**2a. Graphical Representation**

**Instructions:**

Please share your answers filled in-line in the word document. Submit code separately wherever applicable.

Please ensure you update all the details:

**Name: \_\_\_\_\_\_\_\_\_\_\_\_\_ Batch ID:** \_\_\_\_\_\_\_\_\_\_\_

**Topic: Data Visualization**

**Guidelines:**

**1. An assignment submission is considered complete only when the correct and executable code(s) is submitted along with the documentation explaining the method and results. Failing to submit either of those will be considered an invalid submission and will not be considered a correct submission.**

**2. Ensure that you submit your assignments correctly. Resubmission is not allowed.**

**3. Post the submission you can evaluate your work by referring to the keys provided. (will be available only post the submission).**

**Hints: Follow CRISP-ML(Q) methodology steps, where were appropriate.**

1. **Data Understanding: work on each feature of the dataset to create a data dictionary as displayed in the image below:**

Table

Description automatically generated

**Make a table as shown above and provide information about the features such as its data type and its relevance to the model building. And if not relevant, provide reasons and a description of the feature.**

**Problem Statements:**

1. Univariate plots for UNIV data (Plot must have Title, X & Y label)

A) Plot numerical column with 3 different plots ?

B) What are bin parameters? What are the methods to define the number of bins and bin sizes ?

C) Why do density plots exceed the range values of the column ?

D) Plot categorical columns by taking unique values ?

2. Bivariate graphs for UNIV data (Plot must be readable [use rotation], have all labels)

A) Plot 2 numerical columns with scatter plot [use grid] ?

B) 2 Different plots for plotting a numerical column with a categorical column (bar, line) ?

C) How are bar plots different from histogram?

3. Plot multivariate graphs (correlation heatmap, pairplot)

A) Plot for only numerical data ?

B) Plot multivariate graphs for both numerical and categorical columns ?

C) What does it mean when a correlation value says 1? When it is negative? When it is zero?

4. Plot Skewness & Probability distribution for each column of **marks data.** (Hist, box, density)

A) What is normally distributed and What will be the relationship between mean, median & mode ?

B) Which data variables are positively skewed and What will be the relationship between mean, median & mode

C) What are negatively skewed/distributed and What will be the relationship between mean, median & mode

D) What are the distinctive differences between skewness and distribution?