1. Introduction.

The New York City metropolitan region is one of the most important economic regions in the world, as a center of many industries, including finance, international trade, news and traditional media, real estate, education, fashion, entertainment, tourism, biotecnology, law, and manufacturing.

One unique caracteristic of New York city is the diversity of ethnics groups that have entered the United States. So almost all ethnic cuisines are well represented, both within and outside the various ethnics neighborhoods.

Americans and People from all over the world love the culinary diversity available in the city. And one in particular that is very appreciated is the Korean cuisine.

Business Problem.

Based on the introduction presented above, what if someone wanted to open a new Korean restaurant? Where would you recommend that they open it?

Target Audience.

This report is very interesting for investors, real estate owners, entrepreneurs and professional chefs that want to run their own business.

2. Data Section.

Description of the Data.

The following data is required to solve the problem:

A data set , which is a json file (newyork_data.json), that contains a list of boroughs and neighborhoods of New York with geographical coordinates such as latitude and longitude.

A list of Korean Restaurants and their respective geographical coordinates in New York using Foursquare API.

How the data will be used to solve the problem.

The data will be used as follows:

- Load, explore and transform the data into a dataframe to obtain a list of Boroughs and Neighborhoods and their respective geographical coordinates.
- Use geopy library to get the latitude and longitude values of New York City.
- Create a map of New York with neighborhoods superimposed on top.
- Define a query to search for Korean Restaurant that is within 5000 meters from New York city.
- Create a dataframe for Korean Restaurant data only.
- Use machine learning (K-means clustering) to cluster Neighboords with Korean Restaurants.
- Visualize all clusters in a map using Folium.

This data analysis will facilitate one to decide where to open a new Korean Restaurant based on the distribution of the actual ones in New York city.