integrated machine learning

Presenters: Jennifer, Lori, Susana, Shailza

motivation

to test if a machine learning model can predict a country's future happiness score

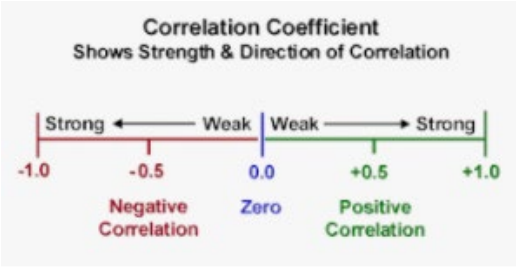
tools used

Google Collab, Jupyter notebook, MongoDB, MongoDB Atlas, HTML/CSS, JavaScript, Heroku, VS Code, Machine Learning with SK Learn and Tensor Flow, Flask, Python etc.

brain storming

```
df_out.corr()
```

	happiness_score	standard_error	gpd_per_capita	Family	life_expectancy	Freedom	government_corruption	Generosity	dystopia_residual
happiness_score	1.000000	-0.177254	0.779171	0.693547	0.734491	0.556414	0.397989	0.168272	0.526343
standard_error	-0.177254	1.000000	-0.217651	-0.120728	-0.310287	-0.129773	-0.178325	-0.088439	0.083981
gpd_per_capita	0.779171	-0.217651	1.000000	0.566283	0.789103	0.331127	0.294856	-0.015402	0.079411
Family	0.693547	-0.120728	0.566283	1.000000	0.569937	0.425377	0.205060	0.071775	0.053264
life_expectancy	0.734491	-0.310287	0.789103	0.569937	1.000000	0.369799	0.249773	0.088002	0.025248
Freedom	0.556414	-0.129773	0.331127	0.425377	0.369799	1.000000	0.492875	0.343389	0.034829
government_corruption	0.397989	-0.178325	0.294856	0.205060	0.249773	0.492875	1.000000	0.289499	-0.023504
Generosity	0.168272	-0.088439	-0.015402	0.071775	0.088002	0.343389	0.289499	1.000000	-0.110632
dystopia_residual	0.526343	0.083981	0.079411	0.053264	0.025248	0.034829	-0.023504	-0.110632	1.000000



$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$$

Which features have the highest correlation to the target happiness score?

GDP per capita, life expectancy and family.

models tested

```
Model: LinearRegression
Train score: 0.7393901378351614
Test Score: 0.8494853874320927

Model: Ridge
Train score: 0.7393880852483157
Test Score: 0.849521505688118

Model: Lasso
Train score: 0.7393882377459953
Test Score: 0.8492500569803096

Model: ElasticNet
Train score: 0.36889495181065124
Test Score: 0.3764101126508038

Model: SVR
Train score: 0.8494933797860317
Test Score: 0.858728773446082

Model: KNeighborsRegressor
Train score: 0.8624256742518872
Test Score: 0.8070943187204942

Model: RandomForestRegressor
Train score: 0.9717845928130108
Test Score: 0.8089776021871344

Model: ExtraTreesRegressor
Train score: 1.0
Test Score: 0.8270806629985477

Model: AdaBoostRegressor
Train score: 0.8362149565161389
Test Score: 0.7889509011693108
```

```
[ ]  1 full_ttf_compare_df = pd.DataFrame({'Country':countries_test_df, 'Actual Score': y_
     2 full_ttf_compare_df
```

	Country	Actual Score	SVR Prediction	Linear Reg Prediction	Lasso Prediction
0	Finland	7.769	6.613931	6.899565	6.897753
1	Denmark	7.600	7.135372	7.020708	7.018174
2	Norway	7.554	7.009200	7.186533	7.183963
3	Iceland	7.494	6.781082	6.926663	6.924602
4	Netherlands	7.488	7.083237	6.942099	6.939602
...	...	...	...	...	...
151	Rwanda	3.334	4.494115	5.119618	5.119137
152	Tanzania	3.231	4.808172	4.736554	4.736947
153	Afghanistan	3.203	3.730025	3.471835	3.474305
154	Central African Republic	3.083	3.517351	3.204270	3.206719
155	South Sudan	2.853	3.631911	3.413541	3.415631

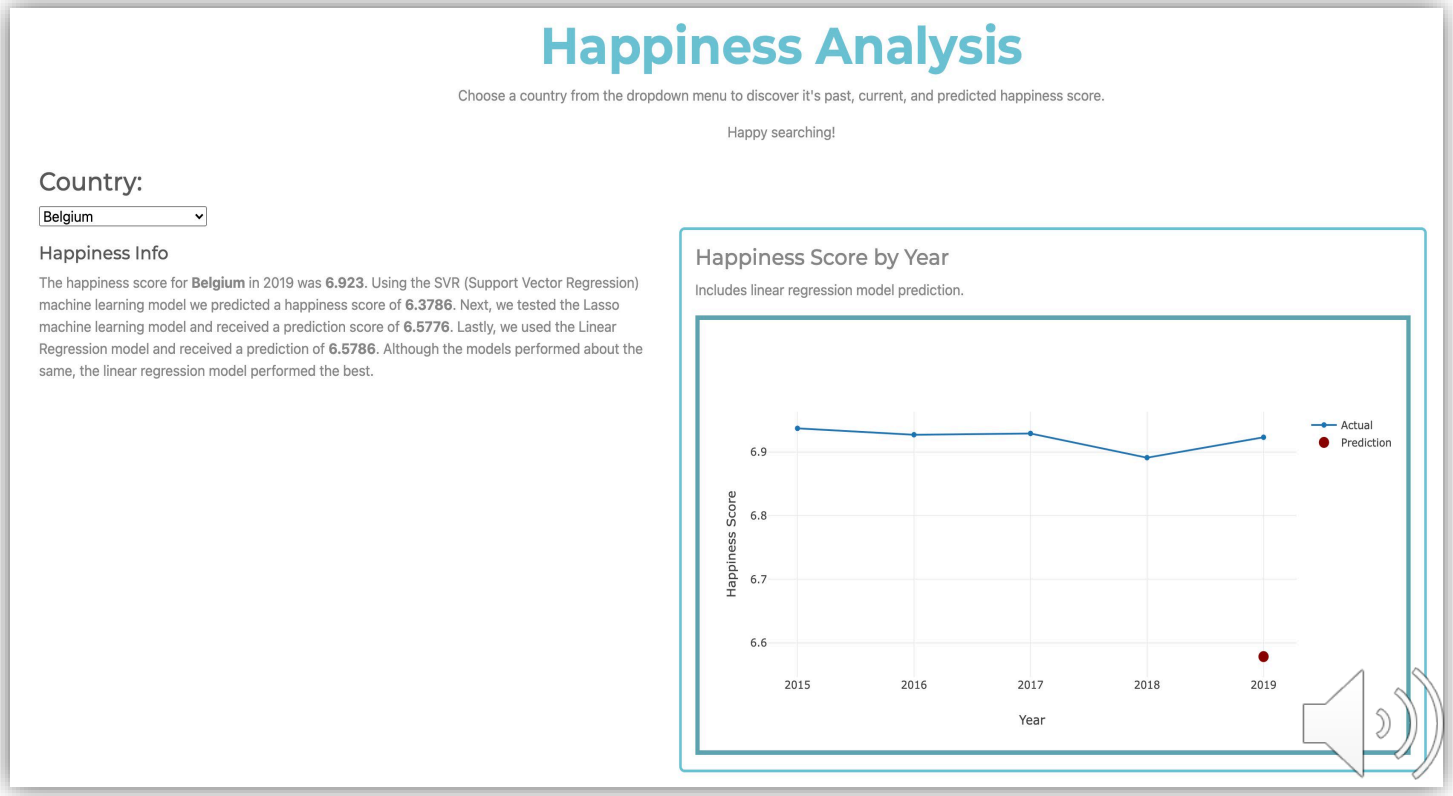
156 rows × 5 columns

models used for our prediction

SVR, Multiple Variable Linear Regression, and LASSO Regression

webpage design

Create an interactive tool to predict a country's happiness score.



heroku deployment Process

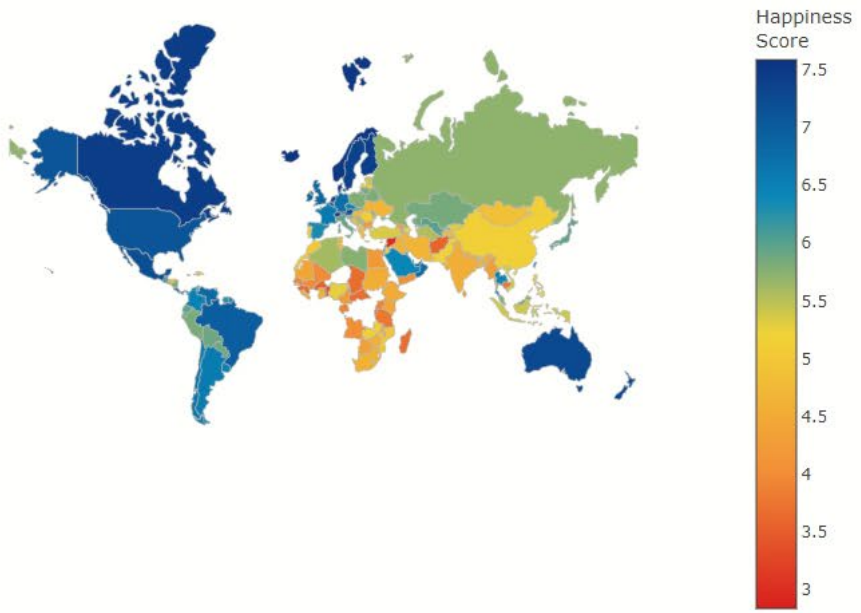
- Create repository
- Determine file structure
- Add Procfile
- Create the Requirements.txt file
- Push to the Repo
- Open Heroku
- Create new app
- Connect Heroku and GitHub
- Deploy

Use the drop down button below to view charts by year.

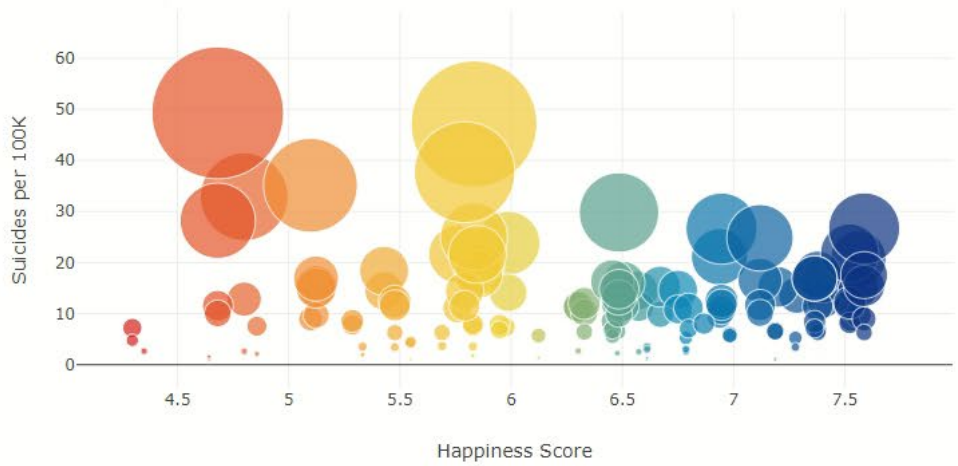
Year:

Select One

World Happiness Map



Happiness Score vs Suicide Rates by Country




thank you