Final Assignment: Applying CRISP-DM Methodology to Credit Card Default Prediction

---

Task 1: Business Understanding

Problem Statement

The credit card industry faces significant challenges related to customer payment defaults, which can lead to substantial financial losses for credit card companies. To address this, the business problem we aim to solve is:
"How can we predict which customers are likely to default on their credit card payments to minimize financial losses and improve risk management?"

Importance

Predicting payment defaults is critical for credit card companies to:

- Reduce financial losses by proactively identifying high-risk customers.
- Optimize credit limits and interest rates for customers based on their risk profile.
- Enhance customer retention by offering tailored financial solutions to at-risk customers.

Stakeholders

- **Credit Card Companies:** To minimize losses and improve profitability.
- **Financial Analysts:** To gain insights into customer behavior and risk factors.
- **Customers:** To receive personalized financial support and avoid defaulting.

Success Metrics

- Develop a predictive model with **90% accuracy** in identifying customers likely to default.
- Reduce the default rate by **15%** within the first year of implementation.
- Achieve a **precision of 85%** to minimize false positives (customers incorrectly flagged as high-risk).

Task 2: Applying CRISP-DM Methodology

1. Business Understanding

The primary goal is to predict customer payment defaults to mitigate financial risks. This involves understanding the business context, defining the problem, and establishing success metrics.

2. Data Understanding

- **Data Sources:**
    - Customer transaction history.
    - Credit scores and demographic information.
    - Payment behavior (e.g., on-time payments, missed payments).
    - Credit utilization ratios.
- **Exploratory Data Analysis (EDA):**
    - Analyze distributions of key variables (e.g., credit scores, payment amounts).
    - Identify correlations between features and the target variable (default status).
    - Detect missing values, outliers, and data quality issues.

3. Data Preparation

- **Data Cleaning:**
    - Handle missing values using imputation techniques.
    - Remove or correct outliers.
- **Feature Engineering:**
    - Create new features such as:
        - *Payment-to-Income Ratio:* Monthly payment amount divided by income.
        - *Credit Utilization Trend:* Change in credit utilization over time.
        - *Missed Payment Frequency:* Number of missed payments in the last 12 months.

- **Data Transformation:**
    - Normalize numerical features.
    - Encode categorical variables (e.g., one-hot encoding).
- **Data Splitting:**
    - Split the dataset into training (70%) and testing (30%) sets.

4. Modeling

- **Model Selection:**
    - Evaluate models such as logistic regression, decision trees, random forests, and gradient boosting machines (e.g., XGBoost).
- **Training:**
    - Train models on the training dataset.
    - Use cross-validation to ensure robustness.
- **Hyperparameter Tuning:**
    - Optimize hyperparameters using techniques like grid search or random search.

5. Evaluation

- **Performance Metrics:**
    - Evaluate models using accuracy, precision, recall, F1-score, and ROC-AUC.
    - Focus on precision to minimize false positives (e.g., customers incorrectly flagged as high-risk).
- **Model Comparison:**
    - Compare models and select the best-performing one based on evaluation metrics.
- **Validation:**
    - Validate the model on the testing dataset to ensure generalizability.

6. Deployment

- **Integration:**

    - Deploy the model into the credit card company's system for real-time predictions.
- **Monitoring:**
    - Continuously monitor model performance and retrain as needed to account for changing customer behavior.
- **Reporting:**
    - Provide stakeholders with dashboards and reports to track key metrics (e.g., default rate, model accuracy).

Task 3: Summary and Expected Outcomes

Summary

In this project, we applied the CRISP-DM methodology to address the business problem of predicting credit card payment defaults. We began by defining the problem and understanding its importance to the business. We then collected and prepared the necessary data, built and evaluated predictive models, and outlined a plan for deployment and monitoring.

Expected Outcomes

- **Improved Risk Management:** The predictive model will enable the credit card company to identify high-risk customers proactively, reducing financial losses.
- **Enhanced Customer Experience:** By offering tailored financial solutions to at-risk customers, the company can improve customer satisfaction and retention.
- **Business Impact:** A 15% reduction in defaults is expected within the first year, saving the company an estimated $1M annually.