

# CREDIT EDA CASE STUDY

Sushant Sarswat (C-38 Batch)

# Scope of Presentation

2

- Problem Statement & Objective
- Method of Approach
- Imputation
- Outliers
- Analyses
  - Univariate
  - Segmented Univariate
  - Bivariate
  - Correlation
  - Previous Data (Merged)
- Major Insights
- Recommendations

# Problem Statement & Objective

3

## **PROBLEM STATEMENT**

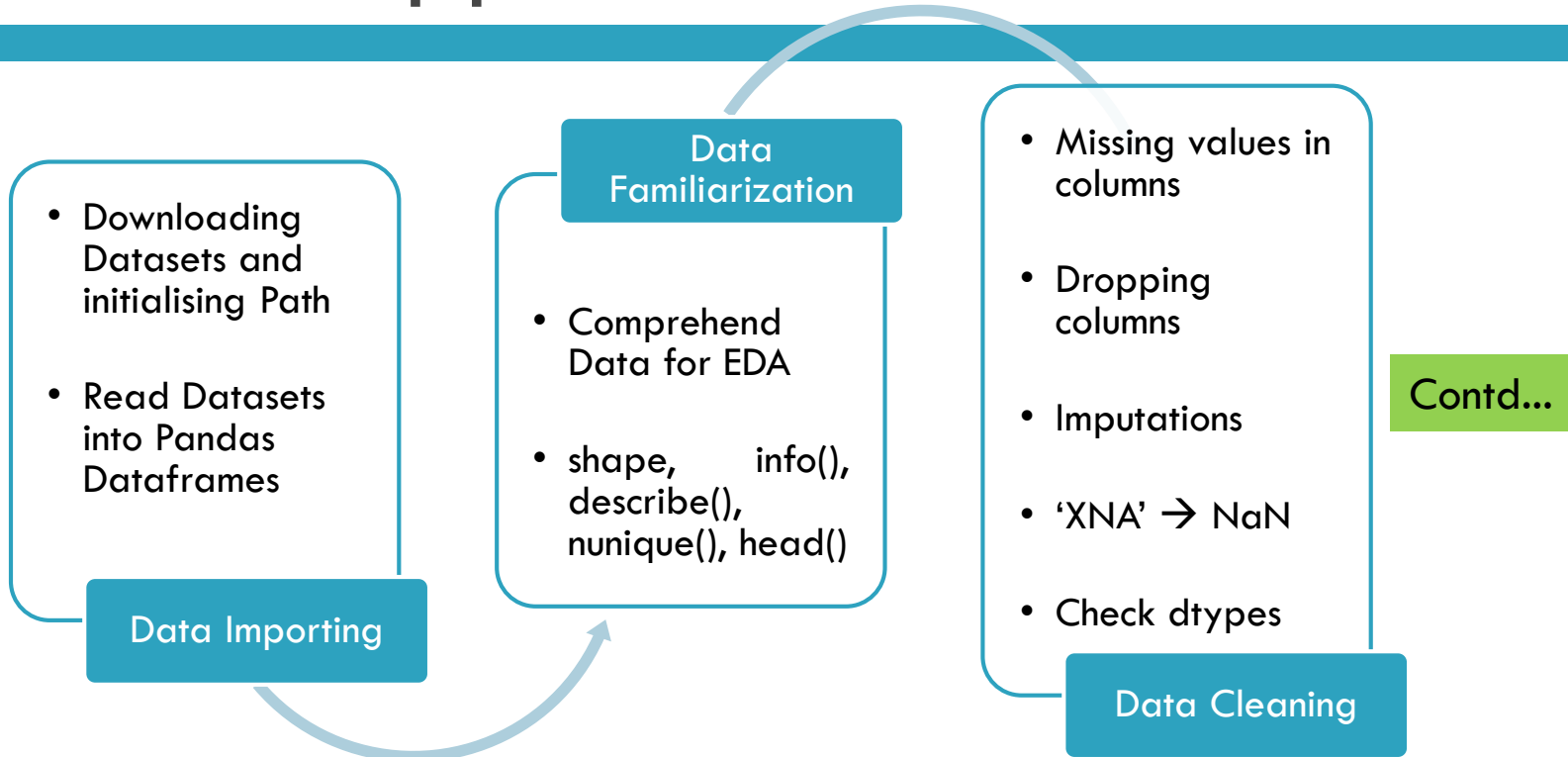
To understand the driving factors (variables) behind loan default (i.e. Variables with strong indicator of default)

## **OBJECTIVE**

- ❖ Ensure disbursement of loan to client likely to pay on time
- ❖ Prevent provisioning of loan to clients likely to default

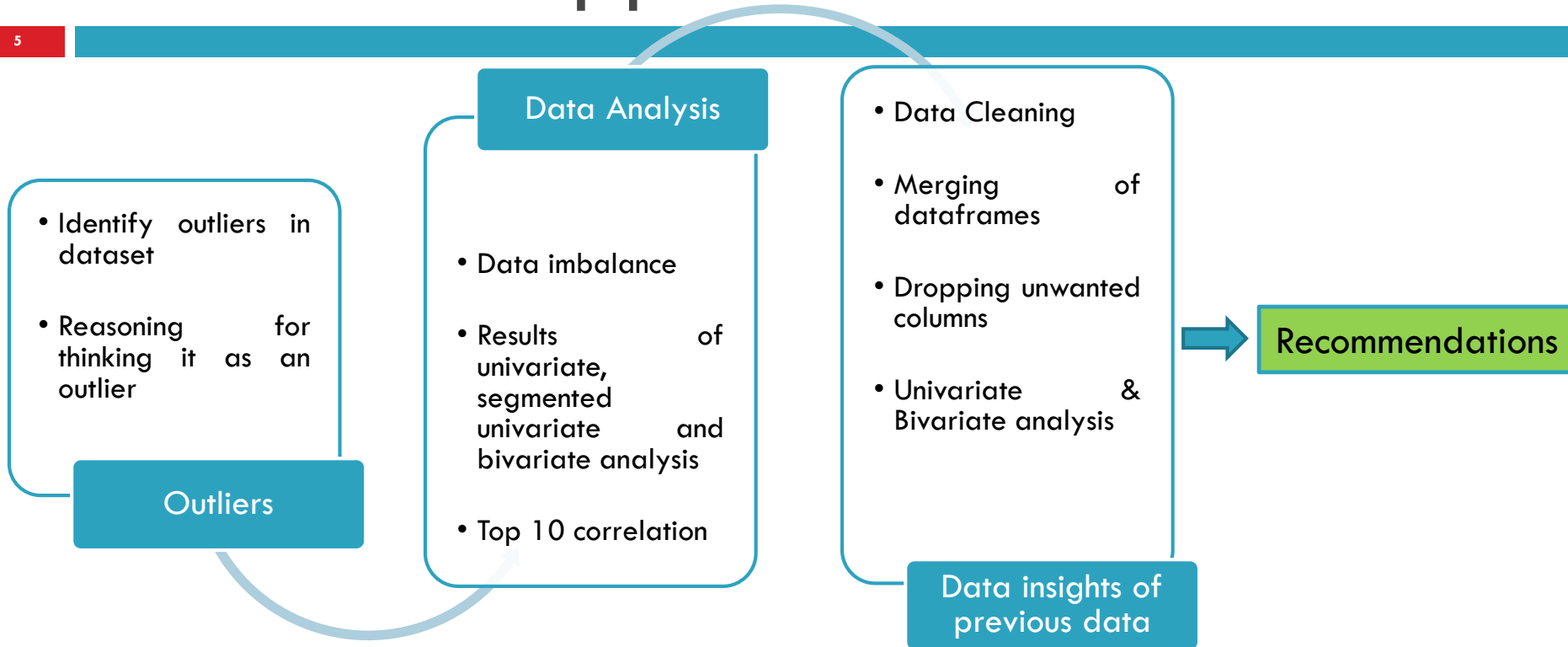
# Method of Approach

4



# Method of Approach

5

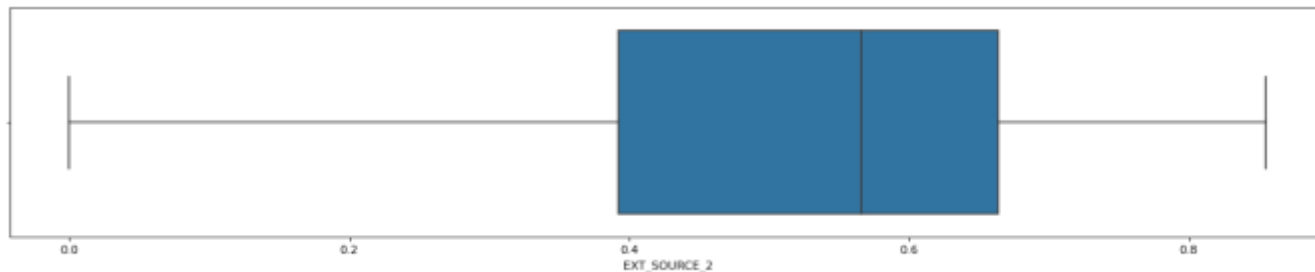


# Imputation (Continuous Variables : For columns with <20% missing Data)

6

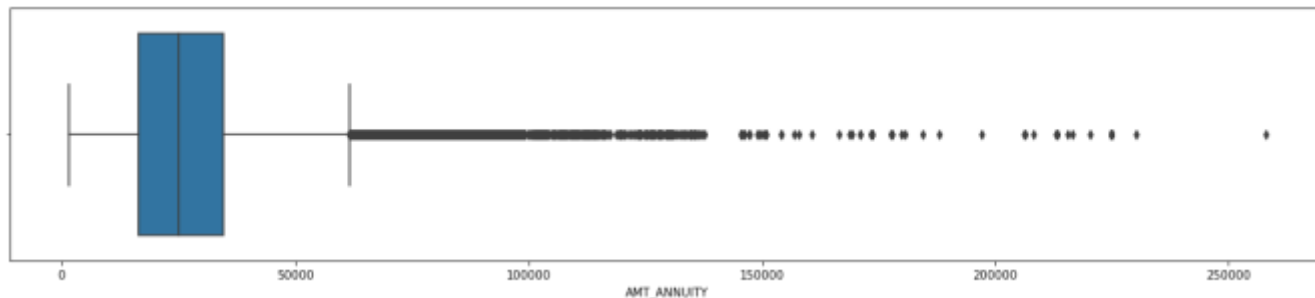
EXT\_SOURCE\_2

Mean and Median almost similar. No Outliers present. Missing values can be imputed with value of 0.565



AMT\_ANNUITY

Outliers present in Data. Missing data may be imputed with median 24903



Note: Replacement of values has not been done in the EDA i.a.w provided guidelines. Only approach has been indicated

# Imputation (Categorical Variables : For columns with <20% missing Data)

7

For Categorical Variables, the imputed values should be the Mode (i.e. Maximum in Frequency)

OBS_30_CNT_SOCIAL_CIRCLE	: 0.0
OBS_60_CNT_SOCIAL_CIRCLE	: 0.0
AMT_REQ_CREDIT_BUREAU_YEAR	: 0.0
AMT_REQ_CREDIT_BUREAU_MON	: 0.0
CNT_FAM_MEMBERS	: 2.0
AMT_REQ_CREDIT_BUREAU_QRT	: 0.0
DEF_30_CNT_SOCIAL_CIRCLE	: 0.0
DEF_60_CNT_SOCIAL_CIRCLE	: 0.0
AMT_REQ_CREDIT_BUREAU_DAY	: 0.0
AMT_REQ_CREDIT_BUREAU_WEEK	: 0.0
NAME_TYPE_SUITE	: Unaccompanied
AMT_REQ_CREDIT_BUREAU_HOUR	: 0.0

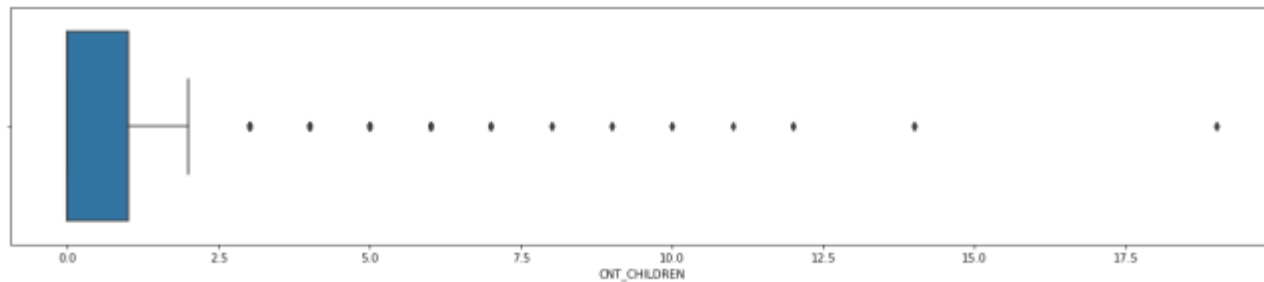
Note: Replacement of values has not been done in the EDA i.a.w provided guidelines. Only approach has been indicated

# Outliers (Numerical Variables)

8

CNT\_CHILDREN

75<sup>th</sup> percentile is at 1.  
Values are mostly lying  
below 3<sup>rd</sup> quartile



AMT\_INCOME\_TOTAL

Single high value outlier  
observed. May be removed  
for better analysis



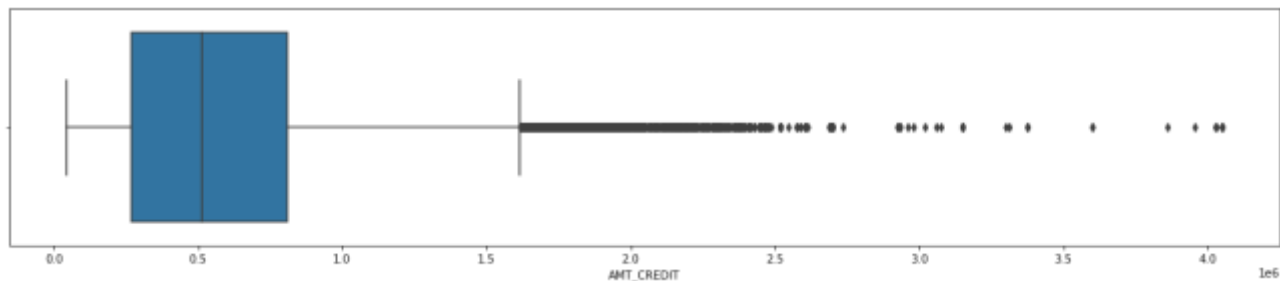


# Outliers (Numerical Variables)

9

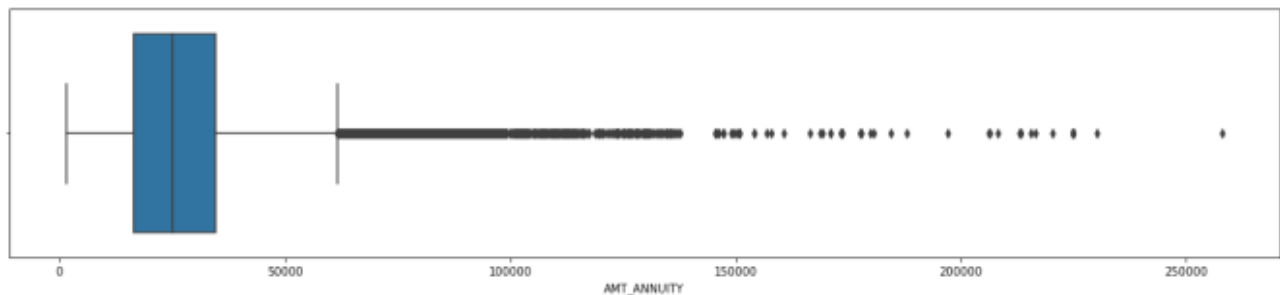
AMT\_CREDIT

1<sup>st</sup> Quartile is slim with respect to 3<sup>rd</sup> Quartile. Therefore, Data is skewed towards first Quartile



AMT\_ANNUIITY

1<sup>st</sup> Quartile is slim with respect to 3<sup>rd</sup> Quartile. Therefore, Data is skewed towards first Quartile

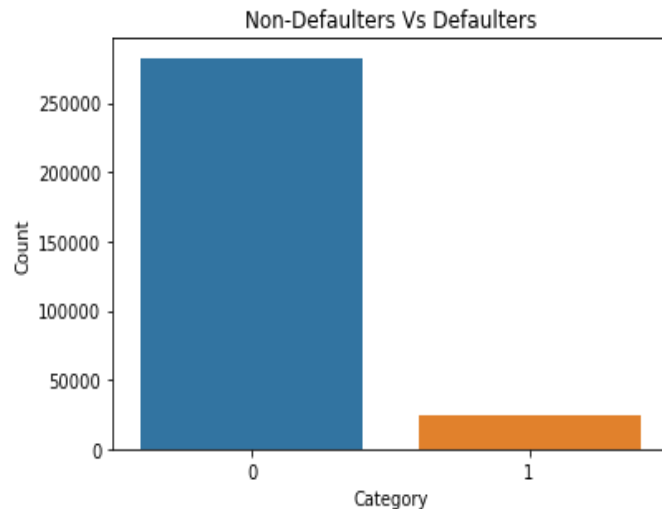


# Analysis (Data Imbalance)

10

Number of Non-Defaulters	: 282682
Number of Defaulters	: 24825
Percentage of Non-Defaulters	: 91.93 %
Percentage of Defaulters	: 8.07 %

Imbalance Ratio	: 11.39
-----------------	---------



Imbalanced data refers to those types of datasets where the target class has an uneven distribution of observations, i.e one class label has a very high number of observations and the other has a very low number of observations

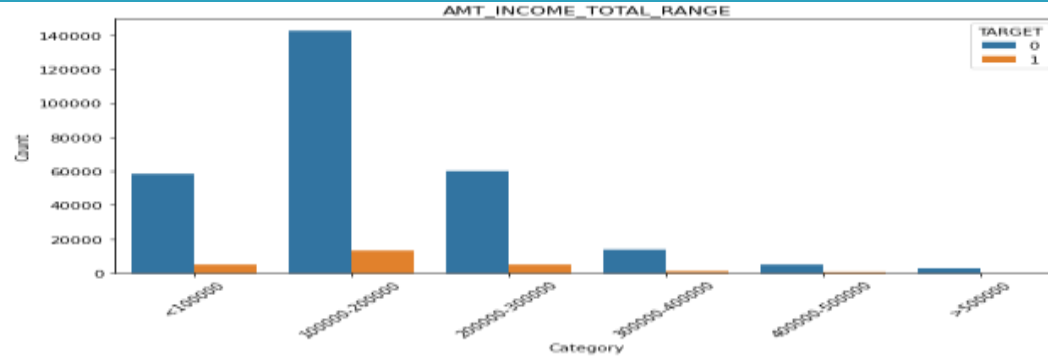
# Analysis (Univariate – Categorical Variables)

11

AMT\_INCOME\_TOTAL\_RANGE

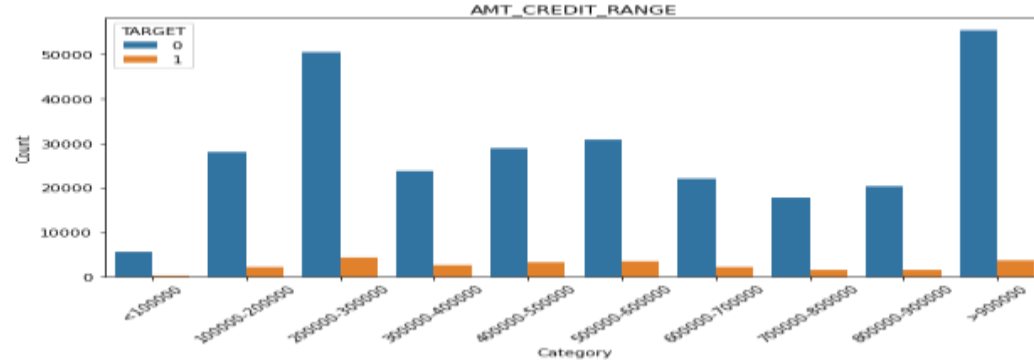
1. Max defaulters are in range 100000 - 200000

2. Min defaulters for range >500000



AMT\_CREDIT\_RANGE

1. Min defaulters for range <100000

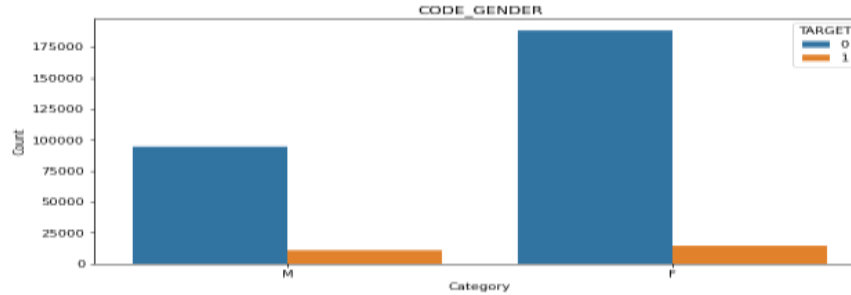


# Analysis (Univariate – Categorical Variables)

12

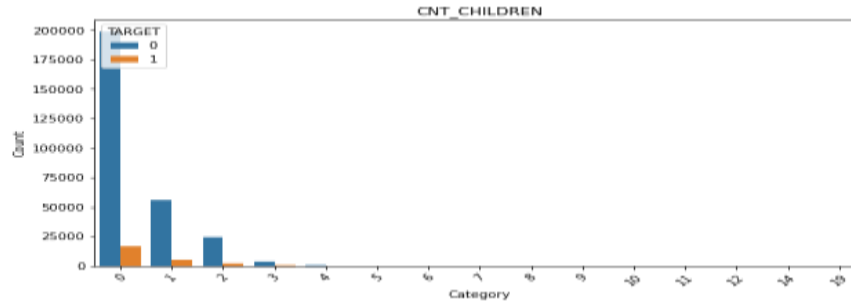
CODE\_GENDER

Males have more defaulters than females percentagewise



CNT\_CHILDREN

The percentage of defaulters tends to increase with number of children of the client

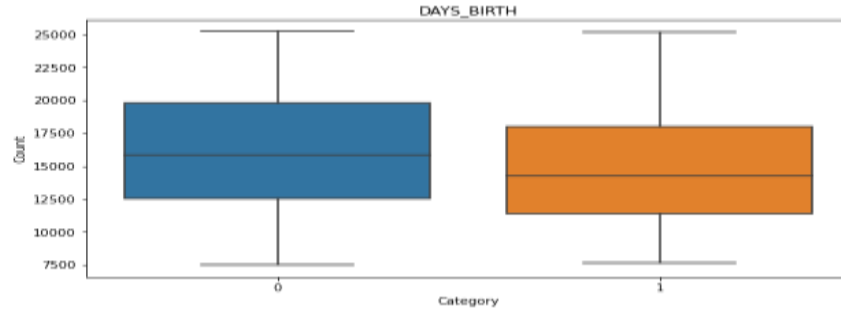


# Analysis (Univariate – Continuous Variables)

13

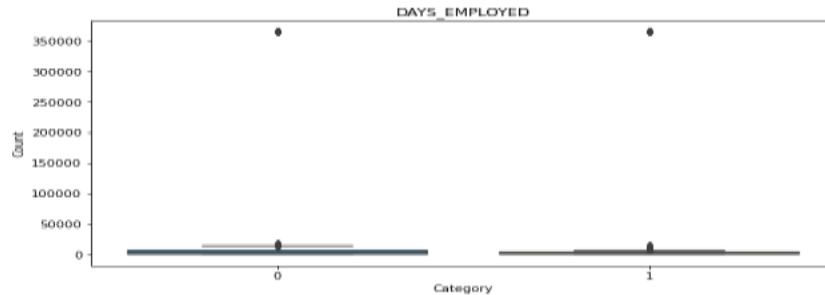
DAYS\_BIRTH

The people with lesser age are more likely to default



DAYS\_EMPLOYED

Removal of outlier will significantly impact boxplot for better analysis

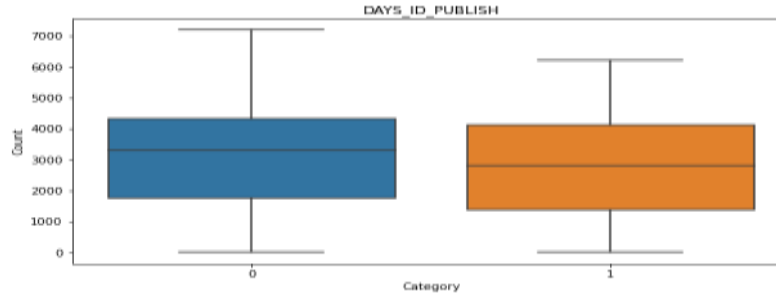


# Analysis (Univariate – Continuous Variables)

14

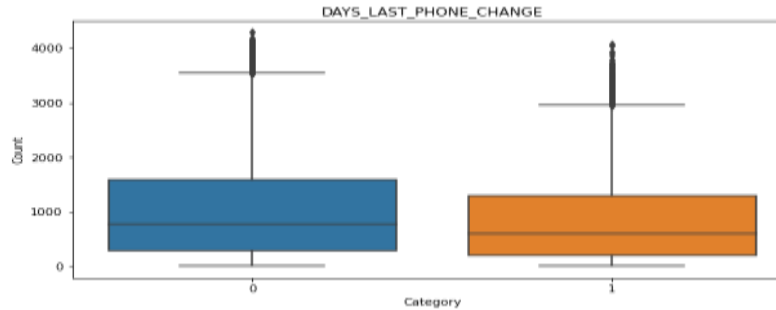
DAYS\_ID\_PUBLISH

Clients who have changed ID recently more likely to default



DAYS\_LAST\_PHONE\_CHANGE

Clients who have changed phone recently more likely to default

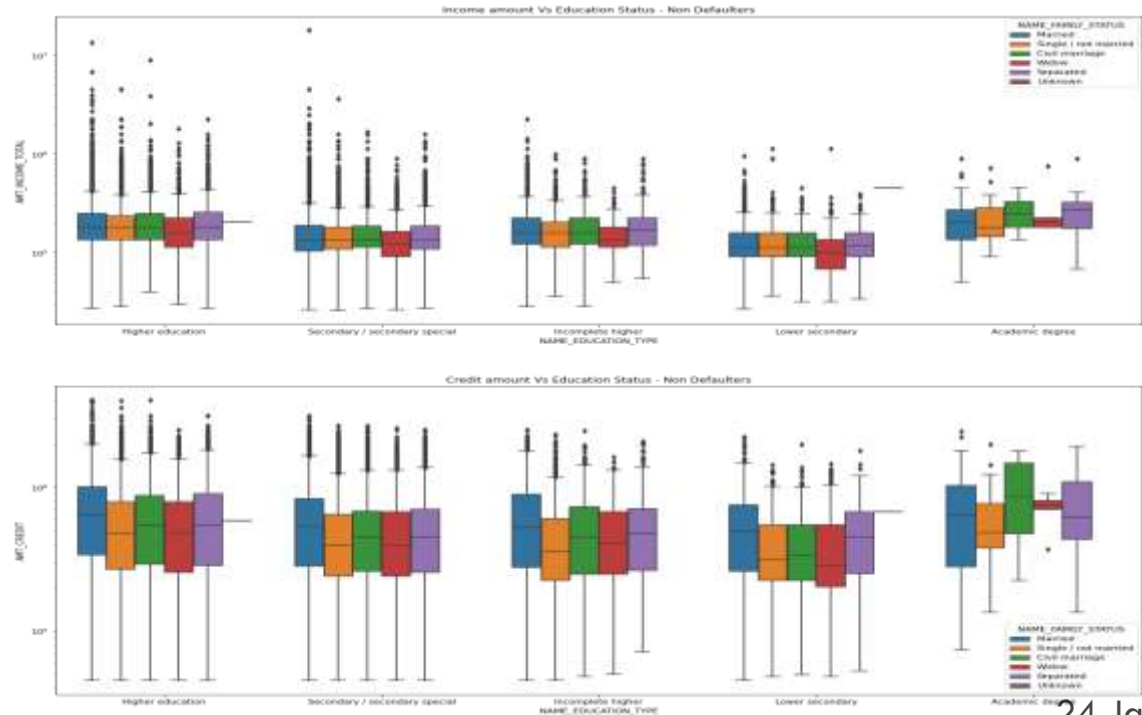


# Analysis (Bivariate – Numerical Variables)

15

## Non-Defaulters

Outliers are not observed in Civil Marriage family having academic degree in both Income amount Vs Education Status and Credit amount Vs Education Status



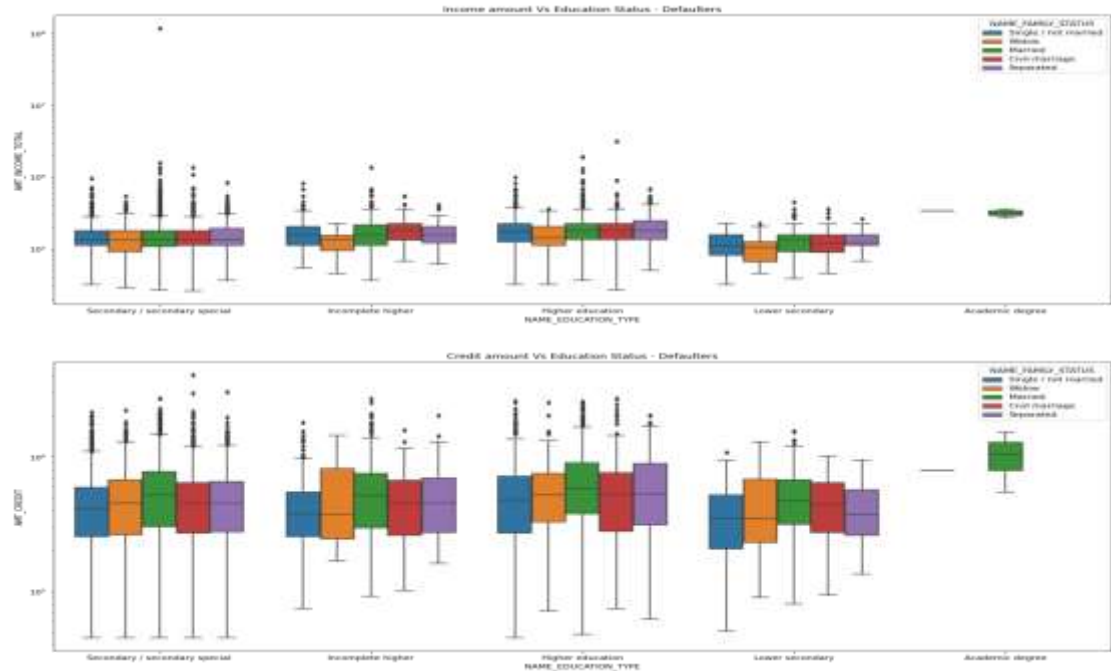
# Analysis (Bivariate – Numerical Variables)

16

## Defaulters

(a) Outliers are not observed in widow family having incomplete higher education in both Income amount Vs Education Status and Credit amount Vs Education Status

(b) Outlier with exceptional income amount observed in married family in Income amount Vs Education Status

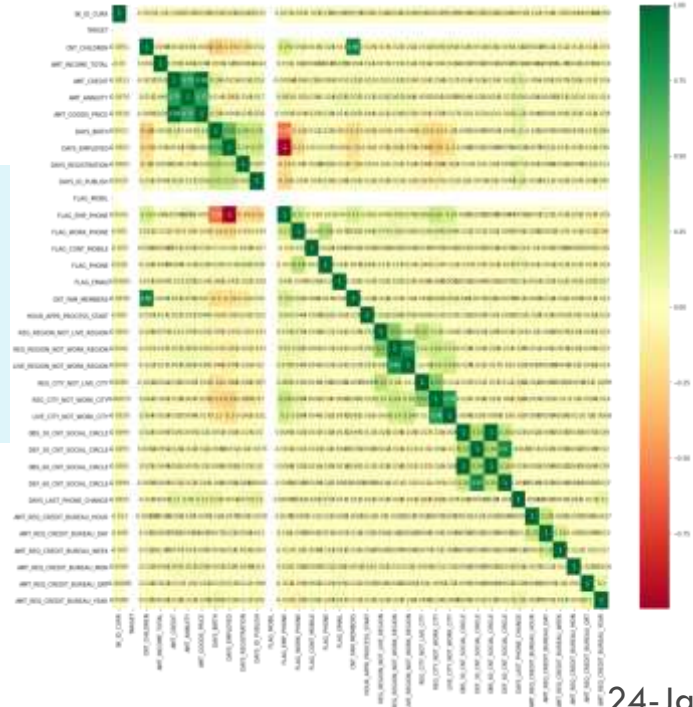




# Analysis (Correlation – Defaulters )

17

OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.998269
AMT_GOODS_PRICE	PRICE_AMT_CREDIT	0.983103
CNT_CHILDREN	CNT_FAM_MEMBERS	0.885484
DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE	0.868994
LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.847885
LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.778540
AMT_ANNUITY	AMT_GOODS_PRICE	0.752699
AMT_ANNUITY	AMT_CREDIT	0.752195
DAYS_EMPLOYED	DAYS_BIRTH	0.582185
REG_REGION_NOT_LIVE_REGION	REG_REGION_NOT_WORK_REGION	0.497937



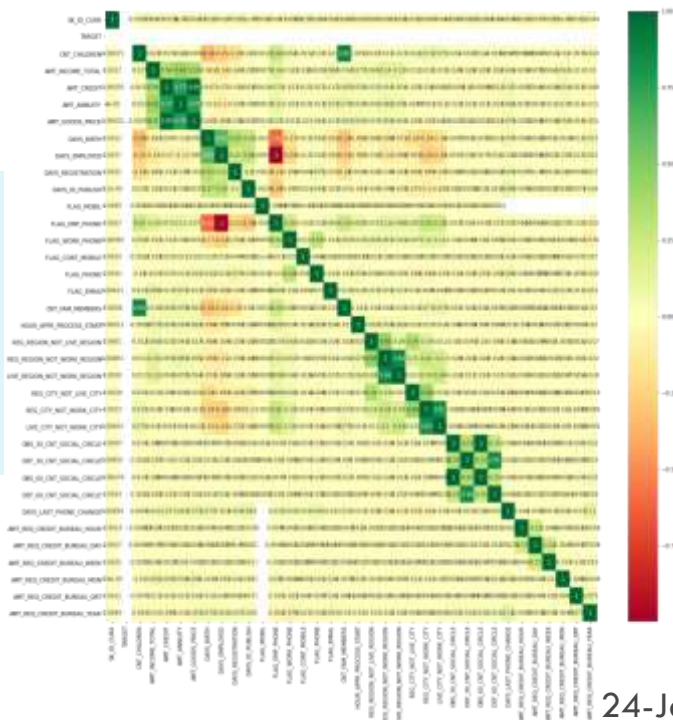
CREDIT EDA CASE STUDY

24-Jan-22

# Analysis (Correlation – Non-Defaulters )

18

OBS_30_CNT_SOCIAL_CIRCLE	OBS_60_CNT_SOCIAL_CIRCLE	0.998508
AMT_CREDIT	AMT_GOODS_PRICE	0.987253
CNT_CHILDREN	CNT_FAM_MEMBERS	0.878570
LIVE_REGION_NOT_WORK_REGION	REG_REGION_NOT_WORK_REGION	0.861861
DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.859289
LIVE_CITY_NOT_WORK_CITY	REG_CITY_NOT_WORK_CITY	0.830381
AMT_GOODS_PRICE	AMT_ANNUITY	0.776686
AMT_CREDIT	AMT_ANNUITY	0.771308
DAYS_BIRTH	DAYS_EMPLOYED	0.626116
REG_REGION_NOT_LIVE_REGION	REG_REGION_NOT_WORK_REGION	0.446101



CREDIT EDA CASE STUDY

24-Jan-22

(Univariate – Previous Data Merged )

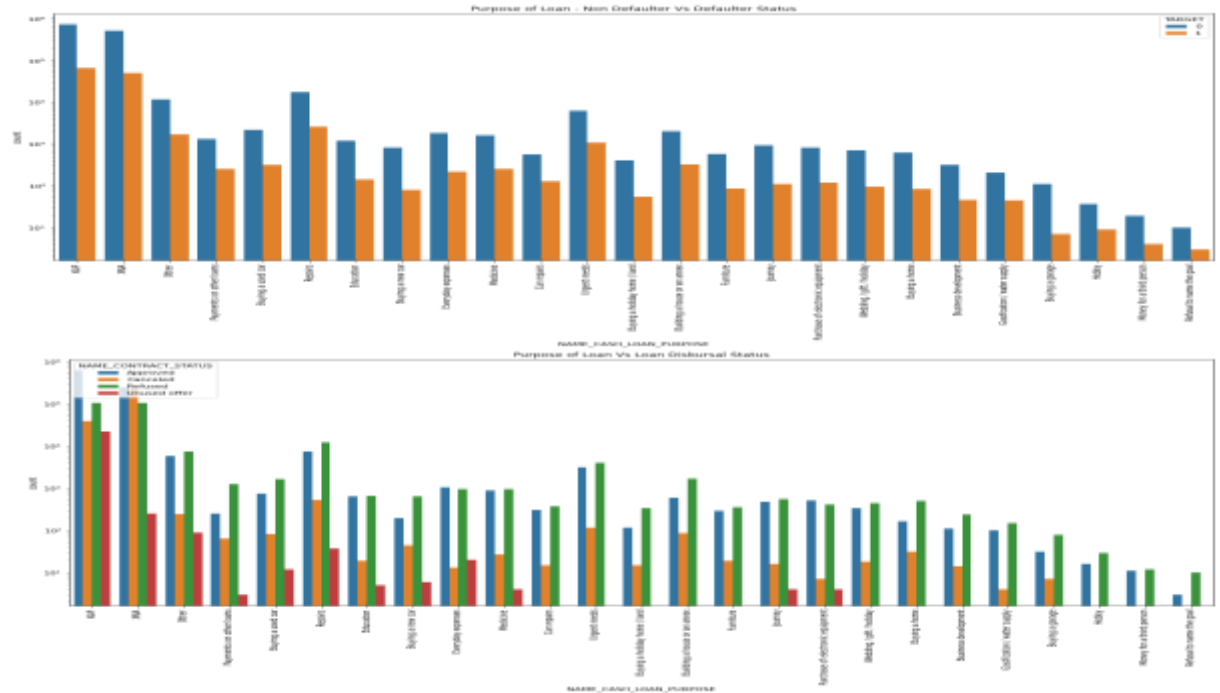
19

## Purpose of Loan

Maximum defaulters are for loans taken for Repairs, Other and Urgent Needs.

### Loan Disbursal Status

Approval of loans is higher than refusal in case of Purchase of electronic equipment.

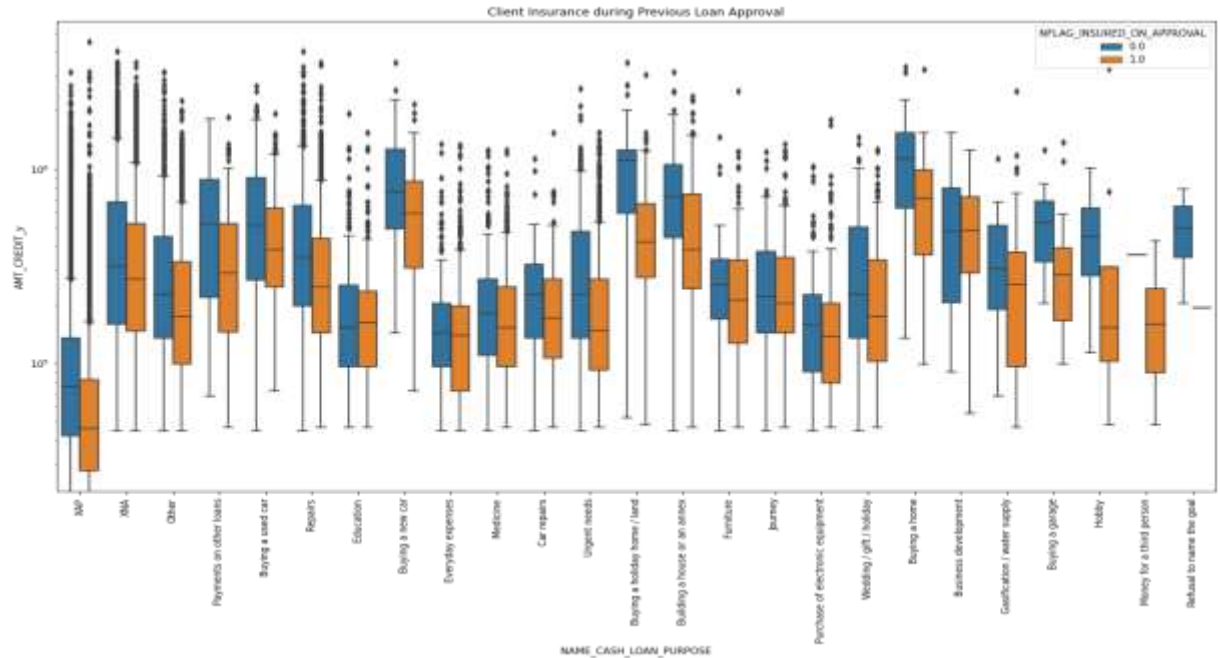


# Analysis (Bivariate – Previous Data Merged )

20

## Client Insurance During Previous Loan

Clients got insured in significant proportions while Buying a New Car, Buying a Holiday home/ Land, Building a house or an annex and Buying a Home



# Major Insights

21

- The percentage of male defaulters is higher i.e. 10.14% Vs female defaulters which is 7.0%
- Car owners are less susceptible to default than non-car owners
- Realty owners are slightly less susceptible to default than non-realty owners
- Percentage of defaulters were higher for clients who did not provide their home phone number
- Percentage of defaulters were higher for clients whose permanent address did not match contact address
- Percentage of defaulters tends to increase with number of children of the client
- Percentage of defaulters tends to decrease with increase in income of the client
- Percentage of defaulters increases with credit amount of loan upto 500000-600000 range and then decreases thereafter

# Major Insights

22

- Percentage of defaulters increases with loan annuity upto 30000-40000 range and then decreases thereafter
- Percentage of defaulters increases for clients with loan for goods worth 300000-400000 range and then decreases thereafter
- Percentage of defaulters tends to increase with increase in family members
- Maximum number of loans were approved for clients who starting applying at 10 AM.
- Minimum loans were applied for clients who started applying around midnight and had high percentage of defaulters
- People having younger age are more likely to default
- Defaulters are likely to change their registration recently in comparison with non-defaulters
- Defaulters are likely to have changed their ID document with which they applied for the loan recently in comparison with non-defaulters

# Major Insights

23

- Defaulters are likely to have changed their phone recently in comparison with non-defaulters
- Cash loans are more popular than revolving loans
- Majority of loan defaulters had come unaccompanied
- Loan defaulters are mainly from Working, Commercial Associate, Pensioner, State Servant category
- Students and Businessmen have very less defaulters
- Maximum defaulters have Secondary education
- Majority of defaulters are labourers
- Max defaulters are from Business Entity Type 3

# Major Insights

24

- In certain categories, loan repayment is much better than loan default viz. Education, Buying a New Car, Buying a holiday home/land and Buying a Garage
- Maximum defaulters are for loans taken for Repairs, Other and Urgent Needs
- Maximum refusal is for loans pertaining to Repairs. This is good as in earlier graph we saw that Maximum defaulters were also from the same category
- Approval of loans is higher than refusal in case of Purchase of electronic equipment
- Clients got insured in significant proportions while Buying a New Car, Buying a Holiday home/ Land, Building a house or an annex and Buying a Home



# Recommendations

25

- Percentage of defaulters were higher for clients who did not provide their home phone number and whose permanent address did not match contact address. Verification of contact details has to be more stringent
- Percentage of defaulters tends to decrease with increase in income of the client. Wealthy clients may be given better offers for loans
- Percentage of defaulters tends to increase with number of children of the client. This may be due to increased expenses and therefore, the applications for loans may be scrutinized for such cases.
- Car owners are less susceptible to default than non-car owners. Status of car ownership may be ascertained from client
- Percentage of defaulters increases with credit amount of loan upto 500000-600000 range and then decreases thereafter. Loans upto said amount may be disbursed with caution
- Percentage of defaulters increases for clients with loan for goods worth 300000-400000 range and then decreases thereafter. Loans upto said amount may be disbursed with caution

# Recommendations

26

- Least percentage of defaulters is for the clients having two family members. Also, the percentage of defaulters tends to increase with increase in family members. May be factored during loan disbursal
- Minimum loans were approved for clients who started applying around midnight and these had high percentage of defaulters also. May be factored during loan disbursal
- Loan defaulters are mainly from Working, Commercial Associate, Pensioner and State Servant category. Loans to these categories may be avoided
- Students and Businessmen have very less defaulters. Loans to these categories may be provided
- In certain categories, loan repayment is much better than loan default viz. Education, Buying a New Car, Buying a holiday home/land and Buying a Garage. These should become the focus areas for safe disbursal of loans.
- Clients got insured in significant proportions while Buying a New Car, Buying a Holiday home/ Land, Building a house or an annex and Buying a Home. This can be used to club insurance offers with loans on such purposes.

# Thank You

27

LIS302DL & Kalman Filter ( $p=0.46972573$ ,  $q=0.0625$ ,  $r=4.0$ ,  $k=0.11743143$ )

“

Without data you're just another person with an opinion.

W. EDWARDS DEMING