Battle Of Neighbourhoods

T.Rukmini Devi

Table of contents:

1.Introduction

Bangalore is called as The Silicon Valley of India because of the large number of information technology companies located in the city which contributed 33% of India's ₹1,442 billion (US$21 billion) IT exports. In recent decades, the city has witnessed rapid growth in population and urbanised area.

While Central Bangalore is the commercial heart of the city, Eastern and South-Eastern Bangalore areas are major hubs for IT and financial companies. Southern and Western parts of the city are mainly residential areas. The neighbourhoods in the Northern and North-Eastern regions are both industrial and residential.

2.Project Description

This project demonstrates an analysis of venues in Bangalore (India) using heterogeneous data sources and data science methods. This requires the extraction, load, transformation and analysis of all data sources contained in the following notebook. Since Bangalore is cosmopolitan there are lots of venues in Bangalore, so we will try to segment venues into **Clusters (based on geographical location)**. We are also particularly interested in helping people who will be visiting Bangalore and want to explore the city. Furthermore, we would also prefer popular locations as there will also be people on a **short-trip.**

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.

Target Audience

- In this project we will try to help individuals travelling to Bangalore by recommending venues segmented in different clusters. Specifically, this report will be targeted to traveller interested in exploring **different Venue** in **Bangalore**, India.

3.Data

1. **Data Sources:**
   Foursquare - a location technology platform that provides various API's under the parent 'Places API' to help you fetch data. There are numerous API's for venues, users, photos, check-ins', list. In this project I have used two of such API's listed below:
   1) Search for Venues
   2) Get Venue Recommendations

2. Data Cleaning:

Both the API's return data in JSON format. Fields like VenueName, Address, Latitude, Longitude and Subcategory. Later category function is used to add category column to the data frame. Two data frames are made using the above API's to form recommended and venues respectively. Also, none of the columns have any missing values in any of the two data frames.

Recommended have 54 rows and 7 columns whereas venues have 187 rows and 7 columns after all the cleaning has been done.

Analysis

1. Exploratory Data Analysis:

Initially the data has been segmented into five cluster i.e. 0,1,2,3 and 4. After clustering the data seaborn has been used to visualize the distribution across the clusters.

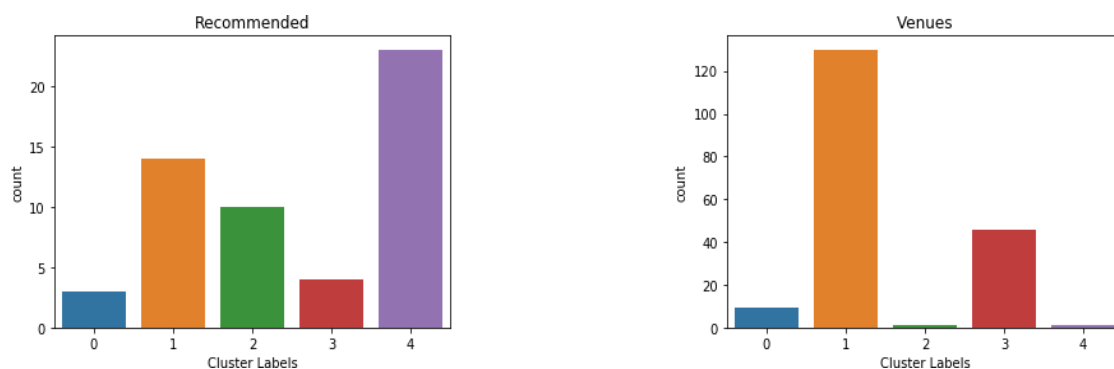## Number of venues in different clusters

Fig 1 Recommended Venues distribution across categories.

From above plots we concluded that cluster 2 and 1 have more venues in recommended and venues respectively.

Plots of Recommended data segmented on behalf of their categories.

Categories are as mentioned below¶
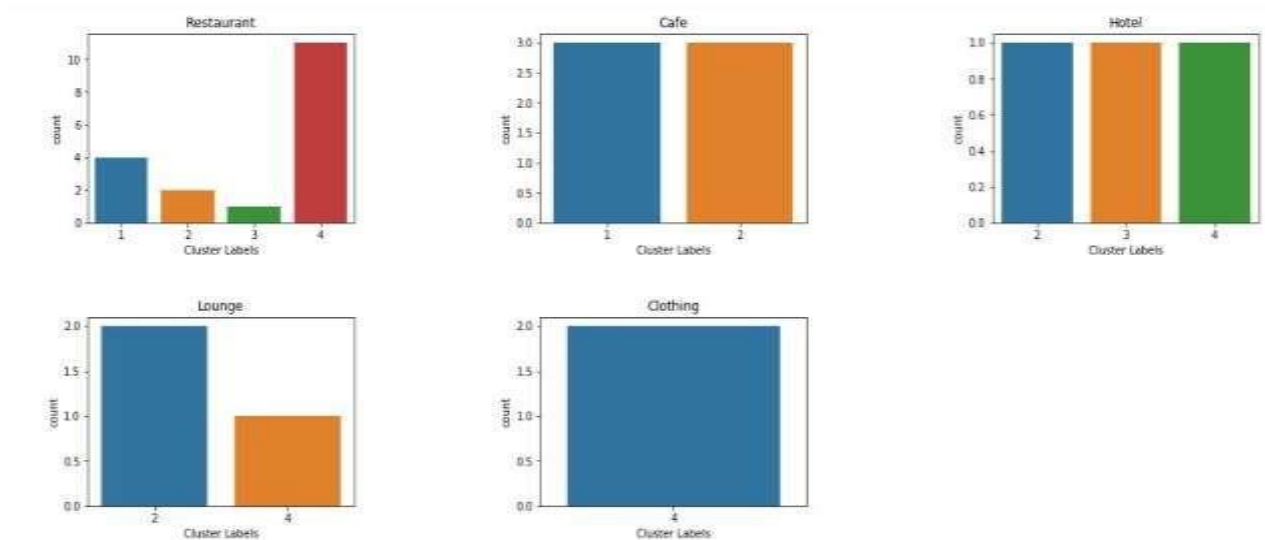- Restaurant
- Clothing Stores
- Cafes
- Lounge
- Hotel

Fig 2 Recommended_venues distribution across categories.

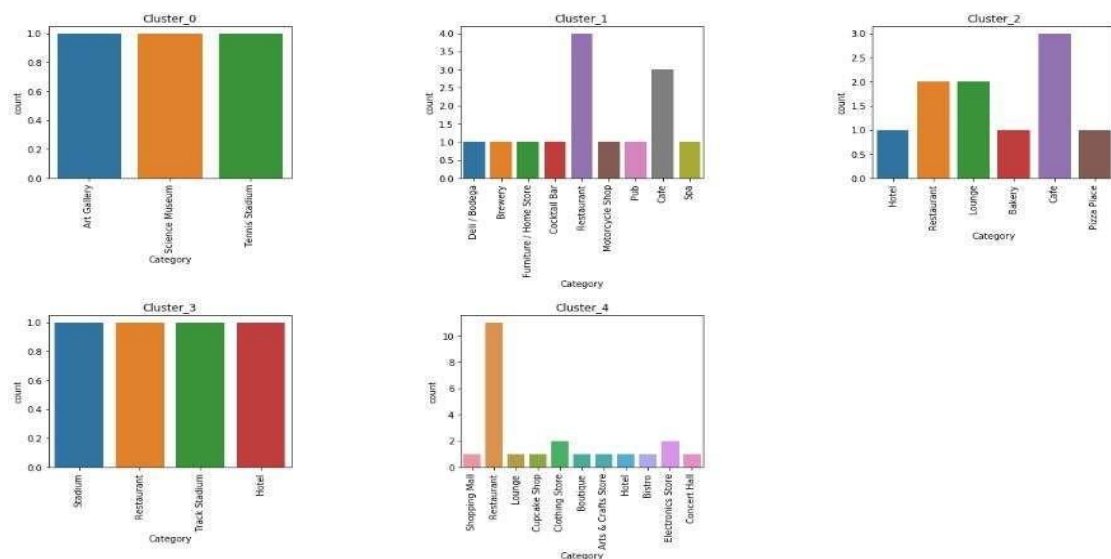# Recommended Data distributed on the basis of the clusters.

Fig 3 Recommended data distribution across Clusters

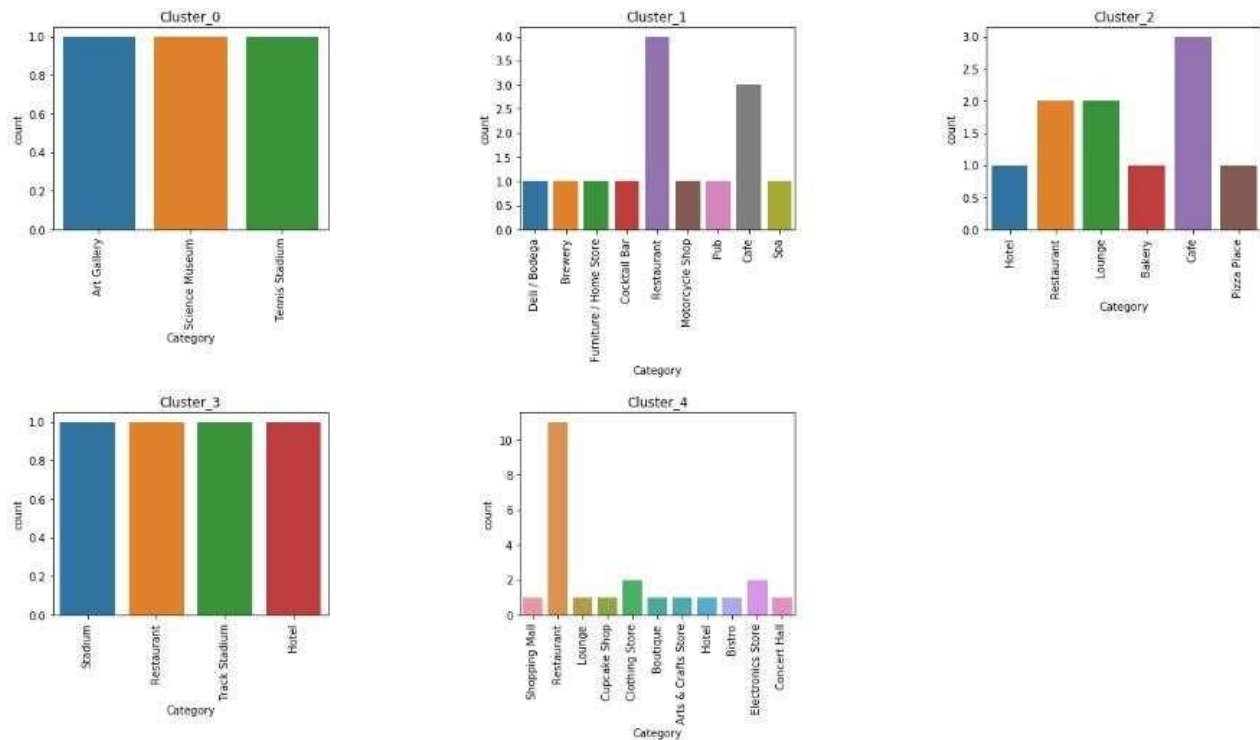# Venues Data distributed on the basis of the clusters.

Fig 4 Venues data distribution across Clusters

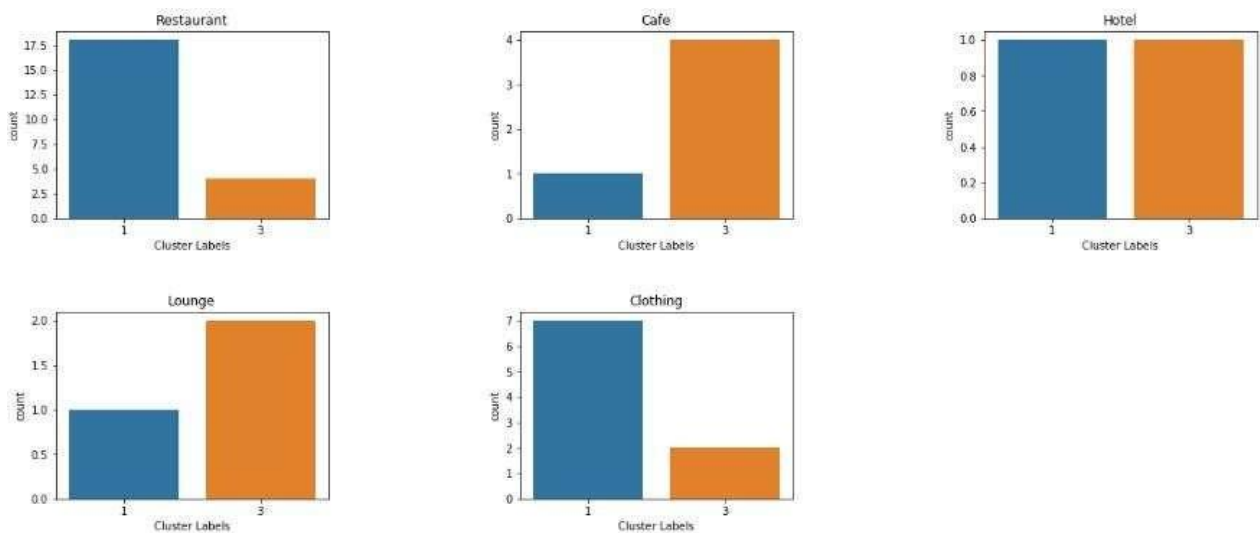# Venues Data distributed on the basis of the categories.

Fig 5 Venues data distribution across categories

Results and Discussion:

- 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:

Purpose of this project was to identify Bangalore venues in order to aid travellers in narrowing down the search for location to be visited. 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 Travellers.

Final decision on optimal visiting location will be made by visitors based on specific characteristics of neighbourhoods 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 neighbourhood etc.