

## SI 507 – Final Project

SICHEN LIU [sichenl@umich.edu](mailto:sichenl@umich.edu)

### Project code:

1. Link to github: [https://github.com/elainesichen/Final\\_project\\_si507](https://github.com/elainesichen/Final_project_si507)
2. Demo video link: <https://www.youtube.com/watch?v=NNVNUFTkpvQ>
3. Special instructions for running code: ①Get daily weather information in Boston for the past six years through National Centers for Environmental Information(NOAA) API key; ②Two options for users to interact: search for restaurants information or weather and bike-sharing information in Boston.
4. Required Python packages:

```
import noaa_api_v2
from time import sleep
import requests
from bs4 import BeautifulSoup
import secret_final
import json
import csv
import plotly.graph_objects as go
import plotly.express as px
import plotly.io as pio
pio.renderers.default='browser'
import Tree_structure
import print_tree
import pandas as pd
```

### Data sources:

1. Yelp(restaurant\_data): Web scraping Yelp restaurant listings using Requests, BeautifulSoup. Get details of restaurants that are listed for the city of Boston, Massachusetts.

#### *Summary of data(json file: use web scraping and caching):*

There are five attributes: ①restaurant's rank, ②name, ③rating(full score is 5), ④review numbers, ⑤more description(In addition to the main cuisine categories, more labels about the restaurant)

The screenshot shows a code editor on the left with a Python script and a dictionary viewer on the right. The code defines a function to scrape restaurant data from Yelp. The dictionary viewer shows a list of restaurant entries with their rank, name, rating, review count, and cuisine categories.

Key	Type	Size	Value
Burger	list	240	[{'rank': '1', 'name': 'Neptune Oyster', 'rating': 4.5, 'reviews_number': 5407, 'more description': ['Seafood', 'Salad', 'Sandwiches']}]
Chinese	list	239	[{'rank': '1', 'name': 'Q Restaurant', 'rating': 4.0, 'reviews_number': 7643, 'more description': ['Bakeries', 'Desserts', 'Gelato']}]
Italian	list	240	[{'rank': '1', 'name': 'Carmelina's', 'rating': 4.5, 'reviews_number': 7643, 'more description': ['Bakeries', 'Desserts', 'Gelato']}]
Japanese	list	233	[{'rank': '1', 'name': 'Douzo', 'rating': 4.0, 'reviews_number': 7643, 'more description': ['Bakeries', 'Desserts', 'Gelato']}]
Mexican	list	237	[{'rank': '1', 'name': 'Citrus & Salt', 'rating': 4.0, 'reviews_number': 7643, 'more description': ['Bakeries', 'Desserts', 'Gelato']}]
Thai	list	134	[{'rank': '1', 'name': 'Pho Basil', 'rating': 4.0, 'reviews_number': 7643, 'more description': ['Bakeries', 'Desserts', 'Gelato']}]

2. National Centers for Environmental Information(weather\_data): Get daily weather information in Boston for the past six years through API key.

#### *Summary of data (json file: use API key)*

There are seven attributes: ①Date: 2015-01-01 to 2020-12-31; ②Min\_temp: Minimum temperature (degrees C); ③Max\_temp: Maximum temperature (degrees C); ④Avg\_temp: Average temperature (degrees C); ⑤PRCP= Precipitation (mm); ⑥SNOW= Snowfall (mm); ⑦AWND= Average daily wind speed (tenths of meters per second);

```

1 [
2   {
3     "date": "2015-01-01",
4     "min_temp": -5.5,
5     "max_temp": 0.6,
6     "avg_temp": -3.3,
7     "prcp": 0.0,
8     "snow": 0,
9     "awnd": 64
10  },
11  {
12    "date": "2015-01-02",
13    "min_temp": -0.5,
14    "max_temp": 5.0,
15    "avg_temp": 1.3,
16    "prcp": 0.0,
17    "snow": 0,
18    "awnd": 57
19  },
20  {
21    "date": "2015-01-03",
22    "min_temp": -5.5,
23    "max_temp": 2.8,
24    "avg_temp": -2.1,
25    "prcp": 15.7,
26    "snow": 36,
27    "awnd": 46
28  },
29  {
30    "date": "2015-01-04",
31    "min_temp": 1.7,
32    "max_temp": 11.1,
33    "avg_temp": 3.5,
34    "prcp": 14.5,
35    "snow": 0,
36    "awnd": 37
37  }
38 ]

```

3. Blue Bikes(bike\_data): Obtain bike usage in Boston every day for the past six years. (Company 'Blue Bikes' shares the data through <https://s3.amazonaws.com/hubway-data/index.html>, and the data is publicly available from Umich's phpMyAdmin).

***Summary of data (csv file :obtain from public resource)***

To enrich the data information by join weather\_data through date.

	A	B
1	date	bike_usage
2	1/1/2015	127
3	1/2/2015	213
4	1/3/2015	142
5	1/4/2015	112
6	1/5/2015	337
7	1/6/2015	267
8	1/7/2015	296
9	1/8/2015	247
10	1/9/2015	275
11	1/10/2015	162
12	1/11/2015	134
13	1/12/2015	279
14	1/13/2015	387
15	1/14/2015	379
16	1/15/2015	287

**Data structure:**

Organize the restaurant\_data into a tree, the data of restaurants are firstly classified by cuisines, and then each cuisine category is classified by binary tree with a star rating of four or above, or below four stars.

- ♦ python file that constructs trees: *Tree\_structure.py*
- ♦ python file that print the tree: *print\_tree.py*
- ♦ JSON file with the tree: *tree\_restaurants.json*

**Screenshots:**

```

This is a summary of information about the restaurant in Boston:
Is it Asian food?
+-Yes: Is it Chinese food?
  +-Yes: It is Chinese food
    rating>=4.0 restaurants: Q Restaurant, Dumpling Xuan...
    rating<4.0 restaurants: Gourmet Dumpling House, Taiwan Café...
  -No: Is it Japanese food?
    +-Yes: It is Japanese food
      rating>=4.0 restaurants: Douzo, Nagomi Izakaya...
      rating<4.0 restaurants: Yamato II, Tsurutontan Udon Noodle Brasserie - Boston...
    -No: It is Thai food
      rating>=4.0 restaurants: Pho Basil, Kala Thai Cookery...
      rating<4.0 restaurants: Thai Basil Restaurant, Montien Thai Restaurant...
-No: Is it Burger food?
+-Yes: It is Burger food
  rating>=4.0 restaurants: Neptune Oyster, Mike' s Pastry...
  rating<4.0 restaurants: The Beehive, Union Oyster House...
-No: Is it Mexican food?
+-Yes: It is Mexican food
  rating>=4.0 restaurants: Citrus & Salt, Tenoch Mexican...
  rating<4.0 restaurants: Casa Romero, Temazcal Tequila Cantina...
-No: It is Italian food
  rating>=4.0 restaurants: Carmelina' s, Giacomo' s Ristorante - Boston...
  rating<4.0 restaurants: Strega Italiano - Back Bay, Maggiano' s Little Italy...

```

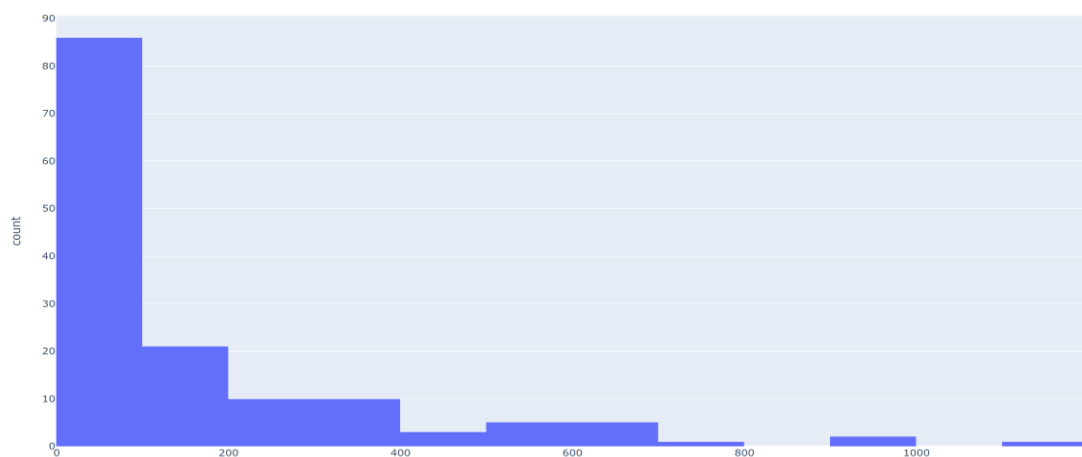
### Interaction and Presentation Options

1. If the user chooses to query restaurant information(for example: mexcian restaurants with star rating of 4 or above)

Rating distribution of eligible restaurants



Distribution of the number of reviews of eligible restaurants



2. If the user chooses to query weather information and bicycles usage. ***For example:***

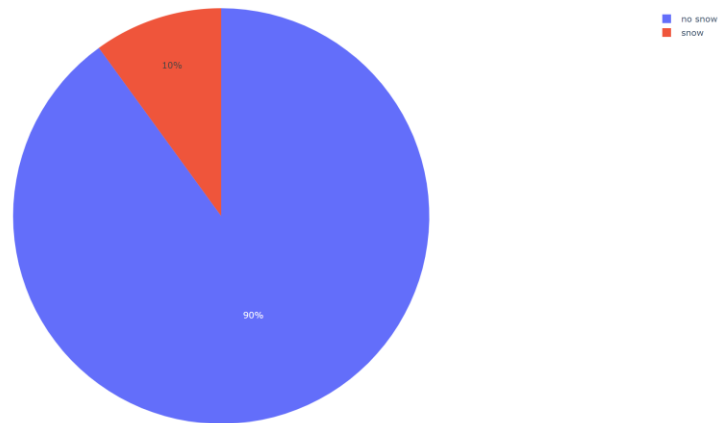
- ♦ The user wants to search for the average temperature on 2020.08.20, then show the change of the month that the date belongs to

the change in avg\_temp during 2020-08



- ♦ The user wants to search for the percentage of snow over a certain period of time

The percentage of snow between 2020-12-01 and 2020-12-30



- ♦ The user wants to search for the bike usage over a certain period of time

The change in bike usage between 2020-07-01 and 2020-08-21

