

CREDIT EDA 'CASE STUDY'

}

[BY INDER BALDAWA]

}

Problem Statement {

1. Identify the factors affecting loan repayment and use EDA to understand their influence on loan default.
2. Determine the relationship between consumer attributes and loan default using EDA.
3. Analyze patterns in loan data to minimize the risk of financial loss for the company and ensure loan applicants who are capable of repayment are not rejected.

}

Steps followed for 'EDA' {

Step 01 DATA UNDERSTANDING

Step 02 Data Cleaning and Manipulation

Step 03 Outlier and data imbalance check

Step 04 Data analysis

Step 05 CONCLUSION

}

DATA UNDERSTANDING < /1 > {



< Importing libraries >

< Checking each column, Index, header, footer >

}

< Identifying data quality issues >

Concepts < /2 > {



< Missing value Imputation analysis >

< Checking the structure and the metadata >

}

< Changing datatypes to date, time, string, int, bool, etc for ease of analysis >

Outlier Check < /3 > {



< Checking the data for outliers that would cause the analysis to be biased>

< Checking for imbalances, ratio, percentage of imbalance >

}

Data Analysis < /4 > {



< Business requirement oriented analysis >

< Correlation between columns >

< Univariate and Bivariate analysis >

< Creating plots to understand the data better and find insight >

}

Conclusion < /5 > {

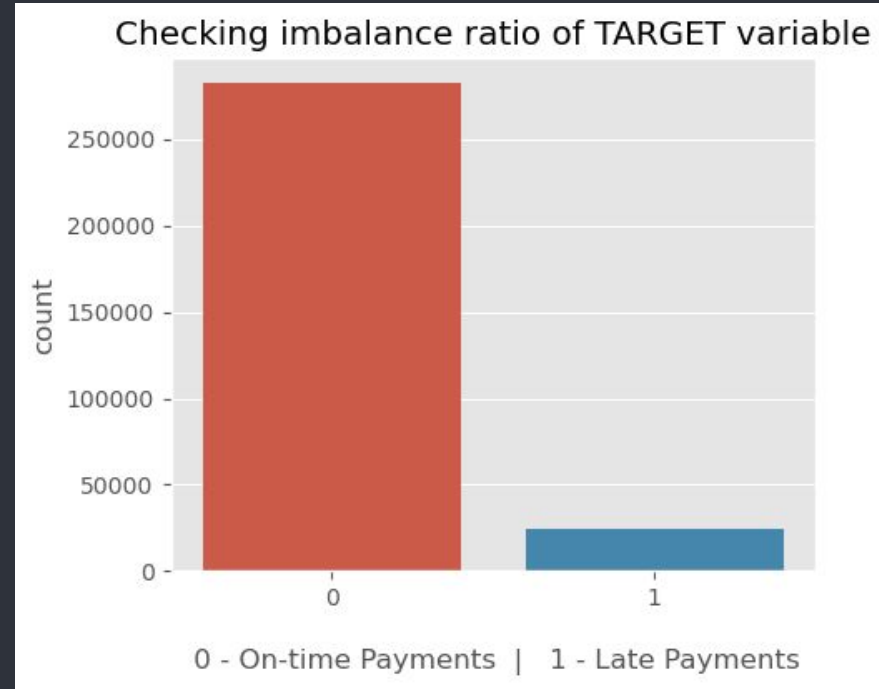
< Explain Business driven, type driven and
data-driven metrics created in previous steps >

< Help in getting better understanding about the
data >

}

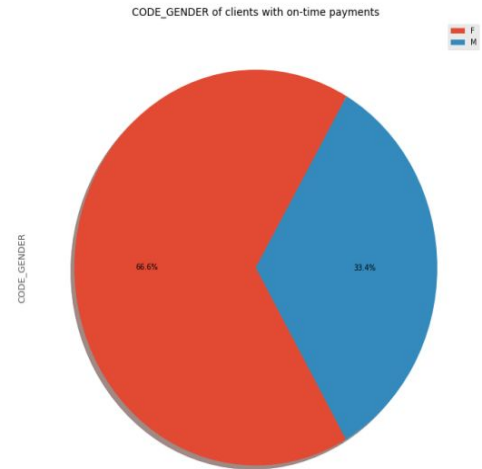
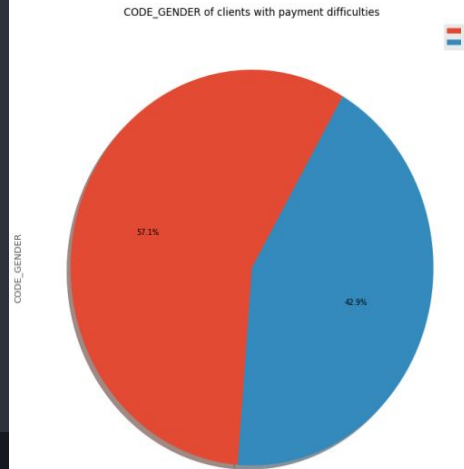
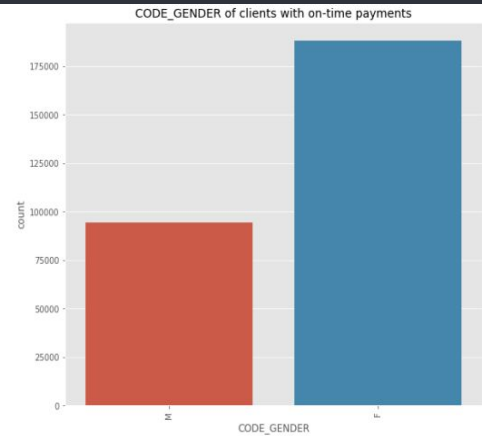
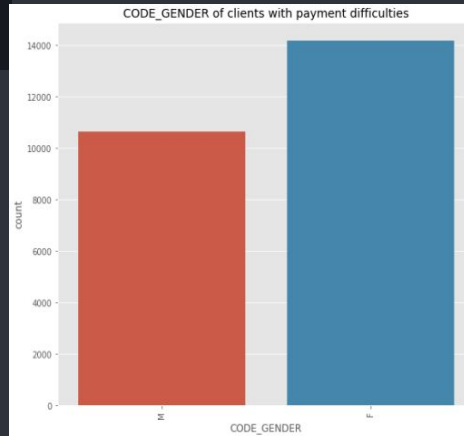
Checking Imbalance for target column 'TARGET'

- The ratio of people who default to those who pay on-time is 11.38
- 1 in every 11 applicant has difficulty in making.



'Gender' as a factor

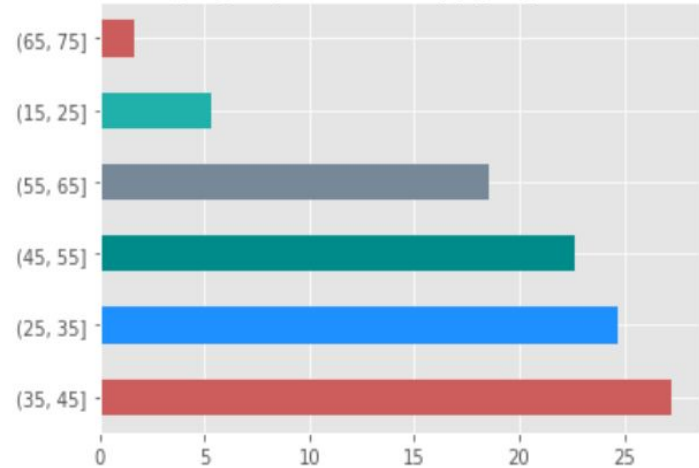
- Females have processed more loans than males.
- Male applicants more likely to default than female applicants .
- Female clients who are working are more likely to make on-time payment



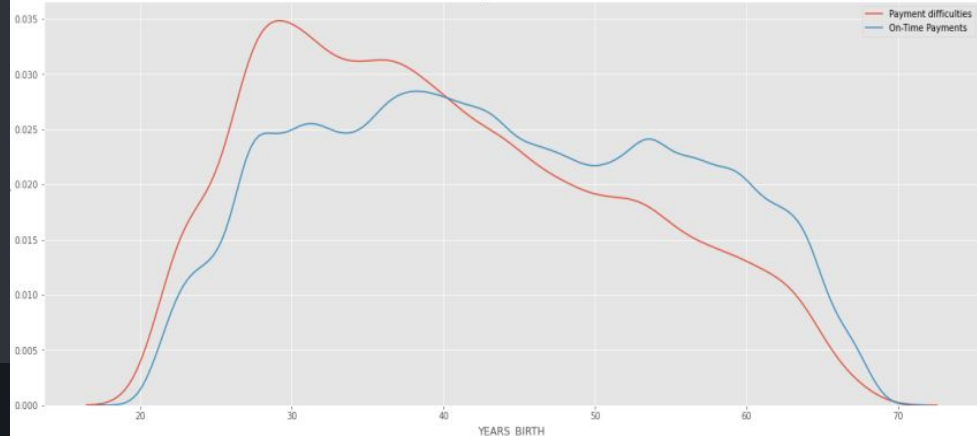
'AGE' as a factor.

- Individuals between the age of 20 and 40 are more likely to Default.
- Individuals with age > 40 are more likely to make on-Time Payments
- Most people belong to the age group of 35-45.

Age group of clients applying for loan

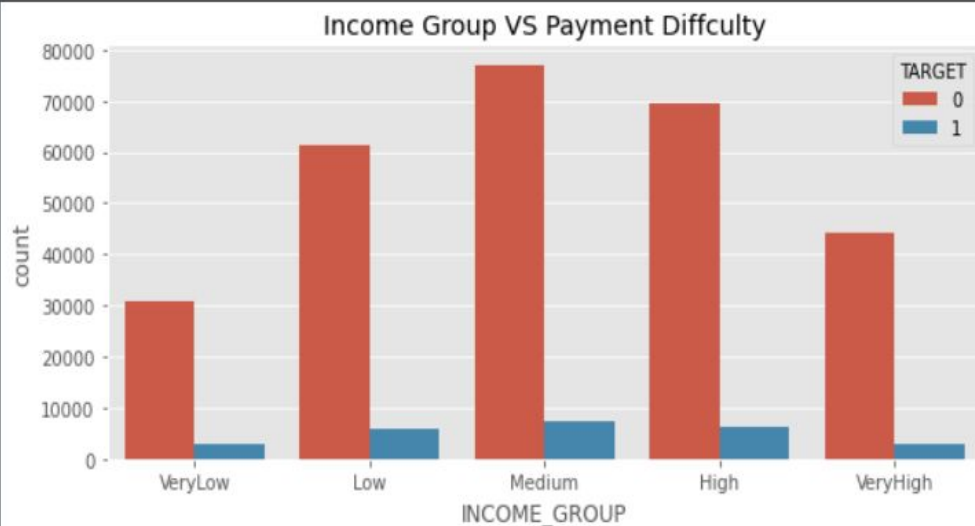


YEARS_BIRTH of clients



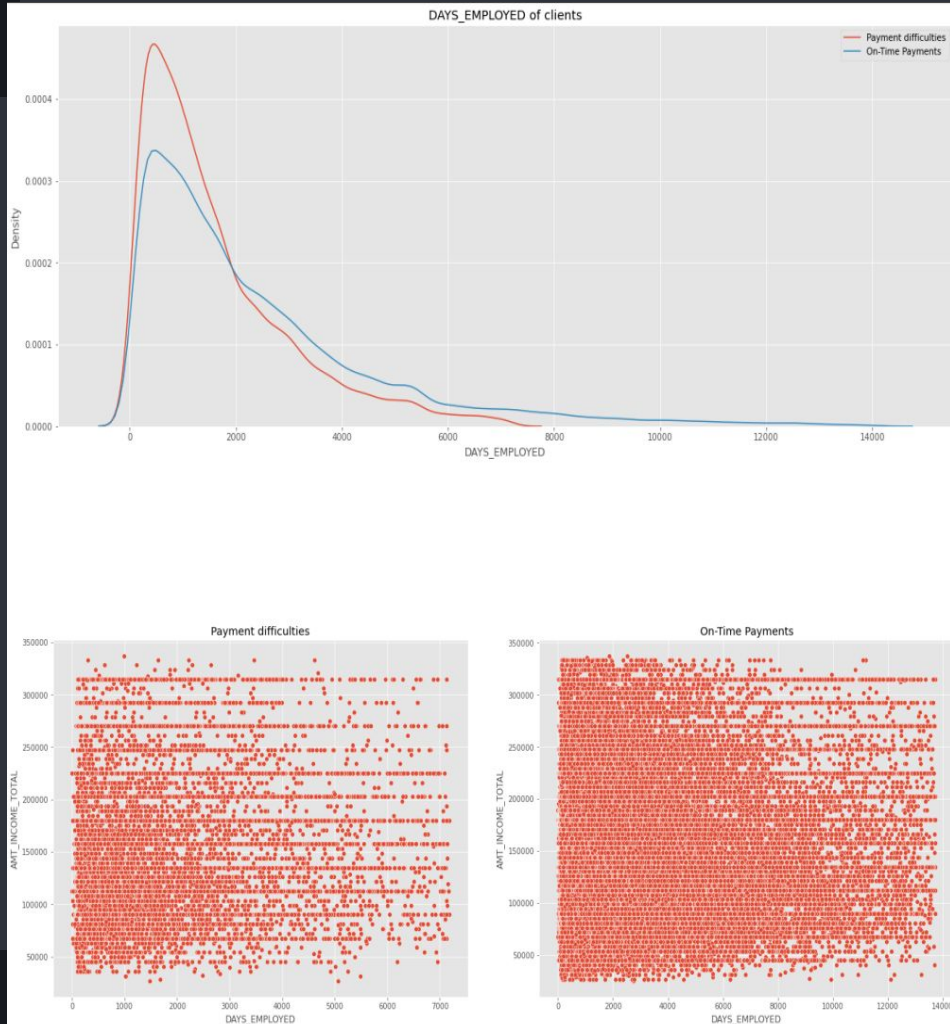
'INCOME' as a factor.

- Most applicants belong to the medium income group.
- Most defaults come from the Medium income group
- Income group 'Low' and 'High' is also a concern.



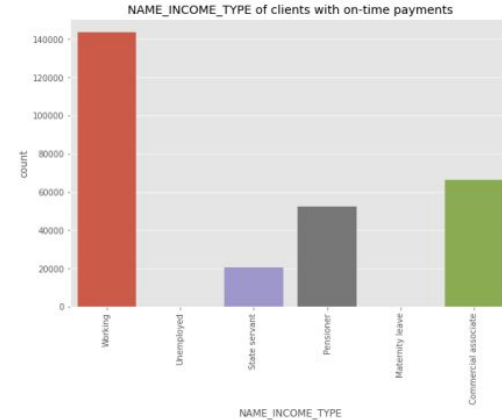
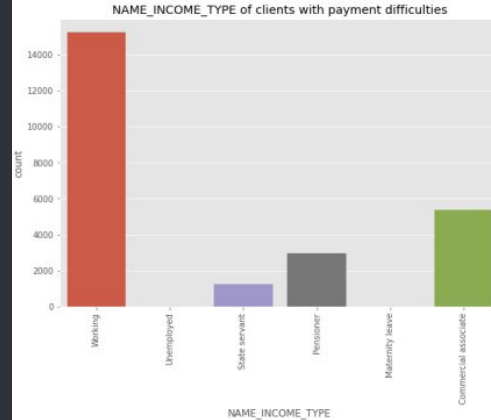
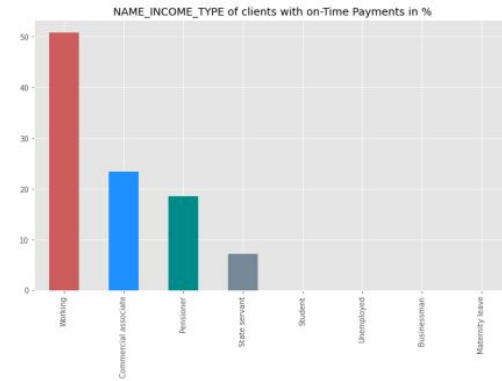
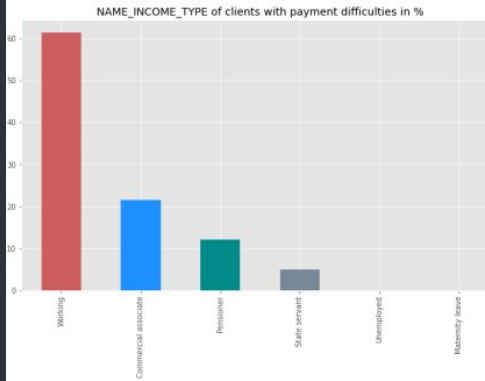
'Employment Years' as a factor.

- Clients who are employed for a long time tend to make their payments on-time.
- Applicants with with more than 4000 days of employment don't default.
- Applicants with with more than 2000 days of employment are least likely to default.



'Job Type' as a factor.

- Most applicants belong to the 'Working' class
- Pensioners have a good track record for on-time payments
- Students and businessmen are less likely to default as compared to others

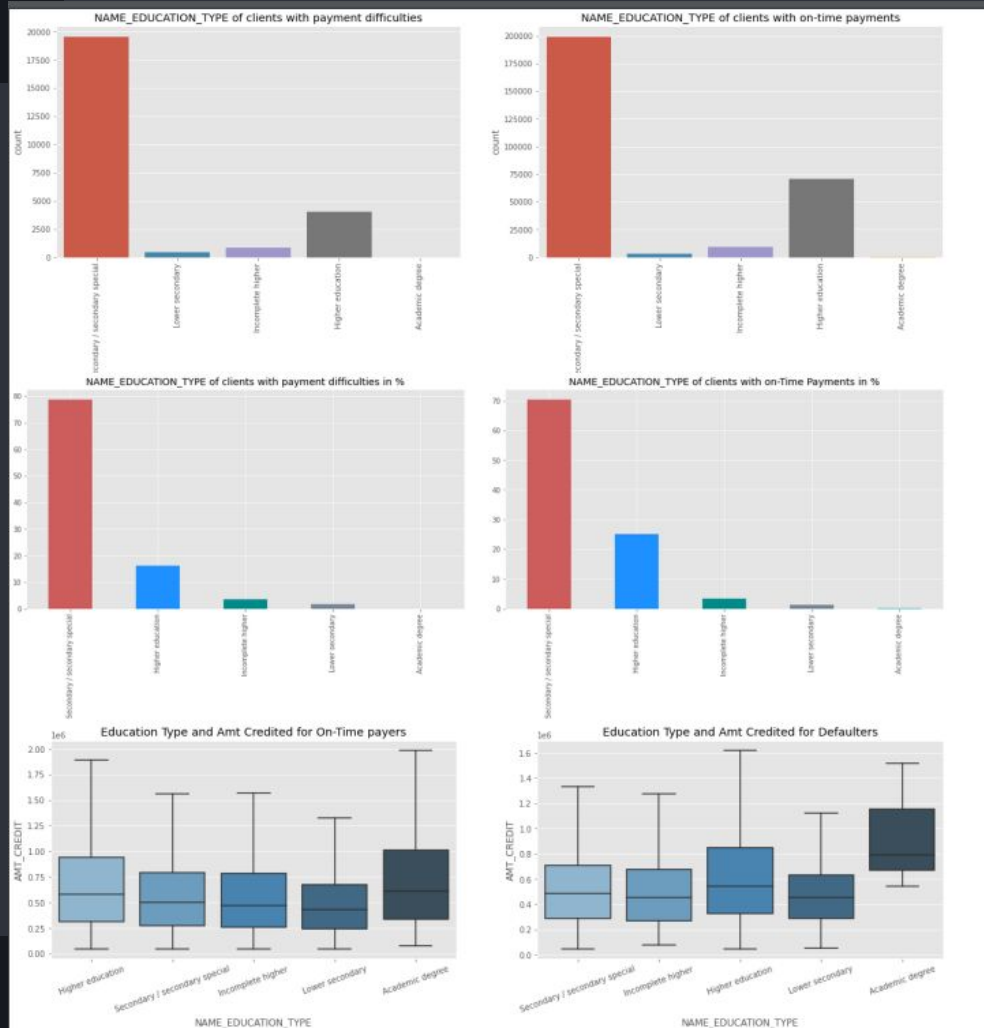


'Education Level' as a factor.

Those with 'Highest Education' are less likely to Default.

Applicants with Academic degree are less likely to Default.

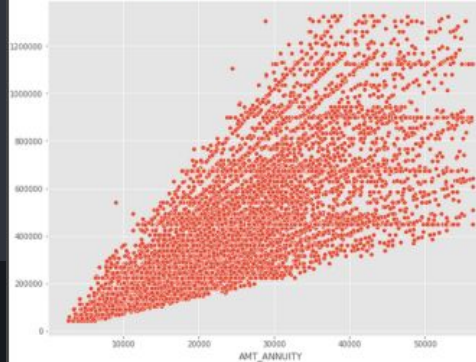
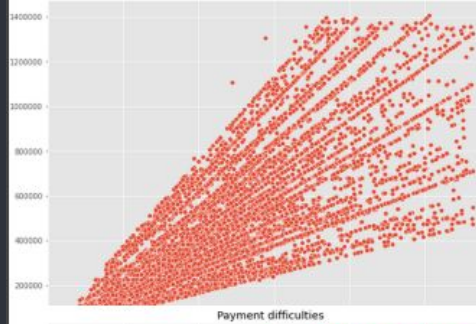
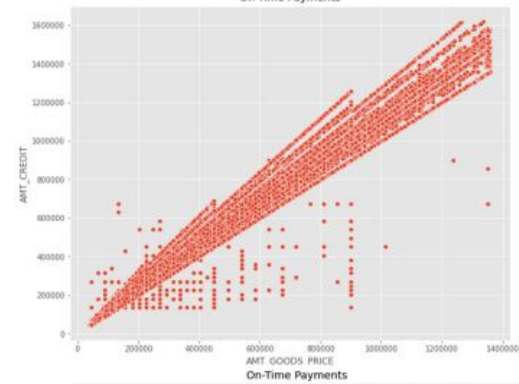
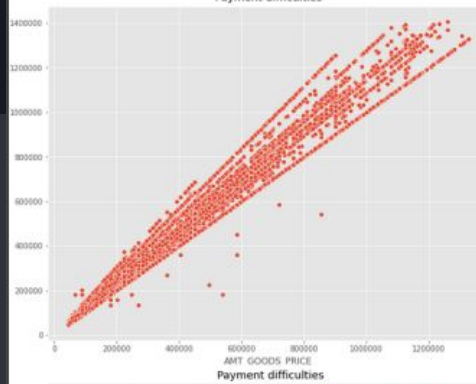
Clients with academic degrees should get higher preference



'CORRELATION' Between Variables

A strong positive correlation was found among the following variables

- AMT_ANNUIITY and AMT_GOODS_PRICE
- AMT_ANNUIITY and AMT_CREDIT
- DAYS_EMPLOYED and AMT_INCOME_TOTAL
- AMT_GOODS_PRICE and AMT_CREDIT
- CNT_FAM_MEMBERS and CNT_CHILDREN
- DAYS_BIRTH and DAYS_EMPLOYED

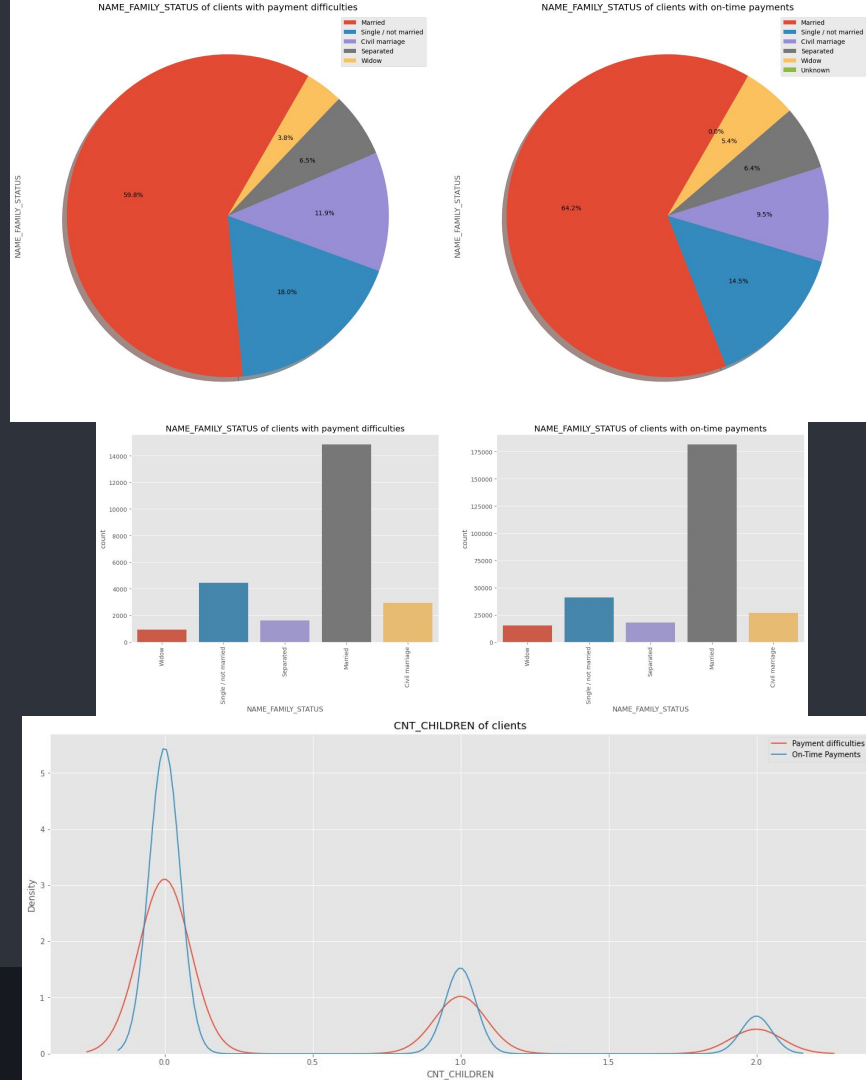


'Family Status' as a factor

People with no children are less likely to default.

Clients who are 'Single/not married' default the most.

Most applicants are 'Married' followed by 'Single' and then 'Not married'

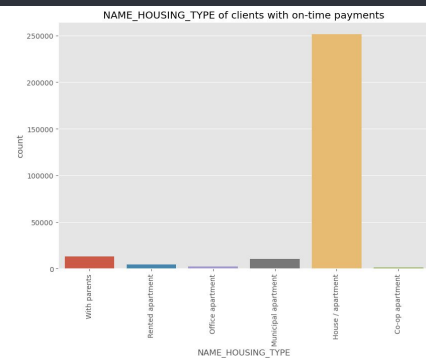
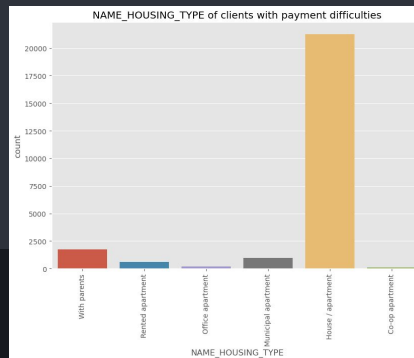
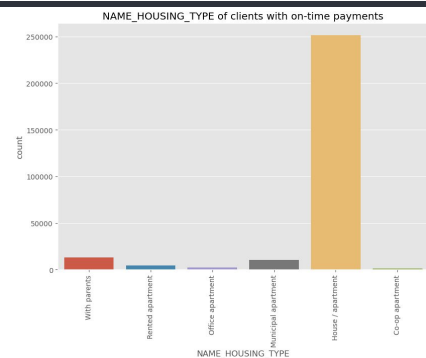
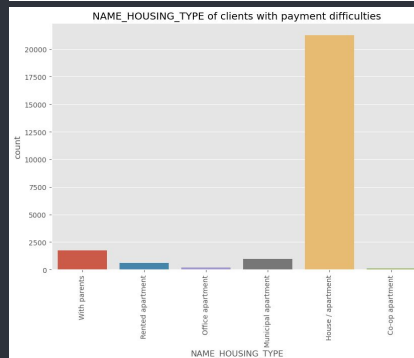
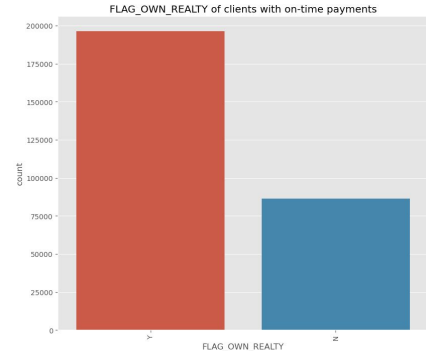
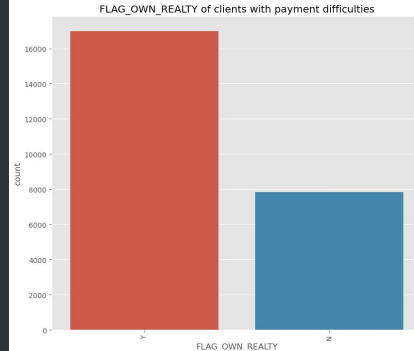


'Housing type' as a factor

Most applicants have a own House.

Those with an Own Home tend to make their payments on time.

People who live in rented homes or with parents tend to default more than those living in own homes or office flats.



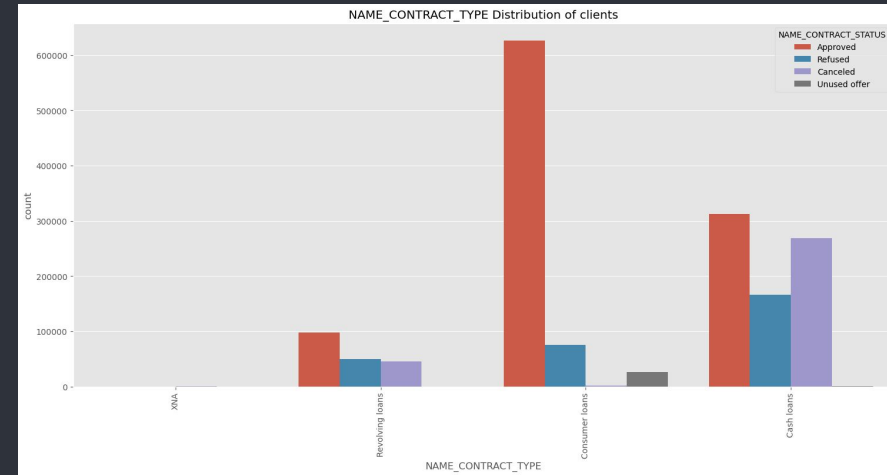
'Types of Loans' requested by customers

More Cash loans have been processed compared to revolving loans.

The most approved type of loan is consumer loan, followed by cash loan.

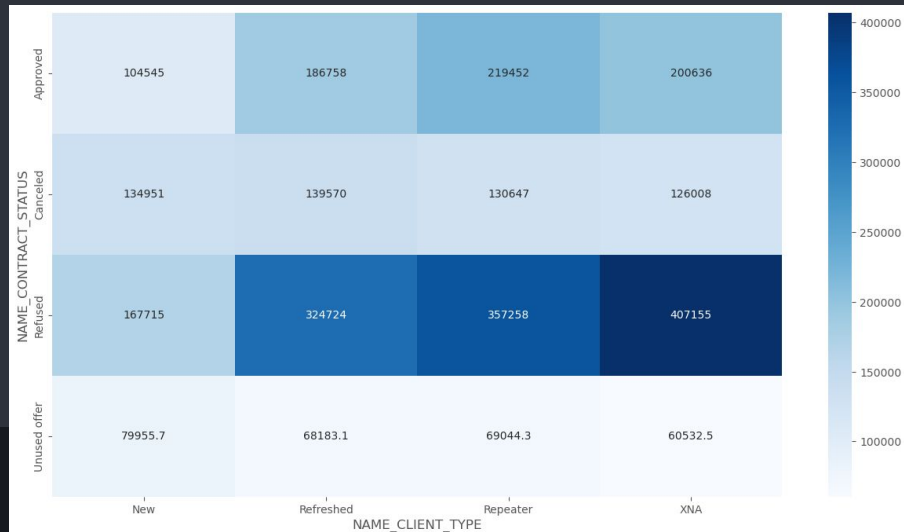
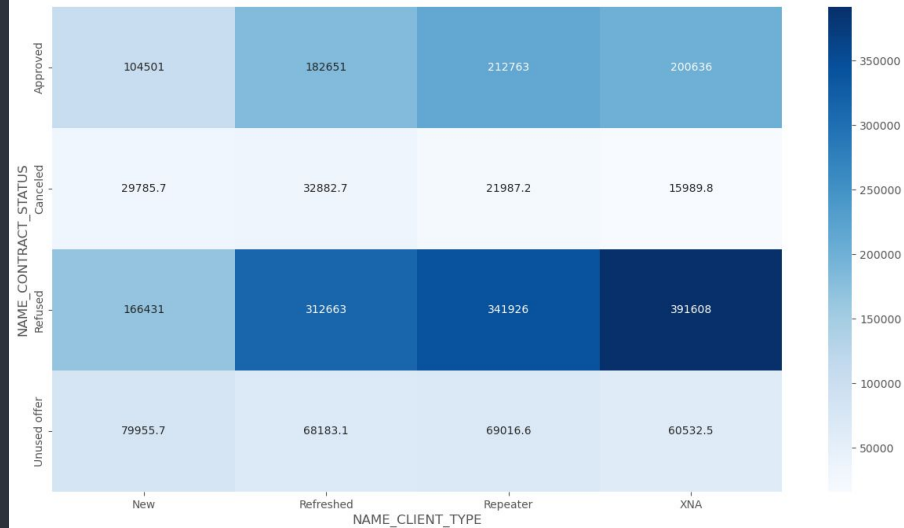
Revolving loan has a higher refusal rate.

More cash loans have been refused than consumer loans.



Reasons for 'Loan Cancellation'

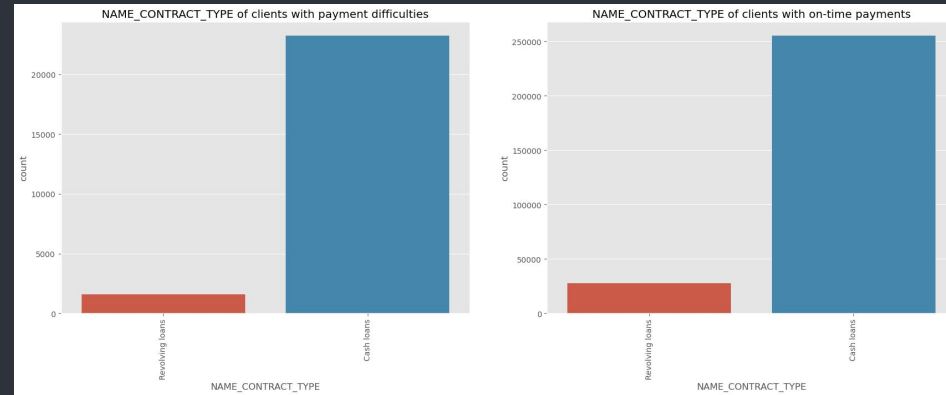
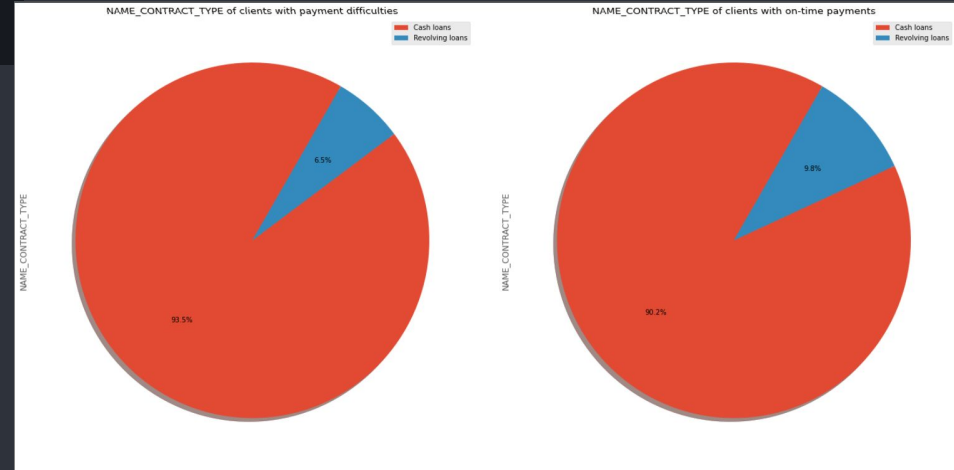
- Loans tend to be canceled when there is Low Unused offer CREDIT AMOUNT.
- Need to investigate why there is credit amount for cancelled and refused loans.
- All cancelled and refused cases have higher value of goods than other categories.



Types of 'Loans Disbursed'

- Most applicants opt for cash loans vs. revolving loans

- Defaults on cash loans are higher compared to those with revolving loans



Conclusion on;

{ 'Individuals more likely to default on their loans are'

- Those with Medium income
- Those in the Age group of 25-35 years, followed by 35-45 years
- Unemployed
- Laborers, salesmen, drivers
- Don't own house, live with parents, or rent flats

}

Conclusion on;

{ 'Individuals who we need to focus on are'

- Those with Previously cancelled, refused, unused loans..
- Students and businessmen with no repayment issues.
- Repeat clients with history of repayment.
- Clients with academic degrees.
- Those Employed for more than 19 years.
- Married individuals.
- Home owners.
- Pensioners.
- Female applicants.

}

```
{
```

```
1
```

```
2
```

```
3
```

```
4
```

```
5
```

```
6
```

```
7
```

```
8
```

```
9
```

```
10
```

```
11
```

```
12
```

```
13
```

```
14
```

```
}
```

Thanks;

```
{
```

‘Do you have any
questions?’

inderbaldawa@gmail.com
+91 8197381400