Capstone Project

THE BATTLE OF NEIGHBORHOODS : Week -1
Search for best coffee shop in New York city

Content

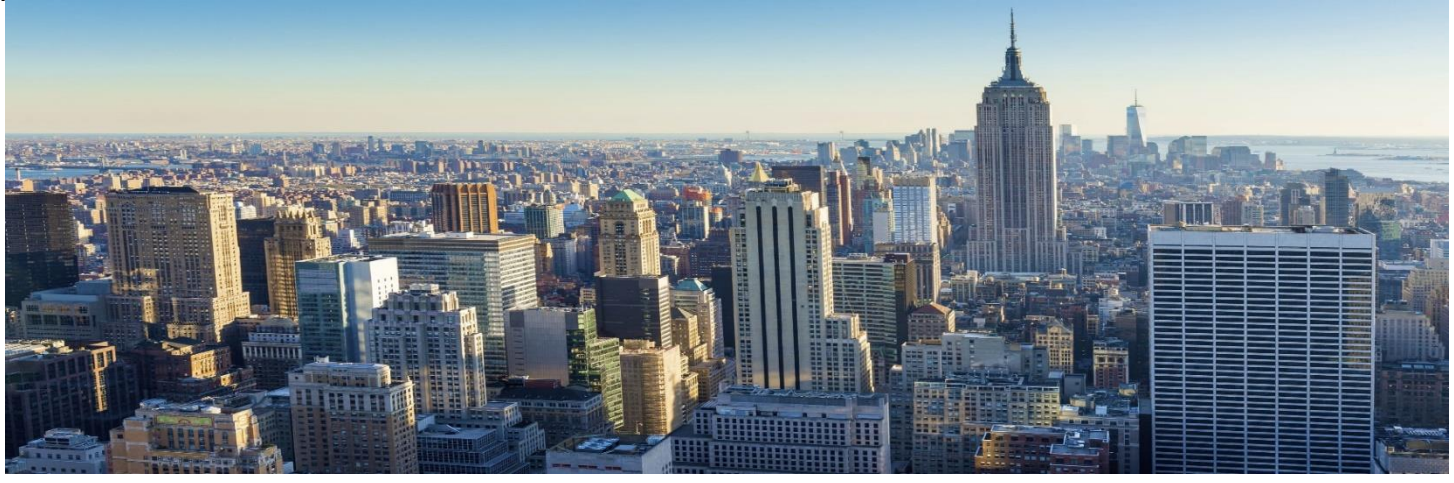
Capstone Project

THE BATTLE OF NEIGHBORHOODS : Week -1
Search for best coffee shop in New York city

1. Introduction/Business Problem

This study will explore the best coffee shop locations throughout the city of New York. New York city is a city that needs no introduction. It is located in New York state with a population of over 8.34 million (census.gov, 2020) and which makes it the most densely populated city in United States. The global financial center index ranked the city as the world's top financial center in 2019 (World Atlas). The city is famous for food and culture and has many public and private schools, bars, restaurants, coffee shops, recreational parks, etc.

However, you probably have heard how hectic life is in the city and overwhelming at times so a nice cup of coffee can help us unwind. But where do we find the best coffee shops or cafe in the city to do just that? Also, stallholders from a west coast wants to expand their coffee business to east coast and have chosen new York city due to population and diversity. Where would be the ideal place for the stallholders to launch coffee shop in the city?



THE BATTLE OF NEIGHBORHOODS : Week -1
Search for best coffee shop in New York city

2. Data Source

In order to analyze and solve the above business problems, data on New York city neighborhoods, boroughs, neighborhood boundaries, latitude, longitude and coffee shops along their ratings and likes are required.

Neighborhoods data: Data provided by the coordinators of "IBM Data Science Professional Certificate" will be utilized to process New York City neighborhoods, boroughs, latitude and longitude. This dataset will be processed, cleaned, and then read it into a pandas dataframe so that it is in a structured format.

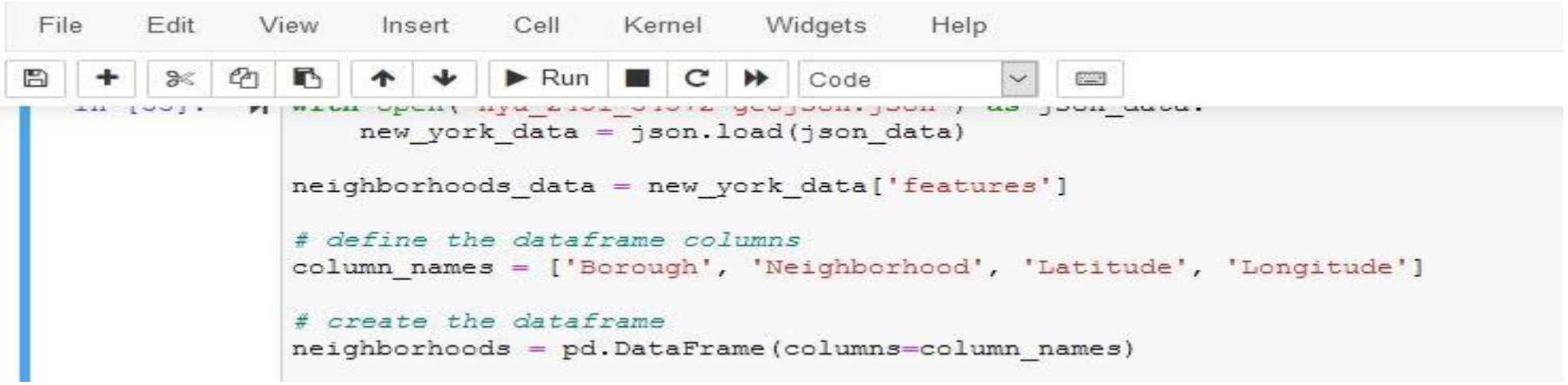
Venues data: Once the neighborhoods data is preprocessed into a dataframe, all data that describes the top venues like restaurants, coffee shops, ratings and likes will be retrieved from Foursquare website via the Request library in Python. The best coffee shops will be then analyzed using Panda libraries.

Neighborhood boundaries: New York city data containing neighborhood boundaries will be obtained from NYC open data source: https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm

Finally I will analyze further and see how many coffee shops there are in each neighborhood and borough and then graph and present the results to the stallholders to make decision.

THE BATTLE OF NEIGHBORHOODS : Week -1

Search for best coffee shop in New York city




```
    new_york_data = json.load(json_data)

neighborhoods_data = new_york_data['features']

# define the dataframe columns
column_names = ['Borough', 'Neighborhood', 'Latitude', 'Longitude']

# create the dataframe
neighborhoods = pd.DataFrame(columns=column_names)
```

```
venues_list=[]
for name, lat, lng in zip(names, latitudes, longitudes):
    print(name)

    # create the API request URL
    url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={
        CLIENT_ID,
        CLIENT_SECRET,
        VERSION,
        lat,
        lng,
        radius,
        LIMIT)

    # make the GET request
    results = requests.get(url).json()["response"]['groups'][0]['items']
```

THE BATTLE OF NEIGHBORHOODS : Week -1
Search for best coffee shop in New York city

2.i Target Audience

Aside water, tea and coffee are the first and second most consumed drinks in the world, respectively. Therefore, Individual and entrepreneurs who are passionate about drinking coffee or opening a coffee shop in New York city would be very interested in this project. The project is also for business owners and stakeholders who want to expand their businesses and wonder how data science could be applied to the questions at hand. Not only stakeholders and business owners but also existing restaurants who are interested in adding coffee to their menu can make use of this project as well.

How the data will be used to solve the problem

- Json file given by the course providers will be loaded, cleaned and processed into a dataframe.
- FourSquare API will be used to locate all venues and then the data will be filtered by coffee shops categories.
- All data present in features label will be generated and descriptive analysis using Panda will be conducted .
- Explore ratings and tips by users and add them to the existing dataframe.
- The best coffee shops will be sorted based on ratings.
- Data will be classified based on kMeans cluster algorithm.
- Finally, analyze and see how many Coffee shops there are in each neighborhood and borough and graph and present the results to the stallholders

- https://data.cityofnewyork.us/City-Government/2020-population/t8c6-3i7b
- World Atlas : https://www.worldatlas.com/articles/the-world-s-top-financial-cities.html
- census.gov: https://www.census.gov/quickfacts/newyorkcitynewyork