

# CREDIT EDA CASE STUDY

By Sumit Bhatia

# PROBLEM STATEMENT.

- ◉ **This case study aims to identify patterns which indicate if a client has difficulty paying their installments .**
- ◉ **This may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.**
- ◉ **This will ensure that the consumers capable of repaying the loan are not rejected.**
- ◉ **Identification of such applicants using EDA is the aim of this case study.**

# ASSUMPTIONS

- ◉ Rows containing XAN and XAP values were dropped if they were in small numbers.
- ◉ If rows containing XAN and XAP values were high in number then these values were assumed as null values.
- ◉ They were not taken into consideration while deriving insights from the data.

# MISSING VALUE AND OUTLIER TREATMENT

## **Application Dataset**

1. Columns containing more than 40% of missing values were dropped.
2. Remaining missing values in continuous columns were imputed by their median.
3. Remaining missing values in categorical columns were imputed by their mode.
4. Outliers, if present in continuous columns were capped by applying upper bound and lower bound with the help of IQR.

## **Previous Dataset**

1. Columns containing more than 45% of missing values were dropped.
2. Remaining missing values in continuous columns were imputed by their median.
3. Remaining missing values in categorical columns were imputed by their mode.
4. Outliers, if present in continuous columns were capped by applying upper bound and lower bound with the help of IQR.

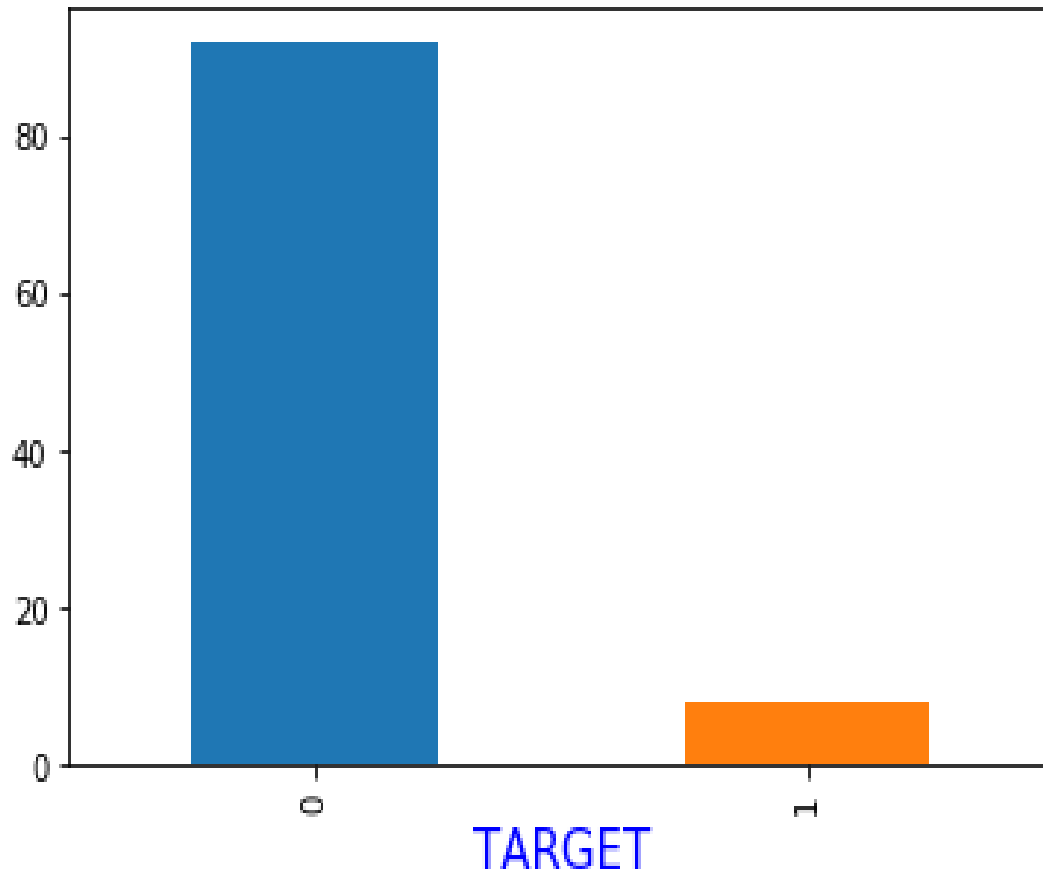
## DATA IMBALANCE

### Graph Showing Data Imbalance in TARGET Variable

As we can see in our data set only app. 8% clients face difficulties in loan repayment, which is quite practical also. Imbalance can be clearly seen in our "TARGET" column. So we will divide our application data set into two.

1. Clients with payment difficulties
2. Clients with no payment difficulties.

Also in "TARGET" column we will replace "1" with "defaulter" and "0" with "non-defaulter".



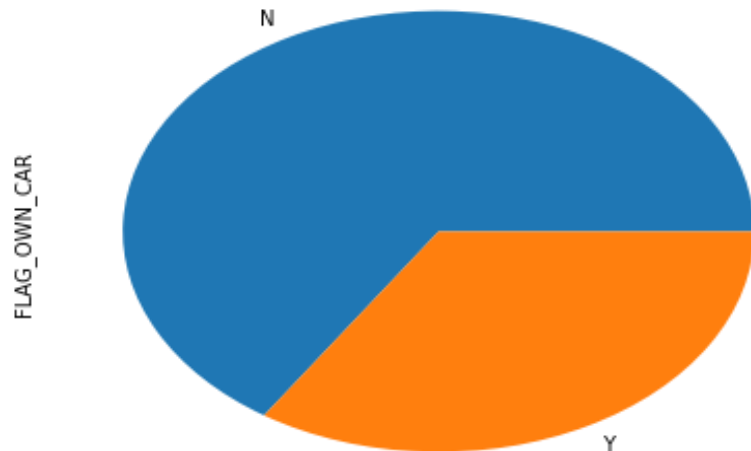
# UNIVARIATE ANALYSIS ON CATEGORICAL COLUMNS IN SEGMENTED APPLICATION DATASET

# CLIENTS WITH OWN CAR WRT TARGET

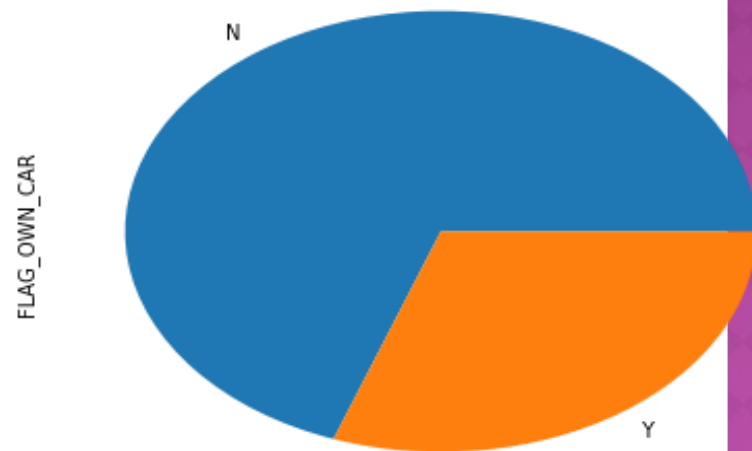
Insights from the graph :

1. More number of people defaulting the loan do not own a car.
2. People who do not own a car are giving higher number of loan applications.

Non-Defaulter



Defaulter



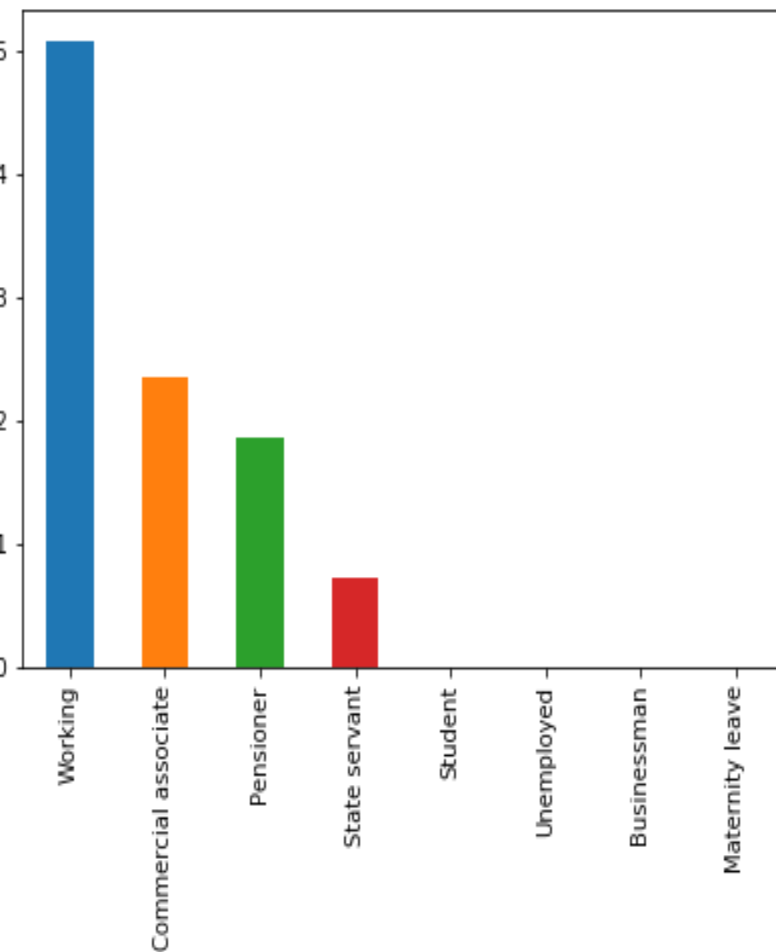
# CLIENTS INCOME TYPE WRT TARGET

Insights from the graph :

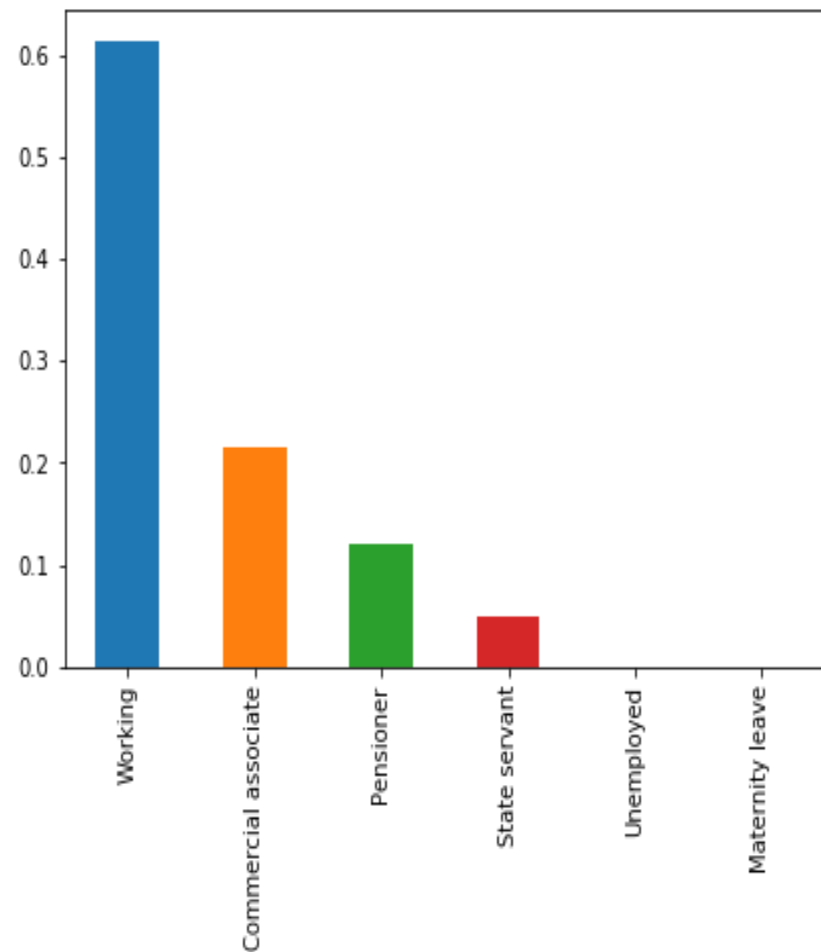
1. Pensioners are less likely to default as compared to other income types.

2. Working people are more likely to default. At the same time, working people are giving maximum business to the bank.

Non-Defaulter



Defaulter





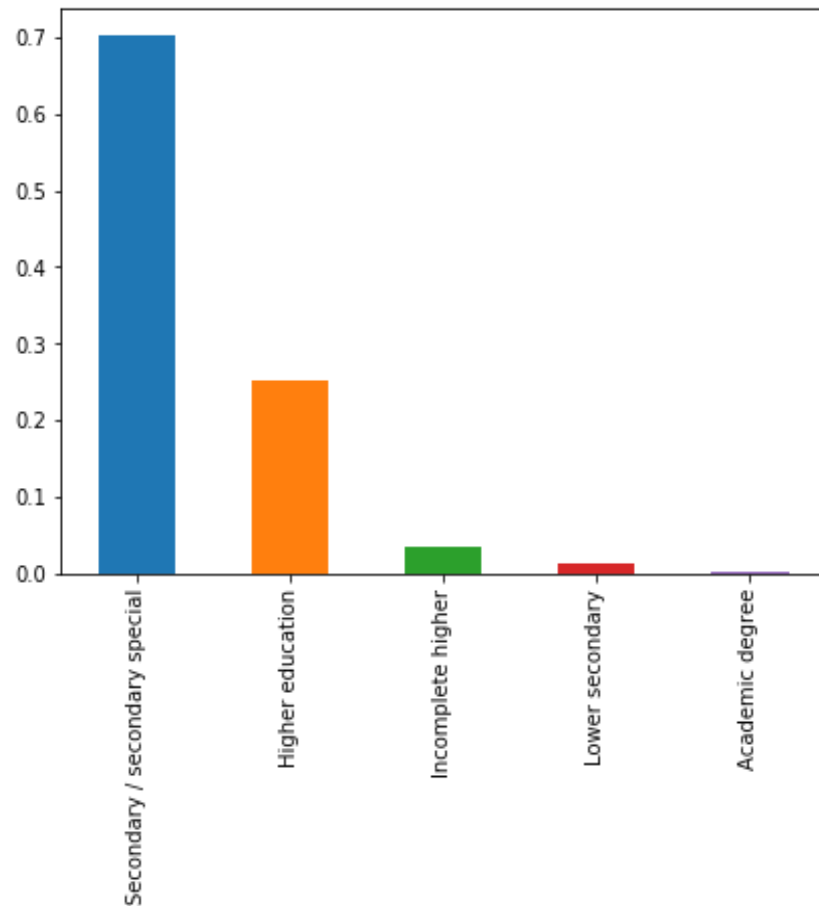
# CLIENT'S EDUCATION WRT TARGET

Insights from the graph :

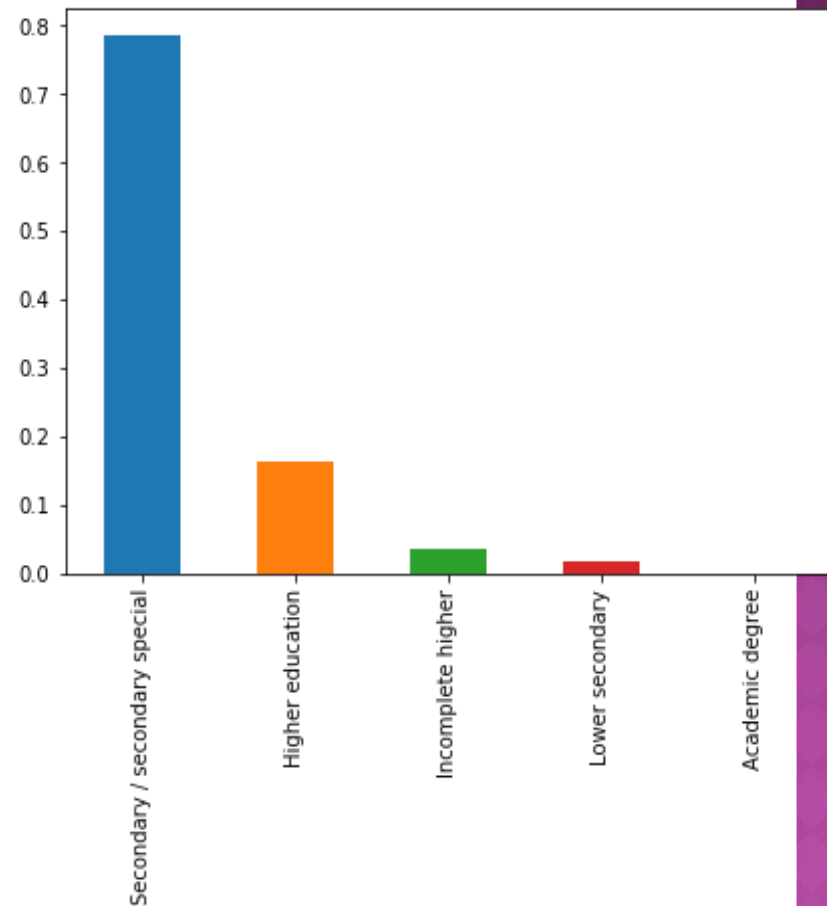
1. People with Secondary/Secondary special education are the ones with highest number of loan applications.

2. People in Higher Education category are less likely to default as compared to other categories.

Non-Defaulter



Defaulter



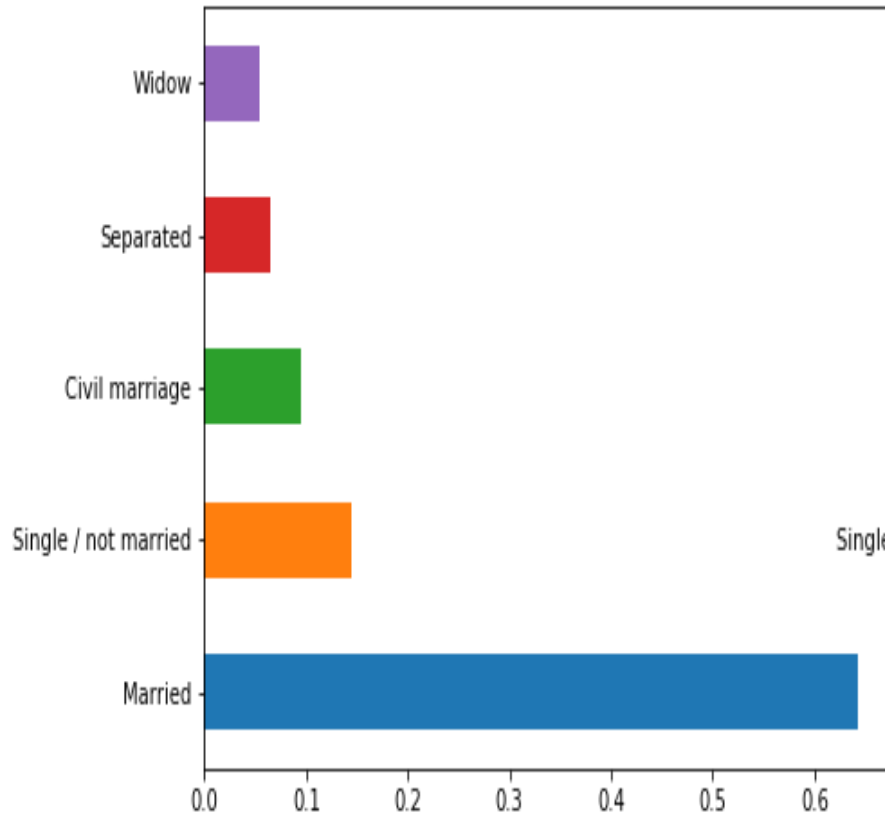
# CLIENT'S FAMILY STATUS WRT TARGET

Insights from the graph :

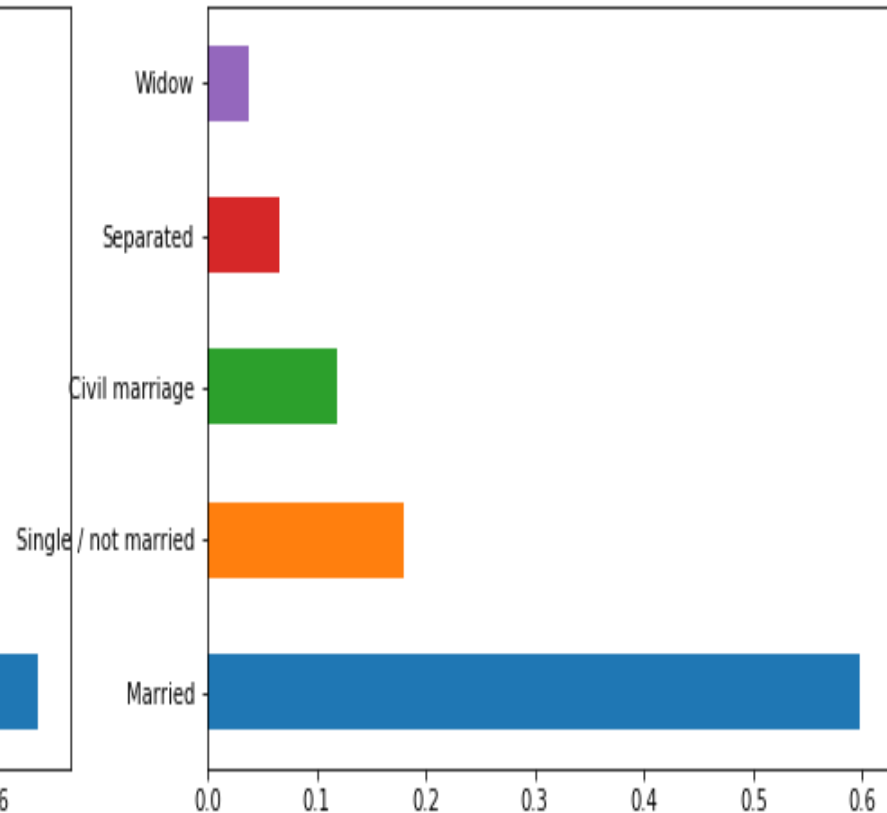
1. Married people are applying for the loans in higher number..

2. Single/Not Married and Civil Marriage people have slightly higher chances of default.

Non-Defaulter



Defaulter



# OCCUPATION TYPE WRT TARGET

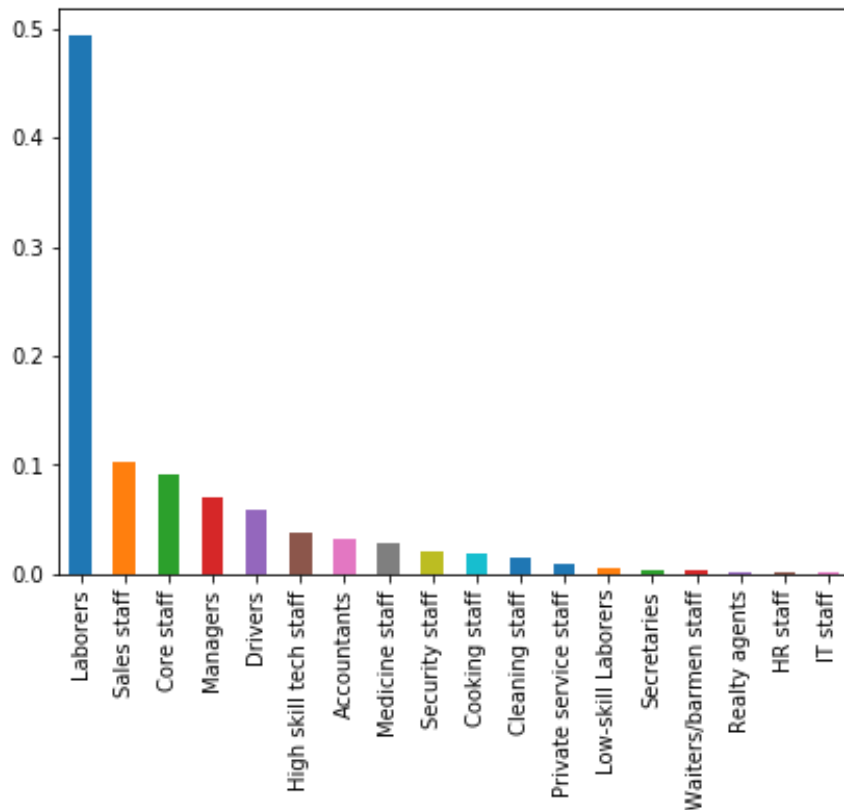
## Insights from the graph :

1. People working in the Laborers category have highest number of loan applications in both cases.

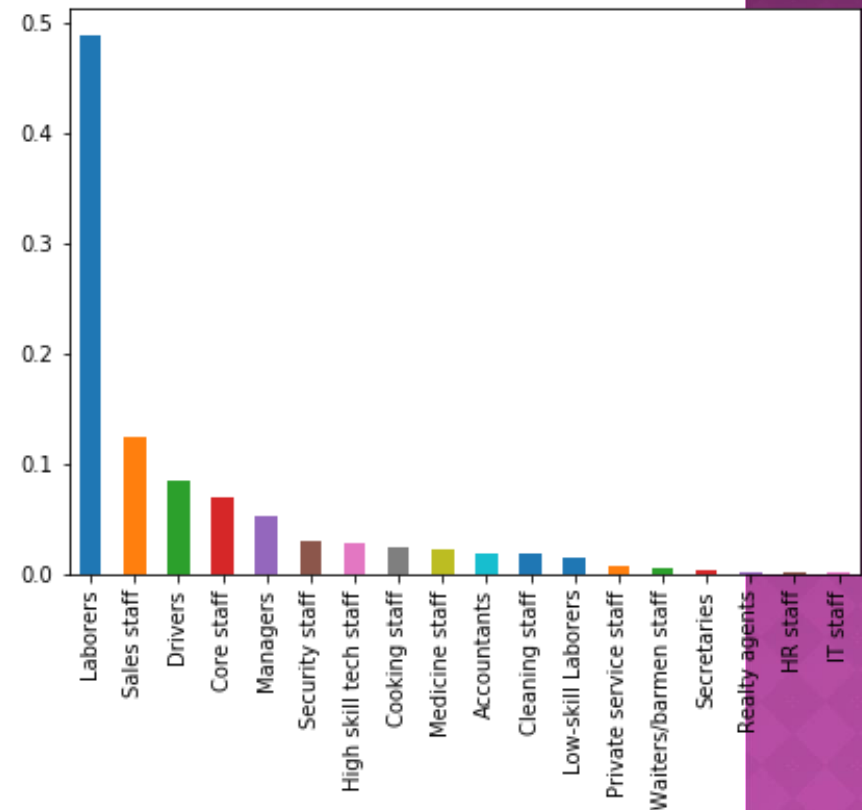
2. Managers have a low chances of loan default.

3. Sales Staff and drivers have a higher chances of loan default.

Non-Defaulter



Defaulter

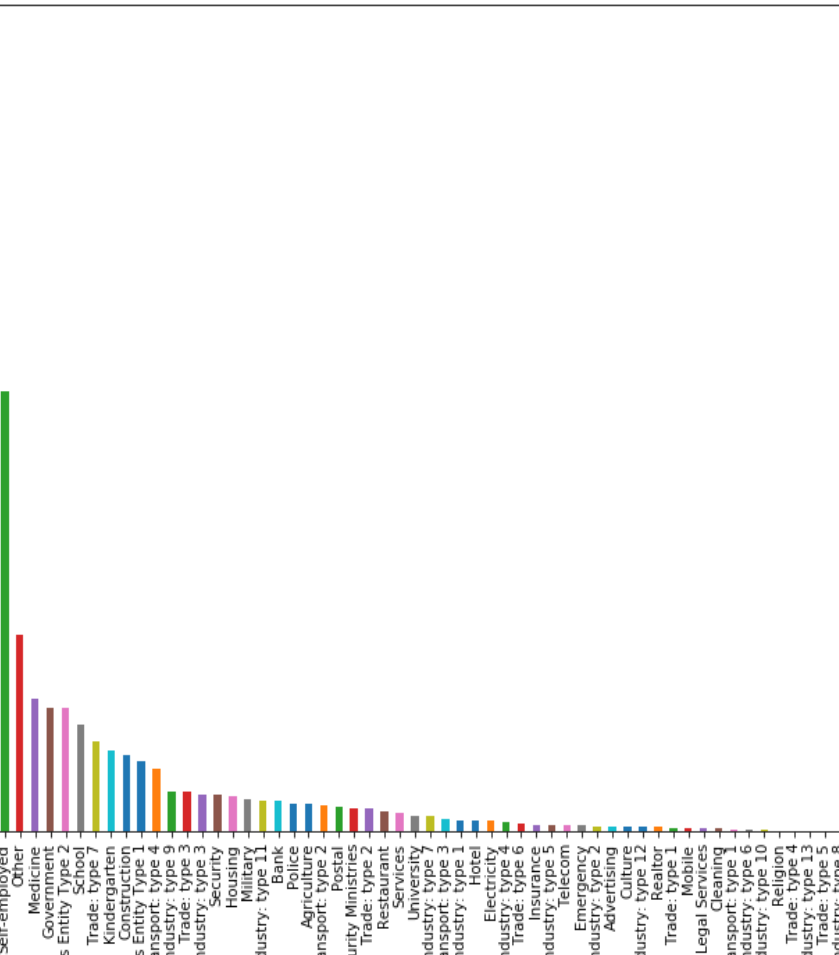


# ORGANIZATION TYPE WRT TARGET

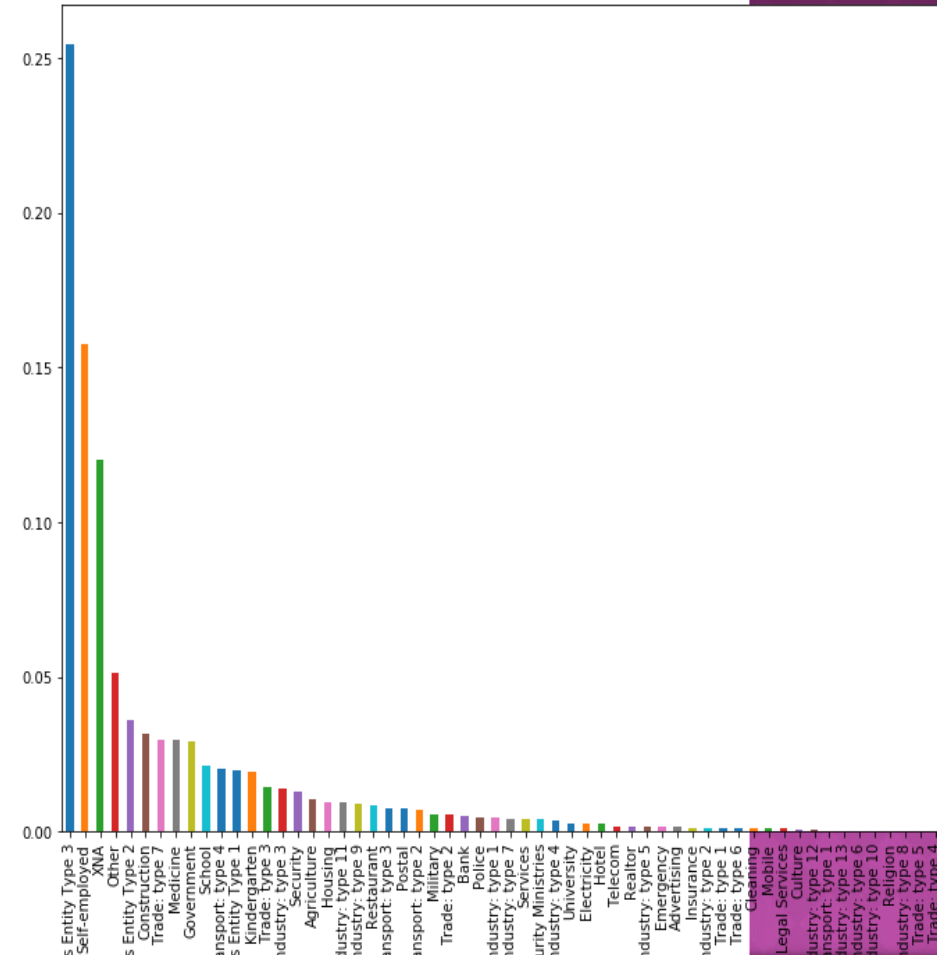
## Insights from the graph :

- Clients with organization type as Business Entity Type 3 and Self Employed are more likely to default.

Non-Defaulter



Defaulter



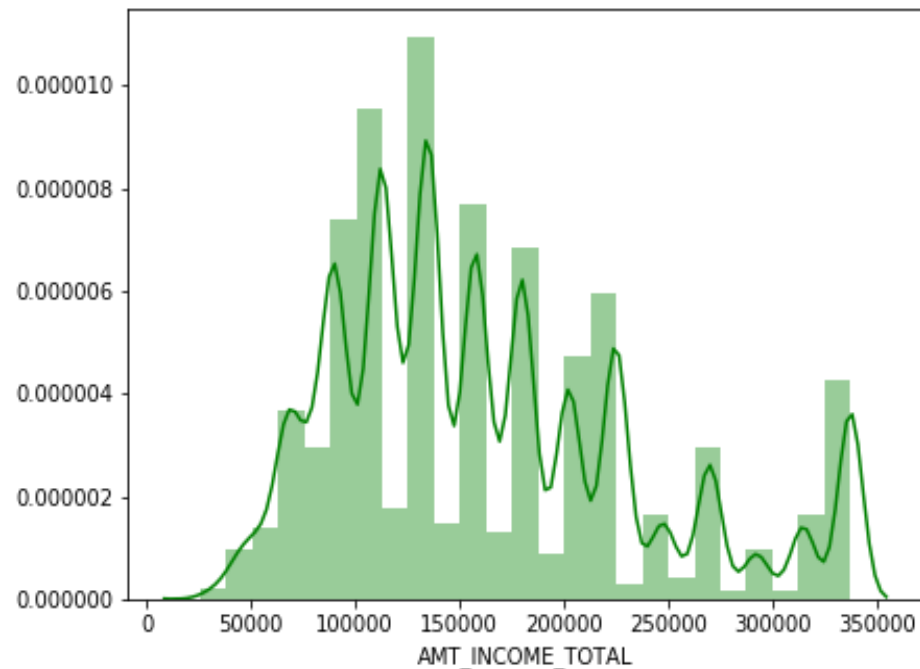
# UNIVARIATE ANALYSIS ON CONTINUOUS COLUMNS IN SEGMENTED APPLICATION DATASET

# INCOME V/S TARGET

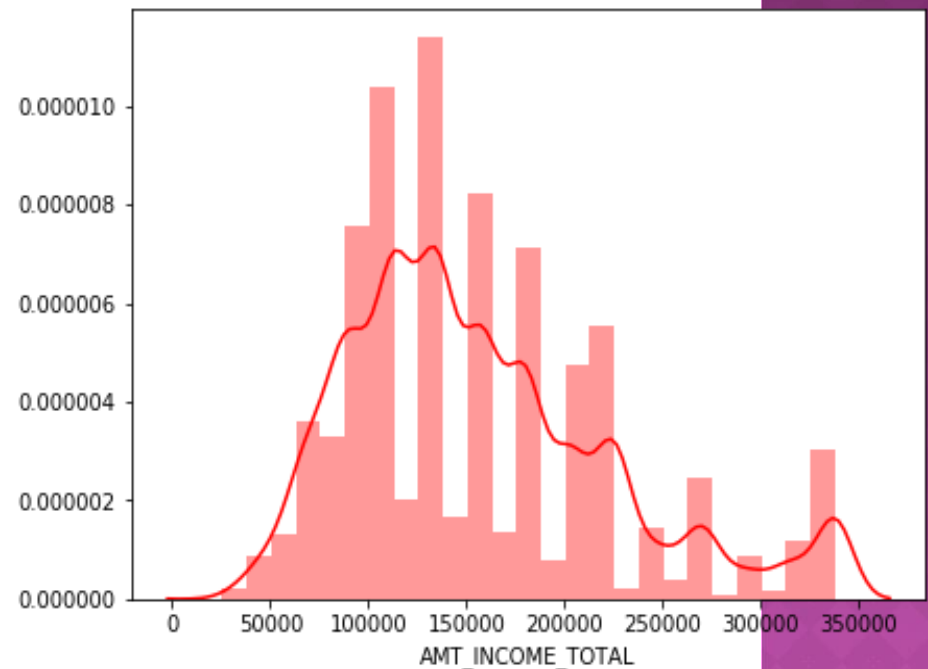
Insights from the graph :

- ❑ People with higher Income Total are less likely to default.

Non-Defaulter



Defaulter

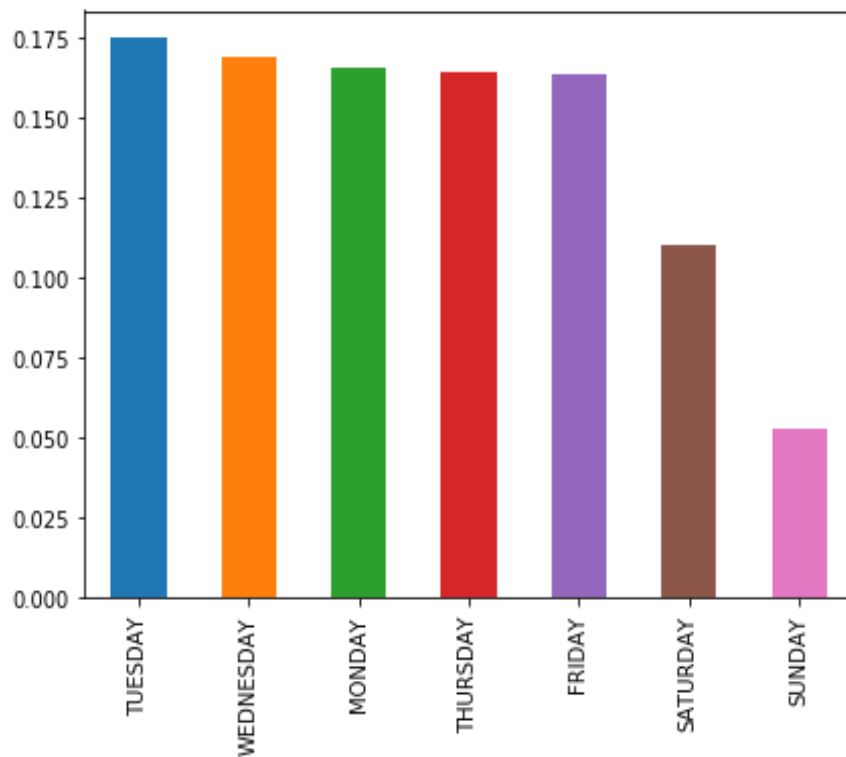


# WEEKDAY PROCESS START WRT TARGET

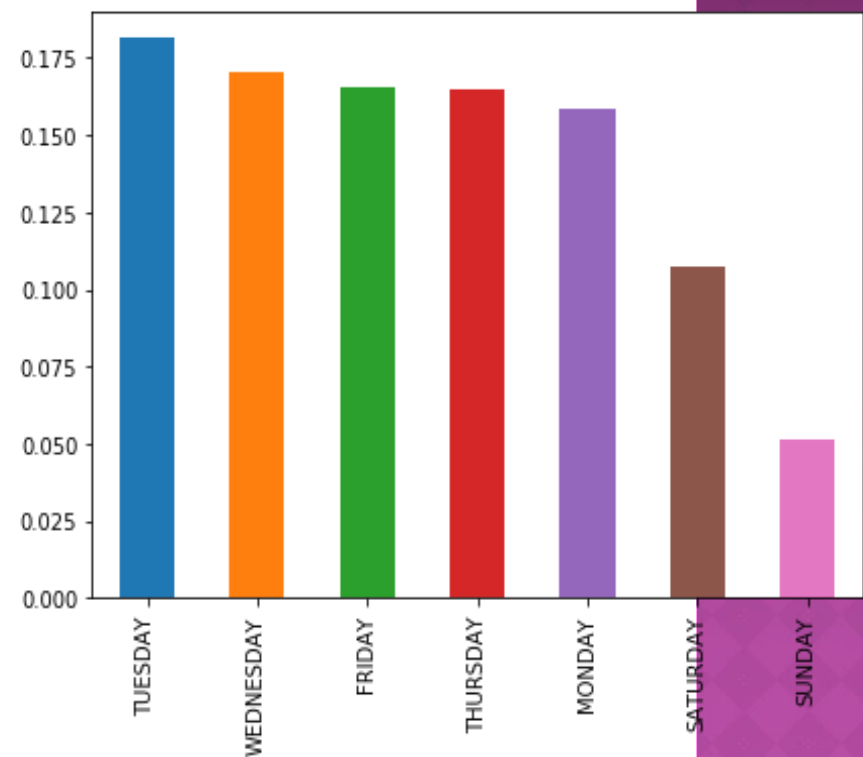
## Insights from the graph :

- ❖ Bank receives more loan applications on weekdays as compared to weekends(Saturday and Sunday).

Non-Defaulter



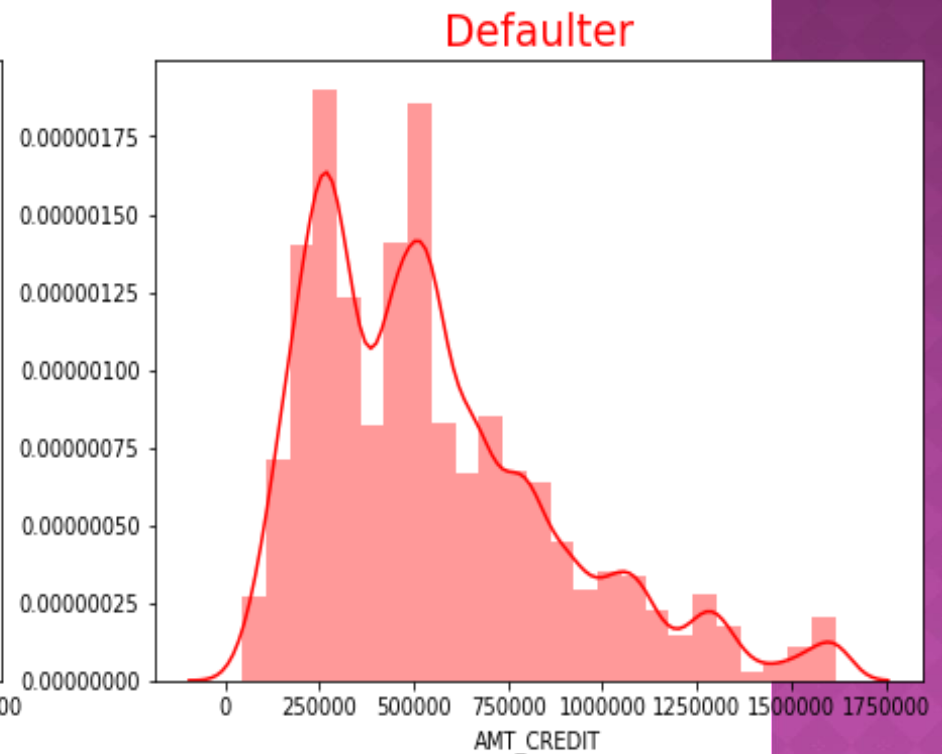
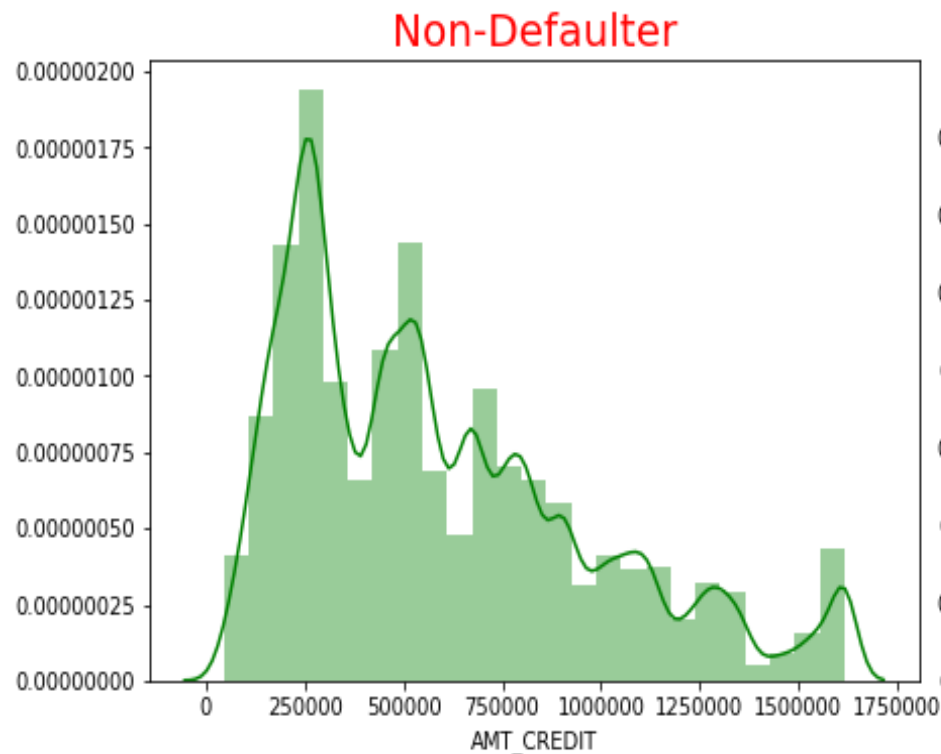
Defaulter



# AMOUNT CREDIT V/S TARGET

Insights from the graph :

- ✓ People with higher Credit amount of loan are less likely to default.





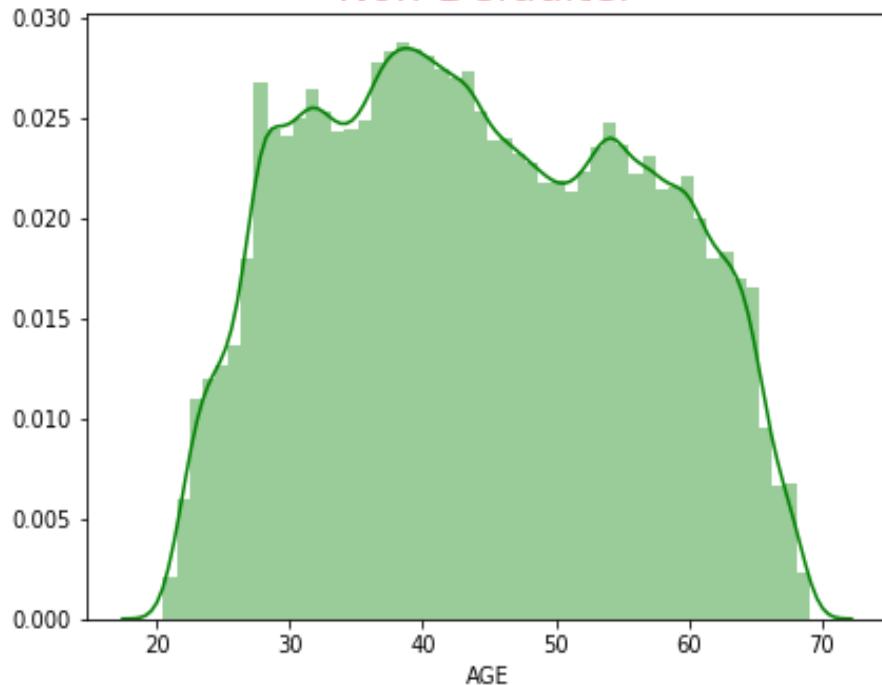
## Insights from the graph :

1. People aged between 25 and 45 are the ones with high number of loan applications.

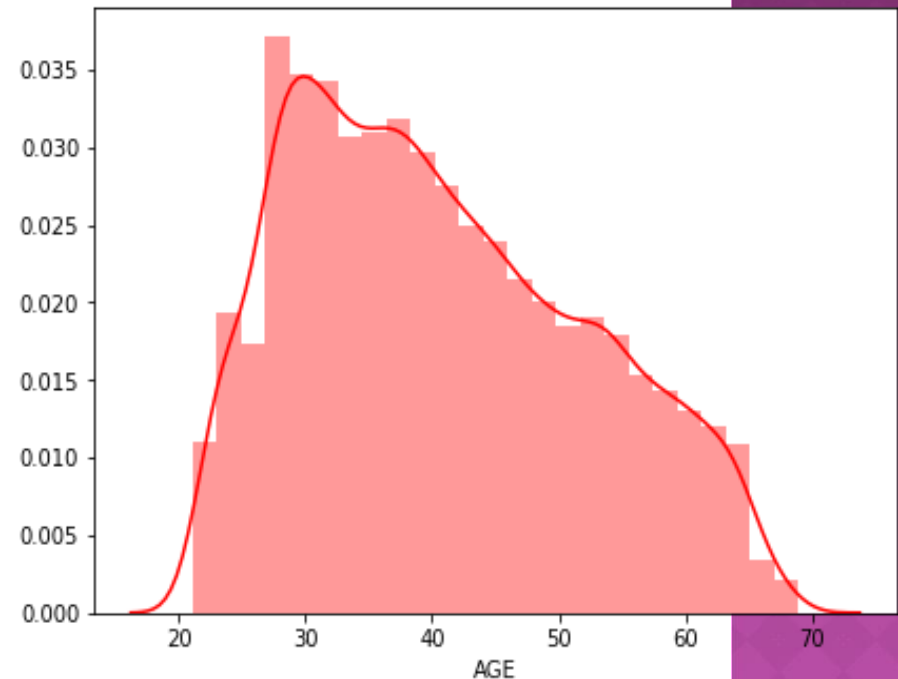
2. People above 50 years of age are less likely to default on their loan.

## AGE V/S TARGET

Non-Defaulter



Defaulter



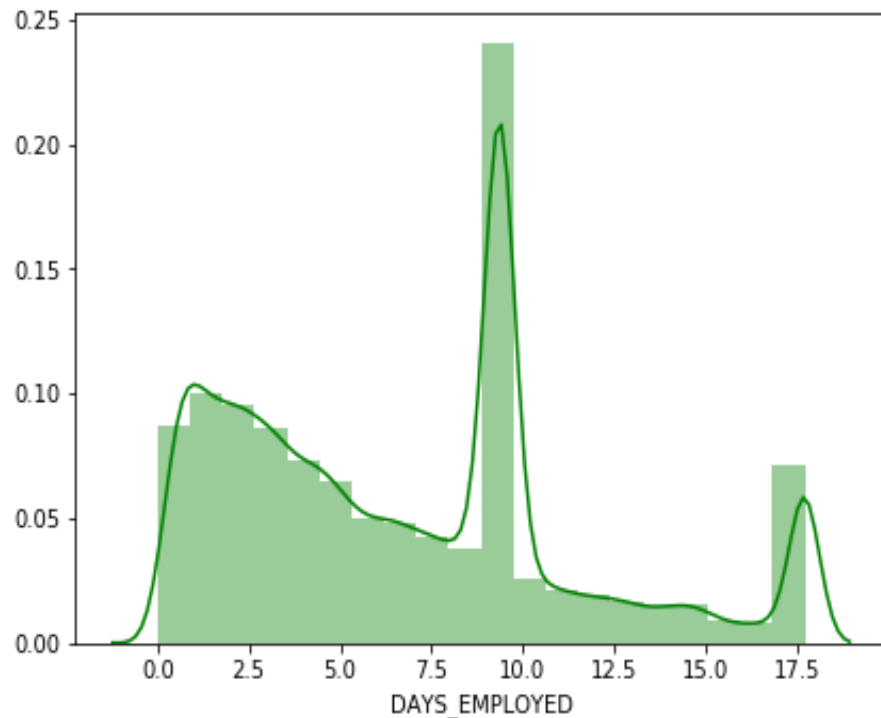
# DAYS\_EMPLOYED V/S TARGET

## Insights from the graph :

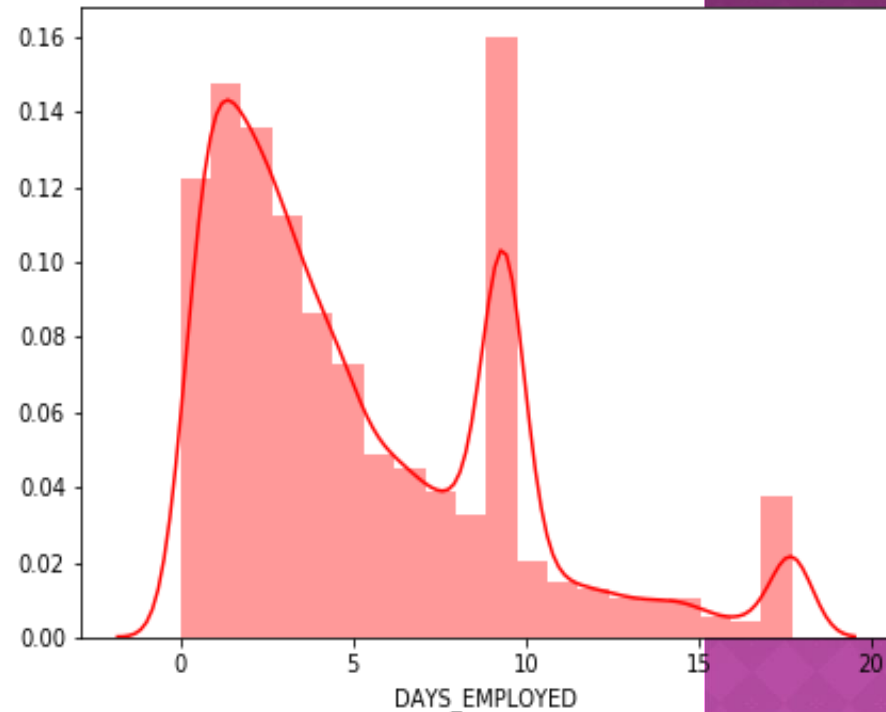
1. People who newly got employed into their current employment (less than 4 years) are the highest among loan defaulters.

2. People working for a long time into their current employment (more than 10 years) are less likely to default.

Non-Defaulter



Defaulter

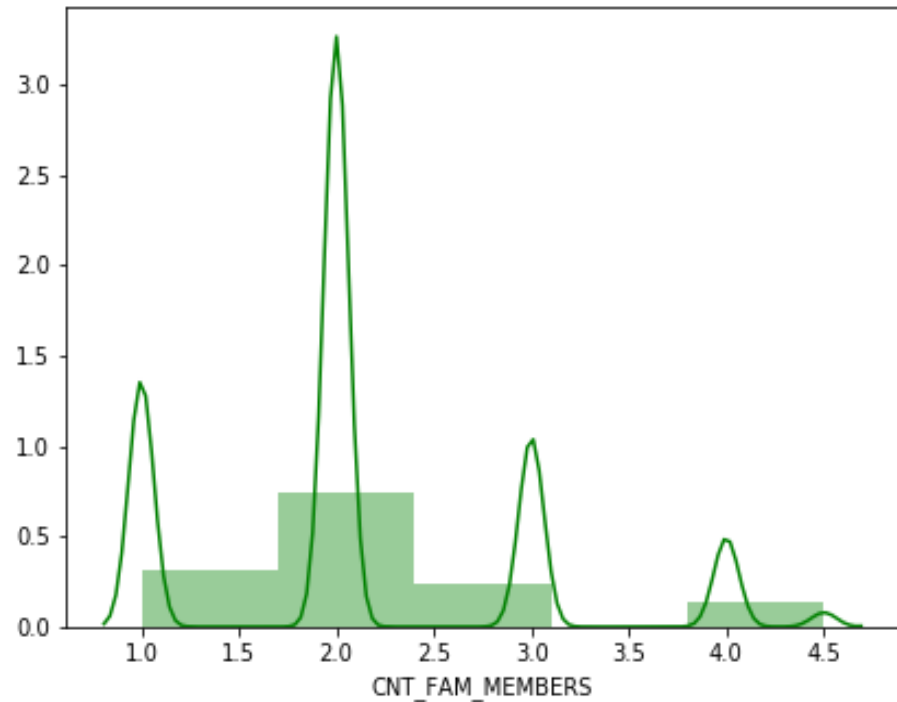


# CNT\_FAM\_MEMBERS V/S TARGET

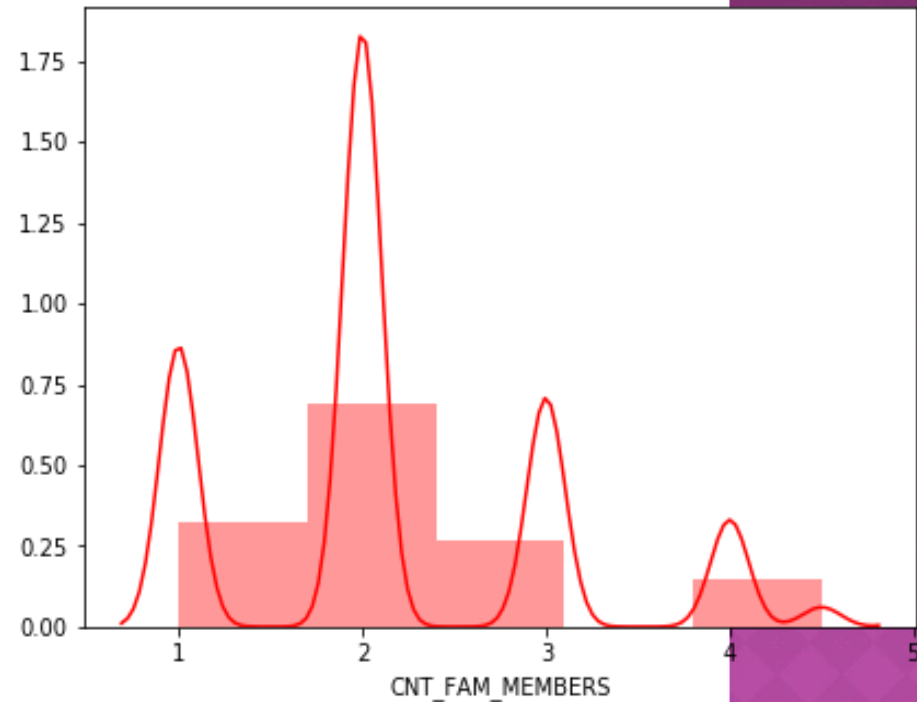
## Insights from the graph :

- People from smaller families have more number of loan applications.

Non-Defaulter



Defaulter

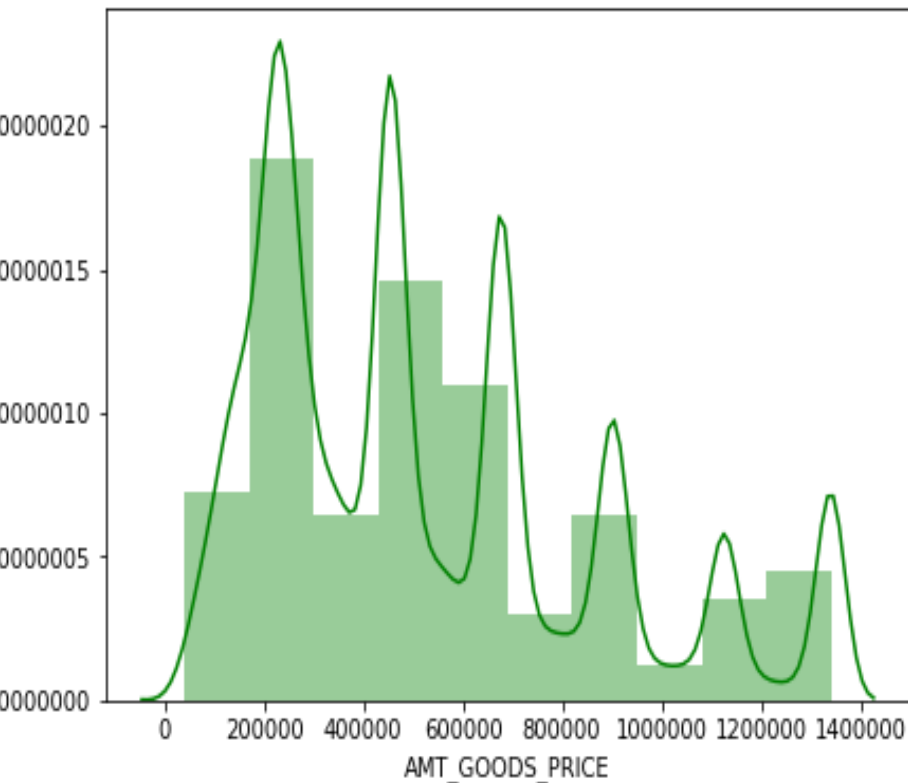


# AMT\_GOODS\_PRICE V/S TARGET

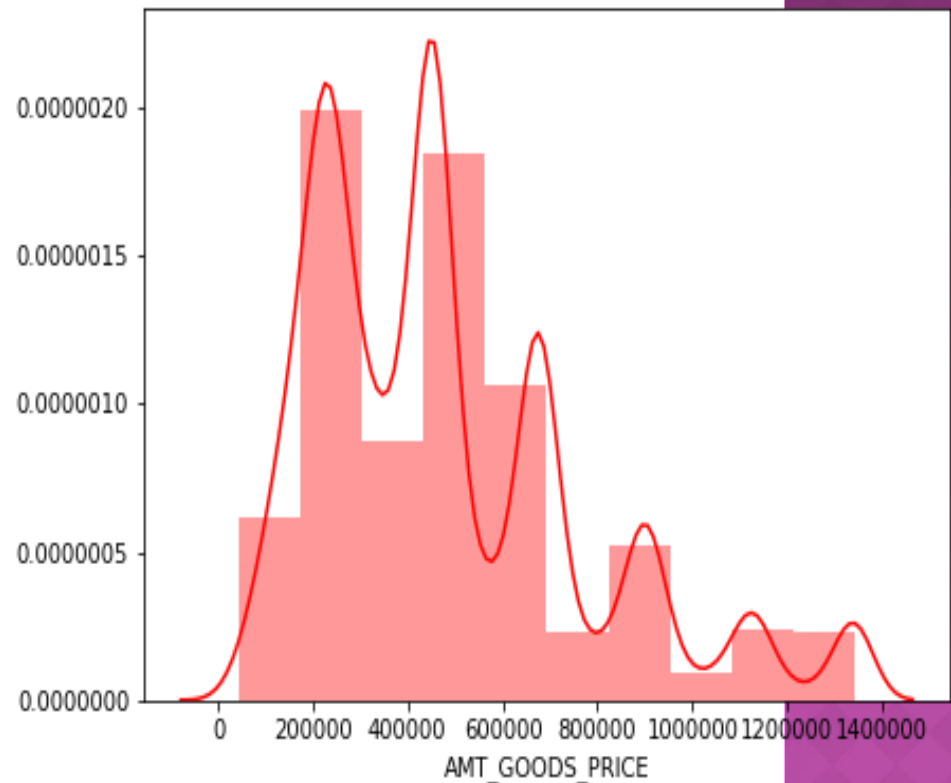
Insights from the graph :

- Higher the price of goods , lesser is the chance of default.

Non-Defaulter



Defaulter

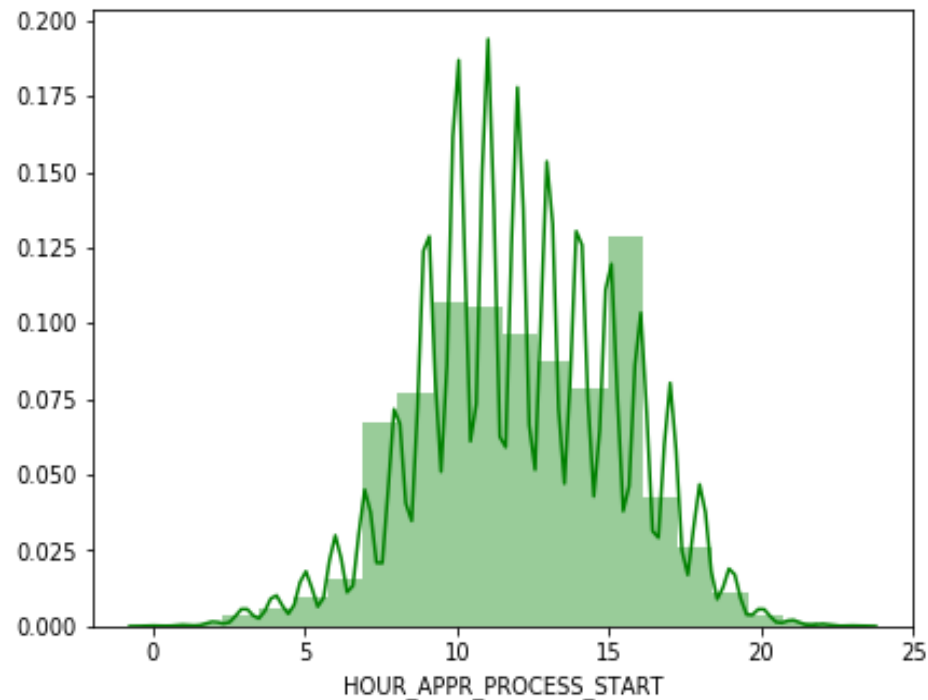


## HOUR\_APPR\_PROCESS\_START V/S TARGET

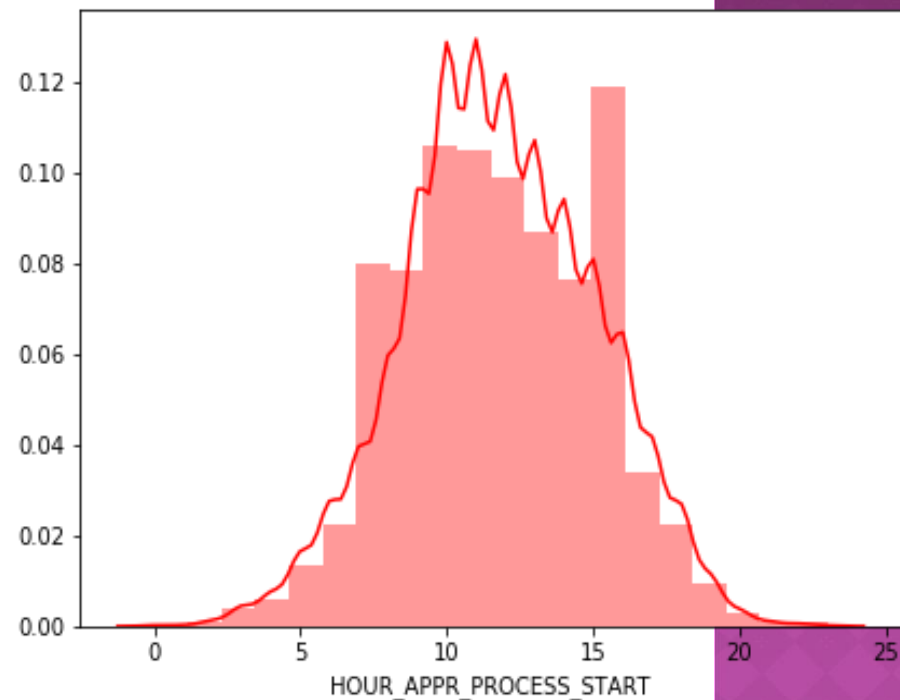
### Insights from the graph :

- Bank received highest number of loan applications between 10 o'clock in the morning to 2 o'clock in the afternoon.

Non-Defaulter



Defaulter

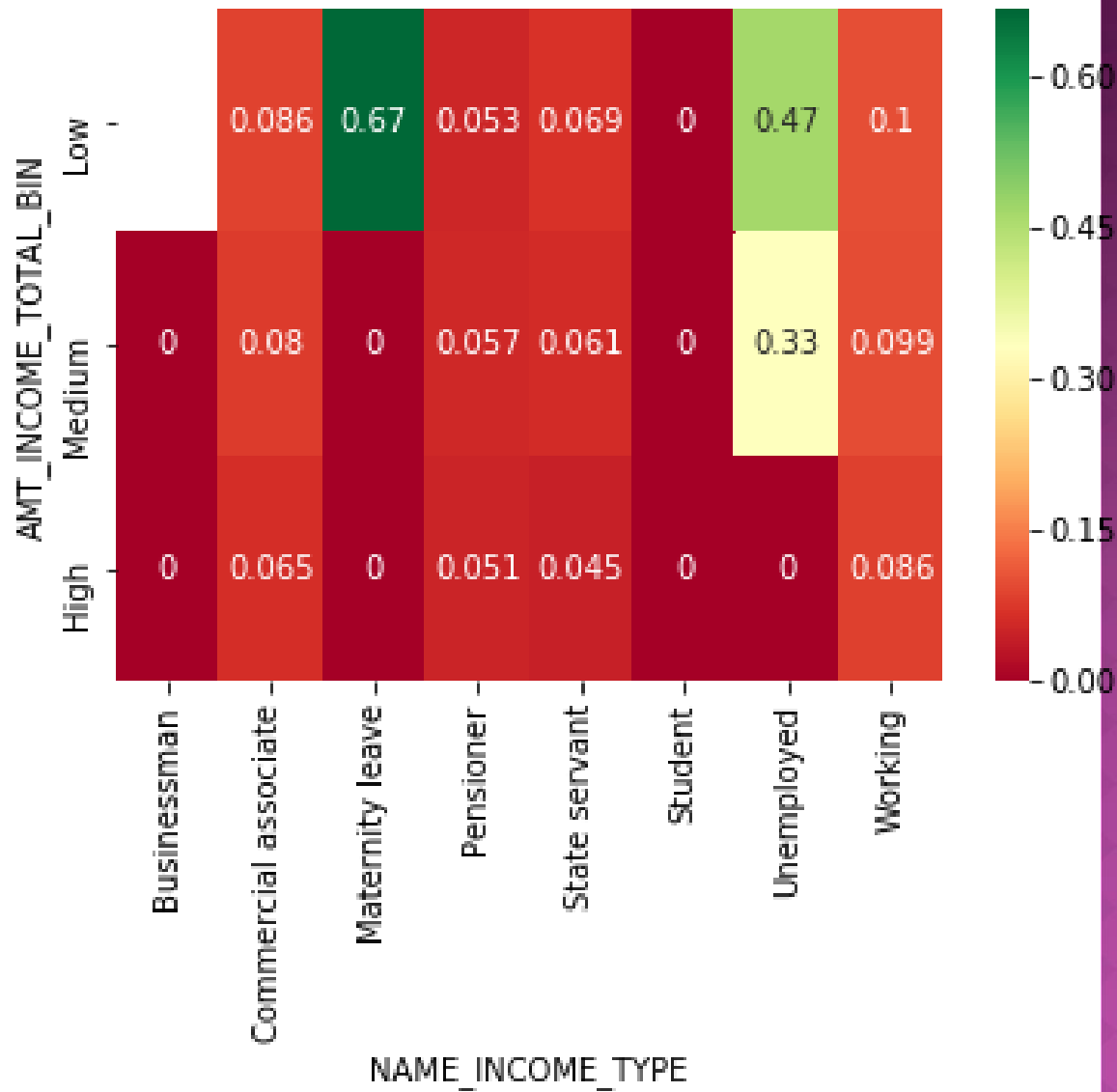


# CORRELATION GRAPH

Insights from the graphs :

1. Clients with income type as maternity leave and having low income are more likely to default.

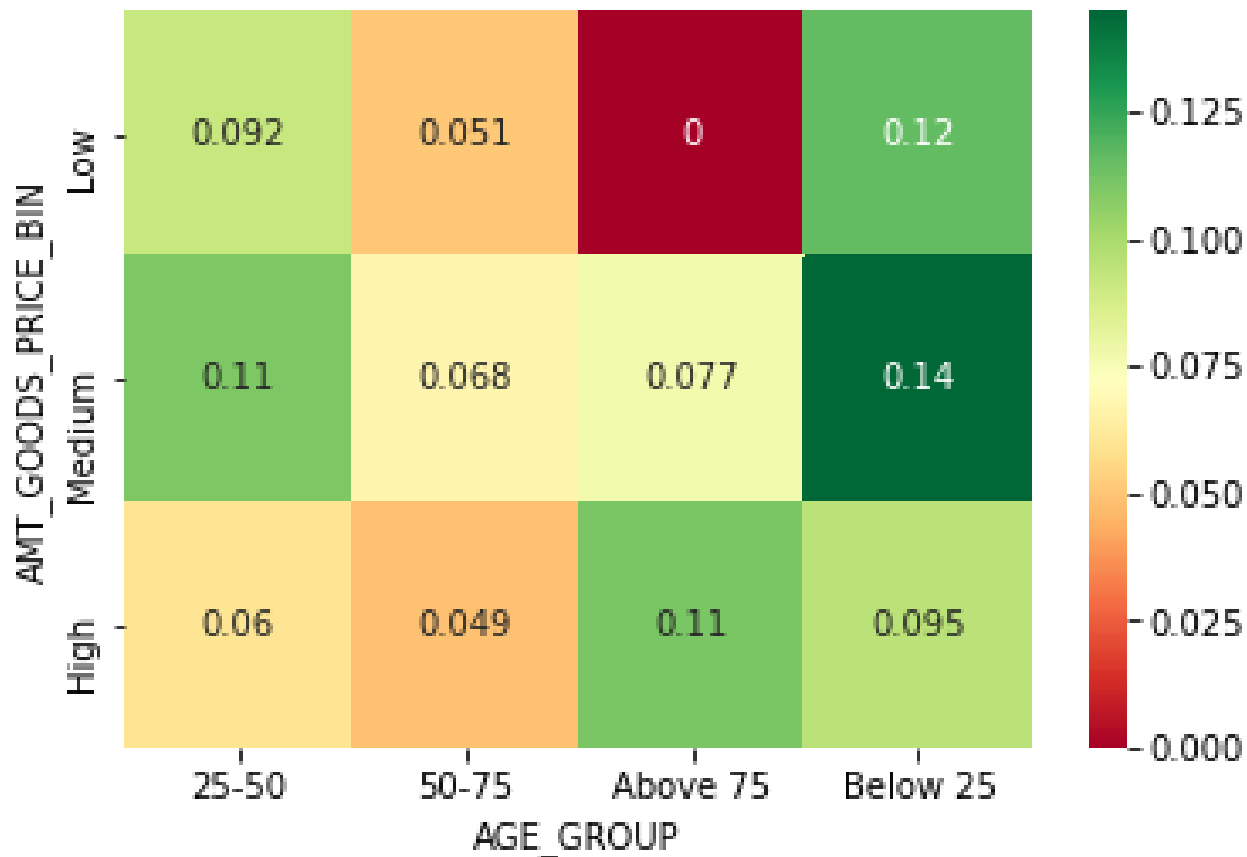
2. Clients who are unemployed are more likely to default.



# CORRELATION GRAPH

Insights from the above graphs :

- Clients below 25 years of age with medium Goods Price are more likely to default.



# ANALYSIS OF CATEGORICAL COLUMNS IN PREVIOUS DATA SET

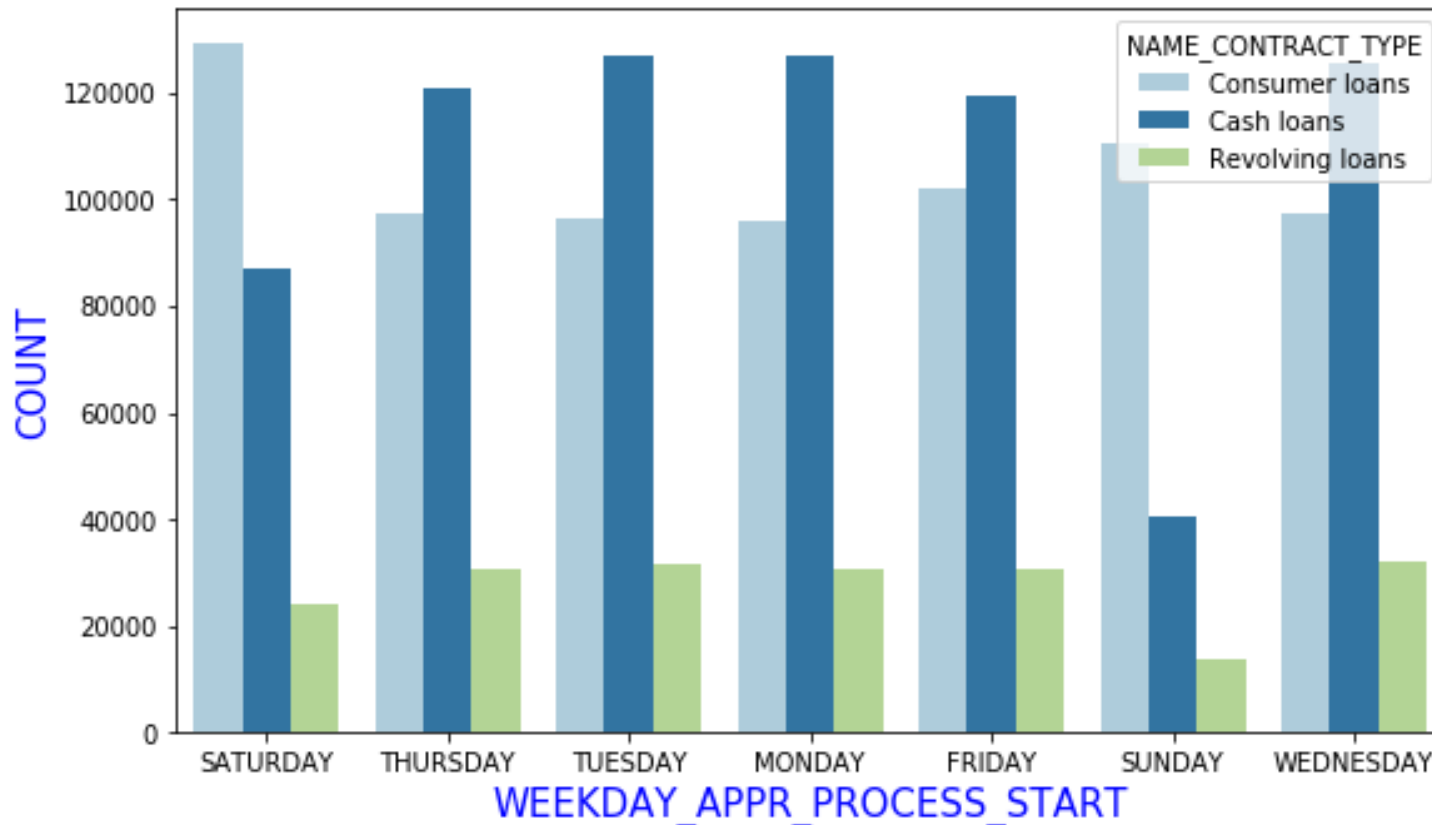


# COUNT PLOT

## WEEKDAY\_APPR\_PROCESS\_START V.S NAME\_CONTRACT\_TYPE

Insights from the graph:

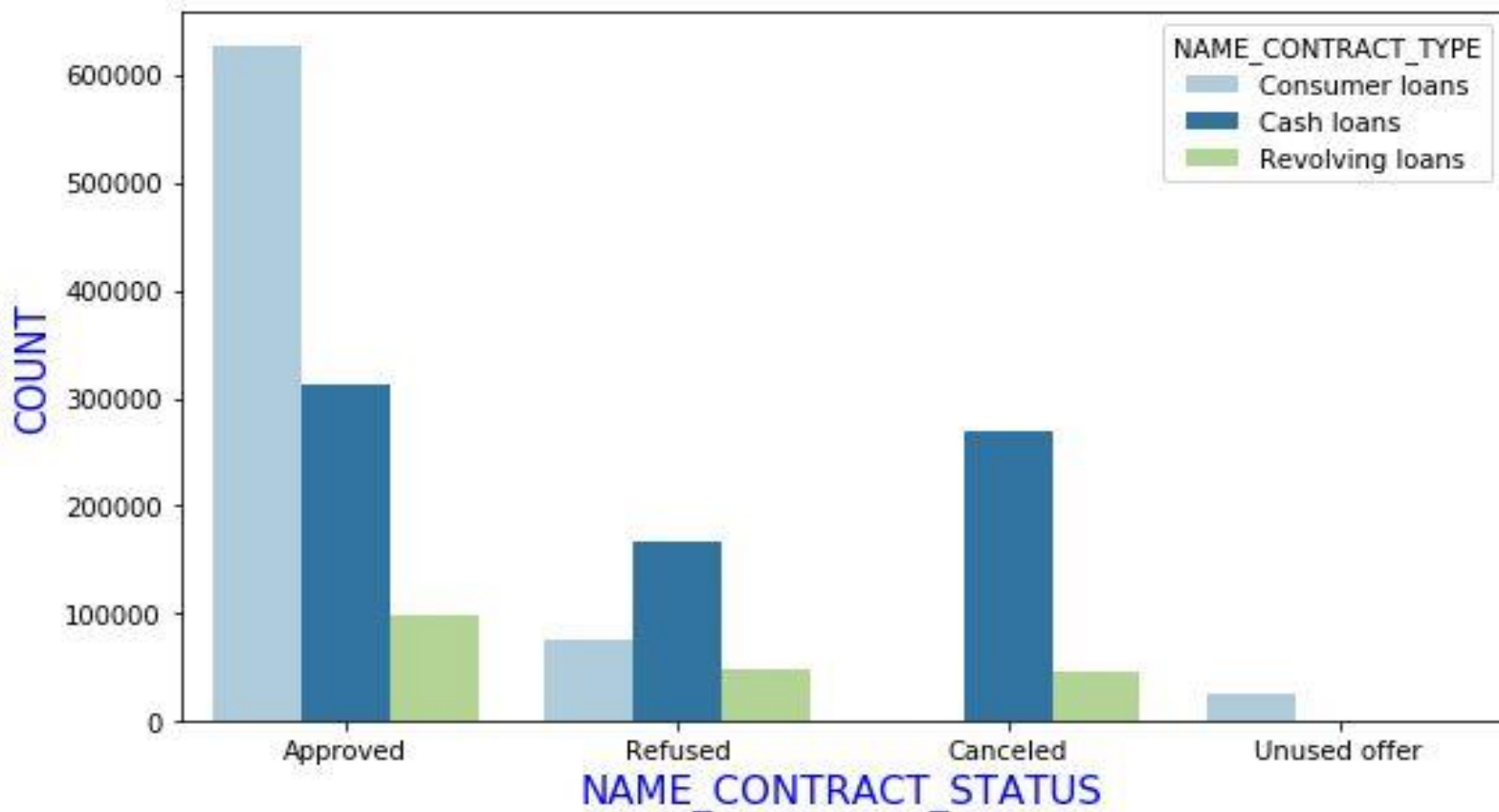
- Bank received relatively higher consumer loan applications on Saturdays and Sundays.



# COUNT PLOT NAME\_CONTRACT\_STATUS V.S NAME\_CONTRACT\_TYPE

Insights from the graph :

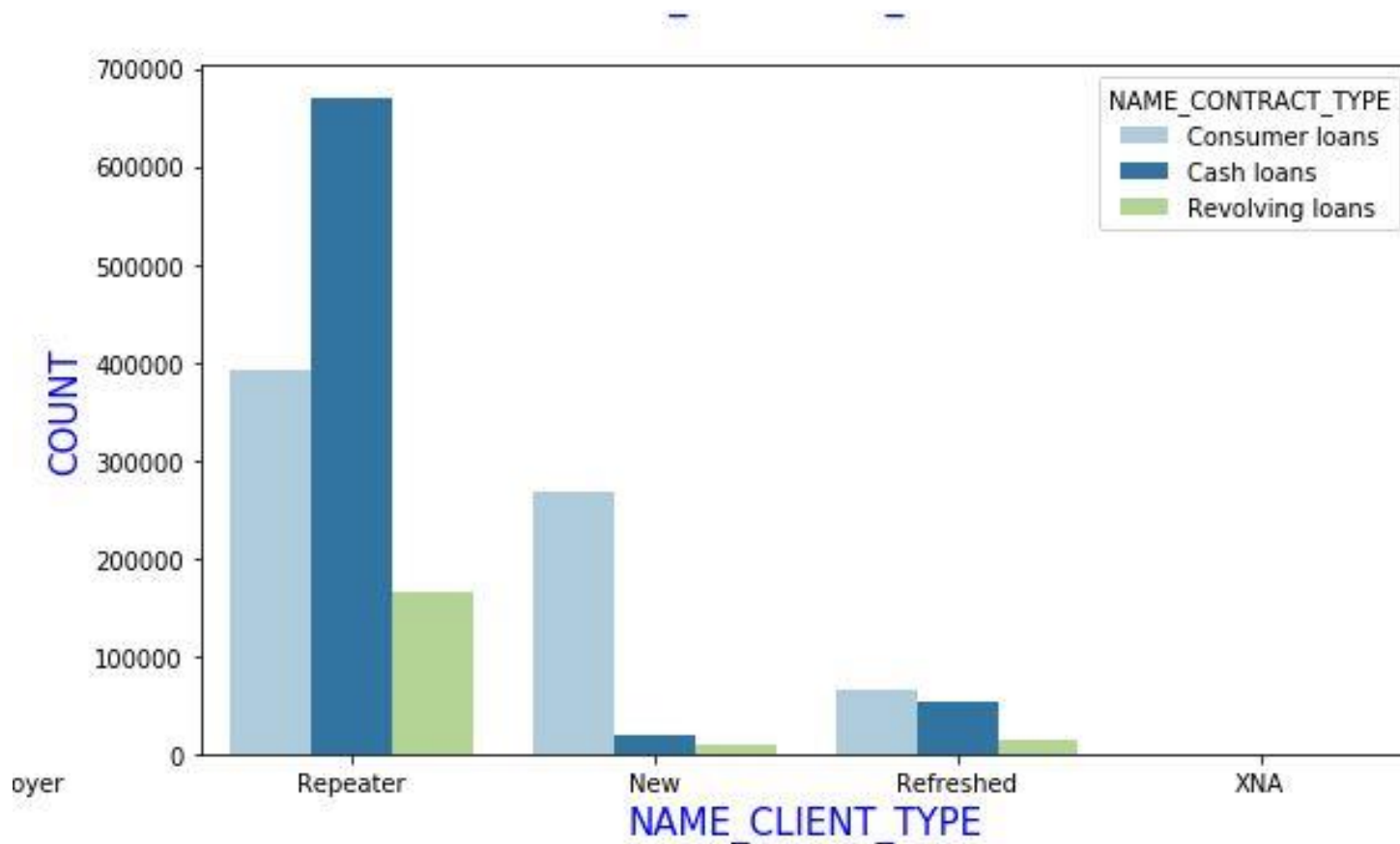
- Consumer loans are highest to be approved followed by cash loans and revolving loans.



# COUNT PLOT ON NAME\_CLIENT\_TYPE V.S NAME\_CONTRACT\_TYPE

Insights from the graph :

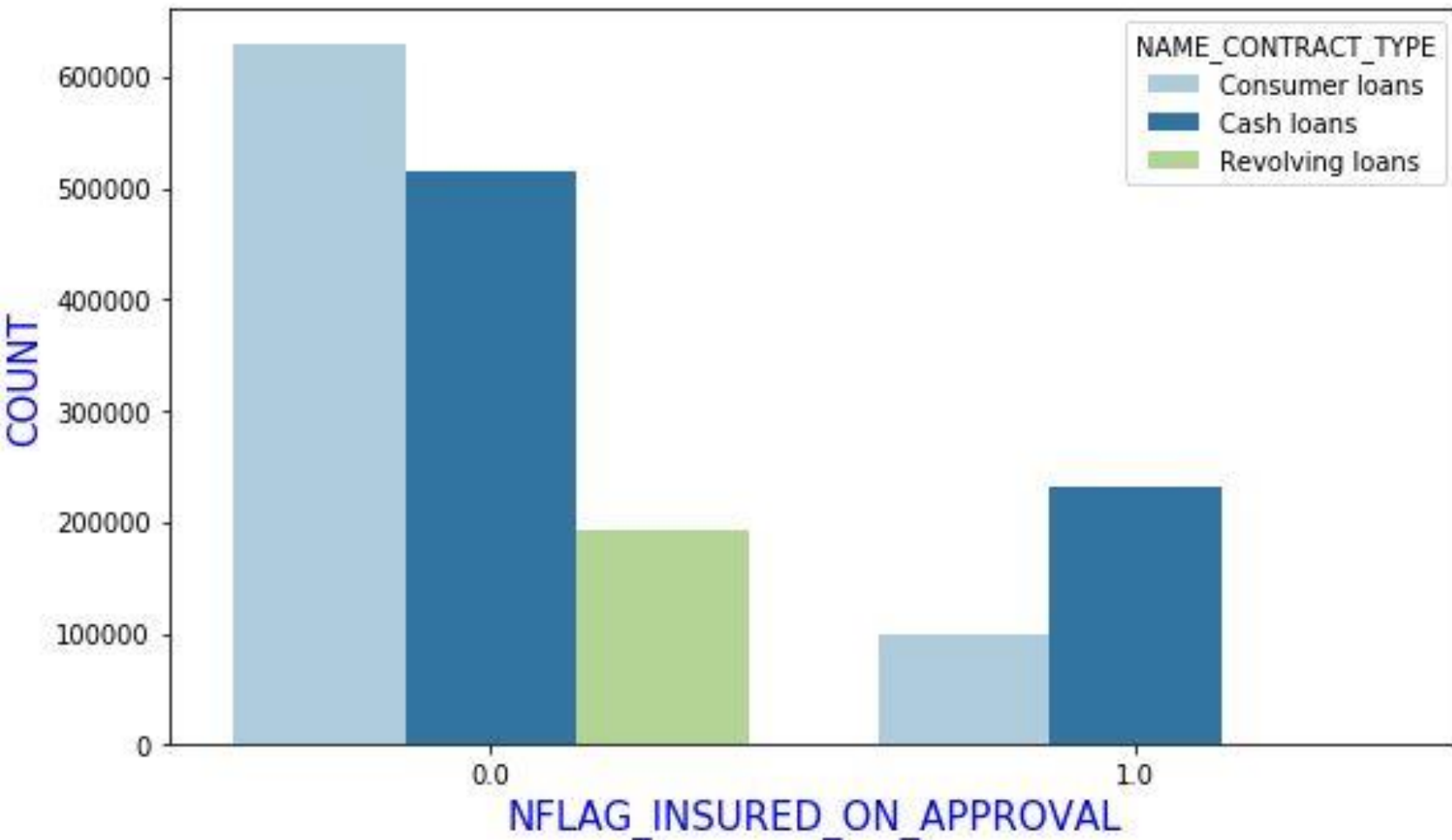
- Repeater Clients are applying for cash loans in comparatively higher number. New clients are opting for consumer loans in higher number.



# COUNT PLOT INSURED\_ON\_APPROVAL V.S NAME\_CONTRACT\_TYPE

Insights From The Graph :

- Majority of clients do not prefer to apply for the insurance.

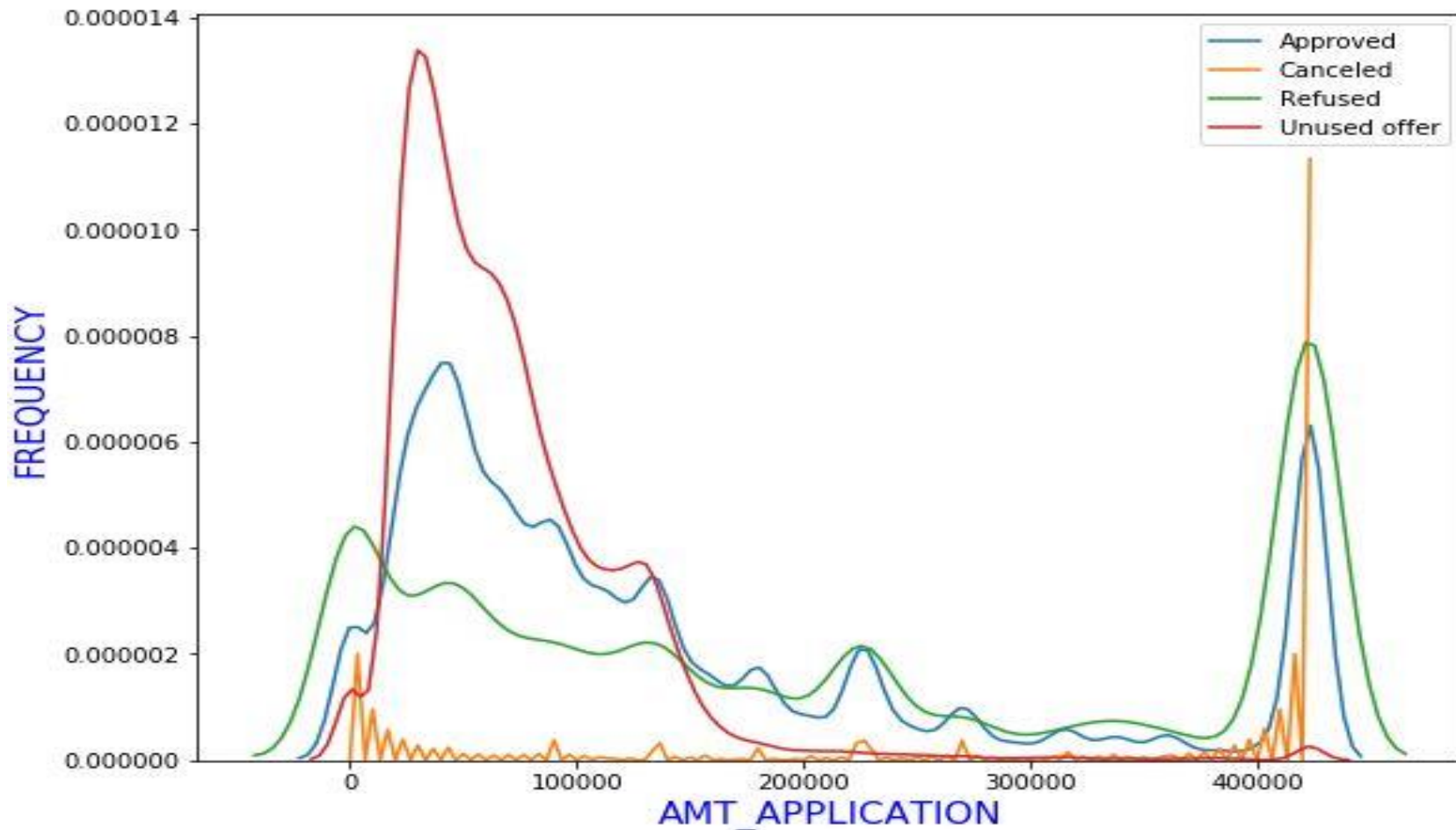


# ANALYSIS OF CONTINUOUS COLUMNS IN PREVIOUS DATA SET

# DISTPLOT AMT\_APPLICATION V.S LOAN STATUS

Insights From The Graph :

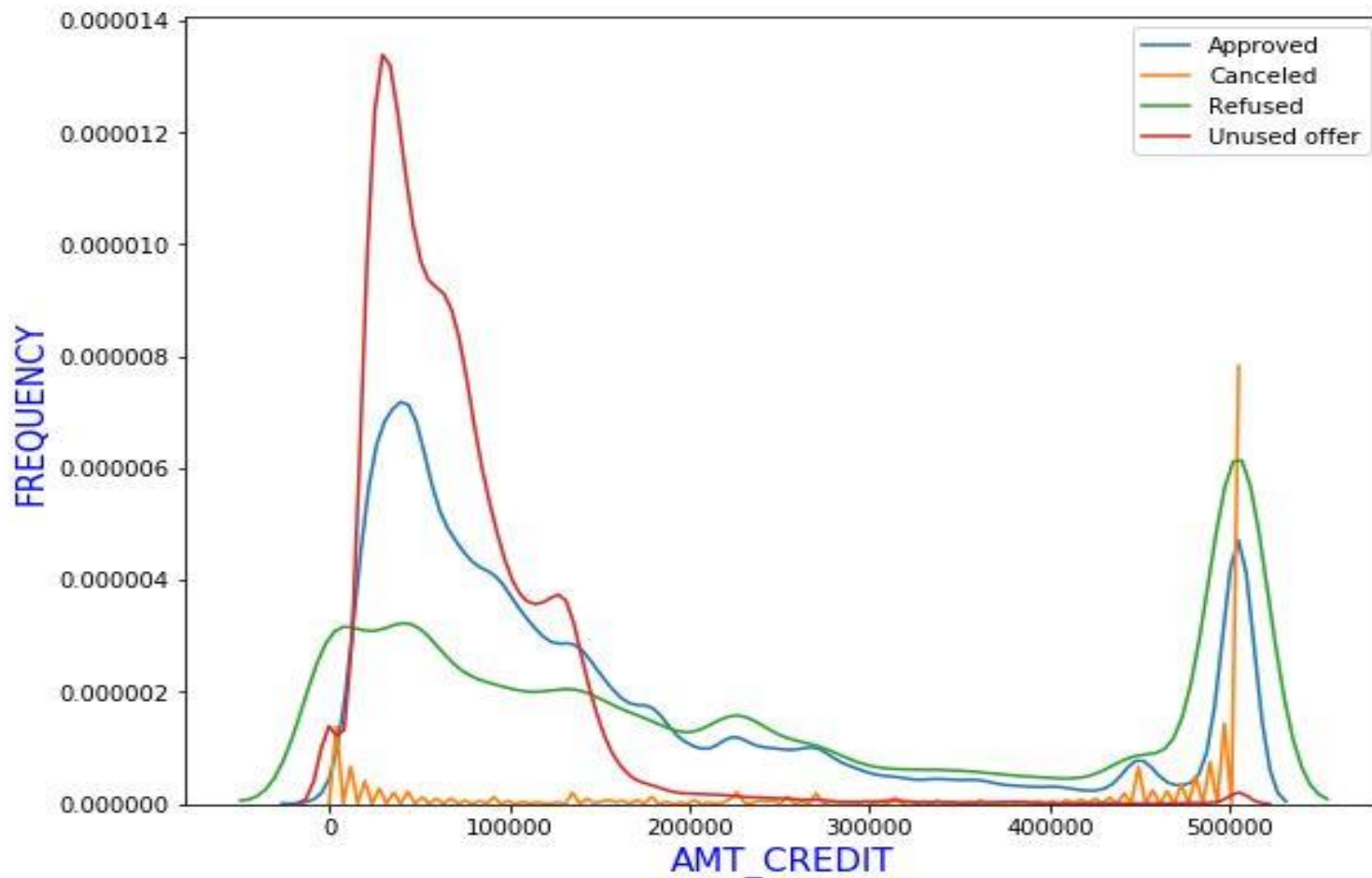
- ❖ With high application amount comes high chances of refusal.



# DISTPLOT AMT\_CREDIT V.S LOAN STATUS

## Insights From The Graph :

- AMT\_CREDIT below 10000 has the highest chances of approval.

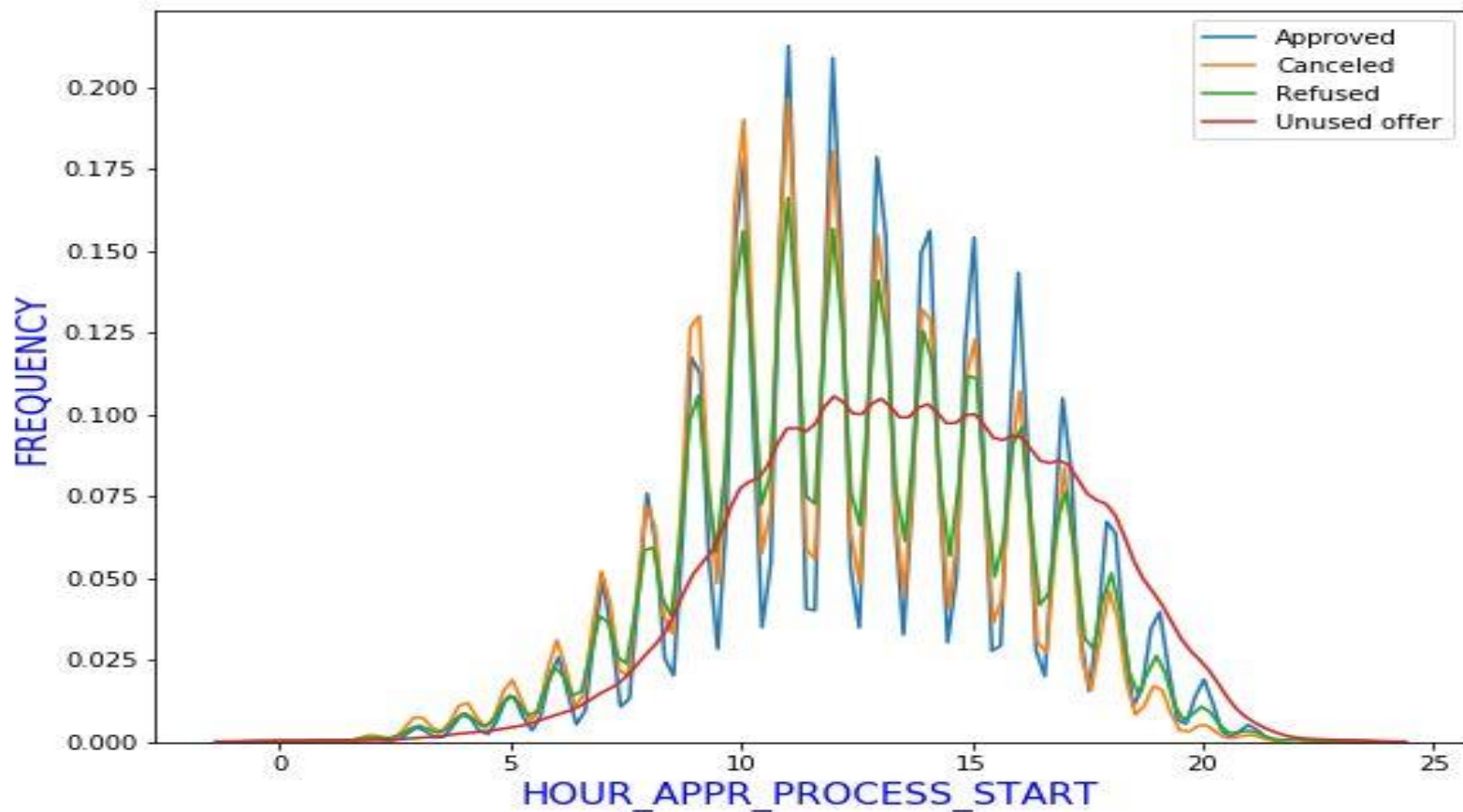


# DISTPLOT

## HOUR\_APPR\_PROCESS\_START V.S LOAN STATUS

Insights From The Graph :

- ❖ Bank receives highest number of loan applications between 10 A.M to 2 P.M.





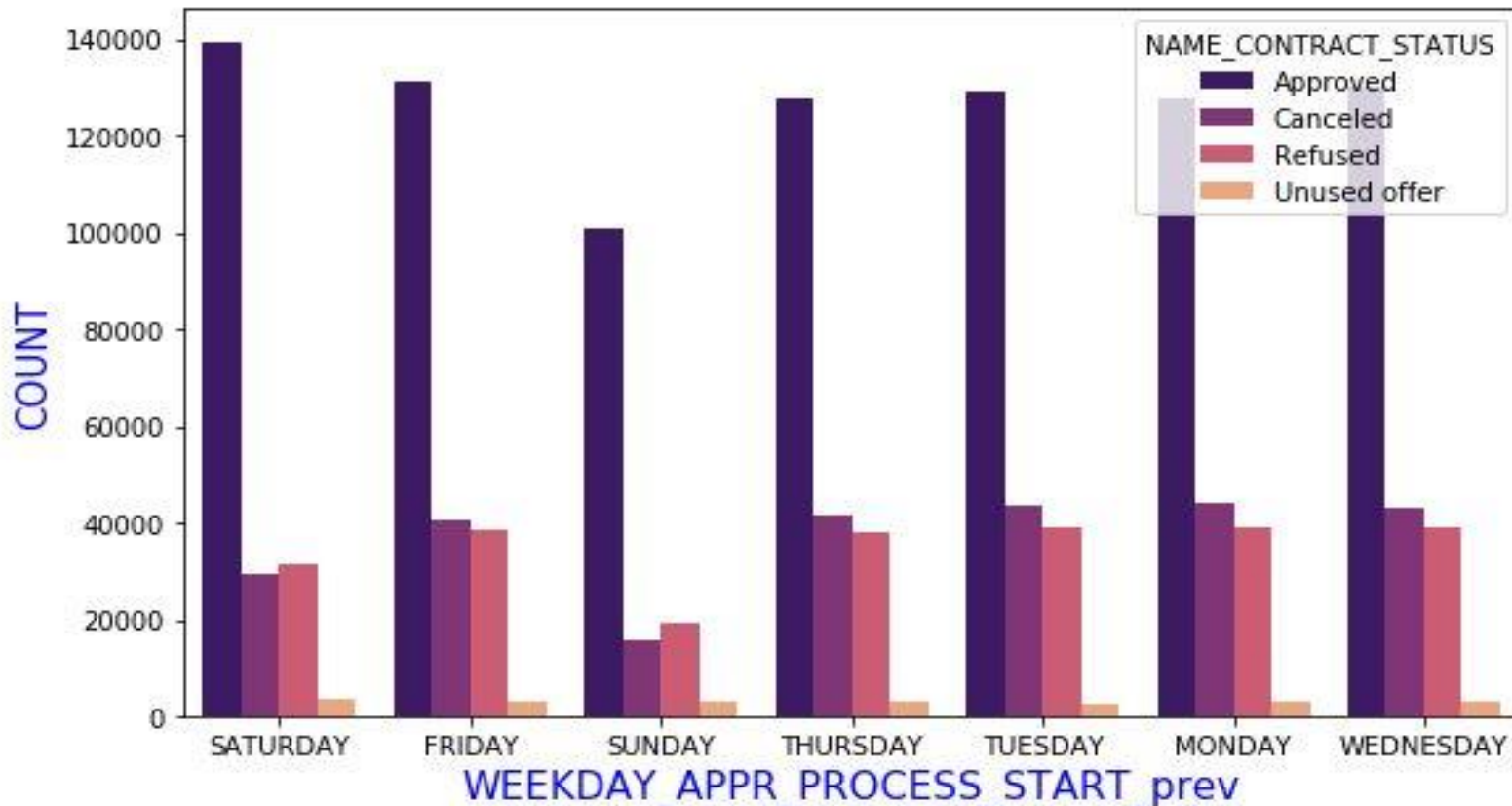
# ANALYSIS OF CATEGORICAL COLUMNS IN MERGED DATA SET

# COUNT PLOT

## WEEK\_APPR\_PROCESS\_START V.S LOAN STATUS

Insights From The Graph :

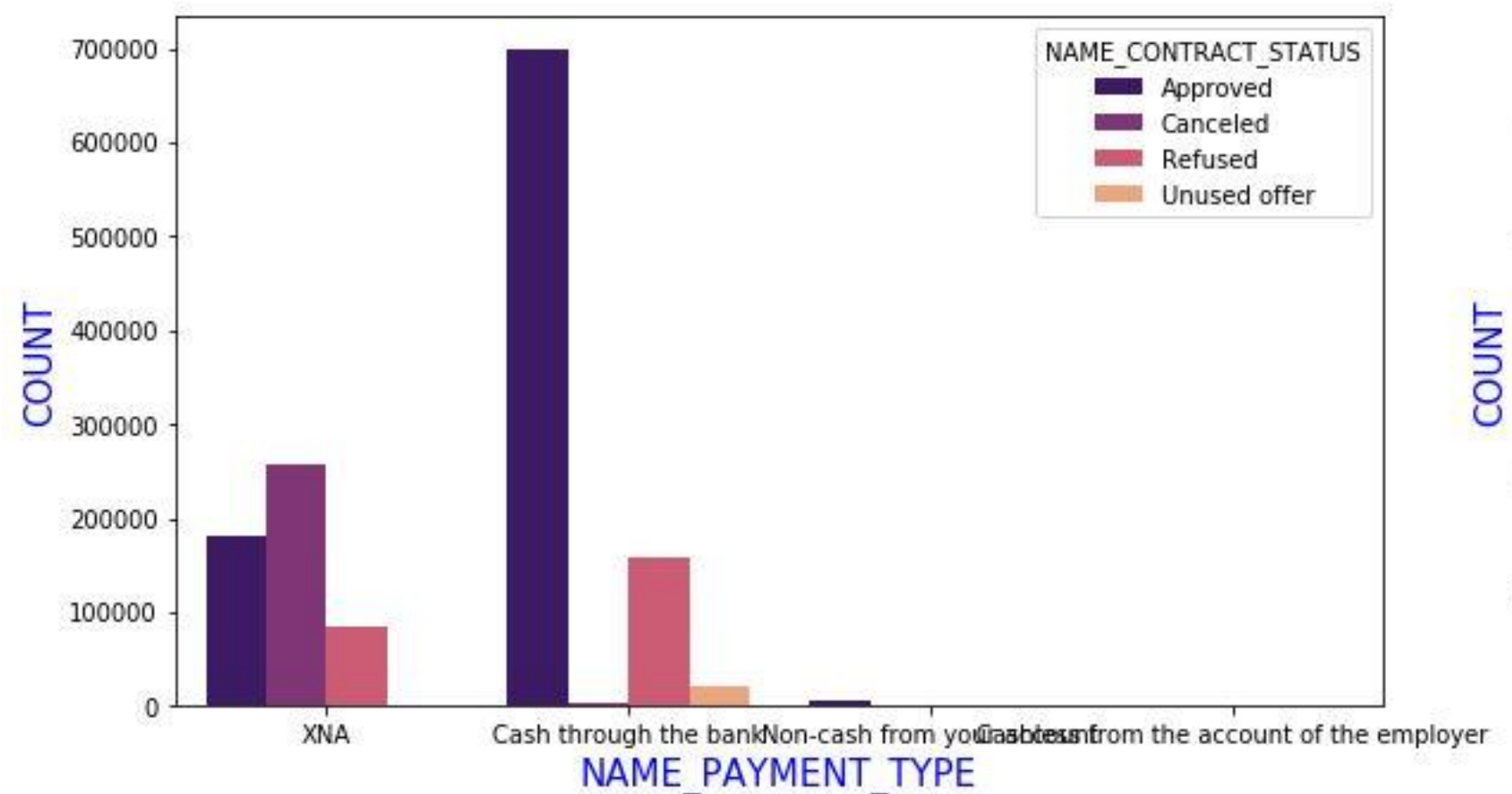
Chances of refusal and cancellations are less on Saturdays.



# COUNT PLOT NAME\_PAYMENT\_TYPE V.S LOAN STATUS

Insights From The Graph :

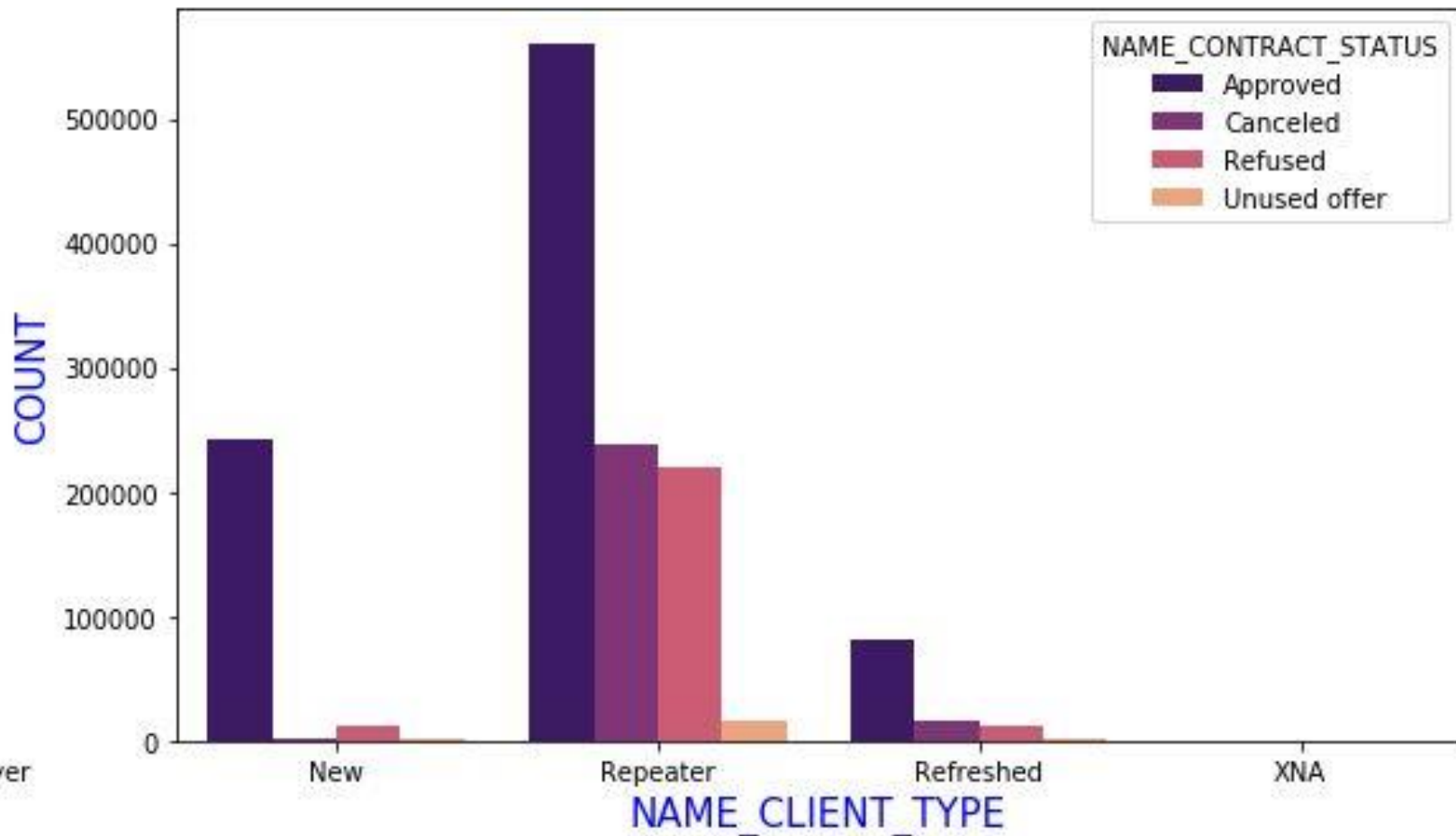
Payment type Cash through bank has a high chance of approval.



# COUNT PLOT NAME\_CLIENT\_TYPE V.S LOAN STATUS

Insights From The Graph :

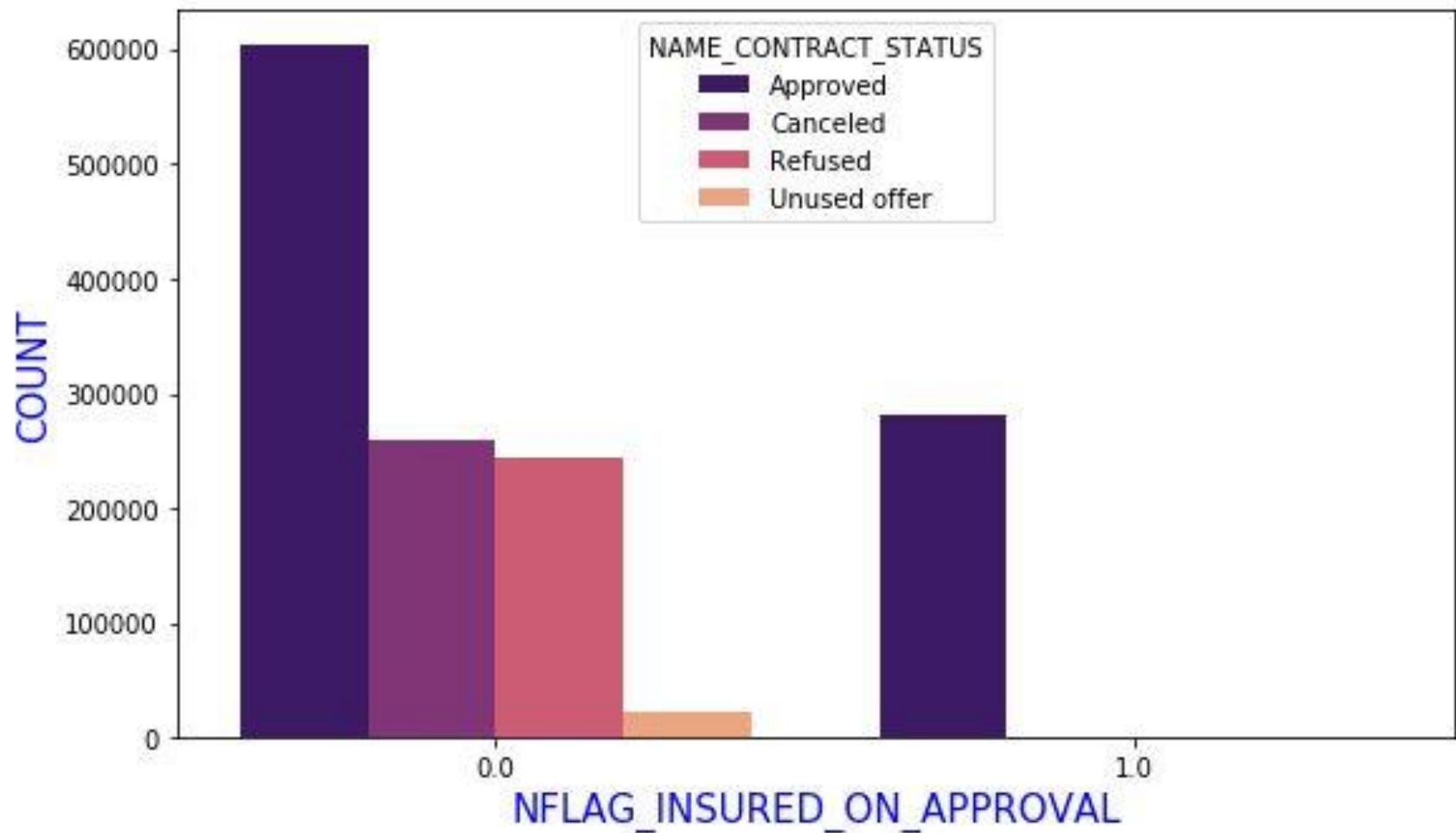
Repeater clients have a very high chance of approval.



# COUNT PLOT INSURED\_ON\_APPROVAL V.S LOAN STATUS

Insights From The Graph :

Clients who have applied for insurance have very low chances of refusal.

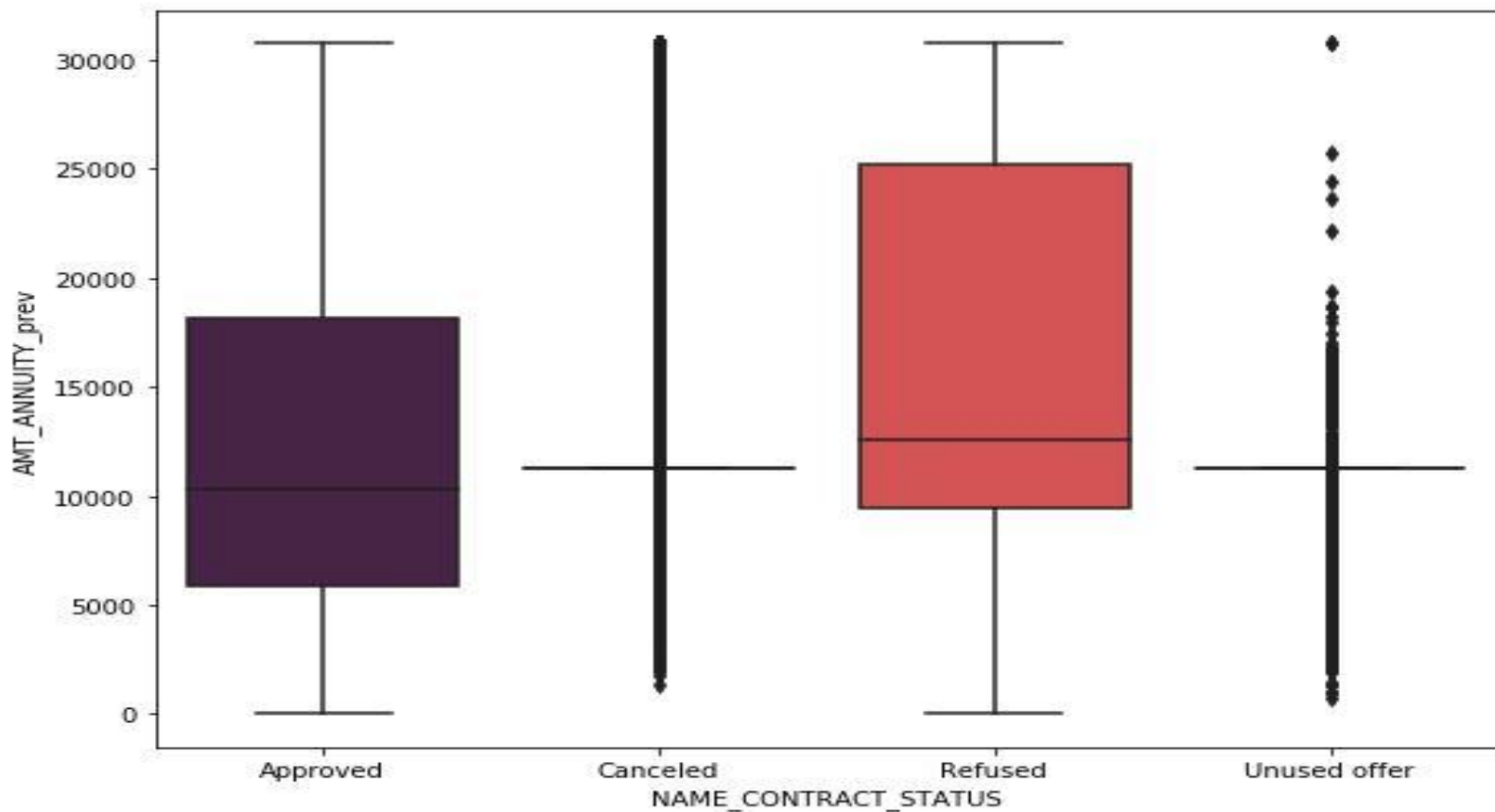


# ANALYSIS OF CONTINUOUS COLUMNS IN MERGED DATA SET

# BOX PLOT LOAN STATUS V.S AMT\_ANNNUITY

Insights from the above graphs :

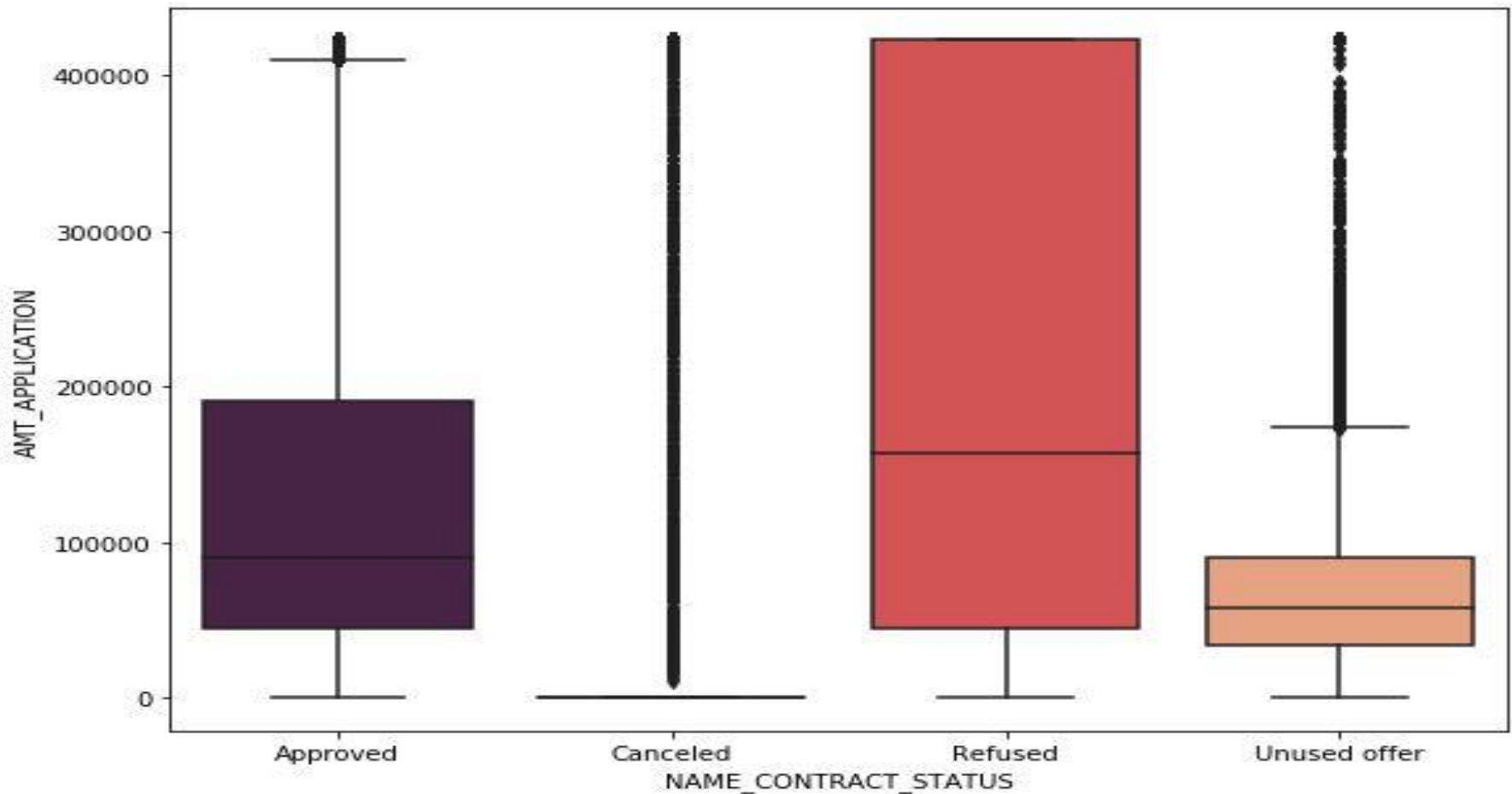
> Lower AMT\_ANNNUITY has higher chances of approval.



# BOX PLOT LOAN STATUS V.S AMT\_APPLICATION

Insights from the graph :

> Lower AMT\_APPLICATION has higher chances of approval and higher

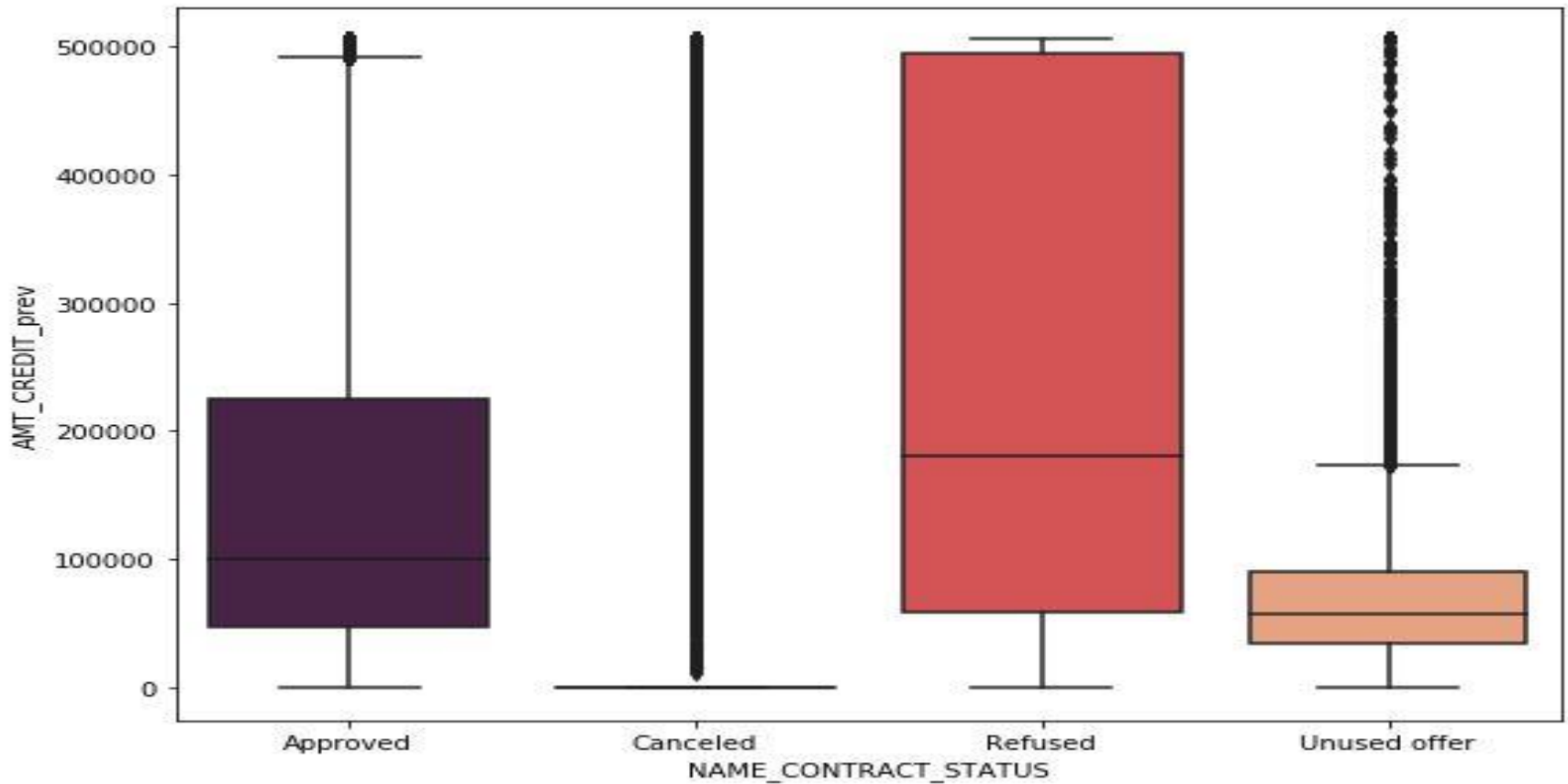




# BOX PLOT LOAN STATUS V.S AMT\_CREDIT

Insights From The Graph :

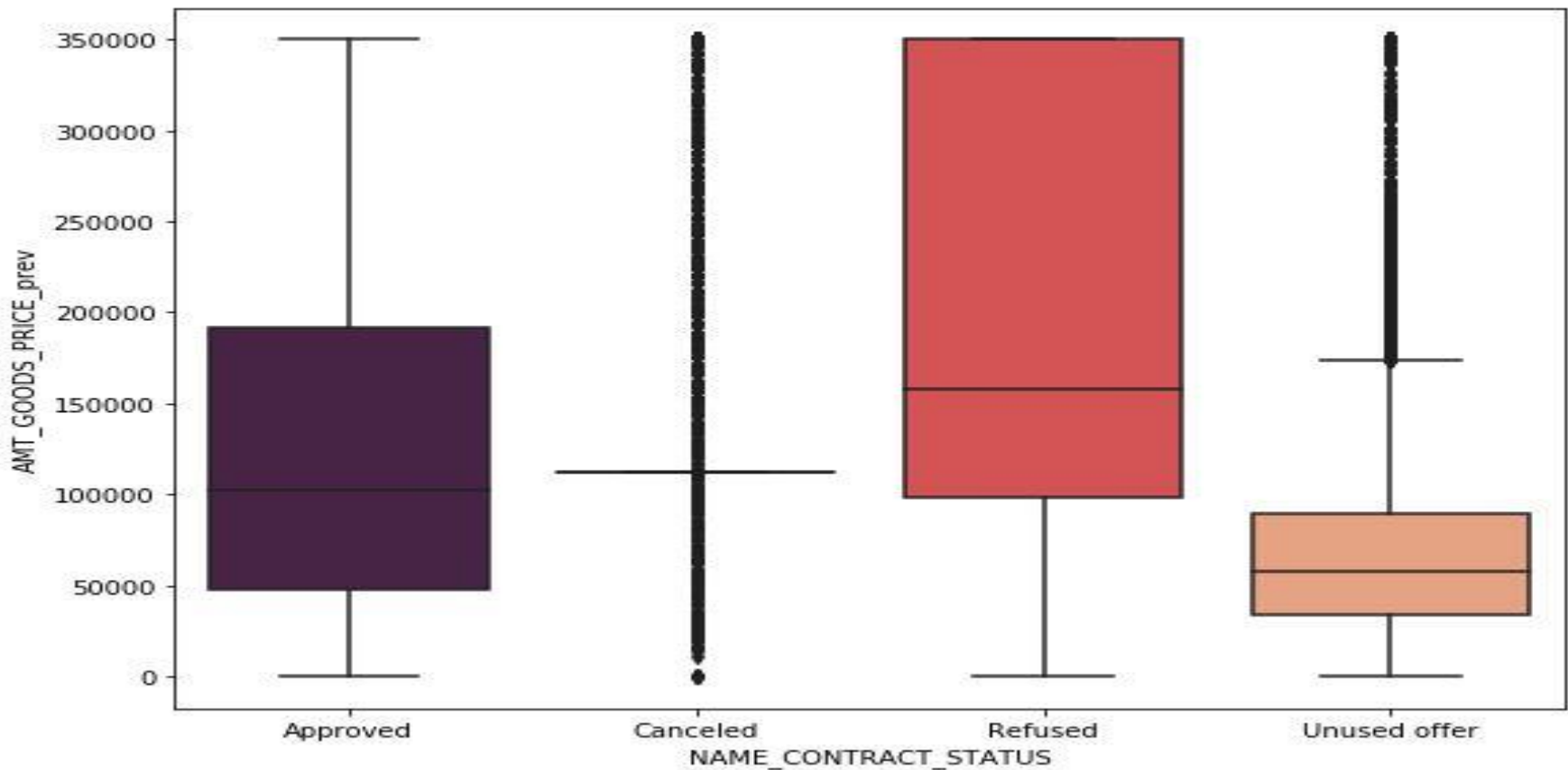
> Higher AMT\_CREDIT has higher chances of refusal.



# BOX PLOT LOAN STATUS V.S AMT\_GOODS\_PRICE

Insights From The Graph :

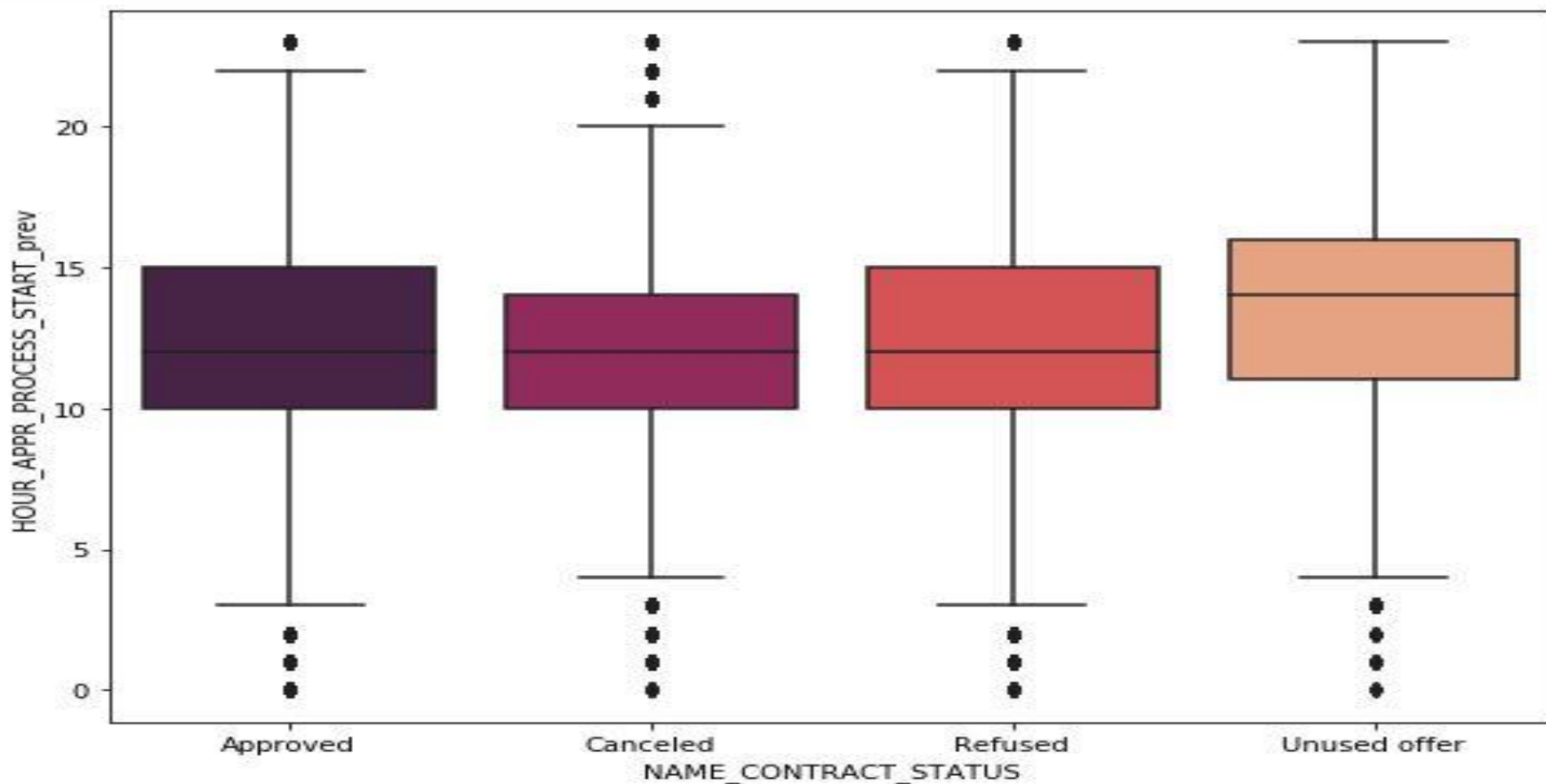
> Higher AMT\_GOODS\_PRICE has lower chances of approval.



# BOX PLOT LOAN STATUS V.S HOUR\_APPR\_PROCESS\_START

Insights From The Graph :

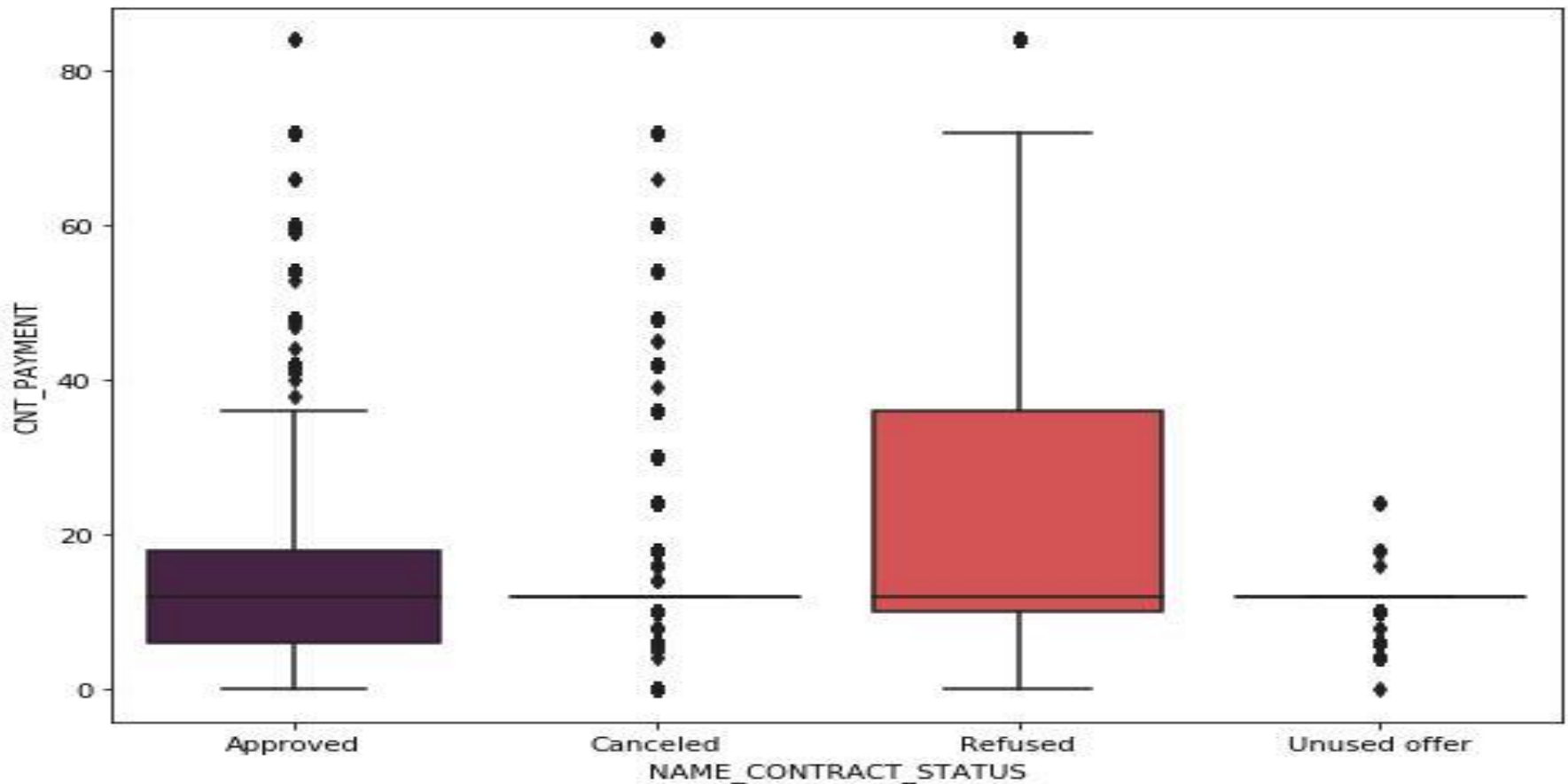
> Higher applications are approved between 10 AM to 2 PM.



# BOX PLOT LOAN STATUS V.S CNT\_PAYMENT

Insights From The Graph :

> Lower loan tenures have lower chances of refusal.



# CONCLUSION

- ◉ Clients who do not own a car are tend to default on loan.
- ◉ Pensioners are less likely to default and working people are more likely to default.
- ◉ Clients with higher education are less likely to default.
- ◉ Single/Not Married and Civil Marriage people have slightly higher chances of default.
- ◉ Managers have a low chances of loan default.
- ◉ Clients with organization type as Business Entity Type 3 and Self Employed are more likely to default.
- ◉ People with higher Income Total are less likely to default.
- ◉ People with higher Credit amount of loan are less likely to default.
- ◉ People above 50 years of age are less likely to default on their loan.
- ◉ People who newly got employed into their current employment(less than 4 years) are the highest among loan defaulters.
- ◉ Repeater Clients are applying for cash loans in comparatively higher number. New clients are opting for consumer loans in higher number.

*THANK YOU*