Capstone report (2)

February 11, 2021

0.1 Introduction: Business Problem

In this project we will use data science tools to identify which borough in New York City has the most unique places to visit.

0.2 Target audience

people looking to explore unique places to visit New York City

0.3 Background of the problem

people looking to explore unique places in New York City usaully dont know which borough in New York City to visit that the most unique places in New York City

0.4 Data to be used for project

foursquare location data (Borough,Borough Latitude,Borough Longitude,Venue,Venue Latitude,Venue Longitude,Venue Category)

0.5 Libraries used for project

requests(library to handle requests),pandas(library for data analsysis),numpy(library to handle data in a vectorized manner),random(library for random number generation),Nominatim(module to convert an address into latitude and longitude values),Image&HTML(libraries for displaying images),json_normalize(tranforming json file into a pandas dataframe library),folium(plotting library)

0.6 Created a dataframe containing name and location of all the boroughs of New York

```
[3]: data = {'Borough':  ['Bronx', 'Brooklyn', 'Manhattan', ' Queens', 'Staten␣
     ↪Island'],
            'Latitude': ['40.8466508', '40.6501038', '40.7896239', '40.7498243',␣
     ↪'40.5834557',],
            'Longitude': ['-73.8785937', '-73.9495823', '-73.9598939', '-73.
     ↪7976337', '-74.1496048']
            }

     df = pd.DataFrame (data, columns = ['Borough','Latitude', 'Longitude'])
```

```
df
```

[3]:
```
       Borough    Latitude    Longitude
0         Bronx   40.8466508  -73.8785937
1      Brooklyn   40.6501038  -73.9495823
2     Manhattan   40.7896239  -73.9598939
3        Queens   40.7498243  -73.7976337
4  Staten Island 40.5834557  -74.1496048
```

0.7 Plotted out map of New York Boroughs

[6]:
```python
latitude = 40.8466508
longitude = -73.8785937

map_borough = folium.Map(location=[latitude, longitude], zoom_start=12)
for borough, lat, lng in zip(df['Borough'], df['Latitude'], df['Longitude']):
    label = '{}'.format(borough)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=6,
        popup=label,
        color='blue',
        fill=True,
        fill_color='blue',
        fill_opacity=1,
        parse_html=False).add_to(map_borough)

map_borough
```

[6]: `<folium.folium.Map at 0x7f278ca2a780>`

0.8 Used Foursquare api to retrieve location data on boroughs giving us the Borough,borough latitude, borough longitude, venue,venue latitude, venue longitude and venue category then converted I formation into dataframe below

[9]:
```
  nearby_venues
```

[9]:
```
     Borough  Borough Latitude  Borough Longitude  \
0      Bronx         40.846649         -73.878593
1      Bronx         40.846649         -73.878593
2      Bronx         40.846649         -73.878593
3      Bronx         40.846649         -73.878593
4      Bronx         40.846649         -73.878593
```

```
..         …                     …                    …
495  Staten Island         40.583454          -74.149605
496  Staten Island         40.583454          -74.149605
497  Staten Island         40.583454          -74.149605
498  Staten Island         40.583454          -74.149605
499  Staten Island         40.583454          -74.149605

                              Venue  Venue Latitude  Venue Longitude  \
0            Congo Gorilla Forest       40.847774        -73.881604
1                     JungleWorld       40.845227        -73.877181
2                    African Lions       40.847058        -73.878024
3                   Grizzly Corner       40.849023        -73.877739
4                        Bronx Zoo       40.853107        -73.878094
..                             …             …                …
495            Alfonso's Pastry Shop     40.545037        -74.160852
496                         Dunkin'     40.560888        -74.135120
497              P.C. Richard & Son     40.589120        -74.166615
498  Piccolino's italian Restaurant     40.551538        -74.149746
499               Hilton Garden Inn     40.614894        -74.176793

             Venue Category
0                       Zoo
1                       Zoo
2               Zoo Exhibit
3               Zoo Exhibit
4                       Zoo
..                        …
495                  Bakery
496              Donut Shop
497       Electronics Store
498       Italian Restaurant
499                   Hotel

[500 rows x 7 columns]
```

0.9 Created a new dataframe containing the only relevant I formation we needed like Borough,category longitude and latitude of venue

```
[12]: df_new

[12]:         Borough           Category    Latitude   Longitude
      0         Bronx                 Zoo   40.847774  -73.881604
      1         Bronx                 Zoo   40.845227  -73.877181
      2         Bronx         Zoo Exhibit   40.847058  -73.878024
      3         Bronx         Zoo Exhibit   40.849023  -73.877739
      4         Bronx                 Zoo   40.853107  -73.878094
      ..            …                   …           …           …
```

```
495   Staten Island                Bakery  40.545037 -74.160852
496   Staten Island            Donut Shop  40.560888 -74.135120
497   Staten Island     Electronics Store  40.589120 -74.166615
498   Staten Island     Italian Restaurant 40.551538 -74.149746
499   Staten Island                 Hotel  40.614894 -74.176793

[500 rows x 4 columns]
```

0.10   Found out there is 155 unique places in New York City

```
[14]:  print('There are {} Uniques Categories In total between the boroughs .'.
         →format(len(df_new['Category'].unique())))
```

```
There are 155 Uniques Categories In total between the boroughs .
```

0.11   Created a new data frame for the Bronx To determine how many unique places does the Bronx have

```
[15]:  df_bronx=df_new[(df_new['Borough'] == 'Bronx')]
       df_bronx
```

```
[15]:     Borough              Category   Latitude  Longitude
       0    Bronx                   Zoo  40.847774 -73.881604
       1    Bronx                   Zoo  40.845227 -73.877181
       2    Bronx            Zoo Exhibit 40.847058 -73.878024
       3    Bronx            Zoo Exhibit 40.849023 -73.877739
       4    Bronx                   Zoo  40.853107 -73.878094
       ..     …                     …         …          …
       95   Bronx        Cuban Restaurant 40.837741 -73.834458
       96   Bronx                   Food  40.871159 -73.863050
       97   Bronx   Vietnamese Restaurant 40.866956 -73.897839
       98   Bronx      Italian Restaurant 40.863019 -73.843607
       99   Bronx           Grocery Store 40.824016 -73.856396

[100 rows x 4 columns]
```

```
[28]:  print('There are {} Uniques Categories In total in Brooklyn  .'.
         →format(len(df_bronx['Category'].unique())))
```

```
There are 42 Uniques Categories In total in Brooklyn  .
```

0.12 Created a new data frame for the Staten Island To determine how many unique places does the Staten Island have

```
[17]: df_staten_island=df_new[(df_new['Borough'] == 'Staten Island')]
      df_staten_island
```

```
[17]:         Borough            Category    Latitude   Longitude
      400  Staten Island            Trail    40.586616 -74.146917
      401  Staten Island    Grocery Store    40.589997 -74.165715
      402  Staten Island   Toy / Game Store  40.581963 -74.166272
      403  Staten Island         Gastropub   40.581222 -74.167654
      404  Staten Island    Clothing Store   40.580279 -74.166153
      ..            ...               ...         ...         ...
      495  Staten Island           Bakery    40.545037 -74.160852
      496  Staten Island       Donut Shop    40.560888 -74.135120
      497  Staten Island  Electronics Store  40.589120 -74.166615
      498  Staten Island  Italian Restaurant 40.551538 -74.149746
      499  Staten Island            Hotel    40.614894 -74.176793

      [100 rows x 4 columns]
```

```
[18]: print('There are {} Uniques Categories In total in Staten Island   .'.
       ↪format(len(df_staten_island['Category'].unique())))
```

```
There are 53 Uniques Categories In total in Staten Island   .
```

0.13 Created a new data frame for the Brooklyn To determine how many unique places does the Brooklyn have

```
[19]: df_brooklyn=df_new[(df_new['Borough'] == 'Brooklyn')]
      df_brooklyn
```

```
[19]:        Borough              Category    Latitude   Longitude
      100  Brooklyn  Caribbean Restaurant    40.649091 -73.949243
      101  Brooklyn               Theater    40.646110 -73.957175
      102  Brooklyn            Restaurant    40.656012 -73.959912
      103  Brooklyn  Caribbean Restaurant    40.654953 -73.959783
      104  Brooklyn                  Café    40.660007 -73.953362
      ..        ...                   ...         ...         ...
      195  Brooklyn   American Restaurant    40.681470 -73.955800
      196  Brooklyn  Caribbean Restaurant    40.680807 -73.949525
      197  Brooklyn           Fish Market    40.677063 -73.969198
      198  Brooklyn                Bakery    40.678468 -73.968684
      199  Brooklyn         Grocery Store    40.676101 -73.971784

      [100 rows x 4 columns]
```

```
[20]: print('There are {} Uniques Categories In total in Brooklyn  .'.
      →format(len(df_brooklyn['Category'].unique())))
```

There are 61 Uniques Categories In total in Brooklyn .

0.14 Created a new data frame for the Manhattan To determine how many unique places does the Manhattan have

```
[21]: df_manhattan=df_new[(df_new['Borough'] == 'Manhattan')]
      df_manhattan
```

```
[21]:        Borough              Category   Latitude  Longitude
      200  Manhattan                 Field  40.790160 -73.955498
      201  Manhattan                Museum  40.785276 -73.957411
      202  Manhattan             Reservoir  40.784213 -73.961058
      203  Manhattan            Art Museum  40.784333 -73.957765
      204  Manhattan                  Café  40.787679 -73.953899
      ..         …                     …          …          …
      295  Manhattan          Liquor Store  40.767272 -73.959544
      296  Manhattan          Concert Hall  40.773359 -73.982373
      297  Manhattan  College Arts Building  40.773749 -73.982829
      298  Manhattan      French Restaurant  40.766829 -73.967579
      299  Manhattan               Theater  40.773354 -73.983827

      [100 rows x 4 columns]
```

```
[22]: print('There are {} Uniques Categories In total in Manhattan  .'.
      →format(len(df_manhattan['Category'].unique())))
```

There are 59 Uniques Categories In total in Manhattan .

Created a new data frame for the Queens To determine how many unique places does the Queens have

```
[23]: df_queens=df_new[(df_new['Borough'] == ' Queens')]
      df_queens
```

```
[23]:      Borough              Category   Latitude  Longitude
      300  Queens          Pizza Place  40.758193 -73.795719
      301  Queens                 Park  40.747715 -73.808877
      302  Queens           Bagel Shop  40.739460 -73.790554
      303  Queens         Movie Theater  40.741098 -73.784097
      304  Queens      Athletics & Sports  40.761575 -73.795557
      ..       …                    …          …          …
      395  Queens          Pizza Place  40.748445 -73.756498
      396  Queens      Tapas Restaurant  40.770271 -73.793495
      397  Queens                 Park  40.772293 -73.769168
      398  Queens          Pizza Place  40.761446 -73.770070
```

```
399    Queens               Gym   40.729120 -73.780555
```

```
[100 rows x 4 columns]
```

[24]:
```python
print('There are {} Uniques Categories In total in Queens  .'.
 ↪format(len(df_queens['Category'].unique())))
```

```
There are 55 Uniques Categories In total in Queens   .
```

0.15 Took each boroughs unique place count and Created another dataframe containing the total unique places under each borough

[25]:
```python
unique_bronx=len(df_bronx['Category'].unique())
unique_manhattan=len(df_manhattan['Category'].unique())
unique_brooklyn=len(df_brooklyn['Category'].unique())
unique_staten_island=len(df_queens['Category'].unique())
unique_queens=len(df_queens['Category'].unique())

df_category={ 'Bronx':[unique_bronx],'Manhattan':[unique_manhattan], 'Brooklyn':
 ↪[unique_brooklyn], 'Staten Island':[unique_staten_island], 'Queens':
 ↪[unique_queens],
}
df_category
df_category_list=pd.DataFrame (df_category, columns =␣
 ↪['Bronx','Manhattan','Brooklyn','Staten Island','Queens'])
df_category_list
```

[25]:
```
   Bronx  Manhattan  Brooklyn  Staten Island  Queens
0     42         59        61             55      55
```

0.16 Used the information on the previous datafram to create a bar chart to check which borough has the most unique places

[26]:
```python
df_category_list.plot.bar( rot=70, title="Number of uniques categories of place␣
 ↪to vist in each Borough");
```

Number of uniques categories of place to vist in each Borough
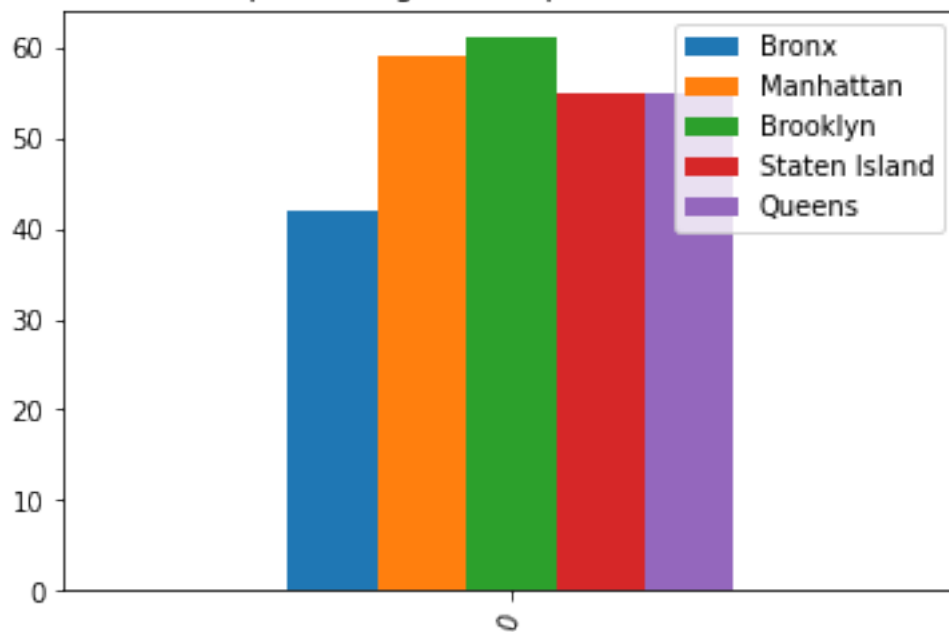
0.17 As we can see Brooklyn has the most unique places Between the rest of the boroughs of New York

[ ]: