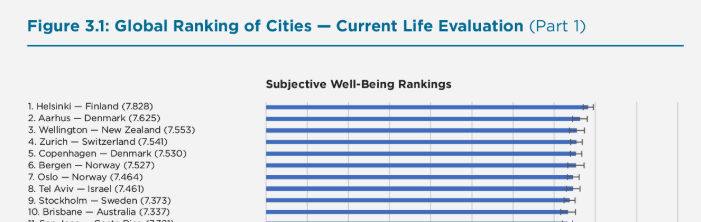
Comparison between northern and southern happiest cities

# Background and Problem

According to The World Happiness Report published on March 2020, the top 3 happiest cities in the world are Helsinki-Finland, Aarhus-Denmark and Wellington – New Zealand.



Source: <https://worldhappiness.report/ed/2020/cities-and-happiness-a-global-ranking-and-analysis/>

One interesting fact is that Helsinki is the 2ND northernmost capital of the World while Wellington is the southernmost capital. The purpose of this report is to find out more insights behind this result - through analyzing the sub-districts clustering of both cities what are the similarity and dissimilarity will be discovered between the happiest cities from the northern and southern countries in the world.

# Data and its sources

The data sources used in the projects are:

1. Sub districts of the city

Wikipedia pages reveal the sub districts of both cities. Sub-district is the basic object in this study from where venues nearby will be discovered.

url of Helsinki sub-district: https://fi.wikipedia.org/wiki/Helsingin\_alueellinen\_jako

url of Wellington sub-district: https://en.wikipedia.org/wiki/Category:Suburbs\_of\_Wellington\_City

1. Foursquare Places API

The developers’ version of Foursquare PLACE API is used to explore the venues nearby each neighborhood. The Places API offers real-time access to Foursquare’s global database of rich venue data.

# Methodology

## Data exploration

|  |  |  |
| --- | --- | --- |
|  | Helsinki | Wellington |
| Population | 650,058 | 215,400 |
| Area | 715.48 km2 | 442 km2 |
| Number of sub districts | 59 | 55 |
| Number of Unique categories | 252 | 159 |
| Number of venues | 1345 | 564 |

While exploring the data, it is observed that Helsinki is 200 km2 bigger area than Wellington, with 3 times more population. The number of the sub-districts are almost same for both cities, while number of unique category – the type of venues in Helsinki is 252, more than 159 in Wellington. In addition, the total number of the venues discovered by Foursquare is almost 3 times more in Helsinki than Wellington.

Based on the figures described above, we can see that Helsinki is a bit bigger than Wellington from both population wise and size wise. The number of the sub-districts of both cities are almost same, however Helsinki has more venues than Wellington, not only from the point of view of number, but also from the type’s perspective.

## Machine learning Algorithm – K means clustering

Next, we will use machine learning algorithm – K means clustering to cluster the sub-districts of both cities.

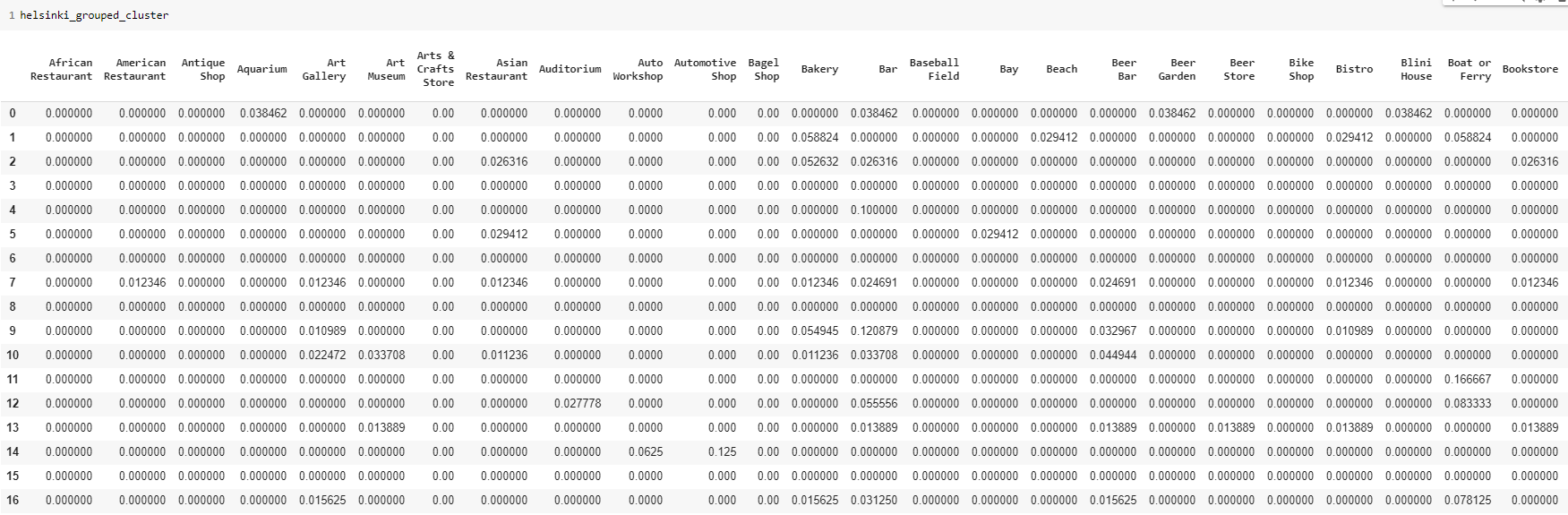
K means clustering is a simple but popular unsupervised machine learning algorithm, which can be simply described as group similar data points together and discover underlying patterns.

### Feature selection

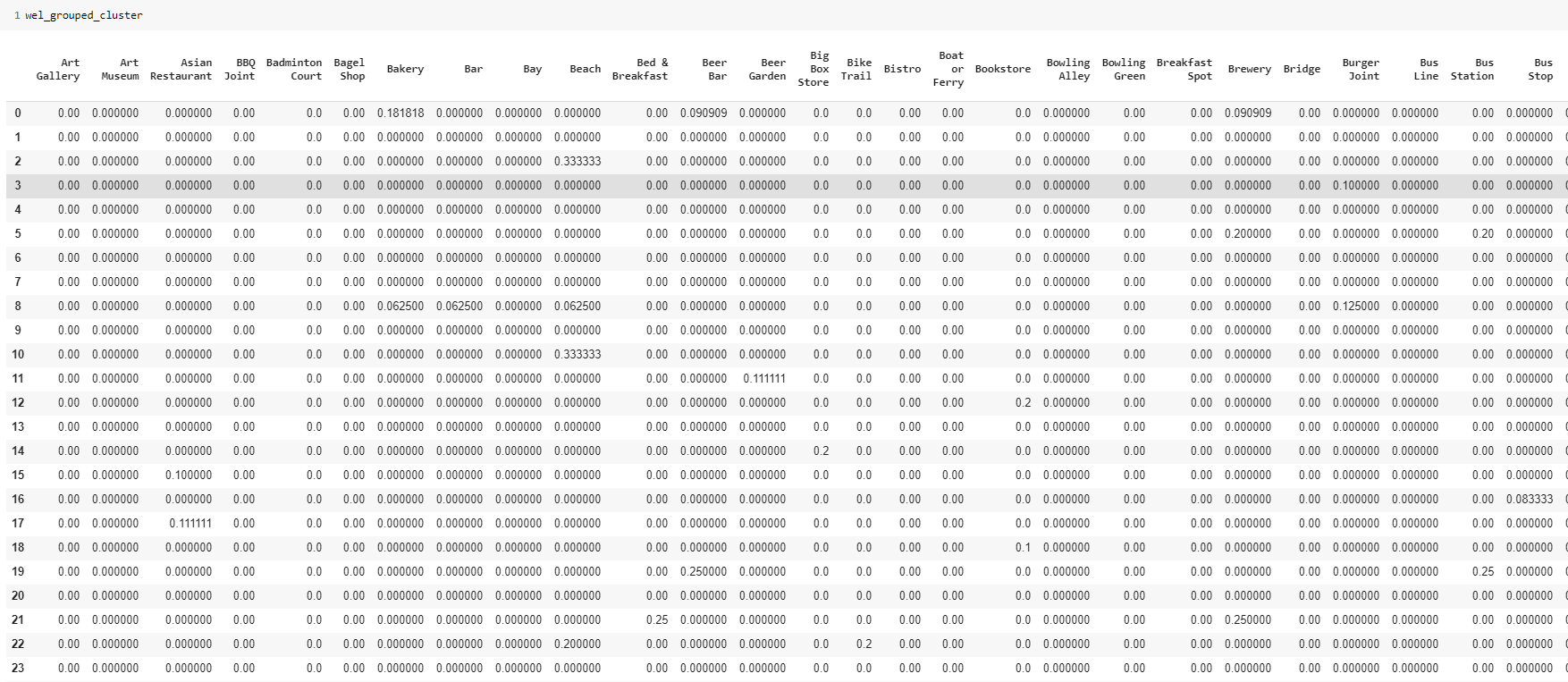
The feature of the data set is decided to use the frequency of the particular type of the venue in the particular sub-strict. As a result, we want to utilize the algorithm to cluster the sub-districts with similar setup of venues. E.g. cluster 1 concerns the sub districts with most stores and restaurants, and cluster 2 covers the sub-districts with more natural scenery e.g. parks and beaches.

During the data preprocessing, one hot encoding approach is used to convert the categorical data – the venue category to numerical value, then the frequency of the category type is calculated accordingly.

Final Feature- Helsinki



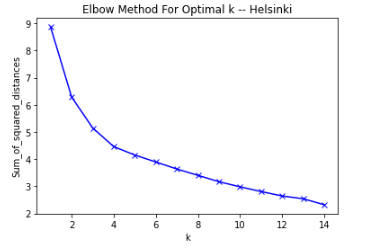
Final Feature - Wellington



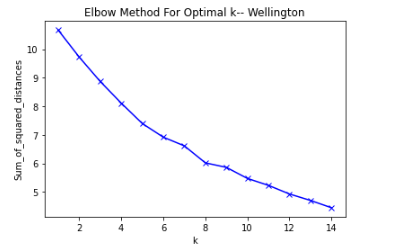
### Optimal K

A very important task in K means clustering algorithm is to find the optimal K, meaning how many clusters we shall divide the data points. The K cannot be too big, but shall still nicely identify the most significant different patterns among the data points. Here we use elbow method for the optimal K selection.

Optimal K choice:



K=4 is selected for Helsinki.



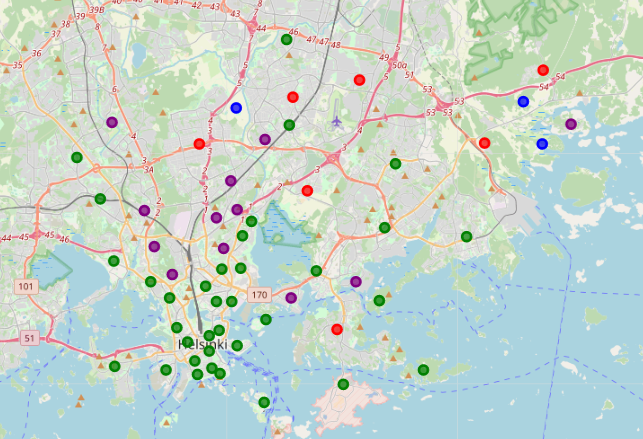
K = 6 is selected for Wellington.

# Results

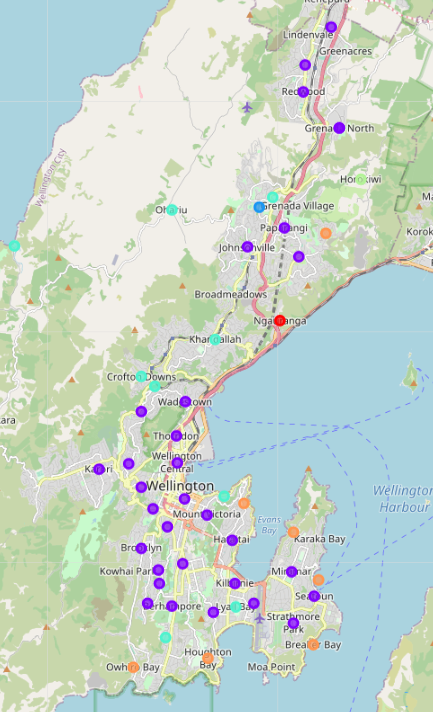
## Clustering

Helsinki sub-districts have been clustered into 4 groups.

|  |  |  |  |
| --- | --- | --- | --- |
| HELSINKI | Number of sub-districts | Color in the map | Percentage |
| Cluster 1 | 12 | purple | 21% |
| Cluster 2 | 3 | blue | 5% |
| Cluster 3 | 7 | red | 13% |
| Cluster 4 | 34 | green | 61% |



|  |  |  |  |
| --- | --- | --- | --- |
| WELLINGTON | Number of sub-districts | Color in map | percentage |
| Cluster 1 | 1 | red | 2% |
| Cluster 2 | 33 | purple | 63% |
| Cluster 3 | 1 | blue | 2% |
| Cluster 4 | 9 | Light blue | 17% |
| Cluster 5 | 1 | Light green | 2% |
| Cluster 6 | 7 | orange | 13% |



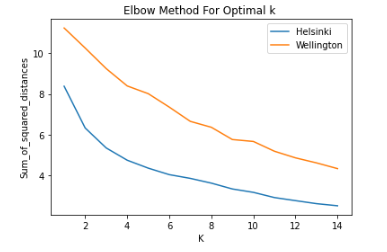
# Discussion

## Similarity

Both cities have similar number of sub-districts, and 60% of which are clustered into one group. This biggest group of sub-districts are near the city center of both cities, where full of restaurants and stores. In the rest sub-districts, two bigger groups are formed in both cities which shares around 15% among all. The remaining clusters concerns less than 3 areas each. As a result, the allocation of the cluster distribution is very similar in both cities.

## Difference

When calculating the optimal K, it is discovered that sub-districts are much more different with each other in Wellington than Helsinki. According to the figure below, to achieve the same level of squared distances between data points Helsinki needs 4 clusters while Wellington needs 14!



This reveals sub-districts are more generalized in Helsinki, while more diversities are appearing in Wellington.

# Conclusion

With this project, we analyze the sub-districts distribution and clustering in the most northern and southern happiest cities of the world, Helsinki and Wellington. By using the K means clustering algorithm, we noticed the sub-district clustering are similar in both cities with the allocation of 60% areas grouped in cluster 1, 15% areas in cluster 2, 15% in cluster 3 and rest. However inside each cluster, the diversity and difference is much bigger in Wellington than Helsinki.