**How to Make the World Happier**

**Introduction**:

Happiness is a popular topic that gets people’s attention-- it is everyone’s goal in life to be happy. “How to be happy?” is a global question that people, leaders, countries, etc. are always asking and trying to find answers to. If there’s a way to analyze, what makes people happy and improve people’s happiness, it would be a great benefit for the government making decisions and meanwhile increasing the overall satisfaction, productivity, and peace of the citizens. Furthermore, this topic also associates with sociological issue such as suicide rate, immigration, crime rate etc. Based on our concern, we chose the following dataset: <http://worldhappiness.report/ed/2018/>.

**Goal:**

Our goal is to find the factors in this dataset that significantly affect people’s happiness, and highlight these factors for both individuals as well as governments (Departments of Happiness for example). Moreover, we want to compare the influence of various factors in different countries. We hope that by narrowing the factors that affect happiness the most, we can focus our efforts on changing those the most.

**Data Description & Questions We Have:**

Our dataset contains the happy score from 156 countries from 2005 to 2017. There are 17 features in our dataset such as GDP, social support, life expectancy etc. Some interesting features include freedom to make life choices, generosity, perceptions of corruption, and happiness score (the output variable of these factors). The factors can split between high-level (governmental policies) and low-level (individual changes). The high-level factors are: GDP, life expectancy, perceptions of corruption, confidence in government, and democratic/delivery quality. The low-level factors are: social support, generosity, freedom to make life choice, positive effects in life, negative effects in life, and income. Thus, our group had the following questions before starting to analyze our data:

**Q1: What factors lead to the changes in happiness?**

Factors that are able to change people’s happiness are able to afford government a clear picture of what policy it can execute to improve the happiness of their citizen. Indeed, each country might be affected by different factors. Also, it is not necessary that countries in the same region are affected the same way. Thus, we have the following questions.

**Q2: Which countries shares similar features?**

Q2 arises from a challenge in our problem: we don’t have enough data points for each country. By grouping the similar countries, we will have more data points to analyze the significant features that impact the happy score. Do we group by region/continent? Do we group by some other factor?

**Q3: What features have significant impact on a particular group of countries?**

Having groups of similar countries allow us to see the features that are weighted more in that specific group. If that factor is controllable, it could be an effective approach to improving the happiness in that group of countries.

**Data Cleaning:**

Certain factors that were included in the original data set either had too many Null values or were a derivative of another factor. Thus, those factors were dropped for the analysis later in the report; these include: democratic quality and delivery quality.

**Exploratory Data Analysis:**

**Summary on Overall Findings:**

The distribution of the happiness score from 2005 to 2017 affords us a clear picture of what could possibly happen in the future. The happiness score ranges from 2.66 to 8.02, indicating that if we want to predict the next year’s worldwide happiness score, it is very likely that the score is within that range. We analyzed the happy score changes by year worldwide, each continent, and some countries as well. Figure 1. shows the overall happiness score changes by year. One abnormality we observed is that in 2005, the happy score was significantly higher than rest of the year. In fact, we have only 27 countries who had the happy score in 2005, where most of the countries are in western Europe and north America. Thus, we removed the data in 2005 when doing further analysis. It is also clear that the range of the happiness score displays an increasing trend. However, this doesn’t specify much for us because it is worldwide and over multiple years.

**Summary of Findings on each Continent:**

We then divided our dataset by continent. Figure 3 to Figure 8 indicates the variability and distribution by year in each region/continent. The range of happiness scores in Europe was the greatest compared to all other continents, and the score in North America and ANZ tended to have the smallest range, but have relatively high happy score. An interesting fact from Figure 6, representing the change of happiness scores in Latin American and Caribbean, suggests a trend that the range tends to be smaller after 2006 with the fact that there’s a huge increase from 2005 to 2006.

**Summary of Findings on some Countries:**

Figure 9 is the graph indicating the changes from 2006-2017 in the US, Canada, Australia and New Zealand. The graph shows that the trends before 2014 in each country were different, but indicates a similar trend from 2014 to 2017 in those four countries. Figures 10 and 11 told us the top 5 and lowest 5 happiest countries in 2017 ranked by their happiness score. All of the top 5 are from Europe and the lowest 5 countries are from Africa and Asia. Figure 12 shows the percentages of the different levels of happiness score by region in 2017. This graph explains one of our previous findings that the range of scores in Europe is big as central and eastern Europe have a high low-score percentage, while western Europe has a high score percentage. We then decided to focus and analyze three countries that were from different levels of happiness scores for the majority of our project: US, China and India (Figure 13 shows our findings). US has steadily decreased in happiness score, India has decreased quite a lot, and China has increased.

**Summary of Findings on some of the Variables by Region**

Figures 14-19 tell us about 1) the correlation between the variables and happiness score, 2) how the level of each variable is distributed by region. In particular, freedom, delivery quality, democratic quality, social support and GDP have clear positive correlations with the happiness score, and corruption has somewhat negative correlation with the happiness score.

**Clustering:**

We divided our data set into two groups. One is from 2006 to 2013, and the other one is 2014 to 2017, as those four years are more relevant to us. We did clustering for the two groups and we found 7 is the optimal number of clusters for both of them. Here, we chose three typical countries: China, India, and the US to see what clusters they are in. From 2006 to 2013, US and India were in the same cluster, while China was in the different one. From 2014 to 2017 all three countries were in three different clusters.

**Feature Selection:**

As there are many variables and factors that could affect happiness score in this data set, we wanted to extract the ones that weigh the heaviest on the score. We started with linear regression and correlation plots to determine which variables might have an interaction between them. However, this quickly become tedious and unhelpful. Finally, we tried both Lasso and SelectKBest and found those algorithms to be the best. Both algorithms gave us very similar results for each cluster so we decided to move forward with SelectKBest as it also provided F-scores to show us the weightage of each variable on the happiness score. Using the F-score we determined the top significant variables for each cluster and thus, each country.

**Solutions and Insights:**

1. **Answers to Q3**

Figure 20 and 25 are the answers to our third question. The same variables were significant for the clusters where all 3 countries (US, India, China) are in 2014-2017. We can conclude that for governmental agencies: GDP and life expectancy are the most significant variables, and for individuals: mainly social support, but also positivity and freedom. Thus, we can conclude that if those variables can be increased in some way, the happiness score will increase accordingly.

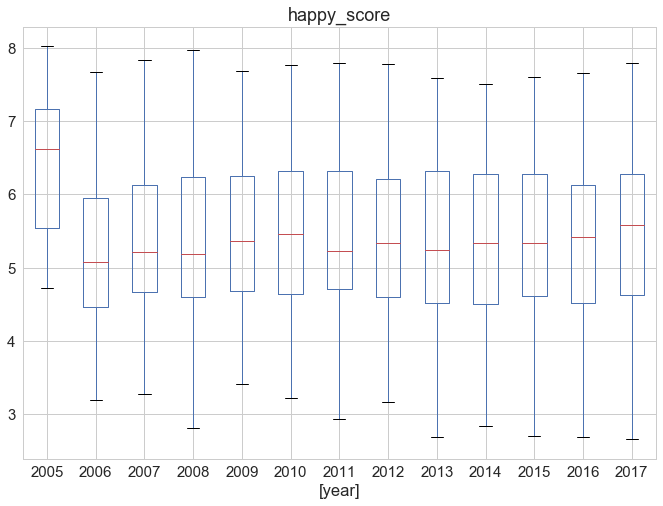
1. **Weight of significant variables in each clusters changes by year**

Now that we’ve known the significant variables from 2014 to 2017, figures 21-24 showed us how the impact of those variables on happiness scores changed overtime. For example, life expectancy used to be significant from 2006 to 2013 in India and US, but from 2014 to 2017, it is not as high as it used to be. One can guess that people in India and the US started to realize that how to live is more important than living longer. The importance of social support also changed over years. Social support made little impact in China, but from 2014 to 2017, it weighed more in determining the happiness. A noticeable fact occurred when we looked at the impact of positivity (laughter and enjoyment) and negativity (worry, sadness, and anger), especially in China: the positivity was no longer as significant as from 2006 to 2013, but negativity, on the contrary, explained a greater part of the happiness score. This might result from the fact that however much worry, sadness and anger can negatively impact people’s happiness has already exceeded the level of positivity that can positively do.

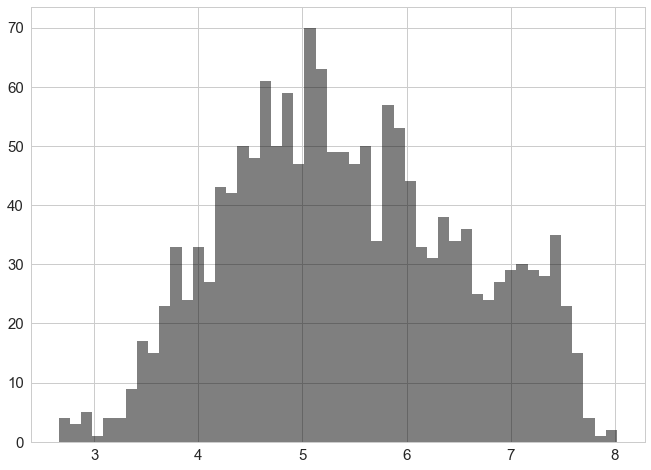
1. **Recommendations**

Based on our findings, the following suggestions are applicable to the three typical countries we picked and the clusters they were in. From governmental perspective, focusing on GDP growth and improving health related entities such as care, pollution, etc. would be most effective in increasing happiness. Reinforcing social support would also be an effective approach. As a person, effective ways to increase happiness include expanding or putting more importance on social support and positive environments in one’s life.

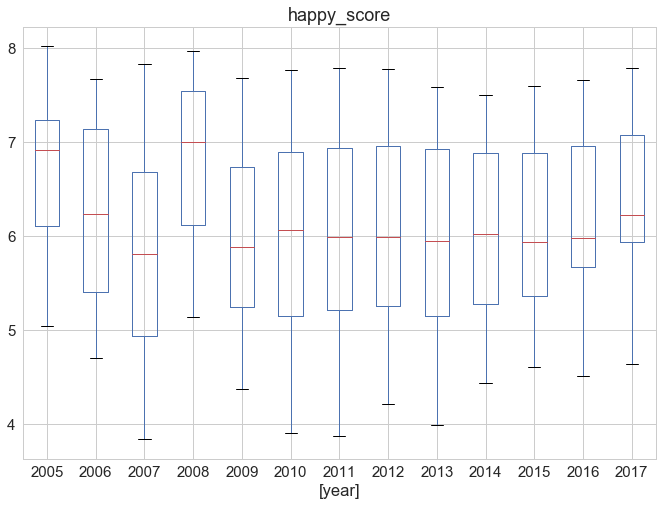
Appendix



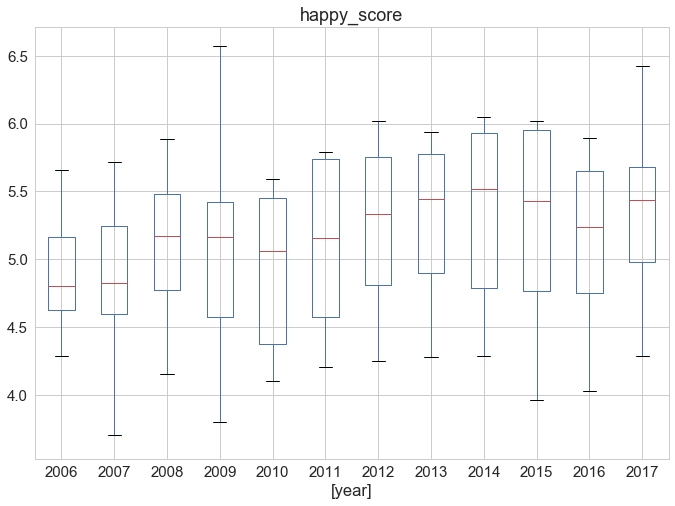
*Figure 1*. Overall Happy Score Changes by Year



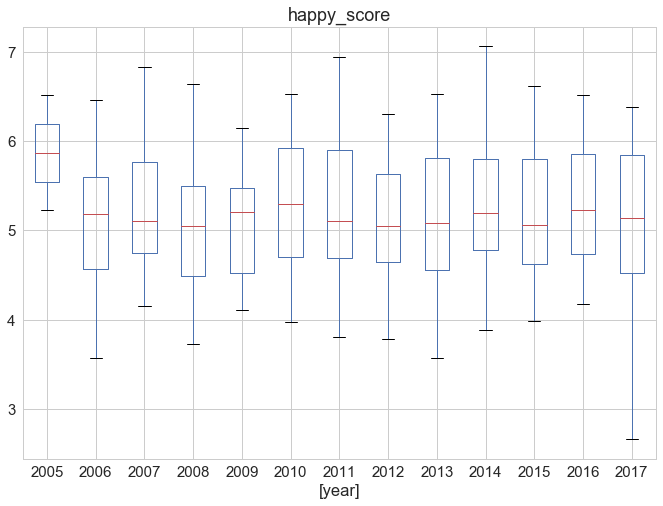
*Figure 2*. Happy Score Distribution



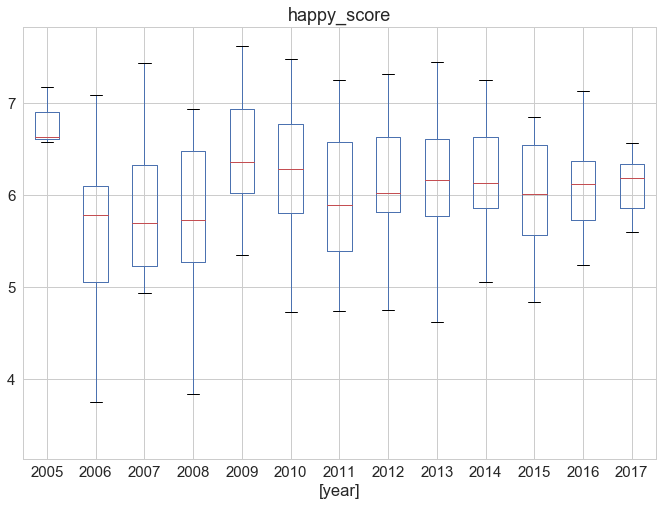
*Figure 3*. Happy Score Changed by Year in Europe



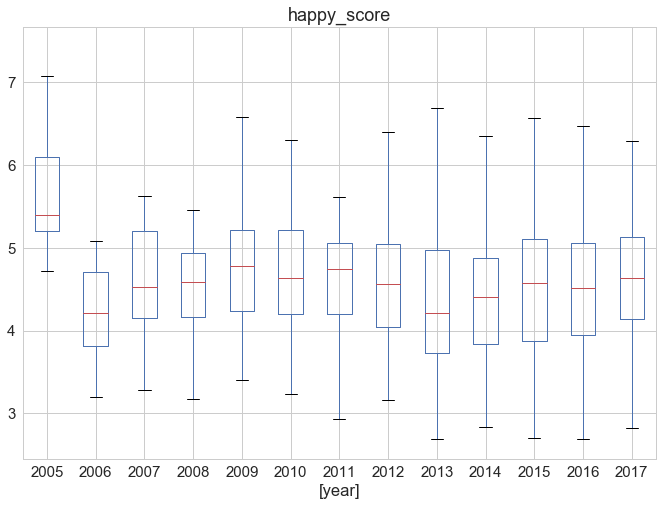
*Figure 4*. Happy Score Changed by Year in Commonwealth of Independent States



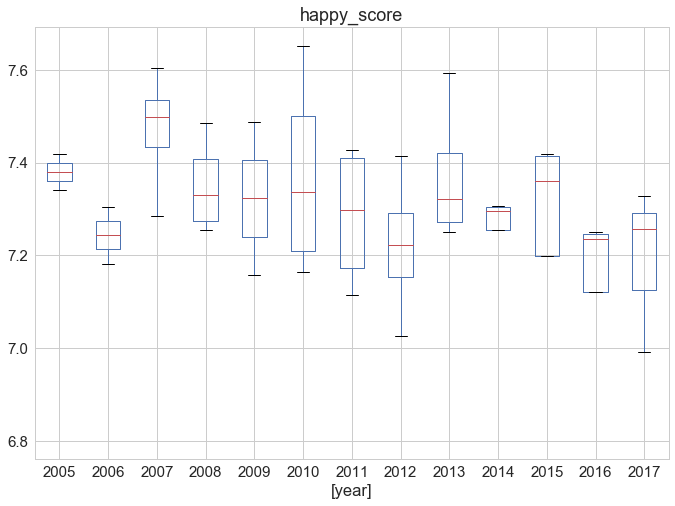
*Figure 5*. Happy Score Changed by Year in Asia



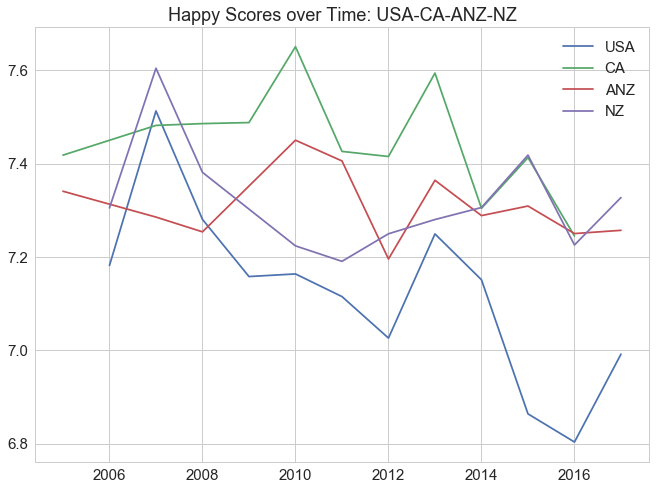
*Figure 6*. Happy Score Changed by Year in Latin America and Caribbean



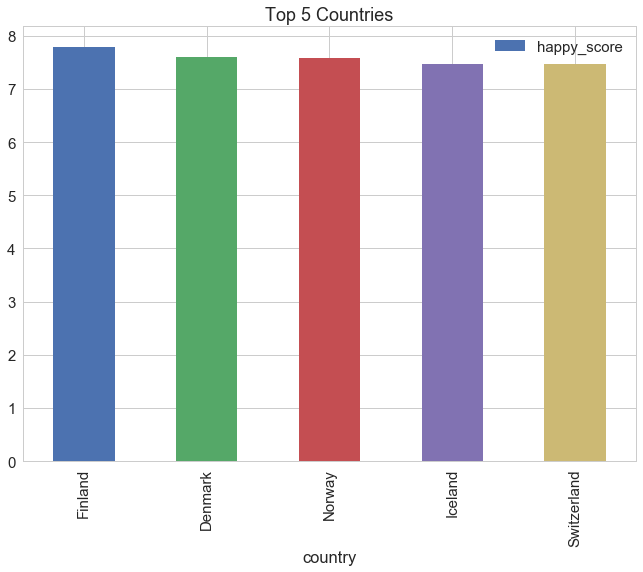
*Figure 7*. Happy Score Changed by Year in Africa



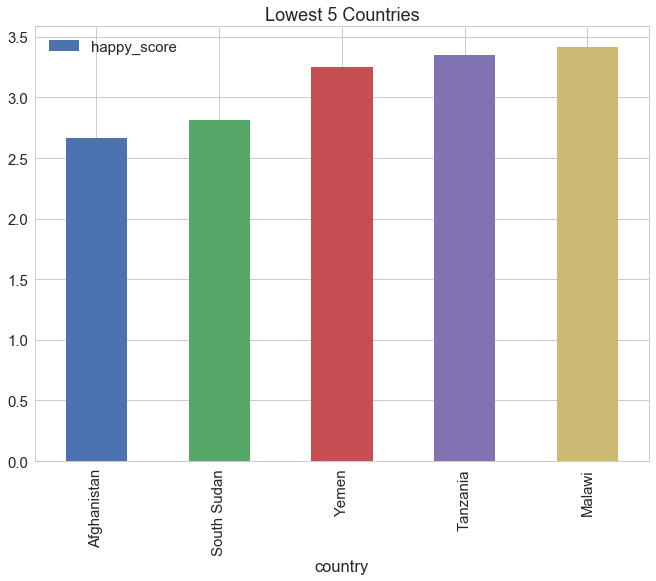
*Figure 8*. Happy Score Changed by Year in North America and ANZ



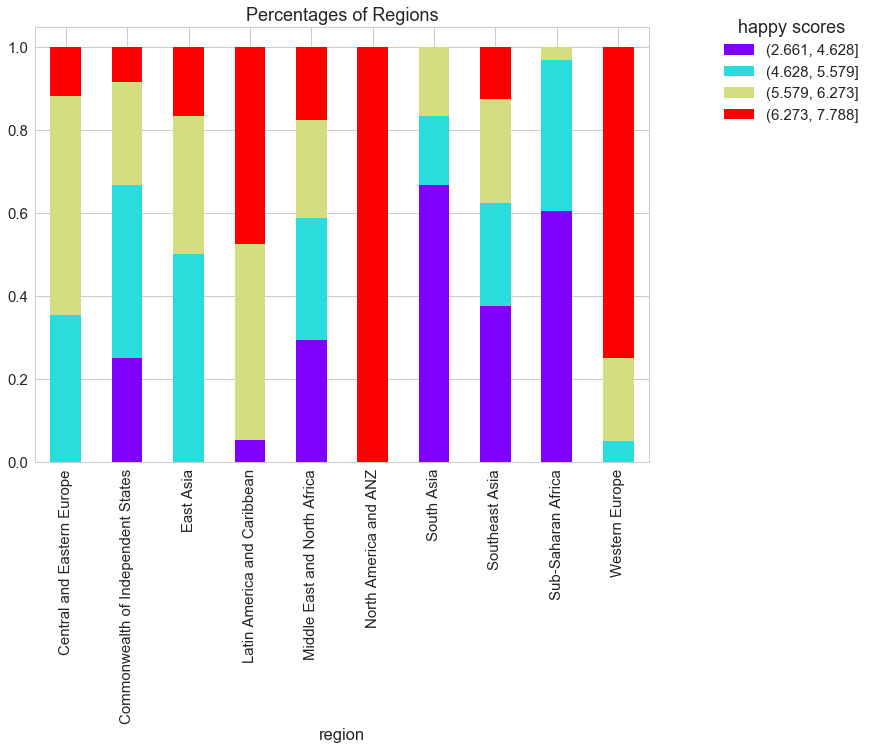
*Figure 9*. Happy Score Changed by Year in US, CA, ANZ, NZ



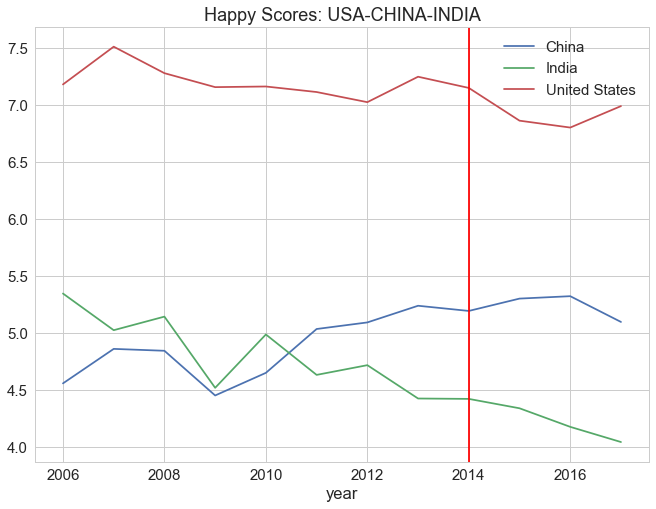
*Figure 10*. Top 5 Countries in 2017 Ranked by Their Happy Score



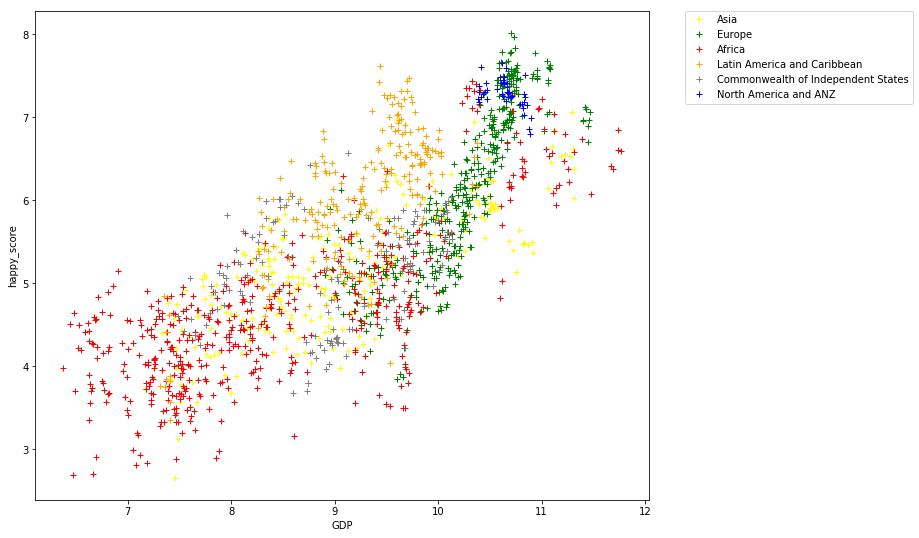
*Figure 11*. Lowest 5 Countries in 2017 Ranked by Their Happy Score



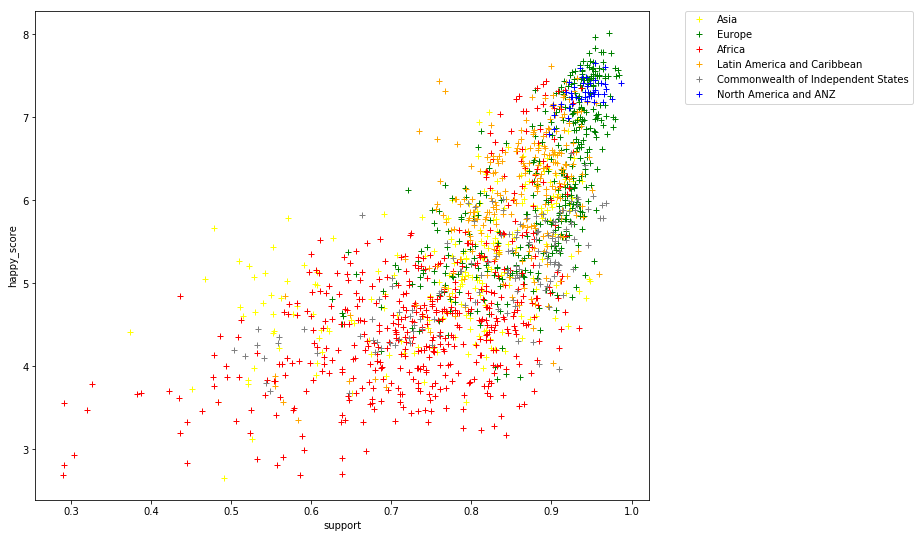
*Figure 12*.Different Levels of Percentage by Region



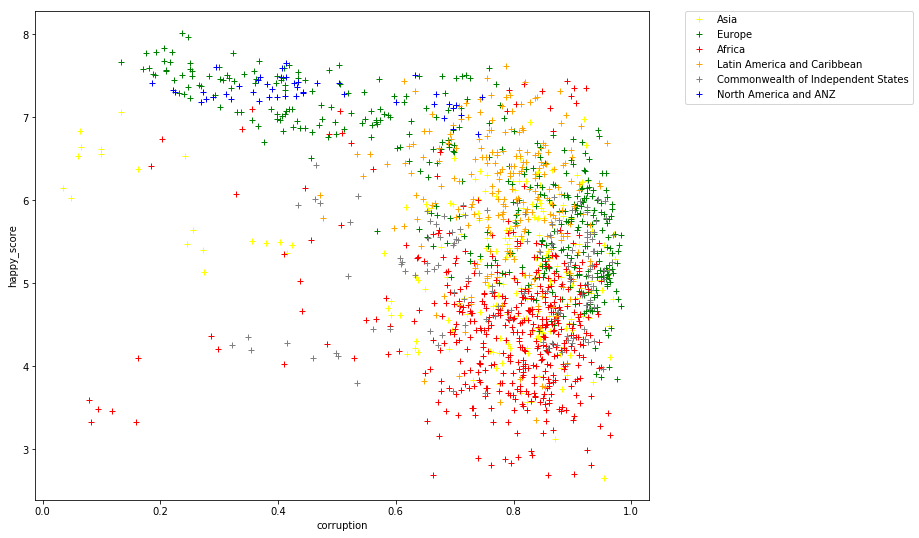
*Figure 13*. Happy Score Changes by Year in US, China and India



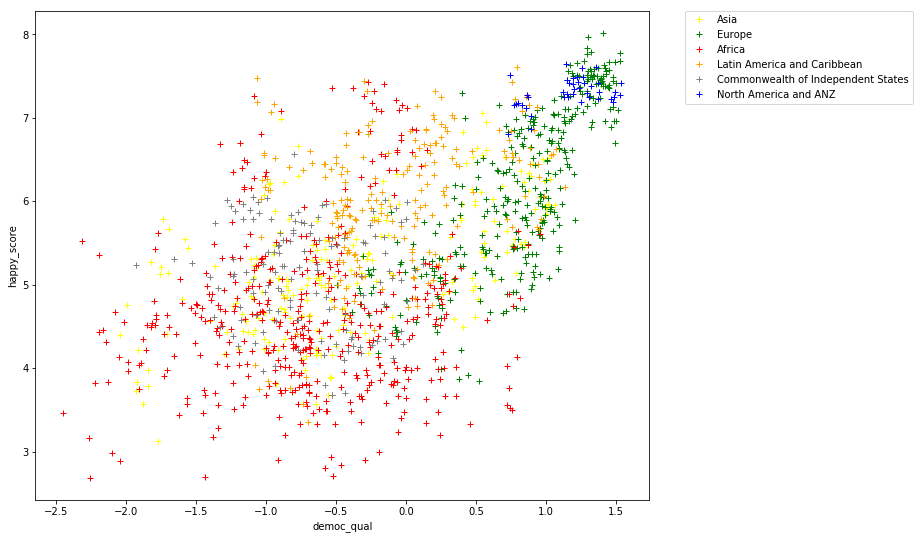
*Figure 14.* GDP effect on happy score by each region



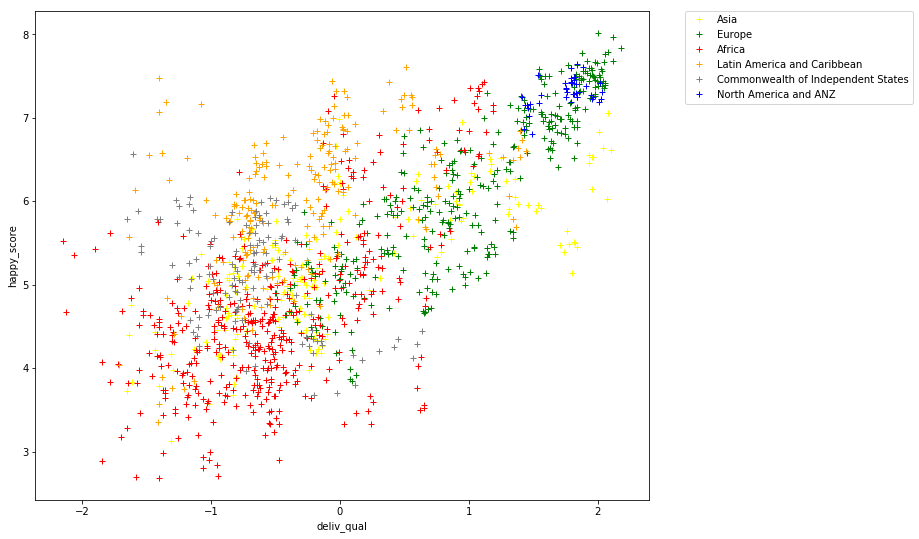
*Figure 15.* Social Support effect on happy score by each region



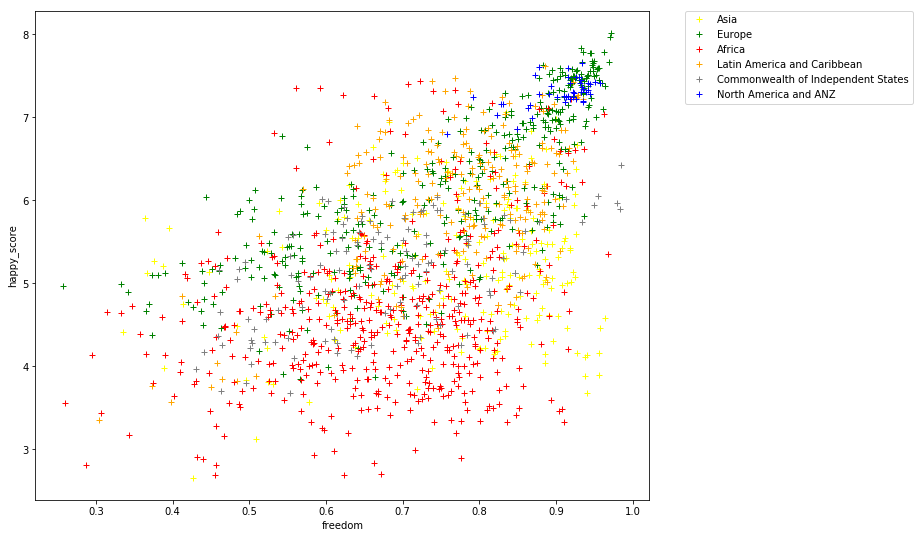
*Figure 16.*Corruption effect on happy score by each region



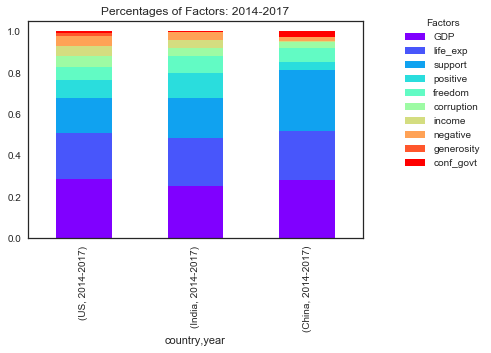
*Figure 17.* Democratic Quality effect on happy score by each region



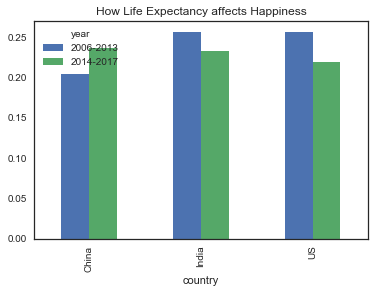
*Figure 18.* Delivery Quality effect on happy score by each region



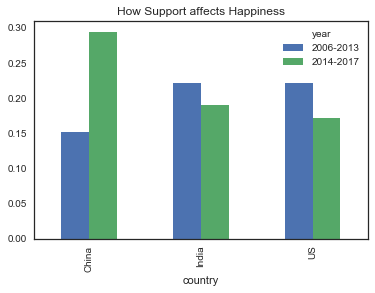
*Figure 19.* Freedom effect on happy score by each region



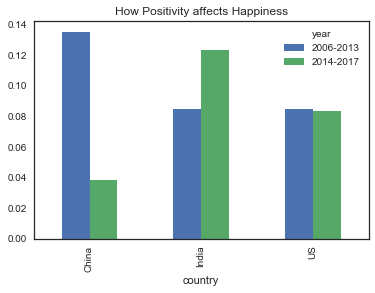
*Figure 20.* Important Variables for three typical countries from 2014 to 2017



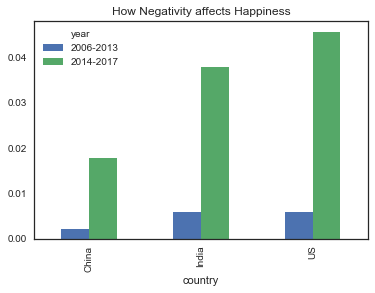
*Figure 21.*Life Expectancy Effect on Happy Score for three typical countries



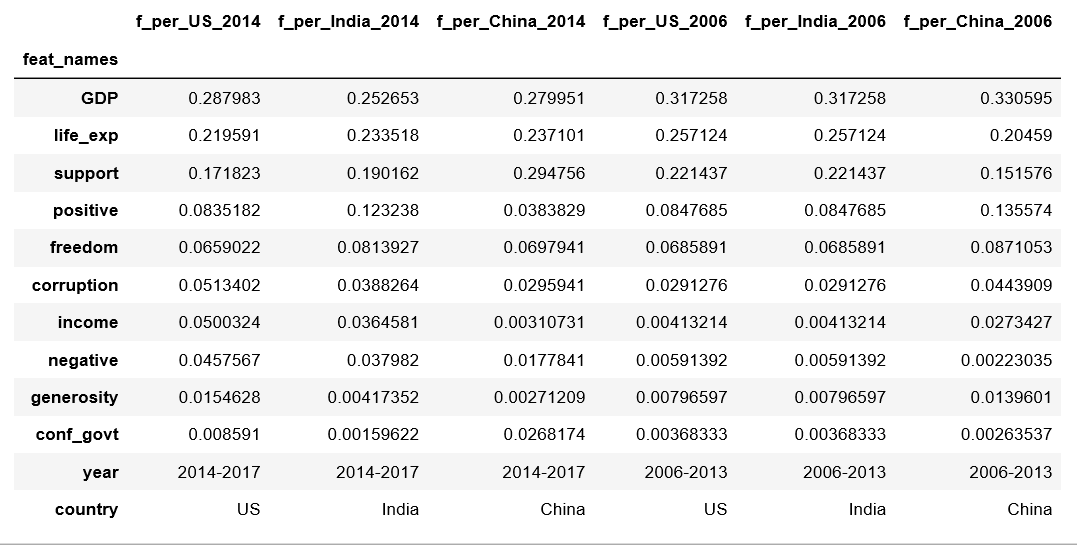
*Figure 22.*Social Support Effect on Happy Score for three typical countries



*Figure 23.*Positivity Effect on Happy Score for three typical countries



*Figure 24.*Negativity Effect on Happy Score for three typical countries



*Figure 25.* Table of the scaled f-scores to determine weight of variables