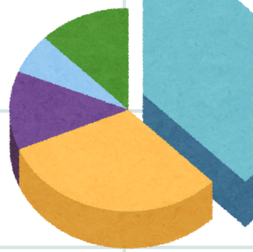
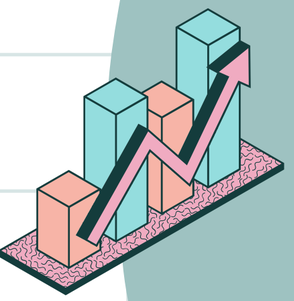
DATA ANALYSIS OF STUDENT SCORES USING PYTHON

By : Biswajeet Prasad





INTRODUCTION

We have a Student data set and we need to analyzes student performance data using Python, pandas, and seaborn. The dataset includes over 30,000+ records with information like gender, study hours, parental education, and test scores in math, reading, and writing. The goal is to clean the data, handle missing values, and visualize key patterns to understand which factors impact student achievement.

🧹 DATA CLEANING & PREPROCESSING

Using pandas, we cleaned the dataset by:

- Dropping unnecessary columns (Unnamed: 0)
- Replacing or removing missing values (fillna, dropna, median)
- Standardizing categorical entries (e.g., replacing "05-Oct" with "5-10" in WklyStudyHours)

This step ensured that the dataset was consistent and ready for analysis.

DATA CLEANING & PREPROCESSING – CODE USED

```
# 1. Importing Required Libraries
import pandas as pd

# 2. Load the dataset
df = pd.read_csv("student_scores.csv")

# 3. Drop unnecessary index column
df.drop(columns=["Unnamed: 0"], inplace=True)

# 4. Replace incorrect value in WklyStudyHours column
df["WklyStudyHours"] = df["WklyStudyHours"].replace("05-Oct", "5-10")
```

DATA CLEANING & PREPROCESSING – CODE USED

```
# 5. Check for missing values
df.isnull().sum()

# 6. Fill missing categorical values
df["EthnicGroup"].fillna("unknown", inplace=True)

# 7. Fill missing numerical values with median
df["NrSiblings"] = df["NrSiblings"].fillna(df["NrSiblings"].median())

# 8. Drop any remaining rows with missing values (optional/if needed)
df.dropna(inplace=True)
```

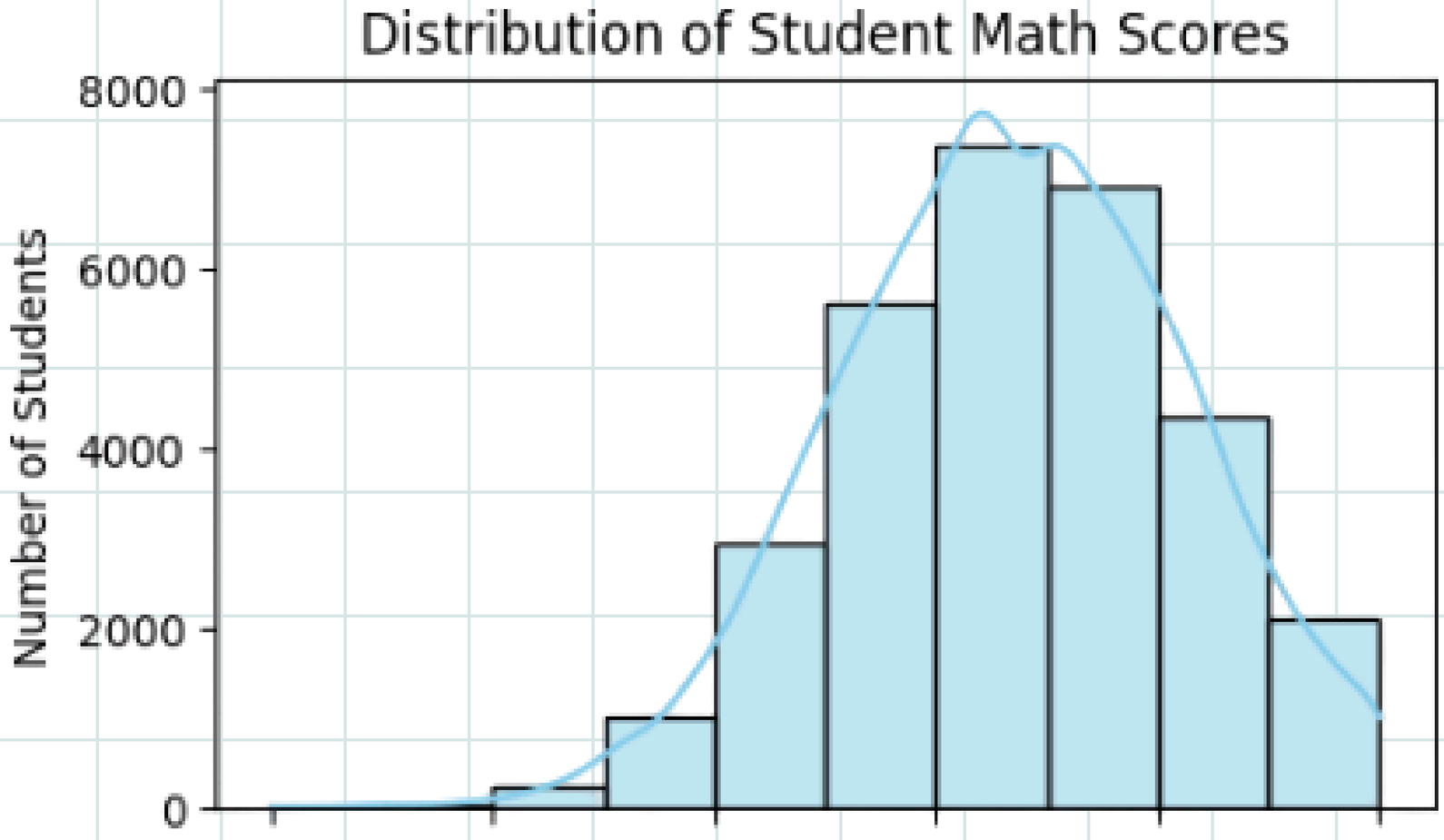
EXPLORATORY DATA ANALYSIS (EDA)

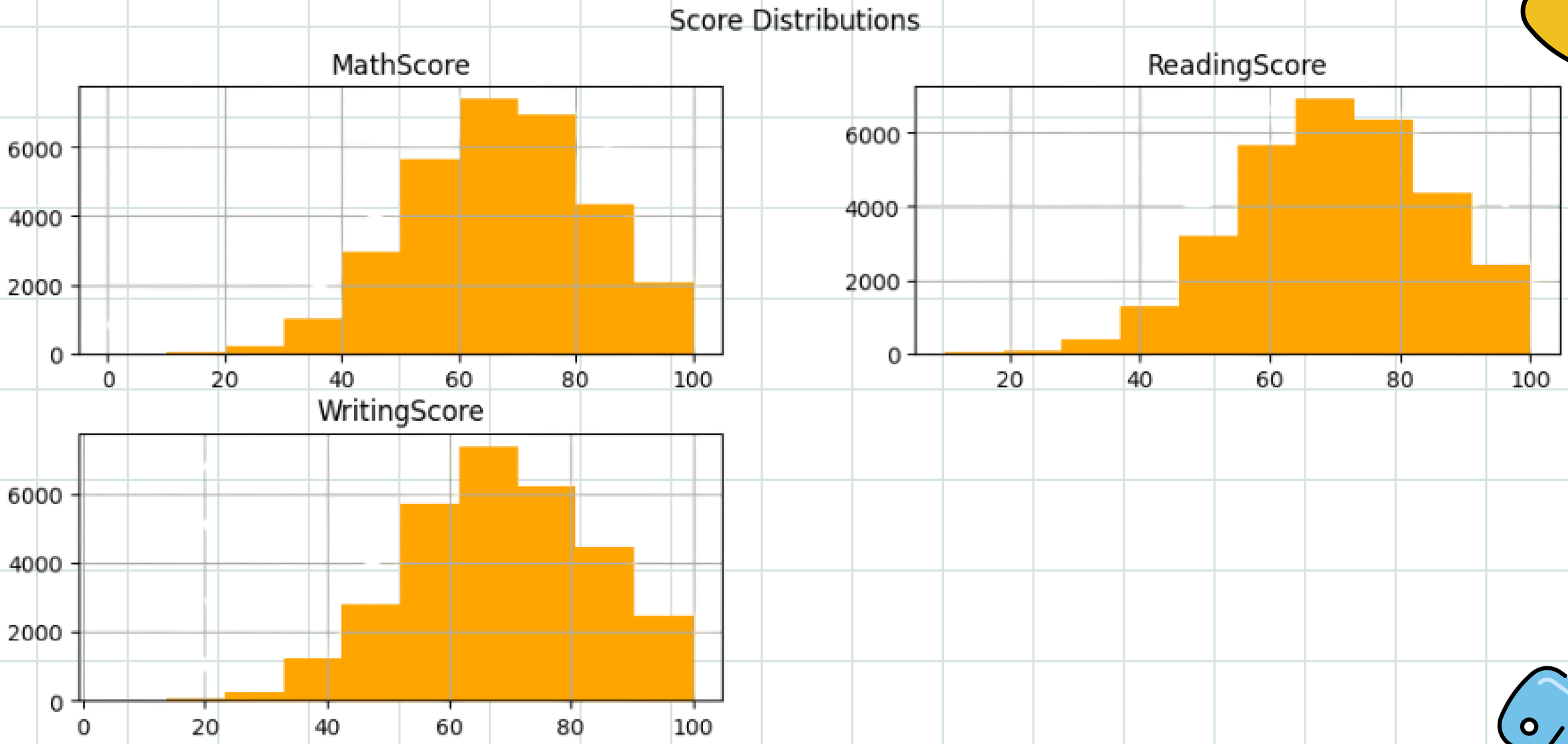
Using pandas, matplotlib, and seaborn, I explored relationships and patterns in the data.

Key insights included:

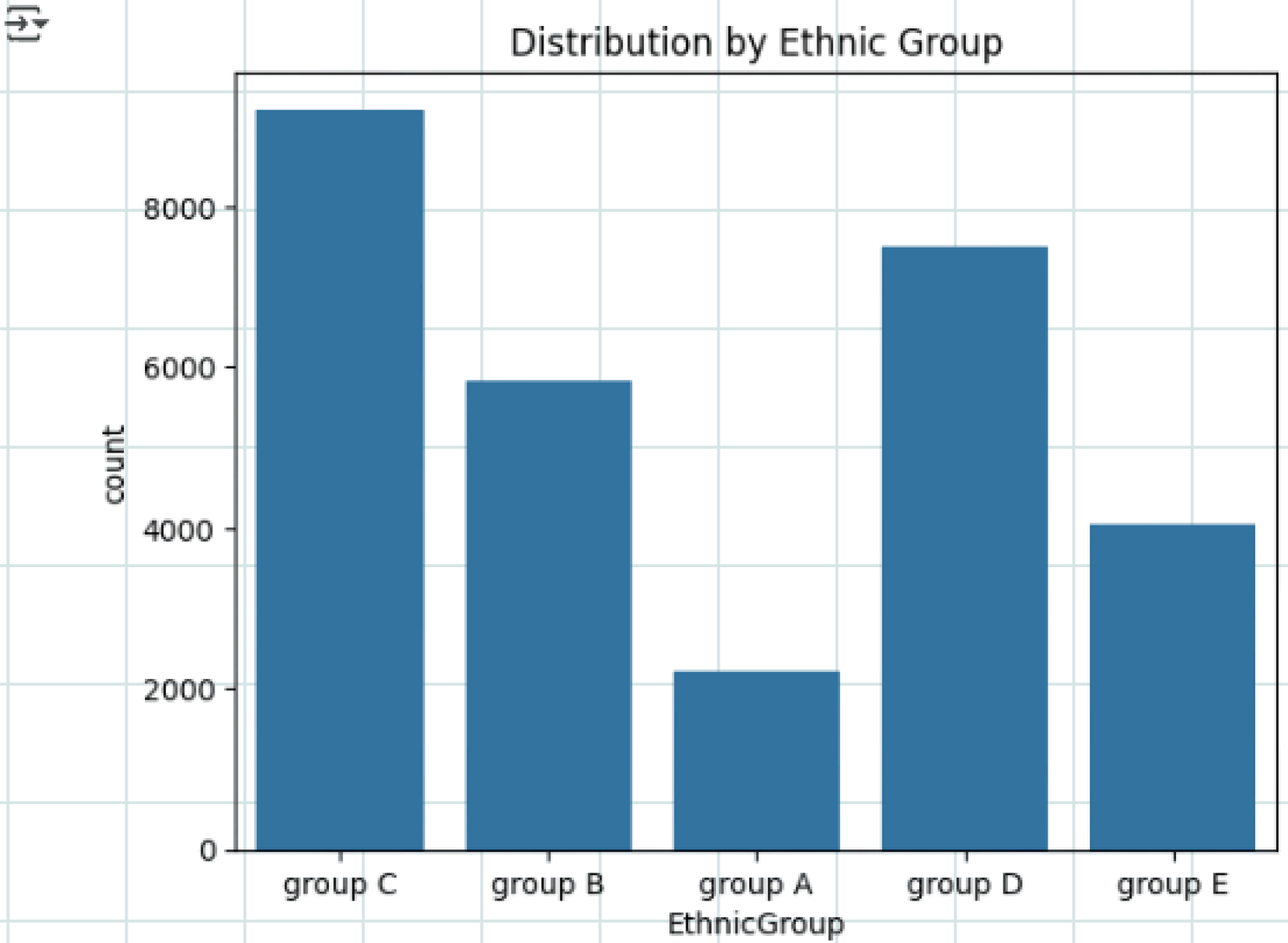
- Distribution of diffrent scores with number of student
- Distribution by Ethnic group
- Reading score by parents education
- Maths Score By Gender
- Maths Score By Study Hour

```
plt.figure(figsize=(5,3))
sns.histplot(df['MathScore'], kde=True, bins=10, color='skyblue')
plt.title("Distribution of Student Math Scores")
plt.xlabel("Score")
plt.ylabel("Number of Students")
plt.show()
```

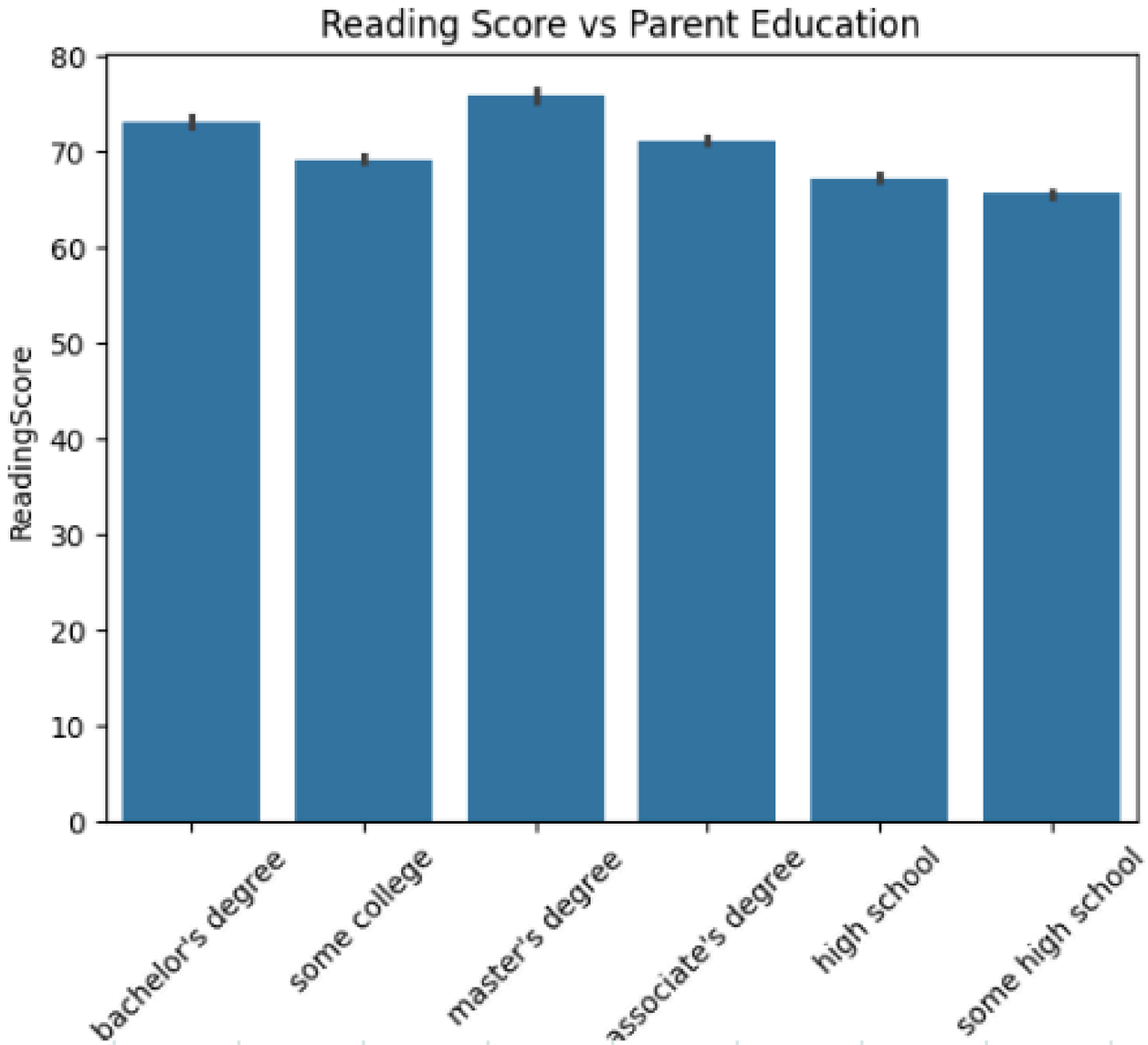


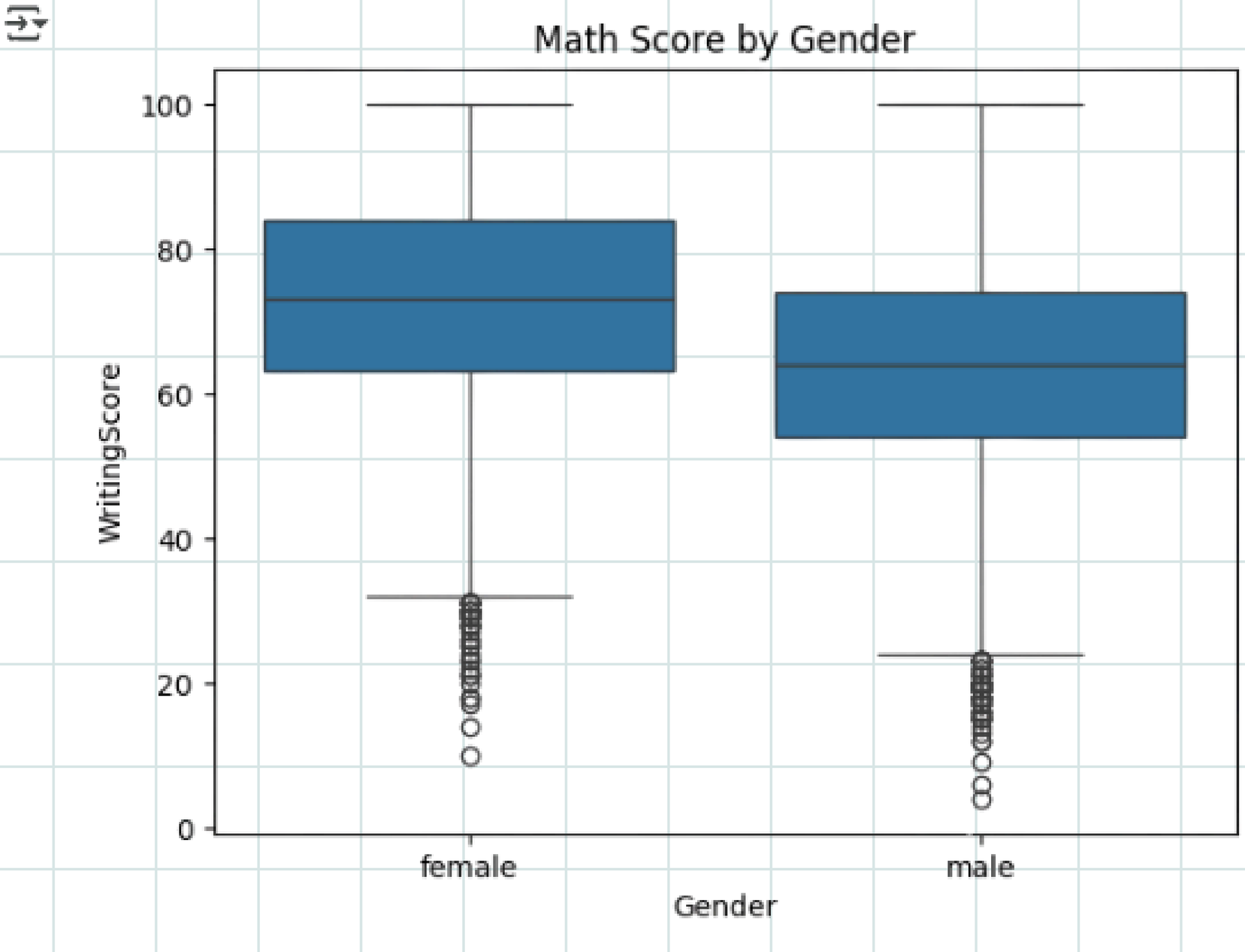

Score Distributions
MathScore
ReadingScore
WritingScore
0
2000
4000
6000
0
20
40
60
80
100

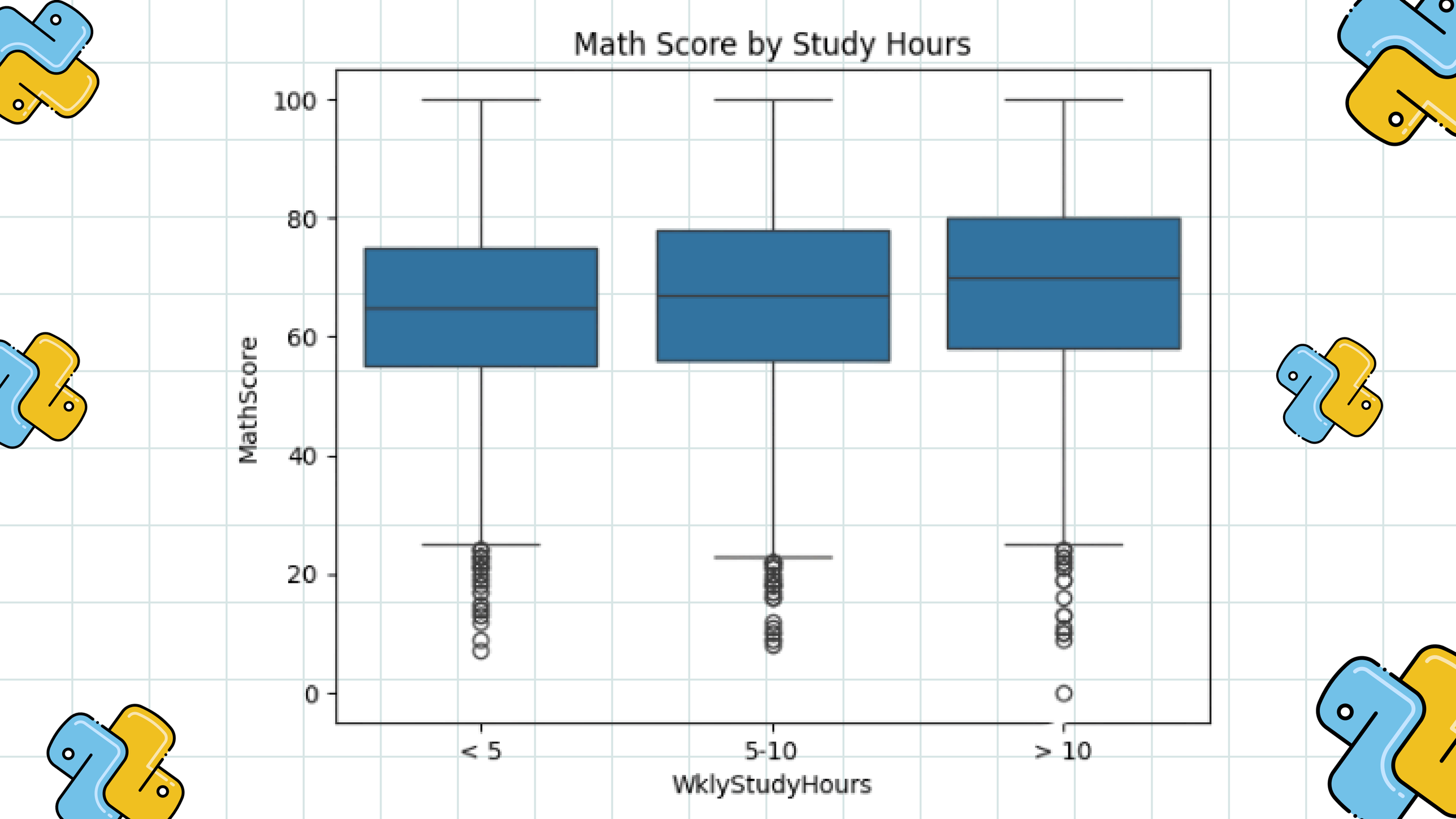
```
[61] sns.countplot(x='EthnicGroup', data=df)
     plt.title("Distribution by Ethnic Group")
     plt.show()
```

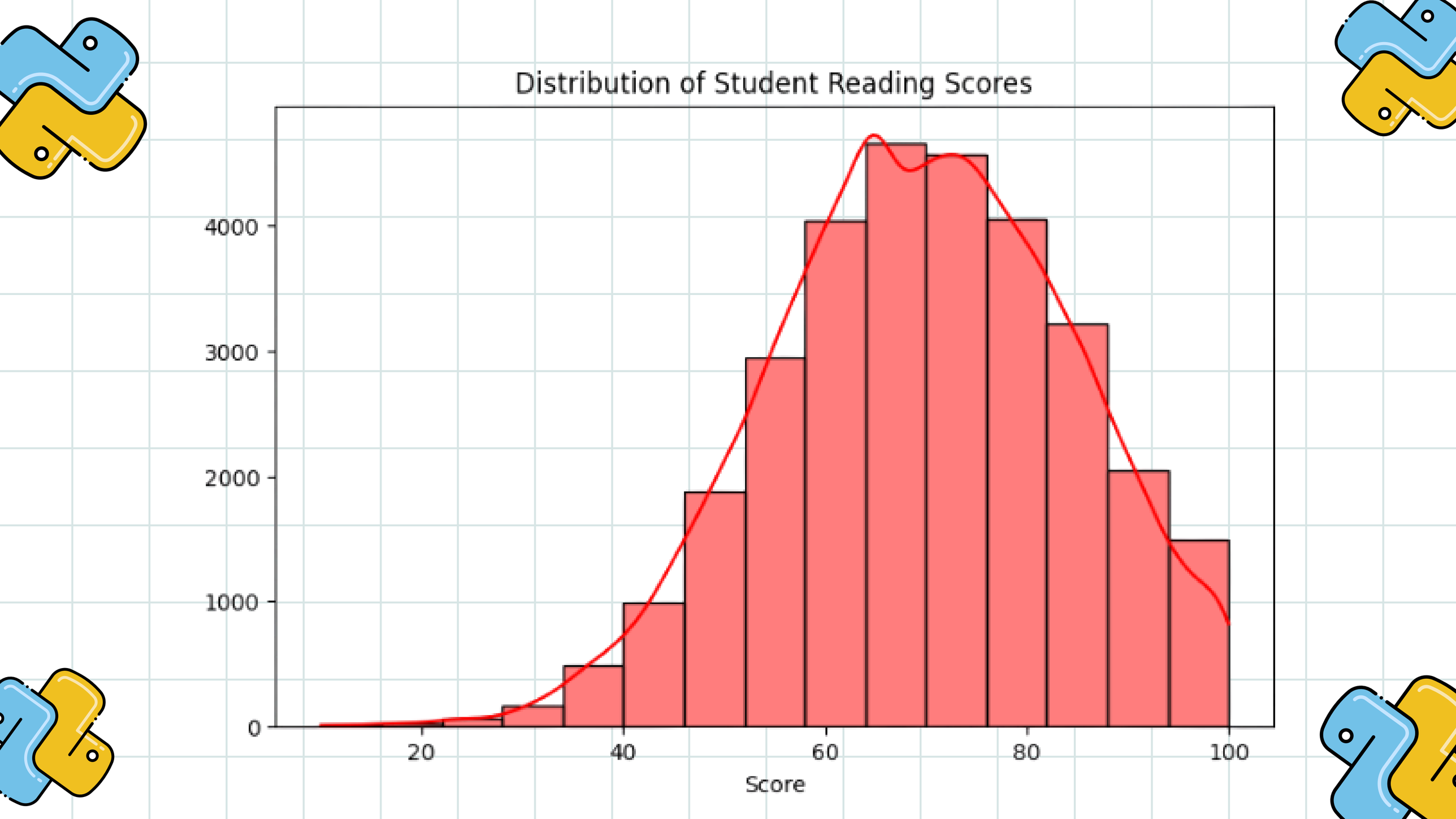


```
sns.barplot(x='ParentEduc', y='ReadingScore', data=df)
plt.xticks(rotation=45)
plt.title("Reading Score vs Parent Education")
plt.show()
```




Math Score by Gender
WritingScore
100
80
60
40
20
0
female
male
Gender


Math Score by Study Hours
MathScore
100
80
60
40
20
0
< 5
5-10
> 10
WklyStudyHours


Distribution of Student Reading Scores
4000
3000
2000
1000
0
20
40
60
80
100
Score

SUMMARY OF FINDINGS

Insight	Observation
📘 Test Preparation	Students who completed test prep scored significantly higher
⏱️ Study Hours	More study hours led to better scores, especially beyond 10 hours/week.
🧮 Subject Correlation	Math, Reading, and Writing scores are strongly correlated
👨‍👩‍👧 Parental Education	Higher parental education is associated with better student performance
👩 Gender Differences	Females slightly outperform over males.
👫 Ethnic Group	Performance varies slightly across ethnic groups, possibly due to access/resources

INSIGHTS THAT MATTER & THE WAY FORWARD

🧠 Smarter Prep, Stronger Performance

- Test preparation and consistent study routines clearly boost scores — structured learning pays off.

🏫 Background Matters, But So Does Support

- Students with less academic family support can excel — if given the right guidance and encouragement.

📊 Data Shows Direction

- High inter-subject correlation suggests a need for balanced skill-building, not just subject-specific focus.

🌍 Equity Isn't Optional

- Slight performance gaps across demographics call for inclusive strategies — equity must be embedded, not added.

🚀 ***What Next? Empower, Personalize, Repeat***

- Expand access to prep programs and peer mentoring
- Promote smart study habits early
- Engage families through outreach
- Personalize learning for diverse needs

THANK YOU!

LET'S CONNECT

WWW.LINKEDIN.COM/IN/DATAWITHBISWAJEET

WWW.GITHUB.COM/DATAWITHBISWAJEET

DATAWITHBISWAJEET@GMAIL.COM