



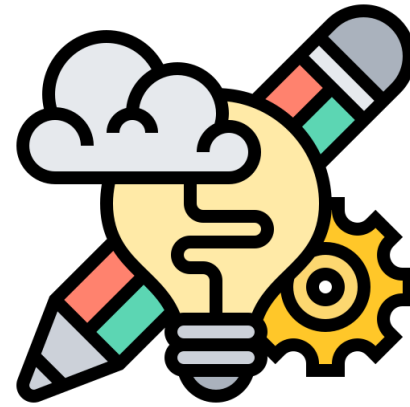
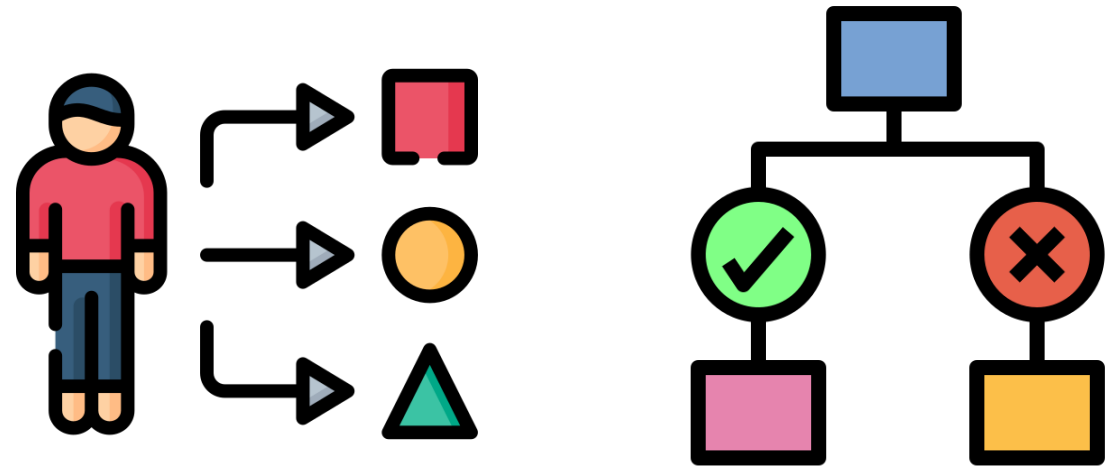
Challenge de Mercado Libre

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...



Filters

Moderado

All

Rule

All

site_id

All

Dominio_normalizado

All

Título

All

Score > 76

True

Fake

All

0.38

Average of Fake

90.46%

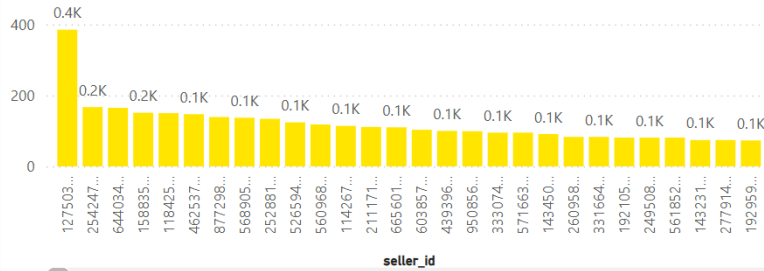
Precision FK_PRICE

93.69%

Precision FK_ATTRIBUTE

100.00%

Precision FK_MODEL



Total Seller

913

Fake Seller

705

site_id	seller_id	Dominio_normalizado	Marca	Rule	Score	Precio	Fake
ARGENTINA	1011731552	SUNGLASSES	Ray-Ban	FK_ATTRIBUTE	0.98	198,700.00	!
ARGENTINA	1011731552	SUNGLASSES	Ray-Ban	FK_PRICE	0.96	74,450.00	!
ARGENTINA	1011731552	SUNGLASSES	Ray-Ban	NOT_Moderated	1.60	397,400.00	
ARGENTINA	1017702952	BODY_SKIN_CARE_PRODUCTS	La Roche-Posay	FK_ATTRIBUTE	0.98	71,355.00	!
ARGENTINA	1017702952	SUNSCREENS	Segunda marca La Roche-Posay	FK_ATTRIBUTE	0.97	109,200.00	!
ARGENTINA	1017702952	FACIAL_SKIN_CARE_PRODUCTS	La Roche-Posay	NOT_Moderated	1.88	218,454.00	!
ARGENTINA	1017702952	SUNSCREENS	La Roche-Posay	NOT_Moderated	1.80	126,410.00	
ARGENTINA	1025299564	CELLPHONES	Motorola	FK_ATTRIBUTE	0.87	659,345.00	!
ARGENTINA	1025299564	CELLPHONES	Motorola	FK_PRICE	0.97	128,571.50	!
ARGENTINA	1025299564	CELLPHONES	Motorola	NOT_Moderated	1.80	1,251,713.00	!



Importante

La información y análisis que serán suministrados proviene de dos fuentes:

- Power BI
- Python



Análisis de Falsos y Modelos

Global

0.11

Average of Fake

90.46%

Precision FK_PRICE

93.69%

Precision FK_ATTRIBUTE

100.00%

Precision FK_MODEL

11% de Falsos tomando el dataset total, incrementa a un 38% con Score > 76%

Argentina

0.11

Average of Fake

90.12%

Precision FK_PRICE

99.21%

Precision FK_ATTRIBUTE

100.00%

Precision FK_MODEL

11% de Falsos tomando el dataset total, incrementa a un 40% con Score > 76%

Colombia

0.10

Average of Fake

82.93%

Precision FK_PRICE

98.31%

Precision FK_ATTRIBUTE

100.00%

Precision FK_MODEL

10% de Falsos tomando el dataset total, incrementa a un 34% con Score > 76%

Brasil

0.11

Average of Fake

92.28%

Precision FK_PRICE

95.27%

Precision FK_ATTRIBUTE

100.00%

Precision FK_MODEL

11% de Falsos tomando el dataset total, incrementa a un 39% con Score > 76%



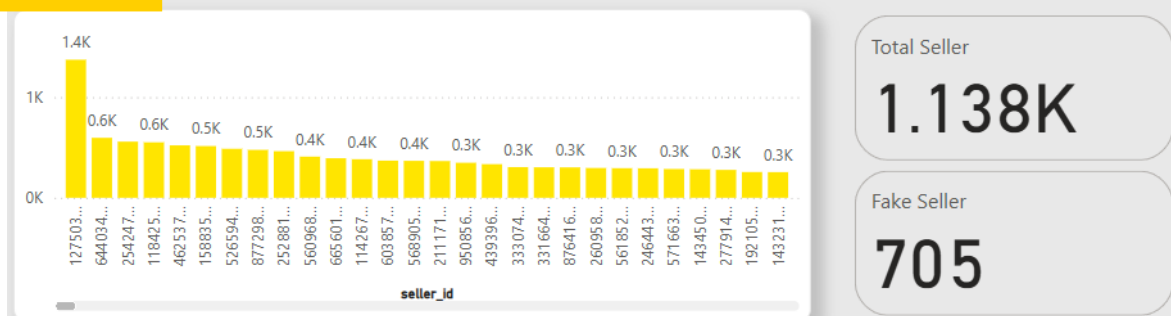
Importante

- Se observó que solamente se presentan fakes con un Score superior al 76%
- FK_PRICE, FK_ATTRIBUTE y FK_MODEL no varían

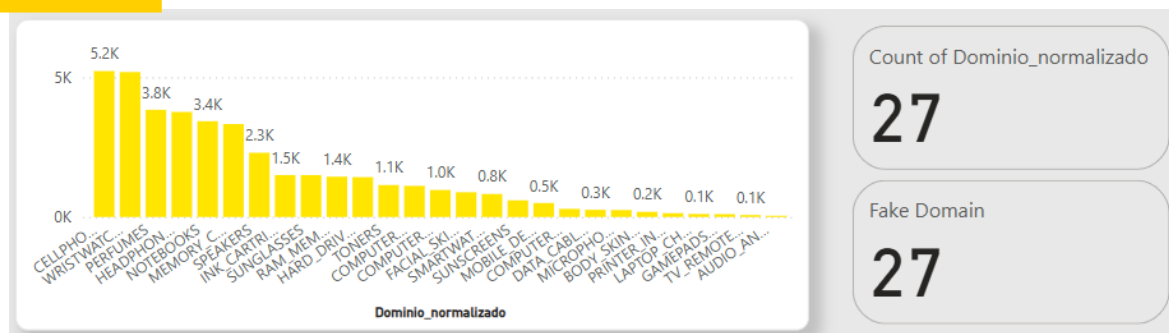


Revisión Vendedor, Dominio y Marca

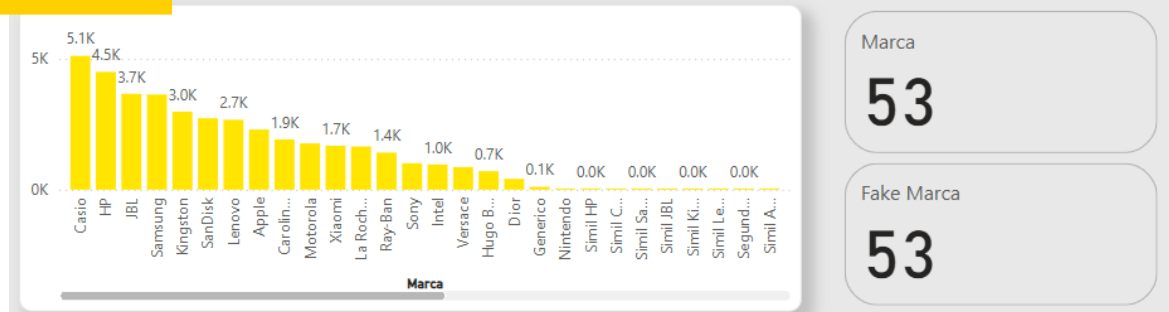
Seller



Dominio



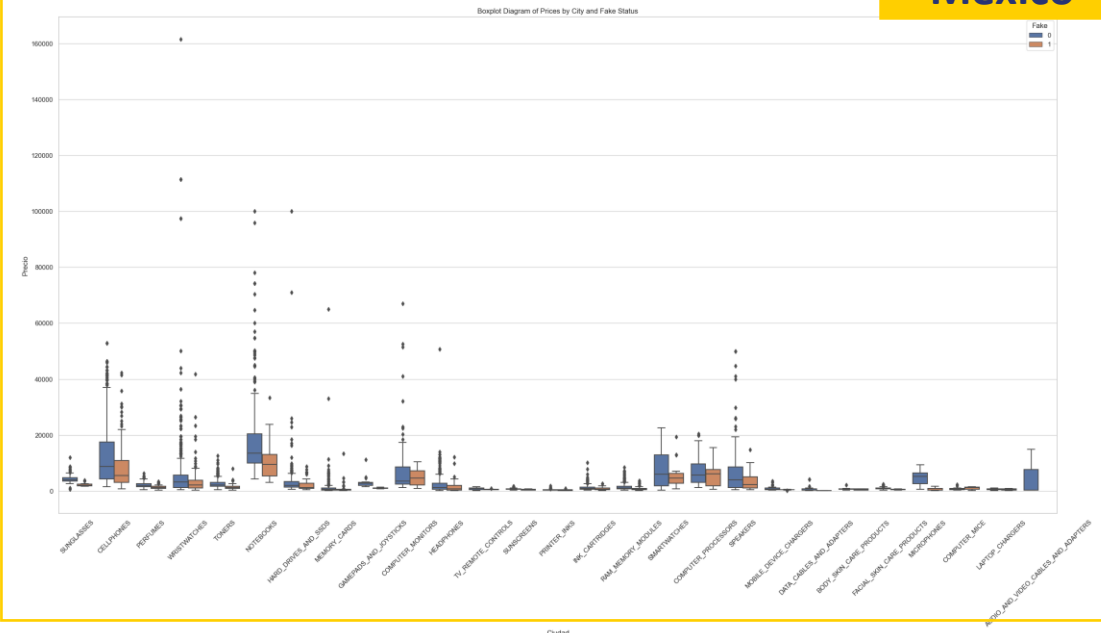
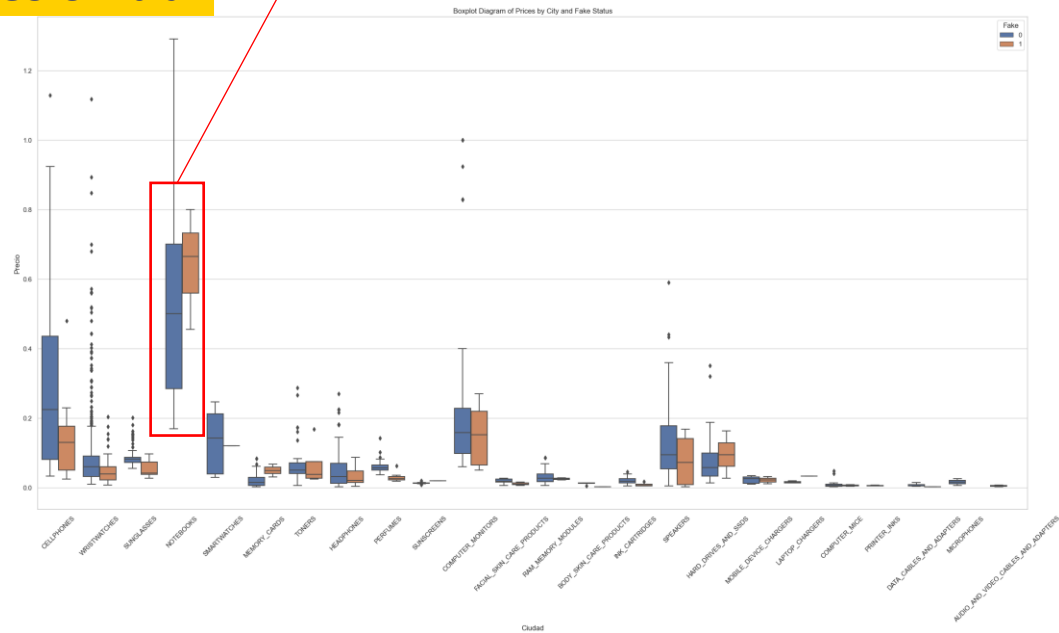
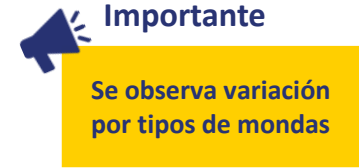
Marca



Importante

Con un Score > 76 se puede observar que 225 Sellers no corren peligro de ser clasificados como Fake.

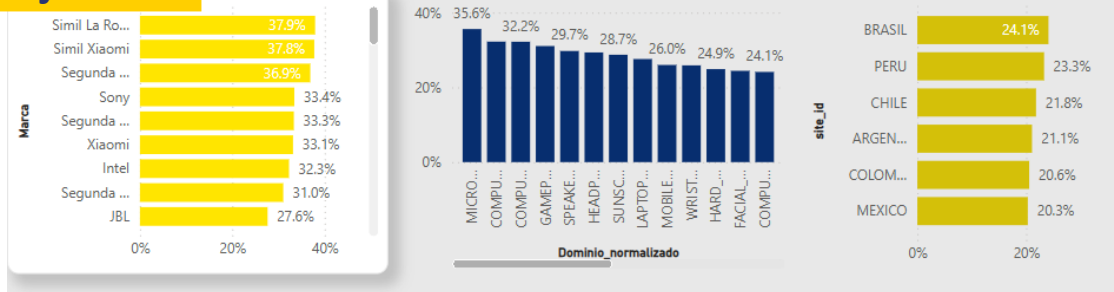
- Sellers, Dominios y Marcas cambian dependiendo el país



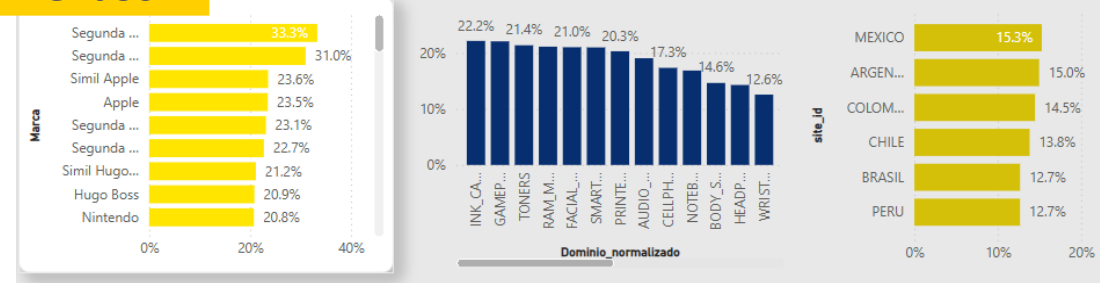


Revisión Léxica

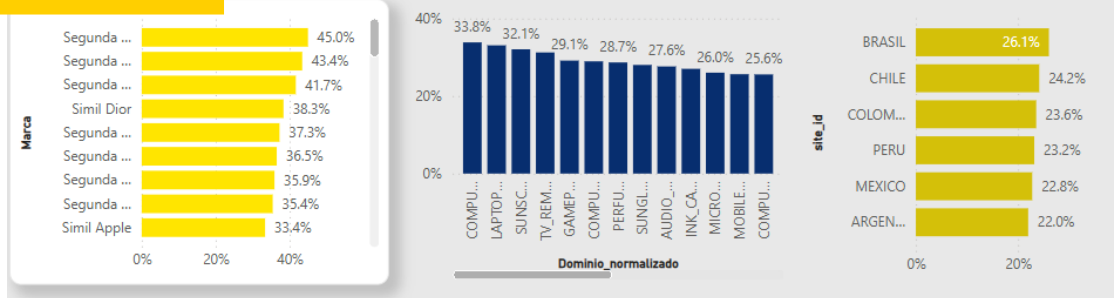
Adjetivos



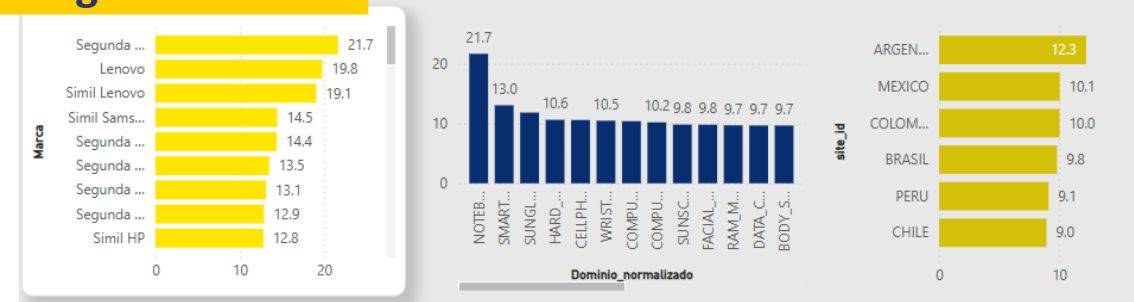
Verbos



Sustantivos



Largo Oraciones



Importante

Los análisis cambian dependiendo los filtros que sean colocados



Total

[illegible]



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



[illegible]

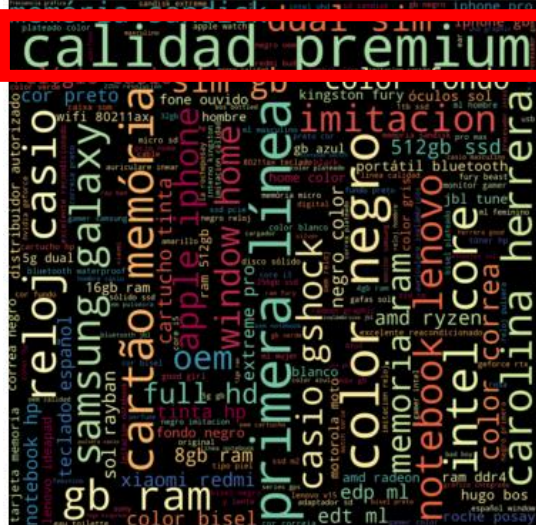
- **Simil**
- **Segunda Marca**

- 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



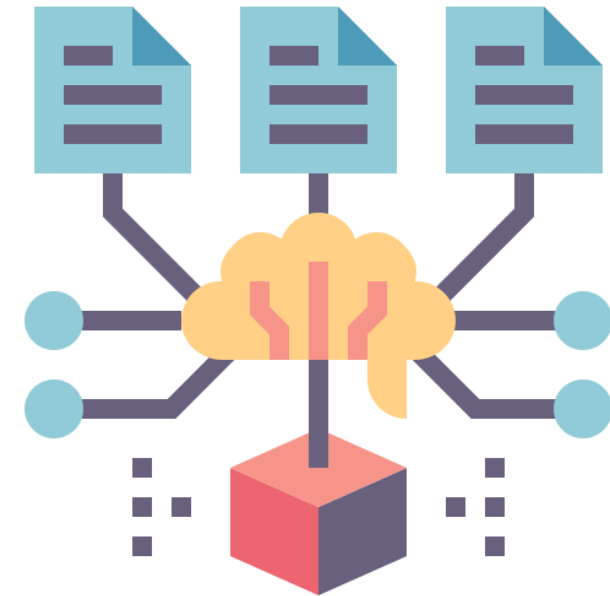
Análisis Numérico – FK_PRICE

Se realizaron dos métodos para lograr entender el precio

Estadística

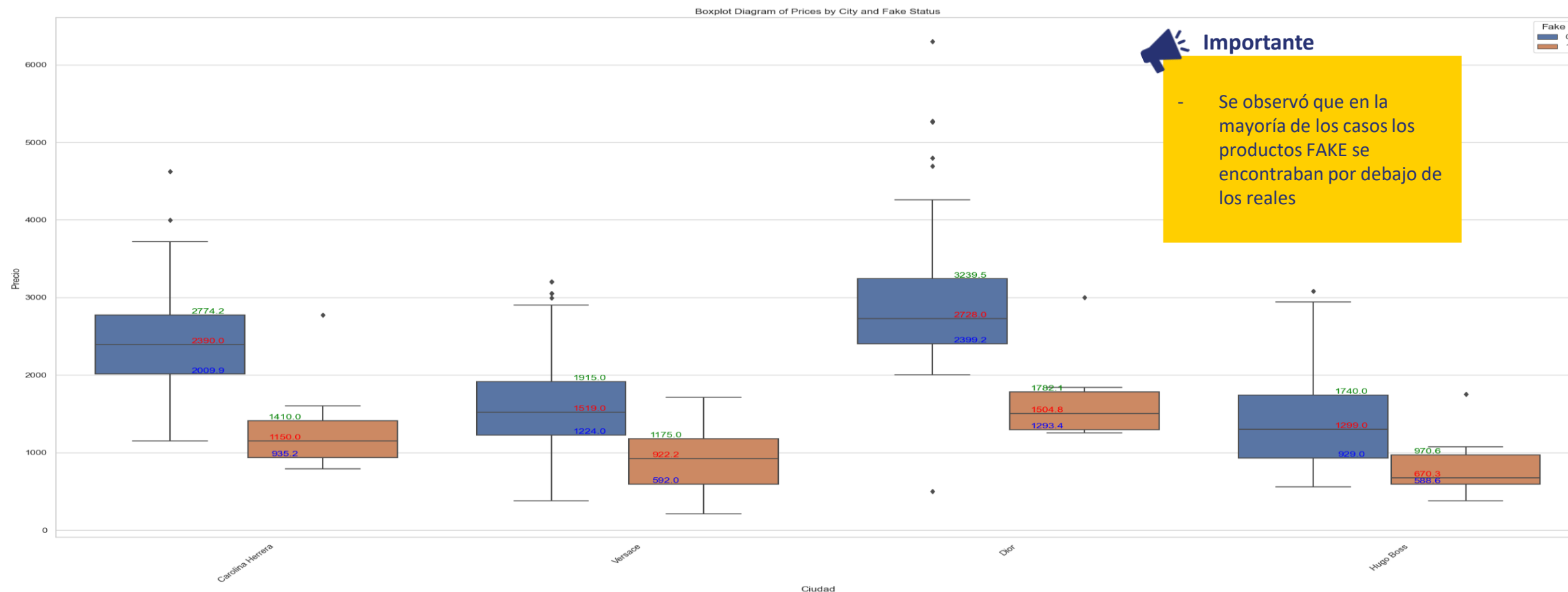


Modelo





Análisis Numérico – Estadístico – FK_PRICE y TEST2



Importante

- Se observó que en la mayoría de los casos los productos FAKE se encontraban por debajo de los reales

Análisis mediante:

- Media
- Media – Desviación
- Media – 2*Desviación
- Mediana
- Percentil 25

Dominio y País

Dominio, País y Marca

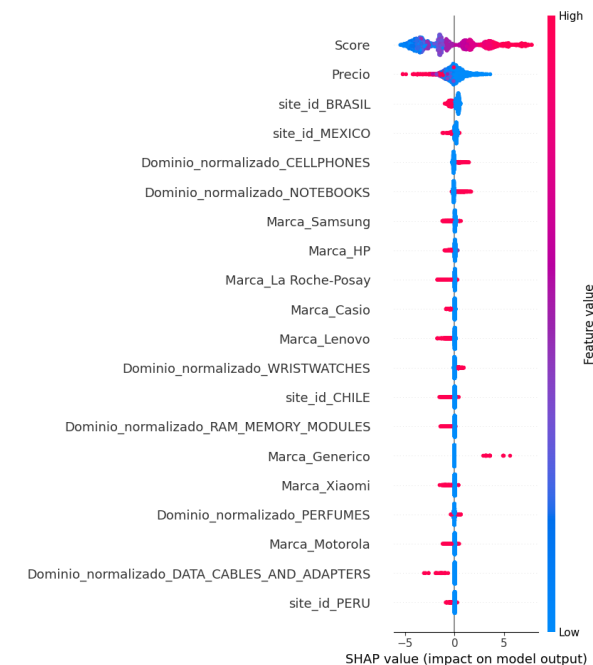
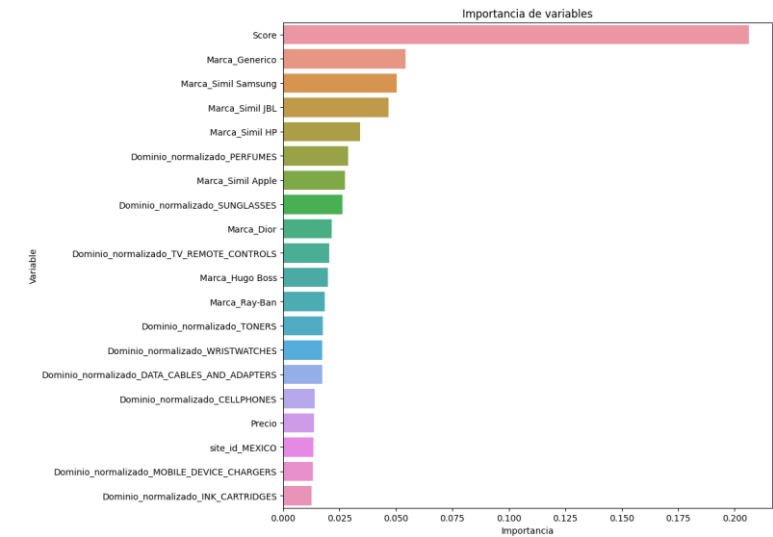
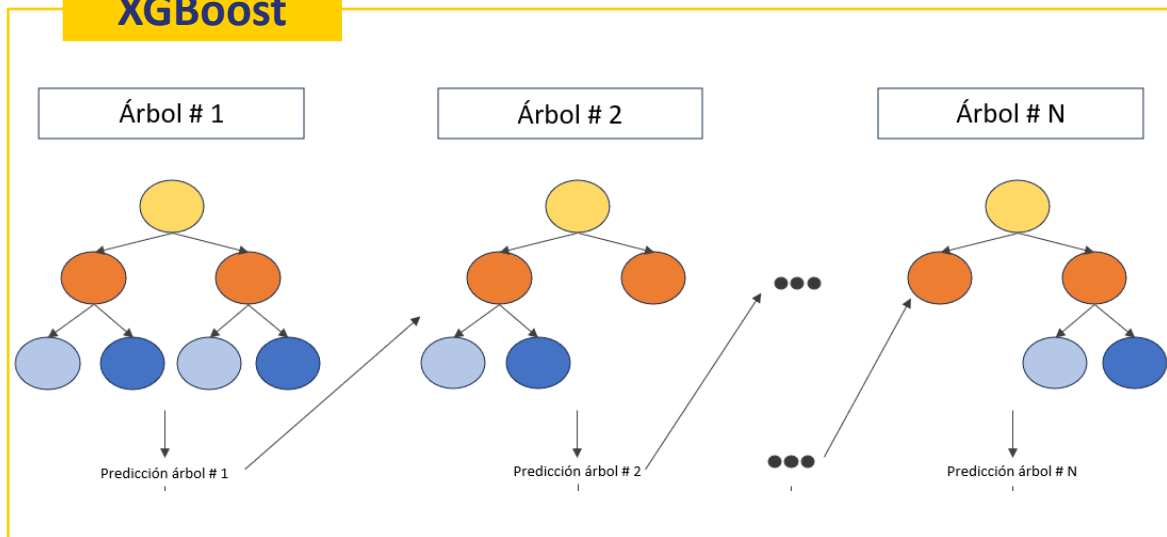
Análisis:

- Media ~ 58% Acc y 54% F-score
- Media – Desviación ~ 75% Acc y 31% F-score
- Media – 2*Desviación ~ 71% Acc y 0.04% F-score
- Mediana ~ 66% Acc y 56% F-score
- Percentil 25 ~ 76% Acc y 56% F-score



Análisis Numérico – Exploración adicional

XGBoost





Preguntas Finales



Solución a Preguntas Adicionales

1

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 la cual los vendedores.

**Forma de identificación:
Análisis por Grafos**

