the flow

A guide to the modern life surrounded by technology

https://withtheflow.herokuapp.com/

By Annabel Cheong

DATA STORY

Data Source: 2020 and 2021 Happiness Dataset

- Gallup World Poll and Lloyd's Register Foundation

- Kaggle

Features (X variables)

- year
- GDP/capita
- healthy life expectancy
- social support
- freedom
- generosity
- corruption

Happiness Score (y variable)

0 to 10

**10 is the happiest*

WEBSITE OVERVIEW

Tableau Visualisations

Machine Learning

Machine Learning Workflow

ETL

Extraction

kaggle

2020 and 2021 dataset

Transformation

python

Pandas

- 2 dataframes: X variables, y variable
- Train, Test, Split
- Predict
- Concatenate to dataframe: X variables, actual, prediction

Load

heroku PostgreSQL

Flask

USER

methods = "GET"

return jsonify (happiness_table)
return jsonify(prediction)

call data
via endpoint

HTML

CSS

JAVASCRIPT

Event listener
(X variables entered)

Data Preparation for Machine Learning

Creating the ML Model

- Variables required: **X_train_scaled**, **y_train**

Testing the ML Model

- Variables required: **X_test_scaled**, **y_test**

Step 1. Read in CSV files via pandas and concatenate dataframes

Additional year column

Step 2. Clean dataframe

Drop NAs, rename columns

Step 3. Categorise y values

Round up y values: e.g. 2-3: 3, 7-8: 8

- Lowest y value: 2.545,
- Highest y value: 7.825

Therefore, **y value categories: 3, 4, 5, 6, 7, 8**

Step 4. Create 2 dataframes; 'X' and 'y' dataframes

X dataframe: Drop y columns
y dataframe: Select only y categorised column

Step 5. Split, Train, Test

```
from sklearn.model_selection
import train_test_split
train_test_split(X, y)
```

X_train, X_test, y_train, y_ test

Step 6. Scale X values

```
from sklearn.preprocessing import MinMaxScaler
X_scaler = MinMaxScaler().fit(X_train)
X_train_scaled = X_scaler.transform(X_train)
X_test_scaled = X_scaler.transform(X_test)
```

Creating the ML Model

Step 1. Import *Random Forest* library

```
from sklearn.ensemble import RandomForestClassifier
```

Step 2. Create the model

```
rf = RandomForestClassifier(n_estimators=770)
```

Step 3. Fit the model

```
rf = rf.fit(X_train_scaled, y_train)
```

Testing the ML Model

Step 4. Test the model

```
model_accuracy = rf.score(X_test_scaled, y_test)
```

Model Accuracy: **76%**

Importance of Features

Test the importance of features in the model

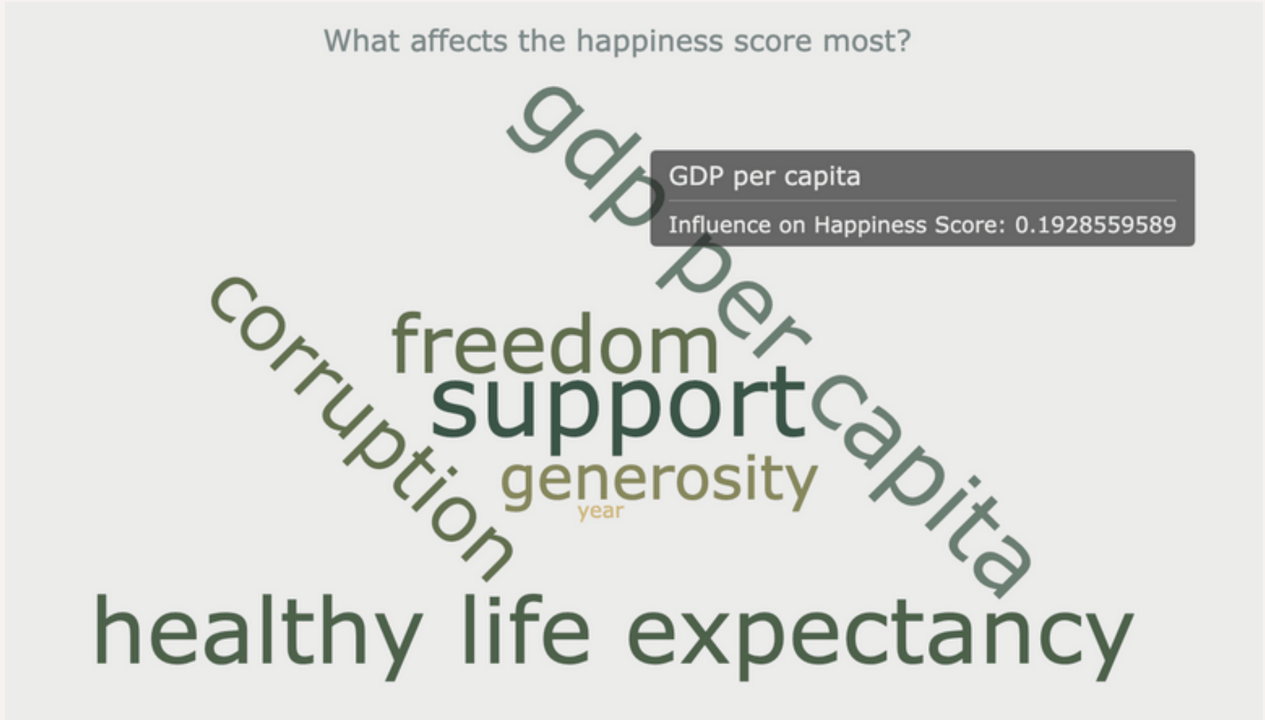
importances = rf.feature_importances_

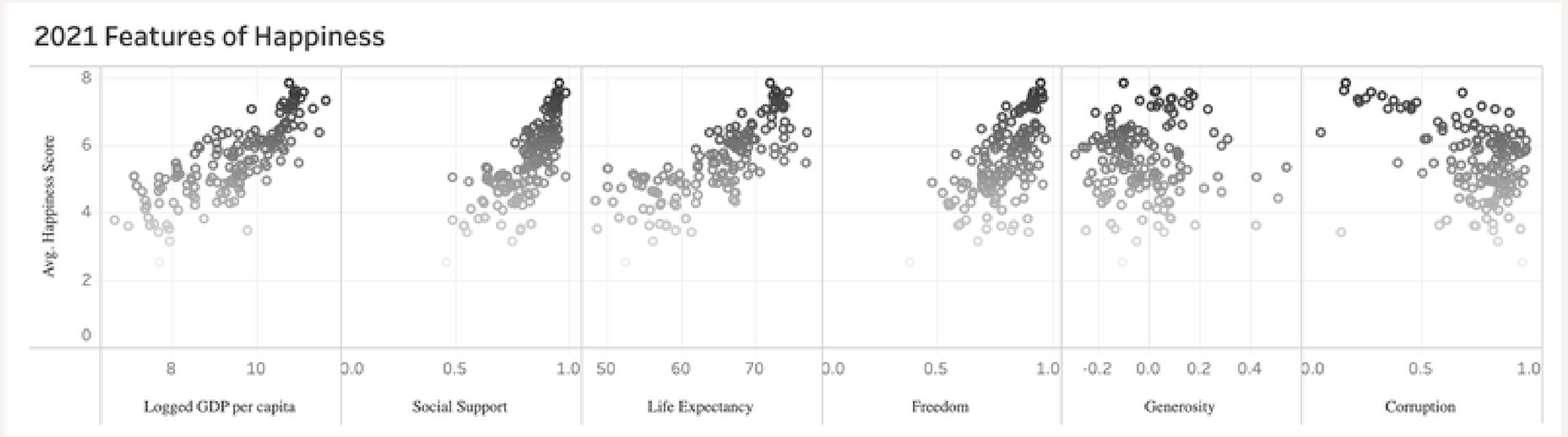
Top 2 Features

- Social Support: 0.199
- GDP/capita: 0.193

Bottom 2 Features

- Generosity: 0.110
- Year: 0.012





Happiness Score Prediction

JAVASCRIPT predictScore.js

Step 1. Javascript Event Listener (D3)

Saved user input x variables

FLASK END POINT app.py

Step 2. Flask end point

Saved x variables via methods = [GET]

Step 3. X scaler

- Scale X user input

Step 4. Predict user happiness score

- Random forest model
- User x user input dataset

JAVASCRIPT predictScore.js

Step 5. Call end point

**that returns jsonify(prediction)* (D3)

Saved user input x variables

Challenge

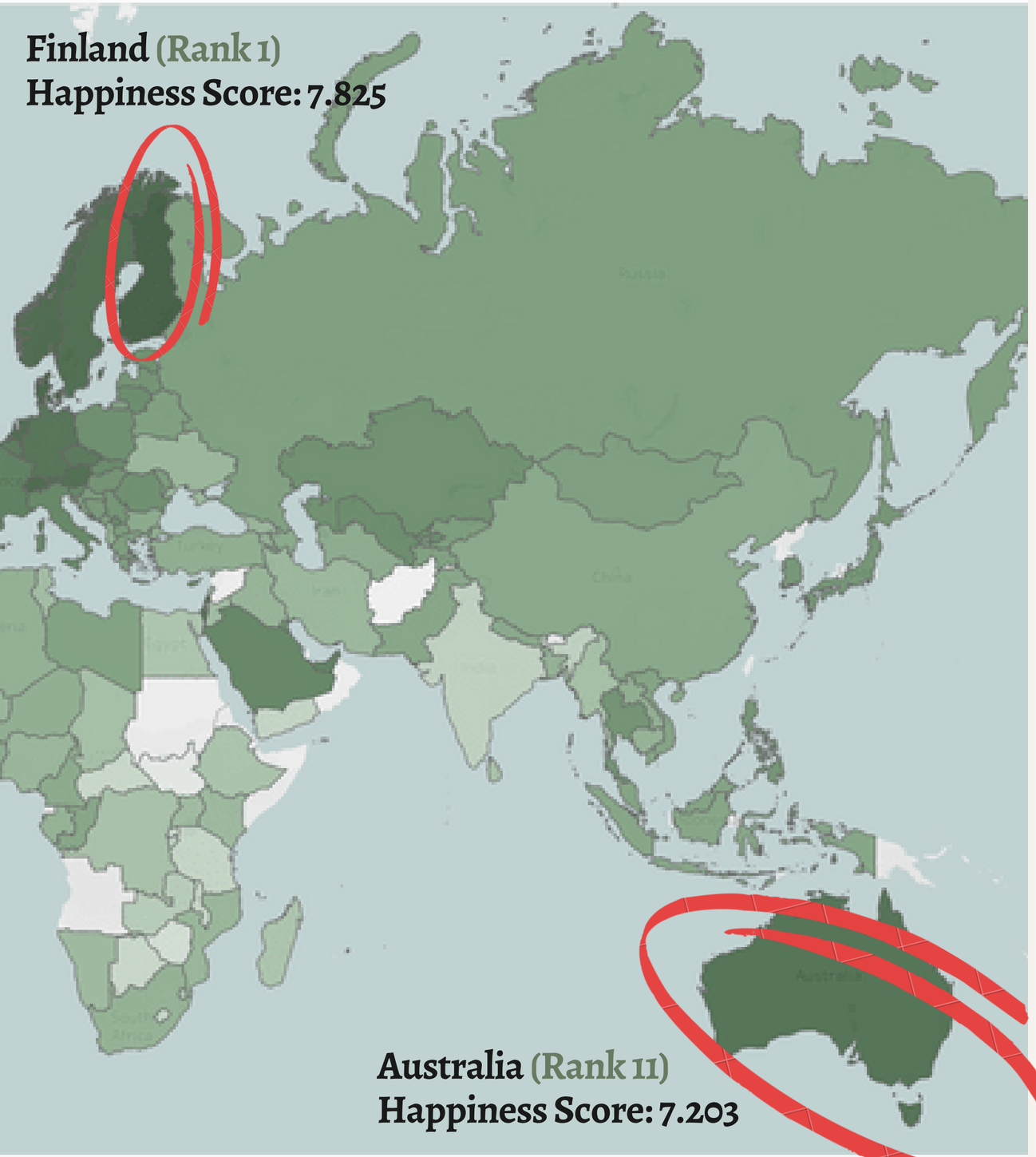
ISSUE

Scaling the user X input dataset will yield values of [0, 0, 0, 0, 0, 0, 0]

Scaling is relative and there are no other X value sets to scale against

RESOLUTION

- Bring in X_train dataset for X scaler
- Transform user X input dataset with X scaler



Conclusion

- Finland is the happiest country in the world
- Australia is 11th most happiest country in the world
- GDP/capita and social support are the most influential features

USER INTERACTION TIME!

Go to:
https://withtheflow.herokuapp.com/page3

- Enter in x variable feature values
- Click: PREDICT HAPPINESS SCORE

QUESTIONS?