Data Variables Explained:

1. Student ID: ID number used to uniquely identify each row inside the student\_list. INT
2. Parental\_level\_of\_education: Shows the highest level of education that a parent received. INT
3. Lunch\_plan: Indicates the level of lunch a student has enrolled into. INT
4. Test\_preparation\_course: Indicates whether the student has successfully completed a test preparation course or not. TEXT
5. Test\_ID: ID Number used to uniquely identify each row pertaining to the test. INT

Pre-Processing Data:

Inside 3 CSV files, there was data that contained a student list, survey results, and test averages. Parental\_level\_of\_education had numerical data that did not explain its nature. The range was from 1 to 6. Also, this particular field inside Excel was formatted as text. To make it not text, I had changed its format into Number and removed the leading decimals. For null values, none were detected. To calculate the data and obtain the averages, I had to either create additional sheets inside of Excel or merge the fields onto one sheet using the XLOOKUP() function. Fields inside the survey table had values that needed decoding. Therefore, I translated the numerical values into categorical values. Upon translation, I made the column easier to read with valuable information rather than just numbers inside a cell. Additionally, I added two extra tables. One labeled subjects and the other tests. Both tables had to generate a surrogate key to relate to the fact table, student\_exam\_results.

Data Analysis:

To answer the first question, I worked inside Excel to generate different tables and charts. When using conditional formatting, I noticed that some scores were extremely high across all three subjects. I noted these as the outliers. The averages were closer to the below 70 range. After conducting this analysis, I created a scatter plot to identify any correlations between parental\_level\_of\_education field and test score averages. Later on, I generated more scatter plots and column charts to compare the results based on gender, test preparation, and lunches. Finally, I determined that a strong correlation exists between test preparation and average test scores along with parental education levels. The analysis behind lunch and test scores correlation remains inconclusive due to lack of related data and possible scientific reasoning. Gender correlation shows that different subjects and test scores are proven to be stronger with one certain gender over the other.

SQL Queries:

select year\_group ,AVG(math\_score) AS "Math Score Average", AVG(reading\_score) AS "Reading Score Average", AVG(writing\_score) AS "Writing Score Average" from student\_exam\_results inner join student\_list on student\_exam\_results.student\_id = student\_list.student\_id group by year\_group order by year\_group;

select gender ,AVG(math\_score) AS "Math Score Average", AVG(reading\_score) AS "Reading Score Average", AVG(writing\_score) AS "Writing Score Average" from student\_exam\_results inner join student\_list on student\_exam\_results.student\_id = student\_list.student\_id group by gender;

filter function or where clause in sql to search and demonstrate outliers

select parental\_level\_of\_education,gender ,AVG(math\_score) AS "Math Score Average", AVG(reading\_score) AS "Reading Score Average", AVG(writing\_score) AS "Writing Score Average" from student\_exam\_results inner join student\_list on student\_exam\_results.student\_id = student\_list.student\_id inner join student\_exam\_survey es on es.student\_id = student\_list.student\_id group by parental\_level\_of\_education, gender order by parental\_level\_of\_education;

select case when parental\_level\_of\_education = 1 then "Some High School" when parental\_level\_of\_education = 2 then "High School" when parental\_level\_of\_education = 3 then "Some College" when parental\_level\_of\_education = 4 then "Associates" when parental\_level\_of\_education = 5 then "Bachelors" when parental\_level\_of\_education = 6 then "Masters" ELSE parental\_level\_of\_education END AS "Level of Education",gender ,AVG(math\_score) AS "Math Score Average", AVG(reading\_score) AS "Reading Score Average", AVG(writing\_score) AS "Writing Score Average" from student\_exam\_results inner join student\_list on student\_exam\_results.student\_id = student\_list.student\_id inner join student\_exam\_survey es on es.student\_id = student\_list.student\_id group by parental\_level\_of\_education, gender order by parental\_level\_of\_education;