



IBM Applied Data Science Capstone

**Opening Shopping Malls for
tourists in Lahore**

Syed Sajjad Ali

Introduction

- ◉ Find neighborhoods in Lahore, Pakistan where there are not enough shopping malls close to the hotels.
- ◉ Tourists don't need to move across taking taxi or buses to do shopping and utilize the time they save exploring the landmarks of the city for which it is gaining popularity day by day
- ◉ Will eventually help the tourism industry to grow in Lahore

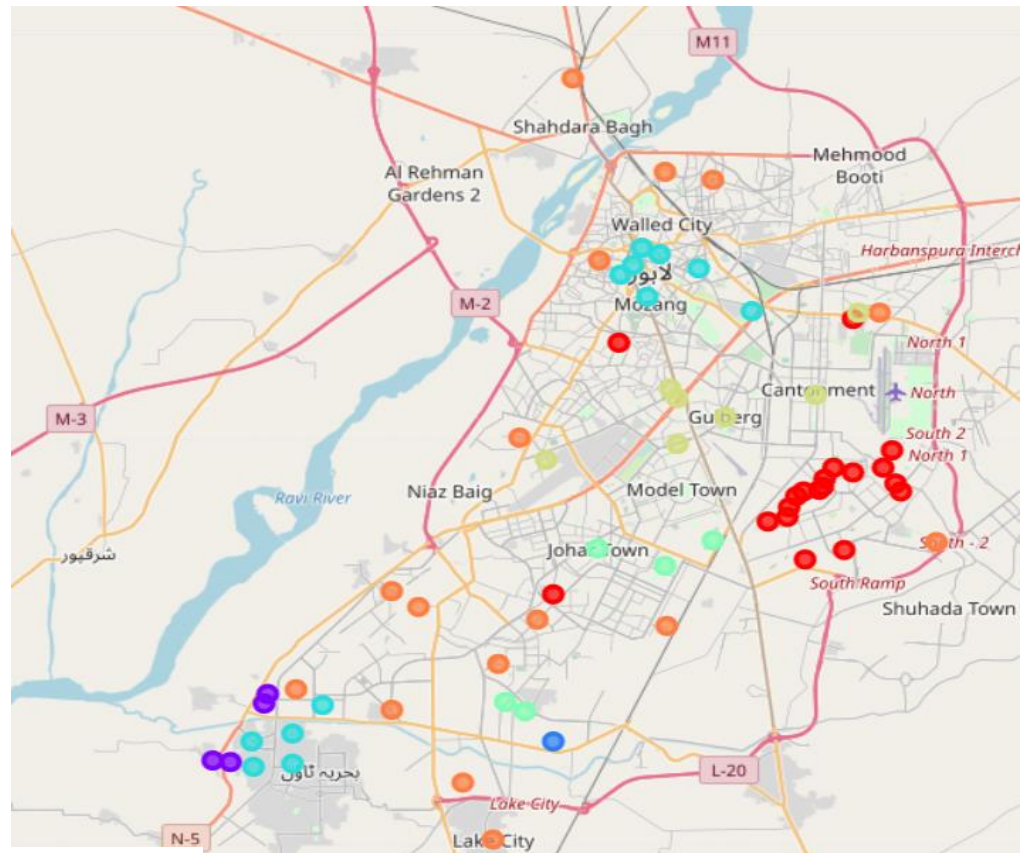
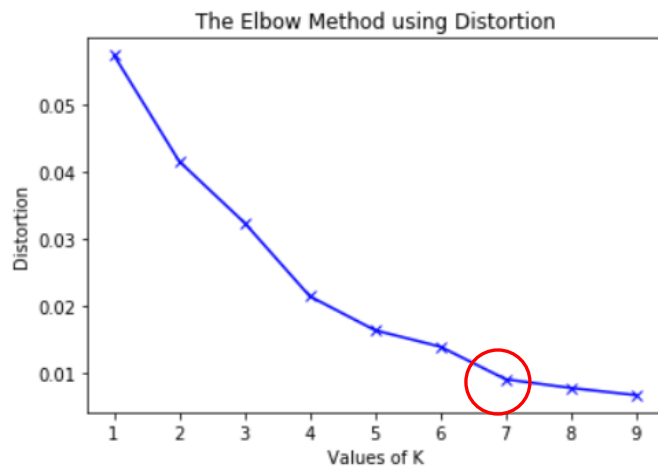
Data

- Lahore suburbs (neighborhoods) dataset from Mapcrow
- Venues dataset for each neighborhood from Foursquare location API
- Python Geocoder Library to get the coordinates of Lahore City

Methodology

- ◉ Importing the dataset from MapCrow
- ◉ Parsing the HTML with BeautifulSoup to get neighborhoods with their coordinates
- ◉ Filtering the data from Lahore District to Lahore City limiting neighborhoods from 159 to 64
- ◉ Get venues for each neighborhood using Foursquare API
- ◉ Group the data by Neighborhoods
- ◉ Apply one hot encoding on the venue categories
- ◉ Group the resultant data by Neighborhoods and calculate mean values for each category
- ◉ Filter data for only Hotels and Shopping Malls
- ◉ Find the optimal number of clusters to be used by using Elbow Diagram
- ◉ Train the model using KMeans Clustering
- ◉ Visualise and examine the clusters
- ◉ Label each cluster and find out the clusters which are of interest

Results



Cluster 0 : Medium/High hotels and Medium/High shopping malls

Cluster 1 : High hotels and no shopping malls

Cluster 2 : No hotels and High shopping malls

Cluster 3 : Medium/High hotels and low shopping malls

Cluster 4 : Low hotels and Medium/High shopping malls

Cluster 5 : No hotels and low shopping malls

Cluster 6 : No hotels and no shopping malls

Discussion

- Cluster 0 gives neighborhoods with a healthy relationship between the hotels and shopping malls since the more hotels are present in the area the more shopping malls are there.
- Cluster 1 gives neighborhoods with a really unhealthy relationship between the hotels and shopping malls since there are no shopping malls in the area although there are hotels
- Cluster 2 gives neighborhoods with no hotels but a presence of shopping malls
- Cluster 3 gives neighborhoods with a really unhealthy relationship between the hotels and shopping malls since there are almost no shopping malls in the area although there are hotels but the number of hotels are lower than that in Cluster 1
- Cluster 4 gives neighborhoods with low hotels but a presence of shopping malls but the presence is low
- Cluster 5 gives neighborhoods with no hotels but a presence of shopping malls but the presence is low
- Cluster 6 gives neighborhoods with no hotels and no shopping malls

Recommendations

- The neighborhoods with the clusters that are "High hotels and no shopping malls" and "Medium/High hotels and low shopping malls" because they satisfy our business problem because in these neighborhoods there is an opportunity to build shopping malls for the tourists
- There are also neighborhoods with no hotels and shopping malls so they can also be targeted in the future if the government has any plans of building hotels in the area

Conclusion

So in conclusion these 16 neighborhoods have potential chance of building shopping malls because of the presence of hotels in the area which will help tourism industry

	Neighborhood	Latitude	Longitude	Cluster Labels	Hotel	Shopping Mall
10	Anarkli	31.5679445	74.3074319	3	0.072727	0.000000
20	Bahria Town	31.3834063	74.1752911	3	0.083333	0.000000
26	Canal Gardens	31.3944129	74.1757053	3	0.090909	0.000000
47	Gawalmandi	31.5723061	74.3176430	3	0.079365	0.015873
70	Jhuggian Ladha Singh	31.3849341	74.1445832	1	0.250000	0.000000
72	Jubilee Town	31.4049640	74.1872397	3	0.125000	0.000000
83	Liaqatabad	31.5647371	74.3028682	3	0.061224	0.000000
88	Maraka	31.3839755	74.1514837	1	0.250000	0.000000
89	Mazang	31.5560352	74.3129047	3	0.081967	0.000000
119	PAEC Foundation Housing	31.3822401	74.1603139	3	0.100000	0.000000
124	Police Lines	31.5668940	74.3329883	3	0.098361	0.000000
127	Punjab Govt Servants Housing Foundation	31.4057118	74.1643178	1	0.250000	0.000000
153	Sukh Chayn Gardens	31.3914660	74.1597798	3	0.100000	0.000000
155	TRICON VILLAGE	31.4093872	74.1657361	1	0.250000	0.000000
160	Urdu Bazar	31.5744573	74.3111106	3	0.075472	0.000000
167	Zaman Park	31.5508554	74.3537354	3	0.093333	0.013333



THANK YOU