



Instituto Tecnológico y de Estudios Superiores de Monterrey

Escuela de Ingeniería y Ciencias

Ingeniería en Ciencias de Datos y Matemáticas

Inteligencia artificial avanzada para la ciencia de datos II

Perfilamiento de clientes y predicción de ventas - Análisis inicial

Arca Continental.

Equipo 6

Ricardo Salinas - A01284657

Oskar Gamboa - A01173648

Marcelo De Luna - A00832239

Erika Martínez - A01028621

Oscar Gutiérrez - A00832992

Profesores:

Edgar Gonzalez Fernandez

Mauricio González Soto

Monterrey, Nuevo León

02 de Octubre del 2024

Índice

1. Planteamiento del Problema	2
2. Hipótesis	2
2.1. Hipótesis: Meses con mayor venta	2
2.2. Hipótesis: Meses con menor venta	2
2.3. Hipótesis: Compra de productos no adquiridos previamente por los clientes	2
2.4. Hipótesis: Productos de lanzamiento	3
2.5. Hipótesis: 20 % de los productos más recientemente vendidos	3
2.6. Hipótesis: Cantidad y tipos de productos	3
2.7. Hipótesis: Cantidad y tipos de clientes	3
3. Métodos usados	4
3.1. Método: Meses con mayor y menor venta	4
3.2. Método: Cálculo de compra de un producto antes no comprado por el cliente	5
3.3. Método: Productos de lanzamiento	5
3.4. Método: Productos Recientemente Vendidos	6
3.5. Método: Cantidad y tipos de productos	6
3.6. Método: Cantidad y tipos de clientes	6
4. Conclusiones	7

1. Planteamiento del Problema

Arca Continental es la segunda embotelladora más grande dentro de Latinoamérica, esta se encarga de distribuir diferentes productos a distintos puntos de distribución. La empresa proporciona información sobre los productos vendidos, incluyendo el producto, el cliente, la fecha y la cantidad, además de información sobre los clientes y la descripción de los productos.

A partir de un análisis de los datos proporcionados, se pretende encontrar qué clientes serían adecuados para incorporar un producto nuevo en su distribución, por lo que es necesario realizar una segmentación de estos clientes y evaluar cómo se han comportado con productos de lanzamiento similares al que se desea incorporar, de esta manera se puede realizar una estrategia efectiva para maximizar la probabilidad de éxito de dicho producto.

2. Hipótesis

2.1. Hipótesis: Meses con mayor venta

H_0 : Los meses de mayor venta corresponden a las temporadas altas de consumo, como diciembre y verano, debido a las festividades y vacaciones.

H_1 : No existe una variación significativa en las ventas de un mes a otro.

2.2. Hipótesis: Meses con menor venta

H_0 : Los meses con menor venta se concentran en los periodos posteriores a las temporadas altas, como enero y febrero, debido a la disminución en el gasto del consumidor.

H_1 : Las ventas son consistentes durante todo el año sin una disminución significativa

2.3. Hipótesis: Compra de productos no adquiridos previamente por los clientes

H_0 : Un cliente realiza una compra de un producto nuevo al menos una vez cada cuatro meses, impulsado por ofertas y campañas de marketing.

H_1 : Los clientes tienden a comprar repetidamente los mismos productos, sin adquirir productos nuevos.

2.4. Hipótesis: Productos de lanzamiento

H_0 : Todos los productos han sido vendidos en años anteriores.

H_1 : Existen productos que no se han vendido en años anteriores.

2.5. Hipótesis: 20 % de los productos más recientemente vendidos

H_0 : El 20 % de los productos más recientemente vendidos corresponde mayoritariamente a productos de temporada o promociones especiales.

H_1 : No existe relación entre el 20 % de los productos más recientemente vendidos y su estacionalidad o promociones.

2.6. Hipótesis: Cantidad y tipos de productos

H_0 : La base de datos contiene al menos 5 tipos diferentes de productos caracterizados por sus características y segmento de mercado

H_1 : La diversidad de tipos de productos en la base de datos es limitada, no existen más de 3 categorías.

2.7. Hipótesis: Cantidad y tipos de clientes

H_0 : El número adecuado para organizar los clientes es de 13 categorías que definen el tipo de negocio que es.

H_1 : El número es menor que 13 categorías con un enfoque diferente.

3. Métodos usados

3.1. Método: Meses con mayor y menor venta

Para identificar los meses con mayor y menor venta se realizaron dos enfoques, el primero fue sumar las unidades vendidas (uni_box) de cada mes por año. Para esto primero se sumaron las unidades vendidas de cada producto por mes y año, agrupando las ventas por el mes correspondiente y sumando las unidades de cada producto.. Posteriormente se graficaron los resultados para poder visualizar la tendencia de las ventas a lo largo del tiempo, obteniendo la siguiente gráfica como resultado, en donde podemos apreciar que el mes con más ventas a través de los años (2019 - 2022) fue el mes de mayo del 2020.

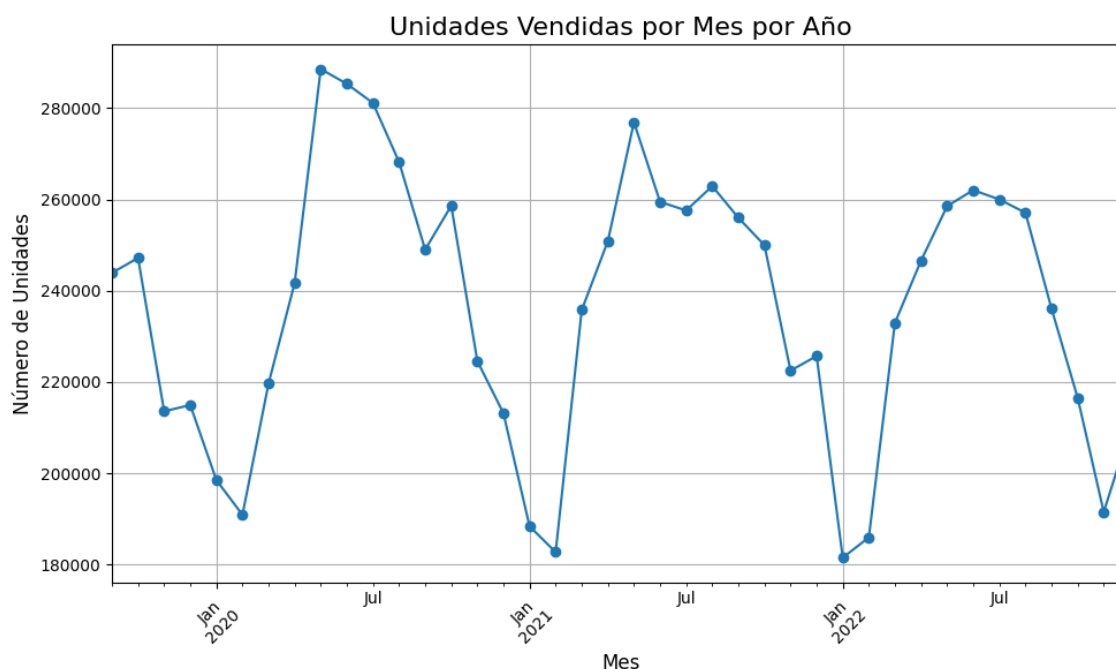


Figura 1: Unidades Vendidas por Mes por Año

Para el segundo enfoque se sumaron las unidades de ventas totales para cada mes sin tener en cuenta el año, de esta forma buscamos identificar que meses del año en general tienden a tener mayor o menor venta acumulada. Para esto se agruparon las ventas únicamente por el número de mes y se sumaron las unidades correspondientes. En esta segunda gráfica podemos apreciar que Septiembre se destaca como el mes con más ventas, seguido de Octubre y que en general los meses con menores ventas son los primeros del año.

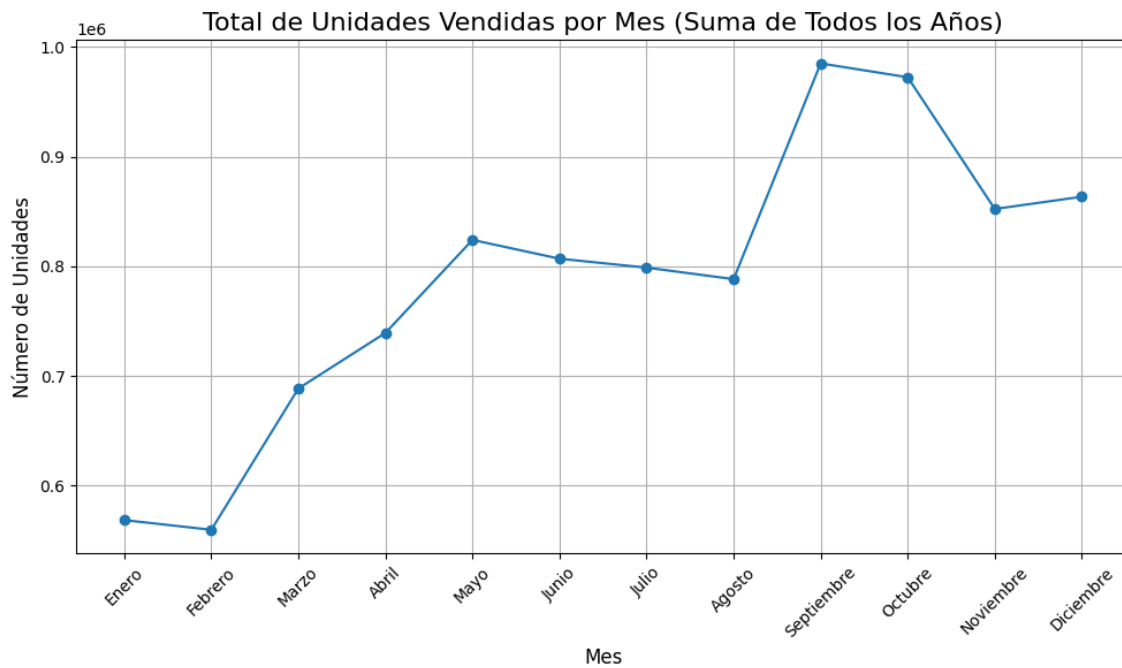


Figura 2: Tota de Unidades Vendidas por Mes

3.2. Método: Cálculo de compra de un producto antes no comprado por el cliente

Primero, se tiene que calcular qué se considera como un producto nuevo o un producto que no se compraba antes. Para esto, se pueden utilizar métodos diferentes. Uno es, por mes, calcular como producto nuevo a un producto que no se compró el mes anterior. Otro método es que un producto se considera nuevo para el cliente si no lo compraba durante un periodo de 4 - 6 meses.

También los resultados de estos métodos pueden cambiar dependiendo de si un producto se calculó por la unidad en la que se compró o simplemente contarlo como 1 si es que se compró un producto. Con estos diferentes cálculos se pueden analizar más cosas para el procedimiento que se requiera hacer.

3.3. Método: Productos de lanzamiento

La definición que se utilizó para producto de lanzamiento es la siguientes: Un producto de lanzamiento es aquel que no ha sido vendido en los años anteriores. Para encontrar los productos que cumplan con esta definición debe filtrarse el dataframe y crearse otros dos a partir de ese, uno con el año que se está evaluando y otro con los años anteriores, luego comprobar qué productos están presentes en el año que

se evalúa que no estén presentes en el otro dataframe. Al final, todos los códigos de los productos que cumplen con esta definición se agregan a una lista para tener fácil acceso a ellos.

Una cosa que cabe mencionar, es que en el dataset donde se dan las descripciones de los productos no están todos los productos de lanzamiento, por lo que no se tiene información de una gran proporción de ellos.

3.4. Método: Productos Recientemente Vendidos

Creamos una variable df la cual contenga las columnas “calmonth” y “material” para poder encontrar los artículos comprados más recientemente, luego con el uso de la variable “calmonth” creamos dos nuevas variables llamadas “fecha ultima venta” y “fecha primera venta”, para poder ver las fechas más recientes de compras y después ordenarlas por fecha de primera compra. Luego con el uso del comando sort values para ordenamos los productos por venta más recientes, luego escogimos solo el 20 % de los productos, finalmente haciendo uso del comando sort values otra vez, usamos la variable “fecha primera venta” y ordenamos los productos dependiendo de su primera venta.

3.5. Método: Cantidad y tipos de productos

Primero agrupamos los productos según las categorías que ya teníamos definidas, como sabor, tamaño y tipo. Luego, calculamos el promedio del indicador en cada grupo, y aplicamos una prueba ANOVA para identificar si había diferencias significativas entre ellos. Los resultados mostraron que ciertos tipos de productos, como los de mayor tamaño o de sabores más populares, tenían un mejor desempeño en el indicador, lo que sugiere que la clasificación de productos sí puede tener un impacto importante, permitiendo enfocar la estrategia comercial en las categorías más rentables.

3.6. Método: Cantidad y tipos de clientes

Para poder dividir los clientes encontramos dos soluciones que dependiendo de los siguientes avances vamos a usar. La primera es usando la columna de categorías que ya incluye la base de datos, esto nos permite dividir los datos en el tipo de cliente que son (por ejemplo abarrotes o farmacias), al final esta manera resulta en 13 categorías, siendo la más popular “abarrotes” con 586 clientes.

Para la segunda opción, usamos las variables que indican la cantidad de parques, supermercados,

hospitales, escuelas y gimnasios que se encuentran cerca de cada cliente. Tomamos la categoría que tuviera la mayor cantidad y agregamos al cliente a un dataframe con el nombre de la categoría. El mayor problema es que no siempre hay una categoría con una mayor cantidad, por lo que hay clientes que están en múltiples categorías. Al final obtuvimos 9 categorías que nos dan una mejor idea de que tipo de clientes concurren en cada tienda.

4. Conclusiones

Se pueden notar ciertas tendencias en las ventas de acuerdo con el análisis de en qué meses se vende más o se vende menos. Además, se realizó un análisis rápido sobre los productos de lanzamiento, para continuar con un análisis más profundo se podría probar utilizando definiciones distintas para producto de lanzamiento. Otro punto crucial para continuar con el análisis es verificar con el socio formador la información sobre los productos faltantes, puesto que el no contar con la descripción de estos productos puede afectar significativamente la calidad de los resultados.