

Zach Mumbauer

DSC 530 Final Paper

Summary of Analysis

The EDA revealed strong trends and relationships between key predictors and happiness scores. Countries with higher GDP per capita and social support generally reported higher happiness scores. Histograms of variables showed normal-like distributions for happiness, GDP, and social support, but revealed outliers in corruption perceptions and generosity. Scatter plots confirmed linear relationships between predictors like GDP and happiness. Regression analysis highlighted that 73.4% of the variance in happiness scores was explained by the chosen predictors, with social support and freedom to make life choices having the largest coefficients.

One limitation of the analysis was the lack of non-linear modeling. While linear models explained a lot of the variance, relationships like diminishing returns were not explored. Additionally, looking at interaction between groups of predictors like how social support might amplify the effect of GDP on happiness, were not tested.

Adding more variables could have enhanced the analysis. Using a separate dataset to include education quality, inequality indices, and mental health metrics could help explain happiness levels. Qualitative variables like cultural norms, political stability, and other events could help interpret the results.

The analysis assumed linear relationships between predictors and happiness scores, equal variances for the t-test, and complete independence between variables. These assumptions helped simplify the modeling process, but they likely have

compounding effects between each other. Happiness is likely influenced by a complex web of interdependent predictors. That type of relationship isn't understood with the linear modelling used in the EDA.

By sourcing the data through Kaggle, the dataset used for EDA was clean and free from any major errors which made the analysis smooth. Understanding the exact causal mechanisms behind the observed correlations is difficult since correlation does not imply causation. The analysis clearly highlights some strong predictors of happiness and validated the hypothesis.