**World-Happiness-Report-Data-Analysis**

Analyzing the World Happiness Report dataset as Part of The EPFL Applied Machine Learning Course (Project 1). The purpose of this project is to exercise and demonstrate acquired abilities. Should the person who is looking at this analysis have any questions or suggestions, do not hesitate to contact me.

## The Data

Our dataset is the 2021 World Happiness Report published by the United Nations Sustainable Development Solutions Network. This data was collected by the Gallup World Poll, using a set of surveys from more than 160 countries in over 140 languages. The surveys aim to analyze data from hundreds of thousands of individuals and investigate the ways in which elements of people’s working lives drive their wellbeing.

Each country is given a happiness score that is the sum of scores from the following 8 criteria:

GDP per capita healthy years of life expectancy, social support (as measured by having someone to count on in times of trouble), Country name, Perceptions of corruption (as measured by a perceived absence of corruption in government and business), Freedom to make life choices, Generosity (perceived freedom to make life decisions generosity), Healthy life expectancy. The countries are then ranked from most happy to least happy based on Ladder score.

## Tasks performed:

### A. Importing, cleaning and numerical summaries

1. Download the data set data.csv from the Resources tab.
2. Import the data as a panda DataFrame.
3. Check the number of observations.
4. Obtain the column headings.
5. Check the data type for each column.
6. Check if there are any missing values.
7. If necessary, remove any observations to ensure that there are no missing values and the values in each column are of the same data type.
8. Obtain the mean, minimum and maximum value for each column containing numerical data.
9. List the 10 happiest countries.
10. List the 10 least happy countries.

### B. Indexing and grouping

1. Use the column 'Country ' to groupby with other variables.
2. Compute the mean Ladder score for each region and rank the regions from most happy to least happy.
3. Compute the number of Country name that have a happiness score/Ladder score above 6.0.
4. Compute the difference between the maximum and minimum Ladder score for each Country. Which country has the largest range of Ladder scores?

### C. Bar plot of the Ladder Score

1. Obtain a horizontal bar plot of the Happiness Score of the top 10 countries. Your bar chart should have the names of the countries are listed vertically along the y-axis and the x-axis should have labels for each number from 0 to 8. Ensure that the chart has an appropriate title and labels.
2. You will now modify the bar chart you obtained in step 1 to turn into a stacked bar chart where the overall happiness score is divided into the seven parts corresponding to the columns:

* Social Support
* GDP
* Health
* Freedom
* Generosity
* Corruption

1. Obtain the same stacked horizontal bar chart as in step 2 but this time instead of the top 10 countries

### D. Histogram of Job Satisfaction

1. Obtain a histogram of the Job Satisfaction using the following categories:

* 40%-50%
* 50%-60%
* 60%-70%
* 70%-80%
* 80%-90%
* 90%-100%

### E. Pairwise Scatter plots

1. Obtain scatter plots of the Ladder Score versus each of the other variables. Your plots should be displayed as multiple plots table and obtained with one command as supposed to separate commands for each plot.

### F. Correlation

1. Obtain the correlation between the Ladder Score and each of the other variables. Which variable has the highest correlation with the Ladder Score?

### G. Probabilities

1. Compute the probability that randomly selected country with Ladder score over 6.0 is from Western Europe. You will have to use pandas to count the appropriate quantities.

### Conclusion

This analysis illustrated that the world’s happiest countries are primarily in North Western Europe, North America, and Australia & New Zealand. It also revealed that Economy (GDP per capita) is the most important factor in evaluating a country’s happiness. Unsurprisingly, the happiest countries and world regions generally tended to be ones with strong and stable economies.

Finland is an example of a country with a high happiness score. The importance of Economy is also strongly positively correlated with those of Health. The Ladder score has a strong negative correlation with Perceptions of corruption.This is expected, since more economic stability and higher GDP per capita generally encourages stable and comfortable family life as well as increases the availability of proper medical resources and healthcare. These factors then weigh more when determining overall happiness.