

# Global Happiness Insights: Prediction with Economic indicators

Presented by:

Yasir Ali Soomro, Parshant Mahajan ,  
Penyaree Wening , and Ninad Pandya



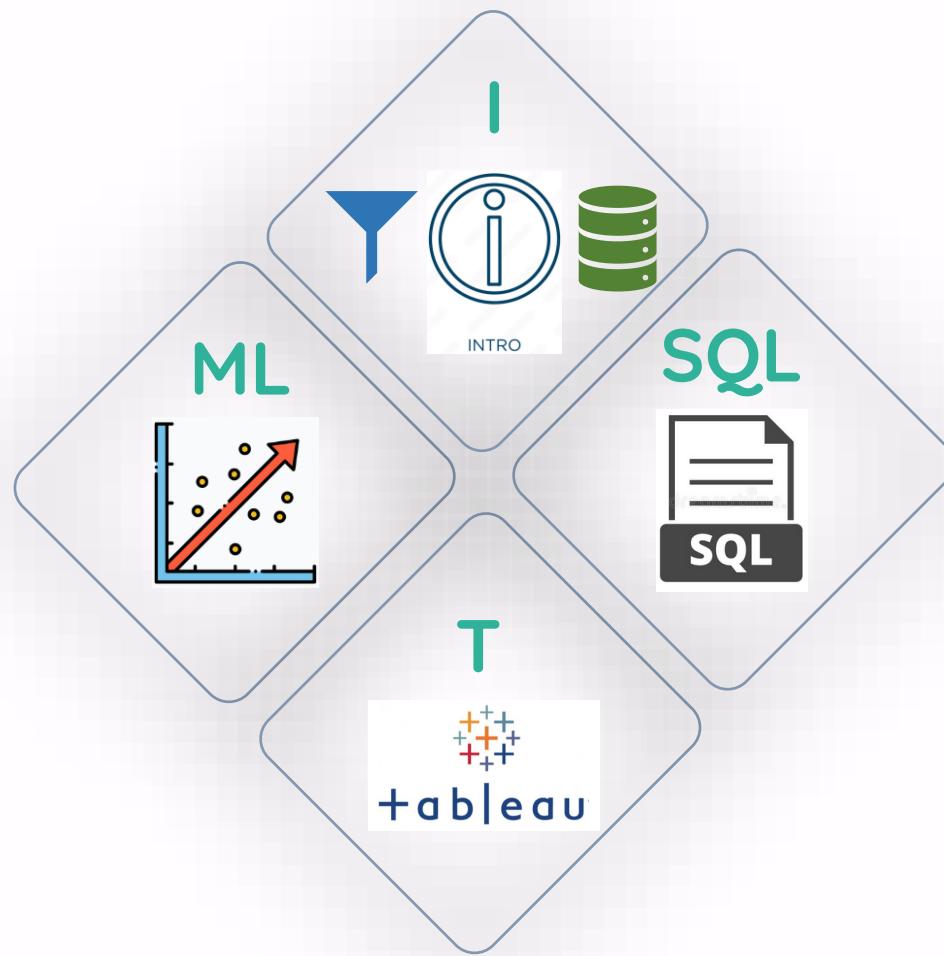
# Table of Contents

## Introduction

Introduction into our project, including purpose, datasets used and intro to other sections of presentation.

## Machine Learning

Happiness prediction analysis using Linear Regression and Random Forest technique.



## SQL Analysis

Data set used, examples of codes used, comparison and creation of relational database.

## Tableau Analysis

Creation of visualizations in Tableau to provide an interactive platform for exploring happiness trends.

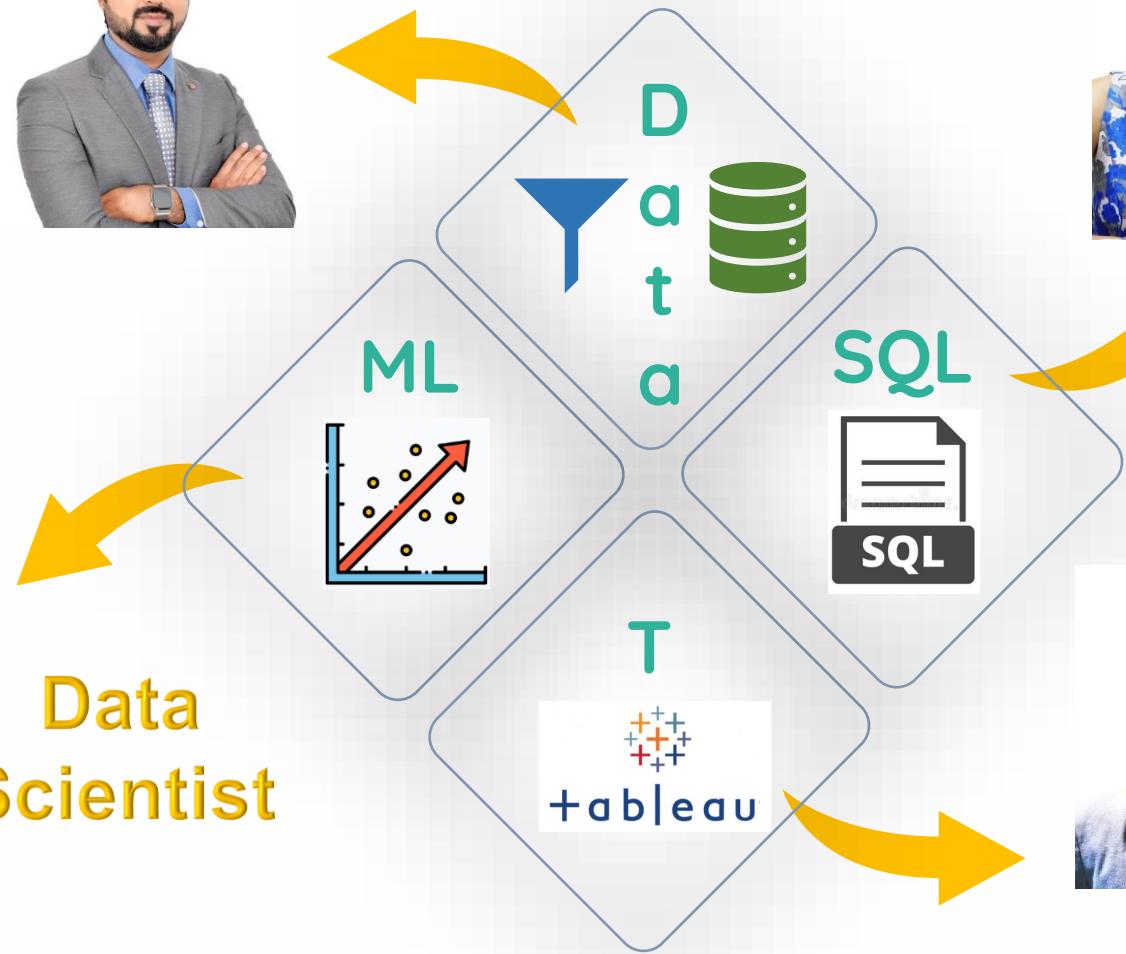
# The Team Roles

Data  
Engineer



Data  
Scientist

Data  
Analyst



# 01

## Introduction

The purpose of this project is to increase understanding of global happiness and well-being by leveraging advanced data analytics, machine learning, and visualization techniques, ultimately providing actionable insights for a happier world.

By analyzing large datasets and applying machine learning algorithms, we can uncover patterns and factors that contribute to happiness scores. These predictive models can help identify key drivers of happiness and inform policy decisions aimed at improving well-being worldwide.

The interactive Tableau Visualization Dashboard will allow users to explore happiness trends across different countries and time periods, enabling a deeper understanding of the factors that influence happiness levels.



- **Source:** Happiness Dataset: <https://worldhappiness.report/data/>
- Unemployment Dataset: <https://data.worldbank.org/indicator/SL.UEM.TOTL.NE.ZS>
- CPI dataset: <https://tinyurl.com/248hffft>

# PROBLEM STATEMENT:

Understanding factors contributing to human happiness is crucial for policymakers, researchers, and society. The World Happiness Report dataset provides valuable insights into global happiness trends, but it also presents challenges that need to be addressed.

- ◆ Addressing these challenges is essential for a more comprehensive understanding of happiness. Moreover, how economic factors effect the happiness score.



# Happiness Factors

The World Happiness Report evaluates and ranks countries based on a variety of factors that contribute to overall happiness and well-being. These factors typically include:

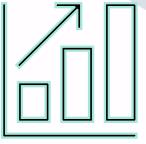
1. GDP per Capita
  2. Social Support
  3. Life Expectancy
  4. Freedom to Make Life Choices
  5. Generosity
  6. Perceptions of Corruption
- +
7. Unemployment Rate
  8. Inflation (CPI)



$$\text{Happiness} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \varepsilon_Y$$



**Unemployment  
Rate**  
**2010-2022**



**Inflation  
Rate (CPI)**  
**2010-2022**



In [8]:

```
# Categorizing countries as per development status and adding new column.

criteria = {
    'Developed': ['Singapore', 'Belgium', 'New Zealand', 'Canada', 'Norway', 'Netherlands', 'Finland', 'Iceland', 'Ireland', 'Germany', 'Sweden', 'Denmark', 'Australia', 'Hong Kong S.A.R. of China', 'Switzerland', 'United Kingdom', 'United States', 'Canada', 'France', 'Germany', 'Luxembourg', 'Japan', 'South Korea', 'Israel', 'Slovenia', 'Malta', 'Austria', 'Spain', 'France', 'Cyprus', 'Italy', 'Czechia', 'Greece', 'Poland', 'Saudi Arabia', 'Lithuania', 'Georgia', 'Hungary', 'Romania', 'Argentina', 'Kazakhstan'],
    'Developing': ['Albania', 'Algeria', 'Armenia', 'Azerbaijan', 'Bahrain', 'Belarus', 'Belize', 'Bolivia', 'Bosnia and Herzegovina', 'Brazil', 'Bulgaria', 'Cameroon', 'Chile', 'China', 'Colombia', 'Costa Rica', 'Croatia', 'Dominican Republic', 'El Salvador', 'Estonia', 'Eswatini', 'Gabon', 'Ghana', 'Guatemala', 'Honduras', 'India', 'Indonesia', 'Iran', 'Jordan', 'Kenya', 'Kosovo', 'Kuwait', 'Kyrgyzstan', 'Latvia', 'Lebanon', 'Libya', 'Malaysia', 'Mauritius', 'Mongolia', 'Montenegro', 'Morocco', 'Namibia', 'Nicaragua', 'Nigeria', 'North Macedonia', 'Oman', 'Pakistan', 'Peru', 'Philippines', 'Portugal', 'Qatar', 'Slovakia', 'South Africa', 'Sri Lanka', 'State of Palestine', 'Thailand', 'Trinidad and Tobago', 'Tunisia', 'Turkiye', 'Ukraine', 'Uruguay', 'Uzbekistan', 'Venezuela', 'Vietnam'],
    'Least Developed': ['Afghanistan', 'Angola', 'Bangladesh', 'Benin', 'Bhutan', 'Botswana', 'Burkina Faso', 'Burundi', 'Cambodia', 'Central African Republic', 'Chad', 'Djibouti', 'Comoros', 'Congo (Brazzaville)', 'Congo (Kinshasa)', 'Ethiopia', 'Gambia', 'Guinea', 'Haiti', 'Laos', 'Lesotho', 'Liberia', 'Madagascar', 'Malawi', 'Maldives', 'Mali', 'Mauritania', 'Mozambique', 'Myanmar', 'Nepal', 'Niger', 'Rwanda', 'Senegal', 'Sierra Leone', 'Somalia', 'South Sudan', 'Sudan', 'Tanzania', 'Togo', 'Uganda', 'Yemen', 'Zambia']
}

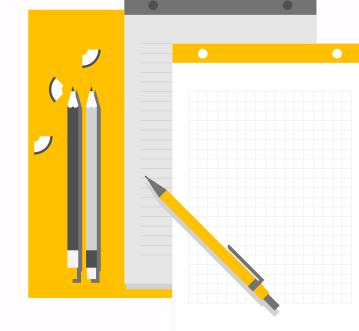
# Function to categorize countries based on criteria
def categorize_country(Country):
    for category, country_list in criteria.items():
        if Country in country_list:
            return category
    return 'Uncategorized' # Default category if not found in any criteria

# Create a new 'Development_Status' column based on the categorization function
df5['Development_Status'] = df5['Country'].apply(categorize_country)

# Display the updated DataFrame
df5
```

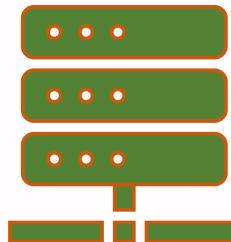
Out[8]:

Year	Happiness Score	Log GDP per capita	Social support	Healthy life expectancy at birth	Freedom to make life choices	Generosity	Perceptions of corruption	CPI %	Unemployment Rate	region	Development_Status
2010	4.758	7.614	0.539	51.1	0.6	0.121	0.707	2.18	Null	Asia	Least Developed
2011	3.832	7.581	0.521	51.4	0.496	0.164	0.731	11.8	Null	Asia	Least Developed
2012	3.783	7.661	0.521	51.7	0.531	0.238	0.776	6.44	1.69	Asia	Least Developed
2013	3.572	7.68	0.484	52.0	0.578	0.063	0.823	7.39	Null	Asia	Least Developed
2014	3.131	7.671	0.526	52.3	0.509	0.106	0.871	4.67	7.91	Asia	Least Developed
...	...	...	...	...	...	...	...	...	...	...	...
2018	3.616	7.783	0.775	52.625	0.763	-0.051	0.844	10.62	Null	Africa	Developing
2019	2.694	7.698	0.759	53.1	0.632	-0.047	0.831	255.3	7.37	Africa	Developing



# Data Cleaning Python

Dictionaiy used  
Pandas  
Numpy





# 02

---

## SQL Analysis

# Creation of various relation database and Analysis

```
CREATE Table Happiness_development_status (
    Country VARCHAR ,
    Year INT ,
    Happiness_Score Float ,
    Log_GDP_per_capita Float ,
    Social_support Float ,
    Healthy_life_expectancy_at_birth Float ,
    Freedom_to_make_life_choices Float ,
    Generosity Float ,
    Perceptions_of_corruption Float ,
    CPI_percentage Float ,
    Unemployment_Rate Float ,
    region VARCHAR ,
    Development_Status VARCHAR
);

SELECT*
FROM Happiness_development_status
```



```
SELECT DISTINCT year, country,Happiness_Score
FROM Happiness_development_status
WHERE year = 2022
ORDER BY Happiness_Score DESC
LIMIT 10;
```

```
SELECT DISTINCT year, country,Happiness_Score
FROM Happiness_development_status
WHERE year = 2022
ORDER BY Happiness_Score ASC
LIMIT 10;
```

Year 2022: Top 10 country high happiness score = Finland,Isael,Denmark,Iceland, Swenden, Netherlands,Norway,Luxembourg,Costa Rica,Lithuania

Year 2022: Top 10 country low happiness score = Afghanistan,lebanon,Sierra Leone,Congo(Kinshasa), Zimbabwe,Malawi,Bangladesh,Botswana,Comoros,Tanzania



# Data Analysis: Some Facts discovered

- Which region has the highest and lowest happiness score?

```
SELECT region, ROUND(AVG(Happiness_Score)::numeric, 3) as Avg_Happiness_Score  
FROM Happiness_development_status  
GROUP BY region  
ORDER BY Avg_Happiness_Score DESC;
```



	region character varying	avg_happiness_score numeric
1	Australian and Oceania	7.245
2	North America	6.977
3	Western Europe	6.757
4	Central America	6.458
5	South America	6.015
6	Central American	5.956
7	Middle East	5.553
8	Eastern Europe	5.529
9	Asia	5.327
10	Caribbean	5.319
11	Africa	4.555
12	Western Africa	4.204
13	AfricaTrinidad and TobagoCaribbean	3.688
14	Souther Africa	2.905

# Data Analysis: Some Facts discovered

- Which countries are the happiest countries each year in the Dataset?

```
SELECT year, country, Happiness_Score
FROM (
    SELECT wh.year, wh.country, wh.Happiness_Score,
           ROW_NUMBER() OVER (PARTITION BY wh.year ORDER BY wh.Happiness_Score DESC) as row_num
    FROM Happiness_development_status wh
    JOIN (
        SELECT year, MAX(Happiness_Score) as max_score
        FROM Happiness_development_status
        WHERE year IN (2022, 2021, 2020, 2019, 2018, 2017,2016,2015,2014,2013,2012,2011,2010)
        GROUP BY year
    ) max_scores
    ON wh.year = max_scores.year AND wh.Happiness_Score = max_scores.max_score
) ranked_data
WHERE row_num = 1
ORDER BY year;
```

[Data Output](#)[Messages](#)[Notifications](#)

	year integer	country character varying	happiness_score double precision
1	2010	Denmark	7.771
2	2011	Denmark	7.788
3	2012	Switzerland	7.776
4	2013	Canada	7.594
5	2014	Denmark	7.508
6	2015	Norway	7.603
7	2016	Finland	7.66
8	2017	Finland	7.788
9	2018	Finland	7.858
10	2019	Finland	7.78
11	2020	Finland	7.889
12	2021	Finland	7.794
13	2022	Finland	7.729

- Which countries are the least happy?

```
SELECT year, country, Happiness_Score
FROM (
    SELECT wh.year, wh.country, wh.Happiness_Score,
           ROW_NUMBER() OVER (PARTITION BY wh.year ORDER BY wh.Happiness_Score DESC) as row_num
    FROM Happiness_development_status wh
    JOIN (
        SELECT year, MIN(Happiness_Score) as min_score
        FROM Happiness_development_status
        WHERE year IN (2022, 2021, 2020, 2019, 2018, 2017, 2016, 2015, 2014, 2013, 2012, 2011, 2010)
        GROUP BY year
    ) min_scores
    ON wh.year = min_scores.year AND wh.Happiness_Score = min_scores.min_score
) ranked_data
WHERE row_num = 1
ORDER BY year;
```

Data Output    Messages    Notifications



	year integer	country character varying	happiness_score double precision
1	2010	Bulgaria	3.912
2	2011	Togo	2.936
3	2012	Madagascar	3.551
4	2013	Egypt	3.559
5	2014	Burundi	2.905
6	2015	Madagascar	3.593
7	2016	Liberia	3.355
8	2017	Afghanistan	2.662
9	2018	Rwanda	3.561
10	2019	Zimbabwe	2.694
11	2020	Tanzania	3.786
12	2021	Zambia	3.082
13	2022	Botswana	3.435

# Which are ten happier and ten least happy countries in year 2022?



```
SELECT DISTINCT year, country, Happiness_Score  
FROM World_happiness  
WHERE year = 2022  
ORDER BY Happiness_Score DESC  
LIMIT 10;
```

Data Output    Messages    Notifications

≡+ 📁 ⏴ 🗂️ ⏴ 🗑️ 📥 📲

	year integer	country character varying	happiness_score double precision
1	2022	Finland	7.729
2	2022	Israel	7.662
3	2022	Denmark	7.545
4	2022	Iceland	7.449
5	2022	Sweden	7.431
6	2022	Netherlands	7.39
7	2022	Norway	7.295
8	2022	Luxembourg	7.228
9	2022	Costa Rica	7.077
10	2022	Lithuania	7.038



```
SELECT DISTINCT year, country, Happiness_Score  
FROM World_happiness  
WHERE year = 2022  
ORDER BY Happiness_Score ASC  
LIMIT 10;
```

Data Output    Messages    Notifications

≡+ 📁 ⏴ 🗂️ ⏴ 🗑️ 📥 📲

	year integer	country character varying	happiness_score double precision
1	2022	Afghanistan	1.281
2	2022	Lebanon	2.352
3	2022	Sierra Leone	2.56
4	2022	Congo (Kinshasa)	3.207
5	2022	Zimbabwe	3.296
6	2022	Malawi	3.356
7	2022	Bangladesh	3.408
8	2022	Botswana	3.435
9	2022	Comoros	3.545
10	2022	Tanzania	3.616

# Data Analysis: Some Facts discovered

- Average inflation rate as per regions and development status:

```
SELECT region, ROUND(AVG(Happiness_Score)::numeric, 3) as Avg_Happiness_Score,  
       ROUND(AVG(CPI_percentage)::numeric, 3)as Avg_CPI_percentage  
FROM Happiness_development_status  
GROUP BY region  
ORDER BY Avg_Happiness_Score DESC;
```

Data Output    Messages    Notifications

	region character varying	avg_happiness_score numeric	avg_cpi_percentage numeric
1	Australian and Oceania	7.245	2.310
2	North America	6.977	2.990
3	Western Europe	6.757	2.218
4	Central America	6.458	2.936
5	South America	6.015	7.984
6	Central American	5.956	1.200
7	Middle East	5.553	5.854
8	Eastern Europe	5.529	4.440
9	Asia	5.327	4.812
10	Caribbean	5.319	3.898
11	Africa	4.555	9.114
12	Western Africa	4.204	0.850
13	AfricaTrinidad and TobagoCaribbean	3.688	1.723
14	Souther Africa	2.905	4.410



# Data Analysis: Some Facts discovered

- Average unemployment rate as per regions and development status:

```
SELECT region, ROUND(AVG(Happiness_Score)::numeric, 3) as Avg_Happiness_Score,  
       ROUND(AVG(Unemployment_Rate)::numeric, 3)as Avg_Unemployment_Rate  
FROM Happiness_development_status  
GROUP BY region  
ORDER BY Avg_Happiness_Score DESC;
```

Data Output    Messages    Notifications

	region character varying	avg_happiness_score numeric	avg_unemployment_rate numeric
1	Australian and Oceania	7.245	5.271
2	North America	6.977	5.803
3	Western Europe	6.757	7.894
4	Central America	6.458	6.309
5	South America	6.015	6.391
6	Central American	5.956	8.240
7	Middle East	5.553	10.706
8	Eastern Europe	5.529	10.960
9	Asia	5.327	4.315
10	Caribbean	5.319	6.557
11	Africa	4.555	8.570
12	Western Africa	4.204	5.585
13	AfricaTrinidad and TobagoCaribbean	3.688	2.640
14	Souther Africa	2.905	1.570



# Data Analysis: Some Facts discovered

- Average social support rate as per regions and development status:

```
SELECT region, ROUND(AVG(Happiness_Score)::numeric, 3) as Avg_Happiness_Score,  
       ROUND(AVG(Social_support)::numeric, 3)as Avg_Social_support  
FROM Happiness_development_status  
GROUP BY region  
ORDER BY Avg_Happiness_Score DESC;
```

Data Output    Messages    Notifications

≡+

	region character varying	avg_happiness_score numeric	avg_social_support numeric
1	Australian and Oceania	7.245	0.948
2	North America	6.977	0.887
3	Western Europe	6.757	0.916
4	Central America	6.458	0.856
5	South America	6.015	0.852
6	Central American	5.956	0.757
7	Middle East	5.553	0.757
8	Eastern Europe	5.529	0.851
9	Asia	5.327	0.812
10	Caribbean	5.319	0.873
11	Africa	4.555	0.759
12	Western Africa	4.204	0.704
13	AfricaTrinidad and TobagoCaribbean	3.688	0.430
14	Souther Africa	2.905	0.565



# Data Analysis: Some Facts discovered

- Average Healthy life expectancy at birth rate as per regions and development status:

```
SELECT region, ROUND(AVG(Happiness_Score)::numeric, 3) as Avg_Happiness_Score,  
       ROUND(AVG(Healthy_life_expectancy_at_birth)::numeric, 3)as Avg_Healthy_life_expectancy_at_birth  
FROM Happiness_development_status  
GROUP BY region  
ORDER BY Avg_Happiness_Score DESC;
```

Data Output    Messages    Notifications

---

	region character varying	avg_happiness_score numeric	avg_healthy_life_expectancy_at_birth numeric
1	Australian and Oceania	7.245	70.354
2	North America	6.977	67.710
3	Western Europe	6.757	70.319
4	Central America	6.458	67.088
5	South America	6.015	65.864
6	Central American	5.956	65.000
7	Middle East	5.553	67.948
8	Eastern Europe	5.529	66.715
9	Asia	5.327	64.374
10	Caribbean	5.319	63.850
11	Africa	4.555	56.674
12	Western Africa	4.204	53.743
13	AfricaTrinidad and TobagoCaribbean	3.688	53.947
14	Souther Africa	2.905	53.380



# Data Analysis: Some Facts discovered

- Average GDP rate as per regions and development status:

```
SELECT region, ROUND(AVG(Happiness_Score)::numeric, 3) as Avg_Happiness_Score,  
       ROUND(AVG(Log_GDP_per_capita)::numeric, 3)as Avg_Log_GDP_per_capita  
FROM Happiness_development_status  
GROUP BY region  
ORDER BY Avg_Happiness_Score DESC;
```

Data Output    Messages    Notifications

	region character varying	avg_happiness_score numeric	avg_log_gdp_per_capita numeric
1	Australian and Oceania	7.245	10.699
2	North America	6.977	10.538
3	Western Europe	6.757	10.718
4	Central America	6.458	9.503
5	South America	6.015	9.411
6	Central American	5.956	9.135
7	Middle East	5.553	9.824
8	Eastern Europe	5.529	9.993
9	Asia	5.327	9.357
10	Caribbean	5.319	9.636
11	Africa	4.555	8.419
12	Western Africa	4.204	7.574
13	AfricaTrinidad and TobagoCaribbean	3.688	7.511
14	Souther Africa	2.905	6.723



# Data Analysis: Some Facts discovered

- Average Perceptions of corruption as per regions and

```
SELECT region, ROUND(AVG(Happiness_Score)::numeric, 3) as Avg_Happiness_Score,  
       ROUND(AVG(Perceptions_of_corruption)::numeric, 3)as Avg_Perceptions_of_corruption  
FROM Happiness_development_status  
GROUP BY region  
ORDER BY Avg_Happiness_Score DESC;
```

Data Output    Messages    Notifications

The screenshot shows a database query results interface. At the top, there are tabs for 'Data Output', 'Messages', and 'Notifications'. Below the tabs is a toolbar with various icons for file operations like new, open, save, and refresh. The main area displays a table with two columns: 'region' and 'avg\_happiness\_score' (with a secondary column 'avg\_perceptions\_of\_corruption'). The 'avg\_happiness\_score' column is highlighted with a red border. The data rows are numbered from 1 to 14, corresponding to different regions.

	region character varying	avg_happiness_score numeric	avg_perceptions_of_corruption numeric
1	Australian and Oceania	7.245	0.342
2	North America	6.977	0.617
3	Western Europe	6.757	0.565
4	Central America	6.458	0.770
5	South America	6.015	0.792
6	Central American	5.956	0.782
7	Middle East	5.553	0.804
8	Eastern Europe	5.529	0.849
9	Asia	5.327	0.759
10	Caribbean	5.319	0.740
11	Africa	4.555	0.790
12	Western Africa	4.204	0.779
13	AfricaTrinidad and TobagoCaribbean	3.688	0.764
14	Souther Africa	2.905	0.808



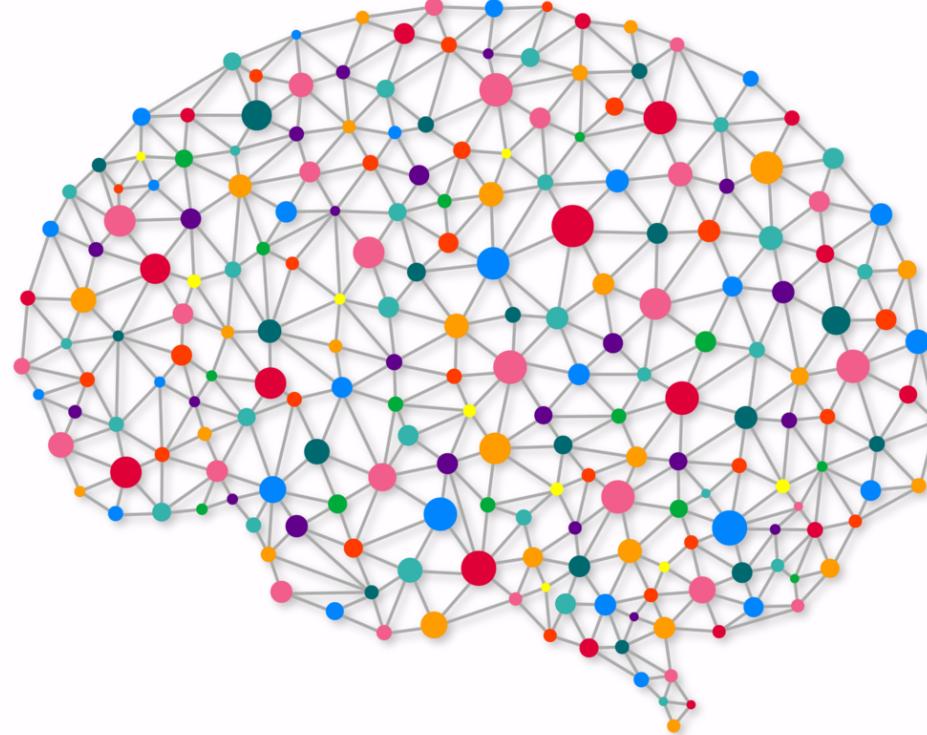
# Canada Performance in single database

```
SELECT *
FROM Happiness_development_status
WHERE country = 'Canada'
ORDER BY year;
```

Data Output    Messages    Graph Visualiser    Notifications

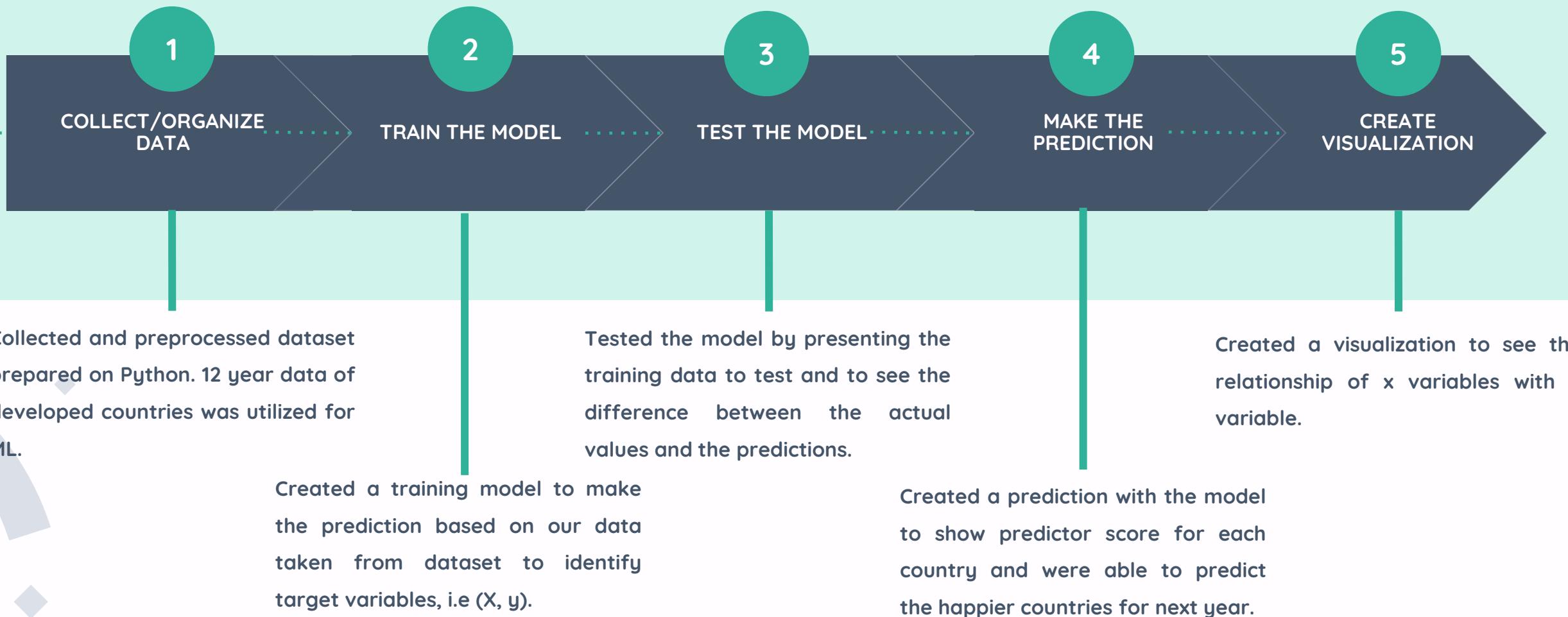
	country	year	happiness_sc	log_gdp_per_c	social_support	healthy_life_expectancy	freedom_to_n	generosity	perceptions_of_crime	cpi_percent_change	unemployment_rate	region	development_status
1	Canada	2010	7.65	10.711	0.954	70.8	0.934	0.227	0.413	1.78	8.18	North America	Developed
2	Canada	2011	7.426	10.733	0.922	70.86	0.951	0.249	0.433	2.91	7.64	North America	Developed
3	Canada	2012	7.415	10.739	0.948	70.92	0.918	0.286	0.466	1.52	7.39	North America	Developed
4	Canada	2013	7.594	10.752	0.936	70.98	0.916	0.312	0.406	0.94	7.14	North America	Developed
5	Canada	2014	7.304	10.77	0.918	71.04	0.939	0.266	0.442	1.91	7.02	North America	Developed
6	Canada	2015	7.413	10.769	0.939	71.1	0.931	0.249	0.427	1.13	6.95	North America	Developed
7	Canada	2016	7.245	10.768	0.924	71.15	0.912	0.207	0.385	1.43	7.04	North America	Developed
8	Canada	2017	7.415	10.786	0.934	71.2	0.945	0.159	0.362	1.6	6.43	North America	Developed
9	Canada	2018	7.175	10.799	0.923	71.25	0.946	0.102	0.372	2.27	5.84	North America	Developed
10	Canada	2019	7.109	10.803	0.925	71.3	0.912	0.107	0.436	1.95	5.69	North America	Developed
11	Canada	2020	7.025	10.738	0.931	71.35	0.887	0.045	0.434	0.72	9.66	North America	Developed
12	Canada	2021	6.939	10.777	0.926	71.4	0.898	0.191	0.384	3.4	7.53	North America	Developed
13	Canada	2022	6.918	10.803	0.929	71.45	0.838	0.222	0.442	6.8	5.28	North America	Developed

# 03



## Machine Learning

# Prediction Process



# Machine Learning- Linear Regression

```
# Define your independent variables (features) and the dependent variable (happiness scores)
X = df_happiness_transformed.drop(['Happiness Score'], axis=1).values

y = df_happiness_transformed['Happiness Score'].values

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=0)

# Create and fit the linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

#Prediction of test set
y_pred = model.predict(X_test)
#Predicted values
print("Prediction for test set: {}".format(y_pred))
```

#Actual value and the predicted value

```
model_diff = pd.DataFrame({'Actual value': y_test, 'Predicted value': y_pred, 'Difference': model_diff['Difference'][0:20]})
```

[29]

...

	Actual value	Predicted value	Difference
0	0.532698	0.813454	-0.280756
1	1.233773	0.947170	0.286603
2	0.677263	0.396174	0.281089
3	0.482804	0.461823	0.020981
4	-1.410611	-0.728655	-0.681956
5	1.561282	0.984675	0.576607
6	-0.631496	-0.573875	-0.057622
7	0.825665	0.915761	-0.090095
8	-0.476697	-0.393144	-0.083553
9	-0.043003	-0.128593	0.085591
10	0.481524	0.345263	0.136261
11	-0.251534	-0.198063	-0.053471
12	0.193674	0.808990	-0.615316
13	-0.959006	-1.136878	0.177872
14	-0.332132	-0.800969	0.468837
15	-0.119763	-0.011383	-0.108380
16	0.652955	0.761094	-0.108139
17	-0.648128	-0.364533	-0.283595
18	-2.816599	-2.502129	-0.314470
19	0.509670	0.386682	0.122987

# ML-Model 1

```
# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print("Mean Squared Error:", mse)
print("R-squared:", r2)
```

```
Mean Squared Error: 0.13565691586692505
R-squared: 0.7638657227464156
```

```
# Display the coefficients of the model to identify influential variables
coefficients = pd.DataFrame({'Feature': X.columns, 'Coefficient': model.coef_})
print("Influential variables:\n", coefficients)
```

Influential variables:

	Feature	Coefficient
0	Log GDP per capita	0.097251
1	Social support	0.229895
2	Healthy life expectancy at birth	0.223856
3	Freedom to make life choices	0.095020
4	Generosity	0.193434
5	Perceptions of corruption	-0.176153
6	CPI %	-0.033693
7	Unemployment Rate	-0.114201

# ML-Model 2

```
# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print("Mean Squared Error:", mse)
print("R-squared:", r2)
```

```
Mean Squared Error: 0.1591784278233005
R-squared: 0.841337320061079
```

```
# Get the coefficients of the model
coefficients = model.coef_

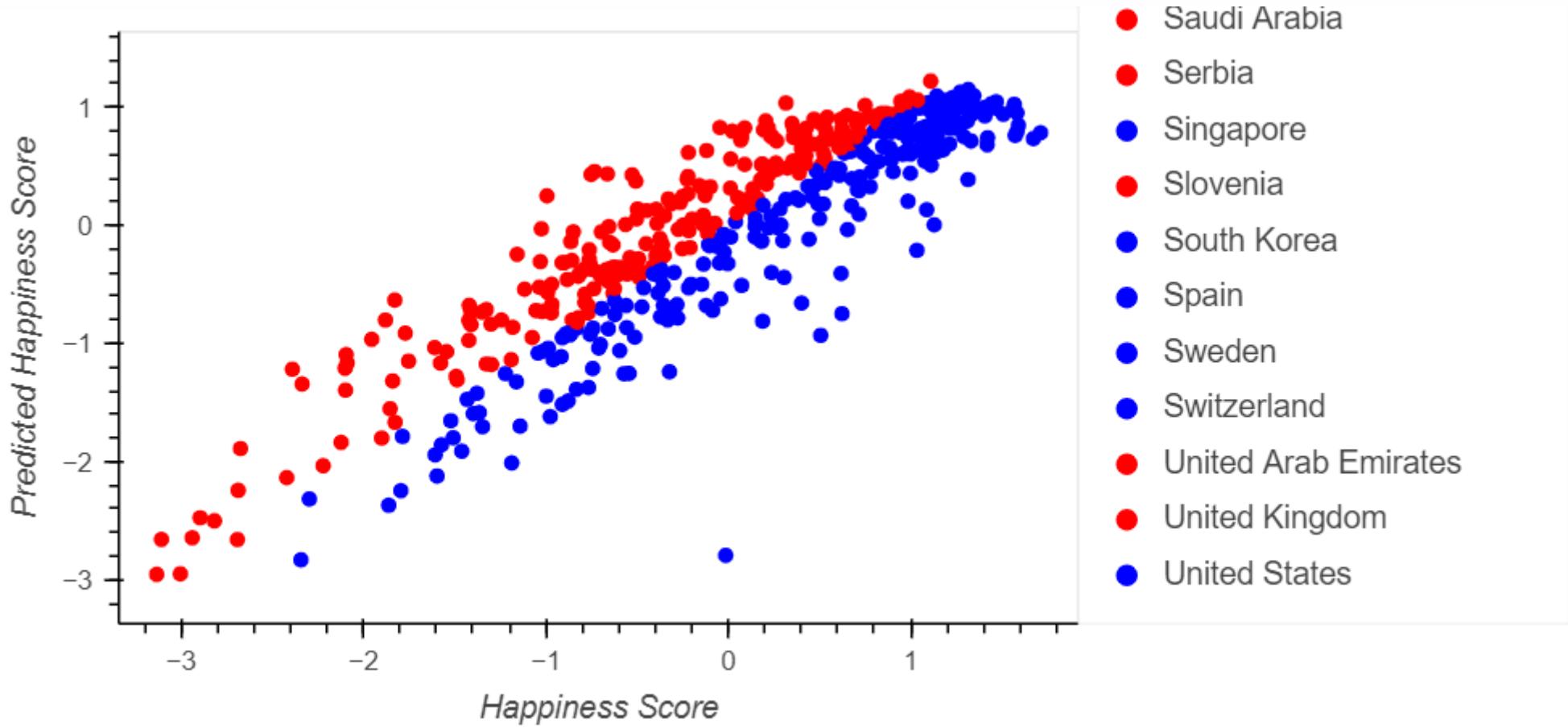
# Display the coefficients along with their corresponding feature names
feature_names = df_happiness_transformed.drop(['Happiness Score'], axis=1).columns
coefficients_df = pd.DataFrame({'Feature': feature_names, 'Coefficient': coefficients})

coefficients_df.head(8)
```

	Feature	Coefficient
0	Log GDP per capita	0.085125
1	Social support	0.279920
2	Healthy life expectancy at birth	0.216557
3	Freedom to make life choices	0.092699
4	Generosity	0.138025
5	Perceptions of corruption	-0.168597
6	CPI %	-0.095736
7	Unemployment Rate	-0.249118

# Scatter Plot

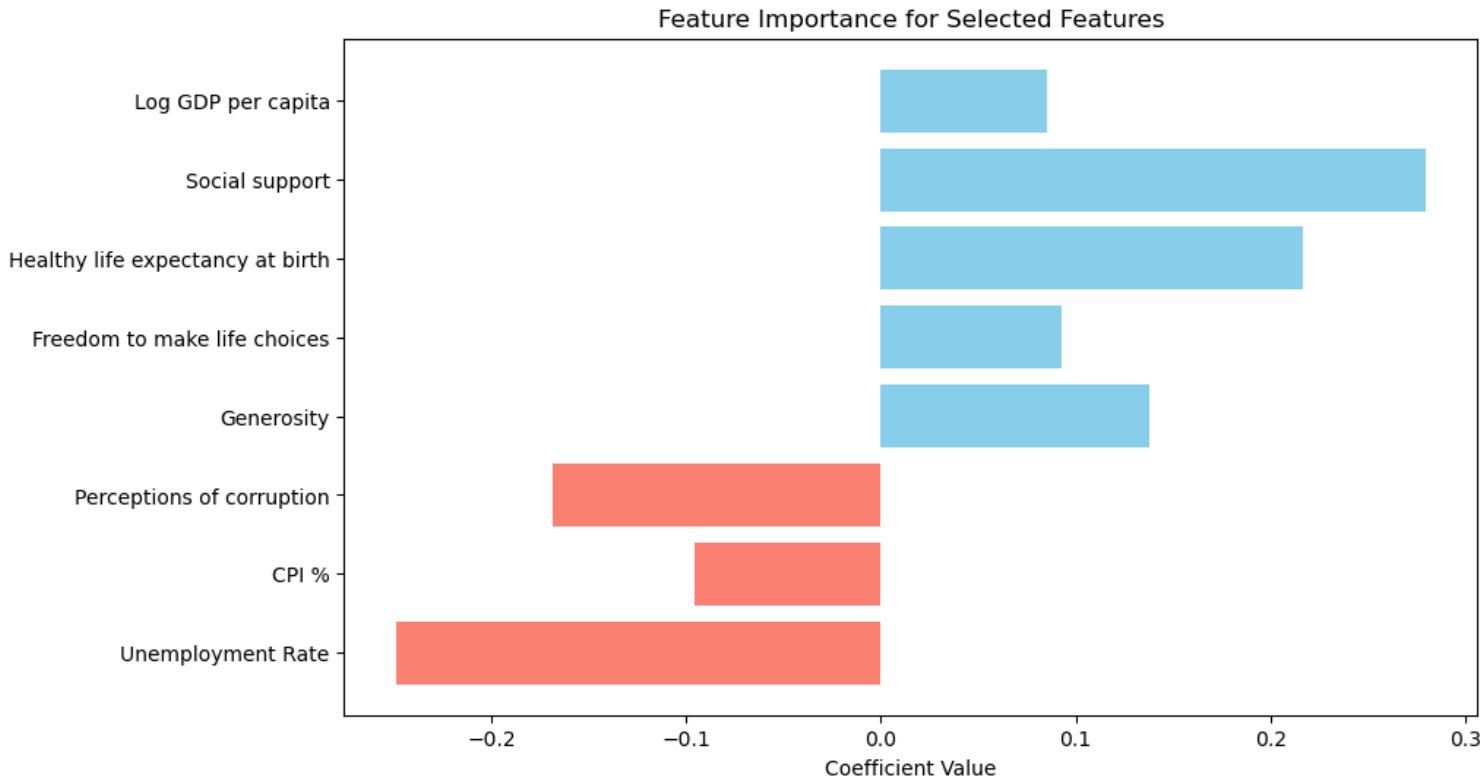
## Actual Vs Predicted score



# Predictive Equation from Machine Learning

$$\text{Happiness} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \varepsilon_Y$$

	Feature	Coefficient
0	Log GDP per capita	0.085125
1	Social support	0.279920
2	Healthy life expectancy at birth	0.216557
3	Freedom to make life choices	0.092699
4	Generosity	0.138025
5	Perceptions of corruption	-0.168597
6	CPI %	-0.095736
7	Unemployment Rate	-0.249118



# ML-Model 3 (Random Forest)

```
# Load the data  
[2] data = pd.read_csv('Developed_nations.csv')

# Encode categorical variables  
[3] data = pd.get_dummies(data, columns=['region', 'Development_Status'])

# Split the data into features (X) and the target variable (y)  
# Exclude the 'Country' column  
[4] X = data.drop(["Happiness Score", "Country"], axis=1)  
y = data["Happiness Score"]

non_numeric_columns = data.select_dtypes(exclude=['number']).columns  
print("Non-numeric columns:", non_numeric_columns)
[5] ... Non-numeric columns: Index(['Country'], dtype='object')

# Split the data into training and testing sets  
[20] X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Train the Random Forest model  
[21] model = RandomForestRegressor(n_estimators=100, random_state=42)  
model.fit(X_train, y_train)

... RandomForestRegressor  
RandomForestRegressor(random_state=42)
```

```
[22] # Evaluate the model  
mse = mean_squared_error(y_test, y_pred)  
r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error: {mse}")  
print(f"R-squared: {r2}")

... Mean Squared Error: 0.06067934176796119  
R-squared: 0.8943771320390587
```

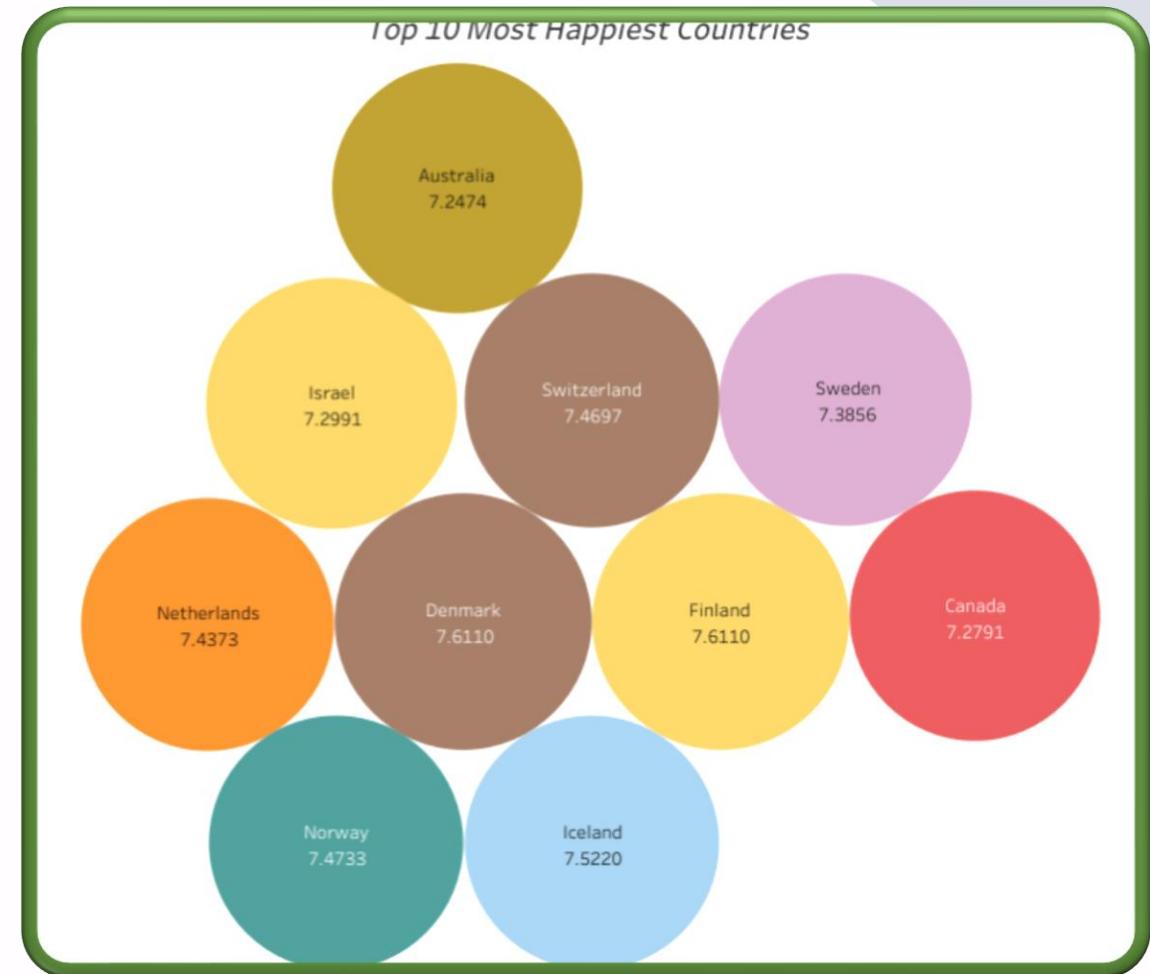
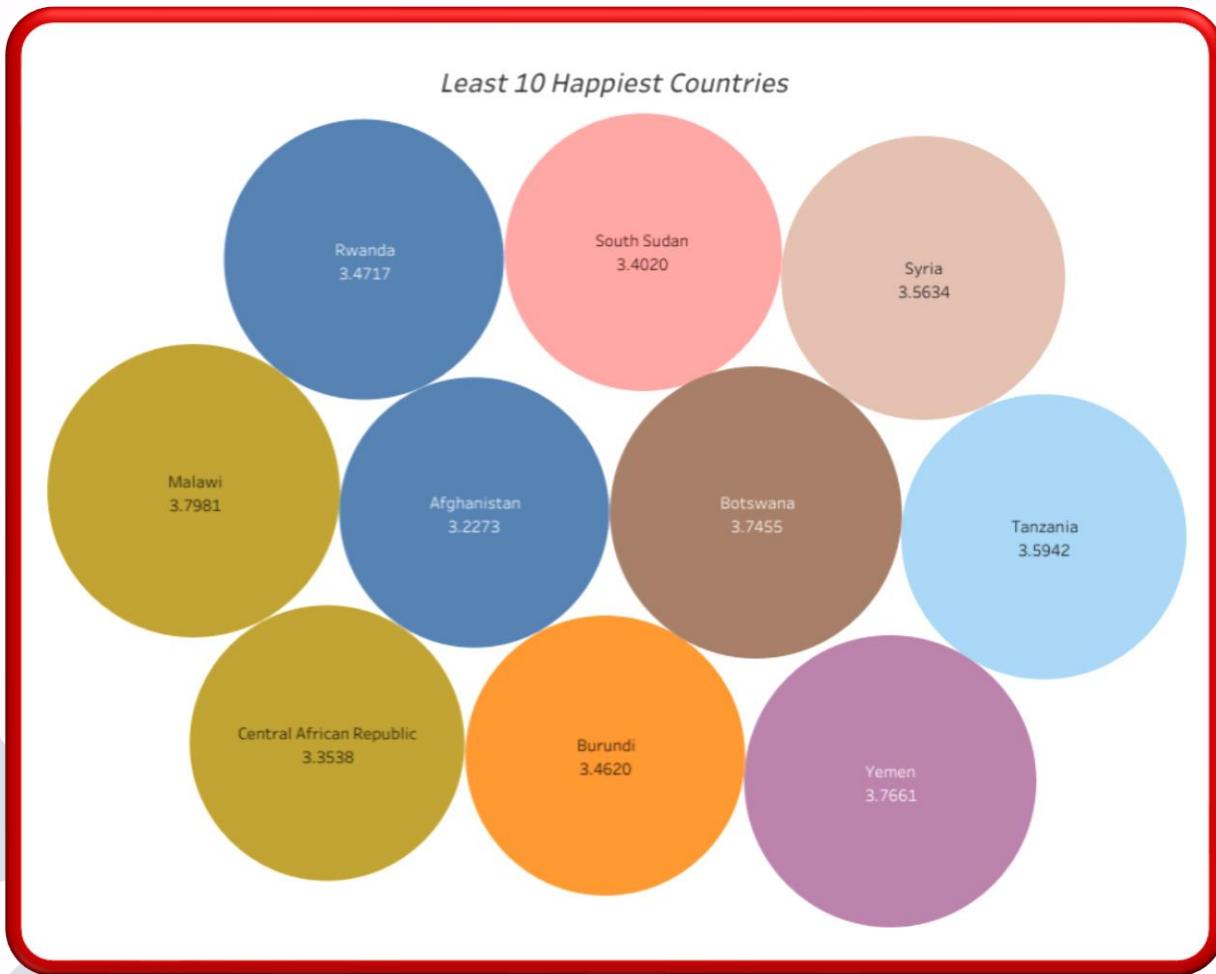
# 04

## Tableau Visuals

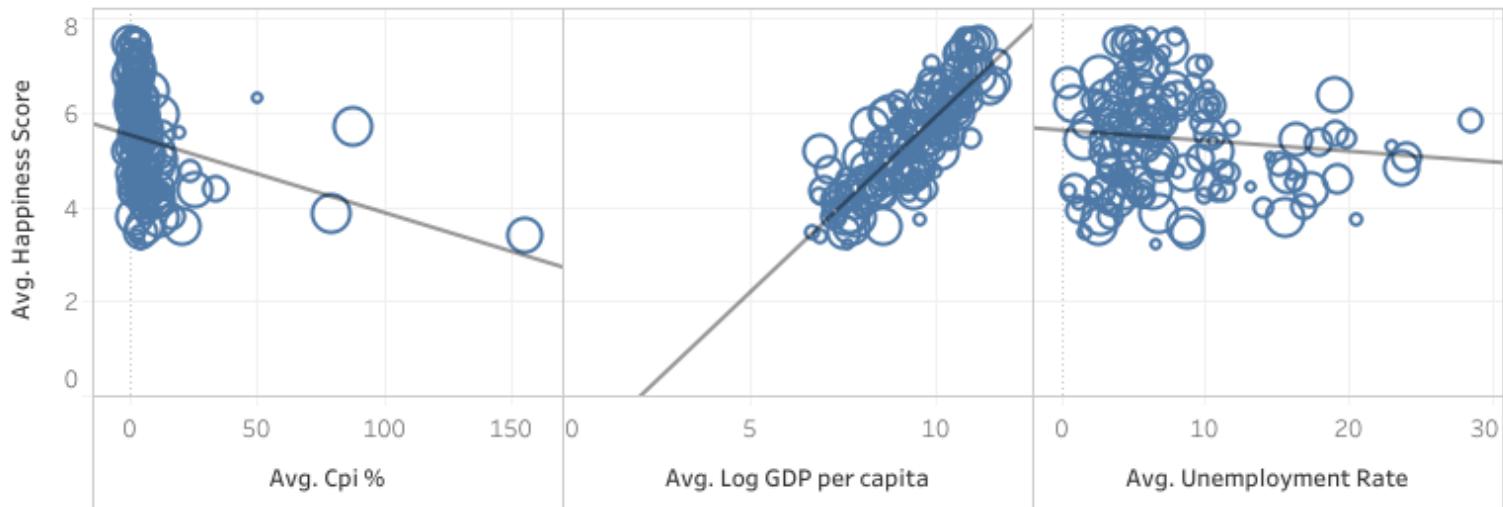


+a**b|eau**

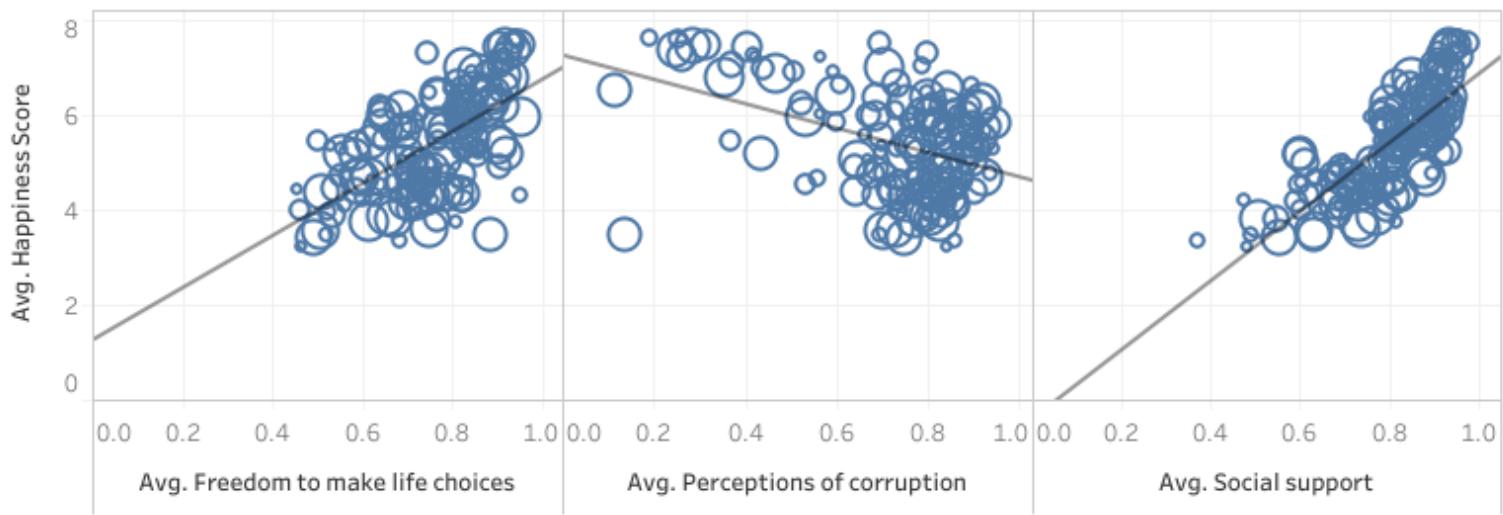
# Some Key Visuals



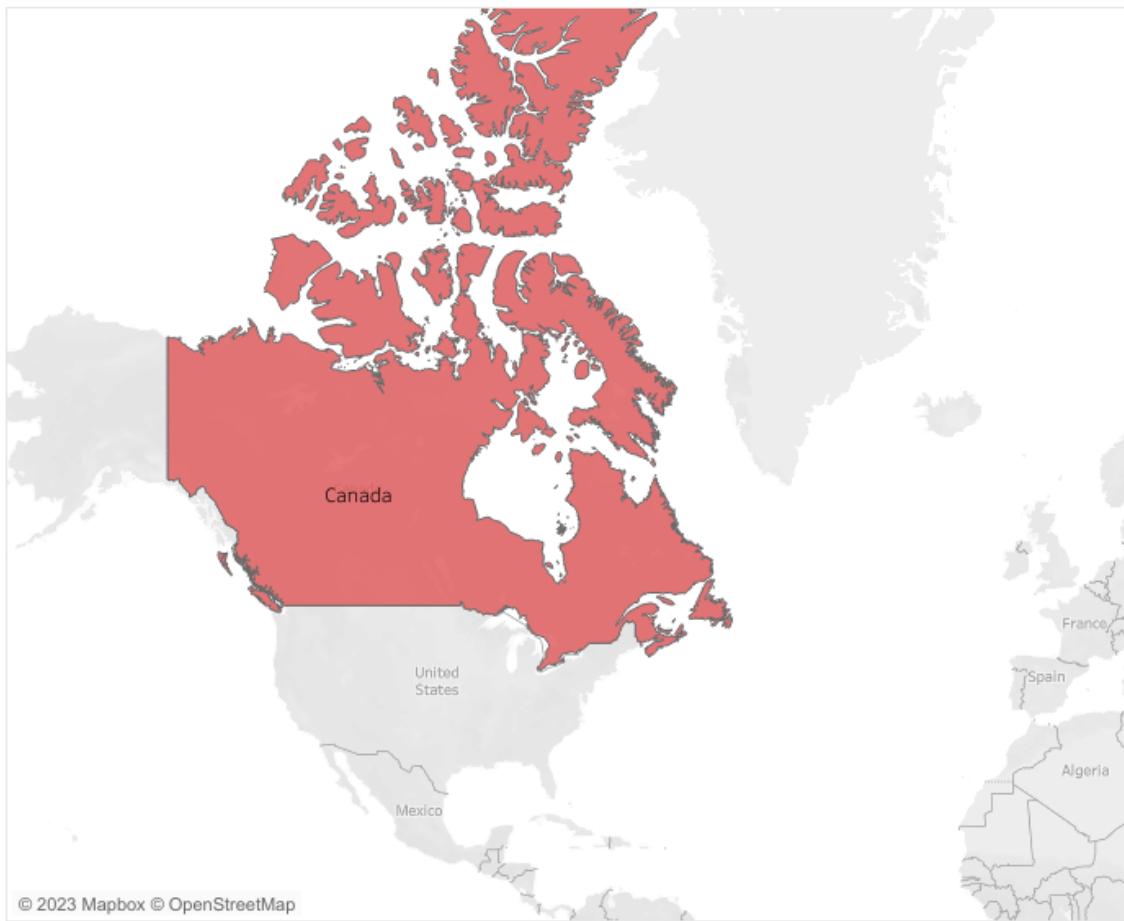
### *HappinessScore\_Vs\_Economic\_Variables*



### *Happiness Score\_Vs\_Social Variables*



*MAP OF CANADA*



*Trend Analysis for Canada (Over the Years)*



# Tabeau Dashboard

- MAIN DASHBOARD
- HAPPINESS COMPARISON DASH
- CANDASH

# This project is useful for?

- This project serves as a valuable resource for various sectors, offering insights into the well-being of populations across the world and providing opportunities for informed decision-making and policy development. Project can benefit various stakeholders and entities, including:
  - **Government and Policy Makers**
  - **Businesses and Corporations**
  - **Non-Governmental Organizations (NGOs)**
  - **Individuals and Communities**
  - **Data Scientists and Analysts**



# Key Takeaways from Questions Answered:

- 1) How does happiness score depend on social and economic factors i.e., GDP per capita, Health & Corruption, Inflation, Unemployment?
- 2) What are the similar facts for the top and bottom 10 countries?
- 3) Which region has most and the least Happy countries in the dataset?
- 4) What is the trend of happiness score for Canada in the last 12 years?

# Limitations

This project has some key limitations and potential challenges in data collection, modeling, and deployment:

## Data Limitations:

1. Data Quality
2. Data Availability
3. Missing Data

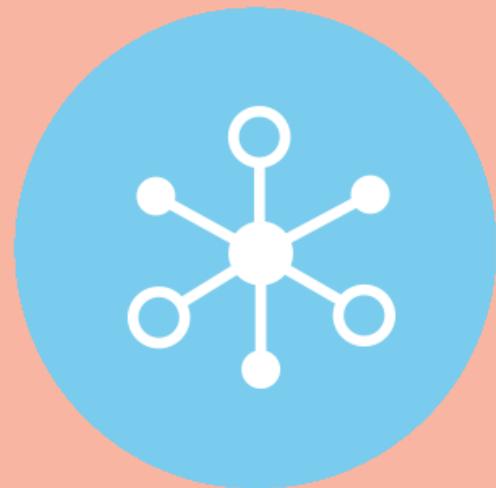
## Modeling Limitations:

1. Simplification
2. Causation vs. Correlation

## Deployment Limitations:

1. Generalization
2. Feedback Loops



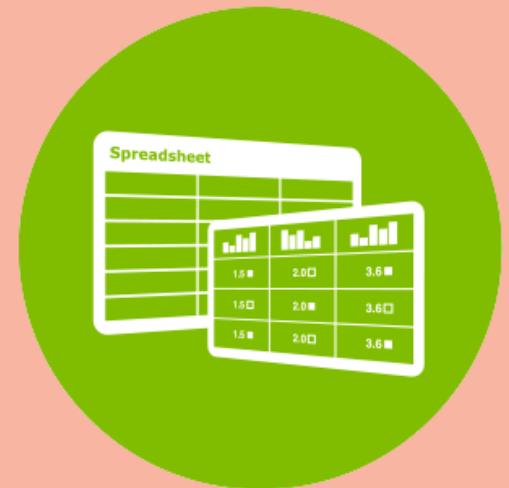


Integrate

Pandas



SQL

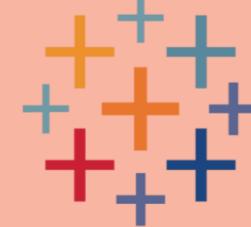


Analyze



Visualize

scikit  
learn





Let's Do It Together: Let's Make the World  
a Happier Place

Thank you!