CHOOSING THE BEST RESTAURANT LOCATION IN NEW YORK CITY

IBM Applied Data Science Capstone

Janrey D. Nevado

TABLE OF CONTENTS

01 INTRODUCTION

02 DATA REQUIREMENTS

03 METHODOLOGY

04 RESULTS

05 DISCUSSION

06 CONCLUSION

INTRODUCTION

Background, Problem, and Target Audience

01

BACKGROUND

Finding a good restaurant location is harder than some people think. Food and service are important to the success of a restaurant, but the location is just as crucial. Some other factors to consider include the following:

- parking,
- accessibility,
- number of competitors, and
- crime rates among others.

PROBLEM STATEMENT

A **new market entrant** is looking to have answers to the following questions:

- Which neighborhood(s) is/are the best location(s) to build a new restaurant considering the following factors:
  - Historical crime rate
  - Number of competitors
- What cuisine dominates in each of the neighborhoods?
- Knowing these things, what cuisine could be introduced to the location?

TARGET AUDIENCE

To answer these questions, the company reached out to our team. The company's objective is to locate the best neighborhood(s) to build a new restaurant in. The company also expects to understand the rationale behind the recommendations.

DATA REQUIREMENTS

Data Collection, Preparation, and Cleaning

02

LOCATION DATA

The data will be primarily used to obtain the number of restaurants by type in every neighborhood.

Source: Foursquare API

CRIME DATA

This data will be used to obtain the number of crimes recorded for every neighborhood in New York City.

Source: NYC OpenData

METHODOLOGY

Exploratory Data Analysis and Clustering

03

METHODOLOGY

Since we wanted to find the neighborhoods with least number of crimes and least number of competitors, *k-means clustering* will be used to cluster neighborhoods that have similar characteristics.

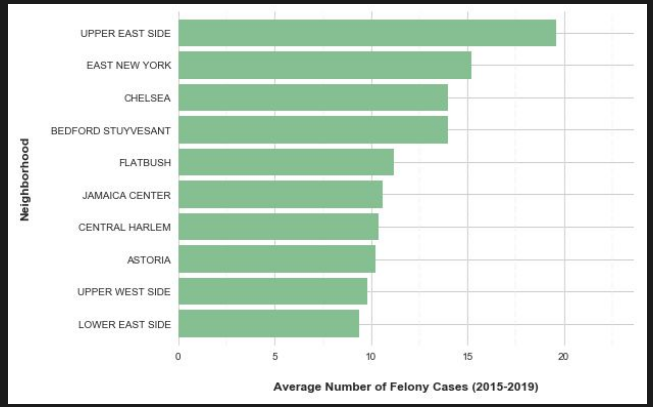
What we would like to obtain are the neighborhoods that have the least number of crimes and least number of competitors.

Number of competitors

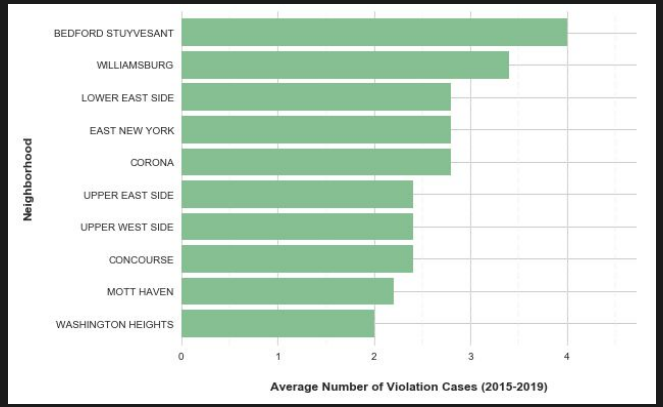
Number of crime cases

TOP NEIGHBORHOODS BY CRIME RATE

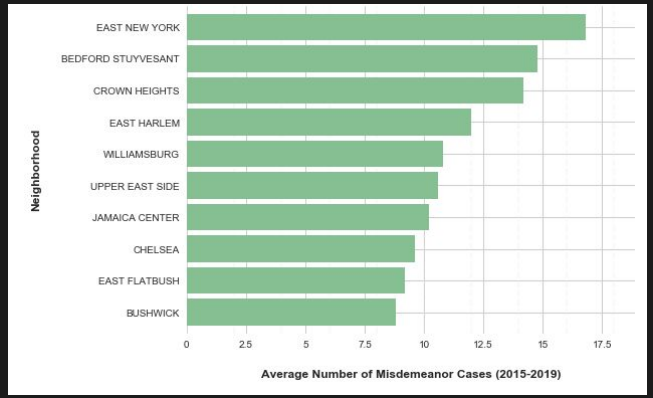
FELONY



VIOLATION

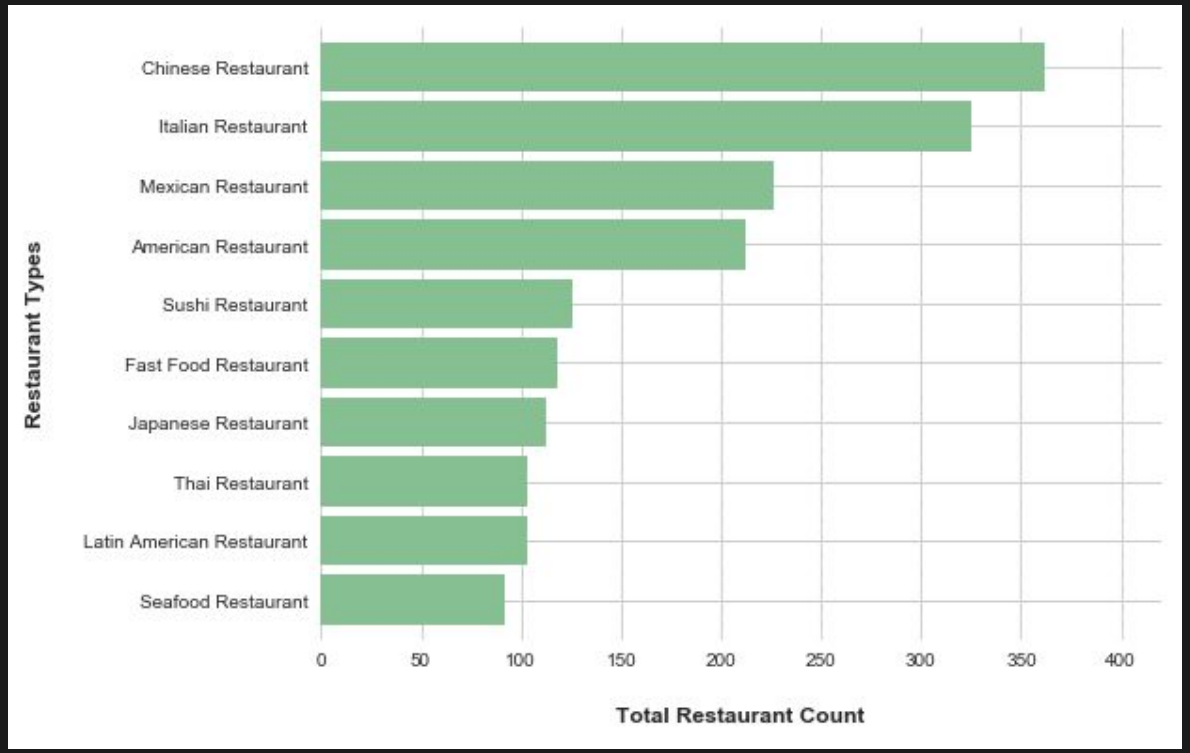


MISDEMEANOR



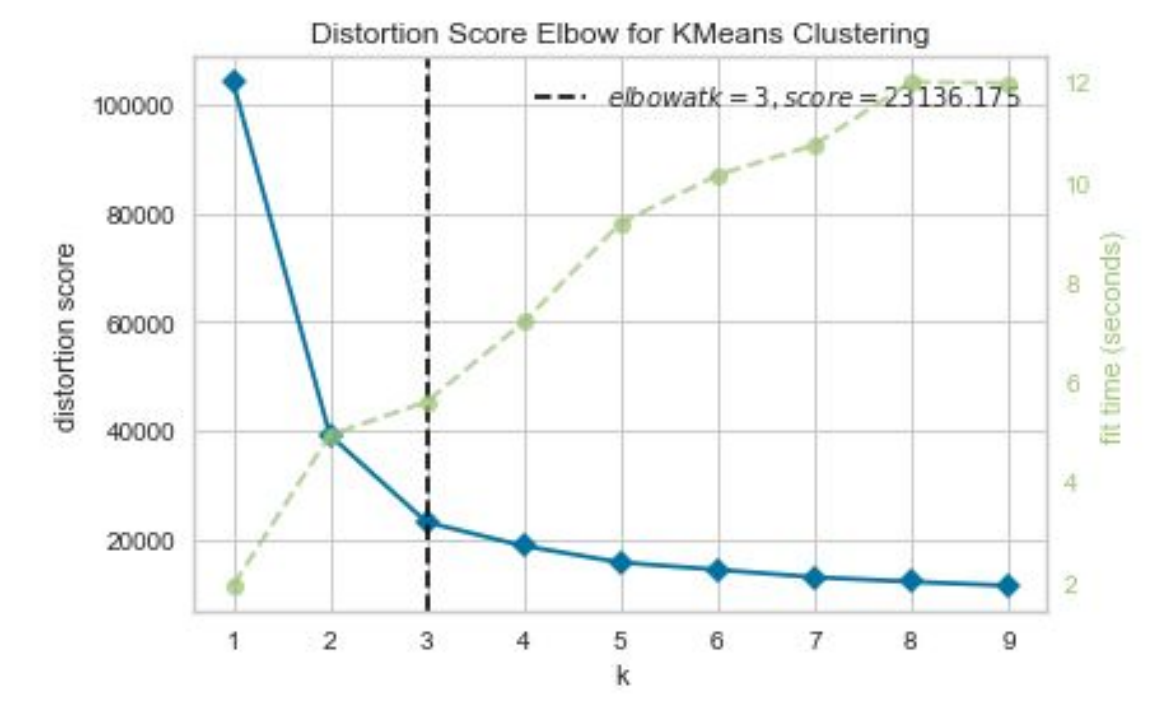
According to the charts, the top neighborhoods by average number of cases are mostly in Manhattan and Brooklyn (**Bedford Stuyvesant**, **East New York**, **Lower East Side**, **Upper East Side**, and **Upper West Side**).

TOP RESTAURANT TYPES IN NYC



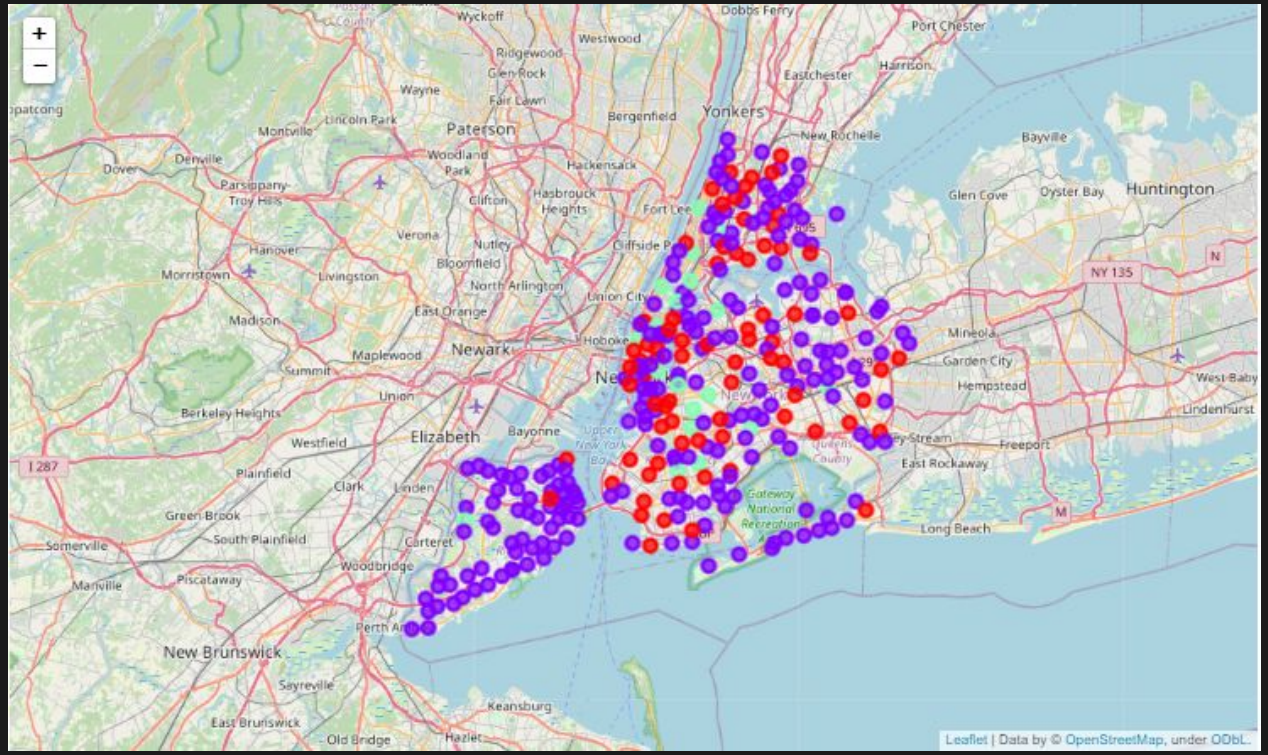
The most common restaurant types in New York City are **Chinese**, **Italian**, **Mexican**, **Fast Food**, **Japanese**, **Thai**, **Latin American**, and **Seafood Restaurants**.

FINDING OPTIMAL K



Elbow method is used to find the optimal number of clusters k. The scoring parameter metric used is **distortion score**, which computes the sum of squared distances from each point to its assigned center. The optimal number of clusters k is **3** using elbow method.

MAP WITH PLOTTED CLUSTERS



RESULTS

Clustering Results

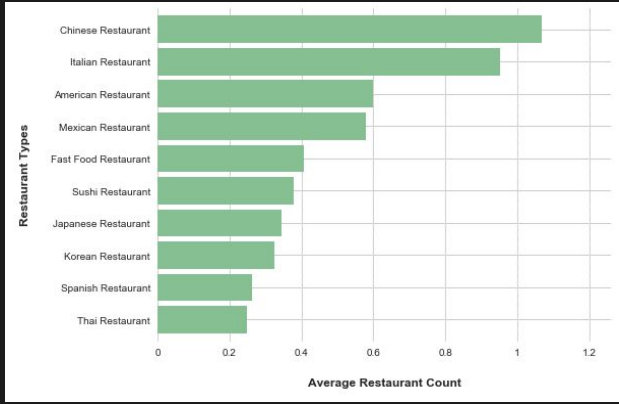
04

CLUSTERING RESULTS

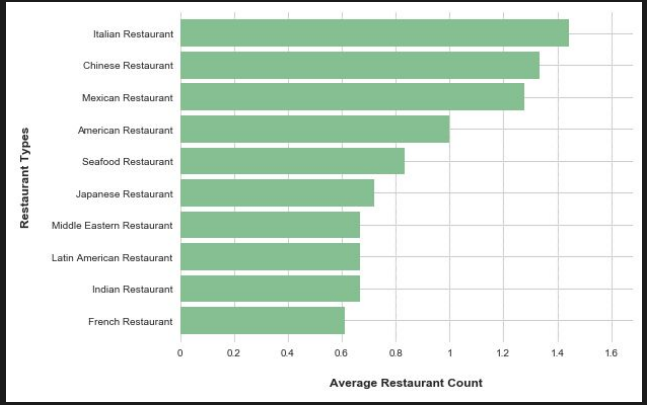
	Neighborhood Count
Cluster 1	74
Cluster 2	212
Cluster 3	18

AVERAGE RESTAURANT COUNT

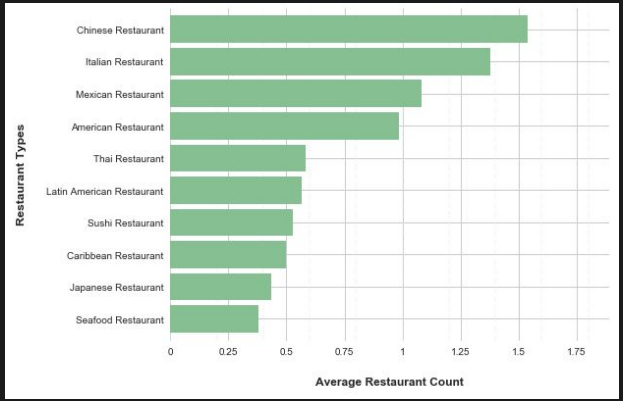
CLUSTER 1



CLUSTER 1



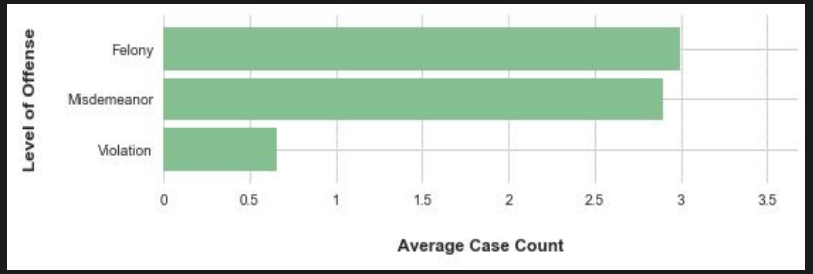
CLUSTER 2



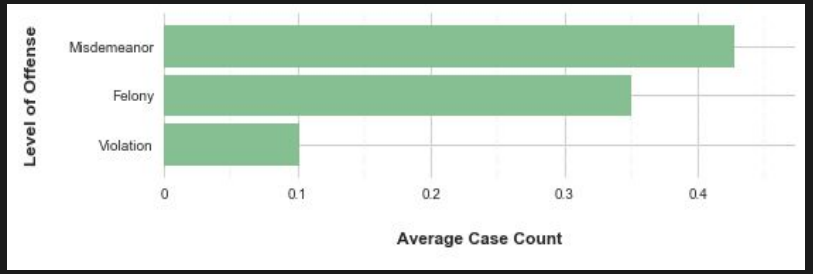
The top restaurant types in all clusters are very similar to the one top restaurant types in all neighborhoods before clustering.

AVERAGE CRIME COUNT

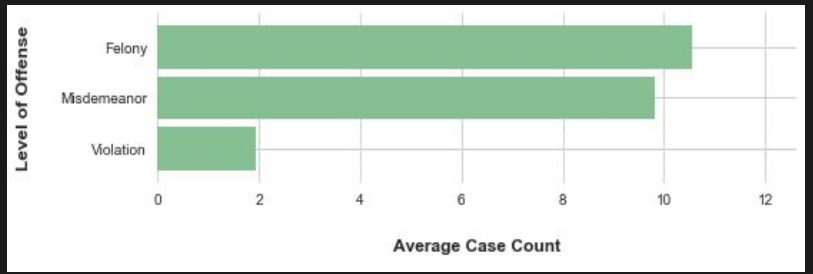
CLUSTER 1



CLUSTER 2



CLUSTER 3



According to the graphs, **Cluster 2** has the lowest average crime count while **Cluster 3** gas the highest average crime count. This suggests that neighborhoods in Cluster 3 are not that safe while neighborhoods in Cluster 2 are pretty safe. **Cluster 1** also has a pretty low average crime count.

DISCUSSION

Summary and Recommendation

05

CLUSTER INTERPRETATION

	Neighborhood Count	Description of Cluster
Cluster 1	74	Relatively high average number of restaurants and low average number of crimes
Cluster 2	212	Relatively low average number of restaurants and very low average number of crimes
Cluster 3	18	Relatively high average number of restaurants and high average number of crimes

ANSWERS TO QUESTIONS

- Which neighborhood(s) is/are the best location(s) to build a new restaurant considering the following factors:
  - Historical crime rate
  - Number of competitors

We can see from the clusters that Cluster 2 neighborhoods have the things that we are looking for because the average restaurant count is the lowest among three clusters, which suggests that competition is not that intense. Crimes are also almost nonexistent in these areas, which is good for the business in the long run.

- What cuisine dominates in each of the neighborhoods?

Chinese, Italian, Mexican, and American Restaurants are the most common restaurants in the cluster, so creating a new restaurant that is similar to these restaurants may not appeal to customers.

- Knowing these things, what cuisine could be introduced to the location?

Going with other Asian cuisine is recommended because the data shows that Asian restaurants are more common and familiar in the area, so it would be easier for the company to attract new customers.

CONCLUSION

Learning and Realization

06

CONCLUSION

In summary, analysis of the data used in this project shows that New York City is still a very attractive location to build a new restaurant because the majority of the neighborhoods have low crime rates and low competition. It is also reinforced that New York City is a really diverse place given the numerous restaurants offering different cuisine.

The analysis done is rather simple, and it will be better if other datasets are also incorporated in the study, like population growth, population density, rent prices, nearness to market, nearness to transportation, and number of tourists among others. Other clustering algorithms can also be considered given the new forms of data that will be added.

Overall, I really enjoyed doing all the courses in this specialization. As a newbie in data science, I really learned a lot, and I think I renewed my passion to solve problems and to come up with solutions that are actionable.

THANKS

Do you have any questions?

janrey.nevado@gmail.com

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.