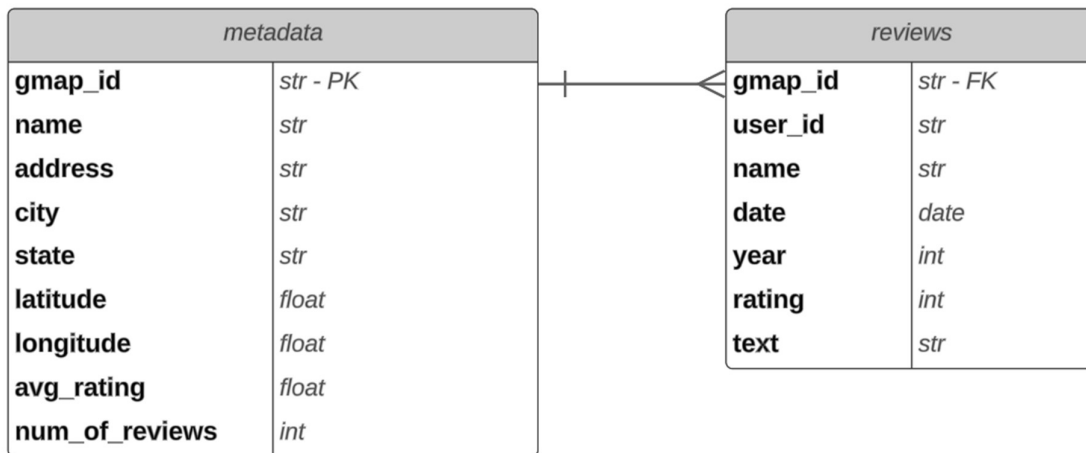


## DIAGRAMA E-R



Este diagrama E-R refleja de forma visual, la estructura de los datos con los que estamos trabajando (después de haberlos limpiado y procesado durante el proceso de ETL), y ayuda a entender las relaciones entre las dos entidades clave: metadata (la información de los restaurantes) y reviews (las reseñas de esos restaurantes).

## Explicación de las entidades

### Entidad metadata (Google Maps):

Esta tabla contiene toda la información básica y geográfica de los restaurantes chinos registrados en Google Maps. Cada restaurante tiene un identificador único (`gmap_id`), que se usa para identificarlo de forma exclusiva. Atributos como el nombre, dirección, ciudad, estado, latitud y longitud permiten no solo identificar el restaurante, sino también posicionarlo geográficamente para análisis espaciales. Atributos clave como `avg_rating` y `num_of_reviews` son resultados agregados que permiten tener una idea del rendimiento o la popularidad de cada restaurante. El objetivo de esta entidad es almacenar toda la información estática sobre cada restaurante y su ubicación, que posteriormente se puede usar para relacionarla con las reseñas.

### Entidad reviews (Google Maps):

Esta tabla almacena las reseñas de los usuarios sobre los restaurantes. Cada reseña está conectada a un restaurante en la tabla metadata a través de la clave foránea `gmap_id`, que indica de qué restaurante es la reseña. La columna `user_id` identifica al usuario que dejó la reseña, lo que permite hacer análisis basados en el comportamiento de los usuarios. Atributos como `date`, `year`, `rating` y `text` proporcionan detalles sobre la reseña en sí: cuándo fue escrita, qué puntuación otorgó el usuario y qué comentarios dejó. Esta entidad permite hacer análisis más detallados sobre el feedback de los usuarios, como entender las tendencias en las reseñas, calificaciones a lo largo del tiempo, o incluso realizar análisis de sentimientos sobre el texto de las reseñas.

### Relación entre metadata y reviews:

La relación entre ambas entidades es de 1 a muchos; indica que un restaurante en metadata puede tener muchas reseñas en la entidad denominada reviews, pero cada reseña está asociada a un solo restaurante. Esta relación es clave para realizar análisis cruzados. Por ejemplo, para obtener el promedio de las calificaciones de las reseñas de un restaurante y compararlo con su `avg_rating`, o incluso analizar la distribución de las reseñas por ciudad o estado.

## DICCIONARIO DE LOS DATASETS LIMPIOS

### metadata (Google Maps):

- **gmap\_id (str - PK):** Identificador único del restaurante de comida china registrado en Google Maps
- **name (str):** Nombre del restaurante
- **address (str):** Dirección completa del restaurante
- **city (str):** Ciudad en la que se encuentra el restaurante
- **state (str):** Estado (de la costa este de EEUU) donde se encuentra el restaurante
- **latitude (float):** Latitud del restaurante
- **longitude (float):** Longitud del restaurante
- **avg\_rating (float):** Calificación promedio que recibió el restaurante
- **num\_of\_reviews (int):** Número de reseñas total que recibió

### reviews (Google Maps)

- **gmap\_id (str - FK):** Identificador único del restaurante al que se refiere la reseña
- **user\_id (str):** Identificador único del usuario que publicó la reseña
- **name (str):** Nombre del usuario
- **date (date):** Fecha en que se publicó la reseña
- **year (int):** Año en que se publicó la reseña
- **rating (int):** Calificación otorgada
- **text (str):** Texto de la reseña