R e s t r { o l o g y }

A Neighborhood Recommender for Upcoming Eateries & Restaurants

Week 1 - Part 2: Data

Restrology, for the purpose of this project, will be restricted to New York City. Therefore, the recommendations made for restaurants/eatries will be from neighborhoods and boroughs of New York city only. We will need a couple of datasets and will integrate them to get the desired outcome which are as follows.

Datasets required for the Project:

1) New York Data (Boroughs + Neighborhoods)

2) Four Square City Guide Data (Venues)

1) New York Data (Boroughs + Neighborhods):

The first Dataset we will be using would contain all the required geographical data about New York city. Namely, we would be using 'Borough', 'Neighborhood', 'Latitude', 'Longitude' among all the other data elements present in the data. For convenience, we would be using the same data set which was provided to us in Week 3 of this course (Applied Data Science Capstone) https://geo.nyu.edu/catalog/nyu_2451_34572 (https://geo.nyu.edu/catalog/nyu_2451_34572). We will use the same link like we did to load this data from where it is downloaded and hosted https://cocl.us/new_york_dataset (https://cocl.us/new_york_dataset)

New York city has a total of 5 boroughs and 306 neighborhoods. In order to segement the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood. The 'Latitude' and 'Longitude' extracted from this dataset will also be pivotal when we use it perform Clustering using K-Means.

All the relevant data is in the *features* key, which is basically a list of the neighborhoods. If we dive into the elements of this *features* key, we will find all of its components. Below is a snapshot of what a single item would look like.

```
{'type': 'Feature',
 'id': 'nyu_2451_34572.1',
 'geometry': {'type': 'Point',
  'coordinates': [-73.84720052054902, 40.89470517661]},
 'geometry_name': 'geom',
 'properties': {'name': 'Wakefield',
  'stacked': 1,
  'annoline1': 'Wakefield',
  'annoline2': None,
  'annoline3': None,
  'annoangle': 0.0,
  'borough': 'Bronx',
  'bbox': [-73.84720052054902,
   40.89470517661,
   -73.84720052054902,
   40.89470517661]}}
```

2) Four Square City Guide Data (Venues):

Foursquare City Guide, commonly known as Foursquare, is a local search-and-discovery mobile app which provides search results for its users. The app provides personalized recommendations of places to go near a user's current location based on users' previous browsing history and check-in history.

Before we can learn how to retrieve data from the Foursquare database, we need to create a developer account. On May 31st of 2018, Foursquare updated their API, and unlike before, now there are some limitations on how many calls you can make to the API. So when you create a developer account, the default type is the sandbox account, with 950 regular calls per day and 50 premium calls per day, and you can retrieve only one photo and one tip per venue. The account used in this project is a personal account which is still free, and with it you get 99,500 regular calls and 500 premium calls.

That is actually 100 times more calls than the default sandbox account. You'll also get access to over 105 million venues or points of interest, but you still only get two photos and two tips per venue, which is just one more photo and tip compared to the sandbox account. The Foursquare API credentials you need to create are: your Client ID and your Client Secret. You will need to pass these credentials every time you make a call to the API. You can go to the onliine Foursquare documentation and click on API and discover endpoints overview, and there you will find the different groups and endpoints available and whether each endpoint falls under a regular or a premium call.

So using the Foursquare API, we can search for specific type of venues or stores around a given location. it is important to remember that for this data, we make a regular call to the API, and if you have a free personal developer account, you can make up to approximately 99 thousand regular calls per day. We can also learn more about a specific venue or store or shop, like their full address, their working hours, and their menu if they have one, and so on. It's also important to remember that for this data, we would need to make a premium call and with the personal developer account, you can make approximately 500 calls per day. Also with the Foursquare API, we can learn more about a specific Foursquare user, their full name, and any tips or photos that they have posted about venues and stores. For this data, a regular call to the API would be made. Furthermore, we can explore a given location by finding what popular spots exist in the vicinity of the location, and for this data a regular call to the API would be made. And finally, with the Foursquare API, we can explore trending venues around a given location. These are venues with the highest foot traffic at the time this regular call to the API is made.