

Term Project

Assignment 12.2

DSC 530

Prof. Matthew Metzger

2023 World Happiness Report Analysis

Carlos Cano

Outcome:

Throughout various analysis techniques that were observed using the variables within the dataset presented by The World Happiness Report 2023, sourced through Kaggle it became evident that there were several direct correlations when data was used to explain what makes and keeps individuals across nations happy.

One of the main relationships was that there was various links to increased life expectancy as it relates to happiness. This was intuitive at first but to view the outcome and run regression analysis as well as calculate p-value to collaborate this information was reinforcing to the published documentation that is reported annually.

Analysis Failures:

The analysis could have been aided by the incorporation of year glances of similar dataset in an expansive view to test if certain variables would change overtime. As the dataset contained quite a bit of information, careful detail had to be paid attention as some of the data had originally been processed and select raw data curated.

Missing Variables:

Among the variable, something that I thought was amiss was more economic data, but as I researched the topic it became evident that the way the data was originally collected from surveys across the world made it hard to gauge holistic numbers versus that of government cited data. Originally a compilation of these datasets

Incorrect Assumption:

Going into this EDA blind, I had assumed that happiness as measured within this dataset would be straight forward and have positive correlations. My inferences in how much this data was positively correlated was shot down immediately upon various analysis. I believe this type of information should be published within the annual reports as the statistical significance is quite important within the various variables and their overall connections.

Challenges:

Originally a compilation of these datasets posed the greatest of challenges when trying to gather data for analysis of this EDA. As such further research was conducted and extrapolated in an attempt to source various data from different parties in an attempt to curate a custom "Master Data" this proved to be extremely cumbersome, and the efforts were given up. I believe further experience in manipulating data may garner a different personal outcome.