# CREDIT EDA CASE STUDY PROJECT

SUBMITTED BY:
ABHINAV CHOUDHARY & AKSHAY SAWANT

# PROJECT OBJECTIVE

*This case study aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.*

# EDA ANALYSIS APPROACH

- **Understanding the Data:** We have used the Jupyter Python Notebook to load and understand the datasets for both Current Applications & Previous Applications. We have gone through the breadth & the depth of the attributes present in the datasets along with their definition to understand the data quality & its spread at a high level.
- **Data Preparation/Cleaning:** After a clear understanding of the problem statement & the given datasets, we moved to the most critical & an essential part of this project i.e. Data Clearing & preparing the data for further analysis.
  In this process, we found quite a few irregularities in the data such as missing values; outliers; incorrect data format & invalid values/records.
  We have then performed all the necessary operations to clean the data as much as we could in order to make it ready for further analysis.
- **Data Analysis:** Finally, we performed various types of Univariate, Bivariate & Multivariate analysis by plotting appropriate graphs with respect to the Target variable. This helped us to draw relevant insights.

***Please note*** *that this presentation is only meant to show our analysis and insights. For the complete step-by-step process, kindly refer to the well commented Python notebook submitted along with this presentation.*

# DATA IMBALANCE
# &
# TOP 10
# CORRELATIONS

- **Data Imbalance:** During our analysis we have tried to understand the data imbalance for the Target variable and it came out to be **10.55** which is the ratio of – customers with **no** payment difficulty v/s. customers with payment difficulty.
- **Top 10 Correlations:** We performed the correlation analysis for the Target variable by plotting the segmented correlation matrix. Here is a list of top 10 closely related attribute pairs with their absolute correlation coefficient values.

*Correlation Matrix plots in the following slides…*

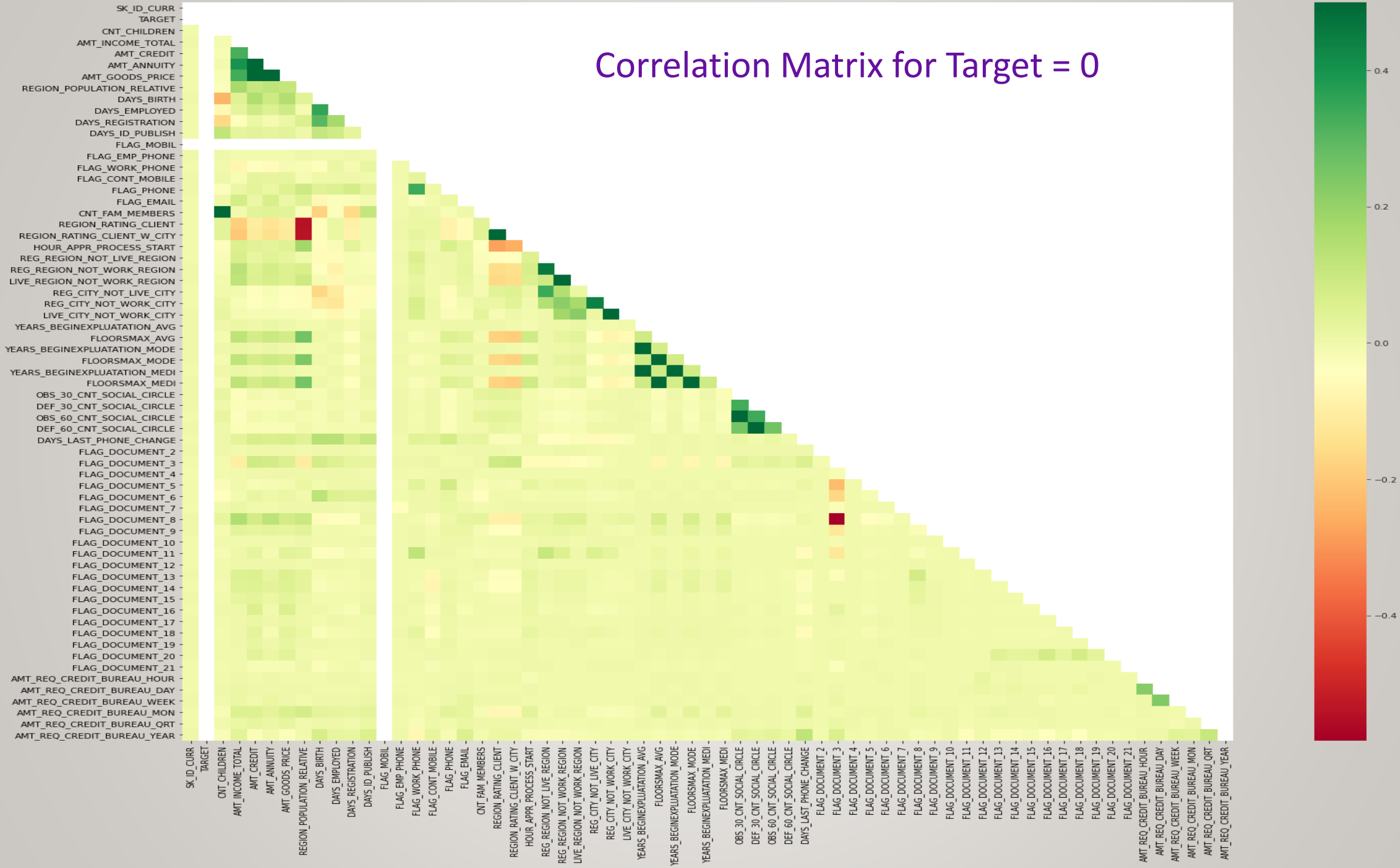
Top 10 Correlation variable pairs for Target = 0

| | | |
|---|---|---|
| AMT_CREDIT | AMT_GOODS_PRICE | 0.986726 |
| AMT_GOODS_PRICE | AMT_CREDIT | 0.986726 |
| FLOORSMAX_MEDI | FLOORSMAX_MODE | 0.987974 |
| FLOORSMAX_MODE | FLOORSMAX_MEDI | 0.987974 |
| YEARS_BEGINEXPLUATATION_AVG | YEARS_BEGINEXPLUATATION_MEDI | 0.993067 |
| YEARS_BEGINEXPLUATATION_MEDI | YEARS_BEGINEXPLUATATION_AVG | 0.993067 |
| FLOORSMAX_AVG | FLOORSMAX_MEDI | 0.997066 |
| FLOORSMAX_MEDI | FLOORSMAX_AVG | 0.997066 |
| OBS_60_CNT_SOCIAL_CIRCLE | OBS_30_CNT_SOCIAL_CIRCLE | 0.998493 |
| OBS_30_CNT_SOCIAL_CIRCLE | OBS_60_CNT_SOCIAL_CIRCLE | 0.998493 |

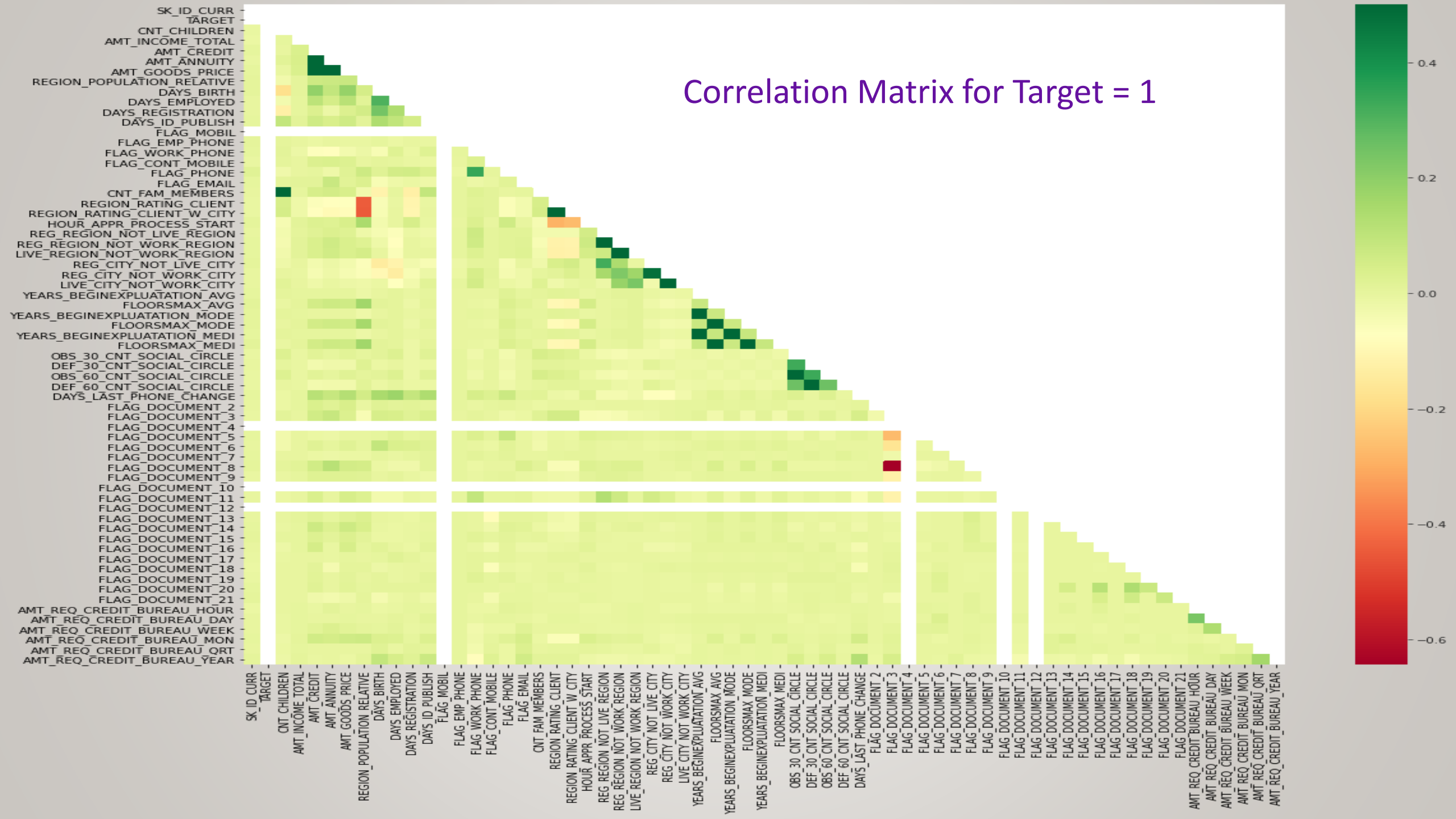
Top 10 Correlation variable pairs for Target = 1

| | | |
|---|---|---|
| FLOORSMAX_AVG | FLOORSMAX_MODE | 0.986833 |
| FLOORSMAX_MODE | FLOORSMAX_AVG | 0.986833 |
| FLOORSMAX_MEDI | FLOORSMAX_MODE | 0.988979 |
| FLOORSMAX_MODE | FLOORSMAX_MEDI | 0.988979 |
| YEARS_BEGINEXPLUATATION_AVG | YEARS_BEGINEXPLUATATION_MEDI | 0.995888 |
| YEARS_BEGINEXPLUATATION_MEDI | YEARS_BEGINEXPLUATATION_AVG | 0.995888 |
| FLOORSMAX_AVG | FLOORSMAX_MEDI | 0.997467 |
| FLOORSMAX_MEDI | FLOORSMAX_AVG | 0.997467 |
| OBS_60_CNT_SOCIAL_CIRCLE | OBS_30_CNT_SOCIAL_CIRCLE | 0.998289 |
| OBS_30_CNT_SOCIAL_CIRCLE | OBS_60_CNT_SOCIAL_CIRCLE | 0.998289 |

# Correlation Matrix for Target = 0



# Correlation Matrix for Target = 1

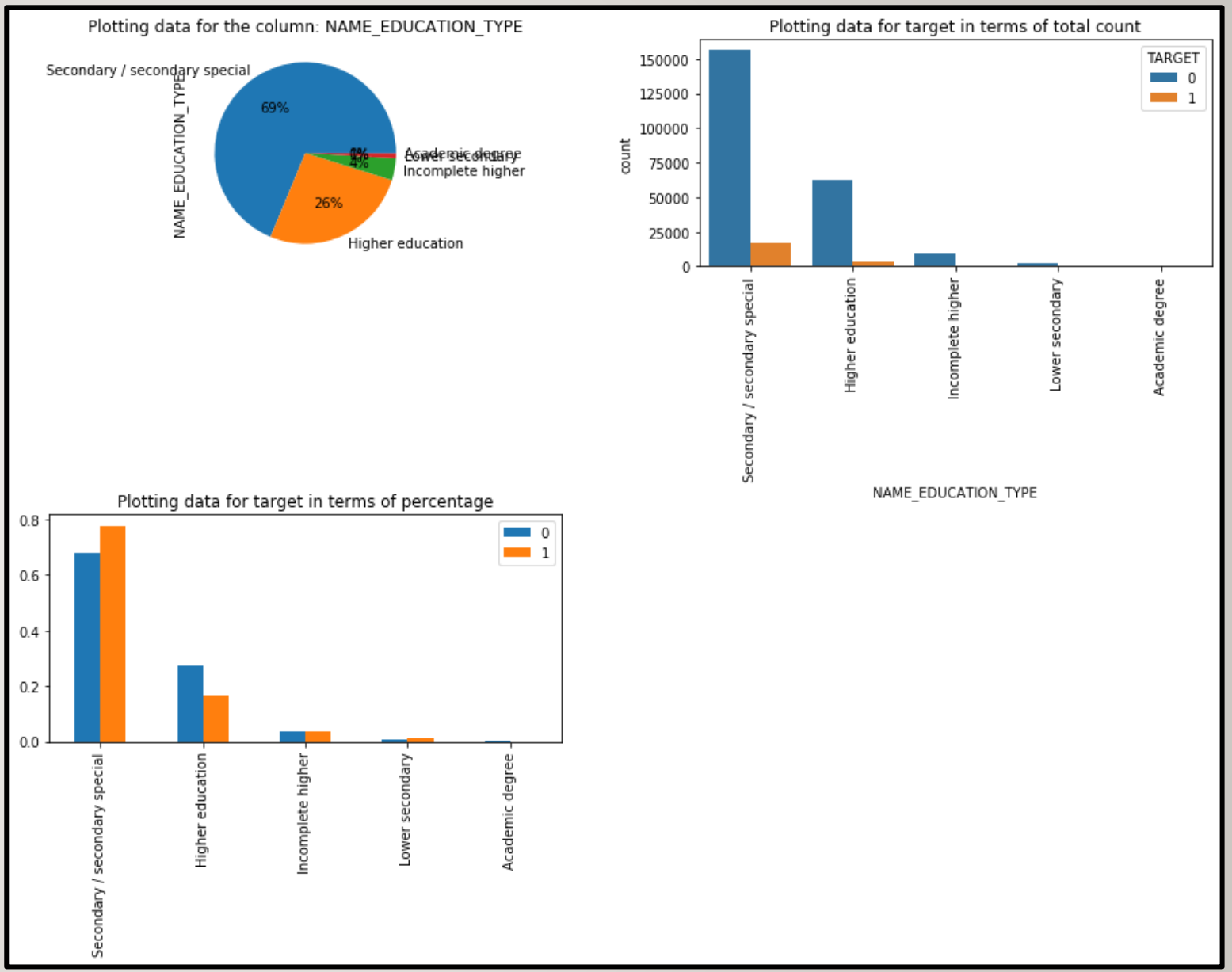


# UNIVARIATE ANALYSIS

# NAME_EDUCATION_TYPE with respect to TARGET variable

**INSIGHTS:**

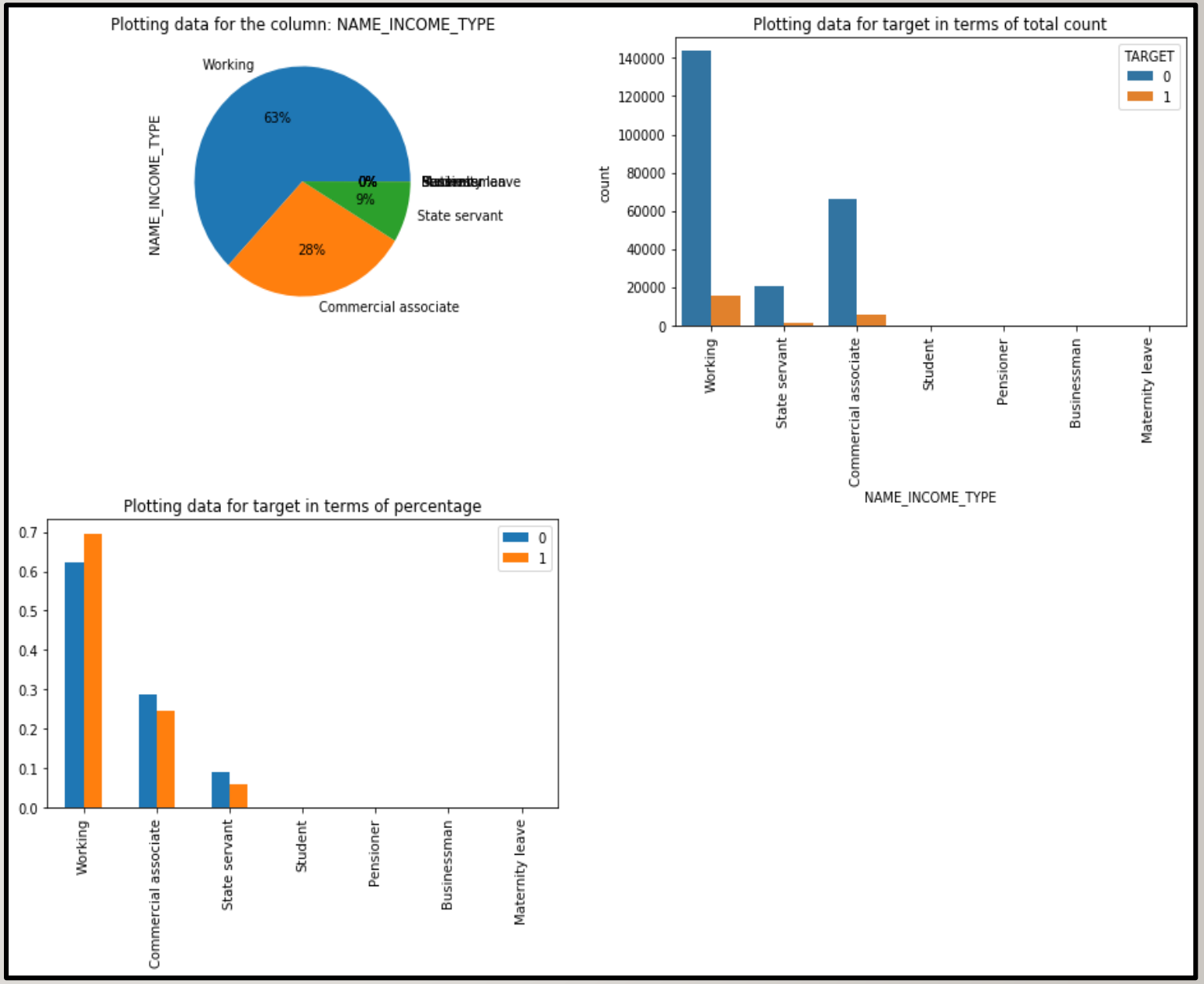
- Clients with Secondary Education are the leading loan applicants followed by clients with Higher Education.
- However, the percentage share of clients with loan payment difficulty is more for Secondary Education clients than any other level when compared with clients who can repay their loans.



# NAME_INCOME_TYPE with respect to TARGET variable

**INSIGHTS:**

- Working class people are leading in loan applications followed by Commercial associates.
- However, the percentage share of clients with loan payment difficulty is more for working class than any other class when compared with clients who can repay their loans.

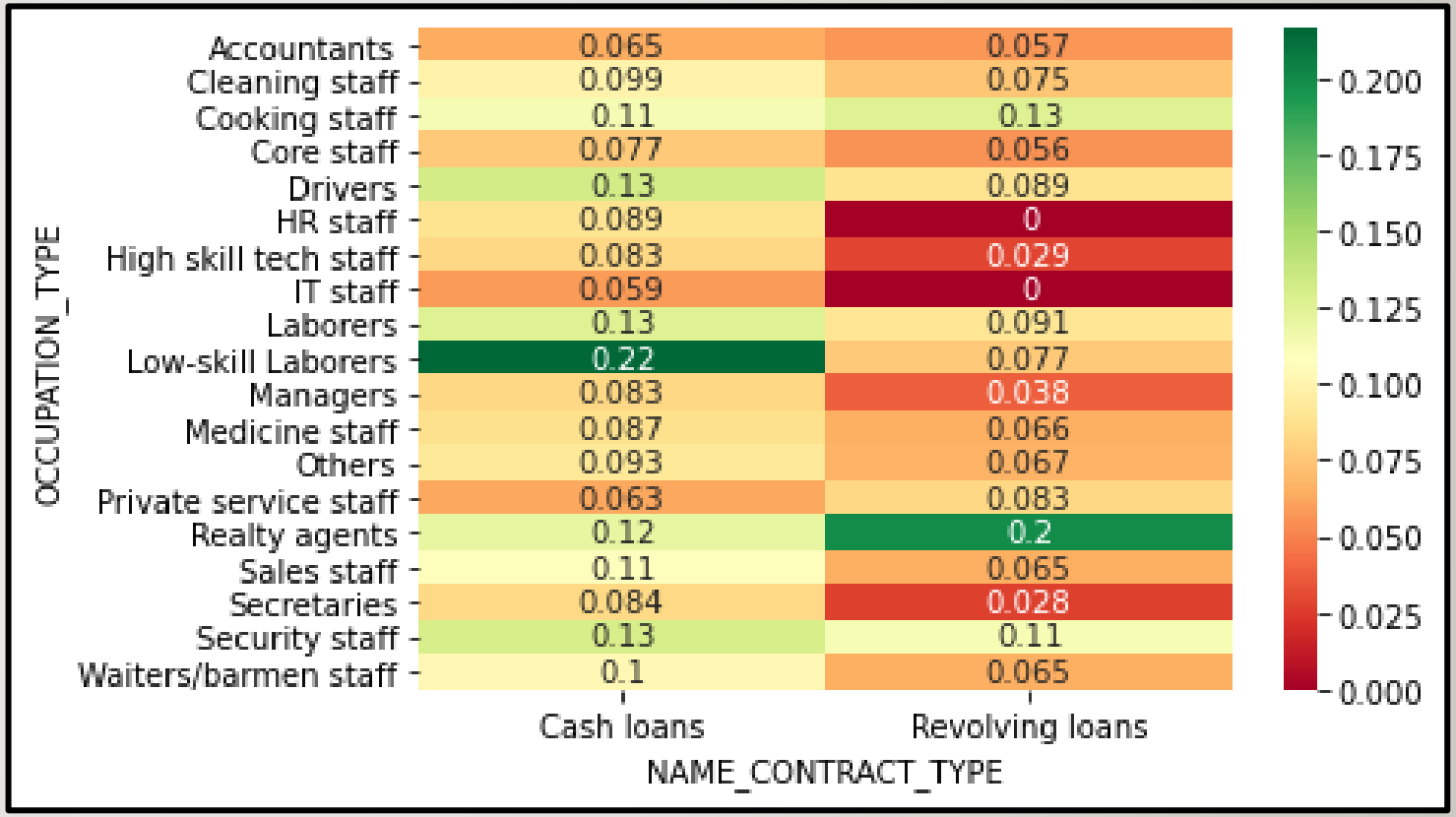


# BIVARIATE & MULTIVARIATE ANALYSIS

# NAME_CONTRACT_TYPE v/s. OCCUPATION_TYPE with respect to TARGET variable

**INSIGHTS:**

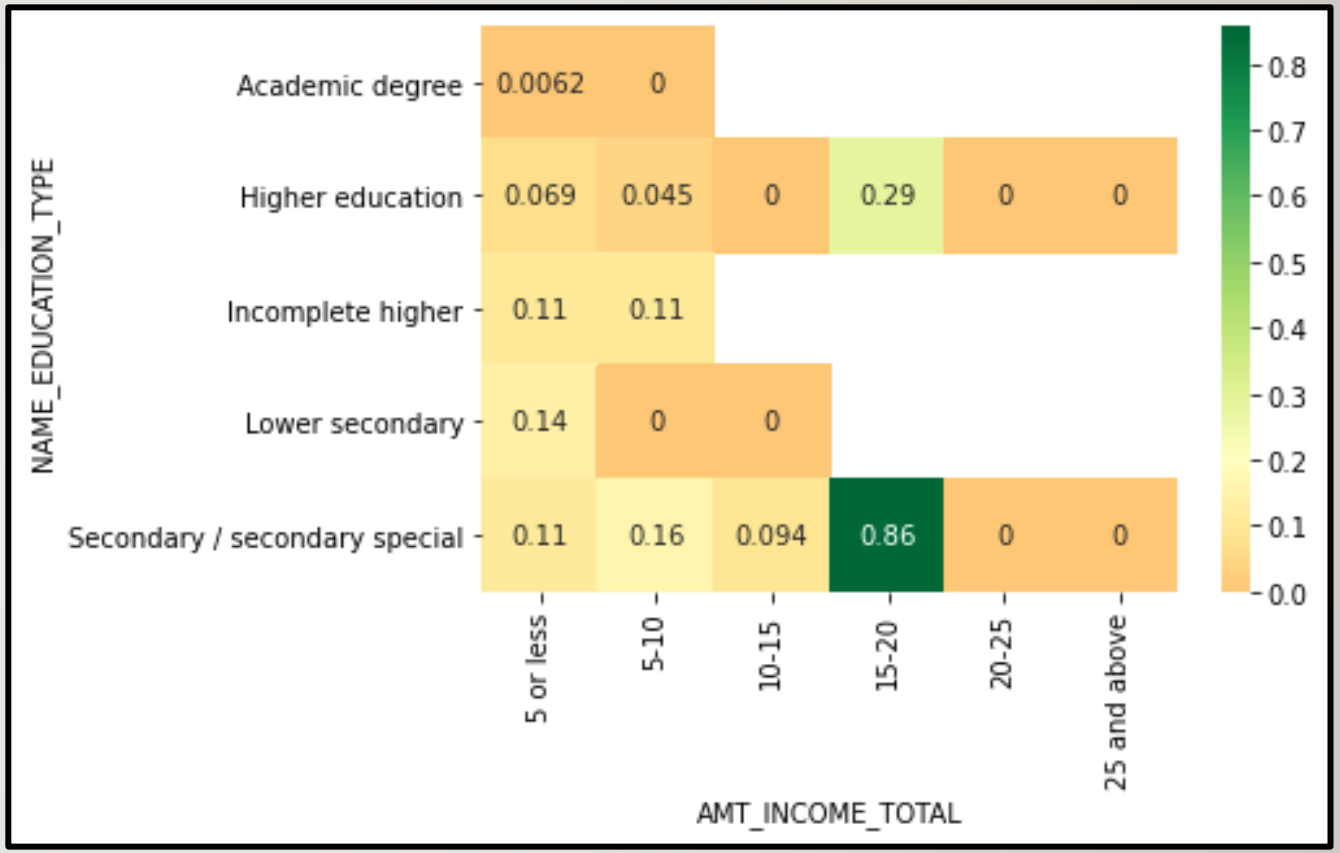
- In case of Cash loans, Low-Skill Laborers have higher difficulty in paying loans whereas IT staffs are able to pay their loans on time.
- In case of Revolving loans, Realty agents have higher difficulty in paying loans whereas HR & IT Staff are able to pay their loans on time.
- In general, people find it more difficult to re-pay their Cash loans than the Revolving loans.



# NAME_EDUCATION_TYPE v/s. AMT_INCOME_TOTAL with respect to TARGET variable

**INSIGHTS:**

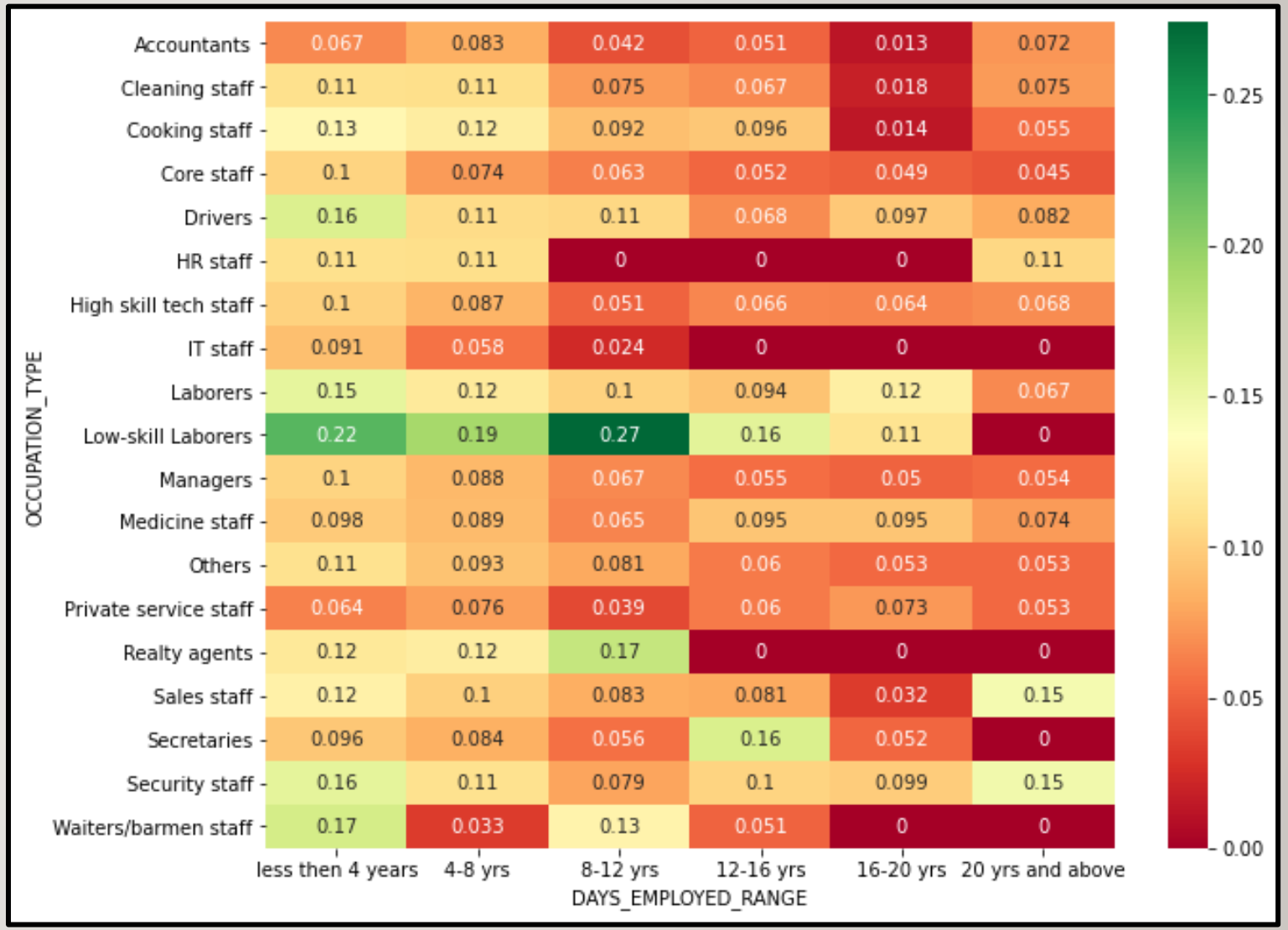
- Clients with income 20 lacs & above are able to pay their loan on time irrespective of education type.
- Clients with Secondary/Secondary special education level & within 15-20 lacs income group finds it most difficult to re-pay their loans on time.



# OCCUPATION_TYPE v/s. DAYS_EMPLOYED_RANGE with respect to TARGET variable

**INSIGHTS:**

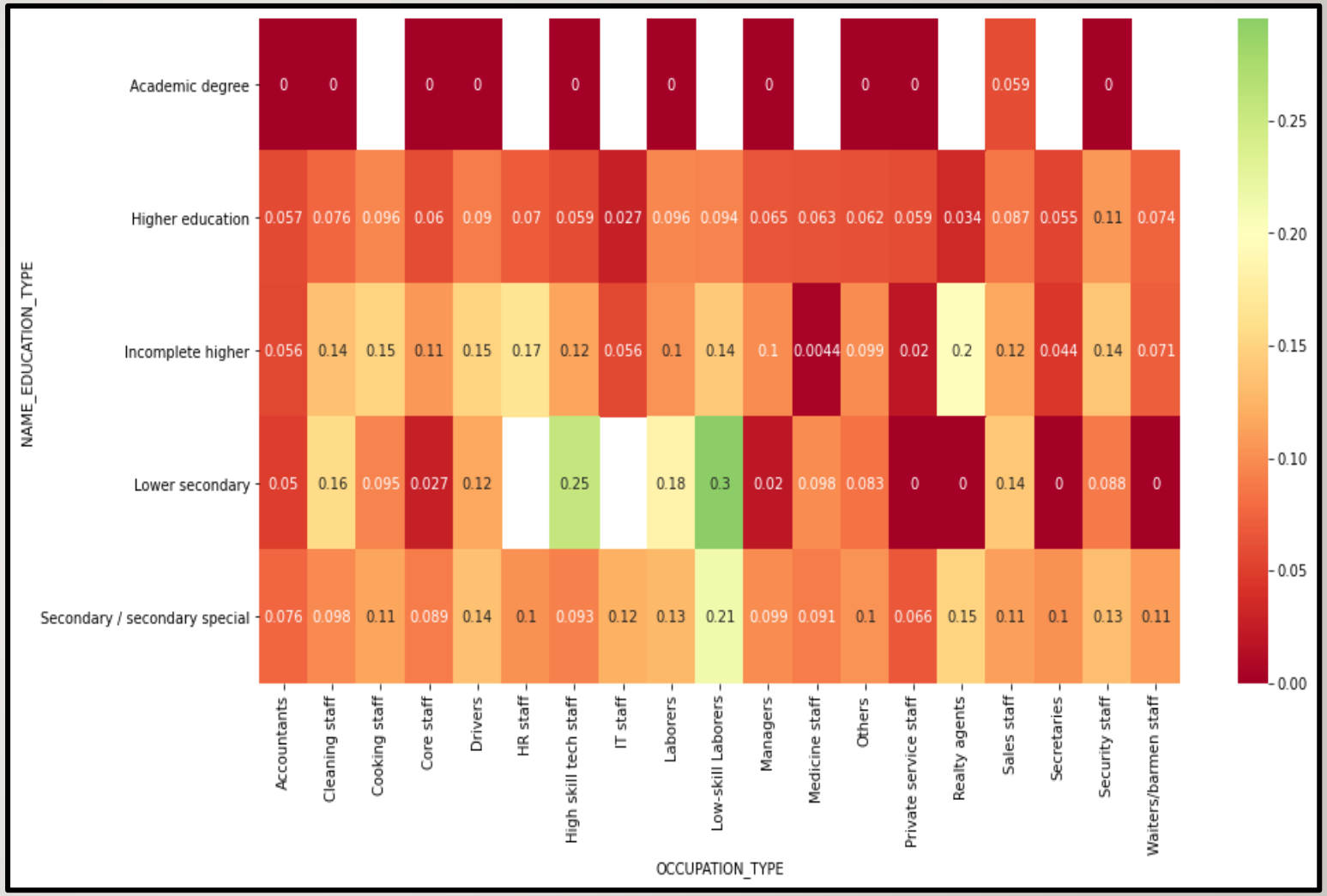
- Low-Skill Laborers with less then 12 years of exp. face more difficulties in repaying their loans then those with more then 12 years of exp.
- Looking at general trend, People with more then 12 years of experience are more capable of repaying loan.



# NAME_EDUCATION_TYPE v/s. OCCUPATION_TYPE with respect to TARGET variable

**INSIGHTS:**

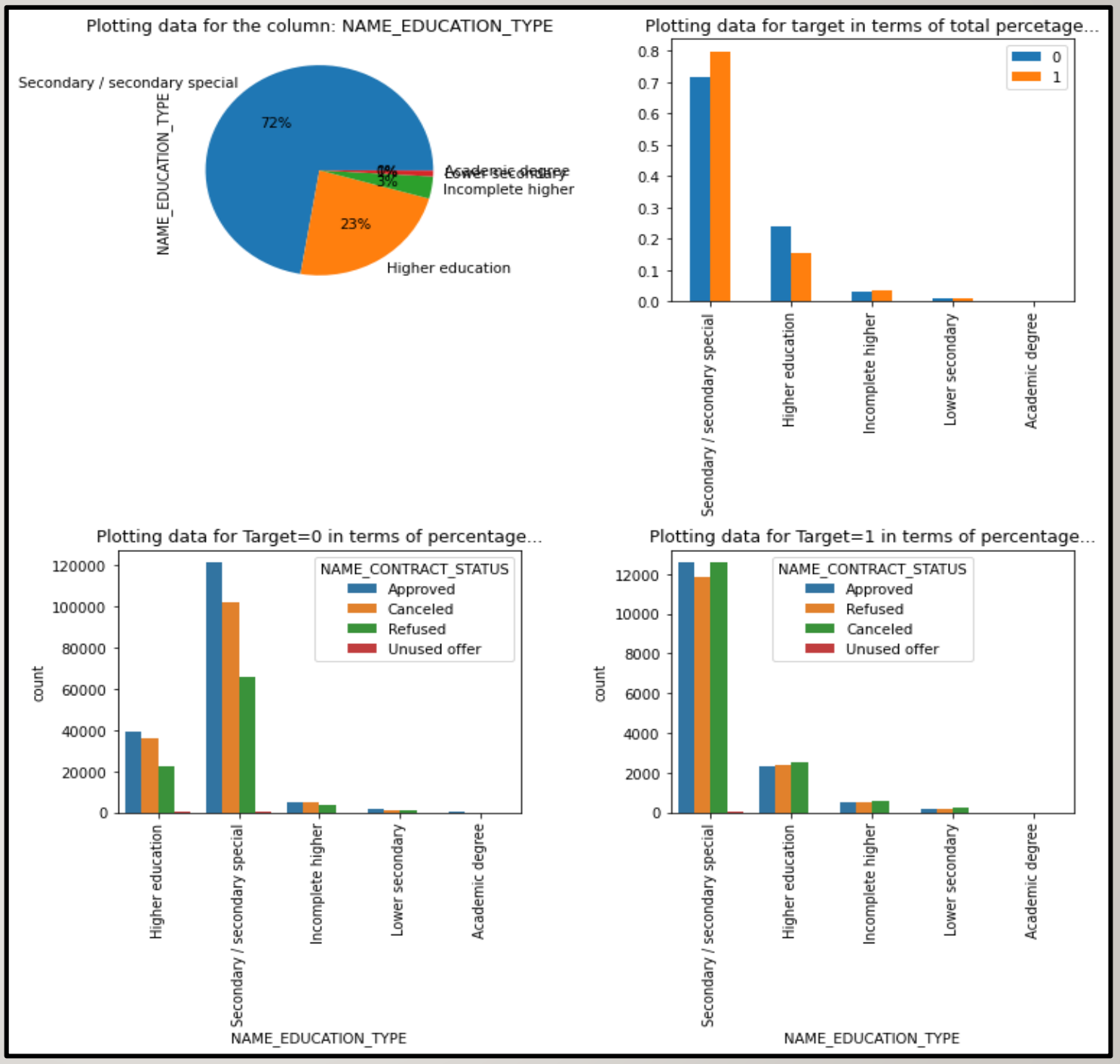
- Low-skill Laborers with lower secondary education finds it difficult to pay their loans on time.



# NAME_EDUCATION_TYPE v/s. NAME_CONTRACT_STATUS with respect to TARGET variable

**INSIGHTS:**

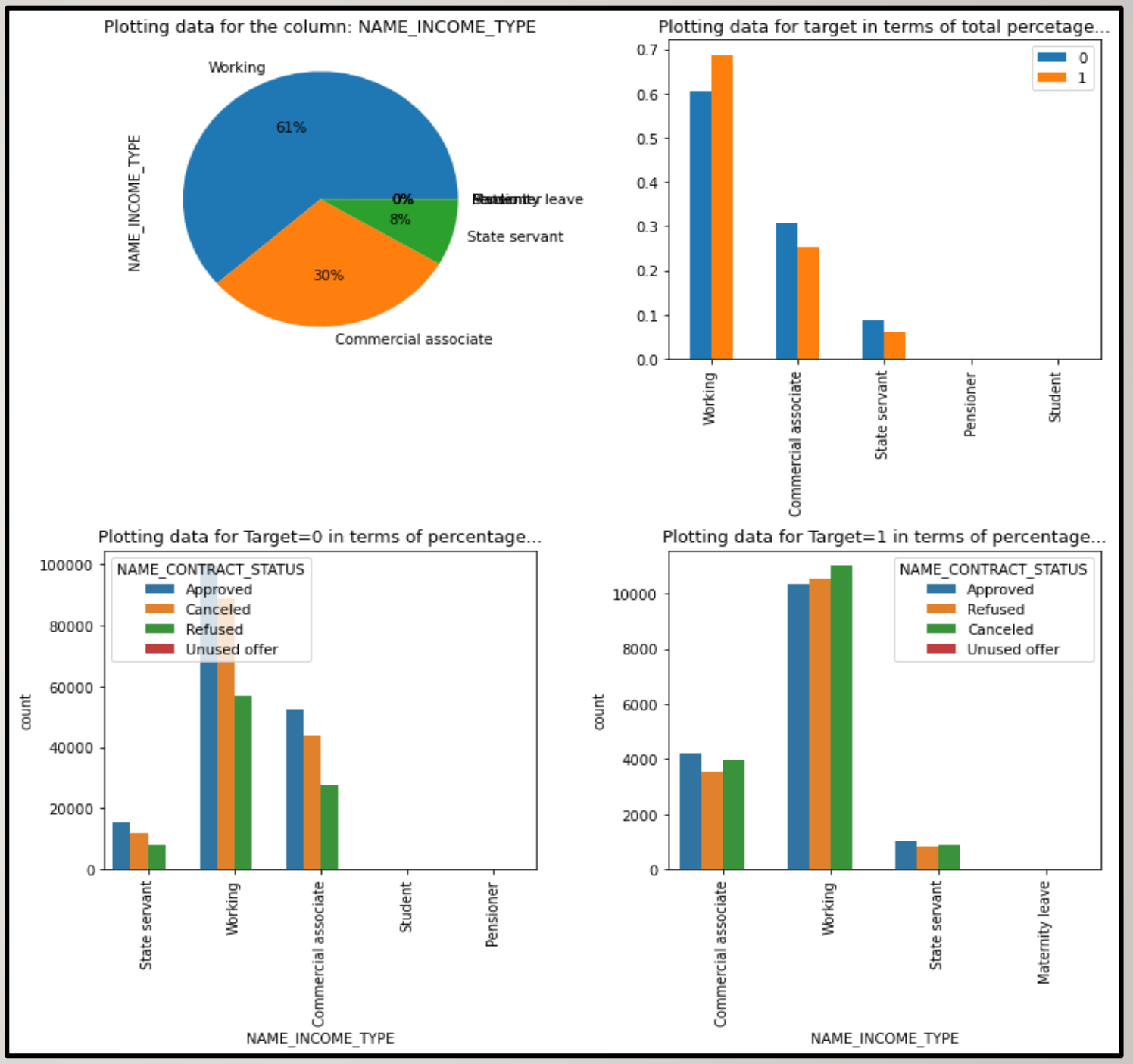
- Secondary/Secondary special are the highest applicants for loan
- However, Secondary/Secondary special have more difficulty in repaying loan then any other categories.



# NAME_INCOME_TYPE v/s. NAME_CONTRACT_STATUS with respect to TARGET variable

**INSIGHTS:**

- Although the working class clients have applied the most(61%) for the loan, they still tend to find it more difficult to re-pay their loans on time.
- That is why working class clients with payment difficulty have higher rejections & cancellations than approvals.

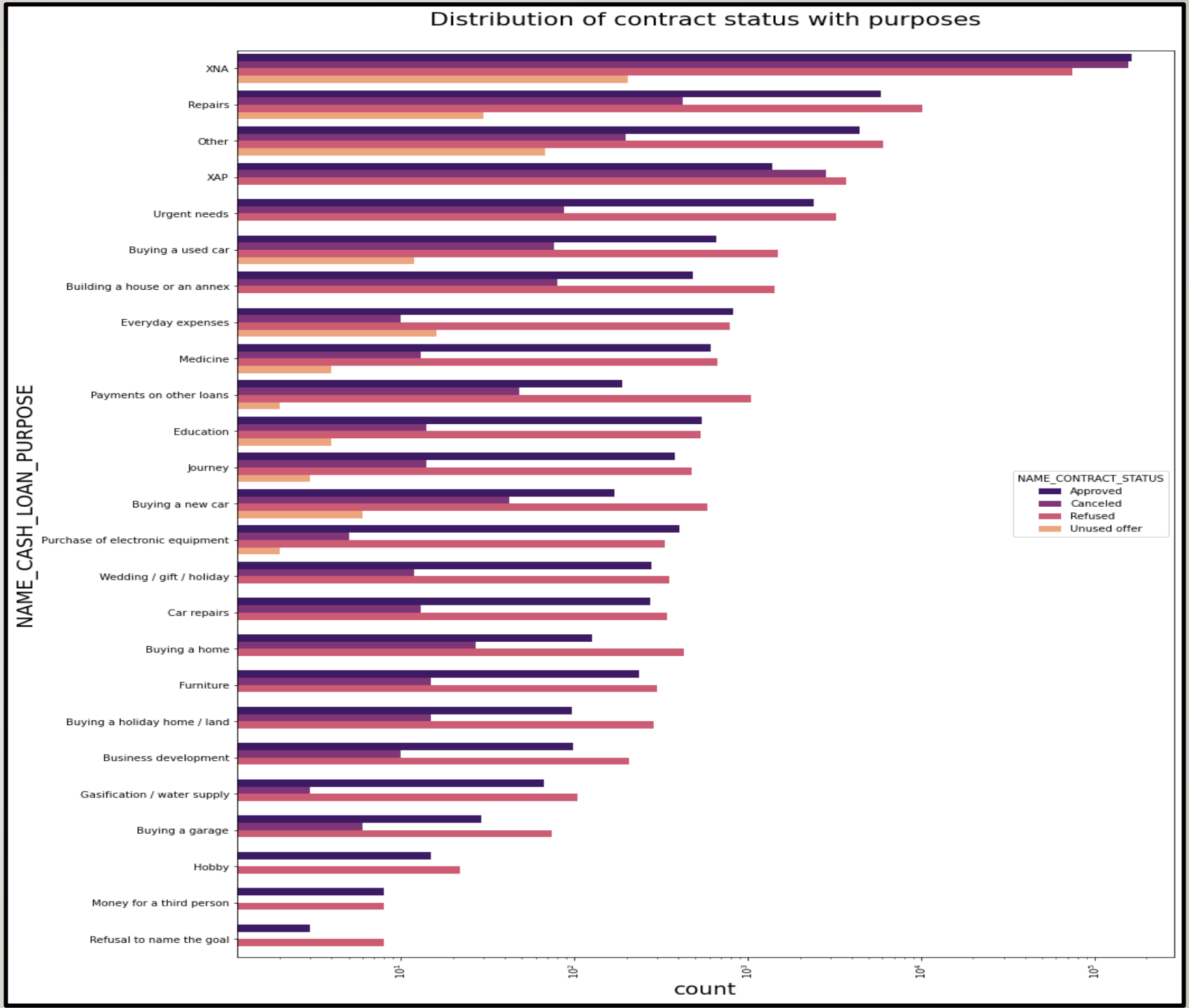


# NAME_CASH_LOAN_PURPOSE v/s. NAME_CONTRACT_STATUS

**INSIGHTS:**

*Assumption: Ignoring some of the values like 'XNA' & 'XAP' which looks irrelevant or junk.*

- Most rejection of loans came from the 'repairs' purpose.
- For 'Education' purposes we have almost equal number of approvals and rejections.
- 'Payments on other loans', 'Buying a House' and 'Buying a used/new car' has significantly higher loan rejections than approvals.



# THANK YOU ☺