Predicting Life Expectancy

Dipta Roy

Overview

What factors have the most influence on Life Expectancy?

Can we predict how long a person will live?

How does obesity/schooling relate to life expectancy?

Who can use this information?

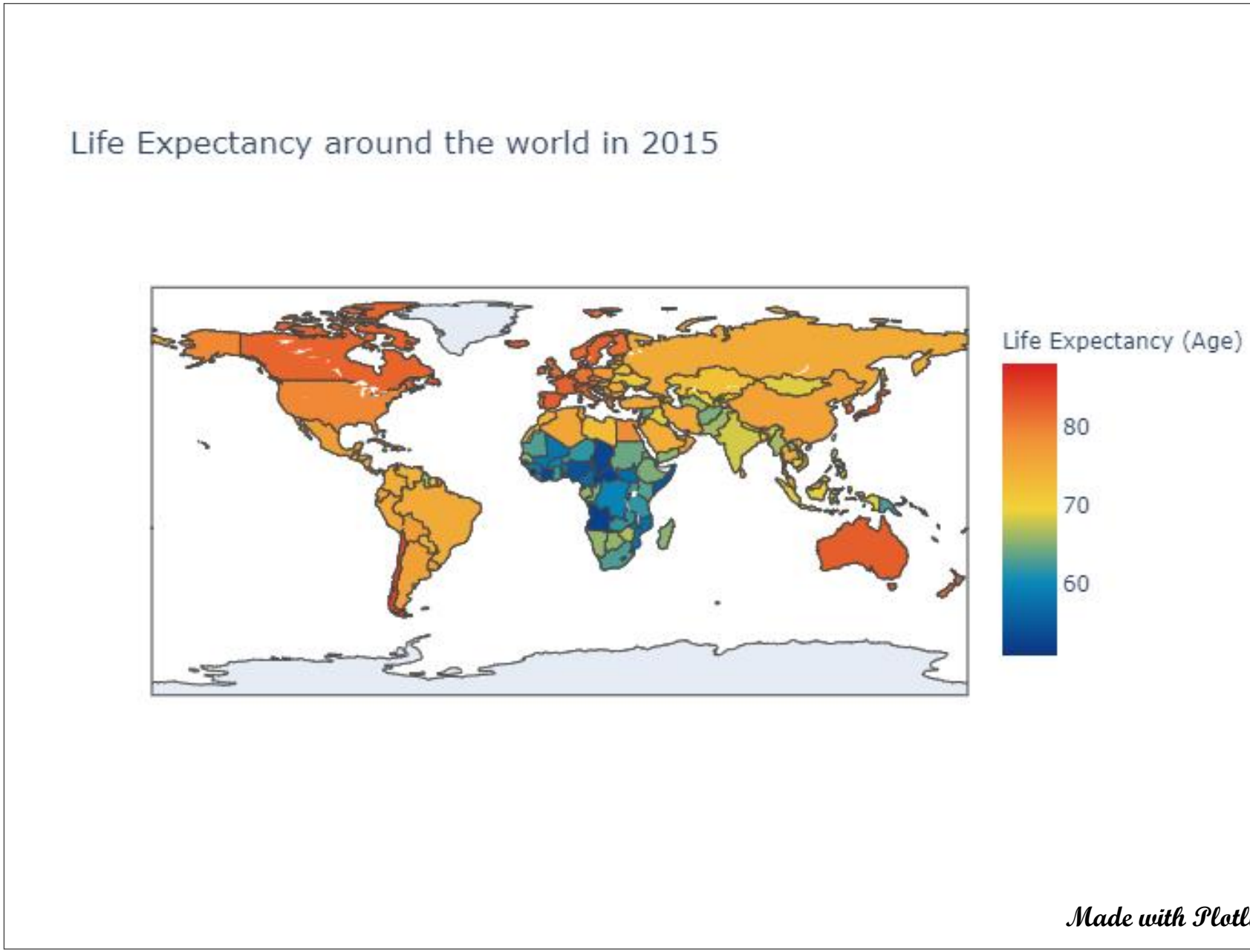
DATA SET!

- FROM KAGGLE
  - Life Expectancy (2000 – 2015) from the World Health Organization
  - Obesity Among Adults by Country (1975-2016) from the World Health Organization
  - Suicide Rates Overview (1985-2016) from UN Development Program, World Bank, and World Health Organization

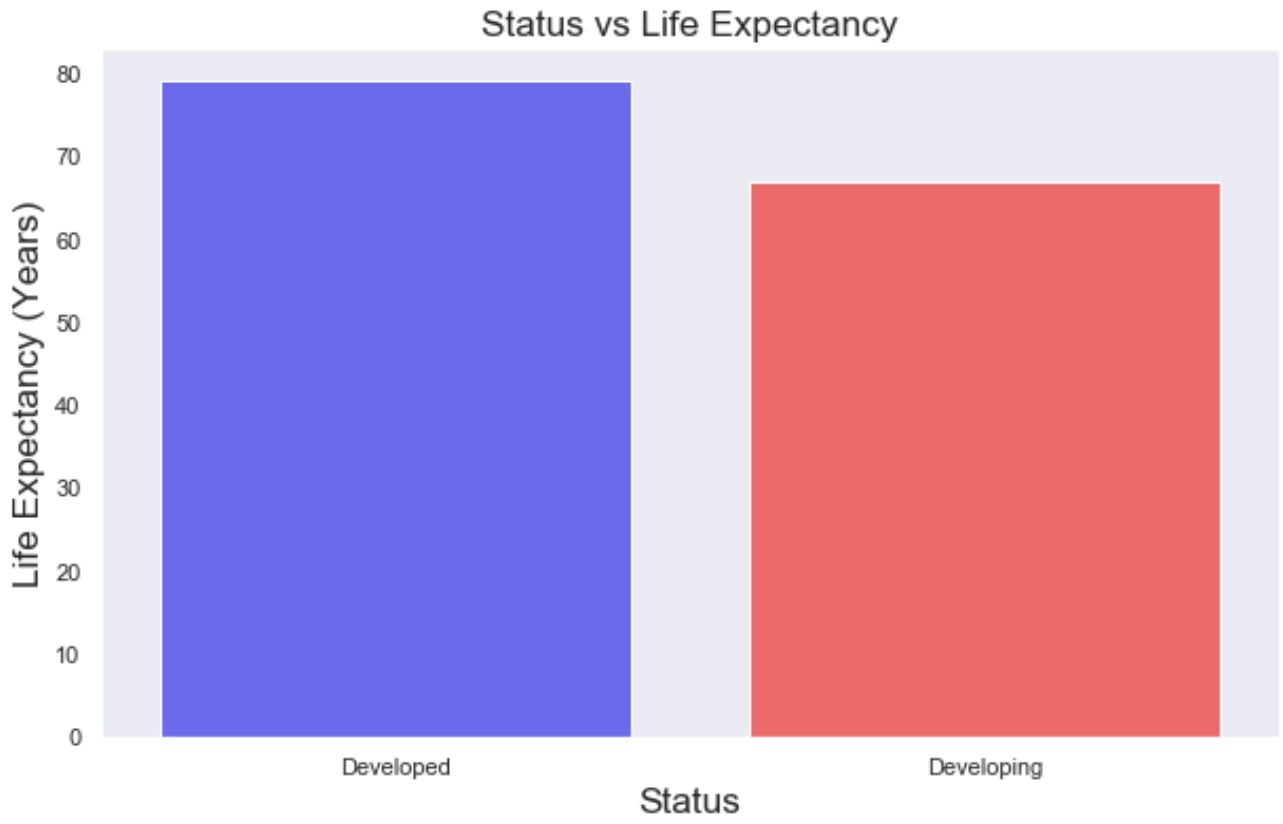



What can we infer from this color map?

- At first glance, it looks like regions close to each other have similar life expectancies.
- Countries who seem more '**developed**' have higher life expectancy
- Countries who have 'free' **healthcare** have higher life expectancy



$$Ho: Life\ Expectancy_{Developed} = Life\ Expectancy_{Developing}$$
$$Ha: Life\ Expectancy_{Developed} \neq Life\ Expectancy_{Developing}$$



First Hypothesis: Life Expectancy of Developed vs Developing Countries

- Two Sampled T-Test
- Since our P-Value is lower than our alpha of 0.05 we can reject our null hypothesis testing
- There is a significant difference in the average Life Expectancy of a developed country and a developing country

Baseline OLS!

- Out of all of the selected features, amount of schooling had the biggest impact on Life Expectancy

OLS Regression Results

Dep. Variable:	life_expectancy	R-squared:	0.682
Model:	OLS	Adj. R-squared:	0.681
Method:	Least Squares	F-statistic:	728.4
Date:	Mon, 22 Jun 2020	Prob (F-statistic):	0.00
Time:	00:55:46	Log-Likelihood:	-7388.1
No. Observations:	2382	AIC:	1.479e+04
Df Residuals:	2374	BIC:	1.484e+04
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	69.1779	0.111	623.090	0.000	68.960	69.396
schooling_scaled	3.3777	0.220	15.363	0.000	2.947	3.809
obesity_scaled	1.6075	0.143	11.208	0.000	1.326	1.889
alcohol_scaled	-0.9721	0.147	-6.599	0.000	-1.261	-0.683
income_composition_of_resources_scaled	2.4960	0.186	13.417	0.000	2.131	2.861
diphtheria_scaled	1.4642	0.124	11.767	0.000	1.220	1.708
percentage_expenditure_scaled	0.7779	0.121	6.446	0.000	0.541	1.014
status_Developed_scaled	1.2027	0.153	7.844	0.000	0.902	1.503

Omnibus:	217.959	Durbin-Watson:	0.378
Prob(Omnibus):	0.000	Jarque-Bera (JB):	768.205
Skew:	-0.419	Prob(JB):	1.54e-167
Kurtosis:	5.653	Cond. No.	4.45

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

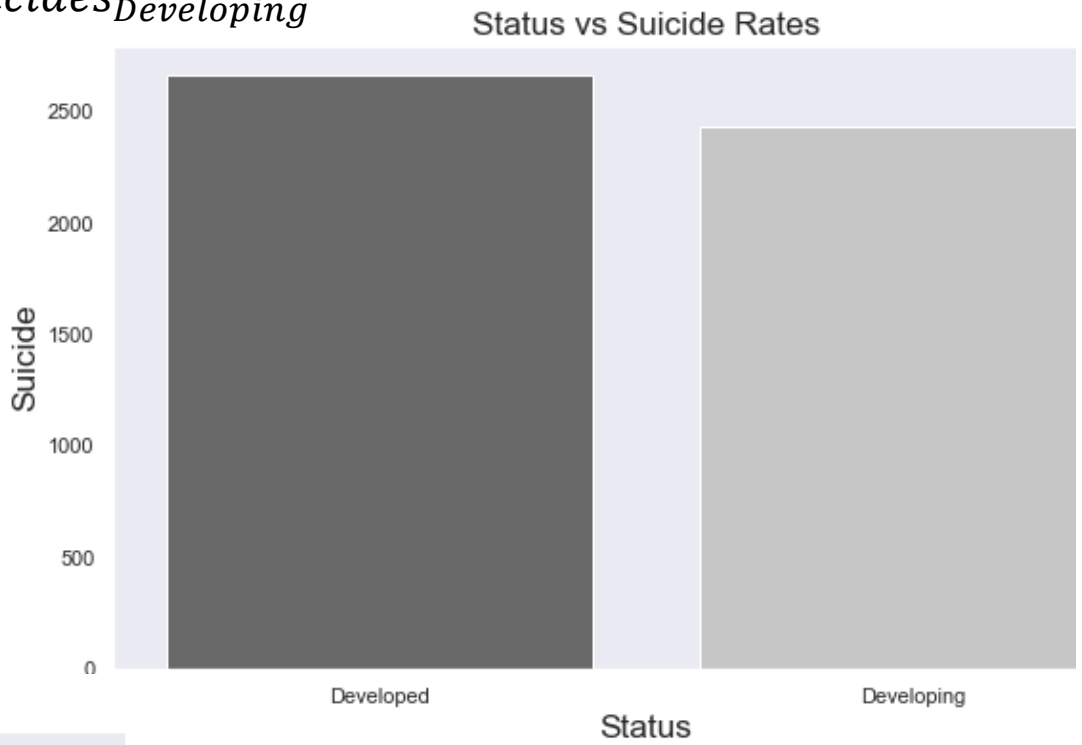
What's going on here?

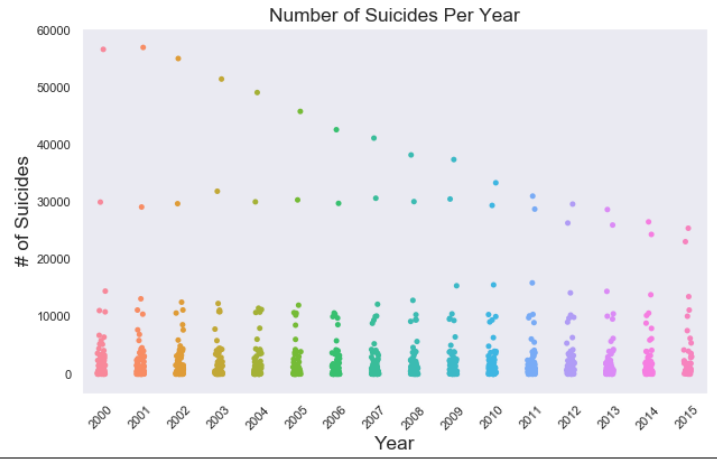
$$Ho: Obesity_{Developed} = Obesity_{Developing}$$
$$Ha: Obesity_{Developed} \neq Obesity_{Developing}$$

Average Obesity Percentage of Developed countries vs Developing countries

- P-Value = 4.95 E -18
- We can reject Null hypothesis.
- There is a significant difference in the average obesity percentage of a developed country and a developing country

$$Ho: Suicides_{Developed} = Suicides_{Developing}$$
$$Ha: Suicides_{Developed} \neq Suicides_{Developing}$$





Number of suicides in Developing countries vs Developed

- P-Value= 0.53
- Fail to reject Null Hypothesis
- There is no significant difference in the average number of suicides in Developed countries and Developing Countries

Modeling to Predict

- Winning model: Baseline with Countries.
- F-test and Recursive Feature Elimination gave the same results as the winning model.
- We can predict the life expectancy of a person

Modelling Type	# of Features	Train RMSE	Test RMSE	R^2
Baseline w/o Country	7	5.44 yrs	5.46 yrs Z: 0.57	0.68
Baseline w/ Country	164	2.09 yrs	2.43 yrs Z: 0.25	0.95
Lasso (alpha=.1)	164	1.81 yrs	2.80 yrs Z: 0.29	0.93

Main Take-Aways and how can we use our data?

- **Schooling** has an influence on life expectancy – policy makers should make school more accessible and more prevalent
- Government **expenditure on healthcare** has a positive correlation, countries with lower life expectancy should increase their healthcare expenditure in order to improve its average lifespan.
- Governments should allocate budget into **physical** and **mental** health to help increase life expectancy

THANK YOU!!
ANY QUESTIONS?