COURSERA CAPSTONE PROJECT

IBM APPLIED DATA SCIENCE CAPSTONE: BATTLE OF NEIGHBORHOODS

OPENING A NEW SHOPPING MALL IN NEW YORK CITY, USA

BY

JACOB THEOPHILUS AYOKANMI

OCTOBER, 2019

INTRODUCTION

Shopping is a great way to relax and enjoy yourselves, after a tedious and exhaustive workday. The weekend is meant for shopping, time for you to exercise your purchasing power and tick items off your shopping list.

In a city like New York, rated as second top technology city in the USA (https://www.businessinsider.com/most-high-tech-cities-in-the-us-2017-8/), the shopping mall is a one-stop venue for all types of shoppers. It is a great opportunity for the effective distribution of goods and services.

Shopping Mall is a great way of generating passive income because it is a lucrative business that retailers will continually pay a huge amount of money to have an outlet in your building.

One of the things retailers, check out when renting a space in a shopping mall is the location and its immediate environment. The better the neighborhood, the more retailer the property developer' will attract.

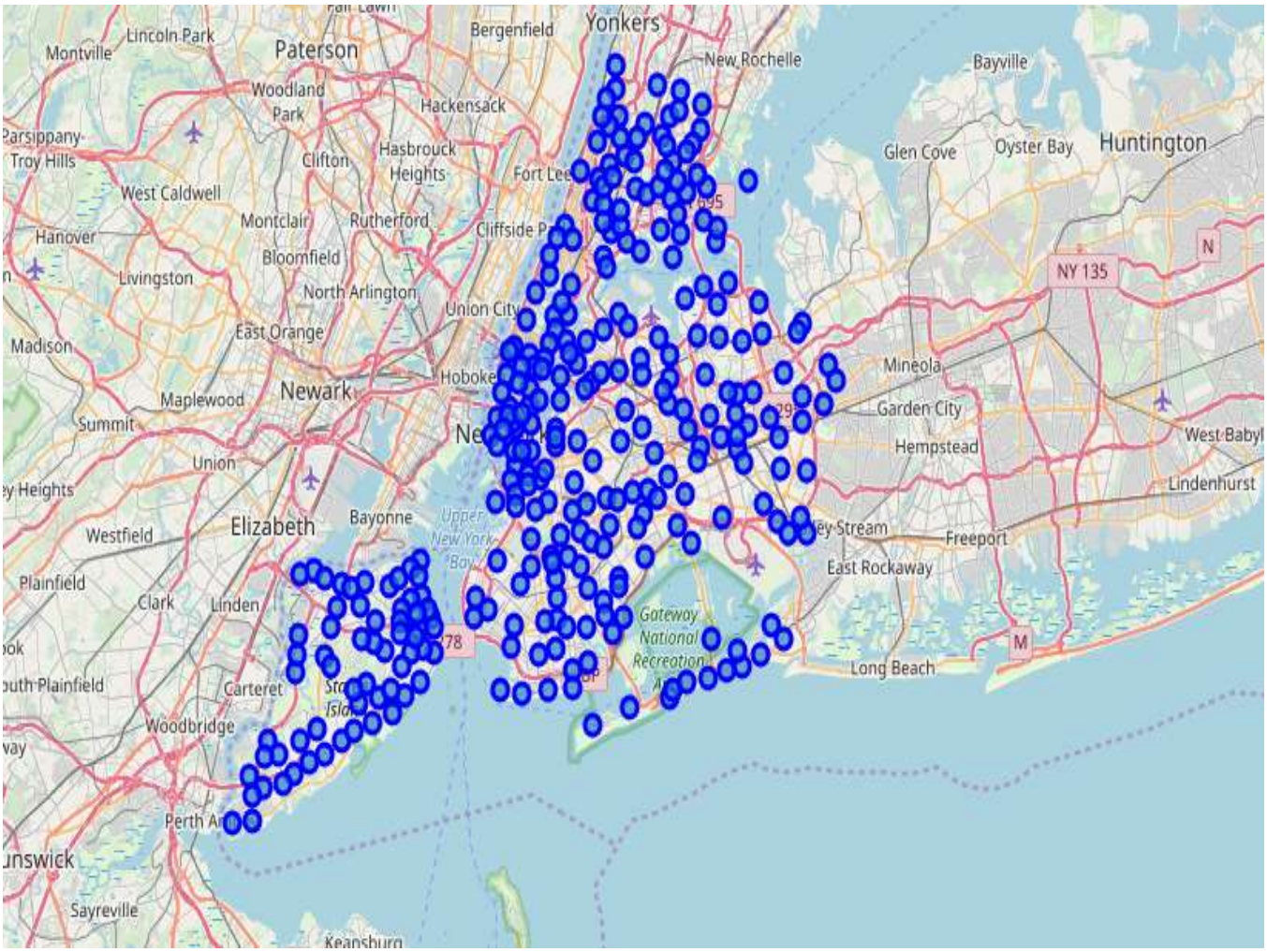
Therefore, it is of great importance to select a good location for building a shopping mall.

BUSINESS PROBLEM

My client, a successful Africa property developer wants to expand his business portfolio by having an extra modern shopping mall in New York City, USA.

Since he has never been to the USA before his problem is to choose a list of possible locations that will be best for opening a shopping mall.

Which neighborhood and locality in New York City are best for opening a shopping mall?



Map of New York with its Neighborhood imposed on it.

TARGET AUDIENCE

1. Business personnel that needs locations in NYC build a shopping mall.
2. Property Developer needs a good location for their clients.
3. Data Science Enthusiasts that want to build learn and implement some of the most used exploratory data analysis techniques.

DATA

The data required to solve this problem include:

1. The list of the entire neighborhood in New York, with their postcode and Borough.
2. The list of all venues in each neighborhood in New York City.

[54]:

	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
5	Bronx	Kingsbridge	40.881687	-73.902818
6	Manhattan	Marble Hill	40.876551	-73.910660
7	Bronx	Woodlawn	40.898273	-73.867315
8	Bronx	Norwood	40.877224	-73.879391
9	Bronx	Williamsbridge	40.881039	-73.857446
10	Bronx	Baychester	40.866858	-73.835798
11	Bronx	Pelham Parkway	40.857413	-73.854756
12	Bronx	City Island	40.847247	-73.786488
13	Bronx	Bedford Park	40.870185	-73.885512
14	Bronx	University Heights	40.855727	-73.910416

DATA SOURCE

1. The data for the list of all neighborhoods in NYC is located on the NYU Spatial Data Repository website: https://geo.nyu.edu/catalog/nyu_2451_34572.

2. I will use Four Square API to obtain the top 100 venues in each neighborhood within a radius of 2000.

METHODOLOGY

1. The NYC Neighborhood Data was obtained from the USA website" https://geo.nyu.edu/catalog/nyu_2451_34572 , in json format. The name of neighborhood, postcodes and borough was extracted from the JSon file and save as a data frame.

2. Geopy library was used to obtain the coordinates for each Borough in the data frame. Two columns were added to the data frame for the Longitude and Latitude.

3. The list of venues in NYC was obtained using Foursquare API and the weigh according to the Shopping Mall.

4. With the help of Unsupervised Machine Learning Algorithms (K means Algorithms), the neighborhoods were segmented into 3 clusters.

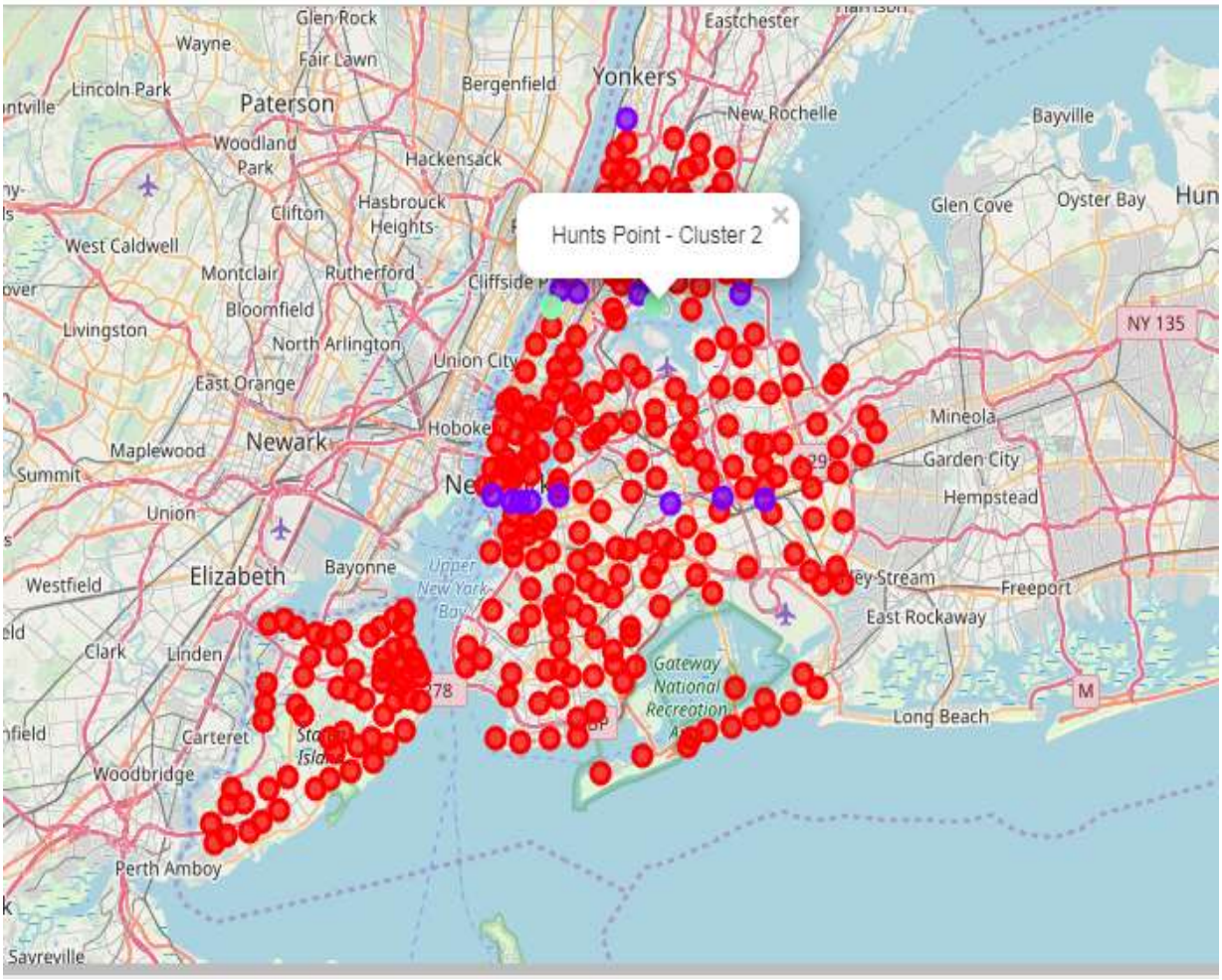
RESULT

The result of the project is as follows:

Cluster 0: No single shopping mall in this neighborhood.

Cluster 1: High/Moderate number of shopping mall in this neighborhood.

Cluster 2: low number of shopping mall in this neighborhood.



DISCUSSION

The majority of the shopping Malls are concentrated around Manhattan, Bronx Whitestone Bridge, and Throg Necks Bridge. Cluster one has the highest number of shopping malls and a small number in cluster two.

On the other hand, cluster zero which comprises of the majority of the neighborhood in New York has no shopping mall. This presents

a great opportunity for property developers to open shopping malls in the neighborhood of cluster zero.

Property developers should avoid neighborhood in clusters one, because of the competition the shopping malls in the neighborhood can generate.

LIMITATION AND SUGGESTIONS FOR FUTURE RESEARCH

1. The city was chosen randomly without any consideration for statistics like population density, and so on. Also, we only considered only one factor i.e. frequency of shopping malls, other factors could influence the location decision of a new shopping mall.
2. We are only limited to the top 100 venues around the 2000 radius due to the limited number of calls we can make with the Foursquare API.

CONCLUSION

In this project, I have:

- Identified a business problem.
- Sourced for the data required.
- Cleaned and prepared the data for analysis.
- Performed machine learning using K-Means Algorithm to cluster the data, and

- Finally arriving at a data-driven recommendation for property developers and property investors.

At the end of this project, the neighborhoods in cluster O are the most preferred locations to open a shopping mall.