

Capstone Project I: Data Wrangling

Data

Restaurant reviews in Toronto from Yelp open dataset is used for the project (<https://www.yelp.com/dataset>). The dataset is pretty clean..

Related files

- review.json: contains data related to the review including business_id, no information about the business.
- business.json: contains data related to the business.

Steps

1. Get Toronto restaurants from business.json by categories and city.
2. Read reviews from review.json by chunk and merge it with Toronto restaurants.
3. Reset business_id and review_id as indexes.
4. Rename columns.