**Prediction of Global Happiness Using Machine Learning**

**Group - 36**

* Venkata Kalyan Ram Ghatti – 0756328
* Priya Kandula – 0754438
* Venkata Hanuman Sai Kumar Kaparaju – 0753837
* Harish Veeramosu - 0755976

**Introduction:**

World Happiness Report is the annual publication of country’s happiness ranking based on the individual surveys. This survey mainly contains articles and rankings of the national happiness based on the responses of the individual’s own lives. For the country’s ranking some random population was selected unbiased and provided with Cantril ladder survey. This contains a ladder rating from 0 to 10, they are asked to rate their own lives ‘0’ being worst possible life and 10 being the best possible life. The main features based on which people were asked to rate are GDP, Healthy life Expectancy, Social support, Freedom of Choice, Generosity and Perceptions of corruption. They were asked to rate from 1 to 10.

Happiness is the one of the most important things in people lives. When it comes to measuring the happiness, things get complicated while indicating the whole country’s happiness ranking based on the group of people. One of hardest thing in measuring happiness ranking is considering both rich and poor family as same. At the same time different things and necessities make the people happy based on different category of their life (Rich poor). As we all know that Happiness is an aspiration of every human being. This can be a measure of social progress. This ranking would help in making some important decisions in future based on views and survey of the people on own lives. This report will also help in increasing the well-being of the people. Happiness can be influenced by genetics, personality and even sometimes sheer luck. By shifting the measurement to happiness than wealth would give us the more overall satisfaction to people.

The main objective of the World Happiness report is to review the science of measuring and understanding people’s well-being and to use the survey measures of life satisfaction to track the quality of lives. In addition to this analysis of life evaluations throughout the world, each world happiness report has variety of importance which would help in improving the lives of the people. This reports mainly focus on the social environments for happiness, as reflected by the quality of personal social connections and social institutions. Our report mainly covering the data from 2017 to 2019. Our dataset mainly consists of social features and their impact on people lives. Some of the key features are GDP per capita, Social support, healthy life expectancy, freedom, generosity, and corruptions.

**Benchmarks:**

Despite the increase in financial health there is low happiness score in some countries. This clearly sets a benchmark to know the wellbeing of people with respective to countries. Which would help the countries to set standard reforms to improve the happiness score based on the analysis, doing this helps in improving the standards of living. This will also have a great impact on performance of people in jobs and works which would directly help the countries to develop in different prospects like social and economics. It also helps to keep a track of people’s real well being. Other countries which are top on the list would do some contributions to lowest happiness score countries based on analysis.

**Methods:**

Happiness is termed as the state of being happy. Our data contains multiple features that can impact the happiness coefficient or the happiness score.

Data comprises of 158 observations and 13 features. The dataset included reports collected by WHO from 2015 to 2019 which consists various features which are:

* **Region**:

Happiness score is calculated for various regions. The dataset contains 10 different regions namely 'Western Europe', 'North America', 'Australia and New Zealand', 'Middle East and Northern Africa', 'Latin America and Caribbean’, ‘Southeastern Asia', 'Central and Eastern Europe', 'Eastern Asia’, ‘Sub-Saharan Africa', 'Southern Asia'.

* **Happiness Rank**:

Every country/ Region is ranked on basis of happiness score with 7.587 being the highest score and ranked the top-most by considering all other features.

* **Happiness Score:**

Happiness Score is calculated on basis of various features used in the data from 2015 to 2019. While happiness score depicts the life expectancy and economy of a country/ Region.

* **Country:**

Happiness Score is calculated for 158 different countries from 2015 to 2019 on basis of 13 features. By calculating happiness score for different countries, we evaluate happiest countries to live in.

* **Economy (GDP per Capita):**

Economy or Gross Domestic Product indicates a country total goods and services for a period. Hence, Economy became a tool to measure a countries happiness coefficient.

* **Family**:

Family plays a crucial role in deciding happiness factor. However, World Happiness Report indicates a major relation between Family, Economy, health, and Happiness Score.

* **Health (Life Expectancy):**

Quality of life is measured differently in different countries. In UK it is measured using 43 different indicators where health, well being and relationship and where we live mattered the most, among others.

* **Freedom:**

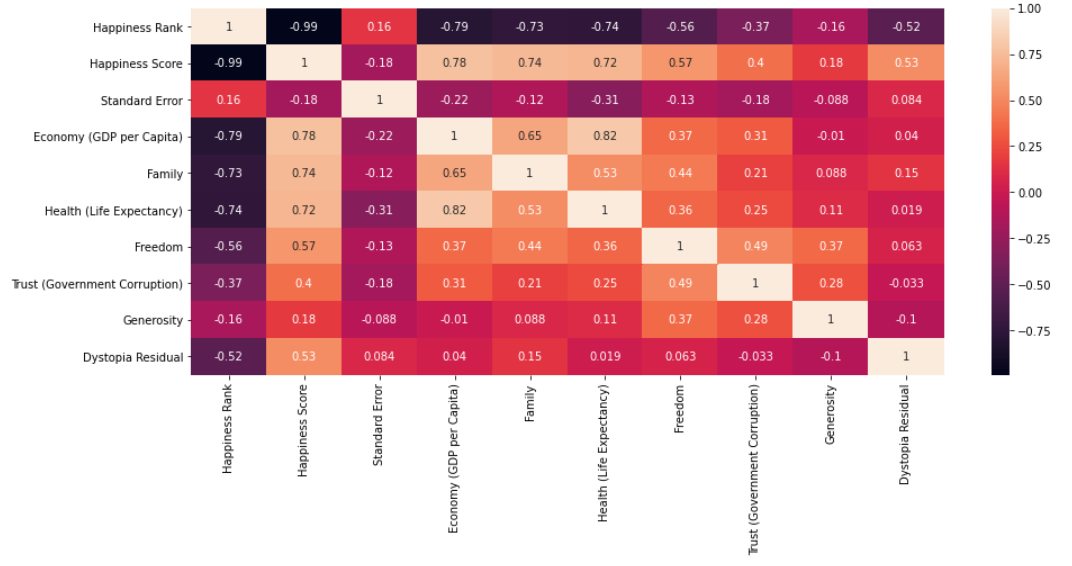
Freedom or life choice of people plays a key role for happiness.

* **Trust (Government Corruption):**

It is a known fact that people invest a lot of money towards government in the form of taxes. People trust the government to utilize these funds for the betterment of people of that country. Happiness Score depends on how the government is taking care of people and their concerns.

**Exploratory Data Analysis:**

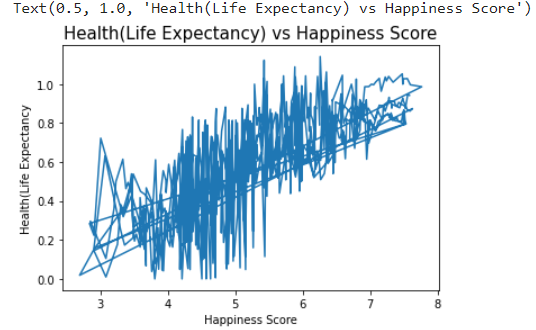
From the below co-relation matrix we understand how a feature impacts on the other feature. For happiness score Family, Freedom, Health (Life Expectancy) and economy (GDP per Capita) are known to play a crucial role.

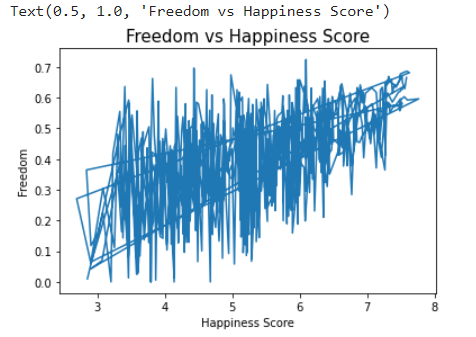
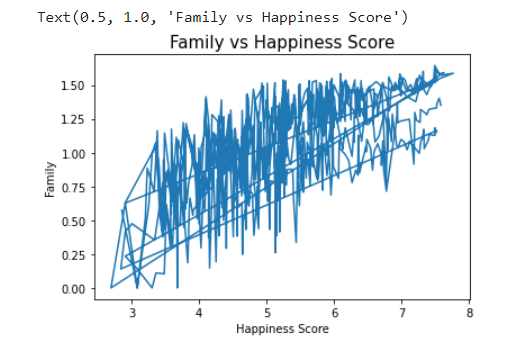


On obtaining key features from the co-relation matrix, we plotted few graphs to show how each feature impacted on the other.

Below are the graphs plotted for key features discussed as above with respect to happiness score. It is evident that lower happiness coefficient resulted in lower economy rates, lesser freedom, average health, and family life.

However, higher happiness score resulted in higher economy rates more freedom and fabulous health and family life. From the visualizations we conclude that happiness score plays a crucial role in everyone’s life.





* To build a model that predicts happiness score we had to club data from five different years (2015-2019).
* Major challenges that we encountered was differences in number of features as they varied between 10 to 13 for 5 different years.
* We worked on renaming the differences in the column names to not to impact the agenda of the model.
* While processing the data we encountered columns with a maximum of null values or columns added no value in predicting the happiness score and dropped them.
* We merged 5 different years of data to obtain a single data frame.

We are using machine Learning algorithms to predict the happiness score. This model can be used to predict happiness coefficient for upcoming years and help in deciding and improving on the features that the country is lacking in with respect to happiness.

We are using 4 different regression methods to predict the happiness score namely:

* + **Linear Regression:**

This regression is simplest regression technique and used for Predictive analysis and a linear methodology for including the connection between the response and predictors or descriptive variable.

* + **Random Forest Regression:**

The random forest is a classification and regression algorithm consisting of many decisions’ trees. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree.

* + **Support Vector Machine Regression:**

Support Vector Machine (SVM) is a very common algorithm in machine learning used in both regression and classification. In the SVR, the straight line is referred to as the hyperplane, and it is the line equation y= mx+c in linear regression. Support Vector Regression is similar to the linear regression. The data points on either side of the hyperplane nearest to the hyperplane are known as Support Vectors for drawing the graph.

* + **Decision Tree Regression:**

Decision Tree Algorithms can be used both for regression and for classification tasks; however, we will work with the Python and scikit with a basic implementation of regression.

When the dependent variable is constant, regression trees are used.

The value of terminal nodes is the mean for regression trees for observations in that region. Therefore, we predict to use the mean value if an intangible dataset falls in that area.

**Results:**

We have built a machine learning using various regression methods of which Decision tree regression model has proven to be the best model in predicting happiness score for 2020 with 99% accuracy when compared to other models which are SVM regression, Linear Regression, Random Forest Regression which predicted happiness score for 2020 with an accuracy of 98%.

|  |  |
| --- | --- |
|  |  |
|  |  |

|  |  |  |
| --- | --- | --- |
| **Regression Model** | **Accuracy** | **Mean\_Squared\_error** |
| **SVM Regression** | 98% | 0.02 |
| **Linear Regression** | 98% | 0.02 |
| **Random Forest Regression** | 98% | 0.02 |
| **Decision Tree Regression** | 99% | `0.02 |

**Discussion:**

* We faced a challenge while merging the data of different years into one data frame. We overcame by data preprocessing methods.
* To predict happiness score we choose Decision Tree Regression model as the model was very accurate.

**Conclusion:**

Building this model, we will be able to predict the happiest country for 2020 to live in. Doing so will help the government to shape the country and improve the lacking features. Using this model, we can retrieve the life expectancy of people and improve their living standards to be happy.

By doing so people will have freedom to choose a country to live on basis of their standard of living, health, GDP, and family.

**Contributions:**

|  |  |
| --- | --- |
| Priya Kandula &  Venkata Hanuman Sai Kumar Kaparaju | * + Finding and Research on Global happiness dataset   + Preprocessing of dataset.   + Building regression methods   + Conclusion and suggestions |
| Harish Veeramosu &  Venkata Kalyan Ram Ghatti | * + Introduction to Global Happiness   + Finding and Research on Global happiness dataset   + Exploratory data analysis   + Conclusion and suggestions |

**References:**

[1]: Research Paper: <https://pure.ulster.ac.uk/ws/files/12575336/BHCI_2018_paper_271.pdf>

[2]: Dataset: <https://www.kaggle.com/unsdsn/world-happiness?select=2015.csv>

[3]: World Happiness Report 2017: <http://worldhappiness.report>