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

**Name: Srujan Nagothu**

**ID: 16341576**

**2. Answer:**

Student’s performance data contains 7 columns with 1000 observations.

**Visualization 1: (Bar plot)**

Let us plot a **Bar plo**t graph between gender and frequency where x-axis showing the gender category and y-axis represents the count of each category of gender.

Chart, bar chart

Description automatically generated

**Analysis Outcome:** From this representation, we can conclude that the female student’s distribution is more when compared to male students in the dataset we have considered for analysis.

**Visualization 2: (Box plot)**

Let us plot the **Box plot** graph with gender along x-axis and scores on different subjects along y-axis.

Chart, box and whisker chart

Description automatically generated

**Analysis Outcome:** From this box plot representation, we can conclude that female students have performed better than male students in writing and reading subjects, whereas male students are performed better in math subject when compared to female students. On a total score consideration, female students achieved better scores when compared to male.

**Visualization 3: (Scatter plot)**

Let’s plot a scatter graph where x-axis contains math score, and y-axis contains the reading score of male and female students.

Chart, scatter chart

Description automatically generated

**Analysis Outcome:** From this scatter graph representation, it is clearing concluded that the female students are better performed than male students in reading whereas male students are superior to female students in math.

**Visualization 4: (Histogram)**

Let’s draw a plot with scores of math, reading, writing and total score on x-axis and its related frequency distribution on y-axis.

Chart, histogram

Description automatically generated

**Analysis Outcome:**  From the above frequency distribution, it is clearly showing that most of the students are getting scores in range of 65-80 in each subject which represents that the average marks obtained by students in each subject will be 65+.

**Visualization 5: (FacetGrid)**

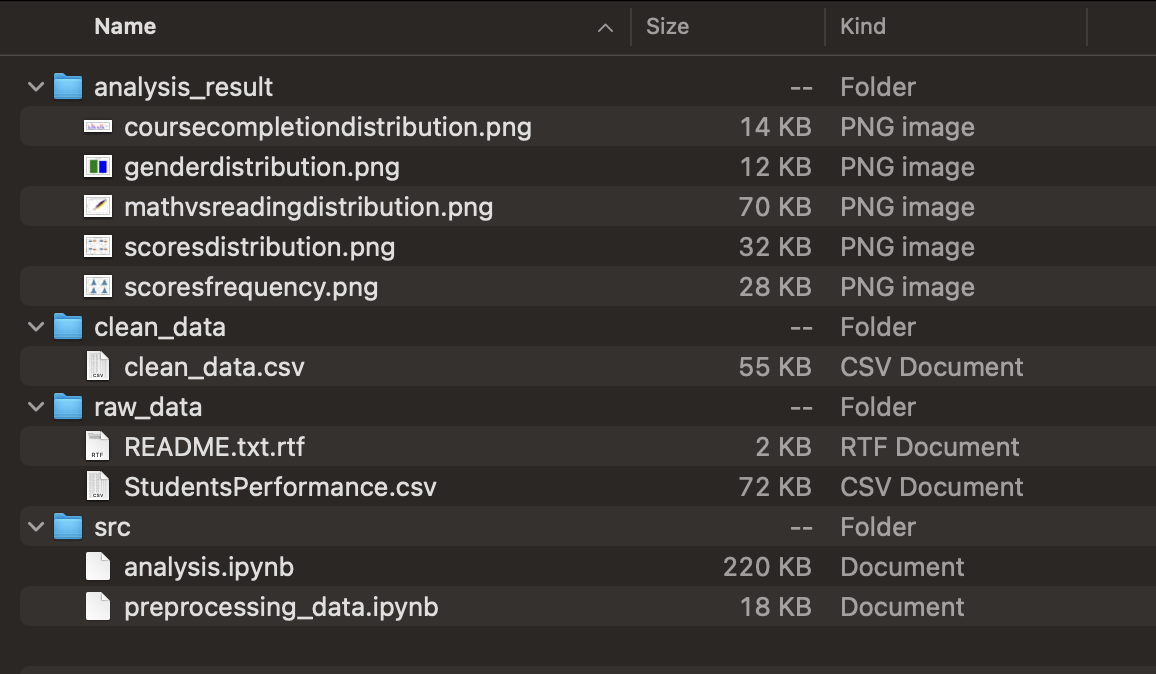
Let’s plot a graph with y-axis representing frequency and x-axis representing a race/ethnicity with a relationship of categorization based on gender and test completion status.

Chart, bar chart, waterfall chart

Description automatically generated

**Analysis Outcome:** This grid representation clearly shows that male and female students who have completed the course are mostly in Group C. And it also shows up that students who didn’t completed the course are in Group D for male students and Group C for female students.

**Folder Structure:**

****