

World Happiness Report 2021 Analysis

Overview

The World Happiness Report 2021 is a comprehensive dataset that captures the happiness levels across different countries and the factors that might influence it. This project aims to explore the dataset, understand its structure, and answer pertinent research questions using statistical analyses.

Project Structure

1. **Dataset:** World Happiness Report 2021 (Source: Kaggle)

2. **Data Exploration:**

- **Variables Analyzed:**

- Ladder score (Happiness score)
- Logged GDP per capita
- Social support
- Healthy life expectancy
- Freedom to make life choices
- Generosity
- Perceptions of corruption

For each variable, the following analyses are conducted:

- Boxplot to observe distribution and outliers.
- Histogram to visualize frequency distribution.
- Descriptive statistics: mean, median, range, and quantiles.
- QQ plot to inspect normality.

3. **Research Questions:**

- **a. Relationship between GDP per capita and Happiness Score**

- Conducted a Pearson correlation.
- Created a scatter plot with a linear regression line to visually inspect the relationship.

- **b. Generosity levels in countries with high vs low perceptions of corruption**

- Divided countries based on median perceptions of corruption.
- Conducted a t-test to compare generosity scores.

- **c. Predicting Happiness Score based on multiple factors**

- Utilized a multiple linear regression model with factors like GDP, social support, life expectancy, freedom, generosity, and perceptions of corruption.

- Checked assumptions through VIF (Variance Inflation Factor) and QQ plots.
- Used stepwise regression for model optimization.

Tools & Libraries Used

- **R:** Primary tool for data analysis and visualization.
- **tidyverse:** For data manipulation and visualization.
- **car:** For statistical functions like variance inflation factor (VIF).

Files Included

- **R file:** Contains the complete code used for data analysis.
- **Excel file:** World Happiness Report 2021 dataset.
- **Google Doc:** Contains detailed analysis of variables and outlines research questions/hypotheses.

Sample Size Details

Typically, the annual sample for each country consists of 1,000 people. If a country had surveys each year, the sample size would amount to 3,000. A sample size ranging from 2,000 to 3,000 provides a reasonably good estimate at the national level.

Key Insights & Findings

The README does not contain direct results, but by executing the provided R code, users can extract specific insights, correlations, and outcomes pertaining to the research questions.

Recommendation for Use

1. Ensure you have the necessary R packages installed ('tidyverse' and 'car').
2. Make sure the dataset path is correctly specified.
3. Run the code sequentially, starting with the data exploration, followed by research questions.
4. Interpret results in the context of the World Happiness Report and the factors analyzed.

Acknowledgements

This analysis is based on the dataset provided by Ajaypal Singh (2021) on Kaggle.