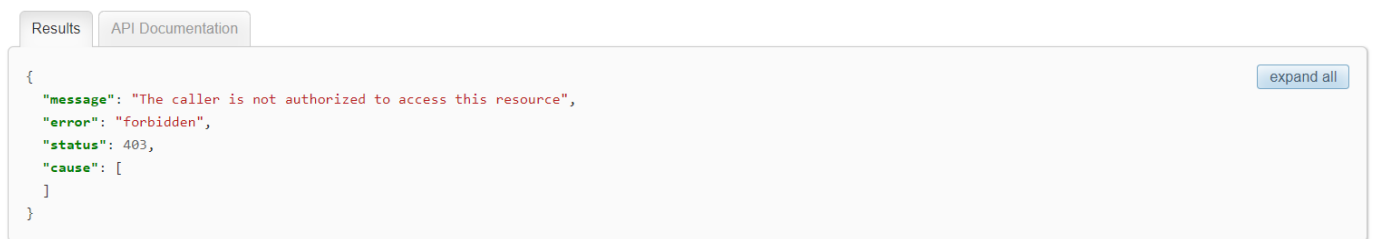


Para el desarrollo del este entregable se trabajará en cuatro grandes bloques:

1. Extracción de los datos
2. Análisis de datos
3. Transformación de los datos
4. Desarrollo de un algoritmo

Para la extracción de los datos se extraerá una muestra de la población en función de las limitaciones por la cantidad de llamados que permite la API (hasta 1000 datos por búsqueda in ID).

NOTA1: si bien se obtuvo una ID_USER, y un COD desde Developers de Mercado Libre, tenía prohibido el uso de ese recurso, o por el tipo de búsqueda. Por ende, como plan B decidí optar por buscar otros elementos en las búsquedas orientadas hacía elementos de electrónica (celulares, televisiones, tablets, notebooks). En la notebook entrega otro result pero el resultado es el mismo no pudiendo obtener más de 1000 registros.



```
{
  "message": "The caller is not authorized to access this resource",
  "error": "forbidden",
  "status": 403,
  "cause": [
  ]
}
```

Copyright © 2021 - MercadoLibre

Como alternativa se podría haber realizado web scrapping de las publicaciones de MELI disponibles desde su buscador, pero se corre el riesgo de perder información y además de identificar erróneamente los campos como precio real y con descuento. Por eso se prosigue mediante el camino de las consultas API.

Las consultas a la API se guardan mediante una función que devuelve un dataframe de los 'results', y este se guarda en un archivo .xls para su uso a posteriori sin necesidad de volver a llamar a la API.

Se unifican los dataframes obtenidos a fin de confeccionar un único DataFrame que contenga todo el contenido extraído de la API, a partir de ello se prosigue con la siguiente sección.

2. Análisis de Datos

En esta instancia se exploran los datos, para conocer la cantidad de nulos, tipos de datos.

Como definición, para ciertos campos, que a futuro es necesario no contar con valores nulos, se procederá a convertir datos en 0 por ejemplo en original_price (si bien no tiene sentido que sea 0, para el procesamiento es mejor que un nulo).

Existen ciertos campos referidos tanto al proveedor, su historia y rating, entre otros, que son diccionarios contenidos en una columna del dataframe. Se identificarán las mismas, y mediante una función "expandiremos" dicho data set. Esto en medida que los valores sean significativos al modelo de negocio, en función de mi propio entendimiento.

3. Transformación de datos

En el punto anterior existen ciertas transformaciones de datos necesarias para poder realizar el análisis pertinente.

Sin embargo, en esta instancia convertiremos ciertas features que serán de utilidad como la antigüedad del vendedor, para no trabajar con dates.

4. Desarrollo de un algoritmo

Para poder realizar un algoritmo apuntando hacia poder determinar mediante los campos de Search las unidades vendidas o 'sold_quantity', este algoritmo será de valor para determinar cuales son los parámetros o atributos que tienen mayor peso para obtener una mejor venta de productos por publicación.

Hay dos caminos para abordar las unidades vendidas, se puede realizar un forecasting conociendo las unidades vendidas de algún producto por categoría a lo largo de un periodo de tiempo, pero como vamos a emplear las publicaciones disponibles de la API.

A priori uno pensaría en: precio, descuentos aplicados, reputación, modo de pago, tiempo de la publicación, stock disponible, nuevo o usado.

Mediante un heatmap, veamos la correlación entre las distintas variables.

Veamos que variables son importantes mediante una regresión lineal, para determinar el peso de las variables antes de proseguir a aplicar modelos más complejos.

De esta manera se puede aplicar a una publicación y estimar la cantidad de unidades vendidas en función de los parámetros de la publicación.

Pequeño esquema para la exploración de datos, se realizarán dos notebooks:

1. EDA: proceso de extracción y transformación de los datos, análisis de los mismos. Respondiendo algunas preguntas de negocio
2. Instanciar un modelo a partir de los datos generados en la parte 1.

