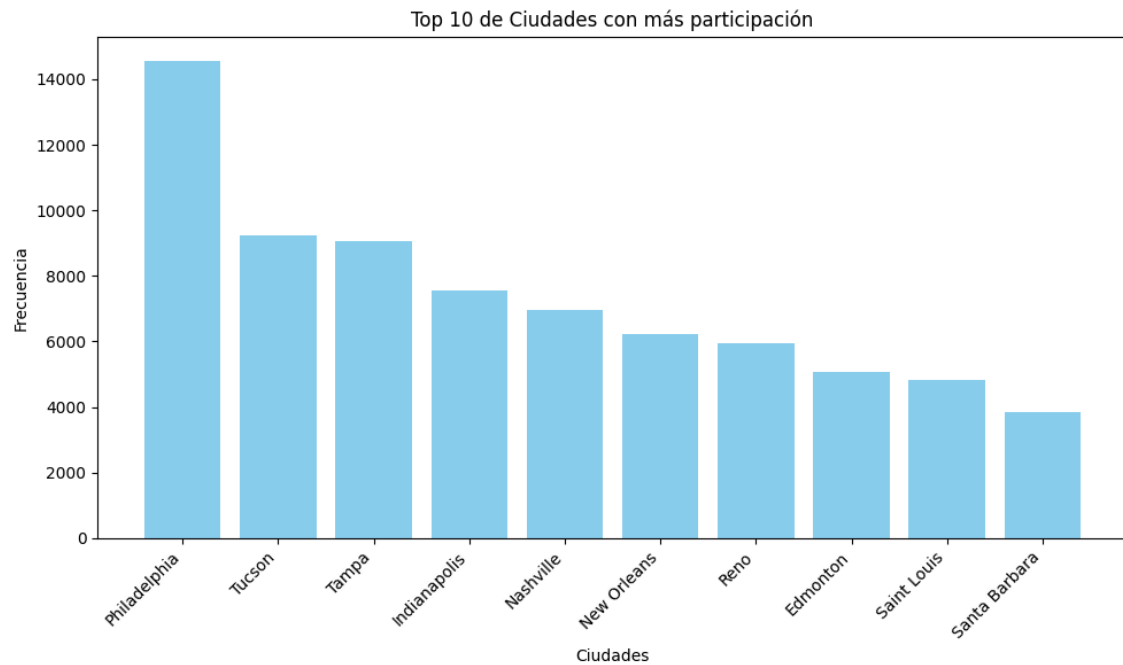
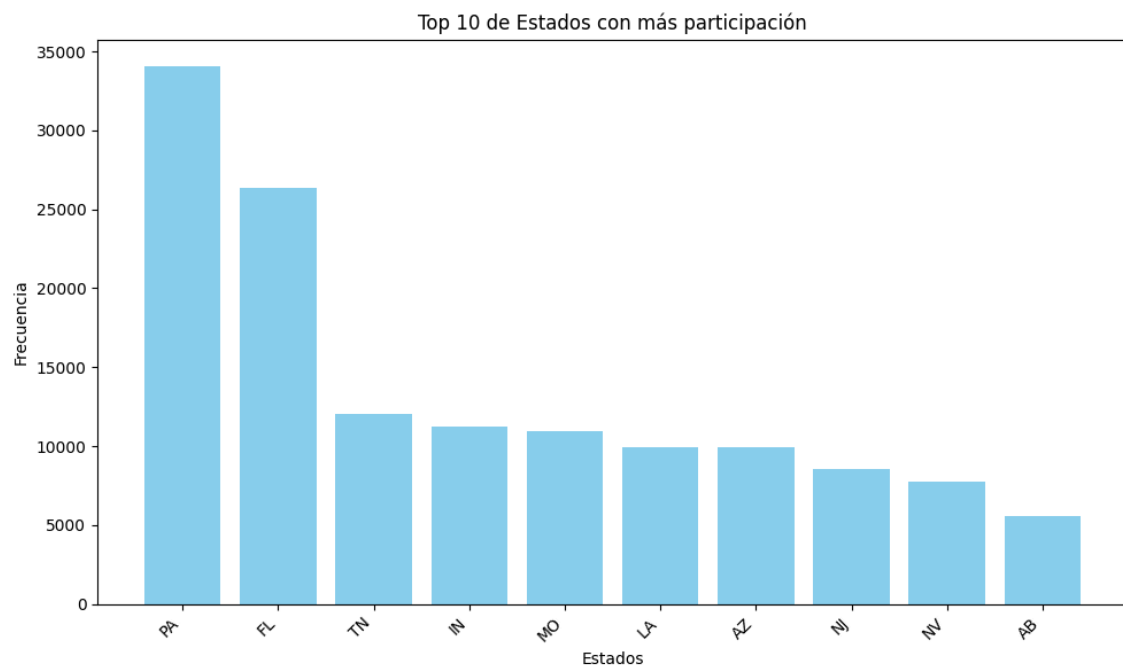


Análisis exploratorio de datos (EDA)

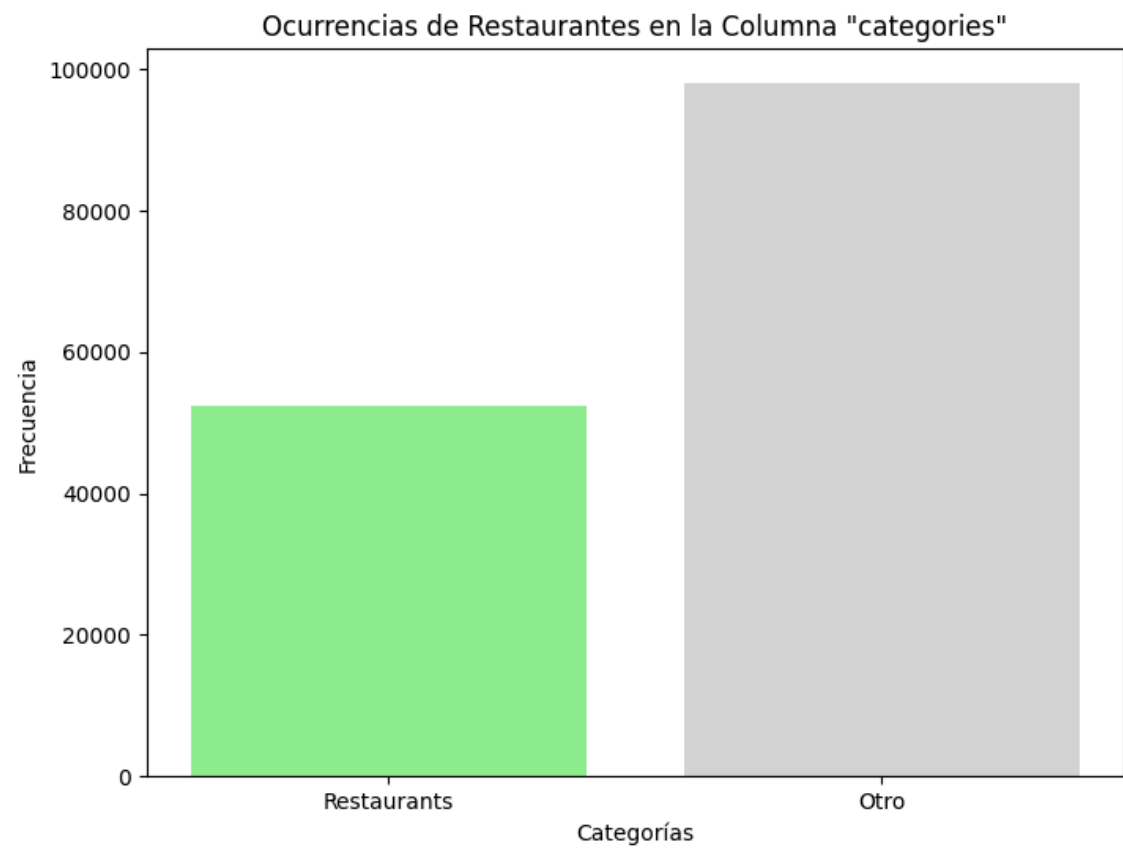
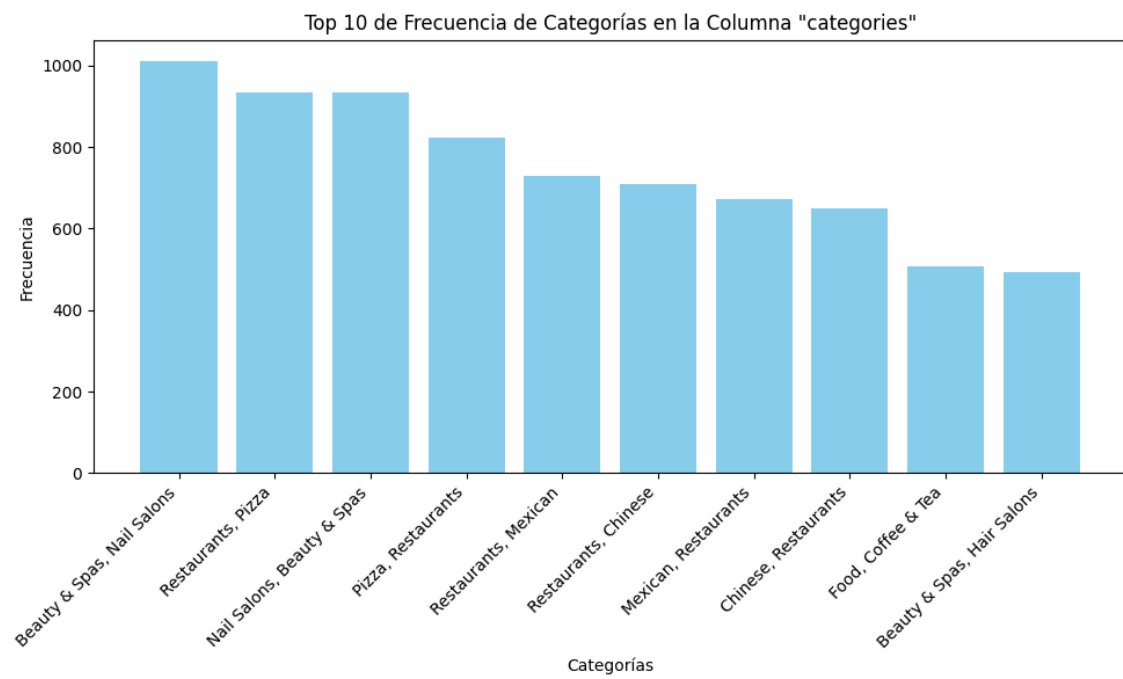
Se uso para analizar e investigar conjuntos de datos y resumir sus principales características, empleando métodos de visualización de datos. Nos ayudó a determinar la mejor manera de manipular los orígenes de datos para obtener las respuestas que necesita, lo que nos permitió descubrir patrones y detectar anomalías



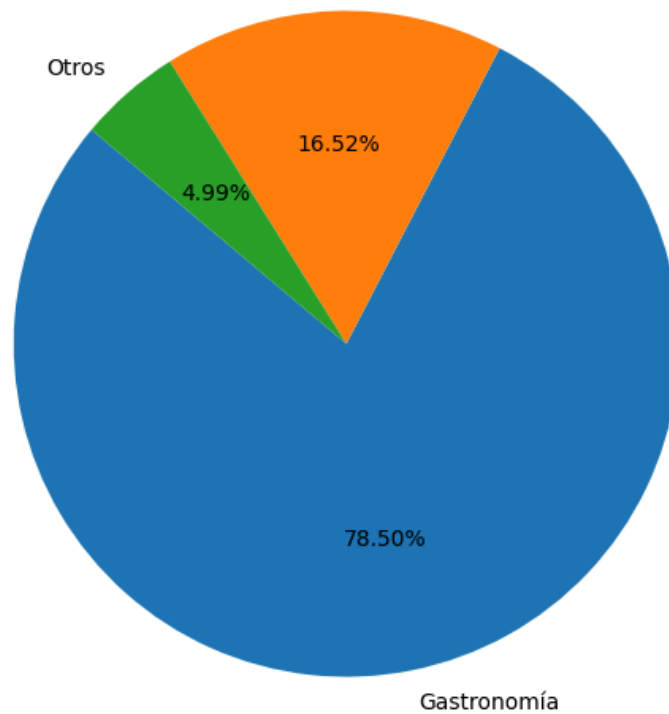
Conocer el top de ciudades con más participación



Estados con mas frecuencia PA (Pennsylvania), FL (Florida), TN (Tennessee)



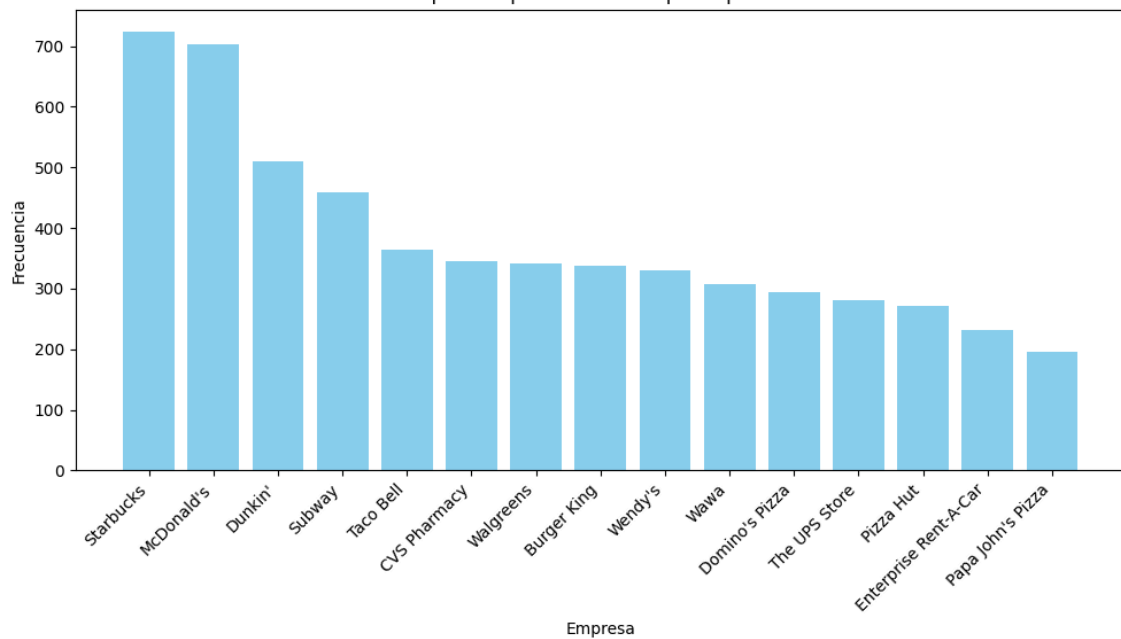
Distribución de ocurrencias por Rubro
Belleza



Se observa que en el top 10 aparecen como predominante 'Restaurants'. A su vez los servicios similares como 'Food', 'Pizza', 'Chinese', 'Mexican', 'Coffe & Te', están también en el top 10.

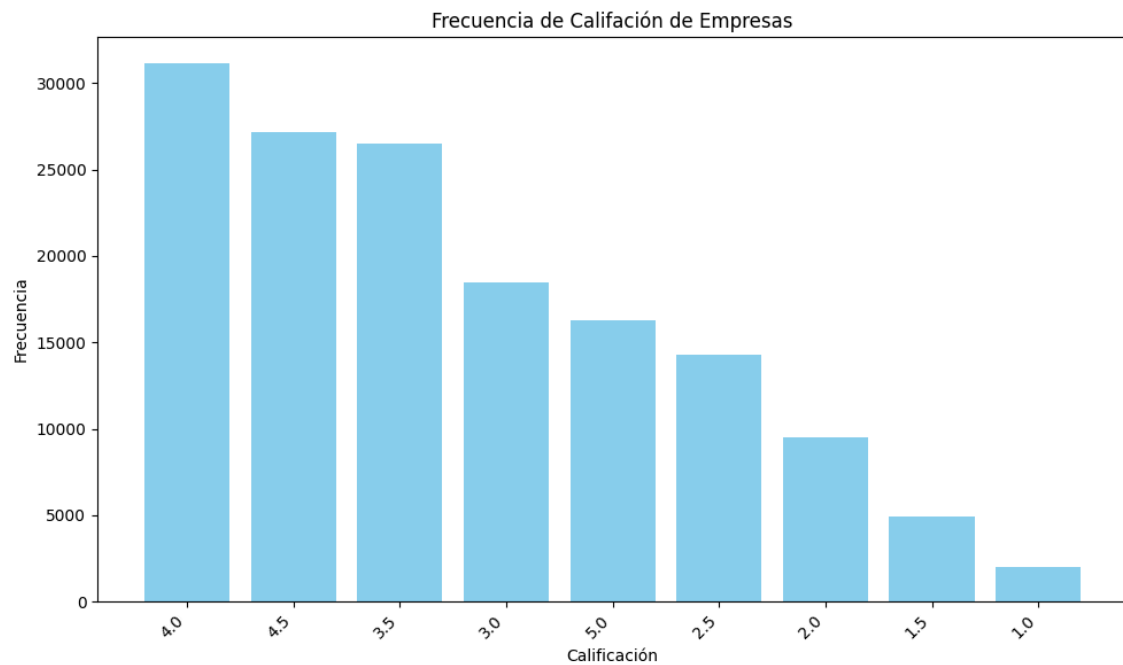
Si agrupáramos esas categorías 'Food', 'Pizza', 'Chinese', 'Mexican', 'Coffe & Te', junto con 'Restaurants' en el Rubro 'Gastronomía' obtendríamos que tienen una participación 118015 ocurrencias en nuestros datos

Top 15 Empresas con más participación

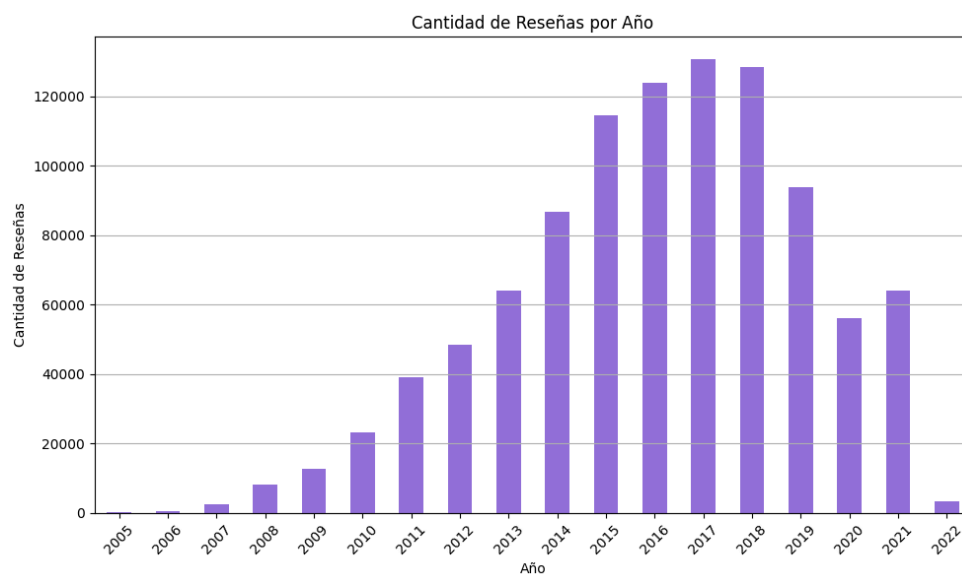


Aunque los restaurantes son dominantes, la presencia de categorías como Pizza, Comida China, Mexicana y Cafeterías sugiere una diversidad de opciones gastronómicas que los usuarios están explorando y evaluando.

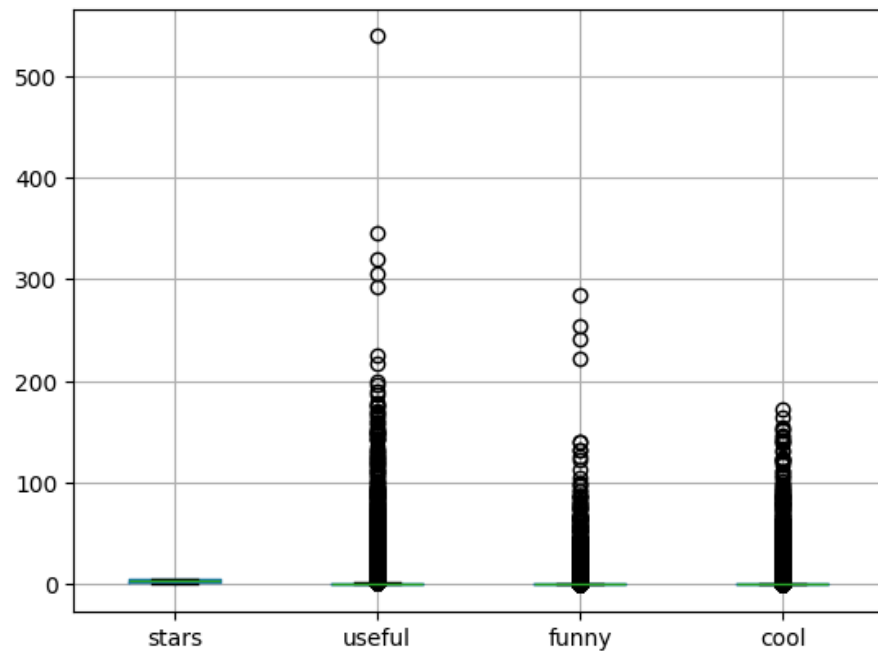
Dominio del sector de la alimentación: La presencia abrumadora de restaurantes y servicios de comida en el top 10 sugiere que el sector de la gastronomía es altamente competitivo y dominante en la plataforma de Yelp. Esto podría indicar una alta demanda y actividad en este mercado.



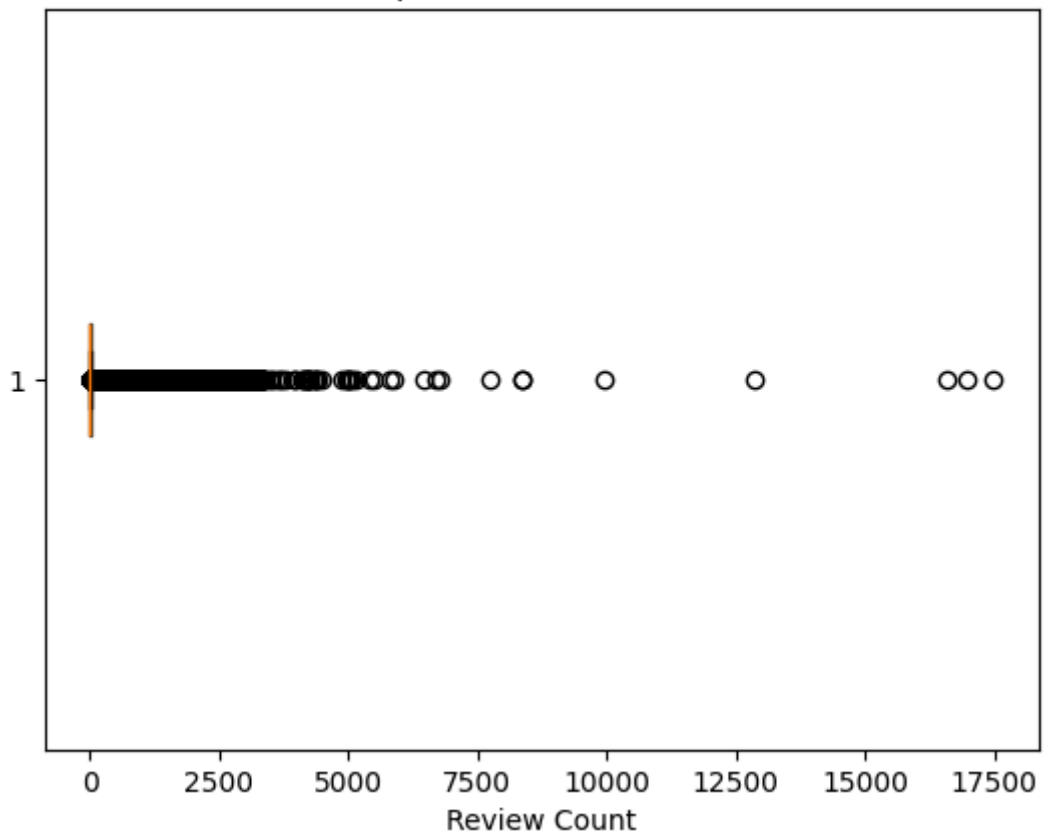
Las calificaciones más comunes son 4.0, 3.5 y 4.5, lo que sugiere que la mayoría de las empresas en el sector de la gastronomía están recibiendo calificaciones relativamente altas. Sin embargo, sería interesante investigar más a fondo las razones detrás de las calificaciones más bajas (3.5 y por debajo) para identificar áreas de mejora para las empresas.



Se determina el rango de años de las reseñas. Del 2005 al 2022.



Boxplot de Review Count



Dentro de la variable useful, funny review_count y fans, los datos se encuentran muy dispersos, para un correcto análisis se realizara una limpieza y se trabajara con los datos no muy dispersos