Capstone Project - The Battle of the Neighborhoods (Week 2)

Applied Data Science Capstone by IBM/Coursera

Taku Sasaki

Table of contents

Introduction: Business Problem

Background

As COVID-19 has spread around the world for more than a year, vaccination is showing signs of convergence. However, there are still many unclear points about COVID-19 infection.

According to statistics, East Asians have a fairly low number of COVID-19 infections. It cannot be concluded that the cause has yet to be determined, whether it is due to genetic characteristics, many people already have immunity, or cultural differences. It is strange that the number of infected people is smaller than in the West in the big city of Tokyo, which has a murderous crowded train commuter.

People infected with COVID-19 vary from region to region. It cannot be said that there are many infected people because of the large population. The risk of COVID-19 infection varies depending on what kind of city you live in or stay in.

Note: For convenience, this document refers to the cities and wards of Tokyo as Borough.

Problem

- The problem is to answer the question of what kind of city is the one with many COVID-19 infections.
- There are various factors that cause a city with many infected people, such as many overseas travelers, many foreigners, and many bars where clusters are likely to occur. The purpose of this survey is to focus on the number of foreign residents and the number of bars to gain insight into the causal relationship with the number of people infected with COVID-19.
- Sample use case:
  - As a person planning to move to Tokyo, find out what kind of city has the same characteristics as a city with few infected people.
  - As a visitor planning a trip to Tokyo, I would like to find a safe city with the same characteristics as a city with few infected people and enjoy eating and drinking.

Data Requirements

- Based on definition of our problem, factors that will influence our decision are:
  - the number of foreign residents in Tokyo
  - the number of COVID-19 test positives in Tokyo
  - the number of existing bars in the neighborhood (any type of bar)

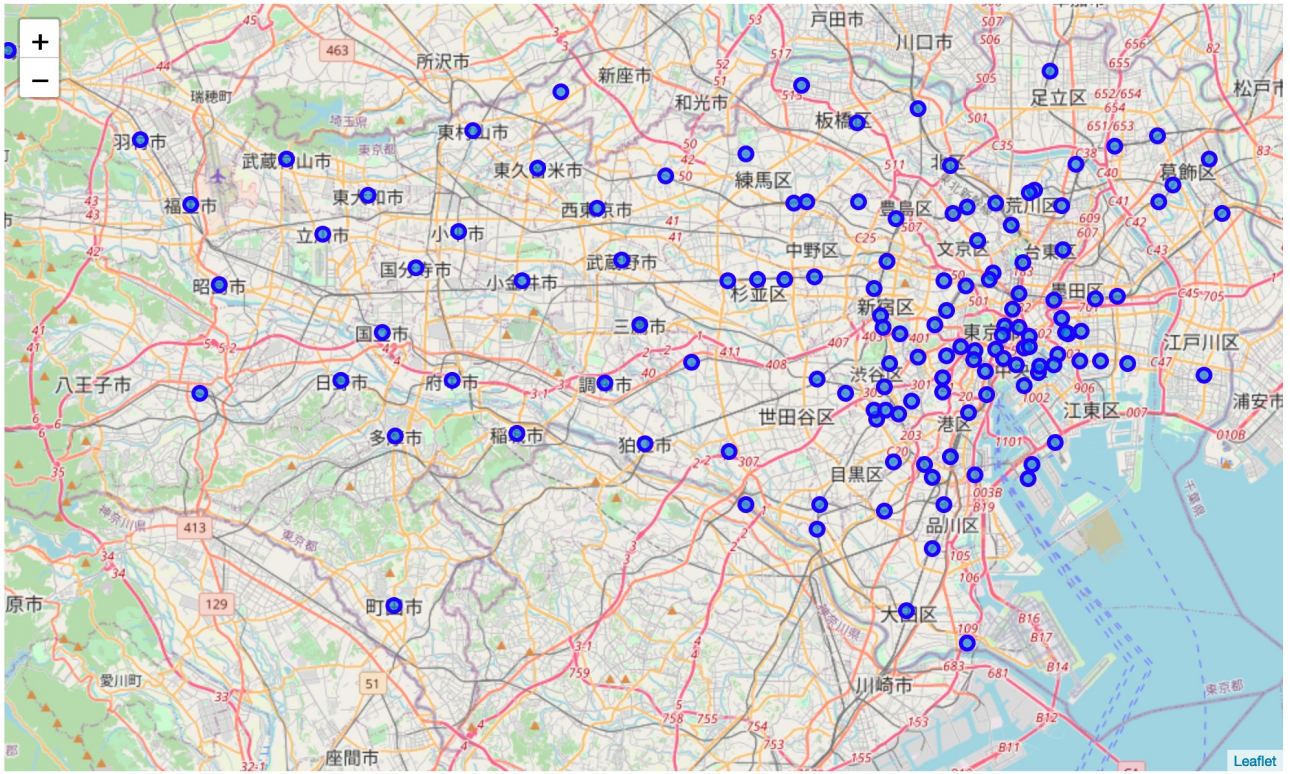
Data Collection

Following data sources will be needed to extract/generate the required information:

- The number of Foreign residents in Tokyo will be obtained from the following site:
  - Source: *"2-4 FOREIGN RESIDENTS BY DISTRICT AND NATIONALITY (2019)" in "TOKYO STATISTICAL YEARBOOK"*
    - by *Statistics Division, Bureau of General Affairs, Tokyo Metropolitan Government*
  - The data is as of 2019.
  - The data is aggregated by Borough, which means ward and city in Tokyo.
  - Foreign residents here mean foreign nationals who are registered according to the Basic Resident Registration Act.
- The number of COVID-19 test positive in Tokyo will be obtained from the following site:
  - Source:
    - *COVID-19 The information website by Tokyo Metropolitan Government*
    - *Tokyo COVID-19 Task Force website (https://github.com/tokyo-metropolitan-gov)*
  - The data is as of the day before yesterday.
  - The data is aggregated by Borough, which means ward and city in Tokyo.
- The list of Special wards and districts in Tokyo
  - Source:
    - *Wikipedia: Special wards of Tokyo*
- The number of bars and their type and location in every neighborhood will be obtained using **Foursquare API**

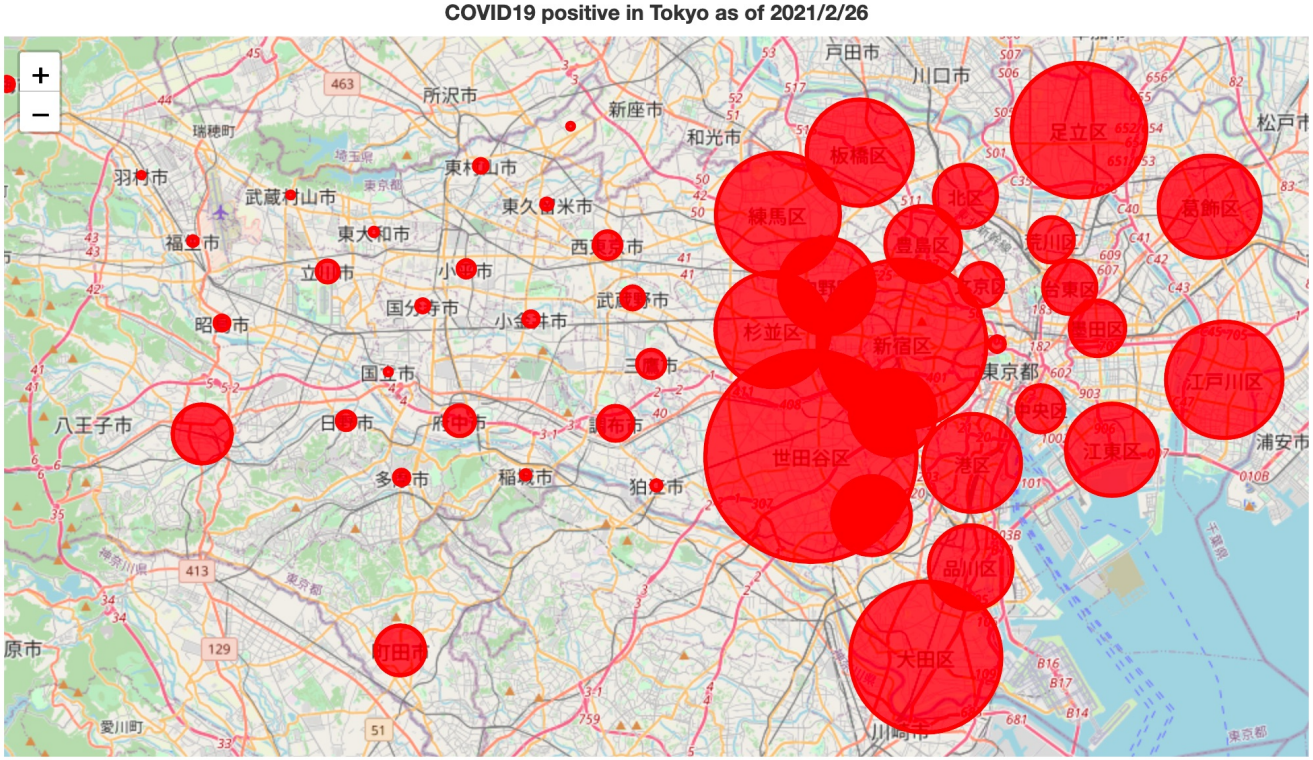
View neighborhoods in Tokyo

[93]:



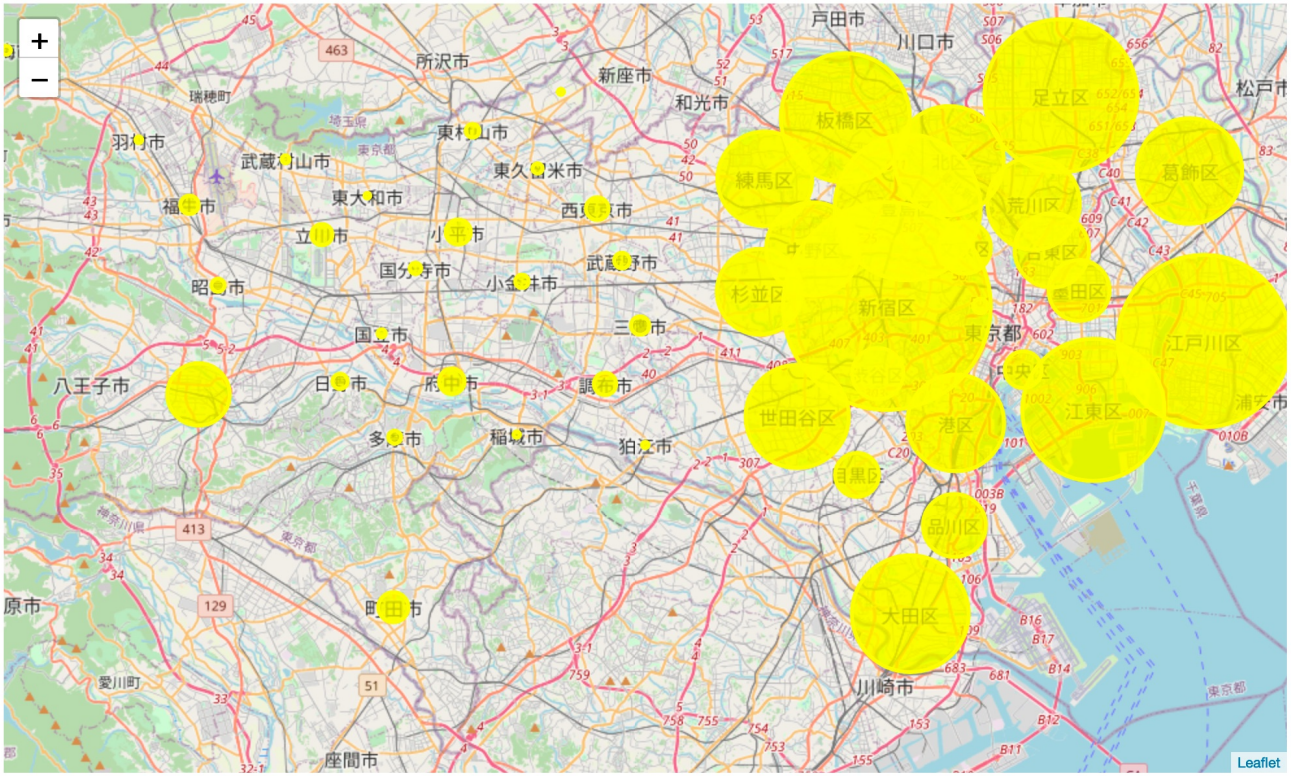
View COVID19 positive in Tokyo

[36]:



View Foreign residents in Tokyo

[37]:



Methodology

1. Analysis Neighborhood bars ¶

The purpose of this step is to investigate what kind of neighborhood belongs to the Borough.

- Use only tokyo_venues_bar
- Use **k-means** to cluster the Neighborhoods bar.
- Use KElbowVisualizer to determine the optimal k value
- Visualize the resulting Neighborhood bar clusters
- List each Neighborhood bar per cluster to investigate

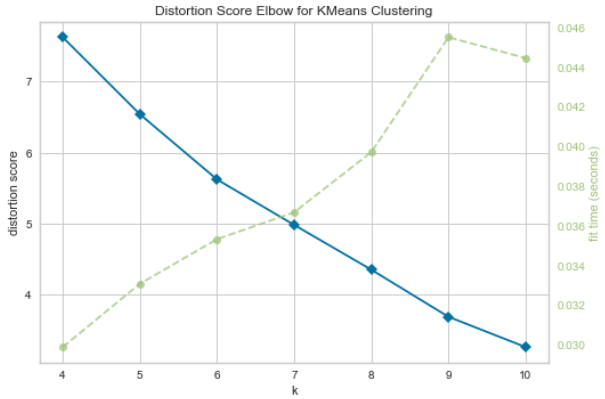
2. Analysis COVID positive, Foreign Residents and bars per Borough ¶

The purpose of this step is to investigate what Borough clusters are susceptible to COVID-19 infection.

- Aggregate each type of bar by Borough by using tokyo_venues_bar
- Merge it with covid_foreign
- Use sklearn.preprocessing.StandardScaler to normalize over the standard deviation
- Use **k-means** to cluster the Borough
- Use KElbowVisualizer to determine the optimal k value
- Visualize the resulting Borough clusters
- List each Borough per cluster to investigate

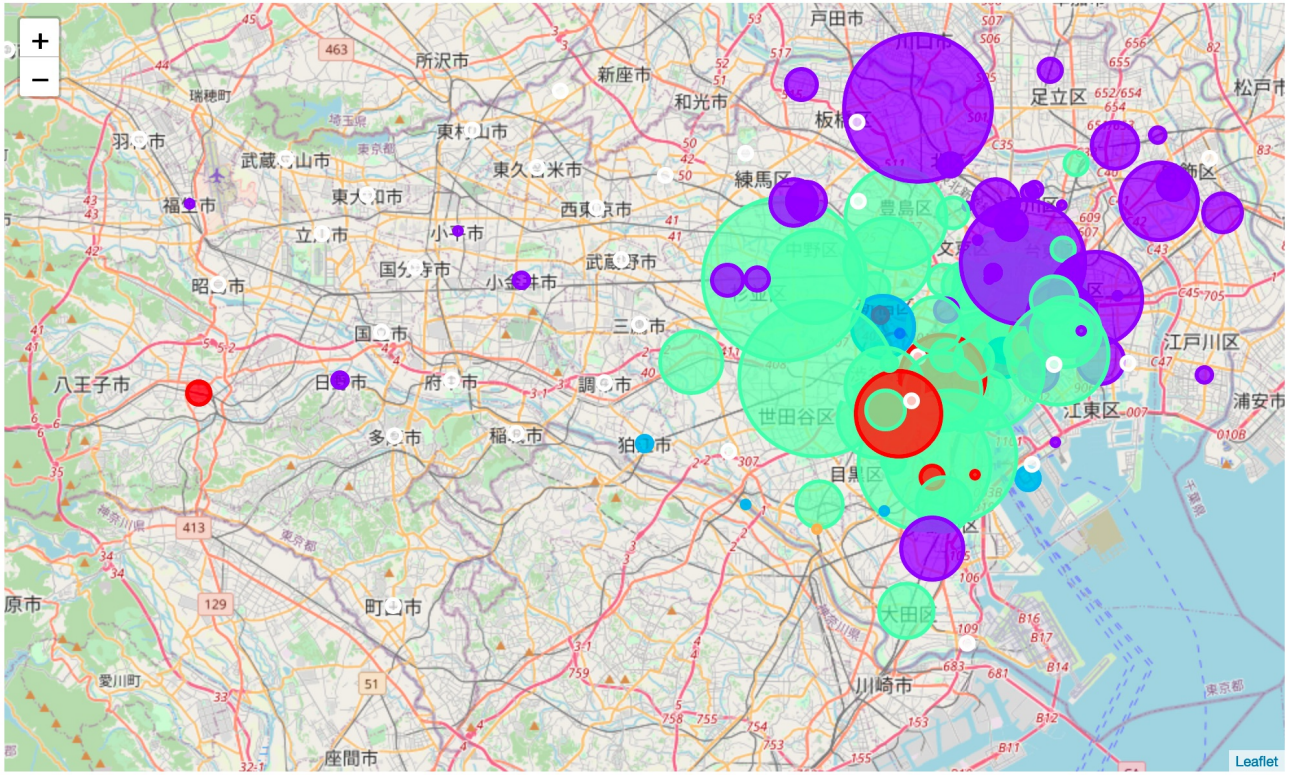
Analysis Cluster Neighborhoods

- So the first step is identify the best “K” using a famous analytical approach: the elbow method.



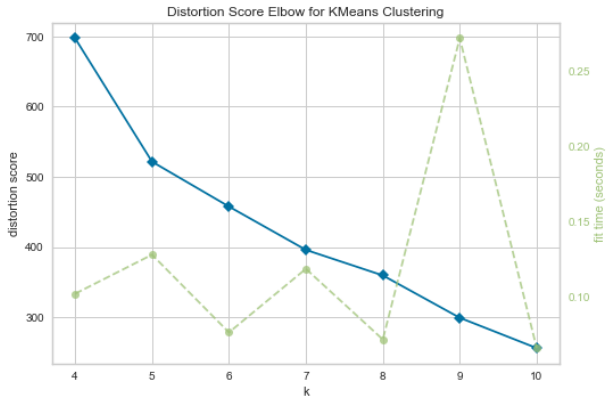
View Neighborhood clusters

[69]:



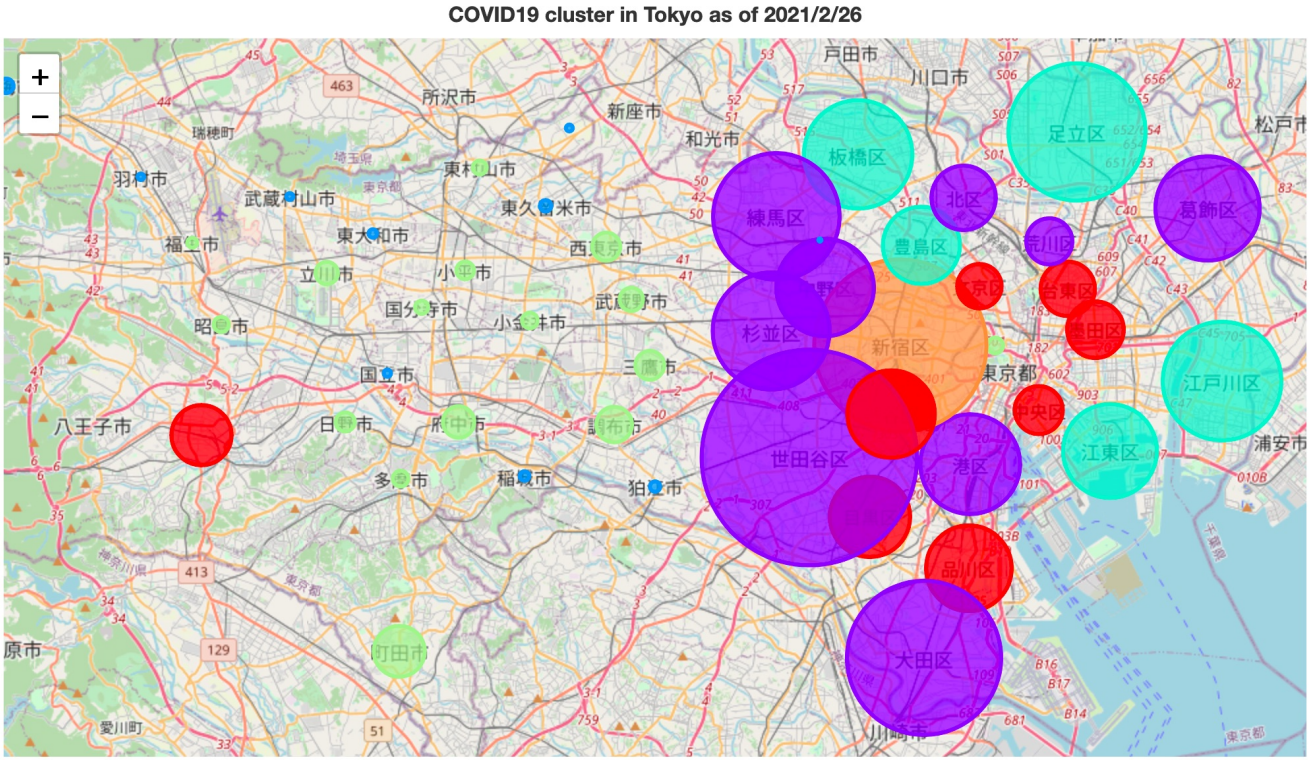
Analysis Cluster COVID positive, foreign residents and bars

- The first step is identify the best “K” using a famous analytical approach: the elbow method.



View COVID Clusters

[86]:



Results and Discussion (1/2)

- **1. Analysis Neighborhood bars ¶**
- Neighborhood cluster view is a cluster created only by the neighborhood bar type. It is possible to grasp the atmosphere of the city by the type of bar.
- Cluster 1:
  - Neighborhood has many expensive bars, and the number of bars is 11 or less.
- Cluster 2:
  - There are many sake bars in the neighborhood, which is often found in downtown, and the number of bars is 19 or less.
- Cluster 3:
  - Neighborhood has many expensive bars, and the number of bars is 8 or less.
- Cluster 4:
  - There are few neighborhoods in downtown and many in the city center, and there are many neighborhoods and bars.
- Cluster 5:
  - Luxury residential area

Results and Discussion (2/2)

- **2. Analysis COVID positive, Foreign Residents and bars per**
- Cluster 1:
  - The number of infected people is relatively small in the city center or in some exceptional suburbs.
- Cluster 2:
  - It is a cluster with Setagaya Ward, which has the highest number of infected people, and there are many bars and foreign residents.
- Cluster 3:
  - Suburban clusters with the fewest bars and foreign residents and the least infected
- Cluster 4:
  - Boroughs with many foreign residents, bars and many infected people in downtown
- Cluster 5:
  - Suburban clusters with few bars and foreign residents and few infected
- Cluster 6:
  - Borough only in Shinjuku Ward, there are many bars, the number of foreign residents is the largest, and the second most infected.

Conclusion

- The problem was to answer the question of what kind of city is a city with many COVID-19 infected people. It could be illustrated by clustering the features of Neighborhood according to the type of bar. Furthermore, Borough could be clustered and illustrated according to the number of foreign residents, the number of bars, and the number of COVID-19 infected persons. It can be speculated that Borough, which belongs to a cluster with a large number of infected people, may be susceptible to infection.

Future research

- The following can be considered as the continuation of future research.
  - The number of foreign residents is from 2019, so it cannot be said that it is valid data. If available, you should investigate with the latest data.
  - It is natural that there are few infected people in areas with a small population and areas with a low population density. The analysis should also take into account the population and population density.
  - In addition to foreign residents, the movement of immigrants and people from overseas should be considered as statistical values.