



NEW RESTAURANT IN NYC

Date: Oct/2020 Created by: Rui Martins

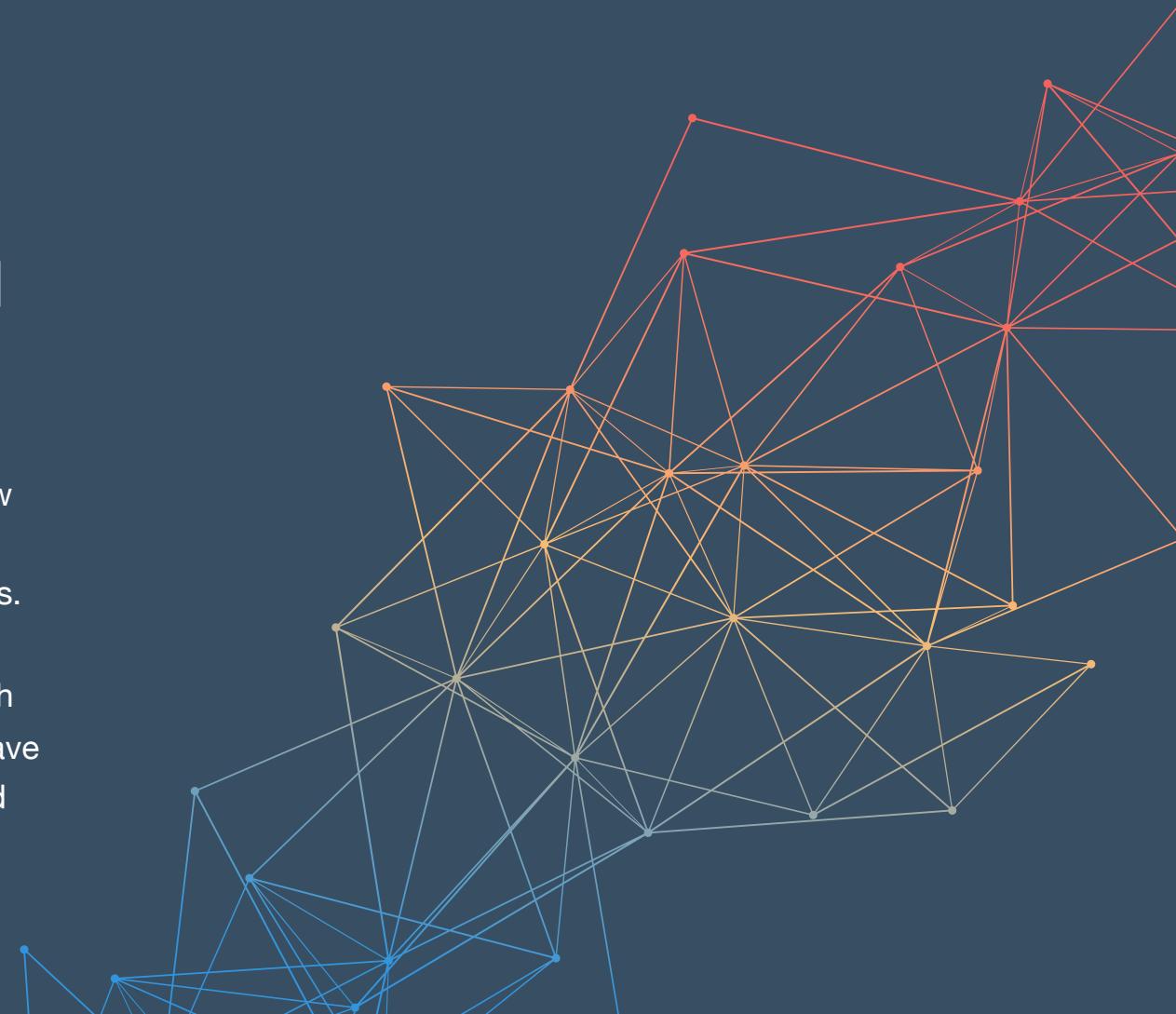
INTRODUCTORY NOTES

Someone looking to open a restaurant in New York City, where would you recommend that they open it?



In this project will try to find an optimal location for a new Spanish restaurant in NYC, we need to find locations (Neighbourhood) that that isn't full of Spanish restaurants.

On the other hand, We will associate the success of each restaurant to the number of 'likes' and their 'rating' we have to find the neighbourhood with the number of rating and their food security



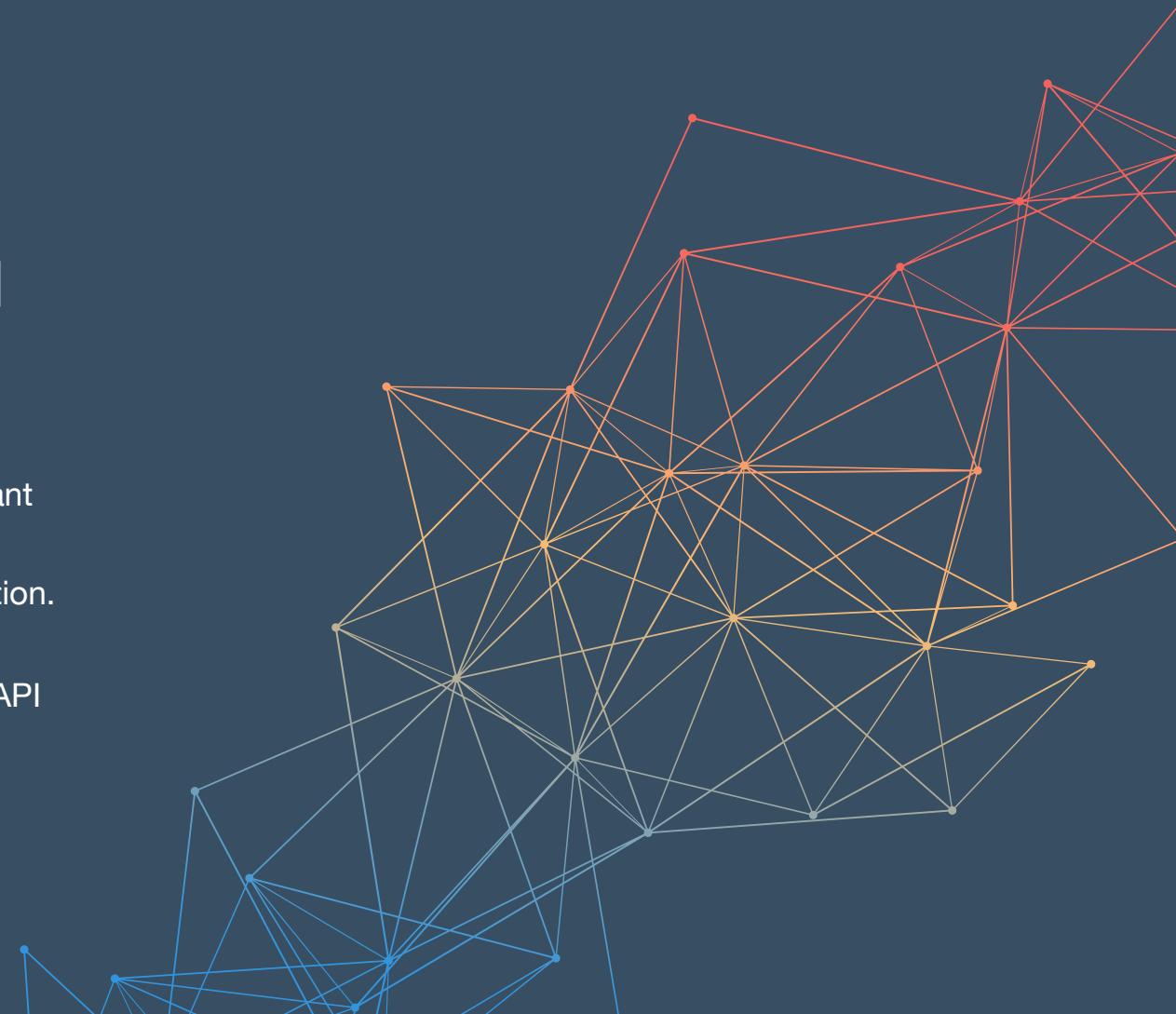
INTRODUCTORY NOTES

Someone looking to open a restaurant in New York City, where would you recommend that they open it?



For food security we will use the New York City Restaurant Inspection Data. This dataset provides restaurant inspections, violations, grades and adjudication information.

Using the information obtained through the Foursquare API we will collect the rating and the number of likes.



DATA ACQUISITION AND PREPARATION

We will collect our data from two sources:

1 - <https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j>

2 - Foursquare API



The first one contains the Restaurant fields:

Names

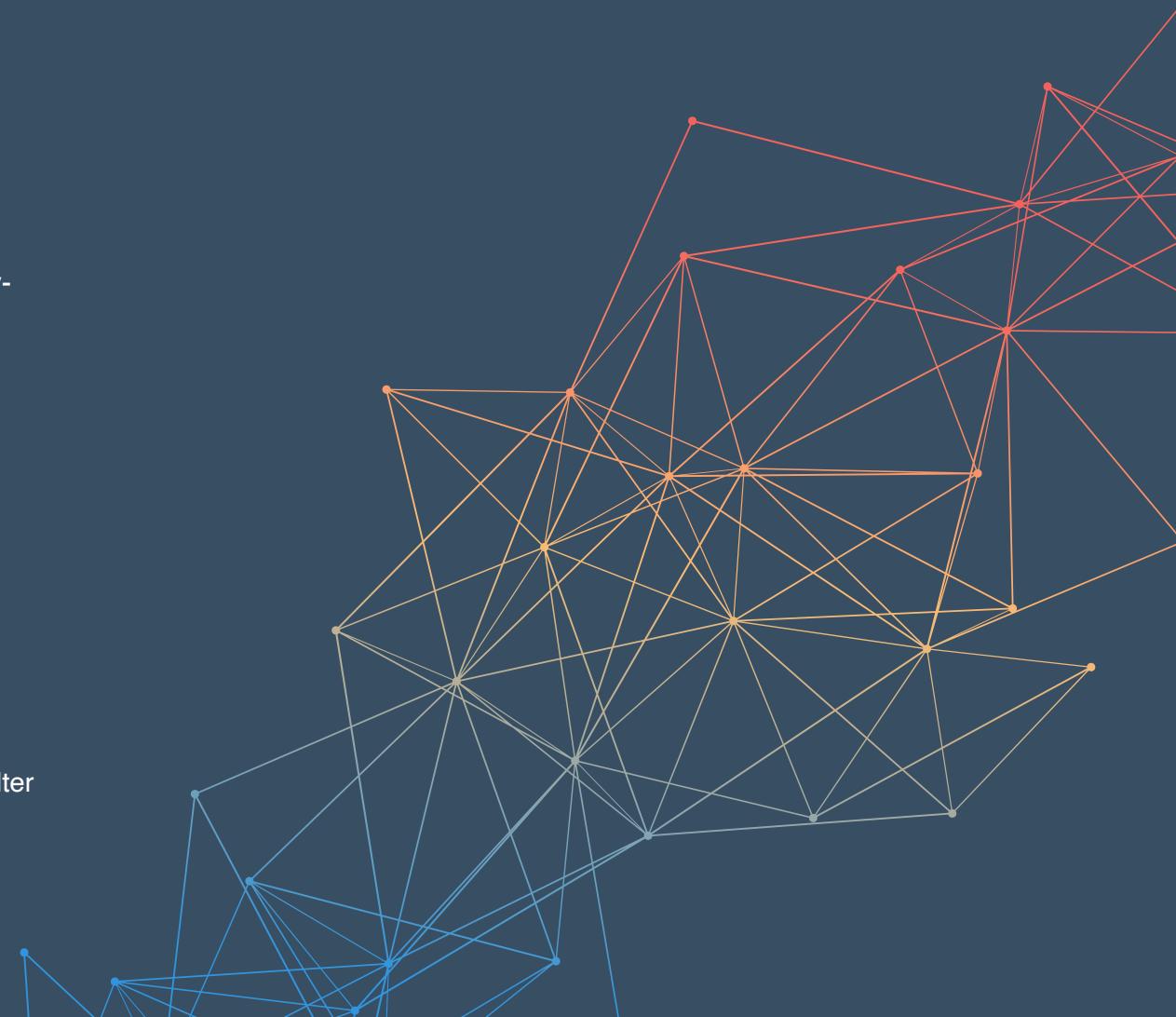
Location (latitude and longitude)

Grade of Inspection

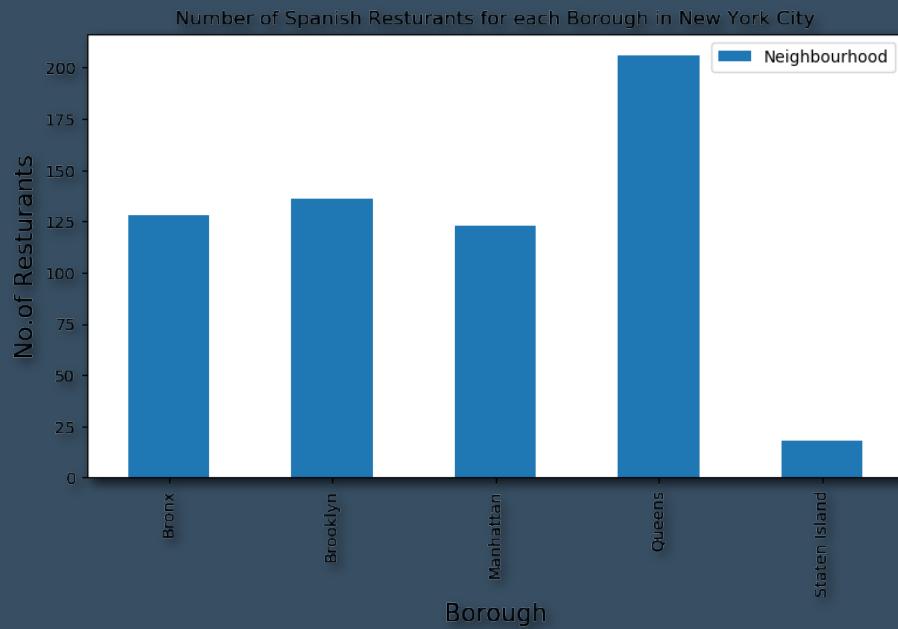
Cuisine

The second one contains the rating and number of likes for each Restaurants.

We will eliminate duplicate restaurants and obtain the number and filter the dataset per Spanish cuisine



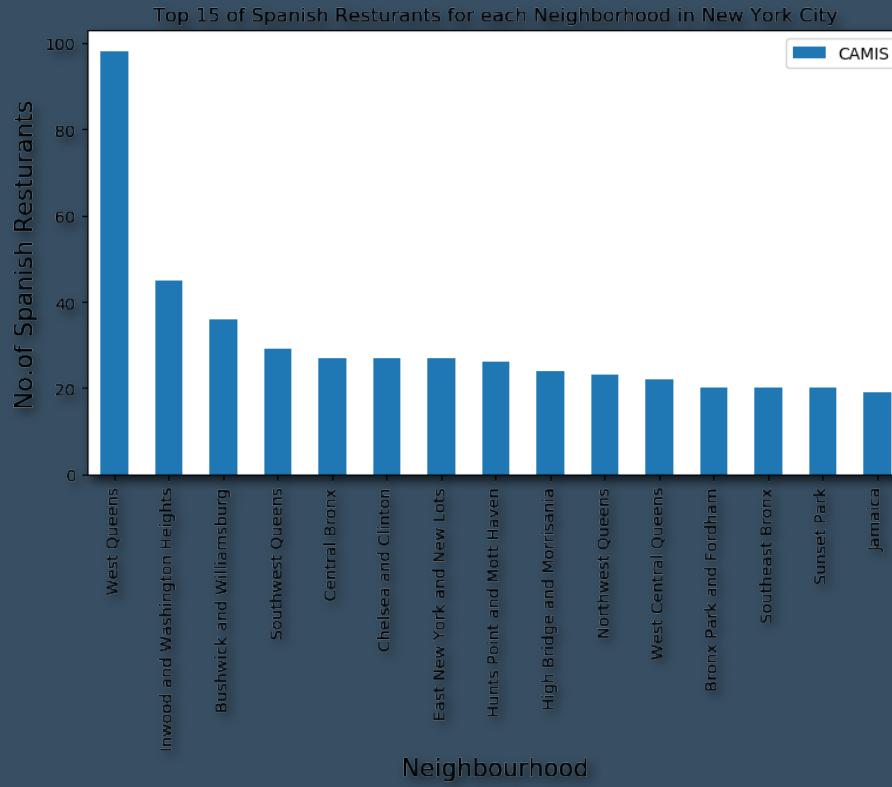
DATA ANALYSIS AND VISUALISATION



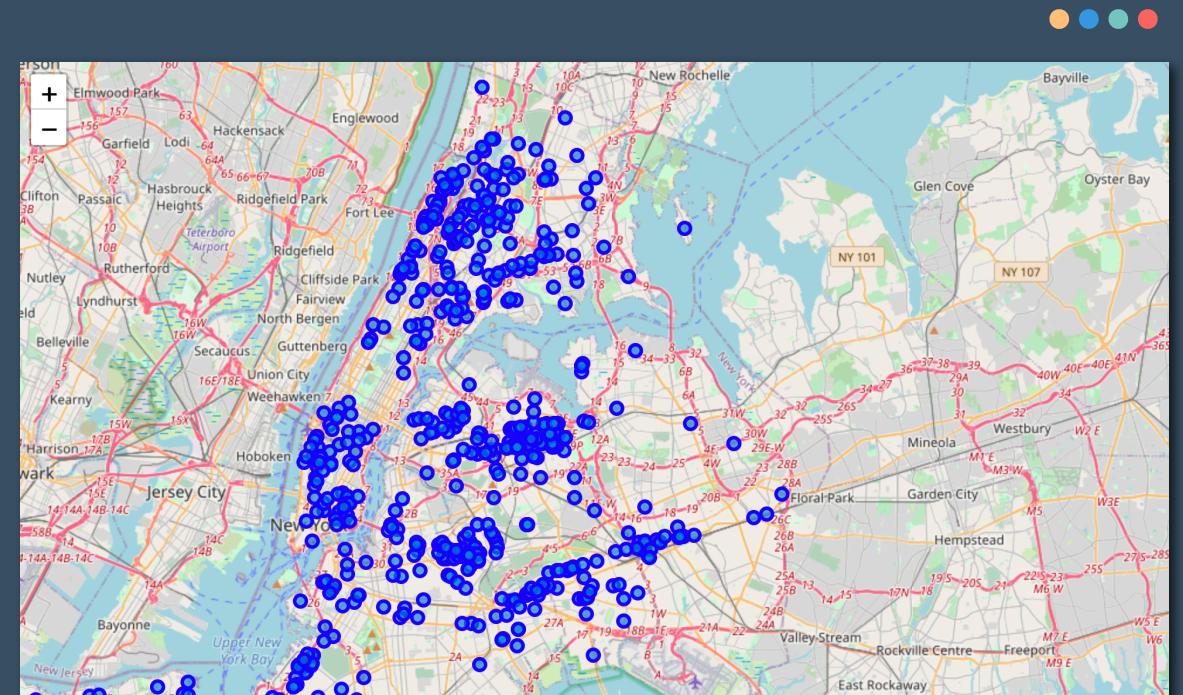
First of all we want to visualise how the Spanish Restaurants are distributed around NYC for each Borough.



DATA ANALYSIS AND VISUALISATION



We see that most restaurants are located in the **West Queens** Neighbourhood.



METHODOLOGY

K-means clustering



K-means clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided.

Data points are clustered based on feature similarity.

The results of the K-means clustering algorithm are:

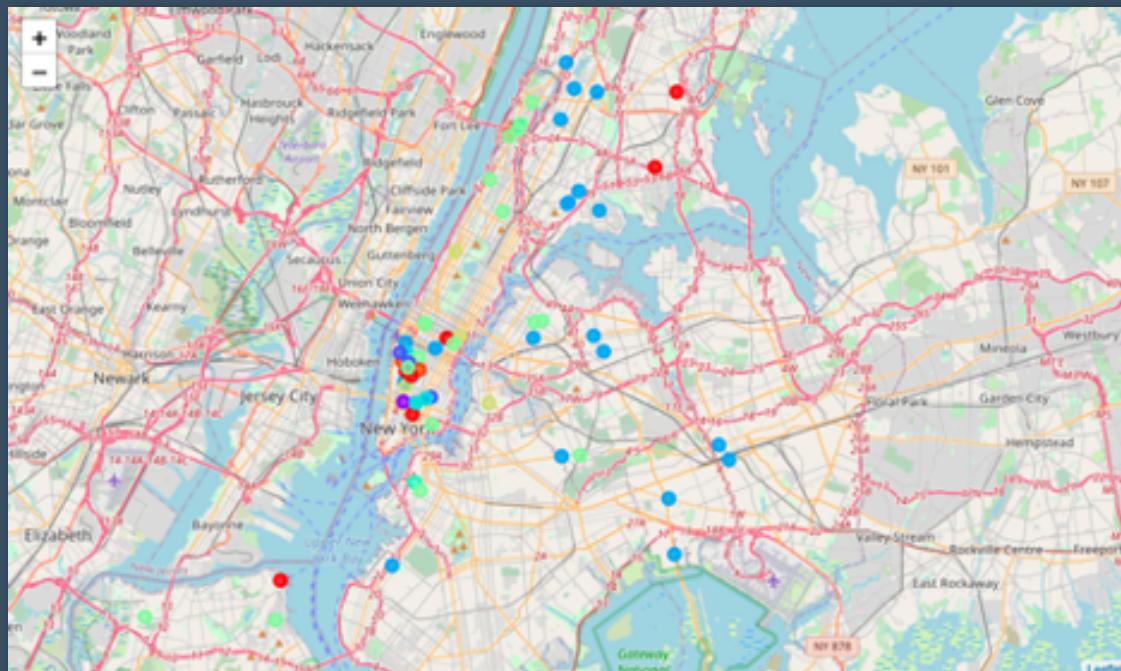
- The centroids of the K clusters, which can be used to label new data.
- Labels for the training data (each data point is assigned to a single cluster)

Rather than defining groups before looking at the data, clustering allows you to find and analyze the groups that have formed organically. The "Choosing K" section below describes how the number of groups can be determined.

I will use trendy recommendation filtering approach in order to make recommendations.

Trendy Recommendation: we use the rating and the number of likes obtained through the Foursquare API and the grade of inspections to clustering restaurants into 10 groups
Based on the result above, the third, sixth and seventh are the clusters with the worst results, so we created a new dataset with the remaining data.

METHODOLOGY



WE will use trendy recommendation filtering approach in order to make recommendations.

Borough.



Trendy Recommendation: we use the rating and the number of likes obtained through the Foursquare API and the grade of inspections to clustering restaurants into 10 groups

Based on the result above, the third, sixth and seventh are the clusters with the worst results, so we created a new dataset with the remaining data.

CONCLUSION

Chelsea and Clinton



We recommend our business partner to open the new restaurant in a trendy Neighbourhood with few Spanish restaurants open.

With that, we have concluded that the best recommendation will be neighbourhood Chelsea and Clinton with a higher Trendy Recommendation, lower competition and easy replication for business expansion.