



# Credit EDA Analysis

---

JAYATI CHOWDHURY

# Problem Statement

---

It's a basic understanding of Credit risk analytic in Banking or Financial domain

We have to Analyze the data to reduce the risk of payment difficulties

Through Exploratory Data Analysis we can solve the real-life Business Problem

So there are two scenarios clearly understandable:-

A) If the applicant is like to repay the loan:-then not approval of loan will be Business loss

B) If the applicant is not like to repay the loan:-then approval of loan will be a financial loss in business

# Business Objective

---

So here company wants to understand the driving factors behind the loan defaulters,

So that company can utilize those data in Risk Analytic

# Overall Approach

---

Data understanding, Data cleaning and manipulation, Data Analysis, Presentation and recommendations, Conciseness and readability of the analysis.

Details steps are below-

- 1.Data Loading & understanding the business behind it
- 2.Finding the Outliers-using Univariate Outliers
- 3.Univariate data analysis
- 4.Bivariate data analysis
- 5.Univariate Analysis on Segmented data

All these analysis are done depending on loan payment difficulties depending on the data on Current Application and Previous Application

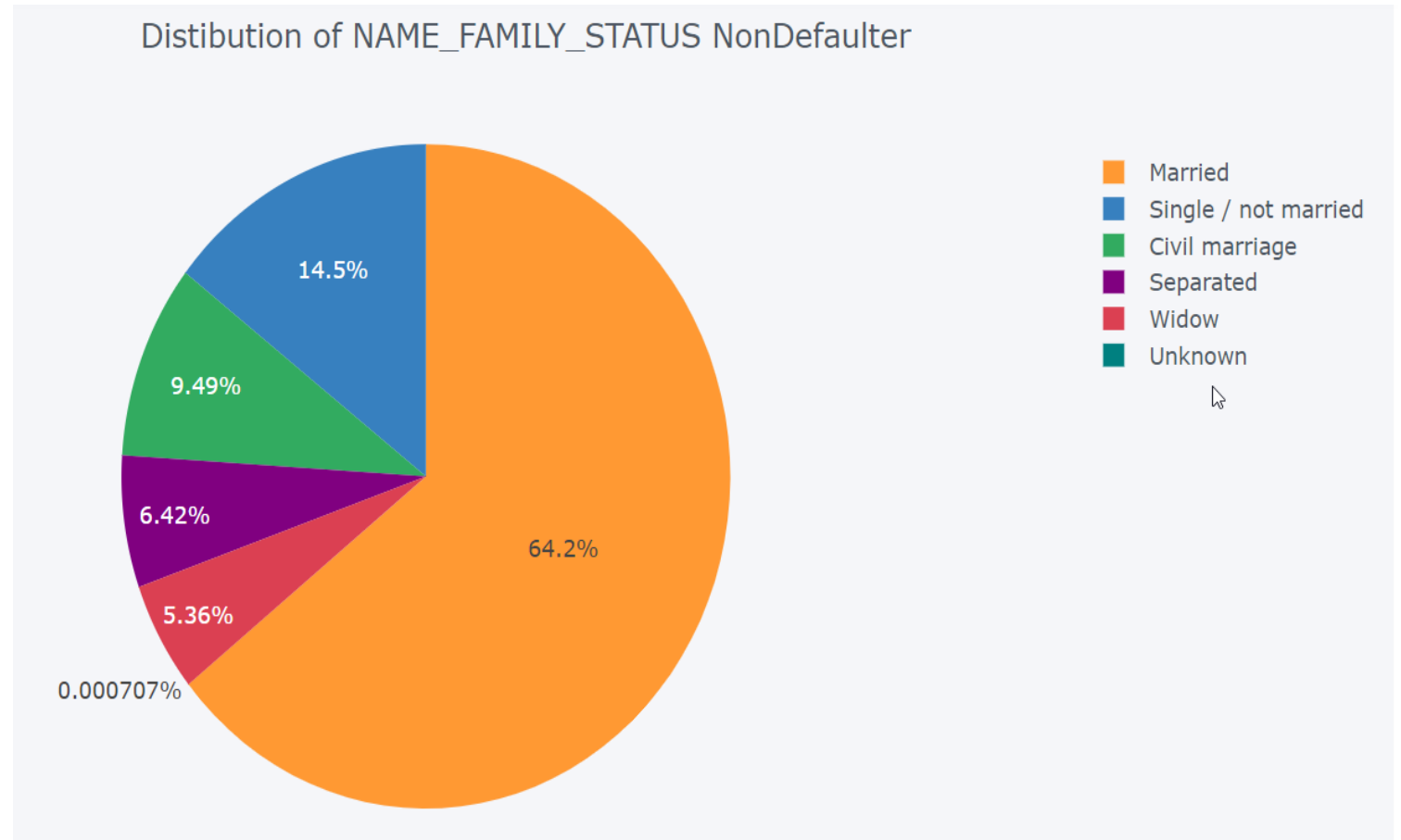
---

# Univariate Analysis on Current Application

# Family Status Distribution against Non- Defaulter

---

By seeing the chart it is clear that Married people are taking more loans. They are responsible & committed enough rather than Single , separated & widow ,so they are important variable for non-defaulters



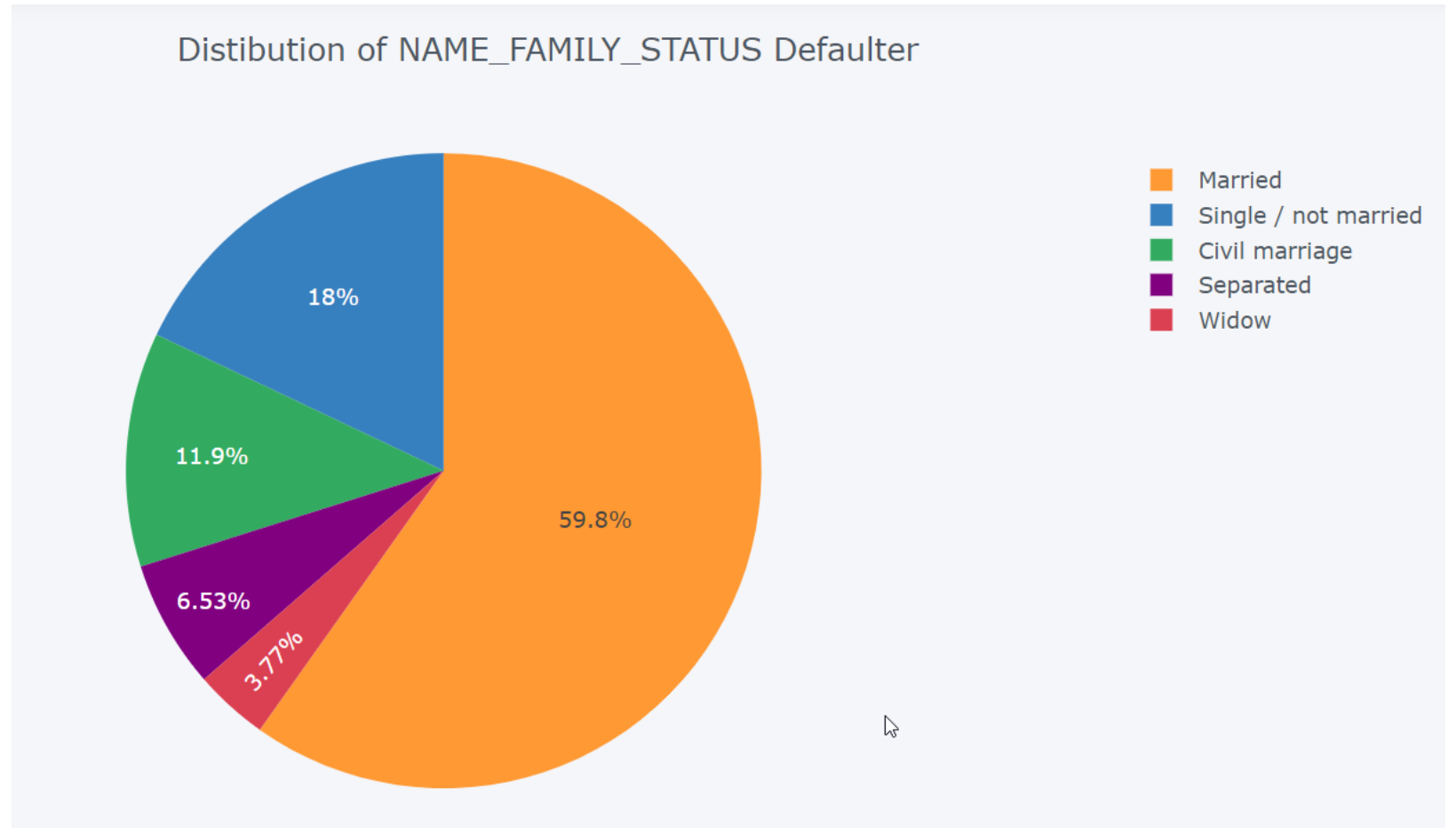
# Family Status Distribution against Defaulter

---

By seeing the chart it is clear that Married people are taking more loans. they are important variable for defaulters also.

## Comparison:-

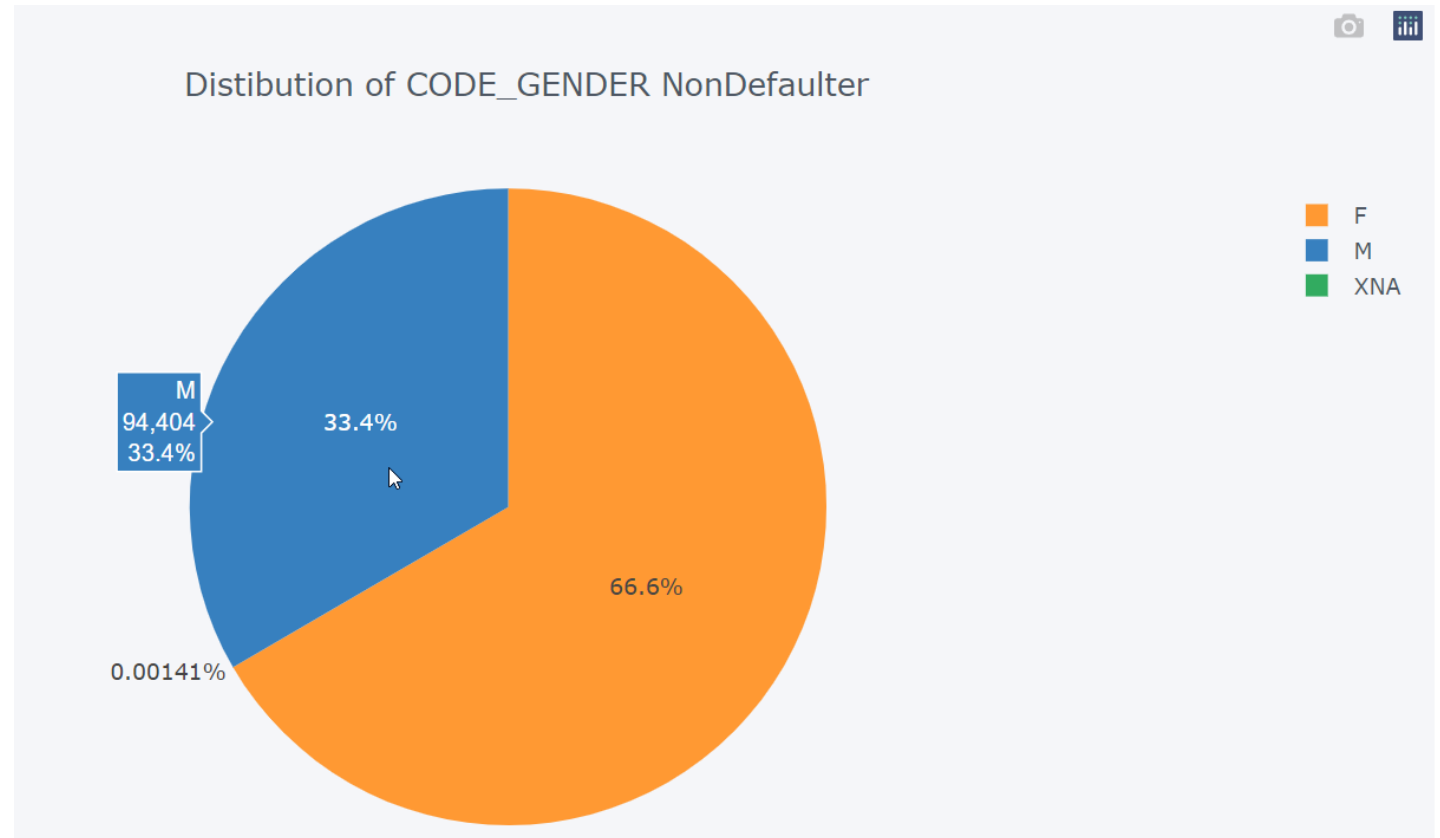
So looks like they are a decision maker for taking loans



# Gender Distribution against Non- Defaulter

---

By seeing the chart it is clear that Females are the driver , as its having more percentage than male as Non Defaulters





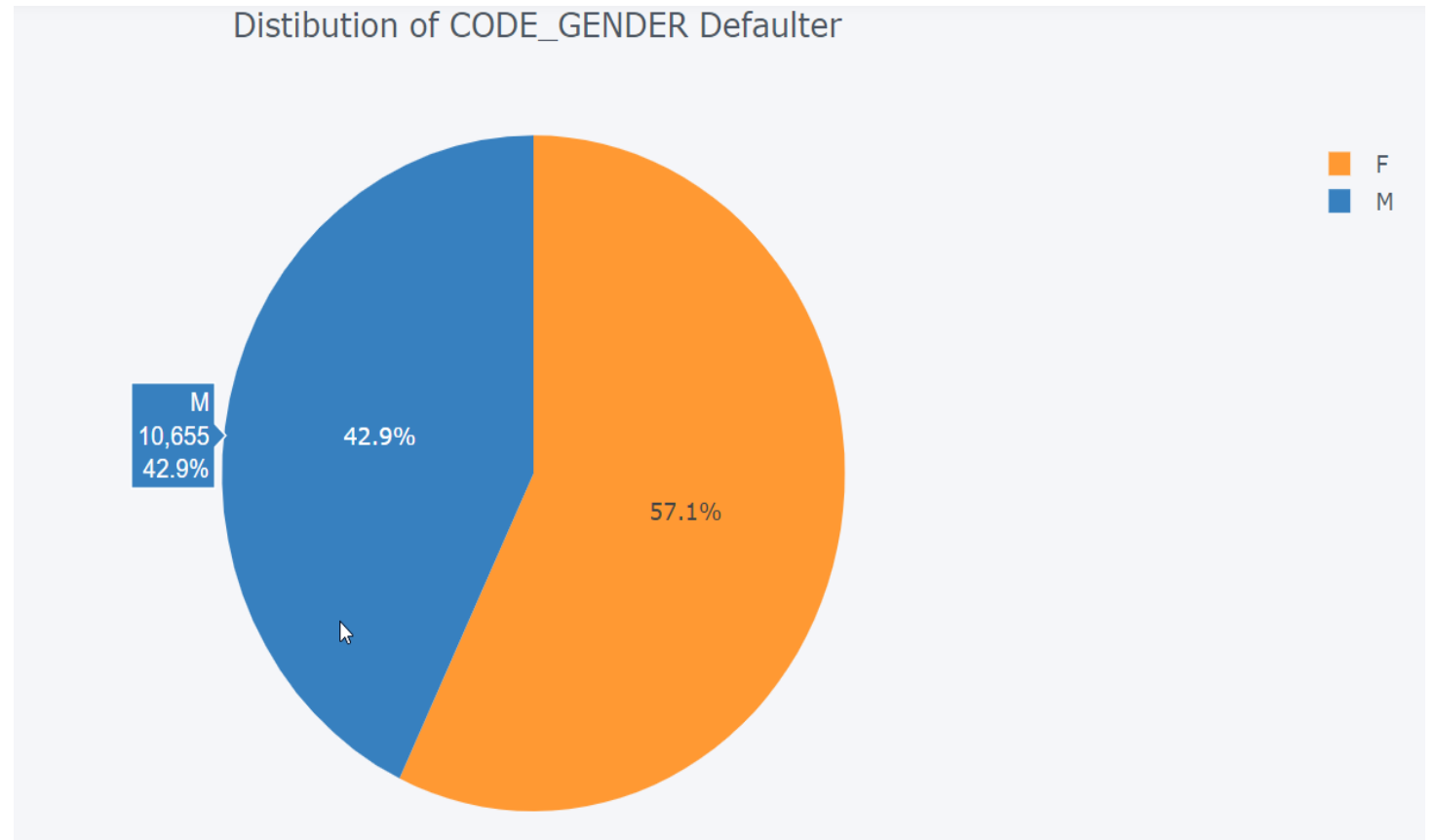
# Gender Distribution against Defaulter

---

By seeing the chart it is clear that Females are the Driver, as its having more percentage than male as Defaulters.

## Comparison:-

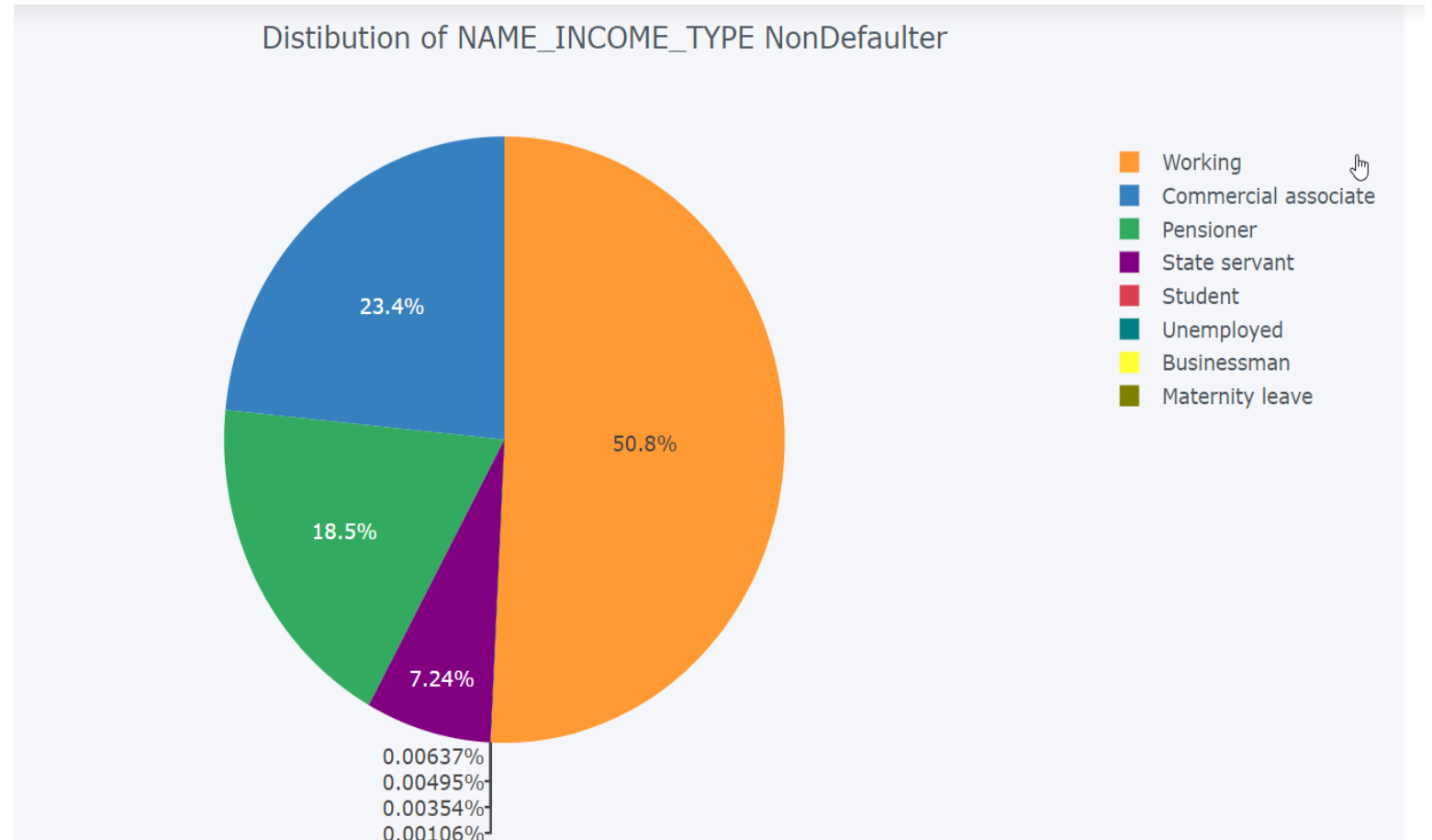
Only comparison is percentage of defaulter male is slightly more than the non defaulter



# Income Type Distribution against Non-Defaulter

---

By seeing the chart it is clear that more working class people are taking more loan and also no difficulties with Payment



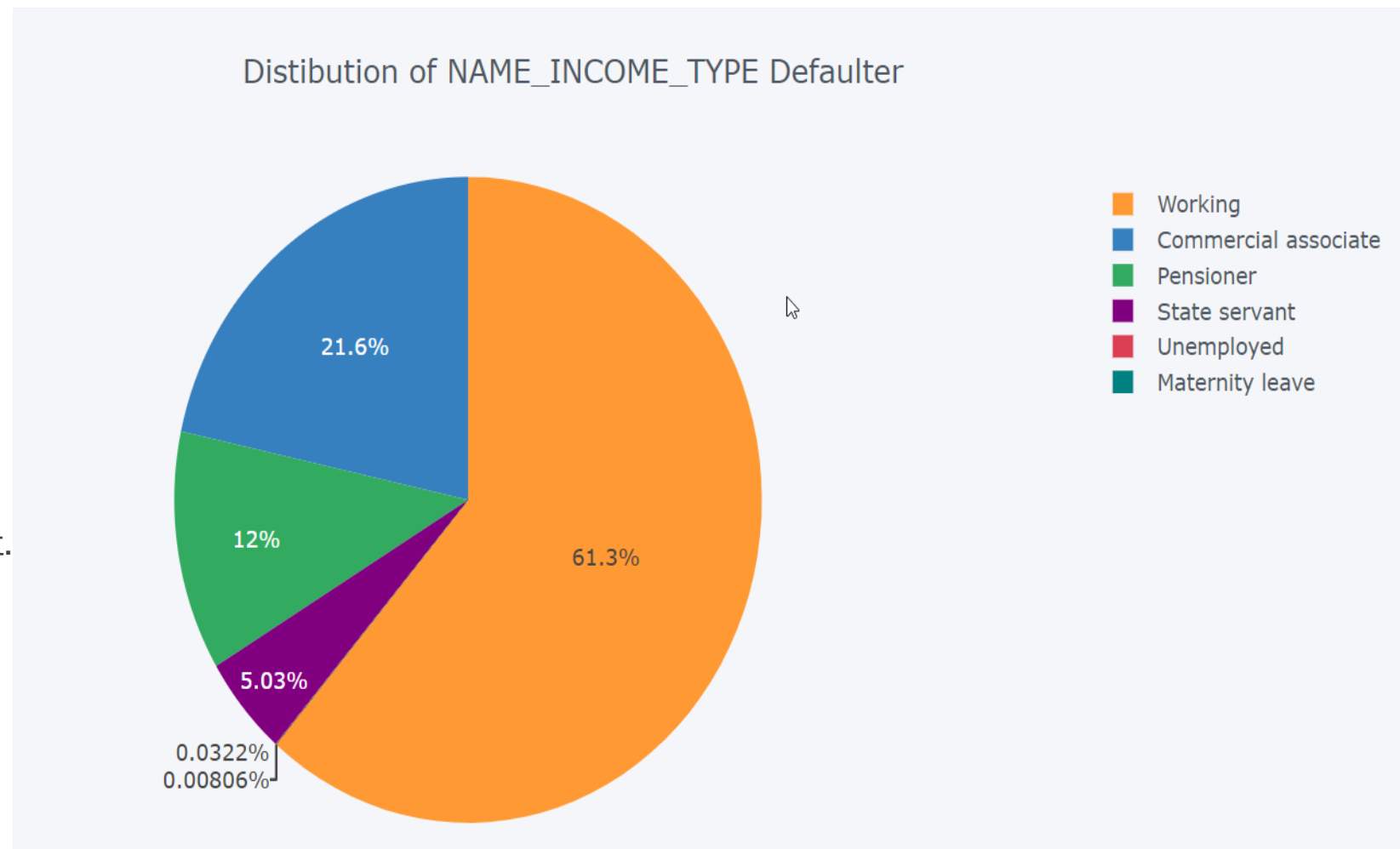
# Income Type Distribution against Defaulter

---

By seeing the chart it is clear that more working class people taking more risk are taking more loan and also facing difficulties with Payment.

## Comparison:-

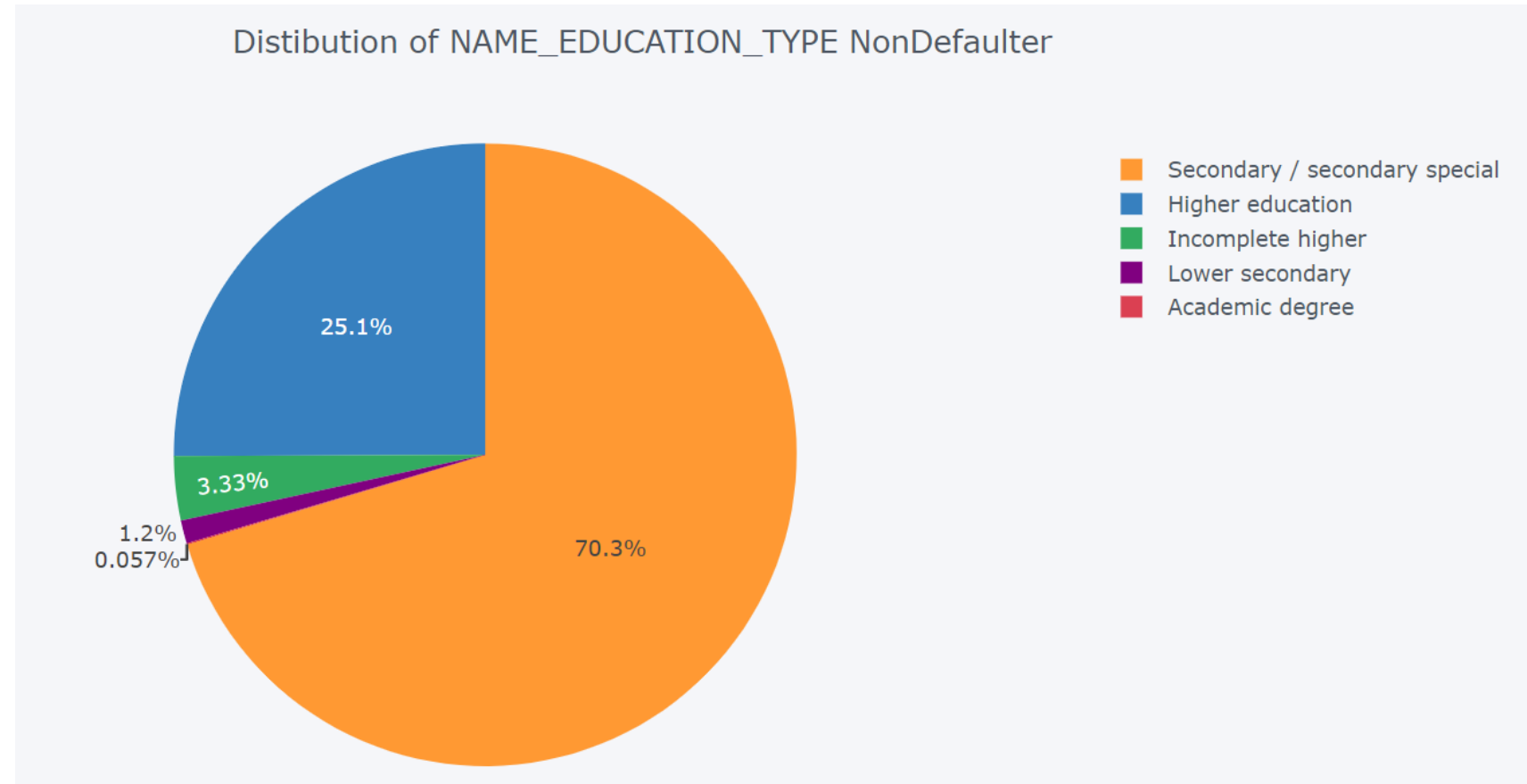
If we compare defaulter chart with non-defaulter in respect with Income Type-its clearly visible pensioners percentage has decreased and working class percentage has increased slight



# Name Education Type Distribution against Non-Defaulter

---

By seeing the chart it is clear that more Secondary Education Holder are the non Defaulter



# Education Distribution against Defaulter

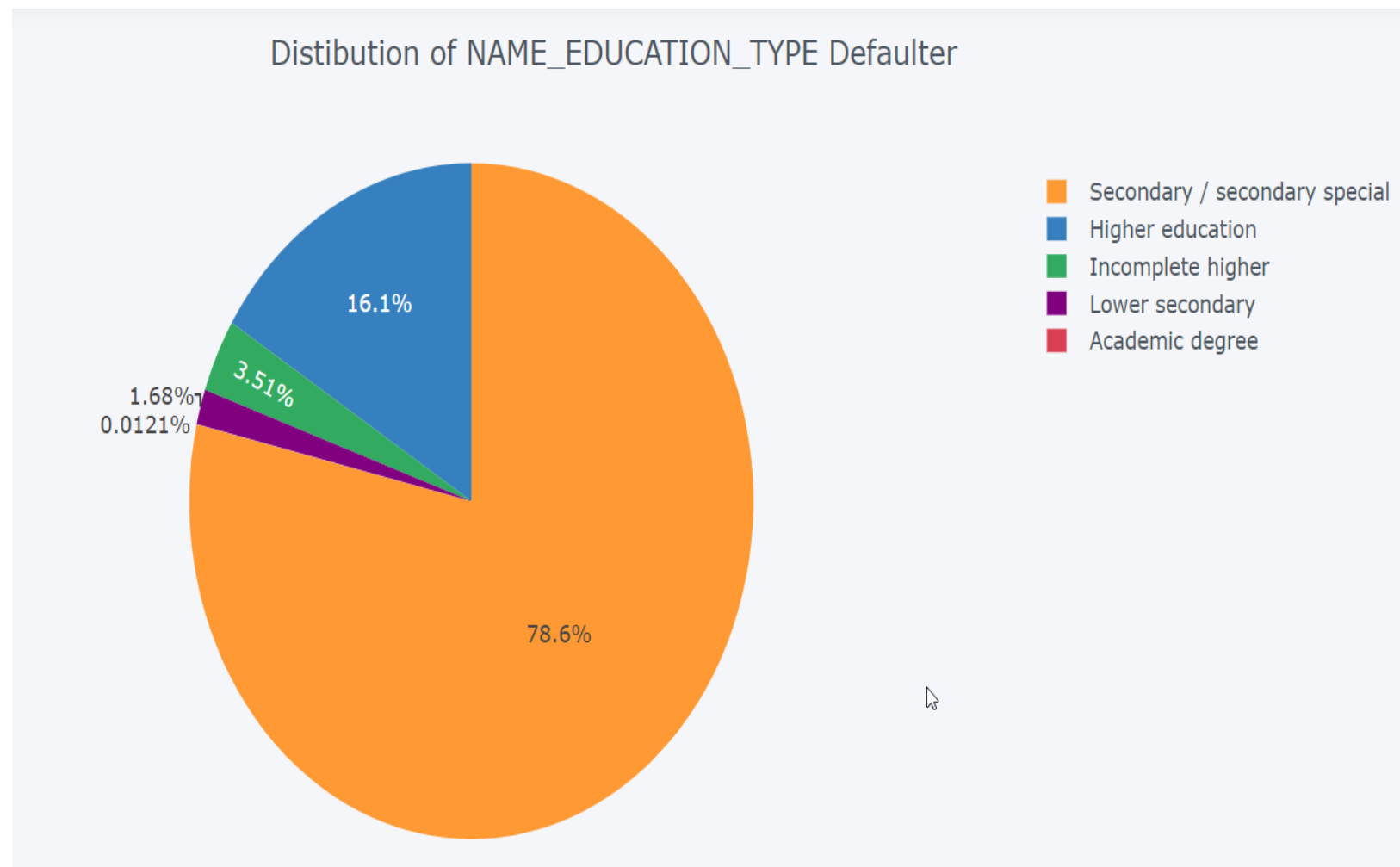
---

By seeing the chart it is clear that more Secondary Education Holder taking more loan and also facing payment difficulties and thus became the defaulters.

.

## Comparison:-

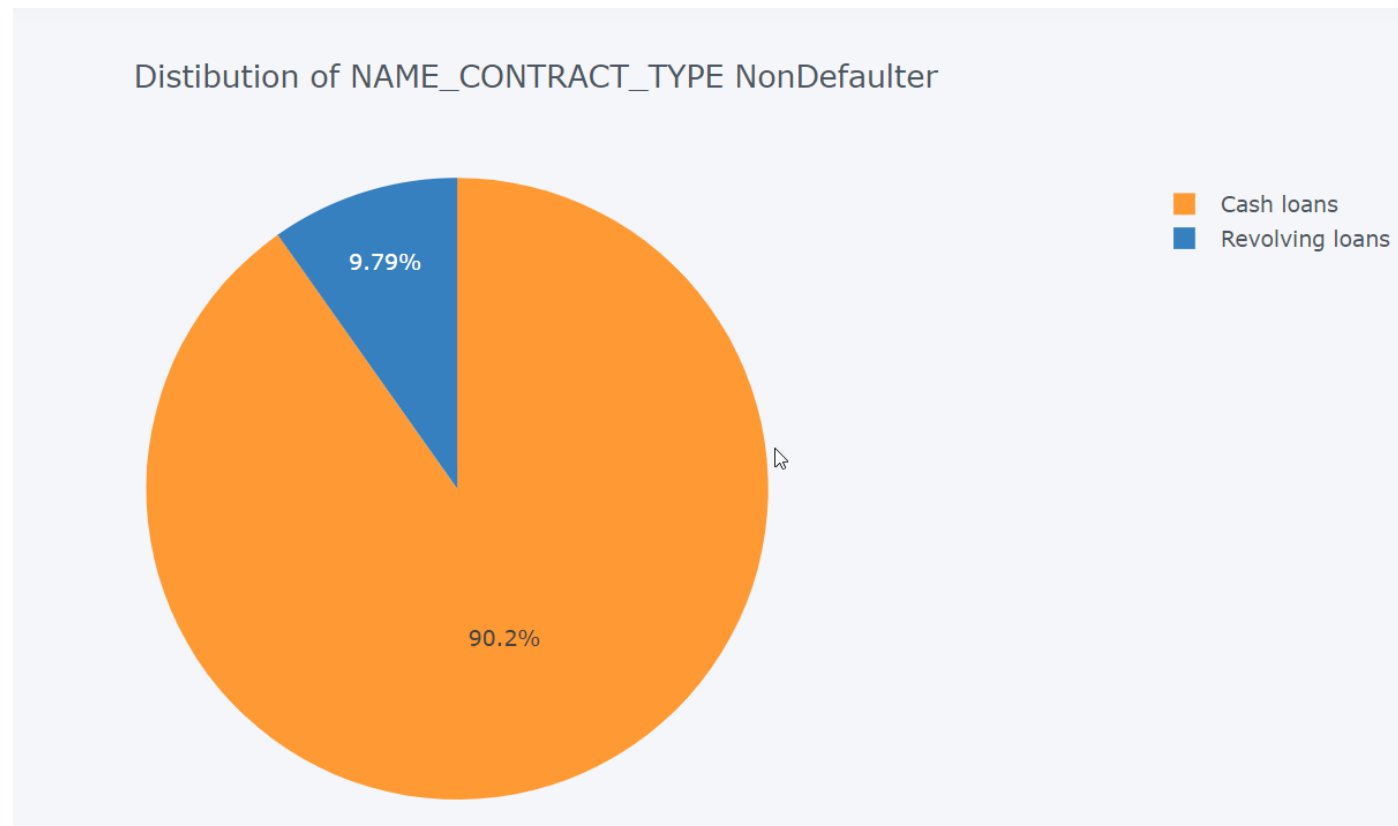
If we compare defaulter chart with non-defaulter in respect with Education. There is a significant percentage decreased in case of Higher Education holder in case taking and paying loans



# Loans contract Distribution against Non- Defaulter

---

By seeing the chart it is clear that people are taking cash loans instead of revolving loans from Bank.



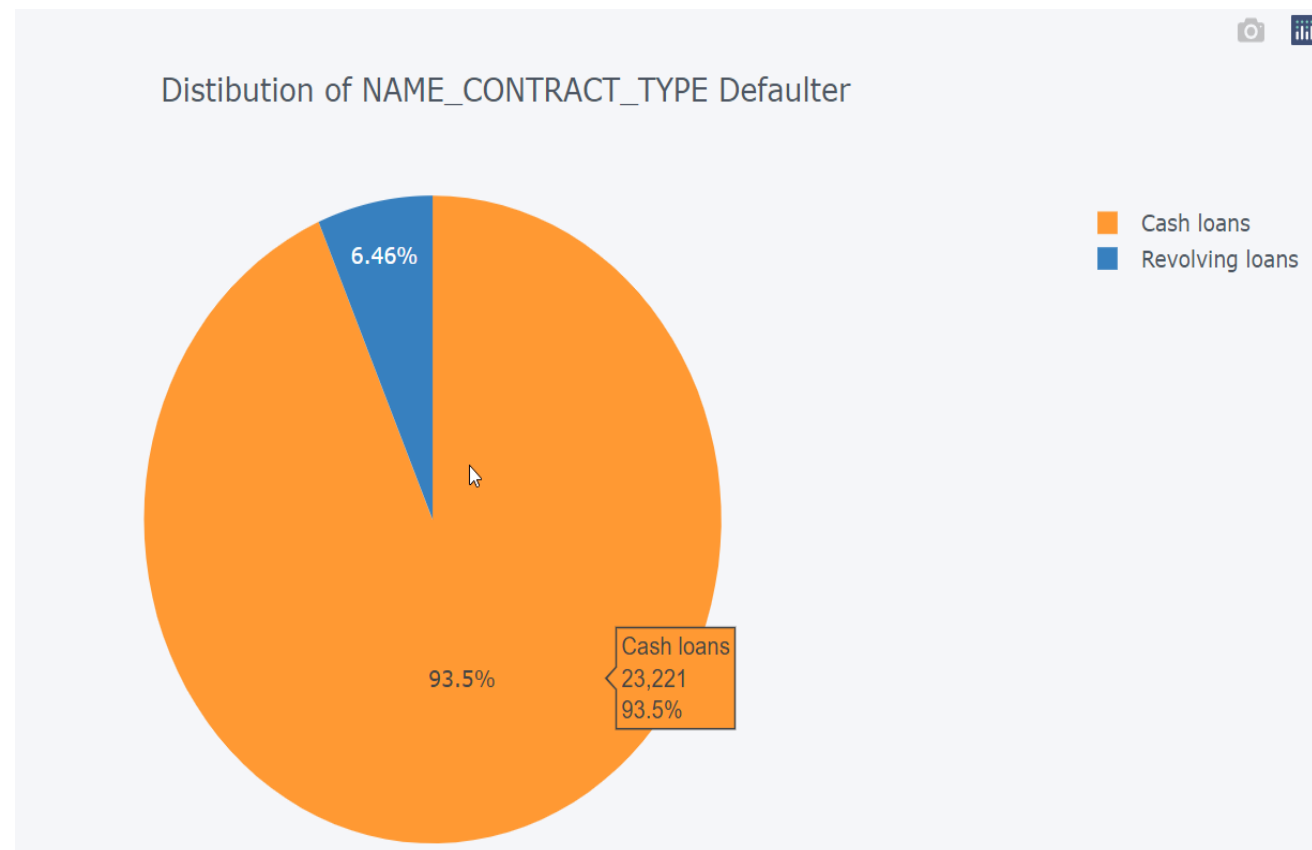
# Loans Contract Distribution against Defaulter

---

By seeing the chart it is clear that people are taking cash loans instead of revolving loans from Bank.

## Comparison:-

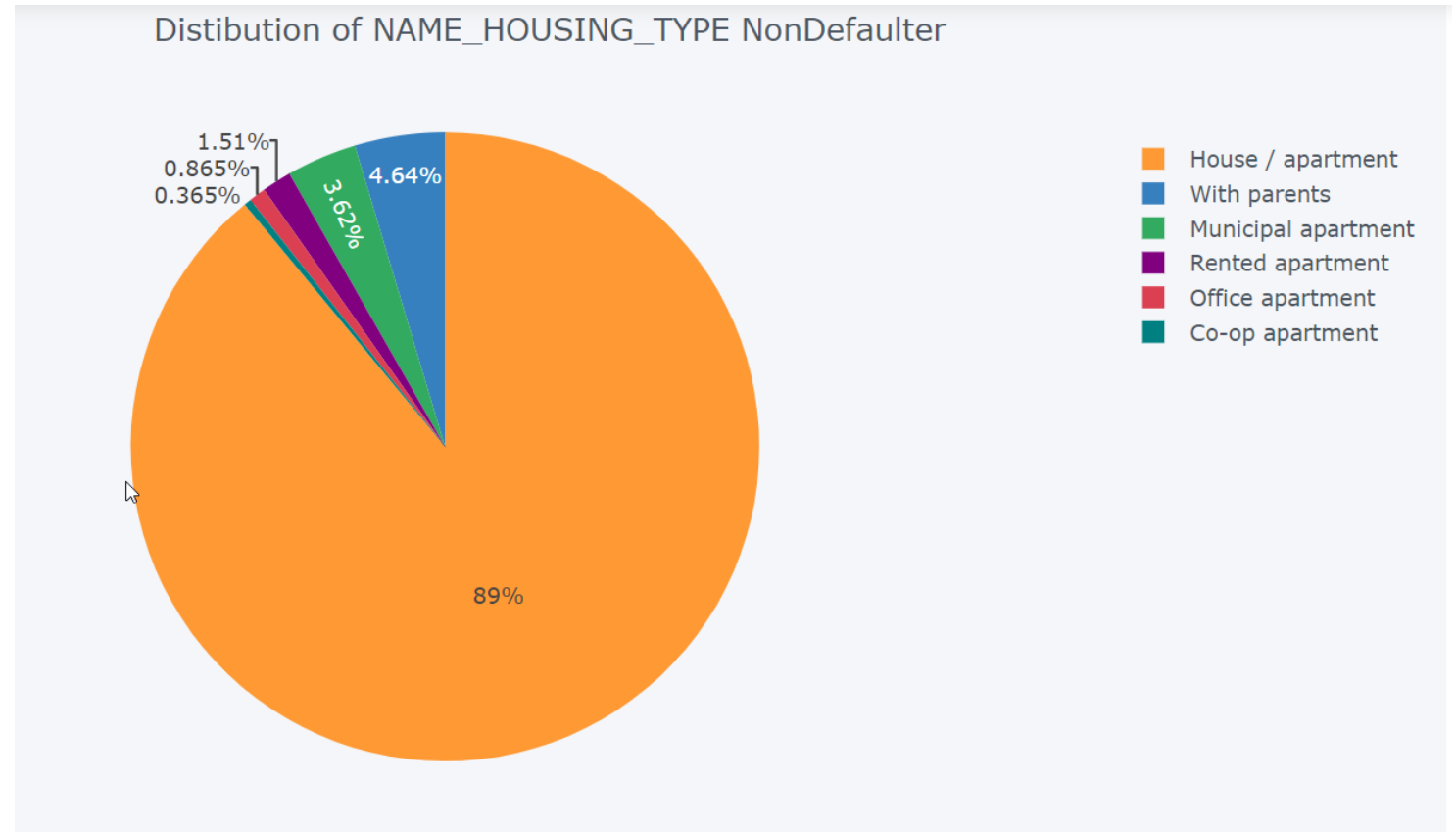
If we compare defaulter chart with non-defaulter in respect with Loan Contract clearly visible 'revolving loans' percentage is significantly lower than 'cash loans'



# Housing Type Distribution against Non- Defaulter

---

By seeing the chart it is clear that people who has their own house or apartment having no difficulties with payment





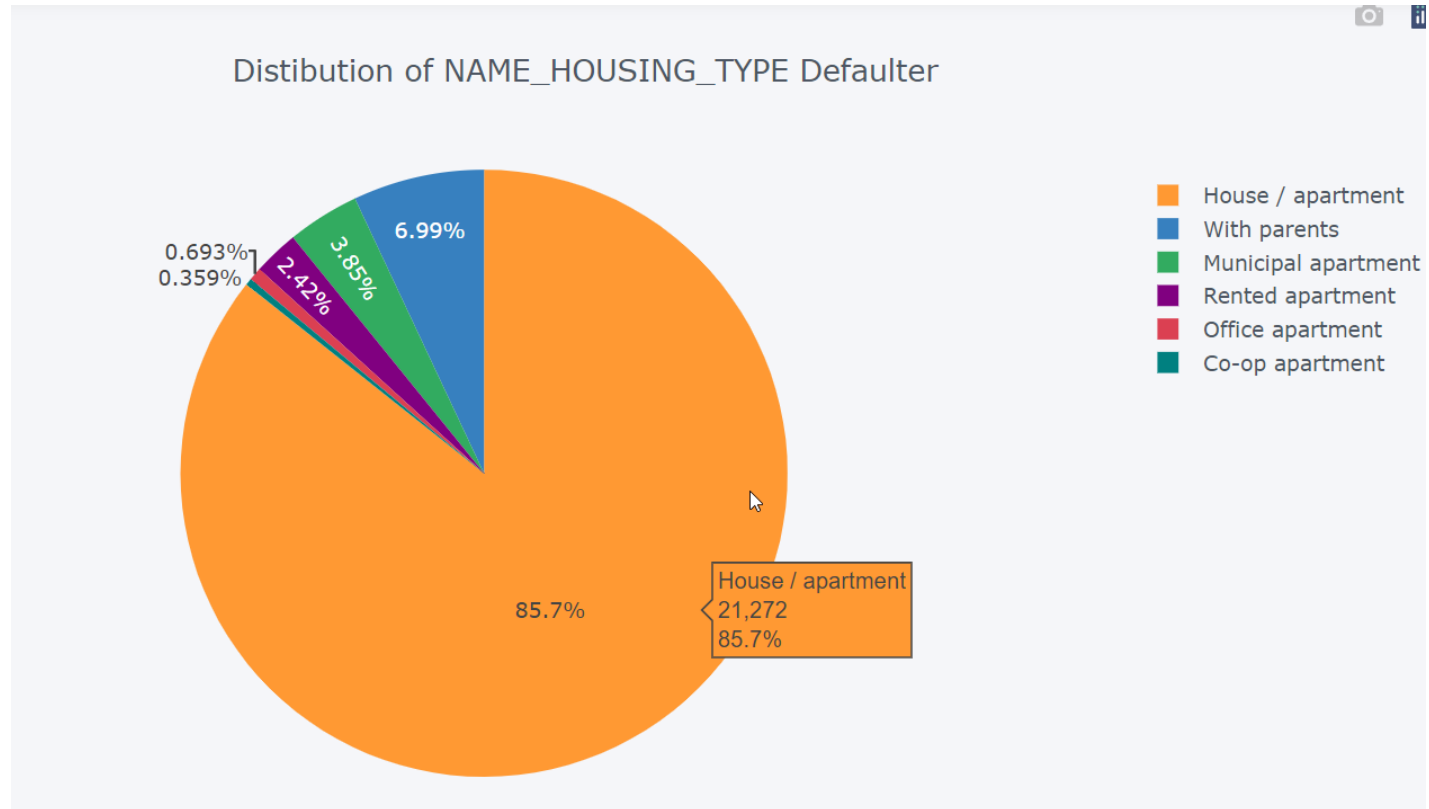
# Housing Type Distribution against Defaulter

---

By seeing the chart it is clear that people who has their own house or apartment already taking more loans

## Comparison:-

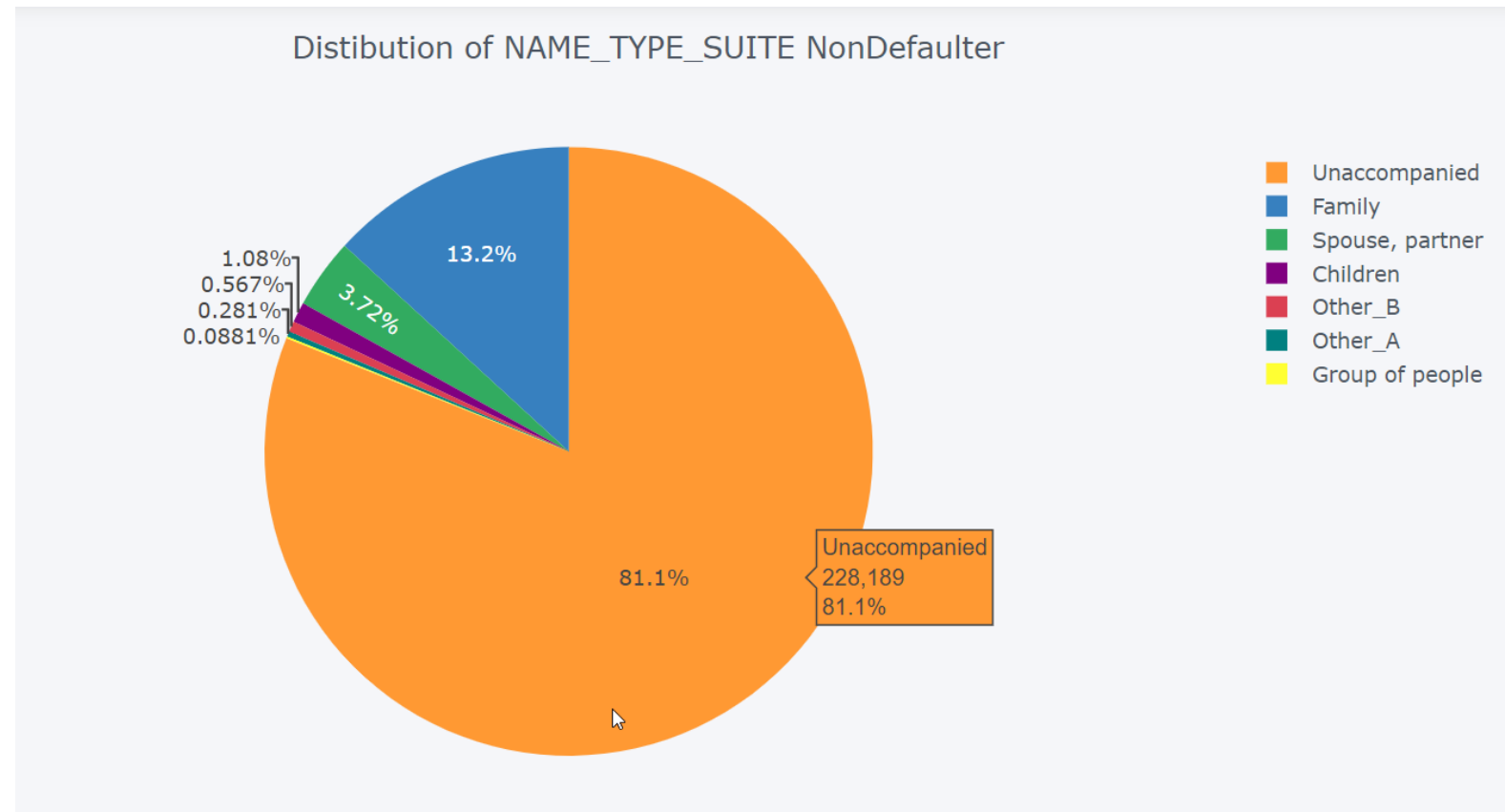
If we compare defaulter chart with non-defaulter in respect with Housing Type-its clearly visible that percentage has increased in case of Rented owners or Living with Parents



# Type Suite Distribution against Non- Defaulter

---

By seeing the chart it is clear that more Unaccompanied people are paying the loan amount in proper time.



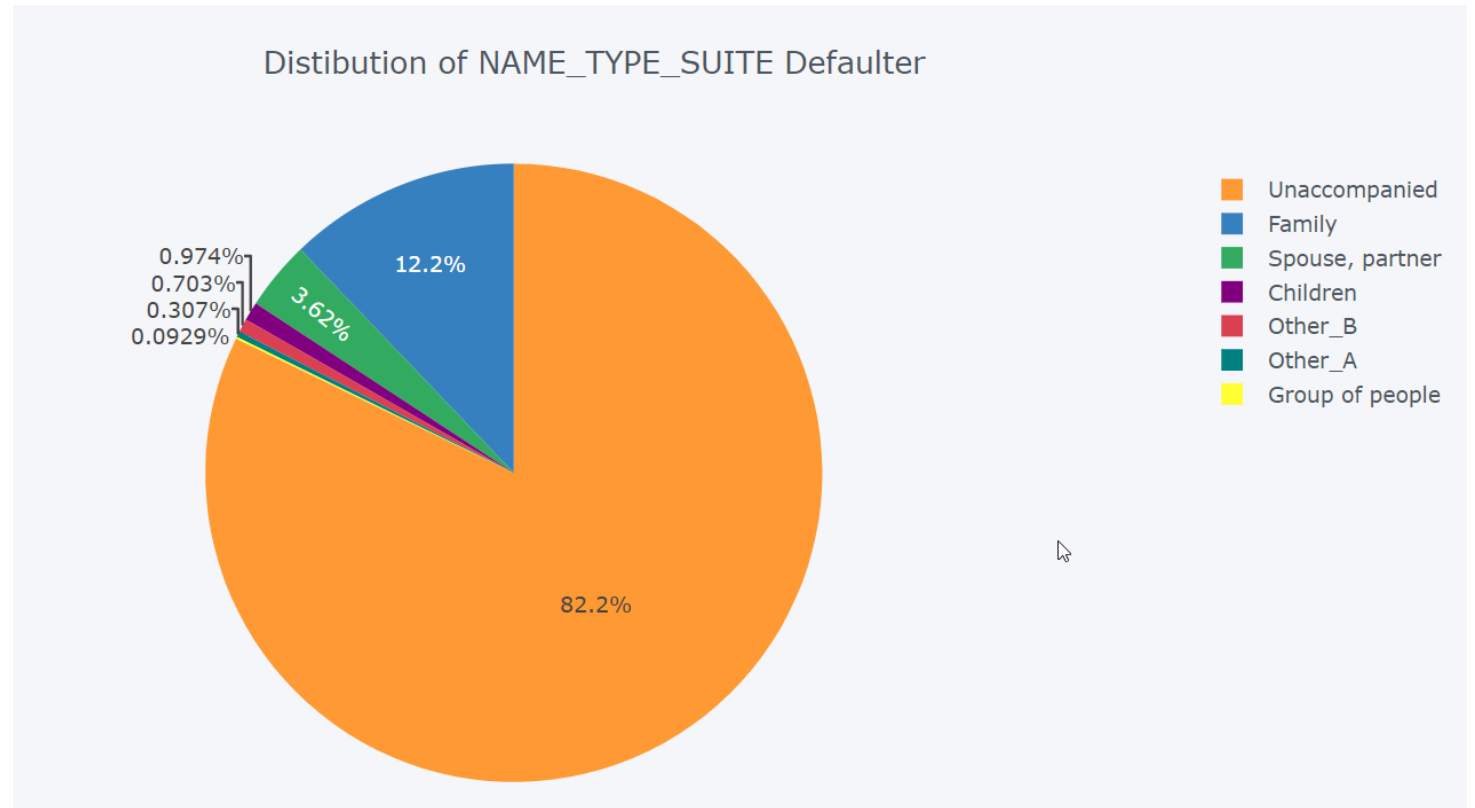
# Type Suite Distribution against Defaulter

---

By seeing the chart it is clear that more Unaccompanied people taking more risk are taking more loan and also facing difficulties with Payment.

## Comparison:-

If we compare daulter chart with non-defaulter in respect with TypeSuite -its clearly visible that Unaccompanied people are more risk taker



---

# Bivariate Analysis of Current Application

# Education Type Vs Credit Distribution against Non-Defaulter

By seeing the chart it is clear married, Civil marriage, Single/not married those are having higher education and taking more credit, though there are some outliers in every education type. But Academic degree holders are having less outlier and they also didn't take credit like Higher Education holder.



# Education Type vs credit Distribution against Defaulter

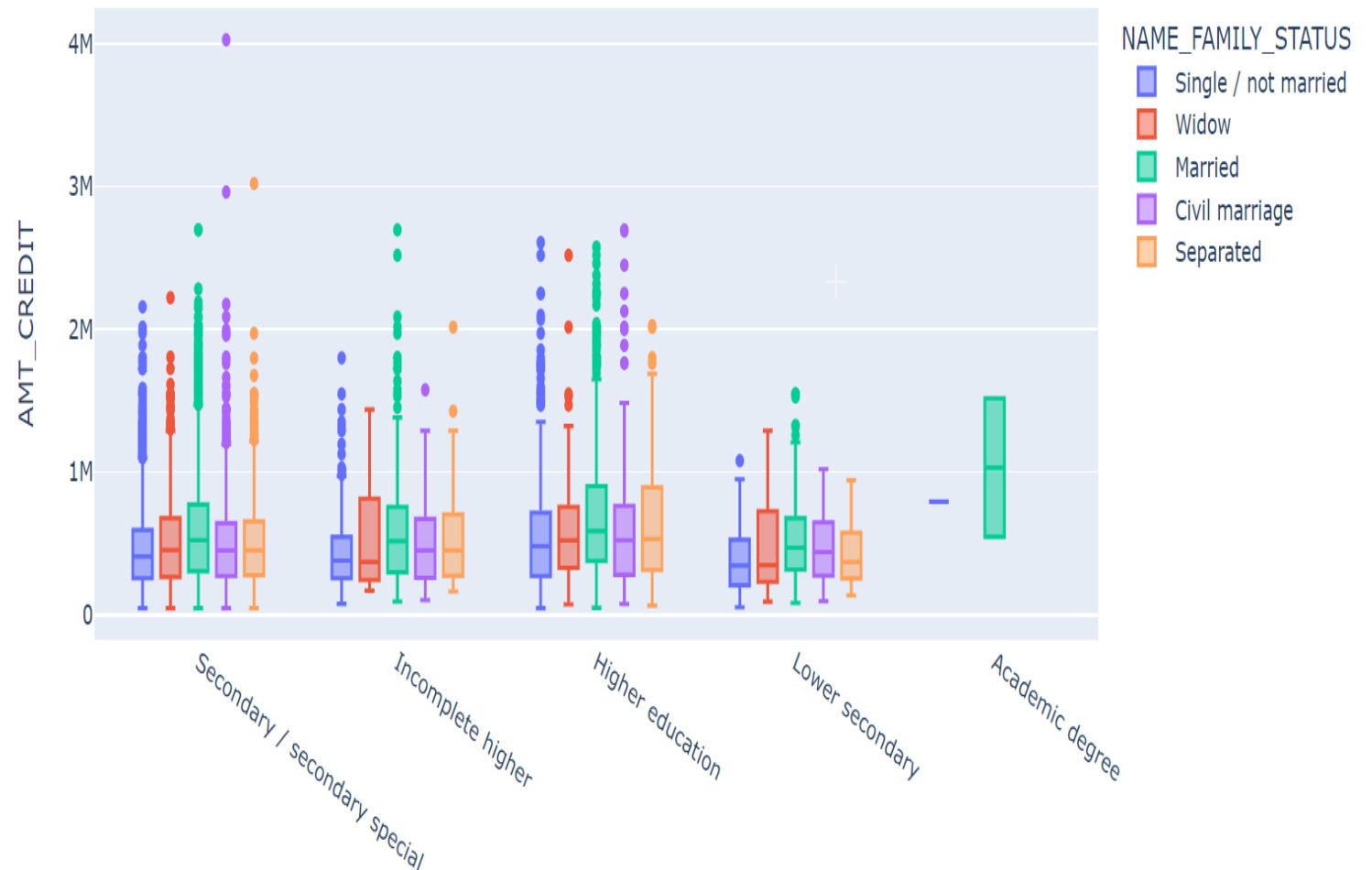
By seeing the chart it is clear married, Civil marriage, Single/not married those are having higher education and taking more credit, though there are some outliers in every education type. But Academic degree holders are having less outlier and they also didn't take credit like Higher Education holder.

.

## Comparison:-

If we compare defaulter chart with non-defaulter in respect with Education Type-its clearly visible outliers are less than non-defaulter and only Married with Academic degree holder is more credit risk taker/defaulter

NAME\_EDUCATION\_TYPE vs AMT\_CREDIT in respect to Defaulter



# Income Type Vs Credit Distribution against Non-Defaulter

By seeing the chart it is clear that there are more outlier in Commercial associate and Businessman specially married / single are the highest , who has no difficulties with payment

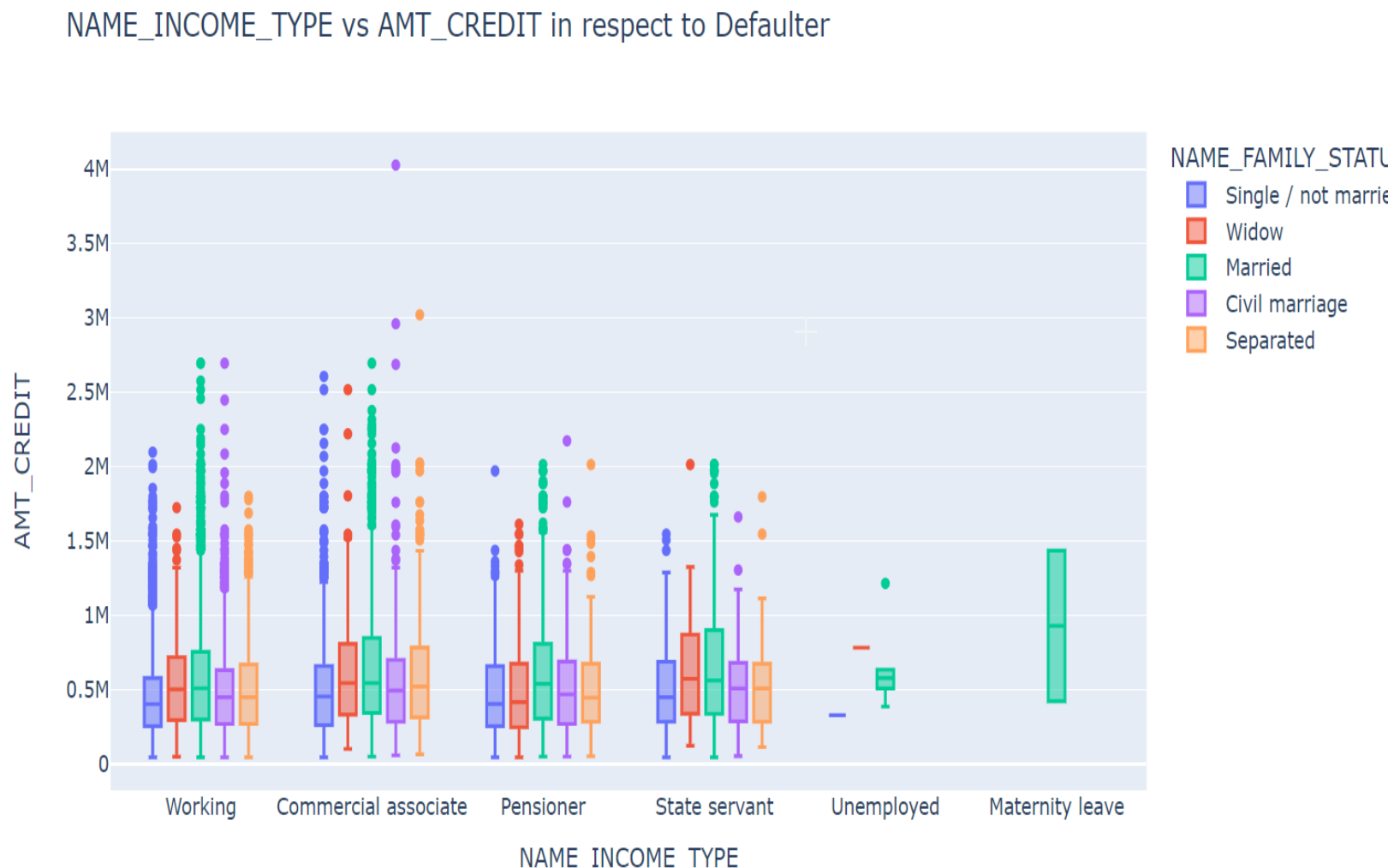


# Income Type Distribution against Defaulter

By seeing the chart it is clear that there are more outlier in Commercial associate and Unemployed people are less risk taker. But married people who are in Maternity leave-more defaulter

## Comparison:-

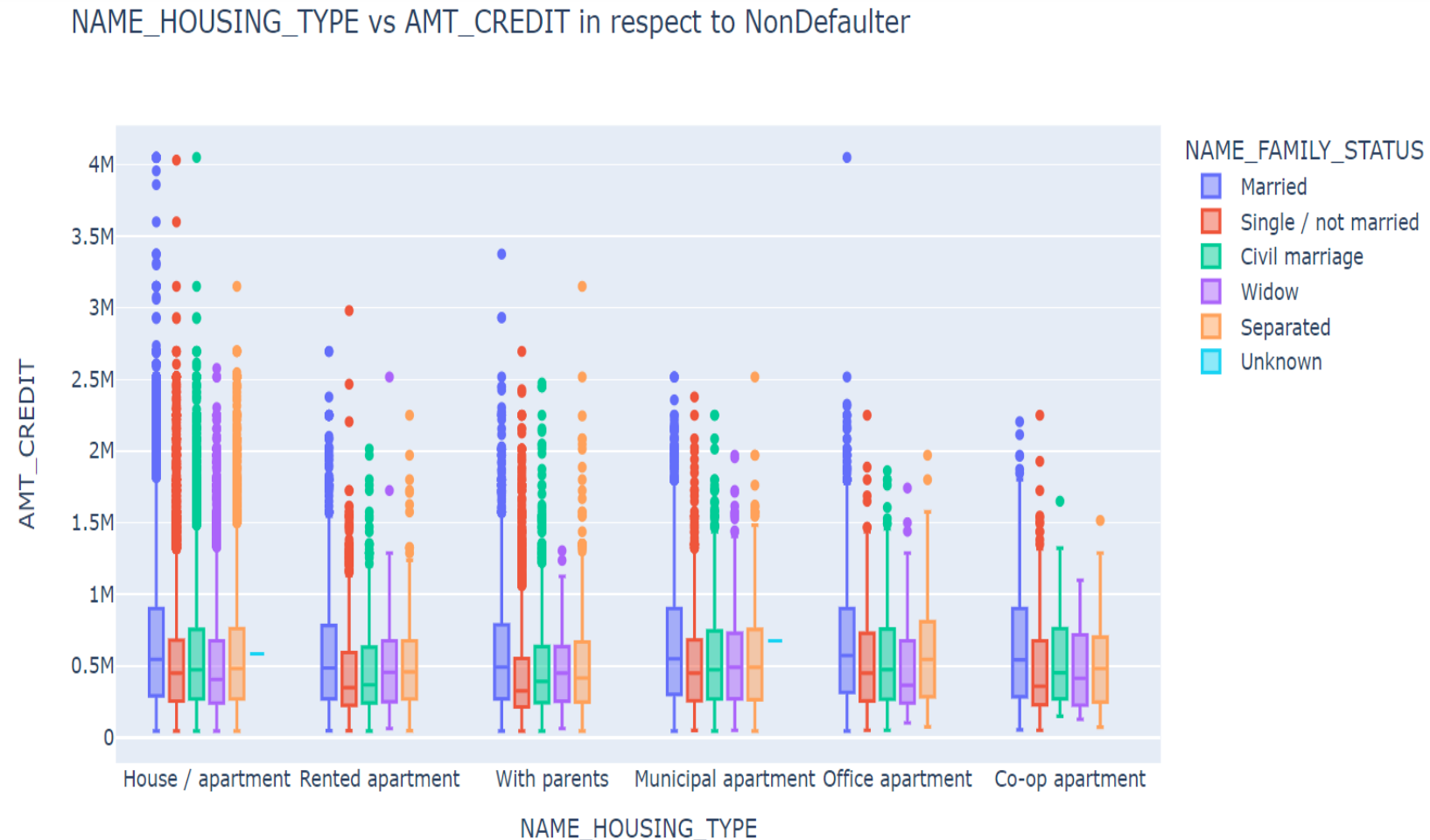
If we compare defaulter chart with non-defaulter in respect with Income Type-its clearly visible that Businessman are not defaulter, so then they paid the loan amount in timely manner





# Housing Type Vs Credit Distribution against Non-Defaulter

By seeing the chart it is clear that more people house/apartment holder is having capability of paying timely and no payment difficulties.

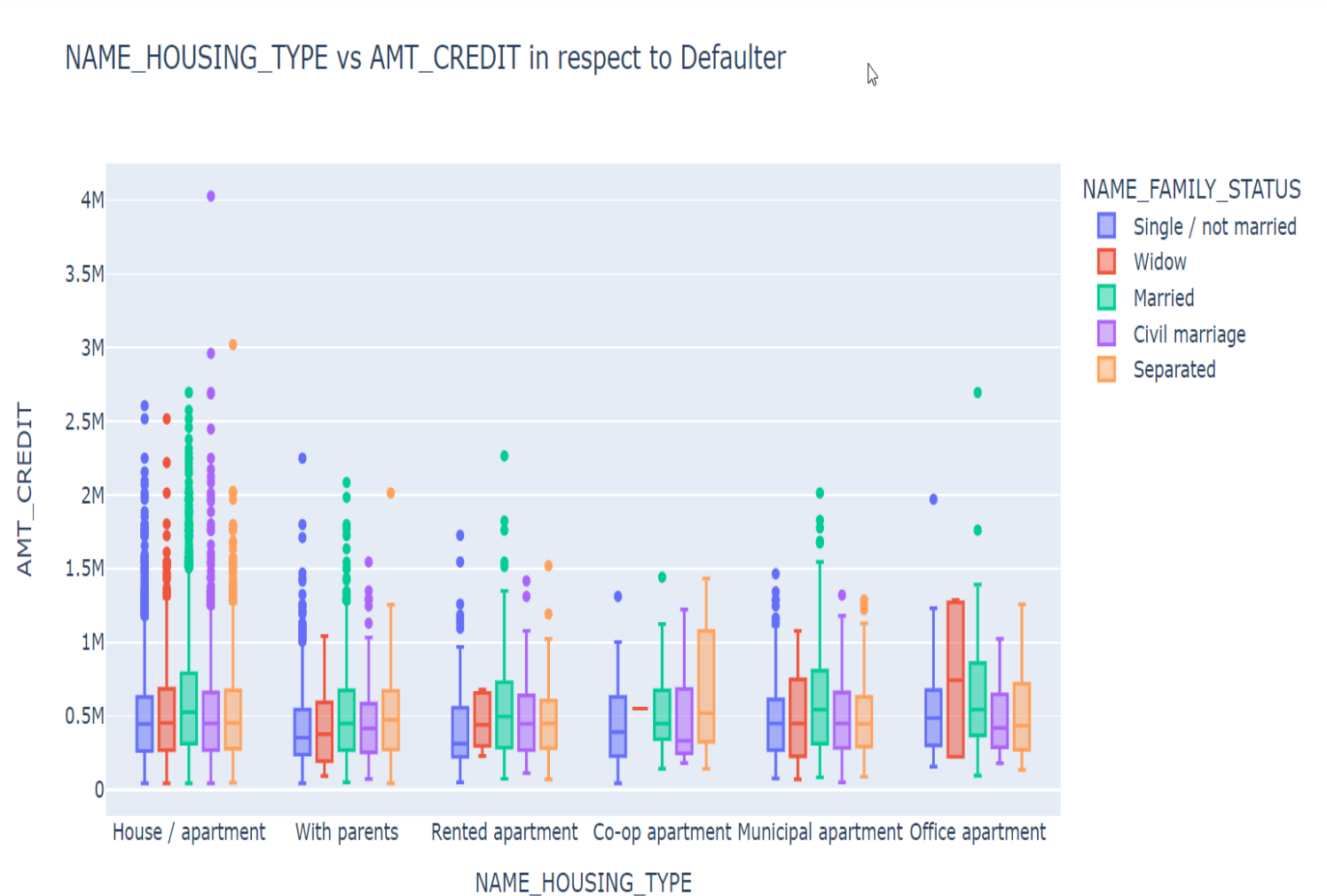


# Housing Type Vs Credit Distribution against Defaulter

By seeing the chart it is clear that more people office apartment holder is having some more difficulties of payment than others.

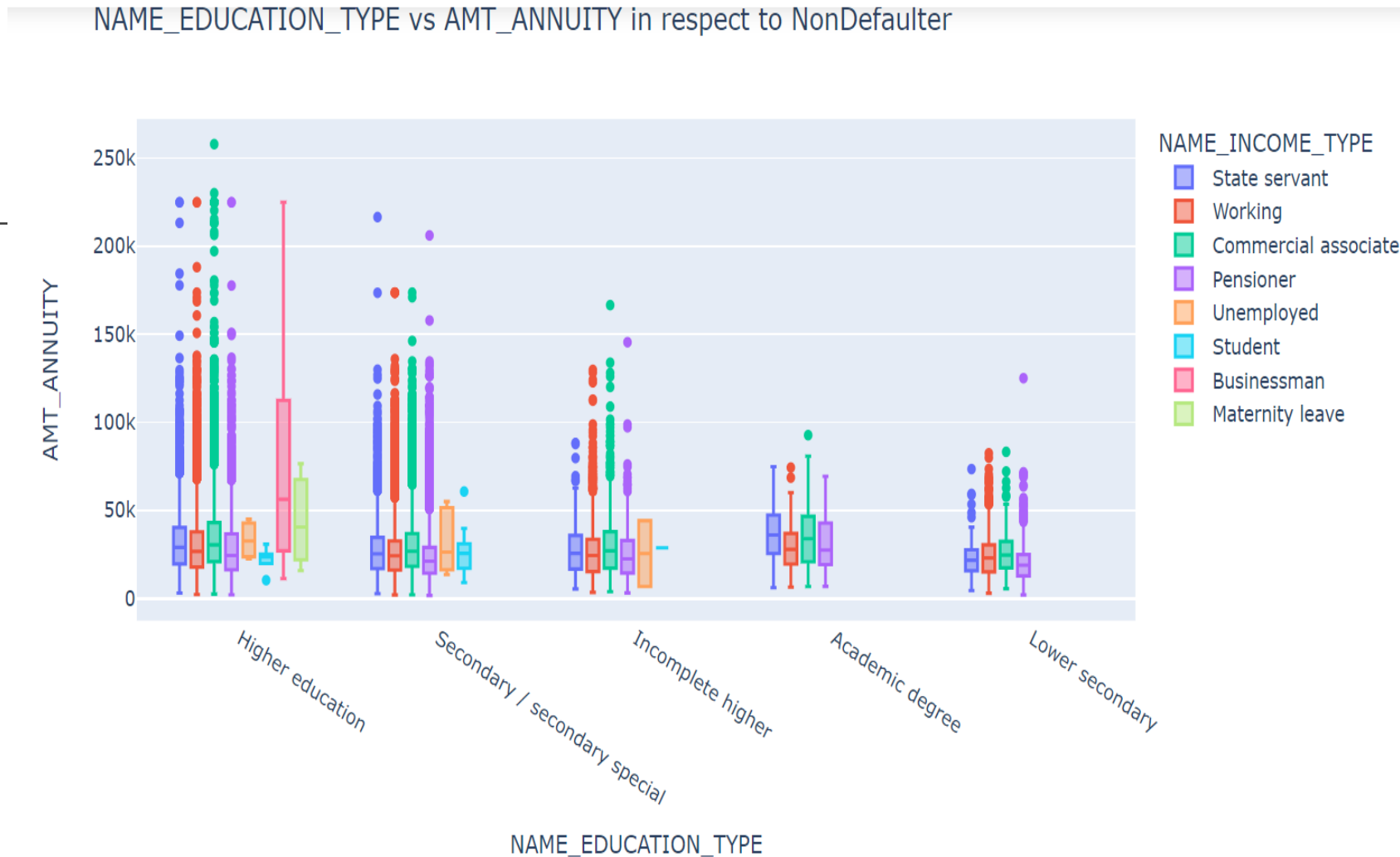
## Comparison:-

If we compare defaulter chart with non-defaulter in respect with Housing Type-its clearly visible that Married & Civil marriage people are more tend to take loan, so they are most influencer here for credit risk



# Education Type Vs Annuity Distribution against Non-Defaulter

By seeing the chart it is clear that Businessman with Higher Education and high annuity having capability of taking loan and paying back in proper time

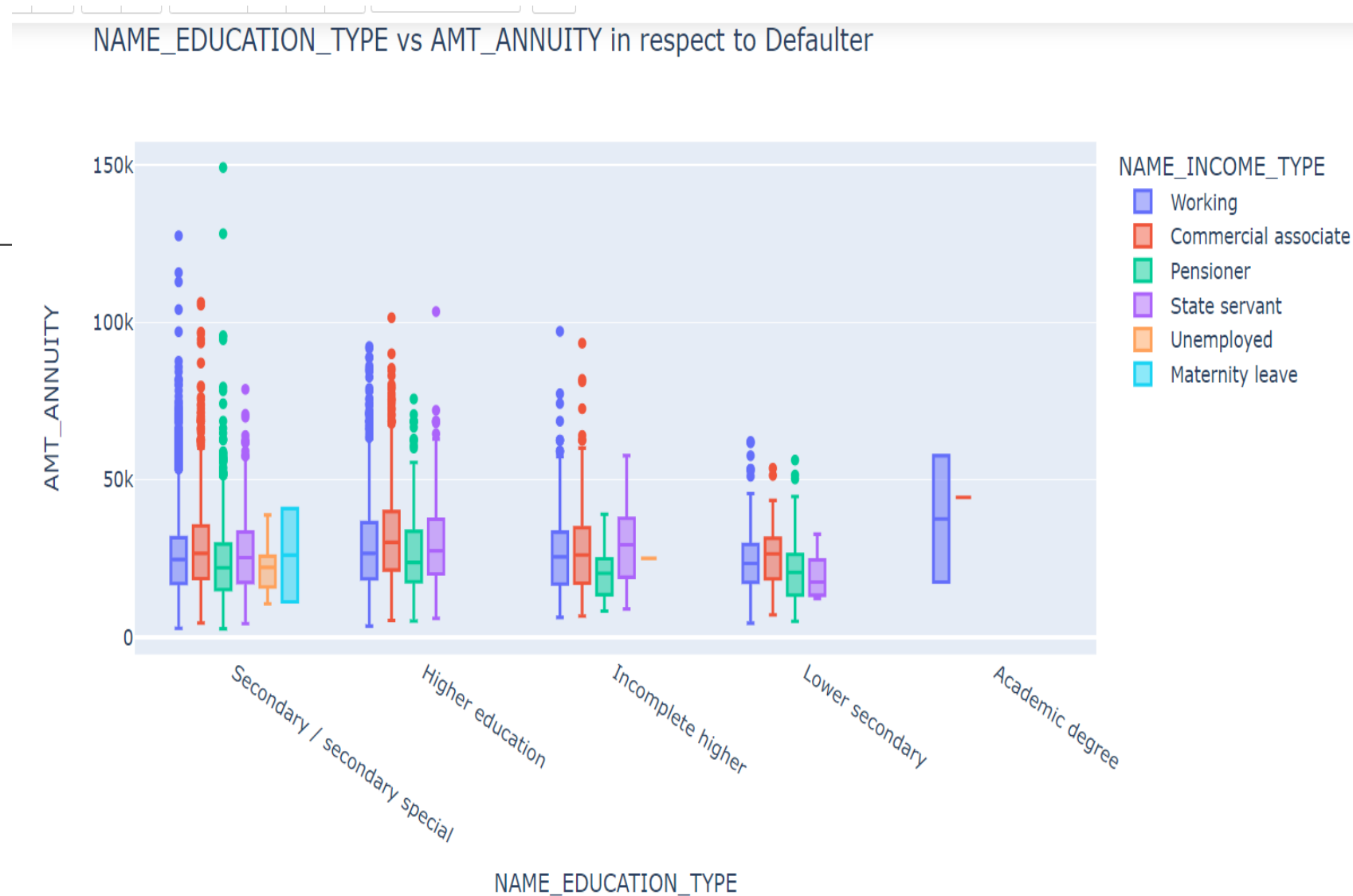


# Education Type Vs Annuity Distribution against Defaulter

By seeing the chart it is clear that working class people who are having academic degree and high annuity facing loan payment difficulties

## Comparison:-

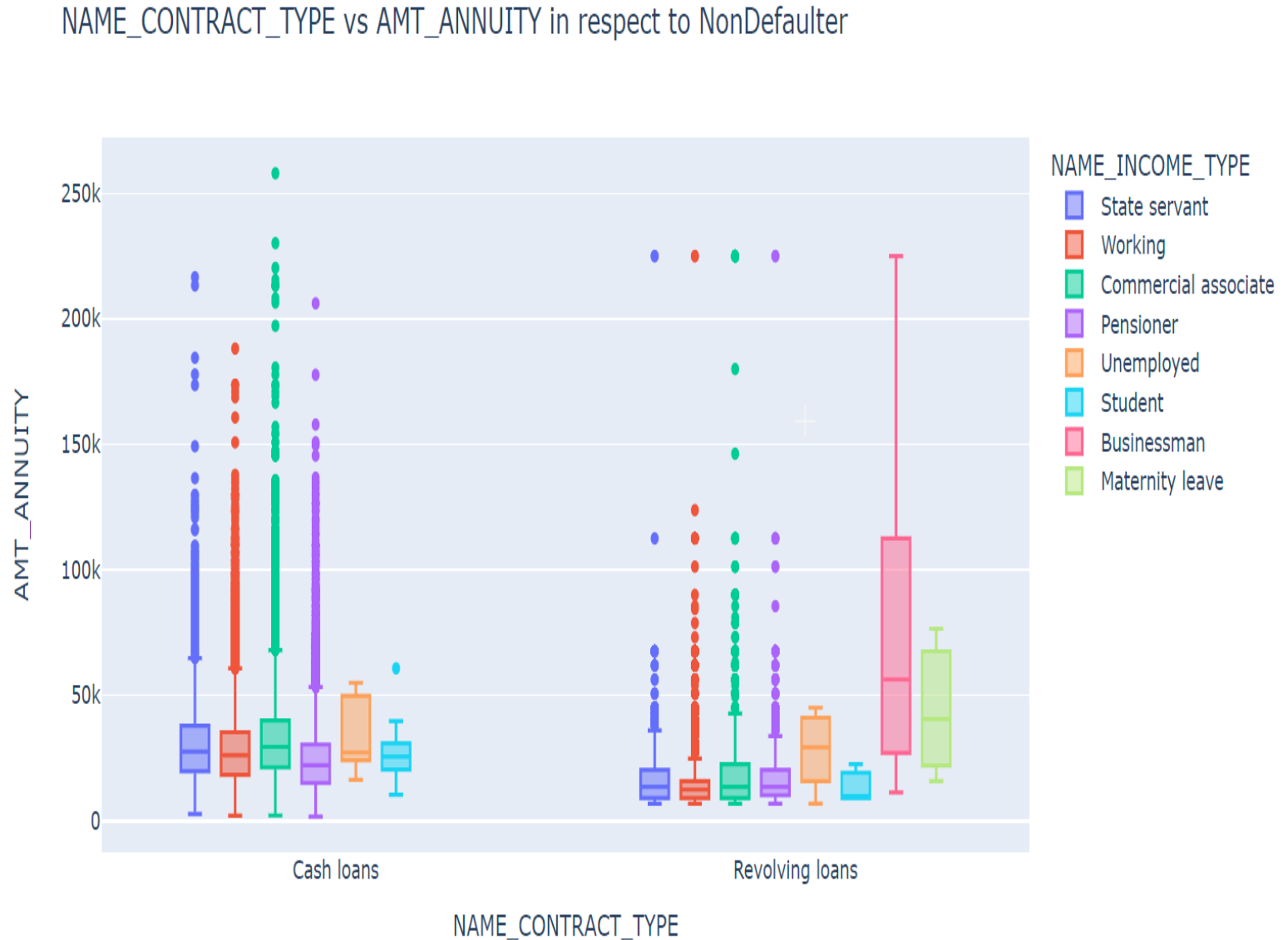
If we compare defaulter chart with non-defaulter in respect with Education Type-its clearly visible people are more likely to be more non defaulter history rather than defaulter history



# Contract Type Vs Annuity

## Distribution against Non- Defaulter

By seeing the chart it is clear that more working-class people are using revolving loans with good Annuity. So, they are more likely to pay in timely manner

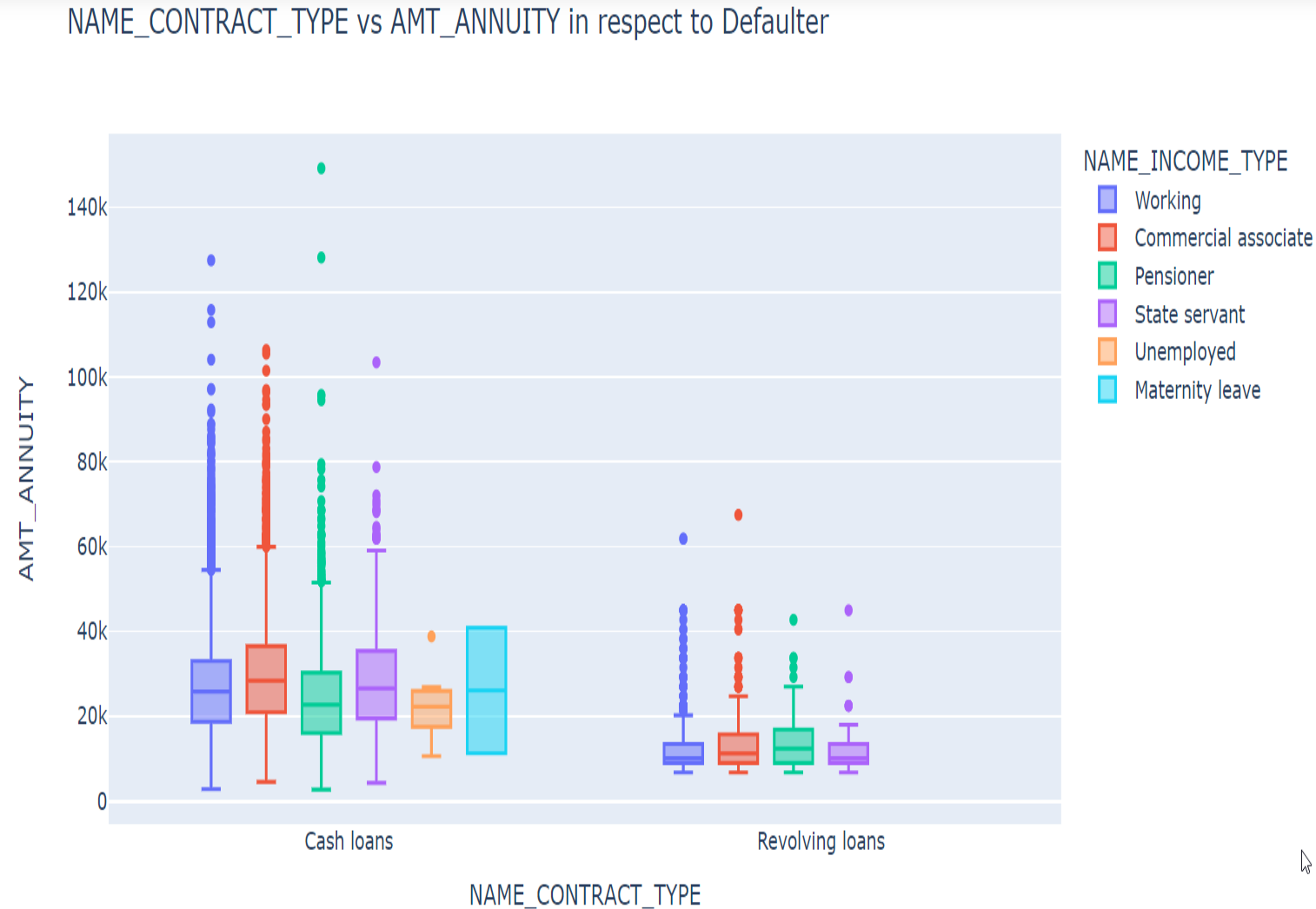


# Income Type Vs Annuity Distribution against Defaulter

By seeing the chart it is clear that more people with maternity leave are using cashloans with good Annuity, but they are facing difficulties in payment

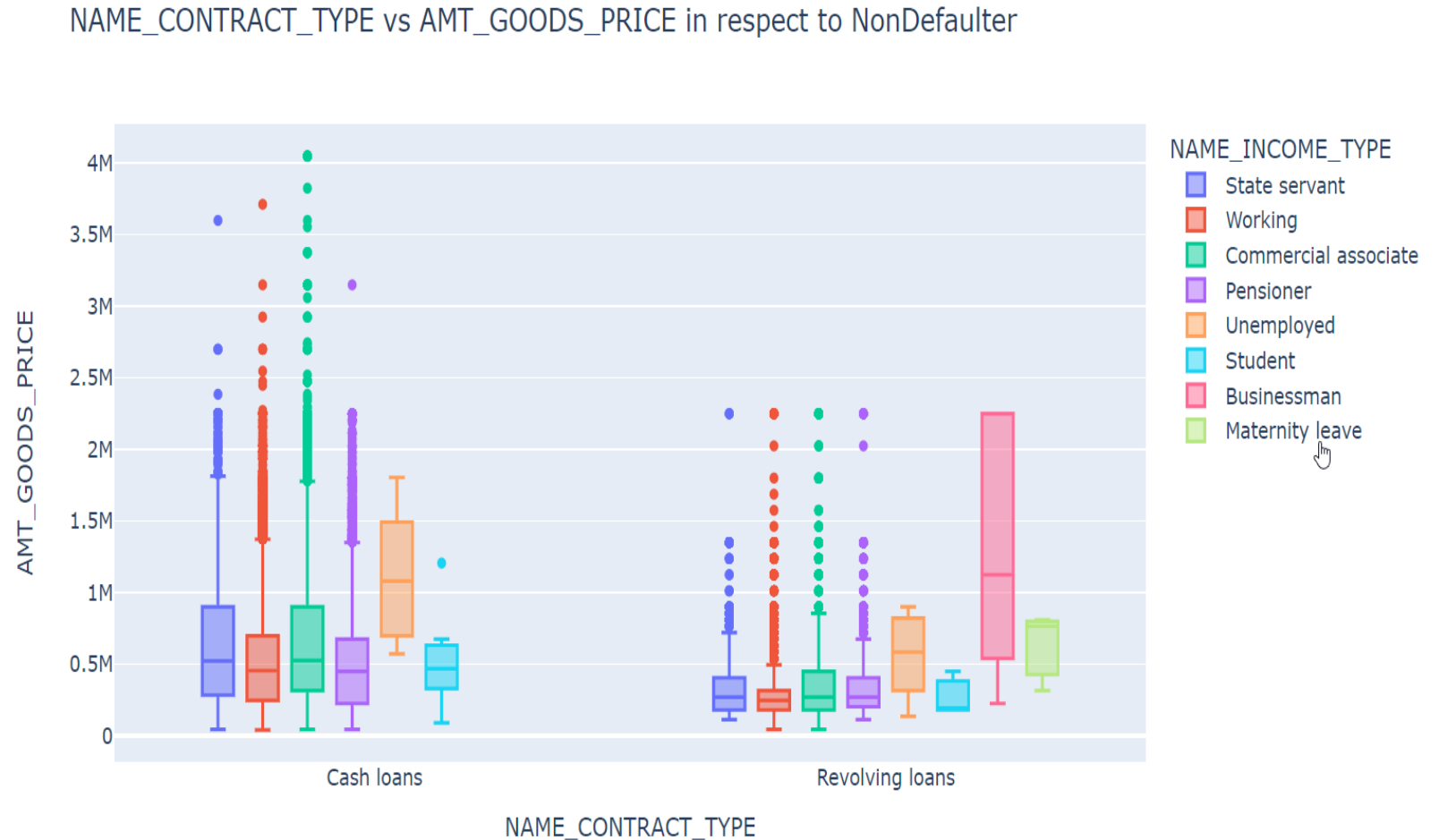
## Comparison:-

If we compare defaulter chart with non-defaulter in respect with Income Type-its clearly visible that cash loan owners rather than revolving loan takers are more like to face payment difficulties



# Contract Type Vs Goods Price Distribution against Non- Defaulter

By seeing the chart it is clear that more Businessman with revolving loan not facing any payment difficulties also even when the Goods prices are high.

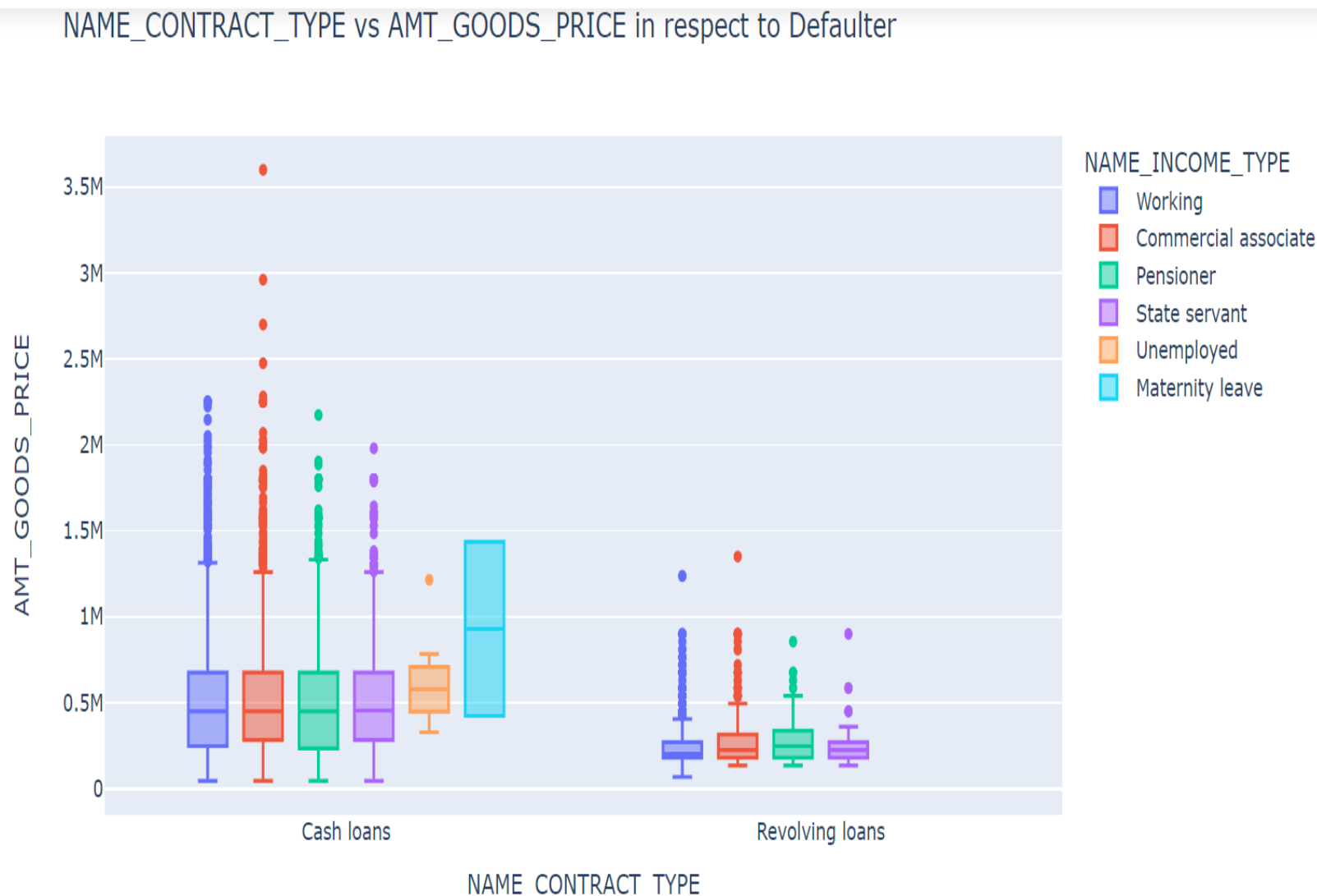


# Contract Type Vs Goods Price Distribution against Defaulter

In case of defaulters also the chart is more or less same like non defaulters

## Comparison:-

If we compare defaulter chart with non-defaulter in respect with Contract Type-its clearly visible that cash loan takers are played the most influencer.

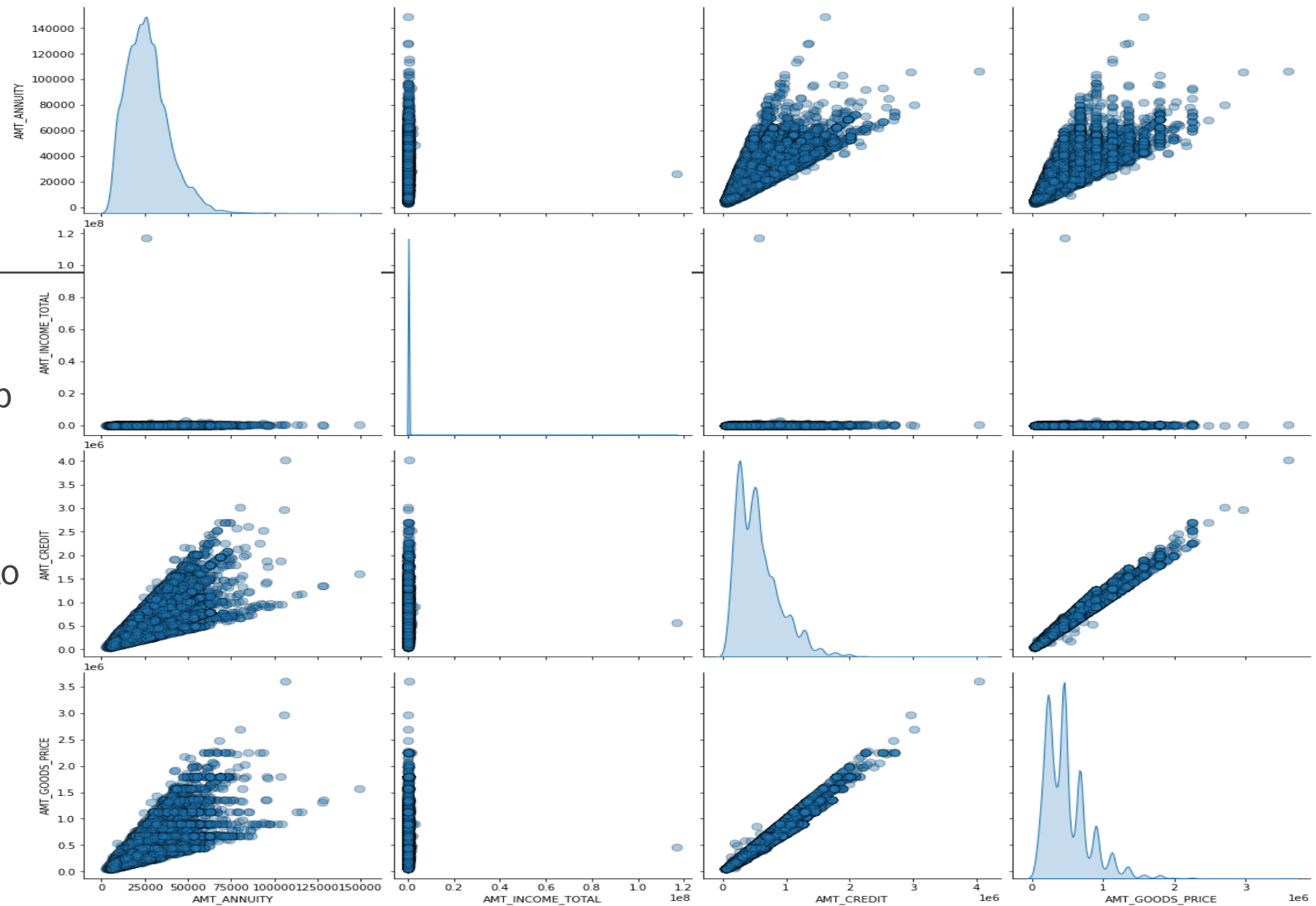




# Pair-plotting for Defaulters

From this pair-plot it is very much clear that relationship between Goods Price and Credit very strong.

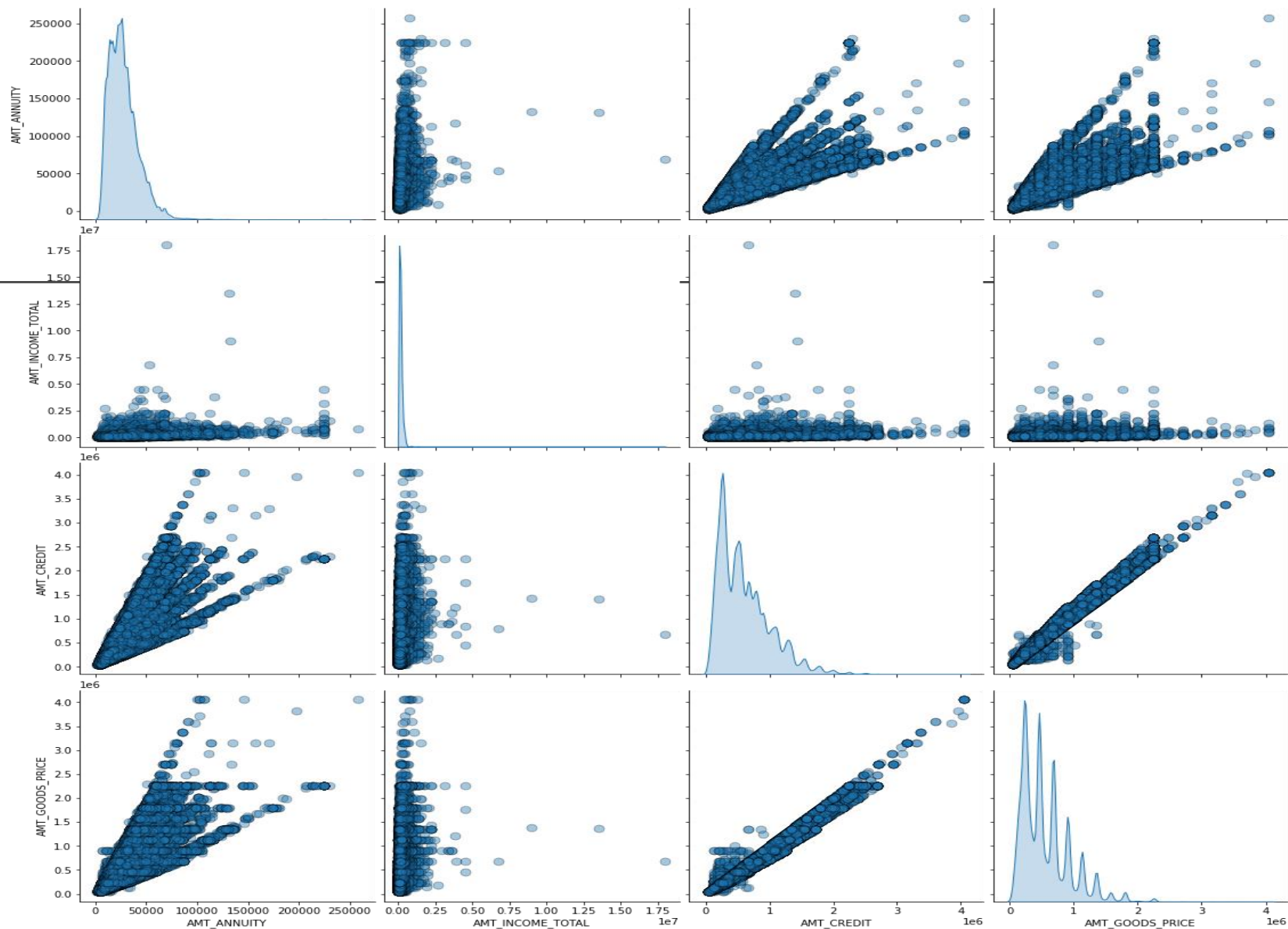
They are highly interlinked to each other.



# Pair-plotting for Non- Defaulter

Same as Defaulter graph

Goods Price and Credit are highly interlinked to each other.



---

# Segmented Univariate Analysis

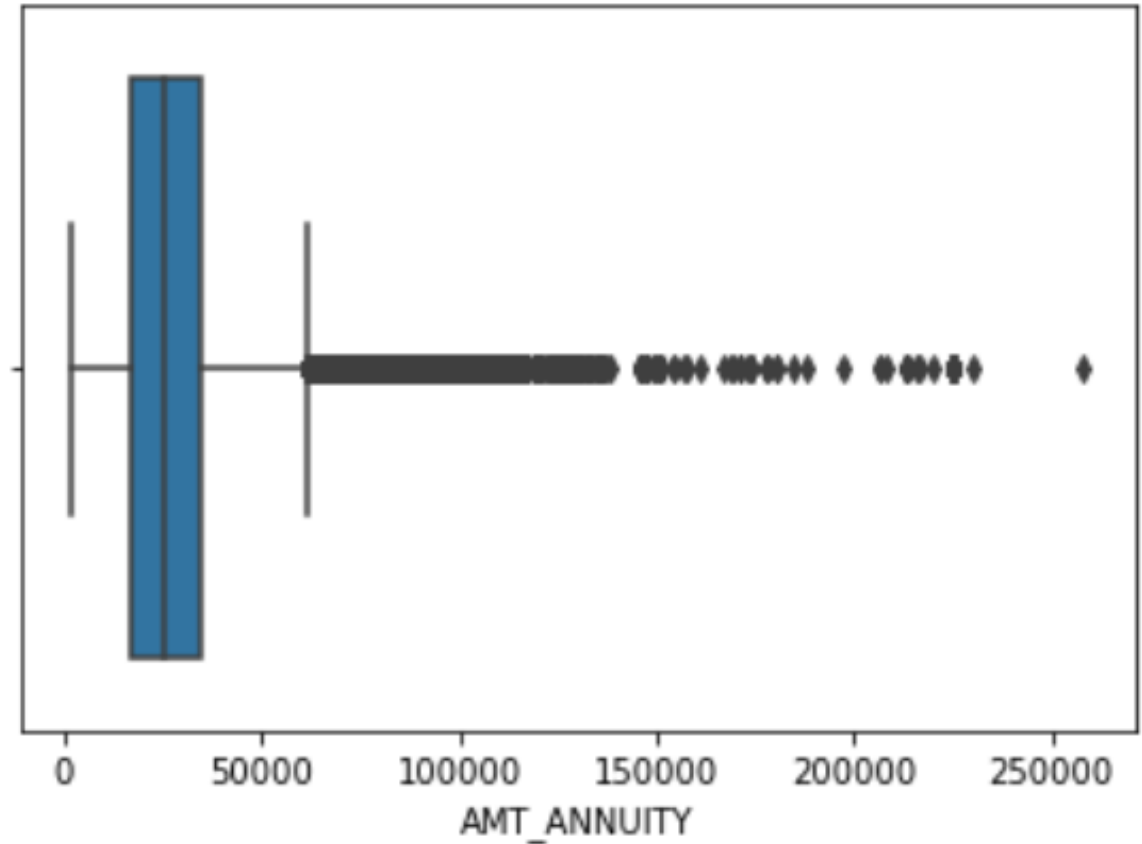
# Annuity Amount Boxplot Distribution against Non- Defaulters & Defaulters

---

There is a high outlier which is more than 250000

And there is more outlier in-between 100000 to 150000

More specifically sub plotting can be visible in next boxplot

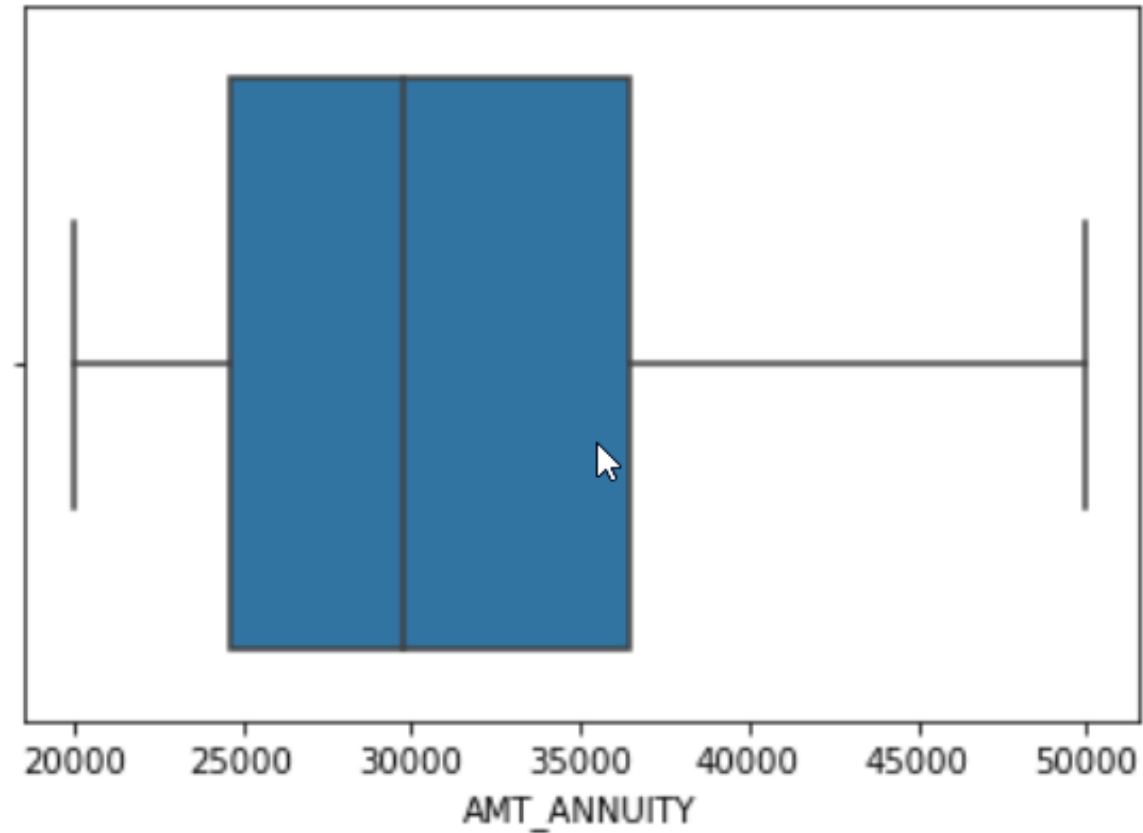


## Income Type Distribution against Non- Defaulter

---

So here its clear that range in-  
between 24000 to 35000 are  
not having outliers

<AxesSubplot:xlabel='AMT\_ANNUITY'>



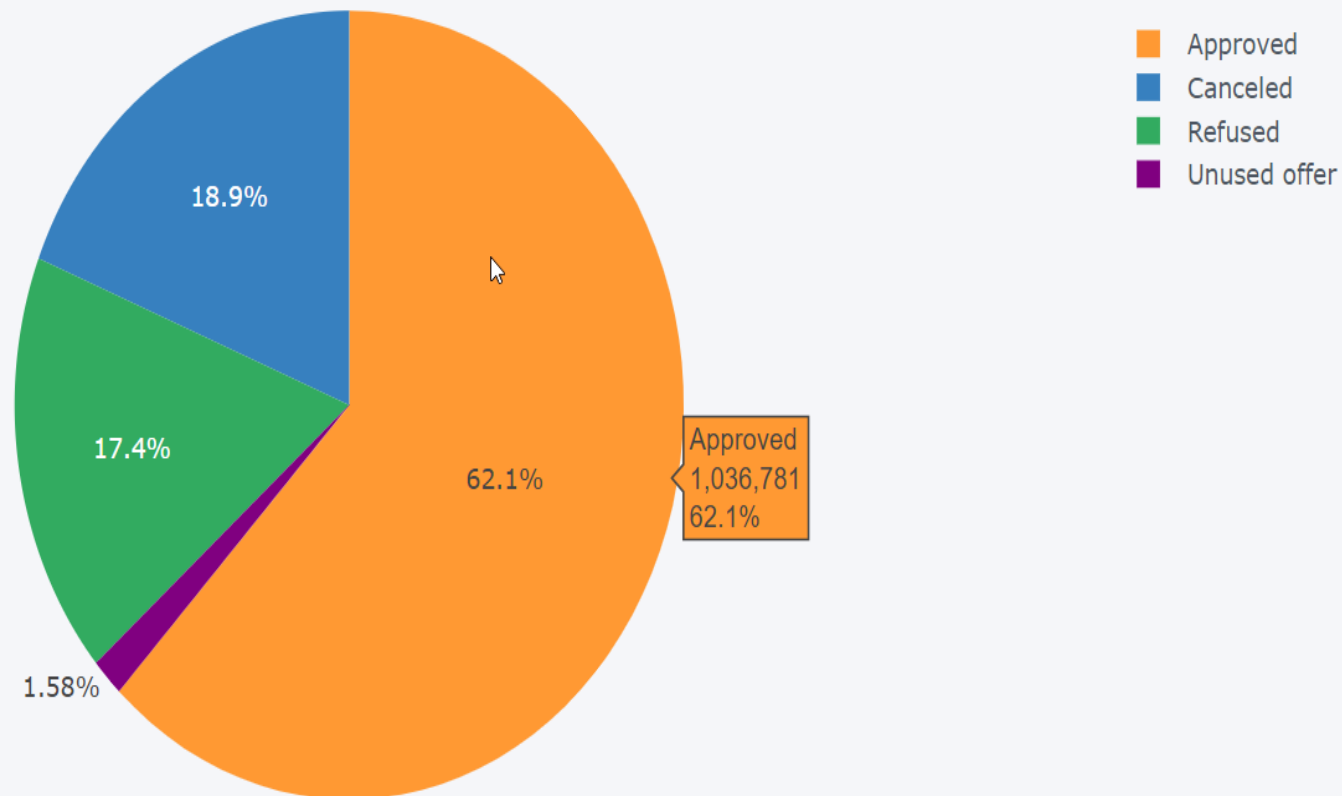
---

# Univariate Analysis on Previous Application

# Contract Status Distribution

In previous application loan approval amount is more and loan refusal percentage is very less .

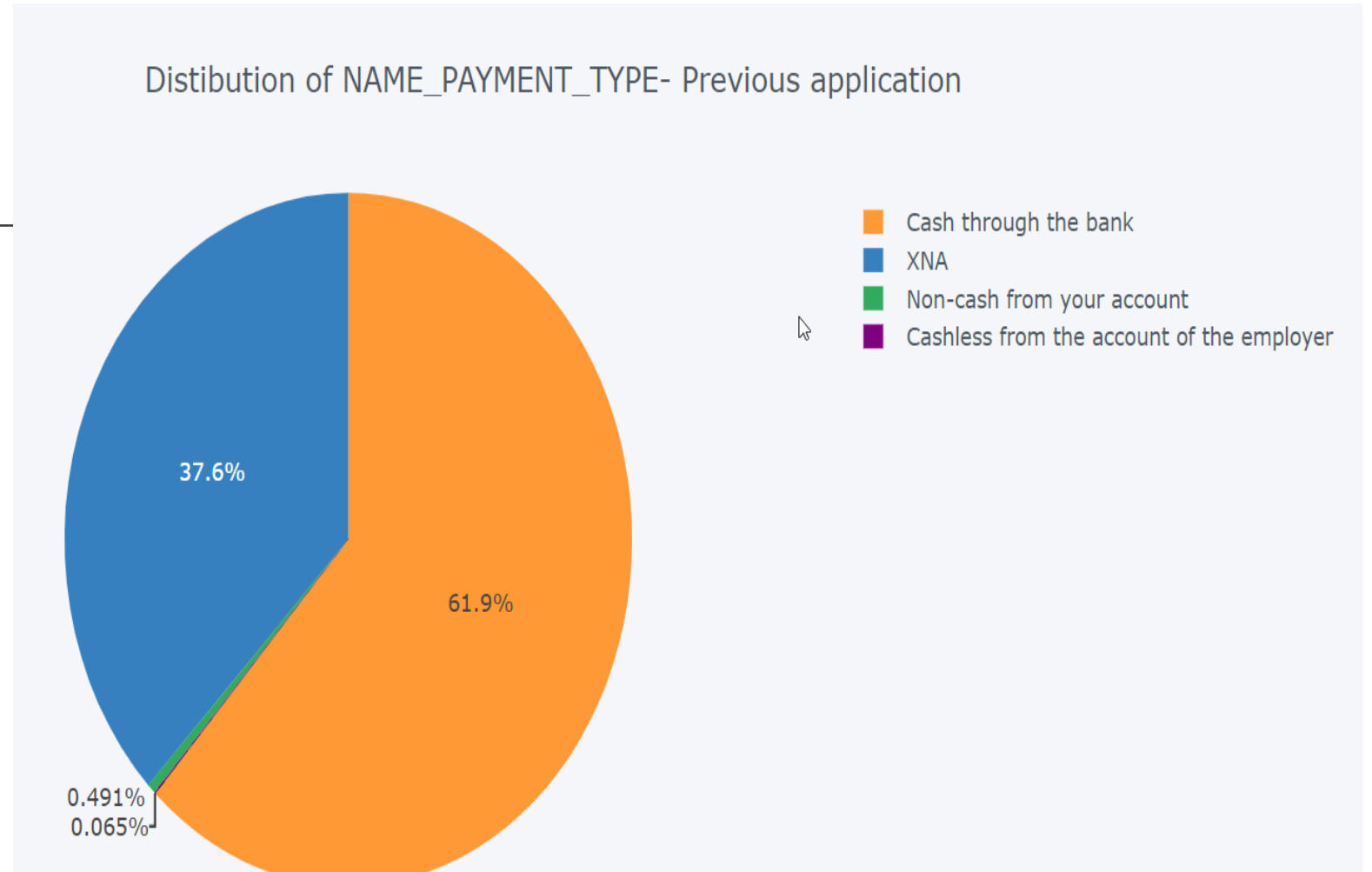
Distribution of NAME\_CONTRACT\_STATUS- Previous application



# Payment Type Distribution

---

In previous application loan cash through the bank is higher, which is more than 50%

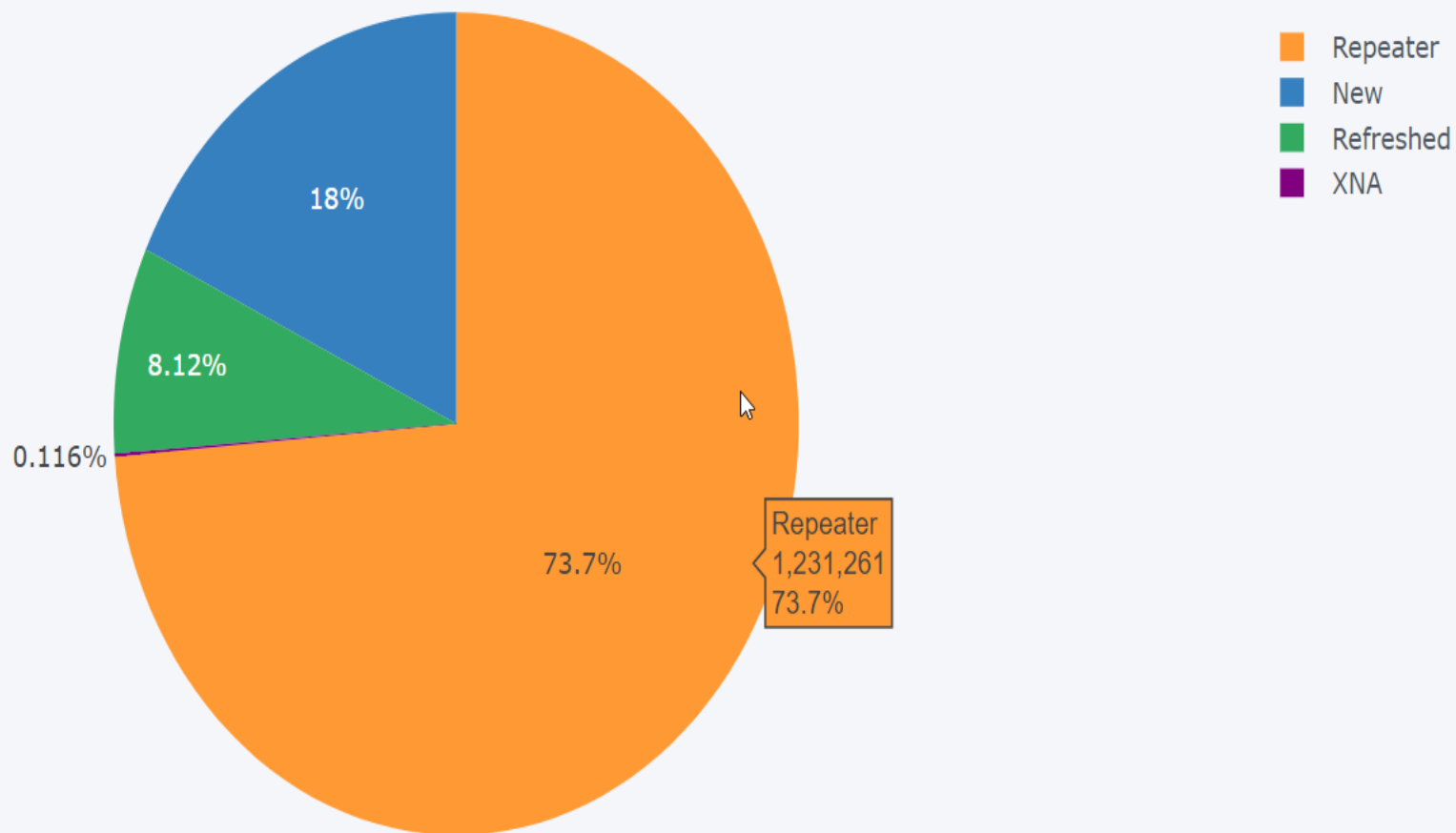




# Client Type Distribution

In previous application loan Repeater is higher, which is more than 50%

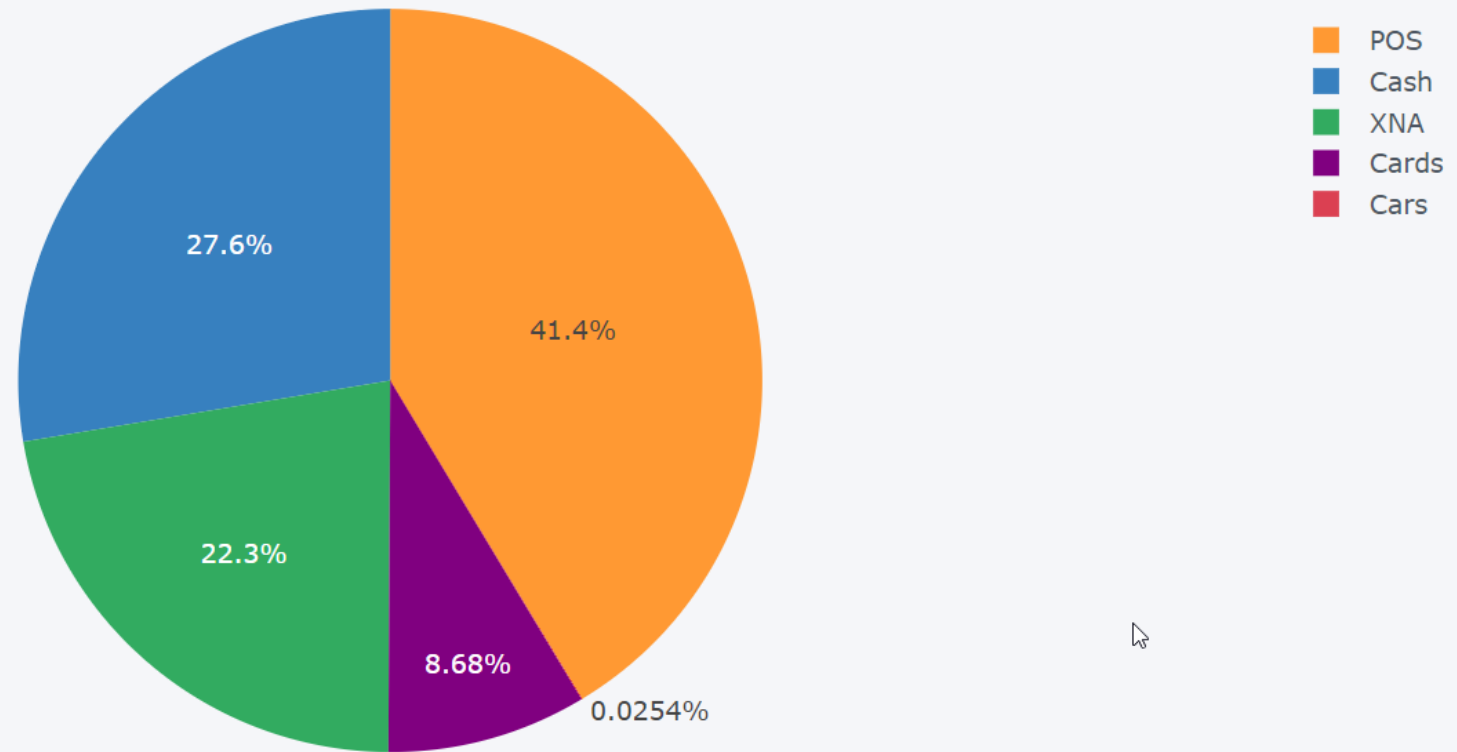
Distribution of NAME\_CLIENT\_TYPE- Previous application



# Portfolio Distribution

In previous application POS  
Portfolio is higher than cash  
portfolio

Distribution of NAME\_PORTFOLIO- Previous application



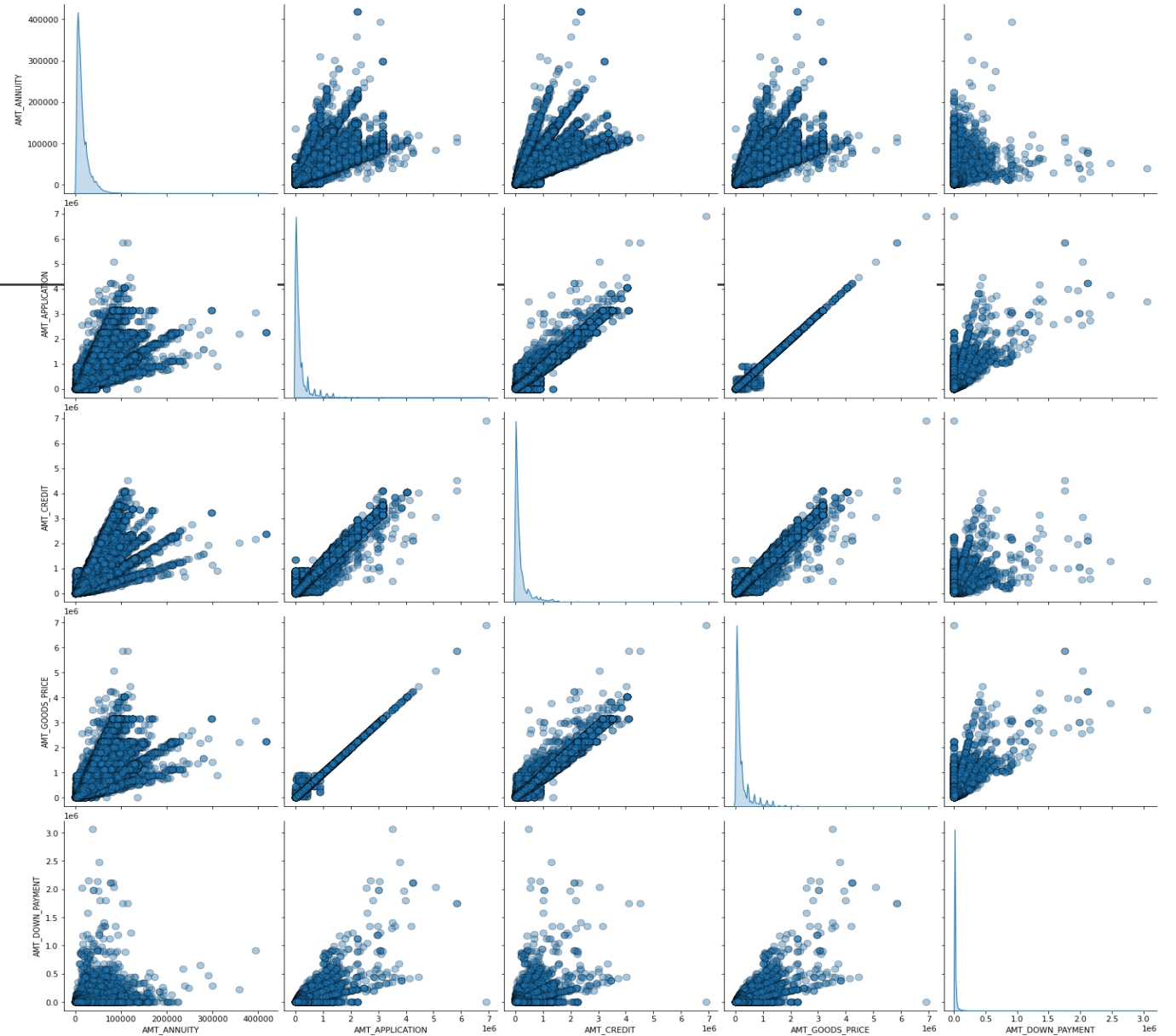
---

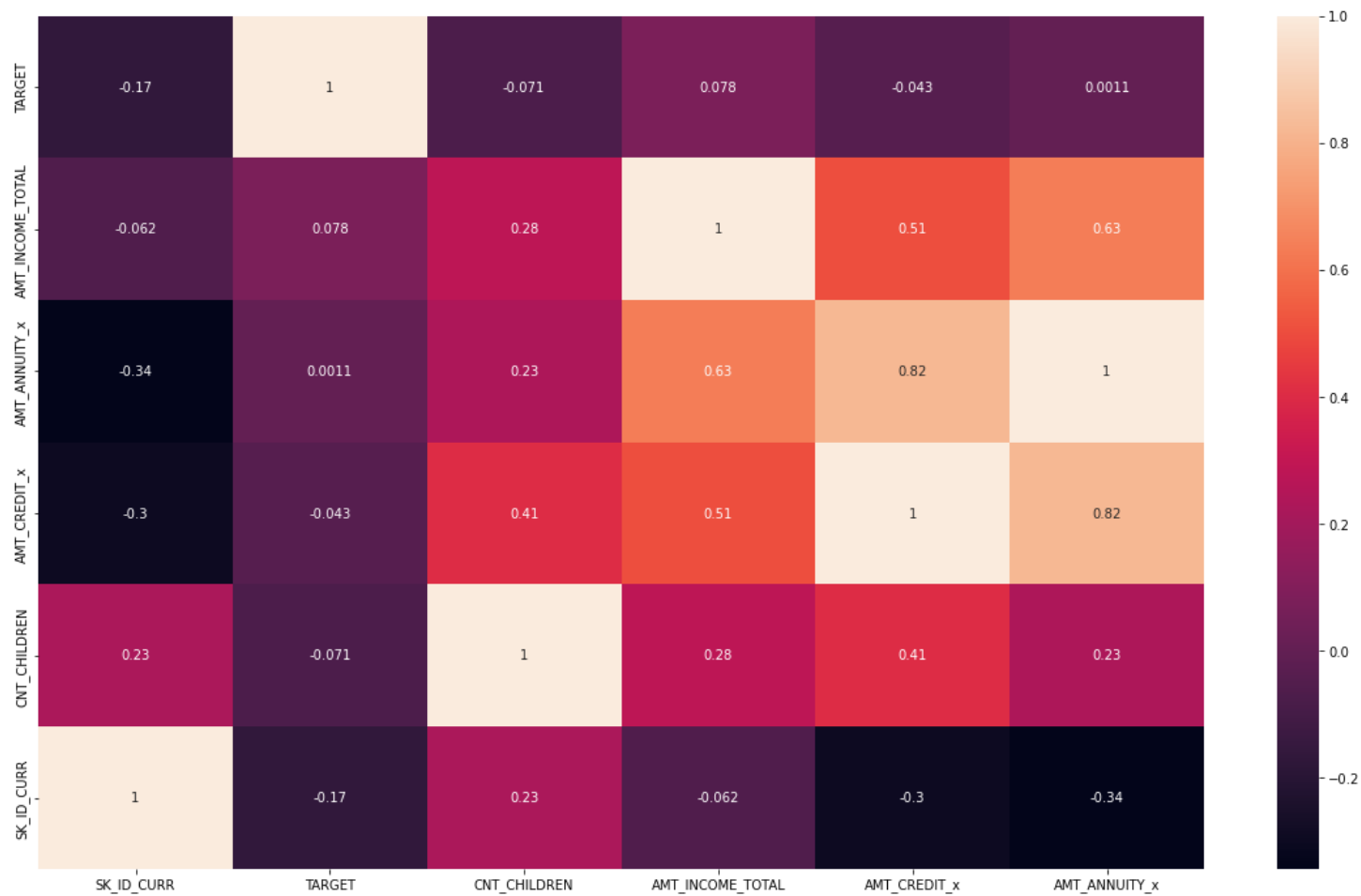
# Bivariate Analysis on Previous Application

# Pair-plotting for Previous Application

Credit amount is interlinked with Goods Price highly

And Amount of application also is highly connection with Goods Price





Income, annuity, credit  
these are the most important  
variables which contributing  
more or are the decision  
maker for a loan defaulter

# Conclusion

---

Comparing Previous Application and Current Application data this can be concluded

- That Credit Amount and Amount of application is highly connected with Goods Price
- From Co-relation –It can be inferred that Income, Annuity & Credit loan are great influencers.

Salary, Credit, Occupation type , Education status these are all the driving factors which can help in Taking loan