Jagdish Galipelli
Data Science Program (International) - September 2021
Jagdish.dataengineer@gmail.com

DATA Summary

307511
Records of Loan Applicants
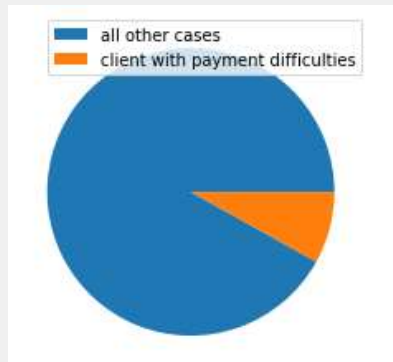
291057
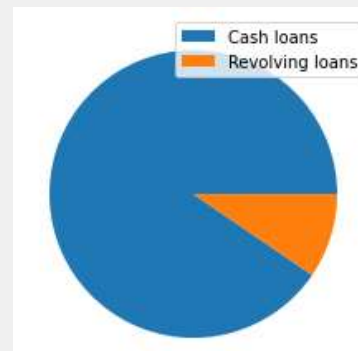Applicants having Previous applications
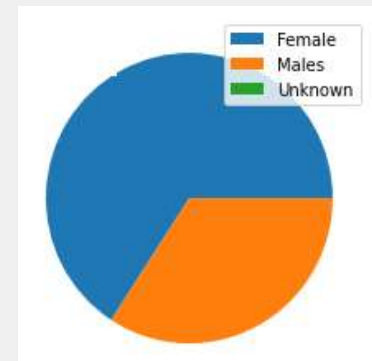
94.64%
Percentage

Data is Imbalanced

91% All other cases
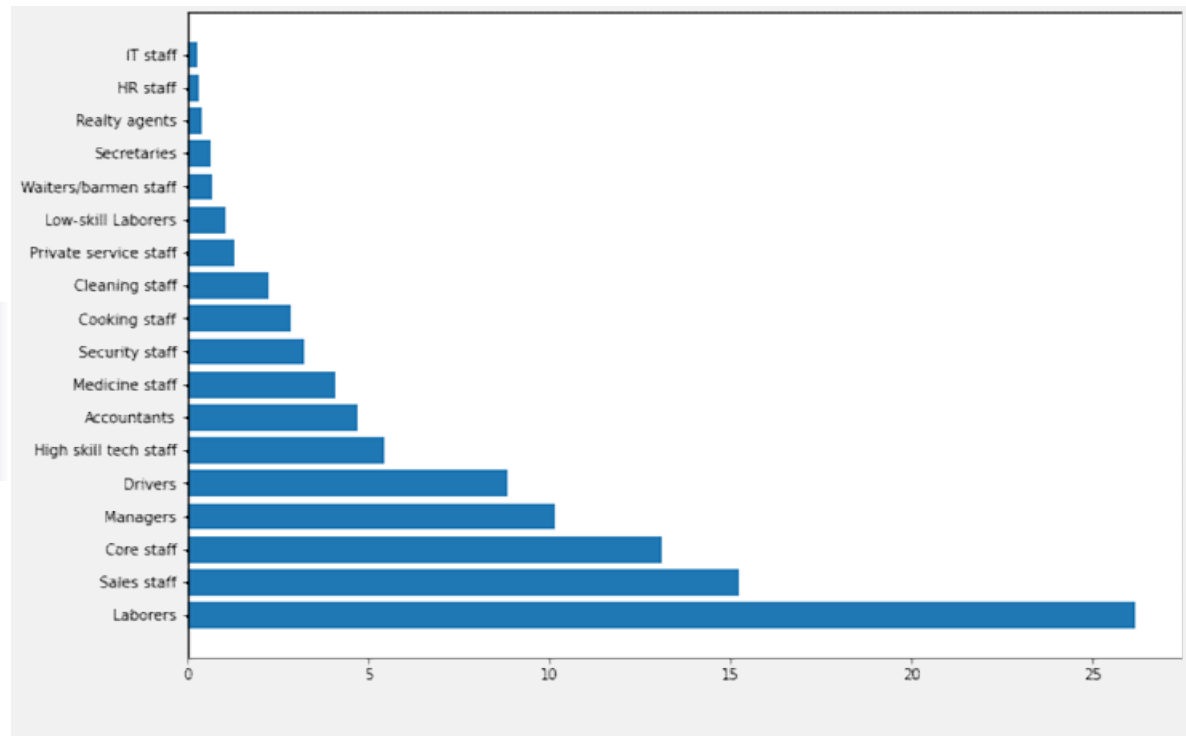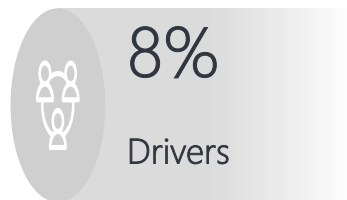8% Clients with Payment difficulties
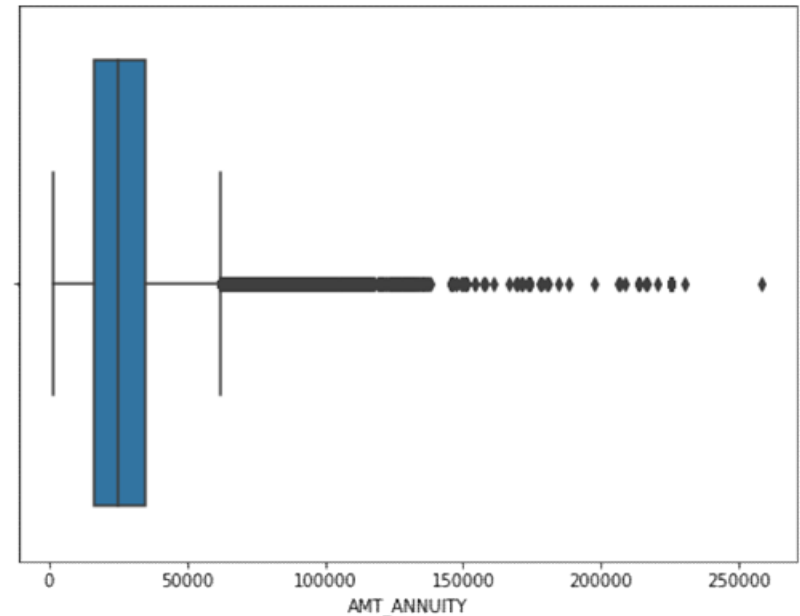

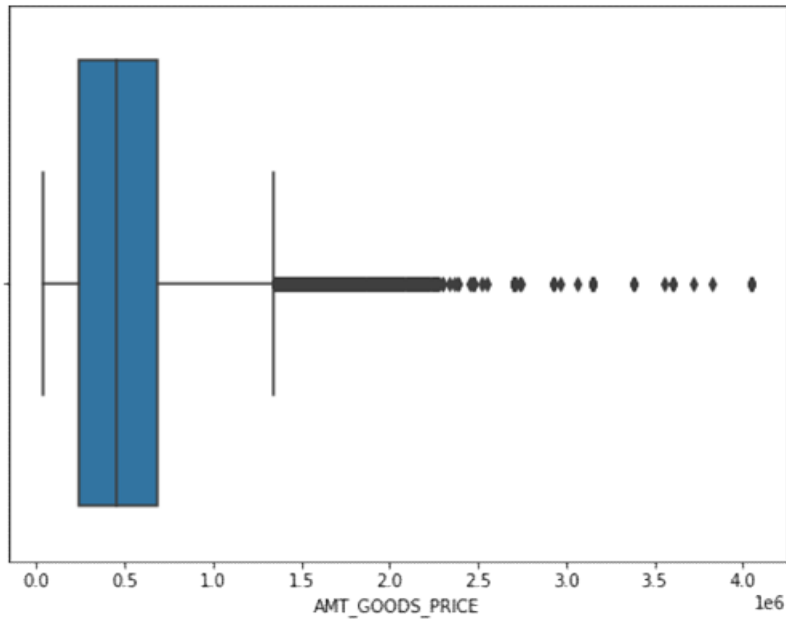90.5% Cash Loans
9.5% Revolving Loans

65% Females
34% Males
<1 Others

Occupation type Analysis

26% Laborers

15% Sales staff

13% Core staff

10% Managers

8% Drivers

5% High skilled t...

IT staff
HR staff
Realty agents
Secretaries
Waiters/barmen staff
Low-skill Laborers
Private service staff
Cleaning staff
Cooking staff
Security staff
Medicine staff
Accountants
High skill tech staff
Drivers
Managers
Core staff
Sales staff
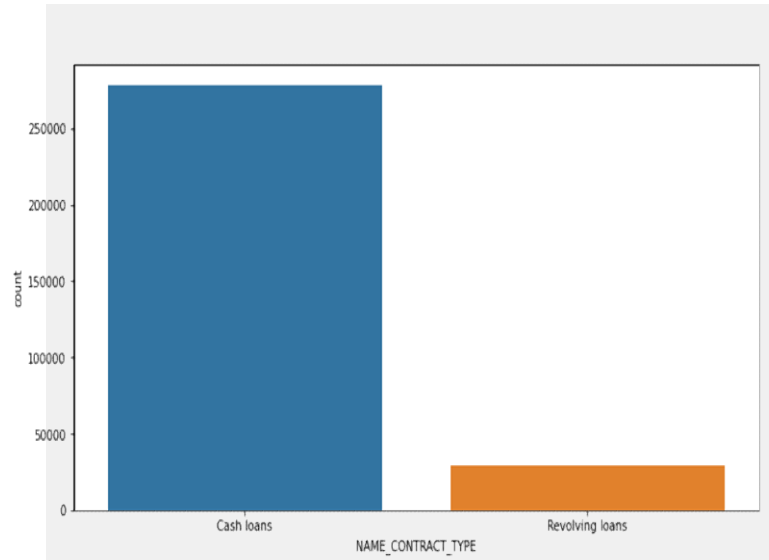Laborers

0   5   10   15   20   25

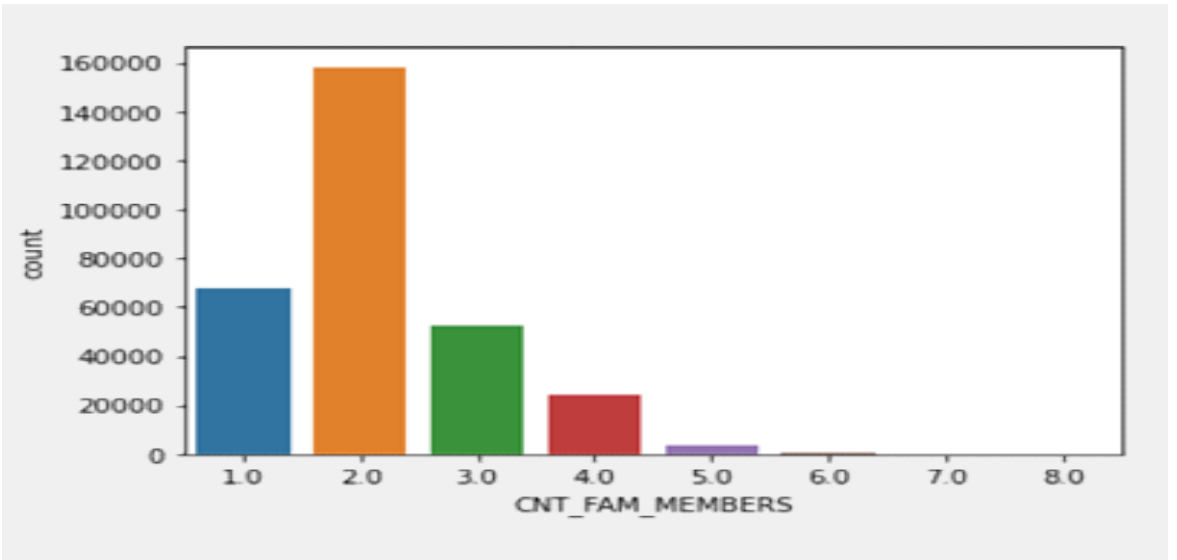# Amount Goods Price and Amount Annuity data distribution

We can see some outliers both in Goods price and Amount Annuity
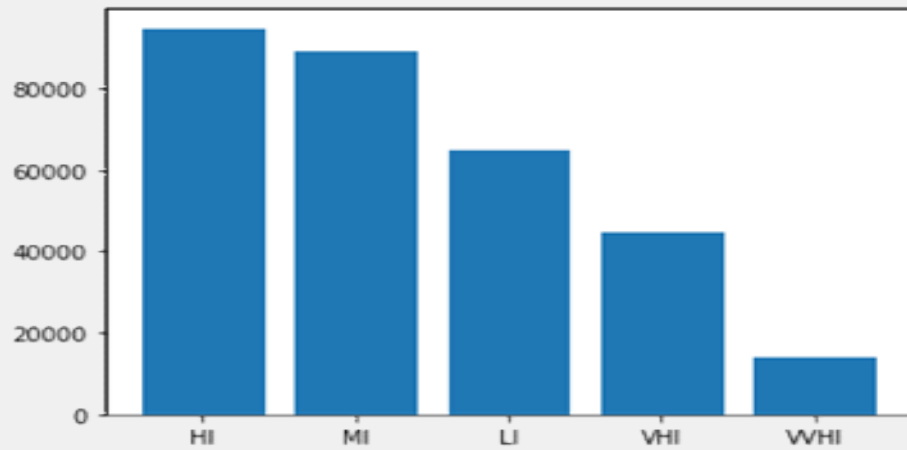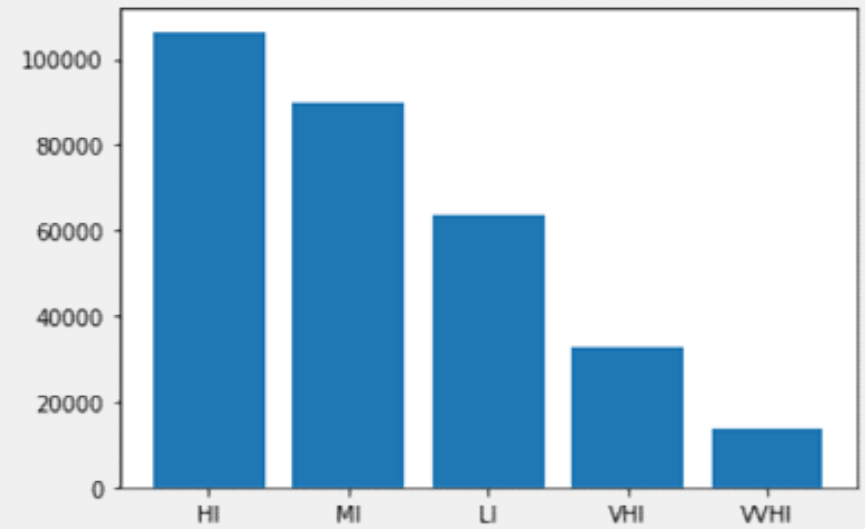
Contract Type

Family members count

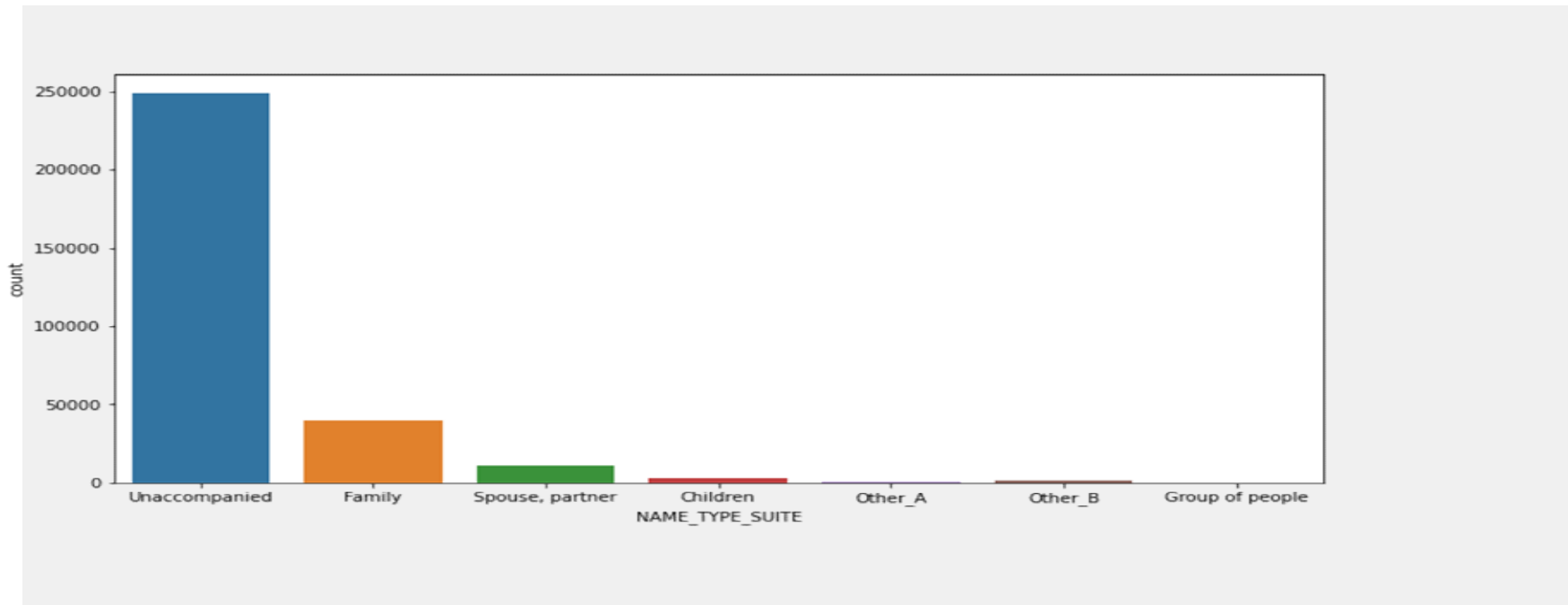Business in terms of Cash loans is higher

Amount Credit bins

Income bins

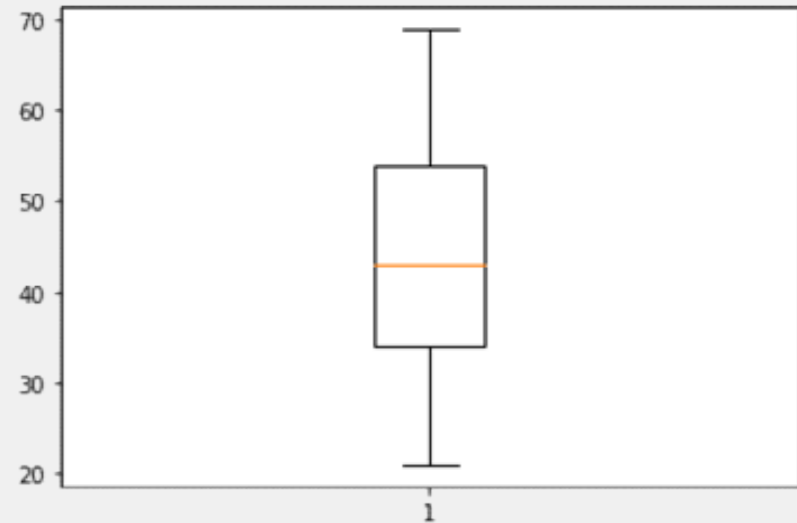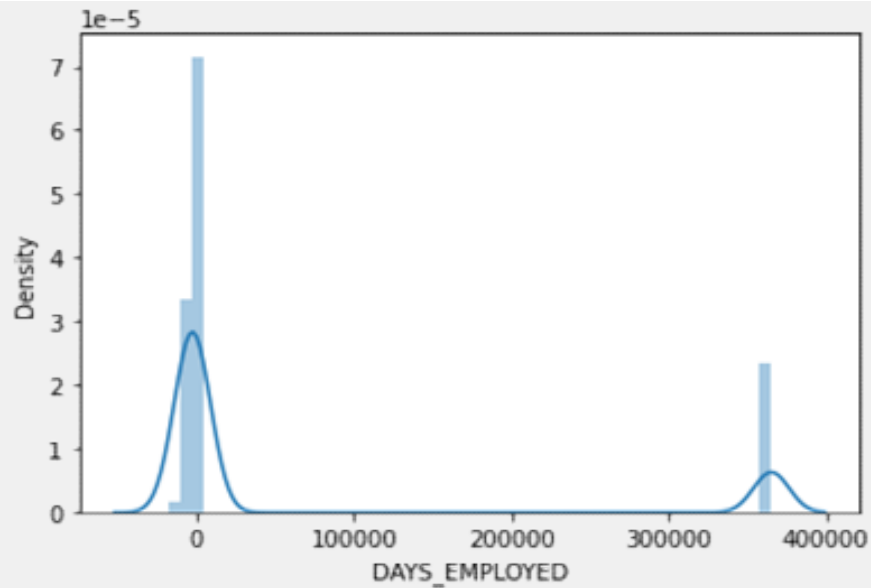High Income and Mid income applicants are more than others

Type Suite Analysis

Majority of them are Unaccompanied category
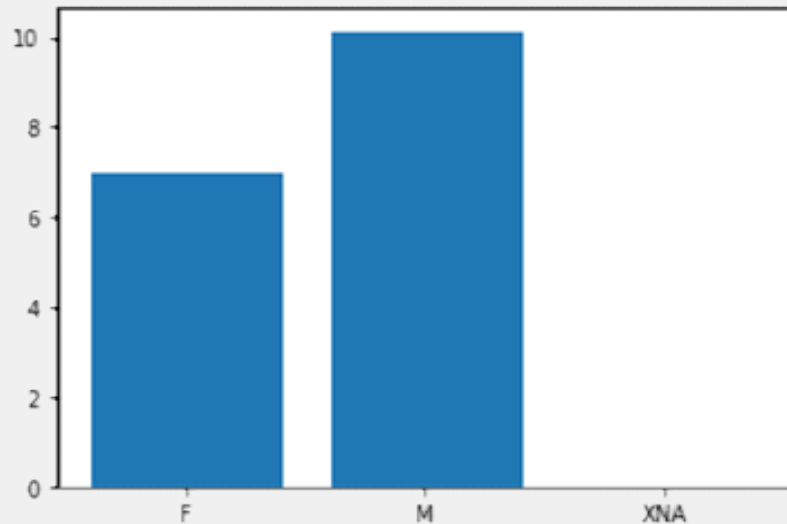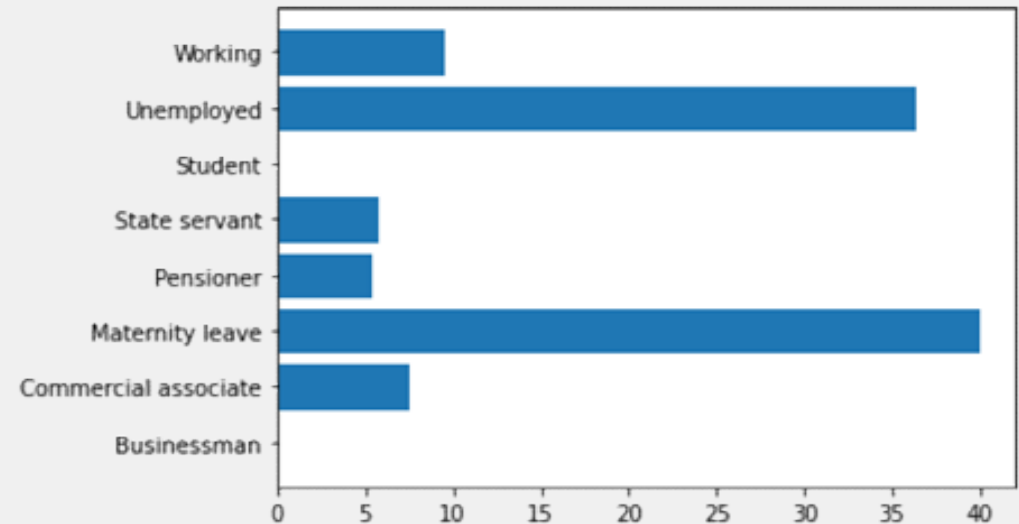
Days Employed

Applicant Age

Some Employees have very high days employed. This might be due to typo error or some default value where data was not available
Applicant age doesn't have outlier values

- clearly shows the Student and Businessman category doesn't have defaulter
- Most of the defaulters are from Unemployed and Maternity category
- Males tend to default more than Females

Education type v/s Target
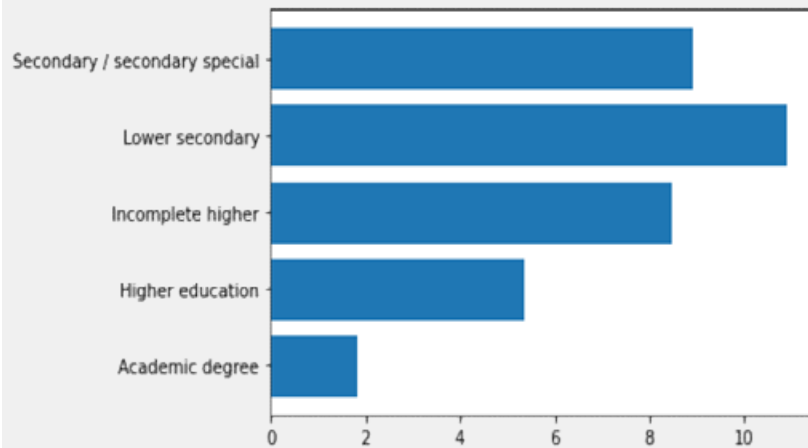
contract type v/s Target
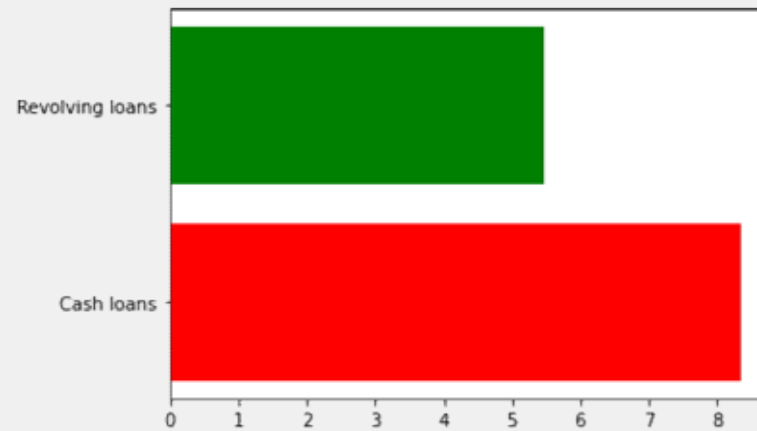
- Cash loans are more defaulters than revolving loans
- People with lower secondary are highest defaulter and then followed by Secondary & Incomplete higher
- People with Academic degree are lowest defaulters

Own Car v/s Target
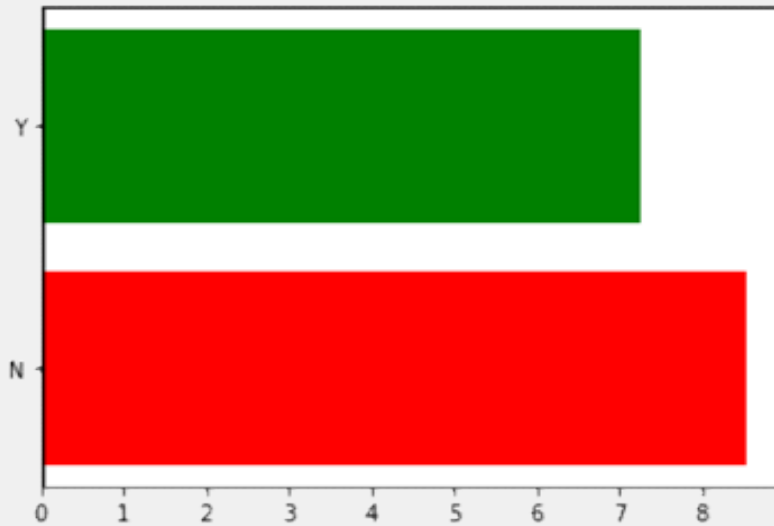
Own Realty v/s Target
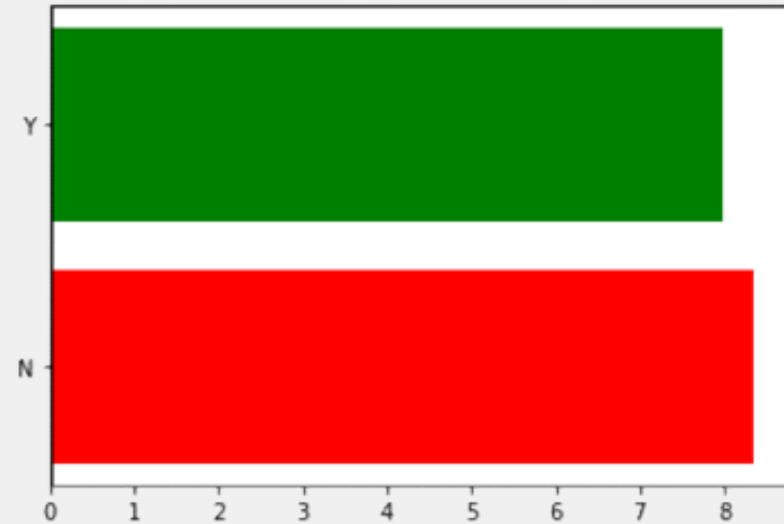
- We can see people with own car are less defaulter than without car people
- Negligible and no much difference wrt defaulting and having own realty
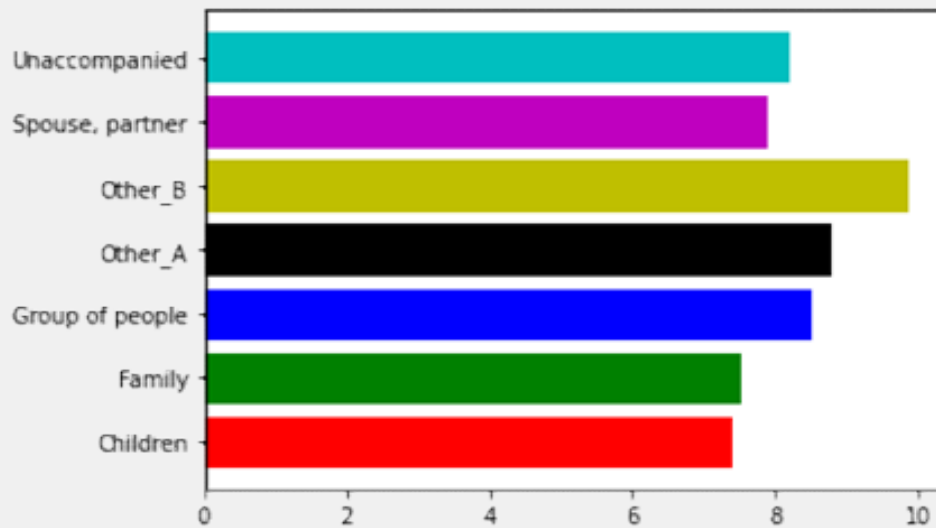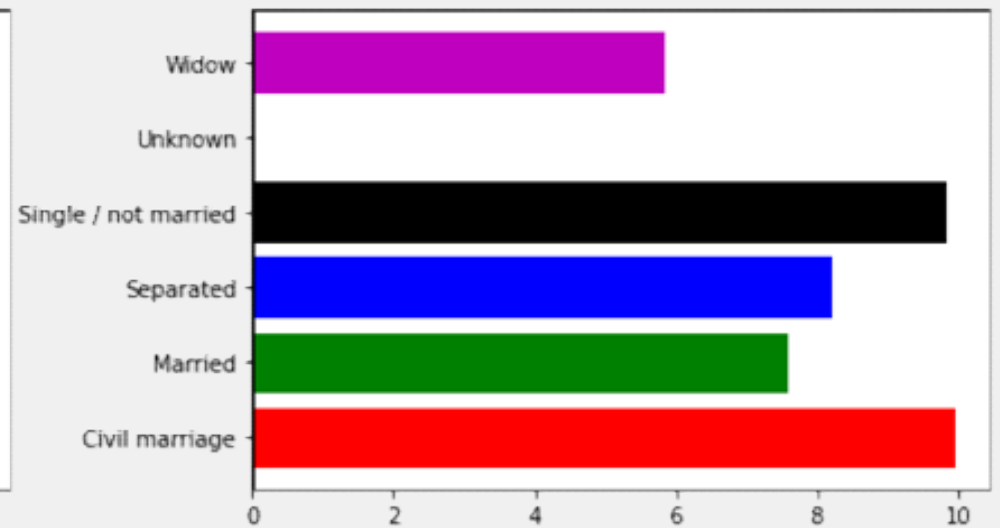
- We can infer that people with civil marriage and single/not married category are more defaulters than others

- clearly we can infer that low-skill laborers tend to default more and accountants are least defaulters
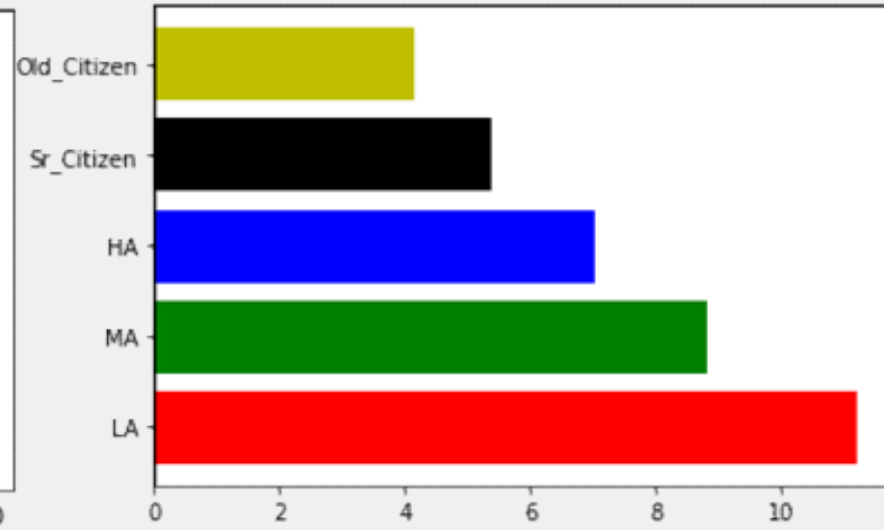
- Middle Age and Higher Age counts are almost same but Higher Age bracket category are lesser defaulters
- LA (Lower age) Age brackets have highest defaulters
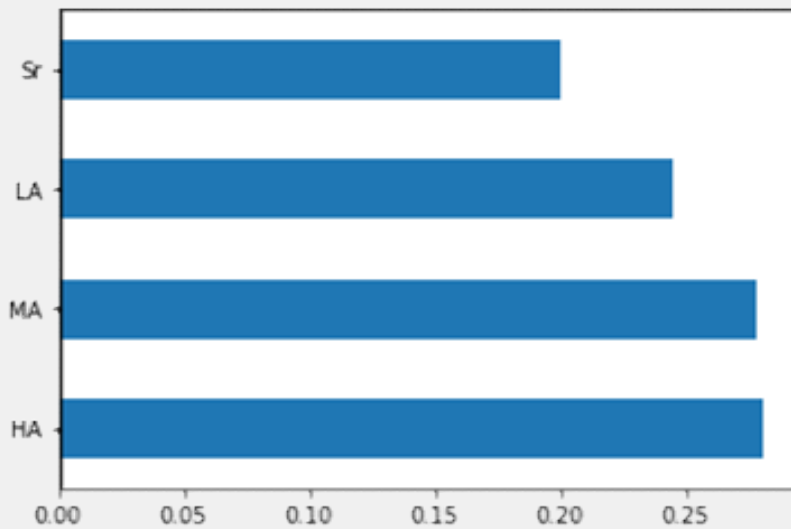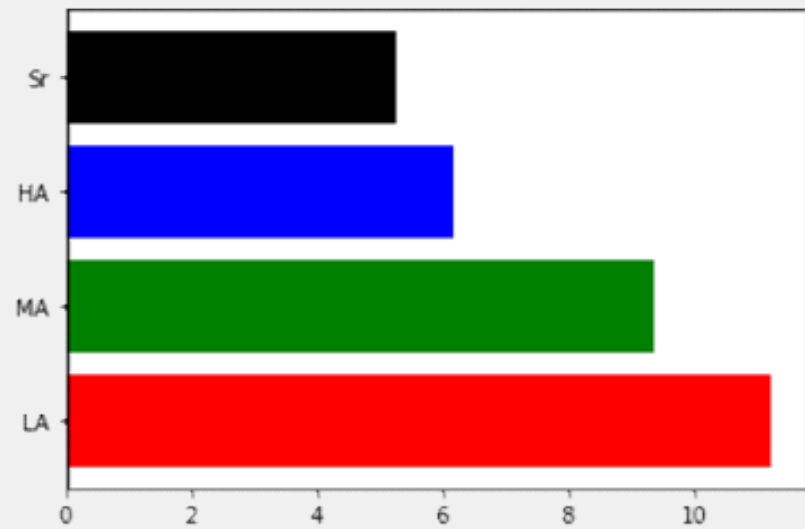
- Middle Age and Higher Age Employment counts are almost same but Higher Age bracket category are lesser defaulters
- LA (Lower age) Age brackets have highest defaulters

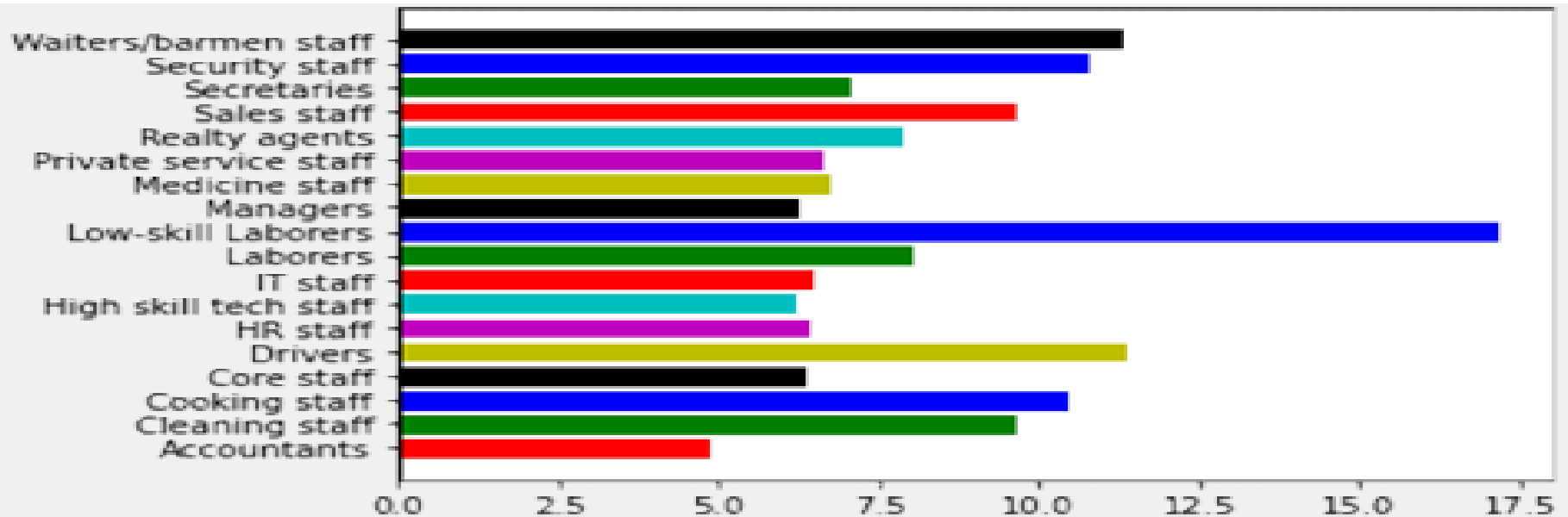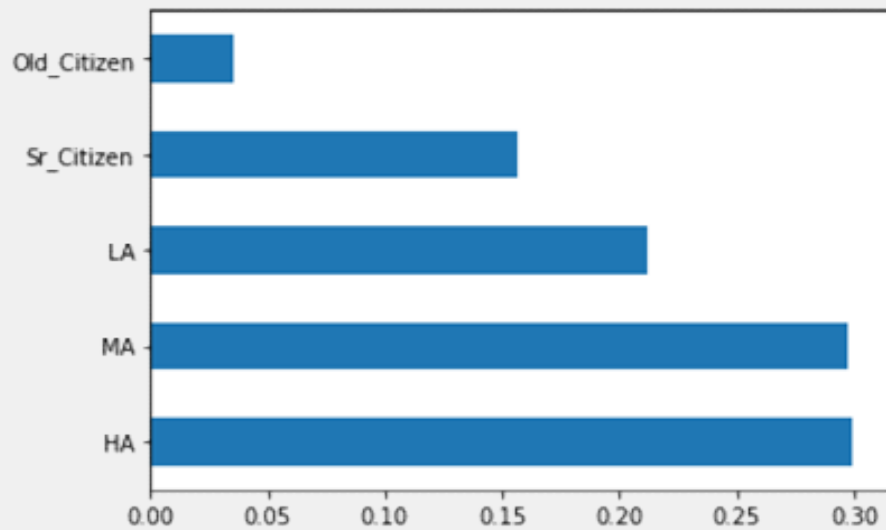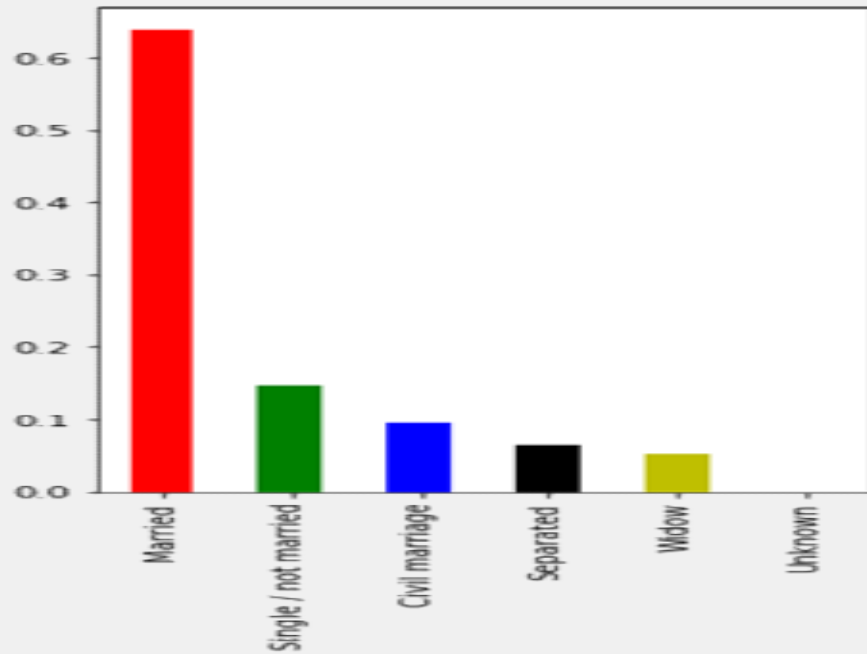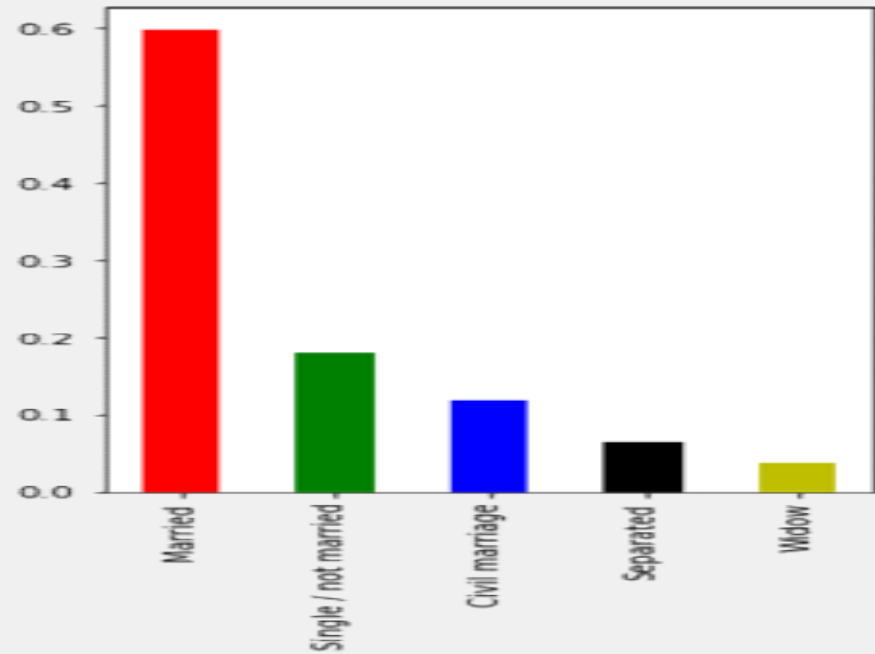- We can infer that people with civil marriage and single/not married category are more defaulters than others
- Married category has less defaulters
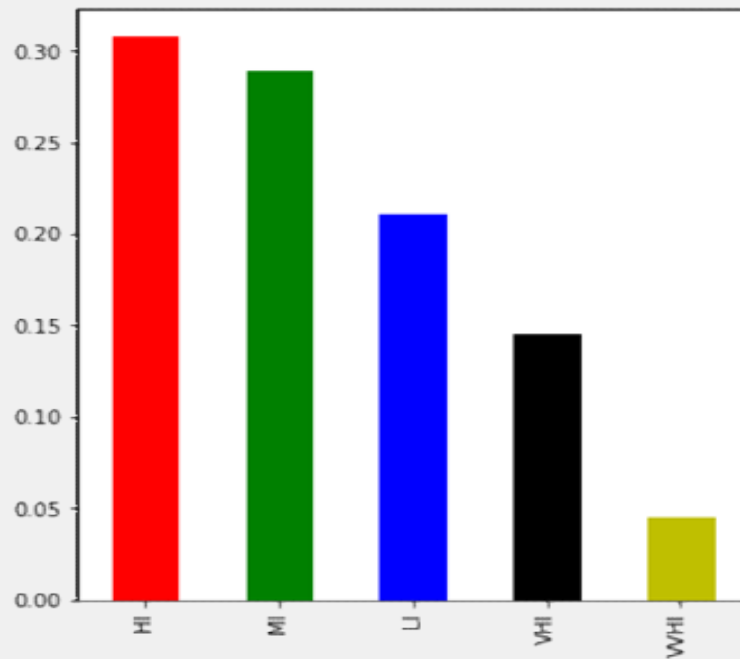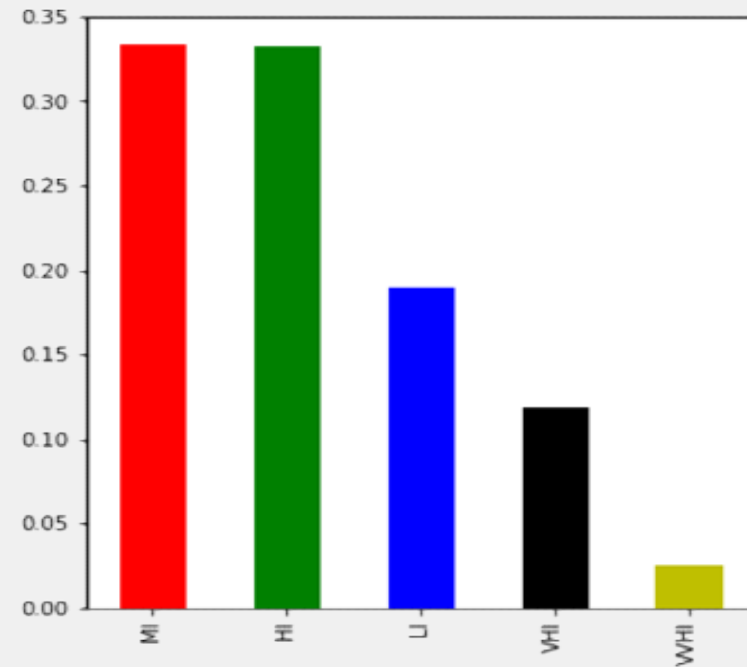
- We can infer that people with middle amount and high amount category are more defaulters than others
- Very very high amount category has less defaulters

- We can infer that people with Low income, middle income and high income category are more defaulters.
- Very very high income category has less defaulters

Top correlated variables

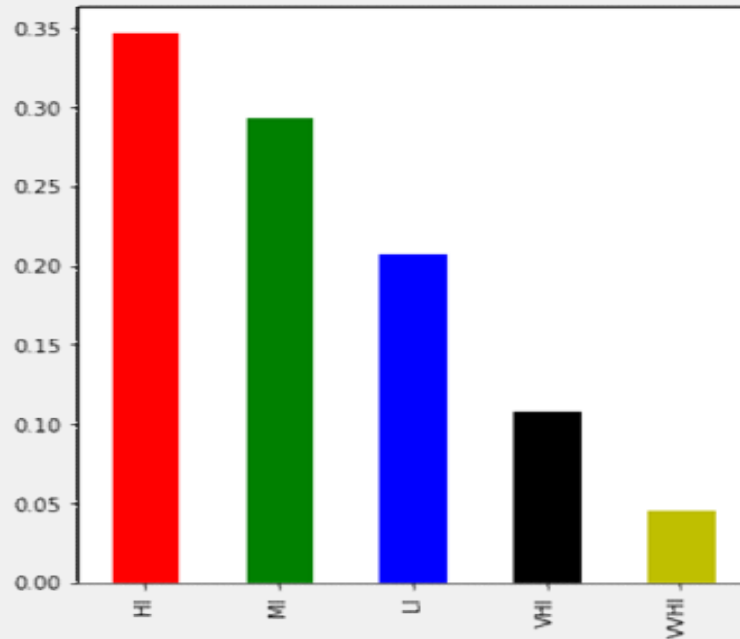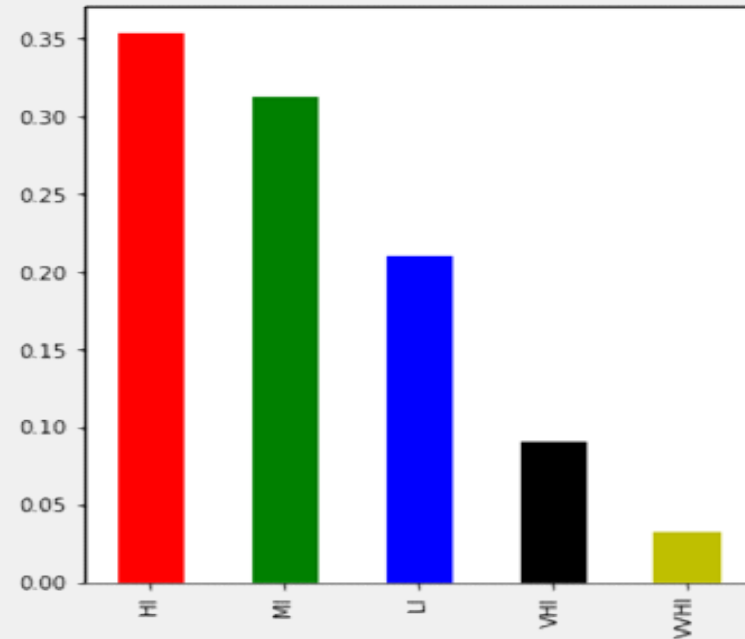| Variable 1	Variable 2	Corr Value
OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.998489543
AMT_CREDIT	AMT_GOODS_PRICE	0.986734292
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.950842214
CNT_FAM_MEMBERS	CNT_CHILDREN	0.876336303
REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	0.86062678
DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.86051738
LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.825574723
AMT_GOODS_PRICE	AMT_ANNUITY	0.774836604
AMT_CREDIT	AMT_ANNUITY	0.770126723
DAYS_BIRTH	DAYS_EMPLOYED	0.623940917
REGION_POPULATION_RELATIVE	REGION_RATING_CLIENT	0.532876531
REGION_POPULATION_RELATIVE	REGION_RATING_CLIENT_W_CITY	0.531535452
REG_REGION_NOT_LIVE_REGION	REG_REGION_NOT_WORK_REGION	0.450803754
REG_CITY_NOT_WORK_CITY	REG_CITY_NOT_LIVE_CITY	0.440409218
REG_CITY_NOT_LIVE_CITY	REG_REGION_NOT_LIVE_REGION	0.339232344
DAYS_REGISTRATION	DAYS_BIRTH	0.331912082
DEF_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.331570806
CNT_CHILDREN	DAYS_BIRTH	0.330937668
DEF_30_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.329338113
EXT_SOURCE_2	REGION_RATING_CLIENT	0.292895312

- Male applicants with previous contract status as "Refused" are most defaulters
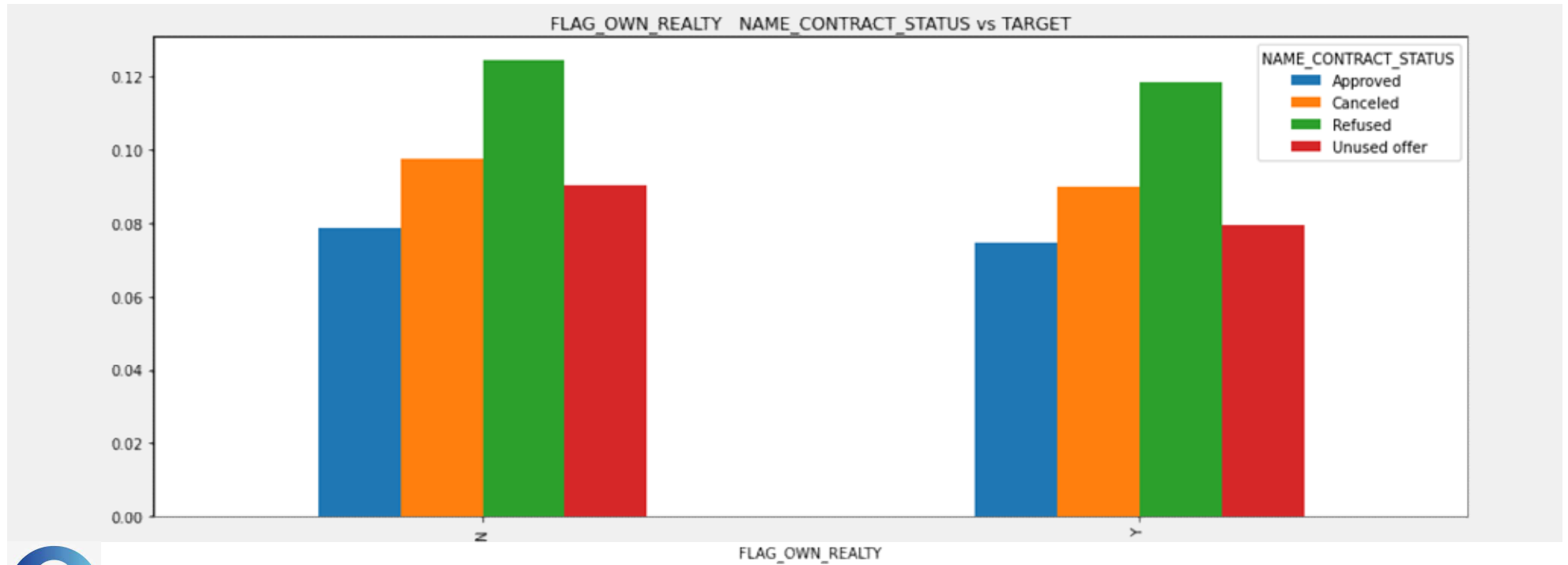- Refused earlier applicants irrespective of gender have most defaulters
- Cancelled previous applications also have defaulters quite high

FLAG_OWN_CAR NAME_CONTRACT_STATUS vs TARGET
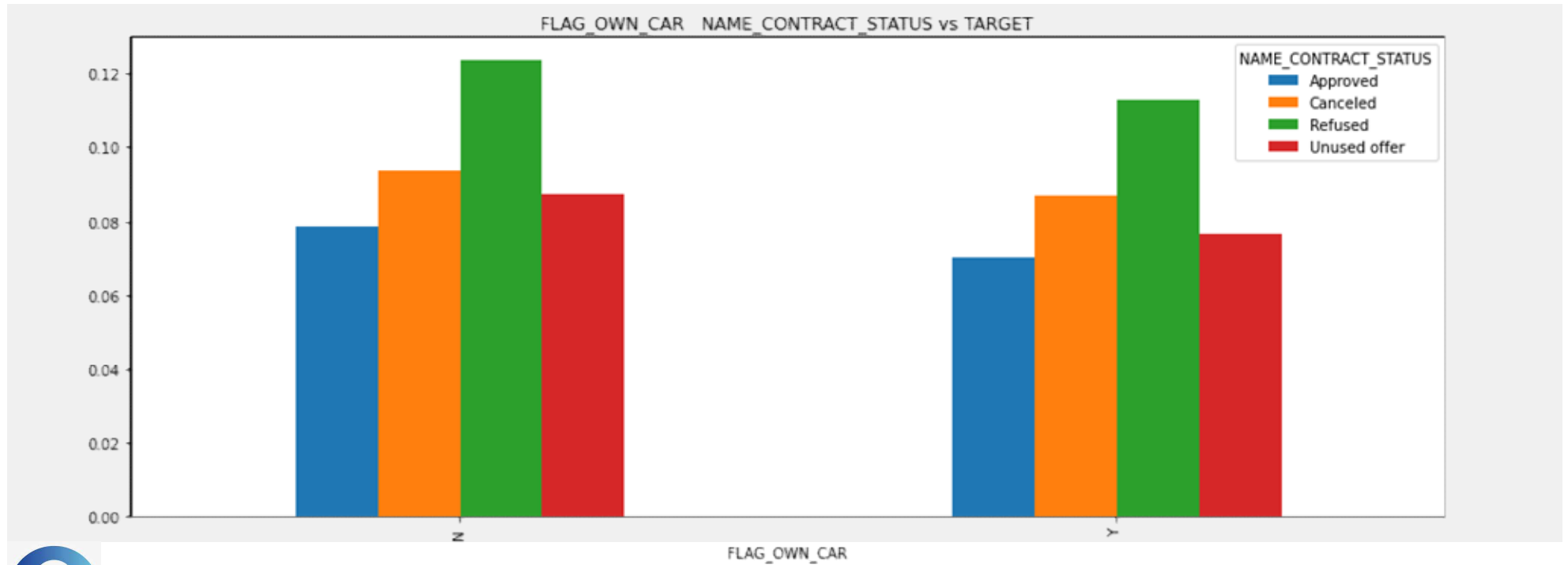
- Slight percentage difference with people having own car.

FLAG_OWN_REALTY   NAME_CONTRACT_STATUS vs TARGET

- Own realty have very slight impact wrt defaulter status

OCCUPATION_TYPE NAME_CONTRACT_STATUS vs TARGET
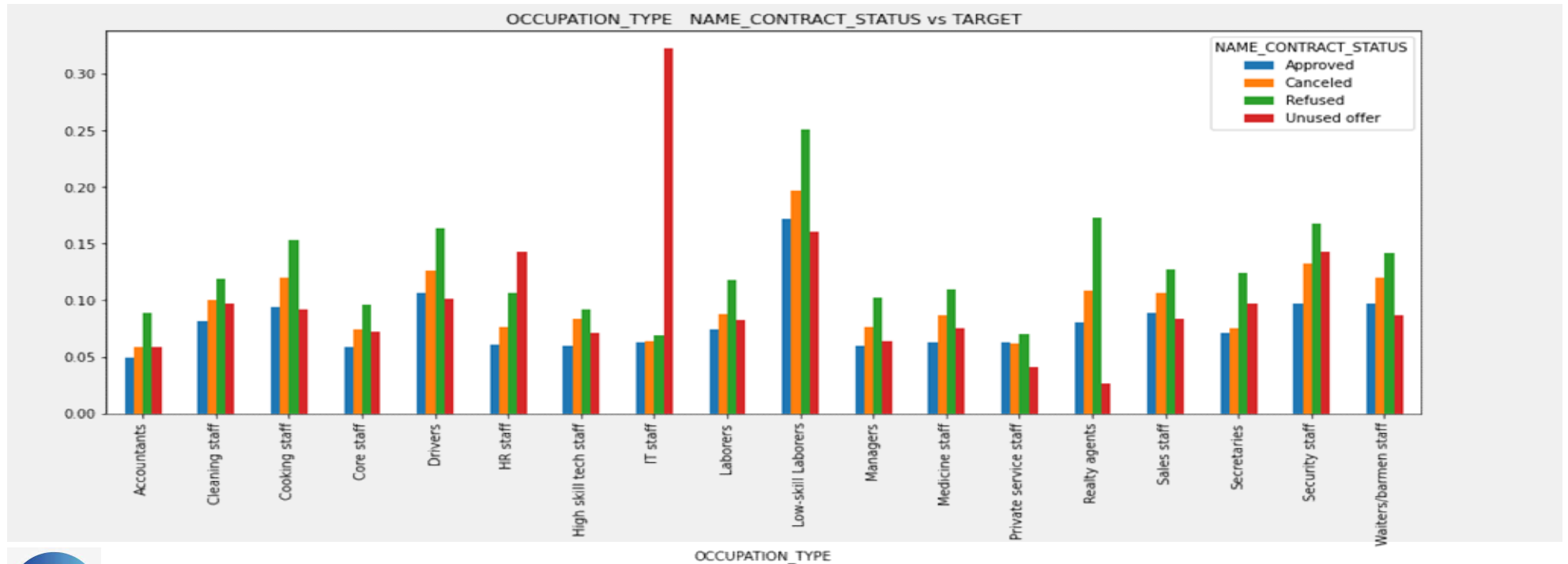
- Many of the IT staff did not used the offer earlier has high defaulters
- Low skill laborer have most refusals in previous app are defaulters
- cooking staff has moderate approval and unused offer
- HR Staff has many unused offers in previous applications
- realty agents who got refusal in earlier in previous application are now defaulters

NAME_HOUSING_TYPE   NAME_CONTRACT_STATUS vs TARGET
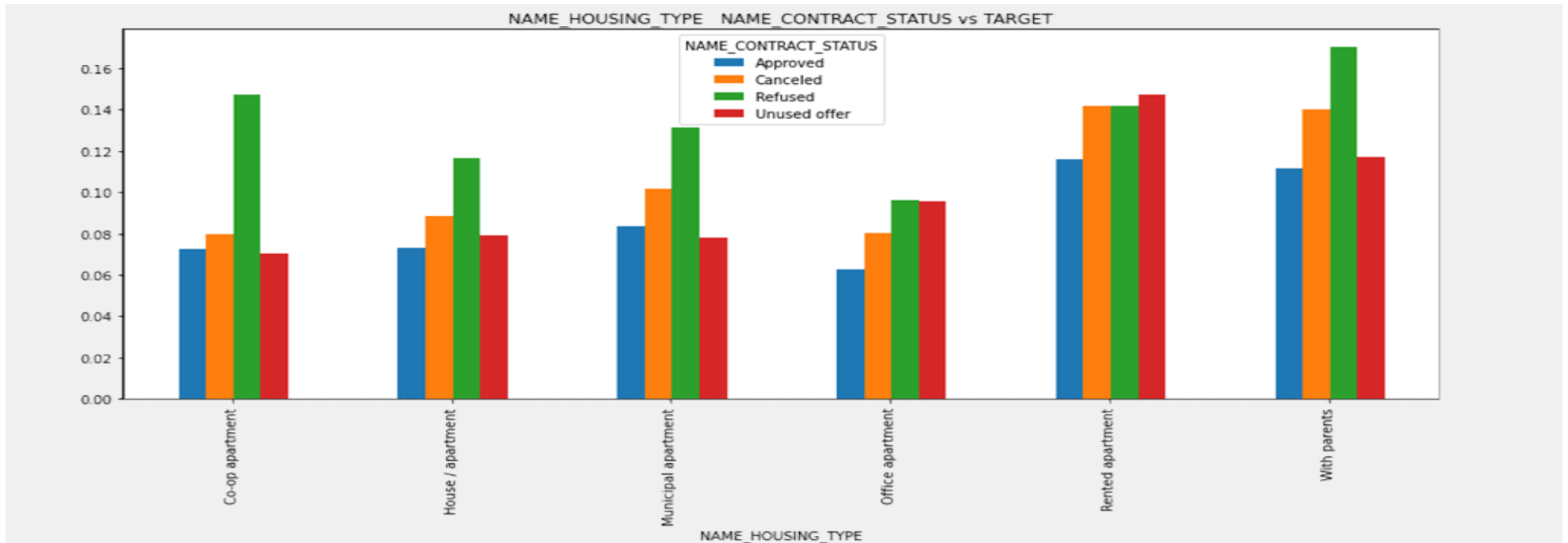
- Applicants with co op apartment has highest rejections ratio
- Applicants staying in rented apartments irrespective of previous application status, they have high defaulters
- Applicants with housing type with parents who had previous refusal has high defaulters
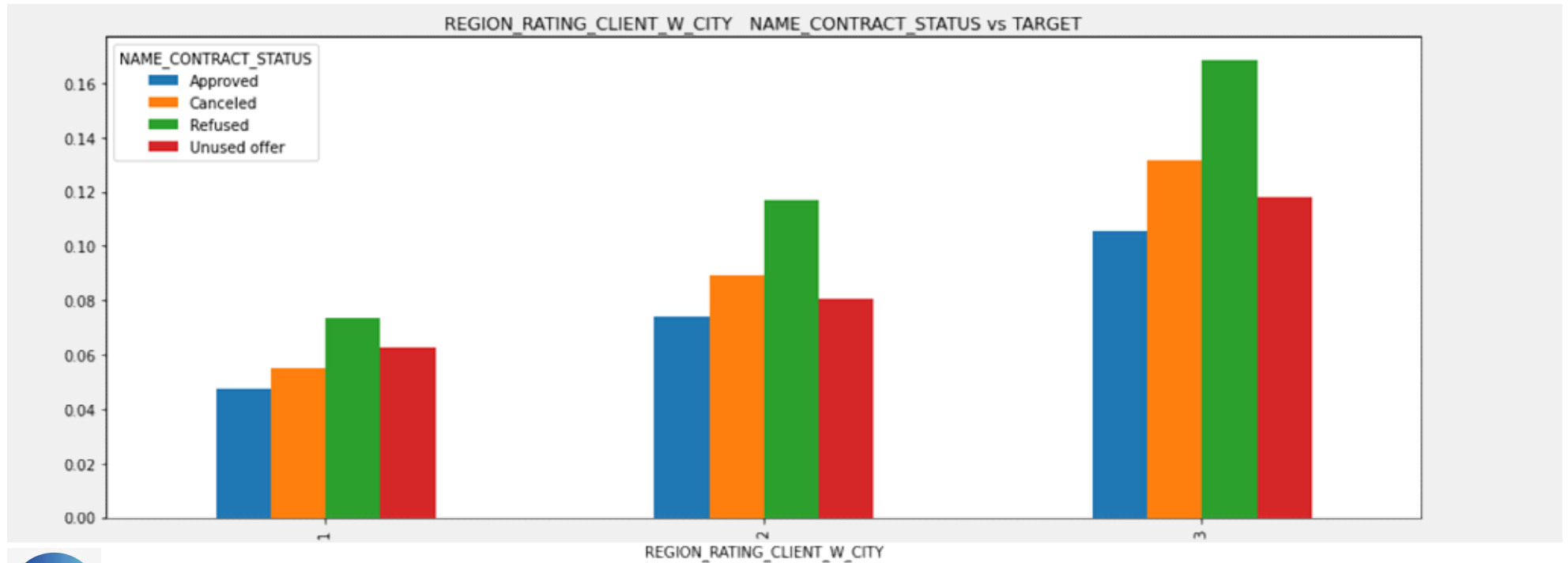
- Applicants staying in region city rating as "3" have highest refusals and also have high defaulters
- Applicants staying in region city rating as "1" have Lowest refusals and also have low defaulters

NAME_PORTFOLIO   NAME_CONTRACT_STATUS vs TARGET
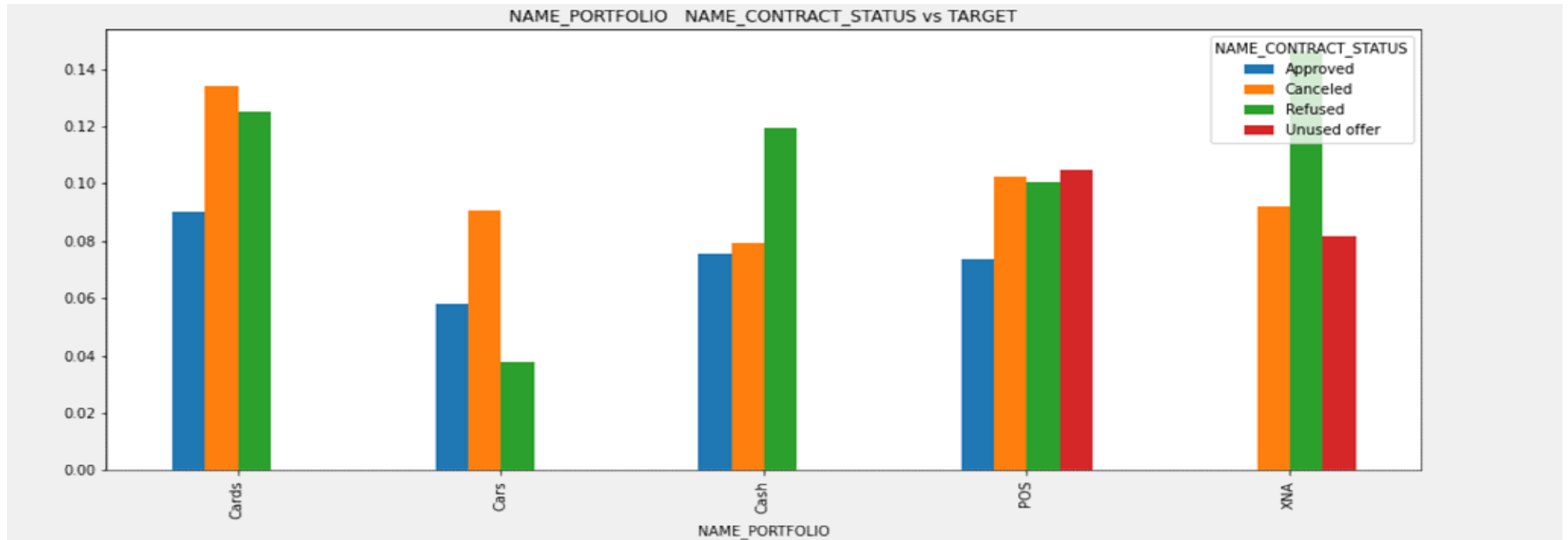
- Cards portfolio has high cancelled/refused applications previously and same are high defaulters
- XNA portfolio has high refused application previously and same are high defaulters
- POS portfolio applications have defaulters and not dependent on previous contract status
- Cash portfolio has high refusal rate

Contract type/Income type/Target
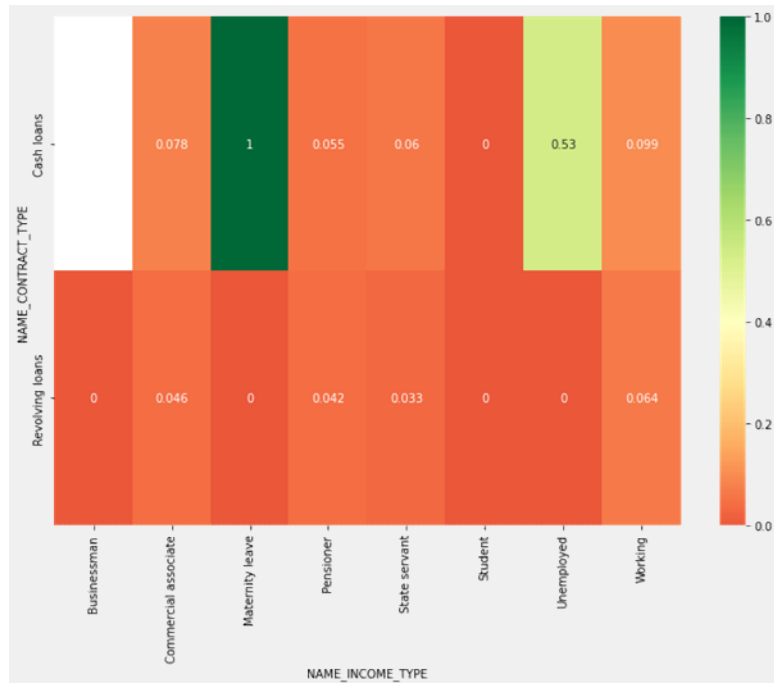
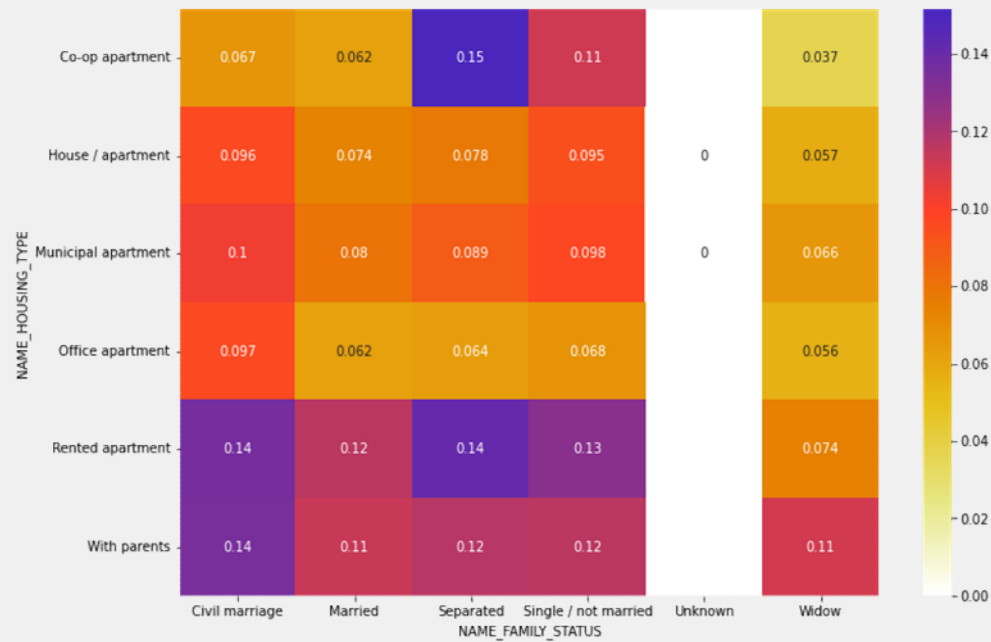Contract type/Education type/Target

- Applicants with Maternity leave and Cash loans have high default rate
- Cash loans for Unemployed also has high defaulters rate
- Lower secondary applicants with Cash loans have high default rate

- Applicants with civil marriage and staying with Parents have high defaulter rate
- Separated applicants and staying in Co-Op Apartment have high defaulter rate
- widow applicants staying in office apartment have low defaulter rate

High level conclusions

❖ Lower education applicants have high defaulter ratio
❖ Applicants with civil marriage and staying with Parents have high defaulter rate
❖ Separated applicants and staying in Co-Op Apartment have high defaulter rate
❖ Widow applicants staying in office apartment have low defaulter rate
❖ Applicants with Maternity leave and Cash loans have high default rate
❖ Cash loans in general have more defaulter rate than others
❖ Cards and Cash portfolio have higher refusals
❖ Applicants staying in rented apartments irrespective of previous application status, they have high defaulters
❖ Many of the IT staff did not used the offer earlier has high defaulters
❖ We can infer that low-skill laborers tend to default more and accountants are least defaulters

THANK YOU