**Principles of Data Science (5530-0001)-Assignment 1**

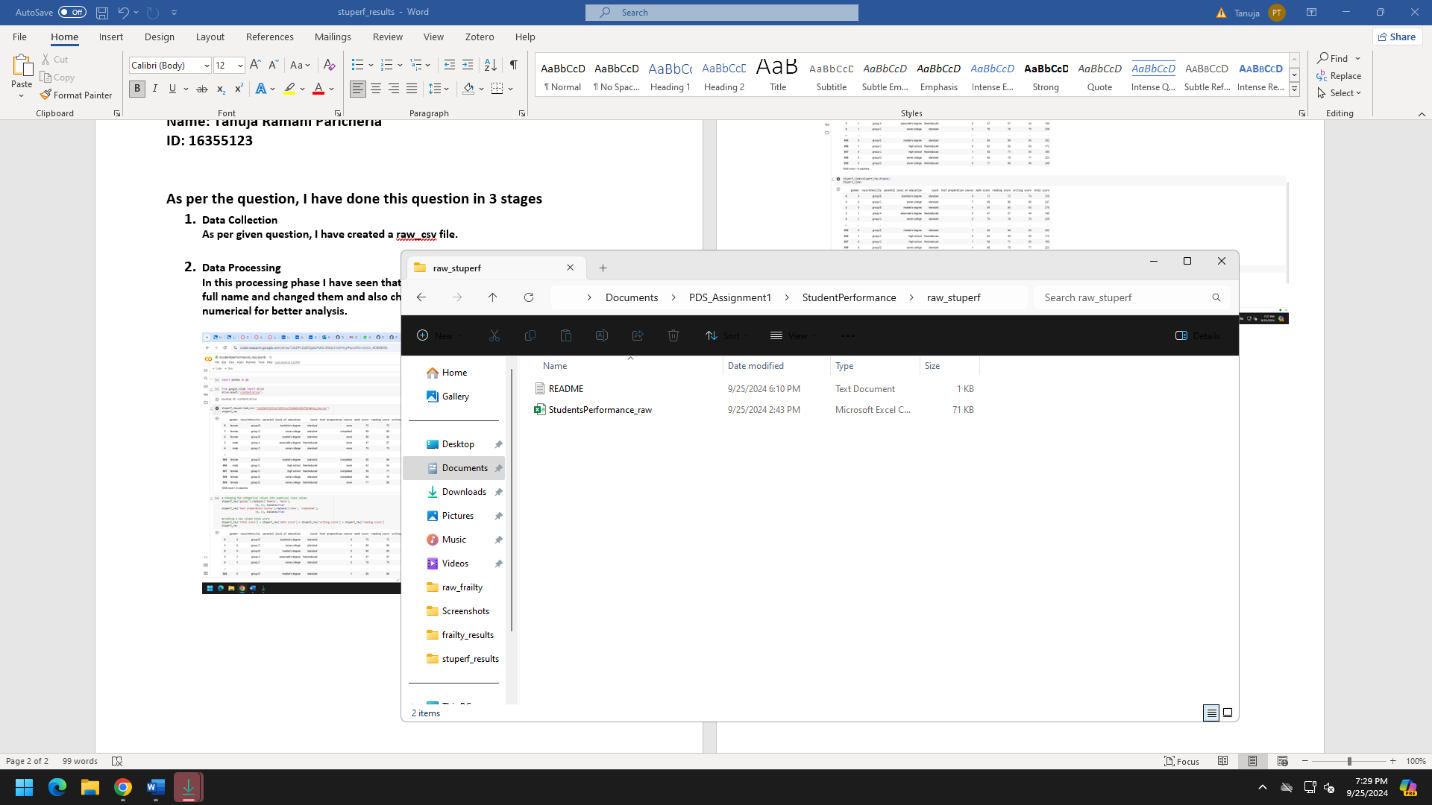
**Name: Tanuja Ramani Paricherla**

**ID: 16355123**

**As per the question, I have done this question in 3 stages**

1. **Data Collection**

**As per given question, I have created a raw\_csv file.**

****

1. **Data Processing**

**In this processing phase I have seen that column names can be much more meaning full name and changed them and also changed the categorical values to numerical for better analysis.**

**A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated**

1. **Data Analysis**

**In this phase, I have done data analysis.**

**I have used matplot lib and seaborn packages**

1. **Pie chart-parental level of education**

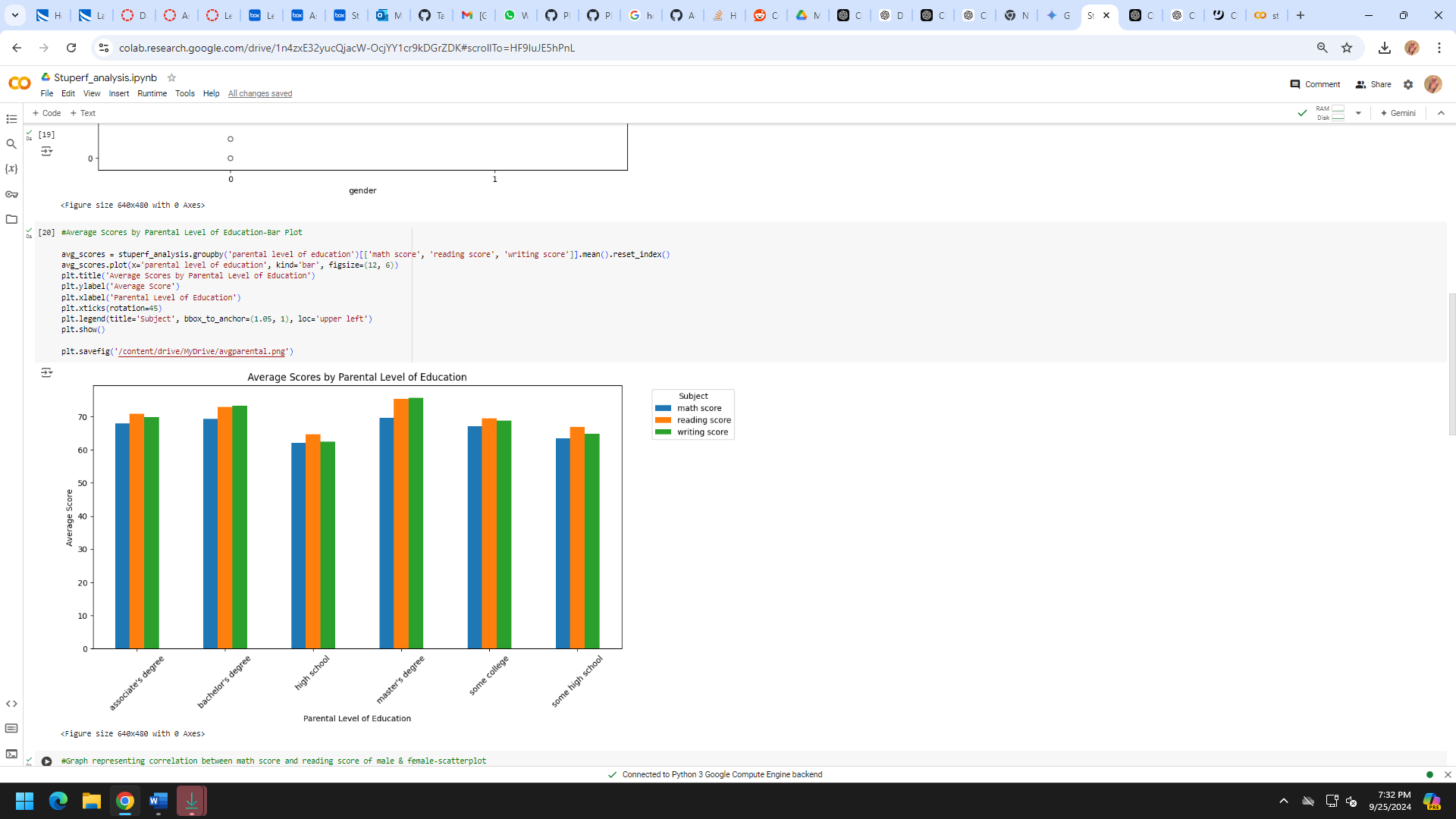
The pie chart shows that a majority of parents in the sample have completed at least some college education, with associate's degrees being the most common. However, a significant portion of parents only have a high school diploma or less, highlighting potential disparities in educational attainment within the community.

**A computer screen shot of a pie chart

Description automatically generated**

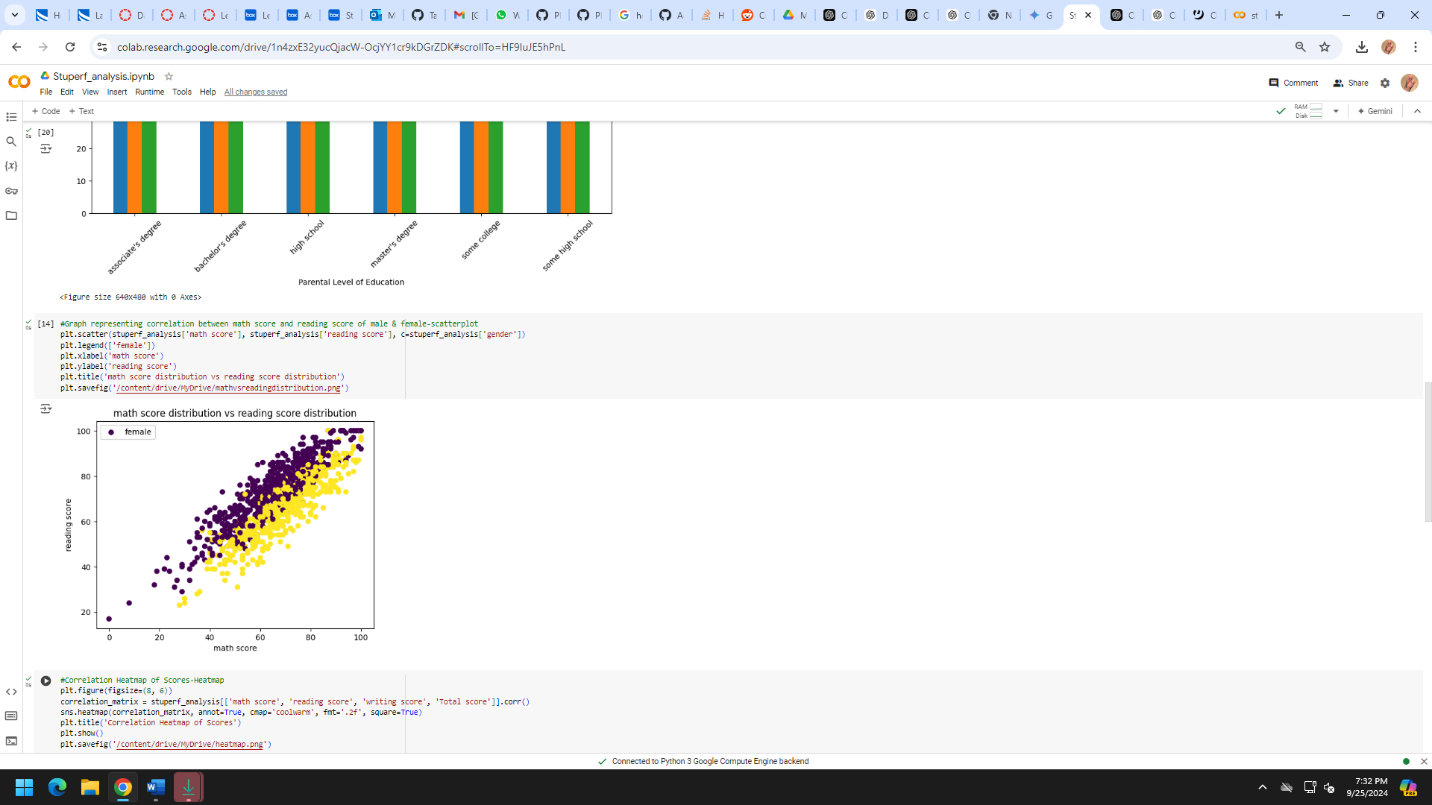
**2.** **Average Scores by Parental Level of Education-Bar Plot**

The graph suggests that there's a general trend where students whose parents have higher education levels tend to have higher average scores in all three subjects. For instance, students with parents who have master's degrees generally outperform students with parents who only have a high school diploma.

****

**3.Graph representing correlation between math score and reading score of male & female-scatterplot**

The scatter plot suggests a general positive correlation between math and reading scores. This means that students who tend to perform well in math also tend to perform well in reading, and vice versa for both geneders.

****

1. **Correlation Heatmap of Scores-Heatmap**

The heatmap suggests that there are strong positive relationships between math, reading, and writing scores. This implies that students who excel in one subject are likely to do well in the others.

The high correlation between the Final score and the individual subject scores indicates that the Final score is a good overall measure of academic performance

**A screenshot of a computer

Description automatically generated**

1. **Performance Based on Race/Ethnicity Bar Plot**

The bar chart shows the average Final score for five racial/ethnic groups (A, B, C, D, and E). While there are some variations, the overall scores are relatively similar across the groups. Group E has the highest average Final score, followed by Group D. Group A has the lowest average Final score

**A screenshot of a computer

Description automatically generated**

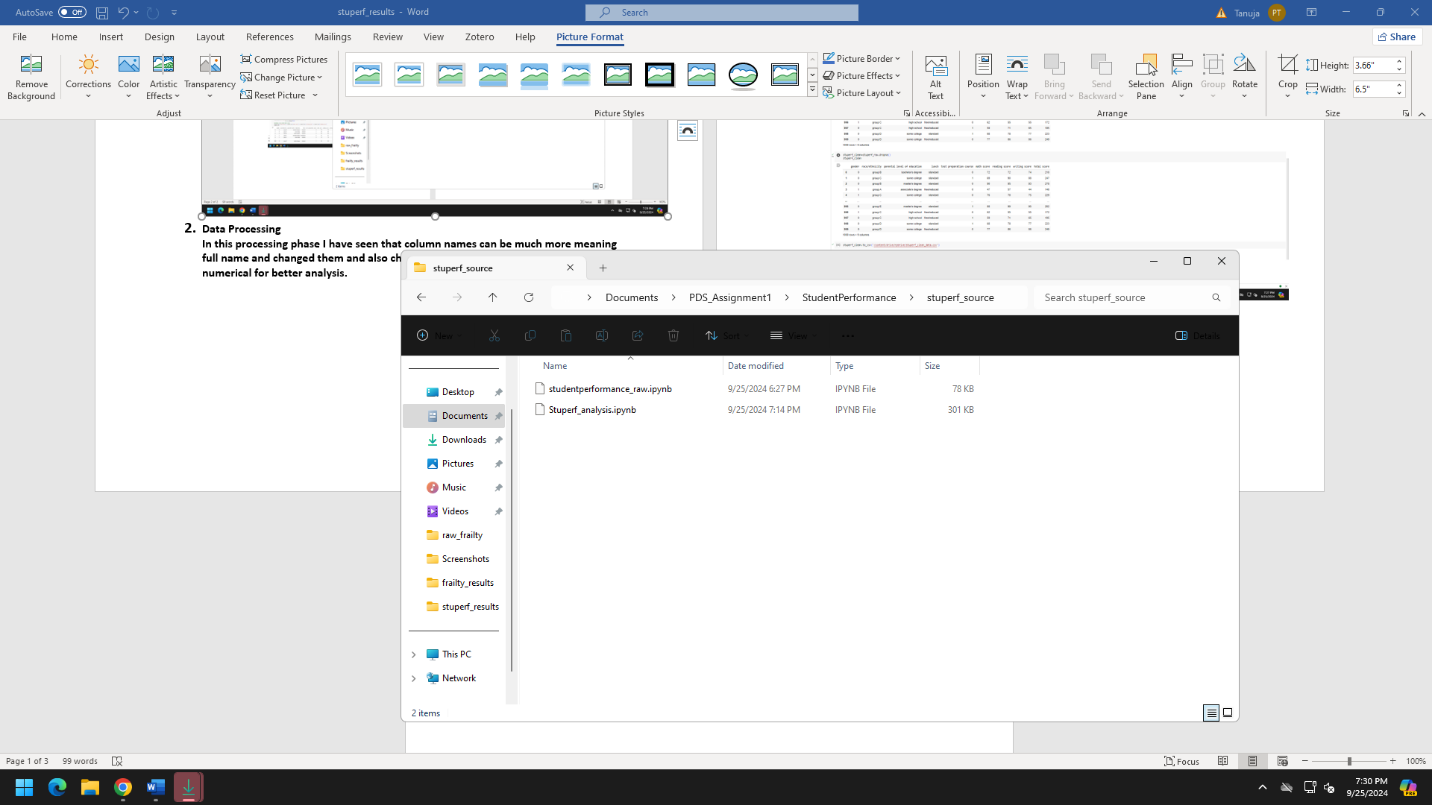
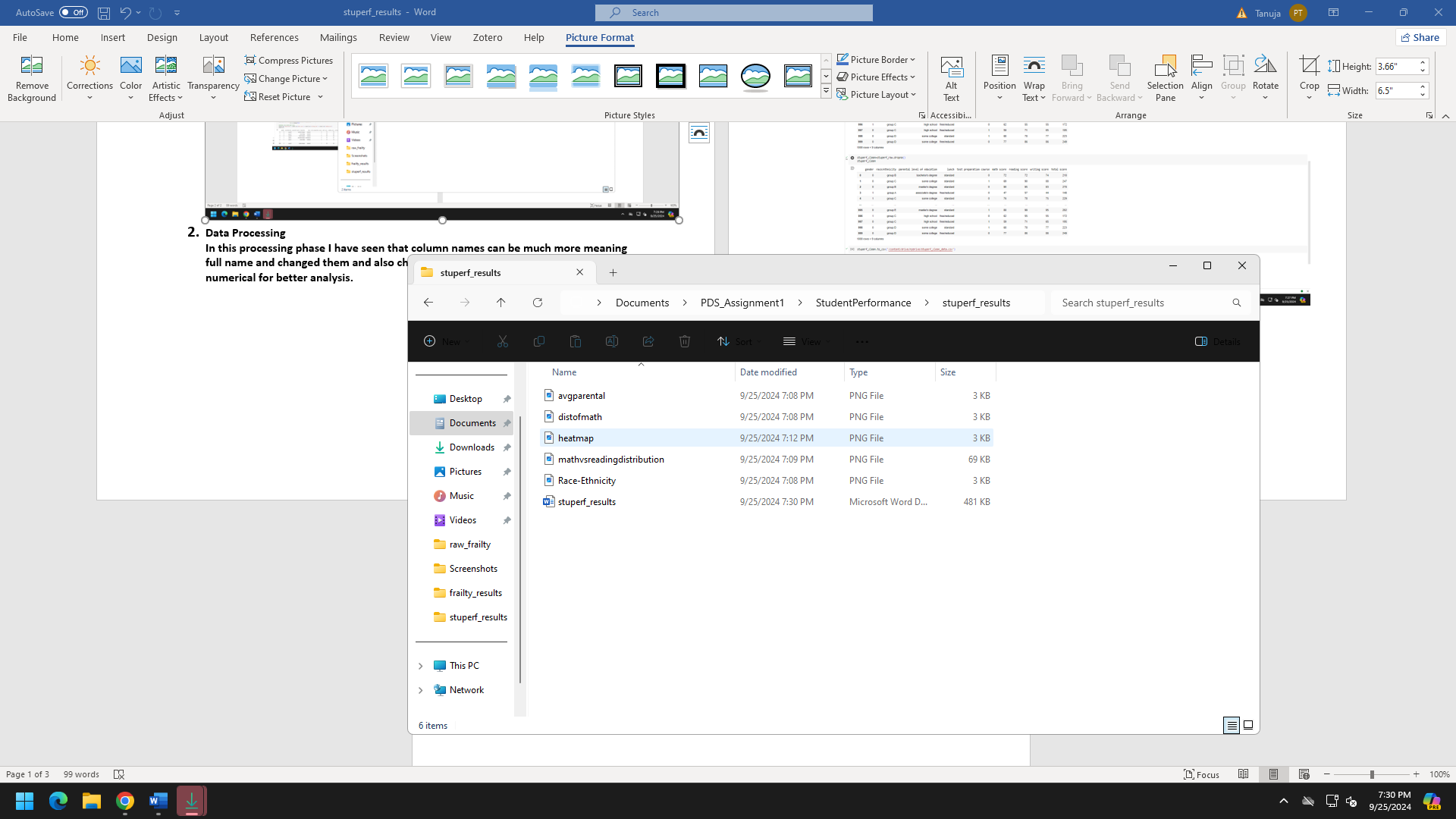
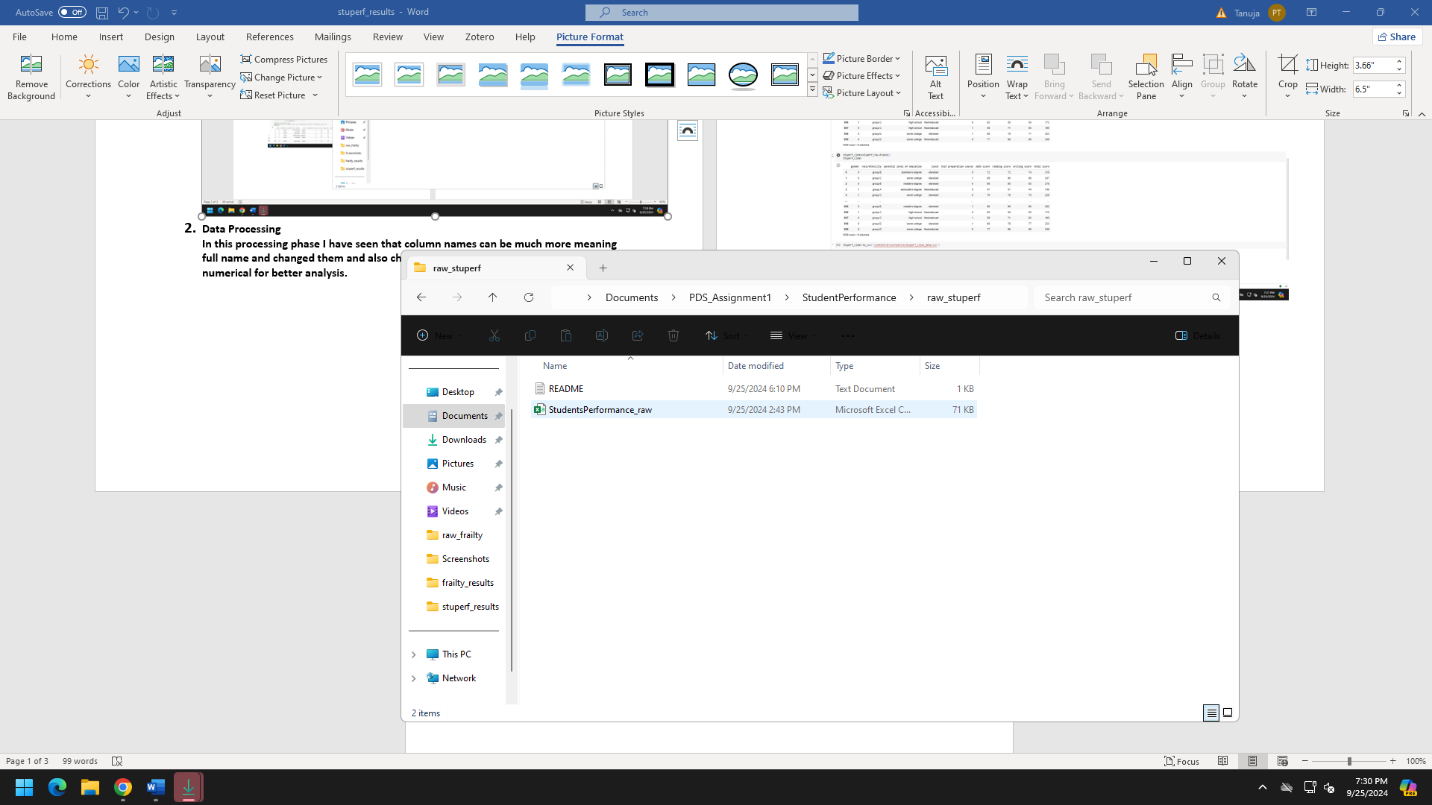
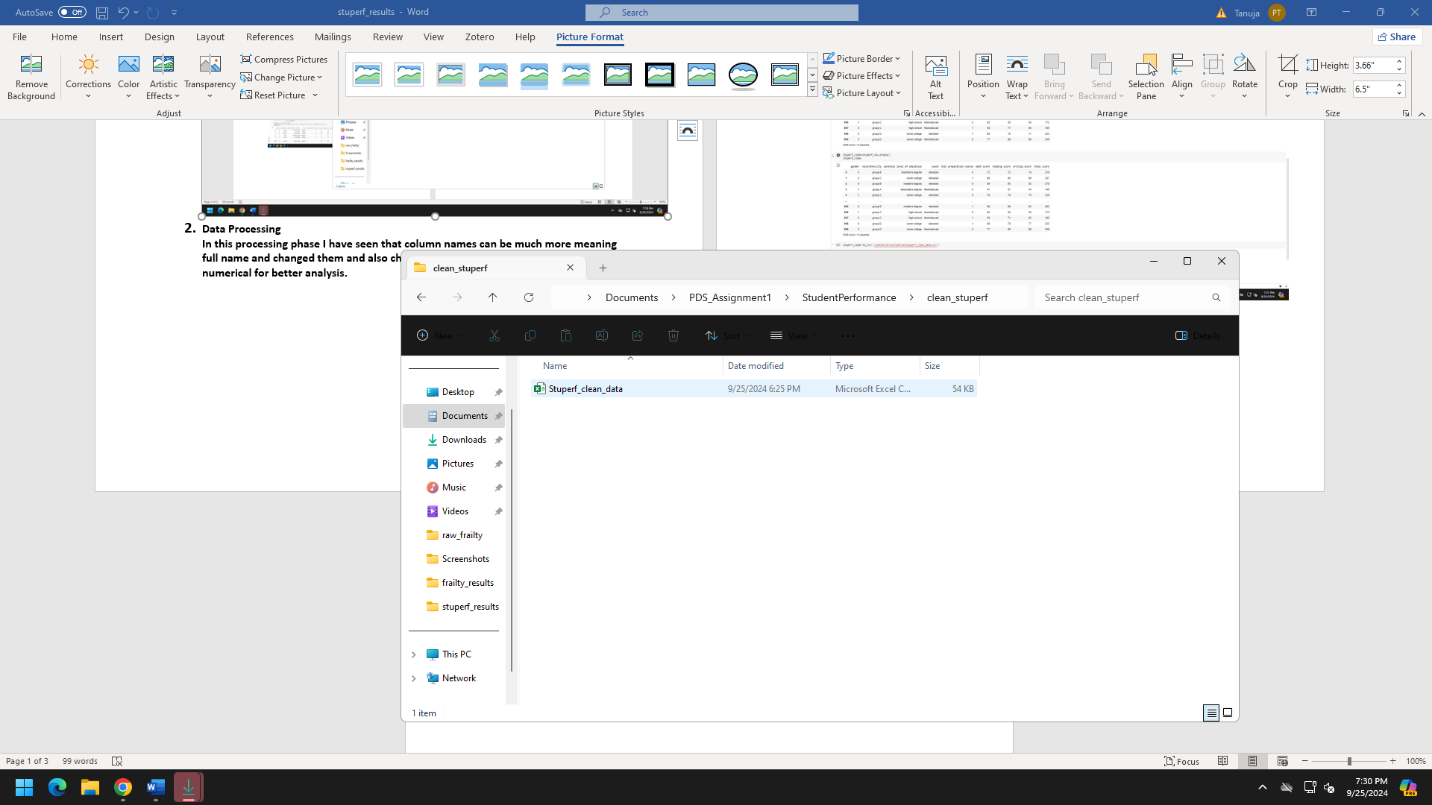
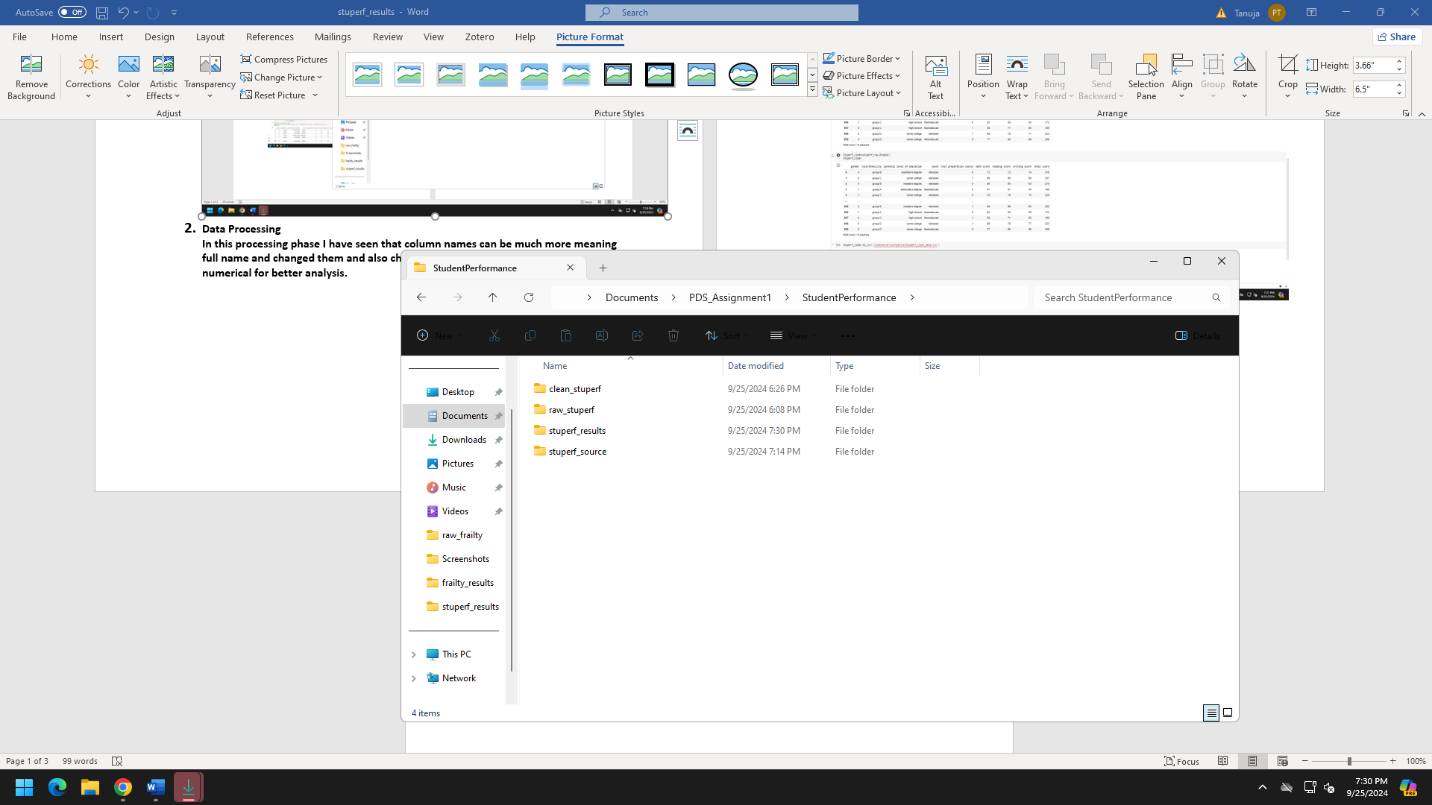
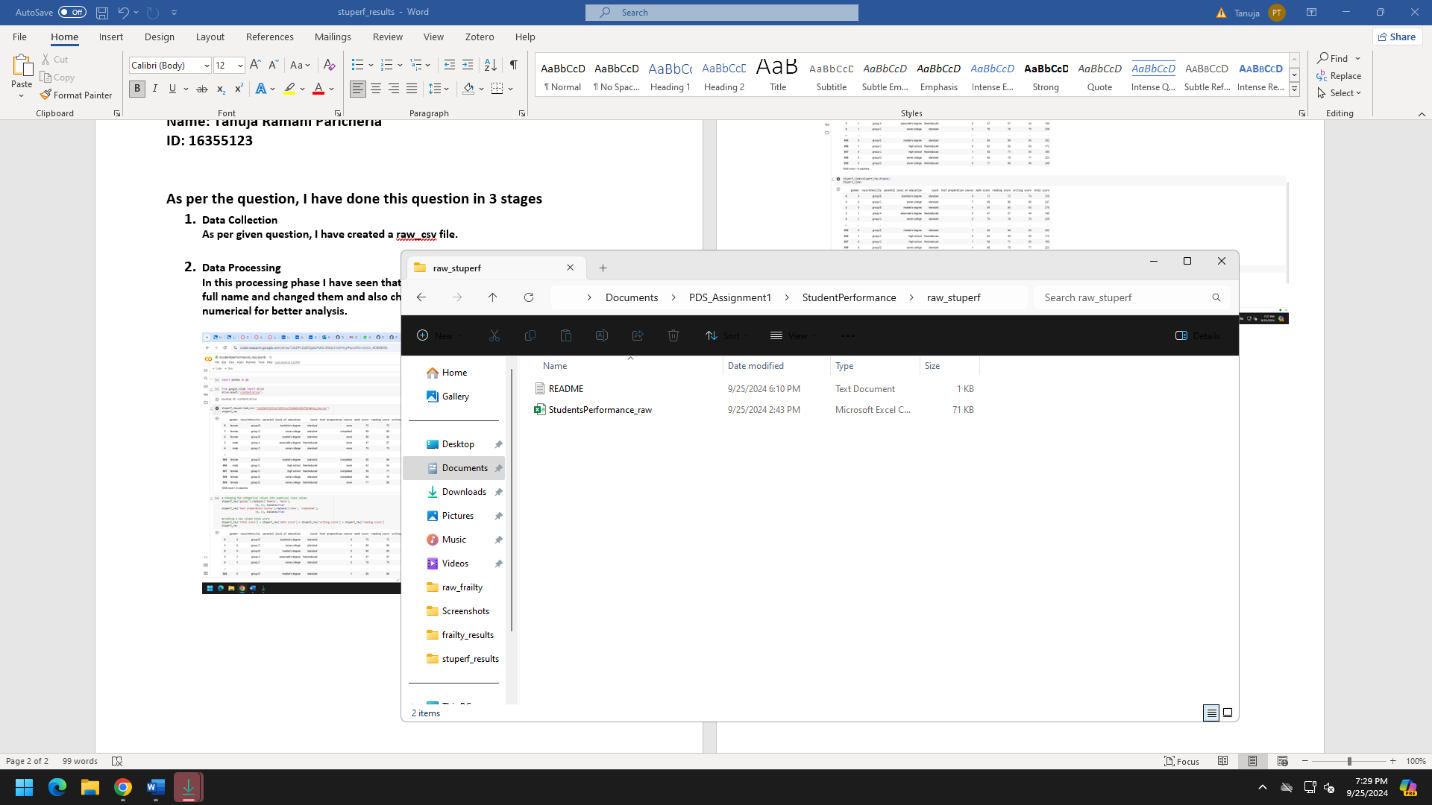
**6.Pairplot**

The scatter plots generally show upward trends, indicating strong positive relationships between the math ,reading,writing and Final score with respective to male and female geneders

**A screenshot of a computer

Description automatically generated**

**Project Structure:**

****