DATA VISUALIZATION PROJECT

Members:

Sujata Joshi
Shagun Ruperia
Shazid Shaik
Karthik Nandam

AGENDA

- Problem Statement
- Objective
- Dataset Information
- Exploratory Data Analysis
- Distribution of numerical features
- Relationships between variables
- Column Description
- Machine Learning Model

PROBLEM STATEMENT

Develop a data visualization project to analyze Global Social Progress Index (SPI), which measures the extent to which countries provide for the social and environment needs of their citizens, aiming to identify trends, patterns, and disparities in social progress across different countries and regions.

OBJECTIVE

- Analyzing factors influencing social progress across countries.
- Investigating the relationship between social indicators and overall progress.
- Examine trends and disparities in social progress worldwide.

DATASET INFORMATION

Dataset Information :

Source : Kaggle, **Size** : 18 KB

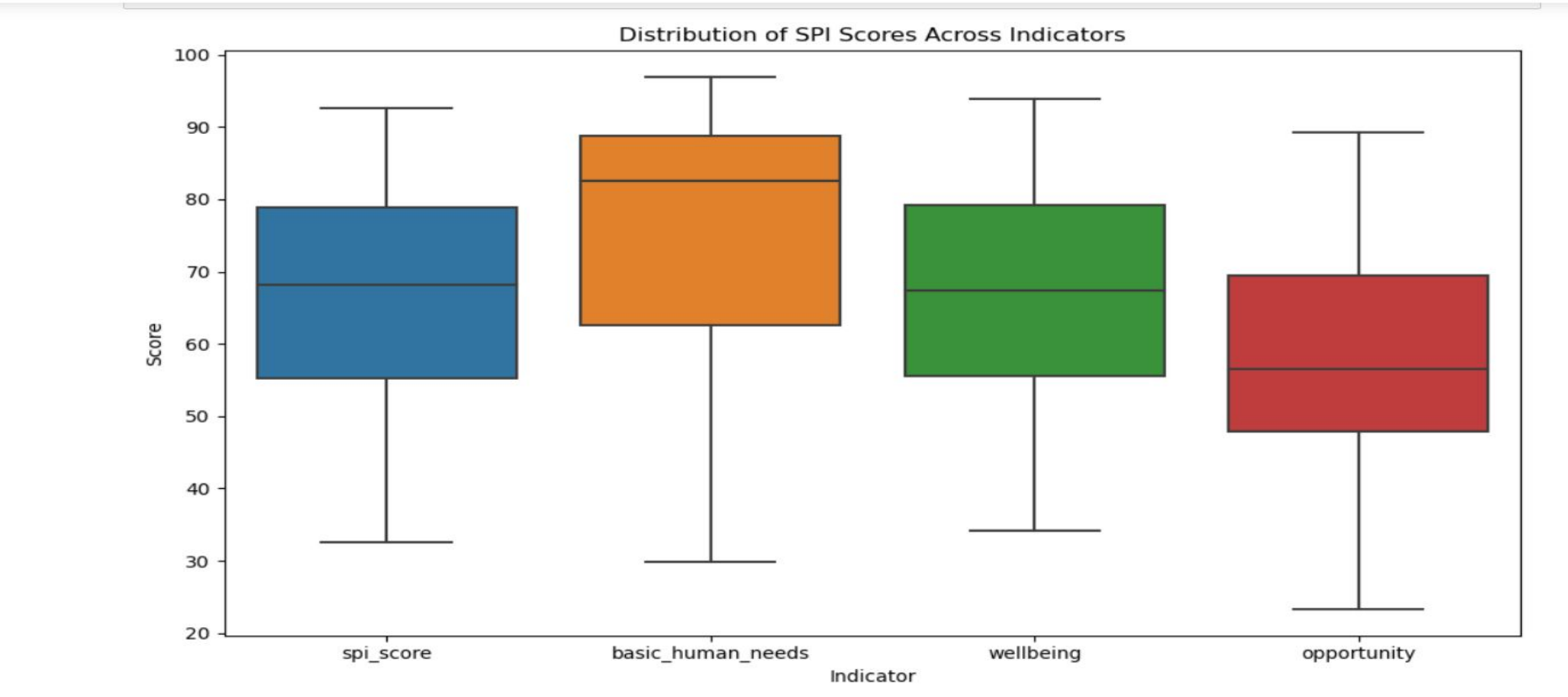
Dataset Description :

- Indicators related to social progress such as basic human needs, access to education, personal freedom and rights, inclusiveness, environmental quality, sanitation, opportunity and safety.
- List of countries and their corresponding social progress index.
- SPI – Measures the well being of a society based on the indicators mentioned above.

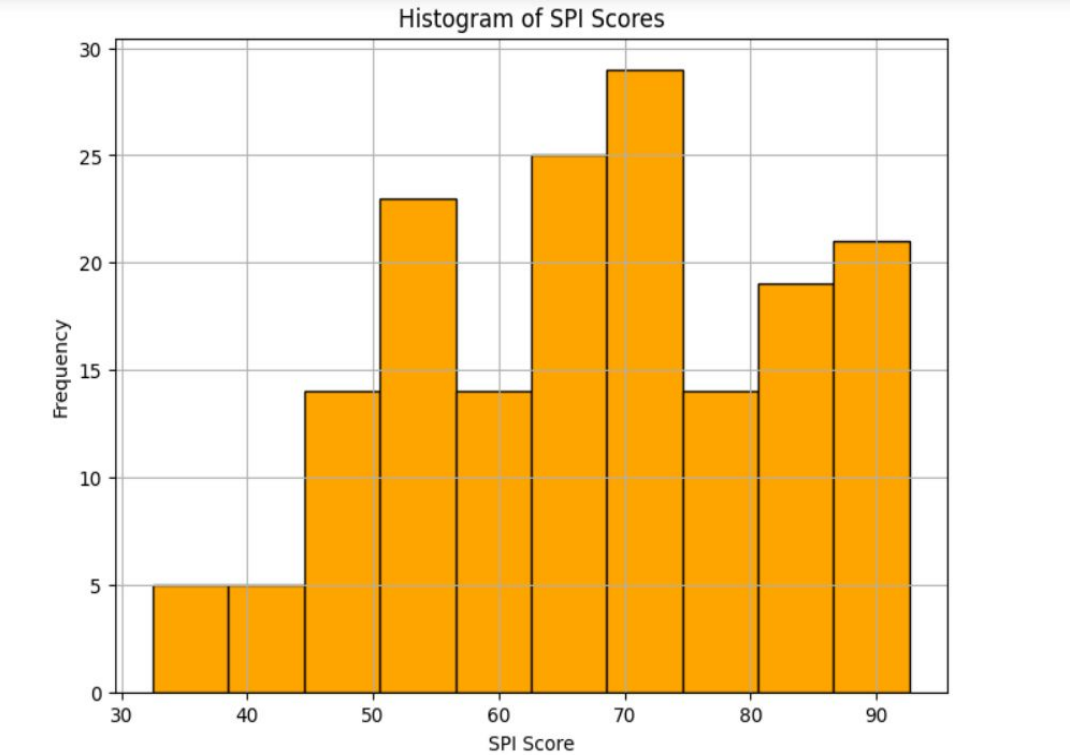
EXPLORATORY DATA ANALYSIS

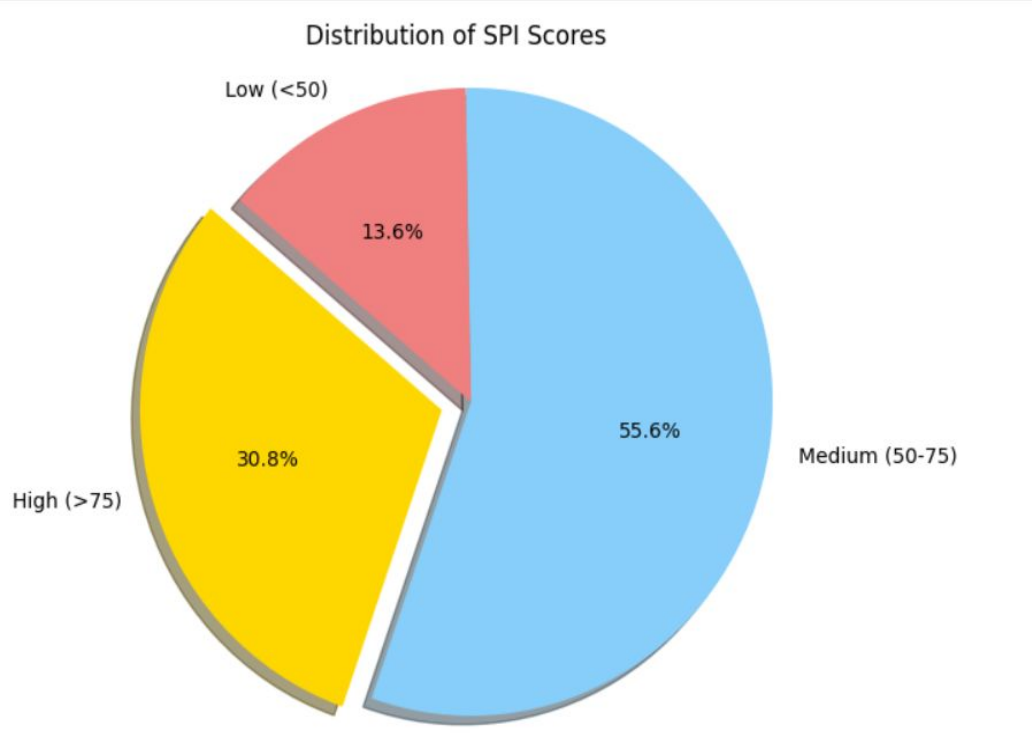
- **Descriptive Statistics** – Compute basic statistics (mean, median, standard deviation, etc.) for each variable to understand the central tendency, dispersion, and distribution of the data.
- **Correlation Analysis** - Investigate the relationships between different variables using correlation coefficients (e.g., Pearson correlation or Spearman correlation).
- By creating plots such as histograms, box plots and scatter plots to visualize relationships between variables.
- **Geographic Analysis** – Mapping the Social Progress Index across countries to visualize regional patterns and disparities.

DISTRIBUTION OF NUMERICAL FEATURES



RELATIONSHIP BETWEEN VARIABLES





COLUMN DESCRIPTION

'spi_rank': rank of the country

'country': name of the country

'spi_score': social progress score

'basic_human *needs'*: *basic human needs*

'wellbeing': *foundations of wellbeing*

'opportunity': *opportunity*

'basic **nutri_ med_ care'**: nutritional and basic medical care

'water_sanitation': water and sanitation

'shelter': shelter

'personal_safety' : personal safety

'access_basic _knowledge' : access to basic knowledge , **'access_info _comms'**: access to information and communication

'health_wellness' : health and wellness

'env_quality' : environment quality

'personal_rights' : personal rights

'personal_freedom _choice': personal freedom and choice , **'inclusiveness'**: inclusiveness, : **'access_adv _edu'** access to advanced education

MACHINE LEARNING MODEL

- **Predictive Modeling using Regression :**

Machine learning models such as linear regression, random forest regression, or gradient boosting regression could be applied to predict the social progress index or its components for different countries. This could help in understanding the factors influencing social progress and predicting the potential impact of various interventions or policies.

MACHINE LEARNING MODEL

- **Predictive Modeling using Classification**

We could also use a model to classify countries into different categories based on their level of social progress. One approach could be to define categories such as “high”, “medium” and “low” social progress based on SPI scores and then use a classification model to predict which category each country falls into.

One suitable classification model is Random Forest classifier which is an ensemble learning method that combines multiple decision trees to make predictions.

Thank You !