"Enhancing Personal Loan Campaigns: A Data-Driven Approach in the Indian Banking Sector"

Agenda

- Identifying key factors that affect a customer’s decision to take a personal loan.
- Analyse the strength of the impact on these factors on personal loan acceptance
- Develop a prediction model with high accuracy to target potential customers effectively in future marketing campaigns

Background of the problem statement

In the competitive Indian banking sector, our client, a leading private bank, seeks to improve its personal loan campaigns through data exploration and predictive modelling. India's personal loan market, valued at over $20 billion, attracts young professionals with rates between 10% to 20%. Amidst intense competition and regulatory oversight, customer acquisition and retention are vital for growth.

With 500+ branches, our client aims to cater to the diverse needs of its large customer base by offering personal loans to depositors. The challenge is to identify potential customers and refine marketing strategies.

The bank, known for safety and reliability, serves 11.5 million depositors through 300+ branches. Leveraging this trust, the bank aims to diversify by offering personal loans, using data-driven marketing.

The primary challenge is to identify potential personal loan customers within the vast depositor base and develop an effective data-driven strategy. The bank's objectives include identifying key factors affecting loan acceptance and building a predictive model for future campaigns.

Lets first discuss a straightforward question that came to our minds, “Why do people take personal loans?”.

There are a variety of reasons for a person's decision to take a personal loan, but some of the common reasons in India are debt consolidation, medical expenses, vehicle financing, large purchases, tax debt, education expenses, etc,...
A person's decision in taking a personal loan may depend on several factors. But with the given dataset, we considered just the given features and some new features with correlation among the previous ones.

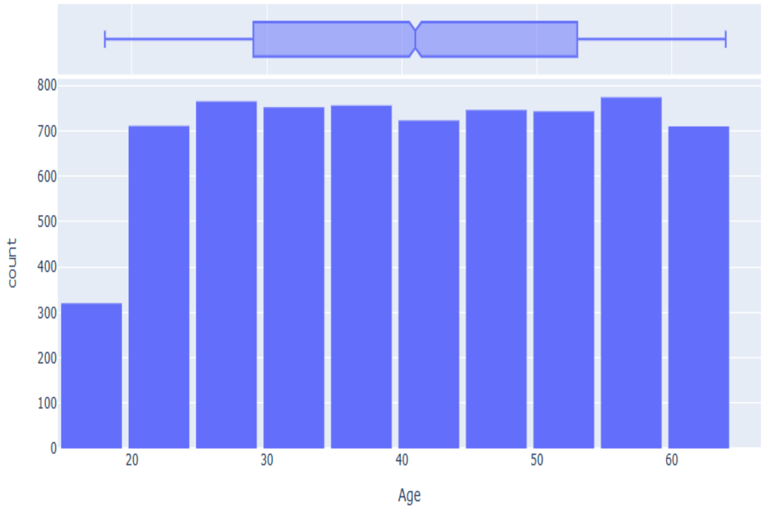
Our approach towards the problem statement

1. <u>Exploratory Data Analysis:</u>
   The dataset under consideration exhibits a uniform distribution across its continuous variables, with no null values present. We used the 'df.describe()' function to obtain a statistical summary of the dataset. This gave key insights into the data's central tendencies and dispersion, including mean, median, 25th percentile, 75th percentile, minimum, and maximum values. Further we dropped the rows where the columns "Average Purchase Value", "Total Value of Returns" and "Last Purchase Value" had negative values, as they cannot be logically. These steps helped us to prepare a relatively clean and reliable dataset, for subsequent analysis and modelling.

2. <u>Approach for potential customer data analysis :</u>
   - Gender & Age
   - Income
   - Occupation
   - Marital Status & Family Size
   - Home value and Home Ownership
   - Total Credit Card Limit and Total Credit Card Balance
   - Total Amount of Online Purchase
   - Number of Days Since Last Online Purchase
   - Locations

Data Exploration : Age Distribution

The following graph shows the age groups in our dataset

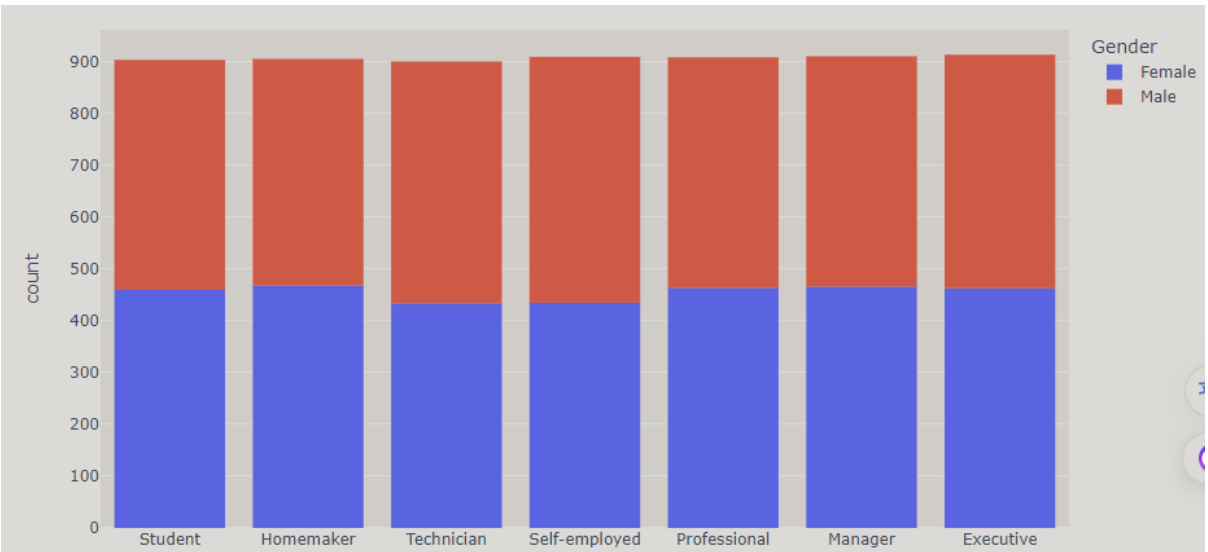


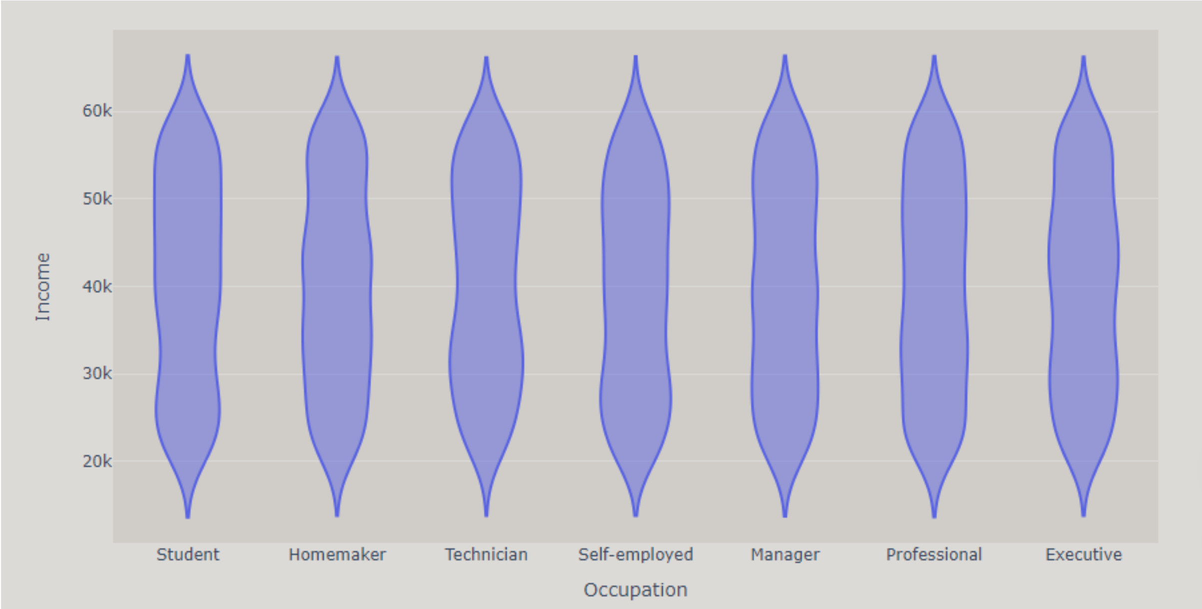
Lenders usually tend to offer a lower interest rate to salaried applicants between 30 and 50 years. This is mainly because of a stable source of income and professional work experience.

However, individuals with ages beyond 50 years usually do not have many working years left. Hence, it can become difficult for them to pay off the higher EMIs to repay the loan before retirement.

The target audience for our marketing should be focused on the age group of 30-50.

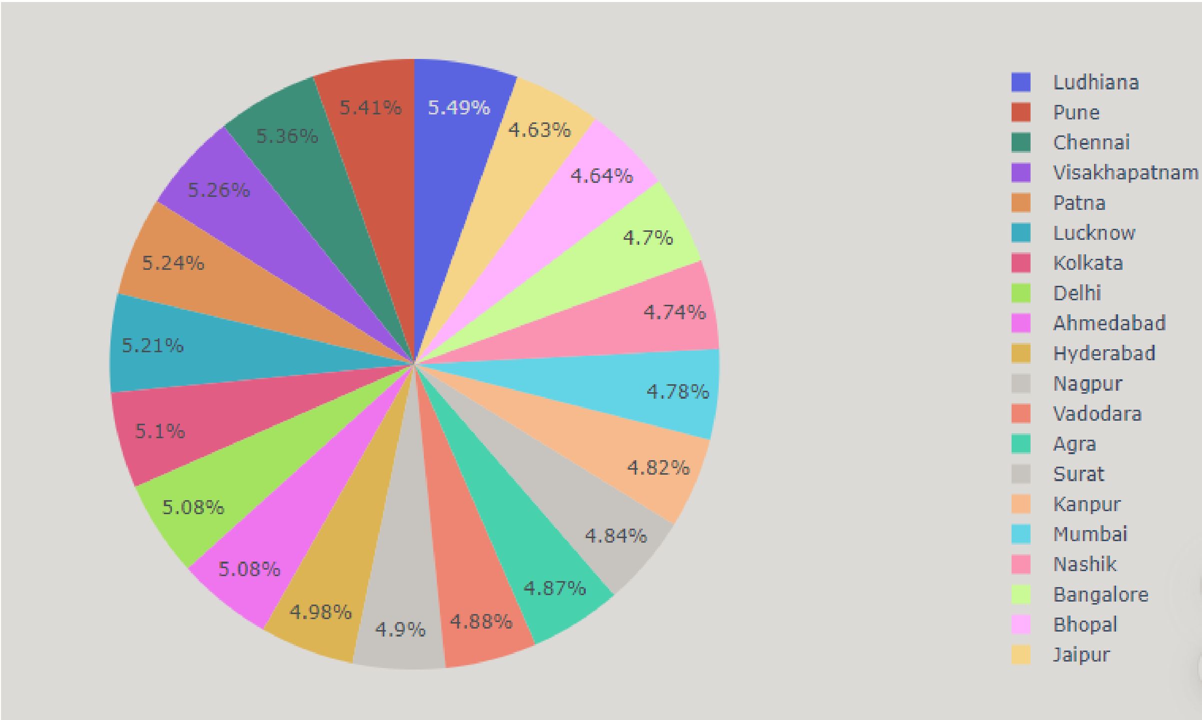
Data Exploration : Occupation



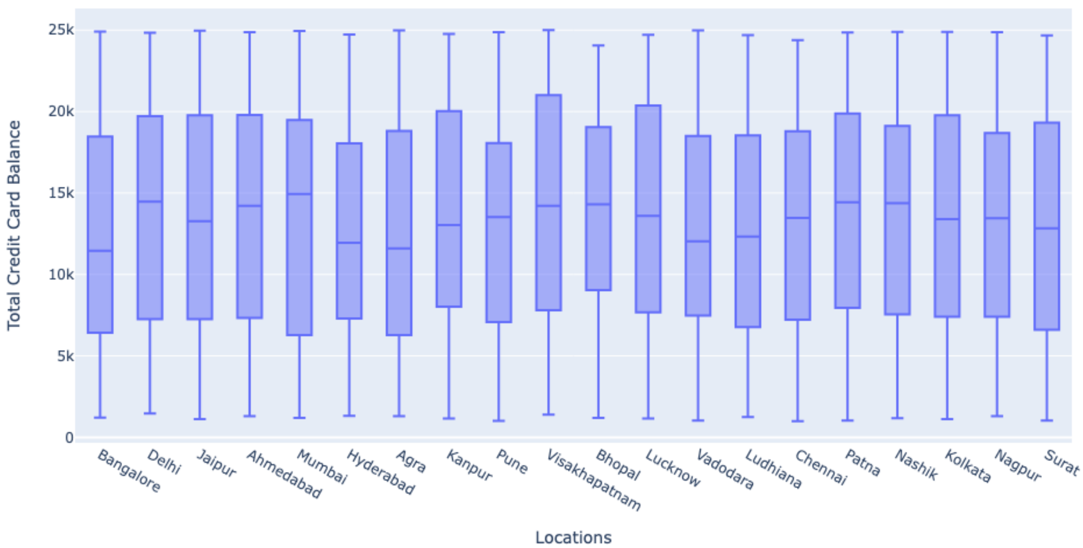


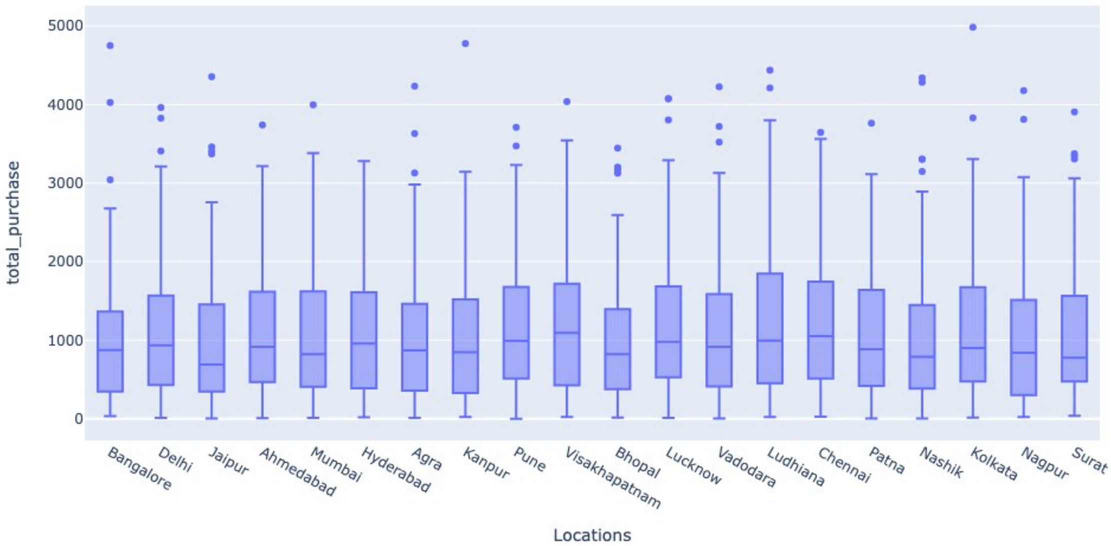
- Students are less likely to take out personal loans compared to other groups because they often have limited income
- Technicians generally have a steady income, which may make them eligible for personal loans. Technicians having lower but steady income are potential consumers as they are credible as well as they require home loans for personal needs.
- Home makers, who may not have a regular income, are less likely to take out personal loans
- Self-employed individuals who have a lower income are the most likely candidates to take a personal loan since they need funds to support their firm. Individuals who have a high salary don't require a personal loan since their business is sufficing.
- Executives,Managers and Professionals typically have stable incomes and may be more likely to qualify for personal loans. They might use loans for various purposes including home renovations, travel, or debt consolidation.

Data Exploration : Locations

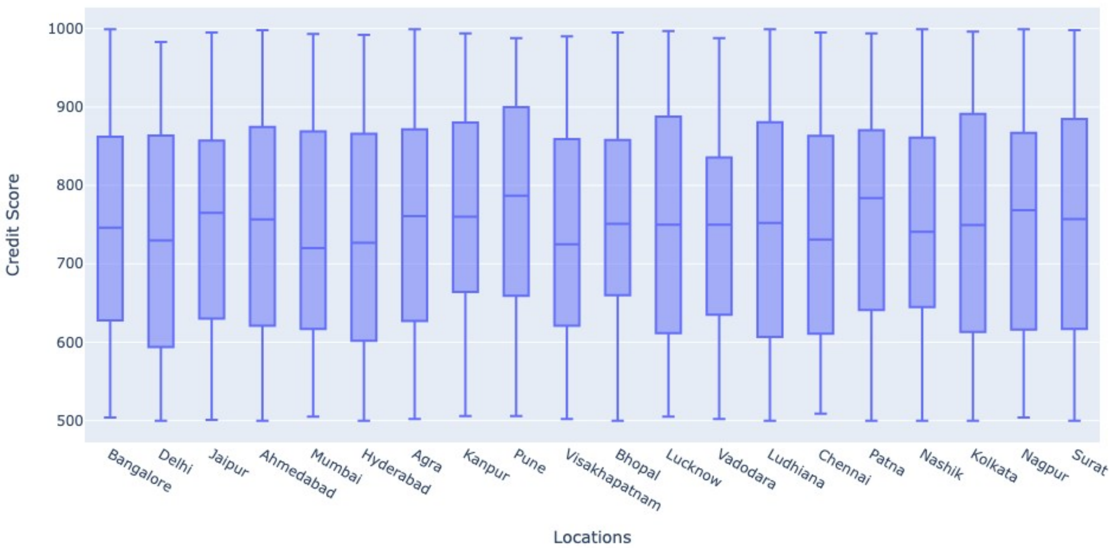


Distribution of customers across locations

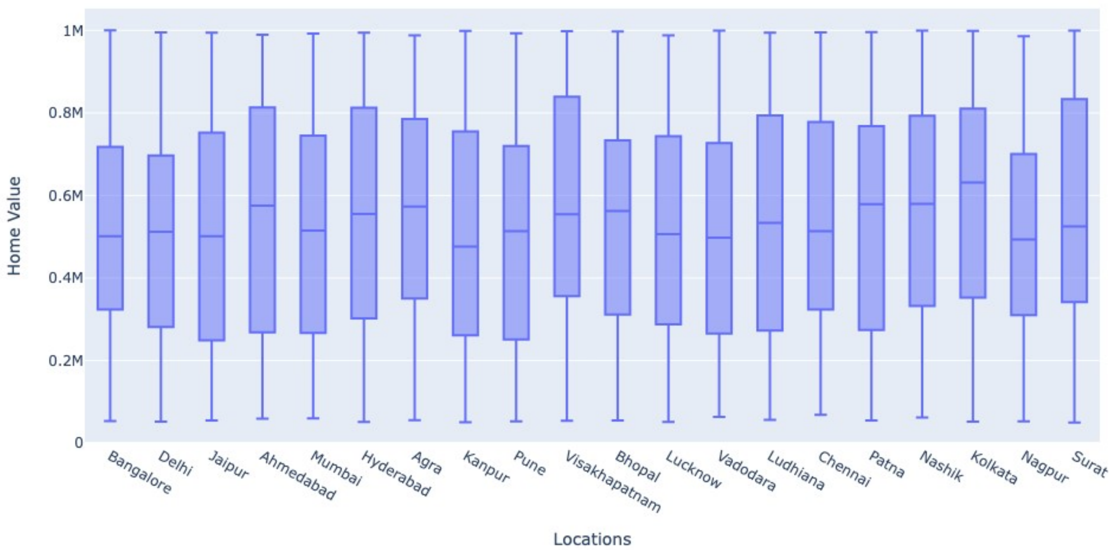



total_purchase
5000
4000
3000
2000
1000
0
Bangalore
Delhi
Jaipur
Ahmedabad
Mumbai
Hyderabad
Agra
Kanpur
Pune
Visakhapatnam
Bhopal
Lucknow
Vadodara
Ludhiana
Chennai
Patna
Nashik
Kolkata
Nagpur
Surat
Locations

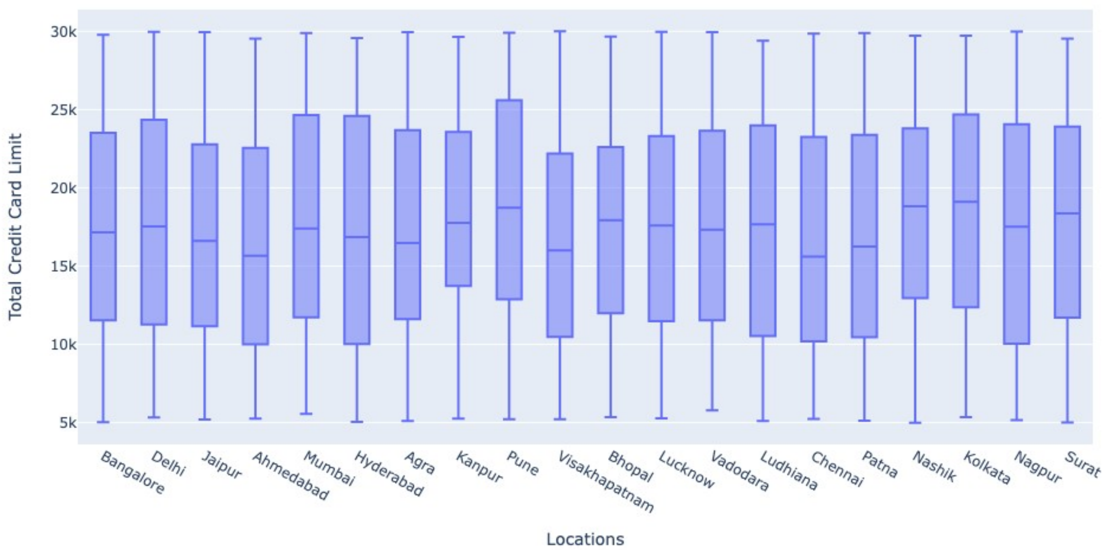
-


Credit Score
1000
900
800
700
600
500
Bangalore
Delhi
Jaipur
Ahmedabad
Mumbai
Hyderabad
Agra
Kanpur
Pune
Visakhapatnam
Bhopal
Lucknow
Vadodara
Ludhiana
Chennai
Patna
Nashik
Kolkata
Nagpur
Surat
Locations

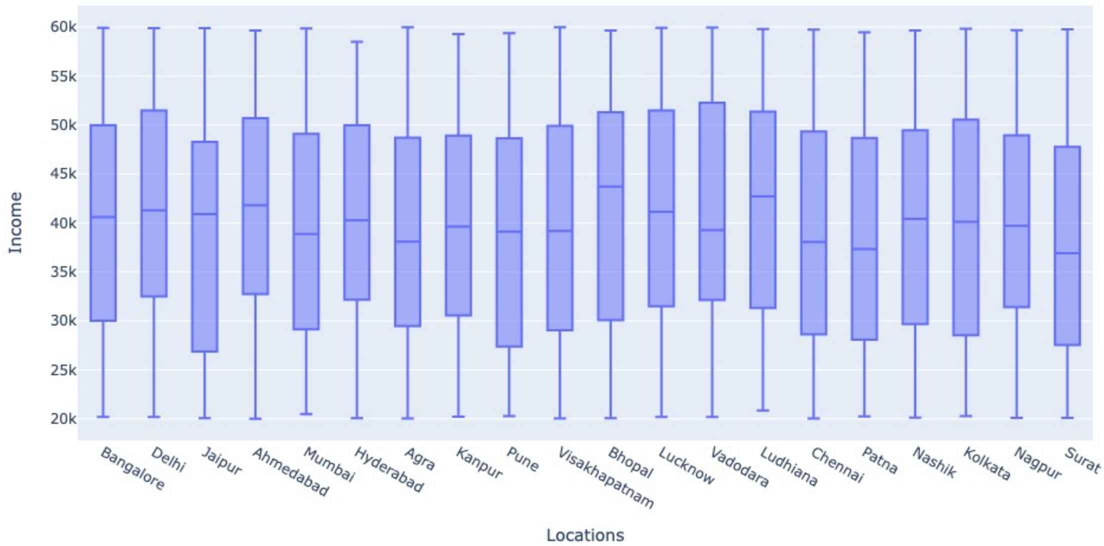
-



Home Value
1M
0.8M
0.6M
0.4M
0.2M
0
Bangalore
Delhi
Jaipur
Ahmedabad
Mumbai
Hyderabad
Agra
Kanpur
Pune
Visakhapatnam
Bhopal
Lucknow
Vadodara
Ludhiana
Chennai
Patna
Nashik
Kolkata
Nagpur
Surat
Locations



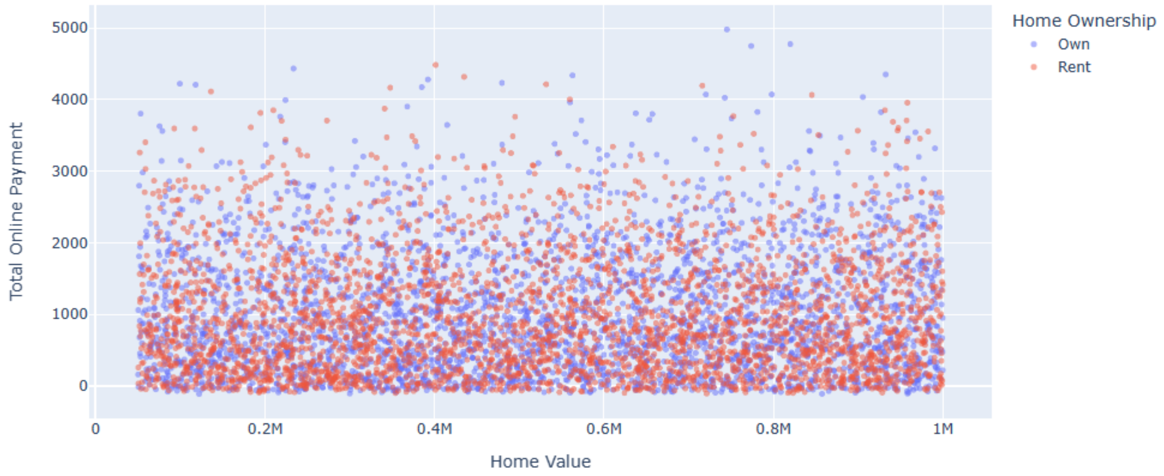
Total Credit Card Limit
30k
25k
20k
15k
10k
5k
Bangalore
Delhi
Jaipur
Ahmedabad
Mumbai
Hyderabad
Agra
Kanpur
Pune
Visakhapatnam
Bhopal
Lucknow
Vadodara
Ludhiana
Chennai
Patna
Nashik
Kolkata
Nagpur
Surat
Locations



The total purchase column was created taking the sum total of products bought and removing the cost of products returned.

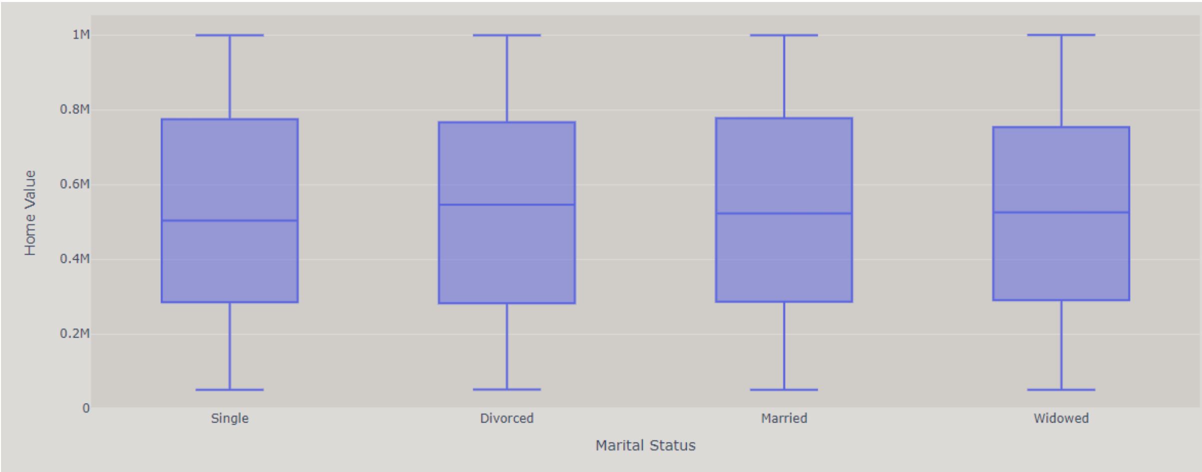
Seeing the total purchase and the total credit card limit(which is proportional to credit score) across different locations, we can see significant changes across locations. More predominantly the change is observed in **Visakhapatnam**, it is the place where people are most likely to take a personal loan.

Data Exploration : Home Values & Home Ownership



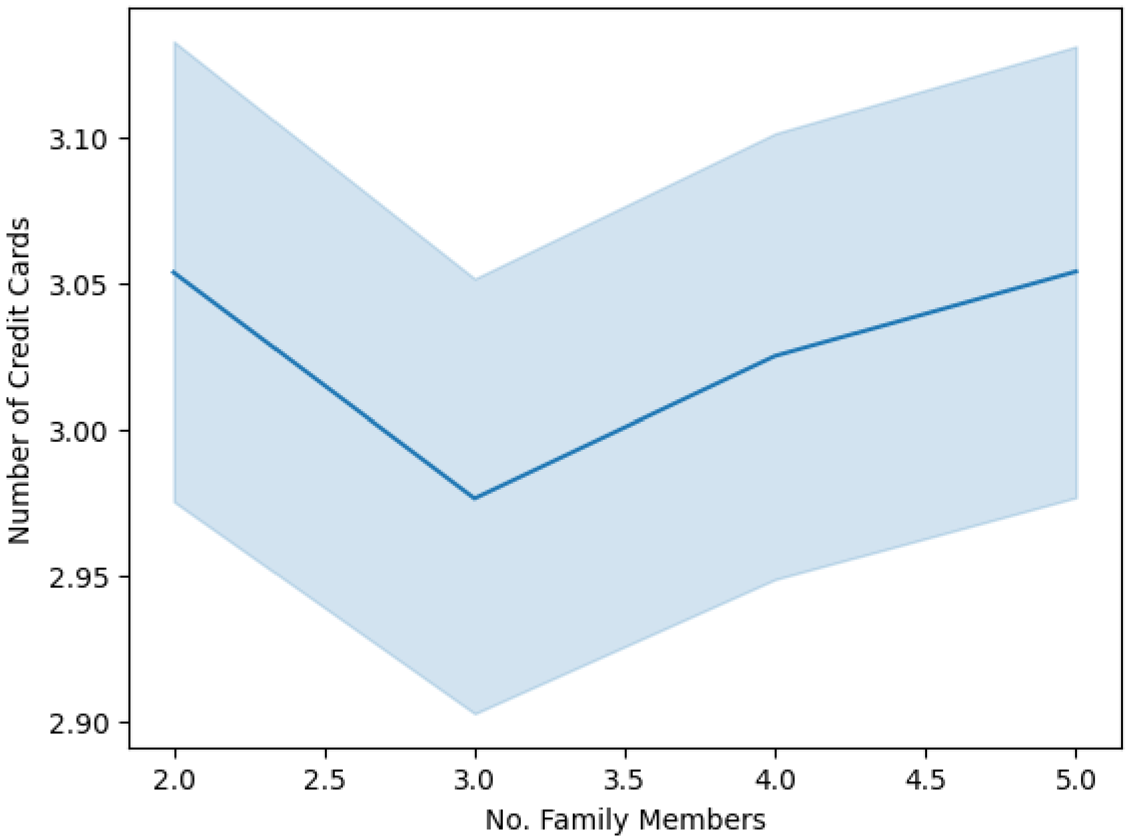
- People living in a high valued rented homes especially those having a high credit score are the most likely customers for taking a personal loan since they have a huge amount to pay and may need money for other necessities
- People having a rented house of high value who also have high online purchases are most likely to take a personal loan since they have high spendings and require more money.

Data Exploration : Marital Status & Family Size



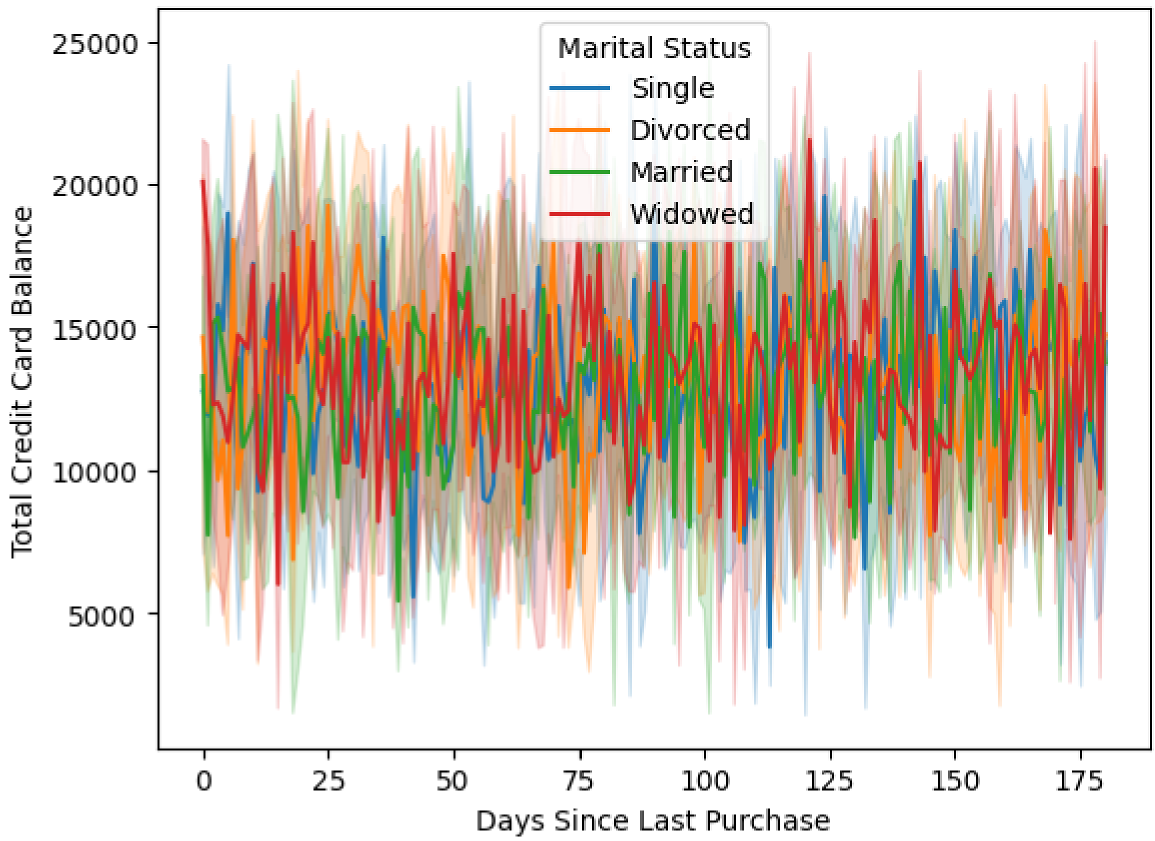
Home Value vs Marital Status

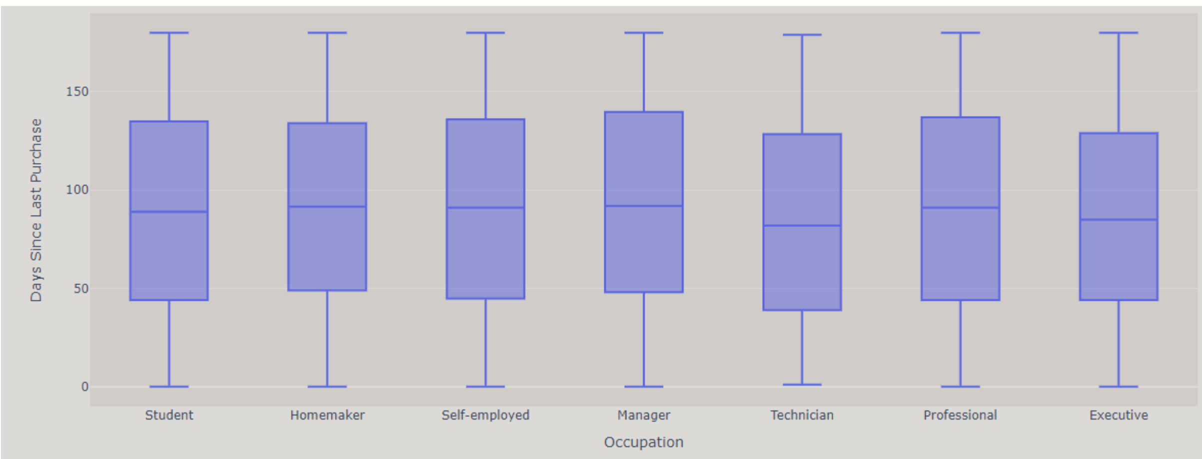
Home Value vs No. of Family Members



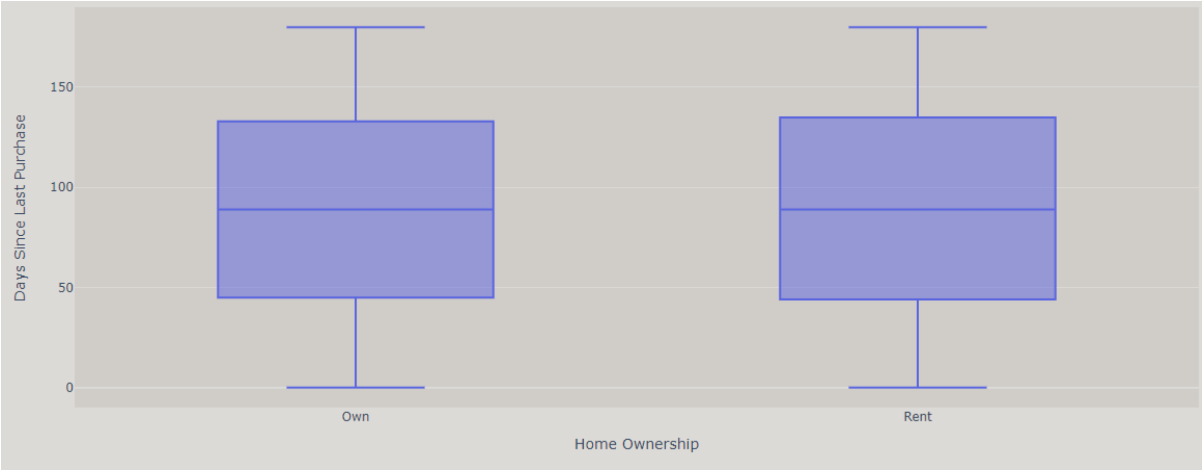
- Families with more people are more likely to take a loan especially when their income is low. Married people require it the most since a lot of money goes into handling the family.
- People with higher number of credit cards are less likely to take a loan since they can pay the balance of one credit card using another, but people with lesser number of credit card and a lower income are more likely to take a loan
- Married individuals having more than 3 members living in rented houses of a high value are most likely to take a loan especially if they are in a stable job such as Professional,Executive or Manager

Data Exploration : Number of Days Since Last Online Purchase

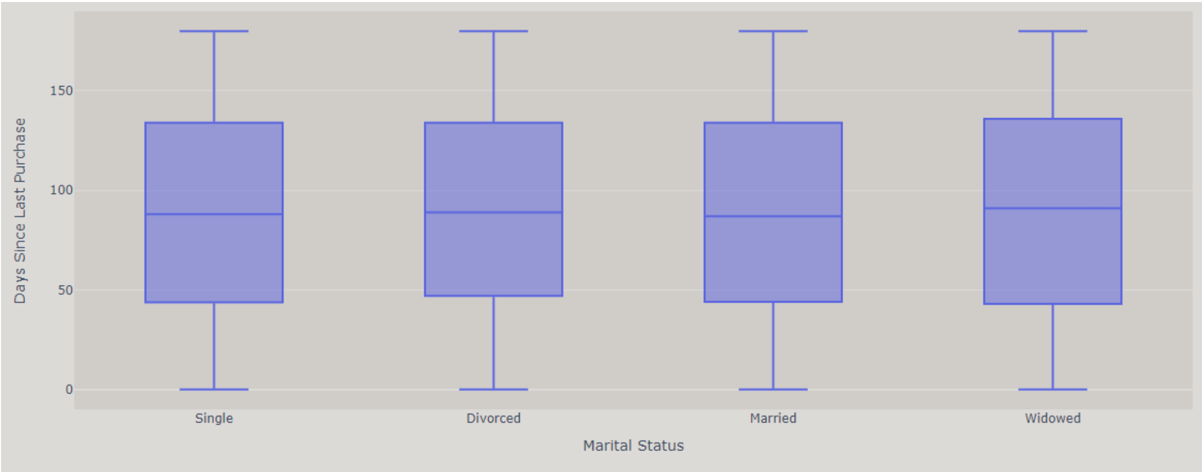




Occupation vs Days Since Last Purchase



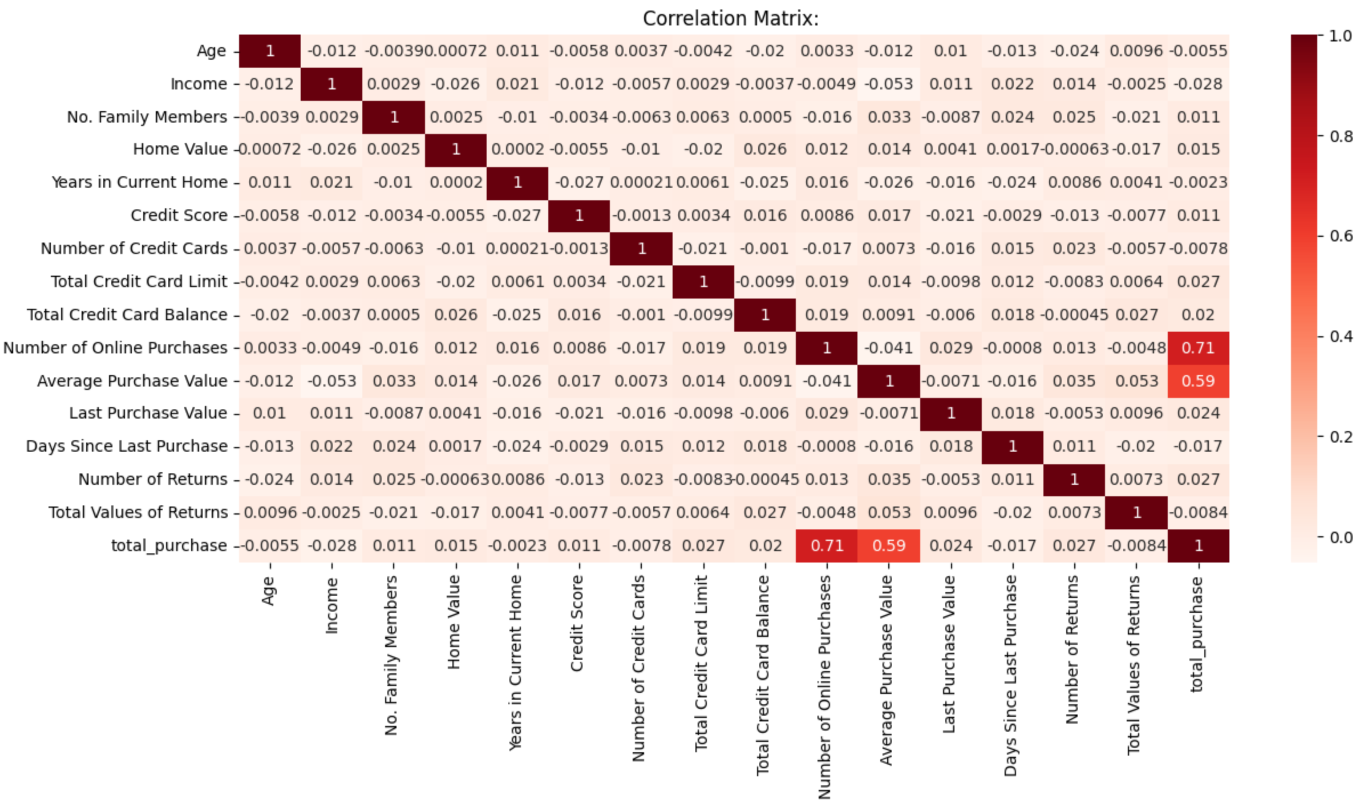
Home Ownership vs Days Since Last Purchase



Marital Status vs Days Since Last Purchase

- People not purchasing items since a long time indicates that they are currently facing financial problems which indicates they are potential candidates for personal loans
- Individuals who are married, live in a rented home, who have their credit card balance greater than their credit card limit and haven't purchased items for a period longer than 60 days are the best potential customers for a home loan

3. Correlation between features:



We plotted a correlation matrix to find out the correlation between the features. It indicated a weak correlation. Heatmap is used to depict the relative difference between the values. Just a few columns shows correlation with each other.
The column 'total_purchase' shows correlation with the columns 'Number of Online Purchases' and 'Average Purchase Value' since we used both these columns to form the 'total_purchase' column.

4. Feature Engineering:

New Features

Credit Utilization Ratio

```
[149] df['Credit_UR'] = df['Total Credit Card Balance']/df['Income']
```

Need For Money

```
[150] df['Need_Money'] = (df['Days Since Last Purchase']**2)*(df['Total Credit Card Balance'])
```

Financial Stability

```
[151] # Apply label encoding to 'education_level' (for demonstration purposes)
      df['Home_encode'] = df['Home Ownership'].astype('category').cat.codes

      # Multiply 'age' with the encoded education level
      df['finan_stable'] = (df['Home Value'] * df['Home_encode'])/df['Income']
```

Extra Amount to be Paid

```
[152] df['ex_to_pay'] = df['Total Credit Card Balance']/df['Income']
```

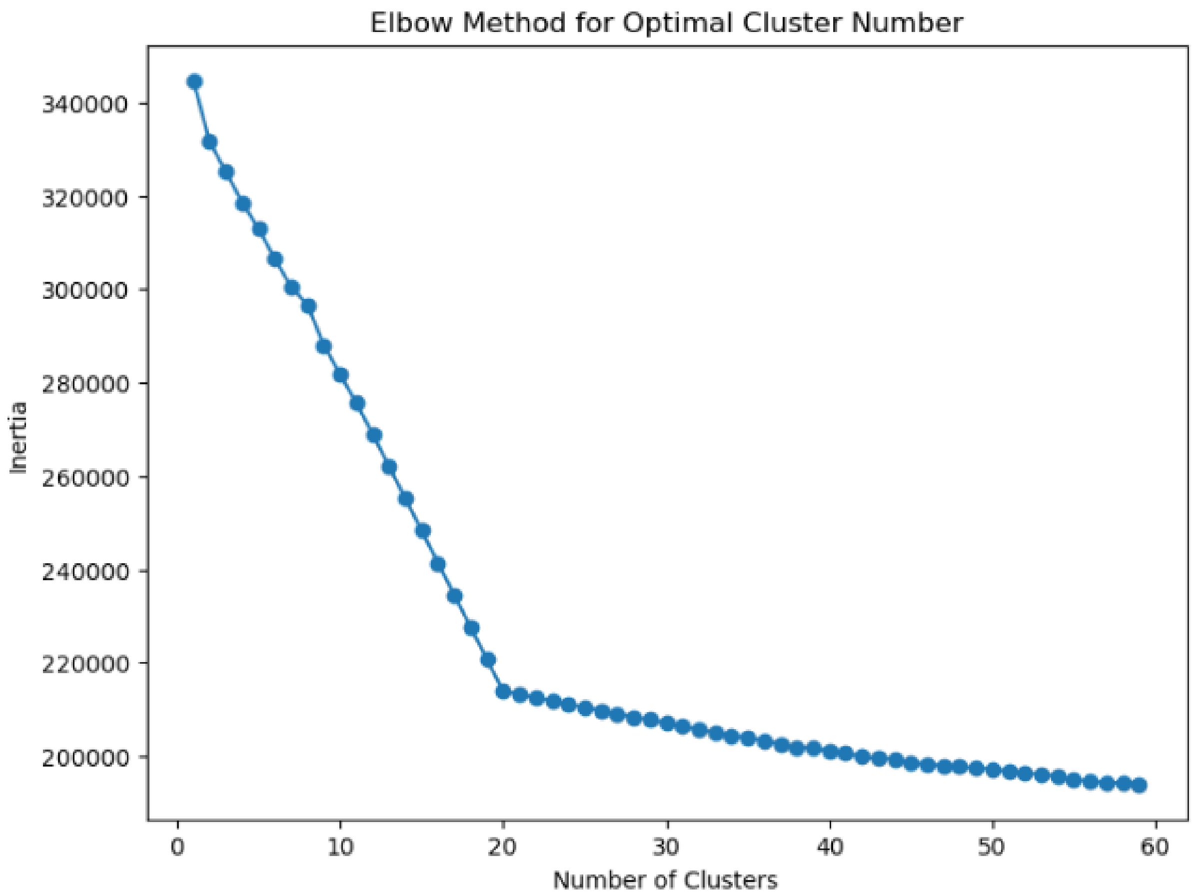
We created 4 new features apart from 'total_purchase' to better analyse the data. They are as follows:

Credit Utility Ratio = (Total Credit Card Balance / Income)
Need For Money = (Days Since Last Purchase^2 / Total Credit Card Balance)
Financial Instability = (Home Value * Home Ownership / Income)
Extra Amount to be paid = (Total Credit Card Balance / Income)

5. <u>Approach:</u>

K-Means Clustering -

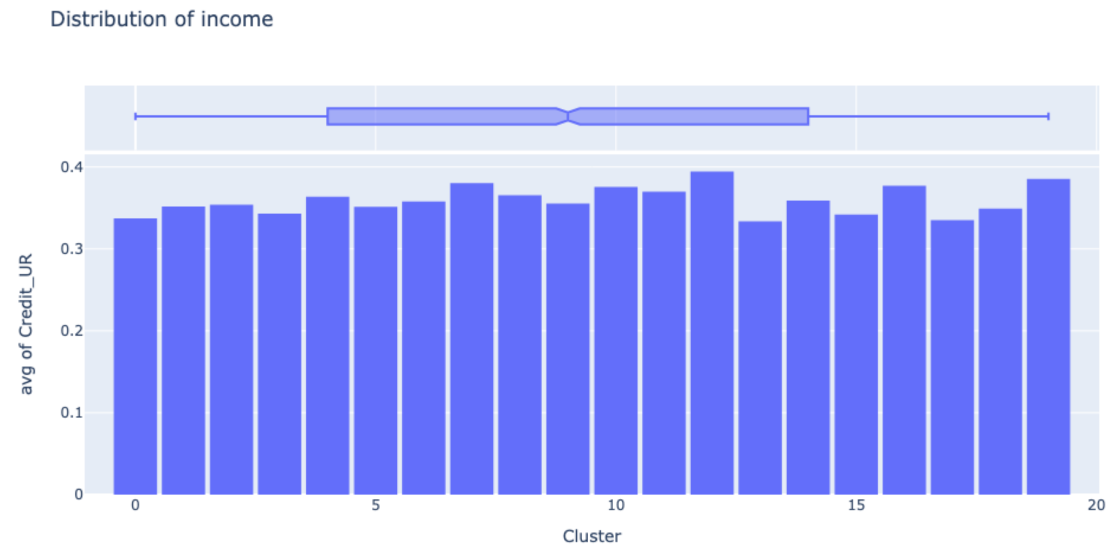
- Analysing all the data and gaining a thorough understanding of the potential customer under each segment of information it is best to create a cluster of potential customers
- A cluster would basically highlight the potential customers in a much better way by taking into considerations all the factors that affect an individual to take a home loan
- Finding the **correlation** among the data helps in removing excess data by realising that various data factors are actually correlated and some may be irrelevant to carry out the analysis
- After this **One-Hot Encoding** was done for each categorical columns namely-**Gender, Location, Occupation, Home Ownership and Marital Status**
- After carrying out this the data was scaled using **Standard Scaler** in order to get bounded values for better accuracy of results and efficient calculation
- By carrying the elbow graph approach it was found that **20 clusters** are possible for our data



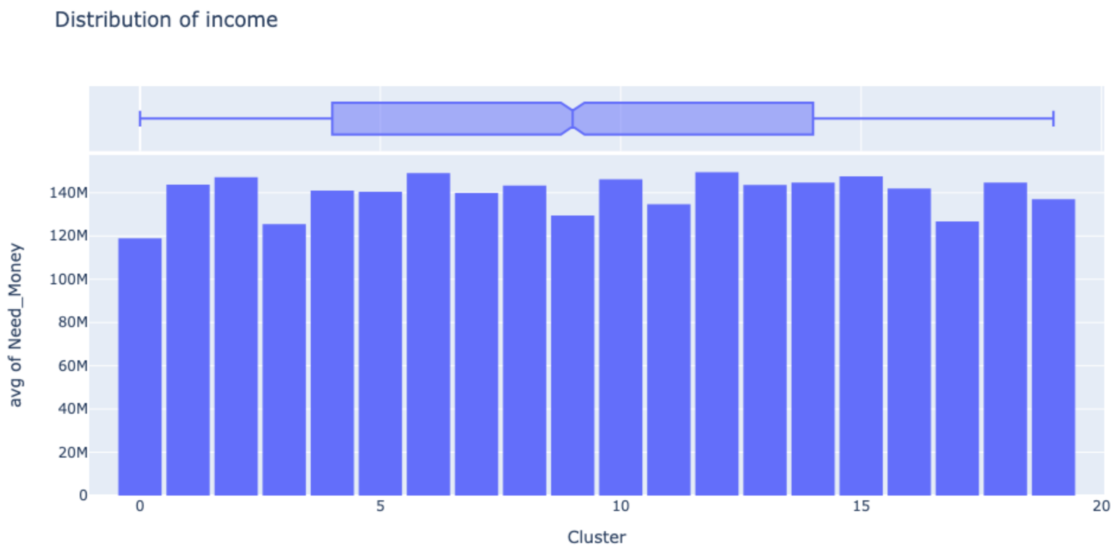
6. Figuring out the target clusters:

From the elbow method for optimal cluster numbers, we obtained the graph elbow at 20 clusters. So we formed 20 clusters.
To decide the target clusters i.e the clusters in which the people were willing to buy a loan, we mainly considered 3 features, “Credit Utility Ratio”, “Need for Money”, and “Financial Instability”. These features took into account all other pre-existing features.

In the feature “Credit Utility Ratio”,

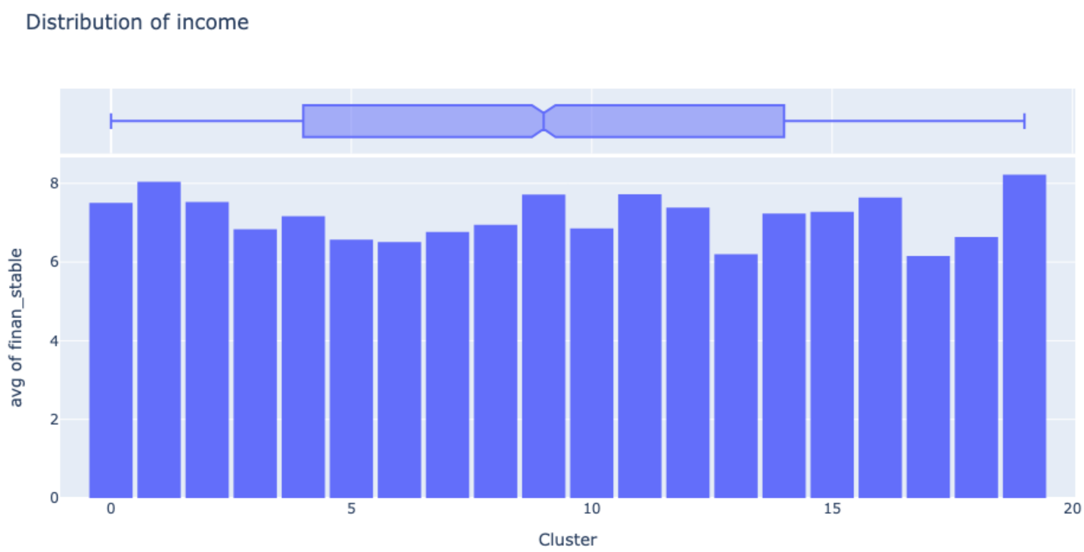


We took into account the clusters which had Credit Utility Ratio greater than 0.38 as there are three clusters exceeding this value, They are: 7, 9, 12
In the feature “Need for Money”:



We considered the clusters which had their values greater than 147M. There are four clusters exceeding this value. They are: 2, 6, 12, 15

In the feature “Financial Instability”:



We took into account the clusters which had Financial Instability values greater than 7.7 as there are four clusters exceeding this value. They are: 1, 9, 11, 19

So a total of 10 clusters can be said to have people with higher chances to buy a personal loan.

Recommendations and insights for the Retail Marketing Department to optimise their future campaigns and improve conversion rates.

Best possible way to optimise for future campaigns are:

1. Implementing a comprehensive system to meticulously monitor and analyse the spending patterns and financial behaviours of customers for a personalised and tailored approach.

2. To track the Credit Card Balance of customers, and focusing on those who are facing difficulties to clear the balance since these are the customers who are in need of extra money which can be provided by personal loans at a lower interest rate.

3. We enable ourselves to provide our clients with a range of services that precisely match their individual financial goals and capacities by exploring the particulars of credit profiles, which includes determining credit card limits and credit score evaluation. This builds a more robust and long-lasting financial relationship.

Customers who generally spend a high volume of money and have high Credit Card Balance will tend to opt for personal loans and banks can also offer them a personal loan by checking their Credit Scores.

Focusing on people who live in **rented homes**, have **high credit balance**, are married or have more family members and who get a stable payment such as **Professional Executives** and **Managers**.