Interview Q&A (With Sample Answers)

Q1. What is feature engineering, and why is it important?

Answer: Feature engineering involves creating new input features or transforming existing ones to improve model performance. It's crucial because better features often mean better predictive power.

Q2. When would you use Label Encoding vs One-Hot Encoding?

Answer:

- Use Label Encoding for ordinal variables (e.g., quality ratings).
- Use One-Hot Encoding for nominal variables (e.g., neighborhood names) without intrinsic order.

Q3. How do you handle missing values for categorical and numerical data?

Answer:

- For numerical data: Replace with mean/median.
- For categorical data: Replace with mode or use "Missing" as a category.

Q4. What are some common transformations to reduce skewness in data?

Answer:

- Log transformation: np.log1p()
- Box-Cox / Yeo-Johnson transformations
- Helps in making distributions more Gaussian for linear models.

Q5. How do you determine which features to keep or remove?

Answer:

- Use .feature_importances_ from tree-based models.
- Remove highly null or low variance columns.
- Use correlation analysis or Recursive Feature Elimination (RFE).