WORLD HAPPINESS REPORT 2023

Analysis of Country Happiness Prediction Using Machine Learning


jovinkaaphelliasalva@gmail.com

THE RESULT

World Happiness Report 2023

jovinkaaphelliasalva@gmail.com

EXPLORATORY DATA ANALYSIS (EDA)

```
import os
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb

#Read CSV File
df=pd.read_csv("/content/WHR2023.csv")
```

```
import pandas as pd

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 137 entries, 0 to 136
Data columns (total 19 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   Country name                          137 non-null    object
 1   Ladder score                          137 non-null    float64
 2   Standard error of ladder score        137 non-null    float64
 3   upperwhisker                          137 non-null    float64
 4   lowerwhisker                          137 non-null    float64
 5   Logged GDP per capita                 137 non-null    float64
 6   Social support                        137 non-null    float64
 7   Healthy life expectancy               136 non-null    float64
 8   Freedom to make life choices          137 non-null    float64
 9   Generosity                            137 non-null    float64
 10  Perceptions of corruption             137 non-null    float64
 11  Ladder score in Dystopia              137 non-null    float64
 12  Explained by: Log GDP per capita      137 non-null    float64
 13  Explained by: Social support          137 non-null    float64
 14  Explained by: Healthy life expectancy 136 non-null    float64
 15  Explained by: Freedom to make life choices 137 non-null float64
 16  Explained by: Generosity              137 non-null    float64
 17  Explained by: Perceptions of corruption 137 non-null  float64
 18  Dystopia + residual                   136 non-null    float64
dtypes: float64(18), object(1)
memory usage: 20.5+ KB
```

```
[9] df.isnull().sum()
```

	0
Country name	0
Ladder score	0
Standard error of ladder score	0
upperwhisker	0
lowerwhisker	0
Logged GDP per capita	0
Social support	0
Healthy life expectancy	1
Freedom to make life choices	0
Generosity	0
Perceptions of corruption	0
Ladder score in Dystopia	0
Explained by: Log GDP per capita	0
Explained by: Social support	0
Explained by: Healthy life expectancy	1
Explained by: Freedom to make life choices	0
Explained by: Generosity	0
Explained by: Perceptions of corruption	0
Dystopia + residual	1

dtype: int64

```
[10] df = df.dropna()
     df.isnull().sum()
```

	0
Country name	0
Ladder score	0
Standard error of ladder score	0
upperwhisker	0
lowerwhisker	0
Logged GDP per capita	0
Social support	0
Healthy life expectancy	0
Freedom to make life choices	0
Generosity	0
Perceptions of corruption	0
Ladder score in Dystopia	0
Explained by: Log GDP per capita	0
Explained by: Social support	0
Explained by: Healthy life expectancy	0
Explained by: Freedom to make life choices	0
Explained by: Generosity	0
Explained by: Perceptions of corruption	0
Dystopia + residual	0

dtype: int64

EXPLORATORY DATA ANALYSIS (EDA)

```
df.loc[df.duplicated()]
```

Country name	Ladder score	Standard error of ladder score	upperwhisker	lowerwhisker	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption	Ladder score in Dystopia	Explained by: Log GDP per capita	Explained by: Social support	Explained by: Healthy life expectancy	Explained by: Freedom to make life choices	Explained by: Generosity	Explained by: Perceptions of corruption	Dystopia + residual

```
[67] df.describe()
```

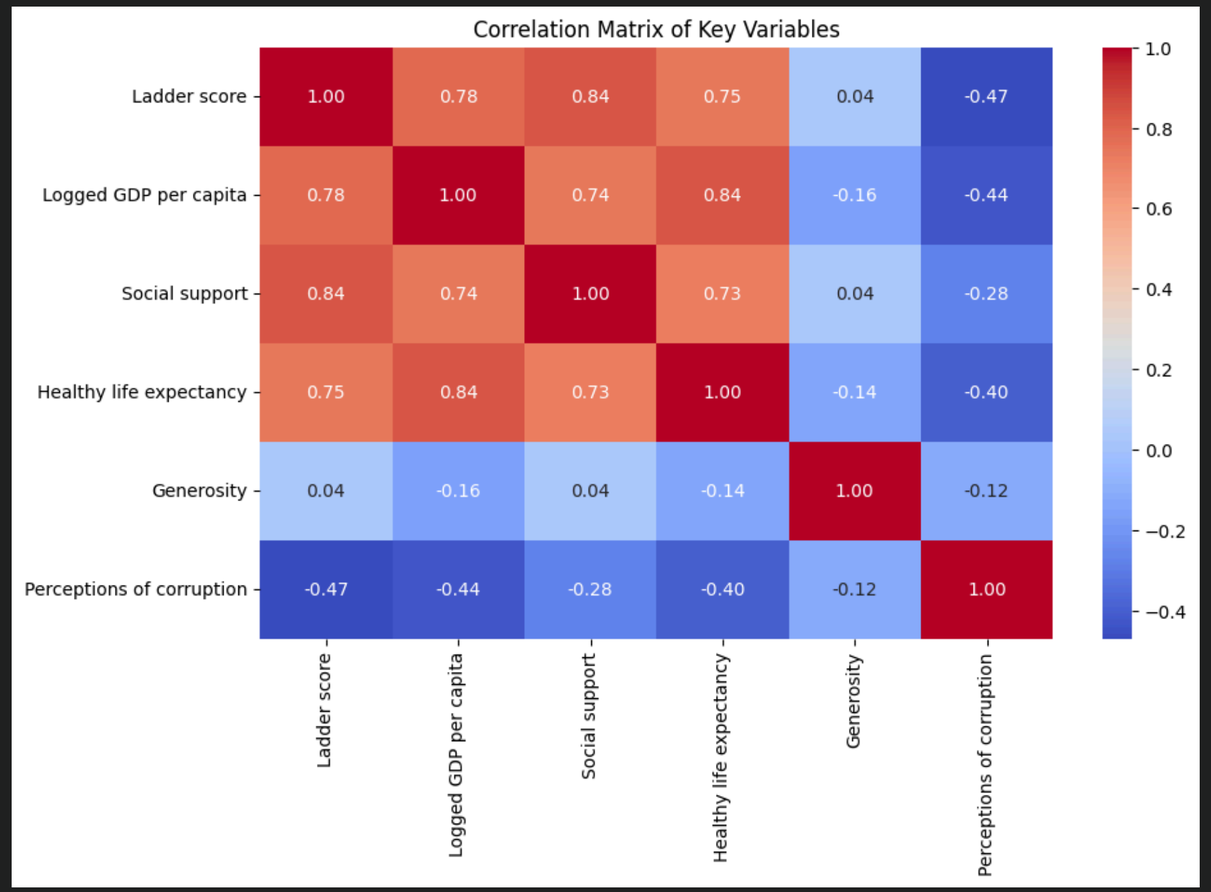
	Ladder score	Standard error of ladder score	upperwhisker	lowerwhisker	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption	Ladder score in Dystopia	Explained by: Log GDP per capita	Explained by: Social support	Explained by: Healthy life expectancy	Explained by: Freedom to make life choices	Explained by: Generosity	Explained by: Perceptions of corruption	Dystopia + residual
count	136.000000	136.000000	136.000000	136.000000	136.000000	136.000000	136.000000	136.000000	136.000000	136.000000	1.360000e+02	136.000000	136.000000	136.000000	136.000000	136.000000	136.000000	136.000000
mean	5.544441	0.064515	5.670772	5.418015	9.455191	0.798632	64.967632	0.788081	0.023566	0.724588	1.778000e+00	1.408919	1.155088	0.366176	0.540912	0.149088	0.146478	1.777838
std	1.142841	0.022996	1.120442	1.166522	1.210107	0.129597	5.750390	0.112498	0.141604	0.177353	2.897251e-15	0.433969	0.327263	0.156691	0.149671	0.075993	0.127009	0.504390
min	1.859000	0.029000	1.923000	1.795000	5.527000	0.341000	51.530000	0.382000	-0.254000	0.146000	1.778000e+00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-0.110000
25%	4.702500	0.046750	4.939750	4.492250	8.587250	0.721000	60.648500	0.726250	-0.071000	0.666000	1.778000e+00	1.097750	0.959750	0.248500	0.458750	0.098500	0.059750	1.555250
50%	5.693500	0.060000	5.824000	5.550500	9.574500	0.826500	65.837500	0.801000	0.002000	0.772500	1.778000e+00	1.451500	1.225500	0.389500	0.557500	0.137500	0.112000	1.848500
75%	6.342500	0.076250	6.452000	6.244750	10.540250	0.896000	69.412500	0.874750	0.117500	0.846000	1.778000e+00	1.798000	1.401250	0.487500	0.656750	0.199250	0.188250	2.078750
max	7.804000	0.147000	7.875000	7.733000	11.660000	0.983000	77.280000	0.961000	0.531000	0.929000	1.778000e+00	2.200000	1.620000	0.702000	0.772000	0.422000	0.561000	2.955000

```
df.head(10)
```

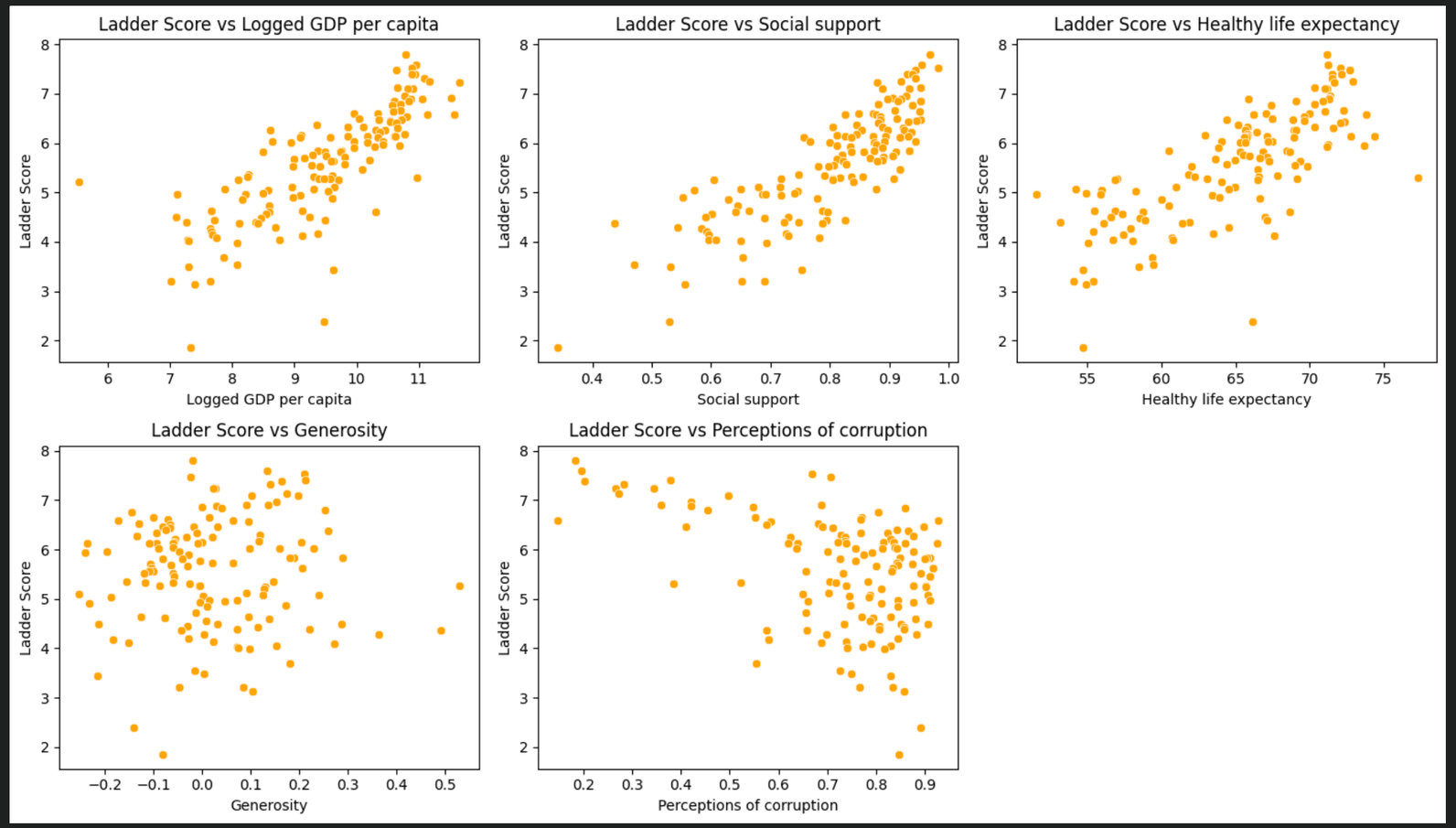
	Country name	Ladder score	Standard error of ladder score	upperwhisker	lowerwhisker	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	...	Ladder score in Dystopia	Explained by: Log GDP per capita	Explained by: Social support	Explained by: Healthy life expectancy	Explained by: Freedom to make life choices	Explained by: Generosity	Explained by: Perceptions of corruption	Dystopia + residual	Latitude	Longitude
0	Finland	7.804	0.036	7.875	7.733	10.792	0.969	71.150	0.961	-0.019	...	1.778	1.888	1.585	0.535	0.772	0.126	0.535	2.363	63.246778	25.920916
1	Denmark	7.586	0.041	7.667	7.506	10.962	0.954	71.250	0.934	0.134	...	1.778	1.949	1.548	0.537	0.734	0.208	0.525	2.084	55.670249	10.333328
2	Iceland	7.530	0.049	7.625	7.434	10.896	0.983	72.050	0.936	0.211	...	1.778	1.926	1.620	0.559	0.738	0.250	0.187	2.250	64.984182	-18.105901
3	Israel	7.473	0.032	7.535	7.411	10.639	0.943	72.697	0.809	-0.023	...	1.778	1.833	1.521	0.577	0.569	0.124	0.158	2.691	31.394800	34.633583
4	Netherlands	7.403	0.029	7.460	7.346	10.942	0.930	71.550	0.887	0.213	...	1.778	1.942	1.488	0.545	0.672	0.251	0.394	2.110	52.243498	5.634323
5	Sweden	7.395	0.037	7.468	7.322	10.883	0.939	72.150	0.948	0.165	...	1.778	1.921	1.510	0.562	0.754	0.225	0.520	1.903	59.674971	14.520858
6	Norway	7.315	0.044	7.402	7.229	11.088	0.943	71.500	0.947	0.141	...	1.778	1.994	1.521	0.544	0.752	0.212	0.463	1.829	64.573154	11.528036
7	Switzerland	7.240	0.043	7.324	7.156	11.164	0.920	72.900	0.891	0.027	...	1.778	2.022	1.463	0.582	0.678	0.151	0.475	1.870	46.798562	8.231974
8	Luxembourg	7.228	0.069	7.363	7.093	11.660	0.879	71.675	0.915	0.024	...	1.778	2.200	1.357	0.549	0.710	0.149	0.418	1.845	49.611277	6.129799
9	New Zealand	7.123	0.038	7.198	7.048	10.662	0.952	70.350	0.887	0.175	...	1.778	1.842	1.544	0.513	0.672	0.230	0.471	1.852	-41.500083	172.834408

10 rows × 21 columns

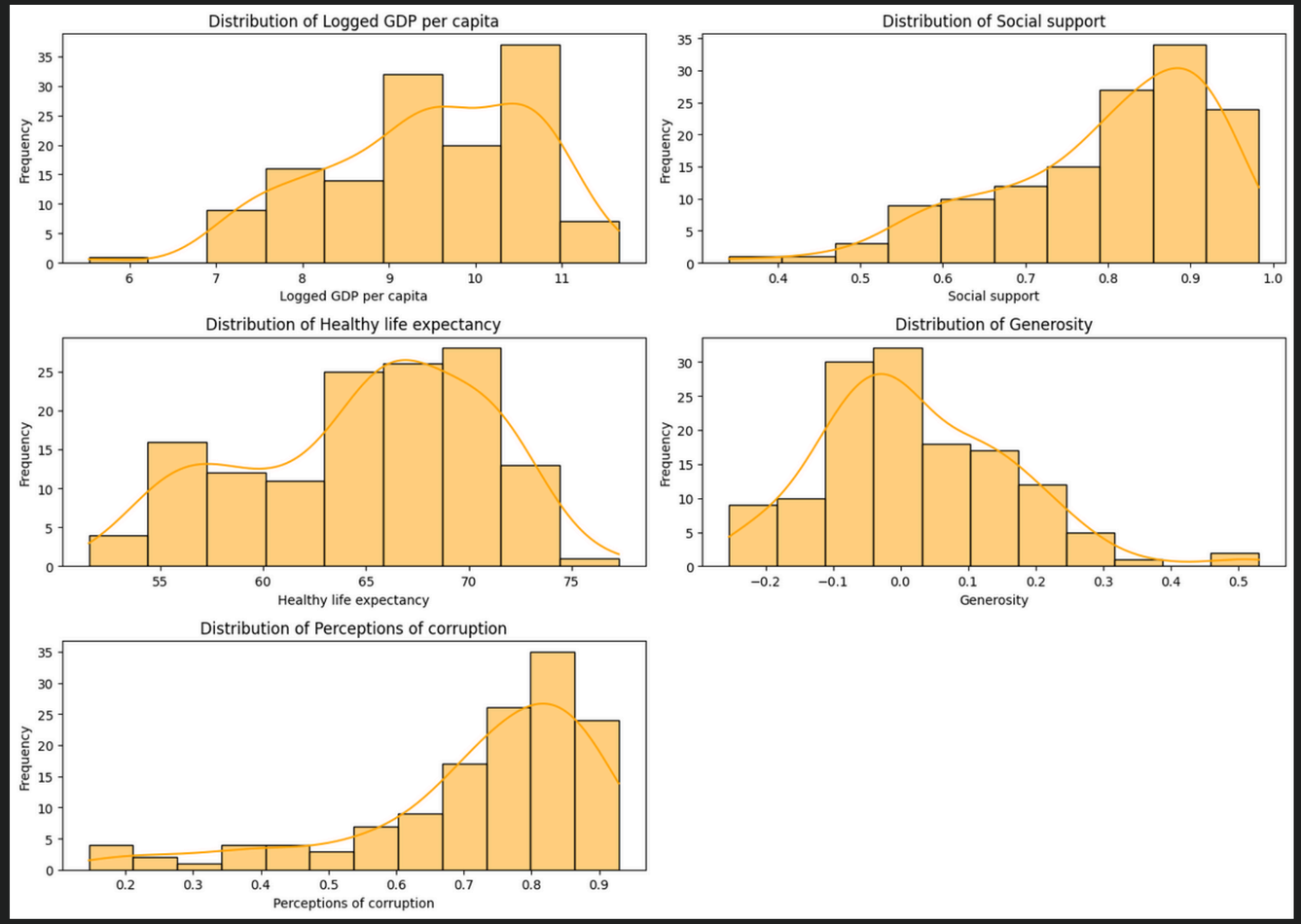
DATA VISUALIZATION



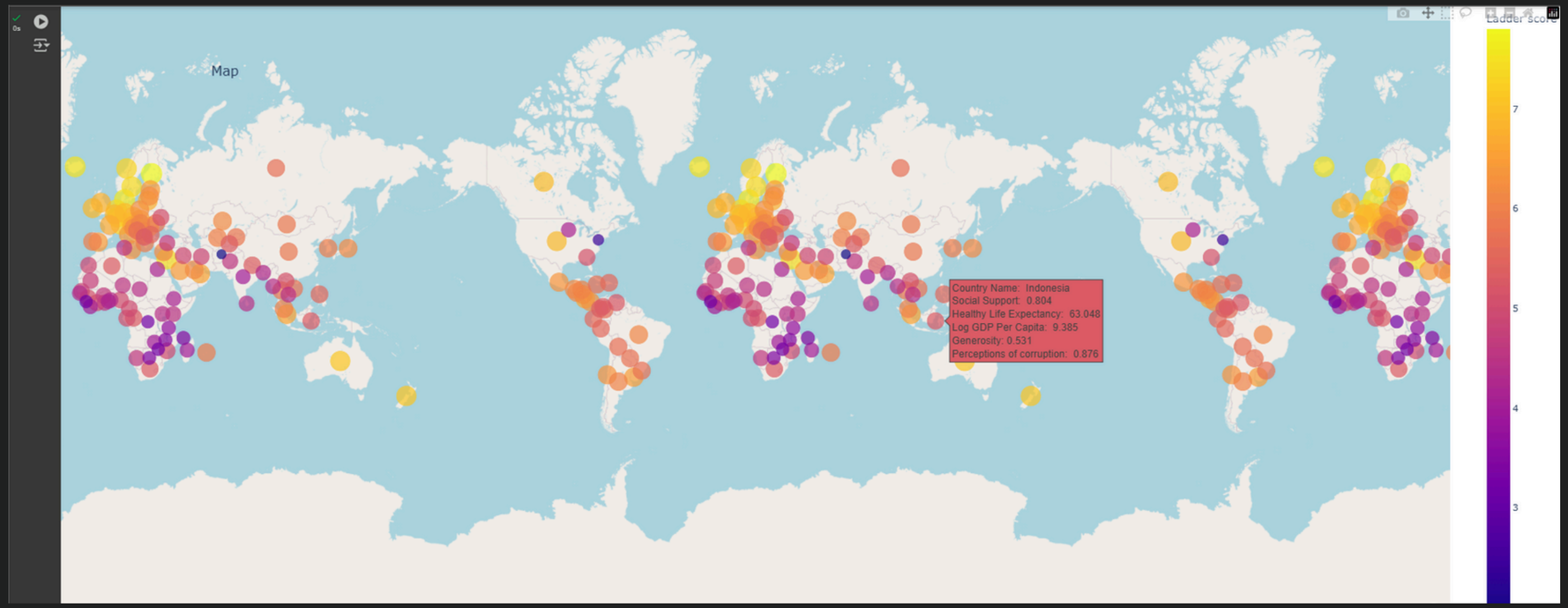
DATA VISUALIZATION



DATA VISUALIZATION



DATA VISUALIZATION

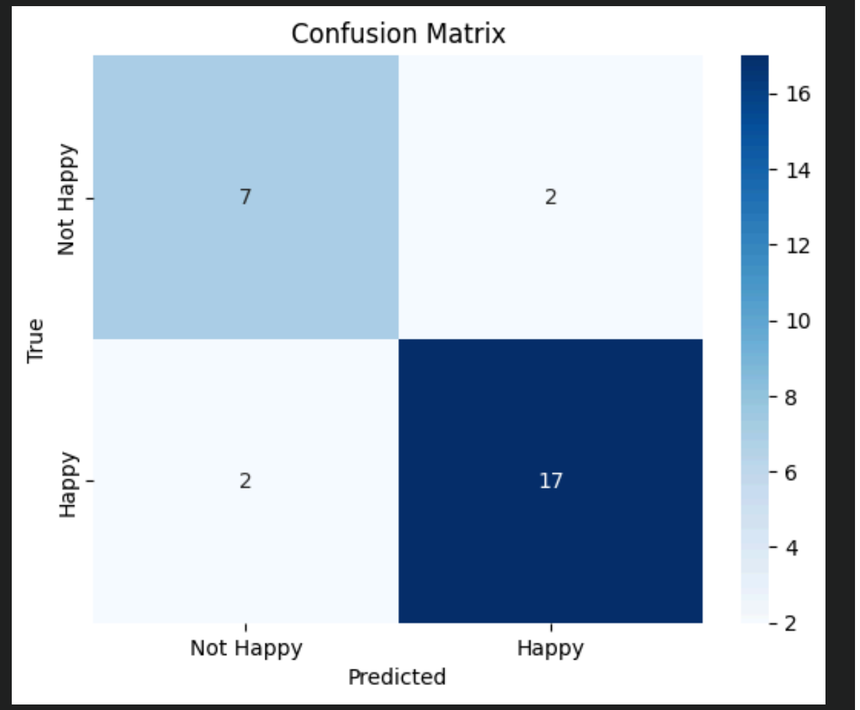


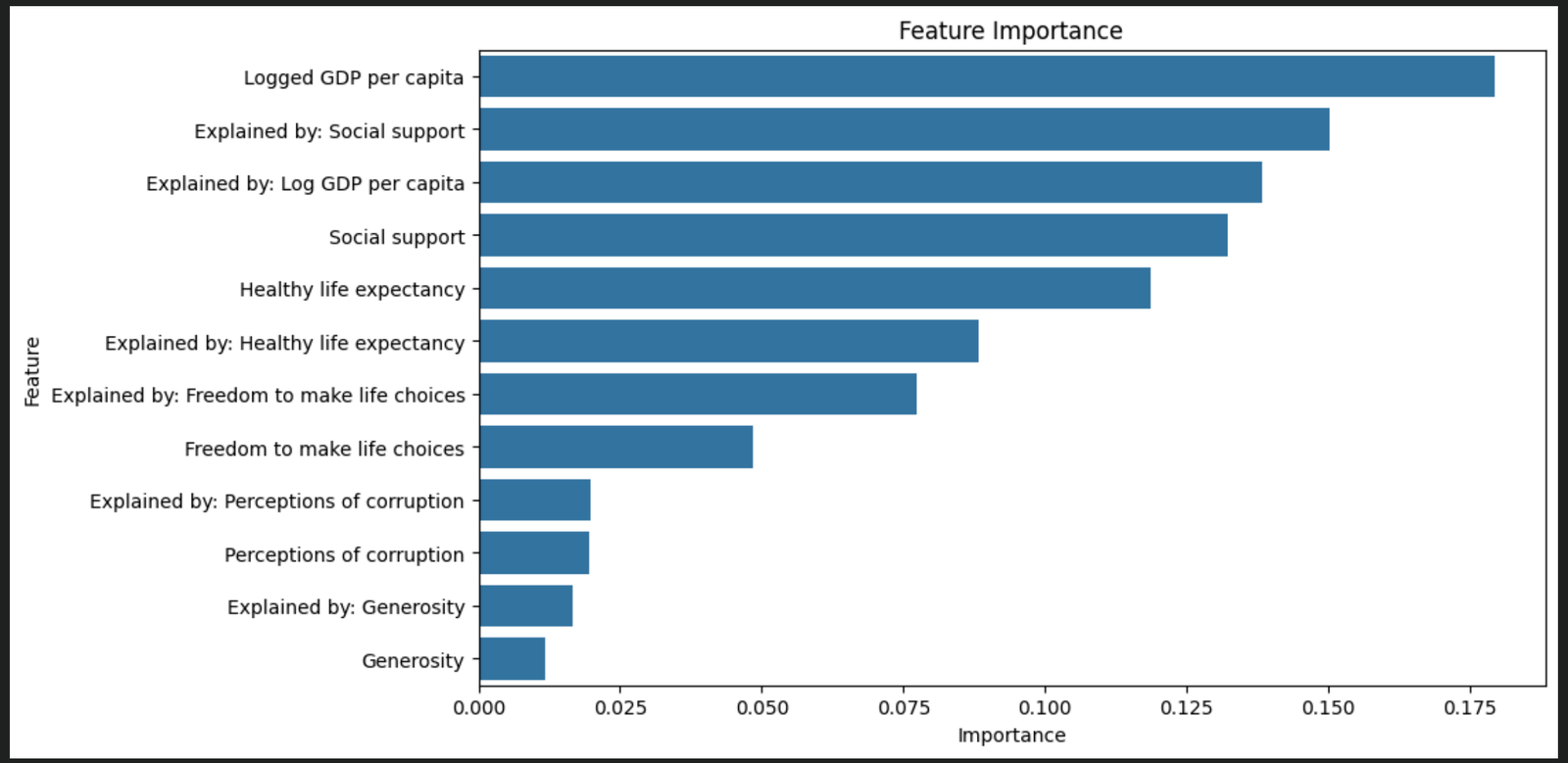
MODEL MACHINE LEARNING (Random Forest Classifier)

```python
# Menentukan target berdasarkan Ladder score median
median_ladder_score = data['Ladder score'].median()
data['Happiness_Label'] = np.where(data['Ladder score'] >= median_ladder_score,
                                   1, 0)# 1 untuk bahagia, 0 untuk tidak bahagia

# Drop kolom target dari fitur
X = data.drop(['Ladder score', 'Happiness_Label'], axis=1)
y = data['Happiness_Label']

# Train-test split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
                                    random_state=42)

# Model training
model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)

# Prediction
y_pred = model.predict(X_test)

# Evaluation
print("\nAccuracy Score:", accuracy_score(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))
```

```
Accuracy Score: 0.8571428571428571

Classification Report:
              precision    recall  f1-score   support

           0       0.78      0.78      0.78         9
           1       0.89      0.89      0.89        19

    accuracy                           0.86        28
   macro avg       0.84      0.84      0.84        28
weighted avg       0.86      0.86      0.86        28
```

MODEL MACHINE LEARNING (Random Forest Classifier)





THE CONCLUSION

After analyzing the dataset from the World Happiness Report 2023, the conclusions are as follows:

1. Logged GDP per capita: This is the most significant factor in predicting happiness, indicating that economic well-being is a key element.
2. Social support: Social support from communities or societies has a substantial impact on a country's happiness.
3. Healthy life expectancy: As the third most important factor, it highlights that a healthy life expectancy is a significant indicator of happiness.

These findings provide insights that a country's happiness is heavily influenced by a combination of economic, social, and health factors. This data can serve as a foundation for developing better public policies.

THANK YOU

Read More