Capstone Project Report
The Battle of Neighborhoods

1. Introduction/Business Problem:

(Introduction to discuss the business problem and who would be interested in this project.)

There is a person who want to live/ stay in a city. And want to choose which city it should choose and then in that city where one should go depending upon ones liking. Either one wants to choose a region to stay where there is a lot of food option (Food Lover). Or one is recreational and want to enjoy the scenery or have fun. Or one would like to live near all the facilities close to each other.

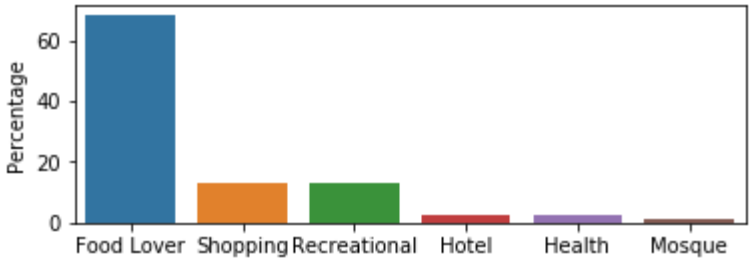

Figure 1

Now in this Project we will analyze two cities of PAKISTAN, one is Lahore and other is Islamabad. And based on the data of FourSquare we will analyze the Differences and Similarities of the options available.

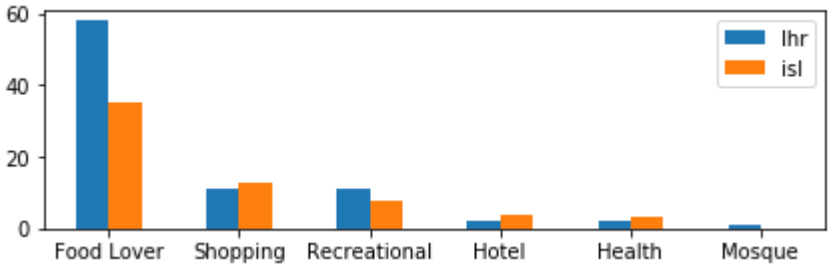

Figure 2

Then one after choosing a city will see which part of the city suits him.

2. Data Description

(Data where you describe the data that will be used to solve the problem and the source of the data.)

a. First *geopy* will be used to find the location (latitude and longitude) of the cities. (Figure 3)

```
Isl_address = 'Islamabad, Pakistan'
geolocator = Nominatim(user_agent="explorer")
location = geolocator.geocode(Isl_address)
Isl_latitude = location.latitude
Isl_longitude = location.longitude
print('The geograpical coordinate of {} are {}, {}.'.format(Isl_address,Isl_latitude, Isl_longitude))
```

```
The geograpical coordinate of Islamabad, Pakistan are 33.6938118, 73.0651511.
```

Figure 3

b. Then using FOURSQAURE Credentials we will get the venues near those cities. (Figure 4)

```
CLIENT_ID = 'D13G4BNF2CYNDJQPLAIPKLNZJM4OG5DG0CE13QVQTH2WFLCF' # your Foursquare ID
CLIENT_SECRET = 'OJXVEOLE2KYEYPZR34SGA4WUIRE1UTGP0C4VJYOUKA1YRJU5' # your Foursquare Secret
VERSION = '20180605' # Foursquare API version

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: D13G4BNF2CYNDJQPLAIPKLNZJM4OG5DG0CE13QVQTH2WFLCF
CLIENT_SECRET:OJXVEOLE2KYEYPZR34SGA4WUIRE1UTGP0C4VJYOUKA1YRJU5
```

Figure 4

c. We will categorize the venues in six using dictionary in new column ‘New Cat’. (Figure 5)

	name	categories	lat	lng	address	New Cat
	Pakistan Monument	History Museum	33.693070	73.068910	Islamabad	Recreational
	Burning Brownie Cafe & Bake Shop	Coffee Shop	33.720508	73.073956	Shop # 66, Beverly Centre	Food Lover
	Lok Virsa Museum	History Museum	33.688942	73.072550	Garden avenue	Recreational
	D.Watson	Pharmacy	33.696279	73.012849	10th Avenue, F-10 Markaz	Health
	Chaaye Khana	Tea Room	33.729388	73.074913	Shop # 11, Block-B, United Bakery Plaza, F-6 M...	Food Lover

Figure 5

d. We will compare the categories of the cities, as shown in Figure 2.
e. We will define the region using kmeans clustering, as shown in Figure 9.
f. Then one can choose the region based on one’s priorities, as in Figure 6.

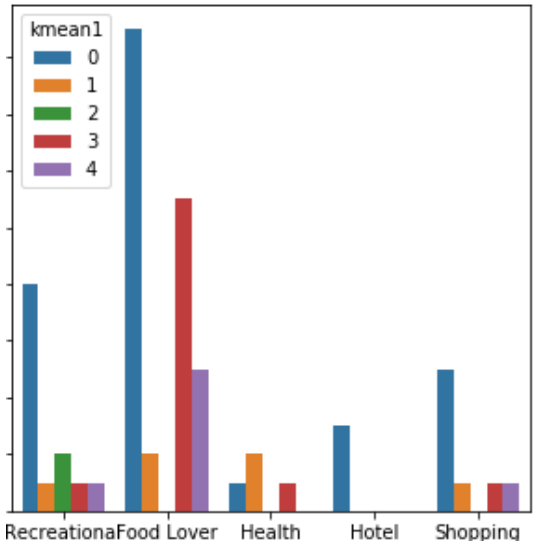

Figure 6

3. Methodology

(Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.)

- We had different categories (as shown in Figure 7).

```
print(set(nearby_venues.categories))
```

```
{'Restaurant', 'Park', 'Fish & Chips Shop', 'Bakery', 'Ice Cream Shop', 'Pakistani Restaurant', 'Coffee Shop', 'Pizza Place',
'Monument / Landmark', 'Chinese Restaurant', 'Fast Food Restaurant', 'Recreation Center', 'Italian Restaurant', 'Middle Eastern
Restaurant', 'Multiplex', 'Asian Restaurant', 'Spa', 'Steakhouse', 'Movie Theater', 'Bookstore', 'Café', 'Food Court', 'America
n Restaurant', 'Gym / Fitness Center', 'Market', 'Portuguese Restaurant', 'Golf Course', 'Convenience Store', 'Indian Restauran
t', 'Mosque', 'Tea Room', 'Department Store', 'Food Truck', 'Rest Area', 'Shopping Mall', 'Mediterranean Restaurant', 'Boutiqu
e', 'Burger Joint', 'Theater', 'Hotel'}
```

Figure 7

- We EDA using dictionary (as shown in figure 8) to make a new category ‘New Cat’.

```
print(dic)
```

```
{'Asian Restaurant': 'Food Lover', 'BBQ Joint': 'Food Lover', 'Bakery': 'Food Lover', 'Bookstore': 'Shopping', 'Boutique': 'Sho
pping', 'Burger Joint': 'Food Lover', 'Café': 'Food Lover', 'Coffee Shop': 'Food Lover', 'Department Store': 'Shopping', 'Fast
Food Restaurant': 'Food Lover', 'Food Court': 'Food Lover', 'Gift Shop': 'Shopping', 'Golf Course': 'Recreational', 'Gym': 'Hea
lth', 'History Museum': 'Recreational', 'Hotel': 'Hotel', 'Ice Cream Shop': 'Food Lover', 'Italian Restaurant': 'Food Lover',
'Lake': 'Recreational', 'Market': 'Shopping', 'Mountain': 'Recreational', 'Multiplex': 'Recreational', 'Nature Preserve': 'Recr
eational', 'Other Great Outdoors': 'Recreational', 'Pakistani Restaurant': 'Food Lover', 'Park': 'Recreational', 'Pharmacy': 'H
ealth', 'Pizza Place': 'Food Lover', 'Restaurant': 'Food Lover', 'Sandwich Place': 'Food Lover', 'Scenic Lookout': 'Recreationa
l', 'Shopping Mall': 'Shopping', 'Snack Place': 'Food Lover', 'Tea Room': 'Food Lover', 'Theater': 'Recreational', 'Trail': 'Re
creational', 'Wings Joint': 'Food Lover'}
```

Figure 8

❖ Used kmeans Unsupervised Clustering Algorithm to make regions. (Figure 9)

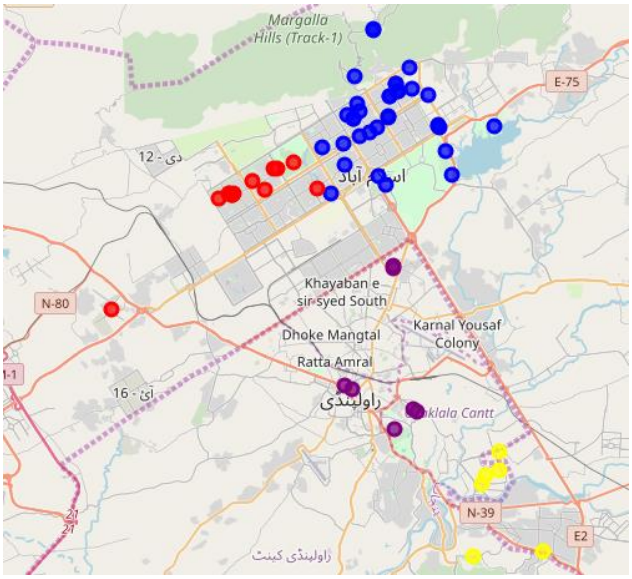

Figure 9

4. Results

(Results section where you discuss the results.)

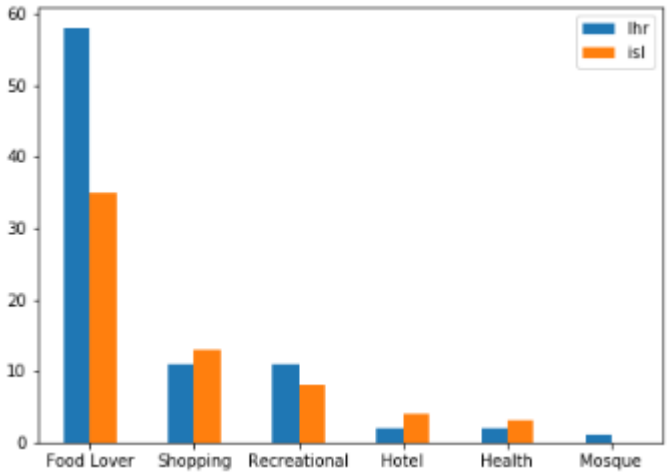

Figure 10

Insights from Figure 10:

★ Lahore has more Food Lover Spots than Islamabad (As per FourSquare App)
★ Lahore has more Recreational Spots than Islamabad (As per FourSquare App)
★ Islamabad is better in hoteling and Health (As per FourSquare App)
★ Lahore and Islamabad are almost equal for Shopping (As per FourSquare App)

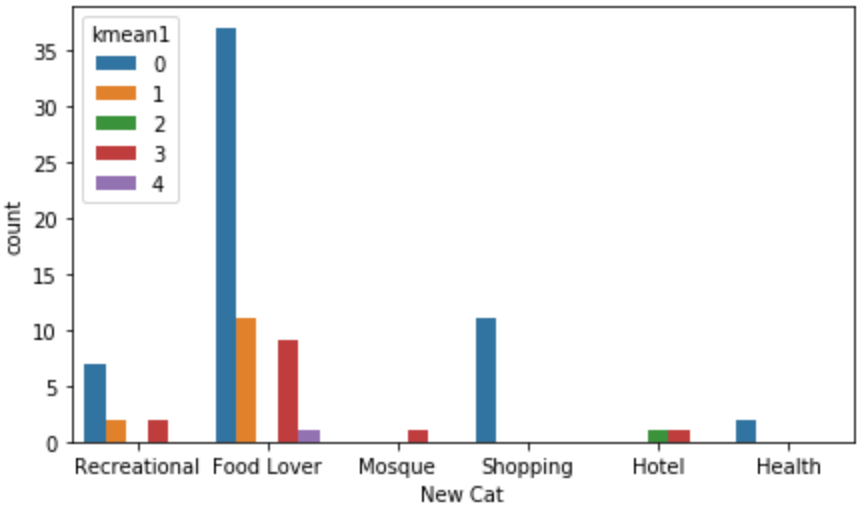

Figure 11: Lahore Data Clustering

Insights from LAHORE Data (Figure 11)

- ★ If you are food lover, you may choose to stay in Cluster 0.
- ★ If you are recreational you may choose Cluster 0.
- ★ If you love shopping, you may choose Cluster 0.
- ★ If you are health conscious then you may choose 0.
- ★ If you want to stay in hotel you have to go to cluster 2 or 3.

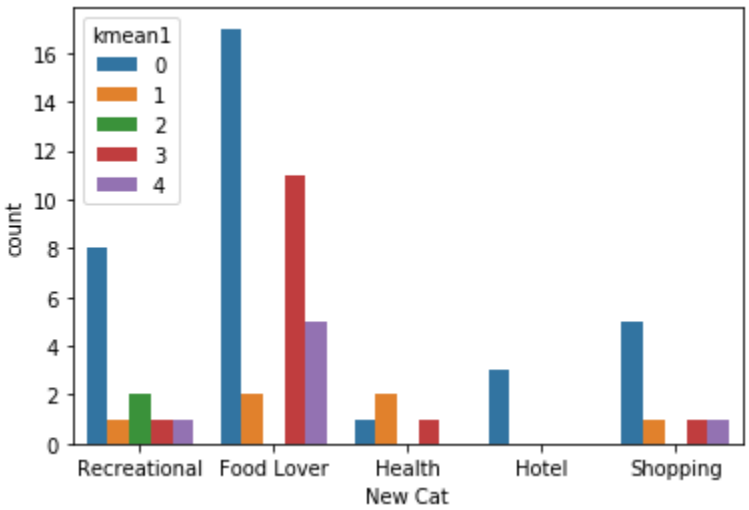

Figure 12: Islamabad Data Clustering

Insights from Islamabad Data (Figure 12)

- ★ If you are food lover, you may choose to stay in Cluster 0 or 3.
- ★ If you are recreational you may choose Cluster 0.
- ★ If you love shopping, you may choose Cluster 0.
- ★ If you are health conscious then you may choose 0,1 or 3.
- ★ If you want to stay in hotel you have to go to cluster 0.

5. Discussion

(Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.)

- ★ From Figure 11, One from Lahore should choose to live in **Cluster 0** for having access to various Categories. (The Blue One). Let's see the Cluster in Map in Figure 13.

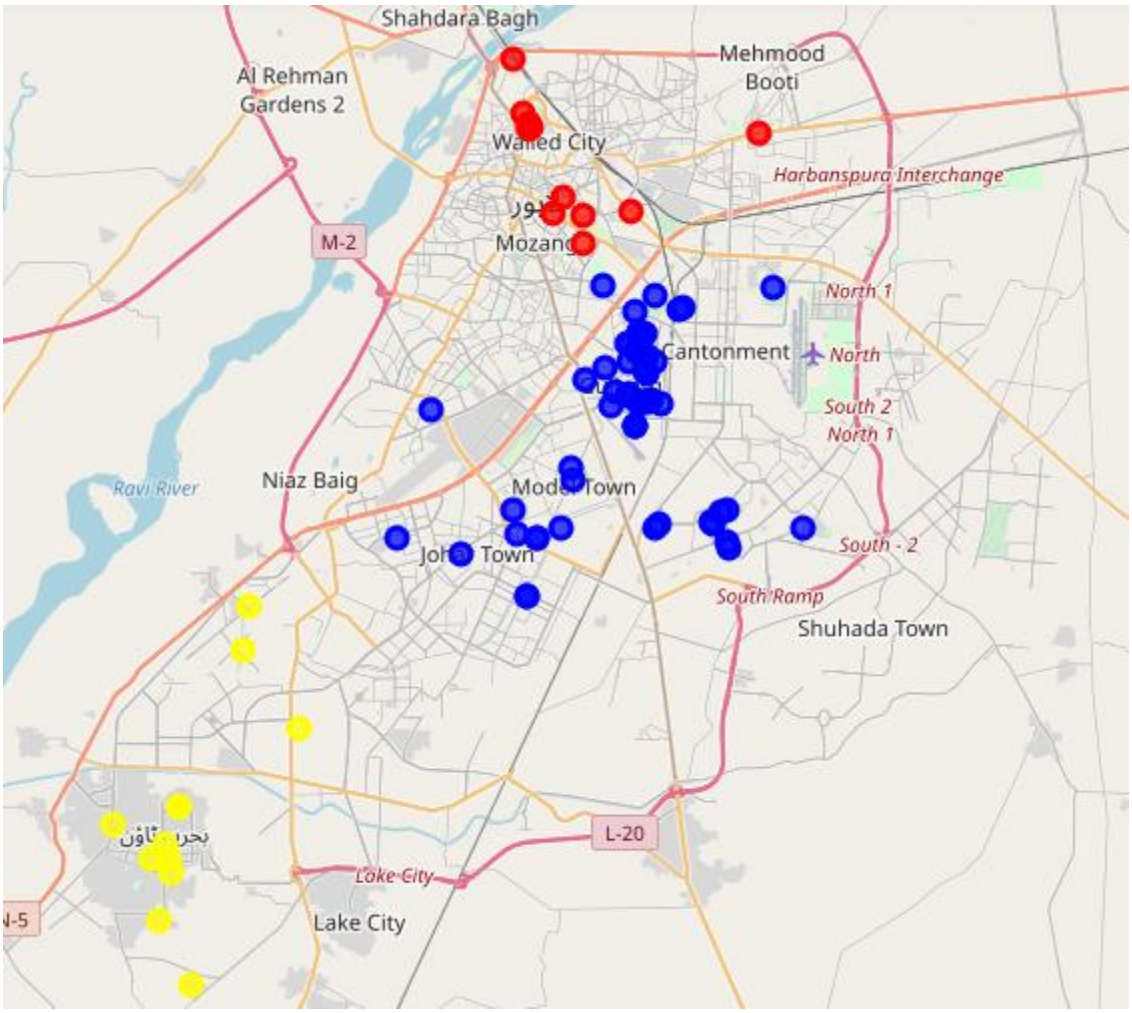

Figure 13

★ From Figure 12, One from Islamabad should choose to live in **Cluster 0** for having access to various Categories. (The Blue One). Let's see the Cluster in Map in Figure 14.

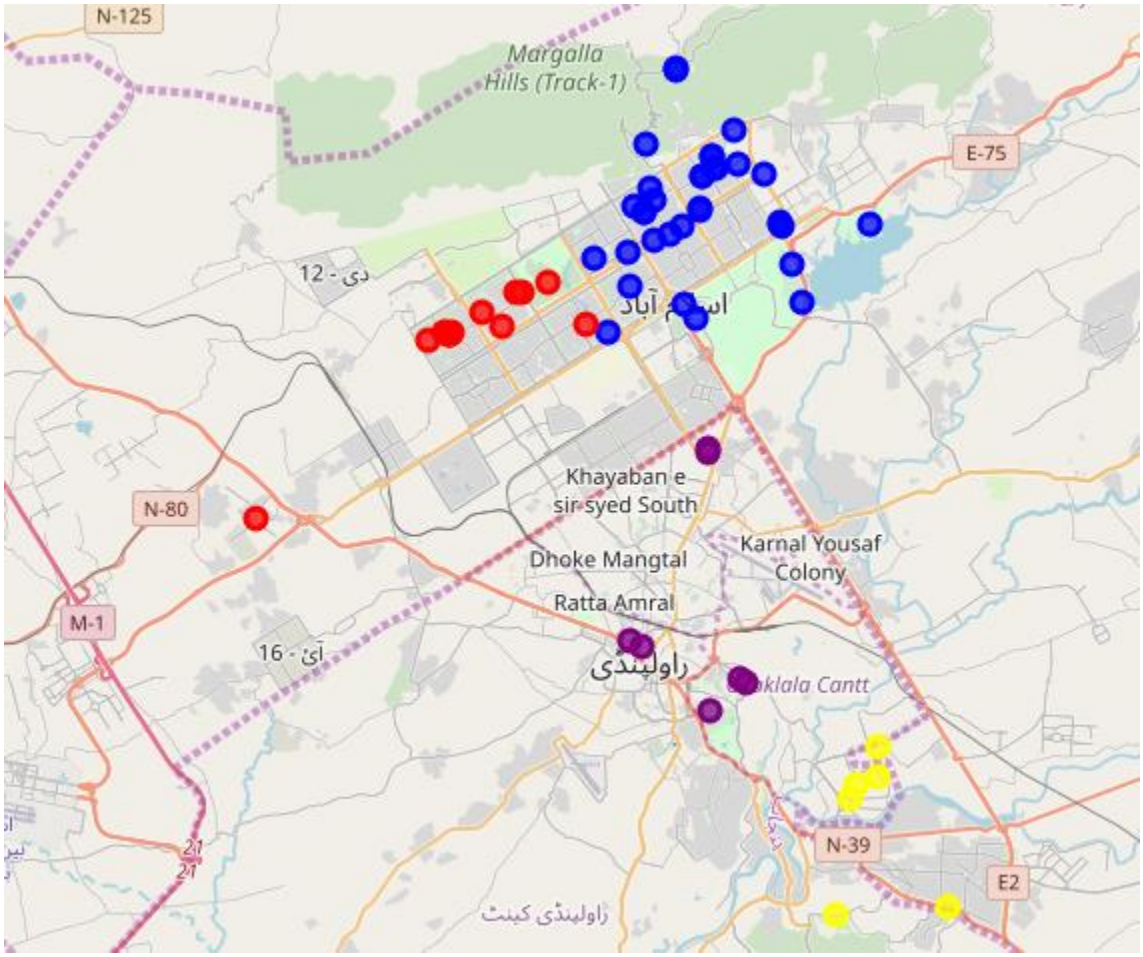

Figure 14

6. Conclusion

(Conclusion section where you conclude the report.)

- ⇨ This report summarizes how we have performed EDA. How we used data to explore the solution of the defined problem.
- ⇨ The both cities are almost equal as per FourSquare data in categories defined.
- ⇨ Through Clustering the excess of nearby spots are shown in blue. Which makes the blue cluster most feasible for multiple category types.
- ⇨ However, a person can choose the cluster depending upon its favorite type of category through figure 11 and 12.