World Happiness Report
Mini Project

CZ1015 Introduction to Data Science and Artificial Intelligence

Gupta Jay • Nguyen Duy Khanh • Tieu Phat Dat
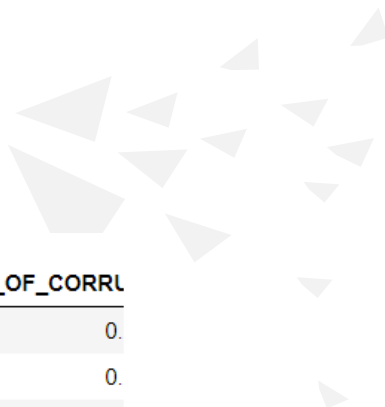
Problem Statements

Data Exploration & Visualisation	Classification & Regression	Data Exploration & Visualisation

What are the top factors which affect happiness?

Are there any wrong perceived factors about happiness?

Create your own Country

Predict it's Happiness Category

Find the most similar Country as yours

Economic Analysis of Singapore over the Years

Democratic Quality

Social Support

Life Satisfaction

Project Dataframe

Out[9]:

	COUNTRY	HAPPINESS_SCORE	GDP_PER_PERSON	HEALTHY_LIFE_EXPECTANCY	SOCIAL_SUPPORT	FREEDOM	GENEROSITY	PERCEPTIONS_OF_CORRU
0	Afghanistan	3.632	1741.687500	52.013329	0.525075	0.445294	0.179054	0.
1	Albania	4.586	11363.095700	68.871552	0.639576	0.726340	0.259975	0.
2	Algeria	5.295	13914.723630	65.604858	0.776977	0.439177	0.128988	0.
3	Angola	3.795	6260.132813	52.460709	0.765275	0.374173	0.106829	0.
4	Argentina	6.388	18807.310550	67.398483	0.905565	0.853390	0.163174	0.

Description of the dataset

COUNTRY : Name of each Country

HAPPINESS_SCORE :A metric measured between 2015 to 2017 by asking the sampled people the question: `"How would you rate your happiness on a scale of 0 to 10 where 10 is the happiest."`

GDP_PER_PERSON : GDP per Capita of each Country in terms of Puchasing Power Parity (PPP) (in USD)

HEALTHY_LIFE_EXPECTANCY : Healthy Life Expectancy at birth are constructed based on data from the World Health Organization (WHO) and WDI.

SOCIAL_SUPPORT : National average of the binary responses (either 0 or 1) to the question `"If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?"`

FREEDOM : National average of binary responses to the question `"Are you satisfied or dissatisfied with your freedom to choose what you do with your life?"`

GENEROSITY : Generosity is the residual of regressing the national average of responses to the question `"Have you donated money to a charity in the past month?"` on GDP per capita.

PERCEPTIONS_OF_CORRUPTION : Perceptions of corruption are the average of binary answers to two GWP questions: `"Is corruption widespread throughout the government or not?"` and `"Is corruption widespread within businesses or not?"`
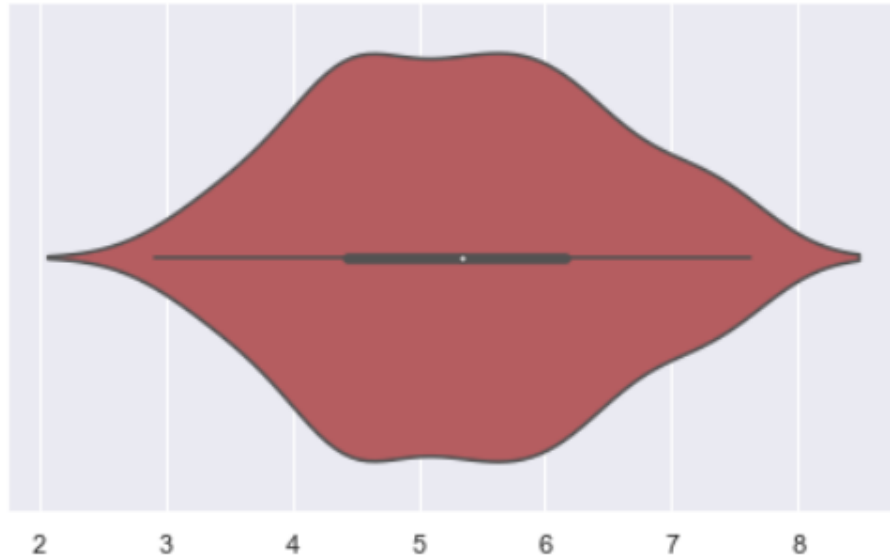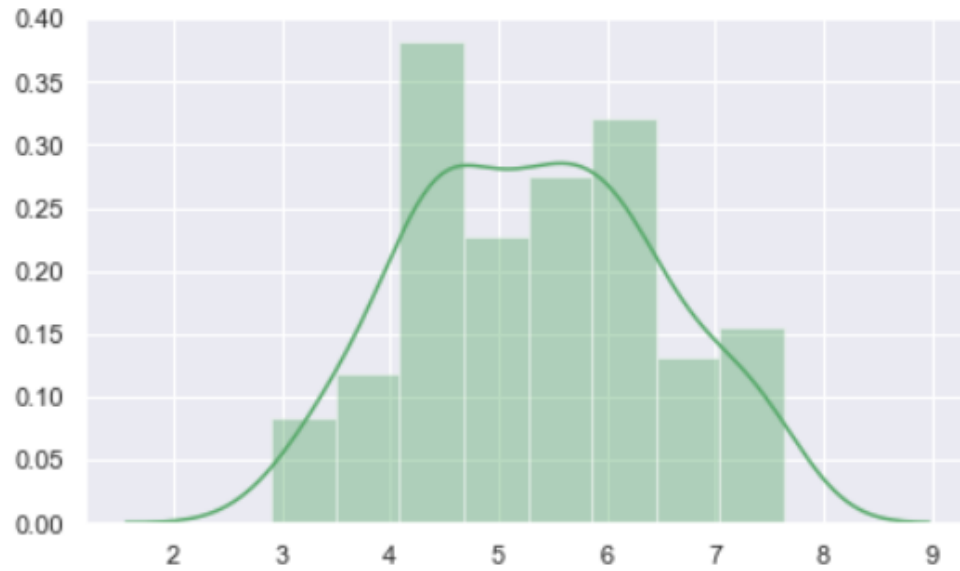
Visualising Happiness

Happiness Index 2018

Happiness Categorisation

```
Out[14]: <matplotlib.axes._subplots.AxesSubplot at 0x14279bfb588>
```


Happy
Happiness Score < 4

Likely Happy / Normal
4 < Happiness Score < 6

Unhappy
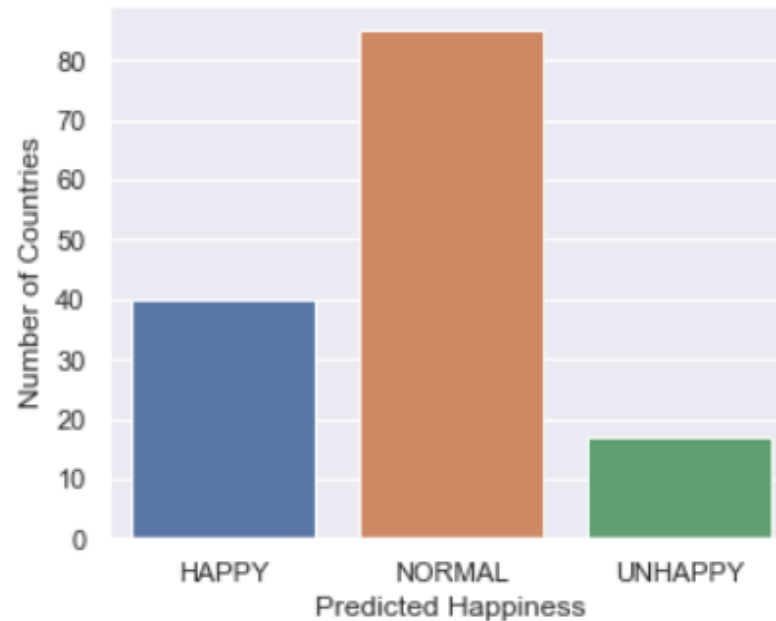Happiness Score > 6

Visualising Happiness

Top factors affecting happiness:

1. GDP Per Capita
2. Life Expectancy
3. Social Support
4. Generosity

Create your own Country using Data Science

Predict **Happiness Category** and the **most similar** Country to our Virtual Country

```
Enter your Country Name: Zootopia
--------------------------------
Let us analyse the feasabiliy [Happiness Category/Score] of Zootopia by entering some important features:


--------------------------------

Reference: Average GDP Per Capita of all Countries: USD 17000, Singapore's GDP Per Capita: USD 81000
Enter the Average GDP Per Capita for your Country (in USD): 40000


--------------------------------

Enter the Average Life Expectancy of your Citizens: (Age) 71.5


--------------------------------

Reference Question: Social Support - If you were in trouble, do you have relatives or friends you can count on to help you when
ever you need them, or not?
Enter the Social Support Index (Choose between 0 and 1, 0 - Lowest : 1 - Highest): 0.95


--------------------------------

Reference Question: Generosity - Have you donated money to a charity in the past month?
How generous are people in your Country? (Choose between 0 and 1, 0 - Lowest : 1 - Highest): 0.94


--------------------------------


Congrats! We can say that the citizens of Zootopia will be HAPPY with a probability of 98.47 %


--------------------------------
```
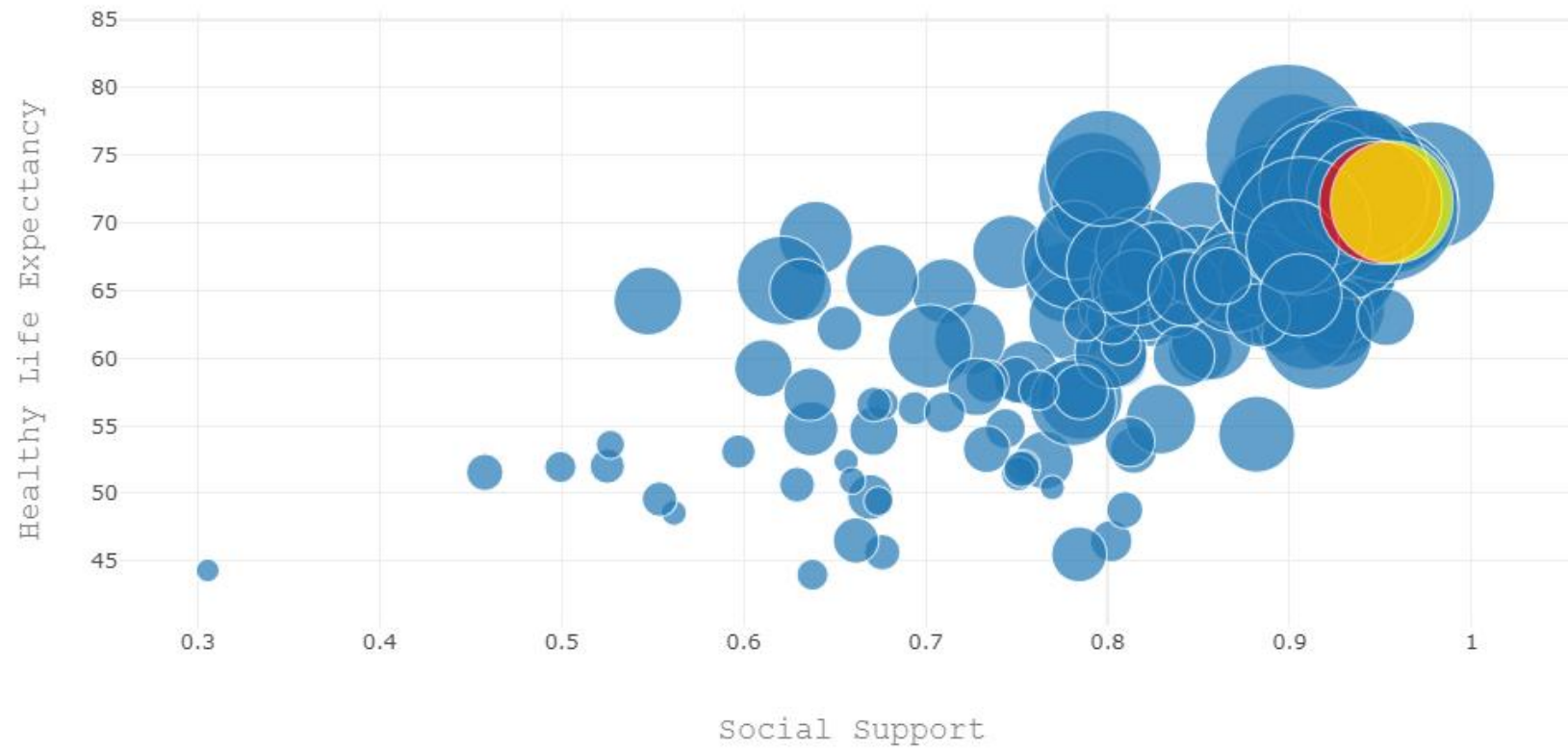

Legend:
Red: Your Country
Yellow: The Country Most Similar to your Country

Finland is the country that has the most similar features to yours

----

Machine Learning
Predicting Happiness

Classification – Random Forest

- Random Forest is used to predict the Happiness Category i.e. Happy, Normal and Unhappy using GDP per Capita, Healthy Life Expectancy, Social Support and Generosity.

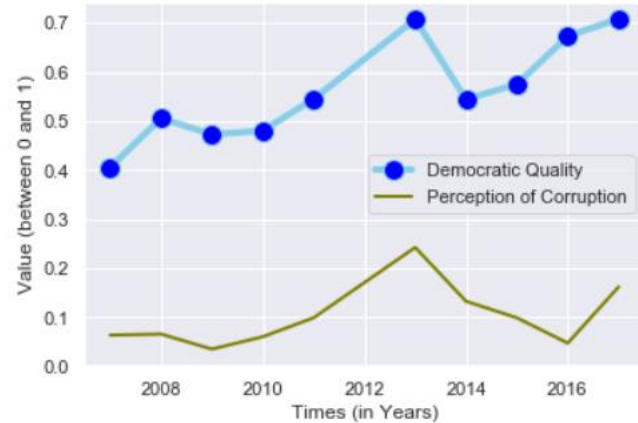- We can get an accuracy of up to **75%** by using the Random Forest Classification Algorithm.

Regression – Linear Regression

- Linear Regression is used to predict which Country is the most similar to our virtual country in terms of GDP per Capita, Healthy Life Expectancy and Social Support.

- We can get an accuracy of up to **70%** by using the Linear Regression algorithm.
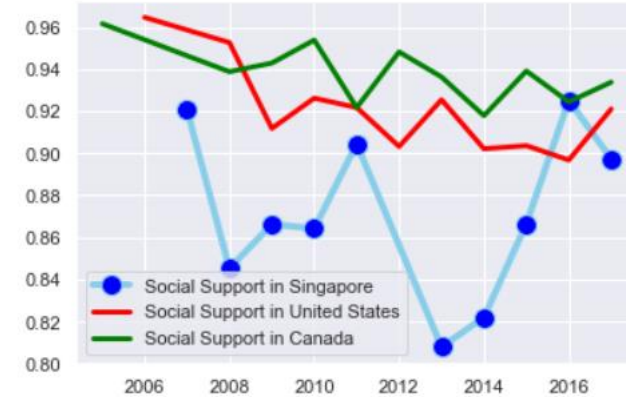
Economic Analysis of Singapore

Democratic Quality

Corruption in Singapore is generally perceived as low. However, a few minor scandals led to an sharp dip in Democratic Quality of Singapore.

We can infer from the following observations that an increase in the Perception of Corruption led to a decrease in the Democratic Quality in the subsequent years.

Social Support

The government has provided schemes and subsidies to lower-income groups. For example, the Workfare Income Supplement Scheme tops up their cash earnings and Central Provident Fund accounts, and helps them build up their retirement savings.

Due to these policies, we can see an increase in Social Support of Singapore from 2013 onwards.

Economic Analysis of Singapore
Life Satisfaction

We can observe a sharp downtrend in life satisfaction from 2014 to 2016.

This can be attributed to the fact that the finance & insurance and health-care sectors didn't perform well in the following years and registered a 1.44 point decline from the previous year.
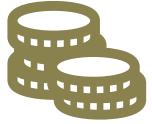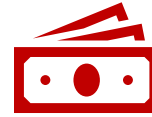
Project Distribution

Gupta Jay	Tieu Phat Dat	Nguyen Duy Khanh
Problem Statements	Problem Statements	Problem Statements
Data Preparation and Cleaning	Data Preparation and Cleaning	Data Preparation and Cleaning
Basic Statistics & Exploratory Data Analysis / Visualisation	Section 2: Create your own Country - Regression Models (Linear Regression, Lasso Regression, Ridge Regression and GridSearchCV)	Section 1: What are the top factors which affect happiness?
Section 1: What are the top factors which affect happiness? Are there any wrong perceived factors about happiness? Section 2: Create your own Country - Classification Model (Random Forest)	Exception Handling	Section 2: Create your own Country - Classification Model (Random Forest)
Economic Analysis of Singapore		

Interesting Observations

Corruption and Democracy

- We infer from the following observations that an increase in the **Perception of Corruption** led to a **decrease** in the **Democratic Quality** in the subsequent years of Singapore.

Money != Happiness

- Out of the top factors identified which contribute to happiness, GDP per Capita is the only factor which relates to money.

- All other factors such as Life Expectancy, Generosity and Social Support contribute significantly in happiness of a Country.

Thank You