The Battle of Neighborhoods

Introduction and business problem

Introduction

- In 2020, we have Tokyo Olympic and expect that many visitors come from over the world. They would see the Olympic game and also may want to go sightseeing in Tokyo and other district. Even if Tokyo is well known city, most of visitor may not know how to go to place where they would go, which area is continence to commute, where is the area we should go for lunch etc. Knowing the Tokyo helps tourist to their visit more comfortable and easy.

Business Problem

- The Objective of this capstone project is to clustering the city of Tokyo, Japan to make the visitors easier to stay during Olympic period. Using the data science methodology and machine learning techniques to provide the useful data to foreigner and visitors in Tokyo.

Data

To get the insight, we will use the following data.

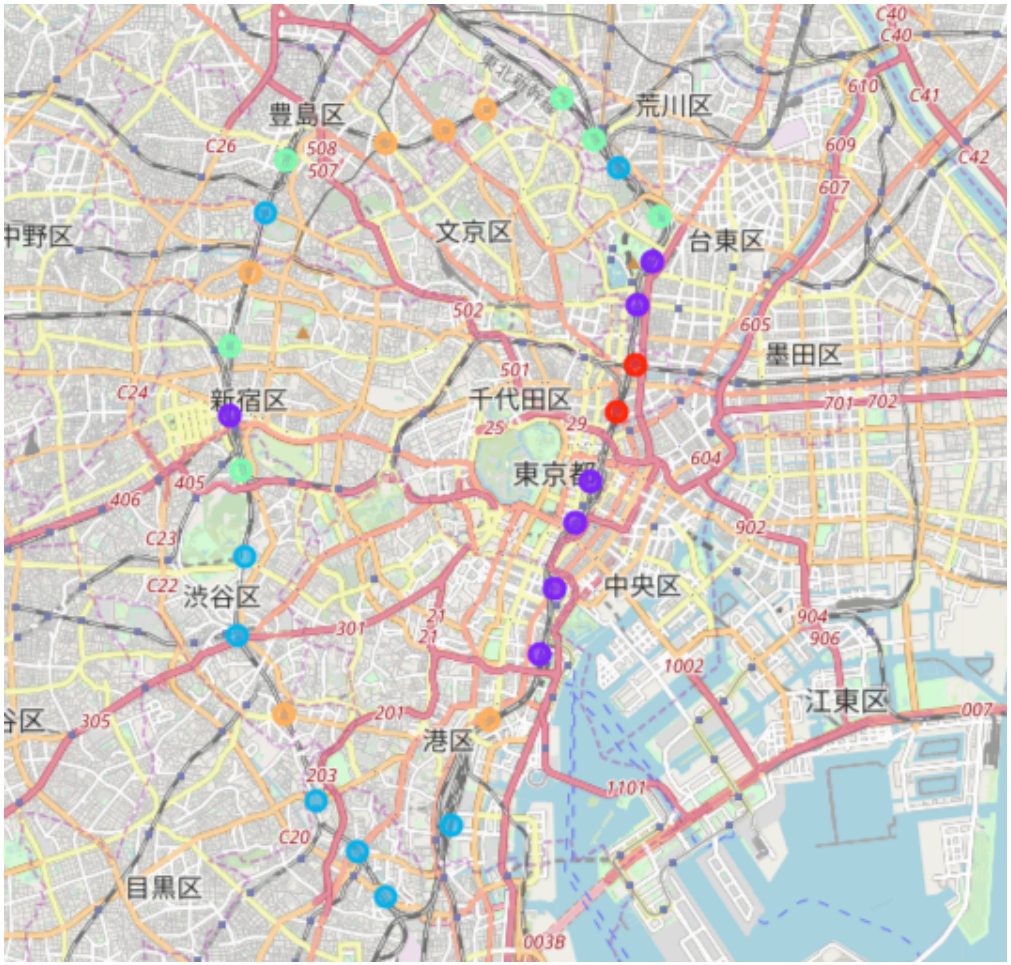
- List of neighborhoods in Tokyo.
- Latitude and Longitude coordinates of those neighborhoods.
- Location data that would be useful for tourist.

Methodology

- Read the Yamanote-Line station coordination and plot into the map to assure all station coordination is properly plotted.
- Using Foursquare API using own client ID and client secret key, collect relative information.
- Using One-Hot-Encoding, analyze the category nearby neighborhoods.
- Clustering in K-Means clustering
- Conjunction Stations data and Clustering data
- Visualize in map.

Plot the key station data in Tokyo

Clustering with Foursquare API



Red color stations have more hotel and accommodation than others.

And purples are second.

Based on my experience living in Tokyo, I thought we have more stations where we have many hotels and accommodation like red and purple.

If you come to Tokyo for Olympic, we better start booking the hotel. Otherwise, you may stay apart from Tokyo area.

Recommendation/Improvement

- In the foursquare API, we may not have latest information because foursquare is not well know in Japan. It will good if I could extract the data from Expedia, hotels.com etc which may have more accommodation info.

- Also if we could get the hotel occupancy rate, we could understand hotels are sufficient during the Olympic period.