

Credit EDA Assignment

Prepared by,
Saumalya Ghosh

Objective of the Case Study

- Find out appropriate client who are not having any payment difficulties in repaying loan. Financial Institutions like Banks/NBFCs have business opportunities if they will enroll these customers into their bucket list.
- Find out applicant who are facing payment difficulties at present or have history of becoming loan defaulter. Financial Institutions like Banks/NBFCs should not approve loan application request from these customers as it seems to be an indicator for potential business loss.
- Analyze data of all the applicants and offer/reject loan request based on merit.

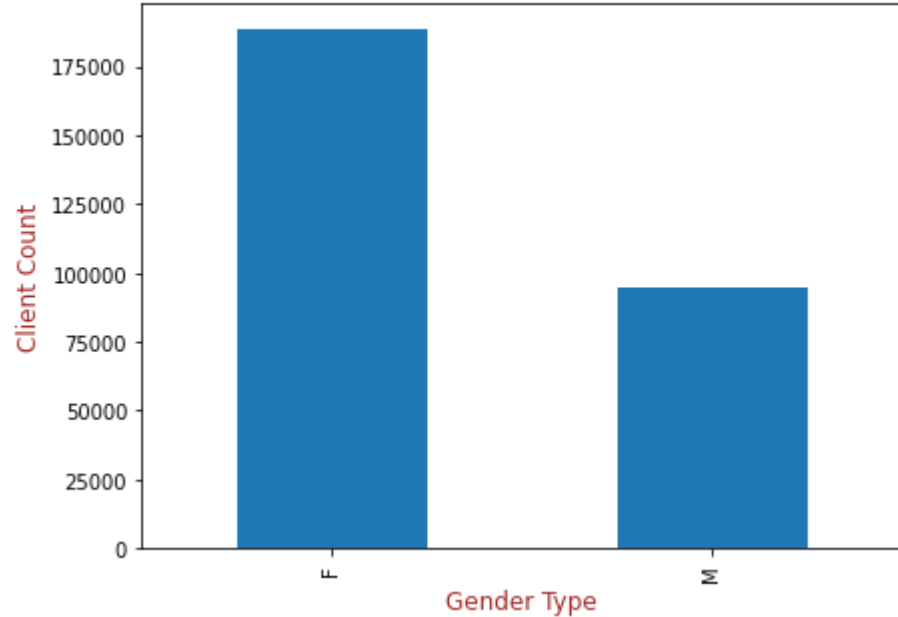
Steps of the Assignment

- Fair understanding of each column description.
- Import necessary libraries
- Load the data
- Basic sanity check of the data
- Check and find out columns consists of Missing Values more than 40%
- Clean/Remove columns having Missing Values more than 40%
- Find out Data Imbalance and Evaluate Ratio of Imbalance
- Assess Univariate and Bivariate Analysis of present and previous loan applications; Find out spread of non-defaulter and defaulter client.
- Find out 5 potential Outliers in the dataset supported by logical reasoning

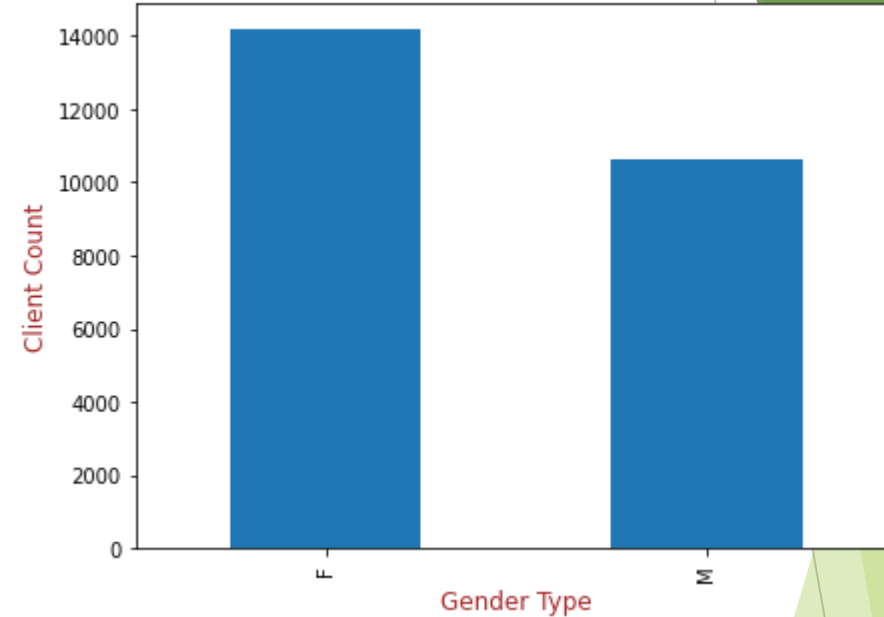
Univariate Analysis on Target Variable

Gender wise Distribution

Gender wise Distribution of Non-Defaulter Client

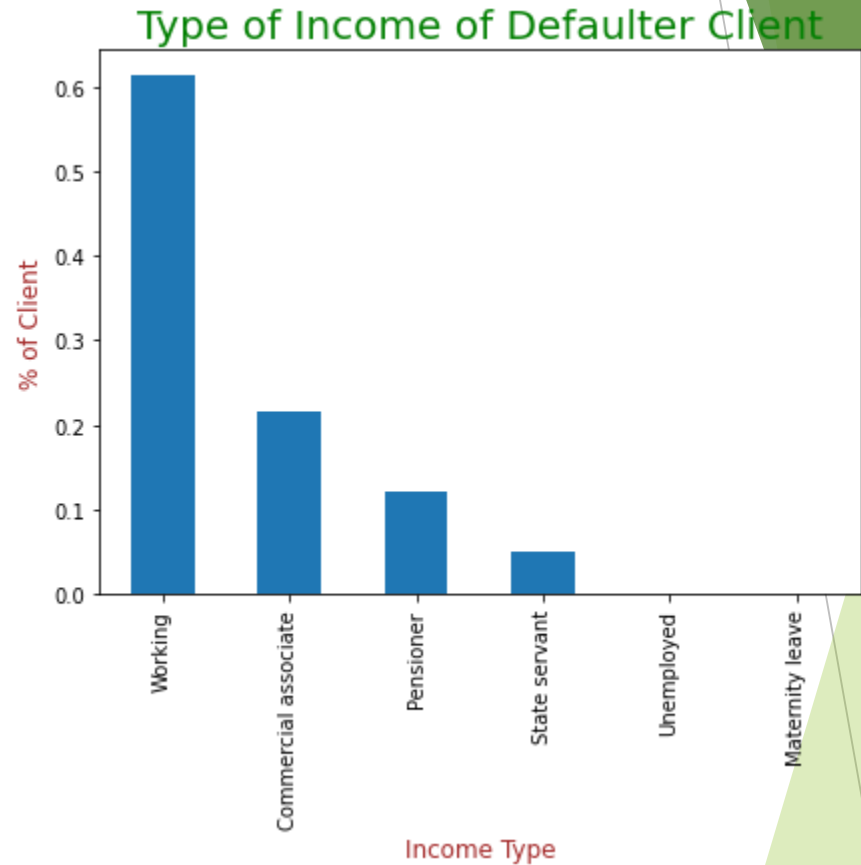
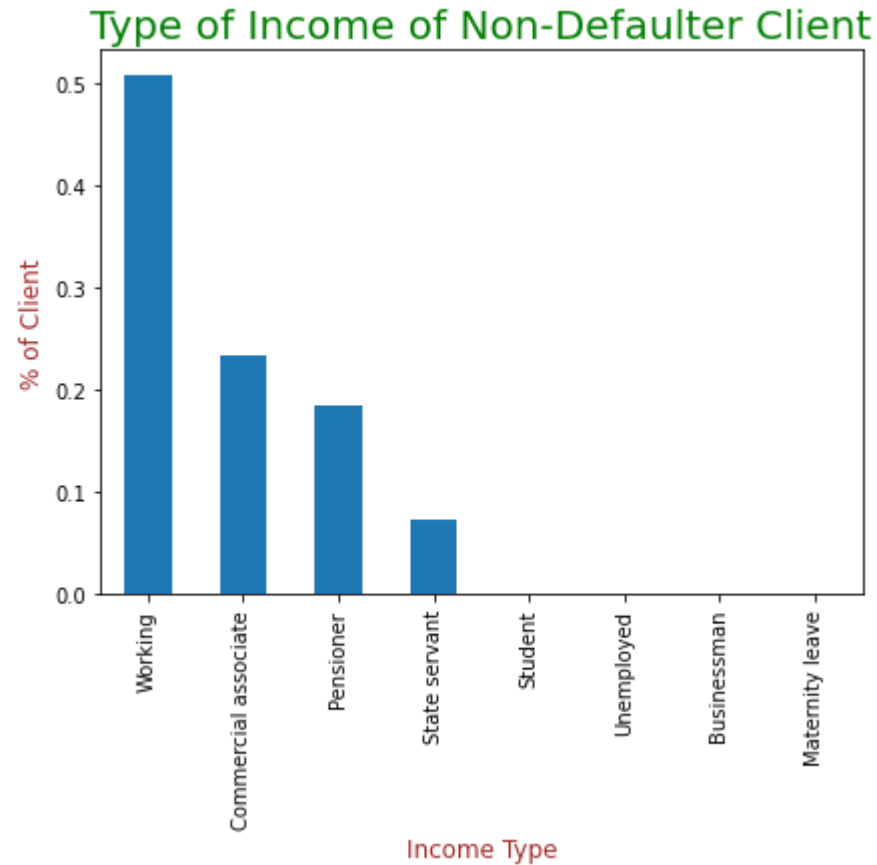


Gender wise Distribution of Defaulter Client



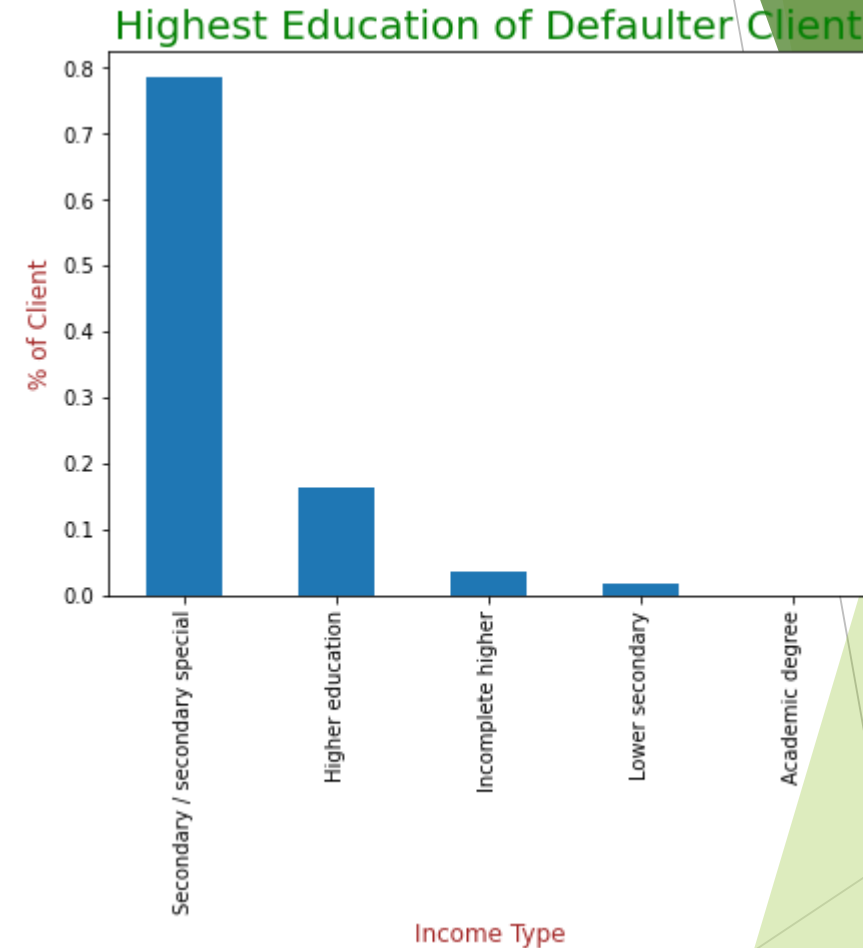
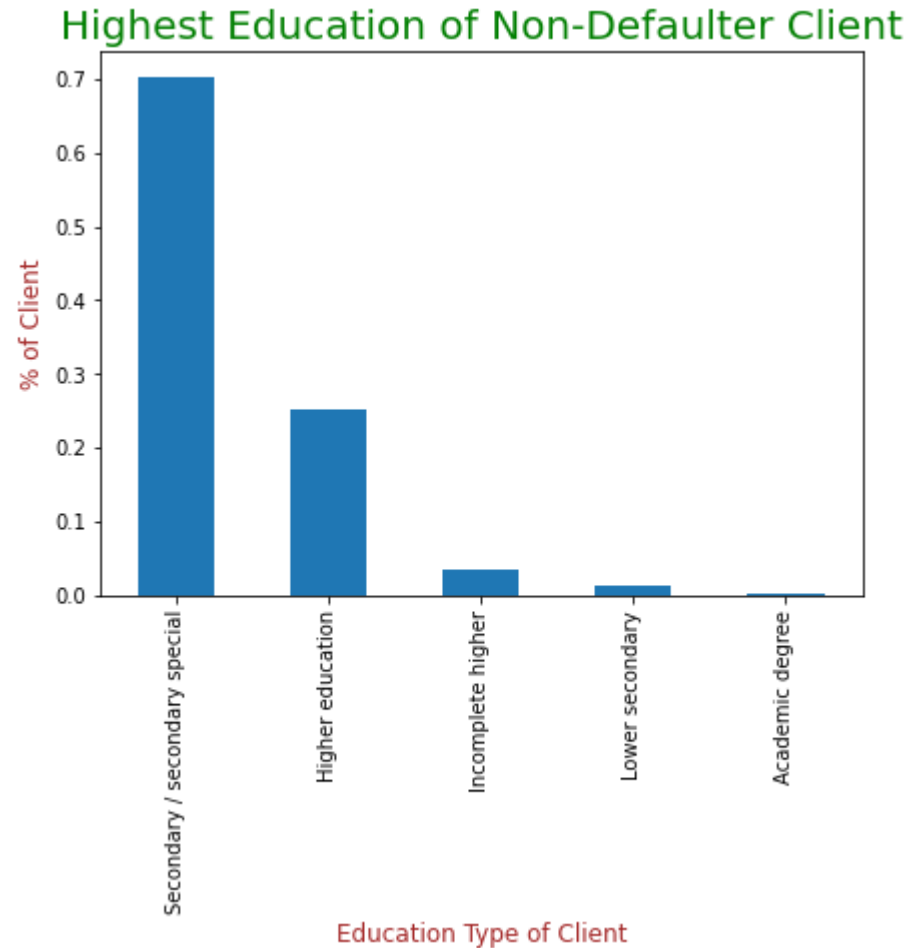
- Non-Defaulter Female Clients are much in number (More than 1.75 L) compared to Male Client (More than 0.75 L).
- On the other hand, Female Client having loan payment difficulties are also higher (Around 0.14 L) than Male Client (0.1 L).
- Financial Institutions should focus on Male applicant who are likely to repay the loan.

Client Income Type wise Distribution



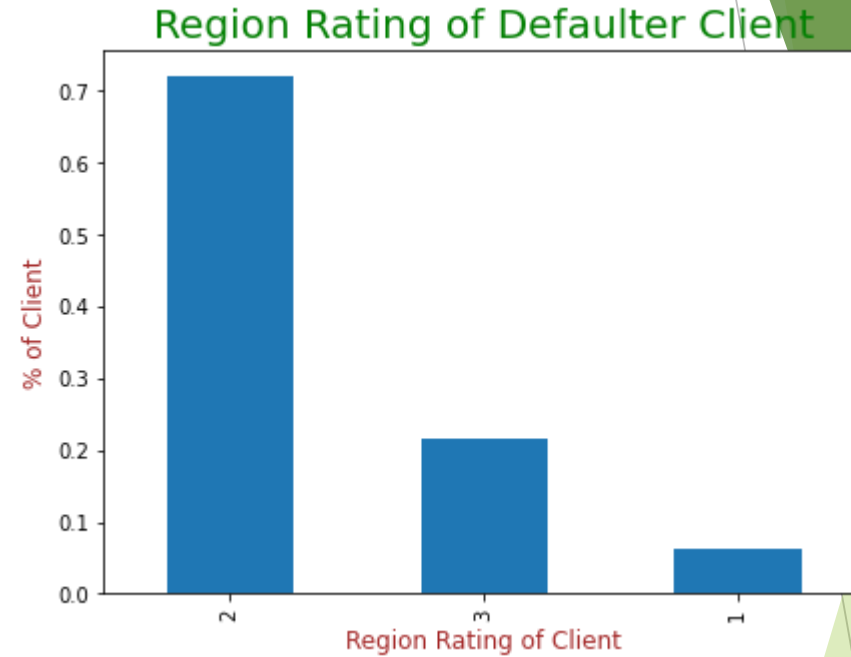
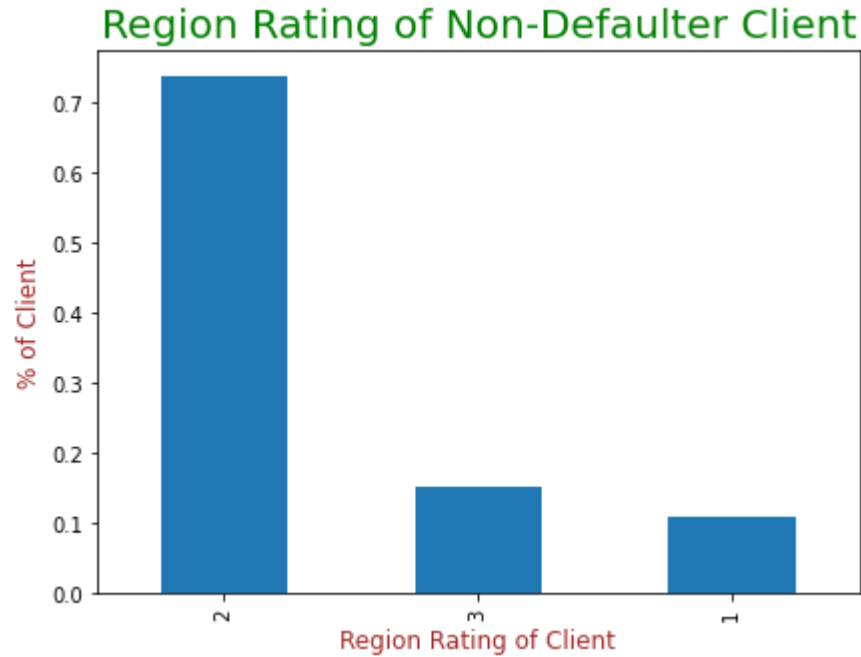
- Working professionals have the highest spread both in Non-Defaulter (~50%) and Defaulter category (~60%)
- Financial Institution should approve loan to Student and Businessman who are likely to repay the loan.

Client Education Type wise Distribution



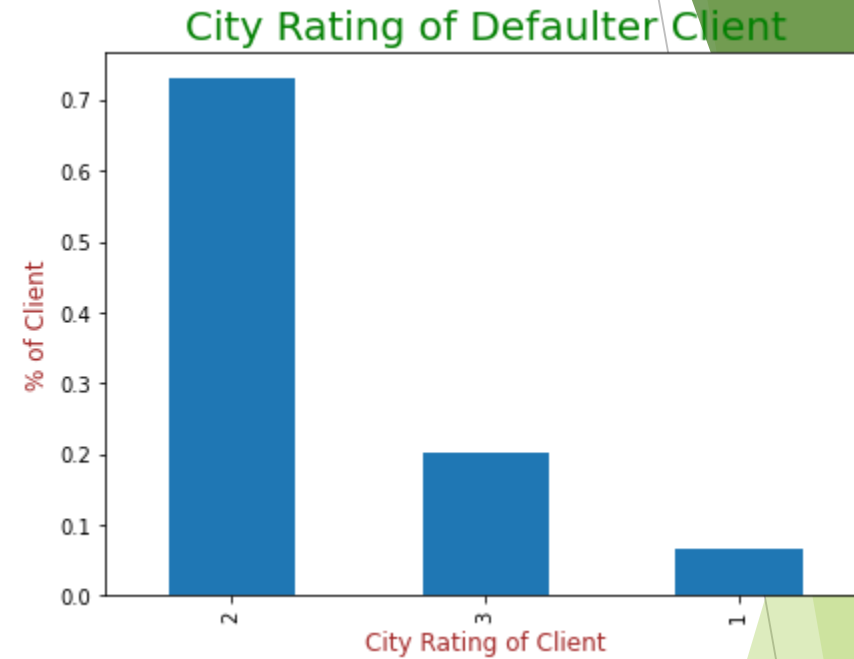
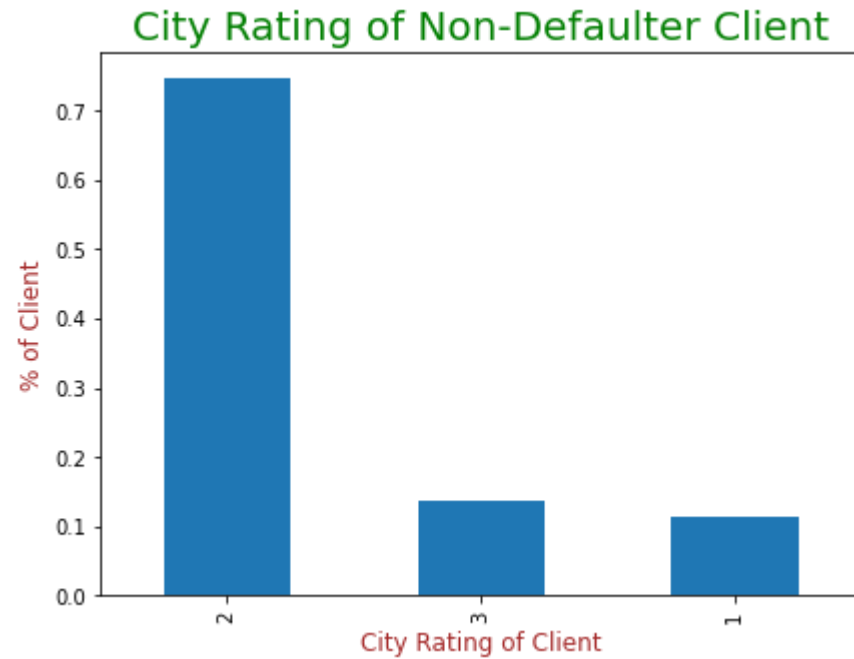
- Very few/Negligible % of Client having Academic degree are found to be loan Defaulter. Financial Institution should process loan to Client having Academic degree

Region wise Distribution



- We can observe from the above Bar chart that, client live in Region 1 have a very less % (Less than 5%) to become a loan defaulter. Hence, it is suggested to provide loan to client living in Region 1.

City wise Distribution

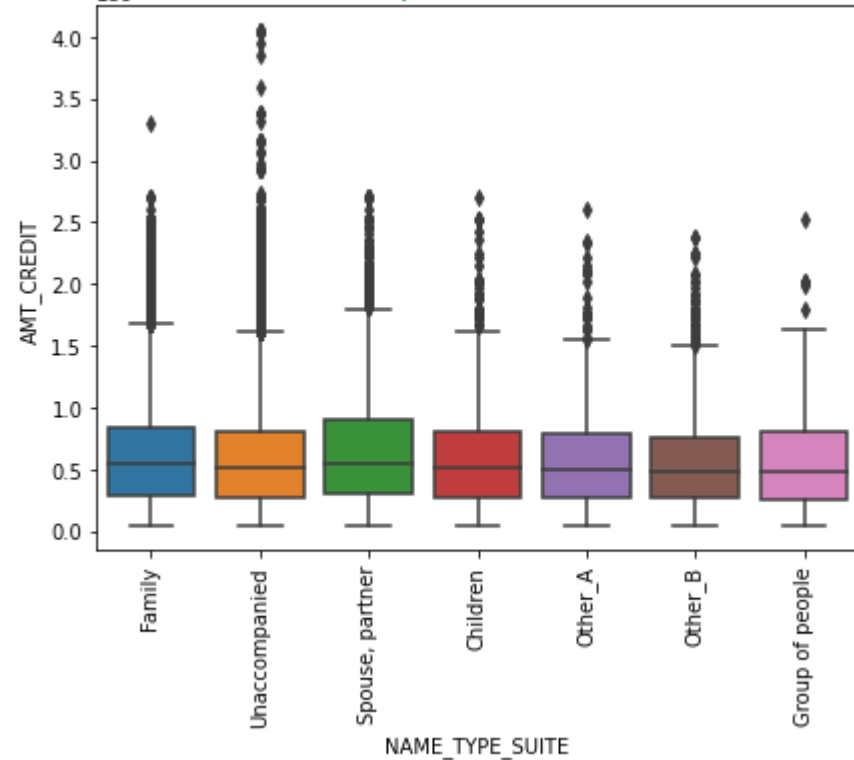


- It has noticed that, client live in City 3 have a tendency to become loan defaulter ($\sim 20\%$) than to repay loan ($\sim 10\%$). Financial Institutions should avoid to sanction loan to these category clients.
- On the other hand, spread of client lives in City 1 have a very less % (Less than 5%) to become a loan defaulter. Hence, loan applications from these clients can be approved on priority.

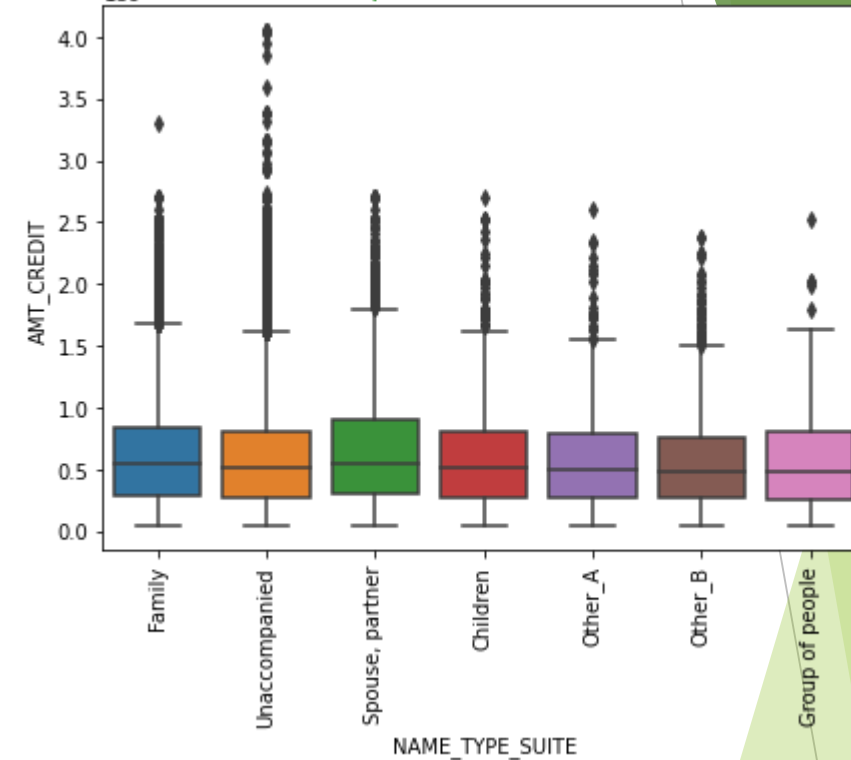
Bivariate Analysis on Target Variable

Client Accompany Status Vs. Loan Credit Amount

Non-Defaulter Client Accompanied Status Vs. Loan Credit Amount

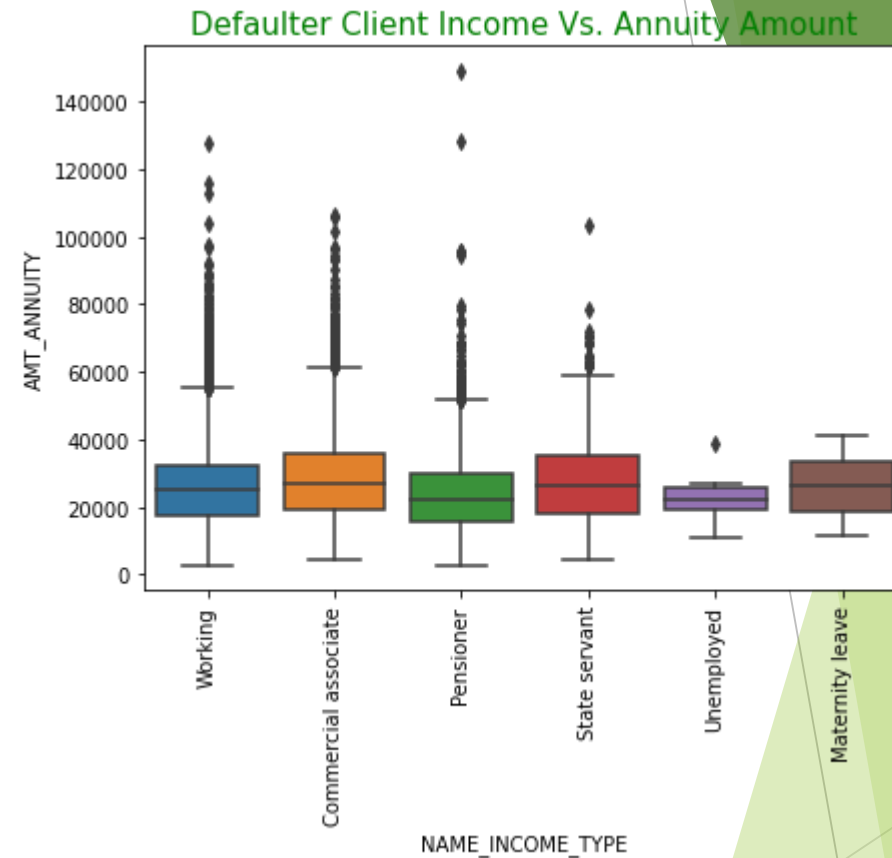
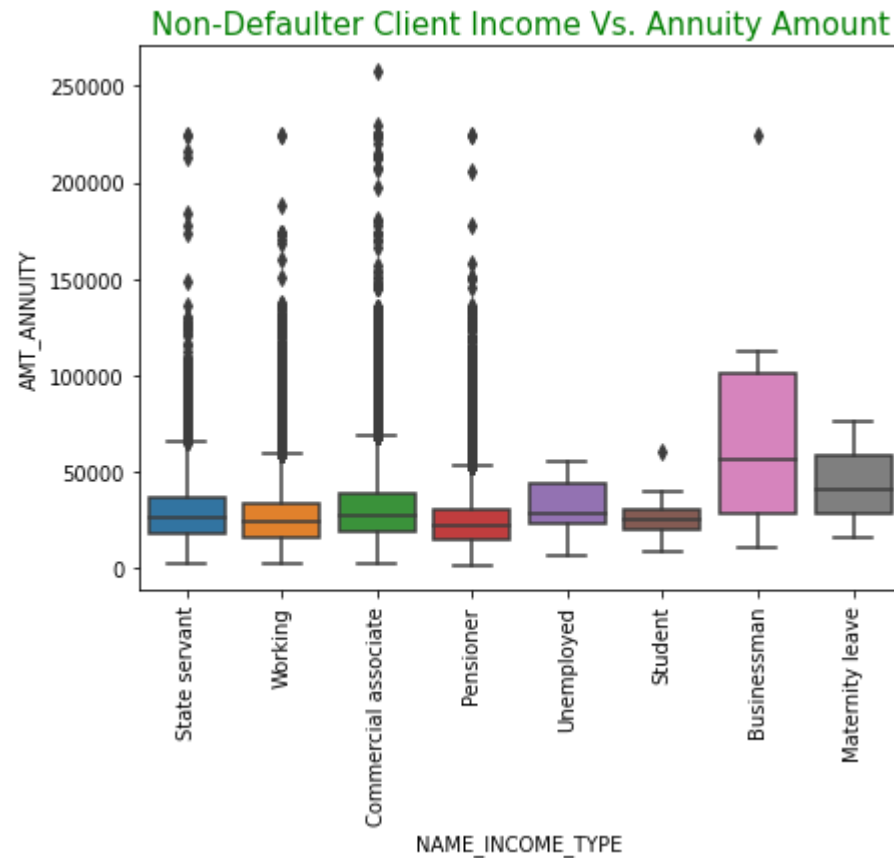


Defaulter Client Accompanied Status Vs. Loan Credit Amount



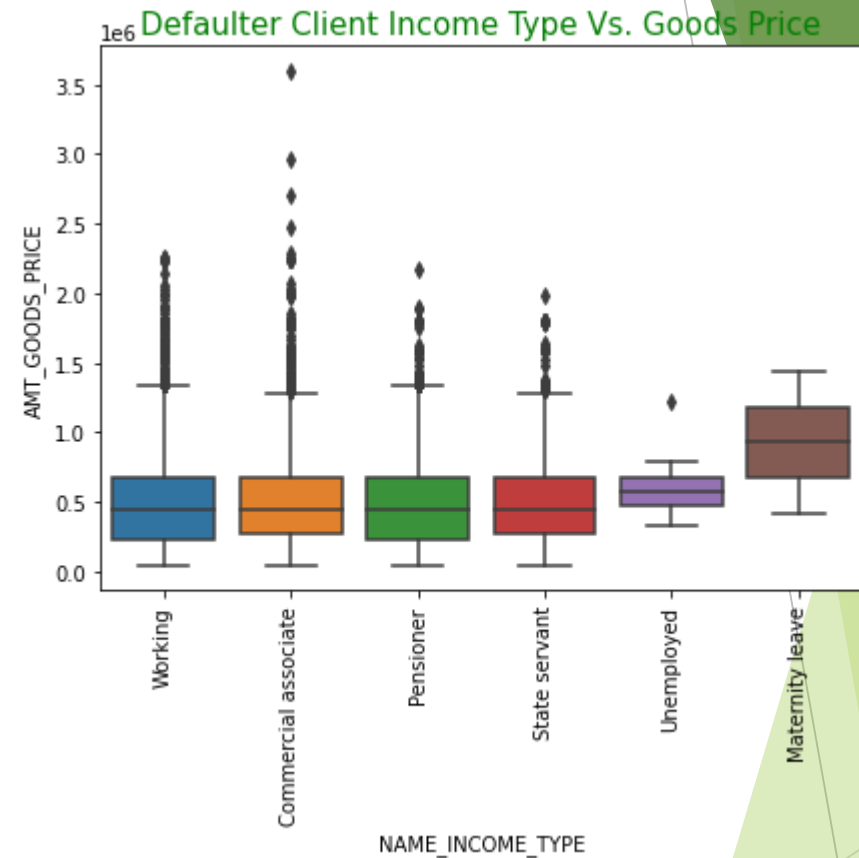
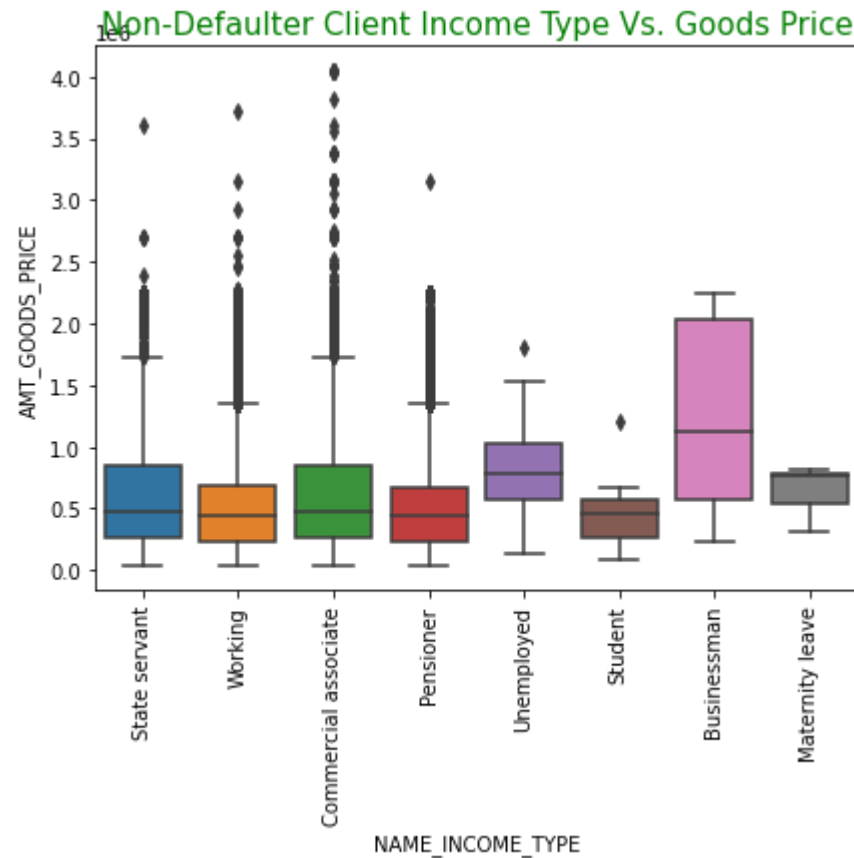
- Non-defaulter Client accompanied by Spouse/Partner have the highest median value and third quartile i.e. upper bound is also higher. Financial Institution should give preference to client accompanied by Spouse/Partner during loan approval.
- Other_B and Group of people falls under Defaulter category have low median value which means they have accessed low credit amount from loan still became loan defaulter.

Client Income Type Vs. Loan EMI Amount



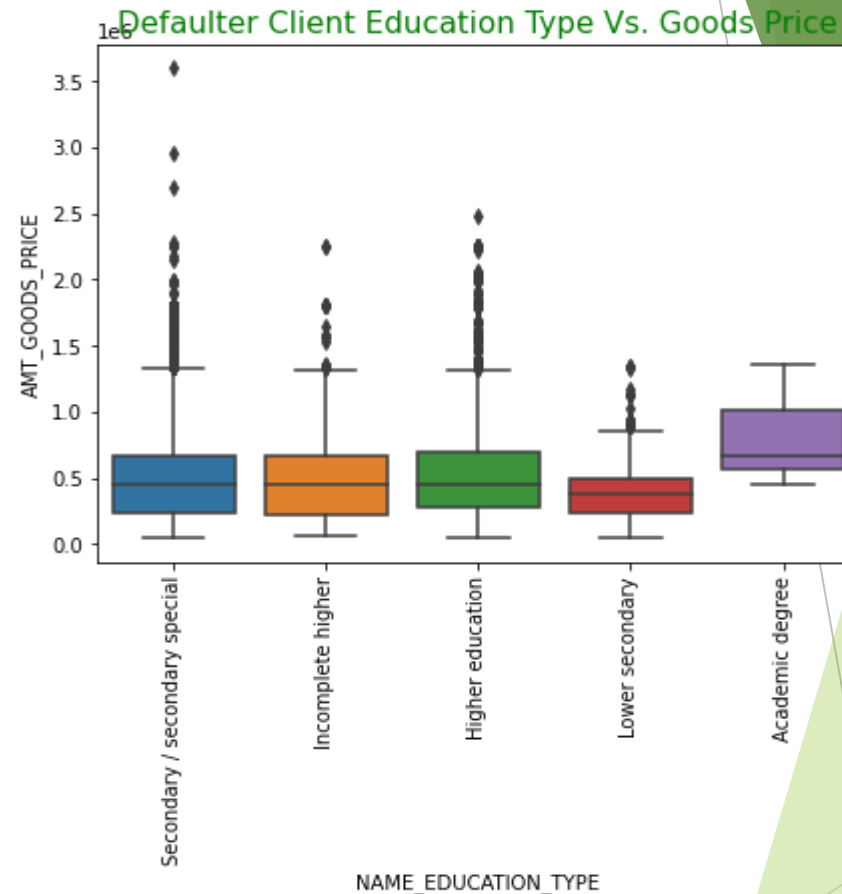
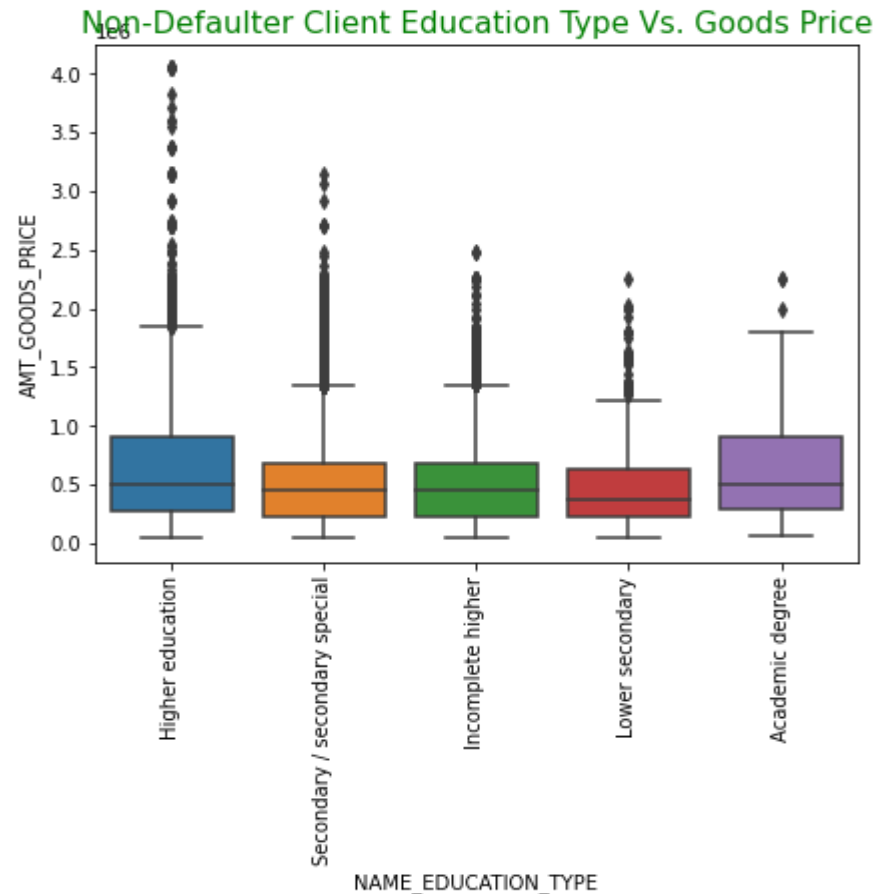
- Loan Annuity (EMI) amount of Businessman found to be the highest and none of them have faced in loan payment difficulties. Banks/NBFCs should give preference to Businessman during loan approval.
- Financial Institutions commercial associate and state servant judiciously as they are most likely to become loan defaulter.

Client Income Type Vs. Goods Price



- Businessman has been given highest amount of loan and they never missed to repay the loan. Financial Institution can focus more on adding businessman customer into their bucket list.
- On the contrary, customer falls under maternity leave category have given a huge amount of loan and they become a loan defaulter. FI should restrict in sanctioning to loan to such customers.

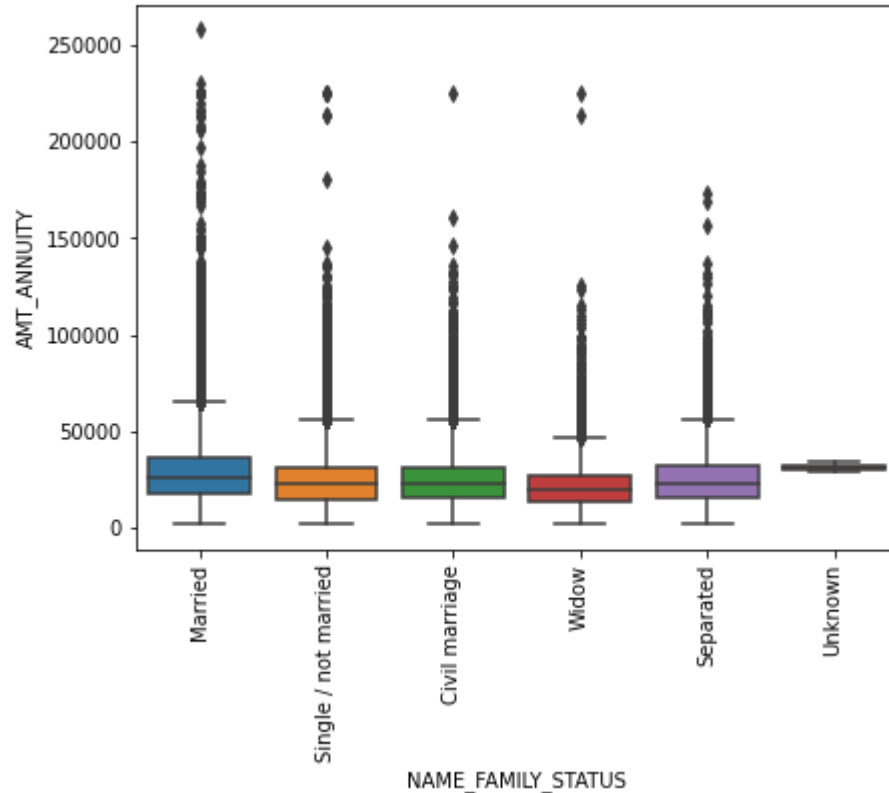
Client Education Type Vs. Goods Price



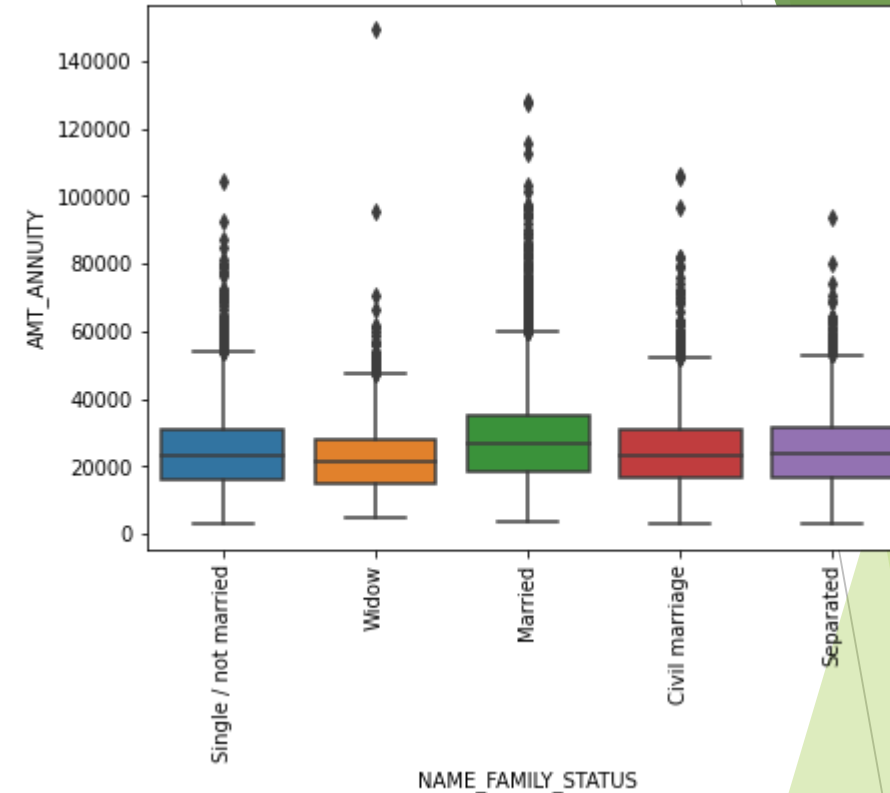
- Client having Higher education and Academic degree have given considerably high loan amount whereas client having academic degree are more likely to be a defaulter. Hence Financial Institutions should prioritize client with Higher education over Academic Degree when considering a high goods price.
- On the other hand, client with Lower secondary education have taken lowest amount of loan.

Client Family Status Vs. Annuity Amount

Non-Defaulter Client Family Status Vs. Annuity Amount

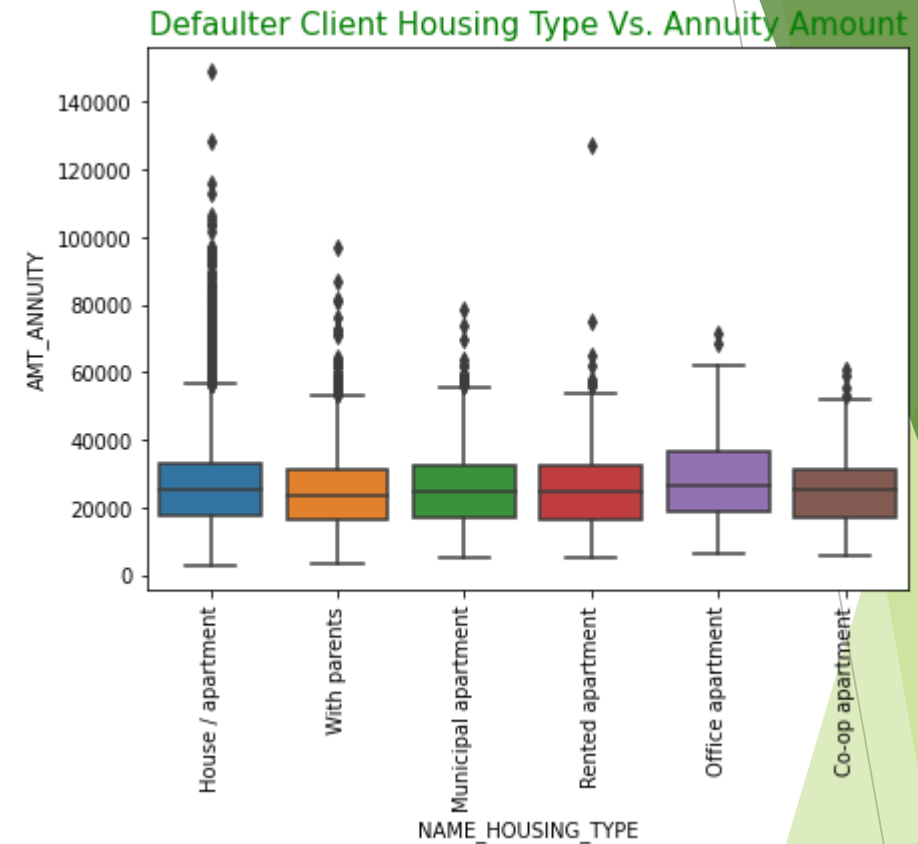
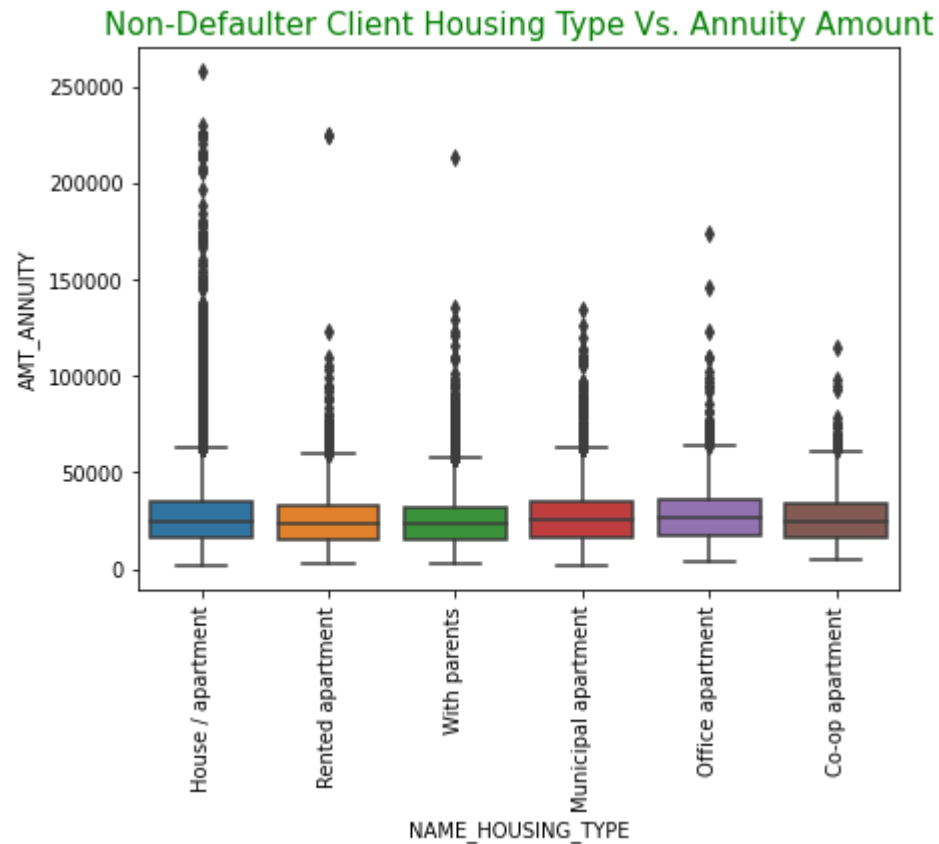


Defaulter Client Family Status Vs. Annuity Amount



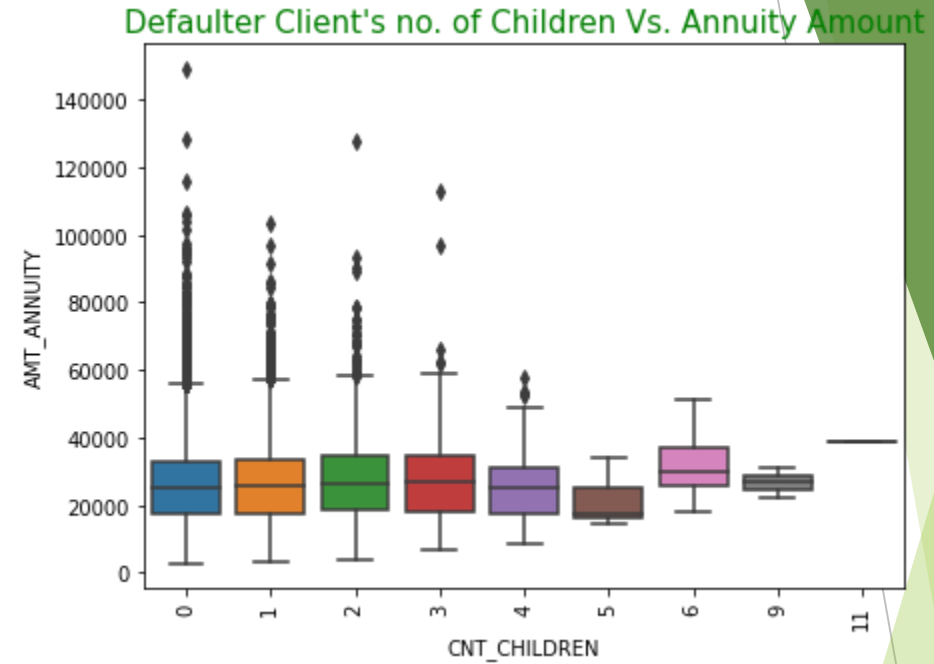
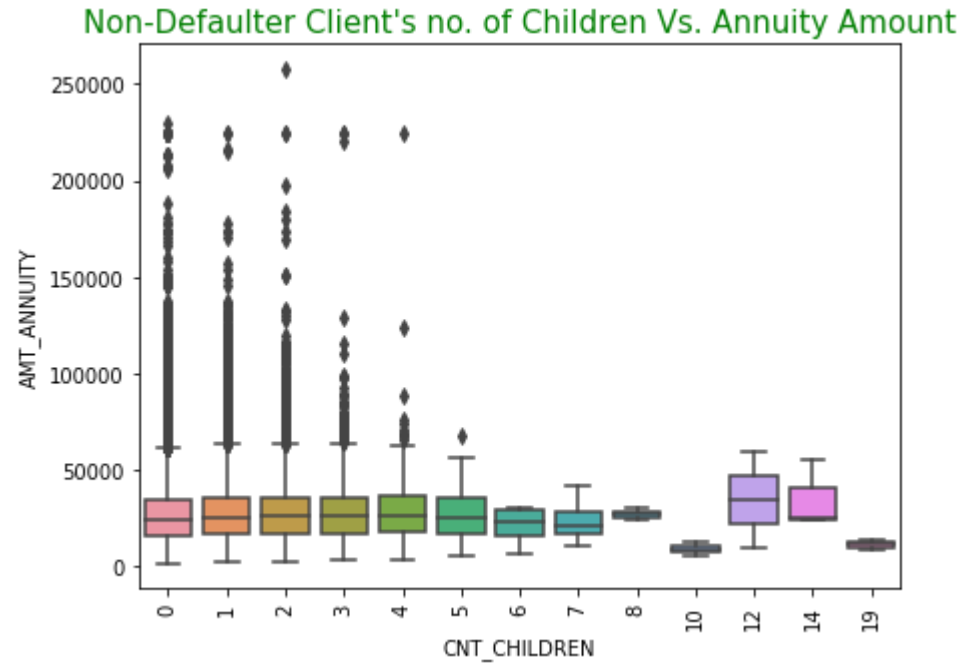
- Both Married and Separated Clients are paying high EMI amount whereas Married client are more often facing payment difficulties. On the other hand, Window clients are paying lower EMI amount but consistent in repaying loan. Hence, it is recommended to approve loan to Separated and Widow Clients predominantly.
- There is one client category 'Unknown' who have not declared their marital status, have never missed to repay the EMI amount.

Client Housing Type Vs. Annuity Amount



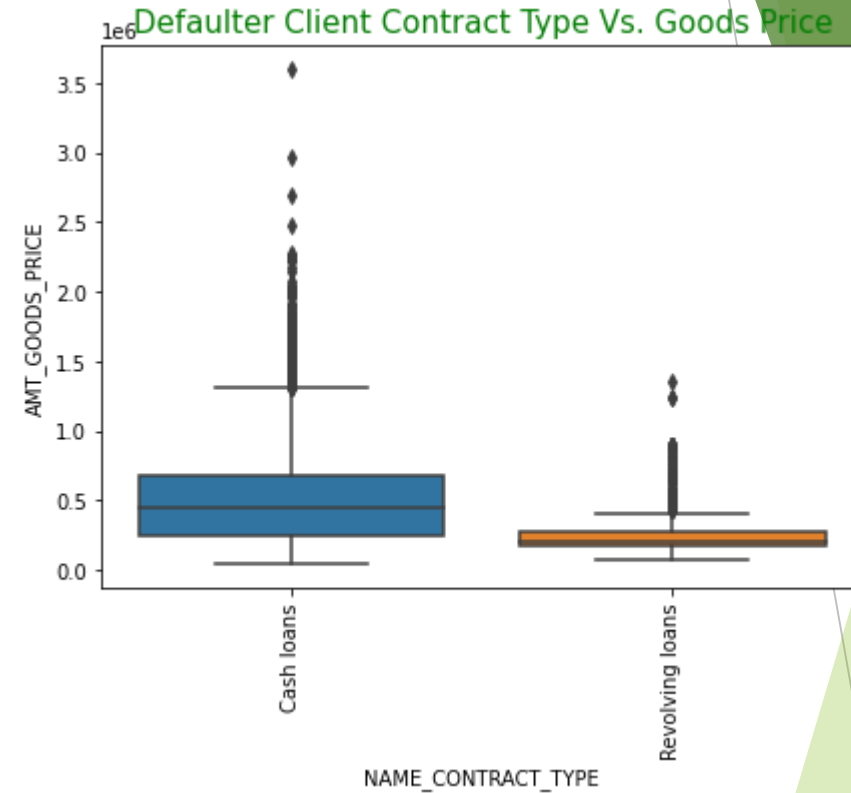
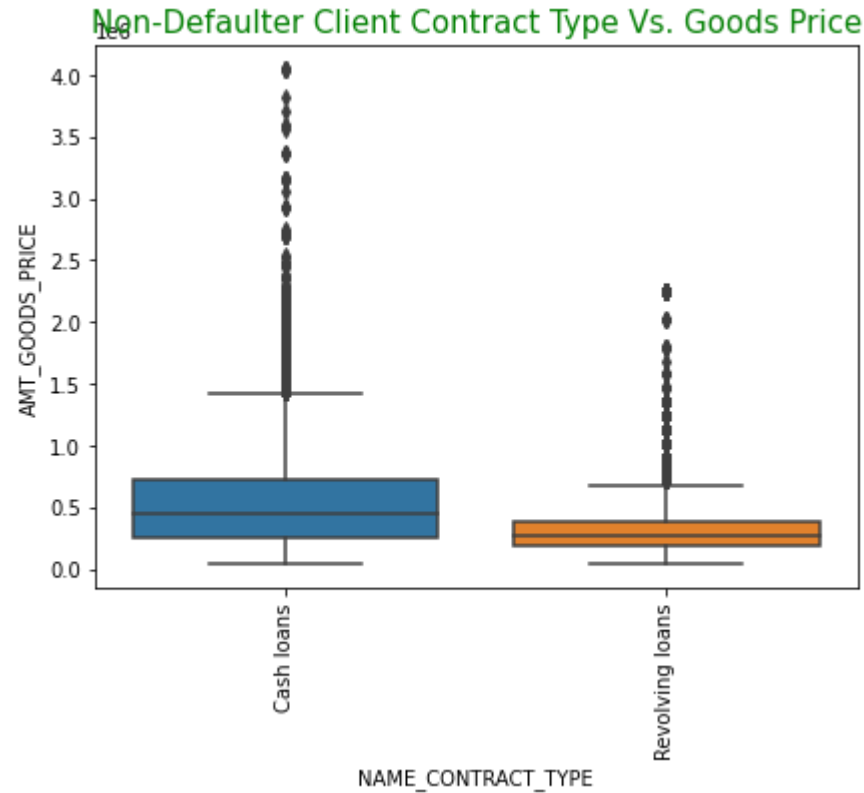
- Clients having own House/apartment and Office apartment are paying highest EMI amount whereas clients having Office apartment are most likely to become loan defaulter. We can observe, a lot of outliers who owned House/apartment as they are paying a very high amount of EMI.
- Preference should be given to clients having House/apartment at the time of approving loan.

Client Children Count Vs. Annuity Amount



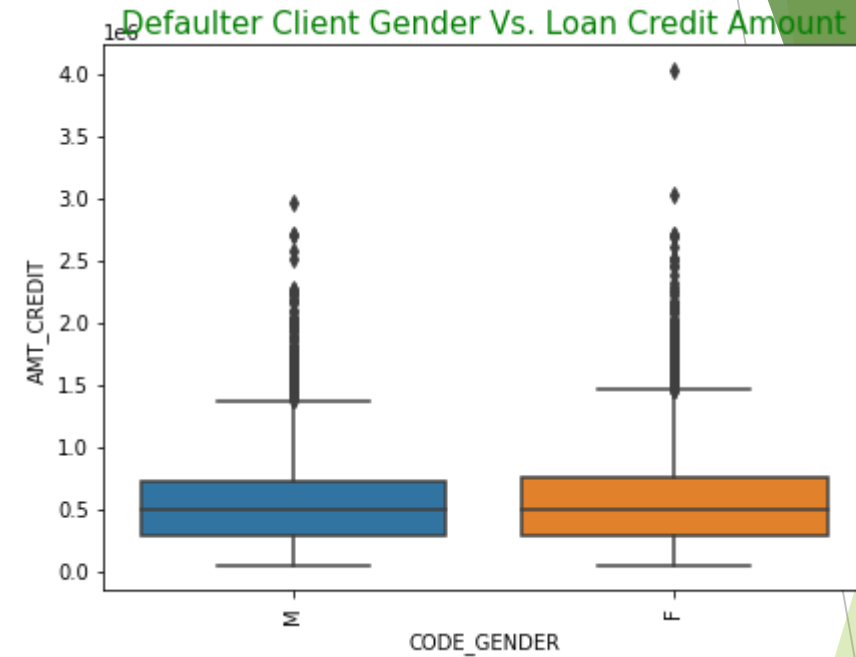
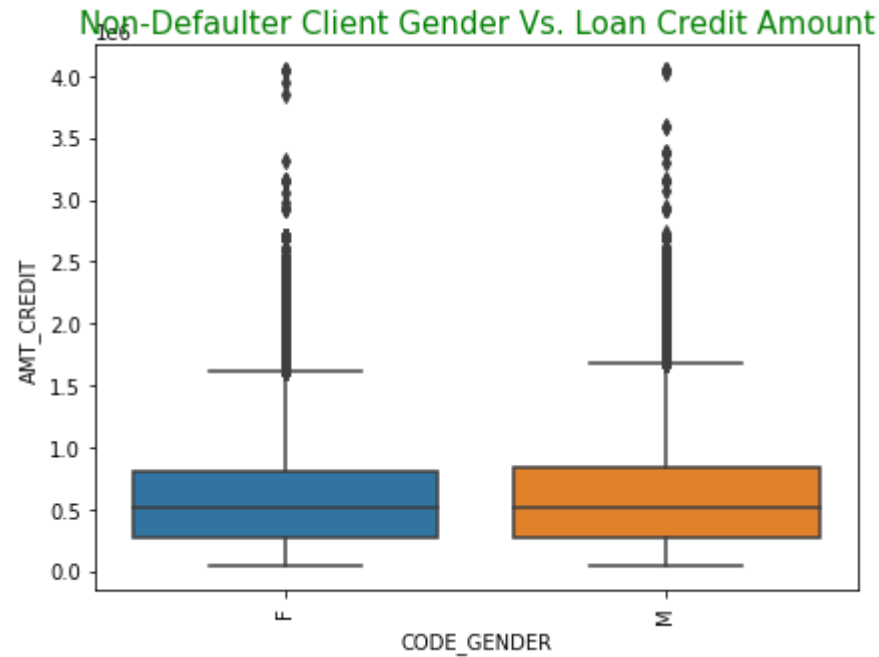
- It has noticed that, Clients having 2 to 4 nos. of children are paying highest EMI amount and maintaining uniformity in repaying loan. Clients having children no. 12 and 14 are also paying very high EMI amount but they are very few in numbers.
- On the other hand, clients having no. of children more than 5 are paying higher EMI amount and not consistent in repaying loan.
- Clients having 2 to 4 children are found to be the most consistent non-defaulter.

Client Contract Type Vs. Goods Price



- It has been observed that, Goods price for Cash loan is always higher than Revolving loan.
- As the Goods price is less for Revolving loan, most of the clients taken Revolving loan are able to repay the loan amount.
- Hence, Financial Institution should assess the risk in issuing Cash loan of higher amount and can restrict approving Cash loans.

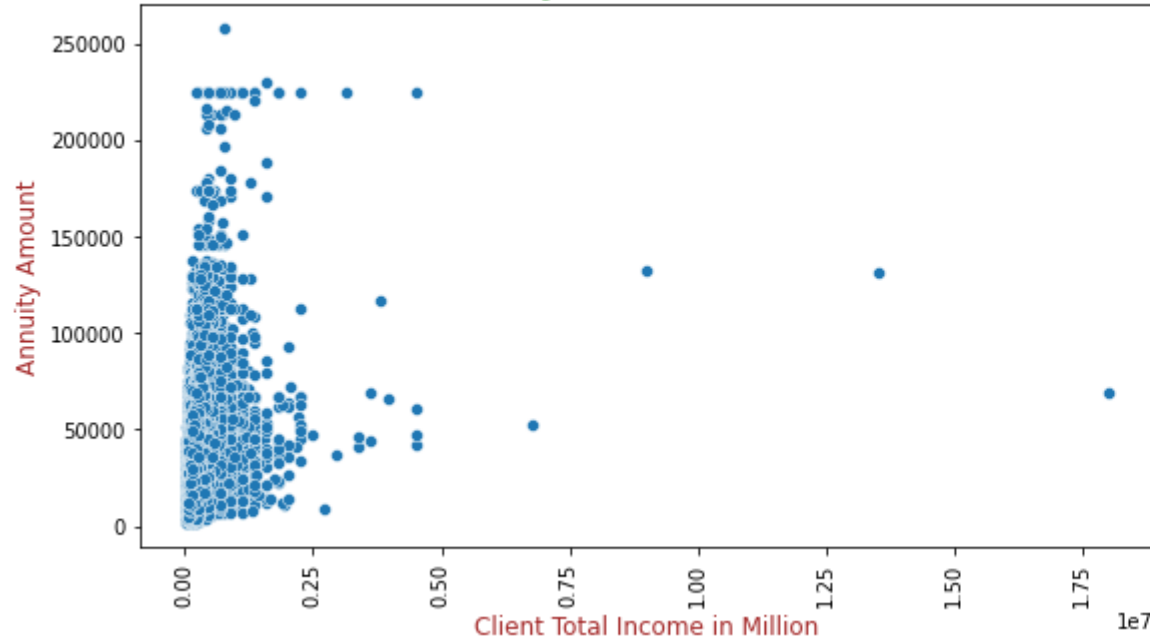
Client Gender Type Vs. Loan Credit Amount



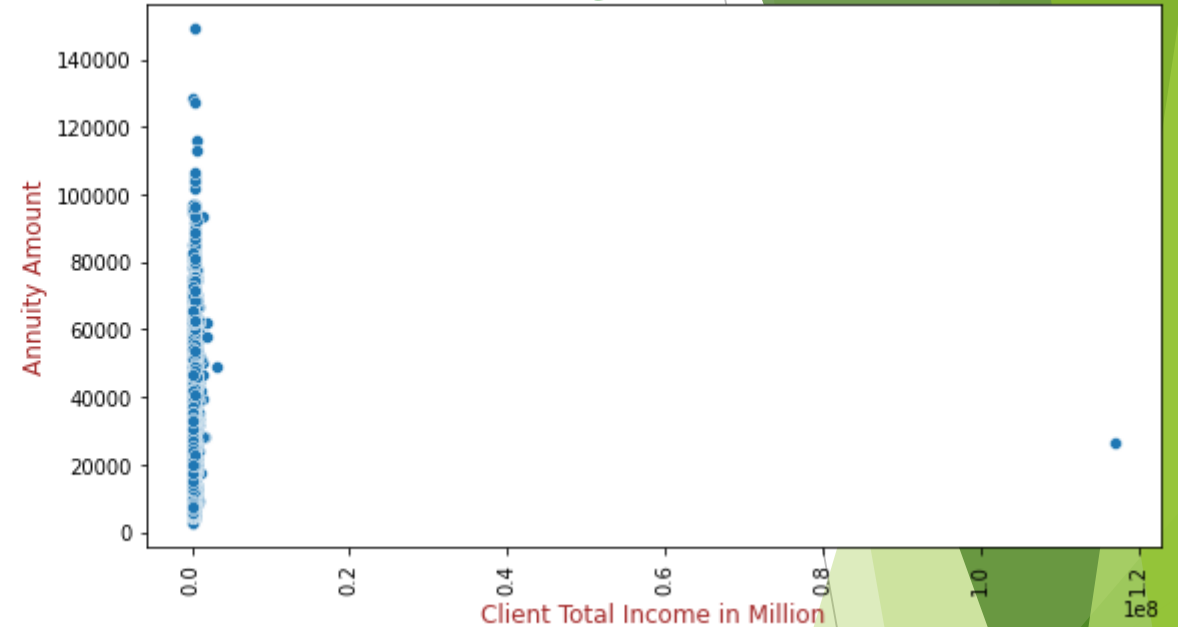
- It has been observed from above Boxplot that, Male clients are eligible to access more credit amount from the money loaned and found to be non-defaulter of loan.
- On the contrary, Female clients who are eligible to access more credit amount from the money loaned are likely to be loan defaulter.
- Hence, Financial Institutions should be given preference to Male clients to access loan credit amount.

Client Total Income Vs. Annuity Amount

Total Income Vs. Annuity Amount for Non-Defaulter Client

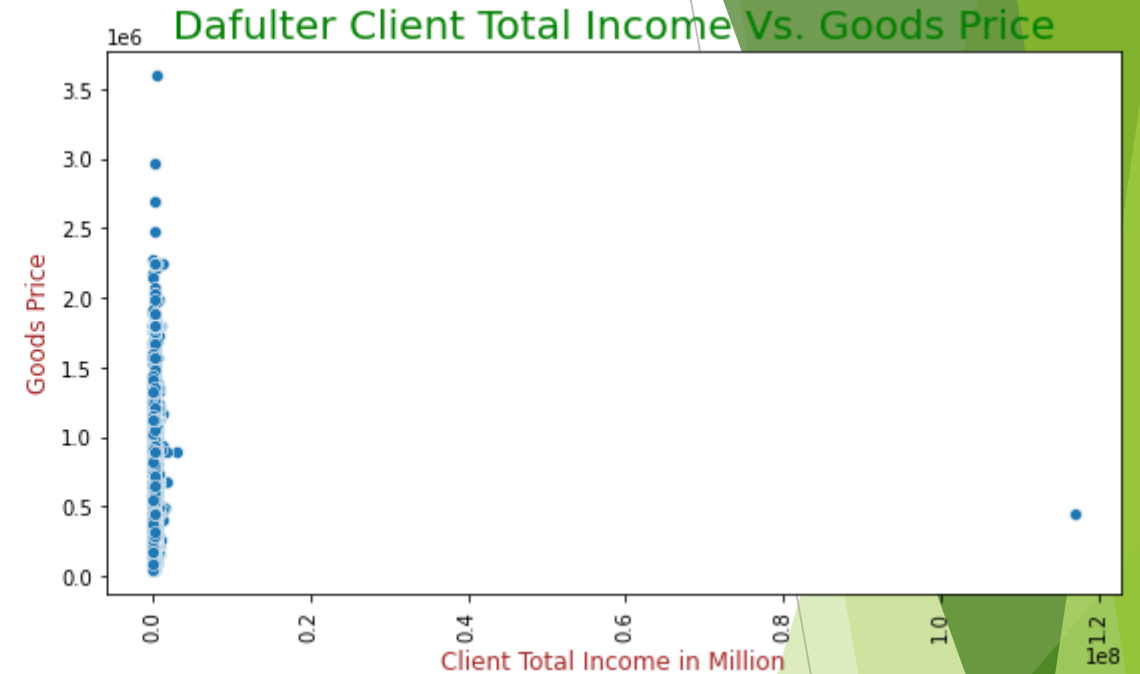
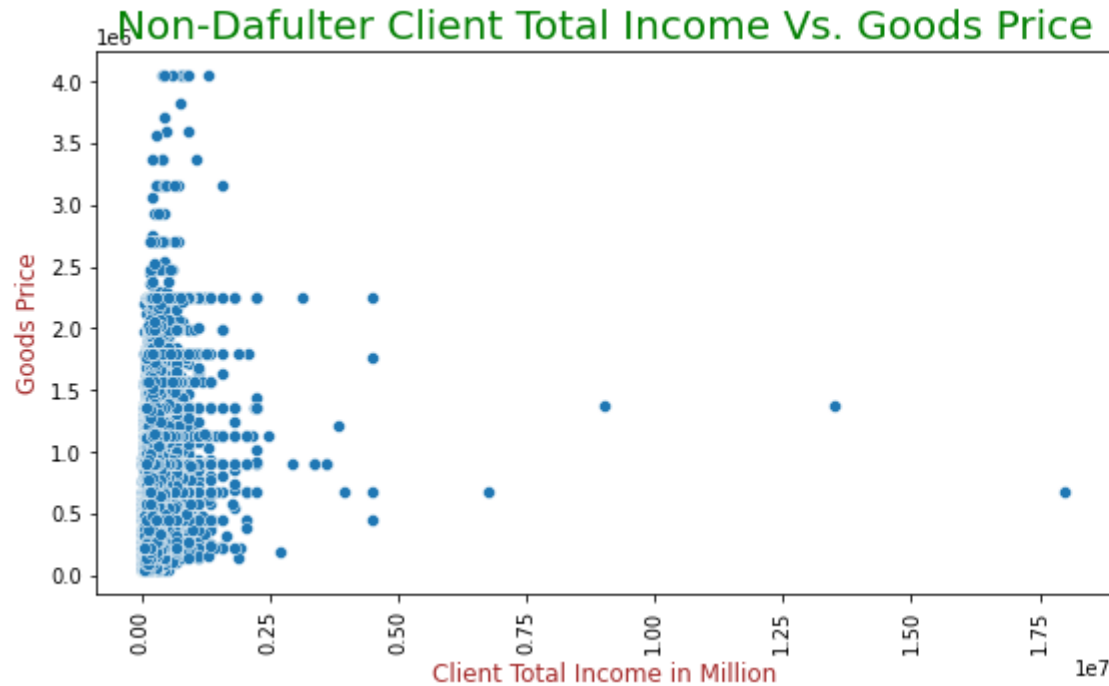


Total Income Vs. Annuity Amount for Defaulter Client



- It has been noticed that, most of the non-defaulter clients falls under income range upto 0.25 Million are paying the highest annuity amount.
- On the other side, clients having income range very less (Within 0.02 Million) are supposed to pay a very high EMI amount and becoming defaulter.
- Hence, it is advisable to approve loan to clients having higher total income.

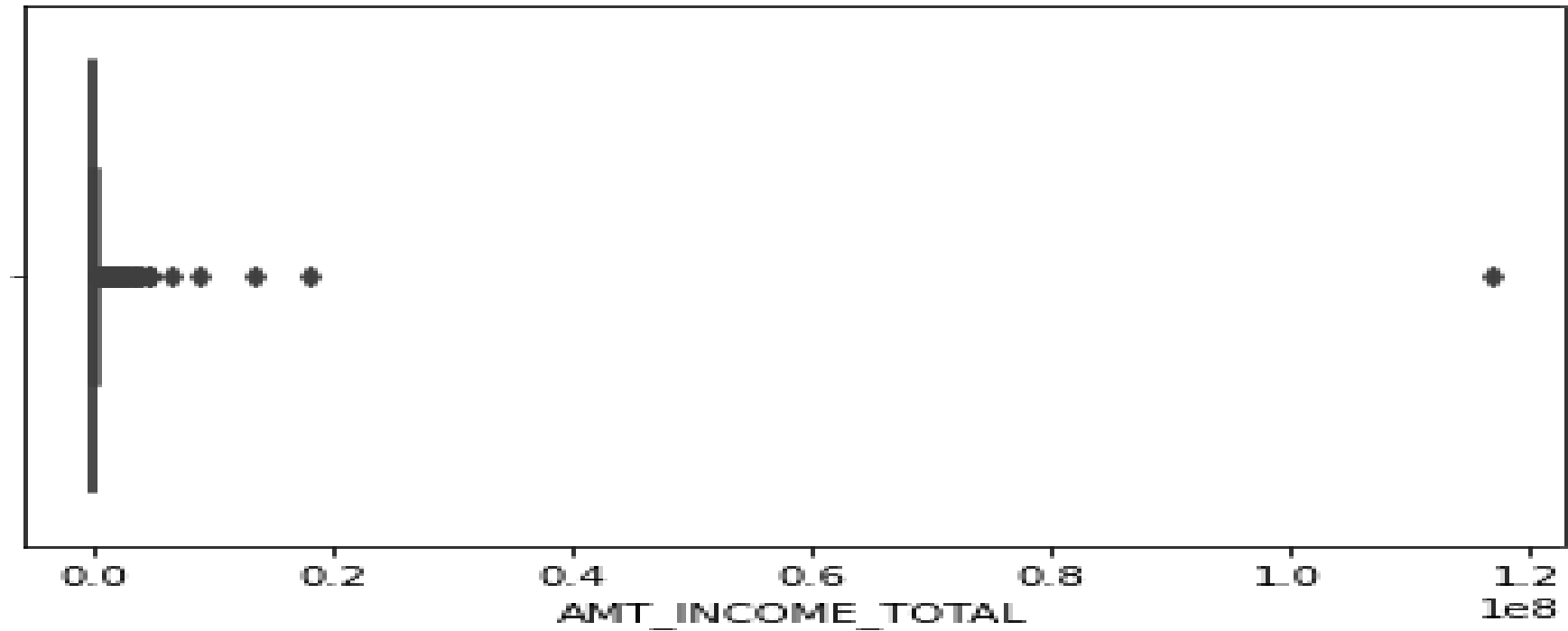
Client Total Income Vs. Goods Price



- It has been noticed that, most of the non-defaulter clients falls under income range upto 0.25 Million have taken higher amount of loan.
- On the other side, clients having income range very less (Within 0.02 Million) have received a higher amount of loan and subsequently became loan defaulter.
- Financial Institution should approve a higher loan amount to clients earning high.

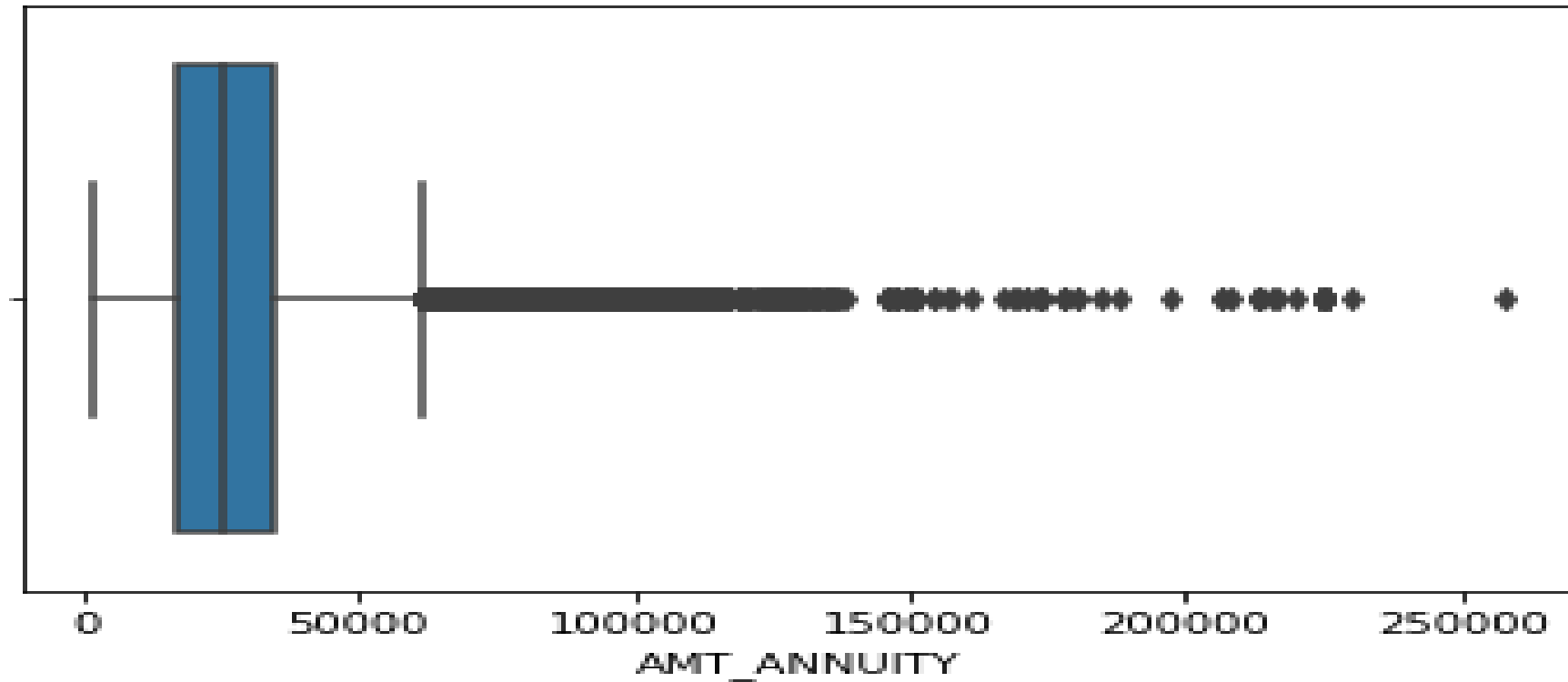
Outliers in Dataframe

Total Income Amount (AMT_INCOME_TOTAL)



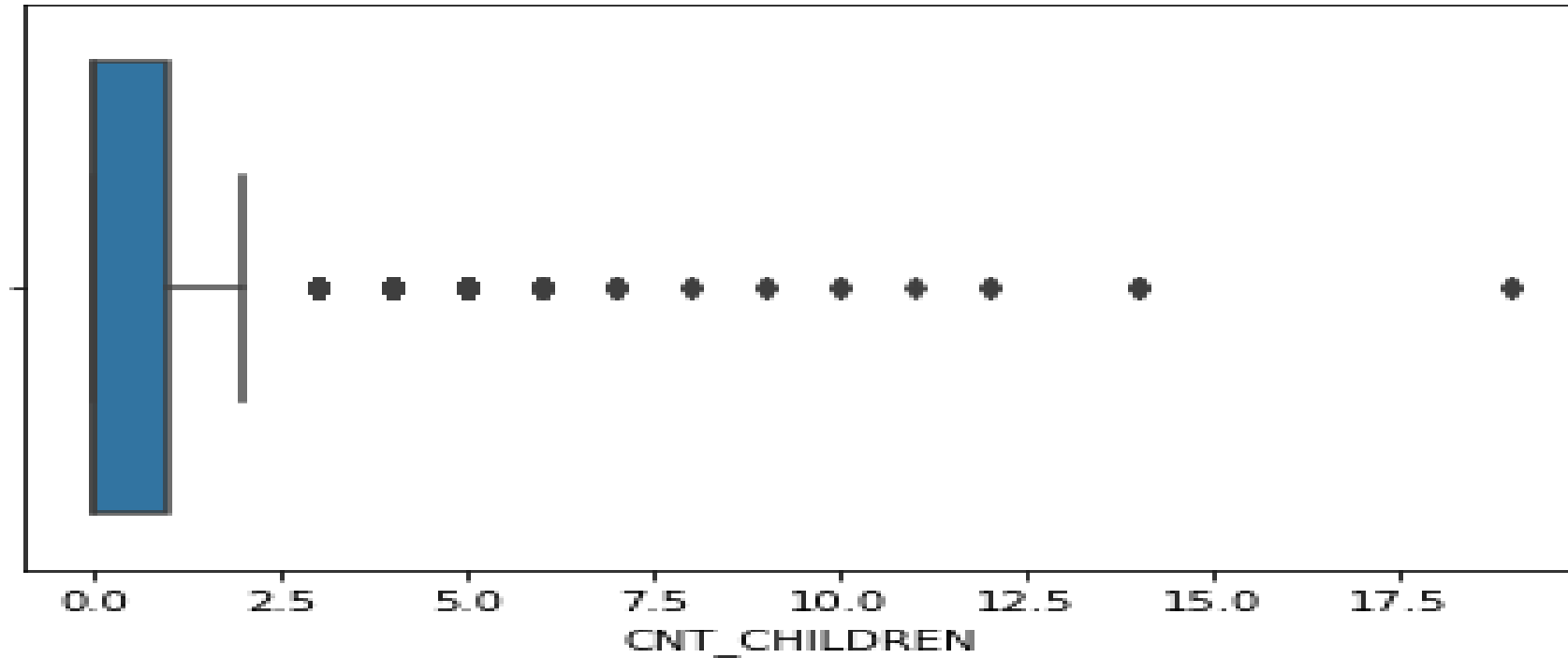
- One of the Client have Income of 117 Million which is found to be an outlier.

Amount Annuity (AMT_ANNUIITY)



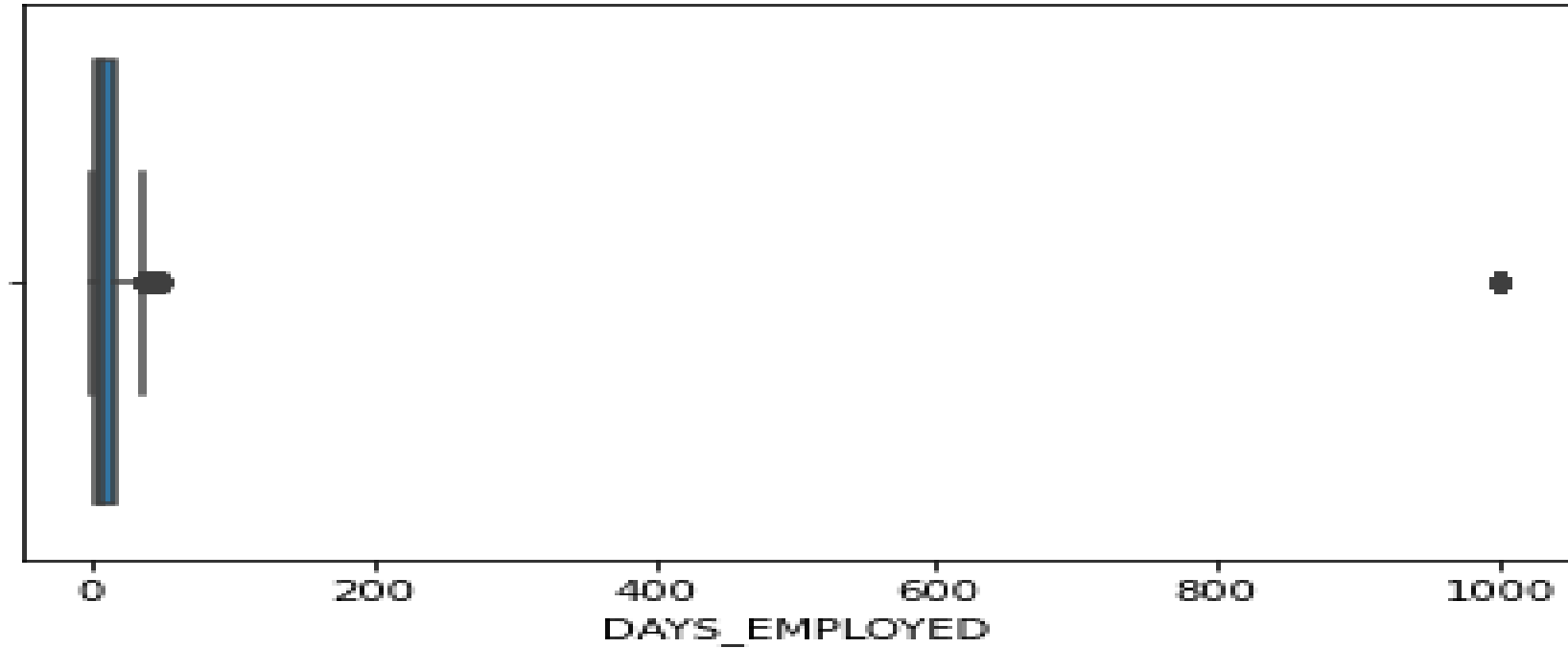
- Annuity Amount of one of the client has found to be 2,58,025 which is an outlier.

Count of Children (CNT_CHILDREN)



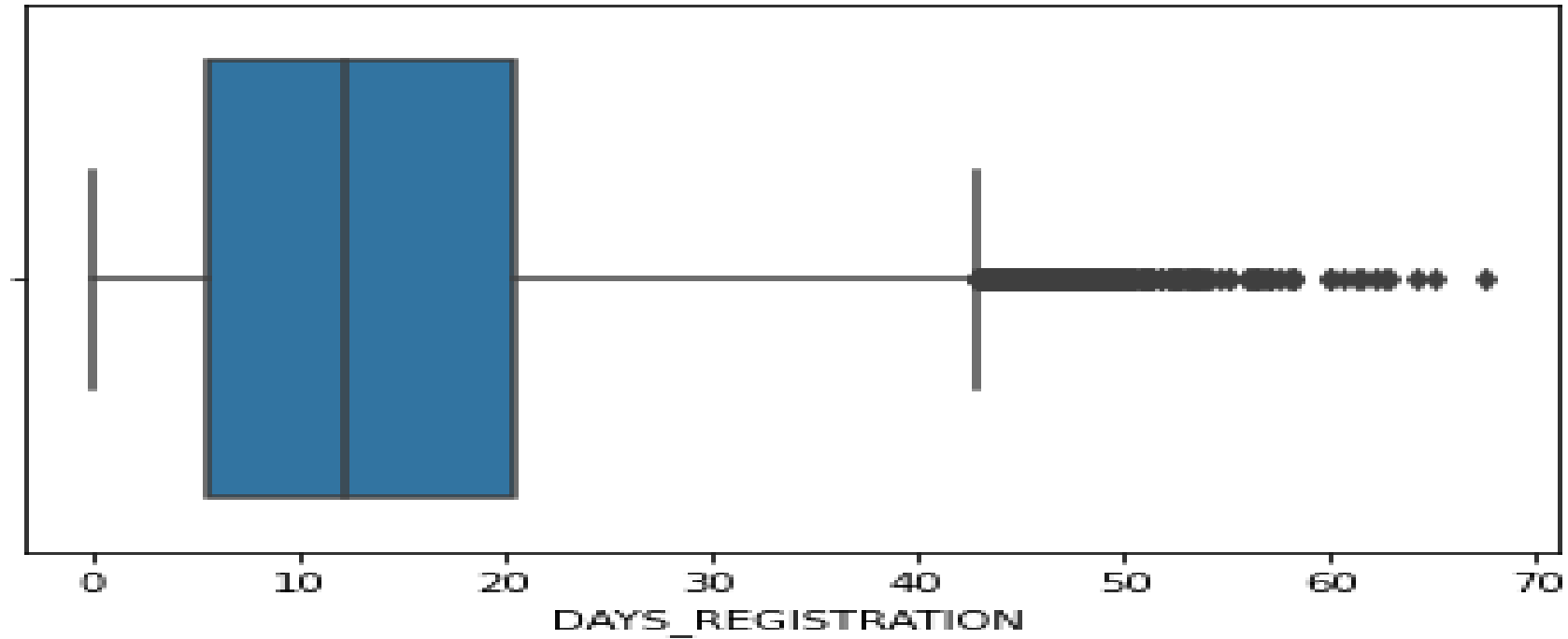
- One of the Client having 19 nos. of Children which found to be an outlier.

Years of Employment (DAYS_EMPLOYED)



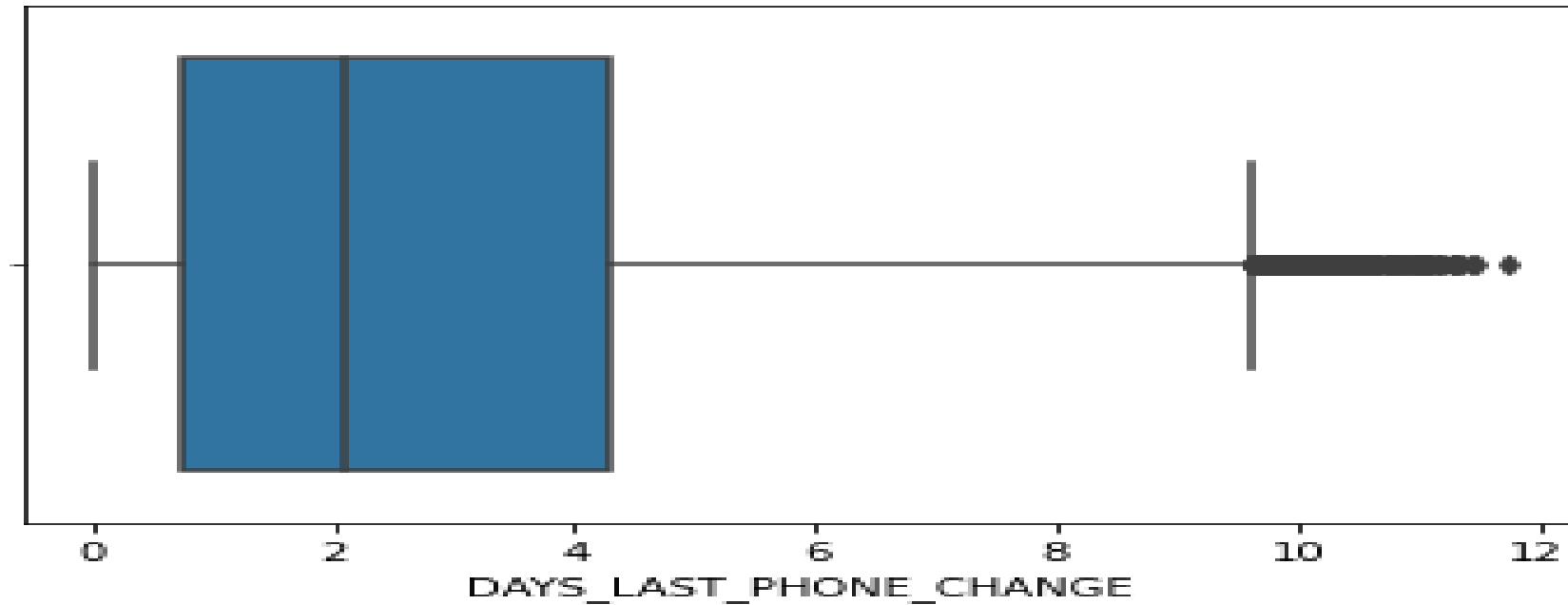
- Years of Employment has a value of 1000 years which is surely an outlier.

Years of Registration (DAYS_REGISTRATION)



- One of the Clint have registered more than 67 years back which is an outlier.

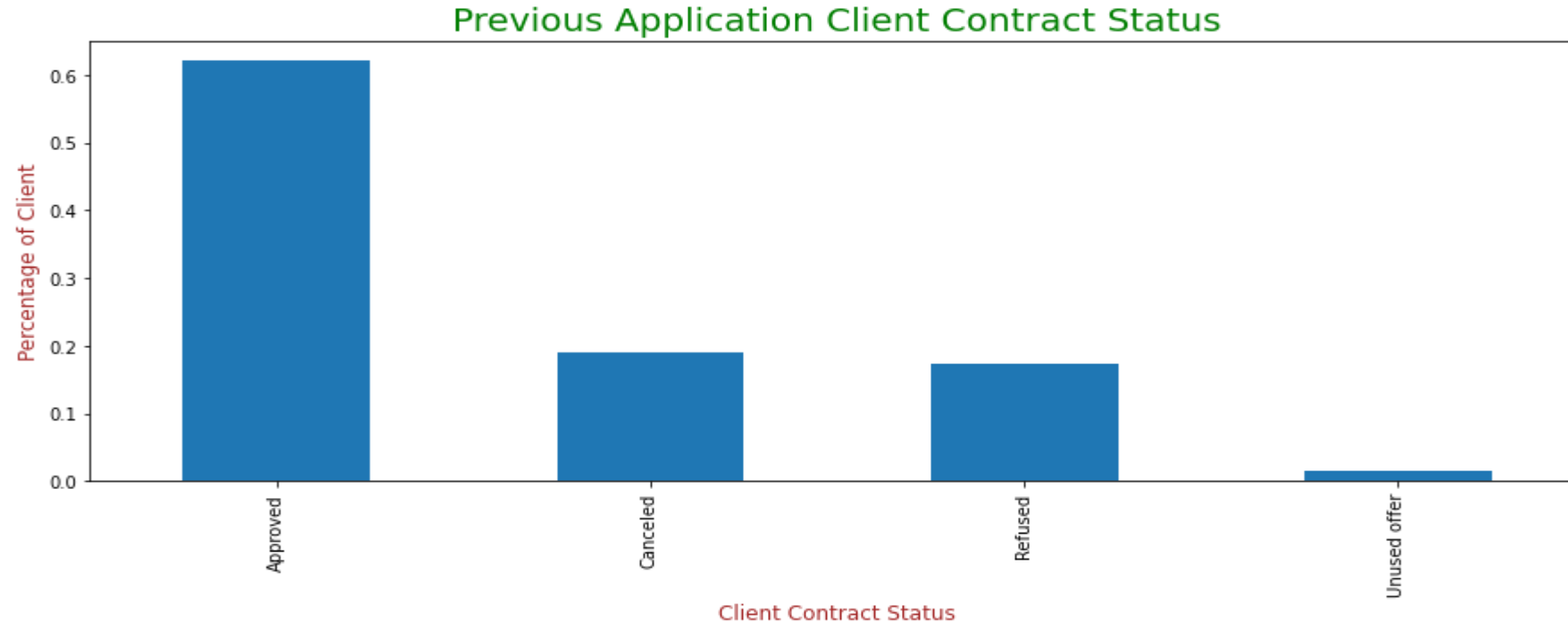
Last Phone Change (DAYS_LAST_PHONE_CHANGE)



- Last Phone Change Days has a value close to 11.75 years which seems to be beyond the normal value range and we can consider this as an outlier

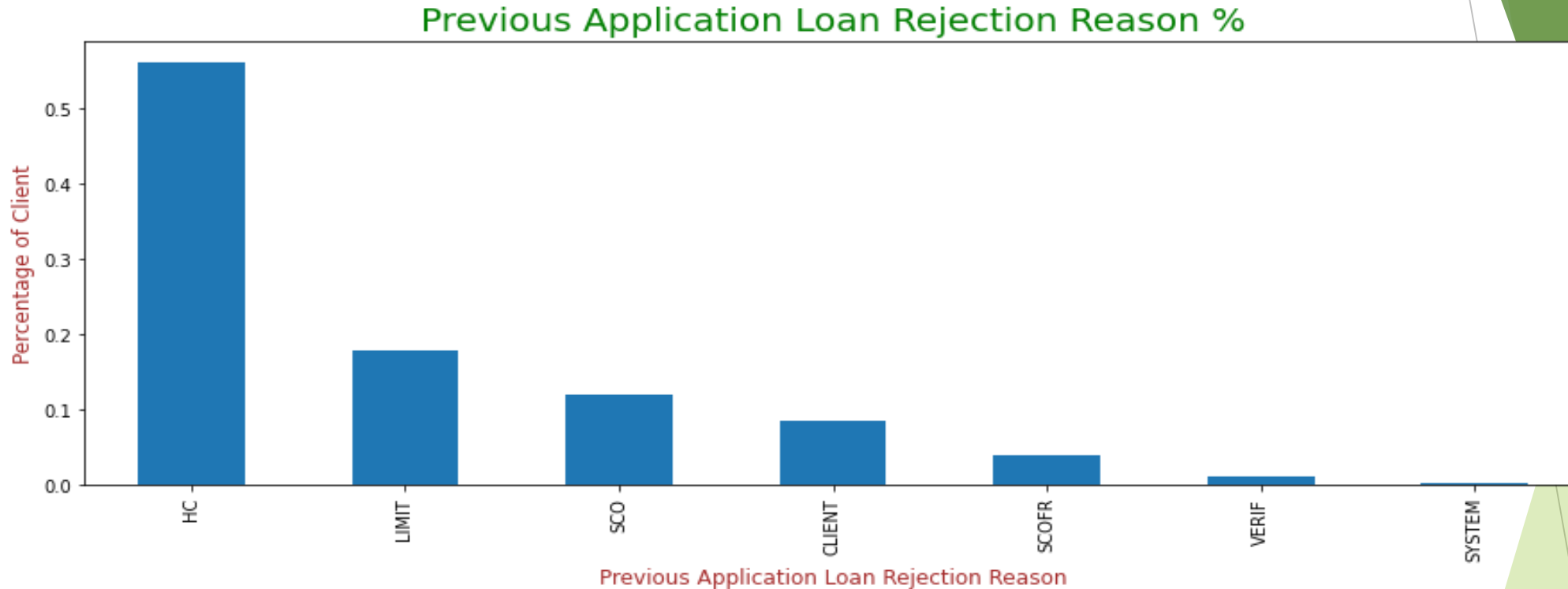
Univariate Analysis on Previous Loan Application

Client Contract Status



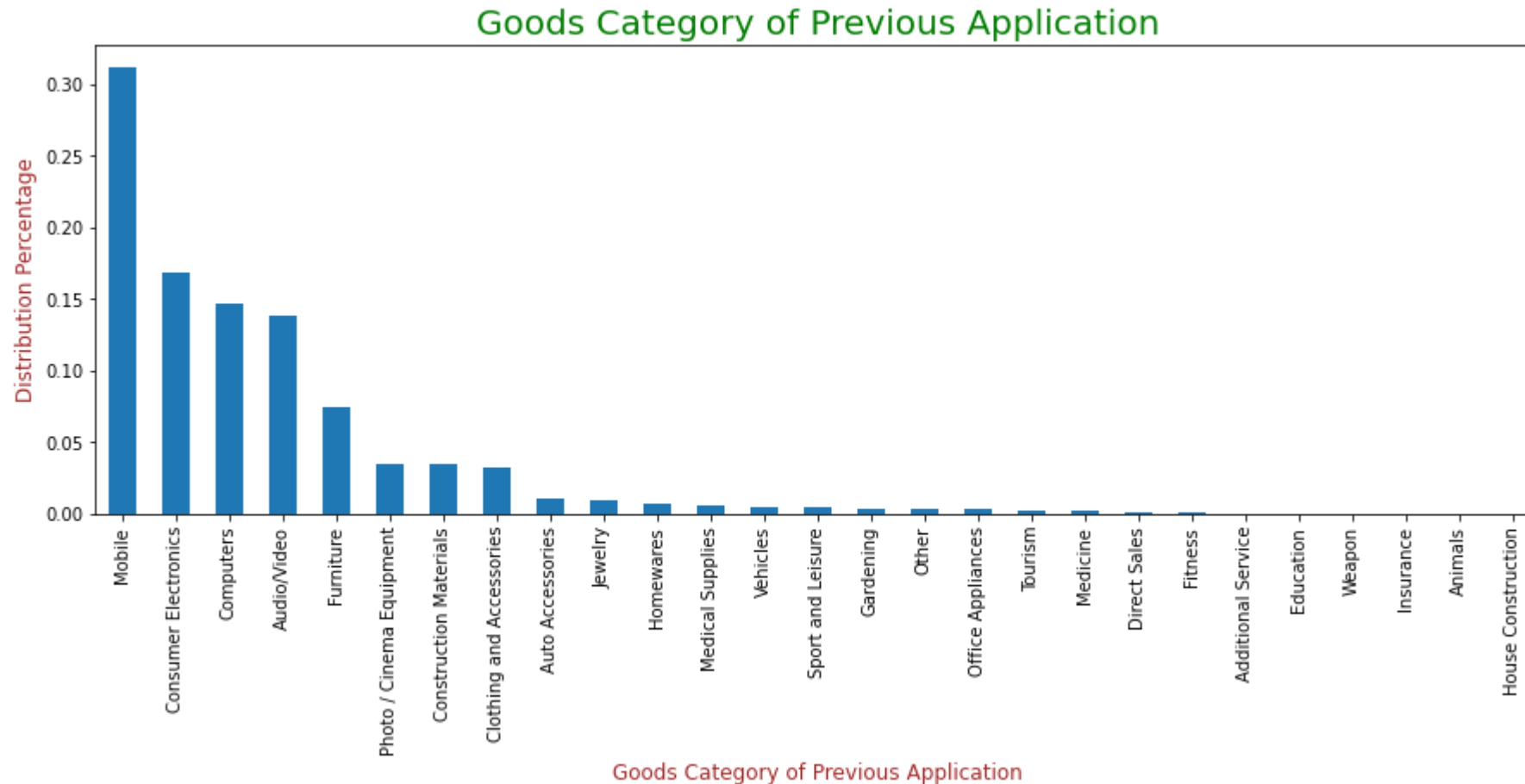
- More than 60% previous loan applications had been approved by the Financial Institution.
- Both Cancelled and Refused applicants percentages are less than 20.
- Loan has been offered to some clients and they have not utilized the offer. Numbers are very less in this category.

Loan Rejection Reason



- Most of the previous loan application got rejected due to HC reason followed by LIMIT and SCO reasons.
- Very few loan application got rejected due to SYSTEM reasons.

Goods Category

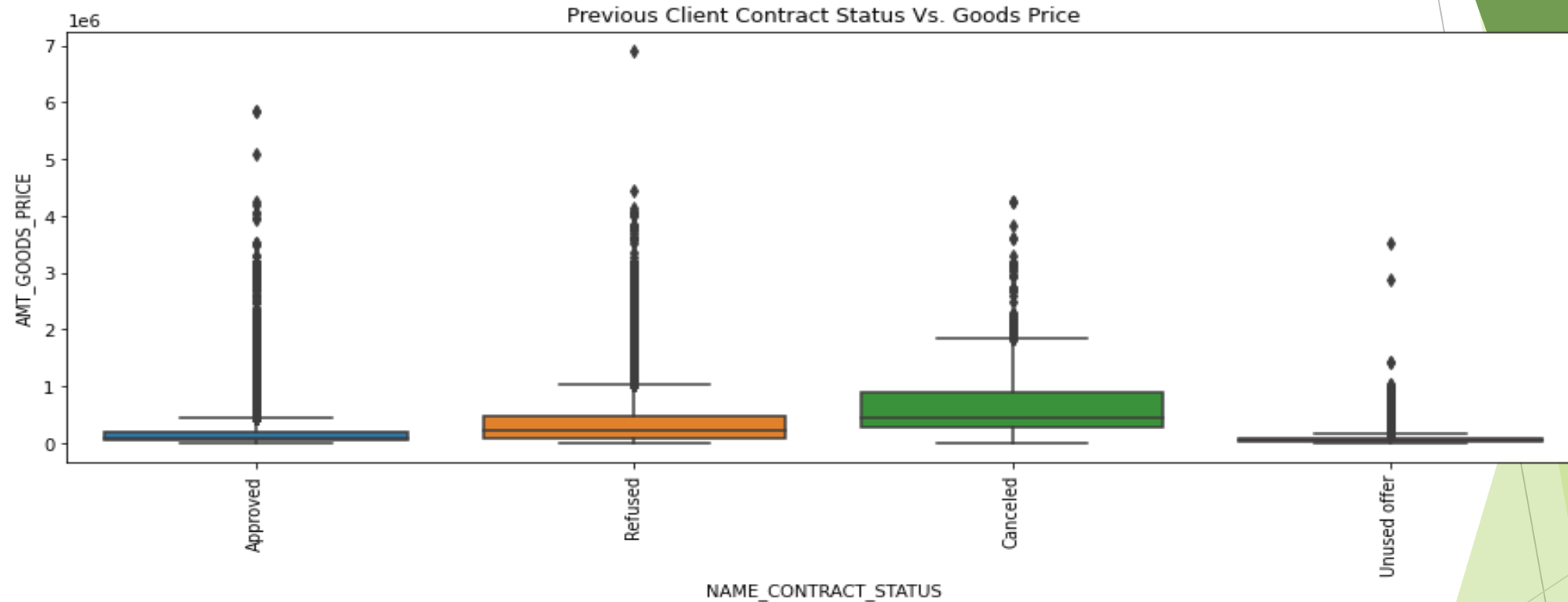


- As per previous application data, loan has taken mostly to purchase Mobile followed by Consumer Electronics & Computers.
- Least no. of loan applicants had applied loan for House Construction.

The background features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, creating a modern and dynamic visual effect.

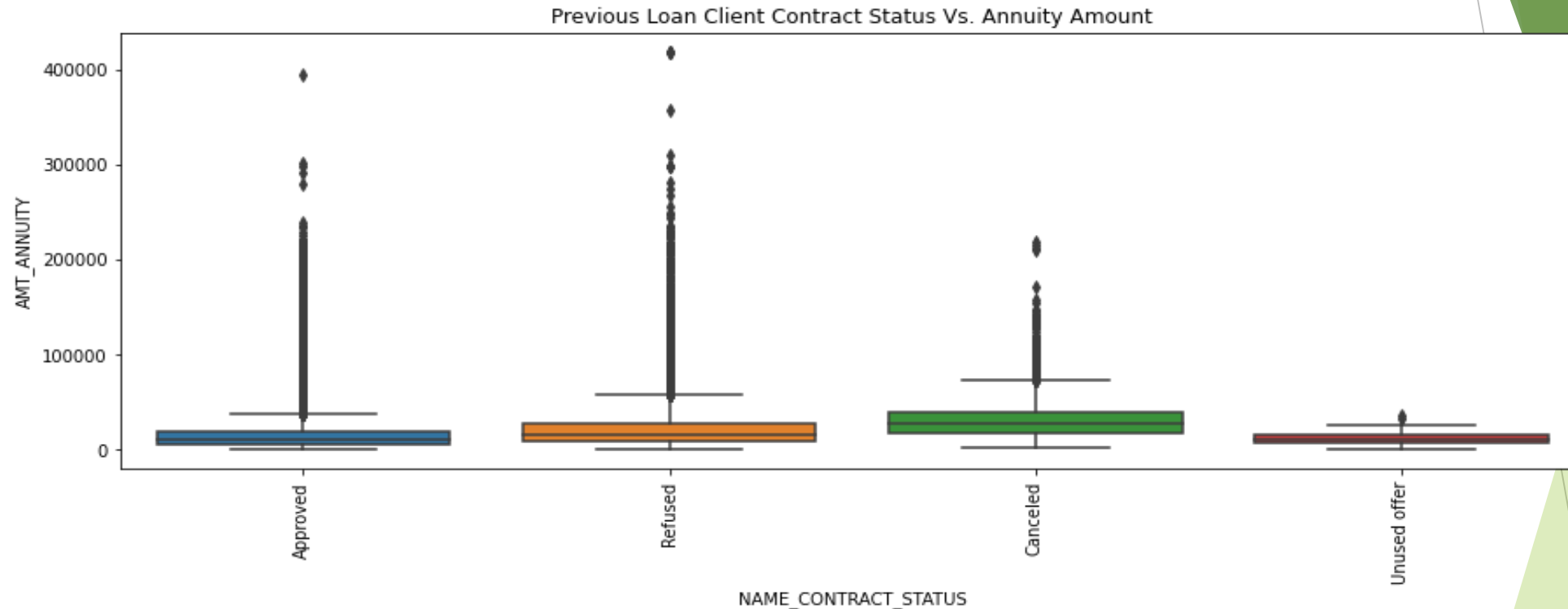
Bivariate Analysis on Previous Loan Application

Client Contract Status Vs. Goods Price



- Loan applied for higher value stand cancelled or refused by the Financial Institution.
- Loan got approved when the loan amount (Goods price) is comparatively lesser.

Client Contract Status Vs. Annuity Amount



- Loan applied for higher value would have high EMI amount. Whenever, EMI amount supposed to be evaluated high, it got cancelled or refused by the Financial Institution.
- Clients whose Loan got approved are paying lesser Annuity Amount.

Summary

Current Application

- Financial Institution should focus on Male applicant who are likely to repay the loan and access can be given for more credit amount on loan applied for.
- Financial Institution can consider approving loan to Businessman and Student who are most likely to repay the loan.
- Financial Institution should prioritize in sanctioning loan to Client having a) higher total income, b) bagged Academic degree, c) accompanied by Spouse/Partner, d) Separated and Widow.
- It is suggested to provide loan to client live in Region 1 and city 1.
- Client falls under maternity leave category should be avoided when sanctioning loan.
- Preference should be given to clients having own House/apartment and children count between 2 to 4.
- Financial Institution should assess the risk in issuing Cash loan of higher amount and can restrict approving Cash loans. Revolving loan of lower amount can be given priority over cash loan.

Summary

Previous Application

- More than 60% previous loan applications had been approved by the Financial Institution.
- Most of the previous loan application got rejected due to HC reason followed by LIMIT and SCO reasons.
- Client has taken loan mostly to purchase Mobile followed by Consumer Electronics & Computers.
- Loan applied for higher value stand cancelled or refused by the Financial Institution.
- Loan got approved when the amount (Goods price) is comparatively lesser.

