



Tecnológico de Monterrey

Proyecto Final - Apple

Materia:

Laboratorio de diseño y optimización
de operaciones (Gpo1)

Alumnos:

A01364643 Abril M. Mondragón Plata
A01362737 Angel Eduardo Pérez Giles
A01364735 Claudia E. López Benítez
A01363924 Gustavo Aparicio Uribe
A01368031 Alan Torres Rachner

Campus: Toluca

Fecha de entrega:

19 de Noviembre del 2021

Profesora:

Ana Luisa Masetto Herrera

Introducción: Contexto general y descripción del proyecto.

A lo largo de la historia el ser humano ha producido y acumulado una gran cantidad de información, ya sea en forma de manuscritos, libros o bitácoras y en la actualidad ese volumen de información solo crece de manera exponencial año con año dando origen a la llamada “Big Data”: enormes volúmenes de datos que no pueden ser tratados de manera convencional, pues superan los límites y capacidades de las herramientas tradicionales para la captura, gestión y procesamiento de datos. El Big Data puede ser una enorme fuente de conocimientos cuando se sabe “minar o extraer” la información valiosa; valiosa para una persona, para una empresa o como es el caso para un proyecto.

Nuestro proyecto consiste en realizar una serie de procedimientos de comprensión, selección, limpieza, modelado y evaluación de datos basándonos en la metodología de minería de datos CRISP - DM por sus siglas en inglés “*Cross-Industry Standard Process for Data Mining*”. El proyecto parte de analizar los datos de venta de la empresa Apple en sus sucursales de toda la república mexicana para después realizar diferentes modelos de predicción de demanda y llegar a una conclusión que represente una mejora al modelo actual de la compañía generando así conocimiento de valor a partir de un conjunto de datos que por sí mismos no son de utilidad.

Etapas 1: Comprensión del Negocio

Situación actual

A inicios del 2020 Apple sufrió grandes pérdidas debido a la pandemia generada por COVID-19, tuvo que cerrar muchos centros minoristas alrededor del mundo, por lo que las ventas se redujeron considerablemente y eso derivó la caída de sus acciones.

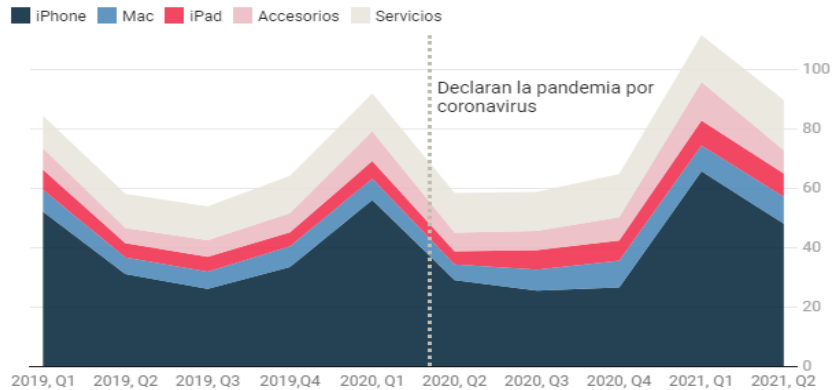
En el segundo trimestre fiscal del 2020 (58.300 millones en ingresos) tuvo un aumento del 2%, pero haciendo también el análisis del tercer trimestre las ventas del iPhone se desplomaron un 7%, se compensó un poco este déficit con un aumento del 16% en los ingresos por servicios (iCloud, Apple Music, Apple TV+) y ventas de iPads y Macs debido al trabajo y estudio en casa; derivado de esto la compañía decidió que no se emitirían predicciones para los siguientes periodos debido a la situación mundial.

Para el primer trimestre fiscal del 2021 Apple logró la mayor cifra de facturación de su historia, con 111.439 millones de dólares en ventas, la mayoría gracias al lanzamiento del iPhone 12; mencionando otros números además del iPhone, los ingresos por computadoras Mac subieron un 21%, las ventas de iPads alcanzaron un aumento del 41%, el segmento de servicios (que incluye desde las ventas de la App Store hasta suscripciones como Apple Music y Apple TV+) también subió, esta vez un 24% y finalmente el mayor incremento fue del segmento de tecnología

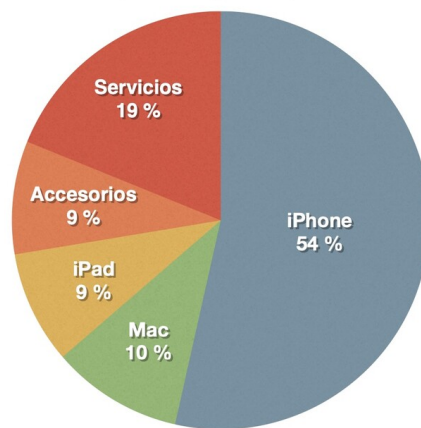
ponible y accesorios para el hogar (Apple Watch, altavoces, entre otros), que subió un 30% todos comparados con el mismo periodo del año anterior.

Apple, antes y después del covid

Ingresos por división (miles de millones de dolares) desde el cuarto trimestre fiscal de 2019 y hasta los resultados presentados en abril de 2021.



Porcentajes de ingresos por división



Apple ya no revela las cantidades en ventas de su producto estrella (iPhone), debido a que desde el 2014 este producto genera arriba del 50% de las ganancias totales anuales de la compañía.

Para septiembre y octubre de cada año Apple organiza dos eventos para promocionar y dar fechas de lanzamiento de sus nuevos productos y softwares, pero para este año no se han confirmado fechas para estos eventos, por lo que los únicos nuevos productos son iMac, iPad Pro, iPhone 12 morado, AirTag y Apple TV 4K, lanzados en Abril del 2020.

Problemática (en términos del negocio)

En el último estado financiero que publicó la compañía se puede ver un incremento en las ventas a pesar del impacto de la pandemia de COVID-19, los resultados favorables se deben a los riesgos que ha tomado la compañía en el área de tecnología y diseño para sus nuevos productos. Evolución que ha tenido Apple para expandir su catálogo de servicios y productos ha posicionado a la empresa como una de las mejores compañías de telecomunicación por su constante evolución.

El análisis de datos es una herramienta que nos va ayudar a comprender cómo se comporta el mercado de Apple tomando en cuenta los gadgets vigentes que maneja la empresa. Los resultados van a ser relevantes para administrar eficientemente las operaciones y las ventas que puede tener la empresa en un futuro.

Problemática (en términos de ciencia de datos)

Los datos presentan estacionalidad en el primer cuarto del año habiendo una alza en las ventas de Apple, decayendo para el segundo cuarto y manteniendo cierta estabilidad el resto del año. La problemática a analizar es una predicción de demanda (regresión) para cada una de las sucursales de Apple a lo largo de la república mexicana, pues se nos solicita conocer cuáles son los productos más vendidos, los menos vendidos y formular una solución para ello. Esa es la pregunta a responder: ¿En cada sucursal cuál es el producto del que se tiene un mayor número de ventas (beneficio)?

Además de lo anterior se podría aplicar la Clasificación de productos ABC la cuál establece que el 80% de las ganancias vienen del 20% de los productos, de esta manera sabríamos de qué productos estamos obteniendo el mayor beneficio. Es una práctica útil de control de los inventarios y se puede realizar para cada sucursal individualmente y así diferenciar cada producto en cada categoría:

- Producto A: Hasta el 85% de las ventas
- Producto B: Más de 85% y hasta el 90% de las ventas
- Producto C: Más de 90% y hasta el 100% de las ventas

Objetivos

Etapas 1: Comprensión del negocio

- Entender y describir los principales problemas que enfrenta la empresa analizada (Apple) en términos del negocio y ciencia de datos para poder establecer los objetivos y un plan de acuerdo a las necesidades del cliente.

Etapas 2: Comprensión de los datos

- Realizar un análisis superficial de datos para poder identificar problemas de calidad en los mismos.
- Entender a grandes rasgos las características de la empresa. Identificar los principales high runners de Apple a través de los años, ¿Han cambiado con el paso de los años? ¿Qué factores se pueden ligar con estos cambios?

Etapas 3: Preparación de los datos

- Realizar una limpieza de los datos con el fin de trabajar únicamente con información confiable.
- Identificar la tendencia en las ventas a través de los años y los factores (variables) que contribuyeron en las mismas.

Etapas 4: Modelado

- Implementar un modelo de pronósticos sobre los siguientes años usando promedios móviles cambiando los promedios al menos 3 veces para reducir el sesgo o error absoluto e implementar un modelos de aprendizaje de máquina que se adecue a los cambios en el tiempo.

Etapas 5: Evaluación (Resultados)

- Mediante el uso de gráficas, comparar el sesgo o error absoluto entre los distintos métodos de pronóstico para ver cual es el que mejor se adecua a la empresa.
- A partir de los resultados obtenidos hacer conclusiones y recomendaciones respecto a la información recabada.

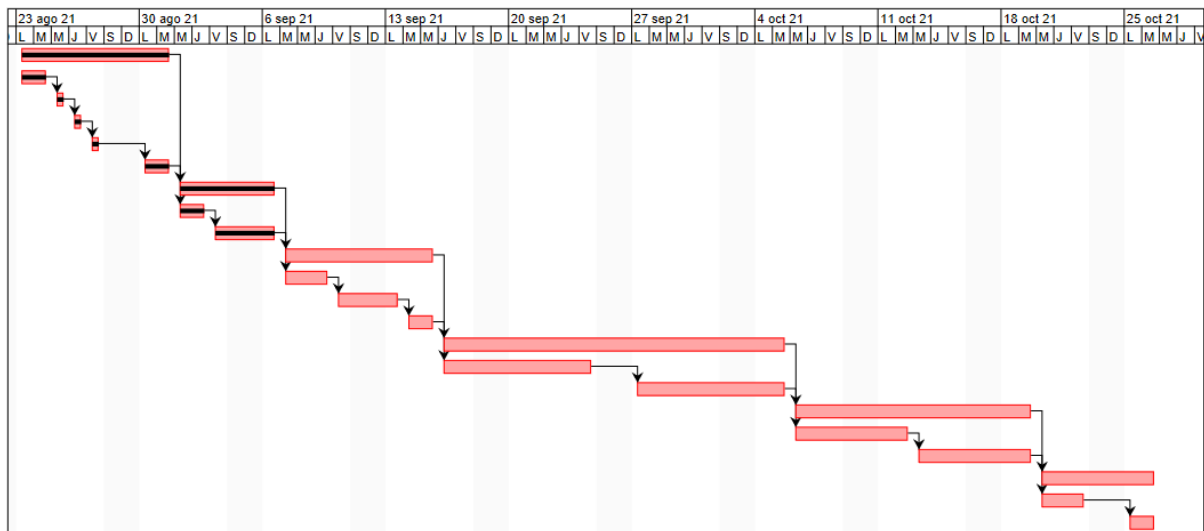
Etapas 6: Implementación

- Mostrar los resultados obtenidos mediante una presentación ejecutiva en la que se muestran los hallazgos encontrados por el equipo donde se presentan los datos más relevantes en la realización del proyecto.

Estructura y plan preliminar

Al dividir el proyecto en 6 etapas como se muestra en el punto anterior y analizar las actividades a realizar, decidimos que todos estaríamos involucrados en las actividades y conforme avanzamos determinar quienes serán los líderes de esa etapa dependiendo de las habilidades que desarrollamos, es por eso que aún tienen un TBD (to be determined). Todo esto se puede observar en el siguiente Gantt de manera clara y gráfica.

		Nombre	Duración	Inicio	Terminado	Completo	Texto1	Predecesores	
1		Etapas 1: Comprensión del Negocio	7 days	23/08/21 8:00	31/08/21 17:00	<input checked="" type="checkbox"/>	Todo el equipo		v s d
2		1) Descripción de la situación actual.	2 days	23/08/21 8:00	24/08/21 17:00	<input checked="" type="checkbox"/>	Abril		
3		2) Entender y describir la problemática (en términos del negocio).	1 day	25/08/21 8:00	25/08/21 17:00	<input checked="" type="checkbox"/>	Gustavo	2	
4		3) Entender y describir la problemática (en términos de ciencia de datos).	1 day	26/08/21 8:00	26/08/21 17:00	<input checked="" type="checkbox"/>	Angel	3	
5		4) Plasmear los objetivos.	1 day	27/08/21 8:00	27/08/21 17:00	<input checked="" type="checkbox"/>	Alan	4	
6		5) Estructurar el proyecto y hacer un plan preliminar.	2 days	30/08/21 8:00	31/08/21 17:00	<input checked="" type="checkbox"/>	Claudia	5	
7		Etapas 2: Comprensión de los datos	4 days	1/09/21 8:00	6/09/21 17:00	<input checked="" type="checkbox"/>	Todo el equipo	1	
8		1) Describir los datos crudos.	2 days	1/09/21 8:00	2/09/21 17:00	<input checked="" type="checkbox"/>	Todo el equipo	6	
9		2) Detectar problemas de calidad.	2 days	3/09/21 8:00	6/09/21 17:00	<input checked="" type="checkbox"/>	Todo el equipo	8	
10		Etapas 3: Preparación de los datos	7 days	7/09/21 8:00	15/09/21 17:00	<input type="checkbox"/>	Todo el equipo	7	
11		1) Limpiar los datos	3 days	7/09/21 8:00	9/09/21 17:00	<input type="checkbox"/>	TBD	9	
12		2) Realizar un Análisis Exploratorio	2 days	10/09/21 8:00	13/09/21 17:00	<input type="checkbox"/>	TBD	11	
13		3) Seleccionar y construir variables	2 days	14/09/21 8:00	15/09/21 17:00	<input type="checkbox"/>	TBD	12	
14		Etapas 4: Modelado	14 days	16/09/21 8:00	5/10/21 17:00	<input type="checkbox"/>	Todo el equipo	10	
15		1) Construir el modelo de promedios móviles	7 days	16/09/21 8:00	24/09/21 17:00	<input type="checkbox"/>	TBD	13	
16		2) Construir un modelo de aprendizaje de máquina	7 days	27/09/21 8:00	5/10/21 17:00	<input type="checkbox"/>	TBD	15	
17		Etapas 5: Evaluación	10 days	6/10/21 8:00	19/10/21 17:00	<input type="checkbox"/>	Todo el equipo	14	
18		1) Construir gráficas para contrastar el desempeño de los modelos.	5 days	6/10/21 8:00	12/10/21 17:00	<input type="checkbox"/>	TBD	16	
19		2) Construir conclusiones dependiendo de los resultados obtenidos.	5 days	13/10/21 8:00	19/10/21 17:00	<input type="checkbox"/>	TBD	18	
20		Etapas 6: Implantación	5 days	20/10/21 8:00	26/10/21 17:00	<input type="checkbox"/>	Todo el equipo	17	
21		1) Crear un reporte final	3 days	20/10/21 8:00	22/10/21 17:00	<input type="checkbox"/>	TBD	19	
22		2) Crear una presentación ejecutiva	2 days	25/10/21 8:00	26/10/21 17:00	<input type="checkbox"/>	TBD	21	



Etapa 2: Comprensión de los datos

Datos crudos

Se tienen más de 100,000 entradas recaudadas de las diferentes sucursales repartidas a lo largo de la república mexicana en un archivo .csv de los años 2018-2019.

Las categorías para los datos son: punto_de_venta, fecha, mes, anio, num_ventas, marca, gamma, costo_promedio, zona, estado, ciudad, latitud, longitud.

Problemas de calidad

- Error en la categoría de año por “anio”
- Mezcla formato de año 2021 con 21
- Mezcla nombre de mes con número de mes
- En el apartado de marca mezcla mayúsculas y minúsculas
- Errores en los números sku

Etapa 3: Preparación de los datos

Como se mencionó en la entrega anterior la marca con la que vamos a trabajar es Apple, conocida compañía en la venta de equipos móviles y otros equipos electrónicos. A inicios del 2020 Apple sufrió grandes pérdidas debido a la pandemia generada por COVID-19, tuvo que cerrar muchos centros minoristas alrededor del mundo, por lo que las ventas se redujeron y eso derivó en la caída de sus acciones.

Al equipo se le fue entregado un conjunto de datos dentro del cual se tienen más de 100,000 entradas recaudadas de las diferentes sucursales repartidas a lo largo de la república mexicana en un archivo .csv de los años 2018-2019. Las categorías para los datos son: punto_de_venta, fecha, mes, anio, num_ventas, sku, marca, gamma, costo_promedio, zona, estado, ciudad, latitud, longitud.

Se pueden observar los siguientes errores de calidad en los datos: Mezcla formato de año 2021 con 21, mezcla nombre de mes con número de mes, en el apartado de “marca” mezcla mayúsculas y minúsculas. Como observación debido a que algunos programas tienen problemas a la hora de leer la letra “ñ” se optó por escribir “anio” para evitar problemas con el programa y por obvias razones.

La limpieza de datos se realizó por categoría, revisando las especificaciones solicitadas para la presentación de los datos en cada una, las categorías que no son mencionadas no tuvieron errores en su registro:

- Correcciones “punto_de_venta”:

Había varias entradas que estaban escritas en algunas ocasiones con mayúsculas y otras con minúsculas, por lo que se cambió a que todas estuvieran escritas con minúscula.

- Correcciones “mes”:

Se mezclaba el formato de número de mes con el nombre del mes, ahora solo se maneja por el número del mes sustituyendo el nombre de mes por su número correspondiente.

- Correcciones “anio”

Se mezcló el formato de año en dos dígitos con el de año en cuatro dígitos, fue cambiado a que todos los años sean de cuatro dígitos.

- Correcciones “marca”:

Se escribió “apple” de varias formas erróneas y mezclando mayúsculas y minúsculas, se cambió a una sola forma para escribir “apple”.

- Correcciones “zona”:

Solo hubo un pequeño error con las mayúsculas al escribir “noroeste”.

- Correcciones “estado”:

Se escribieron nombres de ciudades, por lo que fueron cambiados por su respectivo estado al que pertenecen.

- Correcciones “latitud”:

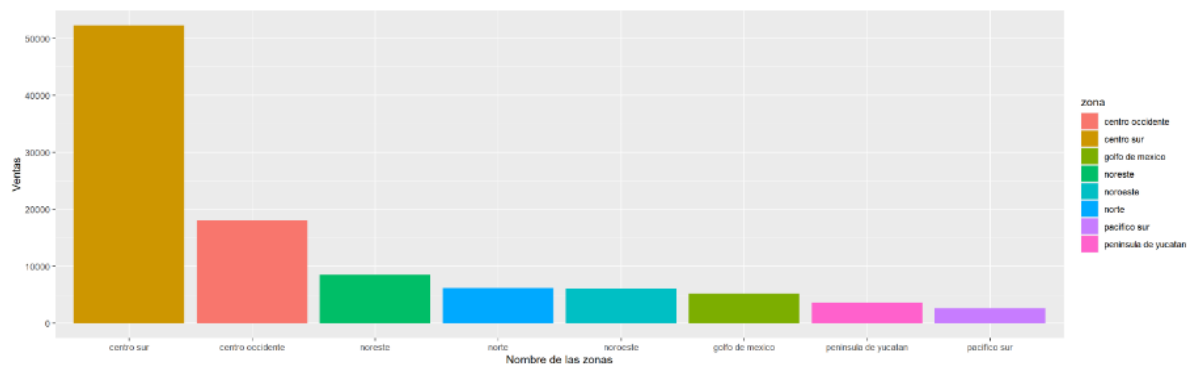
Había un error en la entrada número #19048 donde se puso un e+6 en lugar de un e+1

- Correcciones “longitud”:

Había un error en la entrada número #19048 donde se puso un e+6 en lugar de un e+1.

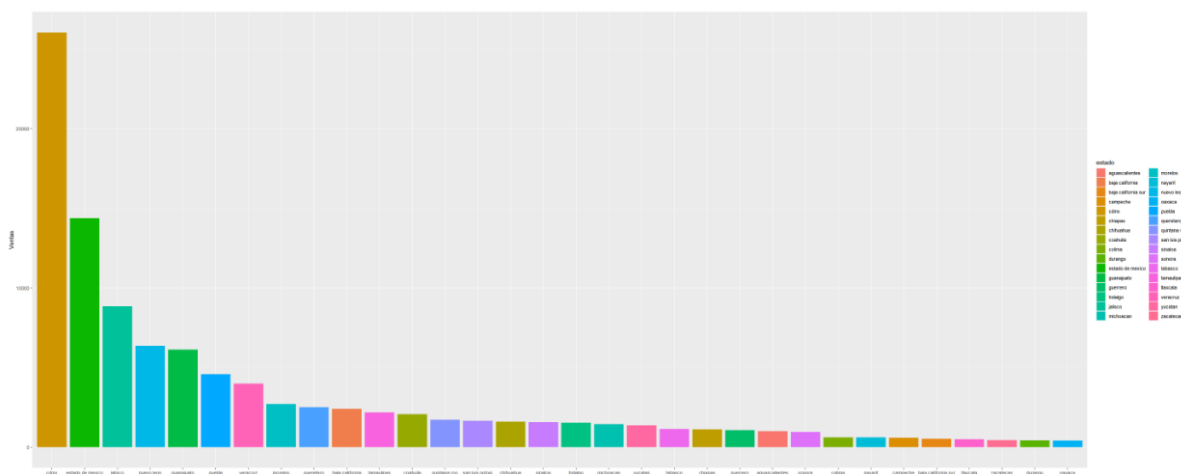
Una vez limpios los datos, realizamos la exploración de los mismos para conocer la situación actual en la que se encuentra la compañía y sus múltiples sucursales de una manera más visual y fácil de entender.

Primero decidimos revisar las ventas de los equipos por zona, para darnos una idea general del panorama.

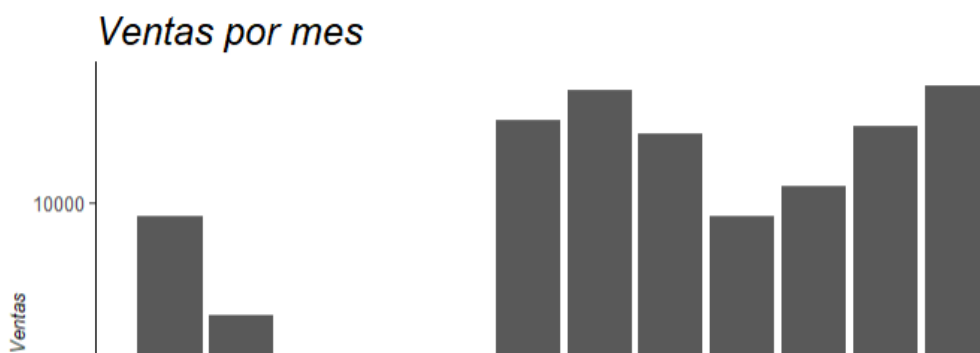


Con la ayuda de la gráfica observamos rápidamente que la zona con el mayor número de ventas es la zona Centro Sur, seguida por la Centro Occidente mientras que la Pacifico Sur es la zona con menos ventas de todas.

Después verificamos cuales eran los estados con el mayor y el menor número de ventas, siendo Ciudad de México el mayor con más de 26,000 ventas y Oaxaca el menor con 400 a lo largo del año que se tiene registrado.



Con estos datos ya podemos afirmar con seguridad que el lugar con mayor número de ventas es la Ciudad de México, por lo que ahora hay que plantearse si queremos mejorar las ventas en cada sucursal o si esperamos mejorar las ventas en los lugares donde ya hay un mercado bien establecido.



En la gráfica anterior se observan del lado izquierdo las ventas de enero a marzo del 2019 y del lado derecho las ventas de junio a diciembre del 2018.

De acuerdo a los datos que tenemos solo podemos evaluar 10 meses, los primeros 3 meses del 2019 y del mes 6 al 12 en el 2018. Se puede ver en la gráfica un decremento de las ventas en el 2019 a comparación de las que tuvimos en el 2018. Esto es debido a que desde el mes 6 del 2019 podemos ver que las ganancias que tiene Apple estaban por arriba de todas las que tuvieron en 2019.

Etapas 4: Modelado

Para poder hablar de modelado es necesario dar una breve introducción a los promedios móviles, pero aún antes de eso, debemos introducir el concepto de pronóstico. El pronóstico en estadística es la estimación sobre lo que se espera que pueda suceder respecto a una variable; esto, en base a un análisis numérico. El pronóstico puede ser utilizado a distintos niveles, permitiendo tomar decisiones para mitigar los efectos negativos que puedan tener ciertas contingencias en caso de que llegaran a concretarse.

Método “Promedio Móvil”

El método de pronóstico “promedio móvil simple” se utiliza cuando se quiere dar más importancia a conjuntos de datos más recientes para obtener la previsión. Cada punto de una media móvil de una serie temporal es la media aritmética de un número de puntos consecutivos de la serie, donde el número de puntos es elegido de tal manera que los efectos estacionales y / o irregulares sean eliminados. El pronóstico de promedio móvil es óptimo para patrones de demanda aleatorios o nivelados donde se pretende eliminar el impacto de los elementos irregulares históricos mediante un enfoque en períodos de demanda reciente.

\hat{X}_t Promedio de ventas en unidades en el período t

Σ Sumatoria de datos

$$\hat{X}_t = \frac{\sum_{t=1}^n X_{t-1}}{n}$$

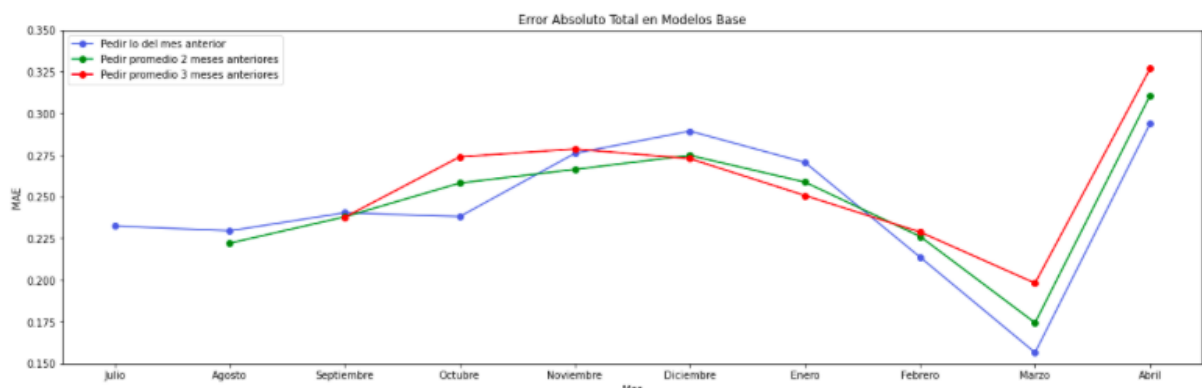
X_{t-1} Ventas reales en unidades de los períodos anteriores a t

n Número de datos

La ventaja del promedio móvil simple es que el indicador es suave y, en comparación con la EMA, es menos propenso a muchas señales falsas. La desventaja es que algunos de los datos utilizados para calcular la media móvil pueden ser antiguos o no estar actualizados. Sin embargo, la EMA y la SMA se utilizan de manera similar: para ver tendencias e identificar áreas de soporte o resistencia. Dado que ninguno de los dos medios es fundamentalmente superior, la cuestión de uno de ellos suele estar determinada por el estilo de negociación o el marco de referencia analítico del usuario.

A continuación se muestran los resultados obtenidos del uso de promedios móviles en nuestro proyecto.

Se realizaron 3 modelos de predicciones y para determinar el más efectivo se calcularon sus errores absolutos respectivamente.



De acuerdo a la gráfica, lo que más convendría utilizar es pedir lo del mes anterior, ya que según los cálculos es el que tiene el porcentaje de error más bajo en casi todos los puntos a excepción de diciembre y enero, pero aún así la diferencia en esos puntos con los otros modelos no es tan grande.

Para Apple según el análisis de las predicciones obtenidas nos arroja que lo más conveniente para los siguientes meses es pedir lo del mes anterior.

El aprendizaje de máquina tiene como objetivo desarrollar un programa que pueda aprender de sí mismo para desarrollar su funcionamiento sin el constante

involucramiento del ser humano para agregar o cambiar restricciones, para este análisis de Apple se pensaron muchos modelos de predicción para conocer la futuras ventas de la compañía en México pero al final se Promedios móviles simples y árbol de decisión.

El modelo de promedios móviles simples nos ayudó a tomar la demanda de los primeros meses para hacer una estimación de los próximos meses y visualizar con claridad el comportamiento de las ventas que tuvo Apple en el 2018 y 2019. Para los promedios móviles simples se tomó en cuenta las ventas por mes del 2018 y 2019 para considerar la participación de los estados y la gama de modelos vendidos en esos años. No sé consideraron otros parámetros de análisis porque se quiso tener un proyecto conciso y que ayudará a pronosticar las ventas mínimas que podría tener Apple en el 2021. Sin embargo, solo se tomaron unos meses del 2018 y el resto de meses del 2019 para realizar los promedios móviles simples de Apple México porque fue la información dada por la empresa.

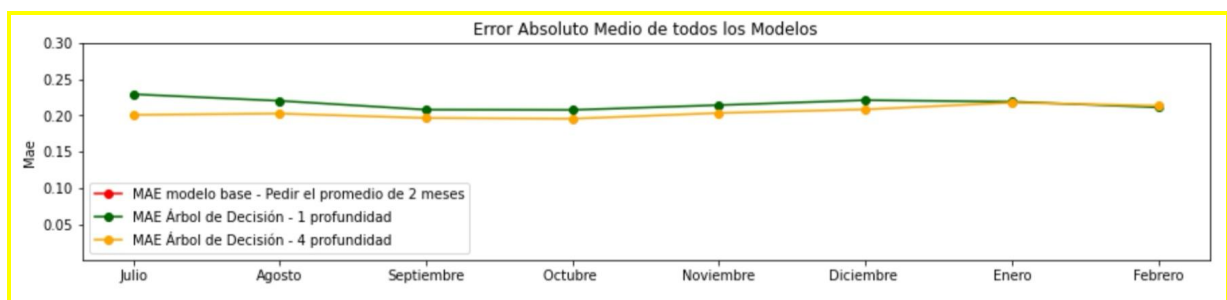
En el análisis de las ventas se tuvo una baja en las ventas predecida en el mes de Febrero y Marzo debido a la información faltante pero los otros meses tuvieron predicciones con errores pequeños por el comportamiento de las ventas estimadas.

Método “Árbol de Decisión”

También se utilizó el modelo árbol de decisión porque toma en cuenta ciertas variables independientes de la base de datos para hacer ramificaciones (grupos y subgrupos) a partir de regresiones, estás ramificaciones clasifican sus resultados con ciertos parámetros que permiten seleccionar con facilidad el resultados más óptimo. Para nuestro análisis se hicieron 8 ramificaciones (particiones) de las ventas mensuales para calcular las medias de cada escenario y tener las métricas finales de cada mes.

Para el árbol de decisión se tuvo que elegir las variables de entrada y las variables de respuesta para elaborar las ocho particiones, se tomó en cuenta las ventas totales de los meses disponibles en el 2018 y 2019 para construir las predicciones de ventas.

Para la matriz final del árbol de decisión se pueden visualizar los resultados de los nueve meses futuros y el modelo base que nos muestra el resume las medias de las observaciones en ese mes.



Etapa 5: Evaluación.

A continuación se muestra cada uno de los 4 métodos de pronósticos utilizados y los resultados obtenidos:

1. Promedios móviles tomando en cuenta los últimos 2 meses. Este método usa el promedio de las ventas de los 2 meses anteriores para hacer un pronóstico.
2. Promedios móviles tomando en cuenta el último mes. Este método usa el promedio de las ventas en el mes anterior.
3. Árbol de decisión con profundidad 1.
4. Árbol de decisión con profundidad 4.

De los modelos propuestos anteriormente se llegó a la conclusión de que el mejor fue el que obtuvimos por el promedio móvil simple del mes anterior, siendo esta la opción con una mejor alternativa para lograr que nuestras estimaciones sobre cuánto debemos pedir en cada tienda sea mejor. Este resultado no fue inesperado en un principio debido a que con este método entre más información se tome en cuenta se obtienen mejores resultados, pero al analizar la situación se concluye que tomar solo un mes es lo más acertado porque detecta las tendencias de compras que a menudo pueden cambiar en periodos menores a un mes.

Referencias:

Comunicado de prensa. (2021, Abril 28). *Apple informa los resultados del segundo trimestre*. apple. Retrieved Septiembre 03, 2021, from <https://www.apple.com/mx/newsroom/2021/04/apple-reports-second-quarter-results/>

El periódico (28 de enero del 2021). La pandemia y el iPhone 12 dan a Apple el mejor trimestre de su historia. <https://www.elperiodico.com/es/economia/20210128/pandemia-iphone-12-dan-apple-11481962>

López, M.. (2021). Resultados financieros del primer trimestre fiscal de 2021: Apple supera los 100.000 millones ingresados en tres meses. Septiembre 2, 2021, de Applesfera Sitio web: <https://www.applesfera.com/apple-1/resultados-financieros-primer-trimestre-fiscal-2021>

La República (1 de mayo de 2020). *Apple reportó que sus ventas de iPhone cayeron 7% desde inicio de la pandemia*. <https://www.larepublica.co/globoeconomia/apple-reporto-que-sus-ventas-de-iphone-cayeron-7-desde-inicio-de-la-pandemia-3000638>

Traders Studio. (2021). ¿Ventajas y desventajas de la media móvil simple (SMA)? 18/10/2021, de Traders Studio Sitio web:
<https://traders.studio/ventajas-y-desventajas-de-la-media-movil-simple-sma/>

GBM Media. (2021). Los promedios móviles en el trading. 18/10/2021, de GBM Media Sitio web:
<https://media.gbm.com/mercados/los-promedios-moviles-en-el-trading/>

Anexos:

Diccionario de Variables:

- punto_de_venta: Diferentes puntos de venta donde la compañía ofrece sus productos
- fecha: Fecha en formato dd/mm/aaaa donde se registra cuando ocurrió la venta
- mes: Permite organizar las entradas por mes en formato de # de mes.
- anio: Permite organizar las entradas por año en formato 20xx
- num_ventas (Unidades): Cantidad de unidades vendidas
- sku: Siglas de “Stock Keeping Unit” o el código que permite identificar un producto del inventario
- marca (Carácter): Marca de la compañía creadora del producto
- gamma (Factor): Categoría de los equipos ofrecidos por Apple
- costo_promedio (Numérico): Costo promedio en el cual se vende producto, existen variaciones en el precio real entre sucursales.
- zona (Carácter): Colonia dentro de la ciudad en la cual se encuentra la sucursal
- estado (Carácter): Estado de la República Mexicana donde se encuentra la sucursal
- ciudad (Carácter): Ciudad donde se encuentra la sucursal.
- latitud: Coordenada de latitud de la sucursal
- longitud: Coordenada de longitud de la sucursal