

A group of four people (three men and one woman) are seated around a table in a meeting. The image is overlaid with a semi-transparent teal layer. The main title 'CREDIT EDA CASE STUDY' is in large white capital letters, underlined with a thin orange line. Below it, the subtitle 'BANK LOAN' is in orange italicized capital letters, set within a teal rectangular box.

# CREDIT EDA CASE STUDY

*BANK LOAN*

# PROBLEM STATEMENT

---

This case study aims to identify patterns which indicate if a client has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

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:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- 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.

# ANALYSIS STEPS

---

Understanding Data



Data Cleaning



Data Imputation



Finding Outliers



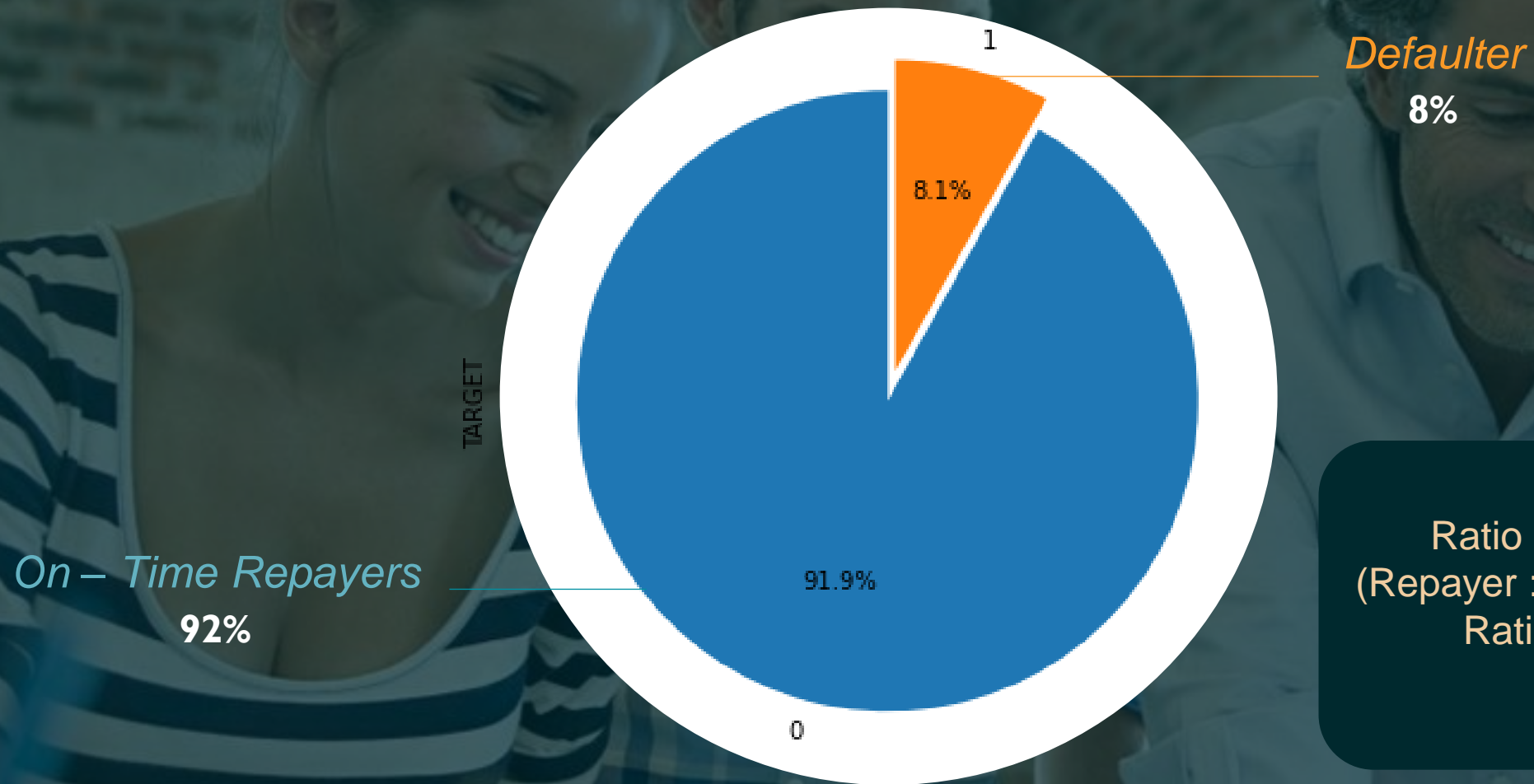
Data Analysis



SUCCESS

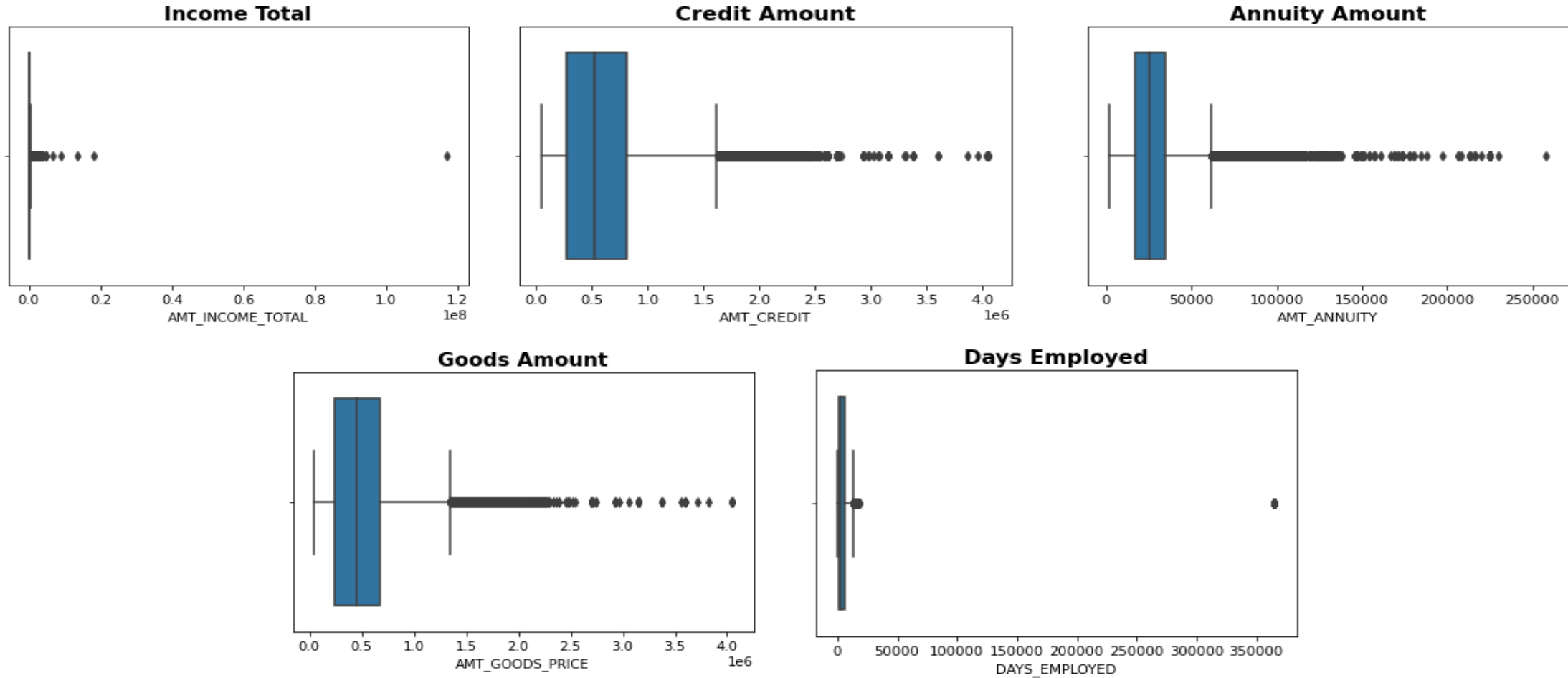


# Data Imbalance



Ratio of Imbalance in %  
(Repayer : Defaulter= 91.9 : 8.1)  
Ratio of Imbalance=  
(11.3 : 1)

# Identifying Outliers

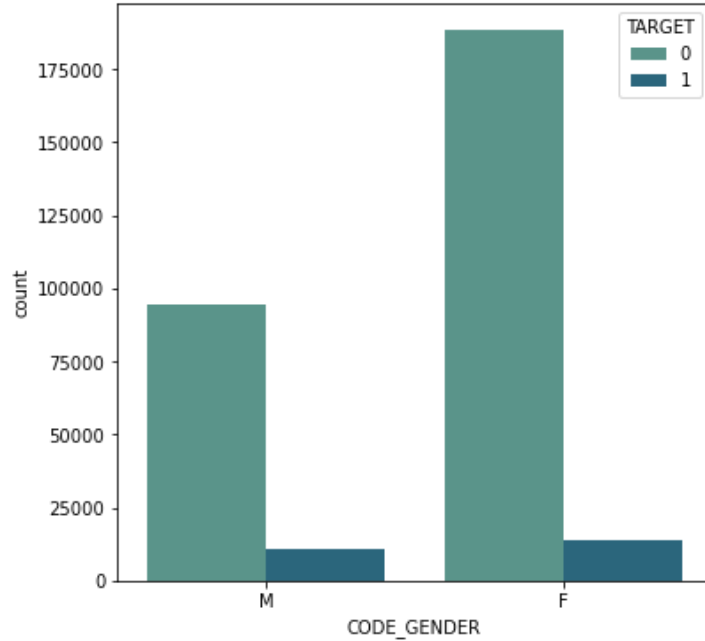


## Inferences:

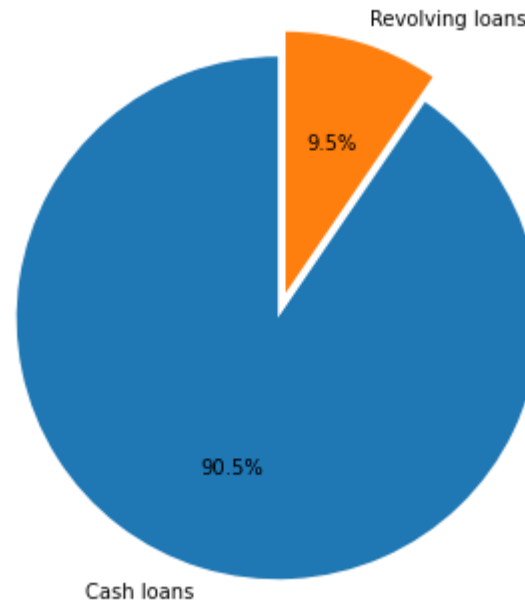
- We can confirm that there are few outliers with high income in Income Total
- As we can see there are about 2000 outliers whose credit amount is higher and being regarded as outliers
- As we can see there are around 1500 outliers in annuity amount column.
- We can see there are few outliers in Goods Price Amount.
- As we can see above 80% quantile all are having same days employed as outliers

# Univariate Analysis

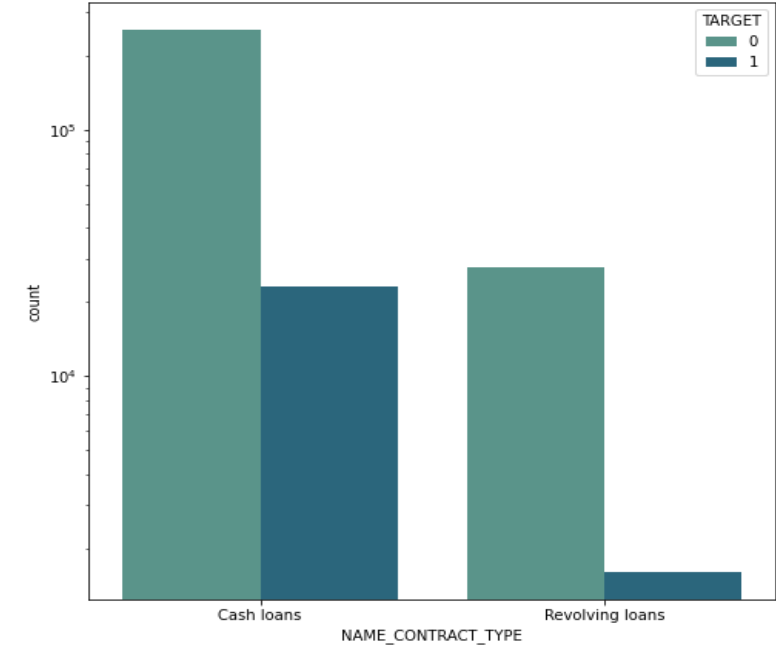
Gender wrt Target



NAME\_CONTRACT\_TYPE



Type of loan wrt Target



## Inferences:

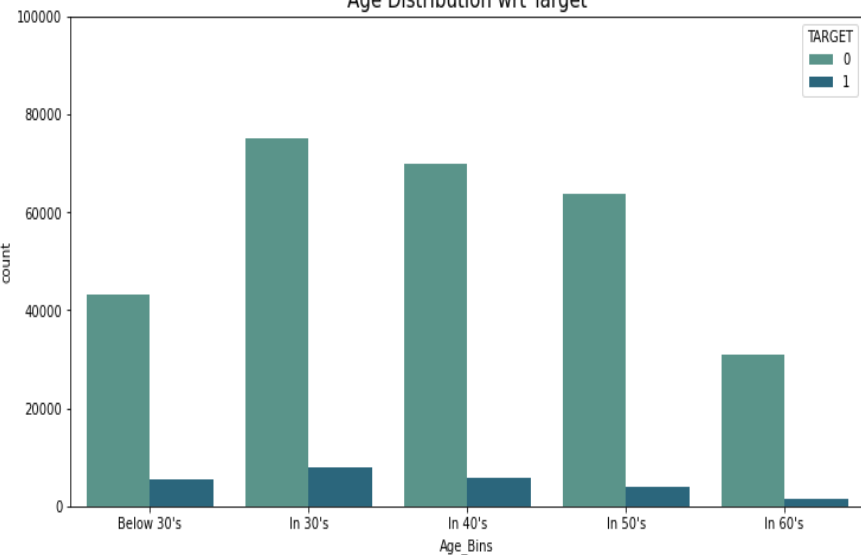
- *We can confirm that more loan applicants are females and most of them tend to repay. So banks should consider giving loans for women.*

## Inferences:

- *We can see around 90% of the loans are Cash Loans*
- *We can see Revolving Loans have very less number of defaulters. So consider giving loans of revolving type*

# Univariate Analysis

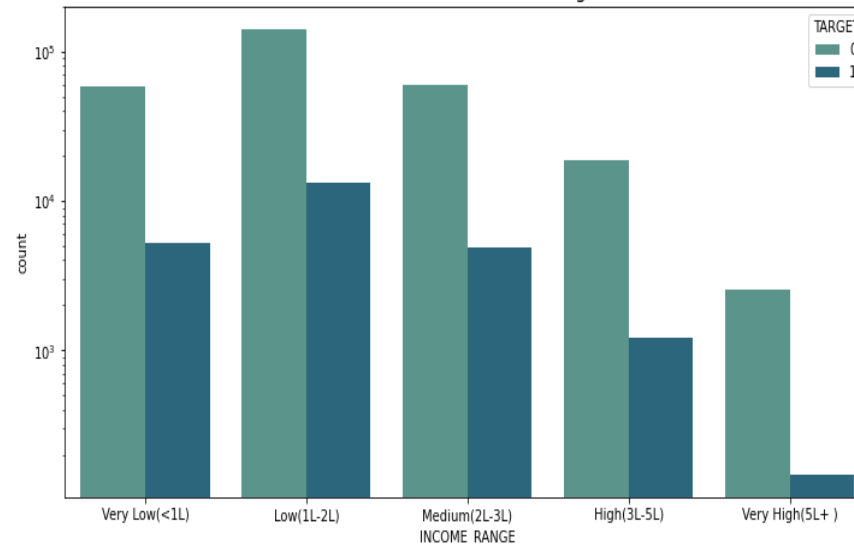
Age Distribution wrt Target



## Inferences:

- *Loan applicants of age in 30's are able to make their payments on time*
- *Considerably above 60s are more percentage of Repaying on time*

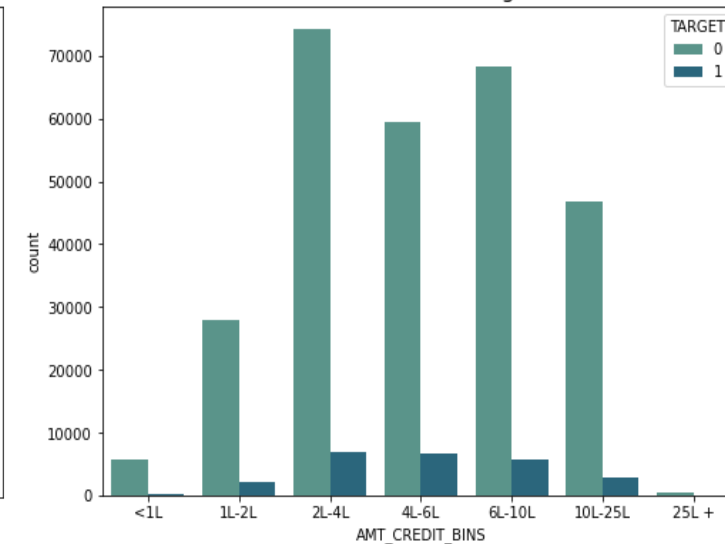
Income Distribution wrt Target



## Inferences:

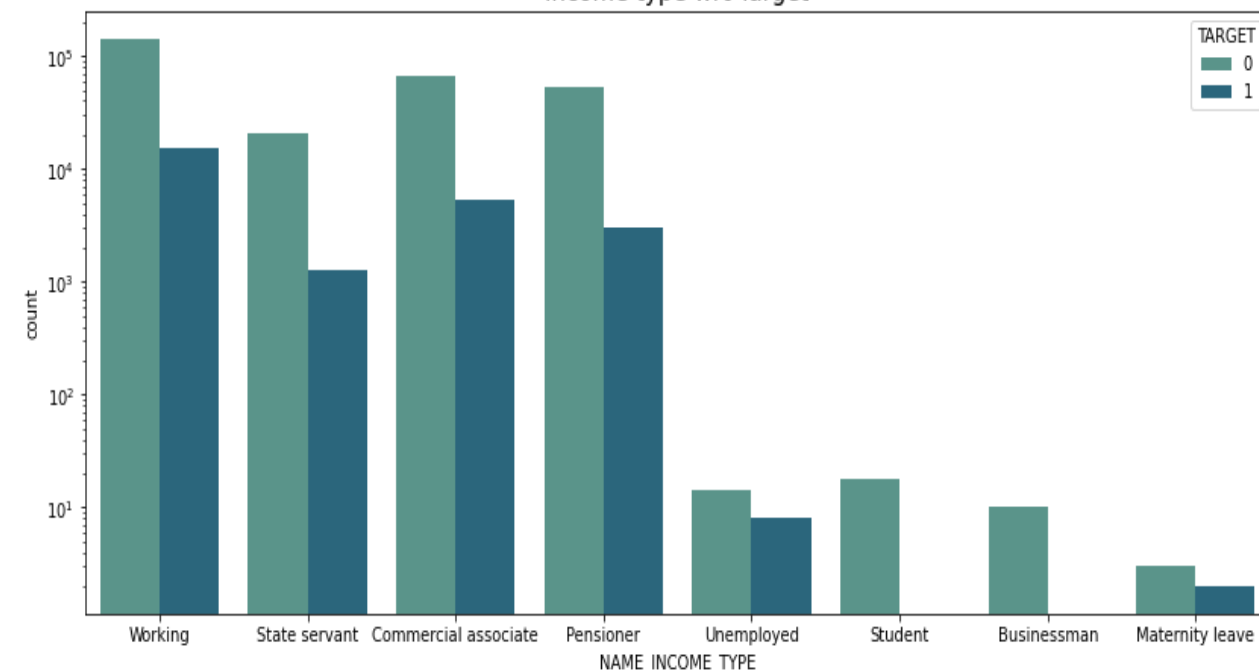
- *Applicants with their income in range (5L+) have less difficulties in repayment on time and able to repay on time.*
- *Here applicants with very high and low credit amount have very less defaulters.*
- *2L-4L credit amount is the highest credit range pickup with more defaulters.*

Credit Amt wrt Target



# Univariate Analysis

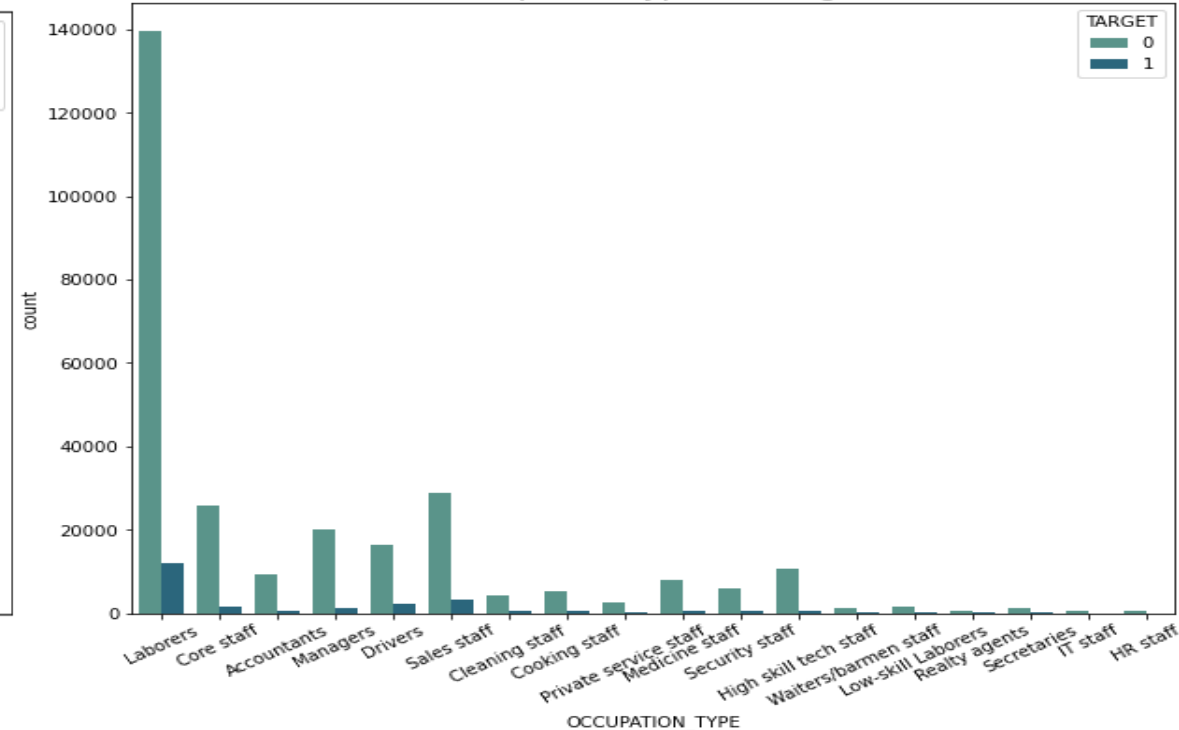
Income type wrt Target



## Inferences:

- *Unemployed & Maternity Income type applicants have more defaulters and should be thoroughly rechecked before confirming loan*
- *Though Student and Business man are less in number they have nil default record. So can be considered for loan*
- *Most of the Loan applicants are Working professionals*

Occupation Type wrt Target



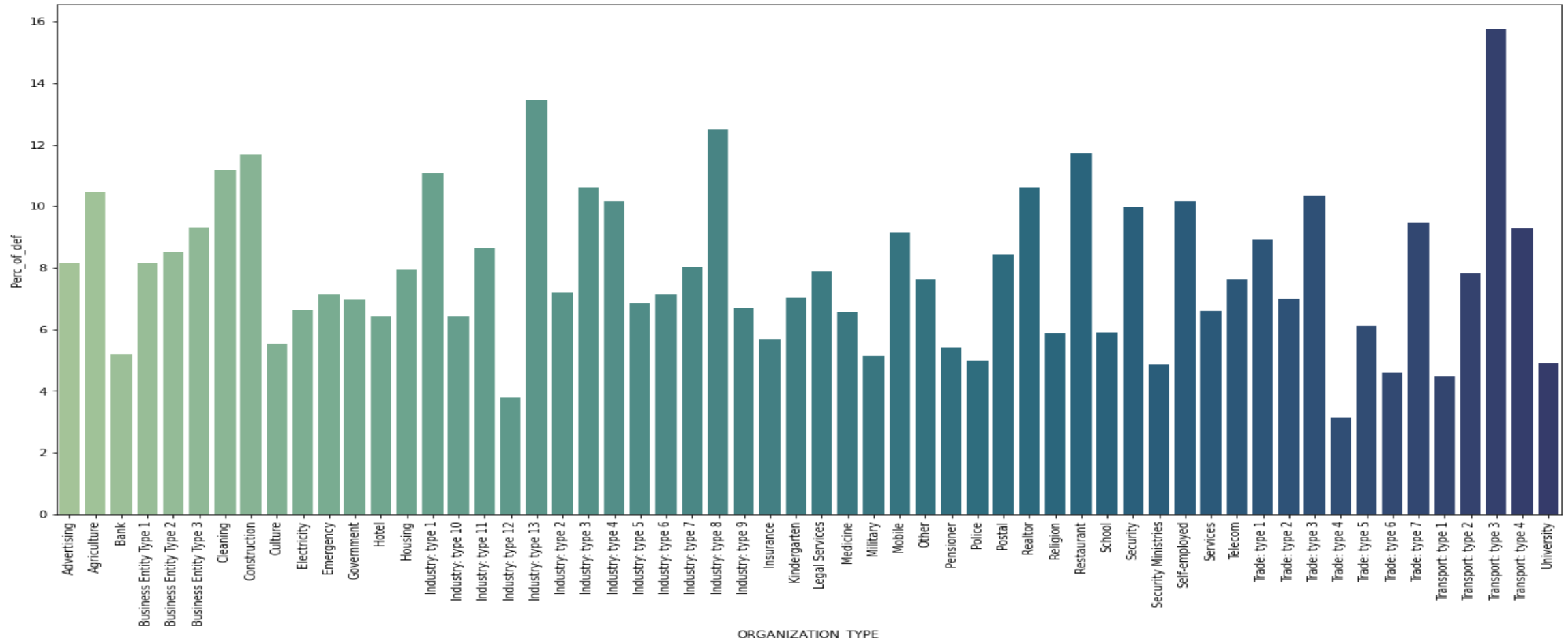
## Inferences:

- *We can clearly understand that laborers are highest in repaying the loan in-time with very less difficulties*



# Univariate Analysis

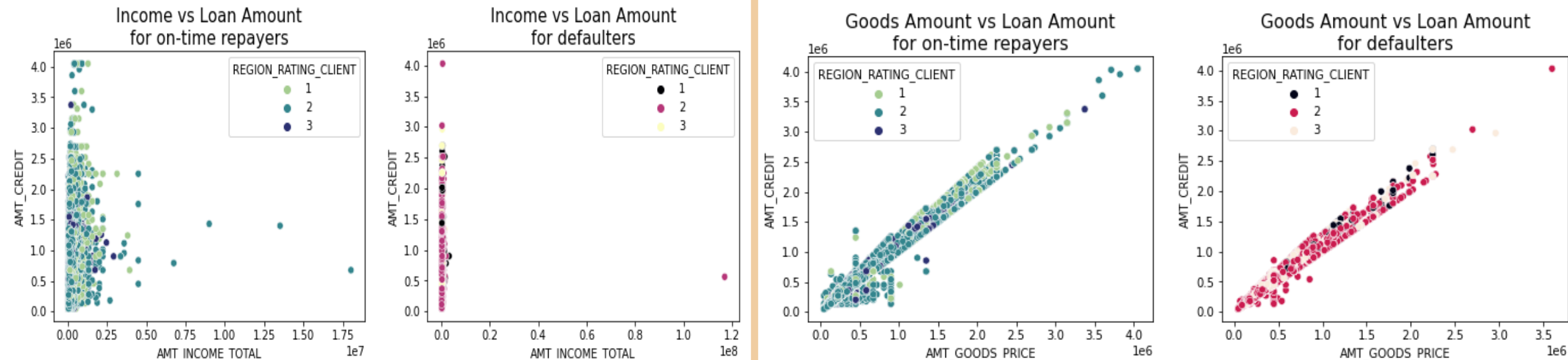
Percentage of defaulters in Organization Type



## Inferences:

- We can clearly say from here that Transport Type 3 are higher in default percentage(16%) followed by Industry Type 13(13%) and Industry Type 8(12.5%).
- The lowest amount of defaulters are Trade type 4 Organization.

# Bivariate Analysis



## Inferences:

- From here we can observe that on time repayers have more income as well as more likely to get more amount of loan where as compared to less income and less credit of defaulters
- There is no Linear correlation between both
- Most of the high income for on-time repayers are from Region Rating 2

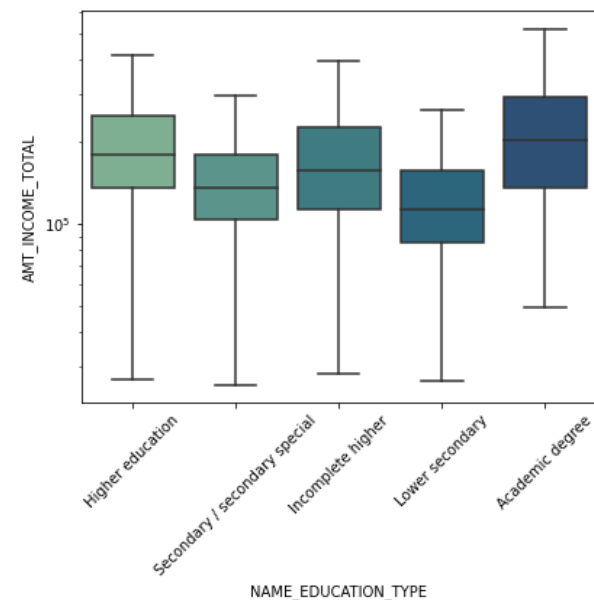


## Inferences:

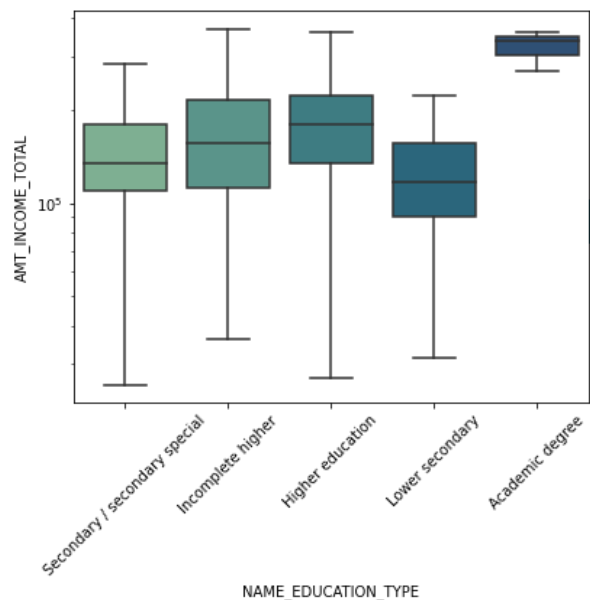
- It clearly shows that both Loan amount and Goods Price are in Linear correlation between them
- We can also see on-time repayers tend to get higher credit corresponding to higher goods price
- Region rating 2 have acquired the highest amount of credit and most of them repay on time

# Bivariate Analysis

Income vs Education for on-time Repayers



Income vs Education for defaulters



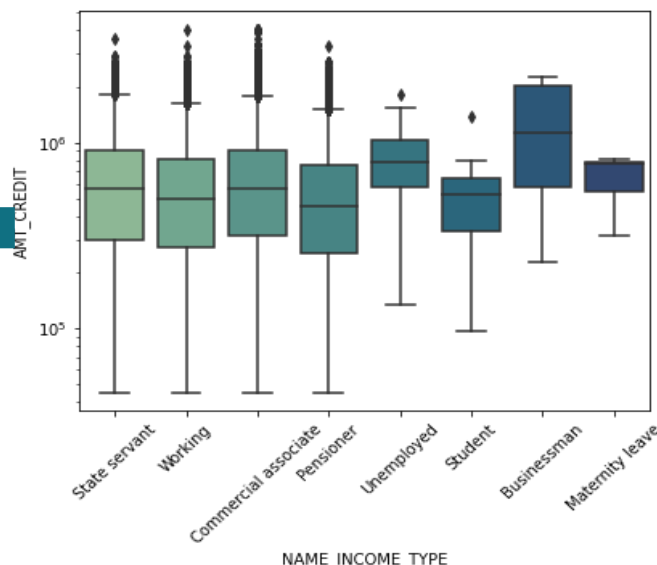
## Inferences:

- As we can see the median income of Academic category aims to be the highest for on-time repayers
- There are also on-time repayers from higher education with more income compared to others

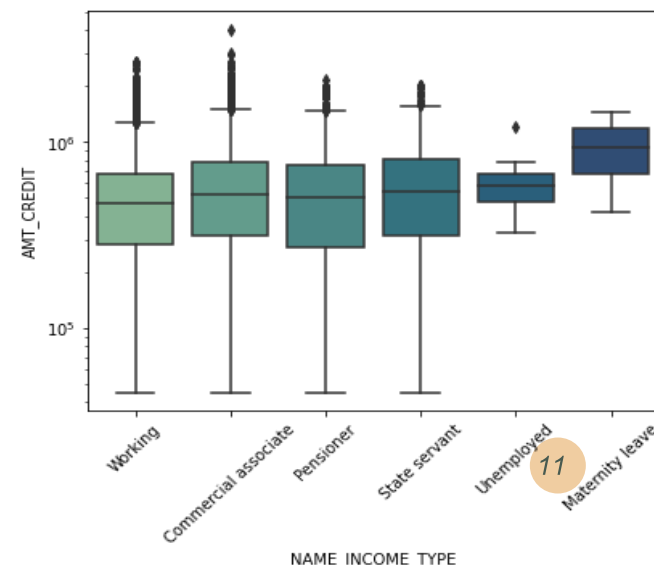
## Inferences:

- As we can see that all Business man are in the higher range of Credit amount takers as compared to other
- We can clearly see that there are no defaulters in Business man Income type.

Income Type vs Credit Amount for on-time Repayers



Income Type vs Credit Amount for defaulters



# Correlation Matrix

-----Top 10 Correlation for Target = 0-----

	VAR1	VAR2	CORR
0	AMT_GOODS_PRICE	AMT_CREDIT	0.99
1	DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.86
2	AMT_ANNUITY	AMT_GOODS_PRICE	0.78
3	AMT_ANNUITY	AMT_CREDIT	0.77
4	Age	DAYS_EMPLOYED	0.63
5	AMT_INCOME_TOTAL	AMT_ANNUITY	0.42
6	AMT_INCOME_TOTAL	AMT_GOODS_PRICE	0.35
7	AMT_CREDIT	AMT_INCOME_TOTAL	0.34
8	DAYS_REGISTRATION	Age	0.33
9	DEF_30_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.33

-----Top 10 Correlation for Target = 1-----

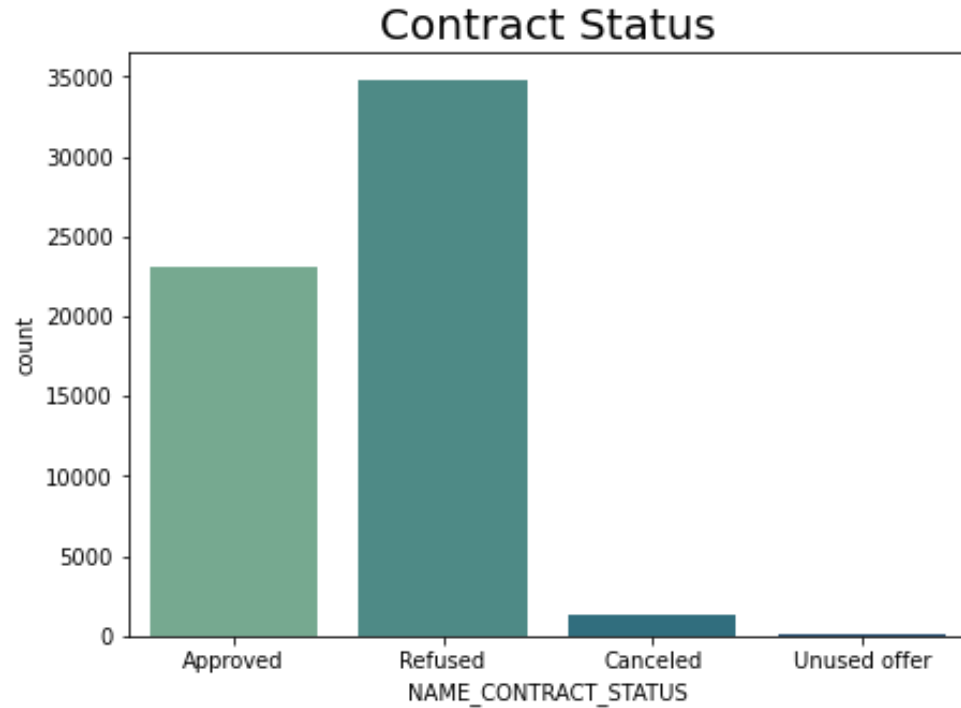
	VAR1	VAR2	CORR
0	AMT_GOODS_PRICE	AMT_CREDIT	0.98
1	DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.87
2	AMT_ANNUITY	AMT_GOODS_PRICE	0.75
3	AMT_CREDIT	AMT_ANNUITY	0.75
4	DAYS_EMPLOYED	Age	0.58
5	OBS_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.34
6	OBS_30_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.33
7	Age	DAYS_REGISTRATION	0.29
8	OBS_60_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.26
9	OBS_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.26

## Inferences:

- *Goods Price vs Credit amount being the one having the highest correlation of about 0.99.*
- *Correlation for Amount Annuity and credit amount is slightly reduced for defaulters.*
- *Correlation for Total Income and Annuity seems to be dropped heavily for defaulters(which is in later position from top 10).*

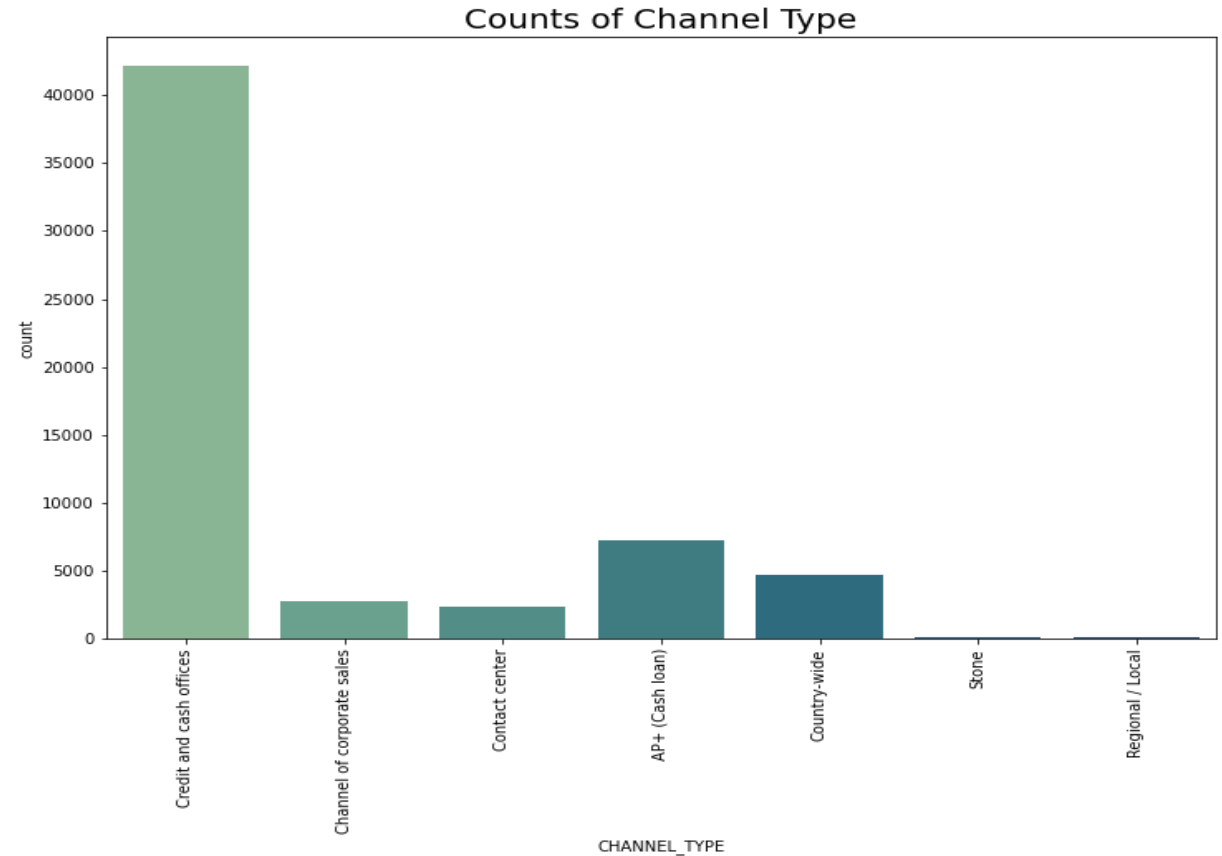


# Merged Data Analysis



#### Inferences:

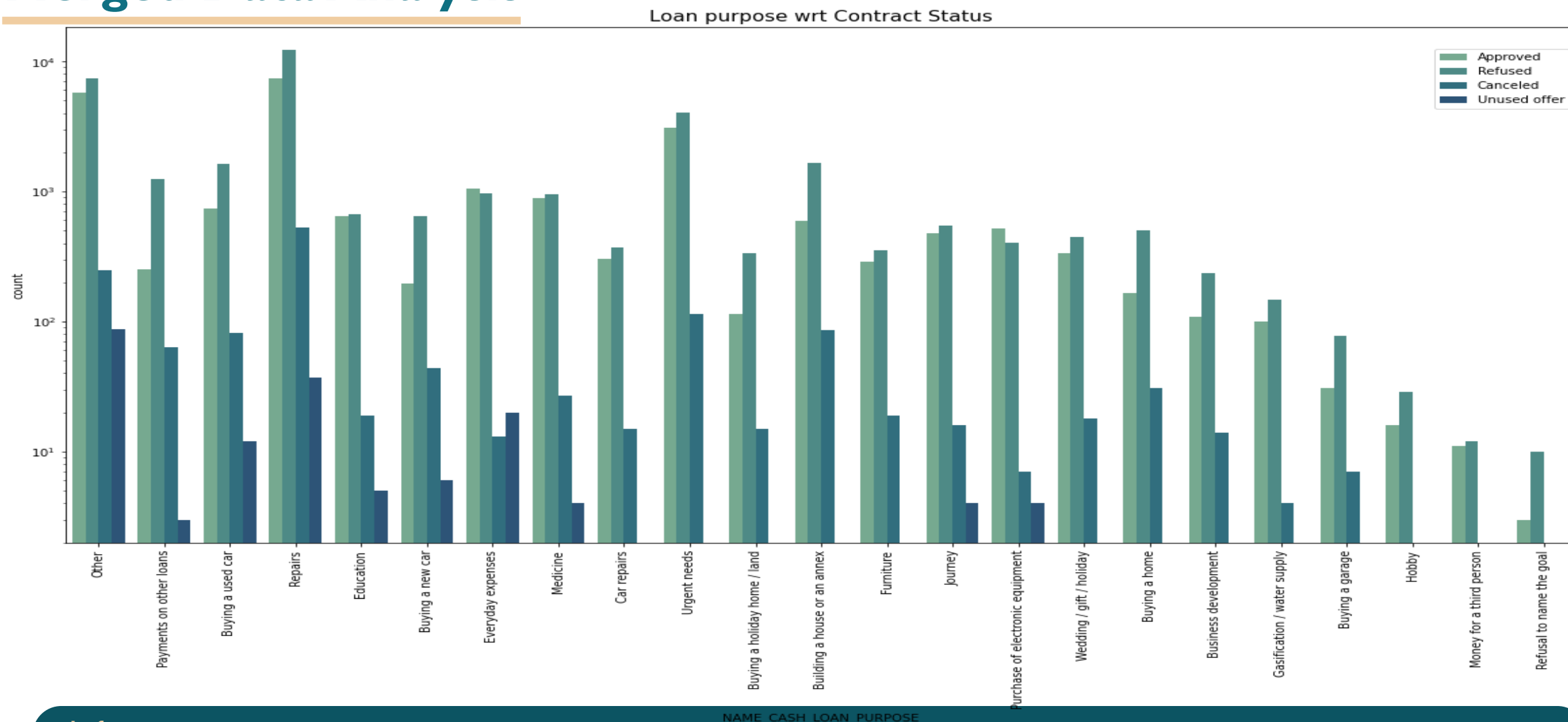
- As we can see most of the applications are Rejected and very few are unused or cancelled



#### Inferences:

- As can be seen most of the previous applications have Credit and Cash offices as their CHANNEL\_TYPE

# Merged Data Analysis

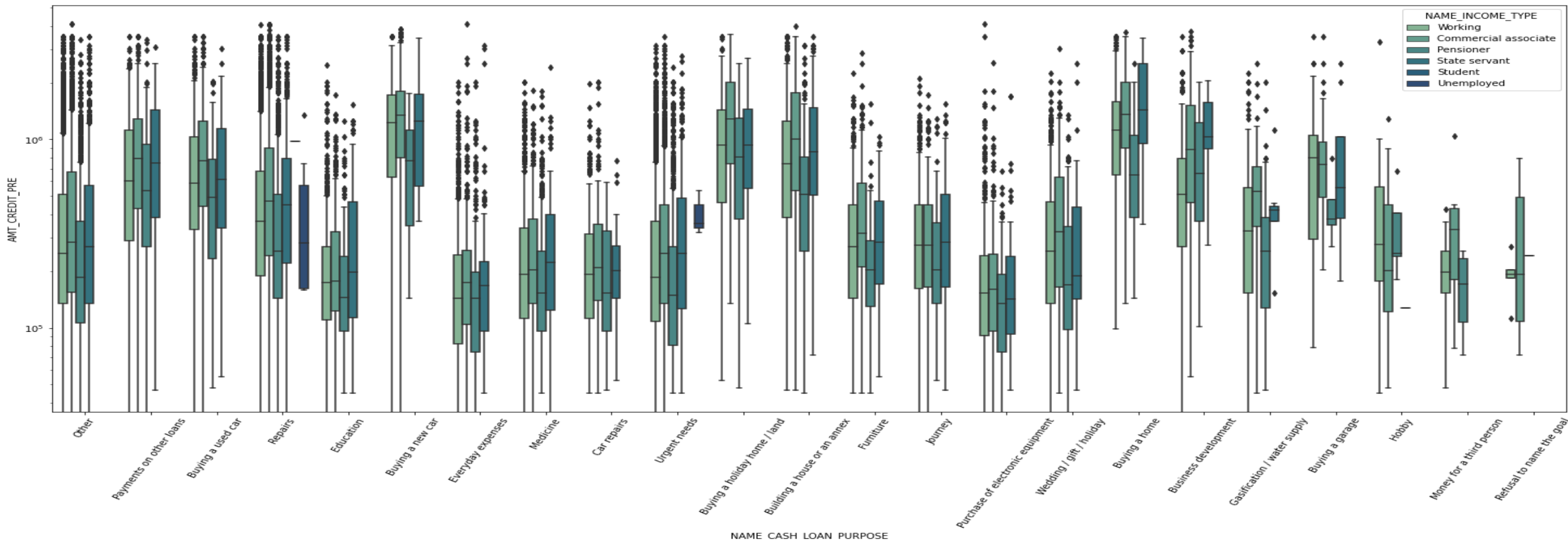


## Inferences:

- *Repairs have the highest number of rejections as in loan purposes.*
- *Paying to other loans , Refusal to name the goal and buying a new car have significant number of rejections when compared to approval rate.*
- *Education and money to a third person have almost equal number of approvals and rejections.*

# Merged Data Analysis

Loan Purpose vs Credit Amount wrt Type of income

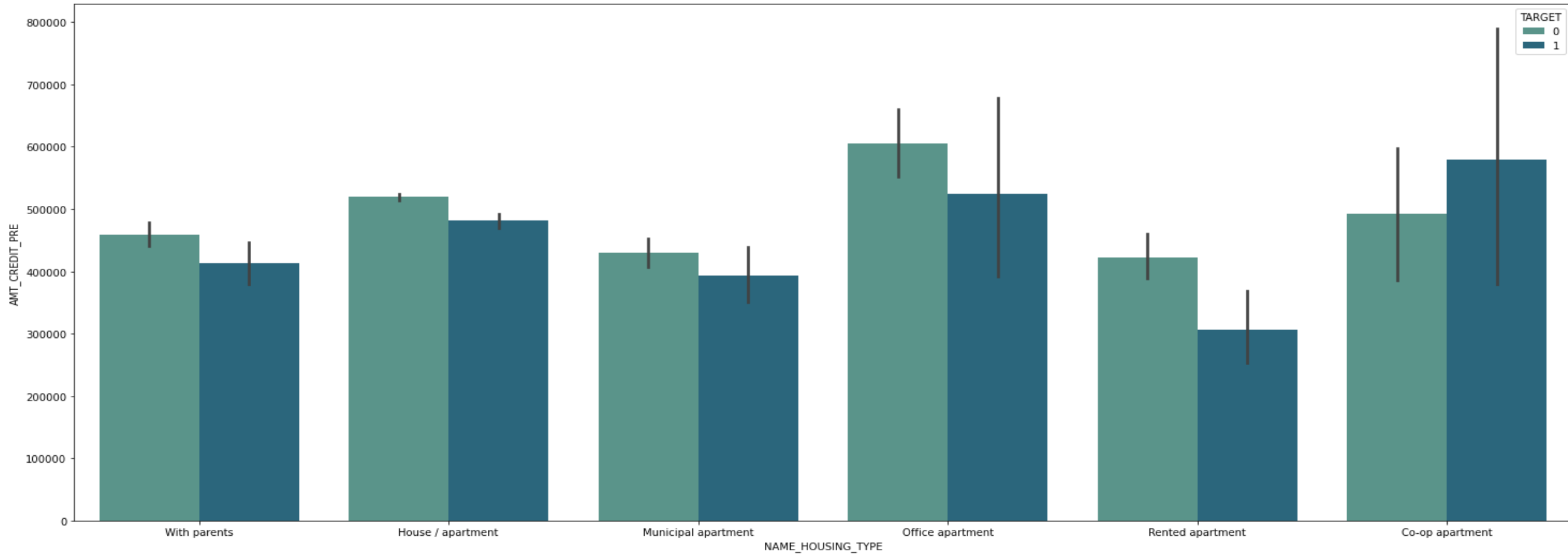


## Inferences:

- As we can see the Loan amount for Buying a home or a land , buying a new car or business developments have their highest credit amount for type of Income types
- Credit amount for Purchase of everyday expenses & Purchase of electronic equip. aims to be the lowest amount among others.
- The Median Credits for state Servant aims to be higher in respective to other Income types.

# Merged Data Analysis

Housing Type vs Credit Amount wrt Target



## Inferences:

- As we can see that Co-op Apartment have higher no: of difficulty of payments wrt higher credit amount. So the bank should avoid giving higher credit for Co-op Apartment.



## Concluding factors for Bank( Repayers)

- ✓ Bank should consider giving more Revolving Loans.
- ✓ Applicants with High Income( greater than 6L ) are less likely to default.
- ✓ Loan applicants with organization Trade Type 4 are less likely to default.
- ✓ Loan takers with Academic degree are less likely to default.
- ✓ Should consider giving loan to more Females.
- ✓ Students and Businessman are less likely to default. So can be considered.



## Concluding Factors for bank(Defaulters)

---

- ✓ Gender male are more likely to default.
- ✓ Unemployed and Maternity leave people are more likely to default.
- ✓ People living in rented apartment are more likely to default.
- ✓ Industry type 3,13,8 are more likely to default.
- ✓ Application with loan purpose as Repairs are more likely to default and most of them are rejected by the bank.
- ✓ Applicants who live in region 3 are more likely to default.
- ✓ Applicants with housing type as co-op apartments should not be considered for loans

# THANK YOU!

---

-  *Gubbala Venkata Hemanth*
-  *hemanthhemu17@gmail.com*
-  *DSC-48 Batch*

