

Credit EDA Case study

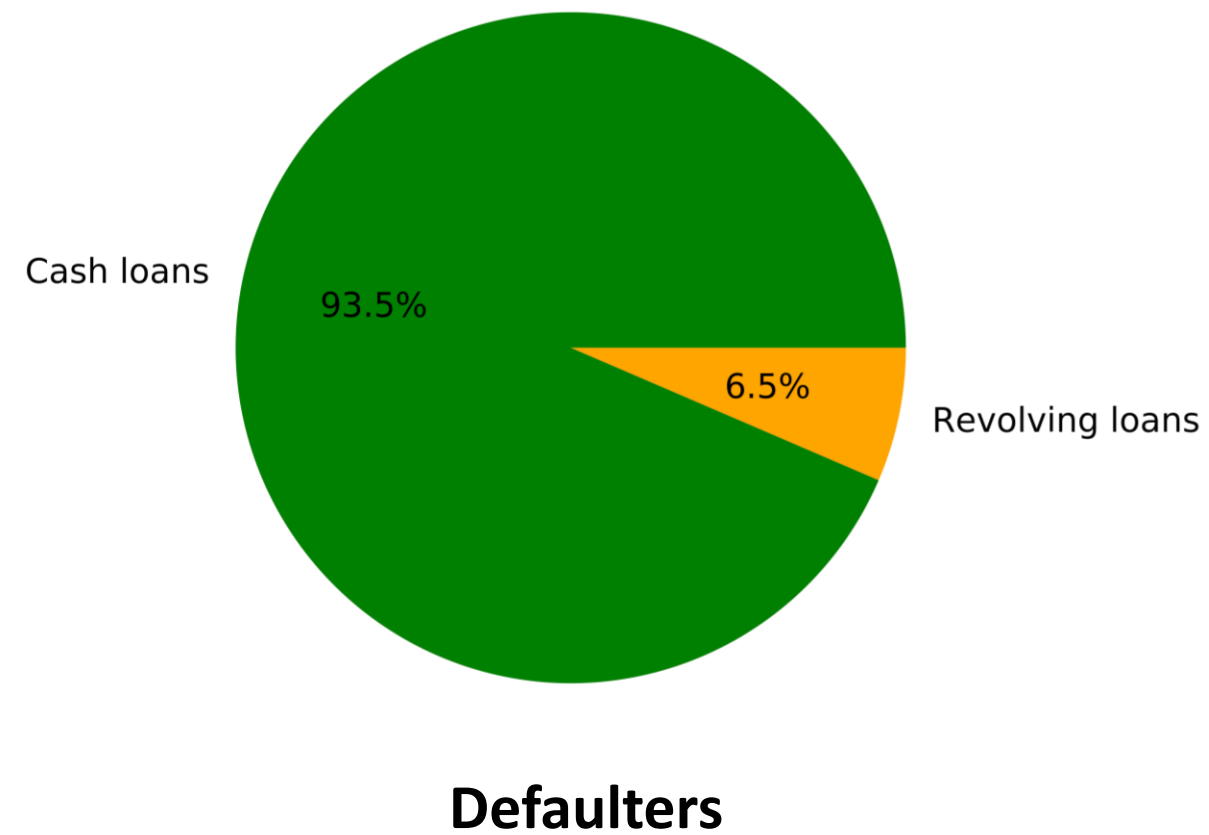
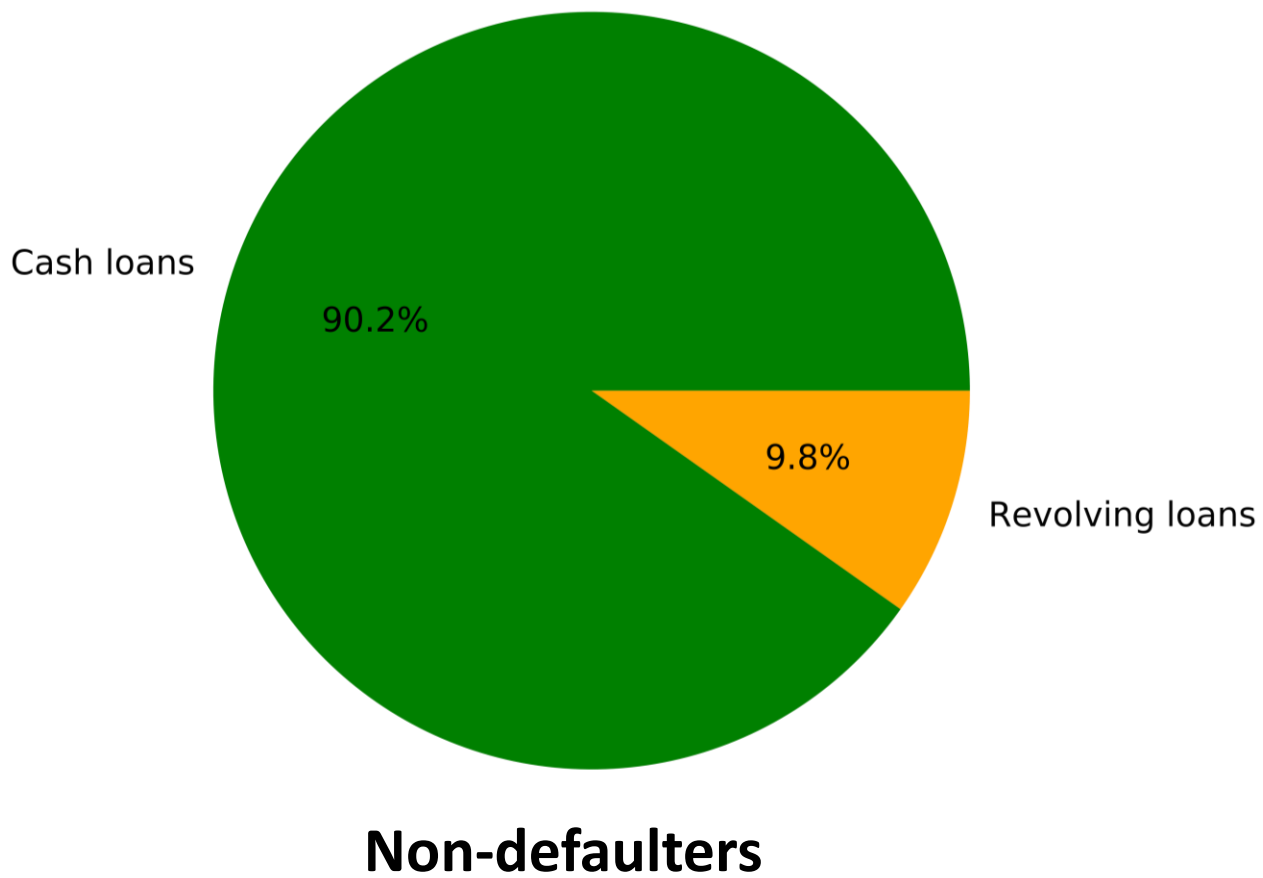
- ***Students :***
 - ***Member 1:***
 - ***Student Name:*** *Hafizullah Mahmudi*
 - ***Student Id:*** *APFE19M00734*
 - ***IIIT Role Number:*** *DDS1950112*
 - ***Member 1:***
 - ***Student Name:*** *Nageswar Reddy*
 - ***Student Id:*** *DDS1950177*
 - ***Role Number:*** *APFE19805553*
- ***16th December, 2019***

Outline

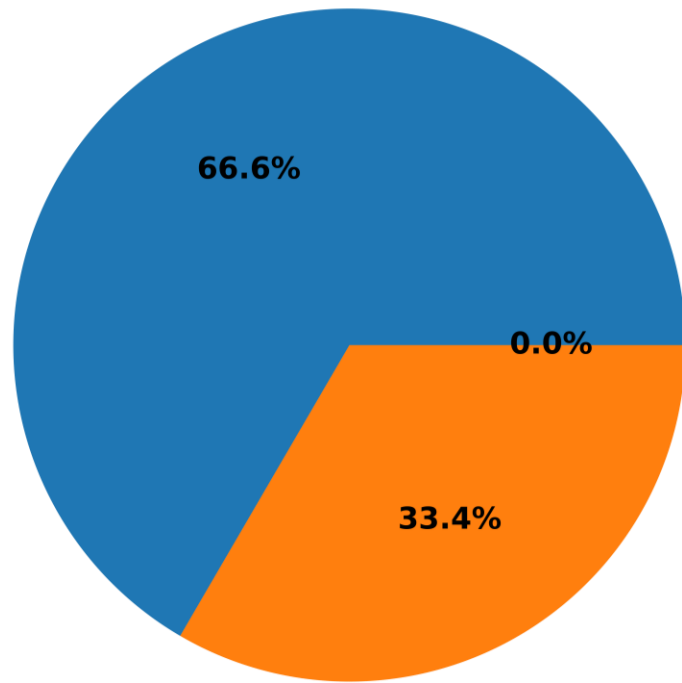
1. Analysis of Current Loan Applications
2. Analysis of Previous Loan Applications
3. Recommendations

Analysis of Current Loan Applications

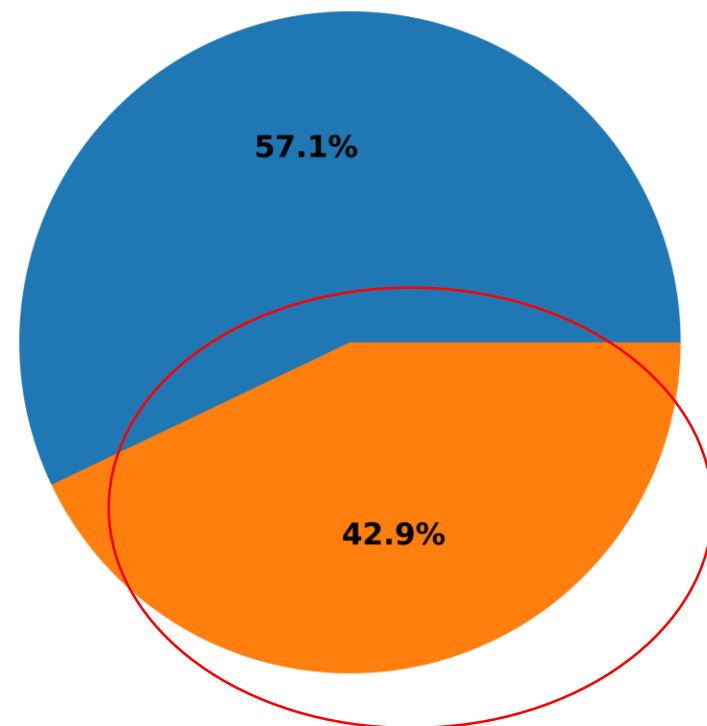
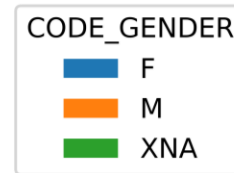
Effect of Contract Type to defaulters



Males are more prone to default than females

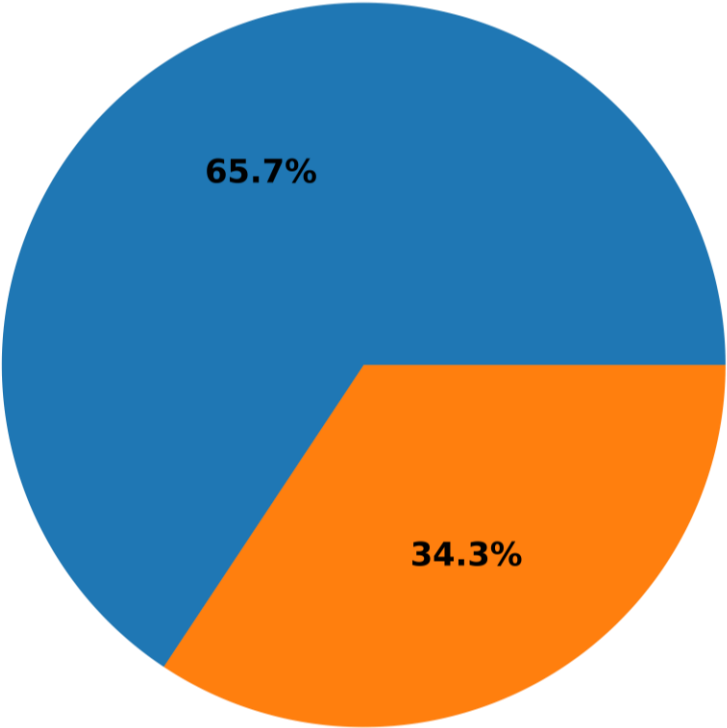


Non-defaulters

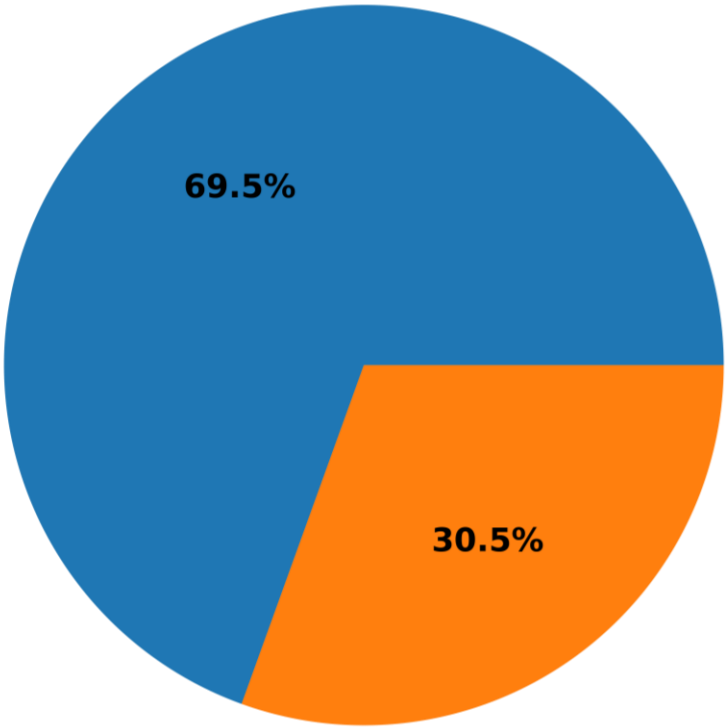


Defaulters

Car owners are more to default than not

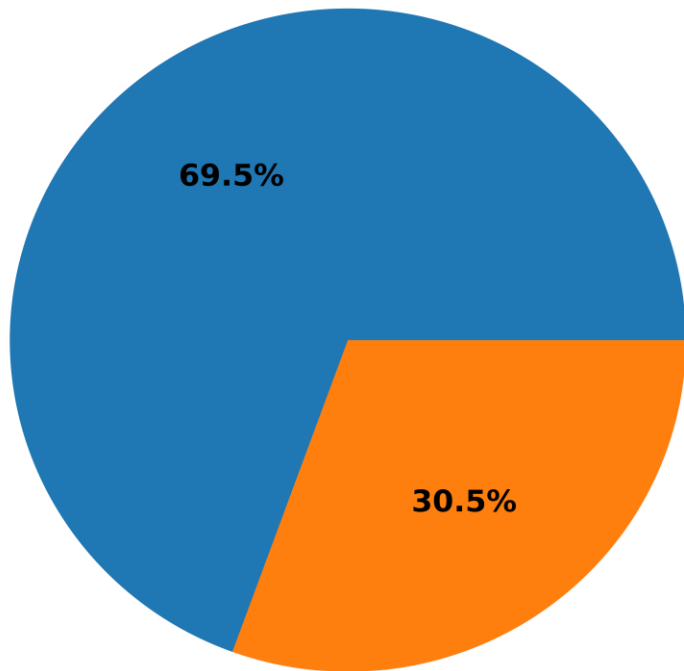


Non-defaulters

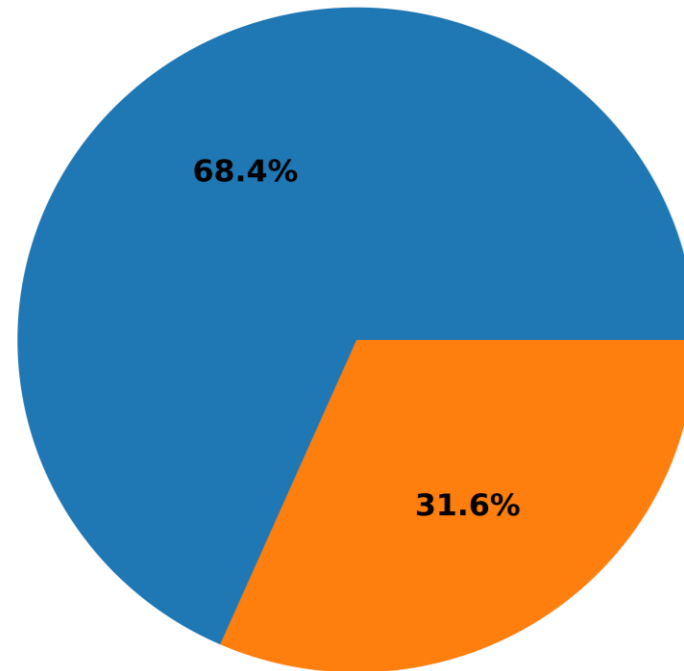


Defaulters

Reality owning in Non-defaulters vs defaulters

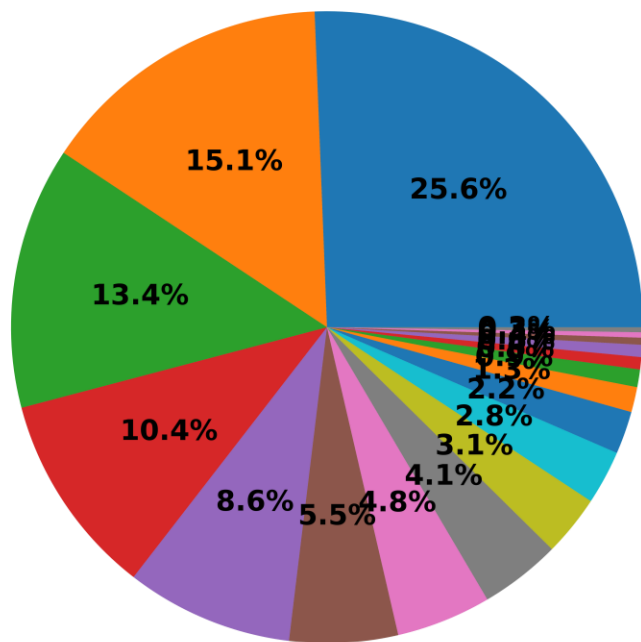


Non-defaulters

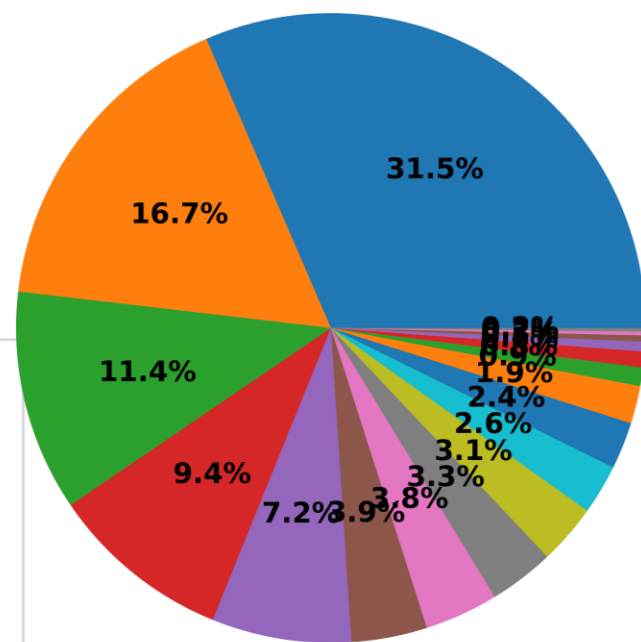


Defaulters

Laborers and sales staff are more probable to default



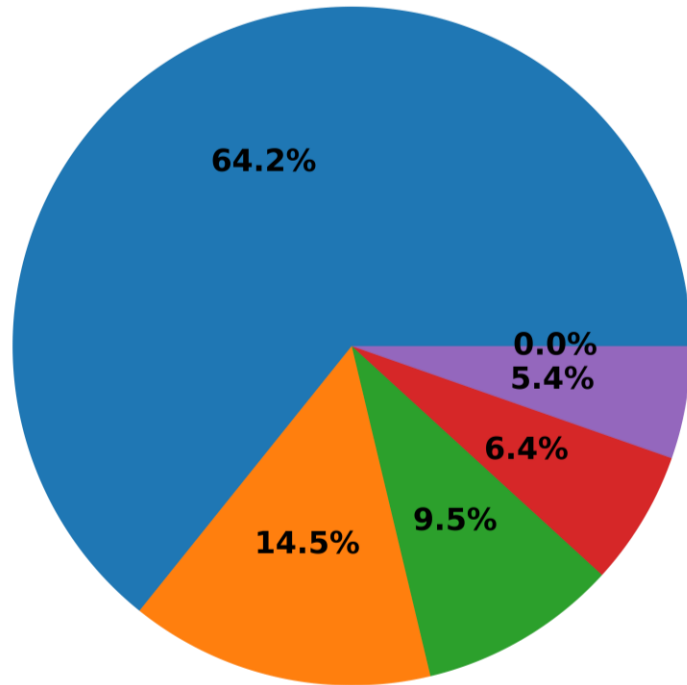
Non-defaulters



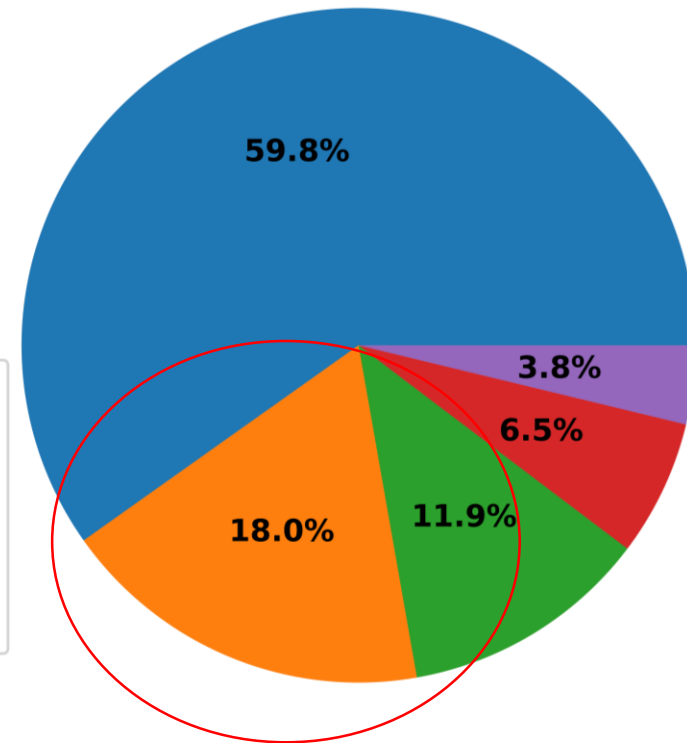
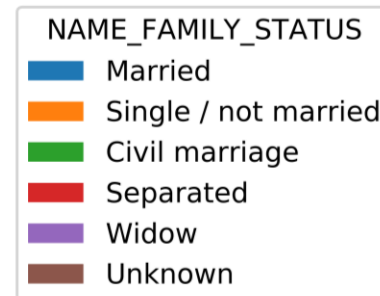
Defaulters



Non-married loan applicants have higher chance to default

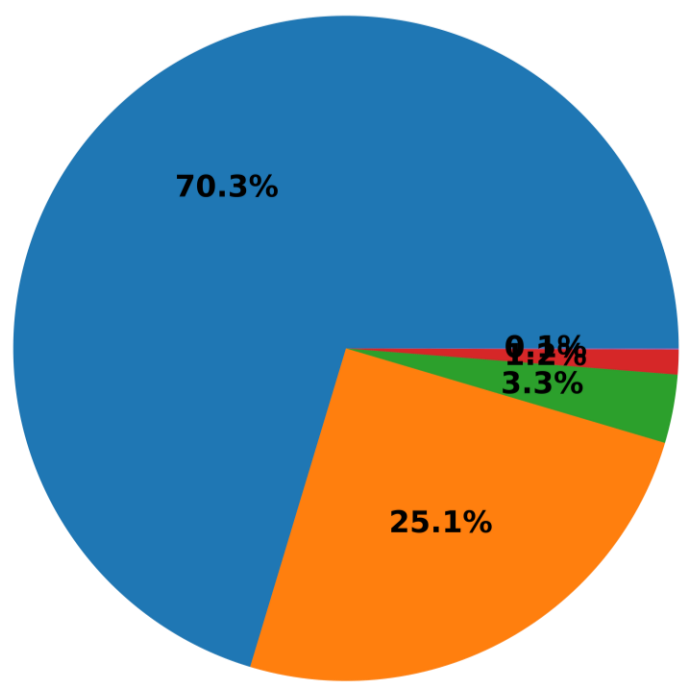


Non-defaulters

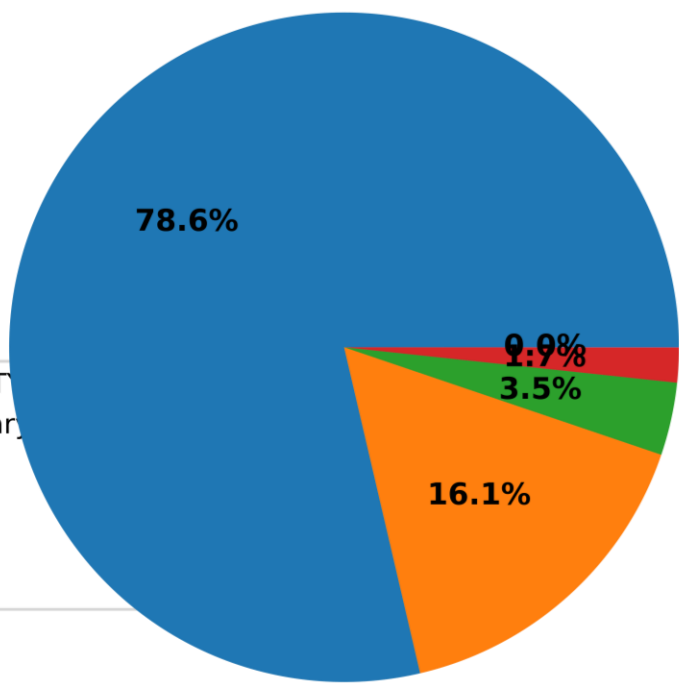
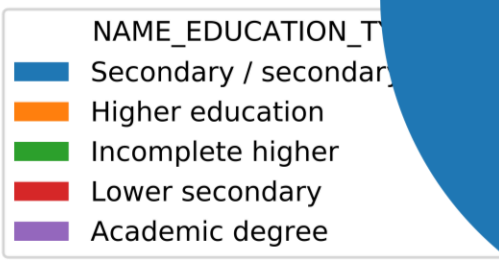


Defaulters

Applicants with secondary education makes high percentage of defaulters

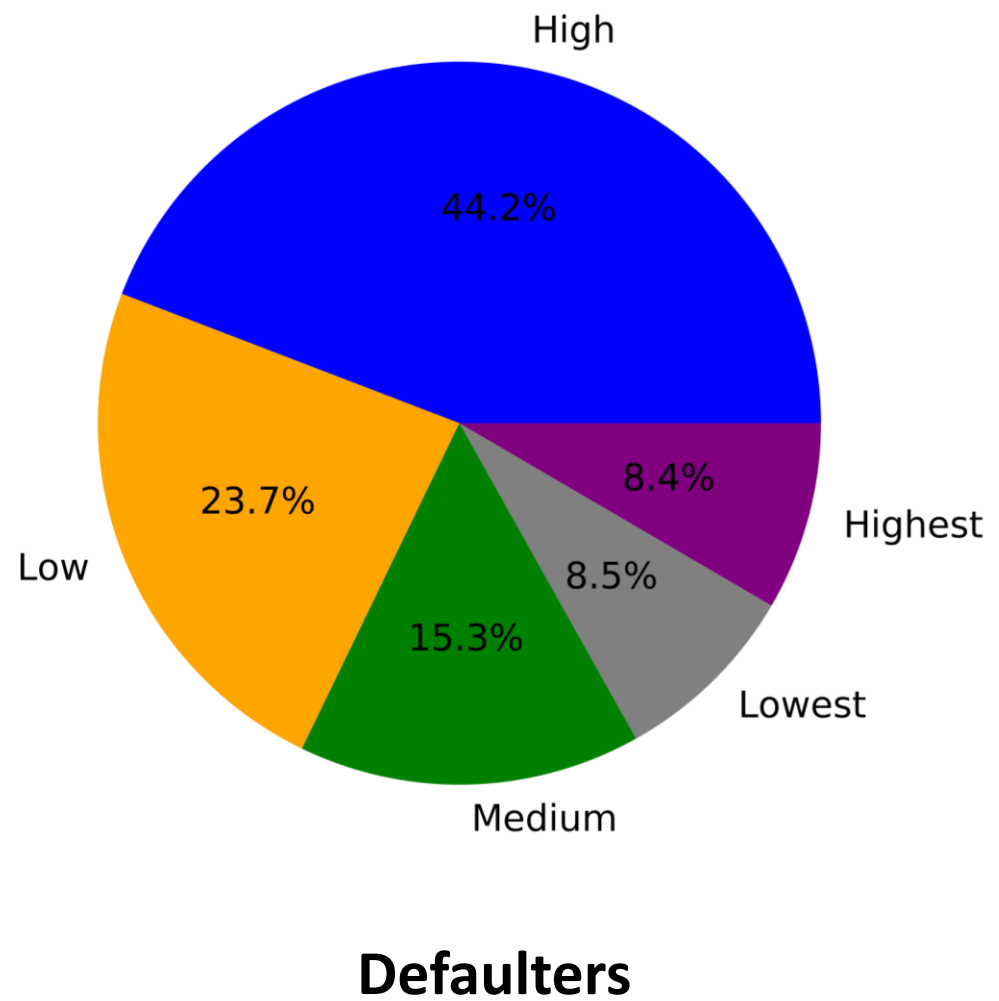
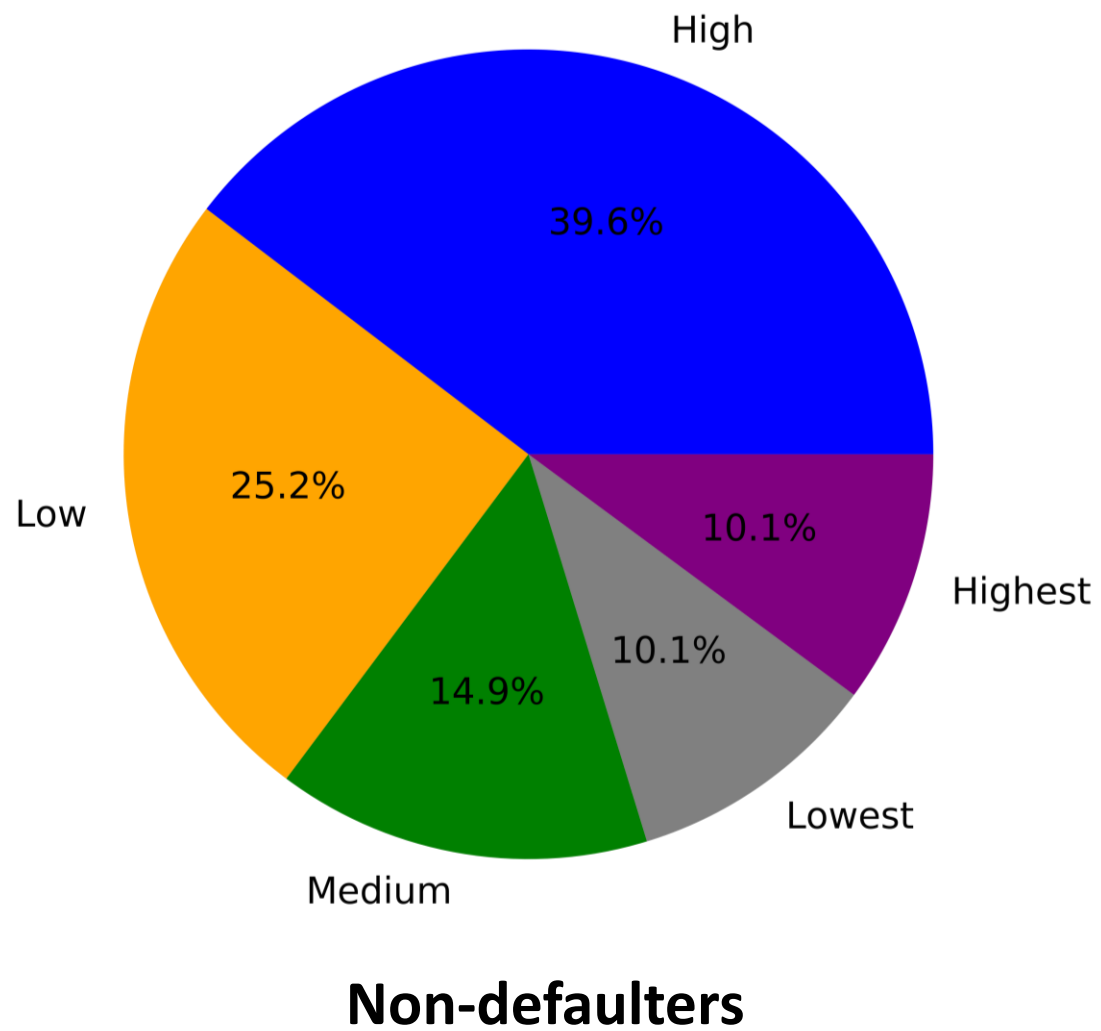


Non-defaulters



Defaulters

About 44% of defaulters have high annuity value that lies between 50th to 90th percentile



Variables with higher factor of affecting loan defaults - 1

Variable	Category of Variable	Non-Defaulters	Defaulters	Observations
Contract type	Cash Loans	90.2 %	93.5%	Cash Loans are far much higher than Revolving Loans for both defaulters and non-defaulters
	Revolving Loans	9.8%	6.5%	
Gender of the client	Male	33.4%	57.1%	Female applicants are receiving more loan than male applicants. But percentage of male defaulter applicants are more than percentage of male in non-defaulters.
	Female	66.6%	42.9%	
Client owns a car/do not own a car	Owens a car	65.7 %	69.5%	Proportion of car holders are similar in both defaulter and non-defaulters applications. However, the number of applications owning a car is comparatively less than non-defaulters.
	Do not own a car	34.3%	30.5%	
Owns Home/Flat	Owens Home/ Flat	65.7 %	69.5%	Proportion of reality owners is almost equal in both defaulter and non-defaulters applications
	Do not own Home/ Flat	34.3%	30.5%	

Variables with higher factor of affecting loan defaults -2

Variable	Category of Variable	Non-Defaulters	Defaulters	Observations
Housing type	House/Apartment	89%	85.7%	The percentage of housing type with House/apartment in defaulters is less than non-defaulters which makes over 85% of all application in both cases
	With Parents	4.6%	7%	
Occupation type	Laborers	25.6%	31.5%	The laborers, sales staff percentage is more in defaulters than non-defaulters. Conversely, the percentage of Core Staff, and Managers are more in non-defaulters than defaulters
	Sales Staff	15.1%	16.7%	
Family Status	Married	64.2%	59.8%	The single and civil marriage percentage is more in defaulters than non-defaulters. Conversely, the percentage of Married is more in non-defaulters than defaulters.
	Single	14.5%	18%	
Education type	Secondary/Secondary Special	70.3%	78.6%	The Secondary/secondary special percentage is more in defaulters than non-defaulters. But, the percentage of Higher education holders is more in non-defaulters than defaulters.
	Higher Education	25.1%	16.1%	

Variables with higher factor of affecting loan defaults -3

Variable	Non - Defaulters	Defaulters	Observations
Client income	Majority of loans are given between 90000 and 200000.	Majority of loans are given between 90000 and 200000.	In both non-defaulters and defaulters income of the client is same, The majority of loans in both defaulters and non-defaulters between 50th and 75th percentile
Credit amount of the loan	Credit range is 250000 and 750000	Credit range is 200000 to 800000	Here Defaulters are having higher Credit range
Loan Annuity			About 44% of defaulters have high annuity value that lies between 50th to 90th percentile

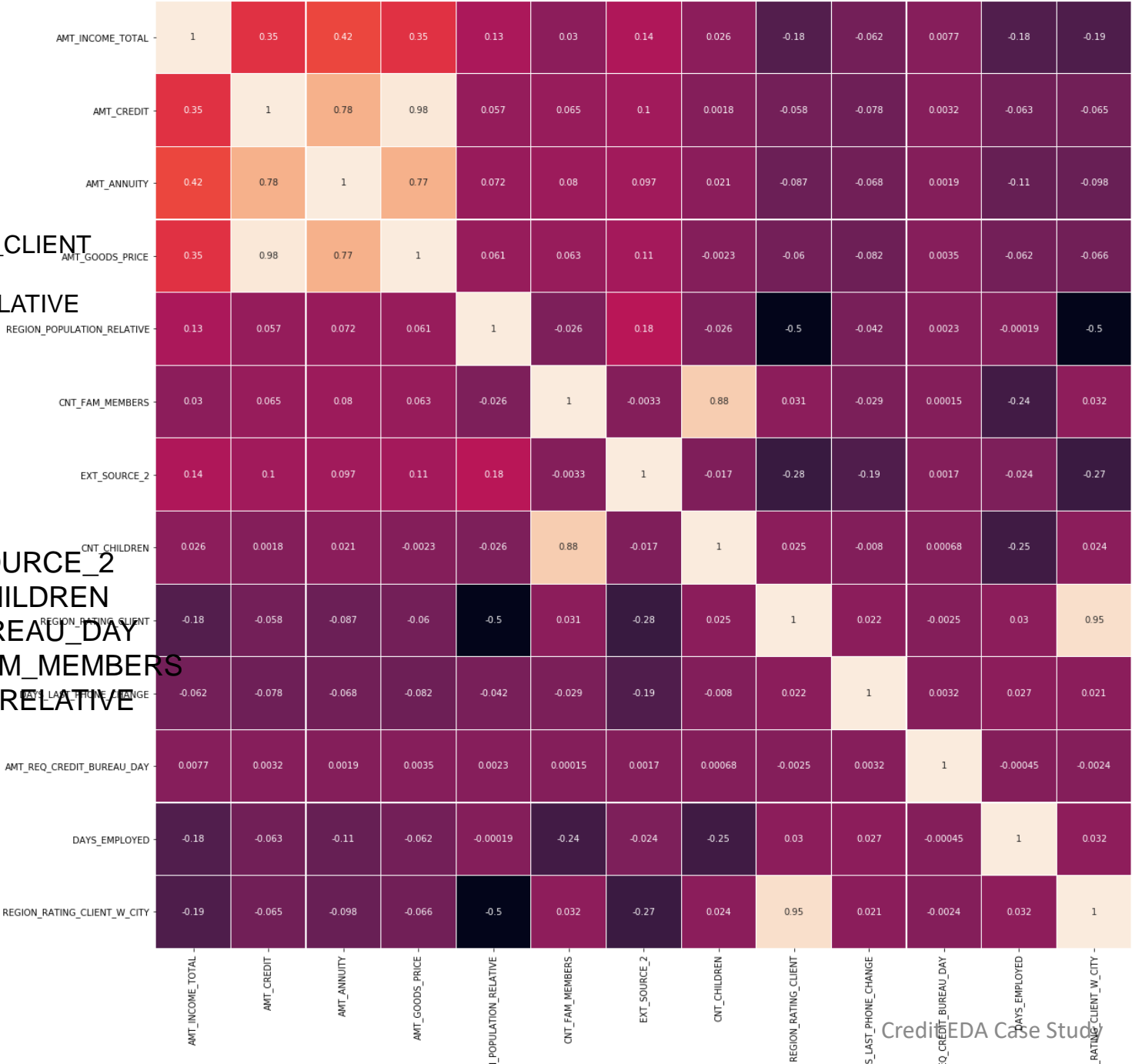
Correlation of variables in non-defaulters

Top 6 positive correlation

- 1.AMT_GOODS_PRICE vs AMT_CREDIT
- 2.REGION_RATING_CLIENT_W_CITY vs REGION_RATING_CLIENT
- 3.CNT_CHILDREN vs CNT_FAM_MEMBERS 0.9
- 4.REGION_RATING_CLIENT vs REGION_POPULATION_RELATIVE
- 5.AMT_ANNUITY vs AMT_CREDIT
- 6.AMT_GOODS_PRICE vs AMT_ANNUITY

Bottom 6 negative correlation

- 1. CNT_CHILDREN vs AMT_CREDIT
- 2. AMT_REQ_CREDIT_BUREAU_DAY vs EXT_SOURCE_2
- 3. AMT_REQ_CREDIT_BUREAU_DAY vs CNT_CHILDREN
- 4. DAYS_EMPLOYED vs AMT_REQ_CREDIT_BUREAU_DAY
- 5. AMT_REQ_CREDIT_BUREAU_DAY vs CNT_FAM_MEMBERS
- 6. DAYS_EMPLOYED vs REGION_POPULATION_RELATIVE



Correlation of variables in defaulters

Top 6 positive correlation

AMT_GOODS_PRICE vs AMT_CREDIT

REGION_RATING_CLIENT_W_CITY vs REGION_POPULATION_RELATIVE

CNT_CHILDREN vs CNT_FAM_MEMBERS

AMT_ANNUITY vs AMT_CREDIT

AMT_GOODS_PRICE vs AMT_ANNUITY

REGION_RATING_CLIENT vs REGION_POPULATION_RELATIVE

Bottom 6 negative correlation

CNT_CHILDREN vs AMT_CREDIT

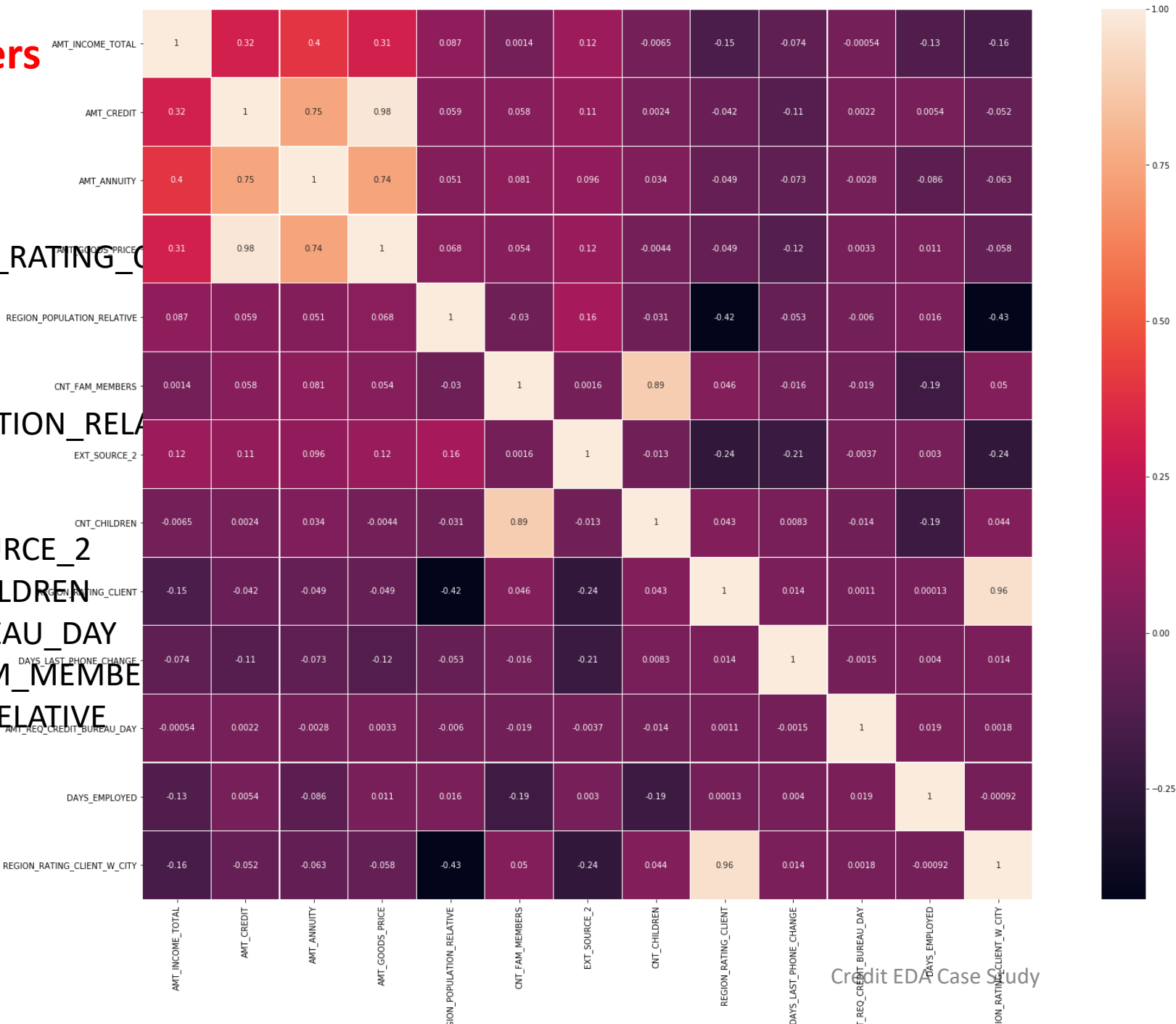
AMT_REQ_CREDIT_BUREAU_DAY vs EXT_SOURCE_2

AMT_REQ_CREDIT_BUREAU_DAY vs CNT_CHILDREN

DAYS_EMPLOYED vs AMT_REQ_CREDIT_BUREAU_DAY

AMT_REQ_CREDIT_BUREAU_DAY vs CNT_FAM_MEMBERS

DAYS_EMPLOYED vs REGION_POPULATION_RELATIVE



Variables with highest correlation between defaulters and non-def

Top 6 variable having positive correlation					
Non - Defaulters			Defaulters		
1.	Price of the Goods	- Credit Amount	1.	Price of the Goods	- Credit Amount
2.	Region Rating by considering City	- Region Rating	2.	Region Rating by considering City	- Region Rating
3.	No of Children	- No of family members	3.	No of Children	- No of family members
4.	Normalized Population of the region	- Region Rating	4.	Normalized Population of the region	- Region Rating
5.	Annuity Amount	- Credit Amount	5.	Annuity Amount	- Credit Amount
6.	Price of Goods	- Annuity Amount	6.	Price of Goods	- Annuity Amount

Observation:
Here we can see that the list of variables having highest positive correlation are same for both Non defaulters and Defaulters

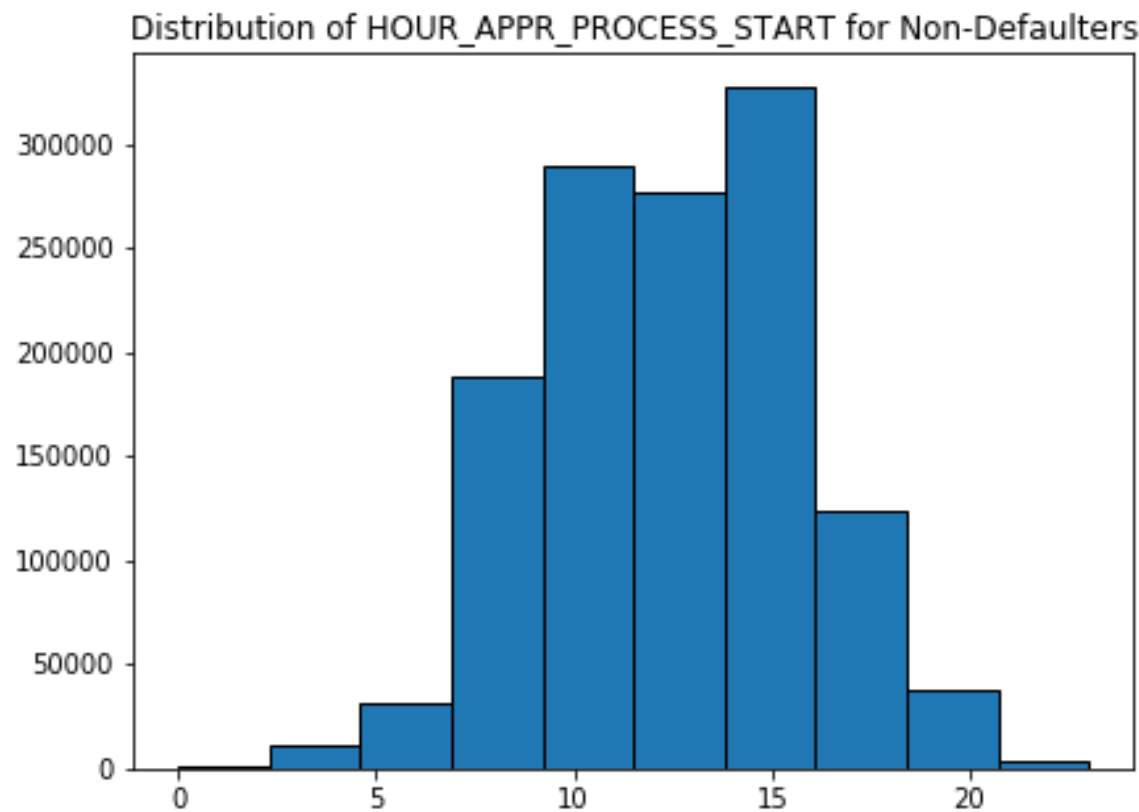
Variables with highest correlation between defaulters and non-def

Top 6 variable having negative correlation			
Non - Defaulters		Defaulters	
1. Price of the Goods	- Credit Amount	1. Price of the Goods	- Credit Amount
2. Region Rating by considering City	- Region Rating	2. Region Rating by considering City	- Region Rating
3. No of Children	- No of family members	3. No of Children	- No of family members
4. Normalized Population of the region	- Region Rating	4. Normalized Population of the region	- Region Rating
5. Annuity Amount	- Credit Amount	5. Annuity Amount	- Credit Amount
6. Price of Goods	- Annuity Amount	6. Price of Goods	- Annuity Amount

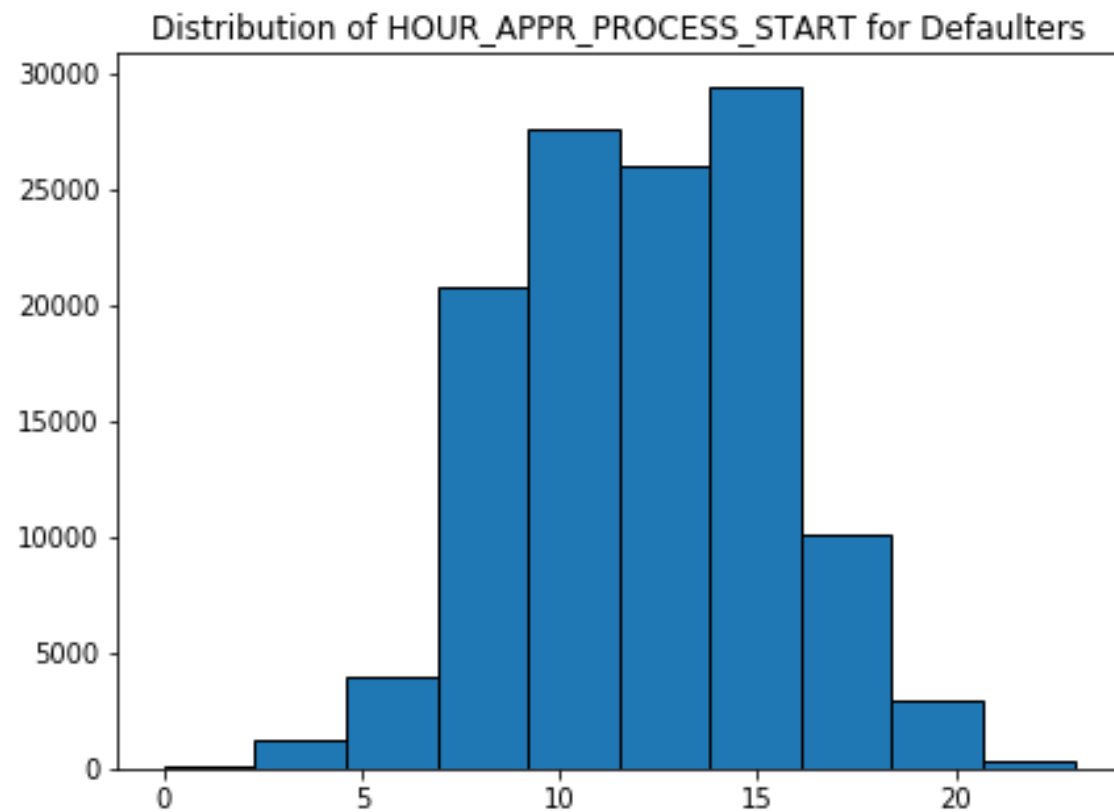
Observation:
Here we can see that the list of variables having highest positive correlation are same for both Non defaulters and Defaulters

Analysis of Previous Loans

Majority of applications are processed during 10 AM to 4PM

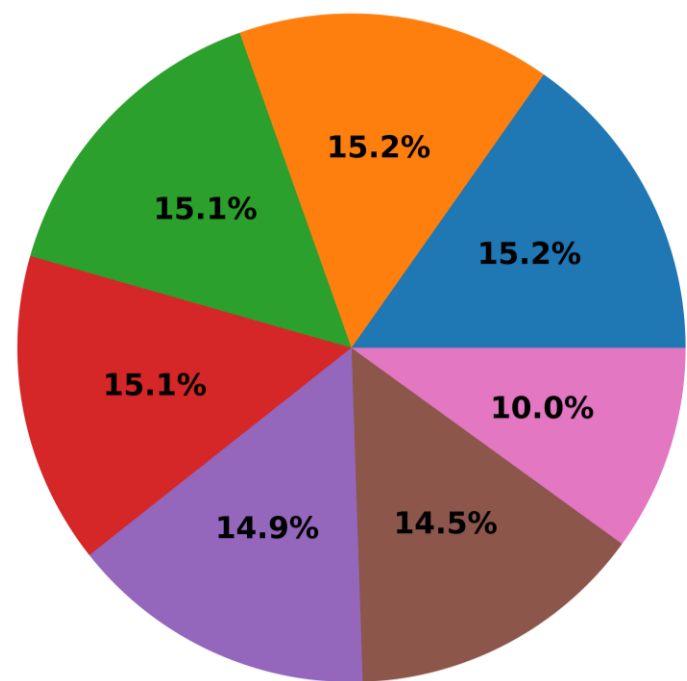


Non-defaulters

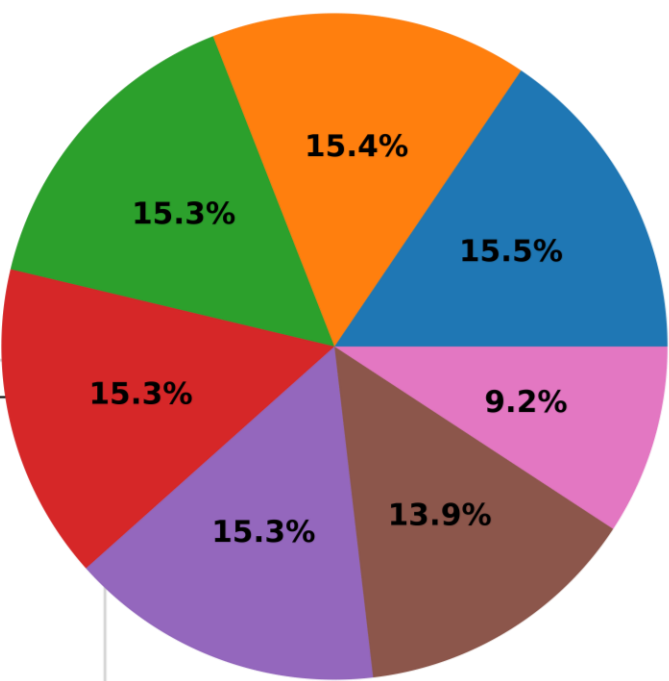
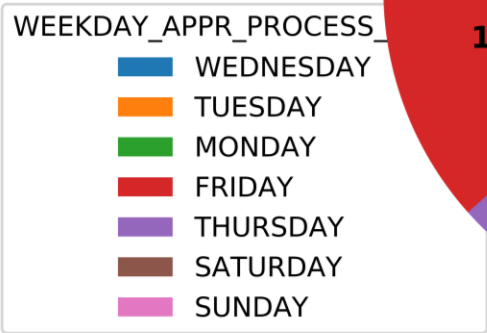


Defaulters

Most applications are processed on Fridays, Mondays, Tuesdays and Wednesdays

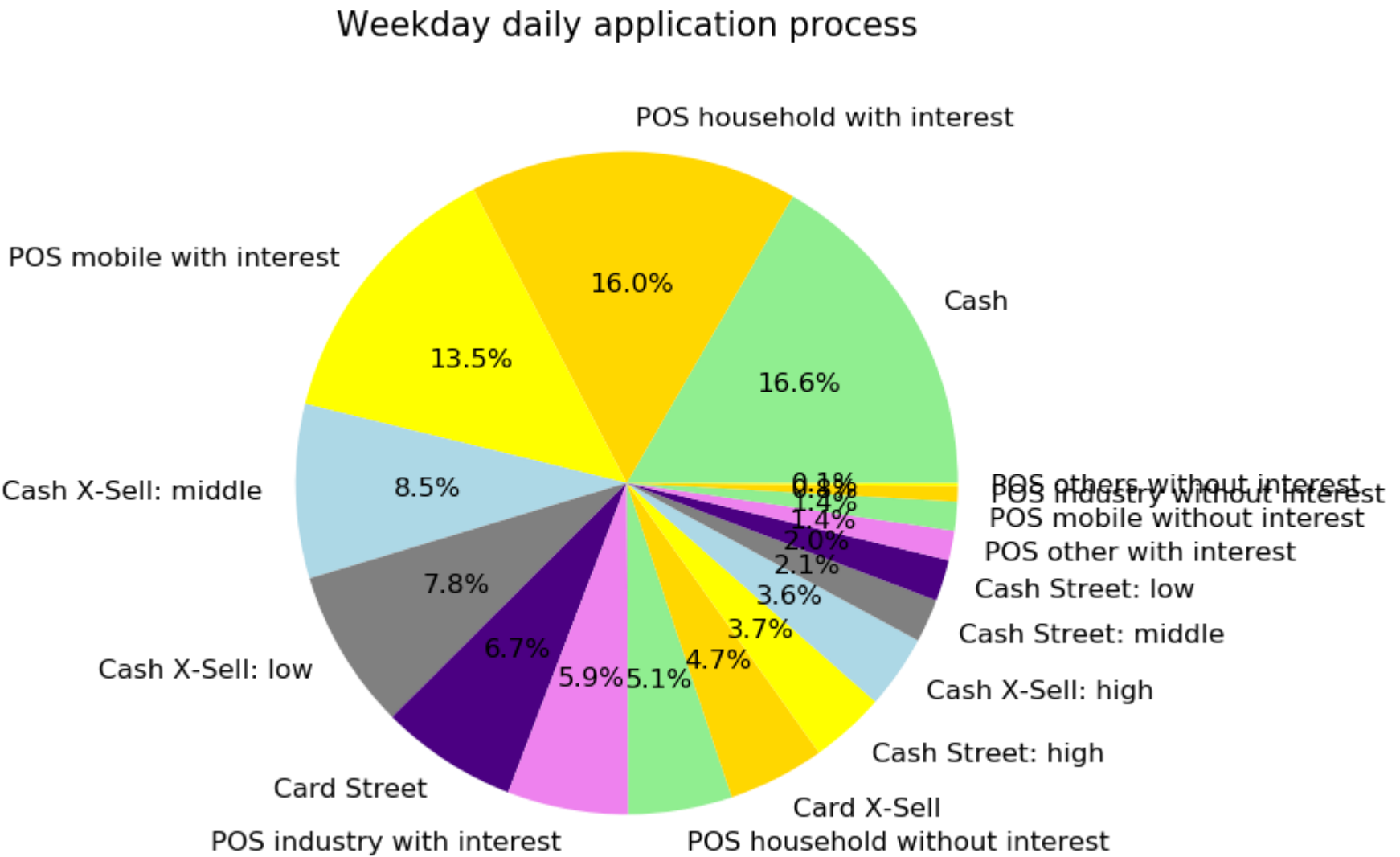


Non-defaulters



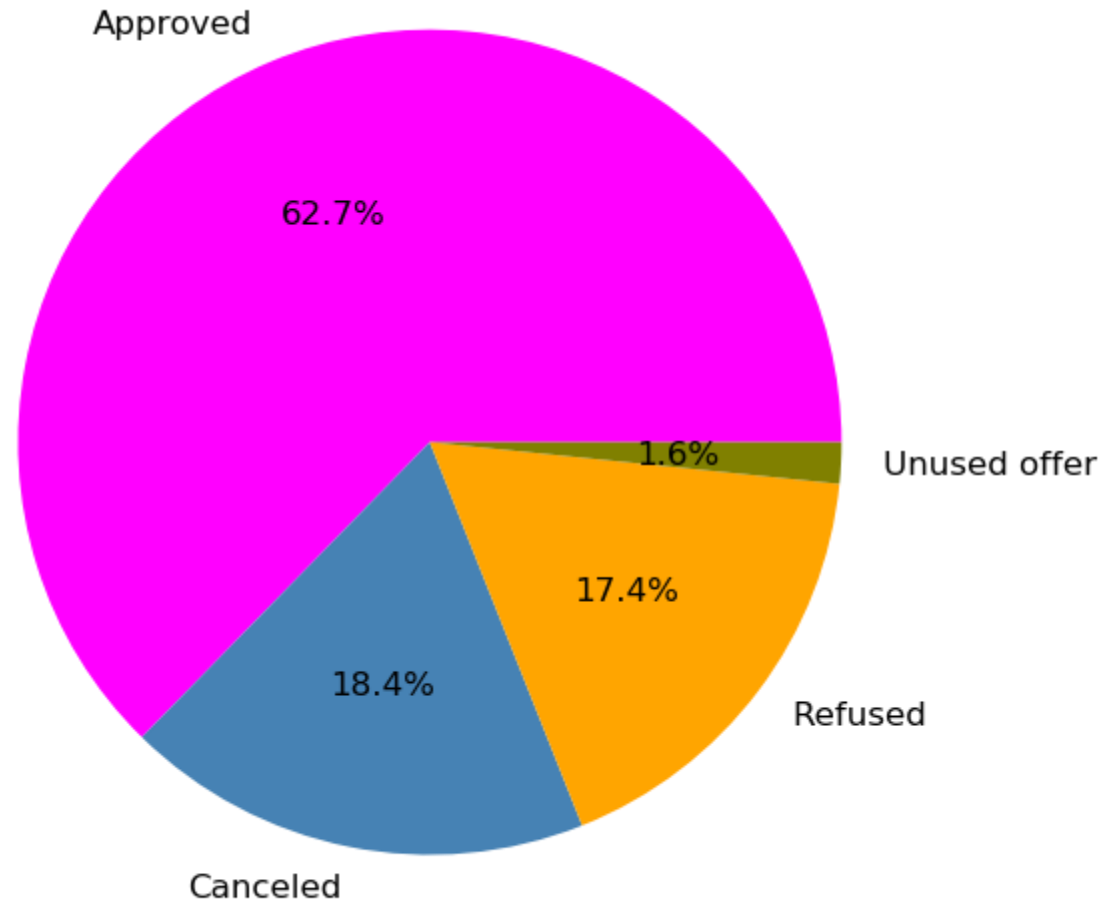
Defaulters

Cash, POS, household with interest, and POS mobile with interest are the top 3 products

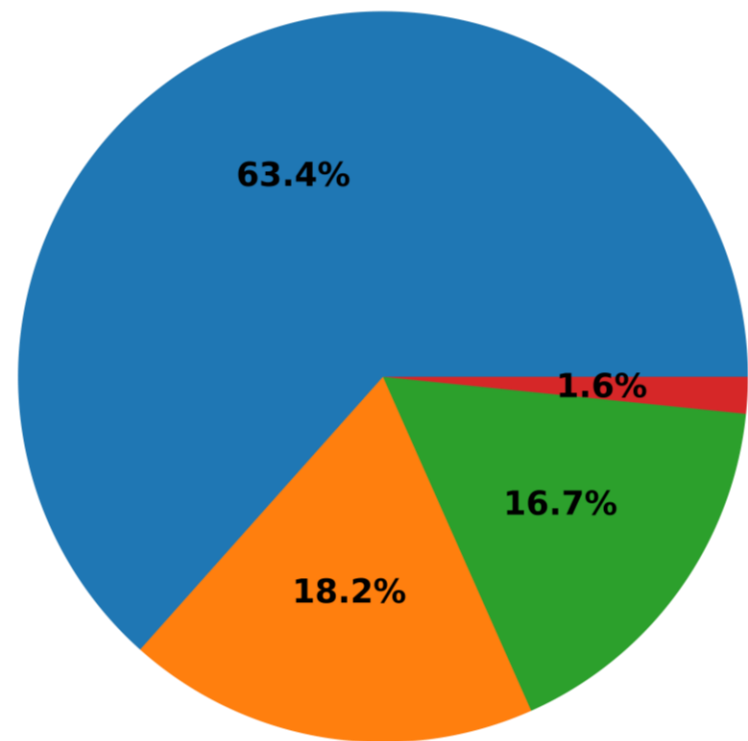


About 63% of previous applications are approved

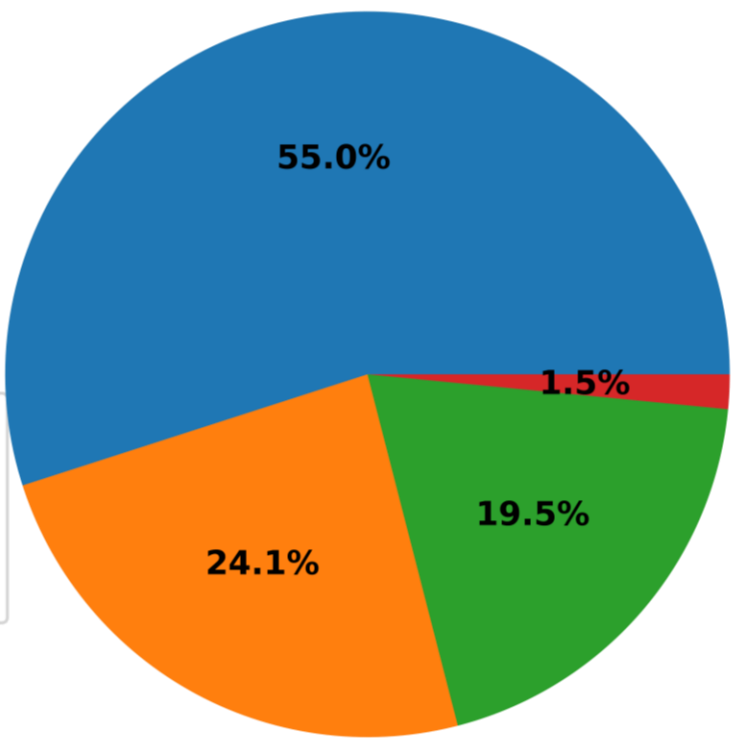
A big number is cancelled and refused and small portion of unused offer



Fewer percentage approved while higher percentage cancelled compared to non-defaulters in previous application

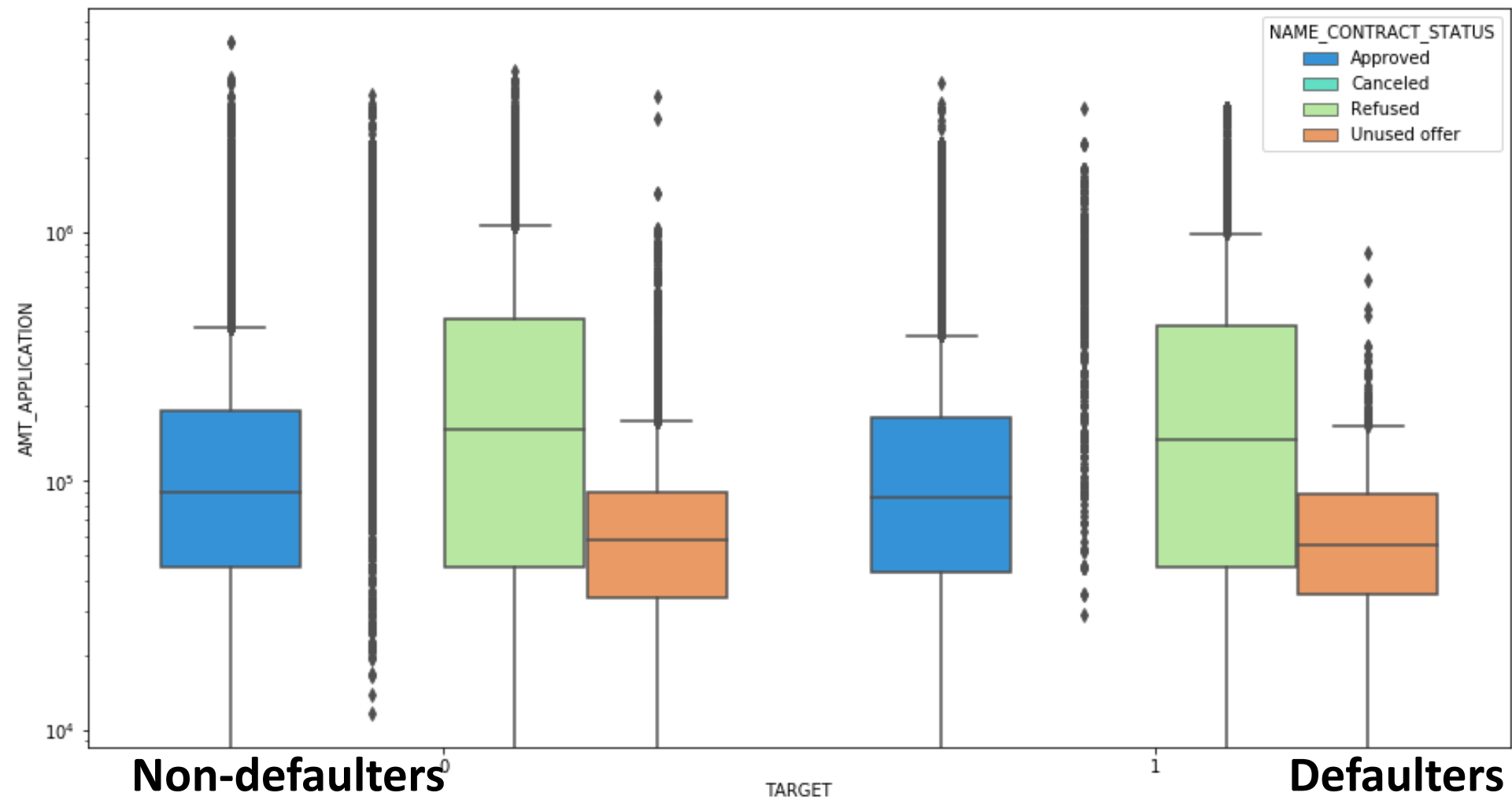


Non-defaulters

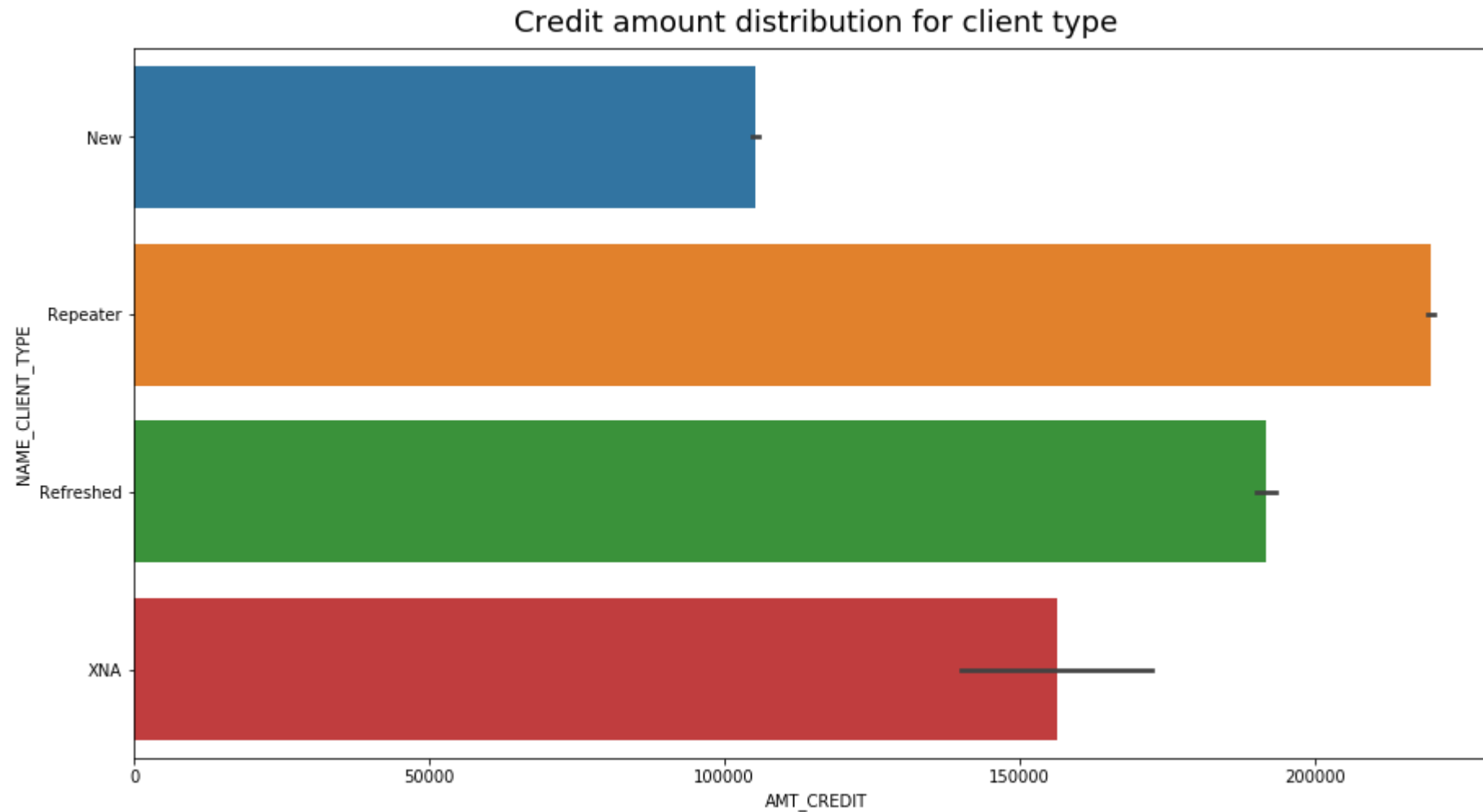


Defaulters

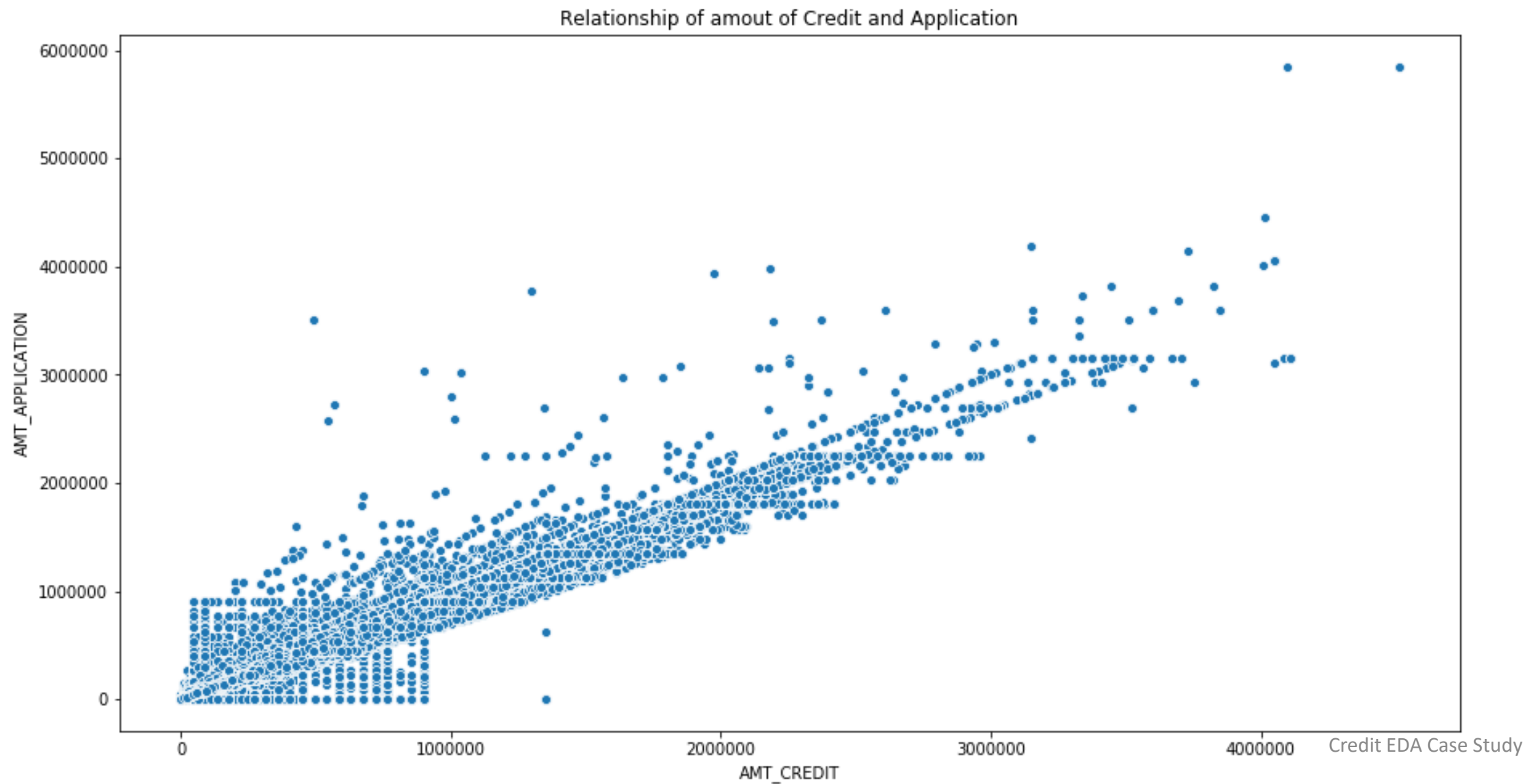
Cash Loans refusals are more when compared with other loans.
No Unused offer reported in Cash loans and Revolving loans.
There are no cancellation reported



Higher number of applications are repeater and refreshed
Client with XNA is not known. It raises question if information is missing or Fraud.

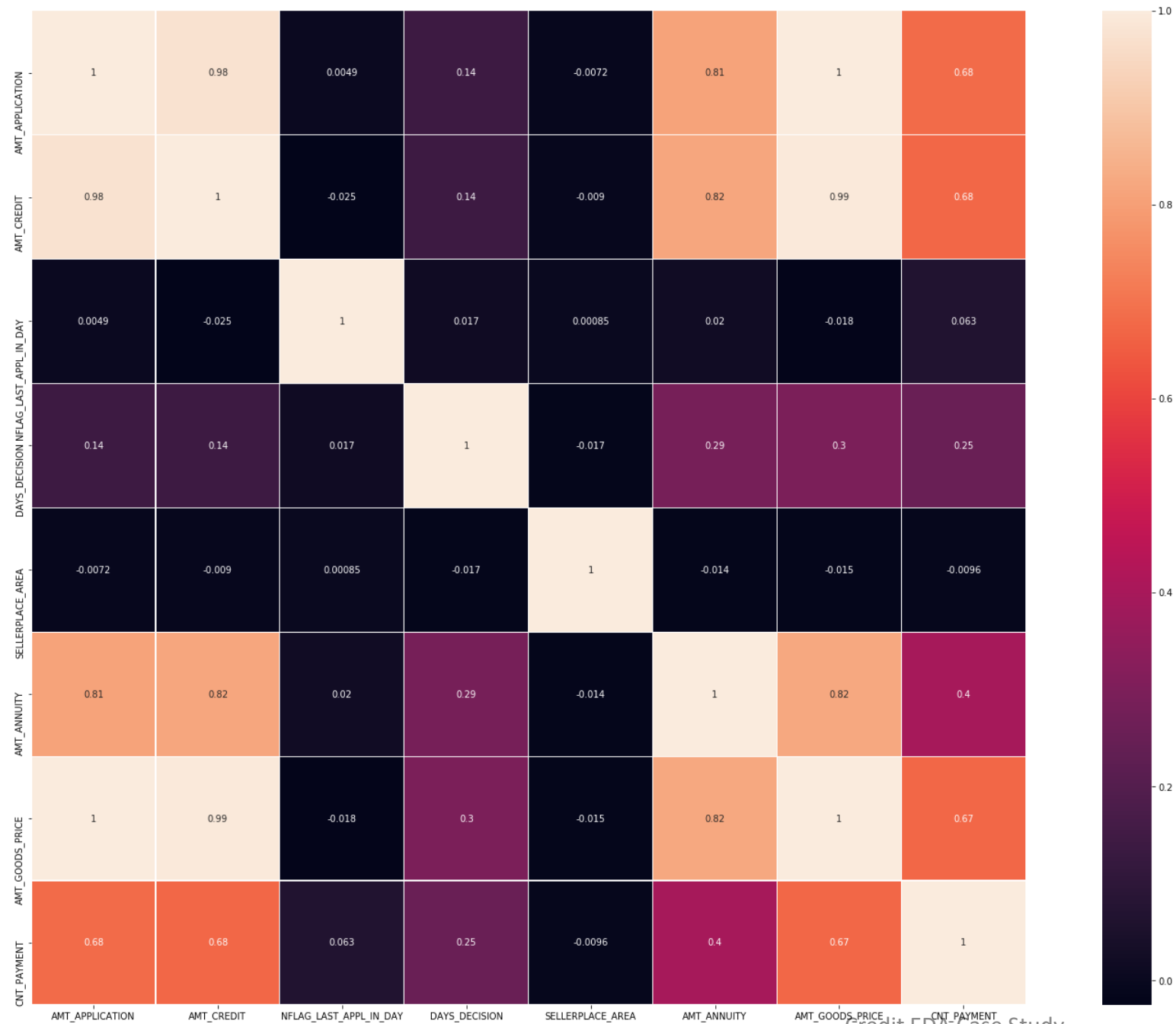


Relation of application amount and credit amount



Loan Amount requested by client originally and final credit loan amount on the previous application given to the client are highly correlated as seen from above scatter and heat maps. This means the loan companies are providing almost closest amount of for which the client has made an initial request for.

We have already seen earlier that consumer loans are the more offered than other loans



Credit EDA Case Study

Variables analyzed in previous application data -1

Variable	Category of Variable	Non-Defaulters	Defaulters	Observations
Hour of application for loan				Most of the applications are processed during 10AM to 4PM
Day of application for loan	Wednesday/Tuesday	15.2%	15.5%	Most of the previous applications are processed on Fridays, Mondays, Tuesdays and Wednesdays . While a lot of applications are also processed on weekends like Sat and Sun When separating by defaulter and non-defaulter no significant difference found
	Monday/Friday	15.1%	15.3%	
Product Combination	Cash	16.6%		Cash, POS, household with interest , and POS mobile with interest are the top 3 products combination for previous loan applications
	POS With Household	16%		
Status	Approved	63.4 %	55%	Fewer percentage approved while higher percentage cancelled compared to non-defaulters in previous application
	Cancelled	18.2%	24.1%	

Variables analyzed in previous application data -2

Variable -1	Variable - 2	Observations
Loan amount	Application status	Cash Loans refusals are more when compared with other loans. No Unused offer reported in Cash loans and Revolving loans. There are no cancellation reported
Applied Loan amount	Amount credited	There is a positive correlation between Applied loan amount & Amount credited
Applied Loan amount	Channel type	Considering the sum of application amount , credit and cash offices had the highest amount in total
Client type	Amount credited	A lot of people have difficulties to get loans due to insufficient credit histories. It shows that old clients are the more favored than others to provide loans by the loan companies. Loans are also given to customers whose data is not available with the Loan companies(Client Type=XNA). This could be a fraud or some error in capturing client information through the system. It is necessary that loan companies capture full and complete information for each client.

Recommendations

Recommendations

	Factors favoring Non default Loans	Factors favoring Default Loans
Contract type	Revolving Loans	Cash Loans
Gender	Female applicants	Male applicants
Own a car or not	Applicant do not own car	Applicant owns car
Housing type	Applicant’s Housing type is House/ Apartment	Applicant staying with parents
Profession	Core staff, Manager	Laborers, Sales staff
Marital status	Married	Singles & civil marriage
Education	Higher Education	Secondary Education or less
Client income	Even though there is no clear difference between Non Defaulters and Defaulters income range, if the income range is in between 90000 and 200000 we should be careful as The majority of loans in both defaulters and non-defaulters between 50th and 75th percentile.	
Loan amount	Here Defaulters are having higher Credit range of up to 800000 whereas it is 750000 for Non defaulters	
Annuity amount	Defaulters are having 44% higher annuity value compared to Non Defaulters	

Recommendations

1. Re-consider loan applications in the following cases that are more prone to defaults
 1. Cash loans vs revolving loans
 2. Males vs female
 3. Car owners vs non-owners
 4. Laborers and sales staff
 5. Single/non-married applicants
 6. Secondary education/people who has newly graduated
2. There are issues with data that needs cleaning
3. Higher number of applications are repeater and refresher in previous application
4. Client with XNA is not known. It raises question if information is missing or Fraud. That should be considered and find the reason behind XNA
5. Research more on cancelled and refused previous applications and find the reason for cancel and refusal. This could help to provide more loans

The End