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.

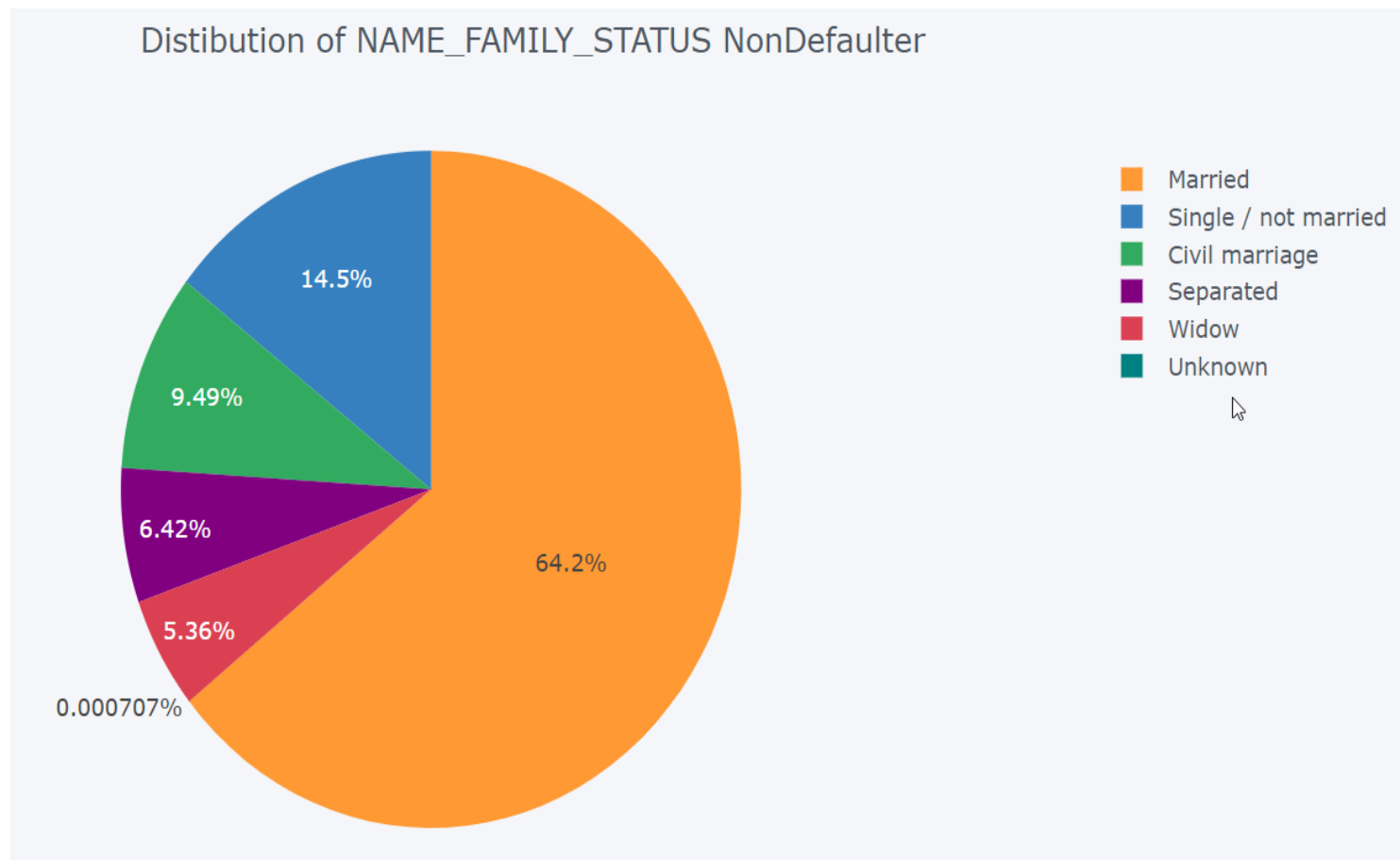So in more details steps are below-

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

Distibution of NAME_FAMILY_STATUS NonDefaulter

- Married
- Single / not married
- Civil marriage
- Separated
- Widow
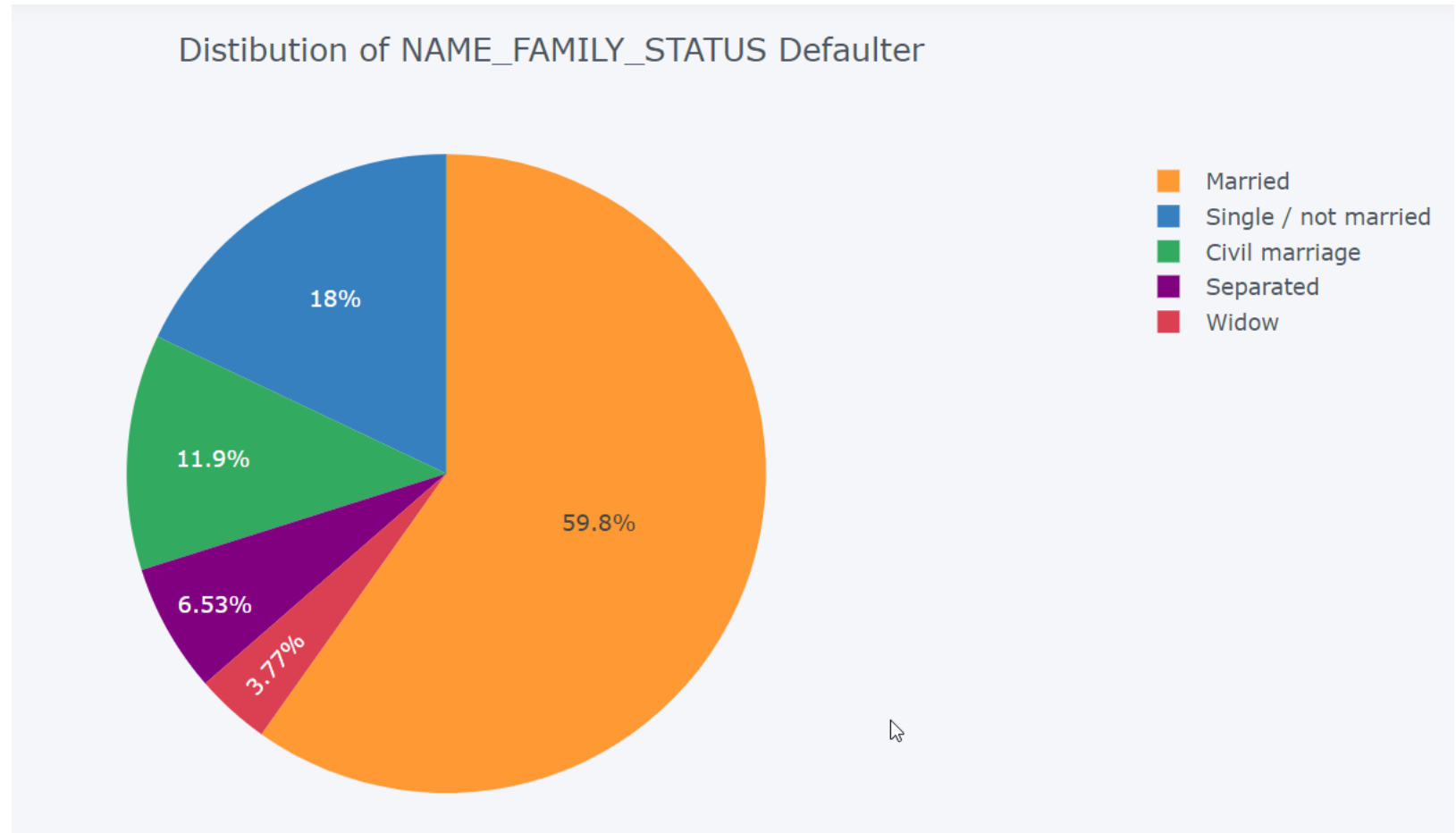- Unknown

14.5%
9.49%
6.42%
5.36%
0.000707%
64.2%

Family Status Distribution against Non-Defaulter

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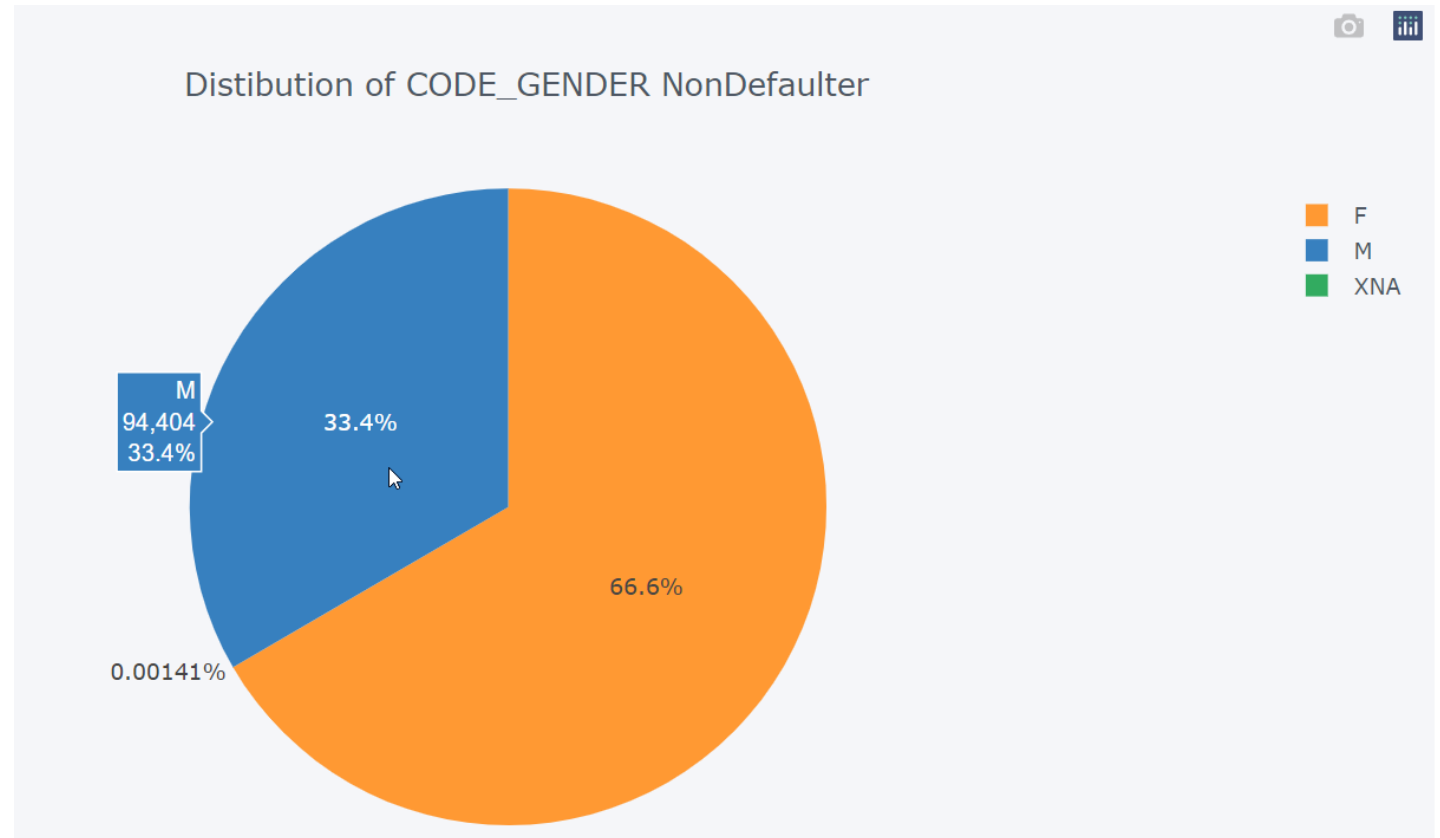
Comparison:-

So looks like they are a decision maker for taking loans

Distibution of NAME_FAMILY_STATUS Defaulter

- Married
- Single / not married
- Civil marriage
- Separated
- Widow

18%
11.9%
59.8%
6.53%
3.77%

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
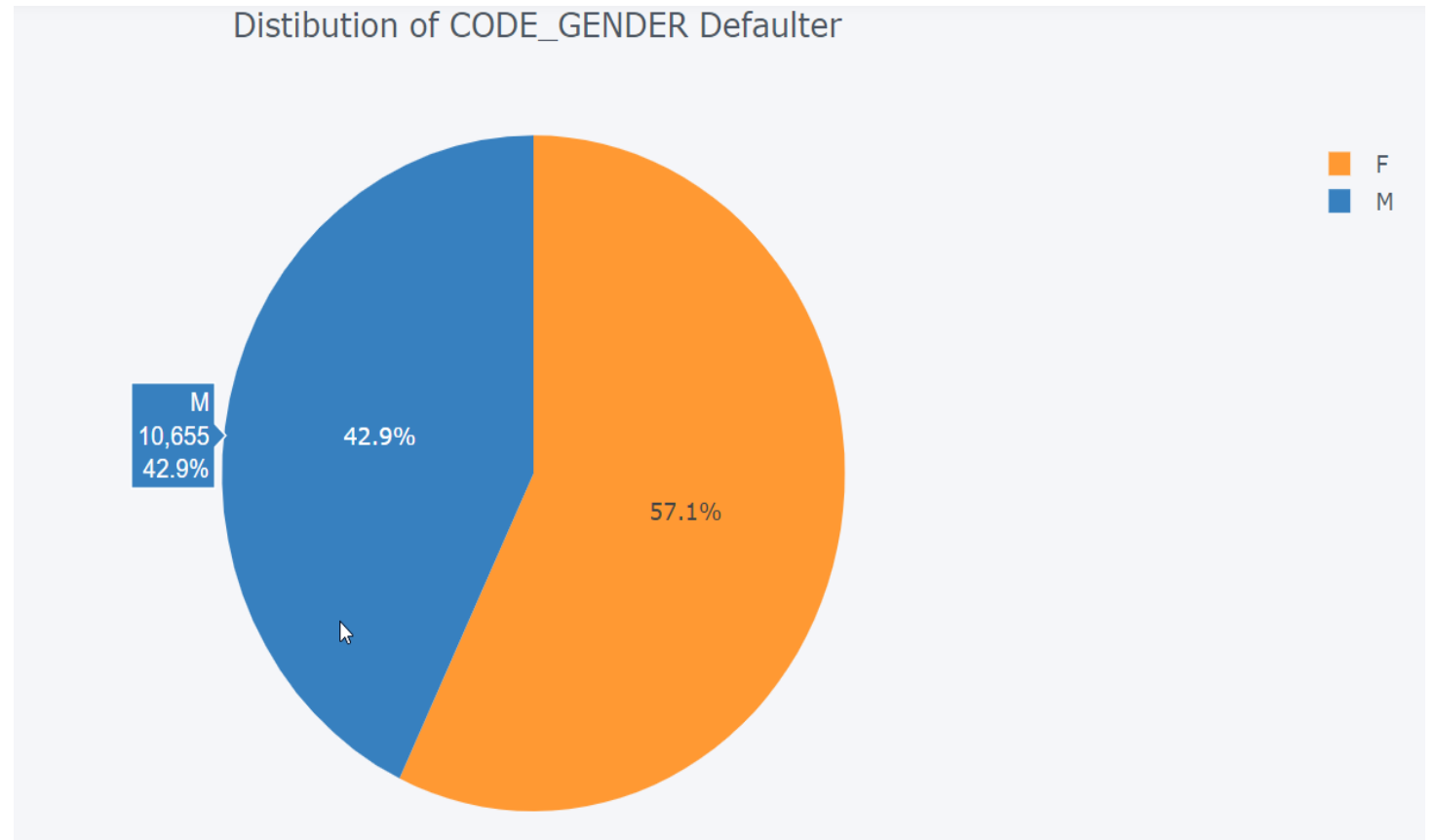

Distibution of CODE_GENDER NonDefaulter

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 Defaulters.
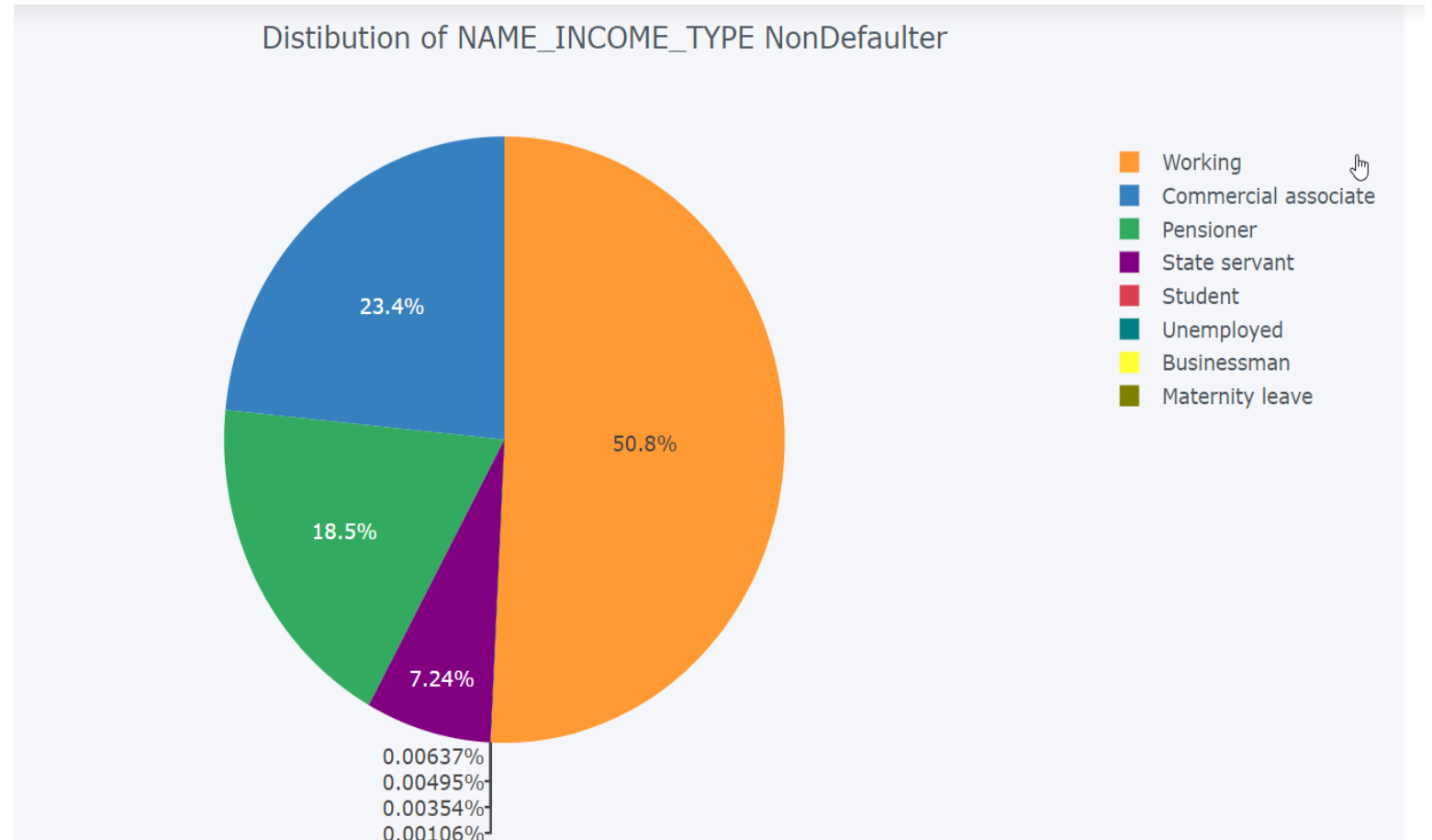
Comparison:-

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

Distibution of CODE_GENDER Defaulter

F
M

M
10,655
42.9%

42.9%

57.1%

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

Distibution of NAME_INCOME_TYPE NonDefaulter

- Working
- Commercial associate
- Pensioner
- State servant
- Student
- Unemployed
- Businessman
- Maternity leave

50.8%
23.4%
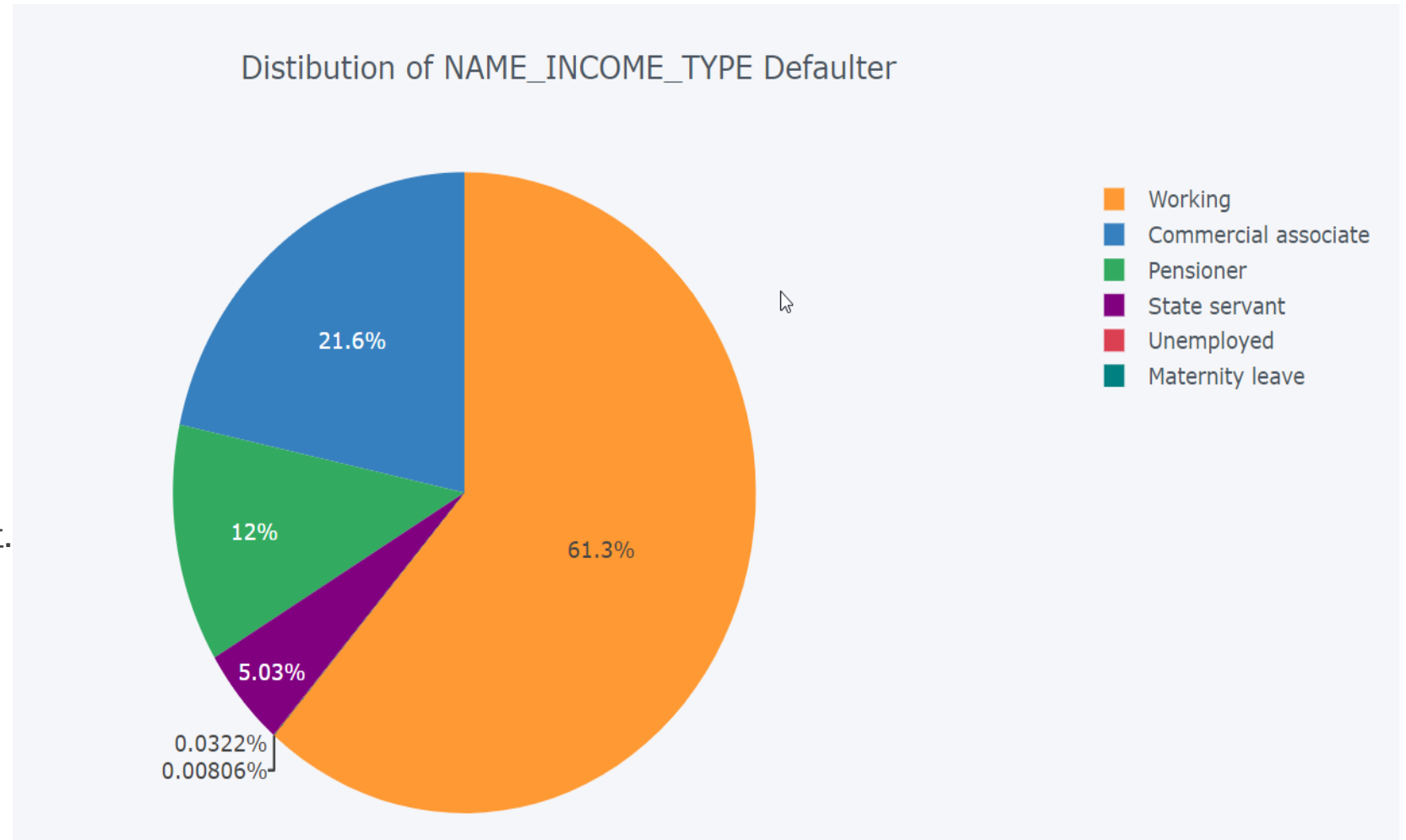18.5%
7.24%
0.00637%
0.00495%
0.00354%
0.00106%

Income Type Distribution against Non-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.
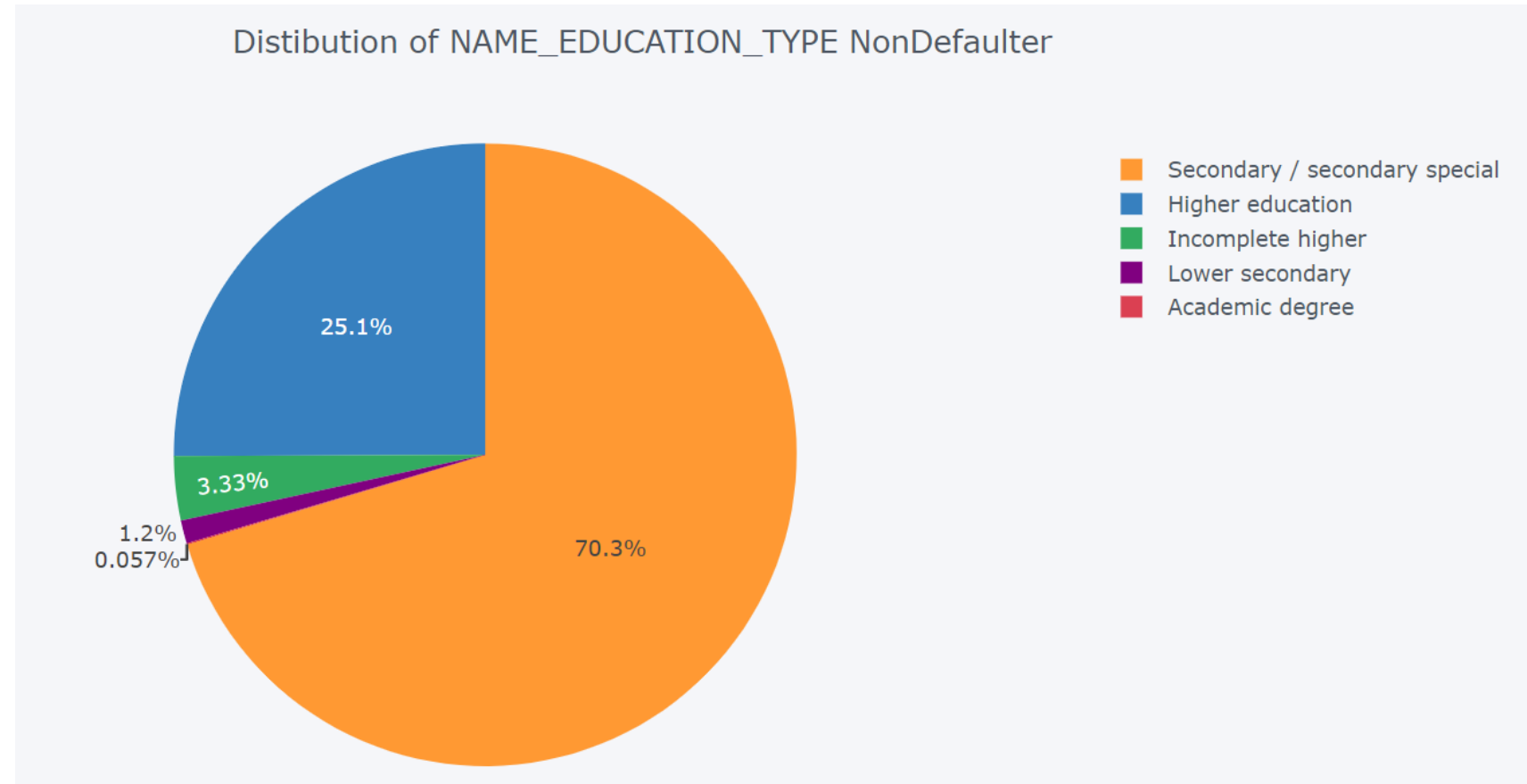
Comparison:-

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

Distibution of NAME_INCOME_TYPE Defaulter

- Working
- Commercial associate
- Pensioner
- State servant
- Unemployed
- Maternity leave

21.6%
12%
61.3%
5.03%
0.0322%
0.00806%

Name Education Type Distribution against Non-Defaulter

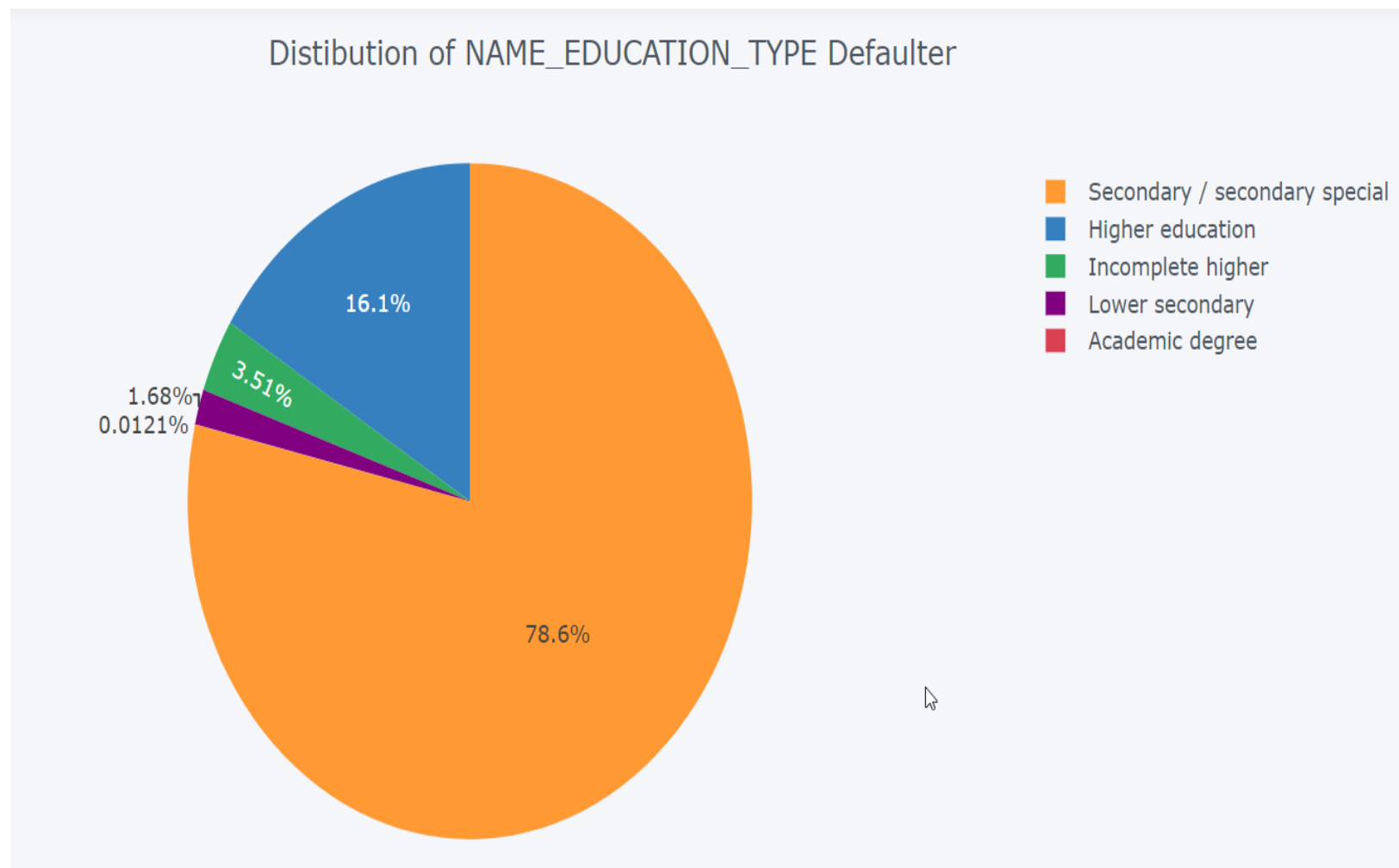By seeing the chart it is clear that more Secondary Education Holder are the non Defaulter

Distibution of NAME_EDUCATION_TYPE NonDefaulter

- Secondary / secondary special
- Higher education
- Incomplete higher
- Lower secondary
- Academic degree

25.1%

3.33%

1.2%

0.057%

70.3%

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

Distibution of NAME_EDUCATION_TYPE Defaulter

- Secondary / secondary special
- Higher education
- Incomplete higher
- Lower secondary
- Academic degree

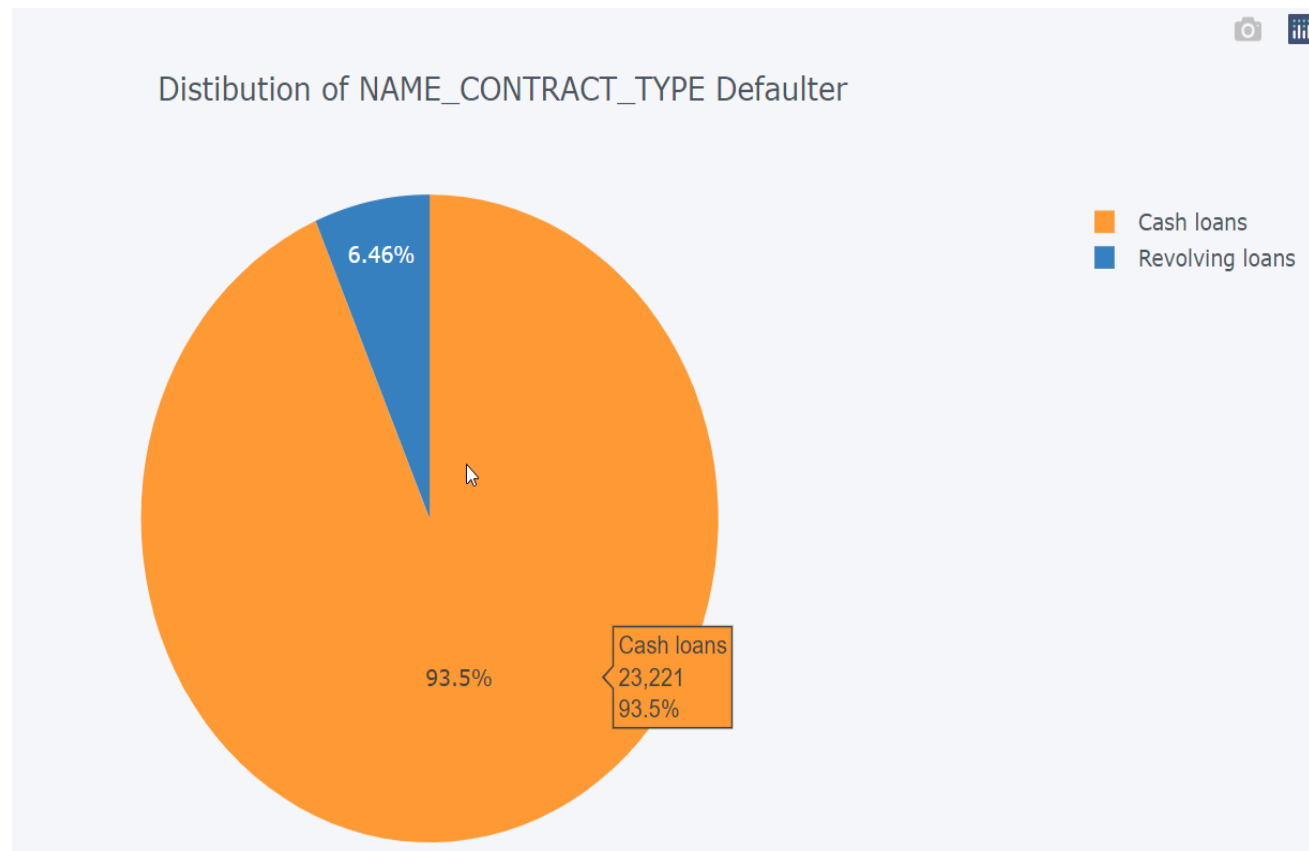16.1%

1.68%
0.0121%

3.51%

78.6%

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.


Distibution of NAME_CONTRACT_TYPE NonDefaulter
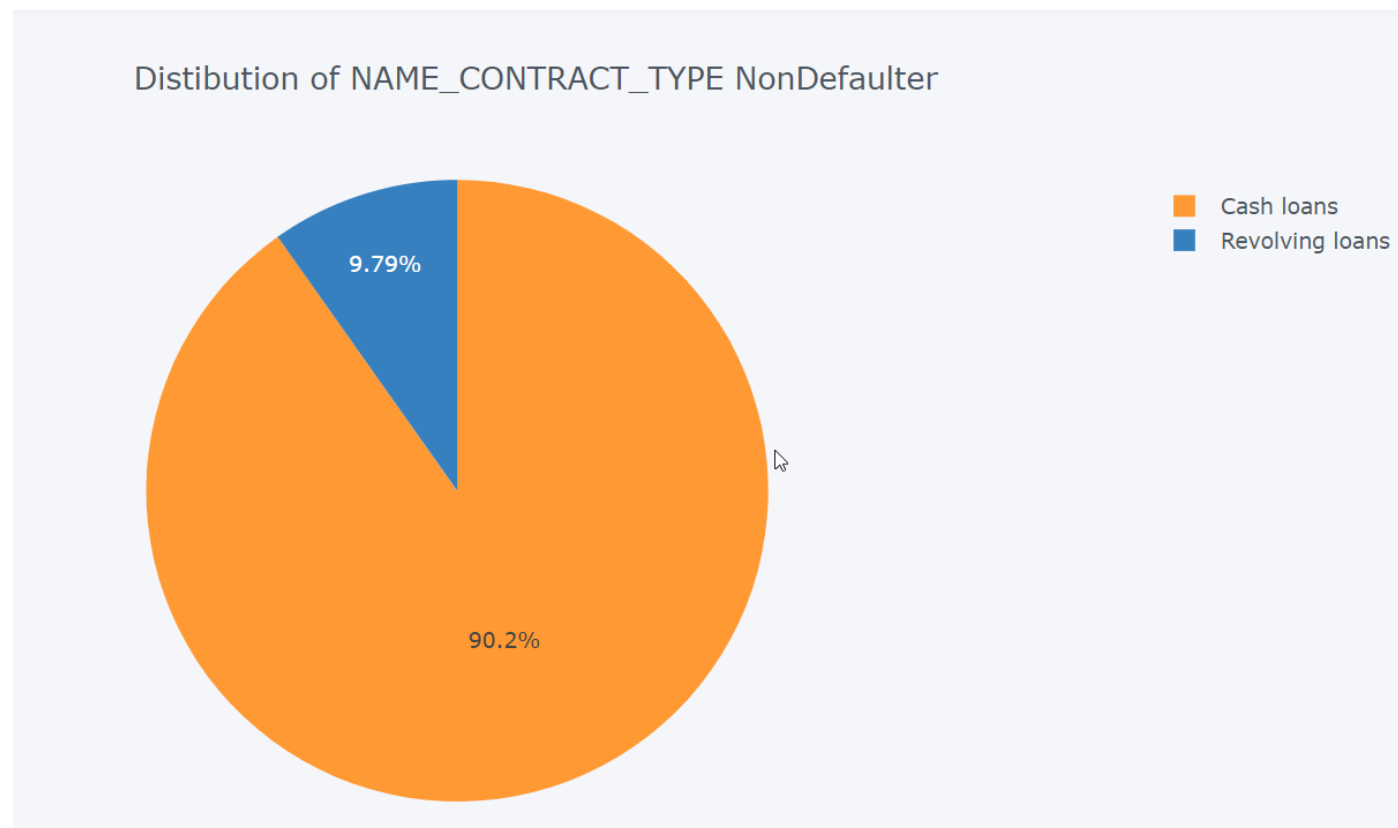
Loans Contract Distribution against Non-Defaulter

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

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'

Distibution of NAME_CONTRACT_TYPE Defaulter

6.46%

93.5%

Cash loans
23,221
93.5%

Cash loans
Revolving 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
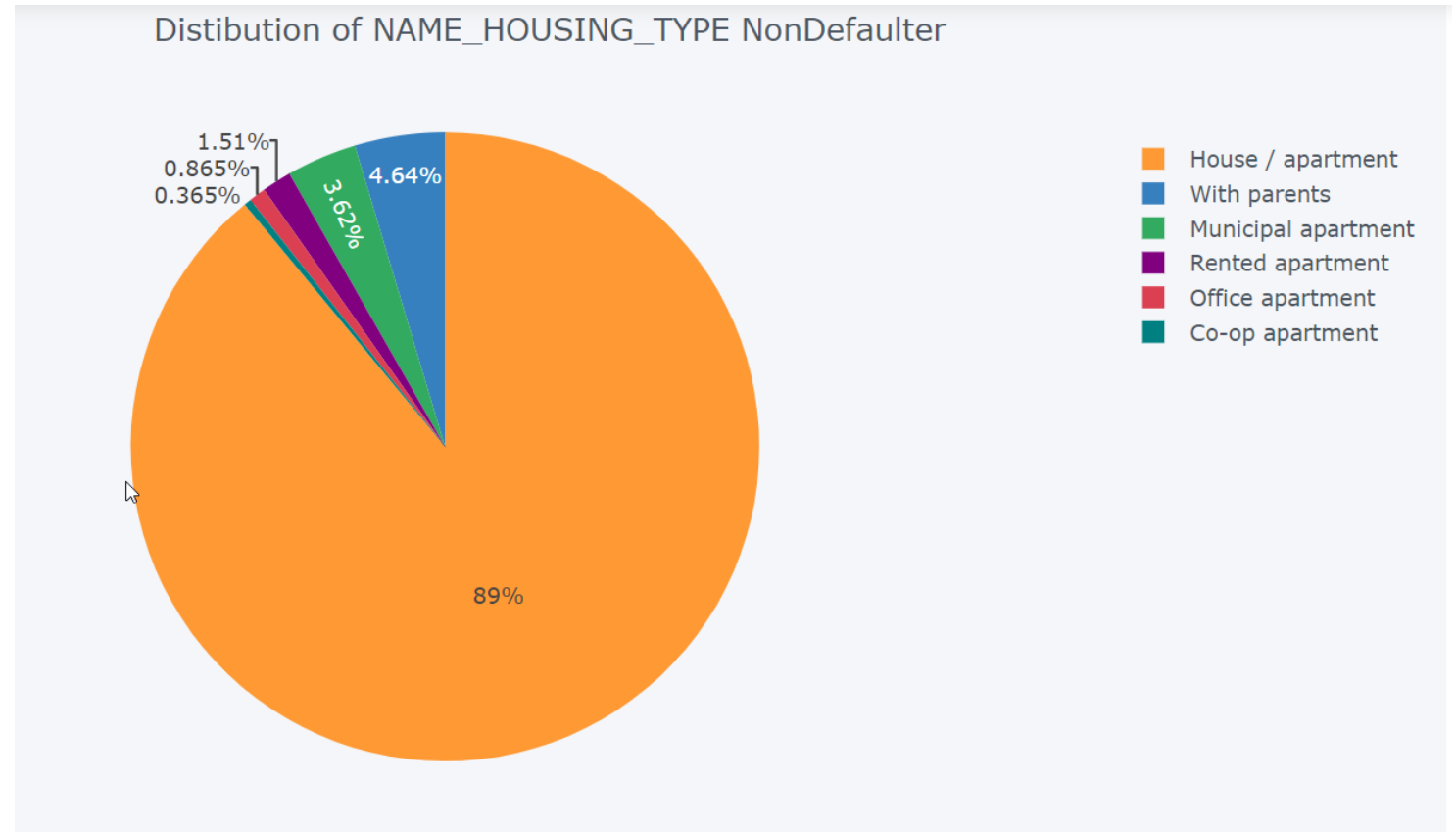
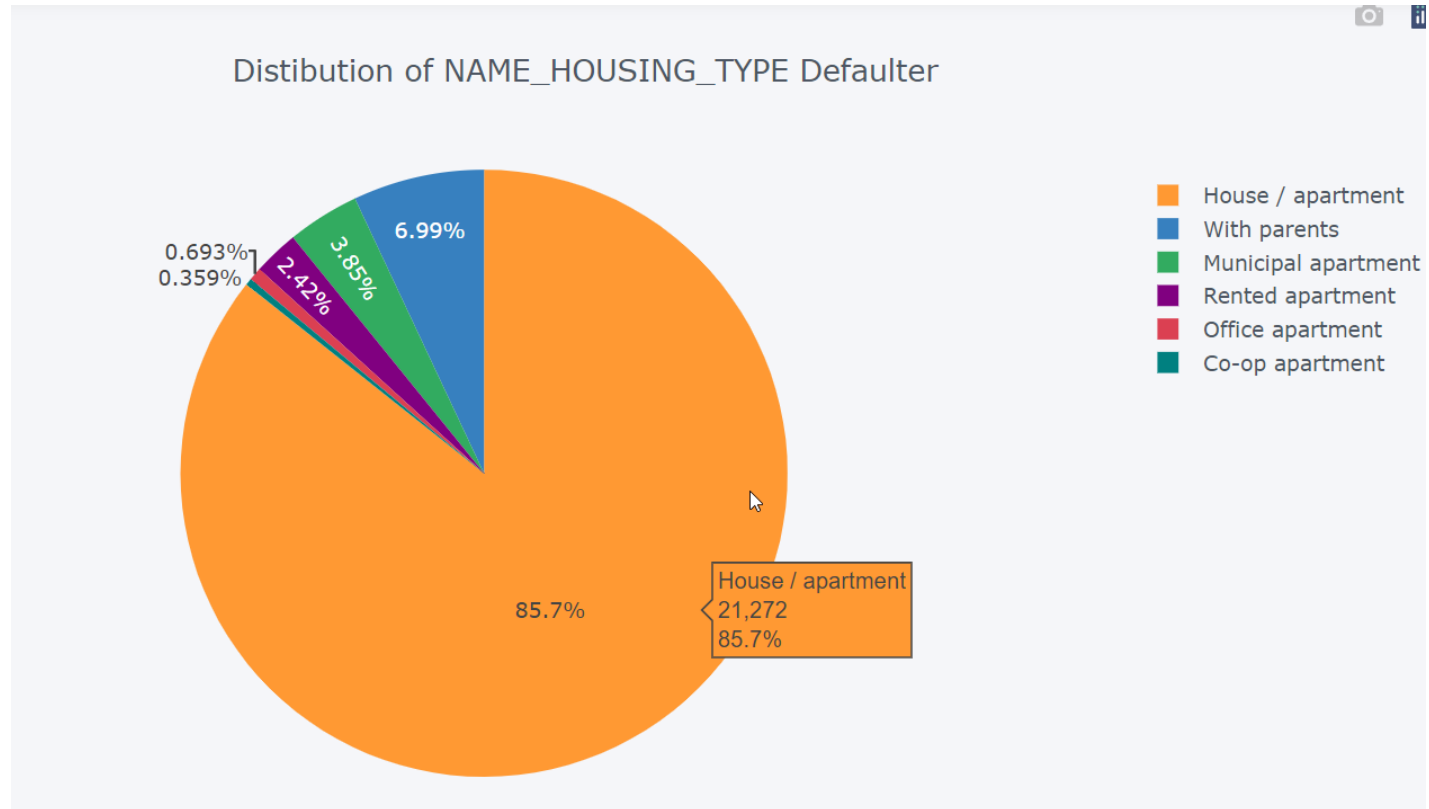Distibution of NAME_HOUSING_TYPE NonDefaulter

Housing Type Distribution against Non-Defaulter

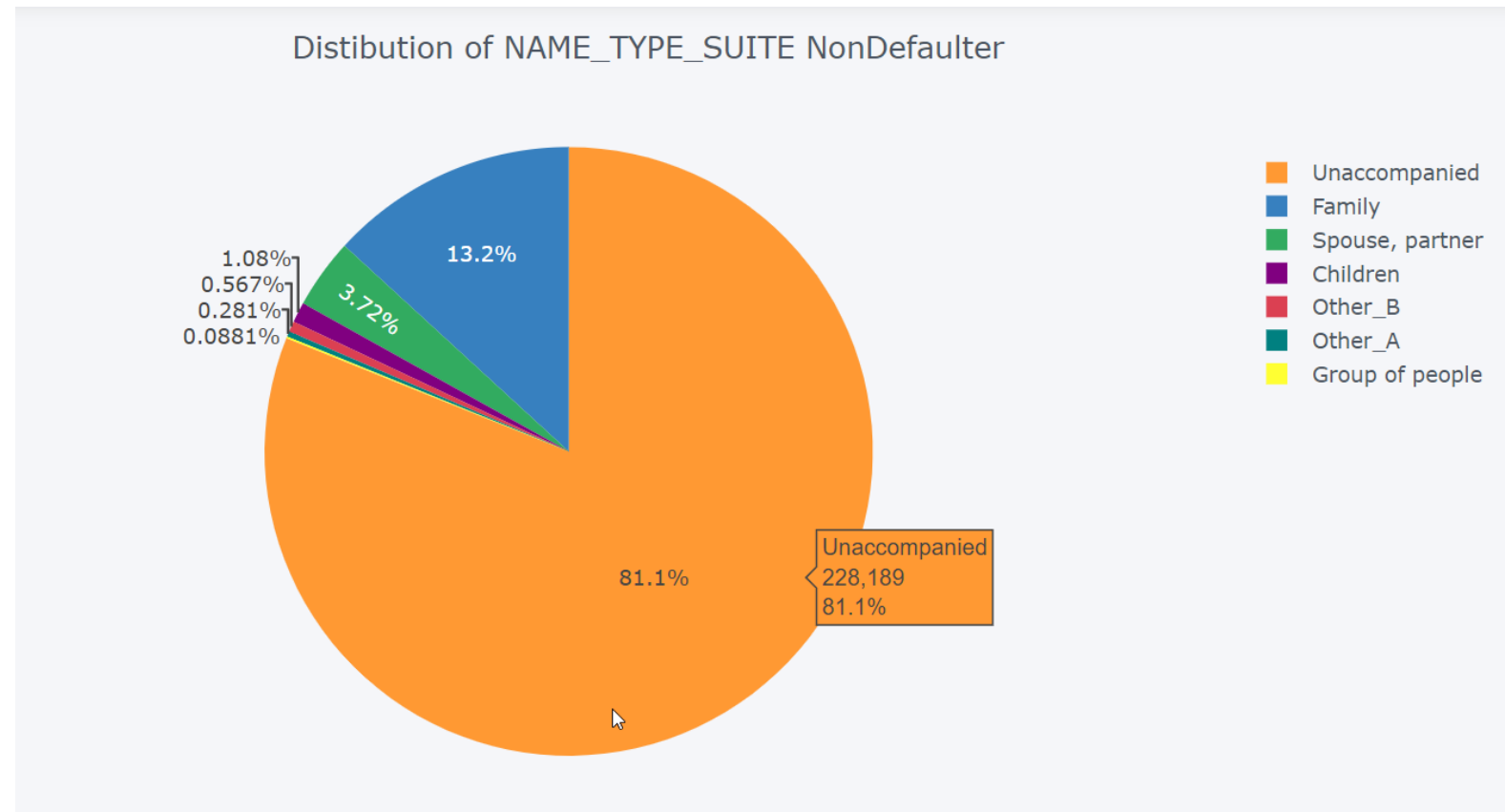By seeing the chart it is clear that people who has their own house or apartment already taking more loans

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

Distibution of NAME_HOUSING_TYPE Defaulter

- House / apartment
- With parents
- Municipal apartment
- Rented apartment
- Office apartment
- Co-op apartment

6.99%
0.693%
0.359%
2.42%
3.85%
85.7%

House / apartment
21,272
85.7%

Type Suite Distribution against Non-Defaulter

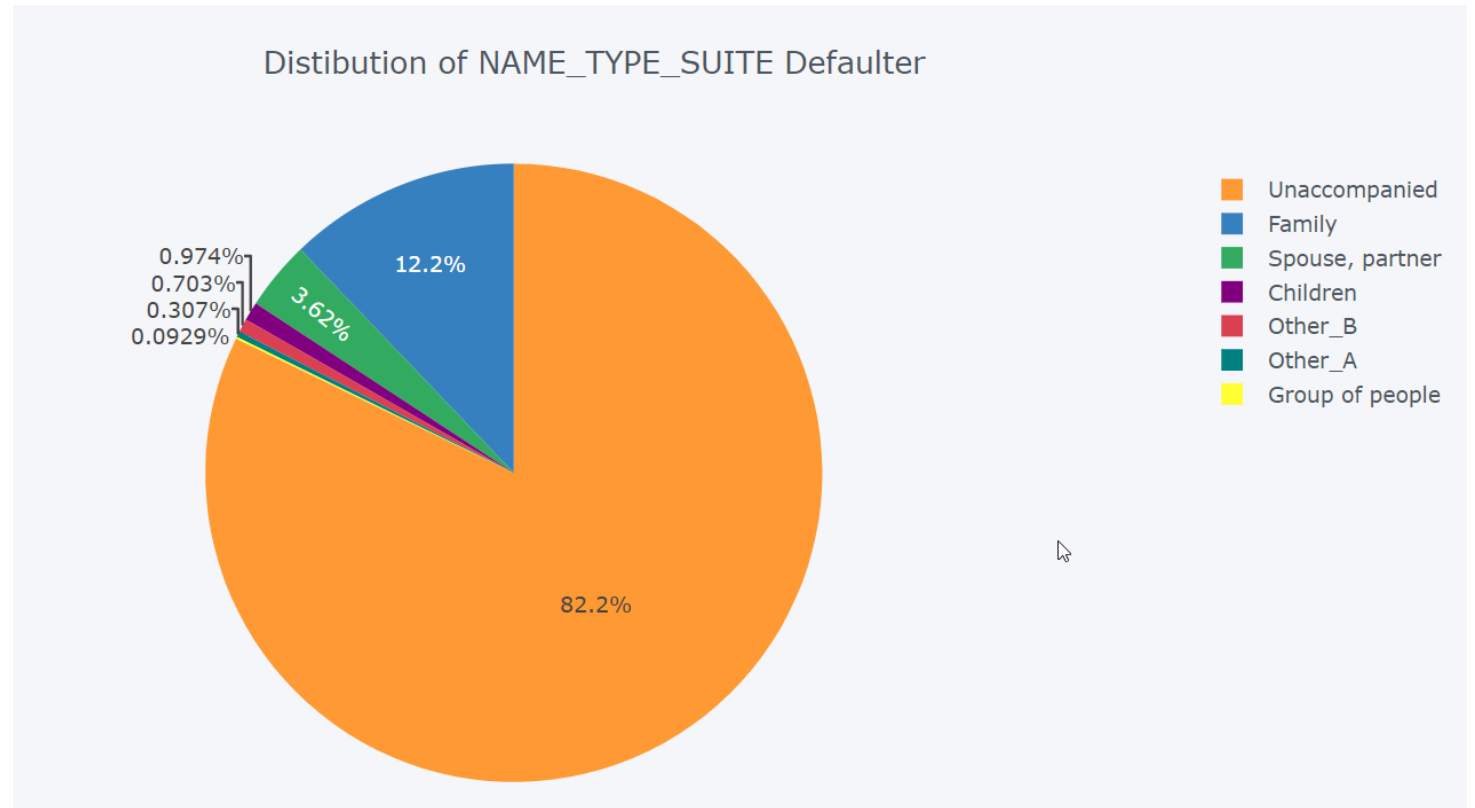By seeing the chart it is clear that more Unaccompanied people are paying the loan amount in proper time.

Distibution of NAME_TYPE_SUITE NonDefaulter

1.08%
0.567%
0.281%
0.0881%

13.2%

3.72%

81.1%

Unaccompanied
228,189
81.1%

- Unaccompanied
- Family
- Spouse, partner
- Children
- Other_B
- Other_A
- Group of people

Type Suite Distribution against Non-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

Distibution of NAME_TYPE_SUITE Defaulter

0.974%
0.703%
0.307%
0.0929%

12.2%

3.62%

82.2%

Unaccompanied
Family
Spouse, partner
Children
Other_B
Other_A
Group of people

Bivariate Analysis of Current Application

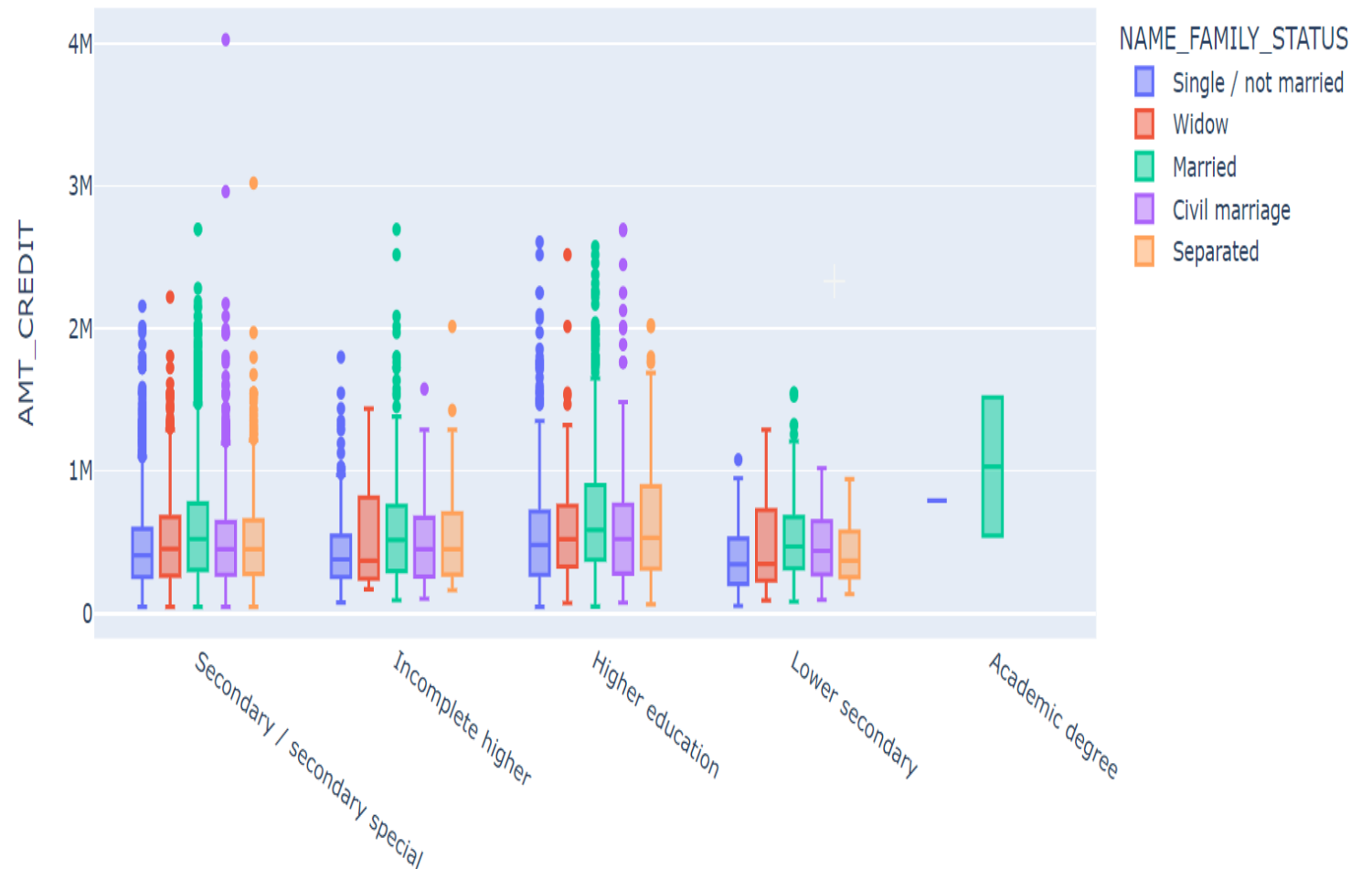# Education Type Vs Credit Distribution against Non-Defaulter

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.

NAME_EDUCATION_TYPE vs AMT_CREDIT in respect to NonDefaulter

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 Non-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
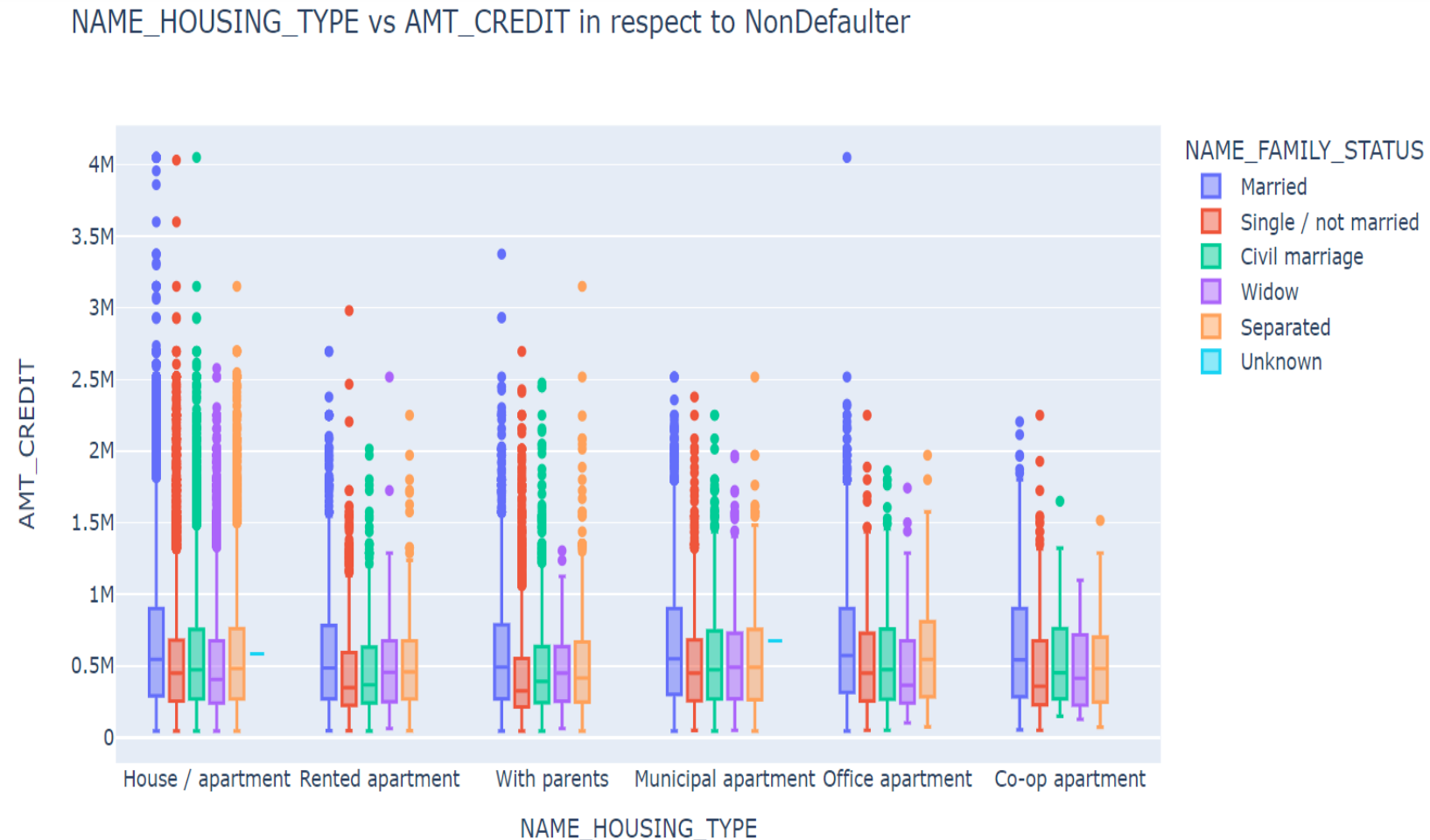
.

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
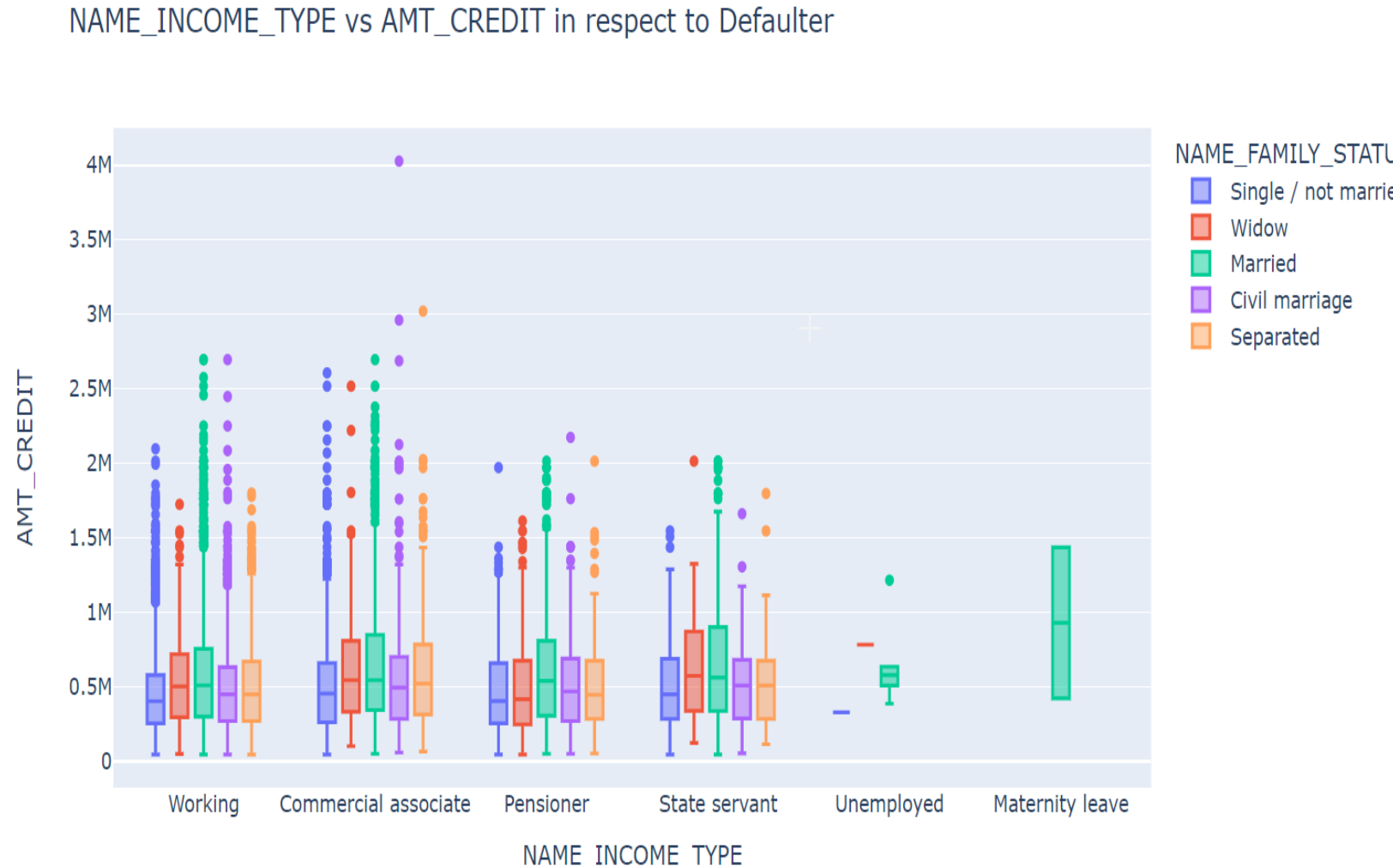
NAME_INCOME_TYPE vs AMT_CREDIT in respect to Defaulter

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.

NAME_HOUSING_TYPE vs AMT_CREDIT in respect to NonDefaulter

Housing Type Vs Credit Distribution against Non-Defaulter

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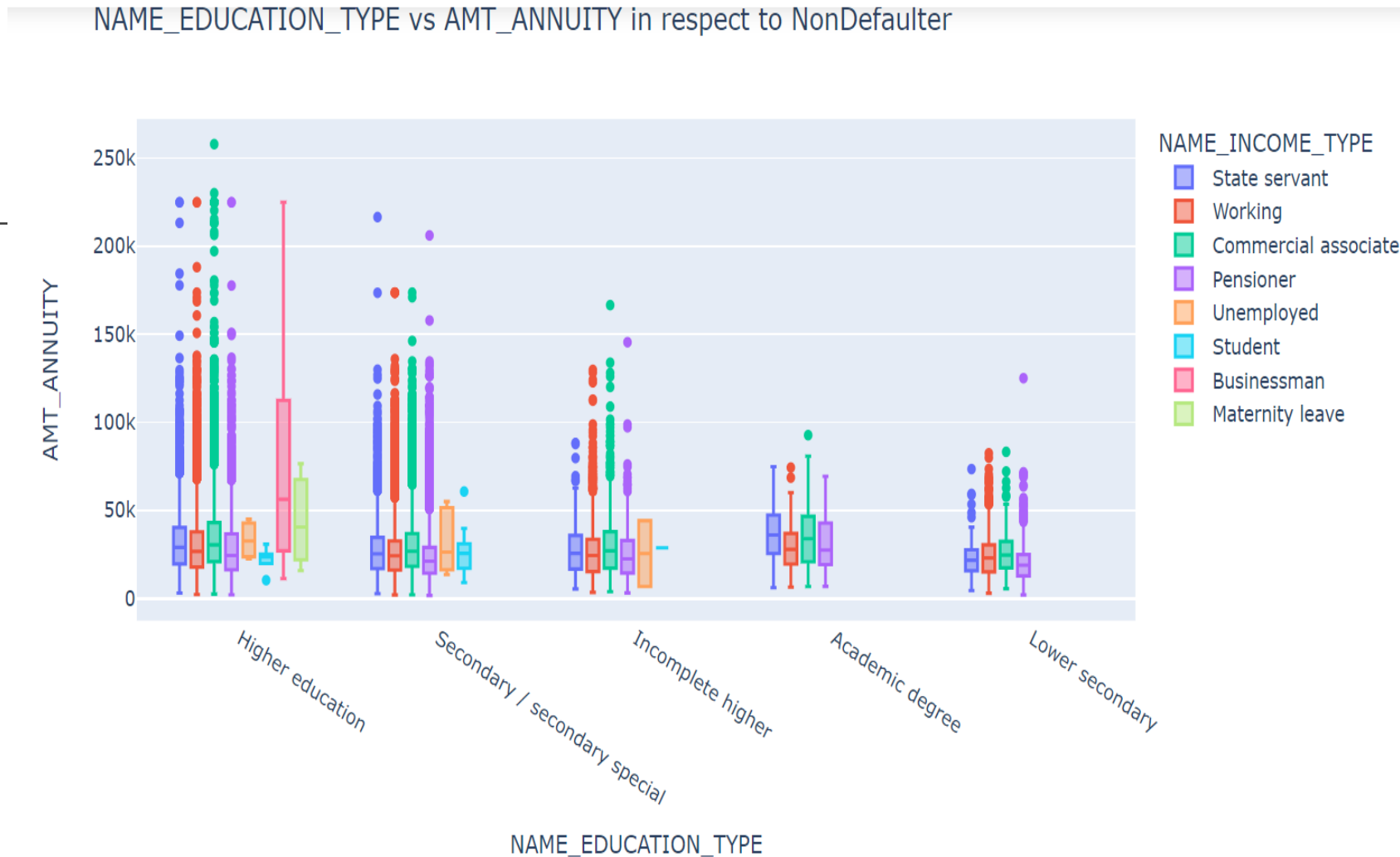
Comparison:-

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
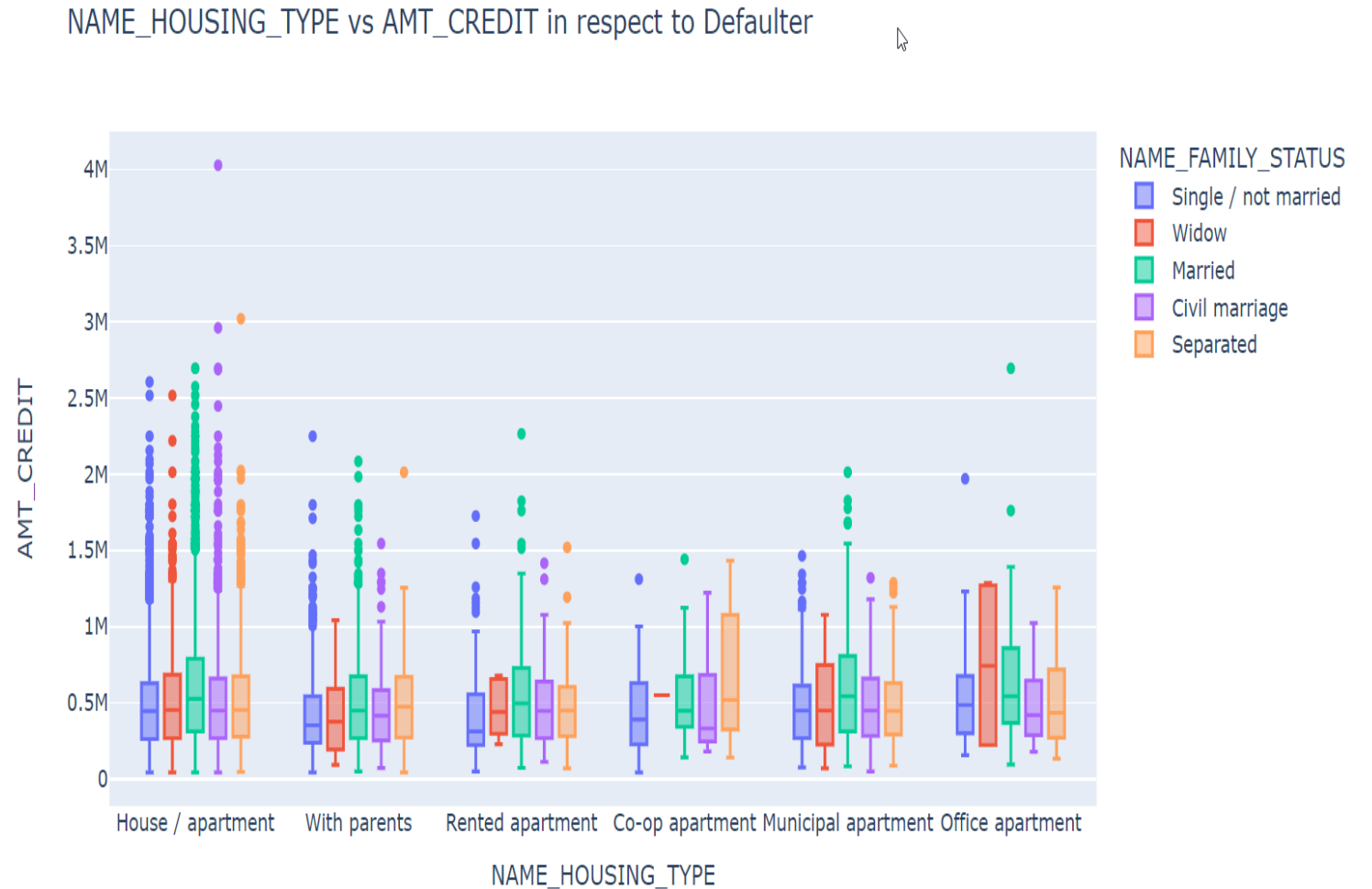
NAME_HOUSING_TYPE vs AMT_CREDIT in respect to Defaulter

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 Non- Defaulter

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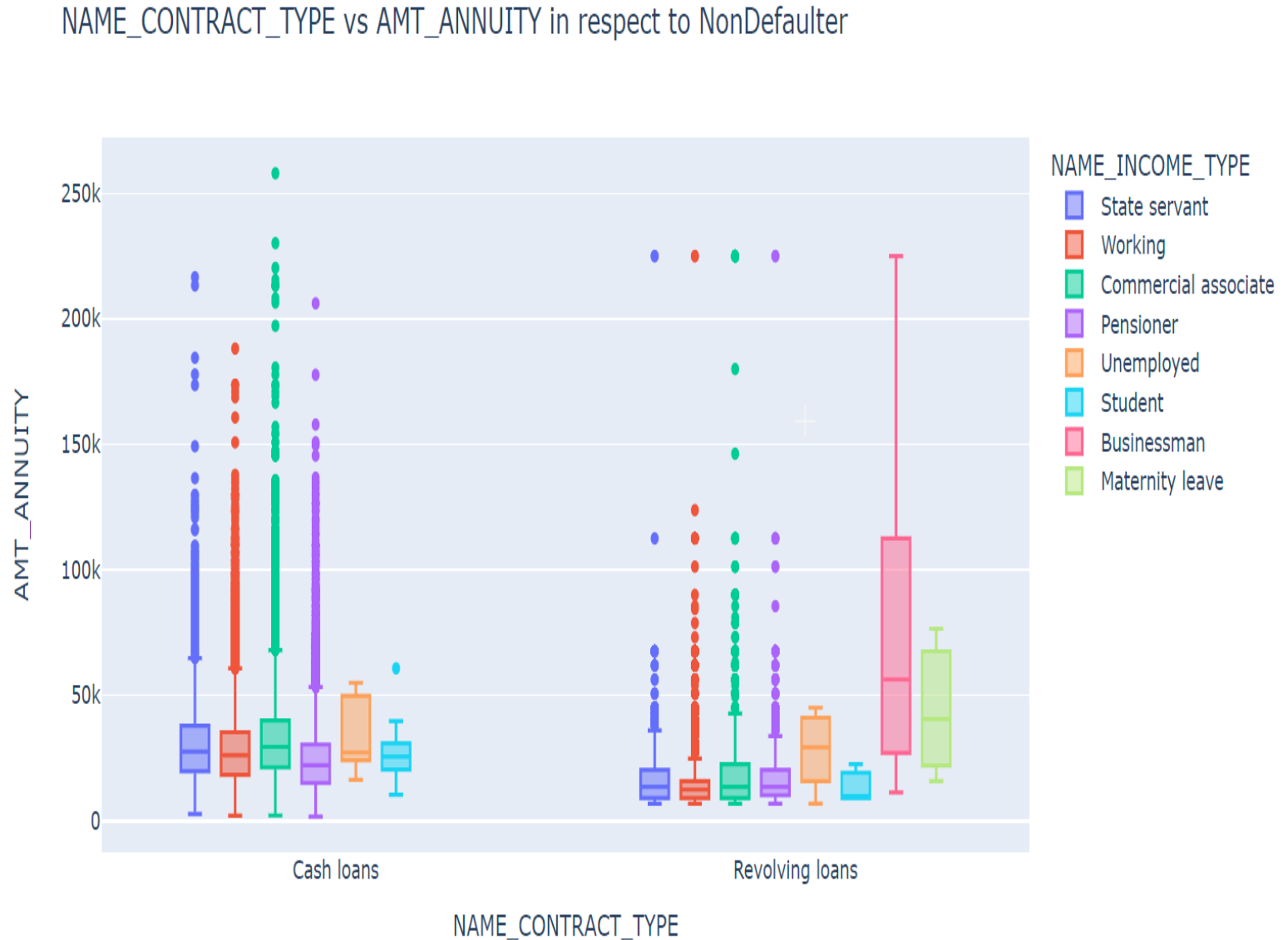
.

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

NAME_EDUCATION_TYPE vs AMT_ANNUITY in respect to Defaulter
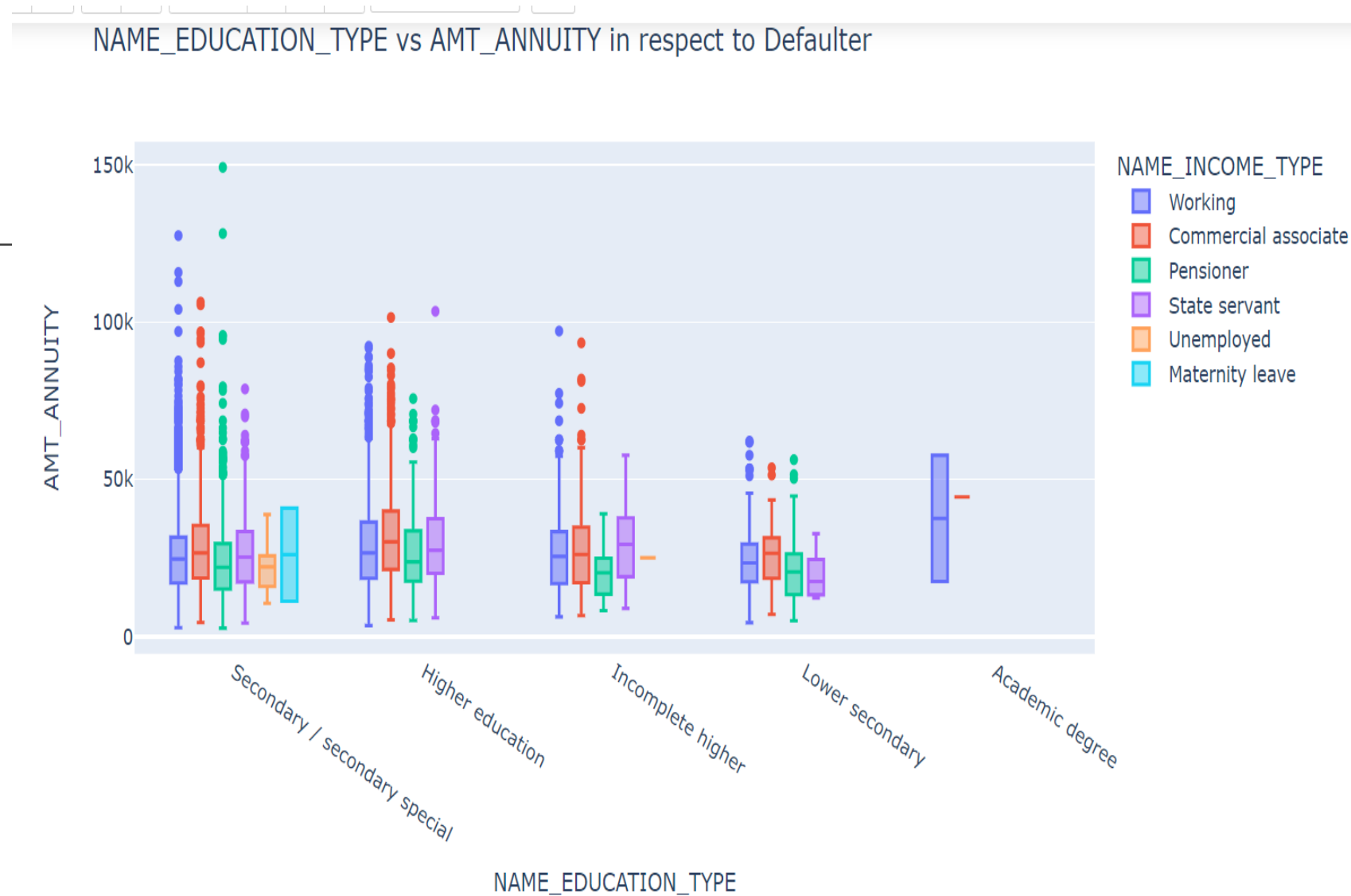
Contract Type Vs Annuity Distribution against Non- Defaulter

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

NAME_CONTRACT_TYPE vs AMT_ANNUITY in respect to NonDefaulter

Income Type Vs Annuity Distribution against Non- 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 difficylties
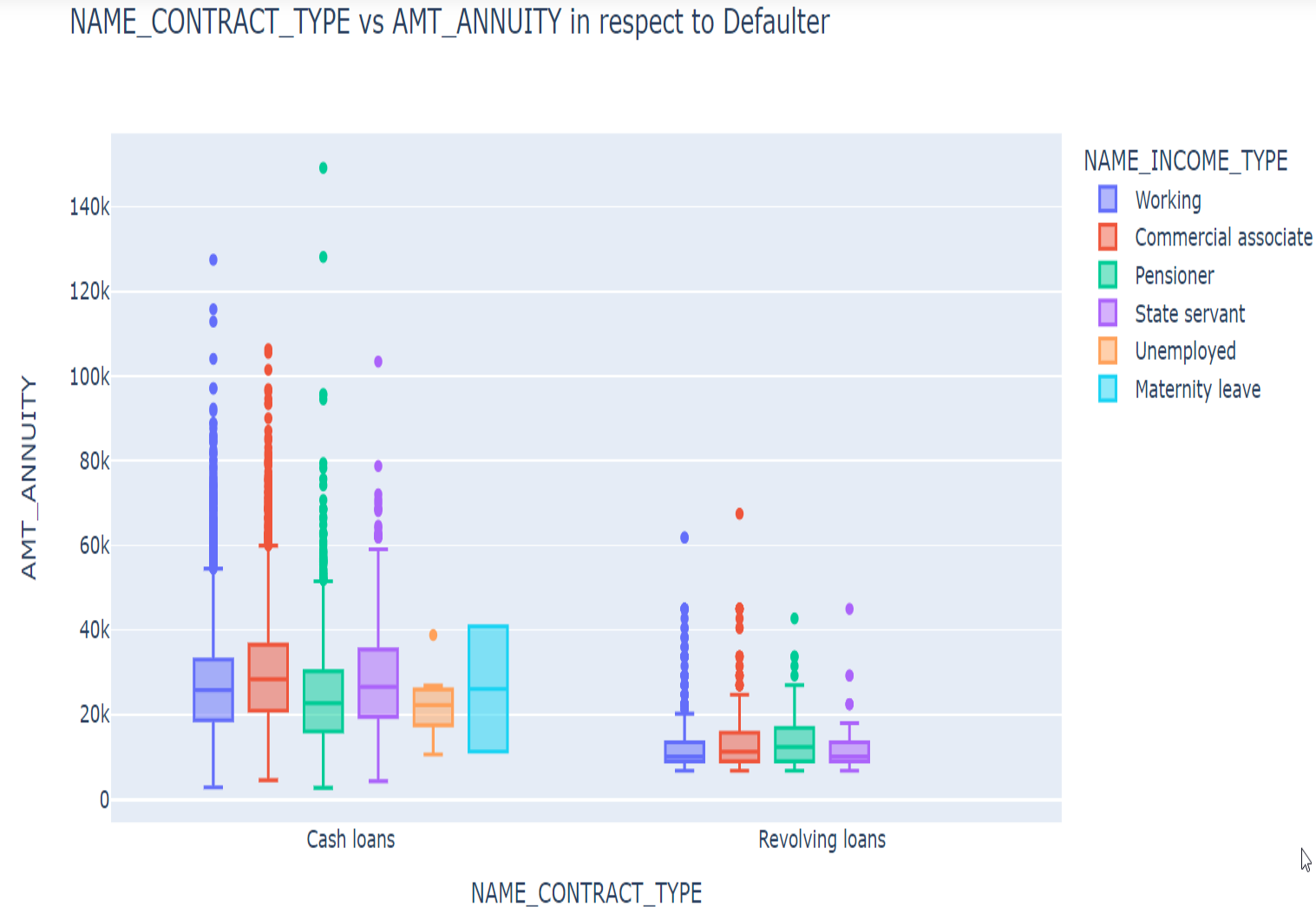
Contract Type Vs Goods Price Distribution against Non- Defaulter

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.


NAME_CONTRACT_TYPE vs AMT_GOODS_PRICE in respect to NonDefaulter
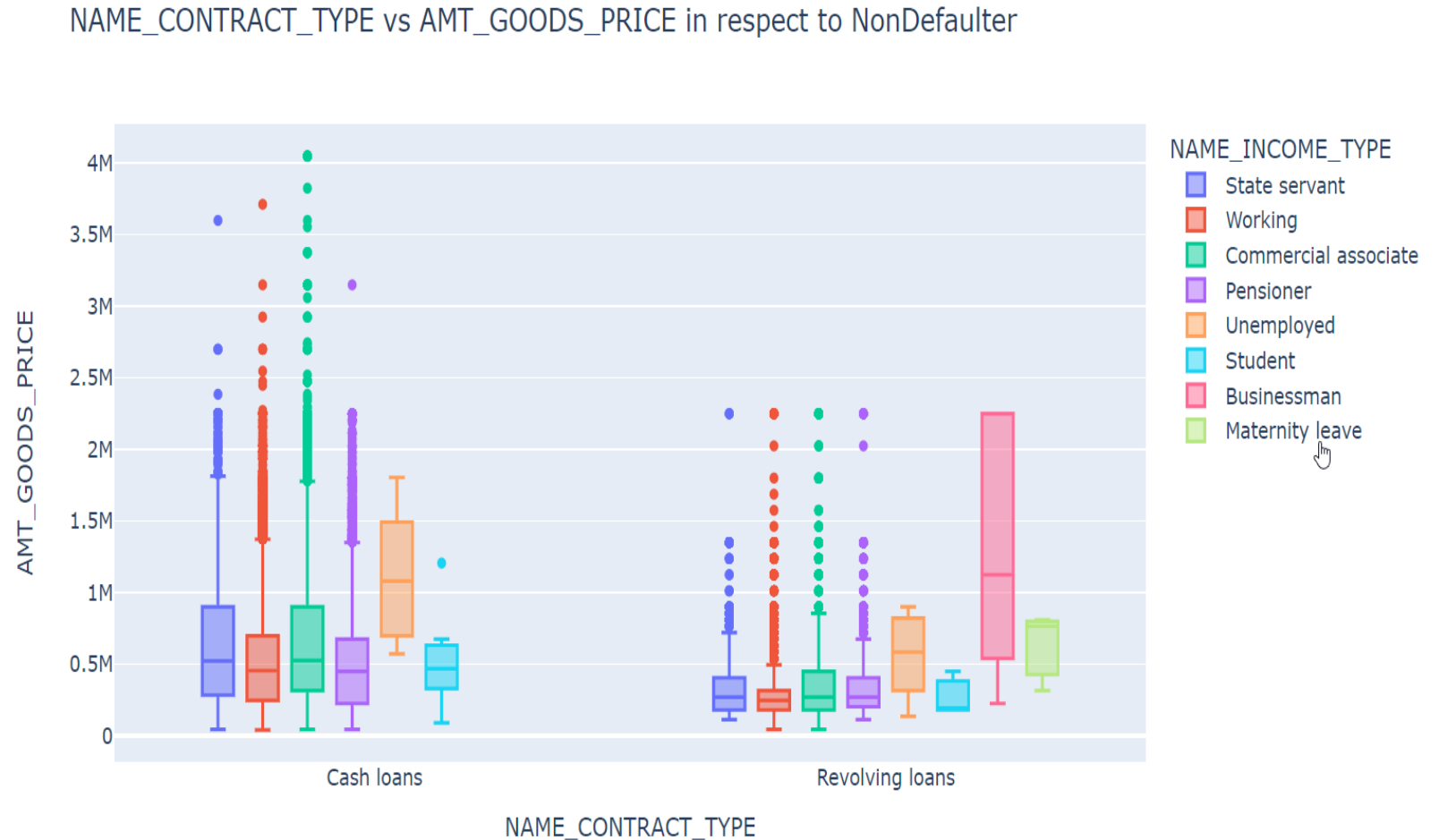
Contract Type Vs Goods Price Distribution against Non- Defaulter

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.

NAME_CONTRACT_TYPE vs AMT_GOODS_PRICE in respect to Defaulter
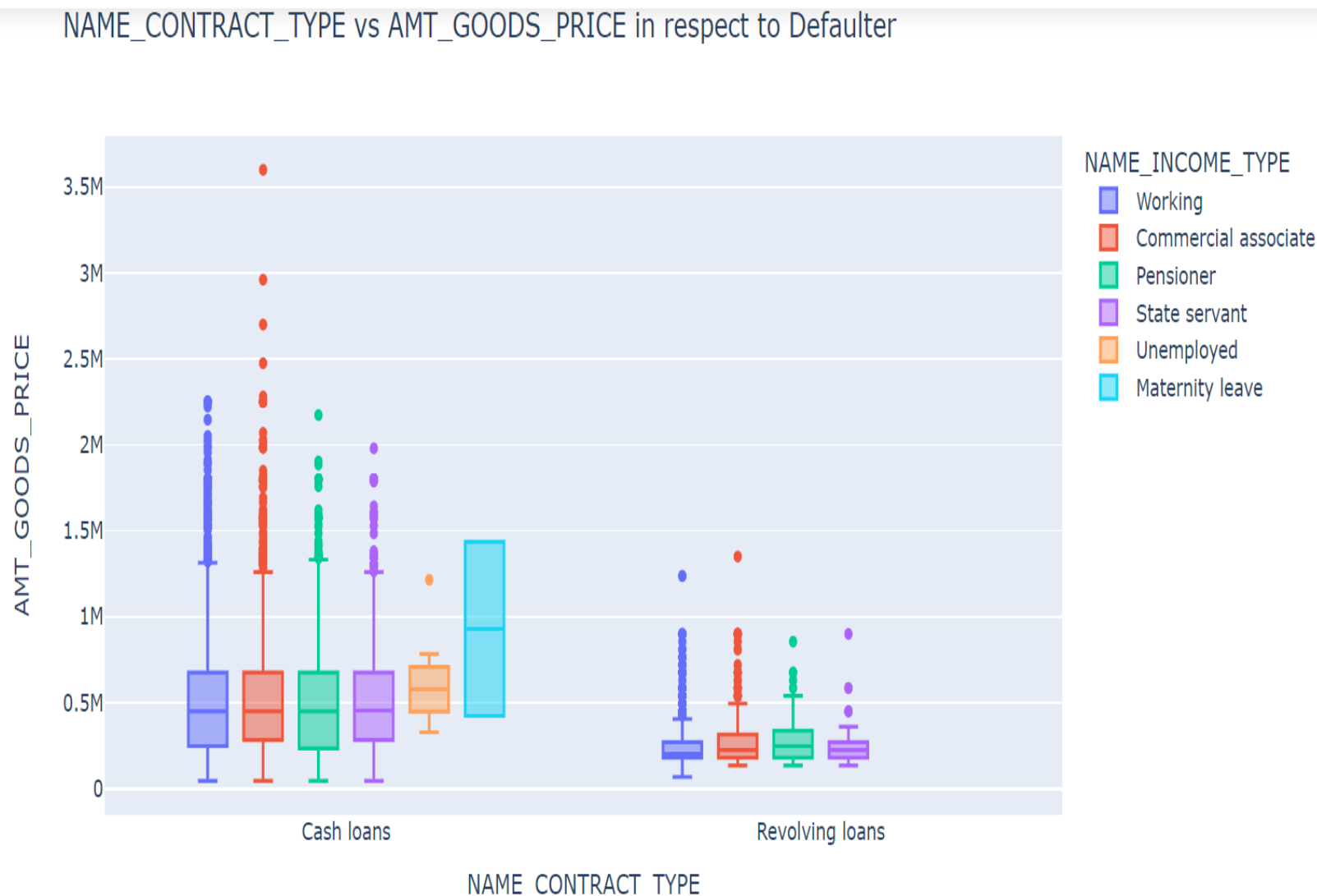
Pair-plotting for Defaulters

From this pair-plot it is very much clear that relationship between Goods Price and Credit very strong.
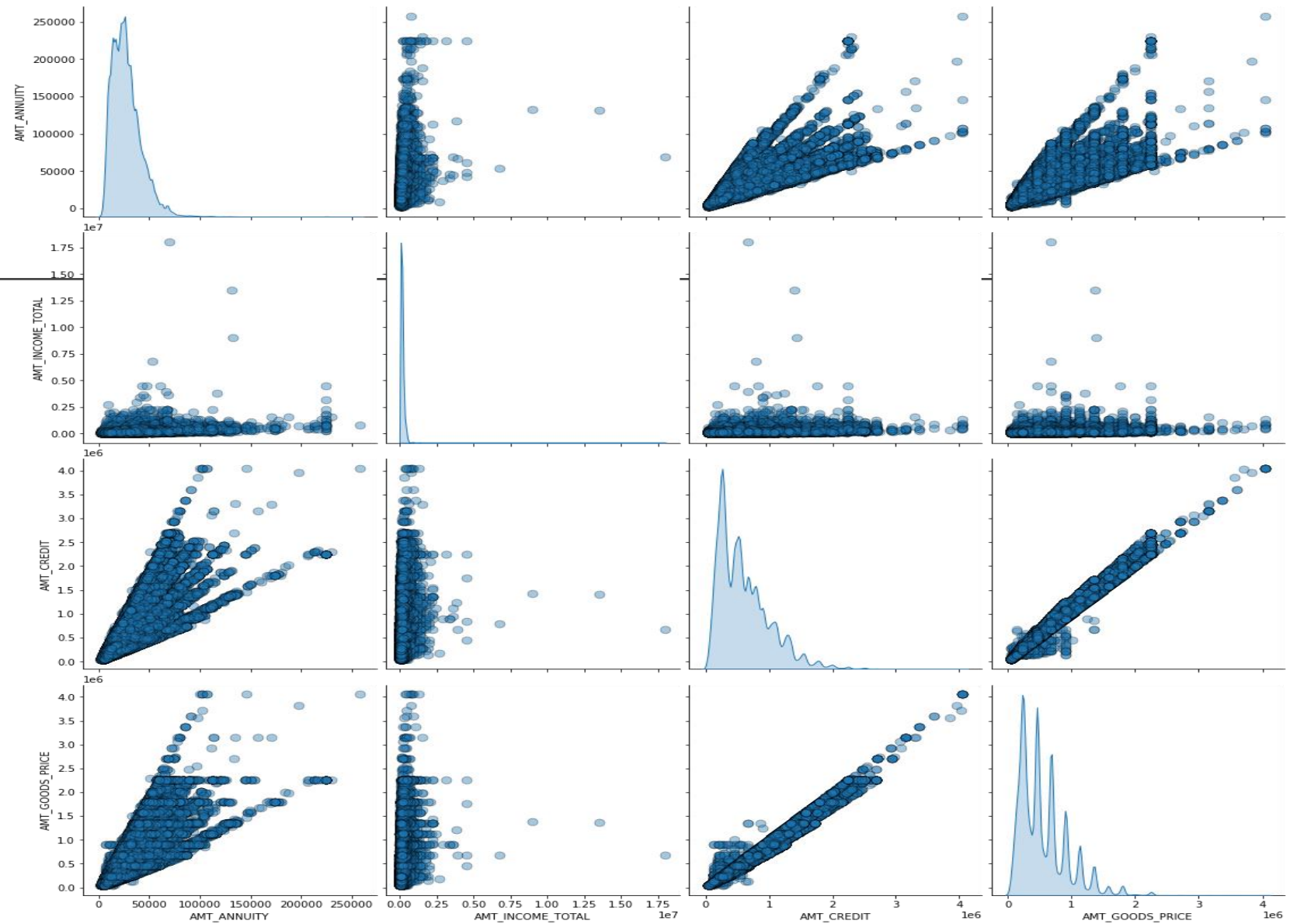
They are highly interlinked to each other.
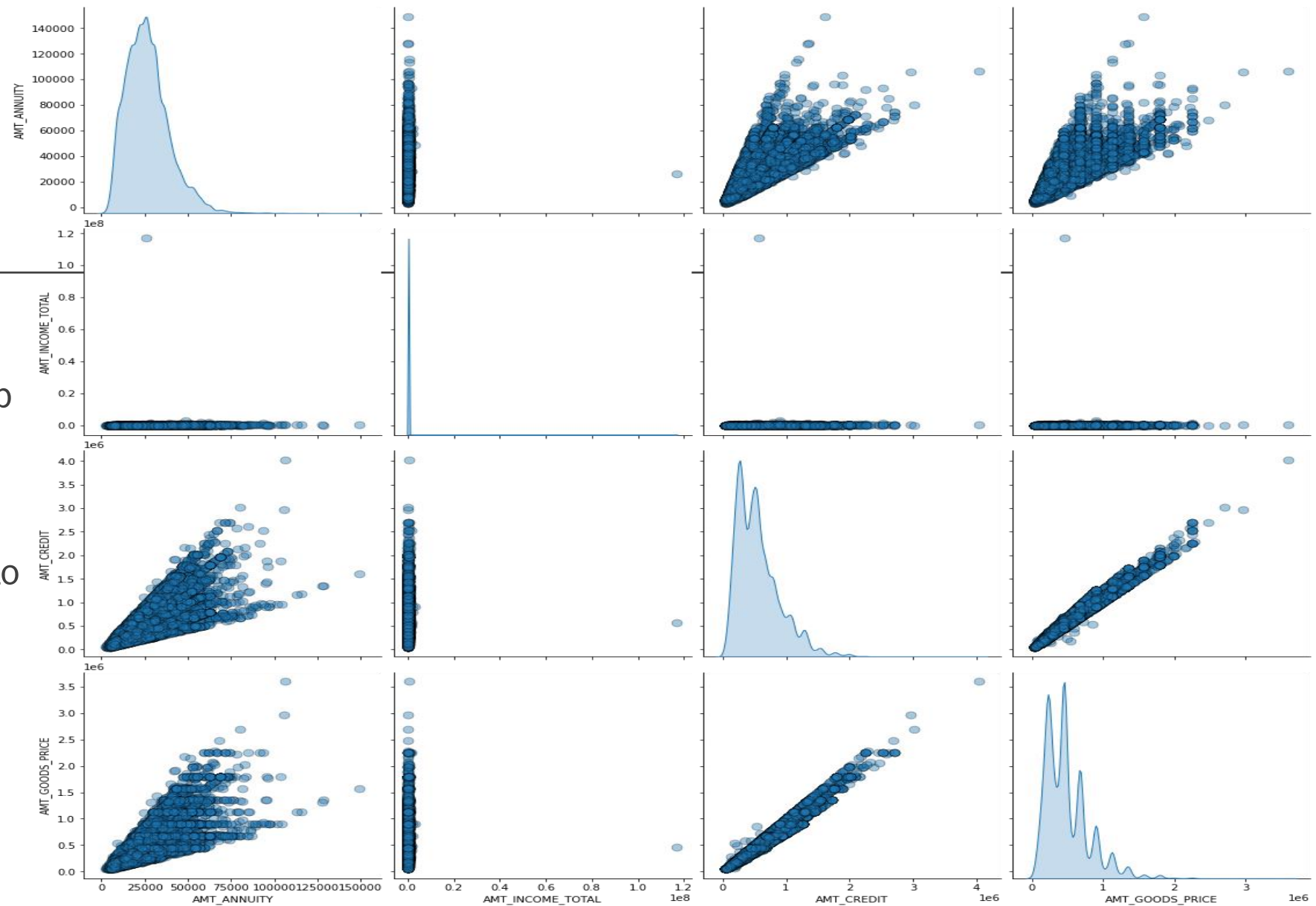
Pair-plotting for Non- Defaulter

Same as Defaulter graph

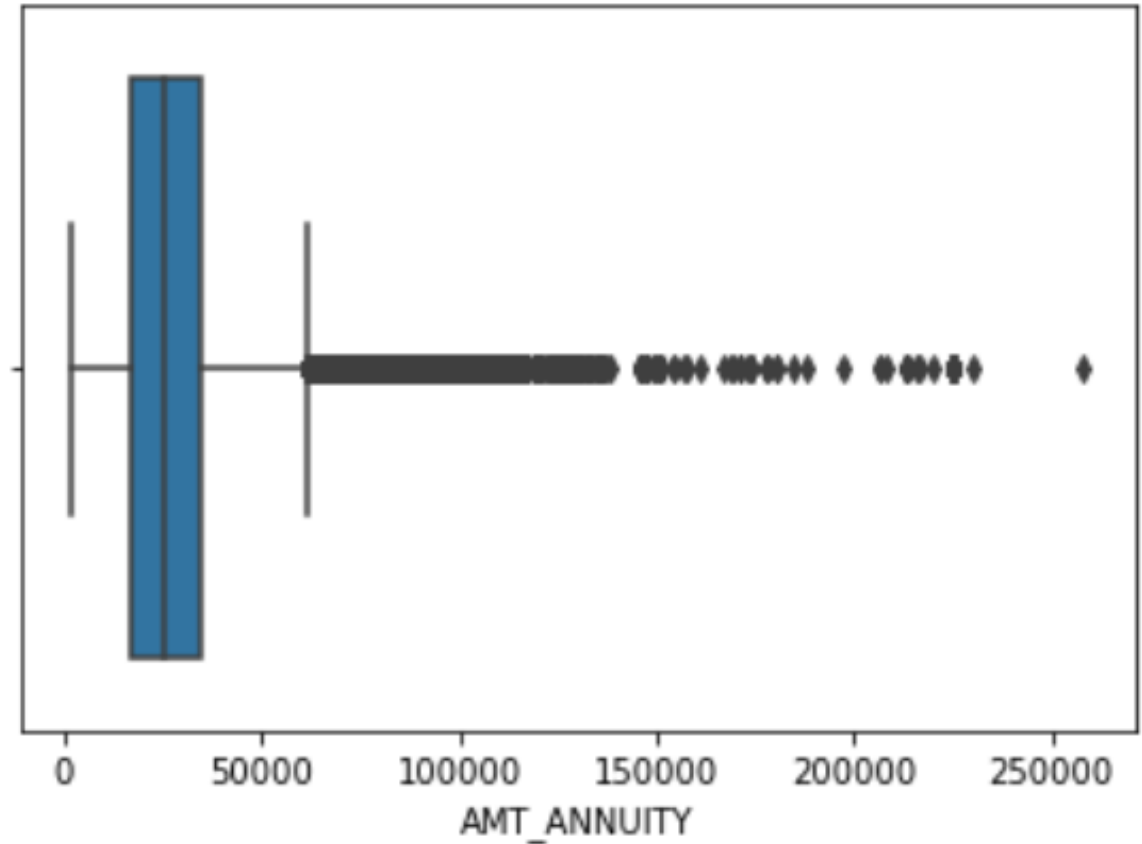Goods Price and Credit are highly interlinked to each other.

Segmented Univariate Analysis

Annuity Amount Boxplot Distribution against Non- Defaulters & Defaulters

There is a high outlier which is more than 250000

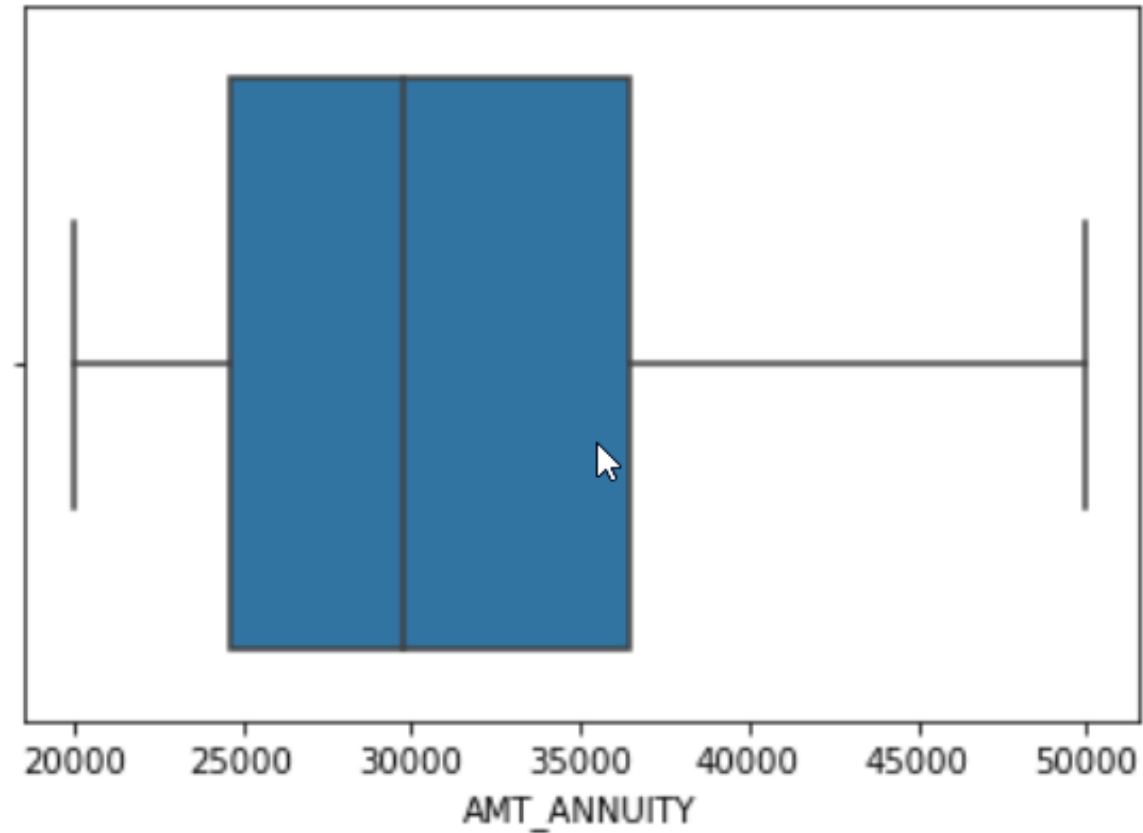And there is more outlier in-between 100000 to 150000

More specifically sub plotting can be visible in next boxplot

AMT_ANNUITY

Income Type Distribution against Non-Defaulter

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

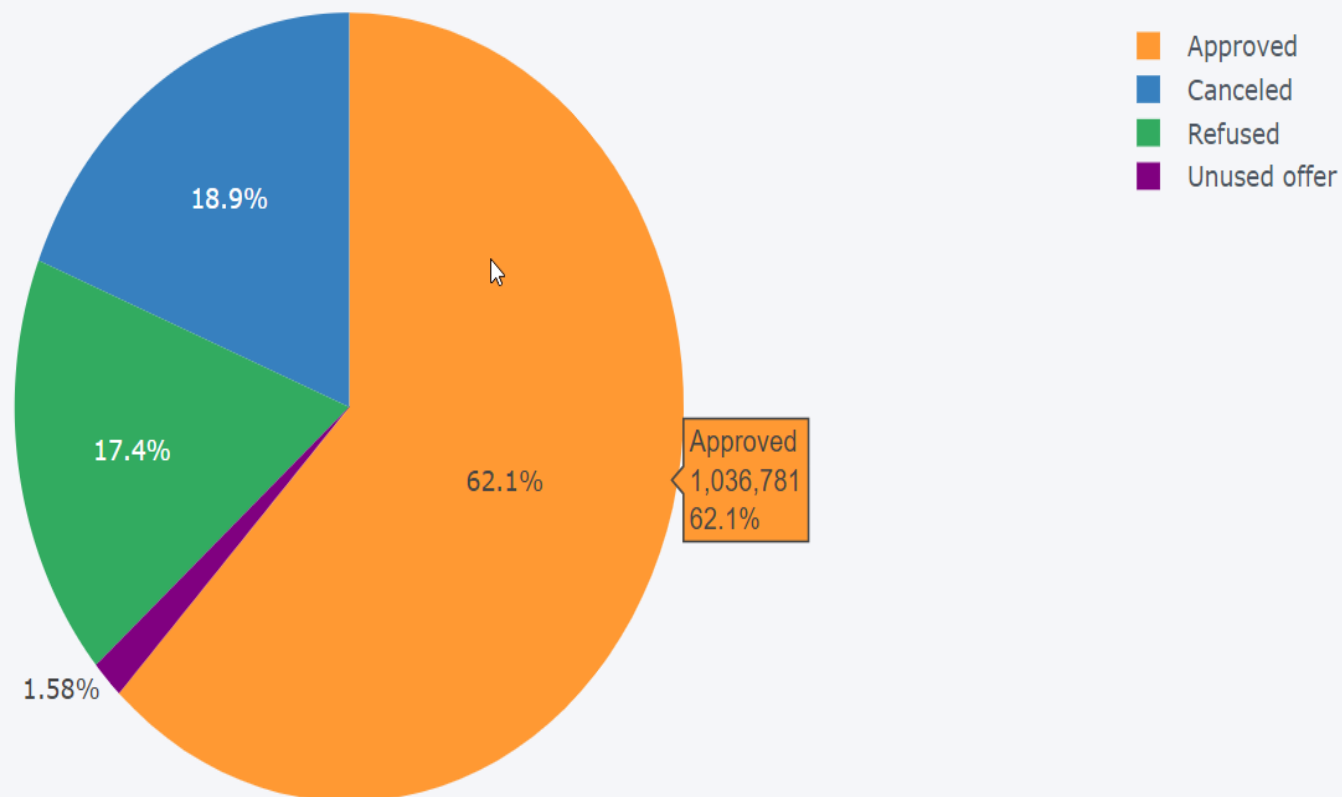`<AxesSubplot:xlabel='AMT_ANNUITY'>`

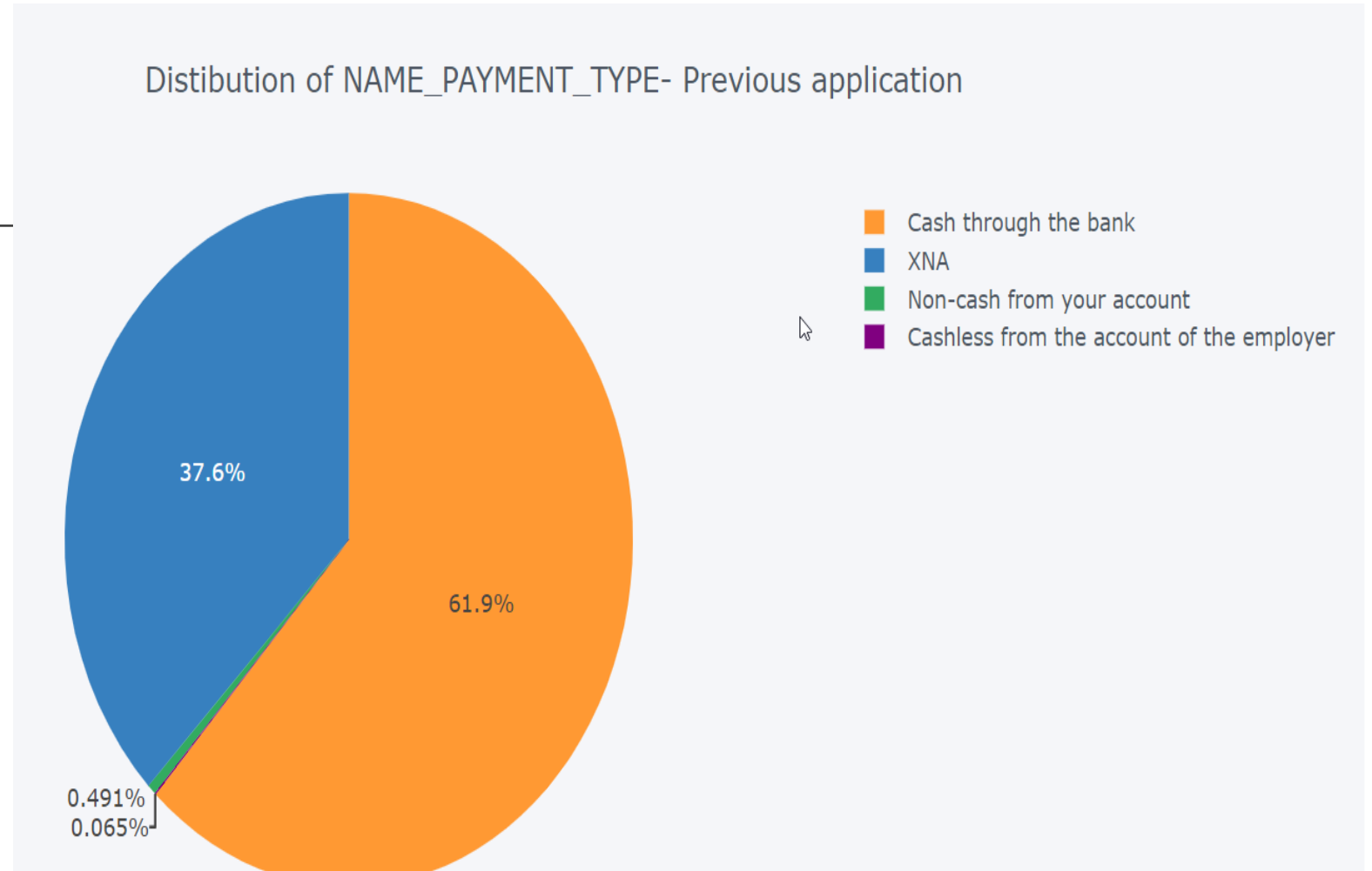# Univariate Analysis on Previous Application

Contract Status Distribution

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

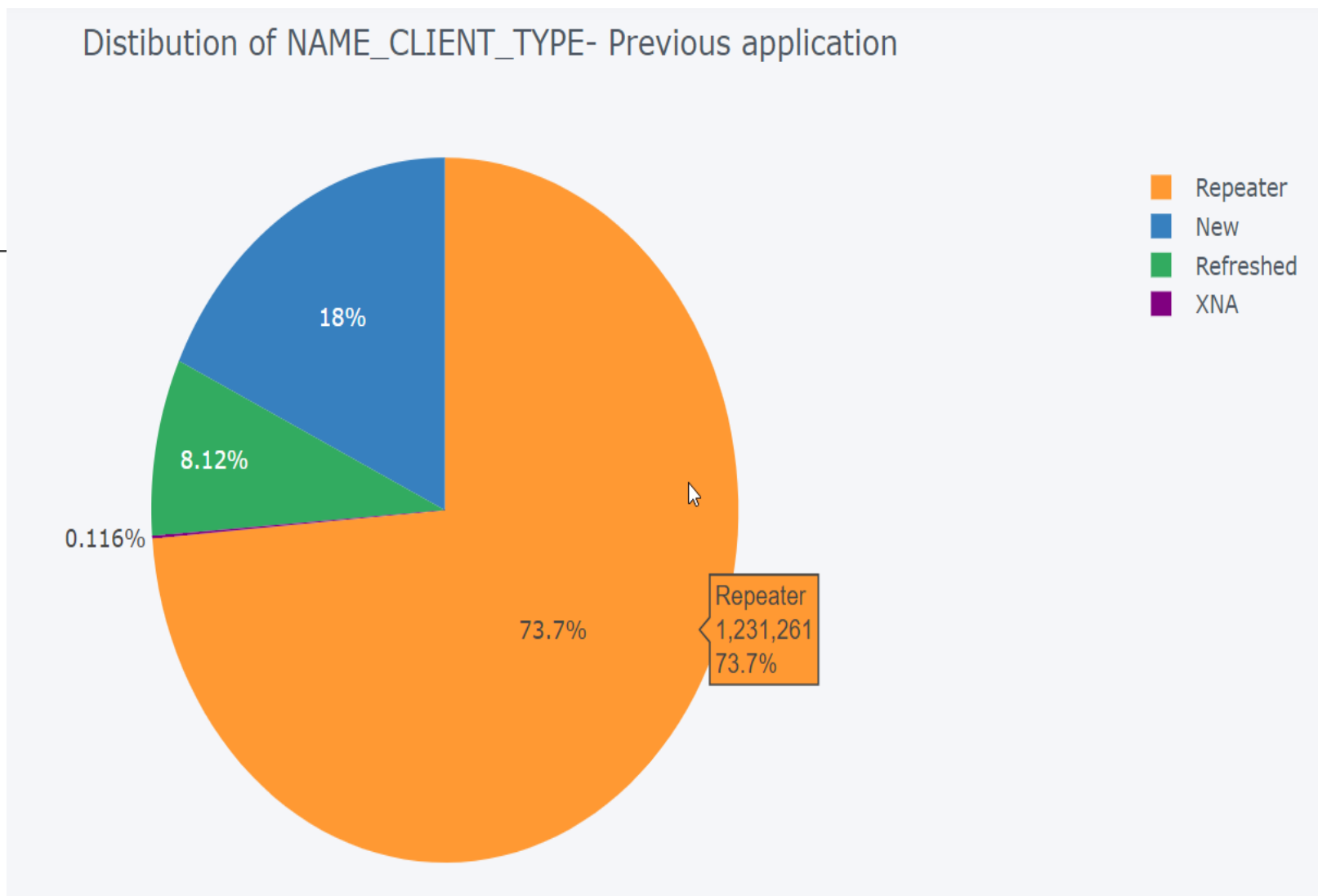

Distibution of NAME_CONTRACT_STATUS- Previous application

- Approved
- Canceled
- Refused
- Unused offer

18.9%

17.4%

62.1%

1.58%

Approved
1,036,781
62.1%

Payment Type Distribution

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

Distibution of NAME_PAYMENT_TYPE- Previous application

- Cash through the bank
- XNA
- Non-cash from your account
- Cashless from the account of the employer

37.6%

61.9%

0.491%
0.065%

Client Type Distribution

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

Distibution of NAME_CLIENT_TYPE- Previous application

- Repeater
- New
- Refreshed
- XNA

18%

8.12%

0.116%

73.7%

Repeater
1,231,261
73.7%

Portfolio Distribution

In previous application POS Portfolio is higher than cash portfolio

Distibution of NAME_PORTFOLIO- Previous application

- POS
- Cash
- XNA
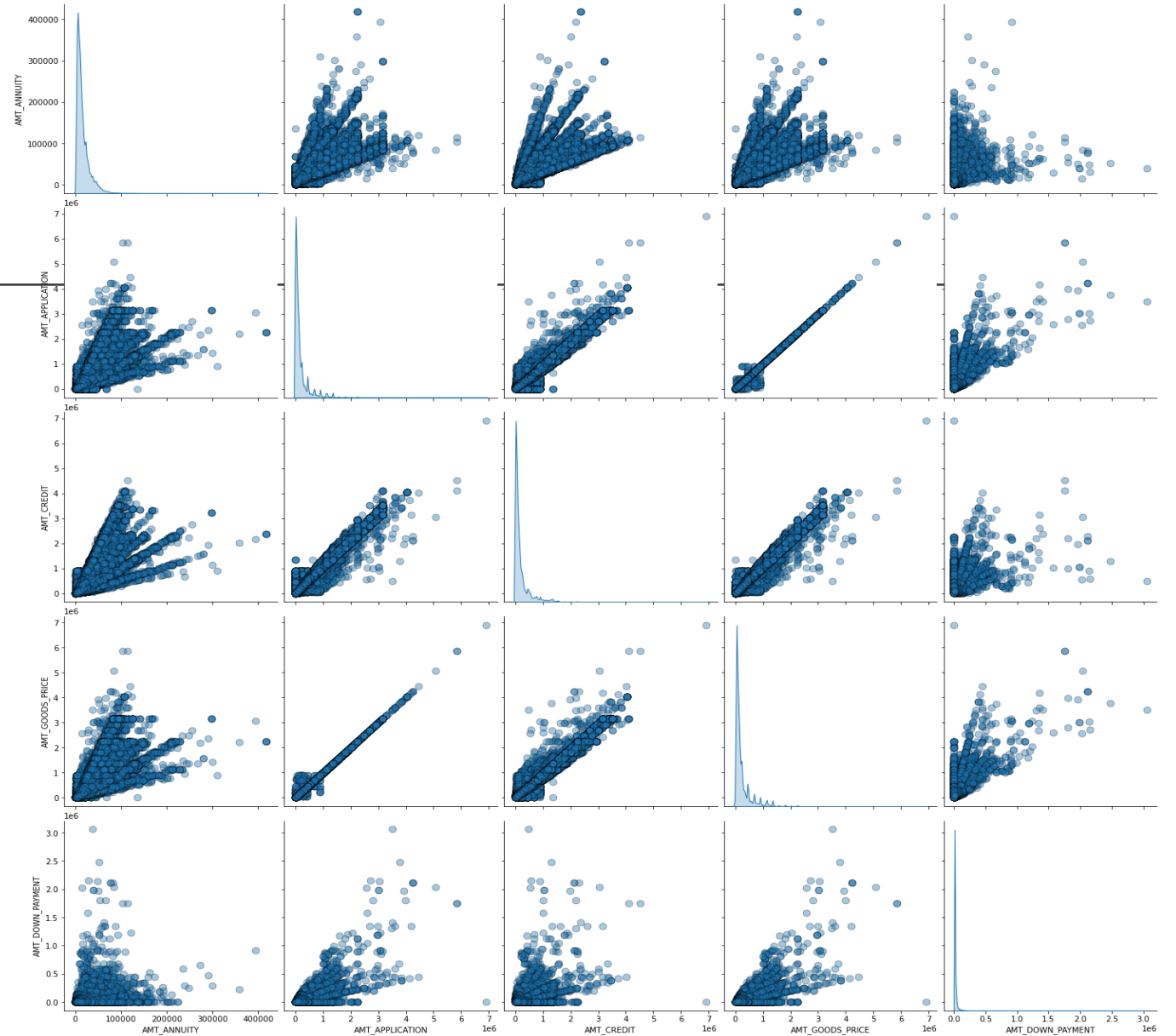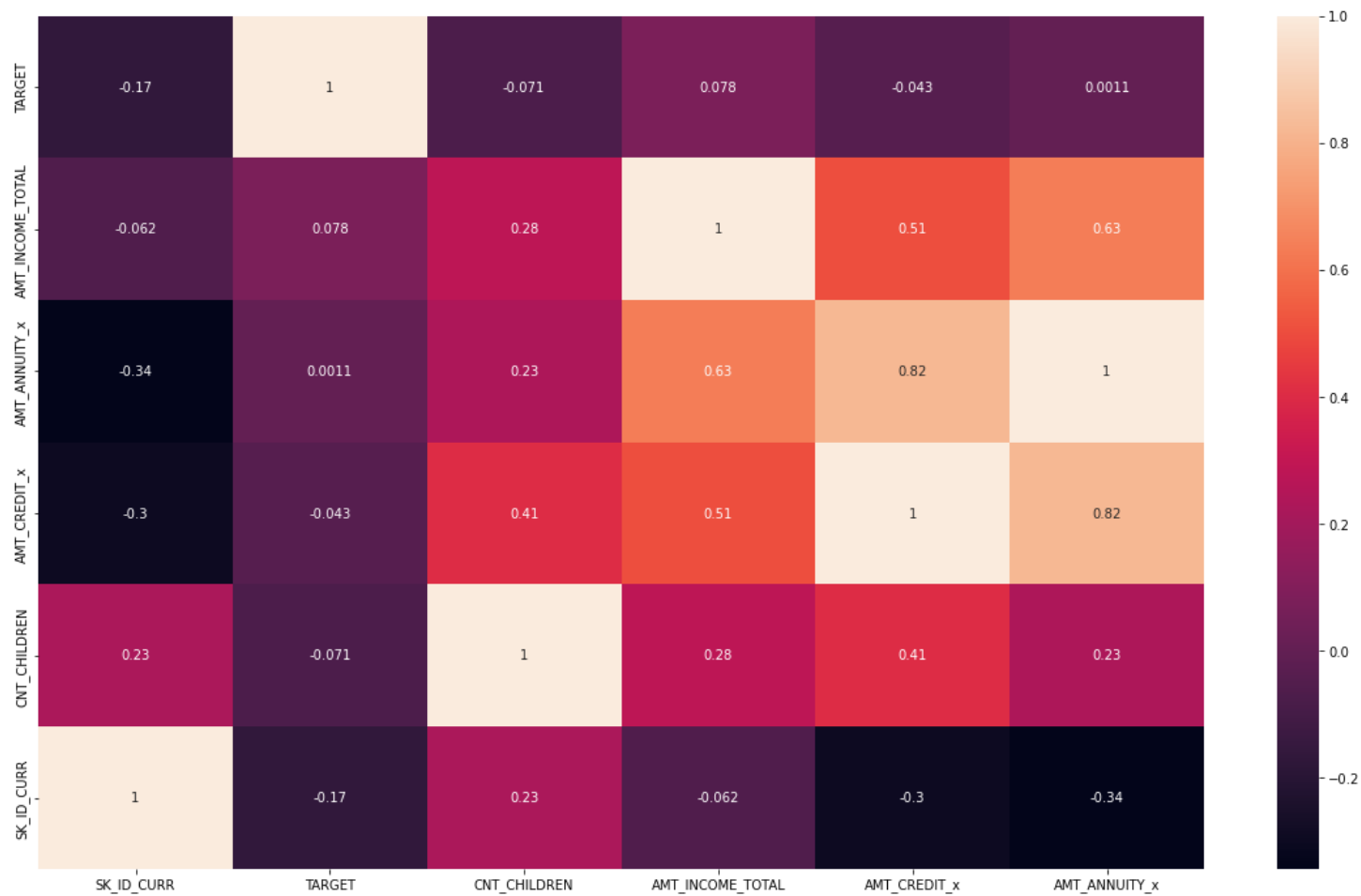- Cards
- Cars

41.4%
27.6%
22.3%
8.68%
0.0254%
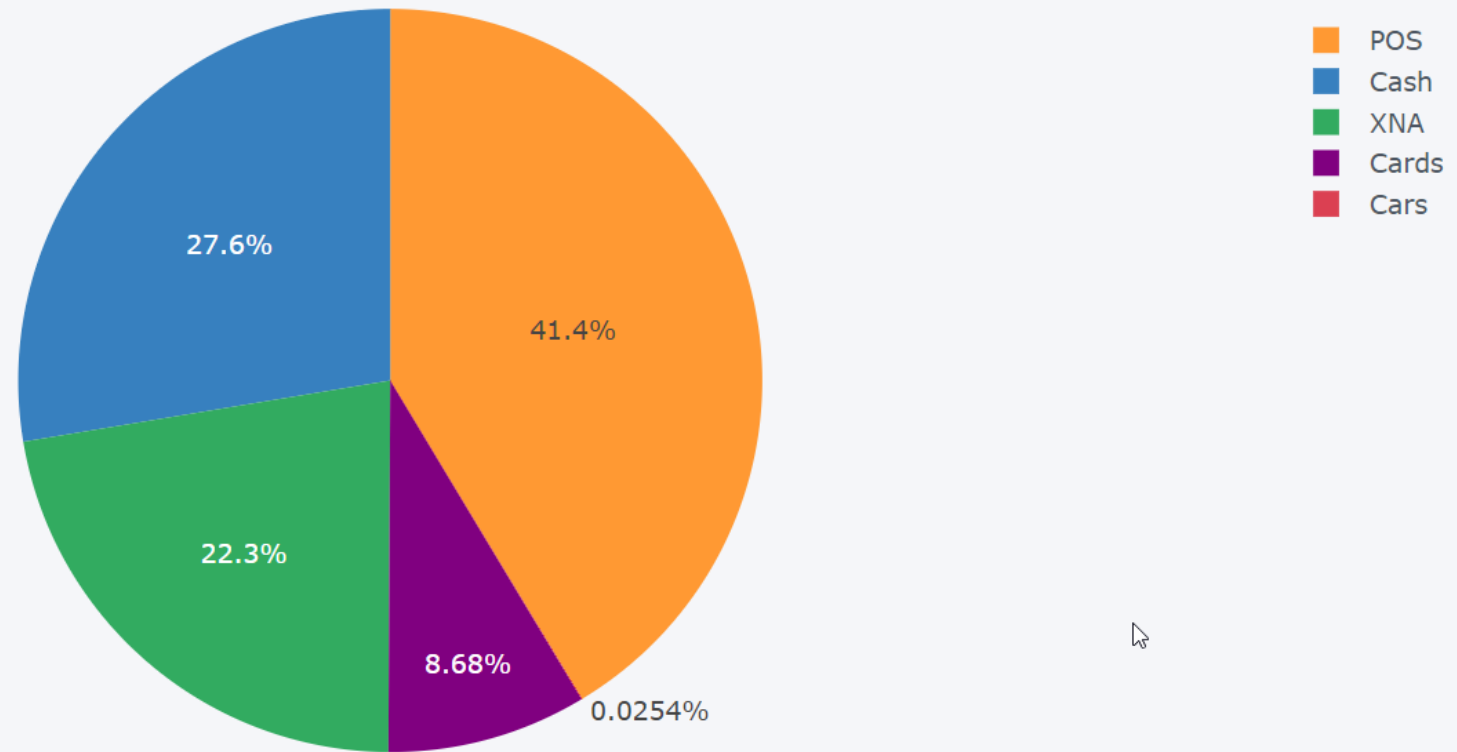
Bivariate Analysis on Previous Application

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