**Capstone Project - The Factors of World Happiness**

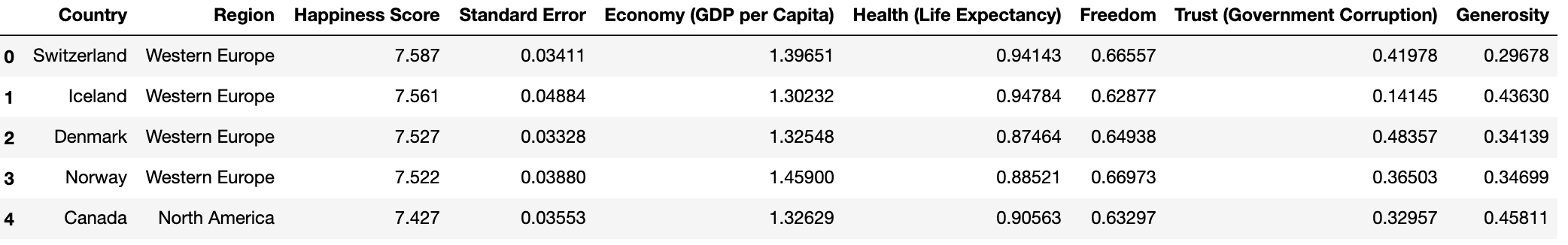
**1. Introduction**

Happiness is closely related to the quality of life and life satisfaction of people. As a citizen of Hong Kong, a city with the longest longevity in the world and good economy, it ranks low in liveability and happiness though. I found the relationships between are quite interesting, and therefore in this project, I would like to use metrics to investigate different factors and their importance to the happiness level. By identifying the significant factors through data analysis, countries and their governors can reflect on their well-being progress for improvements, as well as to create a better environment for its residents to live a happy and quality life.

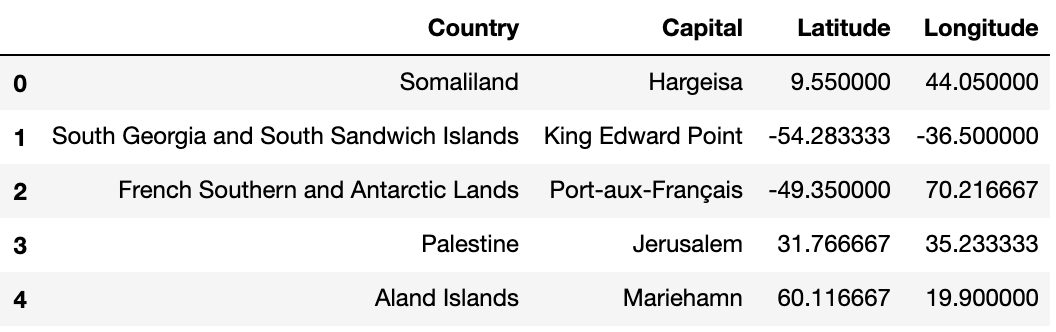
**2. Data acquisition and methodology**

There are two datasets used in this project: 1) 2015 World Happiness Index, and 2) World Capitals GPS. All the data sources are downloaded from Kaggle.

1) 2015 World Happiness Index



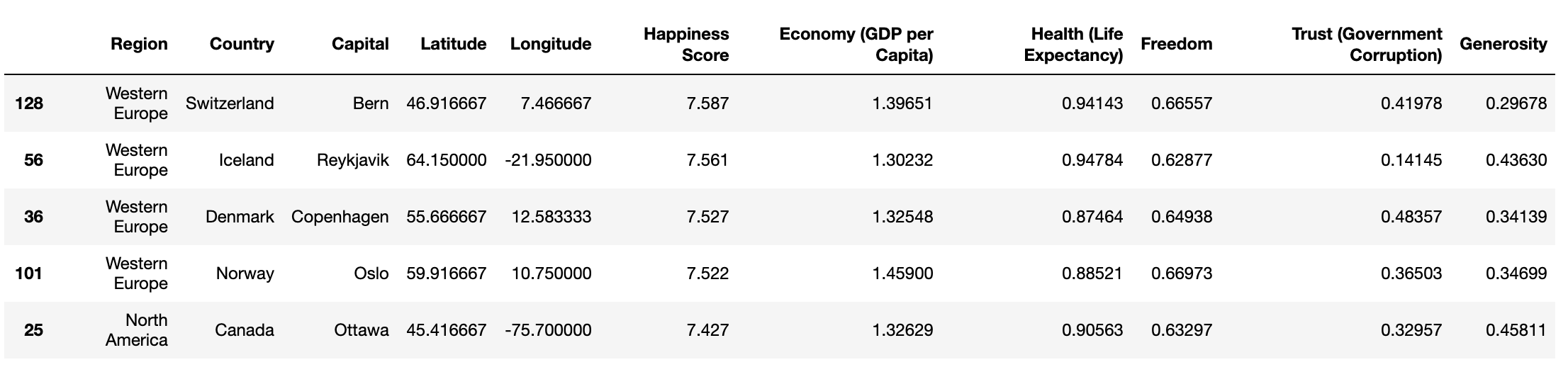
2) World Capitals GPS: It contains location data, especially latitude and longitude, of the capital of 245 countries for creating map visualization.



After cleaning the duplicated entries and missing values, a master data frame was created for further analysis by merging the datasets.

**Feature selection**

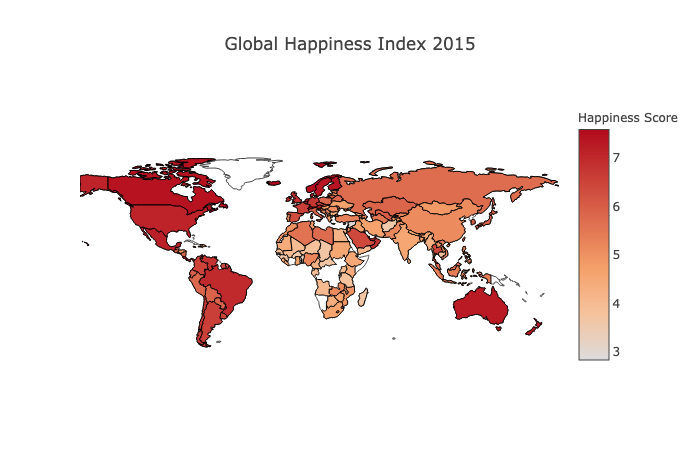
There are 152 records in the master data frame. However, some of the features are dropped because of their low relevance to this project and high correlation. Afterall, 11 features are selected: *Region, Country, Capital, Latitude, Longitude, Happiness score, Economy (GDP per Capita), Health (Life Expectancy), Freedom, Trust (Government Corruption), and Generosity*.



|  |  |
| --- | --- |
| *Variable correlation:* | *Happiness score distribution:* |

**3. Results**

**3.1 Evaluating world happiness score**

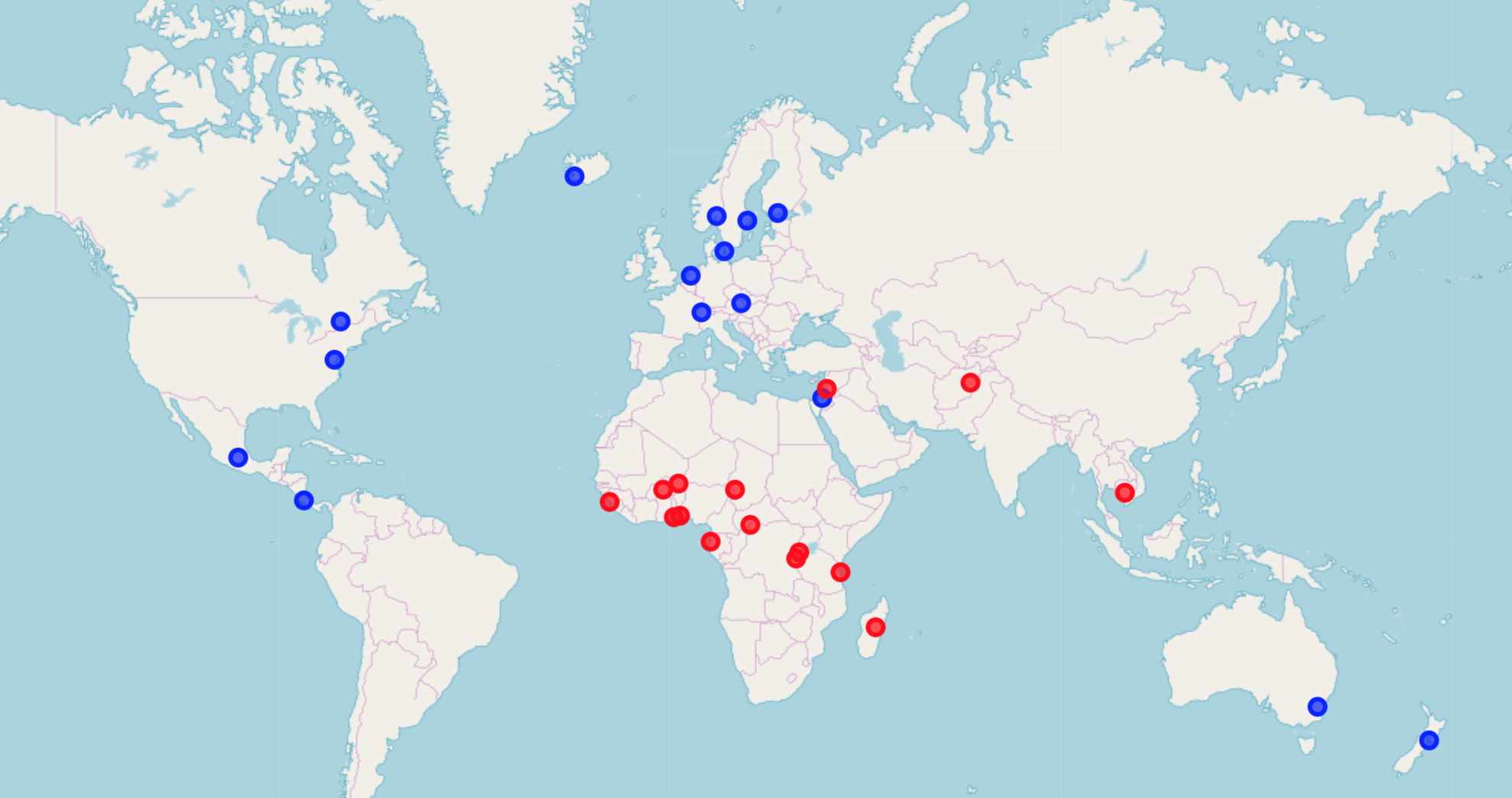


*(Global Happiness Index 2015)*

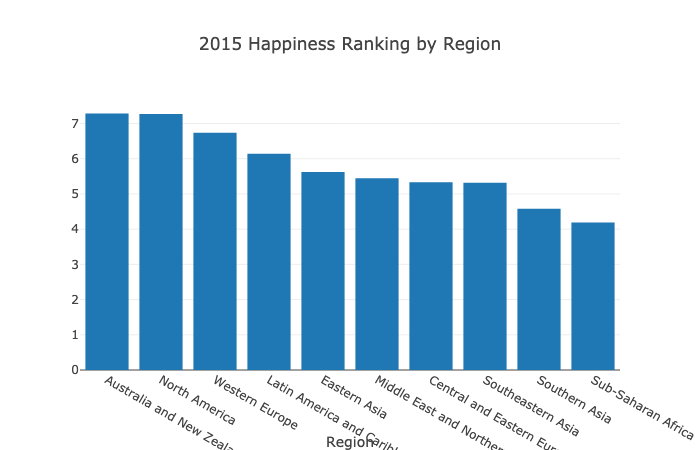
Assuming the mean happiness score (5.41) is the standard. In the dataset, more than half of the countries (52.63%, 80 countries) have a happiness score lower than the mean. Yet the boxplot distribution shows that the happiness level is still within the acceptable range in general.

**3.2 Relationships between happiness and region**

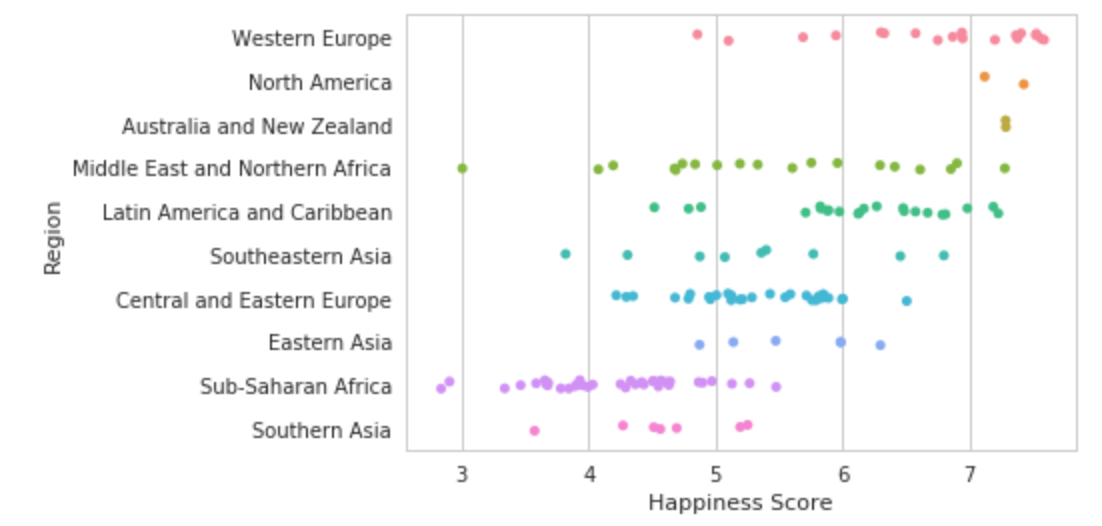
Here is a map showing the distribution of countries with the highest 10% (blue marker) and lowest 10% (red marker) regarding their happiness score. Most of the blue markers and red markers are concentrating in the Western Europe and Sub-Saharan Africa respectively.



The bar chart displays the average happiness score by region, ranking in descending order (the highest score first). The regions with the highest happiness score are Australia and New Zealand, followed by North America; while with the lowest score are Sub-Saharan Africa and Southern Asia.



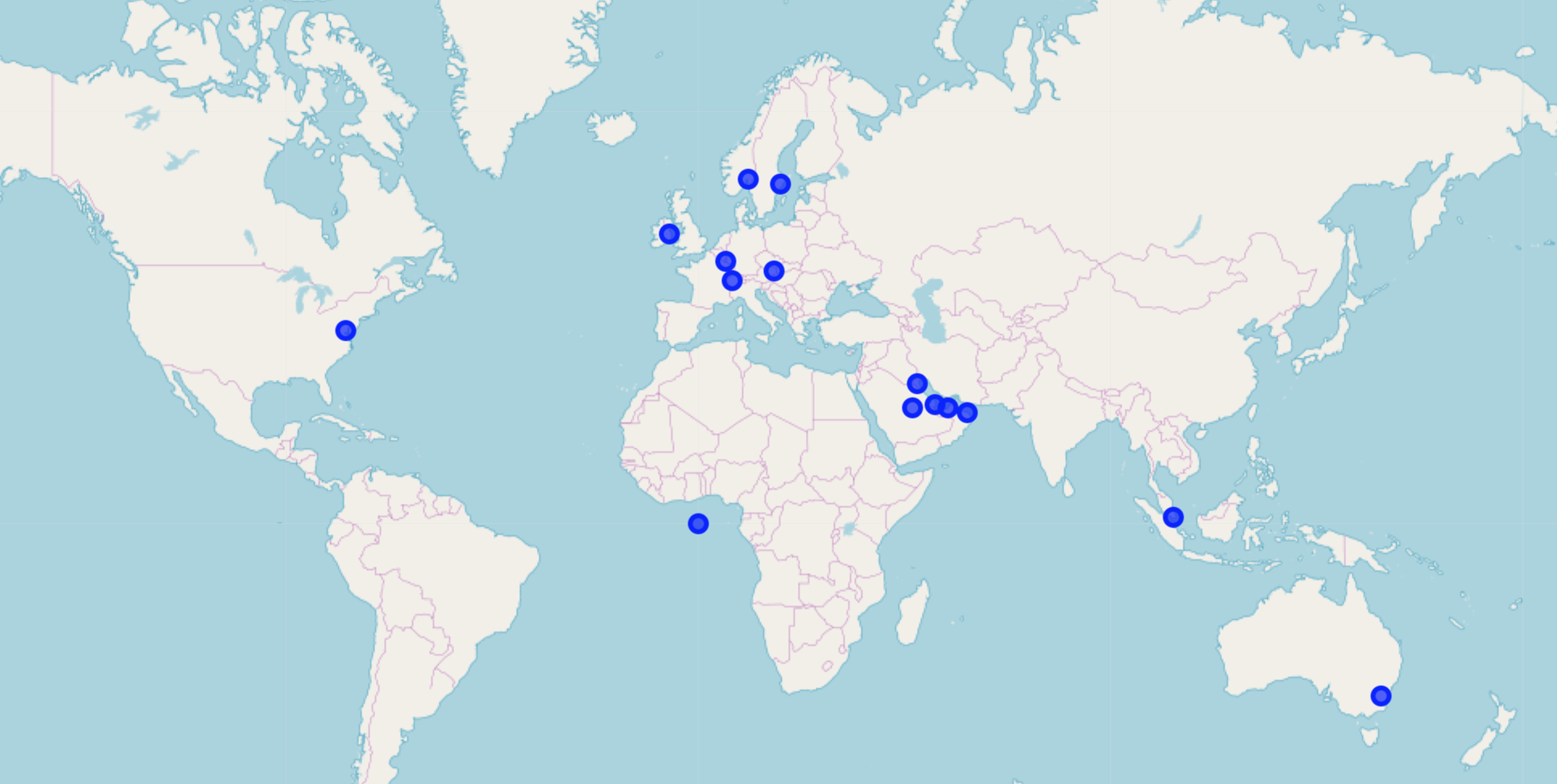
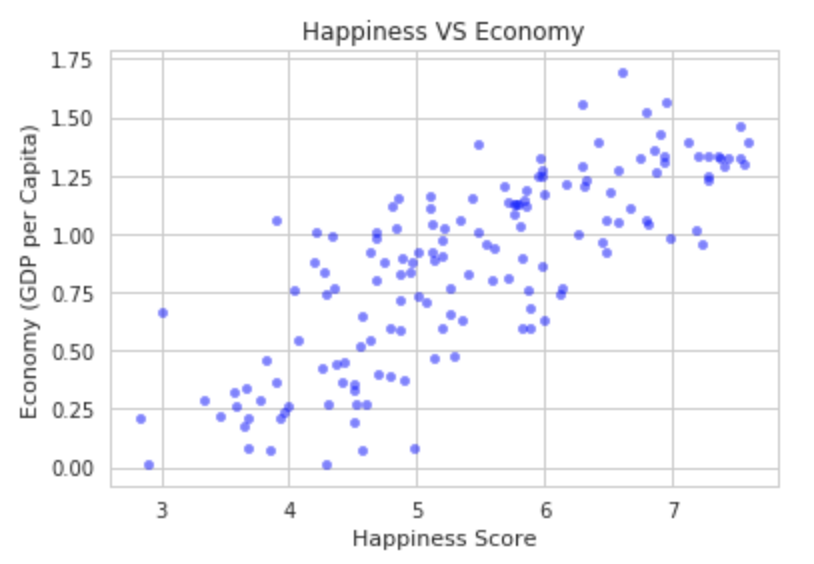
However, when looking in the spread of the regional happiness score, more countries with higher happiness scores tend to concentrate in the Western Europe, and Latin America and Caribbean. Compared to the above bar chart, using strip plot can visualize data distribution without affecting by extreme values or imbalanced data points.



In the following section, I would investigate the relationships between happiness score and 5 different factors (economy, health, freedom, government trust, and generosity), and see how these factors are affecting the extent of happiness score.

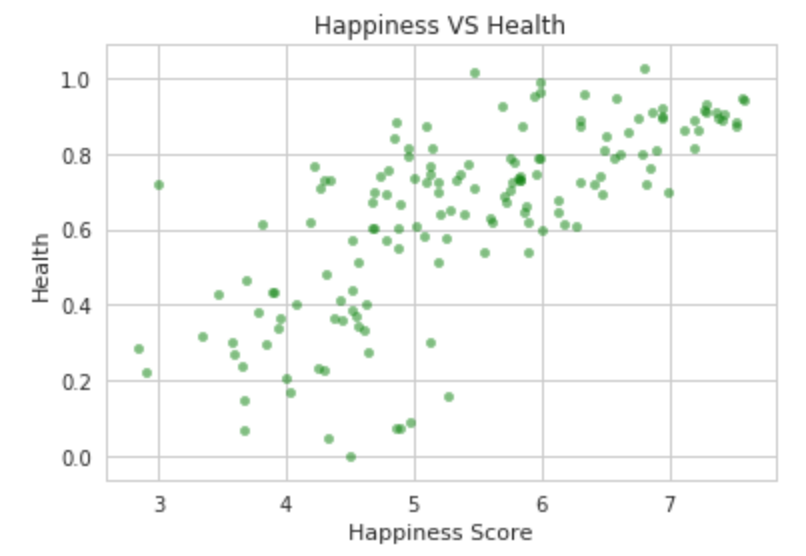
**3.3 Relationships between happiness and economy**

From the diagrams, there is a weak positive correlation between happiness score and economy (GDP per capita) of a country. Counties with higher GDP mainly concentrate in the Europe and Middle East.



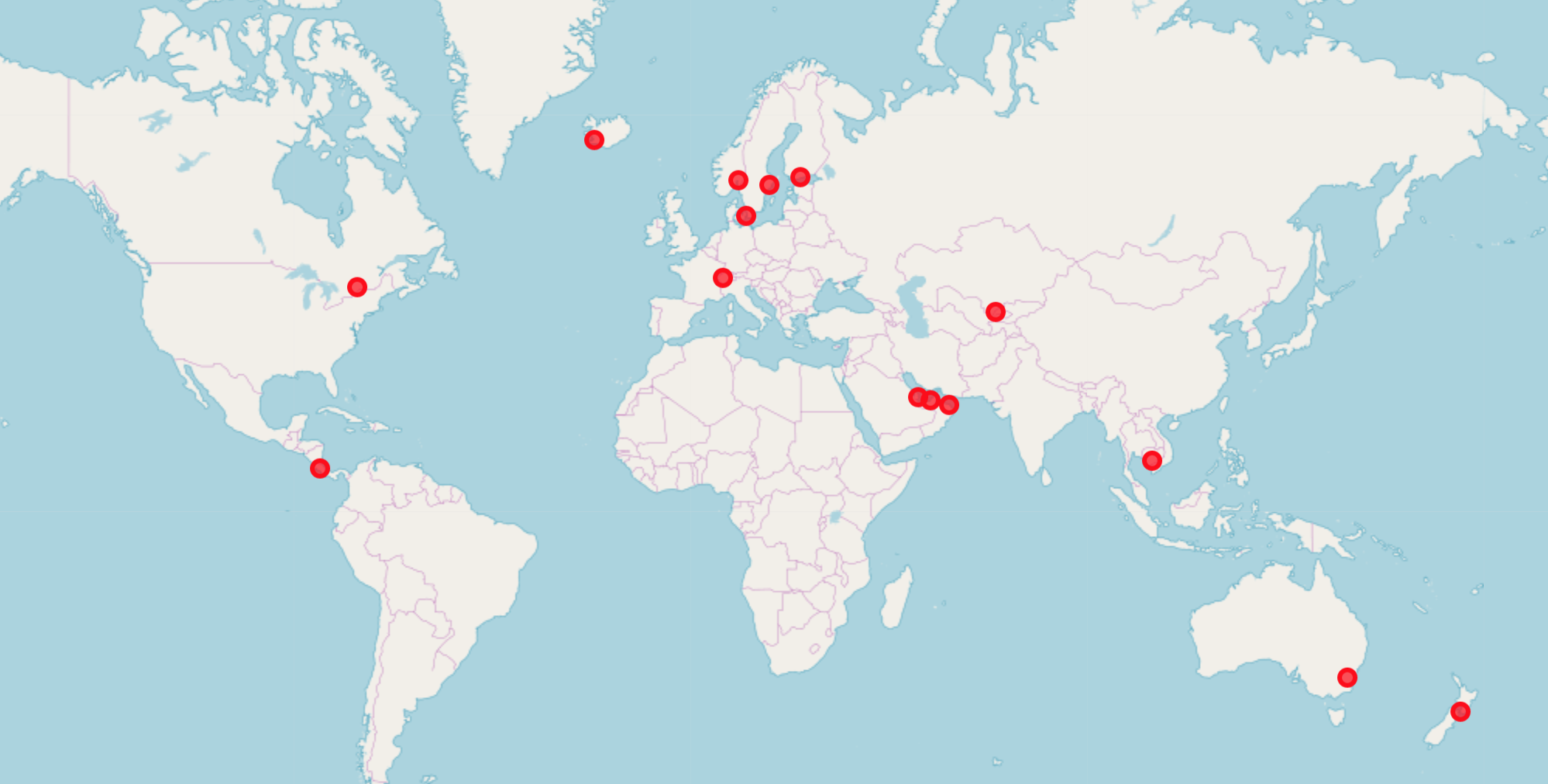
**3.4 Relationships between happiness and health**

From the diagrams, the correlation between happiness score and health (life expectancy) of a country is very weak. Countries with longer longevity are distributed mainly in Europe.



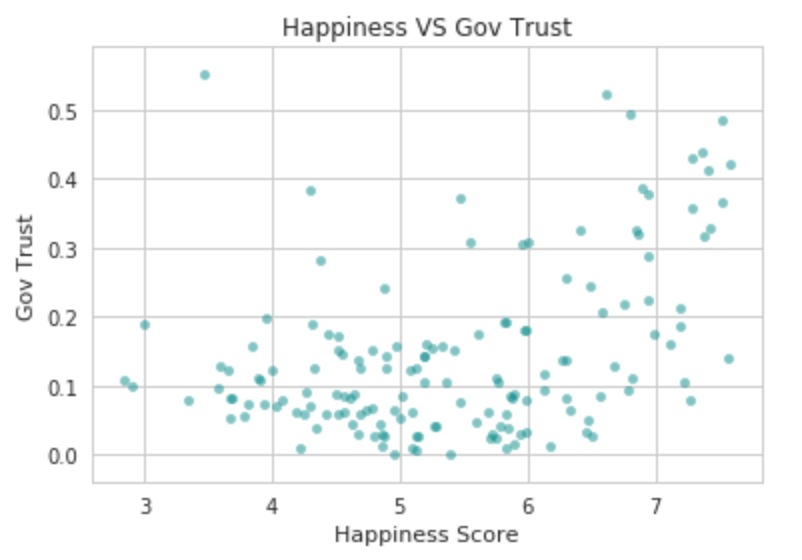
**3.5 Relationships between happiness and freedom**

From the diagrams, it can be concluded that there is random correlation between happiness score and degree of freedom of a country. Counties with higher degree of freedom scatter in different parts of the world.



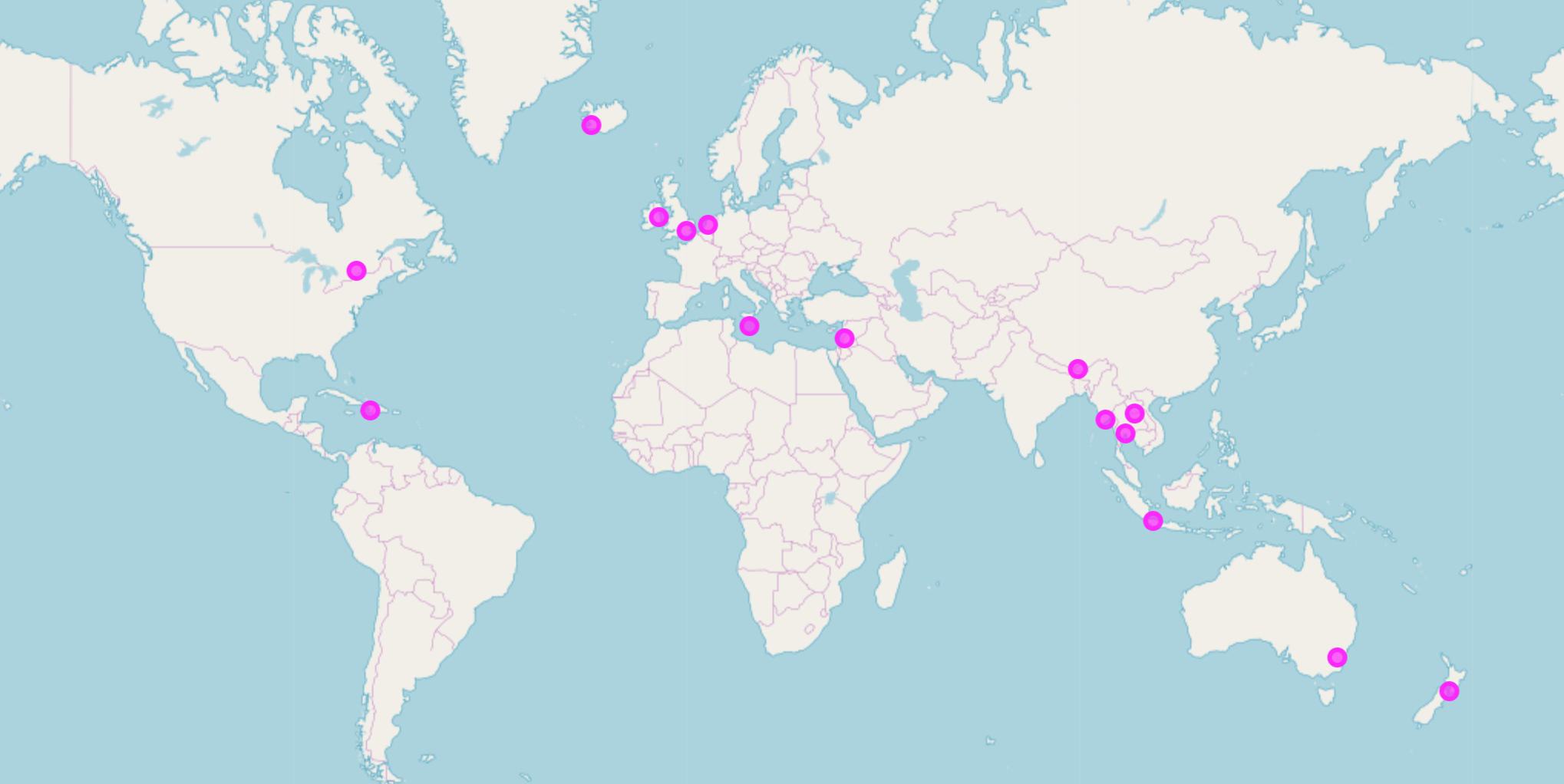
**3.6 Relationships between happiness and government trust**

From the diagrams, it can be concluded that there is no correlation between happiness score and degree of government trust of a country. Counties with higher degree of government trust are distributed in different parts of the world.



**3.7 Relationships between happiness and generosity**

From the diagrams, it can be concluded that there is random correlation between happiness score and generosity of a country. Counties with higher generosity scattered randomly in different parts of the world.



**4. Conclusion and Recommendations**

In conclusion, the economy seems to be the most relevant and has the highest influence to the happiness level among all the factors. If a country would like to improve the well-being of its residents, it is recommended that it should focus on boosting the economy first. But in fact, I reckon that this analysis project is not a comprehensive one since we only used a few variables for prediction. Therefore, for a more accurate and reliable prediction, we suggest including more variables involving both external (e.g. worldwide) and internal (local) factors in future projects.