Phase-1:

1. It seems you already have all the required components for Phase-1.
2. In the notebook, please include clear bold headings (e.g., Data Extraction) as comments or Markdown titles so that supervisors can easily locate relevant sections of code during evaluation.

Phase 2: Preprocessing, EDA & Feature Engineering

Criterion	Description
Data Audit & Availability Check	Need to furnish lines of code on inspecting dataset shape, dtypes, counts, missing values%, duplicates, unique value counts, other relevant, and confirm availability of columns relevant to the task, etc. *As per needs of problem statement/data needs.
Exploratory Data Analysis (EDA)	Need to furnish lines of code and provide relevant summary statistics, distributions, and visualizations (histograms, boxplots, scatterplots, time-series plots where applicable, correlation tables) and derive insights/hypotheses tied to the problem statement and any other relevant. *As per needs of problem statement/data needs or support from the data. Note: If specific types of plots are not applicable include a one-line **Justification (Not Applicable)** in **bold** directly under the relevant notebook section as a **Markdown** cell.
Data Cleaning	Handle missing values (imputation/removal), duplicates, incorrect dtypes, outliers (with justification), irrelevant columns removal, and any dataset-specific fixes. If the dataset requires no cleaning, include a **one-line Justification (Not Applicable)** in **bold** immediately under the relevant notebook section **as a Markdown cell.** *As per the needs of problem statement/data needs or support from the data.

Criterion	Description
Feature Creation / Transformation	Furnish the lines of code to create or transform features that are appropriate to the problem statement and dataset (for example, lag features for time-series forecasting, RFM or aggregate metrics for segmentation, interaction terms for regression, and encodings for categorical variables). Ensure that each feature is relevant, justified, and useful. **NOTE:** If the dataset or problem legitimately requires only a few new features (for example, a Simple Linear Regression model with a single meaningful predictor), clearly document the reasoning with appropriate evidence or explanation. Include the justification in **bold** immediately under the relevant notebook section **as a Markdown cell**. *As per needs of problem statement/data (or support from the data) and ML Model needs.
Feature Selection / Dimensionality Reduction	Furnish the lines of code to apply the required methods to select a subset of features or reduce dimensionality (e.g., Filter Methods, Wrapper Methods, Embedded Methods, feature-importance ranking, correlation filtering, recursive feature elimination, or PCA, etc. whichever are relevant). Very brief description and justification on the approach used to select a subset of features or reduce dimensionality (e.g., Filter Methods, Wrapper Methods, Embedded Methods, feature-importance ranking, correlation filtering, recursive feature elimination, or PCA, etc.) in **bold** immediately under the relevant notebook section **as a Markdown cell**. **OR** If you have not performed any **Feature Selection / Dimensionality Reduction you must mention/justify why not done. Put them** in **bold** immediately under the relevant notebook section **as a Markdown cell** (like the problem statement or the Dataset does not require or supported these aspects…).
Feature Evaluation / Quick Checks	Furnish the lines of code that make simple checks demonstrating that the engineered features are meaningful where applicable. For prediction-based tasks (e.g., regression, forecasting, classification), this may include a **correlation with the target variable**, **simple feature-importance analysis**, or a **quick baseline model test** to show feature usefulness (whichever, you can). For **unsupervised tasks** (e.g., clustering or segmentation), provide a small **logical or data-driven justification** (e.g., better cluster separation, improved interpretability, or domain rationale) instead of correlation-based checks. **OR** If the problem type or dataset does not require or support such checks, then justify with a very few lines**. Put them** in **bold** immediately under the relevant notebook section **as a Markdown cell.**