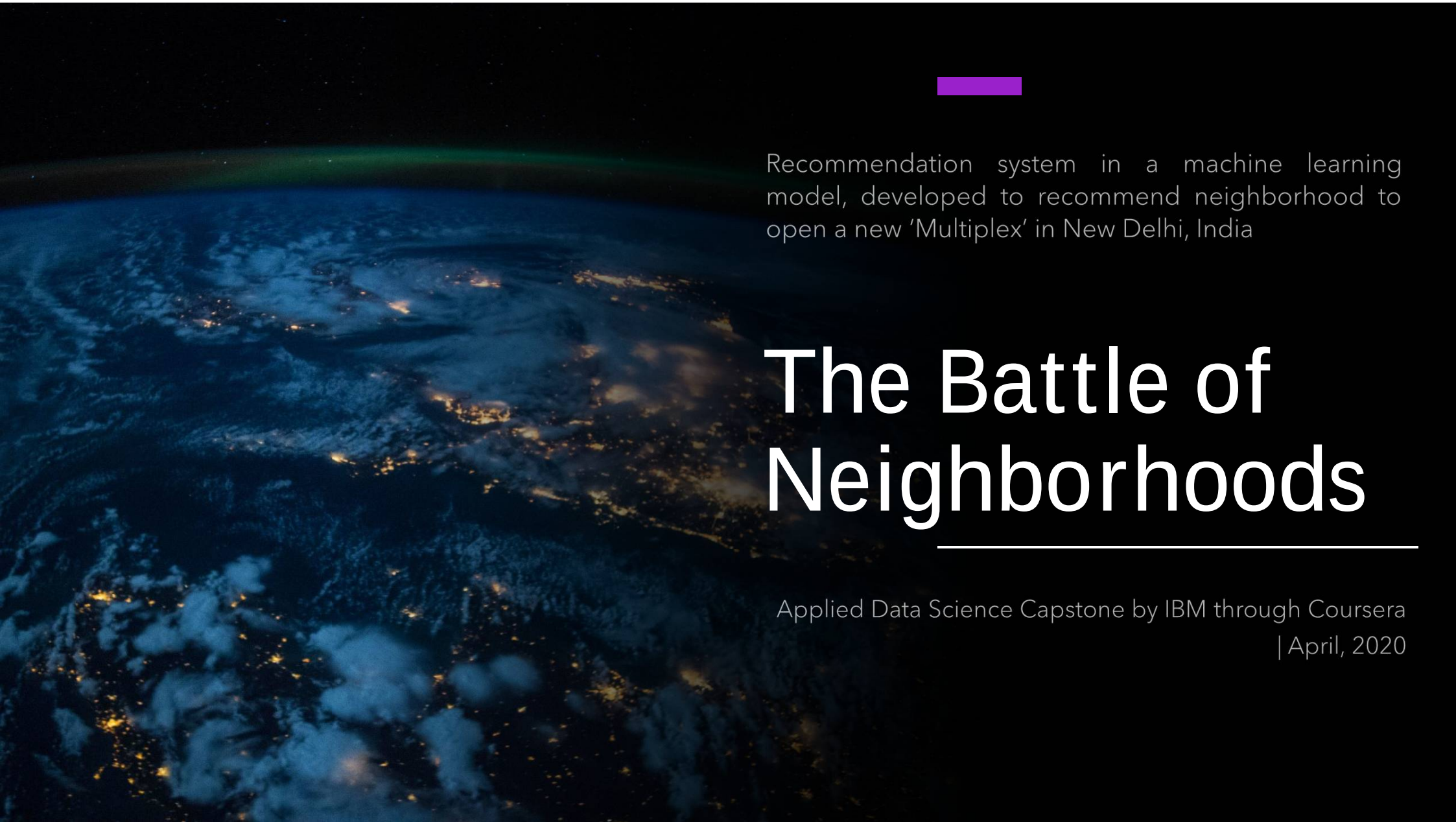


Recommendation system in a machine learning model, developed to recommend neighborhood to open a new 'Multiplex' in New Delhi, India

The Battle of Neighborhoods

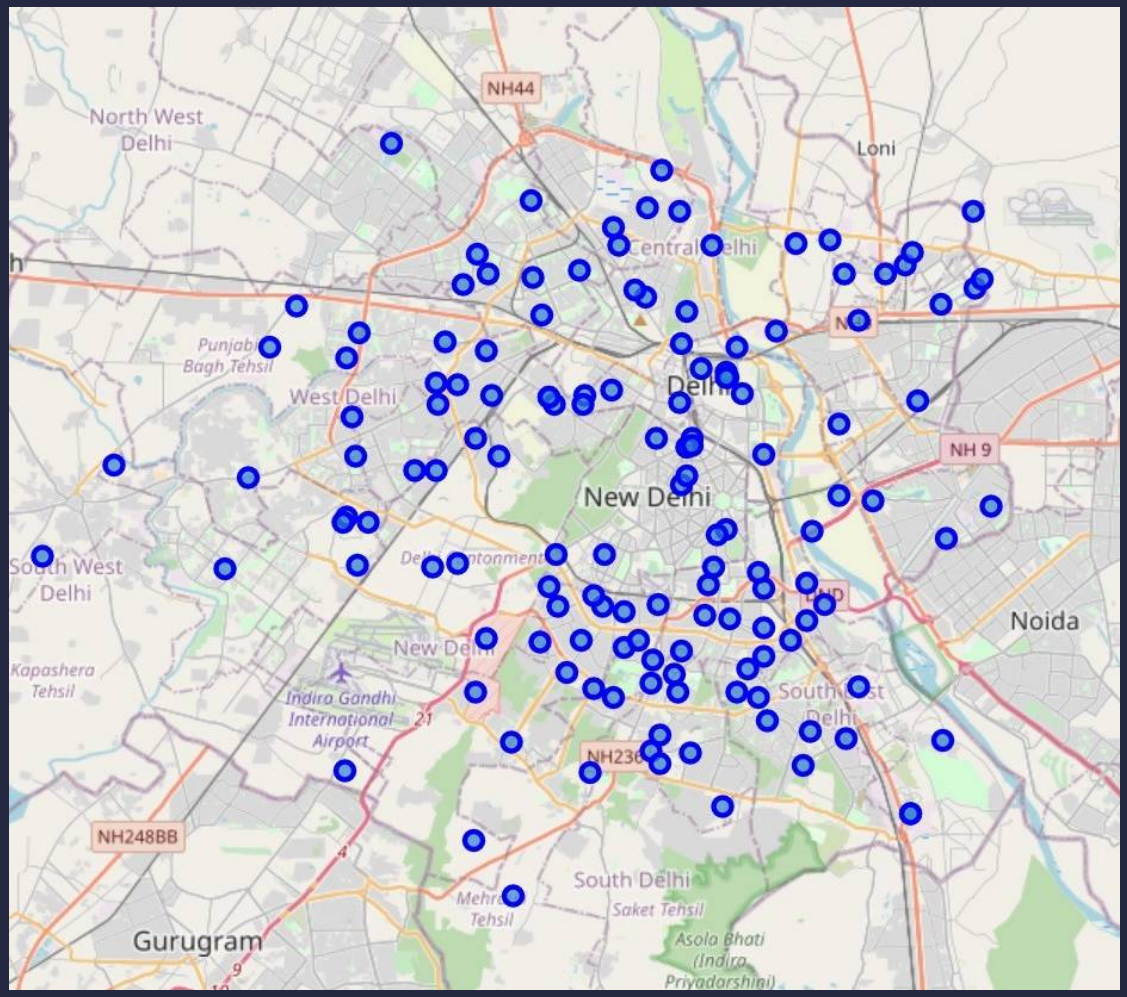
Applied Data Science Capstone by IBM through Coursera

| April, 2020

Introduction

Analysis and selection of best cluster and neighborhood location in New Delhi, India to open a new ‘Multiplex’

Data science methodology and machine learning techniques such as clustering and statistical plotting library of seaborn have been used to provide best solutions at arriving at the best cluster along with illustrative example to derive location in a neighborhood.



Data Analysis – Description of data and its usage

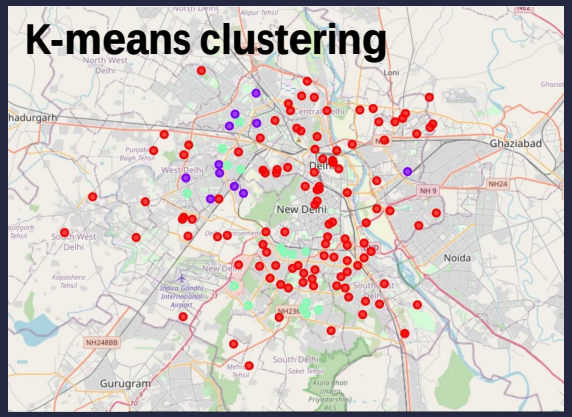
- List of neighborhoods in New Delhi, India

- Latitude and Longitude coordinates of identified neighborhoods for plotting on maps and arriving at the venue data of various neighborhoods.
-
- Use of Foursquare API for fetching the nearest venue locations for defining clusters along with details of venues and their names.

- Venue data, especially that related to ‘Multiplex’.

Methodology

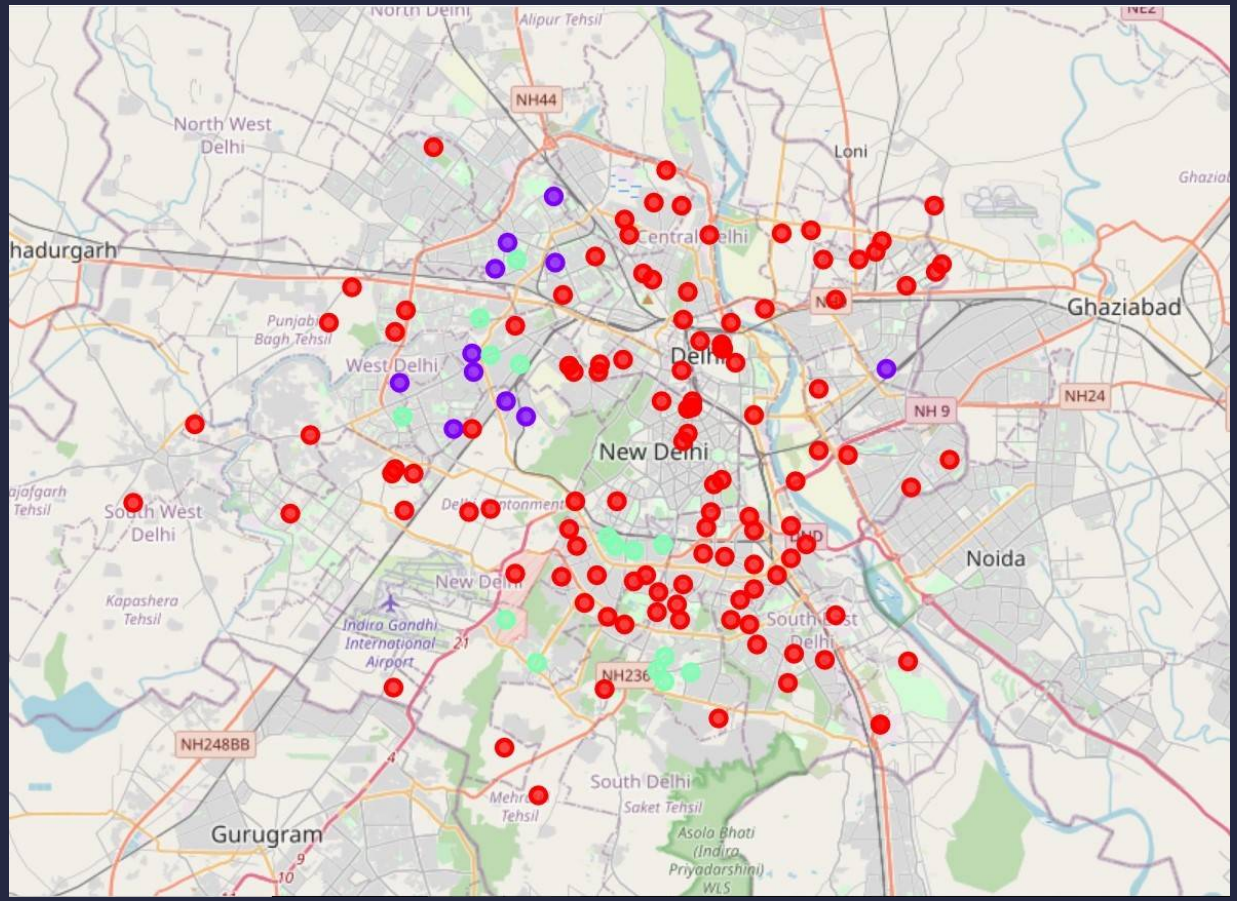
- Web scraping and building dataframe for identifying neighborhoods in New Delhi, India
- Geographical coordinates of identified neighborhoods
- Use of Foursquare API for exploring identified neighborhoods
- K-means clustering
- Statistical Analysis of clusters



Result of the Analysis (1/2)

- Result of k-means clustering: The results shows that amongst 3 clusters, frequency of occurrence of 'Multiplex' can be analyzed as below:
  - Cluster 0: Neighborhoods with least number of multiplexes
  - Cluster 1: Neighborhoods with high concentration of multiplexes
  - Cluster 2: Neighborhoods with moderate concentration of multiplexes

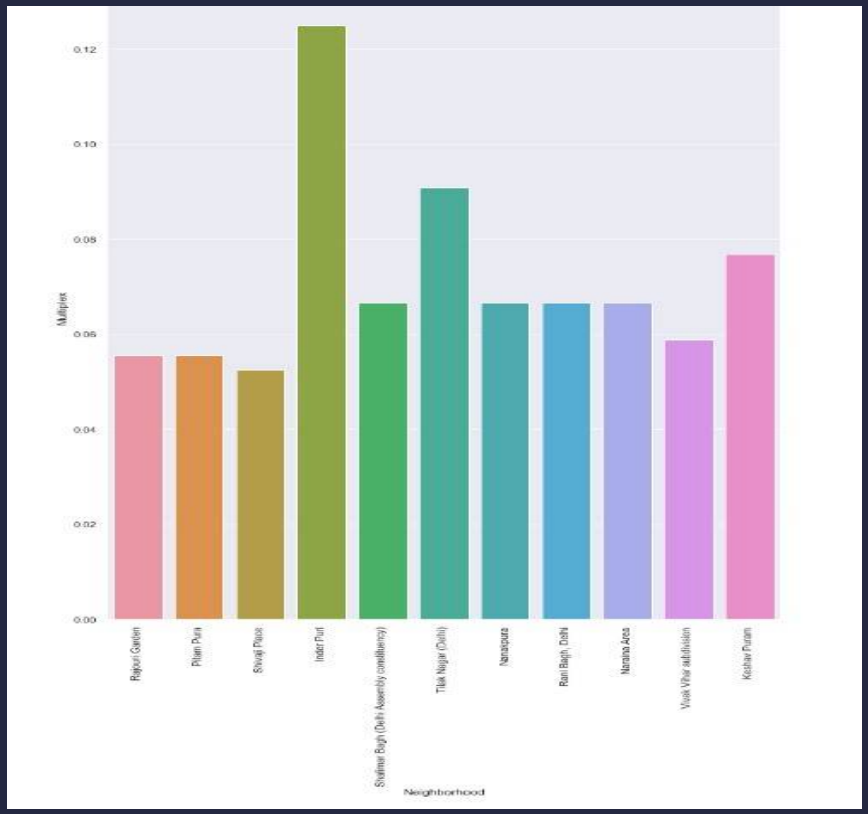
cluster 0 in red color, cluster 1 in purple color and cluster 2 in green color



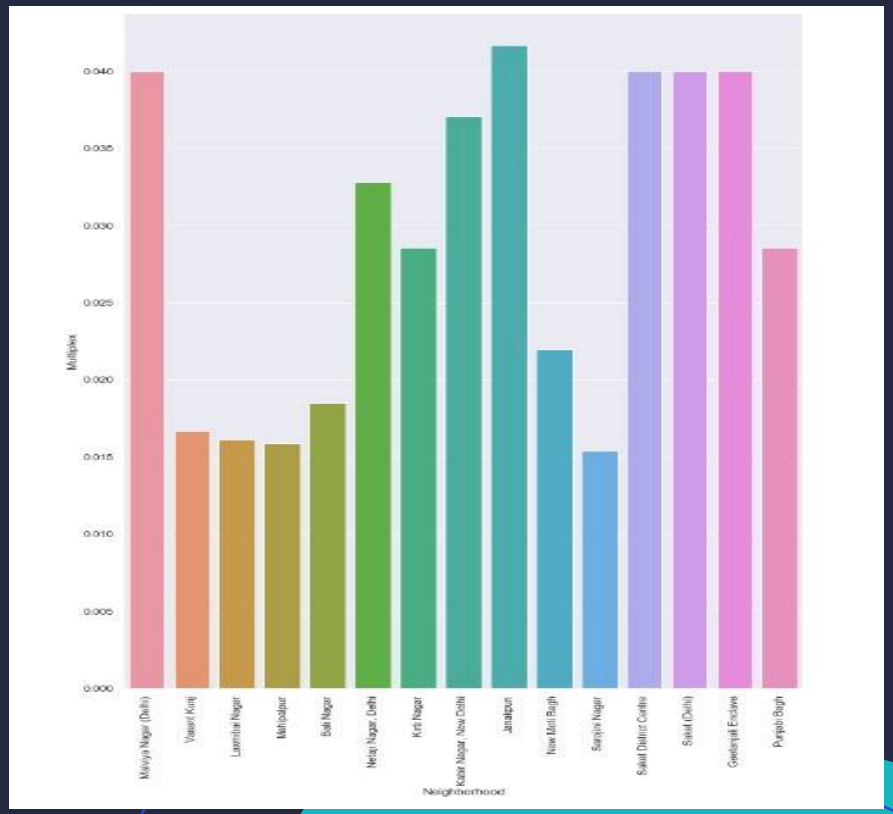
Result of the Analysis (2/2)

The result of statistical analysis shows that almost all neighborhoods in Cluster 1 have similar and high concentration of multiplexes. Further, neighborhoods in Cluster 2 have moderate and varied concentration of multiplexes.

Bar chart for cluster 1 with high concentration of multiplexes



Bar chart for cluster 2 with moderate concentration of multiplexes



Discussion

- Based on clustering and statistical analysis, the result shows that Cluster 1 has limited scope of developing a new multiplex considering high concentration of multiplexes.
- Cluster 2 is identified and shortlisted for further analysis. With this identification, it is found that 15 neighborhoods in Cluster 2 provide opportunity for opening multiplexes.
- As an illustrative case, one neighborhood 'Punjabi Bagh' is explored to arrive at the final decision of opening a multiplex.
- Further analysis of 'Punjabi Bagh' shows that it has restaurants, café, pizza place, hotels, bars, metro stations etc. This states that the location has lot of recreational activities.
- With moderate concentration of multiplex and high level of recreational activities in 'Punjabi Bagh' the location is ideal for developing profitable multiplexes.

Conclusion

- The Project recommends investors and property developers to use the findings made in the report to open new multiplexes across 15 neighborhoods in Cluster 2 that have limited competition.

- Further, investors and property developers can avoid Cluster 1 with high concentration of multiplexes which will result in unnecessary competition and low profit margins.

- The Project has been developed as a powerful data model with increase in accuracy as more and more data is loaded to the model.


thank
you