1. Introduction

a) Background

I hail from India and Indian food is my weakness. As an expat in US I was struggling to find a place to settle where I could find Indian community as well as have access to good Indian food and restaurants. There are apps which help you in making decisions on where to settle , however the experience is broken when it comes down to making decisions based on food habits. I was able to find a good place to live , but the choice of Indian food in the neighborhood was limited, and I had to move somewhere else within a year due to food issues. A good idea would be to figure out a way where locations are segregated based on good food joints , restaurants serving specific community food. That way people would have access to food meeting their palate. This exercise will make a basic assumption that good neighborhood for a community will have good restaurants serving dishes from that community. Idea is to start with a place and then figure out appropriate neighborhood

b) Problem Statement

This project predicts the most suitable places for Indians to live based on Indian community food choice

2. Data Acquisition and Cleaning

a) Data Source

Primary source of data is Foursquare dataset returned when a search service is called. Project starts by calling foursquare search api based on a query . Based on the data returned appropriate data cleaning process is run to filter out only relevant data for our analysis. 30 samples with 8 features were returned (For this project the place chosen was Edison , NJ)

b) Data Cleaning

Step1: Based on the search api response , filter out the venue node within the response node
Step2: Create a search list based on id , name , latitude , longitude and distance . From this list , filter out the category name and Create the required data frame
Step3: For each user tied to the category above , make a call to the user api and retrieve the number of likes for each venue
Step4: Add the number of likes to the data frame created in step 2

c) Feature selection

For this analysis , we will make use of the food joint name , latitude and longitude (to create the folium map) , distance , category name and likes . Number of likes will be used to cluster and segment neighborhood data

3. Exploratory data Analysis

a) Visualization of food joint density vs latitude and longitude

Based on the scatter plot , it seems that the density of food joints is higher in the 40.75,-73.9 location. Please do note that we are dealing with just 30 samples (we are limited with the amount of data returned by foursquare api)
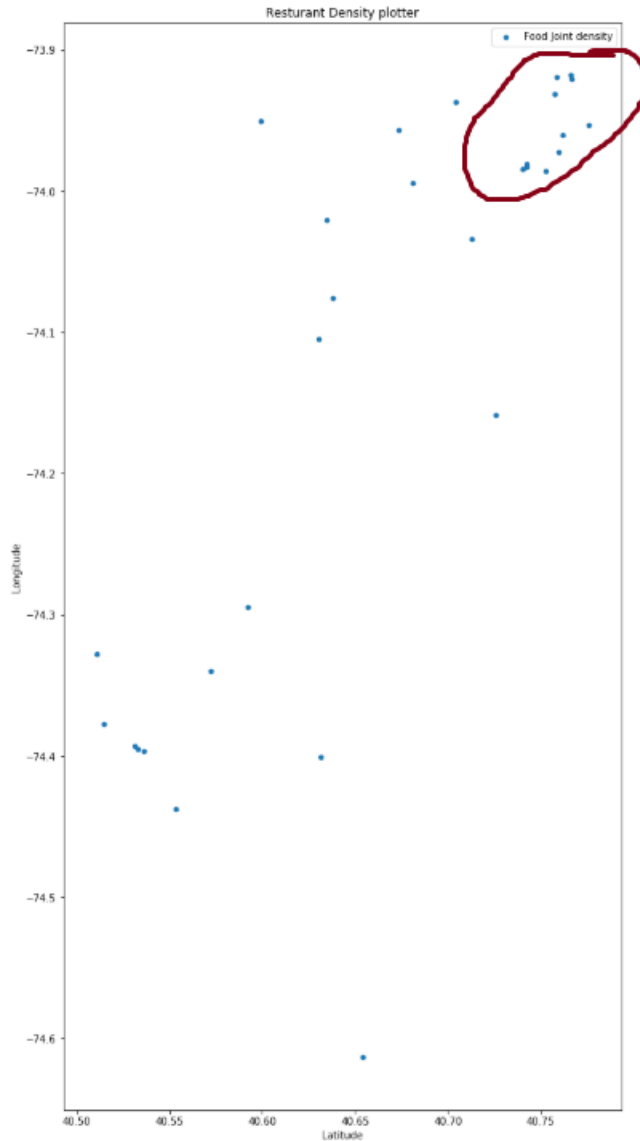
b) Visualization of restaurant categories

It seems that the search query returned some categories which are not tagged with Indian restaurants. Grocery store , hookah bar is also returned . We will keep all these categories (except the NA) as these are related in some way or the other to Indian cuisine
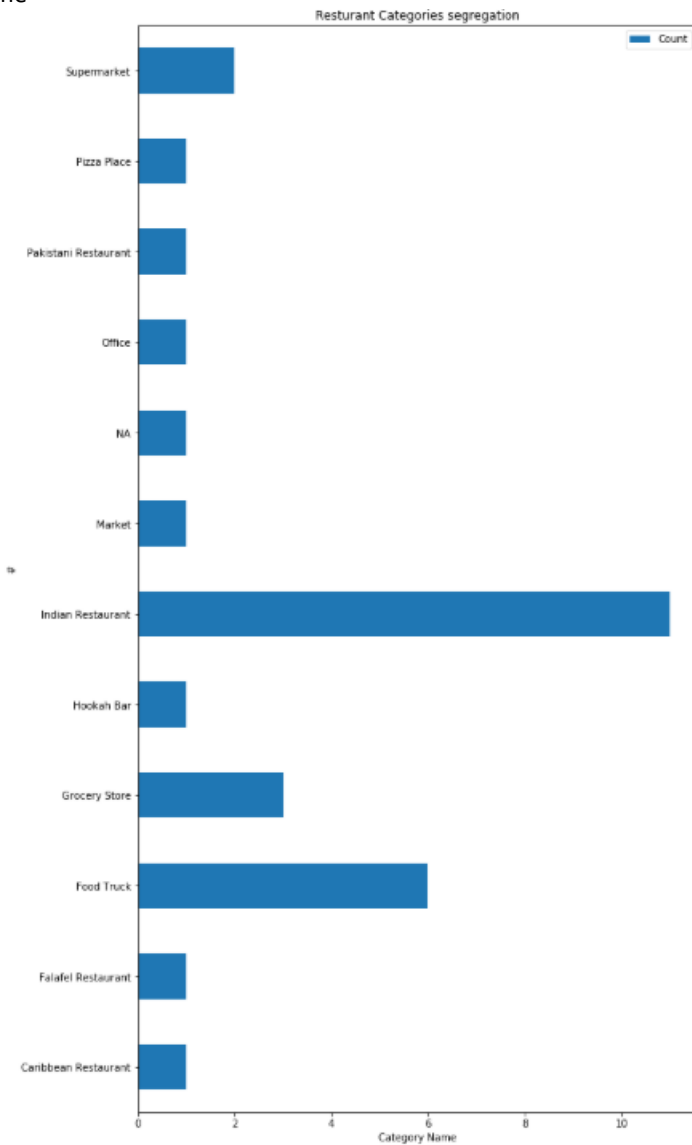
c) Visualization of restaurants popularity based on likes

It seems that OM Real Indian food is the most likes food joint in the dataset . Some of the joints do not have any likes (0) .
Data also suggests that there are little/no review for food sold in carts


Indian Resturants popularity
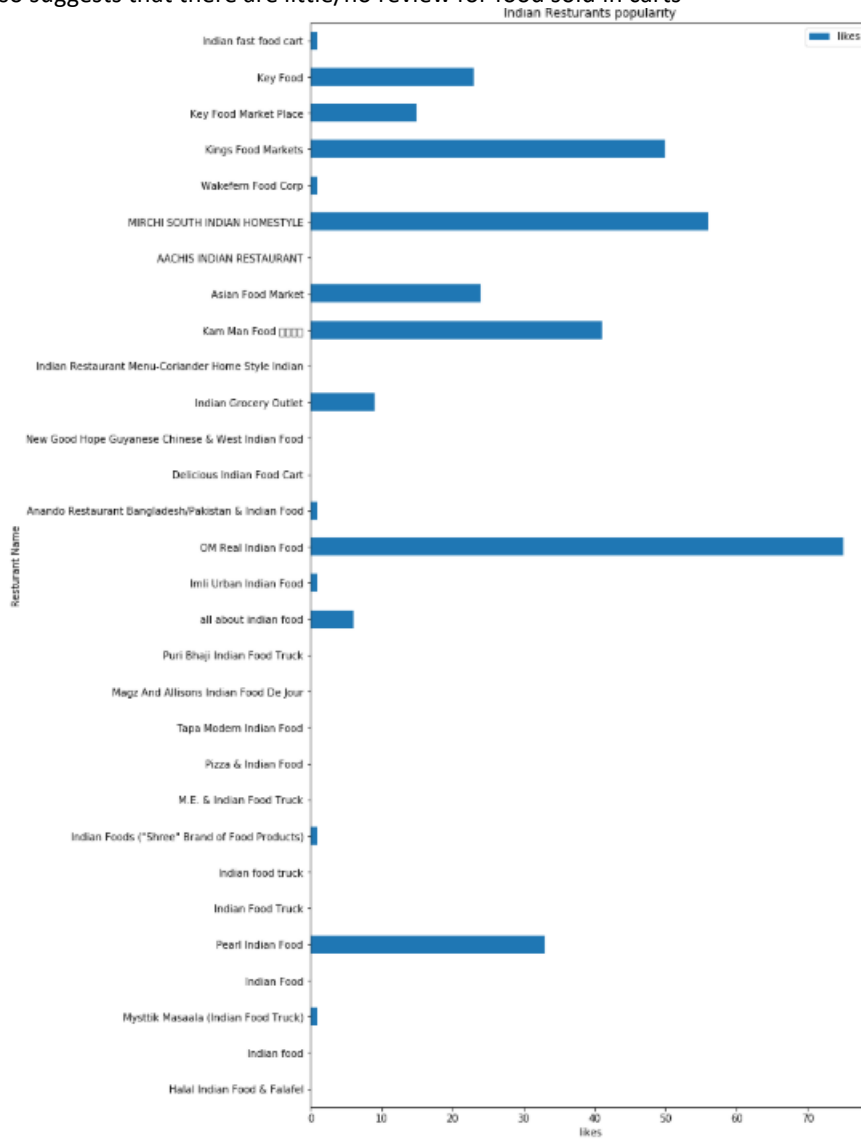
d) **Folium map of restaurants**

A quick map shows the density distribution of the Indian restaurants. Size of the circle is depicting the popularity of the joints. Sats suggests that area near oak tree road and king's county would be a good place for Indian foodies . There are quite a few numbers of places and some good food joints
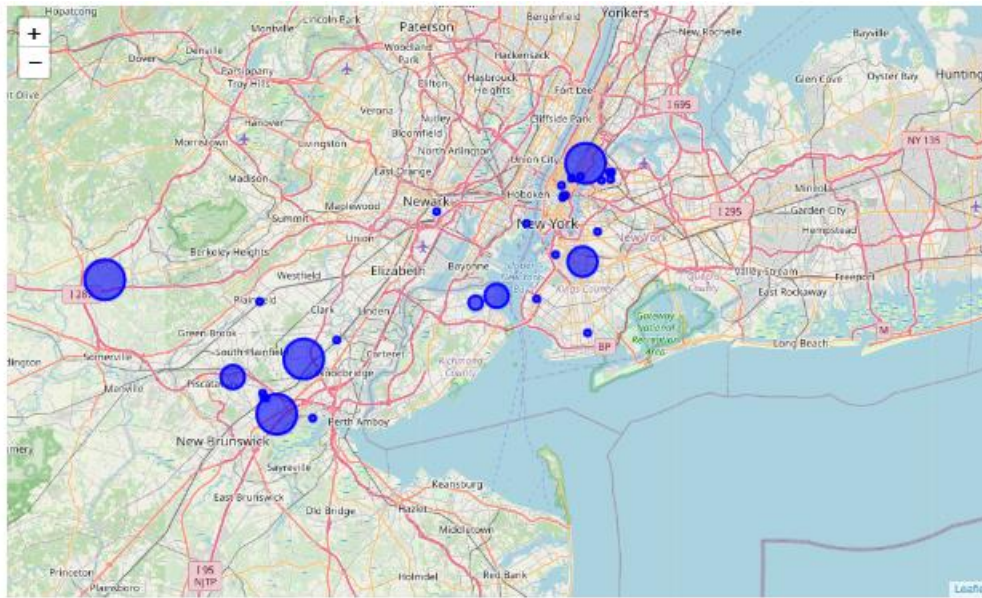
4. Clustering

5 clusters were created based on neighborhood data . It seems that area near king's county and oak tree road would be good places for Indian foodies to look for housing
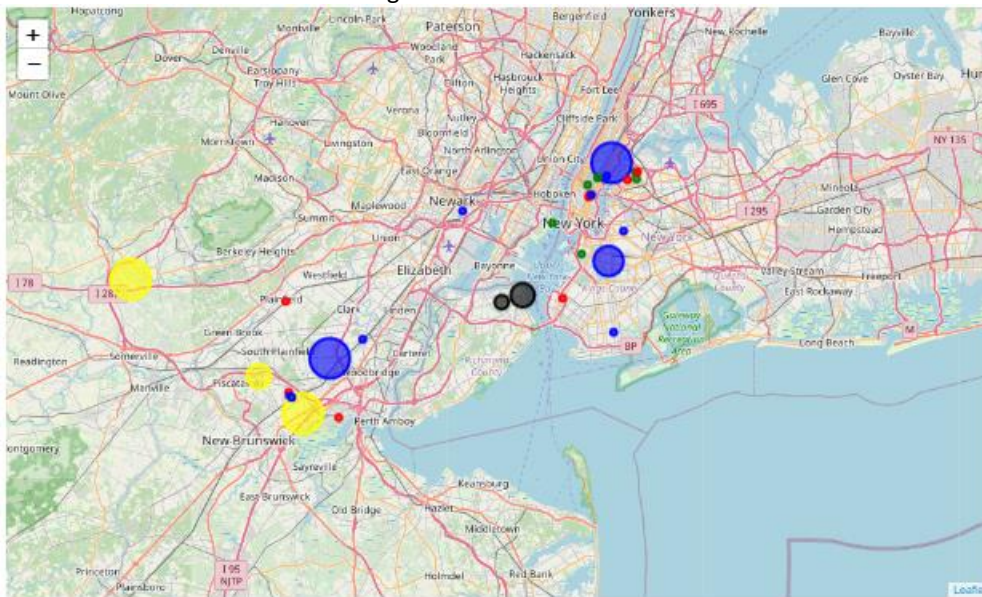
5. Conclusion

It seems that area near king's county and oak tree road would be good places for Indian foodies to look for housing