

Final_Note

August 24, 2019

0.1 The Battle of the Neighborhoods - Week 2

0.2 Introduction & Business Problem :

0.3 Problem Background: Open a new business (Case: Restaurant)

The City of New York, usually called either New York City (NYC) or simply New York (NY), is the most populous city in the United States. With an estimated 2018 population of 8,398,748 distributed over a land area of about 302.6 square miles (784 km²), New York is also the most densely populated major city in the United States. Located at the southern tip of the state of New York, the city is the center of the New York metropolitan area, the largest metropolitan area in the world by urban landmass and one of the world's most populous megacities, with an estimated 19,979,477 people in its 2018 Metropolitan Statistical Area and 22,679,948 residents in its Combined Statistical Area. A global power city, New York City has been described as the cultural, financial, and media capital of the world, and exerts a significant impact upon commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. The city's fast pace has inspired the term New York minute. Home to the headquarters of the United Nations, New York is an important center for international diplomacy.

Situated on one of the world's largest natural harbors, New York City consists of five boroughs, each of which is a separate county of the State of New York. The five boroughs – Brooklyn, Queens, Manhattan, The Bronx, and Staten Island – were consolidated into a single city in 1898. The city and its metropolitan area constitute the premier gateway for legal immigration to the United States. As many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world. New York City is home to more than 3.2 million residents born outside the United States, the largest foreign-born population of any city in the world. As of 2019, the New York metropolitan area is estimated to produce a gross metropolitan product (GMP) of US\$1.9 trillion. If greater New York City were a sovereign state, it would have the 12th highest GDP in the world. New York is home to the highest number of billionaires of any city in the world.

So, the market is truly competitive in New York City and doing business is either a risk or an opportunity.

0.4 Problem description:

Restaurant is a public place. Provide Food and Beverage on a commercial basis. This is open to all to take refreshment, Food and beverage. Everybody can take food and Beverage against money. Restaurant offer service of Food and Beverage desires to satisfy the Guest. Actually, Guest take Rest in restaurant and pay Rent for the having refreshment, food and beverage. Restaurant comes from the word of "Rest and Rent". "Rest & Rent" those words consisting the word restaurant. Where the guest/client/peoples take their Food & beverage The city of New York is known for

its famous restaurant such as French restaurants, Italian restaurants, Asian restaurants, African restaurants, and so on.

0.5 In this project we will analyze the city of New York.

Data 1 : New York City has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will need a dataset that contains the 5 boroughs, the neighborhoods existing in these boroughs and the geographical coordinates of each neighborhood.

Dataset sources:

https://geo.nyu.edu/catalog/nyu_2451_34572

Data 2 : We will use the data relative to the Farmers Markets and Food Boxes.

Dataset sources: - <https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2> - <https://www.grownyc.org/greenmarketco/foodbox>

Data 3 : We will use the dataset related to New York City Population and Cuisines(Restaurants).

Data sources:

- https://en.wikipedia.org/wiki/New_York_City
- https://en.wikipedia.org/wiki/Economy_of_New_York_City
- https://en.wikipedia.org/wiki/Portal:New_York_City
- https://en.wikipedia.org/wiki/Cuisine_of_New_York_City
- https://en.wikipedia.org/wiki/List_of_Michelin_starred_restaurants_in_New_York_City

Data 4: We will use New York city geographical coordinates data and the Foursquare API to explore neighborhoods in New York City.

0.6 I. Download and Explore New York city geographical coordinates dataset

```
[1]: import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files

!conda install -c conda-forge geopy --yes # uncomment this line if you haven't completed the_
→Foursquare API lab
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors
```

```
!conda install -c conda-forge folium=0.5.0 --yes # uncomment this line if you haven't
    ↳ completed the Foursquare API lab
import folium # map rendering library

import csv # implements classes to read and write tabular data in CSV form

print('Libraries imported.')
```

Solving environment: | Killed

Solving environment: | Killed

Libraries imported.

```
[2]: !wget -q -O 'newyork_data.json' https://ibm.box.com/shared/static/
    ↳ fbpbwbovar7lf8p5sgddm06cgipa2rxpe.json
print('Data downloaded!')
```

Data downloaded!

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

[4]: neighborhoods_data = newyork_data['features']

[5]: neighborhoods_data[0]

[5]: {'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]}}
```

0.7 Transform the data

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

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

```
[7]: neighborhoods
```

```
[7]: Empty DataFrame
Columns: [Borough, Neighborhood, Latitude, Longitude]
Index: []
```

```
[8]: for data in neighborhoods_data:
    borough = neighborhood_name = data['properties']['borough']
    neighborhood_name = data['properties']['name']

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

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

```
[9]: neighborhoods.head()
```

```
[9]: 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
```

```
[10]: print('This above dataframe has {} boroughs and {} neighborhoods.'.format(
        len(neighborhoods['Borough'].unique()),
        neighborhoods.shape[0]
    )
)
```

This above dataframe has 5 boroughs and 306 neighborhoods.

```
[11]: neighborhoods.to_csv('Newyork_City_GEO1.csv',index=False)
```

```
[12]: 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.

0.8 Map of New York and it's neighborhood

```
[13]: # 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(neighborhoods['Latitude'],
neighborhoods['Longitude'], neighborhoods['Borough'], neighborhoods['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
```

```
[13]: <folium.folium.Map at 0x7fa67ef9a1d0>
```

0.9 II. Scrapping Population data of New York City and it's neighborhoods

```
[ ]:
```

```
[14]: conda install -c anaconda beautifulsoup4
```

Solving environment: done

```
==> WARNING: A newer version of conda exists. <==
current version: 4.5.11
latest version: 4.7.11
```

Please update conda by running

```
$ conda update -n base -c defaults conda
```

Package Plan

environment location: /home/jupyterlab/conda/envs/python

added / updated specs:

- beautifulsoup4

The following packages will be downloaded:

package	build	
certifi-2019.6.16	py36_1	156 KB anaconda

The following packages will be UPDATED:

certifi: 2019.6.16-py36_1 conda-forge --> 2019.6.16-py36_1 anaconda
openssl: 1.0.2r-h14c3975_0 conda-forge --> 1.0.2s-h7b6447c_0 anaconda

Downloading and Extracting Packages

certifi-2019.6.16 | 156 KB |
| 100%

Preparing transaction: done

Verifying transaction: done

Executing transaction: done

Note: you may need to restart the kernel to use updated packages.

```
[15]: import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files

!conda install -c conda-forge geopy --yes # uncomment this line if you haven't completed the_
→Foursquare API lab
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values

import requests # library to handle requests
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
```

```
import matplotlib.colors as colors
import matplotlib.pyplot as plt

# conda install -c anaconda beautiful-soup --yes
from bs4 import BeautifulSoup # package for parsing HTML and XML documents

import csv # implements classes to read and write tabular data in CSV form

print('Libraries imported.')
```

Solving environment: done

```
==> WARNING: A newer version of conda exists. <==
current version: 4.5.11
latest version: 4.7.11
```

Please update conda by running

```
$ conda update -n base -c defaults conda
```

```
## Package Plan ##
```

```
environment location: /home/jupyterlab/conda/envs/python
```

```
added / updated specs:
- geopy
```

The following packages will be downloaded:

package	build	
certifi-2019.6.16	py36_1	149 KB conda-forge

The following packages will be UPDATED:

```
certifi: 2019.6.16-py36_1 anaconda --> 2019.6.16-py36_1 conda-forge
```

The following packages will be DOWNGRADED:

```
openssl: 1.0.2s-h7b6447c_0 anaconda --> 1.0.2r-h14c3975_0 conda-forge
```

Downloading and Extracting Packages

```
certifi-2019.6.16 | 149 KB |  
->##### | 100%  
Preparing transaction: done  
Verifying transaction: done  
Executing transaction: done  
Libraries imported.
```

```
[16]: pip install lxml
```

```
Requirement already satisfied: lxml in  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages (4.4.1)  
Note: you may need to restart the kernel to use updated packages.
```

```
[17]: pip install lxml
```

```
Requirement already satisfied: lxml in  
/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages (4.4.1)  
Note: you may need to restart the kernel to use updated packages.
```

```
[18]: import requests  
import lxml.html as lh  
import urllib.request
```

```
[19]: source = requests.get('https://en.wikipedia.org/wiki/Demographics_of_New_York_City').text  
soup = BeautifulSoup(source, 'lxml')
```

```
[21]: 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_POPULATION1.csv', 'w') as f:  
    writer = csv.writer(f)  
    writer.writerow(headers)  
    writer.writerows(row for row in rows if row)
```


0.10 Load data from the dataset

```
[22]: Pop_data=pd.read_csv('NewYork_POPULATION1.csv')
Pop_data.drop(Pop_data.columns[[7,8,9,10,11]], axis=1,inplace=True)
print('Data has been downloaded!')
```

Data has been downloaded!

0.11 Remove whitespaces and renaming some columns

```
[23]: Pop_data.columns = Pop_data.columns.str.replace(' ', '')
Pop_data.columns = Pop_data.columns.str.replace('\n','')
Pop_data.rename(columns={'Borough':'persons_sq_mi','County':'persons_sq_km'},_
→inplace=True)
Pop_data
```

```
[23]: NewYorkCitysfiveboroughsvte\n  Jurisdiction\n \
0          The Bronx\n      \n Bronx\n
1          Brooklyn\n      \n Kings\n
2          Manhattan\n      \n New York\n
3          Queens\n      \n Queens\n
4          Staten Island\n      \n Richmond\n
5          City of New York      8,622,698
6          State of New York      19,849,399
7 Sources:[14] and see individual borough article...      NaN

Population\n GrossDomesticProduct\n Landarea\n Density\n persons_sq_mi \
0 1,471,160\n      28.787\n      19,570\n      42.10\n      109.04\n
1 2,648,771\n      63.303\n      23,900\n      70.82\n      183.42\n
2 1,664,727\n      629.682\n      378,250\n      22.83\n      59.13\n
3 2,358,582\n      73.842\n      31,310\n      108.53\n      281.09\n
4 479,458\n      11.249\n      23,460\n      58.37\n      151.18\n
5 806.863      93,574      302.64      783.83      28,188
6 1,547.116      78,354      47,214      122,284      416.4
7      NaN      NaN      NaN      NaN      NaN

squarekm persons/sq.mi persons/sq.km\n
0      NaN      NaN      NaN
1      NaN      NaN      NaN
2      NaN      NaN      NaN
3      NaN      NaN      NaN
4      NaN      NaN      NaN
5      NaN      NaN      NaN
6      NaN      NaN      NaN
7      NaN      NaN      NaN
```

```
[24]: Pop_data.rename(columns = {'NewYorkCitysfiveboroughsvte\n' : 'Borough',
→'Jurisdiction\n':'County',
```