Battle Of Neighbourhoods

By- Prateek Khanna

Recommending venues to people travelling to Bangalore

- This presentation demonstrates a project about analysis of venues in Bangalore (India) using Machine Learning Algorithms and data science methods. This requires the extraction, load, transformation and analysis of all data fetched using Four Square API.

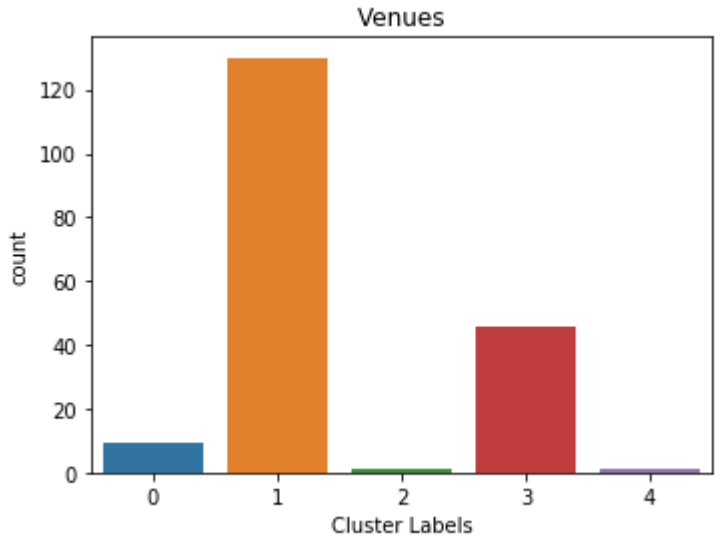
- We will use our data science powers to generate a few most promising neighbourhoods based on these criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by Travelers.

Data gathering and cleaning

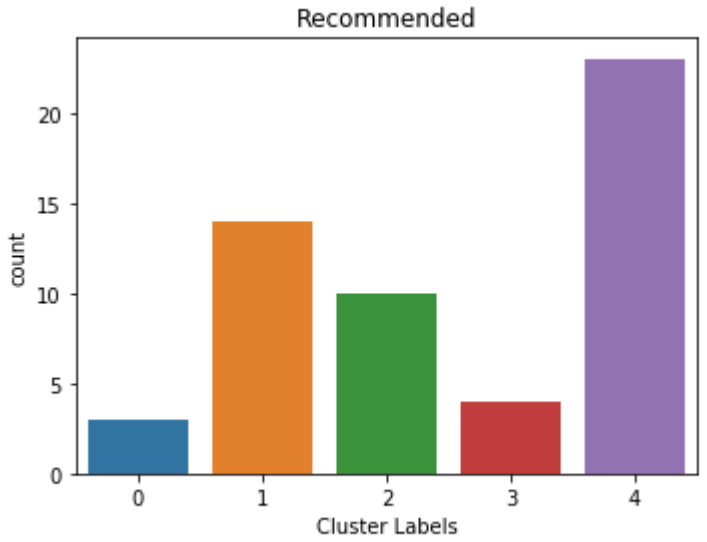
- Data has been gathered using Search for Venues and Get Venue Recommendations Four square API's.
- **Both the API's return data in JSON format. Fields like Venue Name, Address, Latitude, Longitude and Subcategory were fetch for the return data. Later category function is used to add category column to the data frame.**
- **The size of the Data Frames are as follows:**

1. **Recommended(54 Rows, 7 Columns).**
2. Venues(187 Rows, 7 Columns)

Clusters form using K-Means

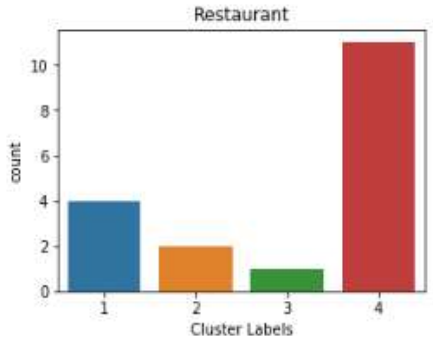


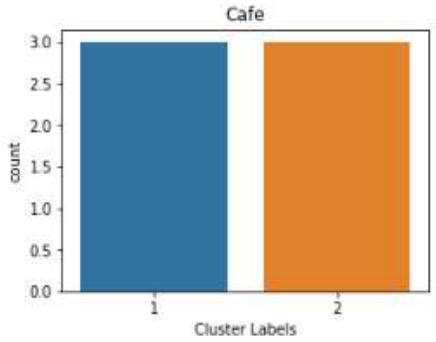
Cluster from using Venues data

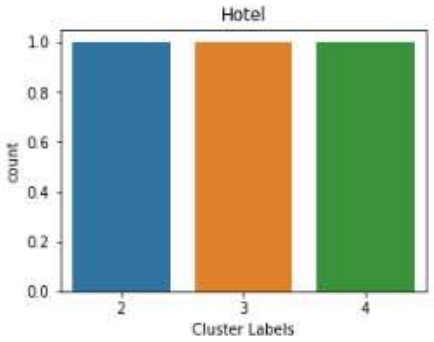


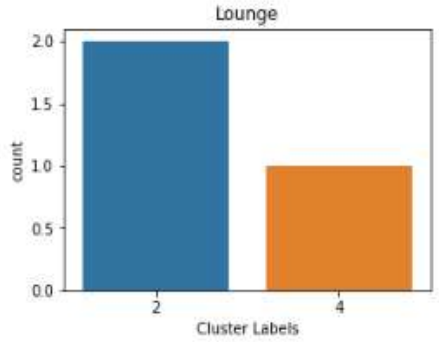
Cluster from using Recommended data

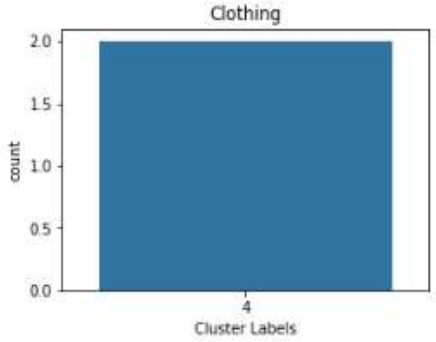
Categorical distribution(of Recommended data frame)



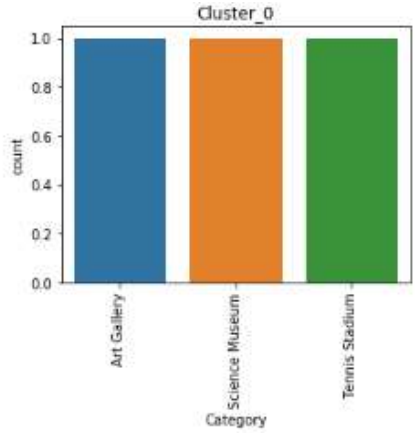


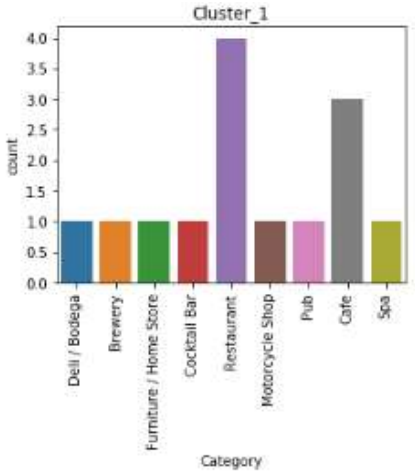


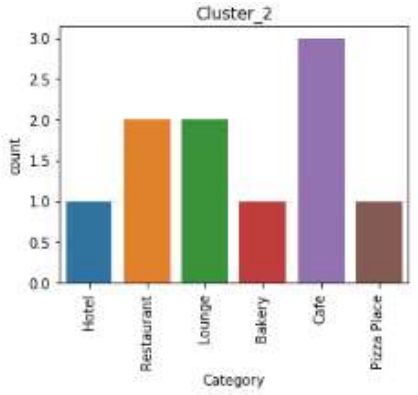


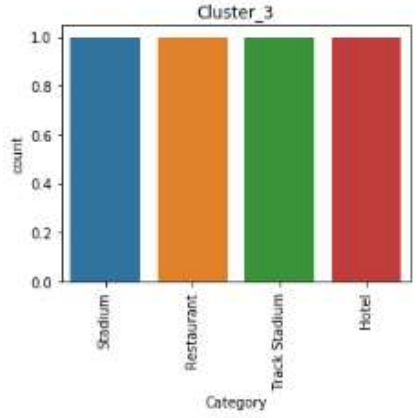


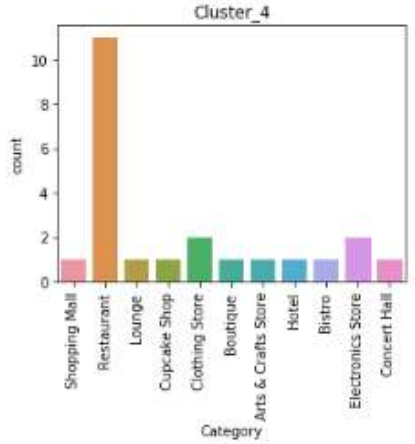
Categories distribution amongst clusters(of Recommended data frame)



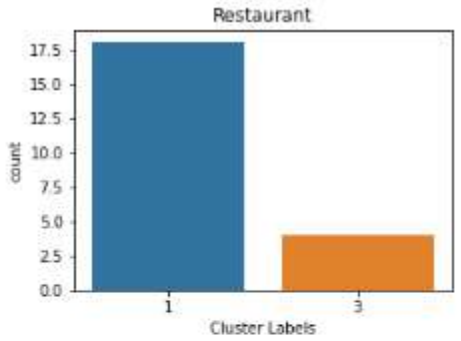


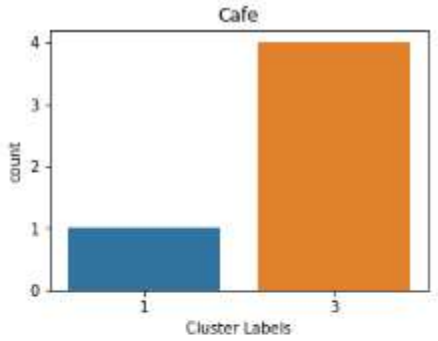


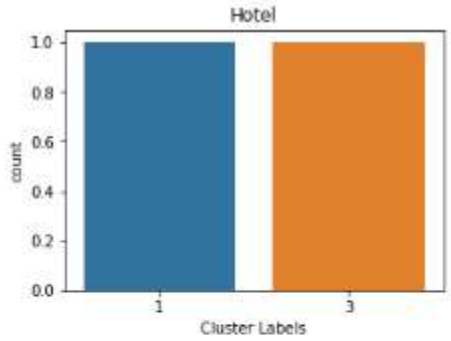


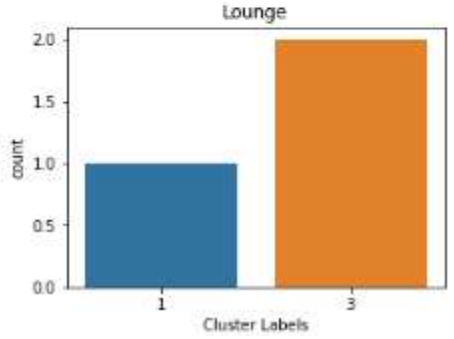


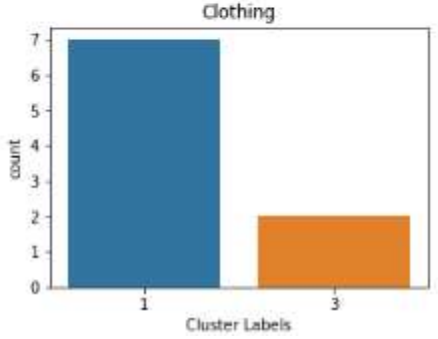
Categorical distribution(of Venues data frame)



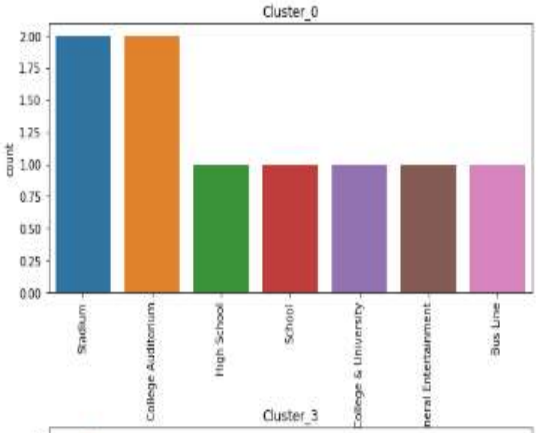


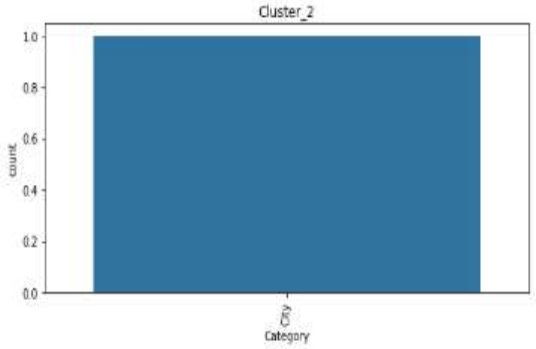


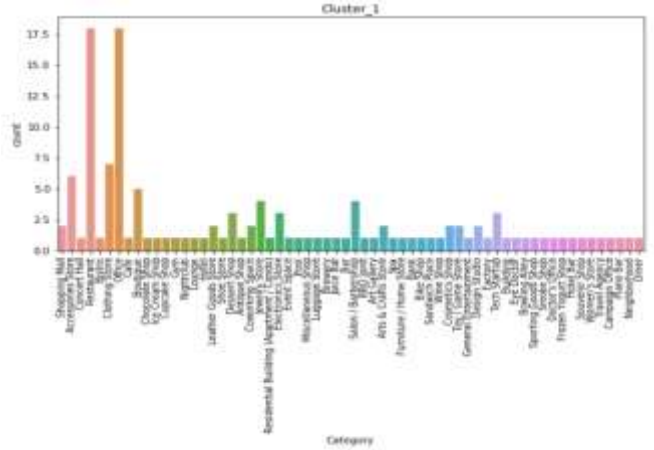


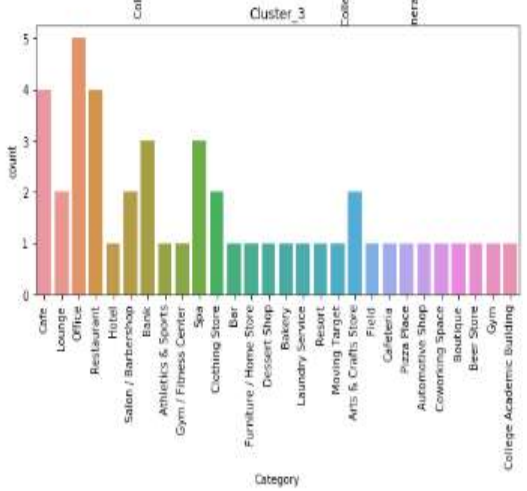


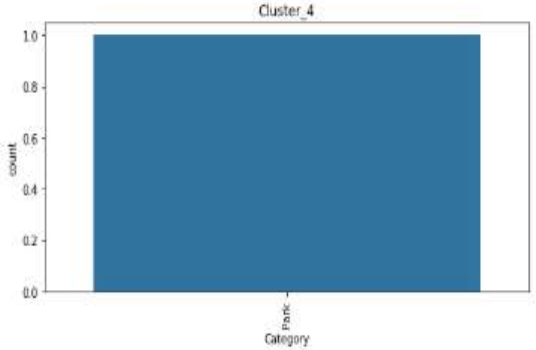
Categories distribution amongst clusters(of Venues data frame)











Results

- Through the analysis done in the notebook we can that **recommended_Cluster_4** contains the popular location to visit when on a strict schedule whereas **Cluster_3** lists the best places to be visited when one has ample amount of time to explore Bangalore.
- Due the difference in the number of places in **recommended** and **venues** data cluster number differ for same location in certain cases.

Conclusion

- By fetching data from Foursquare API we have first identified Recommended Venues that were rated by previous visitors. Clustering of those locations was then performed in order to create major venues of interest (containing greater number of different category locations) and table of those venues were created to be used as reference for final exploration by Travelers.
- Although, final decision on optimal visiting location will be made by visitors based on specific characteristics of neighborhoods and locations in every cluster, taking into consideration additional factors like attractiveness of each location levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.