T1

| **Variable** | **Tipo de variable** | **Descripción** | **Útil?** |
| --- | --- | --- | --- |
| name | Texto | Nombre del comercio | si |
| address | Texto | Dirección del comercio | si |
| gmap\_id | Texto | ID de Google Maps asociado al comercio | si |
| description | Texto | Descripción del comercio | tv |
| latitude | Numérico | Latitud de la ubicación del comercio | si |
| longitude | Numérico | Longitud de la ubicación del comercio | si |
| category | Texto | Categoría o categorías del comercio | si |
| avg\_rating | Numérico | Calificación promedio del comercio | si |
| num\_of\_reviews | Numérico | Número de reseñas del comercio | si |
| price | Texto | Rango de precios del comercio | si |
| hours | Lista de texto | Horario de atención del comercio | si |
| MISC.Service options | Lista de texto | Opciones de servicio ofrecidas por el comercio | tv |
| MISC.Health & safety | Lista de texto | Información sobre medidas de salud y seguridad del comercio | tv |
| MISC.Accessibility | Lista de texto | Información sobre accesibilidad del comercio | tv |
| MISC.Planning | Lista de texto | Información sobre planificación de visita del comercio | tv |
| MISC.Payments | Lista de texto | Información sobre opciones de pago aceptadas por el comercio | tv |
| state | Texto | Estado de apertura y cierre del comercio en el momento de la data | si |
| relative\_results | Lista de texto | IDs de Google Maps de otros comercios cercanos a la ubicación | si (k-vecin) |
| url | Texto | URL de Google Maps del comercio | tv |

T2

| **Variable** | **Descripción** | **Clasificación** | **útil?** |
| --- | --- | --- | --- |
| user\_id | Identificador único del usuario que realizó la reseña. | Identificador | si |
| name | Nombre del usuario que realizó la reseña. | Categórico | si |
| time | Fecha y hora en que se realizó la reseña en formato Unix Timestamp. | Temporal | si |
| rating | Calificación numérica dada por el usuario al negocio, de 1 a 5 estrellas. | Numérico | si |
| text | Texto de la reseña escrita por el usuario. | Textual | si |
| pics | Imágenes adjuntas a la reseña, si las hay. | Categórico | no |
| resp.time | Fecha y hora en que se escribió la respuesta del negocio, en formato Unix Timestamp. | Temporal | tv |
| resp.text | Texto de la respuesta del negocio a la reseña del usuario. | Textual | tv |
| gmap\_id | Identificador de Google Maps asociado con el negocio en cuestión. | Identificador | si |

T3

| **Variable** | **Descripción** | **Clasificación** |  | **Utilidad?** |
| --- | --- | --- | --- | --- |
| business\_id | Identificador único del negocio. | Identificador |  | si |
| name | Nombre del negocio. | Categórico |  | si |
| address | Dirección completa del negocio. | Textual |  | tv |
| city | Ciudad donde se encuentra el negocio. | Categórico |  | si |
| state | Código de dos letras del estado donde se encuentra el negocio. | Categórico |  | si |
| postal code | Código postal del negocio. | Categórico |  | si |
| latitude | Latitud geográfica de la ubicación del negocio. | Numérico |  | si |
| longitude | Longitud geográfica de la ubicación del negocio. | Numérico |  | si |
| stars | Calificación promedio del negocio en estrellas (de 0 a 5), redondeado a 0.5. | Numérico |  | si |
| review\_count | Número de reseñas del negocio. | Numérico |  | si |
| is\_open | Indicador binario de si el negocio está abierto (1) o cerrado (0). | Binario |  | si |
| attributes | Atributos del negocio, como un objeto con pares clave-valor. Algunos valores de atributos también pueden ser objetos. | Categórico |  | tv |
| categories | Lista de categorías de negocios a las que pertenece el negocio. | Categórico |  | si |
| hours | Horario del negocio como un objeto que mapea los días de la semana a los horarios de apertura y cierre. | Temporal |  | tv |

T4

| **Variable** | **Descripción** | **Clasificación** | **útil?** |
| --- | --- | --- | --- |
| review\_id | Identificador único de la reseña | Categórica | si |
| user\_id | Identificador único del usuario que escribió la reseña, referido en el archivo user.json | Categórica | si |
| business\_id | Identificador único del negocio al cual se refiere la reseña, referido en el archivo business.json | Categórica | si |
| stars | Puntaje en estrellas de 1 al 5 dado por el usuario en su reseña | Numérica discreta | si |
| date | Fecha de la reseña en formato YYYY-MM-DD | Temporal | si |
| text | Texto completo de la reseña escrita por el usuario | Textual | tv |
| useful | Número de votos que la reseña recibió como útil | Numérica discreta | si |
| funny | Número de votos que la reseña recibió como graciosa | Numérica discreta | si |
| cool | Número de votos que la reseña recibió como "cool" | Numérica discreta | si |

T5 user

| **Variable** | **Descripción** | **Clasificación** | **útil?** |
| --- | --- | --- | --- |
| user\_id | Identificador único del usuario en Yelp | Identificador | si |
| name | Nombre del usuario | Categórica | si |
| review\_count | Número de reseñas escritas por el usuario | Numérica | si |
| yelping\_since | Fecha de creación de la cuenta de Yelp del usuario en formato YYYY-MM-DD | Temporal | si |
| friends | Lista de identificadores de usuarios que son amigos del usuario | Identificador | tv |
| useful | Número de votos marcados como útiles por el usuario | Numérica | si |
| funny | Número de votos marcados como graciosos por el usuario | Numérica | si |
| cool | Número de votos marcados como cool por el usuario | Numérica | si |
| fans | Número de fans que tiene el usuario | Numérica | si |
| elite | Lista de años en los que el usuario ha sido miembro elite | Temporal/Categ. | tv |
| average\_stars | Promedio del valor de las reseñas escritas por el usuario | Numérica | si |
| compliment\_hot | Número de cumplidos 'hot' recibidos por el usuario | Numérica | si |
| compliment\_more | Número de cumplidos varios recibidos por el usuario | Numérica | si |
| compliment\_profile | Número de cumplidos por el perfil recibidos por el usuario | Numérica | si |
| compliment\_cute | Número de cumplidos 'cute' recibidos por el usuario | Numérica | si |
| compliment\_list | Número de listas de cumplidos recibidos por el usuario | Numérica | si |
| compliment\_note | Número de cumplidos como notas recibidos por el usuario | Numérica | si |
| compliment\_plain | Número de cumplidos planos recibidos por el usuario | Numérica | si |
| compliment\_cool | Número de cumplidos 'cool' recibidos por el usuario | Numérica | si |
| compliment\_funny | Número de cumplidos graciosos recibidos por el usuario | Numérica | si |
| compliment\_writer | Número de cumplidos escritos recibidos por el usuario | Numérica | si |
| compliment\_photos | Número de cumplidos en foto recibidos por el usuario | Numérica | si |

La clasificación de las variables se divide en tres categorías principales:

* **Identificador**: Variables que identifican de manera única al usuario en la base de datos.
* **Categórica**: Variable nominal que describe una categoría del usuario (nombre).
* **Numérica**: Variables que toman valores numéricos y se pueden utilizar en cálculos matemáticos y estadísticos. Estas se dividen en dos subcategorías:
  + **Temporal**: Variable que describe una fecha o tiempo (yelping\_since y elite).
  + **No Temporal**: Variable que describe una cantidad numérica.

La tabla para el archivo "checkin.json" quedaría de la siguiente manera:

| **Columna** | **Descripción** | **Clasificación** | **útil?** |
| --- | --- | --- | --- |
| business\_id | Identificador único del negocio | Categórica, nominal | si |
| date | Fechas y horas en las que se registraron check-ins | Categórica, temporal | si |

Tabla de descripción de variables y clasificación para el archivo "tip.json":

| **Variable** | **Descripción** | **Clasificación** | **útil?** |
| --- | --- | --- | --- |
| text | Texto del consejo dado por el usuario sobre el negocio | Categórica | si |
| date | Fecha en que se escribió el consejo en formato YYYY-MM-DD | Temporal | si |
| compliment\_count | Número de cumplidos totales recibidos por el consejo | Numérica | si |
| business\_id | Identificador de 22 caracteres del negocio en business.json | Categórica | si |
| user\_id | Identificador de 22 caracteres del usuario en user.json | Categórica | si |