COVID Related to Happiness

Yuyan Shi
Samir Epili
Bhavna Kaparaju
Mauricio Morales

Dataset Description
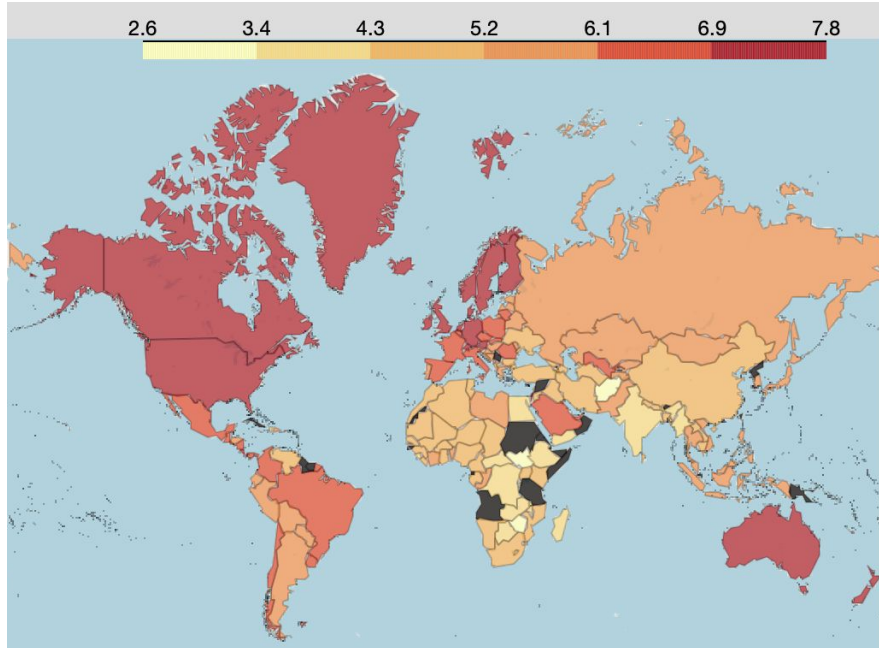
Happiness Index Report

The report includes the happiness index and key variables, such as logged GDP per capita, for each country.
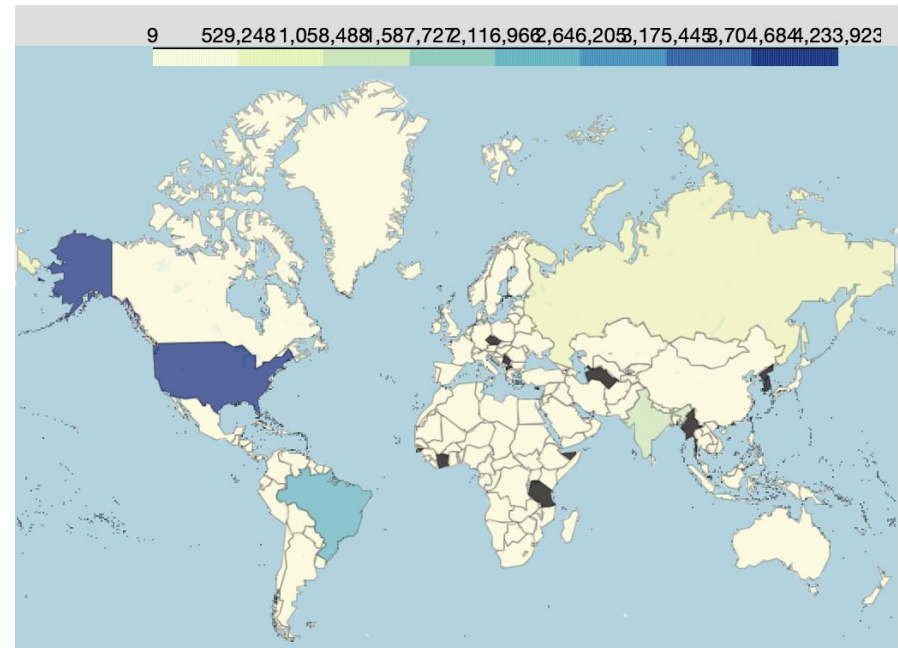Source: Kaggle

COVID-19 Confirmed Cases

The dataset includes cumulative number of confirmed cases for each country, from Jan 22nd to July 26th.
Source: John Hopkins University

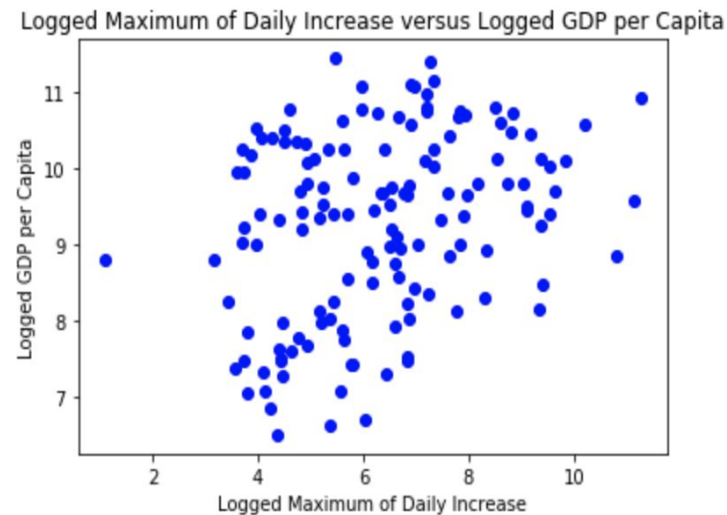# Before the formal model building

Happiness Index Heat Map
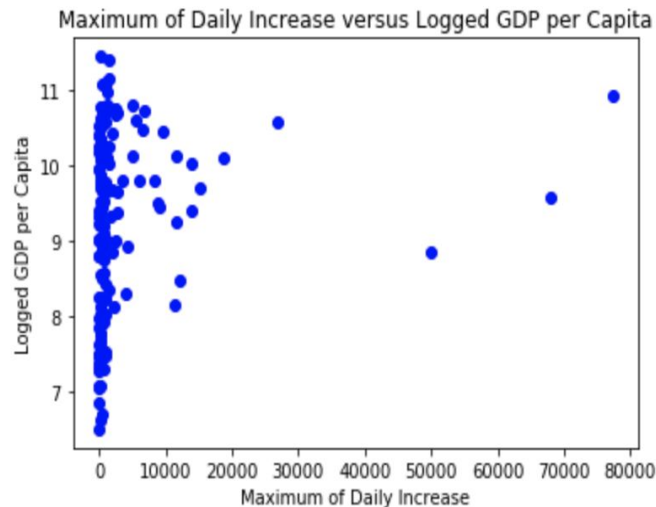
Confirmed Cases (until 7/26) Heat Map

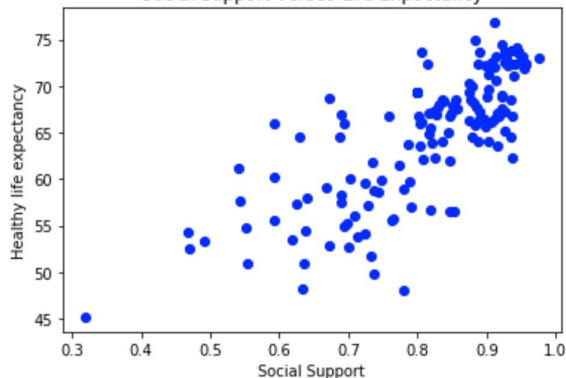We chose the logarithm of max rate, because of......

the homoscedasticity assumption of linear regression model

Maximum of Daily Increase versus Logged GDP per Capita

Logged Maximum of Daily Increase versus Logged GDP per Capita

Some Exploratory Analysis in Happiness Dataset
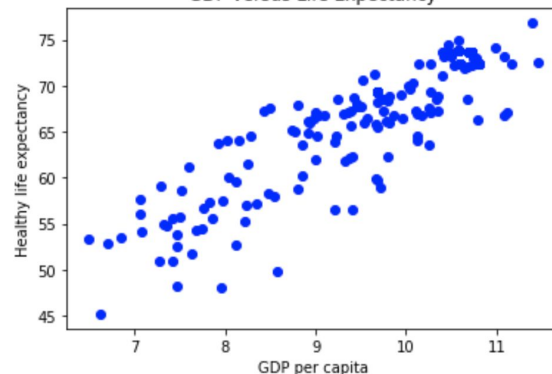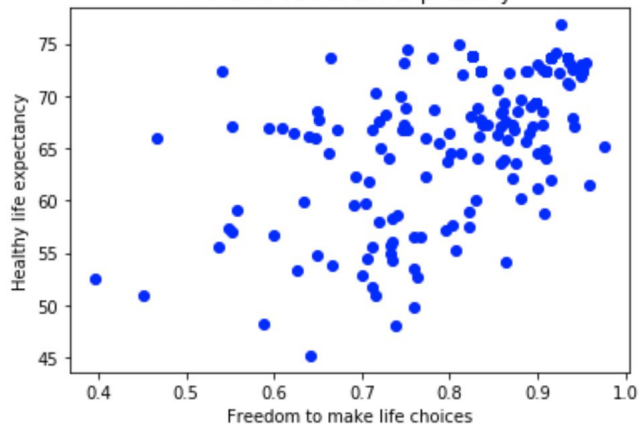
Plotting Max Rate to Variables in Happiness Report

(1/2)
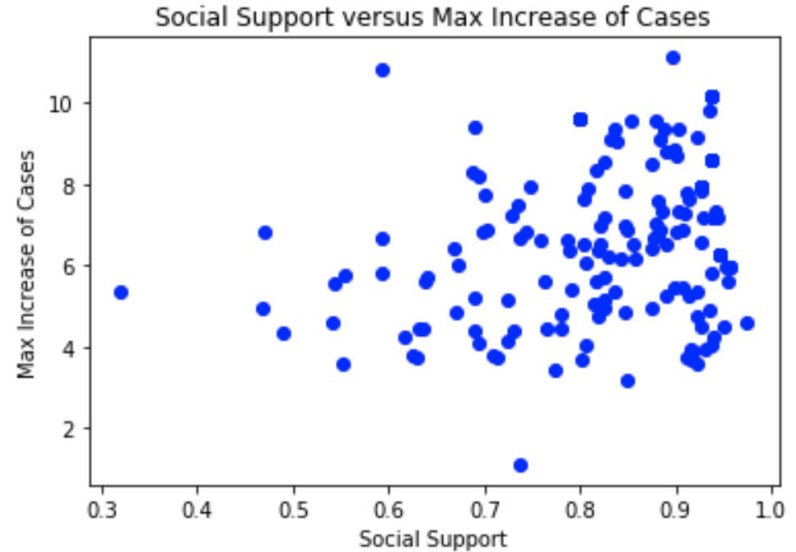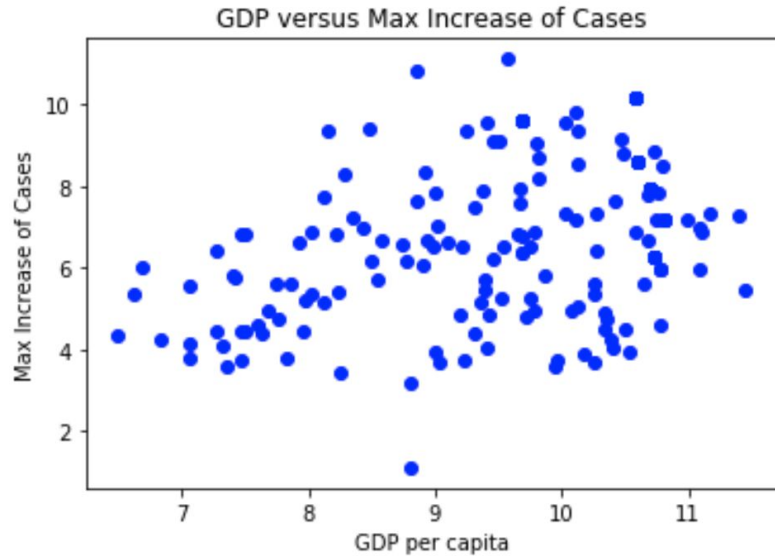
Plotting Max Rate to Variables in Happiness Report
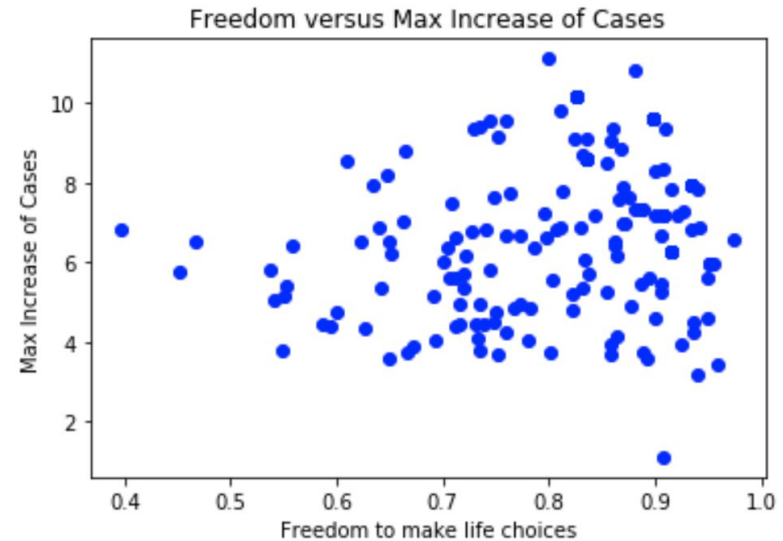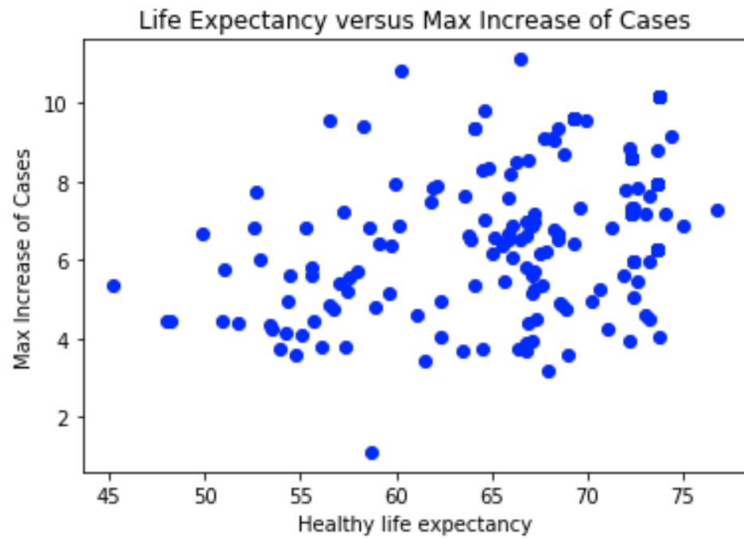
Standard deviations of happiness predictors and maximum rate.

Variables	Standard Deviations
Ladder Score	1.16
Logged GDP per capita	1.12
Social support	0.11
Healthy life expectancy	6.66
Freedom to make life choices	0.11
Maximum increase in infections	2.05

Linear Regression Model of Happiness

```
P-values                                                R-squared:                               0.736
Intercept                       0.001588                Adj. R-squared:                          0.729
Logged_GDP_per_capita           0.005960                F-statistic:                             103.1
Social_support                  0.000063                Prob (F-statistic):                   9.00e-42
Healthy_life_expectancy         0.007293                Log-Likelihood:                        -131.02
Freedom_to_make_life_choices    0.000481                AIC:                                     272.0
Generosity                      0.225126                BIC:                                     287.2
Perceptions_of_corruption       0.047857
--------------------------------------------------------------------------------------------------------
                                coef      std err        t       P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------------------
Intercept                     -3.0104      0.454      -6.631     0.000      -3.908      -2.113
Logged_GDP_per_capita          0.2398      0.082       2.933     0.004       0.078       0.401
Social_support                 2.4049      0.657       3.660     0.000       1.106       3.703
Healthy_life_expectancy        0.0386      0.013       2.949     0.004       0.013       0.064
Freedom_to_make_life_choices   2.3266      0.460       5.054     0.000       1.417       3.236
```

Predicting Max Rate – Linear Regression

Is it possible to predict max rate of coronavirus infections using happiness?

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-0.3658	0.772	-0.474	0.636	-1.892	1.160
Ladder_score	0.9750	0.138	7.084	0.000	0.703	1.247

R-squared:	0.265
Adj. R-squared:	0.260
F-statistic:	50.18
Prob (F-statistic):	6.41e-11
Log-Likelihood:	-284.61
AIC:	573.2
BIC:	579.1

Predicting Max Rate

Using all the predictors of happiness…

R-squared:	0.376
Adj. R-squared:	0.348
F-statistic:	13.46
Prob (F-statistic):	6.55e-12
Log-Likelihood:	-273.09
AIC:	560.2
BIC:	580.8

P-values

|	
Intercept	0.001731
Logged_GDP_per_capita	0.000051
Social_support	0.048782
Healthy_life_expectancy	0.139684
Freedom_to_make_life_choices	0.903185
Generosity	0.725490
Perceptions_of_corruption	0.568015

Predicting Max Rate

Isolating the best predictors...

RMSE: **1.753**

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-4.6920	1.121	-4.186	0.000	-6.908	-2.476
Logged_GDP_per_capita	1.0443	0.120	8.712	0.000	0.807	1.281

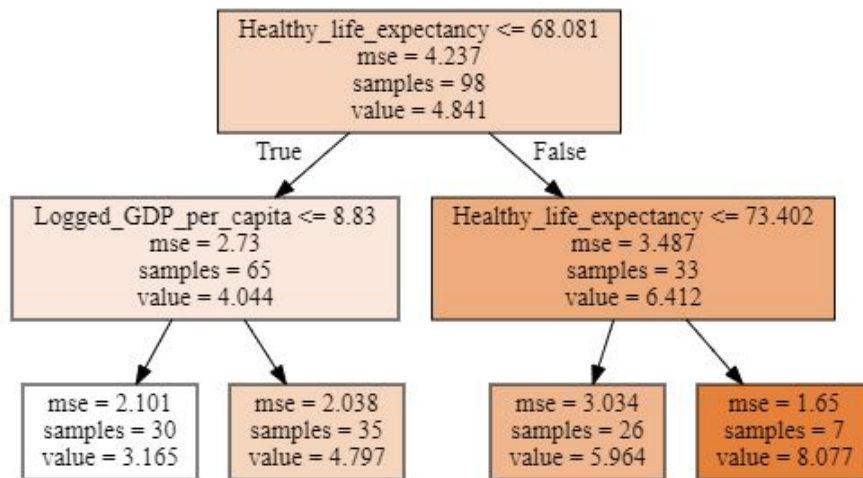|	
R-squared:	0.353
Adj. R-squared:	0.349
F-statistic:	75.90
Prob (F-statistic):	7.94e-15
Log-Likelihood:	-275.63
AIC:	555.3
BIC:	561.2

GDP versus Max Increase of Cases
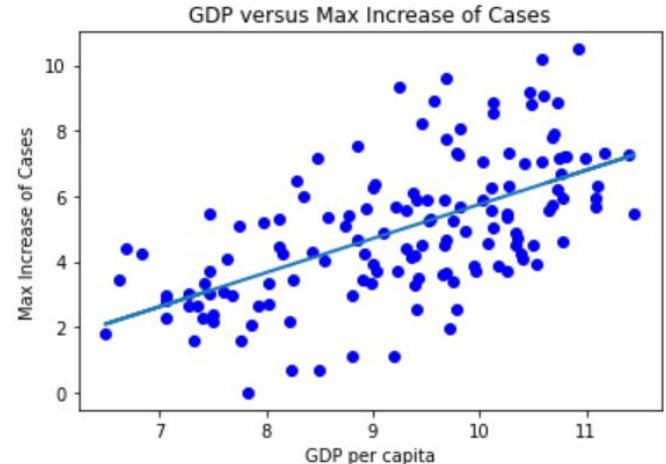
Predicting Max Rate – Regression Tree

What does the regression tree say?

RMSE: 2.10

Mean Absolute Error: 1.7073411875337186
Root Mean Square Error: 2.0959382133392155

Healthy_life_expectancy 0.7739216142767888
Logged_GDP_per_capita 0.22607838572321115
Social_support 0.0
Freedom_to_make_life_choices 0.0

Healthy_life_expectancy <= 68.081
mse = 4.237
samples = 98
value = 4.841

True

False

Logged_GDP_per_capita <= 8.83
mse = 2.73
samples = 65
value = 4.044

Healthy_life_expectancy <= 73.402
mse = 3.487
samples = 33
value = 6.412

mse = 2.101
samples = 30
value = 3.165

mse = 2.038
samples = 35
value = 4.797

mse = 3.034
samples = 26
value = 5.964

mse = 1.65
samples = 7
value = 8.077

Conclusions

- Happiness alone is a poor predictor of coronavirus cases in a country
- Of the happiness predictors, GDP per capita is the most significant in predicting the max rate in a given country
- Less accurate regression tree shows that healthy life expectancy is the most significant predictor
- Positive relationship in countries of high values of both predictors could be due to increased testing => more detected cases
- Uncovering more underlying factors could further improve the model