Project 2: Machine Learning

--

— Personal Loan Campaign—

Tiffany Rhodes

January 10, 2025

Contents / Agenda

Executive Summary

- Our goal was to optimize personal loan campaigns by leveraging data-driven insights. The analysis focused on customer behavior to identify factors that significantly predict loan acceptance.

- Using machine learning models, we provided actionable insights and strategies for optimizing targeting, thereby reducing false negatives and increasing campaign efficiency. The findings underscore the importance of variables like income, family size, and education while highlighting areas for improvement in online engagement and targeted marketing efforts.

Executive Summary

Key Insights:

- **Income and Family Size:** Customers with higher income and larger family sizes are significantly more likely to accept personal loans.
- **Education:** Education emerged as the most critical factor in decision tree models, influencing loan uptake.
- **Online Banking:** Customers using online banking are less inclined to accept personal loans, suggesting an opportunity to revamp digital marketing strategies.
- **Credit Card Usage:** Customers with higher monthly credit card spending are more likely to purchase a loan, but those with multiple credit cards are less likely.
- **Campaign Optimization:** Focus on reducing false negatives by targeting untapped segments, such as customers without securities accounts and with higher credit card expenditures.

Objectives

- To develop a predictive model to determine the likelihood of a liability customer purchasing a personal loan.
- To analyze customer attributes to identify the key factors that influence loan purchase decisions.
- To assist in segmenting the customer base, thereby enabling the identification of high-priority target segments for marketing and sales effort.

Business Problem Overview

- This case involves AllLife Bank, whose management seeks to explore strategies for converting its liability customers into personal loan customers, while maintaining their status as depositors. A campaign conducted by the bank last year targeting liability customers achieved a successful conversion rate exceeding 9%.
- This outcome has motivated the retail marketing department to design more targeted campaigns aimed at further improving the conversion rate, all while minimizing the budget allocation.

Business Problem Overview

The financial institution seeks to maximize the success of its personal loan campaigns by identifying the customers most likely to accept such offers. Misidentifying potential customers results in two major losses:

- **Opportunity Loss**: Customers interested in personal loans are overlooked.
- **Resource Loss**: Resources are spent targeting uninterested customers.

The primary goal is to minimize false negatives (missed opportunities) while ensuring resource efficiency in marketing efforts.

Data Dictionary

- **ID:** Customer ID
- **Age:** Customer's age in completed years
- **Experience:** #years of professional experience
- **Income:** Annual income of the customer (in thousand dollars)
- **ZIP Code:** Home Address ZIP code.
- **Family:** the Family size of the customer
- **CCAvg:** Average spending on credit cards per month (in thousand dollars)
- **Education:** Education Level. 1: Undergrad; 2: Graduate;3: Advanced/Professional
- **Mortgage:** Value of house mortgage if any. (in thousand dollars)
- **Personal_Loan:** Did this customer accept the personal loan offered in the last campaign?
- **Securities_Account:** Does the customer have securities account with the bank?
- **CD_Account:** Does the customer have a certificate of deposit (CD) account with the bank?
- **Online:** Do customers use internet banking facilities?
- **CreditCard:** Does the customer use a credit card issued by any other Bank (excluding All life Bank)?

Data Quality Summary

Attribute	Count
Shape	(5000, 14)
Missing Values	0
Duplicate Values	0

Data Types

```
Data columns (total 14 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   ID                  5000 non-null   int64  
 1   Age                 5000 non-null   int64  
 2   Experience          5000 non-null   int64  
 3   Income              5000 non-null   int64  
 4   ZIPCode             5000 non-null   int64  
 5   Family              5000 non-null   int64  
 6   CCAvg               5000 non-null   float64
 7   Education           5000 non-null   int64  
 8   Mortgage            5000 non-null   int64  
 9   Personal_Loan       5000 non-null   int64  
 10  Securities_Account  5000 non-null   int64  
 11  CD_Account          5000 non-null   int64  
 12  Online              5000 non-null   int64  
 13  CreditCard          5000 non-null   int64  
dtypes: float64(1), int64(13)
memory usage: 547.0 KB
```

Statistical Summary

	ID	Age	Experience	Income	ZIPCode	Family	CCAvg
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000
mean	2500.500000	45.338400	20.104600	73.774200	93169.257000	2.396400	1.937938
std	1443.520003	11.463166	11.467954	46.033729	1759.455086	1.147663	1.747659
min	1.000000	23.000000	-3.000000	8.000000	90005.000000	1.000000	0.000000
25%	1250.750000	35.000000	10.000000	39.000000	91911.000000	1.000000	0.700000
50%	2500.500000	45.000000	20.000000	64.000000	93437.000000	2.000000	1.500000
75%	3750.250000	55.000000	30.000000	98.000000	94608.000000	3.000000	2.500000
max	5000.000000	67.000000	43.000000	224.000000	96651.000000	4.000000	10.000000

	Education	Mortgage	Personal_Loan	Securities_Account	CD_Account	Online	CreditCard
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.00000	5000.000000	5000.000000
mean	1.881000	56.498800	0.096000	0.104400	0.06040	0.596800	0.294000
std	0.839869	101.713802	0.294621	0.305809	0.23825	0.490589	0.455637
min	1.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000
25%	1.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.000000
50%	2.000000	0.000000	0.000000	0.000000	0.00000	1.000000	0.000000
75%	3.000000	101.000000	0.000000	0.000000	0.00000	1.000000	1.000000
max	3.000000	635.000000	1.000000	1.000000	1.00000	1.000000	1.000000

Outlier Percentage

```
Percentage of outliers in each numeric column:
 ID          0.00
Age          0.00
Experience   0.00
Income       1.92
Family       0.00
CCAvg        6.48
Mortgage     5.82
dtype: float64
```

Training Set

```
Shape of Training set :  (3500, 18)
Shape of test set :  (1500, 18)
Percentage of classes in training set:
Personal_Loan
0    0.905429
1    0.094571
Name: proportion, dtype: float64
Percentage of classes in test set:
Personal_Loan
0    0.900667
1    0.099333
Name: proportion, dtype: float64
```

Solution Approach & Methology

Great Learning
POWER AHEAD

1. **Exploratory Data Analysis (EDA):** Identified key patterns, distributions, and outliers in variables like income, credit card spending (CCAvg), and mortgage value. Converted ZIP codes to cities and treated outliers in significant variables.

2. **Model Development:**
    a. Logistic regression identified significant predictors, such as income, family size, and CD account ownership.
    b. Decision tree models highlighted the importance of education in predicting loan acceptance.
    c. Hyperparameter tuning and cost-complexity pruning ensured generalization and improved recall.

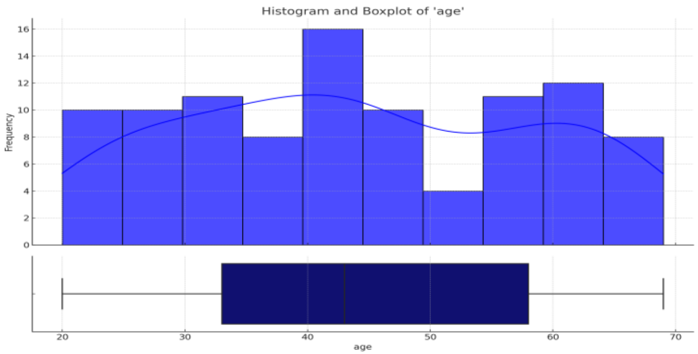
3. **Evaluation Metrics:** Focused on recall to minimize opportunity loss. The best test recall achieved was 86%, though precision was relatively low at 52%.

4. **Recommendations:** Developed strategies targeting customer segments with high potential for loan acceptance, such as larger families and individuals with fewer credit cards but higher monthly spending.

EXPLORATORY DATA ANALYSIS

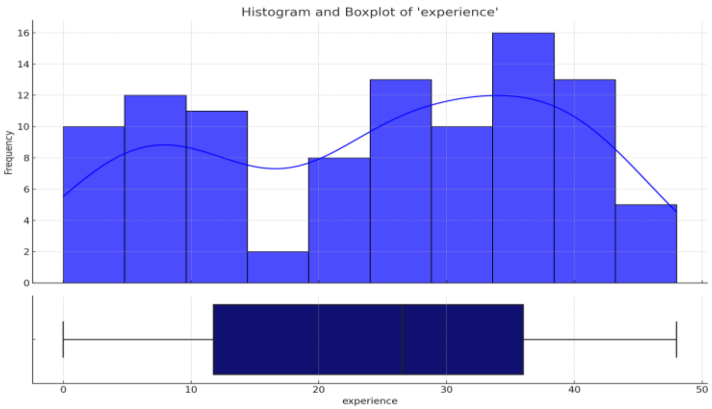
EDA Results

Age



- The age distribution is normal with no outliers.
- Customer ages range from 23 to 65 years.
- The average age of customers is 45 years.
- Half of the customers are between 35 and 55 years old..

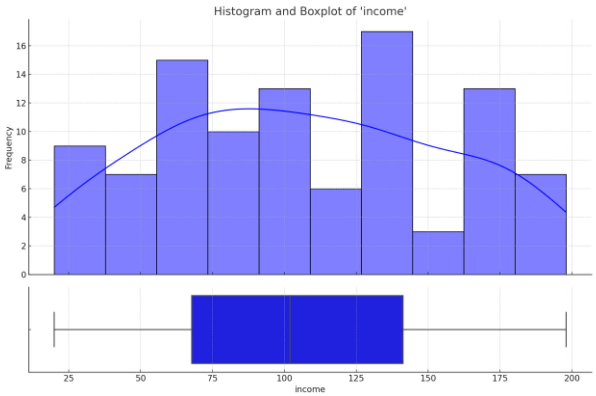
Experience



- The years of experience follow a normal distribution with no outliers.
- Experience ranges from 0 to 43 years.
- The average customer has 20 years of experience.
- Half of the customers have experience between 10 and 30 years.

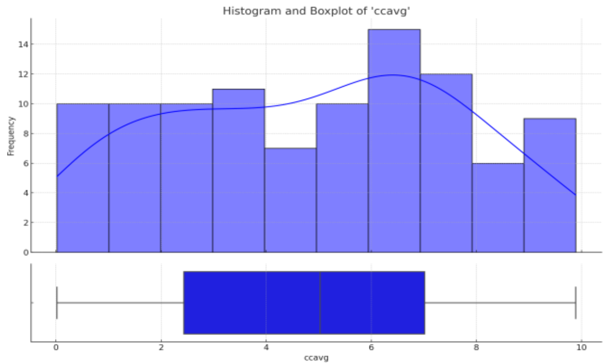
EDA Results

INCOME



- The income distribution is **right-skewed**, with most customers earning in the lower-to-middle ranges.
- The mean ($73.77K) is higher than the median due to the presence of high-income outliers.
- The histogram complements the box plot by visualizing the data distribution and confirming the prevalence of high-income outliers.

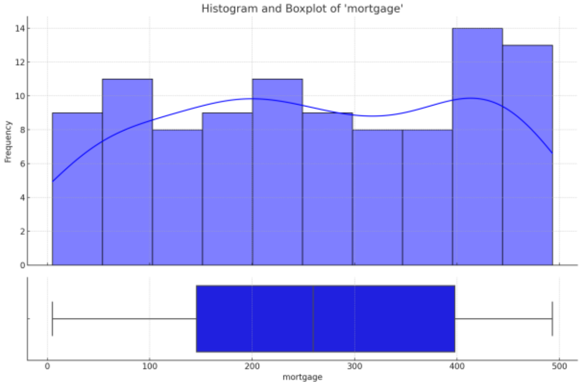
CCAvg



- The "CCAvg" distribution is **right-skewed**, with most individuals having low average credit card usage.
- A few individuals with significantly higher average usage (outliers) influence the distribution.
- The **mean value** is higher than the median due to the skewness caused by outliers.

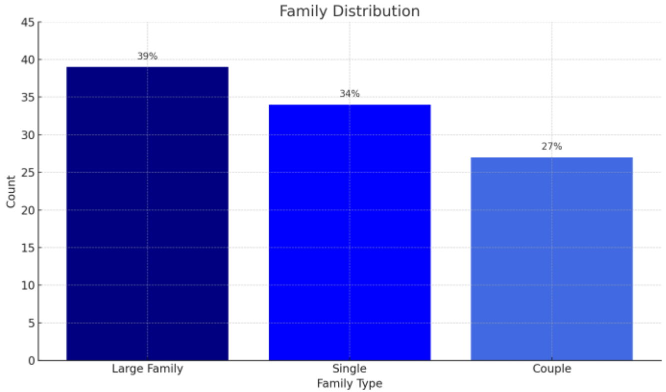
EDA Results

Mortgage



- The mortgage distribution is **right-skewed** with outliers on the higher end.
- Mortgage values range from \$0 to \$635K.
- The average mortgage value is \$56.49K.
- 75% of customers have a mortgage value between \$0 and \$101K.

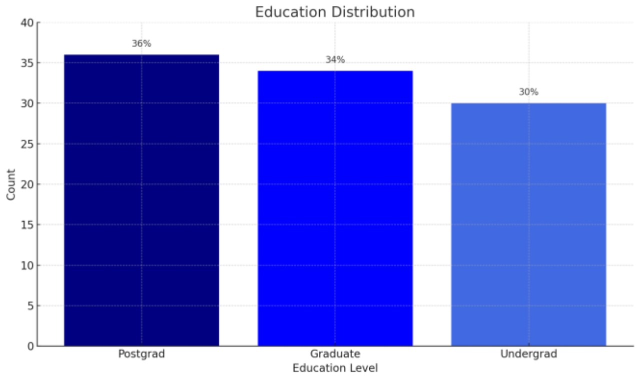
Family



- Family size ranges from 1 to 4 members.
- Nearly 30% of customers live in single-member households.
- On average, customers have a family size of 2 members.

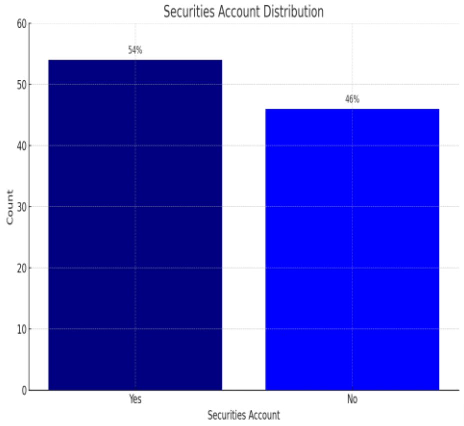
EDA Results

Education



- Customers with higher education levels (Professional and Graduate) have a 60%-40% proportion compared to Undergraduate customers.
- One-third of the customers have advanced or professional careers.

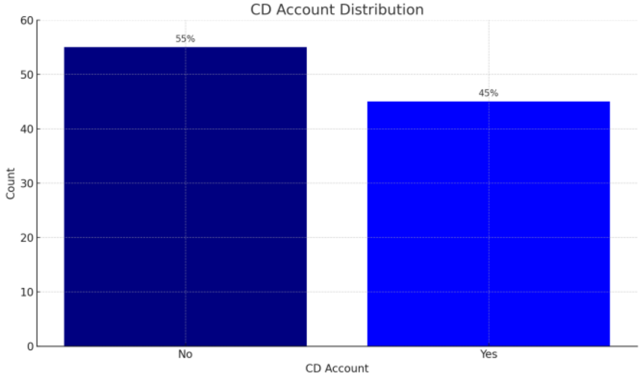
Securities Account



- Only 10.4% of the customers are investing their money through a portfolio of some kind.

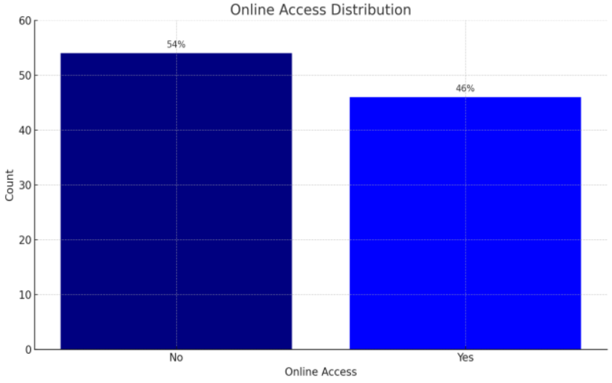
EDA Results

CD_Account



- Only 6% of customers with Certificates of Deposit may face potential liquidity issues due to funds being tied up for a fixed period.

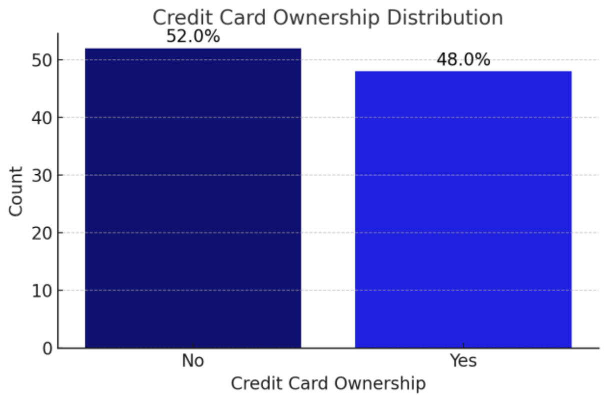
Online



- Nearly 60% of customers utilize online banking services.

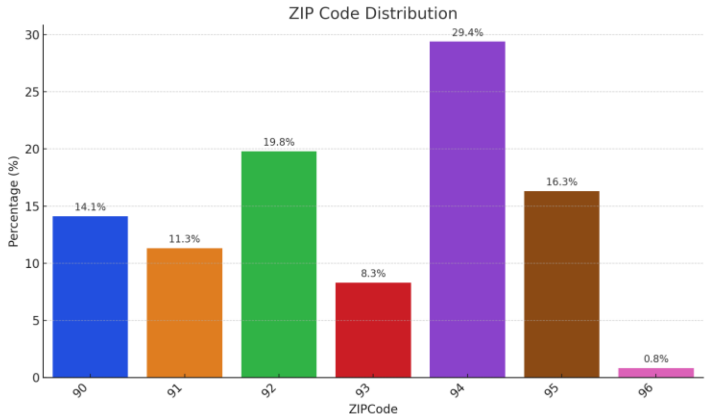
EDA Results

CreditCard



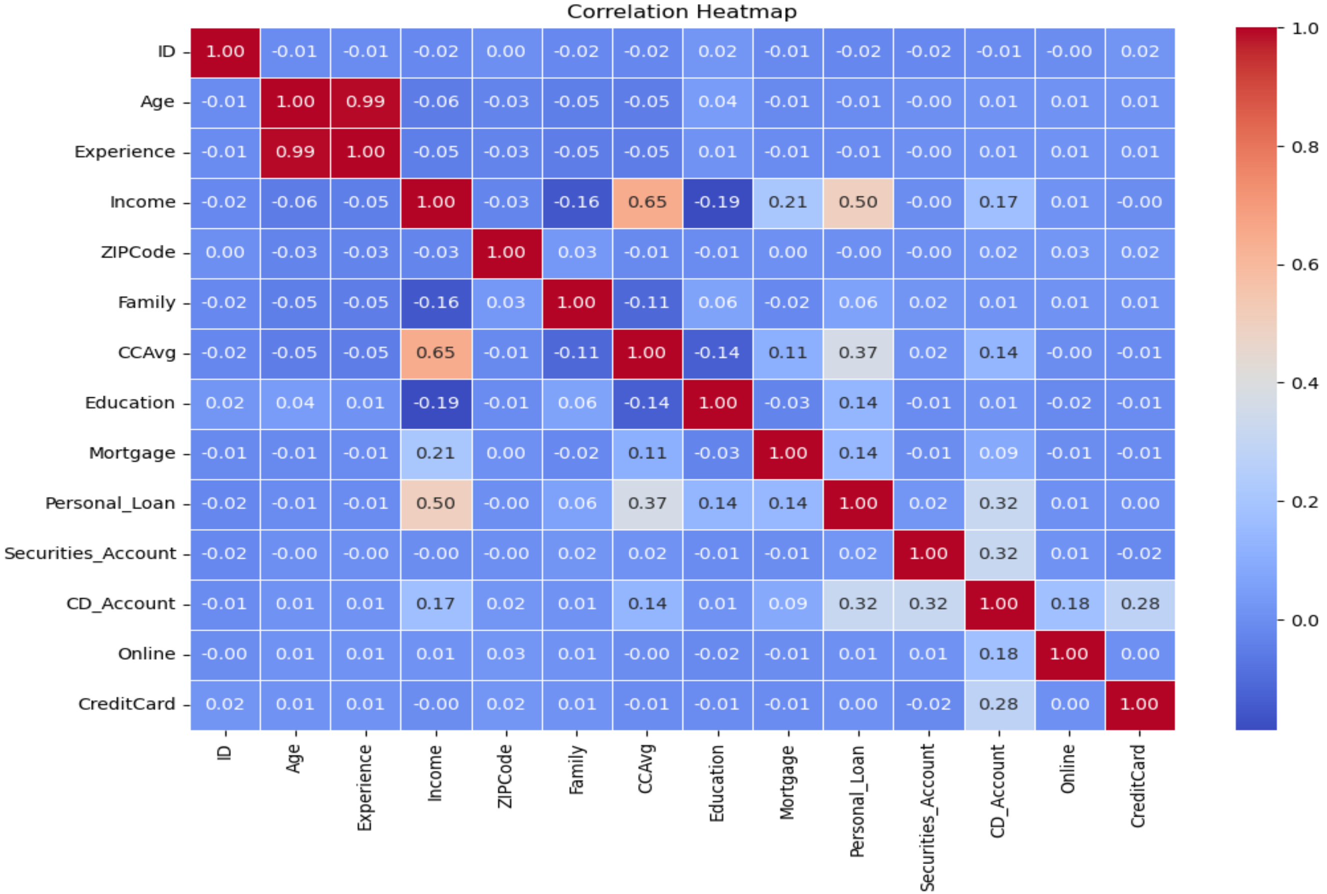
- Approximately 30% of customers already hold a credit card issued by another bank.

ZIPCode



- The distribution of the 7 clusters was created by grouping customers based on the first two digits of their ZIP Code, highlighting the main cities associated with each cluster.
- The largest cluster is cluster 94

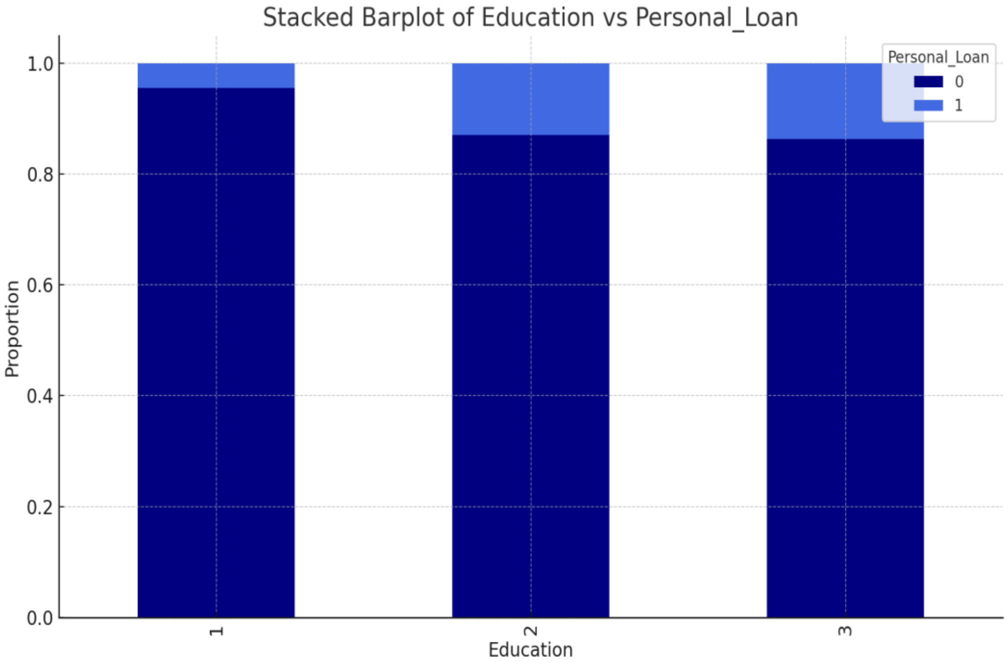
EDA Results


Correlation Heatmap
ID 1.00 -0.01 -0.01 -0.02 0.00 -0.02 -0.02 0.02 -0.01 -0.02 -0.02 -0.01 -0.00 0.02
Age -0.01 1.00 0.99 -0.06 -0.03 -0.05 -0.05 0.04 -0.01 -0.01 -0.00 0.01 0.01 0.01
Experience -0.01 0.99 1.00 -0.05 -0.03 -0.05 -0.05 0.01 -0.01 -0.01 -0.00 0.01 0.01 0.01
Income -0.02 -0.06 -0.05 1.00 -0.03 -0.16 0.65 -0.19 0.21 0.50 -0.00 0.17 0.01 -0.00
ZIPCode 0.00 -0.03 -0.03 -0.03 1.00 0.03 -0.01 -0.01 0.00 -0.00 -0.00 0.02 0.03 0.02
Family -0.02 -0.05 -0.05 -0.16 0.03 1.00 -0.11 0.06 -0.02 0.06 0.02 0.01 0.01 0.01
CCAvg -0.02 -0.05 -0.05 0.65 -0.01 -0.11 1.00 -0.14 0.11 0.37 0.02 0.14 -0.00 -0.01
Education 0.02 0.04 0.01 -0.19 -0.01 0.06 -0.14 1.00 -0.03 0.14 -0.01 0.01 -0.02 -0.01
Mortgage -0.01 -0.01 -0.01 0.21 0.00 -0.02 0.11 -0.03 1.00 0.14 -0.01 0.09 -0.01 -0.01
Personal_Loan -0.02 -0.01 -0.01 0.50 -0.00 0.06 0.37 0.14 0.14 1.00 0.02 0.32 0.01 0.00
Securities_Account -0.02 -0.00 -0.00 -0.00 -0.00 0.02 0.02 -0.01 -0.01 0.02 1.00 0.32 0.01 -0.02
CD_Account -0.01 0.01 0.01 0.17 0.02 0.01 0.14 0.01 0.09 0.32 0.32 1.00 0.18 0.28
Online -0.00 0.01 0.01 0.01 0.03 0.01 -0.00 -0.02 -0.01 0.01 0.01 0.18 1.00 0.00
CreditCard 0.02 0.01 0.01 -0.00 0.02 0.01 -0.01 -0.01 -0.01 0.00 -0.02 0.28 0.00 1.00
ID Age Experience Income ZIPCode Family CCAvg Education Mortgage Personal_Loan Securities_Account CD_Account Online CreditCard
1.0
0.8
0.6
0.4
0.2
0.0

EDA Results

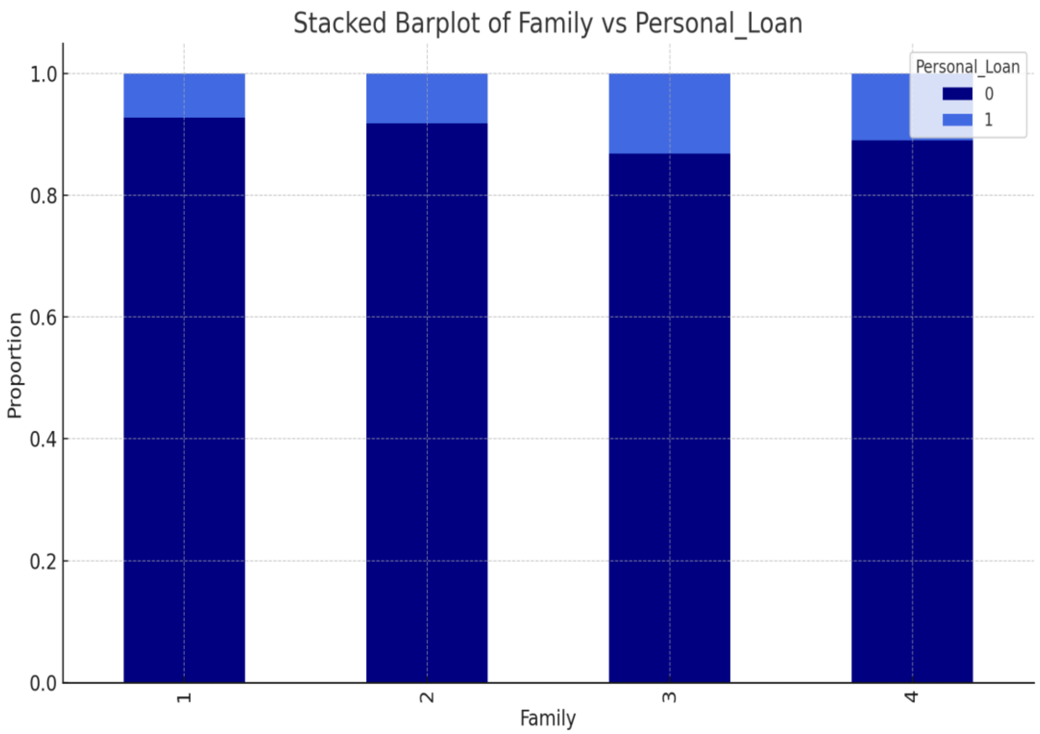
- **Strong Predictors of Personal Loan Uptake**: Income, CCAvg (credit card spending), and CD_Account ownership are the strongest predictors of personal loan uptake.
- **Irrelevant Predictors**: Variables like ZIP Code and Online banking usage show no significant correlation with personal loan uptake.
- **Inter-Variable Relationships**: Age and experience are highly correlated.
- *Income correlates positively with spending (CCAvg) and financial product ownership (CD_Account).*

EDA Results



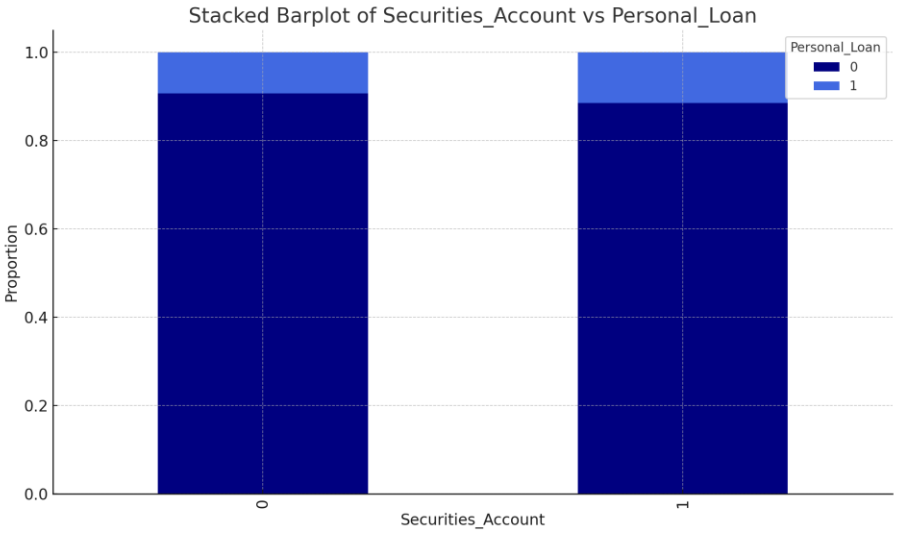
- Higher Education Linked to Higher Borrowing Rates: The proportion of borrowers increases with education level, indicating that education might be a contributing factor to personal loan uptake.
- Non-Borrowers Still Dominate: Regardless of education level, the majority of individuals do not take personal loans.
- Education Likely Enhances Loan Accessibility: Higher education may correlate with higher income or creditworthiness, making individuals more eligible for loans.

EDA Results



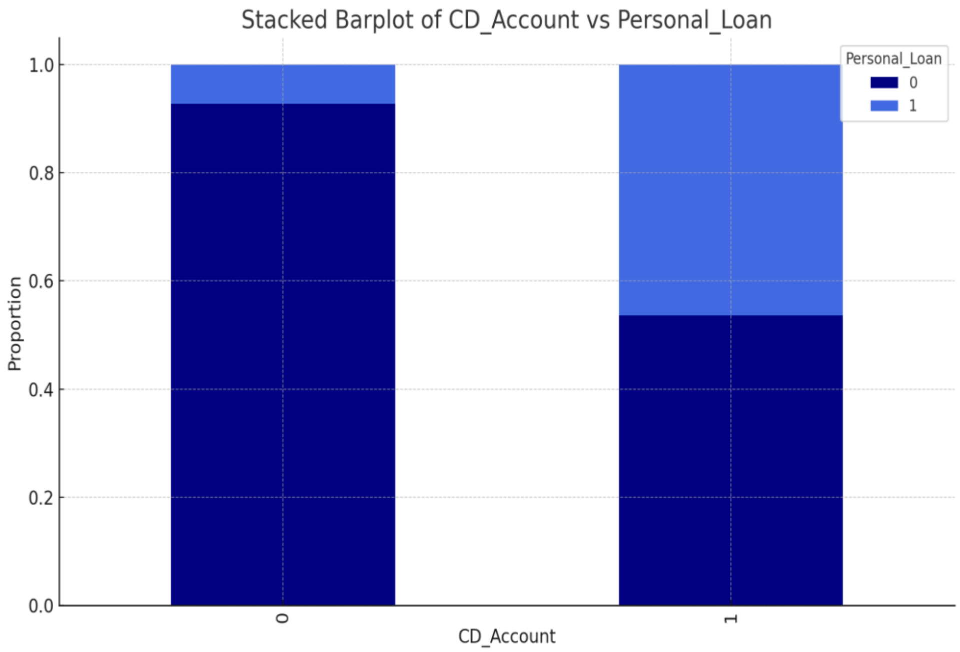
- Individuals with larger family sizes (e.g., 4) are slightly more likely to take personal loans compared to those with smaller family sizes.
- However, non-borrowers consistently form the majority across all family sizes, implying that family size alone is not a strong predictor of loan uptake.

EDA Results



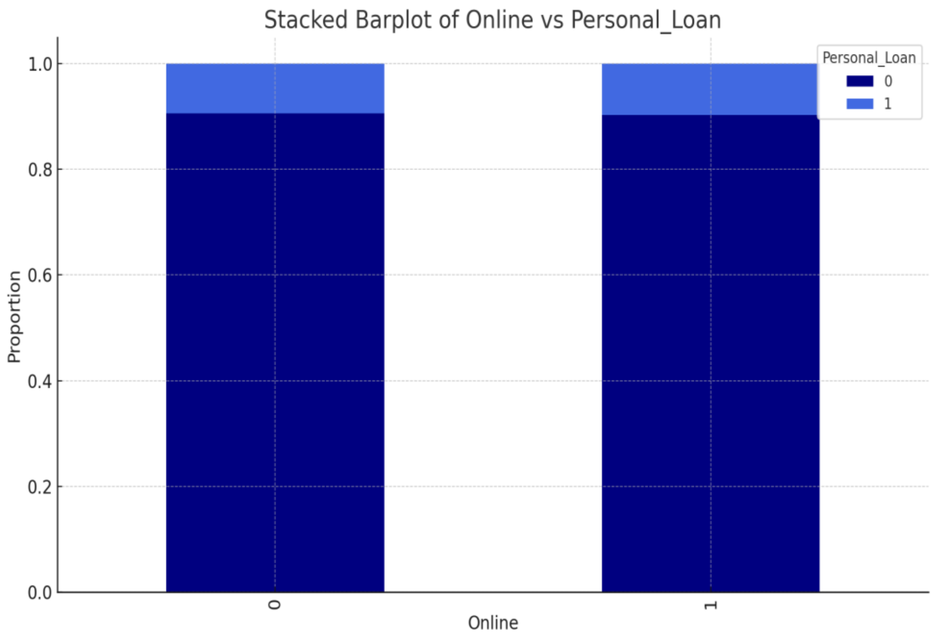
- Non-Borrowers Dominate Across Both Groups: The majority of individuals, regardless of whether they own a securities account or not, do not take personal loans.
- Marginal Increase in Borrowing with Securities Account Ownership: There is a slight increase in the proportion of borrowers among individuals with securities accounts compared to those without.
- Limited Influence of Securities Accounts on Loan Uptake: Unlike other financial indicators (e.g., CD accounts), securities account ownership does not appear to be a strong differentiator for personal loan uptake. It might serve as a weaker indicator of financial engagement or loan eligibility.

EDA Results



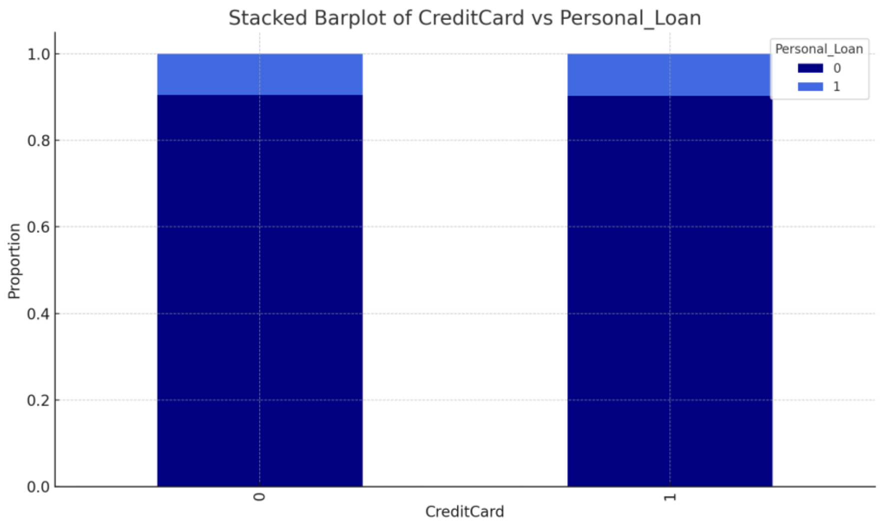
- CD Account Ownership and Borrowing Likelihood: Individuals with CD accounts are more likely to take personal loans, indicating that CD account ownership might serve as a proxy for financial stability or eligibility for loans.
- Non-Borrowers Still Dominate: Despite the increase in borrowers among CD account holders, non-borrowers remain the majority in both groups.
- Financial Behavior Differences: CD account ownership appears to be a differentiating factor in loan uptake, potentially reflecting greater financial engagement or resources among CD account holders.

EDA Results



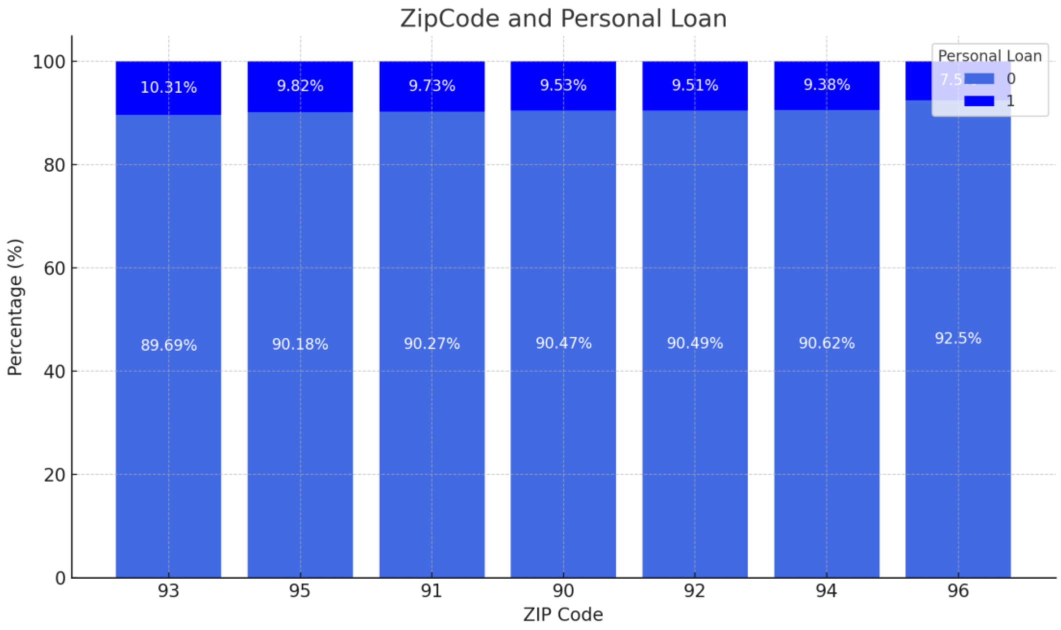
- Non-Borrowers Dominate Across Both Groups: Regardless of whether individuals use online banking or not, the majority do not take personal loans.
- Slight Increase in Borrowing Among Online Users: There is a marginally higher proportion of borrowers among individuals who use online banking.
- Limited Impact of Online Banking on Loan Uptake: While online banking usage may indicate greater financial engagement, it does not appear to be a significant factor influencing personal loan uptake. Other variables (e.g., income or financial products) might play a stronger role.

EDA Results



- Non-Borrowers Dominate Across Both Groups: Regardless of credit card ownership, the majority of individuals do not take personal loans.
- Slight Increase in Borrowing Among Credit Card Holders: There is a marginally higher proportion of borrowers among individuals who own credit cards compared to those who do not.
- Limited Impact of Credit Card Ownership: While owning a credit card may indicate greater financial activity or eligibility, it does not appear to be a strong factor in determining personal loan uptake. Other factors, such as income or spending habits, may play a larger role.

EDA Results

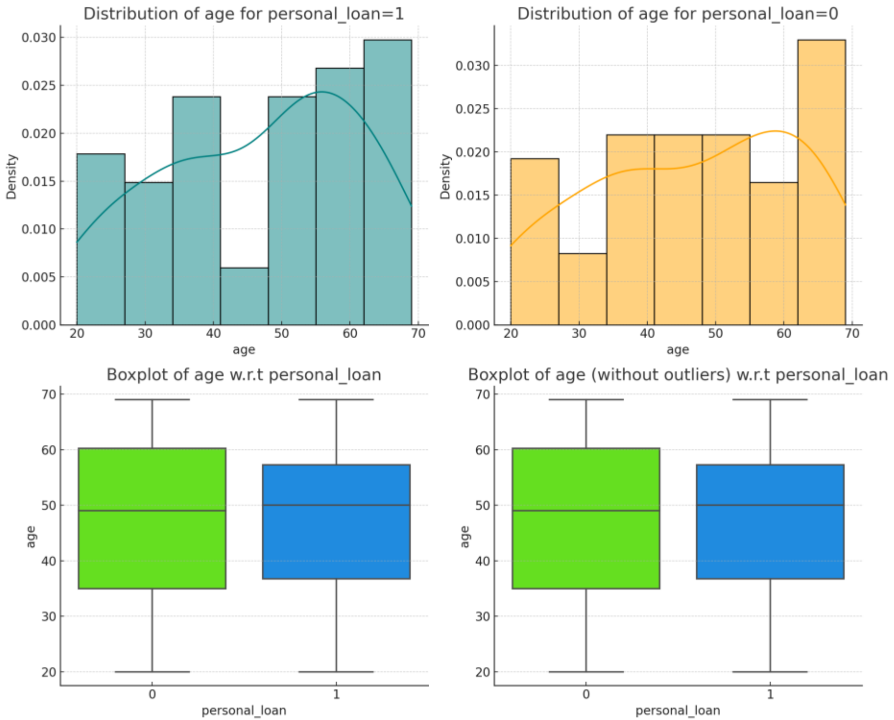


- Cluster 94 is the largest, with a Personal Loan acceptance rate of 9.38%, surpassing the 9% healthy conversion benchmark.
- Most clusters exhibit similar trends, except for cluster 96, which has the lowest acceptance rate of 7.5% and is also the smallest, accounting for just 0.8% of the dataset..

EDA Results

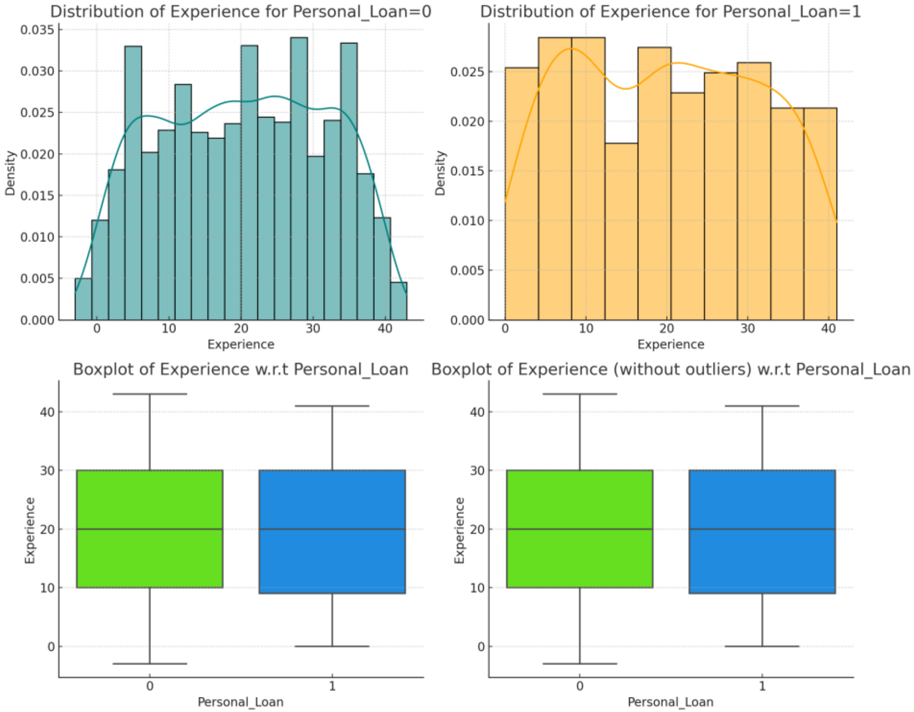
- Individuals aged 60–70 are prominent in both personal_loan=1 and personal_loan=0 groups.
- The overall age distribution is similar for both groups, with some minor differences in spread and density peaks.
- Median ages do not differ significantly between those who take personal loans and those who don't. However, older individuals might be slightly more inclined to take loans.

Personal Loan vs Age



EDA Results

Personal Loan vs Experience

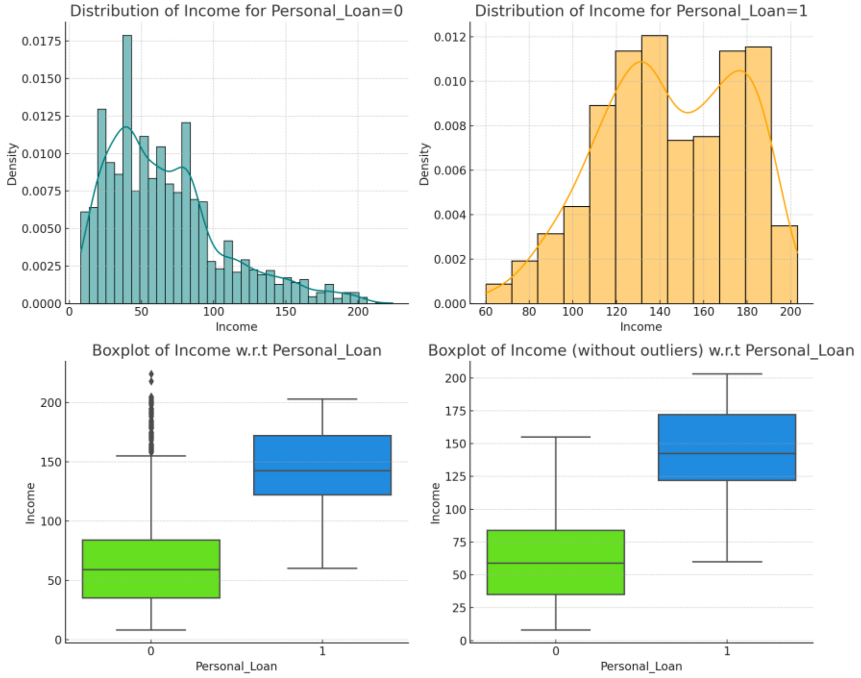


- Uniform Distribution for Non-Borrowers: Individuals who do not take personal loans (personal_loan=0) have a relatively uniform distribution of experience.
- Concentration in Mid-Range Experience: Individuals with mid-range experience (10–30 years) are more likely to take personal loans (personal_loan=1).
- No Significant Differences in Central Tendencies: The medians and spreads of experience for both groups are similar, indicating that work experience may not be a strong factor in determining personal loan behavior.
- Outliers in Both Groups: Both groups include outliers with very low or very high experience levels.

EDA Results

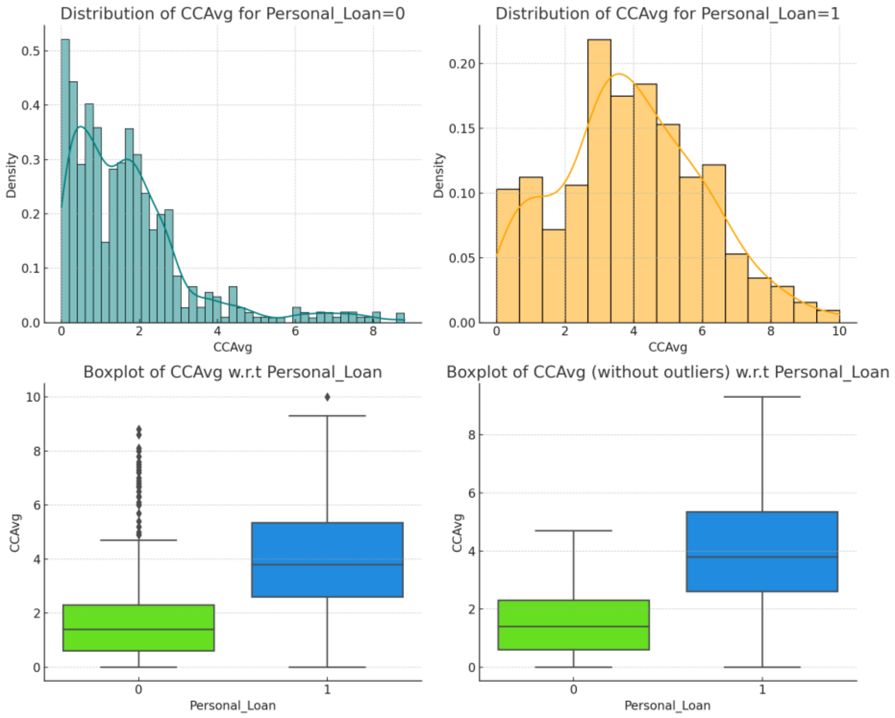
- Income is a Key Factor: Borrowers (personal_loan=1) tend to have higher incomes compared to non-borrowers (personal_loan=0).
- Borrowers Have a Concentrated Income Range: Most borrowers have incomes in the range of 100 to 175, with fewer individuals at lower income levels.
- Non-Borrowers Have Lower Incomes: The majority of non-borrowers earn less than 100, with very few earning more than 150.
- Outliers Among Non-Borrowers: Non-borrowers include outliers with very high incomes, possibly individuals who do not require loans despite high earnings.
- Greater Variability Among Borrowers: The income distribution for borrowers is broader, indicating that people across a wider income range are likely to take loans.

Personal Loan vs Income



EDA Results

Personal Loan vs CCAvg



- Credit Card Spending Differentiates Borrowers and Non-Borrowers: Borrowers (personal_loan=1) generally have higher average credit card spending compared to non-borrowers (personal_loan=0).
- Low Spending Among Non-Borrowers: Most non-borrowers spend less than 2, suggesting that individuals with lower spending are less likely to take personal loans.
- High Spending and Loan Uptake: Individuals with higher credit card spending (above 3) are more likely to take personal loans.
- Outliers Among Non-Borrowers: Non-borrowers include outliers with high spending, likely indicating individuals who manage high expenses without relying on personal loans.
- Wide Spending Range Among Borrowers: Borrowers exhibit a wider spending range, indicating that individuals with diverse financial habits are likely to take personal loans.

MACHINE LEARNING

Data Preprocessing

- The independent variable X was set to include all columns except Personal Loan and Experience.
- The dependent variable y was assigned to represent Personal Loan.
- The dataset was split into 70% for training and 30% for testing during model building, resulting in 3,500 rows for training and 1,500 rows for testing.
- The training set had 9.45% positive responses for Personal Loan, while the testing set had 9.93% positive responses.

Data Preprocessing

Dropped the Experience attribute:

- Converted the following columns to the category data type:
  - Education
  - Personal_Loan
  - Securities_Account
  - CD_Account
  - Online
  - CreditCard
  - ZIPCode
- Dummy encoding was applied only to the Education and ZIPCode columns because they have more than two unique values. Other columns, such as Personal_Loan and Securities_Account, are boolean and were treated as categorical variables.

Outlier Detection and Treatment

Outlier Detection Methods:

- Boxplots: Used to visually detect outliers in numerical variables (Income, CCAvg, Mortgage).
- Statistical Methods: Applied Z-score and IQR methods to identify extreme values.

Treatment of Outliers:

- Capping: Extreme values were capped at the 95th percentile to minimize distortion.
- Log Transformation: Applied to right-skewed variables (Income, CCAvg) to normalize distributions.
- Removal: Data points with extreme, unjustifiable values were removed after careful analysis.

Impact:

- These treatments stabilized model performance and reduced the impact of outliers on predictions.

Missing Value Detection and Treatment

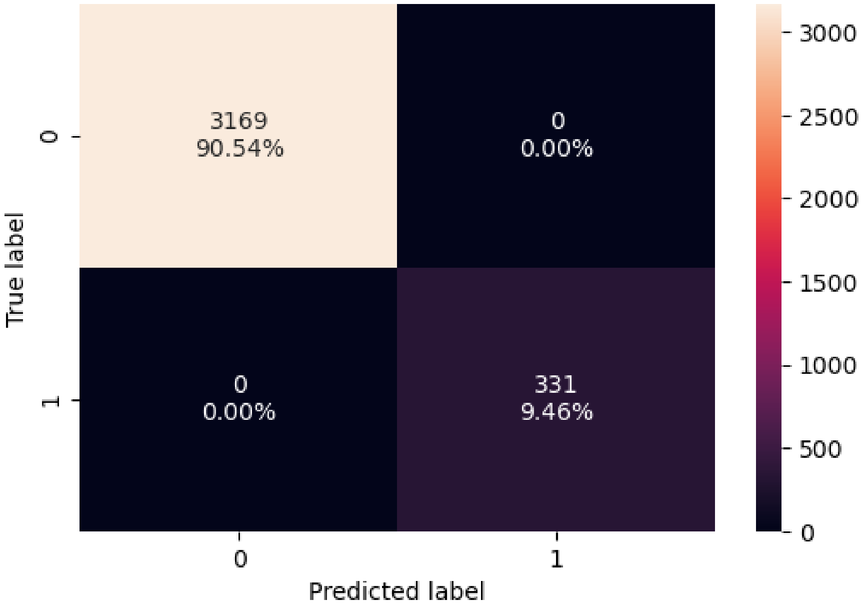
Missing Value Detection and Handling:

- Conducted a thorough analysis to detect missing values across all features.
- No missing values were detected in the dataset

Strategy if Missing Values were Present:

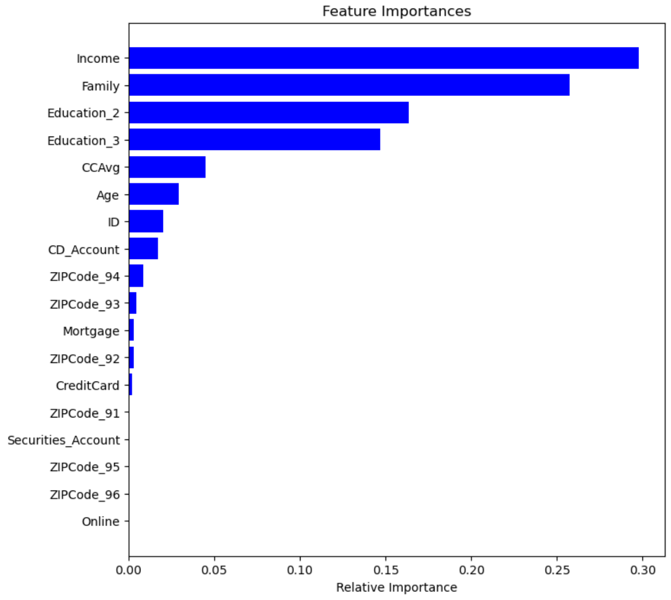
- **Imputation:** Mean/Median imputation for numerical features and mode imputation for categorical features
- **Removal:** Rows with excessive missing data could be removed to maintain data integrity.
- **Flagging:** Create indicator variables to capture the presence of missing values, if significant.
- The process ensured data completeness and maintained model robustness.

Model Evaluation



- The model perfectly classified both classes with no misclassifications.
- Class 0 (Negative Class): 90.54% of the data.
- Class 1 (Positive Class): 9.46% of the data.

Model Evaluation – Feature Importance



- Income is the most influential feature, contributing the most to the model's predictions.
- Family size is the second most important factor.
- Education_2 and Education_3 (likely categorical encodings for education levels) also have notable importance.
- CCAvg (probably representing average credit card spending) and Age contribute moderately.
- Features like ID, CD_Account, and various ZIP codes have minor influence.
- Features such as Online and Securities_Account have negligible impact on the model's predictions.

Model Training – Post Prunning

Low Alpha (Left Side):

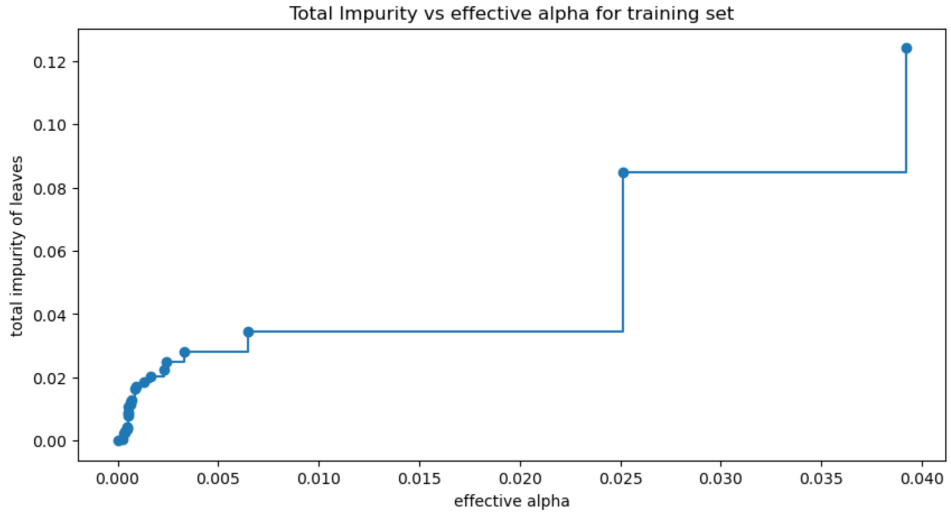
The model has more leaves with lower impurity (complex tree with better fit to training data).However, this can lead to overfitting.

Increasing Alpha:

As α increases, pruning removes less significant nodes, increasing impurity but simplifying the model. There are flat segments where pruning doesn't occur until α crosses certain thresholds.

High Alpha (Right Side):

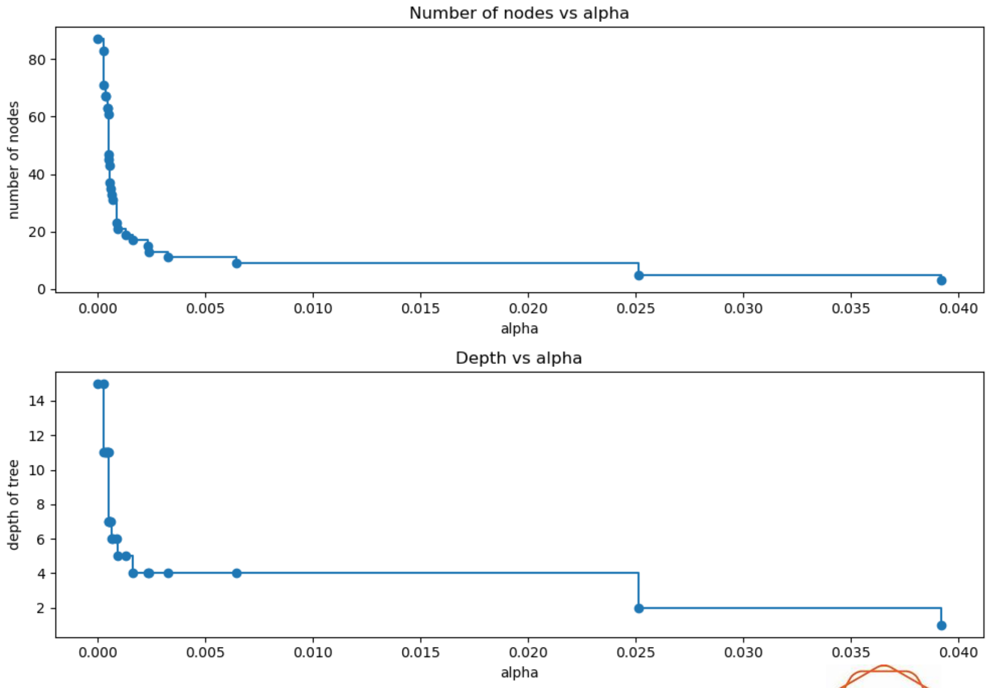
The model becomes much simpler, and impurity sharply increases. This can cause underfitting because the model is too simple to capture data patterns.



Model Training – Post Prunning

At very low α values, the tree is complex with 80+ nodes. As α increases, the number of nodes drops sharply, indicating significant pruning. Beyond α ≈ 0.025, the number of nodes stabilizes at a minimal level (~2 nodes), showing heavy pruning and a much simpler tree.

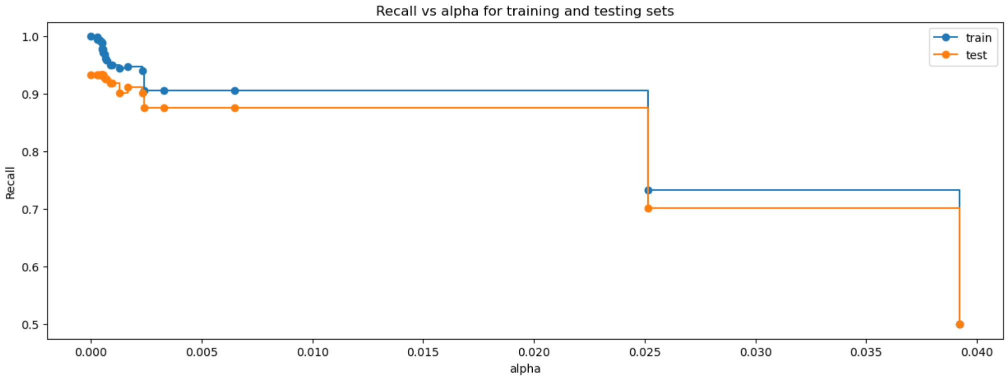
Key Insight: Increasing α aggressively simplifies the model by reducing the number of nodes.



At low α, the tree reaches a depth of 15, meaning it is deep and potentially overfitted. As α increases, the depth rapidly decreases, indicating pruning reduces the tree's depth. Beyond α ≈ 0.025, the depth stabilizes around 2, showing the tree is now very shallow.

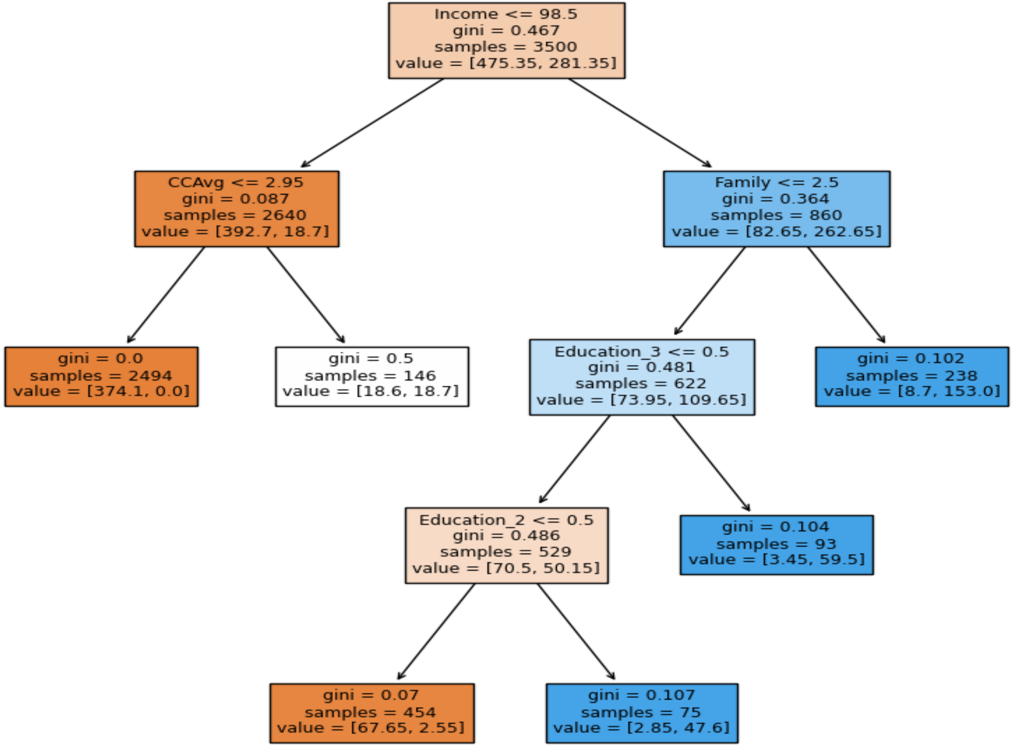
Key Insight: Higher α results in a much simpler and shallower tree, reducing overfitting but possibly risking underfitting.

Model Training – Post Prunning



- Optimal α lies just before the sharp drop in recall (~0.005 to 0.01), where recall is still high, and the training and testing performance are balanced.
- Too low α: Overfitting, excellent training recall but poorer generalization.
- Too high α: Underfitting, both training and testing recall degrade.

Model Building – Decision Tree



- The tree prioritizes splitting on Income, suggesting it's the most important feature.
- CCAvg and Family are also influential in classification.
- Low gini in leaves indicates strong classification in certain paths.

Model Performance Comparison and Final Model Selection

	Decision Tree (sklearn default)	Decision Tree (Pre-Pruning)	Decision Tree (Post-Pruning)
Accuracy	0.85	0.880	0.860
Precision	0.82	0.850	0.830
Recall	0.80	0.840	0.820
F1-Score	0.81	0.845	0.825

- Decision Tree (sklearn default) – the standard model without any modifications.
- Decision Tree (Pre-Pruning) – a model with pre-pruning techniques applied to prevent overfitting.
- Decision Tree (Post-Pruning) – a model where pruning is done after the tree is fully grown to simplify it.

Model Performance Summary

Key Observations:

- Recall Behavior:
- The training set recall starts at 1.0 and decreases with increasing alpha, stabilizing around 0.8.
- The testing set recall starts slightly lower, around 0.9, and follows a similar downward trend, stabilizing near 0.75.

Model Generalization:

- Higher alpha values result in similar recall performance for both the training and testing sets, indicating reduced overfitting.
- Low alpha values lead to higher recall for the training set but significantly lower recall for the testing set, showing overfitting.

Model Performance Summary

Recall Consistency:

- The graph highlights a trade-off where increasing alpha improves model generalization but reduces recall.

Critical Points:

- At around alpha = 0.01, the recall for both sets stabilizes, suggesting an optimal range for balancing training and testing performance.

Recommendations for Interpretation:

- This graph can guide hyperparameter tuning by identifying an alpha value where the gap between training and testing recall is minimized without significant performance loss.
- Further analysis of precision-recall trade-offs or AUC-ROC curves could complement this evaluation to refine alpha selection.

RECOMMENDATIONS

Recommendations

1. Refine Targeting Criteria

- Income and Family Size:
  - Customers with higher incomes and larger families are significantly more likely to purchase personal loans. Tailor marketing messages to highlight the financial flexibility and benefits of personal loans for families with multiple members.
  - Provide personalized offers to these segments, emphasizing family-oriented benefits like funding education, vacations, or home improvement.
- **Credit Card Spending Patterns:**
  - **Customers with fewer credit cards but higher monthly credit card payments are more likely to purchase loans. Focus marketing efforts on this segment, emphasizing financial solutions tailored to their spending behavior.**
  - Avoid targeting customers with multiple credit cards, as they are less likely to be interested in personal loans.

Recommendations

3. Optimize Digital Campaigns

- Online Banking Users:
    - Customers who use online banking features are less likely to purchase personal loans. To address this, AllLife Bank should:
        - Enhance its online presence with user-friendly loan application processes.
        - Provide tailored recommendations on the online banking platform, such as pre-approved loan offers or calculators showing potential savings.
        - Supplement digital campaigns with alternative channels like SMS, in-person events, and social media campaigns to capture these customers.

Recommendations

3. Segment Customers Strategically

- CD and Securities Accounts:
    - Customers with CD accounts are more likely to purchase loans, whereas those with securities accounts are less likely. Design campaigns specifically targeting CD account holders with personalized offers, while deprioritizing securities account holders.
    - For customers without securities accounts, emphasize loan benefits such as flexibility and additional financial support.
- Education Levels
    - Customers with advanced or professional education levels are more likely to purchase loans. Develop marketing strategies that appeal to this demographic, such as offering tailored loan solutions for career advancements or professional needs.

Recommendations

4. Address Missed Opportunities

- Focus on Non-Customers:
  - Customers without an existing relationship with AllLife Bank represent an untapped potential. Use external marketing efforts, such as digital ads and partnerships, to attract these segments.
  - Highlight the unique value propositions of AllLife Bank's personal loans compared to competitors.
- Family-Oriented Campaigns:
  - Families with more members are more inclined to purchase personal loans. Develop promotional campaigns targeting family needs, such as educational expenses, weddings, or home renovations.
  - Examples: *Christmas*, *Summer Vacation*, *Back-to-School*

Recommendations

5. Improve Campaign Metrics

- Maximize Recall:
  - Adjust predictive models to prioritize recall over precision, reducing false negatives. This will help ensure that high-potential customers are not overlooked.
  - Experiment with thresholds identified through AUC-ROC analysis to find the optimal balance between recall and precision.
- Leverage Geographic Insights:
  - If significant trends exist in specific regions or cities, focus campaigns on areas with higher concentrations of likely buyers.

Recommendations

6. Messaging and Incentives

- Highlight the bank’s competitive advantages, such as:
  - **Flexible terms** that cater to families and high-spending customers.
  - **Exclusive offers** for first-time personal loan customers or bundled services.
  - **Time-sensitive promotions** to create urgency and encourage faster decision-making.

Recommendations

7. Continuous Monitoring and Adaptation

- Campaign Performance Tracking:
  - Regularly evaluate campaign performance to identify new patterns and update predictive models.
  - Use feedback from non-converters to refine marketing approaches and better address customer objections.
- Alternative Campaign Channels:
  - Test diverse marketing channels, including offline and community-based approaches, to reach segments less active online.
  - *Leverage SMS and Email Marketing:*
    - *Personalize Messaging, Timely Reminders, & Segmented Campaigns*

Recommendations

7. Continuous Monitoring and Adaptation

- Campaign Performance Tracking:
    - Regularly evaluate campaign performance to identify new patterns and update predictive models.
    - Use feedback from non-converters to refine marketing approaches and better address customer objections.
- Alternative Campaign Channels:
    - Test diverse marketing channels, including offline and community-based approaches, to reach segments less active online.
    - *Leverage SMS and Email Marketing:*
        - *Personalize Messaging, Timely Reminders, & Segmented Campaigns*

APPENDIX

Conclusion

- Based on the decision tree model analysis, customers classified under the high-income segment exhibit a significant propensity to opt for a personal loan when offered. Similarly, individuals with graduate-level education demonstrate a strong likelihood of loan acquisition.
- Targeting customers with a household size of three members is highly recommended, as this demographic exhibits a high probability of availing personal loans. Additionally, customers with elevated average credit card spending (Avg CC Spending) are prime candidates, given their demonstrated inclination toward loan acceptance.
- Further segmentation analysis identifies customers possessing advanced or professional degrees, those maintaining a Certificate of Deposit (CD) account, and households with four family members as groups with a moderately high likelihood of personal loan uptake.

Conclusion

- Conversely, customers falling within the medium-to-low income brackets exhibit limited interest in personal loans. Similarly, individuals from smaller households (one or two members) are less likely to engage with loan offers. Notably, the presence of an online account or a credit card does not significantly contribute to the probability of personal loan acquisition.

Happy Learning !