A white robot hand with red and blue stripes

Description automatically generated



**AI/ML**

**LAB MANUAL**

**Disclaimer: The content is curated from online/offline resources and used for educational purpose only**

**Compare the average yields of a specific crop (e.g., rice) across multiple states**

**Objective:**

Develop and test null and alternative hypotheses to examine differences in crop yields across multiple states

**Equipment Required:**

* Computer with Python and Jupyter Notebook installed
* Dataset: [Agricultural Crop Yield in Indian States Dataset](https://www.kaggle.com/datasets/akshatgupta7/crop-yield-in-indian-states-dataset)
* Python Libraries: pandas, numpy, matplotlib, seaborn, scipy, statsmodels

**Prerequisites:**

* Basic understanding of Python programming
* Knowledge of statistics, including hypothesis testing
* Familiarity with pandas for data manipulation
* Understanding of data visualization using matplotlib and seaborn

**Problem Statement:**

Compare the average yields of a specific crop (e.g., rice) across multiple states to determine if there is a statistically significant difference in crop yields.

**Code:**

# Import necessary libraries

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

from scipy.stats import f\_oneway

# Load the dataset

file\_path = 'crop\_yield.csv'

df = pd.read\_csv(file\_path)

# Display the first few rows

df.head()

# Remove missing values

df.dropna(inplace=True)

# Filter dataset for a specific crop (e.g., Rice)

crop = 'Rice'

df\_crop = df[df['Crop'] == crop]

# \*\*Perform ANOVA Test:\*\* Group data by state and perform an ANOVA test on crop yield.

# Group data by state and calculate yield

states = df\_crop['State'].unique()

groups = [df\_crop[df\_crop['State'] == state]['Yield'] for state in states]

# Perform ANOVA test

f\_stat, p\_value = f\_oneway(\*groups)

print(f'F-statistic: {f\_stat}, P-value: {p\_value}')

# \*\*Visualization:\*\* Visualize the distribution of crop yields across states using a boxplot.

plt.figure(figsize=(12, 6))

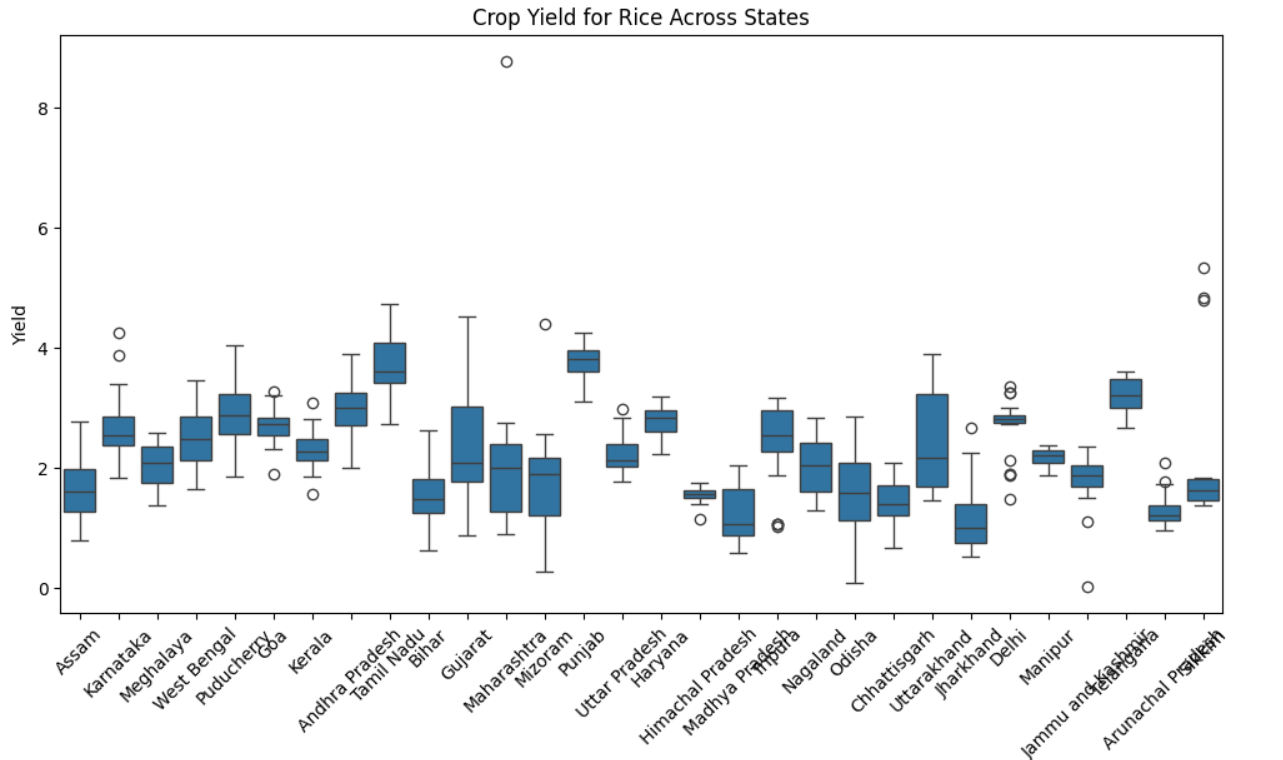
sns.boxplot(x='State', y='Yield', data=df\_crop)

plt.title(f'Crop Yield for {crop} Across States')

plt.xticks(rotation=45)

plt.show()

**Output:**



**Conclusion**

If the p-value is \*\*less than 0.05\*\*, it indicates that the differences in crop yield across states are statistically significant.

If the p-value is \*\*greater than 0.05\*\*, it indicates that the differences in crop yield across states are not statistically significant.

The boxplot helps visualize the variation in crop yields across different states, providing a clear understanding of yield distribution.