

World Happiness Report Dataset

Data Visualization and Prediction

CZ1015 INTRO TO DATA SCIENCE
MINI-PROJECT PRESENTATION

DONE BY: XIAO YANG
WONG CHUN FOONG
KEANE NGUYEN KIM XUYEN

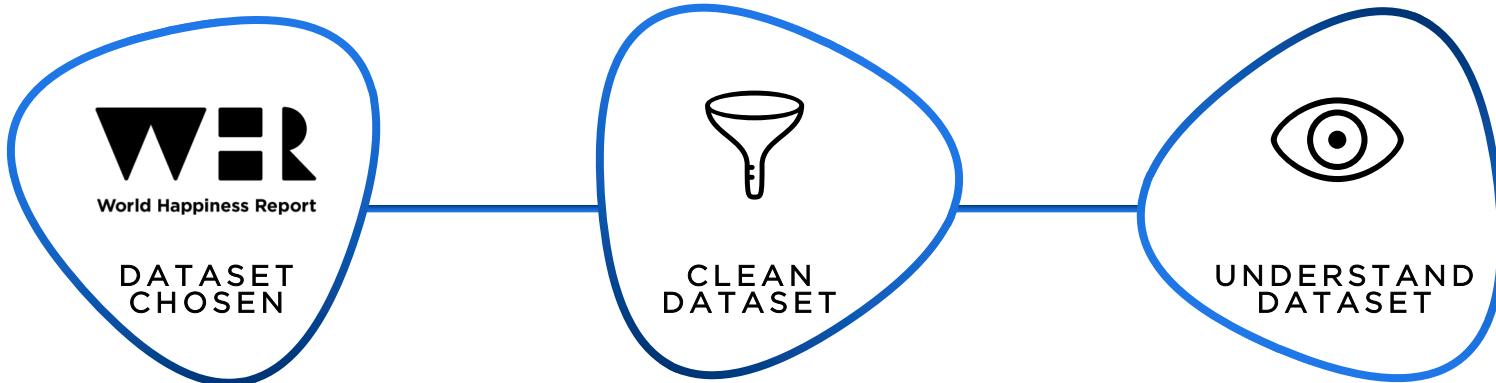


First, let's look into our...

Background

what will be the

World's Happiness Score in the **next 5 years?**



First, we choose the
2019 WHR dataset

Next, we prepare the dataset
by filling N.A.s with Os and
mapping regions and country
codes to respective country

Then, we use the data
to visualize and predict
the data to answer our
above question

Visualization

Analysis

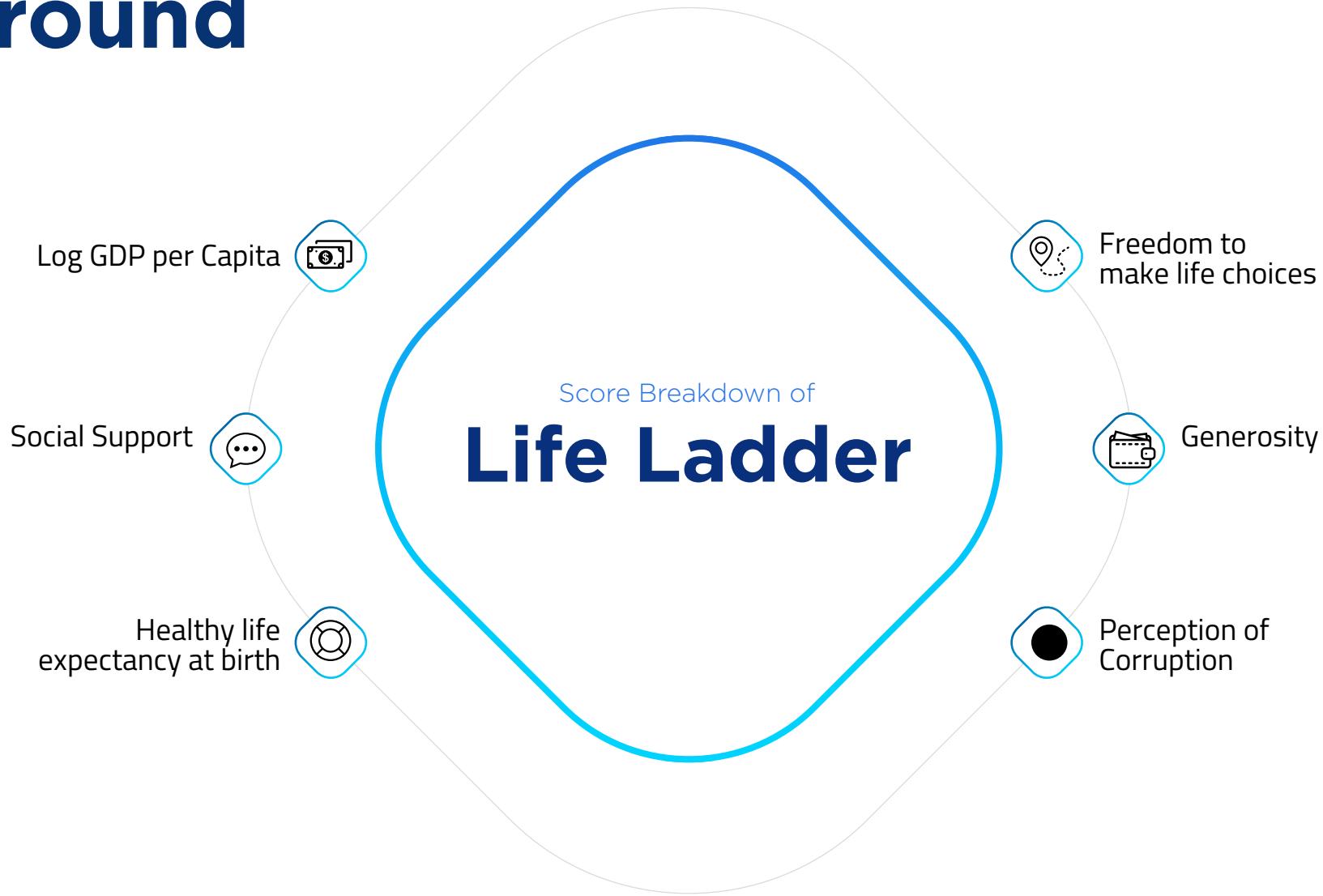
Prediction

Outcome

Conclusion

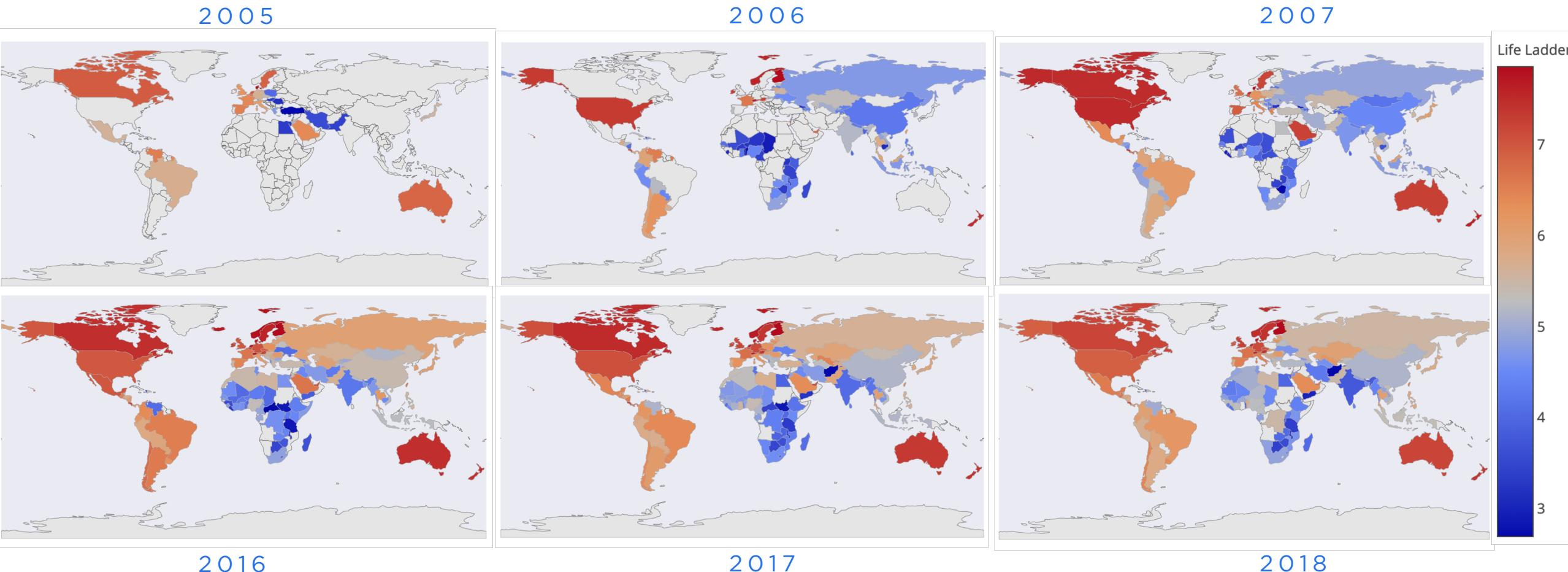
First, let's look into our...

Background



Now, let's take a look at our...

Visualization



2016

2017

2018

Background

Analysis

Prediction

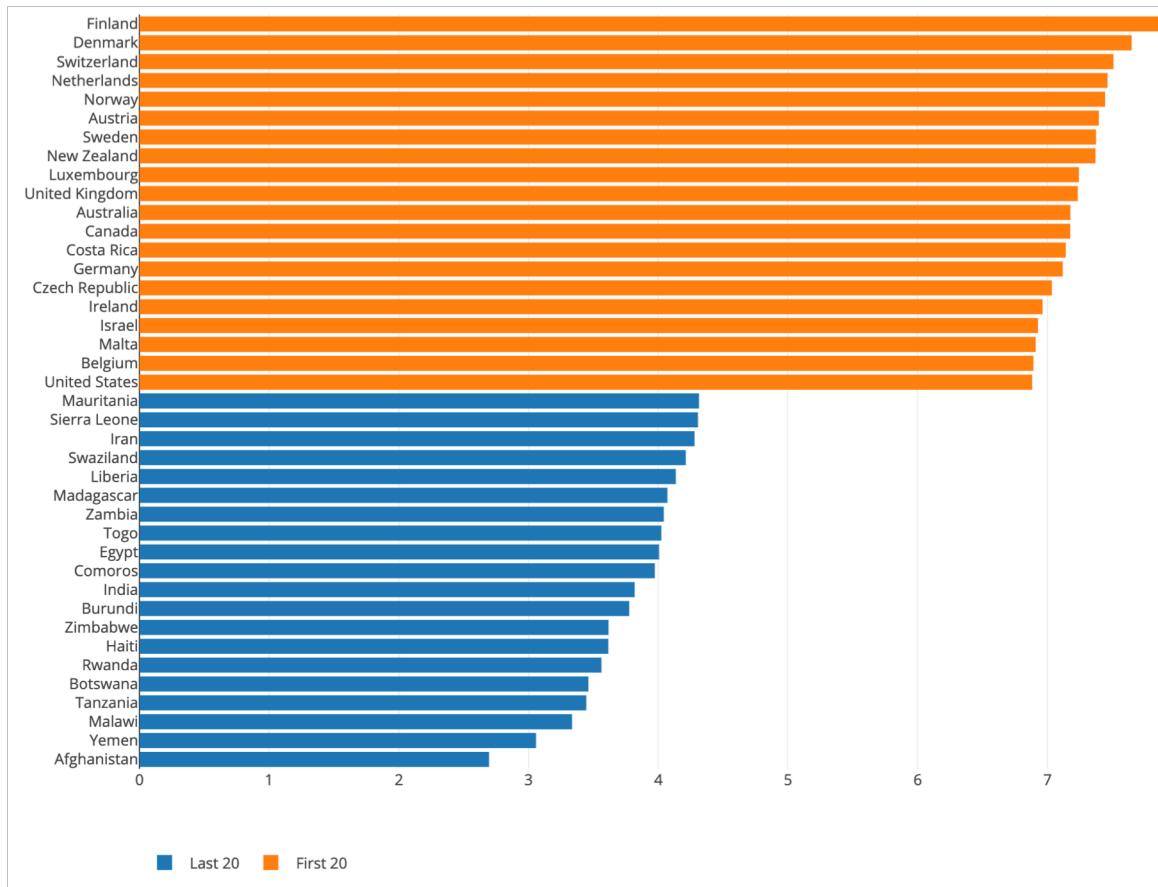
Outcome

Conclusion

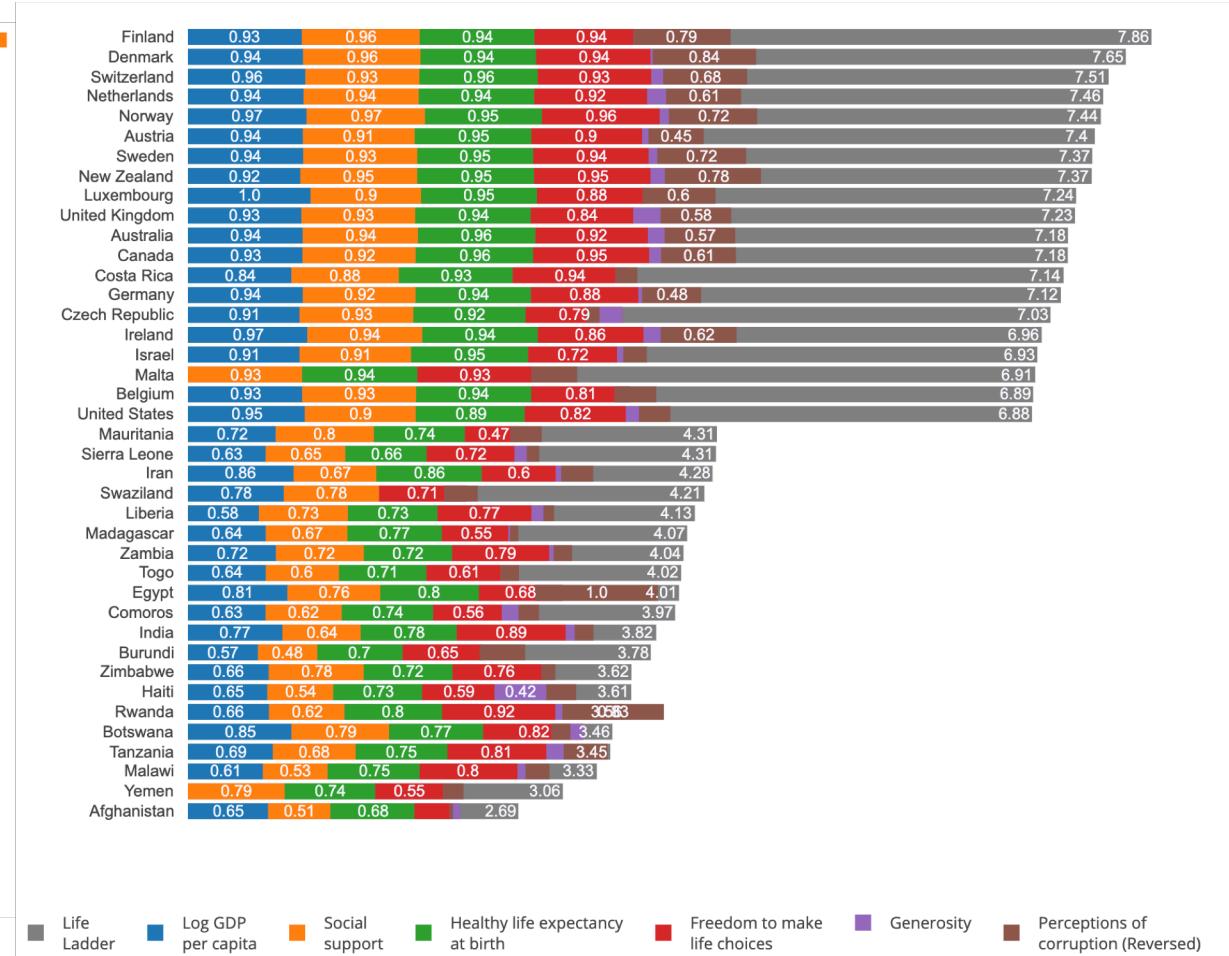
Now, let's take a look at our...

Visualization

Countries' Happiness Scores in 2018 (First and Last 20)



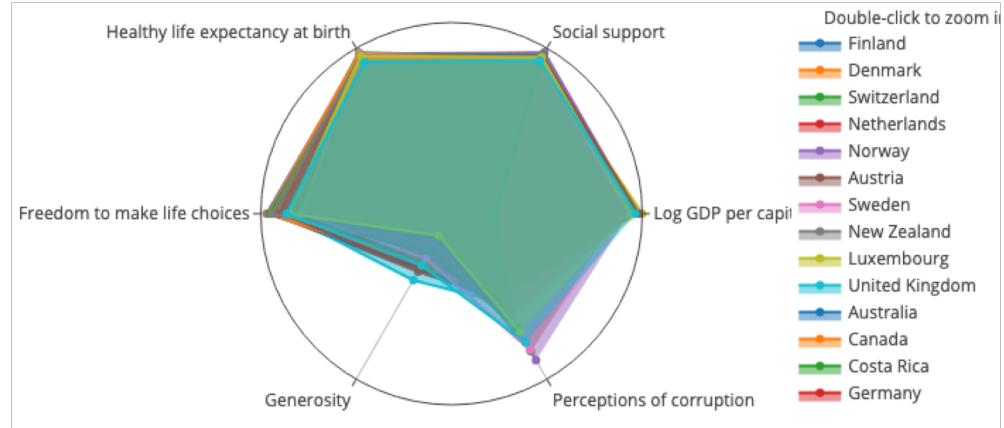
Breakdown of Countries' Happiness Scores in 2018 (First and Last 20)



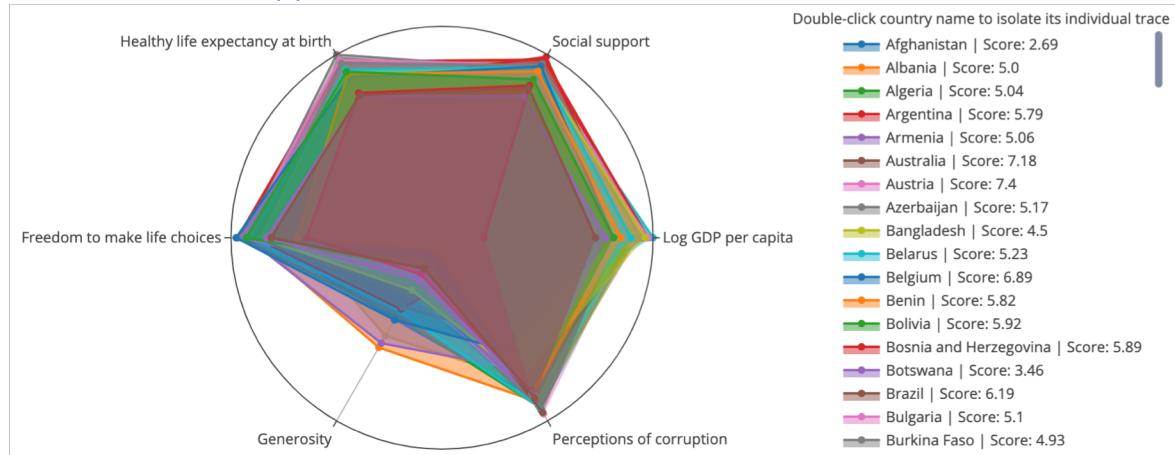
Now, let's take a look at our...

Visualization

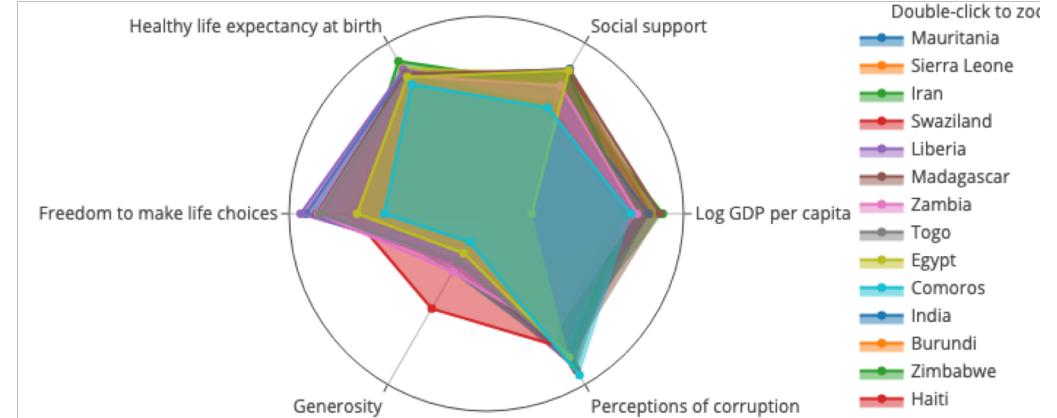
First 20 Happiness Score (Life Ladder) Breakdown into 6 Determinants



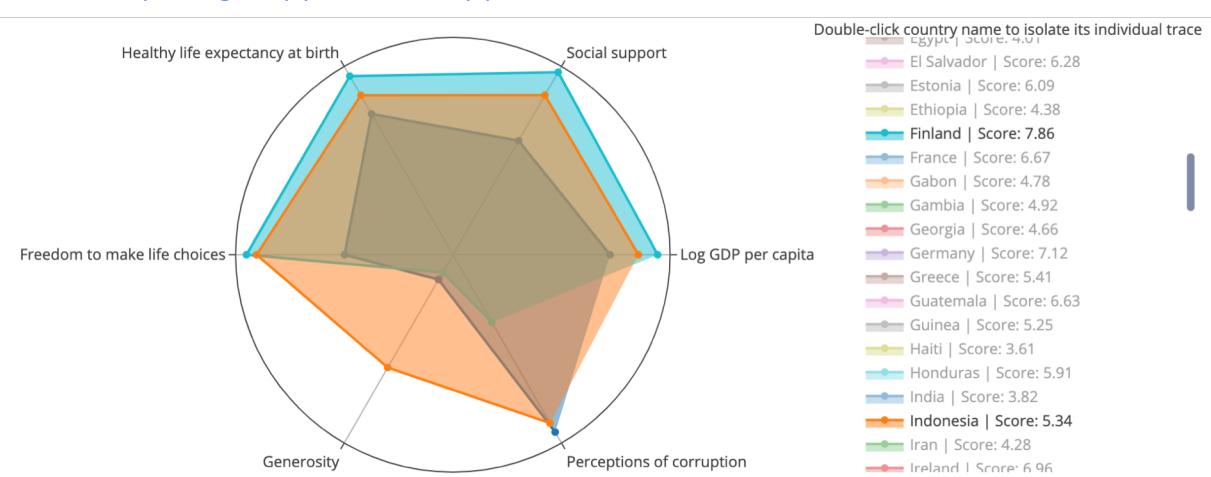
136 Countries' Happiness Score (Life Ladder) Breakdown into 6 Determinants



Last 20 Happiness Score (Life Ladder) Breakdown into 6 Determinants

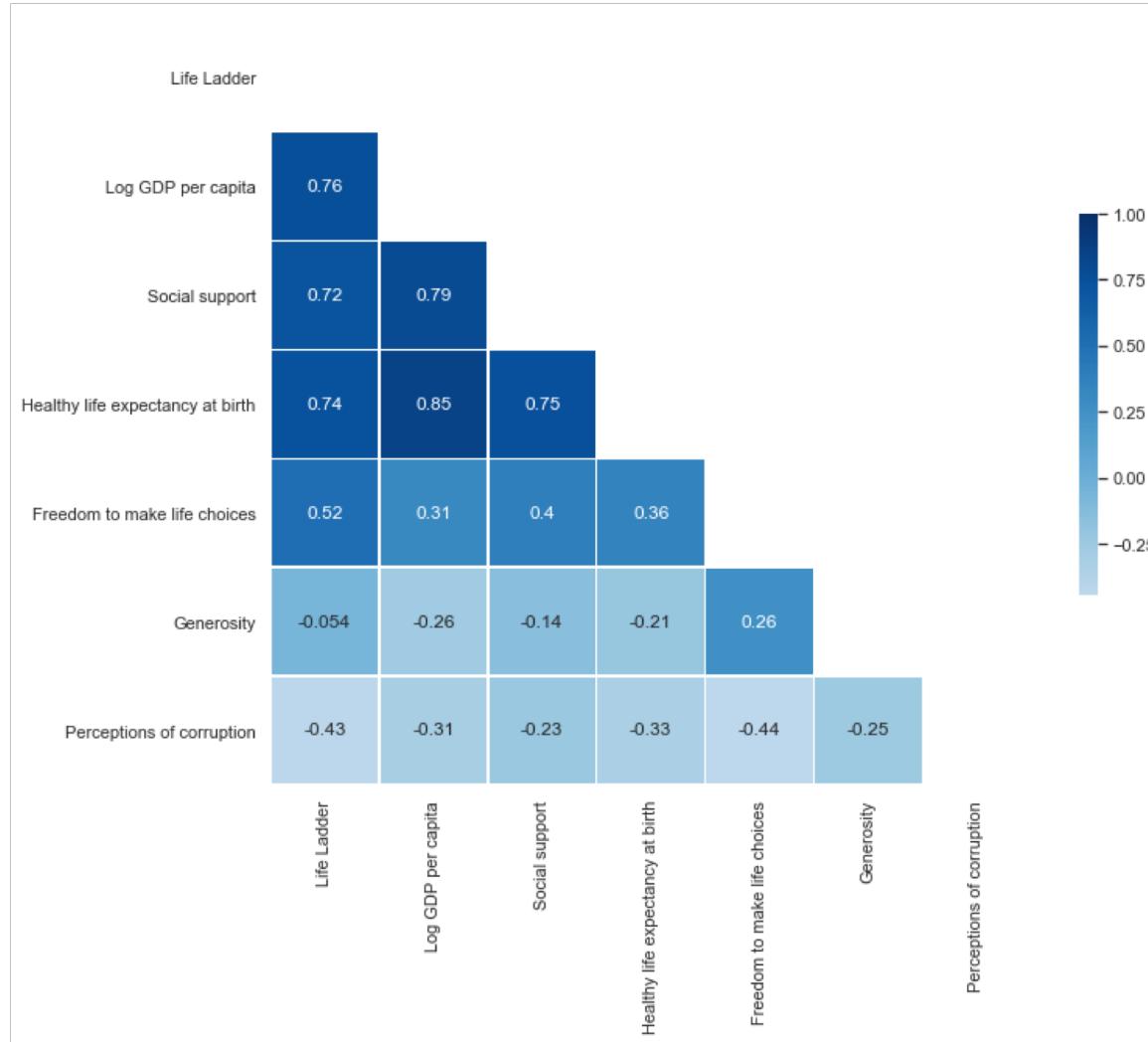


Comparing Happiest, Unhappiest, and Most Generous Countries



Now, let's take a look at our...

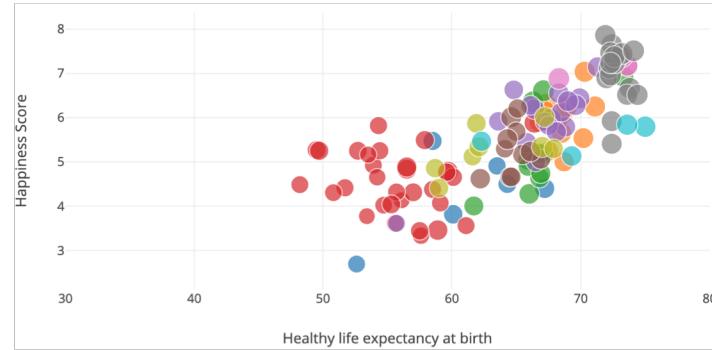
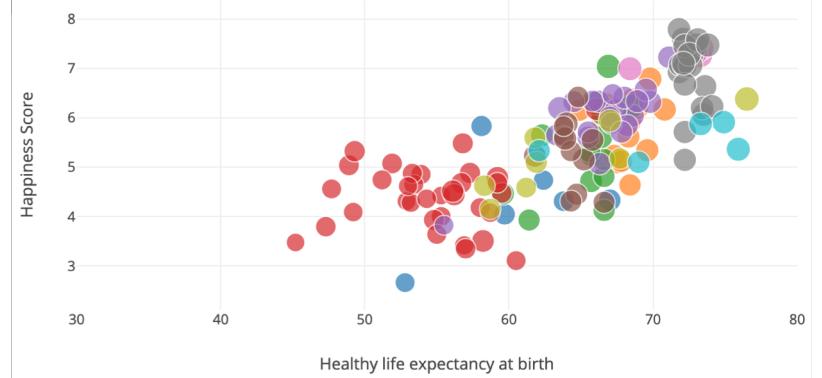
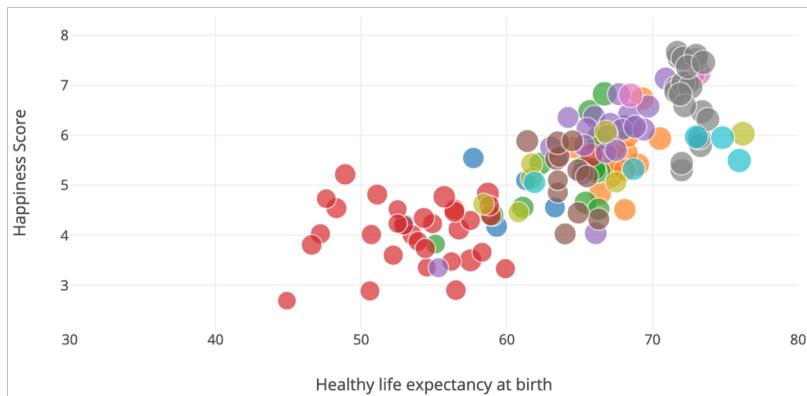
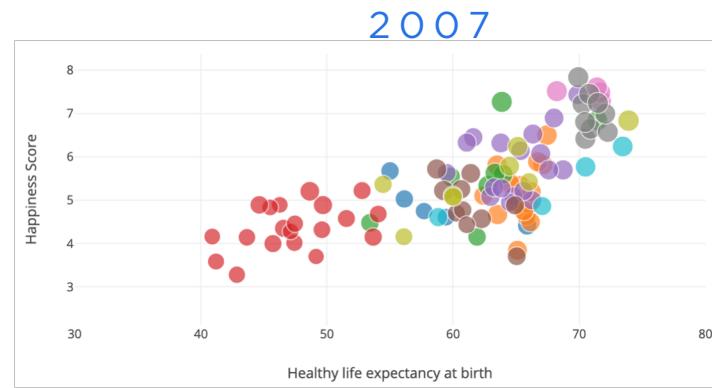
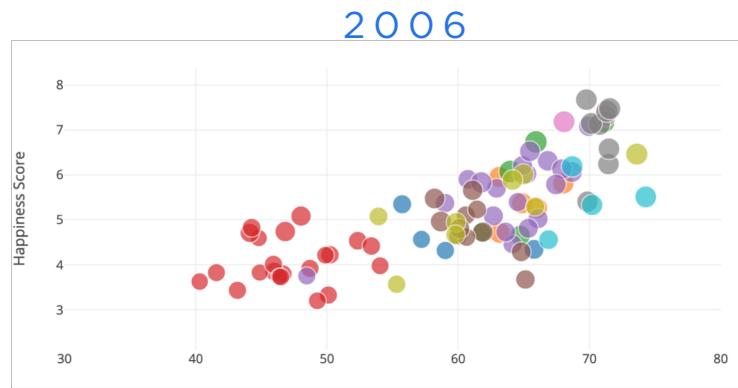
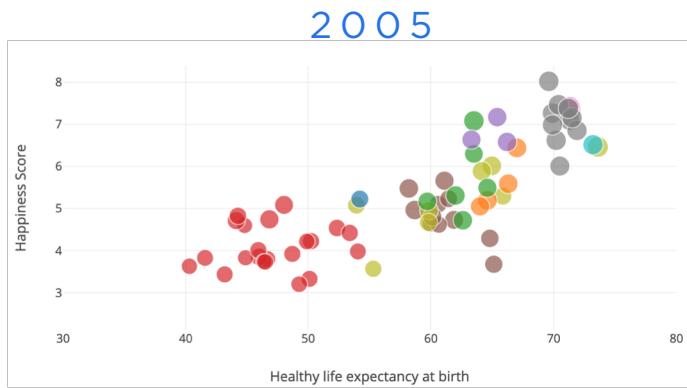
Visualization



Now, let's take a look at our...

Visualization

- South Asia
- Central and Eastern Europe
- Middle East and North Africa
- Sub-Saharan Africa
- Latin America and Caribbean
- Commonwealth of Independent States
- North America and ANZ
- Western Europe
- Southeast Asia
- East Asia



Now, let's take a look at our...

Prediction

1. Predict each index for each country
2. Use the predicted indexes to predict happiness

Now, let's take a look at our...

Prediction

uni-variate linear regression for each column index

	Country name	Year	Life Ladder	Log GDP per capita	Social support	Healthy life expectancy at birth	Freedom to make life choices	Generosity	Perceptions of corruption
1329	Singapore	2006	6.462703	11.088065	0.904329	73.599998	0.756874	0.127692	NaN
1330	Singapore	2007	6.833755	11.133606	0.920632	73.900002	0.866892	0.282540	0.063615
1331	Singapore	2008	6.641957	11.098109	0.845259	74.199997	0.660659	0.035128	0.065775
1332	Singapore	2009	6.144677	11.061897	0.866255	74.500000	0.776382	-0.084892	0.035198
1333	Singapore	2010	6.531402	11.186028	0.864162	74.800003	0.846185	-0.028796	0.060282
1334	Singapore	2011	6.561042	11.226751	0.904474	75.019997	0.821816	-0.159601	0.098924
1335	Singapore	2013	6.533207	11.275893	0.807911	75.459999	0.827103	0.103943	0.242398
1336	Singapore	2014	7.062365	11.301013	0.822033	75.680000	0.834888	0.143034	0.132603
1337	Singapore	2015	6.619525	11.311312	0.866437	75.900002	0.886891	0.139112	0.098944
1338	Singapore	2016	6.033481	11.322025	0.925128	76.199997	0.903736	0.132903	0.047311
1339	Singapore	2017	6.378438	11.356686	0.897350	76.500000	0.926128	0.125963	0.161791
1340	Singapore	2018	6.374564	NaN	0.902841	76.800003	0.916078	NaN	0.096563

Predictions:

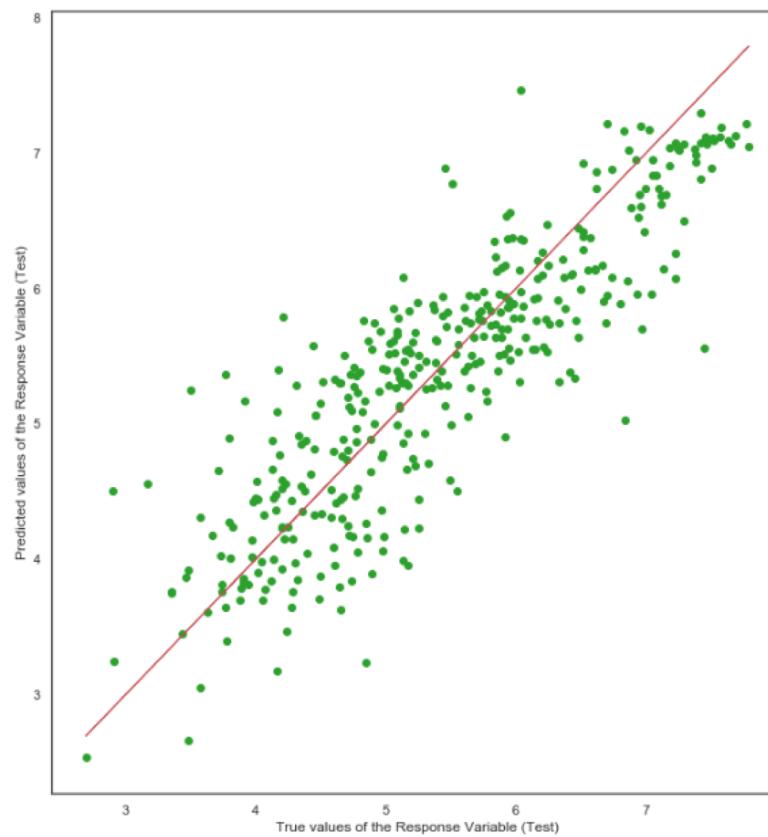
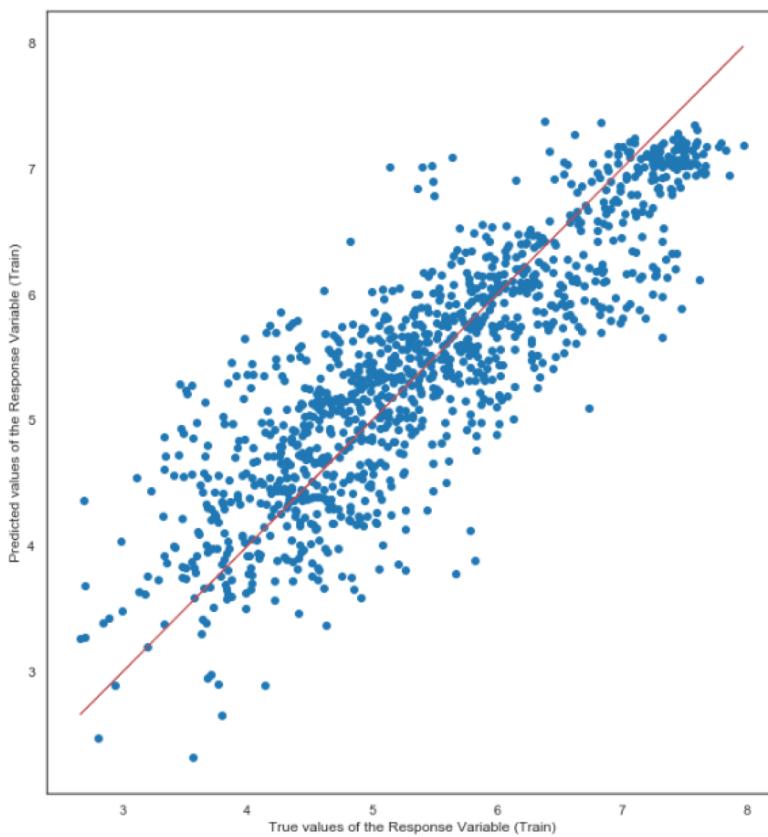
128	Singapore	2019	11.41546	0.877767	77.005642	0.932915	0.104246	0.13983
128	Singapore	2020	11.442071	0.877843	77.261686	0.94686	0.108219	0.145953



Prediction

✓ after predicting the indexes for each country...

Predict happiness for each country by using all 6 indexes

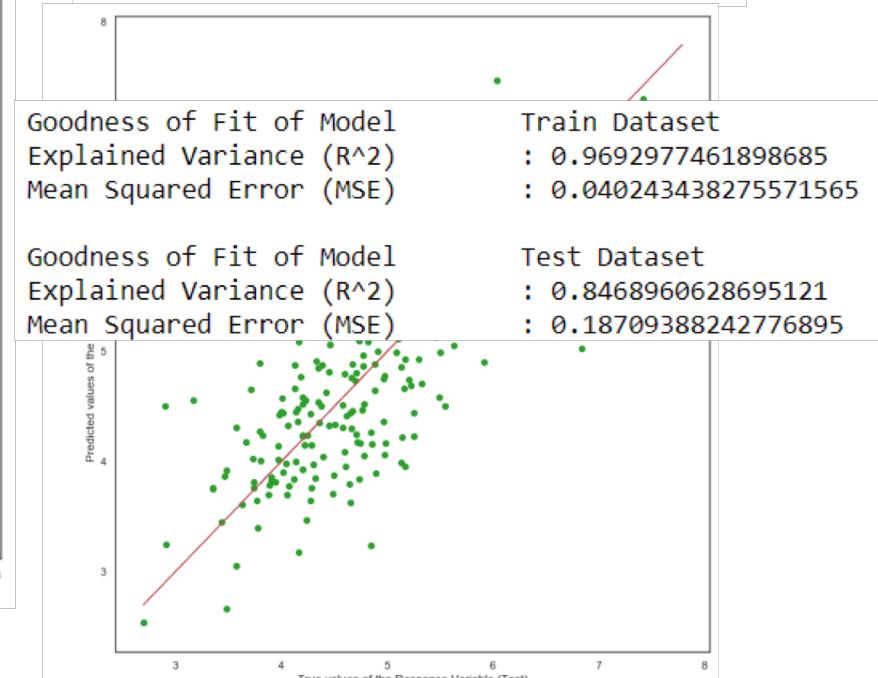
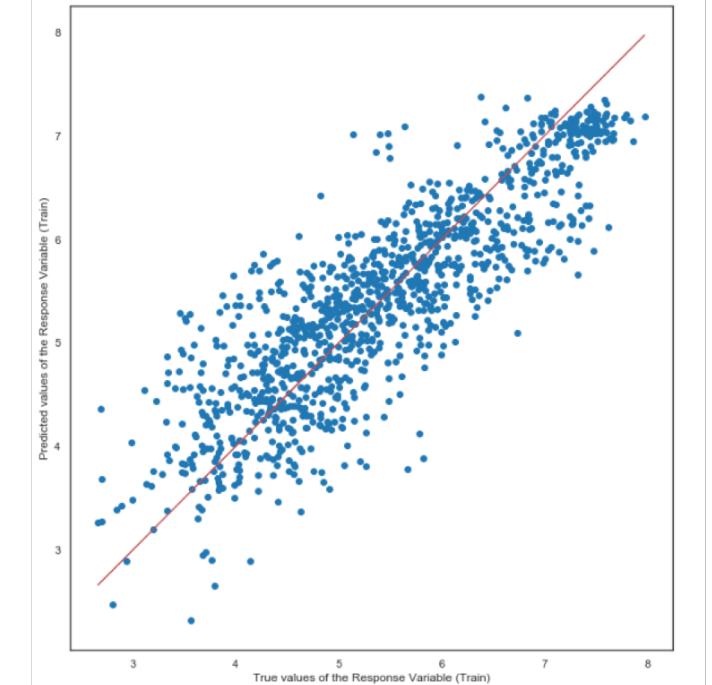
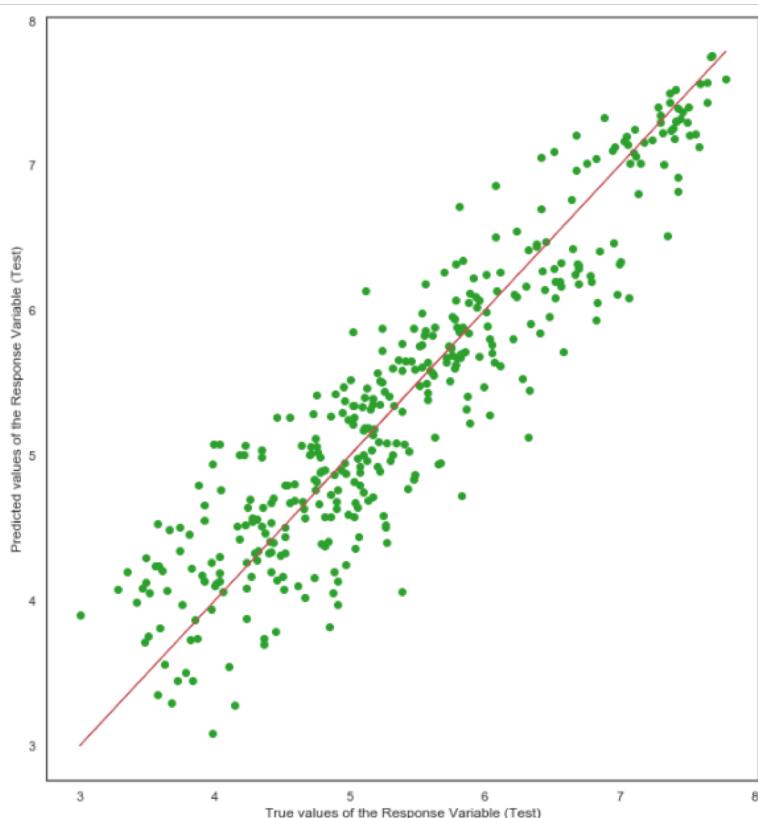
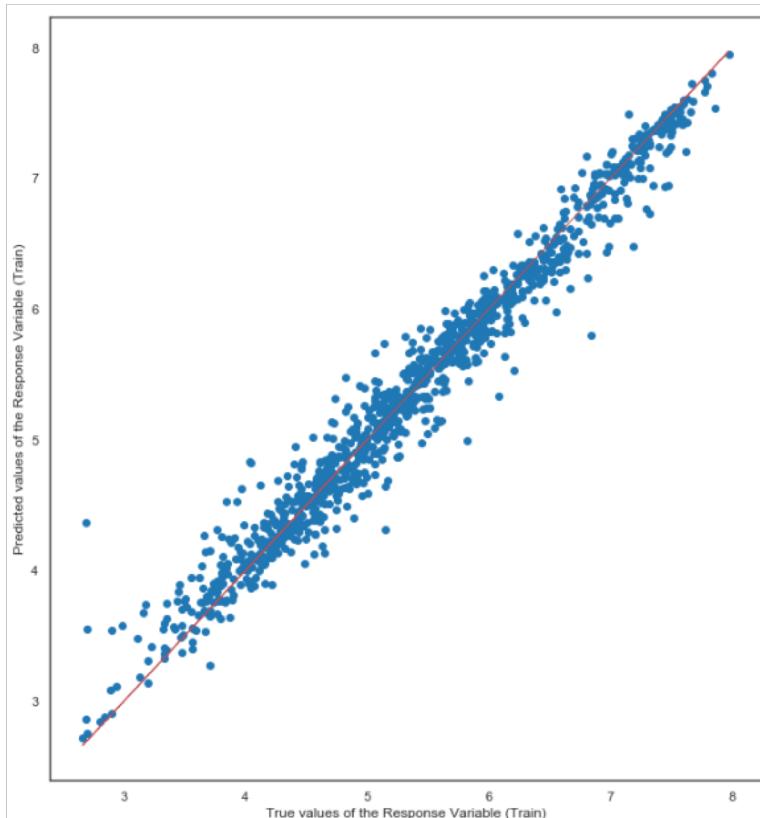


Multi-variate Linear Regression

	Predictors	Coefficients
0	Log GDP per capita	0.338533
1	Social support	2.443221
2	Healthy life expectancy at birth	0.027069
3	Freedom to make life choices	1.179606
4	Generosity	0.614289
5	Perceptions of corruption	-0.510168
Goodness of Fit of Model		Train Dataset
Explained Variance (R^2)		: 0.728699294116204
Mean Squared Error (MSE)		: 0.342262374438454
Goodness of Fit of Model		Test Dataset
Explained Variance (R^2)		: 0.7530677196094658
Mean Squared Error (MSE)		: 0.33772334498685047

Prediction

Random Forest Regression

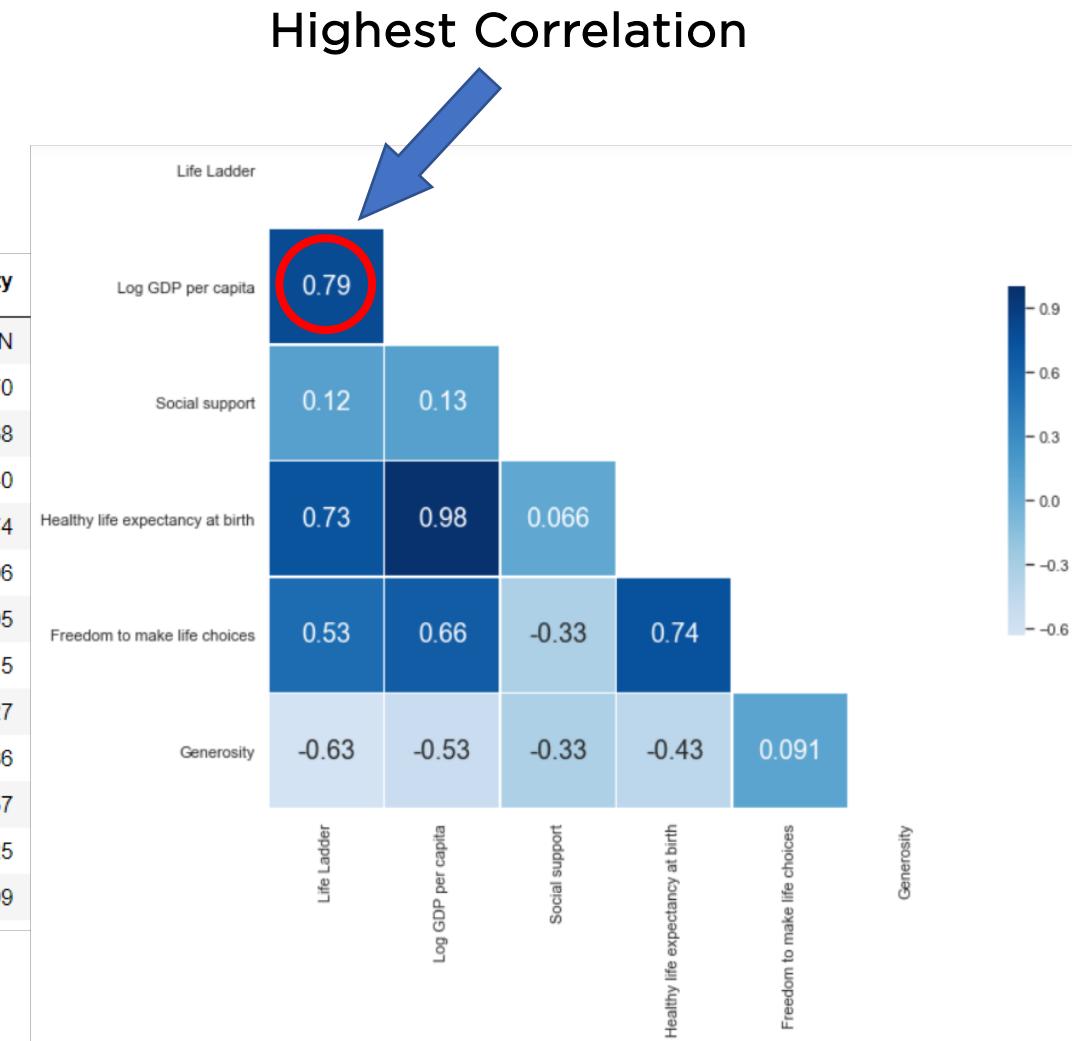


Prediction

Regression is great enough, but stats are not so ideal...

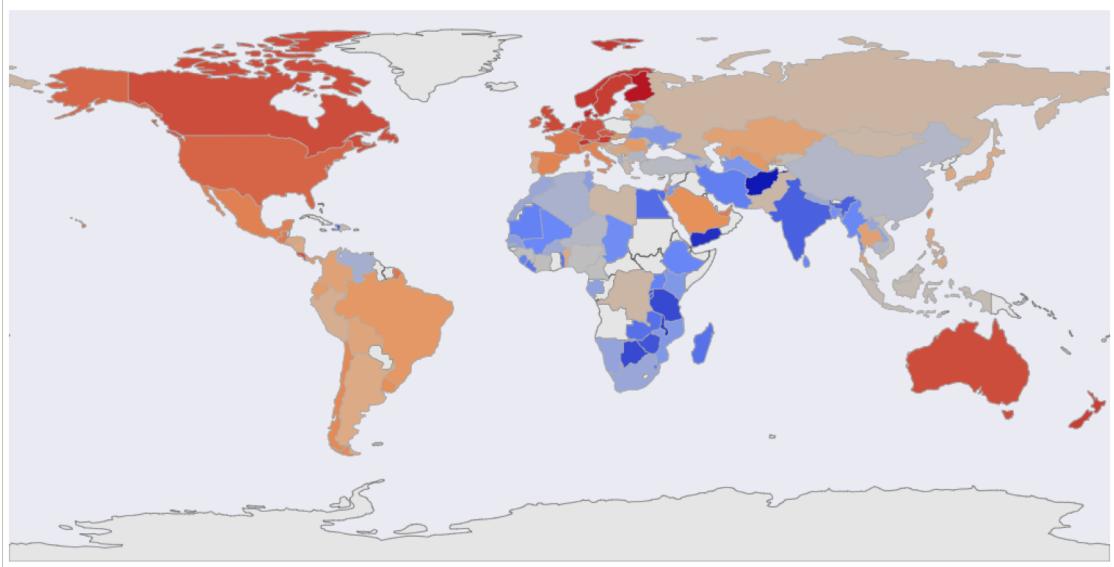
Country name	Year	Life Ladder	Log GDP per capita	Social support	Healthy life expectancy at birth	Freedom to make life choices	Generosity
China	2006	4.560495	8.765778	0.747011	66.879997	NaN	NaN
China	2007	4.862862	8.893610	0.810852	67.059998	NaN	-0.190670
China	2008	4.846295	8.980649	0.748287	67.239998	0.853072	-0.107168
China	2009	4.454361	9.065514	0.798034	67.419998	0.771143	-0.175440
China	2010	4.652737	9.161761	0.767753	67.599998	0.804794	-0.148574
China	2011	5.037208	9.248056	0.787171	67.760002	0.824162	-0.201896
China	2012	5.094917	9.318813	0.787818	67.919998	0.808255	-0.200405
China	2013	5.241090	9.388591	0.777896	68.080002	0.804724	-0.173715
China	2014	5.195619	9.453964	0.820366	68.239998	NaN	-0.232827
China	2015	5.303878	9.515609	0.793734	68.400002	NaN	-0.260586
China	2016	5.324956	9.575045	0.741703	68.699997	NaN	-0.243757
China	2017	5.099061	9.636177	0.772033	69.000000	0.877618	-0.191225
China	2018	5.131434	9.694376	0.787605	69.300003	0.895378	-0.174899

→ Uni-variate Linear Regression

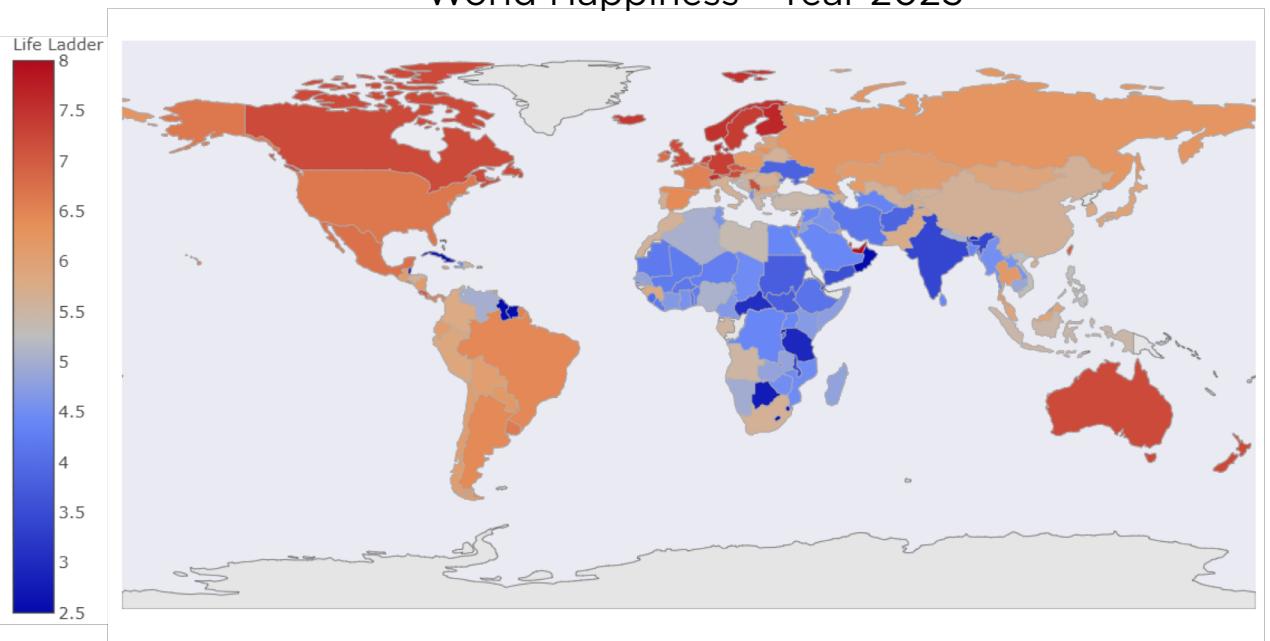


Prediction

World Happiness - Year 2018



World Happiness - Year 2023



So, what is our...

Outcome

We found success in
Random Forest
Model.

However, one interesting point is that multivariate regression model gives unhappy and shorter lifespan results.



WE ACHIEVE
OUR OBJECTIVE!

Finally, for our...

Conclusion



Brighter
future

Learning Points

- Explored possibilities with Plotly visualization tool
- Improved efficiency in python coding
- Gained insights in data analysis

Challenges

- Data cleaning
- Understand data visualizations



Thanks for tuning in!
Any Question?

TEAM CONTRIBUTIONS

