A

Summer Training

Report

On

**Submitted By :-**

**Submitted To :-**

**Kishan Kumar**

**IT Division of CDAC**

**NETPARAM TECHNOLOGIES PVT. LIMITED**

Training held at

**2019 - 2020**

**Neighborhood Recommender for opening a New Business**

**Alok Pandey**

**Krishankant Rathi**

**Gaurav Gupta**

**Tanuj Firoda**

## CONTENTS

|  |  |
| --- | --- |
|  |  |
| **Introduction To Project** | 3 |
| **Literature Review** | 3 |
| **Methodology** | 3 |
| **Problem Formulation and Solution** | 6 |
| **Result, Findings and Conclusion** | 7 |
|  |  |

## 1. INTRODUCTION

### 1.1 INTRODUCTION OF PROJECT

My project provides an interface for an Indian Restaurant Chain to help to decide where it should open its new Restaurant in the City of London, based upon the location of other Indian Restaurant and density of Indian Population living in various areas of London.

### 1.2 BUSINESS PROBLEM

* Our customer is ABC Restaurant, which is an International Indian Restaurant Brand and also a market leader.
* ABC Restaurant has recently planned to expand its business to the United Kingdom, and they want to start this journey from the heart of the UK itself.
* Given the extremely large size and the population of the city, our customer wants to identify the best neighbourhood area to open its first Indian Restaurant covering the majority of the population and facing least competition from other restaurants.
* The problem statement will be: **Which neighbourhood has most Indian population and has lesser number of INDIAN RESTAUARNT’s?**

## 2. LITERATURE REVIEW

* London is one among the cities in the world, where a large Indian Population resides.
* It is referred to as the Arch of the world because of its role as the once world’s most powerful kingdom.
* It has a population of over ten million, making it a megacity and most populous city in United Kingdom.
* Being a demographically diverse city, the needs of the residents are also increasing rapidly.

Hence, any new organization or an existing one should keep up with their pace in supplying the needs of the customers.

* [British Indians](https://en.wikipedia.org/wiki/British_Indians) form the largest ethno-national group in London with a population of around 542,857 or 6.6% of the population. The majority are concentrated in West London, home to London's Hindu community, though populations can be found throughout London.
* As at 2011, the Indian population of Greater London was 542,857 or 6.6% of the population
* In West London, close to [Heathrow Airport](https://en.wikipedia.org/wiki/Heathrow_Airport), resides one of the UK's largest Indian communities. According to the [2001 Census](https://en.wikipedia.org/wiki/United_Kingdom_Census_2001), 39% of the people within the [Ealing Southall](https://en.wikipedia.org/wiki/Ealing_Southall" \o "Ealing Southall) constituency, comprising Southall and nearby areas, are of [British Asian](https://en.wikipedia.org/wiki/British_Asian) origin.

### 3. METHODOLOGY

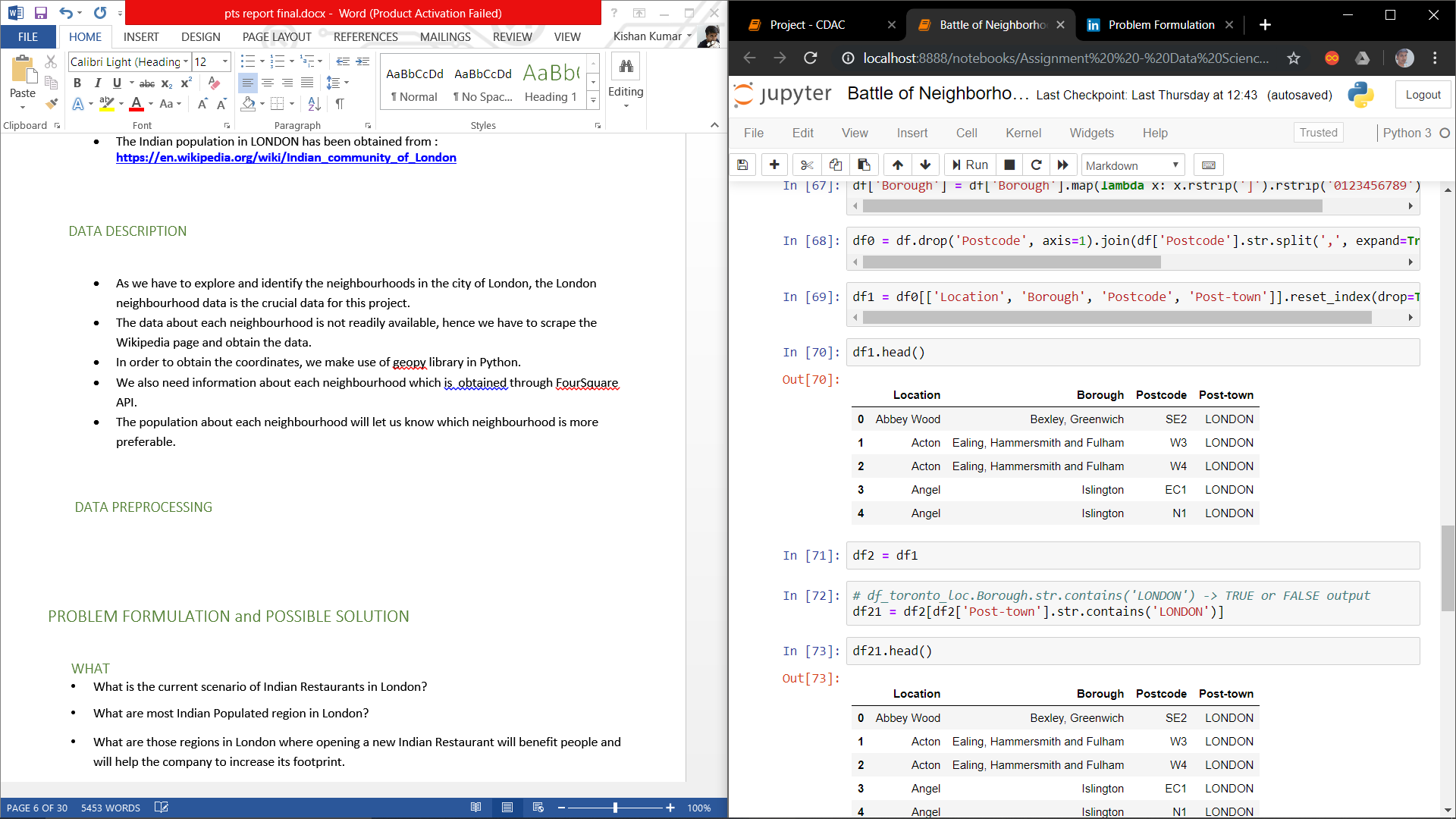
### 3.1 DATA DESCRIPTION

* The data to be used in this project is not readily available. Hence, the data has been obtained from various sources such as
* Foursquare, which is a local search-and-discovery mobile and web based app which provides search results for its users.
* Wikipedia, which has the details about the neighbourhoods in LONDON. [**https://en.wikipedia.org/wiki/List\_of\_areas\_of\_London**](https://en.wikipedia.org/wiki/List_of_areas_of_London)
* The geographic coordinates of each location have been obtained through Geopy([**https://geopy.readthedocs.io/en/stable/**](https://geopy.readthedocs.io/en/stable/))
* The Indian population in LONDON has been obtained from : [**https://en.wikipedia.org/wiki/Indian\_community\_of\_London**](https://en.wikipedia.org/wiki/Indian_community_of_London)

# 3.2 DATA DESCRIPTION

* As we have to explore and identify the neighbourhoods in the city of London, the London neighbourhood data is the crucial data for this project.
* The data about each neighbourhood is not readily available, hence we have to scrape the Wikipedia page and obtain the data.
* In order to obtain the coordinates, we make use of geopy library in Python.
* We also need information about each neighbourhood which is obtained through FourSquare API.
* The population about each neighbourhood will let us know which neighbourhood is more preferable.

### 3.3 DATA PREPROCESSING



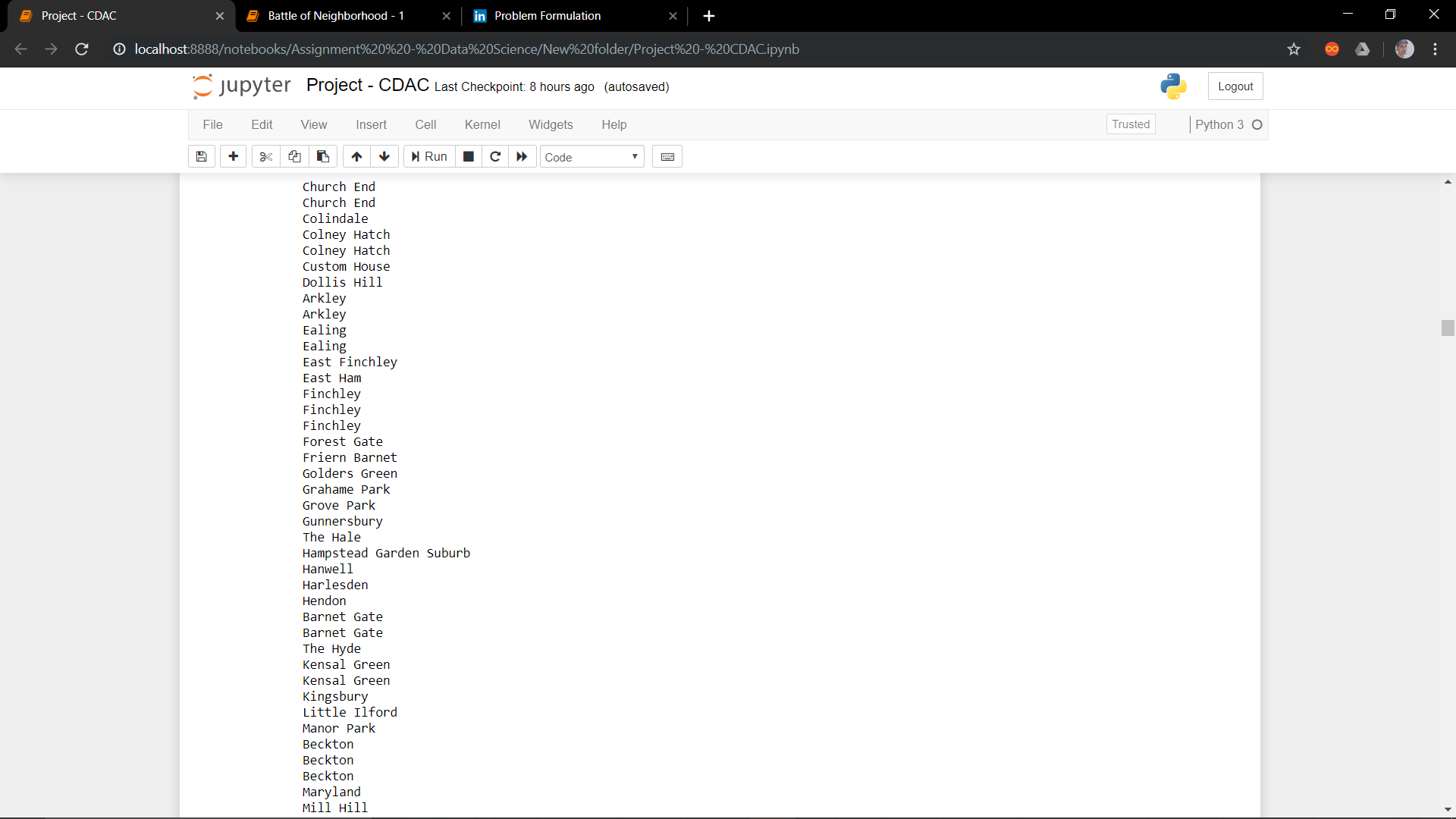
Our processed data from Wikipedia will look the data frame presented above.

# 3.4 EXPLORATORY DATA ANALYSIS

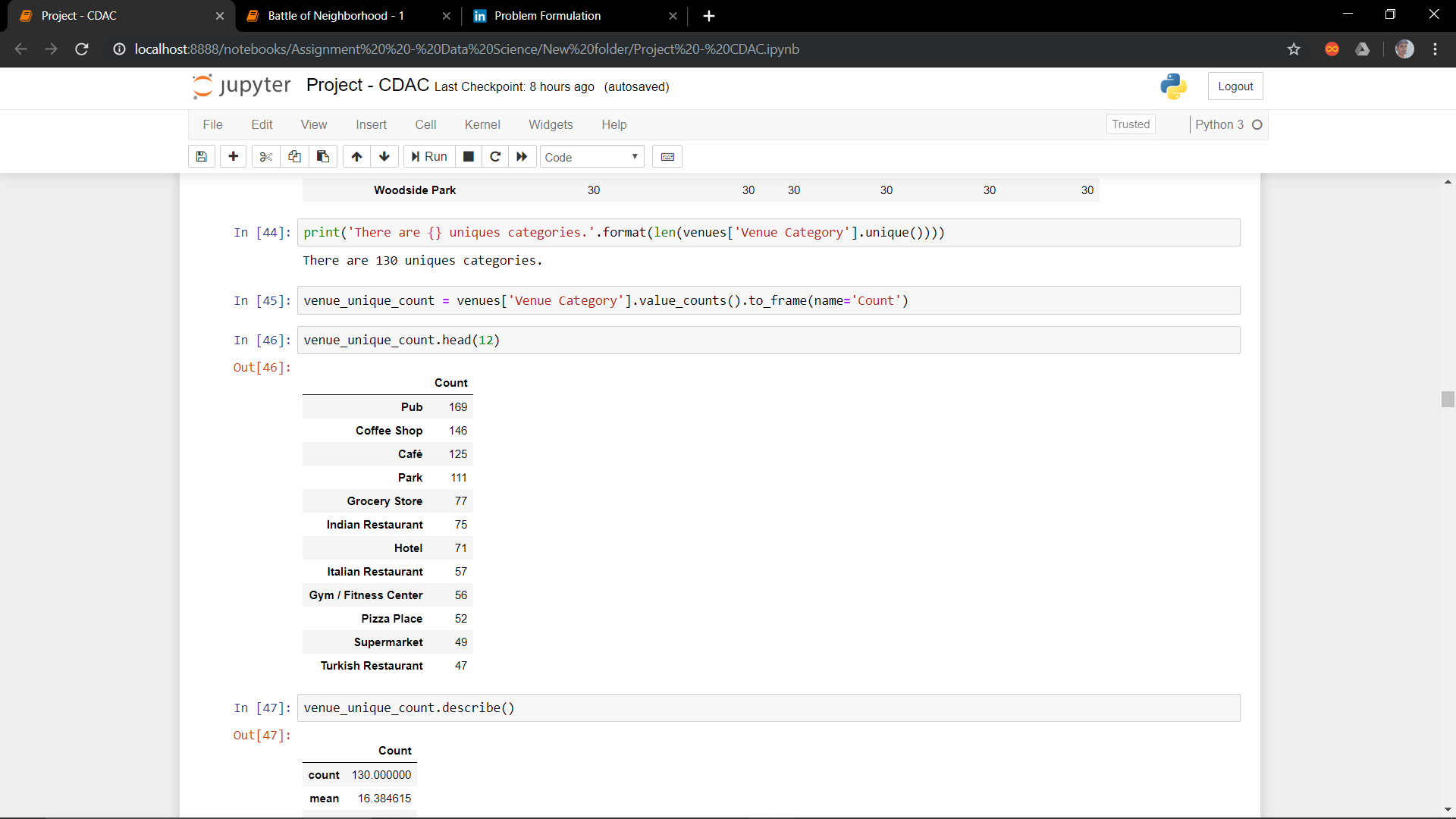
Indian population Percentage in different Borough.



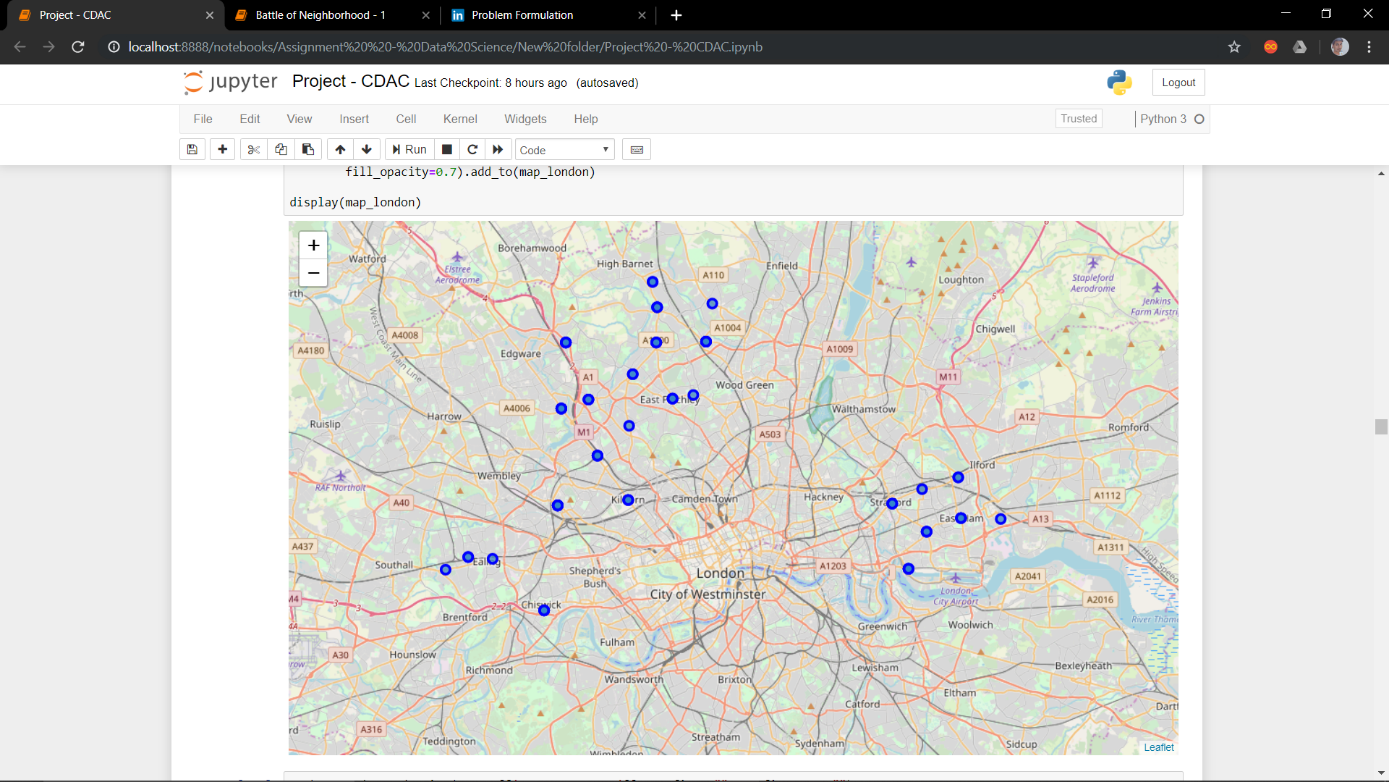
Various Neighborhoods in the Borough where Indian people live



Individual number of Venues of different Businesses in the above Neighborhoods

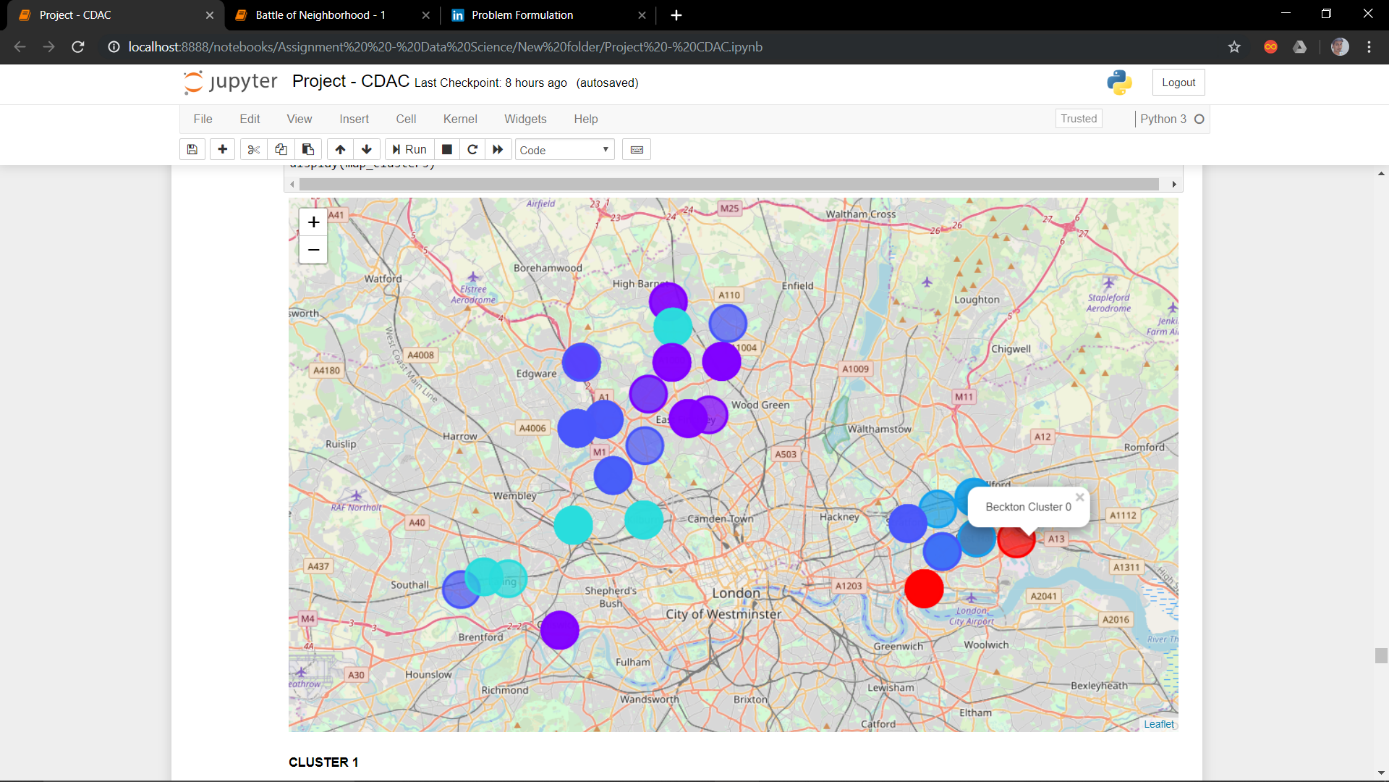


Location of different Neighborhoods in London using



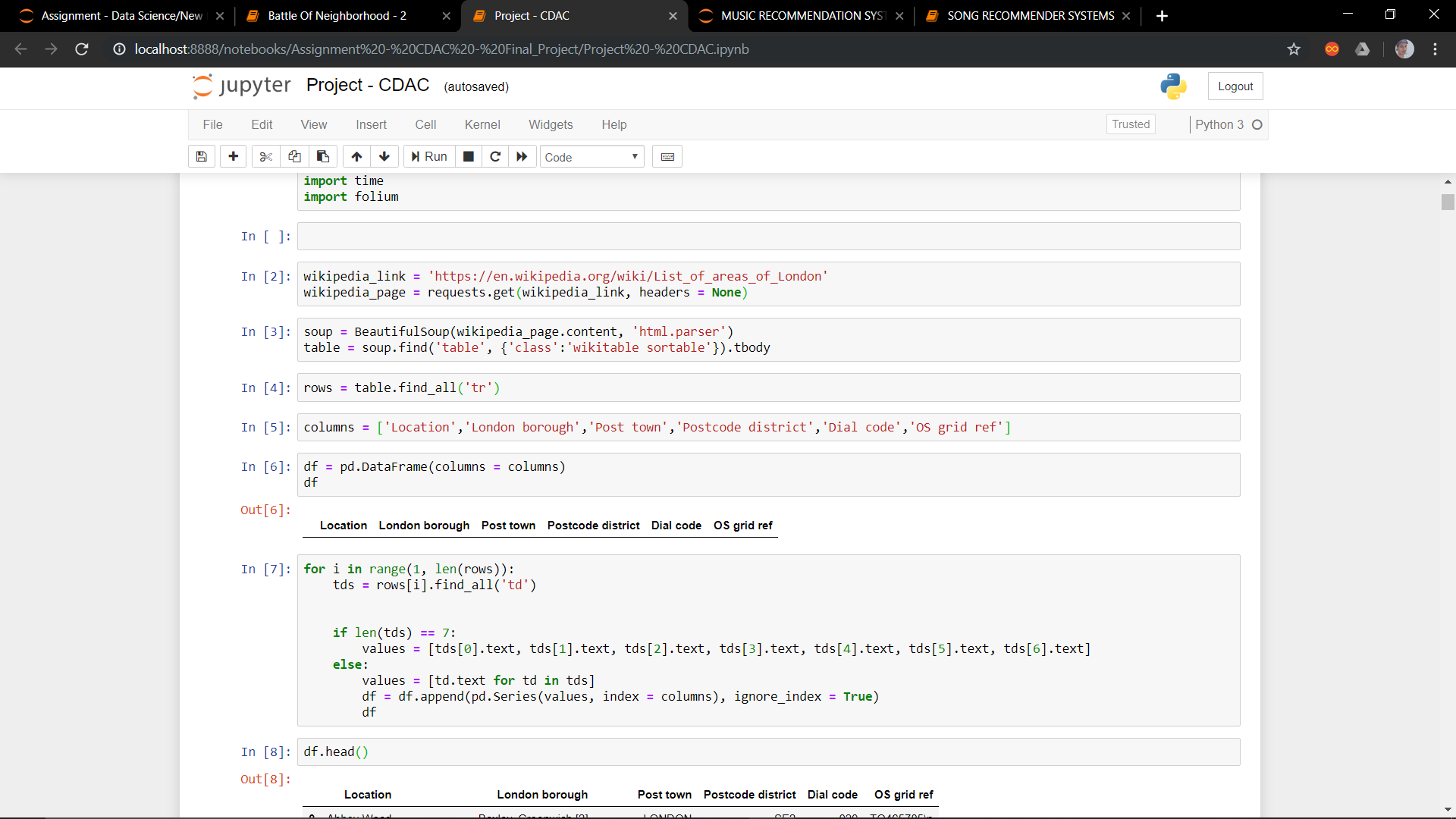
# 3.5 CLUSTERING

Our clustered model will look like

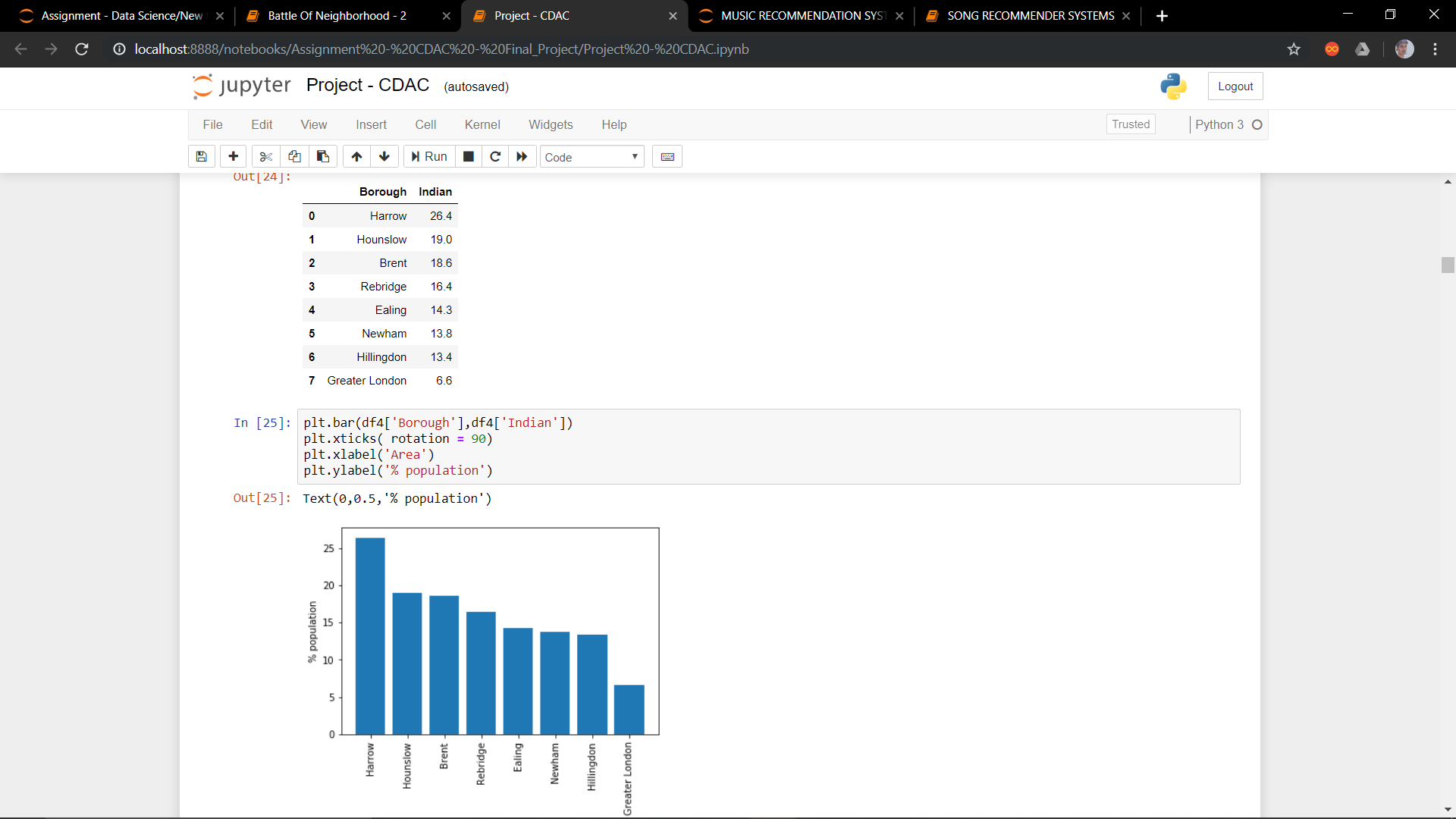


#### 3.6 CODE

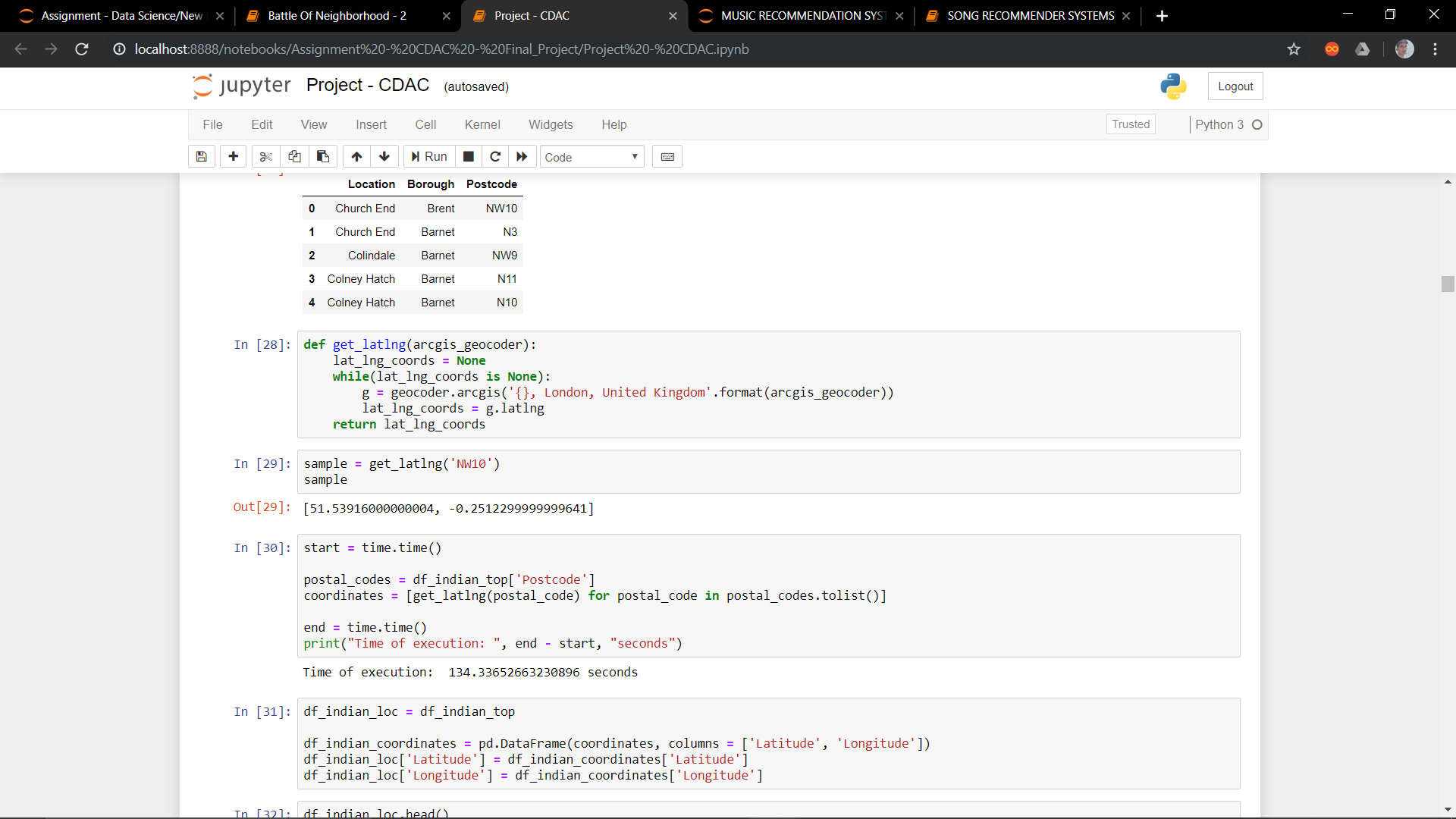
Getting the Data from Wikipedia and inserting it into a data frame



Plotting %Indian Population and different locations using bar graph



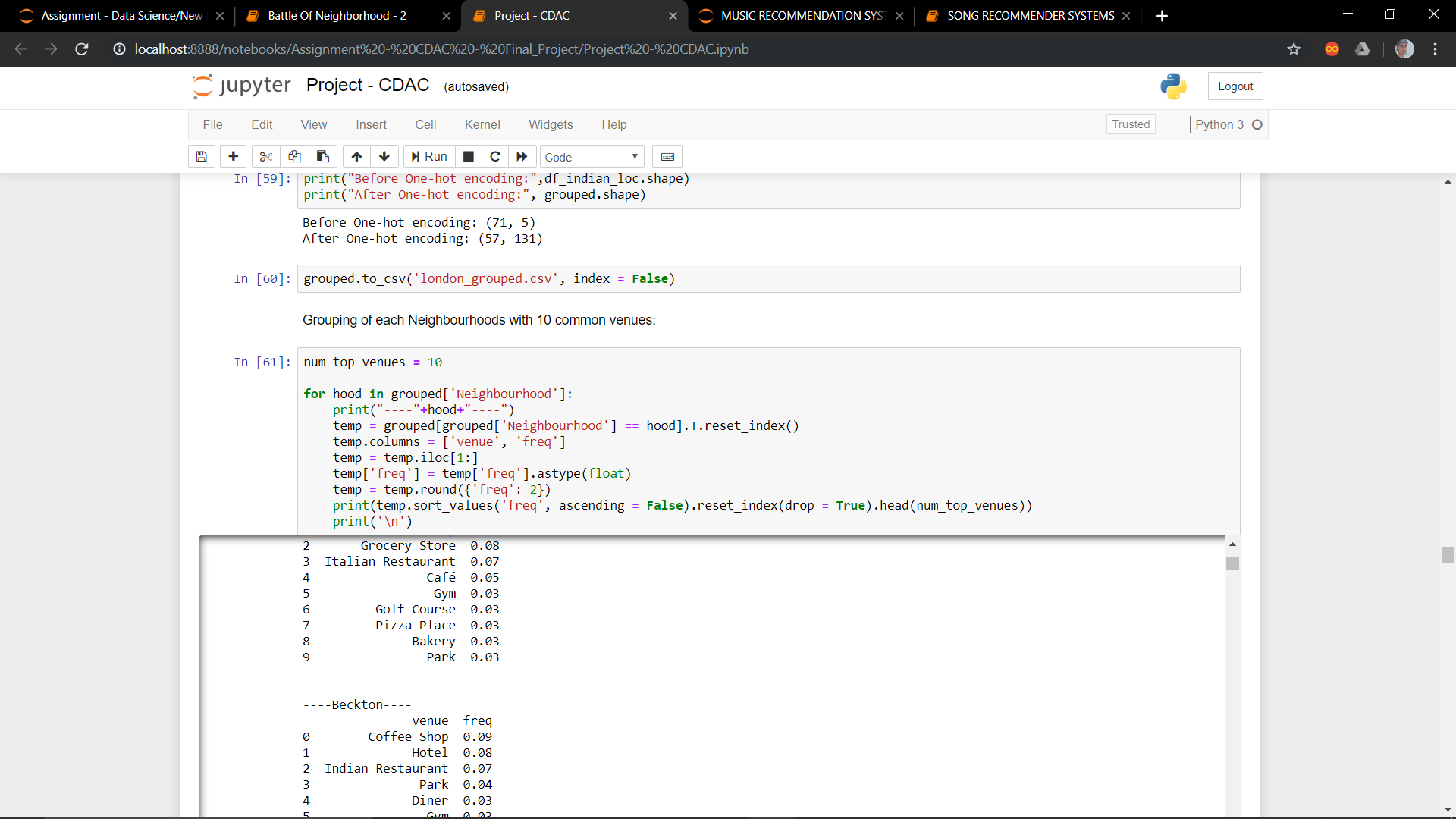
Assigning Latitude and Longitude to each Neighborhood and then store it into a new dataframe



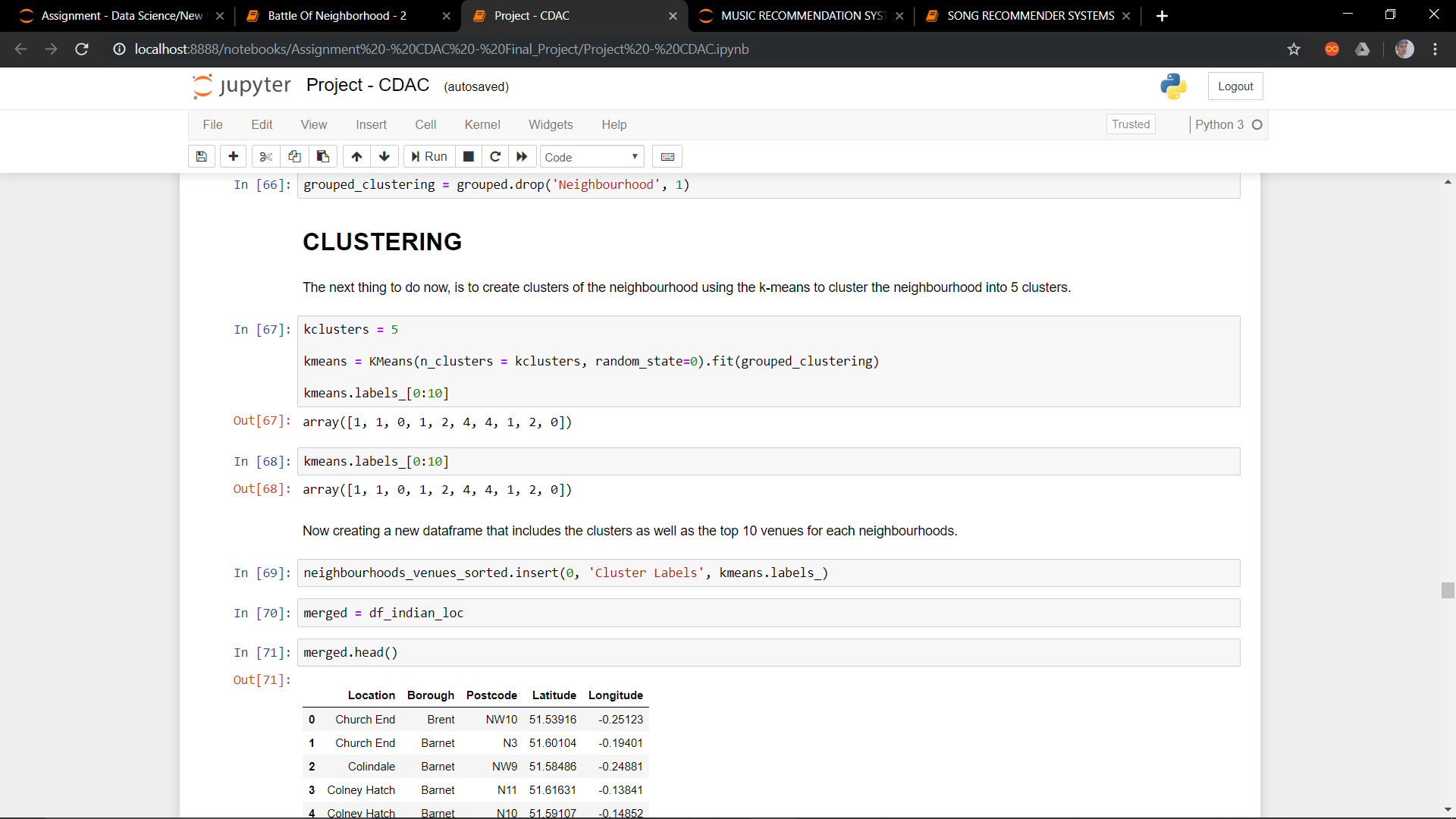
ONEHOT encoding of dataframe obtained in previous step



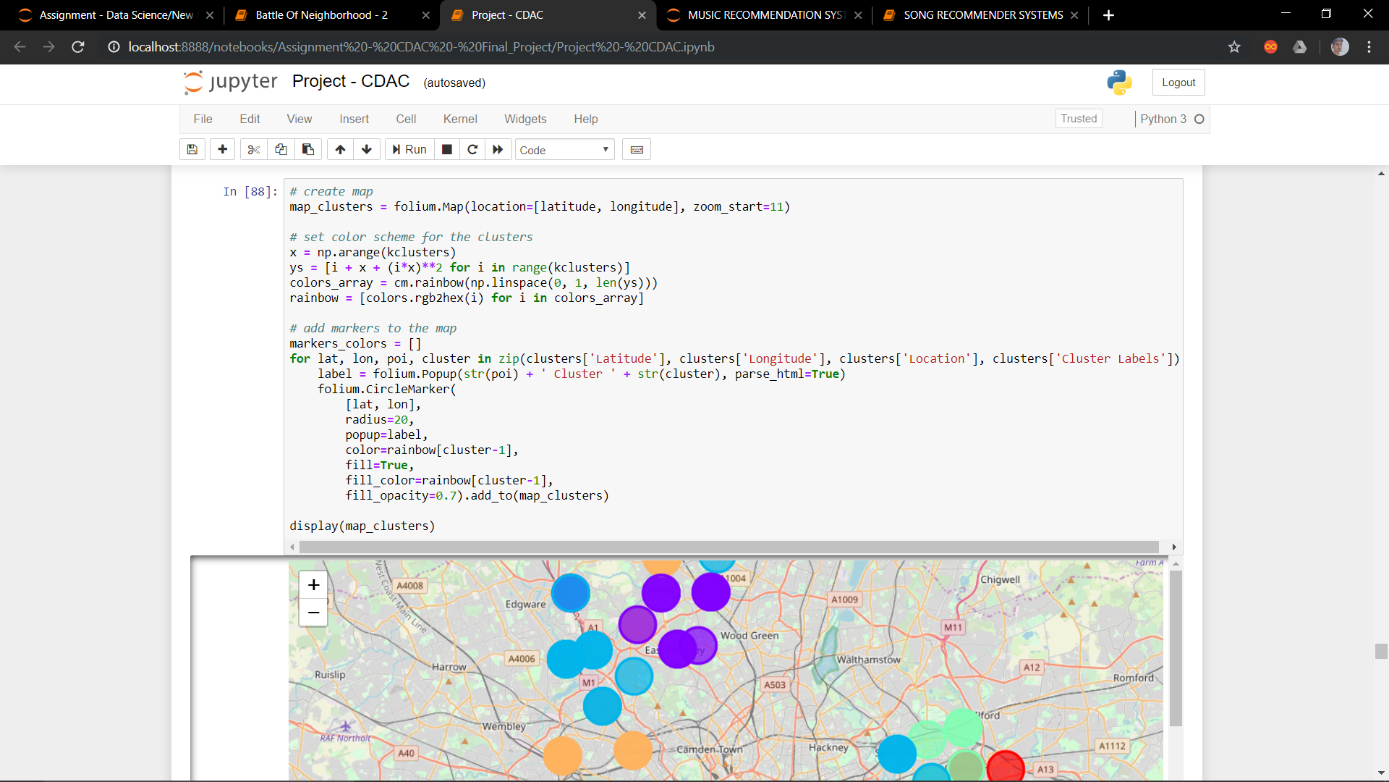
Getting top 10 location from each Neighborhood



Applying clustering to above dataframe



Plotting the different Clusters on the map of LONDON



## 4. PROBLEM FORMULATION and POSSIBLE SOLUTION

### 4.1 WHAT

* What is the current scenario of Indian Restaurants in London?
* What are most Indian Populated region in London?
* What are those regions in London where opening a new Indian Restaurant will benefit people and will help the company to increase its footprint.

### 4.2 HOW

* By analyzing the current density of Indian Restaurant and density of Indian Population we can

Suggest where to open a new Indian Restaurant.

### 4.3 PROPOSED SYSTEM

* First extract the Borough Name along with its location.
* Then assign Longitude and Latitude to each Location using the unique POSTCODE.
* Use foursquare to get venues in a neighborhood.
* Select top 10 venues in each neighborhood.
* Divide all the neighborhoods into different Clusters using K-Means Clustering Method

Analyze each cluster.

## 5. RESULTS, FINDINGS and CONCLUSION

### 5.1 RESULTS

* Pubs, Cafe, Coffee Shops are popular in the London.
* As for restaurants, the Italian Restaurants are very popular in the London.
* Although, the Clusters have variations, a very visible presence is the dominance of pubs.
* *We conclude from above observation that CLUSTERS 3 and 5 are scarce in terms of availability of INDIAN RESTAURANTS*
* Hence, above two mentioned clusters are suitable for opening a new INDIAN RESTAURANT

#### 5.2 FINDINGS AND CONCLUSION

It is very important to note that Clusters 3 and 5 are the most viable clusters to create aN **INDIAN RESTAURANT**. Their proximity to other amenities and accessibility to INDIAN PEOPLE are paramount. These 2 *clusters do not have top restaurants that could rival their standards* if they are created.

In conclusion, this project would have had better results if there were more data in terms of crime data within the area, traffic access and allowance of more venues exploration with the Foursquare (limited venues for free calls).

Also, getting the ratings and feedbacks of the current restaurants within the clusters would have helped in providing more insight into the best locat