

IBM

**Applied Data
Science Capstone
REPORT**

**Capstone Project - The Battle of
Neighborhoods**

Final Project

SUMMARY

- **Introduction**
 - **Data**
 - **Methodology**
 - **Results**
 - **Discussion**
 - **Conclusion**
-

Introduction:

In this project, we will try to help any businessperson who wishes to invest in the city of Baltimore by answering the following questions: What and where?

What: What is the ideal investment in a city?

Where: Where is the optimal place for this investment?

To answer these two questions we must fight a violent battle with data and defeat it, so let us start.

Data:

In this project, we will need one data set that includes:

- The name of the neighborhood.
- Population.

The coordinates of each neighborhood.

The set of data that we will use and you can get from [here](#) contains the following data:

- The name of the neighborhood.
- Count the population in each neighborhood.
- The number of inhabitants in each neighborhood according to their race.
- The number of males and females in each neighborhood.
- The distribution of population in each neighborhood within the age groups and the number of each group.

But it does not contain coordinate data for each neighborhood, so we will need to use other tools to obtain it.

Methodology:

As we mentioned in the introduction, there are two questions that must be answered in order to find the best solution, what and where? Therefore, we asked the first question to the investor, "What is the project

that you want to invest in?", And his answer was, "I would like to build a restaurant that offers Mexican food and drinks".

To answer this question, we used the following methodology:

- **Data preparation:** The data files we have are in "geojson" format so we will extract the required data from them and put them into a data frame, and then we clean it by deleting the lost values.
- **Data completion:** here we lack the coordinates of each neighborhood, so we will rely on the "geopy" package to obtain the coordinates of each neighborhood by its address.
- **Explore the places:** After we have prepared all the data related to the neighborhoods, we still have to explore the places in each neighborhood, such as restaurants, museums, shopping centers, etc. You may think that it is an impossible job, or you have to go to Baltimore to count these places yourself. Do not worry, it is much easier than that, just using Foursquare API, you need:

1. Client ID, Secret ID, Version.

You must register a [Foursquare](#) account to get them.

2. Latitude, Longitude.

You get it from the data frame you have prepared.

- **Clustering:** Finally, we will form a data frame that includes population numbers according to their race and the number of Mexican restaurants in each neighborhood, and then we will apply the k-means algorithm on it.

Results:

At the end of this research, we obtained the following results:

- When analyzing the correlation, we noticed that the Latin and white population correlated with the

number of Mexican restaurants in close proportions, respectively ().

- The neighborhoods were clustered into six clusters.
- The cluster one included 32 neighborhoods, and it contained the following percentages for the total city:
 - 1) 7% of the total number of Mexican restaurants.
 - 2) 20% of the total population.
 - 3) 45% of the total Latinos population.
 - 4) 45% of the total Whites population.
- The cluster two included 4 neighborhoods and also contained:
 - 1) 7.5% of the total number of Mexican restaurants.
 - 2) 6% of the total population.
 - 3) 10% of the total Latinos population.
 - 4) 15.5% of the total Whites population.
- The cluster six included 130 neighborhoods and also contained:

- 1) 52.5% of the total number of Mexican restaurants.
 - 2) 21% of the total population.
 - 3) 21% of the total Latinos population.
 - 4) 23% of the total Whites population.
- These neighborhoods that we have classified as suitable to host a new Mexican restaurant.

Discussion:

We have concluded in this research that three groups of neighborhoods may be suitable for building a Mexican restaurant, but in reality, not all of these neighborhoods are suitable for that, but there are a number of neighborhoods better for such an investment.

From our previous analysis, we extracted a conversation of a number of Mexican restaurants, such

as the white and Latino population as a positive effect, and the black population as a negative effect.

Therefore, we will resort to these factors to choose the best neighborhood to establish our restaurant.

Conclusion:

Finally, I would like to point out that this research can be used to help in selecting the optimal place for any other investment in Baltimore.

In the data set that I used (will be located within the project) many other features that I have not used in this research such as the age of the population, their social status, and their gender.
