## IBM DATA SCIENCE CAPSTONE PROJECT

# **The Best Neighborhood to open a café shop in Brooklyn, New York.**

Table of Contents

[Introduction: 1](#_Toc44553238)

[Business Problem: 1](#_Toc44553239)

[Datasets: 1](#_Toc44553240)

[Methodology: 2](#_Toc44553241)

[Data Understanding / Cleaning: 2](#_Toc44553242)

[Data Modelling: 3](#_Toc44553243)

[Model evaluation: 3](#_Toc44553244)

[Result: 4](#_Toc44553245)

[Discussion: 4](#_Toc44553246)

[Limitations: 4](#_Toc44553247)

[Conclusion: 4](#_Toc44553248)

[References: 5](#_Toc44553249)

[Acknowledgment: 5](#_Toc44553250)

## Introduction:

Brooklyn, A 2nd most densely populated county of US and most populated city of New York borough. More than 1600+ eateries are open (as per nyc.gov site [report data](https://experience.arcgis.com/experience/94760ebb935441be87f21d23a601367e) ) in this covid-19 time phase. Since eating out and sipping a cup of coffee is a major part of people social life. For a businessman, it will be a best choice to open a coffee shop or restaurant for good profit. But the question is, as this place is too big, in which area/neighbourhood, he should open his restaurant or coffee shop which can give him maximum profit.

## Business Problem:

We will be exploring the option to open a coffee shop in Brooklyn neighbourhoods through this project. I will assume that the businessman/ owner wants to open a coffee shop in Brooklyn with the need to attract the customers which will be our focus of the project. Here I will be working on this project in hope to find the most suitable neighbourhood for owner to meet the mentioned conditions.

## Datasets:

We need borough and neighbourhood dataset of the Brooklyn, New York.

Luckily, A data repository site which is publicly available for users to download in structured data formats for New York state. It also comes with the geo location of the neighbourhoods which will be helpful for our project. We will be using this data and filter the dataset accordingly.

<https://geo.nyu.edu/catalog/nyu_2451_34572>

I will be also be using Foursquare API to get the coffee shops details on this neighbourhoods.

## Methodology:

We will work with the above downloaded data and using he foursquare API, we will get the coffee shops details in the Brooklyn neighbourhood and we will explore the shops on the Brooklyn map to understand the data. Then we can make cluster of the venue details and explore the evaluate the results to find out the optimal neighbourhood.

### Data Understanding / Cleaning:

As we have downloaded the data in JSON format, we will have to convert this in pandas data frame to work with it.

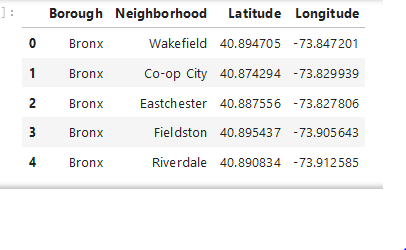


Figure . New York data with Geo location

Also, the data we have downloaded is for New York state. We will have to filter the data for Brooklyn. After filtering the data we can see these neighbourhood in the map to understand the neighbourhood.

A picture containing text, map

Description automatically generated

Figure . Brooklyn neighbourhoods

After this we will use foursquare API to get all the coffee shops details in the neighbourhood of Brooklyn. After converting and cleaning the json response data from JSON, we uploaded the data into pandas dataframe for more readability and further process.

A screenshot of a computer

Description automatically generated

Figure . Brooklyn coffee shops

We can see the Brooklyn neighbourhood’s coffee shops in the map for more understanding and clarity of data.

A picture containing text, map

Description automatically generated

Figure . Brooklyn Coffee Shops

### Data Modelling:

The time has finally come to run the algorithm and solve our problem of finding the best neighbourhood for owner. As we are going to use clustering model which is unsupervised type of machine leaning model. Here we do not have to split the data into train and test dataset as we do in supervised learning models. I will use the K Means clustering method of unsupervised model.

As the data is multi variative we will have to transform/normalize the data before feeding it to the model. For this I have used the StandardScaler transform method.

#### K- Means Clustering:

In K Means Modelling, we set the number of clusters prior to fit the model. In this project, I have chosen the number of clusters as 3 and build the model. Once the model is built, I have feed the above generated data to the model.

### Model evaluation:

Let us see the model evaluation results, we can see the number of coffee shops present in different clusters.

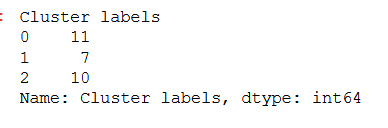


Figure . Number of coffee shops per cluster

Let us also visualize the generated clusters by our model in the map. We can see 3 different coloured dots in the map, each colour represents a cluster.

A picture containing text, map

Description automatically generated

Figure . Brooklyn neighbourhood clusters map

## Result:

After visualizing the clusters on the map and seeing the number of coffee shops present in all the cluster neighbourhoods. It seems the best option would be to open coffee shop on cluster 1 area as this area has less numbers of coffee shops.

## Discussion:

Upon reviewing the above data and mode evaluation results, the cluster 1 will be the best option for owner to open a new coffee shop as there are less number of coffee shops and also the neighbourhood will be less competitive for the coffee shop than other neighbourhoods.

### Limitations:

As we have generated the models based on the number of coffee shops present in the area, this model may not be accurate as there will be more variables present in the real problem statement like the population of the neighbourhood, the neighbourhood area like if there are offices nearby etc.

## Conclusion:

We have used K mean cluster model, where we have decided to cluster the coffee shops in the neighborhood of Brooklyn and find out the best place to open a new coffee shop. This project clearly defines that the user should open his coffee shop in the cluster 1 areas due to the reasons mentioned above.

## References:

<https://geo.nyu.edu/catalog/nyu_2451_34572>

https://developer.foursquare.com/

## Acknowledgment:

I have taken efforts in this project. However, it would not have been possible without the help of the data repository and tools like foursquare etc. I would like to extend my sincere thanks to all of them.