# Student Performance Prediction Report

Name: Danish Ahmad

Roll No.: 202401100400074

# Introduction

This report explores student performance prediction using machine learning techniques. The dataset includes study hours, sleep hours, attendance, previous scores, and extracurricular activities to predict the final exam score. The goal is to understand the key factors influencing student performance through exploratory data analysis (EDA).

## Methodology

1. \*\*Dataset\*\*: A synthetic dataset containing student performance factors was used.  
2. \*\*Exploratory Data Analysis (EDA)\*\*: Visualizations and statistics were used to identify trends.  
3. \*\*Data Cleaning\*\*: Checked for missing values and outliers.  
4. \*\*Visualization Techniques\*\*: Histograms, scatter plots, and heatmaps were used to analyze relationships.  
5. \*\*Findings\*\*: Study hours, attendance, and previous scores have a strong correlation with final scores.

## Code

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

# Load the dataset

df = pd.read\_csv("student\_performance.csv")

# Display the first few rows of the dataset

print("First 5 rows of the dataset:")

print(df.head())

# Check for missing values in the dataset

print("\nMissing values in each column:")

print(df.isnull().sum())

# Display summary statistics of the dataset

print("\nSummary statistics of numerical columns:")

print(df.describe())

# Set the style for seaborn plots

sns.set\_style("whitegrid")

# ---------------------- HISTOGRAM: Distribution of Final Scores ----------------------

plt.figure(figsize=(8, 5))

sns.histplot(df["final\_score"], bins=20, kde=True, color='blue')

plt.title("Distribution of Final Scores")

plt.xlabel("Final Score")

plt.ylabel("Count")

plt.show()

# ---------------------- CORRELATION HEATMAP ----------------------

plt.figure(figsize=(8, 6))

sns.heatmap(df.corr(), annot=True, cmap="coolwarm", fmt=".2f", linewidths=0.5)

plt.title("Feature Correlation Heatmap")

plt.show()

# ---------------------- SCATTER PLOTS ----------------------

# Study Hours vs Final Score

plt.figure(figsize=(7, 5))

sns.scatterplot(x=df["study\_hours"], y=df["final\_score"], color="red")

plt.title("Study Hours vs Final Score")

plt.xlabel("Study Hours")

plt.ylabel("Final Score")

plt.show()

# Sleep Hours vs Final Score

plt.figure(figsize=(7, 5))

sns.scatterplot(x=df["sleep\_hours"], y=df["final\_score"], color="green")

plt.title("Sleep Hours vs Final Score")

plt.xlabel("Sleep Hours")

plt.ylabel("Final Score")

plt.show()

# Attendance vs Final Score

plt.figure(figsize=(7, 5))

sns.scatterplot(x=df["attendance"], y=df["final\_score"], color="purple")

plt.title("Attendance vs Final Score")

plt.xlabel("Attendance (%)")

plt.ylabel("Final Score")

plt.show()

# Previous Score vs Final Score

plt.figure(figsize=(7, 5))

sns.scatterplot(x=df["previous\_score"], y=df["final\_score"], color="orange")

plt.title("Previous Score vs Final Score")

plt.xlabel("Previous Score")

plt.ylabel("Final Score")

plt.show()

# ---------------------- BOX PLOTS ----------------------

# Extracurricular Activities vs Final Score

plt.figure(figsize=(6, 5))

sns.boxplot(x=df["extracurricular"], y=df["final\_score"])

plt.title("Impact of Extracurricular Activities on Final Score")

plt.xlabel("Extracurricular Activities (0 = No, 1 = Yes)")

plt.ylabel("Final Score")

plt.show()

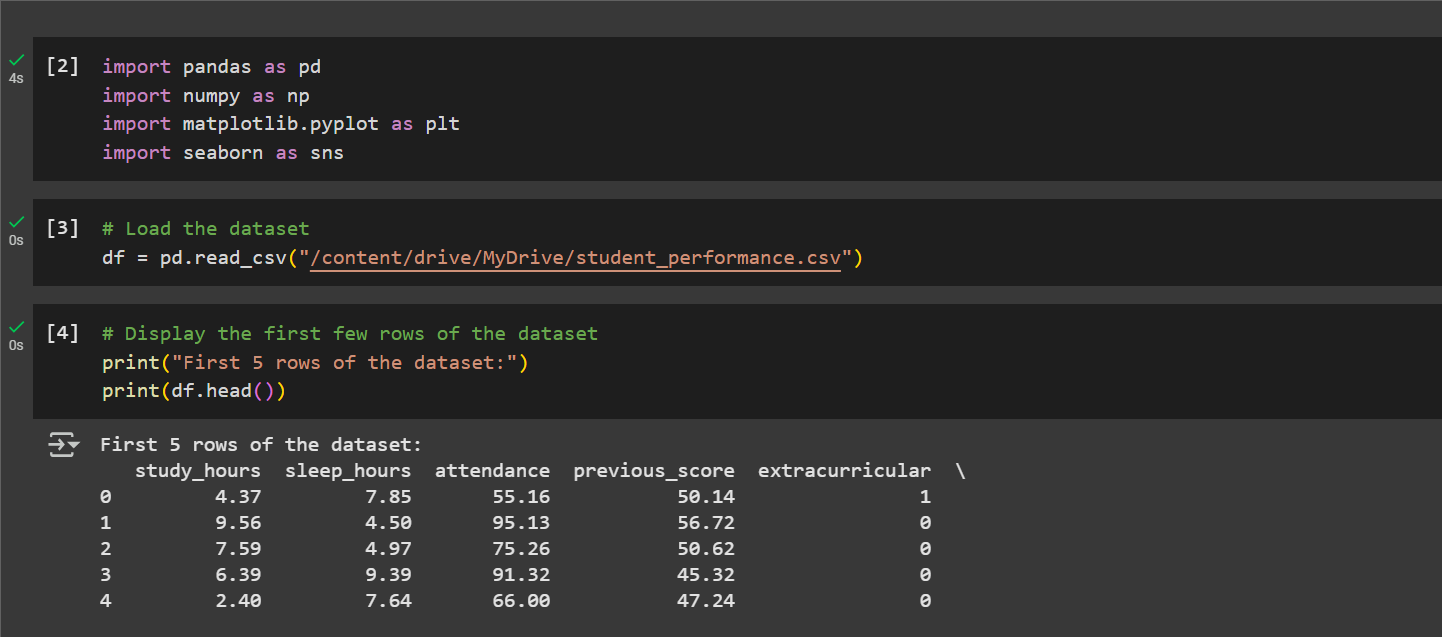
# ---------------------- PAIRPLOT ----------------------

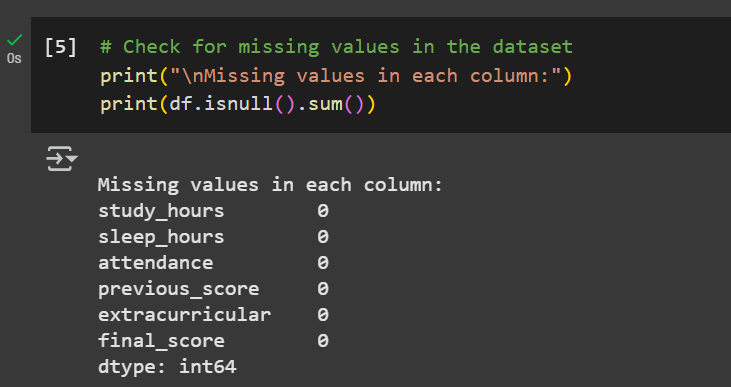
# Pairplot to analyze relationships between multiple features sns.pairplot(df)

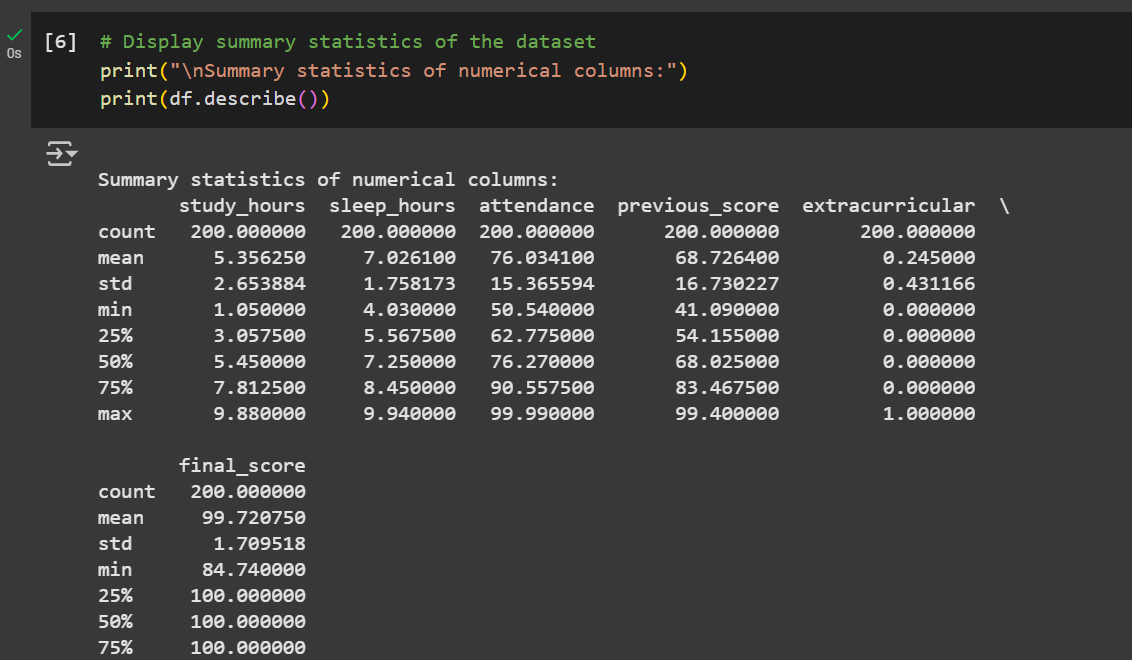
plt.show()plt.title("Distribution of Final Scores")  
plt.show()

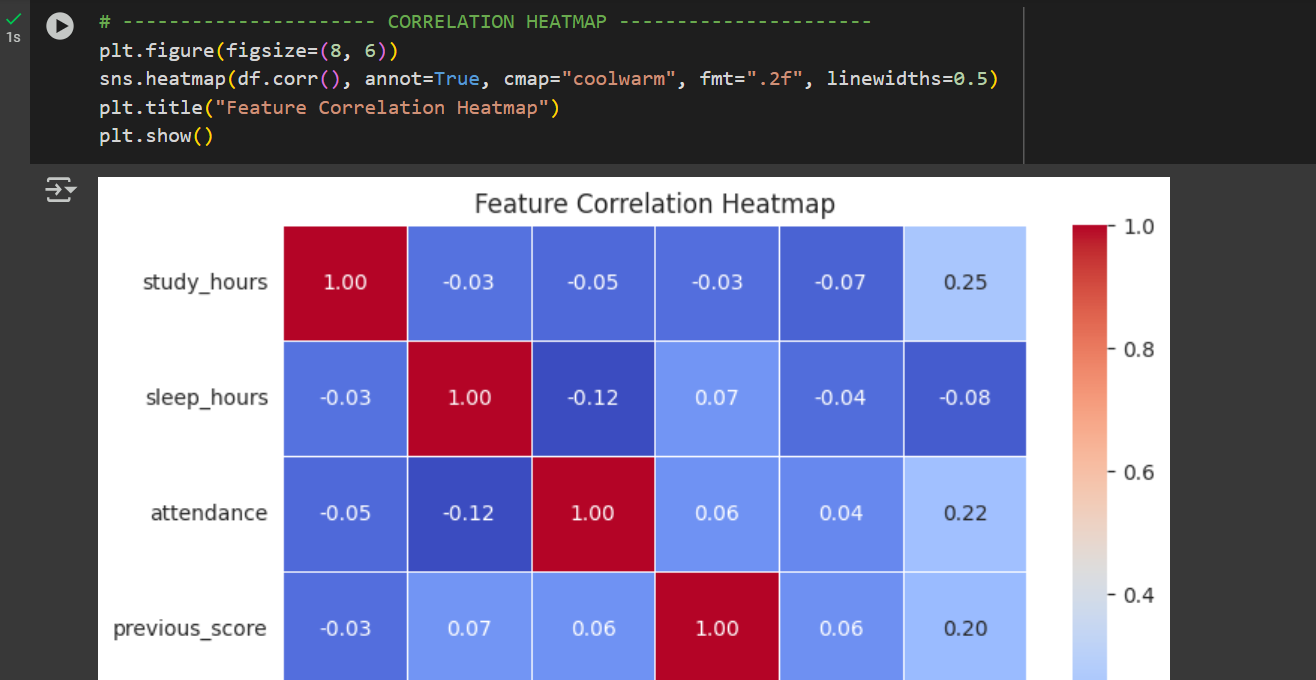
## Output/Result

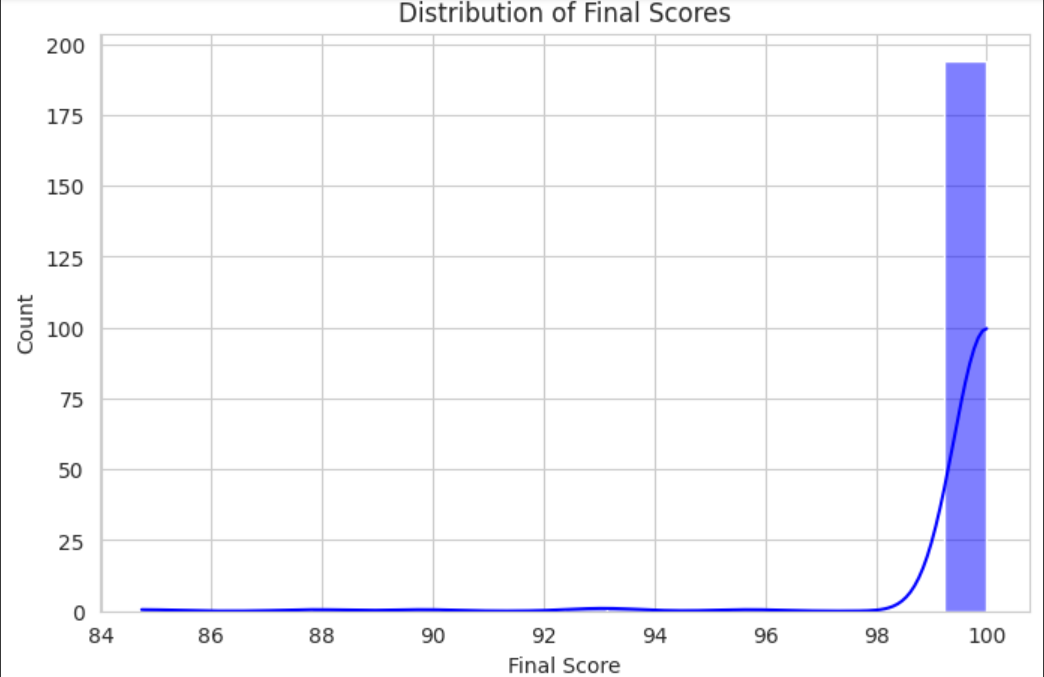
Below is a screenshot of the executed code's results, including graphs and summary statistics. These results highlight the relationships between different factors influencing student performance.

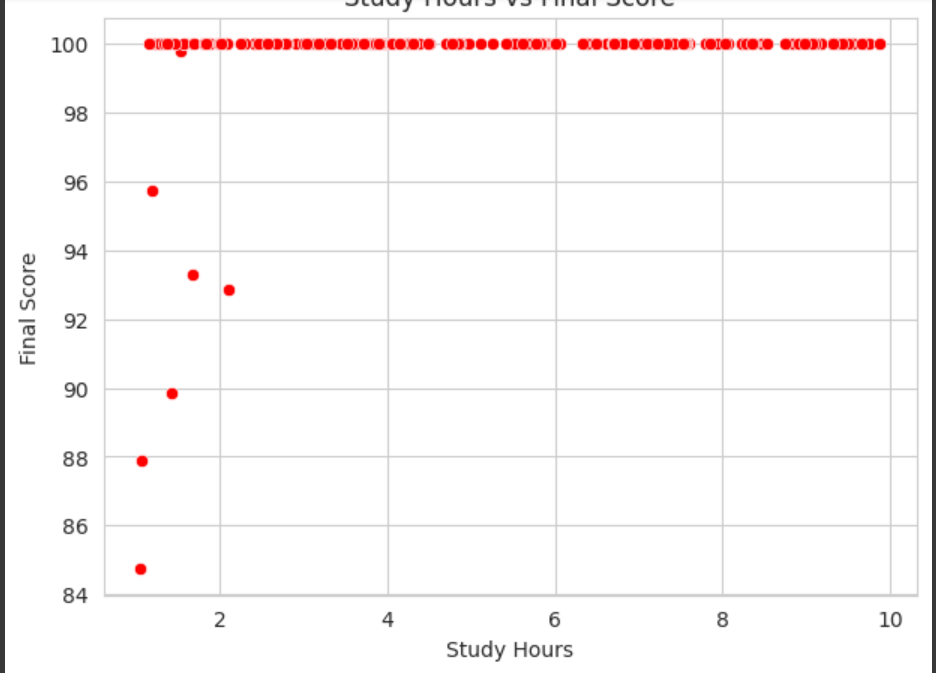


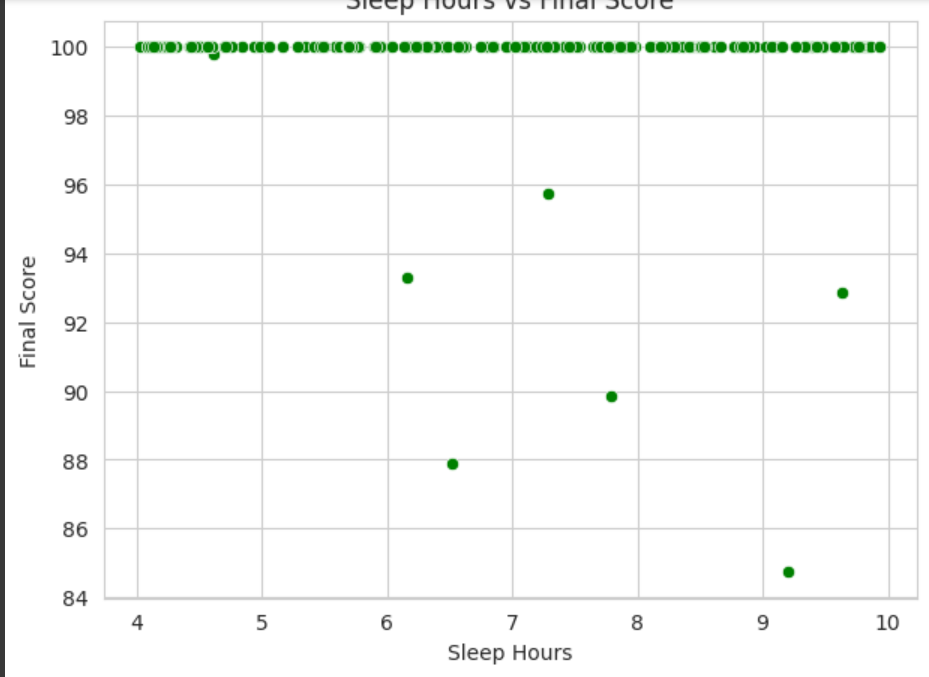












## References/Credits

1. Dataset: Synthetic dataset generated for this project.  
2. Libraries Used: Pandas, NumPy, Matplotlib, Seaborn.  
3. External References: Seaborn and Matplotlib documentation for visualization techniques.