Food court distribution index of Mysore and outskirts

Introduction

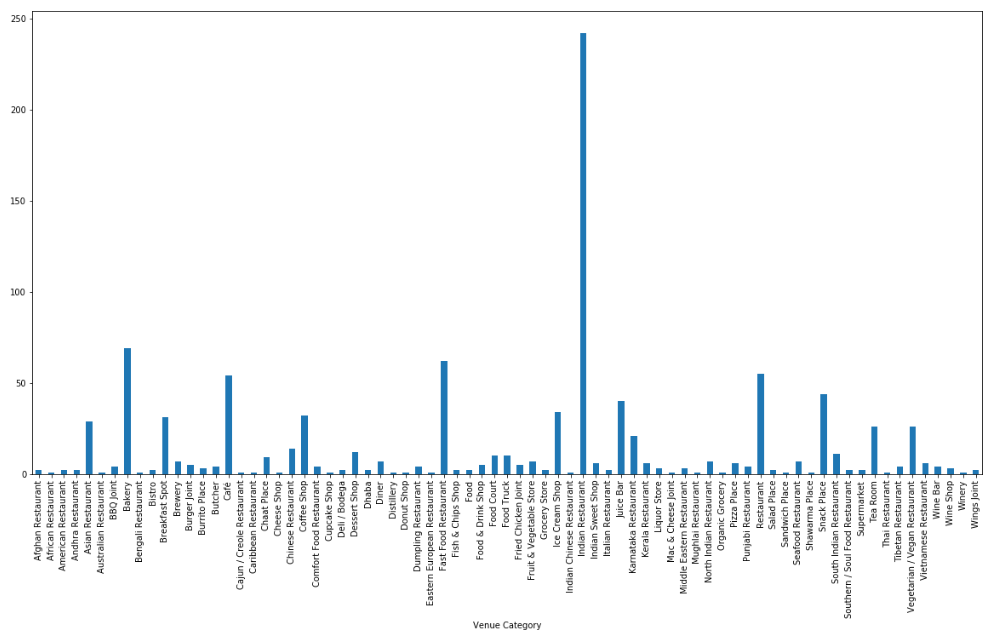
The study mainly attributes towards exploring the neighborhood of Mysore with addressing various types of forum, public outlet, school, play areas, café shop, food court, hospitals, bus depot, metro station etc., which helps people to find out their area of interest within and outskirts of the city. Once the dataset is formed basing on a category of venue, for further clarity they are distributed in various clusters depending on their similarity against a particular type. This clusters are later plotted on a map to have a better insight about the distribution of their desired venues across the neighborhood.

The same code can be used for any variable apart from food for its detail synopsis in the same way as presented below.

Data acquisition and cleaning

The neighborhood dataset has to be classified by its postal code for its further synthesis along with its latitude and longitude values so that the intersection of this dataset can be done with data collected from FourSquare API on the basis of available latitude and longitude coordinates.

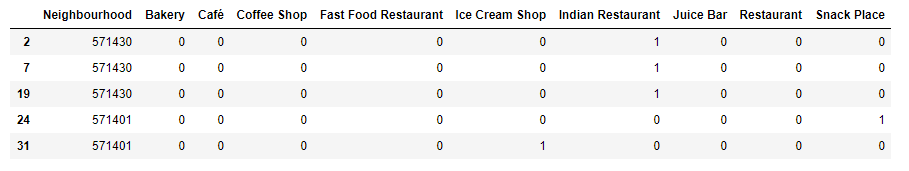
Post obtaining the intersected dataset, a filtering of the data has been done by selecting the category as “food”. The filtered dataset is then plotted on a bar graph to further squeeze it by selecting only top 9 food categories to obtain the basic dataframe for this study.



Venue distribution across neighborhood

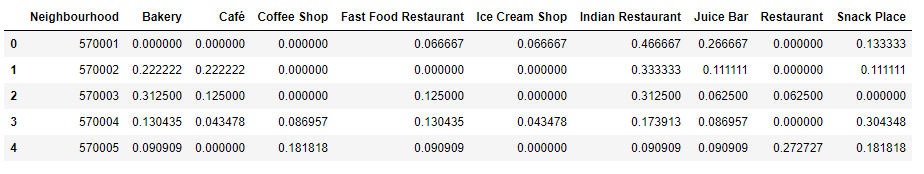
Data Analysis

To analyze the data sequentially, all other columns excepting “Neighbourhood” and “Venue category” are dropped and a fresh dataframe is prepared. Now using dummies all the postal codes are displayed row-wise with their corresponding dummies (0/1) against the listed top 9 categories column-wise. Dummy “1” represents the presence and dummy “0” represents the absence of a category at a given postal code.



Neighborhood dataframe with dummies

Once the dummies are populated over the new dataframe, it is found that several instances of the same postal code get replicated over various category lines which stands difficult for assessment and hence the local mean of each category is calculated against all the postal codes. This final dataframe shows a weight distribution matrix of each category against a particular postal code which is further segregated to various clusters depending on their similarities.



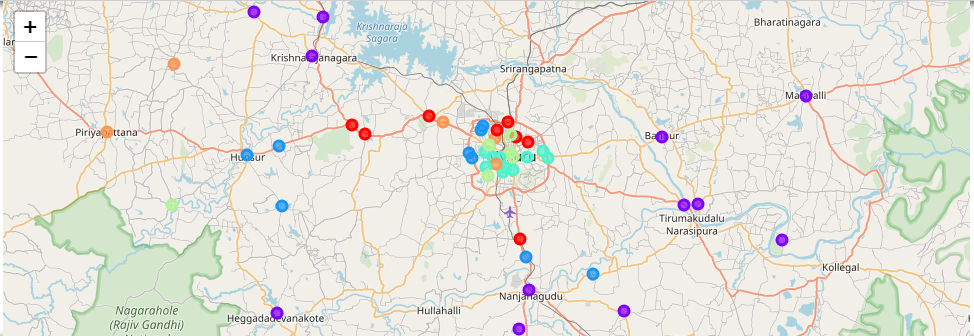
Featureset for neighborhood

Clustering of data

Focusing at the type of above weight distribution matrix, k-means clustering is chosen for its further analysis. To have better analogy of the dataset and its variation with changing cluster levels, the number of clusters are varied between 2 to 8 and finally a value of 6 is chosen for this study. To represent the data distribution against a particular cluster, labels are marked on that particular set. With this the matrix is representing various categories of food outlets at different postal codes against their respective designated labels. At last the discrete distribution of each label is collapsed to a single row by computing the mean of all rows occupied by each of these respective labels. Therefore, the final matrix comes out with possible weight factor distributions of a particular cluster label against the selected top 9 categories.

Discussions

A folium map is constructed basing on the above final dataset which represent the cluster labels with different colors at various neighborhoods of Mysore.



Neighborhood similarity

Looking at the data distribution post clustering, it can be observed that the cluster 0 (ie. red color markers) represents “Indian restaurants” with a higher weight factor and “bakery”, “cafe” with comparatively lower weights inside and outskirt of Mysore city. Whereas, cluster 1 (ie. Purple color) is mainly representing the presence of Indian restaurants at the outskirts of Mysore city. Likewise, other clusters (ie. 2 to 6) gives a detail visual distribution about presence of selected top 9 food categories across various locations of Mysore city and outskirts.