Student Performance Analysis

Project Goal

Take a student survey dataset → clean it → analyze it → build simple dashboards in **Excel, Power BI, and Tableau**.

Main Question:
Identify how **attendance, homework habits, and teacher–parent communication** influence student exam performance. Provide insights to help teachers and administrators support struggling students.

🔗 Dataset Source: Kaggle – Student Performance & Attendance Dataset

---

0. Scope

What drives exam scores?

- **Attendance** – Are students who attend more classes scoring higher?

- **Homework** – Does homework completion (and guardian signature) link to better scores?

- **Communication** – Do students with higher teacher–parent engagement perform differently?

- **Subject/Grade differences** – Which subjects/grades show the biggest performance gaps?

---

1. Datasets

File	Purpose	Key Columns
students.csv	**Who** – Student demographics	student_id, full_name, grade_level
attendance.csv	**Presence** – Daily/subject-wise attendance	student_id, date, subject, attendance_status
homework.csv	**Engagement** – Homework completion & parent signatures	student_id, subject, status, grade_feedback, guardian_signature
performance.csv	**Results** – Exam scores & homework completion %	student_id, subject, exam_score, homework_completion_%
teacher_parent_communication.csv	**Support** – Logs of parent-teacher communication	student_id, date, message_type

📌 **Key Relationships:**

- Primary key: student_id.

- Links:

    - **Attendance & Homework → Performance (Exam Score)**

    - **Communication frequency → Performance improvement**

---

2. Data Cleaning

Power Query for Data Cleaning

All datasets were imported into **Excel's Power Query editor** for systematic cleaning before analysis. Power Query allowed me to:

- **Attendance:**

  - Loaded raw `attendance.csv` into Power Query.

  - Removed duplicate rows by grouping on `(student_id, date, subject)`.

  - Used *Replace Values* to normalize categorical values (e.g., `" Late"` → `"Late"`).

  - Ensured a consistent set of statuses: Present, Excused, Late, Absent, Left Early.

- **Homework:**

  - Converted multiple inconsistent homework status values (like "✓", "pending", "done") into **Complete/Incomplete** using conditional column rules.

  - Replaced null values in **Guardian_Signature** with `"No"`.

  - Trimmed spaces from text fields with *Transform → Format → Trim*.

- **Performance:**

  - Applied filters to cap exam scores at 100.

  - Multiplied decimals in homework completion (0.85 → 85%) with a *Custom Column*.

  - Deduplicated rows and removed trailing blanks from Teacher_Comments.

- **Students:**

  - Extracted the numeric part of `"Grade 1"` → `1` with *Split Column by Delimiter*.

  - Used a small VBA macro later for capitalization of names, but phone number formatting and missing values (`"Missing Number"`) were handled directly in Power Query.

- **Teacher–Parent Communication:**

  - Used *Remove Duplicates* to ensure each message record was unique.

3. Excel (PivotTables)

Attendance

Count of Student_ID	Column Labels						
Row Labels	Absent	Excused	Late	Left Early	Present	Present	Grand Total
Arabic	15204	7485	15096	7452	7557	7543	60337
English	15015	7593	15073	7399	7534	7696	60310
Geography	15112	7474	14912	7527	7622	7614	60261
History	14997	7542	14925	7565	7466	7610	60105
Math	15264	7504	15216	7642	7576	7628	60830
Science	15068	7510	15175	7547	7554	7603	60457
Grand Total	90660	45108	90397	45132	45309	45694	362300

Shows which subjects have the most absences.

Homework

Count of Assignment_Name	Column Labels		
Row Labels	Complete	Incomp	Grand Total
Arabic	5044	5062	10106
English	5100	5109	10209
Geography	5135	5044	10179
History	5030	5080	10110
Math	5066	4998	10064
Science	5006	5106	10112
Grand Total	30381	30399	60780

Subjects with most incomplete homework

Count of Assignment_Name	Column Labels									
Row Labels	A	A+	B	B-	C	C-	D	F		Grand Total
Arabic	1251	1222	1266	1296	1243	1261	1326	1241		10106
English	1263	1307	1261	1317	1277	1274	1279	1231		10209
Geography	1225	1301	1288	1291	1283	1284	1245	1262		10179
History	1273	1307	1236	1244	1279	1236	1264	1271		10110
Math	1284	1253	1205	1277	1262	1220	1275	1288		10064
Science	1237	1319	1335	1214	1241	1231	1242	1293		10112
Grand Total	7533	7709	7591	7639	7585	7506	7631	7586		60780

Subjects with most A's vs. F's.

Count of Assignment_Name	Column Labels		
Row Labels	No	Yes	Grand Total
Arabic	6803	3303	10106
English	6742	3467	10209
Geography	6777	3402	10179
History	6687	3423	10110
Math	6718	3346	10064
Science	6705	3407	10112
Grand Total	40432	20348	60780

Subjects with highest parent signature rates.

Performance

Row Labels	Average of Exam_Score	Average of Homework_Completion_%
Arabic	74.19029251	72.68081003
English	74.40281457	72.81788079
Geography	73.84559671	73.1382716
History	74.32954355	72.63751881
Math	73.9585475	72.19366606
Science	74.417761	73.47340775
Grand Total	74.19058204	72.82456574

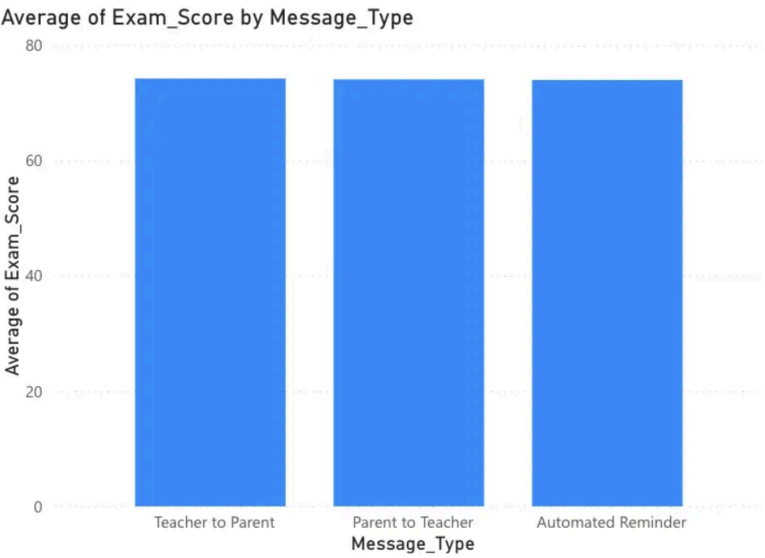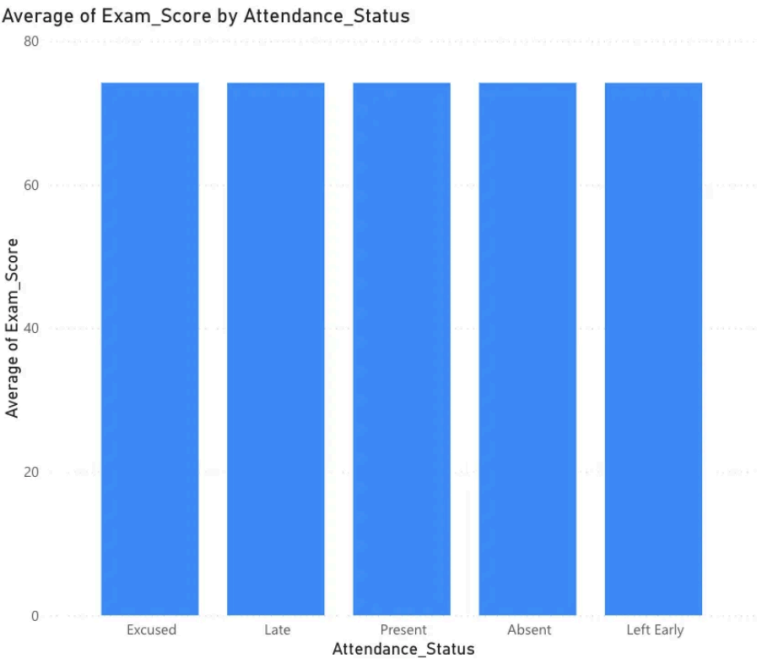Side-by-side analysis: homework completion vs. exam scores.

4. Tableau Visualizations

Attendance vs Subject:

<Attendance vs. Subject>

5. Power BI Visualizations

Average of Exam_Score by Attendance_Status

Average of Exam_Score by Message_Type
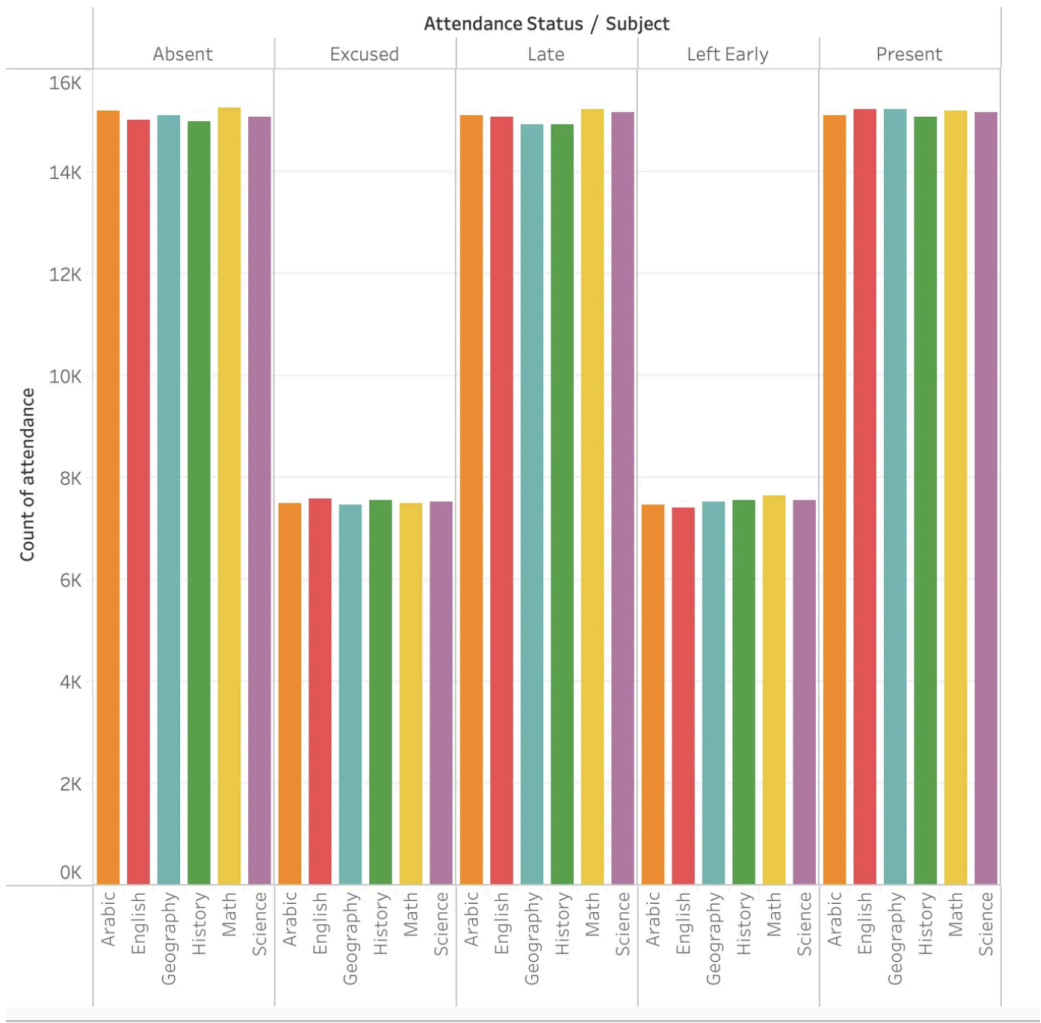
6. Insights & Conclusion

Unfortunately, the dataset used in this project did not yield clear or significant results. From the Excel pivot table analysis, there was no strong relationship observed between homework completion and exam scores. Similarly, Tableau & Power BI visualizations showed only minimal differences:

- **Attendance vs. Subject:** Very small variations across categories (Absent, Late, Excused, Left Early, Present).

- **Attendance vs. Exam Scores:** Students marked "Present" had the highest exam scores, but the difference compared to other categories was very small.

- **Parent-Teacher Communication vs. Exam Scores:** Students with teacher-to-parent communication showed slightly higher exam scores, but again the difference was marginal.

These small differences suggest that this dataset is not well-suited to draw strong conclusions about the factors influencing student performance. A more comprehensive dataset, with clearer variation and additional attributes, would be better for deeper insights.

Nonetheless, this project was valuable for skill development. I successfully applied and strengthened my abilities in:

- **Excel** (data cleaning, pivot tables)

- **Power Query** (data transformation)

- **Tableau** (data visualization and set analysis)

- **Power BI** (dashboard creation)

Overall, while the dataset's limitations restricted the findings, the project provided hands-on experience with essential data analysis and visualization tools.