The Battle of Neighborhoods Week 2

In [ ]:

```
import numpy as np # library to handle data in a vectorized manner
import time
import pandas as pd # library for data analsysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

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


!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 latitud
e and longitude values

!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 folium # map rendering library
from folium import plugins

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

import seaborn as sns

# import k-means from clustering stage
from sklearn.cluster import KMeans
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```
print('Libraries imported.')
```

```
Solving environment: |
```

1. Introduction Section :

Discussion of the business problem and the audience who would be interested in this project.

1.1 Scenario and Background

I am a student living in Singapore. I currently live within walking distance to Downtown "Telok Ayer MRT metro station" therefore I have access to good public transportation to work. Likewise, I enjoy many ammenities in the neighborhood , such as international cousine restaurants, cafes, food shops and entertainment. I have been offered a great opportunity to work in Manhattan, NY. Although, I am very excited about it, I am a bit stress toward the process to secure a comparable place to live in Manhattan. Therefore, I decided to apply the learned skills during the Coursera course to explore ways to make sure my decision is factual and rewarding. Of course, there are alternatives to achieve the answer using available Google and Social media tools, but it rewarding doing it myself with learned tools.

1.2 Problem to be resolved:

The challenge to resolve is being able to find a rental apartment unit in Manhattan NY that offers similar characteristics and benefits to my current situation. Therefore, in order to set a basis for comparison, I want to find a renta unit subject to the following conditions:

```
Apartment with min 2 bedrooms with monthly rent not to exceed US
$7000/month
Unit located within walking distance (<=1.0 mile, 1.6 km) from a
 subway metro station in Manhattan
Area with ammenities and venues similar to the ones described fo
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

r current location (See item 2.1)

1.3 Interested Audience

I believe this is a relevant project for a person or entity considering moving to a major city in Europe, US or Asia, since the approach and methodologies used here are applicable in all cases. The use of FourSquare data and mapping techniques combined with data analysis will help resolve the key questions arisen. Lastly, this project is a good practical case toward the development of Data Science skills.

2. Data Section:

Description of the data and its sources that will be used to solve the problem

2.1 Data of Current Situation

I Currently reside in the neighborhood of 'Mccallum Street' in Downtonw Singapore. I use Foursquare to identify the venues around the area of residence which are then shown in the Singapore map shown in methodology and execution in section 3.0 . It serves as a reference for comparison with the desired future location in Manhattan NY

2.2 Data Required to resolve the problem

In order to make a good choice of a similar apartment in Manhattan NY, the following data is required: List/Information on neighborhoods form Manhattan with their Geodata (latitud and longitud. List/Information about the subway metro stations in Manhattan with geodata. Listed apartments for rent in Manhattan area with descriptions (how many beds, price, location, address) Venues and ammenities in the Manhattan neighborhoods (e.g. top 10) 2.3 sources and manipulation The list of Manhattan neighborhoods is worked out during LAb exercise during the course. A csv file was created which will be read in order to create a dataframe and its mapping. The csv file 'mh_neigh_data.csv' has the following below data structure. The file will be directly

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

read to the Jupiter Notebook for convenience and space savings. The clustering of neighborhoods and mapping will be shown however. An algorythm was used to determine the geodata from Nominatim . The actual algorythm coding may be shown in 'markdown' mode becasues it takes time to run.

mh_neigh_data.tail():

	Borough	Neighborhood	Latitude	Longitude
35	Manhattan	Turtle Bay	40.752042	-73.967708
36	Manhattan	Tudor City	40.746917	-73.971219
37	Manhattan	Stuyvesant Town	40.731000	-73.974052
38	Manhattan	Flatiron	40.739673	-73.990947
39	Manhattan	Hudson Yards	40.756658	-74.000111

A list of Manhattan subway metro stops was complied in Numbers (Apple excel) and it was complemeted with wikipedia data (https://en.wikipedia.org/wiki/List_of_New_York_City_Subway_stations_in_Manhattan) and information from NY Transit authority and Google maps (https://www.google.com/maps/search/manhattan+subway+metro+stations/@40.7837297,-74.1033 for a final consolidated list of subway stops names and their address. The geolocation was obtained via an algorythm using Nominatim. Details will be shown in the execution of methodolody in section 3.0. The subway csv file is "MH_subway.csv" and the data structure is: mhsub.tail(): sub_station sub_address lat long

17 190 Street Subway Station Bennett Ave, New York, NY 10040, USA 40.858113 -73.932983

18 59 St-Lexington Av Station E 60th St, New York, NY 10065, USA 40.762259 -73.966271

19 57 Street Station New York, NY 10019, United States 40.764250 -73.954525

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

20 14 Street / 8 Av New York, NY 10014, United States 40.730862 -73.987156

21 MTA New York City 525 11th Ave, New York, NY 10018, USA 40.759809 -73.999282

A list of places for rent was collected by web-browsing real estate companies in Manhattan : http://www.rentmanhattan.com/index.cfm?page=search&state=results https://www.nestpick.com/search?city=new-york&page=1&order=relevance&district=manhattan&gclid=CjwKCAiAjNjgBRAgEiwAGLlf2hkP3A-cPxjZYkURqQEswQK2jKQEpv_MvKcrIhRWRzNkc_r-fGi0lxoCA7cQAvD_BwE&type=apartment&display=list https://www.realtor.com/apartments/Manhattan_NY A csv file was compiled with the rental place that indicated: areas of Manhattan, address, number of beds, area and monthly rental price. The csv file "nnnn.csv" had the following below structure. An algorythm was used to create all the geodata using Nominatim, as shown in section 3.0. The actual algorythm coding may be shown in 'markdown' mode becasues it takes time to run. With the use of geolocator = Nominatim() , it was possible to determine the latitude and longiude for the subway metro locations as well as for the geodata for each rental place listed. The loop algorythms used are shown in the execution of data in section 3.0 "Great_circle" function from geolocator was used to calculate distances between two points , as in the case to calculate average rent price for units around each subway station and at 1.6 km radius. Foursquare is used to find the avenues at Manhattan neighborhoods in general and a cluster is created to later be able to search for the venues depending of the location shown.

2.3 How the data will be used to solve the problem

The data will be used as follows: Use Foursquare and geopy data to map top 10 venues for all Manhattan neighborhoods and clustered in groups (as per Course LAB) Use foursquare and geopy data to map the location of subway metro stations , separately and on top of the above clustered map in order to be able to identify the venues and ammenities near each metro station, or explore each subway location separately Use Foursquare and geopy data to map the location of rental places, in some form, linked to the subway locations. create a map that depicts, for instance, the average rental price per square ft, around a radious of 1.0 mile (1.6 km) around each subway station - or a similar metrics. I will be able to quickly point to the popups to know the relative price per subway area. Addresses from rental locations will be converted to geodata(

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

lat, long) using Geopy-distance and Nominatim. Data will be searched in open data sources if available, from real estate sites if open to reading, libraries or other government agencies such as Metro New York MTA, etc.

2.5 Mapping of Data

The following maps were created to facilitate the analysis and the choice of the palace to live. Manhattan map of Neighborhoods manhattan subway metro locations Manhattan map of places for rent Manhattan map of clustered venues and neighborhoods Combined maps of Manhattan rent places with subway locations Combined maps of Manhattan rent places with subway locations and venues clusters

3. Methodology section:

This section represents the main component of the report where the data is gathered, prepared for analysis. The tools described are used here and the Notebook cells indicates the execution of steps.

The analysis and the stragegy:

The strategy is based on mapping the above described data in section 2.0, in order to facilitate the choice of at least two candidate places for rent. The choice is made based on the demands imposed : location near a subway, rental price and similar venues to Singapore. This visual approach and maps with popups labels allow quick identification of location, price and feature, thus making the selection very easy.

The procesing of these DATA and its mapping will allow to answer the key questions to make a decision:

```
what is the cost of available rental places that meet the demand
s?
what is the cost of rent around a mile radius from each subway m
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```
etro station?
what is the area of Manhattan with best rental pricing that meet
s criteria established?
What is the distance from work place ( Park Ave and 53 rd St) an
d the tentative future rental home?
What are the venues of the two best places to live? How the pric
es compare?
How venues distribute among Manhattan neighborhoods and around m
etro stations?
Are there tradeoffs between size and price and location?
Any other interesting statistical data findings of the real esta
te and overall data.
```

METHODOLOY EXECUTION - Mapping Data

Singapore Map - Current residence and venues in neighborhood

for comparison to future Manhattan renting place

In [2]:
```
# Shenton Way, District 01, Singapore
address = 'Mccallum Street, Singapore'
geolocator = Nominatim()
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Singapore home are {}, {}.'.format
(latitude, longitude))
```

```
The geograpical coordinate of Singapore home are 1.2784801, 103.849371
7.
```

```
/opt/conda/envs/Python36/lib/python3.6/site-packages/ipykernel/__main_
_.py:3: DeprecationWarning: Using Nominatim with the default "geopy/1.2
1.0" `user_agent` is strongly discouraged, as it violates Nominatim's T
oS https://operations.osmfoundation.org/policies/nominatim/ and may pos
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```
sibly cause 403 and 429 HTTP errors. Please specify a custom `user_agen
t` with `Nominatim(user_agent="my-application")` or by overriding the d
efault `user_agent`: `geopy.geocoders.options.default_user_agent = "my-
application"`. In geopy 2.0 this will become an exception.
  app.launch_new_instance()
```

In [3]:
```
neighborhood_latitude=1.2792655
neighborhood_longitude=103.8480938
```

Dial FourSquare to find venues around current residence in Singapore

In [4]:
```
CLIENT_ID = 'DVCTZDPDYXTS0BRJFPLMHM323APGXNWZI5PLRQ1VC0CFLF1T' # your F
oursquare ID
CLIENT_SECRET = '5NWAGXRLXIXAV0L3DNYY1EPIHNMAAAIZFDFELYSYXL5LFWL1' # yo
ur Foursquare Secret
VERSION = '20180605' # Foursquare API version
```

In [5]:
```
LIMIT = 100 # limit of number of venues returned by Foursquare API
radius = 500 # define radius
# create URL
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&clien
t_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    neighborhood_latitude,
    neighborhood_longitude,
    radius,
    LIMIT)
url # display URL
```

Out[5]:
```
'https://api.foursquare.com/v2/venues/explore?&client_id=DVCTZDPDYXTS0B
RJFPLMHM323APGXNWZI5PLRQ1VC0CFLF1T&client_secret=5NWAGXRLXIXAV0L3DNYY1E
PIHNMAAAIZFDFELYSYXL5LFWL1&v=20180605&ll=1.2792655,103.8480938&radius=5
00&limit=100'
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

In [6]:
```
# results display is hidden for report simplification
results = requests.get(url).json()
#results
```

function that extracts the category of the venue - borrow from the Foursquare lab.

In [7]:
```
def get_category_type(row):
    try:
        categories_list = row['categories']
    except:
        categories_list = row['venue.categories']

    if len(categories_list) == 0:
        return None
    else:
        return categories_list[0]['name']
```

In [8]:
```
venues = results['response']['groups'][0]['items']
SGnearby_venues = json_normalize(venues) # flatten JSON
# filter columns
filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
SGnearby_venues =SGnearby_venues.loc[:, filtered_columns]
# filter the category for each row
SGnearby_venues['venue.categories'] = SGnearby_venues.apply(get_category_type, axis=1)
# clean columns
SGnearby_venues.columns = [col.split(".")[-1] for col in SGnearby_venues.columns]

SGnearby_venues.shape
```

Out[8]:
```
(100, 4)
```

In [9]:
```
# Venues near current Singapore residence place
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```
SGnearby_venues.head(10)
```

Out[9]:

	name	categories	lat	lng
0	Napoleon Food & Wine Bar	Wine Bar	1.279925	103.847333
1	Native	Cocktail Bar	1.280135	103.846844
2	Pepper Bowl	Asian Restaurant	1.279371	103.846710
3	Park Bench Deli	Deli / Bodega	1.279872	103.847287
4	Muchachos	Burrito Place	1.279072	103.847026
5	Mellower Coffee	Café	1.277814	103.848188
6	Freehouse	Beer Garden	1.281254	103.848513
7	Sofitel So Singapore	Hotel	1.280017	103.849813
8	Dumpling Darlings	Dumpling Restaurant	1.280483	103.846942
9	PS.Cafe	Café	1.280468	103.846264

Map of Singapore residence place with venues in Neighborhood - for reference

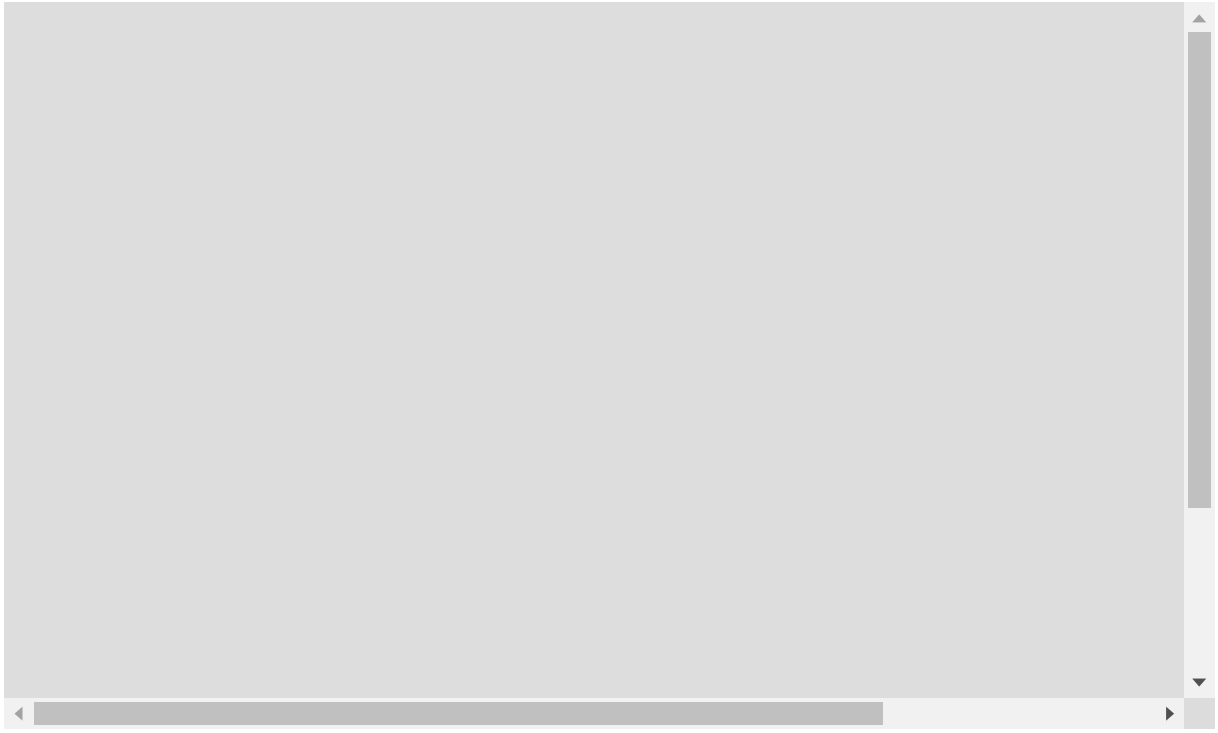
In [10]:
```
latitude=1.2792655
longitude=103.8480938
# create map of Singapore place  using latitude and longitude values
map_sg = folium.Map(location=[latitude, longitude], zoom_start=18)
# add markers to map
for lat, lng, label in zip(SGnearby_venues['lat'], SGnearby_venues['ln
g'], SGnearby_venues['name']):
    label = folium.Popup(label, parse_html=True)
    folium.RegularPolygonMarker(
        [lat, lng],
        number_of_sides=30,
        radius=7,
        popup=label,
        color='blue',
```

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

```
        fill_color='#0f0f0f',
        fill_opacity=0.6,
    ).add_to(map_sg)

map_sg
```

Out[10]:



5.0 DISCUSSION

In general, I am positively impressed with the overall organization, content and lab works presented during the Coursera IBM Certification Course I feel this Capstone project presented me a great opportunity to practice and apply the Data Science tools and methodologies learned. I have created a good project that I can present as an example to show my potential. I feel I have acquired a good starting point to become a professional Data Scientist and I will continue exploring to creating examples of practical cases.

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD

6.0 CONCLUSIONS

I feel rewarded with the efforts, time and money spent. I believe this course with all the topics covered is well worthy of appreciation. This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools.¶ The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision thoroughly and with confidence. I would recommend for use in similar situations. One must keep abreast of new tools for DS that continue to appear for application in several business fields.

In [ ]:

Create PDF in your applications with the Pdfcrowd HTML to PDF API PDFCROWD