Assignment 2. Victor Opili BSE-05-027/2023

1. What is the difference between Univariate Bivariate and multivariate analyses in EDA analyses.

a) Univariate :- A statistical method used to describe and summarize a single variable within a dataset.

- It is crucial step to gain insights into the distribution, central tendency and ~~description~~ dispersion of a ~~vari~~ variable.

b) Bivariate :- A statistical method used to examine the relationship between two variables within a dataset.

- It is essencial for understanding how variables interact and influence each other.

c) Multivariate :- A statistical method used to determine the relationship among multiple variables simultaneously.

- It is essencial for understanding how variables interact and influence each other in a complex manner.

Assignment 2. Victor Opili BSE-05-0129/2022

3. What is an outlier and how to identify them?

- An outlier is a data point that significantly differs from the other observations in a dataset.
- They can indicate variability in your data, errors or interesting phenomena.

- They can be identified using the following methods

a) Visual methods like:
   i. Box plots and Scatter plot.

b. Statistical methods such as the Z-score which indicates how many standard deviations a data point is from the mean.

c. Rule of Thumb.
   Standard Deviation: Any data point ~~are~~ more than two standard deviations away from the mean can be flagged out as an outlier.

Assignment 2. Victor Opili BSE-05-0127/2022

2. During the data preprocessing step, how should one treat missing/null values? How will you deal with them?

a. Identifying missing values

b. Removing missing values: This can be done by dropping rows or columns.

c. Imputing missing values: fill with a specific value

d. Replacing missing values with the most frequent value in the column.