**Principles of Data Science – Assignment -1 – 16344982.**

**Q1.**

**First Stage – Data Collection**

a. The question's raw data is gathered and transformed into a.csv file.

b. The file is kept in a folder called "data\_raw."

**Second Stage – Data Processing**

a. The file is uploaded to Google Colab for processing, and the data is used to generate a data frame.

b. Data has already been cleansed; no pre-processing is required.

A screenshot of a computer

Description automatically generated

**Third Stage – Data Analysis**

a. Reviewed the overall data distribution across all features.

A black rectangular object with a black border

Description automatically generated

b. To determine if there is a pattern between the frail and non-frail categories, each feature's mean was performed with these two groups of women.

A black screen with a black rectangle

Description automatically generated

c. A correlation matrix was also created to check if any aspects might be eliminated to make the data less dimensional.

A black and white rectangular object

Description automatically generated

**Q2 - a**. **Bar plot for parental level education**: It gives information about the number of students and their parents' educational backgrounds, which may have an impact on the results.

A screenshot of a computer

Description automatically generated

**b. Math Scores by gender** - provides for the comparison of math operations scores and the view of gender differences.

A screenshot of a computer

Description automatically generated

**c. plot for reading scores by race/ethnicity** - plays a role in highlighting patterns and differences in data distribution within various groups together with their frequency distribution.

A screenshot of a computer

Description automatically generated

**d. count students for test preparation course** - presents you the number of students who finished the course, which may have an impact on the test results.

A screenshot of a computer

Description automatically generated

**e. pair plot of all scores –** provides in determining any relationship between the subjects.

**A screenshot of a computer

Description automatically generated**