## PRESENTATION

## ON CREDIT EDA CASE STUDY

NEHA SAHU

## WHY WE NEED TO ANALYSE

Credit risk analysis helps company to get a idea about which kind of customers should have provide loan based on their profile. Because in some cases customer proves defaulter they may not able to pay loan on time & company gets loss in business so it's necessary to do some extra work before providing loan to them.

### Which steps needs to follow after getting data

data sourcing & understanding the data

after understanding, data cleaning, checking data qualities, dealing with outliers

data imbalance, according to data imbalance needs to check correlation

univariate, bivariate analysis

afte analyse first data going to other data & merging both the data

univariate, bivariate analysis

based on the analysis recommendations

coclusion

## Approaches & methodology

#### data sourcing & understanding

First go through the basic insights about data, what kind of information we have about clients.

#### data cleaning

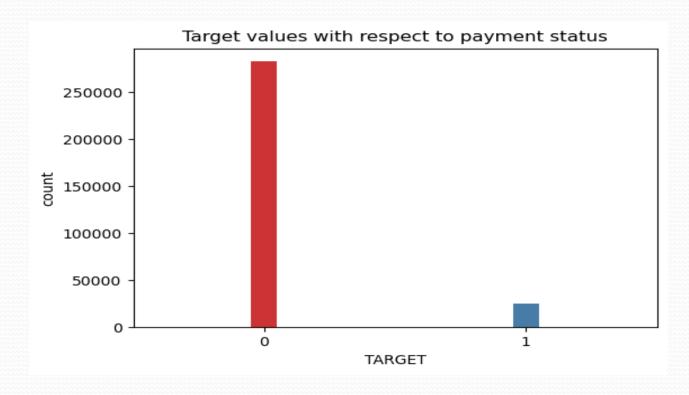
- This step is very important in analysis because after getting the necessary informations only we can get more accurtate analysis.
- > First checked how much % of missing values in data
- If it's more than 40 then it's ok to drop them & if less than 40 then impute those null values with mean median or mode.
- If data is normal distributed then we can go for mean otherwise median or mode will be useful.

### Data imbalance

We would have top most column in every analysis which is called target column so based on target column just devide the datset into 2 datasets & let's check how data distributed

target = 0 , indicates customer haven't payment difficulties target=1 , , indicates customer have payment difficulties

ratio of data imbalance = 11%



## Correlation matrix for target D

	Correlation matrix for target variable 0											
AMT_INCOME_TOTAL -	1	0.34	0.42	0.35	0.17	-0.063	-0.14	-0.065	-0.023	0.14	-0.072	-0.041
AMT_CREDIT -	0.34	1	0.77	0.99	0.1	0.047	-0.07	-0.013	0.0015	0.13	0.036	-0.07
AMT_ANNUITY -	0.42	0.77	1	0.78	0.12	-0.012	-0.1	-0.039	-0.014	0.13	0.028	-0.062
AMT_GOODS_PRICE -	0.35	0.99	0.78	1	0.1	0.045	-0.069	-0.016	0.0036	0.14	0.039	-0.071
REGION_POPULATION_RELATIVE -	0.17	0.1	0.12	0.1	1	0.025	-0.0072	0.052	0.0011	0.2	-0.012	-0.041
DAYS_BIRTH -	-0.063	0.047	-0.012	0.045	0.025	1	0.63	0.33	0.27	0.078	0.2	-0.077
DAYS_EMPLOYED -	-0.14	-0.07	-0.1	-0.069	-0.0072	0.63	1	0.21	0.28	-0.029	0.11	0.023
DAYS_REGISTRATION -	-0.065	-0.013	-0.039	-0.016	0.052	0.33	0.21	1	0.1	0.053	0.1	-0.054
DAYS_ID_PUBLISH -	-0.023	0.0015	-0.014	0.0036	0.0011	0.27	0.28	0.1	1	0.042	0.12	-0.083
EXT_SOURCE_2 -	0.14	0.13	0.13	0.14	0.2	0.078	-0.029	0.053	0.042	1	0.085	-0.19
EXT_SOURCE_3 -	-0.072	0.036	0.028	0.039	-0.012	0.2	0.11	0.1	0.12	0.085	1	-0.067
DAYS_LAST_PHONE_CHANGE -	-0.041	-0.07	-0.062	-0.071	-0.041	-0.077	0.023	-0.054	-0.083	-0.19	-0.067	1
	AMT_INCOME_TOTAL -	AMT_CREDIT -	AMT_ANNUITY -	AMT_GOODS_PRICE -	REGION_POPULATION_RELATIVE -	DAYS_BIRTH -	DAYS_EMPLOYED -	DAYS_REGISTRATION -	DAYS_ID_PUBLISH -	EXT_SOURCE_2 -	EXT_SOURCE_3 -	DAYS_LAST_PHONE_CHANGE -

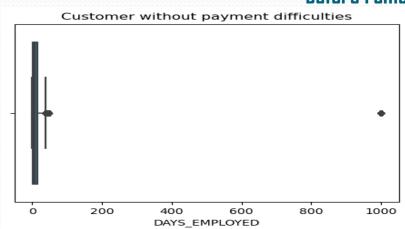
## Correlation matrix for target

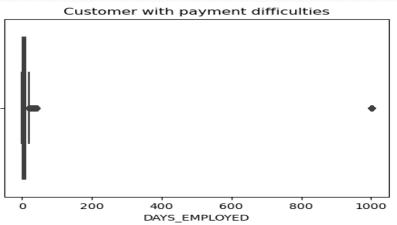
	correlation matrix for target variable 1											
AMT_INCOME_TOTAL -	1	0.038	0.046	0.038	0.0091	-0.0031	-0.015	-0.00016	0.0042	0.0072	-0.015	-0.0024
AMT_CREDIT -	0.038	1	0.75	0.98	0.069	0.14	0.0019	0.026	0.052	0.12	0.078	-0.11
AMT_ANNUITY -	0.046	0.75	1	0.75	0.072	0.014	-0.081	-0.034	0.017	0.12	0.041	-0.08
AMT_GOODS_PRICE -	0.038	0.98	0.75	1	0.076	0.14	0.0066	0.026	0.056	0.13	0.079	-0.12
REGION_POPULATION_RELATIVE -	0.0091	0.069	0.072	0.076	1	0.048	0.016	0.056	0.016	0.17	-0.0098	-0.055
DAYS_BIRTH -	-0.0031	0.14	0.014	0.14	0.048	1	0.58	0.29	0.25	0.11	0.17	-0.11
DAYS_EMPLOYED -	-0.015	0.0019	-0.081	0.0066	0.016	0.58	1	0.19	0.23	0.0011	0.096	-0.0014
DAYS_REGISTRATION -	-0.00016	0.026	-0.034	0.026	0.056	0.29	0.19	1	0.097	0.07	0.086	-0.072
DAYS_ID_PUBLISH -	0.0042	0.052	0.017	0.056	0.016	0.25	0.23	0.097	1	0.06	0.13	-0.12
EXT_SOURCE_2 -	0.0072	0.12	0.12	0.13	0.17	0.11	0.0011	0.07	0.06	1	0.078	-0.21
EXT_SOURCE_3 -	-0.015	0.078	0.041	0.079	-0.0098	0.17	0.096	0.086	0.13	0.078	1	-0.065
DAYS_LAST_PHONE_CHANGE -	-0.0024	-0.11	-0.08	-0.12	-0.055	-0.11	-0.0014	-0.072	-0.12	-0.21	-0.065	1
	AMT_INCOME_TOTAL -	AMT_CREDIT -	AMT_ANNUITY -	AMT_GOODS_PRICE -	GION_POPULATION_RELATIVE -	DAYS_BIRTH -	DAYS_EMPLOYED -	DAYS_REGISTRATION -	DAYS_ID_PUBLISH -	EXT_SOURCE_2 -	EXT_SOURCE_3 -	DAYS_LAST_PHONE_CHANGE -

- 0.6 - 0.4 - 0.2

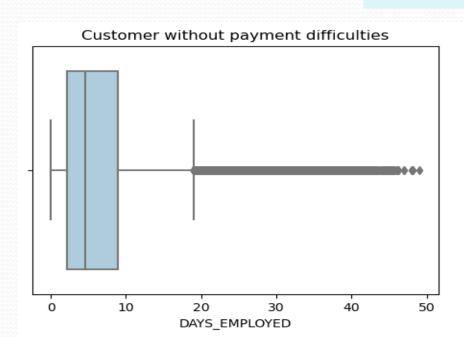
#### univariate numerical analysis ( Dealing with outliers)

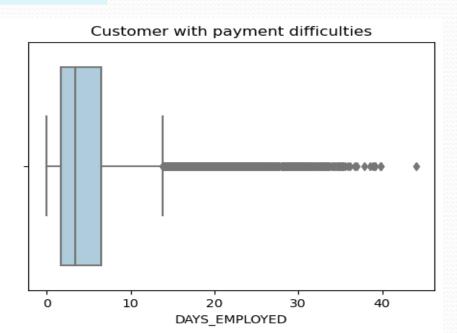
#### before removing the outlier



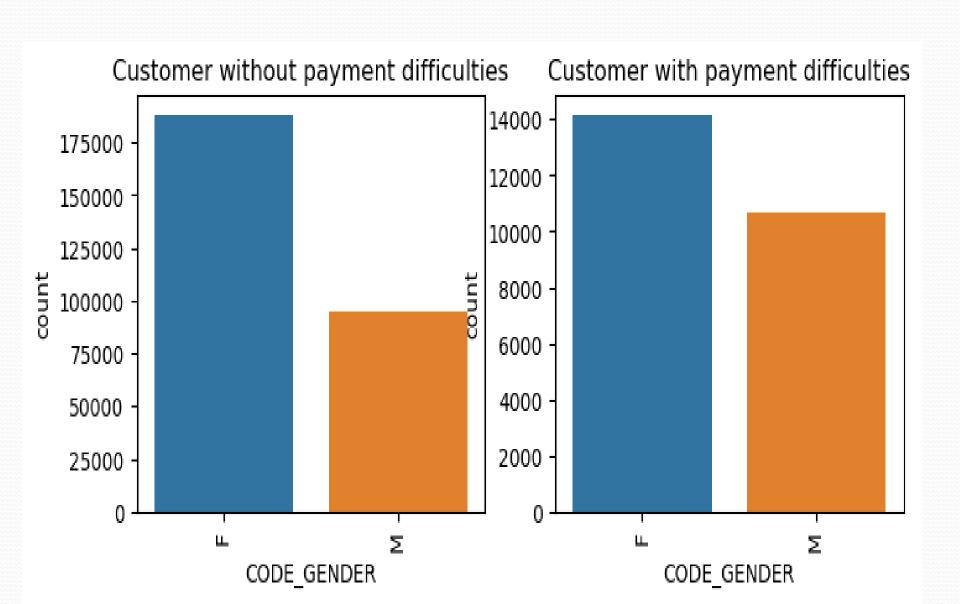


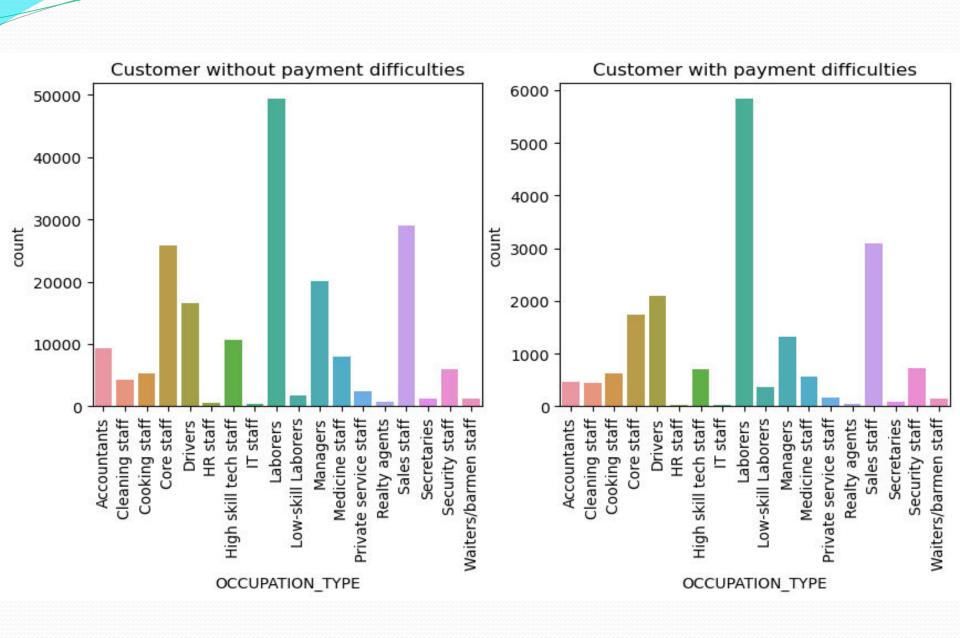
after removing the outlier





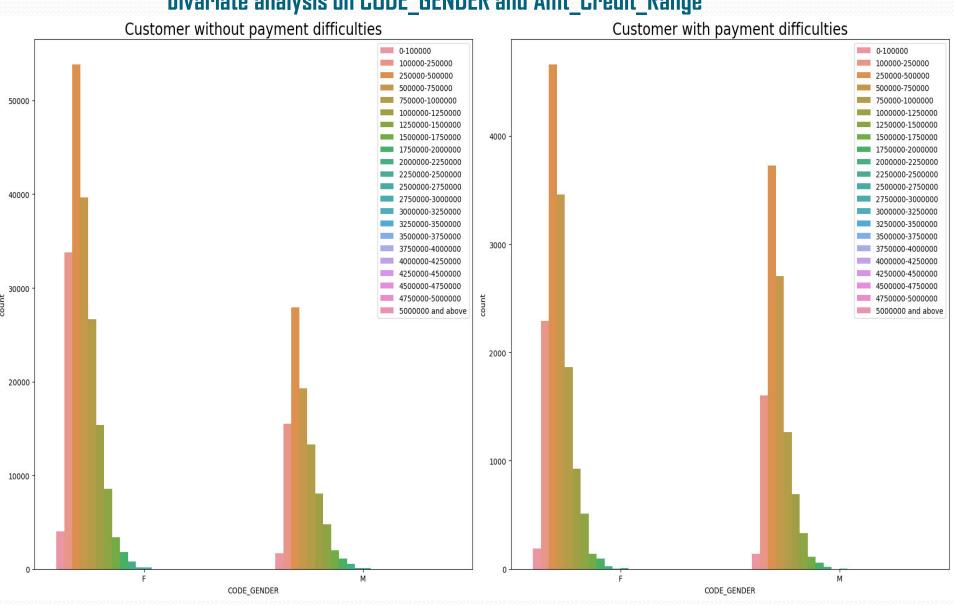
### Unvariate categorical analysis



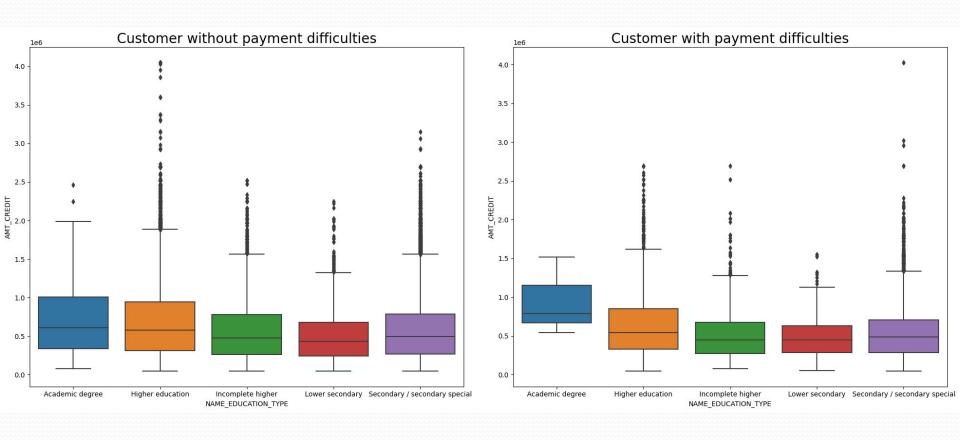


### bivariate analysis (categorical-categorical)

#### bivariate analysis on CODE\_GENDER and Amt\_Credit\_Range



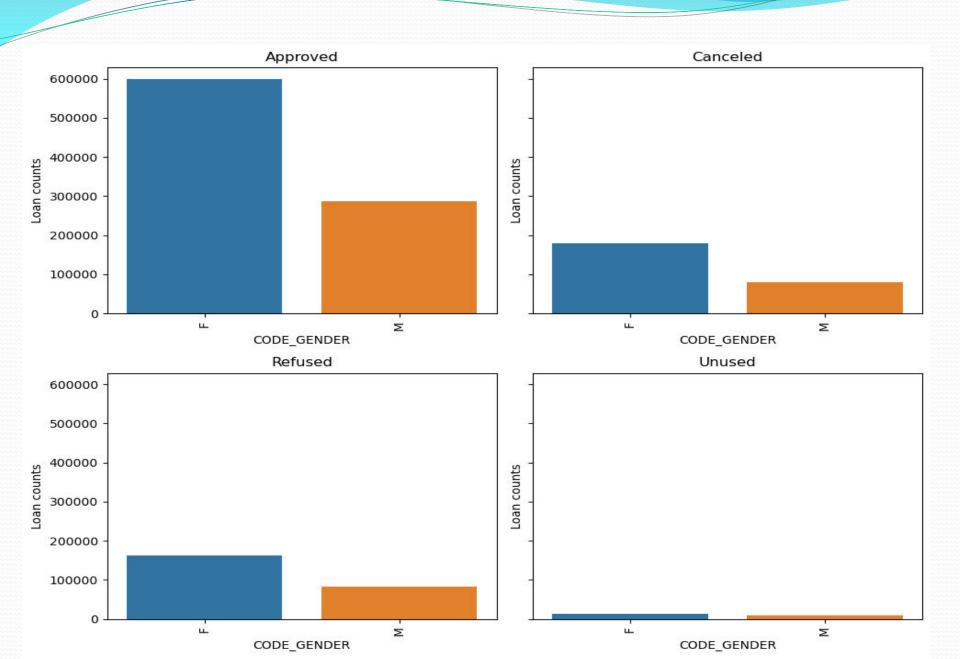
#### (categorical-numerical bivariate analysis)

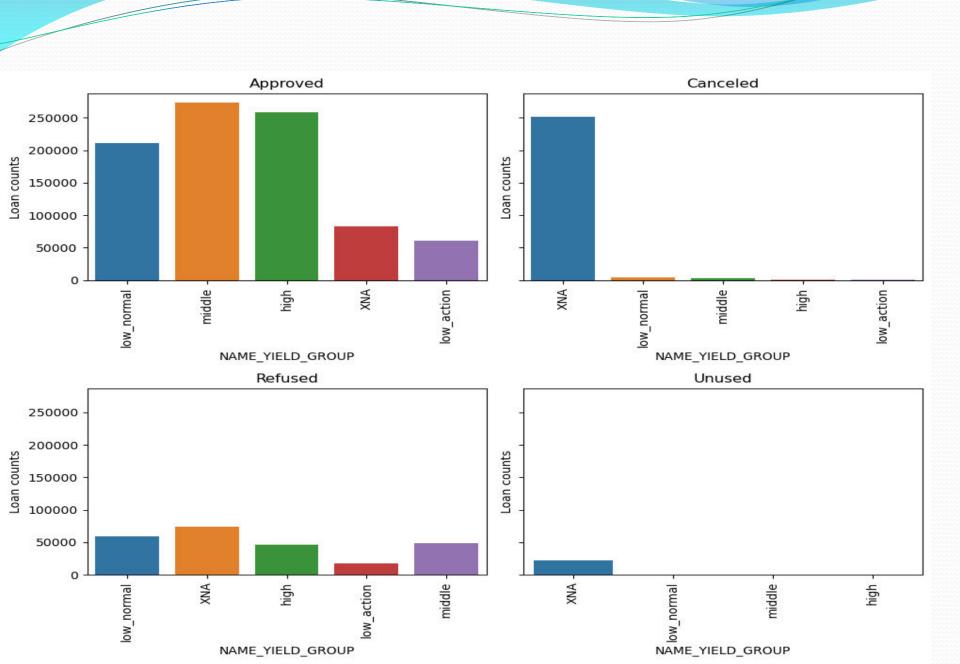


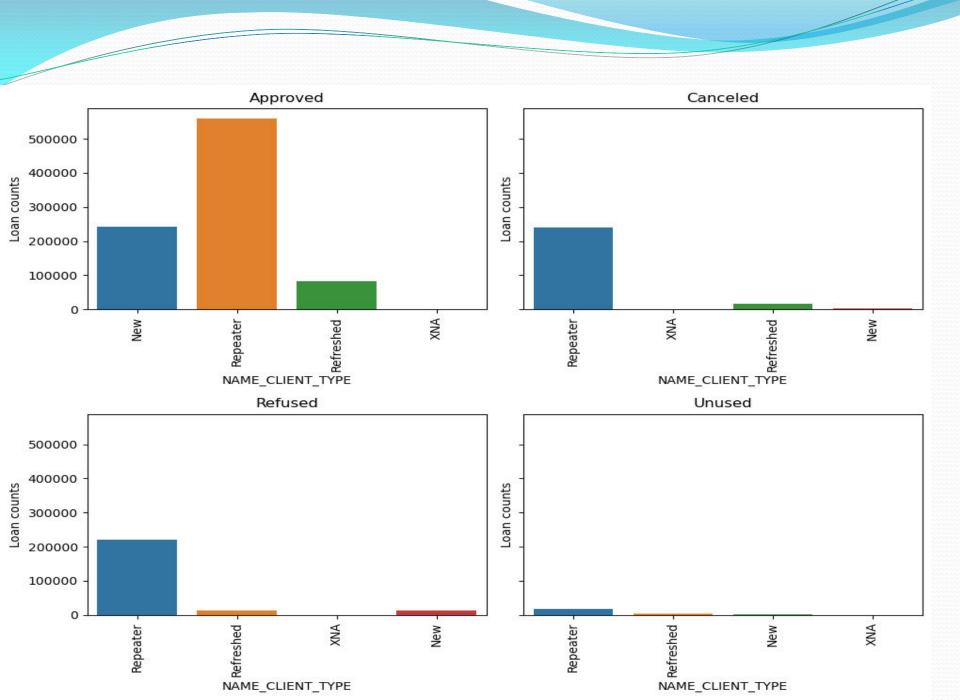
### (numerical-numerical bivariate analysis)



#### After merging both the dataset final outcomes regarding loan approval







### Recommendation

- Female customer can provide as they are less likely to be defaulter.
- Working persons (male & female).
- Married ones can provide.
- Who have their own houses/apartments can get loan.
- Age group of middle, high & normal can provide loan.
- Ones who alrrady taken loan before from the company they also can get loan.
- Refreshed & widow who have unused before approved loan.

# Thamkyou