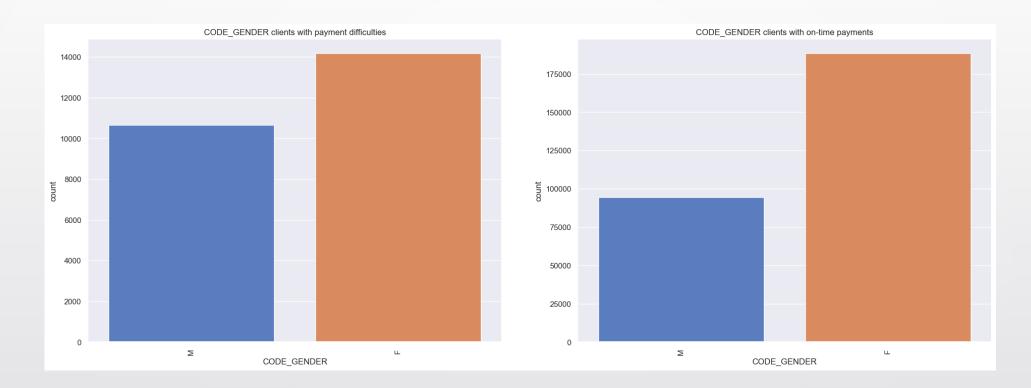
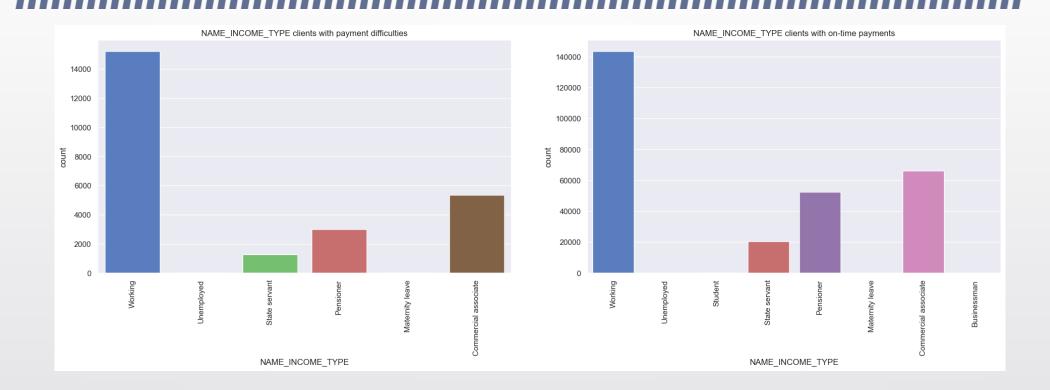
CREDIT EDA CASE STUDY

Patel Rushabh B

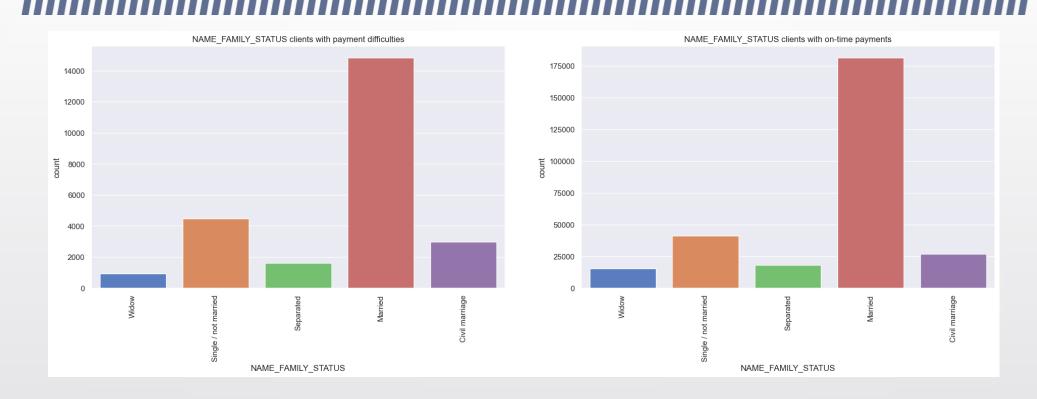
Univariate Analysis of Categorical Variable



 CODE_GENDER column provide that Male client have more payment difficulties than Female, which is weak inference.

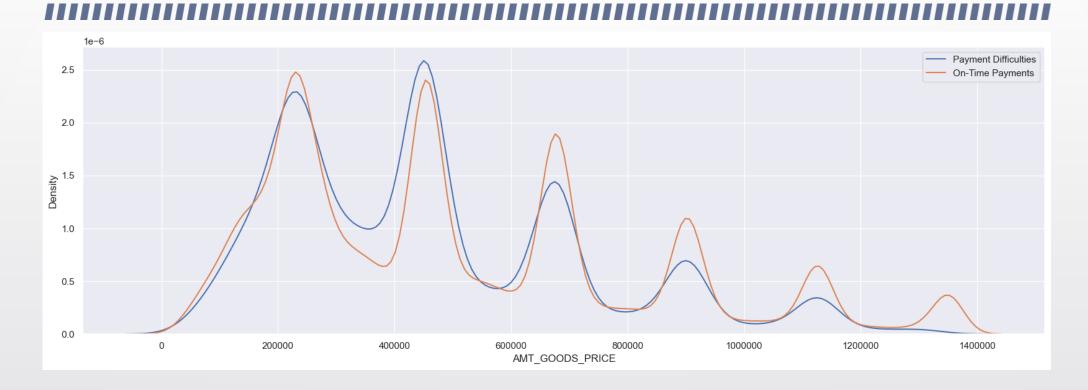


- Student and Business man has no payment difficulties.
- Student have 18 and business man have 10 observations. This is weak correlation.
- Pensioners have more on-time payments.

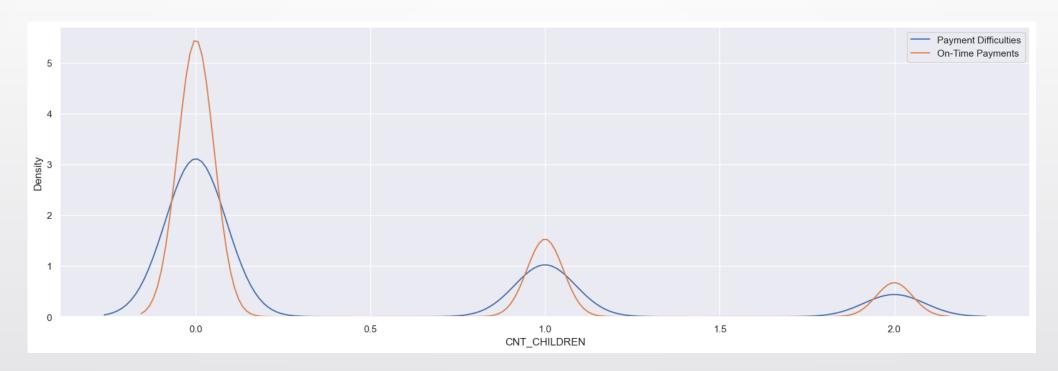


- Married and Widow do on-time payments. But there is no correlation between this.
- Single / Not Married have more difficulties for on-time payments.

Univariate Analysis of Numerical Variables



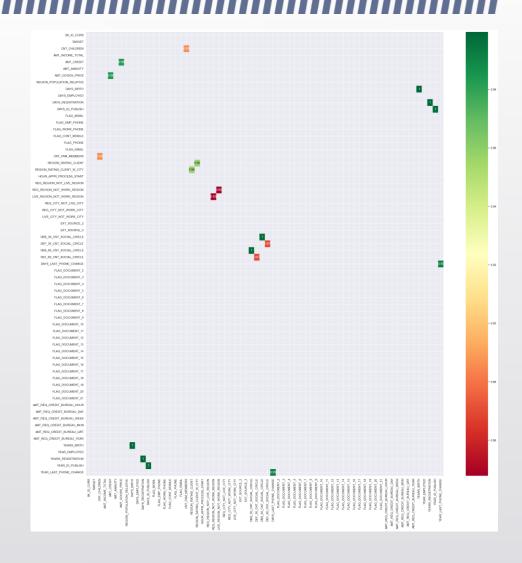
AMT_GOODS_PRICE between 2,50,000 and 5,50,000 there are more clients with Payment difficulties.



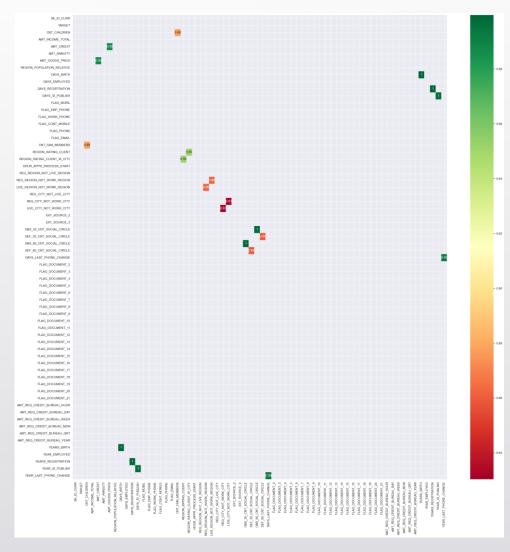
- For CNT_CHILDREN 0 has on-time payments.Those who has 1 or more than that Children also do on-time payments.

Correlation Analysis of Numerical Variables

Correlation Matrix with Payment Difficulties



Correlation Matrix with On-Time Payments

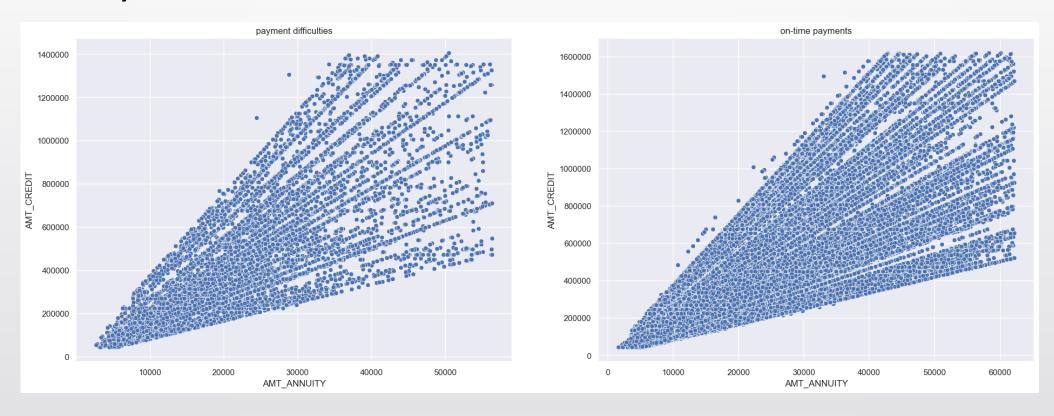


Comparing Correlations between Payment Difficulties and On-Time Payments

- Most of the correlations are same in both the cases except few minor differences.
- AMT_CREDIT & AMT_GOODS_PRICE combination has the highest correlation.
- In Payment difficulties data, highest correlation is between AMT_GOODS_PRICE & AMT_CREDIT for 0.98%
- In On-Time Payment data, highest correlation is between AMT_GOODS_PRICE & AMT_CREDIT for 0.99%

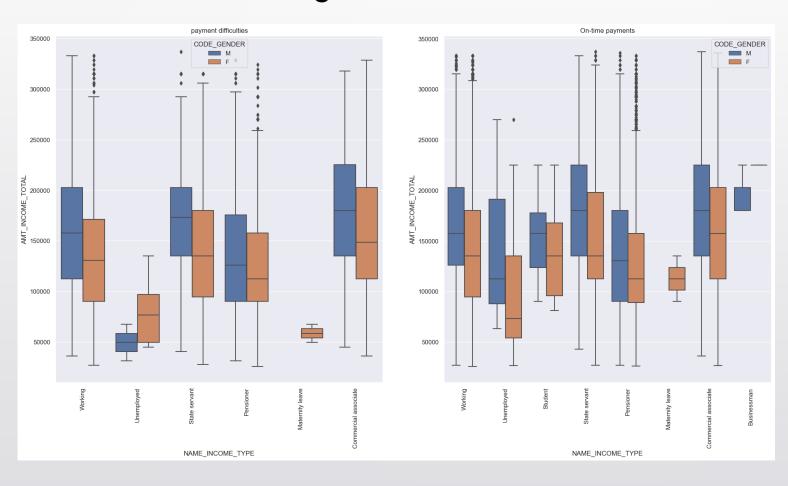
Multivariate Analysis

Analysis of Continuous Vs Continuous variable



- It's a strong correlation between AMT_ANNUITY and AMT_CREDIT.
- As annuity price increases along with that credit amount also increases.

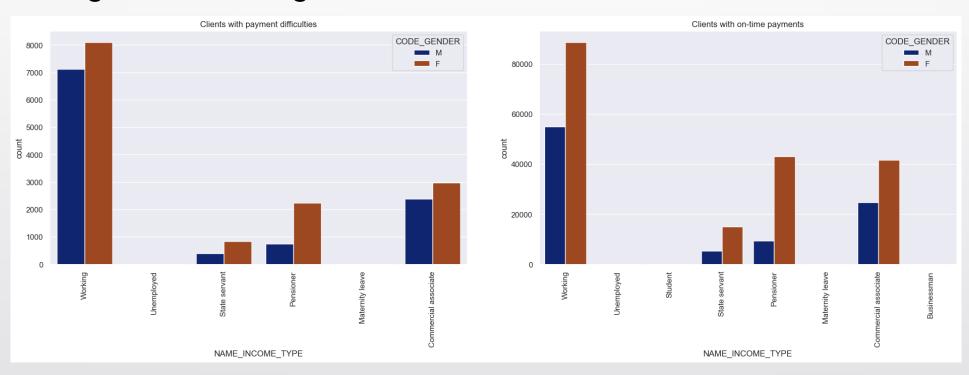
Continuous vs Categorical Variable



Continuous vs Categorical Variable

- Businessman do on-time payments either Male or Female. They don't have payment difficulties. Company must consider this category to give loan.
- Student do on-time payments either Male or Female. They don't have payment difficulties. Company must consider this category to give loan.
- Unemployed clients in Male category have very high income in on-time payments.
- Female client who are on Maternity leave have high income in on-time payments.

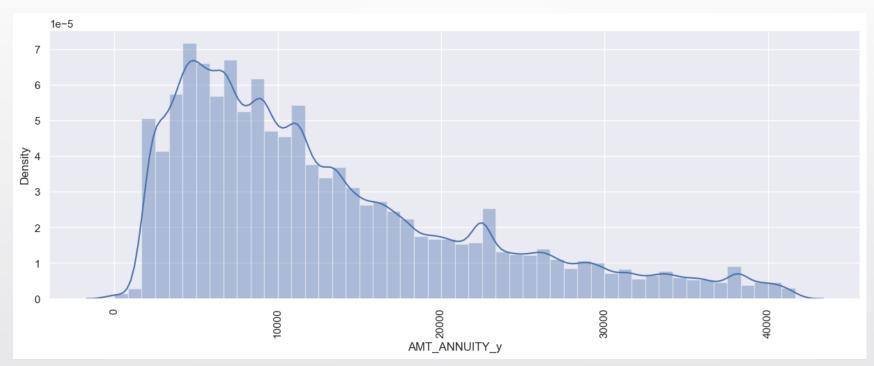
Categorical vs Categorical Variables



- Client who are Pensioner and Female, they have more on-time payments compared to payment difficulties.
- Male working clients do on-time payments but have more payment difficulties.

Analysis after Merging Previous Dataset

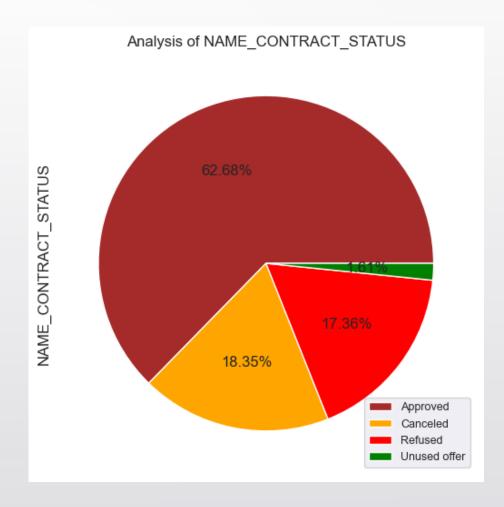
Univariate Analysis on Numerical Variables



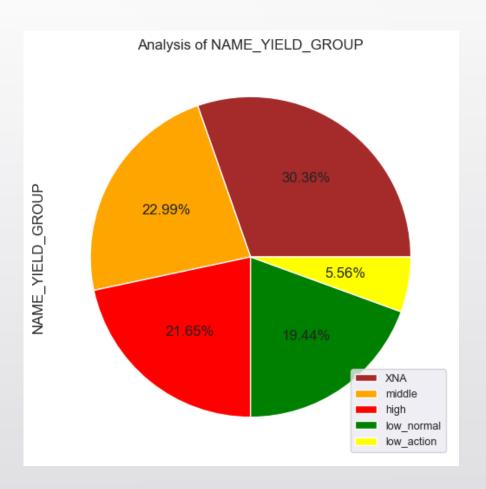
- In previous, loan's annuity increases as number of client decreases.
- Previous loan annuity is less than 8000 to 9000 as we can see the high peak in the beginning.

Univariate Analysis on Categorical variables

- Approval of loan is highest 63%.
- Cancellation of loan is 18% which is second highest.



- XNA interest rate is the highest from others.
- Middle and High interest rate are second and third respectively.



Correlation Analysis of Numerical variables

11111111		,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,								1.0
	AMT_ANNUITY_x	1.00	0.11	0.76	0.76	0.20	0.11	0.13	-0.01	1.0
	AMT_APPLICATION	0.11	1.00	0.12	0.12	0.81	0.98	1.00	0.68	- 0.8
	AMT_CREDIT_x	0.76	0.12	1.00	0.99	0.15	0.12	0.14	0.04	
	AMT_GOODS_PRICE_x	0.76	0.12	0.99	1.00	0.16	0.12	0.14	0.04	- 0.6
	AMT_ANNUITY_y	0.20	0.81	0.15	0.16	1.00	0.82	0.82	0.40	- 0.4
	AMT_CREDIT_y	0.11	0.98	0.12	0.12	0.82	1.00	0.99	0.68	
	AMT_GOODS_PRICE_y	0.13	1.00	0.14	0.14	0.82	0.99	1.00	0.67	- 0.2
	CNT_PAYMENT	-0.01	0.68	0.04	0.04	0.40	0.68	0.67	1.00	
		AMT_ANNUITY_x	AMT_APPLICATION	AMT_CREDIT_x	AMT_GOODS_PRICE_x	AMT_ANNUITY_Y	AMT_CREDIT_y	AMT_GOODS_PRICE_y	CNT_PAYMENT	- 0.0

Correlation on Numerical Variables

- AMT_APPLICATON has highest correlation with AMT_GOODS_PRICE, AMT_CREDIT_y, AMT_ANNUITY_y and good correlation with CNT_PAYMENT.
- AMT_ANNUITY_x has good correlation with AMT_CREDIT_x and AMT_GOODS_PRICE_x.
- AMT_CREDIT_x has solid correlation with AMT_GOODS_PRICE_x and good correlation with AMT_ANNUITY_x.
- AMT_GOODS_PRICE_x has strong correlation with AMT_CREDIT_x and good correlation with AMT_ANNUITY_x.
- AMT_ANNUITY_y has good correlation with AMT_GOOD_PRICE_y, AMT_CREDIT_y AND AMT_APPLICATIONs.
- AMT_CREDIT_y has strong correlation with AMT_GOODS_PRICE_y and AMT_APPLICATION.
- AMT_GOODS_PRICE_y has string correlation with AMT_CREDIT_y and AMT_APPLICATION. It has also good correlation with AMT_ANNUITY_y.
- CNT_PAYMENT has normal correlation with AMT_GOODS_PRICE_y, AMT_CREDIT_y and AMT_APPLICATIONS.

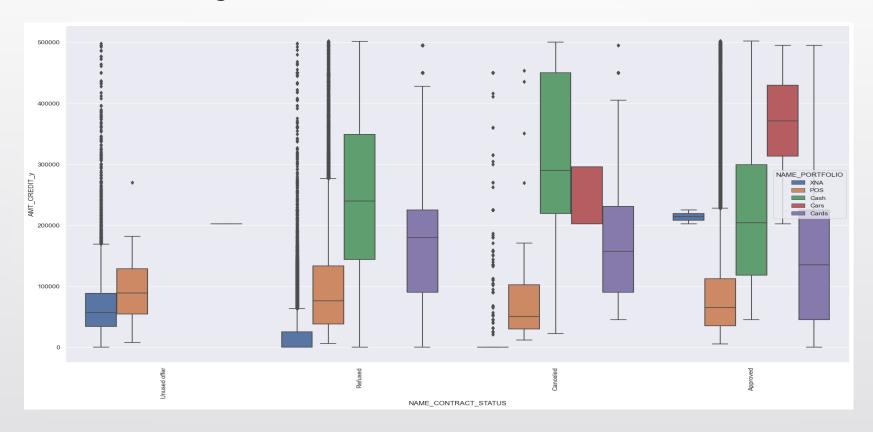
Multivariate Analysis

Continuous vs Continuous Variable



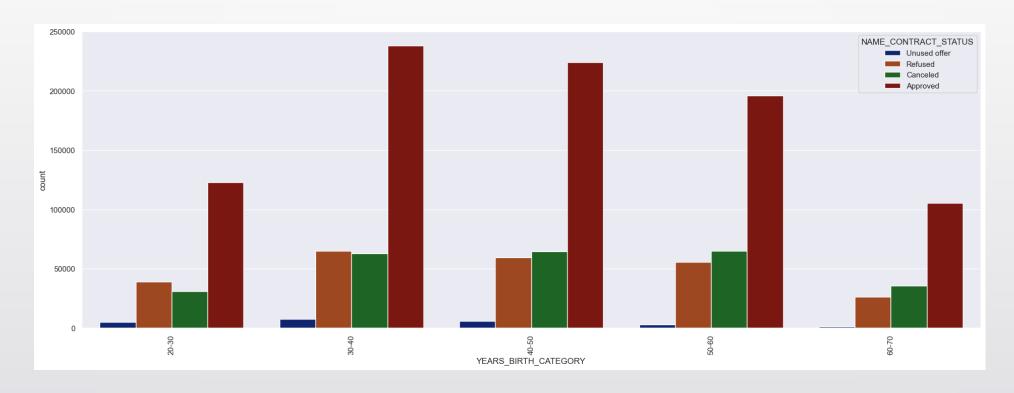
There is strong correlation between Application Amount and Credit Amount.

Continuous vs Categorical Variable



- Clients who have Refused get more median Credit Amount in Cash portfolio.
- Client who have Approved get more median credit in Cars portfolio.
- Clients who have Unused Offer get more median Credit amount as POS portfolio.

Categorical vs Categorical Variables



- Age between 30-40 and 40-50 gets most approvals respectively.
- Age between 60-70 gets least refusals.

Conclusion:

Banks should focus on below points while providing the loan.

- Clients who's age between 30-40 and 40-50.
- Clients who are Married.
- Academic degree in Male category.
- Businessman and Students are the best category to target as they don't have payment difficulties.
- Repeater clients.
- Clients Employment more than 18 years.

THANK YOU