

UNIVERSIDAD PROVINCIAL DEL SUDOESTE

# ELEMENTOS DE APRENDIZAJE DE MÁQUINA Y BIG DATA

Integrantes: Grupo 20

- Lucia ALONSO
- Pastor PAREDES
- Viviana SUAREZ
- Heber ALONSO

<https://github.com/lu8dcf/UPSO-BigData->

## LABORATORIO

### Análisis de Datos de Ventas Mensuales

Informe del análisis de un conjunto de datos que contiene información sobre las ventas mensuales de productos en línea. Utilizaremos herramientas de análisis de datos y visualización sobre el desempeño de ventas.

---

# ÍNDICE

1	Introducción	1
2	Comportamiento de las ventas en los distintos meses	2
2.1	¿Cuál es el ingreso total generado por mes?	2
2.2	¿Cómo variaron las ventas a lo largo de los diferentes meses?	2
2.3	¿Hubo algún mes que se destacó en términos de ventas?	3
3	Optimización de la publicidad y patrón de ventas por hora	5
3.1	¿Cuál es el momento ideal del día para presentar la publicidad y aumentar la probabilidad de compra?	5
3.2	¿Cómo cambian los patrones de ventas por hora a lo largo del año?	5
3.3	¿Hay modificaciones en los patrones de ventas durante las horas de mayor actividad en los distintos meses?	5
3.4	¿En qué ciudades se han registrado las mayores ventas?	6
4	Distribución de ventas por ubicación	7
4.1	¿Cómo se comparan las ventas en diferentes estados o regiones?	7
4.2	¿Existe variación en las ventas por estado a lo largo de los meses?	8
5	Análisis del producto más vendido	9
5.1	¿Cuál es el producto más vendido en general y en cada mes?	9
5.2	¿Qué factores crees que han contribuido al éxito de ese producto en particular?	9
6	Tendencias ventas	10
6.1	¿Existe alguna tendencia o patrón en las ventas que se repita a lo largo de los días en los diferentes meses?	10
6.2	¿Cómo varían las ventas a lo largo de los diferentes días de la semana?	12
6.3	¿Se observa alguna diferencia significativa en las ventas entre los días laborables y los fines de semana?	14
7	Impacto de eventos especiales en las ventas	16
7.1	¿Se ha observado algún aumento o disminución significativa en las ventas en días cercanos a eventos especiales, como días festivos?	16
7.2	¿Qué eventos específicos han tenido un impacto notable en el comportamiento de las ventas y cómo se manifestó ese impacto?	17
8	Pregunta adicional	19
8.1	Análisis de productos cruzados	19
9	Conclusión	21

## 1 INTRODUCCIÓN

Se realizara el análisis de los datos de ventas on-line de accesorios electrónicos (notebooks, auriculares, USB, entre otros) en el lapso de un año es decir de enero a diciembre, de distintos estados de EE.UU. con el objetivo de ver las fluctuaciones que tienen las ventas tanto a lo largo del año como en los distintos horarios, para recabar información sobre en qué momento del día es mejor para publicitar, en que mes/es se tienen más ventas y de que producto donde tal vez sea necesario tener más stock, en que mes hay menor ventas donde seria ventajoso ofrecer ofertas como descuentos o promociones para incentivar las compras y mejorar los ingresos.

Para dicho análisis se tendrá en cuenta:

- \* la cantidad pedida,
- \* las horas y las fechas de los pedidos,
- \* los productos,
- \* los precios,
- \* los estados o ciudades de donde se extrajeron los datos.

El trabajo está realizado en VC con el formato de COLAB.

<https://github.com/lu8dcf/UPSO-BigData-G20.git>

## 2 COMPORTAMIENTO DE LAS VENTAS EN LOS DISTINTOS MESES

### 2.1 ¿Cuál es el ingreso total generado por mes?

En el gráfico 1 están representadas la cantidad total de ítems vendidos por mes, en el grafico 2 son los montos totales que ingreso por la venta de todos los productos por mes.

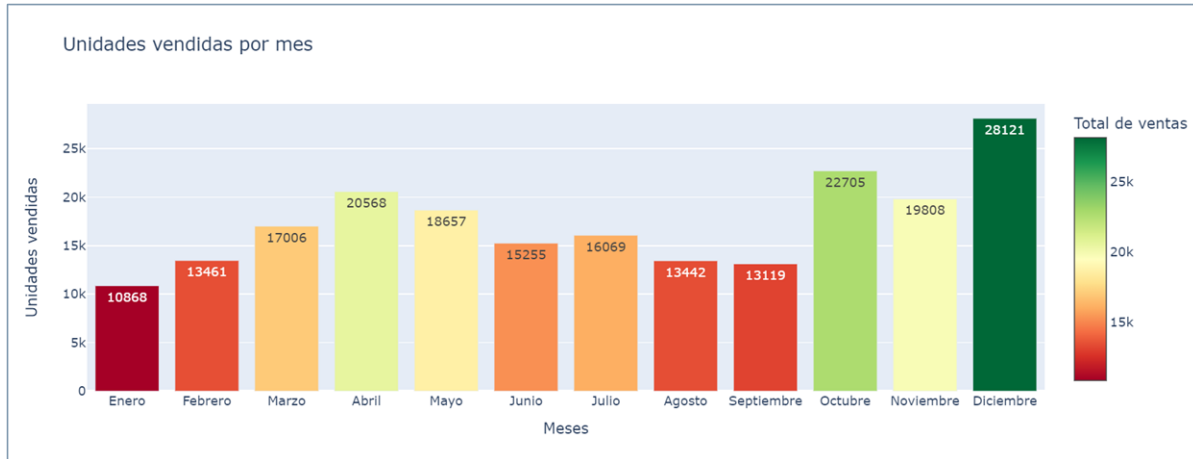


Gráfico 1

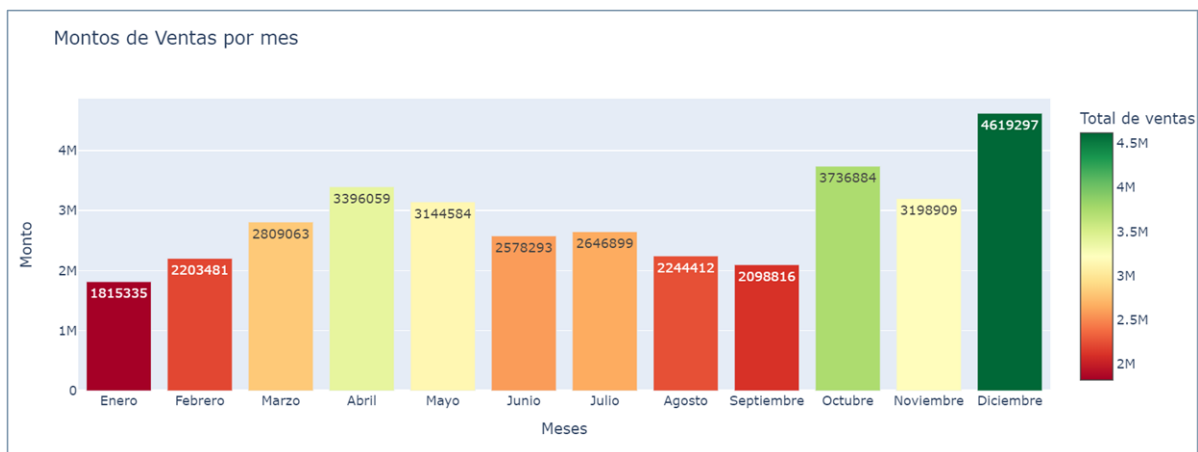


Gráfico 2

### 2.2 ¿Cómo variaron las ventas a lo largo de los diferentes meses?

Para determinar la variación mensual se utilizaron 2 métodos:

**Por Referencia:** Este método es una función `.diff()` de la librería Pandas que devuelve la variación del mes actual en referencia al mes anterior, esto determina si el siguiente mes fue positivo o no en ventas. El mes de enero figura en 0 porque es la referencia inicial. (grafico 3)

# COMPORTAMIENTO DE LAS VENTAS EN LOS DISTINTOS MESES



Gráfico 3

**Por Promedio:** Se calculó el promedio anual de ventas y se restó a las ventas de cada mes, de esta manera se obtiene la diferencia de cada mes con respecto al promedio anual de 2.874.336\$.

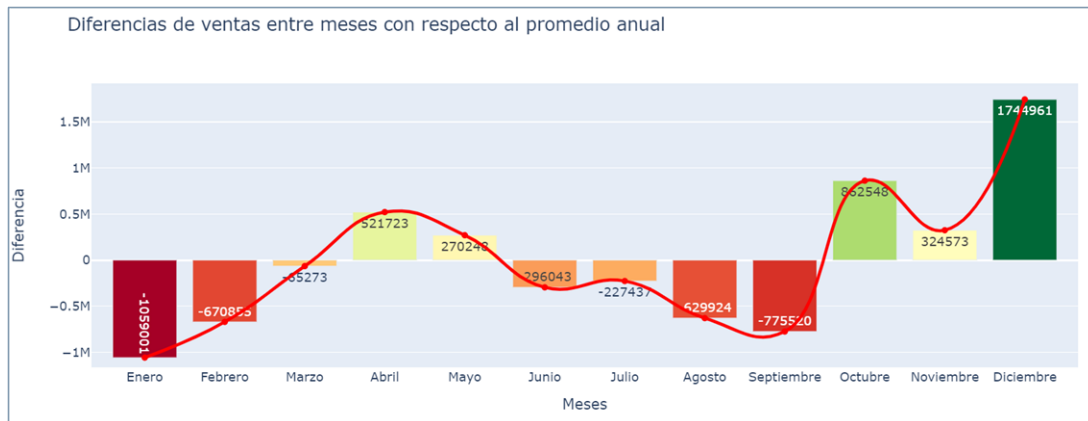


Gráfico 4

Ambos muestran una curva similar en la variación de ventas mes a mes, cabe destacar que el cálculo sobre la cantidad vendida y el monto son similares en la relación gráfica.

## 2.3 ¿Hubo algún mes que se destacó en términos de ventas?

En el gráfico 5 se ve representado en porcentuales la cantidad de ventas mensuales con respecto al anual, determinando claramente que diciembre es el de mayor venta con un 13,4% del anual, de igual manera se puede apreciar cuantitativamente en los gráficos 1 y 2.

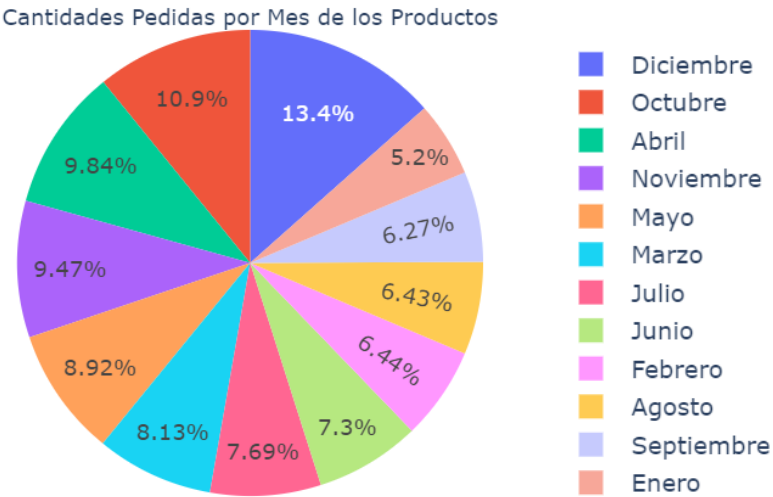


Gráfico 5

### 3 OPTIMIZACIÓN DE LA PUBLICIDAD Y PATRÓN DE VENTAS POR HORA

#### 3.1 ¿Cuál es el momento ideal del día para presentar la publicidad y aumentar la probabilidad de compra?

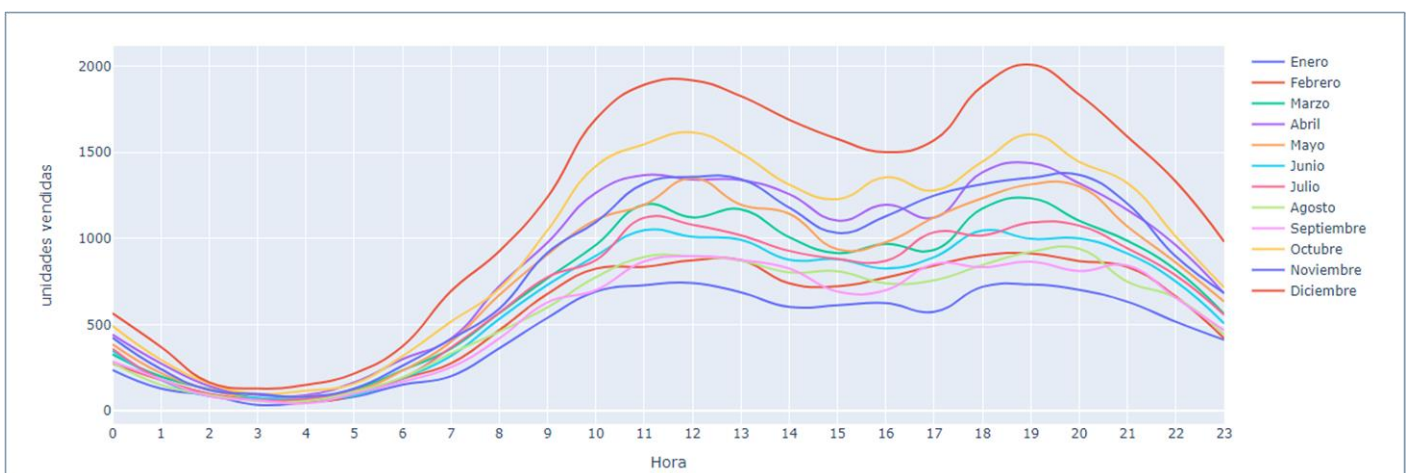
A grandes rasgos en el gráfico 6 se puede observar que los horarios de mayores ventas son entre las 11 y 12 horas am y entre las 19 y 20 horas pm, lo que indica que los mejores horarios para publicitar son entre las 10:30 y 12 horas AM y las 1830 y 20 horas PM.

#### 3.2 ¿Cómo cambian los patrones de ventas por hora a lo largo del año?

No se observa un patrón relacionado a las estaciones del año ya que no tiene relación alguna en los horarios de mayor venta en los distintos meses, a saber, Enero 12hs, Febrero 19hs, Marzo 19hs, Abril 19hs, Mayo 12hs, Junio 11hs, Julio 11hs, Agosto 20hs, Septiembre 12hs, Octubre 12hs, Noviembre 20hs, Diciembre 19hs. Si se podría relacionar el patrón con el horario de descanso como ser el *almuerzo* y *la cena*

#### 3.3 ¿Hay modificaciones en los patrones de ventas durante las horas de mayor actividad en los distintos meses?

Si bien a lo largo del año se mantiene un patrón similar con respecto a las horas donde se produce mayor actividad, pero si tenemos en cuenta las unidades vendidas se observa que en el mes de octubre se produce un salto con respecto al resto del año y que el mes de diciembre tiene un pico aun mayor, este mes es el que más varia con respecto al resto.



Promedios de ventas por hora en los diferentes Meses - Gráfico 6

## 3.4 ¿En qué ciudades se han registrado las mayores ventas?

En el gráfico 7 se puede apreciar que las ciudades con mayor venta son San Francisco 50.239 y lo sigue Los Ángeles 33.289 y New York City 27.932. En el, grafico 8 está representado el total de ventas de las ciudades en porcentajes del total de unidades vendidas.

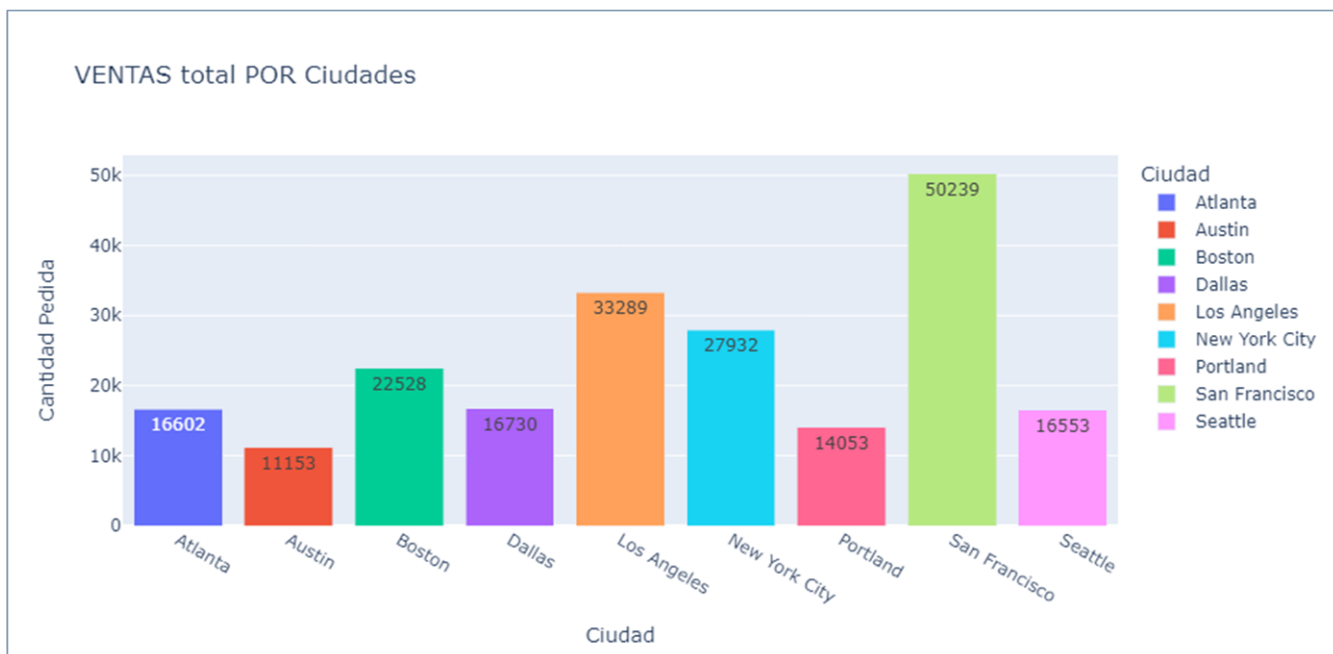


Gráfico 7

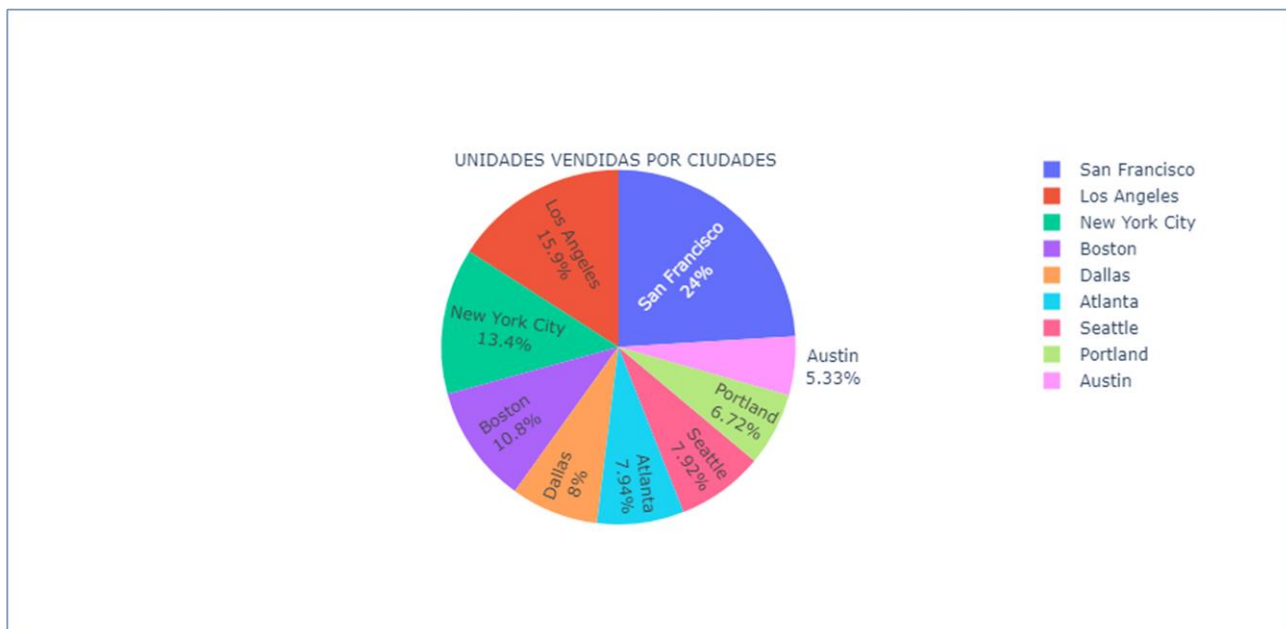


Gráfico 8



## 4 DISTRIBUCIÓN DE VENTAS POR UBICACIÓN

### 4.1 ¿Cómo se comparan las ventas en diferentes estados o regiones?

En el grafico 9 se represento en un mapa los estados de mayor y menor venta con una escala que va de rojo a verde en la cual muestra que el estado de CA sobresale. Y ademas con otro grafico de barras, el 10, el total de ventas por cada estado pero apilados por cada mes de venta. Y en el grafico 11 el total de ventas por estados. En donde se muestra los números del estado de CA de 83.528

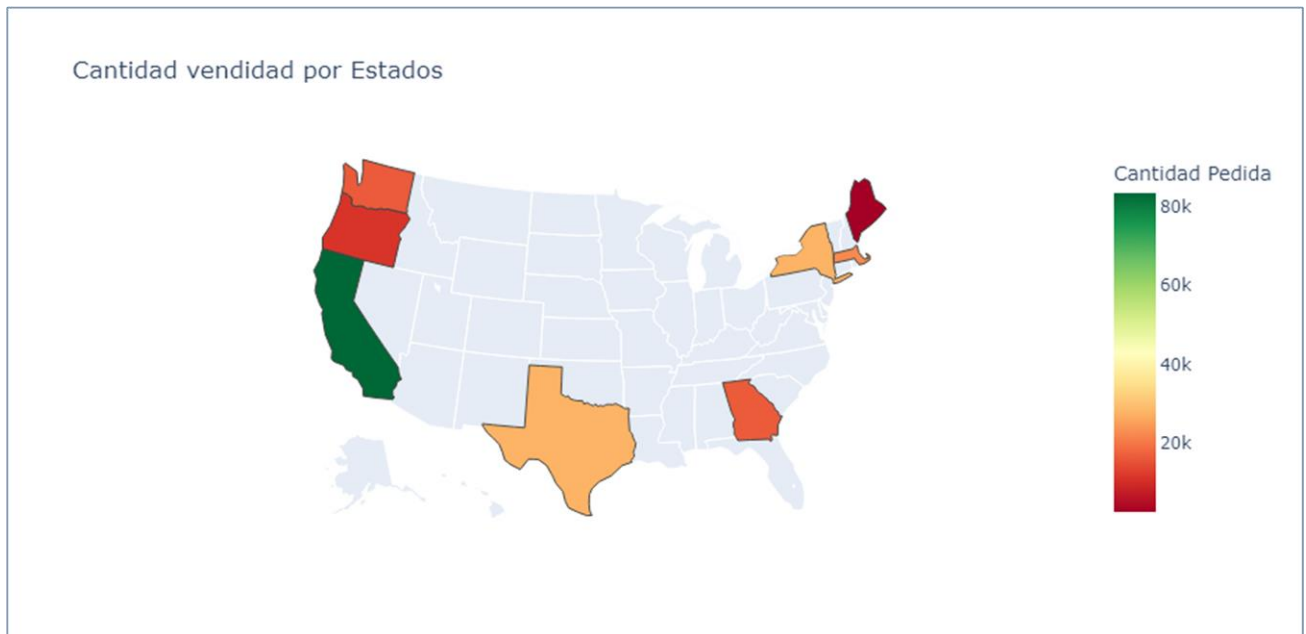


Gráfico 9

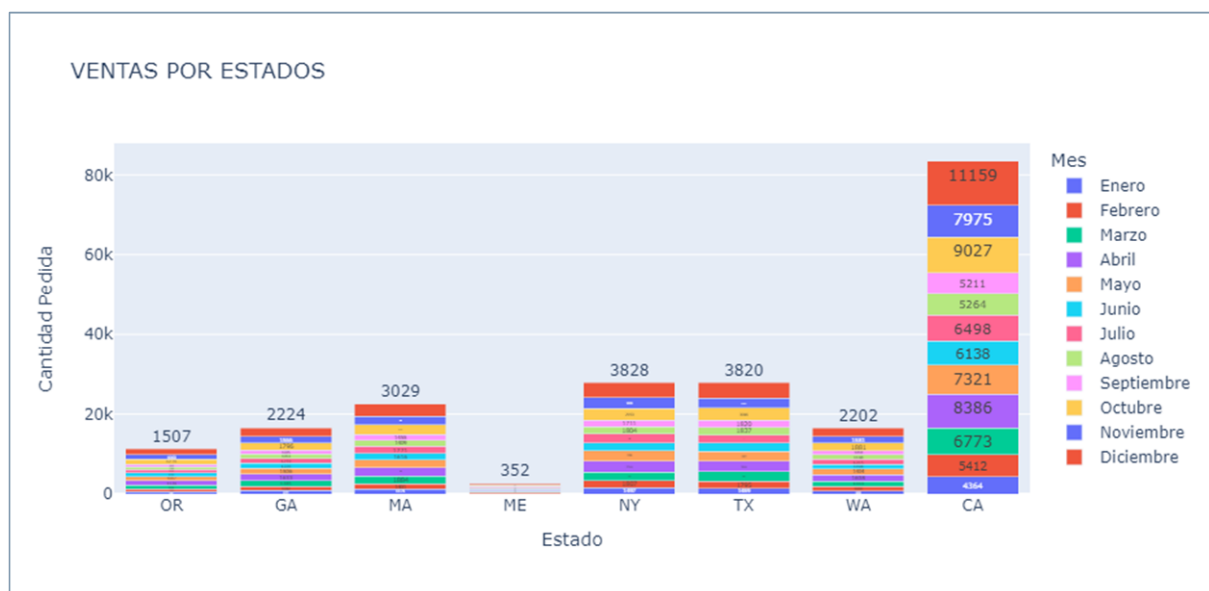


Gráfico 10

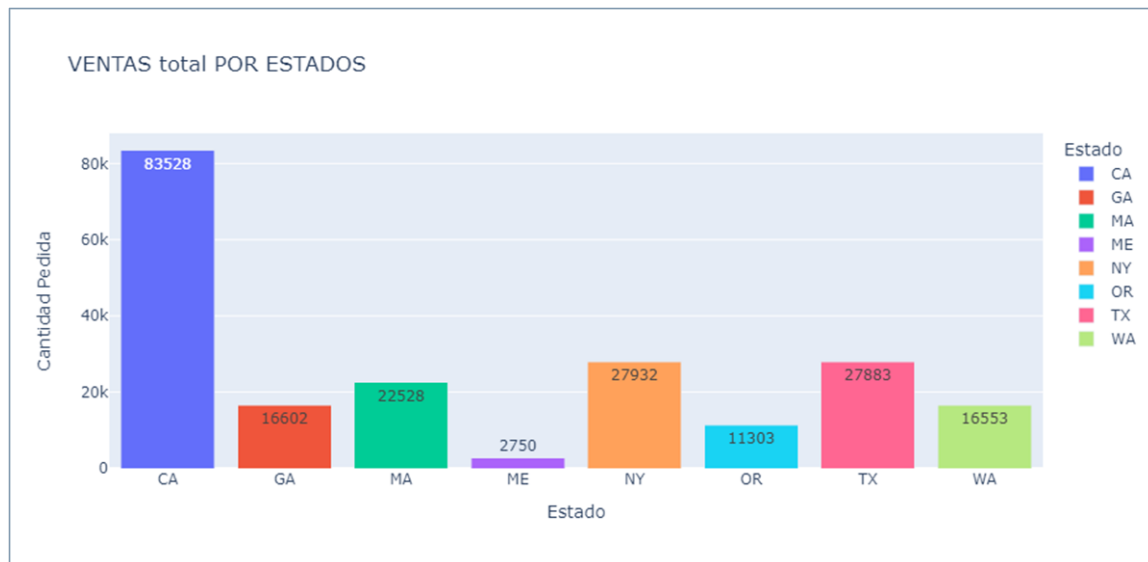


Gráfico 11

## 4.2 ¿Existe variación en las ventas por estado a lo largo de los meses?

Se puede apreciar en el grafico 12 la variacion del estado de CA al estado de WA a lo largo de los meses resaltado una linea de color por cada mes.

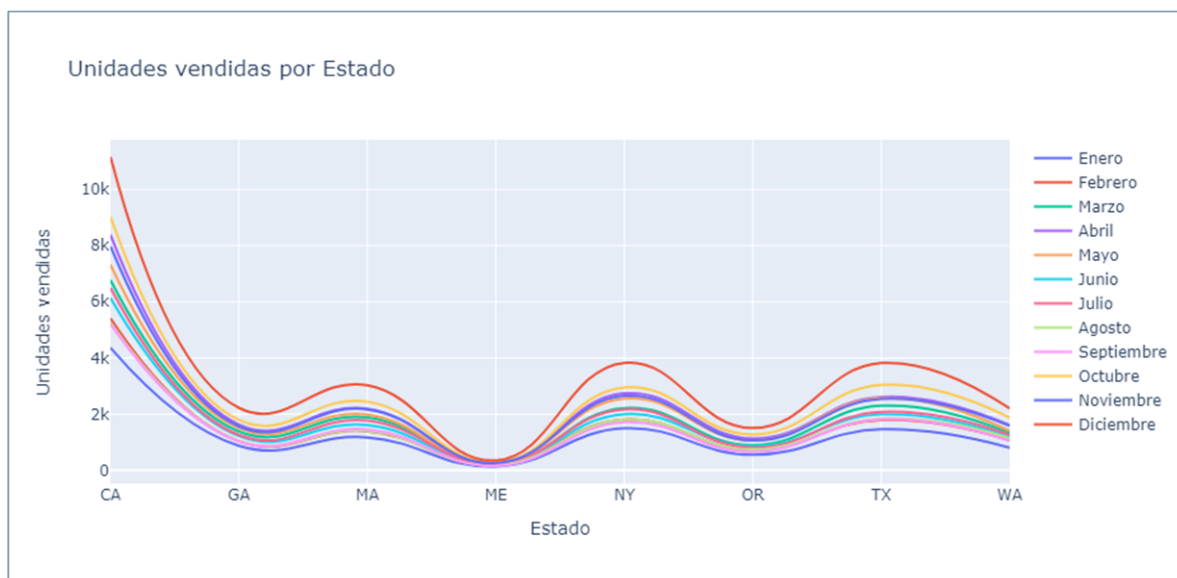


Gráfico 12

## 5 ANÁLISIS DEL PRODUCTO MÁS VENDIDO

### 5.1 ¿Cuál es el producto más vendido en general y en cada mes?

En el siguiente grafico (grafico 13), se ve la cantidad de los productos que han sido más vendidos cada mes, en todos los casos ha sido el mismo producto “AAA Batteries (4-pack)”, ergo la venta anual es del mismo producto con un total de **31.017** ítems vendidos.

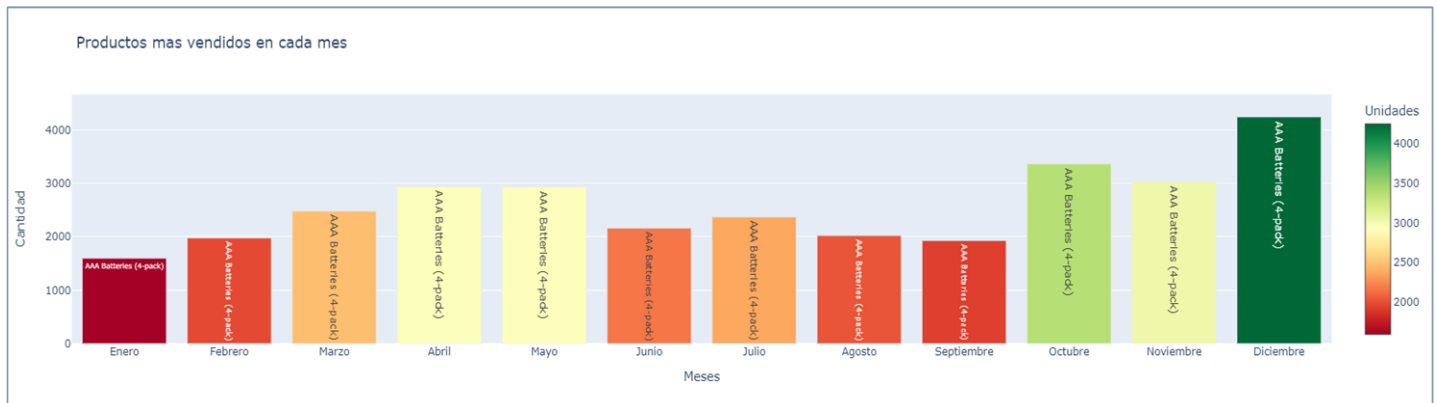


Gráfico 13

### 5.2 ¿Qué factores crees que han contribuido al éxito de ese producto en particular?

No hay nada representativo en el porqué de la mayor venta anual y en todos los meses del mismo producto, pero se pueden tomar las siguientes opiniones:

- a- Por ser un producto económico, es el más económico entre los demás ofrecidos y vendidos.
- b- Es un producto de desgaste, cuando se agota su carga deben ser desechadas y volver a adquirir, o sea el tiempo de vida es menor al de un monitor.
- c- Es producto de uso masivo, es común tener varios controles remotos o pequeños artefactos en el hogar que usen pares de pilas AAA, por ende la demanda es mayor.

## 6 TENDENCIAS VENTAS

### 6.1 ¿Existe alguna tendencia o patrón en las ventas que se repita a lo largo de los días en los diferentes meses?

En el siguiente gráfico se observa que la cantidad de ventas de los productos es muy heterogénea, la gran mayoría de los meses se ubican a partir de las 400 (cuatrocientas) ventas hasta las 800 (ochocientas) de las mismas. Los únicos meses que resaltan son Enero por ser el mes en que menos se vendió y Diciembre por ser el mes en que más unidades se vendió. En caso de Diciembre puede que por las fechas festivas como Navidad sea el causante del aumento en ventas.

Los puntos rojos que hay en cada línea de mes se debe a los picos de ventas que hubo en todo el mes. Respecto a los picos podemos decir que la mayor cantidad de ventas en el día se hicieron entre la tercera y última semana del mes

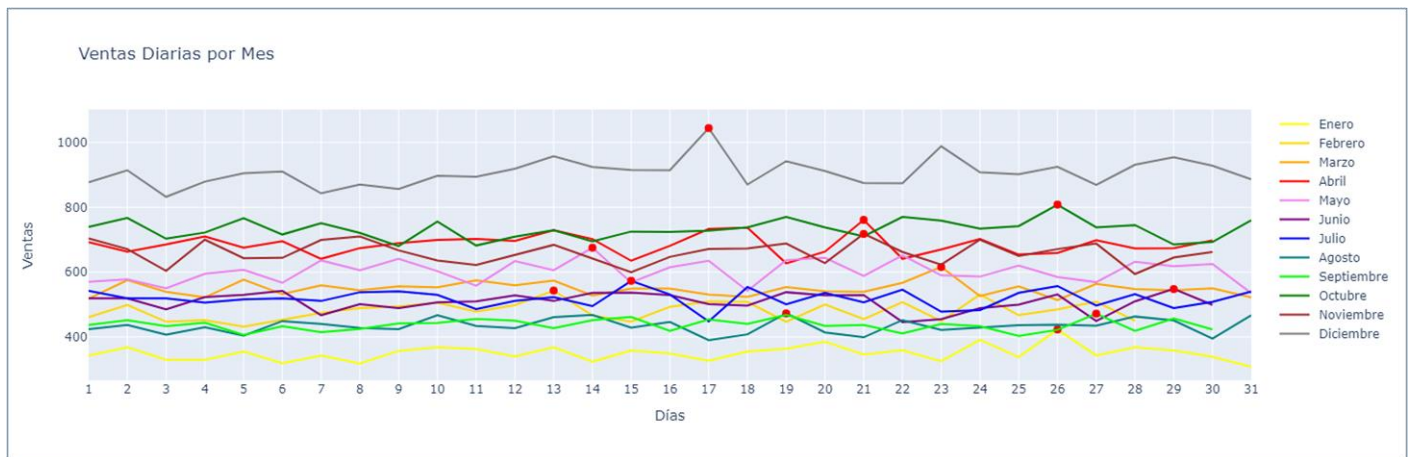


Gráfico 14

En el siguiente gráfico se optó por segmentar mejor el gráfico 14 y ver con mayor claridad. Se realizó un análisis de las ventas agrupadas por semana de cada mes a lo largo del año. De esta forma podemos afirmar que el patrón de ventas se encuentra al final de cada mes, la última semana de cada mes, en su mayoría, sobrepasan con creces a las semanas anteriores del mismo mes. Por ende, las ventas se realizan generalmente al final de cada mes, en la última semana.

## TENDENCIA VENTAS

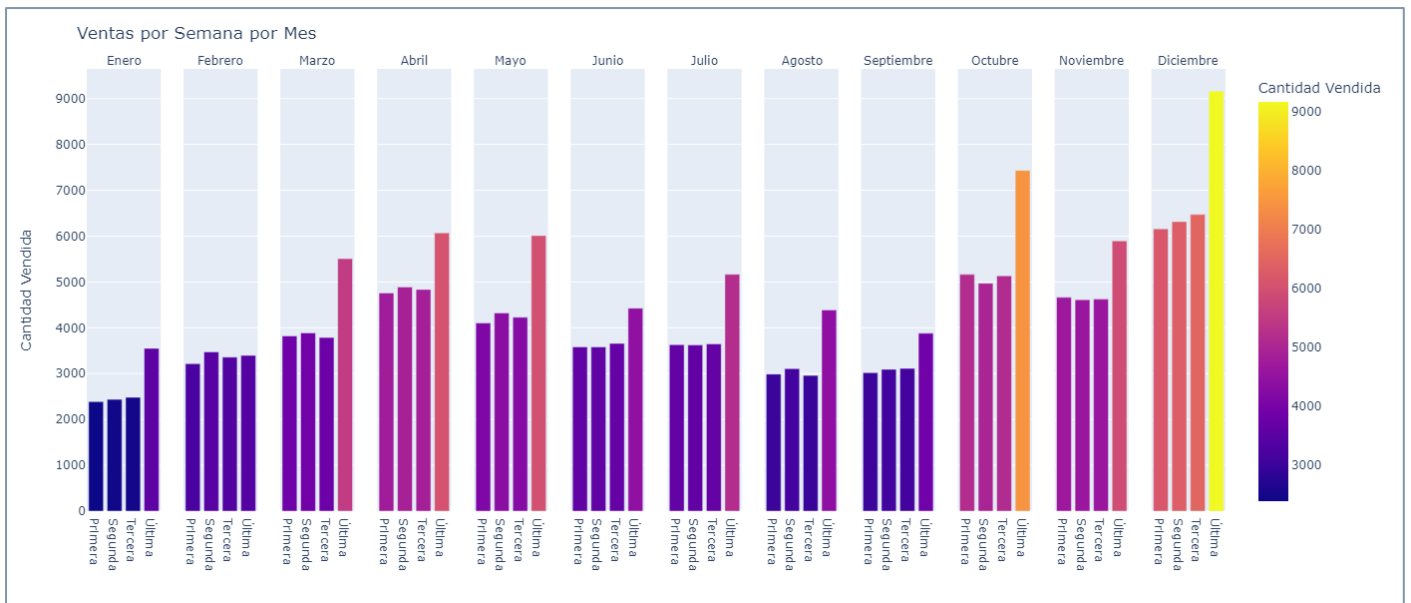


Gráfico 15

En el siguiente gráfico vemos una segmentación igual que en el gráfico 15 (quince), con la única diferencia que posee una disposición apilada, sin embargo, así, no se puede observar con claridad qué semana supera a cuál, pero si podemos saber cuáles son sus cantidades de ventas en cada semana.

Entonces observando cada mes, todos superan la última semana en comparación a las otras tres. Quién no cumple este patrón es el mes de Febrero que por menos de 40 (cuarenta) ventas no pudo superar a las demás semanas. Esto se puede deber a que posee menos días en el mes que los demás meses.

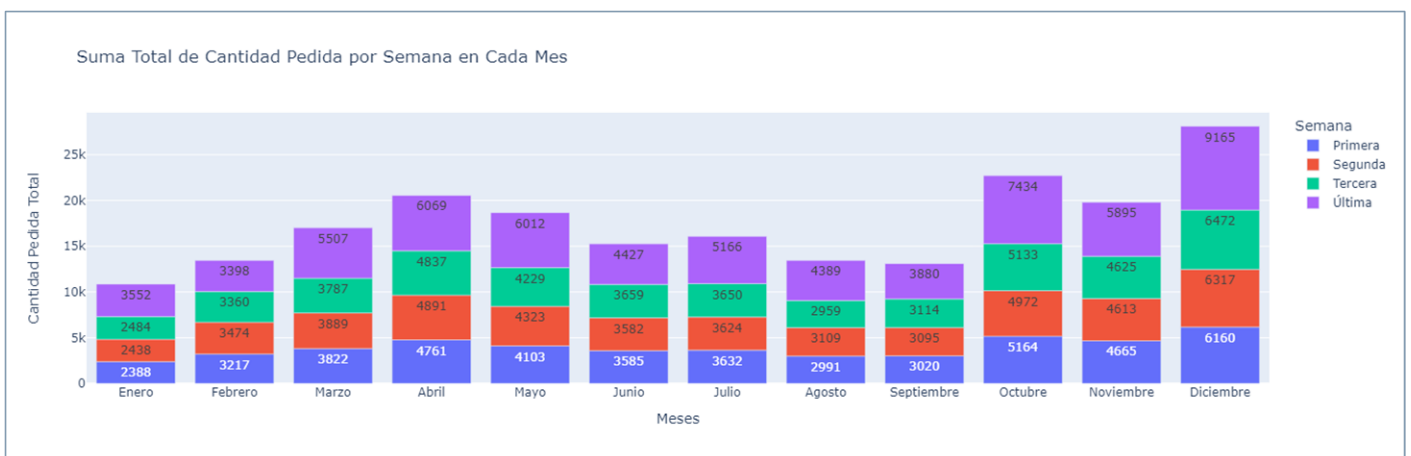


Gráfico 16

Con este último gráfico se cierra la pregunta inicial observando en un gráfico de torta las cuatro semanas y la cantidad que se venden en un porcentaje total para cada semana. La última semana es la que posee mayor porcentaje, por ende, el patrón a lo largo de los días en el mes es una creciente venta en la última semana y la segunda (siguiendo el gráfico 17) semana que es la que le sigue en porcentaje de la última semana.

Cantidades Pedidas por Semana en el año



Gráfico 17

## 6.2 ¿Cómo varían las ventas a lo largo de los diferentes días de la semana?

En el siguiente gráfico se puede ver las cantidades de ventas de los diferentes días de la semana de cada mes, cada día se diferencia con colores y se muestra la venta que hubo en cada día de la semana. En el gráfico 19 (diecinueve) se aprecia mejor.

Cantidades pedidas por día de la semana por mes durante todo el Año

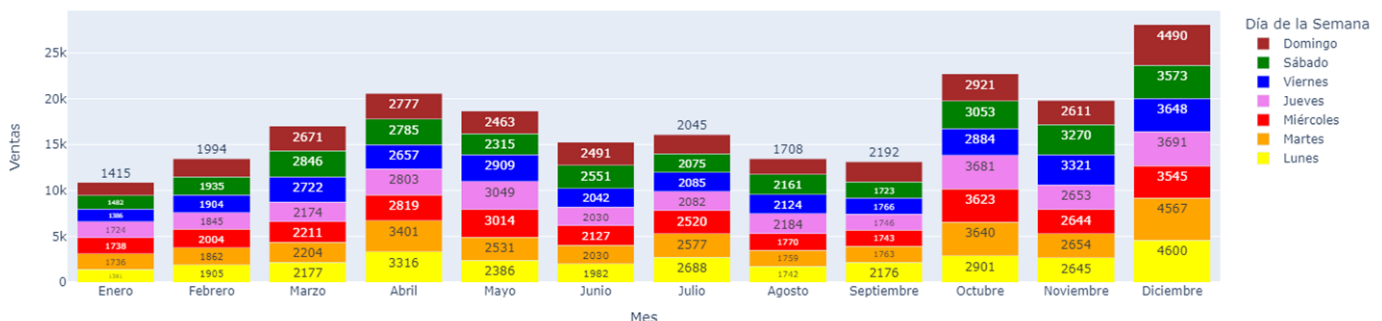


Gráfico 18

En el siguiente gráfico se aprecia mejor la diferencia de ventas en cada día, se puede observar que varían en cada mes la mayor cantidad de ventas, es decir, no hay un patrón claro sobre qué día en particular de la semana se vende más que en otros. A simple vista se observa que los meses de Enero, Febrero, Agosto y Septiembre se vendió más nivelado sin picos en algún día en particular. En Marzo, Junio y Noviembre poseen muestras de ventas altas en Viernes y fines de semana. Para los meses Abril, Mayo, Julio y Octubre se observan ventas en los primeros tres días de la semana. Y en el caso de Diciembre posee ventas en Domingo, Lunes y Martes.

# TENDENCIA VENTAS

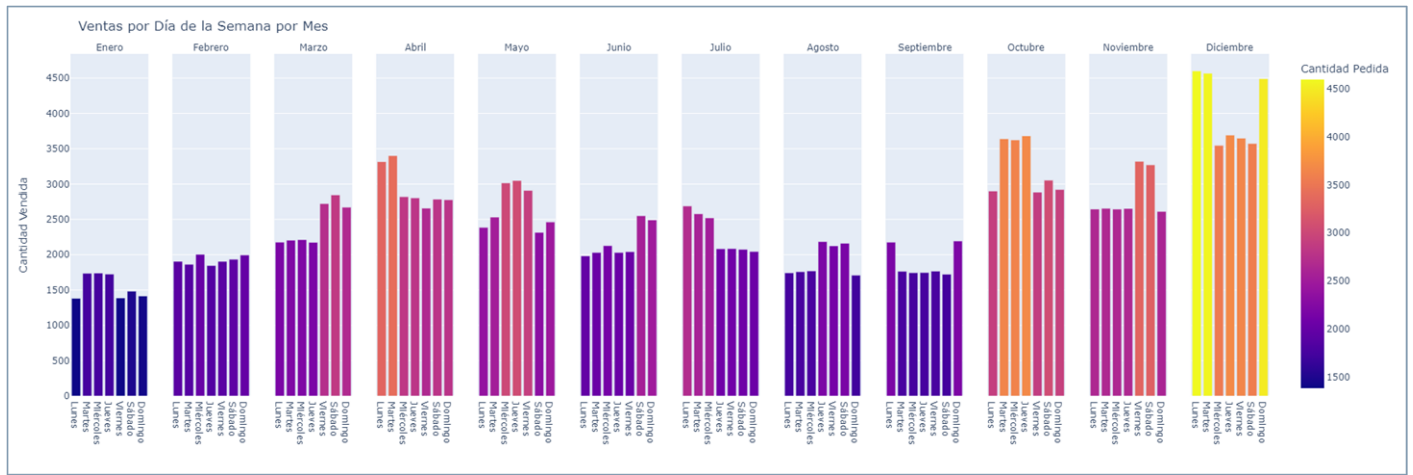


Gráfico 19

En este gráfico se puede observar todas las ventas de cada día de la semana en el año. Los días Martes y Lunes son aquellos que obtuvieron un mayor porcentaje. Y el día que menos vendió fue el Viernes.



Gráfico 20

En el siguiente gráfico se tomó como referencia las ventas totales del Lunes. Y ver gráficamente cómo se encuentran de ventas los siguientes días tomando en consideración el primer día de la semana.

La línea roja demuestra los picos de cada día de la semana en referencia al día Lunes. Así confirmamos que el Martes supera con creces las ventas del Lunes, en el caso del Miércoles tuvo muy pocas ventas. El Domingo en cambio al ser un día de fin de semana que le precede al Lunes se ven unas ventas con muy escasa diferencia.

Con este gráfico se afirma que como el gráfico 20 (veinte) las ventas son mayores en Lunes y Martes. Finalmente dado el gráfico en la mitad de la semana no se registran grandes ventas.

## TENDENCIA VENTAS

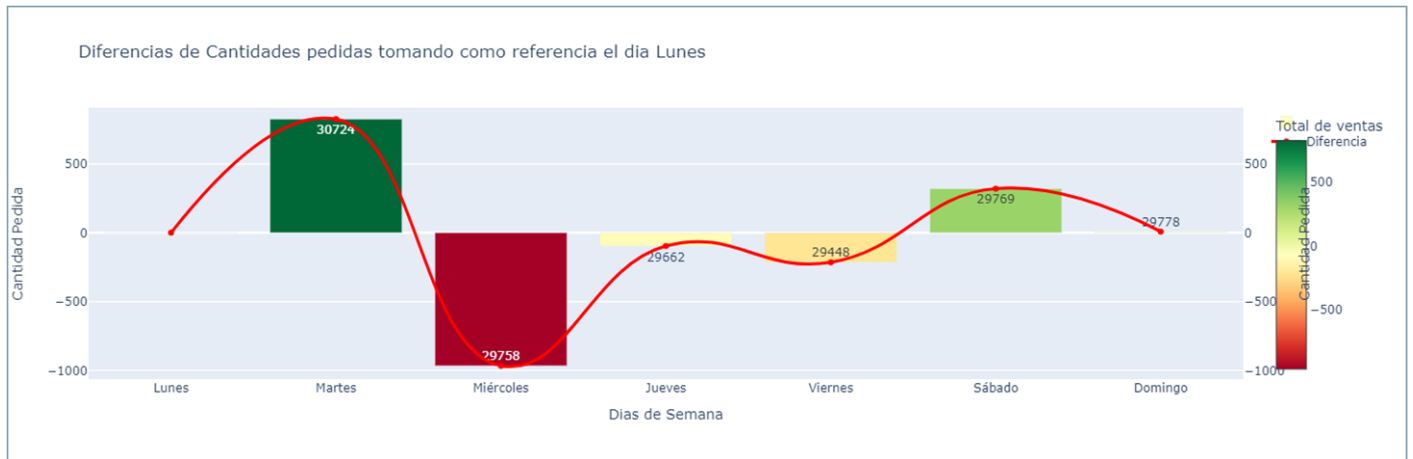


Gráfico 21

En el siguiente gráfico se observan de los distintos días de la semana las ventas en porcentajes en referencia del mínimo y del máximo que se obtuvo de los siete días. El máximo es el Martes y el mínimo es el Viernes. Dado el gráfico se contempla que el Jueves es un día de muy pocas ventas y que los días del fin de semana no varían entre ellos.

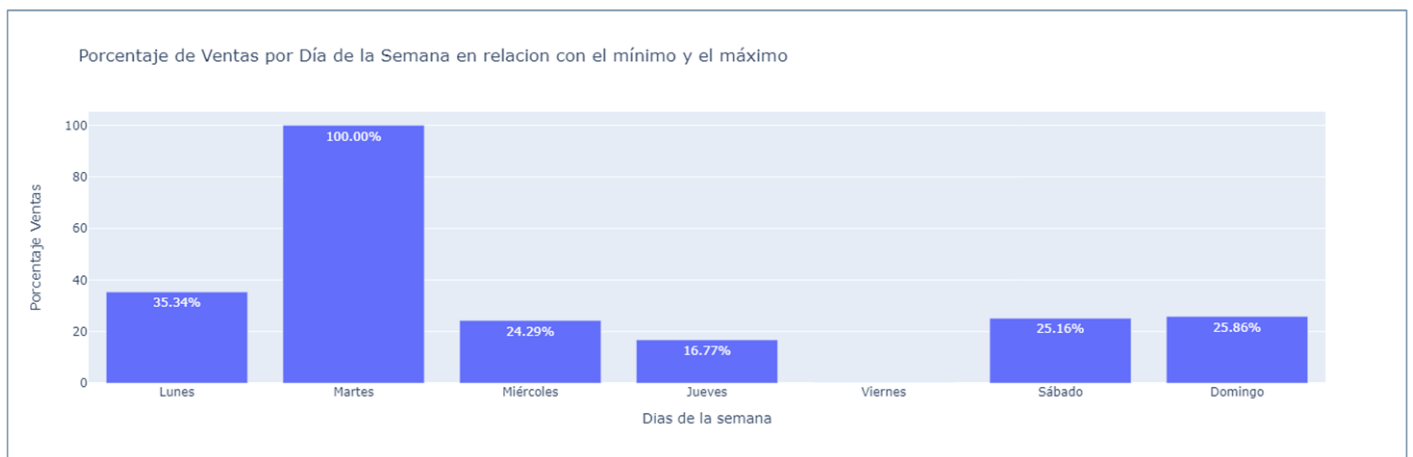


Gráfico 22

### 6.3 ¿Se observa alguna diferencia significativa en las ventas entre los días laborables y los fines de semana?

En los gráficos anteriores, dígame el 18 (dieciocho) y 19 (diecinueve) ya adelantaban los resultados acerca de los días de semana laborales y los no laborales. En el siguiente gráfico se puede observar que los días no laborales poseen pocas ventas a pesar de ser días que normalmente es cuando más se vende, por ser los momentos de óseo de la gran mayoría de las personas.

La línea punteada que se ve en el gráfico marca la media anual en base a la cantidad de pedidos mensuales. Por ende, aquellas barras que superan la línea punteada superan las ventas de la media con creces.



## TENDENCIA VENTAS

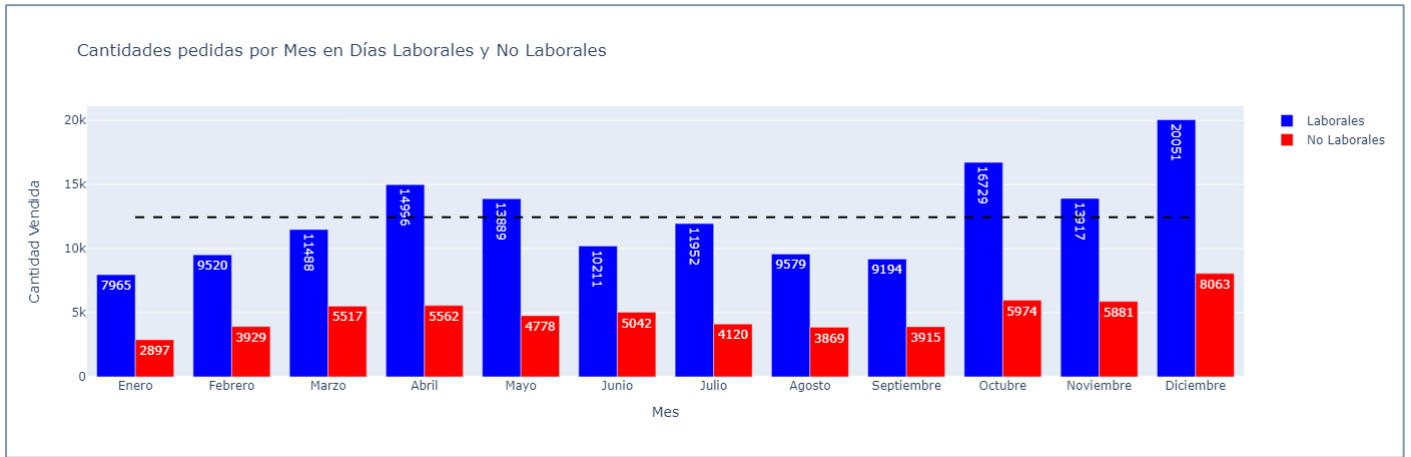


Gráfico 23

## 7 IMPACTO DE EVENTOS ESPECIALES EN LAS VENTAS

### 7.1 ¿Se ha observado algún aumento o disminución significativa en las ventas en días cercanos a eventos especiales, como días festivos?

En el siguiente gráfico de calor se muestra un nivel de color correspondiente a la cantidad total de ítems vendidos cada día. Dadas que las ventas son de los Estados Unidos, se indicaron los días festivos más significativos de la cultura de dicho país, en tal motivo se puede apreciar en el gráfico 24 de cantidades nada significativo por los días festivos, solo un aumento en el *día de la madre*, *día de la recordación* y *navidad*.

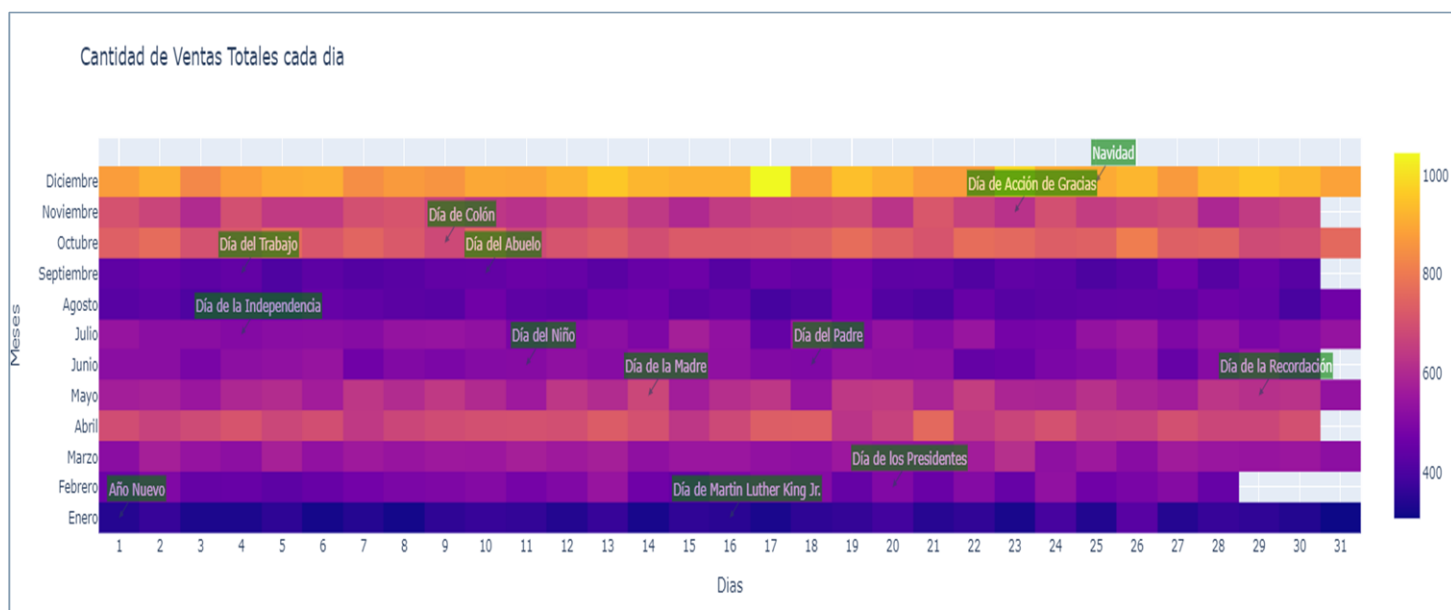


Gráfico 24

Sin embargo, en el gráfico que representa de la misma manera los montos vendidos en cada día se aprecia un aumento en el total de valor de venta, en el *Día del niño*, *Día del Padre*, *Día de la Madre*, *Día de la recordación*, *Día de la Independencia* y *Navidad*. Pero en el gráfico también se ven muchas zonas de aumento de ventas por el mes de Abril y Octubre en los cuales no hay días festivos cercanos.

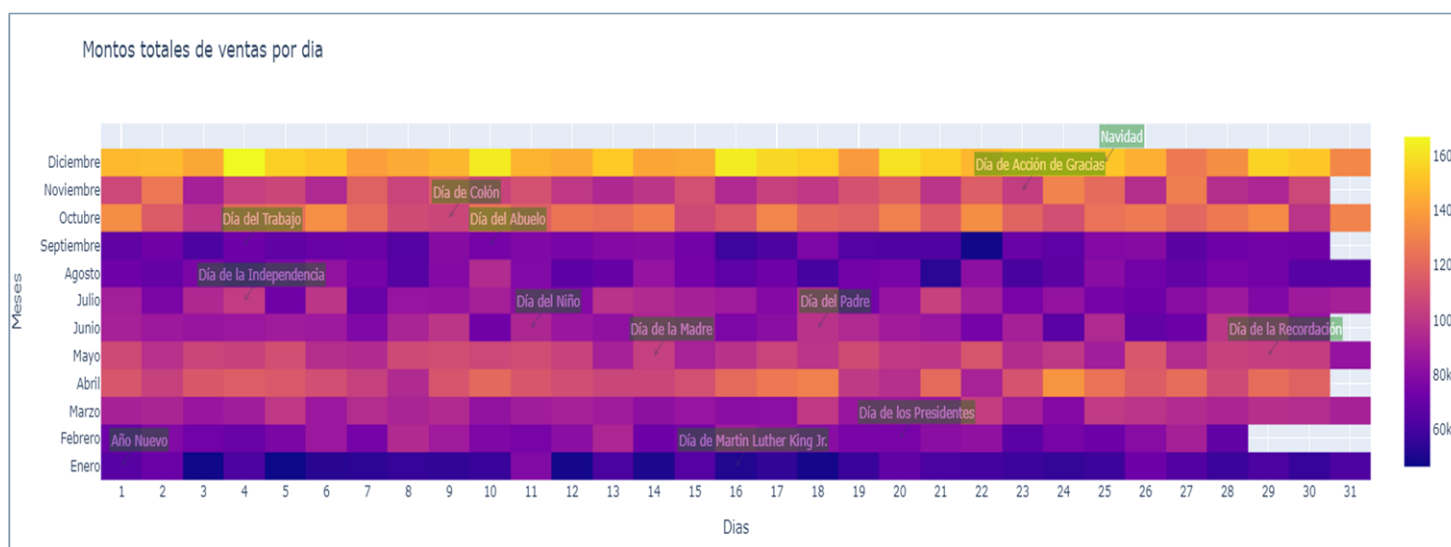


Gráfico 25

## 7.2 ¿Qué eventos específicos han tenido un impacto notable en el comportamiento de las ventas y cómo se manifestó ese impacto?

Notablemente todo el mes de diciembre fue un mes de altos ingresos, el cual podría determinarse por la cercana fecha de navidad. También se podría considerar el aumento los inicios de mayo por el día de la madre. Por otro lado, hay que tomar en cuenta que los productos que están en las ventas son de consumo y de oficina, y no serían causales de obsequios para eventos familiares.

A continuación se describen diferentes eventos que denotan un cambio en los patrones de ventas durante el año.

### *Eventos comerciales:*

- *Columbus Day:* En otoño hay dos grandes temporadas de liquidaciones, la primera de ellas se realiza a mediados de otoño, el mes de octubre por cultura las empresas realizan liquidaciones de sus productos para renovar el stock y evitar así almacenar productos que se desvalorizan o pierden su valor de venta por moda.
- *Summer Sale:* Los primeros días de julio coinciden las liquidaciones de Back to School y del 4 de julio: entre el 1 y el 5 de julio, cercanos al día de la independencia, es otra suba marcada en el grafico 25.
- *President Day:* Cada año se remata la mercancía de invierno entre el 12 y el 15 de febrero, esto se representa en la suba del 13 de febrero.
- *Memorial Day:* Es la liquidación de fin de la temporada primaveral: entre el 28 y el 31 de mayo, Mayo es un mes hermoso en Nueva York, la temperatura es muy agradable y aún no llega la temporada alta así que todo es más económico.

### *Eventos políticos:*

- *Fines de Marzo:* En la última semana de marzo se verifica un leve incremento en las ventas.

El 25 de marzo, Trump anunció que Estados Unidos reconocía la soberanía de Israel sobre los altos del Golán, ocupado por esta desde 1967.

El proceso de destitución del presidente de Estados Unidos, Donald Trump, y la aprobación del proceso en la Cámara de Representantes se convirtieron en los eventos más destacados del país norteamericano durante 2019 y el archivo de la investigación sobre la injerencia rusa en las elecciones decepciona a los demócratas, cercano al 22 de marzo, el fiscal especial Robert Mueller presentó al Departamento de Justicia el informe de la investigación sobre la presunta inferencia de Rusia en las elecciones presidenciales de 2016 y la supuesta relación del equipo de campaña de Trump con Rusia; Para descontento de los demócratas la investigación no encontró ninguna evidencia que poder utilizar en contra de Trump para imputarlo y comenzar un proceso de destitución.

- *23 al 26 de Septiembre:* Leve aumento de ventas en esos días.

Las conversaciones con Ucrania desatan el proceso de destitución, en septiembre se revelaron conversaciones telefónicas entre Trump y el presidente de Ucrania, Volodímir Zelenski, en las que aquel

presionaba a este para que investigase los negocios del exvicepresidente de Estados Unidos, Joe Biden, y su familia en Ucrania a cambio de su apoyo. El 24 de septiembre, la presidente de la Cámara de Representantes, Nancy Pelosi, anunció que se había abierto un proceso de destitución contra Trump por dañar la seguridad nacional y traicionar su juramento presidencial.

- *18 de noviembre, aumento en las ventas del 19 al 25*

El secretario de Estado, Mike Pompeo, anunció que Estados Unidos abandonaba su política anterior, según la cual los asentamientos judíos en la Cisjordania ocupada eran ilegales.

- *Fines de Diciembre: A pesar que diciembre es el mes de mayores ventas los picos se encuentran entre el 16 y el 20.*

El 18 de diciembre, la cámara aprobó la destitución de Trump con 230 votos a favor y 197 en contra, acusando al presidente de abuso de poder y de obstaculizar el funcionamiento del Congreso, que dio comienzo al juicio para la destitución de Trump.

### ***Otros eventos políticos sin repercusión notable en las ventas:***

Ola de dimisiones en la administración de Trump

El 7 de abril, dimitió la secretaria de Seguridad Interna, Kirstjen Nielsen.

En junio, dimitió la portavoz de la Casa Blanca, Sarah Sanders, tras dos años en el cargo. Sanders fue relevada por la directora de comunicaciones de la mujer de Trump, Melania Trump.

El 10 de septiembre, dimitió el asesor de Seguridad Nacional, John Bolton, supuestamente por sus discrepancias con Trump sobre Venezuela.

## 8 PREGUNTA ADICIONAL

### 8.1 Análisis de productos cruzados

El análisis de productos cruzados, también conocido como análisis de asociación, es una técnica utilizada para identificar patrones de compra de productos que tienden a ocurrir juntos en transacciones de ventas. Este tipo de análisis es comúnmente utilizado en el ámbito del comercio minorista y puede proporcionar información valiosa para la toma de decisiones comerciales, como estrategias de marketing, gestión de inventario y diseño de escaparates.

Se analizan las medidas de asociación para identificar patrones significativos. Los conjuntos de productos con un alto soporte, confianza y lift suelen ser los patrones más relevantes. Estos patrones pueden indicar oportunidades para promociones cruzadas, recomendaciones de productos y gestión de inventario.

En este análisis se busca el lift de los productos, se utiliza el dataframe por días en el año, en el cual se toman las ventas diarias y se hace una prueba entre todos los productos, obteniendo un df con las combinaciones de productos y sus valores de lift.

En este análisis se utilizó la librería `itertools`, para obtener las posibles combinaciones obteniendo un array con los diferentes valores, los cuales se plasmaron en un diagrama de calor en el grafico 26, donde se observa claramente las combinaciones que nunca suceden la máxima como:

Productos: [Vareebadd Phone y LG Dryer](#) Lift: 34.63

Y la mínima tal:

Productos: [LG Washing Machine y LG Dryer](#) Lift: 32.43

De todas maneras los valores de lifts mayores a 1 indican que tienen más probabilidades de comprarse juntos que por separado, un valor de 32 es muy positivo y se debe tomar en cuenta, los que se ven en gris en el grafico son valores menores a 1 por tal motivo tienen más probabilidad de venderse solos que en tuplas.

Como conclusión de este punto, es que en cada día hay más probabilidad de vender una [LG Dryer](#) junto a un [Vareebadd Phone](#), que vender una lavadora y secadora de la misma marca ([LG Washing Machine y LG Dryer](#)).

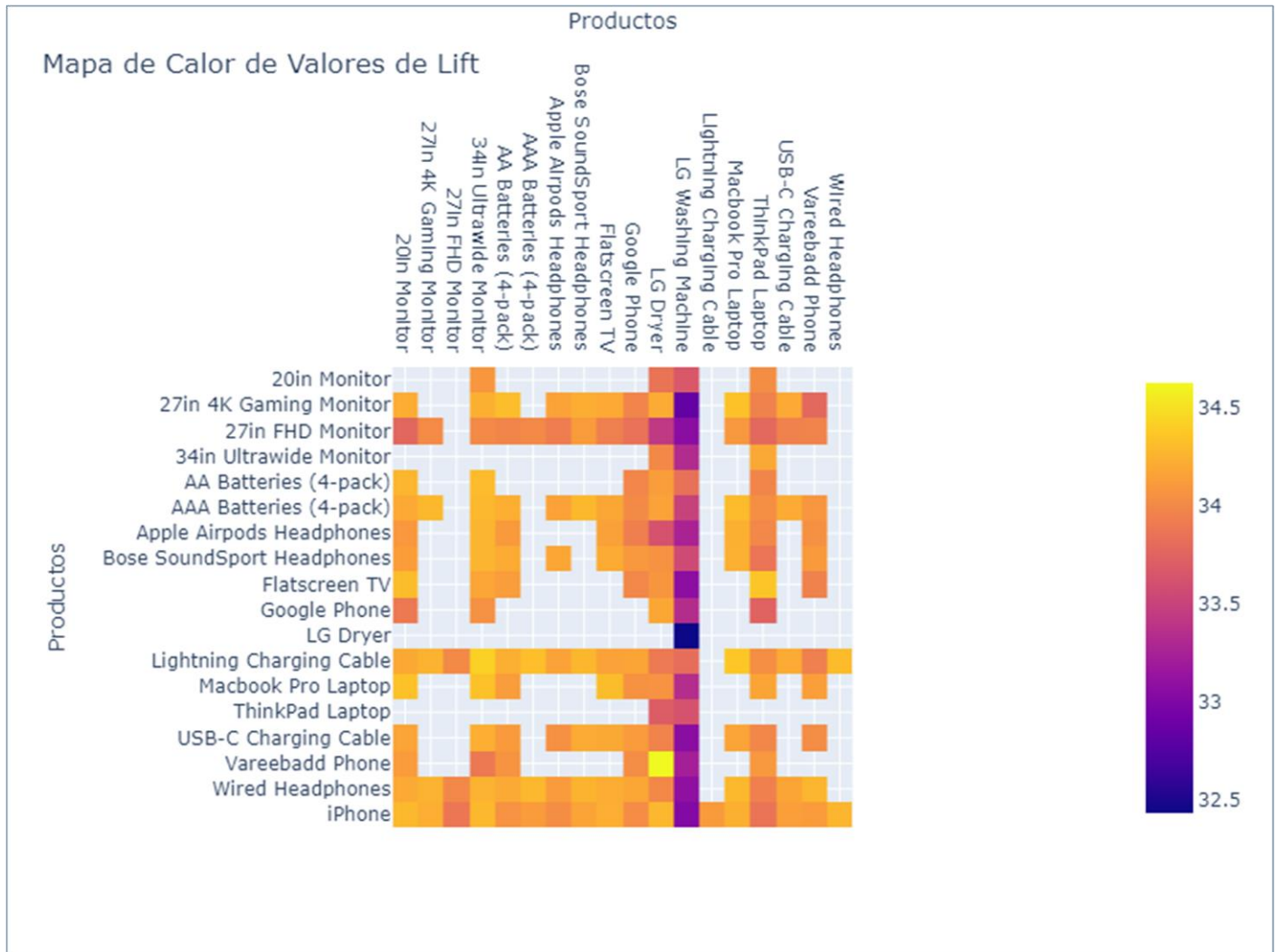


Grafico 26

Como conclusión de este punto, es que en cada día hay más probabilidad de vender una **LG Dryer** junto a un **Vareebadd Phone**, que vender una lavadora y secadora de la misma marca (**LG Washing Machine** y **LG Dryer**).

## 9 CONCLUSIÓN

Desde el punto de vista comercial las diferentes conclusiones en cada sección representadas gráficamente, se pueden apreciar patrones de ventas en referencia a meses, días festivos, efectos políticos, lugares geográficos incluso días del mes, semana y horas en el día, a pesar que la información es simple y con pocos productos ofrecidos, se logró obtener información valiosa, que puede ser utilizada para una estrategia de marketing, ofertas especiales, incluso reservas de stock.

Desde el punto de vista analítico, resulto muy interesante la practica sobre las diferentes herramientas que ofrece la librería Pandas para la manipulación de datos, nos permitió comprender por distintos caminos como llegar a la misma información y lograr corroborar que era el camino correcto.

En la representación el uso de la librería Plotly, nos permitió presentar gráficamente los resultados analíticos ya filtrados. Probamos infinidad de muestras hasta encontrar la más característica que pueda diferenciar los patrones buscados.

Como experiencia fue muy gratificante este laboratorio, independientemente de la complejidad de los datos, es aprender y comprender los diferentes caminos y herramientas para procesar la información disponible.