

PROJECT STAGE 1 (CS 838)

Tarun Bansal, Rohit Damkondwar, Ayush Gupta

Question(s) we want to answer:

1. Do opening-closing hours decide the popularity of a restaurant?
2. If restaurants provide parking facility, are they visited more?

Data Sources:

1. Yelp Dataset (https://www.yelp.com/dataset_challenge)
2. Zomato API (<https://developers.zomato.com/documentation#>)

Extracting Structured Data from Data Sources:

S1 - Zomato Dataset

We made use of Zomato's developer API to get details of top 100 restaurants in major cities across USA, UK and Canada.

We used the following python script to get the above information.

<http://pages.cs.wisc.edu/~tarun/zomato/zomato.txt>

S2 - Yelp Dataset

Yelp provides academic dataset to students. Yelp dataset provides JSON documents containing Restaurant details. Each document contains Restaurant's name, geo location, address, special features offered like Parking, hasTV, goodForKids, etc. These can provide us more insights about the restaurant.

S3 - Text Documents

Yelp also provided text documents containing reviews about all the restaurants. The dataset acquired contains 40 million reviews for 140K restaurants in JSON format.

The reviews received from Yelp -

http://pages.cs.wisc.edu/~rdamkondwar/cs838-ds/yelp_academic_dataset_review.json

Extraction from S3 Text Documents:

The Yelp dataset provided Restaurant details in JSON format. The second dataset was of reviews flattened in a single large file with 40 millions records. We wrote a python program to merge map every restaurant and its reviews and store them in separate files.

The output after merging data - <http://pages.cs.wisc.edu/~rdamkondwar/cs838-ds/yelp/>

Python program - http://pages.cs.wisc.edu/~rdamkondwar/cs838-ds/yelp_parse_py.txt

Open Source tool used:(and brief description)

Python : urllib, json libraries

Jq: Command line JSON processing tool

Link of Structured Data/Text Documents:

S1 - Zomato Dataset

<http://pages.cs.wisc.edu/~tarun/zomato/>

S2 - YELP Dataset

<http://pages.cs.wisc.edu/~tarun/yelp/>

S3 - TEXT Documents

<http://pages.cs.wisc.edu/~rdamkondwar/cs838-ds/yelp/>