

Restaurant Location Recommendation Using Neighborhood Clustering

Introduction: Business Problem

In this project will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an Indian Cuisine restaurant in Delhi, India. Finding a suitable location for restaurants in major cities like delhi proves to be a daunting task. Various factors such as over-saturation or no demand ,for the type of restaurant that the customer wants to open, effect the success or failure of the restaurant. Hence, customers can bolster their decisions using the descriptive and predictive capabilities of data science.

We need to find locations(Neighborhood) that have a **potentially unfulfilled demand** for Indian Restaurant. Also, we need locations that have **low competition and are not already crowded**. We would also prefer location as close to popular city Neighborhood, assuming the first two conditions are met.

We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly Expressed so that the best possible final location can be chosen by stakeholders.

Data Acquisition and preparation

Based on definition of our problem, factors that will influence our decision are:

- number of existing restaurants in the neighborhood (any type of restaurant)
- number of and distance to Indian restaurants in the neighborhood, if any
- distance of neighborhood from popular neighborhoods

In our project we will:

- acquire the names and boroughs of the neighborhoods by scrapping a wikipedia page.
- After we have got the names of all the neighborhoods, we will geocode them using the library geopy.geocoder (Nominatim).
- Next, we use the foursquare API to find all types of restaurants within a 1000 meter radius for every neighborhood.

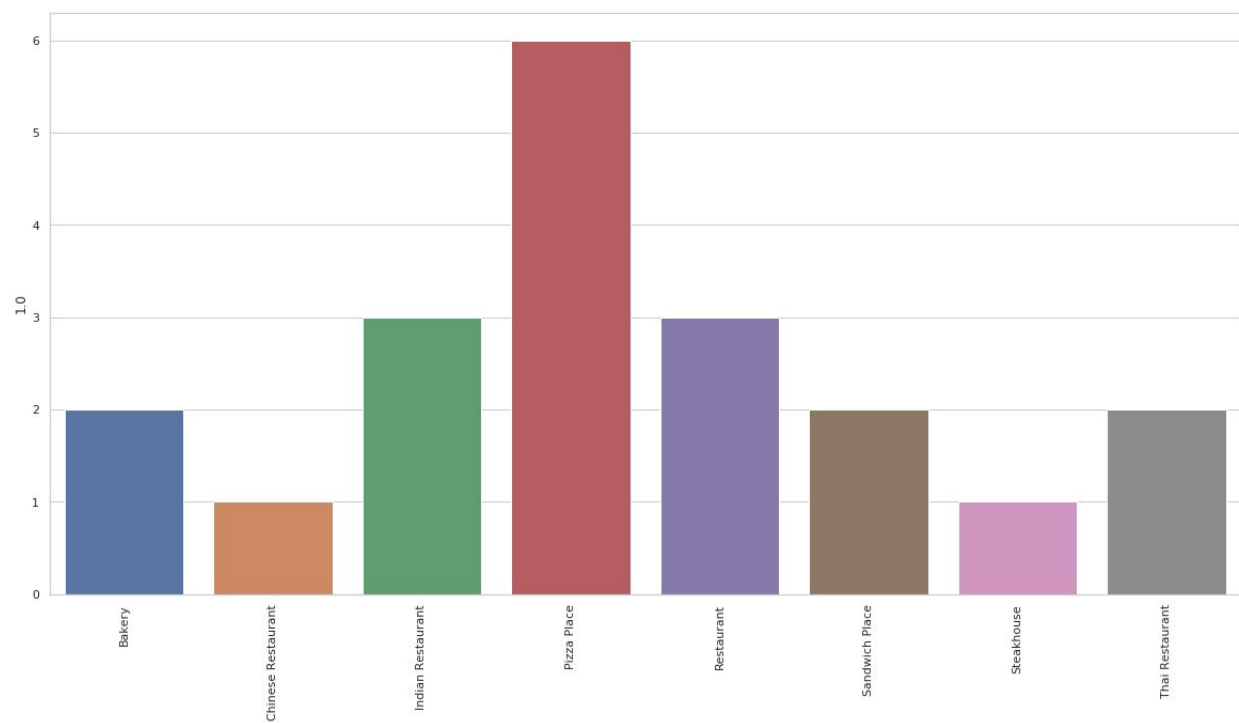
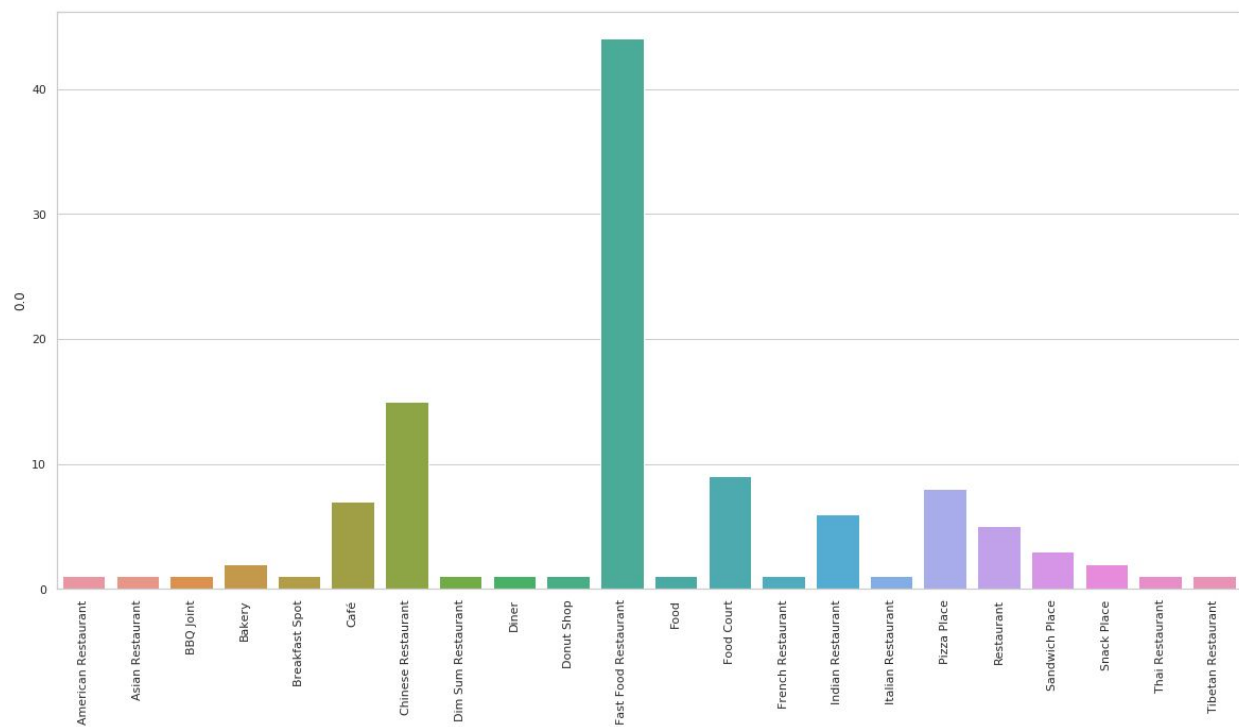
Clustering and Analysis

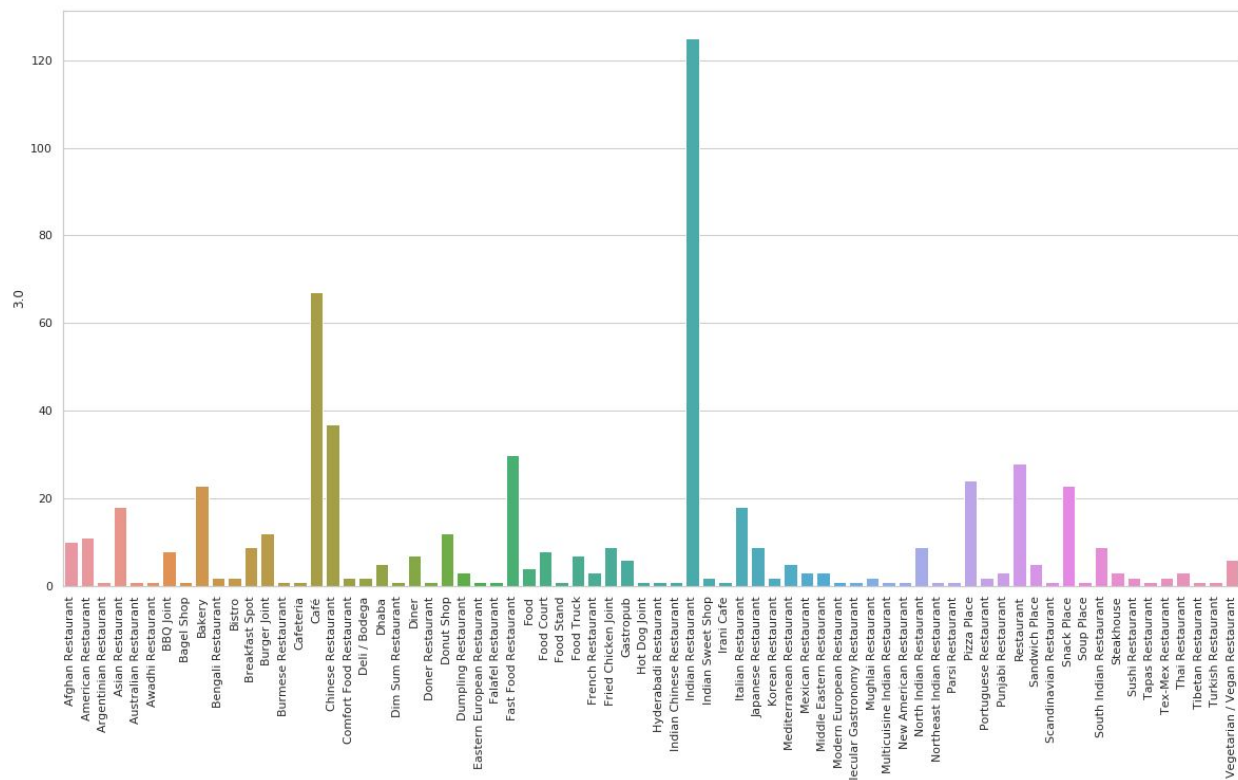
Our goal here is to find the neighborhoods with low density of Indian restaurants. But, how will we decide which neighborhoods, currently operating on minimal number of Indian restaurants, have the potential for growth and which neighborhoods do not.

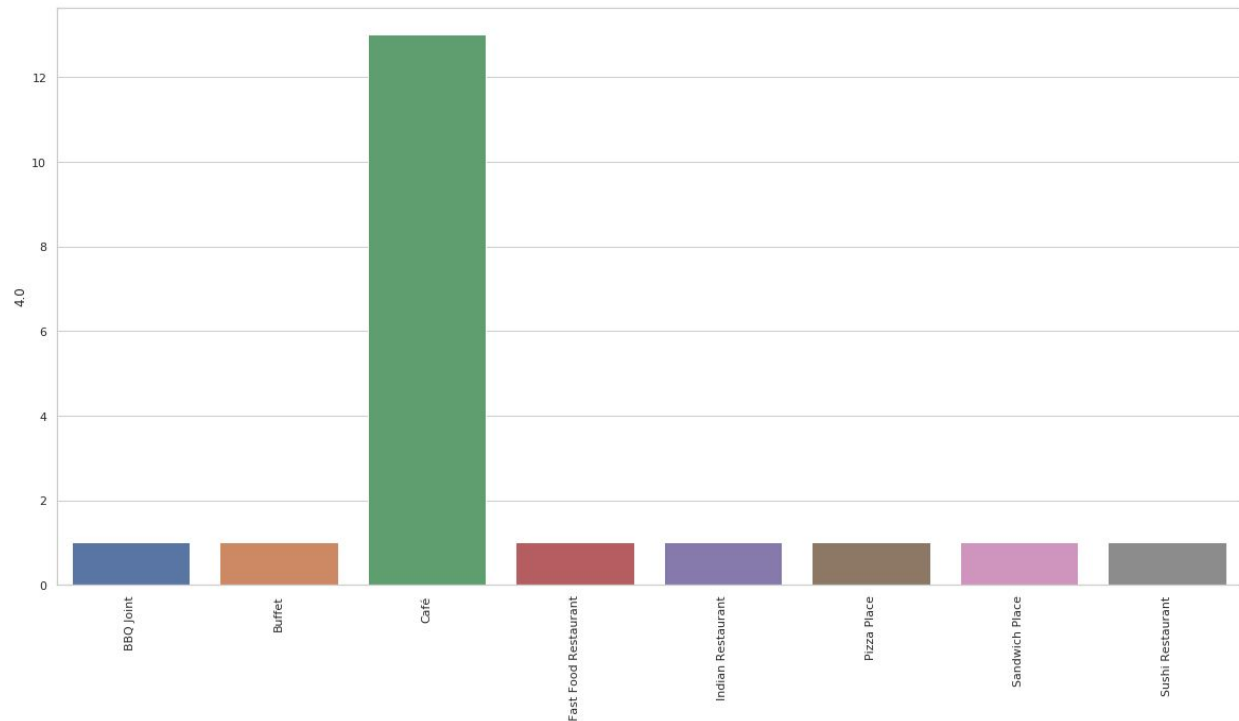
The most intuitive idea would be to find neighborhoods that have similar patterns of restaurant trends.

This can be achieved by clustering the neighborhoods on the basis of the restaurant data we have acquired. Clustering is a predominant algorithm of unsupervised Machine Learning. It is used to segregate data entries in clusters depending on the similarity of their attributes, calculated by using the simple formula of euclidean distance.

We can then analyze these clusters separately and use those clusters that show high trends of Indian Restaurants







Analysing the bar graphs we can clearly see that **clusters 1 and 2** have a high demand for Indian Restaurants

Recommendation

In this section:

- we will, first, analyze the density of the Indian Restaurants in generally for each neighborhood.
- Then we will weed out the neighborhoods that in the highest 70 percentile of density
- Find out the most popular neighborhoods
- Will then try to find remaining neighborhoods that are close to them
- Provide the a detailed recommendation of top 10 neighborhoods

Now, as clusters 1 and 2 have a maximum number of Indian Restaurants, we will focus our search on neighborhoods within these two clusters.

Why?

We know that when we were clustering the neighborhoods the data used contained the mean of all types of restaurants present in the particular neighborhood. Therefore, we can say that the neighborhoods are clustered on their restaurant trends.

Now, clusters 2 and 3 may collectively have the highest number of indian restaurant but there will be some neighborhoods in these clusters which would have a demand for Indian Restaurants, as these neighborhoods are in the same cluster, but would not have enough supply.