

# CREDIT EDA

---

Case Study on Bank loan

Manikandan Natarajan

---

# Business Requirement

- Bank received loan application from various customers
- Identify the type customers have not risk for loan
- Identify the type customer for loan for loan with potential risk
- Identify patterns of existing customer procured the loan approval.

As a Data Analyst, Perform the Exploratory Data Analysis (EDA) method using the python and its libraries like numpy, pandas, matplotlib, seaborn on the data sets.

# Data set

Application dataset contains all the information of customer applied for the loan

Previous application dataset contains information of customer loan data and its status as on date.

Another dataset contains the metadata info the above of 2.

# EDA methodologies

## Data Understanding

- Understand the data given and perform initial level analysis to understand the pattern and trends

## Data Cleaning

- Perform the data cleansing on the irrelevant fields for the business problem

## Data visualization

- Apply various techniques, univariate, Bivariate and Multivariate techniques and create plots.

## Data Analysis and prediction

- On applying all above techniques and understanding the data, will be able to filter out the customers who are eligible for the loan

# EDA analysis overview

Import the data set  
with python pandas  
libraries

Identify null/missing  
data and remove/fill  
the null data.

Modify the column  
data

Create new column  
data

Univariate,Bivariate,  
multivariate analysis

Create plots on the  
column data


Merge datasets

Rule out the  
impossibles.


Find the analysis and  
recommendation for  
the business  
problem.

# Data Cleaning

Perform the analysis on missing values and replaced the with mode/mean/median data



For the Occupation type, impute 'OTHERS' for missing data.



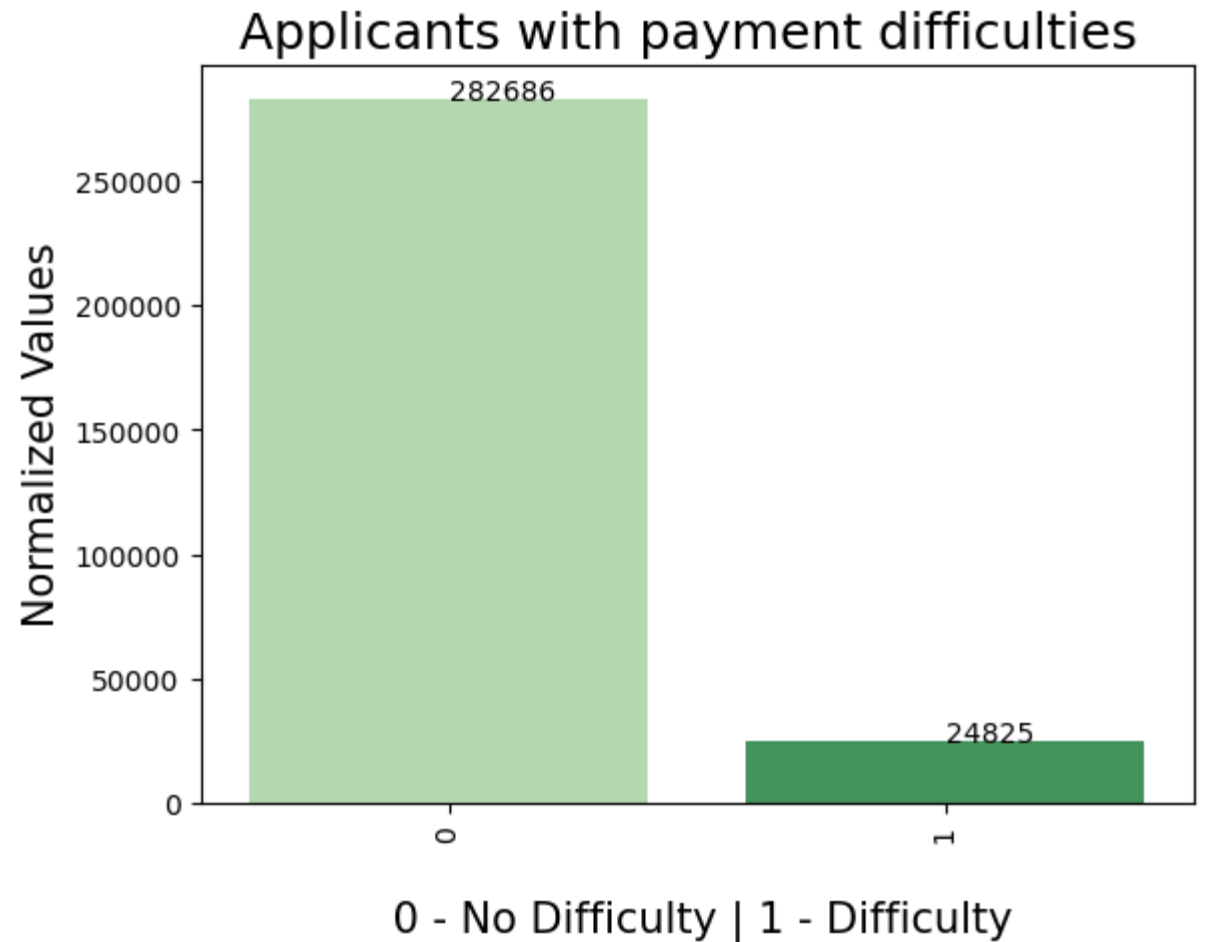
Convert the Days columns to Years based on the column description



Combine all the Flag numerical columns to one Flag column

## Data Analysis

- Plot for Target column with separation of data on payment difficulties
- Ratio between the target 1 and 0 is 11, it implies out of 11 applicant, 1 of the customer is having the payment difficulty



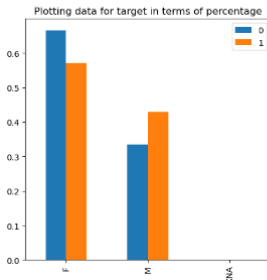
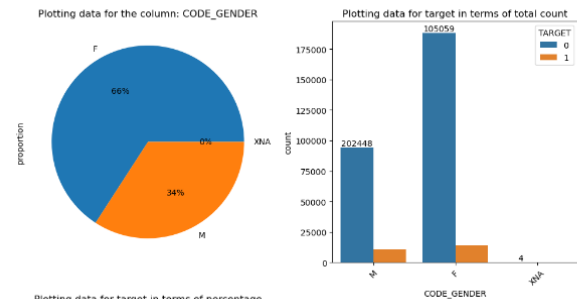
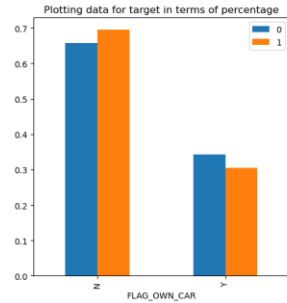
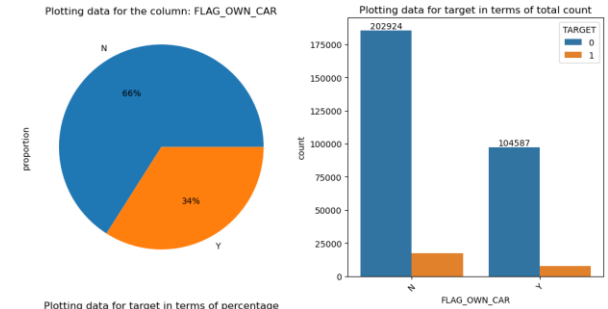
# UNIVARIATE ANALYSIS - application

## CODE\_GENDER

- Female applicant is more than the Male applicant
- Female have less payment difficulty comparing to the men

## FLAG\_OWN\_CAR / REALITY/ HOUSING TYPE

- 69 % of Applicant have own realty
- 66 % don't own car.
- 89 % live in the house/apartment type





# UNIVARIATE ANALYSIS - application

## EDUCATION

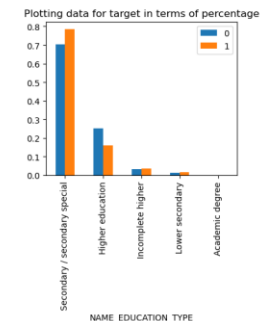
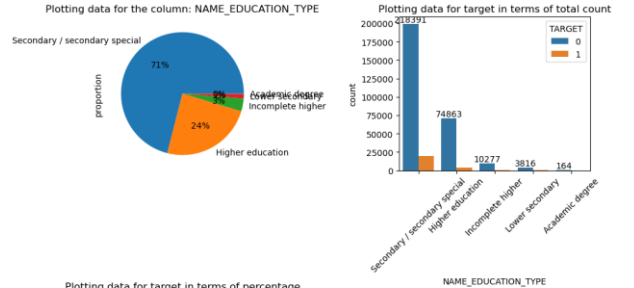
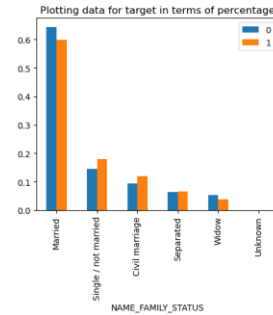
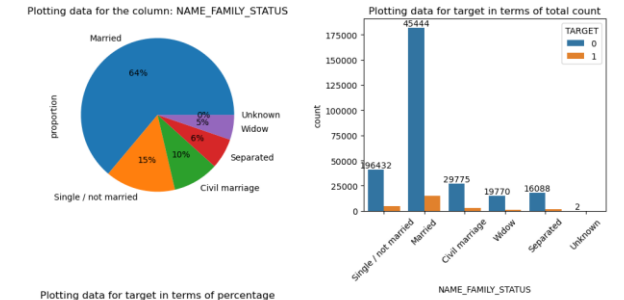
- 71 percentage have complete the secondary and nexy best is higher education of 24 percentage
- - Payment difficulty is higher among the secondary education

## INCOME TYPE

- Working people have higher application and they have higher payment difficulty

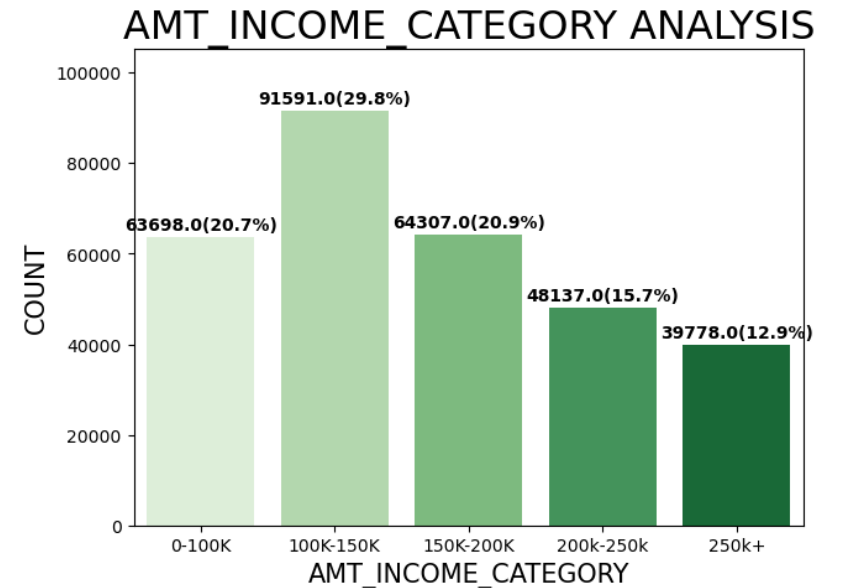
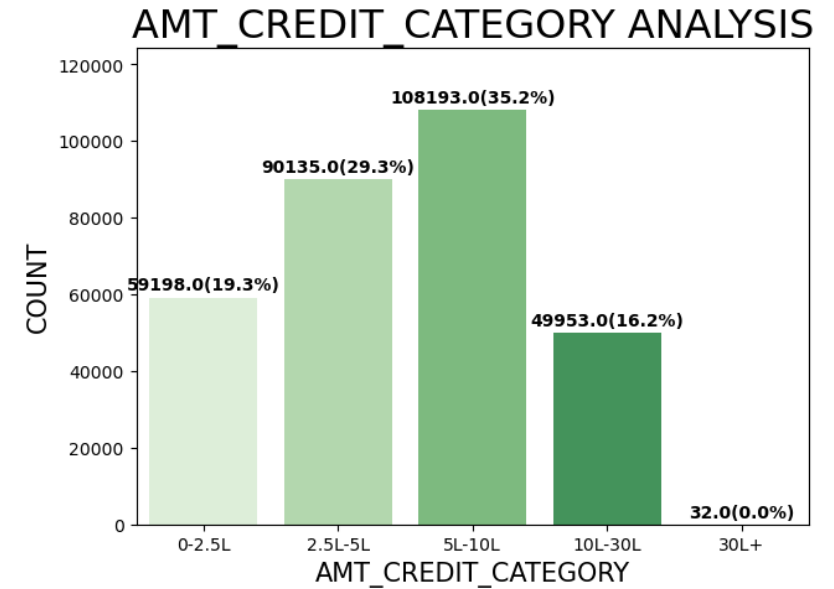
## Family Status

- - Married people application is higher
- - they have approximately same payment difficulty and no payment difficulty percentage.



# AMT category insights

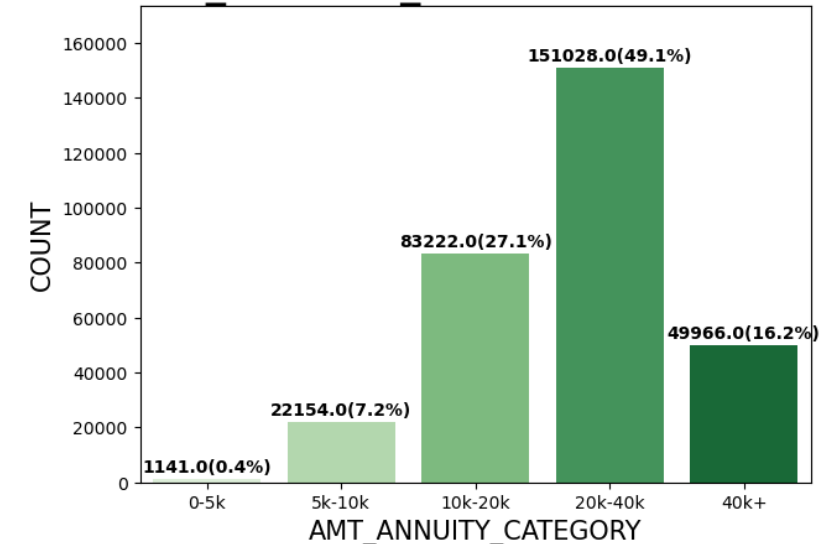
- applicant low salary range of 100k - 150k are in higher in number
- applicant high salary range of 200k+ are in lower in number
- most number of AMT\_CREDIT(LOAN AMOUNT) falls between 5lakh to 10 lakh followed by 30Lakh+ range



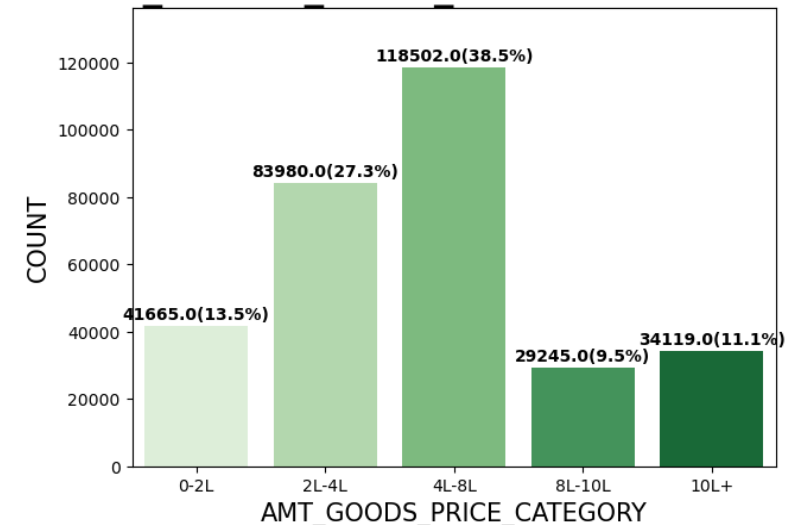
# AMT category insights

- Applicants having low range salary are in need of more Credit which is risk for bank
- Annuity falls under the range of 20k to 40k
- Interesting the amount goods price higher are in category 4L - 8L

AMT\_ANNUITY\_CATEGORY ANALYSIS



AMT\_GOODS\_PRICE\_CATEGORY ANALYSIS



# BiVariate analysis



Cash/Revolving loan is high among the working category



Secondary education is higher for the loan application is risk considering the income type



Married people are higher in cash loan number



46 % of 30-50 age category people applied for loan in most

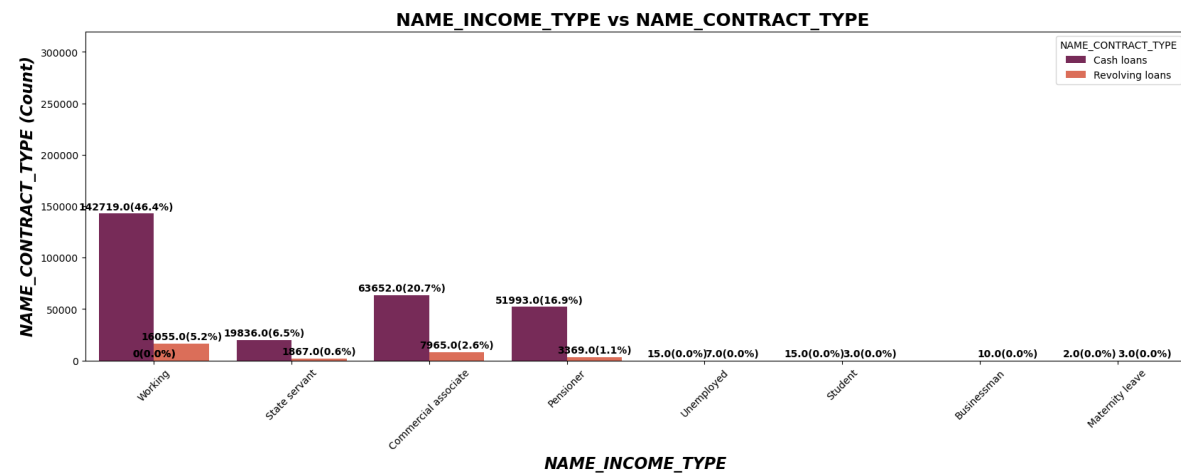
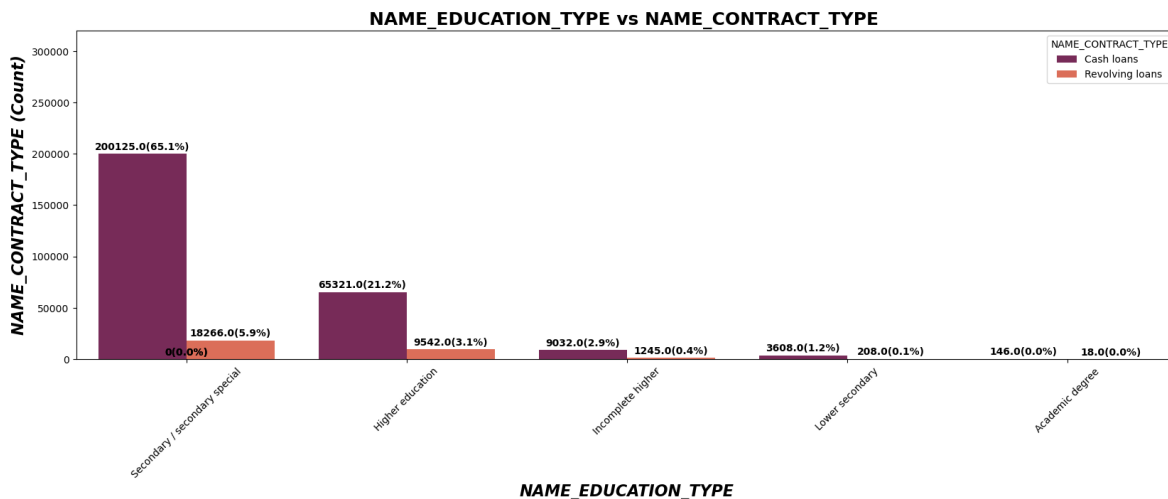


Laborers applied for more loan which is risk considering their income range

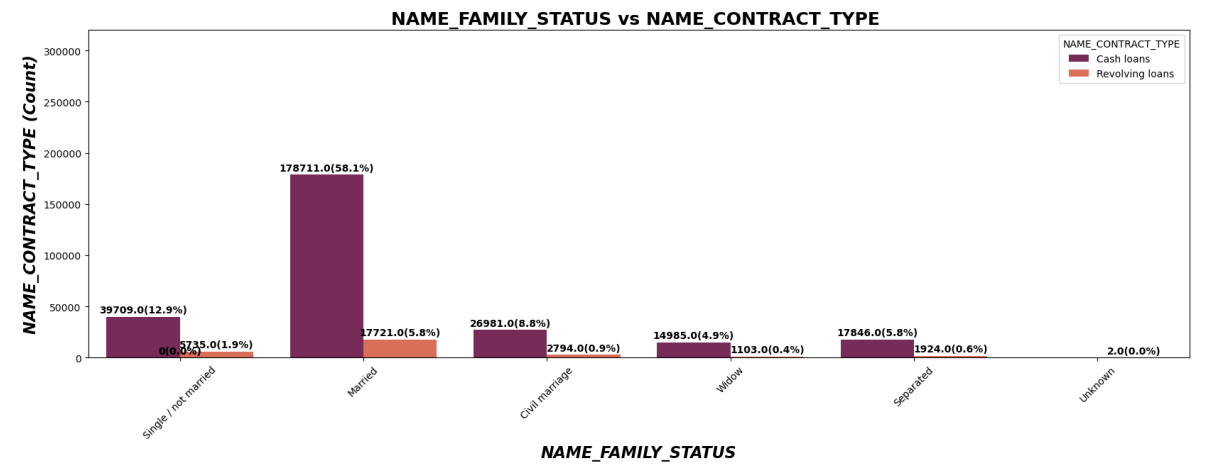
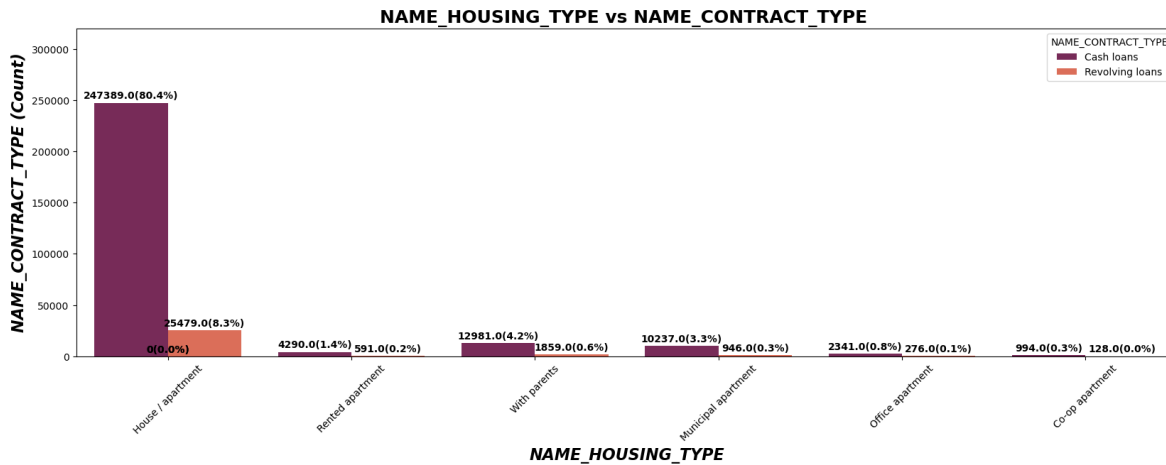


House apartment people are higher in loan cash loan category, less risk for loan repayment

# BiVariate analysis – Plots

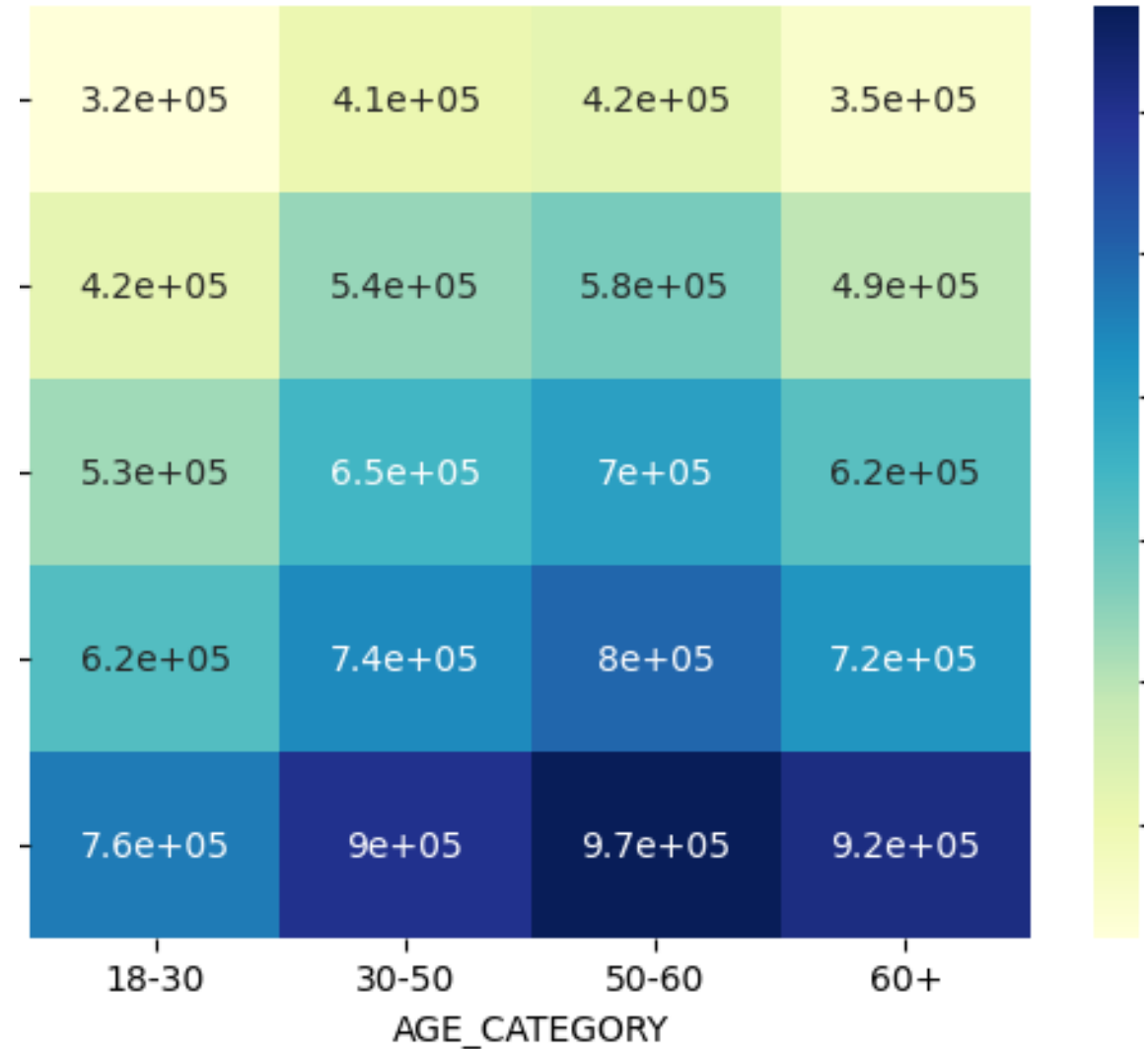


# BiVariate analysis – Plots

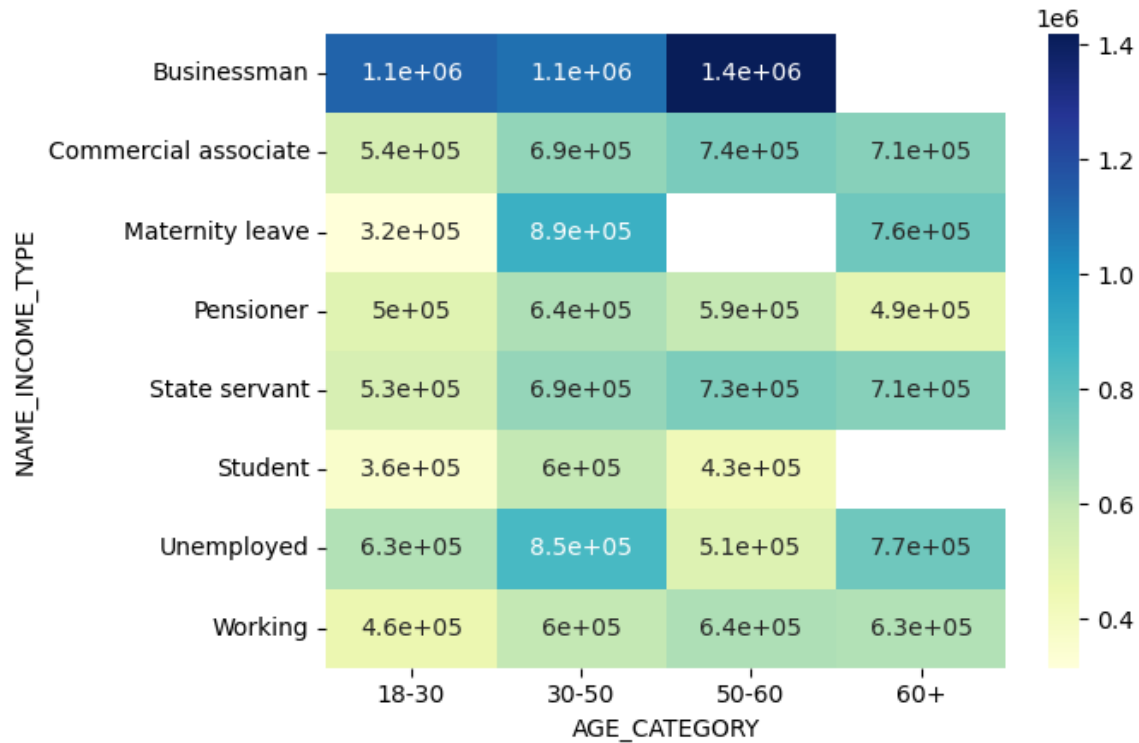


## HEATMAP - AMT\_INCOME\_CATEGORY

- HIGHER SALARY RANGE have higher AMT\_CREDIT it implies, more the salary , the are eligible for more AMT\_CREDIT (LOAN AMT)
- LOWER SALARY RANGE have LOWER AMT\_CREDIT , it implies, the are eligible for only less AMT\_CREDIT (LOAN AMT)
- Similarly the MIDDLE AGE AND UPPER AGE (30-50, 50,60) with high salary range ask for more AMT\_CREDIT
- YOUNG age with less salary have only less AMT\_CREDIT



# HEATMAP - NAME\_INCOME\_TYPE



BUSINESSMAN of all ages have HIGHER AMT\_CREDIT and likely to have longer relationship with bank and risk is lesser



Unemployed all ages have HIGH AMT\_CREDIT, there is risk of loan repayment considering their more stable income

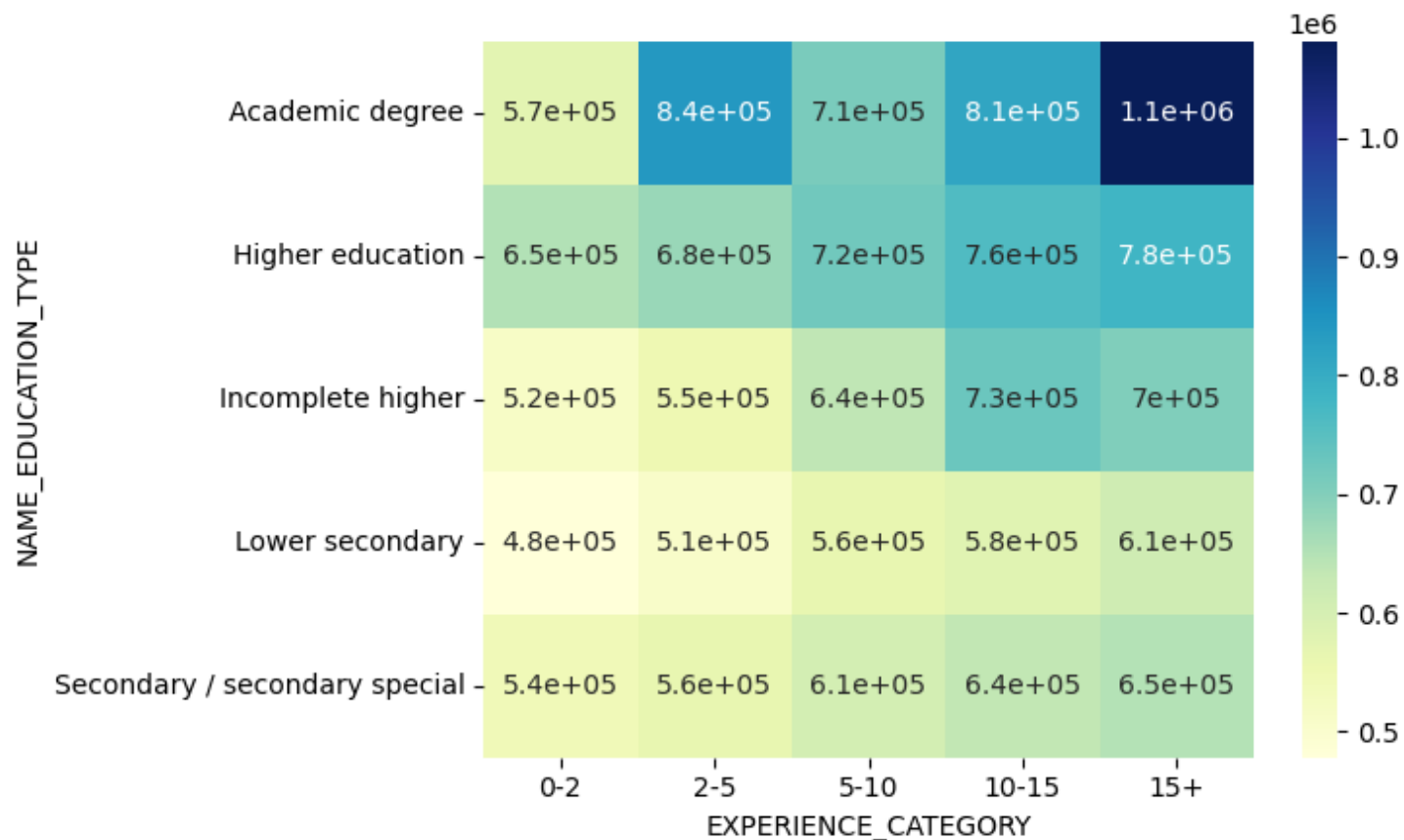


Student of all ages have LOWER AMT\_CREDIT and repayment have less risk considering their income range increase



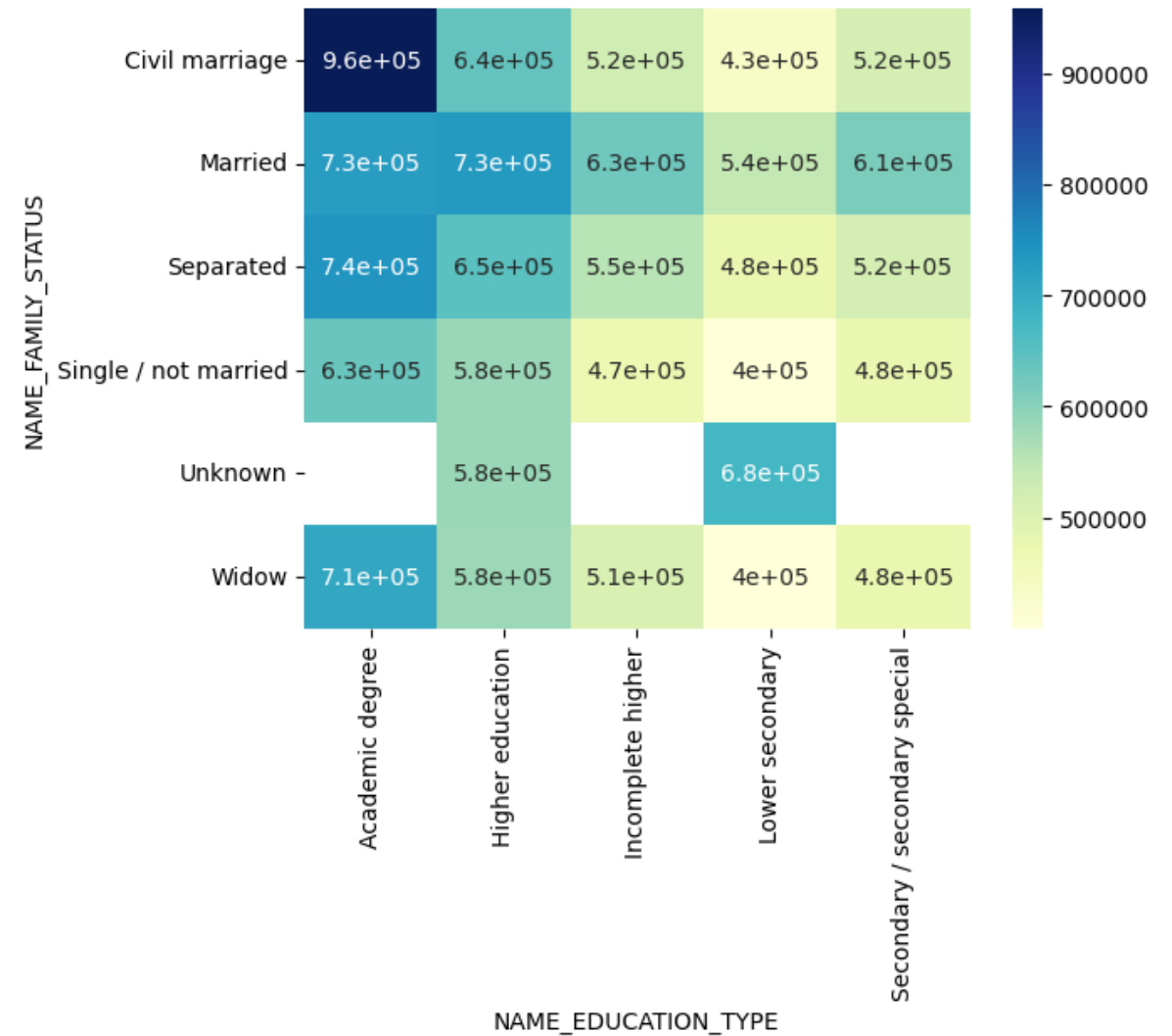
## HEATMAP - NAME\_EDUCATION\_TYPE

- ACADEMIC DEGREE of all experience have HIGHER AMT\_CREDIT and have stable income and risk is less for repayment
- HIGHER EDUCATION of all ages have HIGH AMT\_CREDIT
- Lower secondary all ages have little high AMT\_CREDIT, there is risk of loan repayment



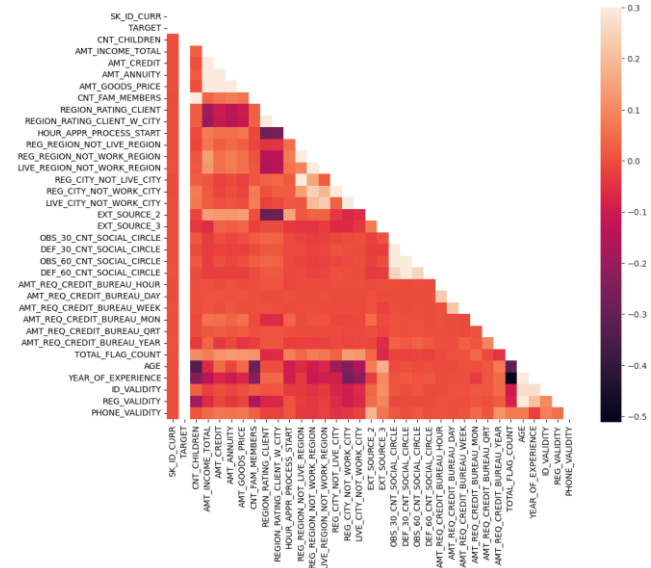
## Heatmap - NAME\_FAMILY\_STATUS

- ACADEMIC degree and civil marriage have HIGHER AMT\_CREDIT
- SEPERATED category have high AMT\_CREDIT, there is risk involved
- Lower secondary and family status UNKNOWN have little high AMT\_CREDIT, there is risk of loan repayment considering their unknown family status



# PAYMENT DIFFICULT

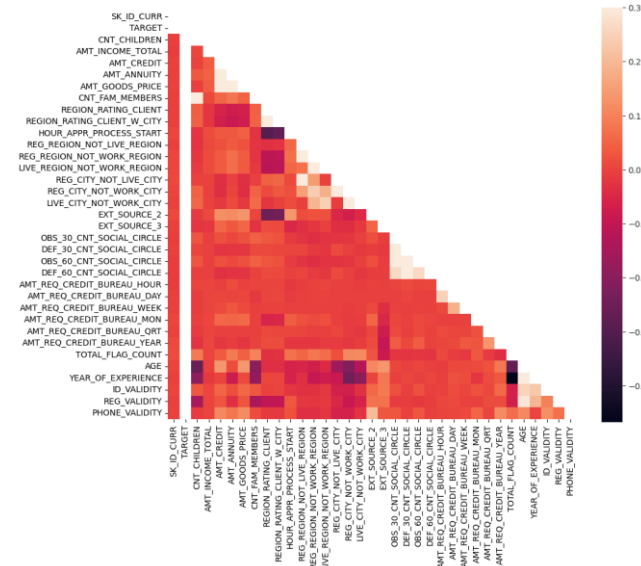
- AMT\_GOODS\_PRICE increases the AMT\_CREDIT value, the more the value of good, the loan amount is expect to increase
- AMT\_GOODS\_PRICE increases the AMT\_ANNUITY alue, the more the value of good, the monthly loan EMI is expect to increase
- AMT\_CREDIT increases the AMT\_ANNUITY value, the more the loan amount, the monthly loan EMI is expect to increase
- CNT OF FAMILY increase when the number of childer in family is more.
- AGE and YEARS of EXPERIENCE implies, more the age , the number of work experience increase



OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998510
OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.998510
AMT_GOODS_PRICE	AMT_CREDIT	0.987022
AMT_CREDIT	AMT_GOODS_PRICE	0.987022
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.950149
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.950149
CNT_CHILDREN	CNT_FAM_MEMBERS	0.878571
CNT_FAM_MEMBERS	CNT_CHILDREN	0.878571
LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.861861
REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	0.861861
DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.859371
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.859371
LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.830381
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	0.830381
AMT_ANNUITY	AMT_GOODS_PRICE	0.776400
AMT_GOODS_PRICE	AMT_ANNUITY	0.776400
AMT_CREDIT	AMT_ANNUITY	0.771276

# Non-Payment difficulty

- AMT\_GOODS\_PRICE increases the AMT\_CREDIT value, the more the value of good, the loan amount is expect to increase
- AMT\_GOODS\_PRICE increases the AMT\_ANNUITY alue, the more the value of good, the monthly loan EMI is expect to increase
- AMT\_CREDIT increases the AMT\_ANNUITY value, the more the loan amount, the monthly loan EMI is expect to increase
- CNT OF FAMILY increase when the number of childer in family is more.
- AGE and YEARS of EXPERIENCE implies, more the age , the number of work experience increase



OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.998270
OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998270
AMT_GOODS_PRICE	AMT_CREDIT	0.982783
AMT_CREDIT	AMT_GOODS_PRICE	0.982783
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.956637
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.956637
CNT_FAM_MEMBERS	CNT_CHILDREN	0.885484
CNT_CHILDREN	CNT_FAM_MEMBERS	0.885484
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.869016
DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.869016
LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.847885
REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	0.847885
LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.778540
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	0.778540
AMT_GOODS_PRICE	AMT_ANNUITY	0.752295
AMT_ANNUITY	AMT_GOODS_PRICE	0.752295
AMT_CREDIT	AMT_ANNUITY	0.752195
AMT_ANNUITY	AMT_CREDIT	0.752195

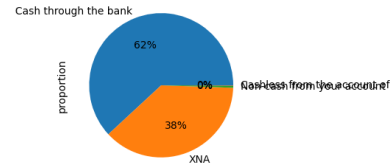
# Previous application – Data Analysis

- Consumer loan are having higher approval rate
- Cash repayment in bank is having higher approval rate
- credit and cash offices are higher in number and approval rate is higher
- Cancellation rate is high in contact center
- consumer and connectivity have higher approval rate and refusal rate

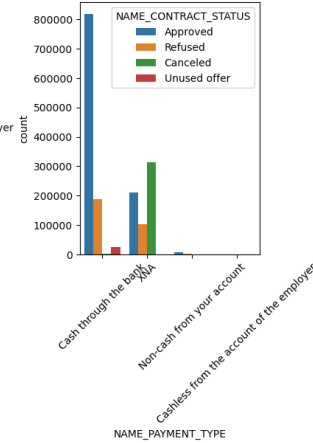
# Previous application – Data Analysis



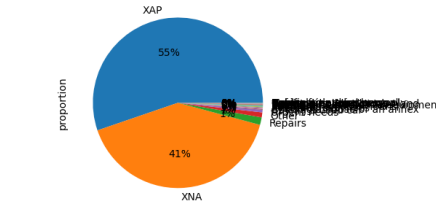
Plotting data for the column: NAME\_PAYMENT\_TYPE



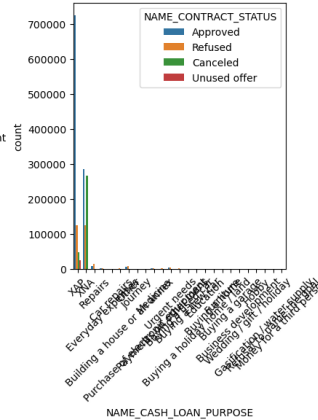
Plotting data for target in terms of total count



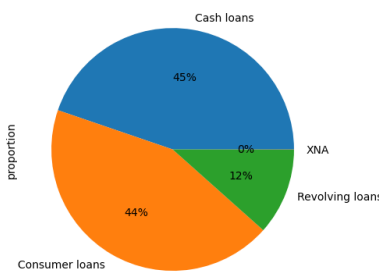
Plotting data for the column: NAME\_CASH\_LOAN\_PURPOSE



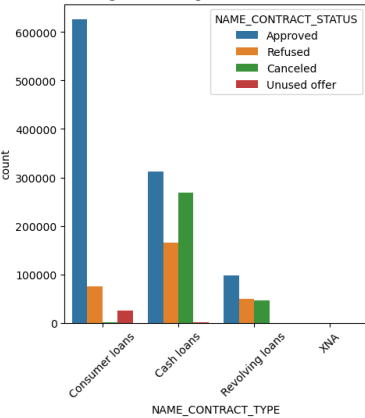
Plotting data for target in terms of total count



Plotting data for the column: NAME\_CONTRACT\_TYPE

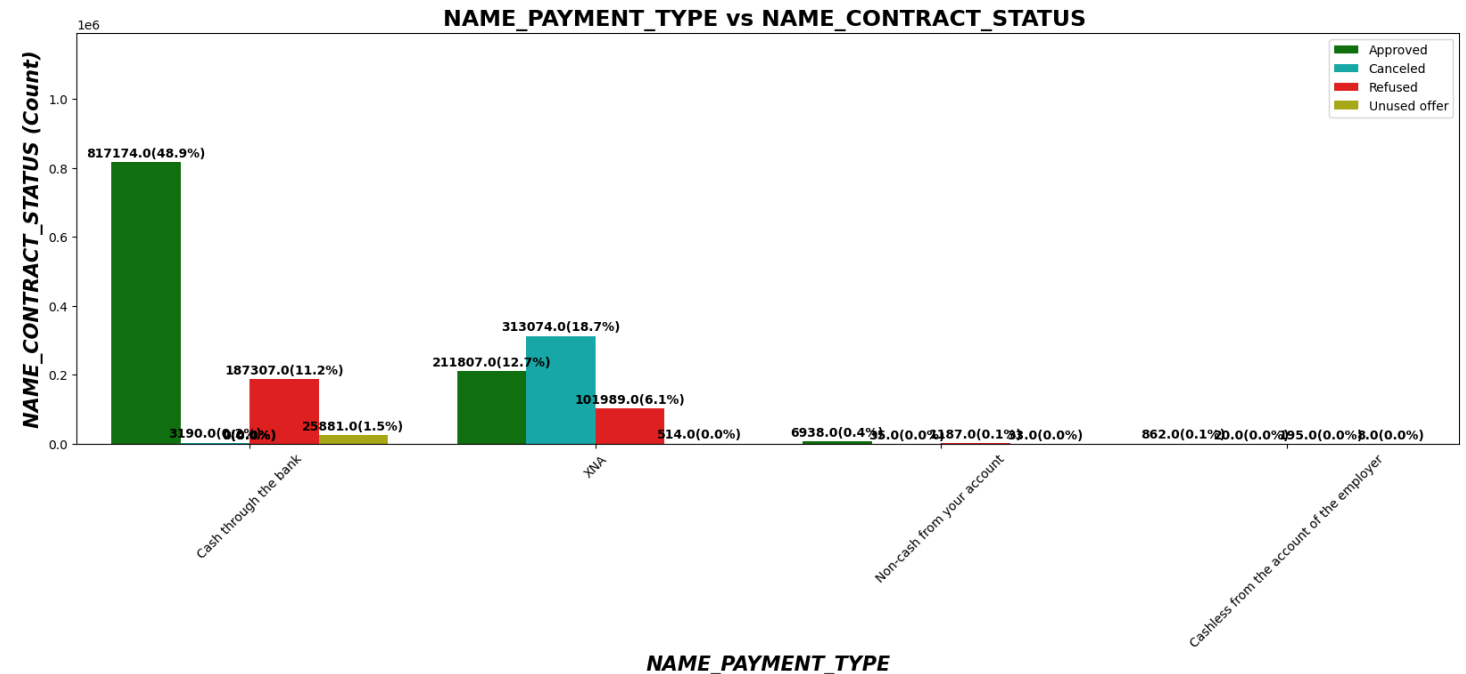


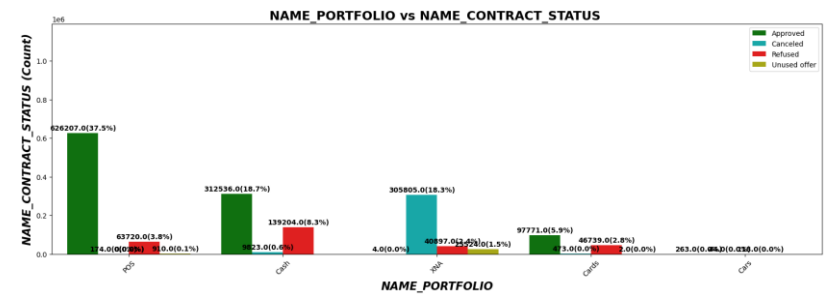
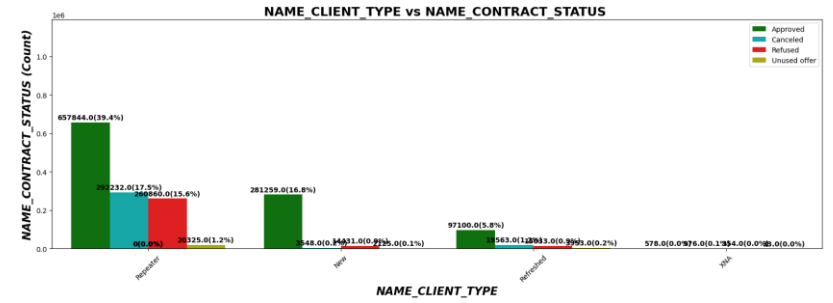
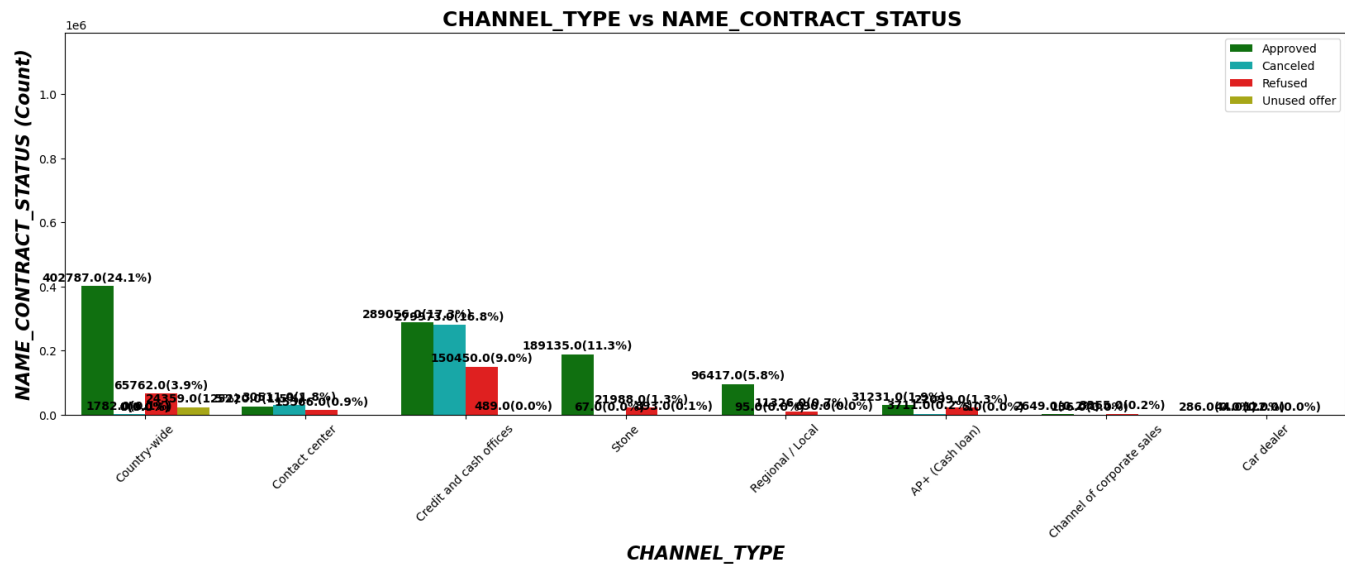
Plotting data for target in terms of total count



## BIVARIATE AND MULTIVARIATE ANALYSIS - previous

- Repeater and New customer approval rate is relatively high.
- Cash through bank customer approval rate is high
- Refusal rate is lesser in new customers
- pos category have good approval percentage
- credit and cash offices have similar approval and cancel percent

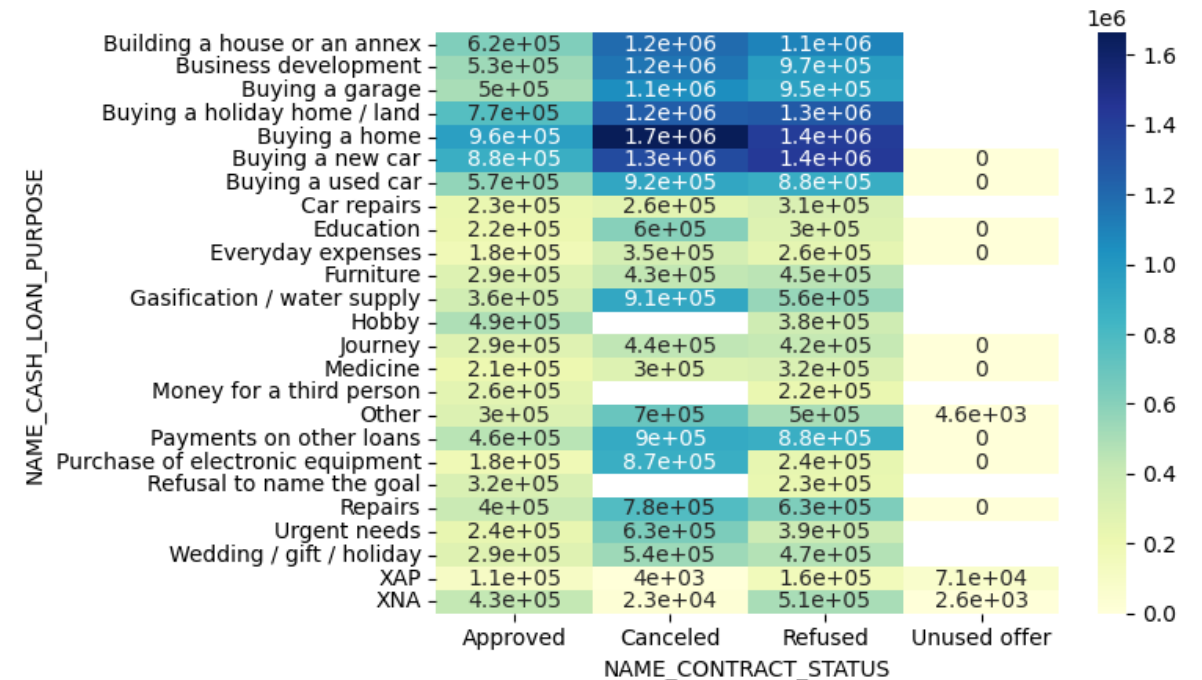






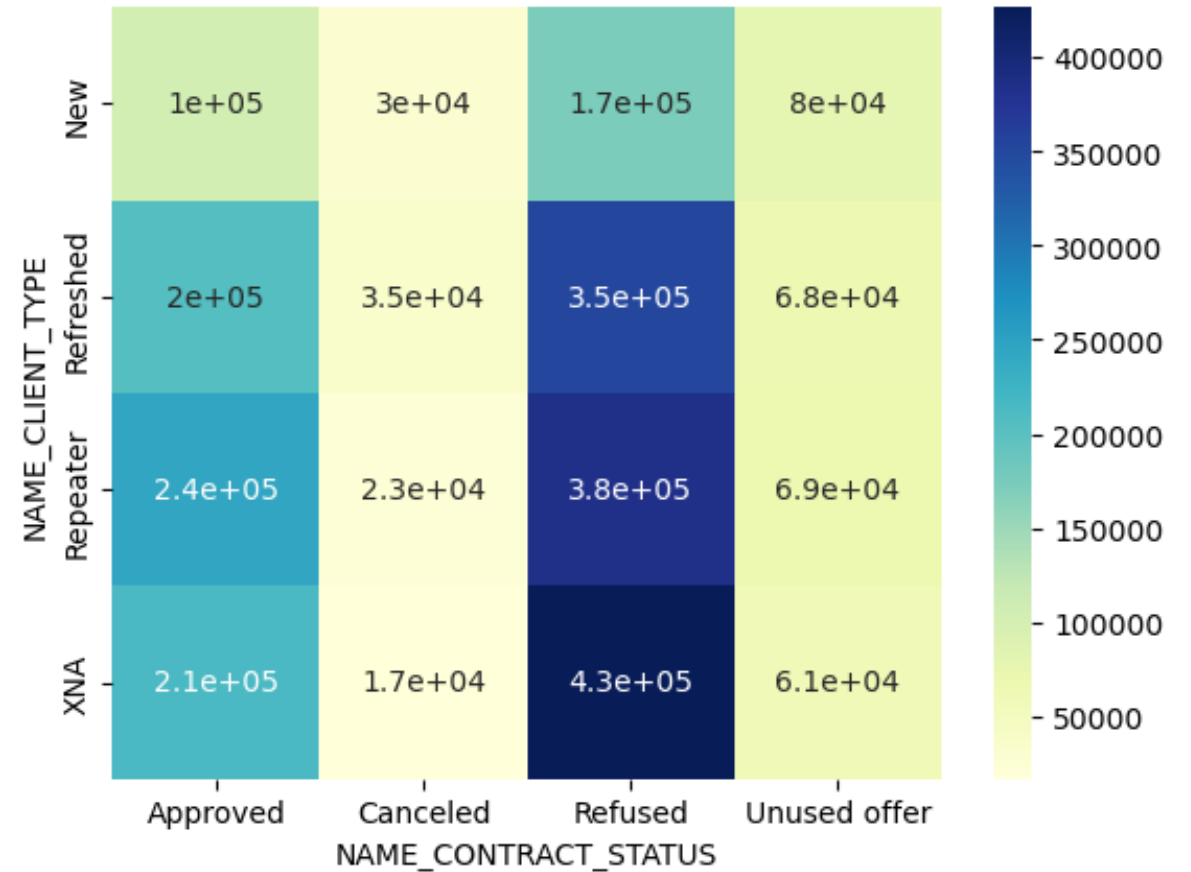
# Heat map – loan purpose

- Home loan and buy new car loan are the most approved category
- Next best is land, its approval rate is good
- Buying old car loan is refused, it implies, used car have less market value and risk is higher
- Everyday expense/Wedding/Gift loan approval rate is lower and most of them are cancelled as well
- Medicine and everyday expense load rejection rate is lower



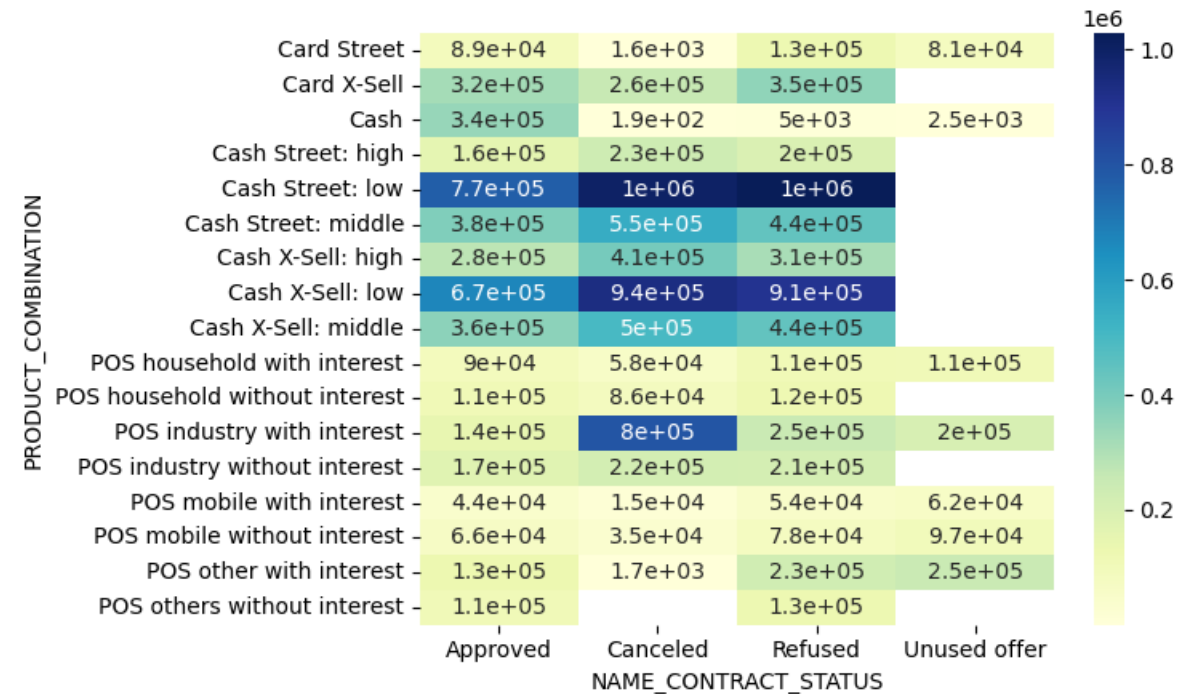
# Heat map – Client Type

- Clients who repeat the loan are likely to get the refused in higher percentage
- Clients who refresh the loan again are also refused.
- Approval rate for new clients are less and equably get rejected
- New clients unused offer also comparatively high compared to other



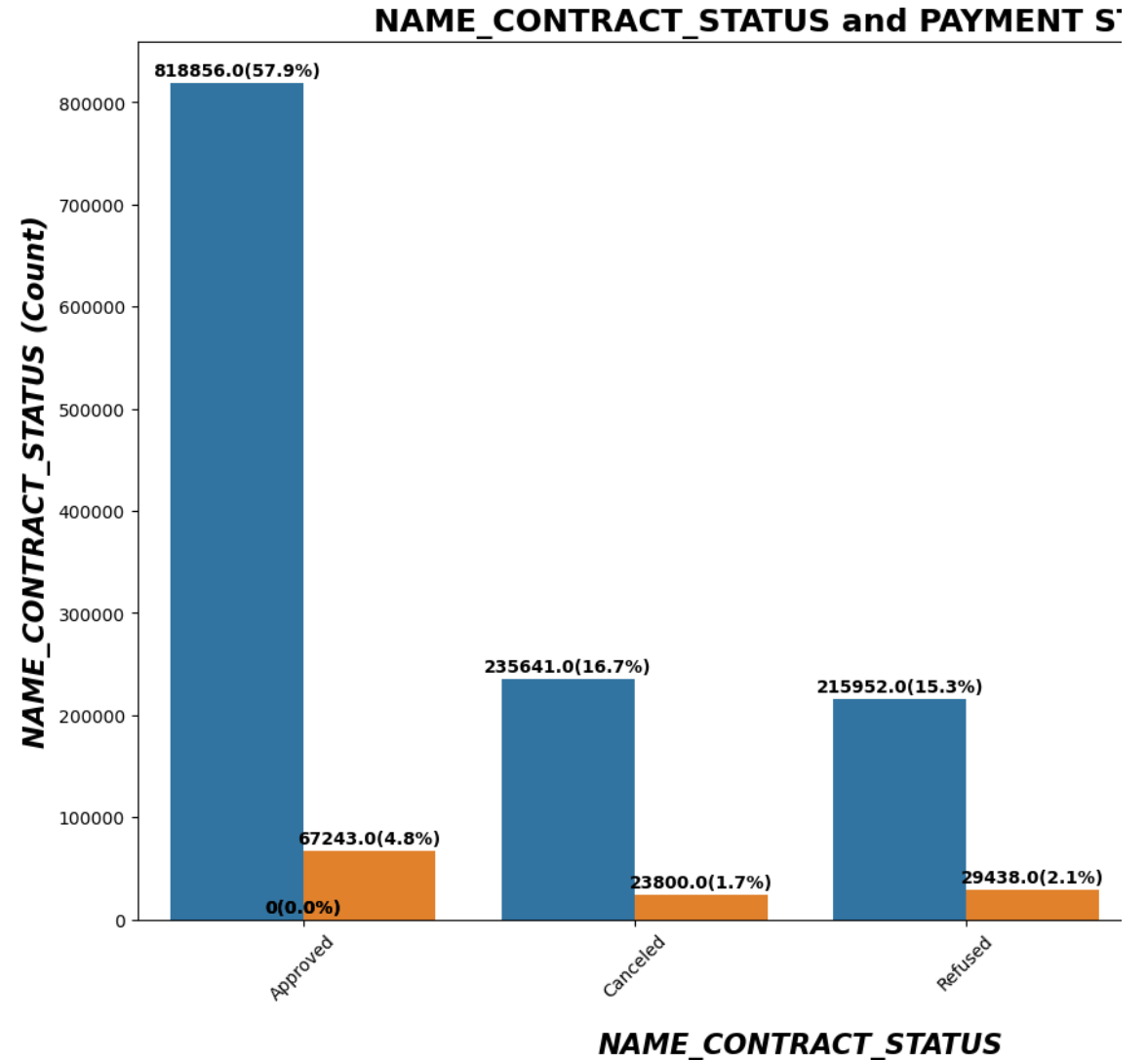
# Heat Map - PRODUCT COMBINATION

- Refusal and Cancellation rate is high for Cash Street- low. Risk is high for this category loan
- Refusal and Cancellation rate is high for Cash X-sell - low. Risk is high for this category loan
- POS industry have cancellation high, Risk is high



## Merge data set – SK\_ID\_CURR NAME\_CONTRACT\_STATUS vs TARGET

- Almost half 51% customer who got approved had no payment difficulties
- 4 % customer approved have payment difficulties and risk is anticipated for those customers to be in default list
- 15 % Customers who got refused last time have less payment difficulties
- Customers who cancelled their loan as well have less payment difficulties



# CONCLUSION and Recommendation based on the EDA analys

---

## No Risk :

- 
- Business income type applicant of above 30+ have no risk in repayment
- 
- Academic degree holder have less risk in repayment
- 
- Working people with work experience above 5 have less risk
- 
- High income and medium AMT credit applicant like to have less risk
- 

## Risk :

- 
- Risk is involved amount the low-income people with high AMT\_CREDIT
- 
- Secondary education applicant with low work experience have high risk
- 
- Used car refusal rate is high, applicant for used car is high risk of repayment
- 
- High AMT\_CREDIT for the low-income customer have risk
-