



CREDIT EDA ANALYSIS

BANK DATASETS ASSESSMENT

KEY POINTS TO PERFORM ANALYSIS

Loan analysis is an evaluation method that determines if loans are made on feasible terms and if potential borrowers can and are willing to pay back the loan.

- ✓ The proper assessment of credit risk can go a long way in reducing the impact of a lender's possible loss.
- ✓ Credit risk assessment helps organizations know whether a borrower can pay back a loan.
- ✓ If a borrower's credit risk is high, their loan's interest rate will be increased.

PROBLEM STATEMENT

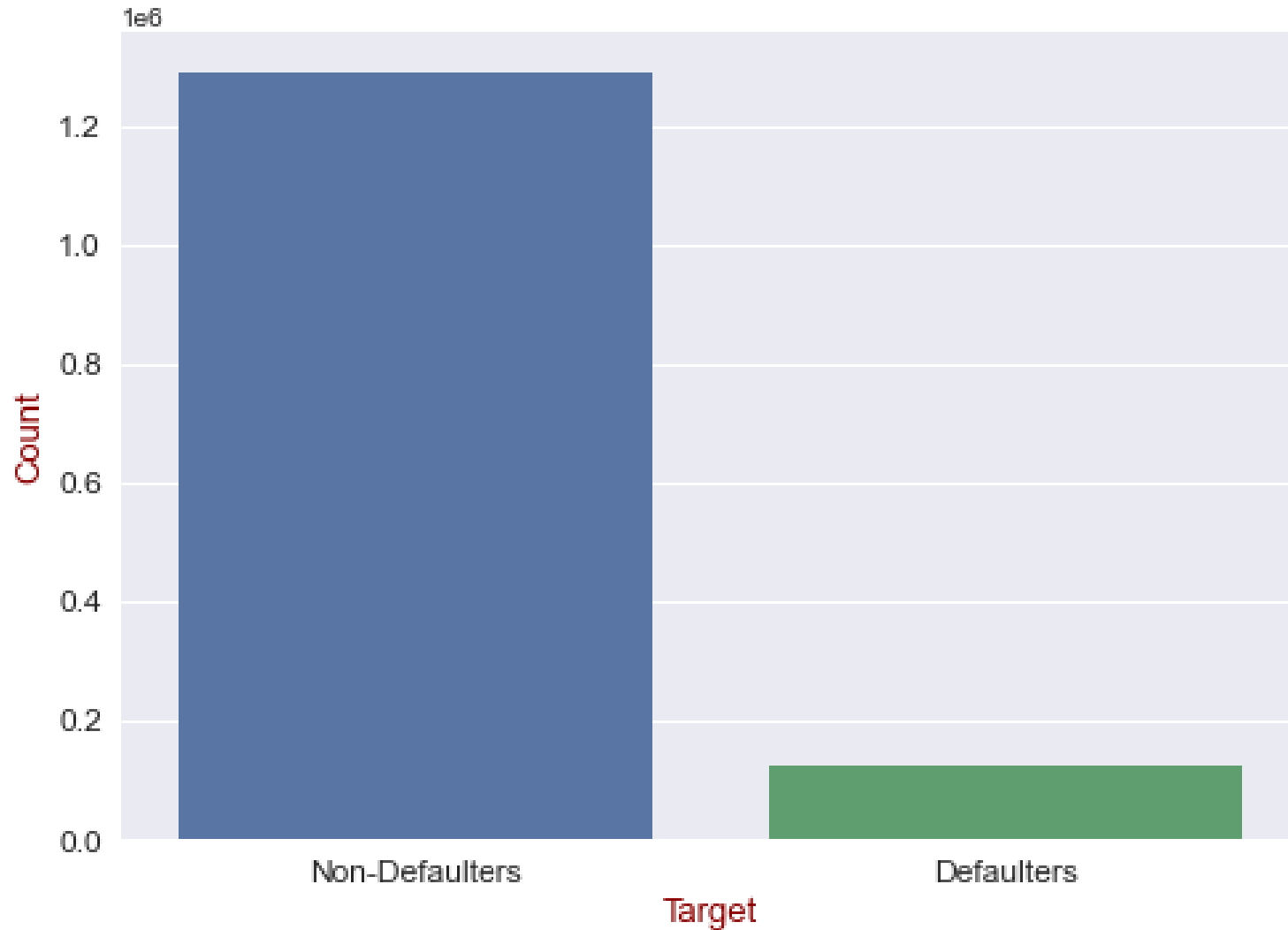
When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- ✓ 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.

STEPS PERFORMED IN THE ANALYSIS

- ✓ Import the necessary libraries.
- ✓ Identify attributes that are key to understand the defaulters.
- ✓ Check the data structure for info, describe the numerical data.
- ✓ Check the null values and impute suitable approach to proceed with dataset.
- ✓ Check the outliers and identify the data imbalance.
- ✓ Check the data types of the attributes and make necessary changes.
- ✓ Visualize the patterns drawn using various attributes and analyse the percentage of defaulters.
- ✓ Identify the strong co-relation existence in the datasets.

Defaulters vs Non-Defaulters Ratio



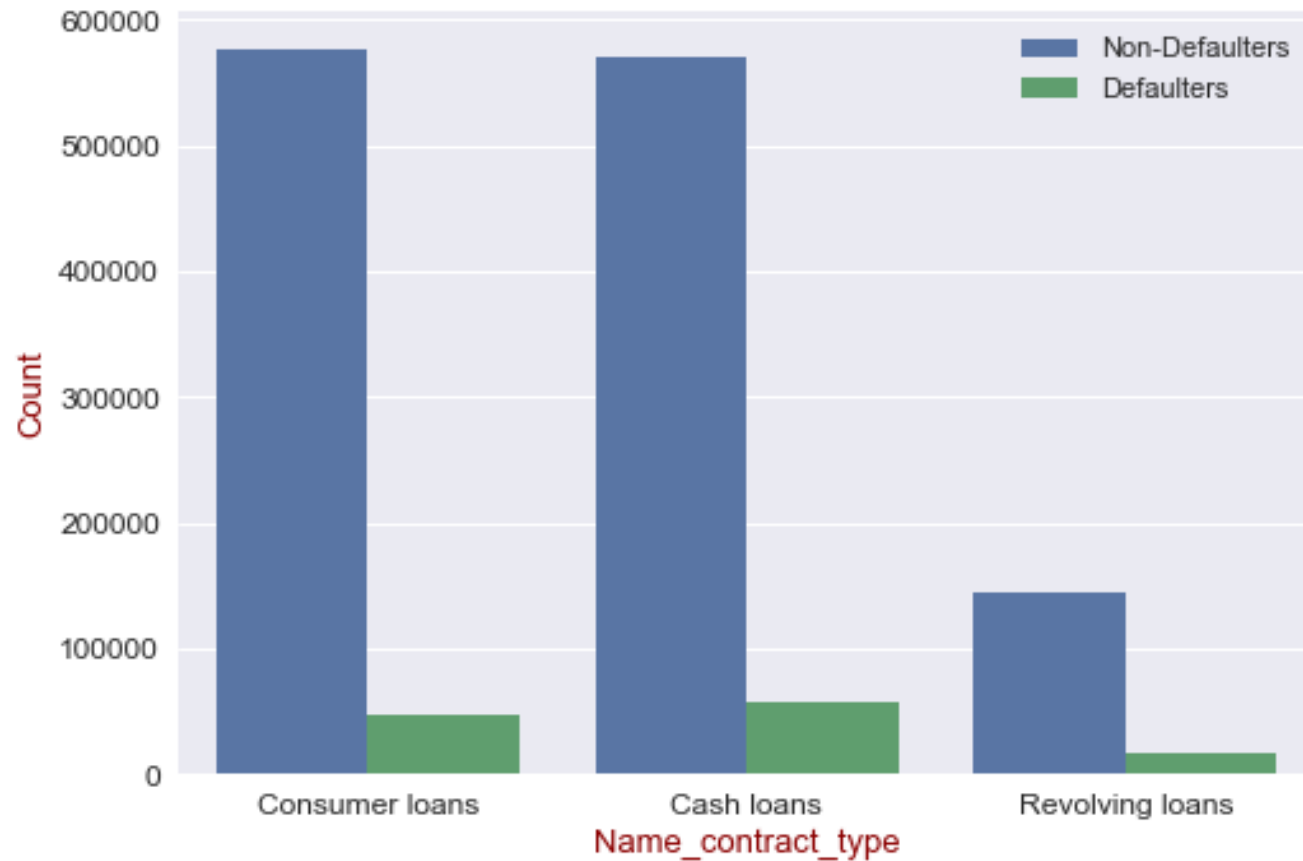
DATA IMBALANCE

Percentage of Defaulters: 8.65

Percentage of Non-Defaulters: 91.35

Since there are less number of defaulters, it is clear that Data is imbalanced with a percentage of 9.

Loan Type

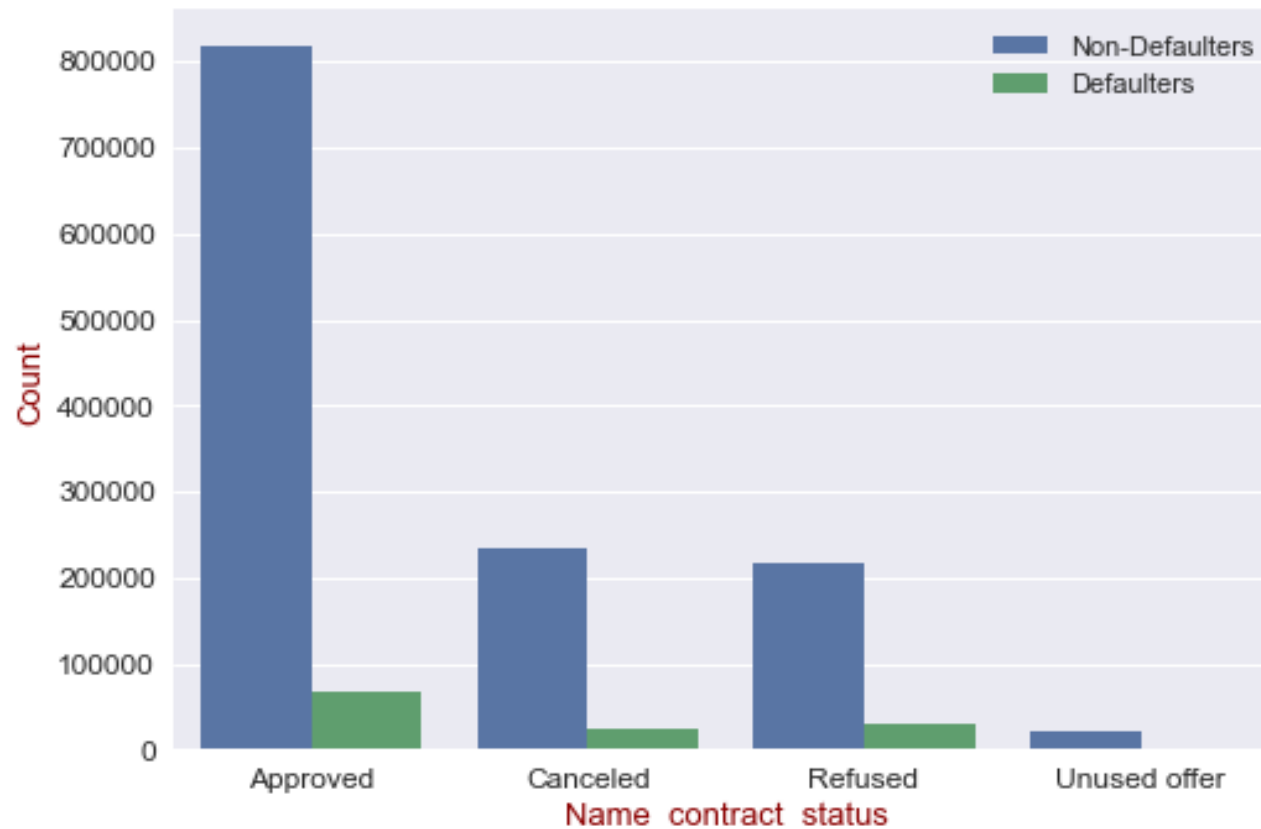


	NAME_CONTRACT_TYPE	Default_Percentage
2	Revolving loans	10.47
1	Cash loans	9.13
0	Consumer loans	7.71

LOAN TYPE

- ✓ On a general count, where count plot is derived, Consumer loans have maximum successful recovery whereas cash loans stands at maximum defaults.
- ✓ Secondly, when compared the individual loan type with its own total count, we can notice that revolving loans stands at maximum defaults.

Status Type

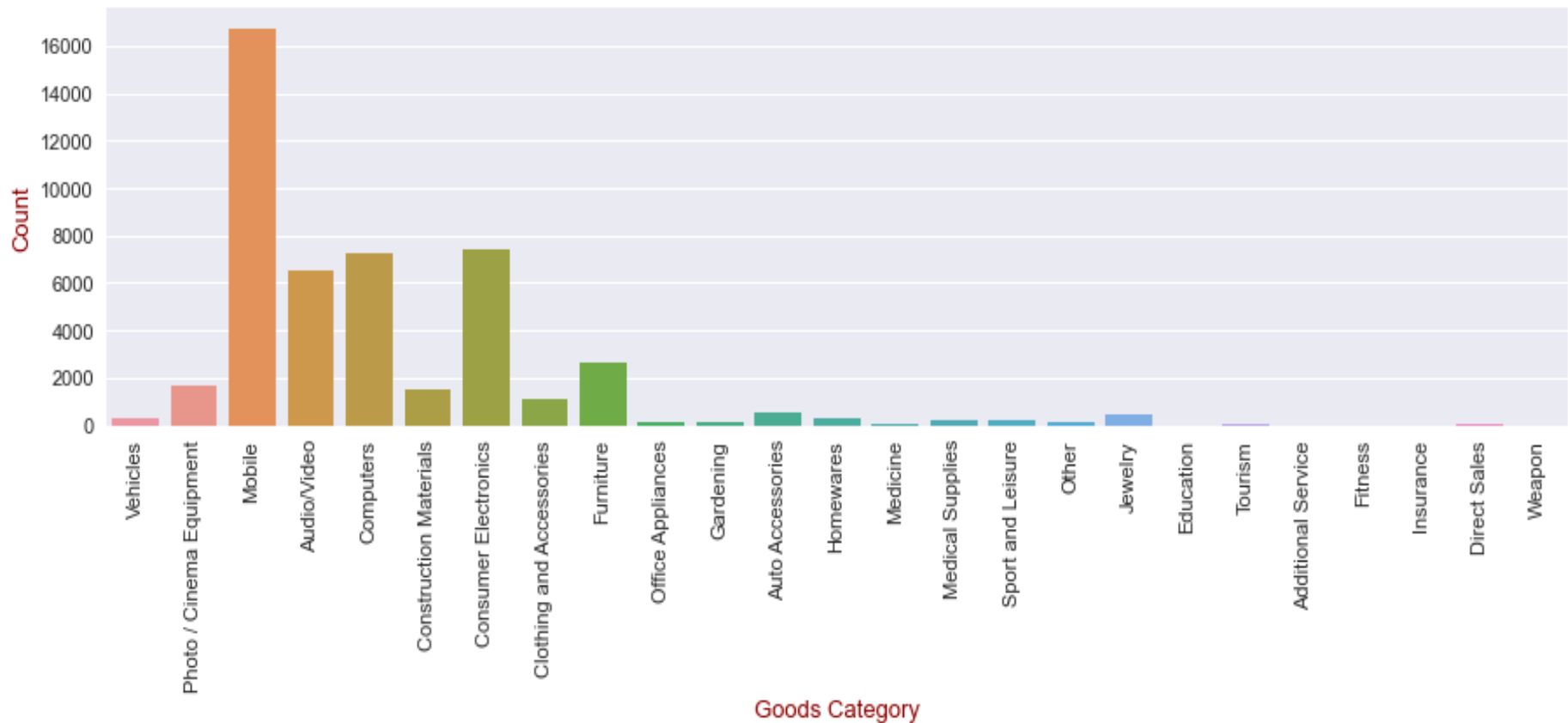


	NAME_CONTRACT_STATUS	Default_Percentage
2	Refused	11.99
1	Canceled	9.16
3	Unused offer	8.25
0	Approved	7.59

LOAN STATUS

- ✓ On a general count, where count plot is derived on various statuses of loan, approved status contribute for maximum defaults.
- ✓ Secondly, when compared the individual loan status with its own total count, it can be observed that refused status stands first.
- ✓ Its clear with percentage derived above on refused loans, that a loan application can't be rejected on the basis of previous performance.

Defaulters by Goods Category

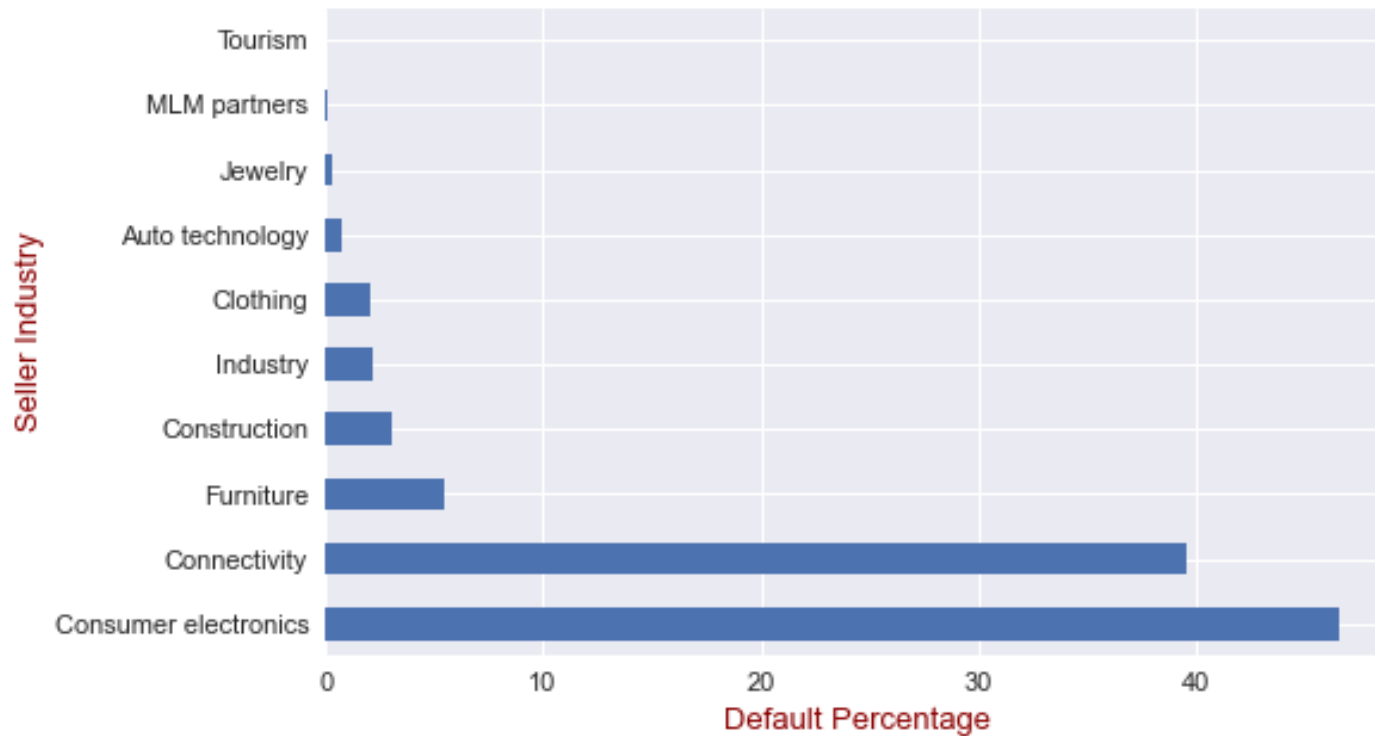


GOODS

	NAME_GOODS_CATEGORY	Default_Percentage
23	Insurance	10.34
0	Vehicles	10.06
1	XNA	9.38
17	Auto Accessories	9.07
14	Jewelry	9.04
4	Mobile	8.65
15	Office Appliances	8.26
21	Direct Sales	8.18
9	Computers	8.10
20	Weapon	7.69
5	Audio/Video	7.66

- ✓ Technology related goods such as Mobile, Consumer Electronics, Computers and Audio/Video holds the highest defaults with total population.
- ✓ When observed with same type count percentage, its the other way around. Automobiles category defaulted more.

Defaulters by Seller Industry

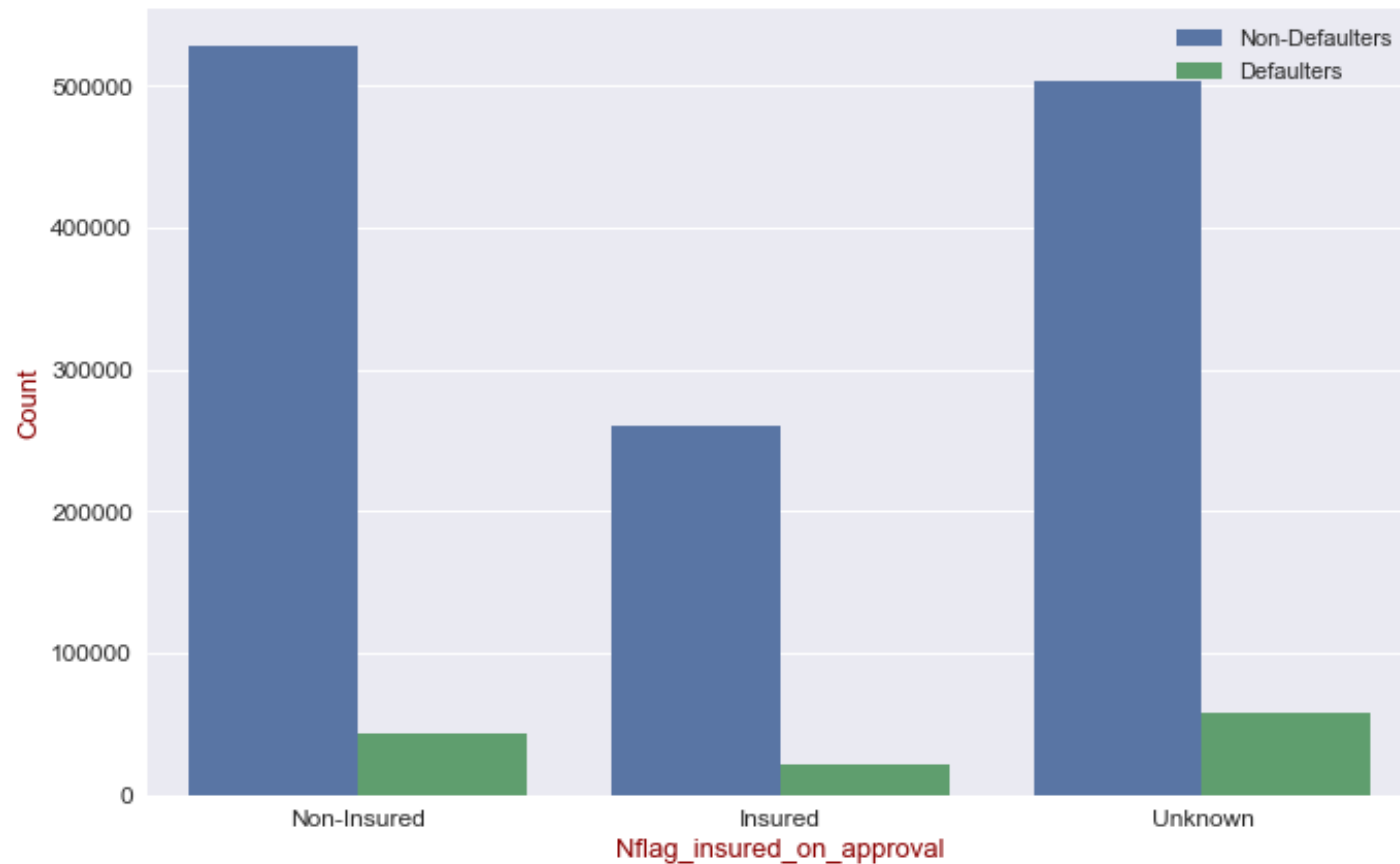


	NAME_SELLER_INDUSTRY	Default_Percentage
0	Auto technology	10.37
1	XNA	9.36
4	Connectivity	9.19
9	Jewelry	8.88
3	Consumer electronics	7.55
7	Industry	7.21
5	Construction	6.57

SELLER INDUSTRY

- ✓ Electronic seller industries holds the highest defaults with total population.
- ✓ When observed with same type count percentage, its Auto-technology which stands high.

Insurance on Approval



NFLAG_INSURED_ON_APPROVAL Default_Percentage

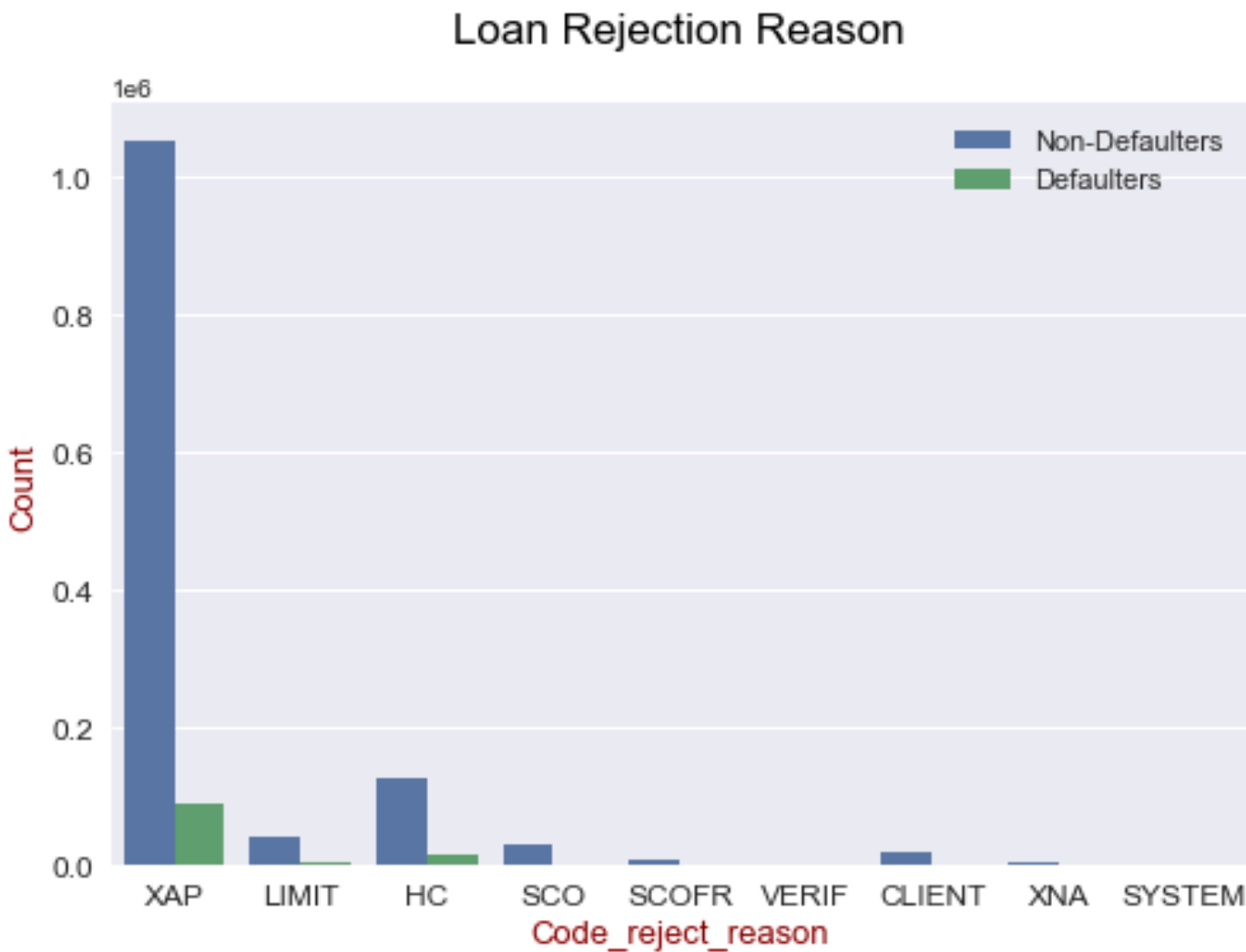
	NFLAG_INSURED_ON_APPROVAL	Default_Percentage
2	Unknown	10.27
1	1.0	7.61
0	0.0	7.58

INSURANCE

It is really important to insure a loan disbursed to borrowers to reduce the default percentage.

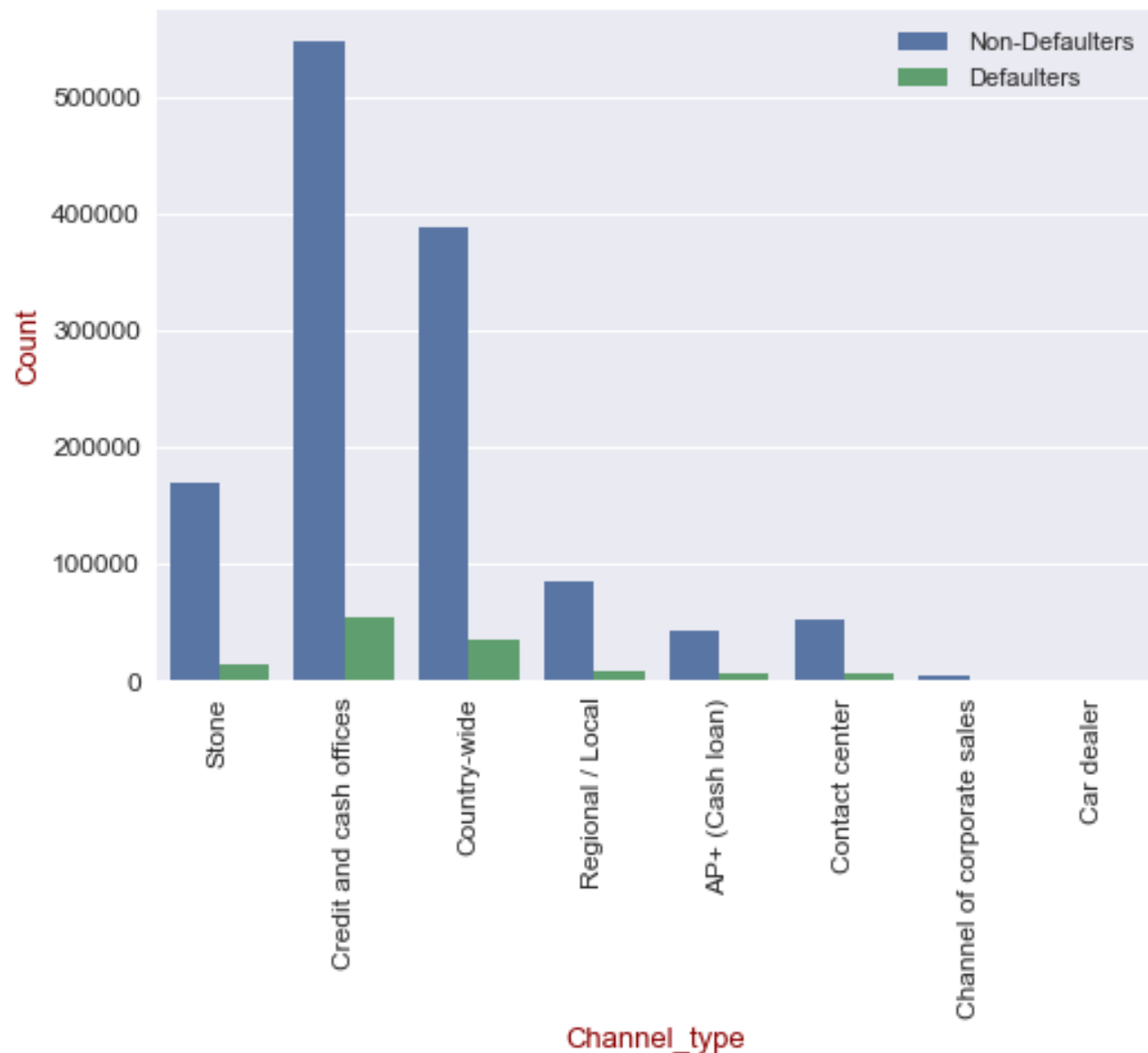
LOAN REJECTION REASON

LIMIT and HC code rejection reasons are at top three on both general count as well as individual population type.



	CODE_REJECT_REASON	Default_Percentage
4	SCOFR	20.93
1	LIMIT	12.56
2	HC	11.73
7	XNA	10.10
3	SCO	10.00
5	VERIF	9.26
6	CLIENT	8.25
0	XAP	7.94
8	SYSTEM	6.25

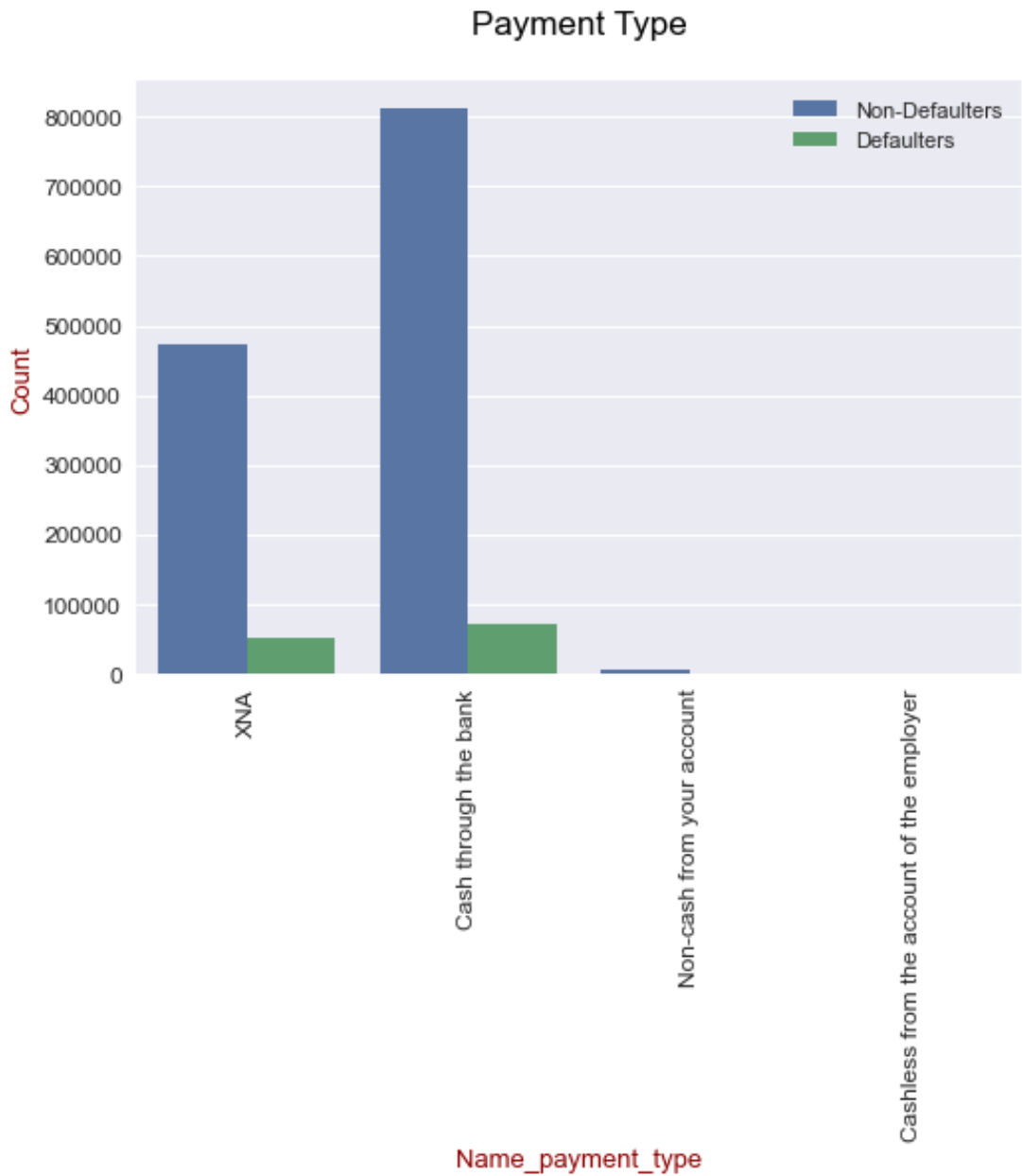
Channel Type



CHANNEL TYPE

- ✓ Credit and cash offices channel type are higher in general count analysis.
- ✓ AP+ (Cash Loan) and Contact centre are highest amongst individual population.
- ✓ Car dealer channel type remains at low profile in both the comparison for defaulters.

	CHANNEL_TYPE	Default_Percentage
4	AP+ (Cash loan)	12.82
5	Contact center	10.27
1	Credit and cash offices	9.01
2	Country-wide	8.25
3	Regional / Local	7.61
0	Stone	7.37
6	Channel of corporate sales	6.38
7	Car dealer	5.17



PAYMENT TYPE

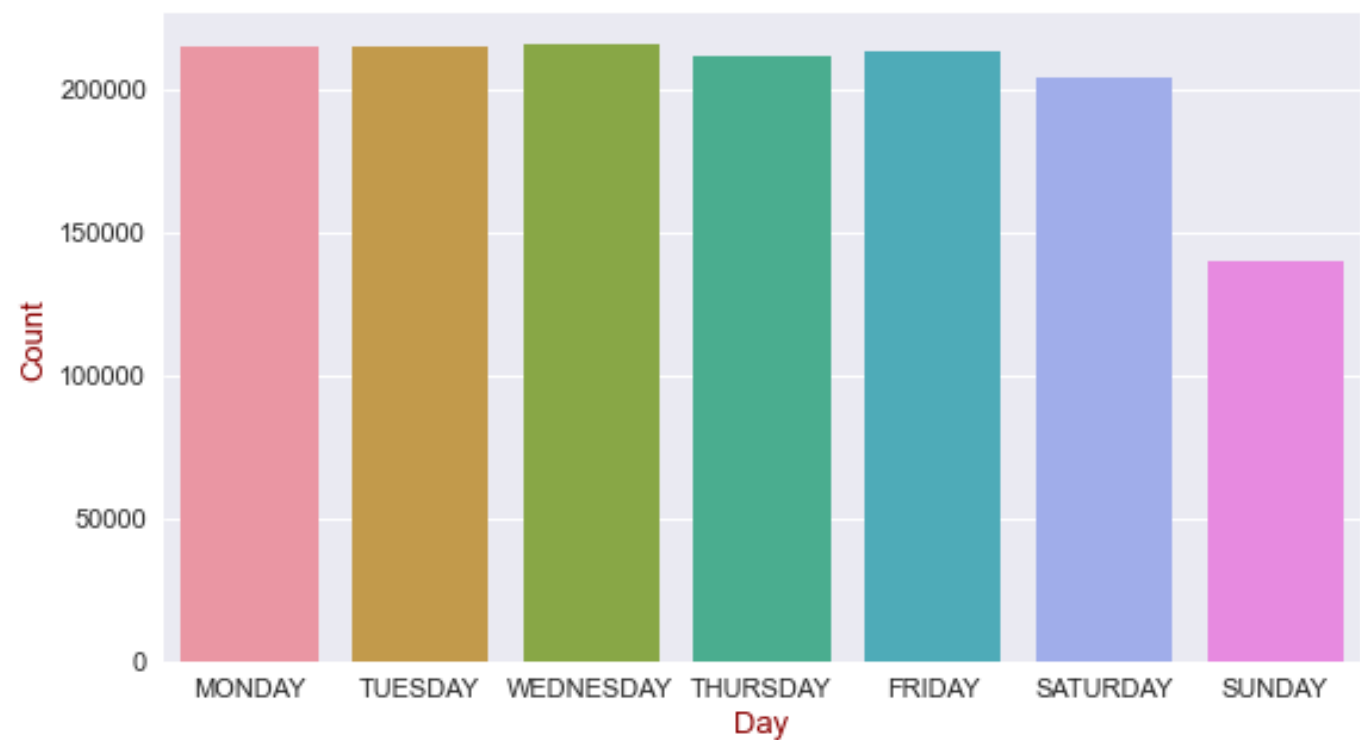
Cashless from the account of the employer is a most safest mode of payment and has least defaulters in both comparison.

	NAME_PAYMENT_TYPE	Default_Percentage
0	XNA	9.62
2	Non-cash from your account	8.21
1	Cash through the bank	8.09
3	Cashless from the account of the employer	8.08

APPLICATION FLOW

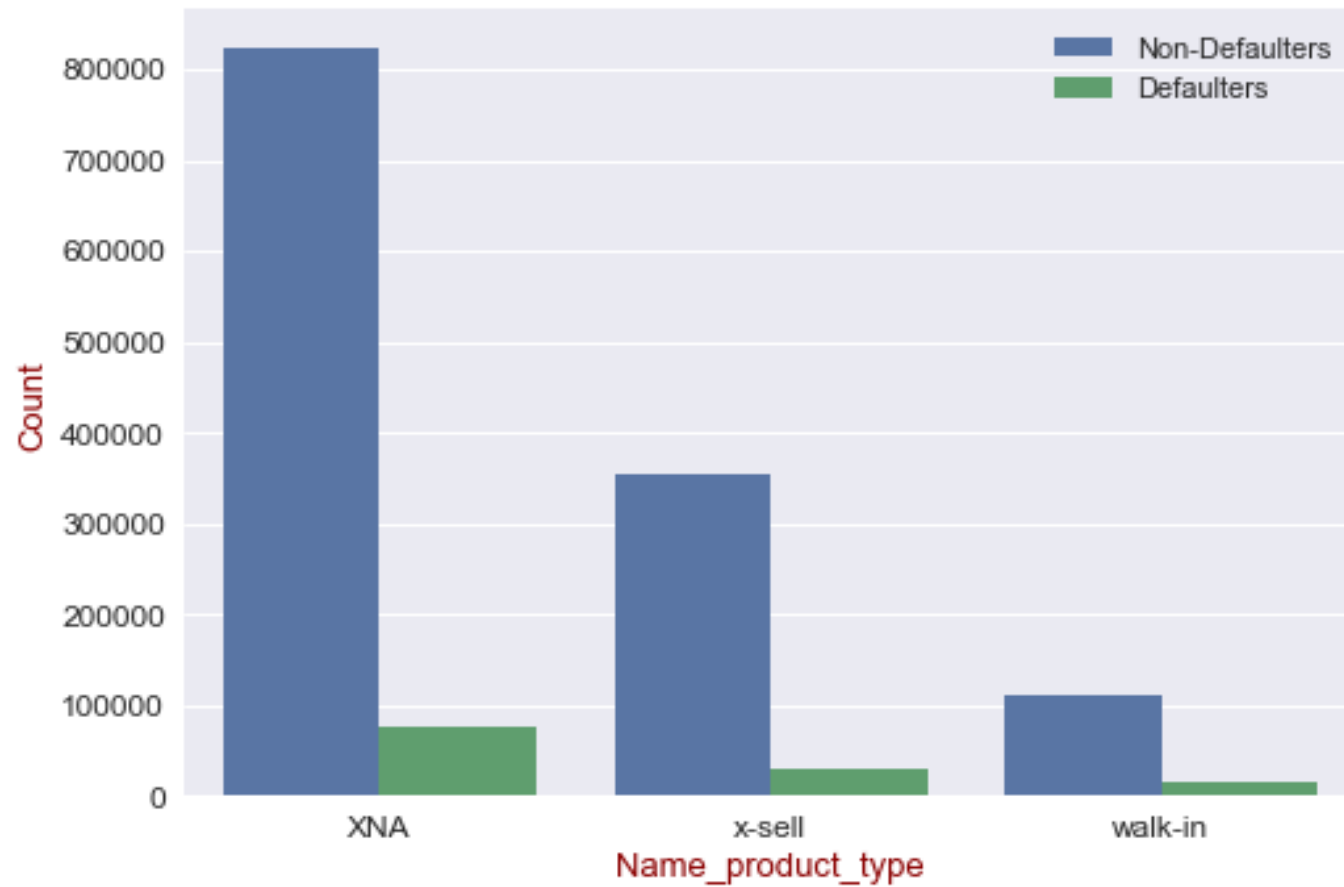
- Percentage of Defaulters' applications received are nearly same each day of the week.
- On a total count, Sunday has received lesser application compared to weekdays.

No. of Applications in a Week



	WEEKDAY_APPR_PROCESS_START	Default_Percentage
3	THURSDAY	8.89
5	MONDAY	8.86
4	TUESDAY	8.76
1	FRIDAY	8.75
6	WEDNESDAY	8.70
0	SATURDAY	8.34
2	SUNDAY	8.05

Product Type



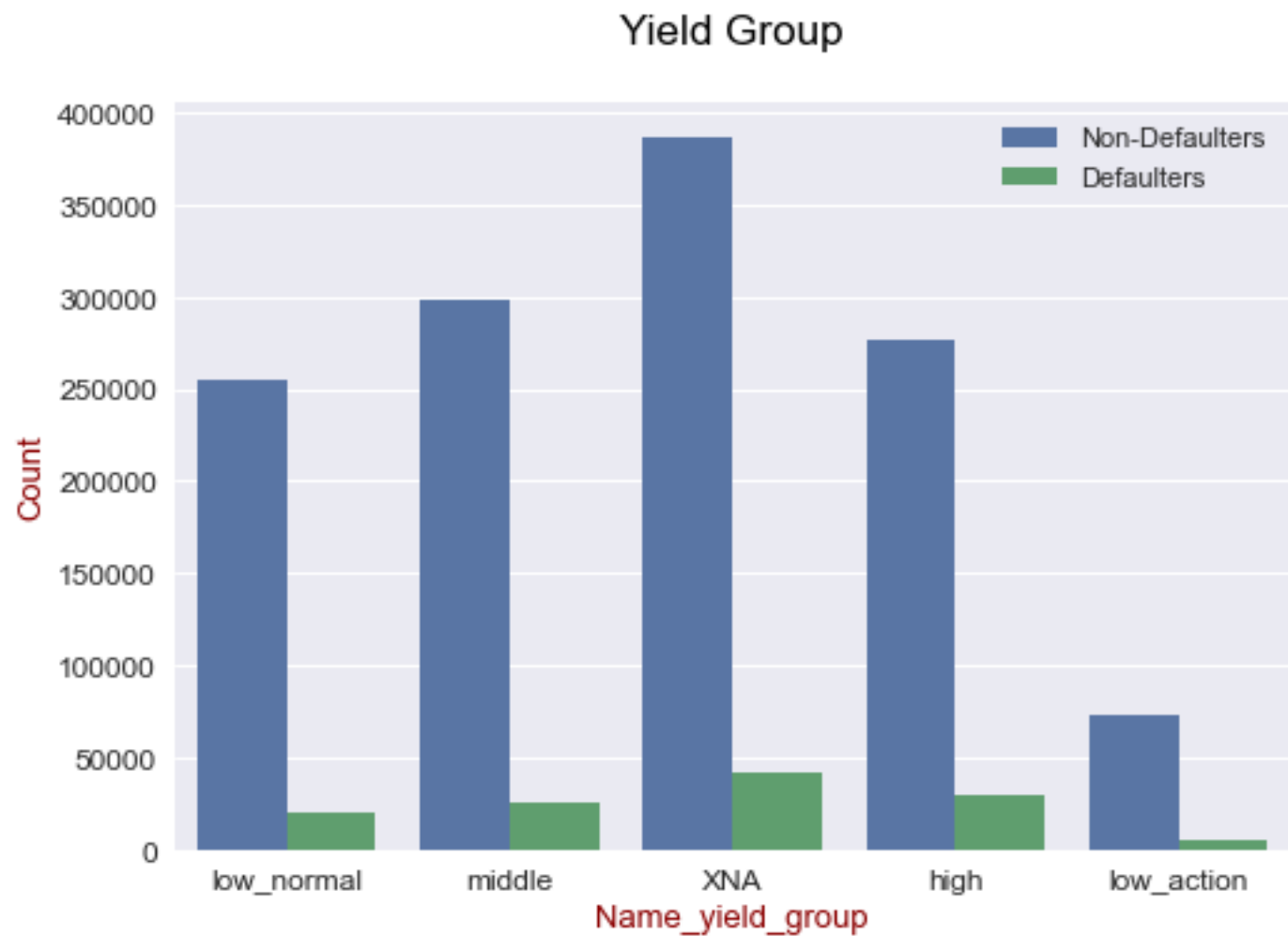
	NAME_PRODUCT_TYPE	Default_Percentage
2	walk-in	12.47
0	XNA	8.35
1	x-sell	8.09

PRODUCT TYPE

- ✓ X-sell (parallelly offered loan when bought other products by client) is more defaulted when compared to Walk-in as per total count.
- ✓ Its the other way that, Walk-in contributes more to defaulters when compared with same population count.

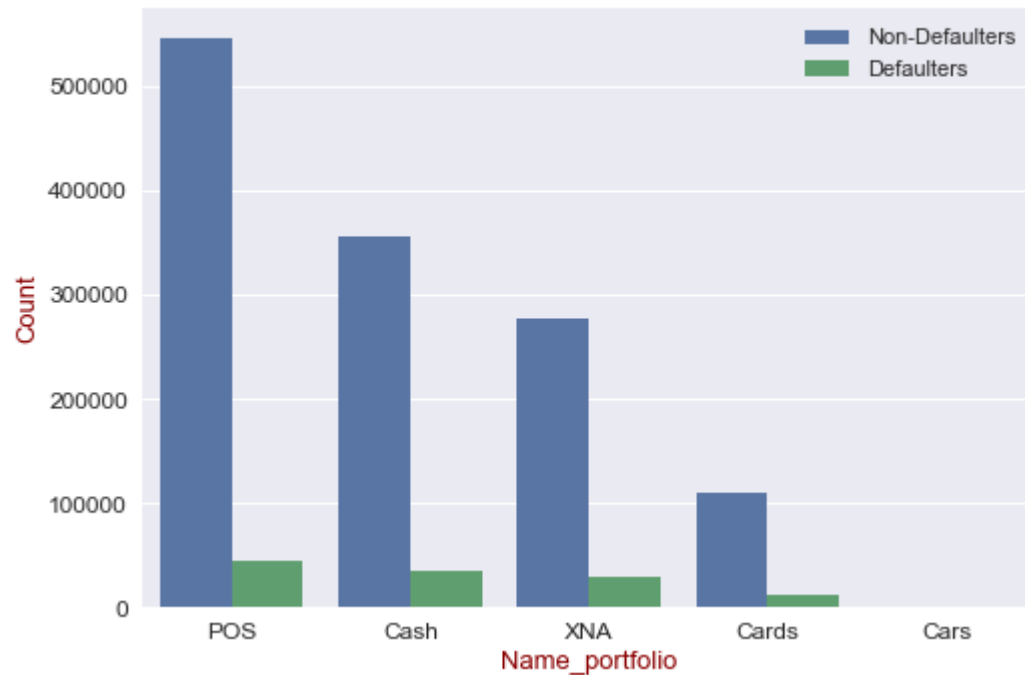
INTEREST ON LOAN

Both high and middle grouped interest categories have same defaulters in both general and individual population count analysis.



NAME_YIELD_GROUP		Default_Percentage
2	XNA	9.86
3	high	9.55
1	middle	8.04
0	low_normal	7.13
4	low_action	6.45

Portfolio



	NAME_PORTFOLIO	Default_Percentage
3	Cards	10.16
2	XNA	9.73
1	Cash	8.88
0	POS	7.64
4	Cars	5.53

PORTFOLIO

- ✓ POS (retailer offered loan at point of purchase) has more defaulters when compared to other portfolio on general count.
- ✓ Cards (third party offered loan at point of purchase) has more defaulters when compared to other portfolio on individual count.

CREDIT Vs DOWN PAYMENT

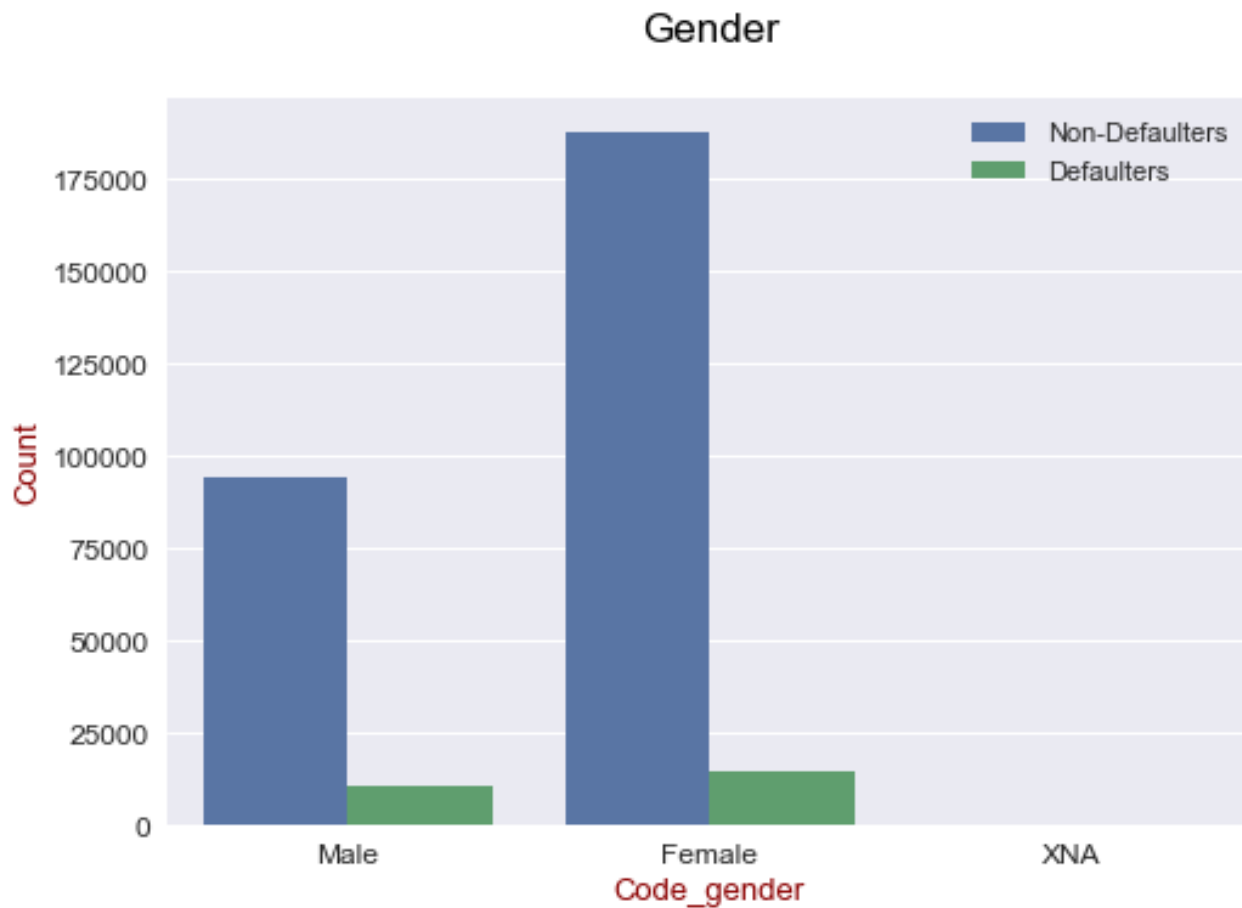
- ✓ Number of defaulters are less for larger amount of credit.
- ✓ For higher down payment, defaulters are less.

Credit vs Down Payment - Non-Defaulters



Credit vs Down Payment - Defaulters





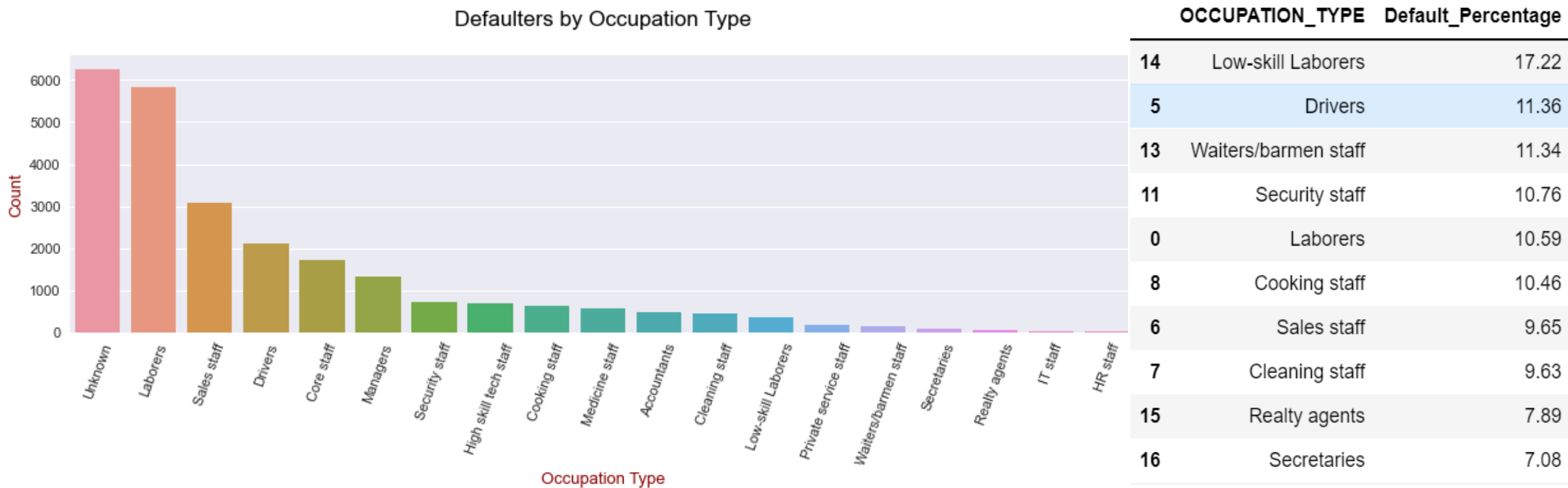
GENDER

- ✓ As per total count and population analysis, female applicants seems to be higher in default zone.
- ✓ When compared with same gender count, its 3% more for male applicants who are likely to default.
- ✓ Overall, female count of applicants are more in dataset.

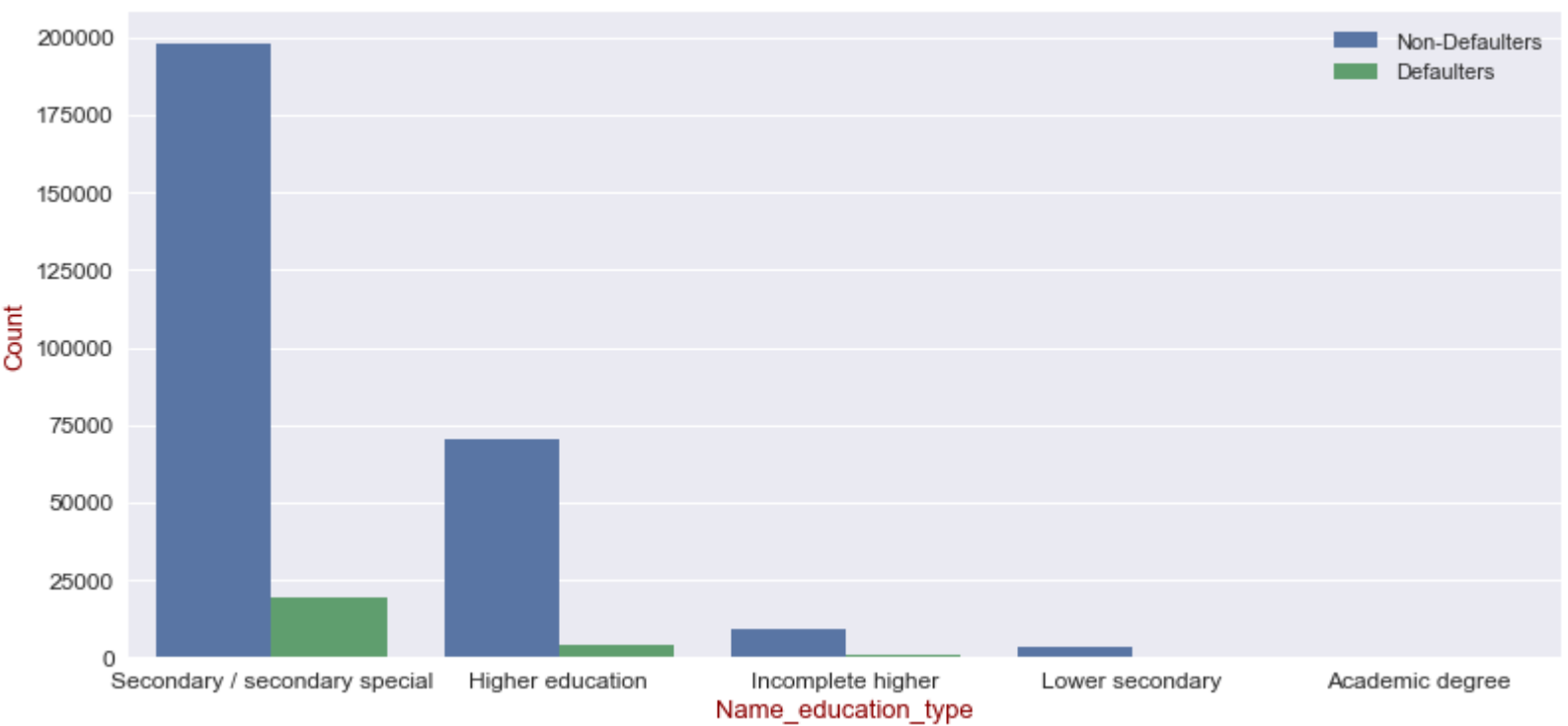
CODE_GENDER		Default_Percentage
0	M	10.17
1	F	7.00
2	XNA	0.00

OCCUPATION

Laborers are more likely to default whereas IT and HR staff are more likely to repay the loan.



Education Type



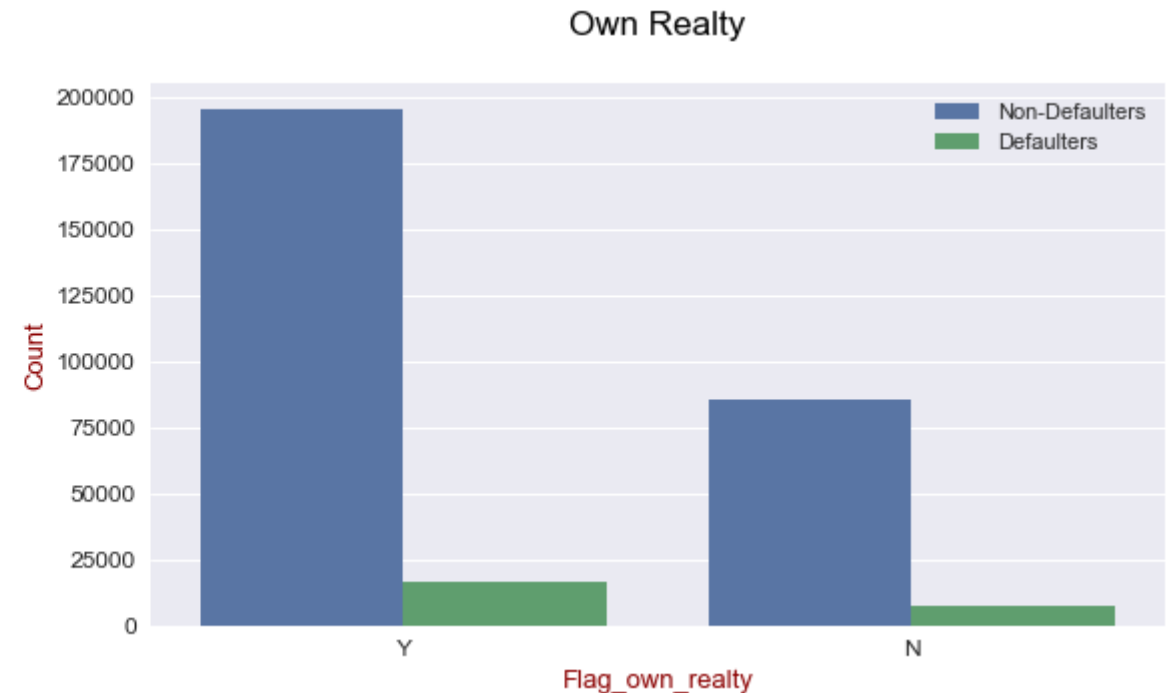
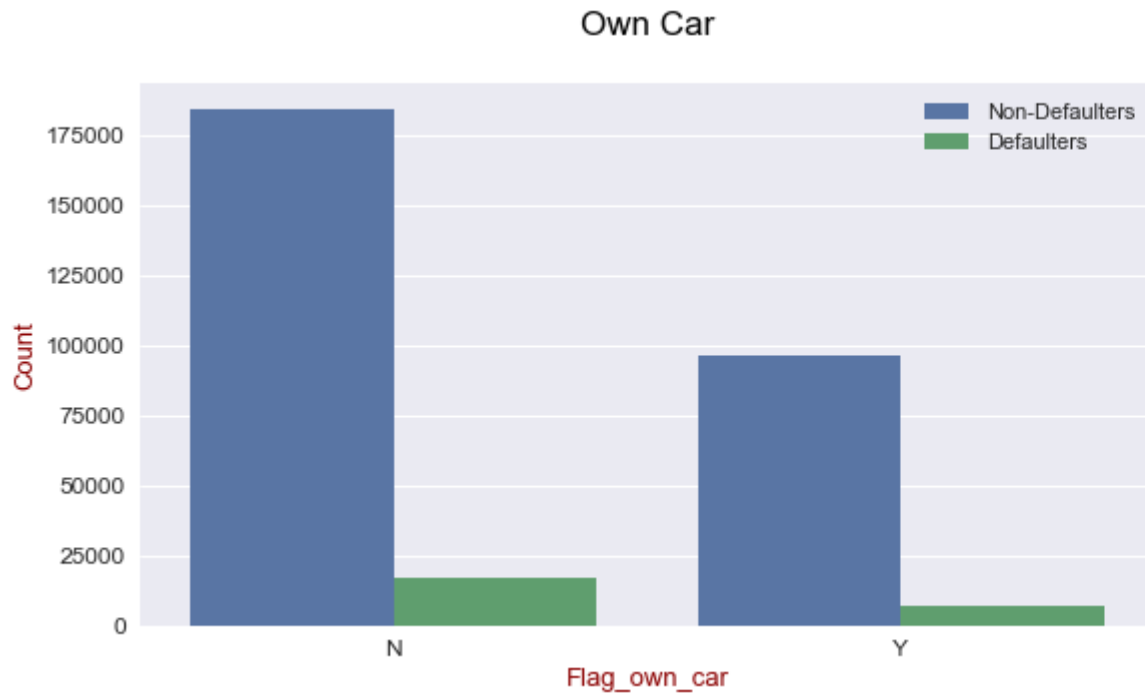
EDUCATION

Secondary education type are likely to default as per total population whereas applicants with lower secondary type of education seems to be at risk of default on same count percentage basis.

	NAME_EDUCATION_TYPE	Default_Percentage
3	Lower secondary	10.95
0	Secondary / secondary special	8.95
2	Incomplete higher	8.50
1	Higher education	5.36
4	Academic degree	1.84

ASSETS

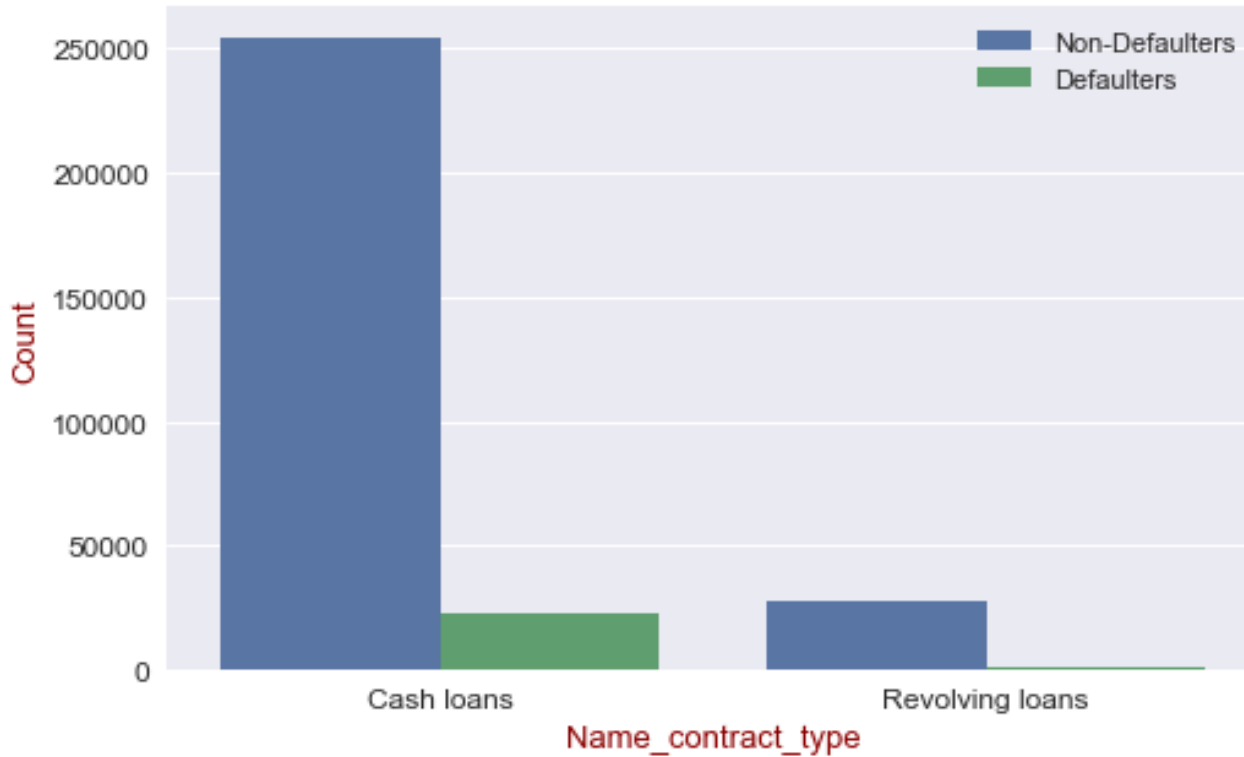
- ✓ Applicants who doesn't own car are likely to default.
- ✓ Applicants who have own realty are likely to default same with a variation of less than 1% with same count analysis.



LOAN TYPE

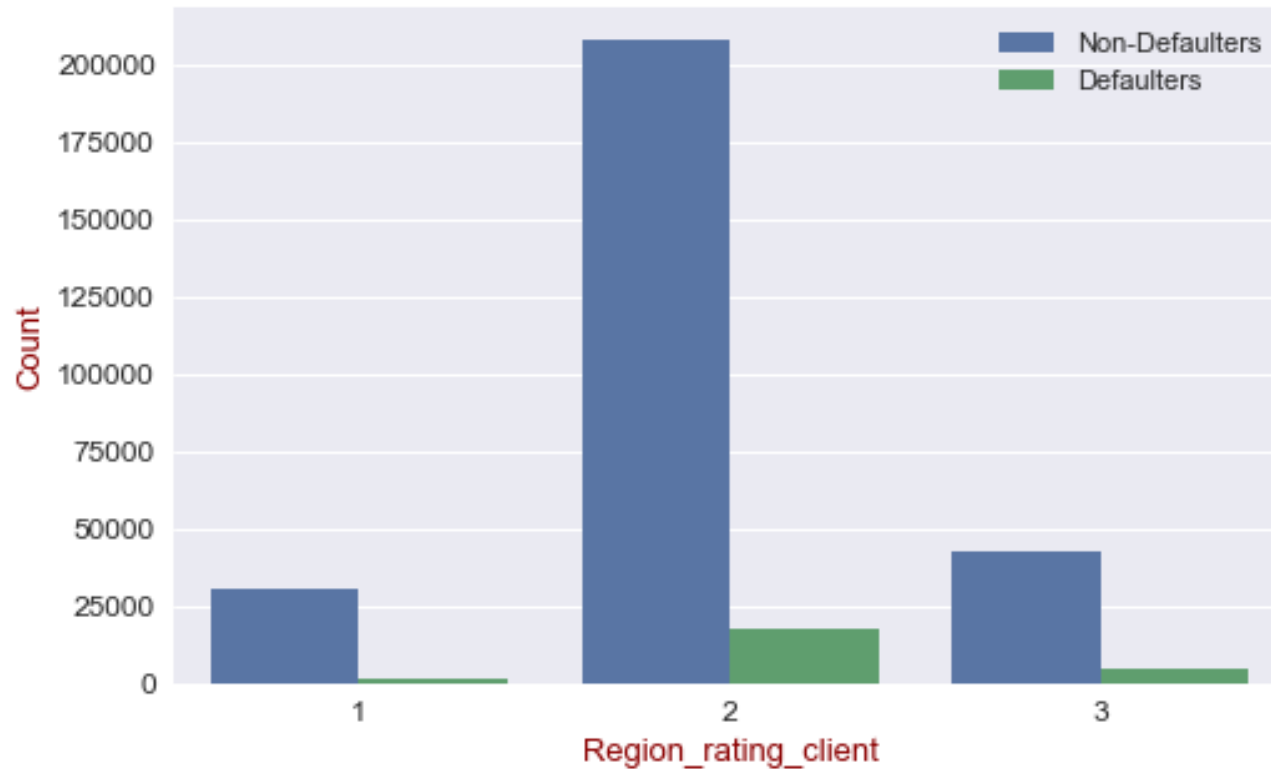
Its the cash loan which is noticed with more likely to default in both comparison.

Loan Type



	NAME_CONTRACT_TYPE	Default_Percentage
0	Cash loans	8.36
1	Revolving loans	5.46

Client Region Rating



REGION_RATING_CLIENT		Default_Percentage
2	3.0	11.13
0	2.0	7.90
1	1.0	4.83

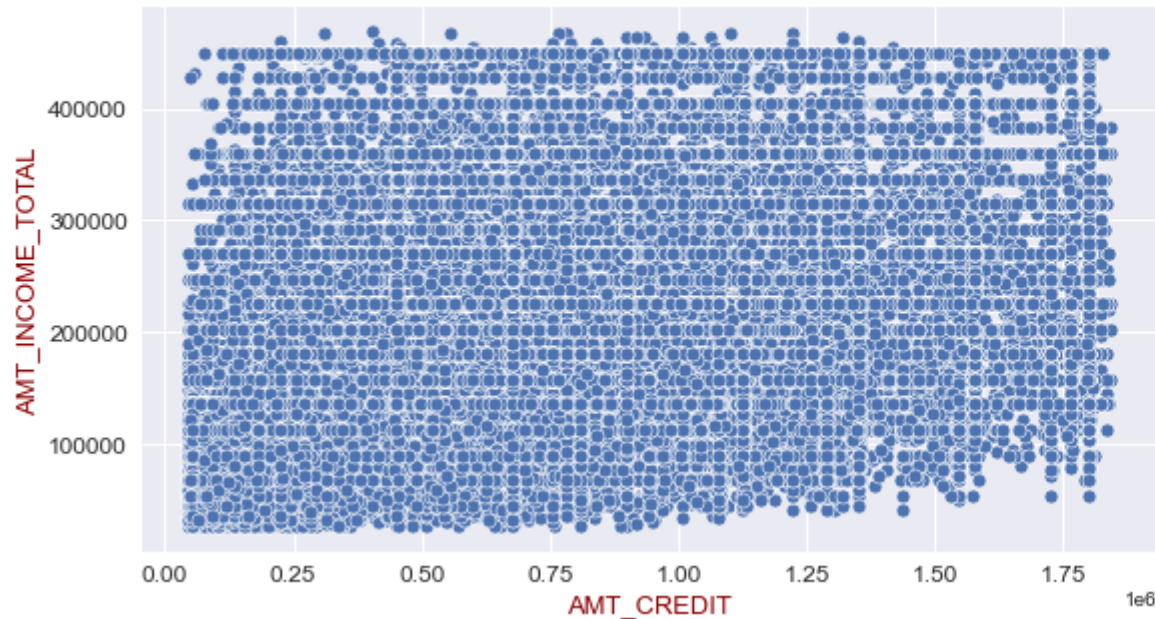
REGION RATING

- ✓ Applicants with region rating 2 are more in number compared to other region ratings.
- ✓ Rating 3 holds highest defaulters in default percentage of same count.

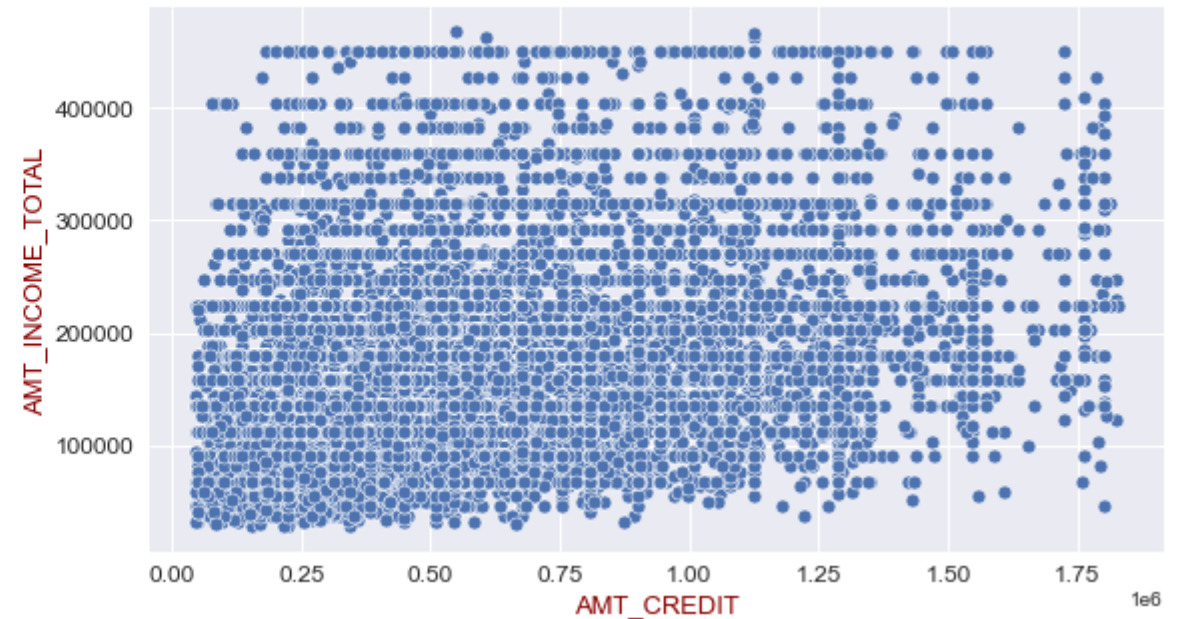
CREDIT AMOUNT Vs TOTAL INCOME

- ✓ Number of defaulters are more even for lesser amount of credit
- ✓ Defaulters are less for higher total income

Total Income Vs Credit - Non-Defaulters



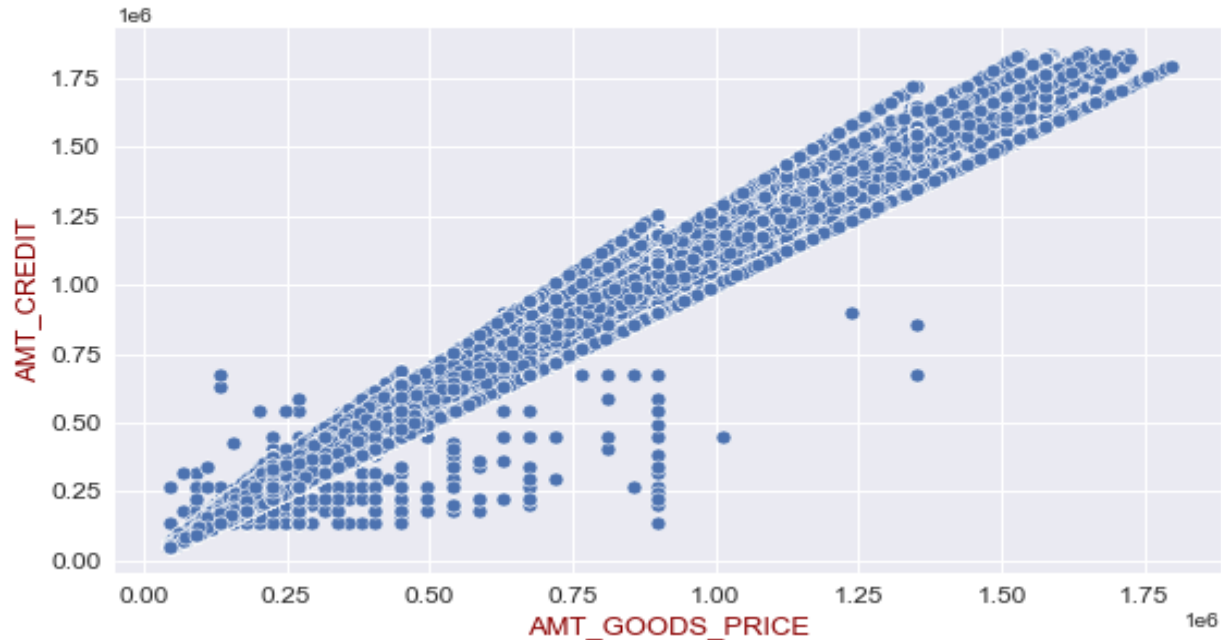
Total Income Vs Credit - Defaulters



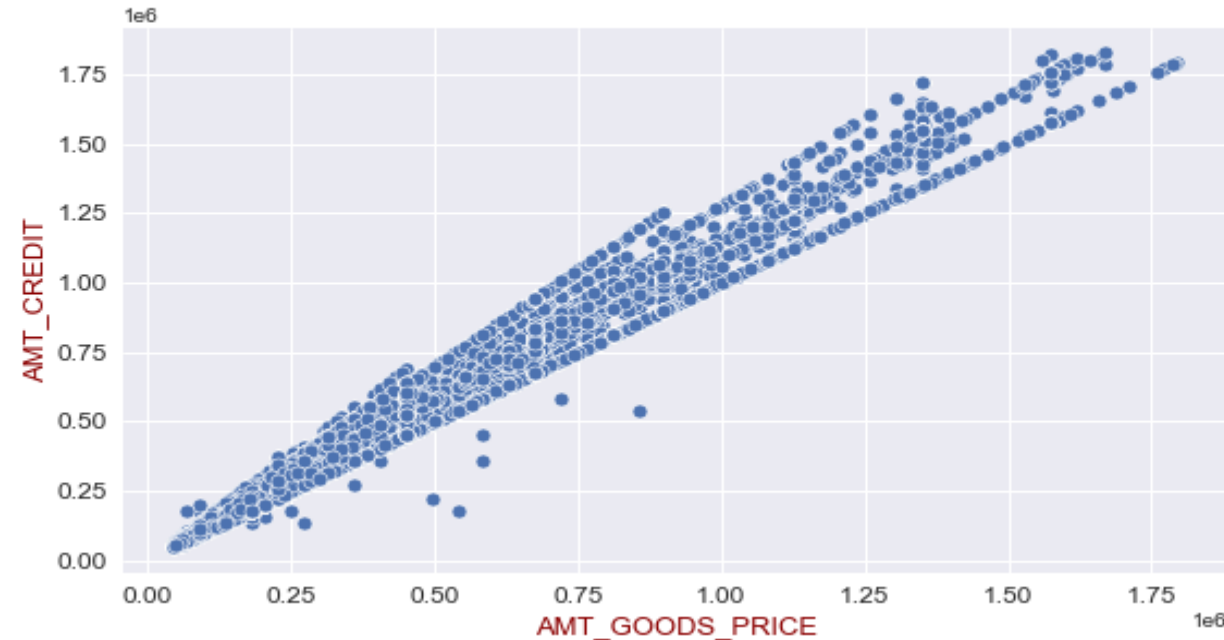
CREDIT AMOUNT Vs GOODS PRICE

- ✓ With above plots, its clear that there is linear positive co-relation between Good price and Credit Amount.
- ✓ Number of defaulters are less with the rise in credit amount as well as goods price.
- ✓ With respect to non-defaulters, the count is more with rise in goods price.

Credit Amount vs Goods Price
Non-Defaulters



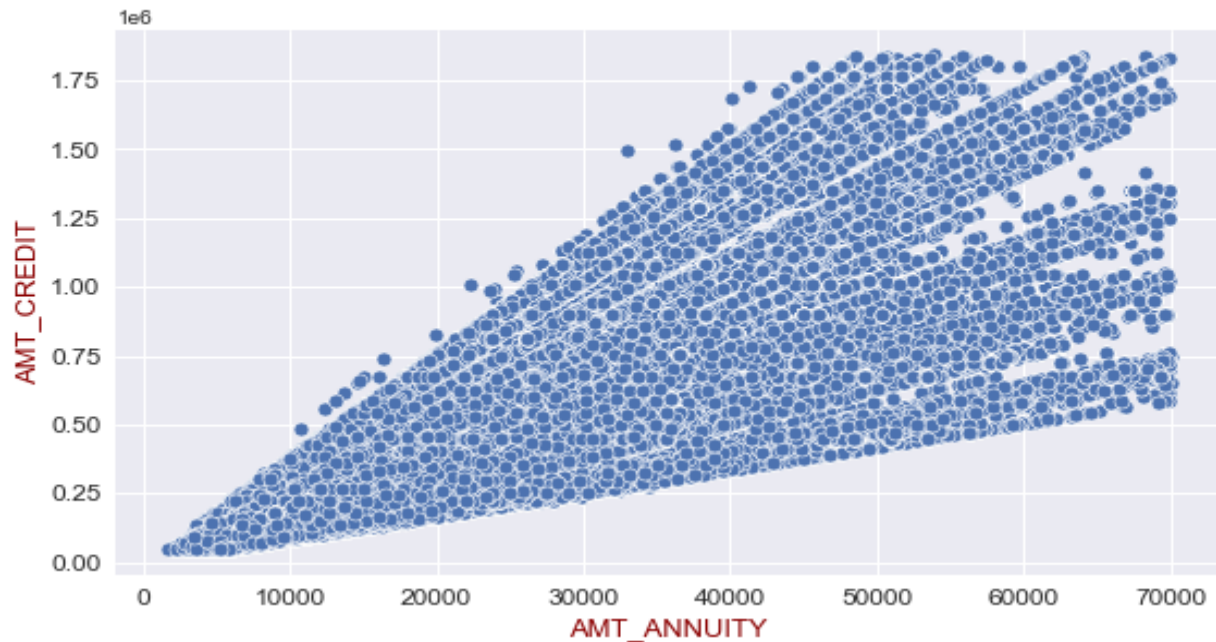
Credit Amount vs Goods Price
Defaulters



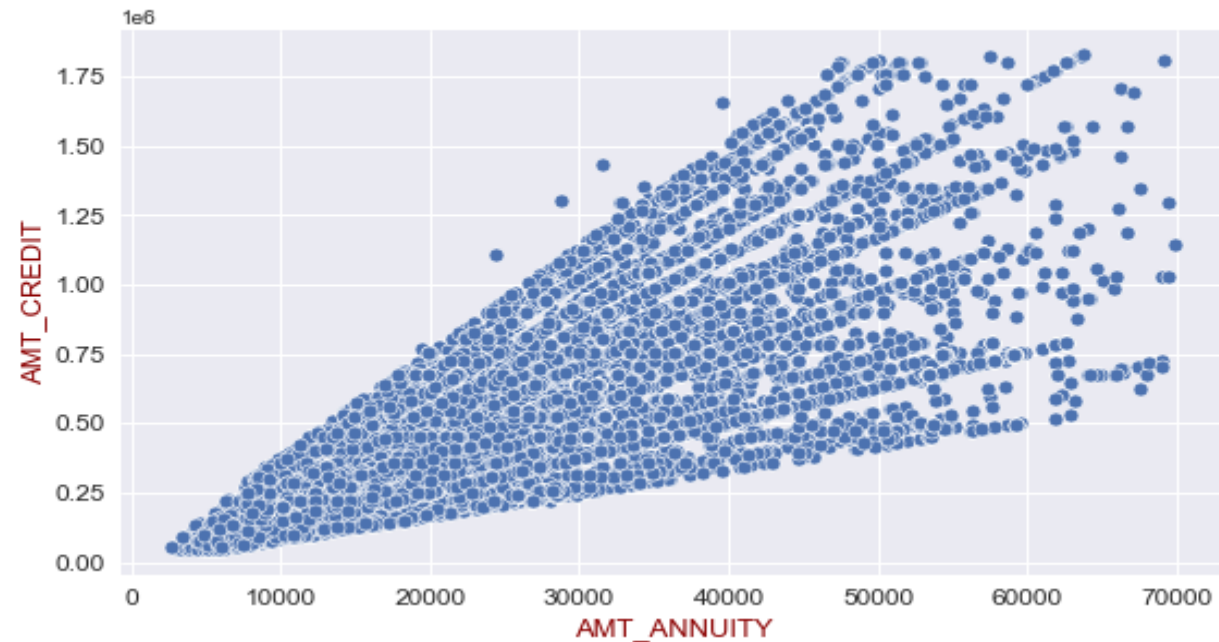
CREDIT AMOUNT VS ANNUITY

- ✓ It's clear that there is linear positive co-relation between Annuity and Credit Amount.
- ✓ Number of defaulters are less with the rise in credit amount as well as annuity.
- ✓ With respect to non-defaulters, the count is more with rise in annuity.

Credit Amount vs Annuity
Non-Defaulters



Credit Amount vs Annuity
Defaulters



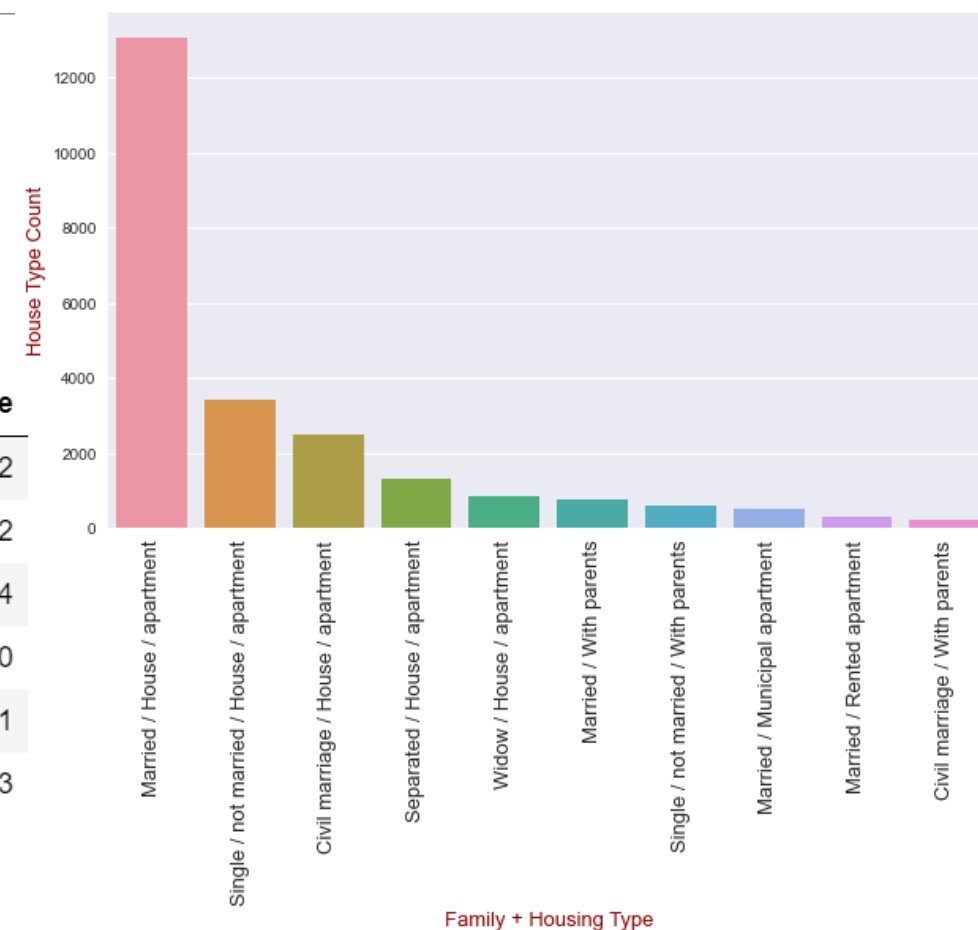
HOUSE TYPE AND STATUS

- ✓ Married and residing in house / apartment are highest likely to default high.
- ✓ Civil marriages are highest amongst more likely to default as per the count of same type.
- ✓ House / apartment are noticed less likely to default when considered individually.

	NAME_FAMILY_STATUS	Default_Percentage
2	Civil marriage	9.96
0	Single / not married	9.84
4	Separated	8.21
1	Married	7.57
3	Widow	5.84

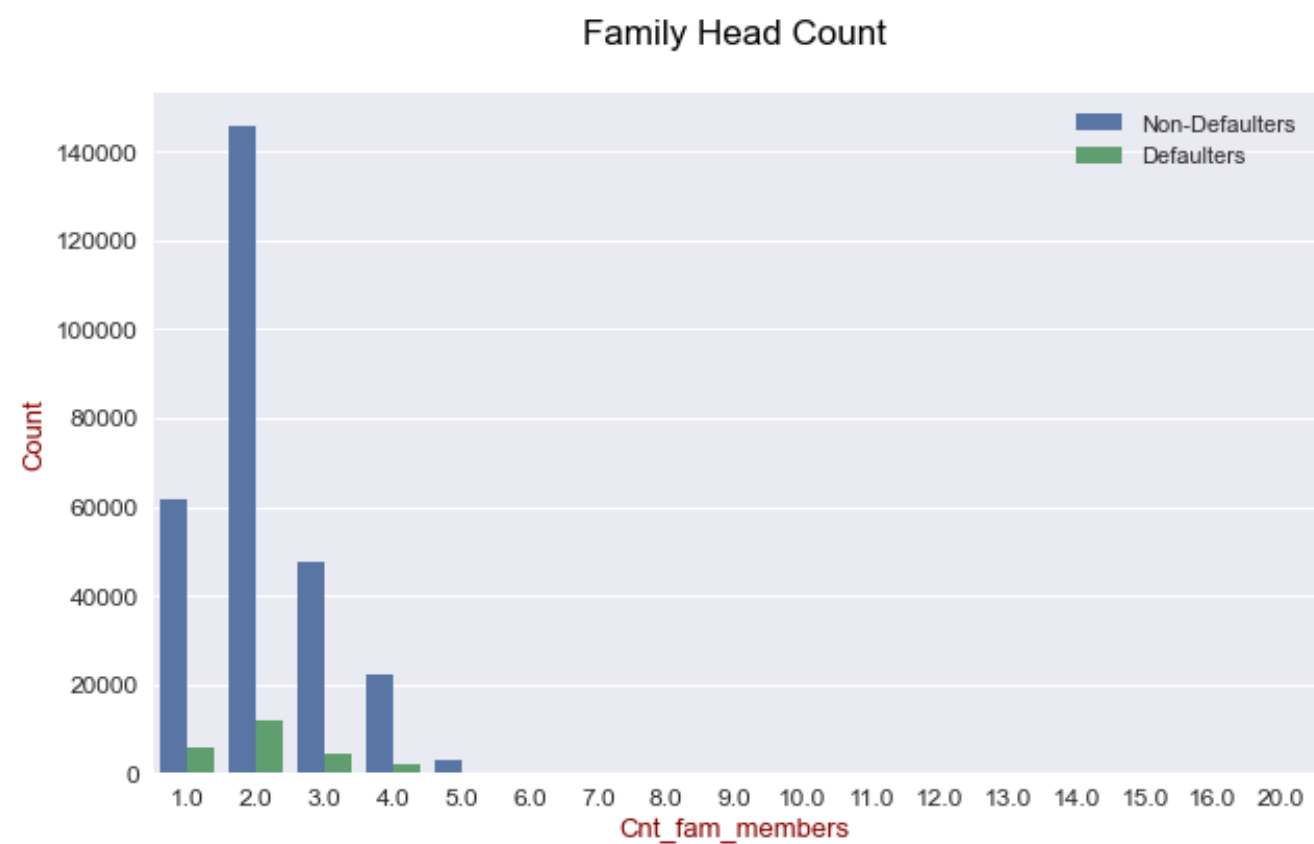
	NAME_HOUSING_TYPE	Default_Percentage
1	Rented apartment	12.32
2	With parents	11.72
3	Municipal apartment	8.54
5	Co-op apartment	8.00
0	House / apartment	7.81
4	Office apartment	6.53

Top 10 Defaulters amongst statuses and housing types combination



FAMILY HEAD COUNT

Families with more than 5 members are more likely to default.

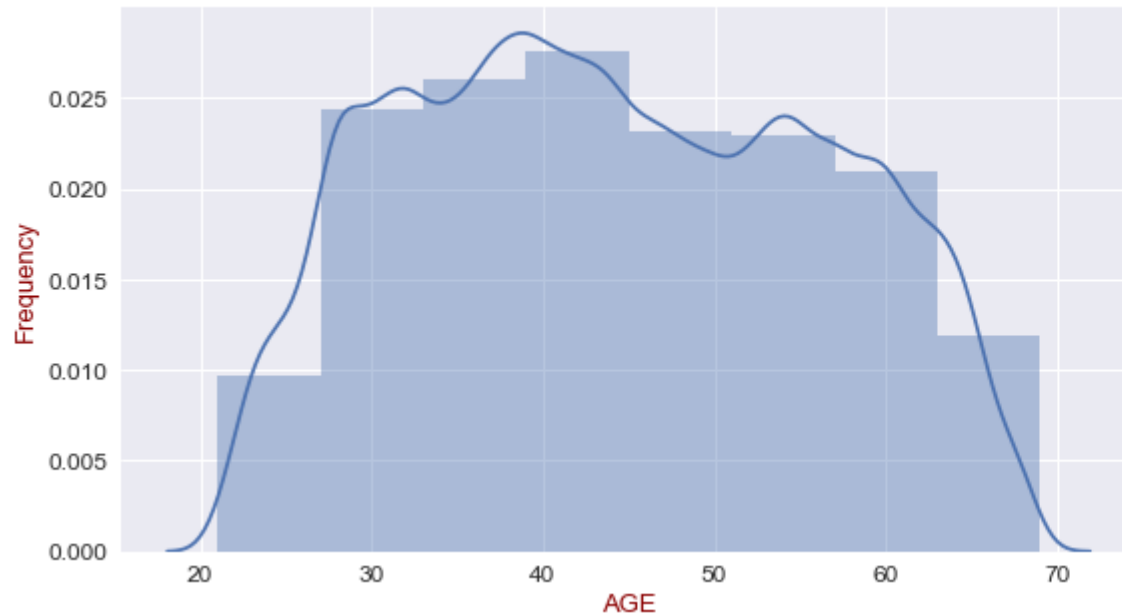


	CNT_FAM_MEMBERS	Default_Percentage
16	11.0	100.00
10	13.0	100.00
9	10.0	33.33
8	8.0	30.00
5	6.0	13.51

AGE DISTRIBUTION

- ✓ Applicants between age grouped within 30 - 40 are more likely to default.
- ✓ Applicants' trend decreases with increase in age for defaulters.

Non-Defaulters Age Distribution



Defaulters Age Distribution

