



AMITY
UNIVERSITY
— Online

Course: Data Modeling and Visualization - TCS

Project: Module 2: Student Analysis

Submitted by:
Hemant Borana



Contents

Approach and Methodology.....	3
Key Findings.....	4
Challenges Faced	5
Learnings and Takeaways	6
Practical Applications	7

Approach and Methodology

When I started this assignment, I first focused on setting up the R environment properly. I installed RStudio and all the required packages like ggplot2, dplyr, readr, tidyr, and rmarkdown. Since I didn't have the dataset, I created a sample Student_Performance.csv file with 100 student records containing all the necessary fields - demographics, scores, study hours, and attendance data.

For data manipulation, I used dplyr functions extensively. I practiced filtering data to find high-performing students, grouping data by gender and parental education to calculate average scores, and creating new calculated columns like average scores across all three subjects. This helped me understand how different factors relate to student performance.

Visualizations Created

The visualization part was interesting. I created five different types of charts:

1. **Bar Chart** - This showed the gender distribution in the dataset. I found that we had 54 female and 46 male students, which is fairly balanced.
2. **Scatter Plot** - I analyzed the relationship between study hours and math scores. The positive trend line clearly showed that students who study more hours tend to score better in math.
3. **Histogram** - The distribution of math scores revealed that most students score around the mean of 68, with a fairly normal distribution.
4. **Box Plot** - Comparing scores across different parental education levels showed that students whose parents have Master's degrees generally perform better and have less score variation.
5. **Line Chart** - Looking at trends across different age groups helped identify patterns in how performance changes as students get older.

Key Findings

Through this analysis, I discovered several important patterns:

- There's a clear positive correlation between study hours and academic performance. Students who dedicate more time to studying consistently achieve higher scores.
- Parental education level appears to be a significant factor. Students with highly educated parents tend to perform better academically, possibly due to better home support and resources.
- Gender differences exist but are not extreme. Female students showed slightly higher writing scores while male students performed marginally better in math.
- The overall score distribution suggests most students are performing at average to above-average levels, with relatively few students at the extreme ends.

Challenges Faced

The main challenge I faced was understanding how different R packages work together. Initially, I had some confusion with dplyr syntax, especially with the pipe operator (`%>%`), but after practice it became clearer.

Another challenge was formatting the R Markdown document properly to ensure all plots displayed correctly and the code chunks ran without errors.

Creating visualizations that were both informative and visually appealing required some experimentation with different parameters like colors, sizes, and themes in ggplot2.

Learnings and Takeaways

This assignment taught me valuable skills in data analysis and visualization using R:

1. How to properly structure and manipulate data using dplyr functions
2. Creating professional visualizations using ggplot2 with appropriate customization
3. Understanding when to use different chart types for different kinds of analysis
4. Documenting analysis in a reproducible way using R Markdown
5. Interpreting data patterns and drawing meaningful conclusions

I now feel more confident in using R for data analysis tasks and understand the importance of proper data visualization in communicating insights effectively.

Practical Applications

The techniques learned in this assignment can be applied to real-world scenarios like:

- Analyzing student performance data to identify at-risk students
- Understanding factors that contribute to academic success
- Making data-driven decisions in educational planning
- Creating reports for stakeholders like teachers, administrators, and parents

Overall, this project provided hands-on experience with the complete data analysis workflow from data loading to final report generation.



Thank You

Contact Details:



hemantpb123@gmail.com



9284494154



www.linkedin.com/in/hemant-parasmal-borana

