CREDIT EDA CASE STUDY

SUBMITTED BY:

RATIK KHANNA

SUBHANJAN ROY

Before Jumping to the Exploratory Data Analyse on Credit lets understand what are the factors involving while giving a loan

▶ Credit Risk: Credit risk is the possibility of a loss resulting from a borrower's failure to repay a loan or meet contractual obligations.

▶ Interest Risk -Interest rate may be reduced

▶ Credit History of the Borrower

▶ Cash flow history and projections for the business

▶ Collateral available to secure the loan

▶ Borrower's occupation and organization status

▶ Income/Education of the borrower

Business Understanding

▶ As we have already discussed on 2nd slide that the factors responsible while giving a loan. Loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter.

▶ So here in this Case Study we will discuss on the parameters which will help us to identify proper customer where we can reduce the risk of loan default.

▶ We are provided with 3 data sets i.e., *application_data, previous_application & columns_description. This will help us to analyse the same.*

# The Following approach is used in doing while Exploratory Data Analysis

▶ **Data understanding-** We first understood the data by viewing its size, shape, null values and important Parmenter (such as mean, median, etc.)

▶ **Preparation and Data Cleaning-** In the cleaning process, the main objective is to remove irregularities from a data set. There are many ways to clean data, but the two most important approaches that you learnt as part of the cleaning step are treatment of missing values and outlier handling.

▶ **Treatment of missing values and outliers** – We treated missing values by removing an entire column with more than 35% missing values and filling with other values like (NaN)

▶ **Data analysis-** Here we used graphs and made Univariate analysis Bivariate and multivariate analysis of various columns and variables to understand relation and correlation of different variables of dataset

▶ **Presentation and Recommendations-** Finally we presented all the findings of analysis in a ppt with our recommendation based upon our analysis.
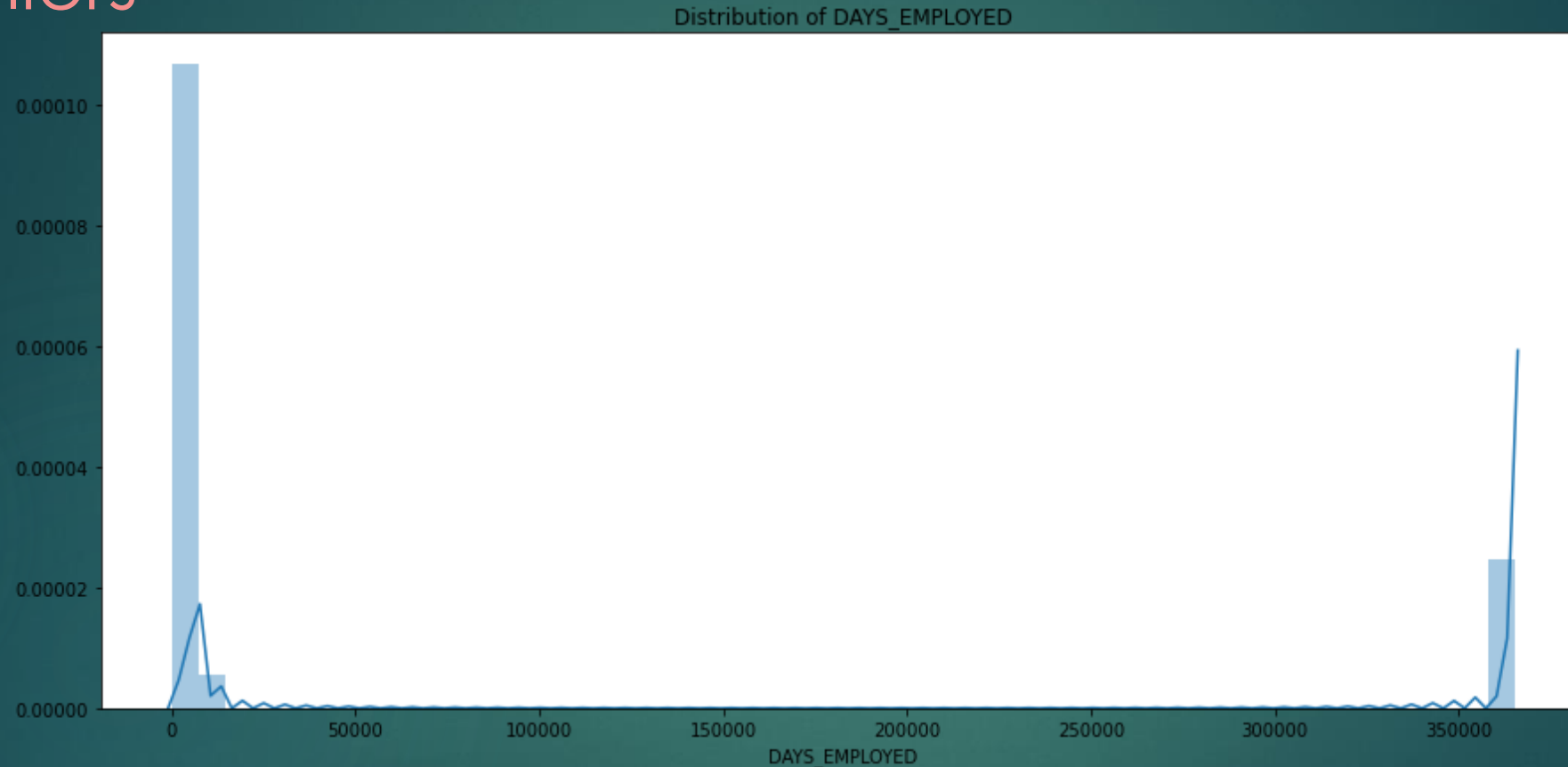
Days Birth in Box Plot to check the outliers

Distribution of Days of Birth
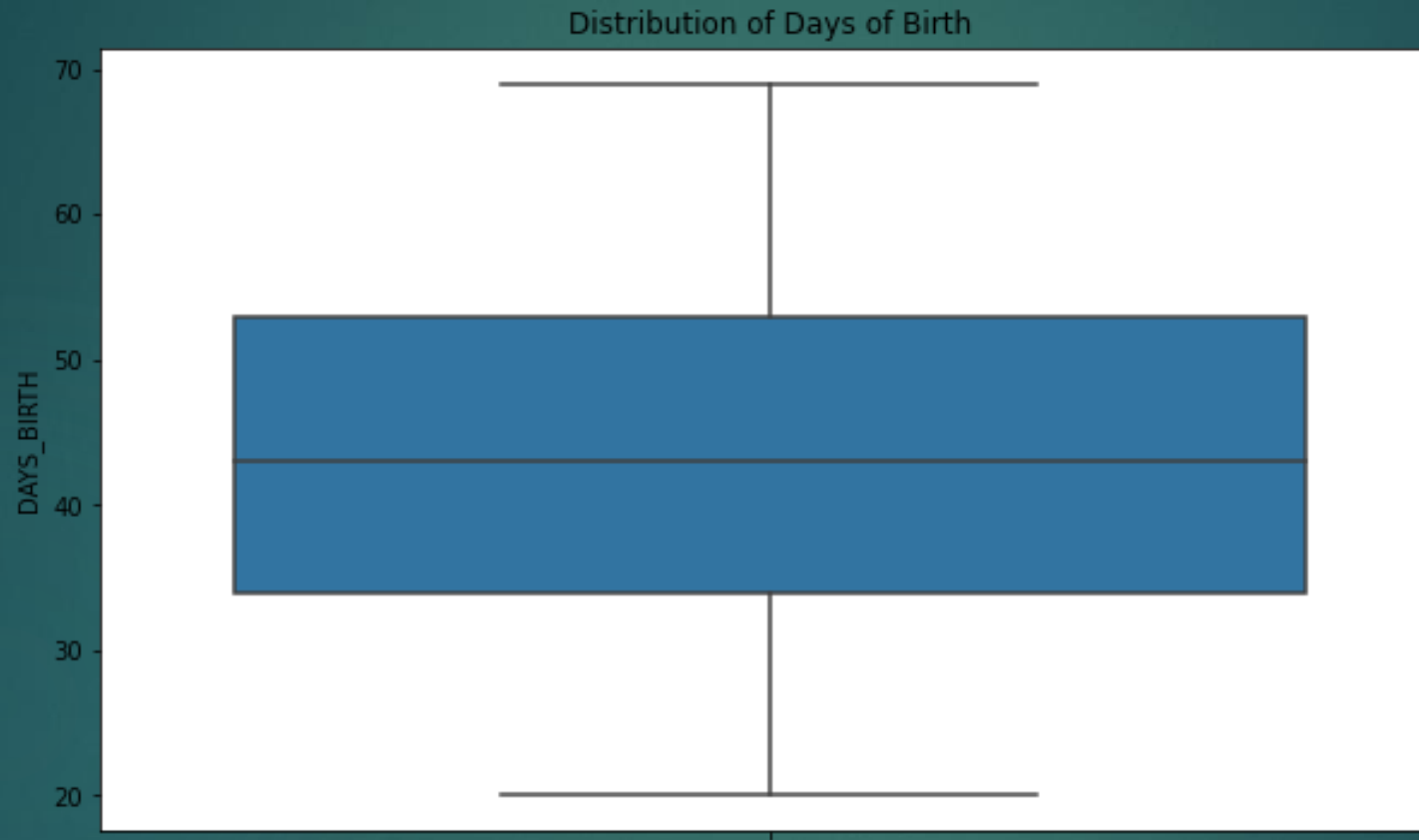
There are no Outlier Present

Annuity Amount & Registration Days in Box Plot to check the outliers
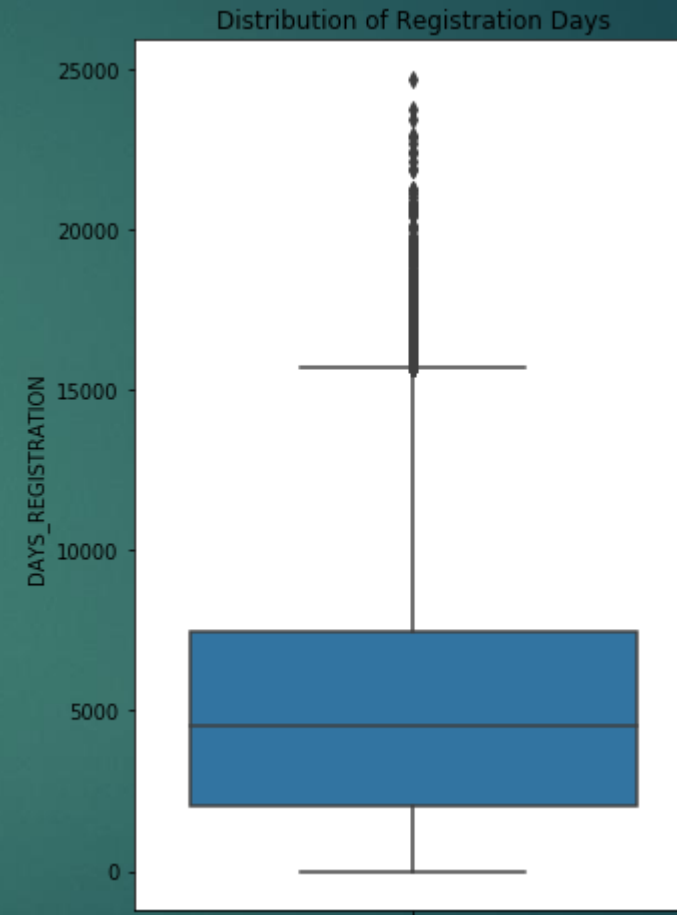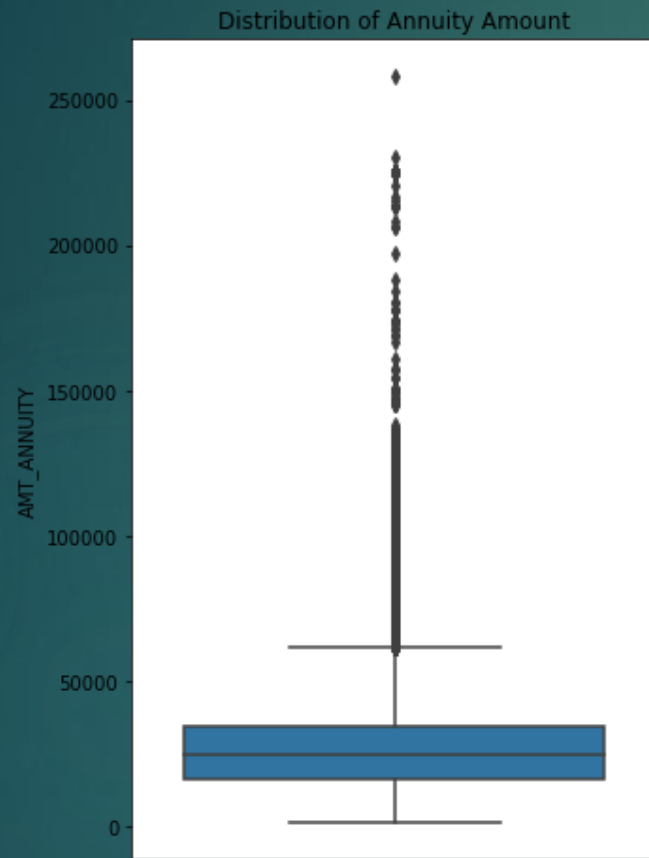
Outliers Present in both the column

# Days Employed in Box Plot to check the outliers

Distribution of DAYS_EMPLOYED

The above plot shows a person's emploment in a company for days before applying the loan. What we can see here is the value greater than 350000 is clearly an outlier. If we convert it to years i.e., 350000/365 is 958 years which is impossible. So practically if a person starts the employment in a company at 20years of age than the person can work only for next 40years attending 60years maximum employment tenure. So if we convert 40years to days than it will come approx 15000Days. So values greater than 15000 days is clearly an outliers.

Lets Plot the Target Distribution with payments
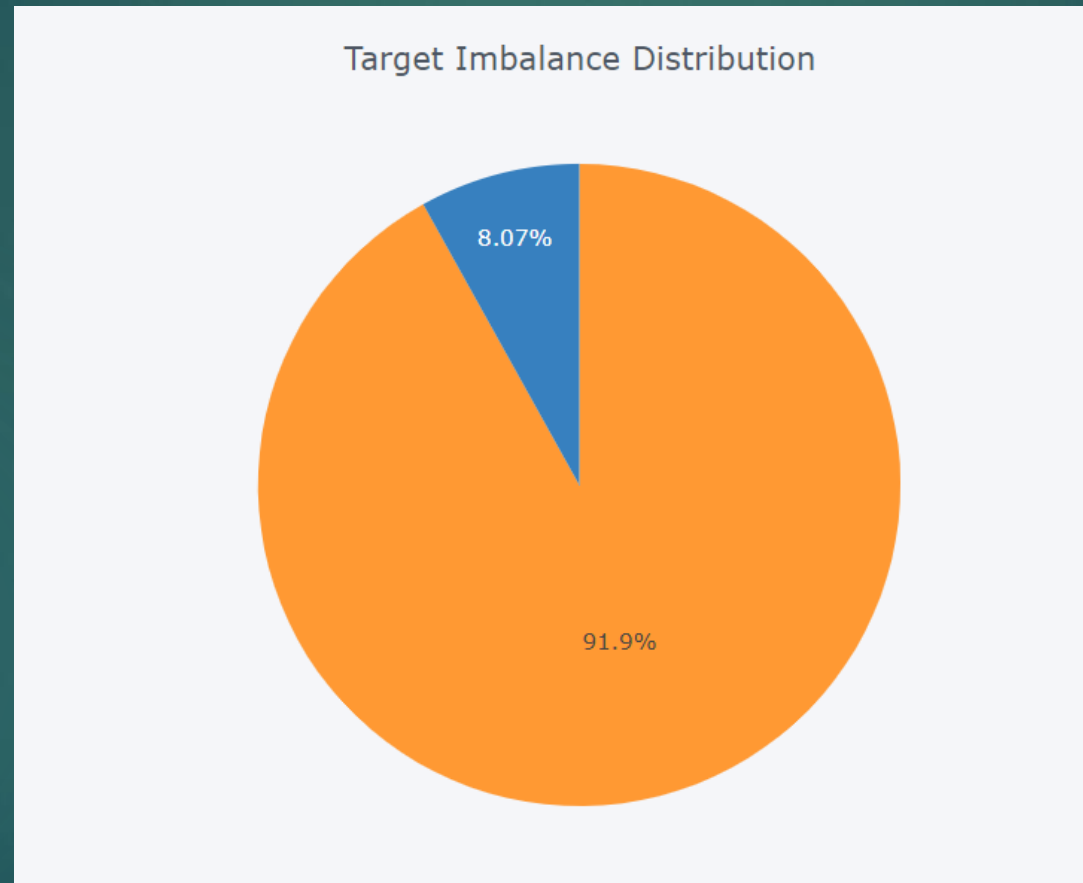
This is highly Imbalanced Dataset as approx. 92% data belong to target value of 0 and only approx. 8% belong to target value of 1. & 92% od people have no payment difficulties whereas 8% has payme difficulties.
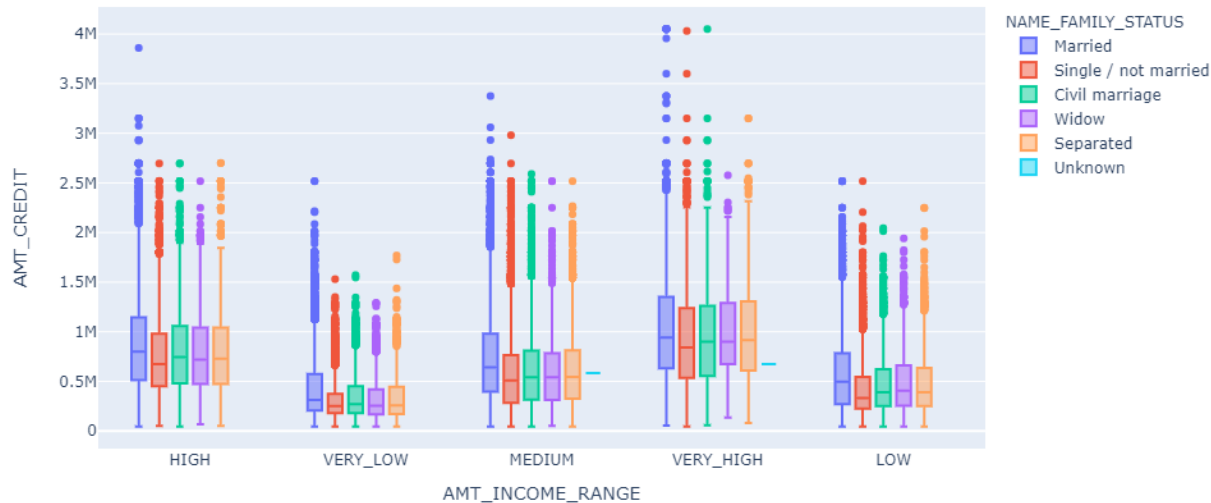
Types of Organizations applied loan, having no loan Payment Difficulties
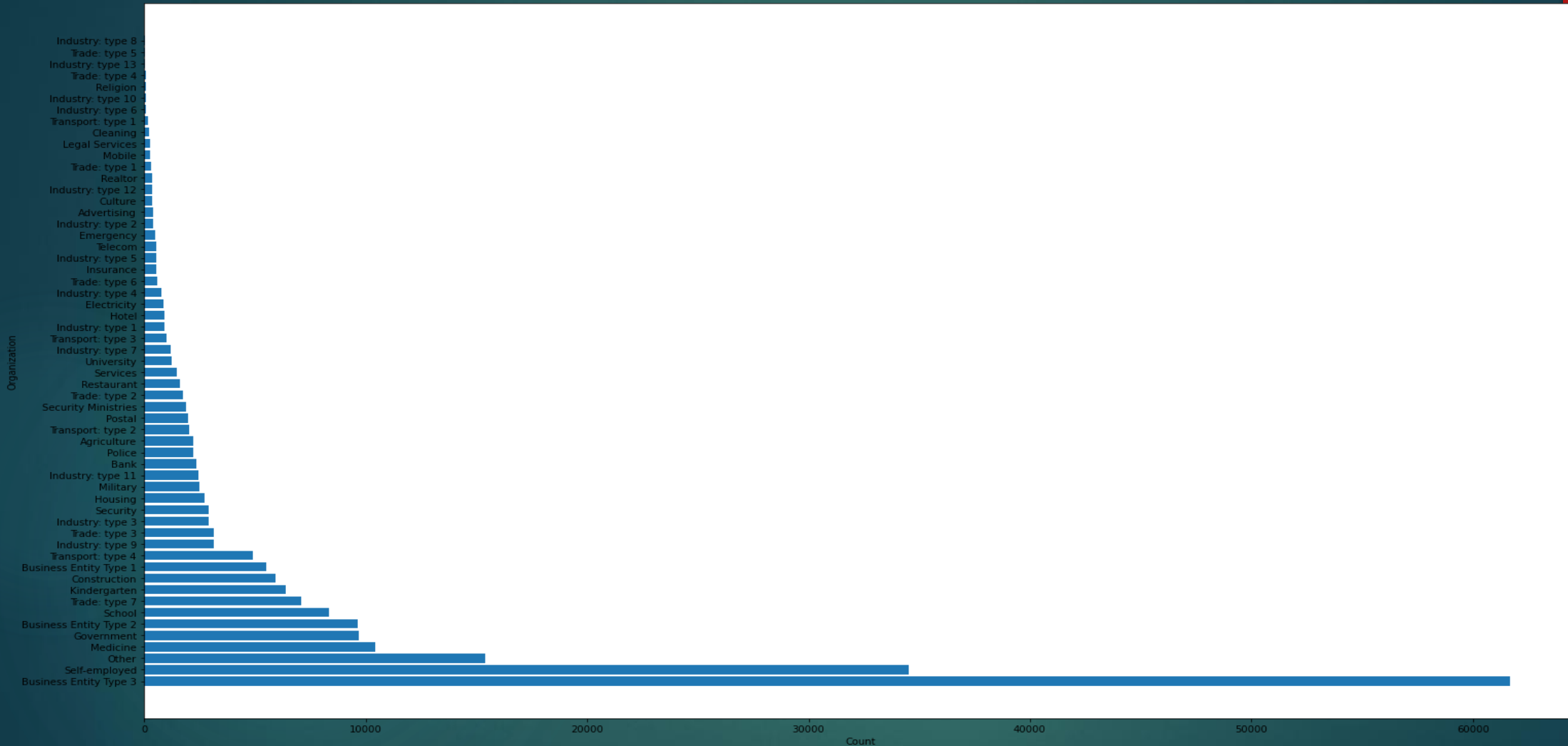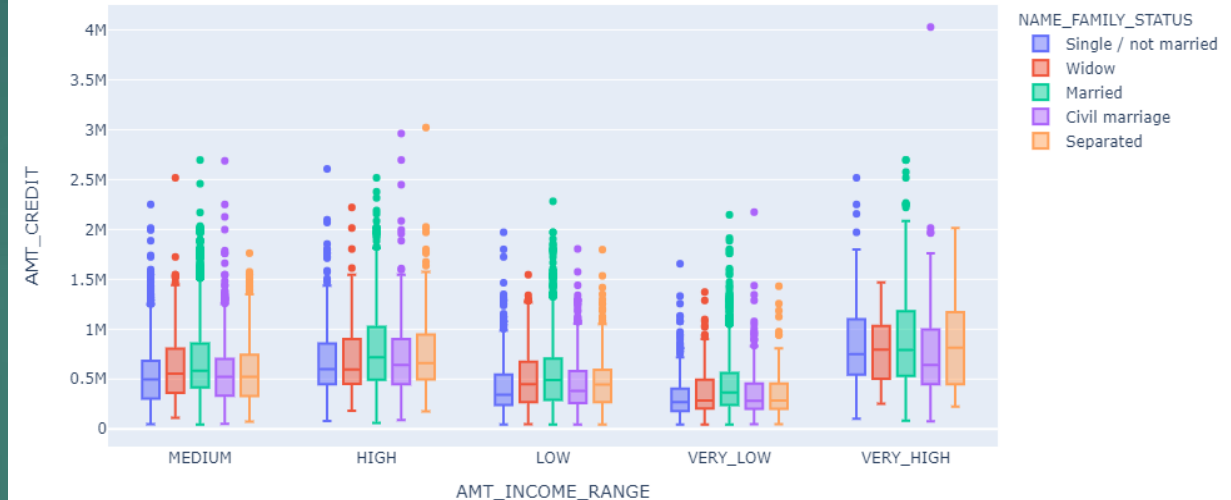

Business Entity 3 have majority of loan application followed by Self employeed and other.

Income Range Vs Credit Amount for Payment difficulties and Non Payment difficulties

Left Side Box Plot: The above plot shows Comparison between Income Range and Credit amount of applicants with No Loan Payment Difficulties. All the box plot has outliers. And people who are having very high income have higher credit amount.

Right Side Box Plot: The above plot shows Comparison between Income Range and Credit amount of applicants with Loan Payment Difficulties. All the box plot has outliers except very high(Widow) and very high(Separated).

Conclusion: Those who are having higher income always avail higher loan amount. And Married People have higher income. But Married people also have difficulty in payment of loan.

Univariate analysis of AMT_ANNUITY of applicates with no loan payment difficulties
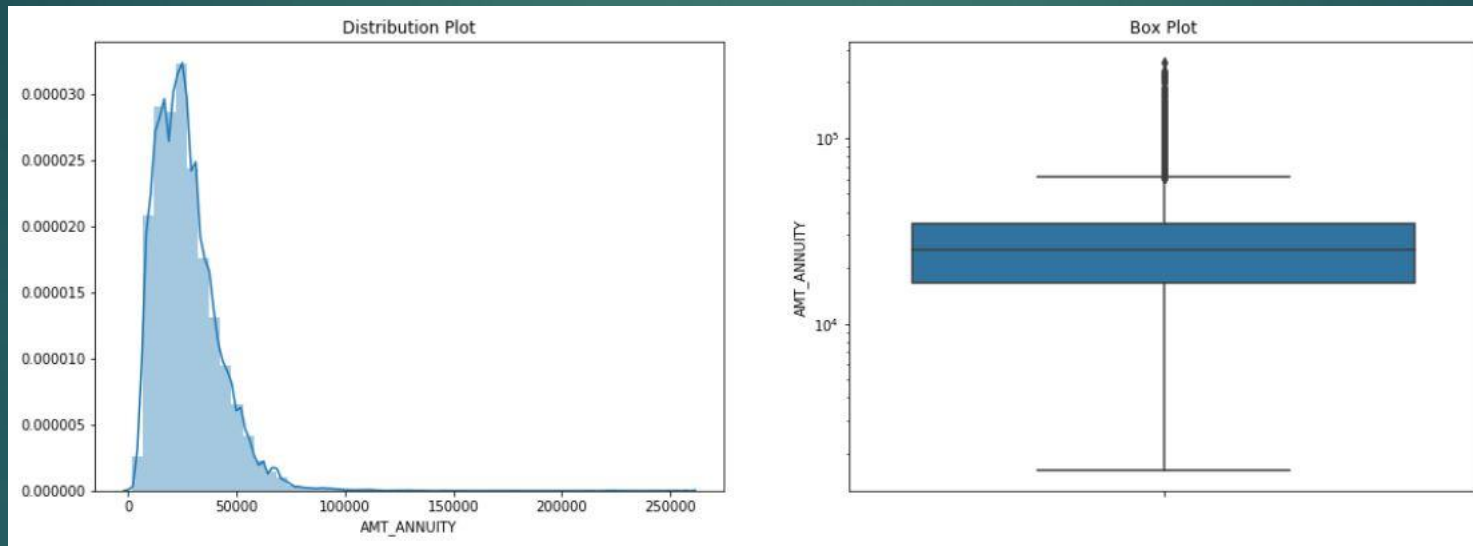
The above plot distribution of Annuity amount and box Plot for annuity amount where people have no payment difficulties. Where we can see mostly annuity amount ranges from 0 to 100000. And the density is very high from 10000 to 50000. In the Box Plot for Annuity amount above has outlier.

Univariate analysis of AMT_ANNUITY of applicates with loan payment difficulties
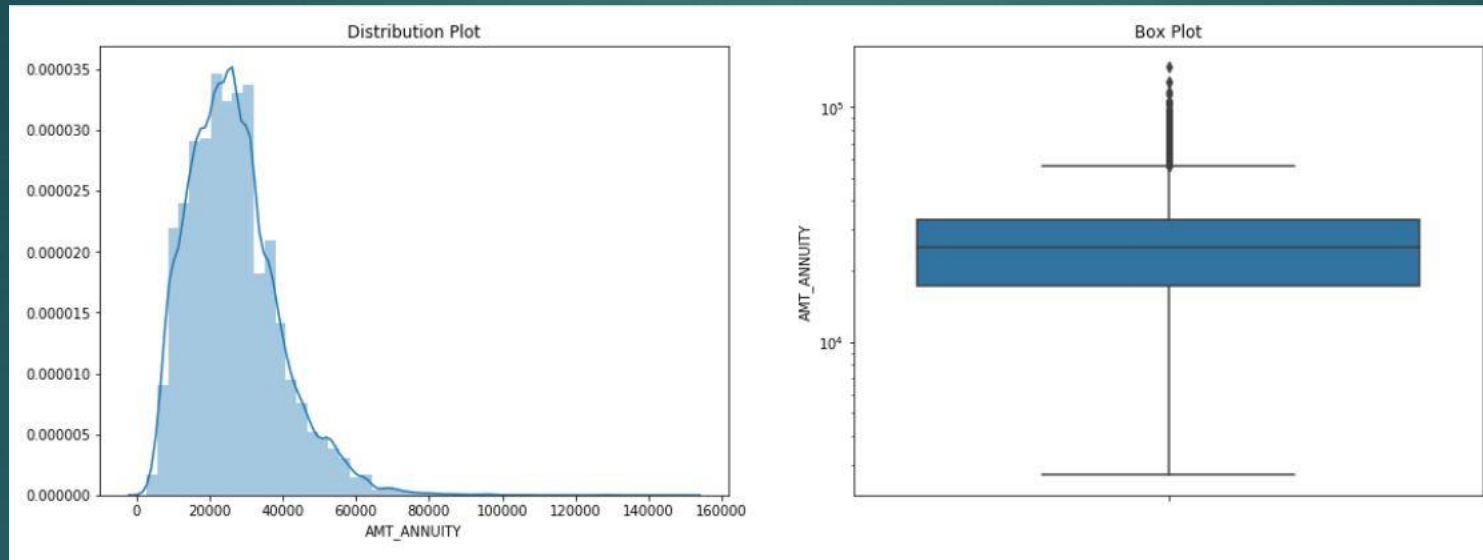

The above plot distribution of Annuity amount and box Plot for annuity amount where people have payment difficulties. Where we can see mostly annuity amount ranges from 0 to 80000. And the density is very high from 10000 to 50000. In the Box Plot for Annuity amount above has outlier.

bivariate analysis of income type and loan payment difficulties with each category
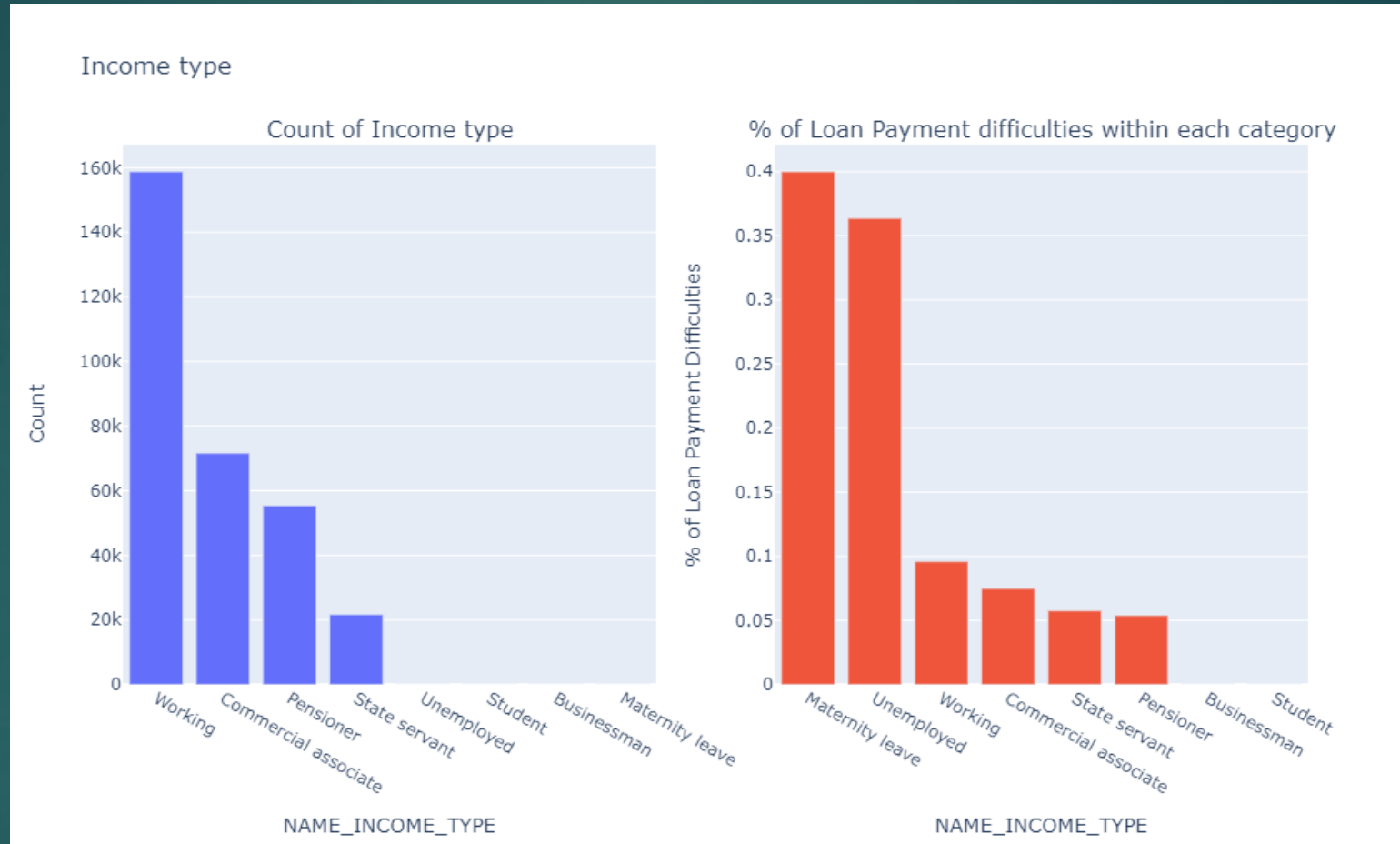
The above plot shows bivariate analysis of income type and loan payment difficulties with each category. Here wecan see income are higher for working followed by commercial associate then pensioner. But payment difficulties mostly seen in maternity leave followed by unemployed than working class.
Conclusion: We can conclude here is people who are earning having income have better loan repayment capacity compared to Maternity leave and unemployed one.
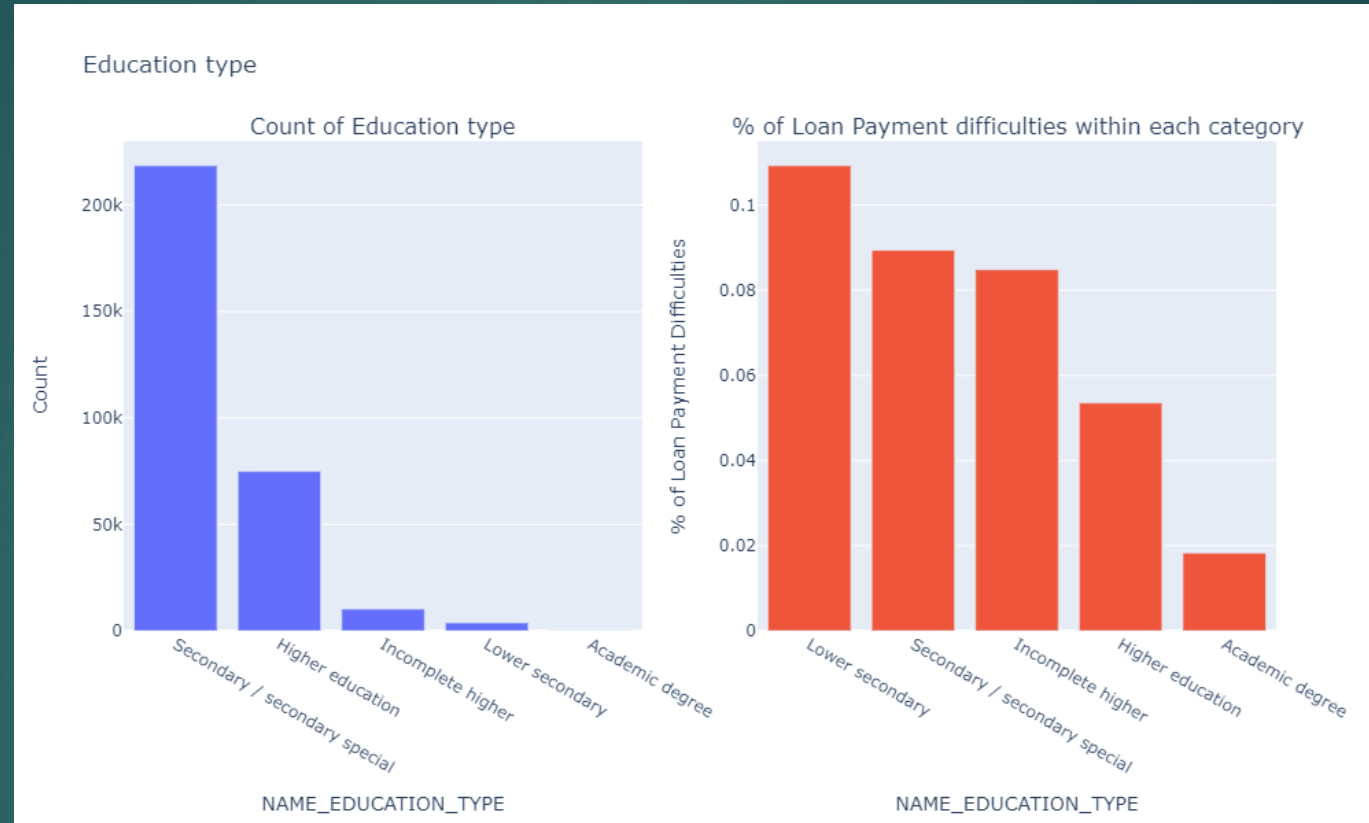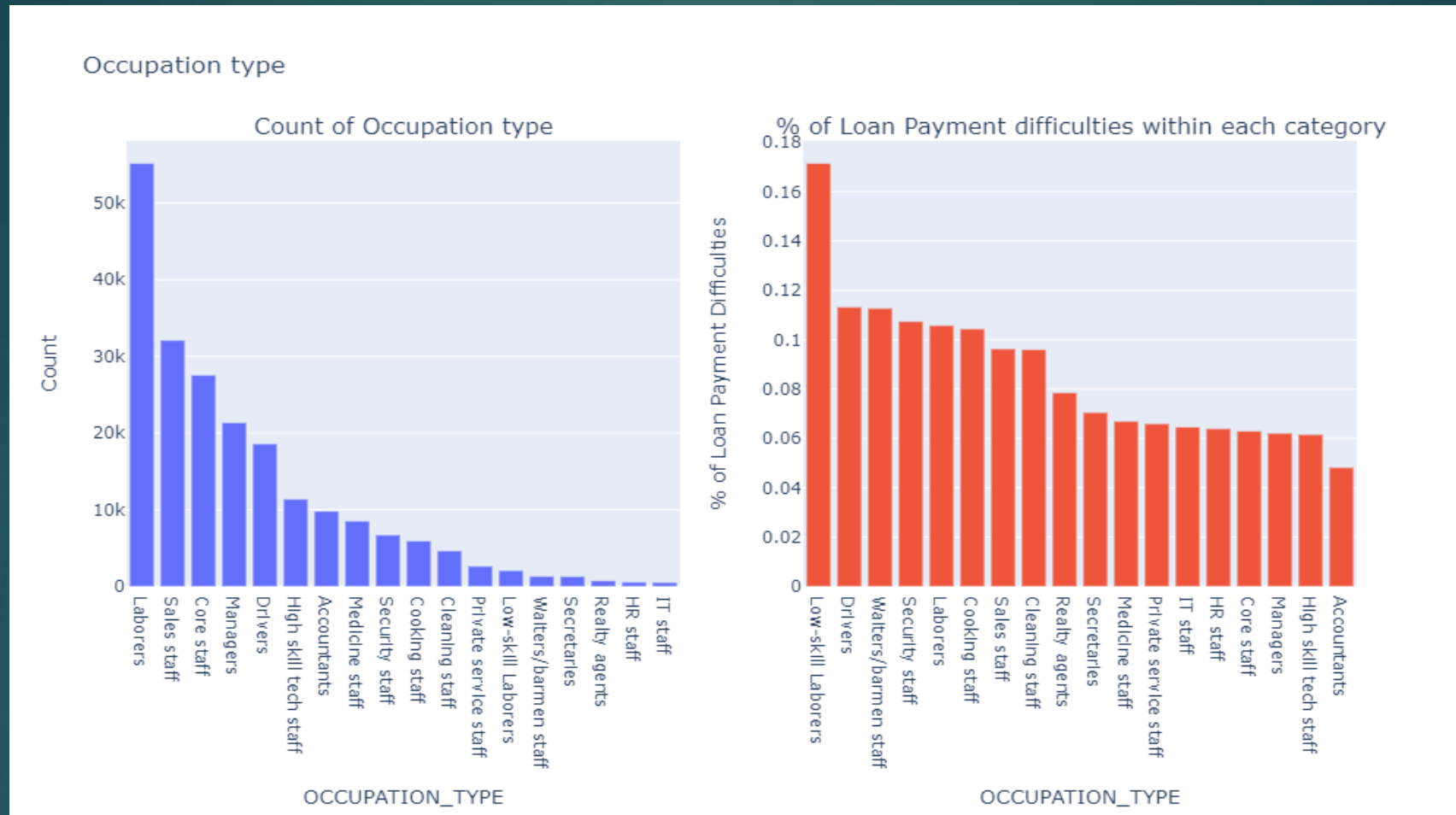
bivariate analysis of education type and loan payment difficulties with each category

The above plot shows bivariate analysis of education type and loan payment difficulties with each category. Here we can see income are higher in Secondary followed by higher education than incomplete. But Loan Payment difficulties are higher in lower education followed by secondary than incomplete.
Conclusion: Its better to give loan to people having higher education, academic degree and secondary education as they have better loan repayment history.

bivariate analysis of occupation type and loan payment difficulties with each category

The above plot shows the bivariate analysis of occupation type and loan payment difficulties with each category. Here majority of people are employed as labourer followed by sales staffs than core staff. And loan payment difficulties are face maximum by Low skilled laborers followed by drivers than waiters.

Conclusion: We can conclude here is It is risky to give loan to low skill laborers, Drivers. So Best option here are High skill tech staffs and accountant.

Contract status of previous application
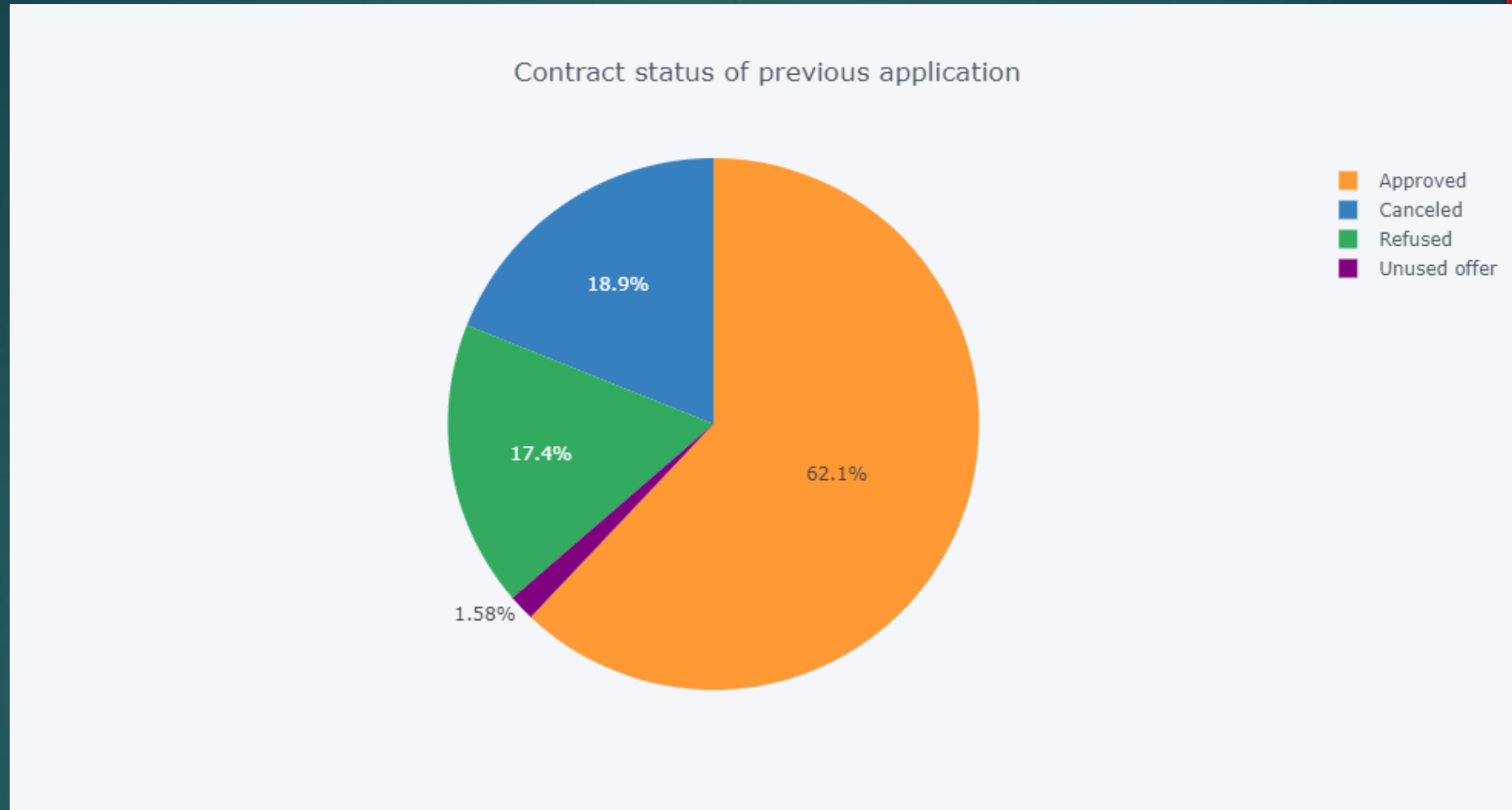
The above plot shows the Contract status of previous application. Here we can see that 62% loans have approved followed by 19% loans were cancelled than 17.4% loans were refused by the financial institute. Conclusion: Contract status of Previous application Category plays a vital role while checking the present loan application. If a client's loan got rejected previously than its give the financial institute a second check on the client's application.

Kind of goods did the client apply for in the previous application


The above plot shows Kind of goods did the client apply for in the previous application. Here we can see maximum number of loan application were for Mobile followed by Consumer Electricals, computers, Audio/Video, Furniture.

Conclusion: Kind of goods did the client apply in prev application gives a clear picture what are the goods where demands are available.

Reasons of previous application rejection

Reasons of previous application rejection
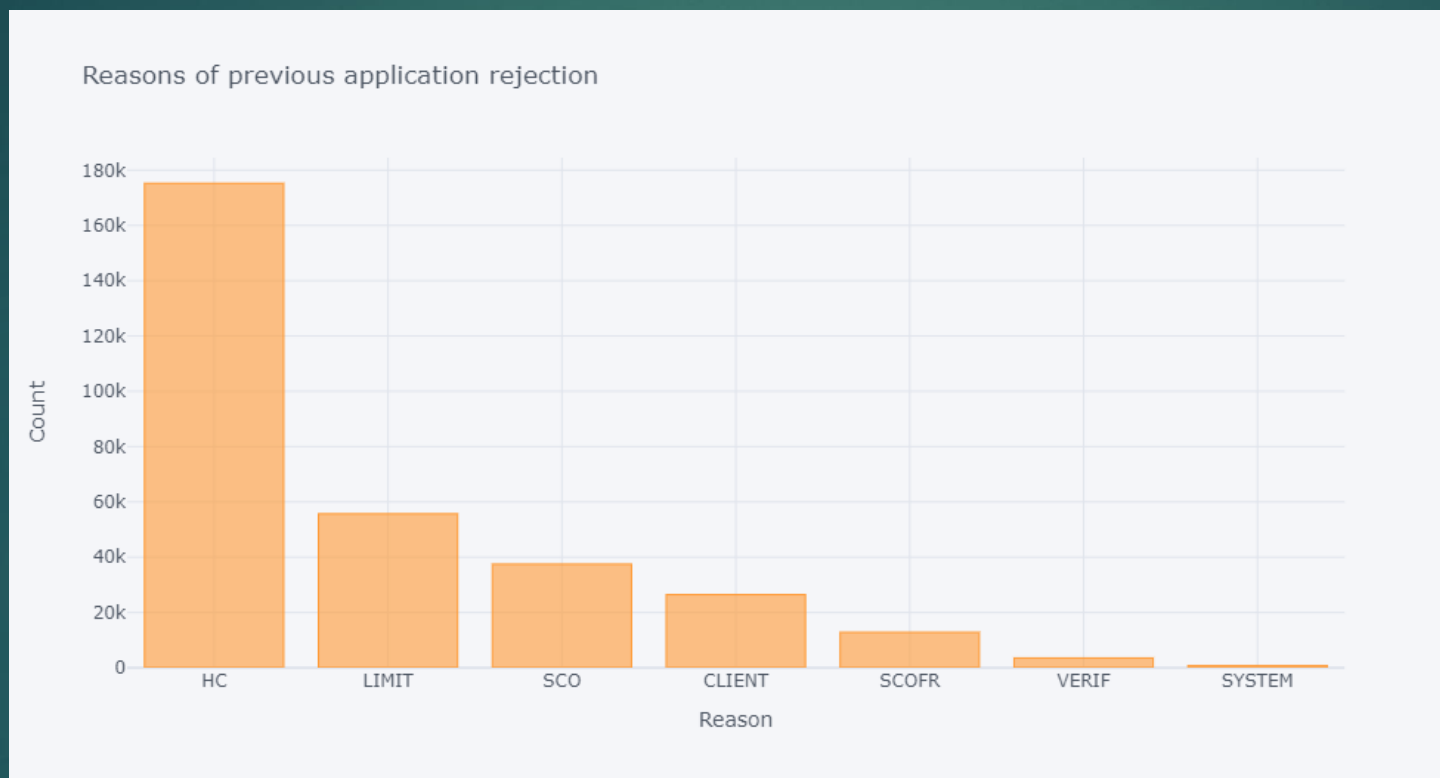
The above plot shows the Reasons of previous application rejection. Here we can see maximum rejection is due to HC followed by Limit than SCO. So Loan should not be given to HC, LIMIT and SCO

Top 10 correlation for the Client with payment difficulties

	VAR1	VAR2	CORRELATION	CORR_ABS
57	AMT_CREDIT	AMT_GOODS_PRICE	0.983103	0.983103
17	AMT_ANNUITY	AMT_GOODS_PRICE	0.752699	0.752699
58	AMT_CREDIT	AMT_ANNUITY	0.752195	0.752195
35	DAYS_BIRTH	DAYS_EMPLOYED	0.582441	0.582441
52	DAYS_REGISTRATION	DAYS_BIRTH	0.289116	0.289116
44	DAYS_ID_PUBLISH	DAYS_BIRTH	0.252256	0.252256
43	DAYS_ID_PUBLISH	DAYS_EMPLOYED	0.229090	0.229090
51	DAYS_REGISTRATION	DAYS_EMPLOYED	0.192455	0.192455
33	DAYS_BIRTH	AMT_GOODS_PRICE	0.135603	0.135603
60	AMT_CREDIT	DAYS_BIRTH	0.135070	0.135070

Here we can see Goods Price have better correlation with amount credit followed by amount credit and good price than amocredit and amount annuity.
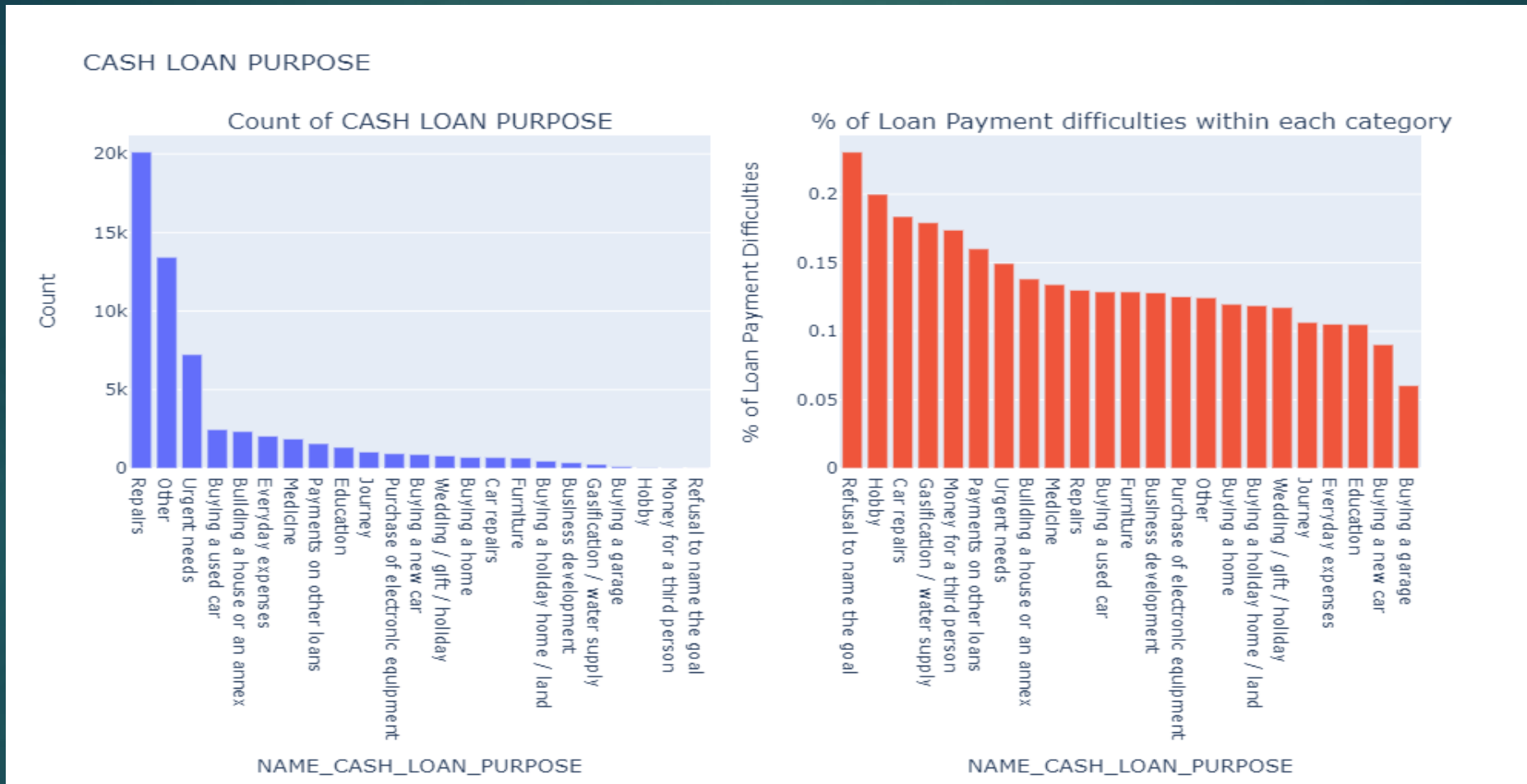
The count of contract status and loan difficulties with in each category

The above plot shows the count of contract status and loan difficulties with in each category. Here we can see that maximum loan application cases are approved than cancelled than Refused. But loan difficulties are mainly faced by Refused, cancelled and unused offer cases.

Conclusion: Loan approved cases have faced lesser payment difficulties which is actually a good sign for giving a loan.

The cash loan purposes and loan difficulties within each category

The above plot shows the cash loan purposes and loan difficulties within each category. Here we can see maximum loans are availed for repair followed by others than Urgent need. Loan difficulties are faced by Refusal to name the goal followed by hobbby than car repairs.

Conclusion: Its very much clear that if a borrower does not have a clear goal for a loan it will end up in loan payment difficulty. Likewise Hobby. It is better to give loan for buying a used car, Garage, Education.

the contract type and loan payment difficulties in each category
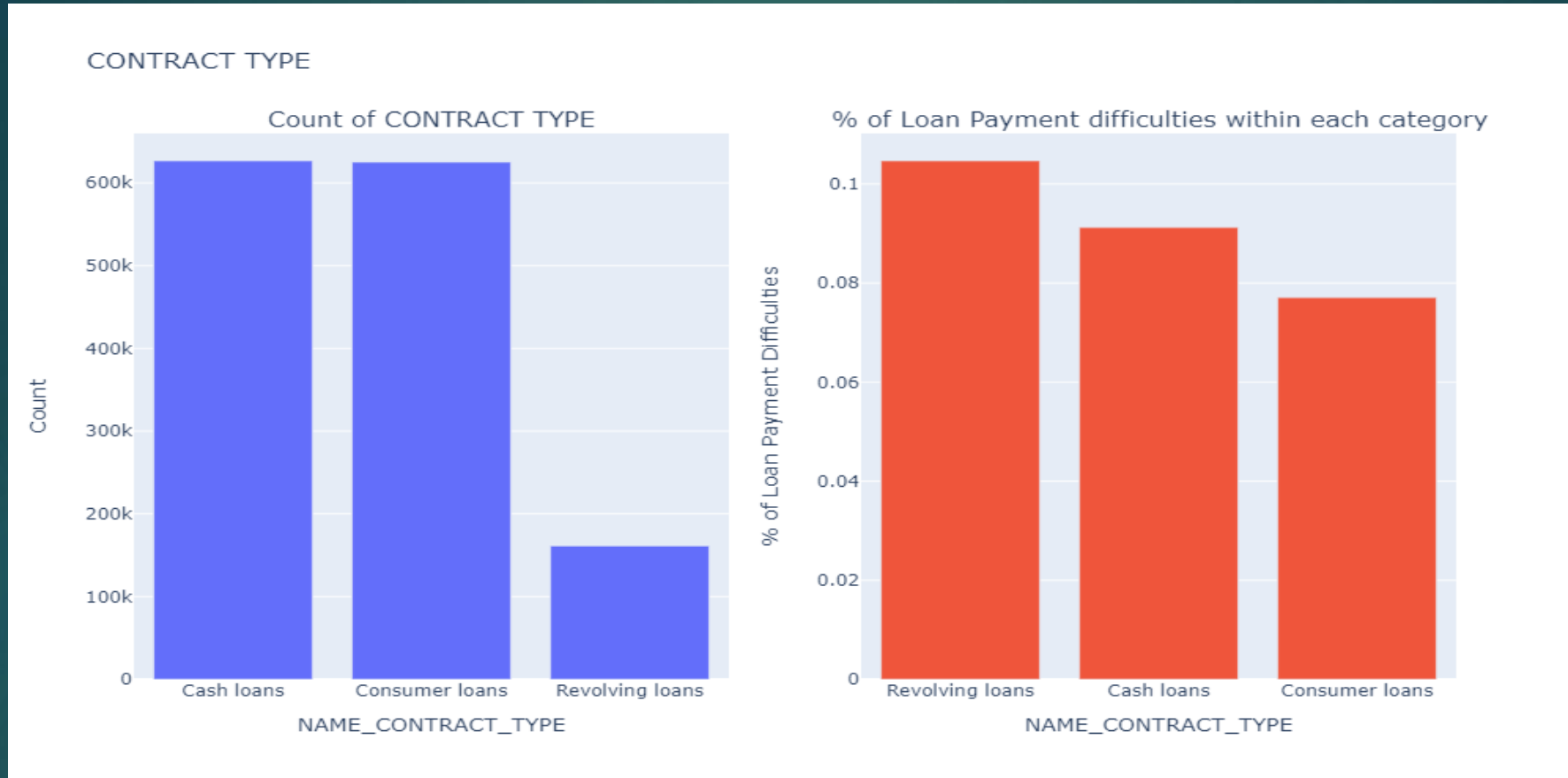
The above plot shows the contract type and loan payment difficulties in each category. Here we can see loans are mainly availed cash loan and consumer loan. And payment difficulties are seen in Revolving loan and Cash loan.

Conclusion: It is better to give consumer loan and cash loan as its repayment difficulties are much better than revolving loan.

Conclusion/ Recommendation

- 92% od people have no payment difficulties whereas 8% has payment difficulties.

- Those who are having higher income always avail higher loan amount. And Married People have higher income. But Married people also have difficulty in payment of loan.

- We can conclude here is people who are earning having income have better loan repayment capacity compared to Maternity leave and unemployed one.

- Its better to give loan to people having higher education, academic degree and secondary education as they have better loan repayment history.

- We can conclude here is It is risky to give loan to low skill laborers, Drivers. So Best option here are High skill tech staffs and accountant.

- Contract status of Previous application Category plays a vital role while checking the present loan application. If a client's loan got rejected previously than its give the financial institute a second check on the client's application.

- Kind of goods did the client apply in prev application gives a clear picture what are the goods where demands are available.

- Loan approved cases have faced lesser payment difficulties which is actually a good sign for giving a loan.

- Its very much clear that if a borrower does not have a clear goal for a loan it will end up in loan payment difficulty. Likewise Hobby. It is better to give loan for buying a used car, Garage, Education.

- It is better to give consumer loan and cash loan as its repayment difficulties are much better than revolving loan.