



## Análisis de datos Huawei

### Entrega Final Proyecto

Laboratorio de diseño y optimización de operaciones

**Profesora: Ana Luisa Masetto Herrera**

Gerardo Gutiérrez Aguirre	A01368146
Juan Ramón Rodríguez Márquez	A01365002
José Manuel González Guadarrama	A01364681
Kevin Alberto Sorzano Mongel	A01368603
Mario Naime Monroy	A01365605
José Alberto Naime Monroy	A01365892

Fecha de entrega: 19 de noviembre de 2021

## **Introducción**

### **Contexto general**

Cuando se habla del sector de las telecomunicaciones , la delimitación conceptual se hace de por momentos difusa debido a que este sector se encuentra inmerso en uno de los campos más dinámicos de la sociedad actual como es todo lo relacionado con las tecnologías para la información ,concepto en el que se engloba todas aquellas actividades económicas y sociales que tienen que ver con el uso de intercambio de información.

La empresa de Huawei fue fundada por el señor Ren Zhengfe inicialmente Huawei empezó como agente de ventas distribuyendo productos PBX importados (sistemas de conmutación centralistas) para una empresa de Hong Kong ,posteriormente Zhengfei vio la oportunidad de realizar ingeniería inversa a esto con el objetivo de poner en marcha una compañía nacional de telecomunicaciones por eso en el año de 1993 Huawei comenzó su propia producción y venta de su PBX (switch digital telefónico) un switch con una capacidad de 10,000 circuitos ,esto le permitió abrirse en el mercado nacional.

### **Descripción del proyecto**

Este proyecto consiste en una base de datos sucios que se nos proporcionó de la empresa de telecomunicaciones Huawei,estos datos se van a tener que limpiar para luego,poder pronosticar la demanda de cada uno de los productos de la marca Huawei en el siguiente mes,en los diferentes puntos de venta en el territorio de México.

## **Etapas 1: Comprensión del Negocio**

### **Descripción de la situación actual:**

Huawei es una empresa china de telefonía móvil que actualmente se encuentra como el tercer más grande vendedor de smartphones en México, ocupando aproximadamente un 14% de la participación en el mercado. Su situación en el mundo se complicó por los bloqueos tecnológicos de Estados Unidos en el gobierno pasado, provocando una caída en sus ventas que hasta este punto no se ha podido recuperar. El panorama para la compañía está lleno de retos entre los que destacan reducir la diferencia con sus competidores, un tema que con la llegada de sus nuevos aparatos y sus estrategias en ventas desean lograr lo más pronto posible.

La importancia de este tipo de proyectos recae en obtener una ventaja competitiva a través del análisis de datos internos para su correcta interpretación y una posterior planeación estratégica a través de los mismos.

### **Descripción de la Problemática**

El negocio de la venta de celulares ha sido una montaña rusa para Huawei, pues empezó como una fuerte amenaza para sus competidores situándose en 2019 como el segundo mayor vendedor de celulares en México con un 16% de participación en el mercado (sólo superado por Samsung), sin embargo las restricciones del gobierno de EUA derivaron en una baja de ventas para el año 2020 y una caída del segundo al tercer lugar en México. El panorama actual no luce alentador es por eso que se busca a través de un análisis de datos, explotar los puntos estratégicos aprovechando los recursos internos para poder figurar de nuevo en los dos primeros puestos de ventas de celular.

Como ingenieros industriales podemos constatar que este tipo de problemáticas son importantes para nuestra carrera ya que muchos de los diferentes empleos que están relacionados con ella tienen que ver con data science y la forma en la que resolvemos problemas y digerimos la información obtenida. Muchos de estos empleos requieren que podamos ubicar, filtrar y encontrar datos específicos y que su rápida localización sea indispensable para poder facilitar el proceso de búsqueda, por lo que es indispensable aplicarlo de la mejor manera.

### **Entender y describir la problemática (en términos de ciencia de datos, tipo de tarea, tipo de datos, etc.).**

Enfrentamos un problema de regresión debido a que manejamos datos cuantitativos y queremos obtener el mejor dato numérico para la correcta predicción de la demanda y buscamos responder a la pregunta “¿Cuál es la forma óptima de trabajo para la organización ya que los costos sean mínimos con la mayor ganancia?”. Mencionado lo anterior, a continuación mostramos material que nos ayudará a entender mejor la problemática de la empresa y con ello poder definir de mejor manera los objetivos.

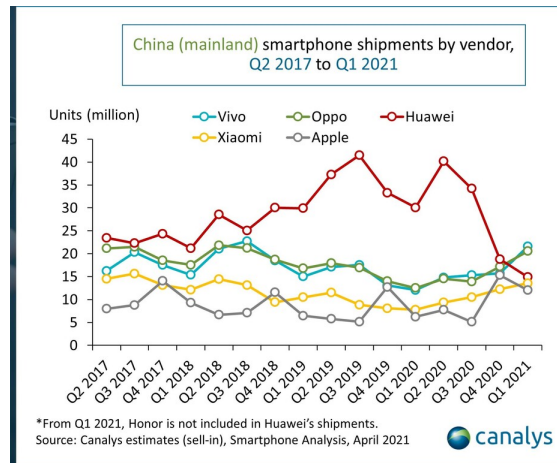


Fig 1: Gráfica de envíos de celulares por proveedores (cuartos de año).

Las estadísticas muestran que Huawei ha pasado de tener 41% del mercado total de smartphones en China a un 16% en los últimos tres periodos mostrados en la tabla 1. Vivo y Oppo, han superado claramente a Huawei para este año 2021 y la tendencia indica que Xiaomi también lo hará.

#### Huawei shipments continue to hemorrhage

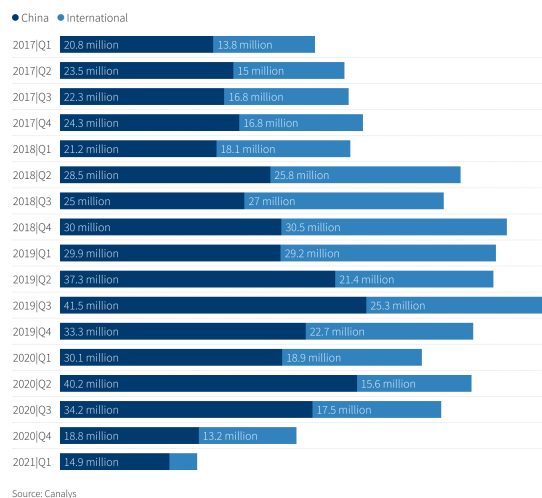


Fig 2: Gráfica de envíos de celulares Huawei

En la tabla 2, podemos observar que el decrecimiento es inminente no solo en China. Sino que a nivel internacional Huawei se está derrumbando. La empresa ha sido claramente afectada por diferentes factores, donde sobresale el veto impuesto por la administración de Trump y mantenido por la nueva de Biden. La empresa está enfrentando una etapa nunca antes vivida, por ende en términos económicos podríamos asumir que se encuentran en una crisis.

## Plasmar los objetivos

Como objetivos personales definimos los siguientes. Los dividimos en primarios y secundarios con el fin de secuenciar el proceso de trabajo para asegurar el correcto funcionamiento del mismo. Nuestro objetivo primario se basa en la correcta realización del proyecto haciendo un buen sistema para la predicción de la demanda por medio de la metodología CRISP DM que sea capaz de resolver o disminuir el impacto de las adversidades que se presentan en la empresa. Los objetivos secundarios los planteamos basándonos en las expectativas de los medios del proceso, pensando de forma sistémica y secuencial tenemos los siguientes puntos para cada una de las etapas:

- **Etapas 1 – Comprensión del Negocio:** descubrir desde el principio factores importantes que pueden influir el resultado del proyecto.
- **Etapas 2 – Comprensión de los datos:** familiarización con los datos. Describir y clasificar datos crudos, detectar problemas de calidad.
- **Etapas 3 – Preparación de los datos:** filtración de ellos. Realizar una limpieza óptima de los datos. Aplicación de ingeniería estadística.
- **Etapas 4 – Modelado:** codificación y análisis de datos. Planteamiento de modelos.
- **Etapas 5 – Evaluación:** evaluación de los modelos propuestos anteriormente y análisis de resultados. Conclusiones.
- **Etapas 6 – Implantación:** aplicación del proyecto y presentación final.

El fin general es que el proyecto sea realizado de manera correcta utilizando las herramientas de programación y procesamiento de datos vistas en clase.

## Estructurar el proyecto y hacer un plan preliminar

Con base en la organización de las actividades del proyecto, se hizo la siguiente distribución y asignación de tareas con tiempos establecidos para posteriormente realizar en diagrama de Gantt y poder visualizarlo de mejor manera:

Al ser un equipo de trabajo de seis integrantes, designamos un líder para cada etapa según nuestros rasgos y habilidades.

*Las organización de equipo se encuentra en el apartado de “Anexos”*

Debido a la extensión del diagrama no pudimos obtener una buena imagen del mismo, sin embargo, en el link a continuación se puede encontrar:

[https://docs.google.com/spreadsheets/d/1EoAdx8eBGwqOO9F\\_BIm3qpjoA02rlqyl/edit?usp=sharing&ouid=114674023006189029027&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/1EoAdx8eBGwqOO9F_BIm3qpjoA02rlqyl/edit?usp=sharing&ouid=114674023006189029027&rtpof=true&sd=true)

## **Etapa 2 Comprensión de los datos**

### **Descripción de los datos crudos**

Los datos con los que vamos a trabajar vienen en un archivo de Excel este archivo de Excel cuenta con 369,618 renglones ,también tiene 14 columnas el archivo cuenta con 14 variables que son:punto de venta,fecha,mes,año,núm ventas, sku, marca, gamma, costo promedio, zona, estado, ciudad, latitud y longitud. Se cuentan con 1902 puntos de venta en 9 zonas distintas, 237 ciudades y existen tres tipos de gama: alta, baja y media. Para mayor duda consultar:

*2. El Diccionario de variables se encuentran en el apartado de “Anexos”*

### **Problemas de calidad con los datos**

A continuación se listan los problemas que se encontraron a la hora de observar los datos:

- Hay un problema con los meses, debido a que algunos están escritos con números y otros están escritos con abreviación.
- En la columna de año, año está escrito como "anio", además hay otro problema debido a que toda la columna debería estar escrita con números de 4 cifras como "2018" y hay algunos años que están marcados como "18".
- Con la variable de zona hay un problema con la ortografía ya que algunos datos están escritos con mayúsculas y otros con minúsculas, además hay unas zonas que están repetidas pero escritas de diferentes maneras
- En la columna de marca también existe un problema de ortografía con las mayúsculas y minúsculas
- Dentro de la columna de estados hay datos erróneos ya que dentro de esta hay registros de ciudades como "Meteppec" y "La Paz", cuando estas no deberían estar ahí porque son ciudades.

## **Etapa 3 :Preparación de los Datos**

### **3.1 Limpieza de datos**

Para la limpieza de los datos, lo primero que se hizo fue llamar la librería “tidyverse” y la librería “rdr”, posteriormente se le pide al programa que haga un resumen de todos los datos del archivo.

A continuación lo que se hizo fue transformar todos los datos de la columna de puntos de venta a minúsculas con la función “tolower”, después se utilizó la función “str\_replace” para reemplazar los puntos de venta que estaban mal escritos a la manera correcta, además también se utilizó esta función para reemplazar las vocales acentuadas a vocales no acentuadas.

Posteriormente, se hace lo mismo con la columna mes, utilizar la función “str\_replace”, lo que se va hacer es que en lugar de tener el mes con las iniciales se va a reemplazar con el número correspondiente de cada mes y después se guarda como variable numérica.

La misma situación aplica para año, reemplazar los años que estén mal escritos con el año escrito en 4 números con “str\_replace”.

A continuación los datos de la columna marca se cambian a minúsculas con la función “tolower” y se aplica lo mismo que en pasos anteriores reemplazar los nombres que estén mal escritos a la palabra “huawei” con la función “str\_replace”.

Para la columna de zonas lo que se hace es transformar todos esos datos a minúsculas con la función de “tolower”.

Para la variable de “estado” lo que se hace es reemplazar el nombre de las ciudades con el nombre de los estados correspondientes, igual con el uso de la función “str\_replace”.

Y lo que se hizo al final es que se muestro un resumen de datos tanto para la latitud como para la longitud de mínimo, máximo, moda, columnas, cantidad, etc.

### **3.2 Análisis exploratorio**

#### **Descripción General de las variables**

Tenemos 34 diferentes modelos de teléfonos móviles de la marca Huawei vendidos en México durante los años 2018 y 2019. Donde según nuestra base de datos todos estos se vendían durante todas las fechas del año, a excepción de de los meses de

Abril y Mayo (4 , 5, respectivamente) ya que no contamos con registros de ventas para esas fechas. Dentro de las variables que consideramos importantes están la gama del teléfono con tres niveles (alto, medio y bajo), pues creemos que es relevante hacer un análisis de los números de venta y los beneficios que se pueden obtener dependiendo de la diferente gama de telefonía. Del mismo modo una variable que estamos seguros será eje de estudio en nuestro análisis es el producto para el cual contamos con 34 diferentes categorías que nos permitirá analizar qué producto supone un beneficio mayor de todos y cuál un menor. Por último y la más importantes variable sobre la cual muchas otras giran es el costo de venta al cliente, pues con este proyecto se busca generar un análisis que permita obtener una ventaja competitiva a través de la información, es por eso que mirar el precio de determinados productos y el cómo se desenvuelve tanto en determinado lugar cómo en determinado tiempo nos hará profundizar en este estudio.

## **Análisis**

Estudiando el “Resumen de Datos I” podemos apreciar que:

A. Existe congruencia entre todos los datos obtenidos. Ya que la variable “Lenght” nos indica que para cada uno de los factores, se obtuvo el mismo número de datos: 369,617.

B. La media de costos de un artículo por cada venta es de 3,815 pesos. Esto indica que la mayor parte de los teléfonos que se venden se encuentran en la categorizada “gama baja” de la compañía.

C. La mediana es de 3,996 pesos. Lo cual nos indica que el precio al que se vende el mayor número de artículos. Este costo de adquisición se refiere a artículos dentro de la categoría de “gama baja” y a su vez en un valor muy próximo a la media respecto al límite superior e inferior de costos.

D. Observamos una incongruencia en el mínimo de los costos de venta. Ya que cero no puede ser el precio de venta de un artículo.

La tabla: “Tabla Gama vs Ingreso Total”, nos presenta la cantidad de artículos vendidos por cada una de las gamas telefónicas de Huawei. El factor más relevante es que 67.78% de los ingresos que obtiene la empresa en México los debe a la “gama baja”. Igualmente, al observar la “Tabla Gama vs Cantidad de unidades totales vendidas” la gama baja tiene 309,999 ventas de unidades en México, lo cual representa más del 83% de las ventas de unidades totales de la empresa. Este dato se aprecia claramente en el “Gráfico I”[1].

Por los resultados obtenidos en los anteriores análisis. Concluimos que la “gama baja”, representa el mayor ingreso y mayor cantidad de ventas por lejos de Huawei



en México. Por ende enfocaremos nuestro estudio en este sector y las diferentes variaciones de las ventas en los diferentes estados y regiones del país.

Prosiguiendo con el estudio. En el “Gráfico II”[2] podemos observar que la mayor cantidad de ventas en México, se encuentran en el centro del país. Donde el centro-sur es quien tiene el mayor porcentaje de las ventas unitarias totales. Esto nos lleva a hacer un estudio por estados, como es presentado en la “Gráfica III”[3] donde se muestra que la ciudad de México y el Estado de México son los Estados donde se acumula la mayor cantidad de ventas en el país. Donde la ciudad de México tiene un total de ventas de teléfonos Huawei.

El “Gráfico IV”[4] es el más interesante de todos. Este nos muestra una comparación del costo promedio de venta de la “gama baja” de teléfonos en cada estado de la república mexicana. Podemos apreciar diferentes valores, donde hay estados como por ejemplo “Ciudad de México (CDMX)” que tienen una menor dispersión del precio de venta en comparación a otros como Zacatecas. A su vez, notamos que el precio de venta promedio varía dependiendo del estado pero no respecto a la zona. Esto quiere decir que según la región donde te encuentres un artículo puede tener un precio inferior o superior a otro, lo cual afecta directamente a las ventas.

De igual forma, este mismo “Gráfico IV”[4] presenta un predominio de la “gama baja” en cuanto a las ventas y que los artículos de “gama media” y “gama alta” son los menos vendidos. A su vez, volviendo al “punto D” del análisis observamos que en los estados: Ciudad de México, Quintana Roo y Yucatán, es donde se presenta el mayor número de ventas por \$0 lo cual son errores en el registro de las transacciones realizadas, queremos destacar que estos son los estados con más repeticiones de este error pero en todas las demás demarcaciones también existen registros del fallo. Otro fallo observado se encuentra en el “Gráfico IV”[4], el cual nos indica una venta nula de artículos en los meses 4 y 5 (Abril y Mayo). Lo cual no nos hace sentido debido a que es muy específico que junto en esos dos meses que son consecutivos no existan registros de ventas de ningún tipo en ninguna localidad autorizada para vender artículos Huawei en todo el territorio nacional.

En conclusión, queremos llevar nuestro análisis enfocado al grueso de las ventas de Huawei en México, la ya mencionada “Gama Baja”. Debido a la variación de esta queremos estudiar el por qué del cambio de precios según la zona para una misma Gama e incluso un mismo modelo de teléfono. Igualmente queremos enlazar lo anterior con un estudio enfocado hacia los meses donde las ventas aumentan y de esta manera poder generar un plan para que el inventario siempre esté acorde con la demanda pronosticada según el historial de venta de estos dos años estudiados.

### **3.3 Ingeniería de Características**

El objetivo de utilizar la ingeniería de datos es generar más variables a partir de nuestros conjuntos de datos, de los datos limpios, con el objetivo de que nos podrían ayudar a poder predecir mejor los resultados.

Como en todos los programas lo primero que se hace es llamar la librería “tidyverse” y nuestra base de datos limpia con la función `read.csv`, posteriormente se mandan llamar los datos para ver como se ven, después se verifica cuantos renglones y columnas tienen, se hace pide una descripción rápida para ver la información se hace un resumen y se asignan diferentes tipos de variable para que se más fácil trabajar con estos datos.

A continuación lo que se hace es convertir las variables cualitativas a índices, primero se hace para los puntos de venta, se crea un data frame donde se guardan los puntos de venta en orden alfabéticamente posteriormente se crea una nueva columna llamada “`pdv_id`” donde se les va a asignar un valor numérico distinto a los diferentes puntos de venta. Posteriormente se hace lo mismo para los años y los meses, se filtra para que nos salga solamente el año y el mes, se crea el índice “`mes_id`” y se les asigna un valor numérico dependiendo del año y mes que corresponde. Lo mismo se hace para el SKU, se filtra para saber cuántos productos únicos hay y se crea el índice `sku_id` para asignar los valores numéricos a los diferentes productos, para la gamma no se hace nada porque sólo hay una.

Ya que se tienen creados estos nuevos conjuntos de datos, lo que se necesita hacer es mezclar estos nuevos conjuntos a nuestros datos originales utilizando la función `left_join`. Primero realizó con los puntos de venta, lo que se hace es que dependiendo del punto de venta se trae la variable “`id`” correspondiente y se crea una nueva variable con esos datos. Se hace lo mismo respectivamente dependiendo del sku y por año y mes, al final se tiene tres nuevas variables o columnas que son “`pdv_id`”, “`mes_id`” y “`sku_id`”.

Para el segundo paso lo que se hace es sacar las ventas totales en base a “`pdv_id`”, “`mes_id`” y “`sku_id`”, debido a que no está completa la serie de tiempo, se necesita completar, posteriormente lo que se hace es crear las combinaciones con la función de “`merge`” para los conjuntos de datos y posteriormente se vuelve a hacer el `left join` con los datos originales ya para saber cuántas ventas hubo en cada punto de venta.

Lo que se hace posteriormente es crear una variable `y_ventas_siguiente_mes` aquí van los datos que se van a predecir que se obtiene a través de recorrer el valor de la variable de `ventas_totales`.

Después se obtienen nuevas características adicionales que nos podrían ayudar en las predicciones, primero se agrupa por mes\_id y por sku\_id para realizar agrupaciones donde se realizan los conteos de las ventas totales en tienda por cada mes y de los promedios de las venta en tienda de cada mes, a continuación se hace lo mismo respectivamente para cada producto, después utilizamos left\_join para añadir estas nuevas variables a la base de datos.

Otro aspecto que nos ayudó fue la información histórica de estas últimas variables que se generaron, en los últimos 3 meses de cada una.

Y al final si se tienen NA se ponen cero en lugar de y así se obtuvieron nuestros datos con ingeniería de características.

## **Etapas 4 : Modelado**

### ***4.1 Promedios Móviles***

#### ***¿En qué consiste?***

Los promedios móviles son promedios calculados a partir de subgrupos artificiales de observaciones consecutivas". 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 esta técnica también supone que la serie de tiempo es estable.

#### ***Características:***

- Es calculado con cierto periodo definido
- Mientras más corto el periodo con el que se trabaja es menos certero
- Mientras más largo el periodo con el que se trabaja es más certero

#### ***Tipos de Promedios Móviles:***

- Simple
- Exponencial
- Suavización

- Ponderado Linealmente

Algunas ventajas que pueden presentar los promedios móviles son algunas de las siguientes:

- Debido a que son pronósticos que son planificados tienen mucho más valor y son más exactos
- Se utiliza para periodos largos que necesiten suavización

## Consideraciones

Para realizar los modelos con promedios móviles lo primero que se tomó en cuenta fue que todos los datos estuvieran limpios, para eso primero empezamos con la lectura de datos, la cual fue primordial realizar la instalación de la paquetería Pandas y abrir la librería sklearn para poder utilizar la función “mean\_absolute\_error”, para después poder importar nuestro archivo con los datos limpios.

Luego se tomó en cuenta llamar los primeros 10 renglones de nuestra base de datos así como la extensión de está para poder visualizar que estemos trabajando con el archivo correcto.

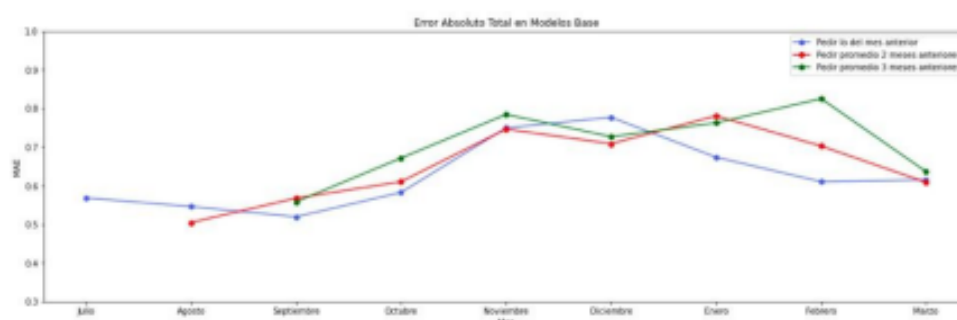
Se visualizó el encabezado de las columnas para posteriormente poder descartar aquellas columnas que no sean de utilidad en el modelo de predicción, es indispensable cumplir con los pasos anteriores para poder empezar con el modelo de promedios móviles.

## Limitantes

Algunos de los limitantes que nos presenta los promedios móviles es que puede ser muy lenta al responder a cambios rápidos o bruscos, también es muy difícil detectar tendencias con estos modelos

## Resultados

Graficamos los tres modelos. Para nuestro análisis respecto al MAE como se muestra en la gráfica de abajo, concluimos que la línea azul la cual representa al promedio móvil de pedir lo del mes anterior es el indicado para los pronósticos.



Respecto a las gráficas de error absoluto medio, hicimos los gráficos del Modelo Base y el Modelo Base - opción 2. En ambos casos se obtuvieron los mismos resultados donde el modelo de pronósticos a utilizar es el que el subgrupo está dado solo por el dato del mes anterior.

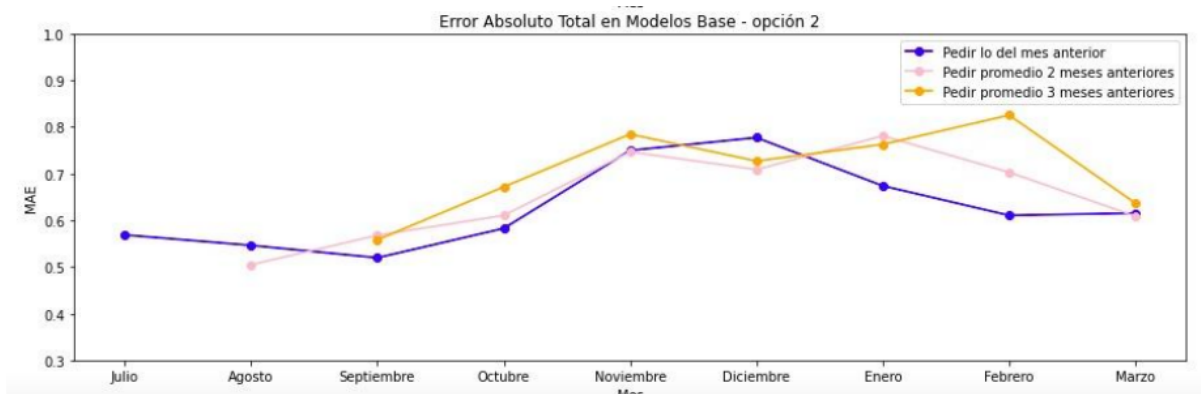


Fig 4: Error Absoluto Total en Modelos Base - opción 2

## 4.2 Aprendizaje de máquina

El aprendizaje de máquina son métodos computacionales a través del aprendizaje de datos para poder mejorar el desempeño de un proceso, este no depende de una conjunto de reglas o instrucciones, sino que este produce las reglas programadas.

El aprendizaje de máquina que nosotros empleamos fue el de diagrama de árbol, el diagrama de árbol es una herramienta que nos permite hallar las relaciones entre un concepto y sus elementos, son los posibles resultados de estas relaciones, le permite a las empresas comparar las diferentes acciones que pueden tomar en base a sus costos y beneficios.

Los diagramas de árbol también se pueden utilizar para el aprendizaje de máquina, para el aprendizaje de máquina se toman en cuenta las observaciones para sobre una generalidad para predecir su valor.

Los componentes del diagrama del árbol son; primero que nada el tronco que es el elemento principal, a este se le van agregando ramas que son los subniveles que puede llegar a tener el diagrama de árbol y el, cada rama contiene un conjunto de atributos de clasificación asociadas que está asociada a una clase específica. Otro elemento que contienen son los nodos, estos nodos pueden ser de tres tipos: de probabilidad, decisión y terminal. Los nodos representan los datos que se tienen.

Algunas ventajas que pueden llegar a tener los diagramas de árbol es que permite tener una visualización de la generalidad y sus detalles, permite un análisis de forma detallada, funciona tanto para datos categóricos como para datos numéricos.

Por otro lado también puede presentar algunas desventajas; cuando hay datos categóricos con múltiples niveles se inclina a favor de los que atributos que con una mayoría de niveles, se pueden volver tediosos y complejos los cálculos debido a la gran cantidad de resultados que se relacionan, puede ser susceptible a errores debido a que se utilizan muchos datos y algoritmos, si hay algún error toma tiempo repararlo por la extensión de los algoritmos.

Para realizar el aprendizaje de máquina se tomaron en cuenta algunas consideraciones, primero que nada que nuestros datos estuvieran limpios, que tuvieran ingeniería de características, el costo del uso de este disminuye con cada punto de datos adicional y puede realizar problemas con múltiples resultados.

Algunas restricciones que presenta el árbol de decisión es que tiene dificultad para trabajar con datos faltantes en los árboles de decisión, otro punto es que desconoce las correlaciones entre atributos en el conjunto de datos, además tienden al sobreajuste de los datos lo que ocasiona una estimación con menos índice de acierto cuando se quieren predecir nuevos casos y otra restricción que se tiene es que si la profundidad los nodos terminales aumentan demasiado, como se tienen más salidas es más difícil comprender las reglas de decisión de árbol debido a que se tienen árboles de decisión más complejos puede ocasionar a que no se adapten bien los datos y le reste capacidad a la interpretación.

## Resultados:

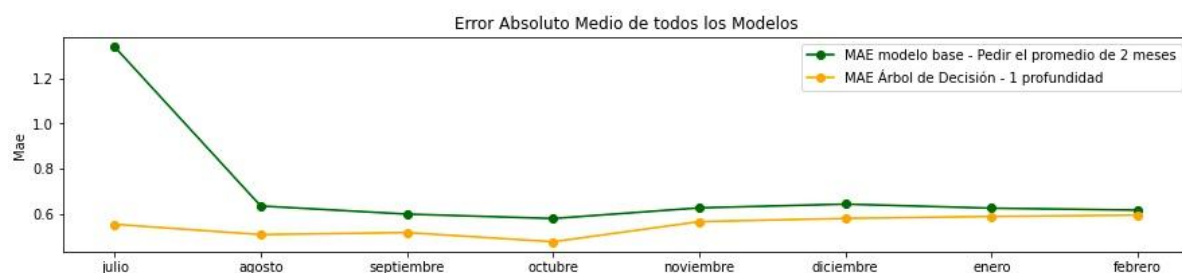


Fig 5: Error Absoluto Medio de todos los Modelos

Para el Árbol de Decisión se utilizaron dos parámetros: un modelo con profundidad 1 y otro con profundidad 5. Comparando las gráficas tomando en cuenta sólo las gráficas del modelo base y Árbol de Decisión con profundidad 1, se eligió el modelo

de MAE Árbol de decisión - 1 profundidad porque es el que tiene mayor consistencia y todos sus valores son inferiores al parámetro anterior.

## Etapa 5 : Evaluación

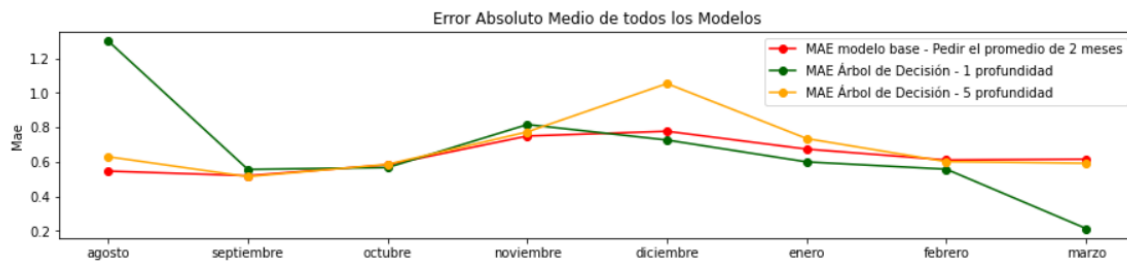


Fig 6: Error Absoluto Medio de todos los Modelos

Después de realizar los diferentes modelos con promedios móviles y con árboles de decisión se llegó a la conclusión que estos tres modelos son los mejores; MAE del mes anterior, Árbol de Decisión - 1 profundidad y Árbol de Decisión -5 profundidad. De estos tres modelos el modelo que le recomendaría a la empresa sería el modelo MAE Árbol de decisión - 1 profundidad debido a que cuenta con un mayor número de repeticiones con menor valor siendo el óptimo de los 3.

Ahora bien, teniendo el modelo elegido y contestando a la pregunta central del inicio del proyecto:

***“¿Cuántas unidades de cada producto de mi marca (Huawei), se van a vender en todos los puntos de venta al siguiente mes de registro?”***

Se obtuvieron los siguientes resultados, que nos muestran la comparación de las ventas anteriores contra el pronóstico de ventas del siguiente mes dividido por sku. Encontramos que existe relación entre ambas cantidades y que es posible comparar las 2 cantidades, véase a continuación:

SKU	Suma de ventas_totales	Suma de y_ventas_siguiente_mes
1	5648	3553
2	6343	4536
3	26083	22984
4	13241	11804
5	6606	4587
6	20586	18820
7	27177	22627
8	11067	9451
9	24767	22052
10	22318	20024
11	4126	2983
12	13646	11793
13	20082	17205
14	7307	7536
15	7104	7326
16	14527	16153
17	5873	6120
18	17786	19257
19	11764	12894
20	17113	18595
21	3796	3796
22	3848	4505
23	2671	3001
24	5631	6423
25	7786	10770
26	6529	9129
27	8680	8961
28	9077	2892
29	6256	1247
30	3121	2429
31	2593	3141
32	2887	3331
33	2619	3205
34	959	3506
<b>Total general</b>	<b>349617</b>	<b>326636</b>

Fig 7: Resultados de ventas con pronósticos por sku

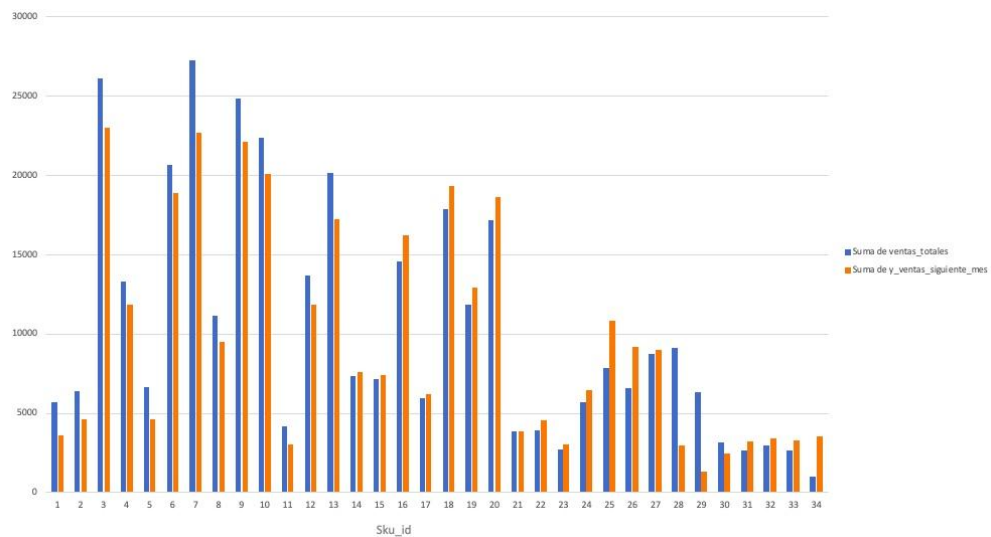


Fig 8: Gráfica de resultados de ventas con pronósticos por sku



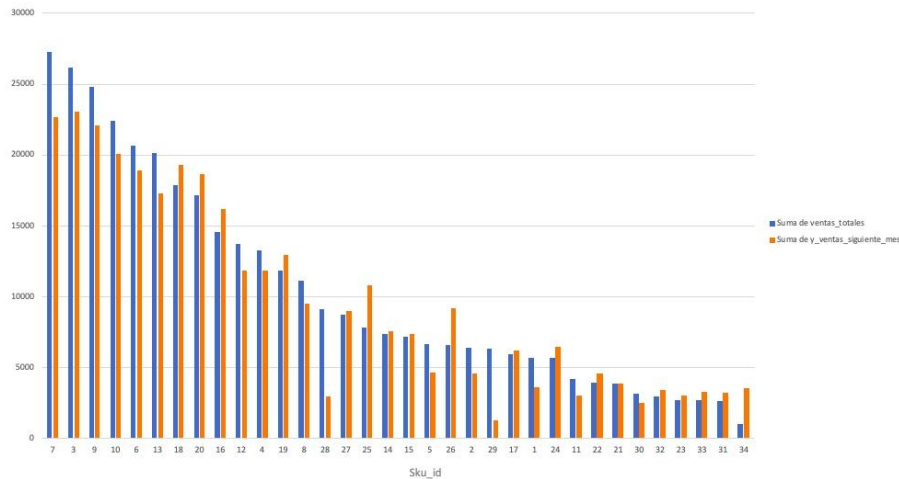


Fig 8: Gráfica de resultados de ventas con pronósticos por sku orden descendente

## Conclusiones

Este proyecto nos ayudó a aplicar todos los conocimientos que aprendimos durante el curso, gracias a los diferentes temas que vimos durante el semestre nos pudimos dar cuenta la importancia que es trabajar primero que nada con una base de datos limpia, ya que para realizar cualquier tipo predicciones, o cálculos es indispensable primero tener una base de datos limpia. Asimismo para el proyecto utilizamos diferentes herramientas como R o python que nos asistieron en nuestra diferentes etapas del proyecto; como por ejemplo el de cómo limpiar una base de datos, qué tipos de métodos de aprendizaje de máquina, cuál modelo es el que más se adapta dependiendo nuestros tipos de datos, además nos pudimos dar cuenta que conociendo y sabiendo los diferentes algoritmos que usamos nos pueden ahorrar mucho el tiempo al realizar cualquier tipo de cálculo que sea necesario. Así que el proyecto nos ayudó a fortalecer todo lo aprendido en clase, los diferentes algoritmos, también nos llevamos diferentes lecciones sobre cómo trabajar en equipo.

Finalmente, fue un proyecto muy enriquecedor en donde nos dimos cuenta que podemos aplicar estos conocimientos en las diferentes áreas en el mundo laboral o para cualquier tipo de situación personal. En la vida profesional siempre se nos presentarán desafíos como este, donde la empresa desea aumentar su utilidad con un costo de inversión mínimo y aplicar proyectos como este de análisis de venta podría hacer que la compañía donde trabajemos consiga su objetivo gracias a un estudio similar al presentado en este proyecto

## Anexos

### Diccionario de variables

- **Punto\_de\_venta:** Es el espacio en el que una empresa establece contacto con su posible cliente potencial.
- **Fecha:** Es un indicador a través del tiempo en que ocurre o se hace algo.
- **Mes:** Periodo que transcurre entre un día y el de igual fecha del mes siguiente.
- **Año:** Periodo a contar de doce meses, a contar desde un día cualquiera.
- **Num\_ventas:** Es el conjunto de ventas de productos o servicios que ha realizado una empresa en un periodo determinado.
- **Sku:** Los códigos sku son elementos fundamentales encargados de llevar el control y gestionar el stock en el almacén.
- **Marca:** Es un nombre, término, señal, un símbolo o una combinación de algunos de ellos que identifica.
- **Gama:** Es un índice que clasifica a los teléfonos con respecto a sus atributos.
- **Estado:** Estado de la república mexicana.
- **Costo\_promedio:** Es el costo por cada unidad dentro de una producción
- **Zona:** Es la extensión de terreno cuyos límites están determinados por razones económicas, políticas, entre otras.
- **Ciudad:** Es una población donde habita un conjunto de personas que se dedican a actividades industriales y comerciales.
- **Longitud:** Distancia angular de un punto de la superficie terrestre, según los meridianos.
- **Latitud:** Distancia angular que hay desde un punto de la superficie de la Tierra hasta el paralelo del Ecuador.

### Gráficas

“Gráfico I”[1].

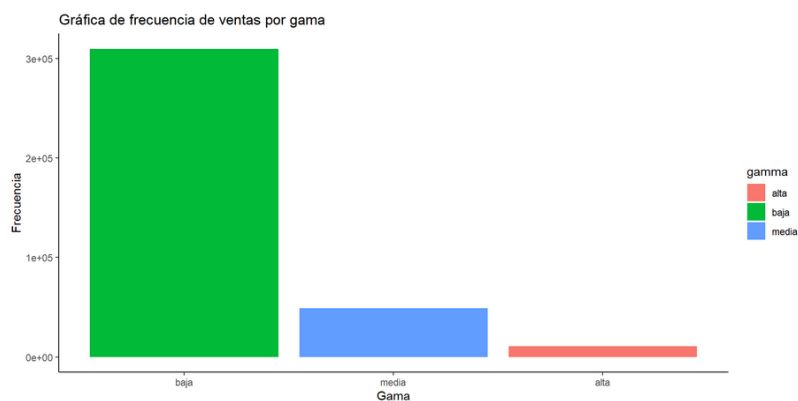


Fig 9: Gráfica de frecuencia de ventas por gama

“Gráfico II”[2]



Fig 10: Gráfica de frecuencia de ventas de gama baja por zona

“Gráfica III”[3]

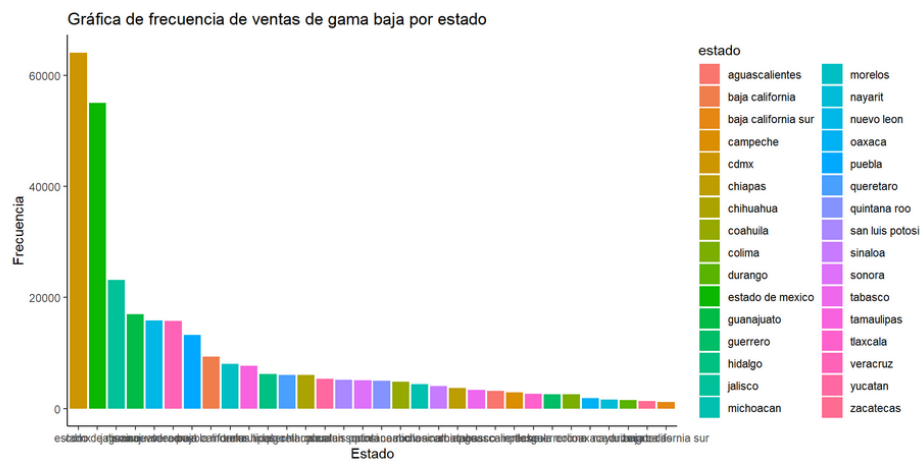


Fig 11: Gráfica de frecuencia de ventas de gama baja por estado

“Gráfico IV”[4]

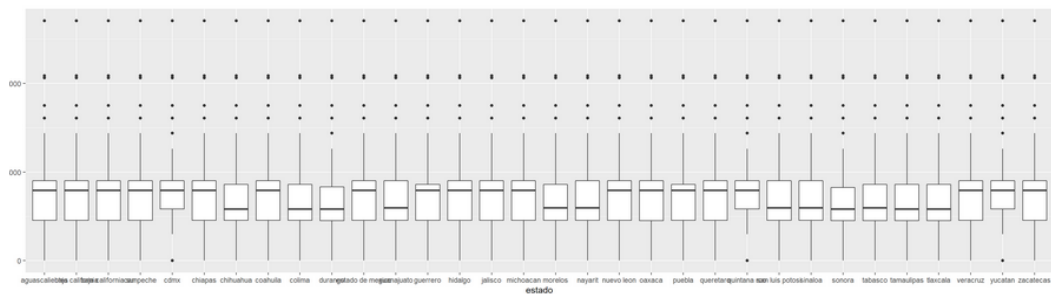


Fig 12: Gráfica de frecuencia de ventas de gama baja por estado

## **Referencias**

- Carlos Fernández De Lara. (2021). *Vamos a regresar al trono, pero nos tomará tiempo: Huawei*. 2021, de Forbes Sitio web: <https://www.forbes.com.mx/tecnologia-vamos-regresar-trono-nos-tomara-tiempo-huawei/>
- Pastor, J. (2021, 30 abril). *Debacle de Huawei en China: el veto provoca la caída en picado de sus ventas incluso en su país de origen*. Xataka. <https://www.xataka.com/moviles/debacle-huawei-china-veto-provoca-caida-picado-sus-ventas-incluso-su-pais-origen>
- BBC News Mundo. (2021, 5 marzo). *Huawei: cómo la falta de acceso a componentes está asfixiando al gigante tecnológico chino*. <https://www.bbc.com/mundo/noticias-56216518>
- I. (2019, 14 febrero). *Qué es el «Machine Learning»*. Iberdrola. Recuperado 18 de noviembre de 2021, de <https://www.iberdrola.com/innovacion/machine-learning-aprendizaje-automatizado>
- A. (2021a, junio 2). *7 ventajas que el Machine Learning le ofrece a su empresa*. American Data. Recuperado 18 de noviembre de 2021, de <https://www.data.cr/2021/06/02/7-ventajas-que-el-machine-learning-le-ofrece-a-su-empresa/>

- Merayo, P. (2020). *Qué son los árboles de decisión y para qué sirven*. Máxima Formación. Recuperado 18 de noviembre de 2021, de <https://www.maximaformacion.es/blog-dat/que-son-los-arboles-de-decision-y-para-que-sirven/>
- *Qué es un diagrama de árbol de decisión*. (s. f.). Lucidchart. Recuperado 18 de noviembre de 2020, de <https://www.lucidchart.com/pages/es/que-es-un-diagrama-de-arbol-de-decision>
- DESCONOCIDO. (-). Introduction to Time Series Analysis. Engineering Statistics Handbook, de Octubre 2021 Sitio web: <https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm> - SUPPORT MINITAB. (2019). ¿Qué es un promedio móvil? Octubre 2021, de MINITAB Sitio web: <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/modeling-statistics/time-series/supporting-topics/moving-average/what-is-a-moving-average>
- DESCONOCIDO. (-). Introduction to Time Series Analysis. Engineering Statistics Handbook, de Octubre 2021 Sitio web: <https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm>
- SUPPORT MINITAB. (2019). ¿Qué es un promedio móvil? Octubre 2021, de MINITAB Sitio web: <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/modeling-statistics/time-series/supporting-topics/moving-average/what-is-a-moving-average/> Lunes 18 de octubre de 2021
- Tokens24 2018. (2018). ¿Qué son los promedios móviles y cómo se utilizan en el comercio?. OCTUBRE 2021, de TOKENS 24 Sitio web: <https://www.tokens24.com/es/cryptopedia/trading/que-son-los-promedios-moviles-y-como-se-utilizan-en-el-comercio>
- SUPPORT MINITAB. (2019). Interpretar todos los estadísticos y gráficas para Promedio móvil. Octubre 2021, de MINITAB Sitio web: <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/modeling-statistics/time-series/how-to/moving-average/interpret-the-results/all-statistics-and-graphs>
- *Promedios Móviles*. prezi.com. (n.d.). Retrieved November 19, 2021, from <https://prezi.com/j9g3cyakg4xs/promedios-moviles/>

- *El sector de las telecomunicaciones*. Telos. (n.d.). Retrieved November 19, 2021, from <https://telos.fundaciontelefonica.com/archivo/numero041/el-sector-de-las-telecomunicaciones/>
- Arteaga, S. (2020, July 11). *Huawei: La Historia de la Compañía Fundada por un ingeniero con 3.000 dólares Y 3 empleados*. ComputerHoy. Retrieved November 19, 2021, from <https://computerhoy.com/reportajes/industria/huawei-historia-672315>
- Domingo, R. (n.d.). *Huawei: Un poco de la historia de esta gran compañía de celulares: Revistas Blogs*. Revistas Blogs | LinkBuilding con Enlaces Naturales. Retrieved November 19, 2021, from <https://revistasblogs.com/consejos-emprendedores/huawei-historia-empresa-celulares/>