**Exploratory Data Analysis (EDA) Report**

**Bank Personal Loan Dataset**

**Advanced Techniques in Data Science Course Project**  
*Prepared by: Masooma Marriam*

**1. Problem Definition**

**Real-World Problem**

Thera Bank aims to expand its personal loan customer base by identifying existing depositors who are likely to accept personal loan offers. By analyzing customer data, the bank seeks to develop targeted marketing strategies to increase loan adoption rates.

***Significance***

* **Business Optimization**: Enhance marketing efficiency by focusing on customers with a higher propensity to accept personal loans.
* **Customer Satisfaction**: Tailor loan offers to meet the needs of potential customers, improving engagement and satisfaction.
* **Operational Efficiency**: Allocate resources effectively by identifying key customer segments for loan promotions.

**2. Asking the Right Questions**

To guide the analysis, the following key questions were formulated:

1. What are the demographic and financial characteristics of customers who accepted personal loan offers?
2. How do factors such as income, family size, and education level influence the likelihood of accepting a personal loan?
3. Are there specific customer segments that exhibit higher acceptance rates for personal loans?
4. What is the correlation between customers' existing financial products (e.g., securities accounts, CD accounts) and their acceptance of personal loans?
5. How can the bank effectively target potential customers for future personal loan campaigns?

**3. Data Collection**

**Source: Bank Personal Loan Dataset from Kaggle.**

**Description: The dataset contains information on 5,000 customers, including demographic details, financial profiles, and their response to a personal loan offer.**

**Features:**

* **ID: Unique customer identifier.**
* **Age: Customer's age in years.**
* **Experience: Years of professional experience.**
* **Income: Annual income in $000.**
* **ZIP Code: Residential ZIP code.**
* **Family: Family size.**
* **CCAvg: Average monthly credit card spending in $000.**
* **Education: Education level (1: Undergraduate, 2: Graduate, 3: Advanced/Professional).**
* **Mortgage: Value of house mortgage in $000.**
* **Personal Loan: Indicator of loan acceptance (0: No, 1: Yes).**
* **Securities Account: Indicator of having a securities account (0: No, 1: Yes).**
* **CD Account: Indicator of having a certificate of deposit account (0: No, 1: Yes).**
* **Online: Usage of internet banking (0: No, 1: Yes).**
* **CreditCard: Usage of bank-issued credit card (0: No, 1: Yes).**

**4. Data Wrangling (Preprocessing)**

Several preprocessing steps were undertaken to ensure data usability:

* **Handling Missing Values**: The dataset was complete with no missing values.
* **Data Type Conversion**: Converted categorical variables (e.g., Education, Family) to appropriate data types for analysis.
* **Feature Engineering**:
  + Created a new feature, **Income\_Category**, to segment customers into income brackets for detailed analysis.
  + Generated interaction terms between features such as Income and Education to capture combined effects.

**5. Exploratory Data Analysis (EDA)**

***5.1 Personal Loan Acceptance Distribution***

* **Objective**: Understand the proportion of customers who accepted the personal loan offer.
* **Insight:** Approximately 9.6% of customers accepted the loan offer, indicating an imbalanced target variable.

***5.2 Demographic Analysis***

* **Age and Experience:**
  + **Objective:** Examine the distribution of age and professional experience among customers.
  + **Insight:** Both age and experience are approximately normally distributed, with means around 45 years and 20 years, respectively.
* **Family Size:**
  + **Objective:** Analyze the distribution of family sizes.
  + **Insight:** Most customers have a family size of 1 to 4 members, with a slight skew towards smaller families.

***5.3 Financial Profile Analysis***

* **Income:**
  + **Objective:** Assess the income distribution among customers.
  + **Insight:** Income distribution is right-skewed, with a mean annual income of approximately $73,000.
* **Credit Card Spending (CCAvg):**
  + **Objective:** Understand customers' average monthly credit card spending.
  + **Insight:** The majority of customers have average monthly spending between $0 and $5,000, with a few high spenders.

***5.4 Relationship Between Features and Loan Acceptance***

* **Income vs. Personal Loan:**
  + **Objective:** Determine how income levels affect loan acceptance.
  + **Insight:** Higher income groups show a greater propensity to accept personal loans.
* **Education vs. Personal Loan:**
  + **Objective:** Analyze the impact of education level on loan acceptance.
  + **Insight:** Customers with advanced/professional education levels have higher acceptance rates.
* **Credit Card Usage vs. Personal Loan:**
  + **Objective:** Examine the relationship between credit card usage and loan acceptance.
  + **Insight:** Customers using the bank's credit card are more likely to accept personal loan offers.

**6. Visual Insights**

**The following visualizations supported the analysis:**

* **Histograms:** Displayed the distribution of customer ages, incomes, and credit card spending, highlighting central tendencies and variability.
* **Bar Charts:** Illustrated the count of personal loan acceptances across different education levels and family sizes, revealing trends in loan acceptance.
* **Scatter Plots:** Showed relationships between income and average credit card spending, as well as income and mortgage values, identifying potential correlations.
* **Heatmaps:** Presented correlation matrices of numerical features, identifying significant relationships between variables influencing loan acceptance.

**7. Conclusion**

The Exploratory Data Analysis (EDA) of the Bank Personal Loan Dataset revealed significant insights into customer demographics and financial behaviors:

* **Customer Profiling:** Identified key characteristics of customers more likely to accept personal loan offers, such as higher income levels and advanced education.
* **Feature Relationships:** Uncovered correlations between variables, aiding in understanding factors influencing loan acceptance.
* **Data Quality:** Assessed the completeness and integrity of the dataset, ensuring reliability for further analysis.