



Project 5

Freedom vs Happiness

Can We Have Both?

SACV Consulting

Shayla Bason . Andrei Muravev . Chris Campbell . Vivek Bavda



Introduction

The United States has been and will continue to be the leader of the world. Both domestically and internationally, our cornerstone value has been freedom. This has driven our policy and government for which we hold responsibility. Implied in this ideal of freedom is that it leads to happiness for us and the world. Given the importance of this assumption to all of us, SAVC Consulting has taken on a multipart project to study the link between happiness and freedom. In thinking about freedom, most people agree that personal freedom leads to happiness. Freedom of Speech, Religion, Association, and Press have few critics. It is economic freedom where the political divide in the U.S. severs the views of Americans. Economic Freedom is Part 1 of this project. Based on the Heritage Foundation's data of American and international economic freedom and the World Happiness Report using Gallup Polling data, this study seeks to determine if one of our cornerstone values, economic freedom, leads to happiness. Using Linear Regression(L1 and L2 Regularization), Random Forest Regressors, and a Neural Network Regressors to model this data, we have used the correlation coefficient or R2 to measure how much of happiness is explained by economic freedom or the lack thereof.

Outline

Problem Statement

Data Sets

Cleaning

EDA

Modeling

Conclusion and Recommendations



Problem Statement

In order to identify and further explore societal/economic traits that correlate to happiness SACV Consulting will first identify which of the ML models listed above can explain the highest degree of variance through it's happiness predictions. This will allow for iterative feature engineering and fine-tuning of the model.



Data Sets

Data set 1:

Index of Economic Freedom 2008 - 21

The Index of Economic Freedom is an annual index and ranking created in The Heritage Foundation and The Wall Street Journal to measure the degree of economic freedom in the world's nations.

Source: heritage.org

Data set 2:

World Happiness Report 2008 - 21

The World Happiness Report is a publication of the Sustainable Development Solutions Network, powered by data from the Gallup World Poll and Lloyd's Register Foundation, who provided access to the World Risk Poll.

Source: worldhappiness.report



Cleaning

Import:

- Read in CSV files
- Standardize and rename like column headers
- Create “country_name” column and merge the two Dataframes on it, dropping all rows without corresponding data

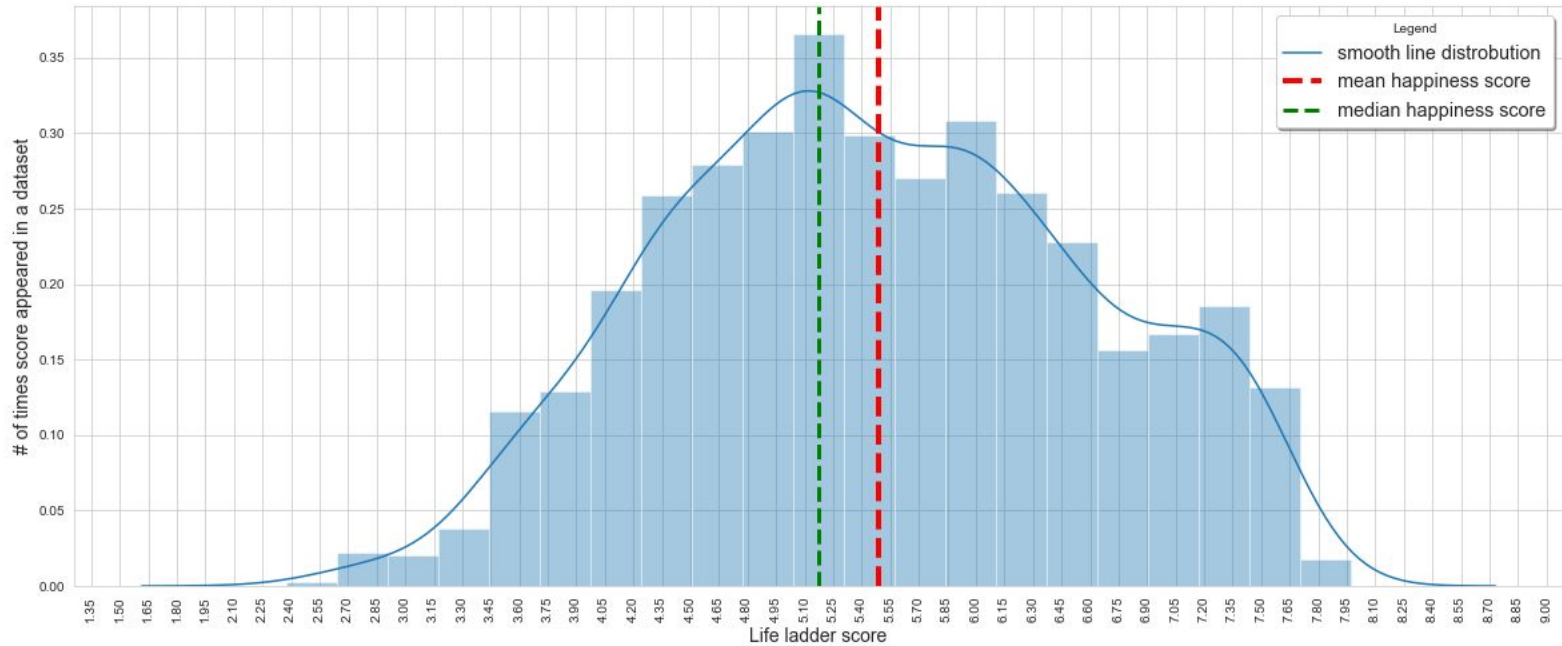
Impute:

- Drop `judicial_effectiveness` and `fiscal_health` columns due to lack of data
- Evaluate the performance of the imputers below, and predict all remaining null values with the highest performer:
 - Mean
 - Median
 - Linear Regression
 - Bayesian Ridge
 - Decision Tree Regressor
 - **Extra Trees Regressor** (*winner*)
 - K Neighbors Regressor

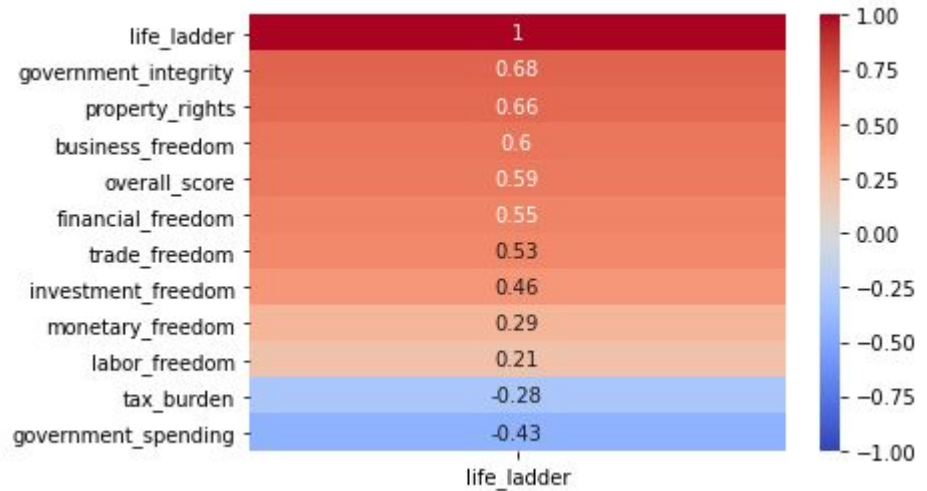


EDA

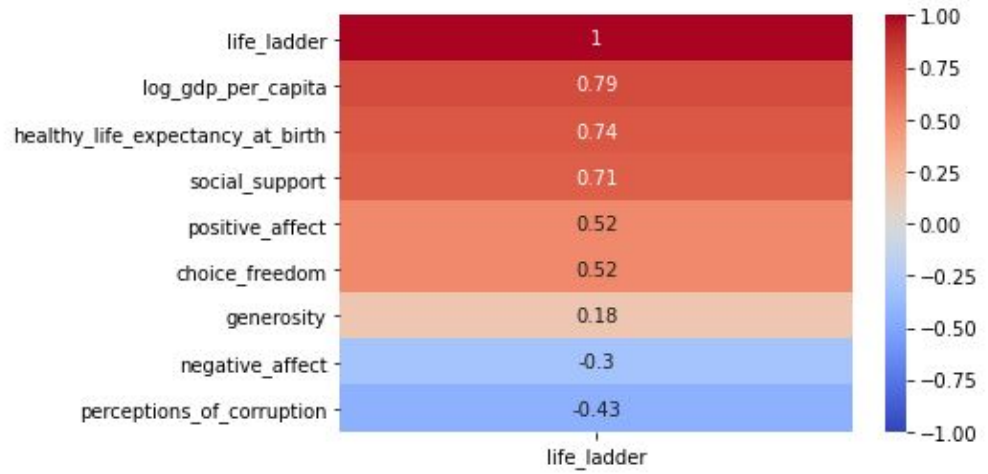
Life ladder distribution



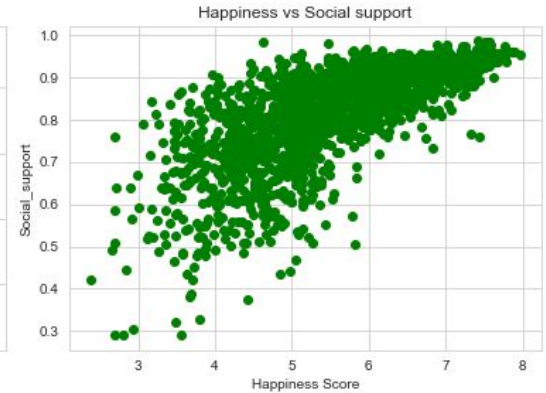
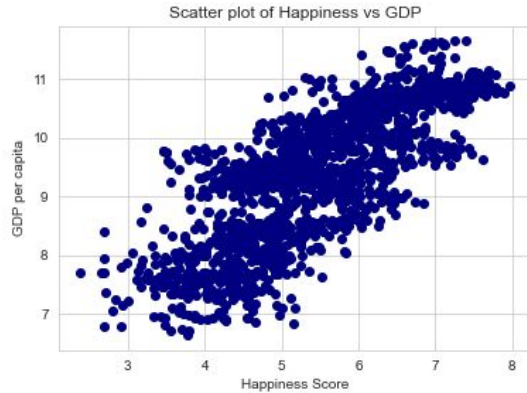
Economic features - Life ladder correlation



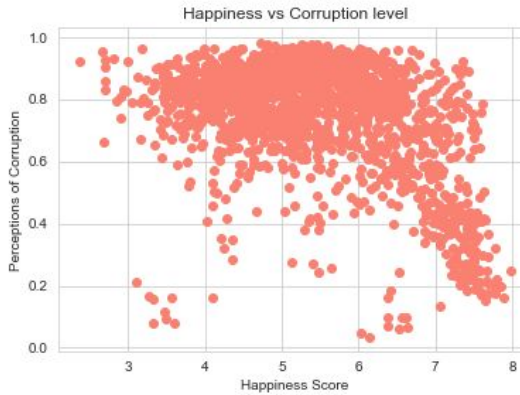
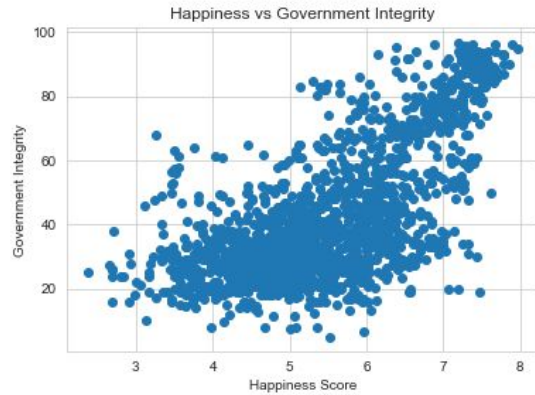
Happiness features - Life ladder correlation



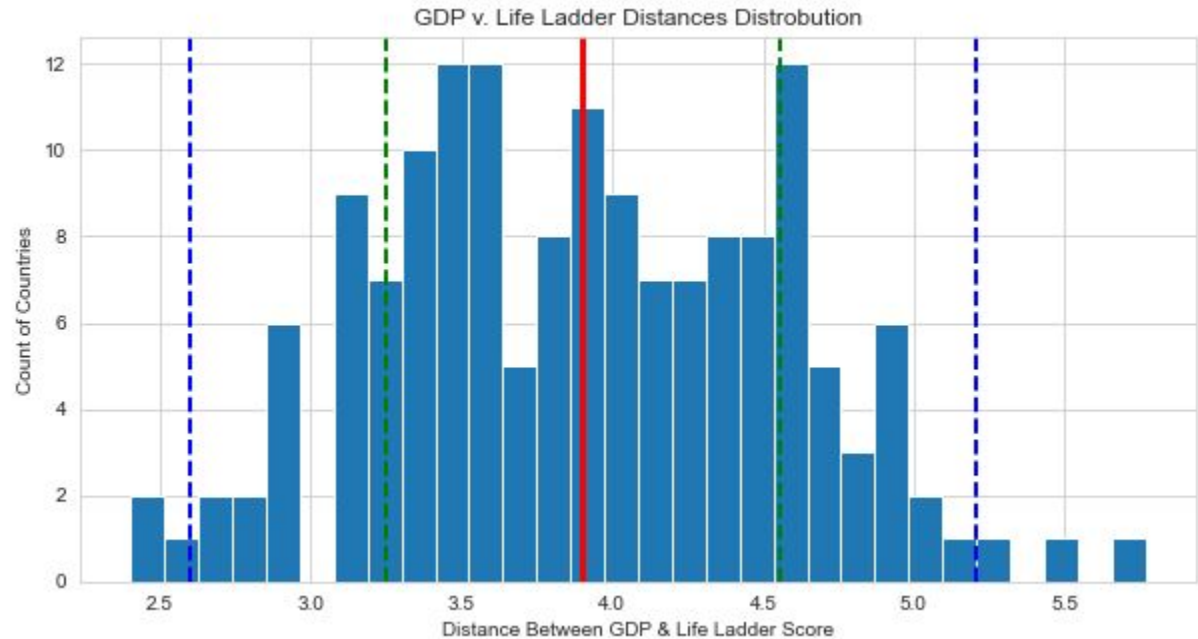
Happiness correlation with best features



Happiness correlation with best features *(continued)*

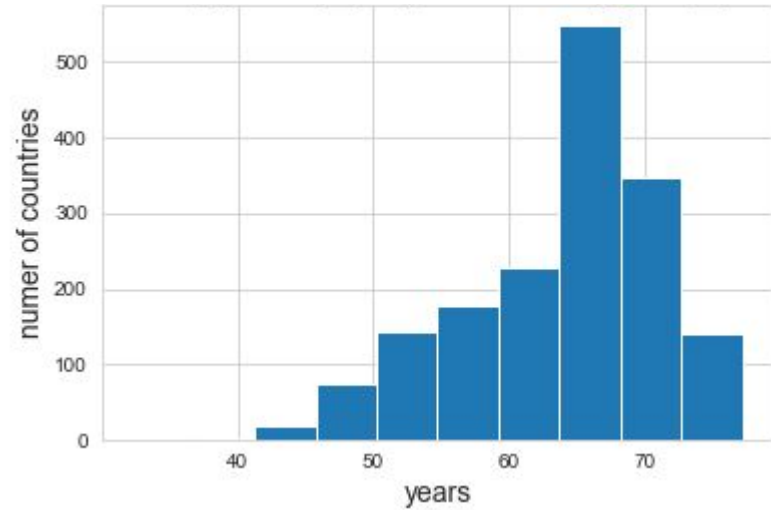


GDP vs Life ladder distance distribution

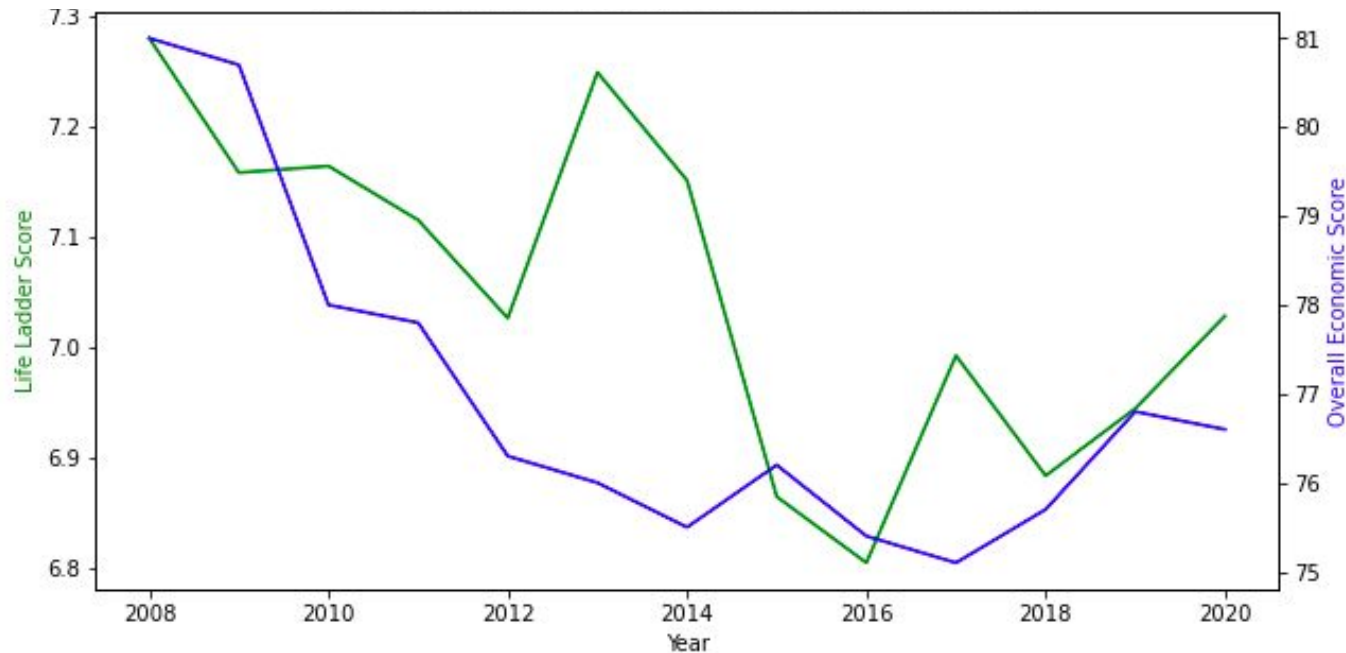




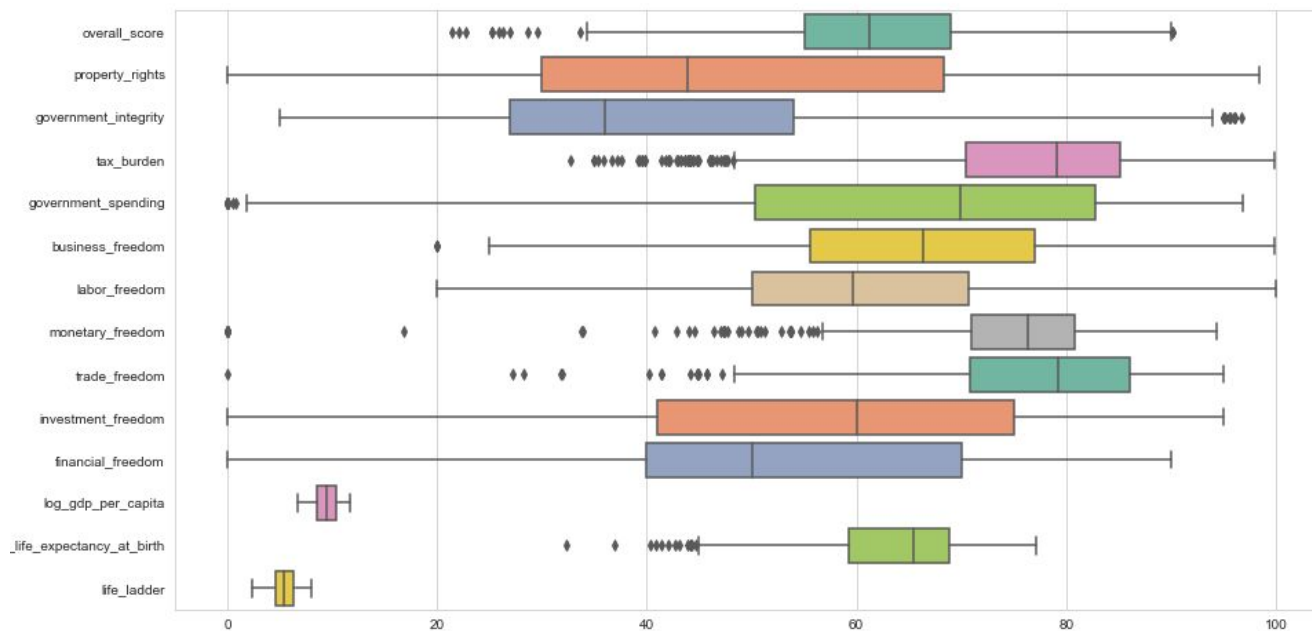
Healthy life expectancy at birth



United States: Happiness and Overall Economic Score by Year

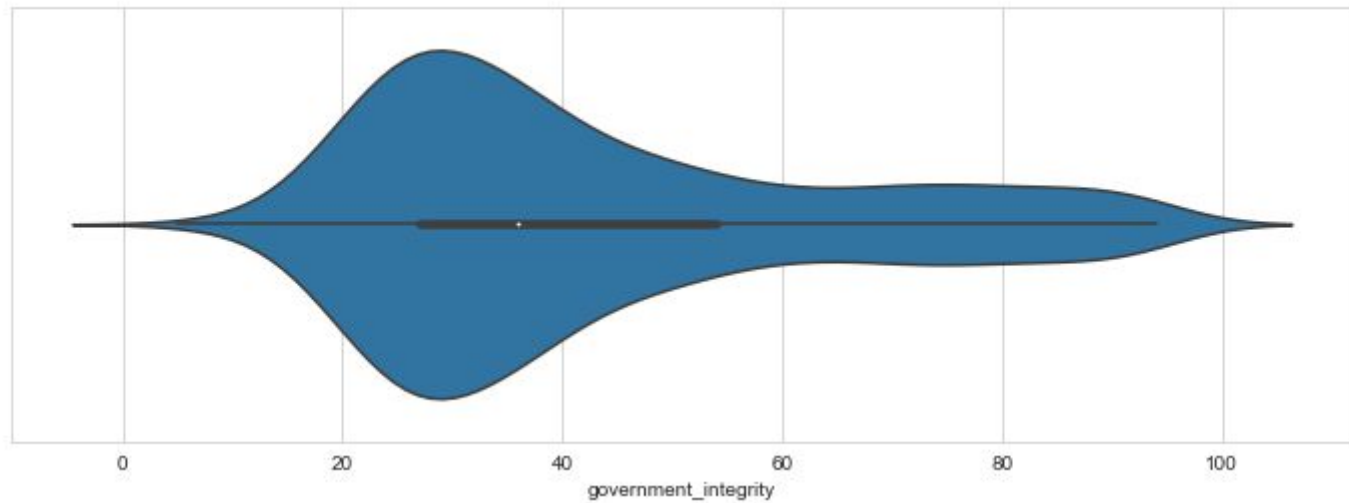


Distribution by feature





Government integrity





Modeling



MODEL	SCORE	MODEL	SCORE
Linear Regression + all features	MSE: 0.28		train score: 0.77
	MAE: 0.41		test score: 0.78
	R2: 0.78	Ridge	MSE: 0.28
			MAE: 0.41
			R2: 0.78
Linear Regression + Economic features	train score: 0.52		
	R2 test score: 0.57		train score: 0.76
			test score: 0.75
Linear Regression + Happiness features	train score: 0.76	Random Forest	MSE: 0.31
	test score: 0.76		MAE: 0.44
			R2: 0.75
	train score: 0.59		
	test score: 0.62		train score: 0.70
Lasso	MSE: 0.47		test score: 0.72
	MAE: 0.56	Ridge Model + Best Features	MSE: 0.34
	R2: 0.62		MAE: 0.46
			R2: 0.72



Conclusion (1 of 2)

- We used four models. Linear Regression, Linear Regression with Lasso, Linear Regression with Ridge, and Random Forest Regressor. Each model has its strengths. Linear Regression is the easiest to explain. Each coefficient on the variable like the gdp per capita is able to explain part of the variation in the dependent variable, which in this case is the life_ladder score. This model is also less likely to overfit , and does not generalize well to new data in comparison to the Random Forest Regressor. Moreover, in terms of a political argument, using a simpler model would allow the information contained to be better explained. While Random Forest is a more powerful model, it can best be explained as a game of 20 questions in a decision tree and can be hard to explain.



Conclusion (2 of 2)

- In terms of settling on a production ready model, the Random Forest Regressor did the best job of predicting happiness, especially on the train score. While Random Forest models generally overfit data, which gives it trouble in generalizing to new data, the R^2 score on the test data was higher than all of our other models. We were able to explain around 86% of the variance in the model. It should also be noted that Linear Regression allowed us to identify several variables that showed greater importance in predicting happiness.