

# Challenge Técnico - Operaciones & Analytics - Moderaciones IT

**MERCADO LIBRE**

**Elaborado por: Santiago Ramírez Chocontá**

**Abril 2025**

## Tabla de contenido

1.	Descripción general .....	2
2.	Tecnologías empleadas .....	2
3.	Premisas .....	3
4.	Análisis y resultados .....	4
4.1.	Productos moderados y no moderados por país .....	4
4.2.	Top 3 marcas moderadas .....	4
4.3.	Top 5 vendedores .....	5
4.4.	Reglas y puntaje asignado por un modelo ML .....	6
5.	Estrategias que pueden ayudar en la identificación de productos falsificados ....	8
5.1.	Modelo de detección de anomalías – DBSCAN .....	8
5.2.	Modelo de detección de anomalías – Autoencoder .....	8
5.3.	Modelo de análisis de sentimientos - NLP .....	9
5.4.	Colaboración con marcas .....	9
5.5.	Campañas al consumidor .....	9
6.	Impacto en el Marketplace .....	10
	Conclusiones .....	11

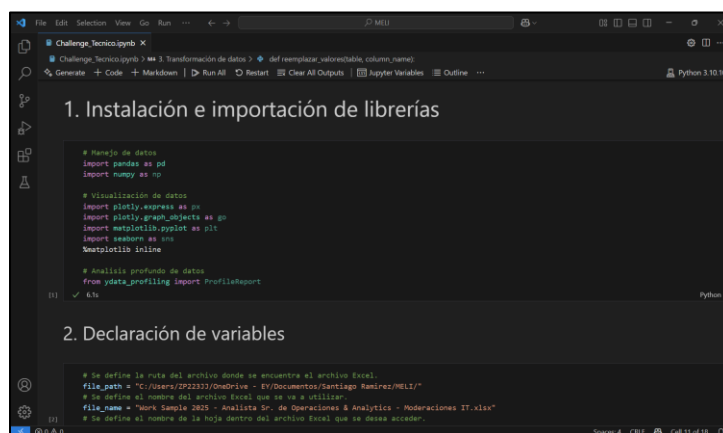
## 1. Descripción general

Este documento presenta los resultados del análisis realizado sobre el conjunto de datos proporcionado, relacionado con la moderación de productos en MercadoLibre. Su objetivo es extraer información clave que mejore la detección de productos falsificados.

## 2. Tecnologías empleadas

Para el desarrollo de este desafío, se utilizaron los siguientes programas:

- Visual Studio Code: Se empleó Python (versión 3.10.10) para analizar los datos, identificando tendencias, anomalías y concentraciones inusuales.



```

1. Instalación e importación de librerías

# Manejo de datos
import pandas as pd
import numpy as np

# Visualización de datos
import matplotlib.pyplot as plt
import matplotlib.pyplot as plt
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

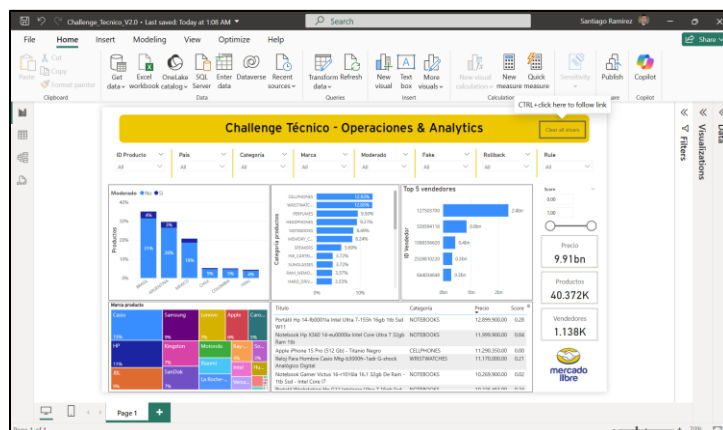
# Análisis profundo de datos
from ydata_profiling import ProfileReport

2. Declaración de variables

# Se define la ruta del archivo donde se encuentra el archivo Excel.
file_path = "C:/Users/CP2333/OneDrive - RY/Documents/Santiago Ramirez/MLLI/"
# Se define el nombre del archivo Excel que se va a utilizar.
file_name = "Work Sample 2025 - Analista Sr. de Operaciones & Analytics - Moderaciones IT.xlsx"
# Se define el nombre de la hoja dentro del archivo Excel que se desea acceder.

```

- Power BI: Esta herramienta facilita la presentación visual y clara de los resultados, permitiendo mostrar la situación actual de los productos y detectar tendencias, anomalías o señales de alerta.



### 3. Premisas

Según el conjunto de datos proporcionado y el análisis realizado, se establecen las siguientes premisas para el desarrollo de este desafío:

- a) Campo "Rule": En la base de datos hay cuatro categorías: FK\_PRICE, FK\_ATTRIBUTE, FK\_MODEL y campos vacíos.
  - FK\_PRICE: Indica si el precio de un producto está por debajo o por encima de ciertos límites, lo que puede llevar a su revisión o rechazo.
  - FK\_ATTRIBUTE: Se refiere a productos que no cumplen con las características o atributos esperados, lo que también puede resultar en revisión o rechazo.
  - FK\_MODEL: Indica si un producto no coincide con las especificaciones del modelo, lo que puede llevar a su revisión o rechazo.
  - Campos nulos: Se refiere a productos que ya han sido revisados y no presentan ninguna moderación, ya que se encuentran dentro de los límites establecidos.
- b) Campo "Score": Es el puntaje asignado por un modelo ML y se interpreta de la siguiente manera:

Es un indicador del riesgo de si un producto no coincide con las especificaciones del modelo. Un score alto indica que se cree que el producto es falso. Un puntaje bajo (33%) indicaría que se tiene pocos indicios de que el producto es falso, por lo cual se considera auténtico.
- c) Campo "Total": Aunque este campo no está definido en el diccionario de datos compartido, se ha observado que todos los registros presentan un valor de 1. Por lo tanto, se omite del análisis y se elimina del conjunto de datos para los análisis a realizar.
- d) Campo "Precio": Se interpreta que los valores observados corresponden a USD.

## 4. Análisis y resultados

El análisis del conjunto de datos proporcionado revela los siguientes números:



6 países



1,138 vendedores



40,372 productos



4,278 productos moderados



53 marcas



27 categorías

### 4.1. Productos moderados y no moderados por país

El top 3 de los países con mayor número de productos está encabezado por Brasil, con 14,142 productos, seguido de Argentina con 11,938 y México con 8,378:

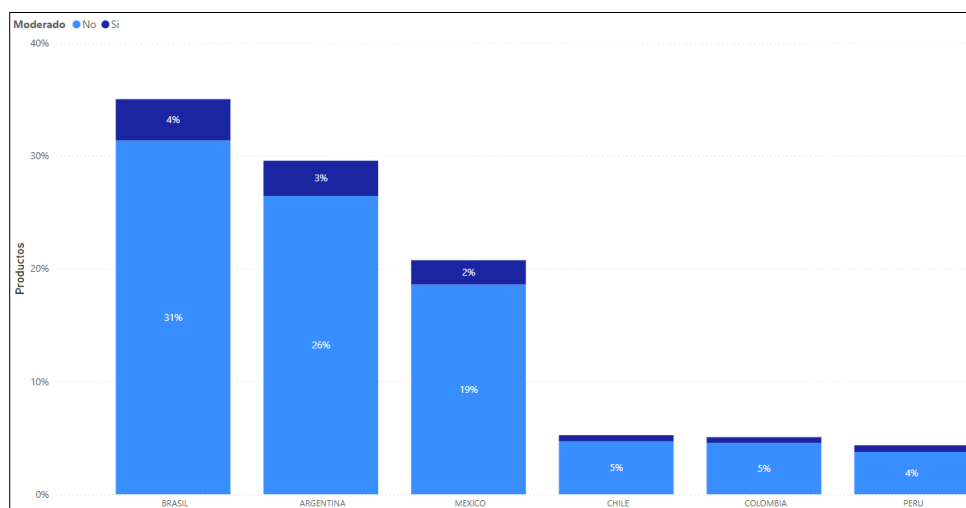


Ilustración 1. Comparación de número de productos no moderados y moderados según el país.

### 4.2. Top 3 marcas moderadas

El valor total de los productos moderados en todos los países alcanza los USD 752 millones. Las principales marcas moderadas por país son:

BRA: 1.39M	ARG: 524.89M	MEX: 3.35M	CHL: 46.34M	COL: 176.51M	PER: 138.84M
<b>SanDisk</b> 17%	<b>hp</b> 16%	<b>hp</b> 14%	<b>CASIO</b> 17%	<b>CASIO</b> 42%	<b>CASIO</b> 19%




<b>JBL</b> 13%	<b>Lenovo</b> 14%	<b>CASIO</b> 12%	 12%	 10%	<b>JBL</b> 14%
<b>CASIO</b> 10%	<b>SAMSUNG</b> 11%	<b>SAMSUNG</b> 11%	 10%	<b>Ray-Ban</b> 8%	<b>SONY</b> 12%

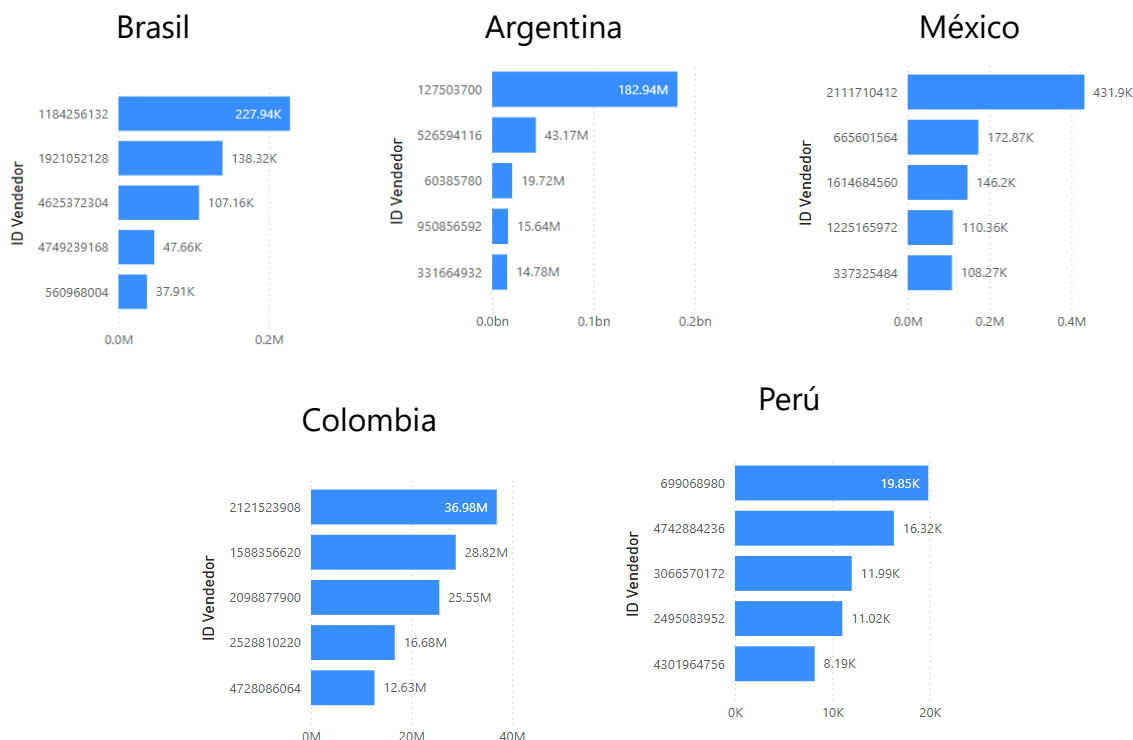
Tabla 1. Precio de productos moderados por país por marca.

Así, Argentina ocupa el primer lugar con USD 524.89 millones en productos moderados, seguida de Colombia en segundo lugar con USD 176.51 millones y Perú en tercer lugar con USD 138.84 millones, también en productos moderados.

Las marcas más destacadas en productos moderados son Casio (11%), HP (10%), Samsung (9%) y JBL (9%), entre otras.

### 4.3. Top 5 vendedores

A continuación, se presenta el acumulado de precios por vendedor para los productos moderados.



Los tres países con los mayores vendedores de productos de precios altos son Argentina, que lidera la lista con USD 182.94 millones, seguido de Colombia con USD 36.98 millones y México con USD 431.9 mil.

Argentina, la marca que más productos ofrece, moderada por el vendedor con el valor más alto, es Lenovo. En Colombia y México, la marca predominante es Apple.



#### 4.4. Reglas y puntaje asignado por un modelo ML

Según los registros del conjunto de datos proporcionados, se ha identificado que el 89.4% de los productos carecen de una regla asociada y presentan un estado de moderación de "No". Esto indica que estos productos son considerados auténticos.

Value	Count	Frequency (%)
FK_PRICE	2579	6.4%
FK_ATTRIBUTE	1267	3.1%
FK_MODEL	432	1.1%
(Missing)	36094	89.4%

Ilustración 2. Frecuencia de las categorías encontradas en el campo Rule.

En los productos marcados como no moderados, se pueden observar los siguientes valores:

```
np.sort(df[df["Moderado"] == "No"]["Score"].unique())
✓ 0.0s

array([0. , 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1 ,
       0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17, 0.18, 0.19, 0.2 , 0.21,
       0.22, 0.23, 0.24, 0.25, 0.26, 0.27, 0.28, 0.29, 0.3 , 0.31, 0.32,
       0.33, 0.34, 0.35, 0.36, 0.37, 0.38, 0.39, 0.4 , 0.49, 0.5 , 0.51,
       0.52, 0.53, 0.54, 0.55, 0.56, 0.57, 0.58, 0.59, 0.6 , 0.61, 0.62,
       0.63, 0.64, 0.65, 0.66, 0.67, 0.68, 0.69, 0.7 , 0.71, 0.72, 0.73,
       0.74, 0.75, 0.76, 0.77, 0.78, 0.79, 0.8 , 0.81, 0.82, 0.83, 0.84,
       0.85, 0.86, 0.87, 0.88, 0.89, 0.9 , 0.91, 0.92, 0.93, 0.94, 0.95,
       0.96, 0.97, 0.98, 0.99])
```

*Ilustración 3. Listado de valores únicos del score donde el estado de moderación es "No".*

Según la ilustración 3, se puede observar que una considerable cantidad de productos presenta puntajes superiores a 0.7 y que están clasificados como no moderados. Al realizar un análisis más detallado de los puntajes, se evidencia que el 28% de los productos no moderados tiene un puntaje superior a 0.5. Esto sugiere la posible existencia de 10,031 productos falsificados que no han sido detectados.

A continuación, se presentan los valores de puntuación (score) correspondientes a cada una de las reglas: FK\_MODEL, FK\_ATTRIBUTE y FK\_PRICE, en el conjunto de datos analizado.

```
np.sort(df[df["Rule"] == "FK_MODEL"]["Score"].unique())
✓ 0.0s

array([1.])

np.sort(df[df["Rule"] == "FK_ATTRIBUTE"]["Score"].unique())
✓ 0.0s

array([0.79, 0.8 , 0.81, 0.82, 0.83, 0.84, 0.85, 0.86, 0.87, 0.88, 0.89,
       0.9 , 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99])

np.sort(df[df["Rule"] == "FK_PRICE"]["Score"].unique())
✓ 0.0s

array([0.77, 0.78, 0.8 , 0.81, 0.82, 0.83, 0.84, 0.85, 0.86, 0.87, 0.88,
       0.89, 0.9 , 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99])
```

*Ilustración 4. Puntajes que puede tomar cada una de las reglas existentes.*

Se observa que la única regla que alcanza un valor de 1 en el campo de "Score" es la regla FK\_MODEL. Por otro lado, la regla FK\_ATTRIBUTE presenta valores que oscilan entre 0.79 y 0.99, mientras que la regla FK\_PRICE muestra puntuaciones en el rango de 0.77 a 0.99.

FK\_MODEL



#### FK\_ATTRIBUTE



#### FK\_PRICE

En el análisis de las cinco principales reglas, se observa consistentemente la presencia de las mismas categorías: CELLPHONES, WRISTWATCHES, HEADPHONES, MEMORY\_CARDS y PERFUMES. Esto sugiere que estas categorías son las más deseadas y comercializadas por los vendedores.

## 5. Estrategias que pueden ayudar en la identificación de productos falsificados

En este apartado se presentan algunas estrategias que pueden ayudar a Mercado Libre en la identificación de productos falsificados. Se proponen enfoques técnicos, como la validación de modelos de Machine Learning y enfoques funcionales que no requieren conocimientos técnicos.

### 5.1. Modelo de detección de anomalías – DBSCAN

DBSCAN es un algoritmo no supervisado basado en densidad que agrupa puntos de datos cercanos y detecta anomalías en regiones de baja densidad. Al aplicarlo a características extraídas de imágenes de productos, como texturas, colores y formas, se pueden identificar patrones de productos auténticos y clasificar aquellos con características atípicas como posibles falsificaciones. Su capacidad para manejar datos de alta dimensionalidad y resistencia al ruido lo convierte en una herramienta eficaz para la detección de anomalías en conjuntos de datos complejos.

### 5.2. Modelo de detección de anomalías – Autoencoder

Esta propuesta consiste en entrenar un autoencoder con una variedad de imágenes de productos auténticos, como zapatillas, para que aprenda a representar sus



características y patrones. Una vez entrenado, el modelo puede evaluar imágenes de productos potencialmente falsificados. Al intentar reconstruirlas, mostrará un alto grado de error si son falsificaciones, lo que indica que no ha captado las características normales. Este enfoque permite identificar productos falsificados de manera efectiva y se puede adaptar continuamente con nuevas imágenes, ofreciendo una solución robusta y escalable para la detección de fraudes.

### **5.3. Modelo de análisis de sentimientos - NLP**

A través de la minería de texto y el análisis de sentimientos, se pueden buscar patrones y tendencias que puedan indicar insatisfacción o desconfianza en productos que podrían ser falsificados. Esto incluirá la detección de palabras clave, análisis de tono y la clasificación de comentarios negativos o sospechosos. Al enfocarse en la voz del consumidor, esta estrategia permitirá a Mercado Libre identificar posibles fraudes basándose en la percepción del cliente, mejorando así la calidad del servicio y la confianza en los productos ofrecidos.

### **5.4. Colaboración con marcas**

Establecer una colaboración estratégica con las marcas, en la que estas puedan proporcionar información detallada sobre las características de sus productos auténticos, sin la necesidad de compartir bases de datos completas. Esta colaboración podría incluir el suministro de imágenes de alta calidad, descripciones y características únicas que diferencian sus productos en el mercado.

En caso de que la colaboración estratégica con las marcas resulte complicada, se sugiere una alternativa en la que las marcas capaciten al personal interno de Mercado Libre. Este enfoque permitiría que el equipo adquiriera el conocimiento y la experticia necesarios para desarrollar reglas de negocio que reflejen los estándares de autenticidad de cada marca.

### **5.5. Campañas al consumidor**

Otra estrategia para mitigar el impacto de los productos falsificados es implementar un programa integral de educación dirigido a los consumidores. Este programa debe centrarse en enseñar a identificar productos falsificados y fomentar la denuncia de productos sospechosos. Para incentivar a los consumidores, se puede establecer un sistema de recompensas para quienes reporten productos sospechosos que, tras una verificación, resulten ser falsificados.

## 6. Impacto en el Marketplace

- ¿Por qué crees que es esencial identificar y eliminar productos falsificados? ¿Cómo afecta esto a los compradores, vendedores y a la plataforma en general?

El problema de los productos falsificados afecta tanto a consumidores como a vendedores. Para los consumidores, la poca capacidad de detectar estos productos puede deteriorar su confianza en las compras, lo que resultará en una disminución del número de compradores y, en consecuencia, en pérdidas de ventas y una reducción de las ganancias del Marketplace.

Por otro lado, muchos vendedores ofrecen productos legítimos, pero sus ventas pueden verse afectadas por el aumento de productos falsificados, que a menudo pasan desapercibidos. Esto provoca una pérdida de credibilidad en sus productos, lo que lleva a los consumidores a buscar alternativas en otros lugares.

- Algunos vendedores buscan evadir las reglas para seguir operando. ¿Cómo crees que podrían hacerlo y qué soluciones propondrías para adelantarnos a esas tácticas?

Aunque Mercado Libre ya cuenta con el "Brand Protection Program", es crucial involucrar a los gobiernos de los 18 países en los que opera. Es necesario buscar el apoyo de más marcas y comunicar los beneficios y el impacto positivo que generan este tipo de programas. Es importante recordar que el problema de los productos falsificados refleja un fenómeno social y económico. En España desde 2019, este fenómeno ha afectado gravemente a diversos sectores, ya que muchos jóvenes se inclinan por artículos falsificados debido a su atractivo valor en el mercado.

Otra opción significativa, considerando mi desconocimiento de las normativas internas de Mercado Libre, sería implementar sanciones económicas severas para aquellos vendedores que incumplan las reglas, como la venta de productos falsificados. De esta manera, se podría bloquear a los infractores y obligarlos a cumplir con las sanciones impuestas.

## Conclusiones

1. En Argentina, Colombia y Perú se ha registrado un aumento en la cantidad de productos falsificados de alto valor. Como resultado, el precio de estos productos en estos países alcanza los 701.54 millones de dólares, lo que representa el 93% del valor total de todos los productos moderados (Ver Tabla 1). Por lo tanto, se sugiere priorizar la característica del precio en lugar de validar el volumen de los productos moderados. Por ejemplo, en Brasil se cuentan con más productos moderados, lo que se debe también al gran volumen de productos disponibles. Sin embargo, si se priorizan los países con una alta cantidad de productos moderados sin considerar aquellos donde los productos son costosos, se podría generar la pérdida de clientes importantes y, en consecuencia, pérdidas económicas significativas.
2. La tendencia más destacada en productos moderados por vendedor se observa en el sector tecnológico, especialmente en artículos como notebooks y teléfonos móviles. Este sector presenta desafíos, ya que la identificación de estos productos puede resultar más complicada debido a su alta similitud con los originales. Por ejemplo, es más fácil reconocer monturas de Ray-Ban falsificadas debido a posibles imperfecciones en su fabricación. En contraste, en el ámbito tecnológico también se enfrenta el problema de la venta de dispositivos robados. Una posible solución para mitigar las falsificaciones en este sector es optar por vendedores que ofrezcan estos productos y que estén verificados por Mercado Libre, lo que garantiza la legitimidad de los productos.
3. Según la información proporcionada, se observa que cuando el campo "Score" toma el valor de 1, se asigna al producto la regla FK\_MODEL. Esta regla indica que un producto no coincide con las especificaciones del modelo. Además, se puede inferir que el algoritmo ML tiene un buen rendimiento, ya que los productos marcados con la regla FK\_MODEL también han sido etiquetados manualmente como "Fake", y no se han registrado productos revertidos.

Por otro lado, en el caso de las reglas FK\_ATTRIBUTE y FK\_PRICE, se observa que, una vez que los productos son marcados, hay situaciones en las que no son clasificados como "Fake", y también se presentan casos de productos revertidos. Es necesario validar el funcionamiento del algoritmo ML para estas

reglas, ya que podría existir un sobreajuste o un bajo ajuste, lo que podría resultar en recalls muy altos o bajos, así como en variaciones en la precisión.

4. Independientemente de las reglas de negocio establecidas, es fundamental definir umbrales específicos para que ciertos valores del score sean automáticamente clasificados como moderados. Esto es especialmente relevante, ya que se han identificado productos con scores de 0.97, 0.98 e incluso 0.99, lo que sugiere una alta probabilidad de que estos productos sean falsificados. Se recomienda que si un producto supera un score del 0.7 sea automáticamente marcado como moderado.