World Happiness!

by: Anita and Chadwick

Overview

Analyzing socio-economic and political data to determine how happy a country's citizens are.

Goals:

1. Use regression to identify the most important factors to overall country happiness.
2. Use the model to predict a country's happiness (for the following year).

Process

Cleaning

pandas

python

EVALUATION
Regression

Analysis
plotly

NumPy

Visualization

Seaborn

matplotlib

sklearn
scipy
qgrid
statsmodels

jupyter

kaggle

Data – Kaggle

- **Dataset of indicators**
  - List of socio-econ. and political factors
  - Citizen sentiment - 2018
- Variables scaled on a 0-1 basis except for GDP per capita and Social Support (0-2)

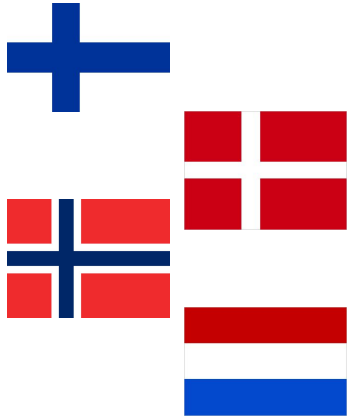
How is happiness measured?

"Nationally representative samples of respondents are asked to think of a ladder, with the best possible life for them being a 10, and the worst possible life being a 0.

They are then asked to rate their own current lives on that 0 to 10 scale"

4 Happiest Countries

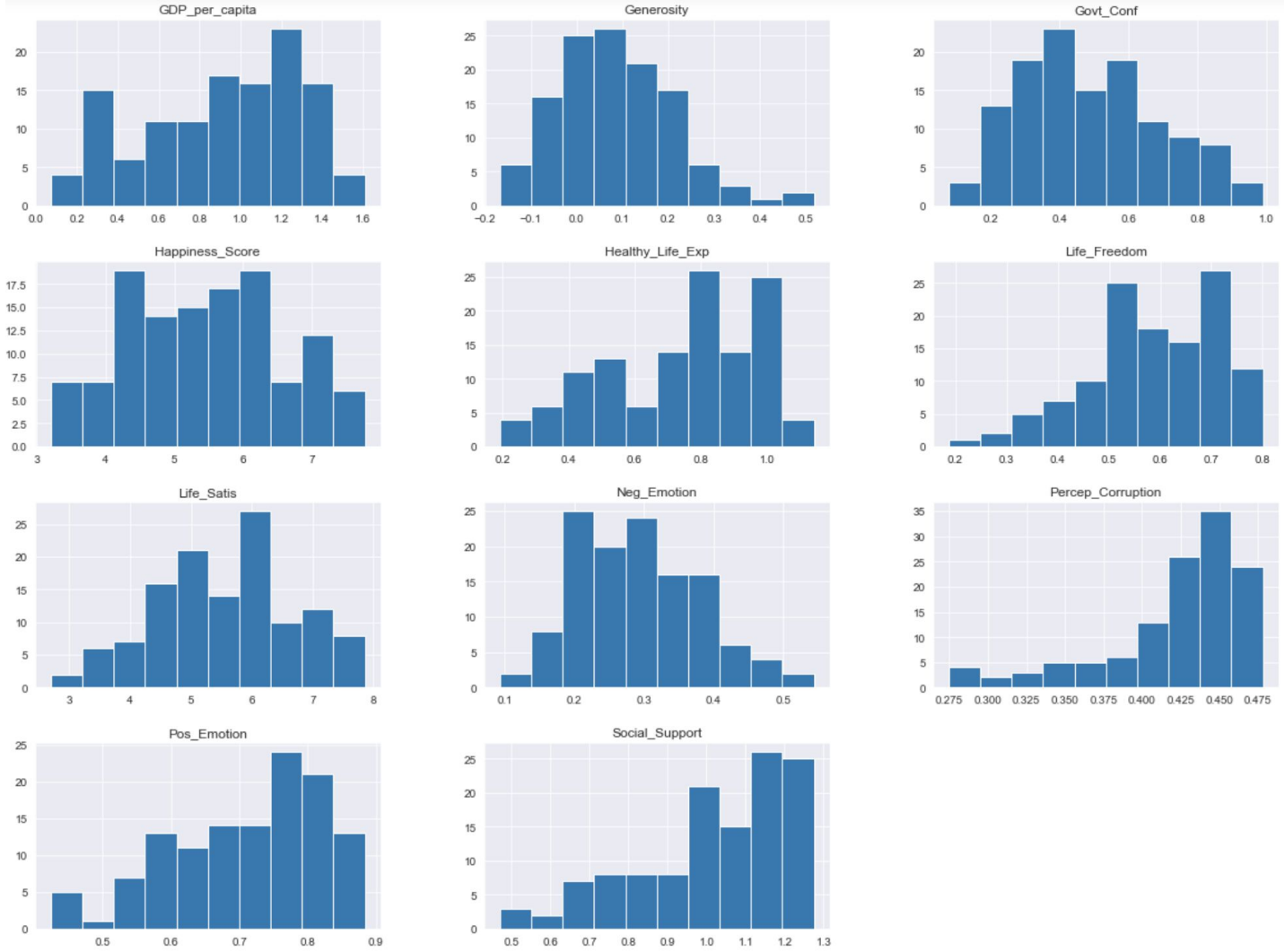
1. Finland
2. Denmark
3. Norway
4. Netherlands



4 Unhappiest Countries

1. Afghanistan
2. Tanzania
3. Rwanda
4. Yemen

Evaluation



Distribution of each indicator:

- ➔ gdp per capita
- ➔ generosity
- ➔ govt confidence
- ➔ happiness score
- ➔ healthy life exp
- ➔ life freedom
- ➔ life satisfaction
- ➔ negative emotion
- ➔ perception of corruption
- ➔ positive emotion
- ➔ social support

Results of EDA

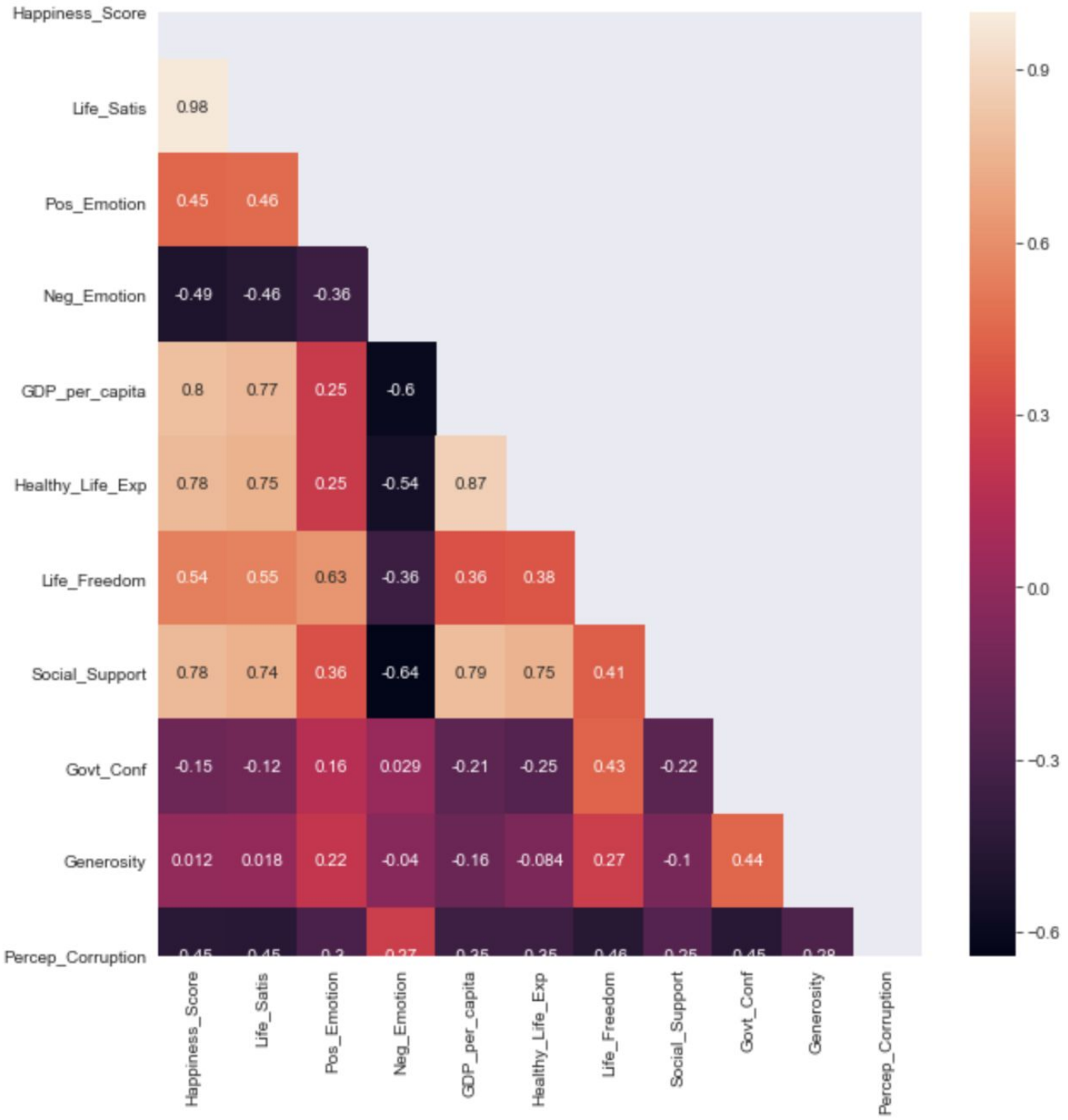
Happiness Score (from 1 to 10)

Mean Score for 2018:

5.47

- Top 3 Strong positive correlation with happiness score
  1. Life Satisfaction
  2. GDP per capita
  3. Social Support, Healthy Life Exp

- Top Negative correlation with happiness score
  1. Negative emotions
  2. Perception of Corruption



First Model Run

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2.7569	1.222	2.256	0.027	0.326	5.188
Pos_Emotion	0.3962	0.700	0.566	0.573	-0.995	1.788
Neg_Emotion	1.8869	0.842	2.240	0.028	0.211	3.562
GDP_per_capita	1.3072	0.355	3.683	0.000	0.601	2.013
Healthy_Life_Exp	0.0355	0.568	0.062	0.950	-1.095	1.166
Life_Freedom	2.3137	0.773	2.992	0.004	0.775	3.852
Social_Support	1.7391	0.529	3.290	0.001	0.687	2.791
Govt_Conf	-1.4448	0.469	-3.083	0.003	-2.377	-0.512
Generosity	0.8085	0.529	1.529	0.130	-0.243	1.860
Percep_Corruption	-4.3797	1.766	-2.481	0.015	-7.892	-0.867

Omnibus:	2.861	Durbin-Watson:	2.056
Prob(Omnibus):	0.239	Jarque-Bera (JB):	2.309
Skew:	-0.375	Prob(JB):	0.315
Kurtosis:	3.196	Cond. No.	82.2

OLS Regression Results

Dep. Variable:	Happiness_Score	R-squared:	0.777
Model:	OLS	Adj. R-squared:	0.753
Method:	Least Squares	F-statistic:	31.75
Date:	Thu, 16 Jan 2020	Prob (F-statistic):	3.69e-23
Time:	13:53:52	Log-Likelihood:	-68.812
No. Observations:	92	AIC:	157.6
Df Residuals:	82	BIC:	182.8
Df Model:	9		
Covariance Type:	nonrobust		

R^2 = .777
77.7% of variance explained by the model

- ➔ Dropped 3 variables where p-value>0.05 variables
  - ◆ the coefficients were not explanatory

- ➔ OLS Linear Regression
- ➔ Life_Satisfaction dropped before running
  - ◆ seen as a proxy for happiness score

Refining and Refitting the Model

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2.7569	1.222	2.256	0.027	0.326	5.188
Pos_Emotion	0.3962	0.700	0.566	0.573	-0.995	1.788
Neg_Emotion	1.8869	0.842	2.240	0.028	0.211	3.562
GDP_per_capita	1.3072	0.355	3.683	0.000	0.601	2.013
Healthy_Life_Exp	0.0355	0.568	0.062	0.950	-1.095	1.166
Life_Freedom	2.3137	0.773	2.992	0.004	0.775	3.852
Social_Support	1.7391	0.529	3.290	0.001	0.687	2.791
Govt_Conf	-1.4448	0.469	-3.083	0.003	-2.377	-0.512
Generosity	0.8085	0.529	1.529	0.130	-0.243	1.860
Percep_Corruption	-4.3797	1.766	-2.481	0.015	-7.892	-0.867

Omnibus:	2.861	Durbin-Watson:	2.056
Prob(Omnibus):	0.239	Jarque-Bera (JB):	2.309
Skew:	-0.375	Prob(JB):	0.315
Kurtosis:	3.196	Cond. No.	82.2

➔ Original 9 variables from first model run

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	3.4376	1.044	3.293	0.001	1.362	5.513
Neg_Emotion	1.6460	0.826	1.994	0.049	0.005	3.288
GDP_per_capita	1.1729	0.267	4.389	0.000	0.642	1.704
Life_Freedom	2.5989	0.647	4.017	0.000	1.313	3.885
Social_Support	1.8188	0.518	3.512	0.001	0.789	2.848
Govt_Conf	-1.4240	0.422	-3.372	0.001	-2.264	-0.584
Percep_Corruption	-5.2601	1.634	-3.218	0.002	-8.510	-2.011

Omnibus:	3.704	Durbin-Watson:	1.956
Prob(Omnibus):	0.157	Jarque-Bera (JB):	3.030
Skew:	-0.415	Prob(JB):	0.220
Kurtosis:	3.320	Cond. No.	64.7

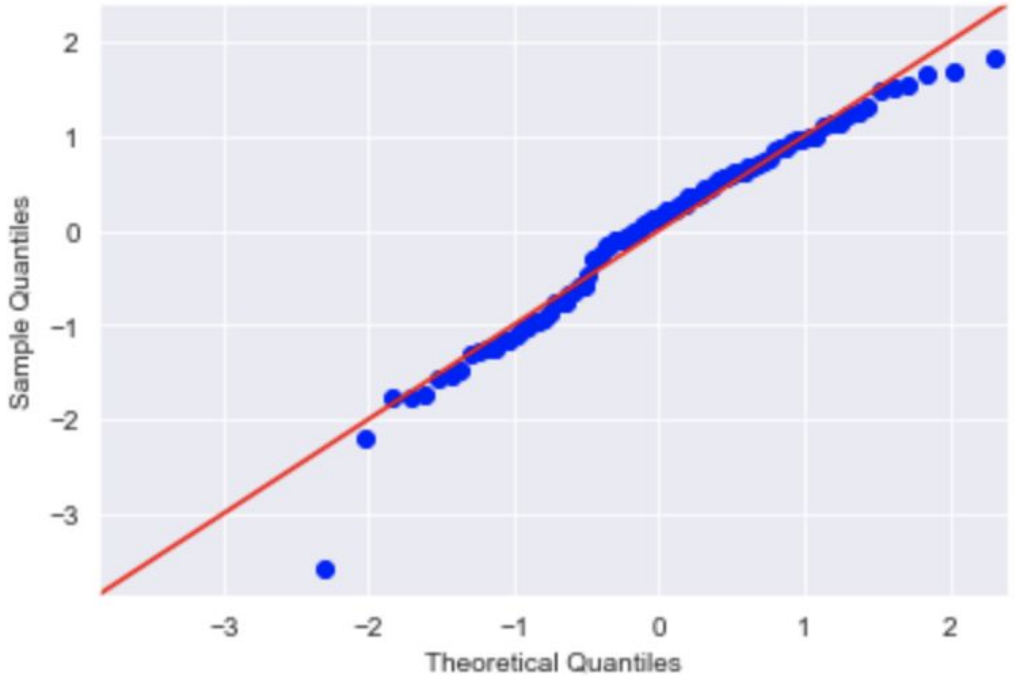
➔ Remaining 6 variables after adjusting based on p-values

➔ Pos_Emotion, Healthy_Life_exp, Generosity are dropped

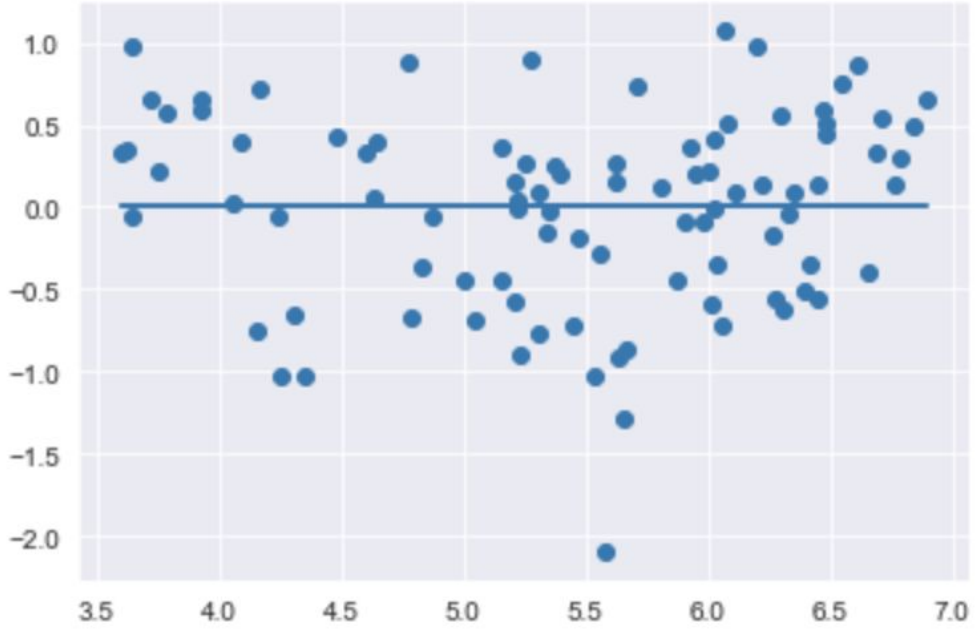
Refining and Refitting the Model

Checking the error terms to see if assumptions are met:

Q-Q Plot: Check for normal distribution in residuals



Check for homoscedasticity in the residuals



Final Model

OLS Regression Results

Dep. Variable:	Happiness_Score	**R-squared:**	0.585
Model:	OLS	**Adj. R-squared:**	0.569
Method:	Least Squares	**F-statistic:**	37.10
Date:	Thu, 16 Jan 2020	**Prob (F-statistic):**	4.59e-15
Time:	14:24:52	**Log-Likelihood:**	-81.976
No. Observations:	83	**AIC:**	172.0
Df Residuals:	79	**BIC:**	181.6
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	8.9060	1.264	7.048	0.000	6.391	11.421
Life_Freedom	5.5551	0.676	8.222	0.000	4.210	6.900
Govt_Conf	-3.4179	0.435	-7.860	0.000	-4.283	-2.552
Percep_Corruption	-11.8233	2.615	-4.521	0.000	-17.029	-6.618

Omnibus:	0.183	**Durbin-Watson:**	1.939
Prob(Omnibus):	0.912	**Jarque-Bera (JB):**	0.278
Skew:	-0.106	**Prob(JB):**	0.870
Kurtosis:	2.811	**Cond. No.**	52.5

$$Y = 8.906 + B_1 5.555 - B_2 3.418 - B_3 11.823$$

Conclusions

$Y = 8.906 + B_0 5.555 - B_1 3.418 - B_2 11.823$

Want a high Freedom score (> .7), medium Gov't Confidence score (< .60), and a low Perception of Corruption score (< .4) to be a top 10 country

Need the scaled data to predict for the following year

Usefulness of Model and Improvements

Biggest Indicators of Happiness:

- Extent to which Freedom plays a role in happiness
- Extent to which Confidence in the Government plays a role in happiness
- Perception of Corruption in the country

Improvements:

- ➔ Further tests to see if assumptions are met
- ➔ Ridge and Lasso regressions
- ➔ Log transformation
- ➔ Using a stricter VIF score
- ➔ Could test on past years' data

Resources:

- Kaggle
- Wikipedia: World Happiness Report (2018)
- original data:
  - from the World Happiness Report published by the Sustainable Development Solutions Network