**EDA:**

1. **Stats-based Analysis**
2. **Graph-Based Analysis**
3. **Pre-processing:**

**Model training**

1. **Training Pipeline**
2. **Prediction pipeline**

**EDA:**

1. **Stats-based Analysis (Mean, median, std)**
2. **Graph-Based Analysis (Different type of graphs)**
3. **Pre-processing:**
4. Handle Missing Value
5. Handel Duplicate Values
6. Basic Cleaning (bring data in to right values/format)
7. Outlier Handling
8. Handel Imbalance Data
9. Feature Selection
10. Dimension Reduction
11. Encoding
12. Scaling

**Handle** **Missing Value:**

1. Random value
2. Forward filling/backward filling
3. Statistical approach (mean. mode, median)
4. End of the distribution
5. Create “your own” ML model to predict missing value
6. Use mean, median, mode to replace the missing value

**Handel Duplicate Values:**

1. Drop the duplicate

**Handling imbalance Dataset**

1. Collect more data
2. Under sampling
3. Over sampling
4. Cluster based oversampling

**Feature Selection**

**Dimension Reduction**

1. (PCA, LDA, TSNE)
2. Split/Merge/Drop/Add

**Outlier Handling:**

1. Detect the Outlier
2. z-Score
3. IQR
4. Boxplot
5. Scatter plot
6. Violin plot
7. Removing Outlier
8. Drop Outlier
9. Replace with other value

**Scaling:**

1. Standardization
2. Min-max scaler

**Encoding**

1. One hot encoding
2. Ordinal Encoding