

## PRIMER FUENTE DE DATOS

En una primera instancia tenemos los datos brindados por Henry, que son Datasets de Yelp y de Google Maps.

En Yelp nos encontramos con reseñas subidas por usuarios de todo tipo de negocios, restaurantes, hoteles, servicios entre otros. Según la experiencia que han recibido.

Google Maps posee una plataforma de reseñas de todo tipo de negocios y tiene integrada en su servicio la localización y mapas

### Datasets

- [Yelp:](#)
  - checkin
  - business
  - review
  - tip
  - user
- [Google Maps:](#)
  - metadata-sitios: Contiene 11 archivos con información de comercios.
  - reviews-estados: Reviews de los comercios, separadas por estado.

## SEGUNDA FUENTE DE DATOS

Para agregar información contundente a nuestro proyecto utilizaremos datos poblacionales de la página oficial del gobierno de los Estados Unidos (API) y paginas relevantes (.csv).

# DICCIONARIO DE DATOS

## YELP

| CHECKIN        |                   |              |
|----------------|-------------------|--------------|
| Nombre columna | Descripción       | Tipo de dato |
| business_id    | Id del local      | Alfanumérico |
| date           | Fecha de registro | Fecha y hora |

| BUSINESS       |  |              |
|----------------|--|--------------|
| Nombre columna | Descripción                                      | Tipo de dato |
| business_id    | Id del local                                     | Alfanumérico |
| name           | Nombre del local                                 | Alfanumérico |
| address        | Dirección del local                              | Alfanumérico |
| city           | Ciudad del local                                 | Alfanumérico |
| state          | Estado del local                                 | Alfanumérico |
| postal_code    | Código postal                                    | Alfanumérico |
| latitude       | Latitud del local                                | Float        |
| longitude      | Longitud del local                               | Float        |
| stars          | Promedio del nivel de estrellas                  | Float        |
| review_count   | Cantidad de reseñas que tiene                    | Int          |
| attributes     | Atributos con los que cuenta el local            | Alfanumérico |
| categories     | Rubro que ocupa el local                         | Alfanumérico |
| hours          | Horarios en los que se encuentra abierto/cerrado | Alfanumérico |

| REVIEW         |   |              |
|----------------|---|--------------|
| Nombre columna | Descripción                                   | Tipo de dato |
| review_id      | Id de la review                               | Alfanumérico |
| user_id        | Id del usuario                                | Alfanumérico |
| business_id    | Id del local                                  | Alfanumérico |
| stars          | Estrella de la reseña                         | Float        |
| useful         | Cuanta gente encontró útil esta reseña        | Int          |
| funny          | Cuanta gente encontró divertida esta reseña   | Int          |
| cool           | Cuanta gente encontró interesante esta reseña | Int          |
| text           | Descripción de la reseña                      | Alfanumérico |
| date           | Fecha de la reseña                            | Fecha y hora |

| TIP              |                                    |              |
|------------------|------------------------------------|--------------|
| Nombre columna   | Descripción                        | Tipo de dato |
| user_id          | Id del usuario                     | Alfanumérico |
| business_id      | Id del local                       | Alfanumérico |
| text             | Reseña corta y rápida              | Alfanumérico |
| date             | Fecha de la reseña                 | Fecha y hora |
| compliment_count | Cuantos cumplidos tiene esa reseña | Int          |

| USER               |  |              |
|--------------------|--|--------------|
| Nombre columna     | Descripción  | Tipo de dato |
| user_id            | Id del usuario   | Alfanumérico |
| name               | Nombre del usuario                                     | Alfanumérico |
| review_count       | Cantidad de reviews que hizo                           | Int          |
| yelping_since      | Desde hace cuanto que está en Yelp                     | Fecha y hora |
| useful             | Cuanta gente encontró útil este usuario                | Int          |
| funny              | Cuanta gente encontró divertido este usuario           | Int          |
| cool               | Cuanta gente encontró interesante este usuario         | Int          |
| elite              | Años que el usuario formó parte de la elite            | Alfanumérico |
| friends            | Amigos del usuario                                     | Alfanumérico |
| fans               | Cantidad de fanáticos que tiene el usuario             | Int          |
| average_stars      | Promedio de estrellas que pone el usuario              | Float        |
| compliment_hot     | Total cumplidos 'hot' recibidos por el usuario         | Int          |
| compliment_more    | Total cumplidos varios recibidos por el usuario        | Int          |
| compliment_profile | Total cumplidos por el perfil recibidos por el usuario | Int          |
| compliment_cute    | Total cumplidos 'cute' recibidos por el usuario        | Int          |
| compliment_list    | Total listas de cumplidos recibidos por el usuario     | Int          |
| compliment_note    | Total cumplidos como notas recibidos por el usuario    | Int          |
| compliment_plain   | Total cumplidos simples recibidos por el usuario       | Int          |
| compliment_cool    | Total cumplidos interesantes recibidos por el usuario  | Int          |
| compliment_funny   | Total cumplidos divertidos recibidos por el usuario    | Int          |
| compliment_writer  | Nro de cumplidos escritos recibidos por el usuario     | Int          |
| compliment_photos  | Nro de cumplidos en foto recibidos por el usuario      | Int          |

## GOOGLE MAPS

| METADATA-SITIOS  |  |              |
|------------------|--|--------------|
| Nombre columna   | Descripción                                      | Tipo de dato |
| name             | Nombre del local                                 | Alfanumérico |
| address          | Dirección del local                              | Alfanumérico |
| gmap_id          | Id del local                                     | Alfanumérico |
| description      | ...  | ...          |
| latitude         | Latitud del local                                | Float        |
| longitude        | Longitud del local                               | Float        |
| category         | Rubros del local                                 | Alfanumérico |
| avg_rating       | Promedio de estrellas del local                  | Float        |
| num_of_reviews   | Cantidad de reseñas                              | Int          |
| price            | Precio   | Float        |
| hours            | Horarios de atención                             | Alfanumérico |
| MISC             | Opciones de servicio                             | Alfanumérico |
| state            | Estado del local                                 | Alfanumérico |
| relative_results |  | Alfanumérico |
| url              | Ubicación y descripción del local en Google Maps | Alfanumérico |

| REVIEW-ESTADOS |                              |              |
|----------------|------------------------------|--------------|
| Nombre columna | Descripción                  | Tipo de dato |
| user_id        | Id del usuario               | Alfanumérico |
| name           | Nombre del usuario           | Alfanumérico |
| time           | ...                          | ...          |
| rating         | Rating que puso en la reseña | Float        |
| text           | Texto de la reseña           | Alfanumérico |
| pics           | Foto de la reseña            | Alfanumérico |
| resp           | Respuestas a la reseña       | Alfanumérico |
| gmap_id        | Id del local                 | Alfanumérico |

## DATA POBLACIONAL EEUU

| EEUU           |                      |              |
|----------------|----------------------|--------------|
| Nombre columna | Descripción          | Tipo de dato |
| NAME           | Nombre del estado    | Alfanumérico |
| POP_2020       | Poblacion del 2020   | int          |
| POP_2021       | Poblacion del 2021   | int          |
| DENSITY_2020   | Densidad del 2020    | Float        |
| DENSITY_2021   | Densidad del 2021    | Float        |
| LASTUPDATE     | Ultima actualización | Fecha        |
| state          | Numero del estado    | int          |

# TABLAS DEL DIAGRAMA ENTIDAD RELACIÓN

## PLATAFORMA

id\_plataforma (PK)

plataforma: A que plataforma corresponde (google / yelp)

## CATEGORIAS

id\_categoria (PK)

categoria: Categoría con la cual se identifica el comercio.

## SERVICIOS

id\_servicio (PK)

servicio: Servicios que brinda el comercio.

## ESTADOS

id\_estado (PK)

estado: Nombre del estado (USA)

avg\_pop: Población del estado.

avg\_pop\_growth: Índice de crecimiento de la población con respecto a años anteriores.

per\_capita\_income: Ingresos de los habitantes por cabeza.

## CIUDADES

id\_ciudad (PK)

id\_estado (FK)

ciudad: Nombre de la ciudad (USA)

## USUARIOS

id\_usuario (PK)

id\_plataforma (FK)

identificador\_usuario: Código interno de la plataforma, que identifica al usuario.

nombre: Nombre del usuario.

reviews\_count: Cantidad de reviews realizadas por el usuario.

fans\_count: Cantidad de fans que posee el usuario.

avg\_stars: Promedio de estrellas en las reseñas realizadas.

compliment\_count: Cantidad de cumplidos que tienen las reviews del usuario.

cool\_count: Cantidad de “cool” que tienen las reviews del usuario (buena onda).

funny\_count: Cantidad de “funny” que tienen las reviews del usuario (gracioso).

## **RESTAURANTES**

id\_restaurante (PK)

id\_ciudad (FK)

identificador\_yelp: Si el restaurante está en yelp, aquí aparecerá su identificador.

identificador\_google: Si el restaurante está en google, aquí aparecerá su identificador.

nombre: Nombre del comercio.

direccion: Dirección del comercio.

latitud: latitud geográfica para geolocalización.

longitud: longitud geográfica para geolocalización.

avg\_rating: Promedio de los rating recibidos.

reviews\_count: Conteo de cantidad de reviews recibidas.

estado: Si el local actualmente funciona o está cerrado.

## **CATEGORIASRESTAURANTES**

id\_categoria\_restaurante (PK)

id\_categoria (FK)

id\_restaurante (FK)

## **SERVICIOSRESTAURANTES**

id\_servicio\_restaurante (PK)

id\_servicio (FK)

id\_restaurante (FK)

## **REVIEWS**

id\_review (PK)

id\_restaurante (FK)

id\_usuario (FK)

identificador\_plataforma: Código interno de la plataforma, que identifica la review..

raiting: Cantidad de estrellas que se otorgó.

review: Texto de la review.

fecha\_hora: Fecha y hora que se registró.

useful: Cantidad de veces que la review fue marcada como “useful” (útil).

funny: Cantidad de veces que la review fue marcada como “funny” (gracioso).

cool: Cantidad de veces que la review fue marcada como “cool” (buena onda).



## **CONSEJOS**

id\_consejo (PK)

id\_restaurante (FK)

id\_usuario (FK)

consejo: Texto del consejo.

fecha: Fecha en que se registró.

compliment\_count: Cantidad de usuarios que le marcaron un cumplido.