

EDA ON LOAN DEFAULTERS

BY HEMALAKSHMI B

Business Understanding

The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Some consumers use this to their advantage by defaulting on loans. Suppose you work for a consumer finance company which specializes in lending various types of loans to urban customers. You have to use EDA to analyze the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.

When the company receives a loan application, the company has to decide for loan approval based

on the applicant's profile. Two types of risks are associated with the bank's decision:

1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
2. If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

To avoid risks associated with lending loan, we can do a detailed Exploratory Data Analysis on the given data sets. This can give us information on defaulters pattern and draw insights from it.

There are three datasets provided:

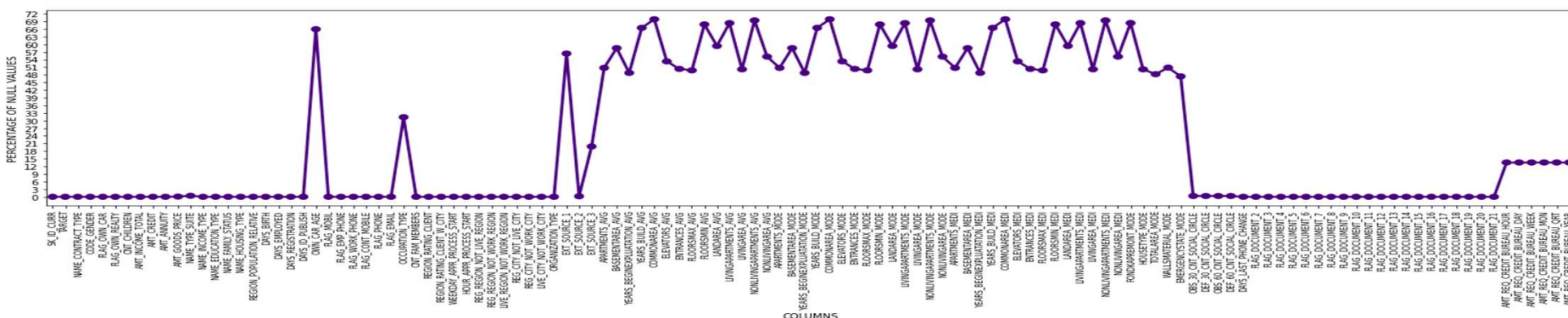
1. columns_description.csv:- Gives information on the columns in both datasets.
2. application_data.csv:- This dataset gives information on loan, difficulty in repayment of loans.
3. previous_data.csv:- This dataset gives information on previous loan applications.

DATA MINING

Sleuthing for insights

HANDLING MISSING VALUES

APPLICATION DATASET MISSING VALUES



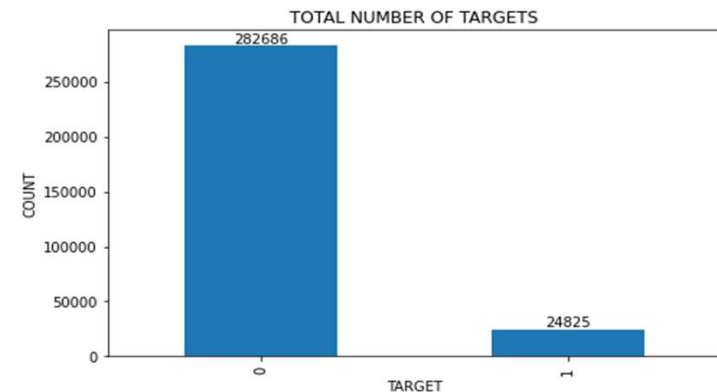
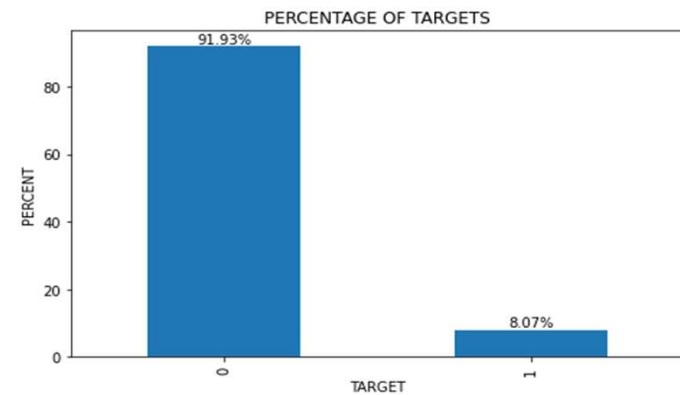
- ▶ Missing values in columns range from 13% to 70%.
- ▶ Features with missing values more than 40% have been deleted. As there are few features with more than 30% null values but could impact our EDA have been retained. There is no said criteria for deleting columns with null values.
- ▶ Few features with normalised values with less detail about it are also deleted.
- ▶ Total 71 columns were finalised for EDA in Application dataset and 26 columns in Previous application dataset

ANALYSING APPLICATION DATA SET

Application data set has a main feature 'TARGET'.
'TARGET' feature tells us whether the given ID has defaulted (TARGET = 1) on repayment of their loan or not.

INSIGHTS:

- ▶ Most of the clients i.e. 91.93% are non defaulters while only 8% of them have defaulted in repayment.
- ▶ This seems a highly imbalanced data.
- ▶ We will have to investigate further, study other features in relation to find out the pattern of defaulters

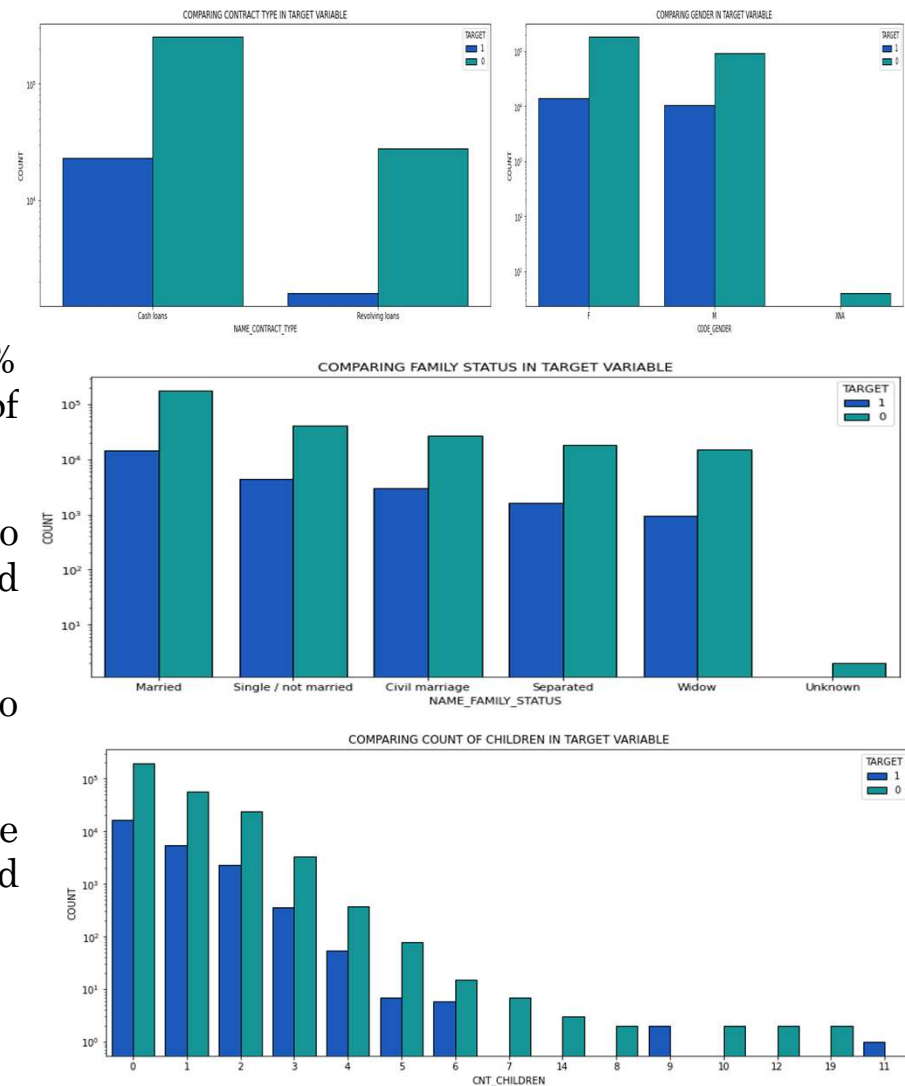


UNIVARIATE ANALYSIS

Univariate is a simple analysis and are done on single variable based on categories like Income type, Occupation type, Gender, organization type etc.

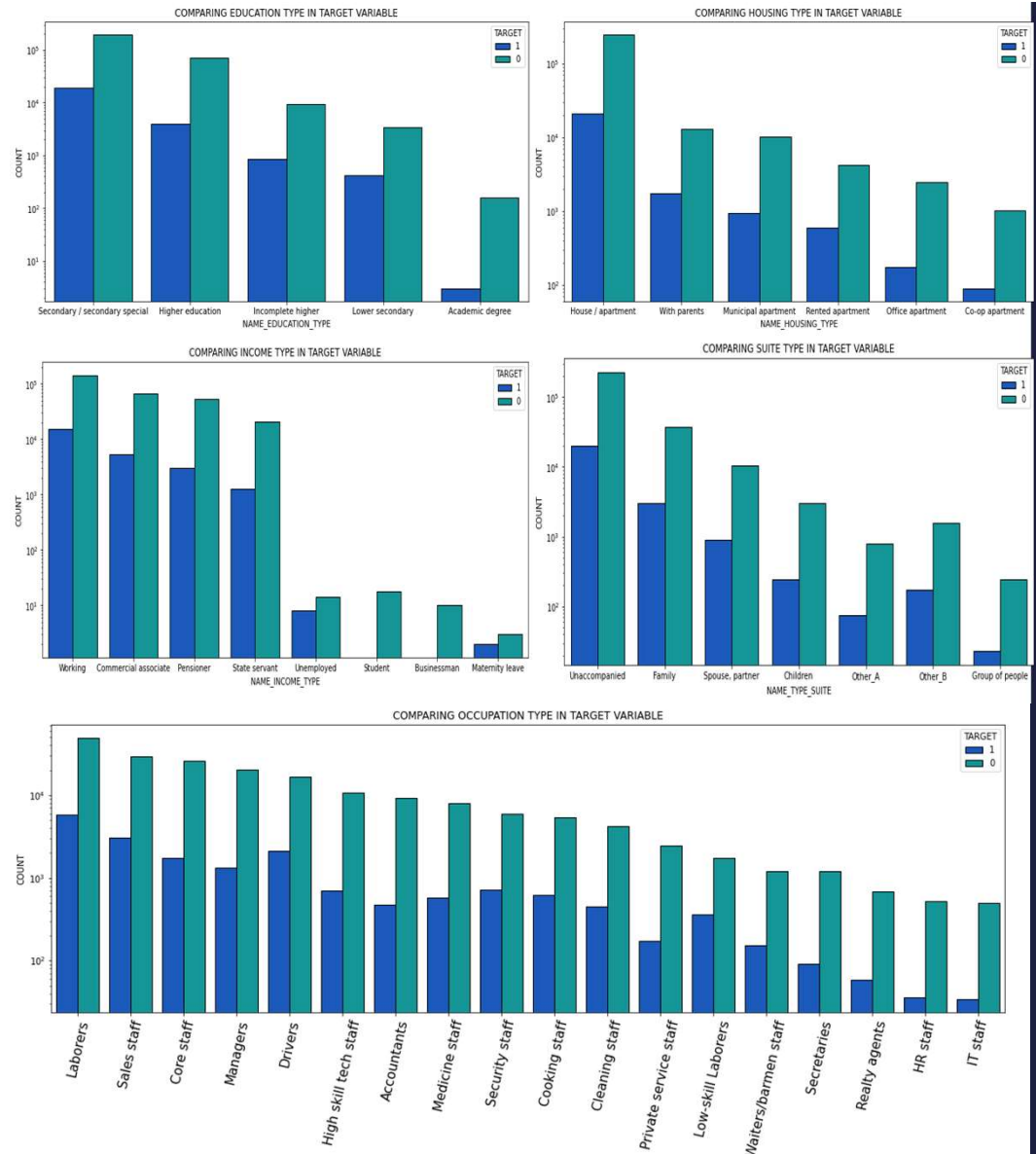
INSIGHTS:

- ▶ 90.48% of clients took cash loans out of which 8.3% were defaulters. 9.52% took revolving loans out of which 5.47% were defaulters
- ▶ 65.83% were females out of which 6.99% tend to default and 10.14% out of 34.16% males had defaulted.
- ▶ 63.88% were married and 5.8% of widows tend to default
- ▶ 70.03% had no children, 7.71% of them were defaulters and 28.57% with 6 children had defaulted.



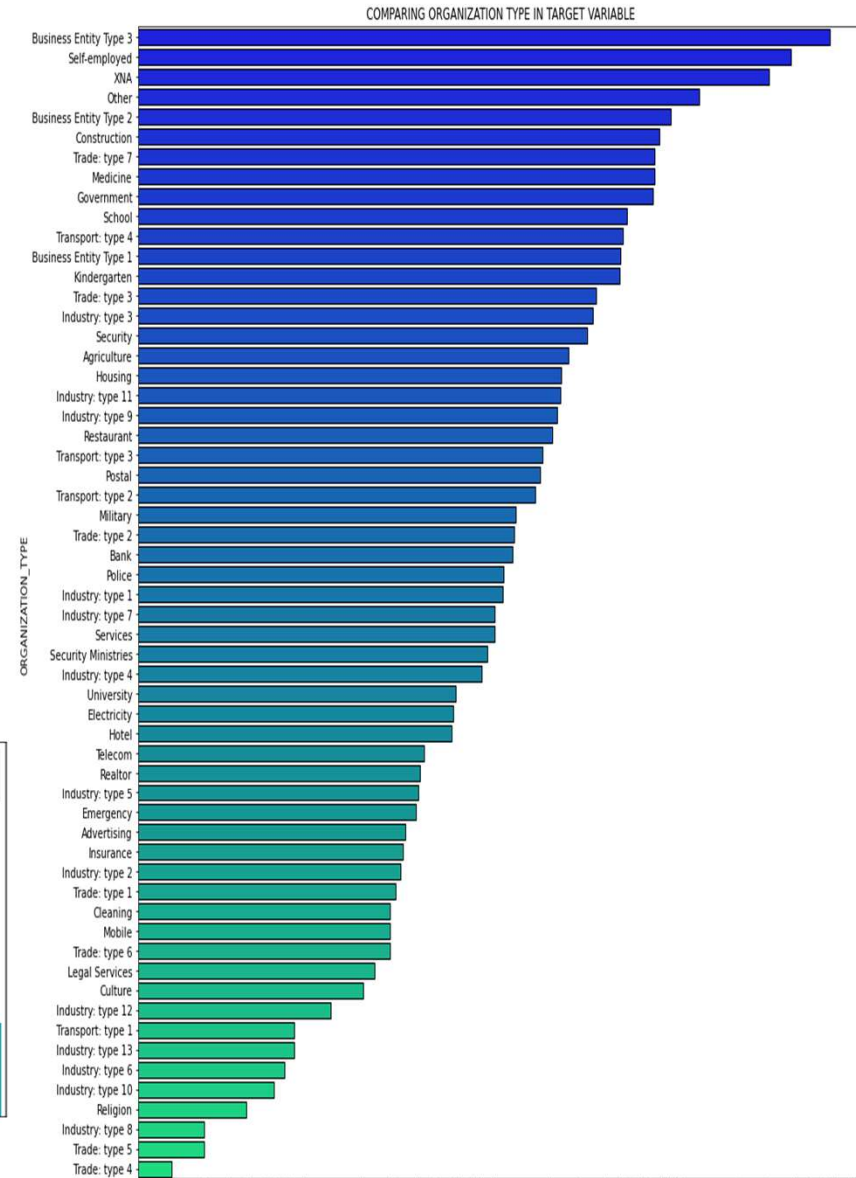
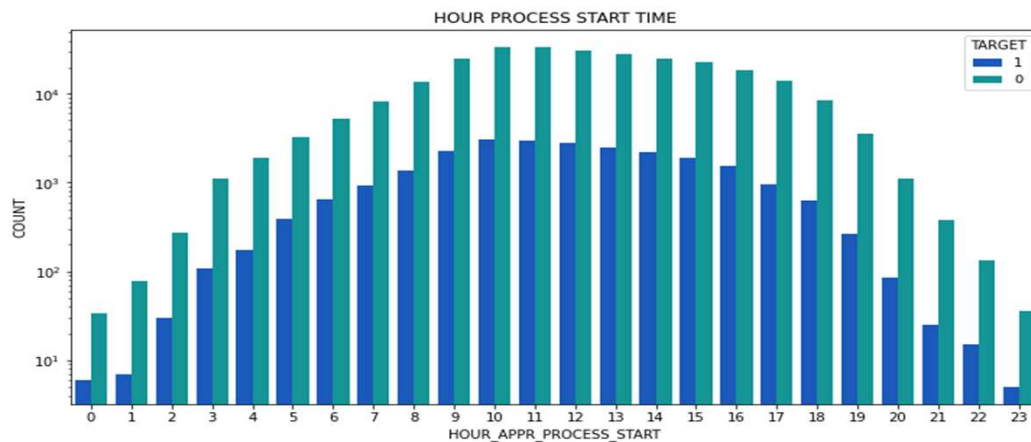
INSIGHTS:

- ▶ 71.01% have Secondary/secondary special education and 8.94% of them have defaulted along with 10.92% of people with Lower secondary education.
- ▶ 51.63% of clients have working income type and 9.59% have defaulted. Student and business type do not seem default.
- ▶ 81% were unaccompanied while applying for loan.
- ▶ Of 88.73% who had house/apartment, 7.8% defaulted. Approximately 12% of those who lived in rented apartment or live with parents have defaulted.
- ▶ Labourers constitute 26.14% of which 10.58% and 17.15% of Low-skill labourers seem to have defaulted.



INSIGHTS:

- ▶ Business Type 3 constitute 22.11% of clients. Of which 9.29% had defaulted. 15.75% of Transport Type 3, 13.43% of Industry Type 13, 12.5% of Industry Type 8 were among top defaulters.
- ▶ Peak hour for most of the process were between 10-13. Most defaulted at 0, 5, 6, 7 and 23 hrs
- ▶ But we need to dig deeper for more stronger insights

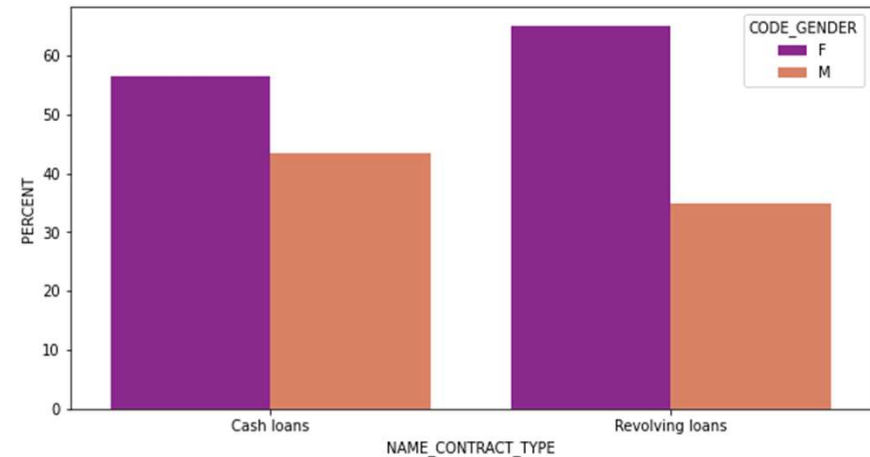
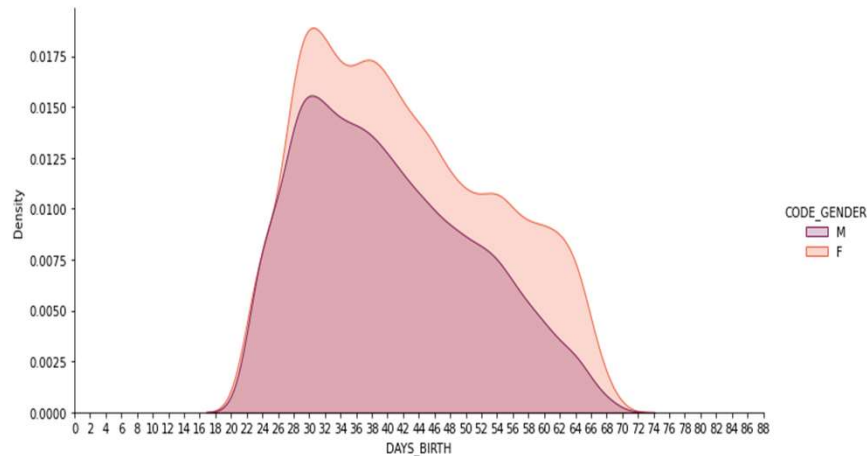
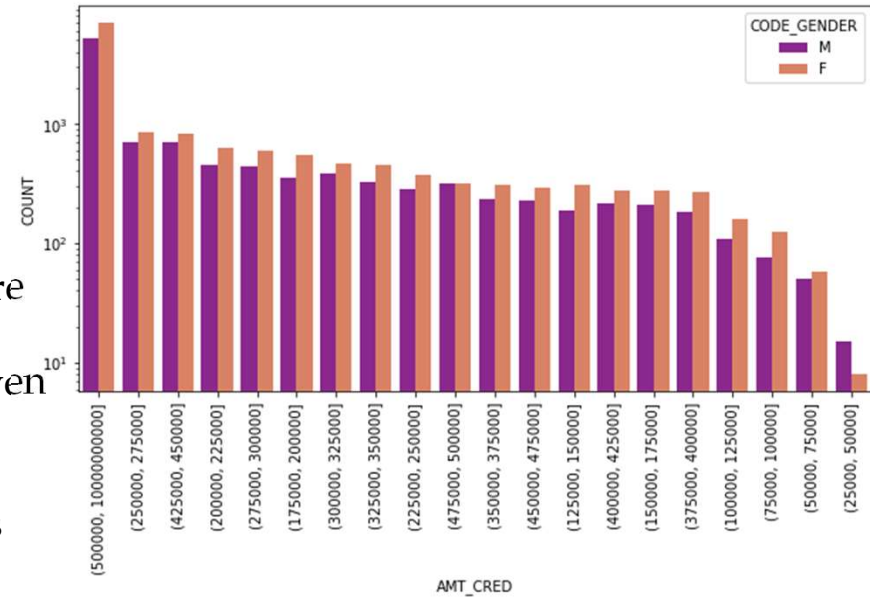


BIVARIATE ANALYSIS

GENDER ANALYSIS

INSIGHTS:

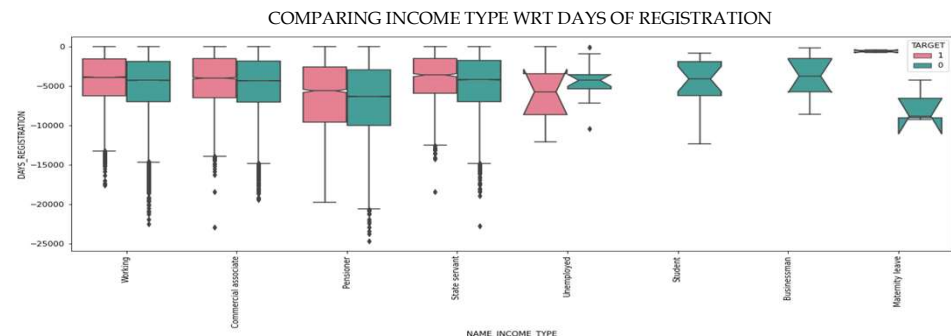
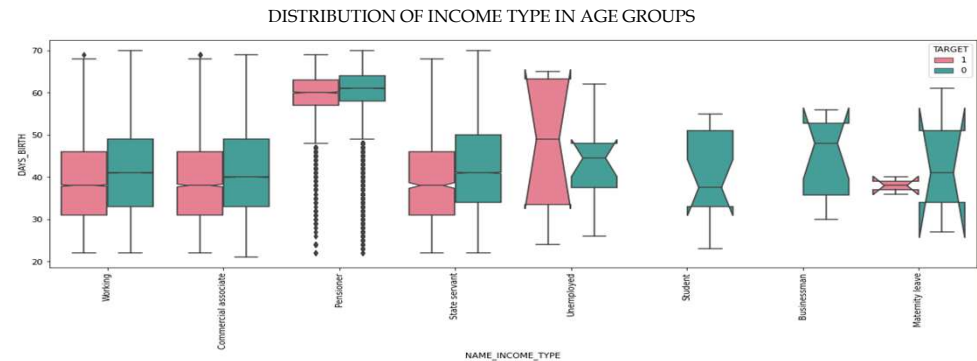
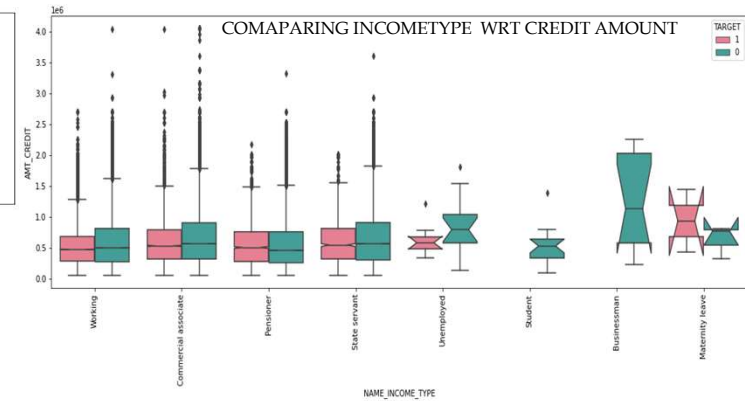
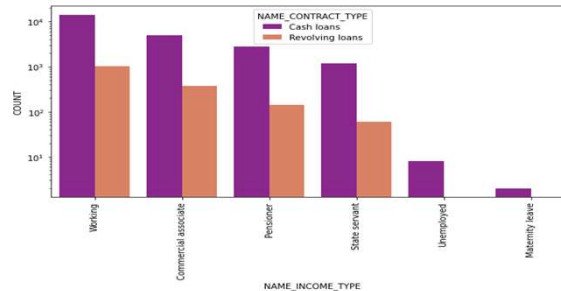
- ▶ 56.53 % Cash loans and 65.02% of revolving loans were taken by females
- ▶ Credit amount taken by females were more in any given range.
- ▶ Most females belonged to 30-40 age group and most males belonged to 25-35 age group. Female count was more even in higher age groups compared to that of males



INCOME TYPE

INSIGHTS

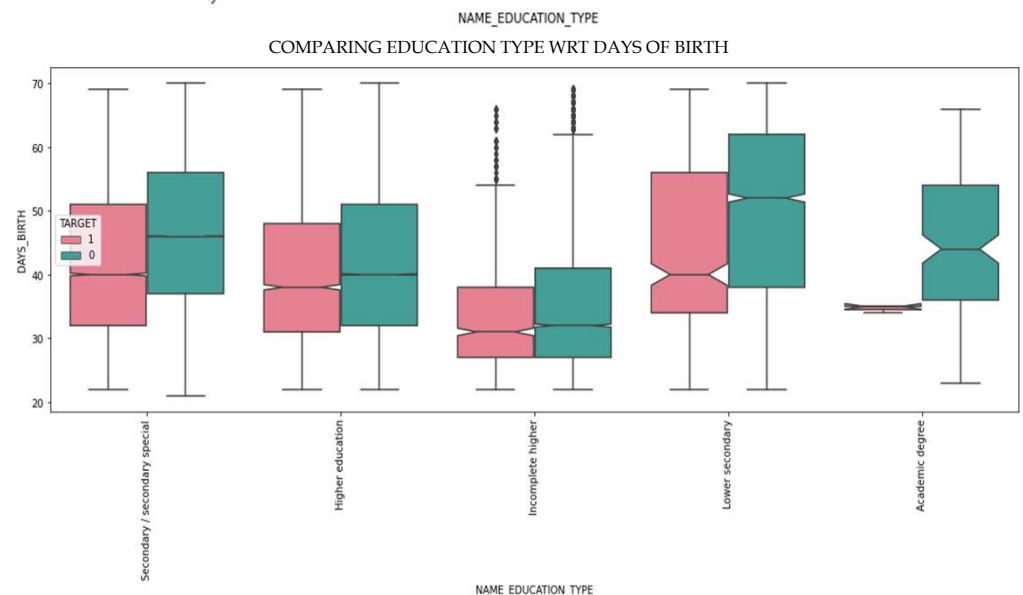
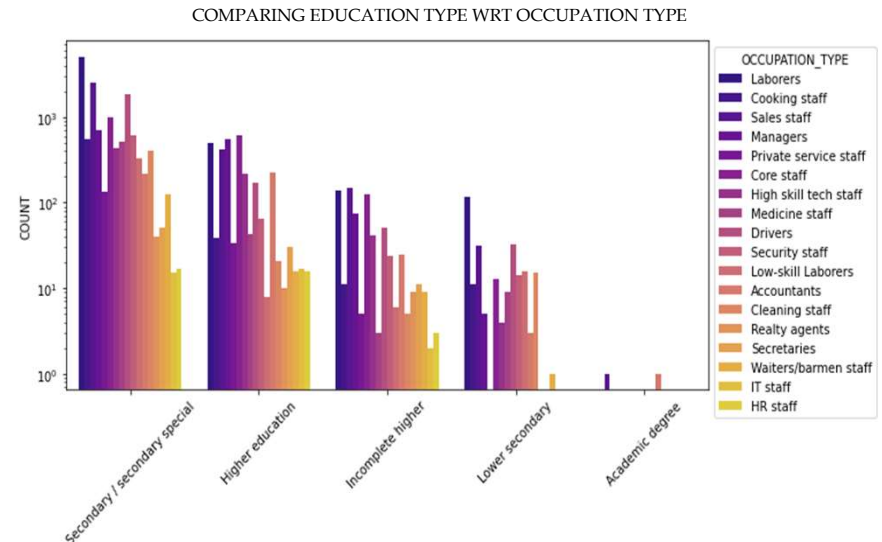
- ▶ Most of them took cash loans but the ones with maternity leave who took cash loans and belonged to age group 35-40 definitely defaulted.
- ▶ Unemployed who took cash loans defaulted more and no defaulters are observed in revolving loans
- ▶ The defaulters with higher credit amount with maternity leave as income type were more
- ▶ The clients with less time between the application and change of registration defaulted more



EDUCATION TYPE

INSIGHTS

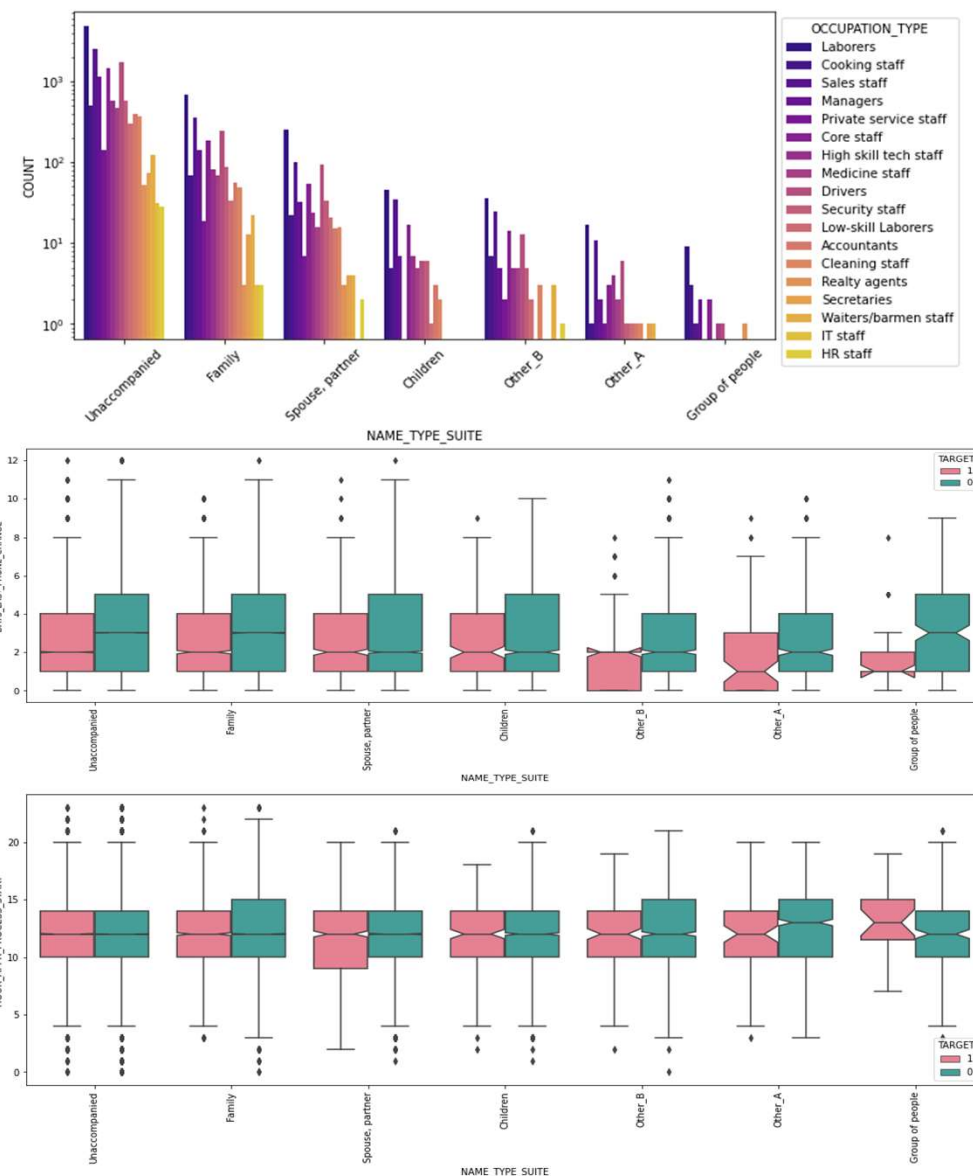
- ▶ The highest defaulters belonged to secondary/secondary special education type and lowest belonged to ones with academic degree.
- ▶ In lower secondary education the IT staff, Private service staff, Realty agents and secretaries did not default.
- ▶ The only defaulters from academic degree belonged to accountants and sales staff. Their age group can also be narrowed down to approximately 34-35 age



SUITE TYPE

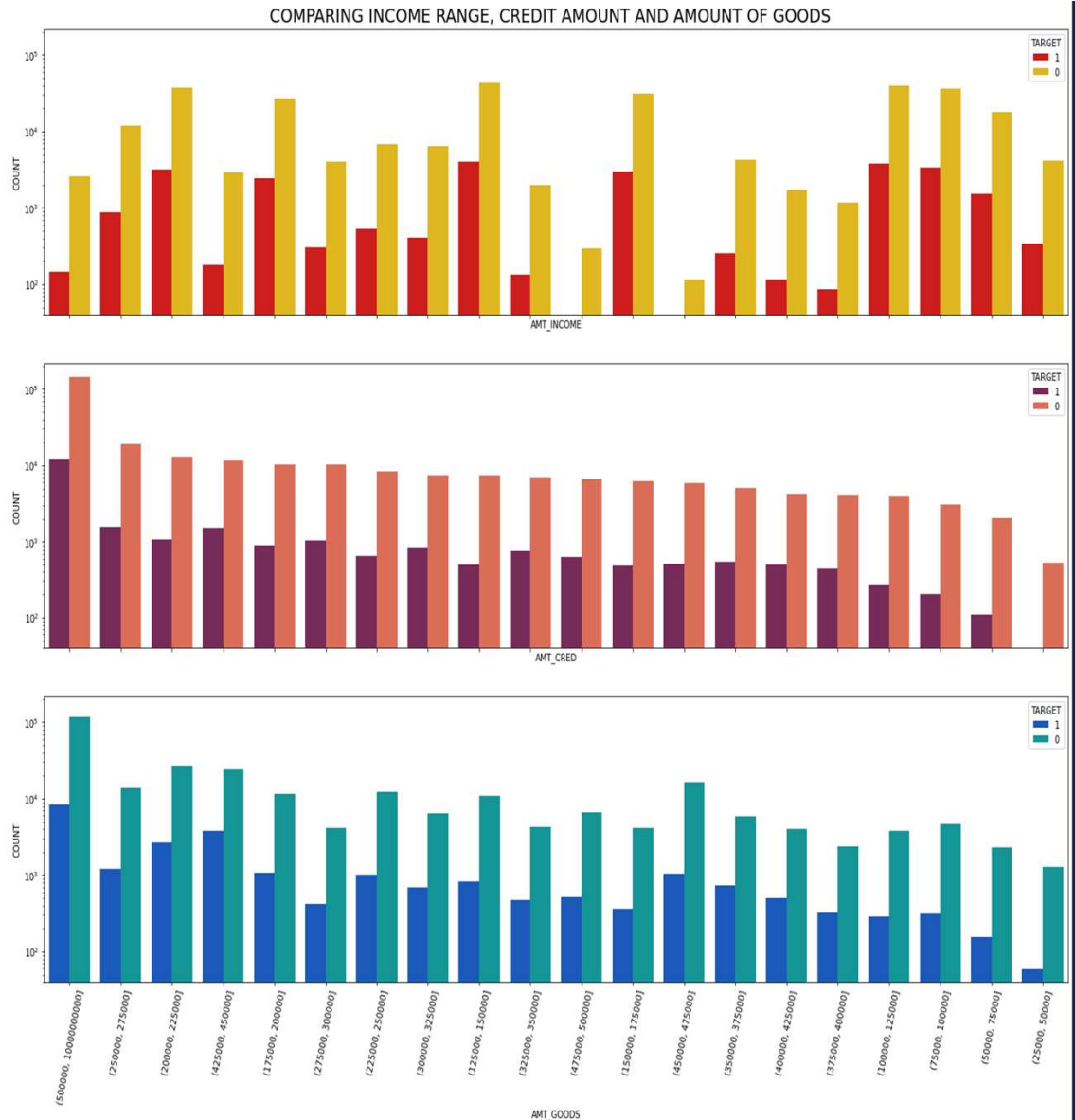
INSIGHTS:

- ▶ In defaulters people with suite type as spouse/partner, never had IT staff as occupation.
- ▶ Suite type Children never had occupation type as Private service staff, Realty agents, secretaries, Waiters/barmen staff, IT staff or HR-staff.
- ▶ With suite type Other_B , no defaulters were from Accountants, Realty agents, Secretaries, IT staff occupation.
- ▶ Realty agents with suite type Group of people always defaulted. Suite type Group of people were never from occupation IT staff, High skill tech staff, Accountants, Waiters/barmen staff, IT staff or HR-staff.
- ▶ It is also observed that defaulters with suite type Other_B, Other_A and Group of People had their last phone changed near to application date.
- ▶ Most of the defaulters with suite type Group of people had applied for loan at 12 hrs.



Comparing the amount of goods and credit taken in the amount range:

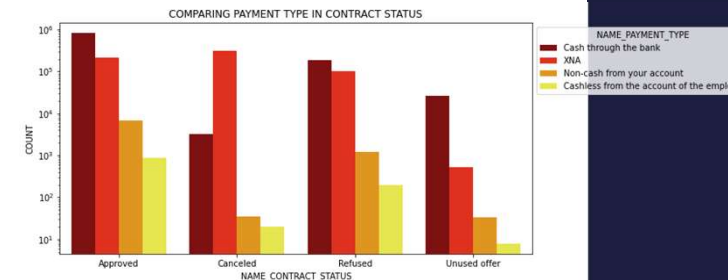
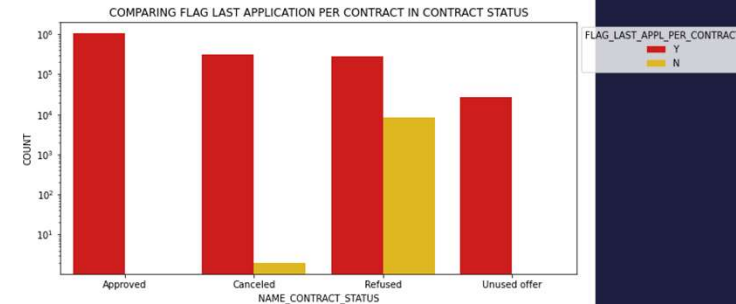
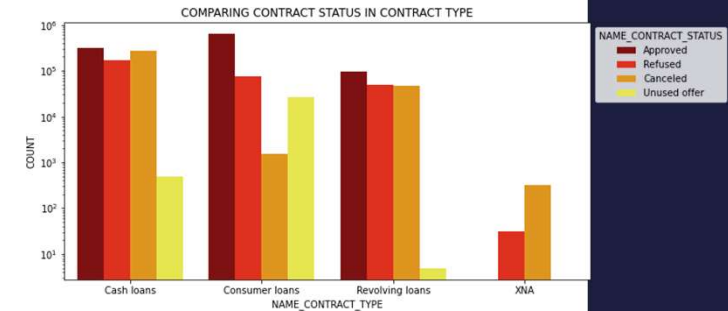
- ▶ We can see that almost 33.58% of defaulters had applied for loan with goods amount in range of (500000, 100000000000) , but 49.22% had taken credit amount in same range.
- ▶ That means credit amount was more than goods price in 15.64% of cases.
- ▶ This parity can be observed in other price range as well



ANALYSING PREVIOUS APPLICATION DATA

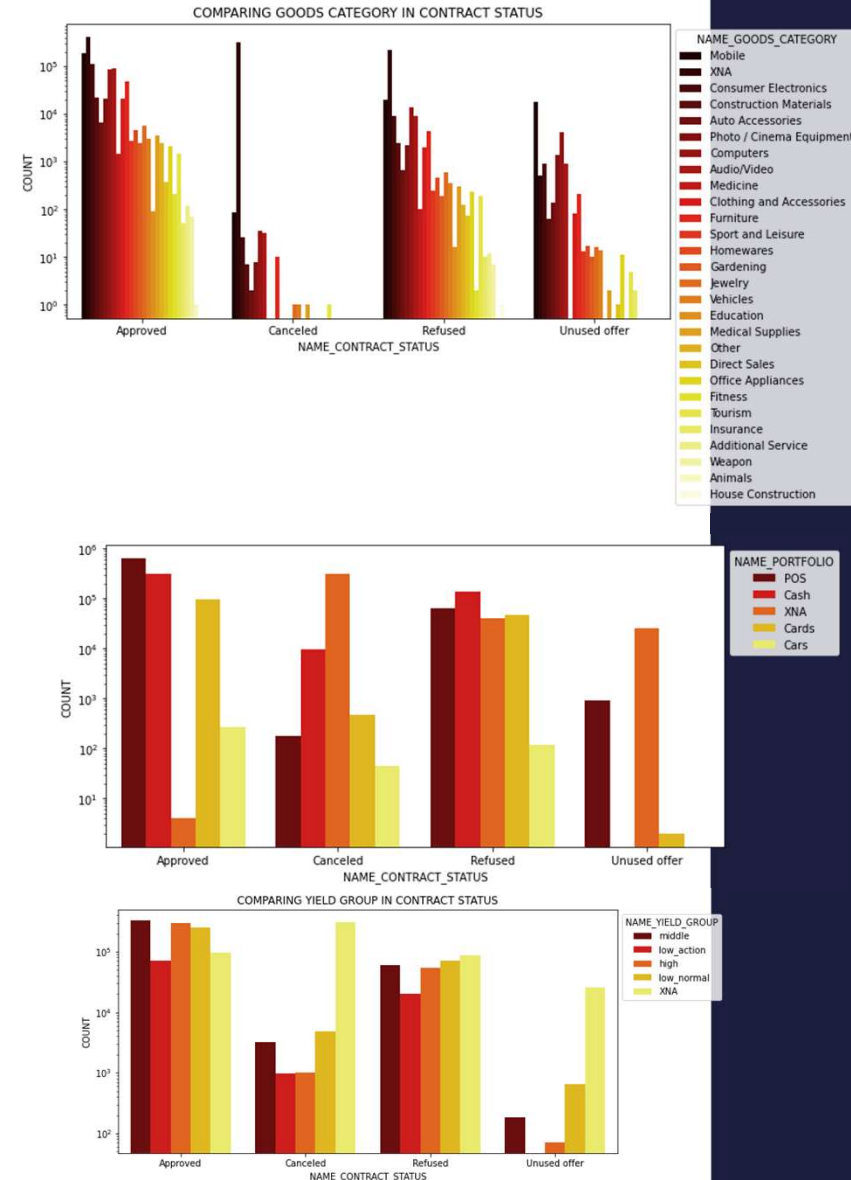
INSIGHTS:

- ▶ In contract type 85.92% of consumer loans were approved and XNA were never approved.
- ▶ All approved status with Flag last application contract were yes. The no were mostly refused.
- ▶ 78.82% of approved had payment type Cash through bank as most. 98.97% with cancelled status belonged to XNA payment type was less approved mostly cancelled or refused.



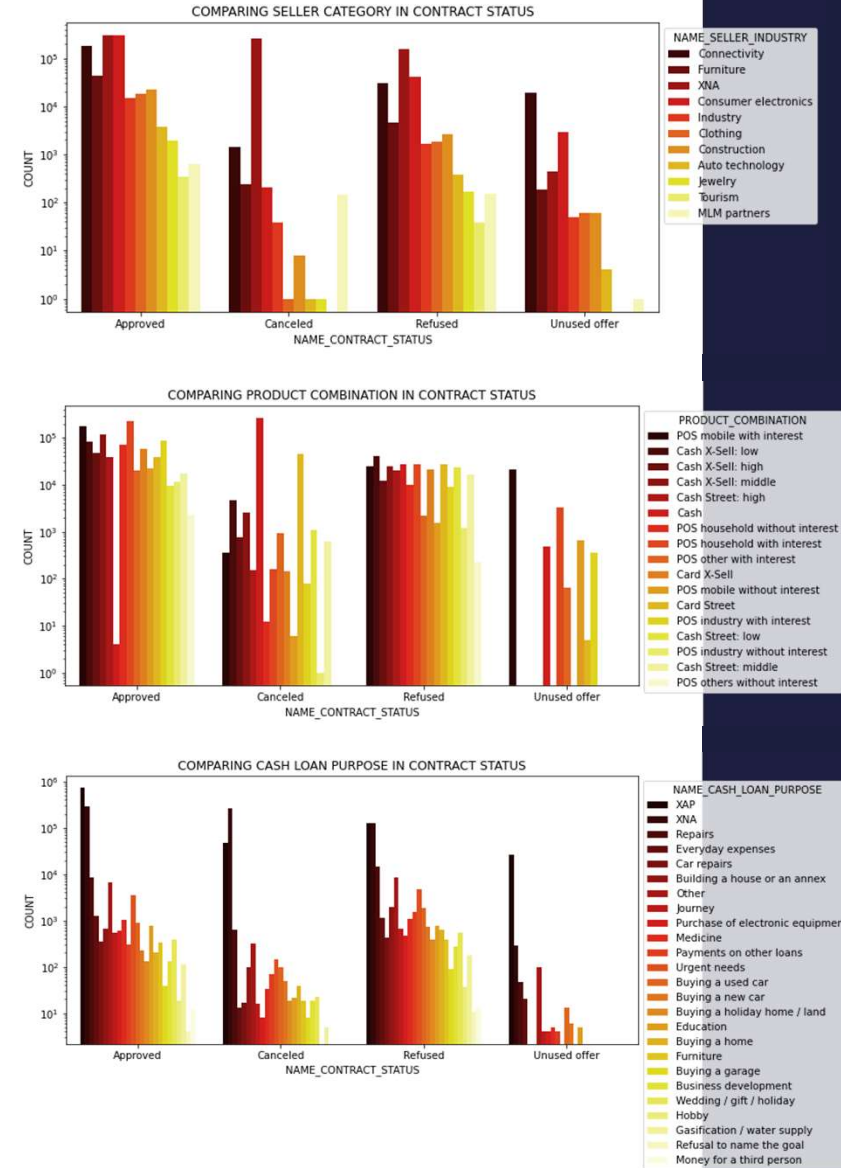
INSIGHTS:

- ▶ Goods category Fitness, Other and Medicine were among top approval rates. Only 81% in Insurance were approved. XNA category covered 99% in cancelled status and only 43% were approved.
- ▶ Portfolio POS was most approved and least approved were XNA. Most refused were cash portfolio.
- ▶ 59.22% Yield type XNA had cancelled. 31.16% of approved were from middle yield followed by 28.84% from high yield.



INSIGHTS:

- ▶ Seller industry type XNA were only 40.90% approved and 36.68% were cancelled. Most unused offers belonged to connectivity. MLM partners were only 65.60% approved.
- ▶ Product Combination POS household with interest constituted 22.50 % of approved status. Cash dominated cancelled status with 81.89%. Approval constituted only 34.49% in Card Street.
- ▶ In cash loan purpose 78.5% of XAP and 42.13% of XNA was approved whereas 80.42% of Payment on other loans and 73.33% of Refusal to name the goal were refused.



CONCLUSION

Final thoughts

CONCLUSION:

- Overall there were more female clients.
- The IT staff, Private service staff, Realty agents and secretaries with lower secondary education, can be trusted.
- Students, businessman never defaulted.
- Banks should pay attention to:
 - The age group of 25-40.
 - People with high credit amount for low goods price and low income.
 - Female defaulters who were associated with high credit amount.
 - People with occupation type as Laborers, Low skill laborers.
 - People who belong to Business type entity 3, XNA, self employed.
 - Realty agents with suite type group of people.
 - Maternity income type with cash loans.
 - People with Academic degree who took higher credits and with occupation type accountants and sales.
- From Previous application:
 - XNA contract type, XNA payment type defaulted more.
 - Lower default rates in product combination as POS industry without interest, seller industry as clothing and tourism, portfolio as cars.
 - Higher default rates were found in Refusal to name the goal, Hobby, Car repairs, Gasification/water supply, Money for a third person and Payments on other loans in that order.