

# 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 [ESTE](#) 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 Michelin (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 Michelin 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 [Kaggle](#) un *dataset* generado a partir de datos recabados de la web de la guía Michelin, cuya información se muestra a continuación. Los datos originales provienen de [ESTA](#) dirección web.

name	blurb	michelin_award	city	country	lat	lon	image	chef	cuisine1
1131 - Abbaye de la Bussière	The exceptional setting of a former abbey is now the	ONE_STAR	La Bussière-sur-Ouche	France	472157221	47241758	https://axwwwgrkdco.cloudimg.io/v7/mtp-cf-images.s3-eu-		Modern Cuisine
114, Faubourg	This chic brasserie within the premises of Le Bristol has	ONE_STAR	Paris	France	488717748	23142125	https://axwwwgrkdco.cloudimg.io/v7/mtp-cf-images.s3-eu-	Loic Dantec	Modern Cuisine
1741	This town house, built under the First Empire, faces the	ONE_STAR	Strasbourg	France	485806	77535777	https://axwwwgrkdco.cloudimg.io/v7/mitp-p-001-delivery.s		Modern Cuisine
1912 - Les Cures Marines	Abutting the casino, the former spa facilities have been re-	ONE_STAR	Trouville-sur-Mer	France	493662411	776574	https://axwwwgrkdco.cloudimg.io/v7/mtp-cf-images.s3-eu-		Modern Cuisine
Abri	Following in the footsteps of the Japanese chefs who set	ONE_STAR	Paris	France	48877819	23492225	https://axwwwgrkdco.cloudimg.io/v7/mitp-p-001-delivery.s	Katsuaki Okiyama	Modern Cuisine
Accents Table Bourse	This place close to the Bourse, run by a Japanese	ONE_STAR	Paris	France	488699134	23404253	https://axwwwgrkdco.cloudimg.io/v7/mtp-cf-images.s3-eu-	Romain Mahi	Modern Cuisine
Agapé	Soft hues, a sculpture by Nathalie Decoster, and paintings	ONE_STAR	Paris	France	48885054	23070328	https://axwwwgrkdco.cloudimg.io/v7/mtp-cf-images.s3-eu-	Benoit Dumas	Modern Cuisine
Alliance	Alliance brings together two alumni of the restaurant Aga-	ONE_STAR	Paris	France	488500098	23532787	https://axwwwgrkdco.cloudimg.io/v7/mtp-cf-images.s3-eu-	Toshitaka Omiya	Modern Cuisine
Ambroisie	Large bay windows looking out onto the lake, an intimate	ONE_STAR	Saint-Didier-de-la-Tour	France	45560381	54875038	https://axwwwgrkdco.cloudimg.io/v7/mtp-cf-images.s3-eu-	André Taormina	Modern Cuisine
Apicius	Apicius is located in a sumptuous 18C private mansion tha	ONE_STAR	Paris	France	488732881	23072835	https://axwwwgrkdco.cloudimg.io/v7/mitp-p-001-delivery.s	Mathieu Pacaud	Modern Cuisine
Apicius	In the heart of the city, upstairs at the Saint	ONE_STAR	Clermont-Ferrand	France	457794304	30827294	https://axwwwgrkdco.cloudimg.io/v7/mtp-cf-images.s3-eu-	Arkadiusz Zuchmanski	Modern Cuisine
Aspic	This bistro is everything a bistro should be. It has	ONE_STAR	Paris	France	488791753	23436245	https://axwwwgrkdco.cloudimg.io/v7/mtp-cf-images.s3-eu-	Quentin Giroud	Modern Cuisine
Au 14 Février	In the heart of this romantically named town, this restaura	ONE_STAR	Saint-Valentin	France	469516902	18550344	https://axwwwgrkdco.cloudimg.io/v7/mtp-cf-images.s3-eu-	Iku Negami	Modern Cuisine
Au Gourmet	The new owner-chef, formerly of Strasbourg's Crocodile, h	ONE_STAR	Drusenheim	France	48751779	79475461	https://axwwwgrkdco.cloudimg.io/v7/mitp-p-001-delivery.s	Ludovic Kientz	Modern Cuisine
Au Souper Fin	The name of the restaurant doesn't lie, with the	ONE_STAR	Frichemesnil	France	496225595	11393488	https://axwwwgrkdco.cloudimg.io/v7/mtp-cf-images.s3-eu-	Eric Buisset	Modern Cuisine

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:
  - *nombre*
  - *descripcion*
  - *estrellas*
  - *ciudad*
  - *pais*
  - *lat*
  - *lon*
  - *chef*
  - *tipo\_cocina\_1*
  - *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 Michelin, pero para ello es necesario complementar los datos originales con datos de otras fuentes.

Para ello vamos a hacer uso de [APIs](#) (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”:

GET ▼ <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

Headers 👁 5 hidden

KEY	VALUE
<input checked="" type="checkbox"/> Authorization	Bearer cnzAL3VRJeC7HZzegxq71E64T1ZJh0OkdoVz... <span>🔴</span>
Key	Value

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?term=delis&latitude=37.786882&longitude=-122.399972 ...

Params

Authorization

Headers (6)

Body

Pre-request Script

Tests

Settings

Query Params

	KEY	VALUE
<input checked="" type="checkbox"/>	term	delis
<input checked="" type="checkbox"/>	latitude	37.786882
<input checked="" type="checkbox"/>	longitude	-122.399972
	Key	Value

En este caso, la búsqueda se realizó para obtener datos del restaurante “Delicatessen”, 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).

```
1 {
2   "businesses": [
3     {
4       "id": "FmGF1B-Rpsjq1f5b56qMwg",
5       "alias": "molinari-delicatessen-san-francisco",
6       "name": "Molinari Delicatessen",
7       "image_url": "https://s3-media3.fl.yelpcdn.com/bphoto/4gRY9zVs83yHvngljSzXyA/o.jpg",
8       "is_closed": false,
9       "url": "https://www.yelp.com/biz/molinari-delicatessen-san-francisco?adjust_creative=k2879mFN9gftT8M8xHW3Q&utm_campaign=yelp_api_v3&utm_medium=api_v3_business_search&utm_source=k2879mFN9gftT8M8xHW3Q",
10      "review_count": 1213,
11      "categories": [
12        {
13          "alias": "delis",
14          "title": "Delis"
15        }
16      ],
17      "rating": 4.5,
18      "coordinates": {
19        "latitude": 37.79838,
20        "longitude": -122.40782
21      },
22      "transactions": [
```

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 Michelin.
- 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
  - Los endpoints son las URLs de una PI que responden a una petición.
- Bearer Token
  - 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.