

# Deals Data Interview Challenge

## Exploración y análisis

En Meli disponibilizamos muchos de nuestros servicios principales en nuestras [apis públicas](#) lo que permite acceder a muchos de los datos de forma sencilla. En la primer parte del desafío el objetivo es realizar un **análisis exploratorio de las publicaciones con descuento** que tenemos en nuestro marketplace.

**Las preguntas a responder y el enfoque del análisis son libres**, como punto de partida se puede utilizar nuestra Api de [search](#)

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

Uno de los resultados obtenidos 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/MLA15400181",
  "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`

A modo de guía, algunas de las preguntas a responder podrían ser por ejemplo:

- ¿Cual es el descuento promedio en distintas [categorias](#) del marketplace?
- ¿Cuantos 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](https://developers.mercadolibre.com.ar/es_ar/api-docs-es)

**Entregable:**

1. Un análisis a modo reporte que debe entregarse en un jupyter notebook
2. El lenguaje utilizado debe ser python, utilizar buenas prácticas es recomendable y puedes importar paquetes propios a la notebook.
3. 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](#) 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.

¿Que 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 scoreo del modelo
- Un notebook que permita ejecutar el pipeline de scoring para distintos ids de ítem.
- Un análisis de performance del modelo

