Said Djama

I started this project by first looking at the dataset and trying to understand what kinds of questions it could answer. The dataset focuses on student performance and includes information like hours studied, sleep hours, previous scores, practice papers, extracurricular activities, and a final performance index. Right away, I noticed that most of the variables were numeric, which made it good for exploring relationships instead of just counts. My main goal became understanding how different habits like studying, sleeping, and practicing questions relate to student performance.

At the beginning, I considered using multiple bar charts to compare averages across different variables. I also thought about showing several charts at once on the page. After sketching a few rough ideas on paper, I realized that this approach made it harder to see direct relationships between variables. It felt cluttered and did not really help answer questions about how one factor changes with another. Because of that, I decided to center the design around a scatterplot since it is much better at showing trends, patterns, and correlations between two variables.

The final visualization uses a scatterplot where the y axis is always the performance index and the x axis can be changed using a dropdown menu. This allows the viewer to explore how performance relates to hours studied, sleep hours, previous scores, or practice papers without needing multiple separate charts. I chose to keep performance index fixed on the y axis so that comparisons across different x variables feel consistent and easy to understand.

Color is used to represent whether a student participates in extracurricular activities. I chose color for this variable because it is categorical and only has two values. Using color makes it easy to compare the two groups without adding extra visual clutter. I avoided using too many colors so the chart would stay readable. Tooltips were added so that when hovering over a point, more detailed information about that student appears. This keeps the main chart clean while still allowing access to all attributes in the dataset.

While exploring the data with this visualization, I noticed several interesting patterns. Students who study more hours generally tend to have higher performance index values, although the relationship is not perfectly linear. Sleep hours also seem to matter, since extremely low sleep often appears alongside lower performance scores. When comparing students with and without extracurricular activities, there does not appear to be a strong negative impact on performance. In many cases, students involved in extracurriculars perform similarly to those who are not, which was somewhat surprising.

Overall, this visualization helped me answer questions that would have been harder to see just by looking at the raw data or a table. The interactive elements made it easier to explore multiple relationships without overwhelming the viewer. The design choices were focused on clarity and exploration rather than complexity, which made the data easier to understand and more engaging to interact with.