



CREDIT RISK ANALYSIS

EDA CASE STUDY

BY
TANISHQ SAMDARSHI

PURPOSE

- ❖ The case study aims to apply EDA techniques for a real business case
- ❖ This case study also aims to develop some understanding of risk analytics in banking and finance services
- ❖ To develop understanding on using data to minimize risk of loss while lending money to clients
- ❖ This case study will also push to gain some knowledge of banking and finance terms



PROBLEM STATEMENT

Understanding the variable which are useful for analysis and finding clients to lend money at minimum risk

When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

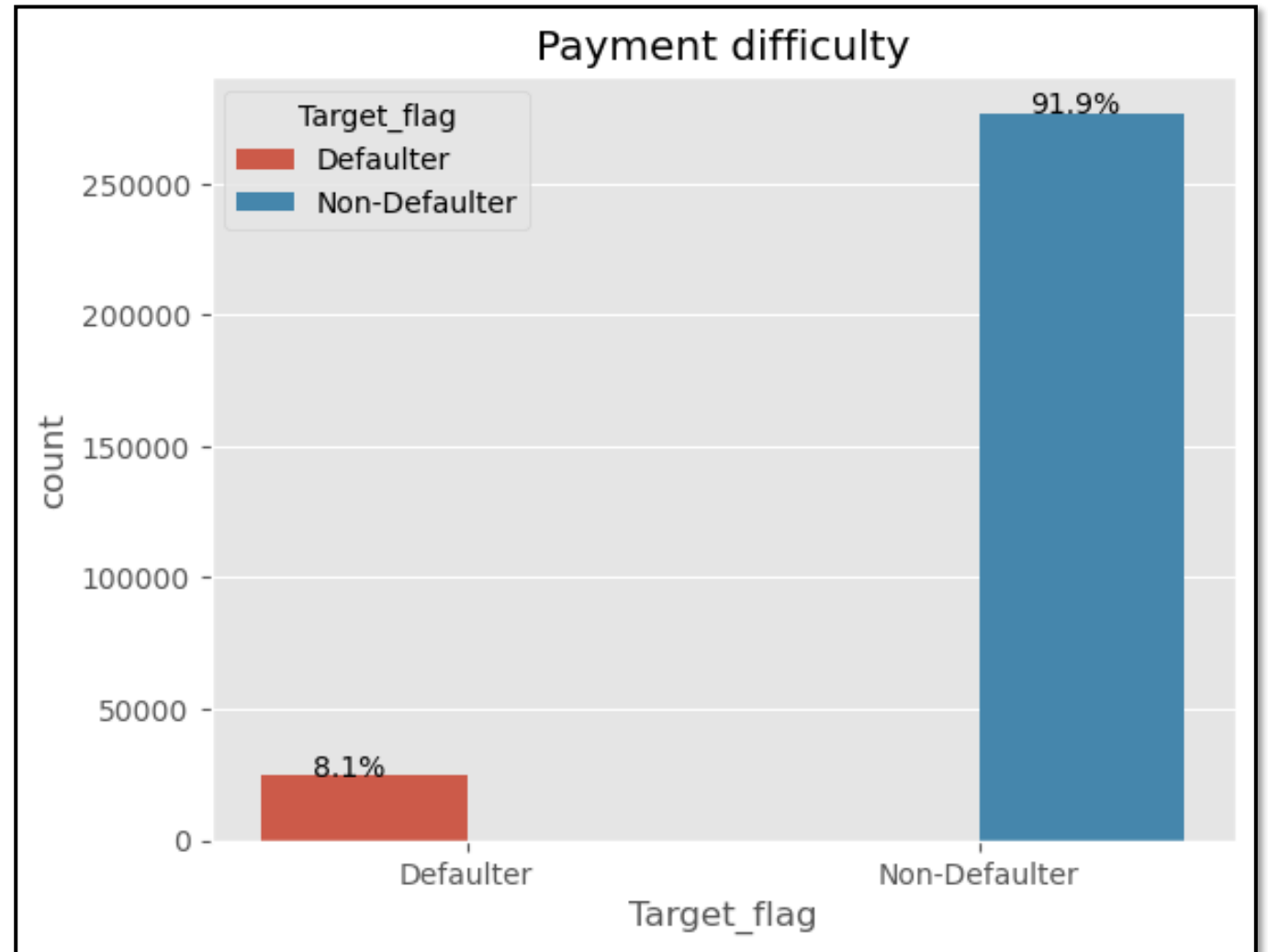
- ❖ If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- ❖ If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.



APPLICATION DATASET ANALYSIS

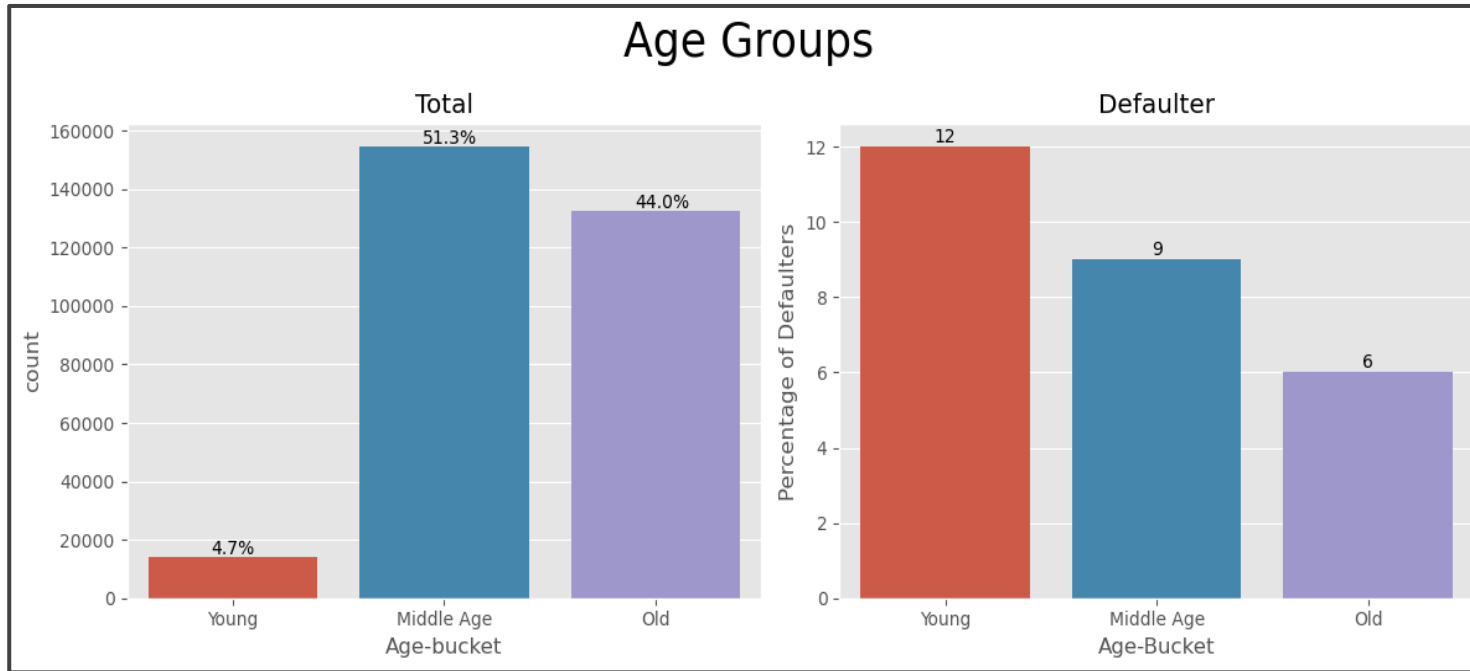
Data Imbalance in Application dataset

- ❑ The data is highly imbalanced as approx 92% are non-defaulters means they 92% clients repay their loans as compared to 8% who cannot repay their loan(defaulter).
- ❑ This trend is a good indicator for bank.
- ❑ It will help bank to minimize the risk



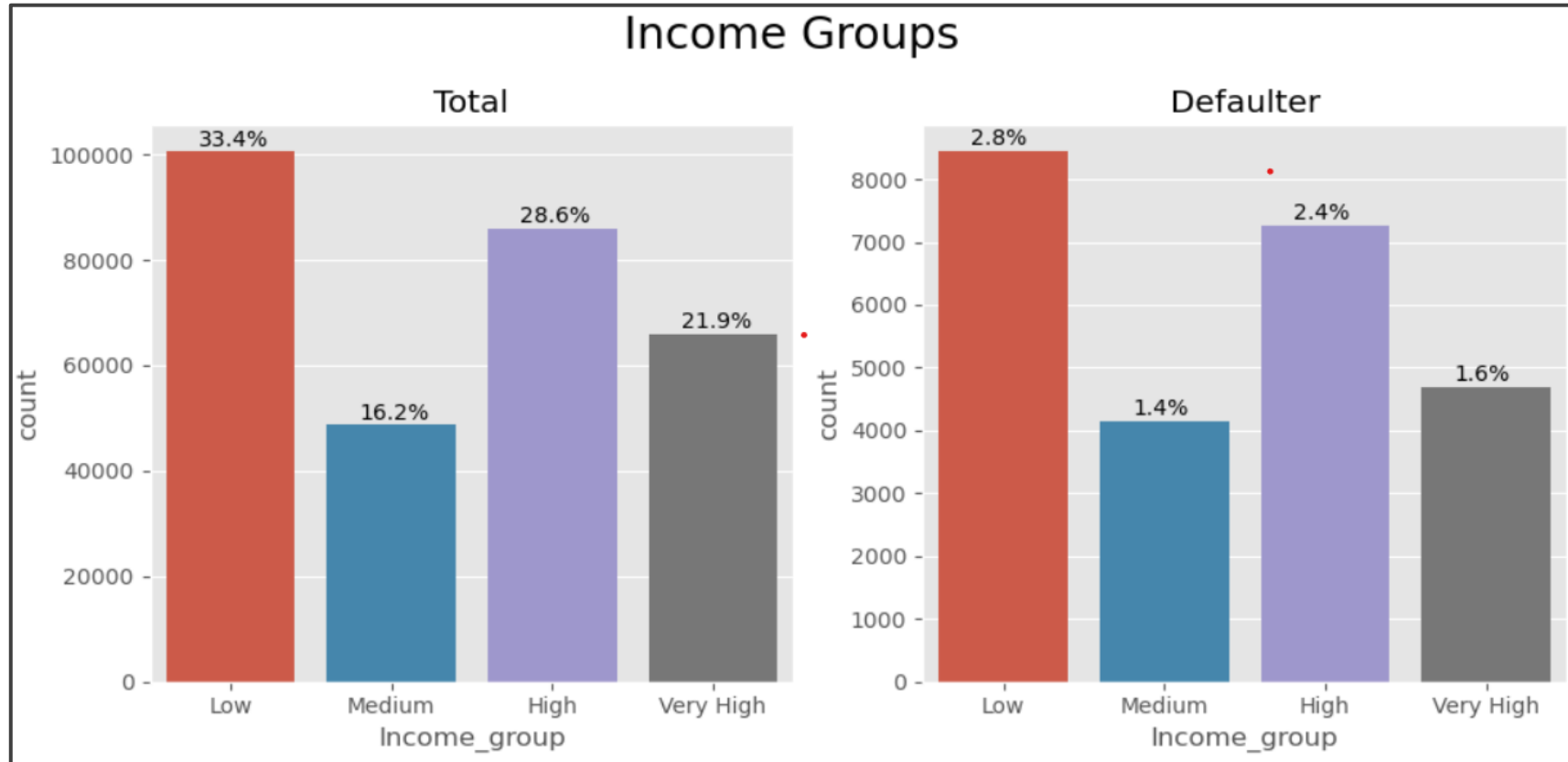
Univariate Analysis

1: AGE-WISE ANALYSIS



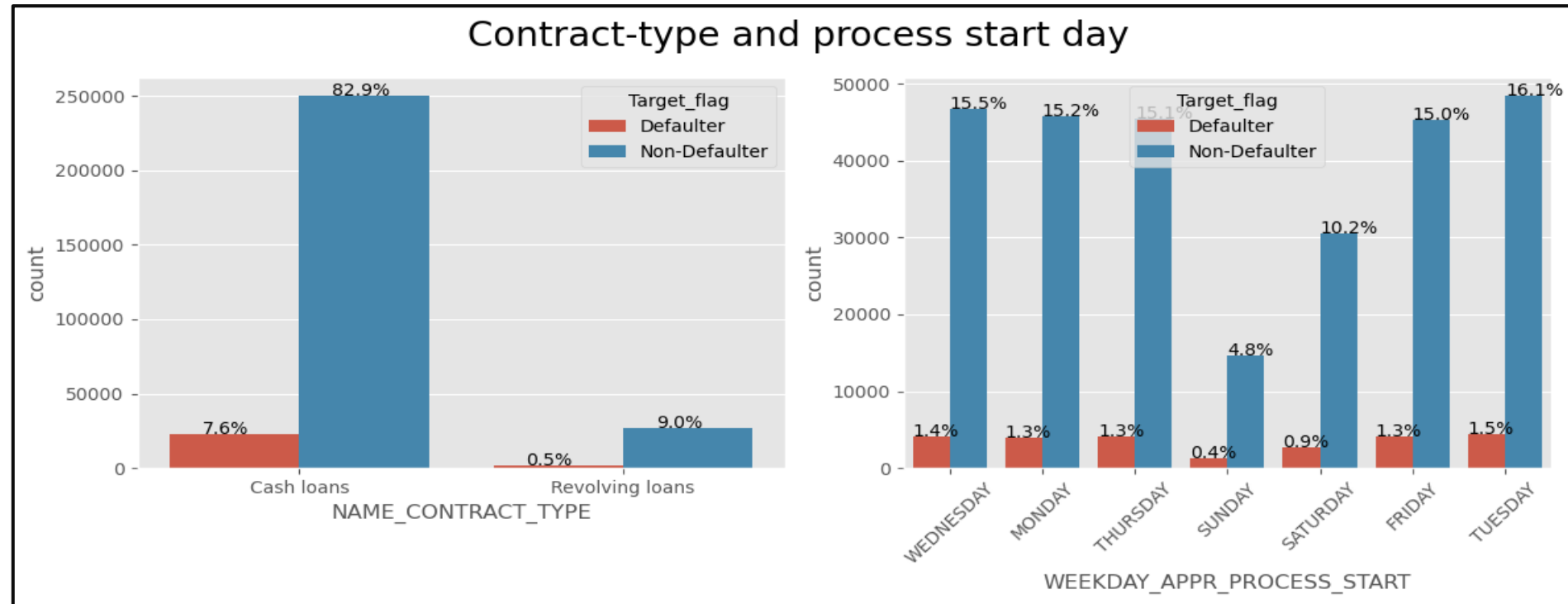
- ❑ 62.5% of the clients belong to Middle Age group , 21.5% to Old group followed by Young at 16.1%.
- ❑ Young people are most likely to be defaulter (~12%) and only 8% of Middle age client followed by 5% of Old age client are defaulted
- ❑ From second graph it's clear that with increasing age tendency to commit default reduces and giving loan to Middle and Old age group is less risky.
- ❑ However that doesn't mean that Young client shouldn't be given loan , their activities should be tracked carefully and all the formalities must be done

2:INCOME GROUPS ANALYSIS



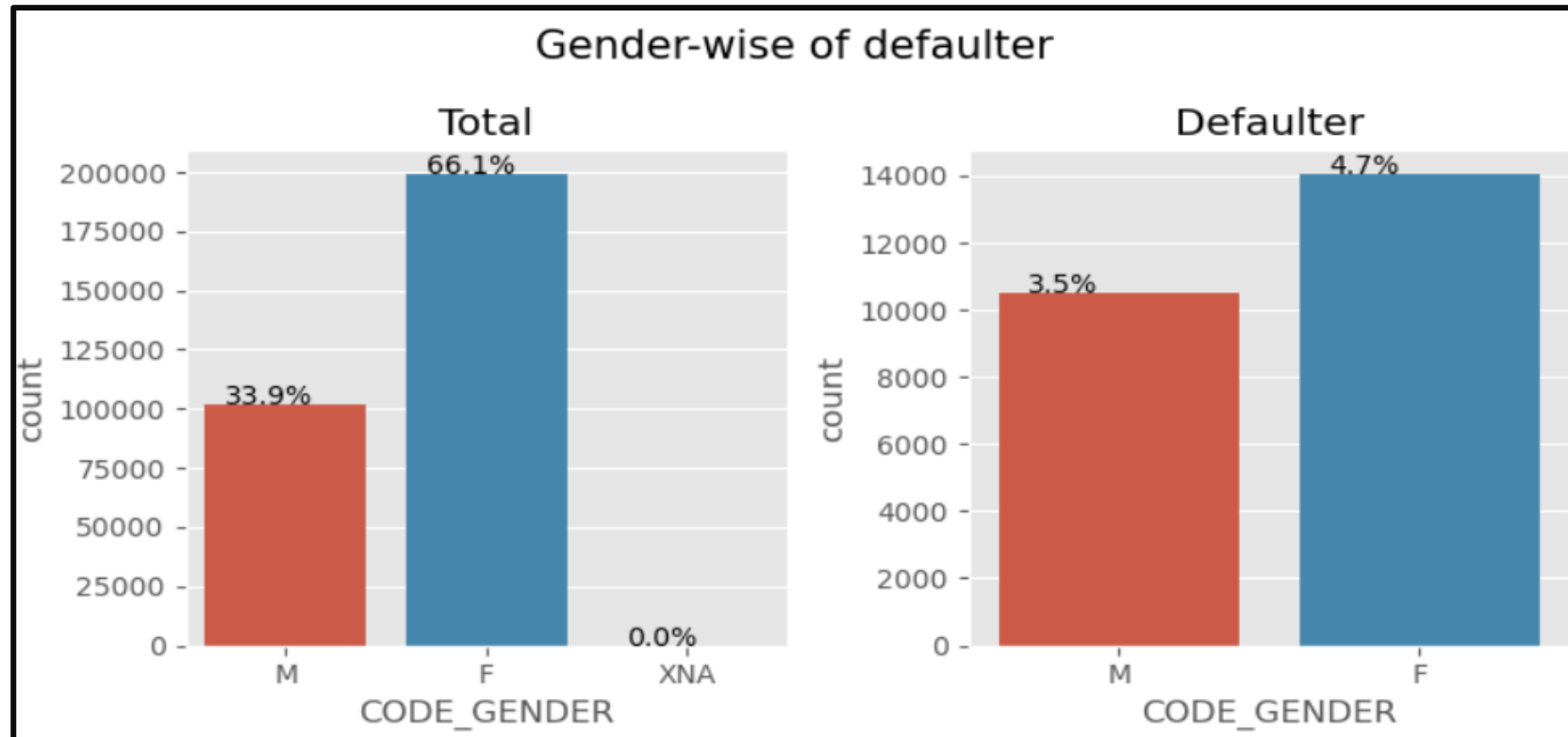
- ❑ There is a high rate of defaulter in 'Low' and 'High' income group.
- ❑ People in 'Medium' and 'Very High' income group.
- ❑ Although the default rate of lower income group is low as they rarely apply for loan

3: TYPE OF CONTRACT & PROCESS START DAY



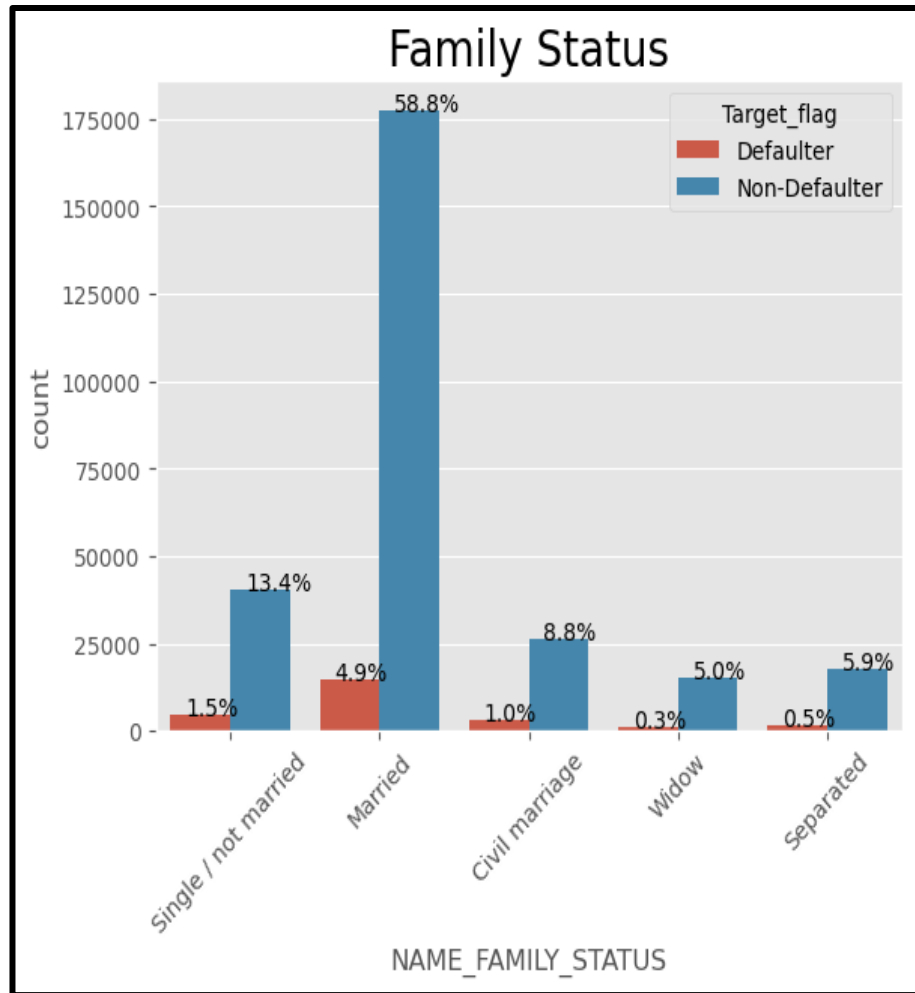
- ~90.5 % loans (82.9% + 7.6%) are cash loans and approx. 83% are non-defaulters .
- 9.5 % loans (9% + 0.5%) are revolving and 9% are non-defaulter.
- All weekdays have similar number of applicants than weekend(Saturday and Sunday) , this indicates that most clients are working.
- These are good indicators for minimising credit risk.

4: GENDER-WISE ANALYSIS



- ❑ The number of females is almost double the number of males
- ❑ Percentage of defaulter in females is more than in males. However, if we look at the ratio of defaulter to gender count, ratio is less for females than males
- ❑ Hence giving loans to females is less risky than males

5:FAMILY STATUS ANALYSIS



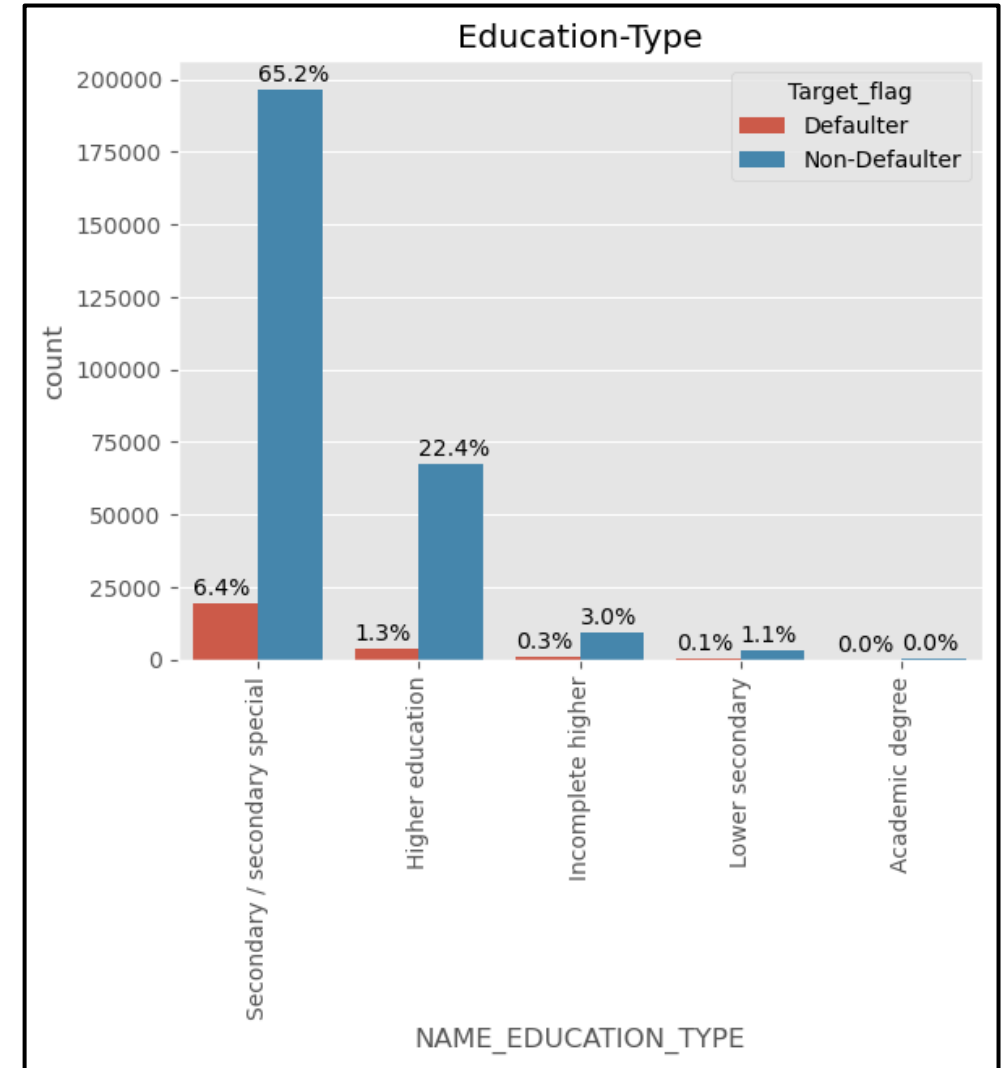
Value	Percentage of Defaulters
Civil marriage	10.000000
Single / not married	9.874194
Separated	8.241165
Married	7.638657
Widow	5.845498

- Approx 64% of loans are applied by married people and amongst them only 7.6% of them are defaulter.
- Maximum defaults are done by Single and Civil marriage clients.
- Widows are least risky however they contribute only 5.5 % of the loan applications.
- Among the clients Married and widows are the two most preferred class.

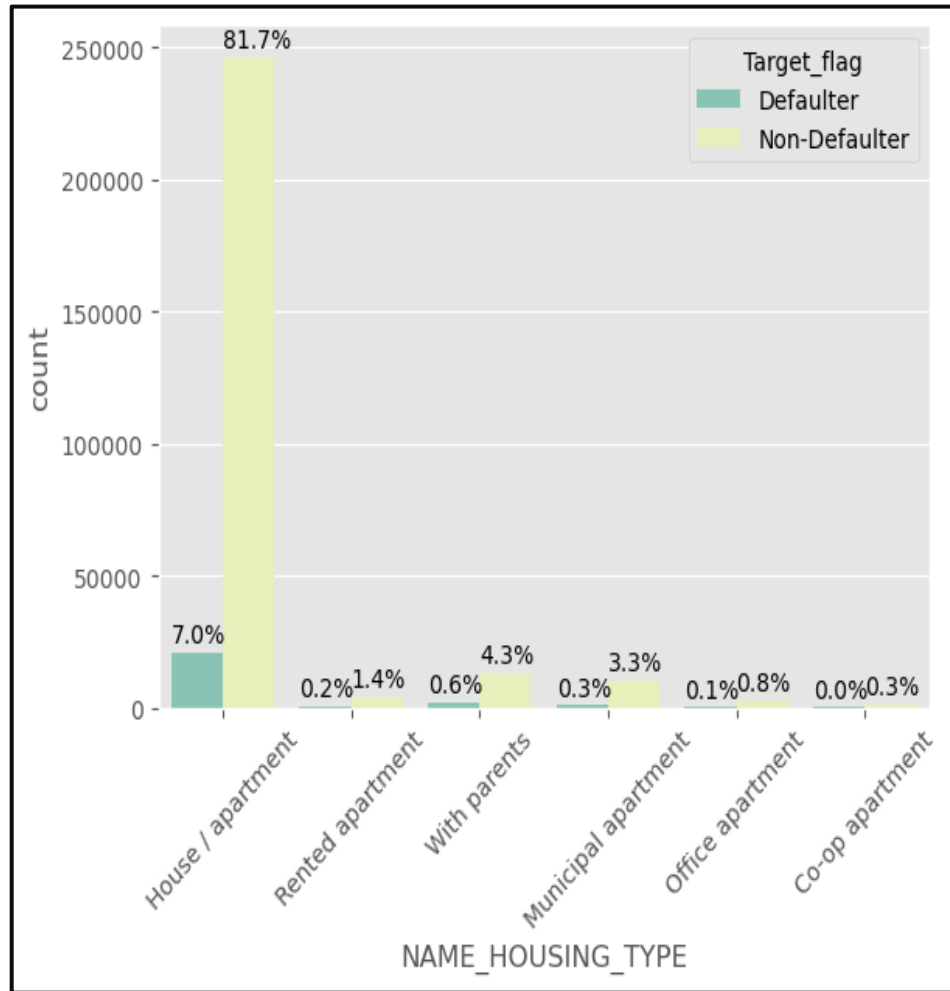
6: EDUCATION-WISE ANALYSIS

- Most of the client have secondary education (~71%) followed by higher education (~24%) .
- From the table we can see that as the education level increases the less likely client is to default
- So while granting loan preference must be given to client with higher education
- However it doesn't mean client with lower education level must be given loan but they must be tracked and all the formalities should be monitored

Value	Percentage of Defaulters
Lower secondary	10.999736
Secondary / secondary special	8.971290
Incomplete higher	8.543515
Higher education	5.444435
Academic degree	1.986755



7: HOUSING ATTRIBUTES ANALYSIS

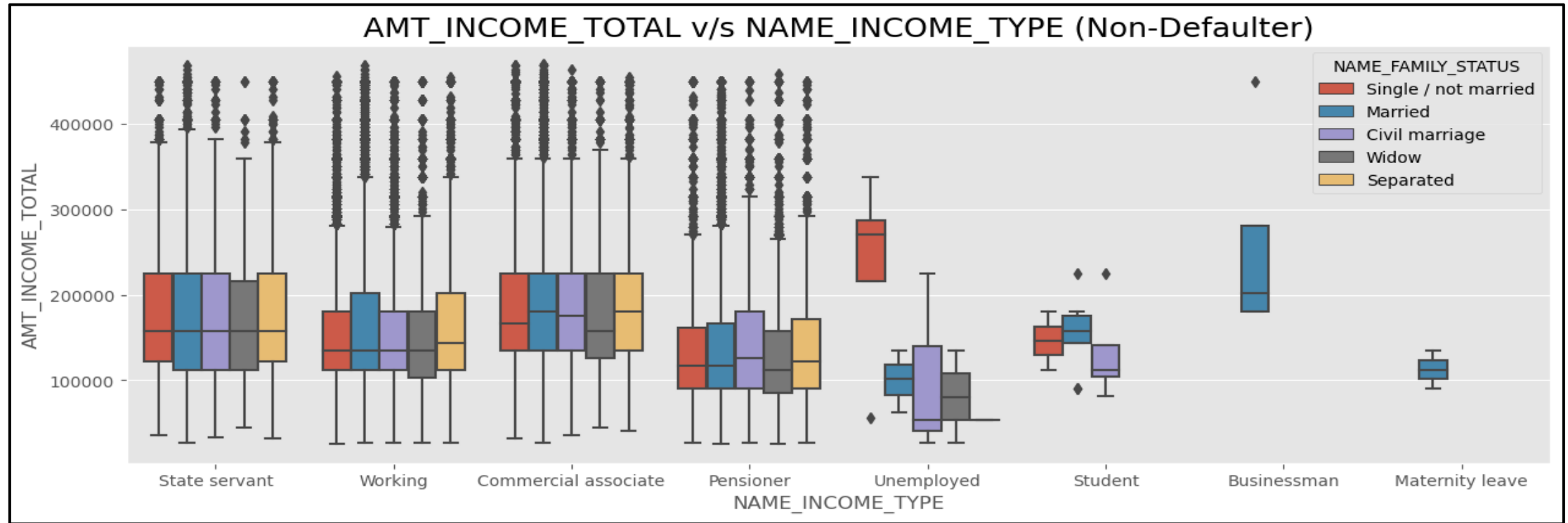


Value	Percentage of Defaulters
Rented apartment	12.41
With parents	11.73
Municipal apartment	8.60
Co-op apartment	7.96
House / apartment	7.87
Office apartment	6.66

- Most of the client live in house/apartment
- Rented apartment clients are very low in number but have highest % of defaulter.
- Of all housing types house/apartment has second lowest default % followed by office apartment but it's very low in number.
- It is clear that house/apartment clients are the most preferred client when housing attribute is considered

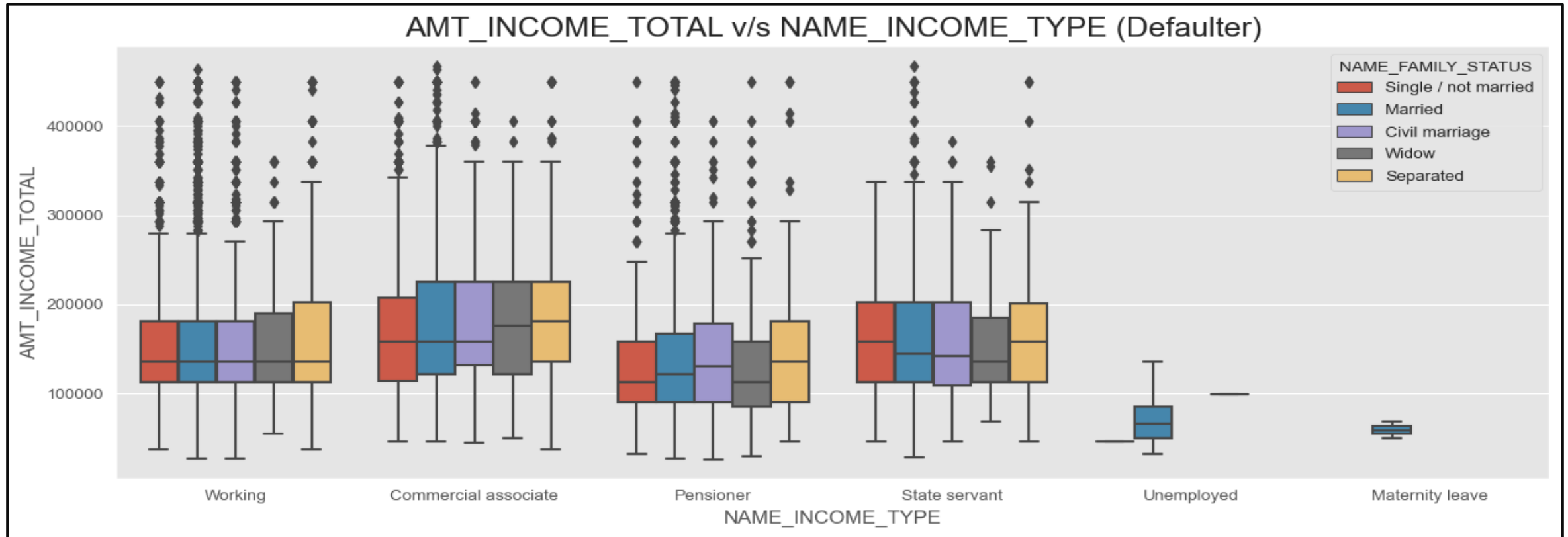
BIVARIATE ANALYSIS

1: TOTAL INCOME V/S INCOME TYPE (for Non-Defaulters)



- ❑ We can see that Income type 'Commercial associate' the income amount is mostly equal with family status. It does contain many outliers.
- ❑ Less outlier are having for Businessman but there income amount is higher.
- ❑ Student type of civil marriage family status are have less income amount than others..

2: TOTAL INCOME V/S INCOME TYPE (for Defaulters)



- ❑ In this plot 'Working category' has IQR for all family status almost same with equal median
- ❑ Less potential outliers are there plot in State servant & Commercial associate category and no outliers for Maternity leave and Unemployed
- ❑ Commercial associate are having more median value of income than any others.

TOP 10 CORRELATION

DEFAULTER

AMT_CREDIT	AMT_GOODS_PRICE	0.98
REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.96
CNT_FAM_MEMBERS	CNT_CHILDREN	0.89
DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.87
LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.85
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	0.78
AMT_ANNUITY	AMT_CREDIT	0.76
AMT_GOODS_PRICE	AMT_ANNUITY	0.75
AGE	DAYS_EMPLOYED	0.58
REG_REGION_NOT_LIVE_REGION	REG_REGION_NOT_WORK_REGION	0.50

NON-DEFAULTER

AMT_CREDIT	AMT_GOODS_PRICE	0.99
REGION_RATING_CLIENT	REGION_RATING_CLIENT_W_CITY	0.95
CNT_FAM_MEMBERS	CNT_CHILDREN	0.88
REG_REGION_NOT_WORK_REGION	LIVE_REGION_NOT_WORK_REGION	0.86
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.86
REG_CITY_NOT_WORK_CITY	LIVE_CITY_NOT_WORK_CITY	0.83
AMT_CREDIT	AMT_ANNUITY	0.79
AMT_ANNUITY	AMT_GOODS_PRICE	0.79
DAYS_EMPLOYED	AGE	0.62
AMT_INCOME_TOTAL	AMT_ANNUITY	0.46

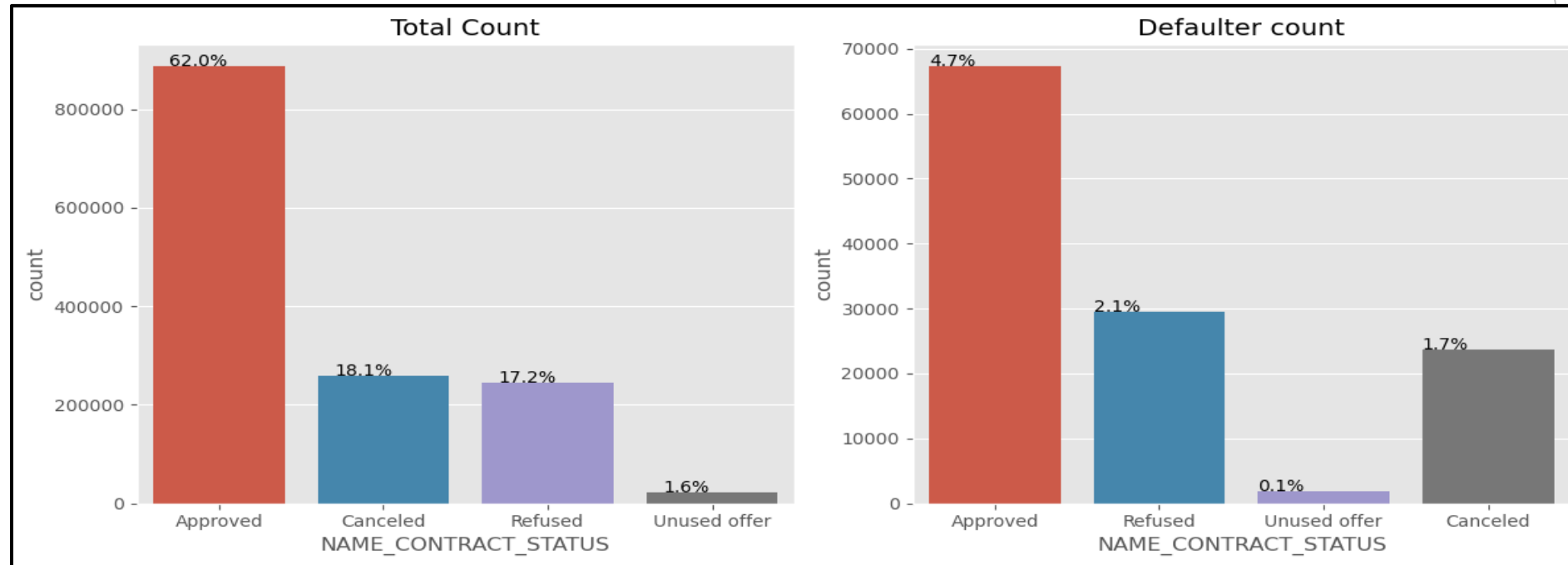
- ❑ Credit amount and prices of goods are strongly correlated.
- ❑ Correlation between credited amount and annuity , annuity and goods prices are similar for both defaulter and non-defaulter
- ❑ Age and days employed correlation is almost same for both .



PREVIOUS APPLICATION DATASET ANALYSIS

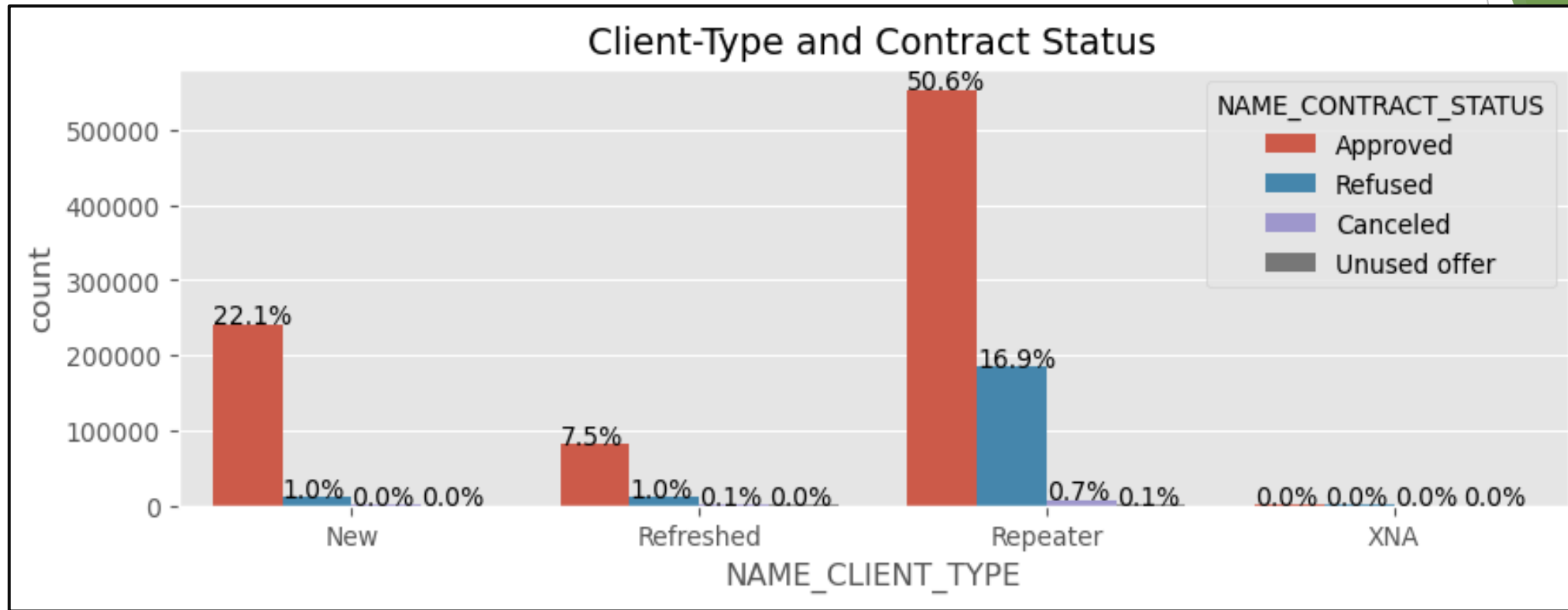
UNIVARIATE ANALYSIS

1: CONTRACT STATUS OF PREVIOUSLY APPLIED DEFAULTERS



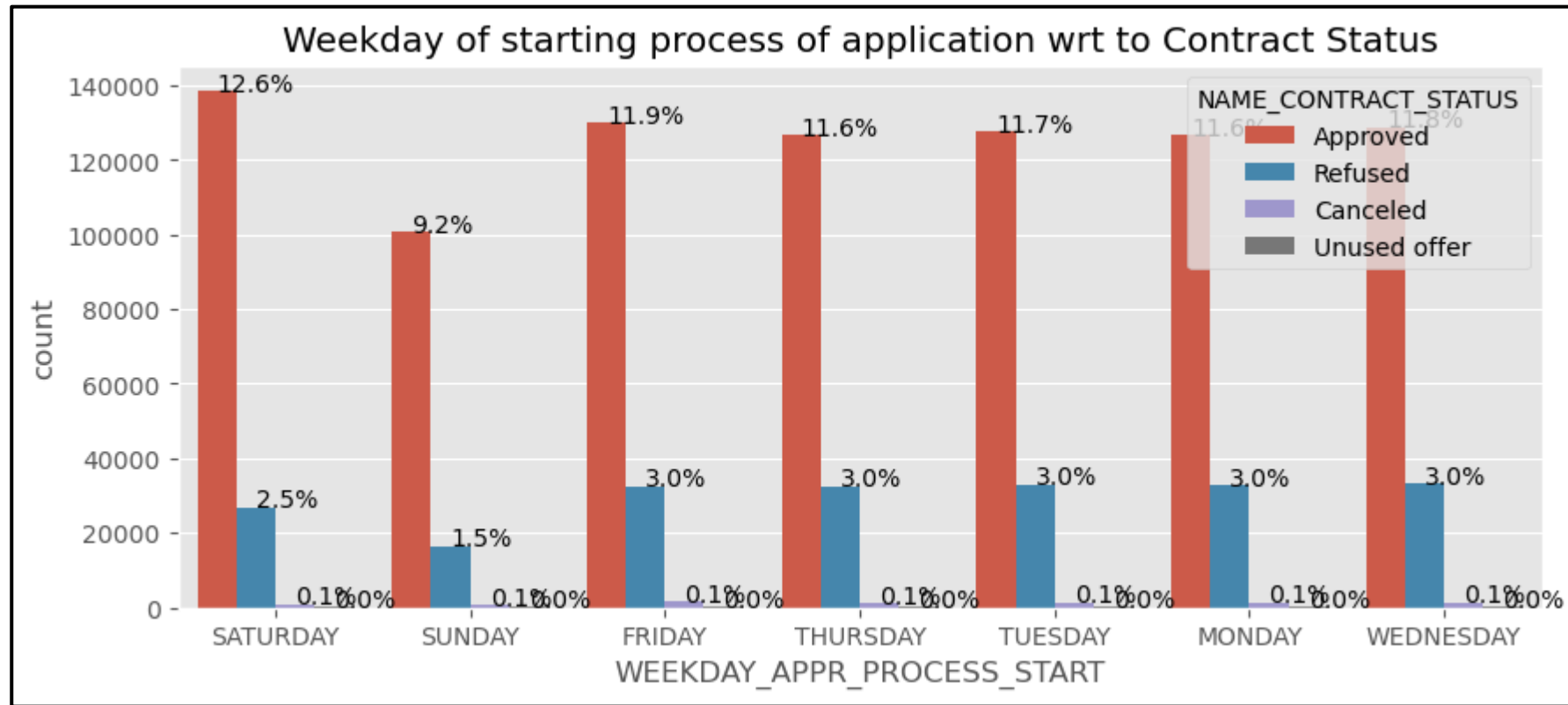
- ❑ 62% of the applicant's previous loan was approved, 18% of loans were cancelled, almost 17% of loans were refused.
- ❑ Amongst the defaulter of current data who have a history of previous application, 4.7% of loans were approved, 2.1% refused and 1.7% were cancelled
- ❑ People who were defaulter and were approved loan should be skeptically considered for new loan

2: ANALYSIS OF CLIENTS TYPE WITH DIFFERENT STATUS



- ❑ Majority of client base are repeater clients
- ❑ Ratio of Approved status to Refused status is high for repeaters considering their large market base.
- ❑ Refreshed clients contribute lowest , not considering XNA clients are their information is not available and their market base is also negligible .
- ❑ Repeater clients must be given preference by the banks for business profit.

3 : PROCESS START DAY WRT CONTRACT STATUS



- ❑ Maximum loans were approved on Saturday, on all other days approved loans lie between 120k to 140k except for Sunday (~100K), which is a good sign for bank.
- ❑ Number of refused loans are almost same everyday except Saturday and Sunday, which is very low.
- ❑ Cancelled and unused offer are almost negligible.
- ❑ Number of loans are getting approved consistently everyday except Sunday

TOP 5 CORRELATION

For Approved , Refused , Unused offer and Cancelled contact status

APPROVED

AMT_APPLICATION	AMT_CREDIT	0.96
AMT_GOODS_PRICE	AMT_ANNUITY	0.83
AMT_CREDIT	AMT_ANNUITY	0.83
AMT_ANNUITY	AMT_APPLICATION	0.81
AMT_APPLICATION	CNT_PAYMENT	0.68

REFUSED

AMT_APPLICATION	AMT_CREDIT	0.98
AMT_ANNUITY	AMT_GOODS_PRICE	0.85
AMT_APPLICATION	AMT_ANNUITY	0.84
AMT_CREDIT	CNT_PAYMENT	0.71
AMT_APPLICATION	CNT_PAYMENT	0.70

UNUSED OFFER

AMT_CREDIT	AMT_APPLICATION	0.98
AMT_ANNUITY	AMT_CREDIT	0.94
AMT_APPLICATION	AMT_ANNUITY	0.94
CNT_PAYMENT	AMT_APPLICATION	0.17
	AMT_GOODS_PRICE	0.16

Cancelled

AMT_CREDIT	AMT_APPLICATION	0.99
AMT_ANNUITY	AMT_GOODS_PRICE	0.82
	AMT_CREDIT	0.82
AMT_CREDIT	CNT_PAYMENT	0.63
AMT_APPLICATION	CNT_PAYMENT	0.61

TOP 5 CORRELATION

For Approved , Refused , Unused offer and Cancelled contact status

- ❑ Correlation between AMT_APPLICATION and AMT_CREDIT is highest in all four contract status.
- ❑ AMT_GOOD_PRICE AND AMT_ANNUITY has second highest correlation for all the cases except 'Unused Offer' wherein AMT_CREDIT and AMT_ANNUITY is second highest.
- ❑ AMT_APPLICATION and CNT_PAYMENT has lower correlation in all cases except 'Unused offer' where CNT_PAYMENT and AMT_GOOD_PRICE is lower

CONCLUSION

- ❑ Bank should target middle age and old age clients as their number is really high and defaulter percentage is low.
- ❑ Banks should give out more cash loans as the number of cash loans applicants is really high as compared to revolving loans and default percentage is low in cash loans.
- ❑ Female pose less risky customer as the ratio of default to total count is less for females than males. Also they present greater market base for banks.
- ❑ Married clients present the highest market base for banks with second lowest default percentage. Banks should give preference to married clients.
- ❑ Education level of an applicant is a good indicator of capability of a client to repay the loan. Higher the education level , more reliable is the client.
- ❑ Repeated business is good for bank . For large share of total population currently their default percentage is low.
- ❑ There is a high correlation for goods price and credit amount in both current application and previous application dataset . This indicates that people are getting required loan amount.