

STUDENT PERFORMANCE DASHBOARD

TEAM MEMBERS:

- MOHAMED WAEL MOHAMED
- NOURHAN MOHAMED AHMED
- MOHAMED ATEF IBRAHEM IBRAHEM
- SEIF MOHAMED HELMY AHMED

TABLE OF CONTENTS

INTRODUCTION	3
TASK_1.....	4
TASK_2.....	5
TASK_3.....	6
CONCLUSION.....	7

INTRODUCTION

This project focuses on analyzing student performance data to uncover meaningful patterns and insights. The primary objective is to support teachers and school administrators in understanding the factors that influence academic outcomes. By cleaning the dataset, applying SQL queries, and building interactive visualizations, the project highlights trends such as the relationship between attendance and GPA, differences across majors, and performance distribution by gender.

Ultimately, the Student Performance Dashboard provides a data-driven approach to monitor academic progress, identify at-risk students, and recommend strategies for improving overall performance.



TASK 1 – DATA PREPROCESSING

The first step in the project was to prepare and clean the dataset to ensure accuracy and consistency. The preprocessing tasks included:

- **Duplicate Removal:** Checked and removed duplicate records to avoid bias in analysis.
- **Missing Values:** Verified that no critical data was missing in the dataset.
- **New Feature Creation:** A new column `PerformanceCategory` was added to classify students based on their GPA:
 - **High Performance:** $GPA \geq 3.5$
 - **Medium Performance:** $2.0 \leq GPA < 3.5$
 - **Low Performance:** $GPA < 2.0$

```
#Checking for duplicates on rows
print(df.duplicated().any())
print(df.duplicated().sum())
False
0
df = df.drop_duplicates()

#Checking for the null values
df.isna().values.any()
```

This preprocessing step ensured that the data was clean, reliable, and ready for analysis in the next stages of the project.

```
categorize performance ( High/Medium/Low)
def categorize_performance(gpa):
    if gpa >= 3.5:
        return "High"
    elif gpa >= 2.0:
        return "Medium"
    else:
        return "Low"

df["PerformanceCategory"] = df["GPA"].apply(categorize_performance)

df.to_csv("student_data_with_performance.csv", index=False)

df.head(10)
```

StudentID	Gender	Age	StudyHoursPerWeek	AttendanceRate	GPA	Major	PerformanceCategory
0	1	Male	24	37	90.75	3.47	Arts
1	2	Female	22	37	74.90	2.32	Education

TASK 2 – SQL QUERIES

To explore and analyze the dataset more effectively, the data was imported into an SQL database. Several queries were executed to extract key insights:

1. Top Students by GPA

- Selected the top 10 students with the highest GPA.
- Helped identify the most outstanding performers.

2. Average GPA by Major

- Calculated the average GPA for each major.
- Highlighted which majors had stronger or weaker academic performance.

3. Attendance Trends

- Grouped students into categories based on their attendance rate:
 - High Attendance ($\geq 80\%$)
 - Medium Attendance ($50\% - 79\%$)
 - Low Attendance ($< 50\%$)
- Revealed the strong link between attendance and overall performance.

4. Average GPA by Gender

- Compared the average GPA of male and female students.
- Showed slight differences in performance across genders.

These SQL queries provided structured insights into the dataset and laid the foundation for the visualization stage.

```
set top performers by subject
query_top = """
SELECT StudentID, Gender, Age, GPA, Major, PerformanceCategory
FROM Students
ORDER BY GPA DESC
LIMIT 10;
"""

top_students = pd.read_sql(query_top, conn)
top_students
```

	StudentID	Gender	Age	GPA	Major	PerformanceCategory
0	30	Female	24	3.99	Science	High
1	181	Male	19	3.99	Business	High

```
query_avg = """
SELECT Major, ROUND(AVG(GPA),2) as Avg_GPA
FROM Students
GROUP BY Major
ORDER BY Avg_GPA DESC;
"""

avg_gpa_major = pd.read_sql(query_avg, conn)
avg_gpa_major
```

	Major	Avg_GPA
0	Business	3.04
1	Arts	3.01
2	Engineering	2.97
3	Science	2.96
4	Education	2.94

```
query_attendance = """
SELECT
CASE
    WHEN AttendanceRate >= 80 THEN 'High Attendance'
    WHEN AttendanceRate >= 50 THEN 'Medium Attendance'
    ELSE 'Low Attendance'
END as AttendanceCategory,
COUNT(*) as StudentCount
FROM Students
GROUP BY AttendanceCategory;
"""

attendance_trends = pd.read_sql(query_attendance, conn)
attendance_trends
```

	AttendanceCategory	StudentCount
0	High Attendance	193
1	Medium Attendance	307

TASK_3

VISUALIZATIONS

To better understand the data, multiple visualizations were created to highlight key trends and patterns:

1. GPA Distribution

- A histogram was plotted to show how GPAs are distributed among students.
- This helped identify the general academic level of the student body.

2. Average GPA by Major

- A bar chart compared the average GPA across different majors.
- Showed which majors had higher or lower performance levels.

3. Attendance vs GPA

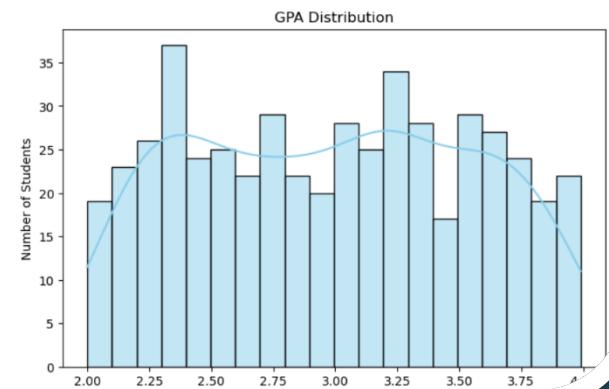
- A scatter plot displayed the relationship between attendance rate and GPA.
- Demonstrated a positive correlation: higher attendance generally resulted in higher GPA.

4. Performance by Gender

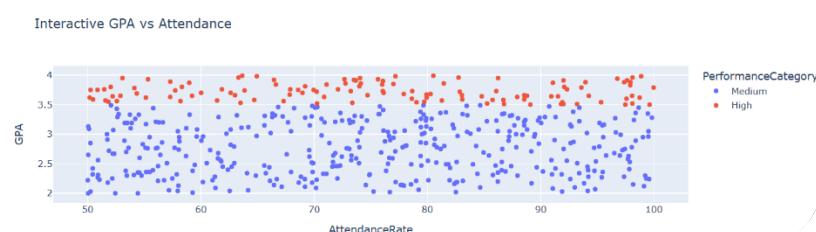
- A count plot was used to compare male and female students across performance categories (High, Medium, Low).
- Helped reveal differences in performance distribution between genders.

These visualizations provided clear and accessible insights into the dataset, making it easier to interpret patterns and communicate results.

```
fig = px.histogram(df['GPA'], bins=20, kde=True, color='skyblue')
plt.title('GPA Distribution')
plt.xlabel('GPA')
plt.ylabel('Number of Students')
plt.show()
```



```
fig = px.scatter(df, x='AttendanceRate', y='GPA',
                 color='PerformanceCategory',
                 hover_data=['StudentID', 'Major'],
                 title='Interactive GPA vs Attendance')
fig.show()
```



CONCLUSION

This project demonstrated how student performance data can be analyzed and visualized to uncover meaningful insights. The Student Performance Dashboard allows teachers and administrators to make data-driven decisions that directly support student success.

TAKEAWAY #1

Students with higher attendance rates consistently achieved better GPAs, confirming that attendance is not just a routine metric but a strong predictor of academic success. Improving attendance policies and motivating students to attend classes regularly can directly enhance overall performance levels.

TAKEAWAY #2

Female students showed slightly higher GPAs on average compared to male students, highlighting subtle gender-based differences in academic achievement. In addition, performance varied significantly across different majors, indicating that some fields of study may require additional academic support or tailored learning strategies to close the gap.

TAKEAWAY #3

The dashboard proved to be a practical decision-support tool for educators and administrators. By monitoring key indicators such as GPA, attendance, and performance categories, schools can identify at-risk students at an early stage. This enables timely interventions, targeted mentoring, and strategic planning to improve overall academic outcomes and ensure better student success rates.