COURSERA CAPSTONE PROJECT

IBM APPLIED DATA SCIENCE CAPSTONE PROJECT REPORT

NEIGHBOURHOOD SEGMENTATION AND CLUSTERING FOR HYDERABAD, INDIA USING LOCATION DATA THROUGH FOURSQUARE API

BY: PRANAV KARAMCHEDU

INTRODUCTION

The love that the city of Hyderabad and its people show towards food and drinks is very well known. It is one the major foodie spots in India and people travel from all over the country to visit some of the famous restaurants which have developed a name for themselves through their quality of service offered and have become brands synonymous with certain foods. The act of opening a restaurant in such a booming region is an interesting opportunity for investors, chefs, and general food lovers in the city. To succeed in such a venture apart from quality in services offered good marketing, catering to the public demand, availability and "brand recognition" and other factors play an important role in deciding how successful and well known the restaurant is and how much attention it attracts. The location of a restaurant plays a major role in the success or failure it. A well-placed restaurant with regards to public demand and

selecting the right neighbourhood are very important which is the main aim of this project.

BUSINESS PROBLEM

The objective of this capstone project is to segment and cluster the different neighbourhoods in the city of Hyderabad to analyse its different locations and select the most optimum location(neighbourhood) to open a restaurant in the city.

TARGET AUDIENCE OF THE PROJECT

The target audience of this project are investors and chefs looking to open a restaurant in the city of Hyderabad.

DATA AND ITS SOURCES

To tackle this problem, we use the following data to complete the analysis:

- List of neighbourhoods in the city of Hyderabad: obtained from the Wikipedia page.
- Geospatial data (latitude and longitude coordinates) of the neighbourhoods and venues in the project.
- Location data (venues, type of venue etc) from the Foursquare API which helps in the process of clustering the neighbourhoods.

METHODOLOGY

The first step in this project is to obtain the neighbourhood data by web scraping the Wikipedia page.

(https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Hyderabad,_India).

This is achieved by using requests, BeautifulSoup and pandas packages available in Python. Then, using the geocoder package we obtain the geospatial data of the neighbourhoods and for the city of Hyderabad (using the Nominatim package of geopy.geocoders. After plotting the different neighbourhoods using the folium package, we use the Foursquare API to get the location data of different venues and complete the exploratory analysis. After we have the final dataset ready, we cluster the neighbourhoods in k clusters (value of k decided using the elbow plot method) using the K-Means Clustering algorithm. Based on the results of the clustering and some further analysis, we come to conclusion as to which neighbourhood in the city is the optimum location to open a new restaurant and decide on the type of restaurant along the way.

RESULTS AND DISCUSSION

From the final table of seven neighbourhoods A.S Rao Nagar would be a good choice to start an Indian Restaurant as it has the joint least number of Indian restaurants in the cluster and it an urban area and also a major commercial and residential neighbourhood in the city of Hyderabad. This analysis can be further improved by selecting a target parameter based on which the suitability of a location can be decided based on the target parameter and further deciding the suitability of a neighbourhood in catering to a particular type of venue. The end location chosen in part of a cluster where similar such restaurants are of good popularity, but in a neighbourhood with not too many such venues, hence reducing the competition from other venues. The target audience of the selected location is also an added benefit which increases the amount of eyes/attention to the restaurant. The neighbourhood recommended at the end though can act as the starting point for further and deep analysis which could eventually result in location which has not only no nearby competition, but also other factors considered, and all other relevant conditions met.

CONCLUSION

The main purpose of the project was to cluster neighbourhoods in the city of Hyderabad to select an optimum location for the opening of a new restaurant. In the analysis we also looked at which type of restaurant (Indian) would be the most popular choice and the end decision was made considering the clustering results, demographic data, and similarity of neighbourhoods. To make a thorough decision additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighbourhood etc can also be considered.