



## **Data Science Challenge**

Data & Analytics Team

El desafío consta de 3 ejercicios independientes que van desde análisis exploratorio, machine learning o el diseño de una solución de data science.

### **¿Qué evaluamos?**

El desafío busca evaluar distintos aspectos como:

- Capacidad analítica y exploración de datos
- Visualización de resultados
- Conocimientos de técnicas de generación de features y modelado
- Análisis de performance
- Buenas prácticas de desarrollo
- Diseño e implementación de Machine learning en producción

### **Algunas reglas y recomendaciones:**

1. La mayoría de los ejercicios se piden resolver en Jupyter notebooks y te recomendamos subirlas a un repositorio de GitHub público para compartir los resultados.
2. No dejes de hacernos preguntas sobre cualquier duda con los enunciados.

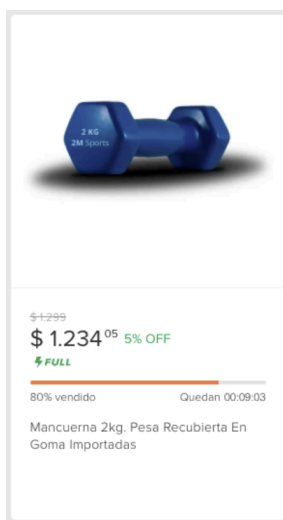
El desafío se analiza de acuerdo al seniority del postulante y teniendo en cuenta también las necesidades particulares de la posición.

# 1. Explorar las ofertas relámpago, ¿qué insights puedes generar?

## Descripción

En conjunto con el desafío te compartimos un archivo llamado ofertas\_relampago.csv el cual posee información de los resultados de ofertas del tipo relampago para un periodo de tiempo y un país determinado.

[Estas ofertas](#) en mercadolibre se pueden ver de la siguiente manera:



Es decir, son ofertas que tienen una duración definida de algunas horas y un porcentaje de unidades (stock) comprometidas.

El objetivo de este desafío es hacer un EDA sobre estos datos buscando insights sobre este tipo de ofertas.

Las columnas del dataset son autoexplicativas pero puedes preguntarnos cualquier duda.

## Entregable

El entregable de este desafío es una Jupyter notebook con el EDA.

## 2. Similitud entre productos

### Descripción

Un desafío constante en MELI es el de poder agrupar productos similares utilizando algunos atributos de estos como pueden ser el título, la descripción o su imagen.

Para este desafío tenemos un dataset “items\_titles.csv” que tiene títulos de 30 mil productos de 3 categorías diferentes de Mercado Libre Brasil

ITE_ITEM_TITLE
Tênis Yate Masculino Estilo Pleno Fr6500 Cinza
Tenis De Academia Super Confortavel Pisada Neu...
Tênis Olympikus Voa Feminino Rosa
Tenis Masculino Under Armour Amortecimento Con...
Tênis 36 Branco Star Universe Original Novo Al...
Cinta Trava Microtex Presilha Sapatilha Fizik ...
Bicicleta Absolute Aro 29 12v Preto/azul - Tam...

### Entregable

El objetivo del desafío es poder generar una Jupyter notebook que determine cuán similares son dos títulos del dataset “item\_titles\_test.csv” generando como output un listado de la forma

ITE_ITEM_TITLE	ITE_ITEM_TITLE	Score Similitud (0,1)
Zapatillas Nike	Zapatillas Adidas	0.5
Zapatillas Nike	Zapatillas Nike	1

donde ordenando por score de similitud podemos encontrar los pares de productos más similares en nuestro dataset de test.

Considere el tiempo de ejecución y analice si la solución es escalable, mostrando el tiempo de ejecución en función del tamaño de la lista a procesar.

obs: Para este ejercicio, no utilice modelos previamente entrenados.

### 3. Previsión de falla

#### Descripción

Los galpones de Full de mercado libre cuentan con una flota de dispositivos que transmiten diariamente telemetría agregada en varios atributos.

Las técnicas de mantenimiento predictivo están diseñadas para ayudar a determinar la condición del equipo de mantenimiento en servicio para predecir cuándo se debe realizar el mantenimiento. Este enfoque promete ahorros de costos sobre el mantenimiento preventivo de rutina o basado en el tiempo porque las tareas se realizan solo cuando están justificadas.

#### Entregable

El objetivo es generar un notebook Jupyter con un modelo para predecir la probabilidad de falla del dispositivo un día antes de la posible falla , se debe utilizar una time series para ayudar el modelo.

El archivo "full\_devices.csv" tiene los valores diarios para los 9 atributos de los dispositivos y la columna que está tratando de predecir se llama 'failure' con el valor binario 0 para no fallar y 1 para fallar, la columna device tiene id de los dispositivos.