

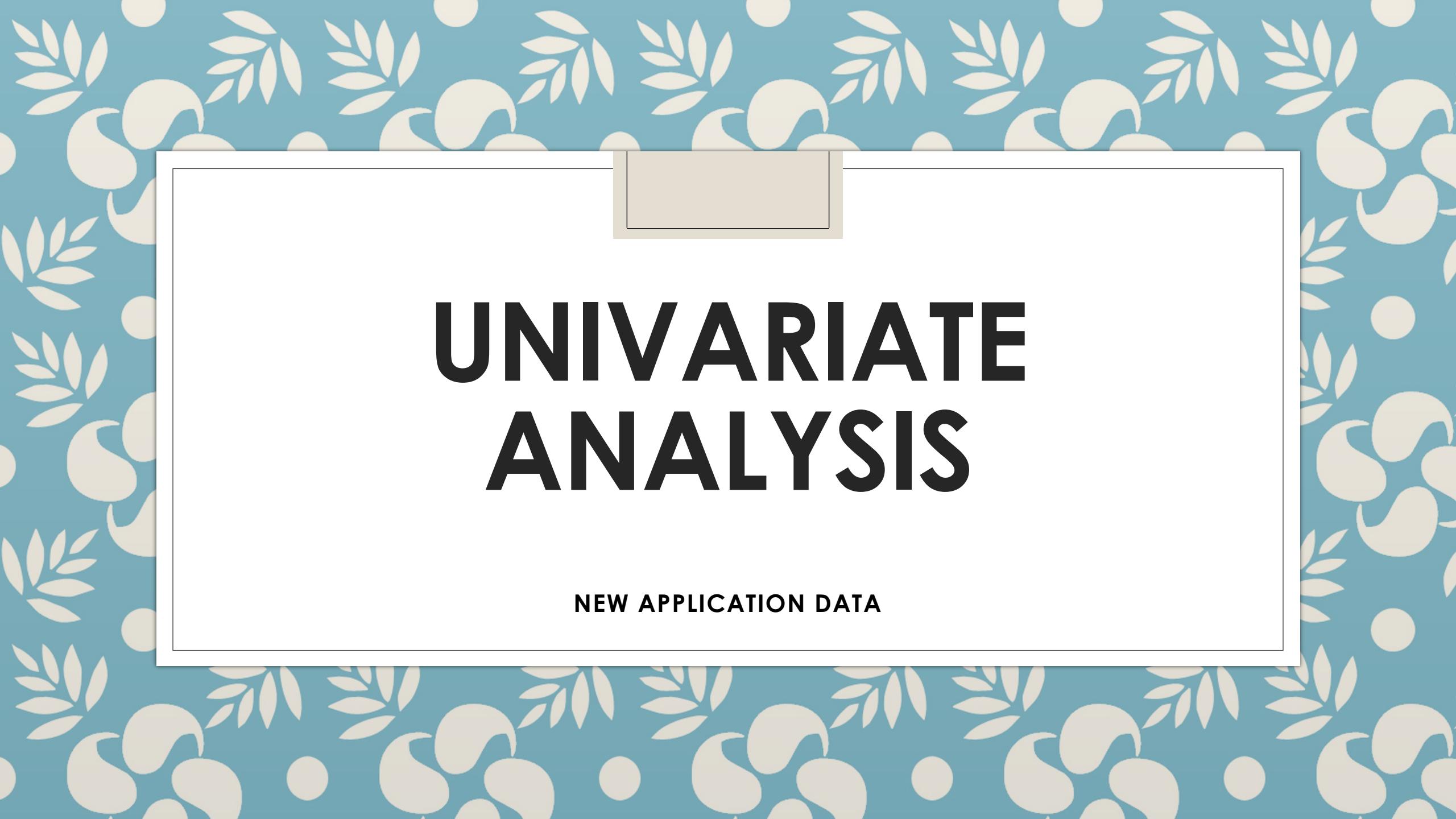
CREDIT EDA

CASE STUDY

UMANG SHUKLA

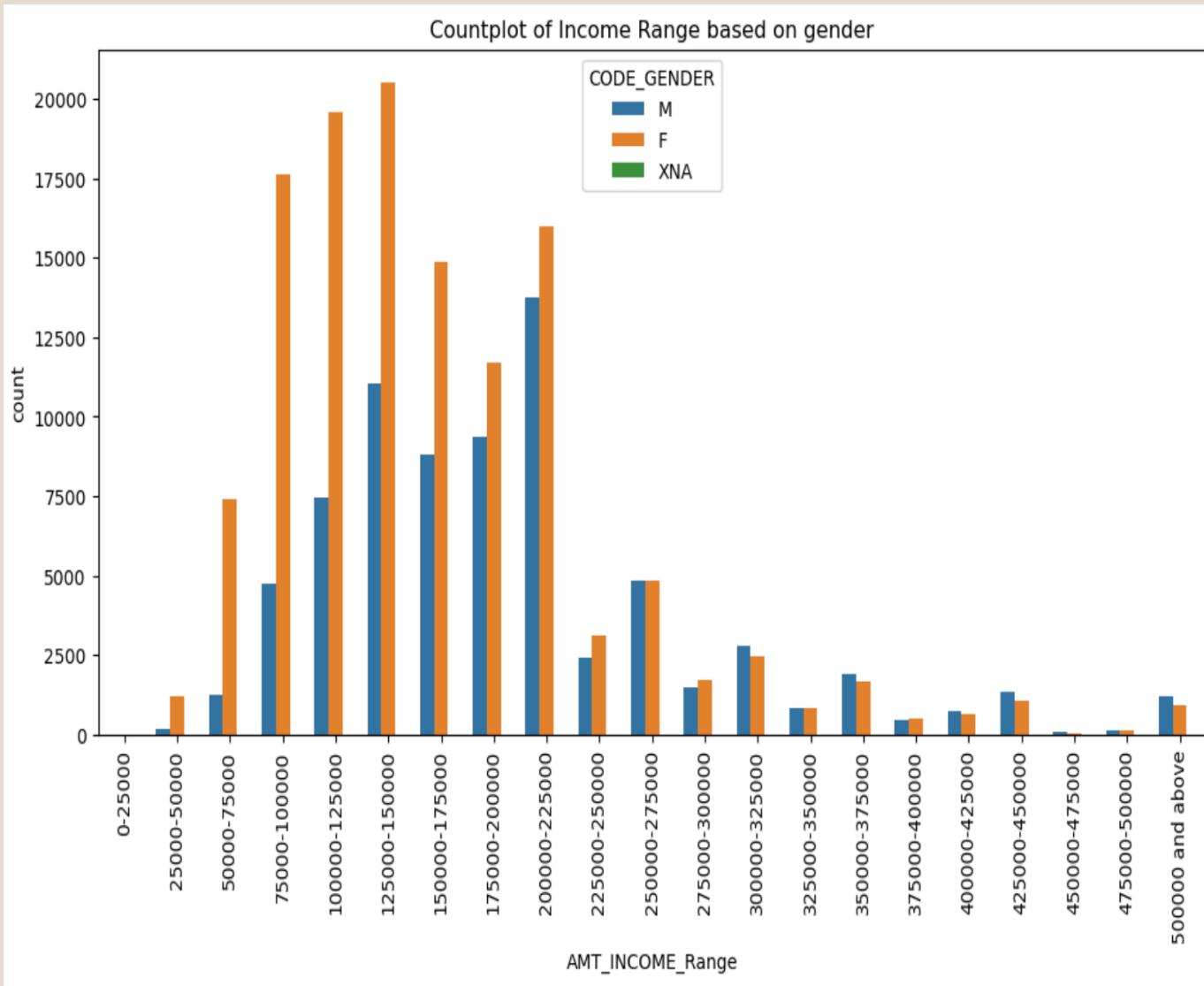
Problem Statement

- Use EDA to analyse and find insights present in the dataset to make sure that the applicants capable of repaying the loan are not rejected.
- When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:
 1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
 2. If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.
- The company wants to understand the driving factors behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

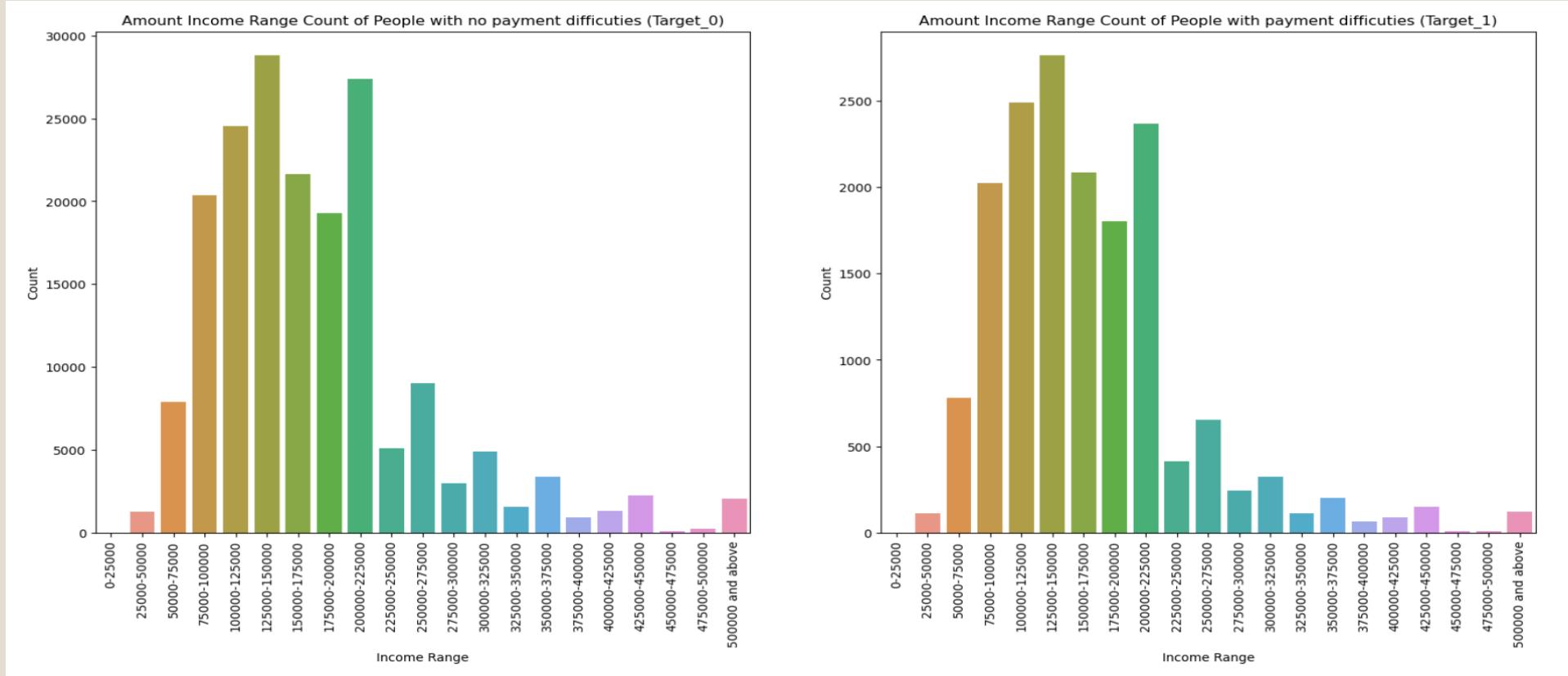


UNIVARIATE ANALYSIS

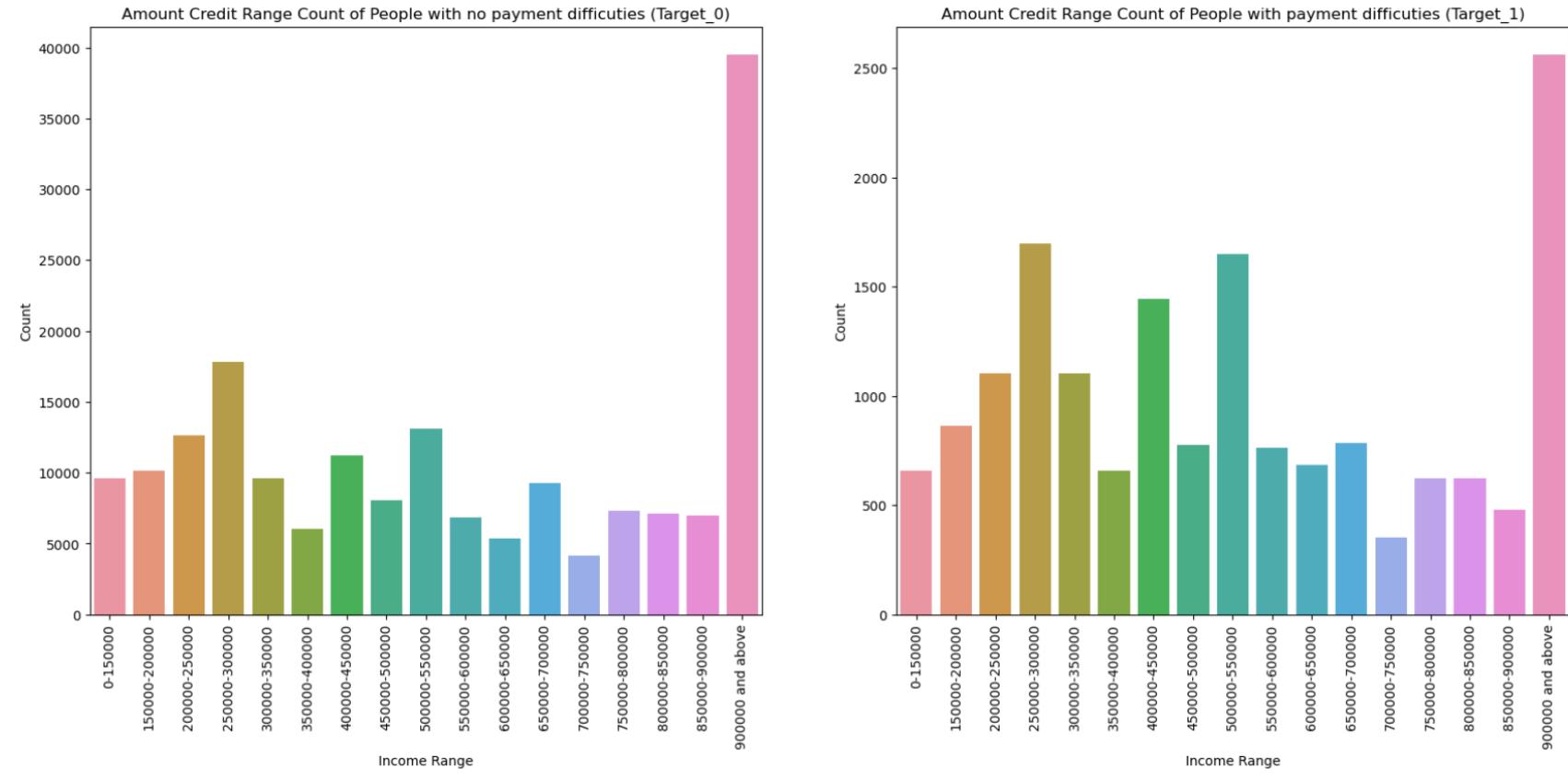
NEW APPLICATION DATA



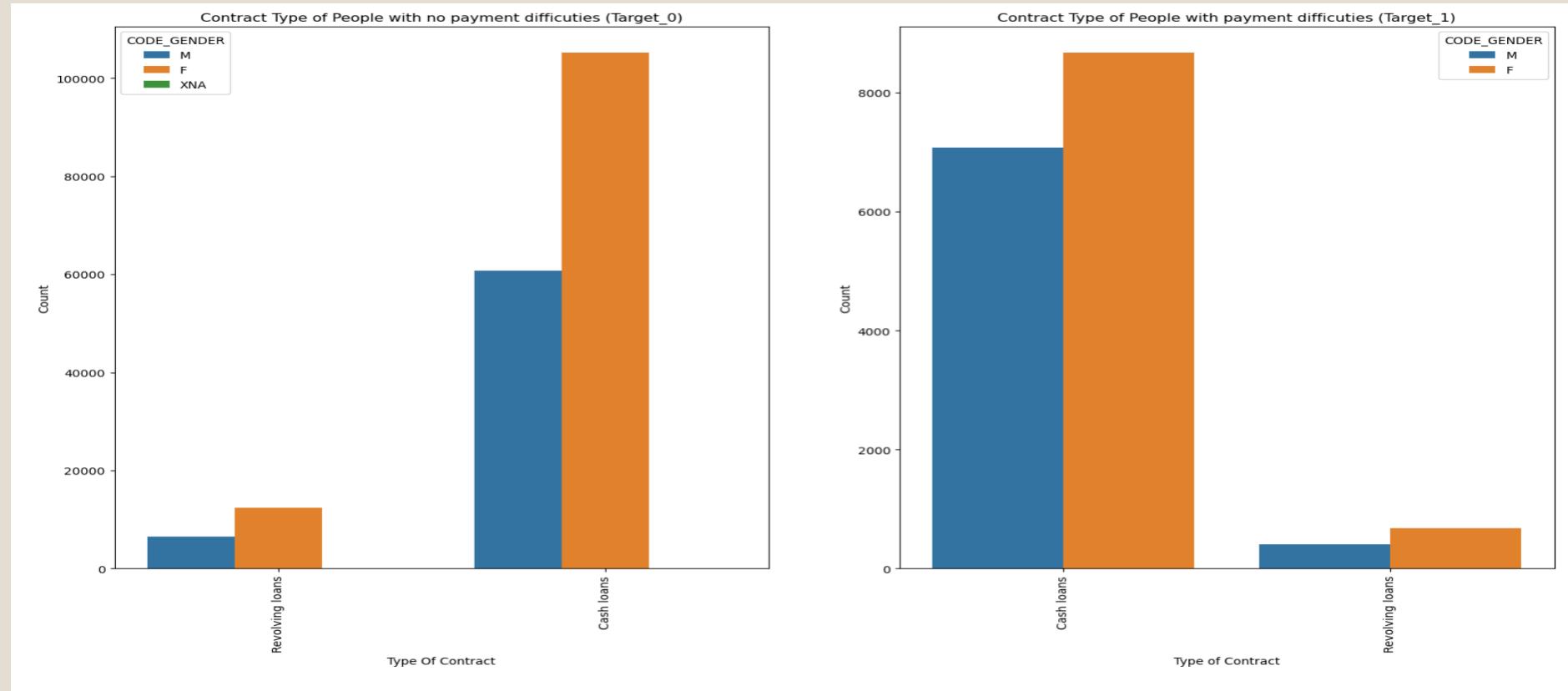
- We can see in the starting range from 25000 to 250000 the income range of females are higher than male.
- After 250000 we can see an equal range of income between male and female and in rise in some ranges.
- We can see a drop in count after range 200000-225000.



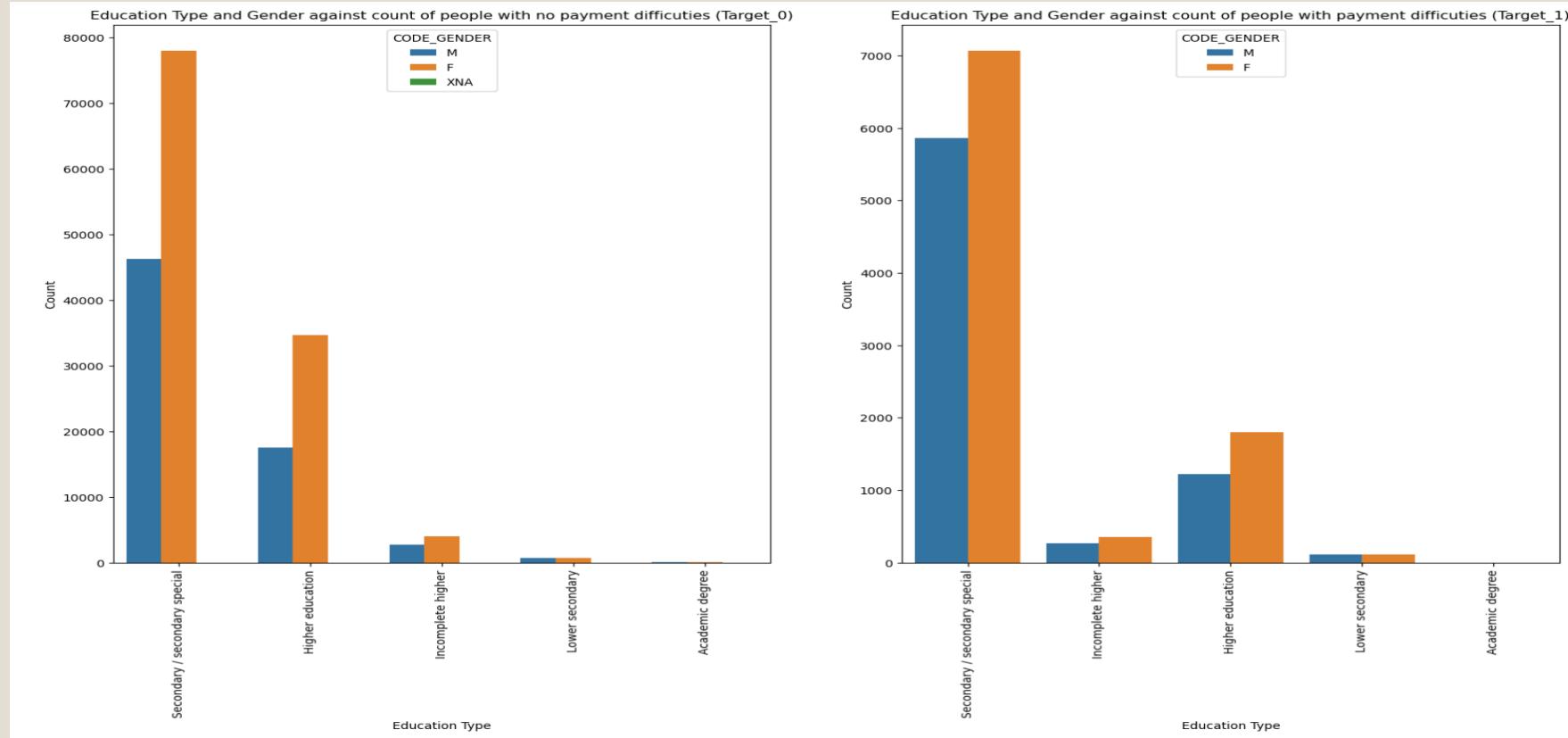
- Income range for both the target variables are similar .
- The income range 100000-125000 have the highest count for both the target variables.



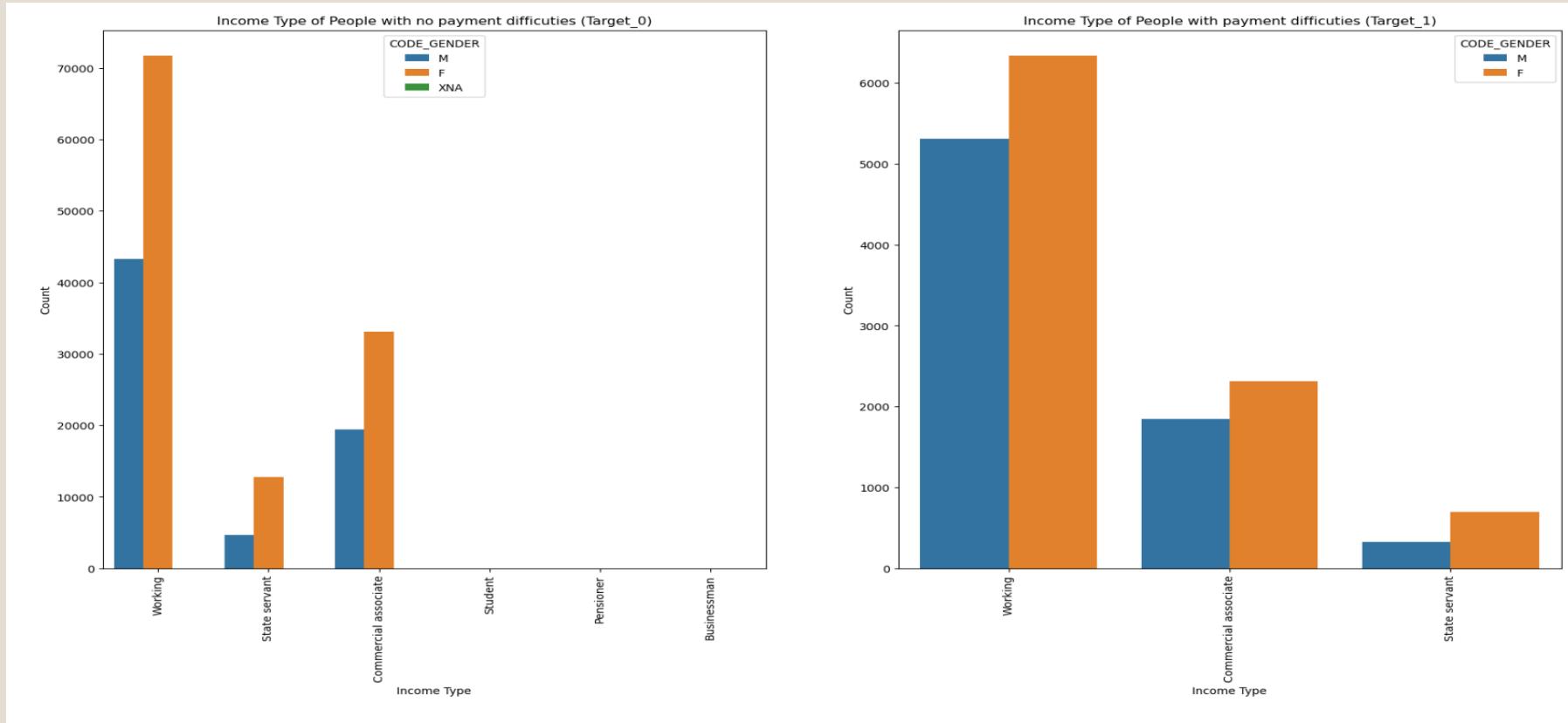
- If we see the Loan Amount Credit-Range target_0 have higher number of amount credit from 200000-550000 and target_1 have lesser number of amount credit in the same range, it means both the targets have maximum number of requirement of loan in this range.
- Credit-Range of 900000 and above have the highest number of requirement in both target variables.



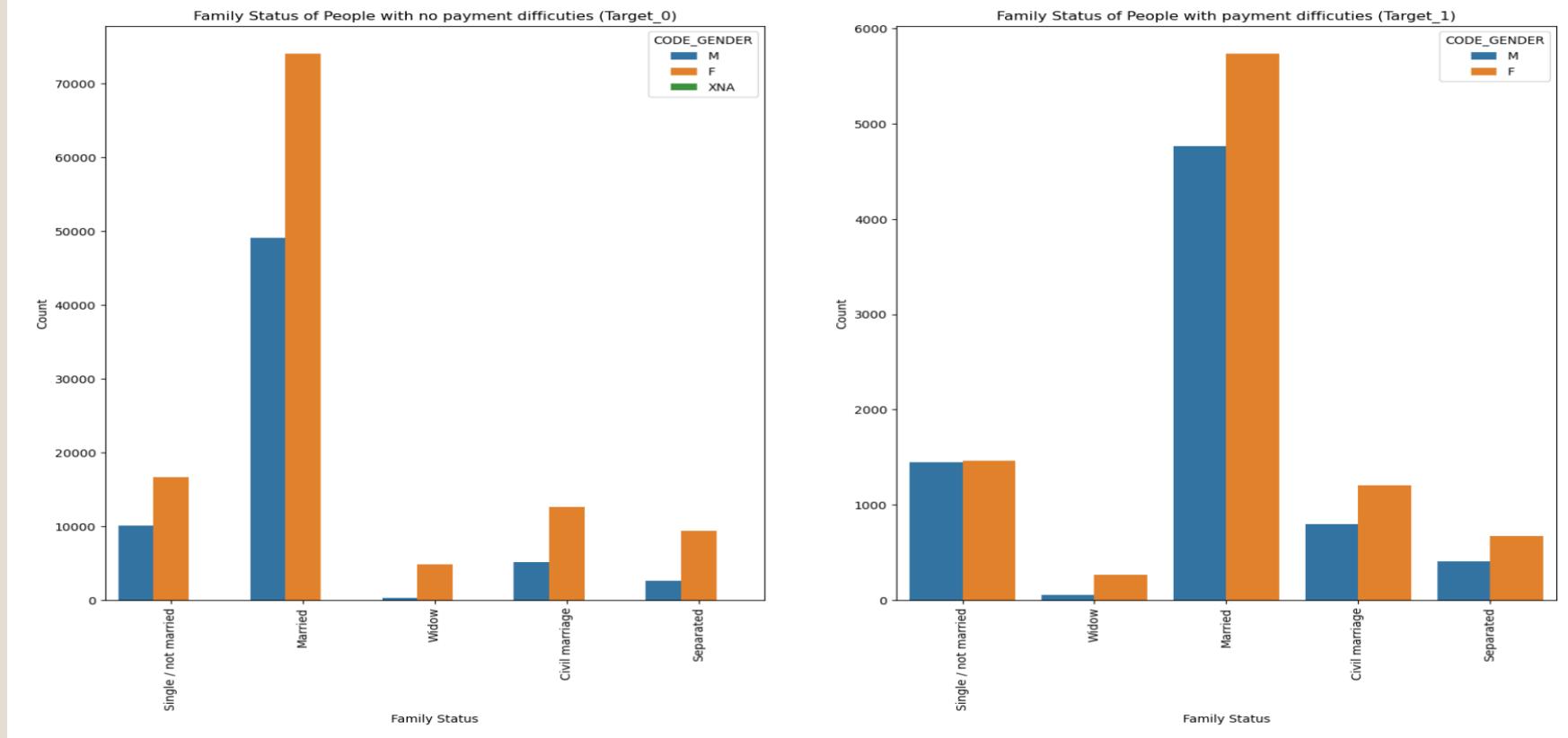
- There is a high requirement of cash loans for both target variables irrespective of the gender.
- We can see that the count of males for cash loans against female with no payment difficulties is less when compared to ratio of count of cash loans for people with payment difficulties.



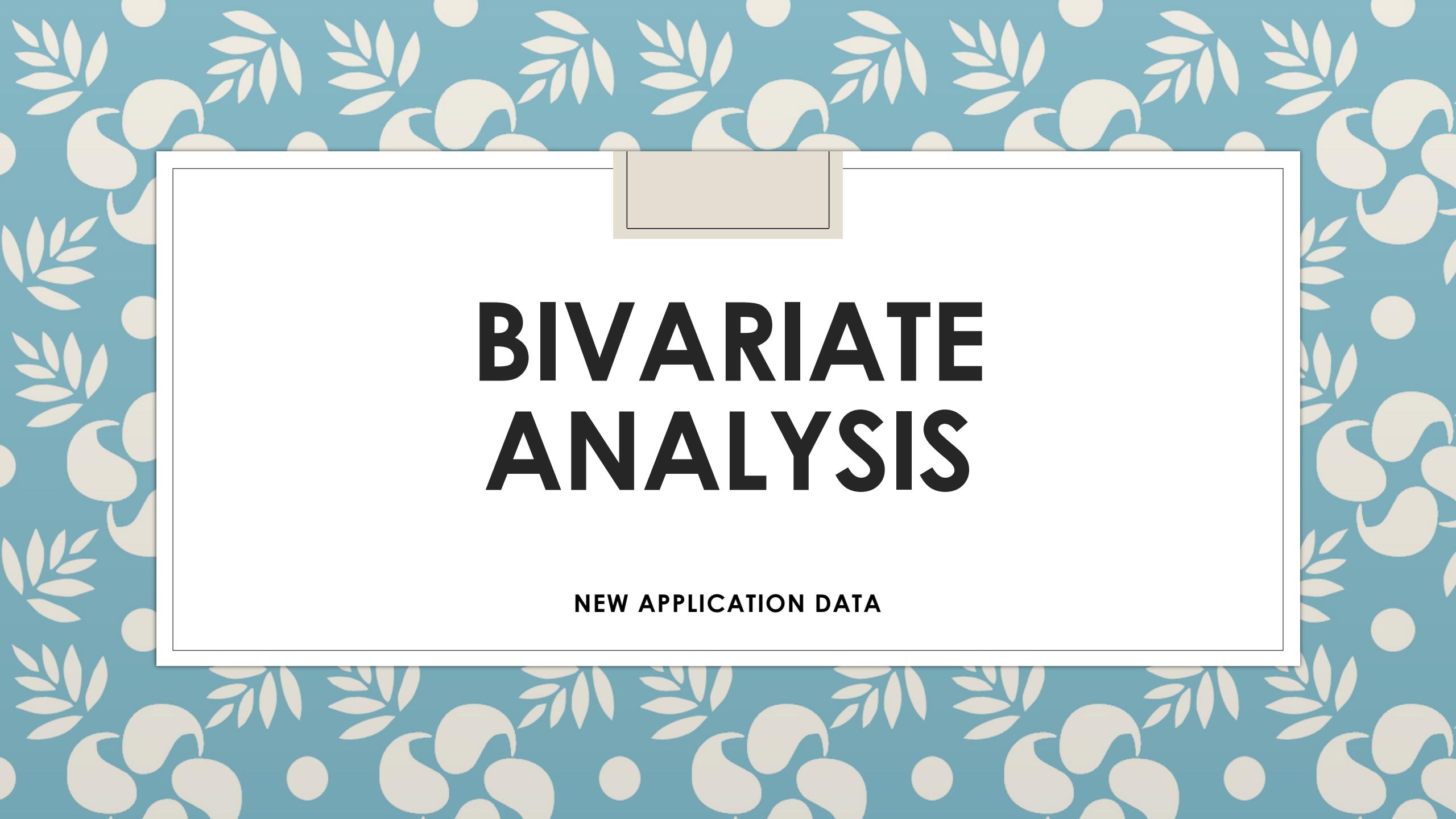
- Education type secondary/secondary special has the highest count for both target variables.
- The count of females in all the education type is higher than males for both target variables.
- The difference between male and female in target_1 is less than the difference in target_0.



- Working professionals have the highest count for both the target variables(target_0 and target_1).
- Count of females for every income type is higher than males for both the target variables(target_0 and target_1).
- There is an absence of Student , Pensioner and Businessman in target_1.

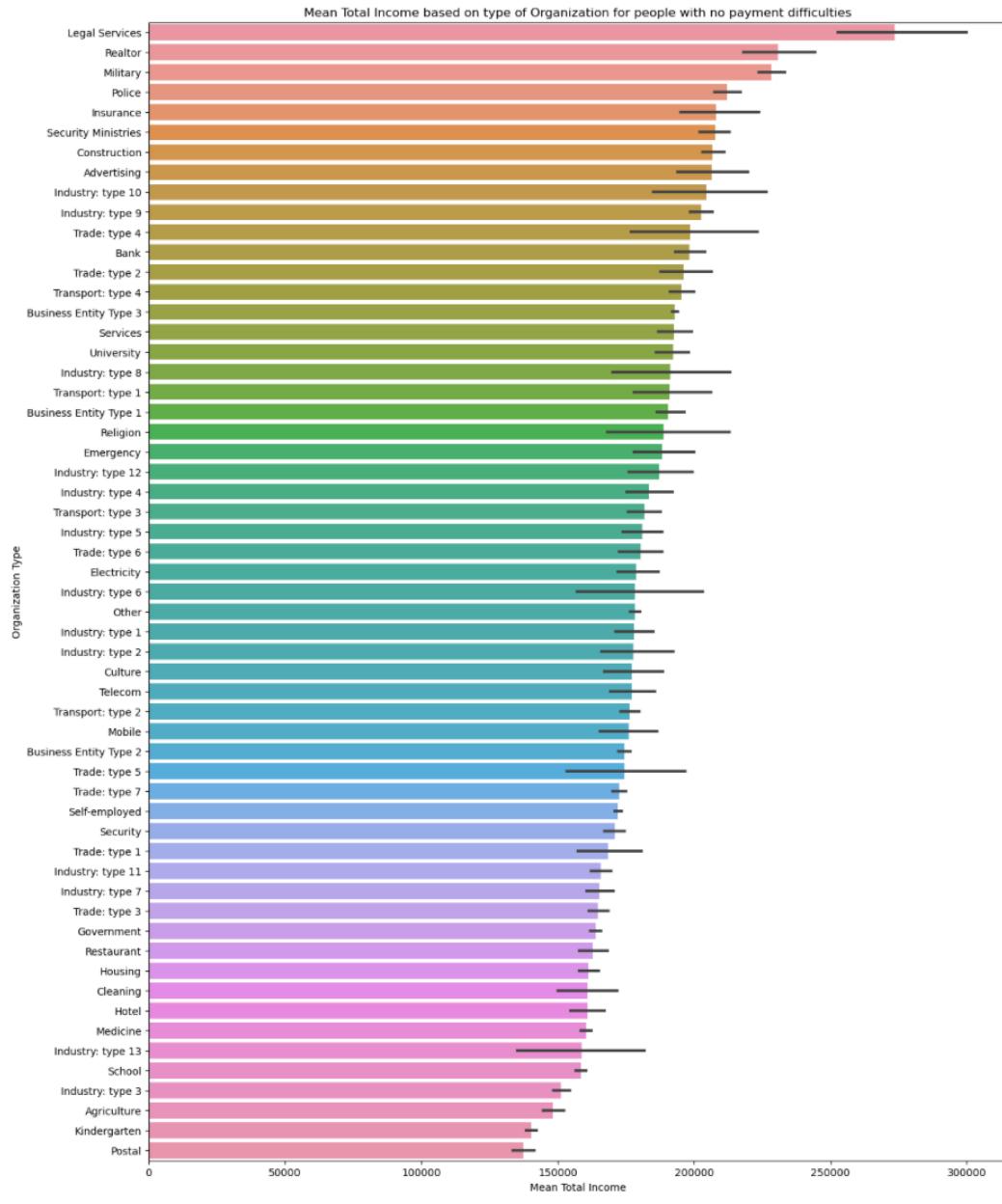


- The count of married people as family status have the highest count for both target variables
- Widowed people have the lowest count for both target variables irrespective of the gender
- Single/not married people have the same ratio of male and female with payment difficulties (target_1)

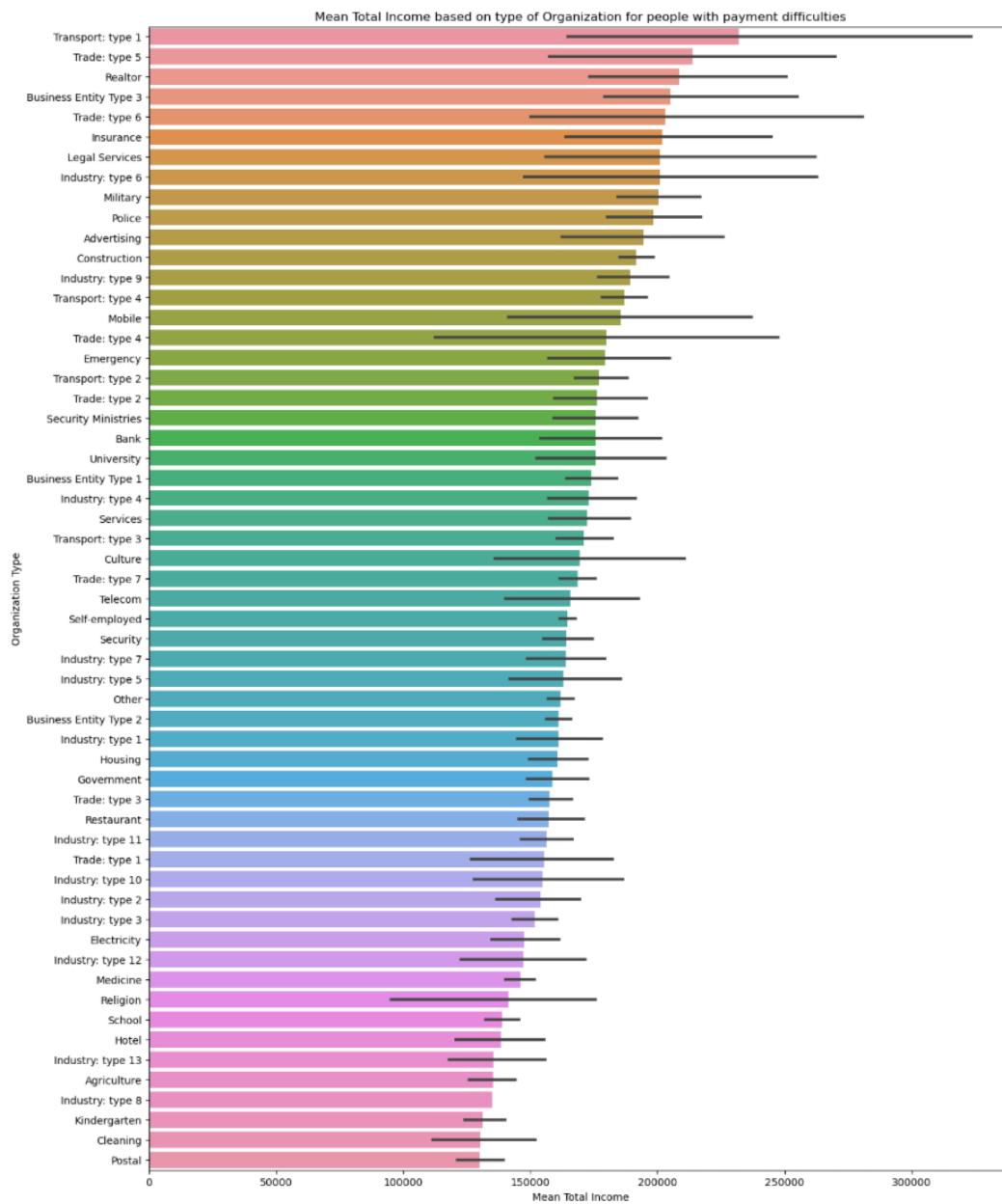


BIVARIATE ANALYSIS

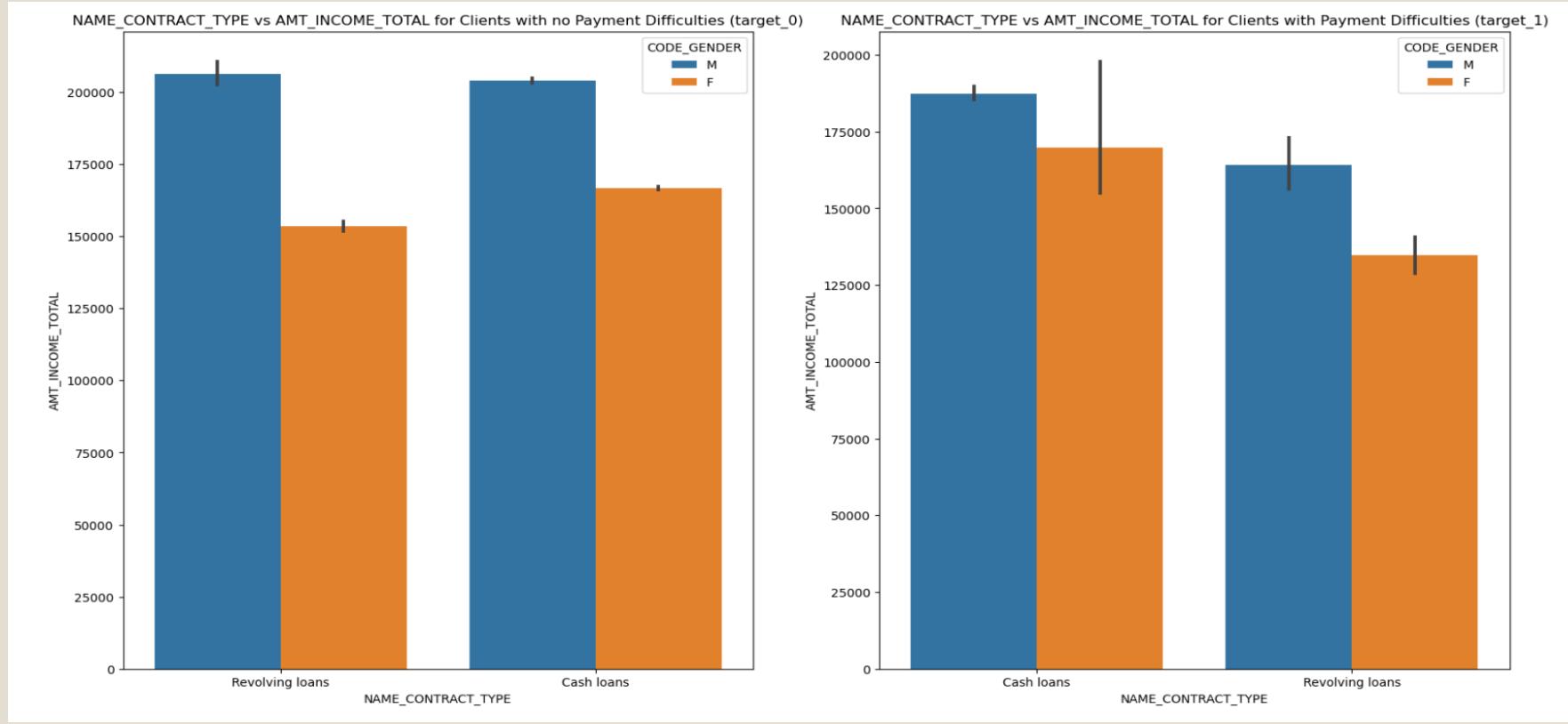
NEW APPLICATION DATA



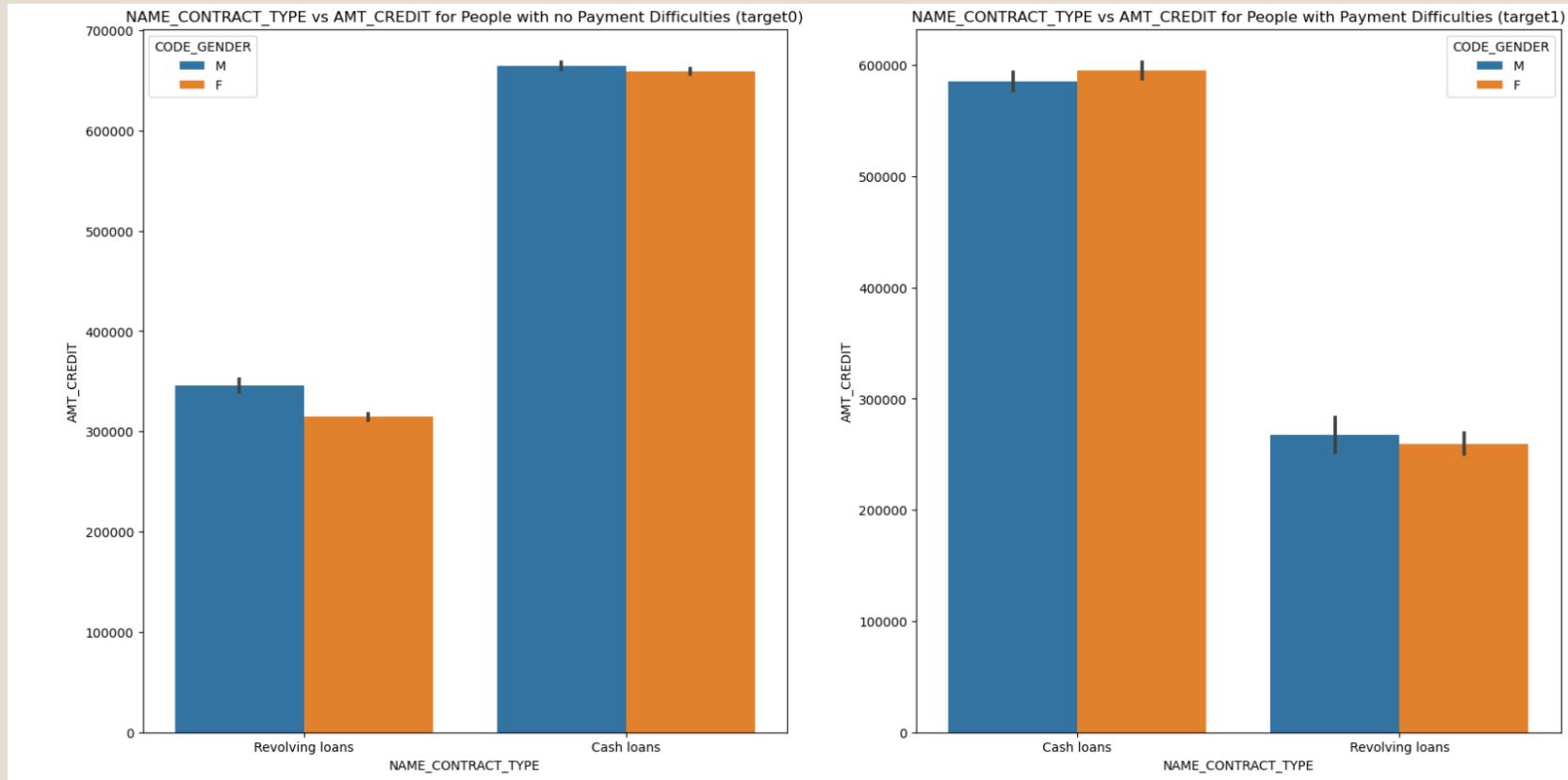
- Clients with organization type Legal Services have the highest average total income for people with no payment difficulties (target_0) around 275000.
- Clients with organization type Postal have the lowest average total income for people with no payment difficulties(target_0) around 140000.
- Although in the previous plots the organization type business entity type 3 has the highest applicants the average income of it is not the highest.



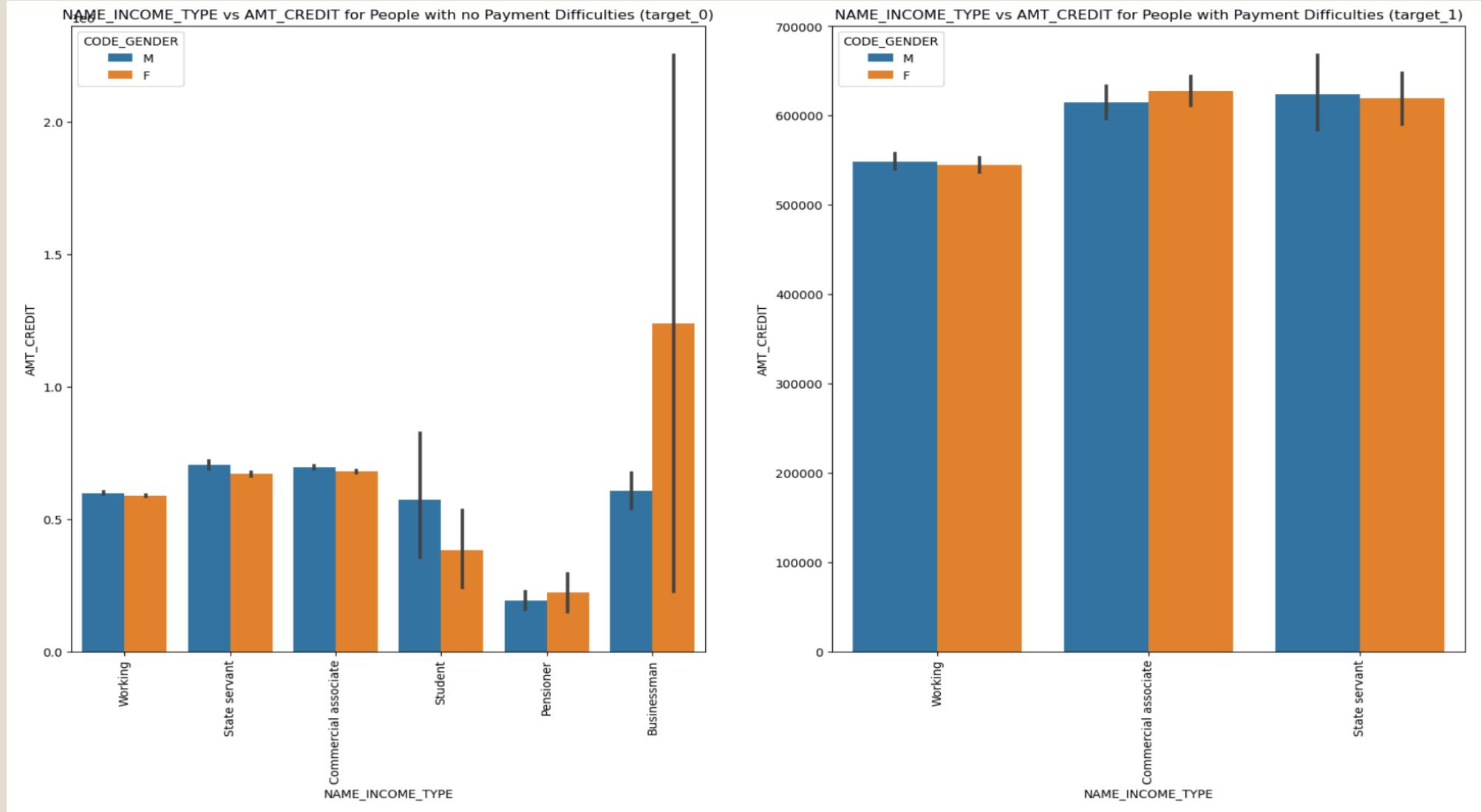
- Clients with organization type Transport Type 1 have the highest average total income for people with no payment difficulties (target_1) around 225000.
- Clients with organization type Postal and Cleaning have the lowest average total income for people with no payment difficulties(target_1) around 135000.
- In target_1, Transport Type 1 has the highest total income amount, whereas in target_0, Legal Services are leading in terms of total income



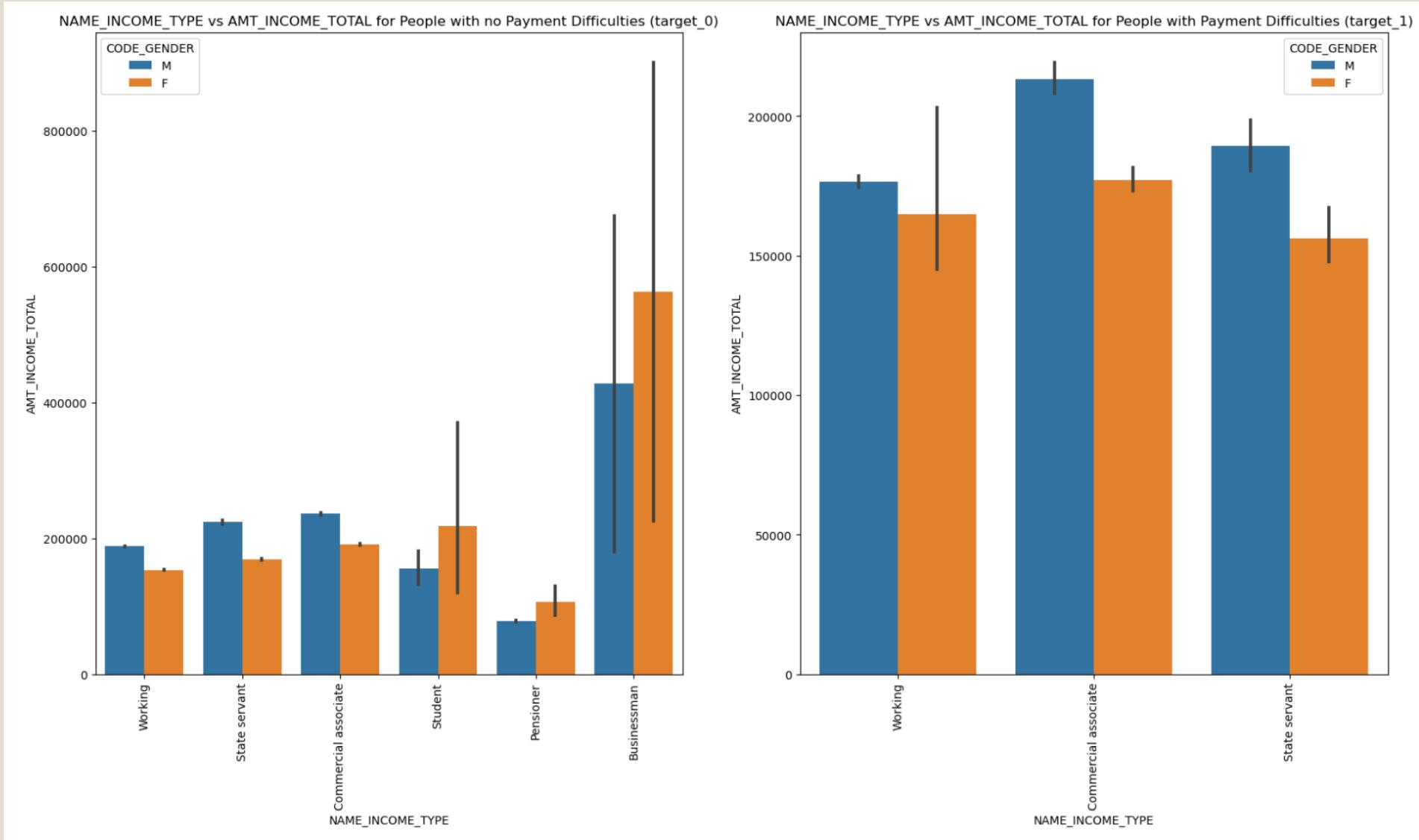
- We can see nearly equal contract type requirement for male and female for clients with no payment difficulties(target_0).
- The income total for clients is nearly the same for target_0 variable irrespective of the loan type.
- The difference between total income applicants of cash loans and revolving loans for target_1 is higher than target_0



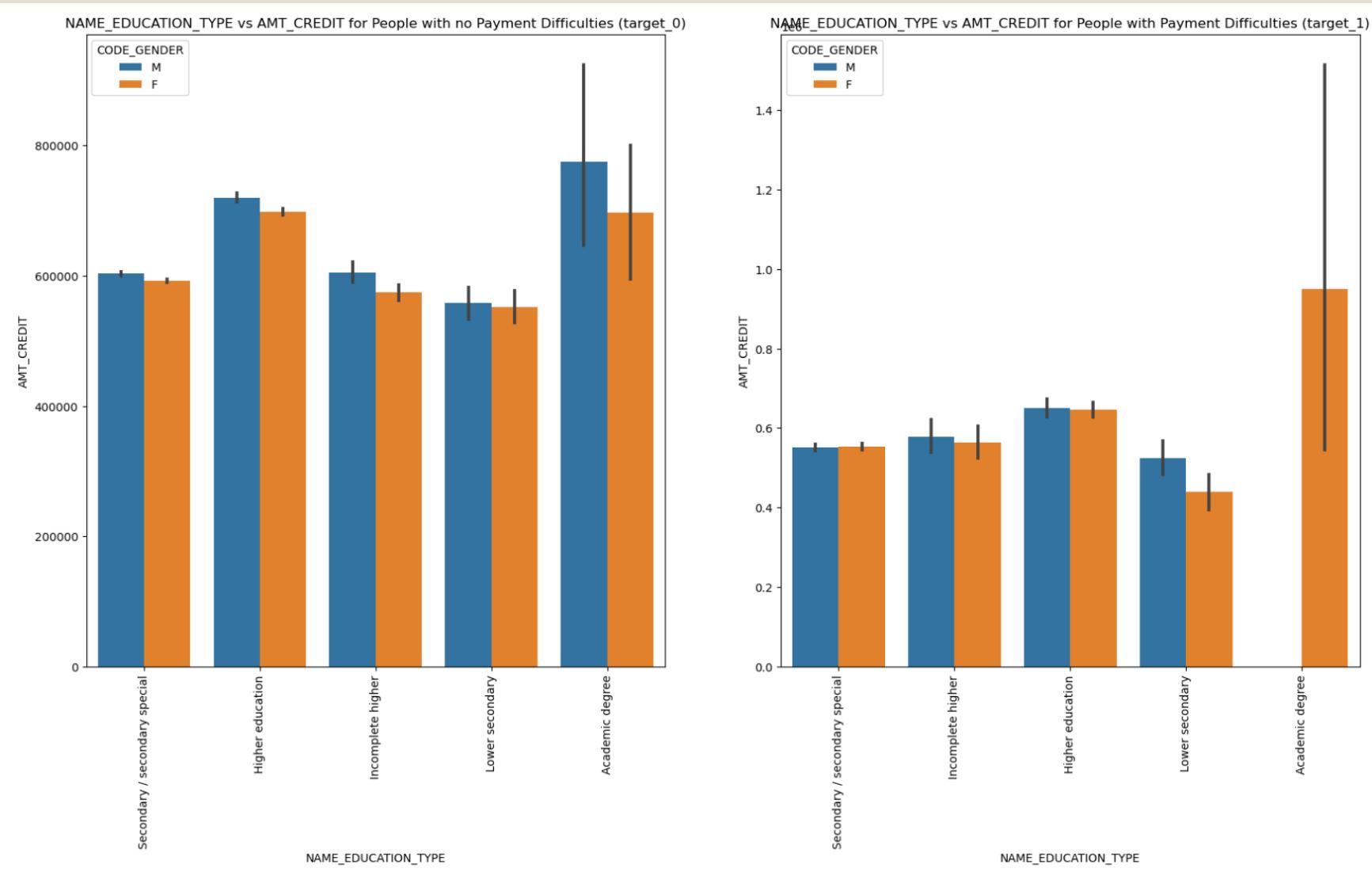
- Cash loan applicants have higher credit in both the target variables.
- Male are usually provided higher credit amount with no payment difficulties for both loan types.
- The ratio of count of applicants for both the target variables for both the loan type is nearly the same



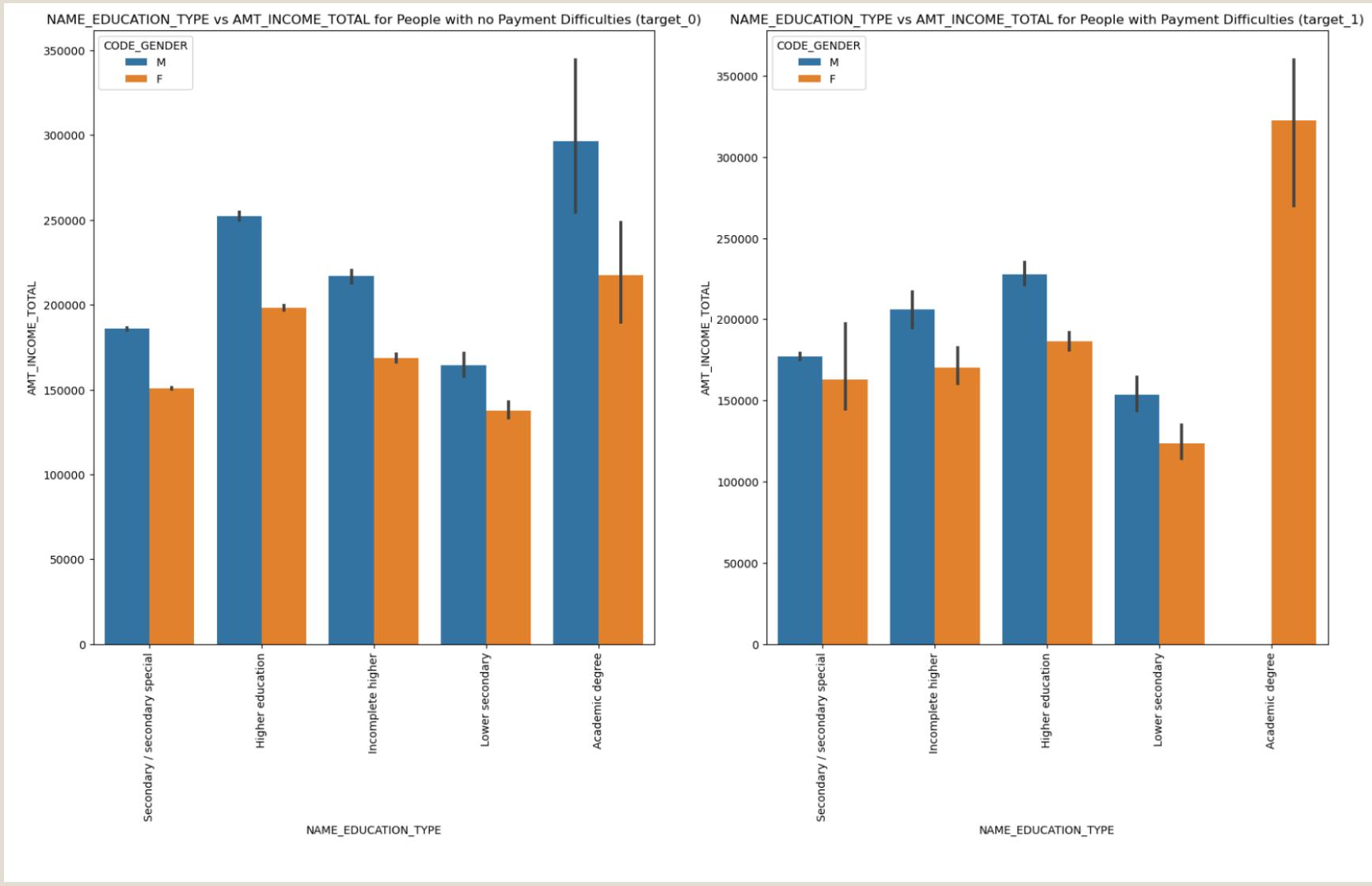
- Target_0 - Businessman income type has the highest total credit amount for female which is unique only in target_0
- Target_0 -Commercial associate and State Servant income type has the highest credit amount for male
- Target_0 -Pensioners income type have the lowest credit amount for female
- Target_1 - Commercial associate and State Servant income type has the highest credit amount for both genders with nearly same value
- Target_1 - Working income type have the lowest credit amount for both genders
- Working and State servant are mutual in both target variables



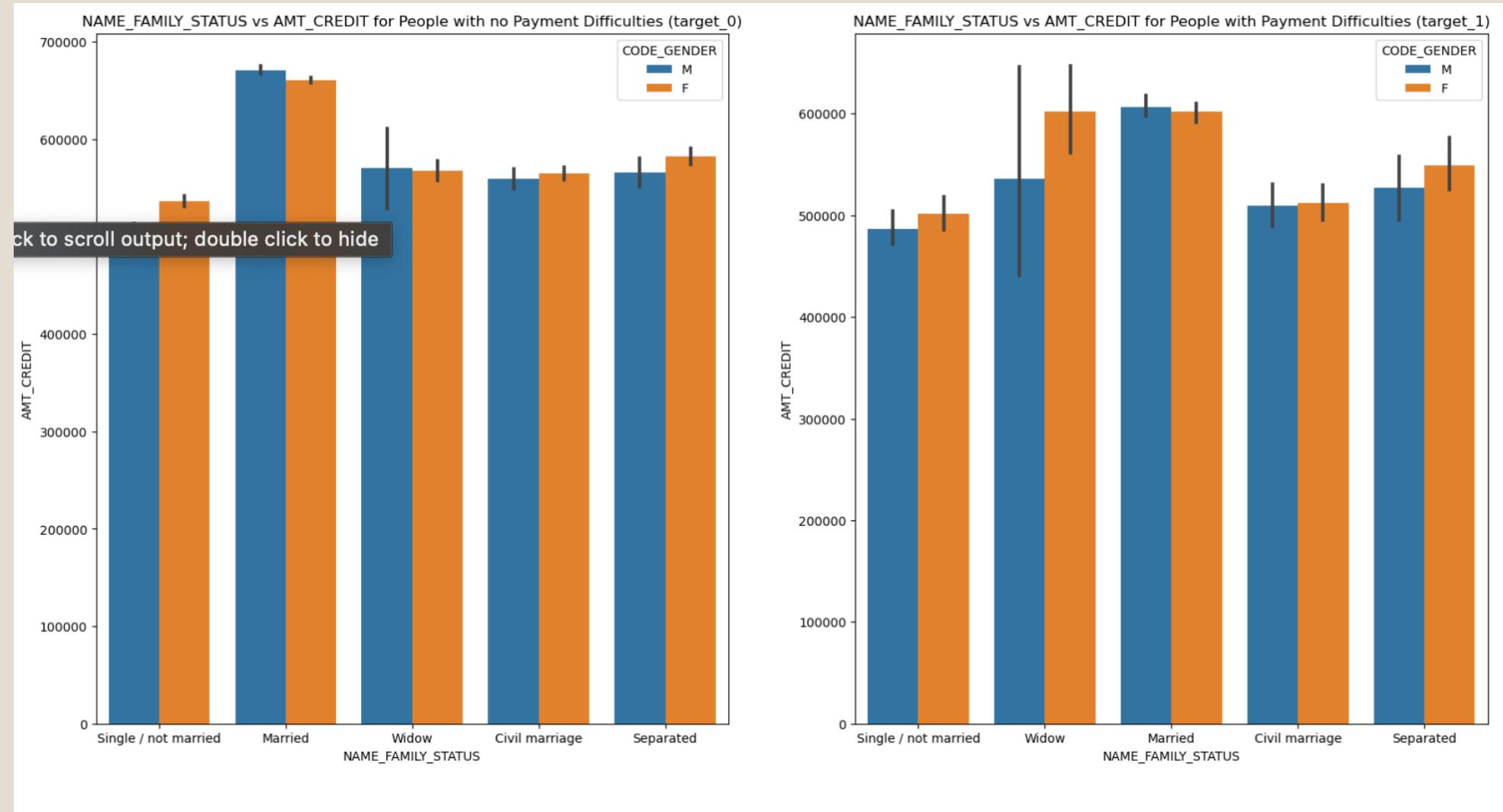
- Target_0 - Businessman income type has the highest total income for both genders which is unique only in target_0
- Target_1 - Commercial associate income type has the highest total income for both genders which is unique only in target_1
- Target_1 - Working income type have the lowest annuity amount for male and State servant for female
- Working and State servant are mutual in both target variables



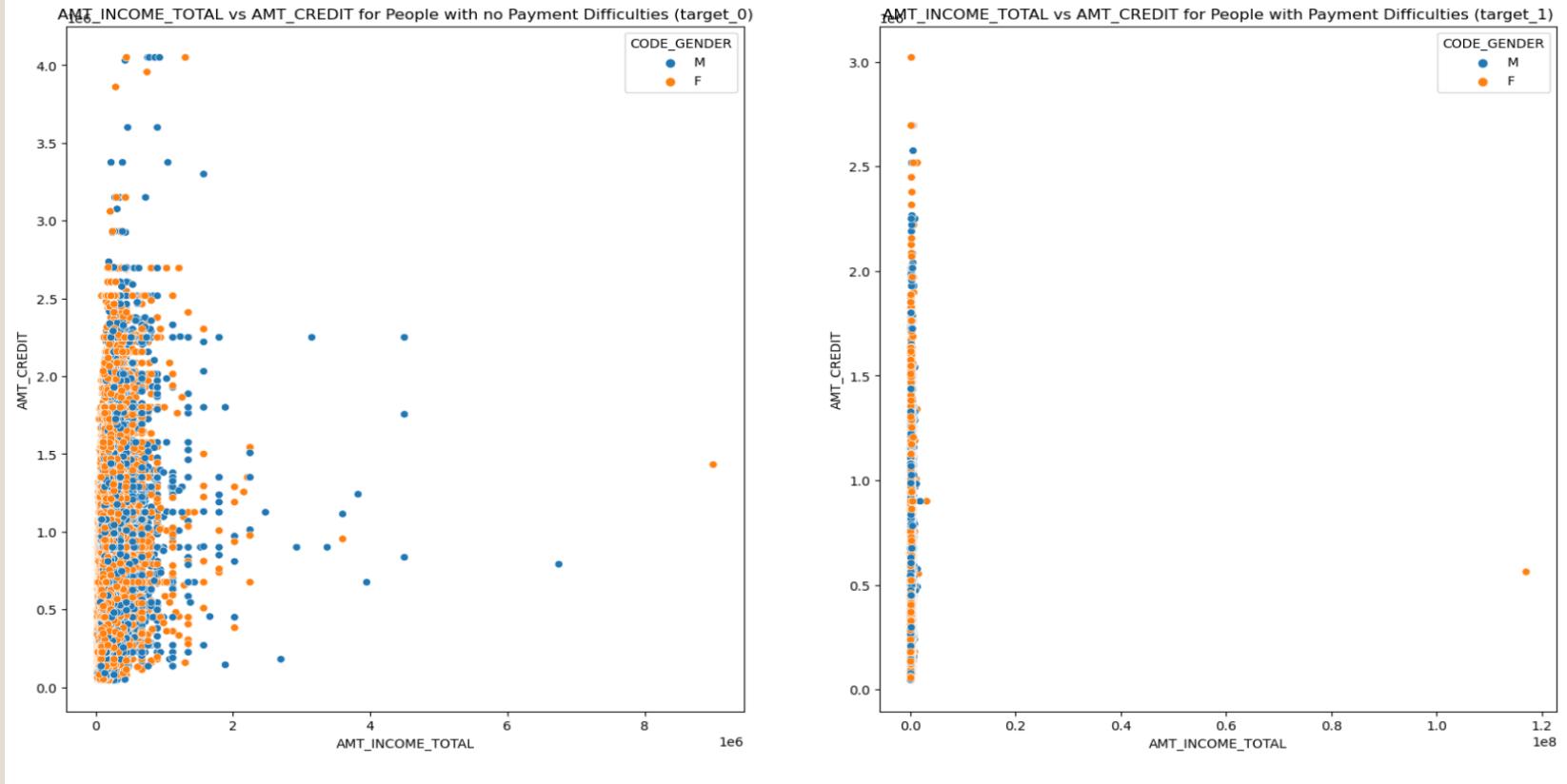
- Target_0 -Academic degree has the highest credit amount for male
- Target_0 -Higher Education and Academic degree type has the highest credit amount for female with nearly same values
- Target_0 -Lower secondary have the lowest credit amount for both genders
- Target_1 - Higher education has the highest credit amount for male with almost same value with female
- Target_1 -Academic degree has the highest credit amount for female
- Target_1 -Lower secondary have the lowest credit amount for both genders
- Target_1 -There is an absence of male people for Academic degree



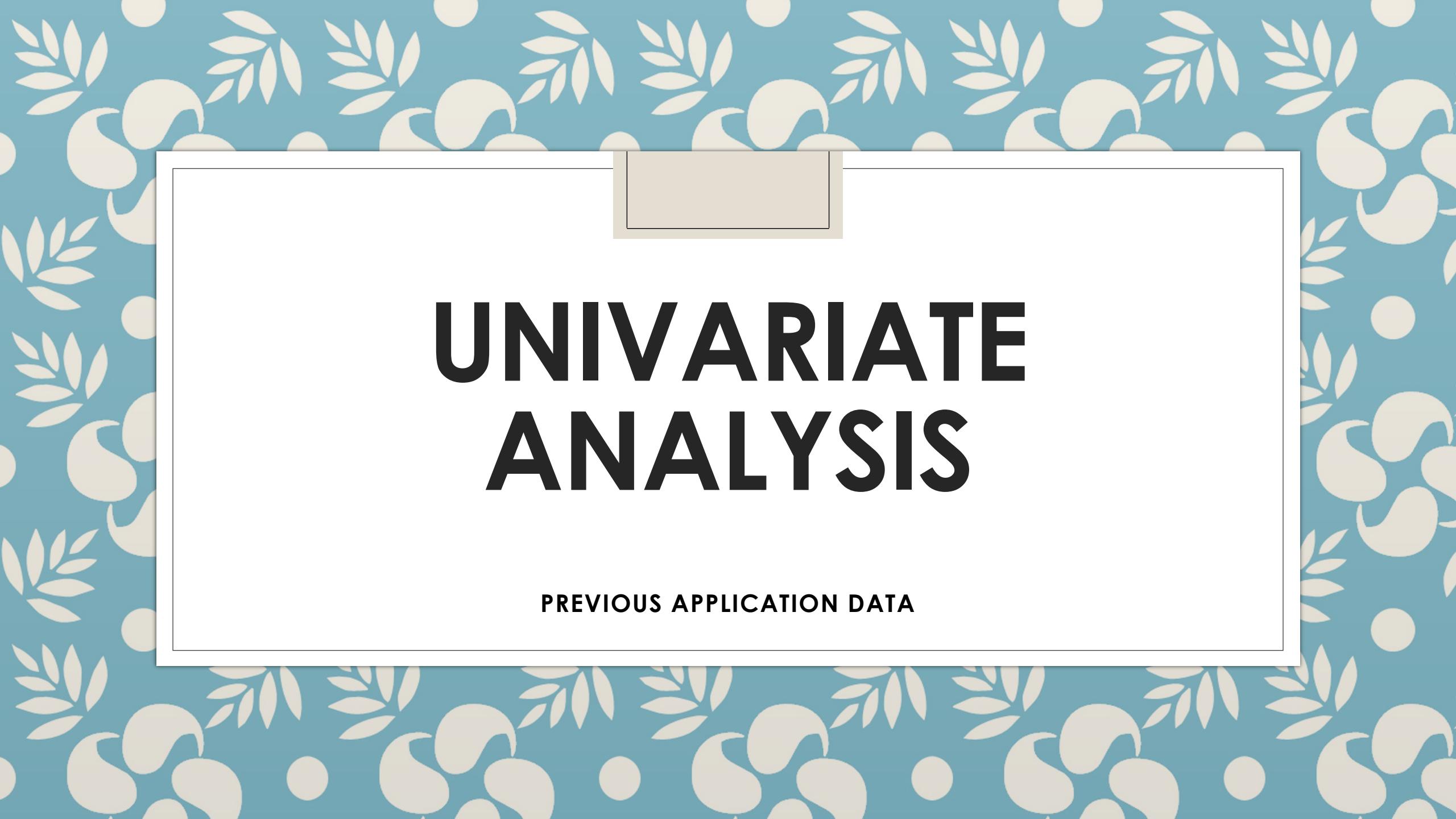
- Target_0 -Academic degree has the highest total income for both genders
- Target_0 -Lower secondary have the lowest total income for both genders
- Target_1 - Higher education has the highest total income for male
- Target_1 -Academic degree has the highest total income for female
- Target_1 -Lower secondary have the lowest total income for both genders
- Target_1 -There is an absence of male people for Academic degree



- Target_0 -Married people have the highest credit amount for both genders with nearly same values
- Target_0 -Single/not married have the lowest credit amount for both genders
- Target_1 - Married people have the highest credit amount for male
- Target_1 -Married and Widow have the highest credit amount for female
- Target_1 -Single/not married have the lowest credit amount for both genders

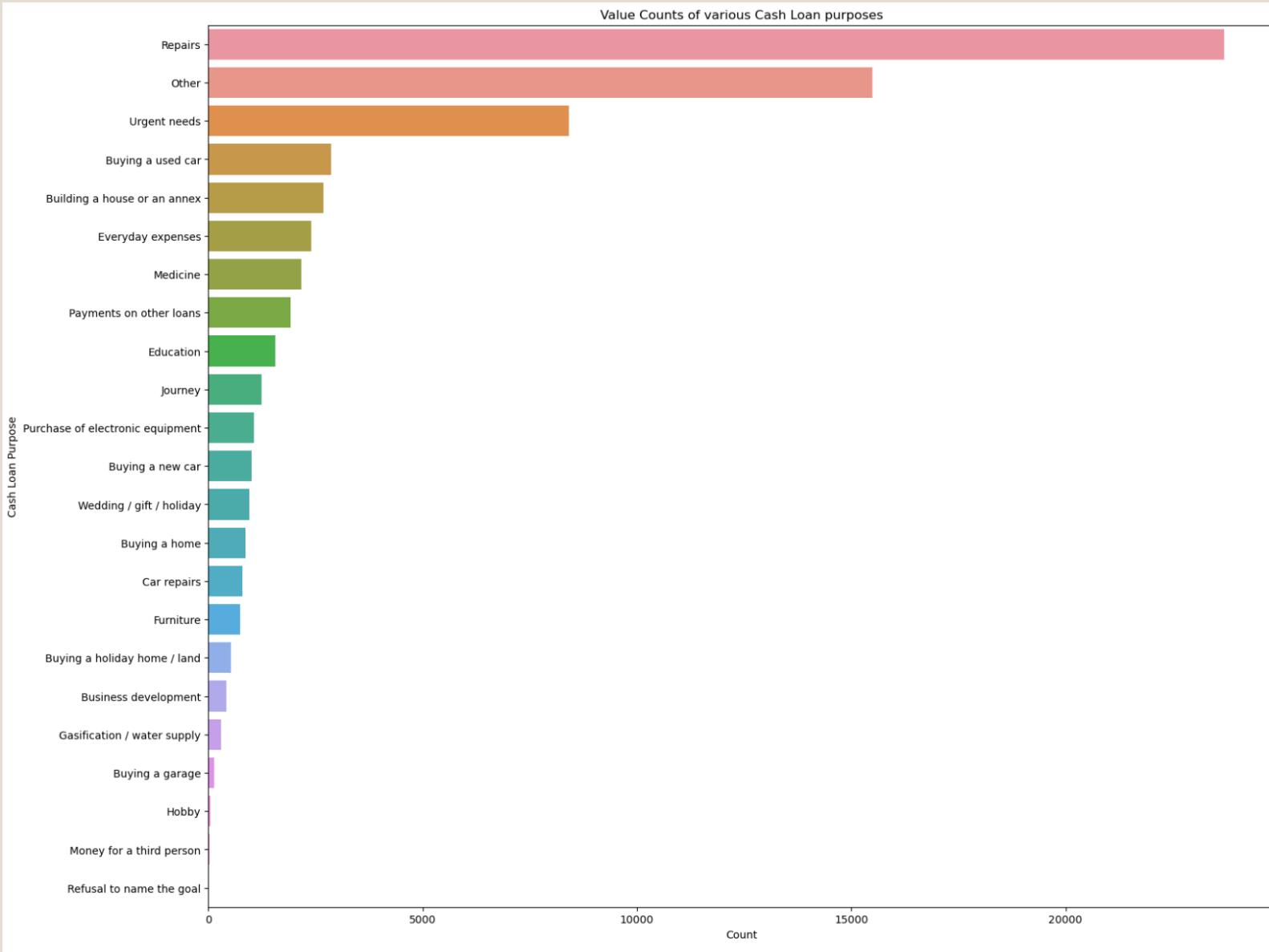


- Target_0 - Loose spread of male and female data can be seen
- Target_0 - Higher total income has no relation to higher loan credit amount
- Target_1 - Tight spread of linear relationship between male and female data can be seen
- Few outliers can be spotted on both target variables

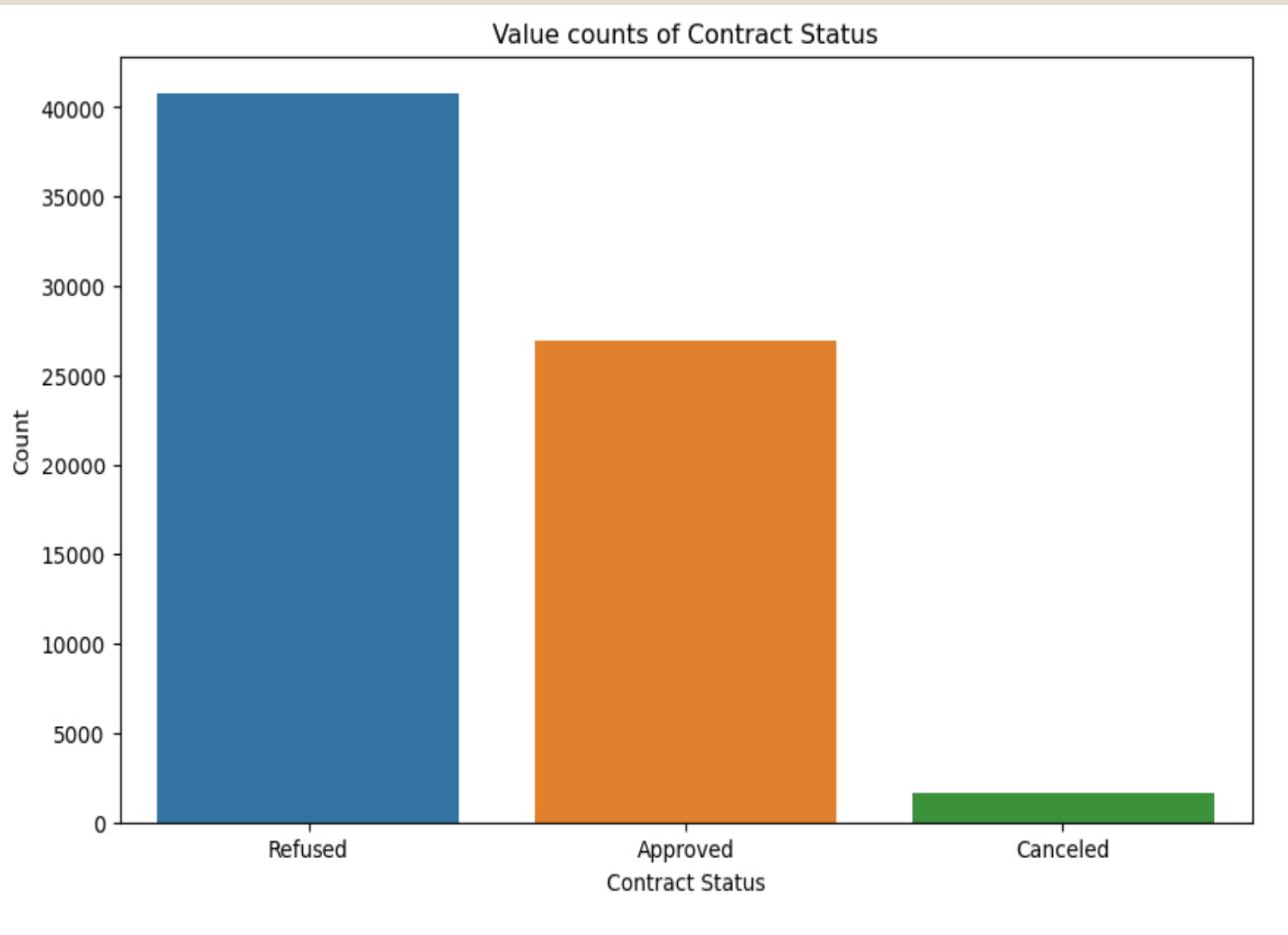


UNIVARIATE ANALYSIS

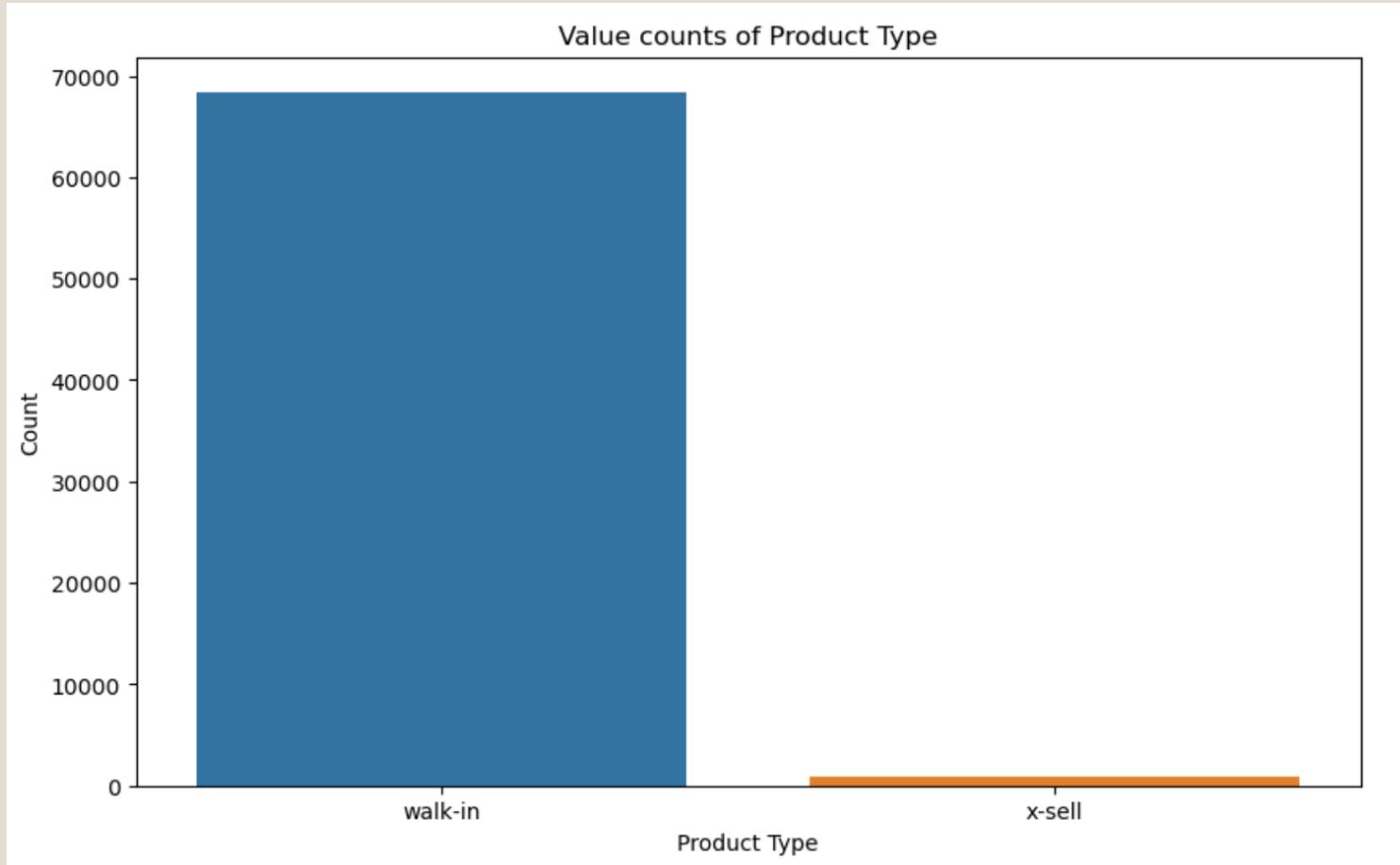
PREVIOUS APPLICATION DATA



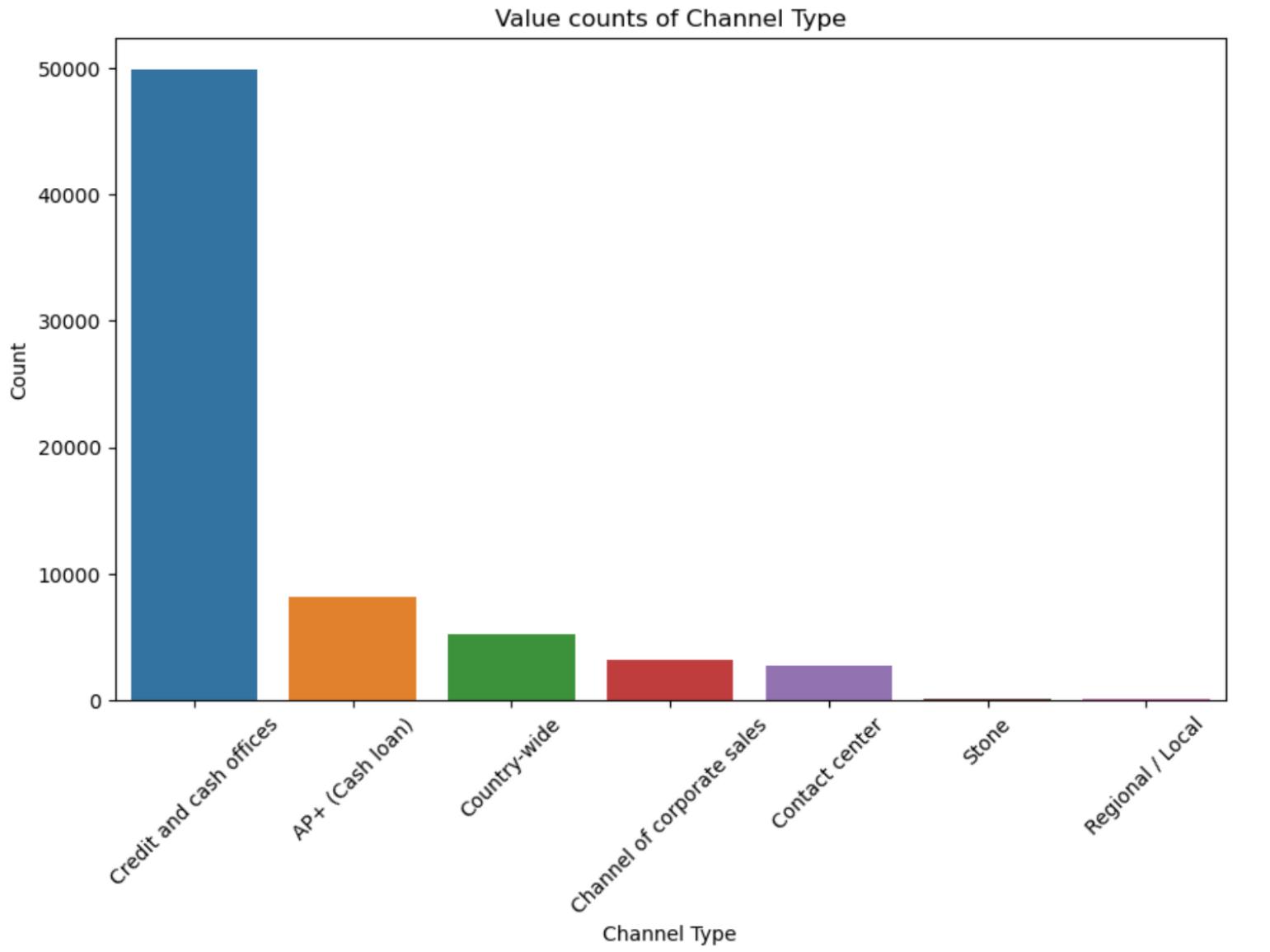
- Loans for Repair purposes have the highest count.
- Buying a car and buying a new house loan requirements are very low.



- Small fraction of contracts are being cancelled.
- The gap between approved and refused contract status is considerably high.
- A lot of contract status are refused this may be due to some underlying factors like certain criteria not being matched.



- Most of the applicants are walk in customers.
- X-sell applications are very low.

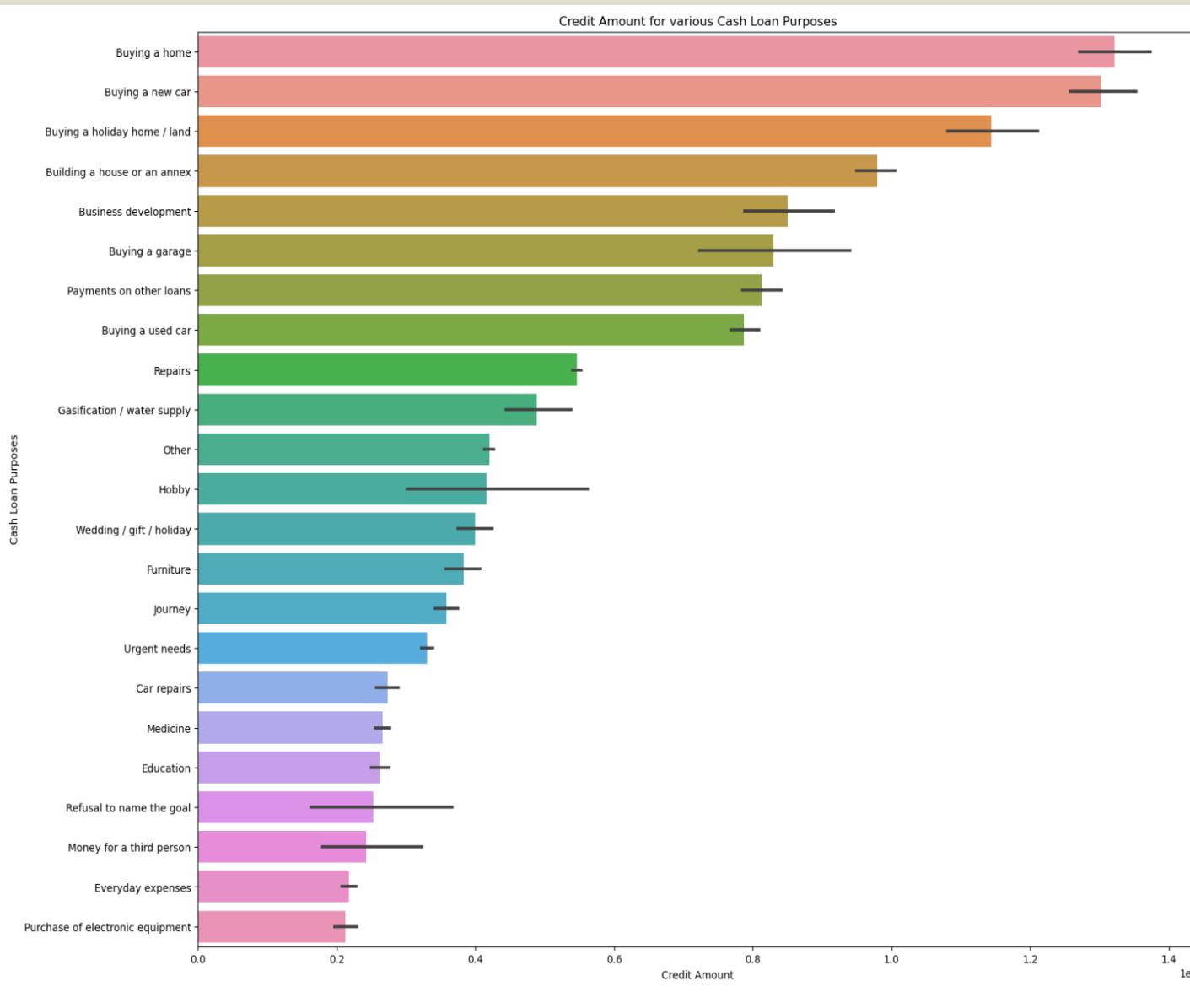


- Credit and Cash offices channel type have the highest value counts.
- Stone and Regional/Local have a very low value count.

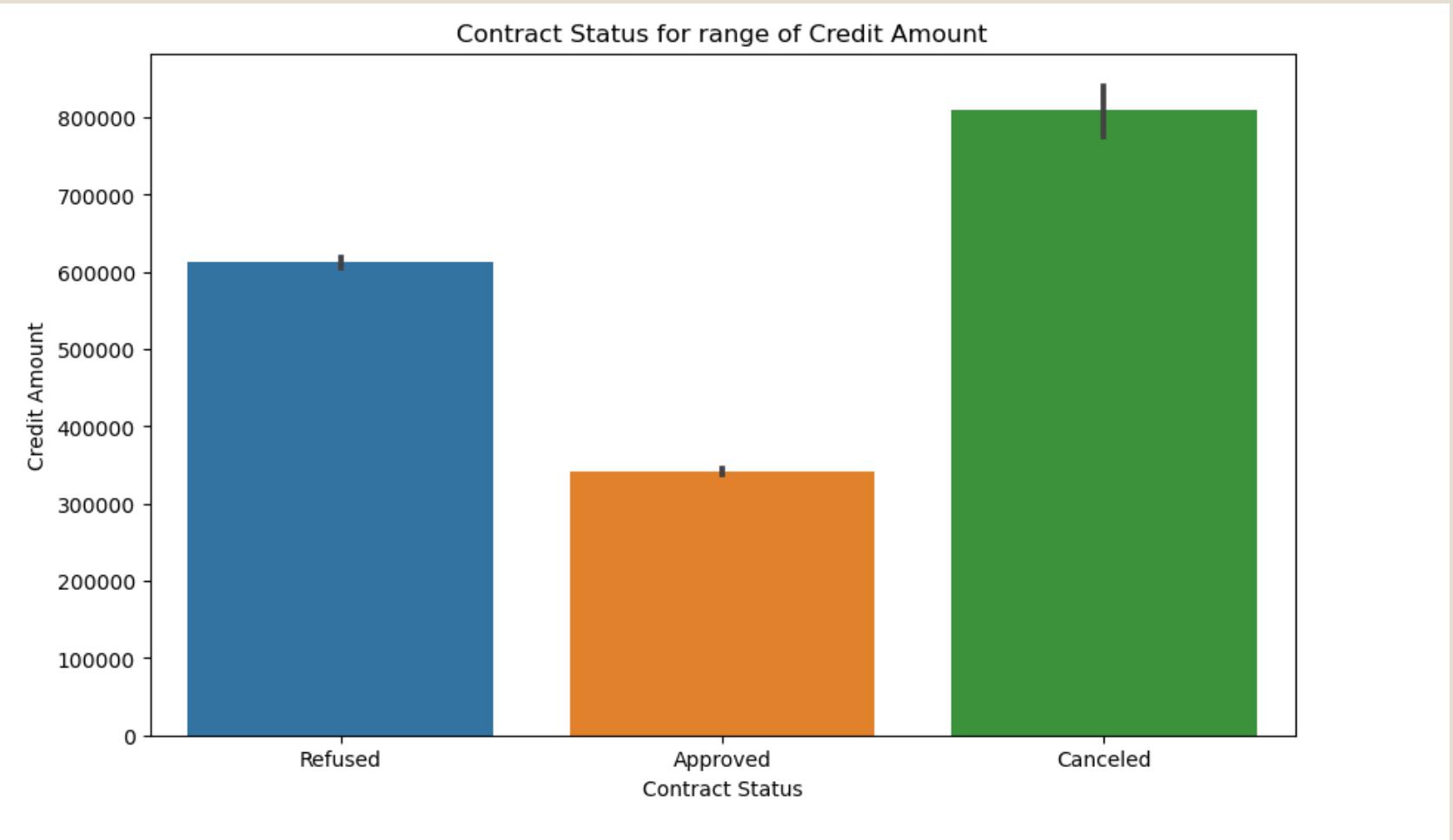


BIVARIATE ANALYSIS

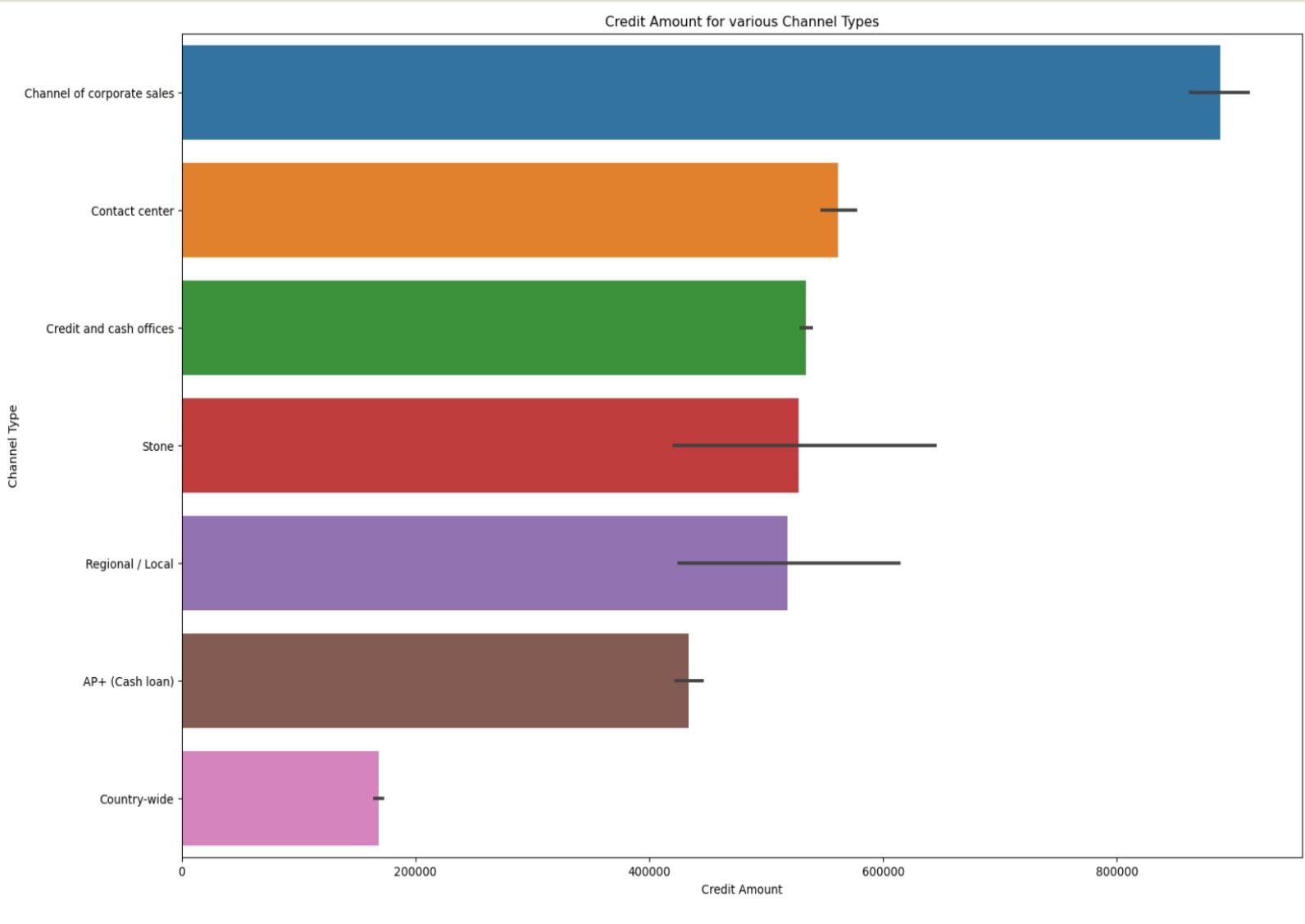
PREVIOUS APPLICATION DATA



- Average Credit amount for buying a home , buying a new car is the highest this may be due to higher purchasing price.
- Average Credit amount for Purchase of electronic equipment and everyday expenses is the least.



- Canceled contracts usually have the highest Credit amount around 800000.
- Approved contracts usually have the lowest Credit amount around 350000.

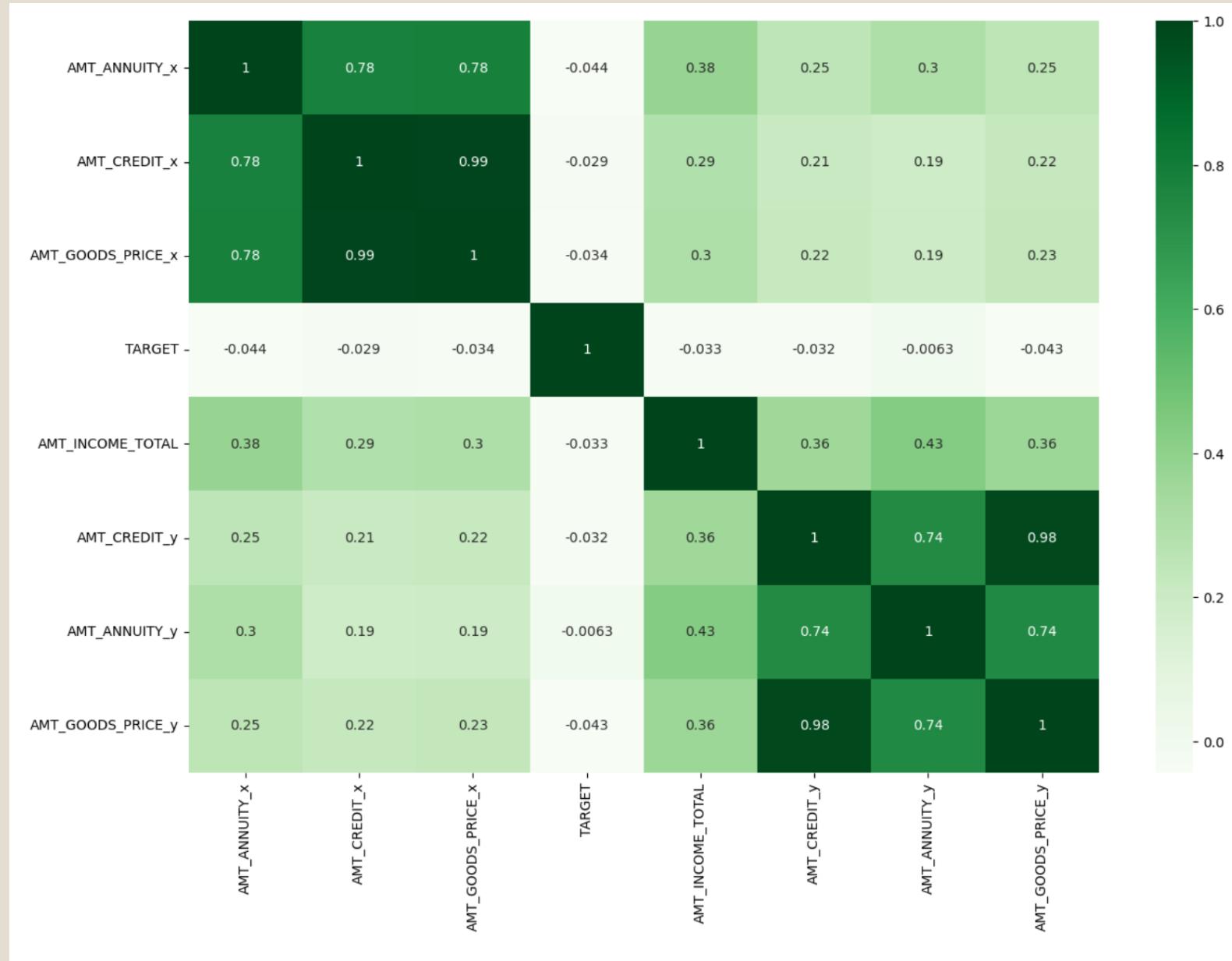


- Channel of corporate sales have the highest credit amount this may be due to higher purchase requirements.
- Country-wide channel type has the lowest credit amount of all.

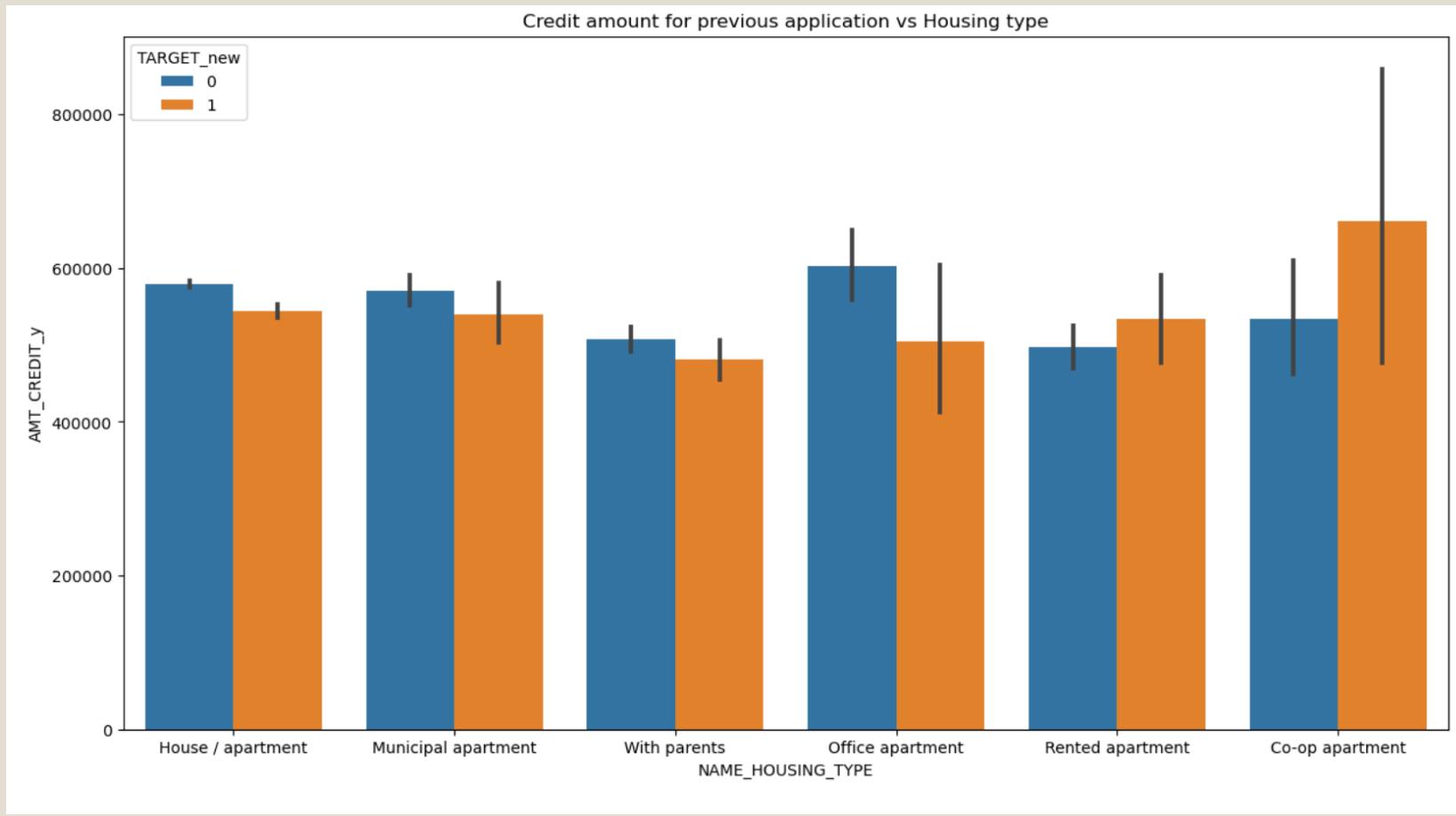


BIVARIATE ANALYSIS

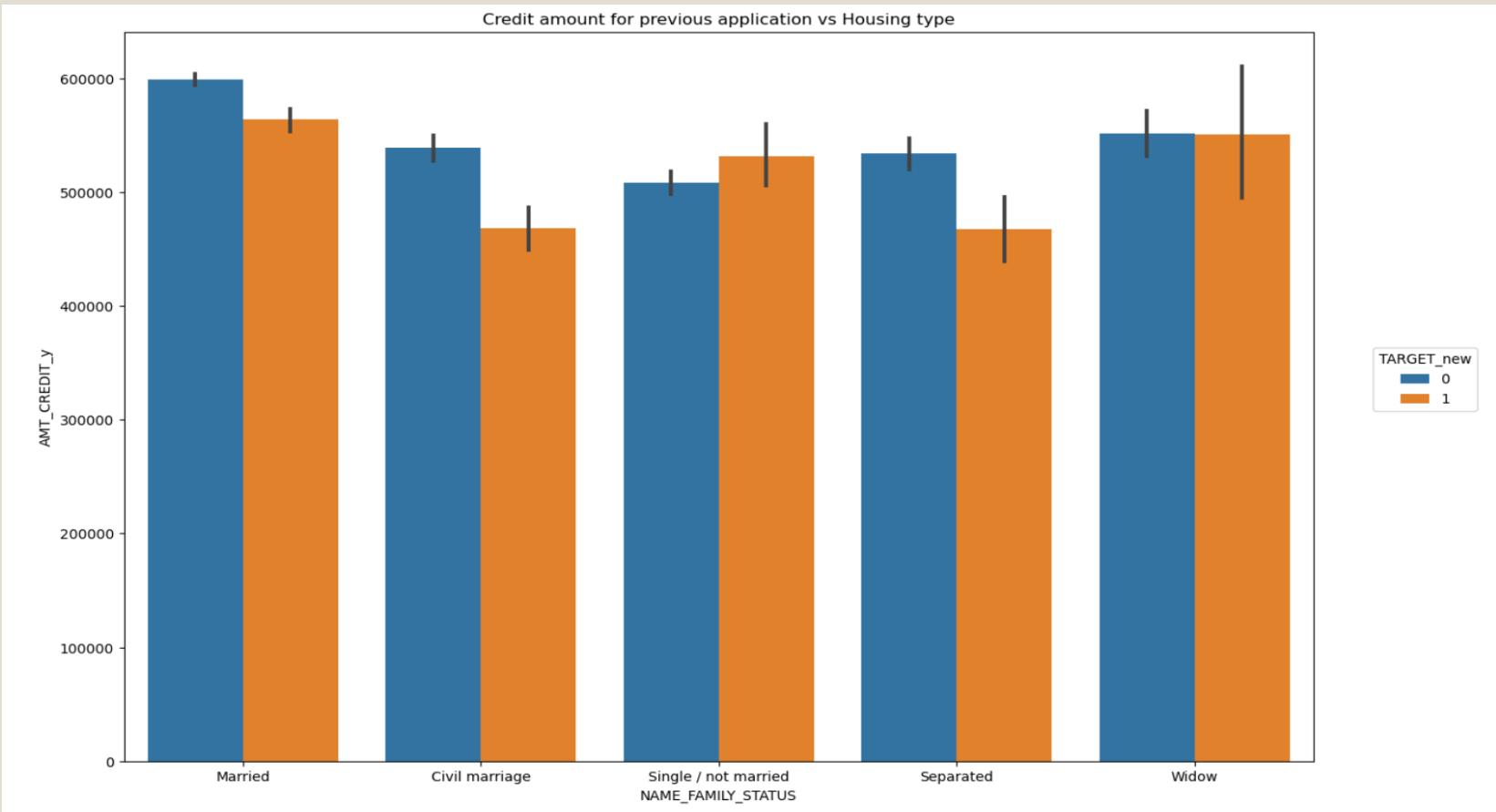
MERGED DATA



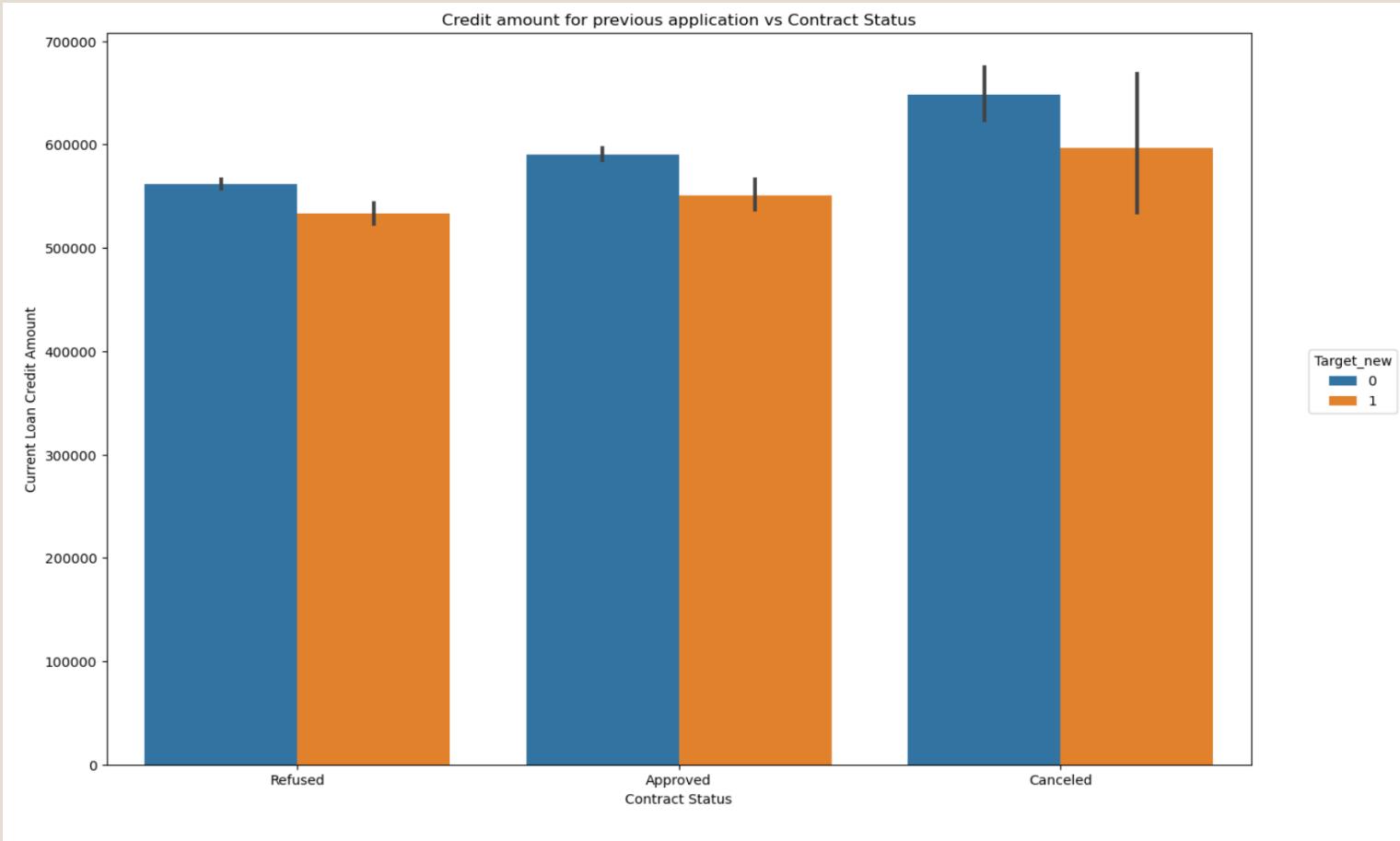
- AMT_ANNUITY_x and AMT_CREDIT_x have a strong positive linear relationship.
- AMT_ANNUITY_x and AMT_GOODS_PRICE_x have a strong positive linear relationship.
- AMT_CREDIT_x and AMT_GOODS_PRICE_x have a very strong positive linear relationship.
- AMT_CREDIT_y and AMT_ANNUITY_y have positive correlation.
- AMT_GOODS_PRICE_y has a positive correlation with AMT_CREDIT_y and AMT_ANNUITY_y.
- Target has the lowest correlation with every other variable



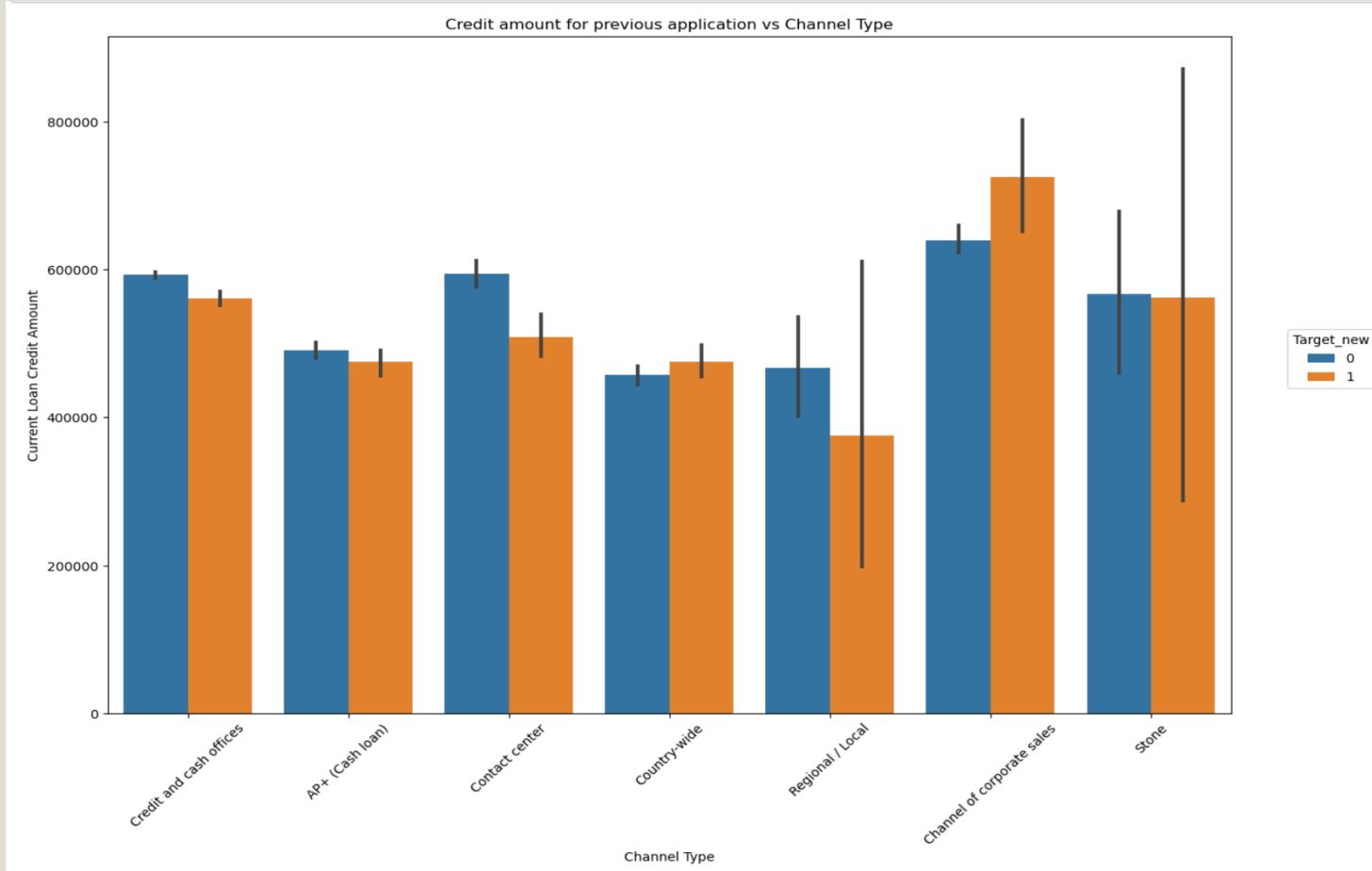
- Target_1 - People with Co-op apartment have the highest credit amount and people living with parents have the lowest credit amount.
- Target_0 - People with Office apartment have the highest credit amount and people with Rented apartment have the lowest credit amount.



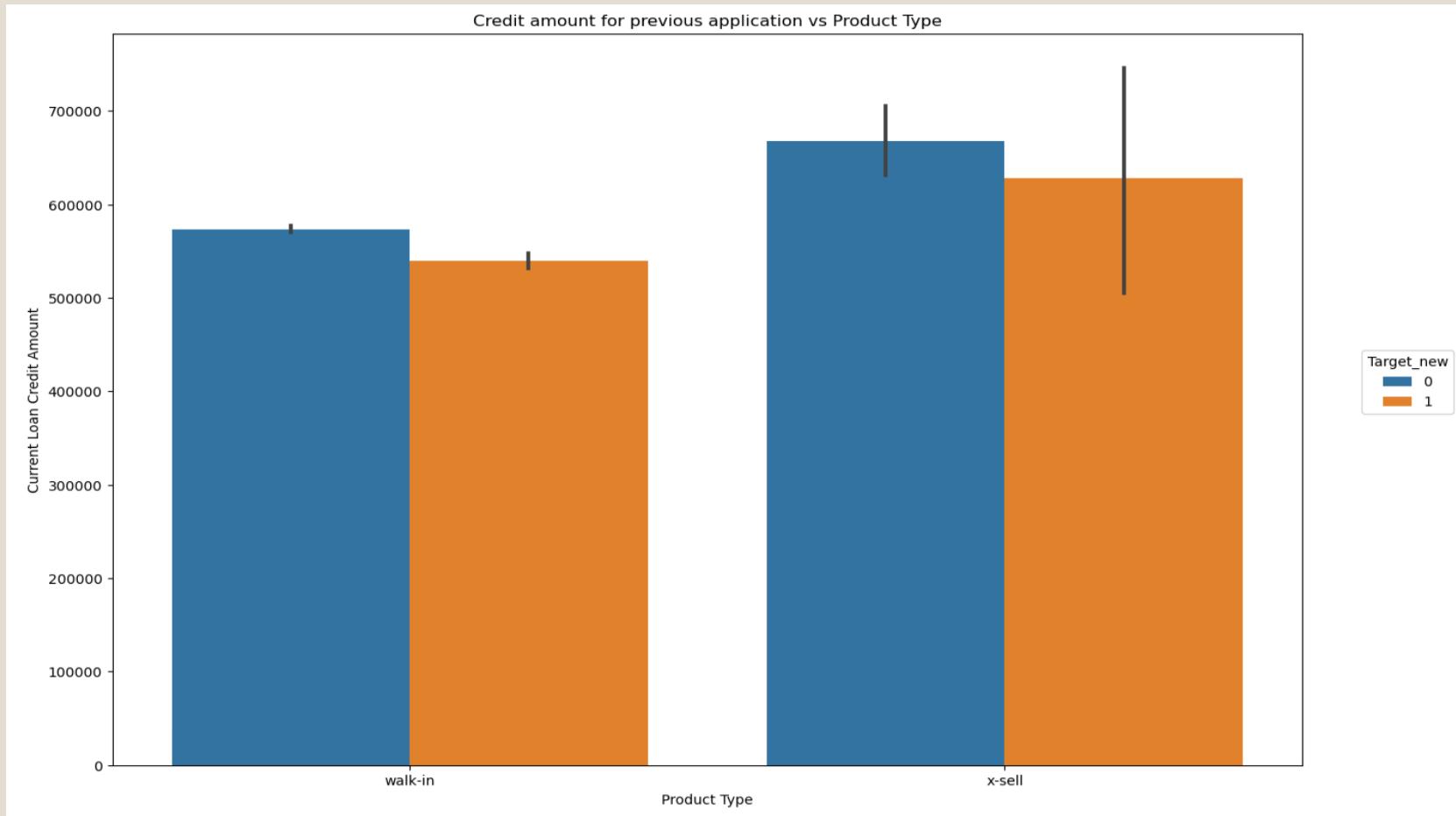
- Target_1 - Married people have the highest credit amount and Single/Not married people have the lowest credit amount.
- Target_0 - Married people have the highest credit amount and Widowed people have the lowest credit amount.



- Target_1 - People with canceled contract status have the highest credit amount and People with refused contract status have the lowest credit amount.
- Target_0 - People with canceled contract status have the highest credit amount and People with refused contract status have the lowest credit amount.



- Target_1 - People with channel of corporate sales have the highest credit amount and People with Regional/Local channel type contract status have the lowest credit amount
- Target_0 - People with channel of corporate sales have the highest credit amount and People with Country-wide channel type have the lowest credit amount.



- Target_1 - X-sell People have the highest credit amount and Walk in people have the lowest credit amount
- Target_0 - X-sell People have the highest credit amount and Walk in people have the lowest credit amount



CONCLUSION

- 1.Compared to men, women are more likely to default. Because men are less likely to be present in target_1, we can boost the Cash Loan credit to them
- 2.Business Entity : Type 3 organisation are the most prone to have defaults. As an alternative, we should concentrate on growing the businesses of the bank, the hotel industry, and the advertising industry as they are less likely to default .
- 3.For bank operations, we can also take into account Legal Services, as they generate the largest income and are less likely to default than when compared to Business Entity: Type-3.
- 4.We can concentrate on Males with Academic Degrees because they make a good amount of income and are relatively less likely to default
- 5.For the benefit of the bank's business, we must increase the number of applications from the Channel of Corporate Sales over Credit and Cash Offices channel and these applicants from Channel of Corporate Sales have the highest Loan Credit Amount.
- 6.We receive the most applications from "walk-in" clients, but since "x-sell" consumers have higher loan amount credit, we can increase the applications of x-sell product type consumers to grow the bank's business



THANK YOU !