



# Credit EDA Case Study

---

YUGANDHAR IPPILI

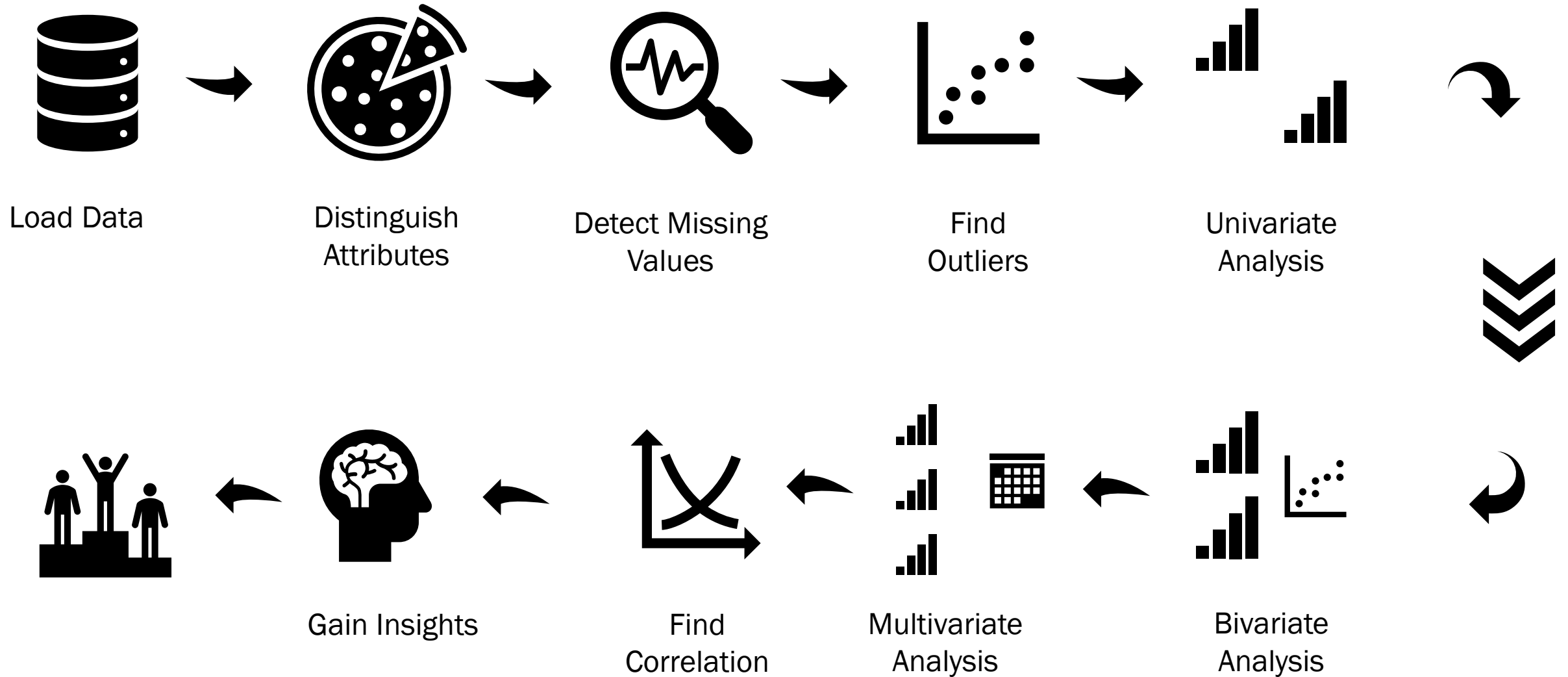
MANASA KARANAM

# Business Understanding

---

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. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.

# Problem Solving Methodology





# Inferences from Application Data

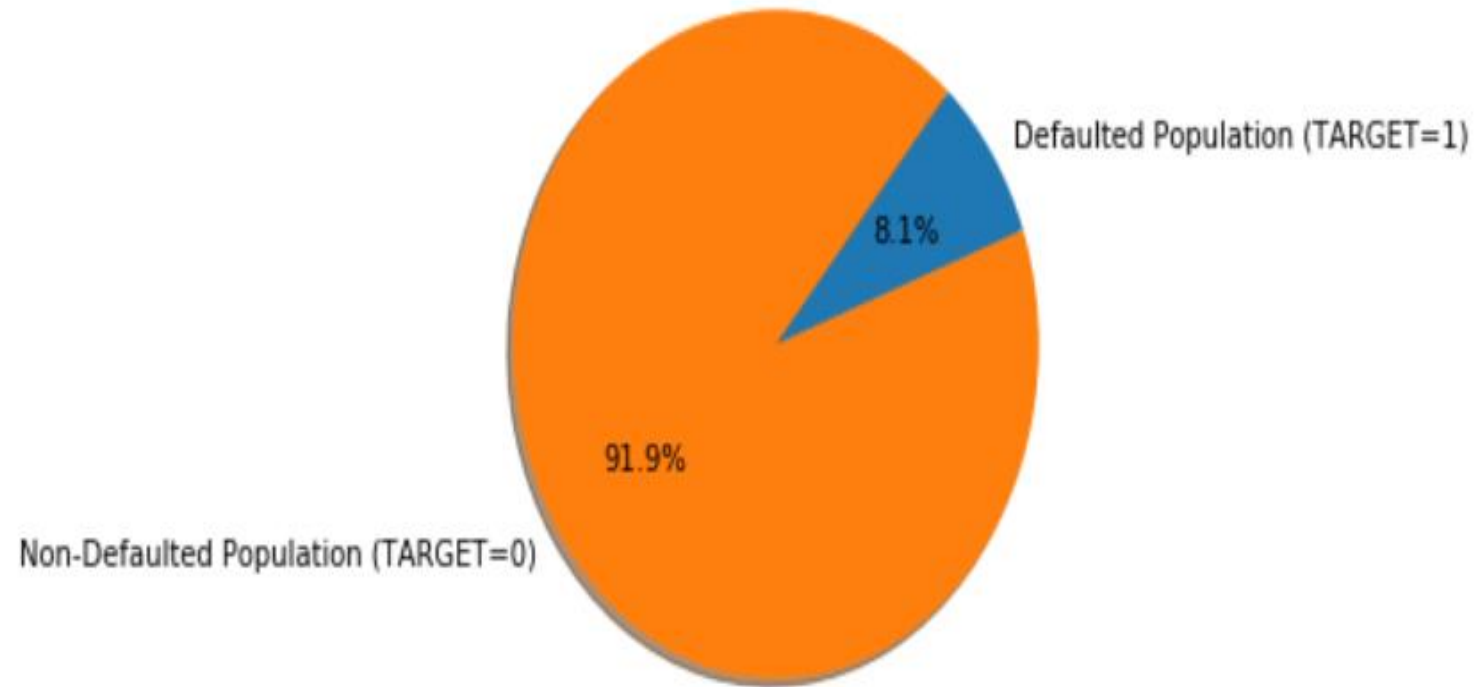
---

# Checking Data Imbalance

As per the pie chart, 91.9% population are non-defaulters in loan payment and 8.1% are the defaulters.

Ratio of imbalance is 11.4

## Data Imbalance



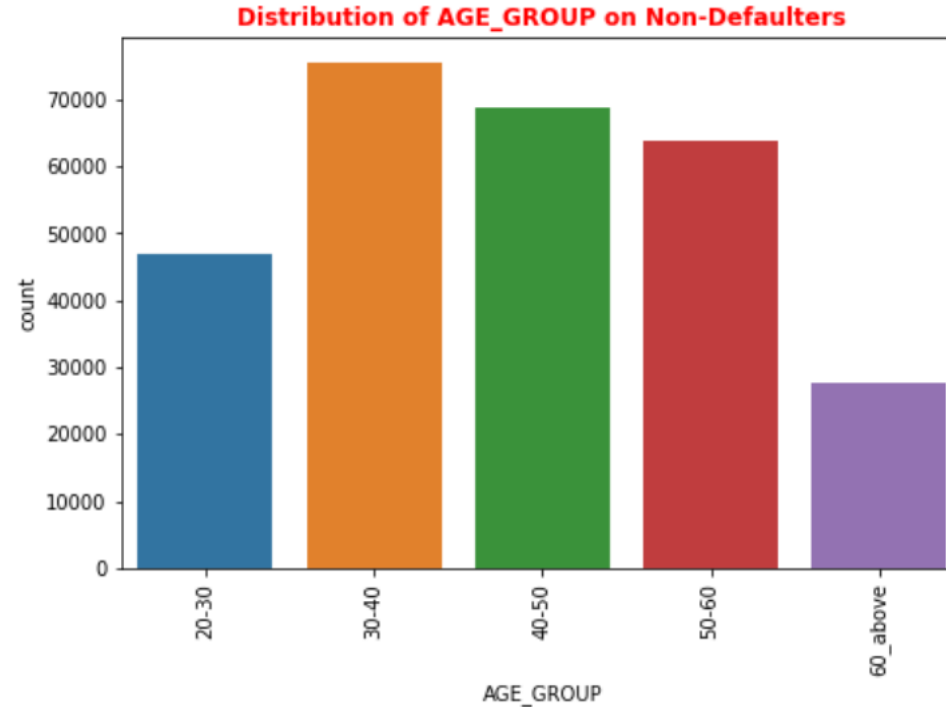
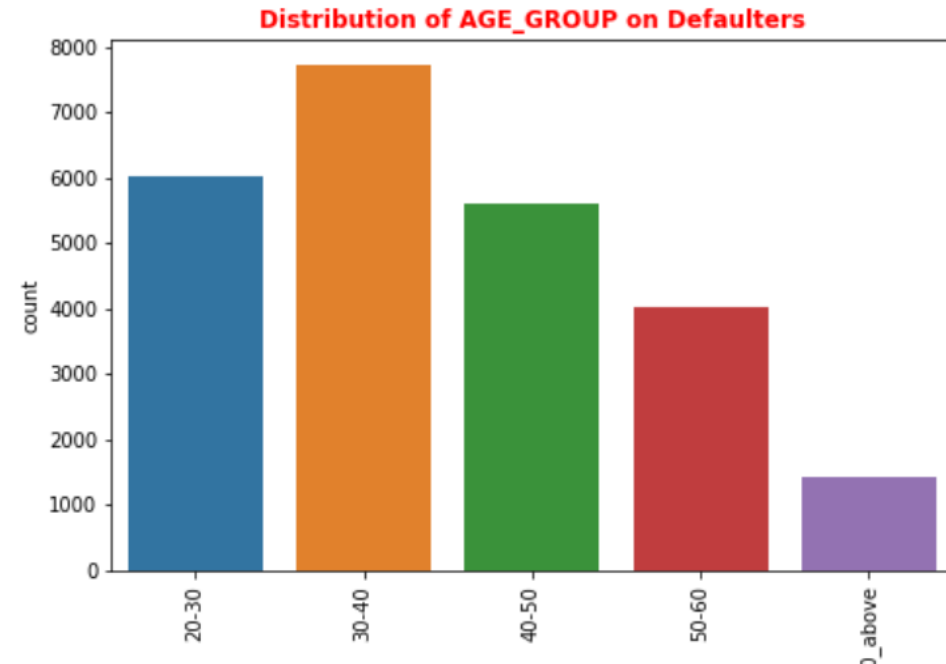
---

# Univariate Analysis For Categorical And Numerical



# Age Group

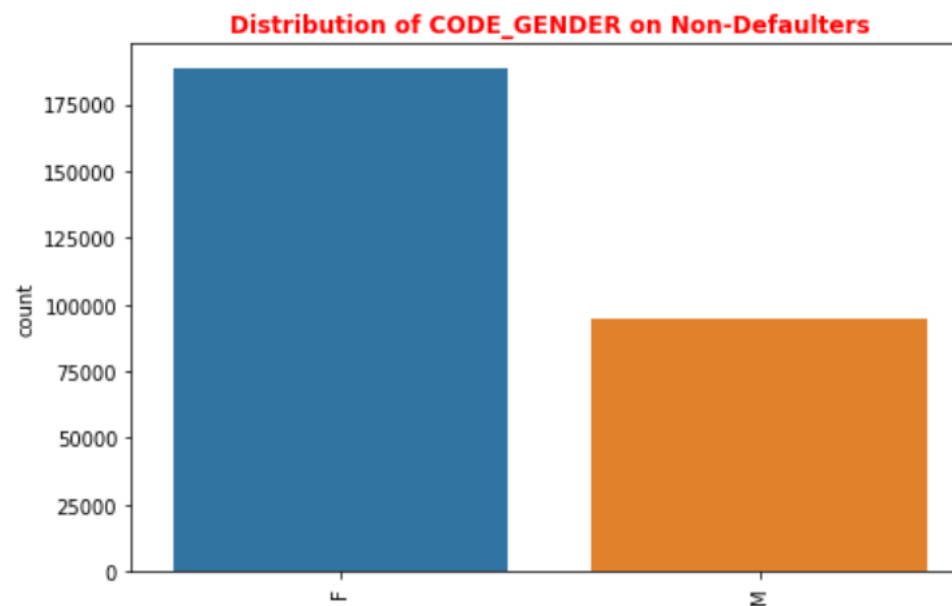
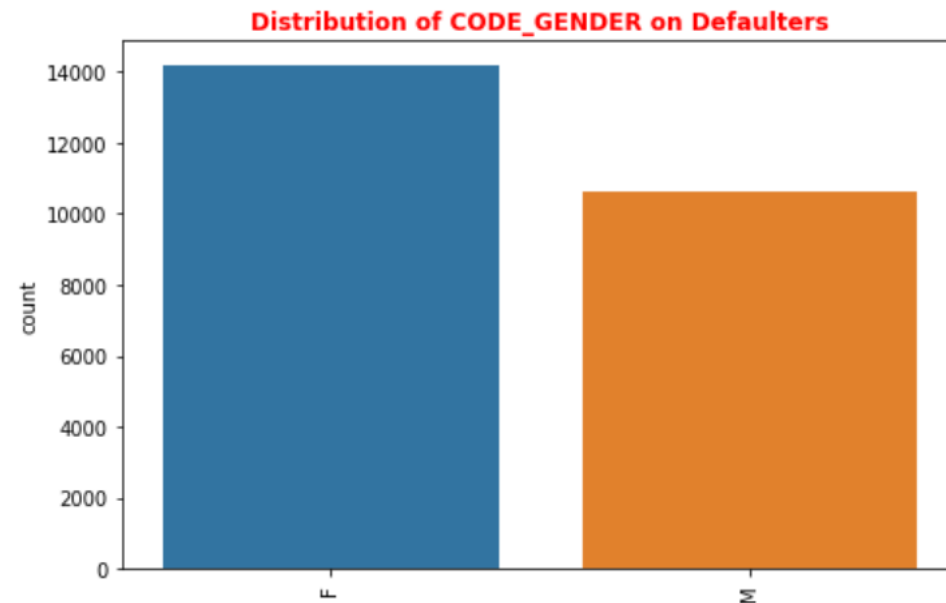
As per the graph, there seems to be an increase in number of people in the age group 20-30 that to the defaulters ( facing payment difficulties) while compared to non-defaulters.



# Gender

Females are the majority in both defaulters and non-defaulter.

Undoubtedly, most female have applied for loan than male.



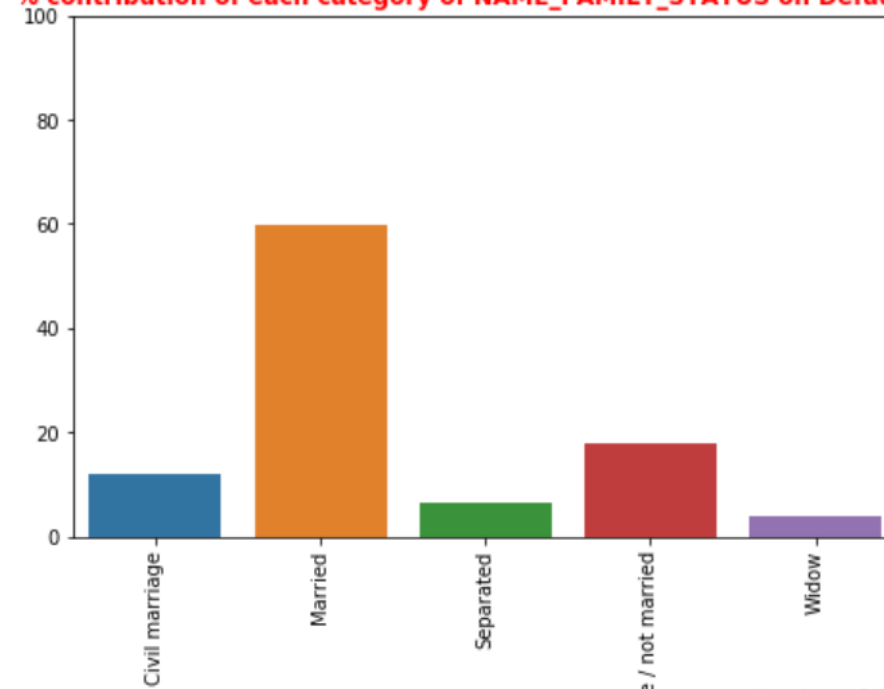


# Family Status

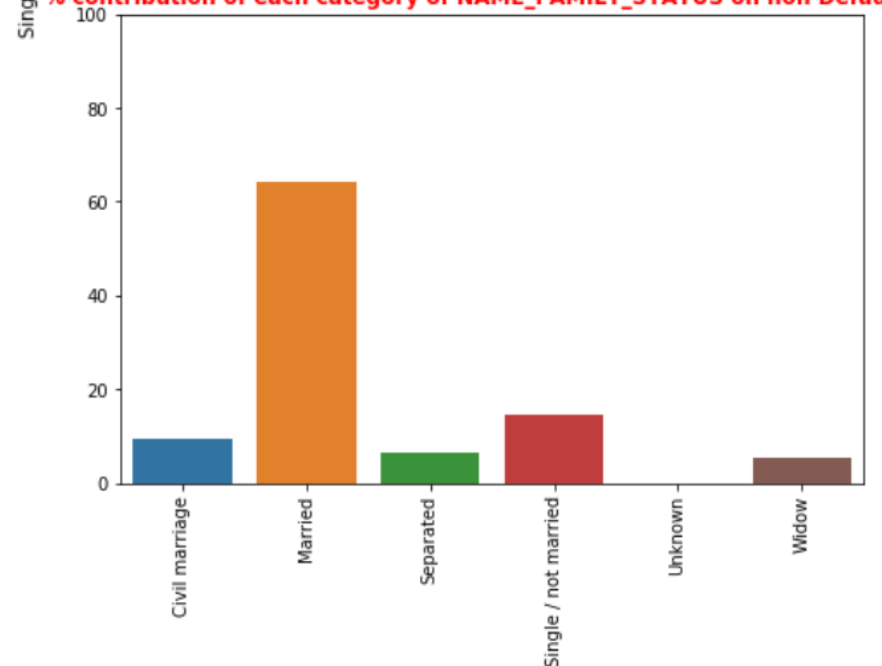
Most of the population that applied for loan are married.

Single / not married is proportionally higher in defaulter population as compared to non defaulted population. This concludes that Single applicants have higher defaults.

% contribution of each category of NAME\_FAMILY\_STATUS on Defaulters



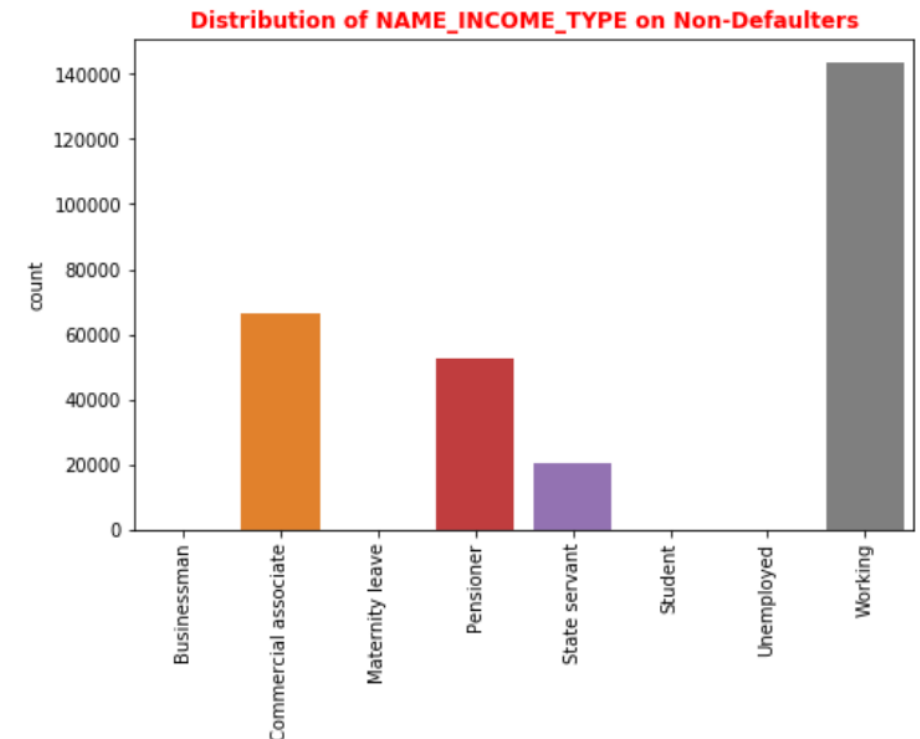
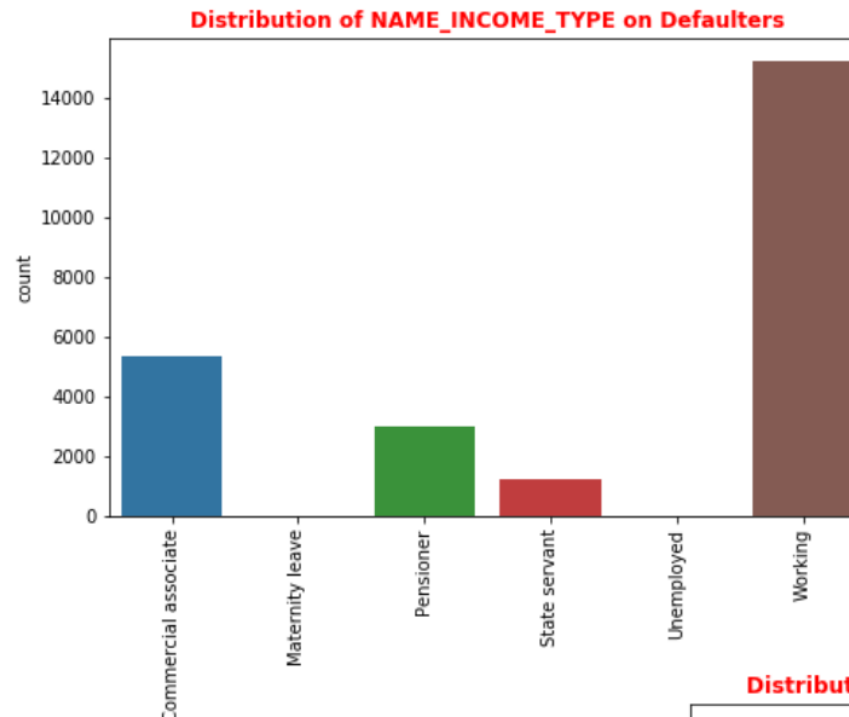
% contribution of each category of NAME\_FAMILY\_STATUS on non Defaulters



# Income Type

It seems like most of the defaulters are the working people, so as in Non-Defaulters. And there seems to be a moderate rate in Commercial associate, Pensioner, and state servant.

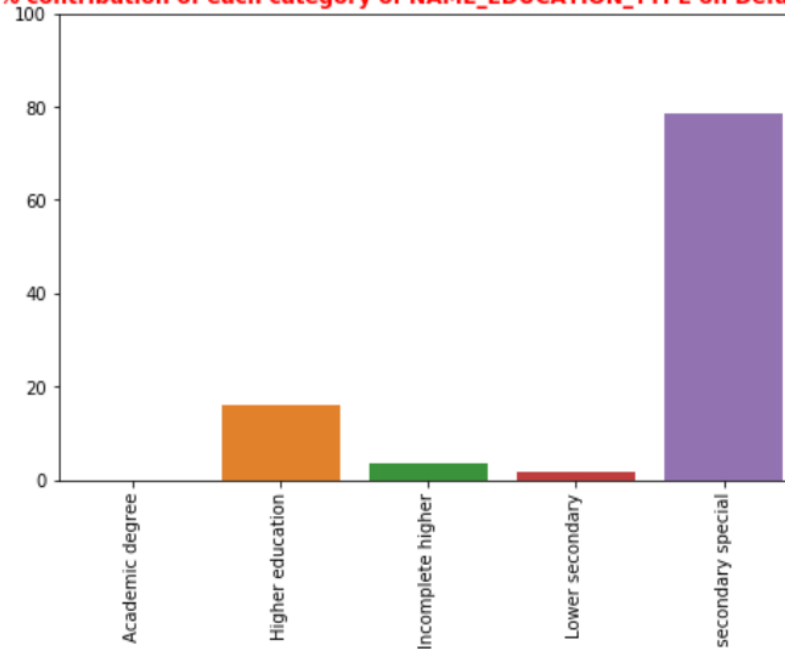
Least count to be seen in Unemployed, Maternity Leave, Businessman and Student.



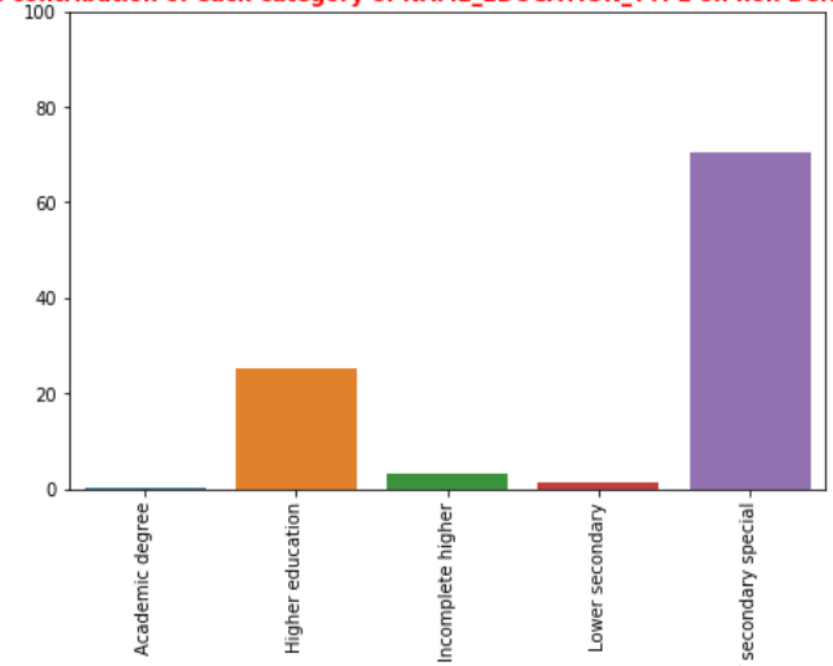
# Education Type

Higher education count is proportionally lesser in defaulted population as compared to non defaulted population. Hence higher the education level, the lower the default rate.

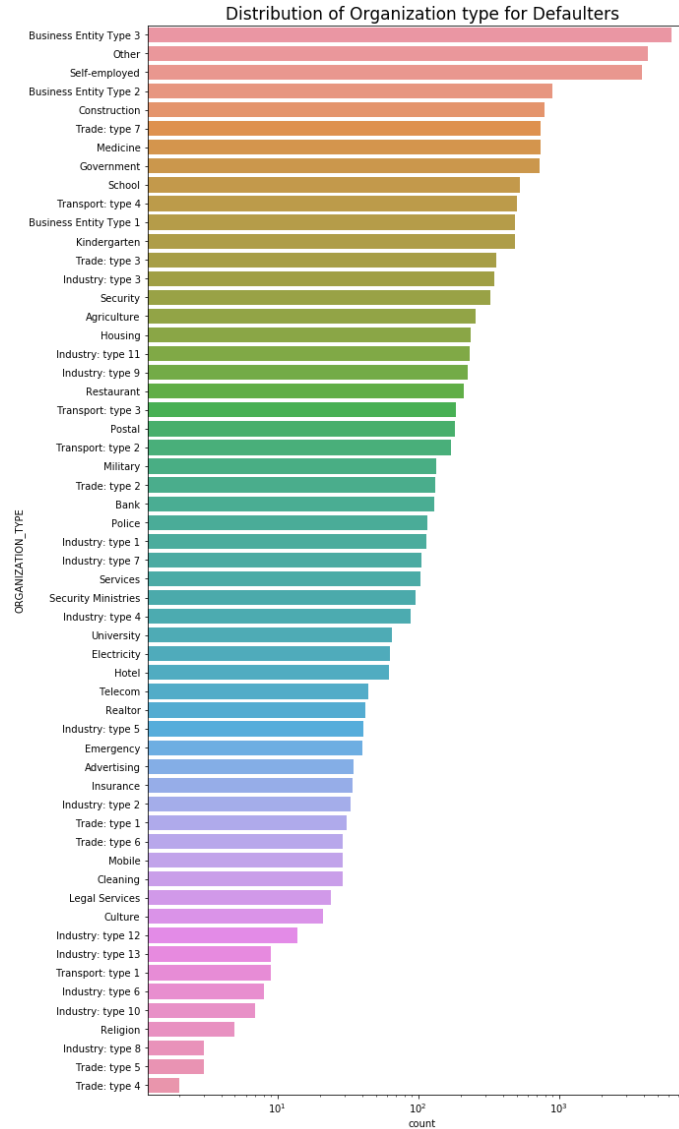
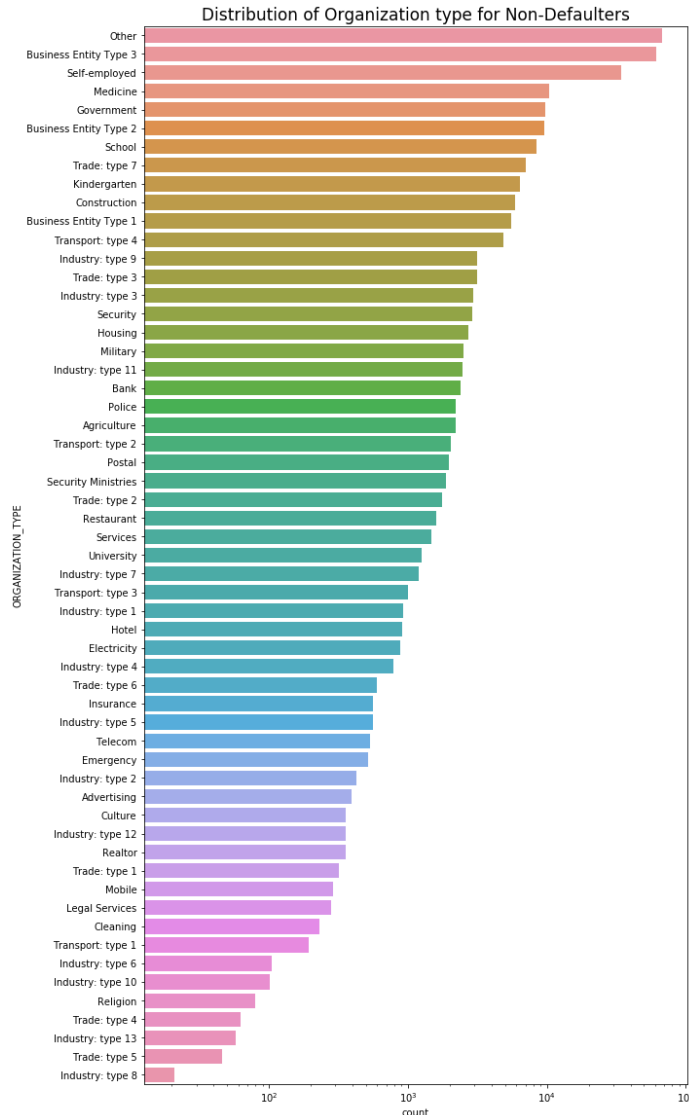
% contribution of each category of NAME\_EDUCATION\_TYPE on Defaulters'



% contribution of each category of NAME\_EDUCATION\_TYPE on non Defaulters



# Organization Type



## For Non - Defaulters:

➤ The highest number of loan applications are from most of the organization type 'Other', 'Business entity Type 3', 'Self-employed', 'Medicine' and 'Government'.

➤ Fewer applications are from Industry type 8, type 5, type 13, type 4, and religion.

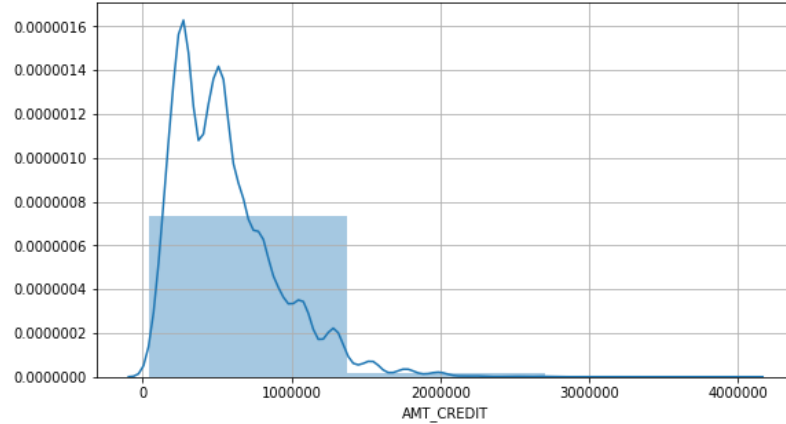
## For Defaulters:

➤ The highest number of loan applications are from most of the organization type 'Business entity Type 3', 'Other', 'Self-employed', 'Business entity Type 2', and 'Construction'.

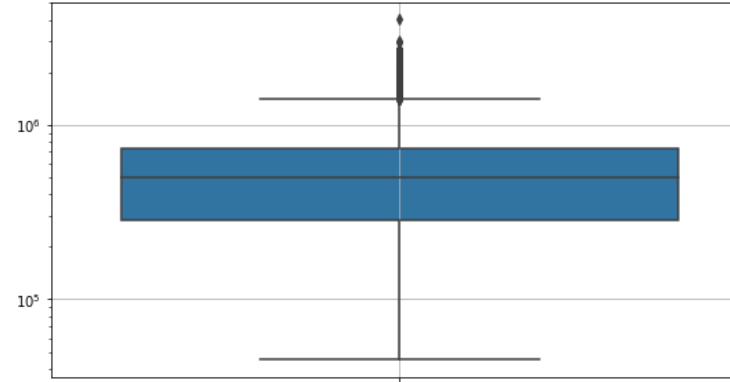
➤ Less applications are from Industry type 4, type 8, type 5, type 10 and religion.

# AMT\_Credit

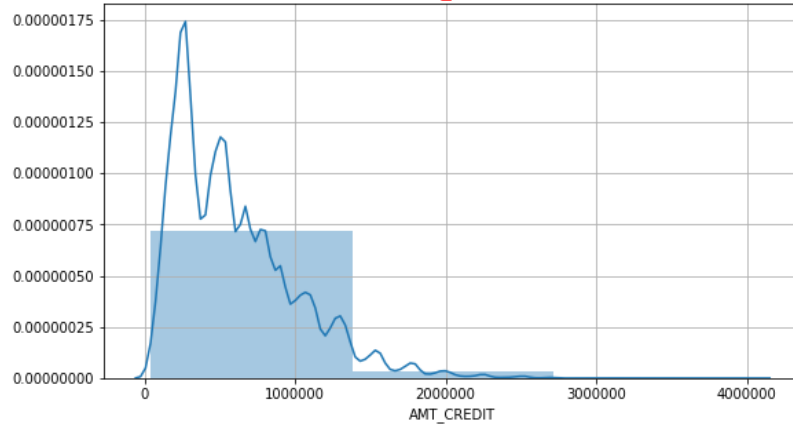
Distribution Plot for AMT\_CREDIT on Defaulters



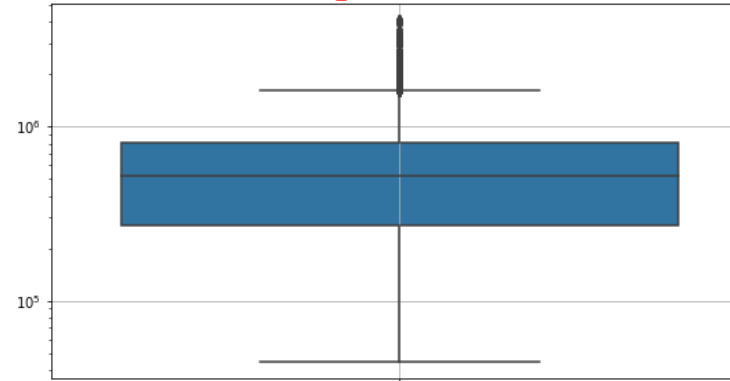
Boxplot for AMT\_CREDIT on Defaulters



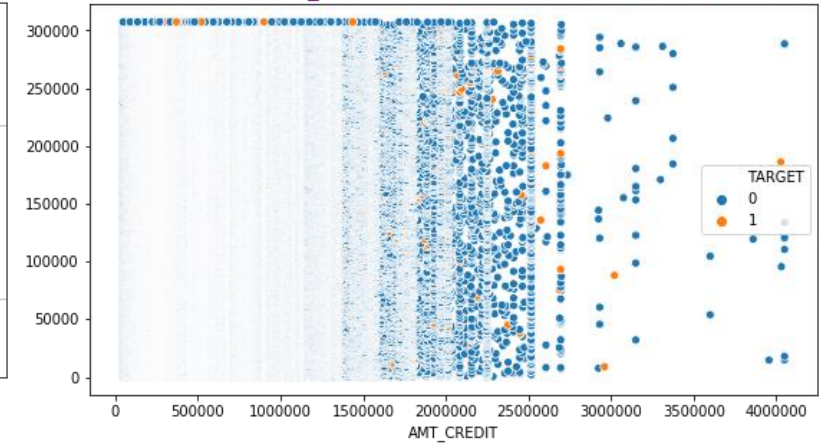
Distribution Plot for AMT\_CREDIT on Non-Defaulters



Boxplot for AMT\_CREDIT on Non-Defaulters



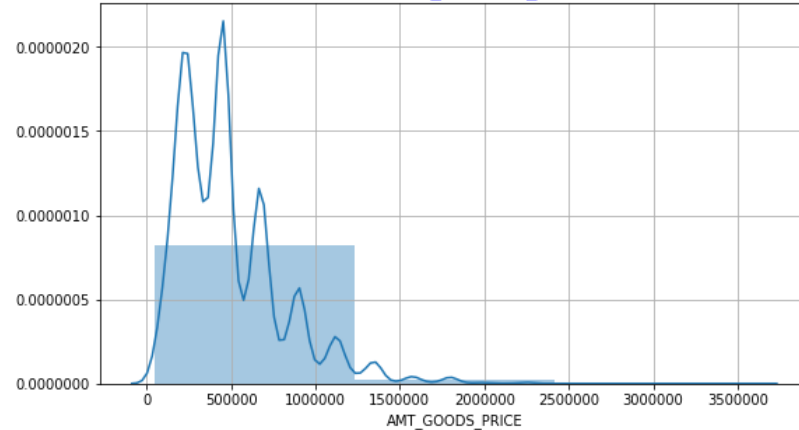
Distribution of AMT\_CREDIT for Defaulters and Non-Defaulters



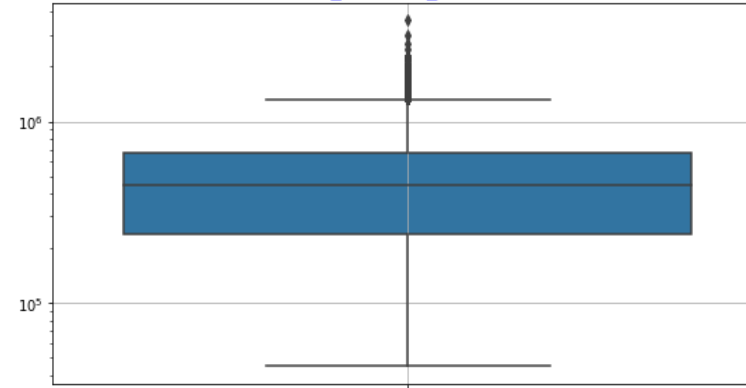
There's a visibility of outliers  
and most of values are in  
between (0.1 lakh to 2 lakh).

# AMT\_GOODS\_Price

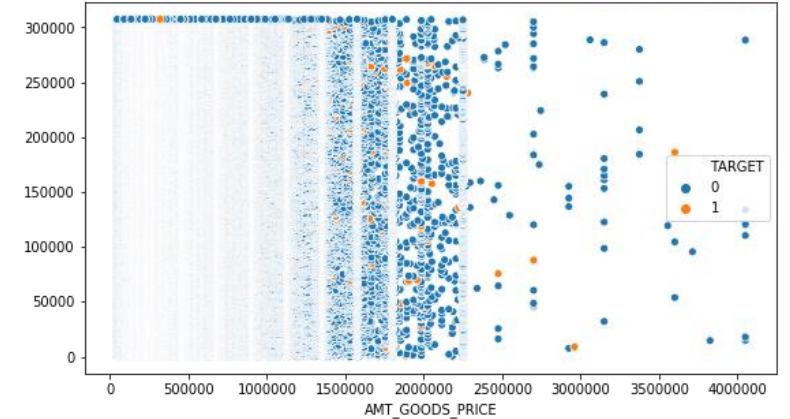
Distribution Plot for AMT\_GOODS\_PRICE on Defaulters



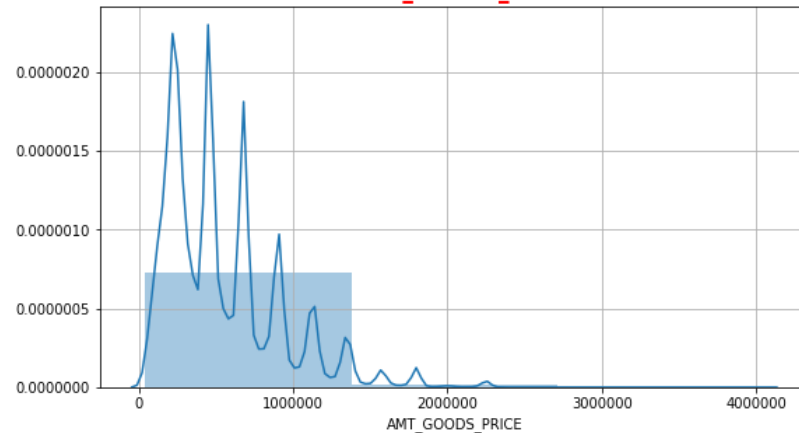
Boxplot for AMT\_GOODS\_PRICE on Defaulters



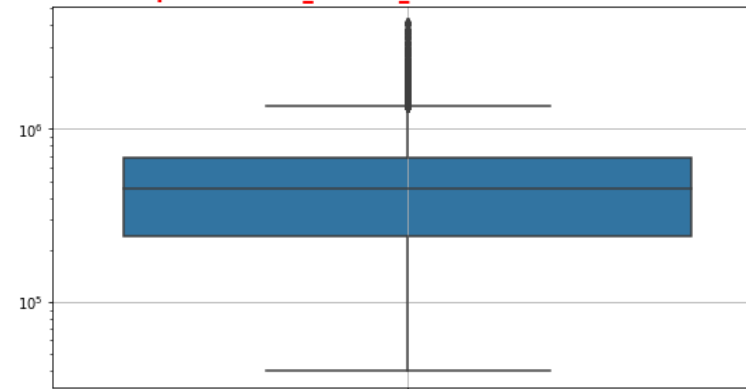
Distribution of AMT\_GOODS\_PRICE for Defaulters and Non-Defaulters



Distribution Plot for AMT\_GOODS\_PRICE on Non-Defaulters



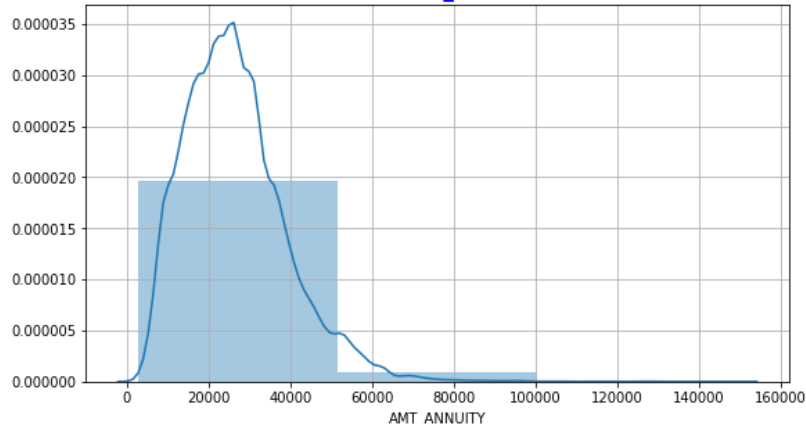
Boxplot for AMT\_GOODS\_PRICE on Non-Defaulters



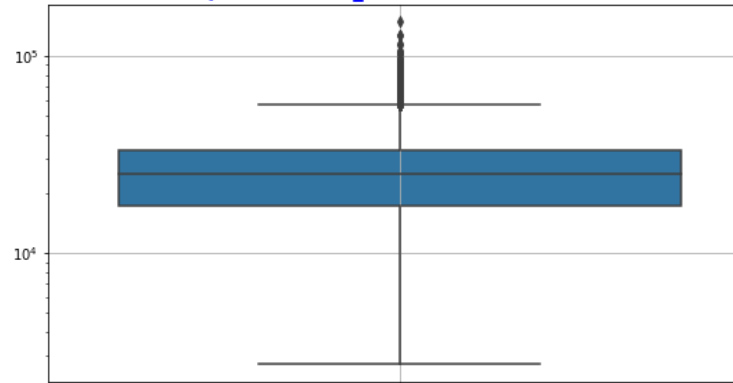
There's a visibility of outliers  
and most of values are up to 1.5  
lakh.

# AMT\_ANNUIITY

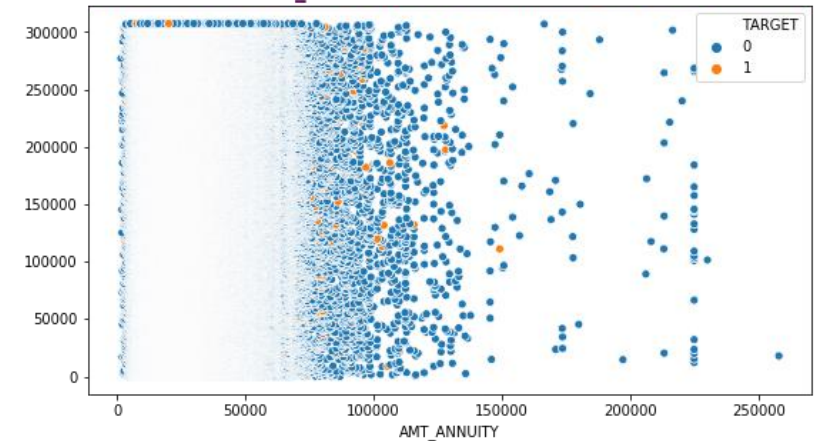
Distribution Plot for AMT\_ANNUIITY on Defaulters



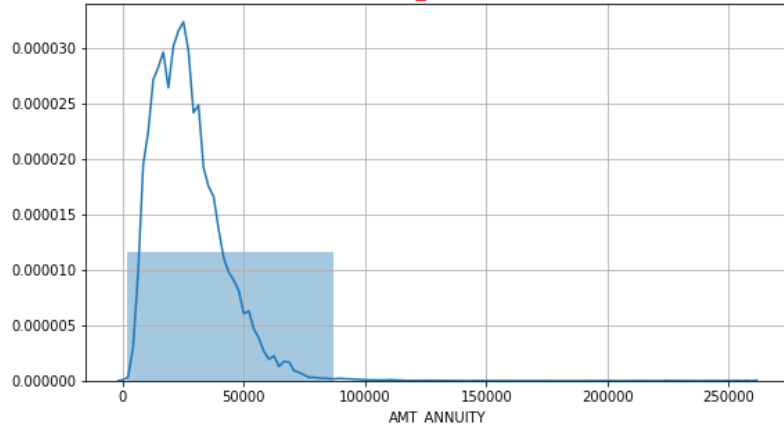
Boxplot for AMT\_ANNUIITY on Defaulters



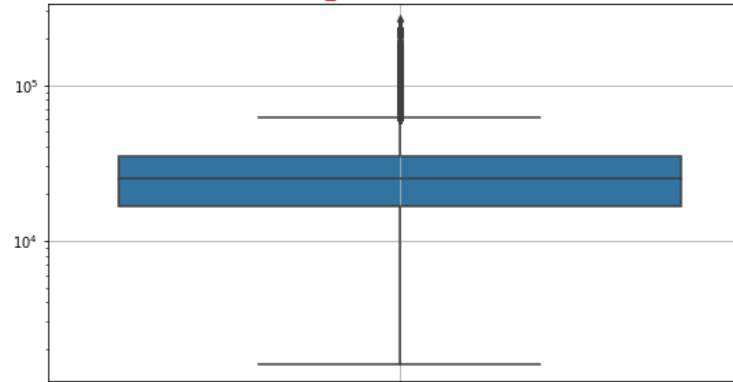
Distribution of AMT\_ANNUIITY for Defaulters and Non-Defaulters



Distribution Plot for AMT\_ANNUIITY on Non-Defaulters



Boxplot for AMT\_ANNUIITY on Non-Defaulters

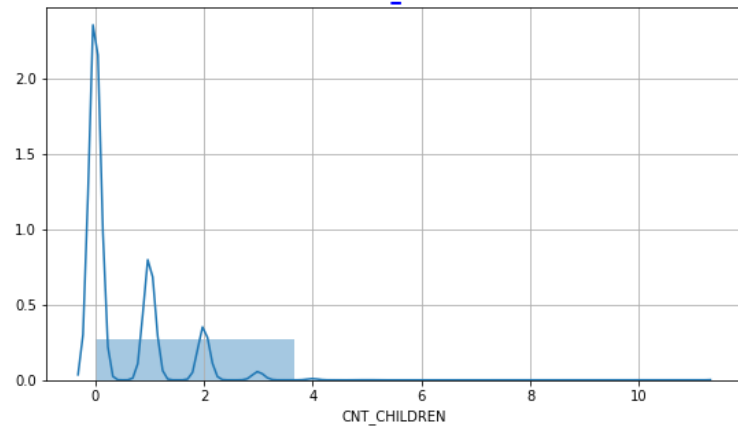


We can observe some outliers and the first quartile is bigger than third quartile for annuity amount which means most of the annuity clients are from first quartile.

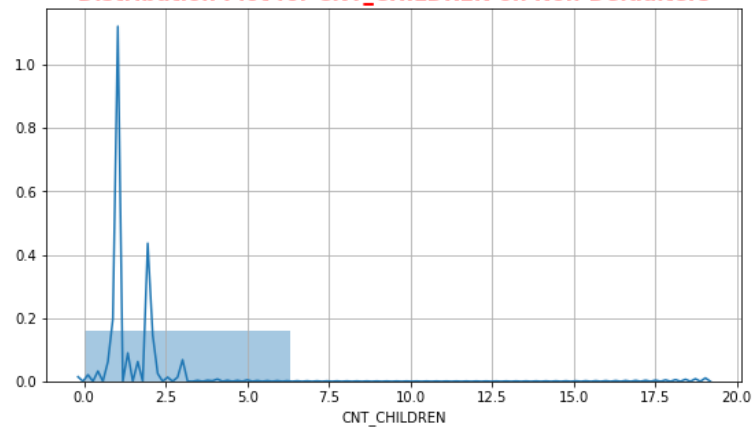


# CNT\_CHILDREN

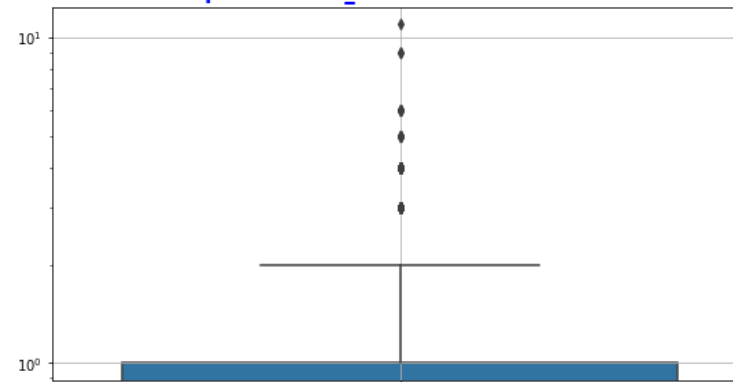
Distribution Plot for CNT\_CHILDREN on Defaulters



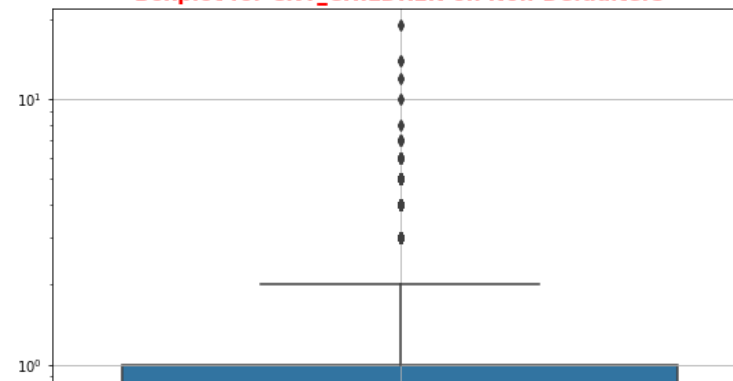
Distribution Plot for CNT\_CHILDREN on Non-Defaulters



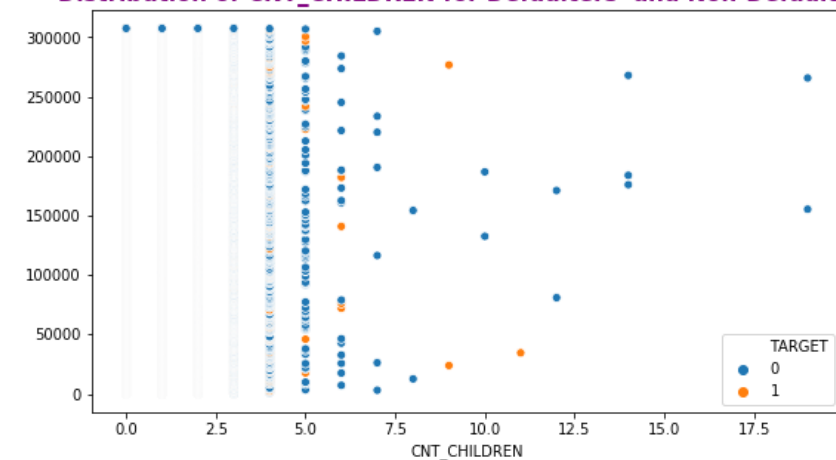
Boxplot for CNT\_CHILDREN on Defaulters



Boxplot for CNT\_CHILDREN on Non-Defaulters



Distribution of CNT\_CHILDREN for Defaulters and Non-Defaulters



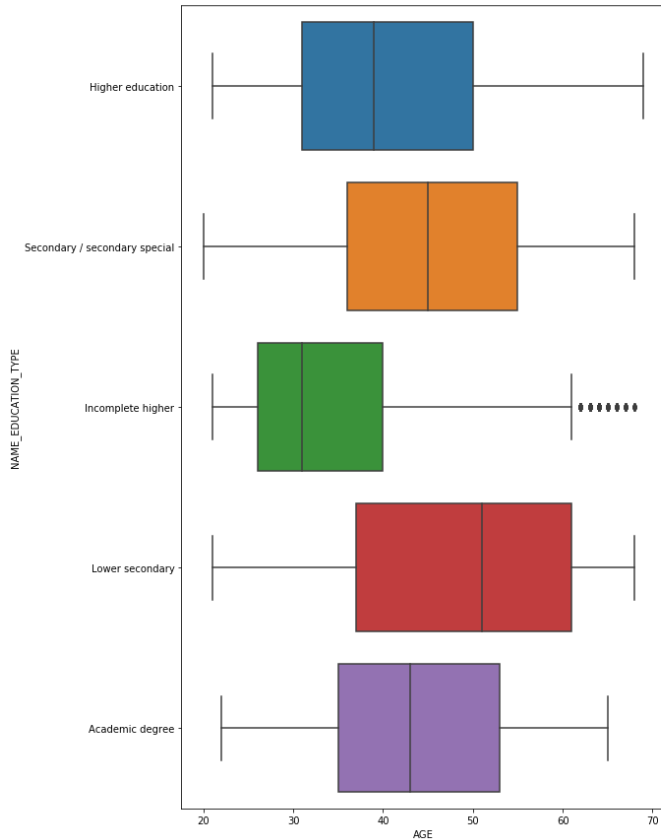
Most people have 1 or 2 child.  
Also, there are few outliers.

---

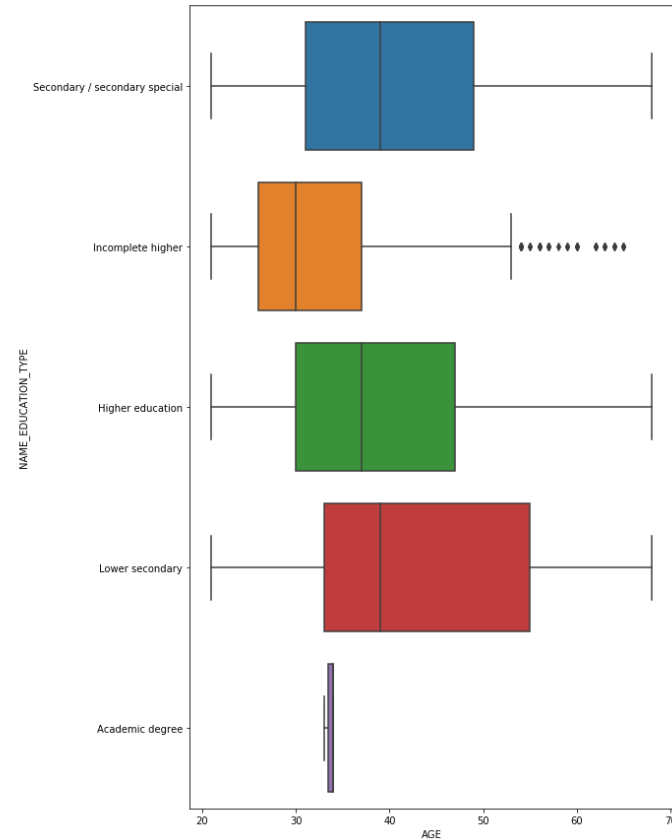
# Bivariate Analysis And Correlation

# Age Vs Education Type

AGE vs NAME\_EDUCATION\_TYPE for Non-Defaulter



AGE vs NAME\_EDUCATION\_TYPE for Defaulter



## For Target = 0

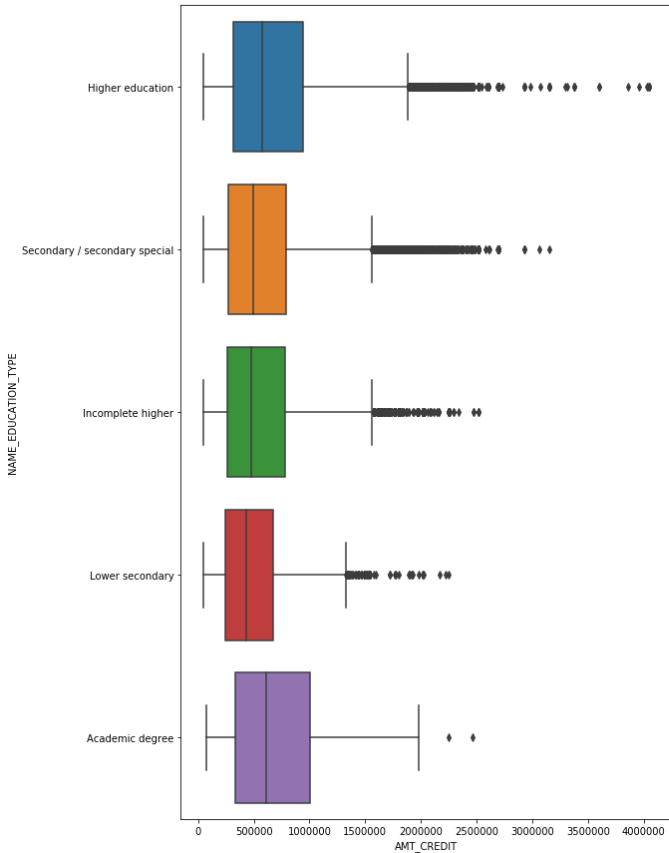
- There is an outliers in Incomplete higher in between age 60-70.
- Apart from Incomplete higher where first quartile lies on age 40, other education types aren't facing much difficulties in loan repayment.

## For Target = 1

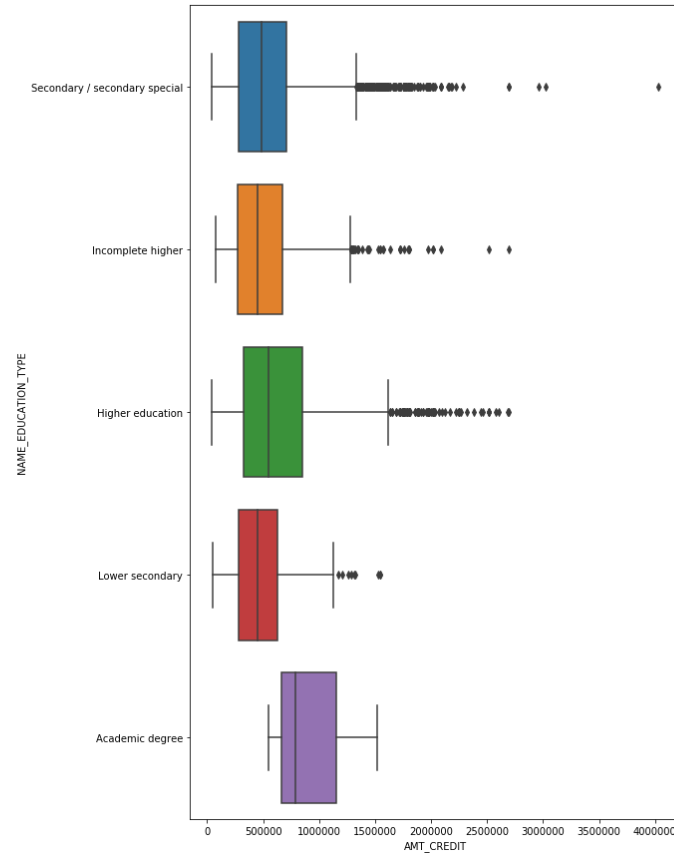
- There is a visibility of an outliers in Incomplete higher from the age group between 50-70.
- People with an age group 30-40 and education type as Academic degree and Incomplete higher seems to be facing difficulties in loan repayment.

# AMT\_CREDIT Vs Education Type

AMT\_CREDIT vs NAME\_EDUCATION\_TYPE for Non-Defaulter



AMT\_CREDIT vs NAME\_EDUCATION\_TYPE for Defaulter



## Non-Defaulter

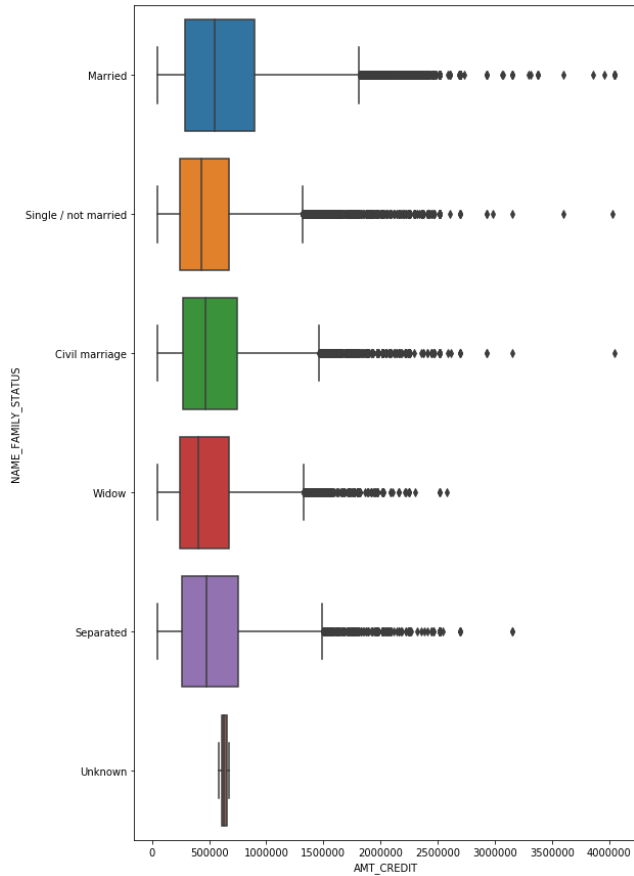
- Academic degree education are having higher credit amount than others.
- Higher education are having more outliers.
- Academic degree is having most of the credits .

## Defaulter

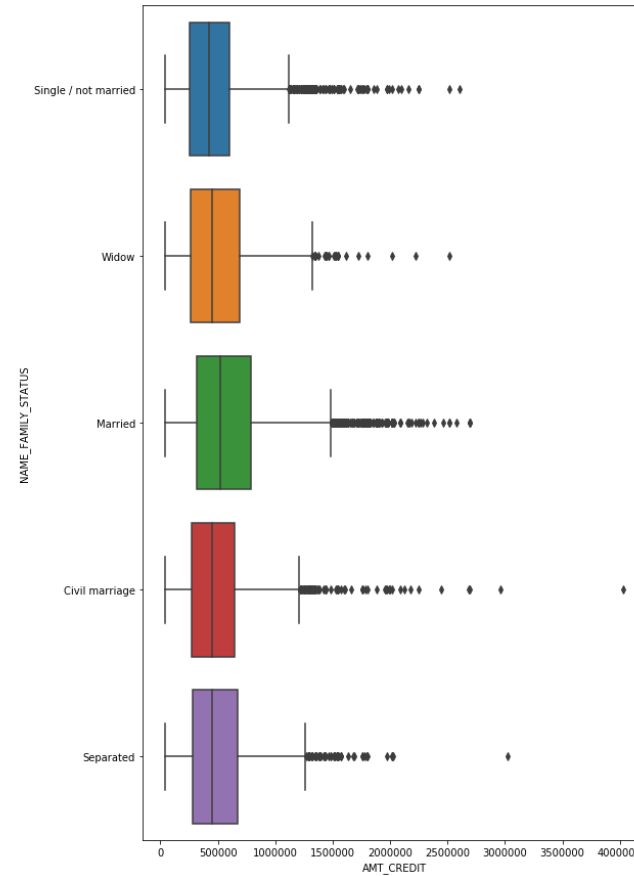
- Academic degree education have higher credit amount than others.
- Higher education seems to have more outliers.
- Academic degree has most credit amount .
- Academic education category has a smaller number of defaulters.

# AMT\_CREDIT Vs Family Status

AMT\_CREDIT vs NAME\_FAMILY\_STATUS for Non-Defaulter



AMT\_CREDIT vs NAME\_FAMILY\_STATUS for Defaulter



## Non-Defaulter

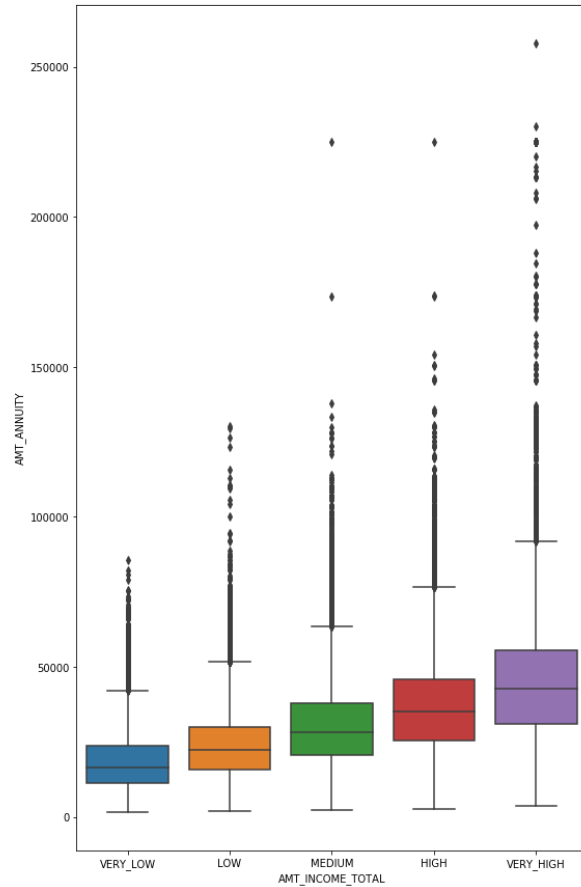
- Married are having higher credit amount than others.
- widow are having less outliers.

## Defaulter

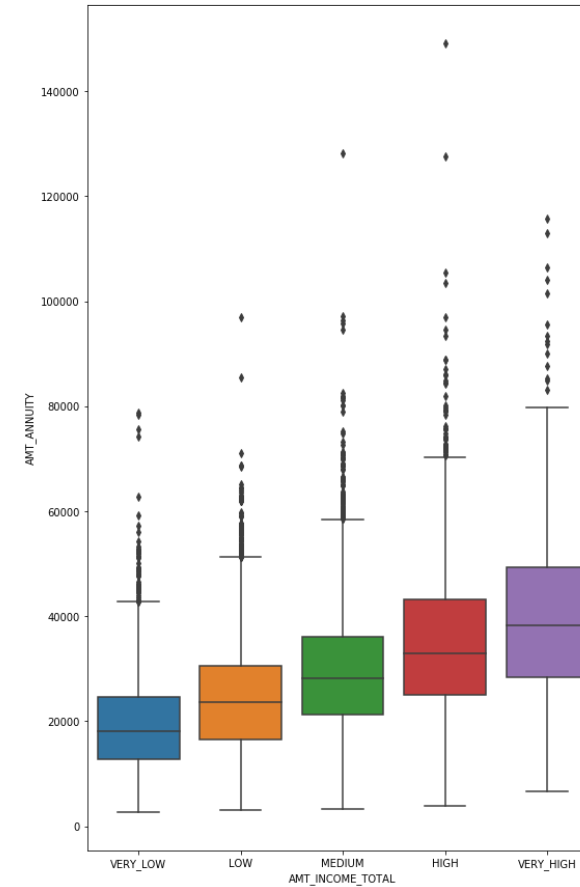
- Married are having higher credit amount than others.
- Civil Marriage seems to have more outliers.
- Civil marriage has most credit amount .
- Defaulters are almost equal to every category.

# AMT\_ANNUITY Vs Income

AMT\_INCOME\_TOTAL vs AMT\_ANNUITY for Non-Defaulter



AMT\_INCOME\_TOTAL vs AMT\_ANNUITY for Defaulter



## Non-Defaulter

- Very High from the amount income total is having more annuity compare to all remaining categories.
- The same category 'very high' is having outlier which is greater than the others.

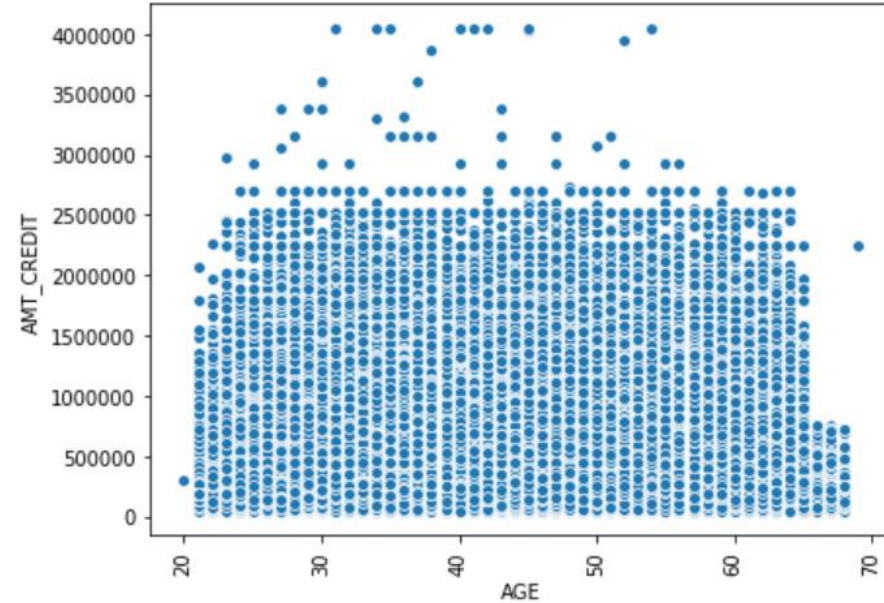
## Defaulter

- Very High from the amount income total is having more annuity compare to all remaining categories.
- The same category 'high' is having outlier which is greater than the others .

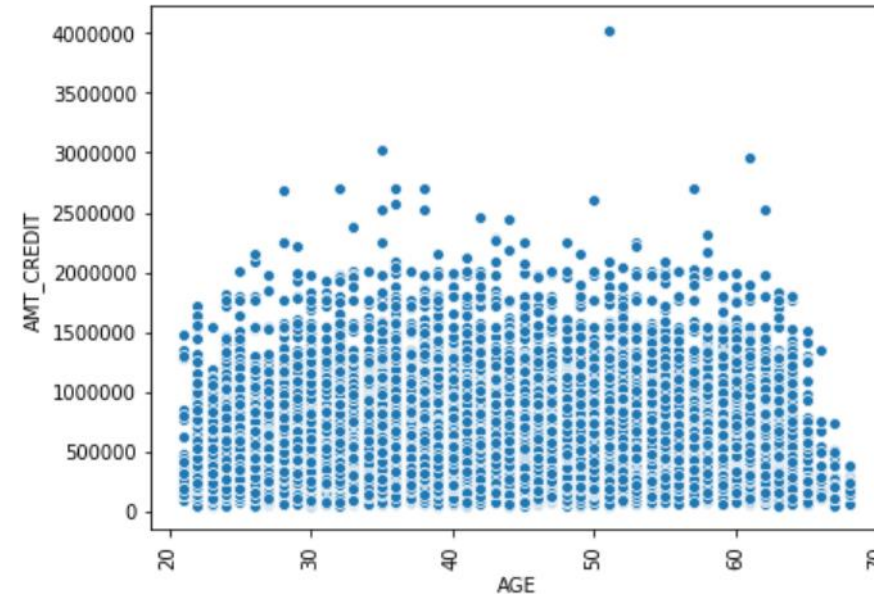
# Age Vs Amt\_Credit

- In Non-Defaulter and Defaulter, the credit amount seems to be less i.e., 50k-100 in the age group of 60-70 while comparing to other age group.
- In defaulters, credit amount seems to be less around 300k in comparison to non- defaulter.

AGE vs AMT\_CREDIT for Non-Defaulter



AGE vs AMT\_CREDIT for Defaulter

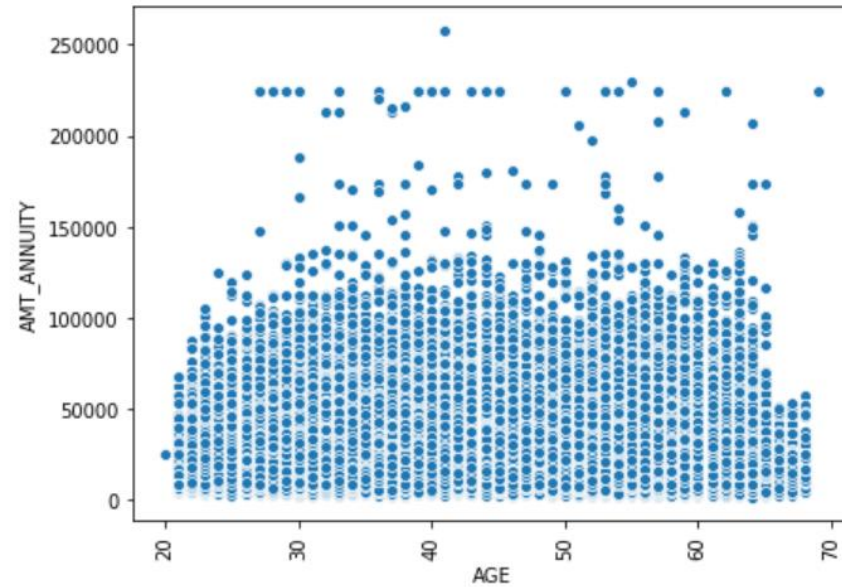




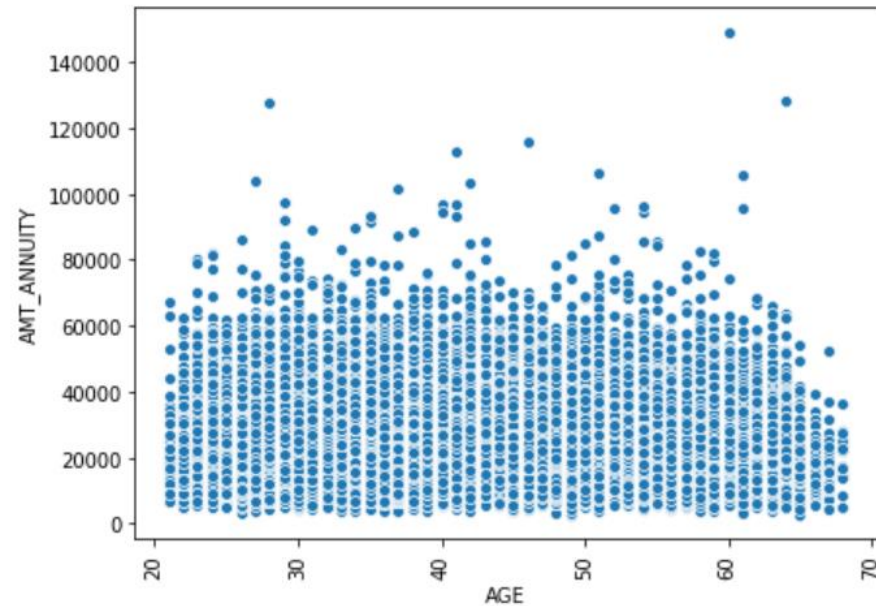
# Age Vs Amt\_Annuity

- People of age group 60-70 have less annuity amount in both cases.
- Annuity amount seems to be higher in Non-Defaulters.

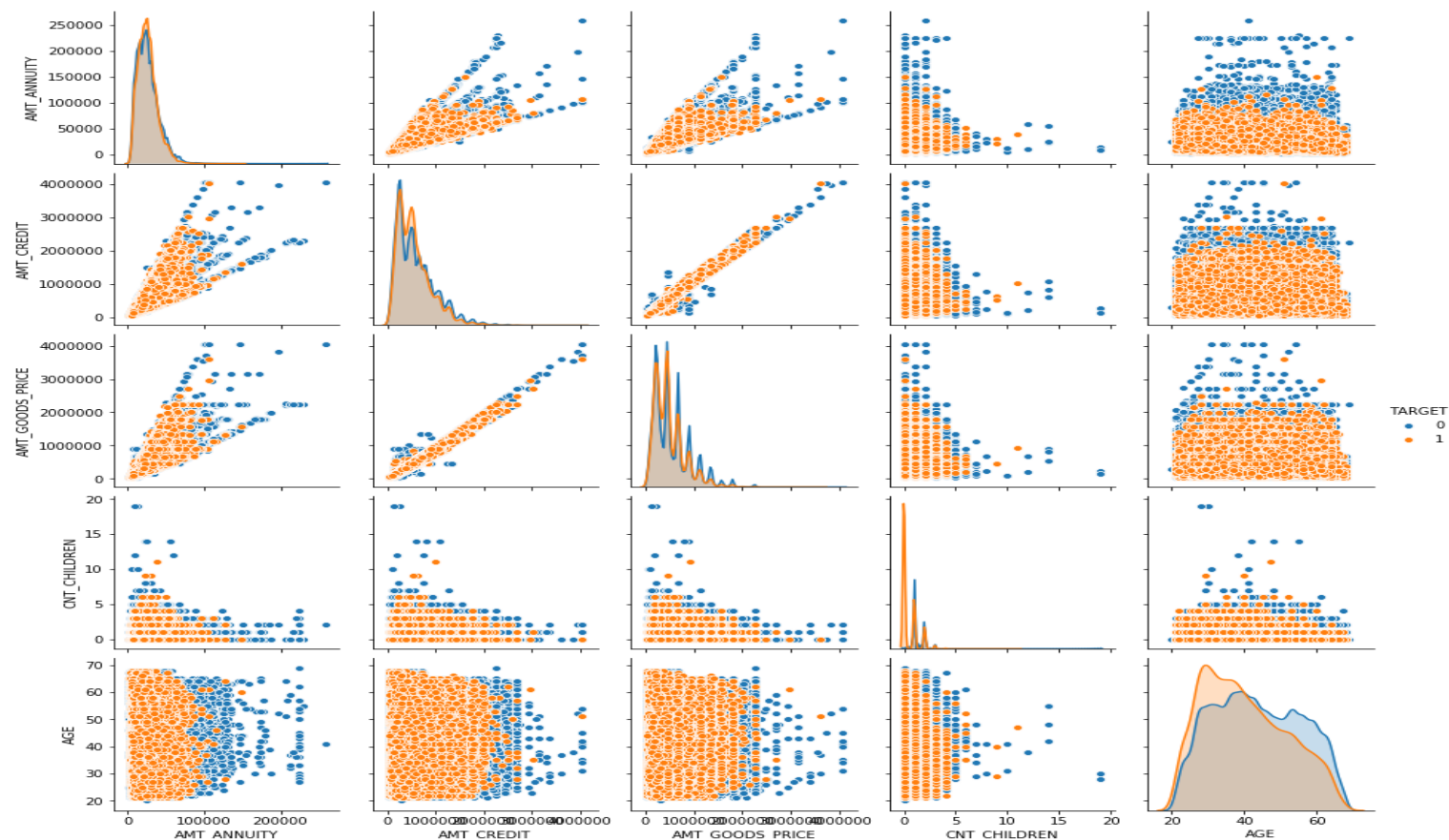
AGE vs AMT\_ANNUIITY for Non-Defaulter



AGE vs AMT\_ANNUIITY for Defaulter



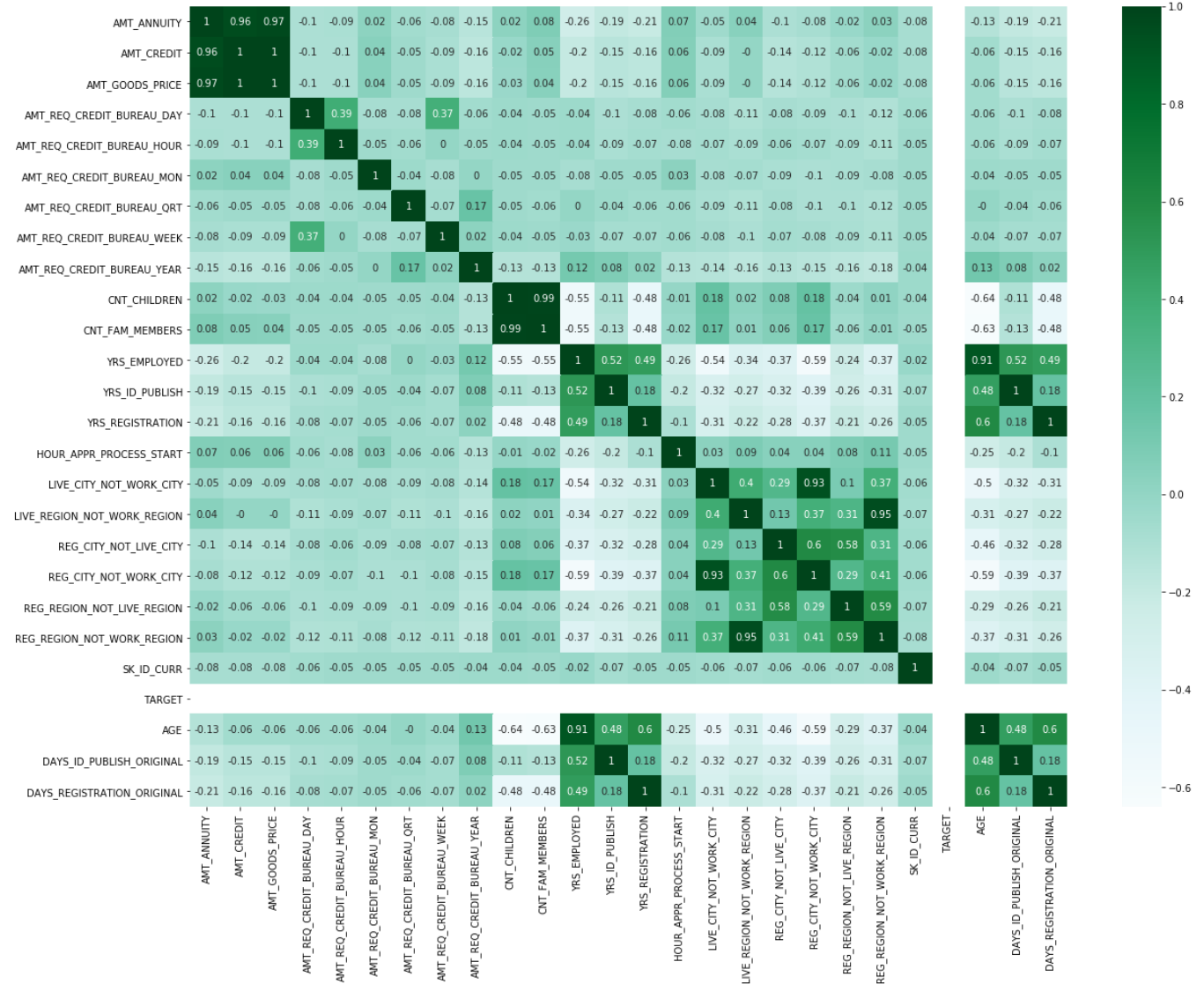
# Pair plot with Hue as Target



# Correlation for Target 0

## Top 10 Correlation for Non-Defaulter

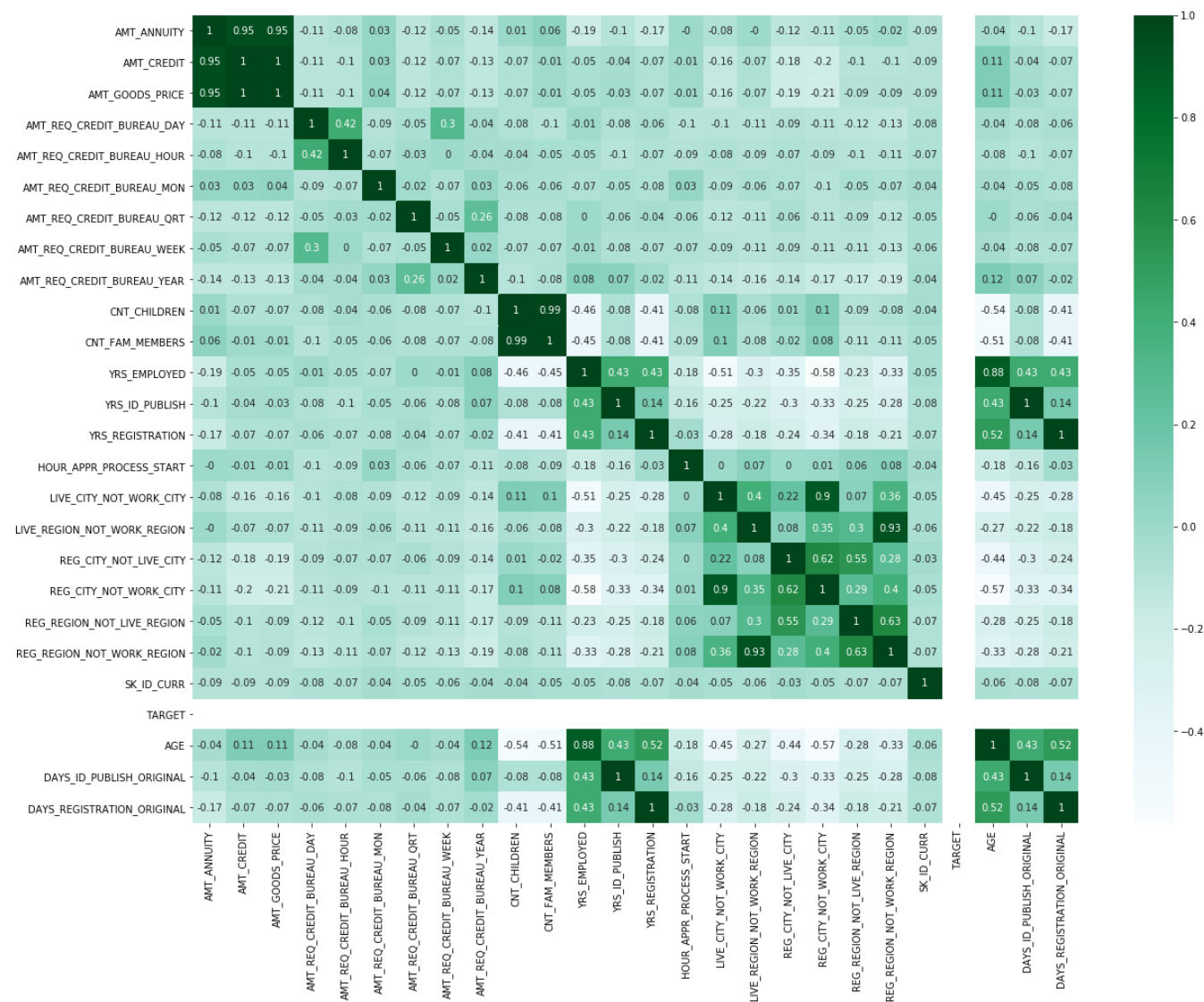
- ('AMT\_GOODS\_PRICE', 'AMT\_CREDIT')
- ('CNT\_CHILDREN', 'CNT\_FAM\_MEMBERS')
- ('AMT\_ANNUITY', 'AMT\_GOODS\_PRICE')
- ('AMT\_ANNUITY', 'AMT\_CREDIT')
- ('LIVE\_REGION\_NOT\_WORK\_REGION', 'REG\_REGION\_NOT\_WORK\_REGION')
- ('REG\_CITY\_NOT\_WORK\_CITY', 'LIVE\_CITY\_NOT\_WORK\_CITY')
- ('AGE', 'YRS\_EMPLOYED')
- ('AGE', 'CNT\_CHILDREN')
- ('CNT\_FAM\_MEMBERS', 'AGE')
- ('AGE', 'DAYS\_REGISTRATION\_ORIGINAL')



# Correlation for Target 1

### Top 10 Correlation for Non-Defaulter

- ('AMT\_CREDIT', 'AMT\_GOODS\_PRICE')
- ('CNT\_CHILDREN', 'CNT\_FAM\_MEMBERS')
- ('AMT\_ANNUITY', 'AMT\_CREDIT')
- AMT\_GOODS\_PRICE', 'AMT\_ANNUITY')
- ('LIVE\_REGION\_NOT\_WORK\_REGION', 'REG\_REGION\_NOT\_WORK\_REGION')
- ('LIVE\_CITY\_NOT\_WORK\_CITY', 'REG\_CITY\_NOT\_WORK\_CITY')
- ('YRS\_EMPLOYED', 'AGE')
- ('REG\_REGION\_NOT\_WORK\_REGION', 'REG\_REGION\_NOT\_LIVE\_REGION')
- ('REG\_CITY\_NOT\_LIVE\_CITY', 'REG\_CITY\_NOT\_WORK\_CITY')
- ('YRS\_EMPLOYED', 'REG\_CITY\_NOT\_WORK\_CITY')

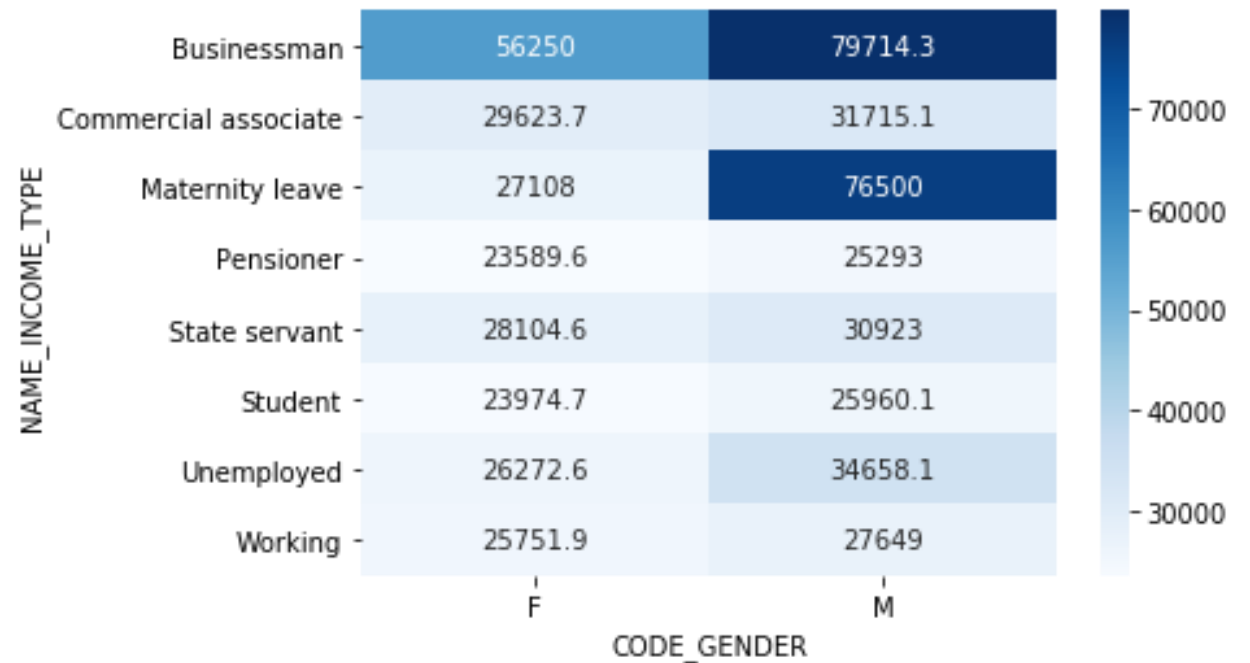


---

# Multivariate Analysis

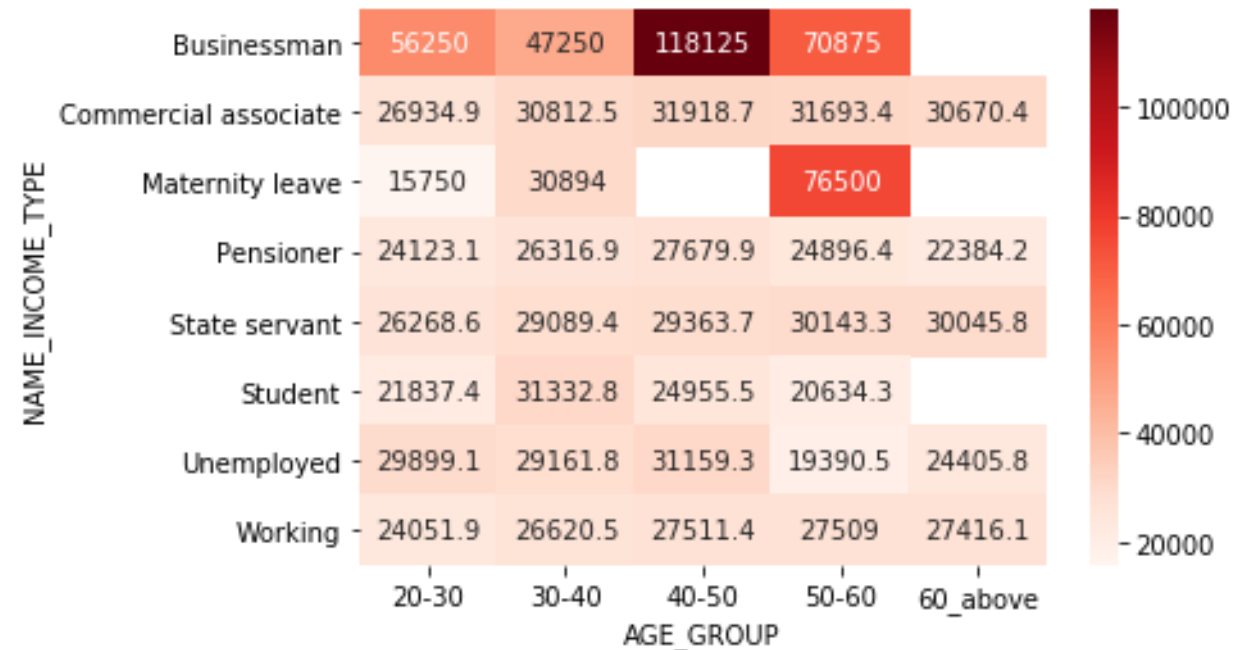
# Income Type, Gender and Annuity

Both Male and Female who are into Business have high annuity and comparatively Male have high annuity. Both Male and Female who are students have very less annuity and comparatively Female has less annuity.



# Income Type, Age and Annuity

Age group of 40-50 years who  
are into business have very high  
annuity.





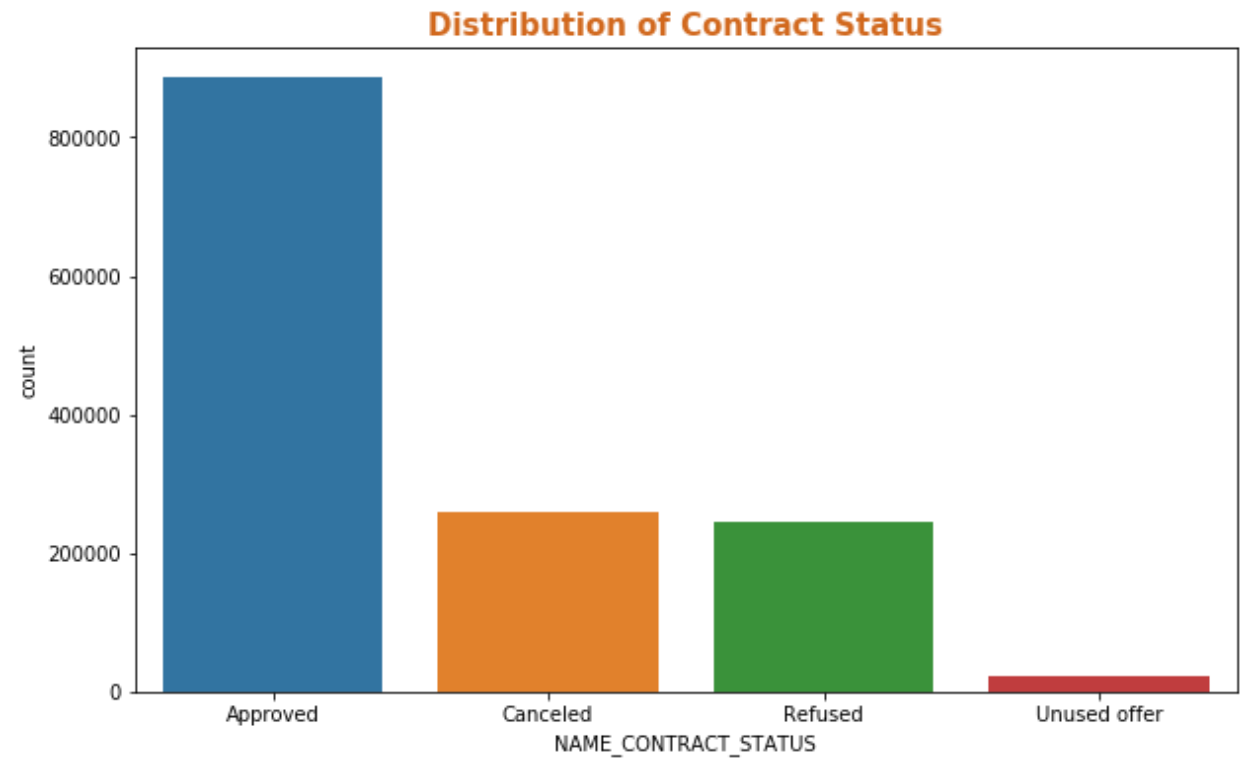


# Inferences from Previous Application Data

---

# Contract Status

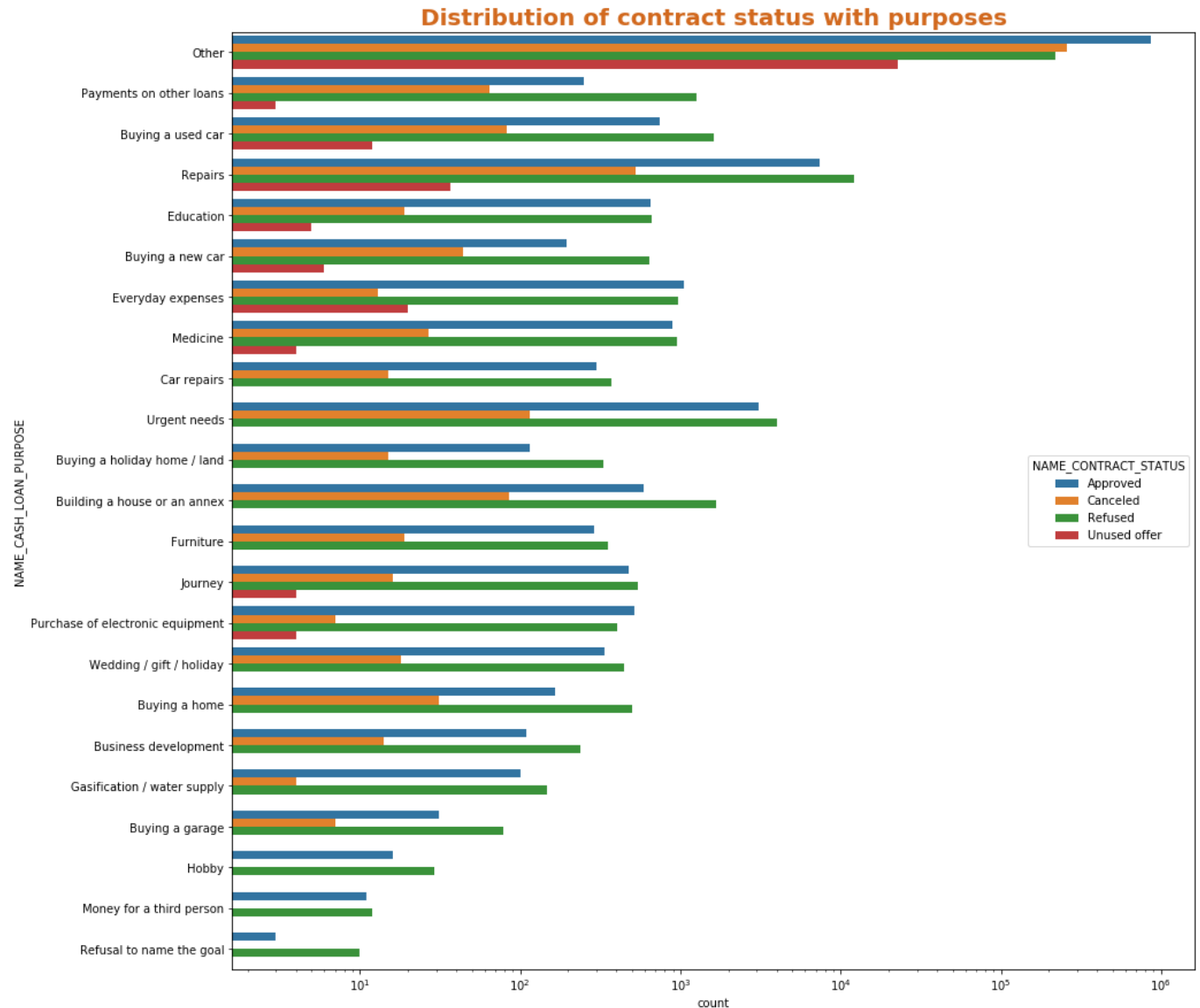
There seems to high amount applications approved.



# Distribution of Contract Status with Purposes

Loan payment is significantly higher than facing difficulties in some scenarios i.e., 'Buying a garage', 'Business development', 'Buying land', 'Buying a new car' and 'Education'.

Loan purposes with 'Repairs' are facing more difficulties in payment on time.

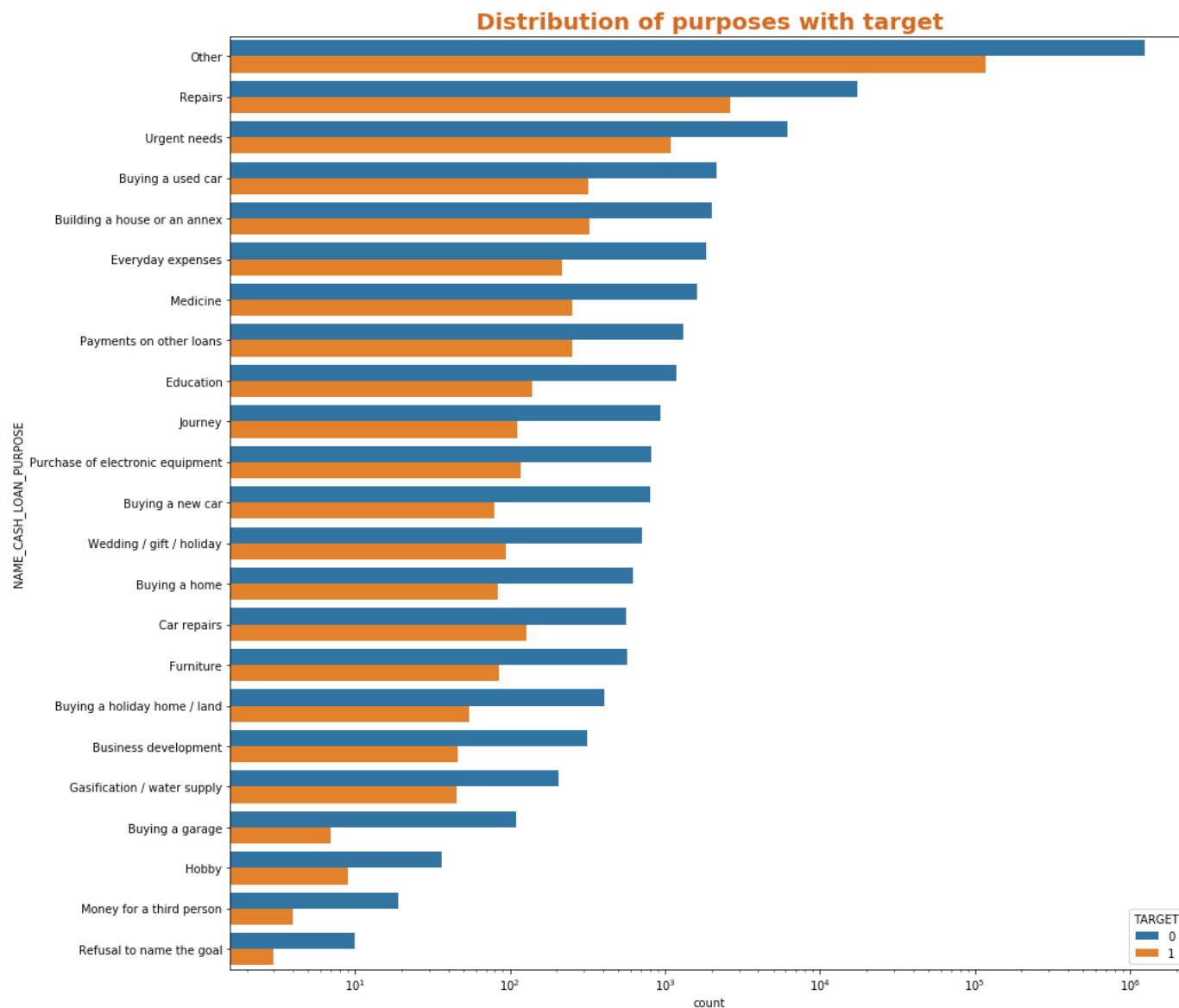


# Distribution of Purposes with Target

Most of the refused applications is from Repairs.

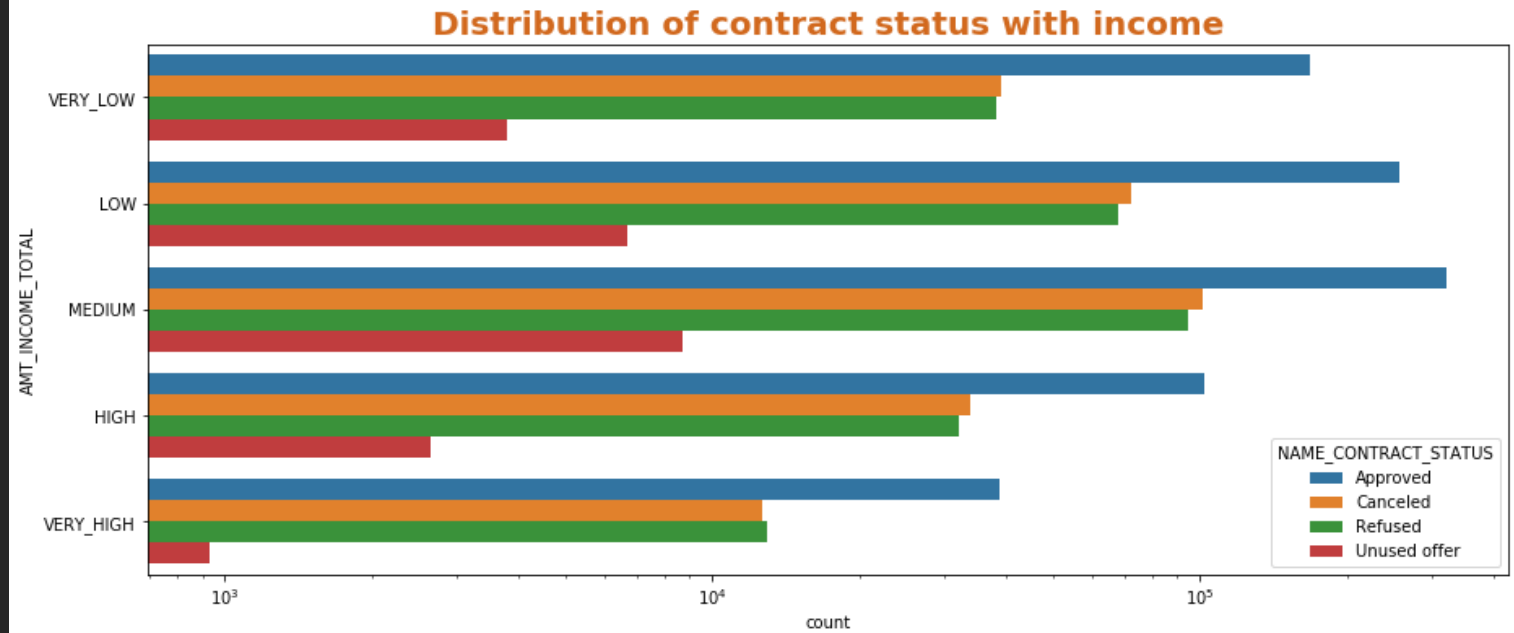
In purchase of electronic equipment and everyday expenses, approved application status is more than refused one.

Approved and Refused status seems to be almost equal in education.



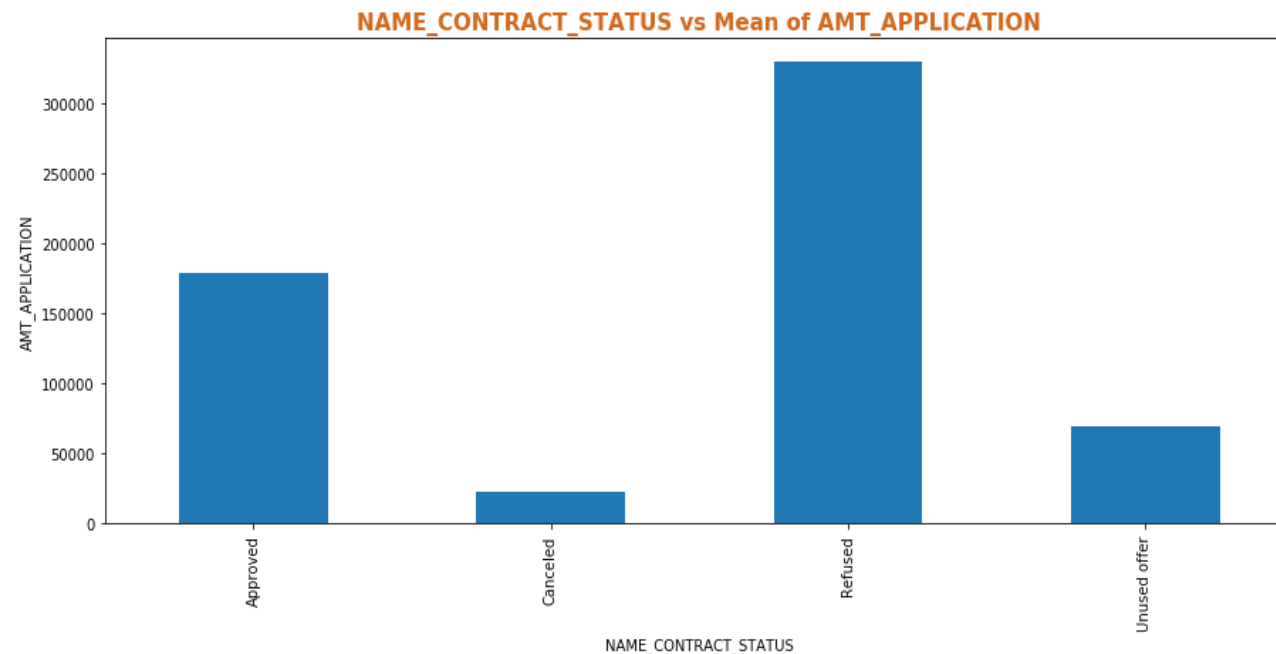
# Distribution of Contract Status with Income

People having high income have most of the applications approved. People having very high income have same number of canceled or refused status.



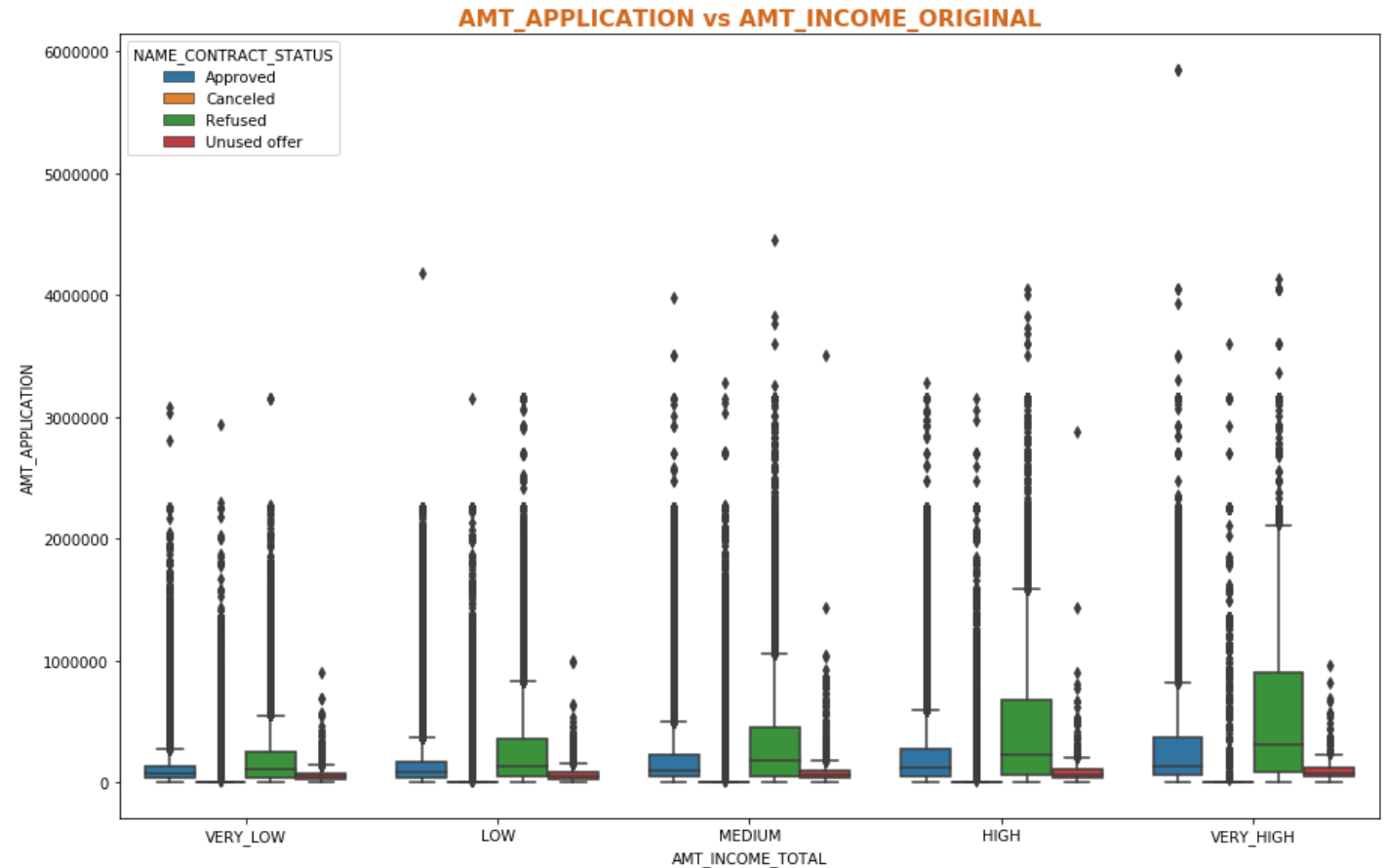
# Contract Status Vs Mean Amt\_Application

In the above plot, application amount is quite high (300k) in the loan application status that are refused.



# Amt\_Application Vs Income Total

- As high the amount application, a greater number of refused applications.
- There are a smaller number of rejected applications for Income category very low.





# Conclusion

---

- Bank should refuse loan application if loan amount is higher than 300k and income is less than 100k.
- Banks should focus less on income type 'Working' as they are having the greatest number of unsuccessful payments.
- Loan purpose 'Repair' is having higher number of unsuccessful payments on time.
- Bank should refuse loan application of Labors because they found as most defaulter.
- Banks should focus more on contract type 'Student' , 'pensioner' and 'Businessman' with housing 'type other than 'Co-op apartment' for successful payments.

**THANK YOU!**