

General

El desafío consta de dos partes, la primera exploratoria y una segunda de machine learning aplicado. Ambas partes están relacionadas y se tendrán en cuenta en la evaluación del desafío

¿Qué evaluamos?

El desafío busca evaluar distintos aspectos como:

- Capacidad analítica y de exploración de datos
- Visualización de resultados
- Conocimientos de técnicas de generación de features y modelado
- Análisis de performance
- Buenas prácticas de desarrollo

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

Exploracion y analisis

Muchos de los servicios principales de MELI son públicos, lo cual permite acceder a los datos de forma sencilla. En la primera parte del desafío el objetivo es realizar un **análisis exploratorio de las publicaciones con descuento** del marketplace. **Las preguntas a responder y el enfoque del análisis son libres**, como punto de partida se puede sugiere utilizar la api de search de mercadolibre, la cual está por detrás del buscador de MercadoLibre.

<https://api.mercadolibre.com/sites/MLA/search?q=tv%204k>

Uno de los resultados para la búsqueda de “tv 4k” es:

```
{
  "id": "MLA842101865",
  "site_id": "MLA",
  "title": "Smart Tv LG ThinQ 4k 43 43um7360psa",
  "seller": {},
  "price": 42999,
  "currency_id": "ARS",
  "available_quantity": 500,
  "sold_quantity": 250,
  "buying_mode": "buy_it_now",
  "listing_type_id": "gold_pro",
  "stop_time": "2040-02-28T04:00:00.000Z",
  "condition": "new",
  "permalink": "https://www.mercadolibre.com.ar/smart-tv-lg-thinq-4k-43-43um7360psa/p/MLA1540018",
  "thumbnail": "http://mla-s2-p.mlstatic.com/989312-MLA40912654718_022020-l.jpg",
  "accepts_mercadopago": true,
  "installments": {},
  "address": {},
  "shipping": {},
  "seller_address": {},
  "attributes": {},
  "differential_pricing": {},
  "original_price": null,
  "category_id": "MLA1002",
  "official_store_id": 1255,
  "catalog_product_id": "MLA15400181",
  "tags": {},
  "catalog_listing": true
},
```

* Las publicaciones **con precio con descuento** son aquellas que poseen el campo `original_price` `!= null`

En esta parte del desafío, las preguntas son abiertas, pero a modo de ayuda estas son algunas de las que se podrían responder:

- ¿Cual es el descuento promedio en distintas categorías del marketplace?
- ¿Cuántos productos con descuento tenemos en televisores? ¿y en celulares?

Te recomendamos leer nuestra documentación para familiarizarte con todos los servicios disponibles:

https://developers.mercadolibre.com.ar/es_ar/api-docs-es

Entregable

En esta etapa del desafío el entregable es una jupyter notebook que resuma el análisis y la exploración a modo de reporte. La notebook con sus dependencias debe estar disponible en un repo público de GitHub que permita ser clonado y reproducir el experimento.

Modelo

De los resultados de búsqueda del search, se pueden obtener datos interesantes sobre cada publicación: un campo valioso es el de 'sold_quantity' que hace referencia al total de ventas históricas de un ítem.

Utilizando los datos de nuestras Apis públicas, el desafío es el de armar un dataset y un modelo que permita predecir con atributos de la publicación el valor de 'sold_quantity'.

Por ejemplo para el id = [MLA842101865](https://api.mercadolibre.com/items/MLA842101865) el valor de sold_quantity es 250.

Te recomendamos empezar mirando estos servicios

- <https://api.mercadolibre.com/sites/MLA/search?q=tv%204k>
- <https://api.mercadolibre.com/items/MLA842101865>

El modelo a utilizar es libre y valoramos la comparación entre distintos caminos.

¿Qué vamos a evaluar?

- Capacidad de generar el dataset
- Los features utilizados y el tipo de modelo
- El análisis de performance del modelo
- El código y las buenas prácticas de desarrollo.

Entregable:

Un repo público de git donde se encuentre

- El pipeline de entrenamiento y scoring del modelo
- Análisis de performance del modelo