ANÁLISIS DE RESTAURANTES

ALUMNOS:

- ALEJANDRO FERNÁNDEZ TRIGO
- JUAN DIEGO VILLALOBOS QUIRÓS



HITO II – INFORME DE SEGUIMIENTO DEL TRABAJO INTELIGENCIA EMPRESARIAL – CURSO 2021/2022

ÍNDICE

- 1. DESCRIPCIÓN DEL TRABAJO
- 2. CONTENIDOS ORIGINALES
- 3. TRATAMIENTO DE LOS DATOS
- 4. OBTENCIÓN DE DATOS COMPLEMENTARIOS
 - A. APIS
 - **B. DATASETS NUEVOS**
- 5. BIBLIOGRAFÍA, GLOSARIO

DESCRIPCIÓN DEL TRABAJO

Trabajo sobre un *dataset* de *Kaggle* que recoge todos los restaurantes de alta cocina alrededor del mundo, con la finalidad de extraer conclusiones de los datos mediante Power BI y complementarlos con otras fuentes como APIs externas.

La presentación del trabajo puede encontrarse en formato PDF en <u>ESTE</u> enlace solo para usuarios autenticados en la Universidad de Sevilla.

La idea original era hacer un análisis sobre los restaurantes y sus calificaciones por países, tipo de comida, etc. pero hemos acotado la investigación a un área concreta; el de los restaurantes de alta cocina clasificados según la Guía Michelín (por su prestigio). Partiendo de esta base nuestra intención será la de hacernos una serie de preguntas sobre los mismos:

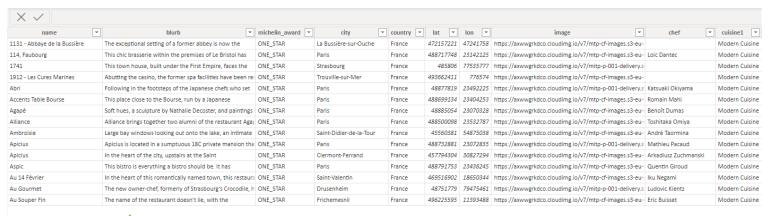
- ¿Qué países obtienen mejores calificaciones y por ende una mejor cocina?
- ¿Qué tipo/tipos de cocina son mejor valorados?
- ¿Existen chefs o tipos de cocina que destaquen por encima?

Mediante la obtención de datos complementarios, queremos ampliar este set de preguntas para cuestionarnos cosas usando las opiniones de clientes, reseñas externas, etc. de forma que podamos ver si:

- ¿Las puntuaciones que asignan entidades como la Guía Michelín se corresponden con la realidad?
- ¿Hay forma de predecir que tipos de cocina o países son más propensos a ser valorados positivamente por los comensales?

CONTENIDOS ORIGINALES

Se extrae de <u>Kaggle</u> un *dataset* generado a partir de datos recabados de la web de la guía Michelín, cuya información se muestra a continuación. Los datos originales provienen de <u>ESTA</u> dirección web.



Puede encontrarse AQUÍ.

TRATADO DE LOS DATOS

El dataset proviene 'limpio' en origen, esto es, no contiene filas incompletas que puedan dificultar el trabajo a priori ni datos en formatos desconocidos o complejos. Tras cargar los datos en *Microsoft Power BI* se hace un primer reconocimiento y se cambian:

- La clasificación en estrellas se modifica para ser numérica en lugar de texto.
- Se elimina la columna 'image' que contiene una dirección web que no nos interesa.
- Se elimina la columna 'url' que contiene una dirección web que no nos interesa.
- Se elimina la columna 'url2' que contiene una dirección web que no nos interesa.
- Se renombran las columnas de forma que pasan a ser:
 - o nombre
 - descripcion
 - estrellas
 - ciudad
 - o pais
 - o lat
 - Se conserva la latitud y longitud, a priori por si nos fuera de ayuda para el uso de APIs externas.
 - o chef
 - o tipo_cocina_1
 - o tipo cocina 2
- Se hace un primer reconocimiento para ver que no contiene errores; no obstante, no todas las entradas contienen el dato 'chef' pero a priori no se considera un problema. Así mismo, algunos restaurantes no indican un 'tipo cocina 2' pero igualmente no es problemático.

OBTENCIÓN DE DATOS COMPLEMENTARIOS

La intención de este trabajo es llegar a ciertas conclusiones sobre la calidad de los restaurantes de alta cocina según los recoge la guía Michelín, pero para ello es necesario complementar los datos originales con datos de otras fuentes.

Para ello vamos a hacer uso de <u>APIs</u> (Application Programming Interface) para recabar datos adicionales de los restaurantes de los que ya disponemos. Haremos uso de Postman para obtener nuevos conjuntos de datos que relacionar con los originales.

APIS USADAS

Para la obtención de más datos (partiendo de los campos contenidos en el dataset encontrado de Kaggle) vamos a usar la siguiente API: Yelp Fusion. Yelp es una plataforma para encontrar los mejores negocios, dados diversos datos como país, ciudad, etc...

Yelp nos permite obtener información de manera sencilla sobre restaurantes, como pueden ser valoraciones o menús, entre otros. Su API nos permite, además, obtener información de restaurantes dadas su longitud y latitud.

Yelp Fusion API

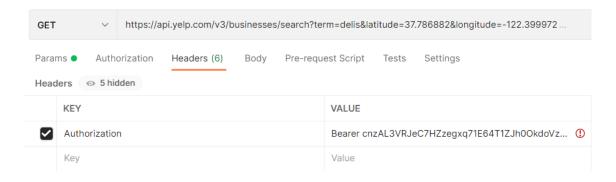
Para usar esta API, debemos de registrar una aplicación en el "dashboard" que Yelp pone a disposición de los desarrolladores. Para ello, accedemos a la siguiente URL: https://www.yelp.com/developers.

Registramos la app (en nuestro caso, la llamamos "Inteligencia Empresarial - Proyecto Universidad de Sevilla"). Una vez la registramos, obtenemos una API Key, la cual usaremos como *Bearer token*. Este *Bearer token* tenemos que usarlo en todas las llamadas a la API.

En relación con los "endpoints", poseemos varios, pero nos centraremos en uno en concreto: https://api.yelp.com/v3/businesses/search. Como parámetros tendremos que usar "term" (*string* que hace referencia al restaurante, como su nombre, por ejemplo), "latitude" y "longitude" (que hacen referencia a su ubicación).

En relación con los límites de uso de la API, esta nos permite ejecutar "endpoints" hasta 5000 veces diarias, por cada API Key, por lo que es más que suficiente para realizar el proyecto (obviamente, cada día se reinicia el número de usos, por lo que no hay problemas en este aspecto).

A continuación, podemos ver cómo quedaría el "header" de la ejecución del "endpoint":



En relación a los parámetros, estos quedarían tal y como se puede ver en la siguiente imagen:

GET	https://api.yelp.com/v3/businesses/search?te	rm=delis&latitude=37.786882&longitude=-122.399972
Params Authorization Headers (6) Body Pre-request Script Tests Settings Query Params		
	KEY	VALUE
\checkmark	term	delis
\checkmark	latitude	37.786882
\checkmark	longitude	-122.399972
	Key	Value

En este caso, la búsqueda se realizó para obtener datos del restaurante "Delicatessem", ubicado en San Francisco. Sin embargo, con modificar los tres valores ("term", "latitude" y "longitude") podemos obtener datos de cualquier restaurante.

Por último, vamos a tratar el objeto JSON que nos devuelve la API. En este caso, tendremos que "parsear" el objeto JSON a CSV, de forma que podamos tratarlo de manera más sencilla en Power BI (nos permite importar objetos JSON, pero pensamos que es mejor usar ficheros CSV, pues es mucho más sencillo a la hora de editar los mismos, así como su tratamiento).

```
E
                https://api.yelp.com/v3/businesses/search?term=delis&latitude=37.786882&longitude=-122.399972
 Params • Authorization Headers (6) Body Pre-request Script Tests Settings
                                                                                                                  Status: 200 OK Time: 1279 ms Size: 20.72 KB Save Response V
Body Cookies Headers (22) Test Results
 Pretty Raw Preview Visualize JSON V
                                                                                                                                                                              ( Q
               "businesses": [
                        "id": "FmGF1B-Rpsjq1f5b56qMwg",
                       "alos: "FmGF18-Rps]qit5056qNwg",

"alias: "molinari-delicatessen-san-francisco",

"name": "Molinari Delicatessen",

"image_uxl": "https://s3-medias.fl.yelpcdn.com/bphoto/4gRY9rVs8JyHvngljSzXyA/o.jpg",

"is_closed": false,

"url": "https://www.yelp.com/biz/molinari-delicatessen-san-francisco?adjust_creative=k2879mFN9gfjtT8M8xHW3Q&
                        categories": [
                                  "alias": "delis",
"title": "Delis"
                        "coordinates": {
    "latitude": 37.79838,
    "longitude": -122.40782
                         transactions": [
                                                                                                                    ⑤ Cookies ☺️ Bootcamp ※ Auto-select agent ▶ Runner ॥ Trash № ③
```

Como podemos comprobar en la captura anterior, el objeto JSON devuelto en la ejecución del "endpoint" posee diversos atributos: "id", "alias", "name", "image_url", "is_closed", "url", "review_count" (número de valoraciones), "rating" (valoración media), "transactions" (comida a domicilio o en restaurante), entre otros. Por tanto, poseemos muchos atributos que nos pueden servir para correlacionar con los disponibles en el dataset de Kaggle.

DATASETS NUEVOS

Para la obtención de nuevos datasets, debido a que la ejecución del "endpoint" solo nos devuelve datos sobre un restaurante, vamos a desarrollar un script en Python que sea el encargado de popular un fichero auxiliar con los datos de diversos restaurantes, con tan solo ejecutarlo una vez. De esta forma, automatizamos la obtención de estos datos, y disminuimos el tiempo necesario para crear otro dataset independiente.

Esta es la situación actual del trabajo, dónde nos encontramos desarrollando la automatización que nos permita llamar a la API y recabar todos los datos de forma automatizada.

BIBLOGRAFÍA, GLOSARIO

Bibliografía:

- Datos obtenidos de Kaggle.
- Datos obtenidos de Yelp junto a su documentación.
- Datos recabados de la web de la Guía Michelín.
- Otros datos presentes en el temario de la asignatura.

Glosario:

- API
 - Una API es el conjunto de protocolos y definiciones que se usan para integrar y desarrollar el software de las aplicaciones. La finalidad de las API es unir sus productos y sus servicios con otros sin saber cómo se implementan, lo que ayuda al desarrollo de las aplicaciones además de ahorrar tiempo y dinero.

API Key

 Las API Keys permiten que servicios diferentes conecten y se comuniquen con otros, de manera sencilla. Esto es, te permite saber cómo una aplicación cualquiera, ofrece un servicio cualquiera.

Dashboard

 Tablero o cuadro de mandos en el que se reflejan gráficamente métricas o KPIs.

• Endpoint

o Los endpoints son las URLs de una PI que responden a una petición.

• Bearer Token

o Método de autenticación seguro para la llamada de APIs.

JSON

 JSON es un formato para el almacenamiento de información estructurada que se utiliza principalmente para la transferencia de datos.