

Marlene Aviles

November 16, 2019

Final Project: EDA

World Happiness Report: 2015 & 2016

Statistical/Hypothetical Question:

I hypothesize that regions with greater freedom to make individual life choices will tend to have a higher perception of happiness.

Outcome:

Many factors can influence the way a person measures their level of happiness. During this analysis, people were asked in a survey to rank the extent to which perception of the following factors contributes to happiness scores. The elements used were freedom, generosity, economy, and family, and the years 2015 and 2016 were compared.

Although it was noted that freedom does have a positive correlation, other factors were found to have a more significant impact on the happiness score. Pearson's correlation was measured, and the top three variables with the highest correlation to happiness scores in the years 2015 and 2016 were found to be economy, family, and health.

Regression analysis was applied to the 2015 dataset. This analysis was conducted between freedom and happiness scores, as EDA showed a positive linear relationship. The dependent variable for this analysis was happiness score, and freedom was used as a predictor. This analysis reflected that the mean absolute error for the variable of freedom is a 0.58 on Happiness score.

It was also noted that the scatter plots support that regions with the highest sense of freedom tend to have a higher happiness score, as reflected in the data collected from western Europe. When looking at the histograms for happiness score, 2015 shows a binomial distribution versus 2016, which indicates a closer to normal distribution. The mean score for both years remained consistent at a 5.38. On the other hand, the sense of freedom did decrease from 0.43 in 2015 to 0.37 in the year 2016.

What do you feel was missed during the analysis?

Additional demographic information, such as the socio-economic status of those individuals surveyed, could have been included in this analysis to provide a better understanding to see if their current situation has any weight on the selection of the scores.

Were there any variables you felt could have helped in the analysis?

The inclusion of a variable that measures the relationship between a person's work environment and their happiness score could assist this analysis.

Were there any assumptions made you felt were incorrect?

It was assumed that the level of sense of freedom a person has would have the most significant effect on their happiness score.

What challenges did you face, what did you not fully understand?

It was difficult to understand how the variables were collected, and it required additional research on my end, to fully understand them.

Sources Cited:

Gallup (n.d.) Tracking the World's Happiness.

Retrieved from <https://www.gallup.com/analytics/247355/gallup-world-happiness-report.aspx>

United Nations Sustainable Development Solutions Network. (2017). World Happiness Report.

[Data file]. Retrieved from <https://www.kaggle.com/unsdsn/world-happiness>