A full report consisting of all of the following components:

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

Business problem discussion

- Starting a restaurant in New York has its own set of legal considerations that are specific both to the restaurant industry and to New York State. These include choosing the proper business entity, obtaining proper state and local licenses and permits, dealing with state and local health and safety regulations, getting adequate insurance, reviewing franchising issues, and dealing with employees.
- New York City's demographics show that it is a large and ethnically diverse metropolis. It is the largest city in the United States with a long history of international immigration. New York City was home to nearly 8.5 million people in 2014,accounting for over 40% of the population of New York State and a slightly lower percentage of the New York metropolitan area, home to approximately 23.6 million. Over the last decade the city has been growing faster than the region. The New York region continues to be by far the leading metropolitan gateway for legal immigrants admitted into the United States.
- The cuisine of New York City comprises many cuisines belonging to various ethnic groups that have entered the United States through the city. Almost all ethnic cuisines are well represented in New York City, both within and outside the various ethnic neighborhoods.New York City was also the founding city of New York Restaurant Week which has spread around the world due to the discounted prices that such a deal offers.In New York City there are over 12,000 bodegas, delis and groceries and many among them are open 24 hours a day, 7 days a week.

Target Customers

The population of New york has grown considerably over the last decades.With it's diverse culture , comes diverse food items. There are many resturants in New york City, each beloning to different categories like Chinese , Indian , French etc.

The real deal is that as much as there are many fine restaurants in New york – Asian, Middle Eastern, Latin, American restaurants and Italian, you can struggle to find good place to start restaurant business without analysis by each component. So as part of this project, we will list and visualize all major parts of New York City that has great restaurants and population density which will ease to our client for opening restaurant in New York.

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

For this project we need the following data :

- **Data source : https://cocl.us/new_york_dataset**
- **Description : This data set contains the required information. And we will use this data set to explore various neighborhoods of new york city.**
- **Data source : https://en.wikipedia.org/wiki/Demographics_of_New_York_City**
- **Description:The Data set contains population growth in New York**
- **Data Source:https://en.wikipedia.org/wiki/Cuisine_of_New_York_City**
- **Description:This data set contains many cuisines belonging to various ethnic groups that have entered the United States through the city.**

- **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.**

Main Components

Exploring New York Data

In this we load the data and explore data from newyork_data.json file. Using either by pandas DataFrame or by Map using geopy and folium libraries as follows.

Load New York DataSet

```
!wget -q -O 'newyork_data.json' https://cocl.us/new_york_dataset
print('Data downloaded!')
```

```
Data downloaded!
```

```
with open('newyork_data.json') as json_data:
    newyork_data = json.load(json_data)
```

```
newyork_data['features']
```

Using Pandas DataFrame:

```
df_newyork.head()
```

:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

```
df_newyork.shape
```

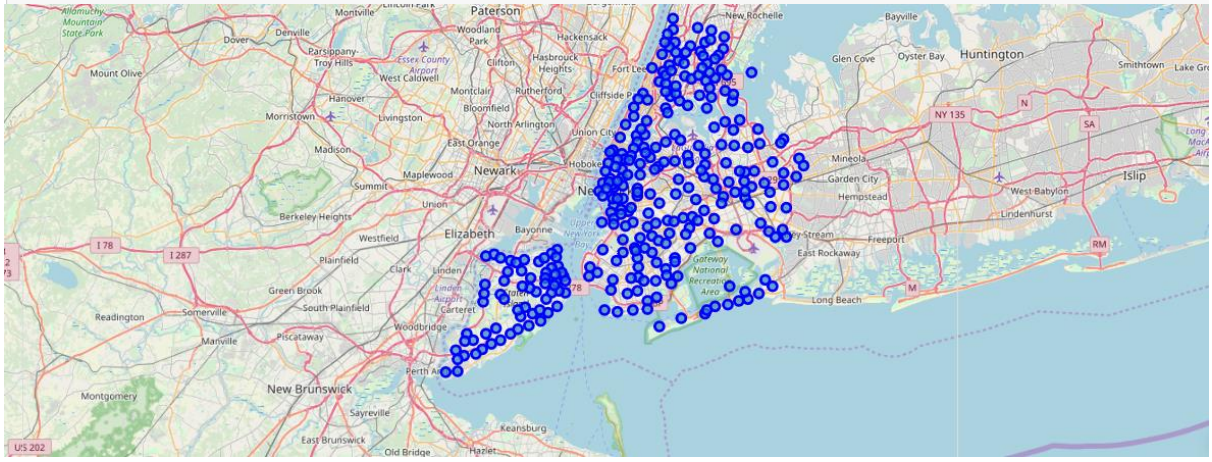
: (306, 4)

Using geopy and folium libraries:

```
address = 'New York City, NY'

geolocator = Nominatim(user_agent="Jupyter")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of New York City are {}, {}.'.format(latitude, longitude))
```

```
The geograpical coordinate of New York City are 40.7127281, -74.0060152.
```

```
# create map of Toronto using latitude and longitude values
map_NewYork = folium.Map(location=[latitude, longitude], zoom_start=10)

# add markers to map
for lat, lng, borough, neighborhood in zip(df_newyork['Latitude'], df_newyork['Longitude'], df_newyork['Borough'], df_newyork['Neighborhood']):
    label = '{}, {}'.format(neighborhood, borough)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_NewYork)

map_NewYork
```



Exploring New York Demographics Data

To analyize New York city Population scrapped the data from NY Demographics Wikipedia pages given above in the data section. We used BeautifulSoup python library. Beautiful Soup is a Python package for parsing HTML and XML documents .

Loading NY DemoGraphics Data

```
from bs4 import BeautifulSoup # package for parsing HTML and XML documents
import csv
website_url = requests.get('https://en.wikipedia.org/wiki/Demographics_of_New_York_City').text
soup = BeautifulSoup(website_url,'lxml')
table = soup.find('table',{'class':'wikitable sortable'})
#print(soup.prettify())

headers = [header.text for header in table.find_all('th')]

table_rows = table.find_all('tr')
rows = []
for row in table_rows:
    td = row.find_all('td')
    row = [row.text for row in td]
    rows.append(row)
```

```
with open('Newyork_Population.csv', 'w') as f:
    writer = csv.writer(f)
    writer.writerow(headers)
    writer.writerows(row for row in rows if row)
```

```
df_NY_Population=pd.read_csv('Newyork_Population.csv')
```

```
df_NY_Population=df_NY_Population.fillna('')
```

```
df_NY_Population
```

:

	Borough	County	Estimate_2017	square_miles	square_km	persons_sq_mi	persons_sq_km
0	The Bronx	Bronx	1,471,160	19,570	42.10	109.04	34,653
1	Brooklyn	Kings	2,648,771	23,900	70.82	183.42	37,137
2	Manhattan	New York	1,664,727	378,250	22.83	59.13	72,033
3	Queens	Queens	2,358,582	31,310	108.53	281.09	21,460
4	Staten Island	Richmond	479,458	23,460	58.37	151.18	8,112
5		City of New York	8,622,698	806.863	302.64	783.83	28,188
6		State of New York	19,849,399	1,547.116	47,214	122,284	416.4

Exploring New York Restaurants Data

Use Forsquare API to Explore NY Restaurant

```
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={}'.format(
        CLIENT_ID,
        CLIENT_SECRET,
        VERSION,
        lat,
        lng,
        radius,
        LIMIT)

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

    # return only relevant information for each nearby venue
    venues_list.append([(
        name,
        lat,
        lng,
        v['venue']['name'],
        v['venue']['location']['lat'],
        v['venue']['location']['lng'],
        v['venue']['categories'][0]['name']) for v in results])

nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
nearby_venues.columns = ['Neighborhood',
             'Neighborhood Latitude',
```

Machine Learning Techniques

We used K means Clustering to Explore NewYork Restaurant Data

Use K means Clustering to Analyze Data

```
# set number of clusters
no_of_clusters = 3


# run k-means clustering
kmeans_model = KMeans(n_clusters=no_of_clusters, random_state=0).fit(df_man_brook_restaurant)

# check cluster labels generated for each row in the dataframe
kmeans_model.labels_
```

```
array([2, 1, 2, 0, 2, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 2, 0, 0, 1, 0, 0,
       2, 0, 1, 0, 1, 2, 0, 1, 0, 1, 2, 1, 2, 0, 0, 0, 2, 0, 1, 0, 0, 0,
       0, 1, 0, 2, 1, 0, 0, 2, 1, 2, 0, 2, 0, 2, 0, 0, 2, 1, 0, 0, 2, 2,
       1, 1, 0, 2, 1, 0, 1, 2, 2, 1, 2, 0, 1, 0, 1, 0, 2, 2, 2, 1, 0, 0,
       0, 1, 2, 0, 2, 1, 0, 2, 2, 0, 2, 2, 0, 2, 1, 2, 0, 2, 2, 0, 1, 2],
      dtype=int32)
```

- **Results section where you discuss the results.**

Tibetan Restaurant	Turkish Restaurant	Udon Restaurant	Ukrainian Restaurant	Vegetarian / Vegan Restaurant	Venezuelan Restaurant	Vietnamese Restaurant	Total	Total Sum
0.000000e+00	0.217391	2.173913e-02	-1.734723e-18	0.478261	-3.469447e-18	3.260870e-01	22.108696	44.217391
1.040834e-17	0.034483	-6.938894e-18	-1.734723e-18	0.034483	-3.469447e-18	-5.551115e-17	9.344828	18.689655
8.571429e-02	0.228571	8.571429e-02	2.857143e-02	0.714286	5.714286e-02	6.000000e-01	32.942857	65.885714

Cluster0 : The Total and Total Sum of cluster0 has smallest value. It shows that the cluster0 has few restaurants in NEWYork.

Cluster1 : The Total and Total Sum of cluster1 has Middle value. It shows that the markets are Cluster 1 is growing restaurants business in NEWYork.

Cluster2 : The Total and Total Sum of cluster1 has highest value. It shows that the Cluster2 has high number of restaurants in NEWYork

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

  In Manhattan and Brooklyn restaurants of cuisines of many countries are available. So if risk can be taken with great menu on board. It also shows people love eating cuisines of various countries.

- **Conclusion section where you conclude the report.**
- Brooklyn and Manhattan has great opportunity of restaurant business. Bronx, Queens are growing in restaurant business