Milestone 2: Data Analysis and Visualization

Objectives

- Analyze the cleaned dataset to extract meaningful health insights.
- Use visualizations to support interpretation and decision-making in a healthcare context.

Tasks & Activities

1. Further Data Cleaning

- Re-examined for residual outliers or inconsistencies post initial cleaning.
- Confirmed absence of missing values and ensured proper encoding.
- Normalization validated using descriptive statistics.

2. Data Analysis

- Performed correlation analysis to assess the strength of relationships between features and the target outcome (cardio).
- Identified key predictors for cardiovascular disease, including:
    - Systolic and diastolic blood pressure
    - Cholesterol level
    - Age
    - BMI (calculated from height and weight)
- Applied statistical tests (e.g., t-test, chi-square) to validate associations between features and disease presence.

3. Data Visualization

- Developed the following visual assets:
    - **Heatmaps** showing correlation between features.
    - **Boxplots** comparing distributions by disease status.
    - **Count plots** for categorical feature distributions.
    - **Scatter plots** to show trends between BMI, age, and cardiovascular risk.
- Proposed the use of **interactive dashboards** (Plotly, Dash, or Tableau) for stakeholder presentations and real-time data exploration.

Deliverables

- **Cleaned Dataset & Analysis Report** detailing preprocessing steps and major findings.
- **Visualizations of Health Trends**:
    - Static and interactive plots demonstrating disease risk patterns.
    - Support tools for clinical decision-making.