Italian restaurants in NYC
Where to eat and enjoy?

Introduction where you discuss the business problem and who would be interested in this project.

The Italian community in New York City is the largest in the country and third at the global level. Additionally, the Italian cuisine is one of the most widely available in the world. Choosing a place to set up an Italian restaurant depends on the competition, density and potential customers to the site. It is my hypothesis that a place with high ratings and likes from the Foursquare database can help setup an operation in the restaurant business.

The potential customer abounds in the region. According to Wikipedia, the largest enclaves of Italian American communities exist in the NYC area.

U.S. communities with the most residents of Italian ancestry[edit]

The top 25 U.S. communities with the highest percentage of people claiming Italian ancestry are:[215]

1. Johnston, Rhode Island 46.7%
2. Monroe, Massachusetts 46.5%
3. Hammonton, New Jersey 45.9%
4. Frankfort, New York (village) 44.7%
5. East Haven, Connecticut 43.1%
6. Roseto, Pennsylvania 41.8%
7. Old Forge, Pennsylvania 41.3%[216]
8. Franklin Square, New York 40.0%
9. North Massapequa, New York 38.9%
10. Frankfort, New York 38.5%
11. Totowa, New Jersey 37.7%
12. Lowellville, Ohio 37.4%
13. Fairfield, New Jersey 37.2%
14. North Providence, Rhode Island 36.6%
15. Thornwood, New York 36.5%
16. South Hackensack, New Jersey 36.3%
17. Hawthorne, New York 36.2%
18. Saugus, Massachusetts 36.1%
19. Nutley, New Jersey 36.0%
20. Jessup, Pennsylvania 35.9%
21. Stoneham, Massachusetts 35.8%
22. Revere, Massachusetts (greatest percentage of any city) 35.7%
23. East Hanover, New Jersey 35.6%
24. Harrison, New York 34.9%
25. Deer Park, New York 34.9%
26. Staten Island, New York (greatest percentage of any county) 34.7%
27. West Paterson, New Jersey 34.3%
28. Valhalla, New York 34.2%
29. Lyndhurst, New Jersey 33.8%

Any restaurateur who wants to open a new Italian eatery can make use of the report in order to pinpoint which borough and neighborhood to locate the restaurant.

Data where you describe the data that will be used to solve the problem and the source of the data.

The data to be used is the Foursquare database with likes and ratings overimposed on the data about boroughs and neighborhoods and its latitude and longitude in order to place each restaurant on a map.

Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

I used the Watson Studio from IBM to analyze the data from the Foursquare API where I created a professional account in order to have more calls from the API than the sandbox account.

The first step is to install all necessary libraries and programs to analyze the data from Foursquare. I used Folium, Nominatin and Geopy to plot the data in a map. In addition, I used pandas and matplotlib. From Foursquare, I get the data using the following script:

```
def geolocation(address):
    geolocator = Nominatim(user_agent="ny_explorer")
    location = geolocator.geocode(address)
    latitude = location.latitude
    longitude = location.longitude
    return latitude,longitude
```

In [38]:

And,

```
def get_venues(lat,lng):

    #set variables
    radius=1000
    LIMIT=10
    CLIENT_ID = 'JSYVL4UF05JDZINLM15K5WB21C0Q0N3XMR5MK4DVBSXLDUJK'# Foursquare ID for me, myself
and I
    CLIENT_SECRET ='CEYJ0SN5PX3QQD2IMXD0IRQFZM2W1GQD4BWVHE0HX2PE24WF' # Foursquare Secret
    VERSION = '20180605' # Foursquare API version as told

    #url to fetch data from foursquare api
    url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={}
,{}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

    # get all the data
    results = requests.get(url).json()
    venue_data=results["response"]['groups'][0]['items']
    venue_details=[]
    for row in venue_data:
        try:
```

```
            venue_id=row['venue']['id']
            venue_name=row['venue']['name']
            venue_category=row['venue']['categories'][0]['name']
            venue_details.append([venue_id,venue_name,venue_category])
        except KeyError:
            pass

    column_names=['ID','Name','Category']
    df = pd.DataFrame(venue_details,columns=column_names)
    return df
```

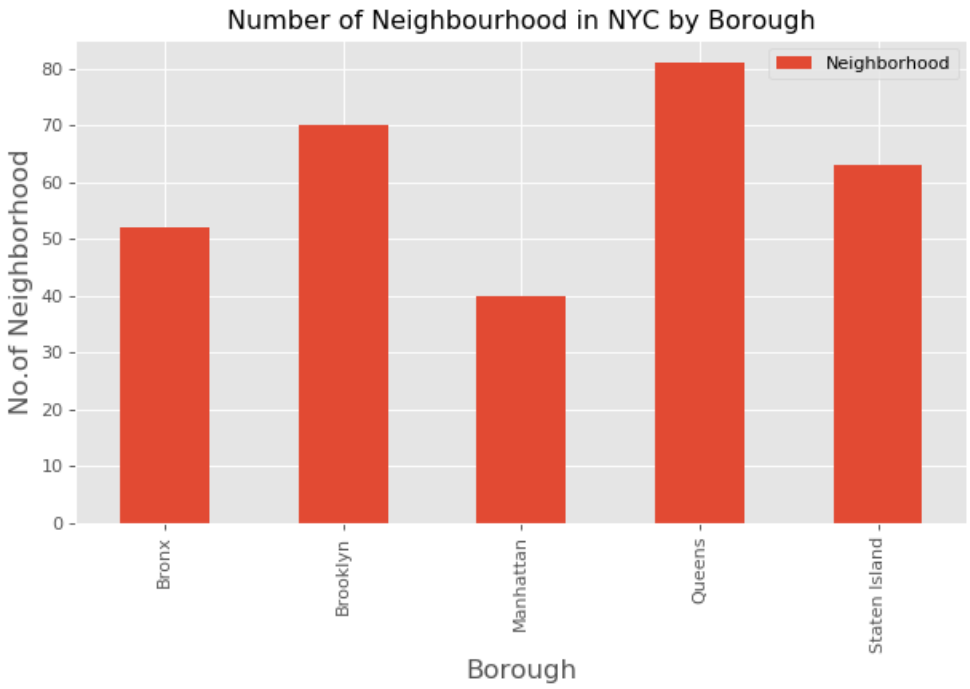
Then I used the data from cocl.us to get the lat and longs of the borough and neighborhoods of NYC.

```
#
def get_new_york_data():
    url='https://cocl.us/new_york_dataset'
    resp=requests.get(url).json()
    # all data is present in features label
    features=resp['features']

    # define the dataframe columns
    column names = ['Borough', 'Neighborhood', 'Latitude', 'Longitude']
    # instantiate the dataframe
    new_york_data = pd.DataFrame(columns=column_names)

    for data in features:
        borough = data['properties']['borough']
        neighborhood_name = data['properties']['name']

        neighborhood_latlon = data['geometry']['coordinates']
        neighborhood_lat = neighborhood_latlon[1]
        neighborhood_lon = neighborhood_latlon[0]

        new york data = new york data.append({'Borough': borough,
                                              'Neighborhood': neighborhood_name,
                                              'Latitude': neighborhood_lat,
                                              'Longitude': neighborhood_lon}, ignore_index=True)

    return new_york_data
new_york_data=get_new_york_data()
new_york_data.head(15)
```

Results section where you discuss the results.

The results are depicted in the following graphs:

The number of neighborhoods by borough is as follows:



Then I used the list of Italian restaurants in order to get the rating from each and average the ratings by borough and neighborhood.

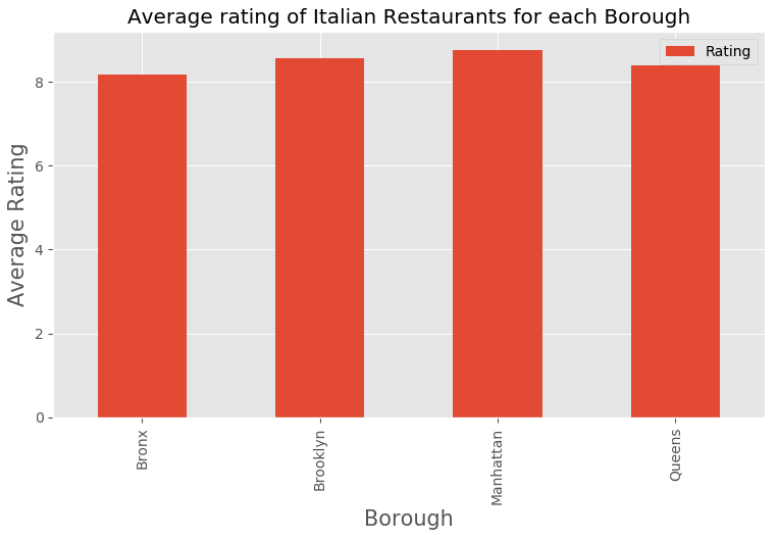
The results I get are the following: the best neighborhood to eat according to ratings are Bushwick and Fort Greene.

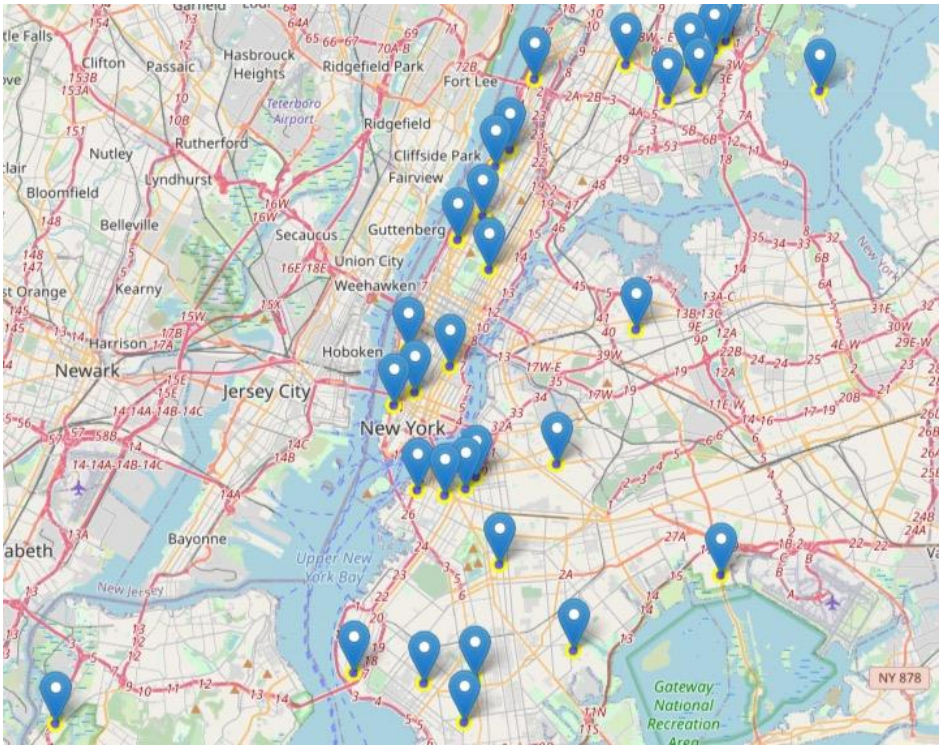
Neighborhood	Average Rating
Bushwick	9.50
Fort Greene	9.40
Cobble Hill	9.20
Boerum Hill	9.10
Greenwich Village	9.05
Hamilton Heights	9.00

Tribeca	8.90
Manhattanville	8.80
Gramercy	8.80
Pelham Parkway	8.70

The best borough to eat in average and where the density is the highest is Manhattan and Brooklyn.

Borough	Average Rating
Manhattan	8.763636
Brooklyn	8.560000
Queens	8.400000
Bronx	8.190909





Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.

I found the data from Foursquare to be sufficient to decide the location of a new restaurant depending of the ratings and likes of the customers who used Foursquare to rate and rank the restaurant

Conclusion section where you conclude the report.

- The best borough to have Italian is Manhattan as it has the highest average in ratings.
- The borough with most Italian restaurants is State Island, therefore the median distance to an Italian restaurant is minimized
- The neighborhood with highest rating is Bushwick