Project Title: Analyzing Global Happiness Trends (2015-2019)

1. Introduction The purpose of this project is to explore global happiness trends over time, identifying key factors that influence happiness scores across different countries and regions. The analysis will cover five years (2015-2019) of data from the World Happiness Report, using advanced statistical techniques and visualization tools to uncover meaningful insights.

2. Project Objectives

- Understand how happiness scores have evolved over time.
- Identify the most influential factors affecting happiness in different regions.
- Compare country-wise happiness rankings and explore regional differences.
- Utilize geospatial and regression analysis to determine correlations between happiness and socio-economic indicators.
- Present findings through an interactive Tableau dashboard.

3. Data Source

The dataset used for this analysis is the **2019 World Happiness Report**, which provides insights into happiness rankings across different countries. The data includes several key socio-economic indicators that contribute to happiness, such as **GDP per capita, social support, life expectancy, personal freedom, generosity, and perceptions of corruption**.

- **Source:** Kaggle - World Happiness Report Dataset
- **Year:** 2019
- **Total Rows:** 156
- **Total Columns:** 9
- **Data Categories:**
    - **Categorical Features:** country_or_region
    - **Numerical Features:** overall_rank, score, gdp_per_capita, social_support, healthy_life_expectancy, freedom_to_make_life_choices, generosity, perceptions_of_corruption

Why This Dataset Was Chosen: The World Happiness Report dataset was selected because it provides a comprehensive view of global well-being metrics over multiple years. It meets all project criteria, including:

- **Geographical coverage**: Data is available at the country level, allowing for regional comparisons and geospatial visualizations.
- **Rich socio-economic insights**: Includes factors such as GDP per capita, social support, and life expectancy, enabling a multifaceted analysis of happiness determinants.
- **Time-series potential**: Data spans five years, allowing for trend analysis and forecasting.
- **Reputable source**: The dataset originates from a well-documented and widely used global study, ensuring credibility and reliability.

DATA PROFILE

6.Data Cleaning

Before conducting analysis, the dataset underwent a **comprehensive cleaning process** to ensure data accuracy and consistency.

Cleaning Steps Taken:

- **Handled Missing Data:** No missing values were found, so no imputation was needed.
- **Removed Duplicates:** No duplicate rows were present in the dataset.
- **Standardized Column Names:** Column names were converted to lowercase and spaces were replaced with underscores for uniformity.
- **Checked Data Types:** Verified that numerical variables were correctly formatted as float64 and categorical variables were kept as object.
- **Saved the Cleaned Dataset:** A cleaned version (cleaned_2019.csv) was created and stored in /DATA/Prepared data/.

7. Descriptive Statistical Analysis

After cleaning, a **descriptive statistical analysis** was performed to understand the distribution of key variables.

Summary Statistics:

- **Happiness Score (score):**
    - **Minimum:** 2.85 (Least happy country)
    - **Maximum:** 7.77 (Happiest country)
    - **Mean:** 5.41
- **GDP Per Capita (gdp_per_capita):**
    - **Minimum:** 0.18
    - **Maximum:** 1.68
    - **Mean:** 0.90
- **Healthy Life Expectancy (healthy_life_expectancy):**
    - **Minimum:** 0.36
    - **Maximum:** 1.14
    - **Mean:** 0.72
- **Social Support (social_support):**
    - **Mean:** 1.19
    - **Indicates a strong correlation with happiness**

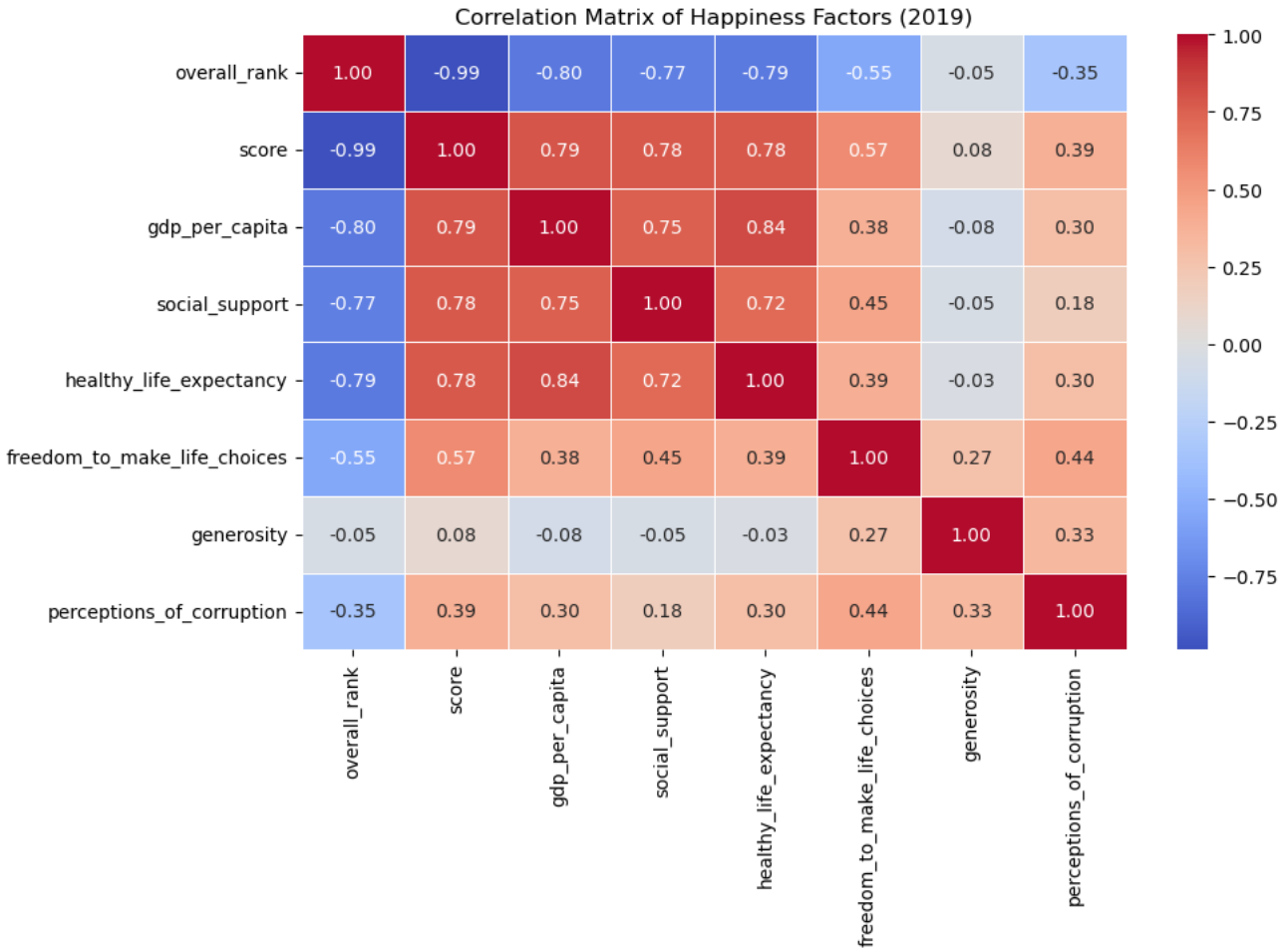
Correlation Analysis

To explore the relationships between happiness and socio-economic factors, a **correlation heatmap** was generated.

- **Economic Prosperity Matters:**
    - **gdp_per_capita (0.79) and healthy_life_expectancy (0.78) have the strongest correlations with happiness (score).**
    - Wealthier countries and those with better healthcare tend to have **happier populations**.
- **Social Support is a Major Factor:**
    - **social_support (0.78) strongly correlates with happiness**, suggesting that **strong community networks lead to higher well-being**.
- **Personal Freedom Contributes to Happiness:**
    - **freedom_to_make_life_choices (0.57) is positively correlated**, indicating that people feel happier in societies where they have more control over their lives.
- **Corruption Perception Plays a Role:**
    - **perceptions_of_corruption (0.39) shows a moderate correlation**, meaning that **lower corruption levels generally lead to higher happiness scores**.
- **Generosity Has a Weak Impact:**
    - **generosity (0.08) shows the weakest correlation**, implying that **happiness is not significantly driven by self-reported generosity**.

Visual Evidence:

A correlation heatmap was generated to visually present these findings. This visualization highlights the **strongest and weakest relationships** between happiness and other factors.



8. Limitations & Ethical Considerations

While the **2019 World Happiness Report** provides valuable insights into global well-being, it is essential to acknowledge **its limitations and ethical concerns**. Understanding these factors ensures that any conclusions drawn from the dataset are **contextually accurate and responsibly interpreted**.

A) Data Collection Biases

The dataset relies on **self-reported survey data**, which introduces potential **biases and limitations**:

- **Subjective Responses:**
    - Happiness is inherently **subjective**, and people may interpret happiness scales differently based on **cultural, personal, or situational factors**.
    - A score of "7" in one country **may not have the same meaning** as a "7" in another.
- **Sampling & Representation Issues:**
    - Some countries may have **limited survey participation**, leading to **underrepresentation** of certain demographics.
    - Rural and underprivileged populations may be **underrepresented** in some regions, skewing the results in favor of urban, wealthier populations.
- **Cultural & Societal Differences:**
    - Different cultures have varying **attitudes toward happiness**, which may impact how individuals respond to surveys.
    - Some respondents may **overstate or understate their happiness levels** due to **social norms or government influence**.

B) Limitations of the Data

Despite being **comprehensive**, the dataset has certain **limitations** that affect its interpretation.

- **Economic Indicators Are Simplified:**
    - GDP per capita is **not a complete measure** of economic well-being.
    - Income inequality is **not considered**, meaning a country could have high GDP per capita but also high economic disparity.
- **Limited Scope of Social & Political Factors:**
    - The dataset includes **perceptions of corruption** but does not assess **actual government policies or political stability**.
    - Other factors such as **education levels, employment rates, or environmental sustainability** are not included.
- **Single-Year Analysis Constraint:**
    - This study focuses only on **2019**, limiting its ability to detect **happiness trends over time**.
    - A **multi-year comparison (2015-2019) could provide better insights** into long-term happiness trends.

C) Ethical Considerations

When working with happiness data, ethical concerns must be **carefully considered**, particularly regarding **data privacy, interpretation, and policy implications**.

- **Privacy & Anonymity:**
    - The dataset consists of **aggregated country-level data**, so individual responses remain anonymous.
    - However, in future studies, ensuring **data protection for survey participants** should remain a priority.
- **Use in Policy & Decision-Making:**
    - Governments and organizations may use happiness scores for **policy recommendations**, but **misinterpretation could lead to ineffective policies**.
    - For instance, **high GDP does not guarantee happiness**, so focusing solely on economic factors may ignore critical social and psychological well-being aspects.
- **Potential for Misuse:**
    - Countries with **lower happiness rankings** may **discredit the report** or use it for political purposes.
    - Organizations using this data must ensure that conclusions are **fair, evidence-based, and do not reinforce stereotypes**.

10. Defining Questions to Explore

To develop meaningful insights from the 2019 World Happiness Report, we will structure our questions based on different exploratory approaches:

-Broad Exploratory Questions

These questions help us understand general happiness trends across countries.

- Which countries had the highest and lowest happiness scores in 2019?
- How does happiness vary across different regions of the world?
- What are the key factors that contribute most to happiness?

-Relationship-Based Questions (Correlations)

We will explore how different socio-economic factors relate to happiness.

- What is the relationship between GDP per capita and happiness?
- How strongly does social support impact happiness levels?
- Does life expectancy correlate with happiness across countries?
- Are perceptions of corruption linked to lower happiness scores?
- Does greater freedom in life choices contribute to higher happiness scores?

-Time-Series & Trends Questions

Since we have data from 2015-2019, we can explore happiness trends over time.

- How have happiness rankings changed over the past five years (2015-2019)?

- Which countries have experienced the biggest improvements or declines in happiness?
- Has the relationship between GDP and happiness strengthened or weakened over time?
- Have corruption perceptions improved or worsened in happier countries over time?

- Comparison-Based Questions

Comparing different groups of countries can provide deeper insights.

- Do higher-income countries always have higher happiness scores?
- Are there significant differences in happiness between developed and developing countries?
- How do happiness factors differ between the top 10 and bottom 10 ranked countries?
- Are countries with similar GDP levels equally happy, or do social factors play a bigger role?