DAY - EDA PART 1 (Overview Interview Questions)

Q) What are the overall steps in Exploratory Data Analysis?

Ans - 4 overall steps in EDA - Step 1 → Data exploration and Preparation.
Step 2 → Missing Value treatment.
Step 3 → Outlier detection and Treatment.
Step 4 → Feature Engineering.

Step 1 - Data exploration and Preparation.
It mainly contains Uni-variate Analysis and Bi-Variate Analysis.

a) Univariate Analysis → Uni means One, so Single Variable Analysis.
i) Measure of central tendency → Mean, Median, Mode
ii) Measure of data spread → Quartile, percentile, Range, IQR, Boxplot, variance, SD
iii) Variation between variables → Covariance, Correlation Coefficient (pearson, spearman)
iv) Measure distribution & peakness → Skewness, Kurtosis.

b) Bivariate Analysis → Bi means Two, Two Variable Analysis.
Majorly two types of Data Variables are there Continuous & Categorical.
Possible combination - i) Continuous Vs Continuous → Correlation Coefficient.
ii) Categorical Vs Categorical → Chi-square test.
iii) Continuous Vs categorical → T test ($n < 30$), Z test ($n > 30$)
ANOVA test.

Step 2 → Missing Value treatment
Techniques for imputing Missing value - i) Continuous Data → Mean/Median imputation
ii) Categorical Data → Mode Imputation.
iii) KNN Imputation.

Step 3 → Outlier Detection & treatment
Outliers are the data point that differs significantly from other observation.
i) Outlier Detection techniques - Percentile, Box plot, Z Score.
ii) Remove Outlier techniques - Capping based on Upper and Lower Range.

Step 4 → Feature Engineering
Three major steps in Feature Engineering - Transformation, Scaling, Feature construction
i) Transformation - To Normalize the data, Methods → i) Log ii) Square root iii) Cube root
ii) Scaling - To standardize the data, Method → Min-Max Scaler. (Normalization, standardisation)
iii) Feature Construction - Creating feature based on original descriptors.
Methods - Binning, Encoding