CREDIT EDA CASE STUDY PRESENTATION

Harshavardhana AS
Abhishek Kumar
Trivikram Chausalkar

AGENDA

- The credit analysis is a crucial task for the bank which ensures lending the loans to the appropriate people thereby reducing the loss due to loan default thereby increasing the business profit of the bank

- The agenda of this case study primarily focuses on various parameters of applicants which drive the condition for loan default.

- Through this case study we suggest the possible factors which results in loan default , so that the bank can refrain from sanctioning loans to those specified categories. It is also equally important that the bank needs to approve the loans of the non-defaulting people so that it might not result in their business loss

PROCEDURE FOLLOWED

1. Handled outliers and removed columns which contribute to more than 40% of null values in application_data.csv file.

2. The appropriate columns with negative values was converted to positive values in application_data.csv using abs() function.

3. Merged both the files ,application_data.csv and previous_application.csv based on the column ‘SK_ID_CURR’ using ‘inner’ join.

4. Performed Univariate , Bivariate and Multivariate analysis on the merged files.

5. Identified Top-10 correlation features of defaulters in both application_data.csv and previous_application.csv files.

6. Performed Z-Test and Chi-Square statistical tests to prove the statistical significance of the claims made.

INSPECTION OF DATA

1. The application_data.csv data sheet contains 3,07,511 rows and 122 columns

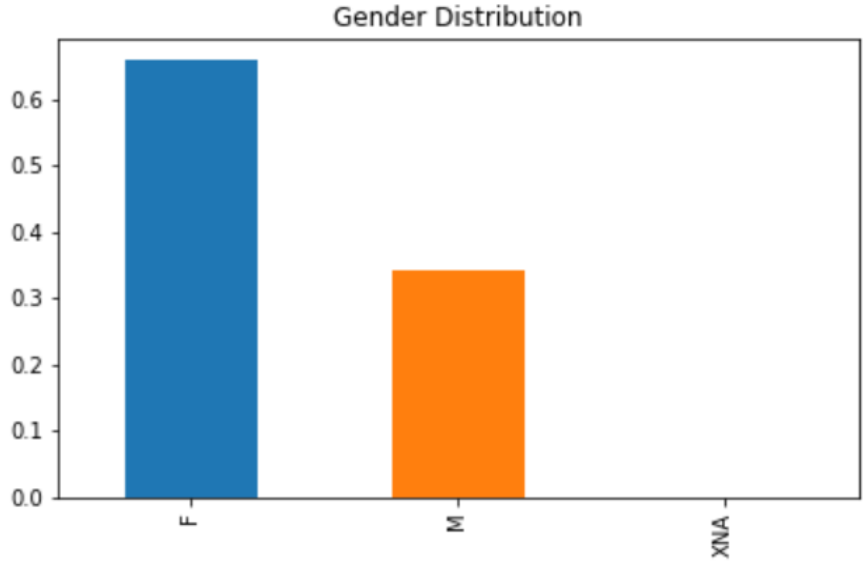
2. Each column in the application_data.csv refers to applicant's details like education, marital status, income, credit loan provided etc.

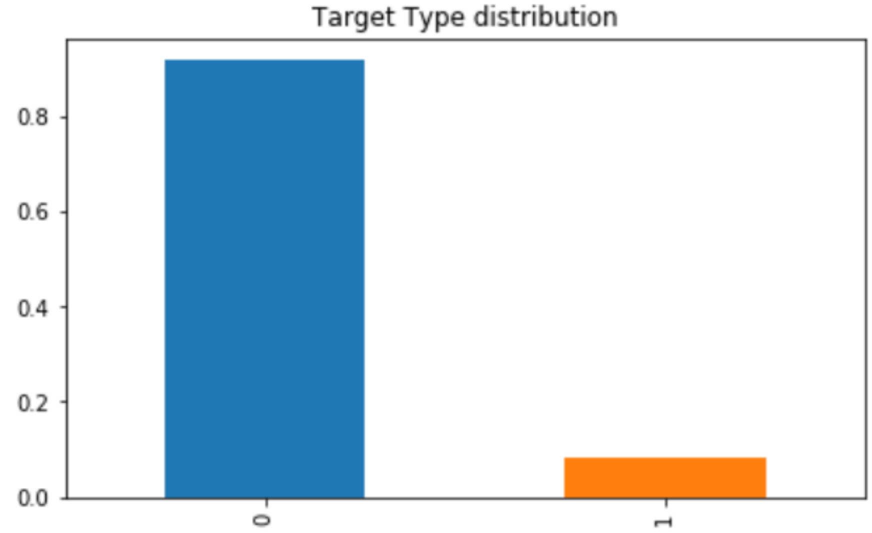
3. The 'TARGET' column in application_data.csv refers to the loan payment status of the clients . The value 0 refers to the client who has difficulty in re-paying the loan and value 1 refers to the client has not-defaulted .It is the column of importance in application_data.csv .

4. The previous_application.csv contains the details of previous status of loan sanctioned to the clients , this data sheet has over 16,70,214 rows and 37 columns.

5. The column 'NAME_CONTRACT_STATUS' in the previous_application.csv contains the details of loan contract status like if the previous loan was approved cancelled, refused or unused at the past for the clients . This is the column of importance in previous_application.csv .

GENDER AND TARGET VARIABLE PERCENTAGE

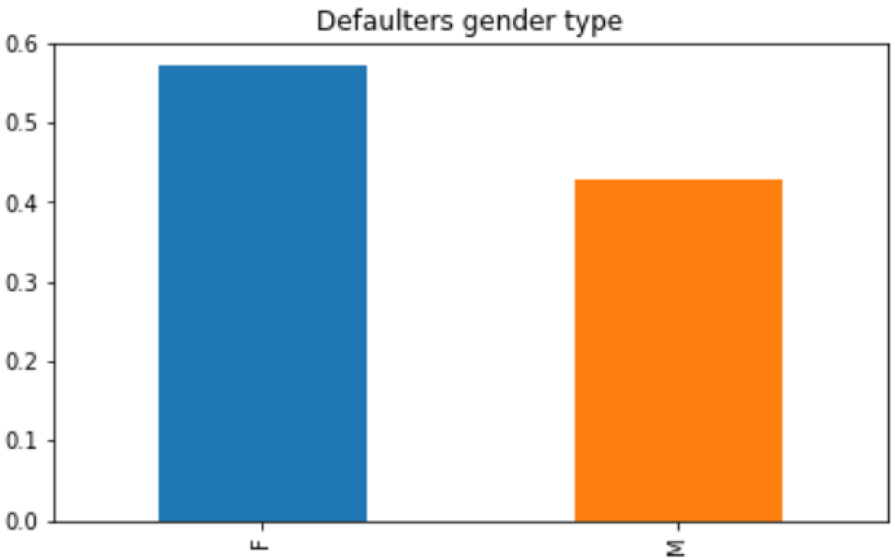


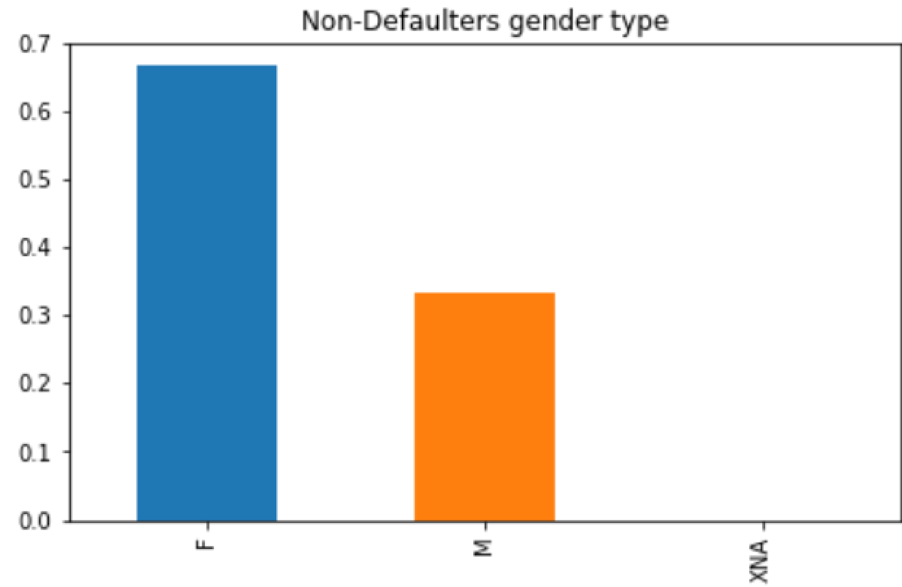


- The above graph clearly indicates that females apply more loan than males.
- In the given data over 65% contribute to females against 35% of males
- From the TARGET distribution over 85% of the people have no difficulty in re-paying the loan

UNIVARIATE ANALYSIS ON PRESENT APPLICATION STATUS

EFFECT OF GENDER

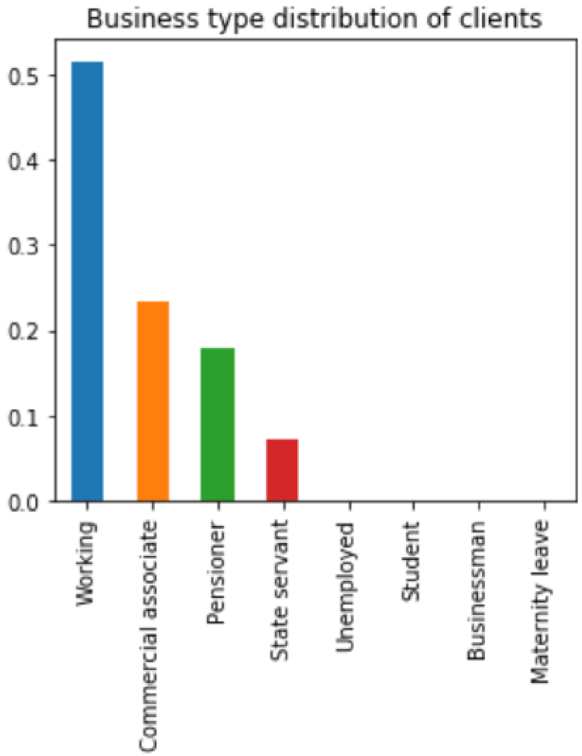


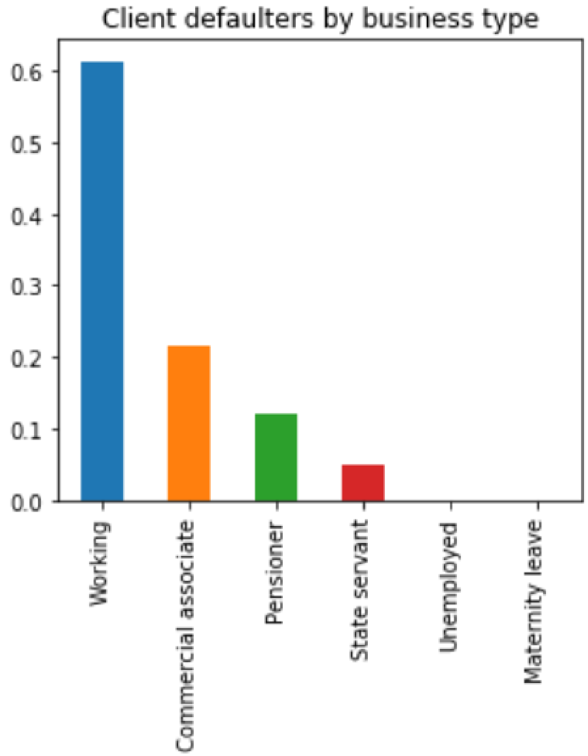


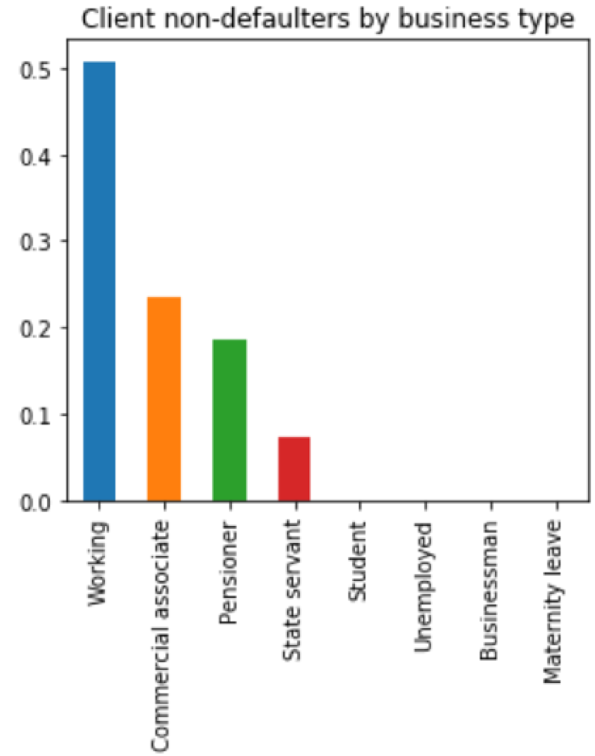
- In the given data , male tends to default more as the percentage of defaulters in male is slightly on the higher side than defaulters.

- In the given data, females have slightly greater proportion in the non-default case when compared to the defaulters.

EFFECT OF BUSINESS TYPES



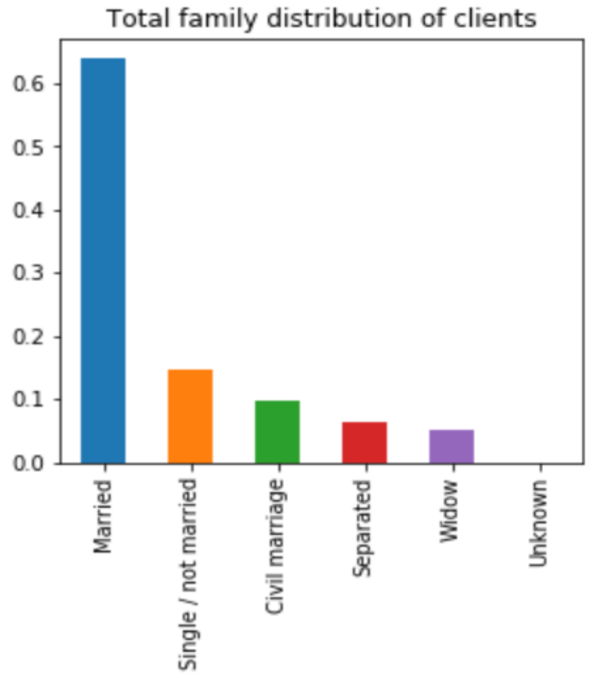


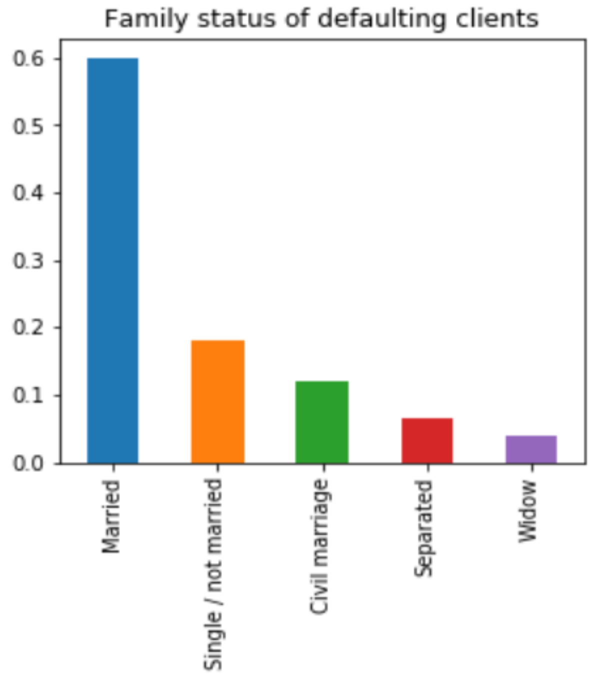


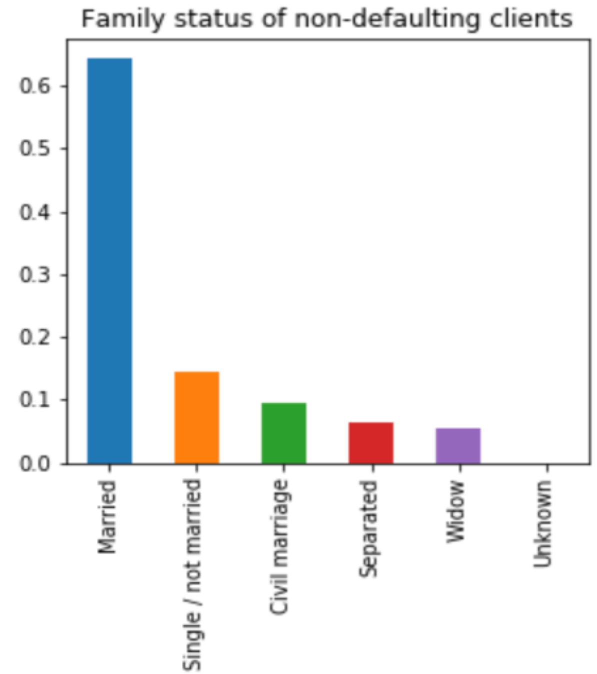
1. The population of Pensioner and State Servant are less likely to default.

2. The maximum people who avail loans falls in the category of ‘Working’ and from the given data, they tend to likely default compared to others.

EFFECT OF FAMILY TYPES



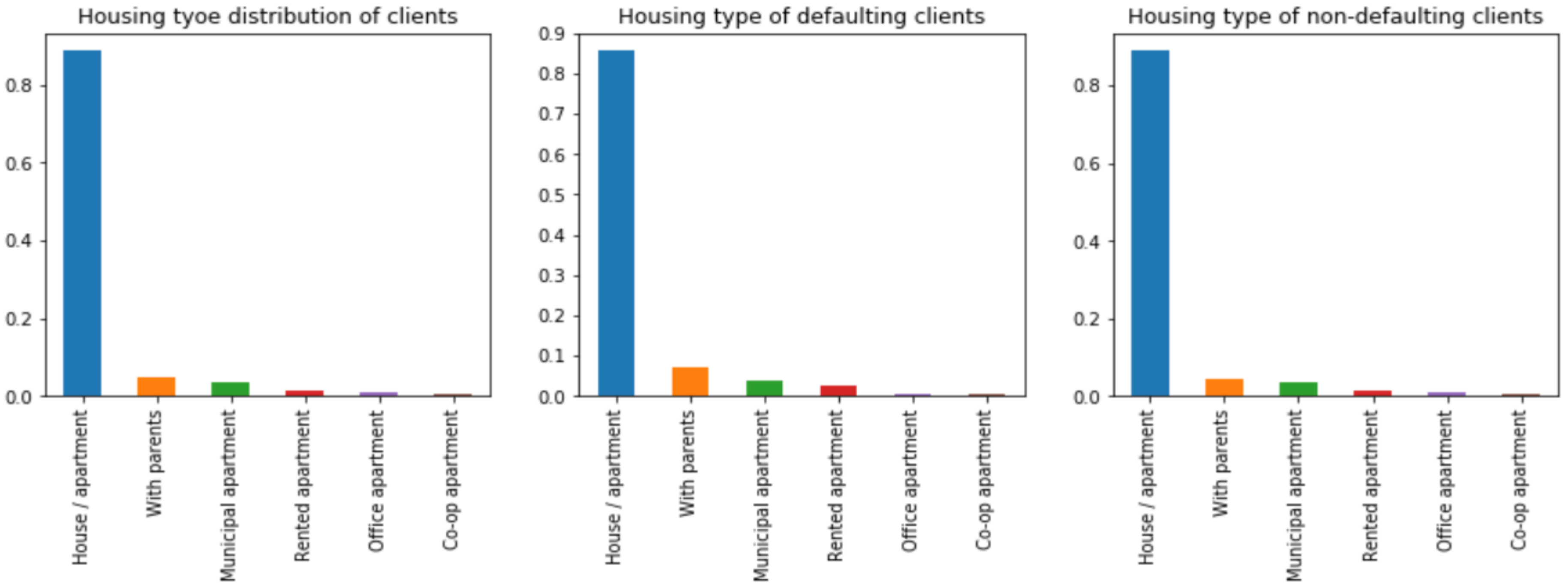




1. Based on the given data, married people tend to fall slightly on the non-default side and same for widow.

2. On the other hand for the given proportion 'civil married' people and 'single not married' tend to default

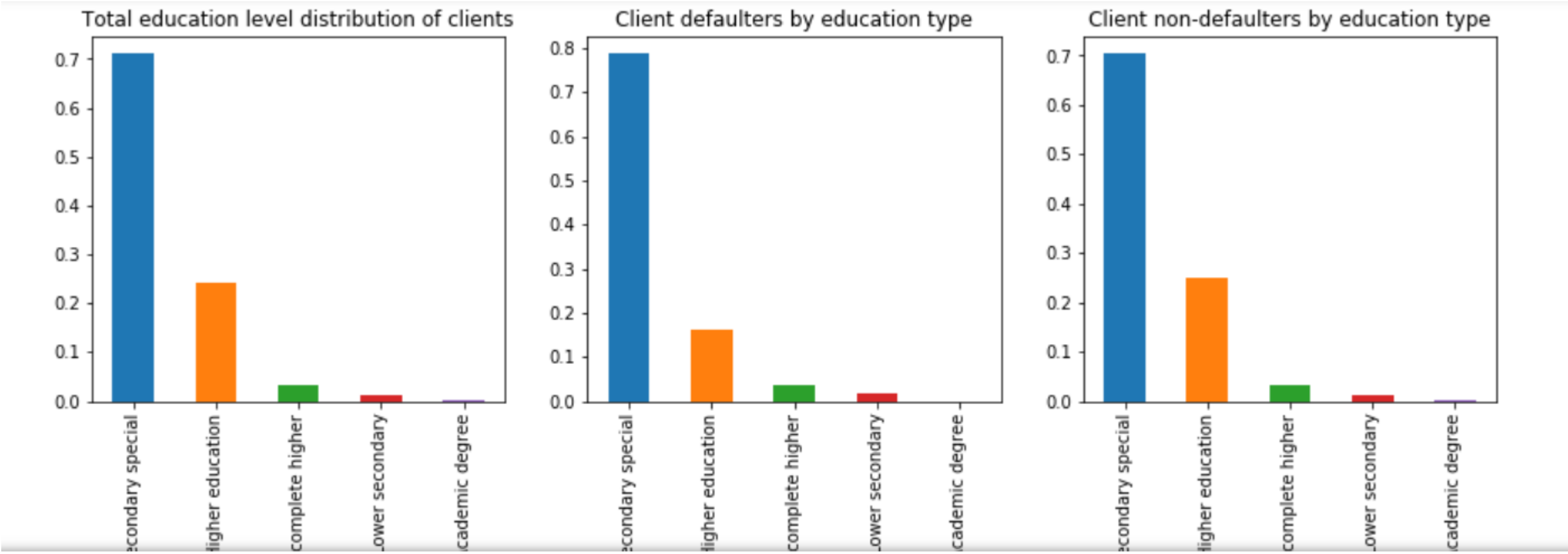
EFFECT OF HOUSING TYPES



1. Most of the loans are concentrated towards the people who live in apartments and the defaulting case is more however the non-default case is also equally dominating.

2. People who live with parents tend to default less (may be the data is less distributed across this feature)

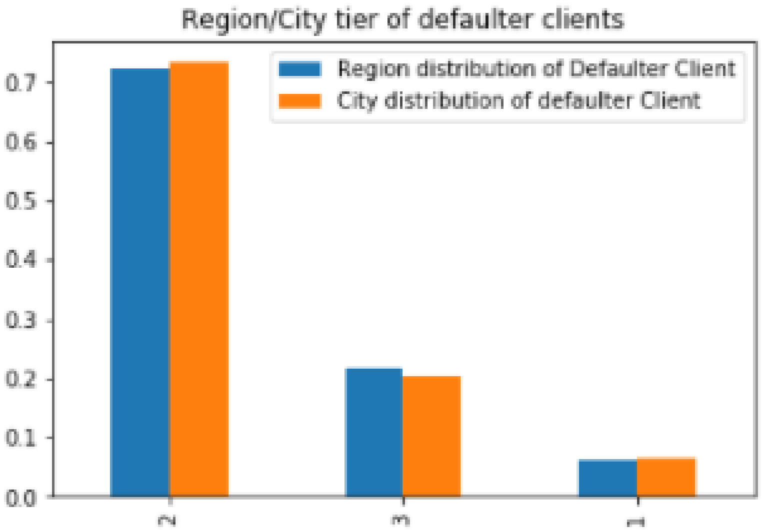
LEVEL OF EDUCATION OF CLIENTS

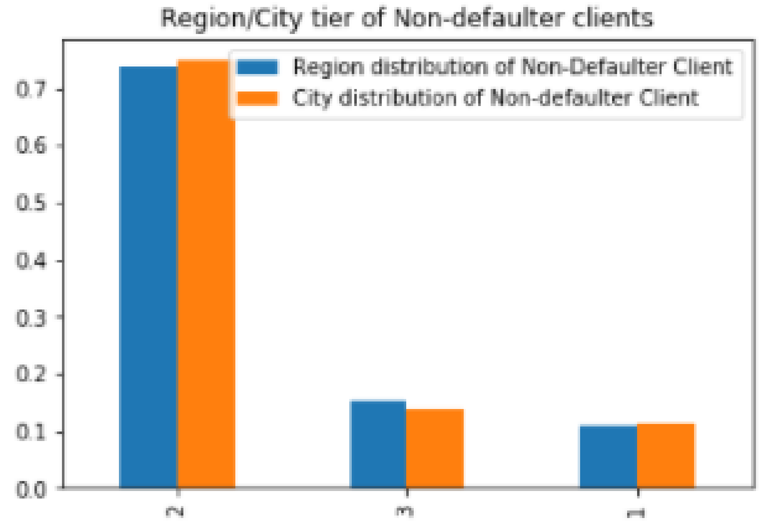


1. In the given proportion people whose education is of secondary special has a slight default rate of 79% against the non-default rate of 69%.

2. Higher educated Clients are less likely to default based on the above bar graph.

BASED ON TIER OF CITY



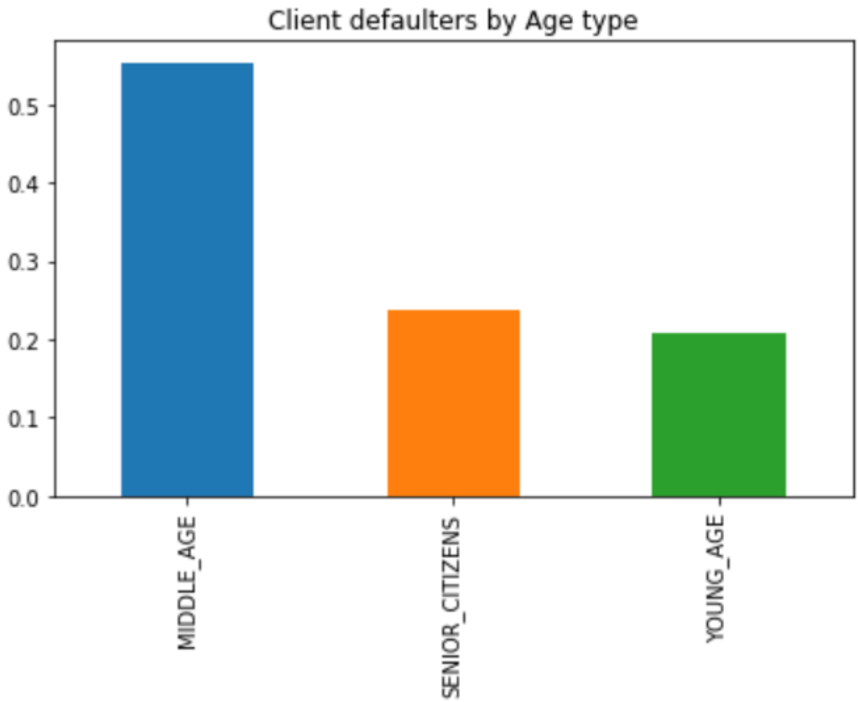


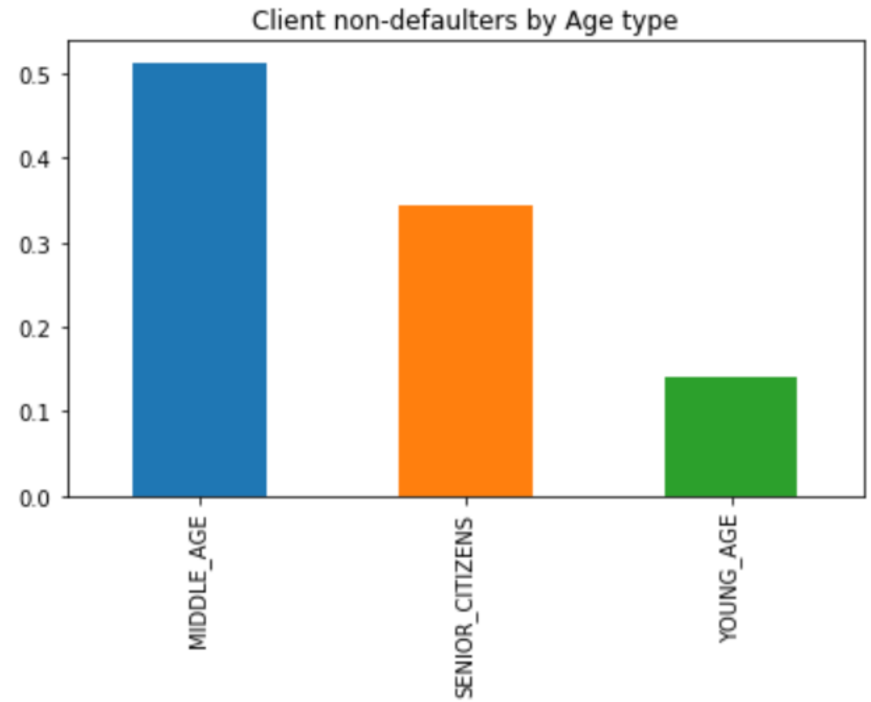
1.Loans are targeted/given to people who come from tier-2 city/region and their default case is higher.

2.We can also see that in tier-3 the proportion of people who default is slightly higher than the people who do not default (for non-default it is roughly 10% and default it is roughly 18% for tier-3)

3.From the given data, people from Tier-1 city are less likely to default as there is a rough 10% difference between defaulters and non-defaulters.

EFFECT OF AGE

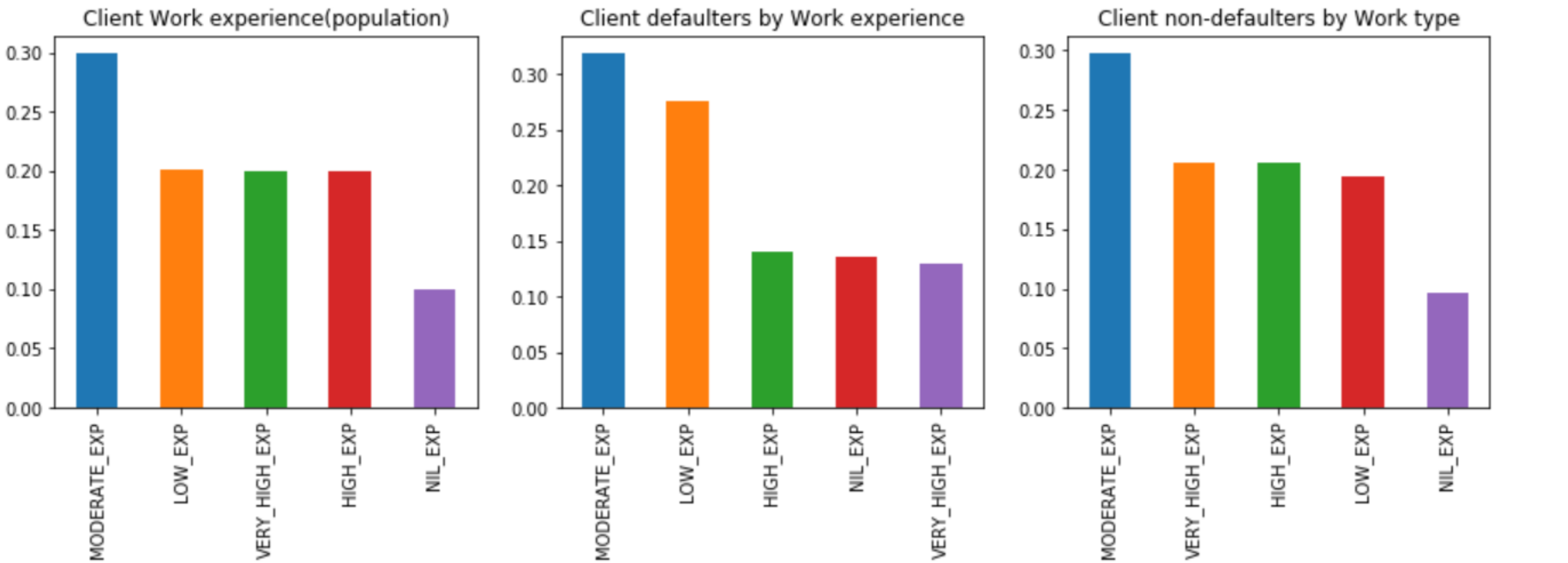




2. The young aged people (Age:20-30) are likely to default based on the above observation.

3. The middle-aged people(Age:30-50) shows no difference as the ratio of defaulters and non-defaulters remain the same.

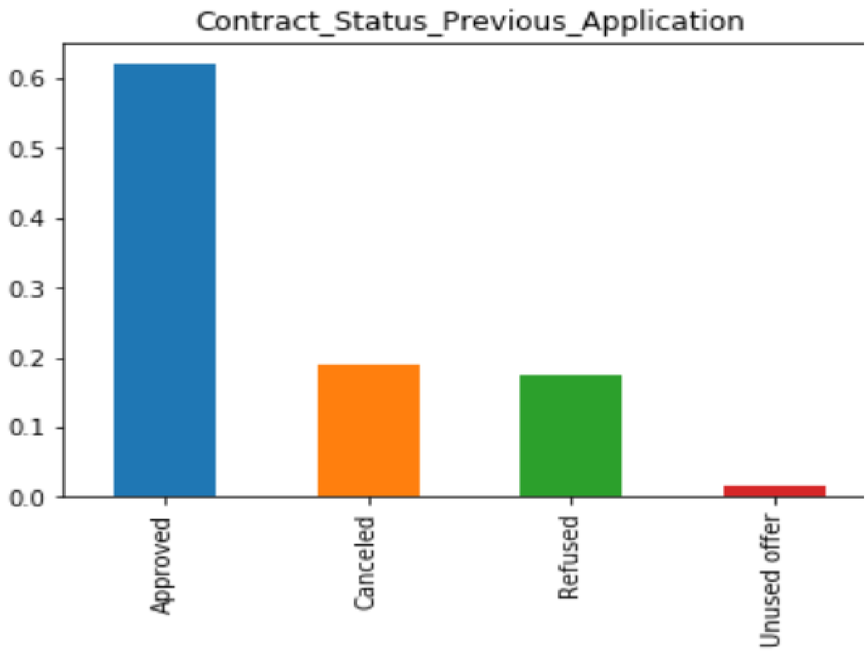
EFFECT OF WORK EXPERIENCE

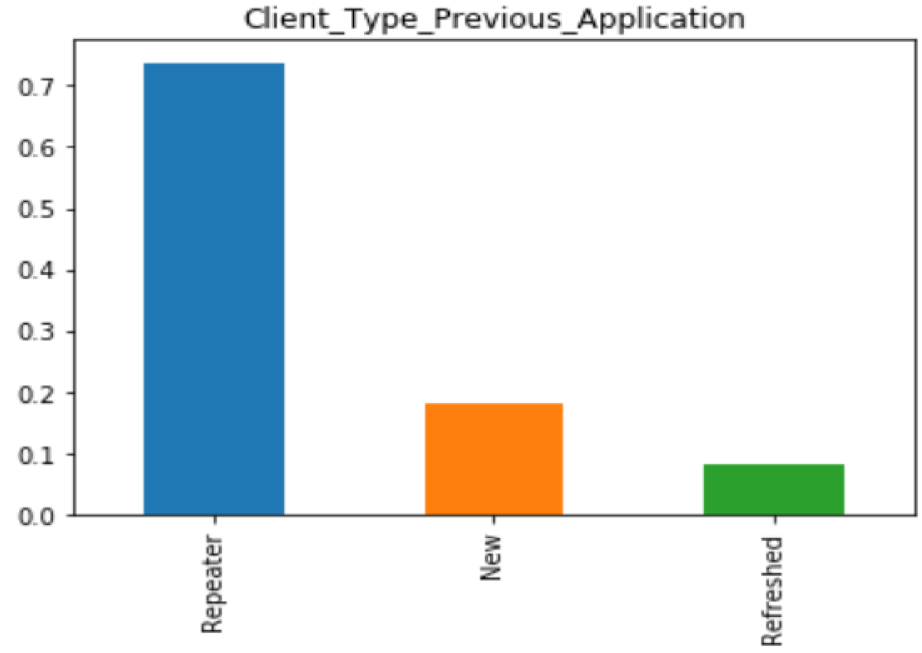


1. From the observation, people with Nil Experience (<1 year experience) and with low experience (between 1- and 3-years experience) has a higher default rates compared to others.

2. People with High experience (between 8- and 25-years experience) and Very High experience (>25 years) are on the less-default side.

ANALYSIS ON CONTRACT STATUS





1.More than 60% of loan has been approved in the previous application followed by cancelled and refused.

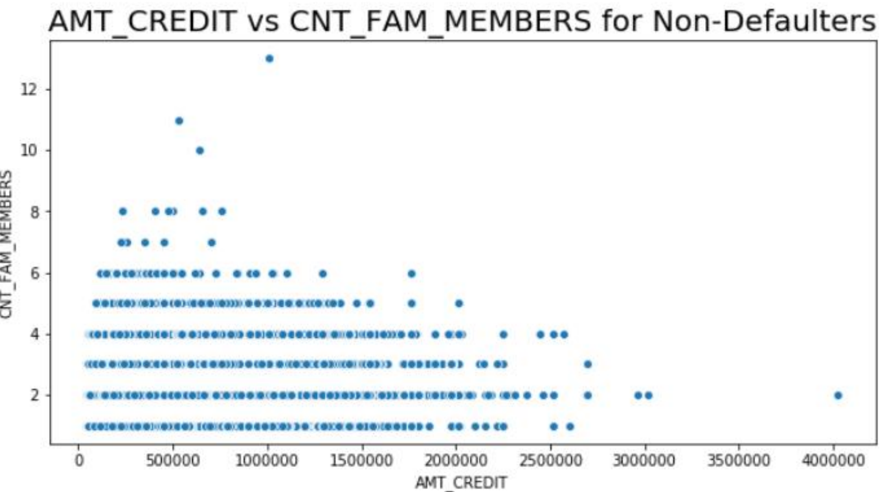
2.Over more than 70% of people are repeaters while availing the loan followed by New people.

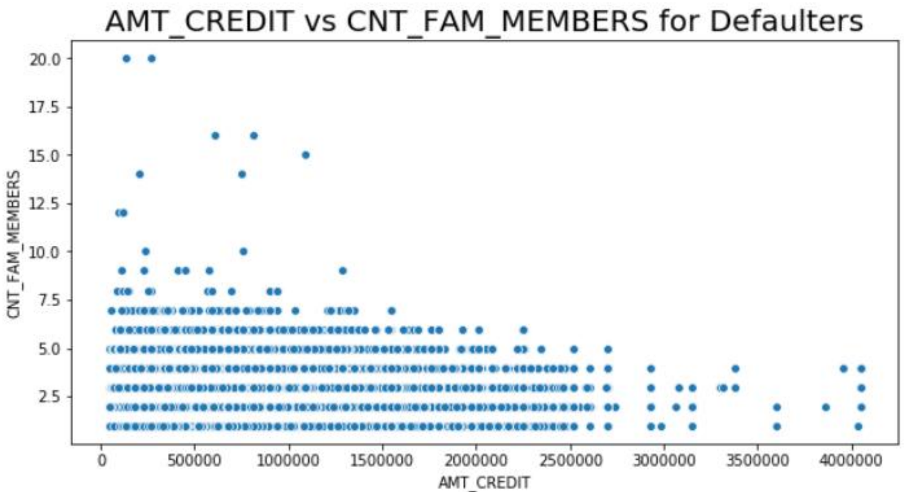
TOP-TEN CORRELATION FEATURES OF DEFAULTERS

	Column 1	Column 2	Correlation
1982	OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	0.998269
370	AMT_GOODS_PRICE	AMT_CREDIT	0.982783
1239	REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.956637
1100	CNT_FAM_MEMBERS	CNT_CHILDREN	0.885484
2044	DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.868994
1487	LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.847885
1673	LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.778540
371	AMT_GOODS_PRICE	AMT_ANNUITY	0.752295
309	AMT_ANNUITY	AMT_CREDIT	0.752195
2388	FLAG_DOCUMENT_6	DAYS_EMPLOYED	0.617646

BIVARIATE ANALYSIS ON MERGED APPLICATION STATUS

ANALYSIS BASED ON CREDIT AND COUNT OF FAMILY

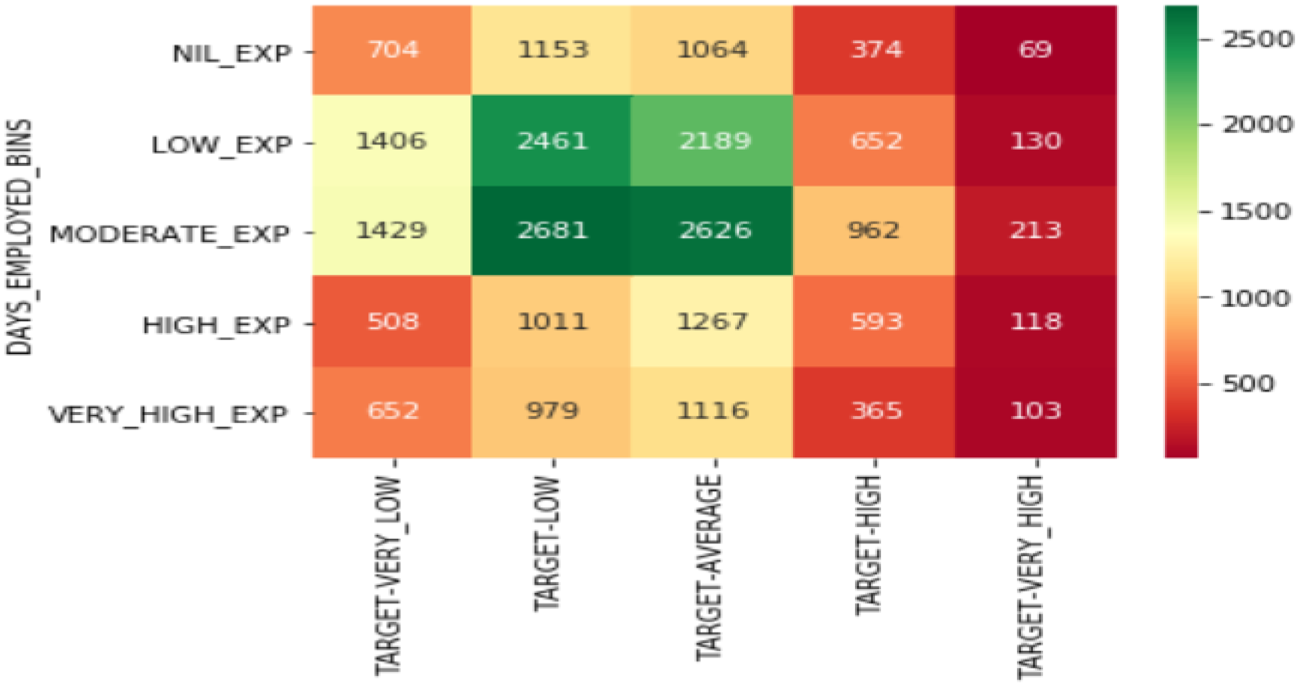




1. The family count with less than 7 tends to default more on the lower CREDIT value sanctioned to the clients.

2. Clients with the family count >7 has proportionally less default cases as the loan payment might be supported by their family.

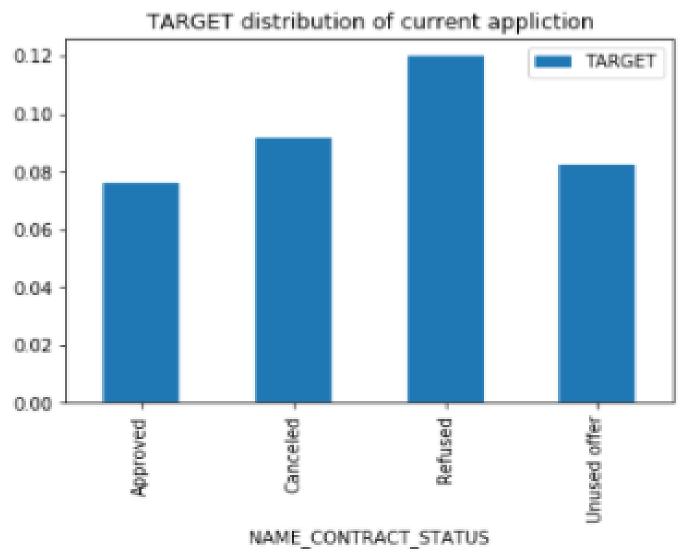
ANALYSING USING HEAT MAP BASED ON EXPERIENCE



1.If we observe from the heatmap it is evident that people who are having moderate or low experience in work tend to default more provided with the credit loan amount.

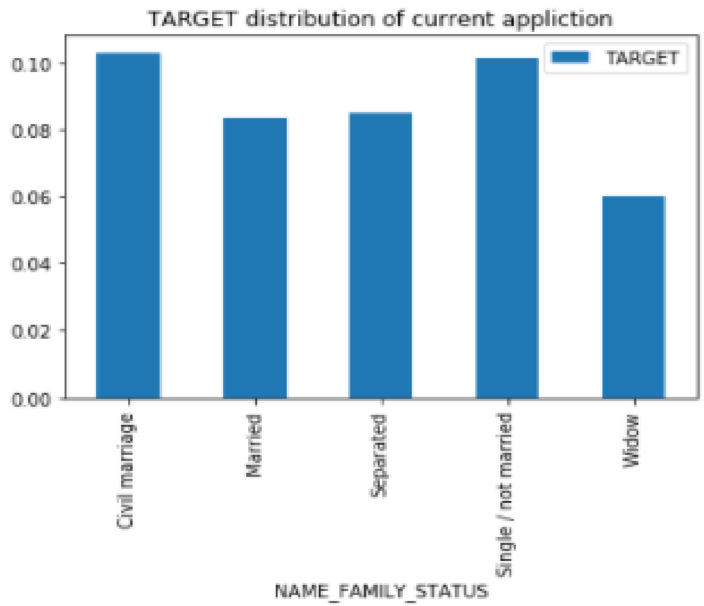
2.However, people with high and very high experience do not default much for the credit amount they are allotted with.

ANALYSIS ON THE MERGED DATA FOR DEFAULTERS



Applications those were rejected previously tend to result in higher percent of default in current application

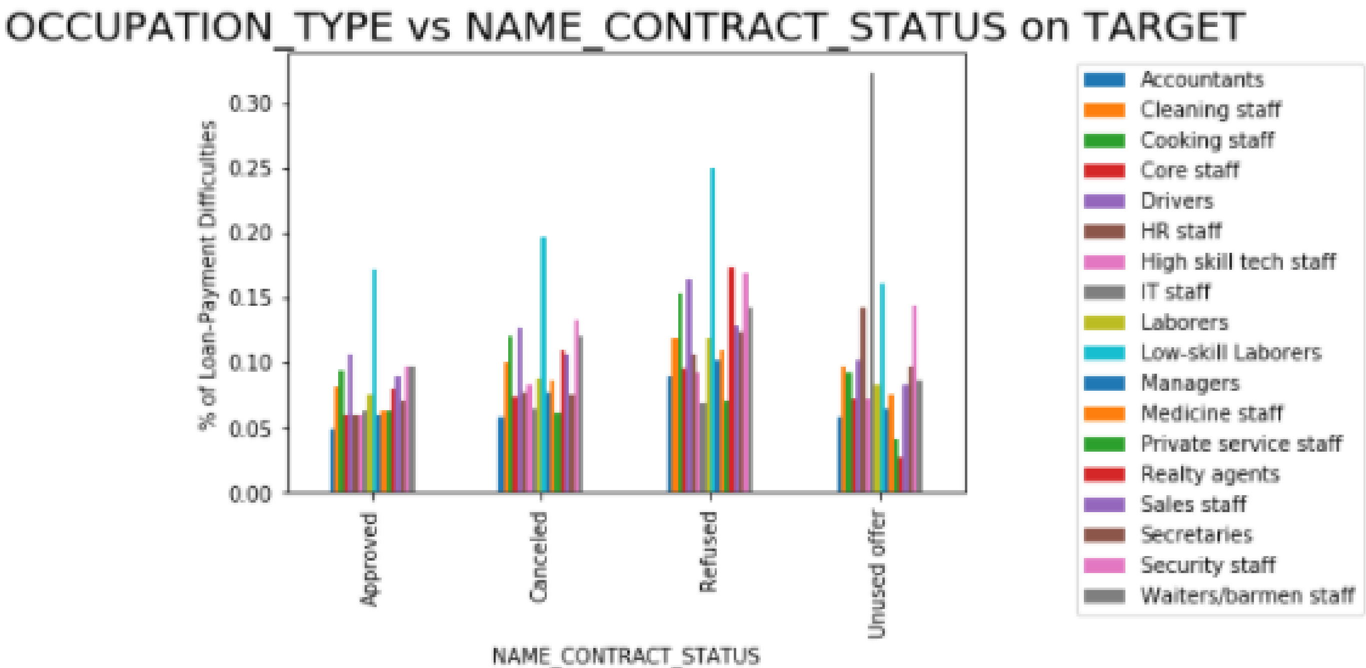
ANALYSIS BASED ON RELATIONSHIP STATUS



People who are civil married and single/not married tend to default more than others

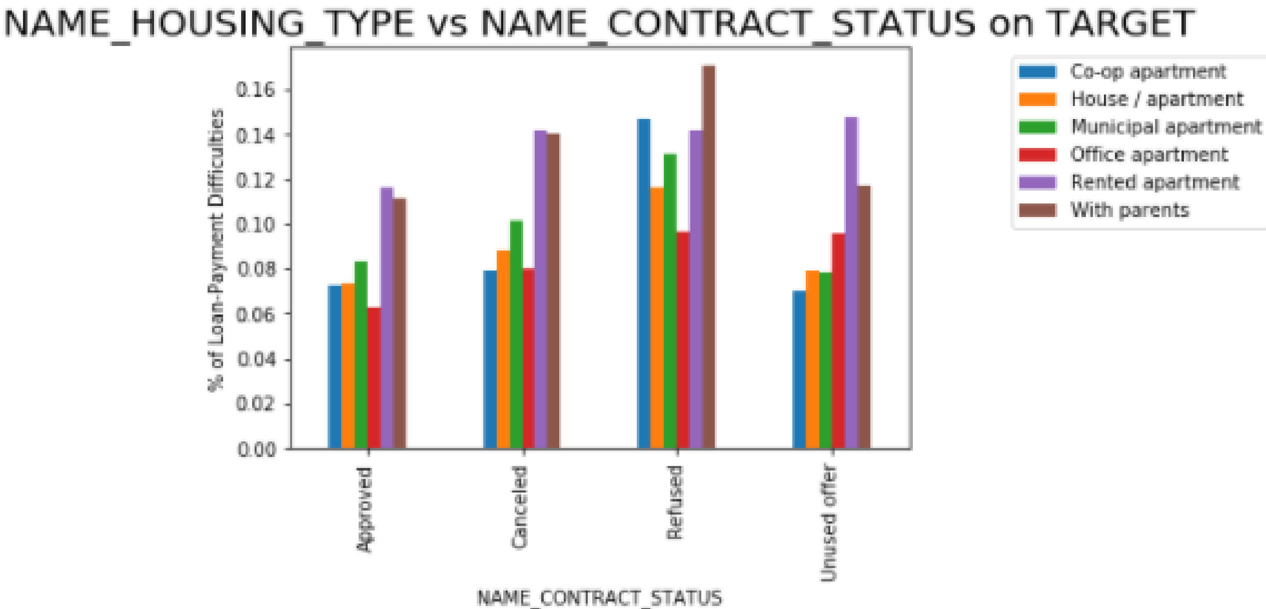
MULTIVARIATE ANALYSIS OF MERGED DATA

OCCUPATION_TYPE VS NAME_CONTRACT_STATUS



1. The people with Low-Skilled Laborers and Laborers whose previous loan application was refused/cancelled tend to default more in the present status

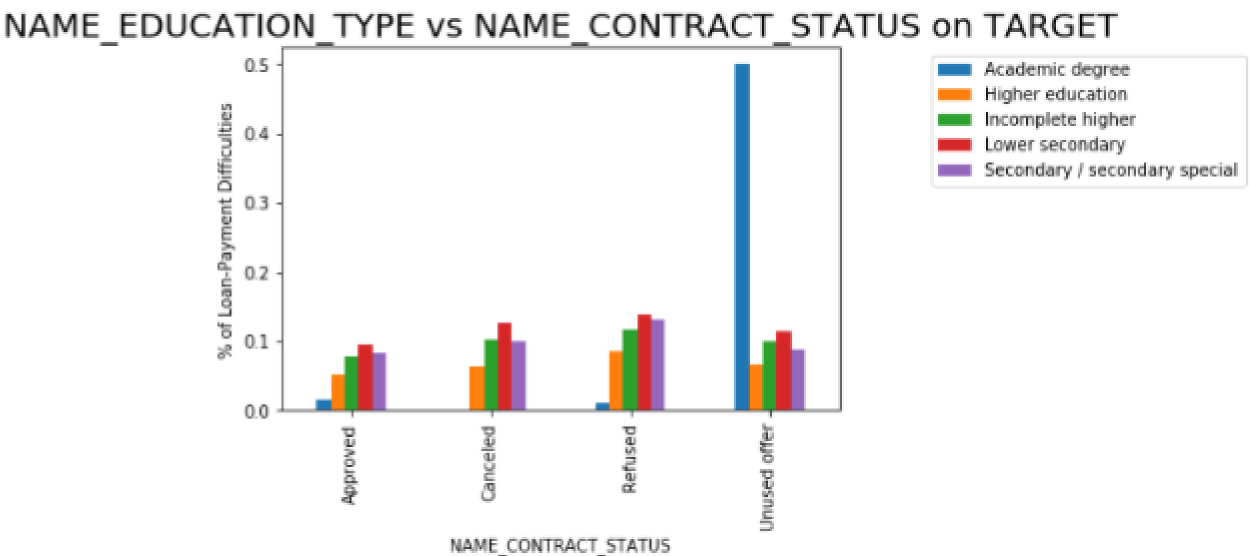
NAME_HOUSING_TYPE VS NAME_CONTRACT_STATUS



1. People whose previous application is ‘Unused’ and staying in ‘Rented Apartments’ has a higher default rate

2. People whose previous application is ‘Refused’ and ‘Cancelled’ and staying with parents has a comparatively higher default rates than other categories.

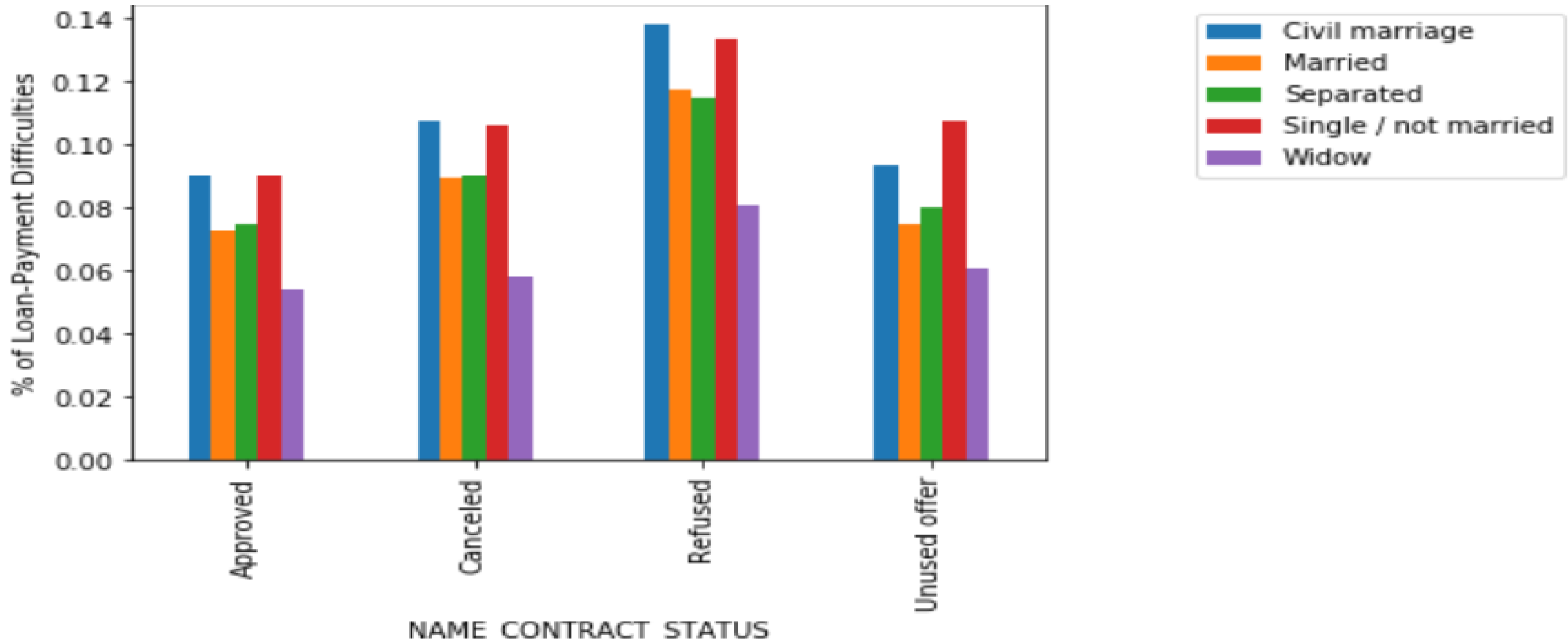
NAME_EDUCATION_TYPE VS NAME_CONTRACT_STATUS



1. People with Academic degree and whose previous application is unused tend to default more.

2. For the people with lower secondary/incomplete higher education whose previous loans are cancelled/refused default more than the higher education people in all aspects

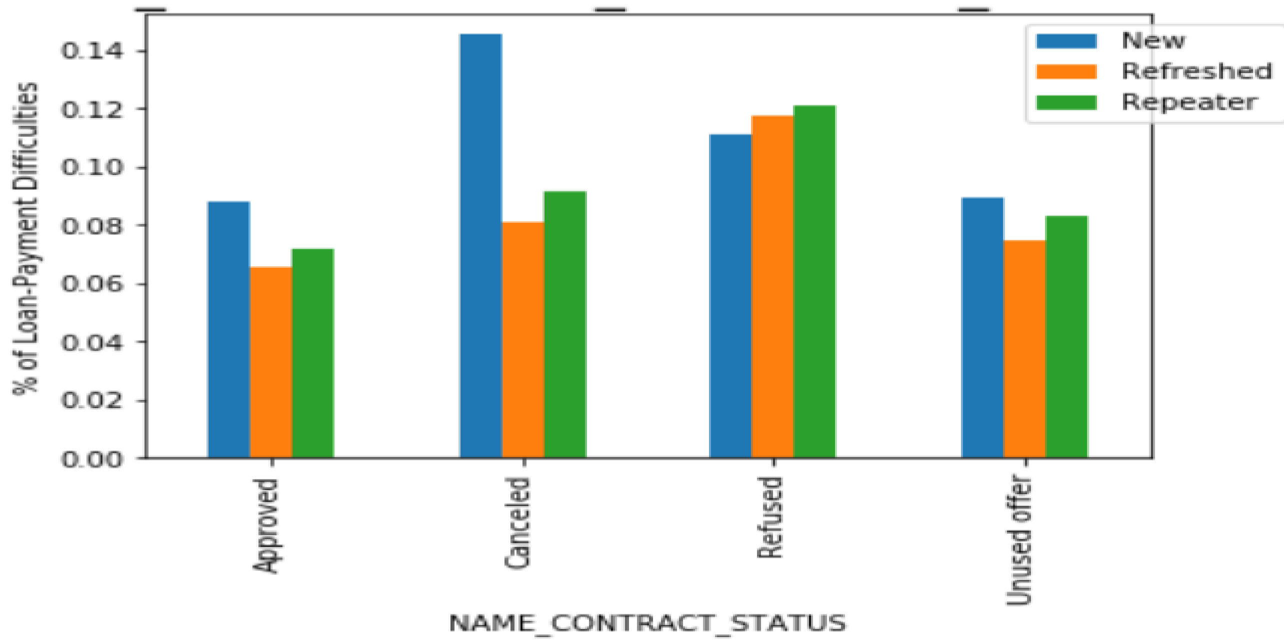
NAME_FAMILY_STATUS VS NAME_CONTRACT_STATUS



1. Civil married people whose application is rejected in the previous application has higher default rate than others

2. Widow people has overall less default rate in comparison to other type of people

NAME_CLIENT_TYPE VS NAME_CONTRACT_STATUS



1. New type of people who tends to default more whose loan is not used or cancelled in the previous application.

2. The refreshed type of users have comparatively lesser default proportion than others

Z AND CHI-SQUARE STATISTICAL TESTS AND CONCLUSION

1.There is a gender dependency on the default cases of loan .

2.There is no statistical evidence to say that the person who owns a car and flat might default . However as we interpreted earlier person who has own flat has a difficulty in paying loan compared to the own car.

3.The loan type has no dependency on default according to the data we have . However we later investigated with the help of previous application data about the loan type cancellation/refusal nature which had on default cases.

4.There is a statistical evidence that there is a relation between city tier and default cases.

RECOMMENDATIONS FOR LOAN APPROVAL

1. Clients whose Business type is Pensioner and State Servant.
2. Clients who are married and widowed are less defaulted.
3. Clients who completed their Higher Level of education.
4. People from Tier-1 city are likely to be considered.
5. Senior Citizen Clients are better targets for loan approval.
6. Clients with High work experience .
7. Clients who opted for consumer loans.
8. Clients who were of refreshed type in previous application.
9. Clients who reside in office apartments are the better targets than other groups.

RISK GROUPS

1. Working people with nil and low experience.
2. Single/not married and civil married people.
3. People with incomplete and lower secondary education level.
4. Proportion of male clients whose previous application was rejected and cancelled.
5. Low-Skill Laborers whose previous application was rejected and cancelled.
6. New Clients whose previous application got cancelled.
7. Clients who live in rented apartments and living with parents whose previous application got refused/cancelled.

THANK YOU !