la dirección 4 de infantiles está el mundo de bebés donde se encuentra la sección de Bebés Moda). Las secciones buscan constantemente nuevos proveedores que tengan productos diferentes e innovadores para ofrecerle al cliente una tienda con mejor surtido, precio y calidad. Liverpool busca servir al cliente todos los días, en todo lugar y toda la vida, al ofrecer la opción más atractiva en servicio, surtido, calidad y valor del mercado mexicano.

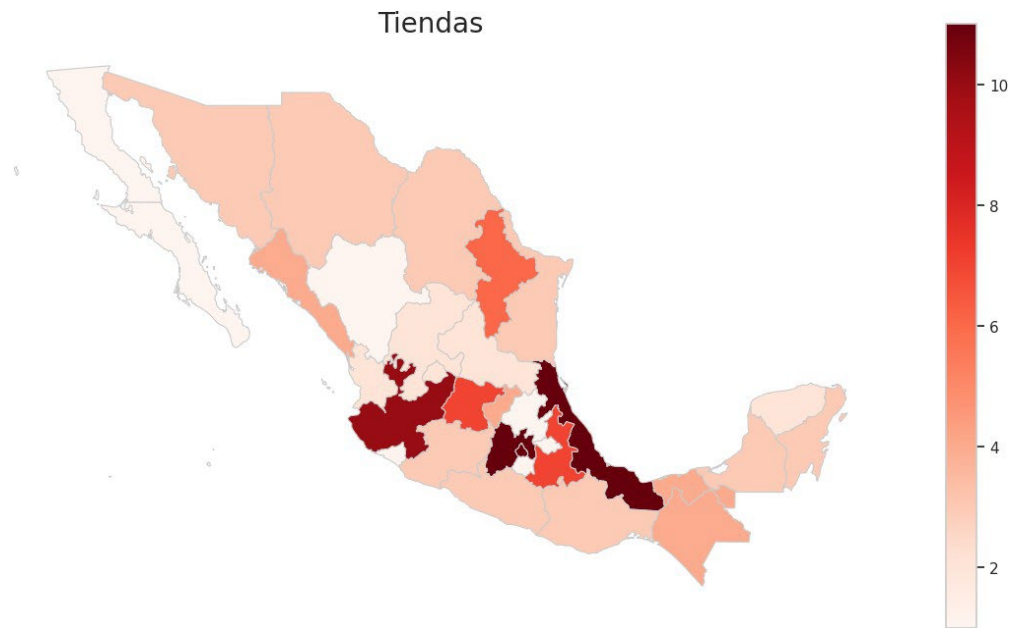


Figura 01 - Número de tiendas por estado de la república mexicana

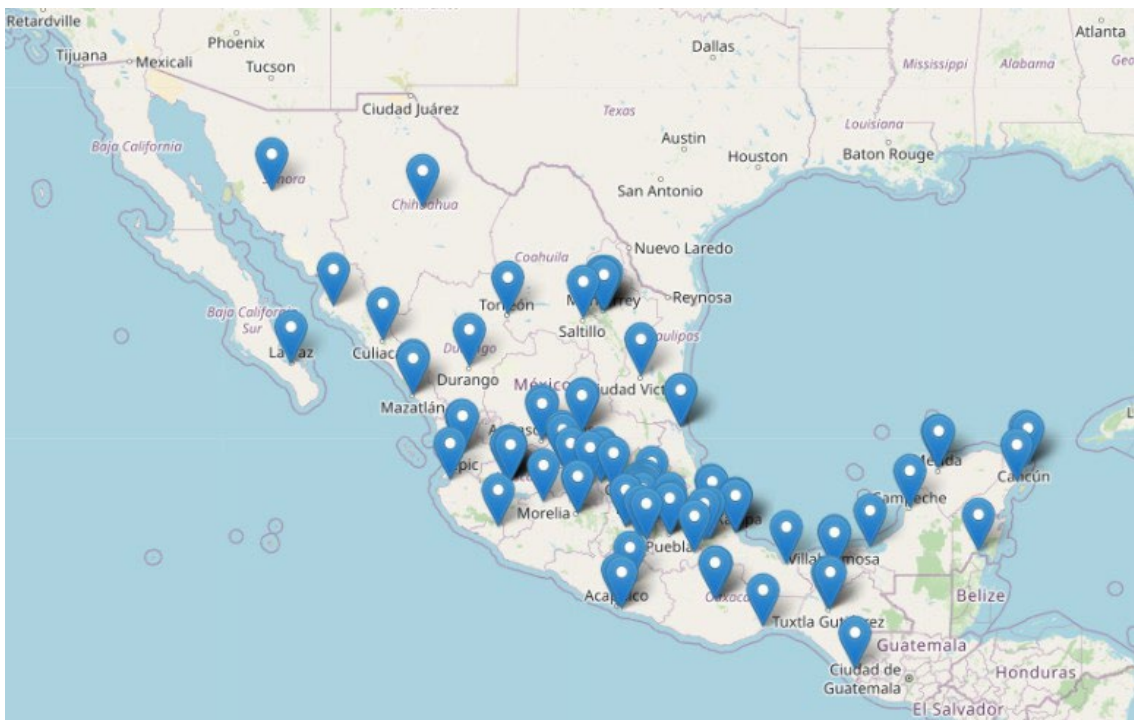


Figura 02 - Ubicación de tiendas Liverpool en la república mexicana

Planteamiento del problema

El problema que se busca abordar considera los siguientes 3 aspectos:

- La actual metodología de agrupación de las tiendas carece de rigor sistemático, además de no ser precisa. Esto hace que las utilidades de la mercancía se vean afectadas.
- Las distintas tiendas de la cadena llegan a tener un sobre inventario o desabasto de mercancía, lo que afecta a la rotación anual por la clasificación en la cual se le coloca con la metodología actual.
- Las distintas tiendas de la cadena solicitan transferencias¹, para poder ingresar o poder deshacerse de la mercancía. Esto genera una redistribución de la mercancía, lo cual eleva los costos logísticos.

Estas situaciones no son ajenas entre sí, y esto nos lleva a la pregunta ¿Es posible reducir los costos logísticos ocasionados por transferencias teniendo un perfilamiento de tienda más adecuado?

Objetivo

El objetivo del presente trabajo es implementar un análisis de aprendizaje máquina no supervisado con el fin de crear un mejor perfilamiento de los puntos de venta, comparado con la técnica actualmente vigente. En consecuencia, esto derivará en un mejor manejo de las transferencias solicitadas por las tiendas, por ende, una mejor asignación de producto, derivando en la reducción de costos logísticos.

Para ello se plantea el siguiente plan de trabajo:

- Planteamiento del problema a resolver.
- Selección de datos secundarios disponibles para el análisis.
- Análisis exploratorio de la información.
- Desarrollo de un análisis de aprendizaje máquina no supervisado que permita el incremento de las ventas de los distintos puntos de venta de la cadena minorista, el cual considerará:
 - Mejorar el actual perfilamiento de los puntos de venta para los distintos casos.
 - Proponer un modelo que reduzca la mercancía que se tiene que discontinuar.

¹ una transferencia es aquel movimiento de mercancía entre una tienda y otra. Esto se realiza con el objetivo de nivelar el inventario.

El trabajo está dividido de la siguiente manera:

Capítulo 1 - Se toman conceptos básicos sobre la historia y estructura de Liverpool, se menciona la metodología actual que se maneja dentro de Liverpool y se mencionan casos prácticos sobre la resolución del problema en cuestión.

Capítulo 2 - Se abordan los conceptos básicos para entender el problema. Se mencionan las diferentes técnicas que existen en la actualidad y se identifica la metodología que se utilizará, explicando las razones para la selección de dicha metodología y por qué no se seleccionó otra metodología.

Capítulo 3 - Se muestra un análisis exploratorio de los datos a través de cálculos y visualizaciones sobre la base de datos obtenida.

Capítulo 4 - Se muestra el análisis de los gráficos presentados en el capítulo 3. En este capítulo se presenta el análisis y la discusión de la metodología.

Capítulo 5 - En este capítulo se discuten los resultados obtenidos y su significado en el proceso actual de perfilamiento de tiendas. También se hace un análisis del nivel de implementación que podrá tener esta herramienta.

Capítulo 6 - Se presentan posibles aplicaciones de este trabajo para que los lectores conozcan la manera en la que pueden replicarlo.

Capítulo 1. Antecedentes y estado de la cuestión

El Puerto de Liverpool, S.A.B., de C.V., es un grupo de tiendas departamentales en México que cuenta además con una financiera. Sus inicios se remontan a 1847 en el centro de la Ciudad de México, en un cajón dedicado a la venta de telas finas y posteriormente, en 1872, el fundador comenzó a importar mercancía embarcada desde el puerto de Liverpool, Inglaterra hacia México. Liverpool abre las puertas de su primera tienda en el centro de la Ciudad de México en 1936 y en 1965 empieza a cotizar en la Bolsa Mexicana de Valores (BMV). En 1988 El Puerto de Liverpool adquiere la empresa Fábricas de Francia, estas tiendas posteriormente se convierten en tiendas Liverpool y más tarde adquiere algunas tiendas Suburbia. Como se puede observar en la Figura 03, El Puerto de Liverpool tiene diferentes negocios, actualmente son 123 tiendas Liverpool, 131 boutiques, 166 tiendas Suburbia, 27 Galerías, 2 centros de distribución y más de 40 bodegas logísticas.

Para el caso particular de Liverpool, la ideología del corporativo indica que “La tienda busca servir al cliente todos los días, en todo lugar y toda la vida”. Busca ser la opción más atractiva en servicio, surtido y valor del mercado mexicano. Hoy, con almacenes en toda la república mexicana, Liverpool ha podido posicionarse en el mercado como una compañía reconocida y respetada por sus competidores. Liverpool sigue creciendo y tiene proyectos muy ambiciosos para el futuro. Actualmente su proyecto más ambicioso es “La Plataforma Logística Arco Norte”, que es un habilitador que permitirá acercar más productos a sus clientes en menor tiempo, mejorando la experiencia de compra. Este proyecto permitirá tener una mejor distribución y control de lo que se ofrece actualmente.

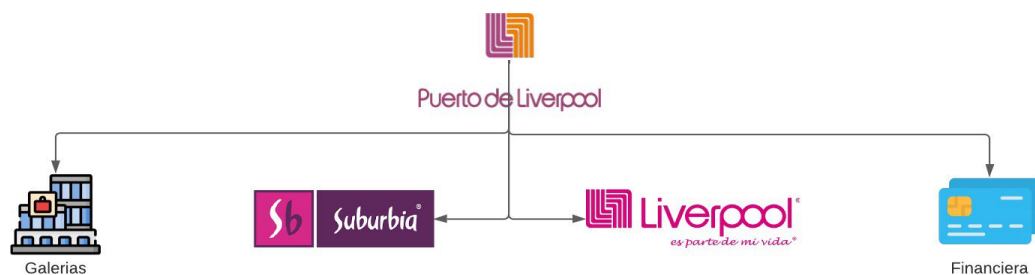


Figura 03 - Estructura de El Puerto de Liverpool

División de Liverpool

Liverpool hace una división geográfica de la república mexicana; los estados de la república son divididos en zonas para tener control de las tiendas que se encuentran en los distintos estados. Para las 123 tiendas actualmente abiertas existen 9 zonas, el beneficio que da esta división es que se puede llevar un mejor control sobre el resultado que se obtiene durante el año. Otro de los beneficios de esta división es que las transferencias entre zonas son menos complicadas ya que no tienen que pasar por el centro de distribución.



Figura 04 - Mapa de México con división de zonas de Liverpool

En la *Figura 05*, se puede observar cómo está dividido Liverpool. En cada tienda Liverpool hay dos divisiones principales para identificar los tipos de productos que se venden:

- Hardline: es toda aquella mercancía que no se puede llevar en una bolsa de Liverpool, es decir muebles, sillas, colchones, lavadoras, caminadoras, etc.
- Softline: es toda aquella mercancía que se puede llevar dentro de una bolsa de Liverpool, es decir ropa, tenis, lociones, secadoras, etc.

Estas 2 áreas se dividen en 9 direcciones para tener un mejor control de los tipos de productos que se ofrecen. Asimismo, las distintas direcciones están definidas en “mundos” y estos a su vez en diferentes secciones. A continuación, se presentan todas las direcciones y el tipo de producto que ofrecen:

- Hardline
 - Dirección 1. Muebles
 - Sillones, escritorios, mesas.
 - Dirección 3. Hogar y diversos
 - Estufas, secadoras, lavadoras, microondas.
 - Dirección 6. Multimedia

- Televisiones, celulares, videojuegos, audífonos, bocinas.
 - Dirección 8. Restaurante y Experiencia Gourmet
 - Comida, dulces, bebidas, alcohol.
 - Dirección 9. Deportes
 - Caminadoras, mancuernas, elípticas.
- Softline
 - Dirección 2. Caballeros
 - Ropa, zapatos, accesorios.
 - Dirección 4. Infantiles
 - Ropa bebe, niño y niña, juguetes, zapatería.
 - Dirección 5. Damas
 - Ropa, zapatos, accesorios.
 - Dirección 7. Cosméticos
 - Perfumes, productos de belleza, secadoras de pelo.

En el caso de la dirección de infantiles existen 3 mundos:

- Mundo Niños
- Mundo Niñas
- Mundo Bebés

Cada uno de estos mundos se divide en diferentes secciones. Un ejemplo de sección es la zapatería infantil, la cual se conforma por Tenis niño y niña, zapatos niña y zapatos niño. Para entenderlo mejor, la *Figura 05* muestra los distintos niveles de desagregación.

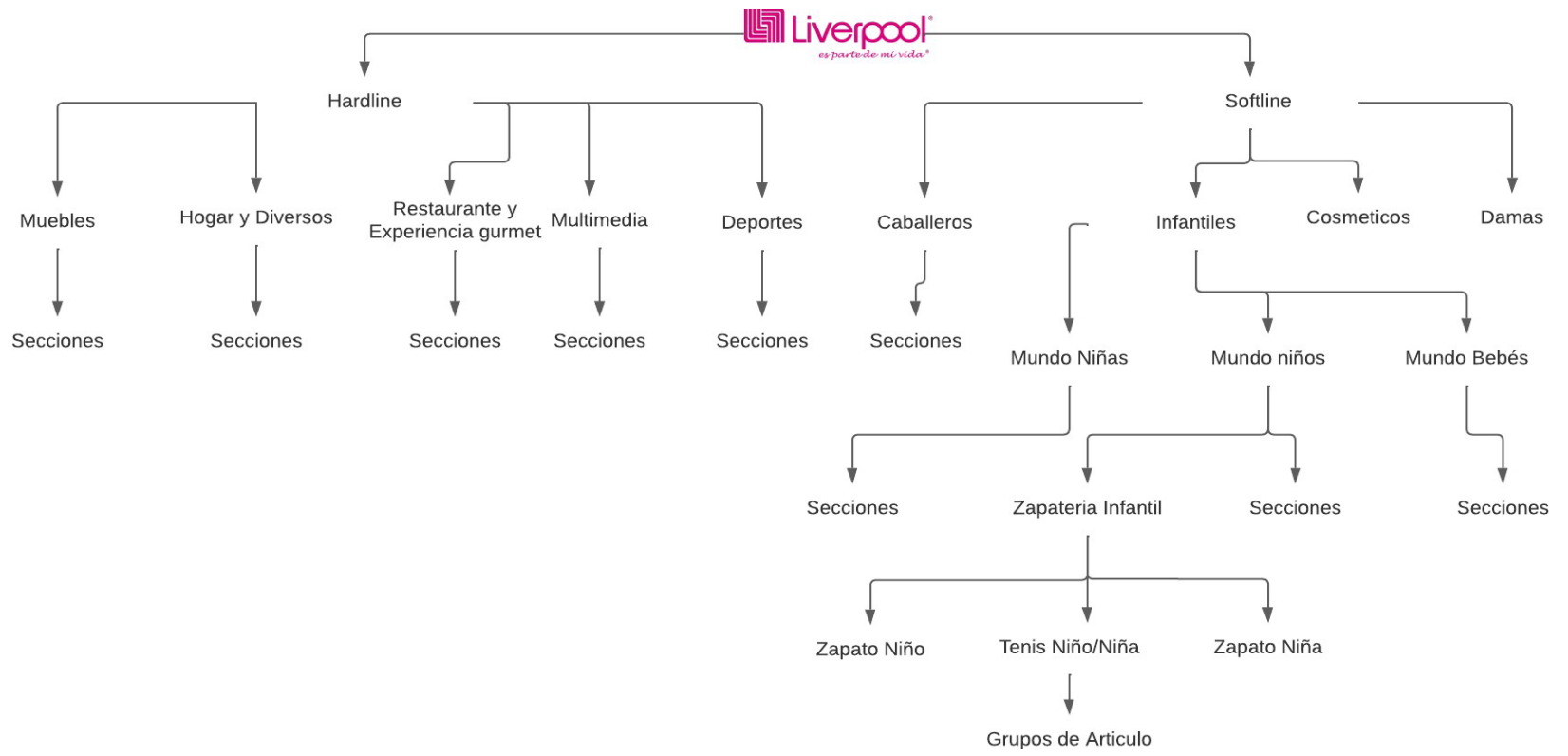


Figura 05 - Estructura de Liverpool

La estructura de las secciones es diferente para cada una y principalmente están conformadas por grupos de artículos. En el caso de la zapatería infantil, la sección de Tenis niño y niña tiene grupos de artículos como: Tenis, Sandalia, calcetas y accesorios. La zapatería tiene un catálogo muy extenso, cuenta con alrededor de 90 marcas que se dividen en tres categorías de mercancía: nacional, importación y marcas propias. Las marcas propias son marcas de Liverpool que se producen en el extranjero para vender el producto en las tiendas, estas incluyen Piquenique, That 's It y Mon Caramel, entre otras.

Pandemia: SARS-CoV-2

Una de las razones por las cuales hubo un gran avance en temas de tecnología fue la pandemia de SARS CoV 2 o mejor conocida como Covid-19. La pandemia comenzó en Hubei, China en diciembre de 2019, a México llegó el 27 de febrero de 2020 y fue hasta mayo de 2023 que la Organización Mundial de la Salud (OMS) da por concluida la pandemia (WHO, 2020). Según datos de la OMS, a nivel mundial existieron 770 millones de casos de los cuales casi 7 millones resultaron en muerte. En México, hubo casi 7.7 millones de casos y 334 mil muertes. En la *Figura 06* se puede observar la gráfica de casos diarios en México.

Sin embargo, la pandemia no fue del todo negativa, ya que debido al cierre de los establecimientos y a que la población se encontraba en aislamiento, se logró que los canales de venta virtuales crecieran. En el caso de Liverpool, las tiendas tuvieron que cerrar sus puertas durante algunos meses. Con el paso del tiempo Liverpool se enfrentó a diferentes retos para poder satisfacer el mercado en línea como lo fue mejorar su servicio al cliente y los tiempos de entrega. Estas mejoras ocasionaron un alza y un avance en el canal de ventas por internet.

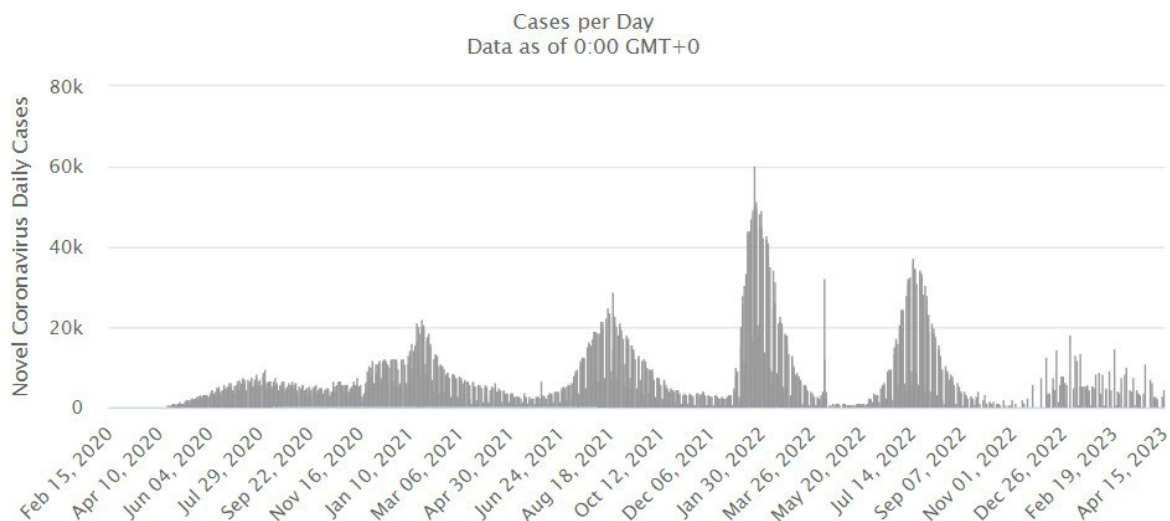


Figura 06 - Casos diarios en México por COVID-19

Metodología actual para el perfilamiento de tiendas

La manera en la que hoy se elabora un clúster dentro de Liverpool es la siguiente:

1. Se obtienen las ventas de los sistemas de Liverpool a nivel sección-marca-centro dentro de un periodo específico (ej. ventas de la marca Nike de la sección 402 durante el segundo semestre de 2019).
2. Se descarga la información en una hoja de Excel.
3. Se filtran y se borran todos los centros a los que no va la marca.
4. Se acomoda de mayor a menor el ratio que se prefiera (ventas o margen).
5. Se realiza un análisis subjetivo de participación con calificación "A,B,C,D" (las letras A suelen ser las mejores tiendas, mientras que las C o D suelen ser las peores).

La manera de hacer un clúster dentro del equipo de Liverpool es ambigua y carece de rigor sistemático. La forma en la que los hace un planeador de la demanda puede ser diferente a la de los demás, lo cual es un tema variable. Incluso las categorías utilizadas no son consistentes. Algunos utilizan categorías adicionales en donde entre más letras "A" tenga una tienda en una calificación, significa que son mejores. Por ejemplo, "AAA" o "AA".

Para poder ejemplificar la situación actual por la que pasa Liverpool a la hora de realizar clústeres, se presentó a 4 planeadores de la dirección de infantiles la siguiente información, la cual comprende un documento que con datos sobre cuánto vendieron, cuánta utilidad presentaron, cuál fue la cantidad de descuentos y de meses sin intereses en pesos mexicanos por cada tienda. También se les presentó la latitud y longitud del centro de ventas. Lo anterior fue para obtener información referente a dos grupos de artículos (Tenis y Sandalia) para el año 2021. El criterio de cada planeador se puede observar en la Tabla 01. Las Tablas 02 a 04 muestran los resultados de los planeadores para una tienda.

Planeador	Rango utilizado	Metodología	Porcentajes
1	AAA-D	Acomodó de mayor a menor las utilidades y sacó la participación acumulada.	AAA-A- 0-25% B- Siguietes 25 tiendas a cluster A C- Siguietes 30 tiendas al cluster B D- Resto de las tiendas
2	AA-D	Análisis Pareto donde se utilizó la participación de las utilidades.	AA- 0-5% A- 5-20% B- 20-50% C- 50-95% D- 95-100%
3	AA-D	Pareto basándose en las utilidades y las ventas	AA- 0-20% A- 20-40% B- 40-61% C- 61-80% D- 80-100%
4	AA-D	Pareto basándose en las utilidades y las ventas	AA- 0-20% A- 20-40% B- 40-61% C- 61-80% D- 80-100%

Tabla 01 - Metodología actual por planner

Datos				Planners (Datos 2021)			
Centro #	Centro	Zona#	Mes	1	2	3	4
0047	Liverpool IRAPUATO	7	1	C	C	D	D
0047	Liverpool IRAPUATO	7	2	C	C	D	D
0047	Liverpool IRAPUATO	7	3	C	C	D	D
0047	Liverpool IRAPUATO	7	4	C	C	D	D
0047	Liverpool IRAPUATO	7	5	C	C	D	D
0047	Liverpool IRAPUATO	7	6	C	C	D	D
0047	Liverpool IRAPUATO	7	7	C	C	D	D
0047	Liverpool IRAPUATO	7	8	C	C	D	D
0047	Liverpool IRAPUATO	7	9	C	C	D	D
0047	Liverpool IRAPUATO	7	10	C	C	D	D
0047	Liverpool IRAPUATO	7	11	C	C	D	D
0047	Liverpool IRAPUATO	7	12	C	C	D	D

Tabla 02 -Tabla asignación planner Sandalia Liverpool Irapuato

Datos				Planners (Datos 2021)			
Centro #	Centro	Zona#	Mes	1	2	3	4
0083	Liverpool LINDAVISTA	2	1	A	A	AA	AA
0083	Liverpool LINDAVISTA	2	2	A	A	AA	AA
0083	Liverpool LINDAVISTA	2	3	A	A	AA	AA
0083	Liverpool LINDAVISTA	2	4	A	A	AA	AA
0083	Liverpool LINDAVISTA	2	5	A	A	AA	AA
0083	Liverpool LINDAVISTA	2	6	A	A	AA	AA
0083	Liverpool LINDAVISTA	2	7	A	A	AA	AA
0083	Liverpool LINDAVISTA	2	8	A	A	AA	AA
0083	Liverpool LINDAVISTA	2	9	A	A	AA	AA
0083	Liverpool LINDAVISTA	2	10	A	A	AA	AA
0083	Liverpool LINDAVISTA	2	11	A	A	AA	AA
0083	Liverpool LINDAVISTA	2	12	A	A	AA	AA

Tabla 03 -Tabla asignación planner Tenis Liverpool Lindavista

Datos				Planners (Datos 2021)			
Centro#	Centro	Zona#	Mes	1	2	3	4
0084	Liverpool COLIMA	4	1	C	C	D	D
0084	Liverpool COLIMA	4	2	C	C	D	D
0084	Liverpool COLIMA	4	3	C	C	D	D
0084	Liverpool COLIMA	4	4	C	C	D	D
0084	Liverpool COLIMA	4	5	C	C	D	D
0084	Liverpool COLIMA	4	6	C	C	D	D
0084	Liverpool COLIMA	4	7	C	C	D	D
0084	Liverpool COLIMA	4	8	C	C	D	D
0084	Liverpool COLIMA	4	9	C	C	D	D
0084	Liverpool COLIMA	4	10	C	C	D	D
0084	Liverpool COLIMA	4	11	C	C	D	D
0084	Liverpool COLIMA	4	12	C	C	D	D

Tabla 04 -Tabla asignación planner Tenis Liverpool COLIMA

Como se puede observar en la Tabla 01, los planeadores utilizaron la misma técnica, pero utilizaron sus propios criterios para determinar los clústeres. Se puede observar que cada uno de ellos tiene un criterio, aunque el criterio entre el planeador #3 y #4 sean casi iguales. En la Tabla 02 a 04 se puede observar que no todos los planeadores le colocan la misma clasificación a cada tienda. Cada planeador tiene un análisis diferente de la sección. Esto deriva en que la asignación de tiendas a algún grupo sea de forma subjetiva. A pesar de ello, la división que se presenta parece ser un criterio importante referente a la toma de decisiones, gracias a ello las asignaciones de producto son más específicas.

También, cabe mencionar que el clúster asignado por un planeador a una tienda es para todo el año, es decir que no se hace una diferencia por mes. Esto puede afectar al resultado de la tienda ya que no todos los meses se tiene el mismo aforo de clientes en un centro. Las tiendas categorizadas entre "AAA" a "B" cuentan con más inventario ya que se espera que tengan ventas exitosas a comparación de las tiendas categorizadas como "C" o "D", las cuales se espera que tengan una venta menor ya. Existen temporadas promocionales importantes donde los centros tienen la oportunidad de vender más mercancía. Las tiendas

al ser categorizadas como inferior por ejemplo “C” o “D” puede que no llegue a cierto potencial, ya que cuenta con inventario limitado dado el cluster colocado. Esto llega a perjudicar el resultado de la tienda porque se le negó venta al cliente por no tener diferente modelaje y una mayor cantidad de producto. Como en el caso contrario, una tienda categorizada como “A” o “B” puede que tenga inventario de sobra y perjudique al resultado de la tienda ya que tiene inventario con poco movimiento. Las tiendas tienen como objetivo de medición varios KPI, los cuales son apoyo para observar el rendimiento de cada una, estos incluyen ventas, margen y rotación. Los planners llegan a realizar una actualización de las categorías como mínimo 1 vez al año, depende de cada uno estar revisando el rendimiento de las tiendas para considerar una recategorización.

Metodología de tiendas del sector retail.

Actualmente, tiendas similares a Liverpool utilizan metodologías comparables para el perfilamiento y clusterización de tiendas. Estas metodologías se basan principalmente en las ventas de cada tienda para agruparlas en distintos clústeres. Además, para la clusterización de las tiendas se consideran otros factores, como la ubicación geográfica y el tipo de productos ofrecidos. También analizan aspectos como los precios de la mercancía y los metros cuadrados de cada departamento.

La distribución de la mercancía puede variar entre las diferentes tiendas del sector retail, ya que cada una puede tener en cuenta diversos aspectos específicos de sus puntos de venta y tipo de negocio. Esta distribución única permite a cada empresa diferenciarse de sus competidores.

Marco teórico propuesto

Ciencia de datos

Cao (2017) menciona que “El concepto de la ciencia de datos fue originalmente propuesto dentro de las estadísticas y la comunidad matemática, su principal función es el análisis de los datos. La ciencia de datos es un nuevo campo transdisciplinario que se basa en sintetizar una serie de disciplinas y cuerpos de conocimiento de estadística, informática, computación, gestión de la comunicación y sociología” (p. 60). La ciencia de datos es la exploración científica de los datos para extraer valores y realizar recomendaciones con estadísticas y modelos matemáticos con el objetivo de apoyar la toma de decisiones. Esta ciencia es la que contiene inteligencia artificial, aprendizaje máquina y minería de datos que conforman parte de los grandes datos.

Para XU et. al. (2021) la ciencia de datos es la metodología relacionada con la realización de la cadena de valor; es la ciencia que influye a otras ciencias y disciplinas que, con el paso del tiempo, permite el surgimiento de nuevas metodologías, teorías o modelos que moldean a las ciencias. Es una ciencia que utiliza modelos para convertir los datos en información, que a su vez se convierte en conocimiento y nos ayuda a tomar decisiones informadas. La ciencia de datos tiene aplicación con diferentes disciplinas.

Esto ha llevado a que la ciencia de datos se haya convertido en una disciplina más completa ya que complementa a estas distintas ciencias. Debido a los aportes de las demás ramas, la ciencia de datos ha mejorado la manera en la que se analizan los datos. Los avances han propiciado relevancia a otras herramientas que están tomando mucha fuerza y relevancia como es la inteligencia artificial. Gracias al apoyo de todas estas herramientas la industria y la economía pueden avanzar de manera significativa.

Herramienta: Python

La herramienta que se utilizó para realizar el algoritmo que se presenta a continuación es Python. Para ello se ocupó una máquina virtual desde el Software as a Service (SaaS) Google Colab. Este intérprete trabaja en línea lo que permite una mayor interconectividad con distintas herramientas, así como una mayor facilidad para compartir el contenido creado ya sea con colaboradores, estudiantes, maestros e internautas, entre otros.

Python es un lenguaje de programación de alto nivel creado en 1991. El énfasis de su filosofía es la legibilidad del código. Es descrito como un código amigable que tiene una gran comprensión de distintas librerías (Van Rossum, 2007). Las librerías que se seleccionaron para el análisis de los códigos y la presentación de los gráficos son: Pandas, Seaborn, Matplotlib.pyplot, Numpy, Sklearn, Plotly.express, geopandas y Pandas Profiling. Algunos de los beneficios que ofrece la herramienta (AWS, 2023), son los siguientes:

1. Es una herramienta gratuita.
2. Es una herramienta eficiente, es fácil de entender.
3. Es una herramienta multiplataforma que se puede utilizar en los sistemas operativos Windows, macOS, Linux y Unix.
4. Cuenta con una biblioteca de códigos reutilizables.
5. Es un lenguaje orientado a objetos.
6. Existe una comunidad muy grande de programadores, que al estar activa, permite a los usuarios solucionar sus dudas y problemas en poco tiempo.

Estado del arte

En este apartado se mencionan tres diferentes documentos que presentan problemas relacionados con el tema que se trata en el presente trabajo. Se analiza la metodología utilizada, así como las consideraciones que se implementaron y finalmente, los resultados que alcanzaron.

Aragón et al. (2023), presentaron un trabajo para la mejora de la segmentación de clientes en la compra de productos electrónicos mediante métodos de aprendizaje no supervisado como el clustering, enfocado al sector del retail. El propósito que presentan es que al realizar una segmentación adecuada de la clientela se pueden tener mejores experiencias de compra, ya que se hace una mercadotecnia más apropiada para cada nicho de mercado, pues los patrones de compra y el consumidor han cambiado sus hábitos de compra a raíz de la pandemia.

La información recabada en la investigación es sobre una empresa peruana llamada Saga Falabella que vende electrónicos. Los datos utilizados corresponden a los productos electrónicos vendidos dentro del periodo de noviembre 2022 a enero 2023. Los datos

recopilados fueron obtenidos de un data warehouse (depósito de datos) de la compañía. Las técnicas que se van a implementar para determinar la segmentación incluyen el clúster jerárquico, k-means y medios. Utilizaron el método RFM (reciente, frecuencia y monto) para poder juntar y homologar los distintos archivos y posteriormente hicieron una limpieza de aquellas variables que no eran relevantes para la investigación.

El clustering, se realizó mediante el lenguaje de programación Python donde implementaron el modelo k-means debido a su capacidad de procesamiento de alto volumen de datos. Utilizaron sobre los datos métodos para determinar el valor óptimo de k, como es el método del codo. Llegaron a la conclusión de que se tenía que segmentar en 5 diferentes grupos, los cuales se asemejan a un análisis 80-20, es decir, el 20% de tu clientela genera el 80% de tus ingresos. Concluyeron que la evaluación de los medios de clusterización fueron los adecuados para la segmentación de clientes y poder tomar mejores decisiones de marketing para cada segmento.

Zamar et al. (2017) presentaron el problema de recolección de pacas después de las operaciones de cosecha de cultivos agrícolas. Los autores definieron la secuencia en la que se recogen los fardos que se encuentran en todo el campo. Mediante el algoritmo k-means restringido se busca minimizar el tiempo de viaje y reducir el consumo de combustible.

Actualmente en la industria de la agricultura se pueden recolectar datos en tiempo real de las operaciones realizadas en el campo. Las principales operaciones sobre el terreno requieren de la coordinación planificada de diversos equipos agrícolas. Las tecnologías pueden proveer información precisa para la agricultura. Se determina que el problema incluye identificar la ubicación de los diferentes sitios de almacenamiento y rutas de recolección.

El problema está relacionado con el enrutamiento de vehículos, la optimización combinatoria de programación y es una generalización del problema del vendedor ambulante (TSP). Se presenta un algoritmo heurístico para resolver el problema después de las operaciones de recolección.

El caso se divide en 2 metodologías. La primera identifica los sitios de los almacenes. La segunda involucra identificar las rutas eficientes. El primer problema es aproximado por el análisis de clusters, el objetivo es la repartición de fardos en k clusters en donde el fardo pertenezca al clúster con el promedio más cercano, como resultado habrá una repartición de los fardos. Se utiliza un algoritmo k-means para el análisis. Dados X puntos, k-means busca la partición de estos n puntos en K que sean menores a la cantidad de puntos. Utiliza funciones que calculan la mínima distancia euclidiana entre los puntos. El resultado fue identificar soluciones casi óptimas mediante el grado de aislamiento de los fardos para ver cuántas veces es seleccionado como el vecino más cercano. Para cada fardo seleccionado se resuelve el problema del vendedor ambulante que visita todos los fardos k^{-1} , la ruta más corta es seleccionada y eliminada del set de nodos. El proceso de repetir hasta que todos los fardos sean seleccionados es un problema heurístico llamado MMROA (min-min min-max route optimization algorithm) ya que cada iteración selecciona el fardo más aislado o menos aislado, minimizando la ruta, lo que deriva en minimizar la ruta que se elige entre los fardos. Para el problema las restricciones se enfocan en el número de clusters (k), la ubicación de las calles y el número exacto de fardos.

Con base en el análisis de la suma de cuadrados se eligió dividir los fardos en 5 clusters, que se identificaron de acuerdo con las 5 restricciones de centros que se determinaron.

Holý et al. (2017) presentaron el problema de la categorización de los productos en la industria retail, que es esencial para el proceso de toma de decisiones. Una de las prácticas más comunes que sigue esta empresa es clasificar los productos de acuerdo con sus características cualitativas y cuantitativas. La clusterización que realiza la empresa está basada en el comportamiento de los clientes.

La información recopilada pertenece a una farmacia en la República Checa. Utilizan un algoritmo genético y se demuestra que, si el número de clústeres es mayor al número original, el método proporciona una mejor estructura de la categorización del producto. El objetivo es encontrar el número de clústeres que minimice el número de productos dentro de un mismo clúster en el carrito de compras. La idea es que el consumidor no comprará más de un producto de cada clúster, por ende, los productos con clúster similar deben de estar homologados en uno solo. El mejor clustering fue dado por la función objetivo mediante un índice aleatorio. La siguiente función fue utilizada para determinar si el consumidor está comprando dos tipos de productos del mismo clúster. Si esta situación se cumple, la función debe ser 1, de lo contrario el resultado es 0.

$$f_{cost}(x) = \frac{1}{n_B} \sum_{b=1}^{n_B} \frac{v_b(x)}{D_B} \quad (1)$$

donde:

V_b es el clúster independiente de x

N_b es el número de canastas

D_b es el número de decisiones.

Mencionan que los mejores resultados referente a los clústeres sucedieron cuando se optimizó la función, al momento de minimizar la. Esto da una aproximación a la solución del uso de un algoritmo genérico heurístico.

Huerta (2011) realizó una segmentación en una empresa de retail especializada. Se enfocó en el tipo de clientes que visitaban los puntos de venta mediante encuestas e información de los clientes. Logró obtener una segmentación de clientes con base en el comportamiento de estos y así tener un mejor conocimiento de sus preferencias y poder atacar mejor a cada segmento.

En conclusión, se obtuvo un método preciso siempre y cuando no se rompieran los diferentes supuestos. Lograron determinar que el comportamiento del cliente se basaba en seleccionar ciertos productos. Los tickets de compra no eran lo suficientemente grandes en cuanto a la cantidad de productos que llevaba el cliente y normalmente se violaban los supuestos.

Capítulo 2. Metodología

En el presente capítulo se mencionan las diferentes metodologías para resolver el problema en cuestión. Estas herramientas están relacionadas con los avances que se han hecho en temas de la Industria 4.0. Posteriormente, se realizará la selección de la metodología a utilizar para obtener una solución que apoye a resolver el problema presentado. Se explican las razones por las cuales la ciencia de datos es la más adecuada.

Logística

A lo largo de los años, Fontalvo et al. (2019) la logística ha evolucionado de algo muy básico a una terminó el cual hoy en día es vital dentro de ciertas empresas. Antes la logística se definía como un concepto de traslado de mercancía entre un punto y el otro. La logística ha evolucionado conforme a las necesidades de los clientes y la evolución tecnológica. Debido a la evolución tecnológica, es posible encontrar definiciones que consideran varias etapas de la logística, por ejemplo, la propuesta por Amr et al. (2019) donde describe cuatro etapas de las cuales son:

- **Logística 1.0:** Dada la primera revolución industrial, la producción manual cambia a ser producida por maquinaria a vapor en el siglo XVII. Posteriormente en el siglo XIX, la logística fue introducida como el movimiento y planeación de las tropas. En 1964 comienza la logística empresarial la cual se enfoca en la distribución de los bienes y se relaciona con las tres P.
 - Localización y destino (Place)
 - Valor del tiempo (Period & Pace)
 - Forma del orden (Pattern)

Su objetivo era la optimización del transporte y el movimiento de los bienes dentro de una organización.

- **Logística 2.0:** Dada la industria 2.0, reconoció la importancia de la producción en masa. Por lo que ante la automatización de la manipulación de la carga se crea la logística 2.0 que es considerada una evolución de la versión 1.0. Posteriormente en los años 80, surge la gestión de la cadena de suministro que incluye la planeación y control de todas las actividades involucradas en las adquisiciones y el abastecimiento, conversión y todas las actividades del manejo de la logística.
- **Logística 3.0:** Aparece en la tercera revolución industrial (1968) dada la manufactura hecha por robots. Surge debido a la creación del sistema del manejo de la logística. Esta considera dos nuevas P's a las previamente mencionadas.
 - Gestión de flujo
 - Agilidad
- **Logística 4.0:** Oyarzún (2018) define esta etapa y dice que, "Ahora, la cuarta se caracteriza por la automatización total de la manufactura mediante sistemas ciber físicos que combinan la maquinaria física con procesos digitales y son capaces de tomar decisiones descentralizadas y de cooperar mediante el internet de las cosas".

Existen otras definiciones para el concepto de Logística. Ramírez (2015) define la logística como la parte de la cadena de suministro encargada de obtener, producir y distribuir material y producto en el lugar acordado y en la cantidad requerida a un precio adecuado. De acuerdo con la cita que realiza Ramírez (2015) “la logística se limita dependiendo de cómo los productos son llevados hacia el consumidor final” y la gestión de la logística hoy en día influye en temas estratégicos, ya que tiene gran relevancia en la economía de cada empresa. Malindžák (2014) explica que el consejo del manejo de la logística, la definió como un proceso de planeación, implementación y un efectivo manejo de las pérdidas, almacenamiento, servicios e información de cómo se consume el producto o servicio. Menciona que existen 6 pilares de la teoría de la logística:

1. La logística como una teoría integral, la que entiende el manejo de los objetivos, flujos de la cadena y la red como un sistema, un método aproximado al sistema.
2. La logística como un objeto del manejo son aquellos flujos de las cadenas y las redes.
3. La logística como variables manejables en tiempo y espacio.
4. La logística como un principio básico del manejo como coordinación de los procesos de la cadena.
5. La logística como un principio básico del manejo de la cadena usando planeación y métodos de calendarización.
6. La logística orientada a la optimización global de las cadenas y las redes.

También menciona que en general la optimización de la cadena logística está basada en la idea de la optimización local, una optimización jerárquica.

Control e implementación de la logística global

Kain, R. et al. (2018) menciona que el término del manejo de la logística proviene del manejo del flujo de materiales desde que es materia prima hasta el cliente y a su parte el flujo de información del control de los movimientos. Es la parte del manejo de la cadena de suministro que planea, implementa y controla de manera eficiente y efectiva el flujo de bienes o servicios entre el punto de origen hasta el punto de destino.

Simchi-Levi, D et al. (2005) menciona que el problema de la oferta y la demanda ha sido uno de los problemas con los que se han tenido que enfrentar las empresas. El contar con la habilidad para poder satisfacerlo es clave para ellas, de lo contrario se pueden presentar varias pérdidas en las ganancias. Varias empresas a lo largo de los años buscan afrontar el problema con estrategias efectivas para evitar incrementos en costo, inventario y tiempos de respuesta. Las empresas tienen que lidiar con la incertidumbre, la demanda nunca se puede pronosticar exactamente. Es por eso que las redes logísticas deben ser diseñadas para eliminar la mayor incertidumbre posible. Una de las razones principales de dicha incertidumbre es la dinámica de la cadena de suministro, la variabilidad en la creación de los pedidos provoca lo que se conoce como un efecto látigo. Ya que la variabilidad de la demanda es estable entre el consumidor y la cadena minorista, pero existe una alta variabilidad en los pedidos que generan dichas cadenas minoristas.

La logística ha evolucionado drásticamente a lo largo del tiempo y se ha logrado un avance referente a la automatización de procesos lo cual permite tomar decisiones para la mejora. La logística dentro de los procesos de Liverpool es vital, ya que su buen manejo permite que la mercancía llegue de la mejor manera a cada centro a los que va cada sección y esto a su vez disminuye los retrasos y evita impactos negativos en los resultados.

En el presente trabajo se considera la logística como el conjunto de procesos que se requieren para garantizar un flujo eficiente y efectivo de los procesos. Gracias a la logística, se puede lograr que los productos sean entregados en las condiciones acordadas al momento de la compra. Este es esencial para que el cliente se encuentre satisfecho, superando los desafíos que surjan durante el proceso.

Inteligencia artificial

La inteligencia artificial hoy en día es de suma importancia, logra resolver problemas complejos de una manera eficiente. Abeliuk et al. (2021) menciona que los comienzos de la herramienta se remontan al año 1956. Un año en el que se lanza un programa llamado "Logic Theory Machine" por Allen Newell, Herbert Simon y Cliff Shaw, el cual era capaz de descubrir demostraciones de teoremas de lógica simbólica. A través de los años la inteligencia artificial ha evolucionado drásticamente y lo cual ha tomado fuerza en los años recientes. Para eso, varios autores han llegado a definir el concepto.

Ongsulee (2017) dice que la inteligencia artificial (IA) es parte de la ciencia de datos; es aquella inteligencia exhibida por las máquinas. Esta se aplica cuando una máquina imita funciones "cognitivas" que los humanos asocian con otros humanos. Actualmente, las capacidades que tiene la inteligencia artificial es imitar el habla humana, juegos estratégicos (p. ej., ajedrez y Go), manejar coches autónomos (marcas como Tesla tienen este tipo de tecnología implementada), enumeración de redes de entrega, simulaciones militares e interpretación de datos complejos, entre otras. Entre los problemas que busca resolver la inteligencia artificial está el razonamiento, el conocimiento, la planeación, comunicación y la habilidad de manipular objetos.

Xu et al. (2021) mencionan que la inteligencia artificial es uno de los pilares de la ciencia de datos. Mientras que Ramachandran (2022) dice que la inteligencia artificial es una metodología de suma relevancia ya que se presenta como una herramienta que ha incrementado y mejorado el desarrollo de los empleados. Gracias a la inteligencia artificial se ha logrado automatizar máquinas con grandes beneficios en varias industrias.

A lo largo del tiempo, la implementación de la inteligencia artificial dentro de las operaciones ha reducido los errores humanos. Como concluye Ramachandran (2022), la inteligencia artificial ha sido altamente adaptada en las organizaciones ya que tiene un rol importante en varias secciones como en recursos humanos para seleccionar candidatos o para analizar los puntos de mejora de rendimiento de la fuerza laboral.

La inteligencia artificial es una herramienta la cual ha tomado fuerza en las últimas décadas. Actualmente es una herramienta la cual se está implementando cada vez más con el objetivo de mejorar procesos y resultados dentro de una organización. Esta ha logrado demostrar dentro de distintas áreas, un potencial, el cual permite un desarrollo de diferentes industrias. En el presente trabajo se considera a la inteligencia artificial como la herramienta que trata de emular comportamientos humanos a través de algoritmos de computadora y el

análisis masivo de datos. Ha demostrado la capacidad para reducir errores humanos provocados por la fatiga, monotonía o sesgos no intencionales. Dicha herramienta puede beneficiar a Liverpool ya que puede resolver problemas y tareas de una manera más eficiente, permitiendo que los procesos sean más ágiles y precisos.

Machine learning

El machine learning conforme a los años, ha evolucionado drásticamente. Latorre, M. (2018) la historia del Machine Learning se remonta los años 50 's, cuando Alan Mathison Turing realiza una prueba la cual media la inteligencia de una computadora, esta prueba sería conocida como el “test de Turing”. Posteriormente el año 1952 Arthur Samuel dio a conocer el primer programa capaz de aprender, un programa de damas que mejoraba cada vez que uno jugaba contra la computadora. Unos años después, en los años 80 's se da a conocer el algoritmo del “vecino más cercano”, un algoritmo que permite adelantar una solución efectiva a los problemas que se le presenten, con el objetivo de volver más precisa la solución y así se pueda obtener un mejor resultado de aquellas operaciones involucradas. Varios autores han definido el concepto, este es de gran relevancia ya que su evolución ha permitido avances en varios sectores como en la medicina y el sector de seguridad. Vázquez, A. M. (2018) considera que es una de las primeras herramientas dedicadas a la identificación de patrones.

Helm et al. (2020) menciona, el Machine Learning, ML o aprendizaje automático en español, es un subconjunto de la inteligencia artificial. El objetivo es un aprendizaje a partir de datos históricos. Además, tiene la capacidad de mejorar el análisis a través de algoritmos que permiten que aprenda. Los algoritmos que se utilizan son capaces de reconocer patrones y con base en los reconocimientos pueden aprender de una manera autónoma. Ongsulee (2017) menciona que el aprendizaje automático está altamente relacionado con la estadística computacional enfocado en la elaboración de predicciones mediante el uso de computadoras. Tiene una fuerte vinculación con la optimización matemática. Es un método para diseñar modelos complejos y algoritmos que se prestan a la predicción. Para Nakaura et al. (2020) el aprendizaje automático es una técnica para analizar y predecir datos. Dada su aplicación se pueden encontrar patrones que pueden ayudar para la implementación de un conjunto de datos. Existen los siguientes tipos de aprendizaje automático:

- **El aprendizaje automático supervisado**, que es aquel código que se entrena donde el resultado es conocido. El algoritmo recibe un conjunto de entradas con sus respectivas salidas y aprende comparando los reales con los resultados. Usualmente es utilizado en aplicaciones donde los datos predicen eventos futuros probables. Rojas (2020) define al aprendizaje supervisado como aquel aprendizaje que se le enseña al algoritmo para alcanzar el objetivo para el que fue planeado. Usa un conjunto de datos clasificados bajo una cierta apreciación o idea para encontrar patrones que pueden aplicarse en un análisis y producir una salida ya conocida. Hay diferentes tipos de algoritmos dentro del aprendizaje supervisado, incluidos:
 - Árboles de decisiones
 - Regresión lineal
 - Clasificación Bayesiana

- Regresión logística
- **El aprendizaje no supervisado**, se usa contra datos que no tienen etiquetas, al sistema no se le da la respuesta correcta. El algoritmo la averigua por sí solo. Este tipo de aprendizaje puede usarse para aprender y establecer líneas de base, perfiles de comportamiento para entidades y con esto encontrar anomalías. El objetivo del algoritmo es explorar los datos y encontrar patrones. Las técnicas más conocidas incluyen mapas autoorganizados, mapeo del vecino más cercano y k-means.
- **El aprendizaje semi supervisado** se usa para las mismas aplicaciones que el supervisado, pero con datos etiquetados y no etiquetados para entrenar al algoritmo.
- **El aprendizaje por refuerzo** se usa en la robótica, juegos y navegación. El algoritmo descubre a través de la prueba y el error las acciones que producen mayores resultados.

En resumen, el machine learning, es aquella que permite que los algoritmos realicen tareas de manera automática, utilizando algoritmos que aprenden de los datos que se le hayan suministrado. Esto se logra mediante la identificación de patrones. Este aprendizaje es la base para que las tecnologías tengan mayor alcance en las actividades que desarrollan. A continuación, se desarrolla a profundidad el aprendizaje utilizado.

Aprendizaje no supervisado

Según Dayan et al. (1999), el aprendizaje no supervisado se remonta años atrás con influencia de ciertos investigadores como Horace Barlow en 1992, quien buscó caracterizar códigos neuronales. Donald Mackay adopta un enfoque ciber teórico en 1956 y David Marr en 1970 realizó el primer postulado sobre el aprendizaje no supervisado con objeto de aprender en su modelo.

Rojas (2020) define el aprendizaje no supervisado como un modelo predictivo, entrenado de manera similar al aprendizaje supervisado, pero con la diferencia de que la comprensión se da en datos no clasificados o etiquetados y descubre los patrones entre grupos similares.

Nakura et al. (2020) dicen que el aprendizaje no supervisado en comparación con el aprendizaje supervisado presenta un problema en el cual no tiene una referencia como su contraparte, si no dadas las iteraciones va encontrando el resultado ideal.

En un artículo presentado por Dike et al. (2018), se menciona que el aprendizaje no supervisado es un algoritmo para separar en sets o clusterizar que se basa en las características de los datos, organiza de manera autónoma o se adapta ya que los datos ingresados no cuentan con un resultado específico, posteriormente menciona que algunos de los beneficios del aprendizaje no supervisado son:

- Puede aproximar cada función sin importar su linealidad.
- Es un lenguaje adecuado para problemas complejos y abstractos.
- El proceso de aprendizaje puede ser jerárquico de lo particular hasta lo más general como son los símbolos.

Nielsen (2016) describe los clústeres jerárquicos como una técnica para hacer un análisis exploratorio de los datos, una técnica de algoritmo no supervisado. Consiste en construir un

árbol de combinación binario, comenzando por los elementos almacenados en conjuntos y procede a fusionar dos subconjuntos cercanos hasta llegar al centro de los elementos.

Este tiene dos aproximaciones, la aglomerativa y la divisiva.

- El acercamiento aglomerativo es el acercamiento donde los puntos de los datos se combinan con los otros.
- El acercamiento divisivo calcula la carga en la que puede ser divisible los datos.

Los clusteres no jerárquicos son aquellos que categorizan el clúster directamente, se pueden encontrar algoritmos que cambian sus centros hasta que todos los puntos estén relacionados a centros. K-means es uno de los ejemplos de los clusteres no jerárquicos que calculan la clasificación en un menor tiempo. Este clúster tiene 4 aproximaciones: partición, basado en densidad, basado en una red y otros acercamientos. Los clústeres de densidad consideran intensamente los espacios de los datos como un clúster.

Este tipo de aprendizaje ha sido relevante para muchos sectores de la industria ya que con ello pueden hacer divisiones como la clusterización y puedan segmentar de mejor manera logrando hacer más eficientes sus procesos. El aprendizaje no supervisado permite que los usuarios diferencien de mejor manera los resultados ya que no se colocan etiquetas y así va aprendiendo constantemente conforme vaya resolviendo los patrones que se le presentan. Este, al poder identificar y analizar patrones puede servir para diferentes aplicaciones como el reconocimiento facial o el reconocimiento de fraudes.

Reducción de dimensionalidad

La reducción de la dimensionalidad es una de las técnicas que se abordan en el aprendizaje no supervisado. Hernández (2016) explica que los métodos de reducción de dimensionalidad son algoritmos que mapean un conjunto de datos. Estos son reducidos por lo cual se permiten hacer una descripción de los datos de una manera más precisa ya que los datos se encuentran acotados.

Existen diferentes tipos de análisis clasificados de la siguiente manera:

- **Análisis de componentes principales (PCA)**
 - Es una técnica estadística que se utiliza para eliminar la redundancia de datos y se basa en el supuesto de linealidad de los datos.
 - Permite la reducción de los datos a través de la ecuación de transformación

$$Y = PX \tag{2}$$

Donde:

P es la matriz ortogonal denominada matriz de representación.

Su objetivo es determinar la P que permita que los datos puedan ser proyectados en una menor dimensión.

P busca la no correlación entre vectores de Y .

- **Análisis de componentes principales probabilístico (PPCA)**
- **Análisis de componentes principales con Kernel y el problema de la preimagen (KPCA)**

- **Incrustación estocástica de vecinos distribuidos t (t-SNE)**
 - Utilizada para visualizar datos de alta dimensión en un menor espacio.
 - Su objetivo es preservar la estructura local de los datos. Mediante las similitudes de dos puntos.

Técnica de clustering

La clusterización no es un tema reciente, dicha metodología ya se menciona en la década de los 70. Previamente abordado por Dalenius en 1950, donde buscaban estimar el valor esperado de una variable aleatoria con una distribución de densidad en distintos intervalos. Dalenius no utilizó el algoritmo para minimizar si no para encontrar el número óptimo de particiones de una clase.

Villardón (2007) define el clustering como un término anglosajón que se refiere a la técnica utilizada para el análisis exploratorio de datos, con el propósito de resolver problemas de clasificación de datos. Los clústeres tienen como objetivo ordenar objetos de forma al grado de similitudes entre ellos.

El análisis clúster intenta identificar de toda manera posible la tendencia de la información para formar grupos, busca la agrupación de grandes datos de información. Las técnicas de clusterización, se pueden diferenciar de la siguiente manera:

- **Clustering jerárquico**

Estos son clústeres formados por la reiteración de patrones usando barridos de arriba hacia abajo o de abajo hacia arriba, existen dos metodologías aglomerativas que son aquellas que utilizan un barrido de abajo hacia arriba. Estos comienzan a formar los clusteres con un simple objeto y luego van iterando, uniéndose a uno de los clústeres cada vez más grande. Los clústeres divisivos son aquellos que utilizan la metodología de arriba hacia abajo para formar clústeres y los va dividiendo en clústeres más pequeños.

- **Clustering de enlace sencillo**

Este método es comúnmente conocido como el método del vecino más cercano. El enlace que se crea entre dos clústeres es el enlace que se encuentra con el miembro más cercano, la distancia más corta. La similitud entre estos clústeres es considerada como igual a la similitud más grande de cualquier miembro de un clúster.

- **Clustering de enlace completo**

Método conocido como el vecino más lejano, este es determinado por la distancia más larga entre uno y otro miembro de dos clusteres diferentes.

- **Clustering de enlace promedio**

También conocido como el método de la variabilidad mínima. La distancia mínima está determinada por el promedio de las distancias de los miembros de un clúster y otro clúster.

- **Reducción iterativa equilibrada y agrupamiento mediante jerarquías (BIRCH)**

Lorbeer et al. (2018) mencionan que este método es uno de los más rápidos en la actualidad. Se caracteriza por la alta calidad y velocidad y supera a los demás

métodos hasta por dos órdenes de magnitud. Este método necesita que se conozcan tres parámetros para ejecutarlo. Se necesita el factor de ramificación Br, el límite T y el conteo de clústeres K. Este método puede ser utilizado para grandes volúmenes de datos, ya que dada su alta velocidad puede realizar clústeres de grandes cantidades de datos en cuestión de segundos.

- **Matrices de distancia**

En este se encuentran las matrices de distancia, las cómo se puede encontrar en el texto presentado por Demey (2011) los tipos de distancias son los siguientes:

- **Minkowski**

La distancia utilizada para medir la similitud entre dos objetos. Checa la similitud dada la suma de las distancias de los distintos objetos.

- **Manhattan**

Es la suma de las diferencias absolutas entre unidades para cada variable. Es mayor que la distancia euclidiana. Diseñada para distancias de mínimo recorrido.

$$d_{(i,j)} = \sum_{t=1}^T |x_{it} - x_{jt}| \quad (3)$$

donde:

T son las distintas variables

x_{ij} número de variables en el punto i

x_{jt} número de variables en el punto j

- **Euclidiana**

Es la distancia más conocida y la que tiene más uso; es la herramienta de cálculo fundamental para métodos multivariados basados en distancias. Esta ocurre cuando se representan las variables cuantitativas en el espacio de coordenadas P .

$$d_{(a,c)} = \sqrt{\sum_{b=1}^P (x_{ab} - x_{cb})^2} \quad (4)$$

donde:

P son los rasgos funcionales

x_{ab} número de rasgos en el punto i

x_{cb} número de rasgos en el punto j

Es aquella distancia que se traslada de un punto A a un punto B

Las distancias son aquello que ayuda al algoritmo de k-means a determinar cuáles son los puntos de los datos que son parte de cada centroide. Esta distancia es relevante para determinar su clasificación dependiendo el rendimiento que pueda tener.

Como parte de la visualización de las características de los clústeres, existe la cardinalidad y la magnitud. La cardinalidad se refiere al número de combinaciones por clústeres. Gráfica todos aquellos puntos de los clústeres en una gráfica de barras. En cuanto a la magnitud, muestra la suma de todas las distancias al centroide de los clústeres y las muestra como una gráfica de barras. Ambas se llegan a comparar, cuando se tiene una cardinalidad alta los clústeres se tiende a tener una magnitud más grande. Cuando la cardinalidad y la magnitud no tienen una correlación, se puede identificar un cluster irregular.

En la *Figura 07* se puede ver una representación de la cardinalidad de dos conjuntos A y B.

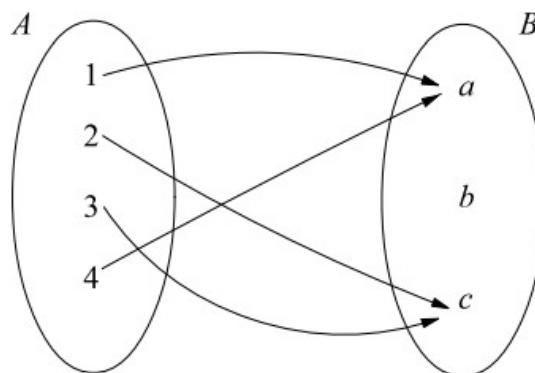


Figura 07 - Representación de cardinalidad

Gracias al clustering muchas empresas han podido segmentar sus datos para obtener un mejor resultado. El clustering permite que los procesos que realiza una empresa sean más eficientes y se obtenga de ello un mejor resultado.

K-MEANS

Al igual que los clusteres jerárquicos, K-Means es parte de los algoritmos de aprendizaje no supervisado. Pham et al. (2005) define k-means como un algoritmo popular de agrupación.

Bock (2007) menciona que una de las aproximaciones más utilizadas está basada en la suma de los criterios cuadrados y el algoritmo que hoy en día es conocido por “k-means” (k-medias), un algoritmo cuyo origen se les atribuye a varios científicos que lo utilizaban bajo diferentes supuestos. Dicho algoritmo trata de aproximar a un número óptimo de “k” (grupos).

Este algoritmo requiere un número de clústeres predefinidos, el cual se puede encontrar mediante el número apropiado para cierto conjunto de datos. El método calcula el número adecuado a través de la prueba y el error. Existen varios métodos para llegar al valor k, que es el número de grupos en los que se va a definir la información. El desempeño del algoritmo se basa en el número k seleccionado. Se puede seleccionar el número k de la siguiente manera:

- Por un set o rango
- Definido por el usuario
- En etapas futuras del proceso
- Valores agrupados a los números de generadores
- Definido por medidas estadísticas
- Valores equiparados al número de clases
- Determinados durante la visualización
- Usando la medida de la colindancia

K-means se utiliza en varias aplicaciones, como la clusterización de los datos. Sirve tanto para segmentar puntos de venta como para determinar los grupos en los que se debe recolectar una cosecha. La herramienta sirve para varios sectores, k-means es una de varias herramientas que existen para segmentar, pero es una de las más populares por lo que varias personas la utilizan para segmentar sus datos usando una clusterización.

La técnica que se utilizará en este trabajo es un algoritmo que usa k-means. Además, se utiliza una técnica de reducción de dimensionalidad que permite ver todas las variables ingresadas en solo dos dimensiones. Este fue seleccionado ya que permite analizar un conjunto grande de datos.

Pruebas de validación

Correa (2021) menciona que es importante tener pruebas que apoyen con la validación de los métodos utilizados de clusterización, pues no determinar un número correcto puede causar heterogeneidad, si es que son pocos clústeres o un número innecesario ya que se pueden crear clústeres con mucha similitud. Estos métodos buscan llegar al número óptimo de clusteres que se debe de tener para representar los datos de una mejor manera y que su clasificación sea más precisa. Existen varios métodos para corroborar el número óptimo de k, entre ellos:

- **Método del codo**

Este método hace un ajuste de los rangos para los valores de k. Se realiza una gráfica mostrando la distancia entre los datos (en el eje Y) contra la cantidad de clusteres (eje X), se puede ver que el número óptimo de k se observa donde hay un codo. En la *Figura 07* se muestra un ejemplo de dicho método, donde una línea marca el número de clústeres óptimos. En la ecuación 05 se puede observar la fórmula utilizada para calcular el punto de inflexión de la gráfica del codo.

$$W_K = \sum_{R=1}^K \frac{1}{n_r} D_r \quad (5)$$

donde:

W es el punto de inflexión de la curva.

k es el número de clústeres.

n_r es el número de puntos en el grupo r .

D_r es la suma de las distancias entre todos los puntos del grupo.

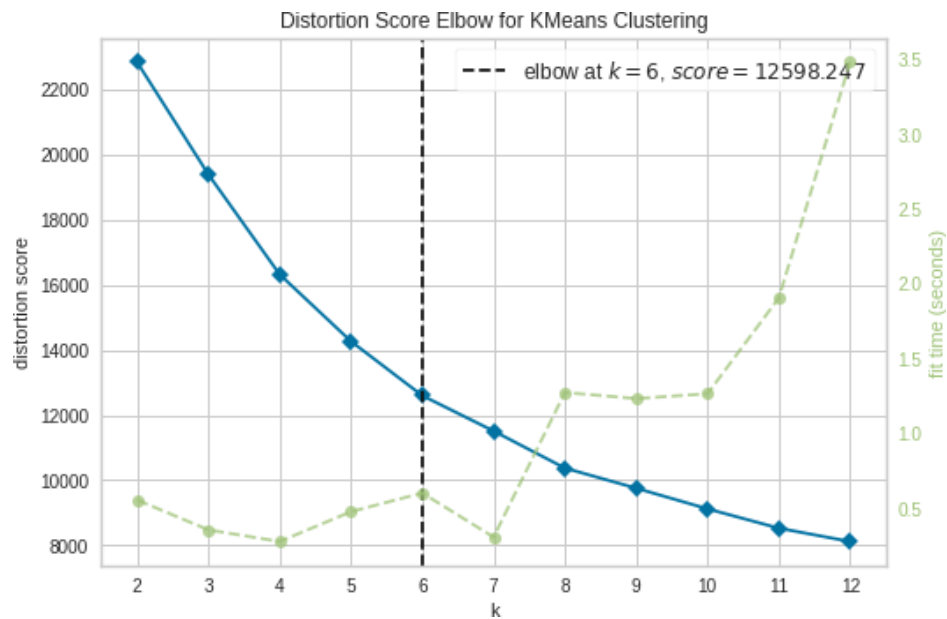


Figura 08- Método del codo para número óptimo de k

- **Método de la silueta**

Este método se utiliza para tener una buena medición de la agrupación de los datos y las distancias entre los clústeres. Muestra la medida entre los distintos centros. En la *Figura 08* se ve un ejemplo del método de la silueta, donde se determina el número de clústeres con sus respectivos coeficientes y magnitudes. En la ecuación 6 se podrá observar el cálculo para el ancho de la silueta.

$$S_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (6)$$

donde:

a es la distancia promedio de x a todos los demás puntos en el mismo cluster.

b distancia promedio de x a todos los demás puntos en el cluster más cercano.

i es el número de observaciones.

S_i es el número de clústeres.

Las observaciones con un S_i grande están bien agrupadas, mientras que las que cuentan con S_i pequeño llegan a ser en dos grupos, las que cuentan con S_i negativo probablemente están ubicadas en un grupo equivocado.

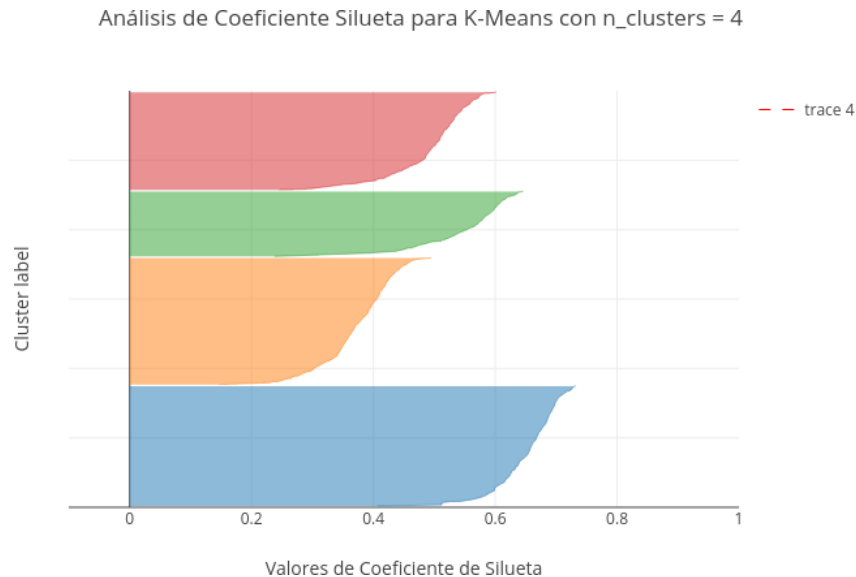


Figura 09- Método de la silueta para determinar k óptimo

Estas comprobaciones nos ayudan a corroborar que el número sea el más adecuado dependiendo de la cantidad de datos ingresados. El criterio de decisión recae en el usuario, pero este tiene el apoyo de un algoritmo. El algoritmo recomienda un número preciso para definir la información, un número más alto o más bajo podría ocasionar rupturas en el número óptimo. Divide un conjunto de datos similar en dos o haciendo un clúster demasiado grande cuando se puede segmentar en más. Para el presente trabajo, el clustering es una técnica que apoya al usuario a poder identificar distintos grupos dados por las similitudes que presentan entre sí. El poder identificar dichas similitudes permite que las segmentaciones sean más precisas.

Para el presente trabajo, el clustering es una herramienta la cual permite hoy en día la clasificación de grupos de datos en diferentes conjuntos. Cada conjunto está conformado por datos que tienen similitudes de tal grado. Esta metodología permite que empresas como Liverpool puedan realizar distribuciones de mercancía con más precisión, ya que se puede identificar qué tiendas son las que más venden ciertos productos y cuáles son aquellas en las cuales ciertos productos no son los más ganadores.

Análisis exploratorio de los datos (EDA)

Esta es una manera de analizar estadísticamente un conjunto de datos para poder observar su distribución y así poder tomar mejores decisiones para la investigación. Ayuda a conocer los datos, sirve para clarificar y obtener las ideas principales para tener un mejor acercamiento a la investigación. Su objetivo es checar la calidad de los datos, hacer una

descripción y una formulación de un modelo para poder ver a profundidad los datos y poder tomar lo más relevante.

Tuckey (1977) menciona que se trata de un trabajo de detective numérico, el cual consiste en un análisis numérico y gráficas cuyo objetivo es describir de la mejor manera posible todos aquellos aspectos de los datos. El analista de datos necesita herramientas para entender los datos proporcionados. En su libro “Exploratory Data Analysis” sugiere varias herramientas para el análisis y encontrar la mejor solución. Entender la diferencia tiene limitantes, no todas las personas son capaces de hacer un análisis sin las herramientas; es necesario ser muy detallista para hacer un análisis efectivo de los datos.

El análisis exploratorio de los datos está basado en gráficos que permiten observar la distribución de los datos, por ejemplo, los histogramas que ayudan a observar la frecuencia de la distribución. Otro gráfico recomendado es la gráfica de cajas y bigotes que nos permite comparar entre ciertos grupos de información como el límite de los datos, la media y los cuartiles. Las gráficas importantes que ayudan a observar la información son las de dispersión y las series de tiempo.

Como parte del EDA, se encuentran los mapas coropléticos los cuales son una representación cartográfica de los datos. Es un gráfico que muestra la densidad de los datos en un espacio geográfico determinado. Estos mapas se usan para mostrar la variación de los datos en un espacio específico. Existen dos tipos de mapas:

- **Los mapas no clasificados** asignan un color con base en el valor de la unidad. Los mapas muestran información, pero carecen de sustento para hacer una discriminación visual.
- **Los mapas clasificados** hacen una clasificación de los datos con base en valores y consideran las áreas geográficas; delinean de una mejor manera las regiones que se le interesan a la variación para poder tomar decisiones más claras y precisas.

Los dendrogramas son parte del EDA, Villardón (2007) los define como “una representación gráfica en forma de árbol que resume un proceso de clustering. Los objetos similares se conectan mediante enlaces cuya posición en el diagrama está determinada por el nivel de similitud/disimilitud entre los objetos”. La decisión sobre el número óptimo de clústeres es subjetiva, especialmente cuando se incrementa el número de objetos.

Los dendrogramas son una reducción de la estructura de los datos. Es la representación gráfica del clustering basado en similitudes entre las distancias que presentan los datos. Este inicialmente comienza con los datos que tienen una similitud de cero y va en cascada, poco a poco, hasta encontrar los grupos de datos que tienen más similitudes.

Son una forma de poder determinar el número de clústeres necesarios para segmentar la información. Mediante las distancias entre los datos logra segmentar los datos en clústeres los cuales fueron definidos por un número k óptimo. Un dendrograma apoya al usuario a observar el clúster de los datos de una manera más visual como se muestra en la *Figura 10*.

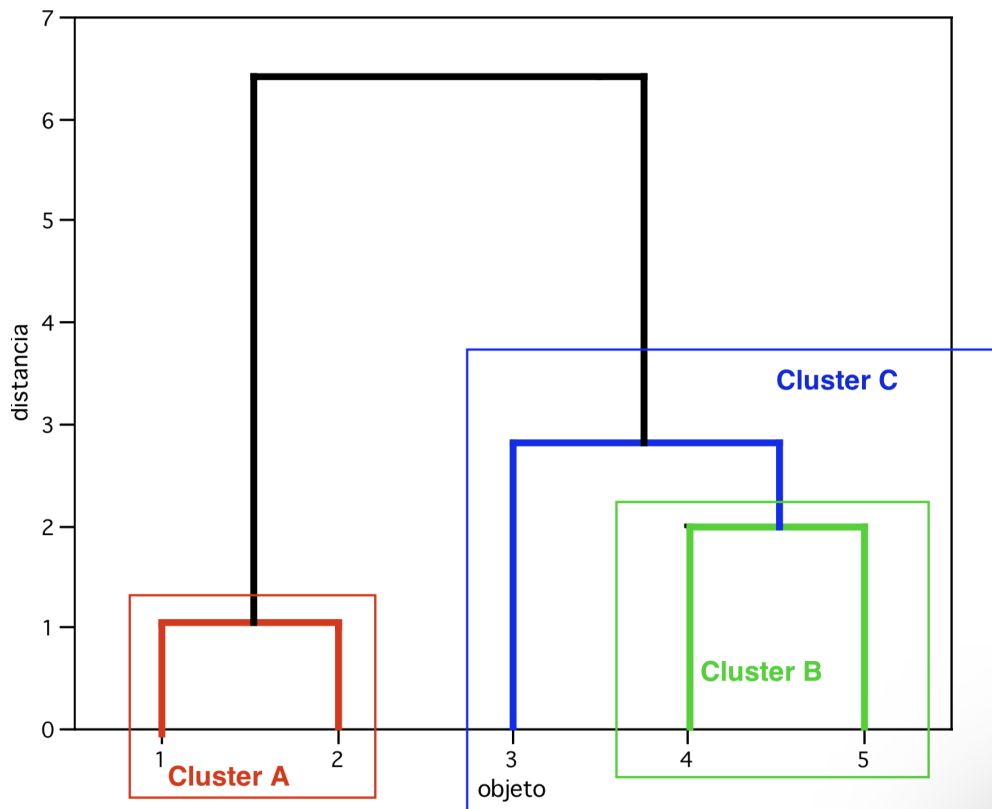


Figura 10 - Ejemplo dendrograma

El objetivo del EDA es que las tablas y las gráficas cuenten su historia de manera autónoma, sirve para que el lector entienda de una manera más simple los datos. El resumen de los datos debe de estar redondeado, las gráficas deben de tener sus ejes bien etiquetados y deben de tener una explicación clara para que el lector no tenga obstáculos para entender lo que observa.

Key Performance Indicator (KPI)

Ríos (2019) menciona que “Un KPI es una herramienta que permite estandarizar argumentos cuantitativos” (p. 24). Es una manera de monitorear el rendimiento de determinado indicador dentro de la empresa. Pursell (2023) dice que es un indicador clave de desempeño que permite medir el éxito de una actividad, estrategia o proceso. Menciona que para una buena medición los KPI deben de ser alcanzables, medibles, relevantes y periódicos. Aclara que existen varios tipos de KPI, por ejemplo, de ventas, marketing, financieros y logísticos, entre otros. Algunos ejemplos de KPI incluyen el retorno de la inversión, la tasa de productividad, la tasa de ausentismo, entre otros. Uno de los principales KPI con los que mide Liverpool sus resultados es la rotación del inventario.

La rotación es un indicador relacionado con el inventario. Este es el que nos indica si se cuenta con el inventario adecuado acorde a la venta. La rotación del inventario permite determinar si una tienda está sobre inventariada o no. Es el número de veces que el inventario se vende o se consume durante un periodo determinado. Un número bajo de rotación significa que el inventario no se consume rápidamente, por lo que se puede decir que hay un sobre inventario. Al contrario, un número alto puede significar que la mercancía se está consumiendo varias veces por lo que se podría decir que hay poco inventario y

conforme hay venta se resurta constantemente. Lo que se busca obtener con el presente trabajo es tener una rotación del inventario mayor a la del año anterior ya que lo que nos puede decir si es que la tienda ha sido más rentable que el año anterior. En la ecuación 07 se podrá observar la manera en la que se puede obtener la rotación anual del inventario.

$$r = \frac{s}{p} \quad (7)$$

Donde:

r es la rotación del inventario anual.

s es la venta anual.

p es el inventario promedio del mes 1 al mes 13 (enero año en curso a enero del siguiente año).

Un factor que influye significativamente en el KPI de rotación es la mercancía de nulo movimiento. En Liverpool, el nulo movimiento se refiere a aquellos productos que no se han vendido durante un período determinado. Estos productos no vendidos ocupan espacio en el inventario, impidiendo la introducción de nuevos productos debido a la falta de espacio disponible. Como resultado, se generan pérdidas en ventas ya que un producto estancado evita la entrada de uno nuevo con potencial de venta.

Hay dos costos principales que resultan del nulo movimiento:

- Costo logístico: Estos son los gastos relacionados con la planificación y el control eficiente del flujo de inventario. Incluyen costos como el transporte, que se incurren al transferir mercancías de un centro a otro.
- Costo de almacenamiento: Son aquellos gastos asociados con el mantenimiento de los inventarios en los centros y centros de distribución.

La rotación es un KPI esencial para medir la eficiencia en la gestión de los inventarios. Un factor que afecta es la mercancía de nulo movimiento, es decir, productos no vendidos que ocupan espacio y generan pérdidas. Estos productos causan costos logísticos y de almacenamiento, por lo que impiden que se introduzcan nuevos productos. La mejora de la rotación puede aumentar la rentabilidad de los centros y las secciones. Optimizar la rotación del inventario es clave para mejorar la eficiencia y rentabilidad.

Datos

Los datos recopilados representan dos grupos de artículos de una de las secciones que conforman la zapatería infantil. En este caso los grupos de artículos Tenis y Sandalia de la sección 402. Los datos recopilados fueron tomados como referencia del último año antes de la pandemia 2019 y el primer año "post pandemia" que fue el 2021. Estos años nos podrán ayudar a observar el comportamiento de ambos grupos de artículos, cuál ha sido su cambio dada la pandemia y cómo ha sido el rendimiento de las tiendas dada la afectación.

Ya recopilada la información de las cifras de ventas, utilidades, costos de meses sin intereses, costo de descuentos, entre otros, por centro, se unificaron los cuatro archivos (Tenis 2021 y 2019 y Sandalia 2021 y 2019) y se hizo una limpieza de datos.

En los archivos de 2019 se borraron las Fábricas de Francia que terminaron operaciones en 2018, así como las tiendas que cerraron durante el periodo 2019-2021. Para ambos años se borraron los rubros de “venta por teléfono” y “ventas por internet”, ya que a partir de 2019 las ventas de dichos rubros ya eran contabilizadas para la tienda más cercana al código postal. Se sumaron dos columnas, “Costo de MSI Dilisa” de tarjetas de la compañía y “Costo de MSI externas” las cuales son de cualquier banco externo estas se convirtieron en “MSI”. También, la columna “Descuento” es la resultante de la suma de 3 columnas “Descuento Tecla”, “Descuento Cupón” y “Descuento Empleado”. Se borraron columnas como “Undering and Overings”, las cuales no aportaban información relevante al análisis. Al unificar los archivos y realizar un pequeño análisis se borraron algunas columnas que no aportaban suficiente información o tenían una alta correlación con otra columna como, por ejemplo, ventas por unidades y ventas en pesos o columnas que por el momento no proporcionan datos relevantes para la diferenciación como el número de sección.

En la *Tabla 05*, se puede observar un muestreo de la base de datos final utilizada. La información de la base de datos está dividida en 4,800 filas y 13 columnas, que son:

- **Grupo_de_Artículo:** Es la descripción del grupo de artículo.
- **Zona#:** Es el número de zona donde se encuentra el centro.
- **Centro#:** Es el número que le asignó Liverpool al centro.
- **Centro:** Es la descripción que le da Liverpool a su centro (Ej. Liverpool Perisur).
- **Año:** Es el año del cual se obtuvo la información de los datos.
- **Mes:** Es el mes del cual se obtuvo la información de los datos.
- **Latitud:** Es la latitud en la que se encuentra el centro.
- **Longitud:** Es la longitud en la que se encuentra el centro.
- **Ventas:** Son las ventas que presenta el centro en el mes y año seleccionados, este se encuentra en pesos mexicanos.
- **Utilidad:** Es la utilidad que presentó el centro en el mes y año seleccionados, este se encuentra en pesos mexicanos.
- **Descuento:** Es el total de descuento aplicado en el mes y año seleccionados, este se encuentra en pesos mexicanos.
- **MSI:** Es el total del costo de meses sin intereses aplicados en el mes y año seleccionados, este se encuentra en pesos mexicanos.

	Grupo de Artículo	Zona #	Centro #	Centro	Año	Mes	Latitud	Longitud	Ventas	Utilidad	Descuento	MSI
0	TENIS	1	1	Liverpool CENTRO	2019	1	19.43	-99.13	475995	166708	159673	15204
1	TENIS	1	1	Liverpool CENTRO	2019	2	19.43	-99.13	250674	97704	54208	1148
2	TENIS	1	1	Liverpool CENTRO	2019	3	19.43	-99.13	414350	166875	53595	14853
3	TENIS	1	1	Liverpool CENTRO	2019	4	19.43	-99.13	499024	202092	53136	14814
4	TENIS	1	1	Liverpool CENTRO	2019	5	19.43	-99.13	634800	208429	193041	18750
5	TENIS	1	1	Liverpool CENTRO	2019	6	19.43	-99.13	553374	198759	158378	9317
6	TENIS	1	1	Liverpool CENTRO	2019	7	19.43	-99.13	741602	271563	196146	23631
7	TENIS	1	1	Liverpool CENTRO	2019	8	19.43	-99.13	1078134	448818	162087	22776
8	TENIS	1	1	Liverpool CENTRO	2019	9	19.43	-99.13	330870	112665	76515	4542
9	TENIS	1	1	Liverpool CENTRO	2019	10	19.43	-99.13	450866	139884	130350	24247
10	TENIS	1	1	Liverpool CENTRO	2019	11	19.43	-99.13	727118	216102	299491	7305
11	TENIS	1	1	Liverpool CENTRO	2019	12	19.43	-99.13	729444	232099	186319	12814

Tabla 05 - Imagen tomada de la Base de datos

En la mercadotecnia se usa la segmentación de clientes con el propósito de dividir a todos los clientes en diferentes segmentos para así poder dirigirse de manera más personalizada a cada uno en cuanto a promociones, precio del producto, modelaje y gustos, entre otros.

En una cadena de retail o minorista, no encuentras al mismo tipo de cliente en todas las tiendas, siempre hay diferencias como es el poder adquisitivo o las necesidades o el flujo de clientes, existen centros comerciales más lujosos que otros.

En esta tesis lo que se busca es hacer una segmentación de los distintos puntos de venta de la cadena minorista. Lo que se quiere es identificar cuáles son los centros con más rotación, flujo de clientes y rendimiento a lo largo del año. Las distintas variables ayudan a determinar cuáles son las mejores tiendas y cuáles son los picos de rendimiento que tienen a lo largo del año.

Capítulo 3. Resultados obtenidos

En este capítulo se pueden observar gráficos que explican el comportamiento de los datos analizados con las herramientas mencionadas. Dicho análisis nos permite ver la situación de la sección, cómo se ha comportado a través de los años mencionados y cuál ha sido su desempeño. Se observa que se realiza un análisis para el perfilamiento de los puntos de venta y que esto pondrá a prueba el planteamiento del problema previamente establecido. Para efectos prácticos se ha decidido solo colocar los gráficos del año 2021, los gráficos del año 2019 se pueden encontrar en el anexo 04 las cuales comienzan desde la *Figura 32*.

Librerías

Las librerías utilizadas en los algoritmos son un apoyo al que puede acceder un usuario dentro del lenguaje de Python para poder acceder a distintas funcionalidades. Gracias a las librerías se puede manipular la data de una mejor manera, lo que permite poder realizar distintos análisis. A continuación, en la Tabla 06, se describen las diferentes librerías que se usaron para realizar la visualización de datos y el uso del algoritmo para la solución de la problemática presentada.

Librería	Descripción	Finalidad
Pandas	Librería que proporciona estructuras de datos y herramientas de análisis de datos.	Esta librería se utilizó para cargar el archivo CSV en Python y para poder manipular el data frame para distintos gráficos.
Seaborn	Librería que apoya a la visualización basada en Matplotlib. Sirve para realizar gráficos más estilizados y complejos.	Librería utilizada para la visualización de los gráficos de violín, dendrogramas y mapas coropléticos.
Matplotlib	Librería que apoya en la creación de gráficos como histogramas, series de tiempo, etc.	Librería utilizada para la visualización de los gráficos de codo, gráficos de bigote y gráficos de radar.
Plotly.express	Es una evolución de la librería Plotly.py, la cual simplifica en menos líneas de código la creación de gráficos complejos.	Librería utilizada para los gráficos de dispersión 3D.
Numpy	Librería para la manipulación de matrices.	Visualización del t-SNE MAP.

Librería	Descripción	Finalidad
Yellowbrick	Librería de visualización que permite gráficos del método del codo.	Librería de donde se importó la visualización de k means.
Scipy	Librería utilizada para cálculos matemáticos, sirve para la optimización y álgebra lineal para resolver problemas.	Librería facilitadora para la distancia euclidiana, principalmente para las gráficas de cardinalidad y magnitud.
Sklearn	Librería para el uso del aprendizaje automático, construida sobre SciPy.	Librería donde se descargo las funcionalidades para el PCA y el algoritmo de K-Means.

Tabla 06 - Librerías usadas

Resultados EDA

Matrices de correlación

En las *Figuras 11, 12, 31 y 32*, se observan las diferentes matrices de correlación de las segmentaciones de la base original. La matriz de correlación permite observar aquellas variables que tienen alta dependencia entre sí.

En estas gráficas se puede apreciar una similitud que es la alta correlación entre las variables de Ventas, Utilidad, Descuento y MSI, esto debido a que son variables con una alta dependencia entre sí. En un caso contrario, existen correlaciones con un valor muy cercano al cero. Un valor cero significa que no hay una correlación entre dichas variables como la correlación entre Mes y Longitud o Longitud y Mes donde se tiene un valor de cero. Es importante tener estas variables dentro del análisis, ya que nos muestran cuáles son aquellos centros que tienen más actividad promocional y aquellos que tienen más utilidad.

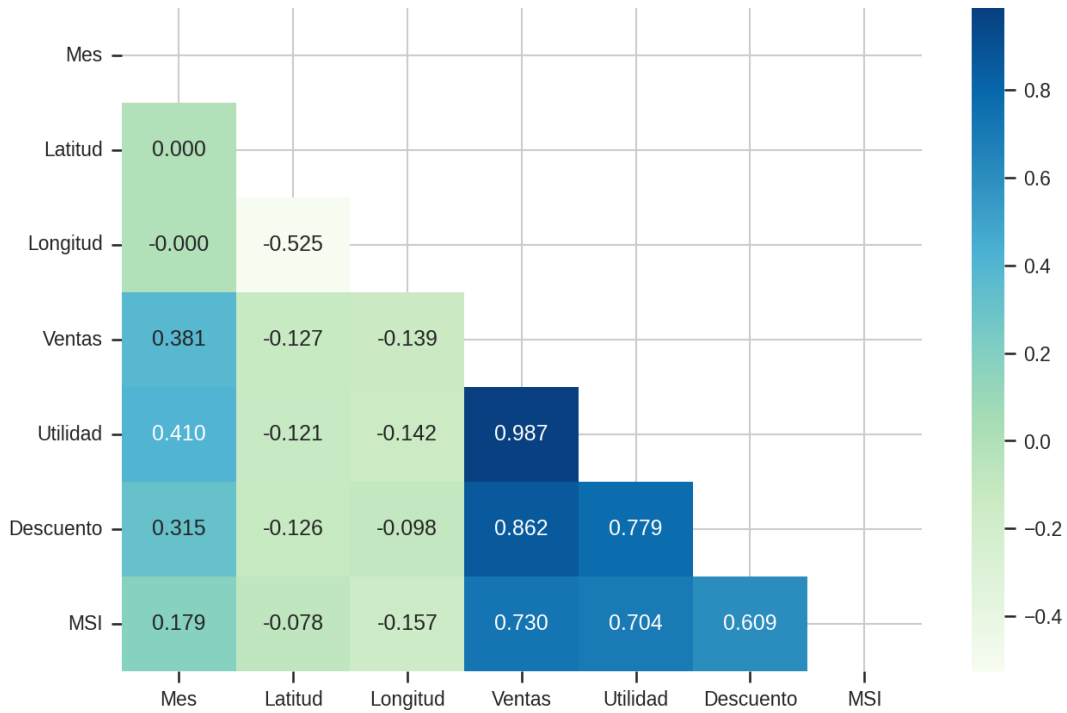


Figura 11 - Matriz de correlación Tennis 2021

Se observa que la intersección con más correlación es la de Utilidad con las Ventas y el caso contrario la variable de longitud y latitud son aquellas que tienen menos correlación con las demás variables.

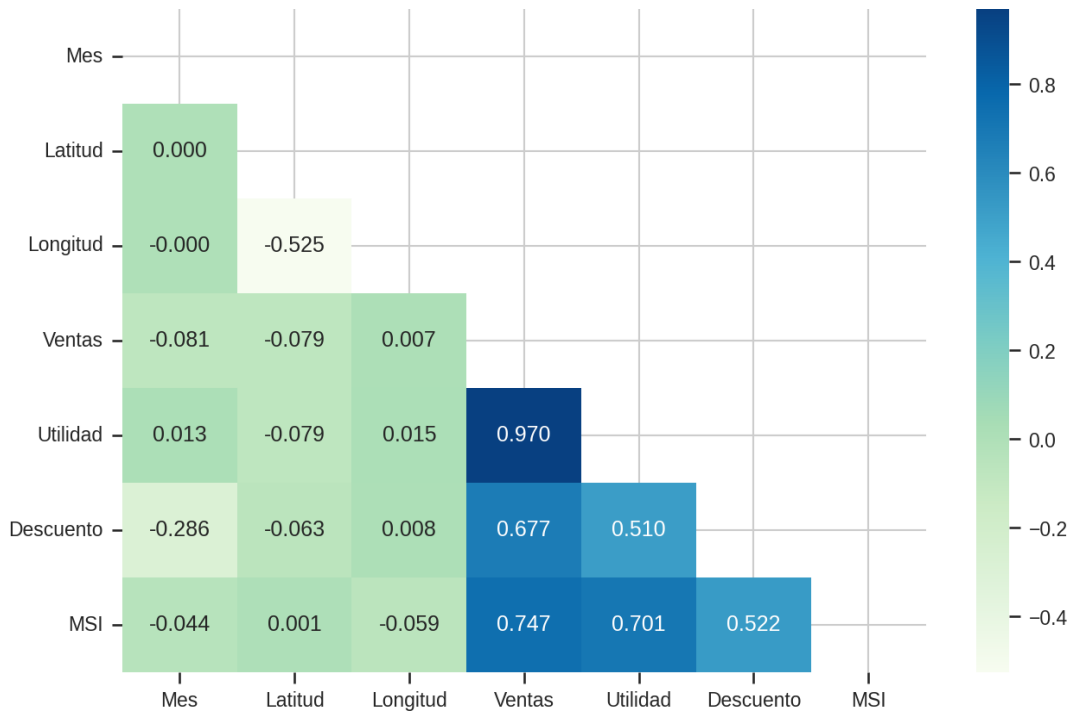


Figura 12 - Matriz de correlación Sandalia 2021

Se observa que la intersección con más correlación es la de Utilidad con las Ventas y el caso contrario la variable de longitud y latitud son aquellas que tienen menos correlación con las demás variables. En el caso de la intersección de Utilidad y mes es la cual tiene menor correlación cercana a cero.

Gráficos de violín

En las *Figuras 13, 14, 33 y 34* se observa el comportamiento de los segmentos previamente mencionados durante los 12 meses del año. En estos gráficos se puede observar la frecuencia de los datos mes a mes.

En la *Figura 13* se puede observar el comportamiento del grupo de artículo Tenis durante el 2021, un año considerado como “nueva normalidad” donde se tuvieron varios regresos a la normalidad como fue el regreso presencial a clases. Aquí se pueden observar varios meses similares, pues no había certidumbre al 100%. Se puede observar un enero muy pequeño, donde las utilidades no fueron tan elevadas. Se tiene como suposición que dicho comportamiento se debe a que en enero de 2021 se estaba pasando por la segunda ola de casos de Covid-19 en México, sin embargo, dicha suposición no se puede tomar como concluyente.

En la *Figura 14* se puede observar el comportamiento de las Sandalias durante 2021, una gráfica similar, pero se ven las diferencias por las afectaciones de la pandemia. En enero se ve una utilidad baja ya que se pasaba por una época de alto contagio; durante noviembre y diciembre se puede observar una utilidad baja debido a una etapa de contagio donde las familias no salían de vacaciones para cuidar su salud. Se pueden observar meses altos como abril y julio lo que se puede atribuir a la temporada vacacional.

En la *Figura 33* se puede ver el comportamiento del grupo de artículo Tenis durante el año 2019. En esta se pueden observar violines más alargados como los de los meses 7, 8, 11 y 12 o violines más pequeños con un cuerpo más redondo como los meses 1, 2, 9 y 10. Los meses de enero, febrero y septiembre son violines cortos ya que estos meses son de alta actividad promocional, es decir, las utilidades bajan debido a promociones de descuentos o meses sin intereses, también se puede relacionar a que las ventas dentro de esos meses no son tan altas. En cambio, en meses como julio, agosto y septiembre los violines son alargados ya que las utilidades dentro de estos meses son mayores, en este grupo de artículo se puede relacionar a que son meses previos al ingreso a clases.

En la *Figura 34* se puede observar el comportamiento de las utilidades del grupo de artículo Sandalia durante el 2019, antes de la pandemia de Covid-19. Aquí se puede observar que los meses con más utilidades fueron marzo, abril y julio. Esto se puede deber a las altas ventas por la temporada vacacional de los alumnos pequeños.

En las figuras previamente explicadas, se puede observar que en diversos meses, existe una utilidad negativa. Esto se debe a que se le aplican grandes descuentos para poder darle salida a la mercancía que se ha rezagado en temporadas anteriores. En cuanto a los tenis o sandalias, suele tener mercancía rezagada de tallas que no son tan frecuentes entre los clientes.

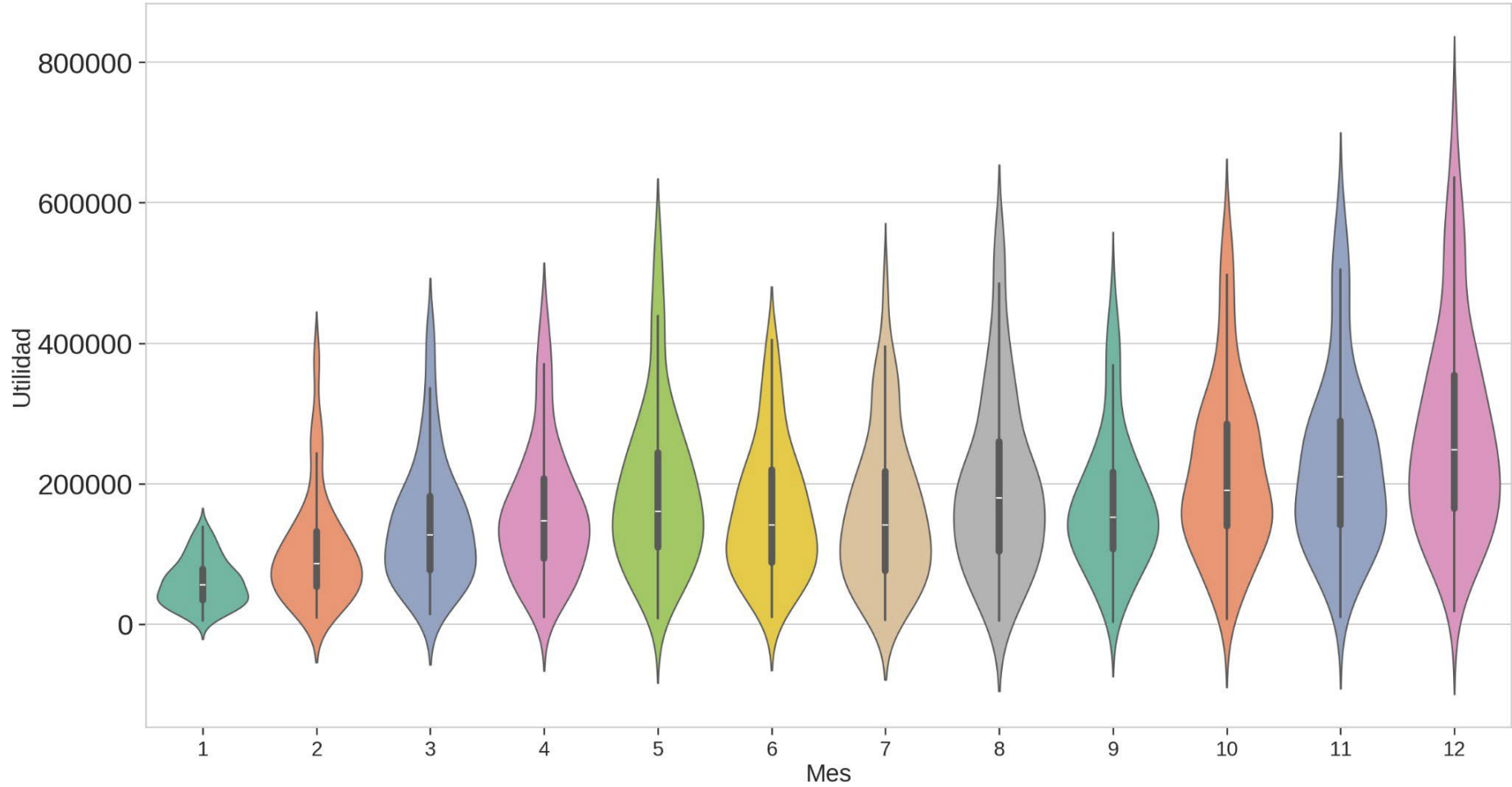


Figura 13- Gráfica de violín Utilidad Tennis 2021

Los resultados muestran que el mes de enero es el que menos utilidades presenta, mientras que el mes de diciembre es el que más utilidad tiene.

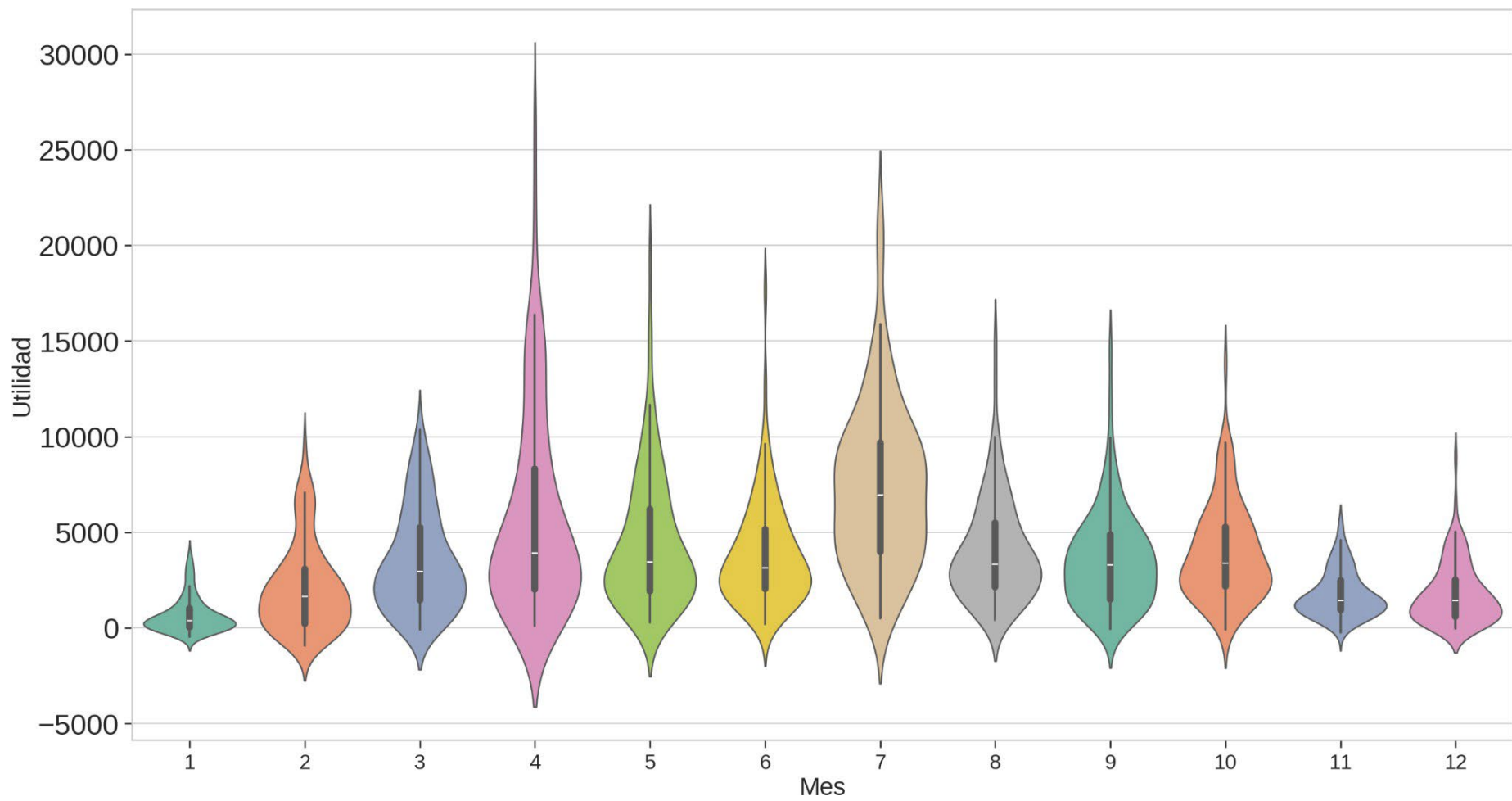


Figura 14- Gráfica de violín Utilidad Sandalia 2021

Se puede observar que los meses con menos utilidades son enero y noviembre, mientras que el mes que más sobresale es el de abril con mayores utilidades que los demás.

Mapas coropléticos

En las *Figuras 15, 16, 35 y 36* se presentan mapas coropléticos que nos permiten observar cómo se encuentra la república mexicana. Se puede observar que las *Figuras 15 y 35* representan la distribución de las utilidades, pero llegan a ser diferentes. La diferencia entre dichas figuras es que la *Figura 15* muestra la totalidad de las utilidades que se presentan por estado. En el anexo 1 se puede observar el desglose de tiendas por estado. Por eso en esta figura se puede ver una gran diferencia entre los estados que tienen más tiendas y los que no. A diferencia de la *Figura 15*, la *Figura 35* representa el promedio de las utilidades por estado. Este gráfico nos explica que no necesariamente los estados que cuentan con más tiendas son aquellos que tienen en promedio más utilidades. Se puede observar en la *Figura 15* que el estado de Hidalgo no representa muchas utilidades a la cadena de tiendas, pero sí se llega a ver por promedio, es uno de los estados que aporta más utilidades por el número de tiendas que tiene actualmente.

En la *Figura 16* se puede observar el mapa, pero en este se resalta la suma de las ventas por el estado. Se observa que los estados que más ventas tienen en promedio son Jalisco, Estado de México y la Ciudad de México.

La *Figura 36* muestra el promedio de los descuentos por estado. Se observa que los estados que más descuentos realizan son la ciudad de México e Hidalgo en promedio, seguidos por los estados de Jalisco y el Estado de México.

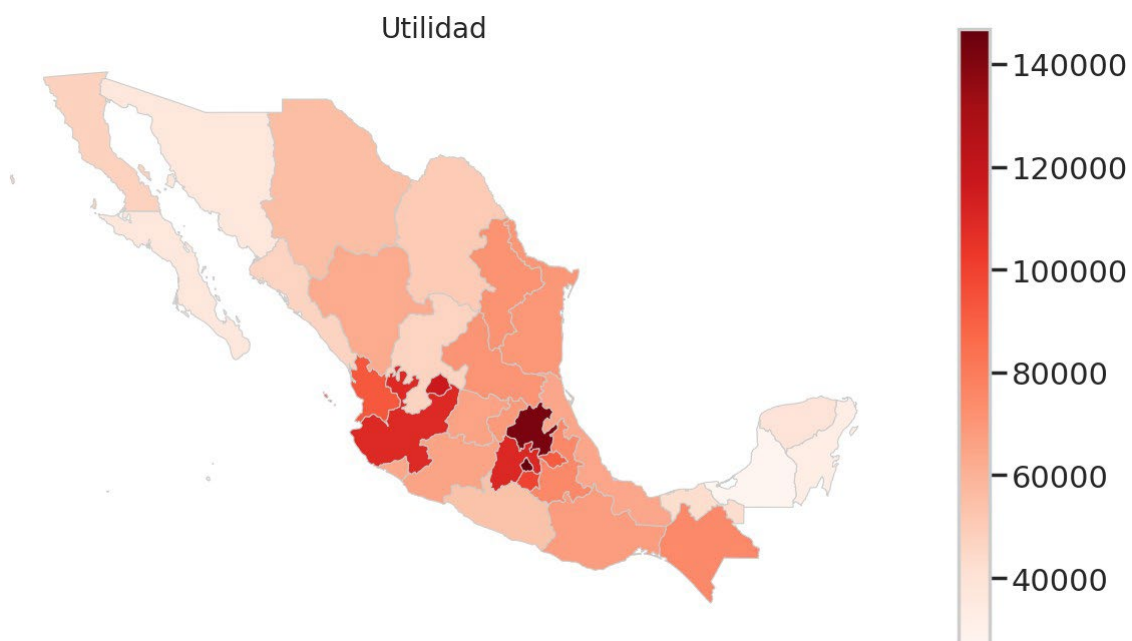


Figura 15 - Mapa coroplético promedio de utilidades por estado

Se observa que los estados con más utilidades en promedio son la ciudad de México e Hidalgo



Figura 16 - Mapa coroplético suma ventas por estado

Se observa que los estados que más ventas tienen en promedio son Jalisco, Estado de México y la Ciudad de México

PCA

En las *Figuras 17, 18, 37 y 38* se muestra el análisis principal de los datos (PCA) donde podemos ver el comportamiento de todas las variables en un plano tridimensional. Lo que muestra es una escala de colores que va desde el color más oscuro que representa el mes de enero hasta el color más brillante que representa el mes de diciembre. Este set de gráficos ayudará a distinguir los distintos patrones de los datos. Los datos están presentados en 4 vistas, la vista isométrica, superior, frontal y derecha, y lo que se busca es representar las combinaciones posibles (x,y), (x,z) y (y,z) con el objetivo de identificar los distintos patrones con más precisión.

- *Tenis 2021 (Figura 17)*: Se puede observar un comportamiento más alargado de las variables en comparación con las del *Tenis 2019* donde los datos están más concentrados. En este caso se puede ver en ciertas vistas que el grupo de datos más oscuro está concentrado mientras que el resto de los datos se encuentran dispersos.
- *Sandalia 2021 (Figura 18)*: Se puede observar un comportamiento distinto a las demás secciones, aquí se puede ver una especie de cuneta. Los meses de enero a marzo se encuentran de cierta forma concentrados de manera seccionada, mientras los demás meses se encuentran reburujado
- *Tenis 2019 (Figura 37)*: En este set de gráficos se pueden observar en las cuatro vistas que los puntos representando al mes de enero, los puntos representando el mes de septiembre y los puntos entre marzo y junio se encuentran concentrados junto a los demás de su color, mientras que los demás meses como son agosto a diciembre se encuentran revueltos en el centro de los gráficos.
- *Sandalia 2019 (Figura 38)*: Se puede observar un comportamiento disperso en las vistas isométricas, (Y,Z) y (X,Y) donde los puntos más oscuros se encuentran concentrados mayormente en un sector del gráfico, mientras que los demás meses se encuentran dispersos y revueltos entre sí.

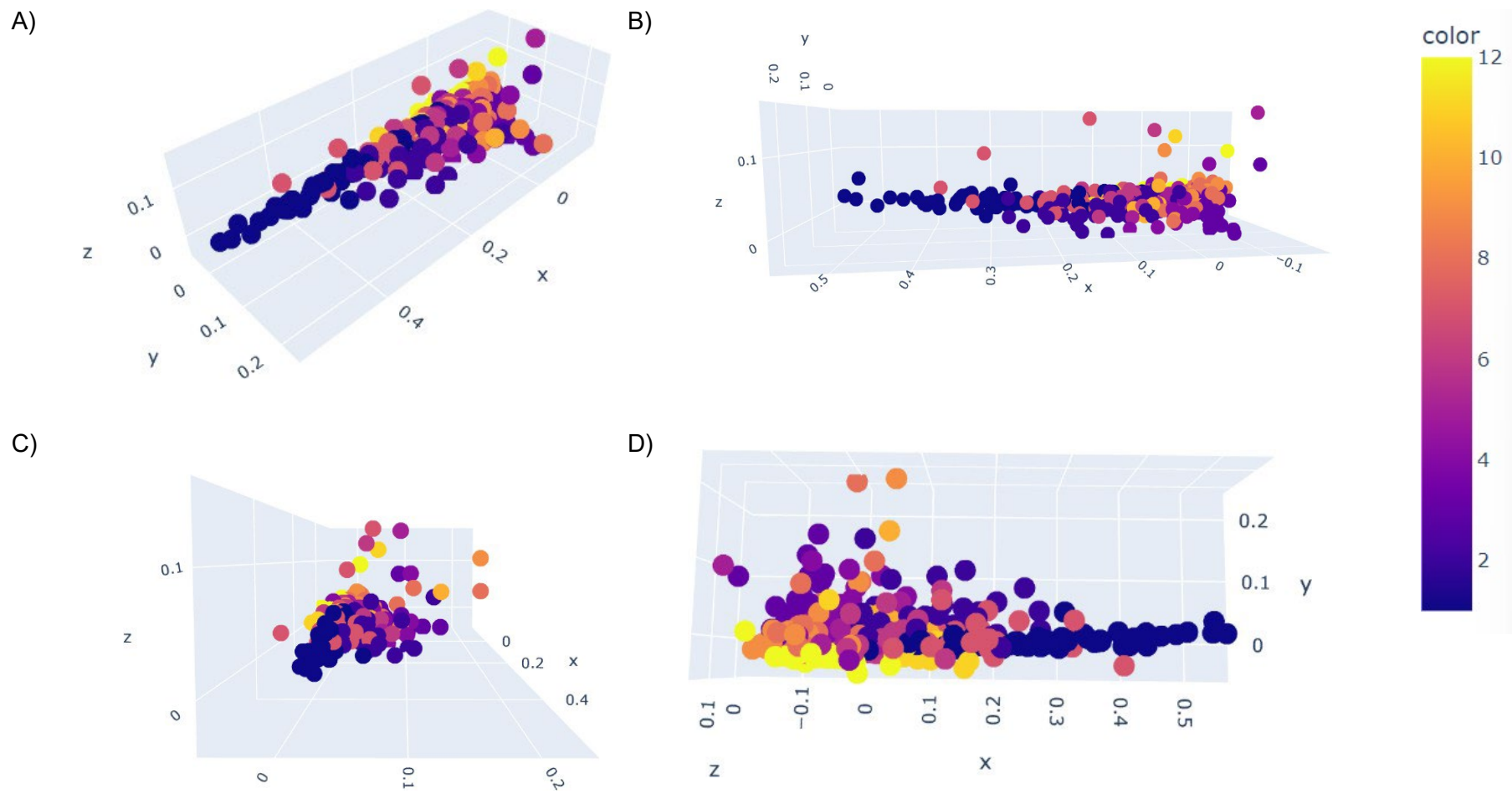
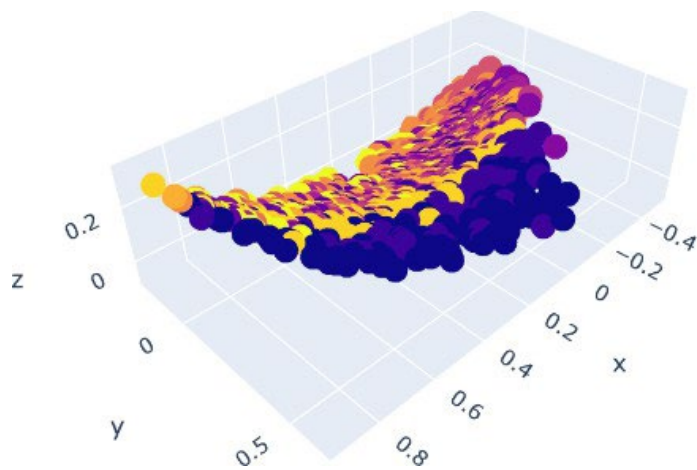
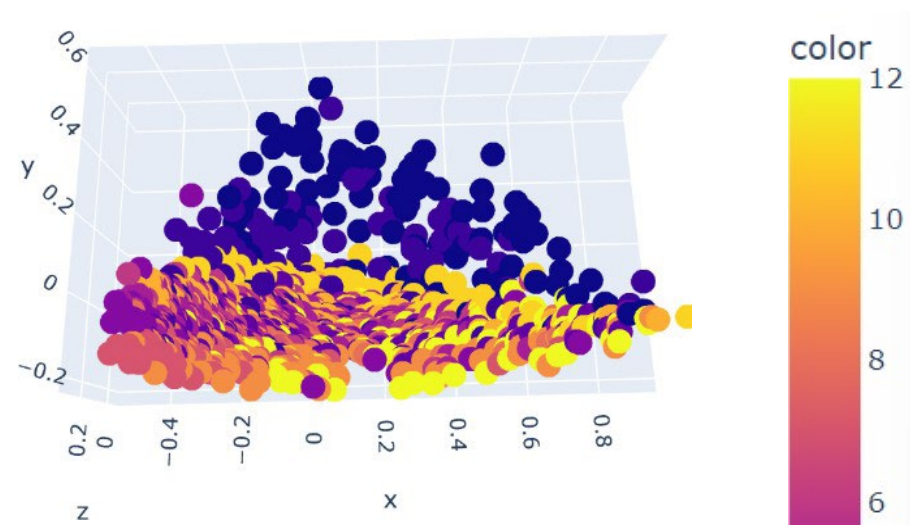


Figura 17 - Vistas de PCA Tenis 2021 A) Vista isométrica, B) Vista superior, C) Vista lateral, D) Vista frontal

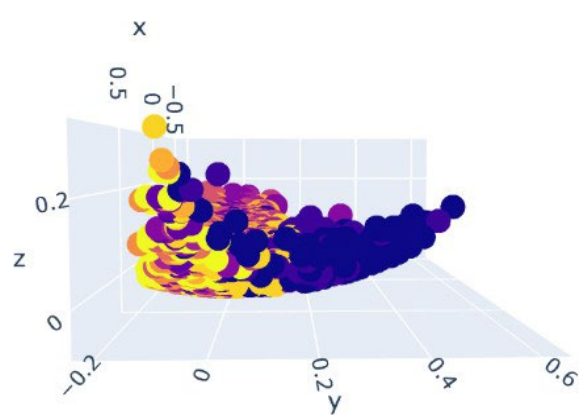
A)



B)



C)



D)

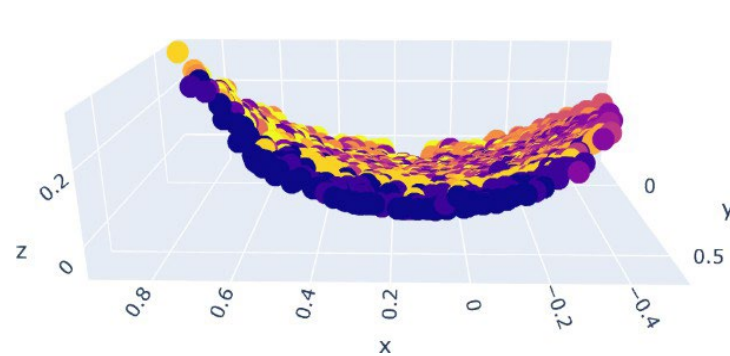


Figura 18 - Vistas de PCA Sandalia 2021 A) Vista isométrica, B) Vista superior, C) Vista lateral, D) Vista frontal

Clusterización

Dendrogramas

En los siguientes dendrogramas, *Figura 19, 20, 39 y 40*, se muestra una manera de clusterizar los datos. En el eje X se pueden observar los centros y en el eje Y se puede observar el grado de similitud, entre más cercano al eje X más similitud existe. Se puede observar que se encuentran cuatro dendrogramas, uno para cada set de datos que fueron previamente filtrados. Las *Figuras 19, 39 y 40* muestran una similitud en la estructura, a diferencia la *Figura 20* que muestra una estructura diferente.

Analizando las *Figuras 19, 39 y 40* como ya se mencionó, cuentan con una estructura similar. Los conglomerados de dichos dendrogramas están acomodados de cierta manera en la que los de la izquierda son aquellos que forman parte de las tiendas con mejor rendimiento, es decir, que se pueden considerar como las mejores tiendas. Se puede observar que estos dendrogramas cuentan con varios conglomerados de tiendas, pero solo 5 son los que pueden ser definidos como clústeres que son los que definen de forma general los datos ingresados.

Analizando la *Figura 20*, se observa un comportamiento diferente a los demás dendrogramas. Se pueden ver claramente los 5 conglomerados principales que construyen los diferentes clústeres. Aquellos conglomerados que cuentan con menos tiendas se pueden considerar como aquellas tiendas que representan mejores resultados a lo largo del año, es decir, las mejores tiendas.

Estos dendrogramas permiten corroborar visualmente el resultado que nos muestra el algoritmo mediante las diferentes pruebas que se realizan para determinar el número óptimo de k que lleva cada sección.

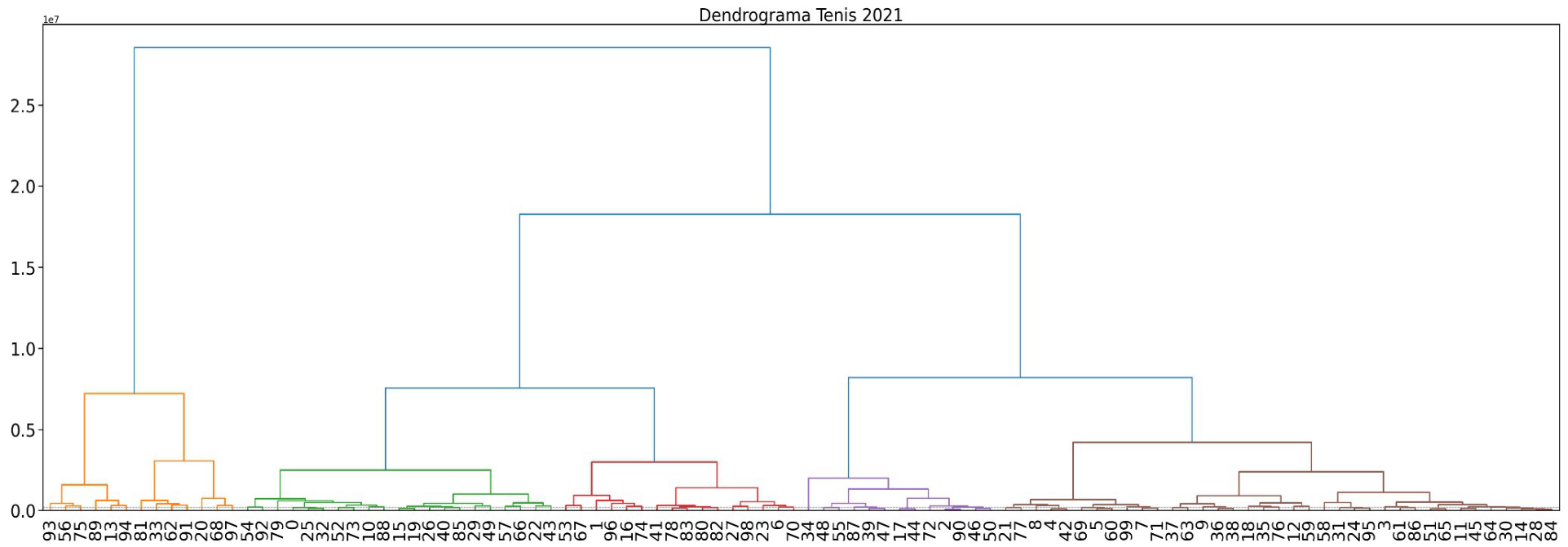


Figura 19- Dendrograma Tenis 2021

Se observa la división en 5 dendrogramas, los cuales son representados por diferentes colores. Cada cluster tiene una cantidad menor de tiendas, siendo el color amarillo y el morado los que aparentan ser los clusters más pequeños y el café siendo el cluster más grande.

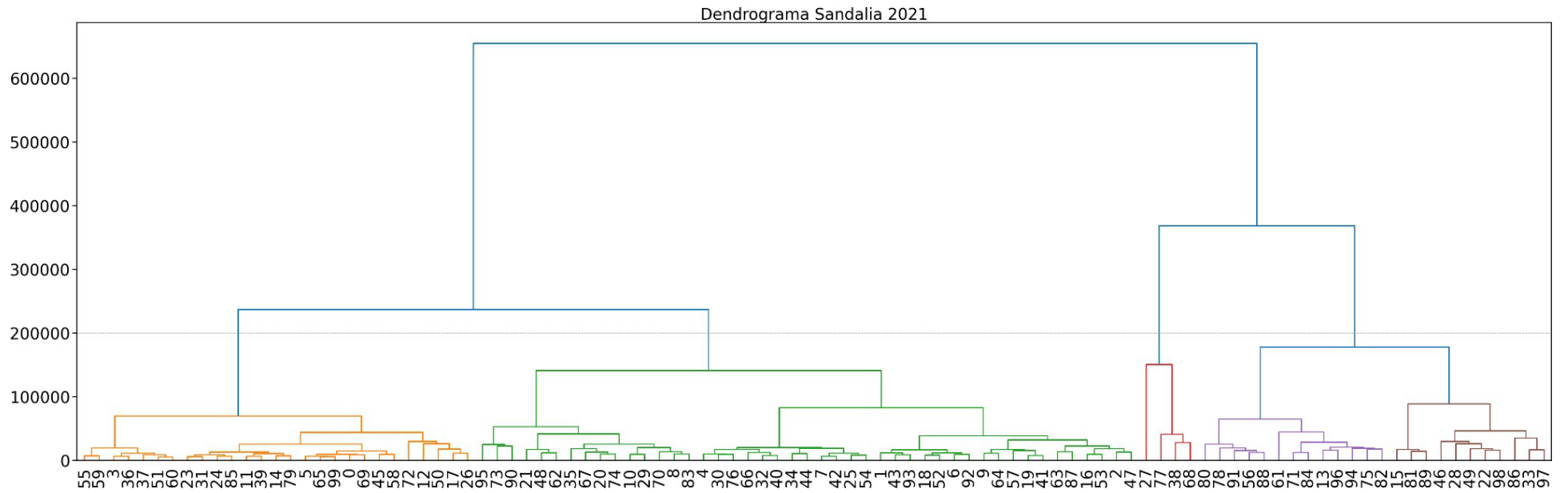


Figura 20- Dendrograma Sandalia 2021

Se observa la división en 5 dendrogramas, los cuales son representados por diferentes colores. Cada cluster tiene una cantidad distinta de tiendas, siendo el color rojo el que cuenta con solo 4 tiendas y el verde siendo el cluster más grande.

En las Figuras 21 a la 30 y de la Figura 41 a 48, se puede observar el método de clusterización de los distintos segmentos de la base de datos original. Se ven las gráficas para determinar el número óptimo de clústeres, se observan gráficas de la distribución de los datos y varias gráficas donde se ven las distintas características de cada clúster dentro del set de datos ingresado.

Método del codo

Para las Figuras 21,22,41 y 42 se utilizó un algoritmo que grafica el número de clústeres con la sumatoria de las distancias. En las diferentes gráficas para cada sección se puede ver que el número óptimo de “k” es 5, este número respectivo de “k” se obtuvo al calcular mediante un algoritmo la puntuación de la distorsión. En esa posición es donde cambia la pendiente radicalmente entre las diferentes sumatorias, es decir, el número óptimo de clusters para cada sección es aquel donde la pendiente tiene un cambio brusco a comparación de las distintas distancias entre “k”.

Tenis 2021

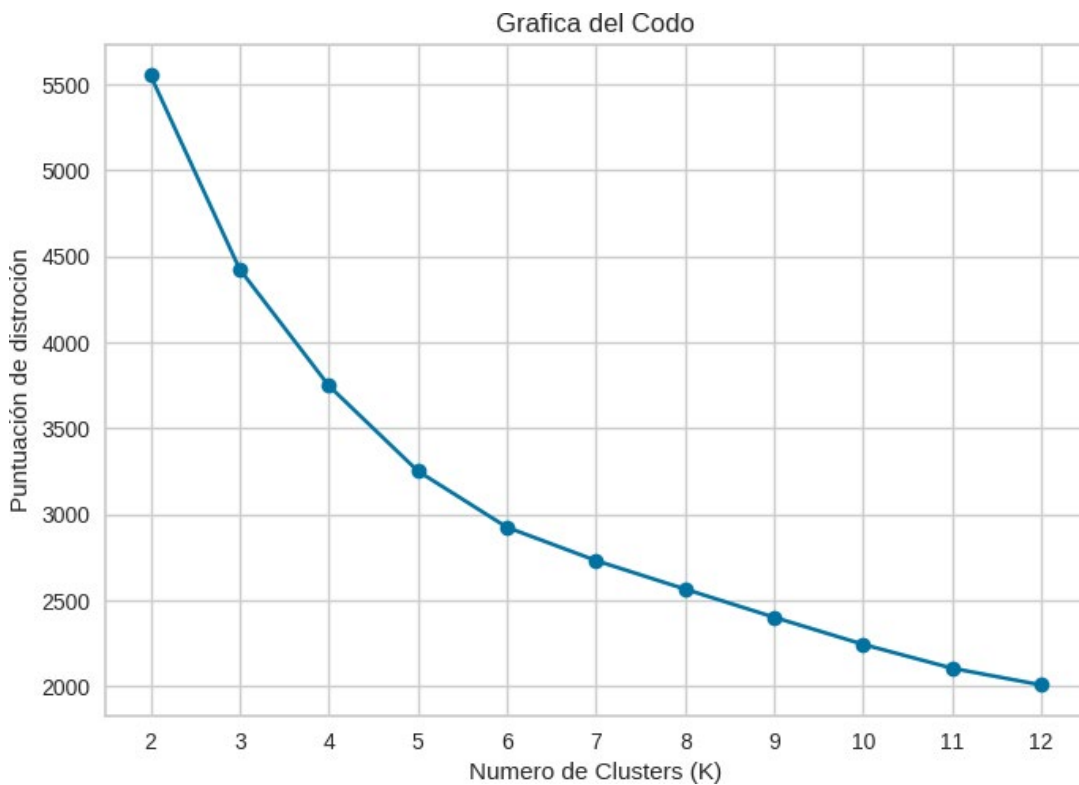


Figura 21- Método del codo Tenis 2021

Se observa que el codo se encuentra en la K #5, la puntuación de dicha K es de 3,251.88

Sandalia 2021

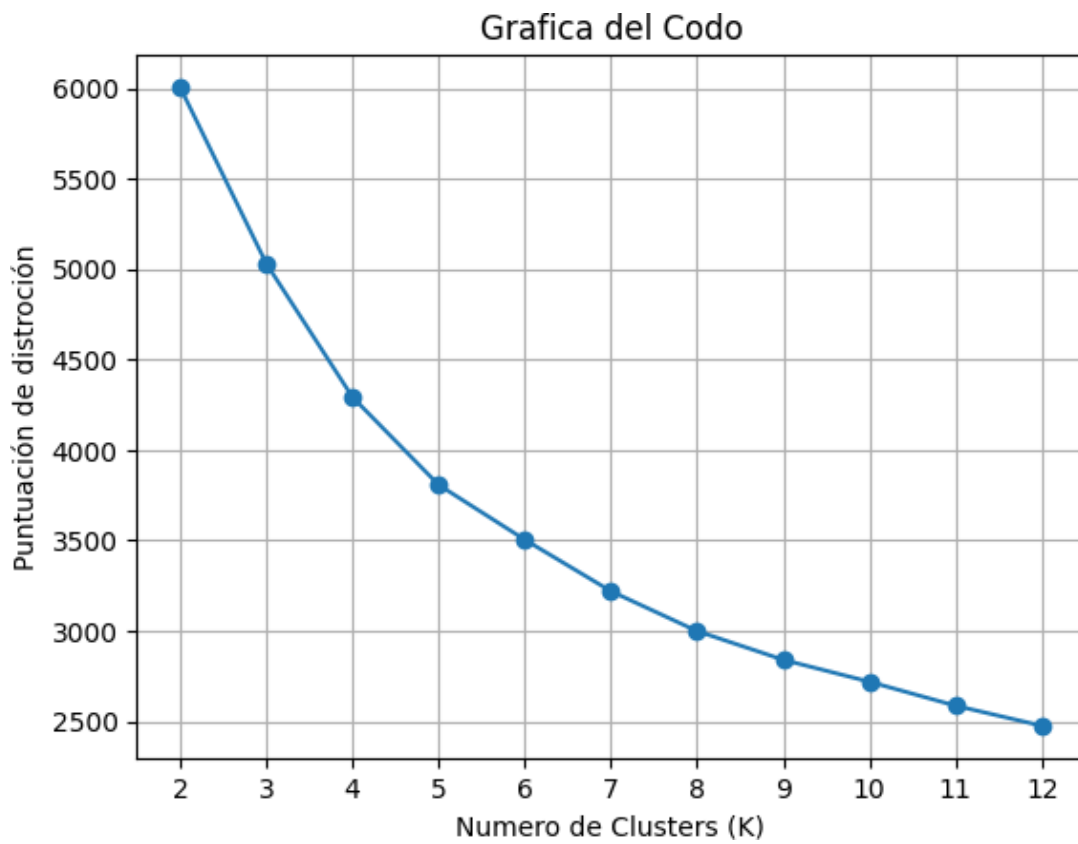


Figura 22- Método del codo Sandalia 2021

Se observa que el codo se encuentra en la K #5, la puntuación de dicha K es de 3,807.5

Visualización

La *Figura 23* muestra la visualización de los clústeres utilizando la metodología t-SNE (t-distributed Stochastic Neighbor Embedding). Se ven las distribuciones de los clústeres en dos dimensiones, donde podemos observar la magnitud de cada clúster y su comportamiento.

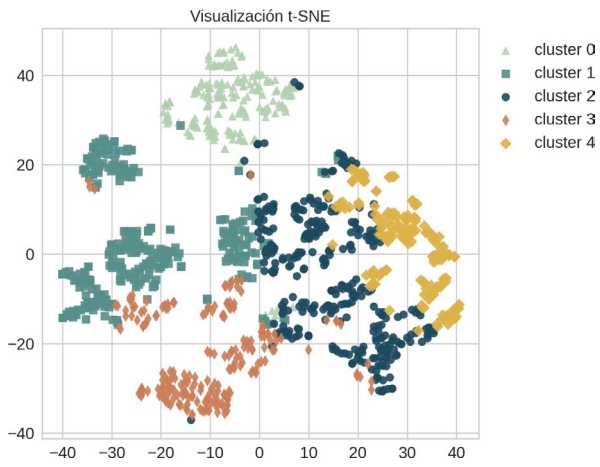
En el caso del inciso A, el cual muestra los clusters de tenis 2019, se puede observar que el cluster con menos magnitud es el cero, mientras que el cluster con más magnitud es el 2. En este también se puede observar la dispersión de los puntos, siendo el cluster 1 el que menos dispersión tiene y el cluster 2 cuenta con los puntos más dispersos.

Para el caso del inciso B, se puede observar los cluster de tenis 2021. Se puede ver que el cluster más pequeño a simple vista es el número 2, mientras que el más grande es el número 1. Se puede observar que el cluster 4 cuenta con puntos compactos, no tiene mucha dispersión como lo es en el cluster 0.

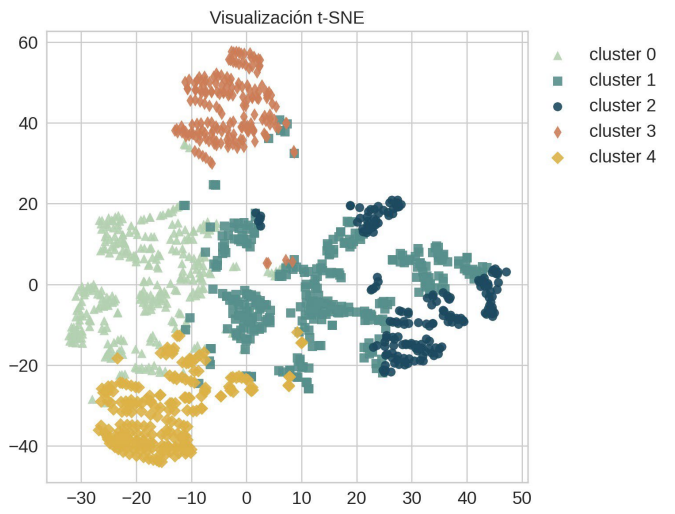
En el inciso C se puede observar el comportamiento de los clusters de sandalia 2019. En este se pueden observar clusters más dispersos y con más magnitud. En cuanto a los clusters 0 y 2 se les puede observar con más magnitud y alta dispersión. En el caso contrario los clusters con menos magnitud y más compactos se pueden observar que son el número 1 y número 3.

En cuanto al inciso D se puede observar el comportamiento de sandalia 2021. En este se puede observar que los clusters 1, 3 y 4 cuentan con una magnitud alta. Mientras que el cluster 2 es el que menos magnitud tiene y es el más compacto.

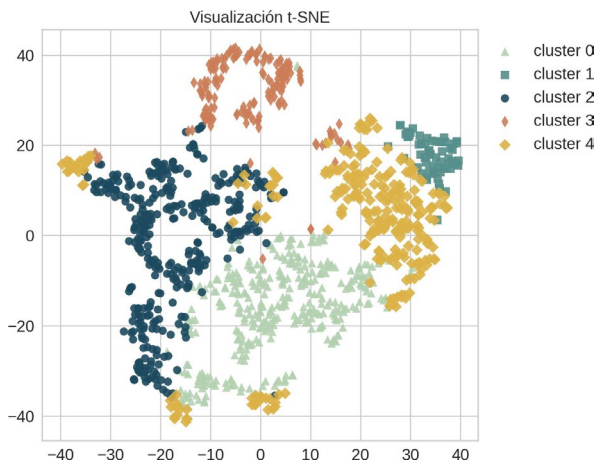
A)



B)



C)



D)

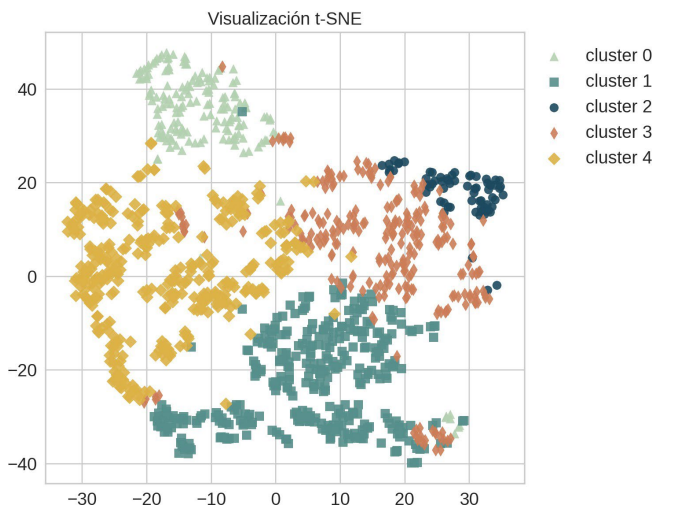


Figura 23- Visualización t-SNE, A) Tenis 2019, B) Tenis 2021, C) Sandalia 2019, D) Sandalia 2021

Características del clúster

En las siguientes figuras se pueden observar las características de los distintos clústeres resultantes del algoritmo. Se muestra la cardinalidad y la magnitud de los distintos clústeres y su relación entre sí.

En la *Figura 24* de la sección *Tenis 2021*, solo un clúster tiene una alta correlación entre magnitud y cardinalidad. El cluster 2 es aquel que se encuentra alineado sobre la diagonal. Se puede observar que los clústeres con más cardinalidad son el 0 y el 1, mientras los que le siguen son el 4 luego el 3 y al final el 2. En cuanto a magnitud, el clúster más grande es el 1, mientras que el siguiente es el clúster número 2 y por debajo se encuentra el clúster 0 con alrededor de 300 puntos, el clúster 3 y 4 cuentan con una magnitud similar.

La *Figura 25* no muestra clusters con alta correlación entre la magnitud y la cardinalidad y se vuelve la sección con menos clústeres alineados, los puntos más cercanos a la diagonal son los clusters 3 y 4. Se puede observar que la cardinalidad de los clústeres 1, 3 y 4 es alta en comparación con los clústeres 0 y 2, con una magnitud que tiene un comportamiento similar a la gráfica de la cardinalidad, el clúster 0 y 2 se encuentran por debajo de los clústeres 1, 3 y 4.

En la *Figura 43* de la sección de *Tenis 2019* se puede observar que el clúster cero tiene una alta correlación al estar sobre la diagonal entre la magnitud y la cardinalidad. Se puede observar que el clúster cero tiene una magnitud más alta que su cardinalidad, el clúster 1 tiene una cardinalidad arriba de los 300 puntos, mientras que en magnitud tiene arriba de 400 puntos. El clúster 2 es aquel con más magnitud y cardinalidad. El clúster 3 tiene una cardinalidad arriba de los 200 puntos y una magnitud abajo de los 400 puntos y es el clúster que se posiciona en el tercer lugar en cuanto a magnitud y el clúster 4, que es el más pequeño en magnitud y cardinalidad, tiene una magnitud similar al clúster 3 y 0.

En la *Figura 44*, se muestra la sección de *sandalia 2019*. Se muestra el cluster 4 con alta correlación entre la magnitud y la cardinalidad. Se puede observar que en cuanto a su cardinalidad el clúster 0 es el mayor, seguido por el clúster 2 con alrededor 330 puntos y le sigue por debajo el clúster 4 con 275 puntos aproximadamente. En cuanto a la magnitud, el clúster 4 es el más grande con alrededor 500 puntos, seguido de los clústeres 0 y 2 con alrededor de 430 puntos.

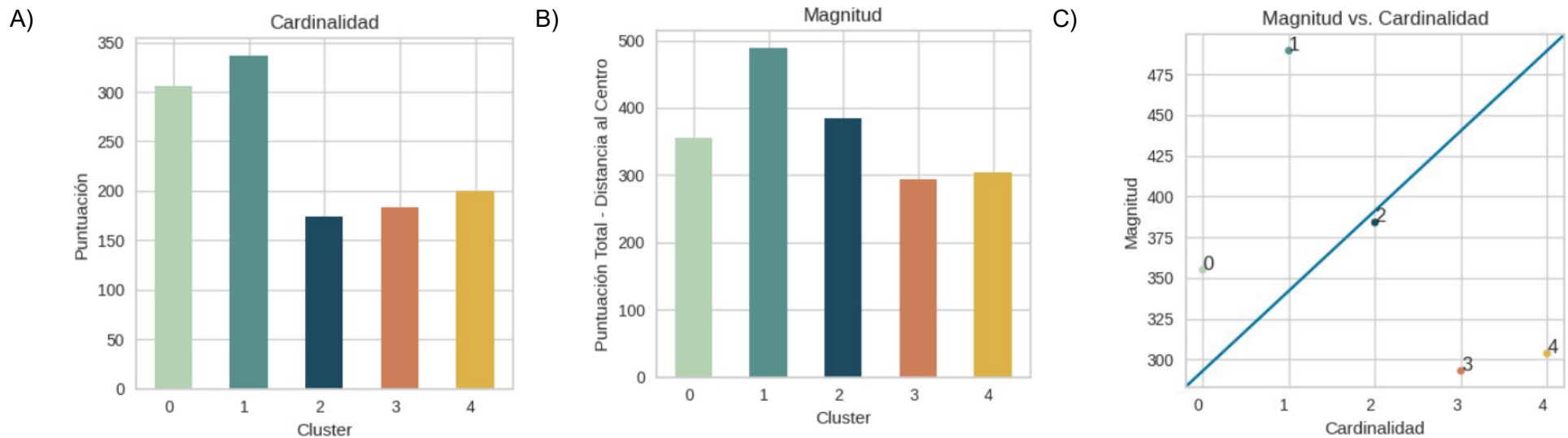


Figura 24- Cardinalidad y magnitud clústeres Tenis 2021, A) Cardinalidad, B) Magnitud, C) Cardinalidad vs. magnitud,

Se puede observar que los clústeres con más cardinalidad son el 0 y el 1, mientras los que le siguen son el 4, luego el 3, y al final el 2. En cuanto a magnitud, el clúster más grande es el 1, mientras que el siguiente es el clúster número 2, y por debajo se encuentra el clúster 0 con alrededor de 300 puntos, el clúster 3 y 4 cuentan con una magnitud similar. En cuanto a la correlación, el cluster 2 se encuentra sobre la diagonal.

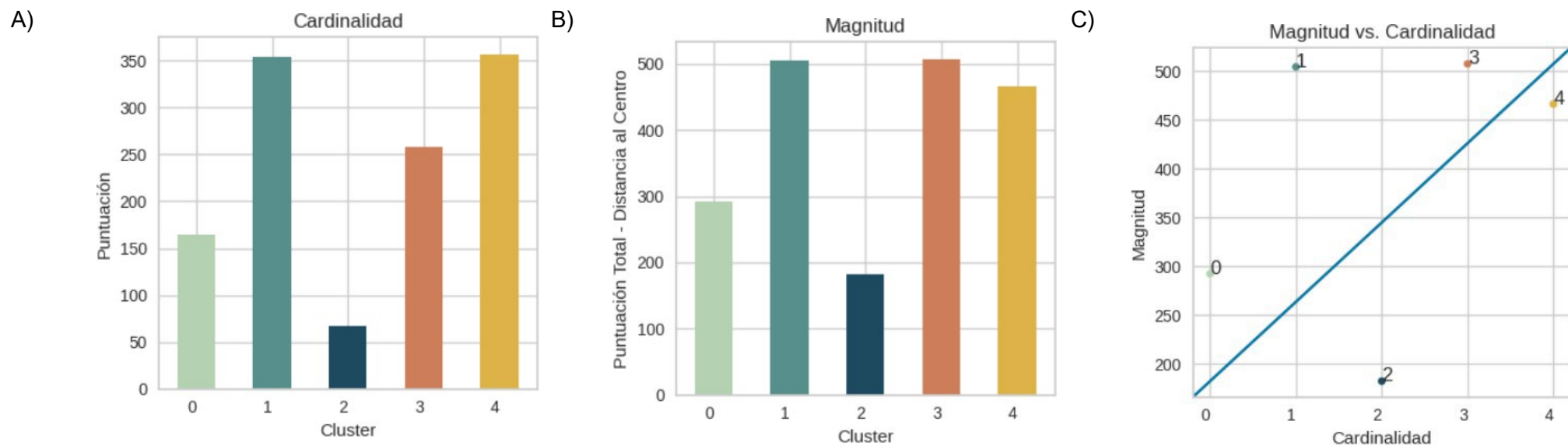


Figura 25- Cardinalidad y magnitud clústeres Sandalia 2021, A) Cardinalidad, B) Magnitud, C) Cardinalidad vs. magnitud

Se puede observar que la cardinalidad de los clústeres 1, 3 y 4 es alta en comparación con los clústeres 0 y 2, con una magnitud que tiene un comportamiento similar a la gráfica de la cardinalidad, el clúster 0 y 2 se encuentran por debajo de los clústeres 1, 3 y 4. Ningún cluster cuenta con una correlación alineada a la diagonal.

En las *Figuras 26, 27, 45 y 46* tenemos gráficos de cajas y bigotes, que nos muestran cuál es la relevancia de las distintas variables dentro de cada clúster

- Para la sección de *Tenis 2021 (Figura 26)* se pueden observar que los clusters 3 y 4 tienen presencia en prácticamente todo el año, mientras que los clusters 1 y 2 cuentan con presencia en casi todos los meses, exceptuando el mes de enero. Mientras que el cluster 0 se encuentra de enero a julio. Son clusters los cuales tienen diferente categoría en las ventas, utilidades, descuento y MSI. Se puede observar que el clúster 0 cuenta con centros con una latitud entre 17° y 22° y una longitud entre -96° y -105° , con ventas principalmente entre 25 y 37 millones de pesos y utilidades entre \$75,000 y \$150,000. Los descuentos y MSI (meses sin intereses) son pequeños. El clúster 1 cuenta con centros dentro de las latitudes 17° y 22° y longitudes entre -96° y -105° , con ventas entre 25 y 77 millones de pesos, utilidades entre \$90,000 y \$375,000, es un clúster que cuenta con centros con actividad promocional alta entre \$50,000 y \$250,000 y MSI de \$2,000 y \$30,000. El clúster 3, igual que el clúster 2, los centros están en latitudes entre 19° y 22° y longitudes entre -96° y -105° . Este clúster cuenta con ventas entre 75 millones y 150 millones de pesos, utilidades entre \$250,000 y \$600,000, actividad promocional entre \$100,000 y \$450,000 y MSI entre \$10,000 y \$50,000. El clúster 3 cuenta con tiendas entre las latitudes 22° y 26° y longitudes entre -105° y -113° , mientras que el clúster 4 cuenta con centros entre latitudes 16° y 21° y longitudes entre -85° y -100° . Ambos tienen ventas entre 0 y 60 millones de pesos y utilidades entre 0 y 350,000 pesos. Sus descuentos están entre 0 y 175,000 pesos y MSI entre 0 y 15,000 pesos.
- Para la sección de *Sandalia 2021 (Figura 27)*, se puede observar más dinamismo en los clusters, solo el cluster 0 y 4 cuentan con presencia todo el año, mientras que el cluster 1 tiene presencia en el primer semestre del año, el cluster 2 entre los meses de febrero y octubre y el 4 con presencia durante el segundo semestre. En cuanto a las ventas, utilidades, descuentos y MSI se puede observar que el cluster 0, 1 y 4 son aquellos que se encuentran bajos, mientras que el cluster 2 es aquel que se encuentra sobresaliendo con números altos. Mientras, el cluster 3 se encuentra entre esos dos segmentos. El clúster 0 los centros se encuentran principalmente entre la ubicación geográfica de 21° y 29° de latitud y longitud de -98° y -105° , cuentan con ventas máximas de \$20,000, utilidades máximas de \$5,000, descuentos máximos de \$4,000 y meses sin intereses de máximo \$400. El clúster 1, sus centros se encuentran entre los 16° y los 23° de latitud y -90° y -105° de longitud, con ventas máximas de \$15,000, utilidades máximas de \$5,200, descuentos máximos de \$6,000 y meses sin intereses máximos de \$250. En cuanto al clúster 2, sus centros están entre los 16° y 23° de latitud y los -93° y -105° de longitud. Tiene ventas máximas de \$50,000, utilidades máximas de \$20,000, descuentos de \$13,000 y meses sin intereses de \$1,000. En el clúster 3, los centros se encuentran entre los 16° y 24° de latitud y entre los -90° y -107° de longitud. Las ventas máximas son de \$20,000, utilidades máximas de \$13,000, descuentos de \$6,000 y \$600 en meses sin intereses. Para finalizar, el clúster 4 tiene presencia durante los meses de julio y diciembre. Los centros se encuentran entre los 16° y 23° de latitud y entre -90° y -106° de longitud. Ventas de \$10,000 como máximo, utilidades entre 0 y \$5,500, descuentos máximos de \$3,750 y meses sin intereses de \$200

- Para la sección de Tenis 2019 (*Figura 45*) se puede observar el cluster 0, 3 y 4 tienen presencia en todo el año, mientras que el cluster 1 tiene presencia en los meses de enero a julio y el cluster 2 de marzo a diciembre, la mayoría del año. En este, se puede observar, el cluster 4 es el que sobresale en Ventas, Descuentos, MSI y Utilidades a comparación de los demás clusters que se encuentran en rangos similares. Se puede observar que el clúster 0, un clúster que cuenta con tiendas entre las latitudes 22° a 26° y longitudes entre -97° y -113°. Son tiendas que tienen ventas entre los \$100,000 y \$600,000, con utilidades entre los \$50,000 y los \$200,000. Estas tiendas tuvieron descuentos y costo de meses sin intereses entre \$100,000 y \$170,000. El clúster 1, son tiendas principalmente centros ubicados entre las latitudes de 17° y 22.5° y longitudes entre los -93° y -106°. Son centros que tuvieron ventas entre los \$100,000 y \$600,000 y utilidades entre \$50,000 y \$275,000 y que tuvieron actividad promocional de descuentos y meses sin intereses hasta \$250,000. El clúster 2 está conformado por centros ubicados entre las latitudes de 17° y 22.5° y longitudes entre los -93° y -106°. Con ventas entre los \$250,000 y los \$750,000, utilidades entre \$75,000 y \$300,000, actividad promocional en descuentos de \$50,000 a \$300,000 y tienen un costo de meses sin intereses de hasta \$35,000. El clúster 3, son centros con ubicaciones en latitudes entre 14° y 23°, longitudes entre los -85° y los -103°, con ventas de hasta \$500,000 y utilidades de hasta \$200,000. La actividad promocional en descuentos es hasta \$150,000 y meses sin intereses hasta \$15,000. El clúster 4, son centros con ubicaciones entre las latitudes de 19° y 22° y longitudes entre los -95° y -105° con ventas entre \$500,000 y \$1,750,000, utilidades entre los \$200,000 y los \$700,000. Fueron centros con actividad promocional entre los \$75,000 y los \$500,000 y meses sin intereses entre los \$7,500 y \$50,000.
- Para la sección de Sandalia 2019 (*Figura 46*) se puede observar que el cluster 3 y 4 tienen presencia los 12 meses del año, mientras que el cluster 0 tiene presencia en 7 meses del año, el cluster 1 tiene presencia en 9 meses y el cluster 2 tiene presencia en el primer semestre del año. En cuanto al dinamismo en las variables de Ventas, Utilidades, Descuentos y MSI, se puede observar que el cluster 1 cuenta con mayor número a comparación de los demás, le sigue el cluster 4 con mayores números que los clusters 0, 2 y 3 que cuentan con números similares en dichas variables. Se puede observar el comportamiento de los clústeres de la siguiente manera. El clúster 0 son centros con latitudes entre los 16° y los 24°, longitudes entre los -85° y los -105°, con ventas hasta los \$20,000 y utilidades máximas de \$5,000. Estos centros tienen descuentos máximos de \$4,000 y meses sin intereses de un máximo de \$500. El clúster 1 tiene presencia en centros ubicados en las latitudes de 16° a 26° y longitudes de -95° a -108°, con ventas entre \$20,000 y \$65,000 y utilidades entre \$5,000 y \$30,000. Los descuentos están entre \$2,000 y \$10,000 y el costo de meses sin intereses entre \$100 y \$2,000. El clúster 2 se puede observar en cuanto a la latitud y longitud se encuentran entre los 17° y 23° y -93° y los -107°, respectivamente. Son centros con ventas máximas de \$15,000, utilidades máximas de \$6,000, descuentos máximos de \$4,500 y meses sin intereses de \$250. El clúster 3 en cuanto a latitud y longitud se encuentran entre los 22° y los 26° y entre -97° y -112°, respectivamente. Tiene ventas máximas de \$20,000, utilidades de \$7,500, descuentos de \$4,000 y meses sin intereses de \$250. El clúster 4, igual que el clúster 3, tiene presencia durante todos los meses, pero en

este caso principalmente entre los meses de mayo y agosto. Su latitud y longitud es entre 18° y los 22° y entre los -85° y los -105° , respectivamente, ventas máximas de \$30,000, utilidades de \$12,000, descuentos de \$8,000 y meses sin intereses de \$500.

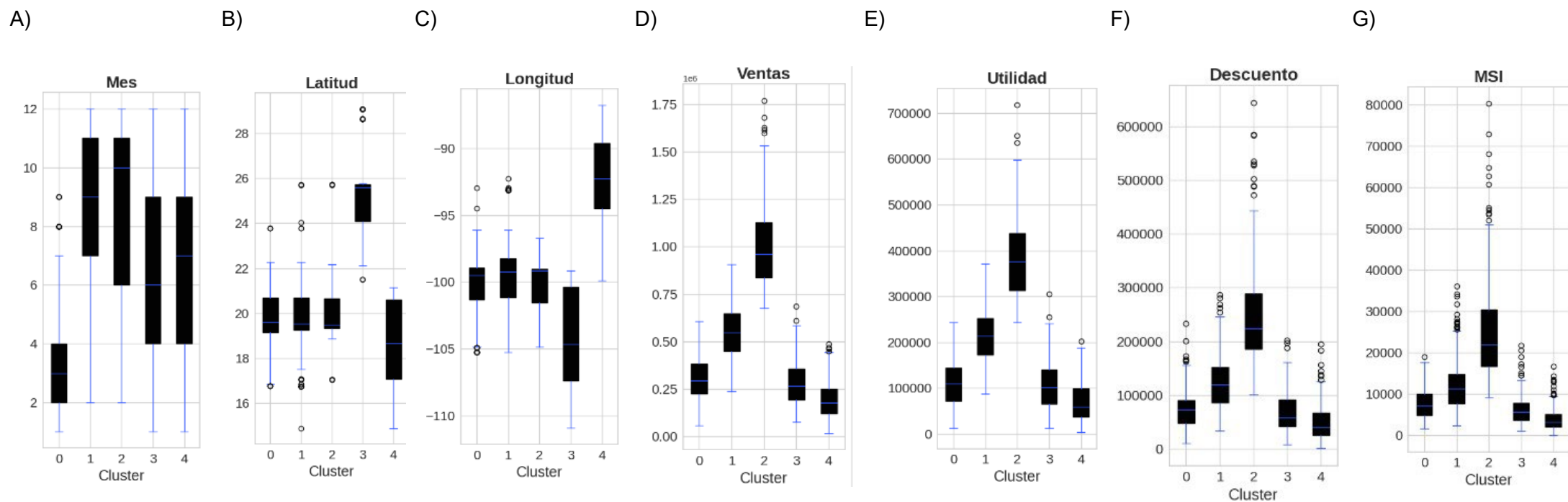


Figura 26- Distribuciones de característica por clúster Tenis 2021, A) Gráfica de cajas y bigotes Mes, B) Gráfica de cajas y bigotes Latitud, C) Gráfica de cajas y bigotes Longitud, D) Gráfica de cajas y bigotes mes Ventas, E) Gráfica de cajas y bigotes Utilidad, F) Gráfica de cajas y bigotes Descuento, G) Gráfica de cajas y bigotes MSI

Se puede observar que el clúster 0 tiene presencia principalmente en los meses de febrero hasta abril con poca relevancia entre meses como enero, mayo, junio y julio. Cuenta con centros con una latitud entre 17° y 22° y una longitud entre -96° y -105°, con ventas principalmente entre 25 y 37 millones de pesos y utilidades entre \$75,000 y \$150,000. Los descuentos y MSI (meses sin intereses) son pequeños. El clúster 1 tiene presencia en todos los meses del año excepto enero. Cuenta con centros dentro de las latitudes 17° y 22° y longitudes entre -96° y -105°, con ventas entre 25 y 77 millones de pesos, utilidades entre \$90,000 y \$375,000, es un clúster que cuenta con centros con actividad promocional alta entre \$50,000 y \$250,000 y MSI de \$2,000 y \$30,000. El clúster 3, igual que el clúster 2, cuenta con meses principalmente de junio a octubre y tiene presencia en todos los meses excepto enero. Los centros están en latitudes entre 19° y 22° y longitudes entre -96° y -105°. Este clúster cuenta con ventas entre 75 millones y 150 millones de pesos, utilidades entre \$250,000 y \$600,000, actividad promocional entre \$100,000 y \$450,000 y MSI entre \$10,000 y \$50,000. El clúster 3 y 4 cuentan con meses en todo el año y principalmente entre abril y septiembre, el clúster 3 cuenta con tiendas entre las latitudes 22° y 26° y longitudes entre -105° y -113°, mientras que el clúster 4 cuenta con centros entre latitudes 16° y 21° y longitudes entre -85° y -100°. Ambos tienen ventas entre 0 y 60 millones de pesos y utilidades entre 0 y 350,000 pesos. Sus descuentos están entre 0 y 175,000 pesos y MSI entre 0 y 15,000 pesos.

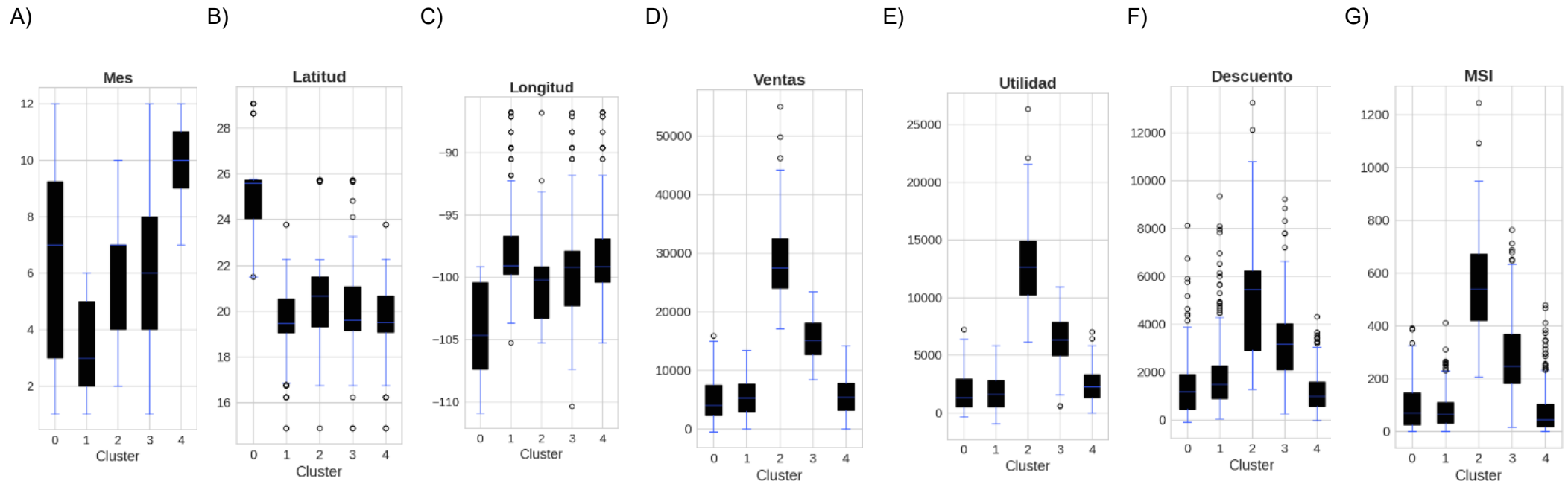


Figura 27- Distribuciones de característica por clúster Sandalia 2021, A) Gráfica de cajas y bigotes Mes, B) Gráfica de cajas y bigotes Latitud, C) Gráfica de cajas y bigotes Longitud, D) Gráfica de cajas y bigotes mes Ventas, E) Gráfica de cajas y bigotes Utilidad, F) Gráfica de cajas y bigotes Descuento, G) Gráfica de cajas y bigotes MSI

El clúster 0 tiene presencia durante todo el año principalmente entre marzo y septiembre. Sus centros se encuentran principalmente entre la ubicación geográfica de 21° y 29° de latitud y longitud de -98° y -105°, cuentan con ventas máximas de \$20,000, utilidades máximas de \$5,000, descuentos máximos de \$4,000 y meses sin intereses de máximo \$400. El clúster 1 tiene presencia en el primer semestre del año. Sus centros se encuentran entre los 16° y los 23° de latitud y -90° y -105° de longitud, con ventas máximas de \$15,000, utilidades máximas de \$5,200, descuentos máximos de \$6,000 y meses sin intereses máximos de \$250. En cuanto al clúster 3 tiene presencia entre febrero y octubre, sus centros están entre los 16° y 23° de latitud y los -93° y -105° de longitud. Tiene ventas máximas de \$50,000, utilidades máximas de \$20,000, descuentos de \$13,000 y meses sin intereses de \$1,000. El clúster 4 tiene presencia durante todo el año igual que el clúster 0, los centros se encuentran entre los 16° y 24° de latitud y entre los -90° y -107° de longitud. Las ventas máximas son de \$20,000, utilidades máximas de \$13,000, descuentos de \$6,000 y \$600 en meses sin intereses. Para finalizar, el clúster 4 tiene presencia durante los meses de julio y diciembre. Los centros se encuentran en entre los 16° y 23° de latitud y entre -90° y -106° de longitud. Ventas de \$10,000 como máximo, utilidades entre 0 y \$5,500, descuentos máximos de \$3,750 y meses sin intereses de \$200.

Capítulo 4. Análisis o discusión de resultados

Chen et al. (2020) mencionan que la inteligencia artificial es una gran herramienta que aporta beneficios de tiempo de procesamiento, exactitud y consistencia. Mencionan que encuesta realizada a un instituto de la fuerza laboral en el año 2018, determinó que esta es una gran herramienta que eleva la productividad. No obstante, es una herramienta que genera duda e incertidumbre entre los trabajadores. Para muchos trabajadores, la IA es algo nuevo y conocen muy poco del tema; piensan que dicha herramienta les quitará su puesto de trabajo (una pregunta que me he planteado en ciertos procesos). Lo que los trabajadores tienen que entender es que la inteligencia artificial está aquí para ser parte del equipo, para ayudarles y no para reemplazarlos. Para este trabajo se consideran las 4 fases por las que tienen que pasar las empresas para implementar la inteligencia artificial dentro de su fuerza laboral propuestas por Chen (Chen et al., 2020). Estas se podrán observar en la *Figura 28*.



Figura 28- Fases para implementar la inteligencia artificial

La inteligencia artificial es un reto para las empresas, pero esto no es algo nuevo, siempre ha sido difícil adoptar y adaptarse a cosas nuevas y poco conocidas. La tecnología avanza todos los días, por eso es importante que las empresas implementen las herramientas de una manera relajada, es decir, que sea clara en cuanto a transparencia, autonomía y privacidad. La inteligencia artificial, si se introduce correctamente, respetando las 4 fases, se volverá un gran activo para la compañía ya que le permitirá procesar con mayor rapidez grandes volúmenes de información variada de una manera consistente.

El algoritmo del presente trabajo se propone como parte de la fase 2, es un algoritmo que retroalimenta al usuario proporcionándole un clúster preciso acorde a los datos ingresados; es como un monitor que le puede dar recomendaciones sobre qué clúster puede ser un centro y en qué meses es deseable tenerlo como una mejor o peor tienda dependiendo de su rendimiento. En esta fase el algoritmo no toma decisiones por sí solo, como en las etapas 3 y 4 pero tampoco es un algoritmo simple como en la fase 1 que apenas se le enseña a formar su propio criterio.

Gráficas de Radar

En las *Figuras 29, 30, 47 y 48* se muestra la relevancia de las variables (Latitud, Longitud, Mes, MSI, Descuento, Utilidad y Ventas) dentro de los cinco clústeres presentados. Esta gráfica muestra que variable es la más destacada dentro del cluster. Se podrán observar los diferentes clusters de diferente color para poder identificarlos. Se puede observar en las Figuras que los clusters tienen distintos comportamientos.

En la *Figura 29* se muestran los clusters para los datos de Tenis 2021. En el cluster 0 las variables de Longitud, Ventas, Utilidad, Latitud, Descuento y MSI son muy similares, destacan casi de la misma manera, mientras que la variable de mes no es tan relevante. En el cluster 1, la variable que más destaca es la de Mes, mientras que la variable de Ventas, Utilidad, Descuento y MSI. El cluster 2 La que más destaca es la de Ventas, mientras que la que menos es la Latitud. En el cluster 3 se puede observar que la variable más destacada es la de la Latitud de las tiendas, mientras que la menos destacada es la variable de Longitud. Para el cluster 4, una variable que destaca a comparación de las demás, en este caso es la Longitud, mientras que la Latitud es la variable que menos destaca.

La *Figura 30* nos muestra la gráfica de los datos de Sandalia 2021. Se observa lo siguiente: el cluster 0 tiene como variable destacada la de Latitud, mientras que la longitud no ha destacado. La variable que más destaca en el cluster 1 es la Longitud, mientras que la variable más destacada es el Mes. El cluster 2 tiene una similitud entre las variables de Ventas y Utilidad, ambas destacan lo mismo, mientras que la longitud no destaca. La variable más destacada en el cluster 3 es la de Ventas, muestra que la que menos destaca es la Latitud. En el cluster 4, la variable más destacada es el mes, mientras que la menos destacada es la Latitud.

En la *Figura 47* se puede observar se puede observar que los distintos clusters para los datos de Tenis 2019. En el cluster 0 se puede observar que la variable más destacada es la de la Latitud de las tiendas, mientras que la menos destacada es la variable de Longitud. En el cluster 1, las variables de Longitud, Ventas, Utilidad, Latitud, Descuento y MSI destacan aproximadamente igual, mientras que la variable de mes no es tan relevante. El cluster 2 la variable que más destaca es la de Mes, mientras que la variable de Ventas, Utilidad, Descuento y MSI. El cluster 3 muestra al igual que el cluster 0 una variable que

destaca a comparación de las demás, en este caso es la Longitud, mientras que la Latitud es la variable que menos destaca. Para el cluster 4, la mayoría de las variables destacan. La que más destaca es la de Ventas, mientras que la que menos es la Latitud.

En la *Figura 48* se puede observar se puede observar que los distintos clusteres para los datos de Sandalias 2019. En el cluster 0 se puede observar que la variable más destacada es la de la Longitud y mes de las tiendas, mientras que la menos destacada es la variable de Latitud. En el cluster 1, las variables de Ventas, Utilidad y MSI. El cluster 2 la variable que más destaca es la longitud, mientras que la variable de Ventas, Utilidad, Descuento y MSI tienen una representación similar. En el cluster 3 la variable más destacada es la de la latitud, mientras que la Longitud es la variable que menos destaca. Para el cluster 4, la mayoría de las variables destacan. La que más destaca es la de Ventas.

Estas gráficas lo que nos muestran es que variables son las que más influyeron en las características de cluster. Lo que muestra es que como se puede observar, las variables que más destacan son aquellas las cuales fueron el criterio del cluster, es decir, si la variable más destacada es la latitud, las tiendas de este cluster son de latitudes similares. A comparación de clusteres donde destacan la variable de ventas, las tiendas de este cluster son aquellas que tuvieron ventas similares.

Características del conglomerado: Medias de características por conglomerado

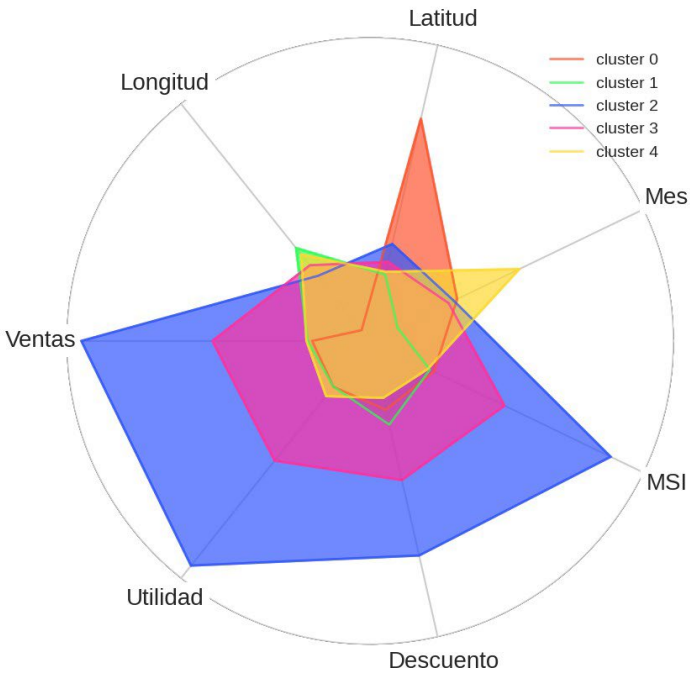


Figura 29- Radar Tennis 2021

Características del conglomerado: Medias de características por conglomerado

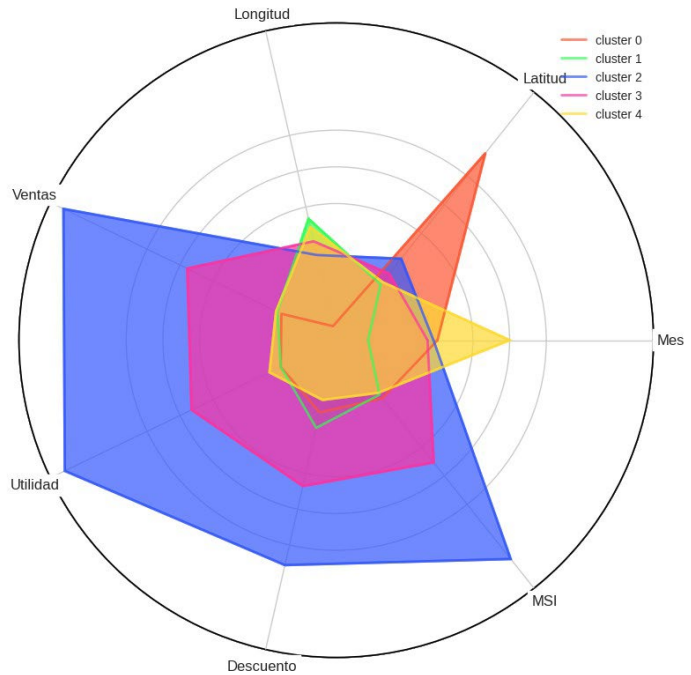


Figura 30- Radar Sandalia 2021

Comparación contra metodología actual

Las Tablas 7 a 12 muestran los resultados que arroja el algoritmo para los grupos de artículos seleccionados. En estas Tablas se pueden ver el centro, su zona, el mes del año, el resultado para los datos de 2021 y 2019 presentado por el algoritmo, el resultado de una clusterización manual hecha por planeadores de Liverpool y la rotación mensual acumulada de cada centro en los años de los datos seleccionados.

Para hacer la comparación se les pidió a 4 distintos planeadores (planners) de la dirección de infantiles que hicieran su división de clusterización con los datos presentados de 2021, se les proporcionó la misma base de datos que se usó en el algoritmo. Para llegar al resultado, cada planeador lo hizo a su manera. Se puede observar que los planeadores utilizan una misma segmentación por año, a comparación del algoritmo que divide la segmentación de la tienda por mes.

Para la comparativa entre la presente técnica y la metodología actual, se analizaron la totalidad de las tiendas. Sin embargo, para efectos prácticos del presente trabajo, se decidió utilizar una muestra de 3 distintos tipos de tienda. De acuerdo con la experiencia que se cuenta en este punto, se seleccionaron tiendas que históricamente muestran un comportamiento de ventas a) Excelente, b) Estándar - Promedio, c) Bajo. Se solicitó a los planeadores calificar estas tiendas utilizando datos de 2019 y 2021 para los 12 meses del año. Finalmente se realizó la clasificación utilizando el presente algoritmo.

Uno de los indicadores comerciales con mayor valor es la rotación del inventario a lo largo de las secciones y los distintos puntos de venta. En las tablas se puede observar que entre más cercano a la AA este el centro mayor rotación se tiene durante el año, mientras que los centros más cercanos a D tienden a tener una menor rotación. Esto sucede ya que el piso de venta para cada sección es finito, por lo que no se puede tener un inventario que cubra las altas ventas que tiene el centro, por lo que constantemente se le tienen que mandar mercancía para que pueda satisfacer las ventas. En el caso contrario, se debe tener cierta mercancía en el piso de venta para que se vea un piso lleno, si no, no es atractivo. En ciertos casos las ventas no logran llegar al inventario a la misma velocidad que el caso contrario.

Se puede observar que el algoritmo no presenta el mismo clúster durante todo el año, esto se debe a que los puntos de venta no tienen el mismo rendimiento todo el año. Tener una clusterización por mes lograr que la profundidad y modelaje de mercancía que se le manda al punto de venta sea de la mejor manera para poder satisfacer al cliente sin sacrificar el índice de rotación del punto de venta. Los resultados completos al detalle se pueden observar en el Anexo 02 y Anexo 04.

Datos				Algoritmo		Planners (Datos 2021)				Rotación Acumulada Anual	
Centro#	Centro	Zona#	Mes	2021	2019	1	2	3	4	2019	2021
0003	Liverpool POLANCO	1	1	B	B	B	A	B	B	0.21	0.02
0003	Liverpool POLANCO	1	2	B	B	B	A	B	B	0.35	0.19
0003	Liverpool POLANCO	1	3	B	B	B	A	B	B	0.51	0.35
0003	Liverpool POLANCO	1	4	A	A	B	A	B	B	0.68	0.57
0003	Liverpool POLANCO	1	5	A	AA	B	A	B	B	0.93	0.81
0003	Liverpool POLANCO	1	6	A	AA	B	A	B	B	1.18	1.03
0003	Liverpool POLANCO	1	7	A	AA	B	A	B	B	1.47	1.27
0003	Liverpool POLANCO	1	8	A	AA	B	A	B	B	1.90	1.56
0003	Liverpool POLANCO	1	9	A	A	B	A	B	B	2.09	1.78
0003	Liverpool POLANCO	1	10	AA	A	B	A	B	B	2.22	2.12
0003	Liverpool POLANCO	1	11	AA	AA	B	A	B	B	2.45	2.51
0003	Liverpool POLANCO	1	12	AA	AA	B	A	B	B	2.76	2.91

Tabla 07- Resultado escenario 1 Tenis

Datos				Algoritmo		Planners (Datos 2021)				Rotación Acumulada Anual	
Centro#	Centro	Zona#	Mes	2021	2019	1	2	3	4	2019	2021
0028	Liverpool VERACRUZ MOCAMBO	5	1	B	B	D	D	D	D	0.22	0.24
0028	Liverpool VERACRUZ MOCAMBO	5	2	B	D	D	D	D	D	0.30	0.39
0028	Liverpool VERACRUZ MOCAMBO	5	3	B	D	D	D	D	D	0.45	0.55
0028	Liverpool VERACRUZ MOCAMBO	5	4	D	D	D	D	D	D	0.58	0.71
0028	Liverpool VERACRUZ MOCAMBO	5	5	D	D	D	D	D	D	0.78	0.88
0028	Liverpool VERACRUZ MOCAMBO	5	6	D	D	D	D	D	D	0.93	1.04
0028	Liverpool VERACRUZ MOCAMBO	5	7	D	D	D	D	D	D	1.04	1.23
0028	Liverpool VERACRUZ MOCAMBO	5	8	D	D	D	D	D	D	1.25	1.40
0028	Liverpool VERACRUZ MOCAMBO	5	9	D	D	D	D	D	D	1.35	1.58
0028	Liverpool VERACRUZ MOCAMBO	5	10	D	D	D	D	D	D	1.42	1.88
0028	Liverpool VERACRUZ MOCAMBO	5	11	D	D	D	D	D	D	1.58	2.13
0028	Liverpool VERACRUZ MOCAMBO	5	12	D	D	D	D	D	D	1.78	2.32

Tabla 08 - Resultado escenario 2 Tenis

Datos				Algoritmo		Planners (Datos 2021)				Rotación Acumulada Anual	
Centro#	Centro	Zona#	Mes	2021	2019	1	2	3	4	2019	2021
0189	Liverpool MONTERREY LA FE	6	1	C	C	B	B	B	B	0.24	0.11
0189	Liverpool MONTERREY LA FE	6	2	C	C	B	B	B	B	0.41	0.26
0189	Liverpool MONTERREY LA FE	6	3	C	C	B	B	B	B	0.67	0.44
0189	Liverpool MONTERREY LA FE	6	4	C	C	B	B	B	B	0.85	0.64
0189	Liverpool MONTERREY LA FE	6	5	C	C	B	B	B	B	1.09	0.88
0189	Liverpool MONTERREY LA FE	6	6	C	C	B	B	B	B	1.26	1.09
0189	Liverpool MONTERREY LA FE	6	7	C	A	B	B	B	B	1.53	1.32
0189	Liverpool MONTERREY LA FE	6	8	C	AA	B	B	B	B	1.95	1.50
0189	Liverpool MONTERREY LA FE	6	9	C	C	B	B	B	B	2.09	1.68
0189	Liverpool MONTERREY LA FE	6	10	AA	C	B	B	B	B	2.22	2.10
0189	Liverpool MONTERREY LA FE	6	11	AA	A	B	B	B	B	2.44	2.52
0189	Liverpool MONTERREY LA FE	6	12	A	A	B	B	B	B	2.79	2.90

Tabla 09 - Resultado escenario 3 Tenis

Datos				Algoritmo		Planners (Datos 2021)				Rotación Acumulada Anual	
Centro #	Centro	Zona#	Mes	2021	2019	1	2	3	4	2019	2021
0117	Liverpool MAZATLÁN	4	1	B	C	C	B	C	C	0.07	0.28
0117	Liverpool MAZATLÁN	4	2	B	C	C	B	C	C	0.20	0.62
0117	Liverpool MAZATLÁN	4	3	B	C	C	B	C	C	0.42	1.19
0117	Liverpool MAZATLÁN	4	4	B	A	C	B	C	C	0.69	1.70
0117	Liverpool MAZATLÁN	4	5	B	A	C	B	C	C	0.84	1.98
0117	Liverpool MAZATLÁN	4	6	B	A	C	B	C	C	0.94	2.41
0117	Liverpool MAZATLÁN	4	7	A	AA	C	B	C	C	1.13	2.93
0117	Liverpool MAZATLÁN	4	8	B	A	C	B	C	C	1.46	3.07
0117	Liverpool MAZATLÁN	4	9	B	A	C	B	C	C	1.69	3.73
0117	Liverpool MAZATLÁN	4	10	A	C	C	B	C	C	1.90	4.14
0117	Liverpool MAZATLÁN	4	11	B	C	C	B	C	C	2.08	4.25
0117	Liverpool MAZATLÁN	4	12	B	C	C	B	C	C	2.22	4.50

Tabla 10 - Resultado escenario 1 Sandalia

Datos				Algoritmo		Planners (Datos 2021)				Rotación Acumulada Anual	
Centro#	Centro	Zona#	Mes	2021	2019	1	2	3	4	2019	2021
0002	Liverpool INSURGENTES	1	1	D	A	A	AA	AA	AA	0.25	0.11
0002	Liverpool INSURGENTES	1	2	D	D	A	AA	AA	AA	0.34	0.29
0002	Liverpool INSURGENTES	1	3	A	AA	A	AA	AA	AA	0.73	0.60
0002	Liverpool INSURGENTES	1	4	AA	AA	A	AA	AA	AA	1.32	1.08
0002	Liverpool INSURGENTES	1	5	AA	AA	A	AA	AA	AA	1.65	1.57
0002	Liverpool INSURGENTES	1	6	A	A	A	AA	AA	AA	1.78	2.01
0002	Liverpool INSURGENTES	1	7	AA	AA	A	AA	AA	AA	2.21	2.17
0002	Liverpool INSURGENTES	1	8	A	A	A	AA	AA	AA	2.37	2.56
0002	Liverpool INSURGENTES	1	9	A	A	A	AA	AA	AA	2.45	2.91
0002	Liverpool INSURGENTES	1	10	AA	A	A	AA	AA	AA	2.66	3.33
0002	Liverpool INSURGENTES	1	11	A	A	A	AA	AA	AA	2.92	3.83
0002	Liverpool INSURGENTES	1	12	C	B	A	AA	AA	AA	3.19	4.16

Tabla 11- Resultado escenario 2 Sandalia

Datos				Algoritmo		Planners (Datos 2021)				Rotación Acumulada Anual	
Centro#	Centro	Zona#	Mes	2021	2019	1	2	3	4	2019	2021
0031	Liverpool MONTERREY	6	1	B	C	B	A	A	A	0.02	0.41
0031	Liverpool MONTERREY	6	2	B	C	B	A	A	A	0.17	0.93
0031	Liverpool MONTERREY	6	3	B	C	B	A	A	A	0.30	1.07
0031	Liverpool MONTERREY	6	4	AA	AA	B	A	A	A	0.71	1.66
0031	Liverpool MONTERREY	6	5	A	A	B	A	A	A	0.94	2.12
0031	Liverpool MONTERREY	6	6	B	AA	B	A	A	A	1.23	3.02
0031	Liverpool MONTERREY	6	7	AA	AA	B	A	A	A	1.60	3.44
0031	Liverpool MONTERREY	6	8	B	A	B	A	A	A	1.76	4.00
0031	Liverpool MONTERREY	6	9	B	C	B	A	A	A	1.82	4.36
0031	Liverpool MONTERREY	6	10	B	C	B	A	A	A	1.89	4.90
0031	Liverpool MONTERREY	6	11	B	C	B	A	A	A	1.96	5.14
0031	Liverpool MONTERREY	6	12	B	C	B	A	A	A	2.02	5.14

Tabla 12 - Resultado escenario 3 Sandalia

Capítulo 5. Resultados y conclusiones

Hoy en día, determinar la demanda que puede tener un punto de venta en un momento específico es prácticamente imposible. Por esta razón, herramientas como las presentadas anteriormente ayudan a perfilar de una mejor manera los puntos de venta, lo que permite optimizar los resultados. El algoritmo proporciona recomendaciones como un monitor, indicando la calificación que debería tener un punto de venta en un momento dado, pero no toma la decisión final; eso depende del usuario (el planificador de la demanda).

La clusterización de los distintos puntos de venta apoyará al planeador para que coloque la mercancía de una mejor manera en las tiendas, en cuanto a modelaje y profundidad. Tener la mercancía bien colocada en los centros correctos aporta beneficios a la sección como es no tener un exceso de inventario y que se le pueda ofrecer al cliente lo que busca, así como que las tiendas tengan una rotación de inventario saludable y que la mercancía no se estanque en los puntos de venta y se tenga que rematar o que se tenga que transferir a otra tienda que haya vendido mejor el producto, lo cual acaba teniendo un impacto no considerado en el costo logístico.

El cliente de la cadena no es el mismo para todos los puntos de venta; depende mucho la ubicación y demografía. Como se puede observar en las *Figuras 38 a la 41*, la diferenciación de los clústeres está influenciada por la longitud y latitud en la que se encuentran los centros. Por esta razón no todos los clústeres tienen la misma presencia en todas las zonas de la cadena. El objetivo del trabajo, determinar una metodología homogénea para obtener la clusterización de los puntos de venta de la cadena minorista. Obtener un resultado que muestre la verdadera clasificación con diferenciación por mes y así poder tener una distribución más precisa acorde a los resultados. Se puede considerar de una manera teórica que lo obtenido generará un mejor resultado para las secciones que repliquen la metodología para poder tener un mejor resultado en las tiendas. De esta manera se tendrá más precisión en lo que se tiene que colocar en las tiendas para que el cliente encuentre lo que necesita en los diferentes meses del año.

La clusterización de las tiendas permite al planeador darse cuenta de que a lo largo del año los diferentes puntos de venta no cuentan con el mismo rendimiento en todos los meses. Con la metodología presentada en este trabajo, se puede definir un proceso homogéneo para determinar la clasificación sin tener que depender totalmente del criterio humano. Los resultados presentados para la segmentación de los puntos de venta demuestran que el rendimiento de los centros puede mejorar durante el año al ser más preciso en el envío de la mercancía que busca el cliente. Al implementar nuevas metodologías que apoyen al usuario a tomar decisiones, se obtendrán mejores colocando la mercancía adecuada en el lugar adecuado. Esto favorecerá al aumento de las utilidades ya que se reduciría la cantidad de mercancía estancada; entre menos mercancía estancada haya en los diferentes puntos de venta menos descuentos se tendrán que aplicar. Se puede concluir que lo presentado en este trabajo cumple con los objetivos e hipótesis planteadas para la reducción de las transferencias entre centros, evitando costos logísticos innecesarios por movimiento de mercancía.

Capítulo 6. Sugerencias para trabajos futuros

Uno de los principales propósitos de este trabajo es que este formato se pueda aplicar y sea escalable en las distintas secciones dentro de la empresa. Esta metodología puede ser utilizada para beneficiar a aquellas empresas que pasen por una situación similar, es decir, puede ser escalable y adaptable siempre y cuando se coloquen las variables necesarias y se haga una limpieza de datos de manera correcta.

Se sugiere definir de la manera más precisa posible los objetivos que se buscan al realizar la réplica. Este trabajo puede ser la base para determinar de una mejor manera si los costos logísticos se ven afectados por una mala distribución de la mercancía; usando el algoritmo se pueden obtener resultados que beneficien al logro del objetivo.

Referencias bibliográficas

1. Zamar, D. S., Gopaluni, B., & Sokhansanj, S. (2017). A constrained k-means and nearest neighbor approach for route optimization in the bale collection problem. *IFAC-PapersOnLine*, 50(1), 12125-12130.
2. Holý, V., Sokol, O., & Černý, M. (2017). Clustering retail products based on customer behaviour. *Applied Soft Computing*, 60, 752-762.
3. Amr, M., Ezzat, M., & Kassem, S. (2019, October). Logistics 4.0: Definition and historical background. In *2019 novel intelligent and leading emerging sciences conference (NILES)* (Vol. 1, pp. 46-49). IEEE.
4. Cao, L. (2017). Data science: challenges and directions. *Communications of the ACM*, 60(8), 59-68.
5. Ongsulee, P. (2017, November). Artificial intelligence, machine learning and deep learning. In *2017 15th international conference on ICT and knowledge engineering (ICT&KE)* (pp. 1-6). IEEE.
6. Rojas, E. M. (2020). Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo. *Revista Ibérica de Sistemas y Tecnologías de Información*, (E28), 586-599.
7. Nielsen, F. (2016). *Introduction to HPC with MPI for Data Science*. Springer.
8. Gülagiz, F. K., & Sahin, S. (2017). Comparison of hierarchical and non-hierarchical clustering algorithms. *International Journal of Computer Engineering and Information Technology*, 9(1), 6.
9. Pham, D. T., Dimov, S. S., & Nguyen, C. D. (2005). Selection of K in K-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219(1), 103-119.
10. Bock, H. H. (2007). Clustering methods: a history of k-means algorithms. *Selected contributions in data analysis and classification*, 161-172.
11. Hernández, J. A. (2016). Métodos de reducción de dimensionalidad: Análisis comparativo de los métodos APC, ACPP y ACPK. *Uniciencia*, 30(1), 115-122.
12. Van Rossum, Guido, et al. *Python Programming language*. En *USENIX annual technical conference*. 2007. p. 1-36.
13. CHATFIELD, Chris. *Exploratory data analysis*. *European journal of operational research*, 1986, vol. 23, no 1, p. 5-13.
14. Romano, J. D., Le, T. T., La Cava, W., Gregg, J. T., Goldberg, D. J., Chakraborty, P., ... & Moore, J. H. (2022). PMLB v1. 0: an open-source dataset collection for benchmarking machine learning methods. *Bioinformatics*, 38(3), 878-880.
15. Stewart, J., & Kennelly, P. J. (2010). Illuminated choropleth maps. *Annals of the Association of American Geographers*, 100(3), 513-534.
16. Forina, M., Armanino, C., & Raggio, V. (2002). Clustering with dendrograms on interpretation variables. *Analytica Chimica Acta*, 454(1), 13-19.

17. Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., ... & Lin, C. T. (2017). A review of clustering techniques and developments. *Neurocomputing*, 267, 664-681.
18. Fontalvo-Herrera, T., De-la-Hoz-Granadillo, E., & Mendoza-Mendoza, A. (2019). Los Procesos Logísticos y La Administración de la Cadena de Suministro. *Saber, ciencia y libertad*, 14(2), 102-112.
19. Castellanos Ramírez, A. (2021). *Logística comercial internacional*. Universidad del Norte.
20. Xu, Z., Tang, N., Xu, C., & Cheng, X. (2021). Data science: connotation, methods, technologies, and development. *Data Science and Management*, 1(1), 32-37.
21. Ramachandran, K. K., Mary, A. A. S., Hawladar, S., Asokk, D., Bhaskar, B., & Pitroda, J. R. (2022). Machine learning and role of artificial intelligence in optimizing work performance and employee behavior. *Materials Today: Proceedings*, 51, 2327-2331.
22. Nakaura, T., Higaki, T., Awai, K., Ikeda, O., & Yamashita, Y. (2020). A primer for understanding radiology articles about machine learning and deep learning. *Diagnostic and Interventional Imaging*, 101(12), 765-770.
23. Lorbeer, B., Kosareva, A., Deva, B., Softić, D., Ruppel, P., & Küpper, A. (2018). Variations on the clustering algorithm BIRCH. *Big data research*, 11, 44-53.
24. Villardón, J. L. V. (2007). *Introducción al análisis de clúster*. Departamento de Estadística, Universidad de Salamanca. 22p.
25. Martinez, I., Viles, E., & Olaizola, I. G. (2021). Data science methodologies: Current challenges and future approaches. *Big Data Research*, 24
26. Donoso, A., Aranda, E., & Ruiz, D. (2022). A new approach based on spectral graph theory to avoid enclosed holes in topology optimization. *Computer Methods in Applied Mechanics and Engineering*, 393.
27. Todd, C. S., Toth, T. M., & Busa-Fekete, R. (2009). GraphClus, a MATLAB program for clúster analysis using graph theory. *Computers & Geosciences*, 35(6), 1205-1213.
28. Lacave, C., Molina, A. I., & Giralt, J. (2013). Identificando algunas causas del fracaso en el aprendizaje de la recursividad: Análisis experimental en las asignaturas de programación.
29. Girón, M. F., López, J. R., & Sornoza, K. J. (2021). Solución al problema del transporte de Aplicación práctica. *Ecuadorian Science Journal*, 5(Esp. 4), 61-73.
30. El-Samak, A. F., & Ashour, W. (2015). Optimization of traveling salesman problem using affinity propagation clustering and genetic algorithm. *Journal of Artificial Intelligence and Soft Computing Research*, 5(4), 239-245.
31. Hernández-Vázquez, J. O., Hernández-González, S., Jiménez-García, J. A., Hernández-Ripalda, M. D., & Hernández-Vázquez, J. I. (2019). Enfoque híbrido metaheurístico AG-RS para el problema de asignación del buffer que minimiza el inventario en proceso en líneas de producción abiertas en serie. *Revista Iberoamericana de Automática e Informática Industrial*, 16(4), 447-458.

32. Taha, H. A. (2004). Investigación de operaciones. Pearson Educación.
33. Correa Henao, M. (2021). Análisis de clúster automático (Doctoral dissertation, Universidad Nacional de Colombia).
34. Malindžák, D. (2015). The Basic Principle of Logistic Theory. Applied Mechanics and Materials, 708, 47-52.
35. Demey, J. R., Pla, L., Vicente-Villardón, J. L., Di Rienzo, J., & Casanoves, F. (2011). Medidas de distancia y similitud. Valoración y análisis de la diversidad funcional y su relación con los servicios ecosistémicos, 384, 47-59.
36. Tukey, J. W. (1977). Exploratory data analysis (Vol. 2, pp. 131-160).
37. Babic, B., Chen, D. L., Evgeniou, T., & Fayard, A. L. (2021). A better way to onboard AI (No. hal-03624592).
38. Huerta Bustos, M. A. (2011). Segmentación de Clientes de Tienda Retail Especialista.
39. Villagra, A., Guzmán, A., Pandolfi, D., & Leguizamón, G. (2009). Análisis de medidas no-supervisadas de calidad en clústeres obtenidos por K-means y Particle Swarm Optimization.
40. AWS (2023). ¿Qué es python?
<https://aws.amazon.com/es/what-is/python/#:~:text=Los%20desarrolladores%20utilizan%20Python%20porque,aumenta%20la%20velocidad%20del%20desarrollo.>
41. Oyarzún, B. R. (2018). LA LOGÍSTICA 4.0. Revista de Marina,(964), 39-44.
42. Dike, H. U., Zhou, Y., Deveerasetty, K. K., & Wu, Q. (2018). Unsupervised learning based on artificial neural network: A review. In 2018 IEEE International Conference on Cyborg and Bionic Systems (CBS) (pp. 322-327). IEEE.
43. Dayan, P., Sahani, M., & Deback, G. (1999). Unsupervised learning. The MIT encyclopedia of the cognitive sciences, 857-859.
44. Hacer data visualization Gracias a plotly. Formation Data Science | DataScientest.com. (2022, August 1).
<https://datascientest.com/es/hacer-data-visualisation-con-plotly#:~:text=offline%20es%20una%20librería%20que,abriéndolos%20en%20el%20navegador%20web.>
<https://yellowbrick-es.readthedocs.io/es/latest/about.html#:~:text=La%20biblioteca%20Yellowbrick%20es%20una,procedimiento%20de%20selección%20de%20modelos.>
45. Álvarez CorrWHO (2020) Who coronavirus (COVID-19) dashboard, World Health Organization. Available at: <https://covid19.who.int/> (Accessed: 01 October 2023).
46. Edera, P. (2023, February 7). Diccionario de Python y todas sus librerías. CIBERNINJAS. <https://ciberninjas.com/python-diccionario/#NumPy>
47. Ríos Jacobo, O. I. (2019). Key Performance Indicators (KPI).
48. Pursell, S. (2023, January 20). Todo sobre LOS KPI: Definición, Tipos y ejemplos. HubSpot. <https://blog.hubspot.es/marketing/que-es-un-kpi>
49. Aragón Gallegos, A. D. C., Cerquin Silva, S. I., Ecurra Yactayo, R. O., & Roncalla Viena, A. L. (2023). Segmentación de clientes para mejorar la experiencia de compra de productos electrónicos en Falabella.

50. Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., ... & Ramkumar, P. N. (2020). Machine learning and artificial intelligence: definitions, applications, and future directions. *Current reviews in musculoskeletal medicine*, 13, 69-76.
51. Vázquez, A. M. (2018). *Introducción a machine learning*.
52. Latorre, M. (2018). *Historia de las web, 1.0, 2.0, 3.0 y 4.0*. Universidad Marcelino Champagnat, 1.
53. Abeliuk, A., & Gutiérrez, C. (2021). Historia y evolución de la inteligencia artificial. *Revista Bits de Ciencia*, (21), 14-21.
54. Simchi-Levi, D., Chen, X., & Bramel, J. (2005). *The logic of logistics. Theory, algorithms, and applications for logistics and supply chain management*.
55. Kain, R., & Verma, A. (2018). Logistics management in supply chain—an overview. *Materials today: proceedings*, 5(2), 3811-3816.
56. Arora, S., Hu, W., & Kothari, P. K. (2018, July). An analysis of the t-sne algorithm for data visualization. In *Conference on learning theory* (pp. 1455-1462). PMLR.

Anexo 01

Lista de las tiendas Liverpool con su respectiva zona y estado de la república mexicana.

Se encuentra en: <https://github.com/DiegoAguirre19/AnexosTesisDiegoAguirre>

Anexo 02

Lista de resultados de Tenis.

Se encuentra en: <https://github.com/DiegoAguirre19/AnexosTesisDiegoAguirre>

Anexo 03

Lista de resultados Sandalias.

Se encuentra en: <https://github.com/DiegoAguirre19/AnexosTesisDiegoAguirre>

Anexo 04

Gráficas de correlación

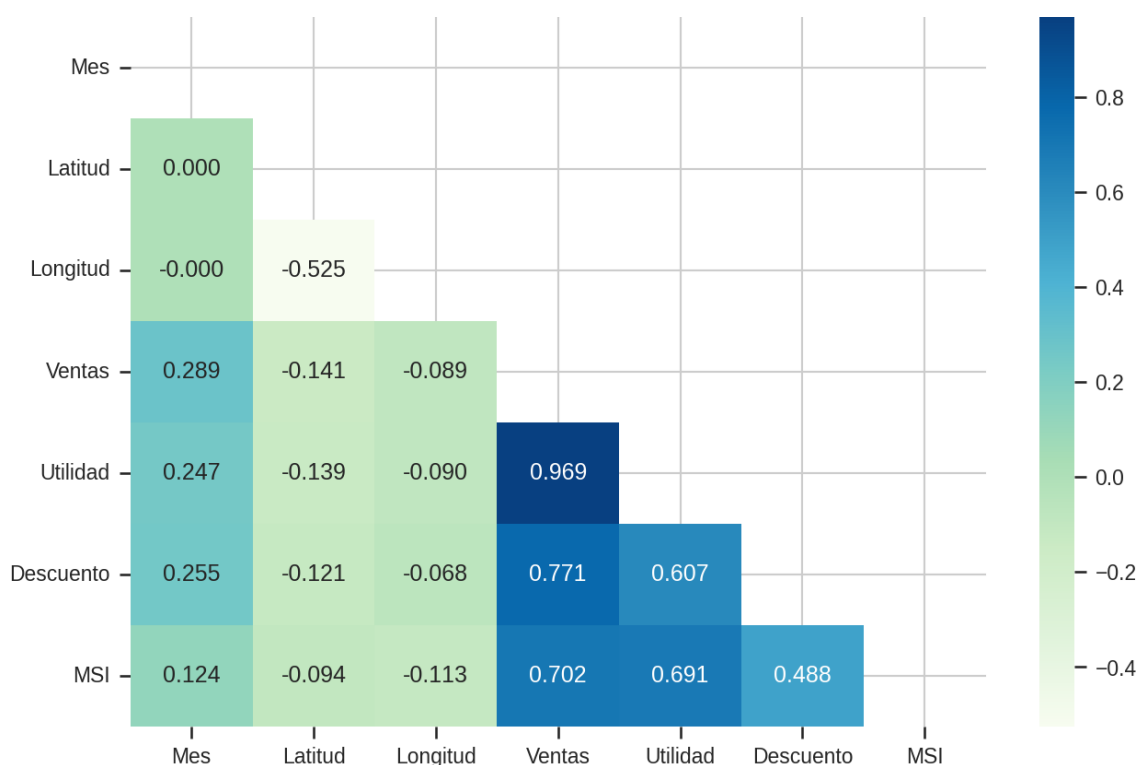


Figura 31- Matriz de correlación Tenis 2019

Se observa que la intersección con más correlación es la de Utilidad con las Ventas_\$ y el caso contrario la variable de longitud y latitud son aquellas que tienen menos correlación con las demás variables.

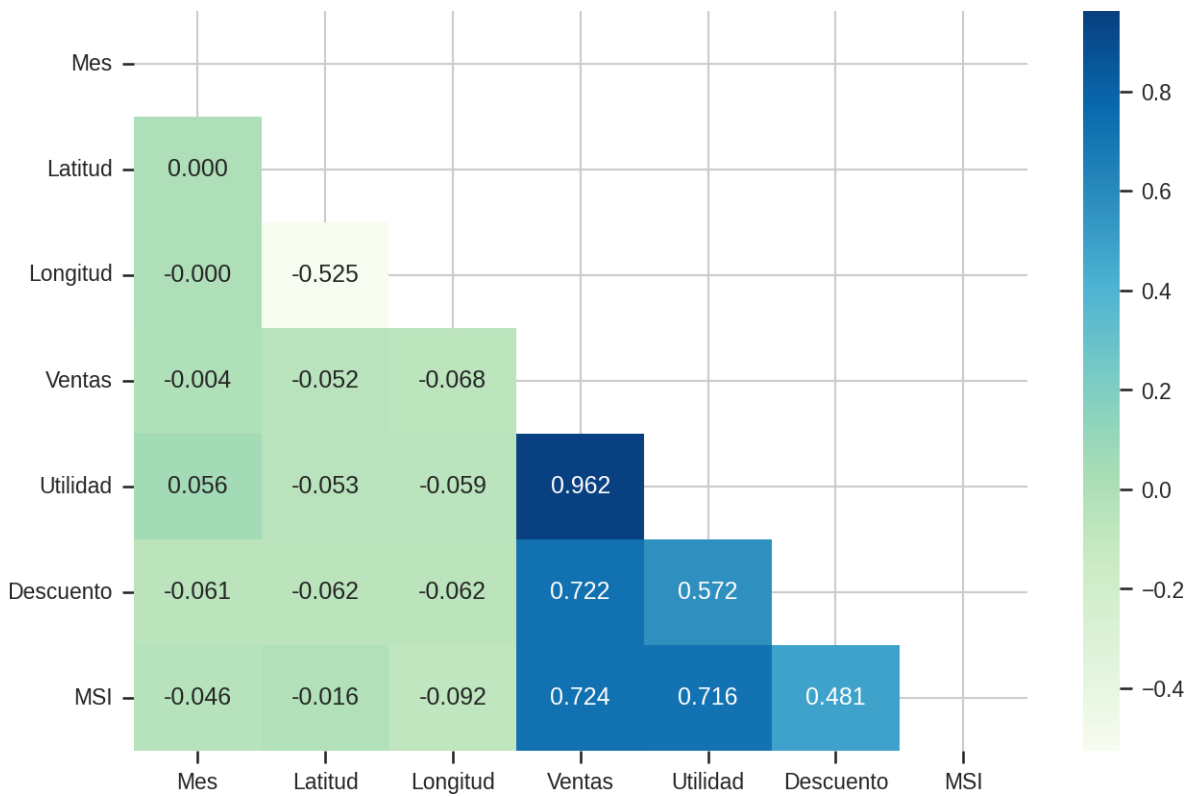


Figura 32- Matriz de correlación Sandalia 2019

Se observa que la intersección con más correlación es la de Utilidad con las Ventas_\$ y el caso contrario la variable de longitud, mes y latitud son aquellas que tienen menos correlación con las demás variables.

Gráficas de violín

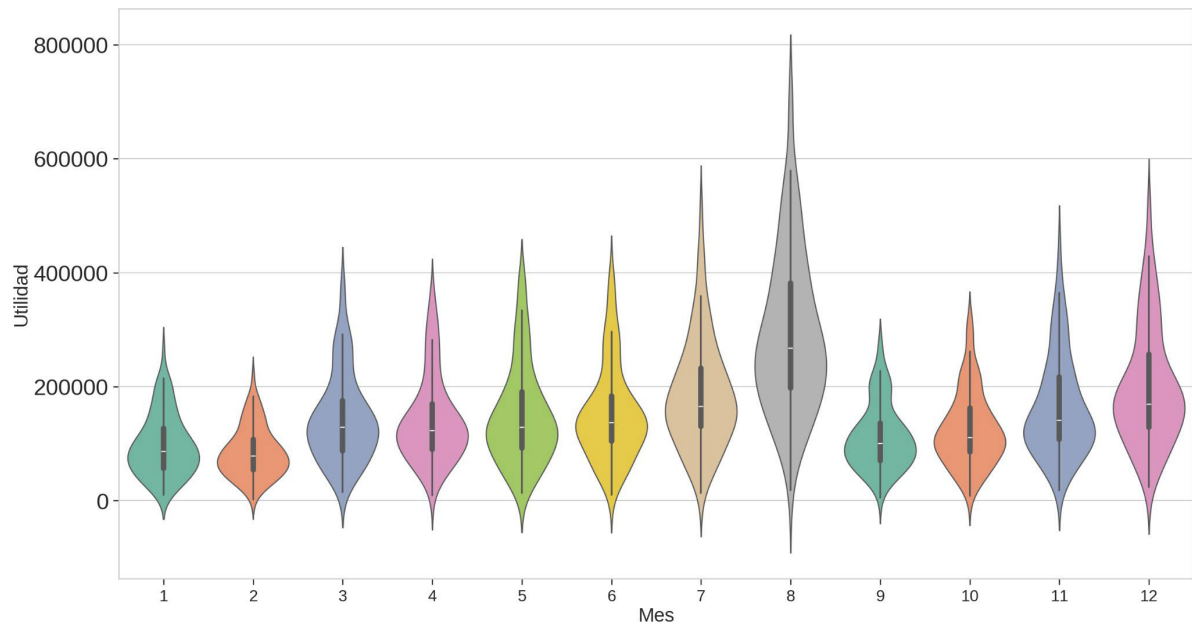


Figura 33- Gráfica de violín Utilidad Tennis 2019

El mes de agosto es el que más utilidades tiene, mientras que el mes con menos utilidades es el mes de febrero. En esta se pueden observar violines más alargados como los de los meses 7, 8, 11 y 12 o violines más pequeños con un cuerpo más redondo como los meses 1, 2, 9 y 10.

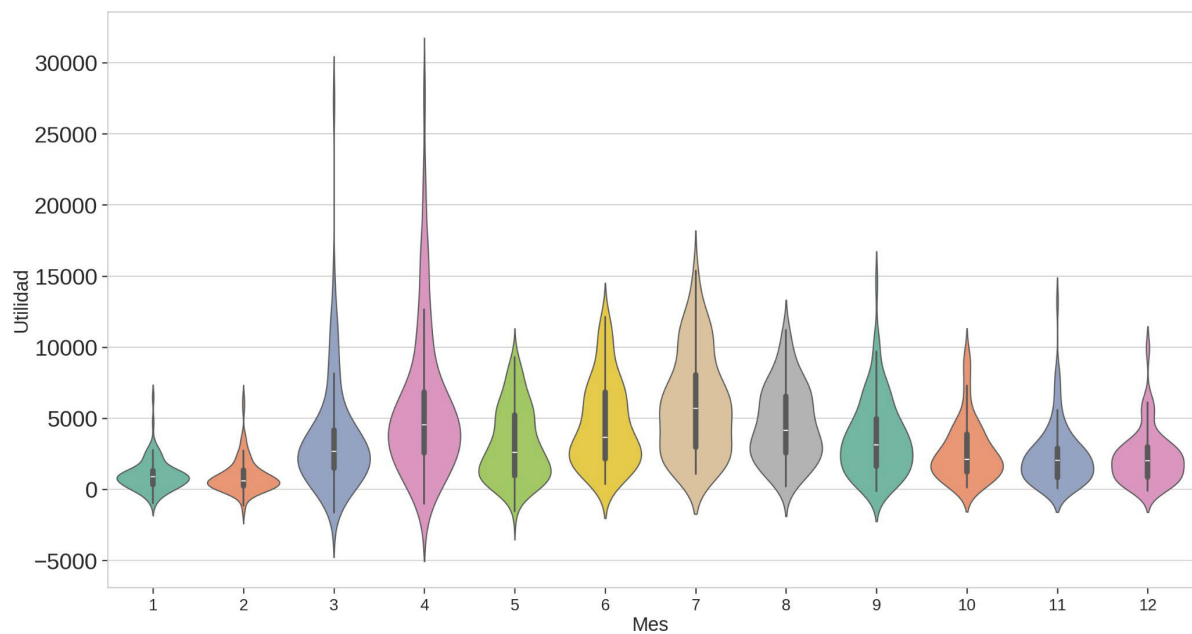


Figura 34- Gráfica de violín Utilidad Sandalia 2019

Los meses de enero y febrero son aquellos los cuales no tienen tantas utilidades como el resto, mientras que los meses de marzo y abril sobresalen de los demás meses.

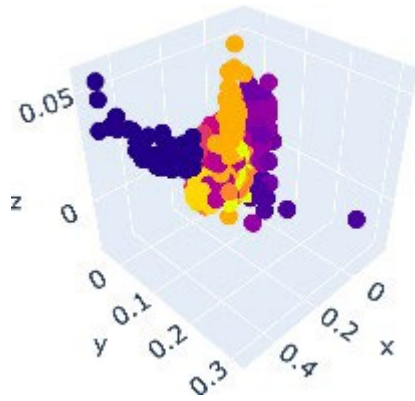


Figura 36- Mapa coroplético promedio de total descuento por estado

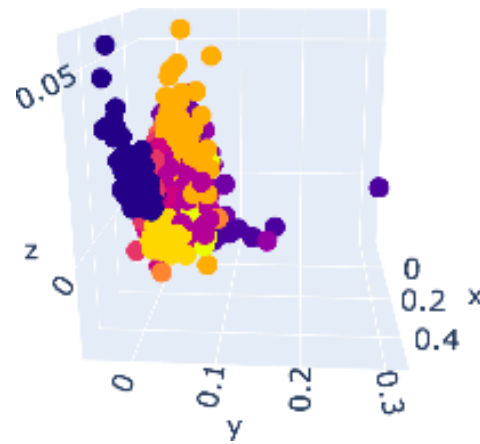
Se observa que los estados que más descuentos realizan son la ciudad de México e Hidalgo en promedio.

PCA

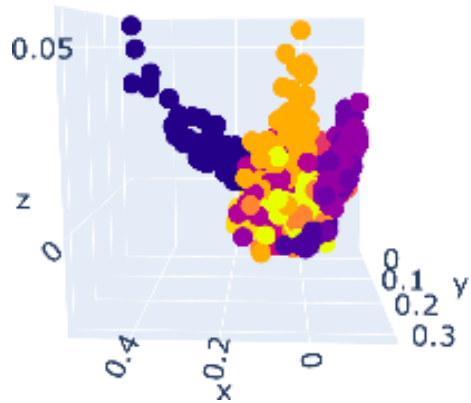
A)



B)



C)



D)

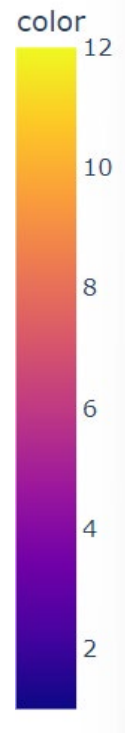
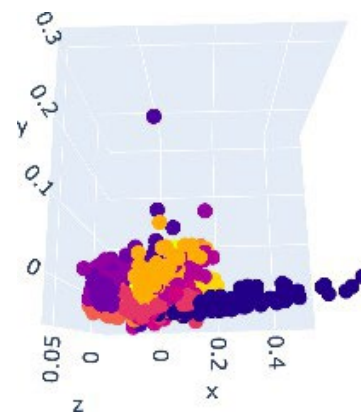
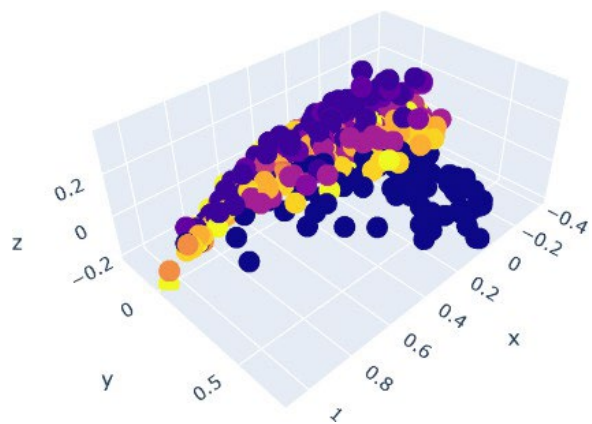
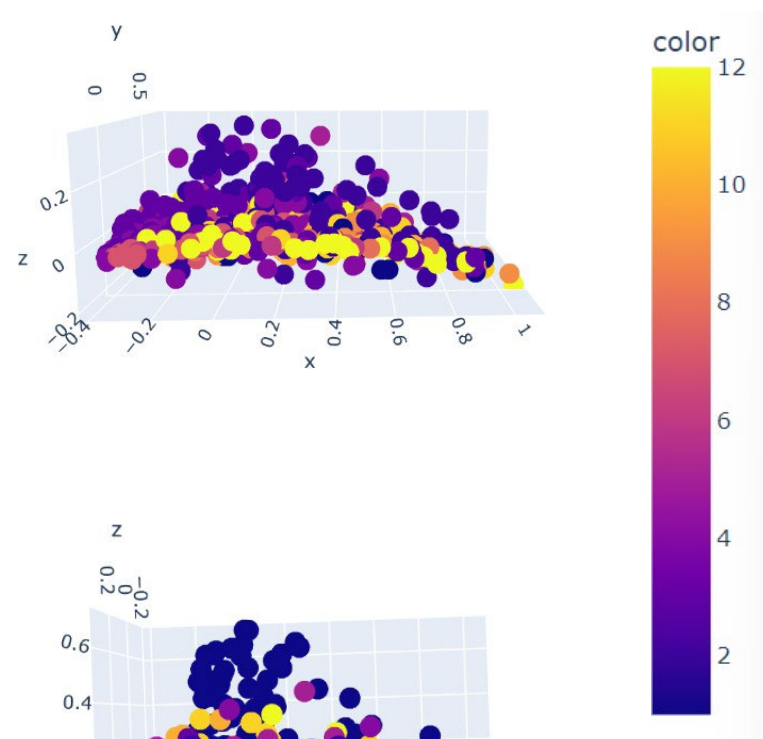


Figura 37- Vistas de PCA Tenis 2019 A) Vista isométrica, B) Vista superior, C) Vista lateral, D) Vista frontal

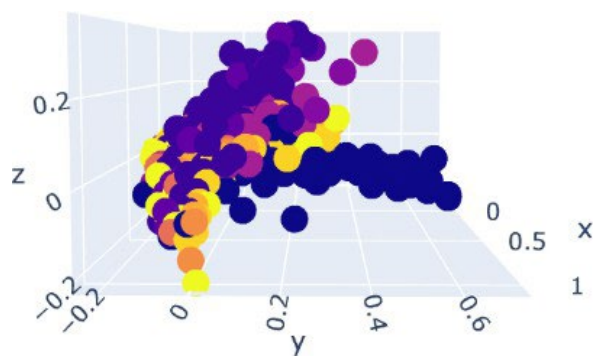
A)



B)



C)



D)

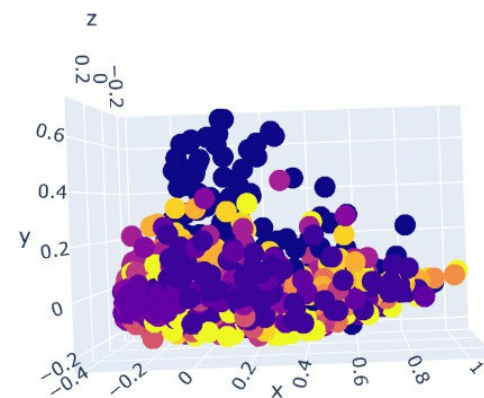


Figura 38- Vistas de PCA Sandalia 2019, A) Vista isométrica, B) Vista superior, C) Vista lateral, D) Vista frontal

Dendrograma

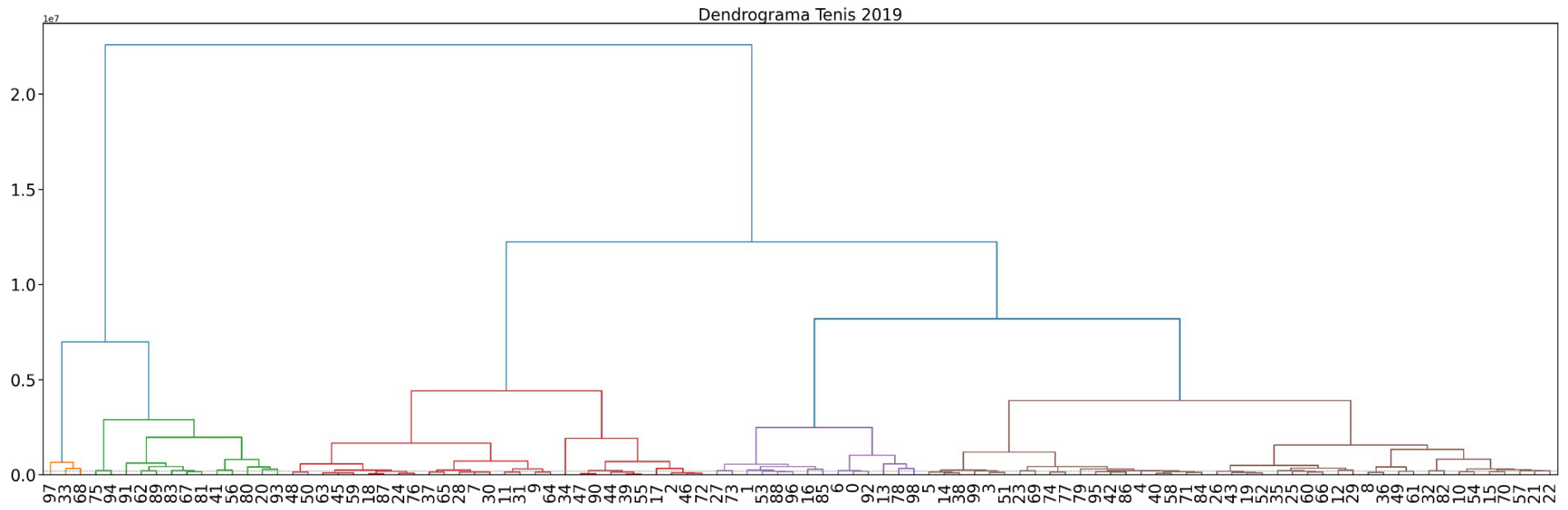


Figura 39- Dendrograma Tenis 2019

Se observa la división en 5 dendrogramas, los cuales son representados por diferentes colores. Cada cluster tiene una cantidad distinta de tiendas, siendo el color amarillo el que cuenta con solo 3 tiendas y el café siendo el cluster más grande.

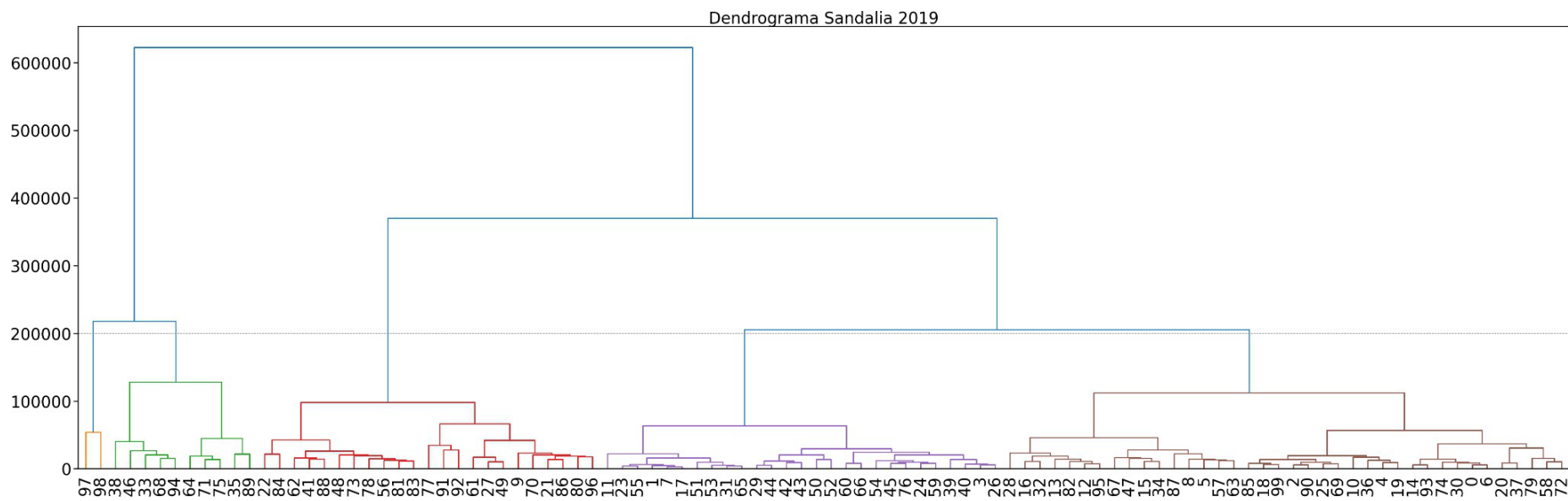


Figura 40- Dendrograma Sandalia 2019

Se observa la división en 5 dendrogramas, los cuales son representados por diferentes colores. Cada cluster tiene una cantidad distinta de tiendas, siendo el color amarillo el que cuenta con solo 2 tiendas y el café siendo el cluster más grande.

Método del codo

Tenis 2019

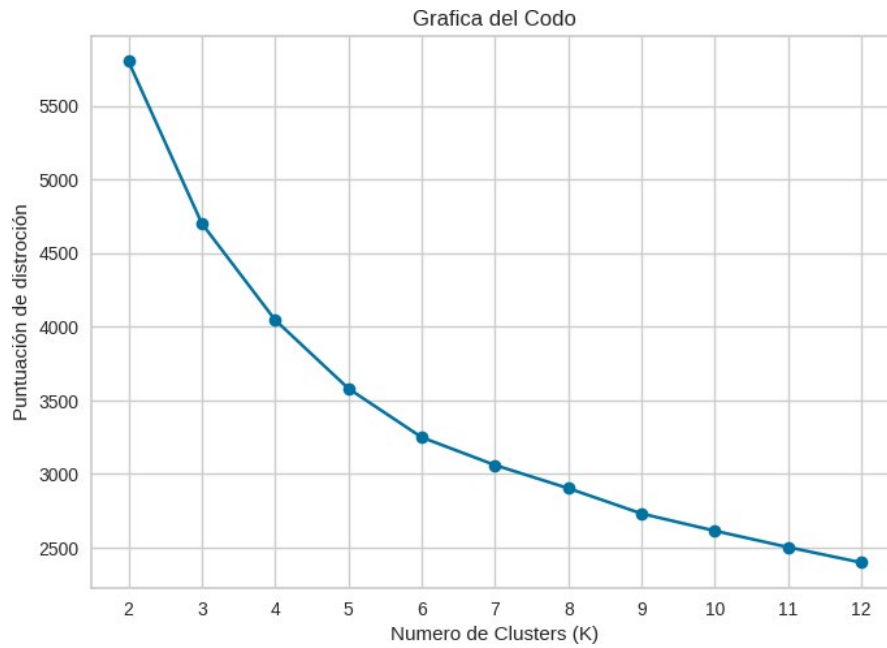


Figura 41- Método del codo Tennis 2019

En la figura se observa que el codo se encuentra en la K #5, la puntuación de dicha K es de 3,580.83

Sandalia 2019

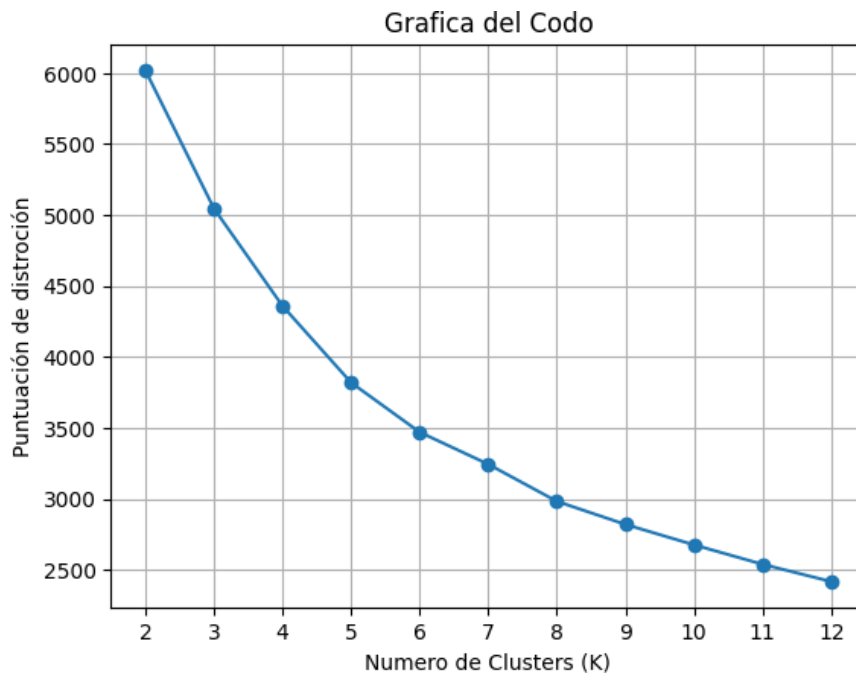


Figura 42- Método del codo Sandalia 2019

Se observa que el codo se encuentra en la K #5, la puntuación de dicha K es de 3,819.

Características del cluster

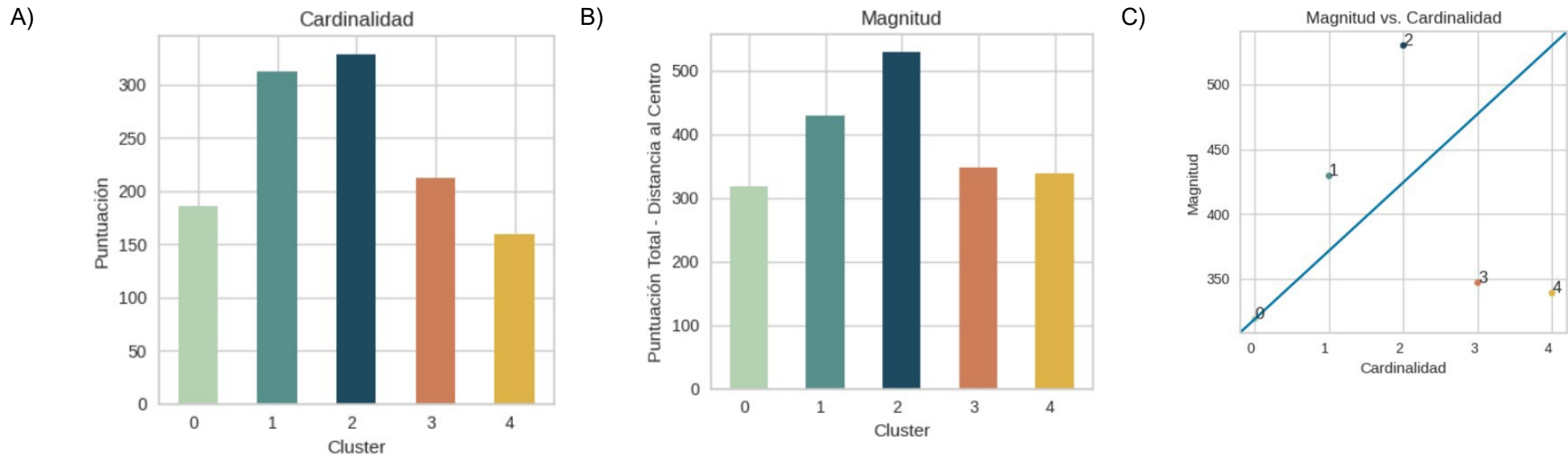


Figura 43- Cardinalidad y magnitud clústeres Tenis 2019, A) Cardinalidad, B) Magnitud, C) Cardinalidad vs. magnitud

Se puede observar que el clúster cero tiene una magnitud más alta que su cardinalidad, el clúster 1 tiene una cardinalidad arriba de los 300 puntos, mientras que en magnitud tiene arriba de 400 puntos. El clúster 2 es aquel con más magnitud y cardinalidad. El clúster 3 tiene una cardinalidad arriba de los 200 puntos y una magnitud abajo de los 400 puntos y es el clúster que se posiciona en el tercer lugar en cuanto a magnitud y el clúster 4, que es el más pequeño en magnitud y cardinalidad, tiene una magnitud similar al clúster 3 y 0.

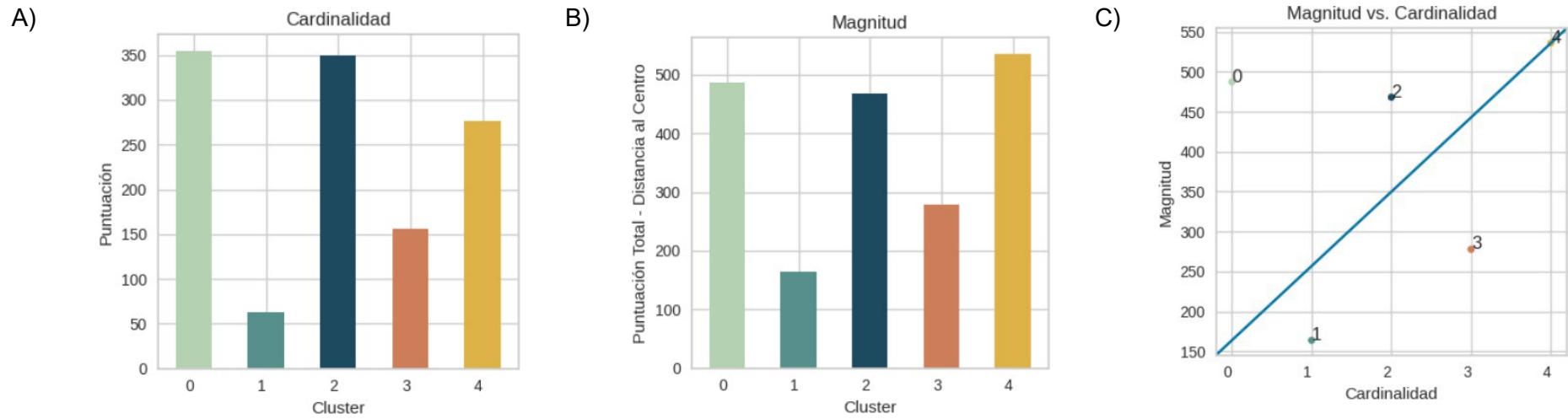


Figura 44- Cardinalidad y magnitud clústeres Sandalia 2019, A) Cardinalidad, B) Magnitud, C) Cardinalidad vs. magnitud

Se puede observar que en cuanto a su cardinalidad el clúster 0 es el mayor, seguido por el clúster 2 con alrededor 330 puntos y le sigue por debajo el clúster 4 con 275 puntos aproximadamente. En cuanto a la magnitud, el clúster 4 es el más grande con alrededor 500 puntos, seguido de los clústeres 0 y 2 con alrededor de 430 puntos. El cluster 4 es el inicio con correlación alta.

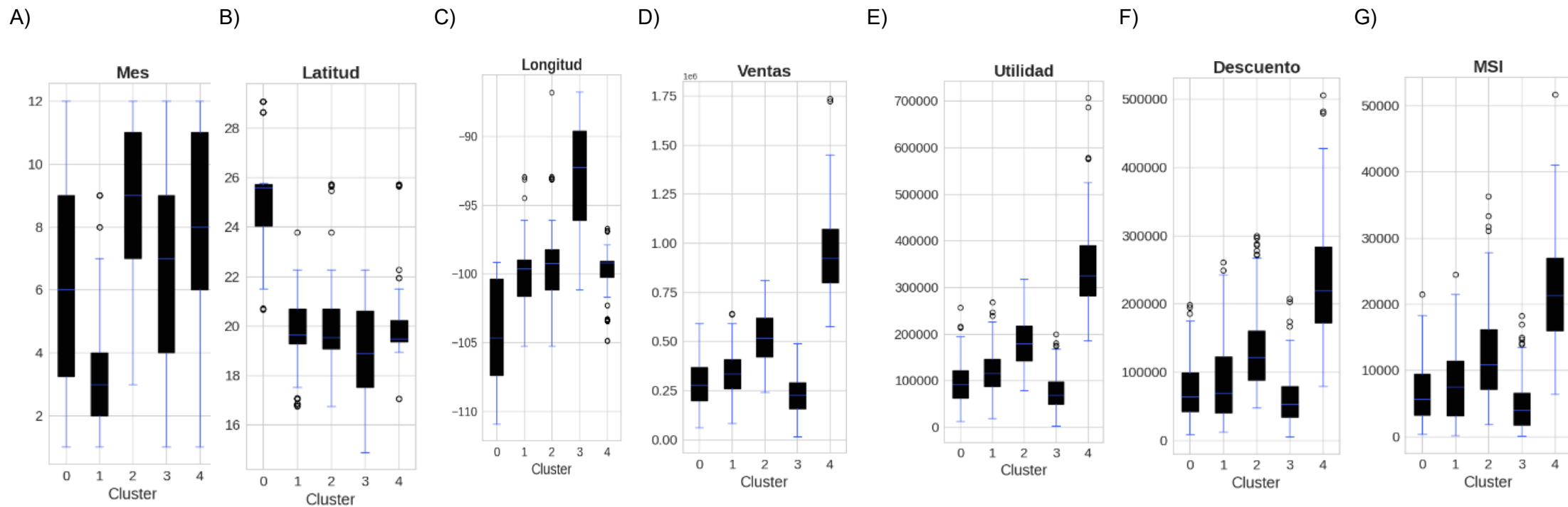


Figura 45- Distribuciones de característica por clúster Tennis 2019, A) Gráfica de cajas y bigotes Mes, B) Gráfica de cajas y bigotes Latitud, C) Gráfica de cajas y bigotes Longitud, D) Gráfica de cajas y bigotes mes Ventas, E) Gráfica de cajas y bigotes Utilidad, F) Gráfica de cajas y bigotes Descuento, G) Gráfica de cajas y bigotes MSI

Se puede observar que el clúster 0 es el que cuenta con presencia en todos los meses, donde predominan los meses de marzo a septiembre. Es un clúster que cuenta con tiendas entre las latitudes 22° a 26° y longitudes entre -97° y -113°. Son tiendas que tienen ventas entre los \$100,000 y \$600,000, con utilidades entre los \$50,000 y los \$200,000. Estas tiendas tuvieron descuentos y costo de meses sin intereses entre \$100,000 y \$170,000. El clúster 1 está conformado por los meses de enero a julio, son principalmente centros ubicados entre las latitudes de 17° y 22.5° y longitudes entre los -93° y -106°. Son centros que tuvieron ventas entre los \$100,000 y \$600,000 y utilidades entre \$50,000 y \$275,000 y que tuvieron actividad promocional de descuentos y meses sin intereses hasta \$250,000. El clúster 2 está conformado por los meses de marzo a diciembre, centros ubicados entre las latitudes de 17° y 22.5° y longitudes entre los -93° y -106°. Con ventas entre los \$250,000 y los \$750,000, utilidades entre \$75,000 y \$300,000, actividad promocional en descuentos de \$50,000 a \$300,000 y tienen un costo de meses sin intereses de hasta \$35,000. El clúster 3 es el que tiene presencia en todos los meses, especialmente entre abril y septiembre. Son centros con ubicaciones en latitudes entre 14° y 23°, longitudes entre los -85° y los -103°, con ventas de hasta \$500,000 y utilidades de hasta \$200,000. La actividad promocional en descuentos es hasta \$150,000 y meses sin intereses hasta \$15,000. El clúster 4 tiene presencia principalmente en los meses de junio a noviembre, pero cuenta con menor presencia en el resto de los meses. Son centros con ubicaciones entre las latitudes de 19° y 22° y longitudes entre los -95° y -105° con ventas entre \$500,000 y \$1,750,000, utilidades entre los \$200,000 y los \$700,000. Fueron centros con actividad promocional entre los \$75,000 y los \$500,000 y meses sin intereses entre los \$7,500 y \$50,000.

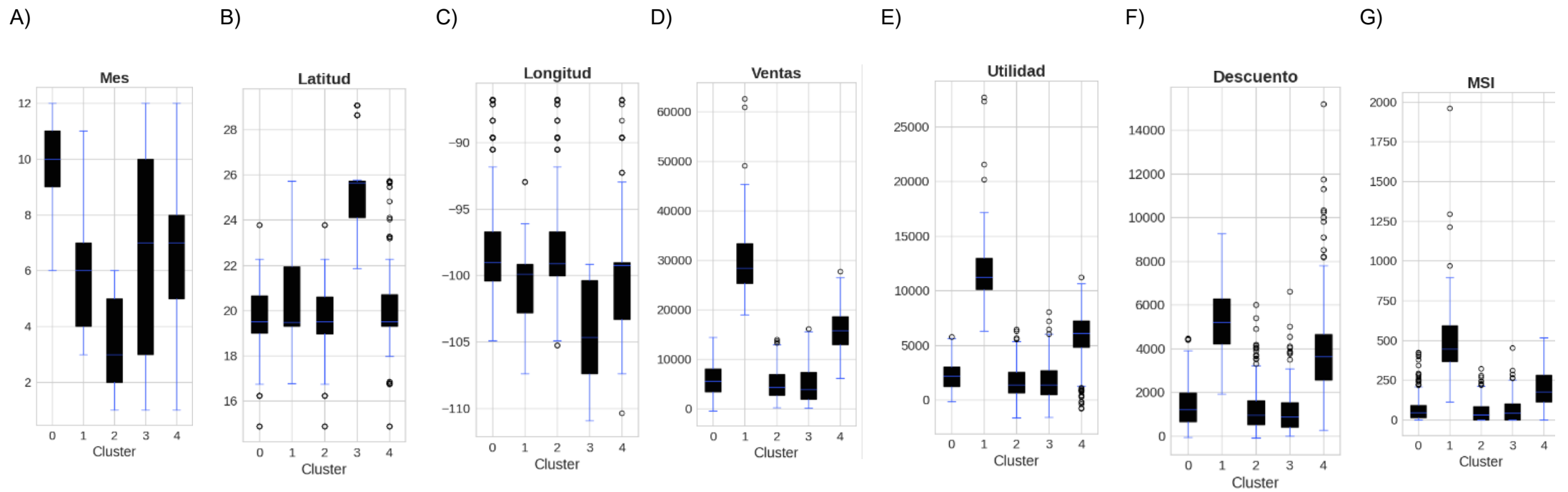


Figura 46- Distribuciones de característica por clúster Sandalia 2019, A) Gráfica de cajas y bigotes Mes, B) Gráfica de cajas y bigotes Latitud, C) Gráfica de cajas y bigotes Longitud, D) Gráfica de cajas y bigotes mes Ventas, E) Gráfica de cajas y bigotes Utilidad, F) Gráfica de cajas y bigotes Descuento, G) Gráfica de cajas y bigotes MSI

Se puede observar el comportamiento de los clústeres de la siguiente manera. El clúster 0 tiene presencia a lo largo del segundo semestre, principalmente entre los meses de septiembre a noviembre, y son centros con latitudes entre los 16° y los 24° , longitudes entre los -85° y los -105° , con ventas hasta los \$20,000 y utilidades máximas de \$5,000. Estos centros tienen descuentos máximos de \$4,000 y meses sin intereses de un máximo de \$500. El clúster 1 tiene presencia de marzo a noviembre, pero predominan los meses de mayo a julio. Son centros ubicados en las latitudes de 16° a 26° y longitudes de -95° a -108° , con ventas entre \$20,000 y \$65,000 y utilidades entre \$5,000 y \$30,000. Los descuentos están entre \$2,000 y \$10,000 y el costo de meses sin intereses entre \$100 y \$2,000. El clúster 2 se puede observar que principalmente cuenta con meses entre febrero y mayo, pero tiene presencia en los meses de enero y junio, en cuanto a la latitud y longitud se encuentran entre los 17° y 23° y -93° y los -107° , respectivamente. Son centros con ventas máximas de \$15,000, utilidades máximas de \$6,000, descuentos máximos de \$4,500 y meses sin intereses de \$250. El clúster 3 tiene presencia durante todo el año, principalmente en los meses de marzo a octubre. En cuanto a latitud y longitud se encuentran entre los 22° y los 26° y entre -97° y -112° , respectivamente. Tiene ventas máximas de \$20,000, utilidades de \$7,500, descuentos de \$4,000 y meses sin intereses de \$250. El clúster 4, igual que el clúster 3, tiene presencia durante todos los meses, pero en este caso principalmente entre los meses de mayo y agosto. Su latitud y longitud es entre 18° y los 22° y entre los -85° y los -105° , respectivamente, ventas máximas de \$30,000, utilidades de \$12,000, descuentos de \$8,000 y meses sin intereses de \$500.

Gráfica de radar

Características del conglomerado: Medias de características por conglomerado

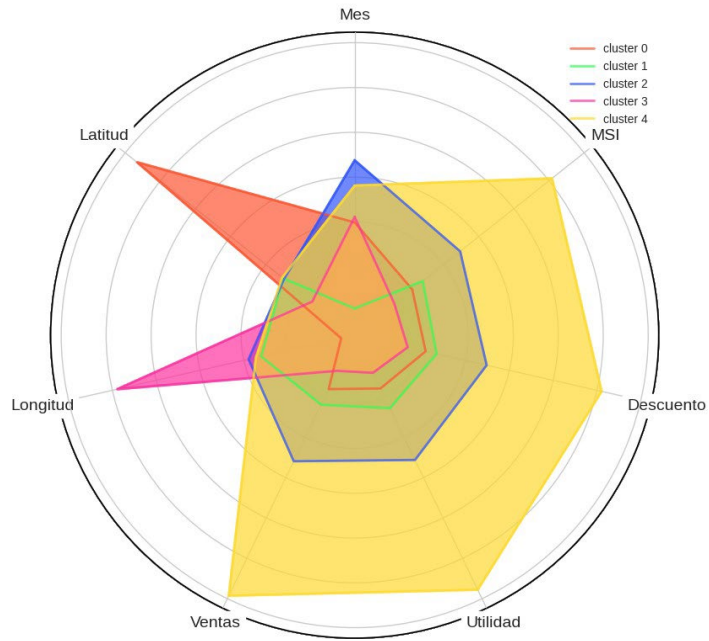


Figura 47- Radar Tennis 2019

Características del conglomerado: Medias de características por conglomerado

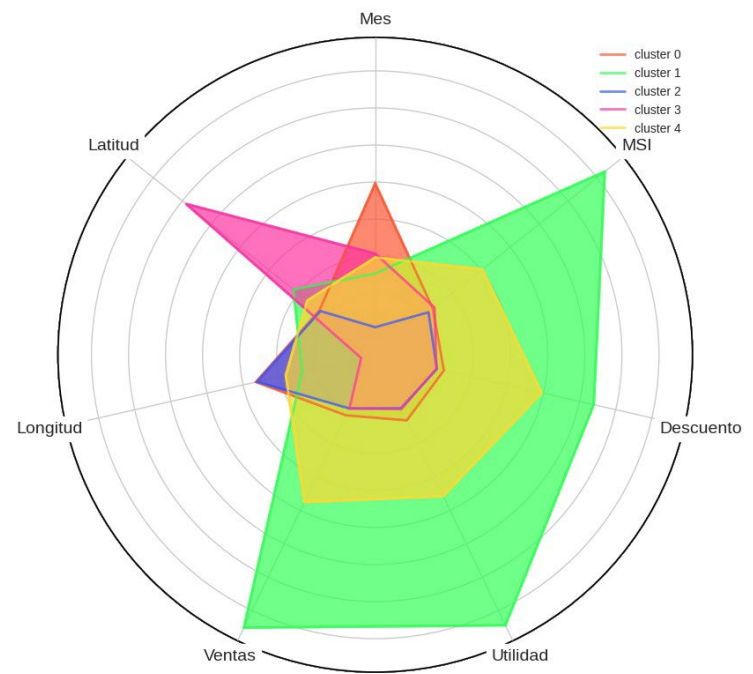


Figura 48- Radar Sandalia 2019