National University
of computer and emerging sciences

Project Report:

World Happiness Report Analysis

Submitted by:

Ayesha Ahmed Baig 20K-0172

Kanza Batool 20K-0423

Tahreem Fatima 20K-0483

Section:

B

Department:

Computer Science

Submitted To:

Dr. Muhammad Nouman Durrani

FAST NUCES, Karachi

Department of Computer Science

Introduction

The World Happiness Report is a comprehensive dataset that evaluates the well-being and happiness of citizens across different countries. This analysis aims to derive meaningful insights from the World Happiness Report datasets from 2015 to 2023. The primary goals include identifying the most influential features affecting a country's happiness and creating a model to recommend improvements for maximum impact.

Objectives

Exploratory Data Analysis (EDA): Understand the structure and characteristics of the World Happiness datasets.
Feature Analysis: Identify the most important features influencing a country's happiness.
Model Building: Develop a model to predict healthy life expectancy based on selected features.
Visualisation: Generate insightful visualisations to support analysis.
Top 10 Happiest Countries: Visualise and compare the top 10 happiest countries for each year.
Model Evaluation: Assess and compare the performance of AdaBoost and Random Forest models.
Predictions: Use the trained model for random predictions and insights.

Methodology

1. Data Loading and Initial Exploration

- **Libraries:** Numpy, Pandas, Matplotlib, and Seaborn were imported for data analysis and visualisation.
- **Data Loading:** Datasets from 2015 to 2023 were loaded into separate dataframes.
- **Data Cleaning:** Null values were handled, and the dataset was cleaned.
- **Column Analysis:** Main and additional columns were described, and histograms were plotted for main columns.
- **Correlation Analysis**: A heatmap was used to visualise the correlation between main columns.
- **Feature Analysis:** Scatter plots were generated for each main feature against the "Ladder score".

2. Data Preparation

- **Combined Dataset:** Datasets from 2020 to 2023 were merged into a single dataset.
- **Data Refactoring:** Columns were renamed for consistency.
- **Visual Representation:** Scatter plots were created for features against the "Ladder score".

- **Top 10 Happiest Countries:** Bar plots were generated to visualise the top 10 happiest countries for each year.

3. Model Building for Predicting Healthy Life Expectancy

- **Selecting Relevant Features:** Features (GDP per capita, Perception of corruption, Social Support, Freedom, Corruption, Dystopia + Residual, Generosity) for predicting healthy life expectancy were chosen.
- **Columns Refactoring:** Columns were renamed for consistency.
- **Conclusion:** Data from 2020-2023 was selected for model building.
- **Model Training and Evaluation:** AdaBoost and Random Forest models were trained and evaluated.
- **Predictions**: Random predictions were made using the trained AdaBoost model.

Results and Analysis

1. Exploratory Data Analysis (EDA)

- **Insights:** The dataset includes various factors influencing happiness, such as GDP, social support, and freedom.
- **Recommendation:** Continue exploring relationships between features to understand their impact on happiness.

2. Feature Analysis

- **Insights:** Social support appears to be the most important feature influencing happiness, while generosity has less impact.
- **Recommendation**: Countries should focus on enhancing social support for greater happiness.

3. Top 10 Happiest Countries

- **Insights:** The top 10 happiest countries vary each year, with consistent performers and occasional changes.
- **Recommendation:** Analyse factors contributing to happiness in top countries for potential implementation elsewhere.

4. Model Building and Evaluation

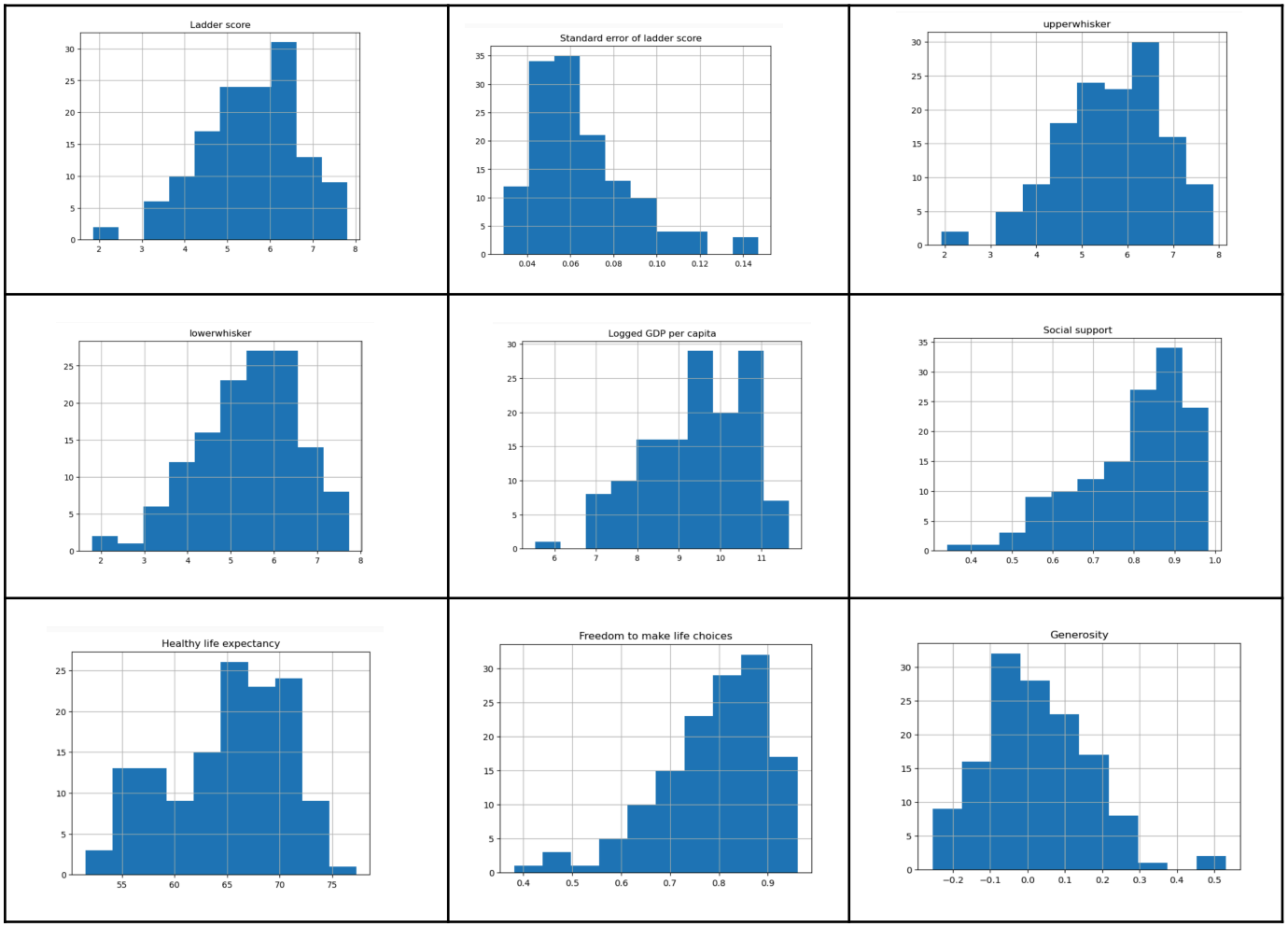
- **Insights:** AdaBoost outperformed Random Forest in predicting healthy life expectancy.
- **Recommendation:** Consider AdaBoost for future predictions on similar datasets.

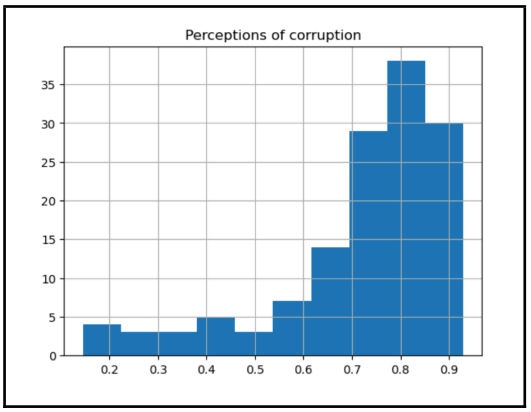
5. Predictions

- **Insights:** Random predictions provide a glimpse into how the model interprets features for life expectancy.
- **Recommendation:** Use the model cautiously for practical applications, considering the inherent uncertainties.

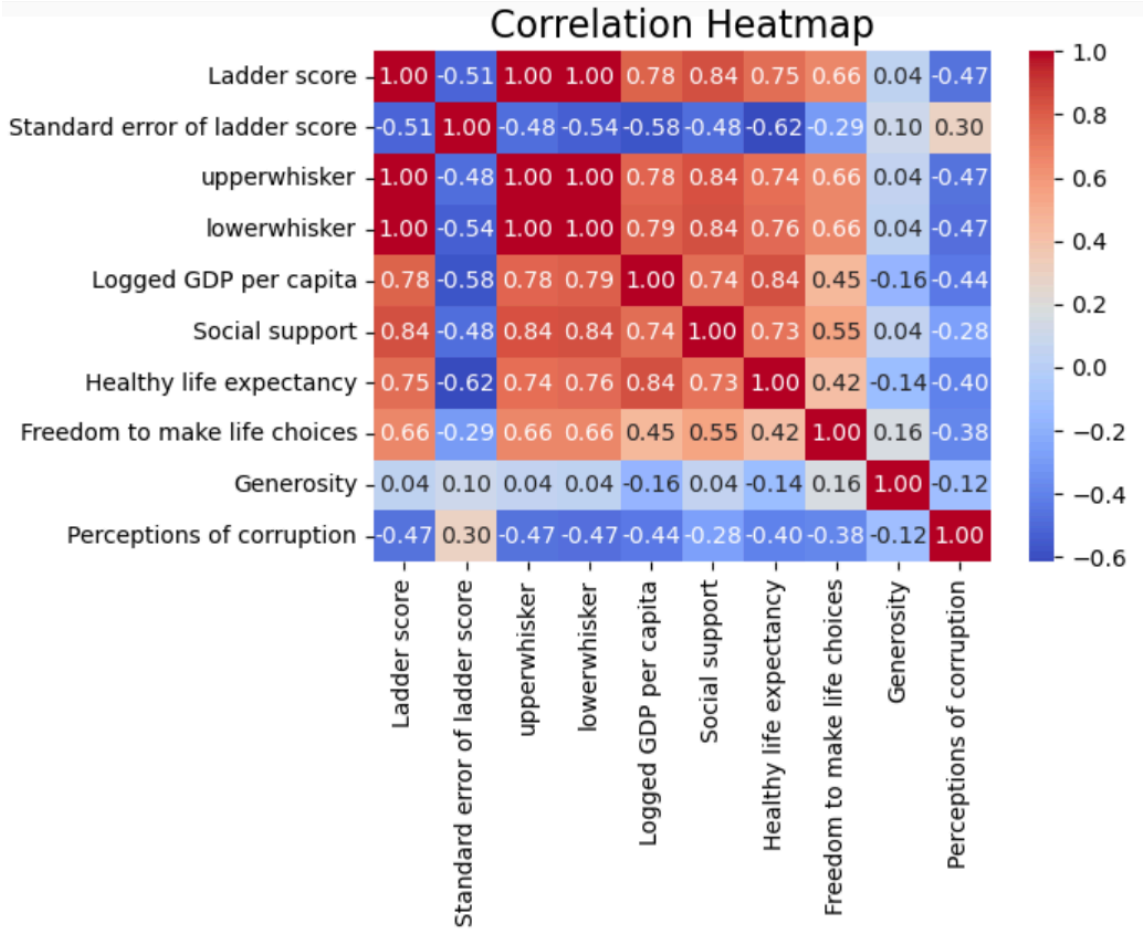
Visualisations

Histograms for Main Columns: Include histograms for key columns to visually represent the distribution of data and identify patterns.

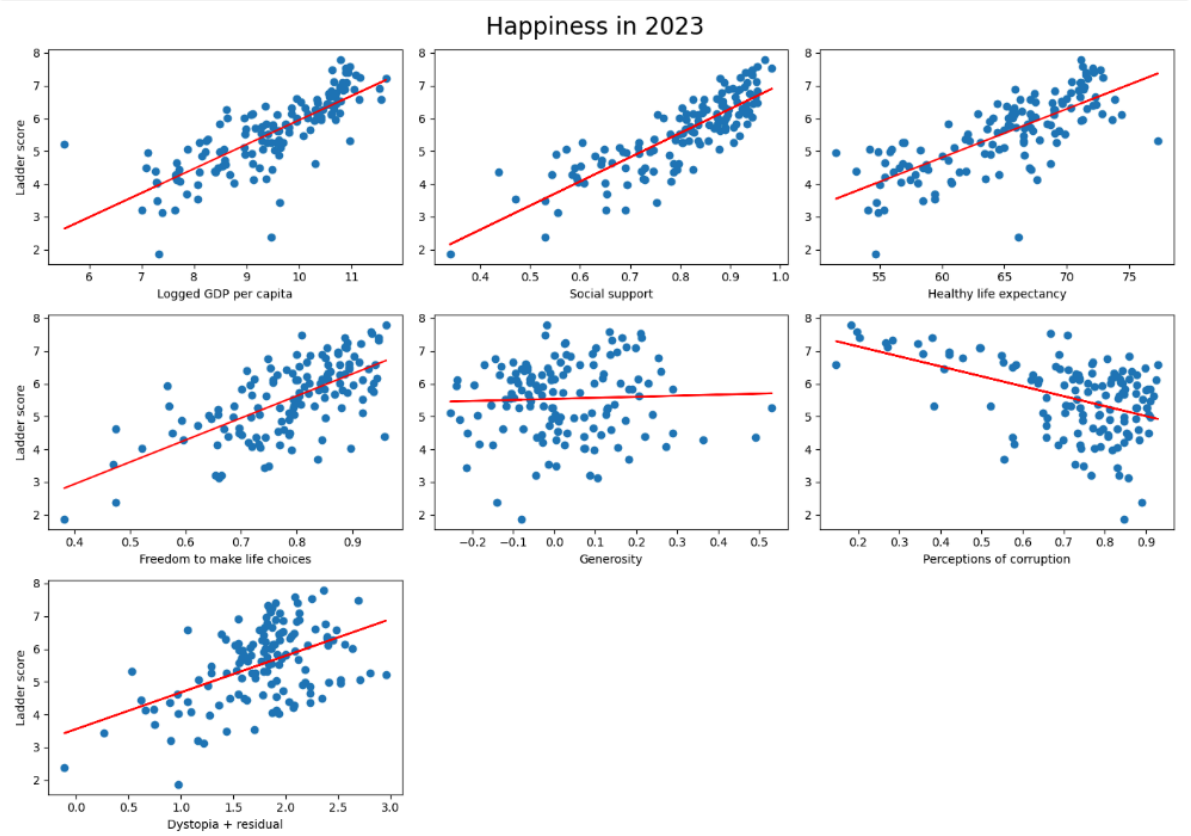


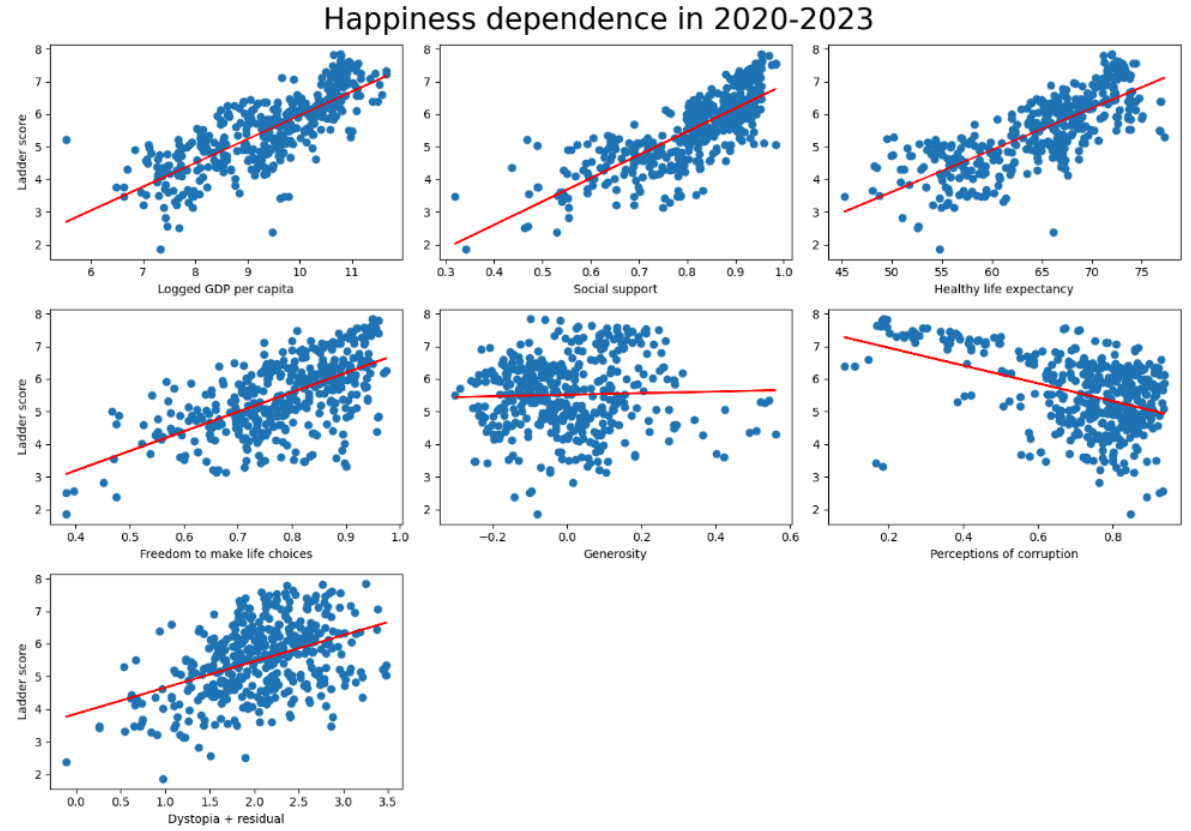


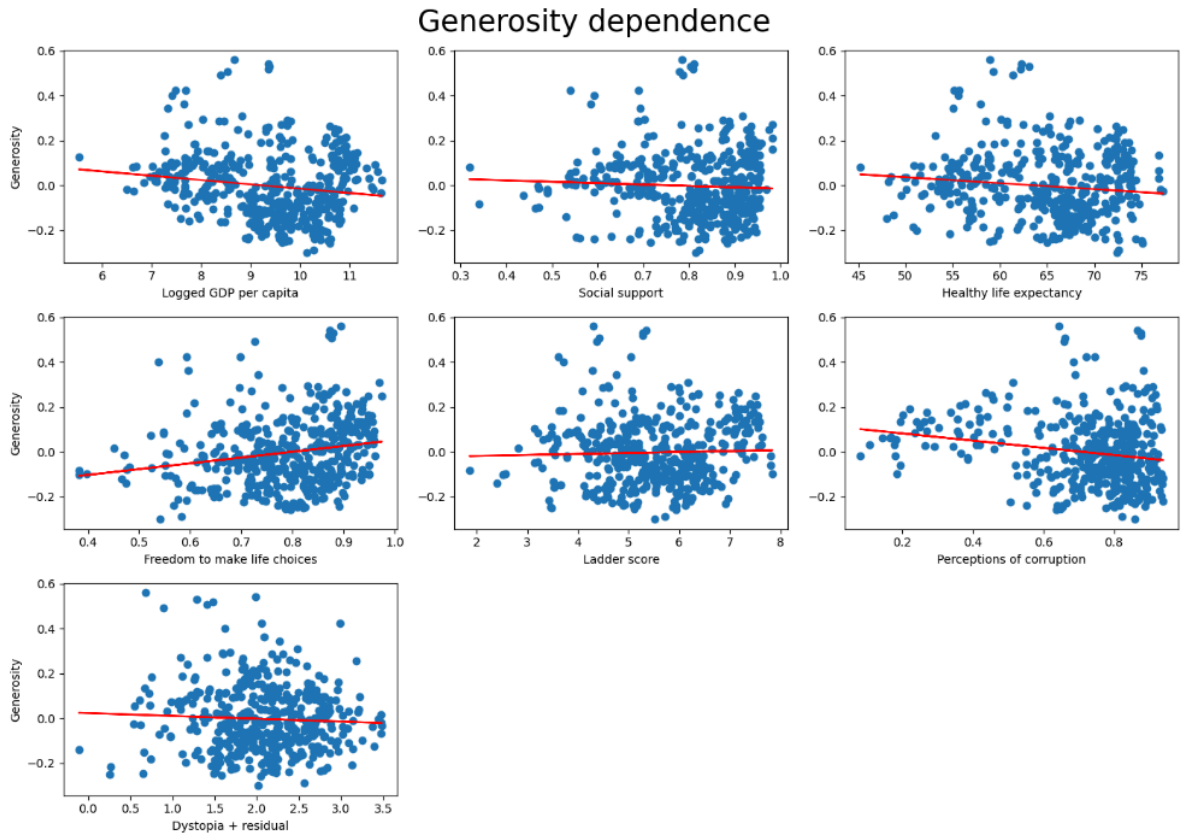
Correlation Heatmap: Display the correlation matrix heatmap to illustrate the relationships between main columns.



Scatter Plots for Feature Analysis: Include scatter plots illustrating the relationship between each feature and the "Ladder score."

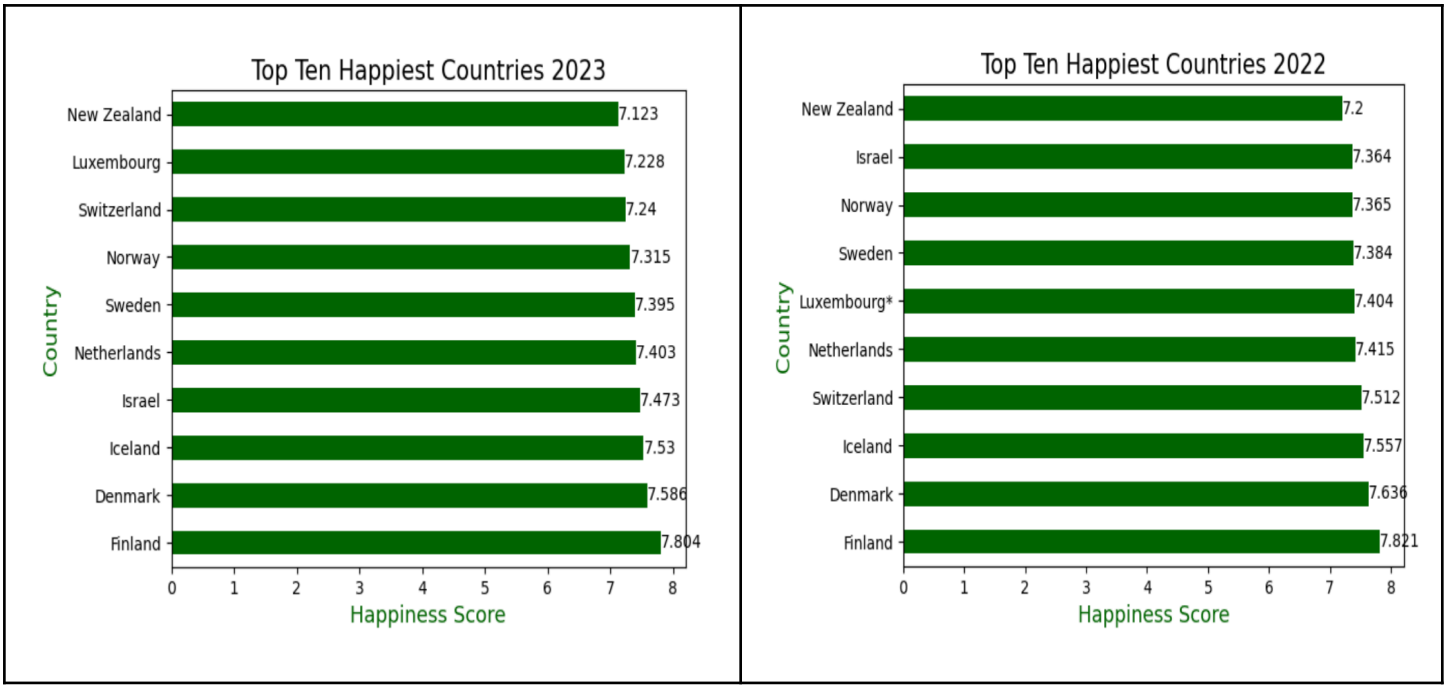


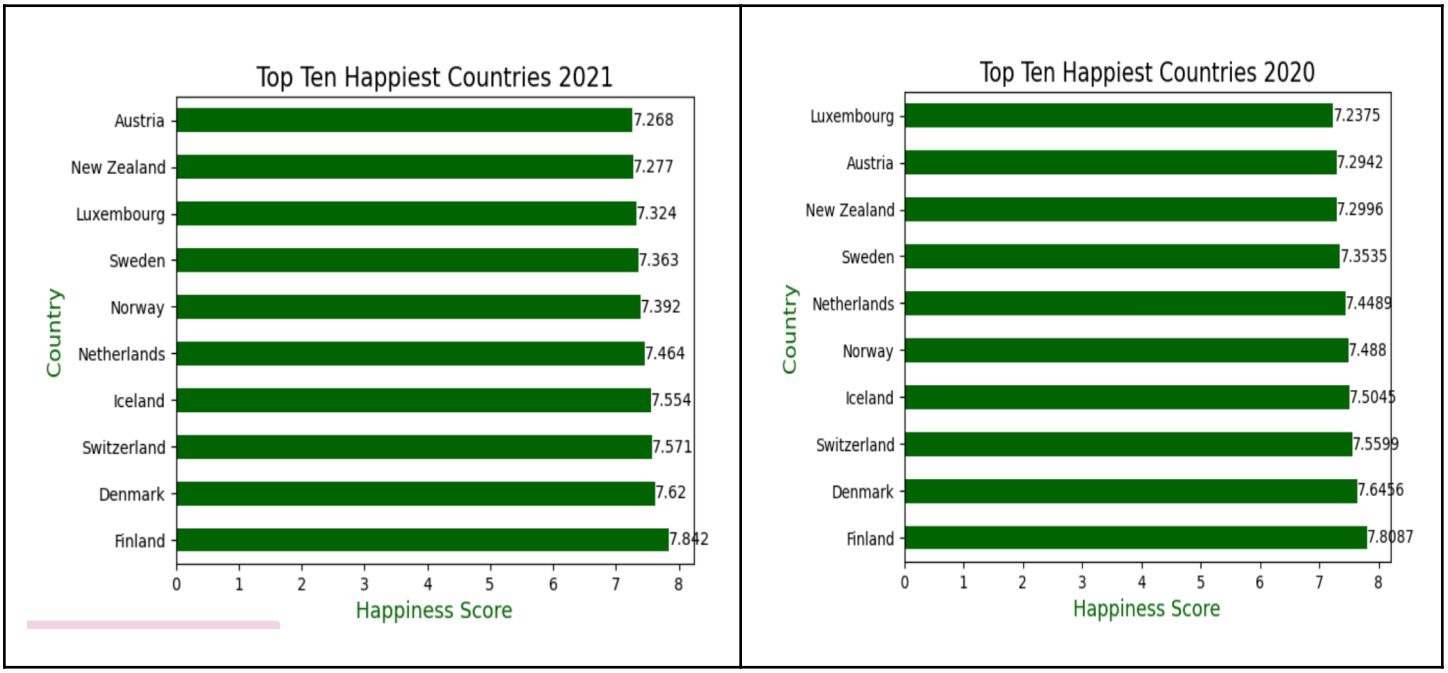




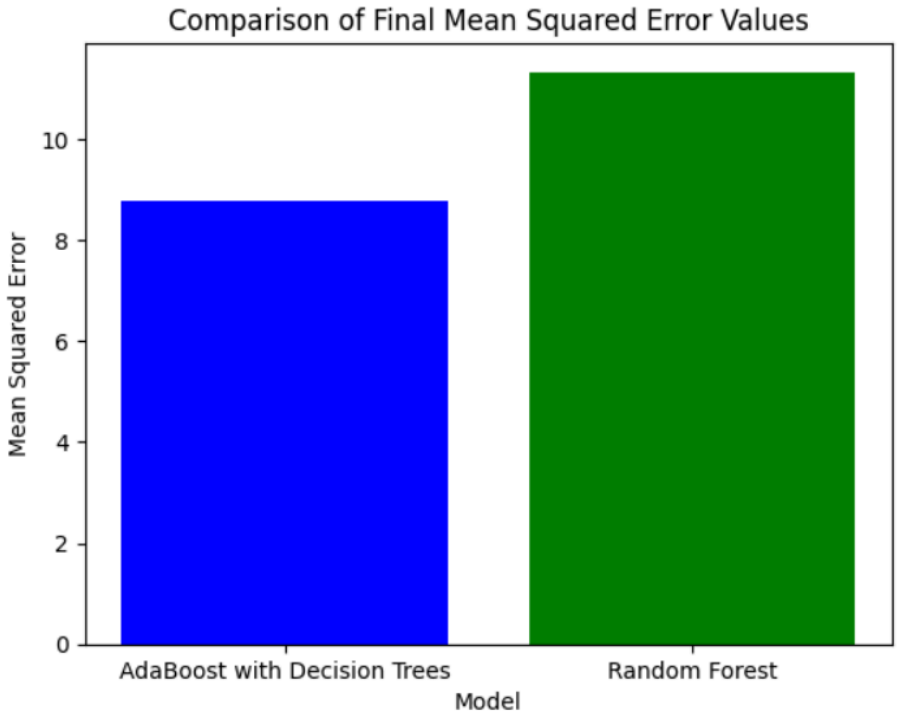
It appears that in countries with higher GDP and increased perception of corruption, generosity tends to be lower. Conversely, individuals tend to exhibit greater generosity in environments where there is more freedom to make life choices.

Top 10 Happiest Countries Bar Plots: Visualise the top 10 happiest countries for each year using bar plots.





Model Evaluation Metrics: Present a comparison of Mean Squared Errors (MSE) between the training and validation sets for both AdaBoost and Random Forest.



Random Predictions Examples: Include random predictions with actual and predicted life expectancy values for a few countries.

```
Example 1: Actual Life Expectancy = 54.05, Predicted Life Expectancy = 55.80020198803277, Difference = 1.75

Example 2: Actual Life Expectancy = 65.656, Predicted Life Expectancy = 66.3122742698462, Difference = 0.66

Example 3: Actual Life Expectancy = 57.34901047, Predicted Life Expectancy = 56.360154321607126, Difference = 0.99

Example 4: Actual Life Expectancy = 57.49607468, Predicted Life Expectancy = 58.15759592899999, Difference = 0.66

Example 5: Actual Life Expectancy = 51.0, Predicted Life Expectancy = 55.80020198803277, Difference = 4.8
```

Conclusion

The analysis of the World Happiness Report datasets provided valuable insights into the factors influencing happiness. The combination of exploratory data analysis, feature analysis, and model building has given us a comprehensive understanding of the dataset. Social support emerged as a crucial factor, and the AdaBoost model showed superior performance in predicting healthy life expectancy.