World Happiness Report

Ahsan Fayyaz, Rezwana Kabita & Shanshan Li

Motivation

- What does happiness mean?
- What variables influence happiness around the world?

Motivation

World Happiness Report is an international survey report published by the United Nations to measure the sustainable development of happiness.

- Uses the data from Gallup World Poll survey for more than 150 countries
  - Surveys 1,000 people in each countries to rate their current lives on the scale 0 to 10
- Uses six factors to explain life evaluation:
  - GDP per capita, social support, life expectancy, freedom, generosity, and government corruption.

Goal

- What is common among these countries?
  - Top most happiest nations in the world?
  - Top least happiest nations in the world?
- Take some data and build a happiness score predictor

Data

- World happiness from World Happiness Report
  - Includes the countries, happiness score, GDP per capita, social support, life expectancy, freedom, generosity, and government corruption.
- COVID19 cases from World Health Organization
  - Includes the countries, and total number of confirmed, deaths and recovered cases.
- Crude suicide rates from Kaggle
  - Includes the countries, and the suicide rate per 100,000 people.
- Drinking water services from Kaggle
  - Includes the countries, and total number of people with access to clean drinking water per 100 people.
- Medical doctors from Kaggle
  - Includes the countries, and total number of people that is a medical doctor per 10,000 people.

Data cleaning

Limitations:

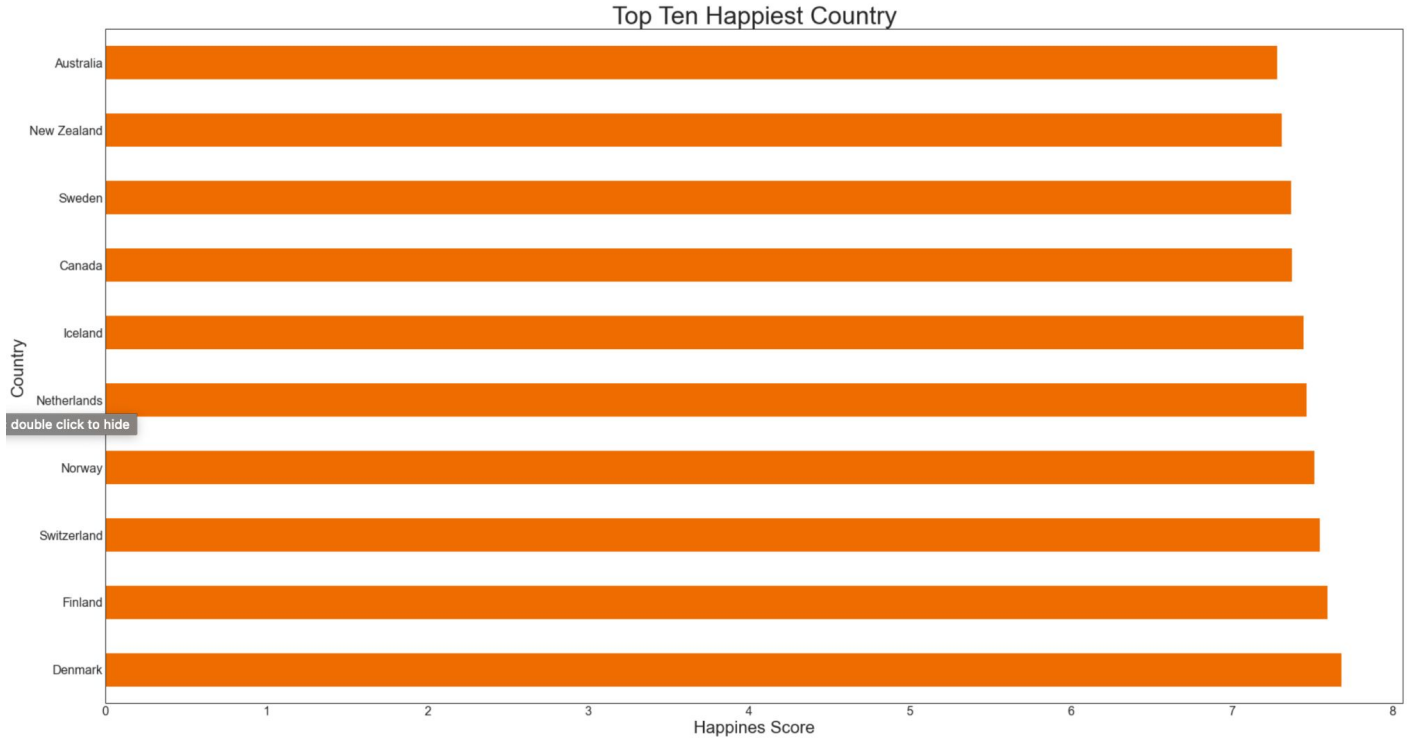
- Multiple datasets from different sources.
- Same countries named differently across datasets

Cleaning of raw datasets:

- Wrote a python script to clean raw data & save it as csv files
- Drop all unnecessary columns and null values.
- Renamed columns
- Group all data in the world happiness dataset by ‘country’
- Combined all other datasets as features into world happiness dataset.

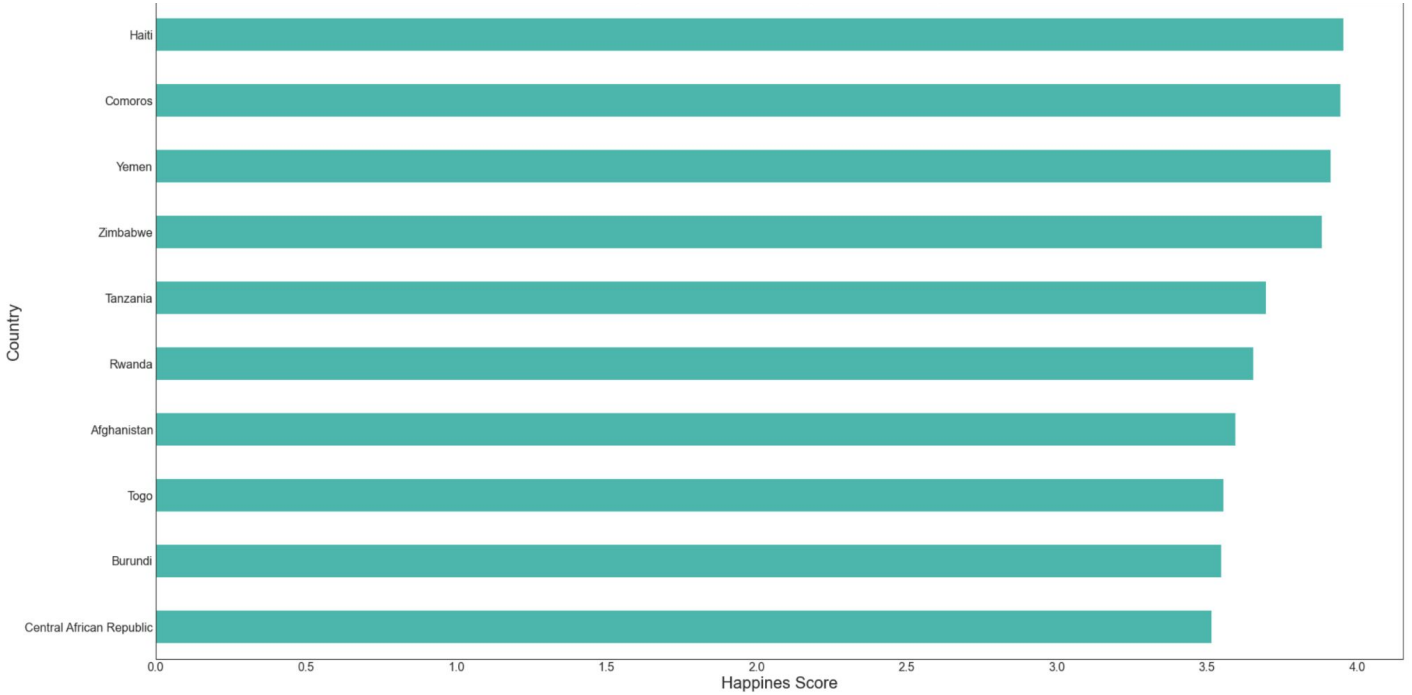
Top Ten Happiest Countries

country	happiness_score
Denmark	7.680305
Finland	7.597216
Switzerland	7.548228
Norway	7.512566
Netherlands	7.466292
Iceland	7.446581
Canada	7.376319
Sweden	7.369527
New Zealand	7.310304
Australia	7.282024



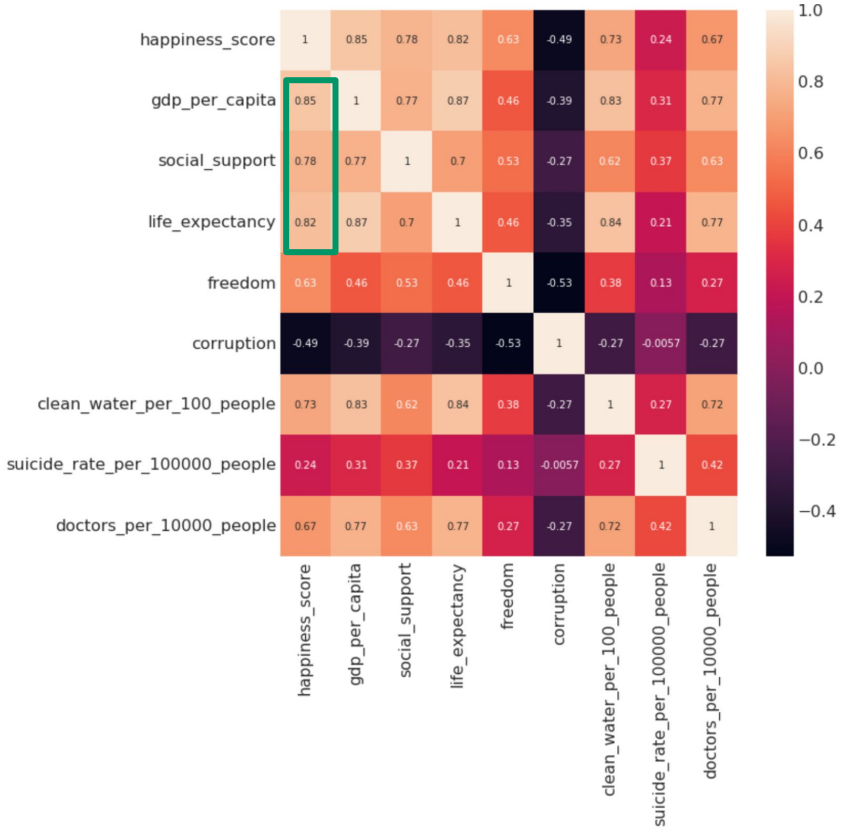
Top Ten Least Happiest Countries

country	happiness_score
Central African Republic	3.514954
Burundi	3.548124
Togo	3.555060
Afghanistan	3.594628
Rwanda	3.654473
Tanzania	3.697284
Zimbabwe	3.882689
Yemen	3.912124
Comoros	3.943963
Haiti	3.954194



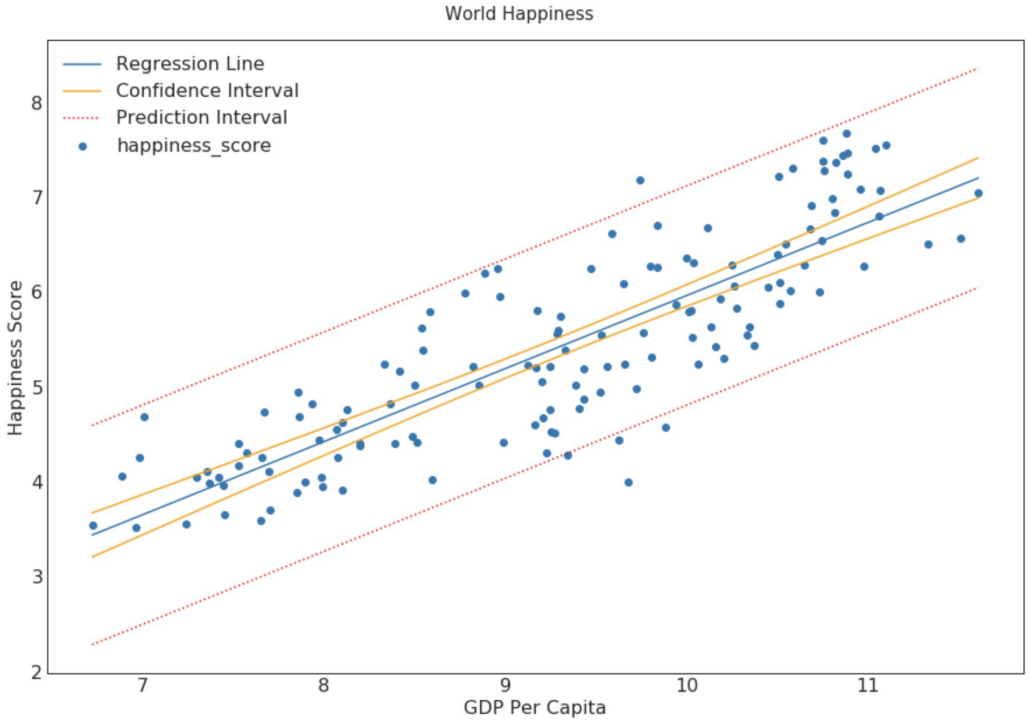
Top Variables Contributing to Happiness

- Economy (GDP per Capita)
- Social Support
- Healthy Life Expectancy



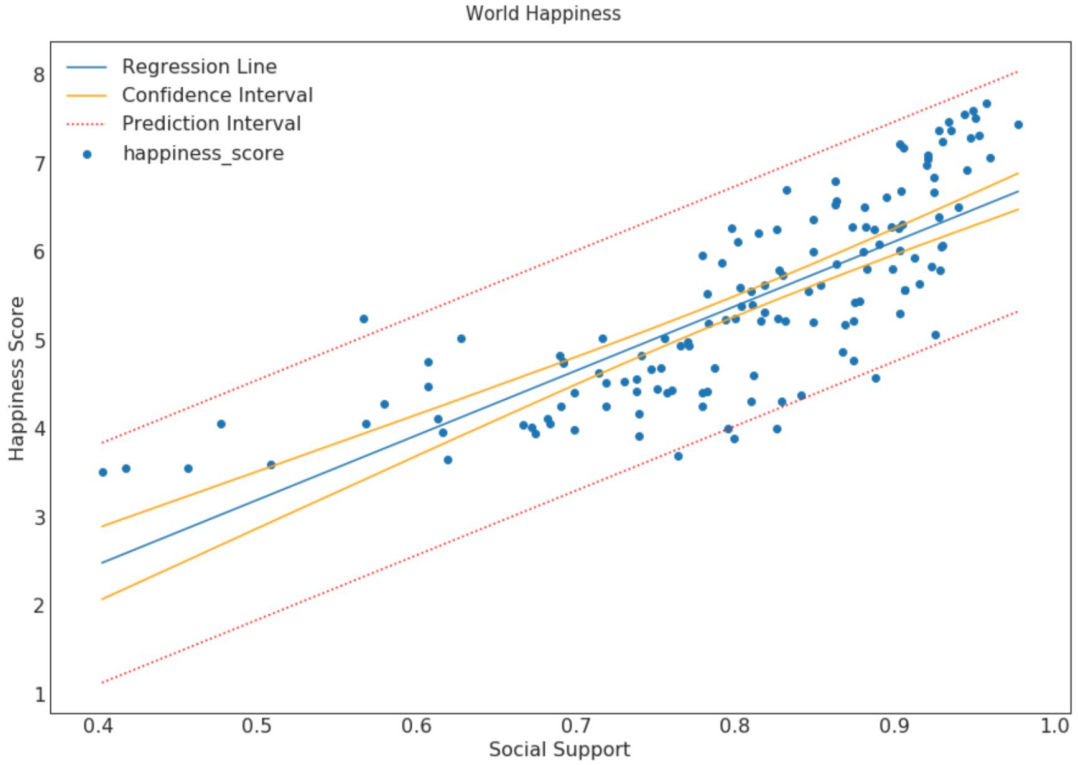
Happiness Score & GDP Per Capita

- Slope = 0.770
- R-squared value = 59.4%



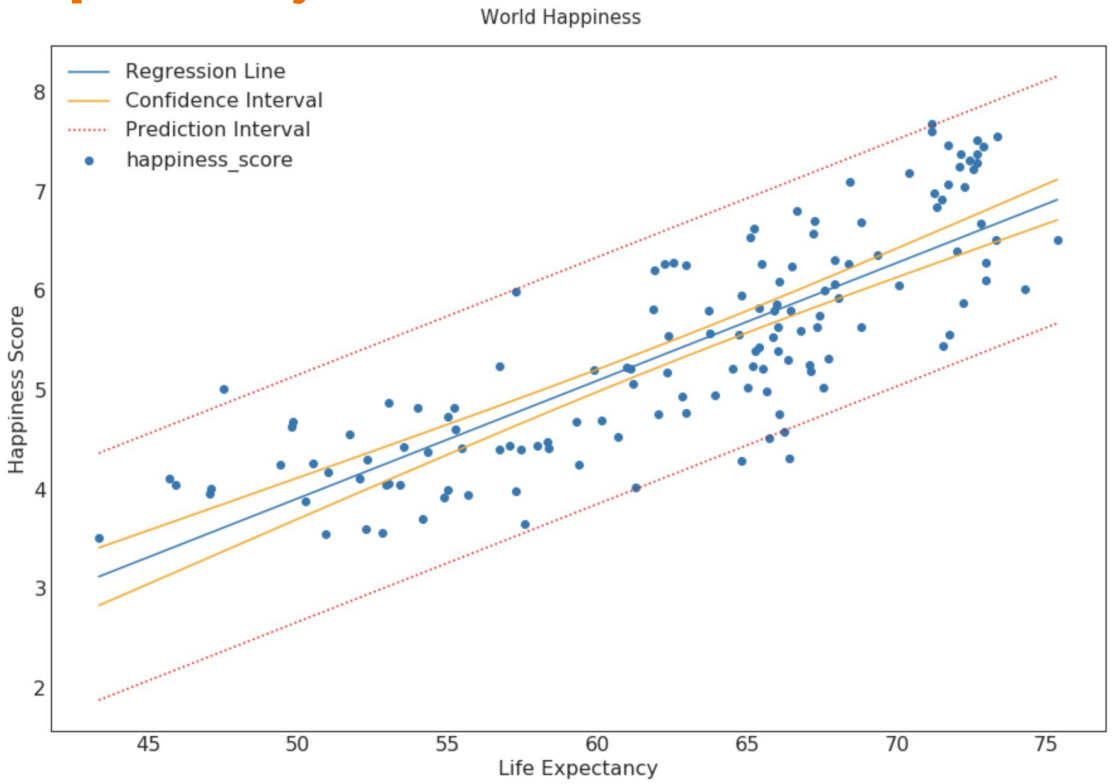
Happiness Score & Social Support

- Slope = 7.30
- R-squared value = 53.42%



Happiness Score & Life Expectancy

- Slope = 0.118
- R-squared value = 1.4%



Variables Used to Measure Happiness

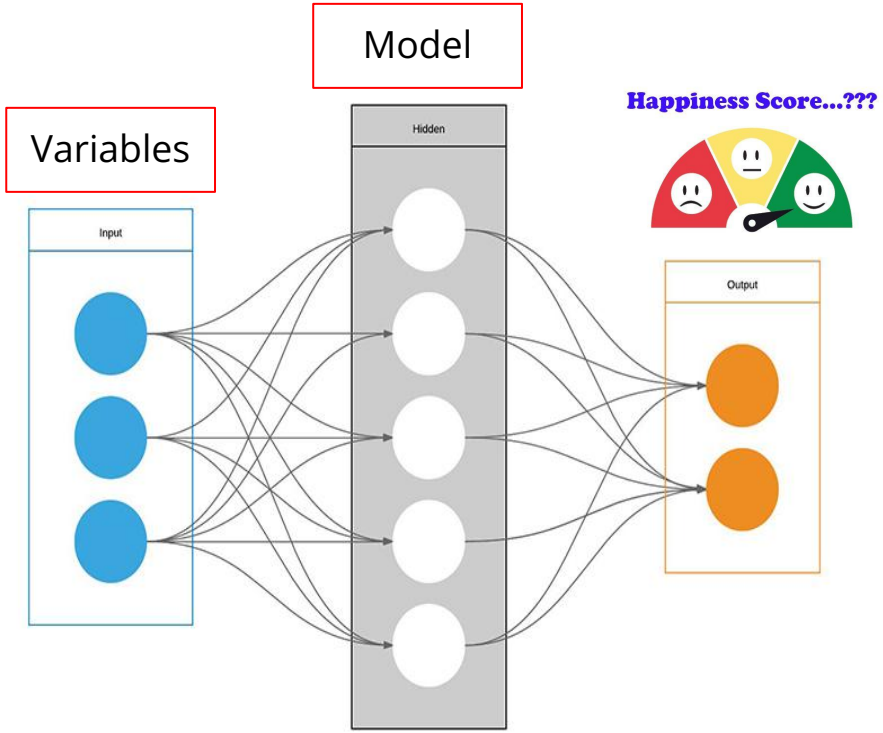
After running our visualizations, we selected these to build our predictive model upon:

- **Economy (Gdp per capita)**
- **Family/Social Support**
- **Healthy Life Expectancy**
- **Freedom**
- **Clean Water per 100 people**
- **Doctors per 10000 people**

Avoided overfitting by using 5-6 variables instead of 10-12

Models

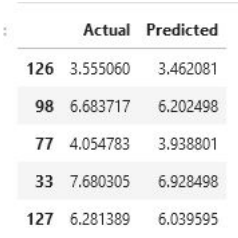
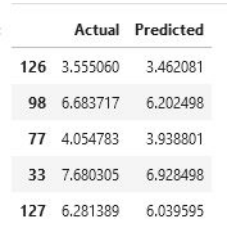
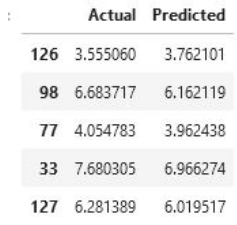
- Happiness Score Predictors
- Data fits Supervised Learning Models
- Continuous Data
- Regression models
  - Linear Regression
  - ElasticNet Regression
  - Ridge Regression
  - Lasso Regression
  - Bayesian Ridge Regression



Evaluation of Models

MODEL	Score/Accuracy (R-squared)
Linear Regression	86.1 %
Linear Regression with CV (10 folds)	69.2 %
ElasticNet Regression	86.1 %
Ridge Regression	65.9 %
Lasso Regression	61.9 %
Bayesian Ridge Regression	84.8 %

Top 3 Most Accurate Predictor Models

Linear Regression	86.1 %	
ElasticNet Regression	86.1 %	
Bayesian Ridge Regression	84.8 %	

Future Work

- Collect more data that is related to happiness
  - Collect data that can fit to classification models
- Perform even more rigorous data-cleaning
  - Make sure no rows are lost when dropping nulls
- Experiment with other Supervised Learning models
- Figure out different metrics to report happiness score
  - What considers happy vs unhappy?
  - What is the boundary line?

THANK YOU!

Any questions?

References

- https://hellogiggles.com/lifestyle/23-things-to-be-happy/
- https://www.countryliving.com/life/g4815/happy-life-quotes/
- https://knowledge.wharton.upenn.edu/article/the-downside-of-happiness/
- https://www.123rf.com/stock-photo/happy_world.html?sti=mjl97ua7wo90576lnh|
- https://enermaps.eu/data-tool/
- https://enermaps.eu/data-tool/
- https://www.pngitem.com/middle/oiTomw_step-1-data-cleansing-and-mining-data-cleaning/
- https://www.latentview.com/blog/predictive-model-data-analytics-efficiency/
- https://icon-library.com/icon/icon-learning-5.html
- https://i.pinimg.com/originals/af/d7/a6/afd7a63b6273439e3ca5927b0e22fca5.jpg
- https://www.flaticon.com/free-icon/accuracy_2519421
- https://adioma.com/icons/machine-learning