**Part 1: Data Cleaning**

**1. Load and Inspect Dataset**

* Load the dataset using Python (pandas).
* Display first 10 rows, data types, and summary statistics.

**2. Identify Missing Values**

* Check missing data per column.
* Visualize missing data using heatmap or isnull().sum().

**3. Handle Missing Values**

* Remove rows/columns if few values are missing.
* Impute missing values using mean, median, mode, or KNNImputer.

**4. Detect Outliers**

* Use **boxplots** or **Z-score/IQR method** to detect extreme values.

**5. Handle Outliers**

* Remove or transform outliers (log, square root, or truncation).
* Justify your approach.

**6. Fix Inconsistent Data**

* Standardize categorical entries ("Male", "male", "M").
* Correct invalid numeric values (negative ages).

**7. Remove Duplicate Rows**

* Detect duplicates and remove if necessary.

**8. Correct Data Types**

* Convert columns to appropriate types (numeric, categorical, datetime).

**9. Validate Numeric Ranges**

* Check that numeric values are within valid ranges (age > 0, score ≤ 100).

**10. Cross-Column Consistency Check**

* Validate relationships between columns (start date ≤ end date).

**Part 2: Data Transformation**

**11. Scaling and Normalization**

* Apply **Min-Max Scaling** and **Standardization** to numeric features.

**12. Encode Categorical Variables**

* Use **One-Hot Encoding** for non-ordinal features.
* Use **Label/Ordinal Encoding** for ordinal features.

**13. Identify Skewed Features**

* Visualize distributions and calculate skewness for numeric features.

**14. Apply Transformations**

* Reduce skew using **log, square root, Box-Cox, or Yeo-Johnson**.

**15. Create Polynomial Features**

* Generate squared or cubic terms for numeric features.

**16. Create Interaction Terms**

* Combine two or more features to create meaningful variables.

**17. Create Domain-Specific Features**

* Use knowledge of the dataset to engineer at least one new feature.
* Example: BMI = weight / height², or area = length × width.

**18. Binning / Discretization**

* Convert continuous features into categories (ex: age groups, income brackets).

**19. Quantile Transformation (Optional)**

* Transform numeric features into uniform or normal distribution.

**20. Truncation (Optional)**

* Set thresholds for extreme values to reduce their influence.