CAPSTONE PROJECT

THE BATTLE OF NEIGHBORHOODS

Looking for the best place to open a middle – high income restaurant in Bogotá

1 INTRODUCTION

After 8 modules of theory information y practical exercises, I have enough tools to be able to carry out a basic behavioral research of certain parameters, using data that can be consulted in public or private institutions. The concepts that will be used in this work will serve for a review from a socioeconomic perspective for decision making.

Now that you have been equipped with the skills and the tools to use location data to explore a geographical location, over the course of two weeks, you will have the opportunity to be as creative as you want and come up with an idea to leverage the Foursquare location data to explore or compare neighborhoods or cities of your choice or to come up with a problem that you can use the Foursquare location data to solve. If you cannot think of an idea or a problem, here are some ideas to get you started:

1. In Module 3, we explored New York City and the city of Toronto and segmented and clustered their neighborhoods. Both cities are very diverse and are the financial capitals of their respective countries. One interesting idea would be to compare the neighborhoods of the two cities and determine how similar or dissimilar they are. Is New York City more like Toronto or Paris or some other multicultural city? I will leave it to you to refine this idea.

2. In a city of your choice, if someone is looking to open a restaurant, where would you recommend that they open it? Similarly, if a contractor is trying to start their own business, where would you recommend that they setup their office?

These are just a couple of many ideas and problems that can be solved using location data in addition to other datasets. No matter what you decide to do, make sure to provide sufficient justification of why you think what you want to do or solve is important and why would a client or a group of people be interested in your project.

Question to be answer: **Looking for the best place to open a middle – high income restaurant in Bogotá**

1.1 BUSINESS PROBLEM

Bogotá is the capitol of Colombia, it`s the most populate city in the country with a total of almost eight million habitants. The total surface is 1.775 km2 and the GDP is in US$201.174 million. The city is divided in 20 localities and has 1.922 boroughs.

Each of these localities has its own particularities, security, infrastructure, access points and socioeconomic difference.

The data will help analyze the best places regarding number of people how life in each borough in order to have a better understanding of the real situation in Bogota and the behavior of the consumers in the city.

1.2 TARGET AUDIENCE

For purposes of the project, the data corresponds to all the official restaurants and bars that are officially inscribe in the commerce chamber of Bogota. Therefore, the small businesses and informal commerce are not being take into account.

The target audience are middle and high income individuals. Which have a high purchasing power and are concentrated in 4 localities, that has a low percentage of habitants.

2 DATA SECTION

The data used for the project is based on official information of the city.

https://datosabiertos.bogota.gov.co/dataset/b0c66a77-3230-4d0c-a119-dead7f9b8b8e/resource/176c0f88-2d75-4d65-a1d6-4de527e7608a/download/egba.geojson

From Wikipedia maps and other sources that has the number, location and type of food of each restaurant and bar inscribed officially in the public and private control institutions.

First step is to install all the libraries and commands that are going to be used.

```
In [1]: !pip install wikipedia
        !pip install folium
        !pip install geopy
        !pip install geopandas
        !pip install geojson
```

```
Requirement already satisfied: wikipedia in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (1.4.0)
Requirement already satisfied: requests<3.0.0,>=2.0.0 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from wikipedia) (2.24.0)
Requirement already satisfied: beautifulsoup4 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from wikipedia) (4.9.1)
Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from requests<3.0.0,>=2.0.0->wikipedia) (2020.6.20)
Requirement already satisfied: idna<3,>=2.5 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from requests<3.0.0,>=2.0.0->wikipedia) (2.9)
Requirement already satisfied: chardet<4,>=3.0.2 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from requests<3.0.0,>=2.0.0->wikipedia) (3.0.4)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from requests<3.0.0,>=2.0.0->wikipedia) (1.
25.9)
Requirement already satisfied: soupsieve>1.2 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from beautifulsoup4->wikipedia) (2.0.1)
Requirement already satisfied: folium in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (0.5.0)
Requirement already satisfied: requests in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from folium) (2.24.0)
Requirement already satisfied: jinja2 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from folium) (2.11.2)
Requirement already satisfied: branca in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from folium) (0.4.1)
Requirement already satisfied: six in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from folium) (1.15.0)
Requirement already satisfied: chardet<4,>=3.0.2 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from requests->folium) (3.0.4)
Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from requests->folium) (2020.6.20)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from requests->folium) (1.25.9)
Requirement already satisfied: idna<3,>=2.5 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from requests->folium) (2.9)
Requirement already satisfied: MarkupSafe>=0.23 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from jinja2->folium) (1.1.1)
```

In [2]:
```
import pandas as pd
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import wikipedia as wp
import requests
from bs4 import BeautifulSoup
import numpy as np # library to handle data in a vectorized manner
import json # library to handle JSON files
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values
import geopandas as gpd

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

import folium # map rendering library
from folium import plugins
from folium.plugins import HeatMap

print('Libraries imported.')
```

```
Libraries imported.
```

In [3]:
```
import wikipedia
```

In [4]:
```
import requests # library to handle requests
import pandas as pd # library for data analsyis
import numpy as np # library to handle data in a vectorized manner
import random # library for random number generation


!pip install geopy
from geopy.geocoders import Nominatim # module to convert an address into latitude and longitude values

# libraries for displaying images
from IPython.display import Image
from IPython.core.display import HTML

# tranforming json file into a pandas dataframe library
from pandas.io.json import json_normalize


! pip install folium==0.5.0
import folium # plotting library

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

Import an install from the official .geojson of Bogota:

In [7]:
```
!wget -q -O 'bogota.json' https://datosabiertos.bogota.gov.co/dataset/b0c66a77-3230-4d0c-a119-dead7f9b8b8e/resource/176c0f88-2d75-4d65-a1d6-4de527e7608a/download/egba.geojson
print('Data downloaded!')
```

```
Data downloaded!
```

In [8]:
```
with open('bogota.json') as json_data:
    bogota = json.load(json_data)
```

In [9]:
```
bogota
```

Out[9]:

```
{'type': 'FeatureCollection',
 'name': 'EGBa',
 'crs': {'type': 'name', 'properties': {'name': 'urn:ogc:def:crs:EPSG::3857'}},
 'features': [{'type': 'Feature',
   'properties': {'OBJECTID': 1,
    'NOMBRE_EST': 'MC DONALD S TUNAL',
    'DIRECCION': 'CL 47 B SUR NO. 24 B 33 lc 213-213',
    'SECTOR_CAT': 'TUNAL ORIENTAL',
    'LATITUD': 4.577897,
    'LONGITUD': -74.130324,
    'SUBCATEGOR': 'L',
    'LOCALIDAD': '06',
    'RNT': 30766,
    'TELEFONO_E': '6538028'
```

3 METHODOLOGY

3.1 BUSINESS UNDERSTANDING

Out[15]:

	Name	Neighborhood	Latitude	Longitude	Borough
0	NaN	TUNAL ORIENTAL	-8.252150e+06	-8.252150e+06	MC DONALD S TUNAL
1	NaN	LAS CRUCES	-8.246424e+06	-8.246424e+06	DOMINOS PIZZA CAJICA
2	NaN	VILLA MAYOR ORIENTAL	-8.251458e+06	-8.251458e+06	BARRA BTA C.C. CENTRO MAYOR
3	NaN	VILLA MAYOR ORIENTAL	-8.251458e+06	-8.251458e+06	CREPES Y WAFFLES S A - RESTAURANTE CENTRO MAYOR
4	NaN	VILLA MAYOR ORIENTAL	-8.251458e+06	-8.251458e+06	DON JEDIONDO SOPITAS Y PARRILLA

In [16]:
```
print('The dataframe has {} boroughs and {} neighborhoods.'.format(
        len(neighborhoods['Borough'].unique()),
        neighborhoods.shape[0]
    )
)
```

```
The dataframe has 497 boroughs and 546 neighborhoods.
```

The first step is to know how many localities and borough are in the city, order the data and cleaning the information that has limited or non-available information.

Next step is to engage the data of the restaurants and bars that are in the city, in order to have one information.

At last, we have to include the data of the number of habitants by localities.

3.2 ANALYTICAL APPROACH

Load Bogota data into pandas data frames.

Link FourSquare API Credentials.

Create functions to retrieve venues, venue details, and location through FourSquare API.

Use functions to retrieve name, borough, neighborhood and ID through the FourSquare API. Use functions to retrieve details such as tips and reviews for the data in the saved .csv file.

In [5]:
```
CLIENT_ID = '' # your Foursquare ID
CLIENT_SECRET = '' # your Foursquare Secret
VERSION = '20180604'
LIMIT = 30
print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID:
CLIENT_SECRET:
```

3.3 DATA EXPLORATION

Localidad	Participación	Habitantes	Porcentaje de Habitantes
ANTONIO NARIÑO	3.3%	108,941	1%
BARRIOS UNIDOS	1.5%	240,960	3%
BOSA	0.0%	646,833	8%
CHAPINERO	27.3%	137,870	2%
CIUDAD BOLÍVAR	0.0%	687,923	9%
ENGATIVÁ	3.0%	874,755	11%
FONTIBÓN	19.3%	380,453	5%
KENNEDY	4.0%	1,069,469	14%
LA CANDELARIA	3.3%	24,096	0%
LOS MÁRTIRES	0.2%	98,758	1%
PUENTE ARANDA	0.5%	258,414	3%
RAFAEL URIBE URIBE	0.3%	375,107	5%
SAN CRISTÓBAL	0.0%	406,025	5%
SANTA FE	8.0%	110,053	1%
SUBA	6.0%	1,174,736	15%
SUMAPAZ	0.0%	6,460	0%
TEUSAQUILLO	6.0%	151,092	2%
TUNJUELITO	0.6%	200,048	3%
USAQUÉN	16.5%	494,066	6%
USME	0.2%	432,724	5%
Total	100%	7,878,783	100%

https://github.com/cerodriguezb1981/Final-module-IBM-capstone/blob/main/Cifras.csv

In [14]:
```
for data in neighborhoods_data:
    borough = neighborhood_name = data['properties']['NOMBRE_EST']
    neighborhood_name = data['properties']['SECTOR_CAT']

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

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

The percentage of populations is not correlated with the participation of the restaurants and bars, the most populated localities has the least participation of businesses, and has also lower income ratio.

3.4 CLUSTERING

Out[13]:

Name	Neighborhood	Latitude	Longitude

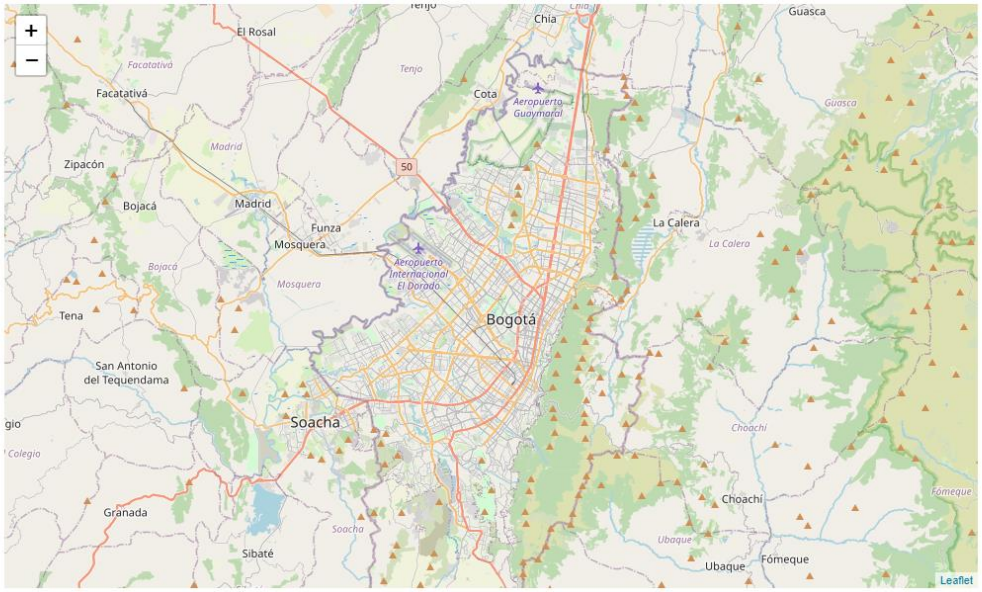
In [14]:
```
for data in neighborhoods_data:
    borough = neighborhood_name = data['properties']['NOMBRE_EST']
    neighborhood_name = data['properties']['SECTOR_CAT']

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

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

In [15]:
```
neighborhoods.head()
```

Out[15]:

	Name	Neighborhood	Latitude	Longitude	Borough
0	NaN	TUNAL ORIENTAL	-8.252150e+06	-8.252150e+06	MC DONALD S TUNAL
1	NaN	LAS CRUCES	-8.246424e+06	-8.246424e+06	DOMINOS PIZZA CAJICA
2	NaN	VILLA MAYOR ORIENTAL	-8.251458e+06	-8.251458e+06	BARRA BTA C.C. CENTRO MAYOR
3	NaN	VILLA MAYOR ORIENTAL	-8.251458e+06	-8.251458e+06	CREPES Y WAFFLES S A - RESTAURANTE CENTRO MAYOR
4	NaN	VILLA MAYOR ORIENTAL	-8.251458e+06	-8.251458e+06	DON JEDIONDO SOPITAS Y PARRILLA

Out[18]:



In [ ]:
```
map_BOGOTA = folium.Map(location=[latitude, longitude], zoom_start=12)

# add markers to map
for lat, lng, borough, neighborhood in zip(df_BOGOTA_complete['Latitude'],
                                           df_BOGOTA_complete['Longitude'],
                                           df_BOGOTA_complete['Id'],
                                           df_BOGOTA_complete['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_milan)

map_BOGOTA
```

In [ ]:
```
# create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=12)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]
# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(bogota_merged['Latitude'], bogota_merged['Longitude'], bogota_merged['Neighborhood'], bogota_merged['Cluster Labels']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=10,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

4 RESULT AND DISCUSSION

After gathering all the data, there is a concentration of bars and restaurants in the high income boroughs. Where the 53% of the business are in four localities, which represents only the 10% of the population the city.

The chain food business has more participations in the localities with low and medium income.

The restaurants and bars with international food are in over 80% located in the high income boroughs.

5 CONCLUSION

To open a restaurant in Bogota, it's very important to know what type of food is going to be served:

International

National

Fusion

Chain food

Is the investor interested in opening a franchise? If so, the options of international restaurants different from fast food is available and a very interesting option. PF Changs had a good experience before the pandemic. It was located in exclusive areas and had a good behavior.

The best places to open a restaurant in the city are in the localities of Chapinero, La Candelaria, Teusaquillo and Usaquen.

6 Further Studies

Using more accurate official information, introduce to the work the prices of each of the restaurants and bar that were in the study.

Include also the taxes and different cost (variable and fix) that will define better a budget regarding the needs of the new project.