**Case Study: Phoenix - Financial Transaction Fraud Detection**

**Objective:** Identify fraudulent financial transactions using machine learning techniques to protect the bank’s assets and customers

**Background:** In the bustling city of Metropolis, the prestigious Phoenix Bank has been a pillar of financial stability for decades. However, a recent surge in fraudulent transactions has put the bank on high alert. The bank’s security team has uncovered a sophisticated fraud ring targeting their customers. The bank’s CEO has tasked the data science team with developing a robust fraud detection model to thwart these cybercriminals and safeguard the bank’s reputation.

**Dataset Overview:** The dataset contains financial transactions made by Phoenix Bank customers over a 2-day period and 284,807 transactions, of which 492 are identified as fraudulent, which translates to a fraud rate of 0.17%.

**Features:**

* **Time:** The number of seconds elapsed between each transaction and the first transaction in the dataset.
* **Amount:** The amount of the transaction, which can be used for cost-sensitive learning.
* **Feature1, Feature2, …, Feature28:** Principal components obtained through PCA (Principal Component Analysis).
* **Class:** The response variable, where 1 indicates a fraudulent transaction and 0 indicates a non-fraudulent transaction.

**Tasks Needed:**

1. **Exploratory Data Analysis (EDA):**
   * Analyze the distribution of the features.
   * Identify any patterns or anomalies in the data.
   * Visualize the data to understand the relationships between different features.
2. **Data Preprocessing:**
   * Handle missing values, if any.
   * Scale and preprocess the features if necessary.
3. **Model Building:**
   * Split the data into training and testing sets.
   * Train multiple models (e.g., Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, etc.).
   * Evaluate the models using appropriate metrics (e.g., Precision, Recall, F1-Score, ROC-AUC).
4. **Model Evaluation and Selection:**
   * Compare the performance of different models.
   * Select the best model based on evaluation metrics.
   * Discuss the trade-offs between different models.
5. **Presentation:** At minimum,
   * Prepare a PowerPoint presentation summarizing the findings from the EDA.
   * Explain the preprocessing steps and the rationale behind them.
   * Discuss the model building process and the performance of each model.
   * Justify the selection of the final model.