PROJECT DOCUMENTATION

EXPLORATORY DATA ANALYSIS USING PYTHON

|  |
| --- |
| TITLE: Exploring Global Happiness Dataset |
| NAME: JEFRINSOBI S |
| COURSE: DA/DS, Offline |
| BATCH: MAY 2025 |

TABLE OF CONTENT

|  |  |
| --- | --- |
| 1. | Introduction |
| 2. | Aim |
| 3. | Problem Statement |
| 4. | Project Workflow |
| 5. | Data Understanding |
| 6. | Data Cleaning |
| 7. | Exploratory Data Analysis (EDA) - Univariate Analysis |
| 8. | Bivariate Analysis |
| 9. | Multivariate Analysis |
| 10. | Overall Insights from Analysis |
| 11. | Conclusion |

Happiness Dataset Analysis Documentation

**INTRODUCTION**

Happiness is one of the most important indicators of human well-being. The World Happiness Report dataset contains country-level data on factors such as GDP, health, freedom, social support, generosity, and perceptions of corruption. This project aims to analyze this dataset using exploratory data analysis (EDA) techniques to identify patterns, correlations, and drivers of happiness across the globe.

**AIM OF THE PROJECT**

- Understand patterns in the World Happiness dataset through Exploratory Data Analysis (EDA).  
- Identify the factors that most impact Happiness Scores across countries.  
- Clean and transform data to enhance quality and accuracy.  
- Visualize trends, relationships, and regional differences in happiness.  
- Apply statistical methods (t-tests, ANOVA, correlation, regression) to validate insights.

**PROBLEM STATEMENT**

Measuring happiness is complex as it is influenced by multiple factors, including economic strength, health, freedom, social support, generosity, and governance. The challenge lies in identifying which features most strongly influence happiness scores and how these patterns vary across different regions and continents.

**PROJECT WORKFLOW**

1. Data Collection: Loaded World Happiness dataset.  
2. Data Cleaning & Preprocessing: Handled missing values, treated outliers, standardized column names.  
3. Exploratory Data Analysis (EDA): Conducted univariate, bivariate, and multivariate analysis.  
4. Feature Engineering: Created meaningful metrics where required.  
5. Statistical Analysis: Applied correlation, t-test, z-test, and ANOVA.  
6. Visualization: Heatmaps, boxplots, bar charts to interpret results.  
7. Insights & Conclusion: Summarized key findings.

**DATA UNDERSTANDING**

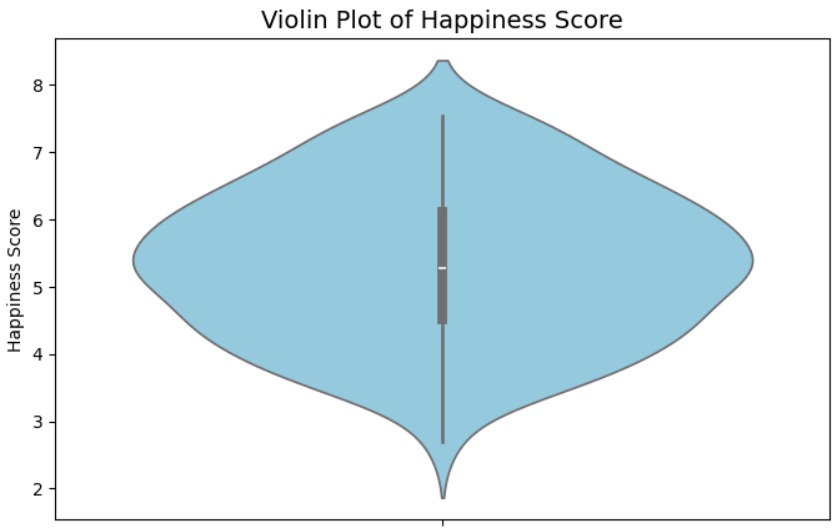
The dataset contains information on multiple countries with happiness scores and associated factors. Key variables include:  
- Happiness Score: The overall happiness rating for each country.  
- GDP per capita: Economic measure of wealth.  
- Health: Life expectancy and healthcare quality.  
- Freedom: Ability of individuals to make life choices.  
- Generosity: Social giving behavior.  
- Perceptions of Corruption: Trust in institutions.  
- Family/Social Support: Strength of social ties.

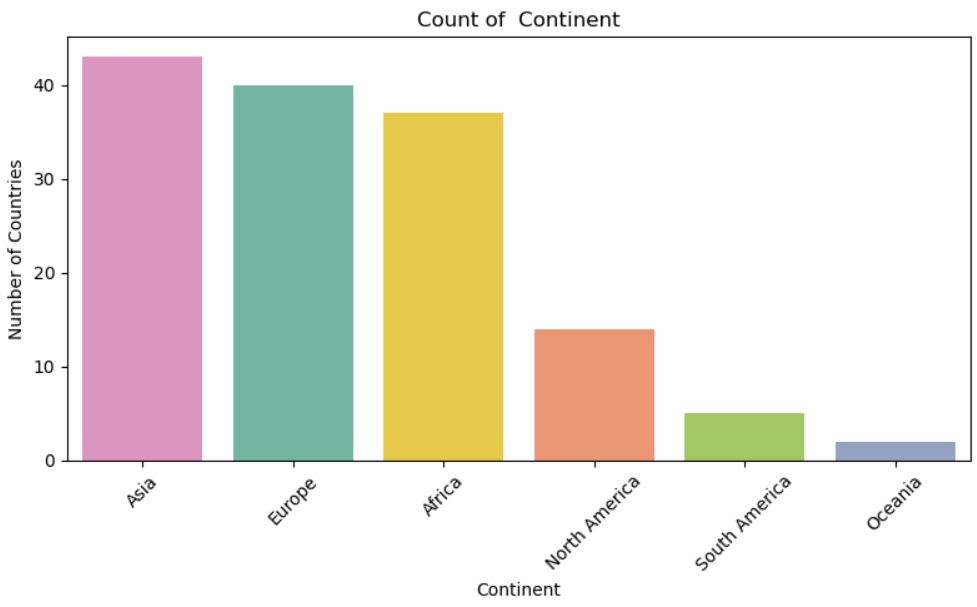
**DATA CLEANING**

- Missing values imputed using mean/median strategies.  
- Outliers checked using boxplots and treated where necessary.  
- Duplicates dropped.  
- Standardized column names for clarity.

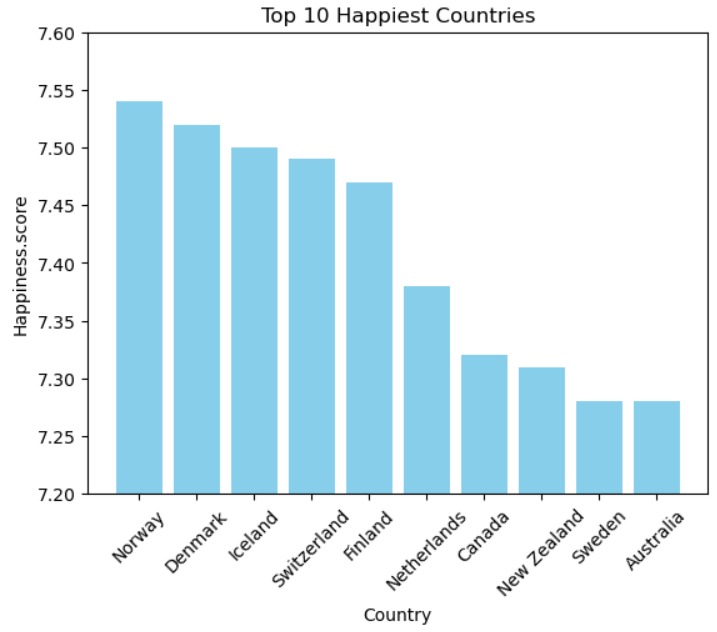
**EXPLORATORY DATA ANALYSIS (EDA)**

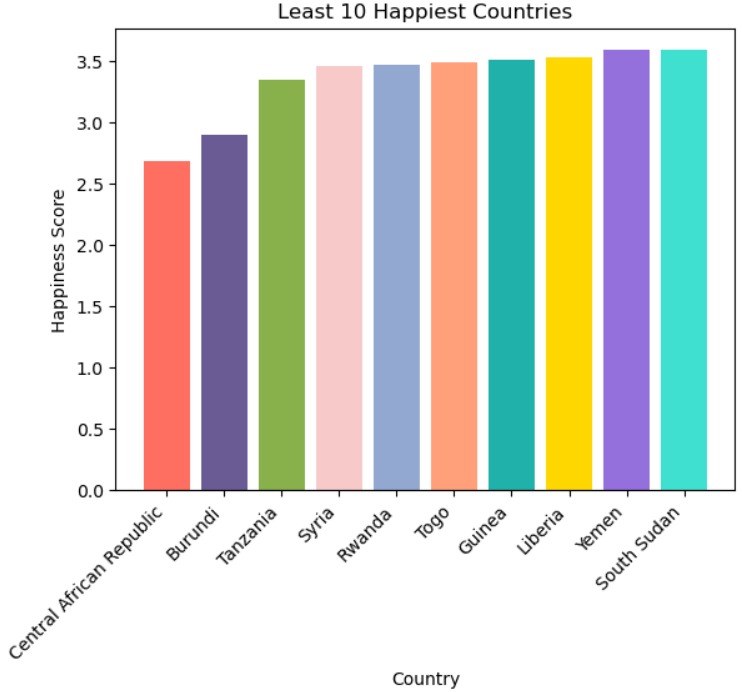
\*\*Univariate Analysis\*\*  
- Most countries score between 5–6 in happiness.  
- But the mean of Happiness score is 5.53,where there is a sharp peak  
- The count plot shows the count of countries in particular continent

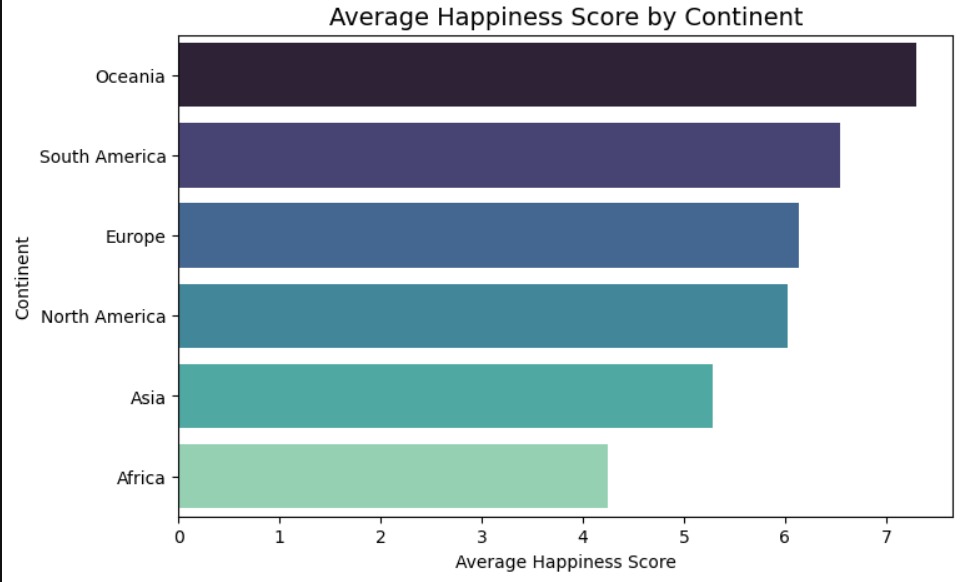




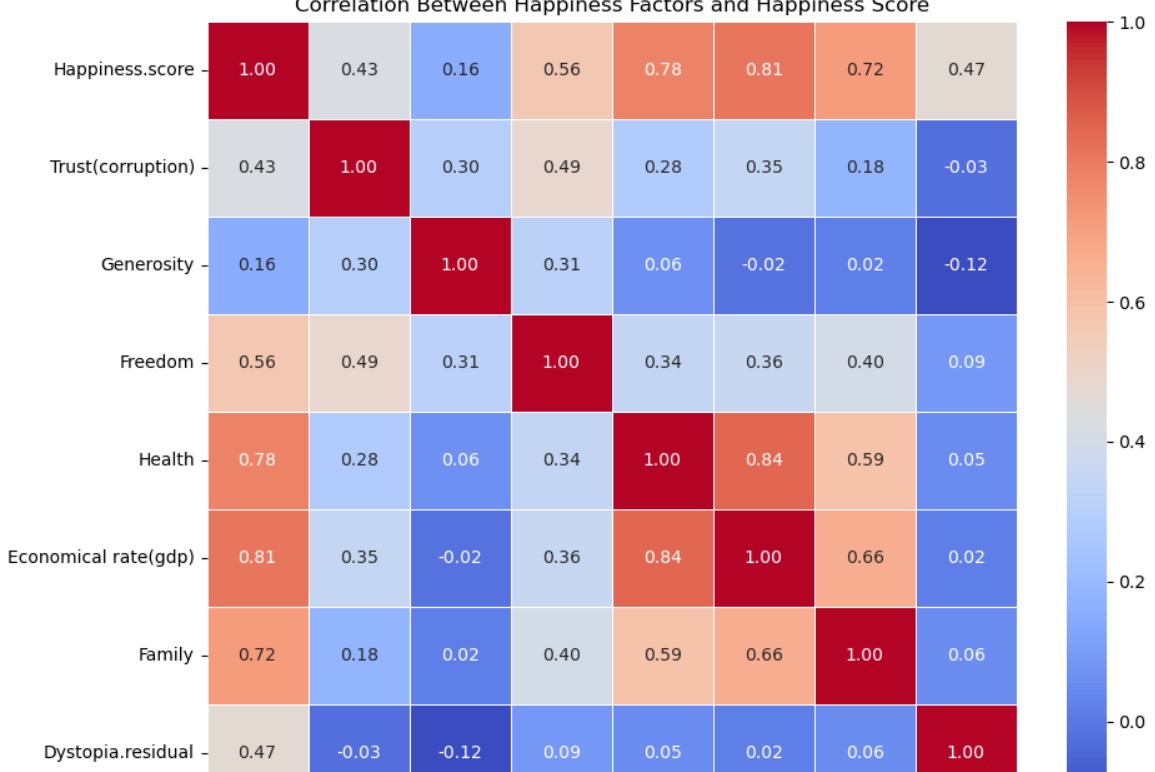
\*\*Bivariate Analysis\*\*  
- It shows how one variable impacts or relates to another.  
- Helps in identifying patterns, trends, and comparisons.  
- Useful for finding correlations and differences between groups.



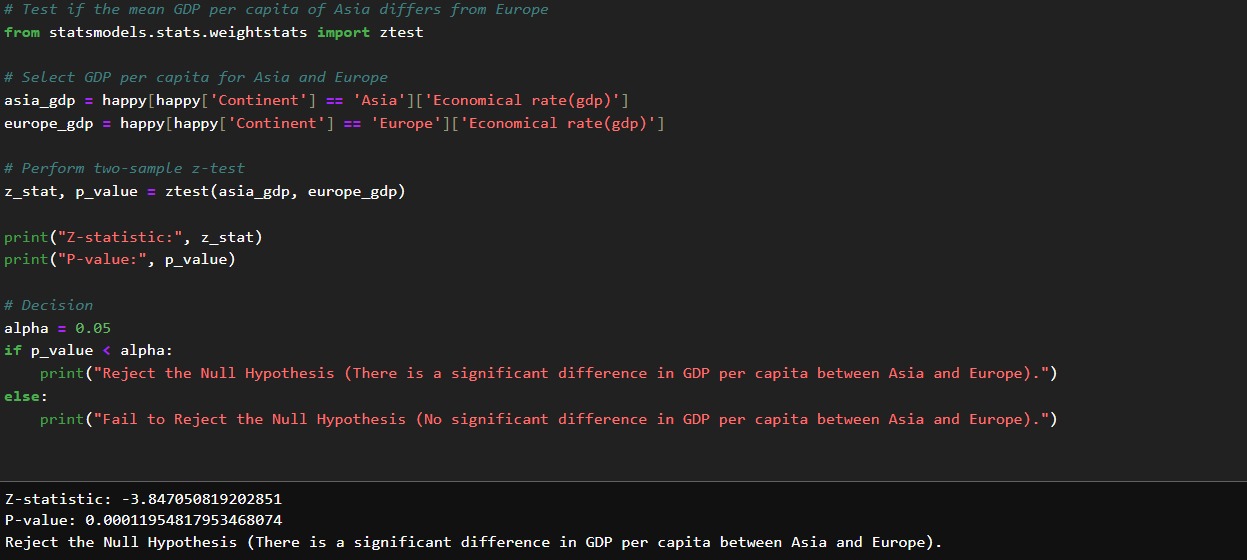




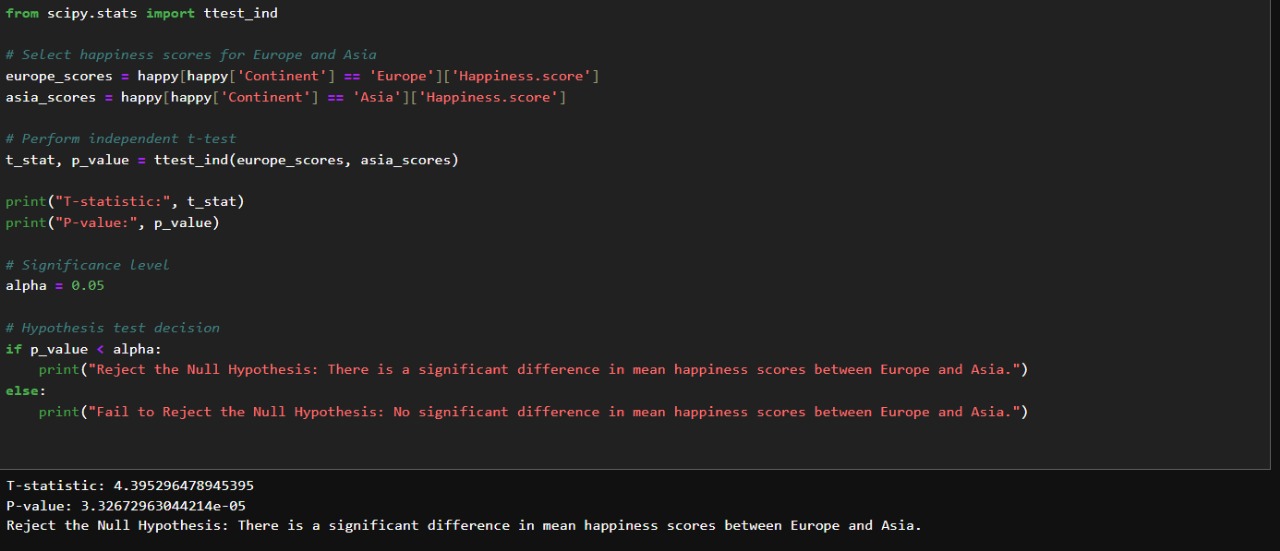
\*\*Multivariate Analysis\*\*  
- The heatmap is used to identify the correlation between the columns  
- The Economical rate is highly correlated with Happiness score.



**HYPOTHESIS TESTING**



- T-test: Europe has significantly higher happiness than Asia.  
- Z-test: GDP per capita is significantly higher in Europe than Asia.  
- ANOVA: Happiness scores vary significantly across continents.  
- Correlation & Regression: GDP, Health, and Freedom are strongest predictors.



**OVERALL INSIGHTS**

1. Data Quality & Preparation Insights:  
-- Missing values handled using mean/median.  
-- Outliers treated using visual inspection and IQR.  
-- Columns standardized for clarity.  
  
2. Univariate Analysis Insights:  
-- Happiness Score lies mostly between 4.5–6.  
-- GDP, Health, and Freedom vary widely.  
  
3. Bivariate Analysis Insights:  
-- GDP strongly correlated with Happiness.  
-- Health and Freedom moderately correlated.  
-- Generosity weak correlation.  
  
4. Multivariate Analysis Insights:  
-- Heatmap shows GDP, Health, and Freedom as strongest drivers.  
-- Regional differences highlight Europe at top, Africa at bottom.  
  
5. Hypothesis Testing Insights:  
-- Significant difference in happiness between continents.  
-- GDP and Health are statistically proven drivers.  
  
6. Business/Policy-Level Takeaways:  
-- Economic development is crucial for improving happiness.  
-- Investments in healthcare and freedom lead to better well-being.  
-- Anti-corruption measures can improve trust and happiness.

**CONCLUSION**

This Happiness dataset project explored factors influencing well-being across countries. The analysis revealed that GDP, Health, and Freedom are the strongest contributors to happiness, while Generosity has minimal impact. Regional differences highlight inequality, with Europe and Nordic nations consistently ranking highest, while Africa shows lower