



Coursera Capstone

IBM Applied Data Science Capstone

MAKOTO Takamatsu



Introduction

Which areas of Tokyo and which businesses are most effective in attracting customers

The main difference between this city and others is that the area of the city is very large. Tokyo is made up of 23 regions. Therefore, it is interesting to note that Tokyo has a variety of regional characteristics. For example, Tokyo-Shibuya(渋谷区) and Shinjuku(新宿区) are known as a town for young people. Also, Chiyoda(千代田区) is the political center of the city and has the National Diet Building and other facilities; Minato(港区) is the center of IT and other businesses. In order to start a business that effectively attracts customers, it is important to first decide which region to start in and which business fits the character of that region.



Objective

- In this project, we study in details the area classification using Foursquare data and machine learning segmentation and clustering.
- The purpose of this project is to segment business trends in each district of Tokyo based on the most common business information obtained from Foursquare API.
- Using segmentation and clustering, we hope we can determine:
the similarity or dissimilarity of cities



Data

- **Restructure the data extracted from wikipedia into a csv file for easy manipulation and reading.**
- **Another aspect to consider in this project is the Foursquare data. That is, although we are using Foursquare data for segmentation and clustering, the amount and accuracy of the data captured cannot determine 100% of the correct classification in the real world.**

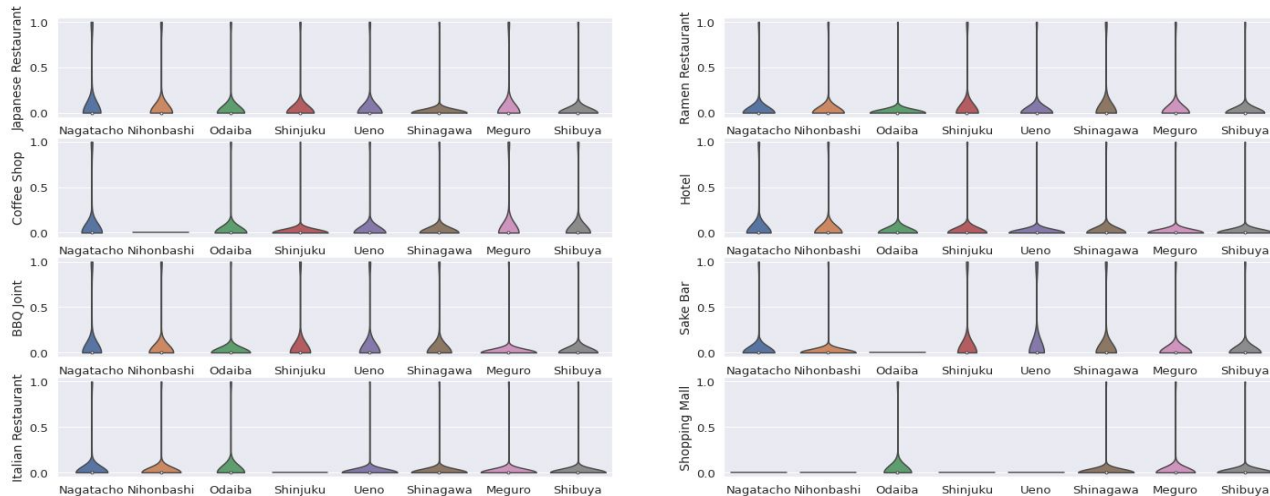


Methods

- 1. The value of the latitude and longitude of the main places of each region from the Wikipedia was scraping. Then use the Foursquare API to explore the neighborhoods for each region. After that, the search function to get the most common venue category of each neighborhood, to group close to the cluster using this feature.**
- 2. The K-means clustering algorithm is used to complete this task. The Folium library also visualizes each district and their business feature clusters.**

Result and Discussion

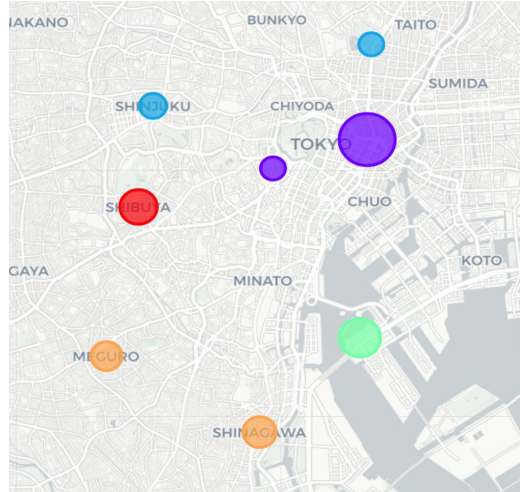
- To determine the characteristics of businesses by district, we examine the relative frequency of businesses of various genres from the Foursquare API.



The relative frequency of business in each region

Result and Discussion

- Cluster similarities or dissimilarities of businesses operating in various districts using K-means clustering algorithm.



Clustering of business trends



Conclusion

- To complete this project, we got a little glimpse of what a real-world data science project might look like. Similarly with real-world problems, we find solutions to those problems by analyzing big data and visualizing it in a visual way.
- We saw the results of scraping web data using the Python library, exploring the major districts of Tokyo using the Foursquare API, and dividing up the districts using the Folium leaflet map.