Github: https://github.com/nlesta/excel_project

Data source: https://www.kaggle.com/datasets/spscientist/students-performance-in-exams

For this portfolio project, I chose to examine student test data. The goal was to create an interactive dashboard that analyzes student performance across various demographic data. Users can utilize slicers to understand specific demographic groups. I also implemented VBA code to enable automatic data import and pivot table/chart updates when needed.

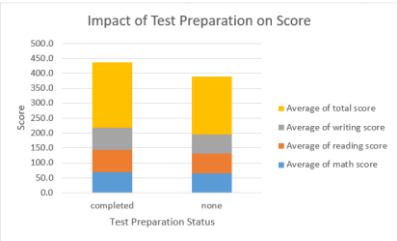
To start, I created a working data sheet to avoid modifying the original data. I then checked for duplicates (none found), formatting issues (none detected), and missing data (none present).

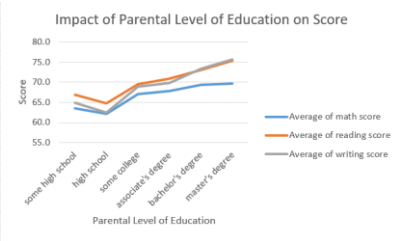
Next, I calculated a total score by summing math, reading, and writing scores. I also categorized students based on their scores using the Excel PERCENTILE function and nested IF statements. This resulted in three categories: "good," "average," and "poor." I repeated this process for all subjects and the total score. The formula is provided below.

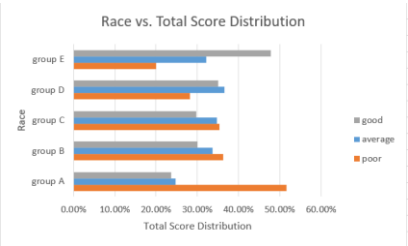
```
=IF(F2<PERCENTILE(F$2:F$1001,0.33), "poor",IF(F2>PERCENTILE(F$2:F$1001,0.66),"good", "average"))
```

I proceeded to analyze potential effects on student scores: completion of test preparation, parental education, and race/ethnicity. As expected, students who completed test prep scored higher. Parental education also showed a positive correlation with student scores. However, some interesting patterns emerged, like students with parents who completed partial high school outperforming those with fully high school educated parents. Also, some score drops appeared for students with parents holding master's degrees, particularly when filtered by gender or race group.

Lastly, I examined score distribution by race. I observed similar distributions for groups B-D, with around 30% in each tertile. However, group A had over 50% of students in the "poor" category, and group E had nearly 50% in the "good" category and only 20% in the "poor" category. The charts are provided below.







To conclude, I developed a dashboard with interactive slicers, allowing stakeholders to explore data trends. Trends deviated significantly depending on selected demographics. Additionally, I added a macro for stakeholders to automate the entire process when working with new data. You can find the code with comments in my GitHub folder. Feel free to test the macro with the test datasets and compare the results!