Phase 3 - Info 2950

Hypothesis 1: Countries that are reported to have an overall higher rate of happiness will experience lower levels of suicide rates.
Analysis: We will run a linear regression model to determine if there is a relationship between happiness score and suicide rates (number of suicides per 100,000 people).This would be our formula: Happiness Score = $\alpha + \beta_{suicide\_rate} * \text{Suicide Rate} + \varepsilon$. We will be looking to see if $\beta_{suicide\_rate} < 0$, to determine if there is a negative correlation between happiness and suicide rates per country.

Hypothesis 2: Smaller population sizes of countries correspond to lower birth rates, higher minimum wages, and lower crime index, which in turn leads to higher happiness scores for countries.
Analysis: We will run a multivariable linear regression to determine the correlation of the inputs of birth rate (number of live births per 1,000 people in the population), minimum wage (in dollars), crime index (out of 100), and population (in number of people) in affecting the output of happiness score. We will test whether $\beta_{population} < 0$, $\beta_{birth\_rate} < 0$, $\beta_{min\_wage} > 0$, and $\beta_{crime} < 0$. Our formula would be: Happiness Score = $\alpha + \beta_{birth\_rate} * \text{Birth\_rate} + \beta_{minimum\_wage} * \text{Minimum wage} + \beta_{crime} * \text{Crime} + \beta_{population} * \text{Population} + \varepsilon$.

Questions for Reviewer

1. When analyzing our hypothesis would we need to translate our data to support the relationship within the hypothesis.
2. Can we include data relationships that may be intuitive and obvious to the reader in our project? For example, can we hypothesize a connection between increasing GDP and happiness score?
3. Can the result of our experiment be inconclusive? How should we factor outliers into our hypothesis?