THE BATTLE OF NEIGHBOOURHOODS

CAPSTONE PROJECT FOR IBM DATA SCIENCE

SUBMITTED BY:

UTKRIST SINGH

[Email-uksingh1213@gmail.com](mailto:Email-uksingh1213@gmail.com)

Submitted on: 08-AUG-2020

INTRODUCTION

Allahabad is one of the major and most ancient cities of Uttar Pradesh , India. Tourists flock in at large numbers and the city remains crowded throughout the year. Hence for the benefit of tourists, a data analysis model that informs them about the neighbourhoods has been worked upon. This project might serve to be a starting point in creating a model useful for Allahabad Tourism. The idea can be extended, of course, to other cities as well.

DATA INSIGHT

Data sets that have been used in this project are simply the pincodes lists from Wikipedia. All other data has been either imported using Foursquare API or derived from original data set.

METHODOLOGY

The methodology used in this project is:

1. Data Input: csv file containing pincodes
2. Data Study: studying the data set, data types etc
3. Data Wrangling: Removing columns that were not needed
4. Data Visualization: Visualising the neighbourhoods on map using folium

1. API Calling: Leveraging Foursquare API to get nearby venues
2. Final Data Wrangling: The data returned by foursquare is converted to pandas dataframe and modified.
3. Clustering: Using k-means clustering to find the busiest neighbourhood.
4. Visualization: Using folium to visualize clusters on map and bar charts for a visual analysis.

RESULTS

After in-depth analysis, the results that followed were:

1. The main land Allahabad has more closely placed clusters than the outskirts.
2. Civil Lines, Katra and Chowk have the most dense clusters and hence are the busiest neighbourhoods of the city.
3. The difference in the density of clusters is less in the main land and it increases as we move to the outskirts.

DISCUSSIONS

The main land Allahabad has more closely placed clusters than the outskirts.

Civil Lines, Katra and Chowk have the most dense clusters and hence are the busiest neighbourhoods of the city.

The difference in the density of clusters is less in the main land and it increases as we move to the outskirts.

CONCLUSION

This project would not have been successful without the co-operation of IBM Skills Network and of course, Coursera. I feel privileged to be a part of this project.

Allahabad is one of the major and most ancient cities of Uttar Pradesh , India. Tourists flock in at large numbers and the city remains crowded throughout the year. Hence for the benefit of tourists, a data analysis model that informs them about the neighbourhoods has been worked upon. This project might serve to be a starting point in creating a model useful for Allahabad Tourism. The idea can be extended, of course, to other cities as well.

Accuracy can be made better and more realistic insights can be added later in this project. The maps and other data visualizations help in analyzing the neighborhoods easily, and they can be made more coherent in future.

THANK YOU!