Smart Lending: Using Analytics to Transform Customer Engagement and Drive Loan Growth

Project 2:Personal Loan Campaign
Course Name: AIML

Date:20/Apr/2024

Contents / Agenda

Executive Summary Key Recommendations

Introduction:

AllLife Bank, a leading U.S. financial institution, is poised to expand its borrower base among its large pool of liability customers. Building on a past campaign's 9% success rate in converting depositors to loan customers, the bank seeks to further leverage predictive analytics to enhance these efforts. By analyzing customer data through decision tree models and refining them with pruning techniques, we have pinpointed key attributes that significantly influence loan uptake. Our optimized model is designed to prioritize high recall, ensuring that potential borrowers are not overlooked, while maintaining cautious precision to avoid inefficient targeting. Implementing this model will enable targeted marketing campaigns, focusing on customers most likely to convert and offering them personalized financial products. This strategic approach promises to not only boost loan conversions but also deepen customer relationships and drive substantial business growth.

Executive Summary Key Recommendations

Enhancing Services Through Strategic Initiatives

Customer Segmentation and Targeting

- **Identify Key Characteristics:** Use the model to pinpoint attributes linked to loan uptake, like income and credit usage
- **Strategic Marketing:** Focus efforts on segments likely to convert, boosting efficiency and rates

Model Monitoring and Updating

- **Continuous Monitoring:** Regularly assess model performance to match expected outcomes
- **Regular Updates:** Refresh the model periodically to stay accurate and relevant with market changes

Risk Management

- **Balanced Approach:** Weigh precision against recall to match the bank’s risk profile
- **Adaptive Strategies:** Dynamically adjust strategies based on analytics and feedback

Product Personalizatio n

- **Customized Offers:** Create loan offers tailored to customer profiles and financial behaviors
- **Promotional Strategies:** Design special promotions targeting specific spending behaviors

Executive Summary Key Recommendations

Enhancing Services Through Strategic Initiatives

Transparency and Stakeholder Engagement

- **Model Interpretability**: Utilize decision tree clarity to explain model decisions, enhancing trust
- **Education and Buy-in**: Inform stakeholders about the model's benefits to secure support

Compliance and Ethical Use

- **Regulatory Compliance:** Ensure the model complies with legal and fair lending standards
- **Bias Auditing:** Regularly check and adjust the model to prevent biases

Feedback and Continuous Improvement

- **Feedback Loops:** Establish systems to incorporate strategy outcomes into ongoing model refinement
- **A/B Testing:** Use A/B testing for optimizing strategies in real-world scenarios

Cross-Selling and Customer Engagement

- **Cross-Selling Initiatives:** Leverage insights for cross-selling, enhancing product uptake
- **Proactive Engagement:** Engage indecisive customers with personalized benefits communication

Executive Summary Conclusion

Implementing this predictive model will not only streamline AllLife Bank's targeting efforts but also enhance the effectiveness of its loan conversion strategies. By continuously refining our approach based on robust analytics, we ensure that our financial solutions are both dynamic and customer-centric. This strategy will not only increase loan uptake but also strengthen customer relationships, positioning AllLife Bank as a leader in personalized banking services

Business Problem Overview

Current Challenge

- **Underutilized Borrower Base:** Despite a large depositor base, only a small percentage of liability customers also take out personal loans
- **Limited Conversion Success:** Past marketing campaigns have achieved a 9% conversion rate, indicating potential for significant improvement

Implications

- **Revenue Limitation:** Current low conversion rates are limiting potential revenue from interest on personal loans
- **Customer Engagement:** Inefficient targeting and personalization in loan offers may lead to missed opportunities in deepening customer relationships

Solution Approach for Business Problem

Solution Approach

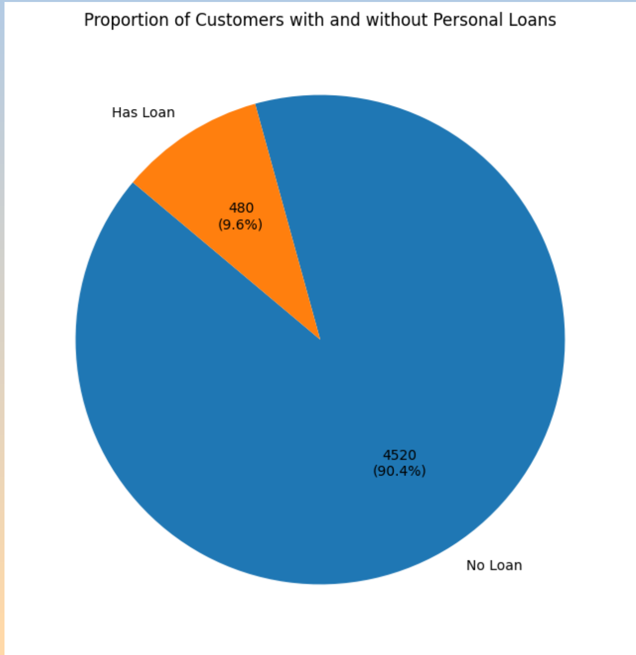
Predictive Analytics Deployment
- Model Development
- Model Optimization

Targeted Marketing Strategy
- Data-Driven Segmentation
- Personalized Campaigns

Engagement and Customization
- Tailored Offers
- Proactive Communication

Continuous Model Refinement
- Regular Updates
- Performance Monitoring

Comprehensive Data Overview



Dataset Composition

- **Scope of Analysis:** Dataset comprises 5,000 customer records with key attributes like age, income, and financial details
- **Data Diversity:** Includes both numerical data (e.g., income, CCAvg, mortgage) and categorical data (e.g., education level, securities account)

Data Completeness

- **Integrity Check:** No missing values, no duplications, ensuring dataset reliability
- **Data Cleaning:** Addressed issues like negative experience years and outliers in income and mortgage values

Data Preparation

Data Type Conversion: Converted the data type of categorical features to 'category'

Feature Engineering: Simplified ZIP Codes to first two digits, converting them to categorical data for analysis

Data Processing Pre-Visualization

Missing Values Checks	• There were no null values in the data set
Anomaly Checks	• 52 rows of negative "Experience" anomalies detected • -1,-2 and -3 entries converted to positive values
Duplication Checks	• There were no duplicate values in the data set
Data Type Conversions	• Converted 7 of the data type of categorical features to 'category'

Data Processing Post-Visualization

Feature Engineering

- Simplified ZIP codes to first two digits, converting to categorical data to decrease granularity and increase analytical value

Dropping Columns

- **ID**: Removed post-duplication check as it offered no analytical value
- **Experience**: Excluded from modeling due to high correlation with Age, reducing redundancy

Outlier Detection

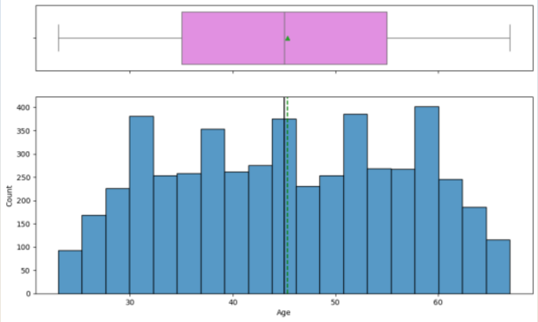
- Conducted meticulous examination for outliers, particularly in features influencing loan acceptance such as Income and CCAvg
- **Key Findings**:
  - **Income**: Identified high-income outliers in non-loan customers; fewer outliers in loan customers suggesting lesser variation in high-income levels
  - **CCAvg**: Significant outliers detected in loan-accepting customers indicating a subset with substantially higher credit card expenditure

Data Preparation For Modeling

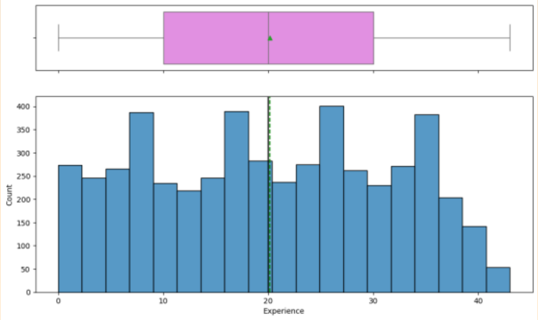
- **Data Split Details**:
  - Training set: 3,500 entries (70% of the dataset) with 478 features
  - Testing set: 1,500 entries (30% of the dataset) with 478 features
- **Class Distribution**:
  - Training set composition: 90.54% did not take personal loans (0), 9.46% took personal loans (1)
  - Testing set composition: 90.07% did not take personal loans (0), 9.93% took personal loans (1)

Univariate Analysis of Age and Experience

Age



Experience



Distribution Shape:

- Age and Experience are nearly uniform, showing diverse customer attributes
- Experience is slightly right- skewed but both are centered around their medians

Variability:

- Broad interquartile ranges for Age and Experience indicate diverse demographics and professional backgrounds among customers

Central Tendency:

- Median Age: Positioned in the mid-40s, indicating a mature customer base
- Median Experience: Approximately 20 years, reflecting considerable professional experience

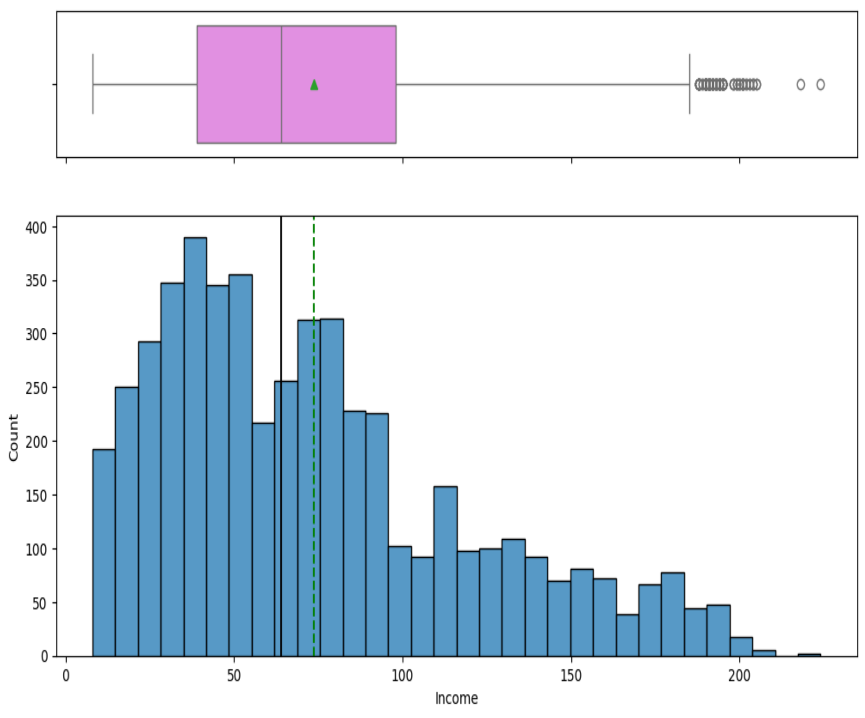
Implications for the Bank:

- Diverse Age and Experience profiles support targeted marketing and product offerings based on life stage and professional background

Univariate Analysis of Income

Income Observations:

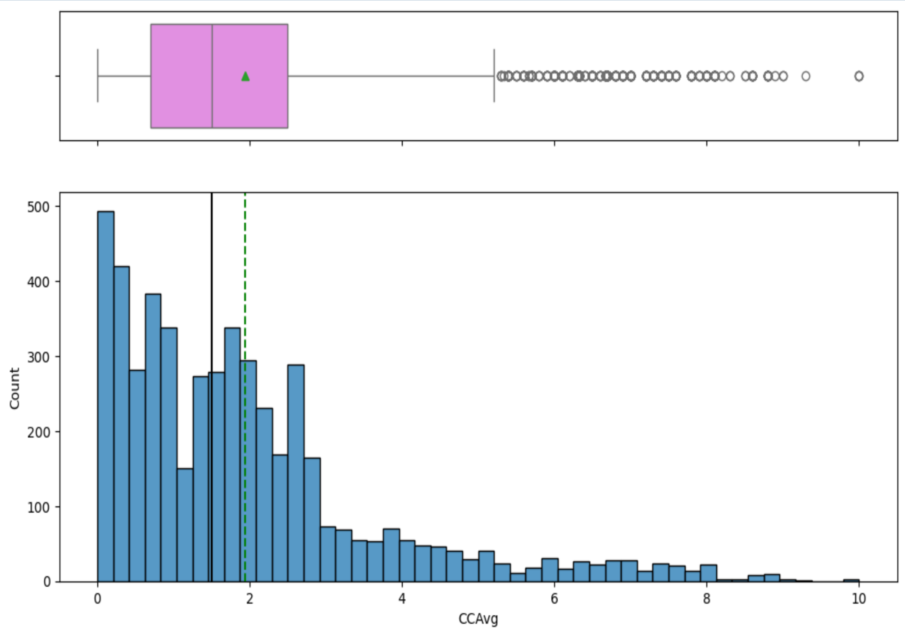
1. **Right-Skewed Distribution:** Most customers earn lower incomes with fewer having high incomes
2. **Central Tendency:** The median income is lower than the mean, characteristic of a right-skewed distribution
3. **Spread and Variability:** Wide interquartile range indicates significant income variability among customers
4. **Outliers:** Numerous outliers on the higher end highlight customers with exceptionally high incomes
5. **High-Income Segment:** The upper quartile has relatively high incomes, relevant for targeting with premium products
6. **Target Market Insights:** Skew towards lower incomes suggests targeting these groups with specific loan products
7. **Economic Diversity:** Highlights the need for diverse banking products to cater to varied economic backgrounds



Univariate Analysis of CCAvg

CCAvg Observations:

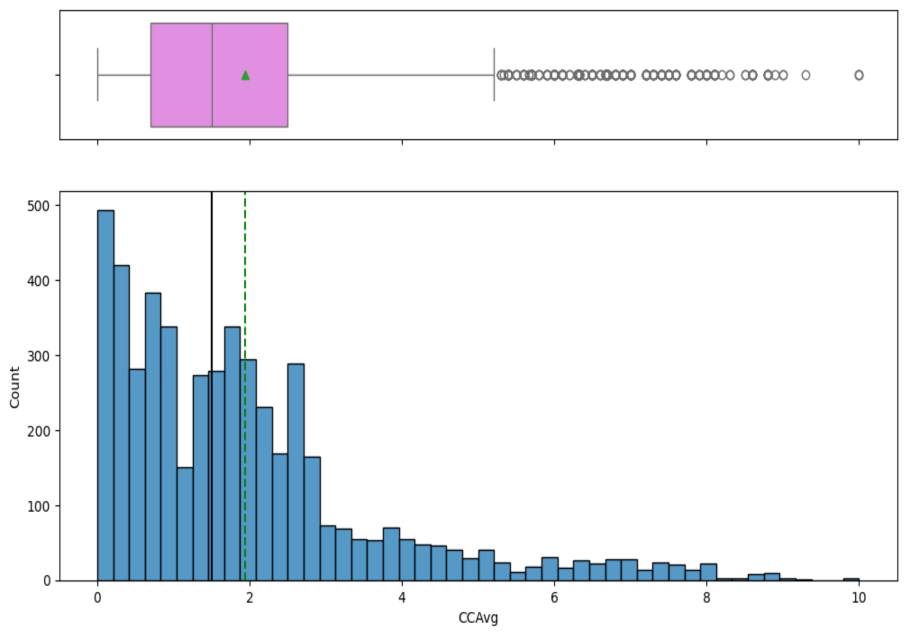
1. **Right-Skewed Distribution**: Majority have lower average credit card spending, with a tail extending to higher spend levels
2. **Central Tendency**: Median spending is below the mean, indicating most customers are low to moderate spenders
3. **Outliers:** Significant outliers in the higher spending range point to a small segment of heavy credit card users
4. **Spread and Variability:** A tight interquartile range but a long right tail suggests a split in spending habits
5. **Typical Spending Range:** Most customers spend between $0 and $2.5K monthly, with a peak around $0.7K
6. **High Spender Focus:** Outliers indicate potential for targeted marketing towards high-value credit card offerings



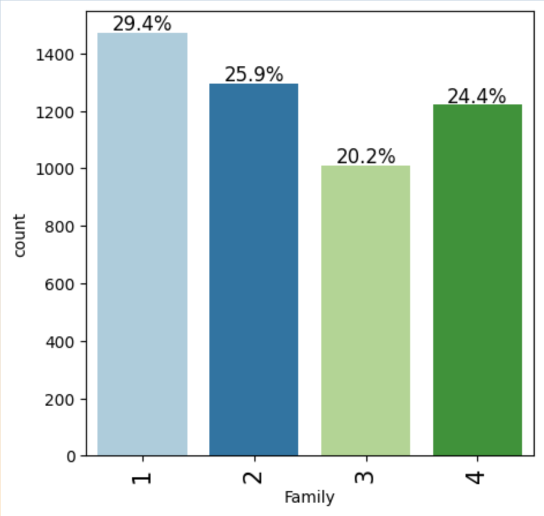
Univariate Analysis of Mortgage

Mortgage Observations:

1. **Right-Skewed Distribution:** A large portion of the dataset, 1538 customers, have zero or very low mortgage values, dominating the lower end
2. **Outliers:** Numerous high-end outliers indicate that while rare, some customers hold significantly high mortgage values
3. **Spread and Variability:** Small interquartile range among non-zero values, suggesting limited variation in mortgage amounts among most customers who hold them
4. **Modeling Implications:** Given many zero values, modeling could improve by treating mortgage presence as binary and amount conditionally
5. **Marketing Potential:** The predominance of customers without mortgages (3462) presents a marketing opportunity for mortgage products

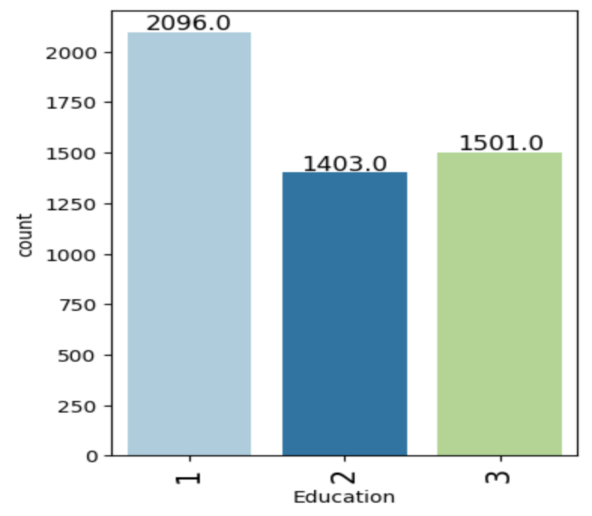


Univariate Analysis Family and Education



Family Size:

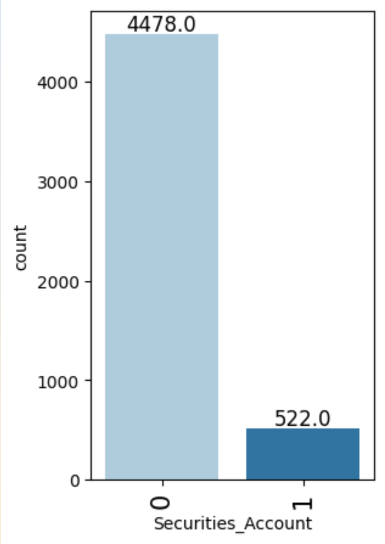
1. Most customers are in family sizes of 1 and 2, with these two categories comprising over half of the customers
2. Smaller family sizes are more prevalent, while larger family sizes (4 members) make up a smaller proportion of the customer base



Education Level:

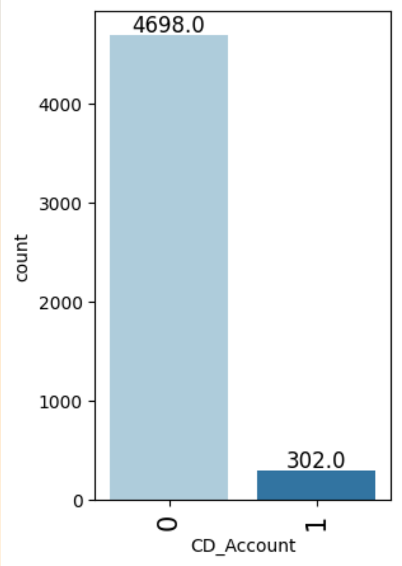
1. A significant number of customers have an undergraduate level of education.
2. Graduate and professional levels also have a considerable representation, indicating a customer base with diverse educational backgrounds

Univariate Analysis Family and Education



<u>Securities Account:</u>

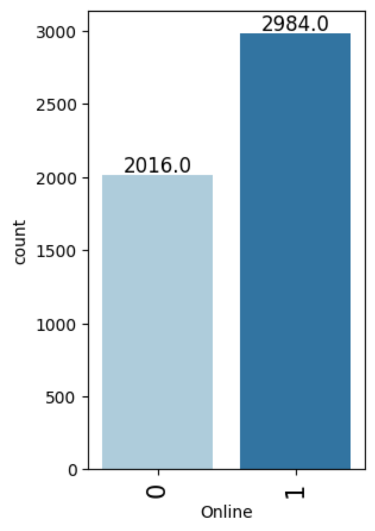
1. A vast majority of customers do not have a securities account with the bank
2. The low prevalence of securities accounts suggests potential for growth in this product area



<u>CD Account:</u>

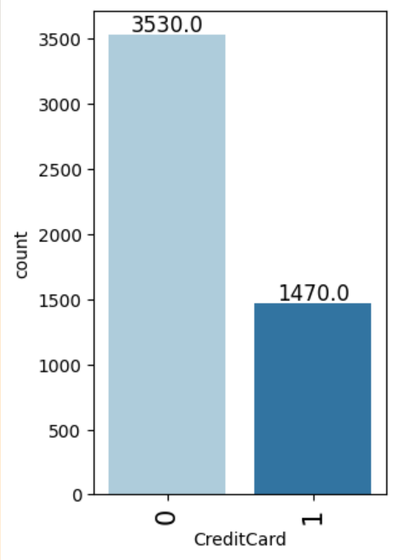
1. Similar to securities accounts, certificate of deposit (CD) accounts are not widely held among the customers, indicating an opportunity for the bank to expand this service

Univariate Analysis of Online Banking and Credit Card



Online Banking:

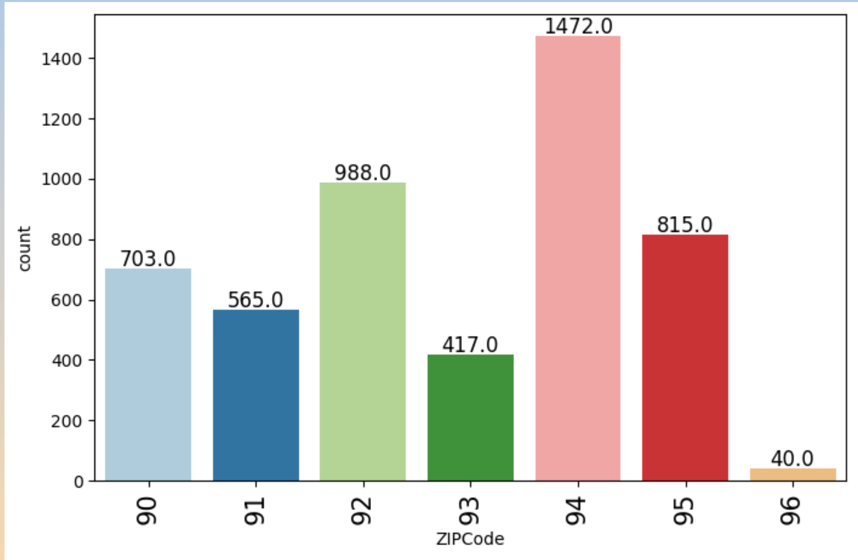
1. There is a fairly even split between customers who use and do not use online banking facilities
2. The significant usage of online banking indicates customer openness to digital banking solutions



Credit Card:

1. A substantial number of customers do not use a credit card issued by another bank, which may imply loyalty or a gap in credit card penetration that the bank could address

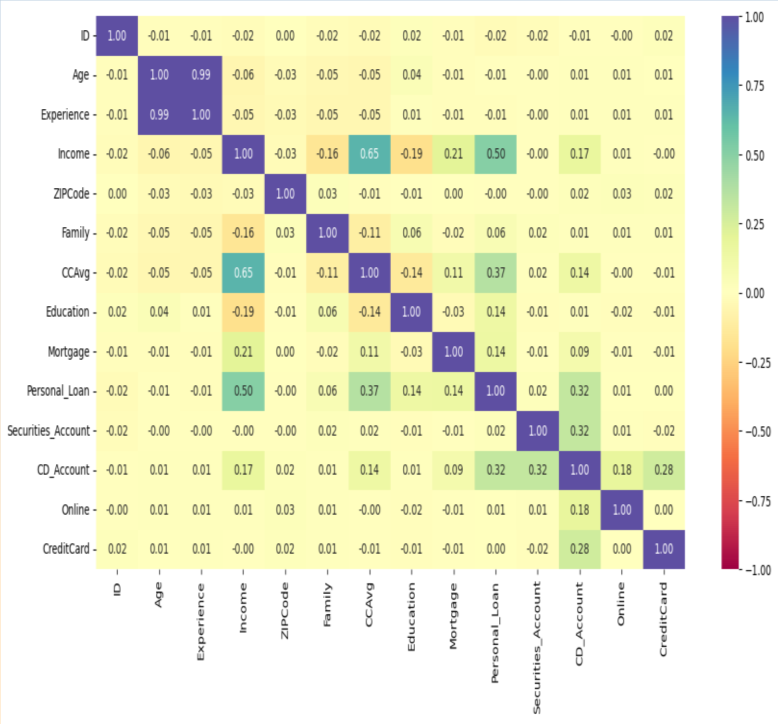
Univariate Analysis of Zip Code



<u>ZIP Code:</u>

1. The distribution of customers across ZIP codes shows a varied geographical spread, with some ZIP codes having a higher concentration of customers
2. This geographical diversity provides opportunities for localized marketing strategies and service offerings

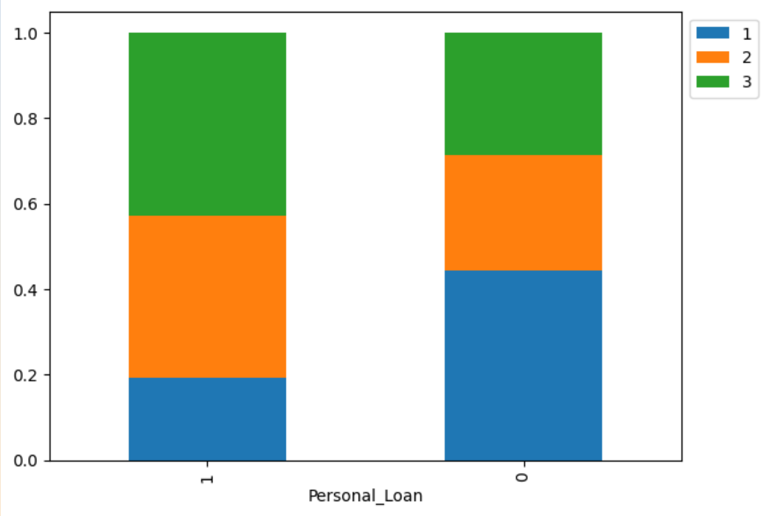
Multivariate Analysis from Heat Map



Key Observations:

1. **Customer Demographics:** Strong positive correlation between age and experience points to targeted opportunities based on life stages; while family size shows a potential link to mortgage uptake.
2. **Income Dynamics:** Income is a central factor, positively influencing personal loan acceptance and credit card expenditure, highlighting higher-income customers as prime targets for financial products
3. **Financial Savvy and Product Usage:** CD account holders' correlations with personal loans and mortgages suggest a segment that's more engaged with a range of financial products, ideal for cross-selling initiatives
4. **Digital Banking Trends:** A positive link between online banking use and credit card possession emphasizes the role of digital platforms in driving service adoption
5. **Geographical Insights:** Weak ZIP code correlations indicate a minimal impact of location on customer banking behaviors, allowing for a broader marketing approach

Bivariate Analysis Education Vs Personal Loan

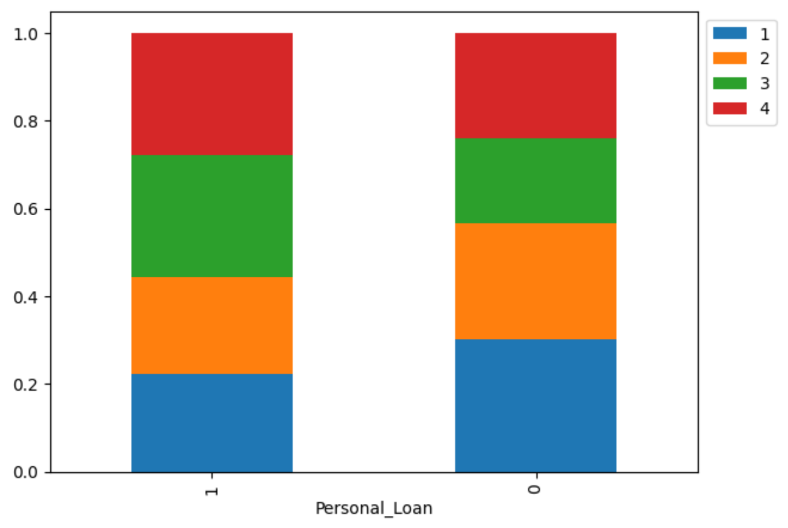


Key Observations of Education vs. Personal Loan:

1. Higher education levels correlate with higher personal loan uptake
2. Graduates and professionals show a greater propensity for loans compared to undergraduates
3. This pattern could guide targeted loan marketing strategies

Education	1	2	3	All
All	2096	1403	1501	5000
Loan (No)	2003	1221	1296	4520
Loan (Yes)	93	182	205	480

Bivariate Analysis Education Vs Personal Loan

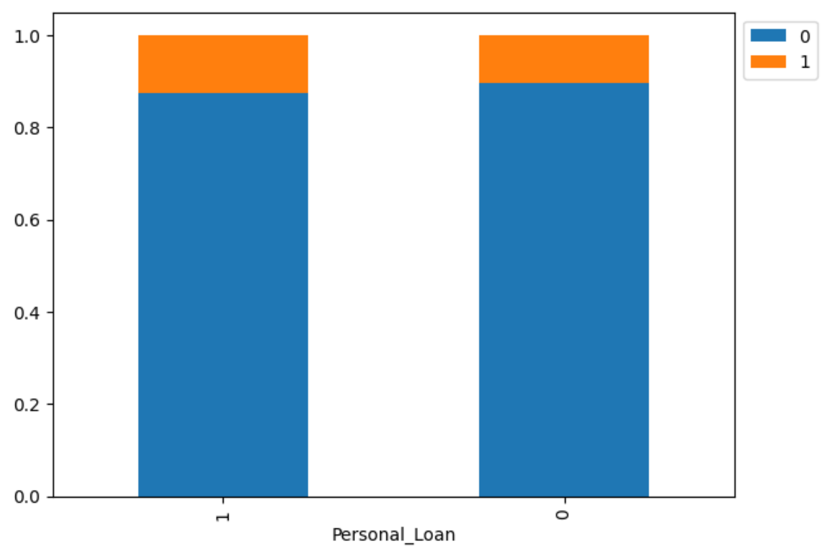


Key Observations of Family Size vs. Personal Loan :

1. Loan acceptance is distributed across all family sizes
2. No single family size significantly dominates loan uptake
3. Family size may not be a primary indicator for personal loan interest

Family	1	2	3	4	All
All	1472	1296	1010	1222	5000
Loan (No)	1365	1190	877	1088	4520
Loan (Yes)	107	106	133	134	480

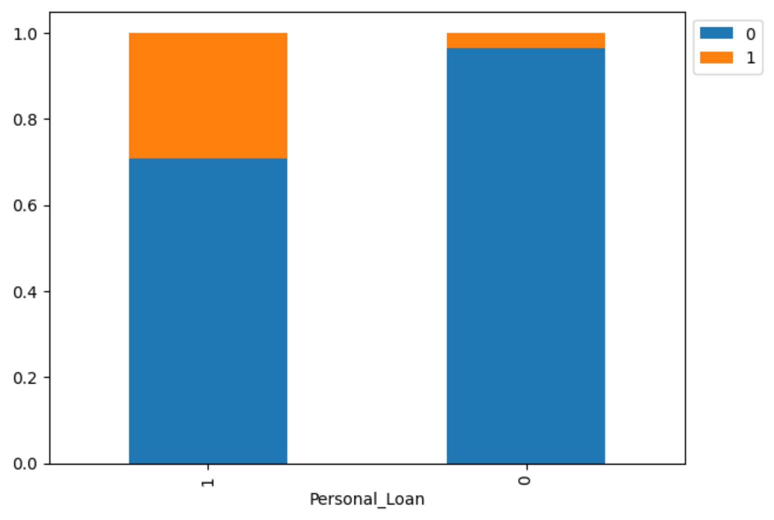
Bivariate Analysis Securities Account Vs Personal Loan



Key Observations of Security Accounts vs. Personal Loan :

1. Securities account ownership is present among both loan and non-loan customers
2. The presence of a securities account does not strongly predict personal loan acceptance
3. The majority do not have securities accounts, indicating other factors influence loan decisions

Bivariate Analysis CD Accounts Vs Personal Loan



Key Observations of CD Accounts vs. Personal Loan :

1. A substantial segment with CD accounts is observed among loan customers
2. The presence of a CD account may suggest a higher likelihood of loan acceptance
3. Most customers do not hold a CD account (4520), pointing to the potential for targeting this product

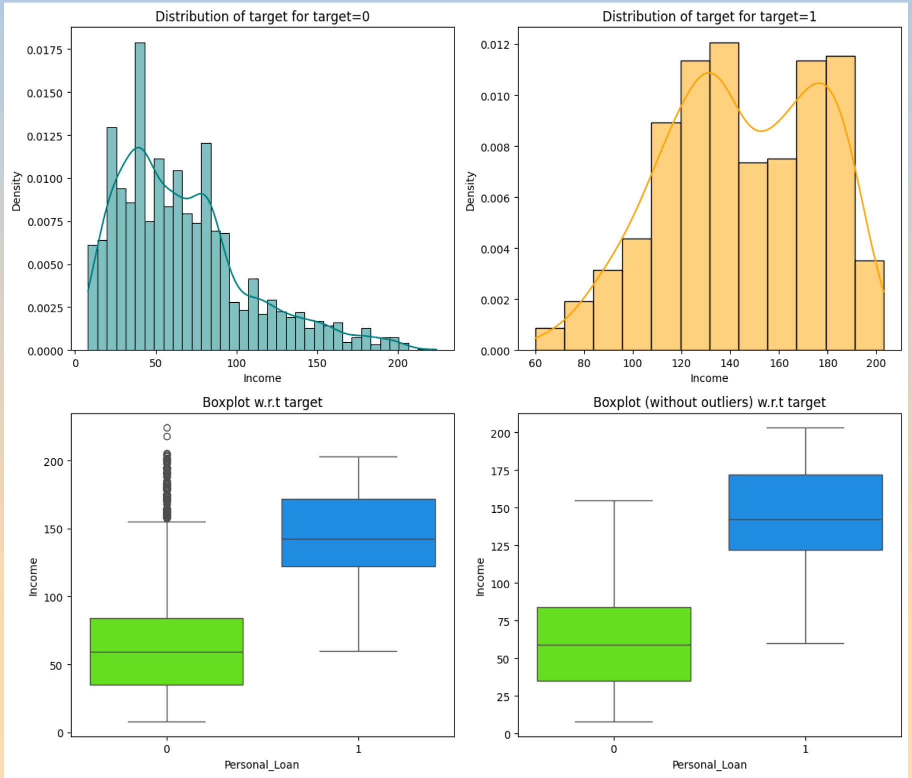
Bivariate analysis indicated that the features 'Online,' 'Zip Code,' 'Age,' and 'Experience' demonstrated minimal relationship with loan acceptance and do not strongly predict personal loan acceptance

Outlier Detection

Features	Outliers present
Income	1.92
CCAvg	6.48
Mortgage	5.82
Personal_Loan	9.60
Securities_Account	10.44
CD_Account	6.04

These outliers can all be indicative of distinct customer segments that the bank could analyze for targeted product offerings or personalized services

Outlier Treatment Income Vs Personal Loan



Income Distribution:

- Both non-loan (target=0) and loan customers (target=1) exhibit skewed income distributions
- Loan customers typically exhibit higher average incomes, highlighting the strong correlation between higher income and loan acceptance

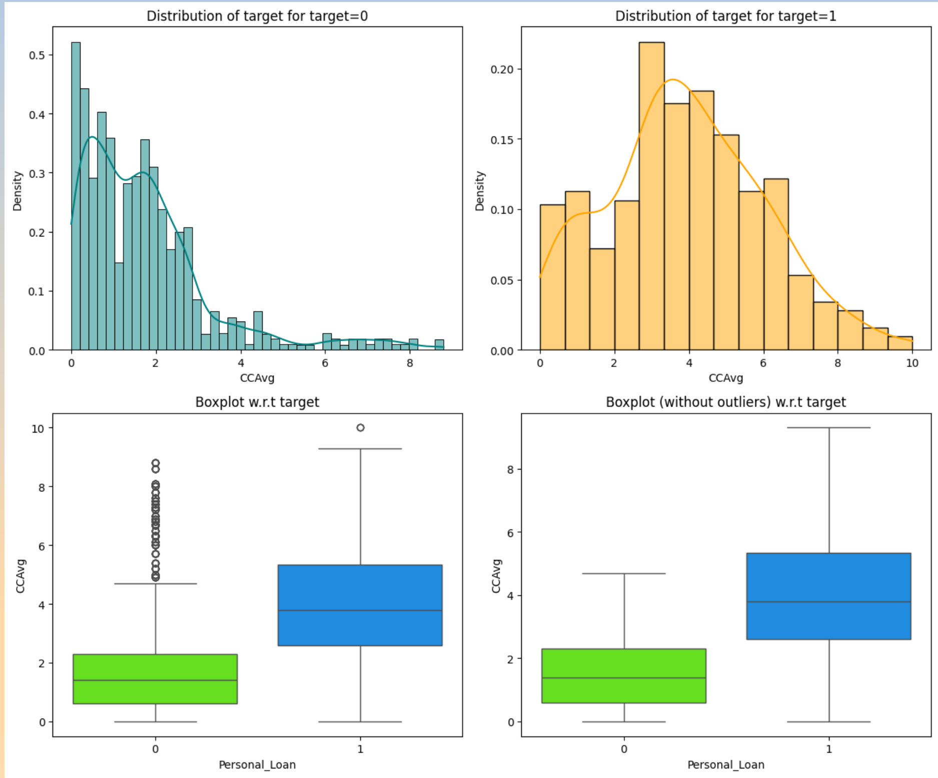
Outlier Analysis:

- **Non-Loan Customers:** Numerous high-income outliers, indicating not all high earners opt for loans
- **Loan Customers:** Fewer prominent outliers, possibly due to less variation in high incomes among this group

Implications for Bank Strategy:

- **Targeting High-Income Individuals:** Data suggests focusing marketing efforts on higher income brackets could enhance loan conversion rates
- **Predictive Model Consideration:** Insights from income distribution, including the outliers, could be pivotal in refining predictive models for loan uptake

Outlier Treatment CCVag Vs Personal Loan

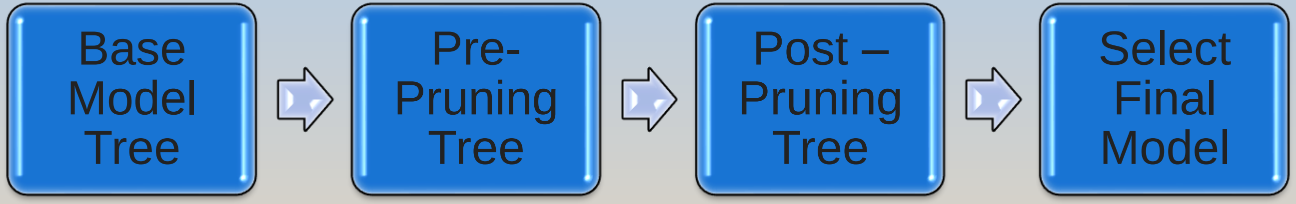


Credit Card Spending Distribution:
- Non-loan customers typically spend less on credit cards.
- Loan customers have higher credit card spending averages

Outlier Analysis:
- Credit card spending outliers among loan customers highlight significant spending variations
- Outliers are kept in the model to capture relevant spending behaviors that may predict loan uptake

Implications for Bank Strategy:
- Dominant spending group evident in both loan accepters and non-accepters
- Increased credit card usage may signal higher loan uptake likelihood
- Correlation between spending and loans to inform future targeted strategies

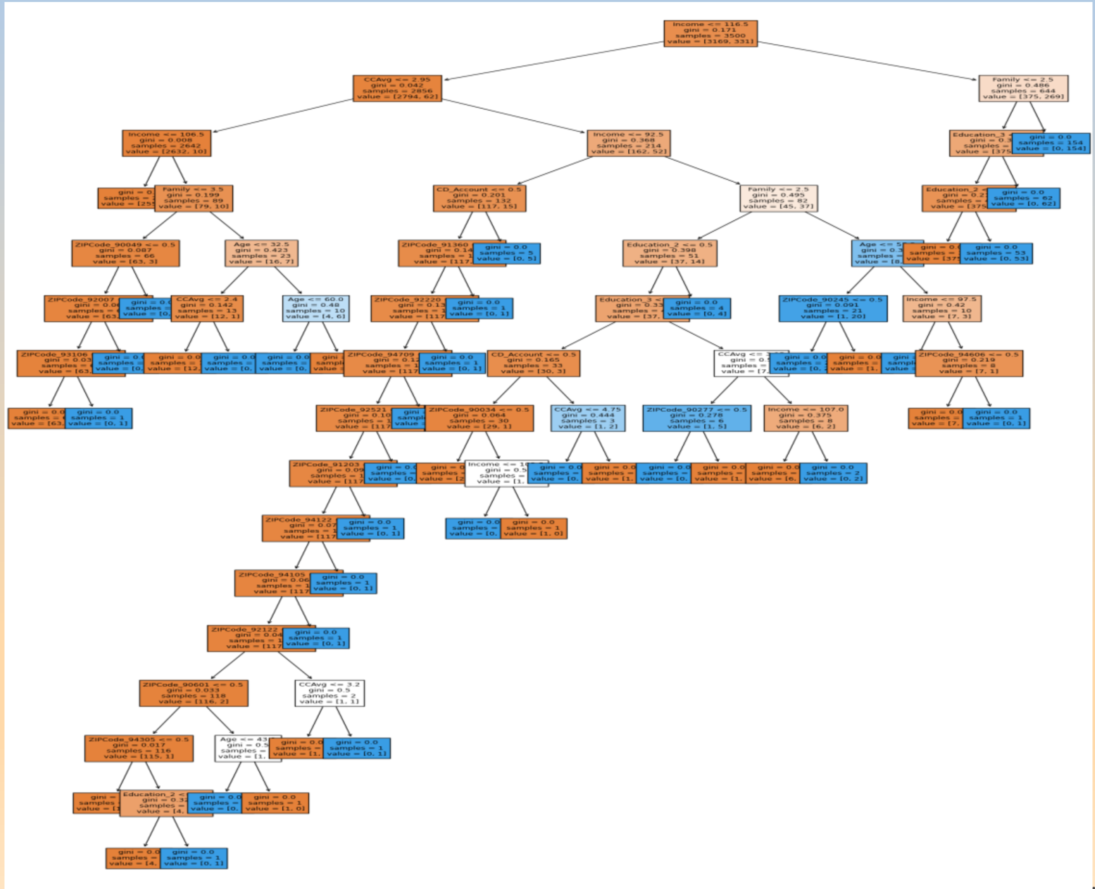
Model Building using Decision Tree



Why Choose Decision Tree?

1. Simplified Decision-Making: Decision trees reflect intuitive decision-making, ideal for visualizing the path from customer attributes to loan outcomes
2. Strategic Insight: Pinpoints key attributes affecting loan uptake, aligning with the goal of improving personal loan conversions through targeted marketing strategies

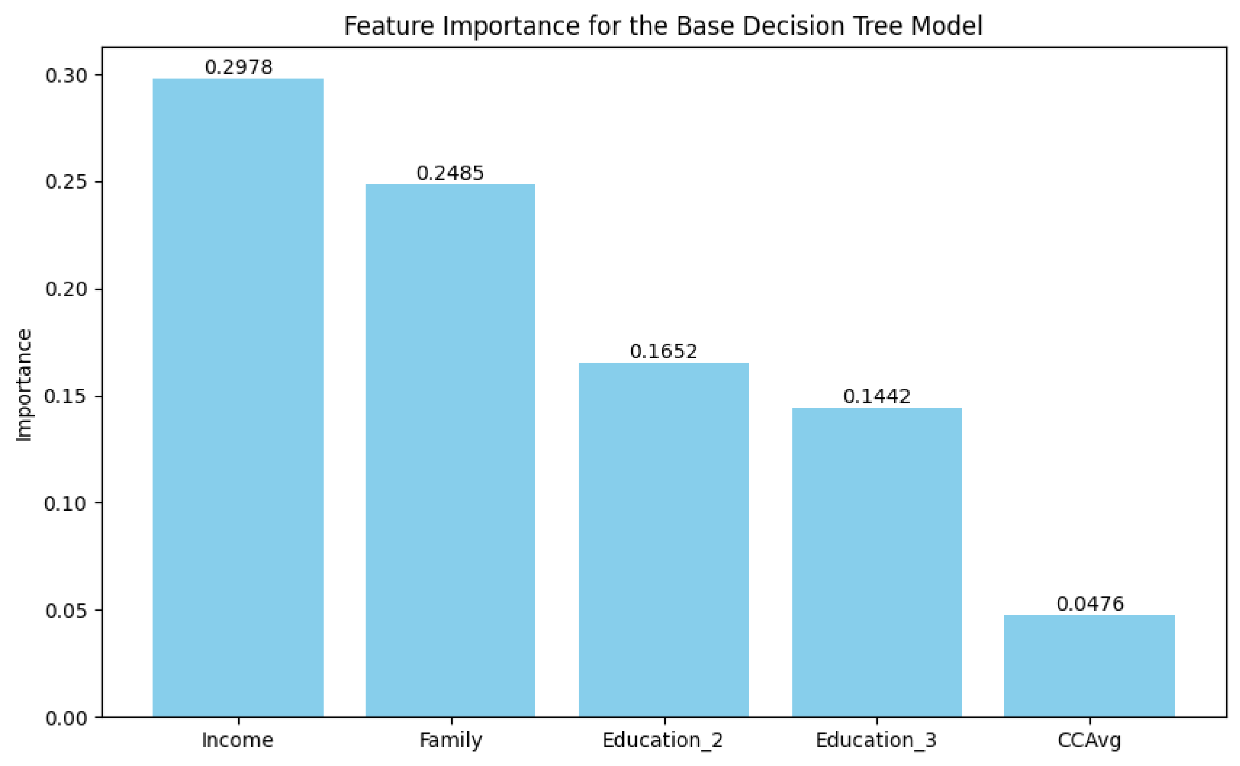
Base Model Building



Base Model Tree Visualization Overview:

The initial decision tree, rooted in 'Income' as the primary node, was constructed applying the Gini impurity criterion and stabilized with a random state set to 1 to ensure reproducibility of results

Base Decision Tree Feature Importance

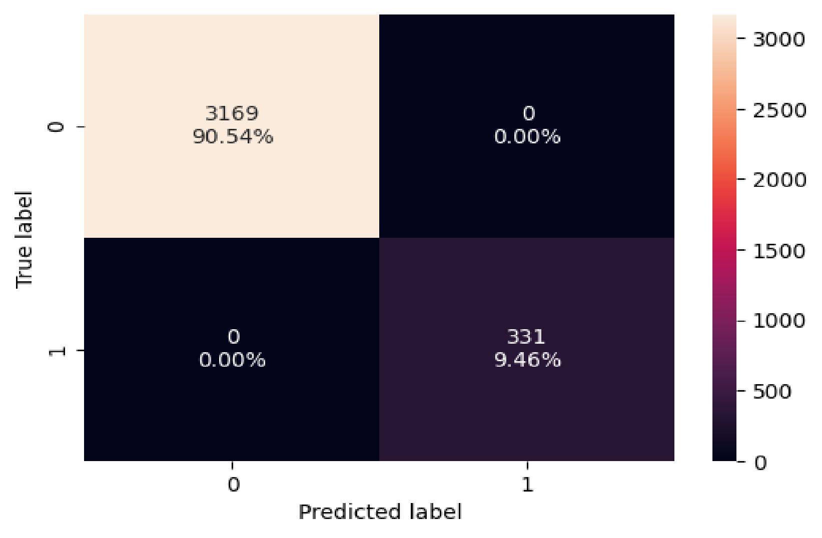


Key Predictors

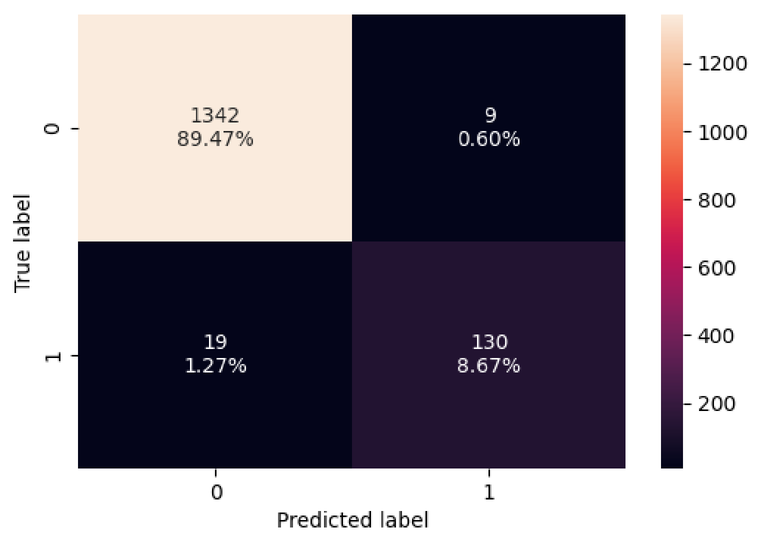
- **Income:** Dominant predictor with a 0.2978 importance score
- **Family Size**: Significant at a 0.2485 importance score
- **Higher Education (Graduate and Professional):** Notable at 0.1652 and 0.1442 importance scores
- **Credit Card Spending (CCAvg):** Minor yet relevant with a 0.0476 score

Base Model Performance Summary

Training Data Performance

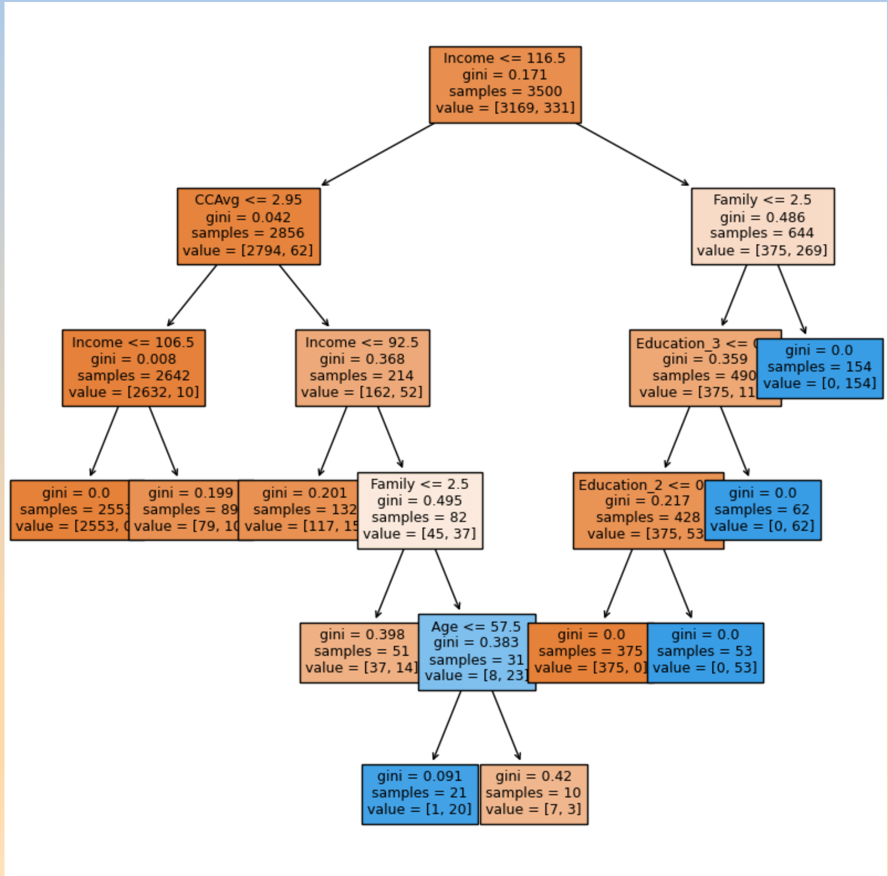


Test Data Performance



	Accuracy	Recall	Precision	F1
Train	1.0	1.0	1.0	1.0
Test	0.981333	0.872483	0.935252	0.902778

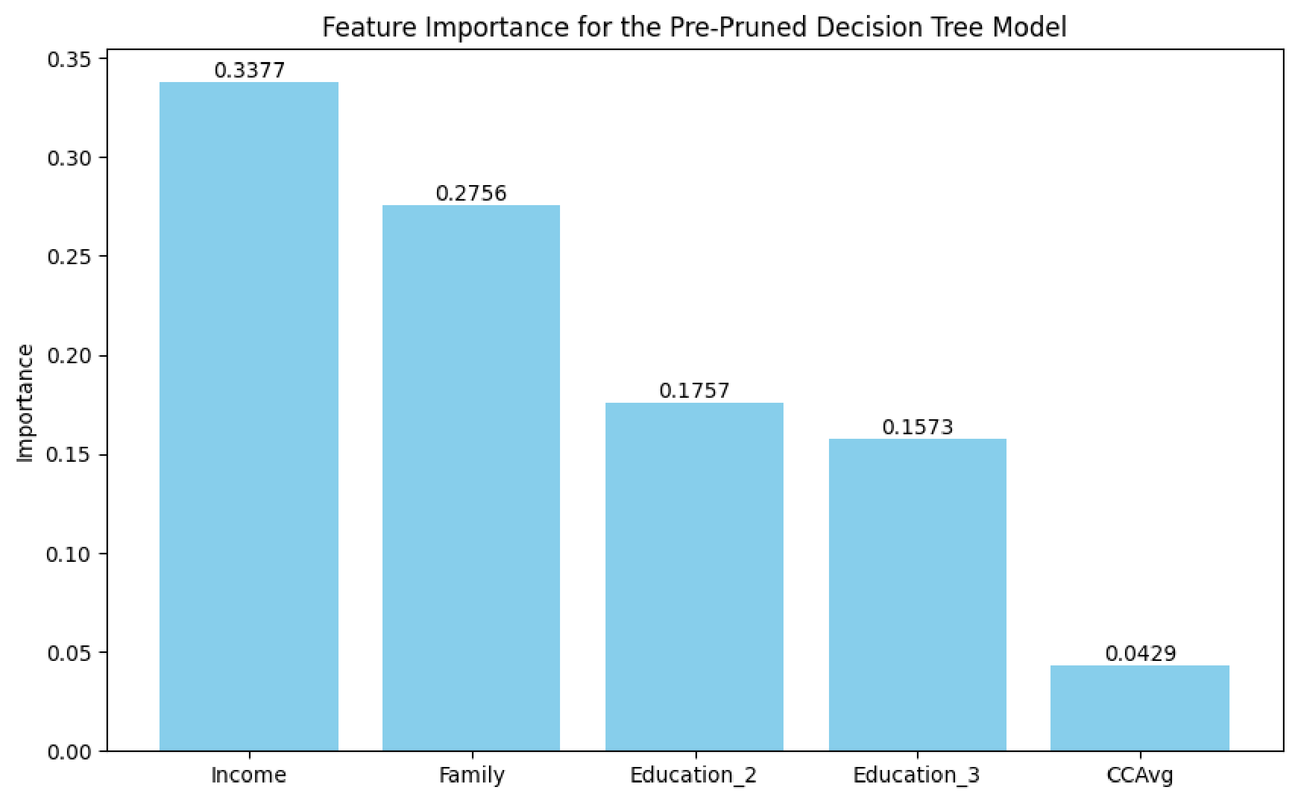
Model Performance Improvement by Pre-Pruning



Pre-Pruned Tree Visualization Overview:

Optimized Decision Tree configured with a max depth of 6, 10 leaf nodes, and at least 10 samples per leaf to balance complexity and generalization; random state set for consistency

Pre-Pruned Decision Tree Feature Importance

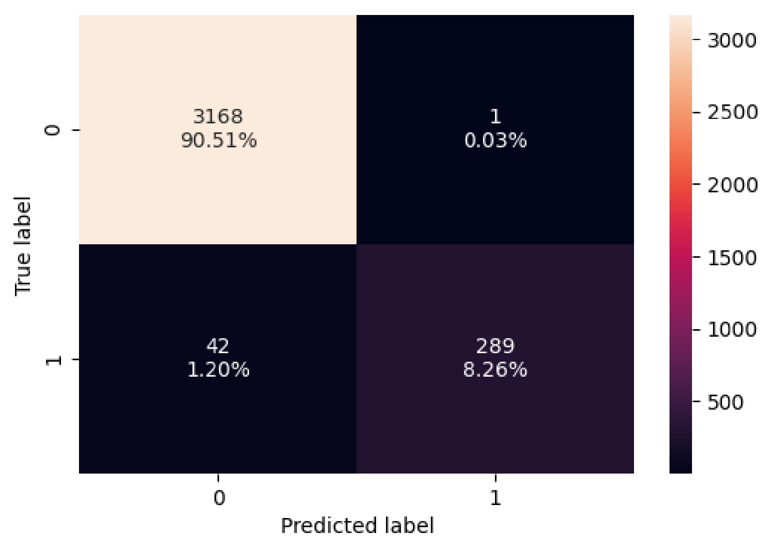


Key Predictors

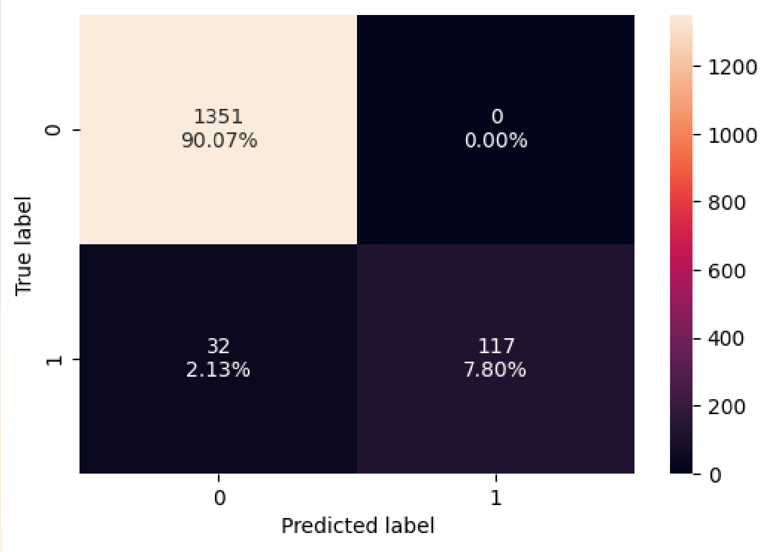
- **Income:** Dominant predictor with a 0.3377 importance score
- **Family Size**: Significant at a 0.2756 importance score
- **Higher Education (Graduate and Professional):** Notable at 0.1757 and 0.1573 importance scores
- **Credit Card Spending (CCAvg):** Minor yet relevant with a 0.0429 score

Pre-Pruned Decision Tree Model Performance Summary

Training Data Performance



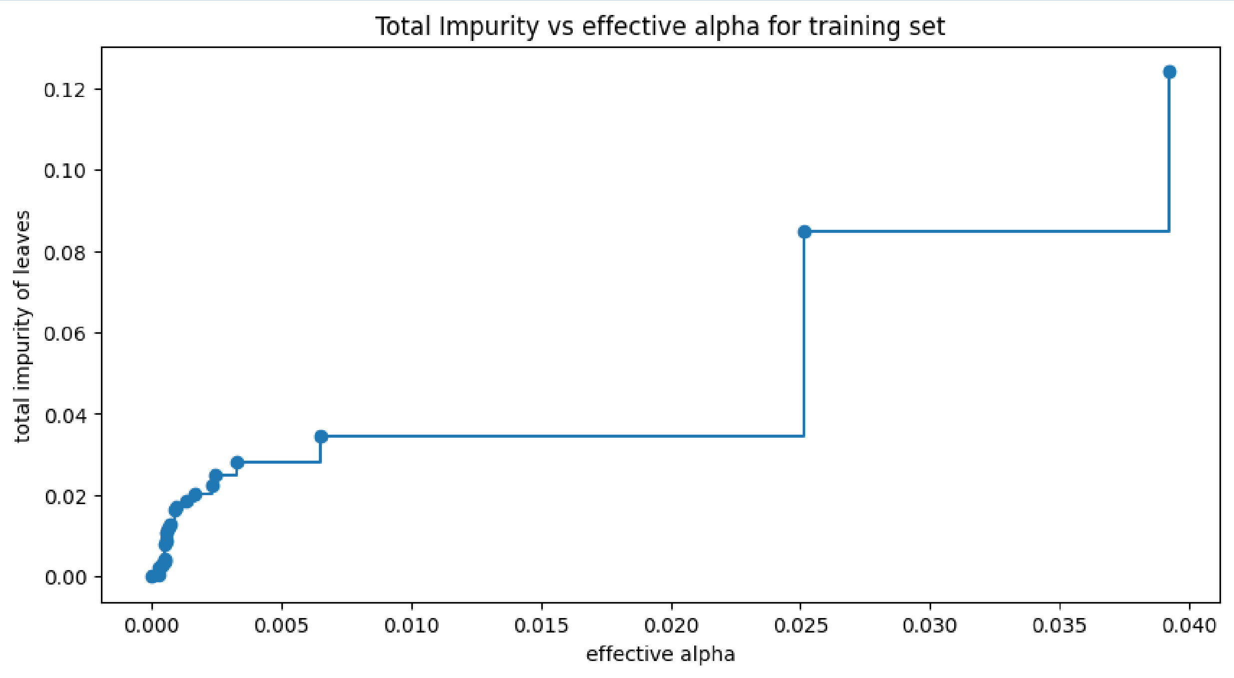
Test Data Performance



	Accuracy	Recall	Precision	F1
Train	0.987714	0.873112	0.996552	0.930757
Test	0.978667	0.785235	1.0	0.879699

Cost Complexity Pruning

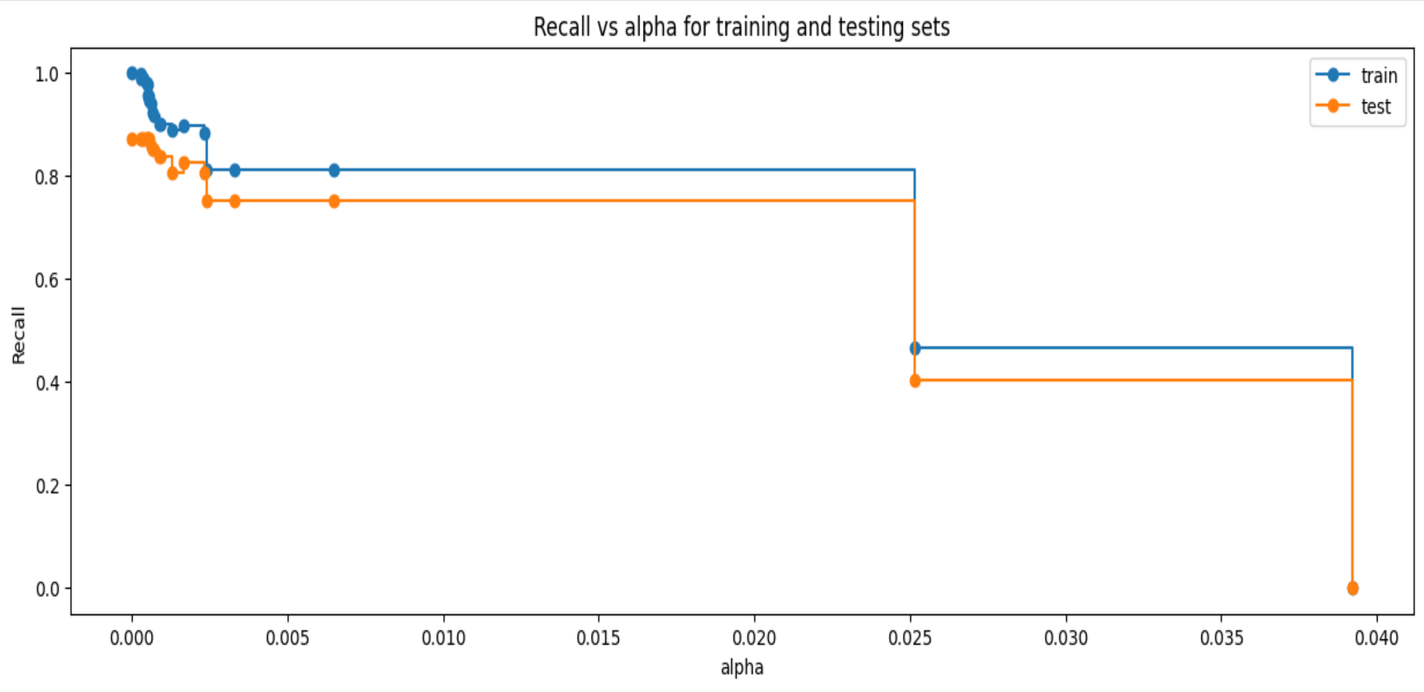
Choosing the Right Alpha for Cost Complexity Pruning



- At the effective alpha value of 0.040, the tree is fully pruned down to just one node
- Impurity begins to increase at an alpha of 0.000
- A range of 0.000 to 0.003 offers optimal alpha values for effective pruning

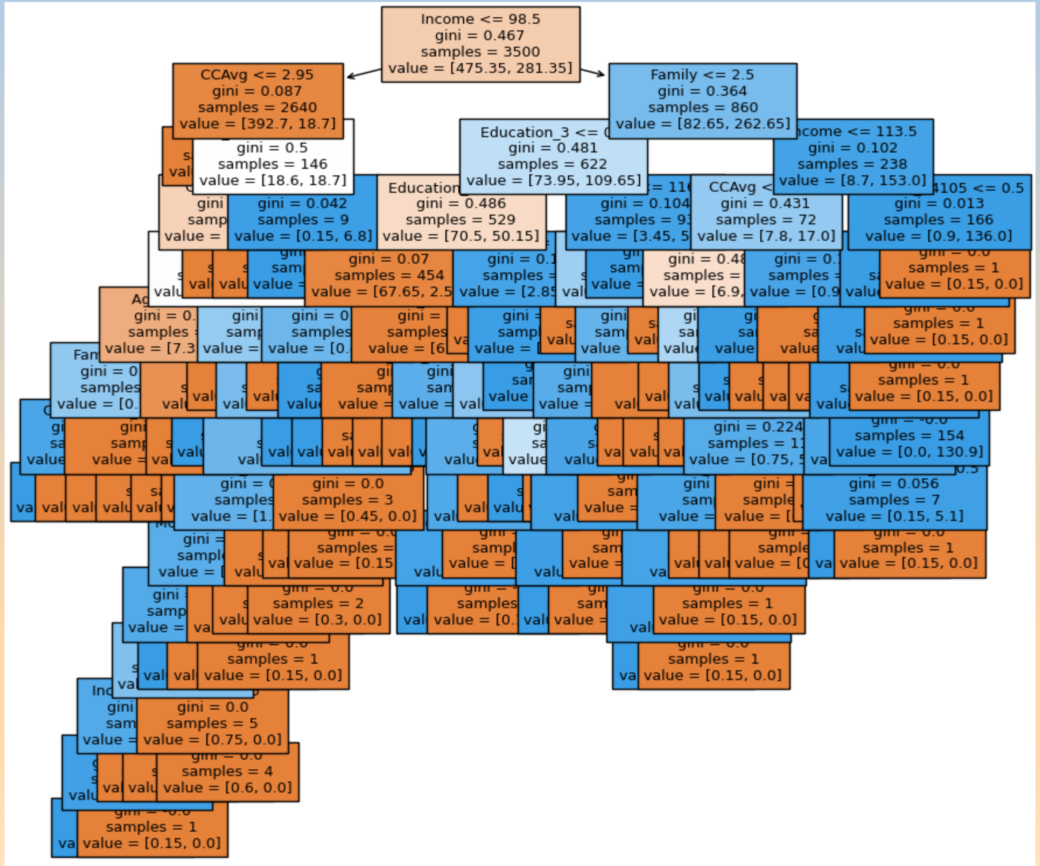
Recall Vs Alpha

Choosing the Right Alpha for Cost Complexity Pruning



First Choice of alpha value is 0.000

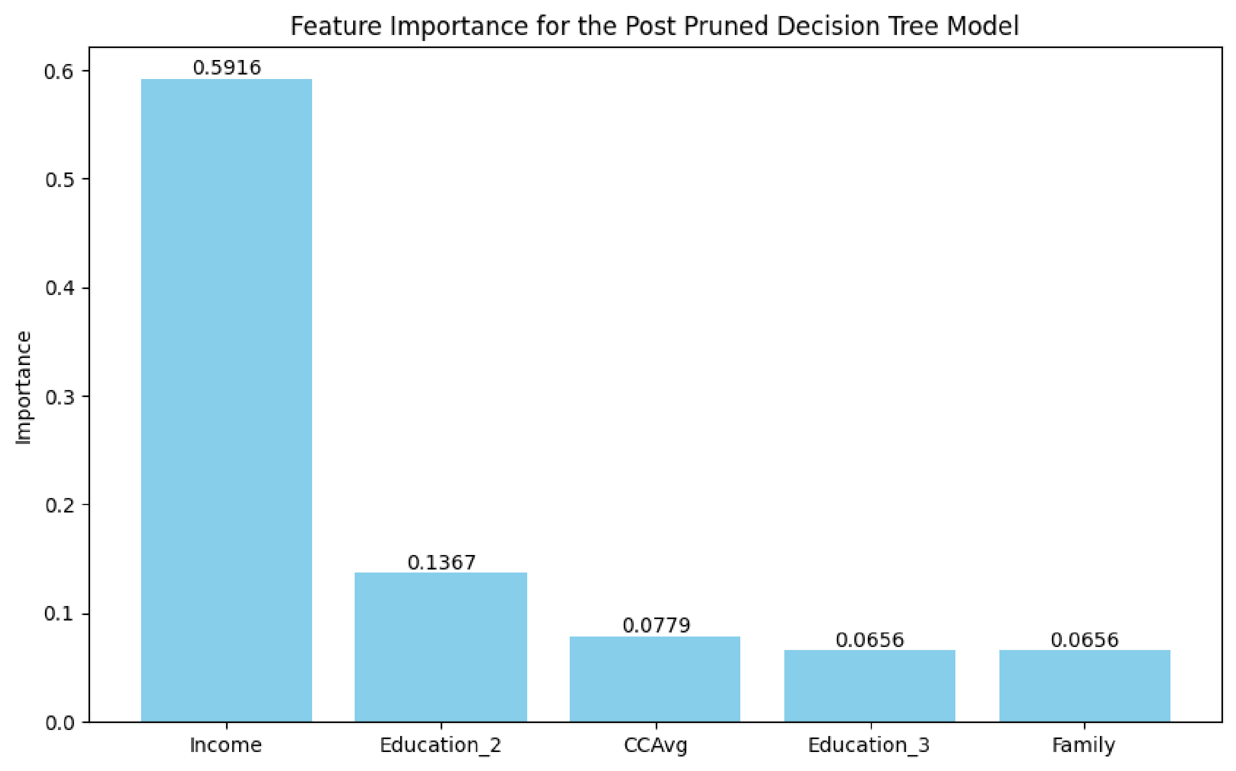
Model Performance Improvement by Post Pruning



Post Pruned Tree Visualization Overview:

- Post-Pruned Decision Tree: Utilizes a Decision Tree Classifier with class weights adjusted to {0: 0.15, 1: 0.85} to account for class imbalance, enhancing focus on the minority class, and set with a consistent random state of 1 for reproducibility
- The chosen alpha value is 0.0

Post Pruned Tree Feature Importance

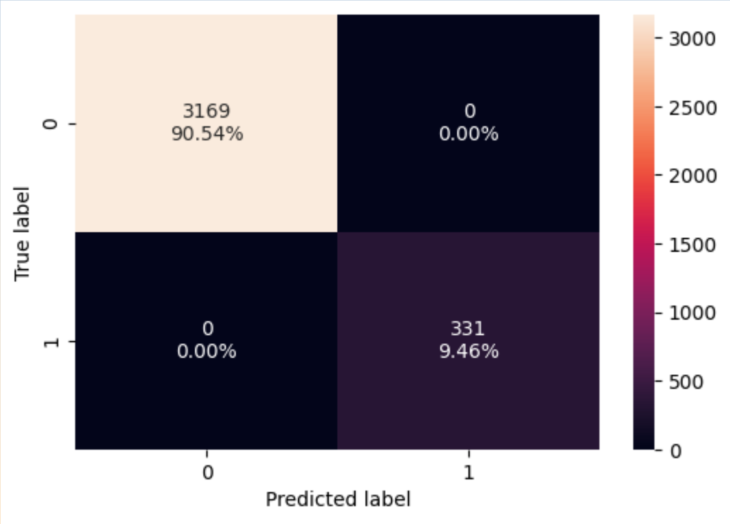


Key Predictors

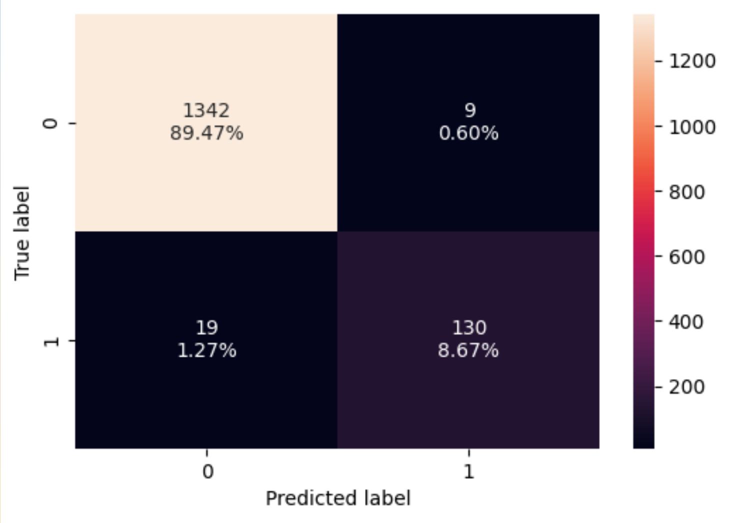
- **Income:** Dominant predictor with a 0.5916 importance score
- **Education (Graduate)** Significant at 0.1367
- **Credit Card Spending (CCAvg):** Notable yet relevant with a 0.0779 importance score
- **Education (Professional) and Family Size :** Minor at 0.0656 importance score

Post Pruned Decision Tree Model Performance Summary

Training Data Performance



Test Data Performance



	Accuracy	Recall	Precision	F1
Train	1.00	1.00	1.00	1.00
Test	0.981333	0.872483	0.935252	0.902778

Comparison of Model Performance Summary

Base Model Metrics

	Accuracy	Recall	Precision	F1
Train	1.000	1.000000	1.000000	1.000000
Test	0.981333	0.872483	0.935252	0.902778

Pre-Pruned Model Metrics

	Accuracy	Recall	Precision	F1
Train	0.987714	0.873112	0.996552	0.930757
Test	0.978667	0.785235	1.0	0.879699

Post-Pruned Model Metrics

	Accuracy	Recall	Precision	F1
Train	1.00	1.00	1.00	1.00
Test	0.981333	0.872483	0.935252	0.902778

Rationale for Selecting the Base Model Decision Tree

Objective Alignment: The primary goal is to maximize the number of loan takers from the bank's pool of customers, efficiently targeting potential loan customers while minimizing unnecessary outreach

Balanced Performance: The base model offers a superior balance of recall and precision, critical for effectively identifying the right customers without excessive false positives

High Recall: The base model's higher recall (0.872483) on test data effectively captures more actual loan takers, crucial for meeting the bank's goal of increasing loan uptake

Effective Precision: Despite being slightly lower than the pre-pruned model, the base model's precision (0.935252) remains strong, ensuring efficient identification of genuine loan takers

Resource Efficiency: The base model's optimal recall and precision balance enhances marketing efficiency, potentially reducing costs and boosting loan conversions

Model Robustness: The base model, free from depth and leaf node limits, captures complex data patterns unlike the other models effectively, suits the bank's varied customer profiles

Strategic Decision Making: The base model provides actionable insights for strategic decision-making, enabling targeted marketing based on reliable predictions of loan uptake

The End