



The Battle of Neighborhoods



Introduction

- To open a pizza place in a neighborhood we need to research different locations in a particular area suitable for booming the business and gaining extra profit. For this we need to analyse the data.
- In this project, the Staten Island, New York data set is used to provide optimal location for a contractor who wants to open a pizza place with the help of machine learning clustering technique and Foursquare API.



Business Problem

- The business problem for the particular project is as follows :

What are the optimal locations in Staten Island, New York for a contractor to open Pizza Place?


- Target audience :
 - The contractors who wants to open pizza place in Staten Island, New York.



Data Discription

Following data is used to solve the business problem :

1. Dataset which contains the data of New York neighborhoods.
2. Coordinates of different neighborhoods in Staten Island, New York.
3. Data related to Pizza Places.



	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585



Methodology

Methodology includes following steps:

1. Data preprocessing
2. Feature extraction
3. Modal creation



Data Preprocessing

- Data preprocessing is important to remove unwanted data in the dataset.
- In New York dataset used in this project contains all the neighborhood data in New York. So data not related to the Staten Island neighborhood is removed.



Feature Extraction

- **One hot encoding** is a process by which categorical variables are converted into a form that could be provided to ML algorithms to do a better job in prediction.
- Use Foursquare API to search for a specific type of venues, to explore a particular venue, to explore a Foursquare user, to explore a geographical location, and to get trending venues around a location.
- Creating new dataframe for venues and using it to create Model.



Modal Creation

- **K-Means Clustering** : k -means is vastly used for clustering in many data science applications, especially useful if you need to quickly discover insights from unlabeled data. Using k -means clusters are created and analysed.



Observations

Most of the Pizza Places are in cluster 2 which are around Tompkinsville, Rosebank, Shore Acres, etc. Lowest Pizza Places are present around neighborhoods in Cluster 0. There are many neighborhoods in Cluster 0. Though most of these neighborhoods are located near the neighborhoods in Cluster 2 and Cluster 1. So to start a Pizza Place in such location is not a good choice. The neighborhoods around Tottenville, Howland Hook, Travis, West Brighton, Sunnyside are some of the optimal locations to open a pizza place.



Conclusion

In this project, I have determined business problem, performed data preprocessing, applied Machine Learning K-Means Clustering method to solve the business problem.