



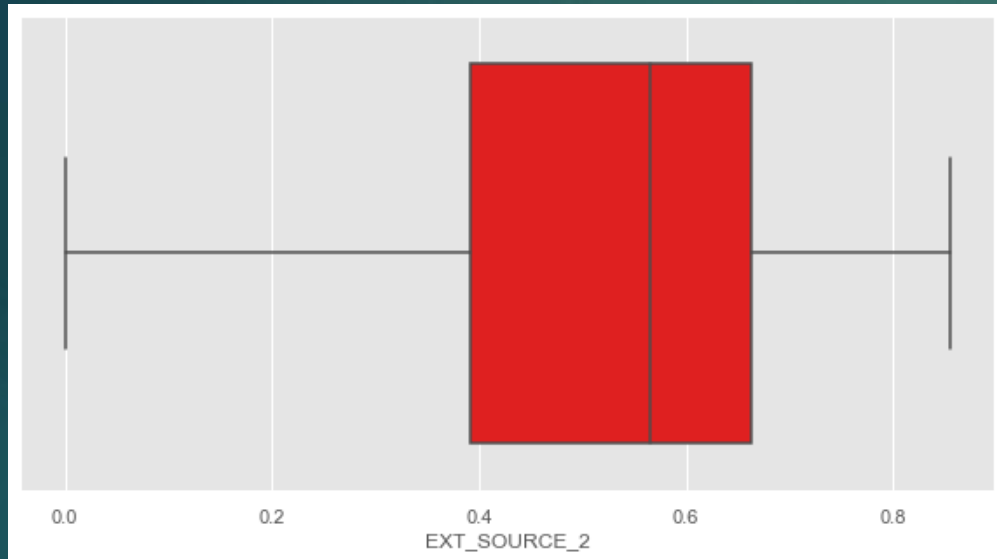
# EDA On Fintech Dataset

- OMPRAKASH DHAKAR

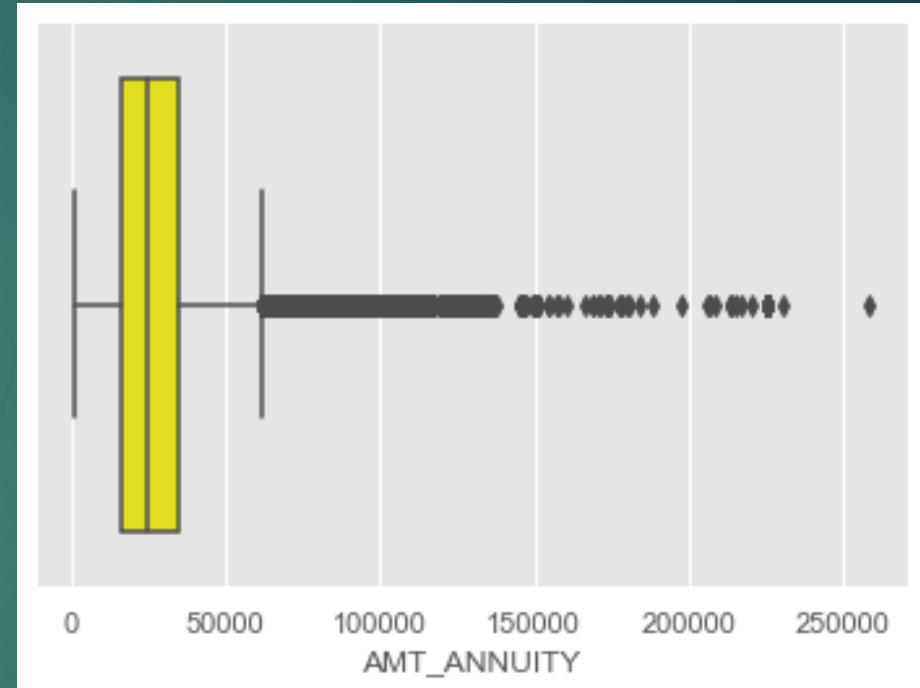
# Introduction

- ▶ This case study aims to give you an idea of applying EDA in a real business scenario. In this case study, apart from applying the techniques that you have learnt in the EDA module, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.

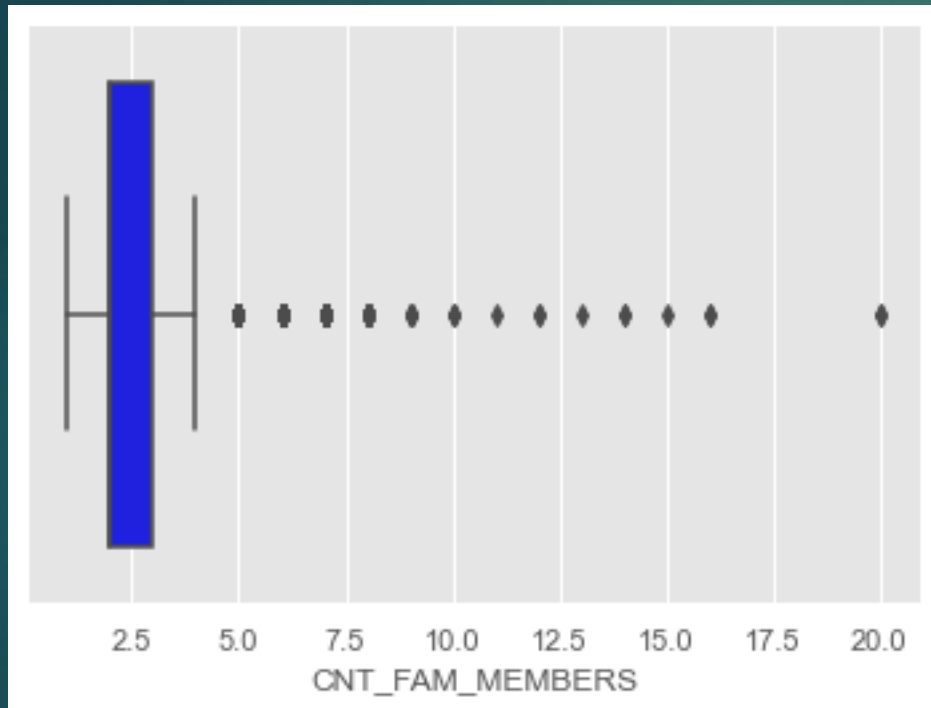
# Outlier Analysis



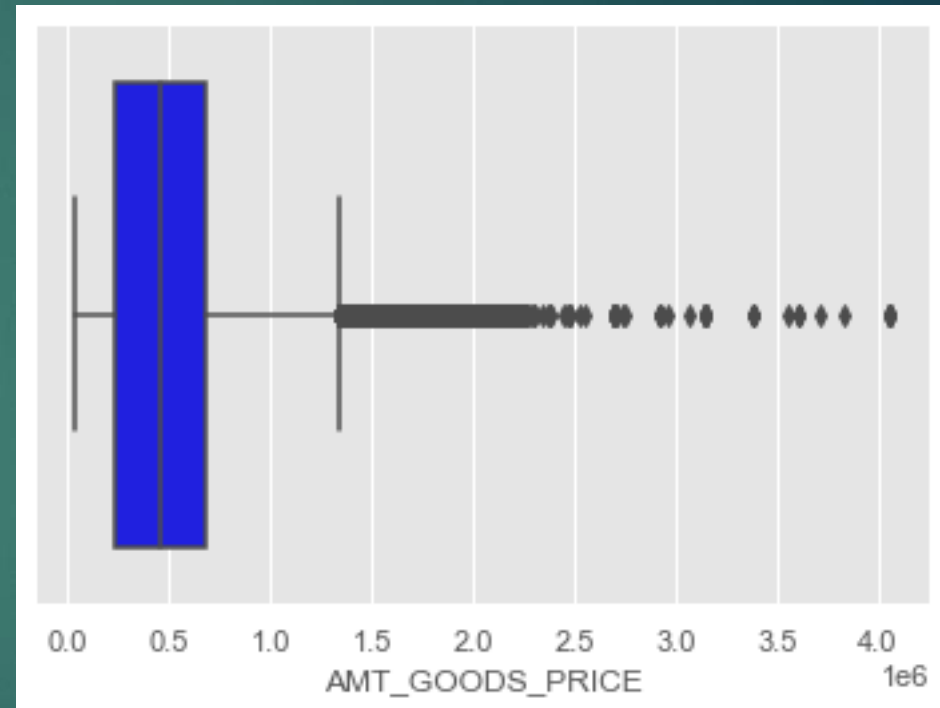
This Box Plot has no outlier so we have choose mean to impute  
Mean = 0.5



This box plot is continuous variable and outliers are present in this plot. So I have choose median to impute  
Median = 24903.00

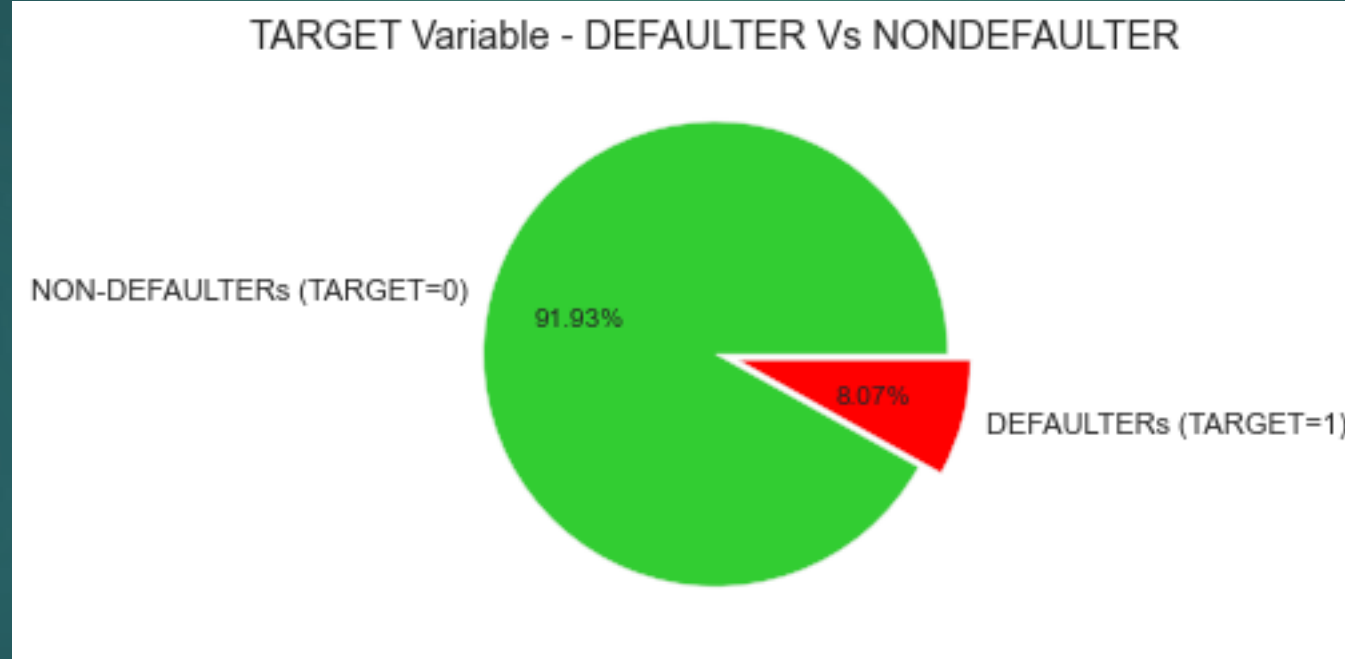


This Box Plot has an outlier for Children's  
in Family Member  
Median = 2

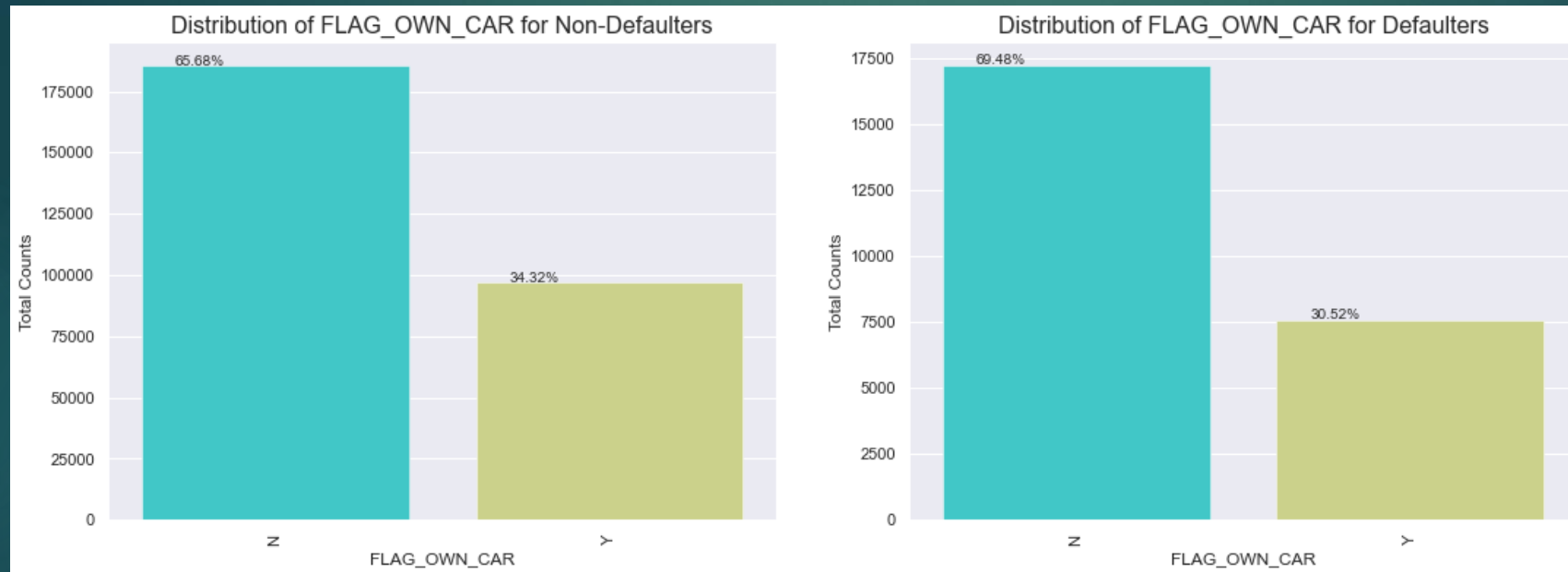


This Box Plot has outlier's for Good's  
Amount  
Median = 450000.00

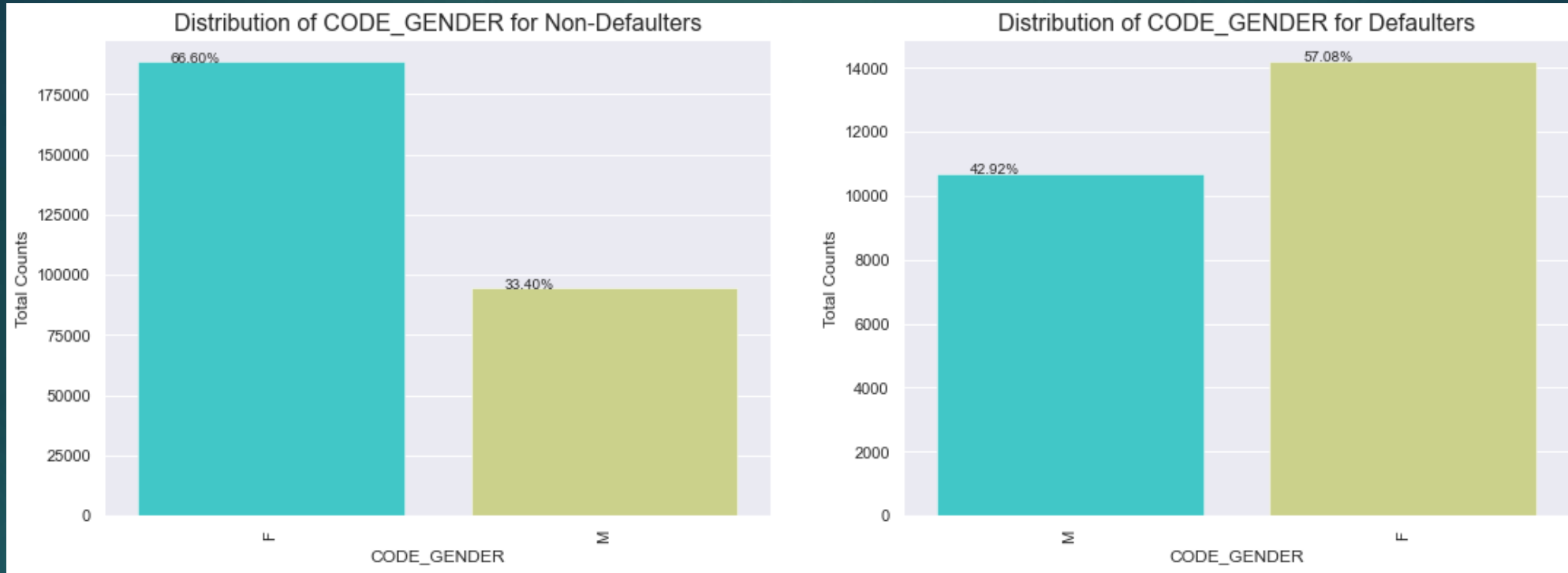
# Target Variable Analysis for Defaulters and Non Defaulters



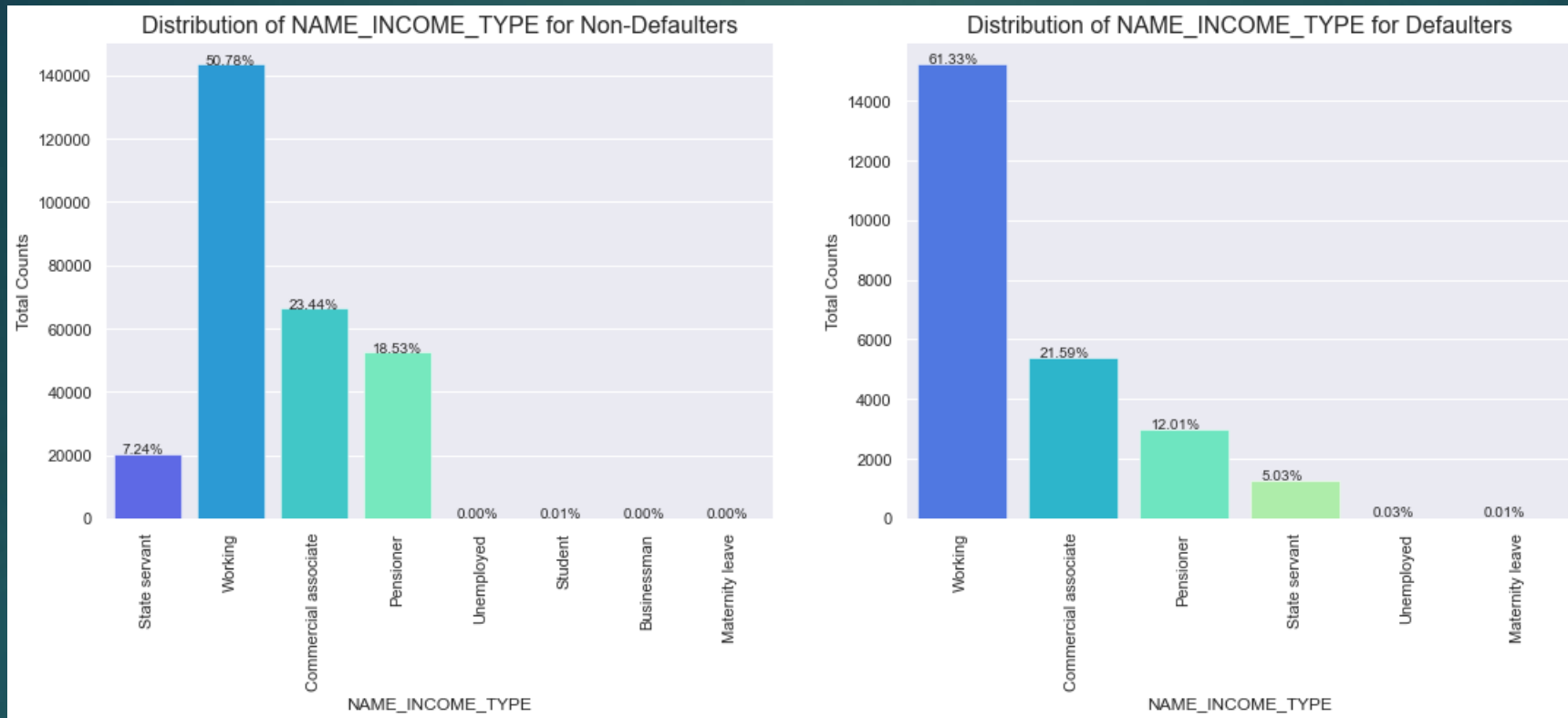
# Uni- Variate Analysis



The plot shows that borrowers who owns a car make a large proportion of both Defaulting (70%) and Non-Defaulting borrowers (65.68%)



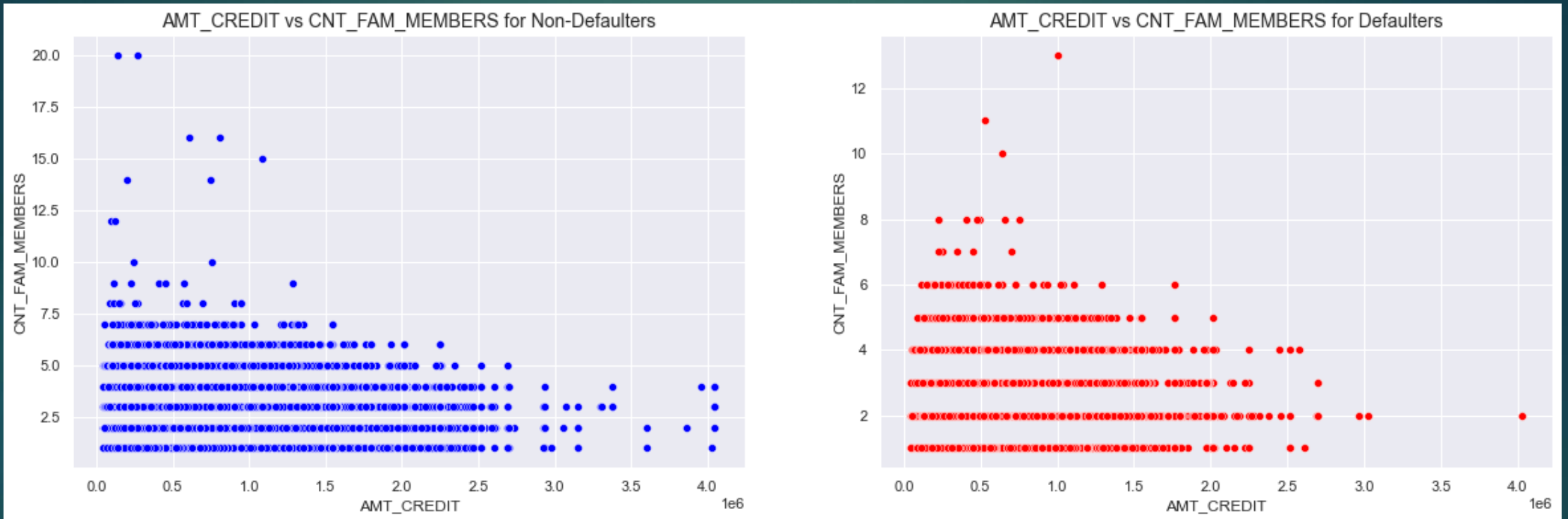
The plot indicates that a higher proportion of female borrowers are Non-Defaulting as compared to male borrowers.



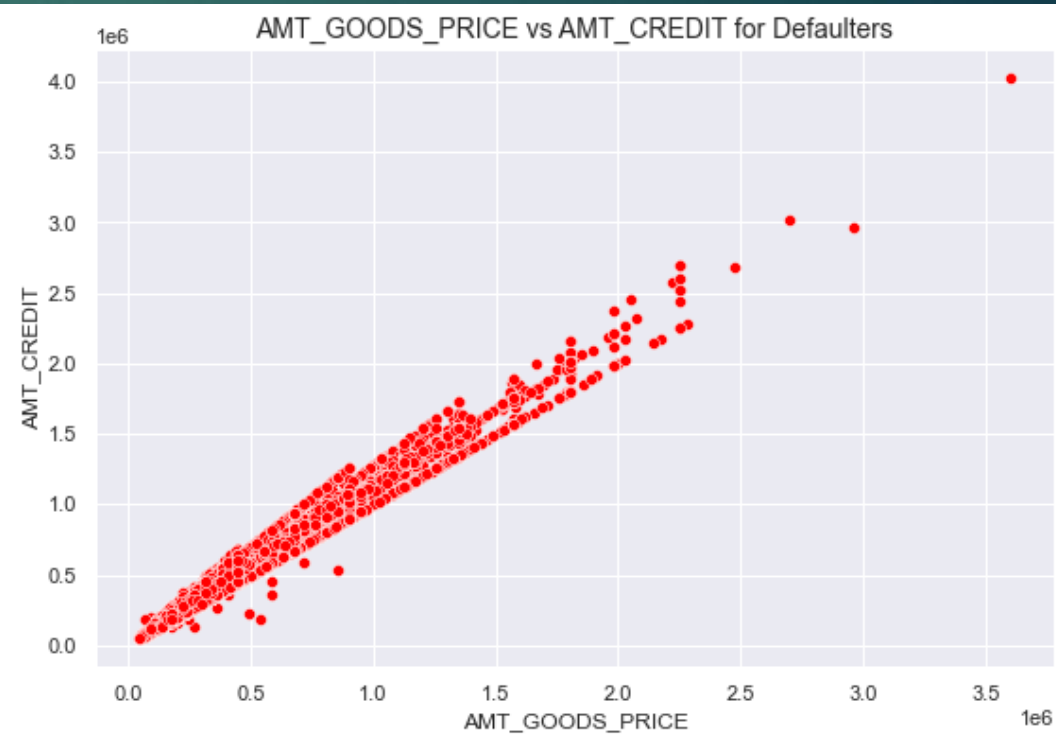
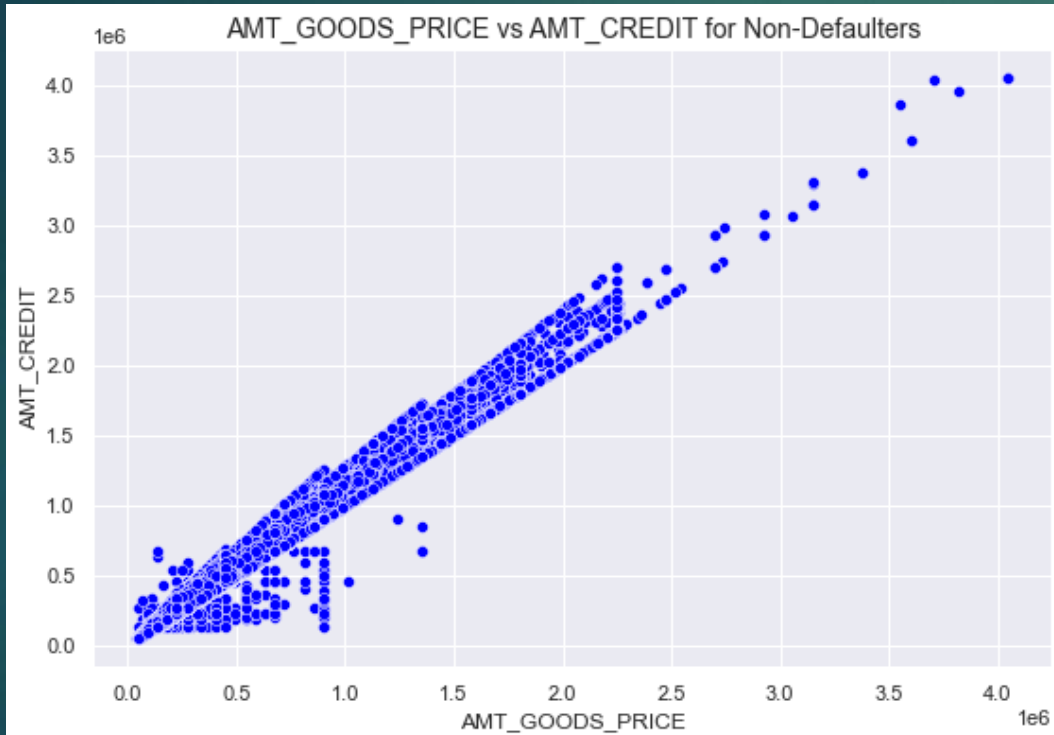
The Plot indicates that student do not default on their loans possibly because they are not required to make payments while they are enrolled in schools. Businessman's also have low default rate. More number of loan are distributed to working class borrowers who make a large portion of both defaulting (61%) and non defaulting (51%) borrowers.



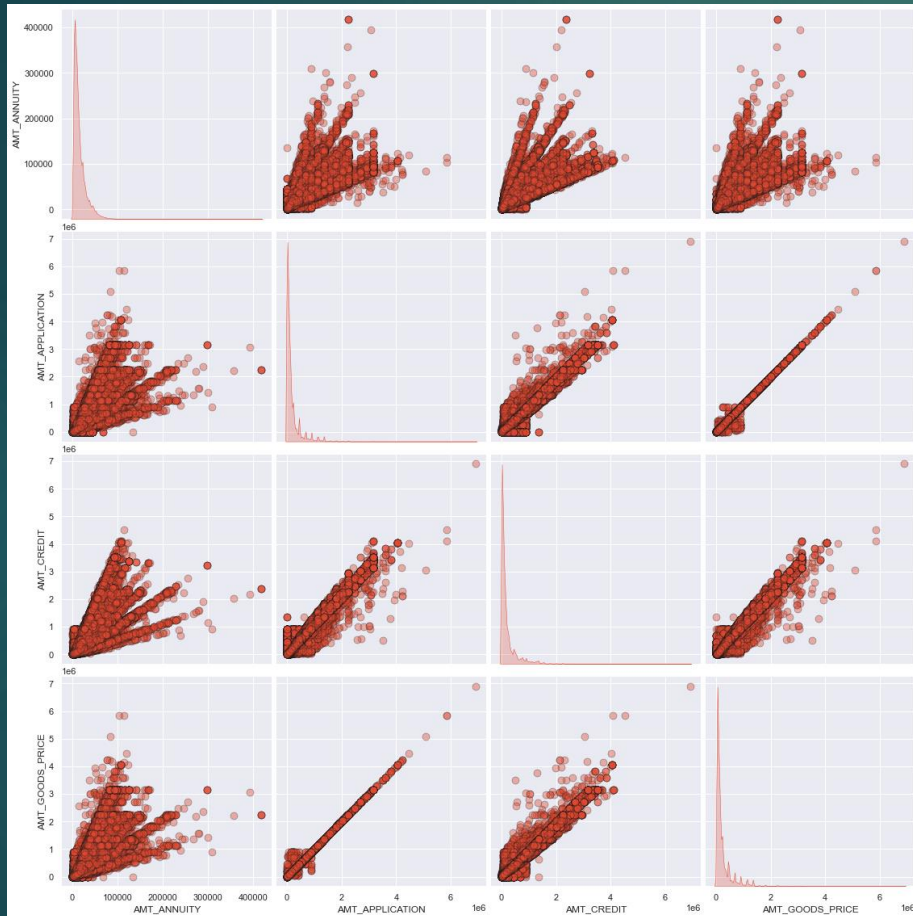
# Bi- Variate Analysis



The Density plot shows that the like hood of default is similar for borrowers with small families and low amount credit values regardless of whether they have loan or not.

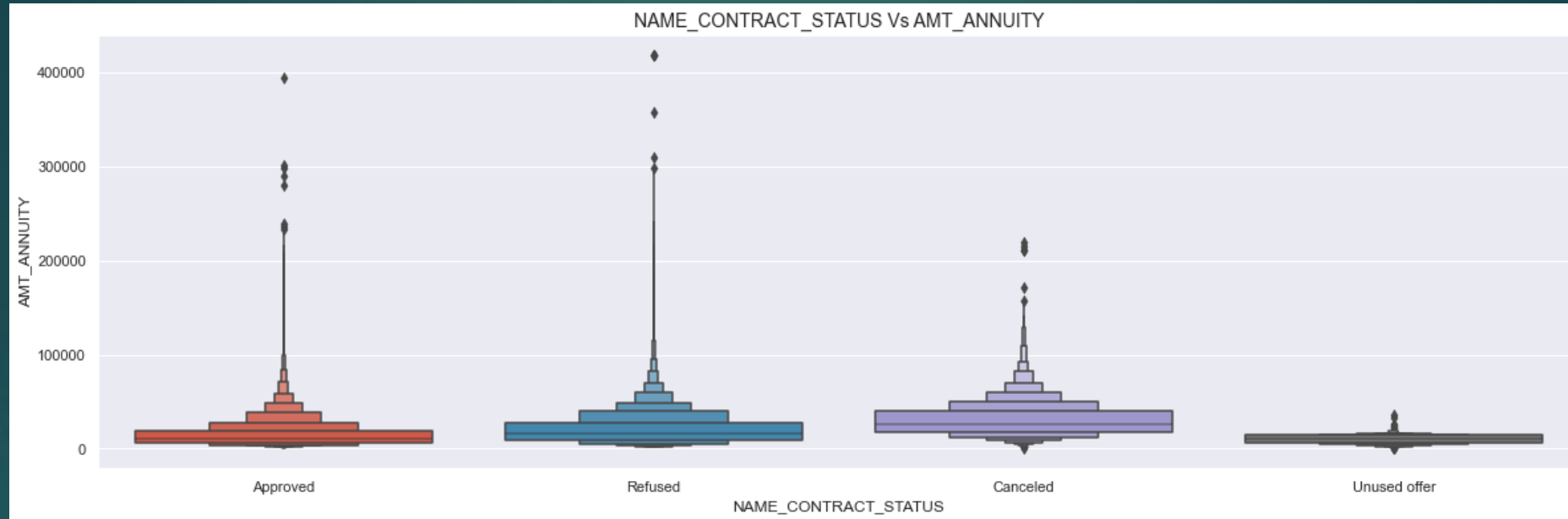


# Bi- Variate (Pair Plot)



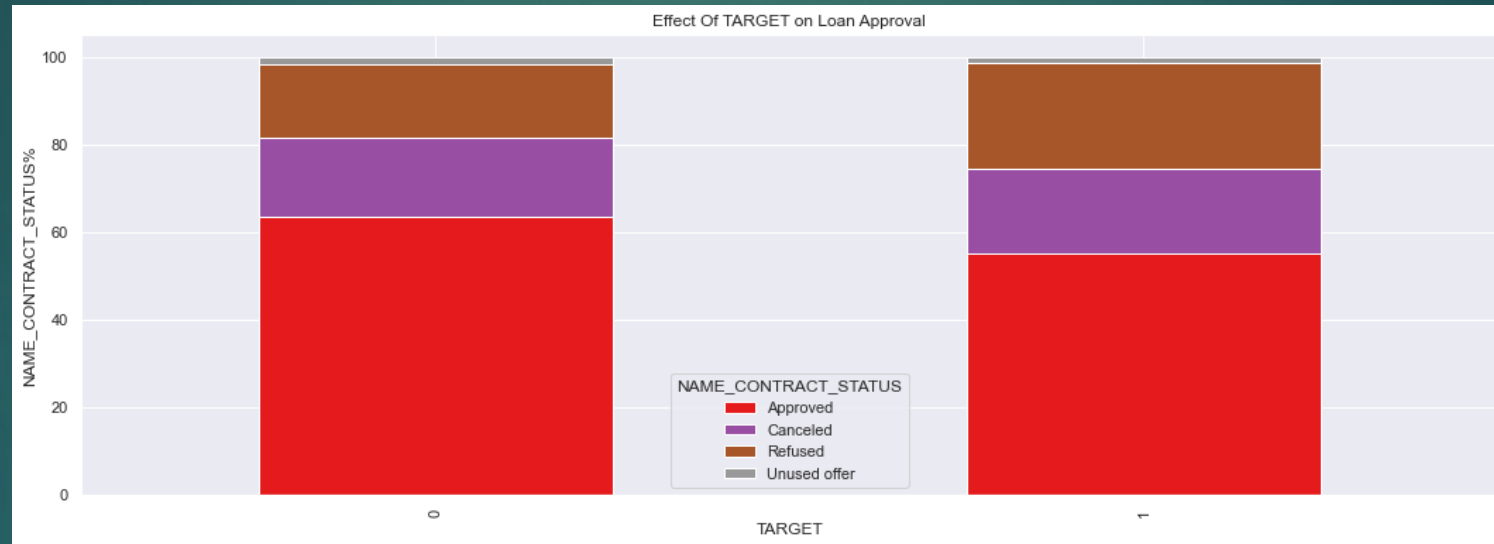
The analysis shows that the annuity of a previous loan application has a strong positive influence on several factors:

1. The amount of credit requested on the previous application
2. The final approved credit amount on the previous application
3. The price of the goods requested on the previous application.



The plot shows that loan applications with lower values of AMT\_ANNUIITY are more likely to be canceled or unused, while applications with very high AMT\_ANNUIITY values are also more likely to be refused. This suggests that there may be a sweet spot for AMT\_ANNUIITY values that results in the highest approval rate.

# Uni-Variate Analysis on Merged Data



There appears to be a relationship between loan approval status and default rates, with borrowers who were approved for a loan earlier having lower default rates compared to those who were refused a loan earlier. However, further analysis is needed to determine the significance of this relationship and to consider other factors that may influence default rates.

# Thank you...

Submitted by **OMPRAKASH DHAKAR**