**Analyzing the Impact of Family Factors on Students’ Academic Success**

DSCP Final Projection

Department of Business Administration 117 楊子嫺 41357001o

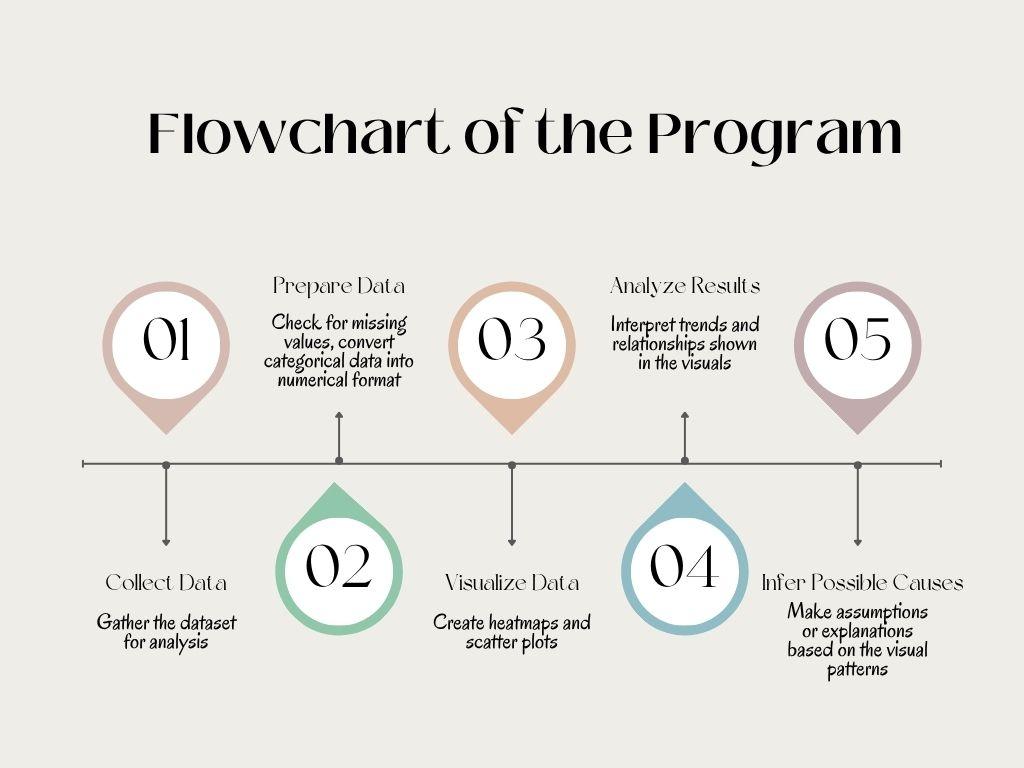
**Introduction:**

Here is my [data source](https://www.kaggle.com/datasets/lainguyn123/student-performance-factors/data) from Kaggle. The dataset contains various features related to student performance, such as Hours Studied, Parental Involvement, and Sleep Hours. Usually, family is considered one of the most influential factors in a person’s development. As a result, I became curious about whether this is also true for students’ academic performance. Hence, I selected family-related and student performance variables to analyze and examine how these factors are related.

**Objectives:**

1. Generate a heatmap to examine the relationships between various factors
2. Create a scatter plot to analyze how family-related factors affect students’ performance in exam in detail
3. Summarize insights from the data analysis

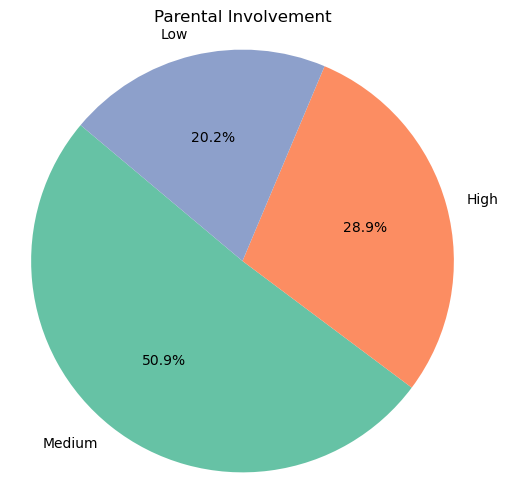
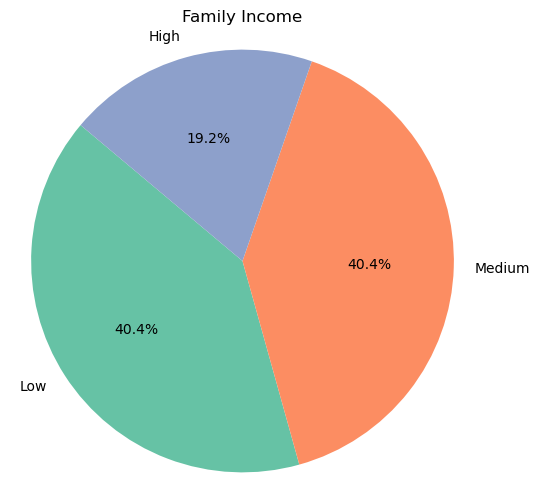
**Flowchart:**

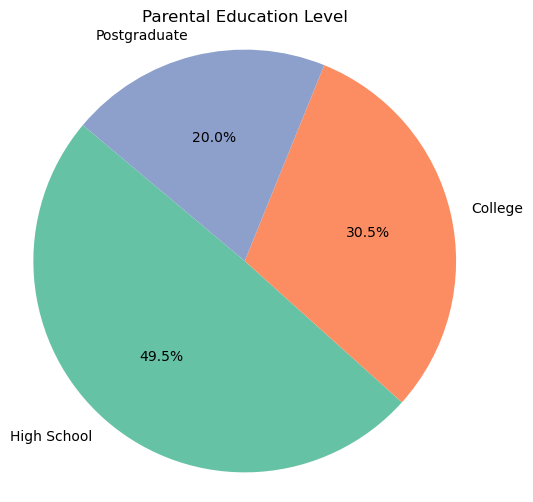


**Data Description:**

In this project, I mainly selected three factors—Family Involvement, Family Income, and Parental Education Level—to analyze. Below are pie charts for each of these three factors separately.

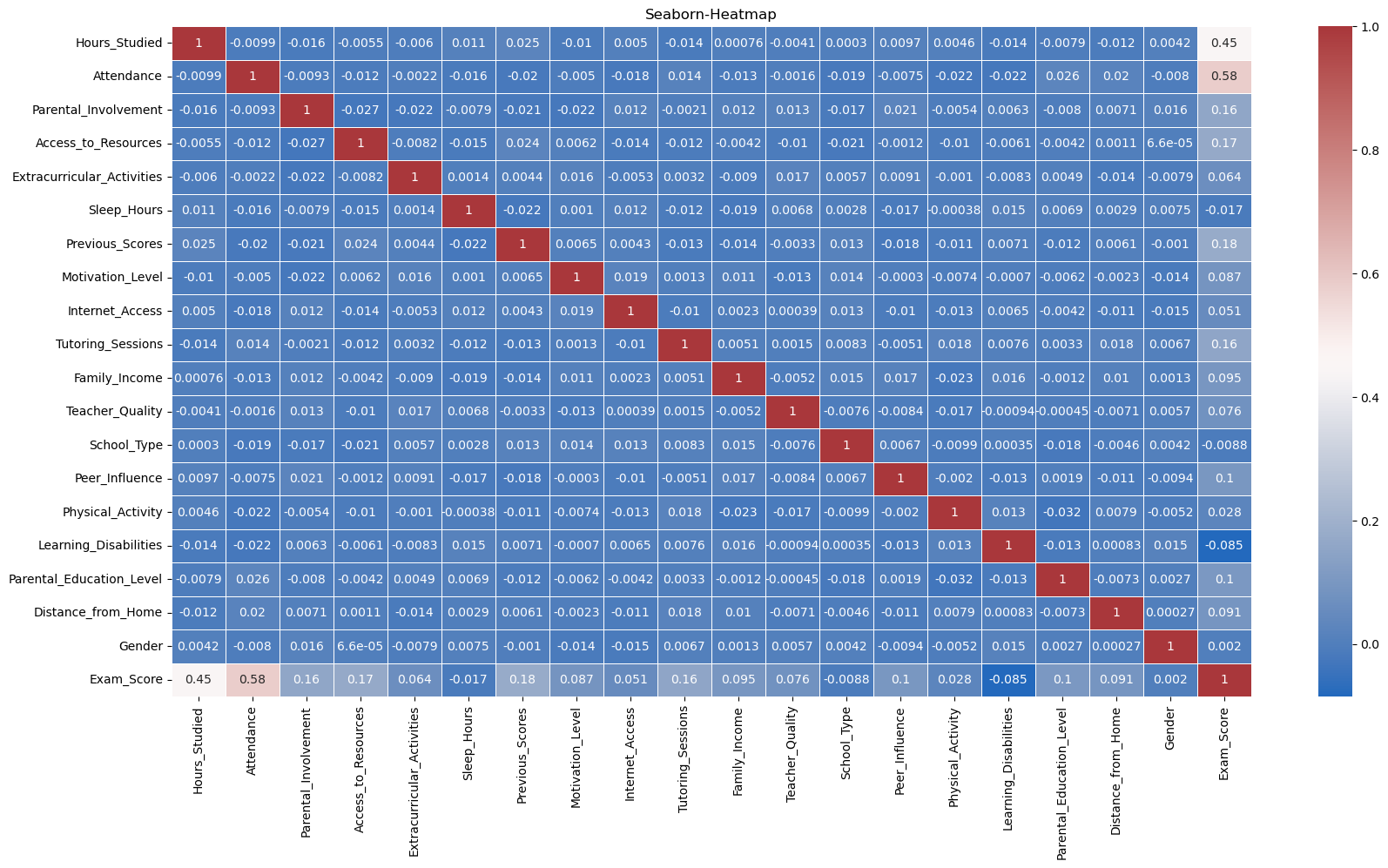
However, after generating these pie charts, I found that the proportions of these three factors vary significantly in the dataset. Take Parental Involvement for example, the low-Parental Involvement group accounts for only approximately 20%, the middle-Parental Involvement group approximately 30%, and the high-Parental Involvement group approximately 50%. Due to the uneven distribution, the analysis results may be biased or skewed due to the imbalance in group sizes.





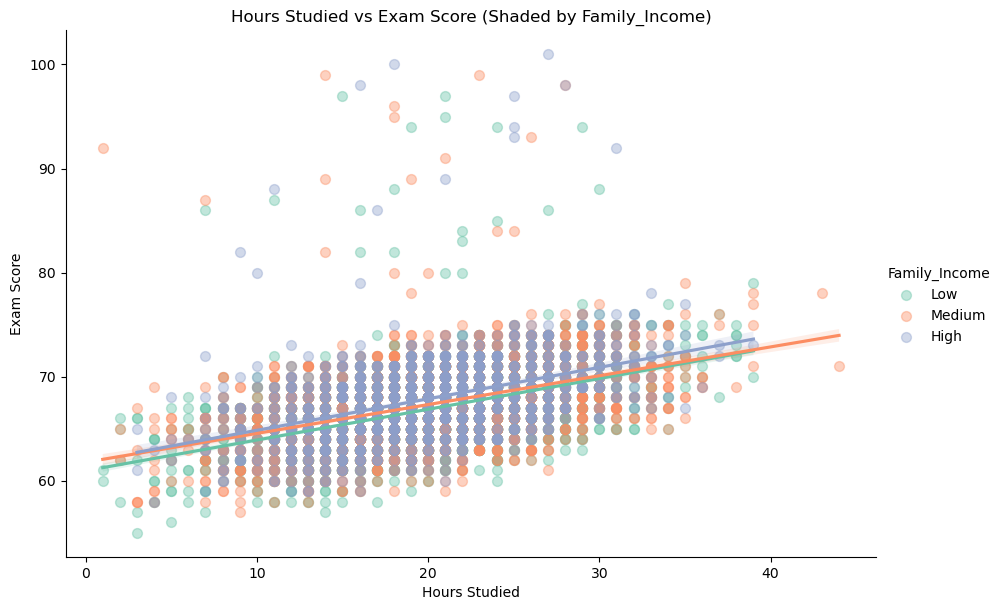
**Data Visualization and Analysis:**

From the heatmap, we can observe the main factors affect Exam Score is Attendance and Hours Studied. However, only weak correlations between each family-related factor and Exam Score. Therefore, I chose to visualize the data using scatter plots to examine the detailed relationships and potential trends that may not be captured by correlation alone.



The scatter plot of Hours Studied vs Exam Score (Shaded by Family Income) shows three regression lines representing different Family Income groups. From these lines, we can see that study time and Exam Scores are weakly positively correlated across all income levels.

Additionally, a notable observation is that the regression line for the high-Family Income group lies above that of the low-Family Income group, indicating that students from higher-income families tend to achieve higher exam scores than those from lower-income families, even when studying for the same amount of time.

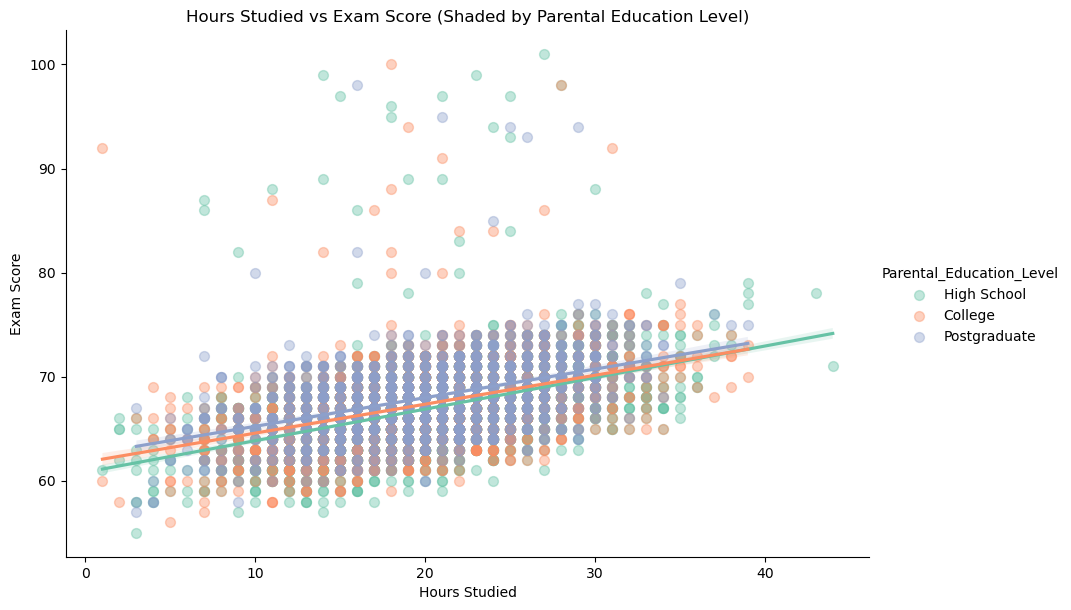


The scatter plot of Hours Studied vs Exam Score (Shaded by Parental Education Level) shows three differently colored regression lines representing different Parental Education Level groups.

It show that students from higher Parental Education Level families tend to achieve higher exam scores than those from lower Parental Education Level families, even when studying for the same amount of time.

However, the slope for the low Parental Education Level group is steeper, suggesting that additional study hours have a greater impact on improving exam scores for students from lower Parental Education Level families.

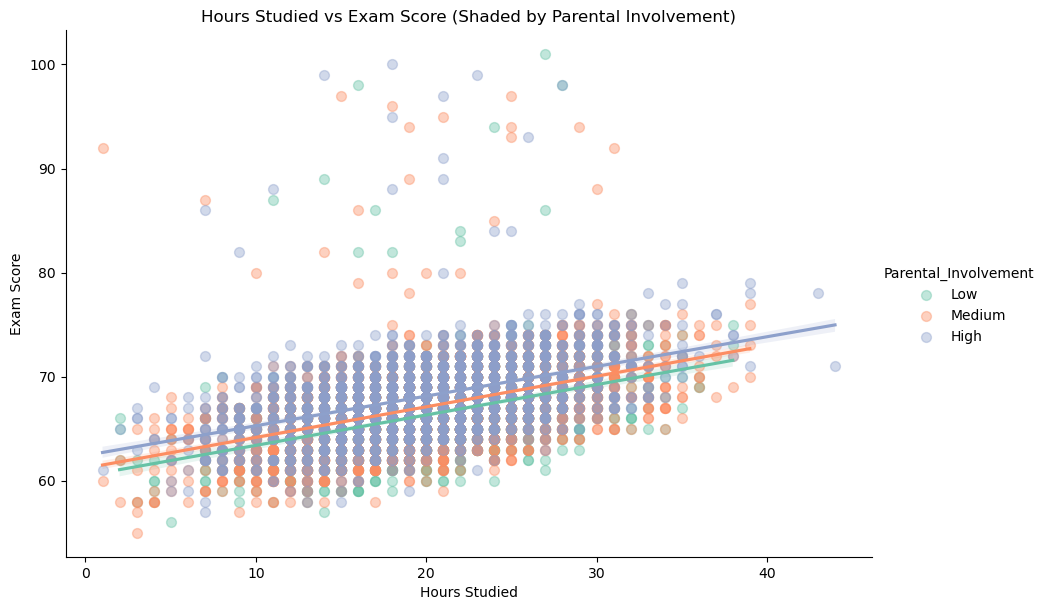
I think that this may imply that while higher Parental Education Level provides an initial advantage, increased effort can significantly boost performance among lower Parental Education Level students.



The scatter plot of Hours Studied vs Exam Score (Shaded by Parental Involvement) shows three differently colored regression lines representing different Parental Involvement groups. I found that the regression lines grouped by Parental Involvement show greater vertical separation compared to those grouped by Family Income and Parental Education Level.

This suggests that parental involvement has a stronger influence on exam scores than family income and Parental Education Level, as the difference in performance between high, medium, and low involvement groups is more pronounced.

In other words, students with highly involved parents tend to perform significantly better than those with less involved parents, even when they study the same amount of hours.



**conclusion:**

Based on the heatmap, we can observe that the family-related factors aren’t the primary determinants of the exam scores. They show only weak positive correlations. Therefore, in order to explore these relationships in greater detail, I generated the scatter plots and regression lines categorized by family income, Parental Education Level, and Family Involvement.

The visualizations reveal that, overall, students from more advantaged families, such as high family income, high parental education level, and high family involvement, tend to achieve better academic performance. This suggests that the family conditions provide an initial advantage. However, this advantage is neither absolute nor permanent. In particular, the slope of the regression lines for students from less advantaged backgrounds indicates that additional study hours can significantly improve performance.

In conclusion, we believe that students from well-off families may begin with a head start. Nonetheless, the advantage is not absolute or significant. If the students from disadvantaged families study hard, the advantage can be reduced—or even overcome.