Data Analysis Mini Course Project

Izza Zahrotunnisa

**RevoU
Mini Course**

Agenda

1. Profile Introduction

2. Project Background

3. Problem Statement

4. Methodology

5. Final Dashboard Result

6. Key Insights

Profile Introduction

I am a driven and analytical Accounting graduate from Universitas Gadjah Mada with strong foundations in finance, accounting information systems, and data-driven decision-making. My experience as Accounting Staff at AIESEC UGM, Media Manager at PALMAE FEB UGM, and Lead Unit Treasurer for KKN-PPM UGM 2023 strengthened my skills in financial reporting, process optimization, and cross-functional collaboration, achieving outcomes such as 100% financial compliance. I am currently advancing my data analytics expertise through training and certifications, with proficiency in Power BI, RStudio, SPSS, Looker Studio, and Excel.

Izza Zahrotunnisa

Accounting Graduate from Universitas Gadjah Mada

Yogyakarta, Indonesia

izzazahratunnisa@gmail.com

https://www.linkedin.com/in/izzazahrotunnisa/

RevoU
Mini Course

Project Background

- This project was completed as part of a five-day short course introducing the fundamentals of data analysis.
- Focuses: Practicing core data analyst skills, including data cleaning, exploration, visualization, and insight generation
- Using publicly available data and tools such as **Google Sheets** and **Google Slides**, the project emphasizes both analytical rigor and clear communication of findings.
- Objective: To build a functional dashboard, uncover meaningful trends, and present actionable insights effectively to stakeholders

Problem Statement

- Organizations often have large datasets but struggle to turn them into meaningful insights.
- This project addresses that challenge by cleaning, exploring, and visualizing data to uncover key trends and patterns.
- Focus: Preparing accurate data, identifying insights relevant to stakeholders, and presenting results clearly through dashboards → demonstrating how data analysis supports informed, data-driven decision-making.

Dataset Information

Kaggle Link:
https://www.kaggle.com/datasets/spscientist/students-performance-in-exams/data

Original Variables

Gender $\longrightarrow$ Student gender (male or female)

Race/Ethnicity $\longrightarrow$ Student ethnic group identifiers (e.g., Group A, Group B, Group C, Group D, Group E)

Parental Level of Education $\longrightarrow$ Highest level of education attained by a student's parent(s) (e.g., some college, bachelor's degree, etc.)

Lunch $\longrightarrow$ Type of lunch provided (standard or free/reduced).

Test Preparation Course $\longrightarrow$ Status of test preparation (none or completed).

Math Score $\longrightarrow$ Student's score in the mathematics exam.

Reading Score $\longrightarrow$ Student's score in the reading exam.

Writing Score $\longrightarrow$ Student's score in the writing exam.

Dataset Information

Kaggle Link:
https://www.kaggle.com/datasets/spscientist/students-performance-in-exams/data

Additional Variables Created

Student ID → A unique identifier assigned to each student for tracking and visualization purposes.

Average Reading Score → The student's reading score normalized or presented for comparative analysis.

Average Math Score → The student's math score restructured for analytical clarity.

Average Writing Score → The student's writing score reformatted to support comparative insights.

Total Score → Sum of math, reading, and writing scores for an overall performance measure.

Average Overall Score → The average of all three subject scores (math, reading, writing), providing a single metric to evaluate overall academic performance.

These additional variables enhance the dataset's utility for ranking, comparative analysis, trend identification, and dashboard visualization.

Methodology

1. Data Acquisition
   ○ The original dataset was obtained from Kaggle ([Students Performance in Exams](#)).
   ○ The dataset includes demographic information and exam scores for students.

2. Data Cleaning and Preparation
   ○ Imported the dataset into Google Sheets.
   ○ Checked for missing values, duplicates, and inconsistencies, and cleaned the data accordingly.
   ○ Added new variables to enhance analysis:
      ■ student ID – unique identifier for each student
      ■ average math score, average reading score, average writing score – to support per-subject analysis
      ■ total score – sum of all three subject scores
      ■ average score overall – mean of all three subject scores

Sorting function used to filter the Top 10 Highest Achieving Students

fx =SORT(A2:K, M2:M, FALSE)

V	W	X	Y	Z	AA	AB
			TOP 10 HIGHEST ACHIEVING STUDENTS			
Student 459	female	group E	bachelor's degre	standard	none	100.0
Student 917	male	group E	bachelor's degre	standard	completed	100.0
Student 963	female	group E	associate's degr	standard	none	100.0
Student 115	female	group E	bachelor's degre	standard	completed	99.0
Student 180	female	group D	some high scho	standard	completed	97.0
Student 713	female	group D	some college	standard	none	98.0
Student 166	female	group C	bachelor's degre	standard	completed	96.0
Student 626	male	group D	some college	standard	completed	100.0
Student 150	male	group E	associate's degr	free/reduced	completed	100.0

Final table of top 10 Highest Achieving Students

Table1

Student ID	Gender	Parental Level of Education	Test Preparation	Total Score
Student 459	female	bachelor's degree	none	300
Student 917	male	bachelor's degree	completed	300
Student 963	female	associate's degree	none	300
Student 115	female	bachelor's degree	completed	299
Student 180	female	some high school	completed	297
Student 713	female	some college	none	297
Student 166	female	bachelor's degree	completed	296
Student 626	male	some college	completed	296
Student 150	male	associate's degree	completed	293
Student 686	female	master's degree	completed	293

Functions used to calculate score ranges and final table of frequency

```
=COUNTIFS(N:N,">50", N:N,"<=60")
```

```
=COUNTIFS(StudentsPerformance!N:N,">=0",StudentsPerformance!N:N,"<=50")
```

Frequency of Students by Score Range

Histogram Values ∨	
T_T Score Range ∨	**Number of Students** ∨
0–50	109
51–60	173
61–70	245
71–80	240
81–90	129
91–100	46
Total of Students	942

Function used to calculate minimum and maximum scores, and quartiles for each subject and gender (candlestick chart)

```
=MIN(FILTER(StudentsPerformance!G2:G, StudentsPerformance!B2:B="Male"))

=MIN(FILTER(StudentsPerformance!G2:G, StudentsPerformance!B2:B="Female"))

=QUARTILE(FILTER(StudentsPerformance!G2:G, StudentsPerformance!B2:B="Male"),1)

=MEDIAN(FILTER(StudentsPerformance!G2:G, StudentsPerformance!B2:B="Male"))

=QUARTILE(FILTER(StudentsPerformance!G2:G, StudentsPerformance!B2:B="Male"),3)

=MAX(FILTER(StudentsPerformance!G2:G, StudentsPerformance!B2:B="Male"))
```

Methodology

3. **Data Transformation and Exploration**
- ○ Calculated aggregate measures such as average scores by gender, parental education level, lunch type, test preparation course, and ethnicity.
- ○ Applied filter functions and sorting to identify top-performing students and their respective statistics and analyze trends across different groups.

4. **Data Visualization**
- ○ Used Google Sheets to create charts including:
  - ■ Bar charts for comparing scores across subjects and demographic groups
  - ■ Pie charts to visualize the proportion of students meeting certain performance thresholds
  - ■ Scatter plots and rankings to highlight top performers and relationships between variables
- ○ Designed dashboards to combine multiple visualizations for easy interpretation of trends and insights.

5. **Insight Generation**
- ○ Analyzed the visualizations to identify patterns and trends in student performance.
- ○ Summarized findings for presentation to stakeholders, highlighting key factors that may influence academic outcomes.

Final Dashboard

Academic Performance Insights: Student Exam Analysis

Confidential
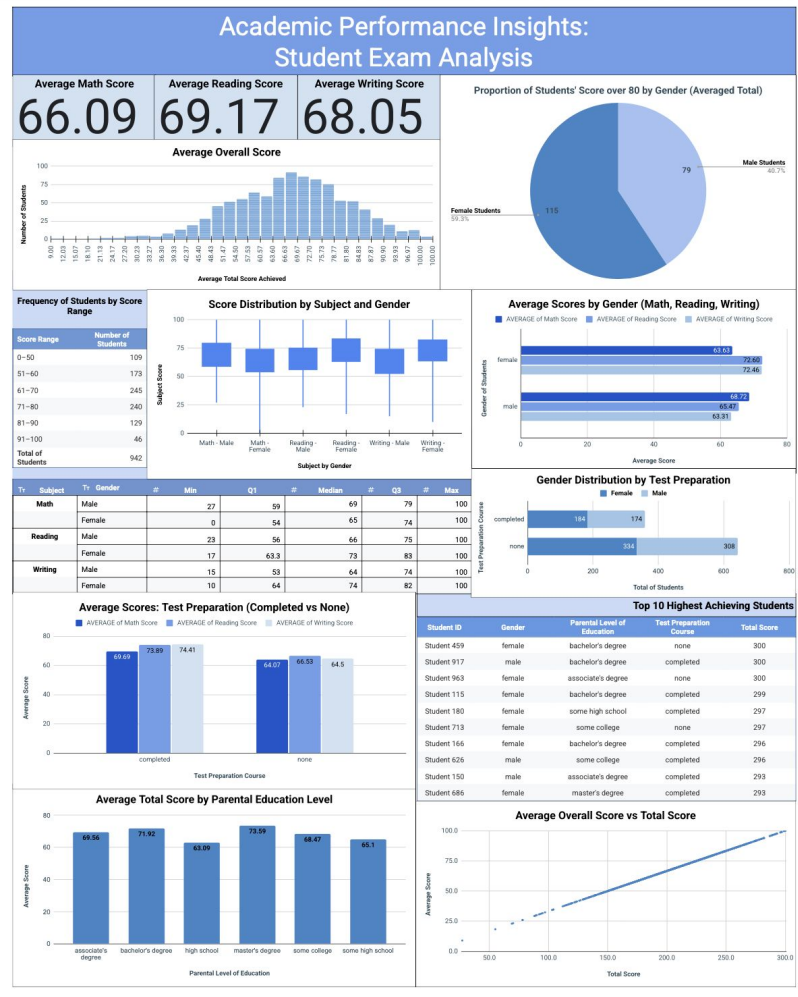
Key Findings

Gender-Based Performance
- Among students scoring above 80:
  - Female: 59.3% (115)
  - Male: 40.7% (79)
- Males perform better in math, while females outperform in reading and writing.
Females show stronger reading performance and higher upper score ranges.

Score Distribution
- Most students fall in the 61–80 range (485 students total).
- Few students reach top scores:
  - 81–90: 129 students
  - 91–100: 46 students
- 109 students score below 50, indicating a need for academic support.

Test Preparation Impact
- Test prep participation is balanced across genders.
- Students who completed test prep score consistently higher across all subjects:
  - Math: 69.69 vs 64.07
  - Reading: 73.89 vs 66.53
  - Writing: 74.41 vs 64.50

High Achievers & Consistency
- 7 of the top 10 students are female.
- Bachelor's degree is the most common parental education among top performers.
- Strong linear relationship between average score and total score, indicating consistent performance across subjects.

RevoU
Mini Course

Conclusion

Student performance is influenced by gender, test preparation, and parental education level.

- Female students excel in language-based subjects and dominate high-achiever groups, while test preparation shows a strong positive association with scores across all subjects.

→ These insights highlight the importance of targeted academic support and data-driven decision-making in education.

Links

GitHub: https://github.com/izzazahrotunnisa/RevoU-DataAnalytics/tree/main

LinkedIn: https://www.linkedin.com/in/izzazahrotunnisa/

Google Sheets Dashboard:

https://docs.google.com/spreadsheets/d/1r0C_Cq4M3NB3iwOiCqlWITYsfCC-ku3whXt b-oZIYSg/edit?usp=sharing

Thank You!

Author's Contact Details

- Izza Zahrotunnisa
- Email: izzazahratunnisa@gmail.com
- Phone Number: +6281319317614 / +97433881815
- Linkedin: https://www.linkedin.com/in/izzazahrotunnisa/