

Proyecto Final

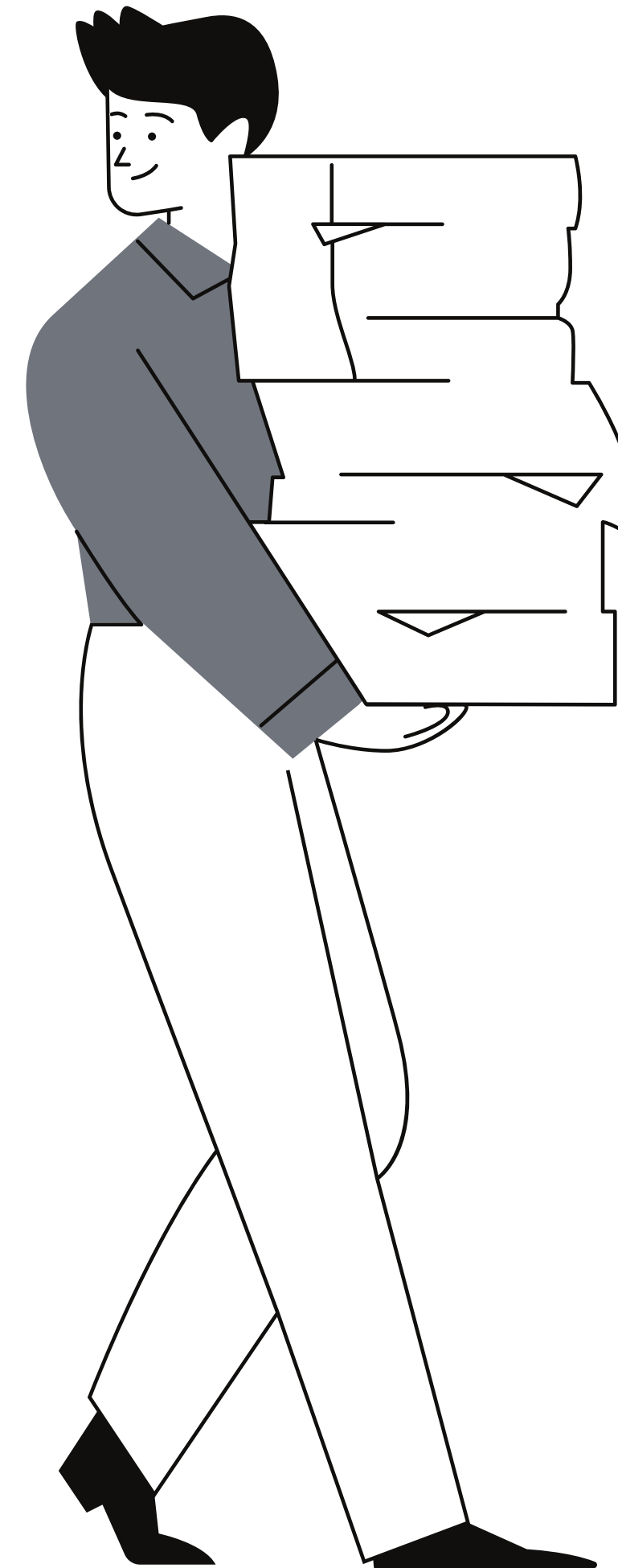
Silvana Armas
Guillermo Rivera
Alejandra Salinas



Introducción

Huawei es un proveedor de infraestructura de tecnologías y comunicaciones y dispositivos inteligentes.

La empresa está comprometida con llevar la tecnología digital a todos para lograr un mundo inteligente y totalmente conectado.



¿Qué representa esto?

Negocio

Preciso control de todos los puntos de venta y un perfecto abastecimiento de los productos demandados.



Ciencia de Datos

Una cantidad abrumadora de datos, considerando el número de sucursales alrededor del país y toda la información que éstas producen.



Desarrollo

- 1 Comprensión del Negocio
- 2 Comprensión de los Datos
- 3 Preparación de Datos
- 4 Modelado
- 5 Evaluación

Comprensión del Negocio

Huawei en México

La firma llegó en 2002 en el sector de las telecomunicaciones, en 2012 incursionó en el área de los smartphones y hasta 2018 la empresa habría vendido 9.39 millones de dispositivos.

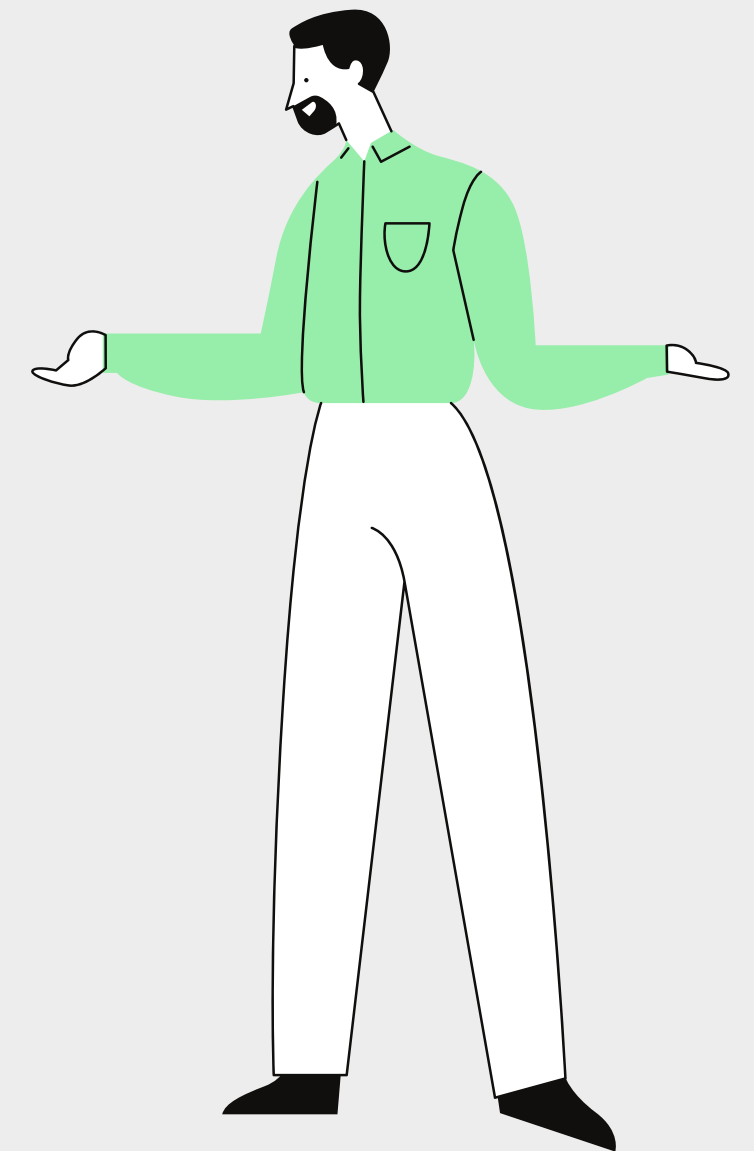


En la primera mitad de 2019, la marca alcanzó una participación de mercado de 12.1% con presencia en todas las gamas (alta, media y baja).

Comprensión de los Datos

Base de datos de 14 variables y 369,617 observaciones.

- | | | | |
|---|---------------------------|----|-----------------|
| 1 | Punto de venta. | 8 | Gama. |
| 2 | Fecha. | 9 | Costo promedio. |
| 3 | Mes. | 10 | Zona. |
| 4 | Año. | 11 | Estado. |
| 5 | Número de ventas. | 12 | Ciudad. |
| 6 | Stock keeping unit (sku). | 13 | Latitud. |
| 7 | Marca. | 14 | Longitud. |

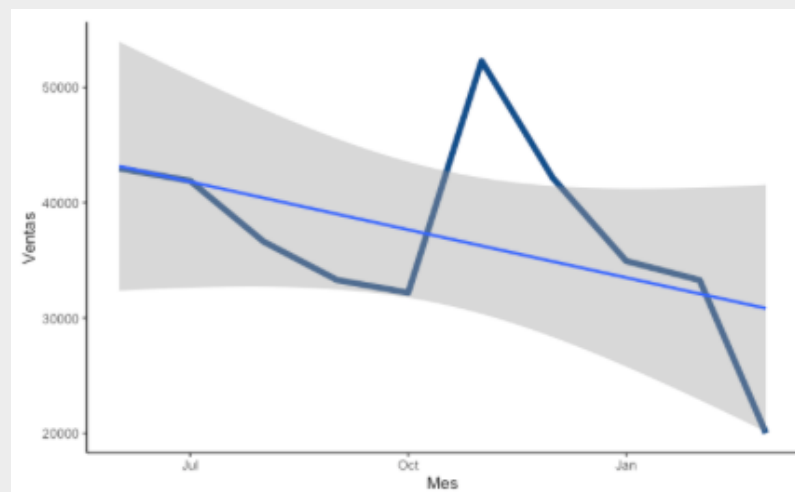


Limpieza de Datos

Crucial de los proyectos de ciencia de datos. La calidad de los datos, su posterior análisis y uso dependen de esto.

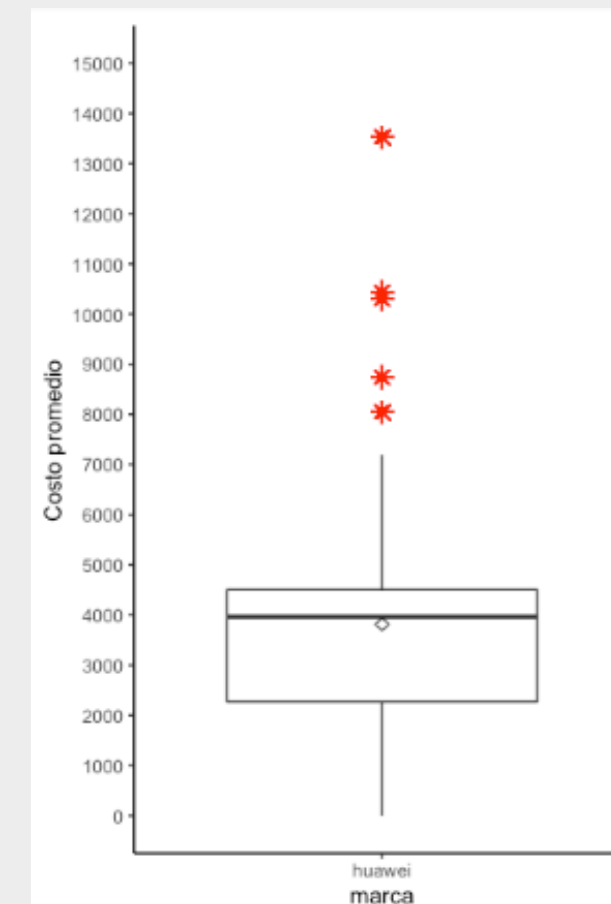
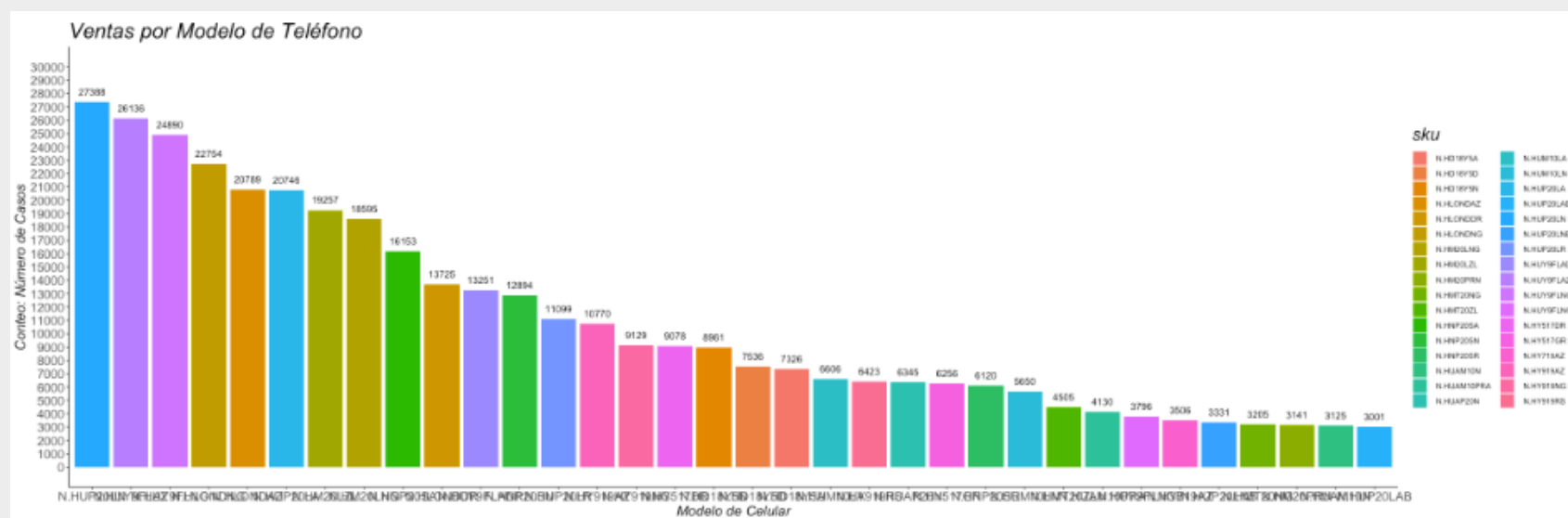
Análisis Exploratorio

Preparación de los Datos

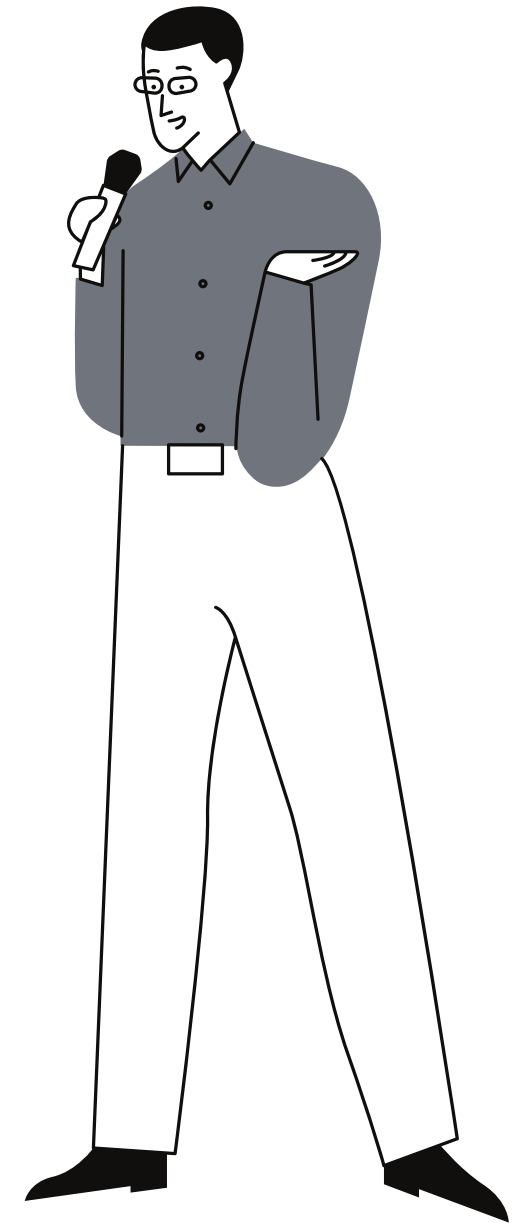


Ventas por mes.

Ventas por modelo.



Costo promedio.



Ing. de características

Consiste en crear características adicionales para un conjunto de datos, con la finalidad de proporcionar información que ayude a diferenciar mejor sus patrones.



1

índices para un mejor manejo de variables cualitativas.

2

Left join, agrupación de ventas y merge.

3

Generar variable respuesta.

Dataset con 24 variables y 645,659 registros.

Modelado

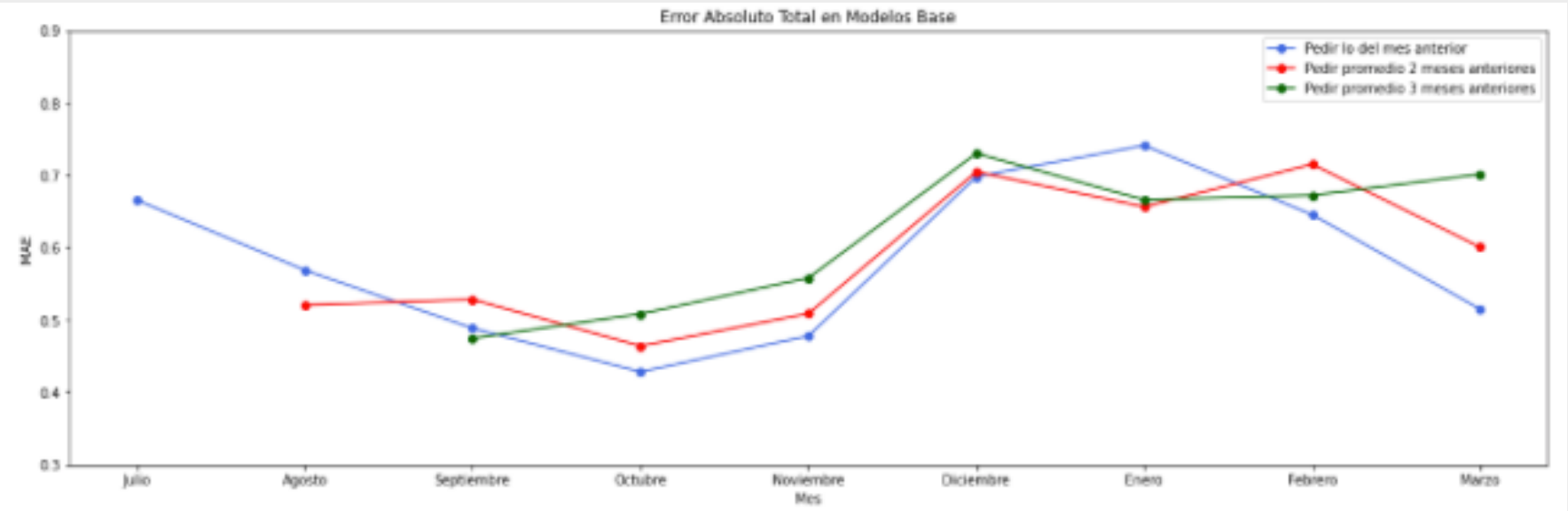
Promedios Móviles

Predicción de una situación futura, basada en datos pasados. Ofrecen un vistazo más aterrizado sobre lo más probable que puede ocurrir.

★ Simples.

★ Ponderados.

★ Dobles.



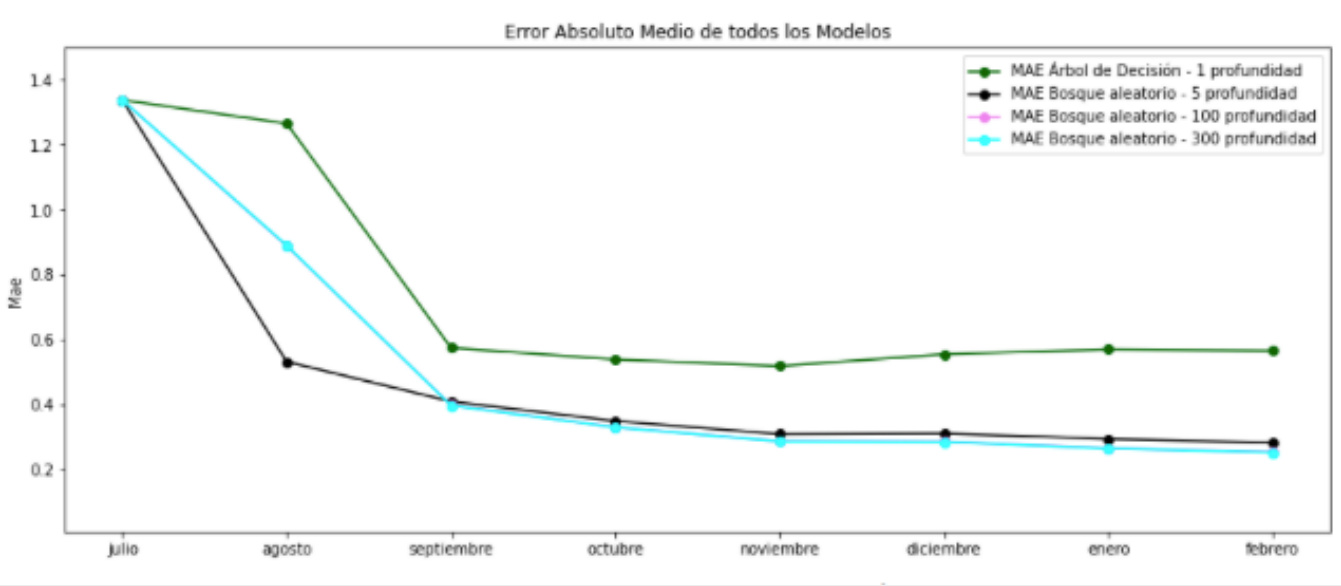
Comparación de MAEs entre promedios móviles.

Modelos Aprendizaje de Máquina

Algoritmos de aprendizaje supervisado.

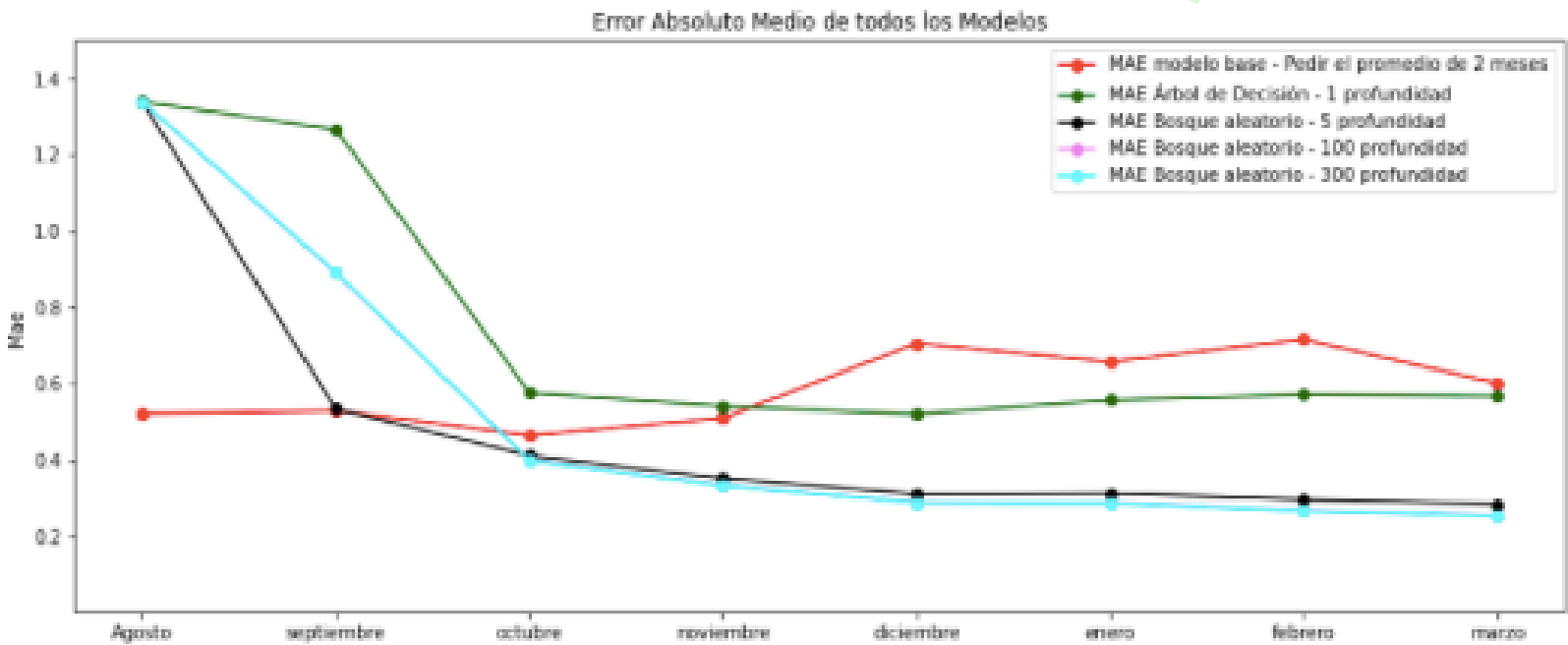
■ Árboles de Decisión.
■ Profundidad de 1.

■ Random Forest.
■ Profundidad de 5, 100, 300.



Comparación de MAEs entre Árbol y Forests.

Evaluación



Comparación de MAEs entre todos los modelos.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
mes	5	0.039115	0.007823	525.84	0.000
profundidad	2	0.002091	0.001045	70.27	0.000
Error	10	0.000149	0.000015		
Total	17	0.041354			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.0038571	99.64%	99.39%	98.83%

ANOVA.

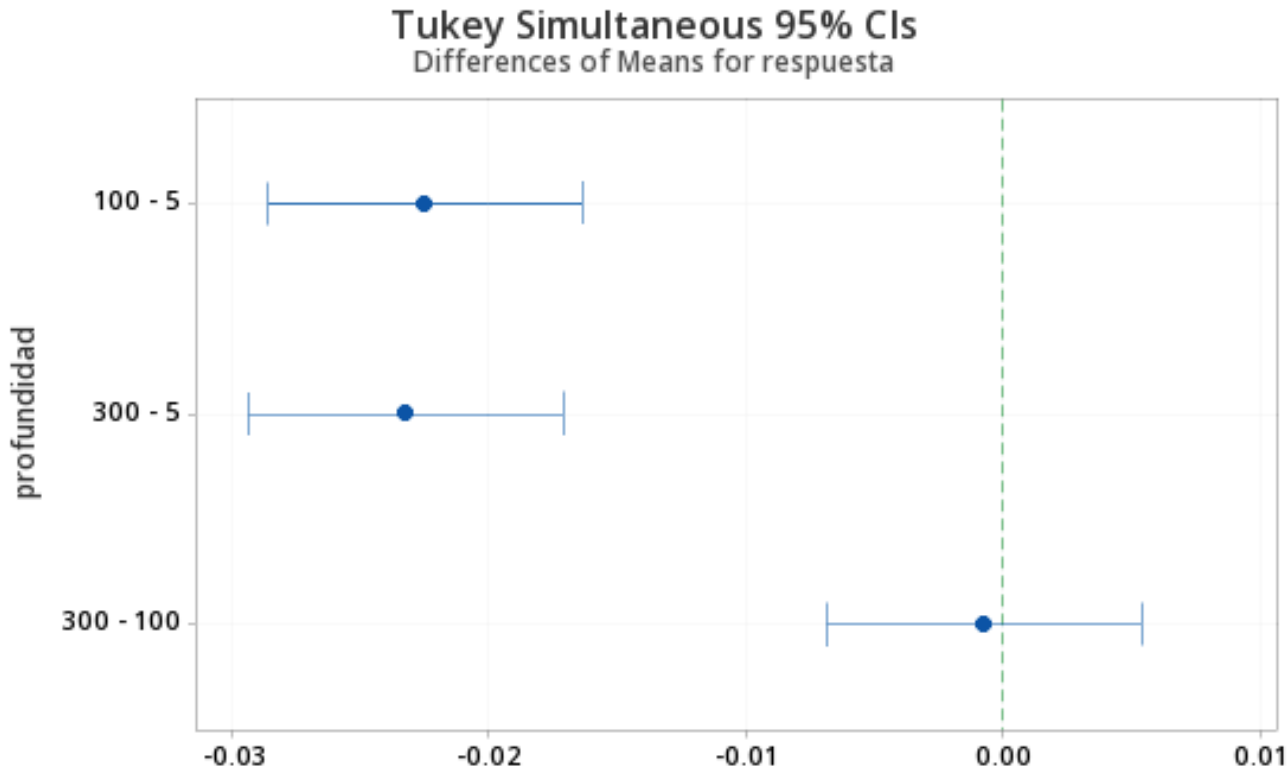
Tukey Pairwise Comparisons: profundidad

Grouping Information Using the Tukey Method and 95% Confidence

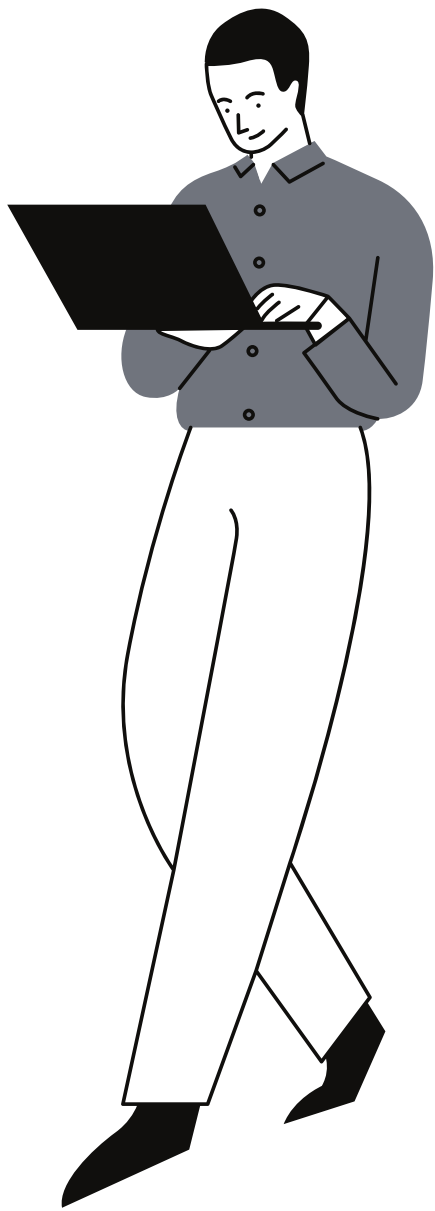
profundidad	N	Mean	Grouping
5	6	0.326696	A
100	6	0.304213	B
300	6	0.303473	B

Means that do not share a letter are significantly different.

Prueba de Tukey.

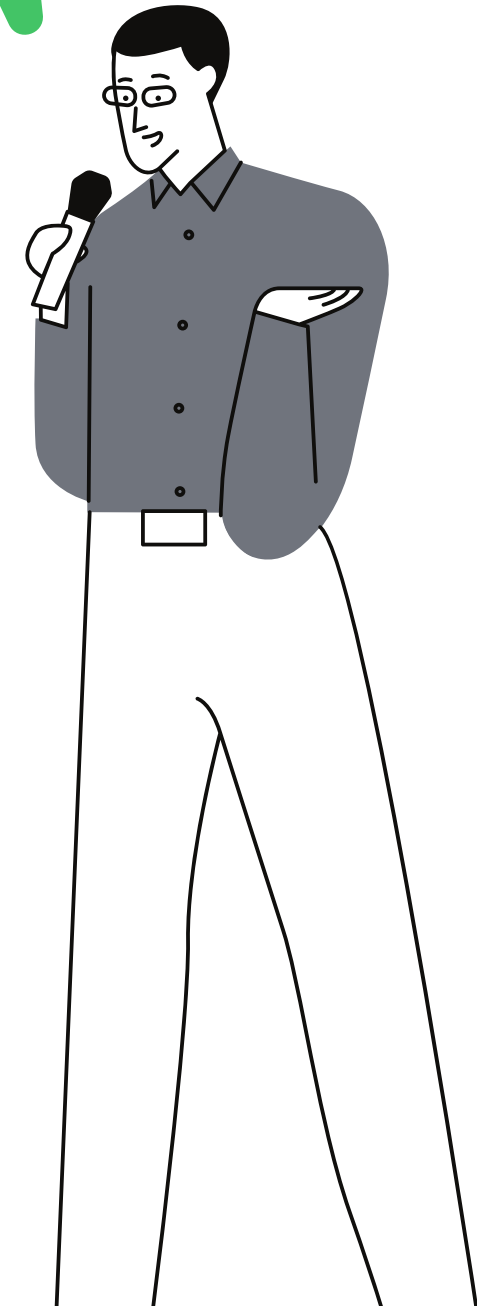


Gráfica de Tukey.



Nuestra
recomendación?

Random
Forest 300!



¡Gracias!



Bonito día.