Phase 3 – Data Exploration & Feature Engineering

Objectives

Analyze the collected data to understand its structure, uncover meaningful insights, and engineer useful features for modeling.

Tasks

1. **Exploratory Data Analysis (EDA)**
    - Compute descriptive statistics (mean, median, standard deviation, etc.).
    - Visualize data distribution and relationships using charts (histograms, scatter plots, correlation heatmaps, box plots, etc.).
    - Identify trends, outliers, anomalies, and potential issues.
2. **Insight Extraction**
    - Highlight important patterns or relationships relevant to your research problem.
    - Describe how these findings guide your next steps (modeling or deeper analysis).
3. **Feature Engineering**
    - Create new attributes from existing raw features (e.g., ratios, aggregated variables, domain-based transformations).
    - Encode categorical variables if needed.
    - Scale/normalize features where appropriate.
    - Justify why each engineered feature might improve performance.

Deliverables

A concise **EDA and Feature Engineering Report** including:

- Key statistics and summary tables
- Visualizations with clear explanations
- A list of extracted insights
- A table of selected features with description and justification