



Escuela de Administración  
Escuela de Ingeniería, Ciencia y Tecnología

Maestría en Business Analytics

Trabajo de Grado  
Marco de trabajo para el desarrollo de la segmentación de clientes para Locatel

Presentado por:

Nicolás Gonzalez Medina

Bogotá, D.C. 15 de septiembre de 2023



Escuela de Administración  
Escuela de Ingeniería, Ciencia y Tecnología

Maestría en Business Analytics

Trabajo de Grado  
Marco de trabajo para el desarrollo de la segmentación de clientes para Locatel

Presentado por:  
Nicolás Gonzalez Medina

Bajo la dirección de:  
Ferney Maldonado López

Bogotá, D.C. 15 de septiembre de 2023

## Contenido

Contenido .....	3
Preliminares .....	5
Agradecimientos .....	6
Declaración de originalidad y autonomía .....	7
Declaración de exoneración de responsabilidad .....	8
Lista de Ilustraciones .....	9
Abreviaturas .....	13
Glosario.....	14
Resumen Ejecutivo .....	16
Palabras clave .....	17
Abstract.....	18
Keywords .....	19
1. Introducción .....	1
2. Objetivos .....	4
3. Alcance.....	5
4. Metodología .....	6
5. Cronograma.....	8
6. Descripción de las fuentes de información .....	9
6.1. Resumen: .....	9
6.2. Base Datos Patología: .....	9
6.3. Análisis Descriptivo Base Datos Locatel:.....	11
6.4. Análisis Ticket Compra Locatel Marzo-Mayo 2022 .....	18
6.5. Análisis Compra Por Género En Locatel:.....	20
6.6. Análisis Ticket Compra Locatel Marzo-Mayo 2022: .....	22
6.7. Análisis Consumo Días Feriados 2022:.....	24
7. Descripción de la Situación organizacional donde se realizará el proyecto (Contexto)29	
7.1. Sector Farmacéutico Colombia:.....	29
7.2. Contexto Locatel:.....	31
7.3. Análisis de participación en ventas por tienda para Locatel:.....	33

7.3.1.	Análisis Tienda Chapinero - Bogotá: .....	34
7.3.2.	Comparativo Tienda Calle 116 vs Tienda Plaza Américas: .....	36
8.	Descripción de la situación estudio de caso y/o problemática empresarial y método y/o estrategia a aplicar para su solución .....	41
9.	Descripción de las alternativas, estrategias y/o acciones que se toman en el análisis de la solución a la problemática .....	43
9.1.	Descripción Alternativas de Proyecto: .....	43
9.2.	Desarrollo Segmentación: .....	45
9.2.1.	Interacción entre variables de Cantidad Ítems y Edad: .....	48
9.2.2.	Interacción entre variables de Cantidad Ítems y Valor Facturado: .....	49
9.2.3.	Interacción entre variables de Tiendas y Edad: .....	51
9.2.4.	Interacción entre variables de Cantidad y Tienda: .....	53
9.2.5.	Análisis de Patología Cuidado Personal: .....	54
9.2.6.	Interacción entre variables de Valor Facturado, Patología y Edad (Grafico Burbujas): .....	57
9.2.7.	Análisis PCA: .....	59
9.2.8.	Resumen Cluster Generados: .....	64
9.2.9.	Buyer Personas para Locatel: .....	69
9.3.	Exploración Segmentación: .....	71
10.	Plan y recomendaciones de implementación y aplicación .....	73
10.1.	Recomendación No 1: .....	73
10.2.	Recomendación No 2: .....	73
10.3.	Recomendación No 3: .....	74
10.4.	Recomendación No 4: .....	74
10.5.	Recomendación No 5: .....	75
10.6.	Recomendación No 6: .....	75
11.	Conclusiones.....	76
	Referencias bibliográficas .....	78
	Anexos Técnicos .....	<b>¡Error! Marcador no definido.</b>

# Preliminares

### **Agradecimientos**

Para este proyecto de grado obtuve conocimientos, herramientas, características y experiencias que me ayudaron a crecer profesional y personalmente, por ello quiero agradecer a mi familiar, dado que por su apoyo puedo cursar estos estudios de maestría, agradecer el director del mismo por los consejos, directrices y apoyo brindado en todo el desarrollo. Por otra parte, agradecer a la universidad y los profesores por el acceso al nuevo conocimiento y la oportunidad de realizar un trabajo exploratorio y práctico en búsqueda de desarrollar las habilidades académicas adquiridas.

### **Declaración de originalidad y autonomía**

Declaro(amos) bajo la gravedad del juramento, que he(mos) escrito el presente Proyecto Aplicado Empresarial (PAE), en la propuesta de solución a una problemática en el campo de conocimientos del programa de Maestría por mi(nuestra) propia cuenta y que, por lo tanto, su contenido es original.

Declaro(amos) que he(mos) indicado clara y precisamente todas las fuentes directas e indirectas de información y que este PAE no ha sido entregado a ninguna otra institución con fines de calificación o publicación.

*Nicolas Gonzalez Medina*

Nombre completo del autor 1

Firmado en Bogotá, D.C. el 15 de septiembre de 2023

### **Declaración de exoneración de responsabilidad**

Declaro(amos) que la responsabilidad intelectual del presente trabajo es exclusivamente de su(s) autor(es). La Universidad del Rosario no se hace responsable de contenidos, opiniones o ideologías expresadas total o parcialmente en él.

*Nicolas Gonzalez Medina*

Nombre completo del autor 1

Firmado en Bogotá, D.C. el 15 de septiembre de 2023

## Lista de Ilustraciones

Ilustración 1. Cronograma detalle de entregas proyecto empresarial Locatel. Creación Propia.....	8
Ilustración 2 Muestra de la base de datos por medicamento y patología. Creación Propia.	10
Ilustración 3 Distribución de la base de datos por Cantidad de Items y Patología. Creación Propia.....	11
Ilustración 4 Resumen tipo de información de la base de datos Locatel. Creación Propia..	12
Ilustración 5 Porcentaje de datos vacíos en la base de datos Locatel. Creación Propia. ....	13
Ilustración 6 Algoritmo KNN variable Genero. Creación Propia. ....	14
Ilustración 7 Resumen de datos numéricos de la base de datos Locatel. Creación Propia ..	15
Ilustración 8 Limpieza de datos numéricos de la base de datos Locatel. Creación Propia. .	16
Ilustración 9 Distribución variable v por Edad y Genero Locatel. Creación Propia. ....	17
Ilustración 10 Distribución ventas por día Locatel. Creación Propia. ....	18
Ilustración 11 Distribución de ventas por Categoría generada por Locatel. Creación Propia. .....	19
Ilustración 12 Distribución por familia de productos generada por Locatel. Creación Propia .....	20
Ilustración 13 Distribución por género de ventas Locatel. Creación Propia. ....	21
Ilustración 14 Distribución de ventas por tienda y mes Locatel. Creación propia.....	22
Ilustración 15 Distribución de ventas por tienda y categoría Junio Locatel. Creación propia .....	23
Ilustración 16 Listado de festivos para Colombia 2022. Creación Propia. ....	24

Ilustración 17 Distribución de ventas por cantidad y categoría del 14 y 15 abril 2023 para Locatel. Creación Propia. ....	25
Ilustración 18 Distribución de ventas por familia productos para el 14 y 15 abril 2022 en Locatel. Creación Propia. ....	26
Ilustración 19 Distribución de ventas por cantidad y categoría del 12 y 13 mayo 2022 para Locatel. Creación Propia. ....	27
Ilustración 20 Distribución ventas por día y tienda Locatel. Creación Propia. ....	33
Ilustración 21 Distribución ventas por Desc. Nivel 2 para tienda Restrepo Locatel. Creación Propia.....	34
Ilustración 22 Ventas por familia de producto de tienda Locatel Chapinero. Creación Propia.....	36
Ilustración 23 Comparativo de ventas por familias de producto tienda Calle 116 y Plaza Américas. Creación Propia .....	37
Ilustración 24 Ventas de tienda Locatel Calle 116. Creación Propia. ....	38
Ilustración 25 Ventas de tienda Locatel Plaza Américas. Creación Propia.....	38
Ilustración 26 Correlación variables normalizadas Base Datos. Creación Propia. ....	46
Ilustración 27 Correlación variables numéricas Base Datos. Creación Propia. ....	47
Ilustración 28 Agrupación de variables Cantidad - Edad. Creación Propia. ....	48
Ilustración 29 Distribución de la agrupación de variables Cantidad - Edad. Creación Propia. ....	49
Ilustración 30 Agrupación de variables Cantidad – Valor Factura. Creación Propia. ....	50
Ilustración 31 Distribución de agrupación de variables Cantidad – Facturación en millones. Creación Propia. ....	50

Ilustración 32 Agrupación de variables Tienda – Edad. Creación Propia.....	51
Ilustración 33 Distribución de agrupación entre variables de Tienda – Edad. Creación Propia.....	52
Ilustración 34 Agrupación de variables Tienda – Cantidad. Creación Propia.....	53
Ilustración 35 Distribución de Agrupación entre variable de Tienda – Cantidad. Creación Propia.....	54
Ilustración 36 Agrupación variable de Patología Cuidado Personal - Edad. Creación Propia. ....	55
Ilustración 37 Distribución Agrupación de Patología Cuidado Personal – Edad. Creación Propia.....	56
Ilustración 38 Diagrama Burbujas para variables de Valor Facturado, Patología y Edad en Tienda Calle 116. Creación Propia.....	58
Ilustración 39 Diagrama Burbujas para variables de Valor Facturado, Patología y Edad Tienda Plaza Américas. Creación Propia .....	58
Ilustración 40 Resumen Componentes Principales. Creación Propia.....	60
Ilustración 41 Clúster creado por PCA. Creación Propia.....	60
Ilustración 42 Productos alta rotación generado por PCA. Creación Propia .....	61
Ilustración 43 Distribución por Tienda y Edad según PCA. Creación Propia. ....	62
Ilustración 44 Gráfica distribución PCA por tienda y edad. Creación Propia.....	62
Ilustración 45 Facturación promedio en millones por rango edad según PCA. Creación Propia.....	63
Ilustración 46 Gráfica distribución facturación promedio por tienda y edad PCA. Creación Propia.....	63

Ilustración 47 Varianza para componentes principales. Creación Propia. ....	64
Ilustración 48 Muestra 1er Clúster. Creación Propia .....	65
Ilustración 49 Muestra 2do Clúster. Creación Propia.....	66
Ilustración 50 Muestra 3er Clúster. Creación Propia .....	67
Ilustración 51 Muestra 4to Clúster. Creación Propia .....	68
Ilustración 52 Distribución Buyer Persona. Creación Propia.....	70

### **Abreviaturas**

- CRM: Customer Relationship Management

## Glosario

- Base Datos Estructurada: Las bases estructuradas poseen un esquema determinado que permite identificar que variables y datos posee, además de poderse organizar y clasificar.
- Customer Journey: Etapas que ejecuta un consumidor para realizar una compra, tiene en cuenta la frecuencia y el objetivo de la misma.
- Data Mining: Es la exploración y valoración de gran cantidad de información o datos, los cuales se procesan con el objetivo de conseguir resultados o conclusiones sobre una problemática o hipótesis.
- Outliers: Dato u observación que se encuentra extraído erróneamente o es incoherente con el promedio de los datos al no tener consistencia con la muestra o población estudiada.
- Share: Participación de ventas que puede cubrir una empresa en un mercado determinado.
- Venta Cruzada: Venta relacionada con productos complementarios con respecto al interés inicial de los consumidores.

**Anexos**

Anexo 1. Mapa tiendas Locatel en Bogotá. Google (s. f.).....33

## Resumen Ejecutivo

Marco de trabajo para el desarrollo de la segmentación de clientes para Locatel

En la actualidad, la empresa Locatel cuenta con información de más de 1.9 millones de clientes sobre aspectos comerciales como lo son el historial de compras, la ubicación geográfica de la transacción y el valor de la misma. La empresa identifica una oportunidad para potencializar su desarrollo comercial a través de las ventas cruzadas (adquisición de productos bajo una necesidad inicial) con el uso del histórico de facturación obtenido, base de datos que recopila rasgos útiles de los consumidores para la generación de una segmentación por variable. Para poder realizar la estrategia de ventas cruzadas, el área de marketing requiere utilizar el reconocimiento e identificación de información clave que pueda utilizar para incentivar el consumo de productos complementarios a las patologías a tratar; por ejemplo, enviar información sobre el tratamiento de patologías, consejos de alimentación, condición física, prevención de enfermedades y descuentos en los productos adicionales que consumen los usuarios por patología.

Lo que se busca con el proyecto es la estructuración de la base de datos, la generación de un resumen interactivo y el modelo predictivo que permitan realizar la segmentación, información con la cual Locatel concluya cómo utilizar, estructurar o modificar su estrategia de CRM y el crecimiento de ventas cruzadas; la estructuración de la base de datos requiere el desarrollo de interacciones entre las variables del histórico de ventas y consultas propias del negocio. La construcción del resumen interactivo se enfocará en mostrar a las áreas comerciales, marketing y demanda el estado de la información recolectada, junto con sugerencias sobre adquisición y gobernanza de datos. Por parte del modelo predictivo, se espera que el resultado generado sea el resumen de la base con

información sobre la segmentación requerida para la estrategia del CRM. El propósito final del proyecto busca que el área de marketing pueda tomar la mejor decisión sobre el enfoque que se deberá utilizar para fortalecer la operación comercial, la estrategia de ventas cruzadas y los diferentes planes de acción que se deban implementar en su estrategia de CRM para cumplir el objetivo.

**Palabras clave**

Cliente, desarrollo comercial, CRM

## **Abstract**

### Framework for the development of customer segmentation for Locatel

Currently, Locatel Company possesses information on over 1.9 million customers pertaining to commercial aspects such as purchase history, geographical transactional data, and transactional value. The company identifies an opportunity to optimize its commercial development through cross-selling (acquisition of products based on an initial need) by utilizing the acquired billing history, a comprehensive database that compiles valuable consumer attributes for generating variable-based segmentation. In order to execute the cross-selling strategy effectively, the marketing department requires the implementation of recognition and identification mechanisms to extract key information that can be harnessed to encourage the consumption of complementary products associated with the ailments being addressed. For instance, disseminating information on disease treatments, dietary recommendations, physical well-being, disease prevention, and discounts on additional products consumed by users based on their respective conditions.

The project's objective is to structure the database, create an interactive summary, and develop a predictive model that will enable segmentation, thereby enabling Locatel to determine how best to structure or modify its CRM strategy and cross-selling growth. The database structuring process necessitates the establishment of interactions between variables derived from sales history and pertinent business queries. The development of the interactive summary will focus on presenting the current state of the collected information to the commercial, marketing, and demand departments, while also providing recommendations on data acquisition and governance. As for the predictive model, it is anticipated that the resulting summary of the database will include the requisite segmentation information

necessary for the CRM strategy. Ultimately, the project aims to empower the marketing department to make well-informed decisions regarding the approach that should be employed to fortify the commercial operation, enhance the cross-selling strategy, and implement various action plans within the CRM strategy to accomplish the defined objective.

**Keywords**

Customer, sales development, CRM.

## 1. Introducción

La cadena de retail para la salud y bienestar Locatel cuenta con presencia comercial en las ciudades de Bogotá, Medellín y Cartagena, en especial con la primera ciudad, cuenta con tiendas distribuidas a lo largo del territorio para cubrir la mayor necesidad demandada por el mercado. Los 16 años de operación que tiene en el país le ha permitido especializarse en el tratamiento de patologías y enfermedades complejas, lo que catapultó al grupo empresarial a ser una cadena de tiendas farmacéuticas y de cuidado personal con gran valor en el mercado. Su infraestructura le permite desarrollar estrategias comerciales y logísticas como distribución en máximo 2 horas para Bogotá, Medellín y Cartagena, ciudades en las cuales cuenta con tiendas físicas, además de envíos a nivel nacional aun sin tener establecimientos comerciales para su despacho como lo es la ciudad de Cali.

La experiencia en el mercado colombiano y el desarrollo comercial de sus establecimientos y productos, hoy en día le permite a Locatel tener una base de clientes aproximadamente de 2 millones de usuarios, quienes depositan su confianza y salud en la compañía, brindando datos socio económicos que le ayudarán a la empresa a estudiar y entender sus diversos nichos de mercado, establecer estrategias de crecimiento y generar valor en la cadena del cuidado de la salud en sus usuarios. Para poder establecerse en el mercado colombiano, Locatel ha generado alianzas con distribuidores y productores clave dentro de la cadena logística de medicamentos, dermocosméticos y consumo masivo; los puntos de venta se encuentran ubicados en lugares residenciales de las ciudades para mejorar el acercamiento al consumidor. Bajo esta situación, la compañía a través de su área de marketing plantea la necesidad de fortalecer su estrategia CRM, que le permita mejorar las ventas cruzadas y generar valor en sus usuarios, a través de comunicaciones proactivas

con indicaciones sobre el cuidado de sus enfermedades y uso preventivo de productos complementarios a los que normalmente consumen.

Con el proyecto de Business Analytics se busca desarrollar el análisis de información de la base de ventas, con el objetivo de generar una segmentación que permita la división de sus cliente según las variables relacionales que se tienen, información con la cual la compañía realizará el entendimiento de las condiciones patológicas y capacidad económica para desarrollar el potencial de consumo de las categorías de salud. La minería de datos se enfoca en generar información a la compañía que funcione como principal herramienta en la toma de decisiones frente a las estrategias comerciales que ejecutarán y conseguir el objetivo de rediseñar su proceso de CRM (aumento de ventas y generación de valor). Una vez se consolide la información necesaria se espera realizar reuniones de seguimiento, para entender si los resultados de las etapas del proyecto siguen orientados con el objetivo inicial.

La idea del procedimiento de minería es conocer el negocio y mercado, la información de la base de datos y su relacionamiento con variables exógenas a la base; luego se trabajará en unificar y limpiar la base de datos con fuentes externas (si se requiere conseguir para el resultado esperado). Además, dentro del conocimiento del negocio será necesario identificar patrones para deducir el perfil de los clientes que la compañía ha podido fidelizar, no desde el detalle de las particularidades de cada usuario, sino desde el usuario ideal o general para la segmentación, con esto poder definir los grupos y enfocar el esfuerzo operativo desde la óptica comercial. El proyecto debe ayudarle a responder las preguntas a Locatel de cómo realizar las posibles estrategias de segmentación y el enfoque de su estrategia CRM, conocer el posible crecimiento de sus ventas a través de un modelo

prospectivo, entendimiento de las generalidades que poseen sus usuarios y tendencias del mercado en general.

## 2. Objetivos

Desarrollar un modelo de segmentación de los clientes que utilice información sobre el nivel de ventas y exponga a los líderes de marketing resultados útiles para que Locatel tome decisiones sobre el desarrollo de su estrategia CRM y el crecimiento de las ventas cruzadas.

Para lograr el objetivo general del proyecto, se establecieron los siguientes objetivos específicos:

- Realizar un análisis descriptivo de las bases de datos para realizar la contextualización y comportamiento comercial que tiene Locatel, a través de data storytelling y minería de datos.
- Construir el modelo no supervisado para la segmentación definida, con el objetivo de establecer los grupos de clientes necesarios en la toma de decisión de Locatel sobre la estrategia CRM, por medio de visualización de datos y analítica descriptiva.
- Crear un reporte para el área de marketing, en el que se detalle la información conseguida a través del modelo, junto con las recomendaciones finales y paso a seguir para continuar el proyecto, utilizando los conceptos de analytics lifecycle y analytics product management.

### 3. Alcance

El proyecto tiene como objetivo generar opciones de segmentación para Locatel, según la necesidad expuesta por la compañía y funciones en la definición de su estrategia comercial con respecto a la implementación del CRM. El proyecto no otorgará la estrategia comercial, proyección de ventas, predicción de crecimiento de los productos comercializados ni los lineamientos para la gestión del CRM, ya que dicha visión debe ser generada en sus procesos operativos por las distintas áreas de la compañía; el proyecto generará el resumen de la información para soportar la toma de decisión por cada una de las partes involucradas. Para el modelo se trabajará con una parte de la base de datos general de Locatel, no se utilizará la totalidad del espectro de clientes por decisión empresarial y con esta muestra se creará la prueba de concepto.

El proyecto de segmentación se enfocará en el uso de la información sobre los productos comercializados y las patologías a las cuales está enfocado su uso farmacéutico. Se requiere hacer la segmentación bajo una enfermedad específica, dado que realizar un estudio de todas las variables no es viable para su análisis y exploración, lo que no generaría valor para la compañía. Por otra parte, el proyecto no entregará información sobre los costos de la puesta en marcha, los cambios operativos y el proceso comercial a desarrollar por parte de Locatel para la implementación de la opción que se generará en el modelo, ya que la decisión de la estrategia a implementar la decidirá Locatel como Product Owner bajo la información generada en la segmentación.

#### 4. Metodología

La metodología seleccionada para realizar el proyecto de minería de datos es el marco de trabajo CRISP (Chapman et al., 2000), dado que la metodología permite flexibilidad en el momento de realizar las actividades en cada una de sus fases, ya que no es excluyente el finalizar la etapa anterior para iniciar la siguiente en su orden. La metodología se enfoca en la generación de soluciones a problemáticas existentes a través de la creación de modelos con los que se pronostican posibles escenarios y así poderse anticipar a los resultados, buscando intervenir áreas y procesos a partir de los datos recolectados. A través de CRISP se logra entregar avances del proyecto, validar con el Product Owner de Locatel si los resultados que se presentan concuerdan con la meta trazada para el proyecto y genera optimización en las entregas.

El desarrollo del proyecto se realizará bajo las 6 etapas del marco de trabajo anteriormente mencionado, las cuales se detallan a continuación:

- Entendimiento del negocio: En esta fase se realizó una reunión con líderes de Marketing, Logística y Recursos Humanos, con el objetivo de entender la misión y visión de la compañía; el estado actual del negocio, proyectos que planean realizar y el objetivo del proyecto. Se alinearon las expectativas del proyecto, la información requerida para el estudio y los pasos siguientes en el proyecto. Como parte del conocimiento, se realizó una visita a la tienda de Cedritos, lugar en el cual se reconoció la estrategia de venta de las tiendas, el proceso operativo y la dinámica de góndolas, promocionales y ubicación de las diferentes categoría de producto.
- Entendimiento de los datos: Para la comprensión de la información, se realizó un ejercicio de visualización, para reconocer las variables de la base de datos, la cantidad

de registros almacenados, los outliers o fenómenos que se presentan en la información y el resumen general del desarrollo de Locatel en el periodo estudiado.

- Preparación de los datos: Una vez entendiendo el negocio y los datos, se procede a retirar la información errónea de las diferentes variables a trabajar, como lo son la edad, el género, las cantidades vendidas y el valor facturado. Además, como trabajo externo a la información de la compañía, se creó una base de datos sobre la patología que trata cada uno de los medicamentos, con el fin de utilizar dicha información en la segmentación a realizar. De los 1,02 millones de registros se dejaron alrededor de 580 mil, dadas las novedades de información que se encontraron.
- Modelación: Se generó el modelo predictivo para realizar la segmentación. En el cual se refleja una muestra de 230 mil registros, generando 4 grupos de segmentación.
- Evaluación: Se comparte los grupos generados con las características de cada uno, para validar si es viable la segmentación realizada.
- Despliegue: Por parte de Locatel, se entregará el modelo a Locatel para que la compañía decida si lo implementan en su estrategia de CRM.



## **6. Descripción de las fuentes de información**

### **6.1. Resumen:**

La información para el proyecto se extraerá principalmente de la compañía, ya que Locatel cuenta con el histórico de sus clientes; información necesaria para el objetivo del proyecto. El proceso de recolección de datos personales para la fidelización por parte de la compañía está fundamentado en la política de tratamiento de datos de Habeas Data y la Ley 1581 del 2012 que define los parámetros a tener en cuenta en el momento de realizar la recolección de manera idónea; informando el objetivo del proceso, quien y como se encargará de su almacenamiento y control. Según ello, la compañía realiza un formulario oficial para todos los clientes que desean inscribirse en la compañía, pueda aprobar el proceso de recolección (Función Pública, s. f.).

En el caso de las fuentes secundarias, se utilizará la teoría de segmentación de clientes, nivel de inversión del sector farmacéutico, entendimiento de la demografía de Bogotá, desempeño comercial de Locatel y el contexto de la competencia en el sector. Se construyen bases de datos propias para el proyecto con información exógena a la de la compañía como lo es la base de datos de las patologías por producto y los días feriados del año 2022 en Colombia. El análisis permitirá conectar los hallazgos de los datos con la realidad que experimenta el sector en el mercado colombiano y la participación de Locatel para enfocar el objetivo del modelo.

### **6.2. Base Datos Patología:**

Al reconocer las variables de la base original, se evidenció la necesidad de conectar los productos con las patologías que tratan, dado que la información no se encuentra adherida en la base de datos, por lo que se realizó una búsqueda externa de cada uno de los

productos incluidos en la base. A pesar que varios artículos comercializados por Locatel no necesariamente sirven para combatir enfermedades (como por ejemplo los alimentos o artículos de cuidado personal), con la investigación realizada se logró categorizar todos los productos vendidos con su respectiva clase de patología. Como se ve en el Anexo técnico A. Base de datos medicamentos por patología, se recopilaron más de 1 millón de registros categorizados en cerca de 220 patologías posibles.

La siguiente es una muestra de la tabla de patologías creada para el proyecto:

<b>EAN</b>	<b>Nombre Producto</b>	<b>Patología</b>
7703991000020	ZYPREXA ZYDIS TAB.10MG X14	Antipsicóticos
300090090074	XANAX TAB.1MG X30	Ansiedad
7861028101215	WELLBUTRIN XL TAB.300MG X30	Antidepresivos
7612797172698	VOLTAREN TAB.150MG X7	Dolor Muscular
4054839015915	VIVERA SUS.ORA.SOB.X8	Digestivo
7702635143512	TRILEPTAL 6% SUS.ORA.X100ML	Epilepsia
7702057711818	SULTAMICILINA MK TAB.750MG X20	Otitis
7707236670177	RELESTAT SOL.OFT.0.5MG/ML X5ML	Oftalmología
33984023185	QUERCETIN COMPLEX SOLGAR CAP.X50	Sistema Inmune
7706440000206	NECROTON CAP.X8	Disfunción Hepática
7703546042000	MUXOL NINOS JBE.X120ML	Bronquitis
7702207127902	LUVOX TAB.100MG X15	Trastorno Psicótico
7707188880167	KLEAN-PREP PVO.SOB.X1	Laxante
7703145906000	INVEGA TAB.6MG X14	Esquizofrenia
7703153028121	GOUTEX CAP.120MG X30	Urinario
7702057010843	FENCAFE TAB.100MG X50	Migrañas
7704039111678	ESOPRAX CAP.20MG X14	Ulcera
7703721148114	BOSU SPORTIVA MINI	Circulación
7591044005623	ARAVA TAB.20MG X30	Artritis

Ilustración 2 Muestra de la base de datos por medicamento y patología. Creación Propia.

Una vez se realiza la unificación de la base de datos de Patologías con la base de datos original de Locatel, se establece la distribución de cada una de las patologías dentro de los datos, arrojando que la patología de Cuidado Personal tiene aproximadamente la tercera parte de las ventas realizadas en el periodo estudiado. Lo que presenta la ilustración

3, es que gran parte de las ventas de Locatel no están garantizadas necesariamente por medicamentos, sino por productos de otro tipo de categorías, los cuales la compañía ha incluido en sus establecimientos para fortalecer la estrategia comercial; familias de productos como son los alimentos (galletas, bebidas, cereales, etc.) maquillaje (labiales, desmaquillantes, mascarillas faciales, etc.), productos dermocosméticos y protección personal, como se evidencia a continuación:

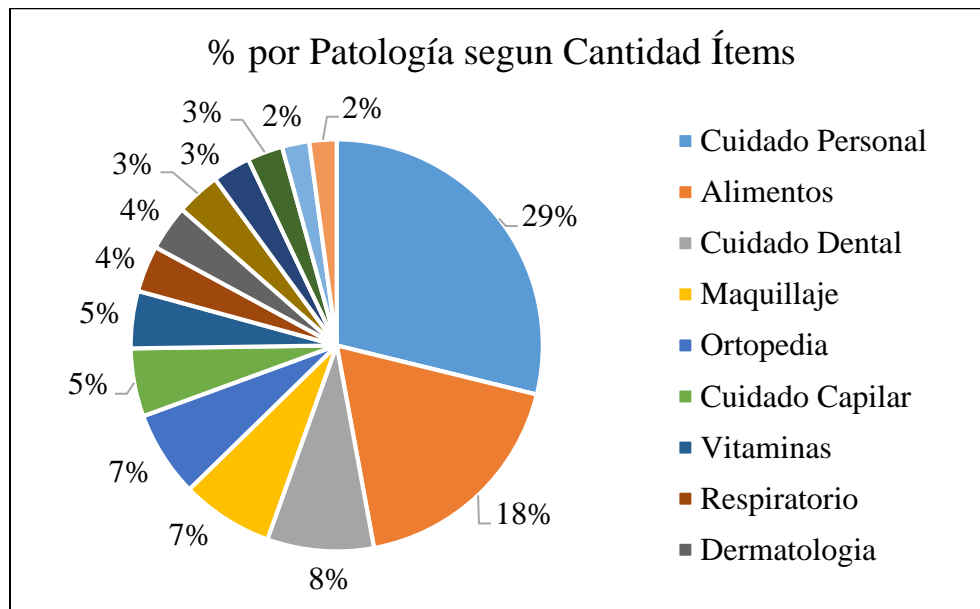


Ilustración 3 Distribución de la base de datos por Cantidad de Items y Patología. Creación Propia.

### 6.3. Análisis Descriptivo Base Datos Locatel:

Para lograr el entendimiento de la información, se realizó un análisis descriptivo de las variables que componían la base de datos entregada por Locatel, las cuales reflejan patrones que permitirán la explicación de algún comportamiento específico. La tipología de información que otorga la compañía está enfocada en el detalle de las transacciones comerciales que se realizan en los canales de distribución de la compañía. Como se puede evidenciar en la Ilustración 4, en la base de datos se encuentran datos numéricos (como lo

son cantidades, valores de compra, edad del usuario) junto con datos categóricos (como lo son las categorías de medicamentos; Farmacéuticos, No Farmacéuticos y Otros). Dicha información permite que en el desarrollo del modelo se pueda combinar variables que presenten una correlación llamativa para el algoritmo ejecute de forma efectiva la segmentación de los diferentes grupos.

<b>Variable</b>	<b>Tipo de Dato</b>	<b>Cant. Datos</b>	<b>Cant. Datos Vacíos</b>
Id Venta	Numérico	1.020.710	0
Fecha Venta	Categórico	1.020.710	0
Ciudad	Categórico	219.098	801.612
Edad	Numérico	1.020.710	0
Género	Categórico	612.042	408.668
Desc. Nivel 1	Categórico	1.020.710	0
Desc. Nivel 2	Categórico	1.020.710	0
Desc. Nivel 3	Categórico	1.020.710	0
Desc. Nivel 4	Categórico	1.020.710	0
Desc. Nivel 5	Categórico	1.020.664	46
EAN	Numérico	1.020.710	0
Nombre Producto	Categórico	1.020.710	0
Cantidad (Ítem)	Numérico	1.020.710	0
Valor Venta sin IVA (Ítem)	Numérico	1.020.710	0
Tienda	Categórico	1.020.710	0
Patología	Categórico	1.020.710	0

Ilustración 4 Resumen tipo de información de la base de datos Locatel. Creación Propia.

El peso de los datos vacíos de cada variable da a entender la magnitud de información que se dejaría de tomar en cuenta para el desarrollo del modelo predictivo. En el resumen anterior, se evidencia que la variable Ciudad solo cuenta con cerca de 220 mil registros, la Descripción Nivel 5 con 1.021 mil y la variable Género con 612 mil registros. Para analizar el paso a seguir en el modelo, se determina el porcentaje por variable de

valores vacíos para todas las variables de la base de datos, en especial para las tres mencionadas anteriormente:

Id Venta	0.000000
Fecha Venta	0.000000
Ciudad	78.534745
Edad	0.000000
Genero	40.037621
Desc. Nivel 1	0.000000
Desc. Nivel 2	0.000000
Desc. Nivel 3	0.000000
Desc. Nivel 4	0.000000
Desc. Nivel 5	0.004507
Ean	0.000000
Nombre Producto	0.000000
Cantidad (Item)	0.000000
Valor Venta sin IVA (Item)	0.000000
Tienda	0.000000
Patología	0.000000

Ilustración 5 Porcentaje de datos vacíos en la base de datos Locatel. Creación Propia.

El primer resumen exploratorio informa que la base de datos no cuenta con toda la información requerida en cada uno de sus registros; las variables Ciudad, Género y Descripción Nivel 5 tienen 78%, 40% y 0.005% de datos vacíos respectivamente, lo que obliga a realizar una limpieza de datos para estas columnas. Para el proceso de limpieza en la variable Género se busca conservar la mayor cantidad de registros posible, por lo que se utiliza el algoritmo K-Nearest Neighbor el cual se utiliza para predecir valores según los datos cercanos del registro. Además de la variable Género, para el anterior algoritmo mencionado se utilizaron como complemento la Cantidad Ítem y Edad del registro. Como se muestra en la ilustración 6, el resultado permite asumir que los datos más cercanos para los registros vacíos en su mayoría son de género masculino, por lo que los datos vacíos para

la variable Género se modificarán con la categoría “M” para los hombres y se completan los datos para dicha variable.

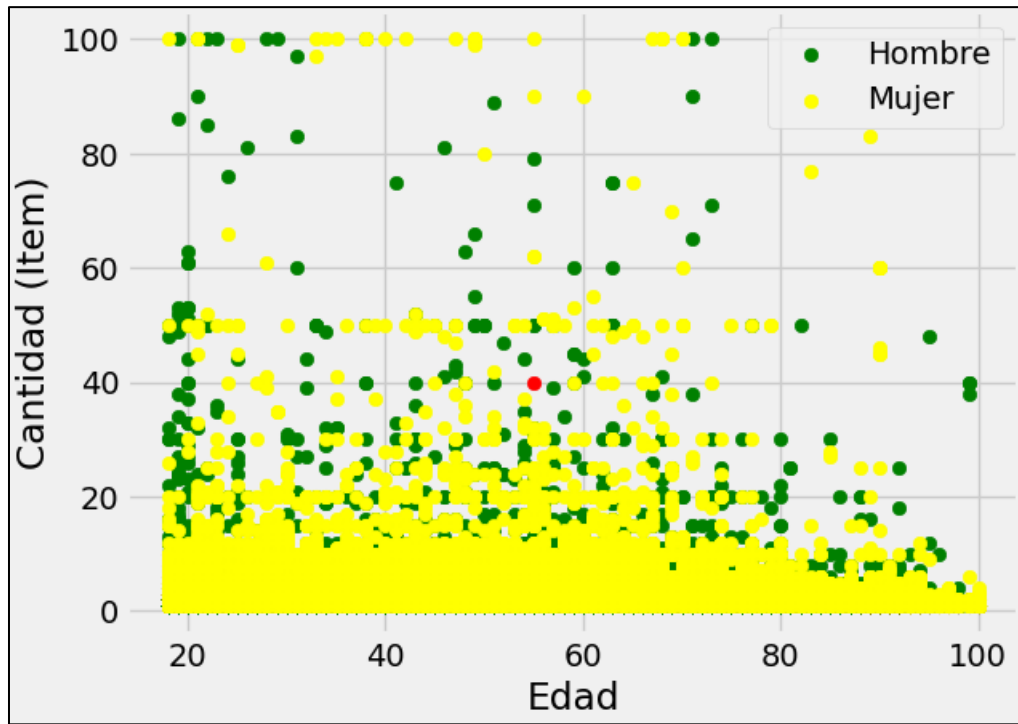


Ilustración 6 Algoritmo KNN variable Genero. Creación Propia.

Por otra parte, las variables Ciudad y Desc. Nivel 5 también fueron reprocesadas dentro de la limpieza de datos. Bogotá al ser la ciudad que tiene la mayoría de las tiendas de la compañía y en la cual se enfocará el estudio del proyecto, se utiliza como dato para diligenciar los registros vacíos de la variable. Por parte de Desc. Nivel 5 se deja el dato “Na” dado que los productos que están referenciados en dicho registro no contemplan la necesidad de tener una quinta categoría, sino que se explican y categorizan por las anteriores 4 grupos que Locatel maneja, por lo que no se podría asignar una categoría en específico.

Con las anteriores variables procesadas, se revalida la información de las variables numéricas, para entender valores nulos o que pudiera dañar el estudio de la muestra al ser

anormales (outliers), por lo que se realizó un resumen con los principales datos numéricos como percentiles, promedio, mínimos y máximos, resumen que se presenta a continuación:

	Edad	Cantidad (Item)	Valor Venta sin IVA (Item)
<b>count</b>	1020710	1020710	1020710
<b>mean</b>	81	1	29315
<b>std</b>	46	2	94611
<b>min</b>	-5955	-400	-15973151
<b>25%</b>	44	1	5900
<b>50%</b>	71	1	15500
<b>75%</b>	122	1	35756
<b>max</b>	122	600	31550100

Ilustración 7 Resumen de datos numéricos de la base de datos Locatel. Creación Propia

Luego del resumen numérico, se puede determinar que existen datos nulos y anormales en las variables de la base, los cuales no pueden ser utilizados dada su racionalidad conceptual; por ejemplo las edades, productos y precios con valor negativo, así como edades de más de 100 años en el 25% de los datos registrados (a pesar de ser posible que un usuario tenga una edad prolongada, la probabilidad de tener un cuarto de muestra con dicha edad es muy poco probable) lo que sugiere un error en el proceso. El resumen se analizó con Locatel y la compañía informa haber tenido posibles errores en el momento de digitar la información de la suscripción, por lo que puede haber datos nulos, como se presentaron anteriormente.

Con el contexto inicial de la base, se desarrolló la limpieza de la información para seleccionar los datos que se ingresarán en el modelo predictivo. Para lo cual, se eliminaron los datos anormales en la variable de Cantidad y Valor Facturado, dado que para el modelo

no es viable tener un valor negativo cuando debe ser positivo (por ejemplo, si se venden unidades de un producto específico, las cantidades deben ser mayores a 0, no tendría sentido si es menor a 0) y dejar el promedio de los datos no refleja la acción que se realizó con el producto. Más adelante, se procedió a intervenir la variable Edad, la cual tiene datos menores a 18 años (dato que no es viable ya que solo pueden registrar usuarios que sean mayores de edad en Colombia) y mayores a 100 años, por lo que se estableció el rango entre 18 y 100 años, para no eliminar el 40% de la totalidad de los registros. A continuación se muestra el resultado final de la intervención con la limpieza de datos en la base final:

	Edad	Cantidad (Item)	Valor Venta sin IVA (Item)
<b>count</b>	514614	514614	514614
<b>mean</b>	47	1	32240
<b>std</b>	17	2	78568
<b>min</b>	18	1	1000
<b>25%</b>	34	1	7983
<b>50%</b>	47	1	17150
<b>75%</b>	60	1	37950
<b>max</b>	100	100	15973151

Ilustración 8 Limpieza de datos numéricos de la base de datos Locatel. Creación Propia.

Dado el resultado del proceso para la limpieza de datos, se inicia el análisis de la información encontrada. La primera variable a analizar es la Desc. Nivel 1 dado que es la primer categoría que tiene Locatel para agrupar los productos; si son medicamentos, suplementos de salud o productos para uso diferente a una necesidad farmacéutica. La compañía realiza dicha agrupación dado que estrategia comercial en los establecimientos de

venta permite el desarrollo farmacéutico y de dermocosmética, aprovechando la necesidad creada en el mercado por la combinación de cuidado personal y salud general.

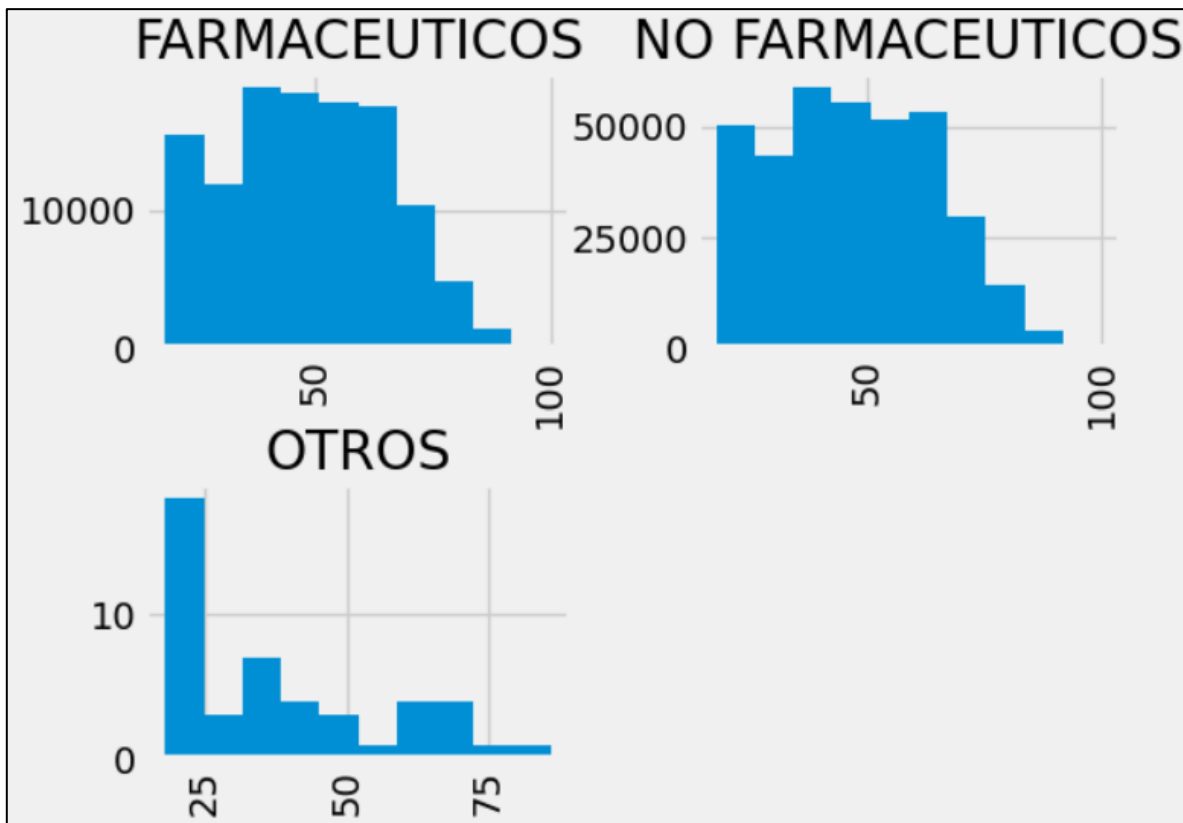


Ilustración 9 Distribución variable v por Edad y Género Locatel. Creación Propia.

Dada la ilustración 9, se puede evidenciar que el promedio de edades de los consumidores de Locatel está en el rango de 40 a 55 años (según la ilustración 8 la media de las edades es de 47 años). Dentro de las subcategorías de la variable analizada, se evidencia que los productos “No Farmaceuticos” son los que tienen mayor representación en las ventas que realiza Locatel, con respecto a los “Farmacéuticos”. El análisis preliminar que se detalla en la ilustración puede ser que Locatel ha enfocado su estrategia comercial y de marketing en generar un amplio mercado de productos y cobertura de necesidades, estrategia por la cual la compañía se considera un retail farmacéutica y no cadena de droguerías.

#### 6.4. Análisis Ticket Compra Locatel Marzo-Mayo 2022

Para entender el negocio desde varios puntos de vista, se analizó la facturación por día y la cantidad de unidades por producto que se pudo atender desde los diferentes canales en el periodo de tiempo determinado, por lo que se comparte a continuación el resumen de consumo por día de la semana de los 3 meses. Con la información recopilada, se identifica un patrón en la compra de los fines de semana, dado que los domingos y lunes se presenta un alto consumo con respecto a los demás días (con excepción del miércoles).

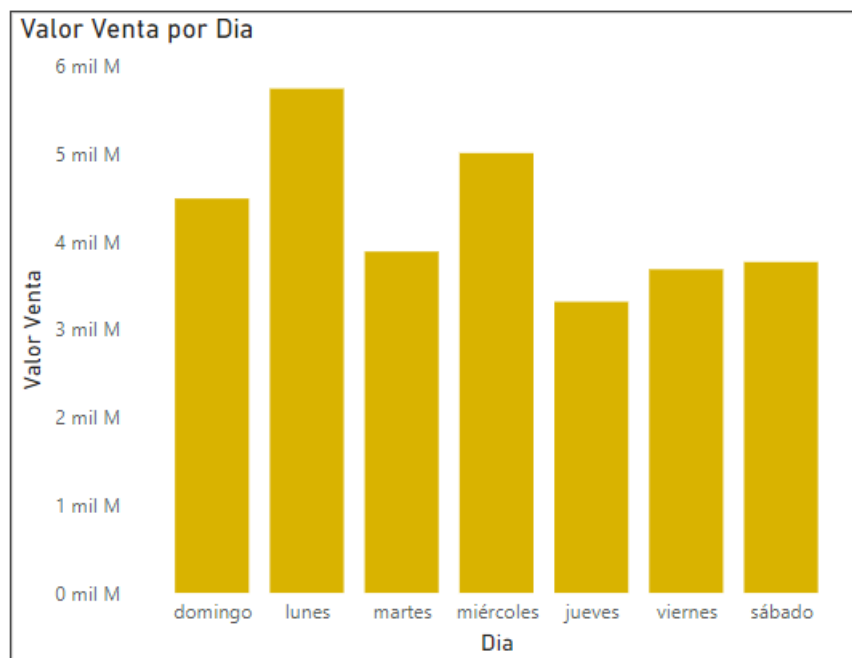


Ilustración 10 Distribución ventas por día Locatel. Creación Propia.

La hipótesis generada es que dichos días tienen un pico elevado de compra por la dinámica social de los días previos (jueves, viernes, sábado y domingo cuando el lunes siguiente es festivo) las personas generalmente se reúnen con amigos, compañeros y familiares para realizan comidas, viajar o ingerir productos a base de alcohol. Para el último bimestre del año 2022 en Bogotá, se registró un número cercano a 30 personas intoxicadas

por la ingesta de bebidas alcohólicas adulteradas, dejando como resultado la afectación de la visión, el sistema nervioso y respiratorio (Bogota.gov.co. 2022). El desarrollo de la actividad social puede resultar en intoxicaciones, ingesta indebida de licor o dolores físicos, efectos que aumentarían el consumo de hidratantes, analgésicos, o alimentos específicos, lo que podría justificar las ventas de Locatel en estos dos días.

A continuación, se comparte el peso por familia de producto para sábados y domingos, en la cual se encuentran las familias de mayor participación en ventas para Locatel, en las cuales se encuentran Éticos (medicamentos bajo receta médica), OTC (medicamentos de venta libre) y Soluciones de Salud (Instrumentos para lesiones, desinfección y protección de heridas).

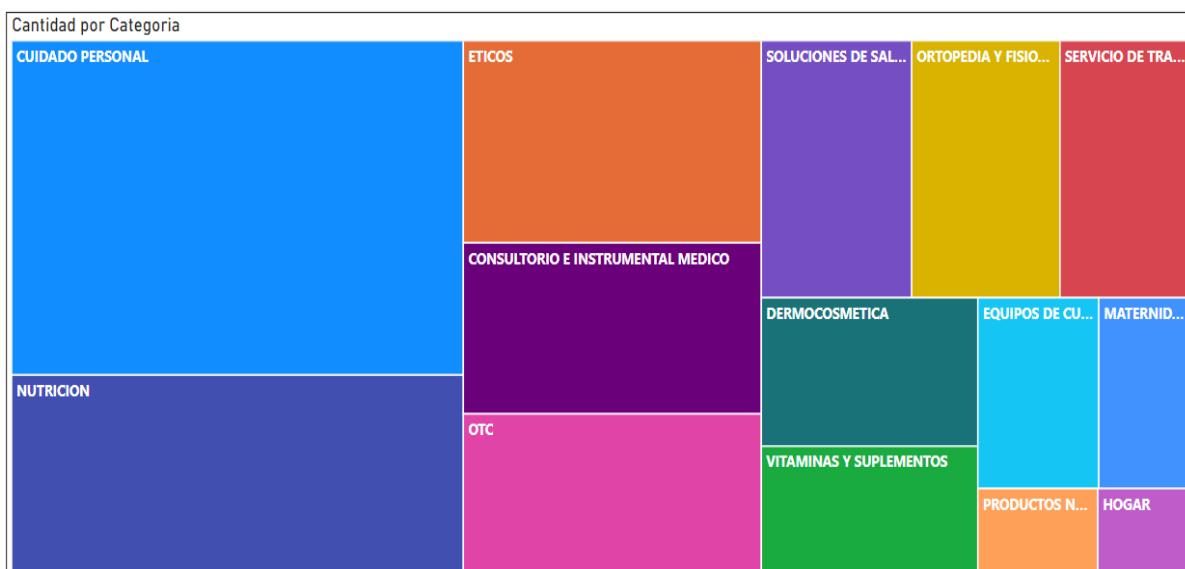


Ilustración 11 Distribución de ventas por Categoría generada por Locatel. Creación Propia.

Sin embargo, al validar los productos consumidos para los fines de semana resulta ser la categoría de cuidado personal (maquillaje, cuidado capilar, cuidado oral, etc.) la que en su mayoría se vende en este periodo por Locatel; el fenómeno que se explica en la base refuta la hipótesis anteriormente planteada y negaría que el comportamiento social pueda

augmentar las ventas de Locatel para productos como analgésicos, hidratantes o anticonceptivos. El hallazgo se explicaría ya que los usuarios adquieren la categoría de Cuidado Personal para su uso durante toda la semana que inicia, en especial para los compromisos laborales, académicos o sociales a desarrollar. La información expuesta en la descripción de la ilustración permite entender que el mercado de dermocosmética está en crecimiento y es vital para Locatel permanecer en ella e innovar con respecto a los productos y canales de distribución de la categoría.

### 6.5. Análisis Compra Por Género En Locatel:

Dado el fenómeno de compra que se expuso anteriormente, se quiere validar porque algunas familias de producto tienen mayor venta que otras y cuál es el tipo de usuario que en su mayoría lo adquiere. Se analizará si el género femenino es quien en su mayoría aporta en las ventas de la compañía y que tipo de productos consume. Por lo anterior, se presenta a continuación el resumen de ventas por familia de producto (subcategoría de agrupación principal categoría de producto) para el análisis propuesto:

Familia Producto	Valor Facturado	Cantidad
HUMECTANTES	595375511	34.575,00
ANTI-ARRUGAS	356660424	10.472,00
REGULAR ADULTOS	270342790	13.194,00
MANUALES	184282674	12.664,00
ESPECIALIZADAS ADULTOS	156661159	9.342,00
TRATAMIENTO	118581624	6.858,00
SPRAY	110758891	6.411,00
PESTANINAS	109736099	4.933,00
BASES	99762076	3.617,00
CORRIENTES ADULTOS	99445293	9.569,00
DESMAQUILLANTES	98264860	7.068,00
MAQUINA CUCHIL.-RPTO.AFEITAR	89378108	3.993,00
<b>Total</b>	<b>4845171656</b>	<b>294.817,00</b>

Ilustración 12 Distribución por familia de productos generada por Locatel. Creación Propia

Al analizar la información sobre la categoría con mayor facturación en el periodo estudiado, la Ilustración 11 expone que Cuidado Personal es la patología que obtiene el mayor valor, aproximadamente el doble de lo que logró facturar la patología de Nutrición, ubicada en el segundo lugar. En el resumen que se presenta en la Ilustración 12, se evidencia la distribución que dicha categoría tiene por artículo, detalle que presenta a los humectantes y anti-arrugas como las dos familias de mayor relevancia.

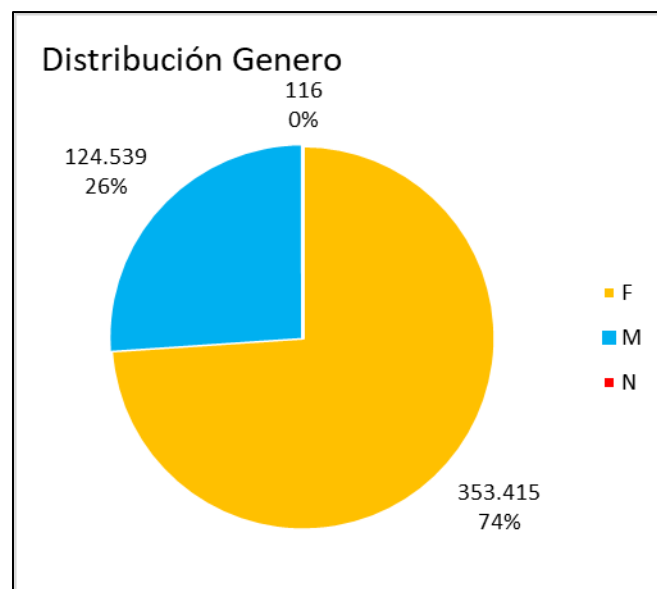


Ilustración 13 Distribución por género de ventas Locatel. Creación Propia.

El resultado de la ilustración 12 se puede explicar por la necesidad del género con más peso en la base de datos, dado que en su mayoría, las mujeres son quienes aceptan su ingreso a la fidelización que la compañía ofrece. A demás, este género triplica en número de usuarios al masculino en el periodo estudiado como lo explica la Ilustración 13 y es el grupo poblacional al cual se enfoca el uso de productos como humectantes, maquillaje y cuidado facial. A pesar que hoy en día los hombres se preocupan por su estado físico y

consumen esta familia de productos, el género femenino es quien mayor detalle presta a su condición física y la oferta de productos es mayor para este grupo poblacional.

### 6.6. Análisis Ticket Compra Locatel Marzo-Mayo 2022:

Se busca analizar el desarrollo comercial de las tiendas en los 3 meses contemplados en la base. Por lo cual, se presenta a continuación el resumen por mes para validar la existencia de algún patrón de compra:

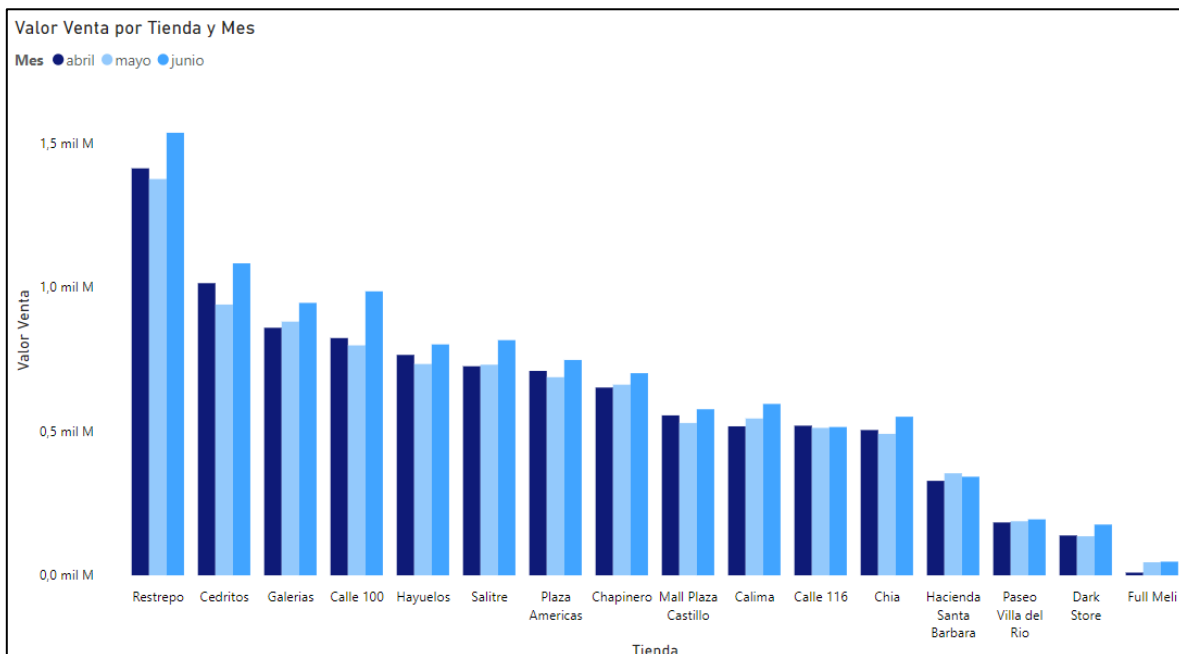


Ilustración 14 Distribución de ventas por tienda y mes Locatel. Creación propia

Se puede evidenciar que no existen desviaciones en la compra para las tiendas y meses; el comportamiento de las tiendas es similar en todos los establecimientos (a pesar del presentado en Calle 100 para junio, tienda en la cual se vende COP 1.000 millones con respecto a los otros dos meses en donde se facturaron 800 millones). Las ventas por mes y tiendas se mantienen estables y es junio el mes que mayor facturación expone para todas las tiendas, incluidas Dark Store (Bodega que factura todos los pedidos realizados por la página web de la compañía) y Full Meli (Pedidos atendidos por la plataforma de

MercadoLibre). El fenómeno de venta en el mes de junio se podría explicar intuyendo que es un periodo vacacional en el que se puede aumentar el consumo de algunas familias de producto relacionadas con dicho periodo de tiempo.

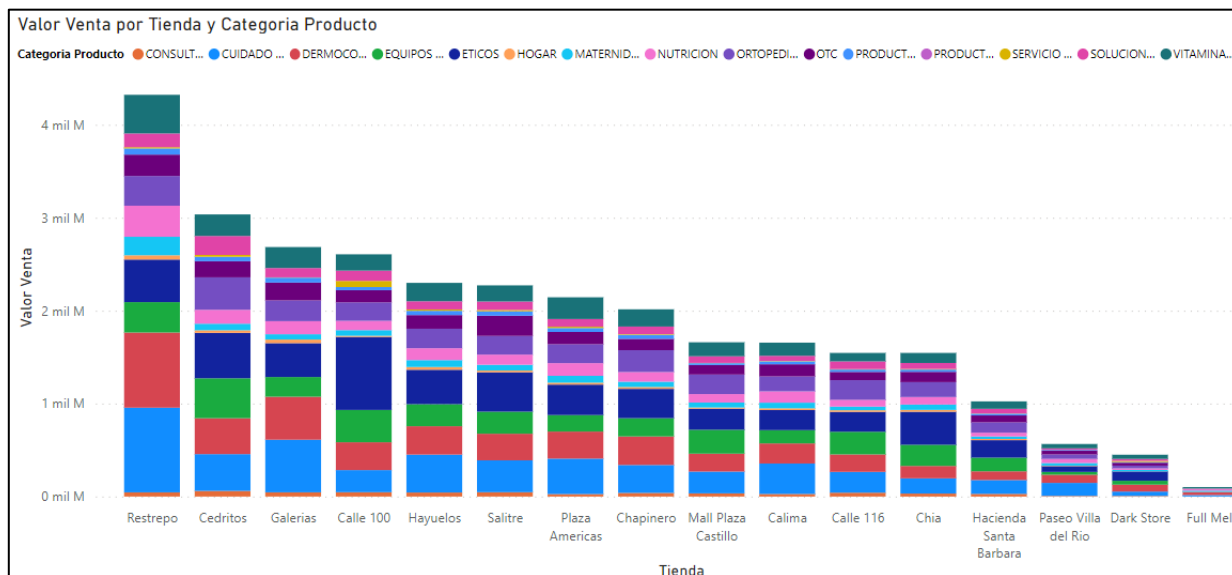


Ilustración 15 Distribución de ventas por tienda y categoría Junio Locatel. Creación propia

Dado el resumen anterior se evidencia que las categorías de Cuidado Personal y Dermatología son las categorías más representativas en la distribución de ventas de las tiendas para junio. En estas dos categorías se encuentran las familias de producto como protectores solares, cremas corporales, humectantes y filtros solares. El consumo de las familias de producto mencionadas se puede asociar a la realización de viajes y cuidado general de la piel que se presenta en el periodo de tiempo evaluado, el cual está referenciado a la primera jornada vacacional del año en el país.

En la validación de los factores exógenos, para la subcategoría de Éticos (medicamentos) las tiendas que mayor venta desarrollan son Calle 100, Chía y Restrepo. El porcentaje que cada una de estas tiendas presenta en la subcategoría se podría entender por la ubicación en la que se encuentran; por ejemplo, la tienda Restrepo se encuentra a 88

metros de distancia del Centro Medico Sanitas EPS, la tienda Chía se encuentra a 1.3 kilómetros de la Clínica Marly, la tienda Calle 100 está en un sector céntrico de instituciones médicas en el que se ubica equidistante al Centro Medico Chicó Navarra, Centro Medico Paralela 103, Compensar calle 94 y la Fundación Santafé. (Google. s. f.).

#### **6.7. Análisis Consumo Días Feriados 2022:**

Para entender el flujo comercial del sector farmacéutico y cosmético, se realizó un análisis del consumo en los días festivos del periodo estudiado. En esta hipótesis se busca determinar si existe una categoría que refleje un pico diferencial de consumo por la fecha o si se mantiene las ventas bajo el promedio general de la compañía. Como contexto inicial, se comparte a continuación las fechas feriadas que se presentaron en el país para el año 2022, los cuales se toman en cuenta desde los lunes hasta los sábados, dado que los domingos ya se consideran festivos en el país:

<b>Fecha</b>	<b>Día</b>	<b>Nombre Festivo</b>
1/01/2022	Sábado	Año Nuevo
10/01/2022	Lunes	Reyes Magos
21/03/2022	Lunes	San José
14/04/2022	Jueves	Jueves Santo
15/04/2022	Viernes	Viernes Santo
30/05/2022	Lunes	Ascensión
20/06/2022	Lunes	Corpus Cristi
27/06/2022	Lunes	Sagrado Corazón
4/07/2022	Lunes	San Pedro y San Pablo
20/07/2022	Miércoles	Día de Independencia
15/08/2022	Lunes	Asunción
17/10/2022	Lunes	Día de la Raza
7/11/2022	Lunes	Día de los Santos
14/11/2022	Lunes	Independencia Cartagena
8/12/2022	Jueves	Inmaculada Concepción

Ilustración 16 Listado de festivos para Colombia 2022. Creación Propia.

En Colombia, la semana con más festivos en el 2022 fue la segunda semana del mes de abril, la cual se conoce como Semana Santa. Durante toda esta semana se conmemora una variedad de celebraciones católicas, en especial el jueves y viernes por el simbolismo religioso e histórico que albergan. Un fenómeno que últimamente se presenta en esta semana es que las personas deciden realizar viajes, visitar familiares o amigos y no necesariamente se restringen al descanso y reflexión que la fecha demanda, lo que significa que categorías referentes a las actividades de esparcimiento deberán aumentar su promedio de venta. A continuación se presenta la distribución de venta por categoría para los dos feriados mencionados anteriormente:

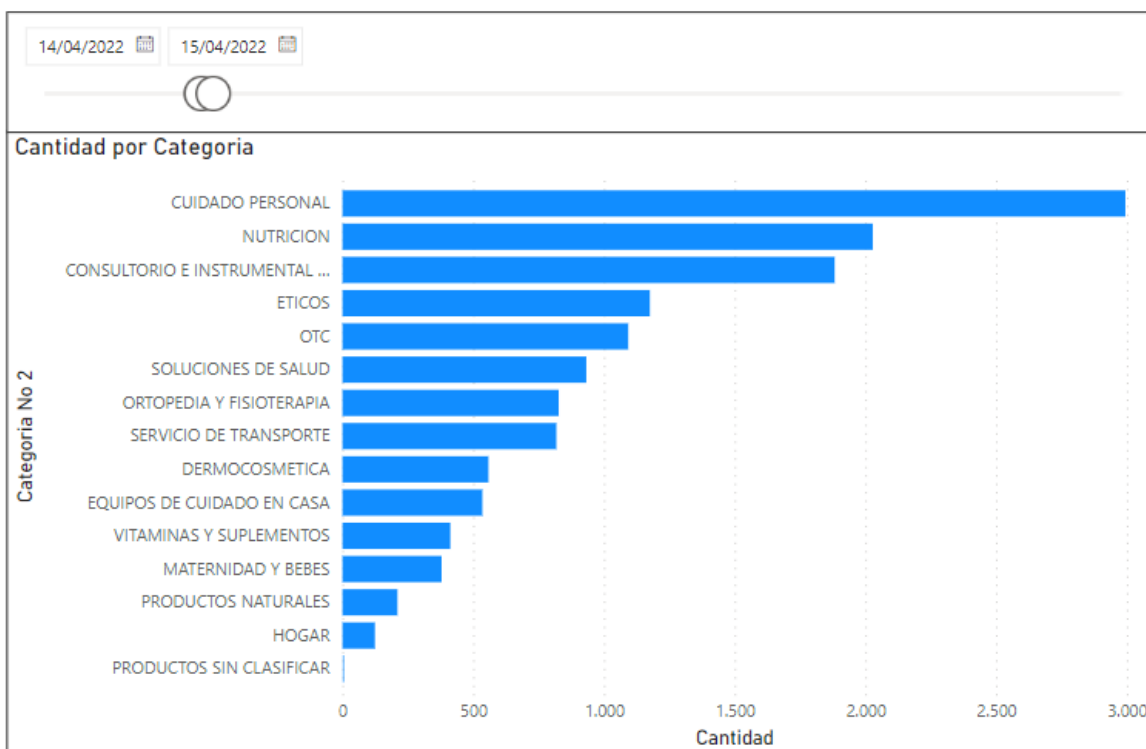


Ilustración 17 Distribución de ventas por cantidad y categoría del 14 y 15 abril 2023 para Locatel. Creación Propia.

Se esperaría que las categorías en general tengan un bajo consumo y sean las líderes aquellas que no sugieran su uso en actividades sociales, dado el significado espiritual que estos dos días representa para un alto porcentaje de la población colombiana. Sin embargo, se puede evidenciar que la categoría Cuidado Personal lidera las ventas en Locatel para el periodo de días estudiado. En esta oportunidad, la categoría Éticos pierde el 3er lugar que sostiene durante el periodo total de la base e ingresa en el top 3 la categoría Consultorio e Instrumental. Las familias más importantes de la tercera categoría son tapabocas, guantes, jeringas y catéteres, lo que permitiría deducir que las urgencias aumentaron al necesitar productos para desinfección y estabilización de salud, intuyendo el traslado de personas o ingestas excesivas de alimentos.

Familia Producto	Valor Facturado	Cantidad
PROTECTORES SOLARES	12327705	220,24
HUMECTANTES	8225928	346,00
V-VARIOS	6861852	89,00
ANTI-ARRUGAS	6503691	110,00
HIDRATANTES	6308170	94,00
A11X-OTRAS VITAMINAS	5253337	95,00
ESPECIALIZADO	4766654	156,00
S01K-LAGRI.ARTIF/LUBRIC.OCULAR	4762150	217,00
SILLAS TRANSPORTE	4725000	4,00
TIRAS	4362810	51,00
N02B-ANALG NO NARC.ANTIPIRET.	4337350	190,96
A11A-POLIVITAMIN.CON MINERAL.	3678458	74,00
AUTOMAT.Y SEMIAUTOMAT.BRAZO	3635214	23,00
REGULAR	3454621	258,64
<b>Total</b>	<b>312641584</b>	<b>13.971,12</b>

Ilustración 18 Distribución de ventas por familia productos para el 14 y 15 abril 2022 en Locatel. Creación Propia.

Para el análisis de los feriados mencionados en la descripción anterior, se estudia la familia de productos de la categoría Cuidado Personal para validar si los productos contemplados indican alguna actividad turística o social. Según la información recolectada,

se puede establecer que durante los días feriados se genera un consumo en categorías de Cuidado Personal como lo son bloqueadores solares, humectantes de piel, vitaminas y anticonceptivos. Por lo cual, se puede inferir que los días feriados son de gran beneficio económico para Locatel, al poder comercializar productos enfocados en el esparcimiento y diversión de la sociedad.

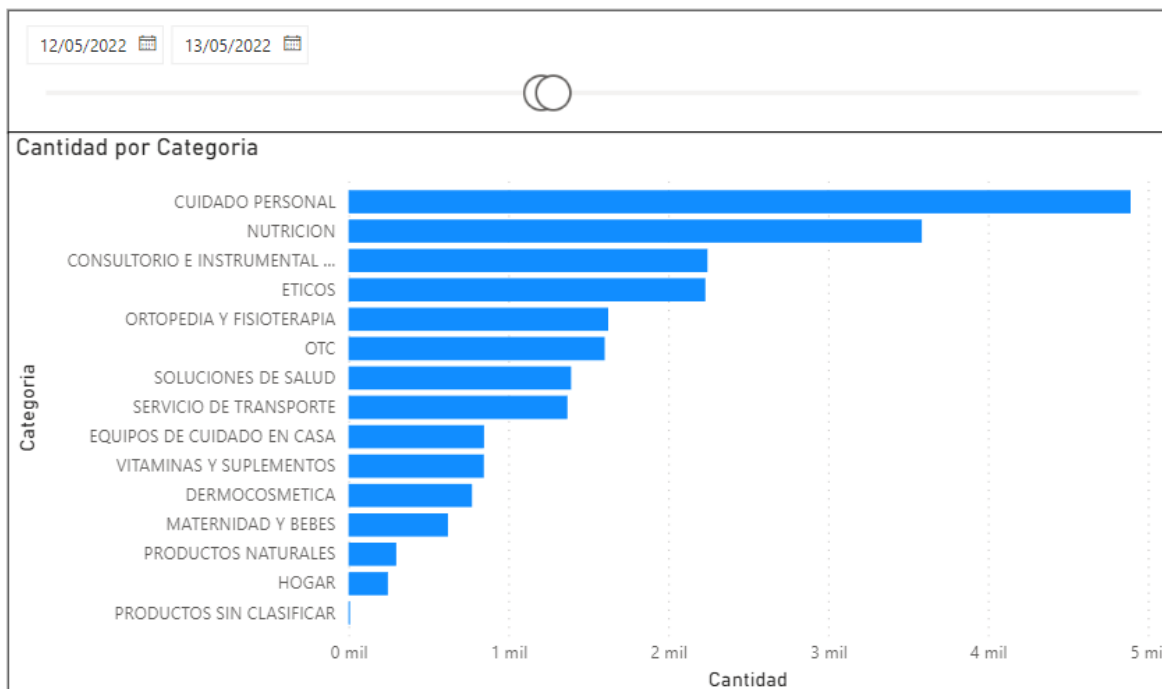


Ilustración 19 Distribución de ventas por cantidad y categoría del 12 y 13 mayo 2022 para Locatel. Creación Propia.

Por otra parte, se requiere comparar el efecto que tienen los festivos en categorías con mayor porcentaje de ventas con respecto a días que no son festivos y confirmar si en épocas ferias se realizan mayores ventas que en los días normales. Se analizará el consumo de los días 12 y 13 de mayo, dado que están en la segunda semana del mes y son fechas programadas para los días jueves y viernes (caso similar al análisis anterior de los festivos). En este periodo de tiempo, según la ilustración 19, las categorías de Cuidado Personal,

Nutrición, Consultorio e Instrumental y Éticos son las más relevantes para Locatel según la cantidad de productos vendidos.

Dado el contraste de los dos periodos de tiempo, se evidencia que las 4 categorías más comercializadas por la compañía concuerdan en los dos casos, pero para el periodo validado de mayo, se evidencia un alza de consumo en los productos de las 4 categorías con respecto al periodo de abril (por ejemplo, Cuidado Personal en el periodo de mayo vendió 2 mil unidades más que el periodo de abril). El contraste realizado de los dos meses refleja un comportamiento similar en consumo y categorías vendidas. Sin embargo, la hipótesis del posible aumento de ventas los días festivos se anula, debido que un periodo de días sin festivo, supera la venta del periodo festivo analizado en este capítulo.

Por lo anterior, validando la información de Locatel, el proceso de recolección es valioso al obtener datos relevantes de estudio; información con la que se puede perfilar el mercado y su cliente potencial. Con el objetivo de establecer la mejor información posible, durante el desarrollo del data mining se elimina la información inconsistente encontrada en las diferentes variables para garantizar que el resultado sea el más acertado posible. El detalle de la información explorado en esta fase, identifica categorías y familias de producto fundamentales para Locatel, junto con el género de mayor peso dentro de la base, información que servirá como contexto para realizar la segmentación y entender el enfoque que podría tomar Locatel para realizar su estrategia comercial.

## **7. Descripción de la Situación organizacional donde se realizará el proyecto**

### **(Contexto)**

#### **7.1. Sector Farmacéutico Colombia:**

La compañía seleccionada para realizar el proyecto empresarial es una empresa que opera en el sector farmacéutico y cuidado personal, convirtiéndose en uno de los más importantes retail farmacéutico del país, con presencia de tiendas físicas en Bogotá, Medellín, Cartagena y distribución a todo el país por medio de sus aliados logísticos. Según el DANE, para el cierre del cuarto trimestre del 2022, la industria manufacturera (en la cual se encuentra contenida la comercialización y fabricación de productos farmacéuticos) tuvo un crecimiento del 9.3% en el PIB nacional respecto al mismo periodo analizado del año 2021, en este rango de tiempo, el sector farmacéutico creció cerca de ocho puntos porcentuales respecto al año anterior (DANE, 2023).

El objetivo del proyecto con la compañía es realizar un entendimiento de sus clientes que le permita seguir desarrollándose comercialmente, para lo que se utilizará la teoría de segmentación desarrollada en Data Mining Techniques in CRM, donde se explora la capacidad de realizar segmentación según la naturaleza del CRM, por retención de clientes o desarrollo de los mismos. (Tsiptsis, 2009). Según el reporte Mapa del retail en Colombia 2022 de la firma EMIS, la compañía cerró el año 2021 con un ingreso operacional alrededor de los COP \$115 mil millones y le permite ocupar el lugar 127 del ranking, el cual domina Tiendas D1 y Alkosto K-tronix. En el sector de Medicamentos y Farmacias, la compañía se encuentra en la sexta ubicación, ranking que lidera la compañía Cruz Verde con más de COP \$3.3 billones en ingresos operacionales (EMIS, 2022).

Locatel para su proceso de crecimiento, desarrolla estrategias de marketing para fidelizar a sus clientes y aumentar las ventas en un mercado competitivo y muy equilibrado en la generación de valor. EMIS en su reporte Health and Beauty Specialists in Colombia generado en el 2023, informa que en el año 2022 el sector retail en Colombia tuvo un crecimiento en el valor de las ventas del 19%, con un cierre de más de COP \$15 trillones, en un mercado que se prevé estar en COP \$20 trillones, las categorías de cuidado facial y maquillaje son las de mayor relevancia en cuidado personal (EMIS, 2022). A través de las estrategias de marketing, la compañía ha podido desarrollar sus canales de distribución, lograr sus objetivos comerciales y ganar mayor presencia o “share” en el mercado farmacéutico.

El sector farmacéutico en Colombia está liderado por empresas que cuentan con presencia en regiones influyentes en el desarrollo del PIB como lo son la Sabana Cundi-Boyacense, la Costa Atlántica, el Pacífico Sur y los Santanderes, tales compañías son Supermercados y Droguerías Olimpia, Copidrogas o Cruz Verde, esta última apalancada por la compañía FEMSA, quien tiene una detallada capacidad en su operación logística (al distribuir bebidas y comercializar en todo el territorio nacional) que le ayuda a fortalecer sus negocios complementarios. Las compañías líderes como Olímpica por la naturaleza de negocio, apoya la operación de droguería con la de retail, dado que dentro del mismo establecimiento comercial se operan los dos canales de distribución. Por su parte, Copidrogas con una facturación en el 2021 de COP \$800 mil millones, tiene el desarrollo de su negocio en la cadena de pequeñas farmacias, de las cuales no es propietaria pero si tiene acuerdo de exclusividad, por lo que otra compañía no les puede vender ningún producto a los drogueros y puede monopolizar en cierta medida la operación y distribución

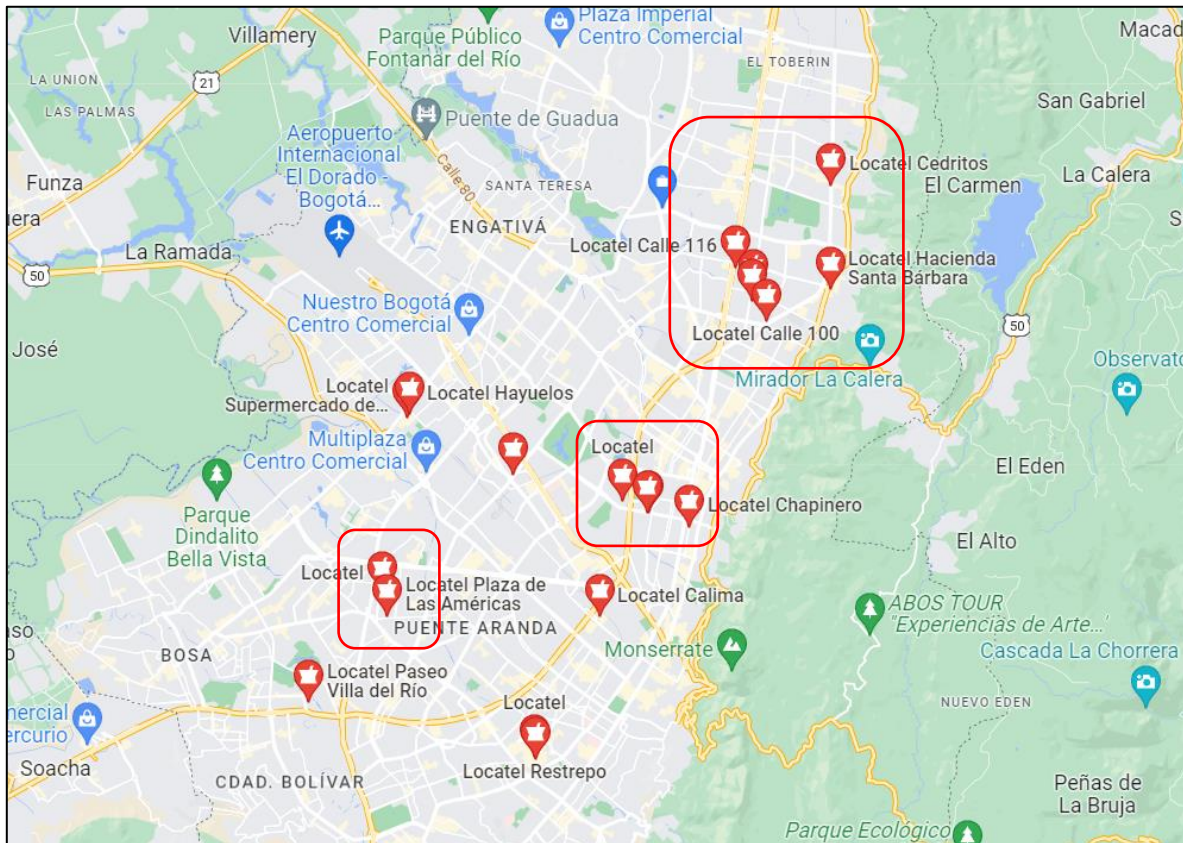
en varias regiones del país, además de no tener que invertir mayor cantidad de dinero en instalaciones comerciales dado que todo debe ser asumido por el propietario del local (EMIS, 2022).

## **7.2. Contexto Locatel:**

Según la naturaleza de negocio de Locatel, al ser un retail farmacéutico y no solamente una cadena de droguerías, su mayor competidor en Colombia es Farmatodo, dado que esta compañía también cuenta con establecimientos propios, con variedad de productos distribuidos a través del local comercial en góndolas y pasillos (formato similar a Locatel). Farmatodo ha logrado establecerse en ciudades principales como Bogotá y Cali, con el propósito estratégico de llegar a 50 tiendas en el país, ha empezado con apertura de tiendas en ciudades intermedias como lo son Bucaramanga y Cúcuta (estrategia que en gran medida está fortaleciendo Locatel). Con ingresos superiores a los COP \$650 mil millones el 2021, Farmatodo se encuentra en las 50 empresas más importantes de retail en Colombia, superando a Locatel e incluidas empresas del sector como Droguerías Colsubsidio, Comfandi y Droguerías Pasteur (EMIS, 2023).

Analizando la información exógena de la compañía, Locatel tiene distribuidas tiendas en Bogotá con cierta agrupación. Si se estudia el primer grupo ubicado en el norte de la ciudad, se logra reconocer que las tiendas están ubicadas equidistantemente entre calle la 100 y la 140, entre Autopista Norte y carrera 7; sector en el cual se encuentran viviendas desde estrato 4 al 6 y generalmente habitan personas con un gasto promedio por consumidor semanalmente mayor a COP \$470 mil agrupando todas las categorías de compra de los usuarios (Ramírez, 2019). Por su equidistancia, el traslado entre las tiendas es aproximadamente de 20 minutos, por lo que les permite reaccionar si requieren enviar un

producto determinado de una tienda a otra que no cuente con inventario para no perder ventas. Analizando capilarmente la tienda ubicada en Hacienda Santa Bárbara, centro comercial ubicado en estrato 6, el cual contiene empresas y tiendas de sectores diferentes al dermocosmético, le permite a Locatel generar ventas cruzadas por el consumo externo a su operación, propio del conglomerado comercial.



Anexo 1. Mapa tiendas Locatel en Bogotá. Google (s. f.)

Explorando grupalmente los demás sectores de la ciudad, se refleja que mantienen el principio de clusterización; no tienen distancias mayores a 6 km, con un promedio de 20 minutos por recorrido entre tiendas (Google. s. f). El sector suroccidental en el que se encuentra otro de los conjuntos de tiendas posee viviendas que oscilan entre el estrato 2 al 4, con 49% de su población categorizada como masculina y el 51% femenina, además el

14% de los hogares de Bogotá se encuentran ubicados en la localidad de Kennedy (Observatorio de Desarrollo Económico. 2019). La ubicación de las tiendas refleja la cercanía de centros comerciales, barrios residenciales y de fácil acceso por las vías principales o de alto flujo comercial como lo son Avenida NQS, Avenida Caracas y Avenida Boyacá (Google. s. f).

### 7.3. Análisis de participación en ventas por tienda para Locatel:

La ilustración que se presenta a continuación, contiene el resumen por día y tienda para evaluar el porcentaje de participación en ventas de cada uno de los establecimientos comerciales; entender cuál aporta mayor peso en las ventas y cual tiene oportunidad para que la compañía realice alguna estrategia comercial. En la gráfica se contemplan los valores facturados por establecimiento comercial y se tienen en cuenta las tiendas de Bogotá, la tienda de Cartagena, la bodega de la tienda online y la venta por la plataforma de marketplace Mercadolibre:

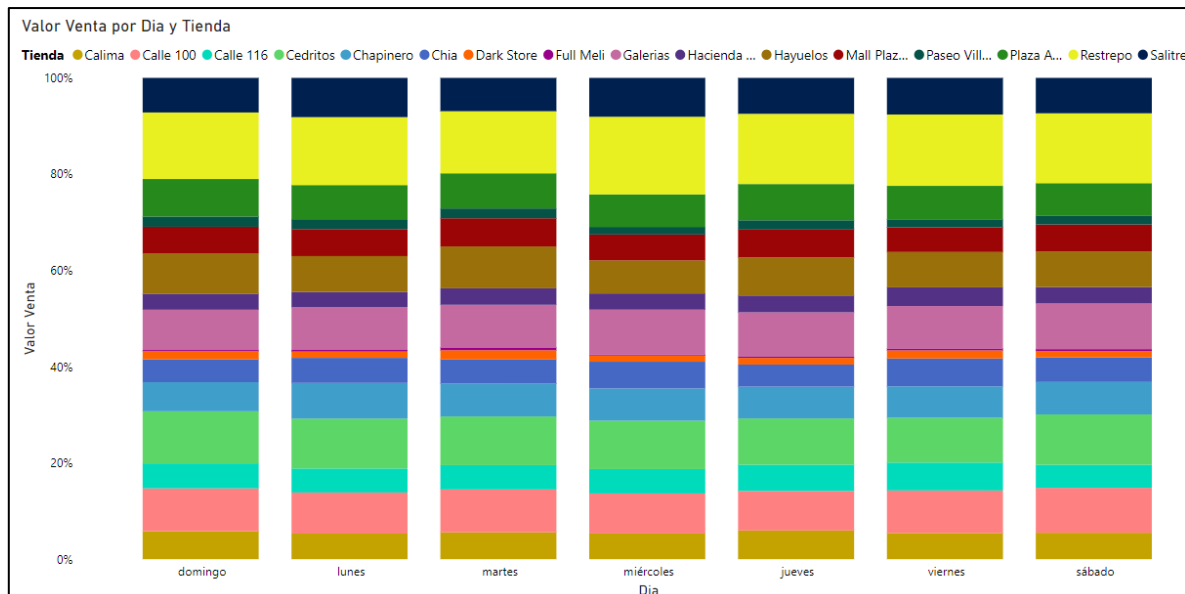


Ilustración 20 Distribución ventas por día y tienda Locatel. Creación Propia.

Con el resumen de la distribución anterior, se evidencia que la tienda de Restrepo es el establecimiento con mayor peso en las ventas de la compañía seguido por Cedritos, Calle 100, Galerías y Hayuelos. En cuanto a las categorías de mayor venta en la tienda de Restrepo, según la distribución de la variable Desc. Nivel 2, las subcategorías que mejor comercialización en valor de venta han tenido son las de Cuidado Persona, la de Dermocosméticos y la de productos Éticos (Entre los que están los medicamentos especializados para tratar enfermedades tales como fenómenos hormonales, depresión, triglicéridos, diabetes, etc.), dato que ayuda a entender la fortaleza que tiene Locatel para las categorías de Care y Dermo:

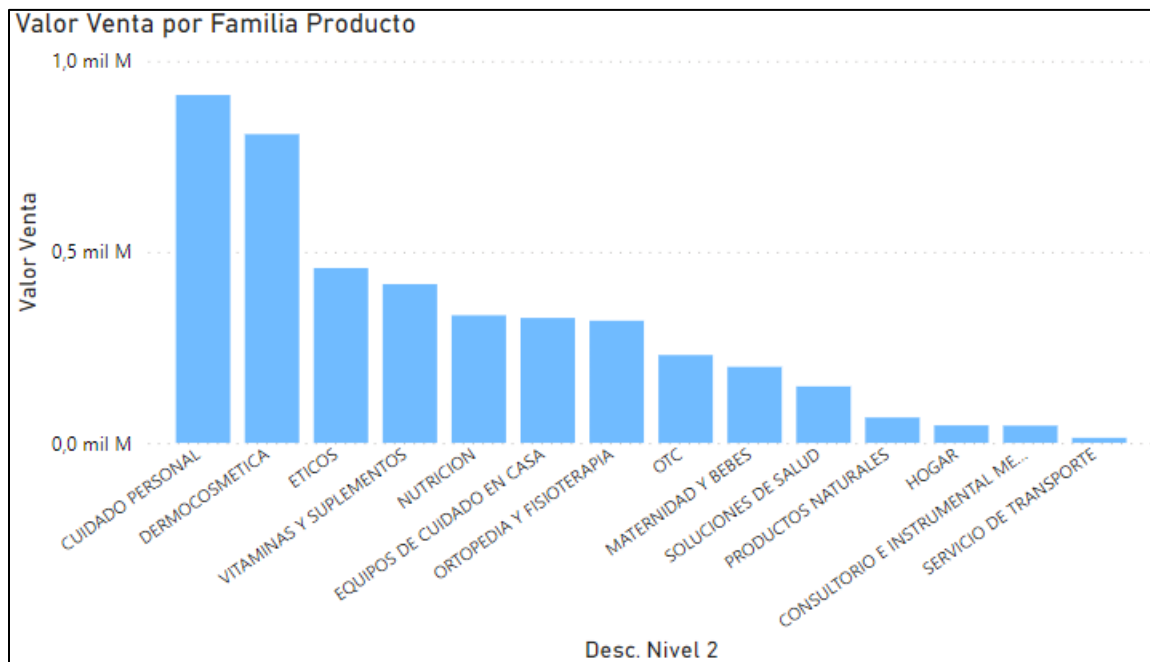


Ilustración 21 Distribución ventas por Desc. Nivel 2 para tienda Restrepo Locatel. Creación Propia.

### 7.3.1. Análisis Tienda Chapinero - Bogotá:

La ubicación geográfica se refleja como una estrategia operativa cuando las empresas logran entender el mercado que desean atender, el perfil de consumidor que tienen y el valor

que puedan generar dentro del sector. Para el análisis geográfico, la zona oriental de la ciudad alberga viviendas y establecimientos en estratos del 4 al 6, en el entorno se encuentran universidades privadas con altos valores de matrícula como la Universidad Javeriana y el CESA, lugares de asistencia masiva como lo es el Parque Nacional, instituciones médicas tales como el Hospital Militar, la Clínica Marly, además de empresas de consumo y servicios como lo es Dollarcity, Hotel Boyacá Real y el Restáurate-Bar El Mono Bandido. (Google. s. f). El nivel de interacción social de la zona tiende a desarrollar categorías con un ticket de compra medio-alto en cuidado personal, vitaminas, artículos para evitar el estado de embarazo, energizantes y alimentos en general, además de la categoría de cuidado personal la cual es la categoría más relevante en el análisis desarrollado.

Dado el contexto geográfico en el que se encuentra la tienda Chapinero, se presenta el resumen de ventas por familias de producto en dicho establecimiento:

Tienda		
LOCATEL CHAPINERO		
Familia Producto	Valor Facturado	Cantidad
PROTECTORES SOLARES	84328503	1.310,00
HIDRATANTES	65456092	960,00
ANTI-ARRUGAS	53241041	935,00
HUMECTANTES	52490782	2.102,00
A11X-OTRAS VITAMINAS	51767227	968,00
AUTOMAT.Y SEMIAUTOMAT.BRAZO	41288236	214,00
COMPRESION MEDIANA	38377517	754,00
A11A-POLIVITAMIN.CON MINERAL.	30871821	637,00
ESPECIALIZADO	26730670	1.022,00
TIRAS	26409706	293,00
V-VARIOS	23359229	529,00
N02B-ANALG NO NARC.ANTIPIRET.	22755760	1.069,30
<b>Total</b>	<b>2014165371</b>	<b>84.414,62</b>

### Ilustración 22 Ventas por familia de producto de tienda Locatel Chapinero. Creación Propia.

Haciendo una validación específicamente de la tienda ubicada en Chapinero, se logra determinar que las familias de producto con mayor facturación son las de protectores solares, hidratantes y anti-arrugas, detalle contrario a la hipótesis anteriormente planteado. El efecto que tiene la ubicación para entender el consumo se da debido que en el sector, además de establecimiento laborales y educativos, se encuentran varias residencias con un perfil determinado; personas mayores a 50 años y con un alto poder adquisitivo que les permite adquirir los productos de dichas familias de mayor valor facturado (Observatorio de Desarrollo Económico. 2019). Locatel ha logrado ubicar sus establecimientos comerciales de forma estratégica según el objetivo de su operación; mantener un ticket de compra promedio según el poder adquisitivo de los consumidores, cautivar el mercado farmacéutico de características específicas como lo es la categoría de cosmética activa y en lugares con un alto flujo de comercio, lo que le permite beneficiarse de la fuerza centrípeta del mercado general desarrollado por el sector económico.

#### **7.3.2. Comparativo Tienda Calle 116 vs Tienda Plaza Américas:**

Se requiere analizar el fenómeno que representa la ubicación de las tiendas en diferentes sectores de la ciudad, con el objetivo de reconocer similitudes dentro del proceso operativo y el diferencial que representa la ubicación geográfica en el desarrollo comercial. Un objetivo operacional que tienen las compañías a la hora de apertura de espacios físicos de venta o distribución de productos es aumentar su utilidad y share de mercado, porque tener tiendas físicas le permite al consumidor no solo adquirir de forma inmediata un producto sino probarlo, sentirlo, reconocerlo y determinar que su necesidad está siendo atendida por los componentes y características de los artículos. Al tener distribuidos los establecimientos

comerciales, Locatel logra tener recordación dentro del mercado y es fácilmente atribuible su core de negocio con la necesidad que presentan los consumidores.

A continuación se presenta el comparativo entre dos tiendas del sector norte (Calle 116) y sector sur (Plaza Américas) para poder analizar el consumo de cada una:

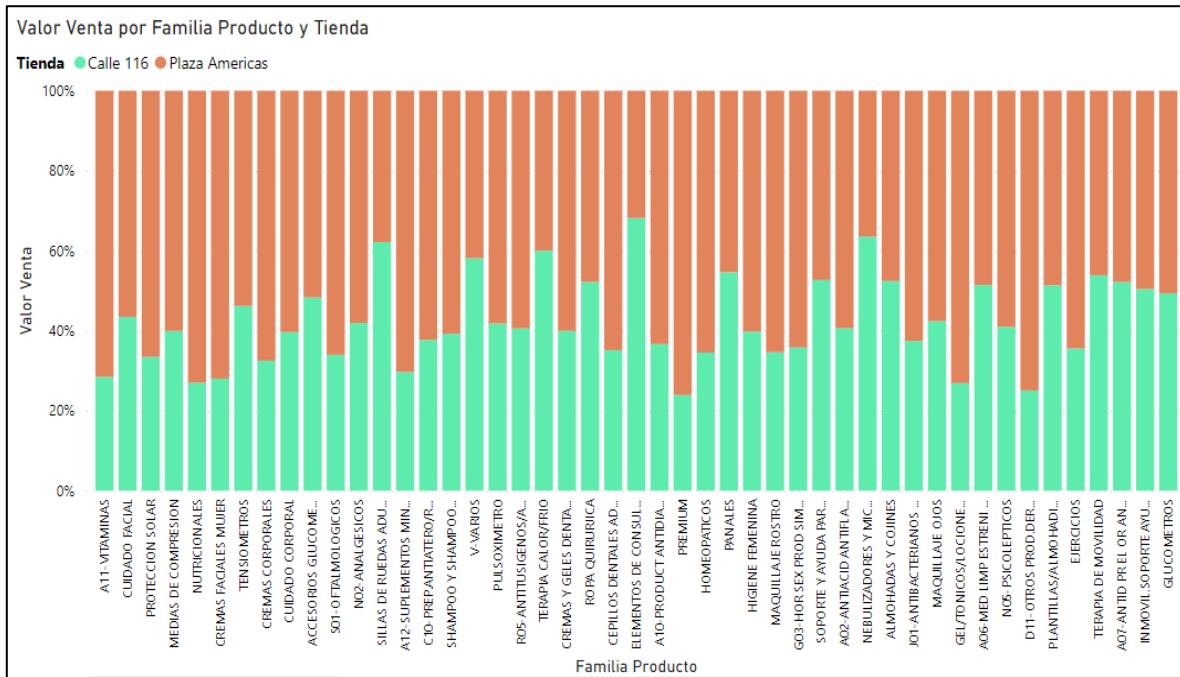


Ilustración 23 Comparativo de ventas por familias de producto tienda Calle 116 y Plaza Américas. Creación Propia

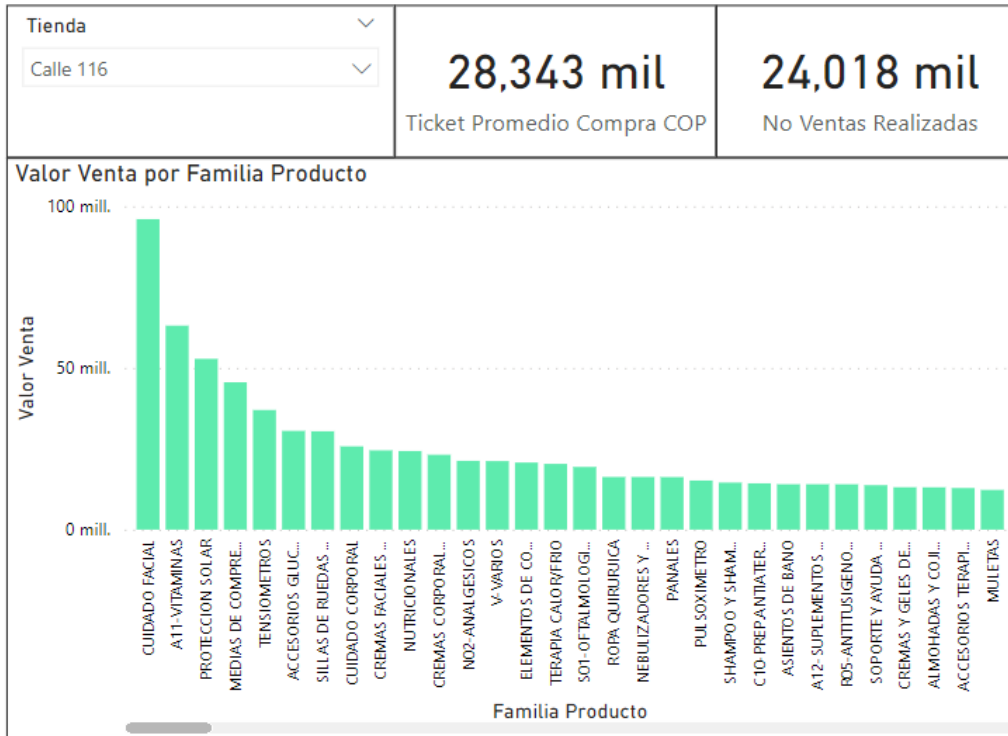


Ilustración 24 Ventas de tienda Locatel Calle 116. Creación Propia.

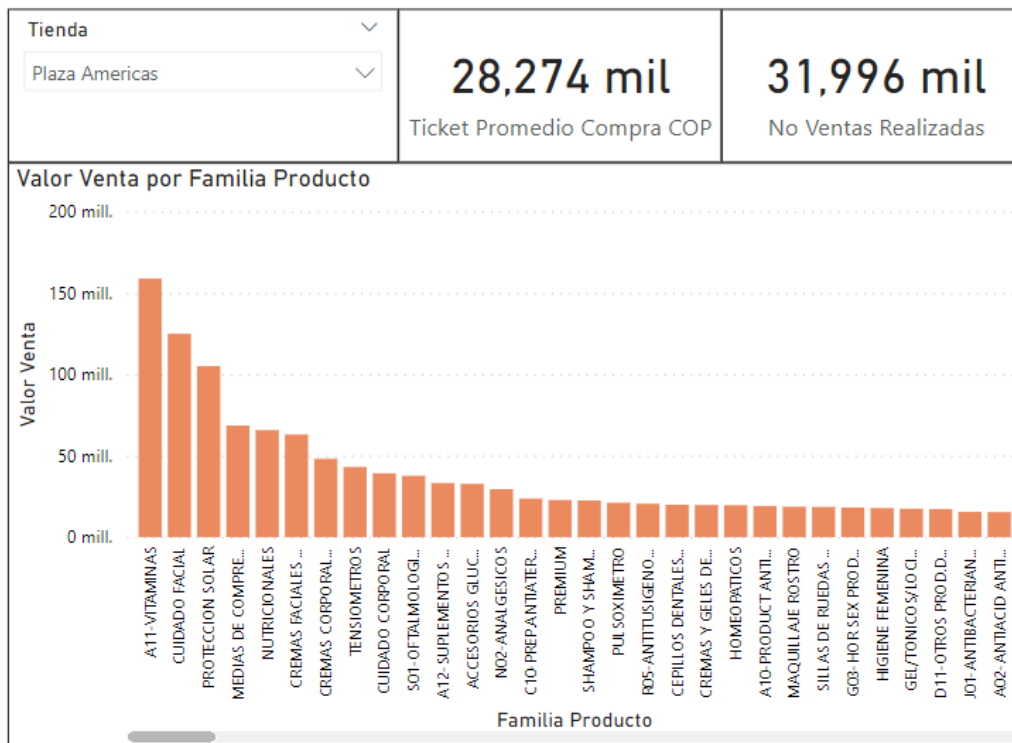


Ilustración 25 Ventas de tienda Locatel Plaza Américas. Creación Propia.

Según el contraste realizado, en cuanto al valor de compra promedio del periodo estudiado, a pesar de la diferencia de estratos que existen en cada una de las zonas, las dos tiendas tienen un ticket de compra similar (COP \$28.343 en Calle 116 y COP \$28.274 Plaza Américas), con una variación del 0.24% lo que refleja un equilibrio en el promedio de compra. Para la tienda de la Calle 116, se realizaron menos ventas con un total de 24.000 facturas, con respecto a Plaza Américas en donde se facturaron alrededor de 32.000, en este aspecto es notable una diferencia en el total y se puede deducir que la ubicación de la tienda del suroriente (la cual se encuentra cerca a un centro comercial con alto flujo de personas) tiene mejor desarrollo con respecto a la del norte (que a pesar de estar sobre una vía principal el flujo de personas presentado no se da por el comercio sino por las actividades económicas de cada cliente). Sin embargo, existe notoria diferencia en las familias de producto vendidas, dado que la más importante en la Calle 116 es Cuidado Facial llegando a los COP \$100 millones vendidos, con respecto a los COP \$150 millones vendidos por la familia de Vitaminas en Plaza de las Américas (siendo Cuidado Facial la segunda en la tienda con COP \$125 millones).

Lo anterior se puede explicar por la distribución demográfica de las zonas; en Usaquén (Calle 116) siendo estrato 5 o 6, el 60% de la población se encuentra laborando y habitando allí tienen la capacidad económica para adquirir productos con un alto valor económico con el objetivo de priorizar la imagen personal para sus empleos (Ramírez, 2019). En la localidad de Kennedy (Plaza de las Américas) el estrato 4 es el más alto que contiene la localidad, se encuentra el doble de hogares con respecto a la otra localidad, lo que permite intuir el interés de las cabezas de hogar por el cuidado y prevención de enfermedades de sus hijos o adultos mayores con la compra de suplementos vitamínicos, familia de producto con el mejor

porcentaje de ventas en la tienda. Este primer ejercicio de contraste por tienda, permite intuir que en algunos casos la ubicación geográfica si podría ser una estrategia de segmentación y de estrategia comercial, dado que en algunos sectores se podría priorizar la distribución de productos más especializados que tendrán mayor probabilidad de venta que en otros sectores.

## **8. Descripción de la situación estudio de caso y/o problemática empresarial y método y/o estrategia a aplicar para su solución**

La necesidad expuesta por la compañía es el fortalecimiento de su estrategia de CRM, de acuerdo con esta oportunidad, se requiere identificar patrones de compra para diseñar una segmentación inicial, agrupando en el modelo predictivo a los clientes por la patología a analizar. En el ejercicio predictivo se incluye la priorización de los productos que tiene mayor venta, con respecto al tipo de enfermedades que puede tratar, junto con la edad y genero del cliente, dado que, el área de marketing busca fidelizar sus clientes con el envío de información sobre el tratamiento de las enfermedades que los consumidores padecen; como los cuidados a tener, productos complementarios a consumir y las promociones comerciales que dichos medicamentos puedan tener en un momento determinado. El desarrollo del proyecto tomará como base de datos las compras realizadas en el trimestre de abril, mayo y junio del 2022 dado que la compañía informa que el comportamiento de los demás trimestres es similar al analizado y se puede tomar como una muestra de la base de datos del año completo.

Dentro del desarrollo del proyecto, se reconocen prácticas generadas por la compañía que le ha permitido afrontar el desarrollo inicial del análisis de datos, como lo son fidelización de los consumidores. La compañía tiene una base de datos estructurada, a la cual constantemente monitorean para actualizar o modificar; sea el caso requerido, lo anterior debido a la constante utilización de los datos de sus clientes para optimizar la gestión del CRM entre actividades comerciales por su plataforma online como en puntos de venta. La estrategia está a cargo de la directora de mercadeo de Locatel, quien generó el

contexto inicial de la estrategia, la forma de recolectar la información y el procesamiento dentro de la estrategia, en la reunión se propone realizar una segmentación de los clientes para que se reconozca el grupo poblacional al que se debe enfocar la estrategia definida de enviar información relevante de las enfermedades y descuentos en los medicamentos que la tratan. Si bien, la necesidad de generar una segmentación que apoye la estrategia de marketing es el principal objetivo del proyecto, en la investigación y análisis de la información se ha podido evidenciar prácticas con oportunidad de mejora en los procesos, las cuales se otorgaran en el entregable final del proyecto.

Para desarrollar un proyecto de data mining y organizar el objetivo del mismo, se utilizará la metodología CRISP (marco de trabajo y entregables directamente de la minería datos) anteriormente mencionado. El objetivo del trabajo requiere una capacitación académica previa sobre la segmentación teórica en el mercado, ejemplos reales sobre la implementación de segmentación en estrategias CRM y la modelación de la información para poder replicar un código, para lo cual se utilizará la literatura del libro *Data Mining Techniques in CRM: Inside Customer Segmentación*. La información teórica se ubicará en documentos académicos, revistas e informes investigativos y desarrollo de códigos sobre modelos no supervisados con enfoque en retail, junto con informes y estudios económicos.

## **9. Descripción de las alternativas, estrategias y/o acciones que se toman en el análisis de la solución a la problemática**

### **9.1. Descripción Alternativas de Proyecto:**

Una de las propuestas de solución para la segmentación es realizarla según el componente activo, dado que varios medicamentos comparten esta descripción según la enfermedad a controlar, por lo que sería de gran utilidad realizar la agrupación desde este concepto. Sin embargo, analizando la opción con la empresa, no se cuenta con esta base de datos. Buscando con fuentes externas como lo es INVIMA, se debe incurrir en un pago para descargar la base de datos que la entidad tiene sobre los medicamentos, aunque no se tiene la garantía que todos los productos de la base puedan estar contenidos en ella, dado que hay algunos productos genéricos y funcionan para varias dolencias o patologías.

Otra opción para realizar el proyecto sería estudiar una muestra más pequeña de la información en la que se reconozca una actividad comercial determinada para una patología específica y poderla comparar con una muestra similar en la que no se realizó la actividad promocional. El objetivo de estudiar estas dos muestras ayudará a determinar si bajo una promoción por patología se evidencia el crecimiento por venta cruzada de productos complementarios y entender si el fenómeno es viable en la operación comercial de la compañía. Sin embargo, la muestra estaría muy segmentada y el modelo no se podría entrenar de buena forma, por lo que el resultado no entregará información nueva o incluso puede realizar una segmentación sin sustento teórico de la información.

Al realizar el contexto geográfico, social y económico en el que se encuentra Locatel, sus establecimientos comerciales y el estado de la competencia, otra opción que se puede explorar para realizar una segmentación es el contexto social en el que se encuentran las

tiendas. La compañía tiene la facilidad de tener tiendas distribuidas por la ciudad, lo que conlleva a que cada una atiende a un contexto social propio, por lo que el estrato, la localidad, el gasto promedio de cada zona, la población y edad promedio de los usuarios puede ser una alternativa para segmentar clientes en su estrategia comercial. La posibilidad de segmentar a través del contexto social requiere un análisis capilar, detallar con un estudio de mercado el entendimiento de la competencia, las categorías de compra, las posibles estrategias de mercado y debería ser liderado por los líderes de operación, quienes conocen la actividad de cada uno de los establecimientos comerciales.

Según lo anterior, la mejor opción para la segmentación será realizarla según la base de datos de patologías creada en propiedad para el proyecto; acorde al historial de compras de la compañía, el camino para llegar a este objetivo es asociar los medicamentos a grupos de enfermedades para identificar cual producto serviría para una campaña comercial. Por la coyuntura del proyecto, se decantó por la asignación de una patología por medicamento, dejando la más importante o referente a la cual es formulada, dado que varios de los productos tienen la capacidad de ser recetados en más de una patología. Dado el nivel de complejidad para poder identificar la enfermedad específica, se realizó una investigación externa por los medicamentos de la base de datos, información que se compartió con el equipo de la compañía para su validación y aprobación.

El realizar la agrupación de medicamentos por enfermedad es una actividad que tiene algunas restricciones con respecto a la fiabilidad del recurso, como por ejemplo el acceso a la información; los productos de dermocosmética, elementos terapéuticos o productos de la categoría consumo masivo no cuentan con una patología a tratar, lo que resulta en crear categorías genéricas para dichos elementos. Por otra parte, se consultará con fuentes oficiales

como Ministerio de Salud, Cámara de Comercio, INVIMA, para determinar la viabilidad de asociar los EAN de los medicamentos con descripción del producto a la hora de solicitar el registro para comercializar los productos.

Según lo anterior, la posibilidad de determinar y encontrar una forma viable para asociar los medicamentos a las patologías tiene varias barreras para la consecución de la información, por ello, se ofrece una solución alternativa para la segmentación de clientes y es utilizar la información de las ventas del periodo estudiado para que la estrategia de CRM se pueda alimentar del data mining. La alternativa es realizar una segmentación por clientes según la edad del comprador, tienda, cantidades y valor de compra. Aunque la alternativa que se oferta es viable, no es el objetivo final del proyecto, sin embargo, la compañía comprende la dificultad para conseguir la información y esta alternativa pueda utilizarse en promociones específicas.

Las alternativas de acción para el proyecto están anidadas a la información inicial que se definió con la compañía, por lo que no involucraría mayor utilización de recursos o gastos. La ventaja inicial es brindar opciones a la compañía para optimizar la estrategia de segmentación para su CRM a través del data mining y fortalecer el proceso de fidelización y desarrollo comercial. La desventaja es no tener la opción que desea el área de marketing para segmentar por patologías las actividades promocionales que desean realizar y el proyecto debe entregar una solución alternativa a la que inicialmente se esperaba.

## **9.2. Desarrollo Segmentación:**

Dado el contexto del negocio, es necesario estudiar y comparar las diferentes variables que son influyentes en el entendimiento de la base de datos para analizar su comportamiento y la viabilidad en el desarrollo de la segmentación. Por lo tanto, a través

del capítulo se analizaran varios escenarios comparativos con el desarrollo de la información pertinente. Para iniciar el análisis, se normalizan las variables de la base y se estudia la correlación entre las mismas, como se presenta a continuación en una muestra de matriz con dimensión de 20 por 20:

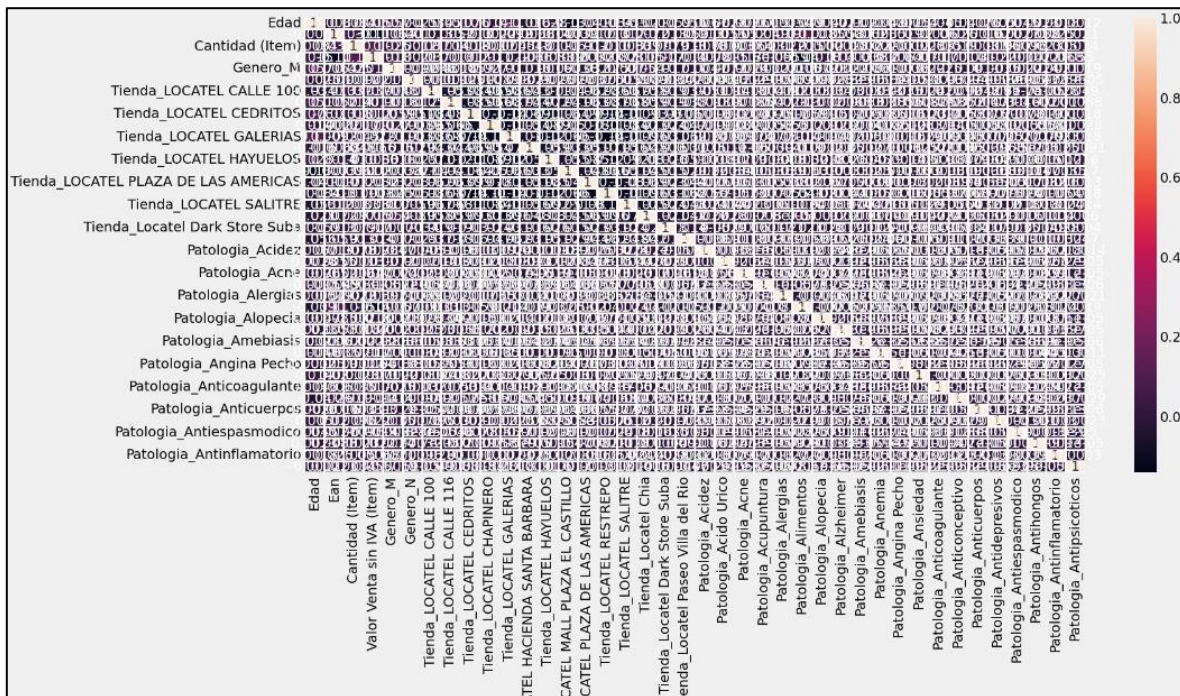


Ilustración 26 Correlación variables normalizadas Base Datos. Creación Propia.

En el análisis de la totalidad de la matriz de correlación para las variables normalizadas no se logra determinar una correlación llamativa, dado que todas están cercanas a 0, lo que significa que no existe una relación proporcional o inversa entre las variables. En este punto del análisis ninguna de las variables tiene fuerte correlación, lo que un aumento o disminución de una variable no afectará la otra y no se podría utilizar para la segmentación dado que el resultado no tendría valor analítico para el proyecto. Con este resultado en el momento no es viable iniciar con alguna de las variables dado que no generará un resultado congruente para la segmentación.

A continuación se muestra el resumen de correlación existente entre las variables numéricas de la base, como en la anterior ilustración, se puede identificar que ninguna de las variables tiene una correlación positiva o negativa con las demás. Siendo muy cercana a 0 (en donde no existe relación alguna) las variables de mejor comportamiento correlacionado son cantidad y valor de venta facturado: (a mayor cantidad de productos vendidos, aumenta el valor facturado):

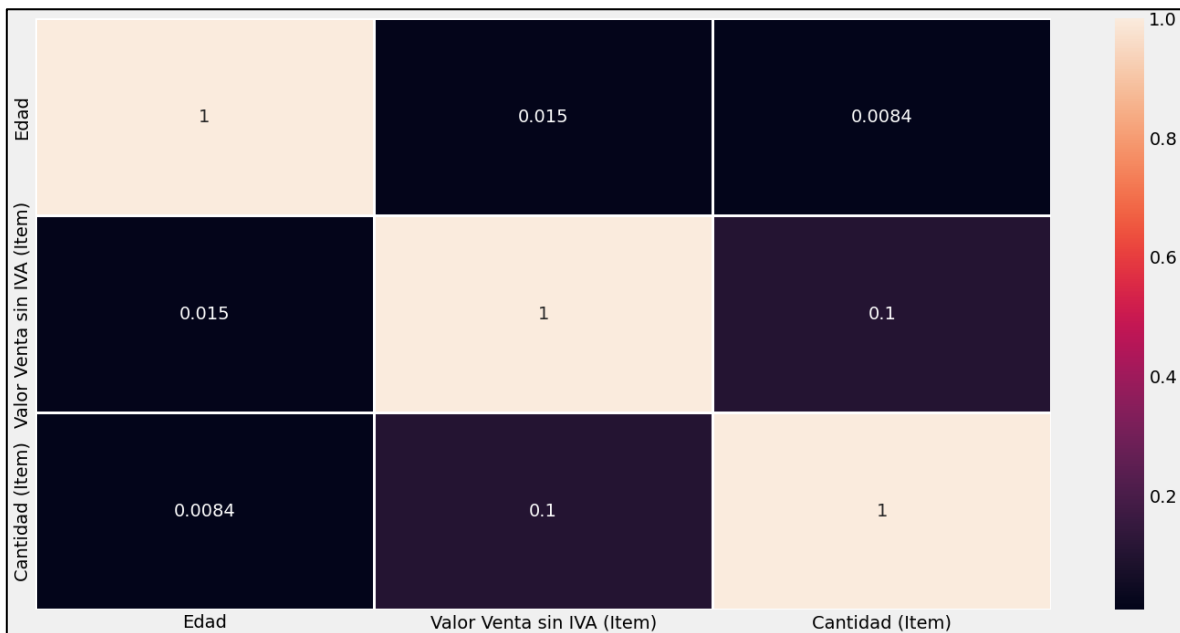


Ilustración 27 Correlación variables numéricas Base Datos. Creación Propia.

Para el esquema de la segmentación se utilizó el modelo de procesamiento k-means, dado que el modelo determinará el número de clúster a crear por las variables analizadas e interacciones realizadas entre ellas. En el modelo de segmentación, se busca la interacción de algunas variables para encontrar el mejor conjunto que tenga un comportamiento idóneo para la segmentación, por lo que se validarán algunas variables numéricas de la base de datos. Según lo anterior, se procederá a comparar un conjunto de variables para entender el efecto de su unificación y resultados.

### 9.2.1. Interacción entre variables de Cantidad Ítems y Edad:

En el primer ejercicio se analizó la variable Cantidad de Ítems con la Edad, arrojando como sugerencia la creación de 4 clúster, por lo que genera el resultado de la gráfica que se expone a continuación:

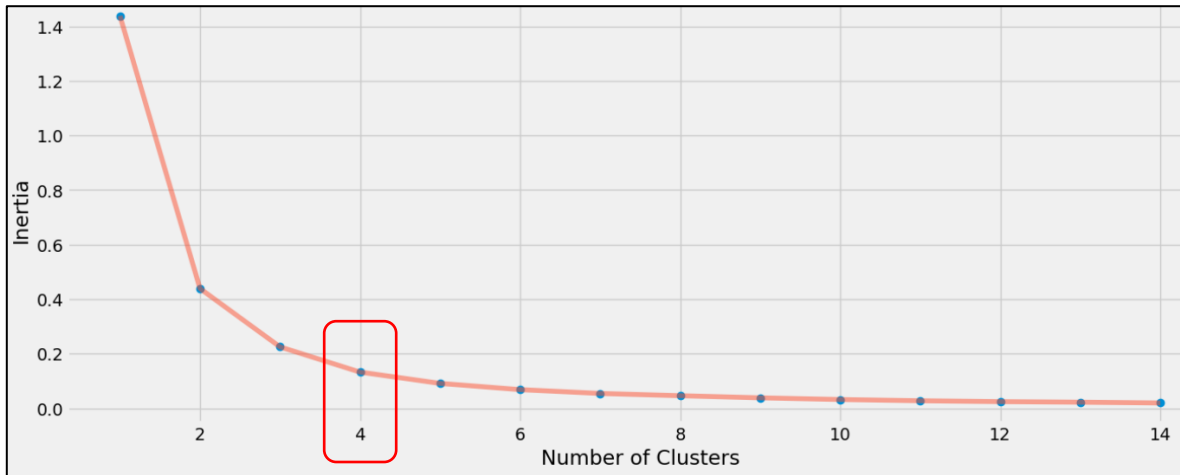


Ilustración 28 Agrupación de variables Cantidad - Edad. Creación Propia.

Al modelo se le asigna la distribución de los registros por grupo según las dos variables evaluadas anteriormente. El resultado de la distribución muestra en los 3 primeros grupos que las cantidades por edad tienen un comportamiento muy similar; dado que refleja un promedio de compra de 15 unidades vendidas por transacción. La distribución de cantidades para el último grupo cumple con el fenómeno expuesto en los tres primeros, aunque la diferencia de este grupo es el rango de edades en el que oscila su rango, ya que en los primeros grupos el rango de edades está alrededor de los 16 años, mientras en el cuarto grupo su rango está entre los 35 años. El fenómeno del cuarto grupo se explica dado que la mayor cantidad de usuarios de la base de datos de la empresa está distribuida entre los 18 y 65 años (Según la ilustración 8 el 75% de las edades están entre 18 y 60 años de

edad), por lo que para el ultimo clúster el modelo debe agrupar mayor cantidad de edades para compensar el déficit de usuarios en el rango de la variable:

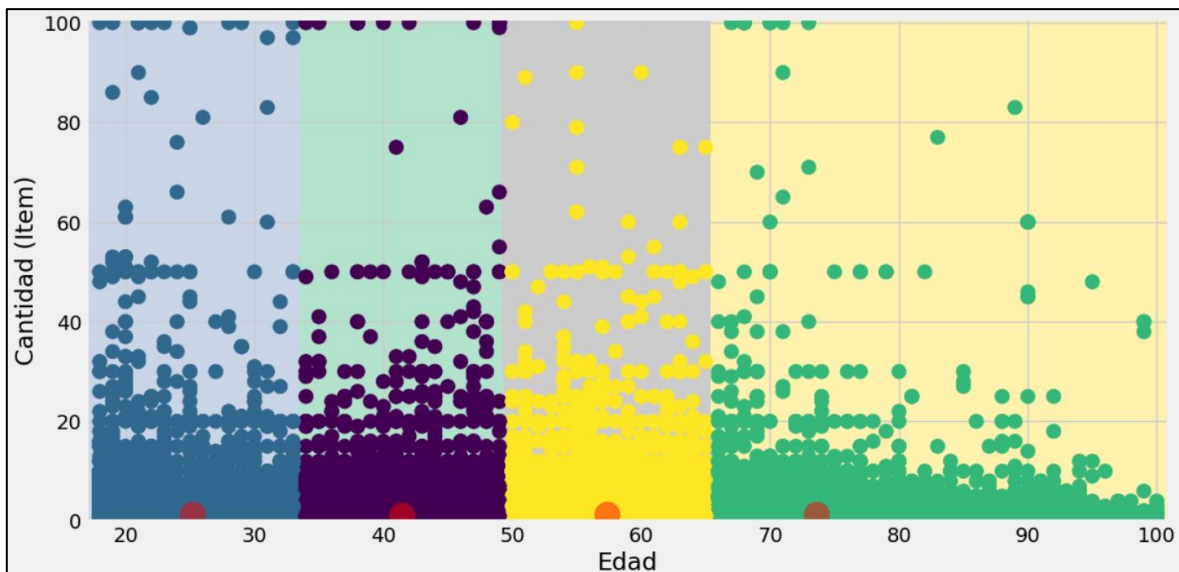


Ilustración 29 Distribución de la agrupación de variables Cantidad - Edad. Creación Propia.

La comparación entre edad y cantidad de productos vendidos refleja las unidades adquiridas por factura y edad, en la cual se muestra que el 80% de las compras están entre 1 y 20 artículos por compra entre 20 y 80 años. Sin embargo, la gráfica no permite determinar un patrón lógico a utilizar, dado que una de las compras de más de 60 artículos pudo haber sido un evento específico o tomarse como un outlier al verse tan alejado de la media de compra. Por lo anterior, este grupo de variables es una opción aceptable para iniciar una segmentación, aunque se aconseja realizar la comparación con otro grupo de variables para analizar si su interacción genera un resultado con mayor entendimiento.

### 9.2.2. Interacción entre variables de Cantidad Ítems y Valor Facturado:

Como se evidencia en la Ilustración 27, la correlación entre Cantidad Ítems y Valor Venta es el mejor comportamiento obtenido por la correlación de las variables, por lo que

se analizarán en conjunto para un nuevo ejercicio de segmentación. Una vez se agregan las dos variables al modelo, el resultado sugiere la creación de 6 clúster:

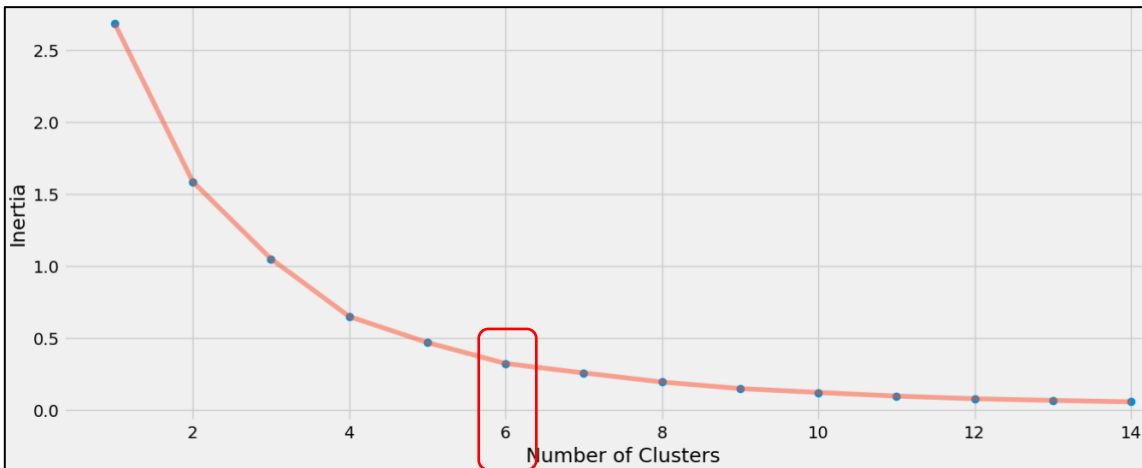


Ilustración 30 Agrupación de variables Cantidad – Valor Factura. Creación Propia.

Una vez se determinan los grupos a generar, se evidencia la distribución de cantidades por valor total de factura, imagen que se comparte a continuación:

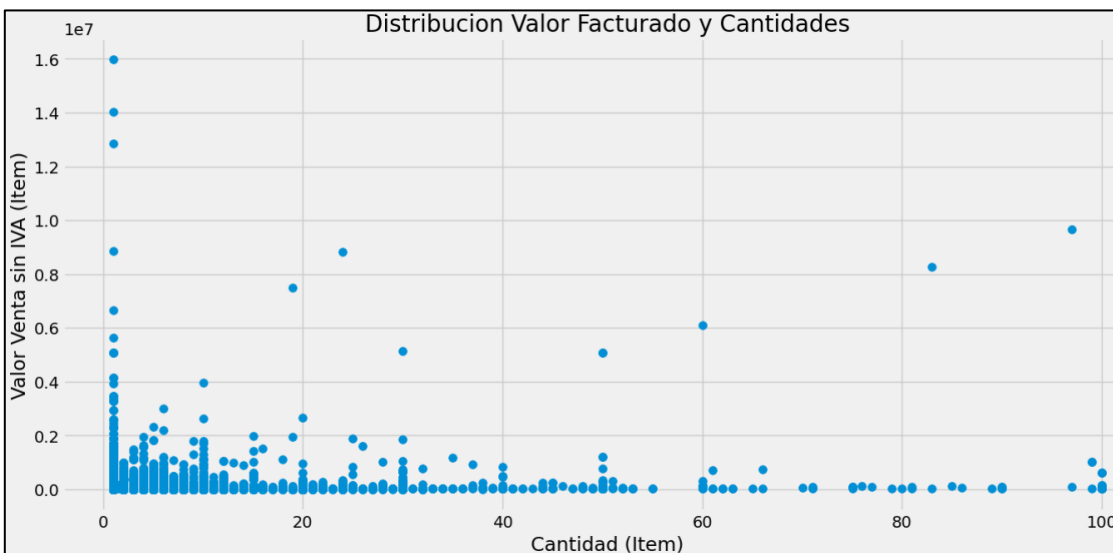


Ilustración 31 Distribución de agrupación de variables Cantidad – Facturación en millones. Creación Propia.

Con la última comparación realizada, se evidencia que la mayoría de compras realizadas están cerca de las 20 unidades, más allá de dicho rango las cantidades solicitadas

son muy poco frecuentes que se adquieran. Por su parte, la facturación no supera los COP \$400 mil, aunque existen algunas compras con valores superior a este límite, lo cual se puede intuir dado que las compras realizadas son productos especializados con alto valor económico, explicando así la variación con respecto a la media. A pesar de tener un comportamiento estandarizado, los valores altos no permiten generar una segmentación clara que contenga los comportamientos de la unión de variables; el resultado si genera una segmentación pero no logra ser detallada ya que más allá de las cantidades y valor asociados no hay un actor determinante en la distribución de los grupos, lo cual sugiere el análisis de un nuevo grupo para determinar mejores resultado de segmentación.

### 9.2.3. Interacción entre variables de Tiendas y Edad:

Para el tercer ejercicio se requiere juntar las variables de Tienda y Edad, Para el modelo se requiere analizar capilarmente una tienda, por lo que se escoge la de Restrepo, dado que según la ilustración 20 es la tienda con mayor porcentaje de ventas realizadas en el periodo estudiado. La unión entre las dos variables mencionadas establece la opción de crear 5 clúster, dado el resultado de la siguiente distribución:

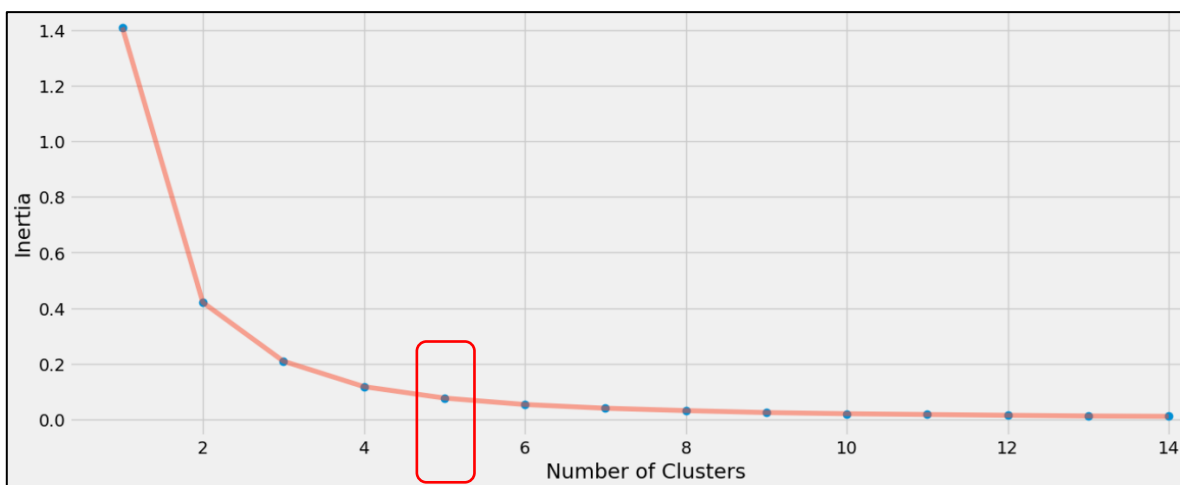


Ilustración 32 Agrupación de variables Tienda – Edad. Creación Propia.

Una vez se determina la cantidad de clúster, se procede a correr el modelo para determinar cómo se distribuyen los datos en cada uno de los grupos. A continuación se presenta el resumen de la iteración de las variables con la cantidad de clústeres, en el cual se identifican los 5 segmentos generados y la cantidad de datos.

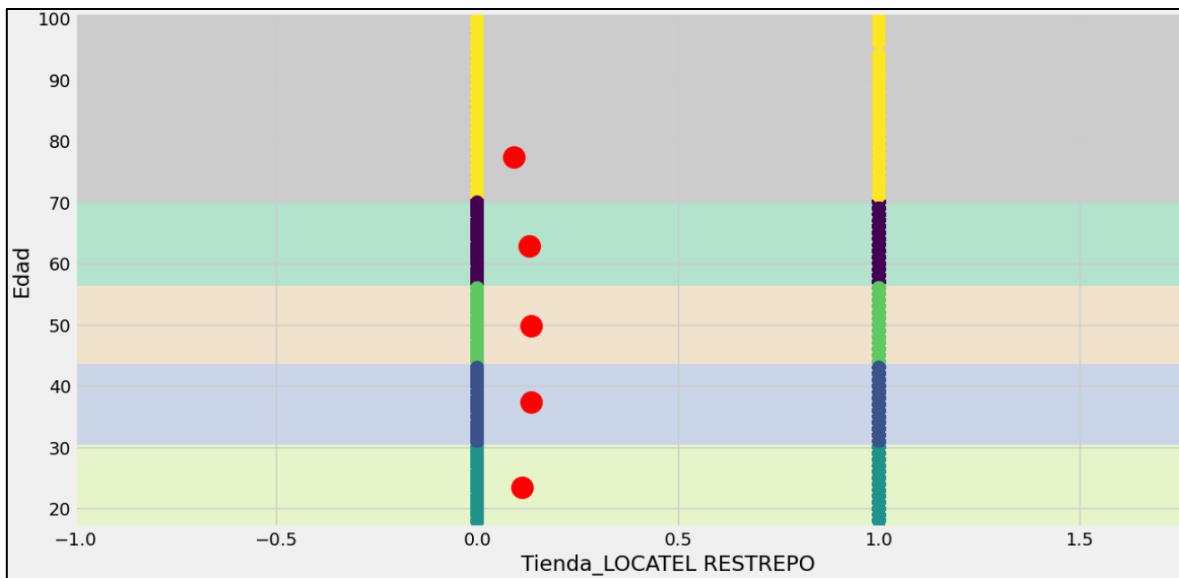


Ilustración 33 Distribución de agrupación entre variables de Tienda – Edad. Creación Propia.

Con el resumen de la interacción de variables, se puede evidenciar que el modelo fragmenta los grupos en rangos de edades muy similares para la Tienda Restrepo. Aunque el modelo encuentra que las edades mayores a 70 años tienen características similares que no permiten fragmentar en mayor número de clústeres la muestra estudiada. El modelo sugiere la opción de contrastar la variable normalizada con otra numérica y tener un resumen de posible segmentación, dado que el resultado no contempla un patrón de compra para la tienda según las edades agrupadas que permite entender una segmentación de la información y determinar que categorías, edades o productos podrían ser más rentables para la compañía.

#### 9.2.4. Interacción entre variables de Cantidad y Tienda:

Continuando con el estudio de la Tienda Restrepo, la variable a unir en el modelo es la Cantidad Items, el resumen de la iteración se comparte a continuación:

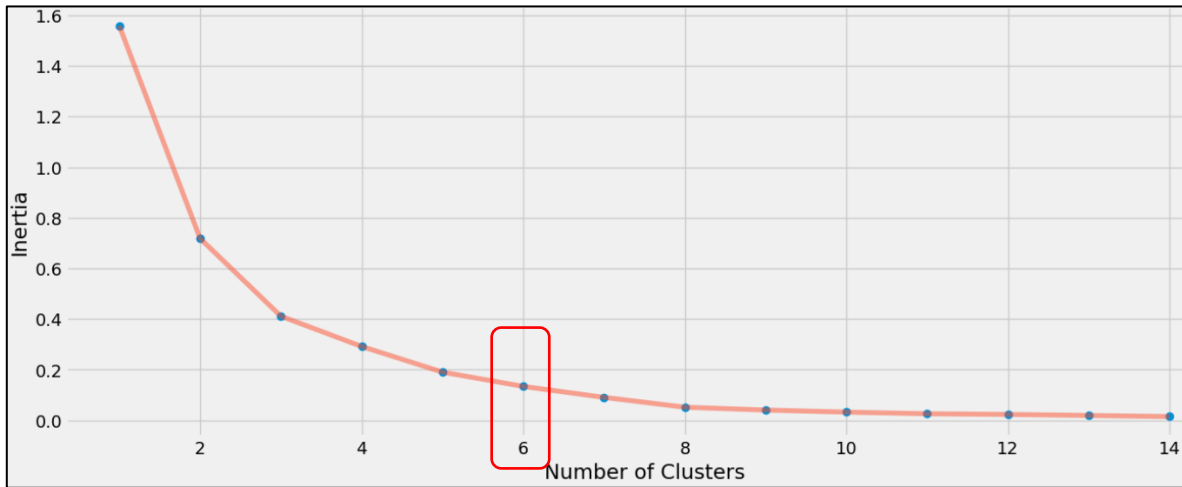


Ilustración 34 Agrupación de variables Tienda – Cantidad. Creación Propia.

Según la gráfica obtenida, el número de clústeres a generar serán 6. Con el número de segmentos determinados, se procesa nuevamente el modelo con las nuevas variables y se genera la siguiente distribución de registros para cada uno de los grupos generados.

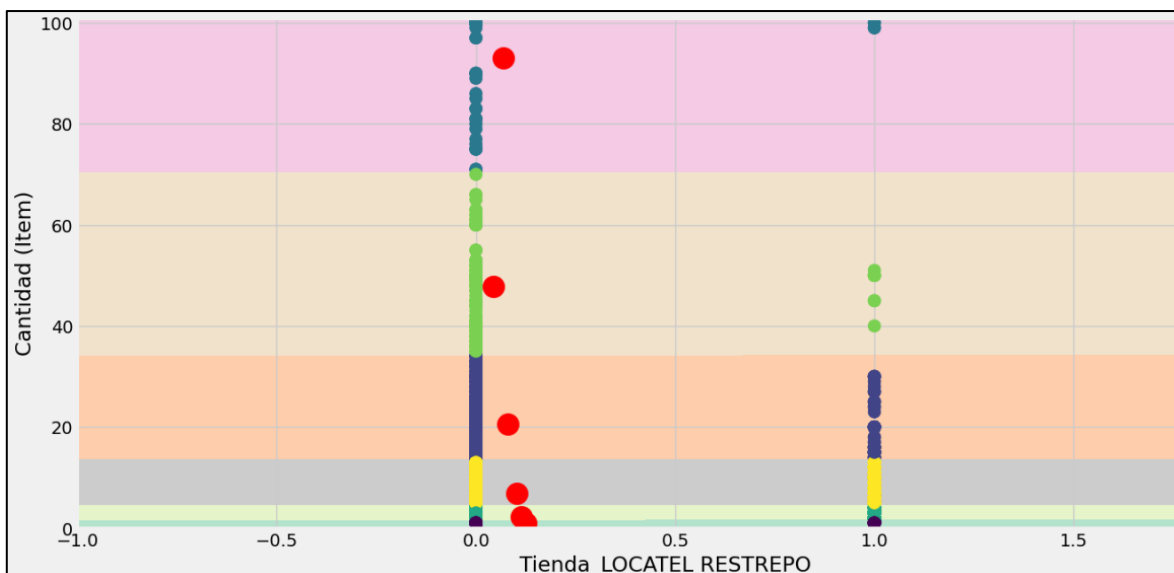


Ilustración 35 Distribución de Agrupación entre variable de Tienda – Cantidad. Creación Propia.

En esta unión de variables, la distribución que realiza el modelo refleja que los grupos no cuentan con un porcentaje equitativo de los datos, dado que, si bien el grupo 1 cuenta con registros de cantidades menores a 10 ítems por factura (número que estaría dentro del rango promedio de compra) el grupo 6 que tiene más de 75 ítems por compra pero no tiene la misma cantidad de registros agrupados para el segmento. El resultado de este algoritmo se podría utilizar con mejor efectividad que los anteriores, dado que si permite una mejor visual de los resultados generados y se podría utilizar para segmentar. Sin embargo, la agrupación sería muy capilar y habría que analizar con mayor profundidad los registros que está acoplado en cada uno de los clúster para entender el comportamiento y no sería escalable si se unificara la totalidad de establecimientos comerciales, solo serviría si se decide trabajar una tienda en específico.

#### 9.2.5. Análisis de Patología Cuidado Personal:

Continuando con el objetivo del proyecto, para estudiar las patologías asociadas al consumo, las variables a analizar son la Edad con la patología consumida de cada registro.

Para entrenar el modelo se debe seleccionar una de las patologías que mayor representación se tenga en la base, por lo que la categoría a analizar será Cuidado Personal, ya que la ilustración 3 demuestra que dicha categoría es la de mayor peso en las ventas de la compañía. La categoría Cuidado Personal es la más importante dentro de Locatel no solo por el número de ventas realizadas durante el periodo estudiado; se explica por el crecimiento del mercado de cosmética activa que está experimentando el país y las nuevas necesidades de cuidado que experimentan no solo las mujeres, también los hombres aumentan su consumo (EMIS, 2022).

El resultado de la nueva distribución realizada entre la categoría Cuidado Personal y Edad tiene un comportamiento similar al estudio realizado entre Edad y Cantidad Items, ya que se generan 4 grupos de segmentación. A continuación se muestra el resumen de clústeres:

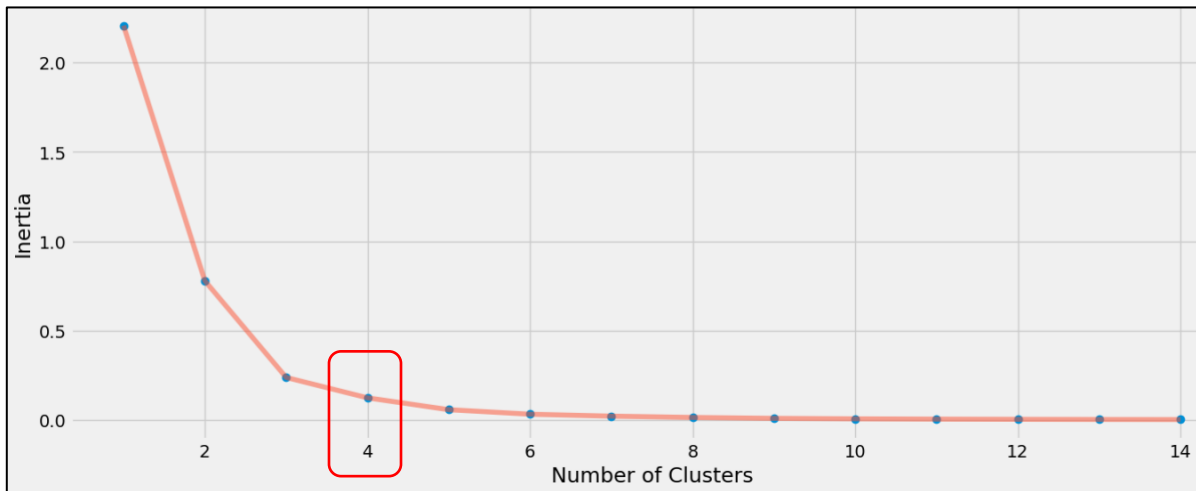


Ilustración 36 Agrupación variable de Patología Cuidado Personal - Edad. Creación Propia.

Con el número de clústeres establecidos, se realiza nuevamente la codificación de las dos variables a estudiar para determinar su distribución en el modelo, a continuación se presenta el modelo de agrupación generado:

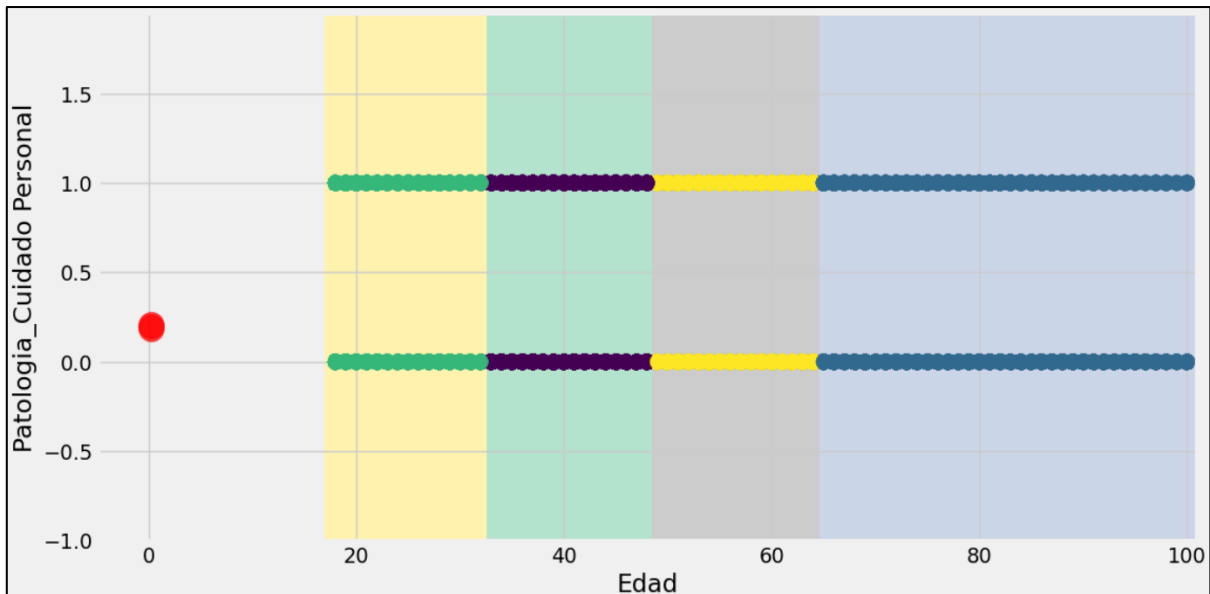


Ilustración 37 Distribución Agrupación de Patología Cuidado Personal – Edad. Creación Propia.

Dado el resultado de la agrupación, se generan segmentos similares a los anteriores grupos de variables; no se logra determinar características propias de cada clúster. En el último grupo generado se evidencia que están consignadas las edades mayores a 65 años (según la ilustración 8 se agrupan el 25% de los datos). A pesar que se refleja una distribución homogénea, los grupos no permiten entender un fenómeno detallado de la información que cada uno agrupa, ni las cantidades y pesos que tiene cada grupo en la venta de Cuidado Personal.

Por lo cual, se sugiere analizar otro tipo de procedimiento para analizar y agrupar la información, ya que la interacción de la información de las variables y el algoritmo K-means no determinan un resultado contundente para generar la opción de segmentación al integrar una variable normalizada, como si se evidenció un mejor desarrollo al ingresar dos variables numéricas (Cantidad Ítems y Edad).

### **9.2.6. Interacción entre variables de Valor Facturado, Patología y Edad (Grafico Burbujas):**

Al necesitar la exploración otros procedimientos de agrupación para validar su efectividad, se utilizará el gráfico de burbujas, en el cual las dimensiones de los datos significarán el tamaño de las circunferencias. Para este procedimiento, se compilaron el valor de la compra y las edades para estudiar cual es el comportamiento de las patologías más relevantes de las Tiendas 116 y Plaza Américas (de acuerdo a la ilustración 23 se tomaron las 6 patologías con mayor facturación de cada tienda en el periodo). En el resumen de las gráficas se puede evidenciar que para las dos tiendas, los productos categorizados como Cuidado Personal son los más relevantes; para Calle 116 se obtiene una facturación de 140 millones, entre edades de 50 a 53 años, mientras que para Plaza de las Américas 210 millones, en edades entre 43 y 46 años. A demás de Cuidado Personal, las tiendas comparten dos patologías similares; Maquillaje y Diabetes, las cuales en encuentran desde 42 hasta 48 años y facturación entre 40 y 50 millones para Maquillaje, en Diabetes el rango está entre 49 a 58 años y facturación de 45 a 50 millones.

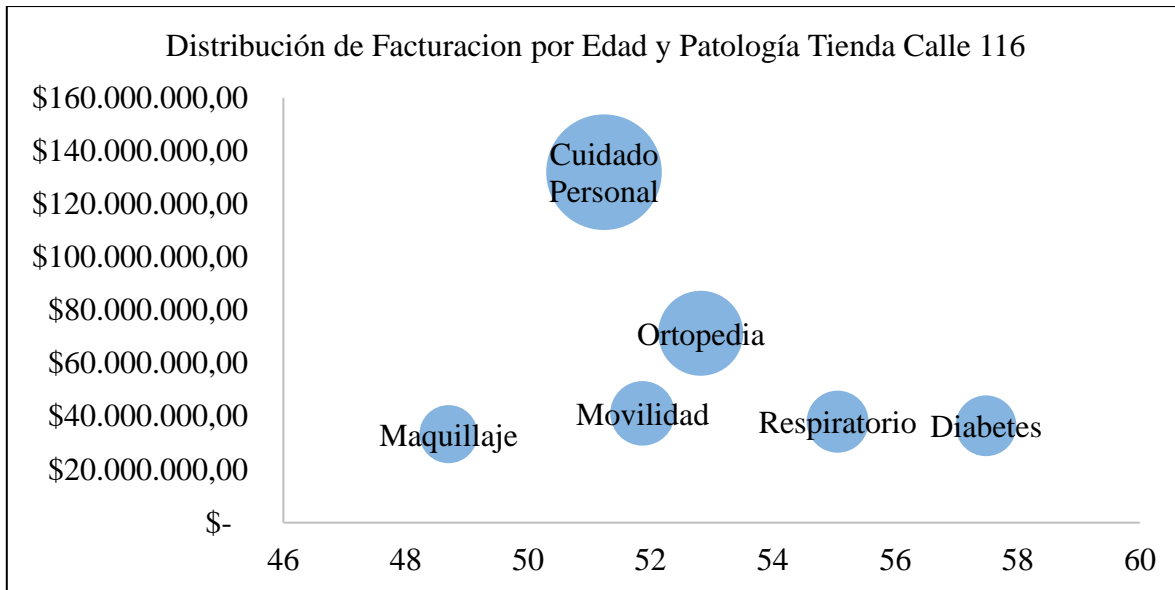


Ilustración 38 Diagrama Burbujas para variables de Valor Facturado, Patología y Edad en Tienda Calle 116. Creación Propia.

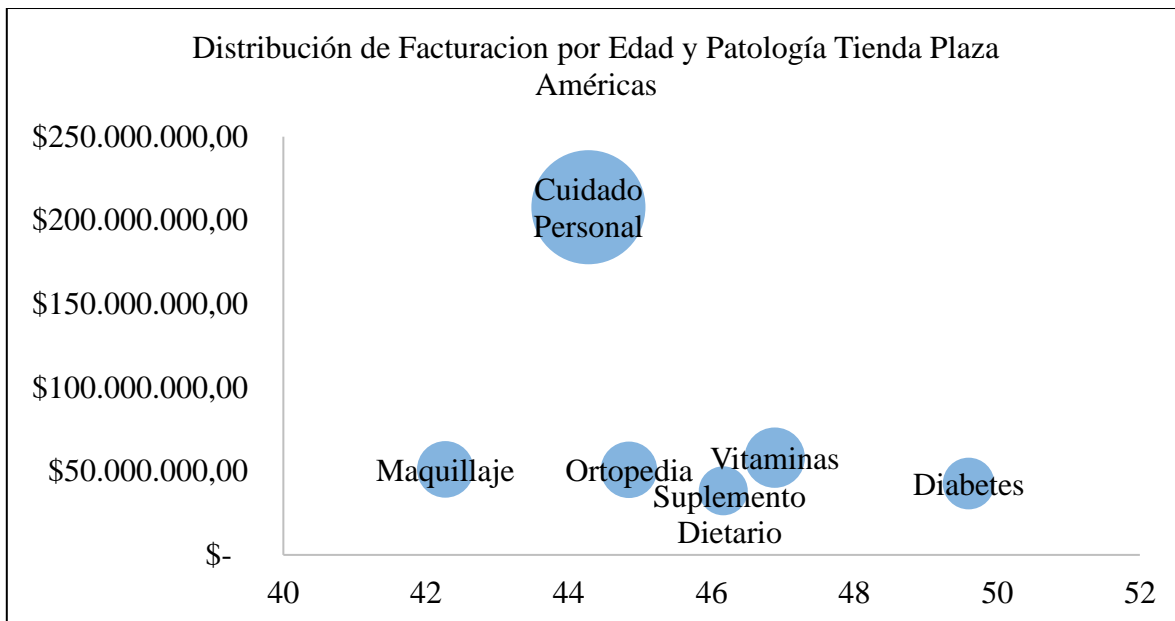


Ilustración 39 Diagrama Burbujas para variables de Valor Facturado, Patología y Edad Tienda Plaza Américas. Creación Propia

El resultado del proceso permite deducir que las categorías de dermocosmética son las que en su mayoría le permiten a Locatel tener un alto nivel de facturación en las tiendas (como lo son Cuidado Personal y Maquillaje). En este resultado se resume el análisis de los

hallazgos generados a través del proyecto, ya que rectifica la posibilidad de enfocar los esfuerzos de segmentación para Locatel a través de las categorías de producto. El resultado de este modelo expone una opción de segmentación la cual podría utilizar la compañía para enfocarse en alguna patología que requieren trabajar en cada una de las tiendas; hacer campañas específicas según el objetivo comercial que se trace por mes o año para los establecimientos comerciales y con el entendimiento del contexto social de los establecimientos comerciales.

#### **9.2.7. Análisis PCA:**

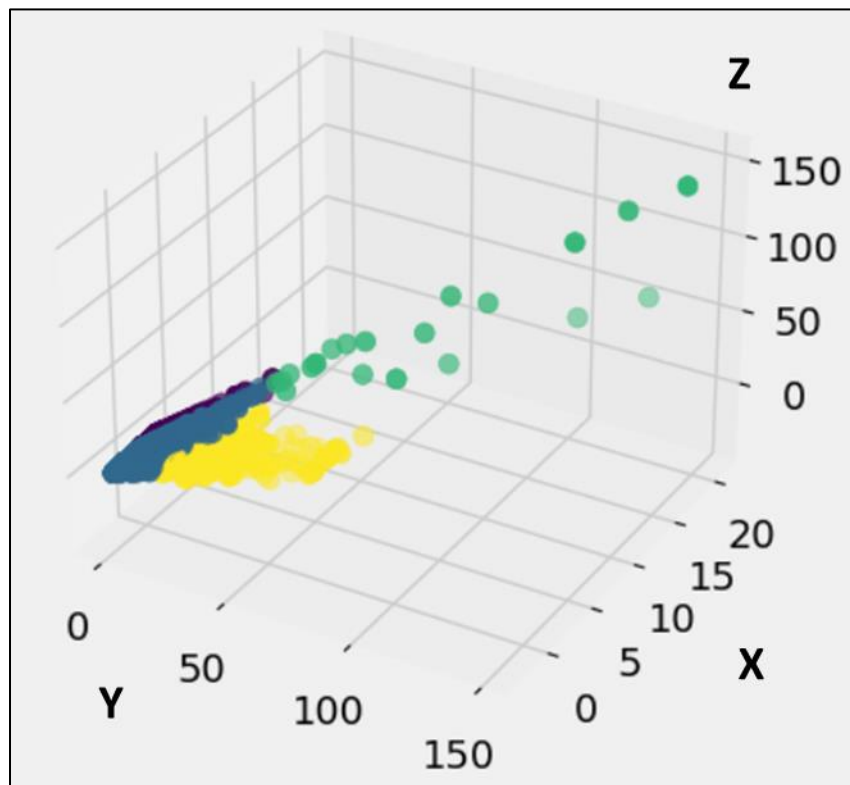
Con el ánimo de analizar otro tipo de procesos, se integra en el estudio el Análisis de Componentes Principales (Principal Component Analysis en inglés) el cual permite disminuir las dimensiones de un estudio y centrarse en las que mayor peso tienen dentro del modelo a través de los vectores de la matriz generada de la covarianza de las variables, dado que en el primer componente se agrupa la mayor cantidad de variaciones en los datos y los demás componentes reúnen cada vez menos variaciones, por ello se genera la reducción de espectros. Se utilizan las variables de Edad, Género, Tiendas, Cantidad y Valor Facturado, entrenando el modelo con 10 componentes y 4 clústeres. Una vez generado el resumen del PCA, los componentes más relevantes son el género masculino, patología de Cuidado Personal, Alimentos, las tiendas de Cedritos, Galerías y Hayuelos.

El resultado de PCA permite que el proyecto se enfoque en las variables más representativas de la base. En este ejercicio, al utilizar 10 componentes principales se reducen las dimensiones iniciales que contemplan los datos en alrededor de 230 espectros, como se muestra a continuación:

```
{'PC0': 'Valor Venta sin IVA (Item)', 'PC1': 'Edad', 'PC2': 'Cantidad (Item)', 'PC3': 'Genero_M', 'PC4': 'Patologia_Cuidado Personal', 'PC5': 'Tienda_LOCATEL GALERIAS', 'PC6': 'Tienda_LOCATEL CEDRITOS', 'PC7': 'Patologia_Alimentos', 'PC8': 'Tienda_LOCATEL CEDRITOS', 'PC9': 'Tienda_LOCATEL HAYUELOS'}
```

#### Ilustración 40 Resumen Componentes Principales. Creación Propia

En el ejercicio el modelo se entrenó con 4 clústeres para validar la distribución de los datos, en los cuales el modelo va asignando los vectores necesarios para cada uno de ellos, los datos representados en la Ilustración 41 son 3 dimensiones compuestas por edades en el eje y, cantidades en el eje x y valor facturado en el eje z:



#### Ilustración 41 Clúster creado por PCA. Creación Propia

Dado el resumen que se utilizó en la clusterización de PCA, se identifica que la mayoría de datos están hasta los 50 años, con un promedio de compra de unidades adquiridas entre 5 y 15 ítems por facturación. Para la muestra que se selecciona es el género masculino quien mayormente está representado en la muestra y que aparece dentro de los componentes principales en la ilustración 40. Los productos con mayor presencia dentro de la muestra

son productos de cuidado personal, intravenosos y respiratorios, el detalle de los principales 20 productos se presentan a continuación:

<b>Nombre Producto</b>	<b>Patología</b>
CRE.ELEMENTAL CORP.ACTE.MACADAMIA X500ML	Cuidado Personal
GASA ALFASAFE ESTERIL NO TEJIDA 3X3 X6	Infecciones
CRE.ELEMENTAL CORP.TE VDE.Y PEP.X500ML	Cuidado Personal
JERINGA MEDISPO INSULINA 1ML 27G X1/2	Intravenosas
CANULA BIOPLAST NASAL OXIG.2M	Respiratorio
JERINGA ALFASAFE 3ML 23G X1	Intravenosas
TAPABOCAS BEGUT ADULTO EMPQ.IND.AZX50	Protección Medica
JERINGA ALFASAFE INSULINA 1ML 29G X1/3	Intravenosas
SONDA MEDEX NELATON URETRAL CAL 12	Cuidado Intimo
AGUA BRISA SIN GAS X600ML	Alimentos
JERINGA ALFASAFE TUBERCULIN 1ML 25GX5/8	Intravenosas
CRE.DEN.COLGATE LUMINOUSWHITE CBON.X106G	Cuidado Dental
UVAS CHEVERES TRIUNFO CUB.X40G	Alimentos
JERINGA MEDISPO 5ML 21G X1 1/2	Intravenosas
TONICO FACIAL ELEMENTAL LBR.ALC.X250ML	Cuidado Personal
JERINGA ALFASAFE 10ML 21G X1 1/2	Intravenosas
OF.TARRITO ROJO JGB TRADI.GRT.50G X380G	Vitaminas
DIBEN VAINILLA X200ML	Suplemento Dietario
GASEOSA COCA COLA X400ML	Alimentos
CHOCTINA.JUMBO MANI X40G	Alimentos

Ilustración 42 Productos alta rotación generado por PCA. Creación Propia

Los productos reflejados en la gráfica anterior pueden servir para que Locatel fortalezca sus estrategias comerciales dado que son productos de categorías de alta demanda como lo son los alimentos. A demás, la posibilidad de trabajar con las categorías de dermocosmética (Cuidado personal, dental, dermatología) como fuente principal de la segmentación se fortalece ya que PCA refleja el peso que los productos de la categoría tienen dentro de la base de datos. Según el resumen del análisis de componentes principales, se contemplan las tiendas Cedritos, Galerías y Hayuelos (sugeridas por el PCA) en rangos de edad para evaluar la cantidad de perfiles de clientes por tienda, propuesta que se evalúa en el cluster número 1 generado en el algoritmo anterior:

Tienda	18-25	26-40	41-55	56-70	71-85	>85	Total
<b>CEDRITOS</b>	6.583	12.537	17.907	18.130	8.279	990	64.426
<b>GALERIAS</b>	4.949	13.286	16.575	19.813	9.670	921	65.214
<b>HAYUELOS</b>	6.761	19.166	20.897	17.259	4.609	457	69.149
<b>Total</b>	18.293	44.989	55.379	55.202	22.558	2.368	198.789

Ilustración 43 Distribución por Tienda y Edad según PCA. Creación Propia.

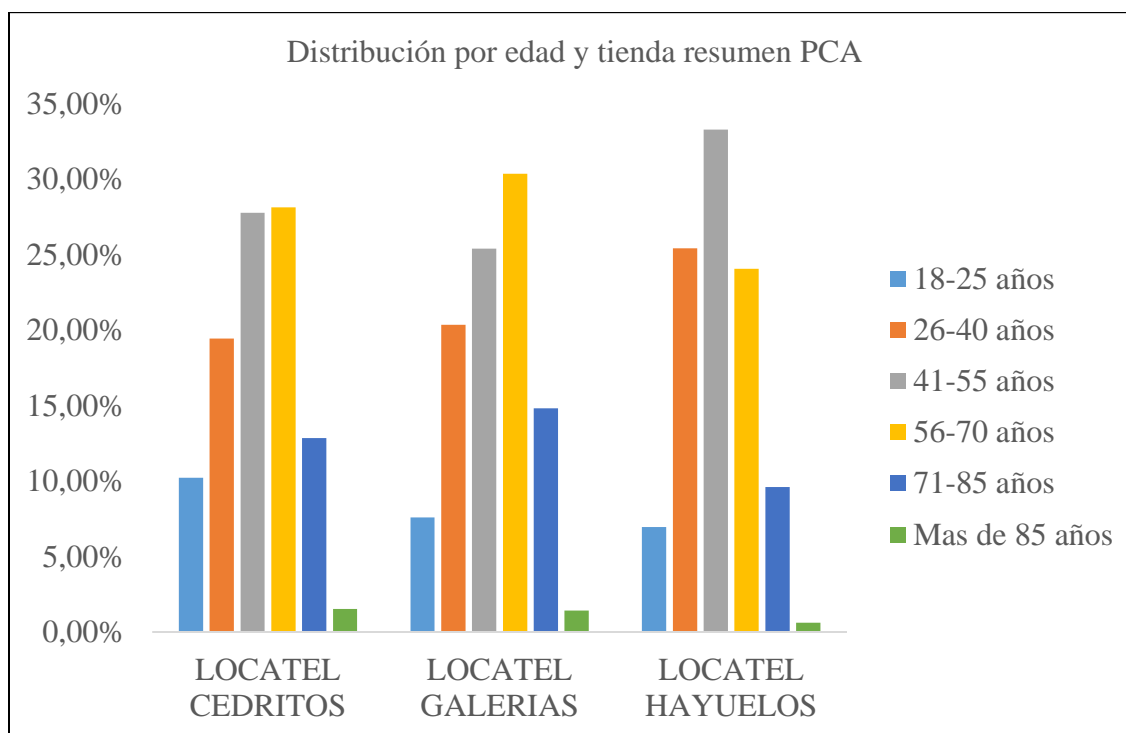


Ilustración 44 Gráfica distribución PCA por tienda y edad. Creación Propia.

El resultado demuestra que las tiendas comparten una distribución normal, estando el mayor peso de edades entre 26 y 70 años de edad. El rango de edad determinado se puede utilizar para generar una estrategia de segmentación por tienda, dado que se puede reconocer la edad del cliente objetivo. Para la tienda Hayuelos, el objetivo está enfocado en los usuarios con edad entre 26 a 55 años, mientras que en Cedritos y Galerías tienen como objetivo los usuarios con edades entre 41 a 70 años.

Tienda	18-25	26-40	41-55	56-70	71-85	>85	Promedio
<b>CEDRITOS</b>	11,67	12,74	17,17	18,12	8,39	1,37	11,58
<b>GALERIAS</b>	6,70	9,72	13,29	15,97	7,98	0,84	9,08
<b>HAYUELOS</b>	4,91	9,58	13,15	10,78	4,36	0,28	7,18
<b>Promedio</b>	7,76	10,68	14,54	14,96	6,91	0,83	9,28

Ilustración 45 Facturación promedio en millones por rango edad según PCA. Creación Propia.

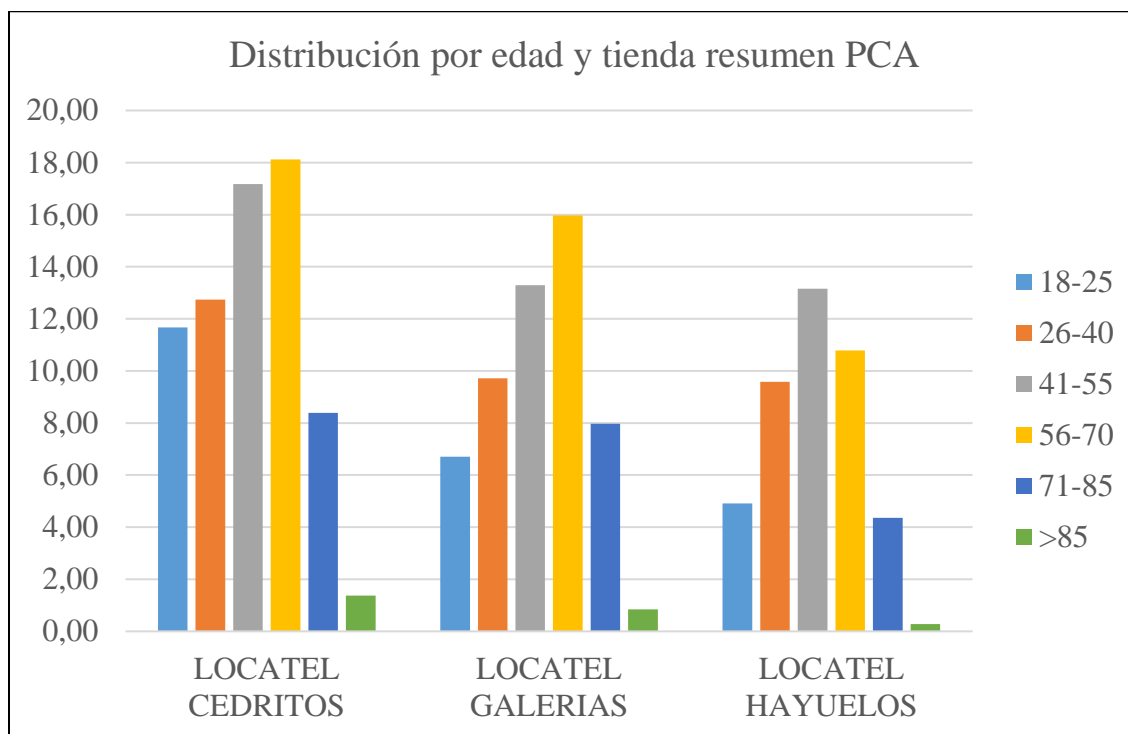


Ilustración 46 Gráfica distribución facturación promedio por tienda y edad PCA. Creación Propia.

La distribución del promedio facturado por tienda evidencia que el rango de 41 a 70 años es aproximadamente el 40% de las ventas realizadas en los 3 establecimientos comerciales analizados, evidenciando la tendencia que podría trabajar Locatel para otras tiendas en su enfoque de estrategia comercial. En los 3 establecimientos comerciales se puede reconocer una distribución normal de las ventas promedio por rango de edades, siendo la tienda Cedritos (sector de estrato 5) la tienda con mayor valor facturado, teniendo un promedio total de venta por COP \$11.5 millones, COP \$2.5 millones por encima de la

segunda tienda con mayor facturación la cual es Galerías (sector de estrato 3). El resumen realizado para este ejercicio le permitirá a Locatel tener una opción de segmentación detallada; enfocar la estrategia comercial en los productos de Cuidado Personal y Vitaminas, con promociones comerciales apoyándose en el consumo de golosinas para generar la necesidad en el rango de edades de los 40 a 70 años de edad en todas las tiendas de Bogotá.

La siguiente es la varianza de los 10 primeros componentes principales:

	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Comp8	Comp9	Comp10
Edad										
Cantidad (Item)	0	-1	-0	0	0	-0	0	0	-0	-0
Valor Venta sin IVA (Item)	1	0	-1	0	-0	0	0	0	-0	0
Genero_M	1	0	1	0	0	0	0	0	0	0
Genero_N	0	-0	-0	-1	-0	-0	0	0	0	0
...	-0	0	-0	0	0	0	-0	-0	-0	-0
...	...	...	...	...	...	...	...	...	...	...
Patologia_Venas Varices	0	0	0	0	0	0	-0	-0	-0	-0
Patologia_Verruga	0	0	0	0	0	-0	0	-0	0	-0
Patologia_Vertigo	0	-0	0	-0	0	-0	0	-0	-0	0
Patologia_Vitaminas	0	-0	0	-0	0	-0	0	-0	-0	0
Patologia_Zona Intima	-0	0	-0	0	0	-0	0	-0	-0	-0

Ilustración 47 Varianza para componentes principales. Creación Propia.

### 9.2.8. Resumen Cluster Generados:

A continuación se presenta una muestra de cada uno de los 4 clúster generados, en los cuales se tiene en cuenta la patología, género y los promedios de edad, cantidad y valor facturado por cada categoría (Para el ejercicio se tomaron los datos con facturación media mayor a COP \$100 mil):

Patología	Genero	Edad Media	Cant Media	Valor Facturado Media
Antipsicóticos	F	44,4	22	649.214
Control Peso	F	47,4	11	85.899
	M	49,5	16	215.949
Crecimiento	F	32,0	35	1.211.925
	M	55,3	33	1.199.683
Déficit Atención	F	42,0	25	385.525
	M	77,0	10	183.500

Encefalopatía	F	40,5	13	277.288
	M	34,0	15	330.525
Endometriosis	F	51,7	150	930.013
	M	21,0	10	77.550
Glucemia	F	47,3	22	44.204
	M	52,5	44	302.778
Hipercolesterolemia	F	46,3	14	274.793
	M	53,1	11	203.256
Miastenia	F	52,5	30	608.700
Movilidad	F	48,2	10	287.358
	M	51,3	10	298.498
Parkinson	F	43,8	18	218.443
	M	56,1	19	133.540
Signos Vitales	F	39,0	10	272.647
	M	57,7	10	235.826
Síndrome De Sjogren	F	61,5	13	199.763
	M	60,0	30	253.200
Testosterona	F	49,0	15	336.075
Trombosis	F	37,4	10	382.250
	M	39,0	10	377.525

Ilustración 48 Muestra 1er Clúster. Creación Propia

En el clúster 1 se evidencia que la patología de Crecimiento es la que mejor promedio de facturación tiene con un valor de COP 1.2 millones, en el cual, el comportamiento entre géneros es similar, dado que en promedio se diferencian solo por 2 unidades a favor de las mujeres, caso diferente con la edad promedio, en la cual la de las mujeres ronda los 33 años y en los hombres los 55 años. En el análisis del clúster, además de Crecimiento, la Endometriosis (afectación en sistema reproductivo femenino) es la categoría que tiene mayor cantidad de unidades vendidas de todo el grupo, siendo las mujeres con edad promedio de 51 años las consumidoras de la categoría, en la cual solo se encuentra el medicamento METRINELLE TAB.2MG X28 con valor unitario COP \$77.550, el cual se vendió únicamente en la tiendas de Cedritos y Calle 100 (Sectores con estratos entre 4 y 6). En este primer clúster se encuentran otras categorías con un alto ticket

promedio de compra como lo es los Antipsicóticos, Déficit de atención, Encefalopatía, Glucemia y Trombosis.

<b>Patología</b>	<b>Genero</b>	<b>Edad Media</b>	<b>Cant Media</b>	<b>Valor Facturado Media</b>
Calcificación	F	28,0	20	155.882
Colesterol	F	55,0	20	159.075
Cuidado Personal	F	51,0	51	135.516
	M	34,7	48	131.870
Digestivo	F	32,0	20	79.302
	M	68,0	20	100.252
Presión Arterial	F	61,0	10	126.060
Respiratorio	F	45,0	10	95.725
	M	80,0	10	163.026
Terapia Física	F	61,0	10	326.681

Ilustración 49 Muestra 2do Clúster. Creación Propia

El resumen del segundo clúster la categoría con mayor promedio de valor facturado es Terapia Física con COP \$326 mil y 10 unidades promedio por compra; siendo únicamente las mujeres las usuarias que consumen esta categoría en edad promedio de 61 años, el producto que se vende en este grupo es MANTA TERMICA PLENIA VINOTINTO 60X70 únicamente en la tienda de Hayuelos. En el grupo analizado aparecen las patologías de Calcificación y Colesterol las cuales no habían sido relevantes para el estudio dado que en la base general no obtienen un peso significativo; sin embargo, si se refleja un fenómeno a estudiar, dado que la edad promedio de los usuarios que adquieren productos de calcificación está sobre los 28 años, edad que no sugiere la necesidad de utilizar productos de esta categoría (dado de la descalcificación en su mayoría se ve reflejada en personas con más edad), lo que permite inferir que la compra la realizan usuarios diferentes a los que consumen el producto. Una de las categorías más representativas de la base es Cuidado Personal, la cual se encuentra determinada en este grupo con edad promedio de las mujeres en 51 años y hombres con 34 años; para los dos

grupos poblacionales el promedio de productos vendidos está alrededor de 30 unidades siendo los protectores solares, las crema faciales y los jabones los productos de mayor relevancia para el grupo.

<b>Patología</b>	<b>Genero</b>	<b>Edad Media</b>	<b>Cant Media</b>	<b>Valor Facturado Media</b>
Alimentos	F	45,2	72	178.560
	M	48,0	64	152.656
Alopecia	F	41,8	10	162.023
	M	45,0	10	72.520
Alzheimer	F	48,5	10	53.950
	M	52,0	10	178.400
Antipsicóticos	F	48,0	10	466.050
Colon Irritable	F	41,3	13	190.781
Crecimiento	F	33,0	15	477.425
	M	75,0	30	1.101.750
Cuidado Personal	F	46,8	61	158.749
	M	51,5	52	132.648
Diabetes	F	51,7	23	131.338
	M	56,0	26	149.508
Epilepsia	F	20,0	15	16.900
	M	35,0	20	152.200
Hospitalización	F	47,4	15	98.075
	M	42,5	20	195.189
Movilidad	F	47,6	10	94.056
	M	54,4	10	233.852
Próstata	F	19,0	10	150.550
	M	54,5	10	59.725
Recuperación Tejidos	F	18,0	40	152.800
	M	23,0	10	177.100
Trastorno Psicótico	M	51,0	10	165.350

Ilustración 50 Muestra 3er Clúster. Creación Propia

En el tercer clúster generado la categoría con mayor relevancia en el grupo es Crecimiento, facturado únicamente en la tienda Chapinero, dado que tiene una venta promedio para el género masculino por COP \$1.1 millones (con respecto a los COP \$477 mil por parte del género femenino), con respecto a las demás categorías las cuales tienen promedios de venta que oscilan entre los COP \$120 mil y COP \$230 mil (solamente

Antipsicóticos aparece por encima de los COP \$460 mil). La categoría de Antipsicóticos vuelve a aparecen entre las de mayor peso en un grupo, dado que ya había aparecido con el clúster número 1, con el cual no comparten las mismas tiendas pero si un sector de la ciudad, dado que el primero están las tiendas de Calle 116 y Calle 100 y en este tercer grupo está la tienda de Chapinero, aproximadamente 7 km de distancia entre las tiendas más lejanas anteriormente mencionadas (Google. s. f). El sector en el que se encuentran ubicadas las 3 tiendas es un punto centripeto con respecto a los centros de salud, en los cuales se tratarían problemas hormonales y mentales, lo que podría aprovechar Locatel para enfocar el esfuerzo estratégico en alinear la venta de los medicamentos directamente con los hospitales, clínicas y consultorios médicos de una forma institucional, con descuentos especiales para dichos centros y ganando mercado nuevo en enfermedades con un alto grado de complejidad.

<b>Patología</b>	<b>Genero</b>	<b>Edad Media</b>	<b>Cant Media</b>	<b>Valor Facturado Media</b>
Presión Arterial	F	40,2	12	153.915
	M	47,0	10	180.000
Saturación Oxigeno	F	60,5	10	74.874
	M	24,0	10	112.983
Spa	F	21,0	10	142.815
	M	65,0	10	142.815
Triglicéridos	F	46,0	10	141.050

Ilustración 51 Muestra 4to Clúster. Creación Propia

Para este último clúster aparece una categoría nueva con un importante peso en el grupo y es Presión Arterial, la cual tiene una facturación promedio en los hombres de COP \$180 mil con edad promedio de 47 años y para las mujeres una media de facturación alrededor de los COP \$154 mil, con edad promedio de 40 años. De igual forma, otra de las categorías con mayor peso en el grupo es la categoría Spa, con una facturación promedio de COP \$142 mil tanto para hombre como mujeres, solo existe un cambio en las edades promedio,

dado que para las mujeres está sobre los 21 años y en hombres sobre los 60 años, los productos que se comercializan son KIT DHEMS VIT C+HIALURONICO+ ROLLER.

Las patologías que se presentan como muestra para este grupo se encuentran en su mayoría en las tiendas de Chapinero, Plaza Américas y Hayuelos.

### **9.2.9. Buyer Personas para Locatel:**

La teoría de Buyer Persona permite a las compañías diseñar de forma idónea sus productos y estrategias comerciales, dado que entendiendo al consumidor se entiende su necesidad (Upwork, s. f.). La intención de esta evaluación es representar el prototipo de cliente ideal de las tiendas, funciones como estrategia comercial y perfilar los prototipos de clientes que se pretende tener y que se encuentren en los establecimientos comerciales priorizados. Al ser una determinación propuesta, no propia de la base de datos, la determinación de los Buyer Persona se hará bajo la modalidad de decisor (usuario que consume directamente el producto para solucionar su necesidad propia) buscando que se ejemplifiquen en algunos prototipos las diversas edades, enfermedades, necesidades y productos que se pueden encontrar comúnmente en el mercado farmacéutico.

Según la información encontrada en los clústeres del PCA, se planea utilizar el Buyer Persona como otra opción de segmentación para Locatel, por lo que a continuación se comparte la tipología de clientes y el comportamiento en las tiendas definidos bajo un conocimiento previo del mercado:

- a. Joven mujer menor de 30 años, sufre de hipoglucemia y le preocupa no quedar embarazada.
- b. Mujer mayor de 40 años, interesada en su imagen personal y el cuidado de sus hijos.

- c. Hombre mayor de 60 años, con dolencias musculares, óseas y con inicios de diabetes.
- d. Hombre entre 30 y 40 años, interesado en mantener su energía y alimentarse sanamente, sufre constantemente de afectaciones respiratorias.

Con los 4 perfiles determinados a evaluar, se realiza una comparación de las condiciones de cada Buyer Personas y la base de datos después de ser tratada bajo la limpieza de datos. Según la interacción de los clústeres y cada cliente ideal, se muestra a continuación el resumen de la distribución general de los 4 prototipos de clientes para las tiendas Cedritos, Galerías y Hayuelos (la probabilidad de encontrar cada Buyer Persona en cada una de los establecimientos comerciales evaluados) evaluados en el cluster 1 determinado en el algoritmo:

<b>Tienda</b>	<b>Buyer A</b>	<b>Buyer B</b>	<b>Buyer C</b>	<b>Buyer D</b>
Cedritos	30%	24%	25%	21%
Galerías	24%	23%	28%	25%
Hayuelos	23%	28%	20%	29%

Ilustración 52 Distribución Buyer Persona. Creación Propia.

Se puede evidenciar que cada uno de los grupos (los porcentajes son la probabilidad de encontrar cada Buyer Persona por cada tienda, realizando filtros en las variables de edad, género y patología, generando un ponderado entre las variables que cumplan las condiciones de cada tipo de buyer persona) tiene particularidades en cuanto a las variables que lo componen; para la tienda Hayuelos su cliente objetivo se encuentra entre los 30 y 40 años, con capacidad económica promedio e interesados en Cuidado Personal, Vitaminas y Alimentos. Por el lado de Cedritos su enfoque está en jóvenes con interés en Salud Sexual y Diabetes. Para Galerías la estrategia de segmentación deberá ir enfocada en adultos mayores de 60 años, interesados en mantener un estado de salud óptimo por sus dolencias

generales. La creación de los perfiles para el Buyer Persona permite describir fenómenos o comportamientos que la base de datos no logre reflejar, incluso fortalecidos por el conocimiento del negocio de los usuarios involucrados en su diseño.

### **9.3.Exploración Segmentación:**

Según las diferentes opciones que se estudiaron con el algoritmo de k-means, podemos evidenciar que algunos grupos analizados no generan un resumen claro y acorde a la necesidad de la segmentación, ya que el resultado generado refleja sesgos informativos y conlleva a tener información incompleta de los registros por grupo, lo que no permite reconocer a cabalidad el enfoque que debería tener cada uno de los clústeres. Sin embargo, el análisis de las variables Cantidad Ítems y Edad obtiene un desarrollo de segmentación inicial que se podría analizar para desarrollar su mercado según el grupo poblacional de cada uno de los clústeres. En el modelo de burbujas se realiza una agrupación por patología que le serviría a Locatel para realizar actividades comerciales según los medicamentos de dicha categoría, comprendiendo características como las edades en las que podría enfocar su estrategia. Por parte del protocolo PCA se enfoca capilarmente en los componentes más importantes que para el caso de la base serán las 3 tiendas establecidas que pudieran ser el foco de Locatel para determinar acciones comerciales y aumentar las ventas teniendo en cuenta las patologías, la edad y el género a configurar.

En el desarrollo del protocolo PCA, el resultado del buyer persona expone algunas opciones de intervención más allá de la segmentación por tienda; los alimentos toman fuerza dentro de la venta en Locatel, como lo presenta la ilustración 42, dentro de los productos más importantes de la base hay bienes comestibles que puedan jalonar las ventas de categorías más importantes para la compañía y no tengan el desarrollo esperado. De

igual manera, la ilustración 43 permite evidenciar en la muestra tomada que el mayor porcentaje de edades de clientes de la compañía está entre 30 y 70 años, insigth que le permitirá a marketing y comercial determinar las estrategias de impacto por categoría que vaya alineado a la edad determinada, dado que se reconoce el grupo poblacional en el cual la estrategia tendrá la mayor probabilidad de éxito.

## **10. Plan y recomendaciones de implementación y aplicación**

### **10.1. Recomendación No 1:**

La primer recomendación para la compañía es la estructuración de los procesos para la recolección, gobernanza y procesamiento de datos sobre la información recopilada de los clientes, productos y ventas, dado que en el data mining desarrollado, varios registros relevantes del estudio se pierden al ser información que desvía el resultado del mismo, como lo son los datos outliers, nulos o vacíos. Es importante que desde su sistema de CRM se garantice que en la creación de clientes, los responsables del cargue de información aseguren que sea completa y fidedigna, el sistema no debe permitir que los datos se guarden si hay alguna información errada, debe devolver el proceso hasta realizarse satisfactoriamente; por ejemplo utilizar fecha de nacimiento, reglas en el sistema para que no salgan cantidad y valor de producto negativo y el género sea obligatorio en el sistema para poder avanzar con el registro, sea cualquier categoría sin distinción. El ajuste del proceso no tendrá ningún costo adicional dado que la compañía cubre el costo del auxiliar, más el costo de su base de datos y sistema CRM.

### **10.2. Recomendación No 2:**

La segunda recomendación es adquirir información externa para la estrategia de CRM conforme al objetivo de segmentar por patología, ya que si bien el estudio demuestra una clara superioridad en ventas de Cuidado Personal, la compañía no cuenta con la información detallada de sus productos para determinar que enfermedades pueden combatir. La compañía podría evaluar la contratación de una empresa de estudios de mercado para recolectar la mayor cantidad posible de información por medicamento y

patología; ya con la información condensada se puede automatizar el modelamiento de la información para adquirir los grupos de interés en la estrategia comercial. De no ser posible la adquisición de un estudio de mercado, se podrían asignar dentro de la práctica de algún pasante para que se pueda levantar la base de datos con fuentes oficiales y públicas, incluso a través de los diferentes proveedores de Locatel.

### **10.3. Recomendación No 3:**

La tercera recomendación es que Locatel utilice el contexto y segmentos generados en el proyecto para evaluar los posibles escenarios en la creación de su estrategia comercial para aumentar sus ventas. Dado que la cantidad de información que recopilan es una fuente robusta para mantener una estrategia basada en el análisis de datos, la compañía debe continuar la exploración de data mining y generar e incluir variables como dirección de residencia, estratos, personas a cargo de los usuarios que ayudarán a enfocar de mejor forma los modelos a crear. Similar a la anterior recomendación, la compañía podría buscar a un pasante con habilidades en programación y análisis de información que entrene el modelo o diseñe un nuevo modelo sobre información relevante que pueda adquirir la compañía.

### **10.4. Recomendación No 4:**

La cuarta recomendación es la creación del aplicativo móvil para mantener el objetivo de ser omnicanal que tiene Locatel. Con la creación de la app, la compañía podría llegar a nuevos mercados como lo es el juvenil; a pesar de no tener un fuerte porcentaje en las compras si pueden ser facilitadores de personas con edad avanzada a quienes les cuentas el uso tecnológico, además, esta población se enfocan en realizar compras por internet y que lo puedan realizar fácilmente por su dispositivo móvil, por lo que se podría estimular el

crecimiento de este nicho de mercado. Dentro del aplicativo, se puede ofrecer la oportunidad de fidelizarse con Locatel y los datos serán tomados desde la creación del usuario que se debe realizar para utilizar la aplicación, lo que garantizará que los datos recolectados son fiables y se optimiza el proceso de autenticación.

#### **10.5. Recomendación No 5:**

La quinta recomendación es realizar campañas comerciales y de fidelización para actualizar y modificar los datos de los usuarios. Es importante generar con regularidad esta actividad, dado que la información que se presenta en el mercado como lo es la vivienda, el número celular y correo electrónico en mayor medida es modificada con una alta probabilidad de recurrencia. Al generar esta estrategia, la compañía contará con información real que le permitirá estudiar de mejor forma su mercado y apoyar con nuevas ideas su estrategia de CRM.

#### **10.6. Recomendación No 6:**

El trabajo desarrollado con la data mining le permitirá a Locatel explorar varios escenarios que servirán como fuente de crecimiento para las ventas. En el buyer persona se determinan los grupos poblacionales más importantes en las ventas y el comportamiento de los mismos; en que tienda compran, que tipo de productos adquieren y las cantidades promedio de cada uno, por su parte, cada cluster tiene diferentes productos que son los más relevantes, enfoque que se puede tener para reconocer el público objetivo al que Locatel aspirará a comercializar. Se deben tener en cuenta las varias propuestas y generar nuevo escenarios que se quieran explorar, incluso para reconocer si una nueva categoría o producto tendría oportunidad en el nicho de mercado y cuál sería su posible participación.

## 11. Conclusiones

A través de la investigación realizada (benchmarking) sobre las prácticas de estrategias de CRM, el objetivo que tiene Locatel para realizar una segmentación por patologías y consumo de medicamentos es una práctica poco explorada y se debe garantizar que los datos de cada registro, compra o usuario estén dentro de los parámetros de calidad establecidos que aseguren su completitud y fiabilidad. La estrategia requiere mayor información a los cálculos comerciales y rotación de los productos, dado que se debe tener en cuenta aspectos fundamentales como componentes activos, laboratorios, regulaciones del país, datos económicos y demográficos del mercado analizado. Por lo cual, la compañía deberá fortalecer las bases de datos para poder realizar investigaciones con mayor eficacia y generar estudios de mercado para integrar el análisis de variables exógenas que influirán en su estudio y estrategia a desarrollar.

Según el desarrollo del proyecto se puede concluir que la utilización del data mining tiene variedad de enfoques y usos dentro de cualquier sector económico y productivo, aspecto que se debe seguir explorando dentro de las organizaciones, en especial las empresas categorizadas como retail, dado el nivel de información que constantemente captura y que está en crecimiento por el desarrollo de dicho sector económico. Es importante trasladar la toma de decisiones bajo los análisis de datos para fundamentar de mejor forma el resultado esperado. El modelo de segmentación generó varios escenarios, los cuales agrupan el comportamiento general de la base de datos, información que Locatel debe utilizar para describir el tipo de cliente por tienda, su comportamiento de compra y la estrategia a realizar para aumentar las ventas cruzadas.

El sector farmacéutico y cuidado personal tiene un auge robusto dentro de Colombia; las categorías que se producen y comercializan en las compañías del gremio reflejan la capacidad que posee el mercado de adquirir productos de alto valor, con gran calidad de materiales y uso especializado. El informe final compartido con la compañía, resume el desarrollo del proyecto, el resultado y los principales hallazgos, con el cual la compañía tomará decisiones sobre sus procesos y estrategia. Por lo tanto, Locatel debe utilizar el reconocimiento que posee en el consumidor y la experiencia del mercado para seguir explorando nuevos nichos y productos, con el objetivo de alcanzar mayor participación en el mercado, mejorar su visibilidad en regiones intermedias del país y poder aumentar las utilidades.

El presente estudio para la segmentación se enfocó en la ciudad de Bogotá, dado que en la base de datos en su mayoría las tiendas pertenecen a esta ciudad, solo se tiene el registro de una en otra ciudad. Con el crecimiento que tiene la compañía es necesario seguir realizando la fidelización y recolección de los datos de las ciudades a las cuales están llegando sus establecimientos comerciales, para poder comparar los comportamientos de Bogotá con las demás ciudades, evaluar si se comportan de igual forma las categorías más importantes en la ciudad o se perfilan nuevas categorías. Con esta información, se puede hacer relación de variables y entender si las estrategias comerciales que se necesitan implementar se pueden realizar para todas las tiendas o para algunos grupos de ellas.

### Referencias bibliográficas

- Casariego-Sarasquete, N. M. (2019). *Metodología de análisis y segmentación de clientes usando secuencias de comportamiento* [MasterThesis].  
<https://reunir.unir.net/handle/123456789/9481>
- Chapman et al (2000). *CRISP-DM 1.0*.
- Companies—Company Profile—Tearsheet*. (s. f.). Recuperado 22 de mayo de 2023, de  
<https://www-emis-com.ez.urosario.edu.co/php/companies?pc=CO&cmpy=2612891>
- DANE (2023, Febrero). *Boletín Técnico Producto Interno Bruto (PIB) IV trimestre 2022pr*.  
Recuperado 14 de mayo de 2023, de  
[https://www.dane.gov.co/files/investigaciones/boletines/pib/bol\\_PIB\\_IVtrim22\\_produccion\\_y\\_gasto.pdf](https://www.dane.gov.co/files/investigaciones/boletines/pib/bol_PIB_IVtrim22_produccion_y_gasto.pdf).
- Distrito refuerza operativos por intoxicación con licor adulterado | Bogota.gov.co*. (2022).  
Recuperado 10 de junio de 2023, de <https://bogota.gov.co/mi-ciudad/salud/distrito-refuerza-operativos-por-intoxicacion-con-licor-adulterado>.
- Flórez López, R. (2005). *Análisis de la eficacia de los modelos de árboles de decisión para la gestión de políticas promocionales centradas en la segmentación de clientes*.  
Universidad de Sevilla. <https://idus.us.es/handle/11441/80105>
- Gomez Gonzalez, D. (2016). *Diseño de un programa de fidelización, basado en un modelo CLV para segmentación de clientes*.  
<https://repositorio.unal.edu.co/handle/unal/59057>
- Google (s. f.). *[Tiendas de Locatel en Bogotá según Google Maps]*. Recuperado el 17 de mayo de 2023, de  
<https://www.google.com/maps/search/tiendas+locatel/@4.700786,-74.0489785,14z>.

*Ley 1581 de 2012—Gestor Normativo—Función Pública.* (s. f.). Recuperado 14 de mayo de 2023, de

<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=49981>.

*How to Perform Market Basket Analysis.* (2022, Noviembre 4). 365 Data Science.

<https://365datascience.com/tutorials/python-tutorials/market-basket-analysis/>.

*Kennedy segunda localidad más poblada de la ciudad, con el 15% de los habitantes.*

(2019, Octubre 28). Observatorio de Desarrollo Económico. Recuperado el 17 mayo de 2023, de <https://observatorio.desarrolloeconomico.gov.co/dinamica-economica/kennedy-segunda-localidad-mas-poblada-de-la-ciudad-con-el-15-de-los-habitantes>.

Nik. (2021, Febrero 16). Pandas get\_dummies (One-Hot Encoding) Explained • datagy.

*Datagy.* <https://datagy.io/pandas-get-dummies/>.

*Market Basket Analysis / Kaggle.* (s. f.). Recuperado 14 de mayo de 2023, de

<https://www.kaggle.com/code/xvivancos/market-basket-analysis>

*Principal Component Analysis (PCA) in Python Tutorial.* (s. f.). Recuperado 14 de mayo de 2023, de <https://www.datacamp.com/tutorial/principal-component-analysis-in-python>.

Ramírez, J. L. M. (2019). *Usaquén es la localidad con la tasa global de participación laboral más alta de la ciudad.* Recuperado el 17 mayo de 2023, de

[https://observatorio.desarrolloeconomico.gov.co/sites/default/files/files\\_articles/bol-etin\\_usaquen.pdf](https://observatorio.desarrolloeconomico.gov.co/sites/default/files/files_articles/bol-etin_usaquen.pdf)

*Reportes—Mapa del retail en Colombia 2022—Resultados EL RANKING DE LAS 400*

*EMPRESAS MAS IMPORTANTES DEL RETAIL.* (s. f.). Recuperado 14 de mayo de

2023, de [https://www-emis-com.ez.urosario.edu.co/php/search/docpdf?doc\\_id=760901841](https://www-emis-com.ez.urosario.edu.co/php/search/docpdf?doc_id=760901841).

*Reportes—Health and Beauty Specialists in Colombia*. (s. f.). Recuperado 14 de mayo de 2023, de <https://www-emis-com.ez.urosario.edu.co/php/search/docpdf>.

Tsiptsis, Konstantinos K., y Antonios Chorianopoulos. (2011). *Data Mining Techniques in CRM: Inside Customer Segmentation*. John Wiley & Sons.

*What Is a Buyer Persona? Basics and Examples for 2023 | Upwork*. (s. f.). Recuperado 27 de mayo de 2023, de <https://www.upwork.com/resources/buyer-personas>.

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1), 3-14.  
<https://doi.org/10.1111/j.2041-210X.2009.00001.x>.