SC1015
MINI PROJECT

Team 5
Jasmine Goh, Lau Fan Yu, Michael Ho

CONTENTS PAGE

01 Motivation/Rationale

Our quest for a high economic wellbeing to escape the rising cost of living may have potentially caused :

- Elevated stress levels
- Deteriorating mental health
- Increasing suicide rates

NTU set up University Wellbeing Office (UWO) to take care of students' mental wellbeing in April 2021.

Dataset: Suicide_Rates_Overview.csv

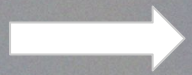
How does a country's economic well being affect suicide rates among different demographic categories?

02: Data Preparation and Cleaning

Dataset: Suicide_Rates_Overview.csv

- Filter relevant variables and put in DataFrame

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27820 entries, 0 to 27819
Data columns (total 12 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   country             27820 non-null  object
 1   year                27820 non-null  int64
 2   sex                 27820 non-null  object
 3   age                 27820 non-null  object
 4   suicides_no         27820 non-null  int64
 5   population          27820 non-null  int64
 6   suicides/100k pop   27820 non-null  float64
 7   country-year        27820 non-null  object
 8   HDI for year        8364 non-null   float64
 9   gdp_for_year ($)    27820 non-null  object
 10  gdp_per_capita ($)  27820 non-null  int64
 11  generation          27820 non-null  object
dtypes: float64(2), int64(4), object(6)
memory usage: 2.5+ MB
```



```
sData = suicideData[[
    'country',
    'year',
    'sex',
    'age',
    'suicides/100k pop',
    'HDI for year',
    'gdp_per_capita ($)'
]]
```

02: Data Preparation and Cleaning

Why did we choose?

- **suicides/100k pop instead of suicides_no**

  To have a more accurate basis of comparison of the suicide rates between countries, we used Suicides/100k pop instead of Suicides_no as each country has a different population.

- **gdp_for_year ($) instead of gdp_per_capita ($)**

  As each country's population is different. GDP Per Capita makes it relatively easier to compare across countries and to adjust for different levels of purchasing power from one country to the next.

02: Data Preparation and Cleaning

- Not every country counts their HDI every year
- **2014** is the most recent year with highest amount of country with HDIs

```
sData = sData[(sData['year'] == 2014 )]
```

- Drop countries with incomplete data in the year 2014

```
sData = sData.dropna()

sData = sData.reset_index(drop=True)
```

03: Exploratory Data Analysis

There are a total of 75 countries in our dataset. We intend to take the top 20% of countries in terms of economic wellbeing and suicides/100k population. In numerical terms, we will take the top 15 countries.

Economic Wellbeing: GDP & HDI

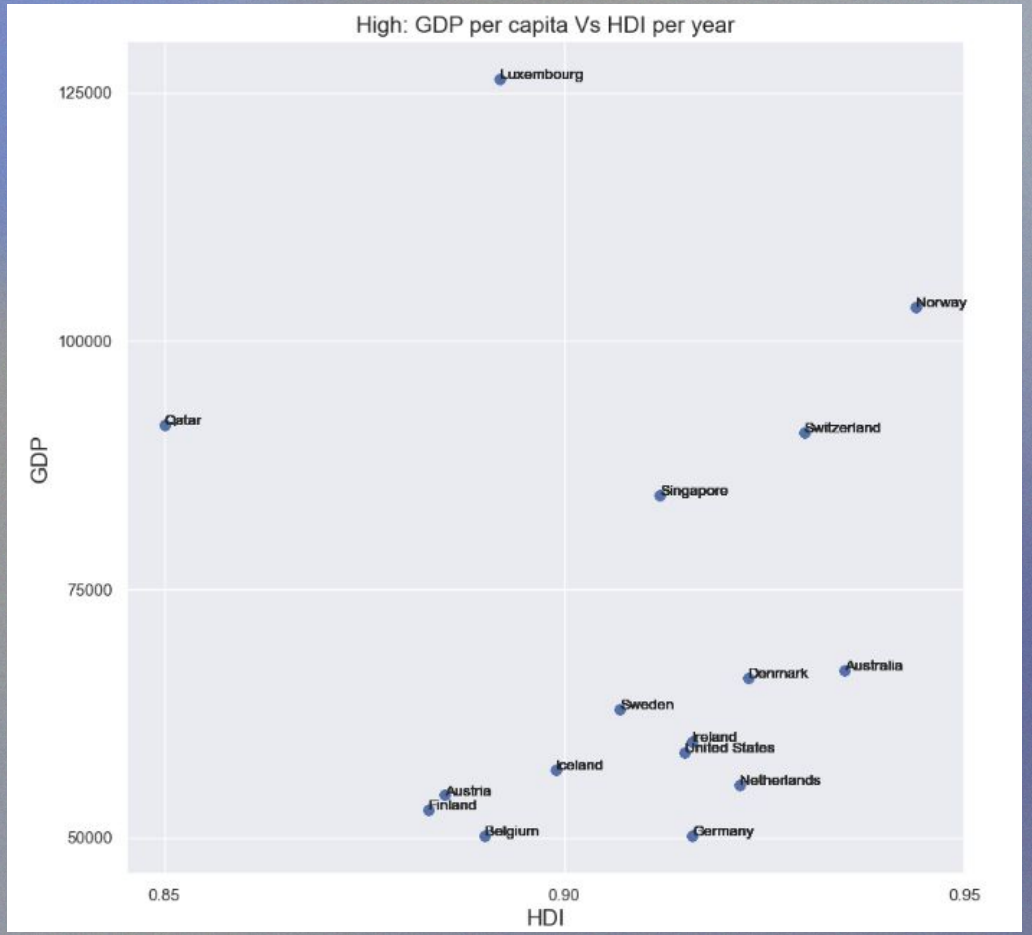
- Countries with the Highest Economic Wellbeing
- Countries with the Lowest Economic Wellbeing

Suicide Rate

- Countries with the Lowest Suicide Rate
- Countries with the Highest Suicide Rate

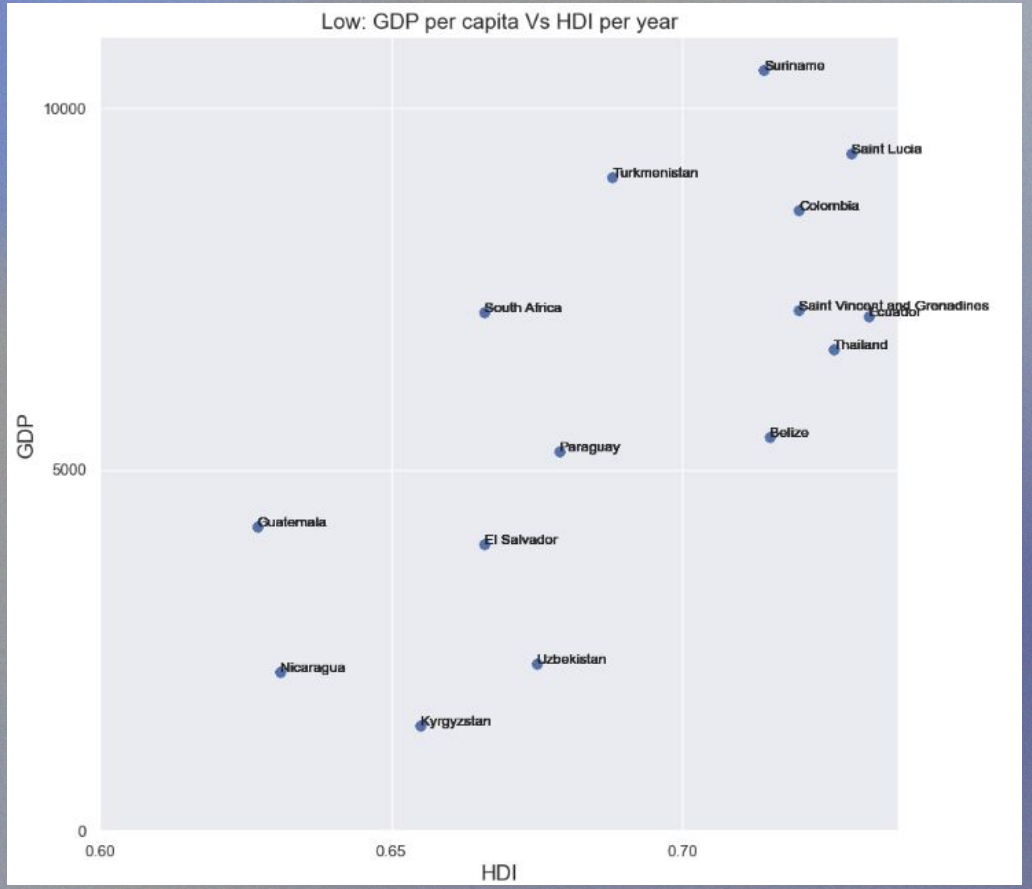
Economic Wellbeing:
HIGH GDP & HDI

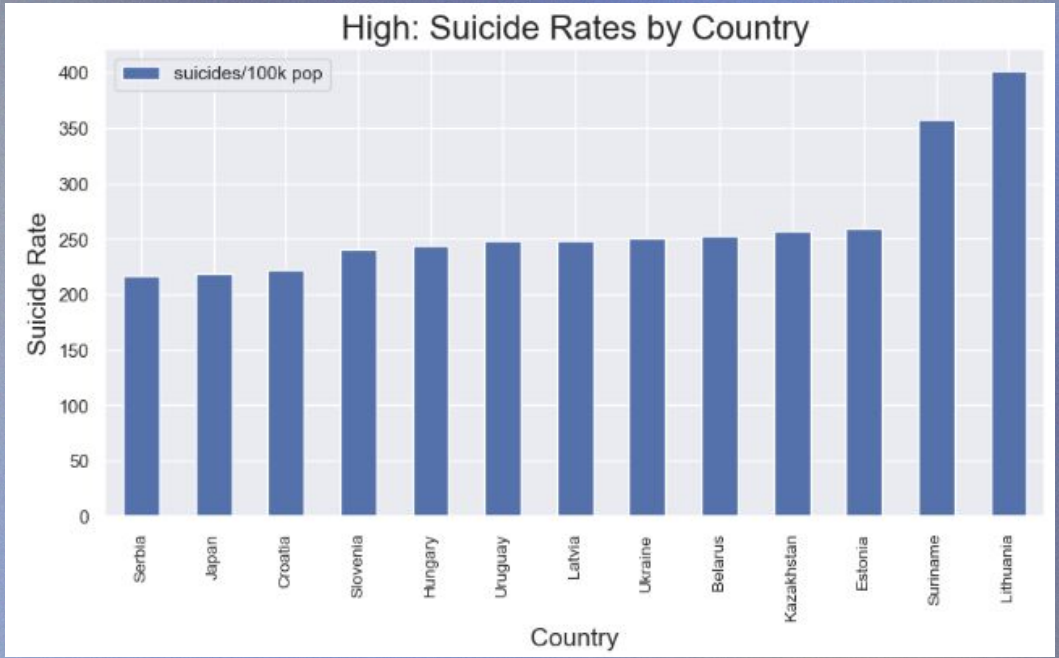
Dropped countries with
GDP per capita < 50000 and HDI < 0.85



Economic Wellbeing:
LOW GDP & HDI

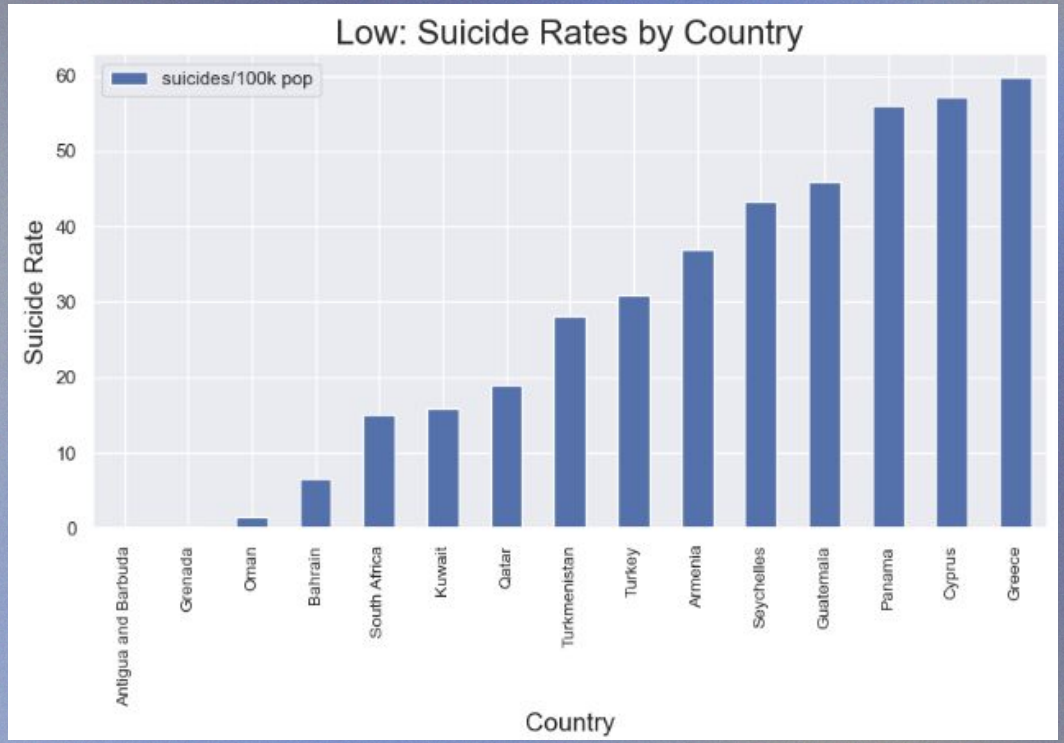
Dropped countries with
GDP per capita > 32000 and HDI > 0.70





Suicide Rate :
HIGH SUICIDE RATE

Dropped countries with
Suicide/100k pop < 210



Suicide Rate :
LOW SUICIDE RATE

Dropped countries with
Suicide/100k pop > 60

Summary of Economic Wellbeing & Suicide Rate

Highest economic wellbeing

-

['Australia', 'Austria', 'Belgium', 'Denmark', 'Finland', 'Germany', 'Iceland', 'Ireland', 'Luxembourg', 'Netherlands', 'Norway', 'Qatar', 'Singapore', 'Sweden', 'Switzerland', 'United States']

Lowest economic wellbeing

-

['Belize', 'Colombia', 'Ecuador', 'El Salvador', 'Guatemala', 'Kyrgyzstan', 'Nicaragua', 'Paraguay', 'Saint Lucia', 'Saint Vincent and Grenadines', 'South Africa', 'Suriname', 'Thailand', 'Turkmenistan', 'Uzbekistan']

Highest suicide rate

-

['Belarus', 'Croatia', 'Estonia', 'Hungary', 'Japan', 'Kazakhstan', 'Latvia', 'Lithuania', 'Serbia', 'Slovenia', 'Suriname', 'Ukraine', 'Uruguay']

Lowest suicide rate

-

['Antigua and Barbuda', 'Armenia', 'Bahrain', 'Cyprus', 'Greece', 'Grenada', 'Guatemala', 'Kuwait', 'Oman', 'Panama', 'Qatar', 'Seychelles', 'South Africa', 'Turkey', 'Turkmenistan']

Economic Wellbeing vs its Suicide Rate

Compare the Economic Wellbeing and the Suicide Rate

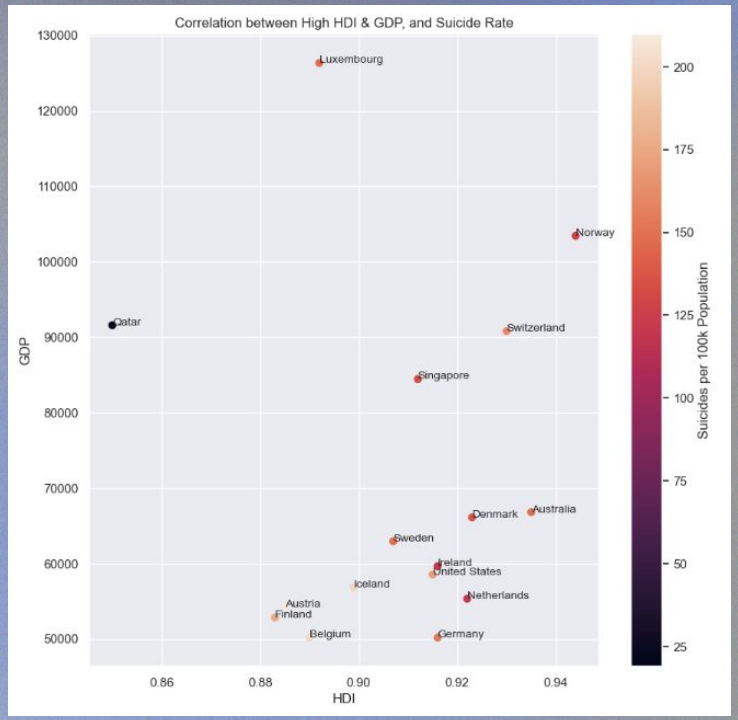
Countries with the highest economic wellbeing and lowest suicide rate
-
['Qatar']

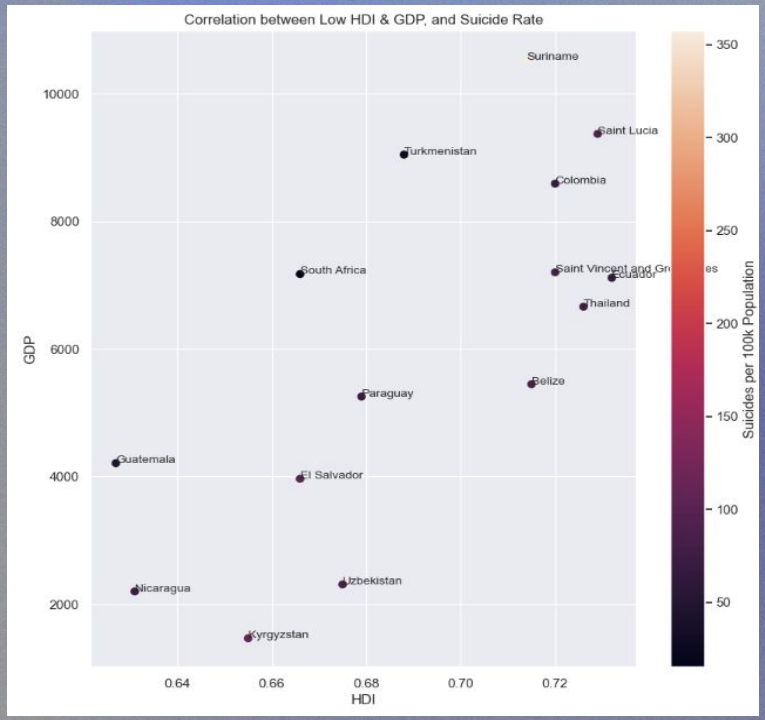
Countries with the highest economic wellbeing and highest suicide rate
-
[]

Countries with the lowest economic wellbeing and highest suicide rate
-
['Suriname']

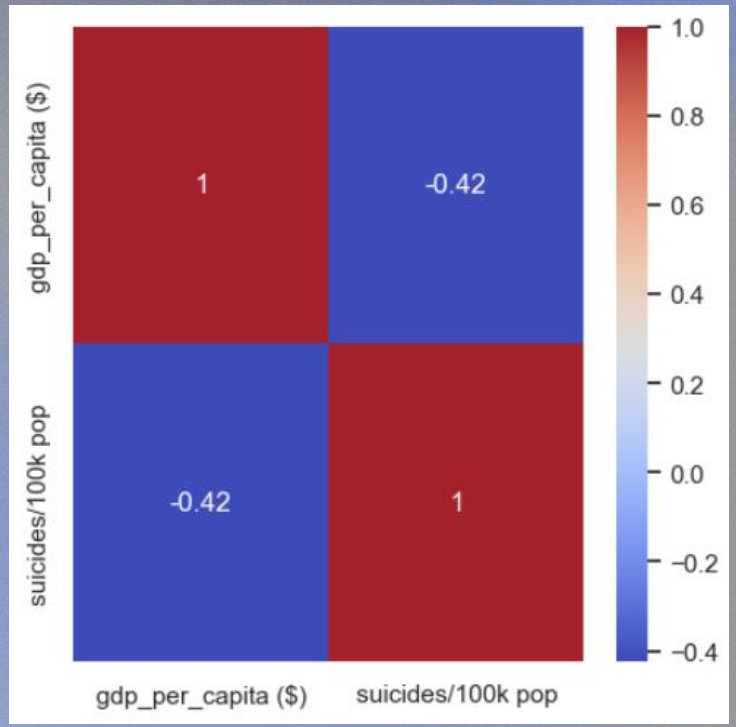
Countries with the lowest economic wellbeing and lowest suicide rate
-
['Guatemala', 'South Africa', 'Turkmenistan']

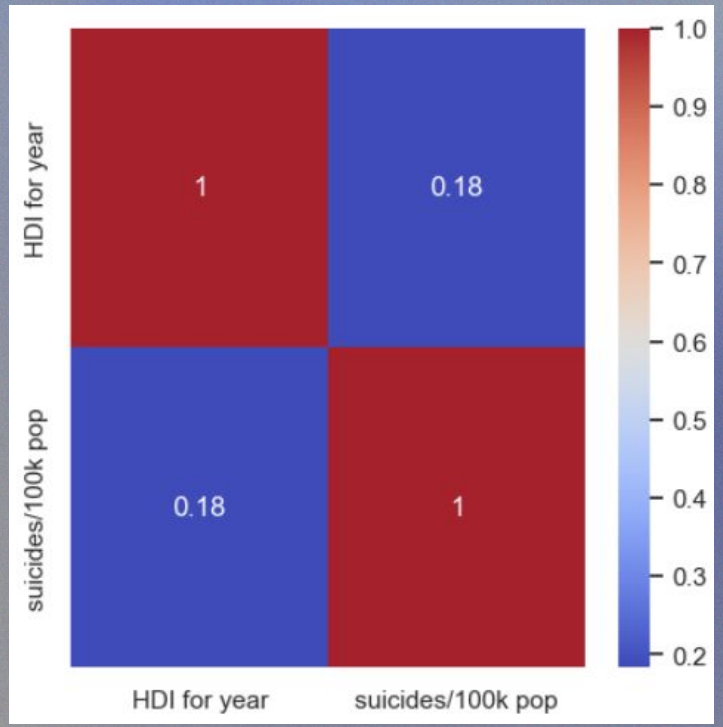
Correlation of HDI&GDP, and Suicide Rate



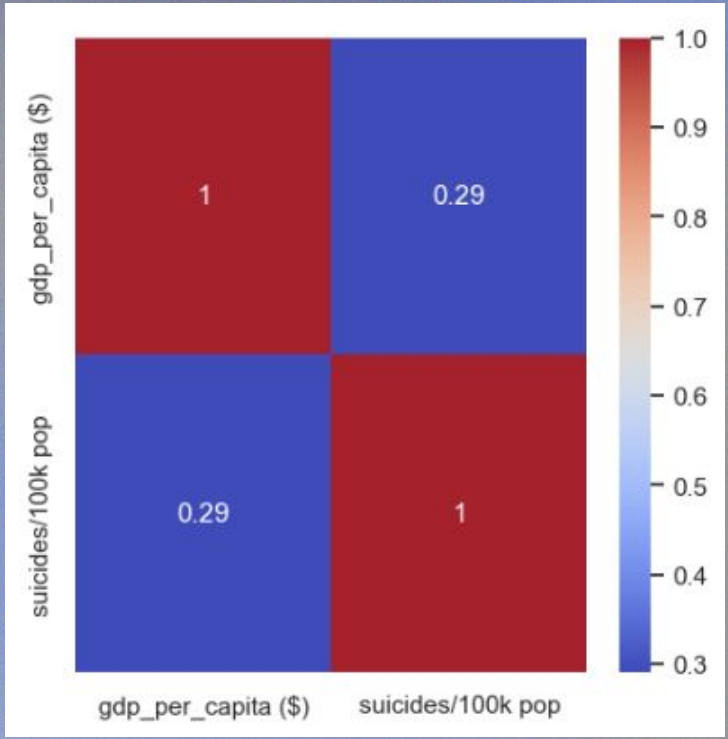


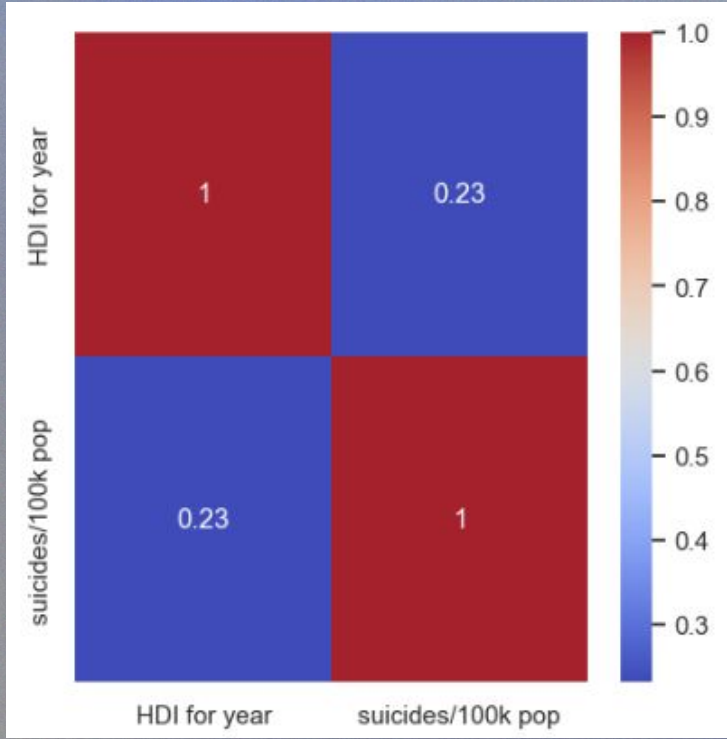
HEATMAP : High Econ vs Suicide Rate





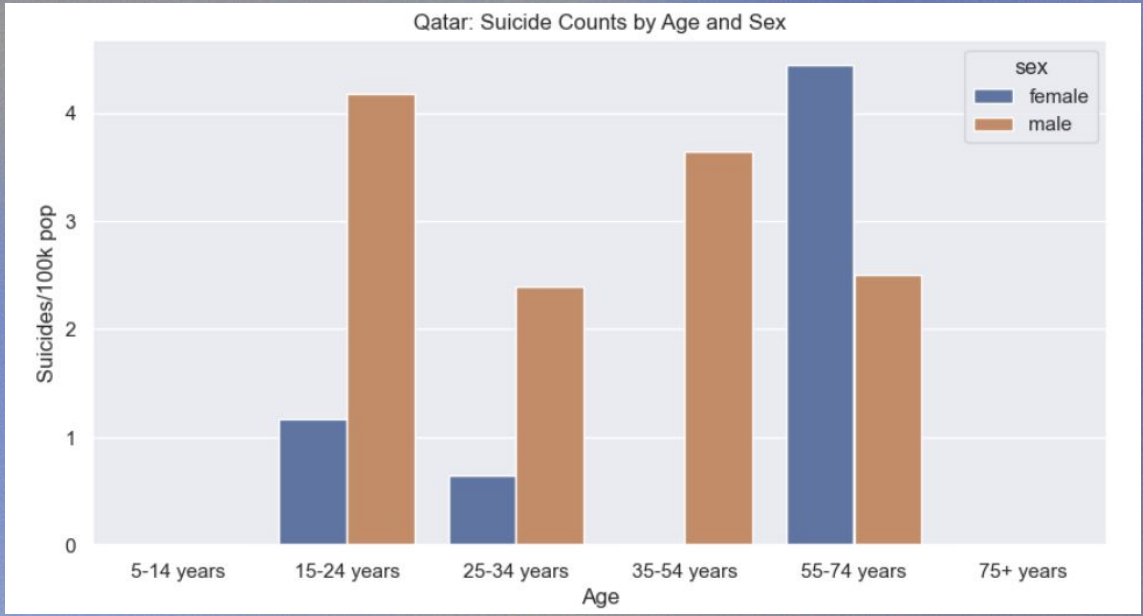
HEATMAP : Low Econ vs Suicide Rate





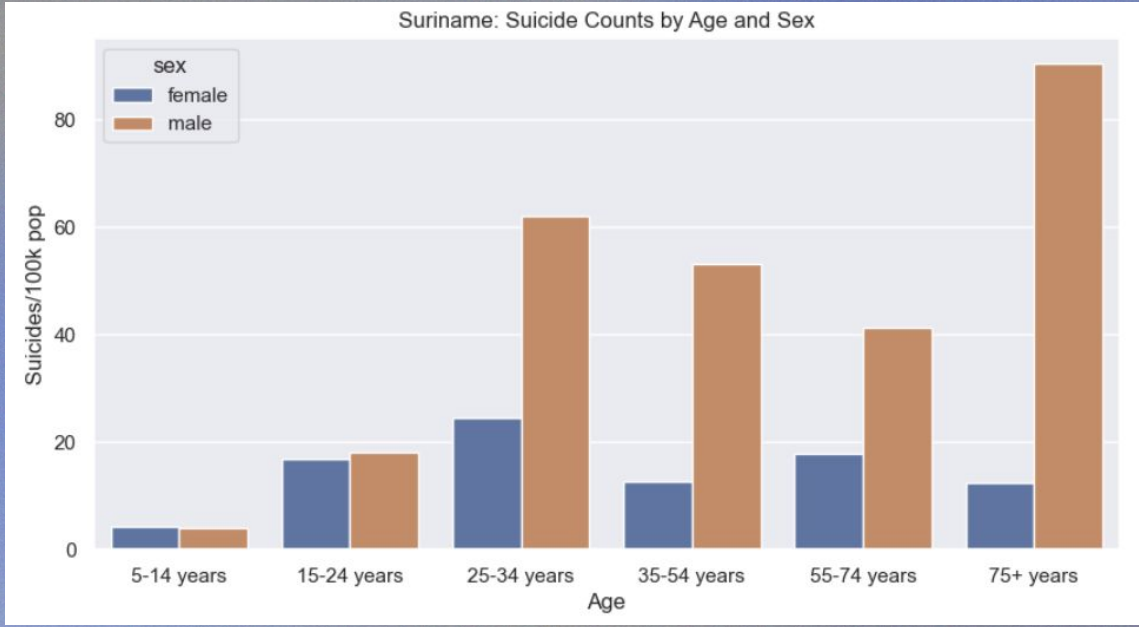
Demographic: High Econ Low Suicide

```
Countries with the highest economic wellbeing and lowest suicide rate
-
['Qatar']
```

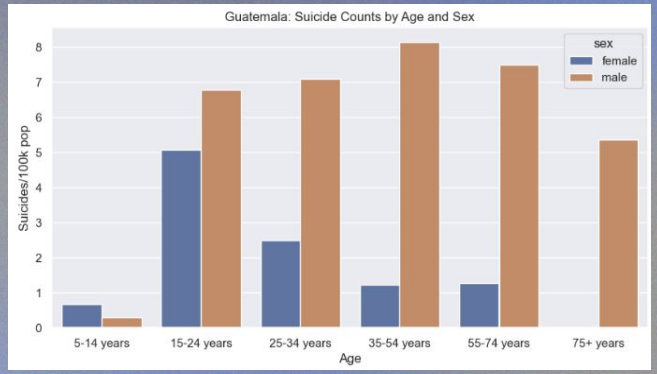


Demographic: Low Econ High Suicide

```
Countries with the lowest economic wellbeing and highest suicide rate
-
['Suriname']
```

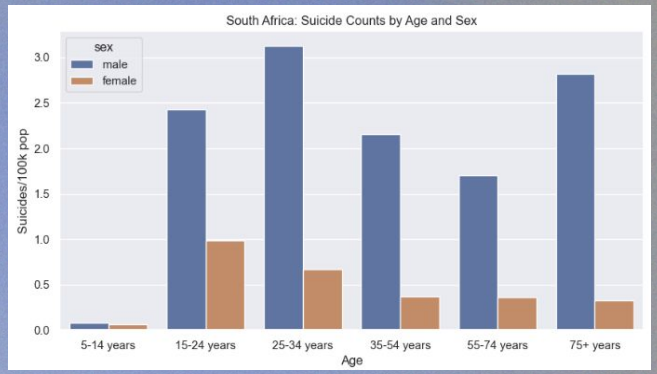


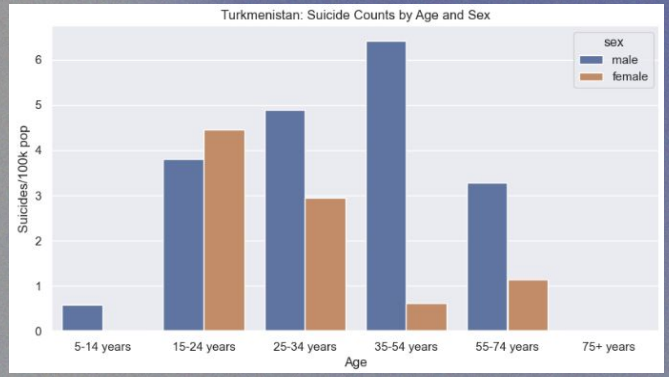
Demographic: Low Econ Low Suicide



Countries with the lowest economic wellbeing and lowest suicide rate -

['Guatemala', 'South Africa', 'Turkmenistan']





04: Machine learning techniques

Linear Regression of Suicides throughout the years in Singapore

```
Singapore = suicideData[[
    'country',
    'year',
    'sex',
    'age',
    'suicides/100k pop',
]]
Singapore  = Singapore[(Singapore['country'] == "Singapore" )]
```

	country	year	sex	age	suicides/100k pop
22400	Singapore	1985	male	75+ years	144.65
22401	Singapore	1985	female	75+ years	77.55
22402	Singapore	1985	male	55-74 years	31.86
22403	Singapore	1985	female	55-74 years	19.98
22404	Singapore	1985	male	25-34 years	18.54

```
Train Set : (24, 1) (24, 1)
Test Set  : (7, 1) (7, 1)
Intercept of Regression         : b =  [1247.35383338]
Coefficients of Regression      : a =  [[-0.61527502]]
```

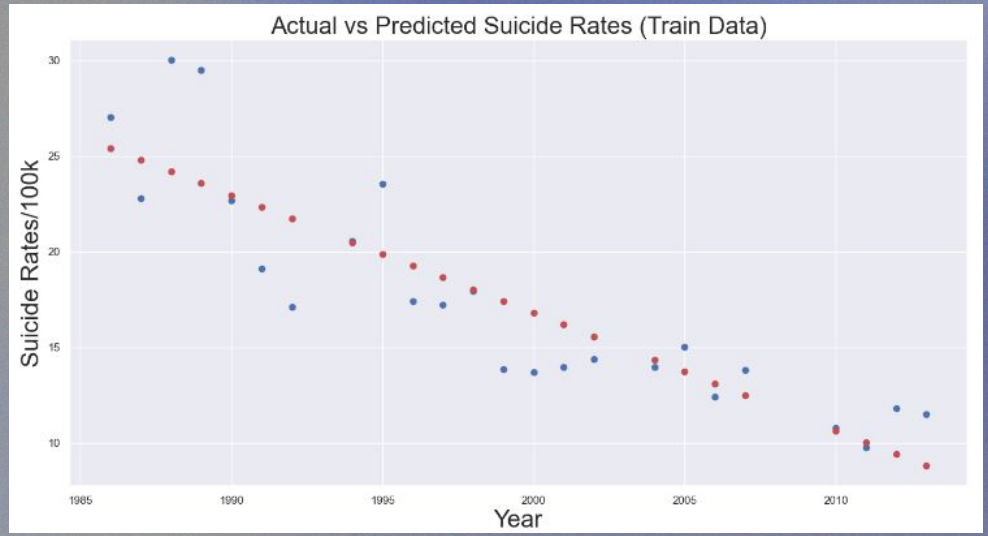
Split data into -> train: 80% test: 20%

Linear Regression of Suicides throughout the years in Singapore

Actual (Blue) vs Predicted Suicides Rates (Red) throughout the years in Singapore

Train Data

Goodness of Fit of Model	Train Dataset
Explained Variance (R^2)	: 0.8129937560527714
Mean Squared Error (MSE)	: 6.210716587845216
Root Mean Squared Error (RMSE)	: 2.492130933126351

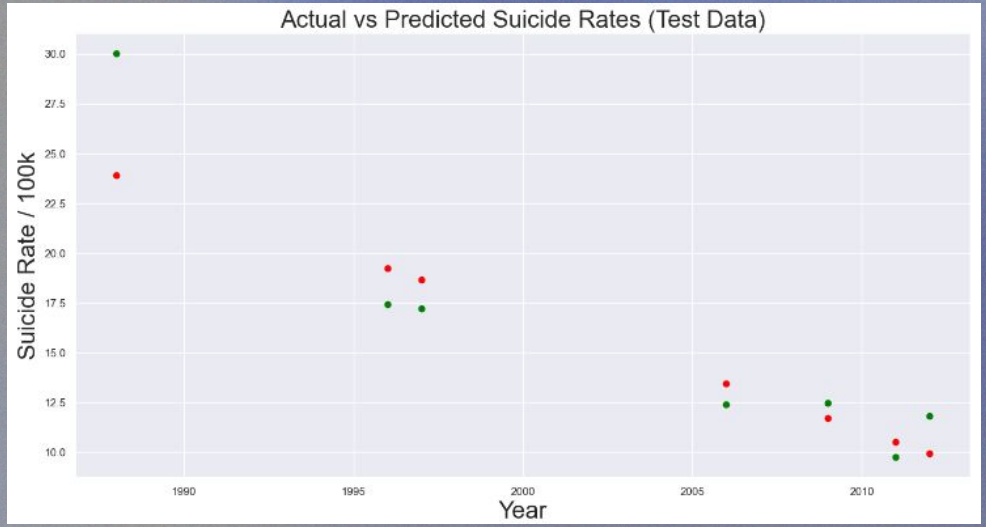


Linear Regression of Suicides throughout the years in Singapore

Actual vs Predicted Suicides Rates throughout the years in Singapore

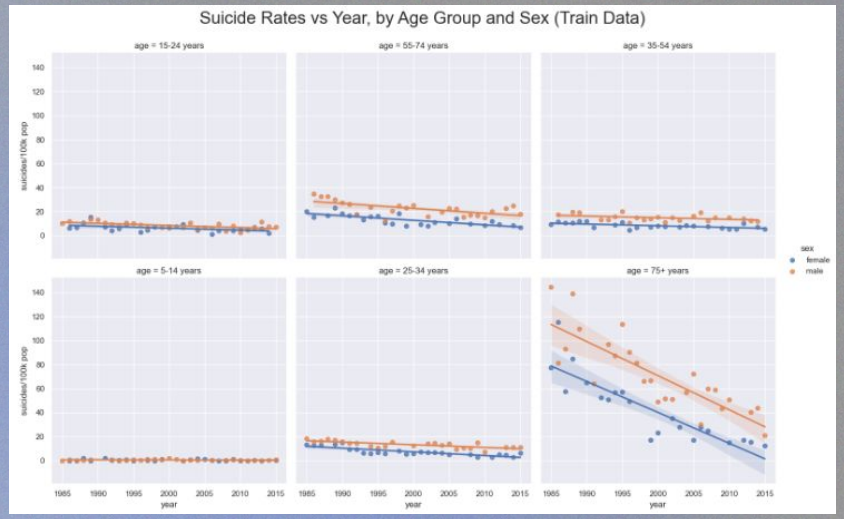
Test Data

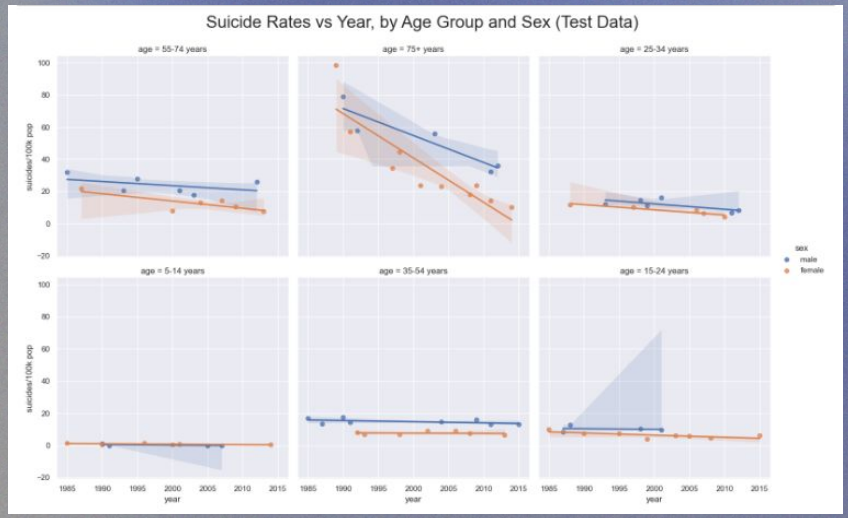
Goodness of Fit of Model	Test Dataset
Explained Variance (R^2)	: 0.8279740117603024
Mean Squared Error (MSE)	: 6.930591439963626
Root Mean Squared Error (RMSE)	: 2.632601648552934



Breakdown for different age groups across the years

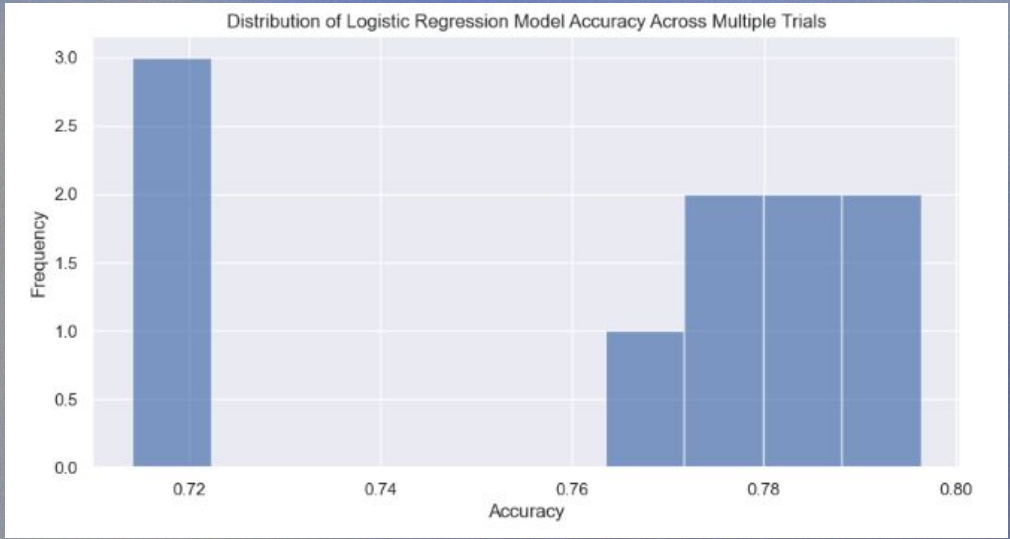
```
# split the data into training and testing sets
train_data, test_data = train_test_split(Singapore, test_size=0.2)
```





Logistic Regression of Data *(New Method)*

- ★ Predict gender of those likely to die by suicide
- ★ Able to provide more targeted support
- ★ Converts age into string type
- ★ Uses LabelEncoder & OneHotEncoder
- ★ Converts gender into numerical values 0 and 1



Accuracy: 77.05%

05: Data driven insights & recommendations

Project Outcome:

- Very weak correlation between a country's economic wellbeing and its suicide rate
- Since a relationship exists, a country's economic wellbeing can be considered a contributor towards suicide rate

Something Interesting:

- Distribution of suicide/100k pop varies greatly across countries
- Intangible factors contributing to this difference (Potentially : Culture, Biases)

Data-driven INSIGHTS and recommendations towards the target problem

Based on Machine Learning:

- Men are more likely to die by suicide than women
- Potentially caused by economic downturns; men work in more vulnerable industries
- Enforce policies that support men's mental health & reduce stigma associated with seeking help

SC1015

THANK YOU