Analysis on Venue Ratings

By Sujata Chavan

Introduction

- Foursquare is a platform where users can visit venues and provide feedback to give other users insight on quality.
- Using this information combined with location information, we can analyse venues in Queens, New York to identify areas of interest.
- This data can be useful for various use cases such as:
  - People looking for a good location to live
  - Tourists looking for areas to visit
  - People wanting to find a good location to open a business

Data collection and preparation

- To carry out this investigation, the following data was collected from various sources:
  - Neighborhood information for Queens, New York, provided by Coursera
  - Venue locations provided by Foursquare
  - Ratings for each venue also provided by Foursquare
- From the collected data, only venues that contained a rating were used. Also, due to the limitation of Foursquare sandbox account, very few data points were able to be collected.
- The scope of the project was reduced to only include areas where data could be collected.
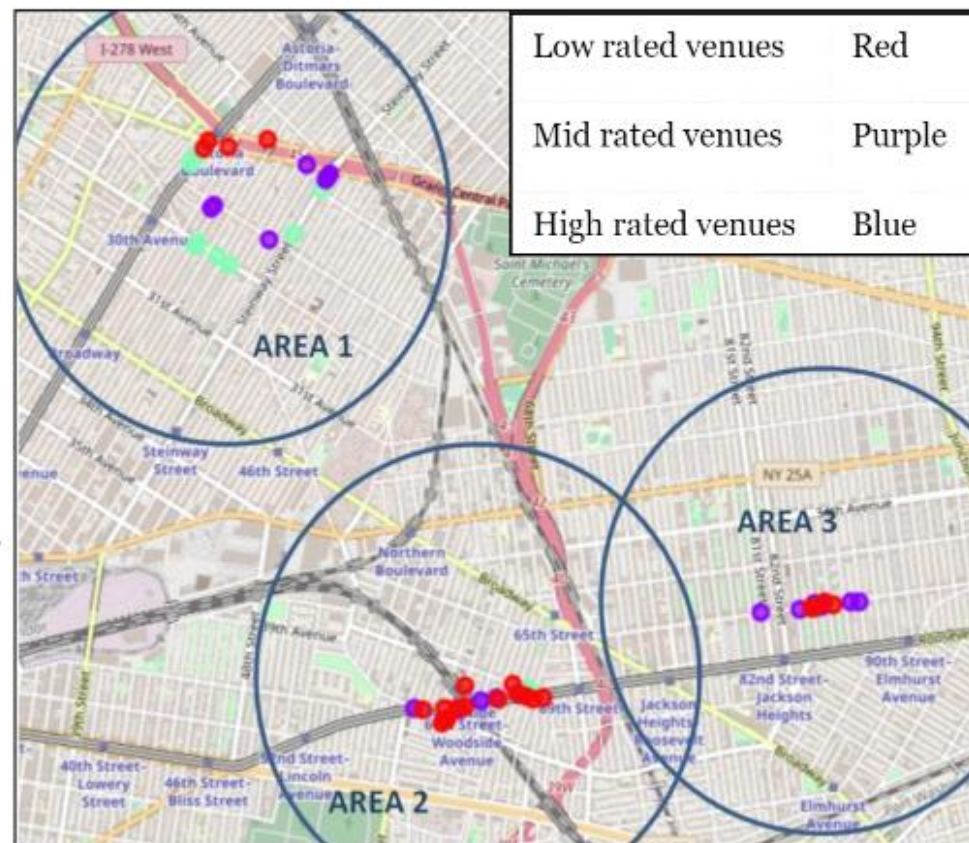
Clustering of ratings

- The data was classified into three clusters based on their ratings
- The mean rating of each cluster is shown below (with a maximum available rating of 10)

Cluster label	Description	Mean
Cluster 0	Low rated venues	7.80
Cluster 1	Mid rated venues	8.45
Cluster 2	High rated venues	9.00

Map representation

- We can now group the data into geographical areas and profile each one based on the ratings received.

- Area 1: the most high rated venues, with some mid and low rated venues to the North.

- Area 2: mainly low rated venues.

- Area 3: a mixture of low and mid rated venues.

Low rated venues	Red
Mid rated venues	Purple
High rated venues	Blue

Conclusion

- With the limited data obtained from Foursquare, venues found were clustered into three distinct areas (indicated on the previous slide)
- Clustering this data into groups of low, medium and high quality venues based on user ratings, the mean of each cluster was:
  - Cluster 0: 7.80
  - Cluster 1: 8.45
  - Cluster 2: 9.00
- Area 1 contained the highest proportion of high quality venues, while still containing a mixture of low to mid-range venues in the northern side.
- Area 2 contained the highest density of venues, most of them being low range venues.
- Area 3 contained a mixture of low and mid-range venues, with no high range venues.
- The order of preference for locations to live based on venue quality would be area 1, area 3, then area 2.