

The Visa Fr Lisa Project



DATA COLLECTION / CLEANING

The loan dataset was collected from Galaxy Bank's customer database. It contains 5,000 rows of data, each representing a different customer. The dataset has 14 columns, each of which represents a different feature about the customer. The dataset did not have any empty cells, so I made sure that each feature was assigned the correct data type.

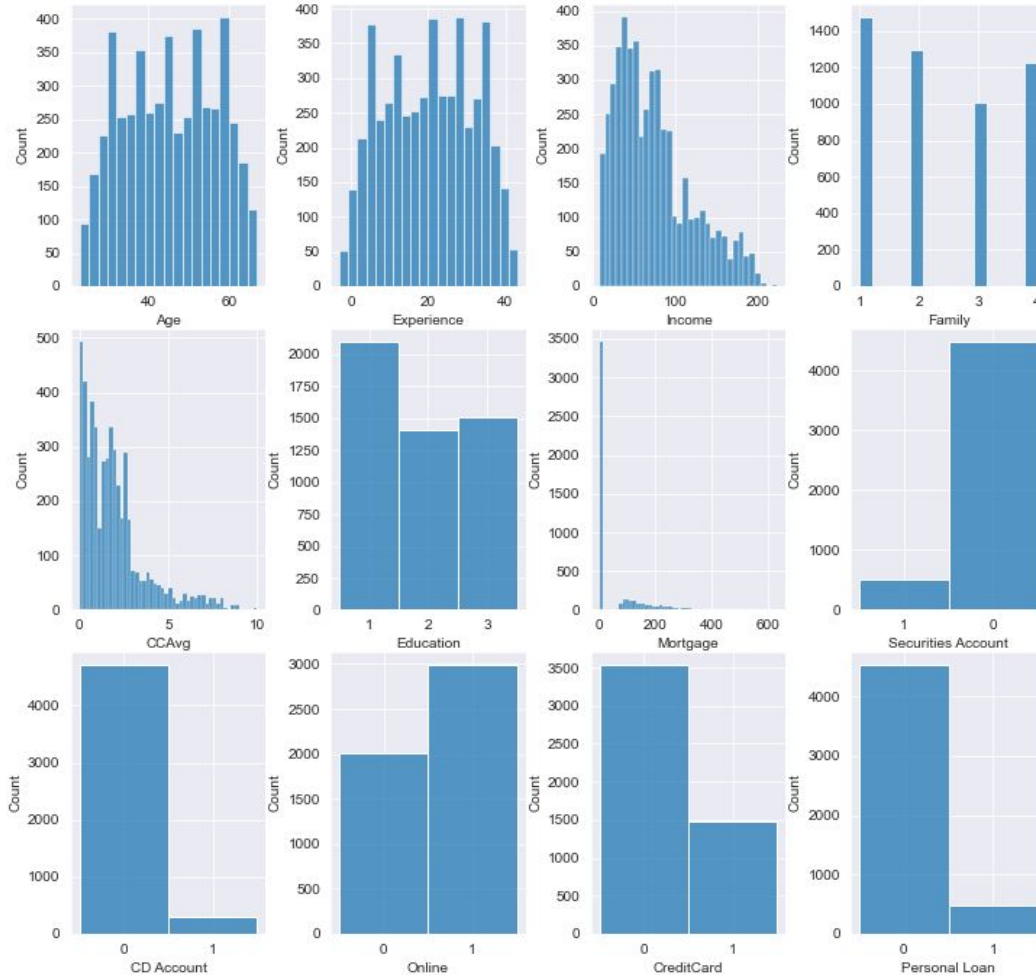
Here is a list of the features in the dataset:

- **ID:** A unique identifier for each customer.
- **Age:** The customer's age in years.
- **Experience:** The customer's years of experience in their current job.
- **Income:** The customer's annual income in US dollars.
- **Zip Code:** The customers Zip code
- **Family:** The number of the family members.
- **CCAvg:** The customer's average credit card spending per month in US dollars.
- **Education:** The customer's highest level of education.
- **Mortgage:** The number of Mortgages a customer has.
- **Personal Loan :** Tells if a customer accepts the Loan or not
- **SecuritiesAccount:** Whether or not the customer owns any securities, such as stocks or bonds.
- **CDAccount:** The amount of cash deposits the customer has in their bank account.
- **Online:** whether the customer does Online Banking or not.
- **Credit Cards:** The number of credit cards the customer has.



DATA EXPLORATION

Histograms of the Columns in the dataset



Histograms: In the loan dataset, the Age and Experience features are almost normally distributed, while the **CCAvg**, **Income**, and **Mortgage** features are skewed to the left. This means that the majority of the values for these features are closer to zero. The other features in the dataset are categorical, which means that they can only have two possible values. As a result, these features do not carry as much information as the numerical features.

Here is a more detailed explanation of the different types of distributions:

- **Normal distribution:** A normal distribution is a symmetrical distribution that is bell-shaped. The majority of the values in a normal distribution are centered around the mean, and the values gradually decrease as you move away from the mean.
- **Skewed distribution:** A skewed distribution is a distribution that is not symmetrical. The values in a skewed distribution may be clustered towards one end of the distribution, or they may be spread out evenly.

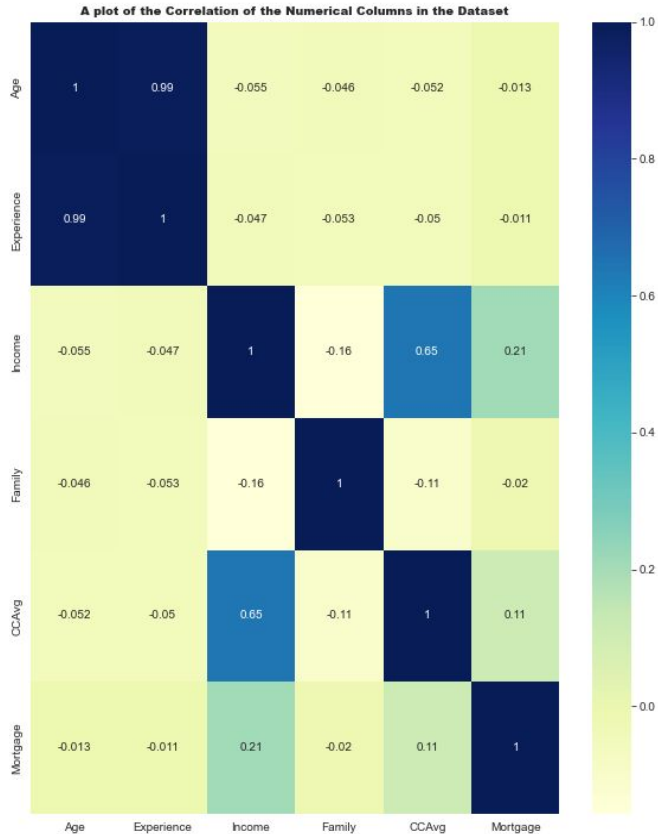
Scatter Matrix

The **Age** and **Experience** features in the dataset are highly correlated, which makes sense because as people get older, they tend to gain more experience. The **Income** and **CCAvg** (Average Credit spent) features are also correlated, because people with higher incomes tend to have more disposable income to spend on credit. The other features in the dataset are not as correlated with each other.

Here is a more detailed explanation of correlation:

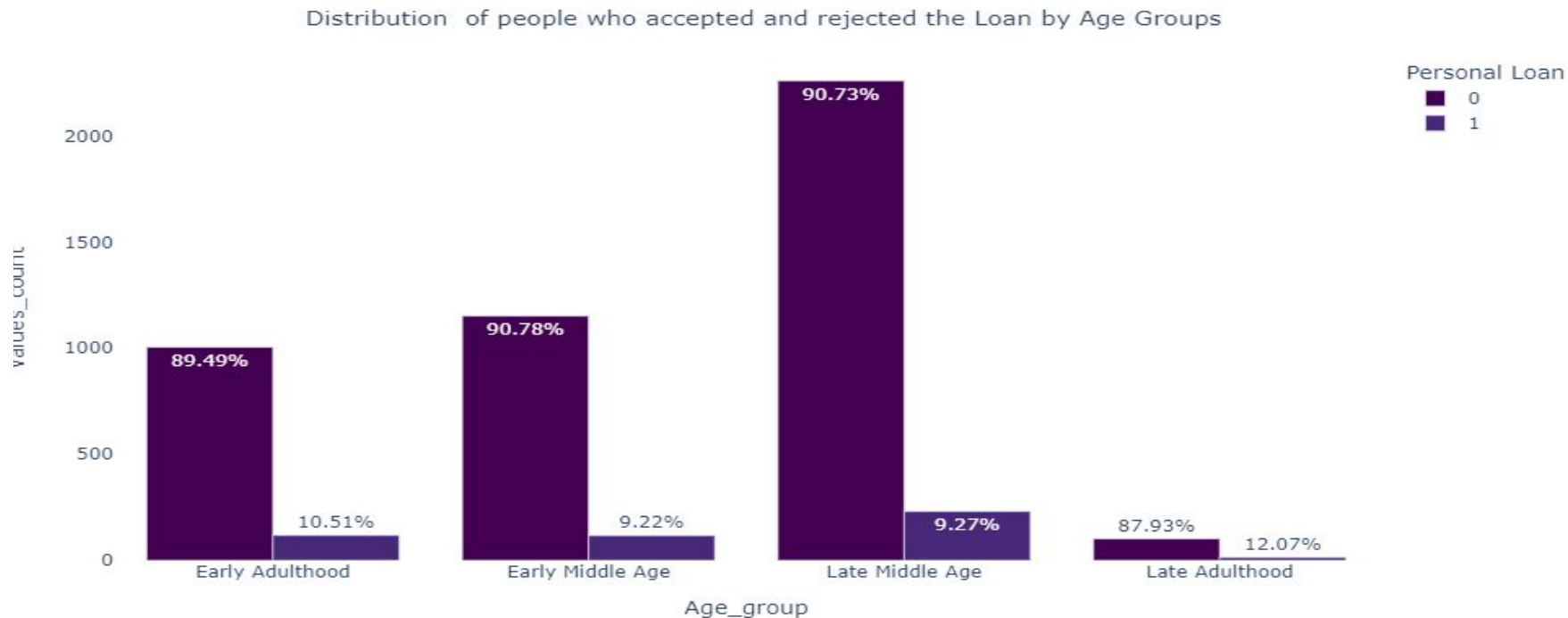
Correlation: Correlation is a measure of how two variables are related to each other. A correlation coefficient can range from -1 to 1 . A correlation coefficient of 1 indicates that the two variables are perfectly correlated, meaning that they move in the same direction at the same time. A correlation coefficient of -1 indicates that the two variables are perfectly negatively correlated, meaning that they move in opposite directions at the same time. A correlation coefficient of 0 indicates that the two variables are not correlated at all.

In this case, the correlation between **Age** and **Experience** is **0.99**, which is a **very strong correlation**. The correlation between **Income** and **CCAvg** is **0.65**, which is also a **strong correlation**. The correlation between the other features in the dataset is much weaker.



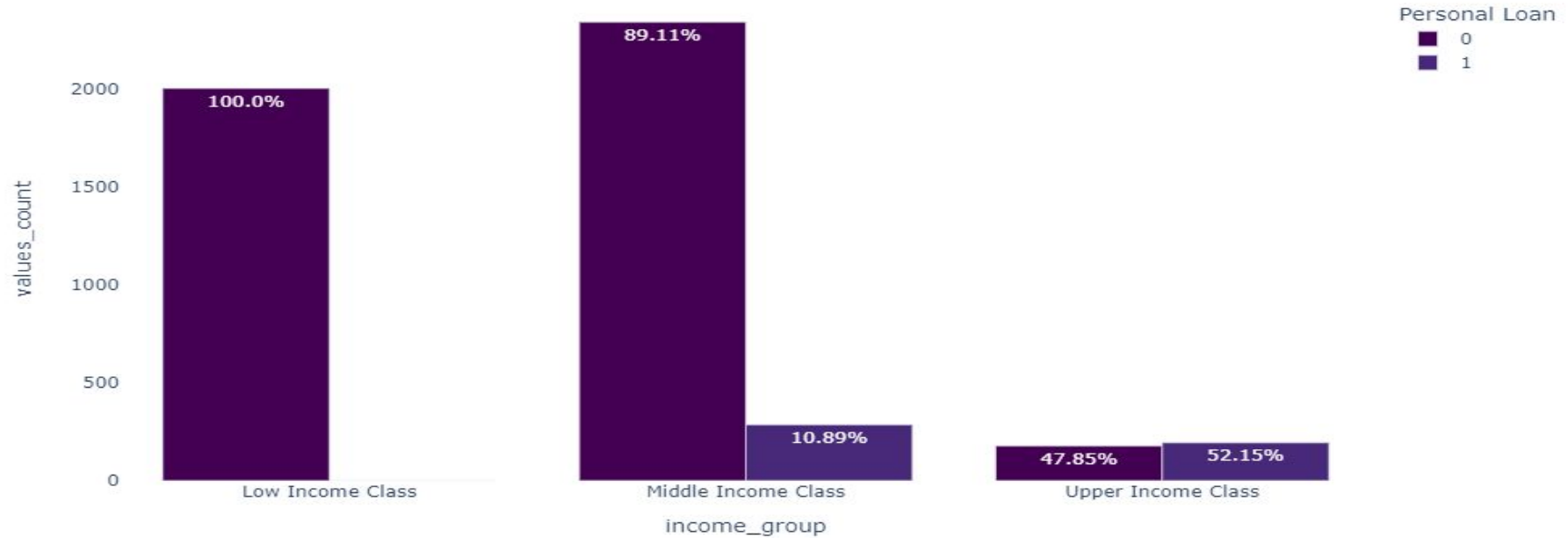


DATA VISUALIZATION



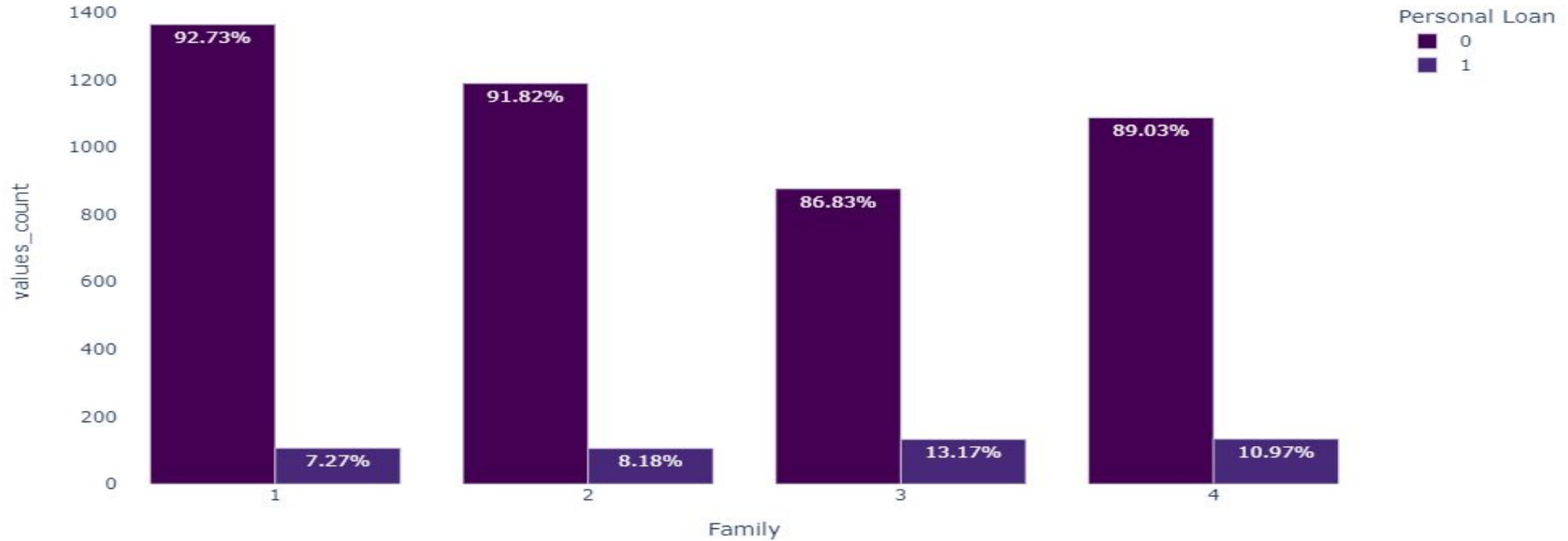
The majority of Galaxy Bank's customers are between the ages of 45 and 64. This age group had the highest number of customers who accepted loans, even though customers between 65 and above had the highest acceptance rate. However, all age groups were represented among loan recipients.

Distribution of People who accepted and rejected the Loans by income group



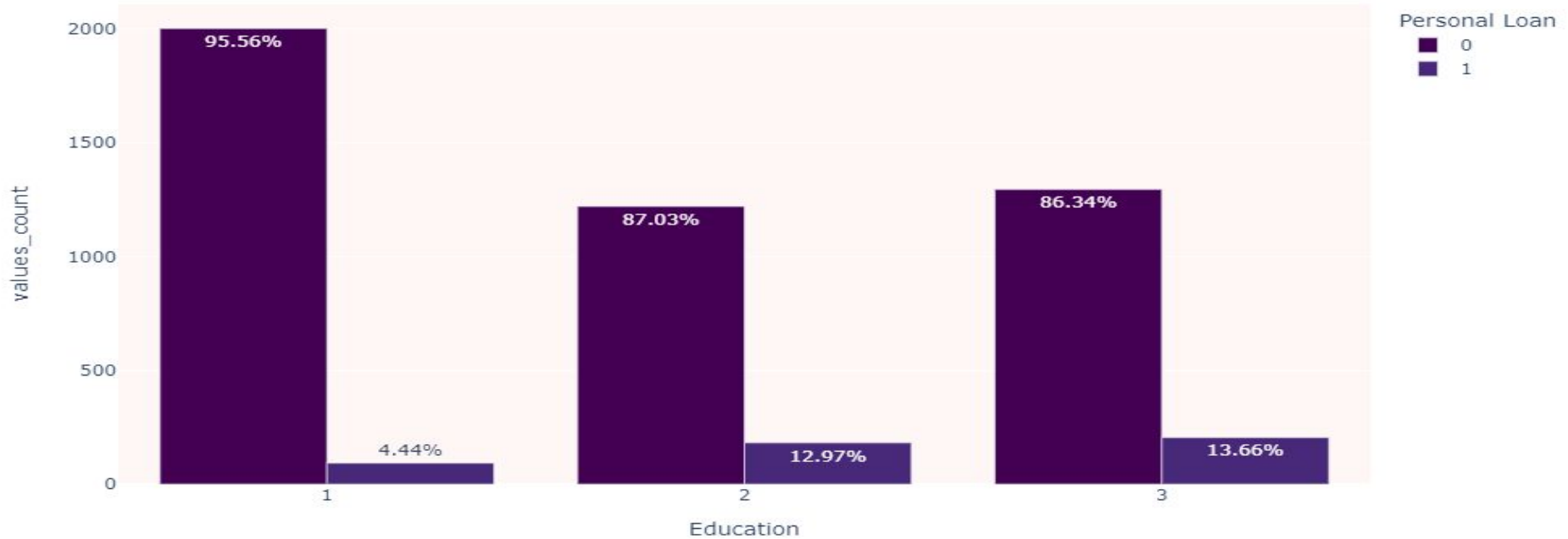
The data shows that loans were only accepted by customers in the middle-income and upper-income groups. No customers in the lower-income group accepted a loan.

Distribution of People who accepted and rejected the Loans by Family group



All family groups were approved for loans, and there was no significant difference in approval rates between family groups..

Distribution of People who accepted and rejected the Loans by Education



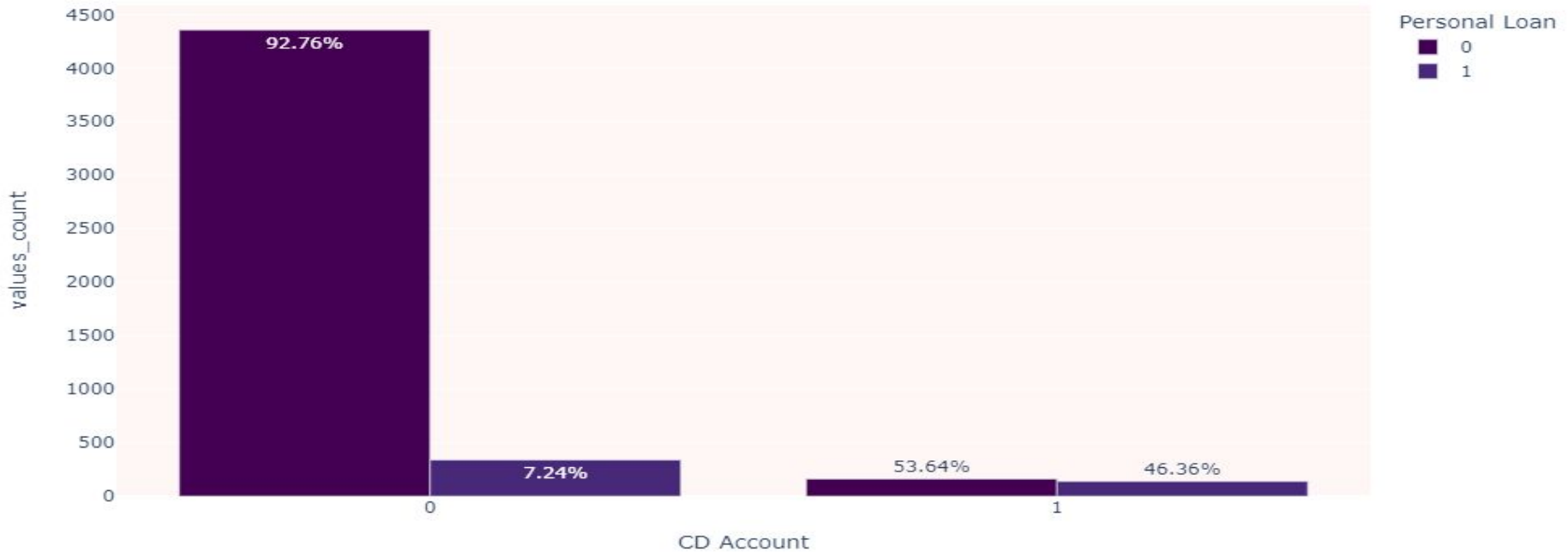
The dataset shows that there are three levels of education. The acceptance rate for loans increased as the level of education increased.

Distribution of People who accepted and rejected the Loans by Securities



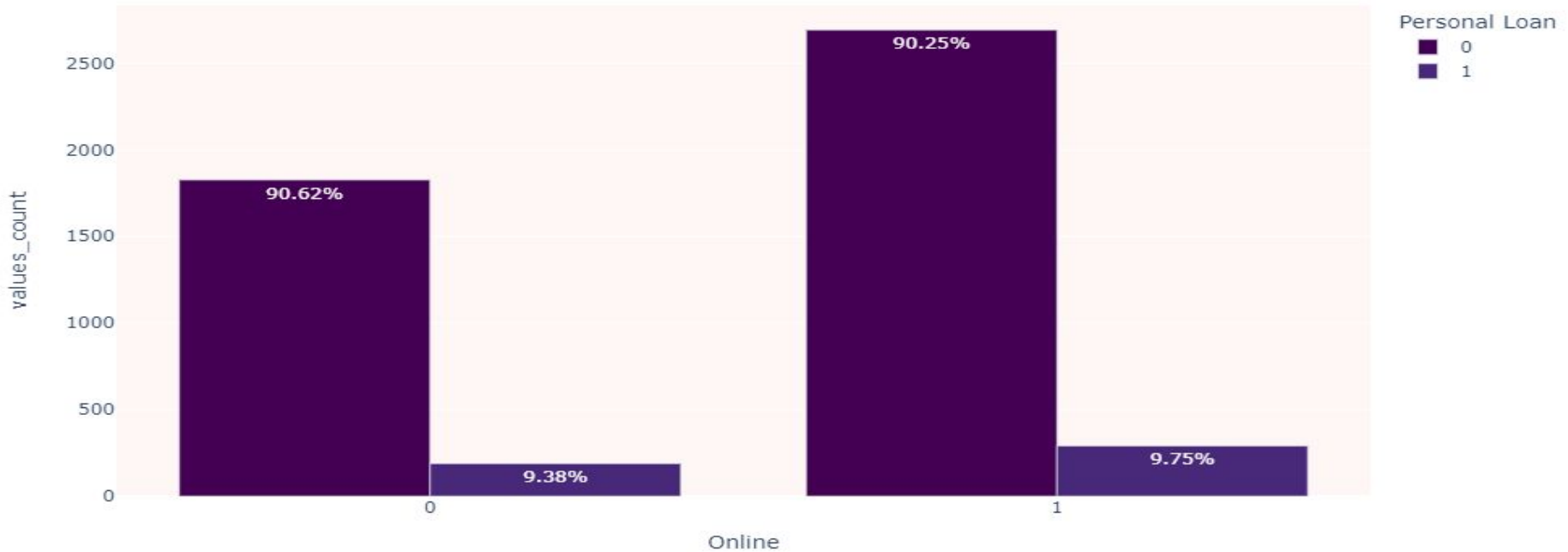
The data shows that the majority of customers do not have any securities, even among those who accepted loans. However, customers who have securities had a higher acceptance rate for loans.

Distribution of People who accepted and rejected the Loans by CD Accounts



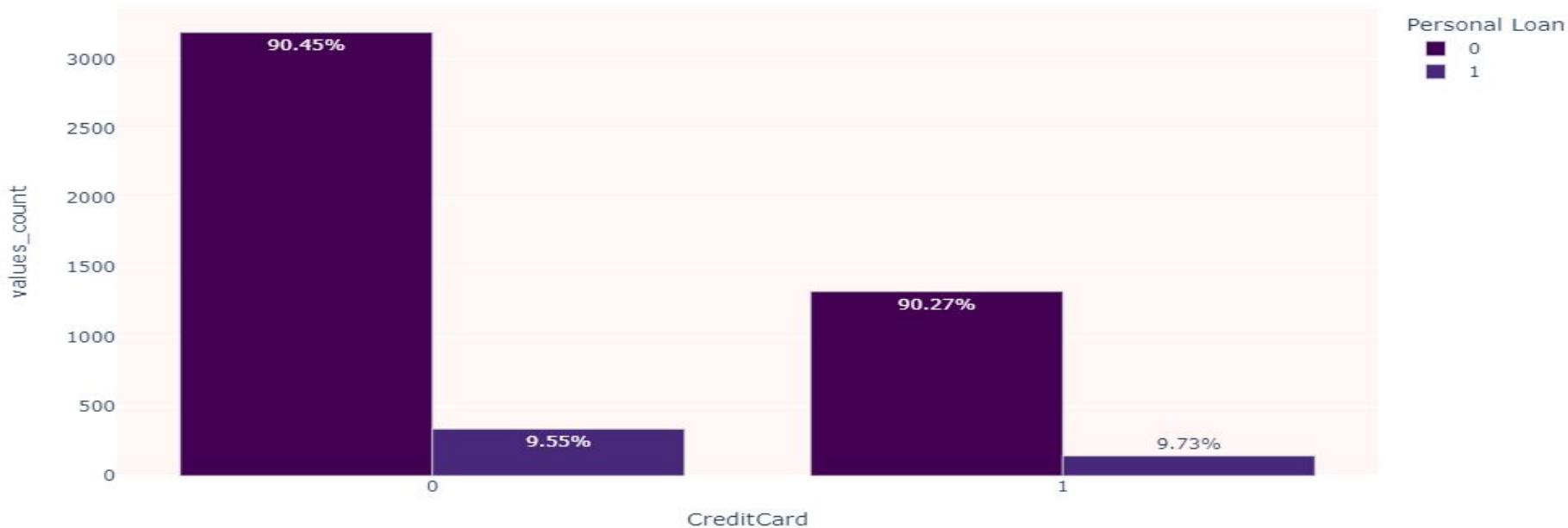
The data shows that the majority of customers who did not accept loans did not have cash deposits. The same is true for customers who accepted loans, as most of them did not have cash deposits either. However, customers who did have cash deposits had a higher acceptance rate.

Distribution of People who accepted and rejected the Loans by customers who use Online Banking



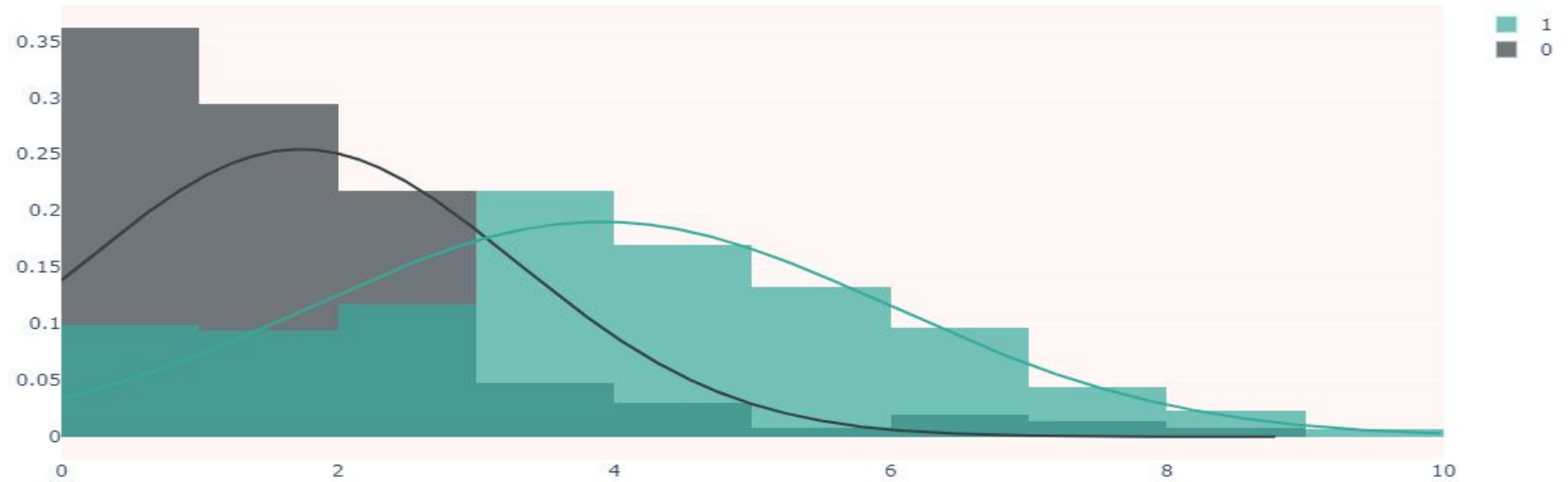
The data shows that the acceptance rate for loans is higher among customers who use online banking than among customers who do not use online banking.

Distribution of People who accepted and rejected the Loans by customers who have Credit Cards



The data shows that customers with credit cards have a higher acceptance rate for loans than customers without credit cards. In general, most customers do not have credit cards.

CCAvg for Customers



On average, customers who accepted loans had a higher average monthly credit card spending than customers who did not accept loans.



MACHINE LEARNING

INITIALIZING THE MODELS AND TRAINING THE DATA

To identify customers who would likely accept loans, I trained 6 models from the scikit-learn library on the data. I then evaluated the accuracy, precision, recall, and F1 score of each model.

The Table to the right is the table that shows how each model performs on Unseen data.

	Accuracy	Precision	Recall	F1
Logistic Regression	0.909	0.634615	0.314286	0.420382
Naive Bayes	0.901	0.524194	0.619048	0.567686
KNN	0.899	0.625000	0.095238	0.165289
Random Forests	0.983	0.978261	0.857143	0.913706
Support Vector Machine	0.911	0.648148	0.333333	0.440252
Decision Trees	0.989	0.989583	0.904762	0.945274

RE-EVALUATION OF MACHINE LEARNING MODELS

From the previous models, the models that fared better was the Random Forests and Decision Trees. After re-evaluating the models with hyper-parameters tuning, the best model for this classification task is the Random Forests

	Accuracy	Precision	Recall	F1
Random Forests	0.989	0.989583	0.904762	0.945274
Decision Trees	0.979	0.937500	0.857143	0.895522



COMMUNICATION

RECOMMENDATION

Galaxy Bank can improve its customer acceptance rate by:

- *Using machine learning to classify customers into income classes:* They can then focus their marketing on the middle and upper-income classes, as these groups have a higher acceptance rate.
- *Identifying customers with higher educational levels:* These customers have a higher acceptance rate, so Galaxy Bank should target them more.
- *Identifying customers who do not have securities, cash deposits, or credit cards:* These customers may be more likely to accept a loan, so Galaxy Bank should consider mapping them to either the middle or upper-income classes.

Galaxy Bank can also improve its customer acceptance rate by:

- *Making the application process more streamlined and easy to understand:* This will make it more likely that customers will complete the application and submit it.
- *Providing more information about the loan terms and conditions:* This will help customers understand what they are agreeing to when they take out a loan.
- *Offering competitive interest rates and fees:* This will make Galaxy Bank's loans more attractive to customers.

By taking these steps, Galaxy Bank can improve its customer acceptance rate and attract more customers.

The background is a solid dark blue. In the top right corner, there is a decorative pattern of overlapping triangles in various shades of blue, including a lighter sky blue and a darker navy blue.

THANK YOU!!!!