

# PROJECT 1

## • THE WORLD HAPPINESS

By Soukhna Wade  
Bellevue DSC680  
Applied Data Science

---

# DATA SOURCES: WORLD HAPPINESS REPORT 2017

- Country: Name of countries
- *Happiness Rank: Rank of the country* based on the Happiness Score
- Happiness Score: Happiness measurement on a scale of 0 to 10
- Economy: Value of all final goods and services produced within a nation each
- Family: Importance of having a family
- Life Expectancy: Importance of health and amount of time people expect to live
- Freedom: Importance of freedom in each country
- Generosity: The quality of being kind and generous
- Trust: Perception of corruption in a government
- Dystopia Residual: Plays as a reference
- Continents



+

•

○


# Cleaning Dataset

Data cleaning is the process of detecting and correcting corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate, or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.

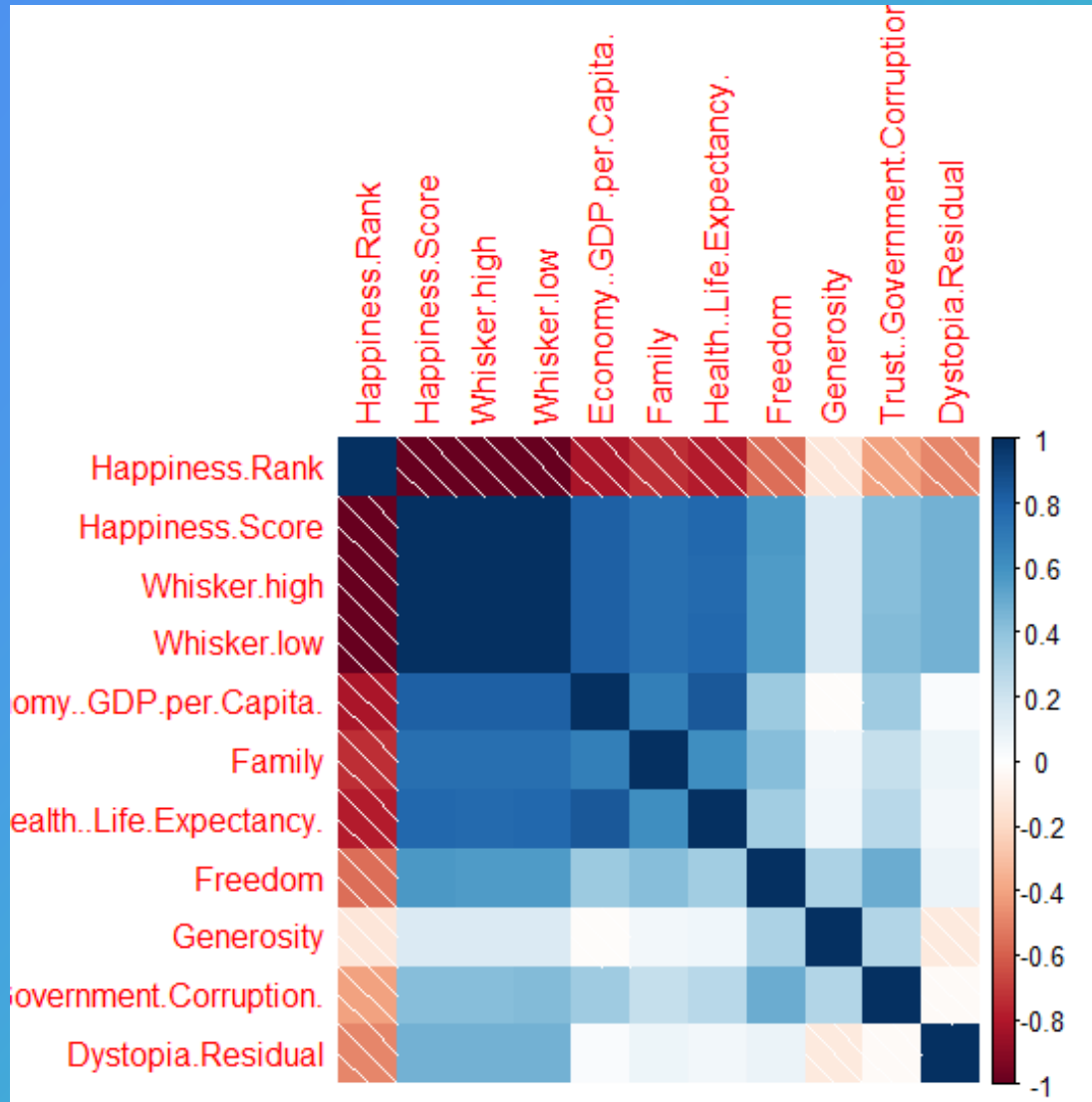


# Visualization

visualization's purpose is the communication of data. Visualization transforms from the invisible to the visible. In this section, we use different variables to determine their correlation

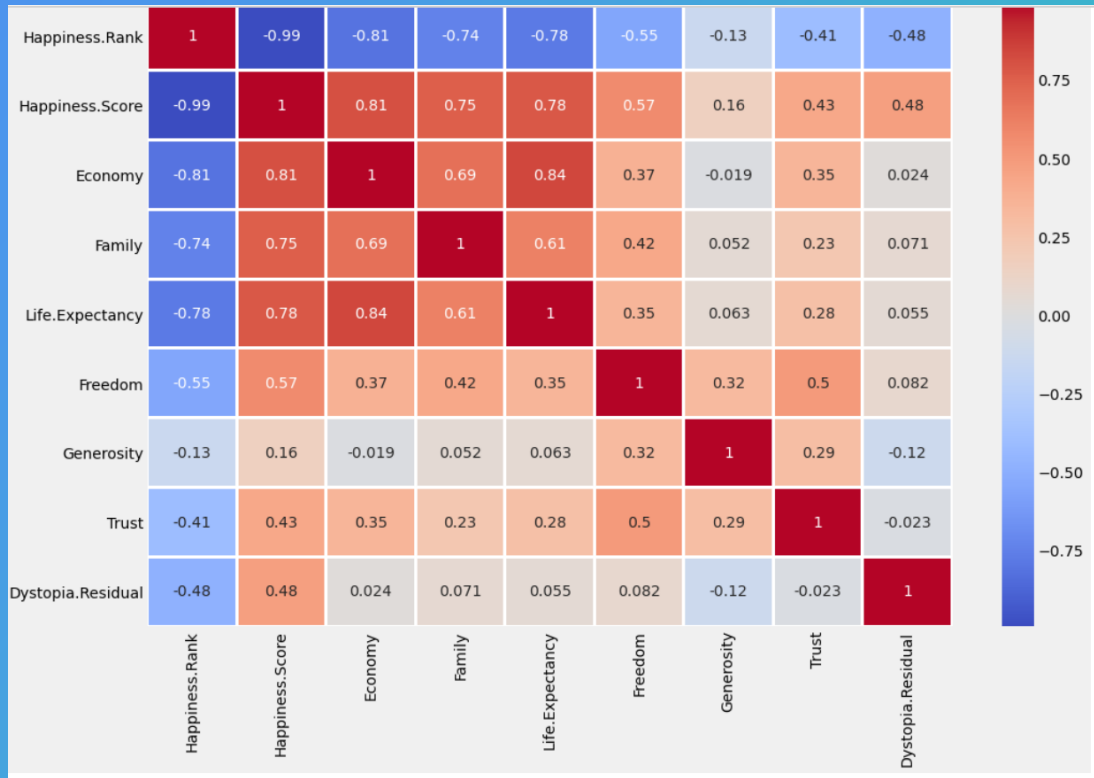


# Correlation plot between numerical variables



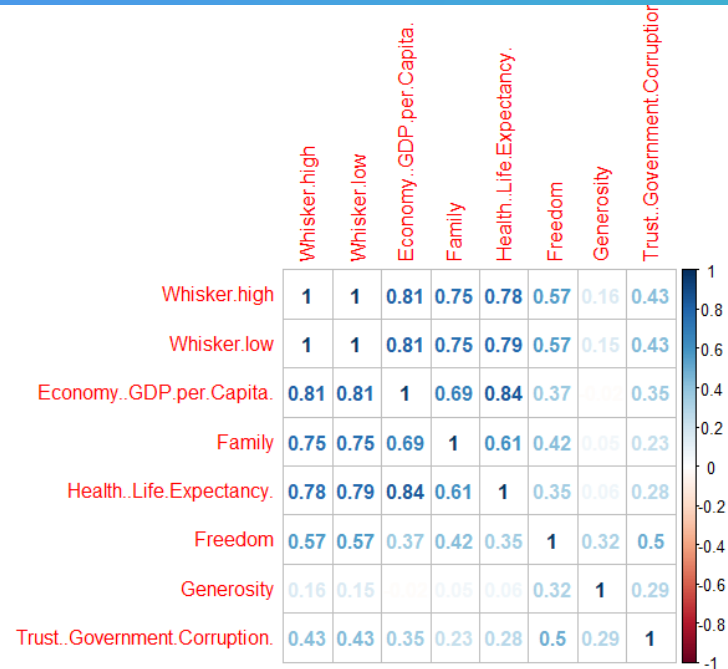
There is an inverse correlation between the Happiness Rank and all the other numerical variables. The lower the happiness rank, the higher the happiness score, and the higher the other seven factors that contribute to happiness.

# Heatmap



The heatmap of correlation among the variables displays the color palette in the side represents the amount of correlation among the variables. Therefore, lighter shade represents a high correlation. You can see that the happiness score correlated with the economy, family, and life expectancy. It is least correlated with generosity

# Removing happiness Rank




The correlation plot shows that the economy, life expectancy, and family play the most significant role in contributing to happiness. Trust and generosity have the lowest impact on happiness scores.



# Prediction

We implement several machine learning algorithms to predict happiness scores. First, we split our dataset into training and test set. The dependent variable is happiness score, and the independent variables are economy, family, life expectancy, freedom, generosity, trust, and dystopia residual.





# COEFFICIENTS


coefficients

	Coefficient
Economy	1.000016
Family	0.999884
Life.Expectancy	1.000109
Freedom	1.000070
Generosity	1.000102
Trust	0.999772
Dystopia.Residual	0.999935



# Conclusion

The summary shows that all independent variables have a significant impact, and the adjusted R squared is one. There is a linear correlation between dependent and independent variables. Also, the sum of the independent variables is equal to the dependent variable which is the happiness score. This is the justification for having an adjusted R squared equal to one. As a result, multiple Linear Regression will predict happiness scores with 100 % accuracy.



# References

1. <https://worldhappiness.report/archive/>
2. <https://www.kaggle.com/unsdsn/world-happiness>
3. <https://www.kaggle.com/pinarkaya/world-happiness-eda-visualization-ml/data#Linear-Regression>
4. <https://www.kaggle.com/javadzabihi/happiness-2017-visualization-prediction/report>
5. <https://www.kaggle.com/roshansharma/world-happiness-report?select=2017.csv>
6. <https://www.kaggle.com/pinarkaya/world-happiness-eda-visualization-ml#Random-Forest>
7. <https://www.kaggle.com/dhanyajothimani/basic-visualization-and-clustering-in-python/data>
8. <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>
9. [https://en.wikipedia.org/wiki/The\\_Happiness\\_Hypothesis](https://en.wikipedia.org/wiki/The_Happiness_Hypothesis)
10. <https://www.theworldcounts.com/happiness/the-definition-of-happiness-in-psychology>
11. Statistical Appendix for Chapter 2 of the World Happiness Report 2020 John F. Helliwell, Haiyang Huang, Shun Wang, and Max Norton February 29, 2020