EDA Credit Analysis - By Gautami

How consumer finance can avoid signing loans to defaulters

Problem Statement

Company

A consumer finance company specializing in urban loans faces challenges in approving loans for individuals without sufficient credit history, risking financial losses from defaults. The company aims to minimize these risks through Exploratory Data Analysis (EDA) to make informed loan approval decisions.

Context

The company confronts the dilemma of distinguishing creditworthy applicants from potential defaulters due to insufficient credit histories. The provided dataset includes information on loan applications, highlighting clients with payment difficulties and those with timely payments. The focus is to identify patterns that influence loan default risks.

Problem statement

The challenge is to employ EDA to find patterns that help identify applicants likely to repay or default on loans, thereby guiding the company's loan approval decisions. The objective is to discover key indicators of loan default, enabling the company to manage its loan portfolio and assess risks effectively, and decide on appropriate actions such as loan denial and identifying fraud.

Challenges deep-dive

Challenge 1

Data Quality and Completeness

One of the primary challenges in any EDA is handling missing, incorrect, or inconsistent data. In the context of loan applications, crucial information might be missing or inaccurately reported such as income levels, employment history, or previous loan repayment records.

Challenge 2

High Dimensionality

Loan application datasets often contain a vast number of variables, including applicant demographics, financial information, loan characteristics, and historical payment data. High dimensionality can make the analysis complex and computationally intensive.

Challenge 3

Bias and Fairness

There is a risk of bias in the data, which can lead to unfair loan approval decisions. For instance, if the historical data on which the EDA is based contains biases against certain groups of people, the analysis might inadvertently perpetuate these biases.

Solution

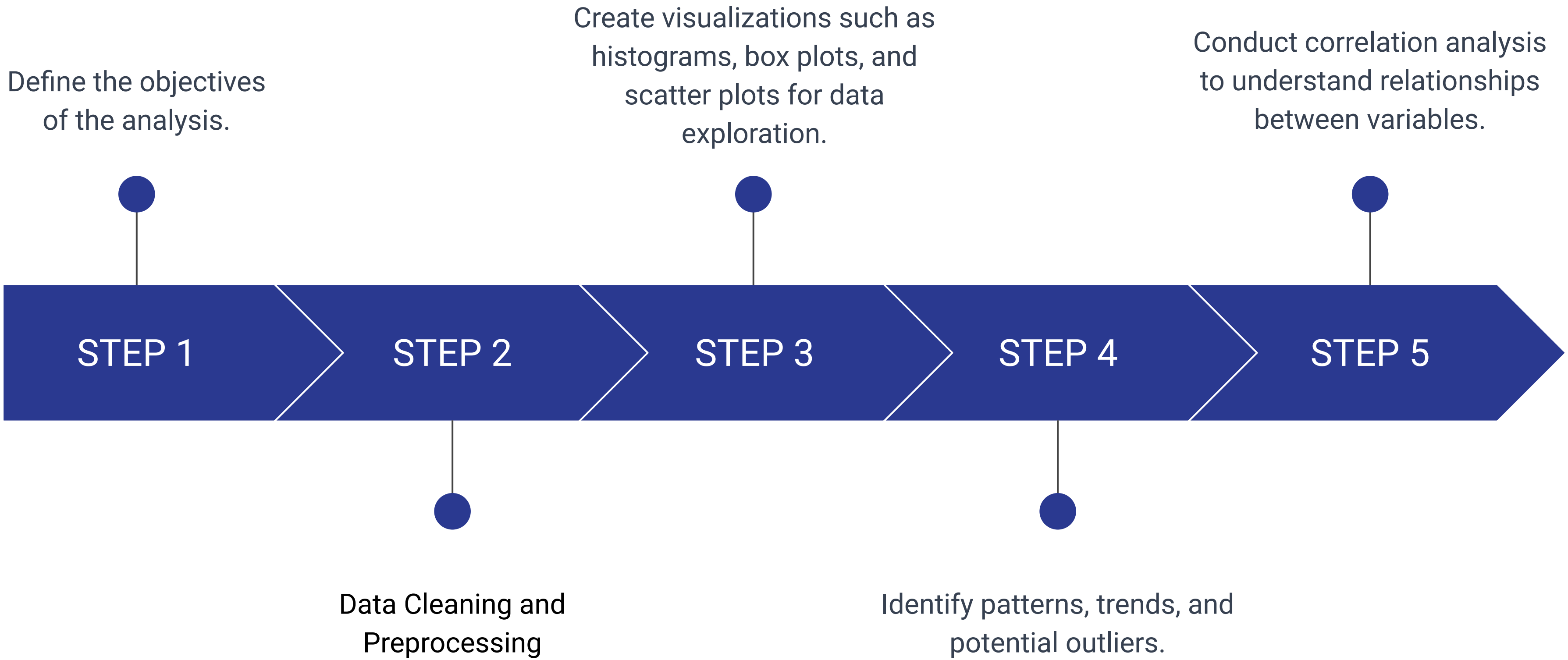
To make the most of it !

Data Quality and Completeness: Implement data cleaning and imputation techniques to handle missing or incorrect values, ensuring a more accurate representation of applicant information.

High Dimensionality: Utilize feature selection methods, such as correlation analysis or dimensionality reduction techniques to streamline the dataset and focus on the most relevant variables for loan approval decision-making.

Bias and Fairness: Properly address biases in the dataset, employ fairness and establish transparent, ethical lending policies to ensure fair and unbiased loan approval processes.

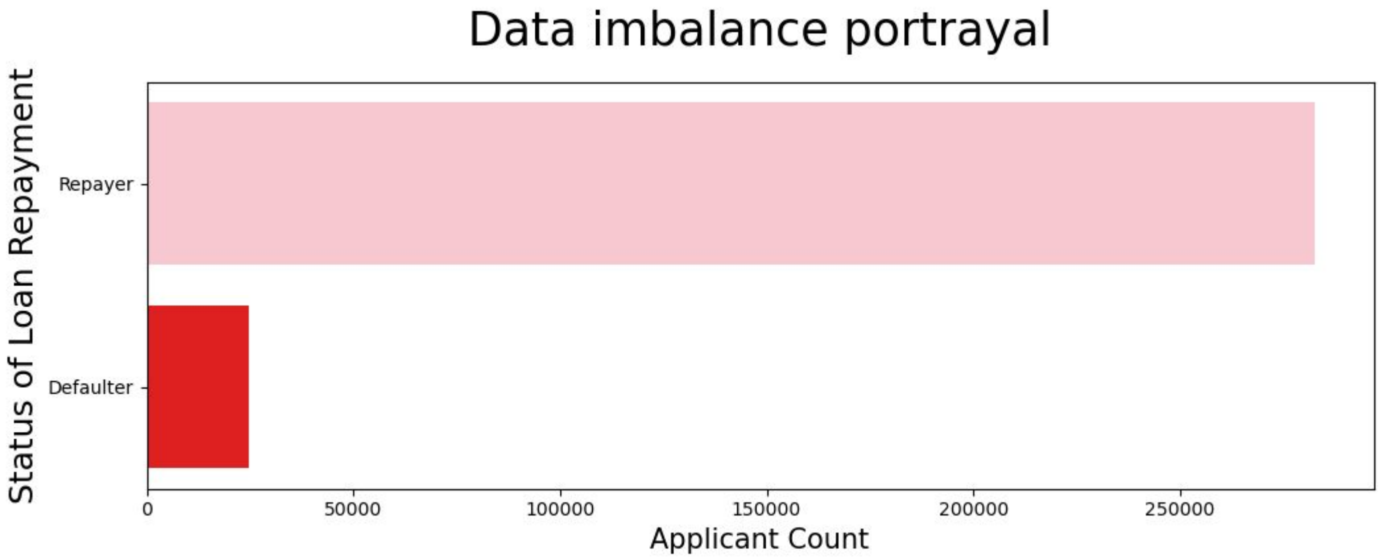
Implementation


Define the objectives of the analysis.
STEP 1
STEP 2
Data Cleaning and Preprocessing
Create visualizations such as histograms, box plots, and scatter plots for data exploration.
STEP 3
STEP 4
Identify patterns, trends, and potential outliers.
Conduct correlation analysis to understand relationships between variables.
STEP 5

Overall visual of the data provided:

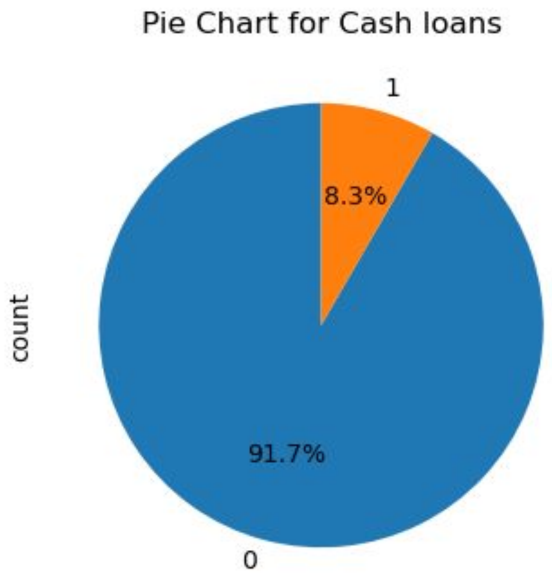
Total Percentage of people repaying the loan is 91.93%

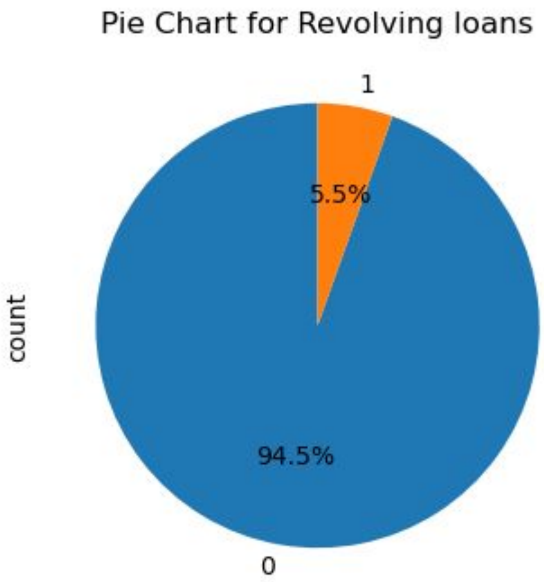
Total Percentage of people becoming defaulters is 8.07%



Visual analysis of loan type w.r.t. Repayers/Defaulters:

- In case of Cash Loans, 8.3% were defaulters whereas 91.7% were on-time payers
- In case of Revolving Loans, 5.5% were defaulters whereas 94.5% were on-time payers





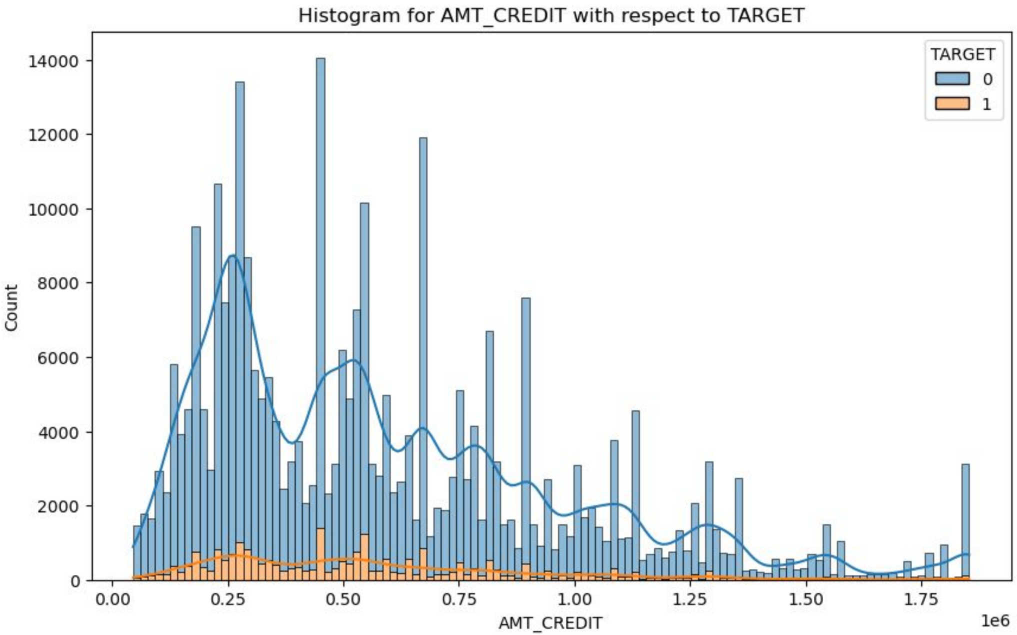
0 represents Defaulters , 1 represents Repayers

Visual analysis of pending credit amount w.r.t Repayers/Defaulters:

- The highest number of defaulters (count between 1000-2000) had credit pending of amounts upto 5 Lakh

- Most defaulters seems to lie within the range of pending credit between 2.5 Lakh to 7.5 lakh

- We don't see any promising pattern as to defaulters relating to pending credit

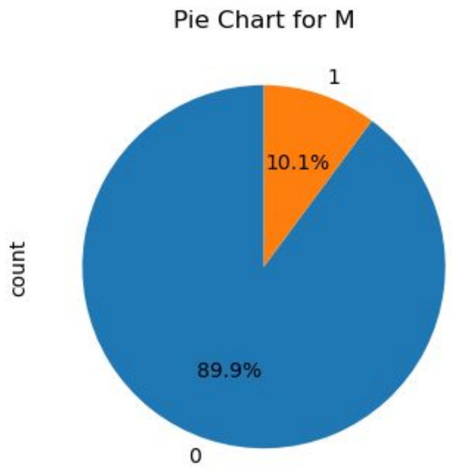


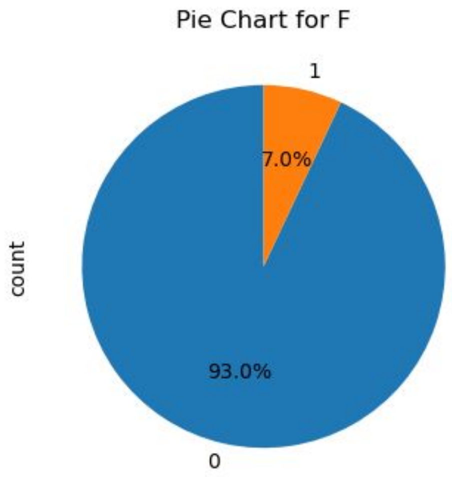
1 represents Defaulters , 0 represents Repayers

Visual analysis of applicant's gender w.r.t Repayers/Defaulters:

- Accounting to female applicants, defaulters were found to be 7% of the total loan bearers

- Accounting to male applicants, defaulters were found to be 10.1% of the total loan bearers

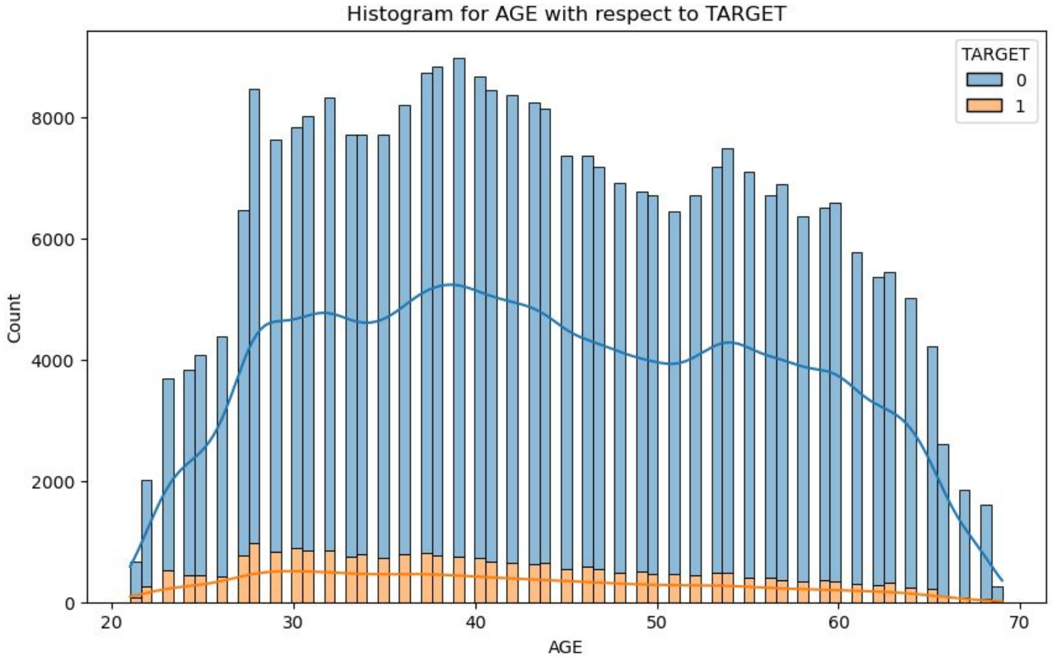




1 represents Defaulters , 0 represents Repayers

Visual analysis of applicant's age w.r.t Repayers/Defaulters:

We cannot infer anything specific by this plot as there is not meaningful correlation as to why an applicant of certain age will repay or default the given annuity

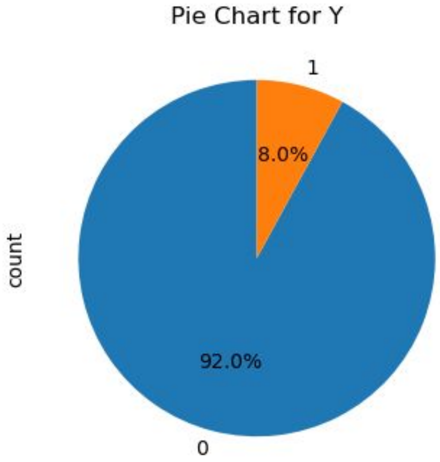


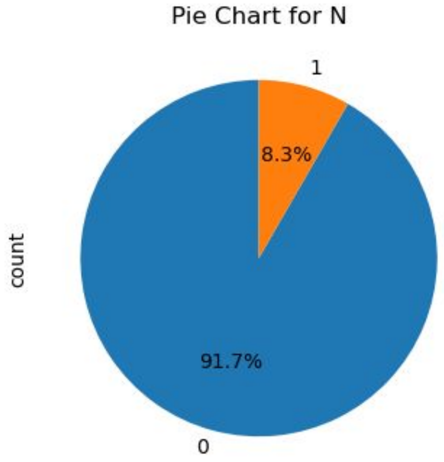
1 represents Defaulters , 0 represents Repayers

Visual analysis of applicant who owns real estate w.r.t Repayers/Defaulters:

- 8.3% of people who own real estate were found to be defaulters

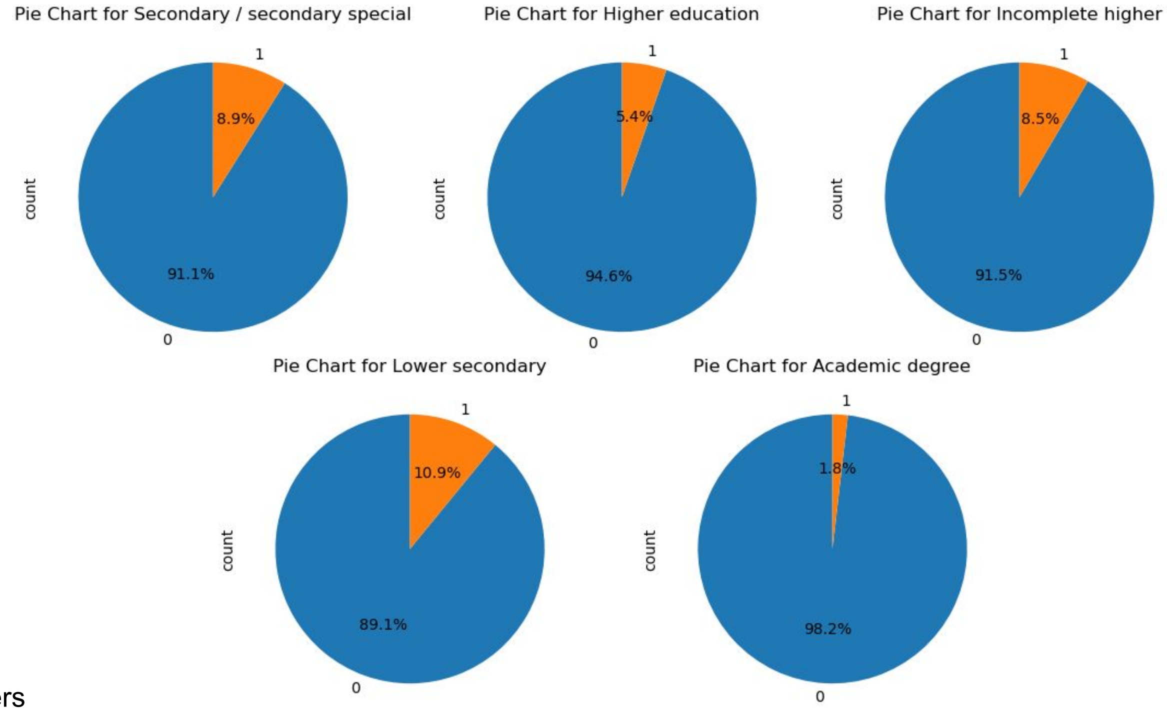
- 8.0% of people who don't own real estate were found to be defaulters





1 represents Defaulters , 0 represents Repayers

Visual analysis of applicant's education type w.r.t Repayers/Defaulters:



- People part of lower secondary education have the highest percentage of defaulters - 10.9%
- After comparing all of the charts we can say that at 1.8%, people with academic degrees have the least tendency to be a defaulter

1 represents Defaulters , 0 represents Repayers

Visual analysis of applicant's income type w.r.t Repayers/Defaulters:

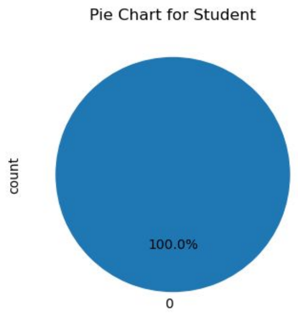
- Most of the loan applicants are working

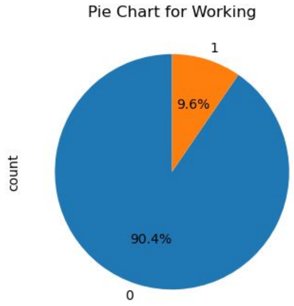
- Top top defaulters are - applicants on maternity leave (40%) and unemployed applicants (36.4%)

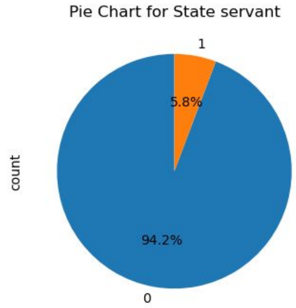
- There are some defaulters that constitute 10% of some of the categories like Pensioner, Commercial associate etc.

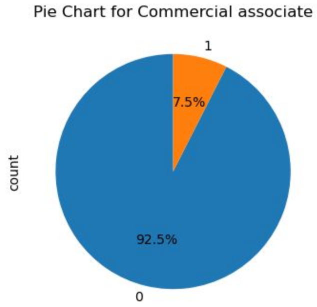
- Best categories to consider for loan will be Student and Businessmen as they have no record of anu default

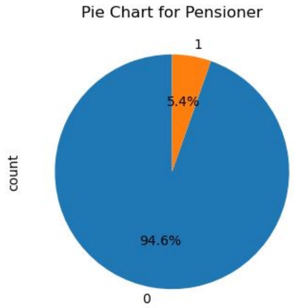
1 represents Defaulters , 0 represents Repayers

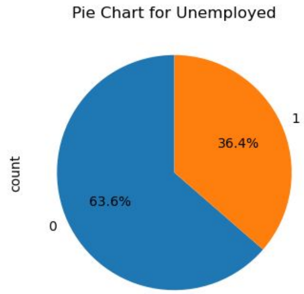


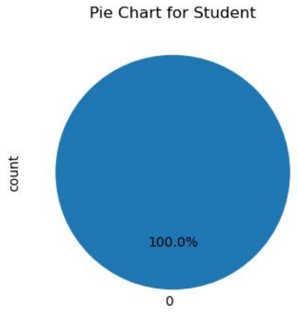






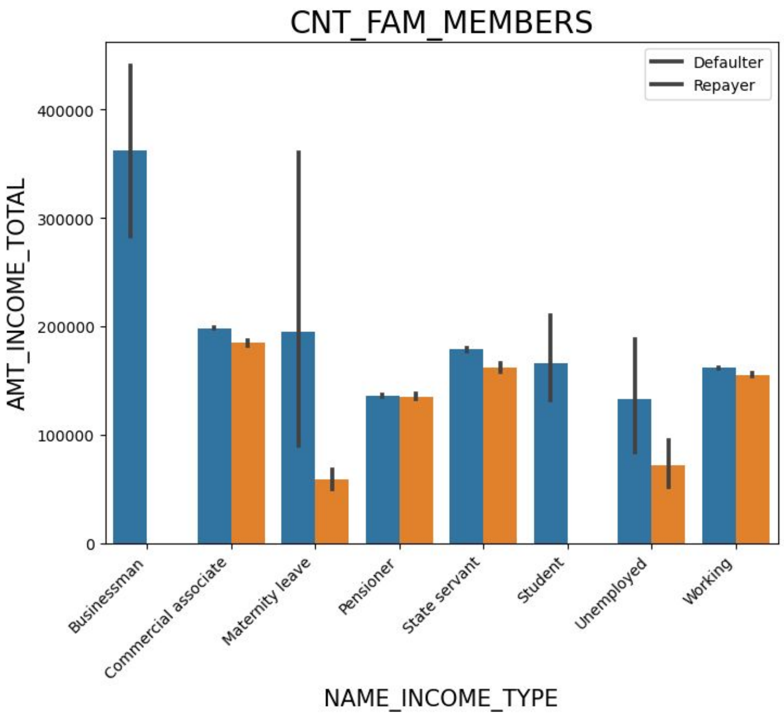






Visual analysis of applicant's total income and income type w.r.t Repayers/Defaulters:

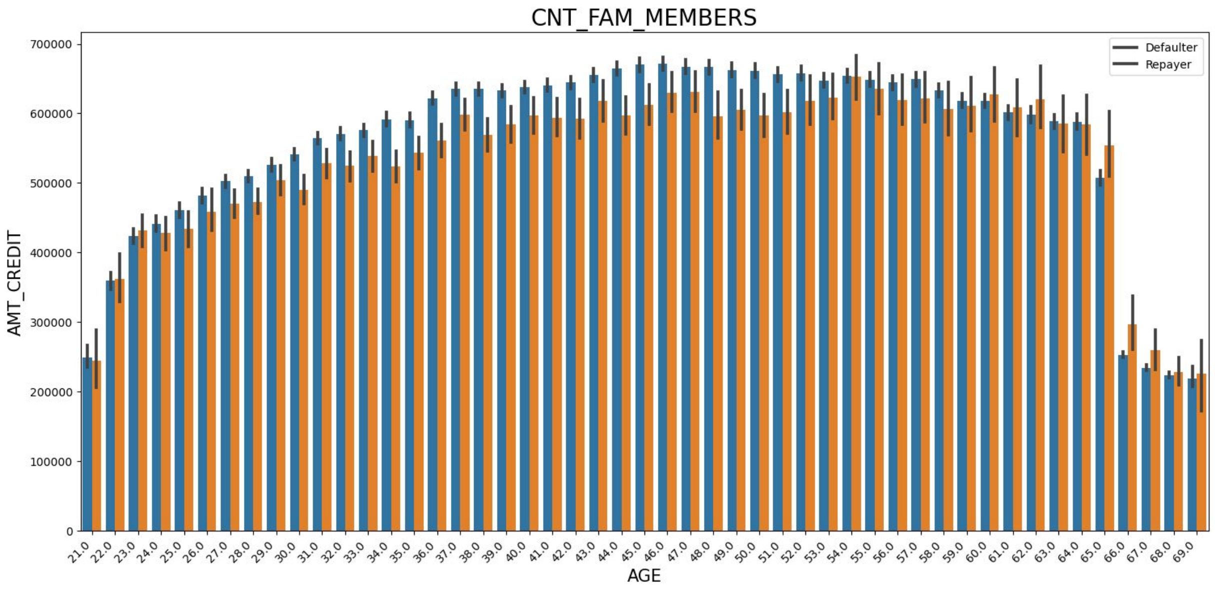
- In case of Businessman, the income seems start from beyond 2.5 Lakh.
- There is no range of income where this income category has seen defaulters, hence can be the best people to be considered for loans.
- In case of Commercial associate the defaulters income is seen to be below 2 lakh altogether.
- Repayers have there income starting from 2 lakh and above.
- In case of Maternity leave, defaulter's income range is more than 0.4 Lakh and less than 0.6 Lakh.
- Repayers in the same category have income in range starting from 0.9 Lakh to 2.9 Lakh.
- In case of Pensioner, repayers and defaulters are present in the same income range hence its difficult to infer if the defaulters are influenced by income at all.
- Another best category to consider for loans is of 'Student' as they have zero defaulters.
- In case of unemployed, defaulters have their highest income touching 1 Lakh which may be in other forms and not exactly as income.
- Repayers in the same category lie in the better income range which is 0.8 lakh to 1.8 Lakh



Visual analysis of applicant's pending credit and age w.r.t Repayers/Defaulters:

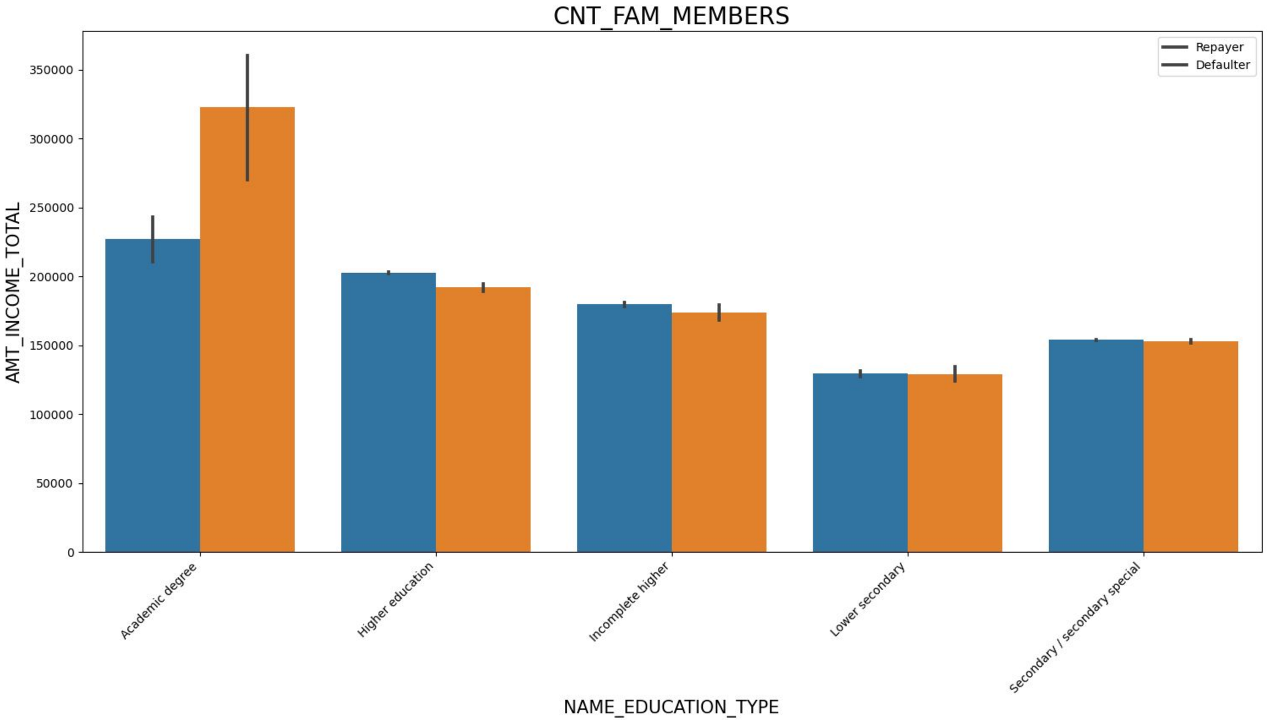
- We have the least amount of defaulters in the age range of 65+ years and a few in 21-22 years old range.

- The people having the highest amount of pending credit amount are the ones with age rage of 43-49 years and 54-62 years.



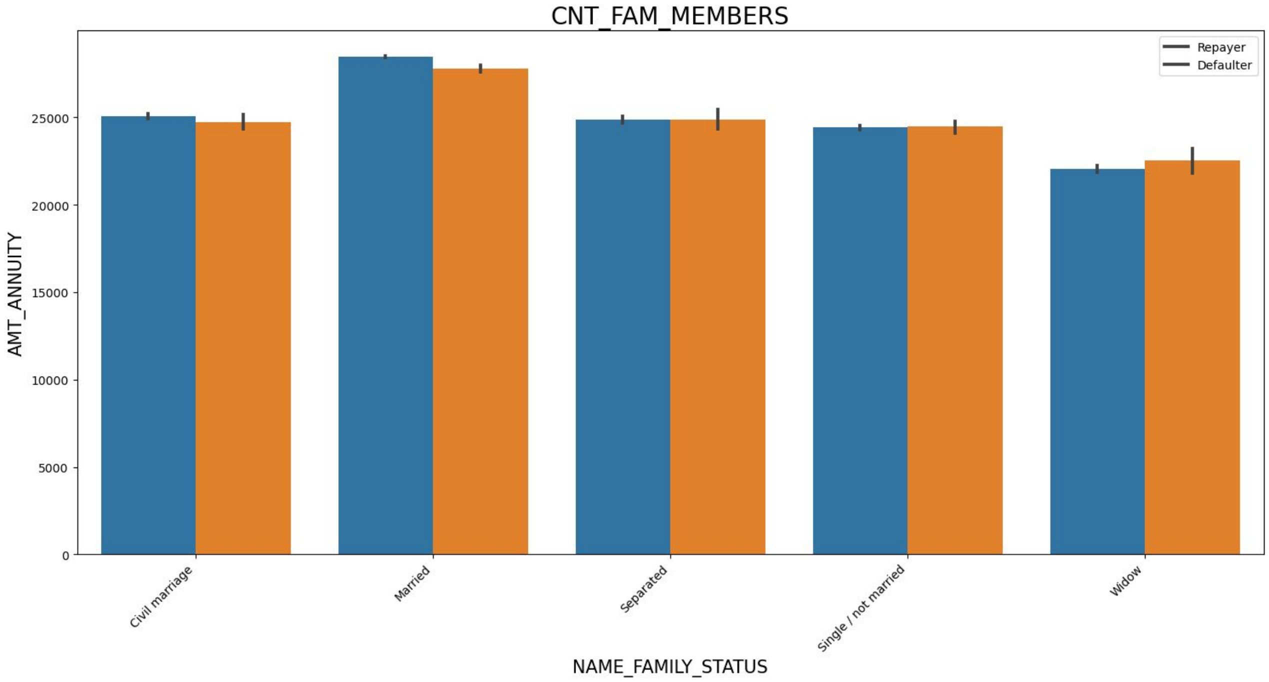
Visual analysis to investigate if education levels impact income:

- The highest number of defaulters are under the education type 'Academic degree' with total income amount of 2.7 Lakh to 3.6 Lakh
- The least amount of defaulters lie under the category of 'Lower Secondary' and with the income range of 1.3 Lakh to 1.4 Lakh



Visual analysis to see how family status affects annuity:

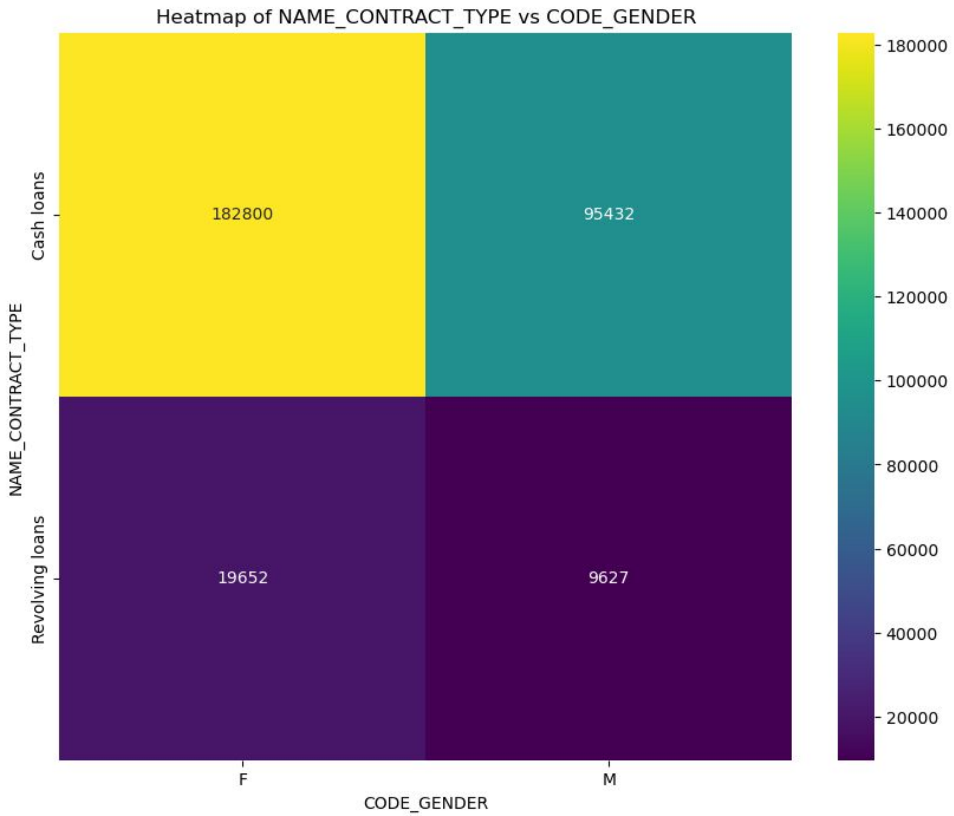
After plotting the bivariate analysis for family status and annuity amount we can say that there is no specific pattern that may influence one to become a defaulter, hence family status is not one of the ground to qualify an applicant for a loan.



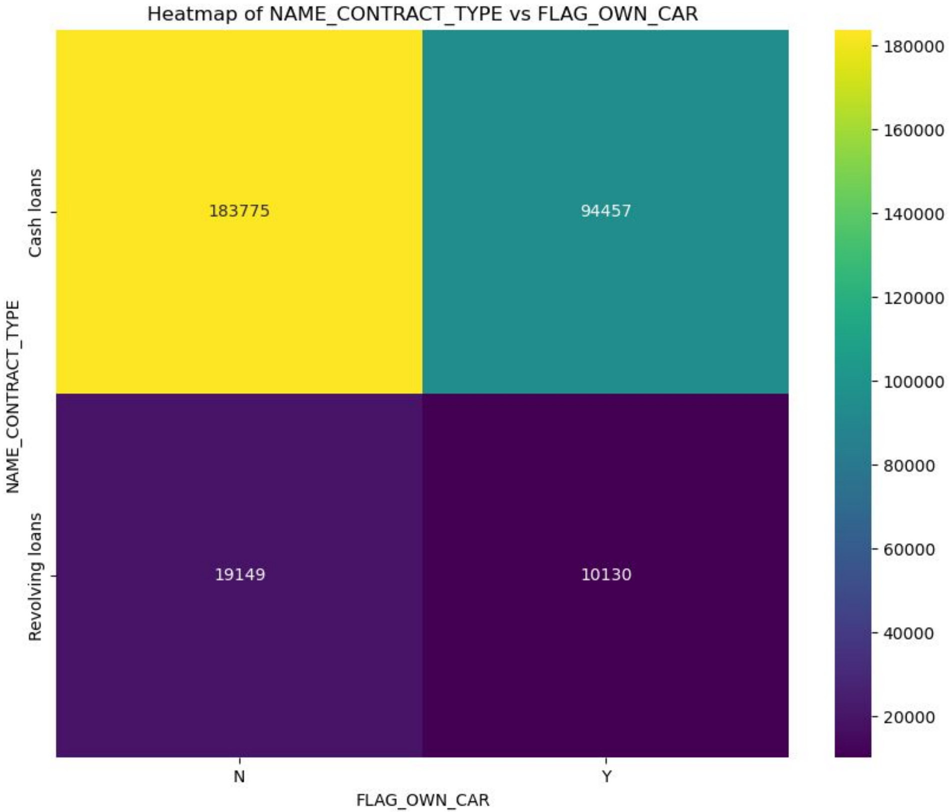
Visual analysis between contract type and applicant gender:

- Overall the most amount of Cash loans are the ones taken by females.

- Males with revolving loans have the least count in the entire data set.



Visual analysis between contract type and applicant with car:

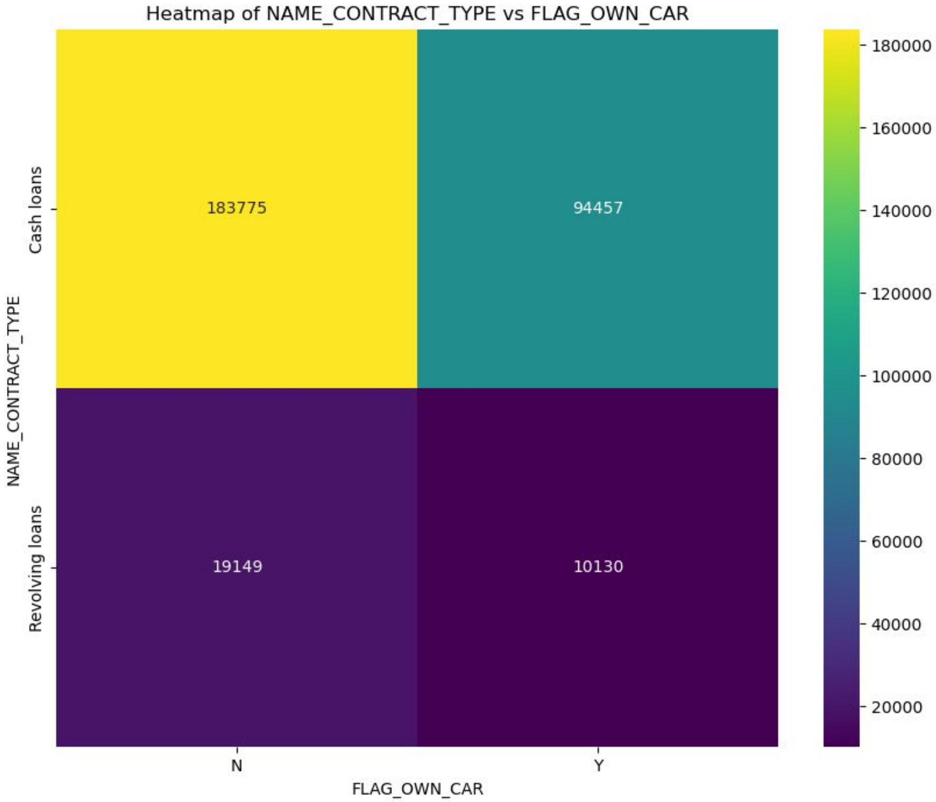


- People who don't own cars are more inclined towards cash loans.

- People who own cars are equally distributed between cash loans and revolving loans.

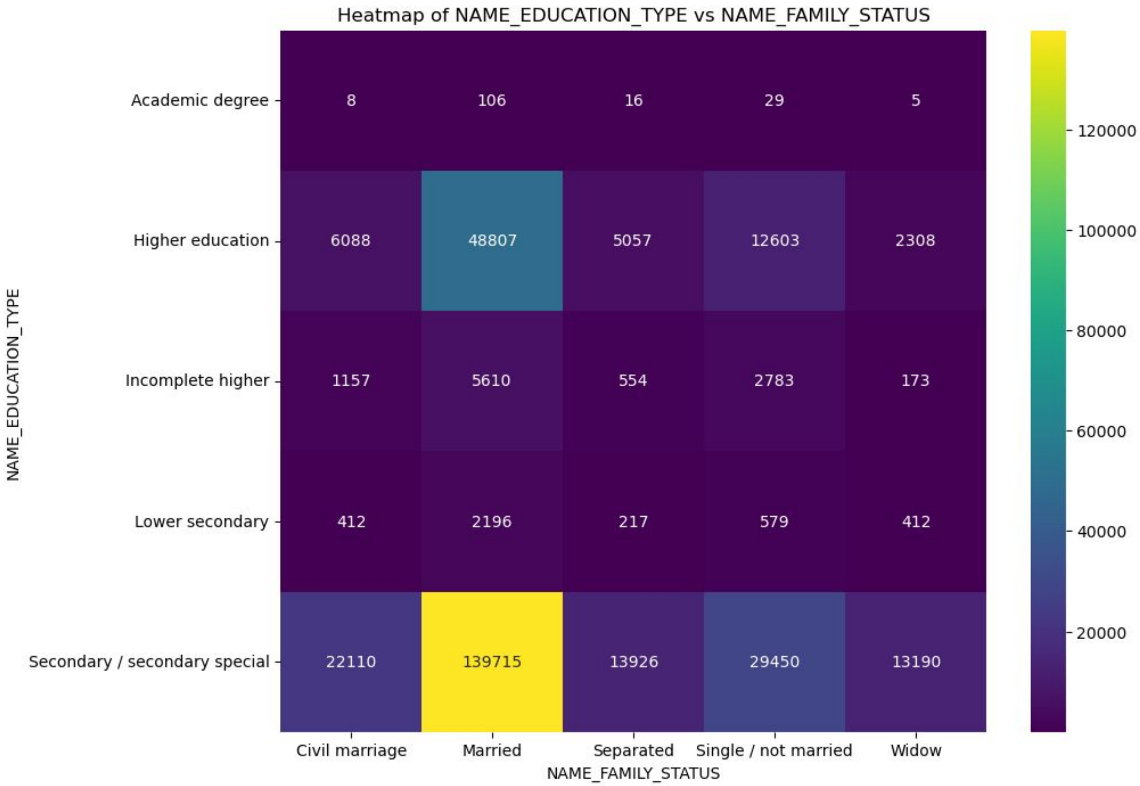
Visual analysis between contract type and applicant with real estate:

- People who do not own any real estate and apply for loans are comparatively less in number.
- In those few people who do apply, more incline towards cash loans than revolving loans.
- People who own real estate tend to apply for loans a lot more. A very large number of them prefer cash loans over revolving loans.



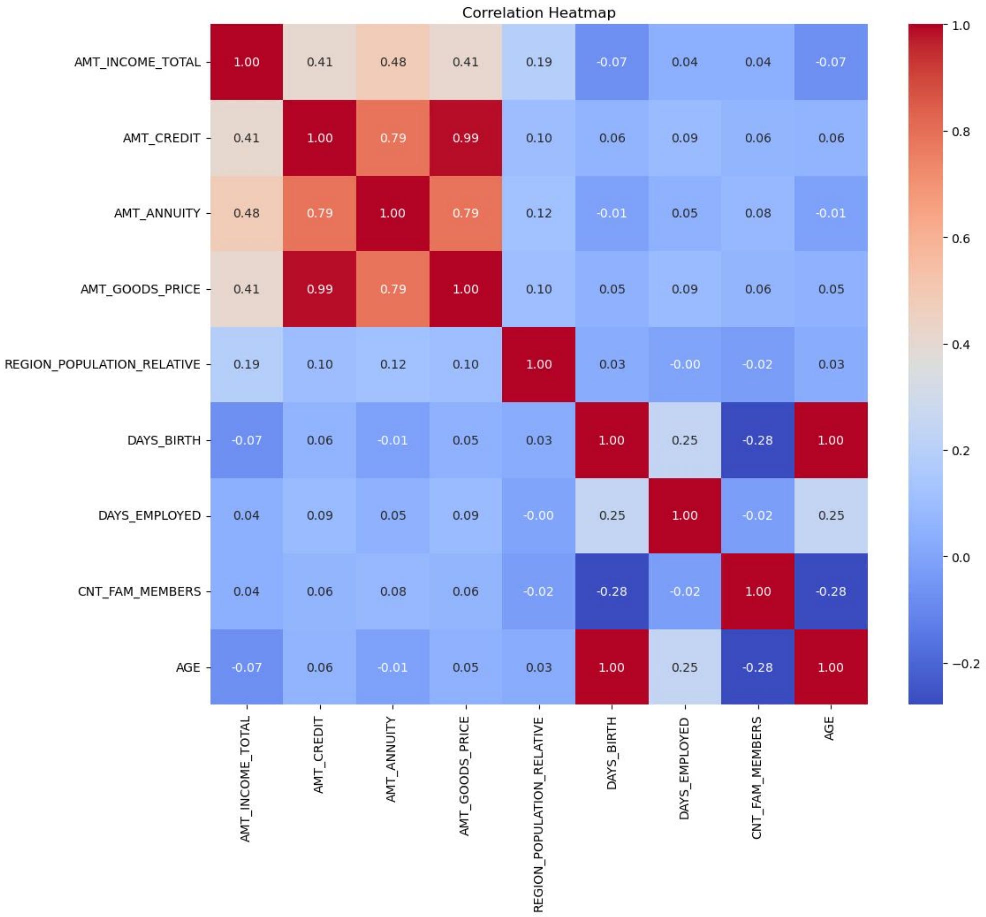
Visual analysis between level of education and family status of applicant:

- Most of the applicants come from married background and have secondary level of education.
- The second category of high applicants are married and attained higher education.
- Least amount of applicants are the ones who are widowed and have an academic degree & people with civil marriage and holding an academic degree.



Visual analysis between potential correlations:

- Pending credit amount & Goods Price: Applicants seeking higher loan amounts inclined to purchase more expensive goods, leading to a positive correlation with goods prices.
- Pending credit amount & Annuity Amount: As the loan amount increases, the associated annuity amount may also rise, reflecting the financial commitment proportional to the loan size.
- Annuity Amount & Goods Price: Annuity amounts could be influenced by the cost of goods, as higher-priced goods might result in larger fixed monthly payments.
- Total Income & Annuity Amount: Individuals with higher total income may qualify for larger loans , reflecting their capacity to manage higher monthly payments.
- Pending credit amount & Total Income: Applicants with higher total income may seek larger loans, reflecting their financial capability to handle more significant credit obligations.
- Total Income & Goods Price: Individuals with higher total income might have the financial capacity to purchase more expensive goods, leading to a positive correlation with goods prices.



Visual analysis of income, annuity and credit:

Density plot for total income amount between defaulters and repayers:

- People with higher income were able to repay better than people with lower income

Density plot for total credit amount between defaulters and repayers:

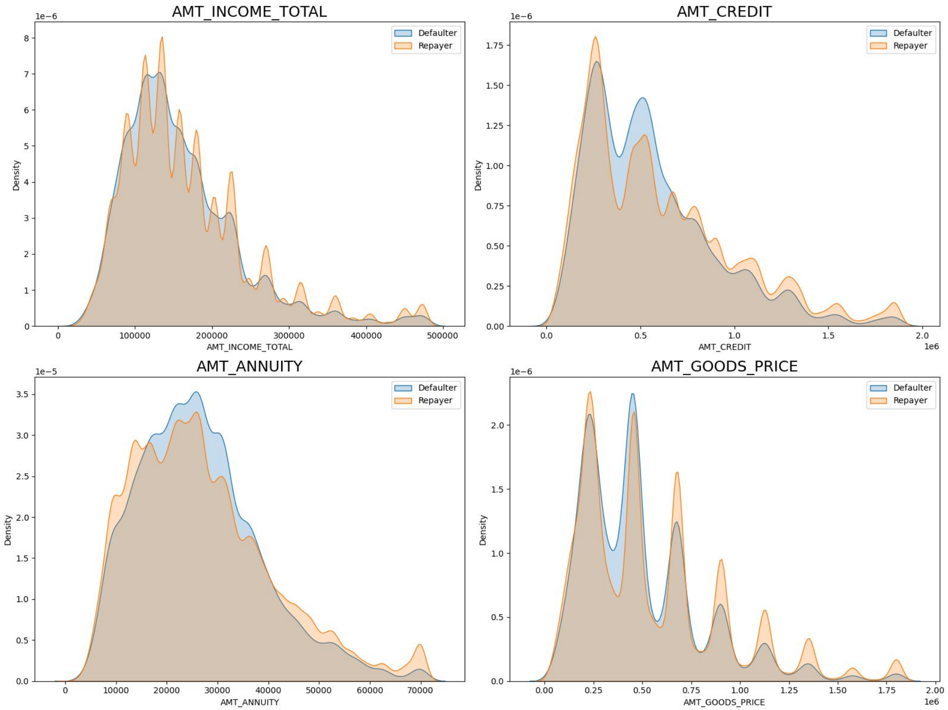
- People having pending credit upto 5 Lakhs turned out to be more of defaulters than repayers
- People who had less pending credit where able to repay the loan taken

Density plot for annuity amount between defaulters and repayers:

- People having annuity between 1.5 Lakh to 3.3 Lakh had higher number of defaulters.

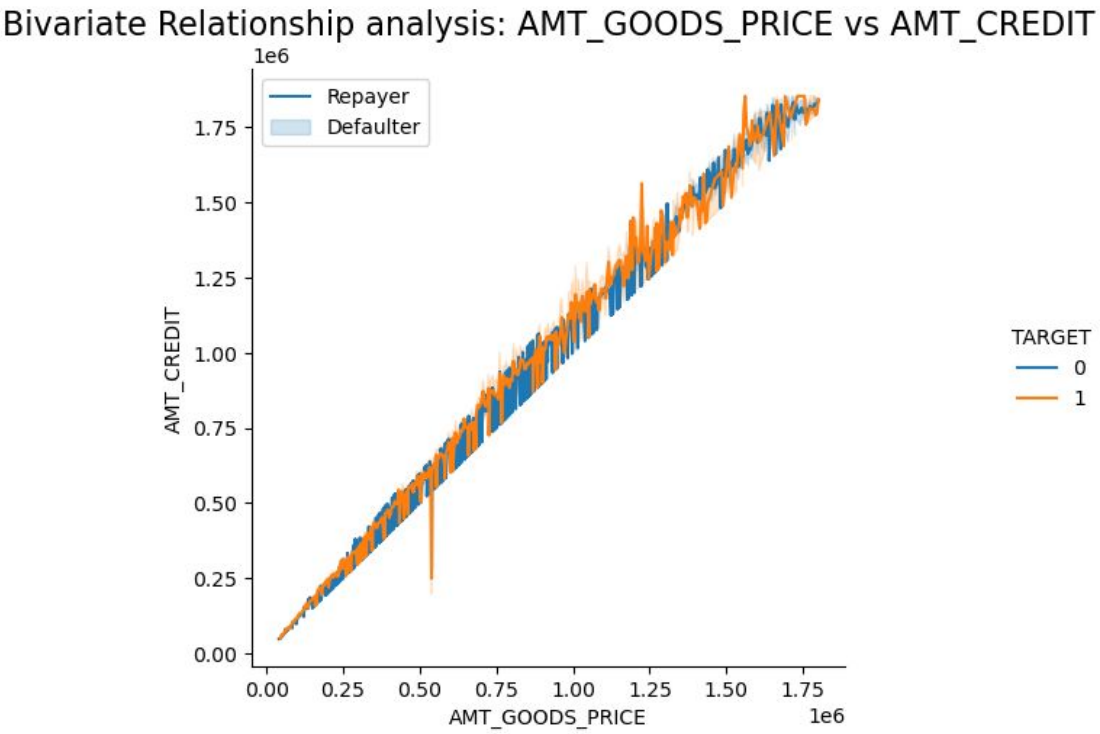
Density plot for total goods price between defaulters and repayers:

- The only time where defaulters took over repayers in total was when the goods price amounted around 5 Lakh.
- As the goods price tend to increase, the defaulters tend to decrease.



Visual analysis of goods price on pending credit:

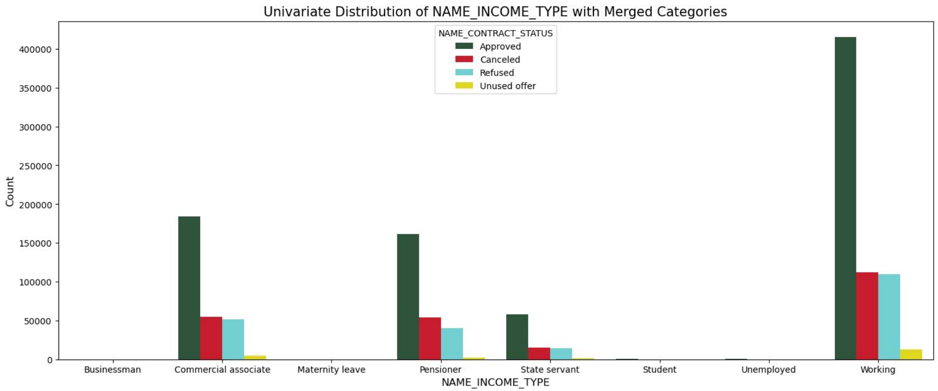
As amount is on the rise (greater than 7.5 Lakh) we can see defaulters overpowering repayers.



Visual analysis of income in previous and current application dataset:

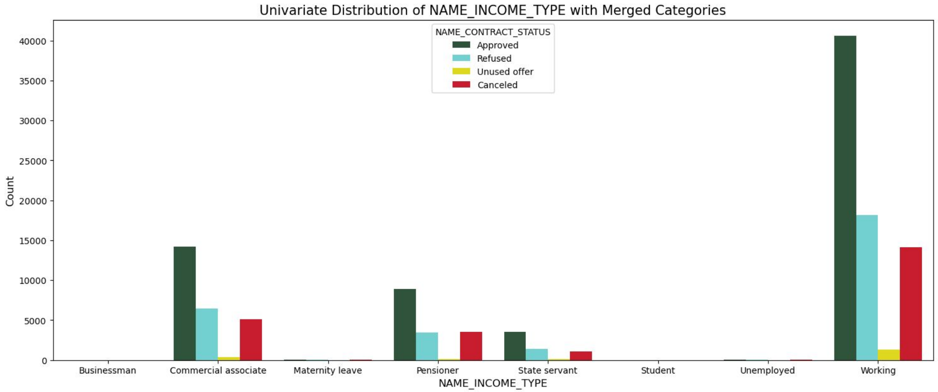
Repayers analysis:

- Most approved loans lie in th working category.
- Least approved loans lie in the state servant category.



Defaulters analysis:

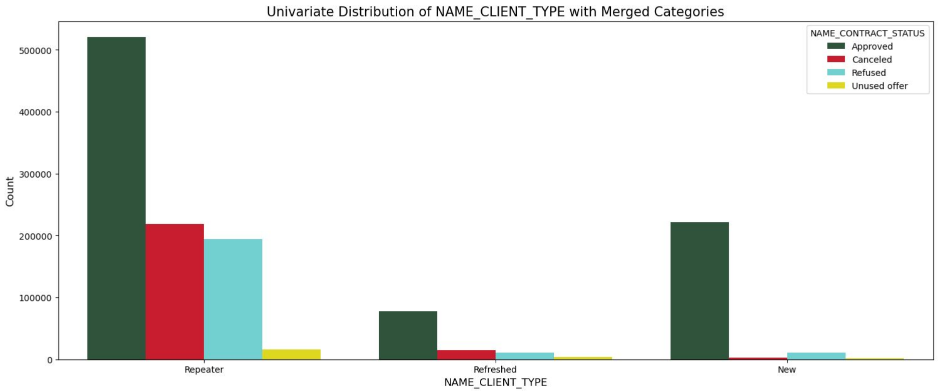
- Most approved and most cancelled loans were from working category.
- Least approved and least cancelled loans were from state servant.



Visual analysis of client in previous and current application dataset:

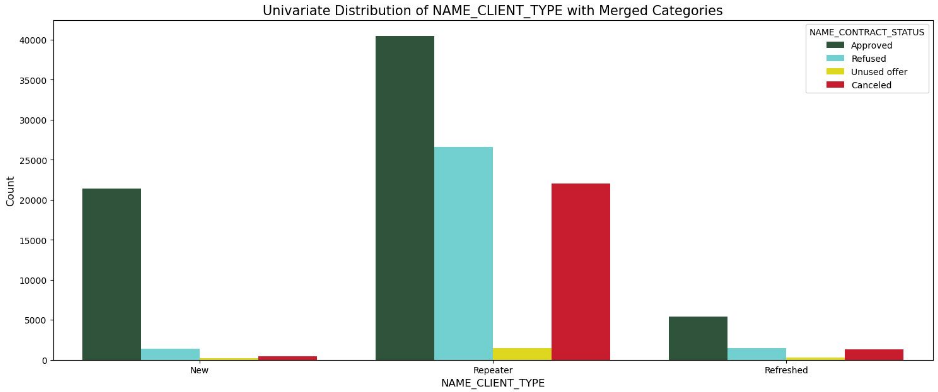
Repayers analysis:

- Most approved loans are from repeater clients
- Least approved are from refreshed clients
- Least cancelled loans are from new clients



Defaulters analysis:

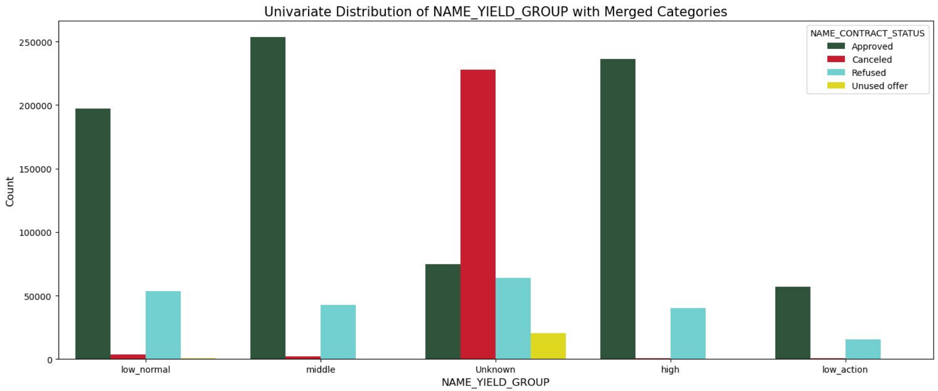
- Most approved loans are from repeater clients.
- Least cancelled loans are from new clients.
- Least approved loans are from refreshed clients.



Visual analysis of yield in previous and current application dataset:

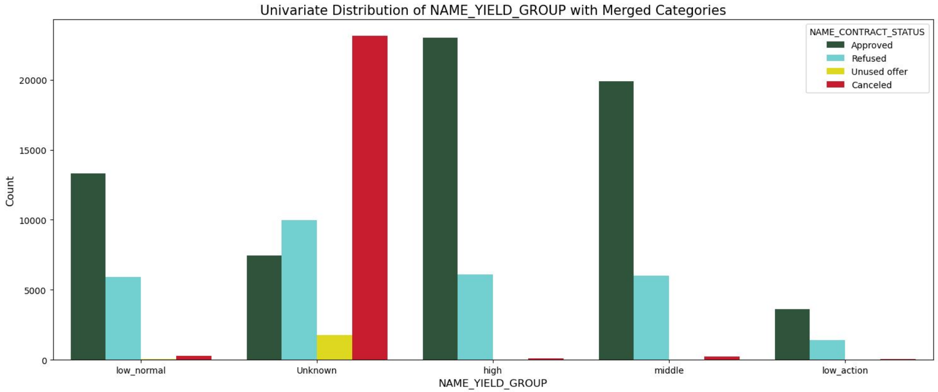
Repayers analysis:

- Most approved loans are which have average yield.
- Least approved loans are which have low yield.
- Most cancelled loans are which had yield in not mentioned.



Defaulters analysis:

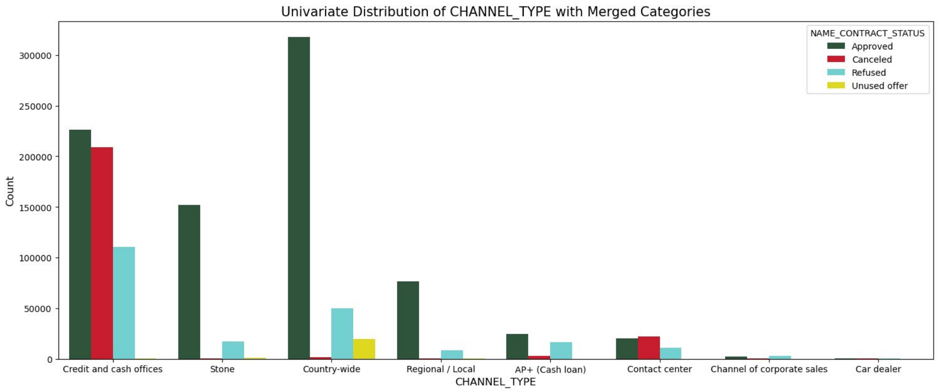
- Most approved loans are from hield yield group.
- Most cancelled loans are from unknown yield group.
- Least approved loans are from low yield group.



Visual analysis of channel in previous & current application dataset:

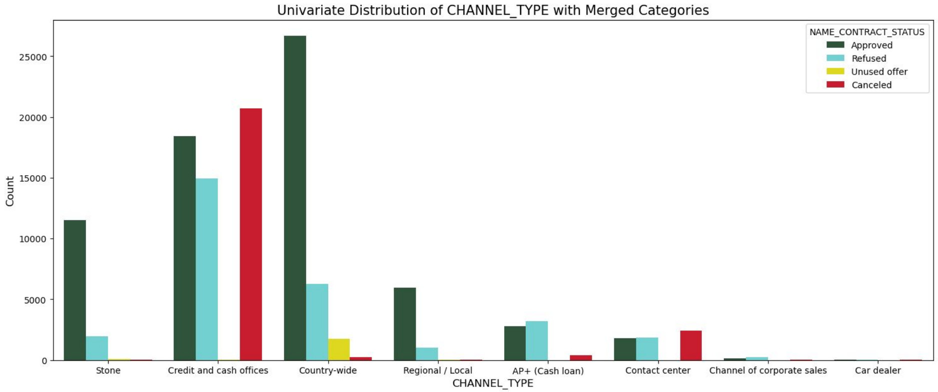
Repayers analysis:

- Highest loan approval is from country wide channel type.
- Lowest loan approval is from car dealers
- Highest cancellation of loans is from credit and cash offices.



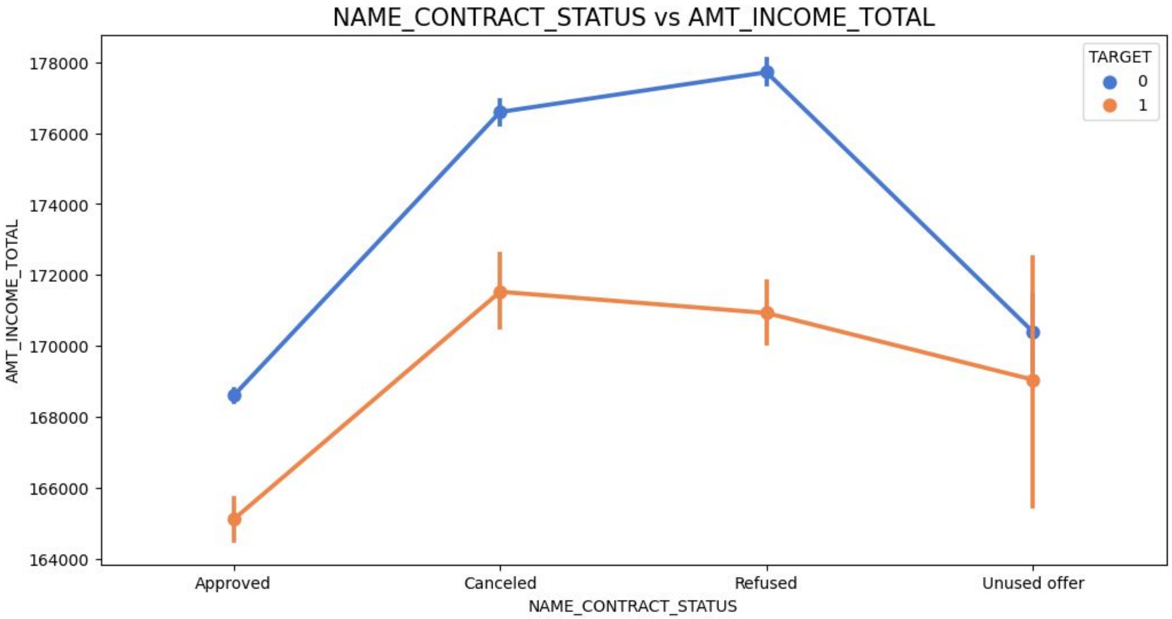
Defaulters analysis:

- Highest loan approvals are from country wide channel.
- Highest cancellations are from credit and cash offices.
- Lowest approvals are from car dealers and channel of corporate sales



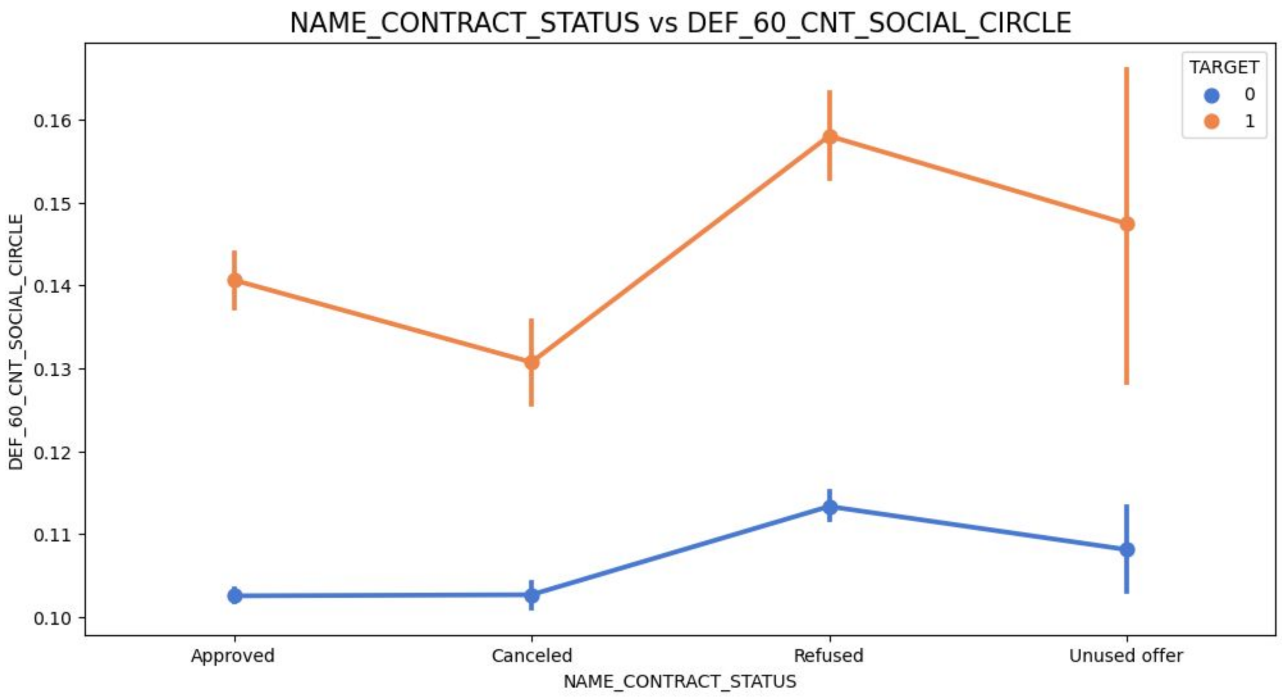
Miscellaneous analysis:

Even after clients that have income in the higher ranges they still are the highest defaulters for majority of the time



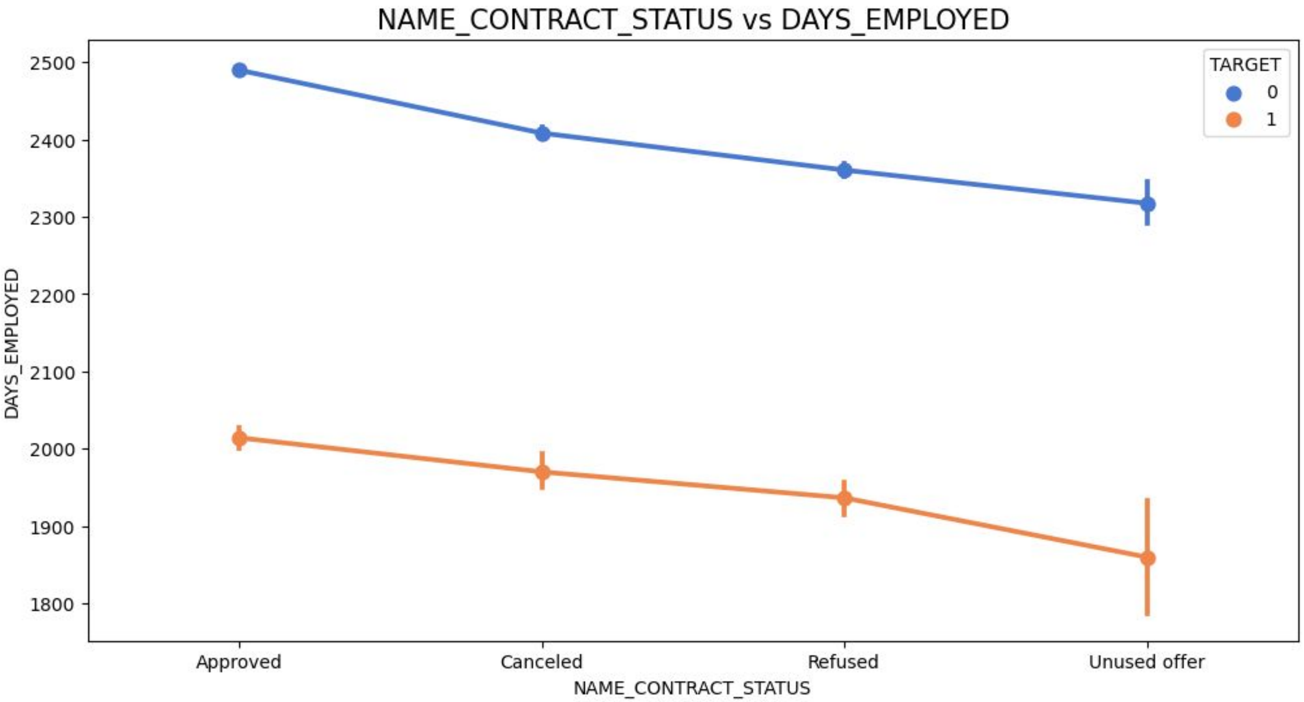
Miscellaneous analysis:

People who have more than 60 days past due defaulters in their area tend to be careful about their repayments



Miscellaneous analysis:

People who are recently employed tend to be repayers but people with older employments have a larger number of defaulters



Conclusion:

- The comprehensive analysis reveals that 91.93% of loan applicants are timely repayers, while 8.07% default on their loans.
- Cash loans experience a slightly higher default rate (8.3%) compared to revolving loans (5.5%).
- The risk of defaulting does not show a clear pattern concerning the amount of pending credit, with most defaulters having pending credit amounts between 2.5 to 7.5 lakh.
- Gender-wise, males have a higher default rate (10.1%) compared to females (7%).
- Educational background impacts default rates, with individuals holding lower secondary education posing the highest risk (10.9%), whereas those with academic degrees are least likely to default (1.8%).
- Employment status is a significant determinant, with the highest default rates observed among applicants on maternity leave (40%) and unemployed applicants (36.4%). Yet, students and businessmen exhibit the lowest risk, having no record of default.
- Income analysis suggests that higher income applicants are generally more reliable, with defaulters typically earning less.
- Age-wise, older applicants (65+ years) and the very young (21-22 years) show lower default rates.
- However, certain correlations, such as between pending credit amount and goods price, suggest that higher loan amounts are sought for more expensive goods, indicating potential risk factors.
- Despite these findings, no definitive pattern emerges to predict default based on factors like family status or income alone, underscoring the complexity of assessing loan default risk.

Conclusion:

- **Factors if an applicant will be Repayer:**

  Higher total income emerges as a decisive factor, indicating a positive correlation between income level and the likelihood of being a repayer.

  Applicants with stable employment and higher income demonstrate a more reliable repayment pattern.

- **Factors if an applicant will be Defaulter:**

  The analysis points to various risk factors contributing to loan default, including lower income levels, certain employment statuses (e.g., being on maternity leave or unemployed), and educational background, with lower secondary education posing higher default risks.

  These factors collectively indicate a multifaceted set of circumstances influencing the likelihood of being a defaulter.

Conclusion:

- **Factors to mitigate any default risk leading to business loss:**

  Applicants with higher pending credit amounts, particularly in the range of 2.5 to 7.5 lakh, pose a potential risk. To mitigate this risk, loans could be considered for such applicants, but with the imposition of higher interest rates to offset potential default losses.

  Additionally, caution should be exercised when dealing with applicants on maternity leave or those currently unemployed.

- **Suggestions:**

  Offering lower interest rates to applicants with strong financial profiles and higher rates to those with riskier profiles may help balance business objectives with risk mitigation. Conducting a comprehensive review of employment and income stability, along with educational background, will provide a more nuanced understanding of an applicant's creditworthiness. Additionally, exploring targeted financial education initiatives for applicants with higher default risks could contribute to better financial decision-making and potentially reduce default rates.

Thank You