Battle of Neighborhoods

Subtitle

Introduction

- The emergence and rising popularity of AirBnb has inititated a new investment oppurtunity for the Pakistanis. People are keen towards purchasing properties in localilties nearest to popular tourist venues so that they can be made available for rent through AirBnB. In comparison to keeping a property occupied by tenants, the purchased property yields a higher rental earning from an investor point of view if rented out on AirBnB.

- My project is targetted to help such investors. I shall be using my knowledge from the ibm data science professional courses to identify and choose a neighborhood best suited for buying a property in Lahore. My selection will be based on the locality of the neighborhood. The closer a neighborhood is to the popular venues of Lahore, the higher will be its chances that it will be preferred for renting by tourists.

Data

- 1.Zameen Property Data.csv

- 2.Venues from Foursquare location data

Methodology

1. Data Cleansing

- As the csv file imported, has all the properties listed for sale in the country of Pakistan and our project is solely based on one city Lahore, therefore we know that we shall read many unneeded details from the csv file. For that purpose, we have to filter out the file in detail and get rid of any columns that play no role in our analysis. For this purpose, we drop all such columns.

- Another cleansing applied is towards the rows in the dataframe read from csv file. We keep focused on only those properties that are located near the neighborhoods in the battle and are constructed houses. We need constrcuted houses because the objective is to purchase a property and put it on AirBnB for rental purposes. Therefore,we keep only those houses that are located nearby the three neighborhoods chosen. Also all unconstructed properties and flats etc are ignored.

2.K Means Clustering

- In the next step, we shall use KMeans clustering algorithm to group all the houses for sale located nearby and inside the three neighborhoods. When this is done, all properties will be classified as belonging to one neighborhood according to its nearness of location to a neighborhood. For this activity we have chosen to make three clusters. Therefore all rows selected from the above step will be classified in three clusters according to KMeans clustering

3. Discovering Venues Around Clusters Using FourSquare

- Now we shall use the credentials of our foursquare account to discover venues around each cluster. The objective is to find out a count of venues near to each neighborhood therefore we shall take the average latitude and the average longitude of each cluster. It shall represent the centre point of a neighborhood. Then, we shall keep a 5 kilometer radius and get all venues from Foursquare. FourSquare will give details about each venue, it category, its coordinates and even its rating, We are concerned about the count of venues near each cluster so we will first discover all venues and then count them for comparison.

4. Choosing a Neighborhood

- For each cluster, I shall compare the count of venues retreived by FourSquare. The neighborhood with the highest venues around will be selected for a property selection whereas the other two neighborhoods will be considered as the losers of the battle of neighborhoods.

Results and Discussion

- The csv file read from dataset contained 168447 properties listed for sale in Pakistan. The battle of choosing a neighborhood for purchasing a property for AirBnb purpose was kept between Gulberg Town, Bahria Town and Lake City of Lahore City only.

- Data cleansing was perfromed and 4786 properties were chosen to be of value for this project.

- KMeans clustering was used to group the 4768 properties in three clusters. Each cluster contains properties that are located close to the centre of a neighborhood hence based on locality. They were shown on map

- Foursquare API was used to find out venues nearby each cluster. We see that the neighborhood of Gulberg Town has 100 venues nearby it while Bahria Town has 12 venues and Lake City neighborhood has just 4 venues'

- A tourist person always wants to visit nearby iconic places so he would prefer to reside somewhere nearby to all major venues. If we want to purchase a property to put on AirBnB so that tourists may book it for rent, then Gulberg Town is the neighborhood selected

Conclusion

- We have compared three neighborhoods of the city of Lahore and utilized clustering algorithm as well location data from FourSquare to find solution to a business problem. This project has helped me revise all concepts studied in data science so far. And it has been interesting to make an analysis on the real time data on a city of my own country Pakistan.