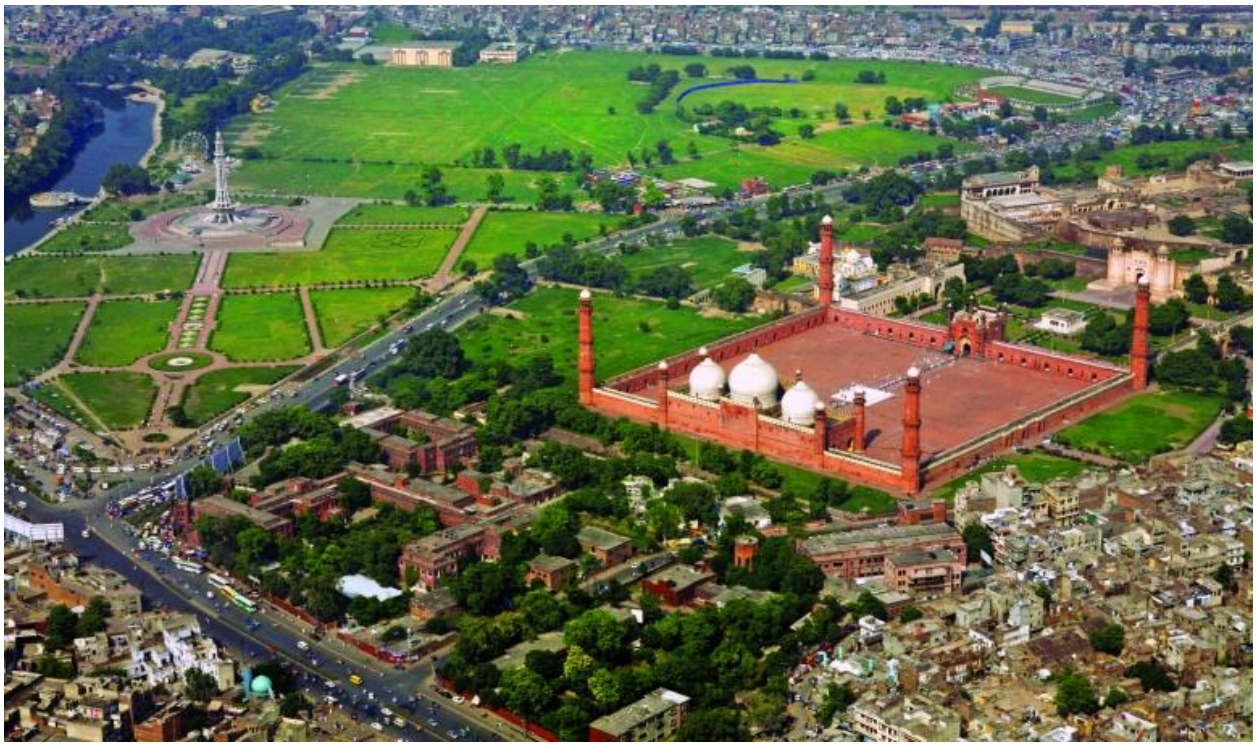
Contents

Introduction/Business Problem

Brief Description about Lahore

Lahore is the capital city of Punjab Province with more than 11 Million according to the 2017 Census and its keep increasing due to immigrants from the other cities. Lahore exerts a strong cultural influence over Pakistan. The city also hosts much of Pakistan's tourist industry, with major attractions including the Walled City, the famous Badshahi and Wazir Khan mosques and Sikh shrines. Lahore is also home to the Lahore Fort and Shalimar Gardens, both of which are UNESCO World Heritage Sites.



Problem Description

The Lahore City is extremely rich and diverse culture. Since 2015, 274 Union Councils have existed in the City District of Lahore.

Let’s assume that you live in any neighborhood council of Lahore and You love your neighborhood, mainly because of all the great amenities and other types of venues that exist in the neighborhood, such as restaurants, shopping mall, historic site. And now you have an offer letter from a great company on the other neighborhood council of the city and great opportunity. However, it is far away from your current residence and you want to relocate yourself in another union council. Wouldn't it be great if you are able to determine neighborhoods on the other side of the city that are the same as your current neighborhood, and if not, perhaps similar neighborhoods that are at least closer to your new job?

The aim of this project is to explore the Neighborhood Councils of Lahore city to find the 10 most common venues in each Union Council.

This information provided by this report would be useful for people who are interested in relocating to a different part of the city and are interested in finding new neighborhoods that are highly similar to their existing neighborhood.

Data

Neighborhood Councils

The Data required for this project mainly extracted from the wikipedia

In order to explore the neighborhood councils, firstly we have to extract the list of councils from https://en.wikipedia.org/wiki/List_of_towns_in_Lahore

```
Union Council No 1 Begum Kot Shahdara
Union Council No 2 Yousif Park
Union Council No 3 Kot Kamboh
Union Council No 4 Shamsabad Shahdara
Union Council No 5 Chah Jhabbay Wala
…
…
…
…
Union Council No 269 Bhoptian
Union Council No 270 Araiyan
Union Council No 271 Jia Bagga
Union Council No 272 Raiwind Rural
Union Council No 273 Raiwind Urban
Union Council No 274 Bablian Otar
```

Geolocation

Our Data set did not contain location data. So we have to extract the geolocation of each council from geopy library. Unfortunately, it returned location of only 123 council and we had to drop the other council due to unavailability of location

	Union Council	Name	NL	location	point
0	8	Faisal Park	Faisal Park,Lahore Pakistan	(Faisal Park, Makhan Pura, China Scheme, لاہور...	(31.6037915, 74.35944568430989, 0.0)
1	11	Majeed Park	Majeed Park,Lahore Pakistan	(Abid Majeed Road (Featherston Road), St. John...	(31.5223002, 74.3723858, 0.0)
2	12	Qazi Park	Qazi Park,Lahore Pakistan	(Qazi Muhammad Isa Road, Faisal Town, Johar To...	(31.4756793, 74.3053802, 0.0)
3	18	Hanif Park	Hanif Park,Lahore Pakistan	(Hanif Garder Street 1, Sharif Pura, Haji Pura...	(31.5871834, 74.4201815, 0.0)
4	19	Siddique Pura	Siddique Pura,Lahore Pakistan	(Jamia Masjid Siddique-e-Akbar, Canal Bank Roa...	(31.5748716, 74.42893671669796, 0.0)
...	...	...	...	...	...
118	262	Shamkay Bhattian	Shamkay Bhattian,Lahore Pakistan	(Shamkay Bhattian, Lahore District, پاک, پنجاب...	(31.3305959, 74.1076545, 0.0)
119	263	Manga	Manga,Lahore Pakistan	(Manga, Lahore District, پاکستان, 55270, پنجاب...	(31.3070462, 74.0485141, 0.0)
120	270	Araiyan	Araiyan,Lahore Pakistan	(Faisalabad-Sheikhupura-Lahore Road, Toll Plaz...	(31.4783407, 73.2237338, 0.0)
121	271	Jia Bagga	Jia Bagga,Lahore Pakistan	(Jia Bagga, Lake City Main Boulevard, Fazaia H...	(31.3304027, 74.2667851, 0.0)
122	272	Raiwind Rural	Raiwind Rural,Lahore Pakistan	(Raiwind, Lahore District, پاکستان, 55150, پنجاب...	(31.2442237, 74.215911, 0.0)

123 rows × 5 columns

FourSquare API

To find out the nearby venues in Lahore city, we will use foursquare api. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the Councils. The information obtained per venue as follows:

	Name	latitude	longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Faisal Park	31.603791	74.359446	Optp	31.601458	74.358464	Restaurant
1	Majeed Park	31.522300	74.372386	Aramish	31.524728	74.371419	Spa
2	Majeed Park	31.522300	74.372386	GOGO Resturant	31.522615	74.368309	Restaurant
3	Majeed Park	31.522300	74.372386	Shell	31.521304	74.368287	Gas Station
4	Majeed Park	31.522300	74.372386	Haider Bakery	31.524237	74.376035	Bakery

Methodology

Web Scrapping

To start the analysis, I used Beautiful Soup library to scrap the neighborhood councils in Lahore and converted it into pandas DataFrame after cleaning.

	UCNo	Name
0	1	Begum Kot Shahdara
1	2	Yousif Park
2	3	Kot Kamboh
3	4	Shamsabad Shahdara
4	5	Chah Jhabbay Wala

Location Extraction

The location points of each union council was obtained from geopy library using user agent "myGeocoder" , which is one the best open source geolocation library available. And RateLimiter, which is convenient function to delay between geocoding calls, was used. Out of 272 union council it retrieved only 124 location. Our data contains missing geolocation points so we have to remove the rows containing null value in location.

	UCNo	Name	NL	location	point	latitude	longitude
0	8	Faisal Park	Faisal Park,Lahore Pakistan	(Faisal Park, Makhan Pura, China Scheme, لاہور...	(31.6037915, 74.35944568430989, 0.0)	31.603791	74.359446
1	11	Majeed Park	Majeed Park,Lahore Pakistan	(Abid Majeed Road (Featherston Road), St. John...	(31.5223002, 74.3723858, 0.0)	31.522300	74.372386
2	12	Qazi Park	Qazi Park,Lahore Pakistan	(Qazi Muhammad Isa Road, Faisal Town, Johar To...	(31.4756793, 74.3053802, 0.0)	31.475679	74.305380
3	18	Hanif Park	Hanif Park,Lahore Pakistan	(Hanif Garder Street 1, Sharif Pura, Haji Pura...	(31.5871834, 74.4201815, 0.0)	31.587183	74.420181
4	19	Siddique Pura	Siddique Pura,Lahore Pakistan	(Jamia Masjid Siddique-e-Akbar, Canal Bank Roa...	(31.5748716, 74.42893671669796, 0.0)	31.574872	74.428937
...	...	...	...	...	...	...	...
119	262	Shamkay Bhattian	Shamkay Bhattian,Lahore Pakistan	(Shamkay Bhattian, Lahore District, پنجاب, پاک...	(31.3305959, 74.1076545, 0.0)	31.330596	74.107654
120	263	Manga	Manga,Lahore Pakistan	(Manga, Lahore District, 55270, پنجاب, پاکستان...	(31.3070462, 74.0485141, 0.0)	31.307046	74.048514
121	270	Araiyan	Araiyan,Lahore Pakistan	(Faisalabad-Sheikhupura-Lahore Road, Toll Plaz...	(31.4783407, 73.2237338, 0.0)	31.478341	73.223734
122	271	Jia Bagga	Jia Bagga,Lahore Pakistan	(Jia Bagga, Lake City Main Boulevard, Fazaia H...	(31.3304027, 74.2667851, 0.0)	31.330403	74.266785
123	272	Raiwind Rural	Raiwind Rural,Lahore Pakistan	(Raiwind, Lahore District, 55150, پنجاب, پاکست...	(31.2442237, 74.215911, 0.0)	31.244224	74.215911

124 rows × 8 columns

Start analyzing neighborhood using location data. However, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in

Nearby Venues

Foursquare data is very comprehensive and it powers location data for Apple, Uber etc. For this business problem I have used, as a part of the assignment, the Foursquare API to retrieve information about the popular spots around these 5 Major Districts of Tokyo. The popular spots returned depends on the highest foot traffic and thus it depends on the time when the call is made. So we may get different popular venues depending upon different time of the day. The call returns a JSON file and we need to turn that into a data-frame.

I set the LIMIT parameter to 100, which would limit the number of venues returned by the Foursquare API and the radius of 500 meter. Here is a head of the list of Nearby Venues.

	Name	Name Latitude	Name Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Faisal Park	31.603791	74.359446	Optp	31.601458	74.358464	Restaurant
1	Majeed Park	31.522300	74.372386	Aramish	31.524728	74.371419	Spa
2	Majeed Park	31.522300	74.372386	GOGO Resturant	31.522615	74.368309	Restaurant
3	Majeed Park	31.522300	74.372386	Shell	31.521304	74.368287	Gas Station
4	Majeed Park	31.522300	74.372386	CSD Mall Road	31.526227	74.371276	Department Store

Result Analysis

We use One Hot Encoding, use the neighborhood to group data, and find out the top ten venues present in each neighborhood.

	Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Akram Park	Asian Restaurant	Breakfast Spot	Zoo	Food	Convenience Store	Department Store	Dessert Shop	Diner	Doctor's Office	Fast Food Restaurant
1	Al Faisal Town	BBQ Joint	Brewery	Shopping Mall	Asian Restaurant	Breakfast Spot	Food	Department Store	Dessert Shop	Diner	Doctor's Office
2	Amin Park	Gourmet Shop	Restaurant	Pakistani Restaurant	Travel & Transport	Gas Station	Garden	Clothing Store	Coffee Shop	Comfort Food Restaurant	Convenience Store
3	Azam Market	Historic Site	Accessories Store	Bakery	Market	Flower Shop	Convenience Store	Department Store	Dessert Shop	Diner	Doctor's Office
4	Baghban Pura	Cemetery	Fast Food Restaurant	Zoo	Flower Shop	Convenience Store	Department Store	Dessert Shop	Diner	Doctor's Office	Fish & Chips Shop

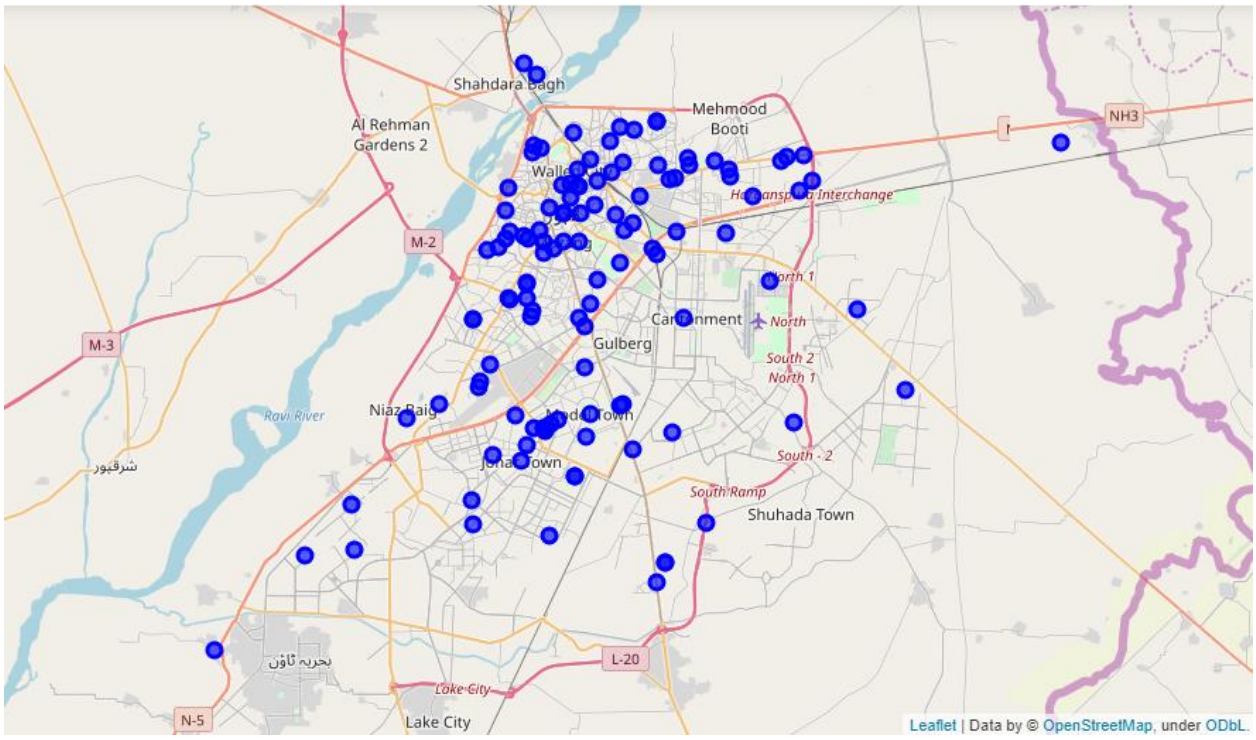
We have some common venue categories in the neighborhoods. We use the unsupervised learning K-means algorithm to cluster the neighborhoods. K-Means algorithm is one of the most common method for clustering in unsupervised learning.

We use a k_cluster value of 5 to split the neighborhoods into 5 different clusters based on the similarity they have concerning the venues they contain.

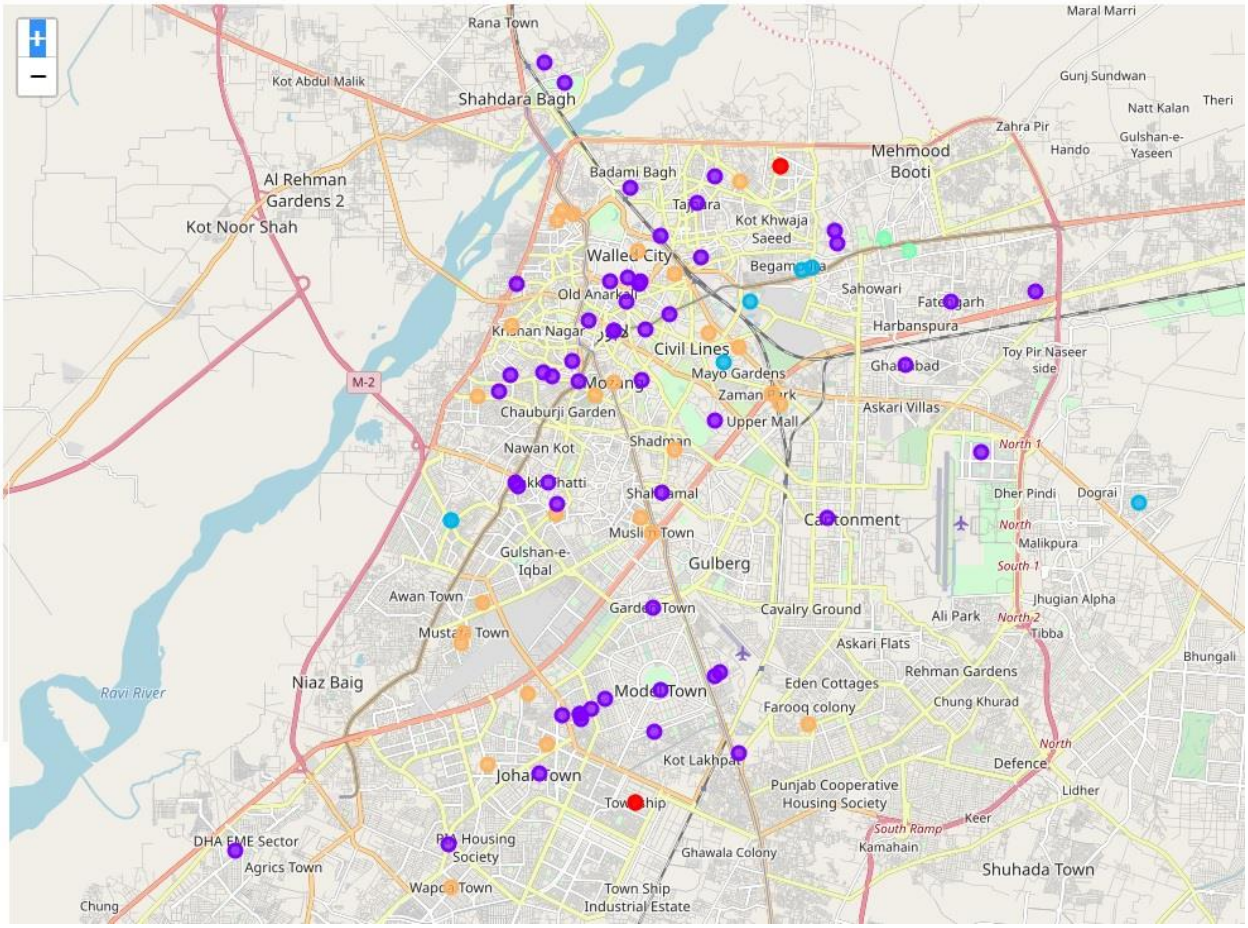
latitude	longitude	altitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
31.603791	74.359446	0.0	0	Restaurant	Zoo	Flea Market	Comfort Food Restaurant	Convenience Store	Department Store	Dessert Shop	Diner	Doctor's Office	Fast Food Restaurant
31.603791	74.359446	0.0	0	Restaurant	Zoo	Flea Market	Comfort Food Restaurant	Convenience Store	Department Store	Dessert Shop	Diner	Doctor's Office	Fast Food Restaurant
31.522300	74.372386	0.0	1	Chinese Restaurant	Gas Station	Department Store	Restaurant	Spa	Fish & Chips Shop	Coffee Shop	Comfort Food Restaurant	Convenience Store	Dessert Shop
31.475679	74.305380	0.0	1	BBQ Joint	Brewery	Shopping Mall	Asian Restaurant	Breakfast Spot	Food	Department Store	Dessert Shop	Diner	Doctor's Office
31.574872	74.428937	0.0	1	Bus Station	Zoo	Flower Shop	Convenience Store	Department Store	Dessert Shop	Diner	Doctor's Office	Fast Food Restaurant	Fish & Chips Shop

Visualization and Exploration

Folium is a python library that can create interactive leaflet map using coordinate data.



1: Before Clustering



2: After Clustering

Conclusion and Discussion

I have created 5 different clusters and each cluster have similar neighborhood venues such as restaurants, parks and historic places. The purpose of this project was to give the options where to relocate in the City of Lahore.

As a recommendation to those who plan to relocate, location selection is only one fundamental problem to think over. Final decision on relocation will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / status of every neighborhood etc.

References

https://en.wikipedia.org/wiki/List_of_towns_in_Lahore

https://developer.foursquare.com

https://pypi.org/project/geopy/