1.In an analytics project, what is the role and importance of data exploration and data preparation? How should data be presented?

Importance of data exploration:

Exploration allows for deeper understanding of a dataset, making it easier to navigate and use the data later. Successful exploration begins with an open mind, reveals new paths for discovery, and helps to identify and refine future analytics questions and problems.

Importance of data preparation:

Data preparation ensures accuracy in the data, which leads to accurate insights. Without data preparation, it's possible that insights will be off due to junk data, an overlooked calibration issue, or an easily fixed discrepancy between datasets.

Data will be presented as follows:

R and Python are the most common languages used for exploration. Data exploration doesn’t require coding for exploration through No-Code Platforms.

2.What insights can summary statistics provide?

Summary statistics is a part of descriptive statistics that summarizes and provides the gist of the information about the sample data. Statistics generally deals with the presentation of information quantitatively or visually.

3.A variety of data transformations and imputation methods exist. Please describe two of these methods.

Transformation Methods:

Constructive: The data transformation process adds, copies, or replicates the data.

Destructive: The system deletes fields or records.

Imputation methods:

Mean imputation:

Simply calculate the mean of the observed values for that variable for all individuals who are non-missing. It has the advantage of keeping the same mean and the same sample size, but many, many disadvantages. Pretty much every method listed below is better than mean imputation.

Substitution:

Impute the value from a new individual who was not selected to be in the sample. In other words, go find a new subject and use their value instead.