

A decorative graphic on the left side of the slide, consisting of a network of white lines and small circles on a blue gradient background, resembling a circuit board or a neural network structure.

# CREDIT EDA CASE STUDY

BY SOUMITA PACHHAL

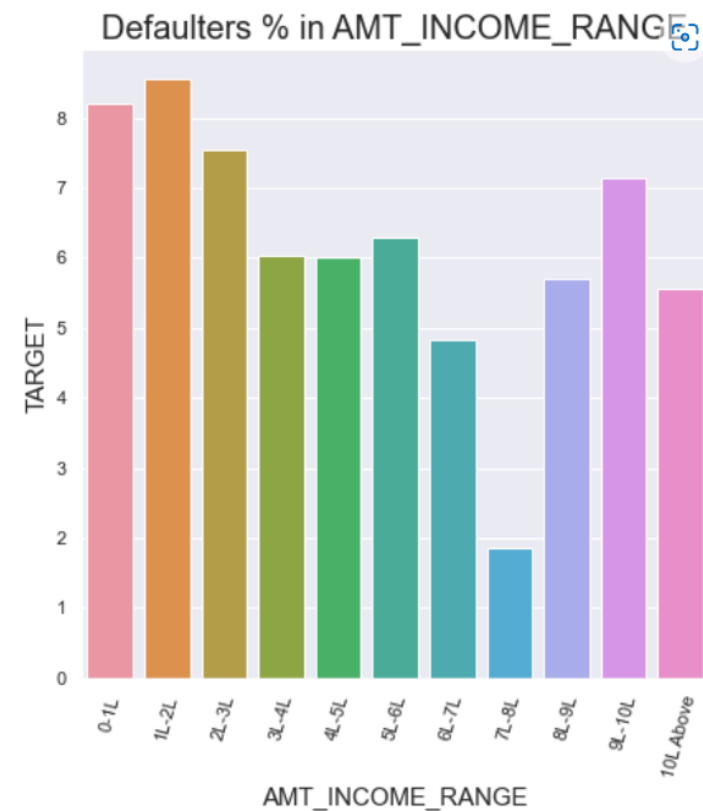
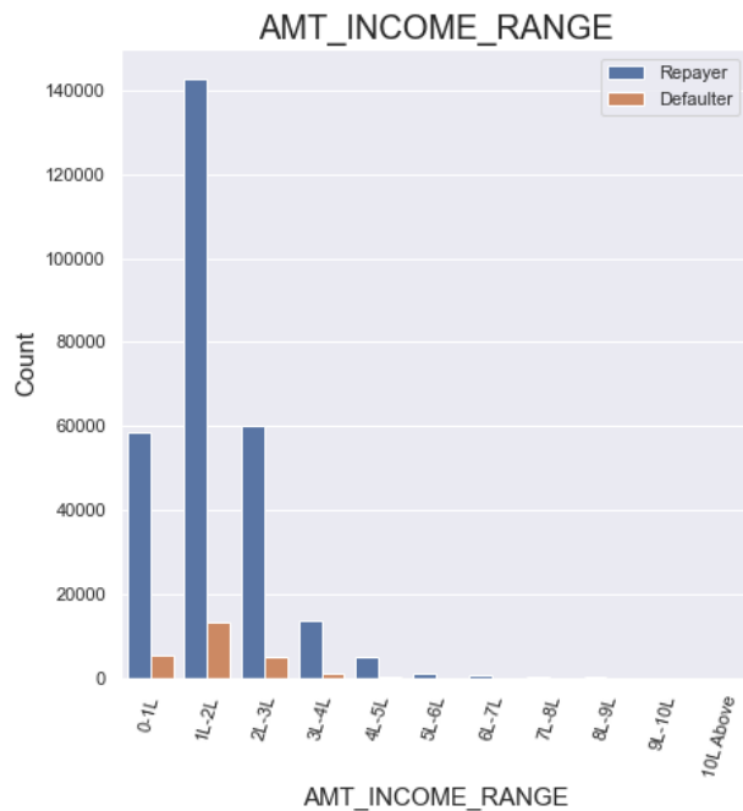
The background is a blue gradient. In the corners, there are white line-art illustrations of circuit boards or neural network structures, consisting of lines and small circles.

# CATEGORICAL UNIVARIATE ANALYSIS

# DISTRIBUTION OF INCOME RANGE

Points to be concluded from the graph on the right side.

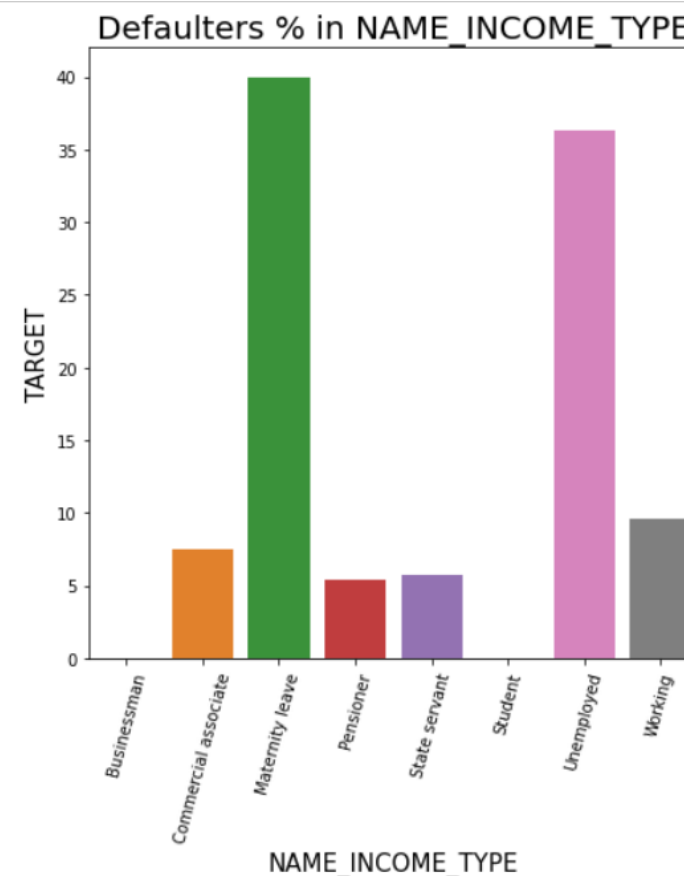
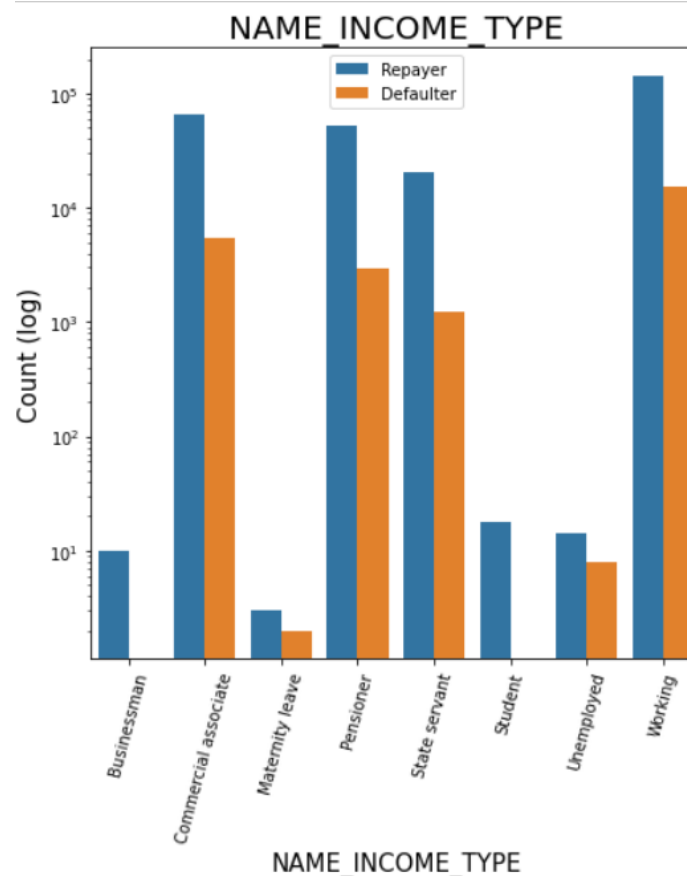
- Majority of the applications have an Income total of less than 3 Lakhs
- Application with Income less than 3 Lakhs has a high probability of defaulting
- Applicants with an Income of 7-8 Lakhs are less likely to default.



# DISTRIBUTION OF INCOME TYPE

Points to be concluded from the graph on the right.

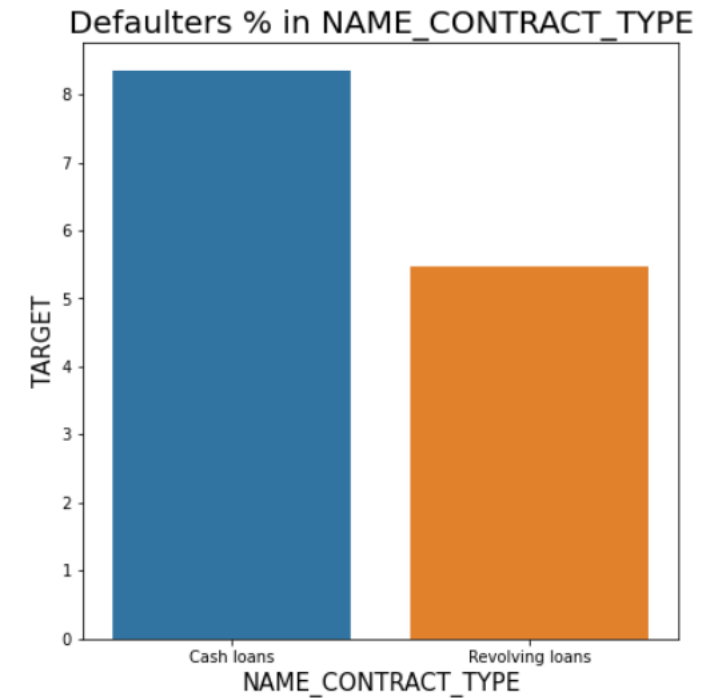
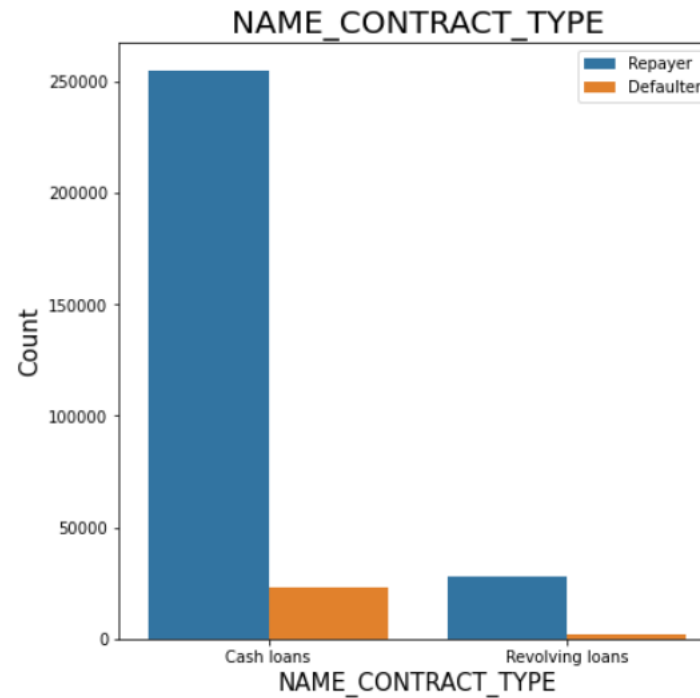
- Most of applicants for loans income type is Working, followed by Commercial associate, Pensioner and State servant.
- The applicants who are on Maternity leave have defaulting percentage of 40% which is the highest, followed by Unemployed (37%). The rest under average around 10% defaultees.
- Student and Businessmen though less in numbers, do not have default record. Safest two categories for providing loan



# DISTRIBUTION FOR CONTRACT TYPE

Points to be concluded from the graph on the right.

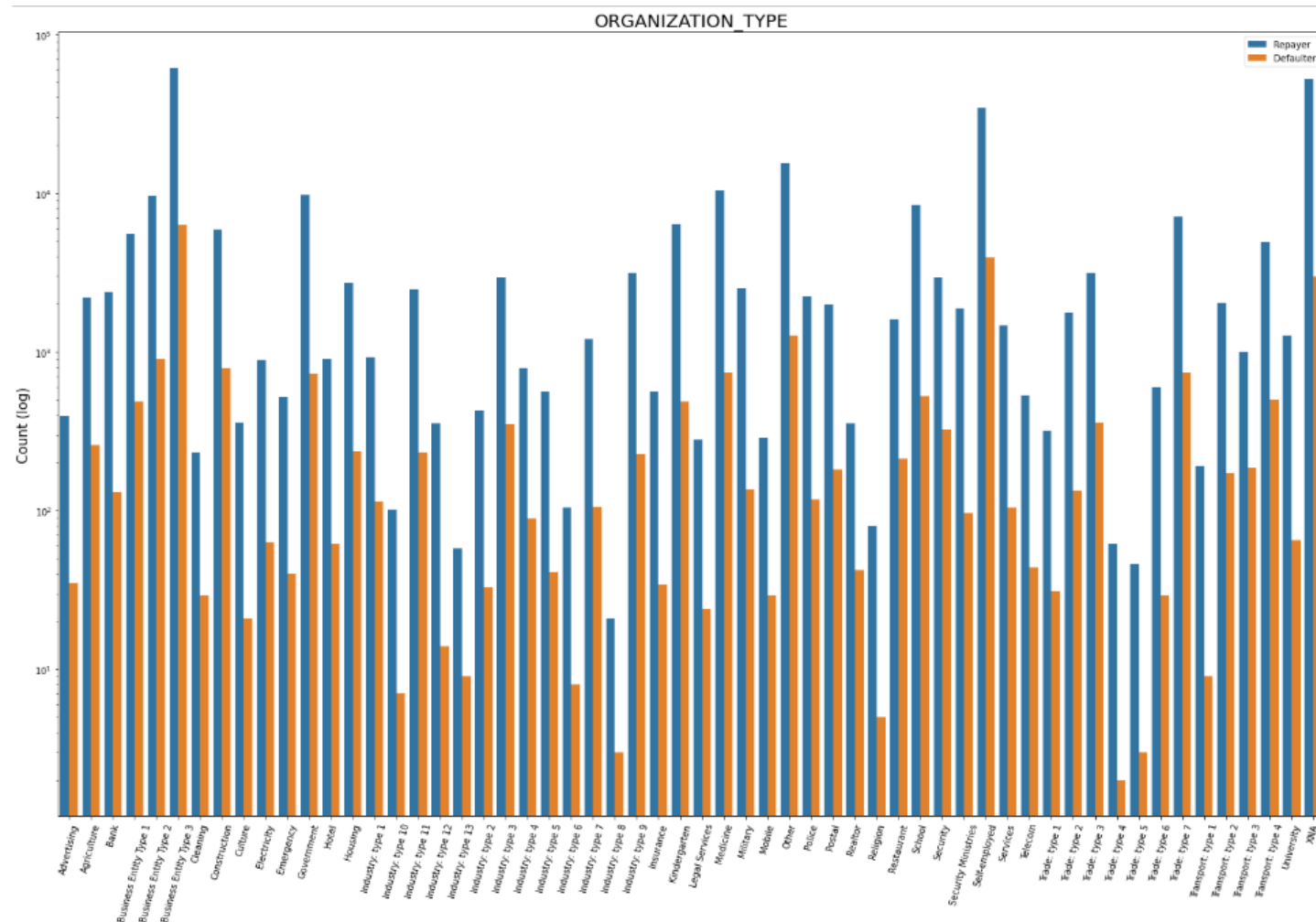
- Revolving loans are just a small fraction (10%) from the total number of loans
- Around 8-9% Cash loan applicants and 5-6% Revolving loan applicant are in defaulters



# DISTRIBUTION OF ORGANIZATION TYPE

Points to be concluded from the graph on the right.

- Organizations with highest percent of defaulters are Transport: type 3 (16%), Industry: type 13 (13.5%), Industry: type 8 (12.5%) and Restaurant (less than 12%).
- Self employed people have relative high defaulting rate, to be safer side loan disbursement should be avoided or provide loan with higher interest rate to mitigate the risk of defaulting.
- Most of the people application for loan are from Business Entity Type 3
- For a very high number of applications, Organization type information is unavailable(XNA)
- It can be seen that following category of organization type has lesser defaulters thus safer for providing loans: Trade Type 4 and 5, Industry type 8



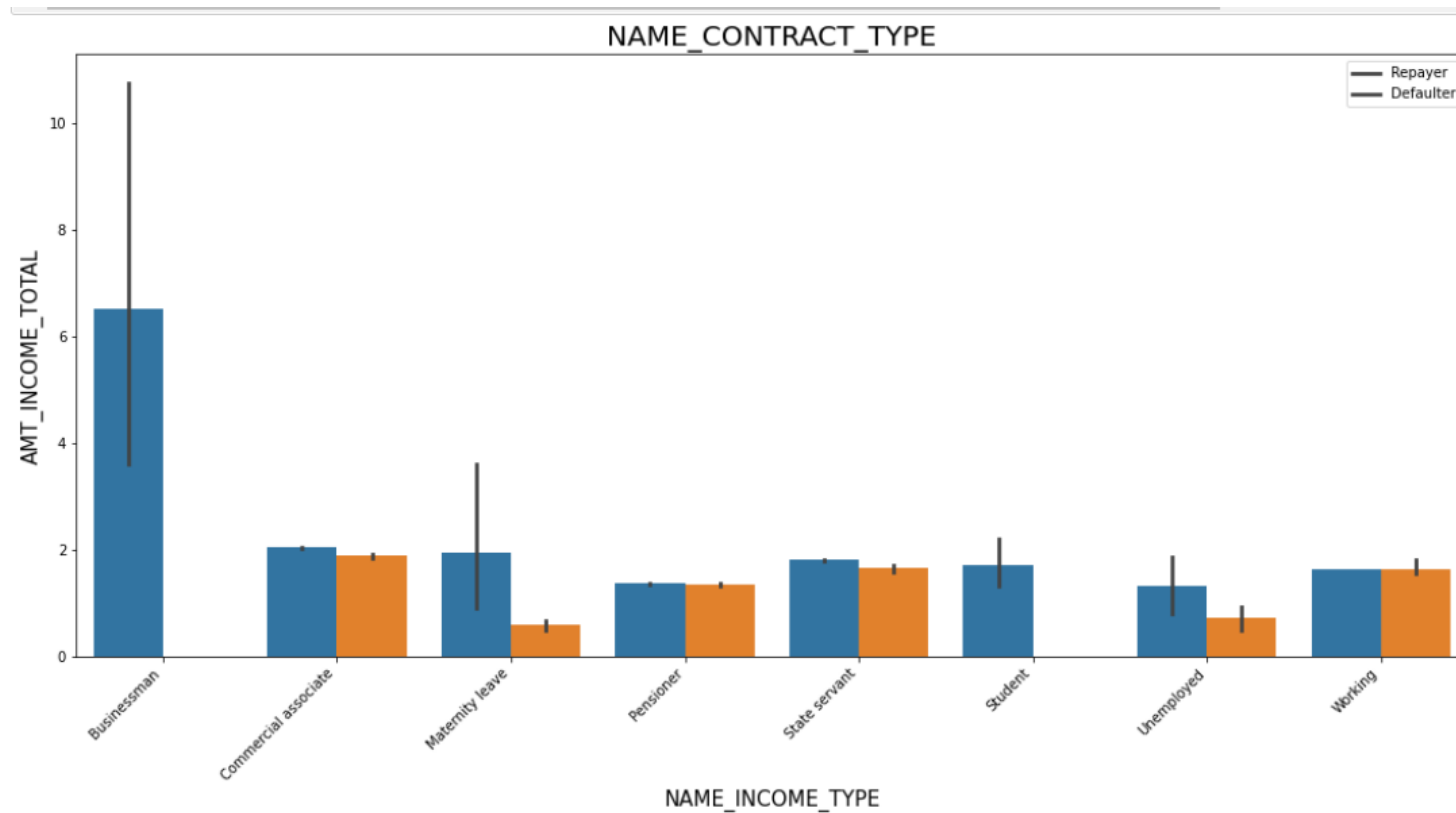
The background is a blue gradient. In the corners, there are white line-art decorations resembling circuit boards or neural networks, with lines and small circles.

# CATEGORICAL BIVARIATE OR MULTIVARIATE ANALYSIS

## Income type vs Income Amount Range

Points to be concluded from the graph on the right side.

- It can be seen that Businessman income is the highest and the estimated range with default 95% confidence level seem to indicate that the income of a Businessman could be in the range of slightly close to 4 lakhs and slightly above 10 lakhs





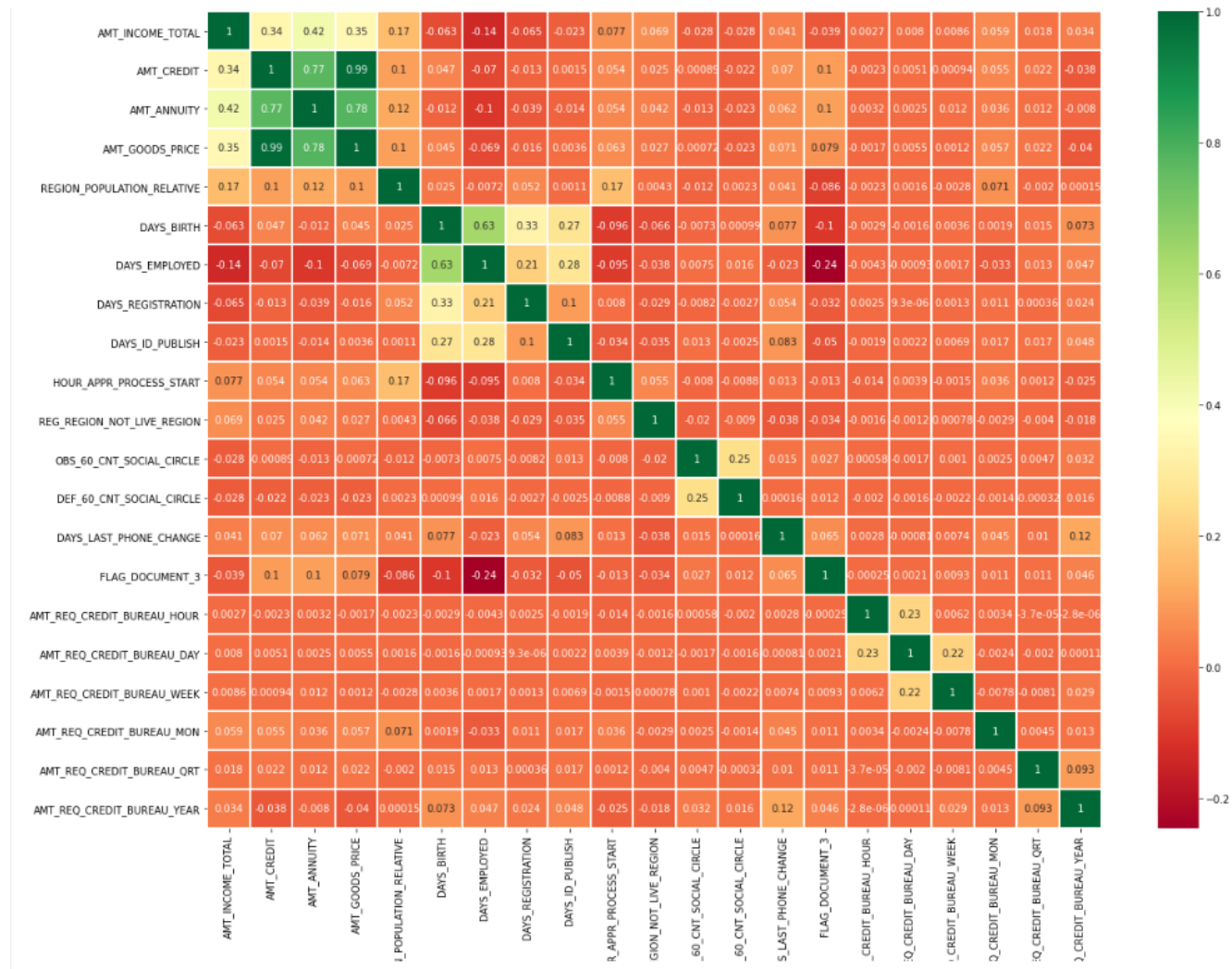
## Heatmap to see linear correlation among Repayers

Points to be concluded from the graph on the right side.

1. Credit amount is highly correlated with:

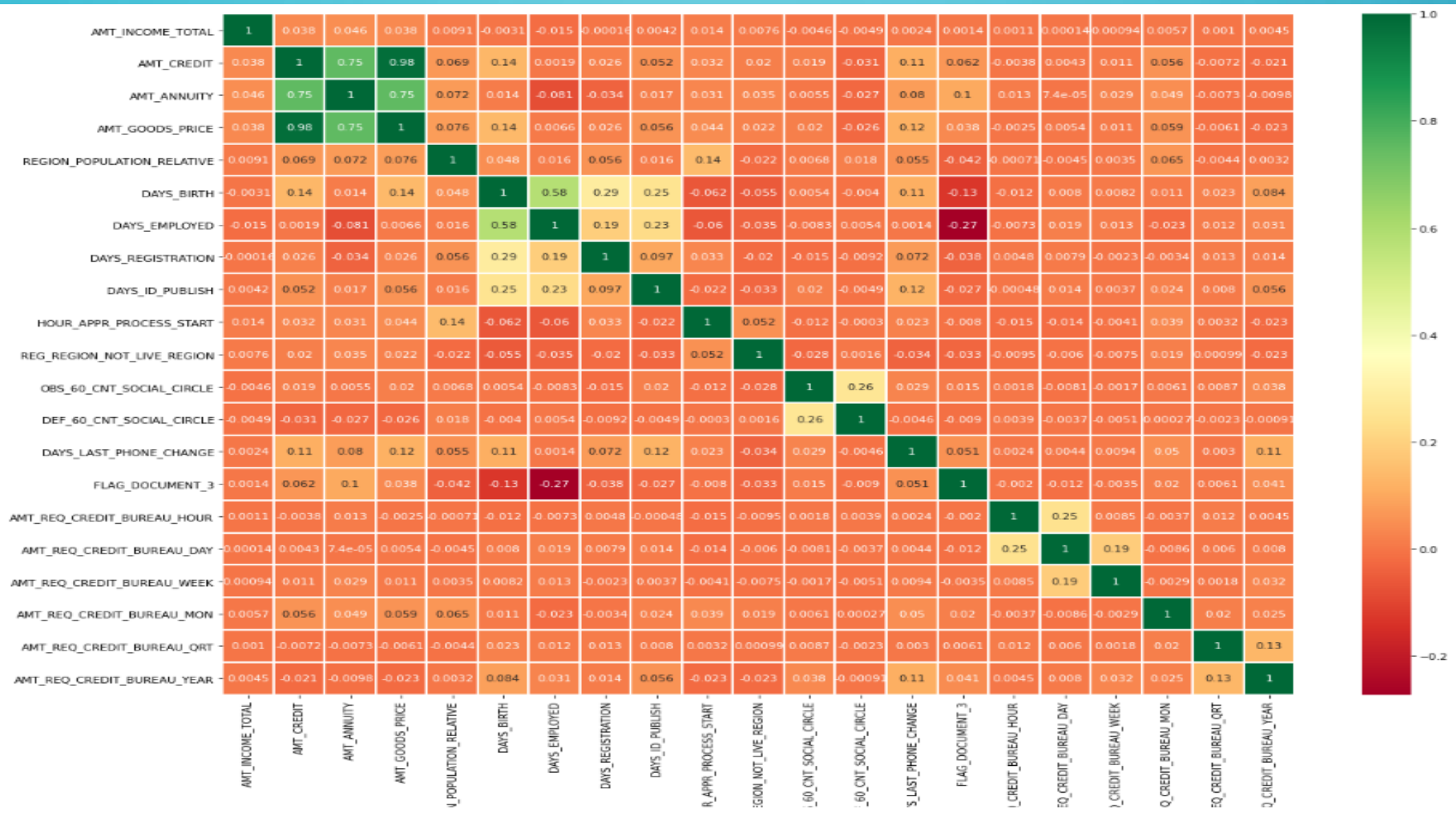
- Goods Price Amount
- Loan Annuity
- Total Income

2. We can also see that repayers have high correlation in number of days employed.



# CORRELATION

The background is a blue gradient with decorative white circuit-like lines in the corners. These lines consist of straight segments and small circles, resembling a stylized electronic circuit board or data network. The lines are located in the top-left, top-right, bottom-left, and bottom-right corners of the image.



# CORRELATION

**Points to be concluded from the graph presented before.**

- **Credit amount is highly correlated with good price amount which is same as repayers.**
- **Loan annuity correlation with credit amount has slightly reduced in defaulters(0.75) when compared to repayers(0.77)**
- **We can also see that repayers have high correlation in number of days employed(0.62) when compared to defaulters(0.58).**
- **There is a severe drop in the correlation between total income of the client and the credit amount(0.038) amongst defaulters whereas it is 0.342 among repayers.**
- **Days\_birth and number of children correlation has reduced to 0.259 in defaulters when compared to 0.337 in repayers.**
- **There is a slight increase in defaulted to observed count in social circle among defaulters(0.264) when compared to repayers(0.254)**

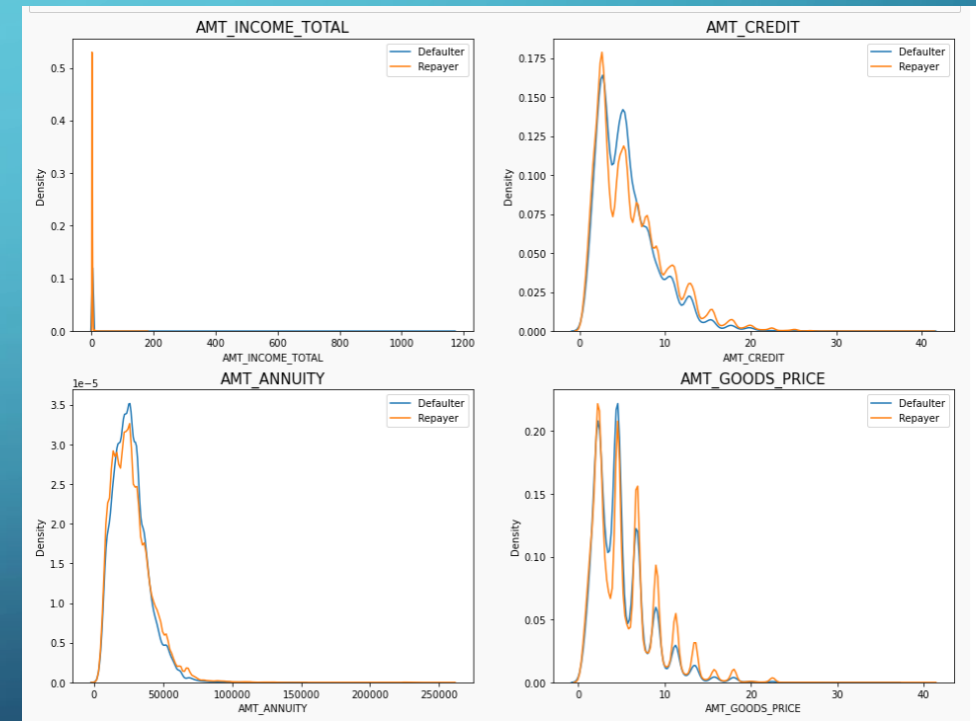
The background is a blue gradient with decorative white circuit-like lines in the corners. These lines consist of straight segments and small circles, resembling a stylized electronic circuit or data network.

# **NUMERICAL UNIVARIATE ANALYSIS**

# DISTRIBUTION PLOT

Few points can be concluded from the graph.

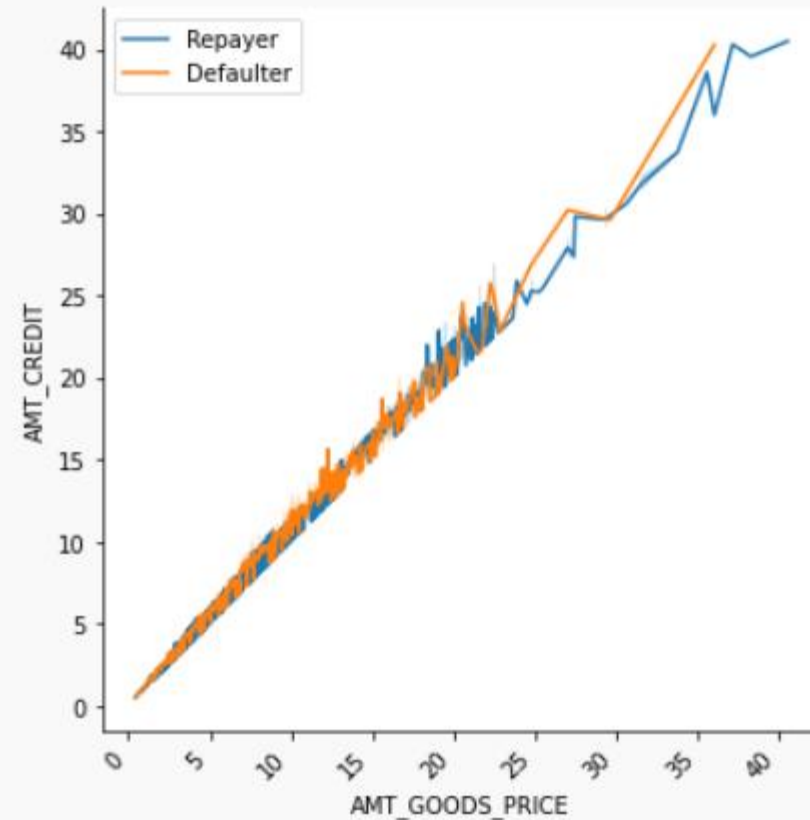
- Most no of loans are given for goods price below 10 lakhs
- Most people pay annuity below 50K for the credit loan
- Credit amount of the loan is mostly less then 10 lakhs
- The repayers and defaulters distribution overlap in all the plots and hence we cannot use any of these variables in isolation to make a decision



# NUMERICAL BIVARIATE ANALYSIS

Few points can be concluded from the graph.

- When the credit amount goes beyond 30 Lakhs, there is an increase in defaulters

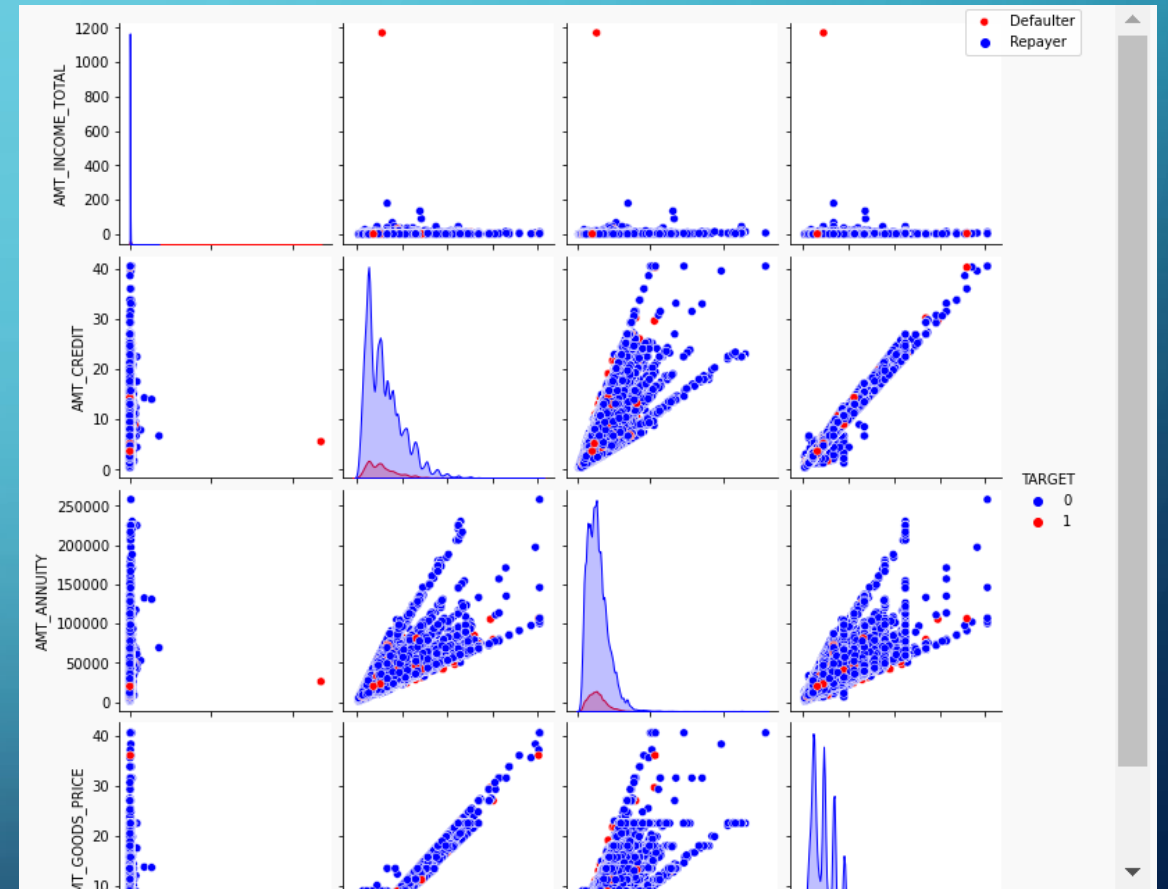




## *Pairplot between amount variable to draw reference against loan repayment status*

**Few points can be concluded from the graph.**

- **When Annuity Amount > 15K and Good Price Amount > 20 Lakhs, there is a lesser chance of defaulters**
- **Loan Amount(AMT\_CREDIT) and Goods price(AMT\_GOODS\_PRICE) are highly correlated as based on the scatterplot where most of the data are consolidated in form of a line**
- **There are very less defaulters for AMT\_CREDIT >20 Lakhs**

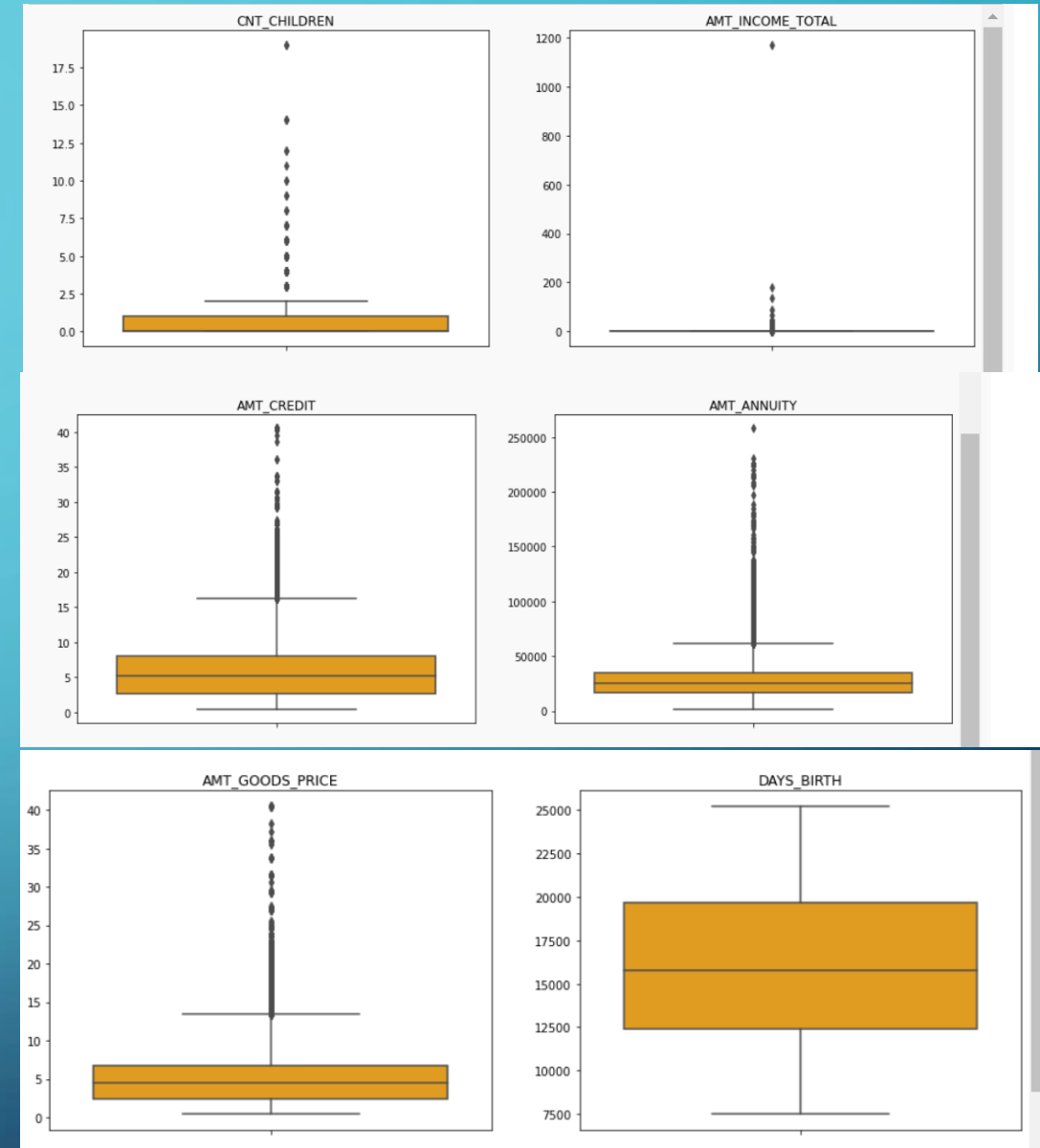




# BOXPLOT FOR APPLICATION\_DATA

Few points can be concluded from the graph.

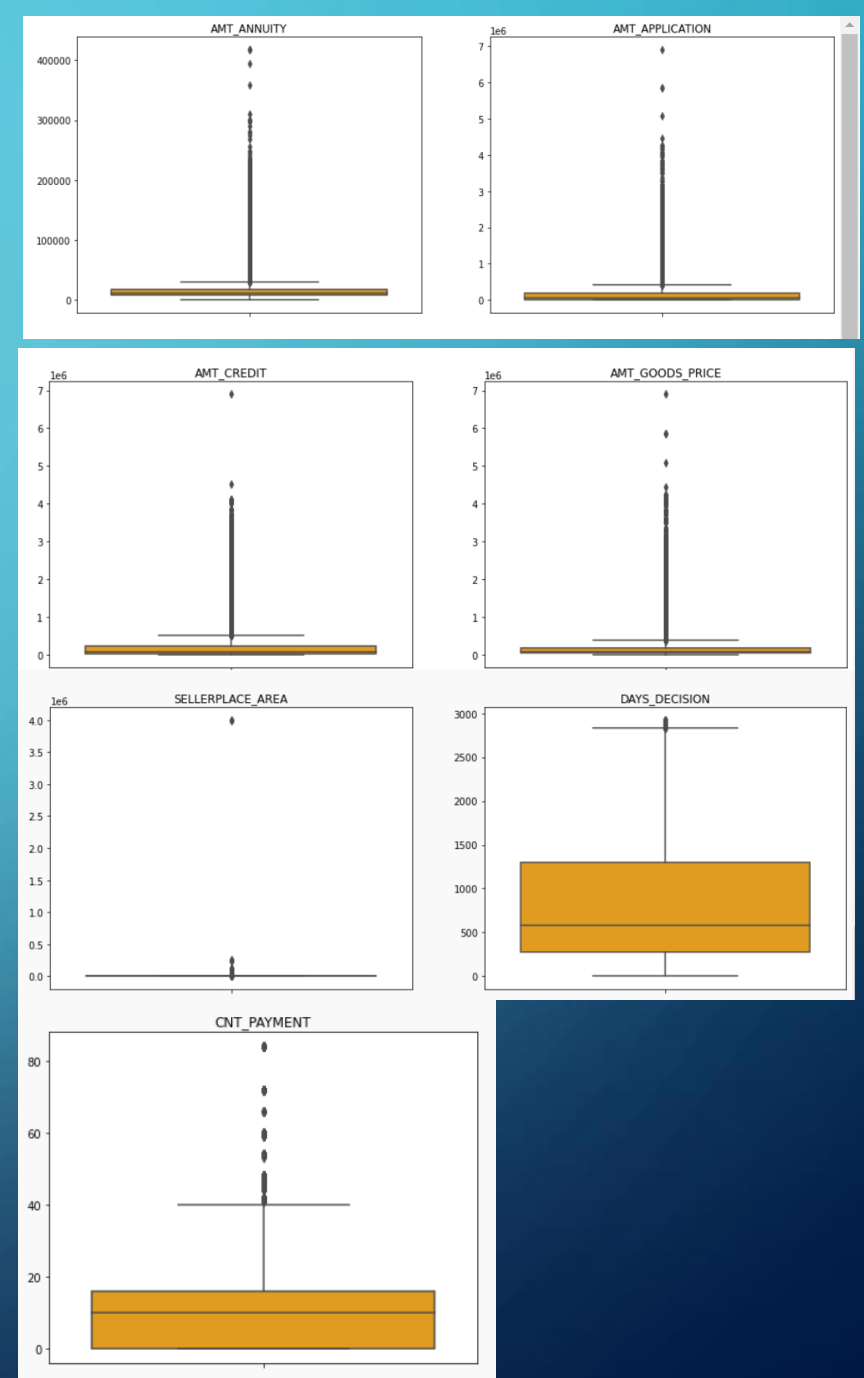
- AMT\_ANNUITY, AMT\_CREDIT, AMT\_GOODS\_PRICE, CNT\_CHILDREN have some number of outliers.
- AMT\_INCOME\_TOTAL has huge number of outliers which indicate that few of the loan applicants have high income when compared to the others.
- DAYS\_BIRTH has no outliers which means the data available is reliable.
- DAYS\_EMPLOYED has outlier values around 350000(days) which is around 958 years which is impossible and hence this has to be incorrect entry.



# BOXPLOT FOR PREVIOUS\_APPLICATION

Few points can be concluded from the graph.

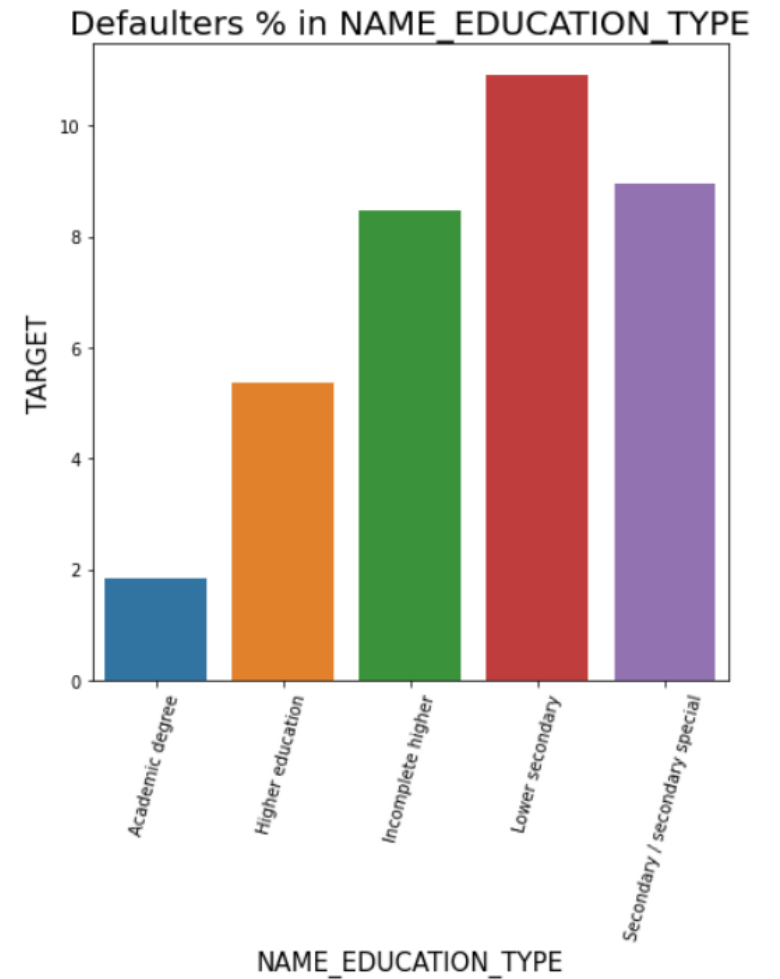
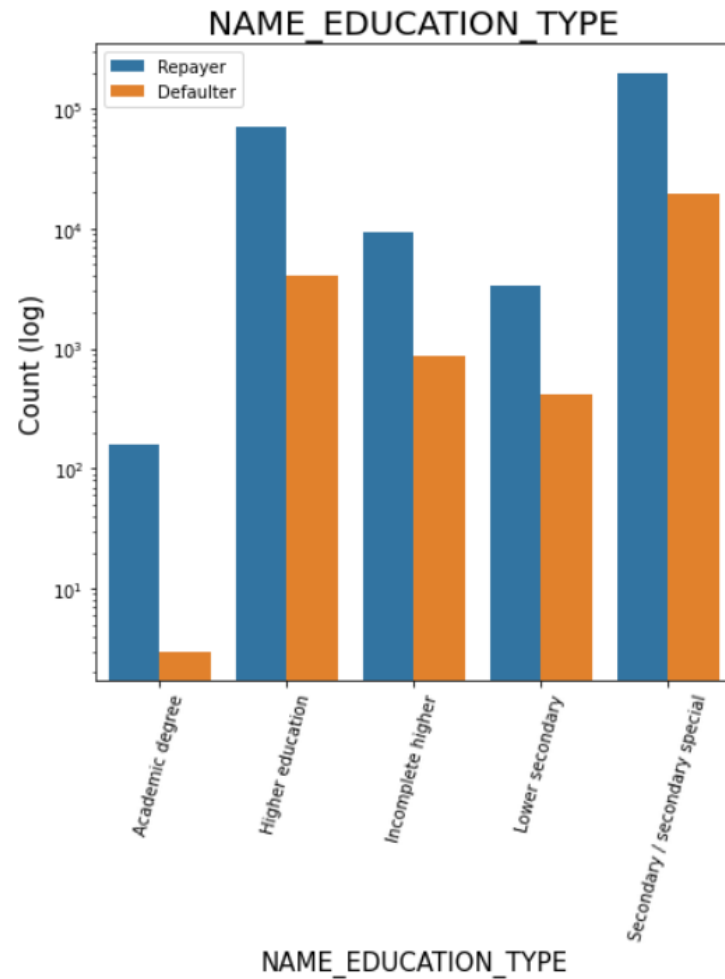
- **AMT\_ANNUITY, AMT\_APPLICATION, AMT\_CREDIT, AMT\_GOODS\_PRICE, SELLERPLACE\_AREA** have huge number of outliers.
- **CNT\_PAYMENT** has few outlier values.
- **DAYS\_DECISION** has little number of outliers indicating that these previous applications decisions were taken long back.



## Analyzing Education Type based on loan repayment status

Few points can be concluded from the graph.

- Majority of clients have Secondary/secondary special education, followed by clients with Higher education.
- Very few clients have an academic degree
- Lower secondary category have highest rate of defaulting around 11%.
- People with Academic degree are least likely to default.



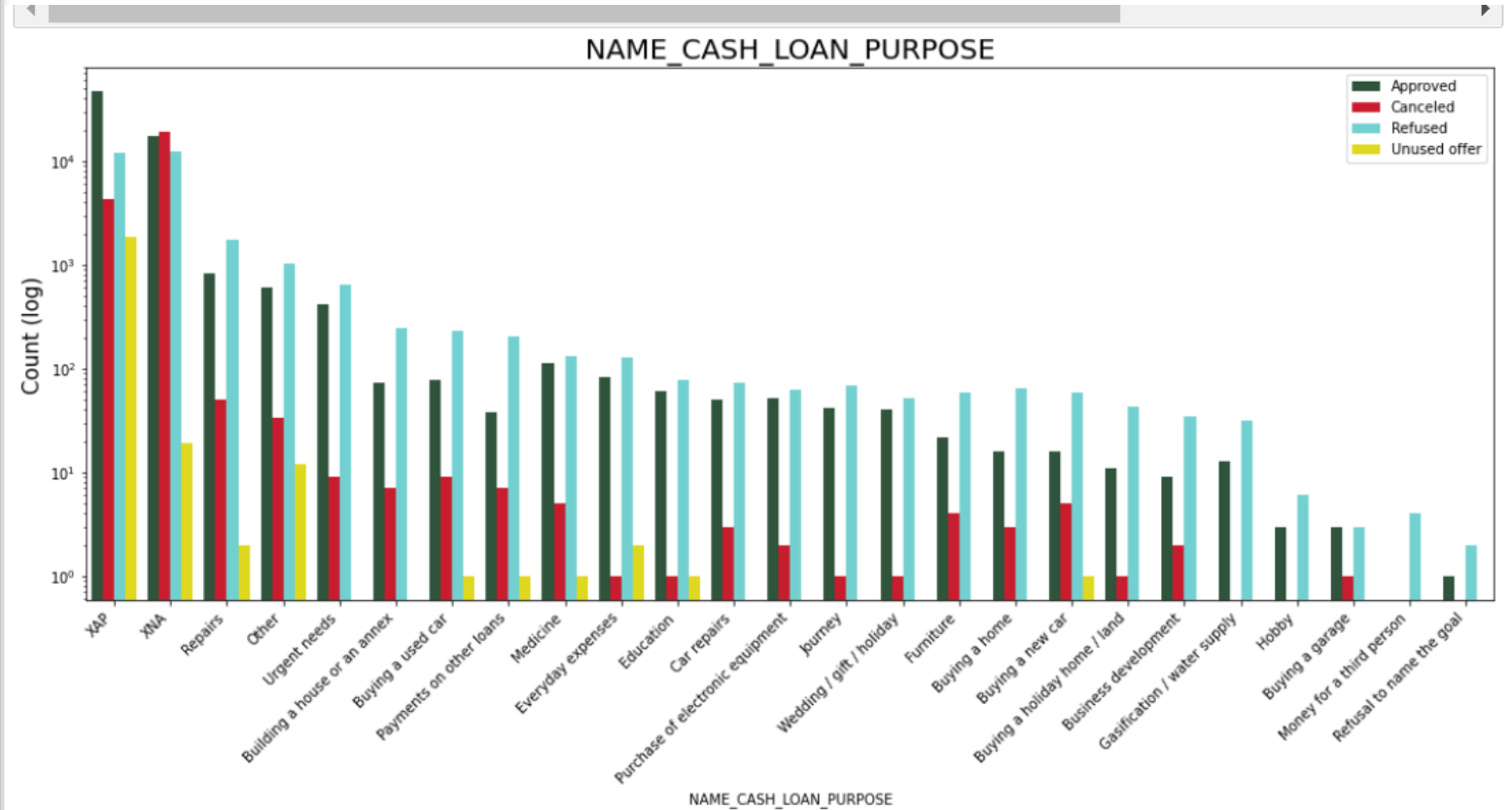
The background is a blue gradient with decorative white circuit-like lines in the corners. These lines consist of straight segments and small circles, resembling a stylized electronic circuit or data flow diagram.

# UNIVARIATE ANALYSIS AFTER MERGING PREVIOUS DATA

## CONTRACT STATUS VS PURPOSE OF THE LOAN

Few points can be concluded from the graph.

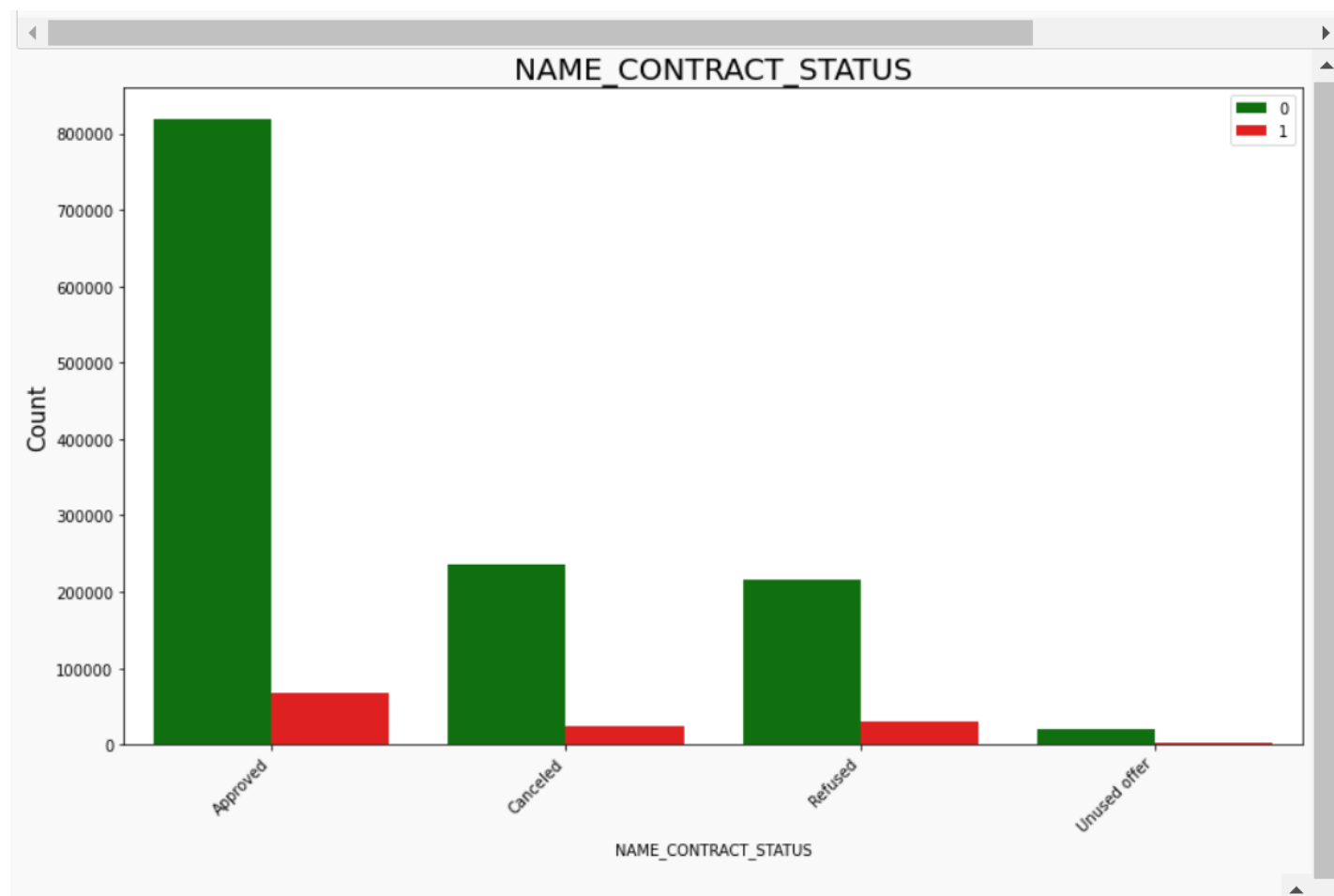
- Loan purpose has high number of unknown values (XAP, XNA)
- Loan taken for the purpose of Repairs looks to have highest default rate
- Huge number application have been rejected by bank or refused by client which are applied for Repair or Other. from this we can infer that repair is considered high risk by bank. Also, either they are rejected or bank offers loan on high interest rate which is not feasible by the clients and they refuse the loan



### *Contract Status based on loan repayment status*

**Few points can be concluded from the graph.**

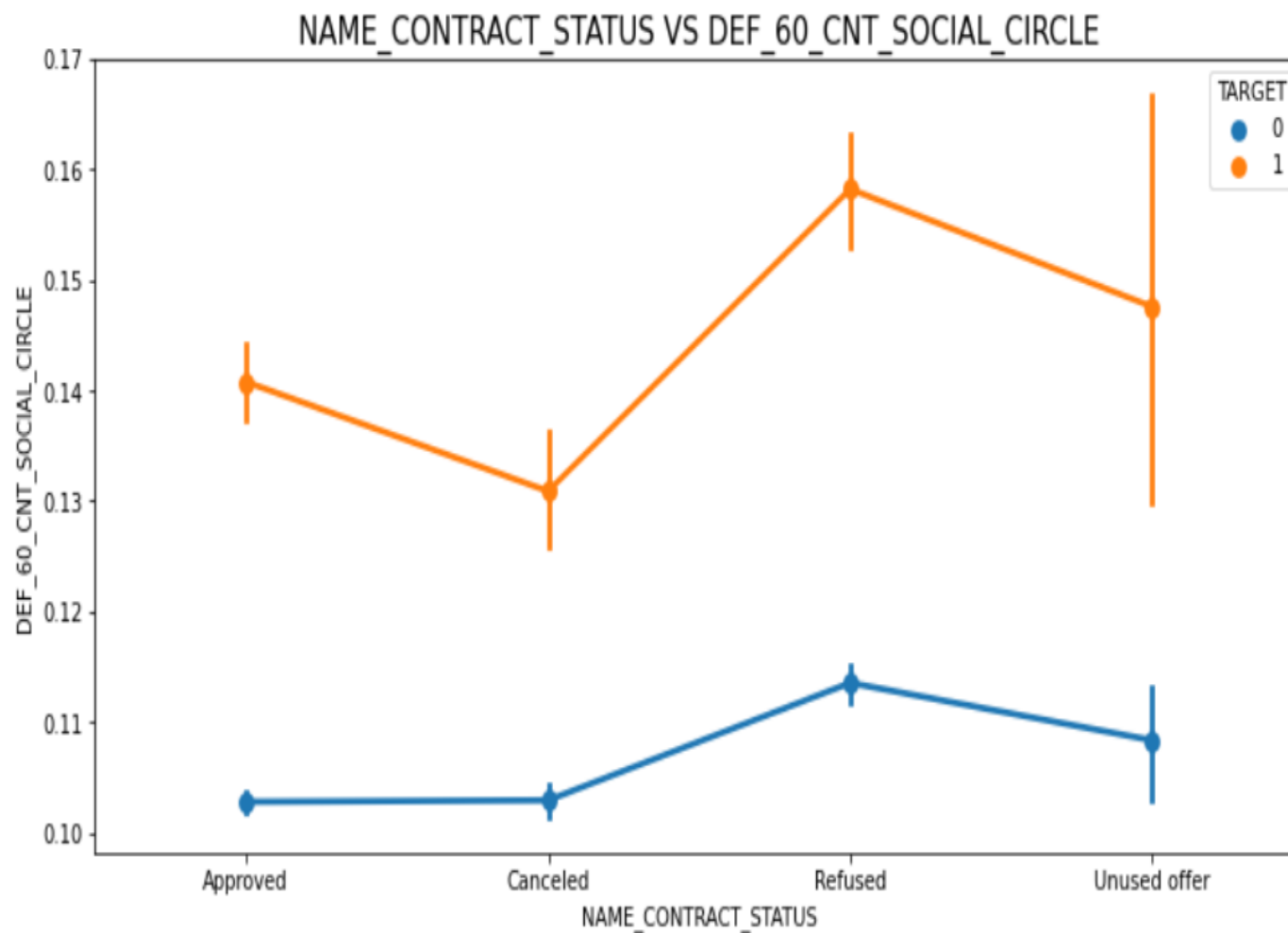
- **90% of the previously cancelled client have actually repayed the loan. Revising the interest rates would increase business opportunity for these clients**
- **88% of the clients who have been previously refused a loan has payed back the loan in current case.**
- **Refusal reason should be recorded for further analysis as these clients could turn into potential repaying customer.**



**Relationship between people who defaulted in the last 60 days being in the client's social circle and contact status**

**Few points can be concluded from the graph.**

- Clients who have average of 0.13 or higher their DEF\_60\_CNT\_SOCIAL\_CIRCLE score tend to default more and thus analysing client's social circle could help in the disbursement of the loan



# CONCLUSION

Decisive Factor whether an applicant will be Repayer:

- NAME\_EDUCATION\_TYPE: Academic degree has less defaults.
- NAME\_INCOME\_TYPE: Student and Businessmen have no defaults.
- REGION\_RATING\_CLIENT: RATING 1 is safer.
- ORGANIZATION\_TYPE: Clients with Trade Type 4 and 5 and Industry type 8 have defaulted less than 3%
- DAYS\_BIRTH: People above age of 50 have low probability of defaulting
- DAYS\_EMPLOYED: Clients with 40+ year experience having less than 1% default rate
- AMT\_INCOME\_TOTAL: Applicant with Income more than 700,000 are less likely to default
- NAME\_CASH\_LOAN\_PURPOSE: Loans bought for Hobby, Buying garage are being repayed mostly.
- CNT\_CHILDREN: People with zero to two children tend to repay the loans

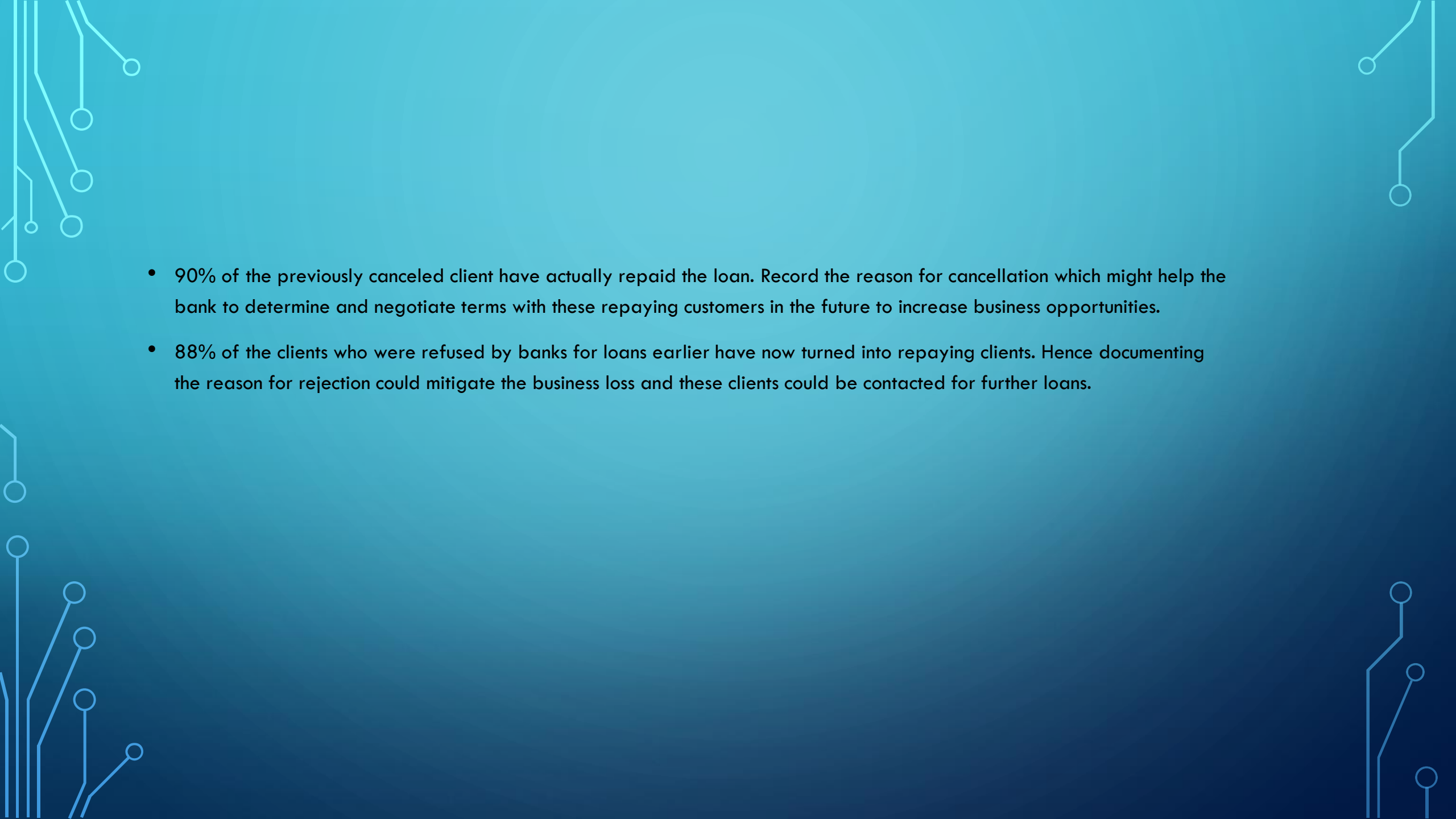


### Decisive Factor whether an applicant will be Defaulter

- **CODE\_GENDER:** Men are at relatively higher default rate
- **NAME\_FAMILY\_STATUS :** People who have civil marriage or who are single default a lot.
- **NAME\_EDUCATION\_TYPE:** People with Lower Secondary & Secondary education
- **NAME\_INCOME\_TYPE:** Clients who are either at Maternity leave OR Unemployed default a lot.
- **REGION\_RATING\_CLIENT:** People who live in Rating 3 has highest defaults.
- **OCCUPATION\_TYPE:** Avoid Low-skill Laborers, Drivers and Waiters/barmen staff, Security staff, Laborers and Cooking staff as their default rate is huge.
- **ORGANIZATION\_TYPE:** Organizations with highest percent of loans not repaid are Transport: type 3 (16%), Industry: type 13 (13.5%), Industry: type 8 (12.5%) and Restaurant (less than 12%). Self-employed people have relative high defaulting rate, and thus should be avoided to be approved for loan or provide loan with higher interest rate to mitigate the risk of defaulting.
- **DAYS\_BIRTH:** Avoid young people who are in age group of 20-40 as they have higher probability of defaulting
- **DAYS\_EMPLOYED:** People who have less than 5 years of employment have high default rate.
- **CNT\_CHILDREN & CNT\_FAM\_MEMBERS:** Client who have children equal to or more than 9 default 100% and hence their applications are to be rejected.
- **AMT\_GOODS\_PRICE:** When the credit amount goes beyond 3lakhs, there is an increase in defaulters.

Factors that Loan can be given on Condition of High-Interest rate to mitigate any default risk leading to business loss:

- **NAME\_HOUSING\_TYPE:** High number of loan applications are from the category of people who live in Rented apartments & living with parents and hence offering the loan would mitigate the loss if any of those default.
- **AMT\_CREDIT:** People who get loan for 3-6 Lakhs tend to default more than others and hence having higher interest specifically for this credit range would be ideal.
- **AMT\_INCOME:** Since 90% of the applications have Income total less than 3Lakhs and they have high probability of defaulting, they could be offered loan with higher interest compared to other income category.
- **CNT\_CHILDREN & CNT\_FAM\_MEMBERS:** Clients who have 4 to 8 children has a very high default rate and hence higher interest should be imposed on their loans.
- **NAME\_CASH\_LOAN\_PURPOSE:** Loan taken for the purpose of Repairs seems to have highest default rate. A very high number of applications have been rejected by bank or refused by client in previous applications as well which has a purpose as repair or other. This shows that purpose repair is taken as high risk by bank and either they are rejected, or bank offers very high loan interest rate which is not feasible by the clients, thus they refuse the loan. The same approach could be followed in future as well.

- 
- The background is a blue gradient with decorative white circuit-like lines in the corners. These lines consist of straight segments and small circles, resembling a stylized electronic circuit or data network.
- 90% of the previously canceled client have actually repaid the loan. Record the reason for cancellation which might help the bank to determine and negotiate terms with these repaying customers in the future to increase business opportunities.
  - 88% of the clients who were refused by banks for loans earlier have now turned into repaying clients. Hence documenting the reason for rejection could mitigate the business loss and these clients could be contacted for further loans.



THANK YOU