

# Capstone Project - Analysis of Busan, South Korea

By Chaesang Jung

<b>Introduction</b>	<b>1</b>
<b>Business Problem</b>	<b>1</b>
<b>Data Preparation</b>	<b>2</b>
<b>Methodology</b>	<b>3</b>
Data collection - Stations	3
Data collection - Venues	4
<b>Results</b>	<b>5</b>
Indian restaurants	5
Finding the number of clusters	5
Cluster analysis	6
<b>Discussion and Conclusion</b>	<b>7</b>

## Introduction

Busan(<https://en.wikipedia.org/wiki/Busan>) is the 2nd largest city in South Korea and it has been the role of the second capital of South Korea as it is located on the opposite side of Seoul. The population is more than 6 million and it's connected well with other countries with Gimhae international airport ( [https://en.wikipedia.org/wiki/Gimhae\\_International\\_Airport](https://en.wikipedia.org/wiki/Gimhae_International_Airport) ), so many people are visiting this city.

## Business Problem

The goal of this project is to understand the Busan metro city area thoroughly using subway stations, restaurants, and other venues. As the city is organized well with subway stations, we will compare areas accordingly.

The target audience will be tourists or the people who are new to this city especially from other countries. We will provide indirect information using venues grouped by the station for these questions.

- Where are the Japanese restaurants?
- Where can we find houses near Universities?
- Where can we find Indian venues?

## Data Preparation

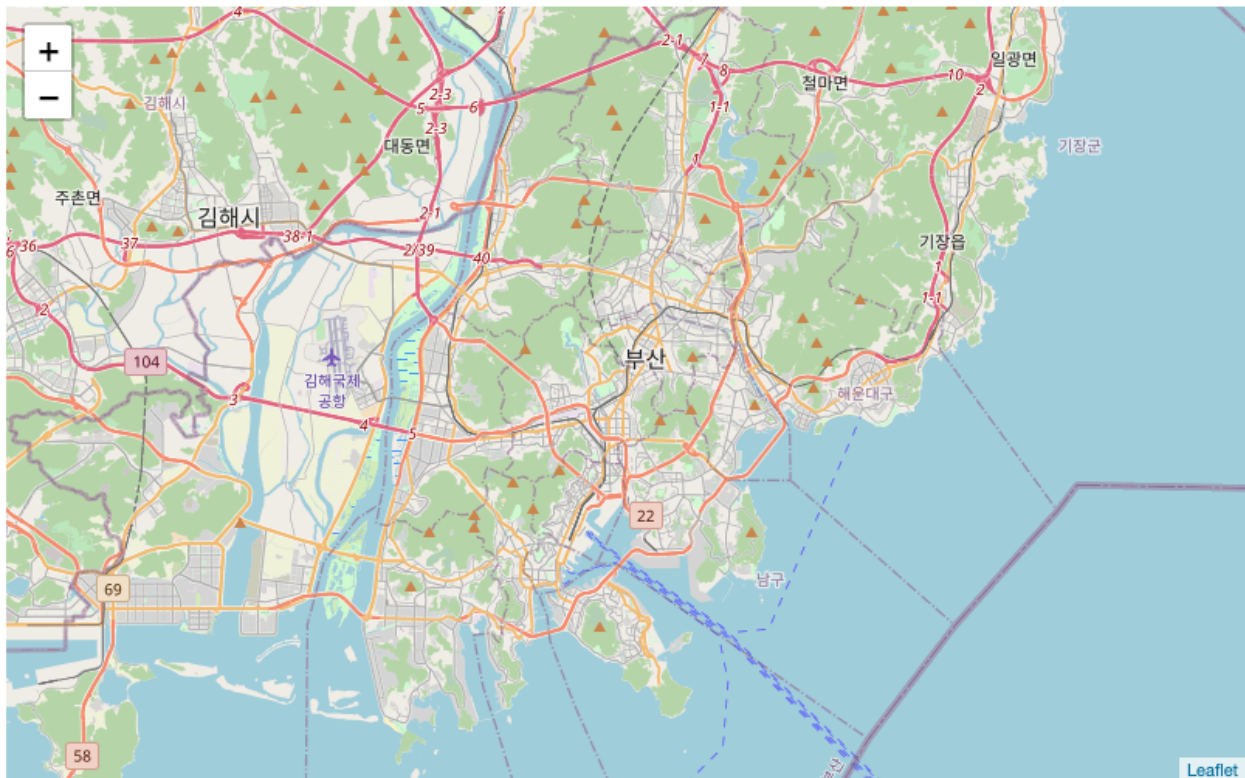
Busan is located at N35.166668, E129.066666

(<https://www.latlong.net/place/busan-south-korea-27604.html>). We will extract all the necessary data from **foursquare APIs** with proper category filters. Here are relevant categories from the Foursquare developer page

(<https://developer.foursquare.com/docs/build-with-foursquare/categories/>).

- Light Rail Station : 4bf58dd8d48988d1fd931735
- Metro Station : 4bf58dd8d48988d1fc931735
- College & University : 4d4b7105d754a06372d81259

Here is the base map of Busan.



For the first step, to get all stations and university locations, I used the maximum(30km) radius. Then we will treat each station as a separate neighbor and collect venues with a 1km radius. Finally, we will cluster stations, and try to find interesting things from a tourist's point of view.

## Methodology

### Data collection - Stations

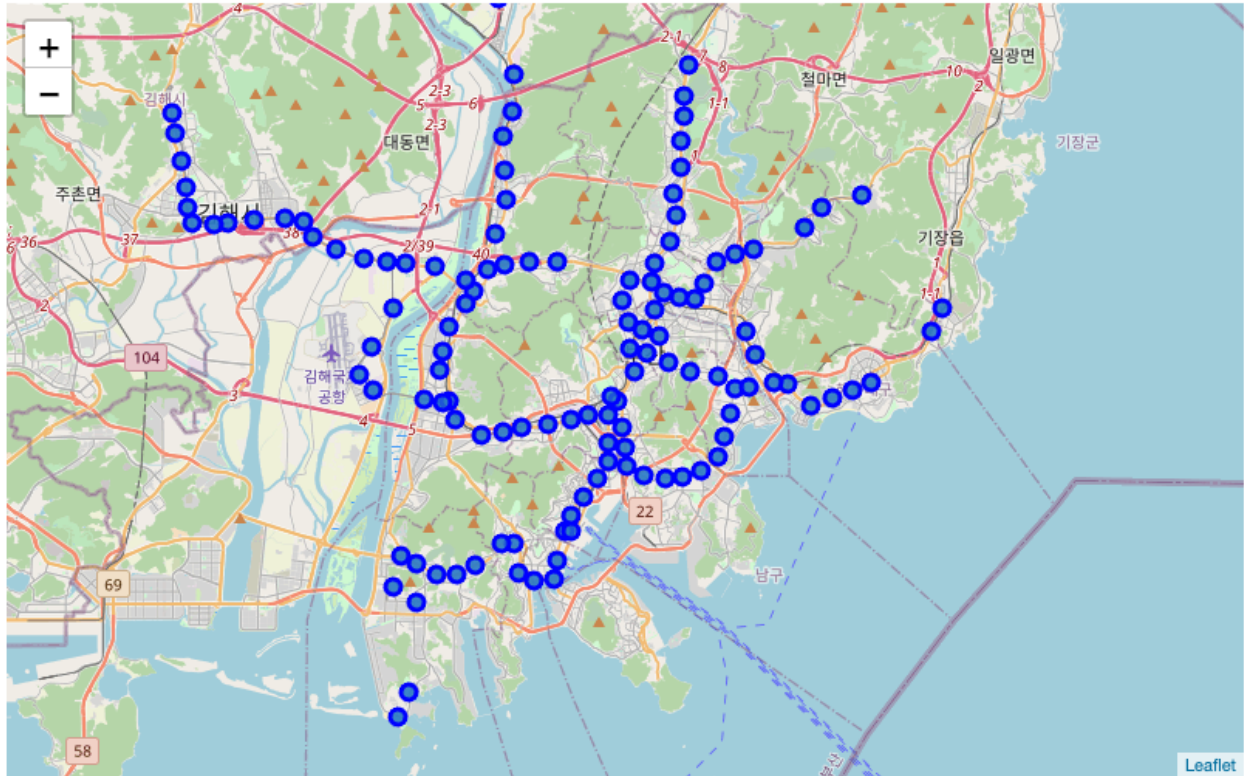
Starting from the center location of Busan, using a 30km radius, we could collect 100 metro stations and 33 Light rail stations as venues. Then we built the station's dataframe.

➤

	name	categories	lat	lng
0	Bujeon Stn. (부전역)	Metro Station	35.162565	129.062973
1	Seomyeon Stn. (서면역)	Metro Station	35.157718	129.059181
2	Yeonsan Stn. (연산역)	Metro Station	35.186107	129.081523
3	Yangjeong Stn. (양정역)	Metro Station	35.173009	129.071133
4	Sajik Stn. (사직역)	Metro Station	35.199044	129.065065
...	...	...	...	...
128	Royal Tomb of King Suro Stn. (수로왕릉역)	Light Rail Station	35.232799	128.872129
129	Gimhae Nat'l Museum Stn. (박물관역)	Light Rail Station	35.240002	128.871828
130	Yeonji Park Stn. (연지공원역)	Light Rail Station	35.249597	128.869291
131	Presbyterian Univ. Stn. (장신대역)	Light Rail Station	35.259436	128.867172
132	Kaya Univ. Stn. (가야대역)	Light Rail Station	35.266786	128.865176

133 rows × 4 columns

Here is the updated map of Busan with stations marked.



## Data collection - Venues

For each station, using a 2km radius, we collected the top 100 venues and analyzed the characteristics of each station. Here are samples of top venue categories of each station.

	Station	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Anpyeong Stn. (안평역)	Fast Food Restaurant	Korean Restaurant	Coffee Shop	Dessert Shop	Farmers Market
1	BEXCO Stn. (벡스코역)	Coffee Shop	Korean Restaurant	Hotel	Multiplex	Seafood Restaurant
2	Baesan Stn. (배산역)	Coffee Shop	Fast Food Restaurant	Seafood Restaurant	Wings Joint	Sushi Restaurant
3	Banyeo Agricultural Market Stn. (반여농산물시장역)	Farmers Market	Coffee Shop	Art Gallery	Wings Joint	Department Store
4	Beomeosa Stn. (범어사역)	Café	Coffee Shop	Moroccan Restaurant	Golf Course	Toll Plaza

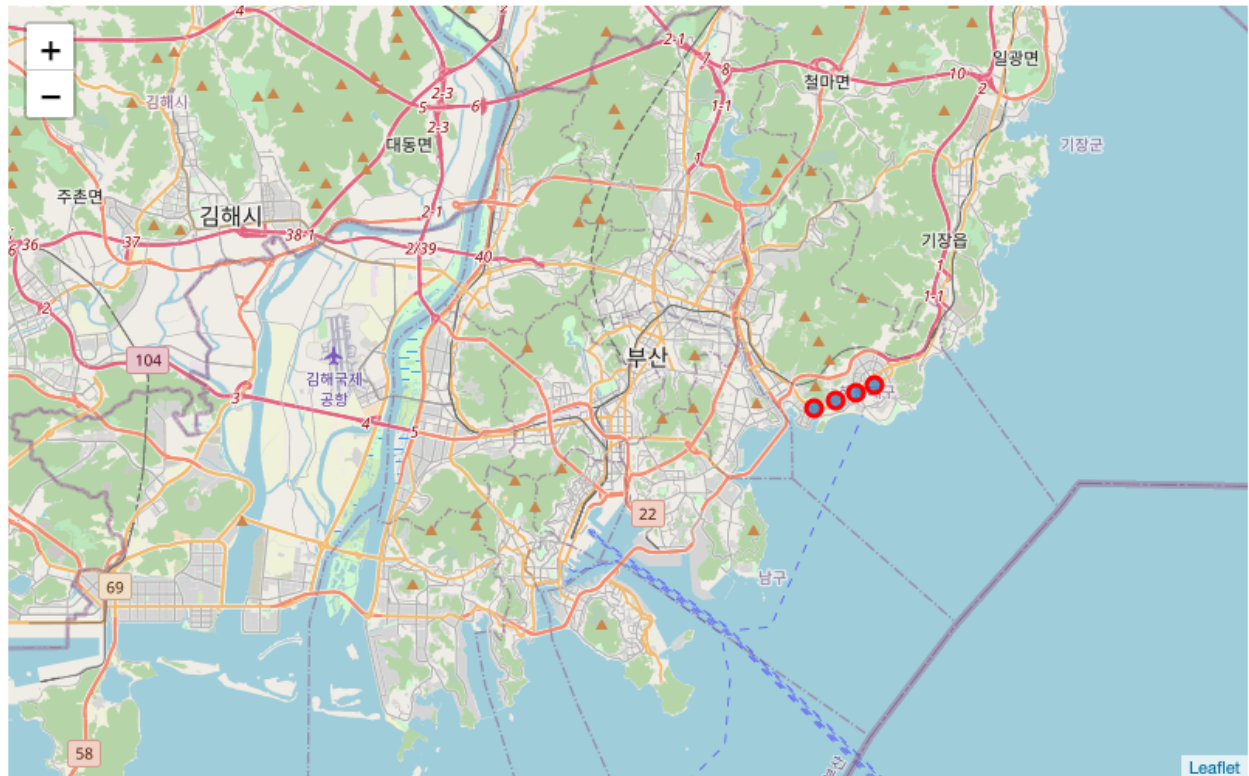
And here is the frequency of the most common venues of each station.

1st Most Common Venue	Station
Coffee Shop	69
Bakery	10
Korean Restaurant	9
Fast Food Restaurant	8
Café	7

## Results

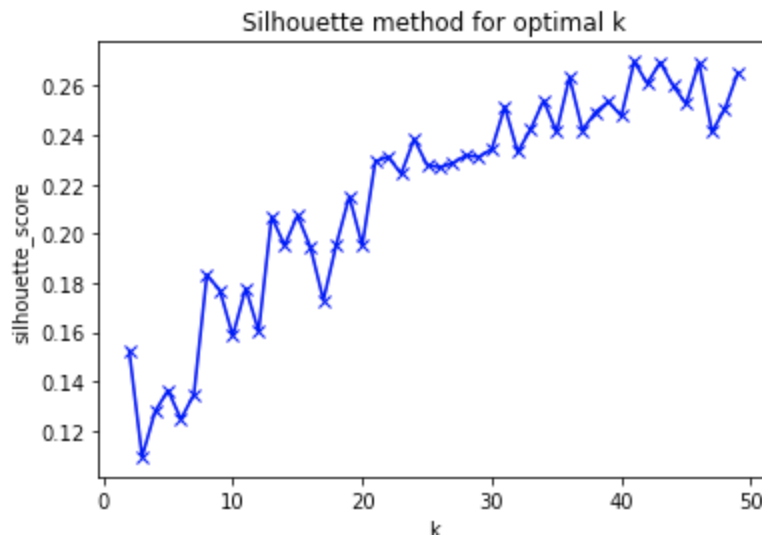
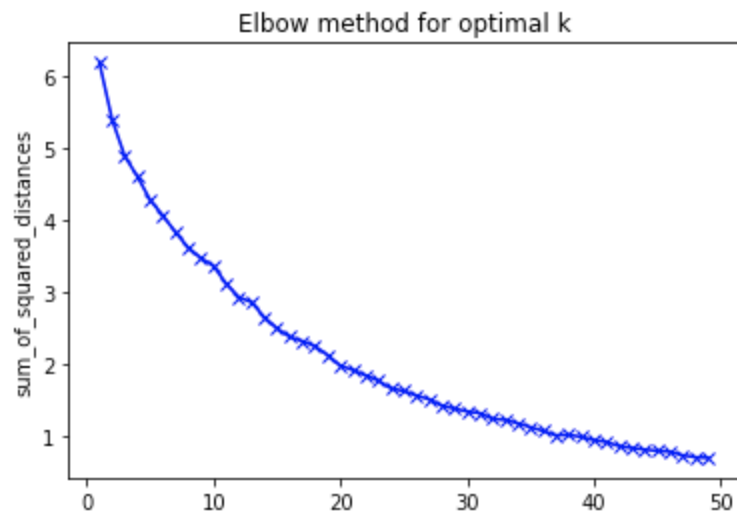
### Indian restaurants

We investigated Indian restaurants, but there are only 4 Indian restaurants are available in Foursquare data. This area is a new downtown near Haeundae beach.



### Finding the number of clusters

To find an optimal value for applying 'k-means' as an unsupervised machine learning algorithm, I applied 2 most popular methods, 'The Elbow Method' and 'The Silhouette Method'.

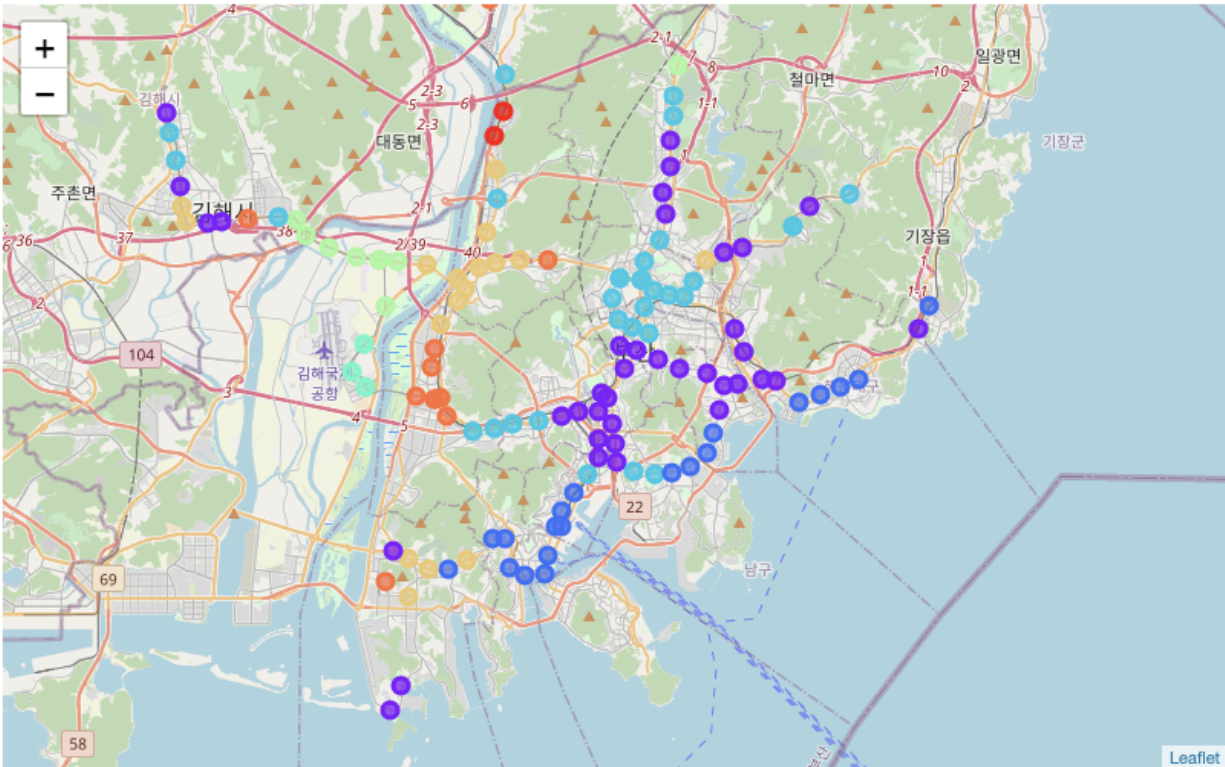


Near  $k=8 \sim k=10$ , there is a small elbow in the first graph and local peak can be found in  $k=8$  in the second graph, I chose  $k=8$  as an optimal value for number of clusters.

### Cluster analysis

With k-means clustering, all the stations are grouped into 8 clusters as below.





And here are short analysis about each cluster excluding outliers.

Cluster 1 (purple) : downtown

Cluster 2 (dark blue) : venues near beach

Cluster 3 (light blue) : residency area near downtown

Cluster 6 (yellow) : new town near river

Cluster 7 (orange) : airport / bus terminal

## Discussion and Conclusion

Through cluster analysis for the station, we can identify the area with similar venues. The area can be divided by clustering stations. 1) Near the beach area, 2) downtown 3) residency area, 4) new town near the river, and 5) airport area. This analysis can give a high-level overview of the Busan area to the people who will visit this city.

All the venues are collected from FourSquare developer API, so the collected data are biased with Foursquare's popularity at Busan area. Here are a few findings.

- The most popular venue types are cafe.
- Not all the Busan's stations are covered.
- College and university information is not available.

- Korean restaurants are the most common popular venue type, and there are also more specific venues e.g. Bossam/Jokbal restaurants, Bunsik restaurants, Gukbob restaurants, and so on. A similar thing happened for Japanese restaurants and Sushi restaurants.