Report

November 12, 2019

1 Introduction

1.1 Business Problem

Determining the most popular areas and places in *Karāchī, Pākistān* for incoming tourists and immigrants.

1.2 Stakeholders

- The State Government : for potential foreign currency inflow
- Immigrants : to know the best areas to liven and the best places to go (since they are new here)
- Tourism firms : to run their business and take the opportunity to build client relationships

2 Data

1. **Area names** and their respective **postal codes** from *Karāchī Metropolitan Corporation's* website:

Note: Being a resident of Karāchī myself; upon closer inspection one or more postal-codes are incorrect.

[1]:
```
import pandas as pd
import numpy as np
!pip install lxml
url = 'http://www.kmc.gos.pk/Contents.aspx?id=13'
df = pd.read_html(url)
```

```
Collecting lxml
  Downloading https://files.pythonhosted.org/packages/ec/be/5ab8abdd8663c0
386ec2dd595a5bc0e23330a0549b8a91e32f38c20845b6/lxml-4.4.1-cp36-cp36m-manylinux1_
x86_64.whl (5.8MB)
     |                  | 5.8MB 29.8MB/s eta 0:00:01
Installing collected packages: lxml
Successfully installed lxml-4.4.1
```

[2]:
```
#By trial, error and inspection the most relevant table is retained:
df = df[13]
df.head()
```

[2]:

```
                                                    0            1
0                                         POSTAL CODE          NaN
1  AREA  POSTAL CODE  AIRPORT  72500  BALDIA TOWN…          NaN
2                                                AREA  POSTAL CODE
3                                             AIRPORT        72500
4                                         BALDIA TOWN        75760
```

Now cleaning and formatting this data:

[3]:
```
df.drop([0,1,2], inplace = True)
df.rename(columns = {0:'Area', 1: 'Postal Code'}, inplace=True)
df.reset_index(inplace = True)
df.drop(columns = 'index', inplace = True)
df.head()
```

[3]:

	Area	Postal Code
0	AIRPORT	72500
1	BALDIA TOWN	75760
2	BOARD OF SECONDARY EDUCATION	75150
3	CANTT	75530
4	CITY GPO	7100

[4]:
```
df
```

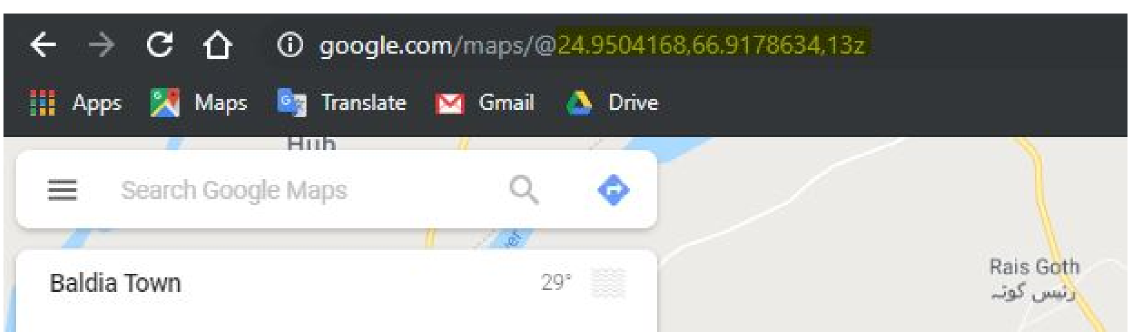
[4]:

	Area	Postal Code
0	AIRPORT	72500
1	BALDIA TOWN	75760
2	BOARD OF SECONDARY EDUCATION	75150
3	CANTT	75530
4	CITY GPO	7100
5	CLIFTON	75600
6	COD	75250
7	DARUL-ULOOM	75180
8	DEFENCE SOCIETY	75500
9	EXPORT PROCESSING ZONE	75150
10	FEDERAL B AREA	75950
11	GULSHAN-E-IQBAL	75300
12	HABIB BANK	75650
13	HOTEL METROPOLE	75520
14	JINNAH POST GRADUATE MEDICAL CENTER	75510
15	KARACHI GPO	74200
16	KARACHI UNIVERSITY	75270
17	KEEMARI	75620
18	KORANGI CREEK	75190
19	KORANGI GPO	74900
20	LANDHI COLONY	75160
21	LIAQATABAD	75900

22	LIYARI	75660
23	MALIR CANTT	75070
24	MALIR CITY	75050
25	MANGHOPIR	75890
26	MANORA	75640
27	MARIPUR(CE)	75780
28	MARIPUR(FA)	75750
29	MEHMOODABAD	75460
30	MODEL COLONY	75100
31	MURAD MEMON GOTH	75040
32	NATIONAL CEMENT INDUSTRY (DALMIA)	75260
33	NAZIMABAD GPO	74700
34	NEW KARACHI	75850
35	NEW TOWN GPO	74800
36	NORTH NAZIMABAD GPO	74600
37	ORANGE TOWN	75800
38	P.C.S.I.R	75280
39	P.E.C.H.S	75100
40	PAKISTAN MACHINE TOOL FACTORY	75760
41	PAKISTAN NAVAL ARMAMENT DEPOT	75790
42	PAKISTAN STEEL MILLS	75800
43	PAKISTAN STEEL MILLS TOWN SHIP	75200
44	PORT MUHAMMAD BIN QASIM	75020
45	QUAIDABAD	75120
46	RAFA-E-AAM SOCIETY	75210
47	S.I.T.E	75700
48	SADDAR GPO	74100
49	SHAH FAISAL COLONY	75230
50	SHAHRA-E-FAISAL	75350
51	SHER SHAH COLONY	75730
52	SINDH GOVERNOR HOUSE	75580
53	NISHTER ROAD	74550

2. Manually forming a data frame for *geo-coordinates*, by manually inputting the names on **Google Maps/Search** and copying the coordinates:

Note: Some coordinates were truncated or rounded-off to maintain uniformity in the data points



[5]:
```
gdf = np.array([[24.9008, 67.1681],
                [24.9525, 66.9550],
                [24.9238, 67.0283],
                [24.8547, 67.0435],
                [24.8511, 67.0017],
                [24.8270, 67.0251],
                [24.8915, 67.1328],
                [24.8456, 67.1662],
                [24.8043, 67.0577],
                [24.8294, 67.2417],
                [24.9275, 67.0641],
                [24.9180, 67.0971],
                [24.9453, 66.9336],
                [24.8465, 67.0241],
                [24.8524, 67.0429],
                [24.8511, 66.9995],
                [24.9418, 67.1207],
                [24.8788, 66.8790],
                [24.8033, 67.1239],
                [24.8399, 67.1411],
                [24.8406, 67.1948],
                [24.9057, 67.0446],
                [24.8784, 67.0103],
                [24.9596, 67.2252],
                [24.8771, 67.1933],
                [24.9265, 66.9514],
                [24.8445, 66.9199],
                [24.8690, 66.9156],
                [24.8700, 66.9204],
                [24.8528, 67.0748],
                [24.9023, 67.1892],
                [24.9191, 67.2496],
                [24.9049, 67.1021],
                [24.9094, 67.0253],
                [24.9999, 67.0648],
                [24.8924, 67.0521],
                [24.9372, 67.0423],
                [24.9517, 67.0023],
                [24.9561, 67.1261],
                [24.8688, 67.0614],
                [24.8394, 67.2513],
                [24.9450, 66.9322],
                [24.8203, 67.3398],
                [24.8723, 67.3358],
                [24.7696, 67.3324],
                [24.8588, 67.2220],
                [24.8795, 67.1746],
```

```
                [24.9053, 66.9928],
                [24.8599, 67.0241],
                [24.8797, 67.1599],
                [24.8604, 67.0689],
                [24.8844, 66.9842],
                [24.8507, 67.0261],
                [24.8853, 67.0283]])
gdf = pd.DataFrame({'Latitude':gdf[:,0],'Longitude':gdf[:,1]})
gdf.head()
```

[5]:

	Latitude	Longitude
0	24.9008	67.1681
1	24.9525	66.9550
2	24.9238	67.0283
3	24.8547	67.0435
4	24.8511	67.0017

Merging `gdf` with `df`:

[6]:
```
df = df.join(gdf)
df.head()
```

[6]:

	Area	Postal Code	Latitude	Longitude
0	AIRPORT	72500	24.9008	67.1681
1	BALDIA TOWN	75760	24.9525	66.9550
2	BOARD OF SECONDARY EDUCATION	75150	24.9238	67.0283
3	CANTT	75530	24.8547	67.0435
4	CITY GPO	7100	24.8511	67.0017

3. Obtaining data from `Foursquare` to obtain popular places; according to user's feedback and corresponding to the *geo-data* obtained previously.

Importing necessary libraries:

[7]:
```
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 pandas␣
 ↪dataframe

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

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

```
#!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
```

Forming the map of Karāchī:

[8]:
```
kc = [24.8607, 67.0011]#Karachi's coordinates from Google
mp = folium.Map(location= kc, zoom_start=10.5)

# add markers to map
for pc, lat, lng, ar in zip(df['Postal Code'], df['Latitude'], df['Longitude'],␣
↪df['Area']):
    label = '{}'.format(ar)
    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(mp)

mp
```

[8]: <folium.folium.Map at 0x7fdf79b51630>

Defining `Foursquare` credentials:

[9]:
```
CLIENT_ID = 'EBZBKBMHOC0LNVSTB3HQ5HSFKGJGSZJ0X2N2QR4D4YDBTUMI' # your␣
↪Foursquare ID
CLIENT_SECRET = 'ASKVML0XTCDWA3CVSFHM4YKXHNZWHYZC4GMVWXQNUWRBZUQZ' # your␣
↪Foursquare Secret
VERSION = '20180604'
LIMIT = 10
```

Defining function that gives a data frame for **top 10** venues for each area within a **10 km** radius

[10]:
```
def getNearbyVenues(names, latitudes, longitudes, radius=10000):

    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 = ['Area',
                  'Neighborhood Latitude',
                  'Neighborhood Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

Using the defined function:

```
[11]: vs = getNearbyVenues(names=df['Area'],
                          latitudes=df['Latitude'],
                          longitudes=df['Longitude']
                         )
```

```
AIRPORT
BALDIA TOWN
BOARD OF SECONDARY EDUCATION
CANTT
CITY GPO
CLIFTON
COD
DARUL-ULOOM
```

DEFENCE SOCIETY
EXPORT PROCESSING ZONE
FEDERAL B AREA
GULSHAN-E-IQBAL
HABIB BANK
HOTEL METROPOLE
JINNAH POST GRADUATE MEDICAL CENTER
KARACHI GPO
KARACHI UNIVERSITY
KEEMARI
KORANGI CREEK
KORANGI GPO
LANDHI COLONY
LIAQATABAD
LIYARI
MALIR CANTT
MALIR CITY
MANGHOPIR
MANORA
MARIPUR(CE)
MARIPUR(FA)
MEHMOODABAD
MODEL COLONY
MURAD MEMON GOTH
NATIONAL CEMENT INDUSTRY (DALMIA)
NAZIMABAD GPO
NEW KARACHI
NEW TOWN GPO
NORTH NAZIMABAD GPO
ORANGE TOWN
P.C.S.I.R
P.E.C.H.S
PAKISTAN MACHINE TOOL FACTORY
PAKISTAN NAVAL ARMAMENT DEPOT
PAKISTAN STEEL MILLS
PAKISTAN STEEL MILLS TOWN SHIP
PORT MUHAMMAD BIN QASIM
QUAIDABAD
RAFA-E-AAM SOCIETY
S.I.T.E
SADDAR GPO
SHAH FAISAL COLONY
SHAHRA-E-FAISAL
SHER SHAH COLONY
SINDH GOVERNOR HOUSE
NISHTER ROAD

[12]:
```python
# one hot encoding
oh = pd.get_dummies(vs[['Venue Category']], prefix="", prefix_sep="")

# add area column back to dataframe
oh['Area'] = vs['Area']

# move area column to the first column
fixed_columns = [oh.columns[-1]] + list(oh.columns[:-1])
oh = oh[fixed_columns]
```

Grouping the *One Hot encoded* data frame:

[13]:
```python
gp = oh.groupby('Area').mean().reset_index()
```

Defining function for most popular venues, **adapted from the Lab(s), as is most of the complex coding you see in this notebook**:

[14]:
```python
def return_most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:num_top_venues]

num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Area']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
area_venues_sorted = pd.DataFrame(columns=columns)
area_venues_sorted['Area'] = gp['Area']

for ind in np.arange(gp.shape[0]):
    area_venues_sorted.iloc[ind, 1:] = return_most_common_venues(gp.iloc[ind, :
 ↪], num_top_venues)
```

3 Methodology

3.1 Exploratory Data Analysis

Let's take a cursory look at the number of venues retrieved in each area:

[15]:
```
vs.groupby('Area').count()
```

[15]:
```
                                   Neighborhood Latitude  \
Area
AIRPORT                                               10
BALDIA TOWN                                           10
BOARD OF SECONDARY EDUCATION                          10
CANTT                                                 10
CITY GPO                                              10
CLIFTON                                               10
COD                                                   10
DARUL-ULOOM                                           10
DEFENCE SOCIETY                                       10
EXPORT PROCESSING ZONE                                 7
FEDERAL B AREA                                        10
GULSHAN-E-IQBAL                                       10
HABIB BANK                                             5
HOTEL METROPOLE                                       10
JINNAH POST GRADUATE MEDICAL CENTER                   10
KARACHI GPO                                           10
KARACHI UNIVERSITY                                    10
KEEMARI                                                5
KORANGI CREEK                                         10
KORANGI GPO                                           10
LANDHI COLONY                                         10
LIAQATABAD                                            10
LIYARI                                                10
MALIR CANTT                                           10
MALIR CITY                                            10
MANGHOPIR                                             10
MANORA                                                10
MARIPUR(CE)                                           10
MARIPUR(FA)                                           10
MEHMOODABAD                                           10
MODEL COLONY                                          10
MURAD MEMON GOTH                                      10
NATIONAL CEMENT INDUSTRY (DALMIA)                     10
NAZIMABAD GPO                                         10
NEW KARACHI                                           10
NEW TOWN GPO                                          10
NISHTER ROAD                                          10
NORTH NAZIMABAD GPO                                   10
ORANGE TOWN                                           10
P.C.S.I.R                                             10
P.E.C.H.S                                             10
PAKISTAN MACHINE TOOL FACTORY                          4
PAKISTAN NAVAL ARMAMENT DEPOT                          4
```

PAKISTAN STEEL MILLS	8
PAKISTAN STEEL MILLS TOWN SHIP	5
PORT MUHAMMAD BIN QASIM	6
QUAIDABAD	10
RAFA-E-AAM SOCIETY	10
S.I.T.E	10
SADDAR GPO	10
SHAH FAISAL COLONY	10
SHAHRA-E-FAISAL	10
SHER SHAH COLONY	10
SINDH GOVERNOR HOUSE	10

Area	Neighborhood Longitude	Venue \
AIRPORT	10	10
BALDIA TOWN	10	10
BOARD OF SECONDARY EDUCATION	10	10
CANTT	10	10
CITY GPO	10	10
CLIFTON	10	10
COD	10	10
DARUL-ULOOM	10	10
DEFENCE SOCIETY	10	10
EXPORT PROCESSING ZONE	7	7
FEDERAL B AREA	10	10
GULSHAN-E-IQBAL	10	10
HABIB BANK	5	5
HOTEL METROPOLE	10	10
JINNAH POST GRADUATE MEDICAL CENTER	10	10
KARACHI GPO	10	10
KARACHI UNIVERSITY	10	10
KEEMARI	5	5
KORANGI CREEK	10	10
KORANGI GPO	10	10
LANDHI COLONY	10	10
LIAQATABAD	10	10
LIYARI	10	10
MALIR CANTT	10	10
MALIR CITY	10	10
MANGHOPIR	10	10
MANORA	10	10
MARIPUR(CE)	10	10
MARIPUR(FA)	10	10
MEHMOODABAD	10	10
MODEL COLONY	10	10
MURAD MEMON GOTH	10	10
NATIONAL CEMENT INDUSTRY (DALMIA)	10	10

NAZIMABAD GPO	10	10
NEW KARACHI	10	10
NEW TOWN GPO	10	10
NISHTER ROAD	10	10
NORTH NAZIMABAD GPO	10	10
ORANGE TOWN	10	10
P.C.S.I.R	10	10
P.E.C.H.S	10	10
PAKISTAN MACHINE TOOL FACTORY	4	4
PAKISTAN NAVAL ARMAMENT DEPOT	4	4
PAKISTAN STEEL MILLS	8	8
PAKISTAN STEEL MILLS TOWN SHIP	5	5
PORT MUHAMMAD BIN QASIM	6	6
QUAIDABAD	10	10
RAFA-E-AAM SOCIETY	10	10
S.I.T.E	10	10
SADDAR GPO	10	10
SHAH FAISAL COLONY	10	10
SHAHRA-E-FAISAL	10	10
SHER SHAH COLONY	10	10
SINDH GOVERNOR HOUSE	10	10

	Venue Latitude	Venue Longitude \
Area		
AIRPORT	10	10
BALDIA TOWN	10	10
BOARD OF SECONDARY EDUCATION	10	10
CANTT	10	10
CITY GPO	10	10
CLIFTON	10	10
COD	10	10
DARUL-ULOOM	10	10
DEFENCE SOCIETY	10	10
EXPORT PROCESSING ZONE	7	7
FEDERAL B AREA	10	10
GULSHAN-E-IQBAL	10	10
HABIB BANK	5	5
HOTEL METROPOLE	10	10
JINNAH POST GRADUATE MEDICAL CENTER	10	10
KARACHI GPO	10	10
KARACHI UNIVERSITY	10	10
KEEMARI	5	5
KORANGI CREEK	10	10
KORANGI GPO	10	10
LANDHI COLONY	10	10
LIAQATABAD	10	10
LIYARI	10	10

MALIR CANTT	10	10
MALIR CITY	10	10
MANGHOPIR	10	10
MANORA	10	10
MARIPUR(CE)	10	10
MARIPUR(FA)	10	10
MEHMOODABAD	10	10
MODEL COLONY	10	10
MURAD MEMON GOTH	10	10
NATIONAL CEMENT INDUSTRY (DALMIA)	10	10
NAZIMABAD GPO	10	10
NEW KARACHI	10	10
NEW TOWN GPO	10	10
NISHTER ROAD	10	10
NORTH NAZIMABAD GPO	10	10
ORANGE TOWN	10	10
P.C.S.I.R	10	10
P.E.C.H.S	10	10
PAKISTAN MACHINE TOOL FACTORY	4	4
PAKISTAN NAVAL ARMAMENT DEPOT	4	4
PAKISTAN STEEL MILLS	8	8
PAKISTAN STEEL MILLS TOWN SHIP	5	5
PORT MUHAMMAD BIN QASIM	6	6
QUAIDABAD	10	10
RAFA-E-AAM SOCIETY	10	10
S.I.T.E	10	10
SADDAR GPO	10	10
SHAH FAISAL COLONY	10	10
SHAHRA-E-FAISAL	10	10
SHER SHAH COLONY	10	10
SINDH GOVERNOR HOUSE	10	10

	Venue Category
Area	
AIRPORT	10
BALDIA TOWN	10
BOARD OF SECONDARY EDUCATION	10
CANTT	10
CITY GPO	10
CLIFTON	10
COD	10
DARUL-ULOOM	10
DEFENCE SOCIETY	10
EXPORT PROCESSING ZONE	7
FEDERAL B AREA	10
GULSHAN-E-IQBAL	10
HABIB BANK	5

```
HOTEL METROPOLE                                  10
JINNAH POST GRADUATE MEDICAL CENTER              10
KARACHI GPO                                      10
KARACHI UNIVERSITY                               10
KEEMARI                                           5
KORANGI CREEK                                    10
KORANGI GPO                                      10
LANDHI COLONY                                    10
LIAQATABAD                                       10
LIYARI                                           10
MALIR CANTT                                      10
MALIR CITY                                       10
MANGHOPIR                                        10
MANORA                                           10
MARIPUR(CE)                                      10
MARIPUR(FA)                                      10
MEHMOODABAD                                      10
MODEL COLONY                                     10
MURAD MEMON GOTH                                 10
NATIONAL CEMENT INDUSTRY (DALMIA)                10
NAZIMABAD GPO                                    10
NEW KARACHI                                      10
NEW TOWN GPO                                     10
NISHTER ROAD                                     10
NORTH NAZIMABAD GPO                              10
ORANGE TOWN                                      10
P.C.S.I.R                                        10
P.E.C.H.S                                        10
PAKISTAN MACHINE TOOL FACTORY                     4
PAKISTAN NAVAL ARMAMENT DEPOT                     4
PAKISTAN STEEL MILLS                              8
PAKISTAN STEEL MILLS TOWN SHIP                    5
PORT MUHAMMAD BIN QASIM                           6
QUAIDABAD                                        10
RAFA-E-AAM SOCIETY                               10
S.I.T.E                                          10
SADDAR GPO                                       10
SHAH FAISAL COLONY                               10
SHAHRA-E-FAISAL                                  10
SHER SHAH COLONY                                 10
SINDH GOVERNOR HOUSE                             10
```

Some areas have less than 10 venues. From their names it can be seen these are mostly industrial or non-general public-visiting areas:

Now let's have a glance at the venues(`vs`) dataset itself:

[16]:
```
vs
```

[16]:
```
              Area  Neighborhood Latitude  Neighborhood Longitude  \
0          AIRPORT                24.9008                 67.1681
1          AIRPORT                24.9008                 67.1681
2          AIRPORT                24.9008                 67.1681
3          AIRPORT                24.9008                 67.1681
4          AIRPORT                24.9008                 67.1681
..             …                    …                       …
499   NISHTER ROAD                24.8853                 67.0283
500   NISHTER ROAD                24.8853                 67.0283
501   NISHTER ROAD                24.8853                 67.0283
502   NISHTER ROAD                24.8853                 67.0283
503   NISHTER ROAD                24.8853                 67.0283

                                      Venue  Venue Latitude  \
0                                 Pizza Max       24.905053
1                    Butler's Chocolate Cafe       24.901609
2                         14th Street Pizza       24.910596
3                Ramada Plaza Hotel Pool BBQ       24.894331
4                          California Pizza       24.907824
..                                      …            …
499  NAPA - National Academy of Performing Arts       24.851907
500                             Zahid Nihari       24.860282
501                           Atrium Cinemas       24.856148
502                             Karachi Club       24.844083
503                                 Xander’s       24.866432

     Venue Longitude         Venue Category
0          67.182587            Pizza Place
1          67.166128            Coffee Shop
2          67.096607            Pizza Place
3          67.156555              BBQ Joint
4          67.109657            Pizza Place
..             …                      …
499        67.021652   Performing Arts Venue
500        67.031918   Pakistani Restaurant
501        67.030312              Multiplex
502        67.029199            Social Club
503        67.077803                   Café

[504 rows x 7 columns]
```

It appears that food venues are very popular around the city.

Next let’s check the number of unique venue categories in this dataset:

[17]:
```
print('There are {} uniques categories.'.format(len(vs['Venue Category'].
→unique())))
```

```
There are 53 uniques categories.
```

3.2 Machine Learning Algorithm(s)

K-Means Clustering Algorithm is used to Cluster Areas that are alike (number of clusters are arbitrarily set to 5:

[18]:
```python
# set number of clusters
kclusters = 5

cl = gp.drop('Area', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(cl)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

[18]:
```
array([0, 0, 0, 4, 4, 0, 0, 2, 4, 3], dtype=int32)
```

Merging Results with the original dataframe, `df`:

[19]:
```python
# add clustering labels
area_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

merge = df

# merge toronto_grouped with toronto_data to add latitude/longitude for each␣
↪neighborhood
merge = merge.join(area_venues_sorted.set_index('Area'), on='Area', how='right')

merge.head() # check the last columns!
```

[19]:
```
                         Area Postal Code  Latitude  Longitude  \
0                     AIRPORT       72500   24.9008    67.1681
1                 BALDIA TOWN       75760   24.9525    66.9550
2  BOARD OF SECONDARY EDUCATION       75150   24.9238    67.0283
3                       CANTT       75530   24.8547    67.0435
4                    CITY GPO        7100   24.8511    67.0017

   Cluster Labels 1st Most Common Venue 2nd Most Common Venue  \
0               0           Pizza Place  Gym / Fitness Center
1               0    African Restaurant                 Diner
2               0         Ice Cream Shop           Pizza Place
3               4  Pakistani Restaurant        Ice Cream Shop
4               4         Ice Cream Shop              BBQ Joint

  3rd Most Common Venue 4th Most Common Venue  5th Most Common Venue  \
```

```
0              BBQ Joint         Shopping Mall                 Bakery
1           Shopping Mall             Juice Bar          Burger Joint
2   Fast Food Restaurant             Donut Shop                 Diner
3    Japanese Restaurant              Multiplex  Performing Arts Venue
4              Multiplex    Chinese Restaurant  Performing Arts Venue

  6th Most Common Venue 7th Most Common Venue 8th Most Common Venue  \
0            Coffee Shop        History Museum  Fast Food Restaurant
1            Pizza Place            Food Court  Fast Food Restaurant
2   Pakistani Restaurant    Chinese Restaurant  Gym / Fitness Center
3     Chinese Restaurant                  Café              BBQ Joint
4                   Café   Japanese Restaurant      Asian Restaurant

  9th Most Common Venue 10th Most Common Venue
0     Falafel Restaurant          Grocery Store
1   Pakistani Restaurant             Donut Shop
2          Shopping Mall              BBQ Joint
3            Social Club      Frozen Yogurt Shop
4            Social Club             Steakhouse
```

Visualizing the Clusters on map:

[20]:
```
# create map
cmap = folium.Map(location= kc, zoom_start=11)

# 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(merge['Latitude'], merge['Longitude'],
 ↪merge['Area'], merge['Cluster Labels']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(cmap)

cmap
```

[20]: <folium.folium.Map at 0x7fdf78195908>

[21]:
```
c0 = merge.loc[merge['Cluster Labels'] == 0, merge.columns]
c1 = merge.loc[merge['Cluster Labels'] == 1, merge.columns]
c2 = merge.loc[merge['Cluster Labels'] == 2, merge.columns]
c3 = merge.loc[merge['Cluster Labels'] == 3, merge.columns]
c4 = merge.loc[merge['Cluster Labels'] == 4, merge.columns]
```

[22]:
```
c2
```

[22]:
```
                  Area Postal Code  Latitude  Longitude  Cluster Labels  \
7           DARUL-ULOOM       75180   24.8456    67.1662               2
20        LANDHI COLONY       75160   24.8406    67.1948               2
23          MALIR CANTT       75070   24.9596    67.2252               2
24           MALIR CITY       75050   24.8771    67.1933               2
30         MODEL COLONY       75100   24.9023    67.1892               2
31     MURAD MEMON GOTH       75040   24.9191    67.2496               2
35         NEW TOWN GPO       74800   24.8924    67.0521               2
45            QUAIDABAD       75120   24.8588    67.2220               2
46   RAFA-E-AAM SOCIETY       75210   24.8795    67.1746               2
49   SHAH FAISAL COLONY       75230   24.8797    67.1599               2

   1st Most Common Venue 2nd Most Common Venue 3rd Most Common Venue  \
7              BBQ Joint                  Café           Pizza Place
20  Fast Food Restaurant             BBQ Joint           Golf Course
23             BBQ Joint  Fast Food Restaurant      Airport Terminal
24           Pizza Place             BBQ Joint           Golf Course
30           Pizza Place             BBQ Joint                Bakery
31             BBQ Joint                  Café  Fast Food Restaurant
35             BBQ Joint  Gym / Fitness Center    Falafel Restaurant
45             BBQ Joint                  Café  Fast Food Restaurant
46           Pizza Place             BBQ Joint        History Museum
49           Pizza Place             BBQ Joint        History Museum

   4th Most Common Venue 5th Most Common Venue 6th Most Common Venue  \
7         History Museum      Asian Restaurant    Frozen Yogurt Shop
20                  Café           Pizza Place           Coffee Shop
23                  Café           Pizza Place                Market
24                Bakery             Juice Bar           Coffee Shop
30           Golf Course           Coffee Shop  Fast Food Restaurant
31      Airport Terminal           Pizza Place           Coffee Shop
35              Tea Room                  Café           Pizza Place
45      Airport Terminal           Pizza Place           Coffee Shop
46  Gym / Fitness Center    Frozen Yogurt Shop                  Café
49  Gym / Fitness Center    Frozen Yogurt Shop                  Café

   7th Most Common Venue 8th Most Common Venue 9th Most Common Venue  \
```

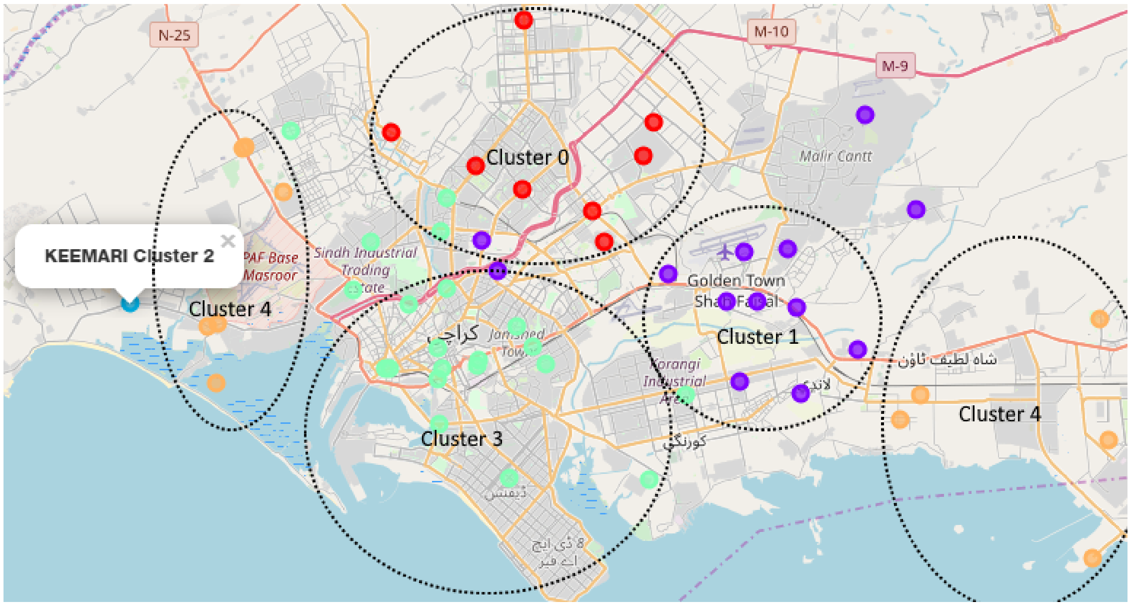
```
7            Coffee Shop    Falafel Restaurant  Gym / Fitness Center
20            Toll Plaza    Falafel Restaurant  Gym / Fitness Center
23           Coffee Shop      Department Store            Toll Plaza
24  Fast Food Restaurant            Toll Plaza    Falafel Restaurant
30            Toll Plaza    Falafel Restaurant  Gym / Fitness Center
31            Toll Plaza    Falafel Restaurant  Gym / Fitness Center
35    Chinese Restaurant  Pakistani Restaurant  Fast Food Restaurant
45            Toll Plaza    Falafel Restaurant  Gym / Fitness Center
46           Coffee Shop    Falafel Restaurant         Grocery Store
49           Coffee Shop    Falafel Restaurant         Grocery Store

   10th Most Common Venue
7           Grocery Store
20          Grocery Store
23                   Farm
24          Grocery Store
30          Grocery Store
31          Grocery Store
35          Grocery Store
45          Grocery Store
46            Golf Course
49            Golf Course
```

4 Results & Discussion

4.1 Observations

Note: The screen shot may not match your run of my code because the areas are clustered differently sometimes.

- The clusters appear to be clustered geographically and distributed roughly symmetrically in the city, as illustrated below:



- The coastal area, *Keemari*, is one of its kind (Cluster 2)
- By inspecting the data frames named `ci` and inline with the preceding discussion, I now name the clusters as follows:
  - `c0` "**Northern** Karāchī district; the place for **Bakeries**"
  - `c1` "**North-Eastern** Karāchī district; the place for **BBQ** and **Pizza**"
  - `c2` "**Western** Coastal Karāchī; **Beach** spot"
  - `c3` "**Central** and **South** Karāchī; for **Ice-Cream** and **Cafē** joints"
  - `c4` "**Western** and **Eastern** Karāchī; **Beach** and **Recreation** spots + **BBQ** joints"

4.2 Recommendations

- The data for popular venues can be further enhanced by forming a trending venues table for different seasons, months, weeks and days.
- The best number of clusters should be chosen via the elbow method.
- Cluster 2 should intuitively be merged with Cluster 4.

5 Conclusion

- Eating joints are generally the most popular venues, followed by Beach and Recreational spots
- The clusters are not only similar by venue types, but the clusters appear to have a geographical pattern as well.