



# Defaulters vs non Defaulters

APPLICATION DATA  
ANALYSIS

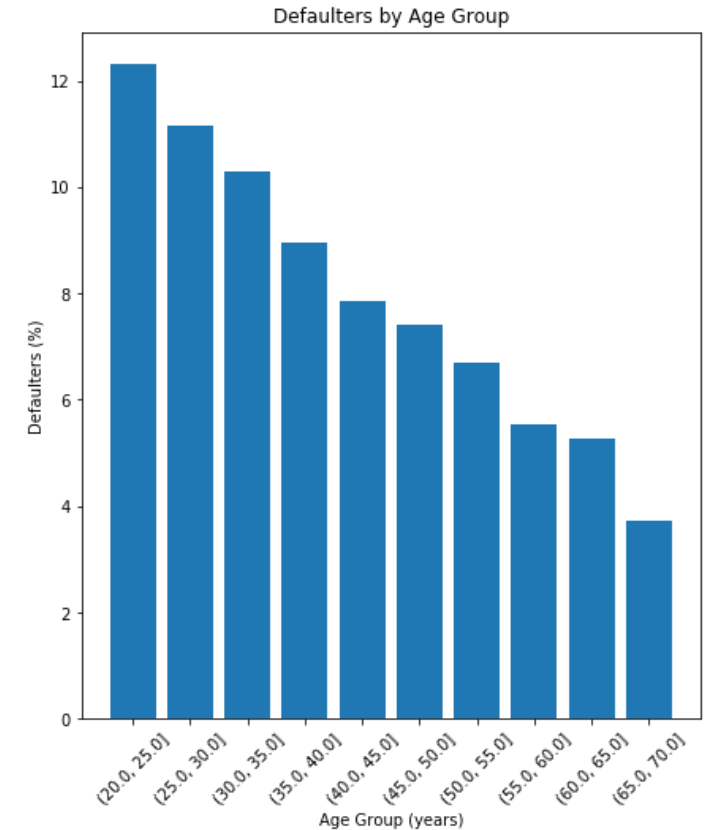
# Effect of Age on Target

---

1. We found that percentage of being a defaulter tend to decrease as getting older.

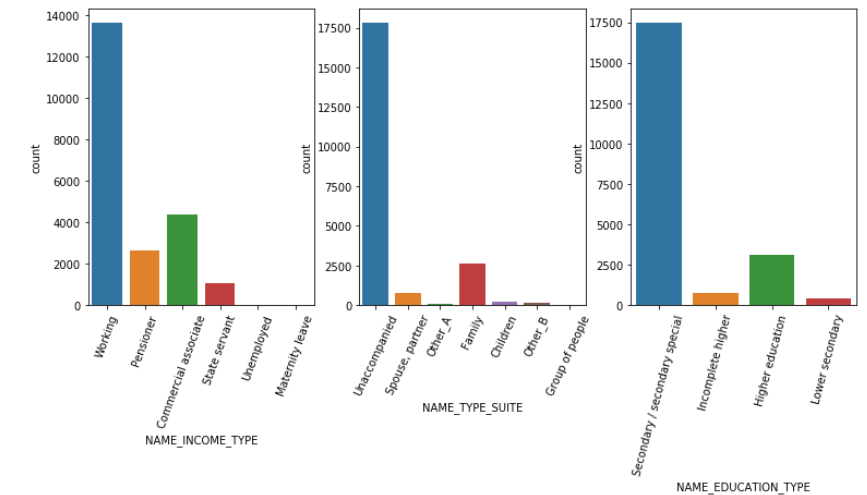
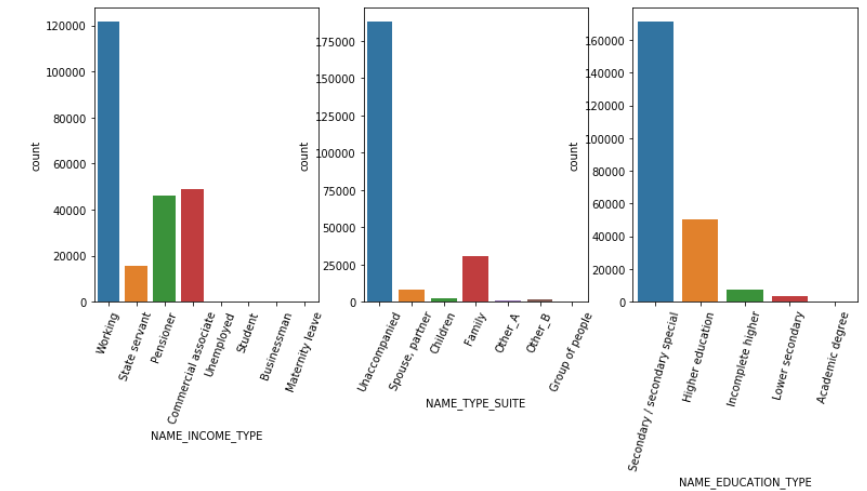
**2. Data provided to us covers the age group from 20 to 70 years.**

**3. Clients who lie in Age Group 20-25 were 12% more likely to be defaulter and least were in age group 65-70 with approx. 3%.**



# Clients likely to be defaulters :

1. Upper plot is for Target 0( who repayed the loan) and below plot is for Target 1(Defaulters)
2. We can conclude that income type group like student and businessman are less likely to be defaulters
3. Meanwhile clients with academic scholar as Education degree are less likely to be Defaulter.



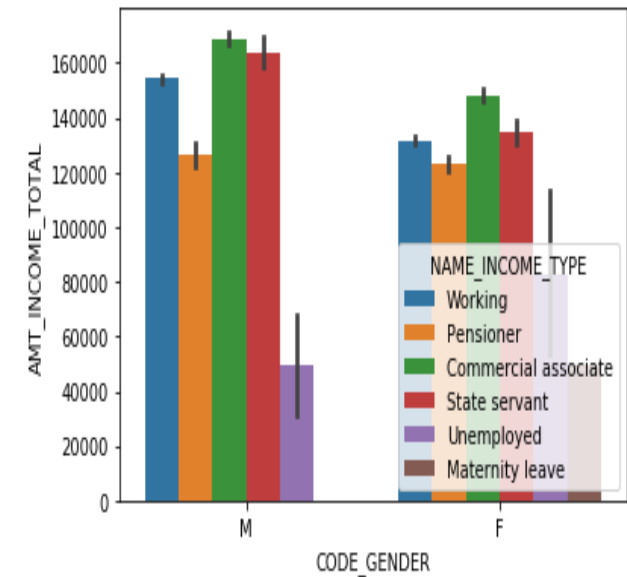
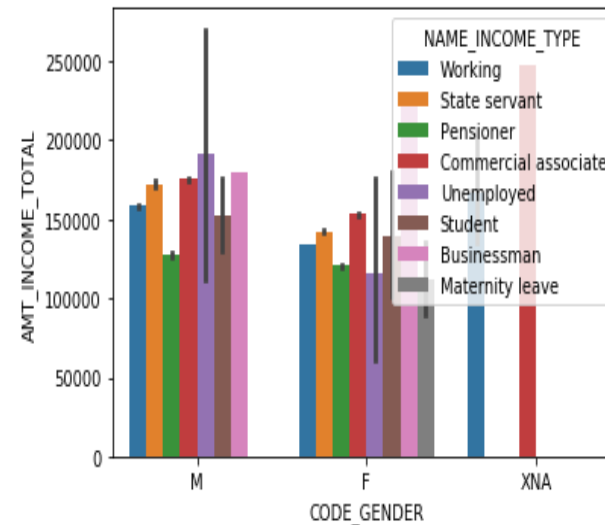
# Effect of Gender , income on Target

I. 1<sup>st</sup> figure shows TARGET 0 and 2<sup>nd</sup> figure is for TARGET 1

II. we can conclude that :

Top 3 likely to be defaulters based on income type w.r.t gender and total income is

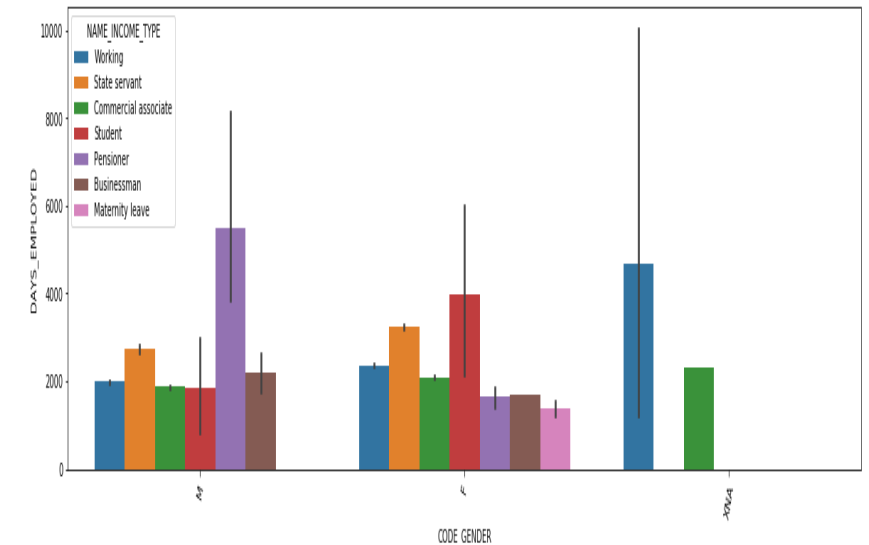
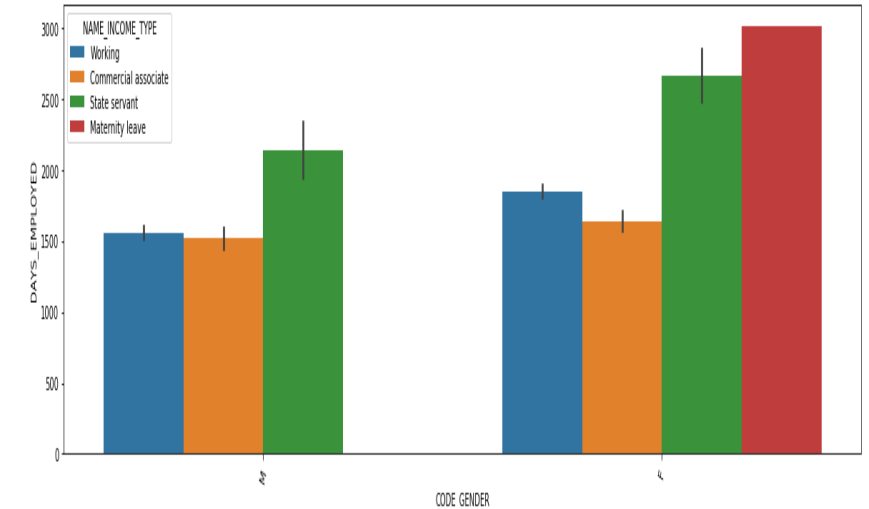
1. Commercial Associate
2. State Servant
3. Working client

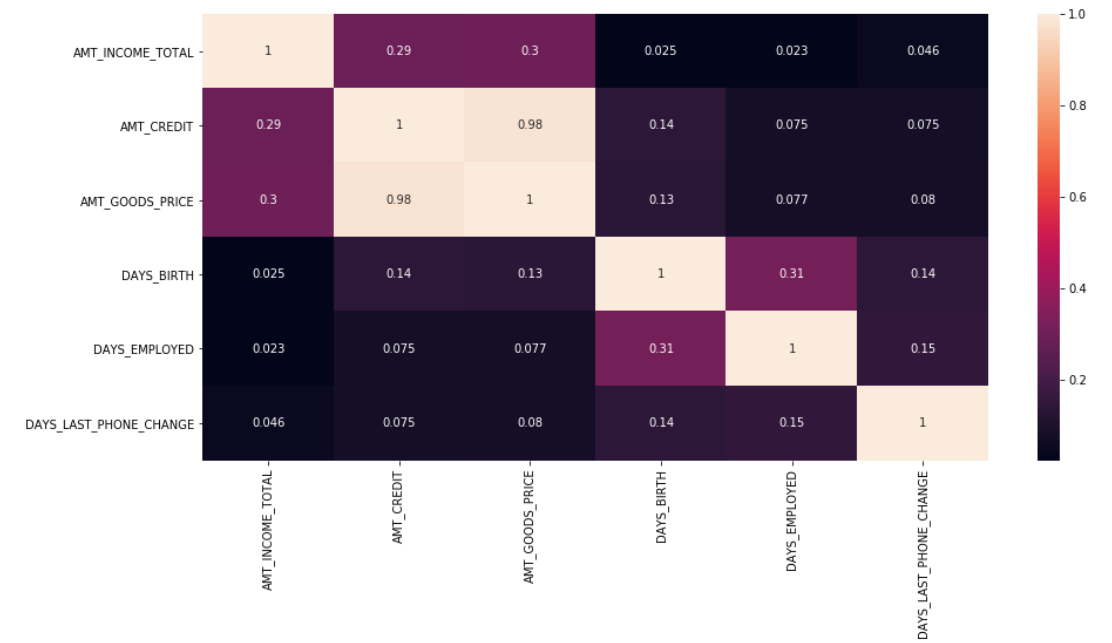
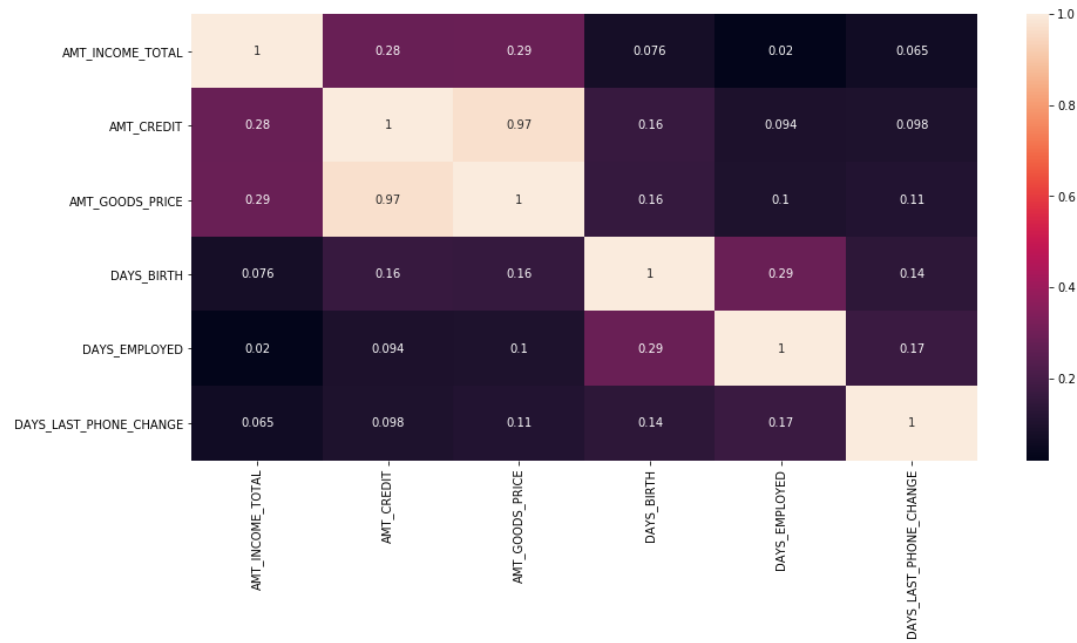


# Effect of Gender , Days employed and income type together on Target

**1.from plots we can conclude that state servant(male and female) and Female on maternity leave are most experienced in case of defaulters.**

**2.Meanwhile for non defaulters , most experienced are pensioner in males and student in female.**

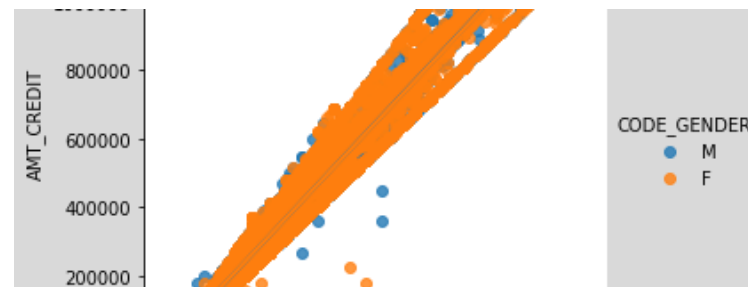
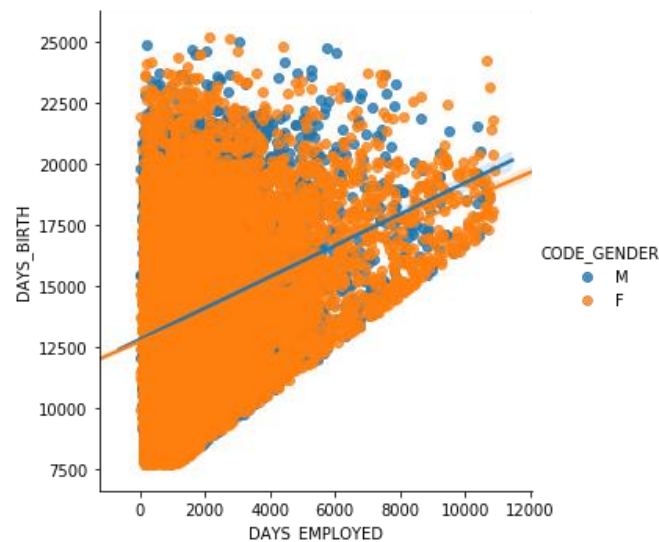
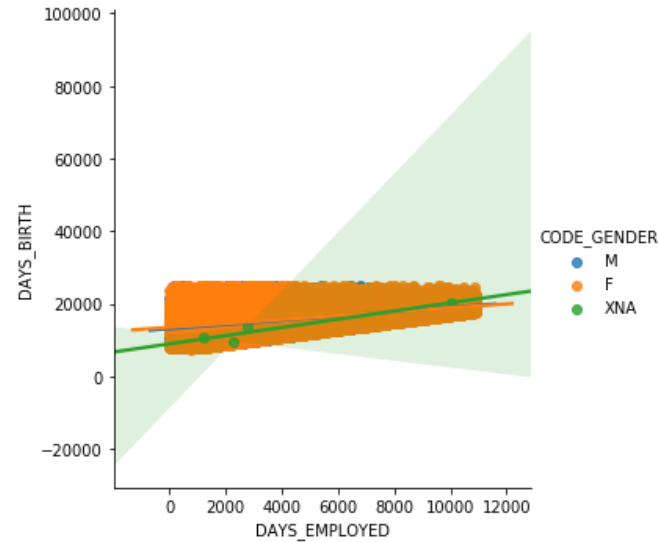
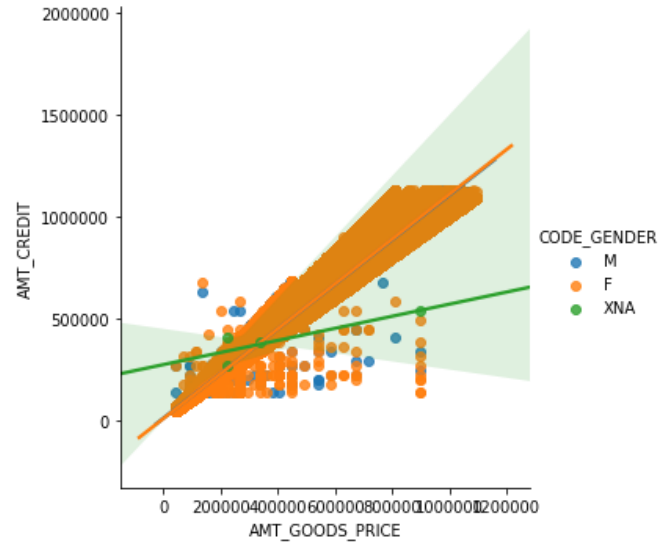




# Correlation Matrix for Target 0 and Target 1

# Top 2 Correlation variable

1. AMT CREDIT and AMT GOODS Price
2. DAYS Birth and DAYS Employed
3. Upper two plots are of TARGET1
4. Below two plots are of TARGET 0



# Previous Application

---

ANALYSIS



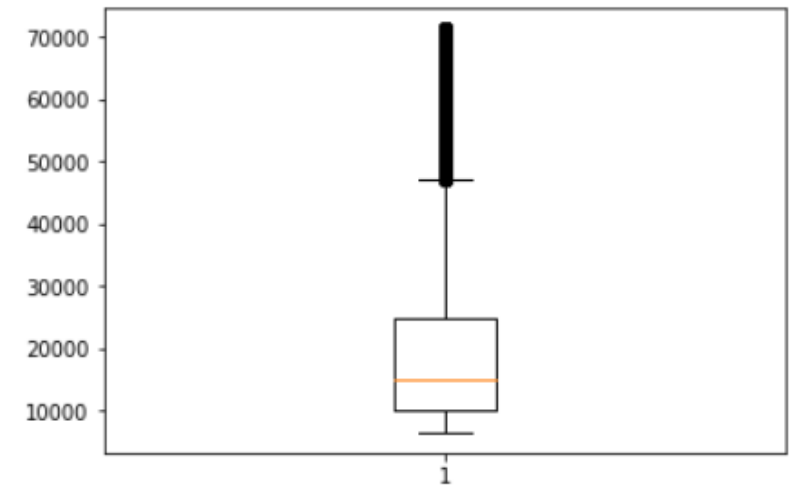
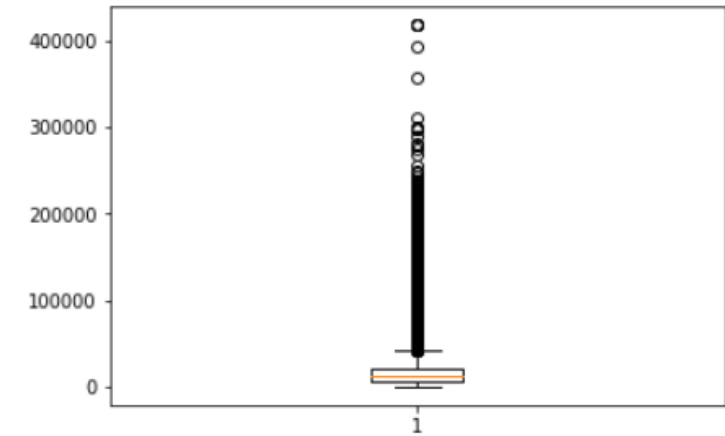
# Outliers

---

We had used the AMT\_ANNUIITY column as the base column for the analysis and while looking deeper into the data for annuity amount, we have seen there were few outliers into the data:

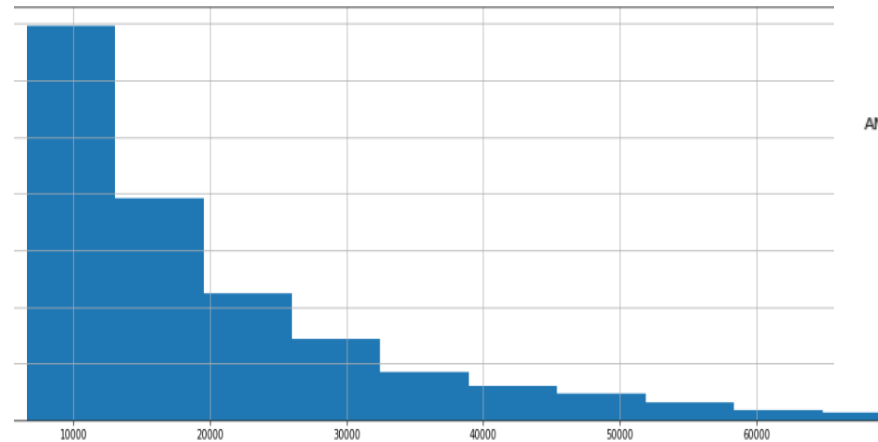
To remove the outliers, we have divided the data into different percentiles:

The first quartile of data seems holding too many low values, hence to remove the outliers we have selected a range between 25<sup>th</sup> and 99<sup>th</sup> percentile:



# Correlation

Correlation between annuity amount and application amount is very high which translate that the final disbursed amount is very near to the applied amount:

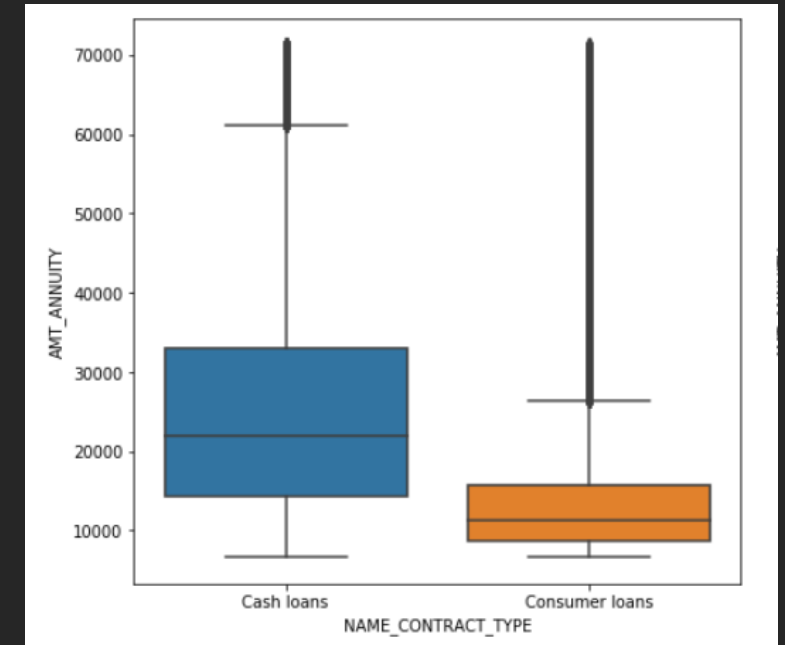
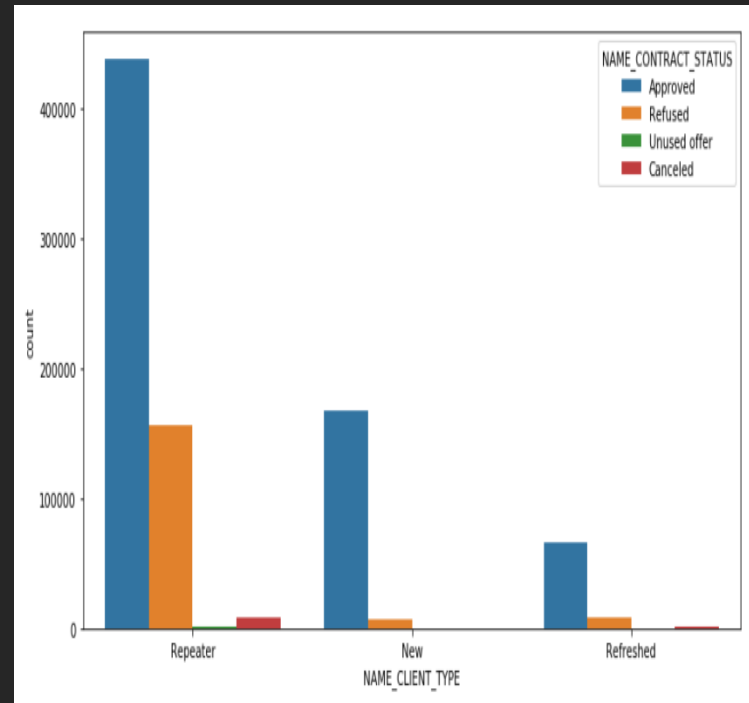


	AMT_ANNUIITY	AMT_APPLICATION	AMT_CREDIT	AMT_GOODS_PRICE	CNT_PAYMENT
AMT_ANNUIITY	1	0.81	0.8	0.81	0.41
AMT_APPLICATION	0.81	1	0.99	1	0.74
AMT_CREDIT	0.8	0.99	1	0.99	0.77
AMT_GOODS_PRICE	0.81	1	0.99	1	0.74
CNT_PAYMENT	0.41	0.74	0.77	0.74	1

# Analysis

From the bivariate analysis we have found that the cash loans are the main driver, however the consumer loans have higher annuity amount:

Rejection rate is high for the repeated customers:



Thank You