Global Rice Production Analysis

Team 4:

Hari Prasannaa Thangavel Ravi (G28549456)

Jin Hyuk Son (G22379534)

Pranjal S. Wakpaijan (G40217310)

Predicting rice production outcomes is a critical aspect of global food security and agricultural sustainability. Understanding the factors that influence rice yield helps farmers and policymakers make informed decisions, ultimately improving food supply and nutrition worldwide. This project, titled "Global Rice Production Analysis," utilizes a dataset named “Explore Data on Agricultural Production,” which contains approximately 13,500 rows and roughly 40 columns. We aim to explore the intricate dynamics of rice production by analyzing key parameters such as land area used for cultivation, harvested area, climatic conditions, and factors related to waste generation. We will begin with exploratory data analysis (EDA) to clean the dataset, addressing missing values and outliers while examining data distributions and inter-relationships among variables. By evaluating various factors such as soil quality, rainfall patterns, and agricultural practices we aim to build predictive models that can estimate rice yields. This study not only addresses real-world challenges in rice cultivation but also provides valuable insights to enhance agricultural strategies, optimize resource allocation, and promote sustainable practices. The findings will support rice producers in improving yield predictions and ultimately contribute to enhancing global food security.

SMART Questions:

How can we use exploratory data analysis to identify key patterns in global rice production data over the past 60 years, and subsequently apply statistical analysis to validate these findings and quantify significant trends, with the goal of completing EDA and statistical analysis within the next month?

Specific:

1. What has been the annual average yield of rice in the USA from 1961 to 2022?
2. How many countries have achieved an overall production of above 25 Million tons and also have a top quartile performance rating?

Measurable:

1. How did the rice production volumes change annually over the 62-year period?
2. What percentage of the total rice produced is exported by each country?
3. What is the correlation between land parameters, such as the area used for cultivation and the area harvested, and yield?

Achievable:

1. What climatic or economic factors correlate with significant changes in rice production during this timeframe?
2. Which country has shown the most improvement in average production over the past ten years?

Relevant:

1. How do regional differences in rice production impact overall national output?
2. In sectors affected by COVID-19, how did the production change from Q1 2020 to Q2 2020?
3. What is the average supply chain waste and how does it compare to the overall production?

Time-bound:

1. What trends can be identified in rice production and yield over specific decades?
2. What were the top-producing countries in the last quarter, and have they consistently outperformed over the last three years?

Github link: https://github.com/hariravi-ds/DATS6101-Team4

Source of data: Our World in Data, https://ourworldindata.org/agricultural-production#explore-data-on-agricultural-production