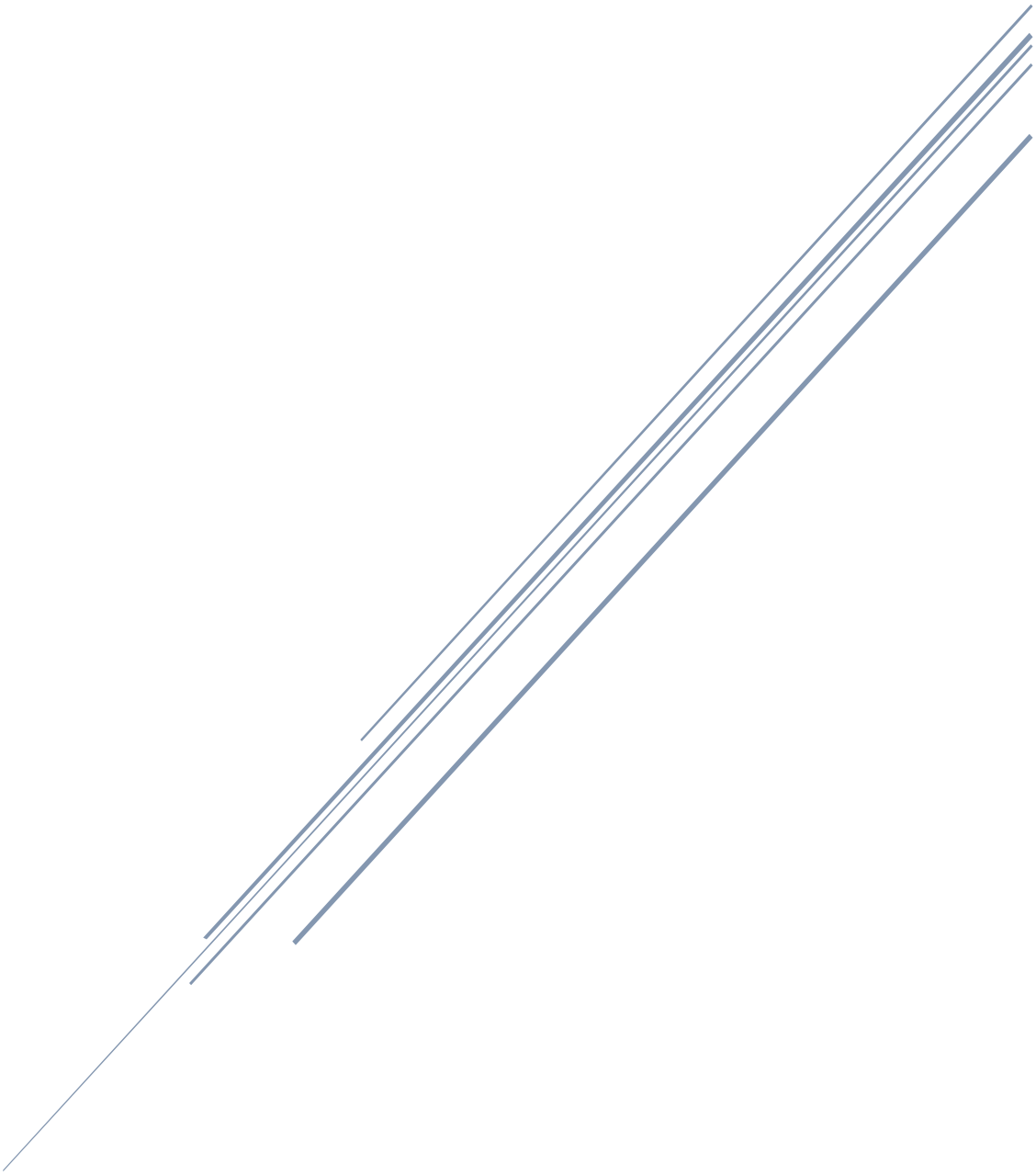
# FINAL CAPSTONE COURSERA

2019



Danaska Alonso Martínez

# INTRODUCTION

The goal of this study is to locate the best area or neighborhood to place a Spanish restaurant in Manhattan. The customer will be someone willing to start a business of this nature in the Big Apple.

# DATA

In order to achieve this objective, we will study all the neighborhoods of Manhattan and their purpose, the location of the main restaurants and the location of the current Spanish restaurants.

The data was collected from a complete source with neighborhoods of Manhattan. However, data as latitude and longitude of the different neighbourhood was not included. Therefore, we used the API of Google Maps to get the related data to the neighborhoods.

The data involved in the services of the city was extracted from Foursquare and used it in the algorithm.

# METHODOLOGY

The core of our study is the use of the classification algorithm K-means which has a loose relationship to the k-nearest neighbor classifier, a popular machine learning technique for classification that is often confused with k-means due to the name. Applying the 1-nearest neighbor classifier the cluster centers obtained by k-means classifies new data into the existing clusters.

Our research can be divided into two parts. In the first part we will discuss the neighborhood study of Manhattan.

## Manhattan's Neighborhoods

Related to the Manhattan's neighborhoods, we found that the city can be separated in five areas due to the different services that they offer. The separation has been possible thanks to the application of our algorithm k-means and the division in five clusters. The classification would be:

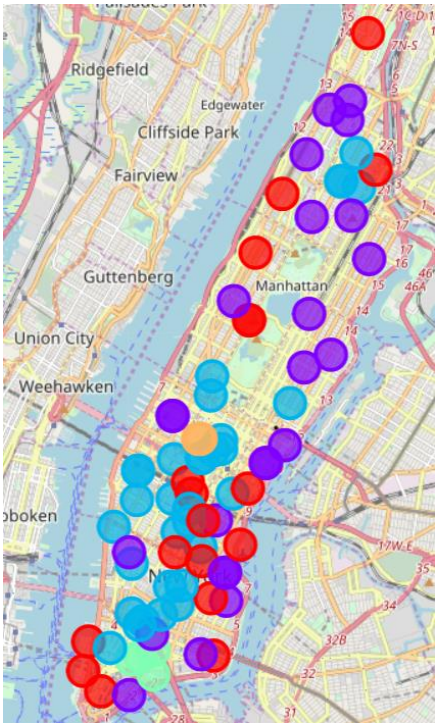
1. Cluster 0 (Red). Resident neighborhood with 20 neighborhoods.

2. Cluster 1 (Purple). Outer area services neighborhood with 26 neighborhoods.

3. Cluster 2 (Blue). Main area services with 32 neighborhoods.

4. Cluster 3 (Green). South area services with 4 neighborhoods involved.

5. Cluster 4 (Orange). Cultural services with 2 neighborhoods.




*Map 1. Map of Manhattan with the different neighborhoods classified by K-means.*

As it is showed in the picture above, the main area of services is in the south of the city, location where will be placed our business.

## Manhattan's Restaurants

Again, we will use the k-means algorithm for this purpose. The final classification would be:

1. Cluster 0 (Red). American Restaurants.

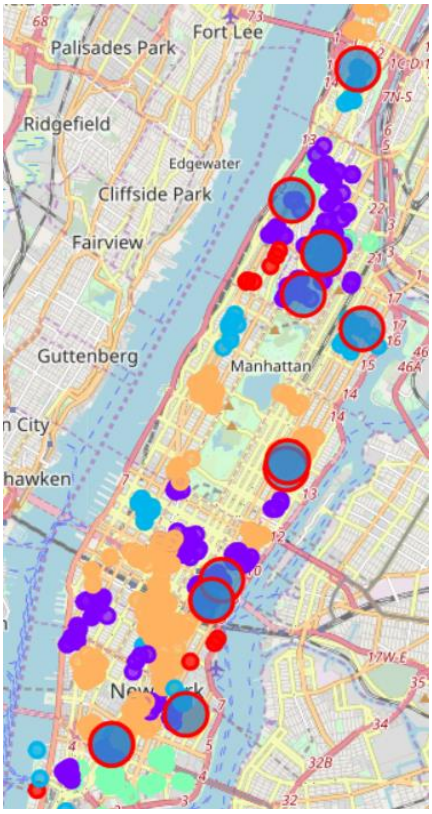
2. Cluster 1 (Purple). American Restaurants.

3. Cluster 2 (Blue). Mexican Restaurants mainly.

4. Cluster 3 (Green). Asiatic Restaurants, mainly Chinese.

5. Cluster 4 (Orange). Italian – American Restaurants.

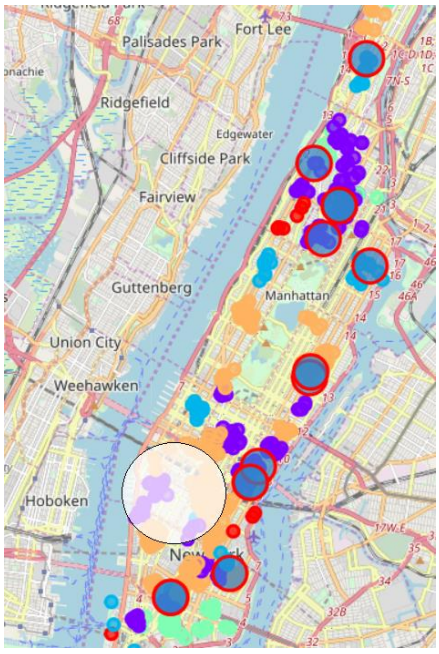
Exploring this data, we found that the Spanish restaurants are located all around but more often in the Cluster 1.




*Map 2 Different clusters classified by K-means in relation with the restaurant types.*

## Results and Conclusions

The selection of the area will be inside of our cluster of main services (Cluster 2 of Map 1) due to it is a good area of services, main Europeans restaurants as Italians.




*Map 3 Final location of the new Spanish Restaurant in Manhattan*