

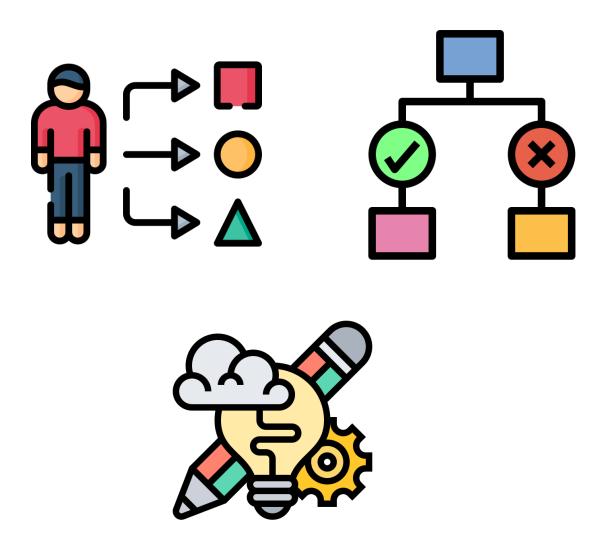
Challenge de Mercado Libre

Carlos Andrés Másmela Pinilla



¿Qué problema se quiso resolver?

Insights clave que puedan mejorar la detección de productos falsificados





Supuestos a partir de la información entregada





Preparación de los datos



Preparación de los datos

- Los datos no contaban ni con duplicados ni con valores faltantes
- Solo fue modificado la parte de texto para poder realizar distintos análisis



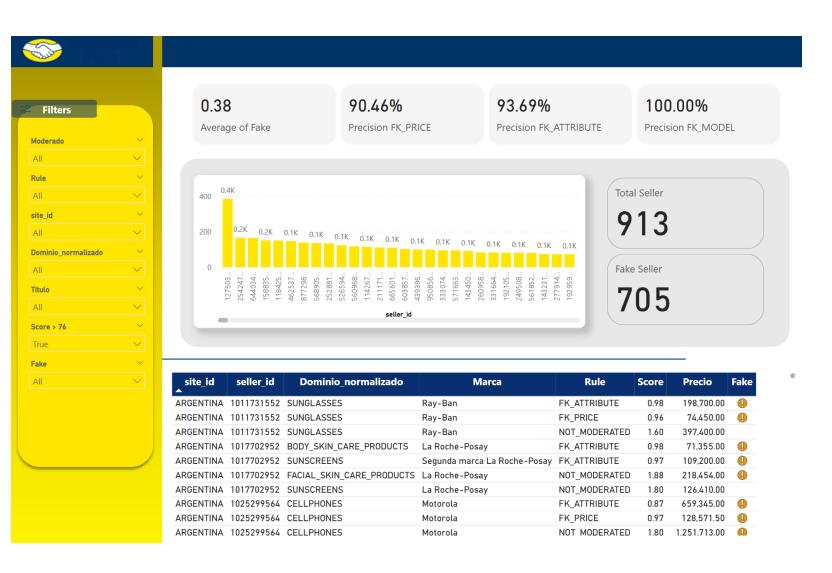




Análisis Exploratorio de los Datos (EDA)



Antes de empezar...



La información y análisis que serán suministrados

fuentes:

proviene de dos

- Power BI
- Python



Análisis de Falsos y Modelos



/ Importante

- Se observó que solamente se presentan fakes con un Score superior al 76%
- FK_PRICE,FK_ATTRIBUTEy FK_MODELno varían



Revisión Vendedor, Dominio y Marca





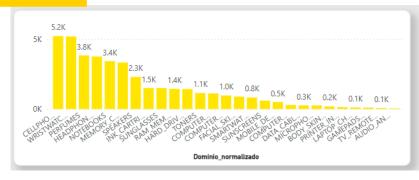
Total Seller

1.138K

Fake Seller

705

Dominio



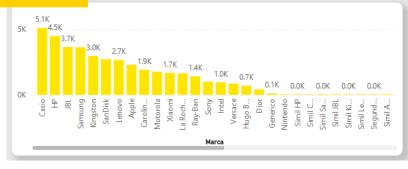
Count of Dominio_normalizado

27

Fake Domain

27

Marca



Marca **53**

Fake Marca

53

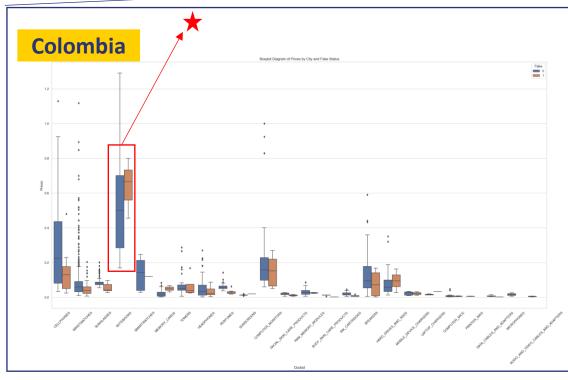
<u>/ Importante</u>

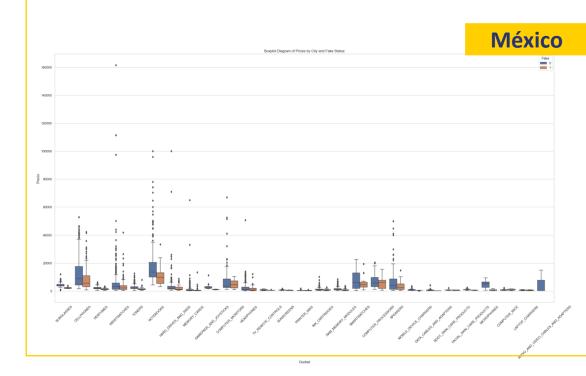
- 76 se puede
 observar que
 225 Sellers no
 corren peligro
 de ser
 clasificados
 como Fake.
- Sellers,
 Dominios y
 Marcas
 cambian
 dependiendo
 el país



Revisión Precio





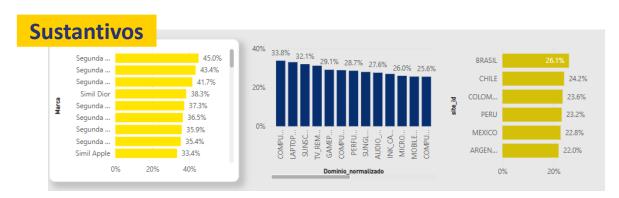


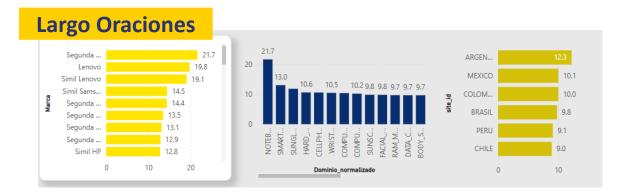


Revisión Léxica









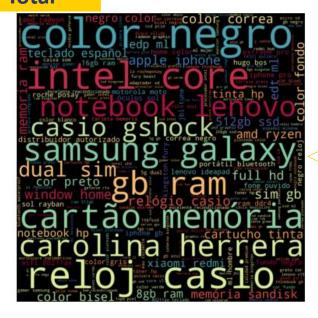
/ Importante

Los análisis cambian dependiendo los filtros que sean colocados

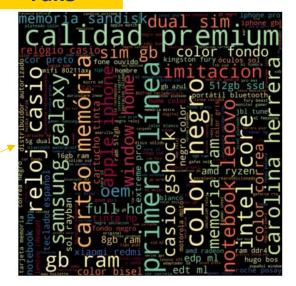


Revisión Léxica

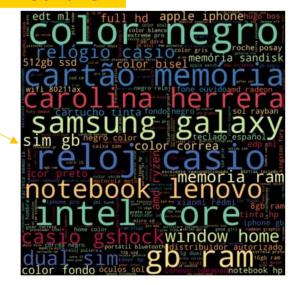
Total



Fake



Not Fake





Análisis



Análisis Inicial

Consideraciones

- FK_MODELO sólo aparece cuando SCORE es
 1
- Sólo se presenta problemas para identificar si es FAKE o no con SCORE superior a 76%, menor o igual siempre es Real





Análisis Léxico – FK_ATTRIBUTE

Título



Marca

- Simil
- Segunda Marca

- Hay 721 filas con Titulo con las características indicadas
- Hay 190 filas con Marca con las características indicadas
- Hay 0 filas con las dos al mismo tiempo
- Se han identificado 911 en total de 1267, lo que corresponde al 71.90% de FK ATTRIBUTE
- Se observa que hay ROLLOUT cuando el modelo predice algo que no se encuentra con las características antes mencionadas



Análisis Léxico – FK_TEST1

Título



/ Importante

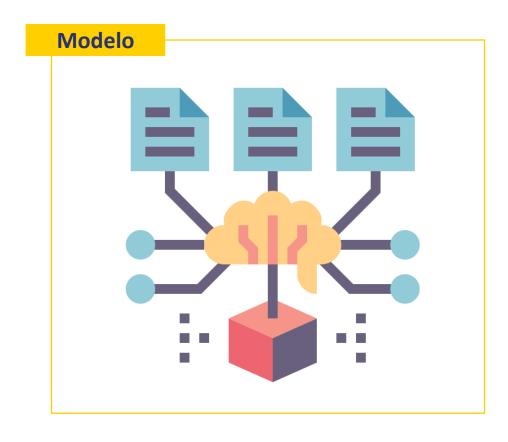
- Siempre que se encuentra Calidad Premium en FK_TEST1 dicho modelo lo identifica como falso al igual que siempre es falso en FAKE
- Se encontraron 146 filas con dicha característica



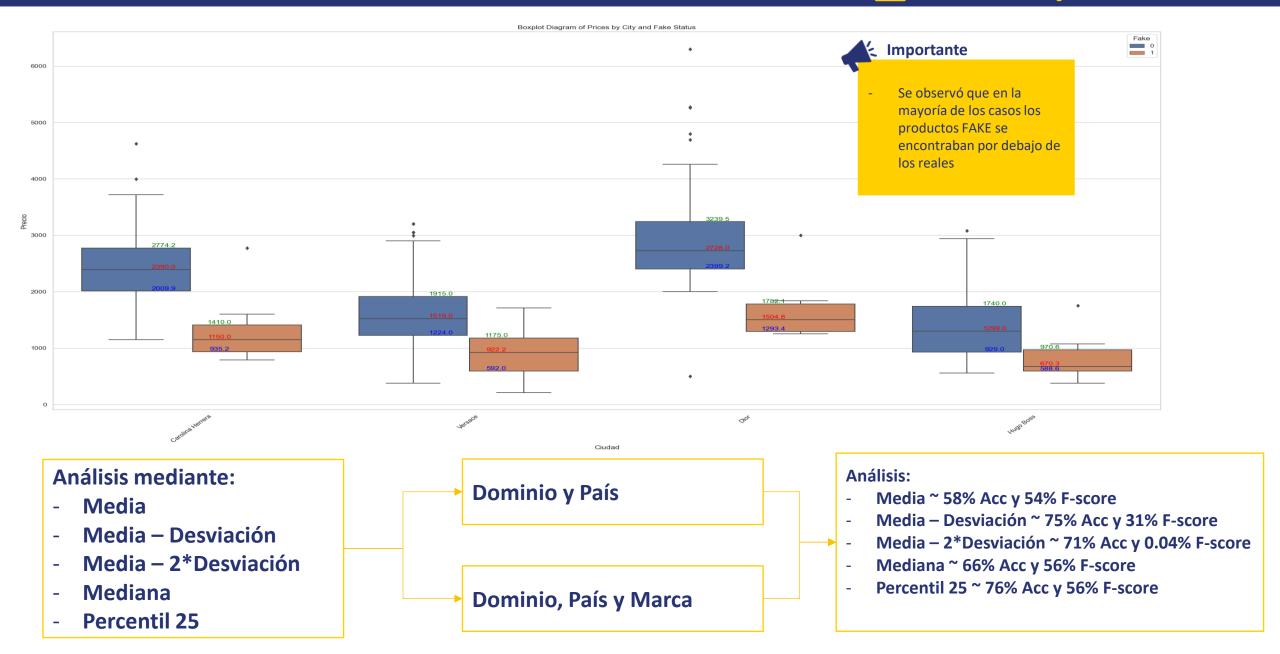
Análisis Numérico – FK_PRICE

Se realizaron dos métodos para lograr entender el precio



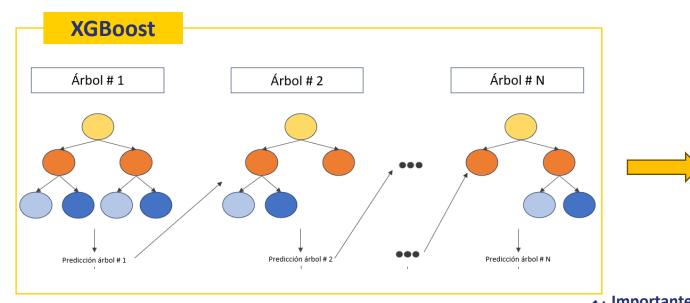


Análisis Numérico – Estadístico – FK_PRICE y TEST2



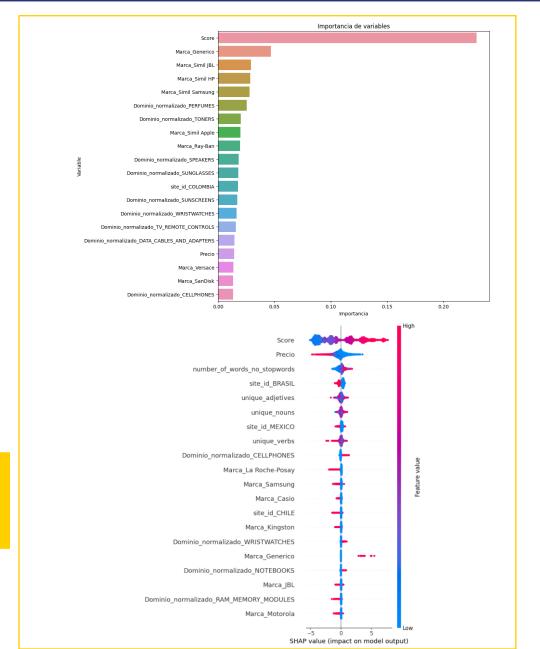


Análisis Numérico – Exploración adicional



Importante

Se emplearon las variables enviadas y algunas de las generadas que fueron explicadas





Preguntas Finales



Solución a Preguntas Adicionales



Es importante identificar y eliminar productos falsificados porque:

- Aumenta credibilidad en la página para comprar
- Genera seguridad al momento de comprar
- Aumenta la cantidad de ventas general
- Los vendedores de productos reales no se ven afectados por productos fake ya que usualmente son mas económicos

2

Emplearía dos tácticas que no fueron presentadas en el caso:

- Análisis de imágenes



- Números de serie del producto o Certificados



3

Creando nuevas cuentas es la forma con las cual los vendedores.

Forma de identificación: Análisis por Grafos

