
Student Performance Analysis

Data-Driven Exploration of Factors Affecting Exam Scores

Presented by : Jobina K v

Introduction

- **This project analyzes a student performance dataset using Python, NumPy, and Pandas.**
- **The dataset includes students' demographics, parental education, lunch type, test preparation, and exam scores.**
- **NumPy is used to perform fast numerical operations like calculating totals and averages of Math, Reading, and Writing scores.**

Purpose is :

- **Explore score distributions and relationships.**
- **Compute quick statistics to reveal factors affecting performance.**

Problem Statement

- **Educational institutions need to understand how background factors affect student performance.**
- **Key factors include :**
 - **Gender**
 - **Parental education**
 - **Lunch type**
 - **Test preparation course.**
- **The challenge is to identify patterns in the data.**
- **Insights will help design strategies to improve student outcomes.**

Proposal Solution

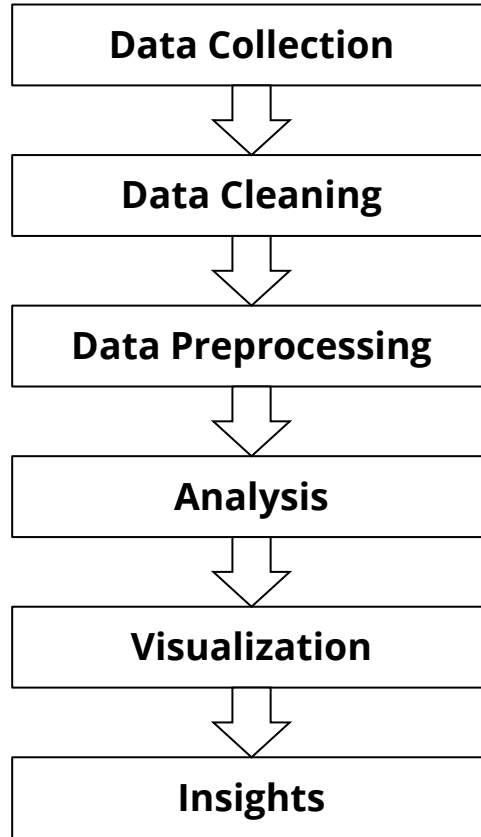
- **Analyze student performance data using NumPy and Pandas.**
- **Calculate total and average scores for Math, Reading, and Writing.**
- **Examine effects of Gender, Parental Education, Lunch Type, and Test Preparation.**
- **Visualize insights with bar charts to identify patterns and guide improvements.**

Dataset Overview & Structure

- The Students Performance dataset includes records of 50 with details about Gender, Lunch, Math Score, Reading Score, Writing Score.
- Purpose is Compute average scores, standard deviation, and compare gender wise performance.

```
Data columns (total 8 columns):  
#      Column                               Non-Null Count  Dtype  
---  -  
0     gender                               50 non-null     object  
1     race/ethnicity                       50 non-null     object  
2     parental level of education          50 non-null     object  
3     lunch                                 50 non-null     object  
4     test preparation course              50 non-null     object  
5     math score                           50 non-null     int64  
6     reading score                        50 non-null     int64  
7     writing score                         50 non-null     int64  
dtypes: int64(3), object(5)  
memory usage: 3.3+ KB
```

Workflow



Tools Used

- **Software:** Python, Google Colab
- **Libraries:** pandas, matplotlib, NumPy

Implementation

Step 1: Selected the first 50 student records from the dataset for analysis.

Step 2: Checked data types, missing values, and duplicates to ensure data quality.

Step 3: Converted Math, Reading, and Writing scores into NumPy arrays.

Step 4: Calculated total and average scores for each student and each subject.

Step 5: Performed grouping and analysis by Gender, Parental Education, Lunch Type, and Test Preparation Course.

Step 6: Visualized results using bar plots :

- **Average scores by Gender**
- **Average scores by Parental Education**
- **Average scores by Lunch Type**
- **Average scores by Test Preparation Course**

Data Info

Data columns (total 8 columns):

#	Column	Non-Null Count	Dtype
0	gender	50 non-null	object
1	race/ethnicity	50 non-null	object
2	parental level of education	50 non-null	object
3	lunch	50 non-null	object
4	test preparation course	50 non-null	object
5	math score	50 non-null	int64
6	reading score	50 non-null	int64
7	writing score	50 non-null	int64

dtypes: int64(3), object(5)

memory usage: 3.3+ KB

Statistics Summary

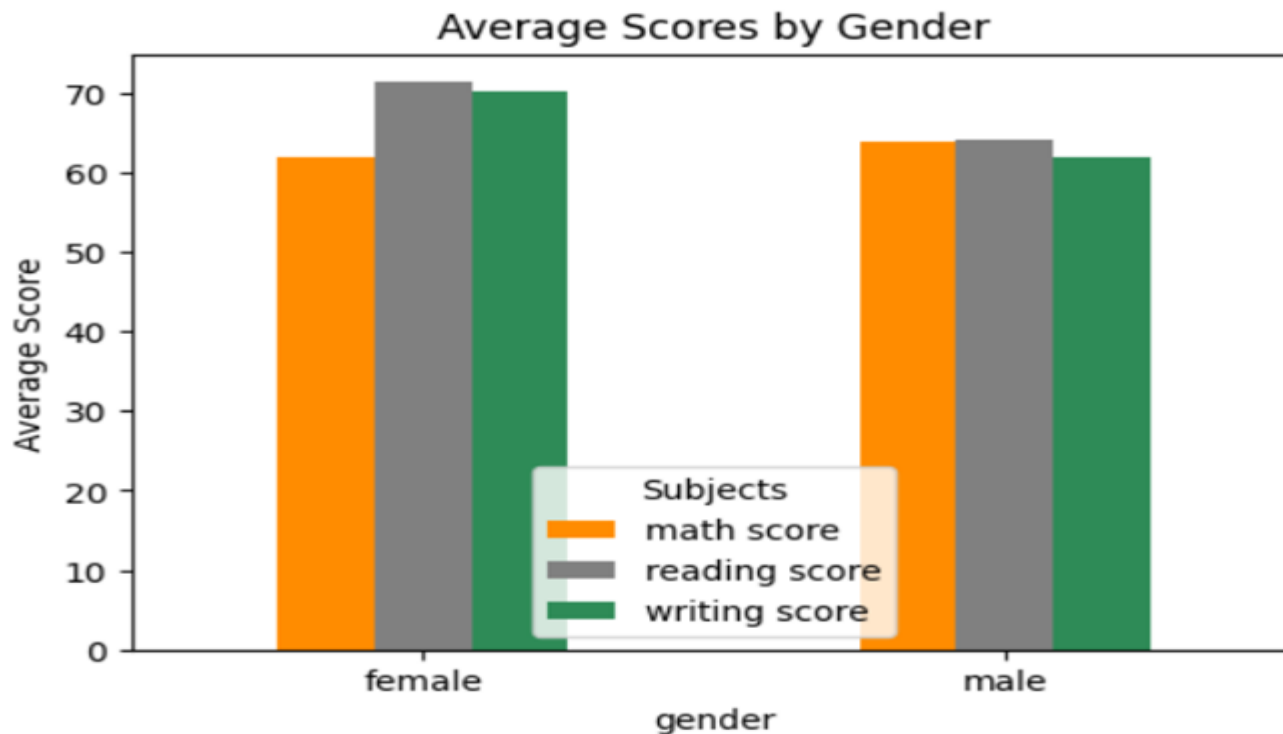
	math score	reading score	writing score
count	50.000000	50.000000	50.000000
mean	62.840000	67.940000	66.440000
std	15.163544	14.261995	14.785763
min	18.000000	32.000000	28.000000
25%	54.250000	57.250000	57.000000
50%	65.000000	70.000000	67.500000
75%	71.750000	75.000000	76.000000
max	97.000000	95.000000	93.000000

Data Cleaning

	0
gender	0
race/ethnicity	0
parental level of education	0
lunch	0
test preparation course	0
math score	0
reading score	0
writing score	0
dtype: int64	

```
df.duplicated().sum()  
np.int64(0)
```

Average Scores by Gender

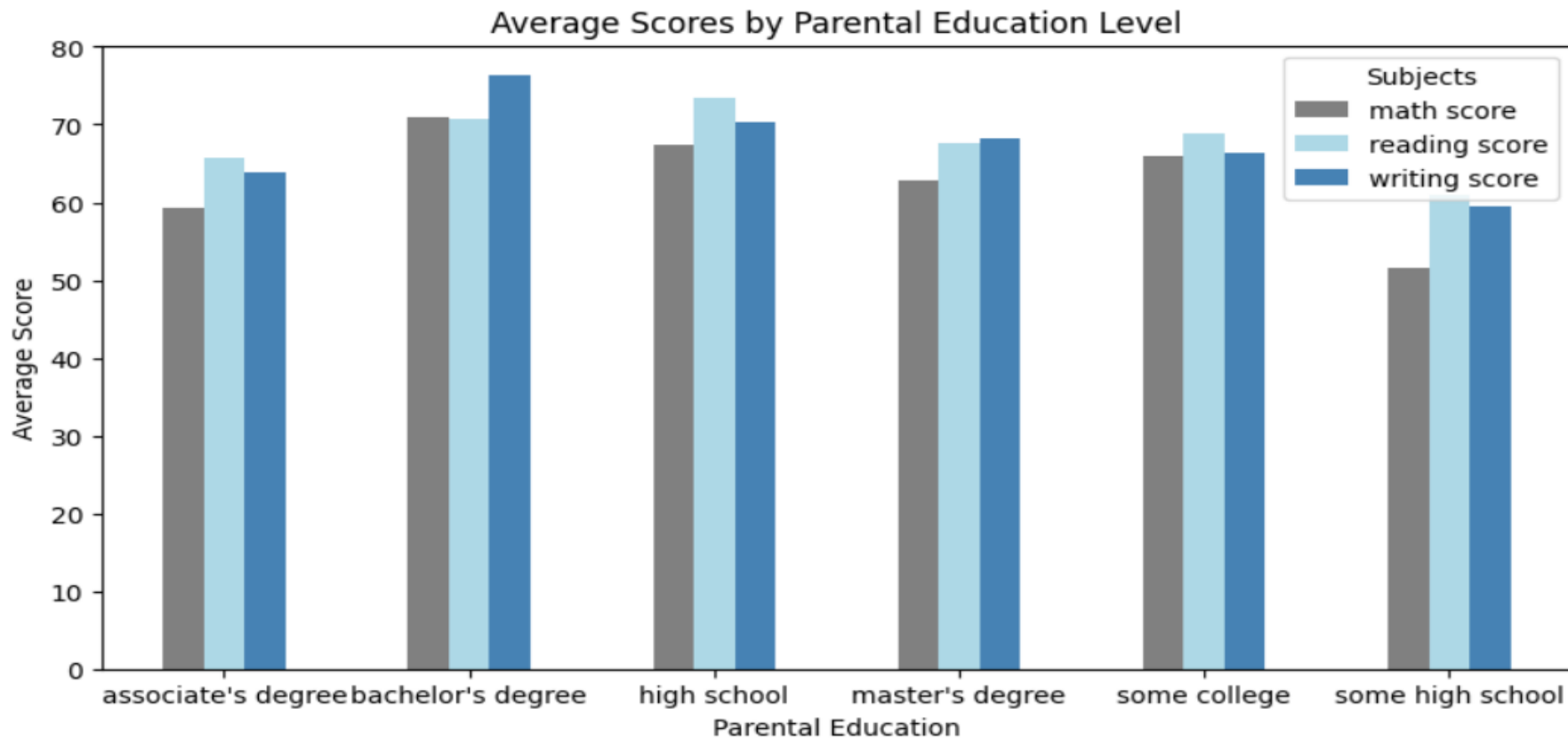


- **This Bar chart showing average Math, Reading, and Writing scores for Male and Female students.**

Key Insight :

- **Female students have higher average scores across all subjects compared to males.**
- **Highlights gender differences in academic performance.**

Average Scores by Parental Education Level

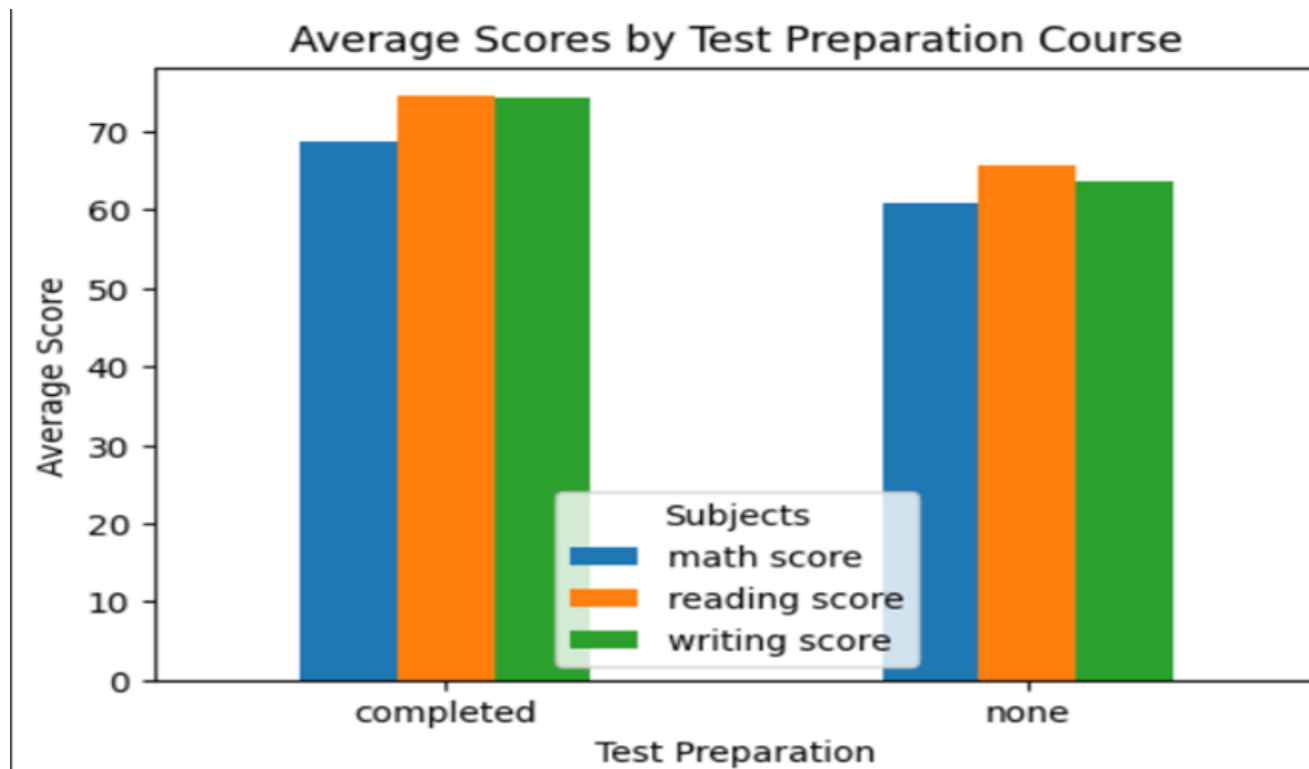


- **This Bar chart showing average Math, Reading, and Writing scores grouped by parental education.**
- **It Showing the Relationship between parental education and student performance.**

Key Insight :

- **Students with parents who have higher education (Bachelor's or Master's) usually get better scores.**
- **Parental education seems to affect how well students perform in school.**
- **Extra support can be given to students whose parents have lower education levels to help improve their performance.**

Average Scores by Test Preparation Course

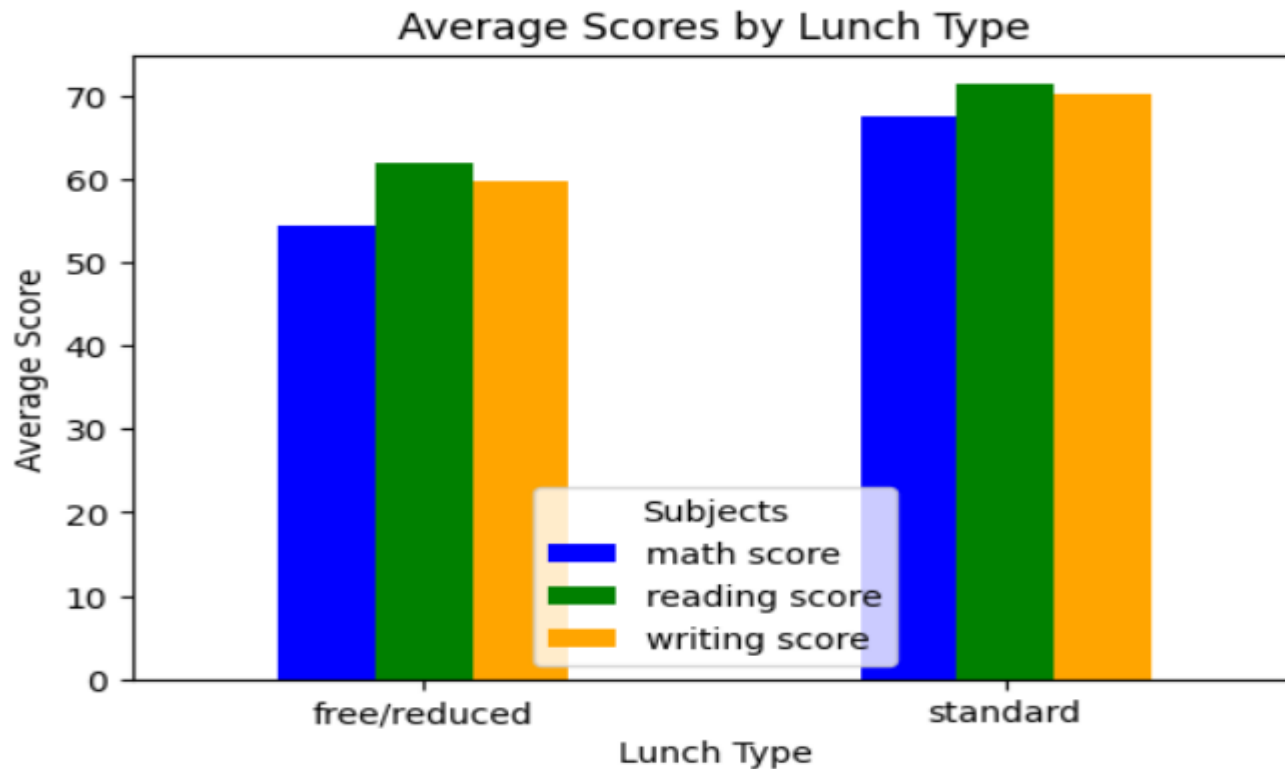


- **This Bar chart showing average scores for students who completed or did not complete the test preparation course.**
- **It Showing the impact of completing the test preparation course on student performance.**

Key Insight :

- **Students who completed the course generally perform better in all subjects.**
- **Suggests that test preparation courses positively affect performance.**
- **Schools can encourage more students to complete test preparation courses.**

Average Scores by Lunch Type

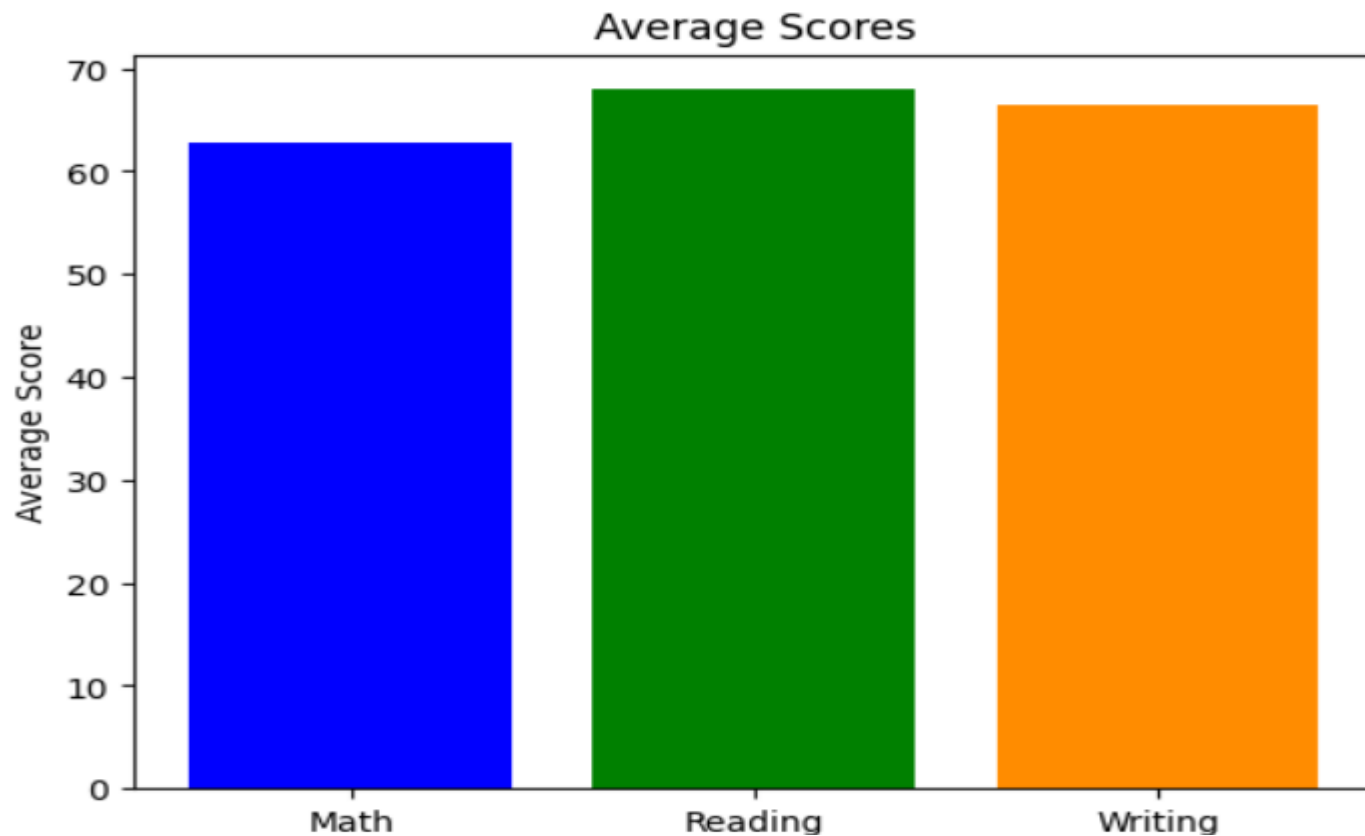


- **This Bar chart showing average scores for standard lunch vs free/reduced lunch.**
- **It Showing the Influence of lunch type on academic performance.**

Key Insights :

- **Students with standard lunch generally perform slightly better than those with free/reduced lunch.**
- **Can guide schools to provide better support for students with free/reduced lunch.**

Average Scores by Subject



- **This Bar chart of average Math, Reading, and Writing scores.**
- **This Showing the overall average performance of students in each subject.**

Key Insight :

- **Reading has the highest average score.**
- **Writing is slightly lower than Reading.**
- **Math has the lowest average among the three subjects.**
- **Helps identify areas where students may need more focus or support.**
- **It clearly shows which subject students perform best in and which subject has lower average marks.**

Result

- **Test preparation course helps students score higher in all subjects.**
- **Students with standard lunch score slightly better than those with free/reduced lunch.**
- **Students with more educated parents tend to score higher.**
- **Female students perform slightly better than male students on average.**
- **Students perform best in Reading, followed by Writing, and slightly lower in Math.**

Conclusion

- **This analysis helped us understand the main factors affecting student performance in exams.**
- **In this dataset we found that:**
 - **Test preparation course and parental education are key factors linked to higher scores.**
 - **Gender and lunch type show some influence but are less important.**
 - **Scores across Math, Reading, and Writing are positively correlated.**
- **These findings can help students, teachers, and parents focus on proper preparation, supportive learning environments, and targeted guidance to improve academic performance.**

Reference

- **Students Performance dataset – Kaggle**
- **Python & Library References – Pandas, Matplotlib, NumPy**

THANK YOU