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

**Name: Ankitha Srirama Reddy**

**ID: 16355578**

**2. Answer:**

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

**Visualization 1: (Bar plot)**

Let's create a bar graph connecting gender and frequency, with the x-axis representing the gender category and the y-axis the number of instances of each gender category.

Chart, bar chart

Description automatically generated

**Analysis Outcome:** We can infer from this representation that, in the dataset we have used for study, the distribution of female students is greater than that of male students.

**Visualization 2: (Box plot)**

Let's create a box plot graph where the y-axis represents topic scores and the x-axis represents gender.

Chart, box and whisker chart

Description automatically generated

**Analysis Outcome:** This box plot illustration leads us to the conclusion that while male students outperformed female students in arithmetic, female students outperformed male students in writing and reading. When the whole score was taken into account, female students fared better than male students.

**Visualization 3: (Scatter plot)**

Let's create a scatter graph using the math scores of the male and female students on the x-axis and the reading scores on the y-axis.

Chart, scatter chart

Description automatically generated

**Analysis Outcome:** It is evident from this scatter graph representation that male students outperform female students in math while female students outperform male students in reading.

**Visualization 4: (Histogram)**

Let's plot the math, reading, writing, and overall scores on the x-axis, along with the frequency distribution that corresponds to each score on the y-axis.

Chart, histogram

Description automatically generated

**Analysis Outcome:**  The frequency distribution shown above clearly demonstrates that the majority of students are receiving scores in the 65 to 80 range for each topic, indicating that the typical mark for each subject will be 65 or above.

**Visualization 5: (FacetGrid)**

Let's draw a graph with the frequency on the y-axis, the race or ethnicity on the x-axis, and the relationship between the classification based on gender and test completion status.

Chart, bar chart, waterfall chart

Description automatically generated

**Analysis Outcome:** The grid representation makes it very evident that Group C has the majority of the male and female course-completion students. Additionally, it reveals that Group D for male students and Group C for female students represent those students who did not finish the course.

**Folder Structure:**

**A screenshot of a computer

Description automatically generated**

**README.txt:**

**A close-up of a document

Description automatically generated**