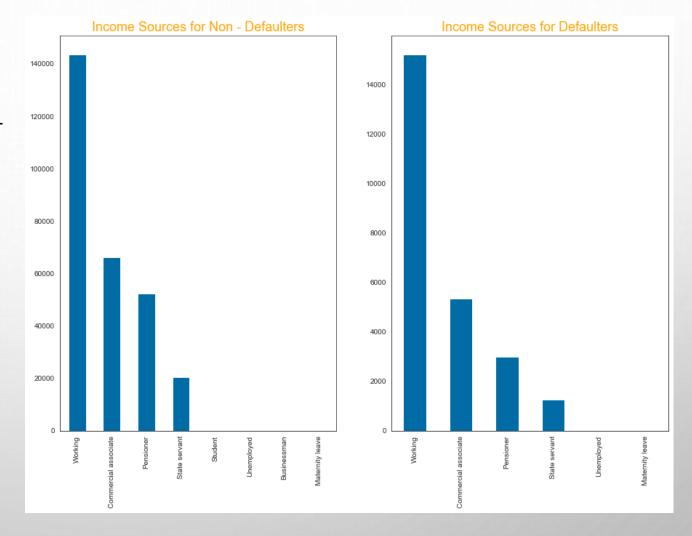# CREDIT EDA CASE STUDY

BY KETAN SONI AND SMIT JAKHOTIA
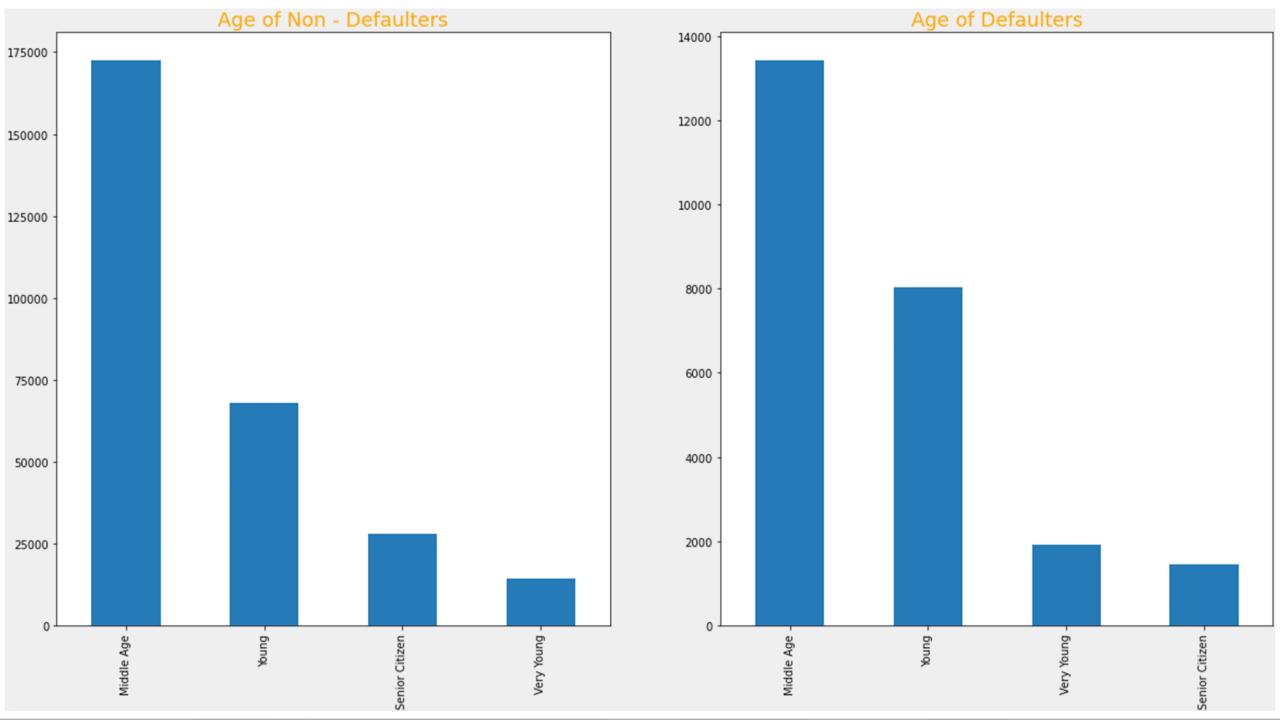
CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET 1 (DEFAULT) AND TARGET 0 (NON - DEFAULT) FOR NEW APPLICATION DATASET

DISTRIBUTION OF INCOME SOURCE

- FOR INCOME TYPE WORKING, COMMERCIAL ASSOCIATE THE NUMBER OF CREDITS ARE HIGHER THAN OTHERS.

- LESS NUMBER OF CREDITS FOR INCOME TYPE STUDENT , UNEMPLOYED AND MATERNITY LEAVE.

- WE ALSO OBSERVE A DECREASE IN THE DEFAULTERS WHO ARE PENSIONERS AND AN INCREASE IN THE DEFAULTERS WHO ARE WORKING WHEN COMPARED THE PERCENTAGE OF BOTH DEFAULTERS AND NON-DEFAULTERS.
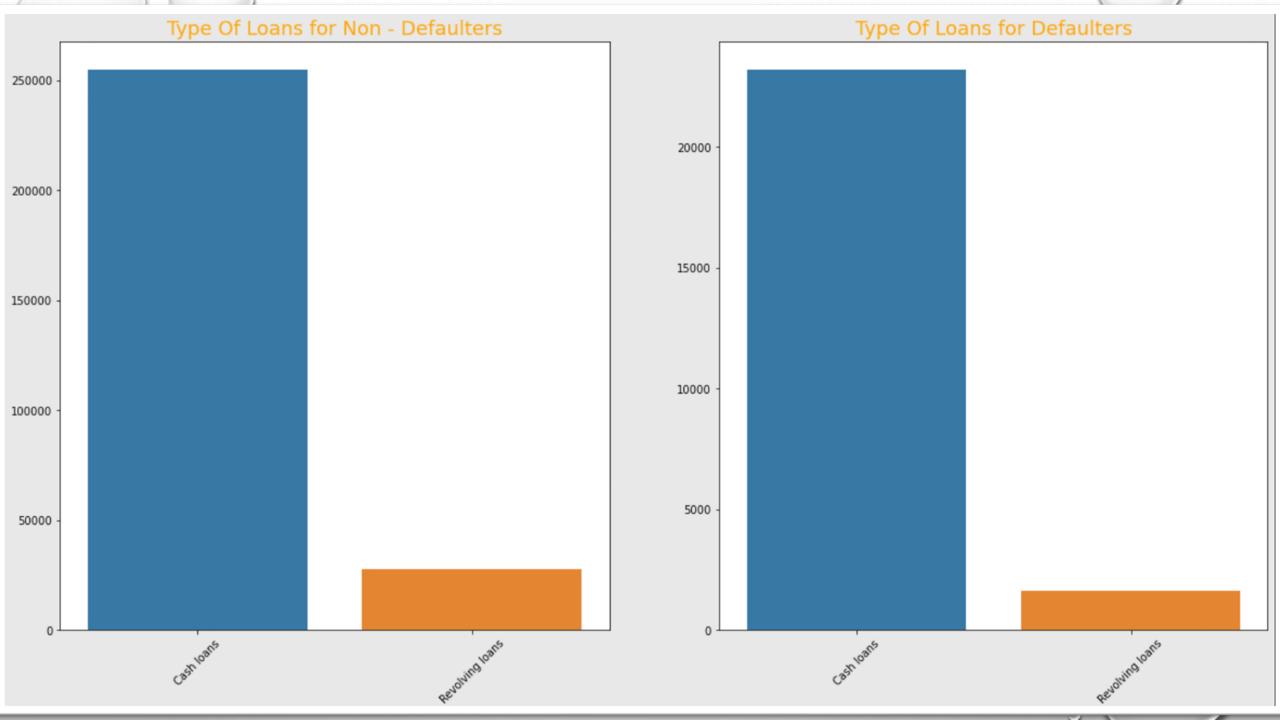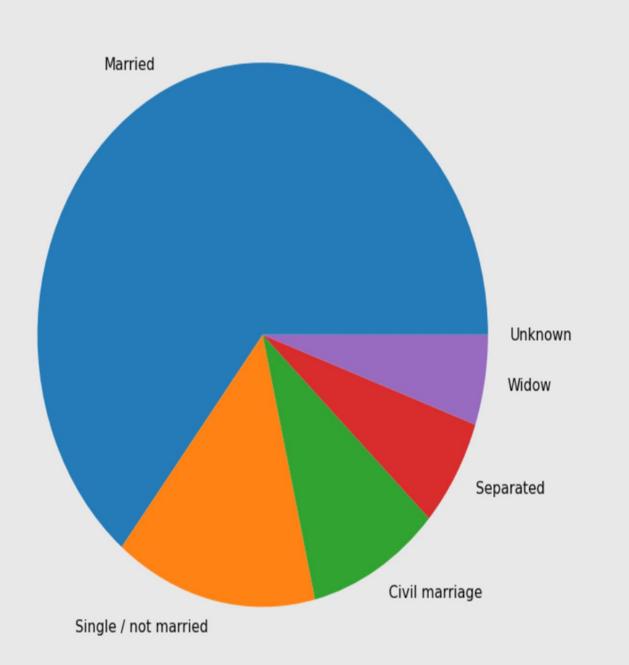
DISTRIBUTION OF AGE GROUPS

- WE OBSERVE THAT THERE IS AN INCREASE IN THE DEFAULTERS WHO ARE YOUNG IN AGE WHEN COMPARED TO THE PERCENTAGES OF DEFAULTERS AND NON DEFAULTERS

Type Of Loans for Non - Defaulters
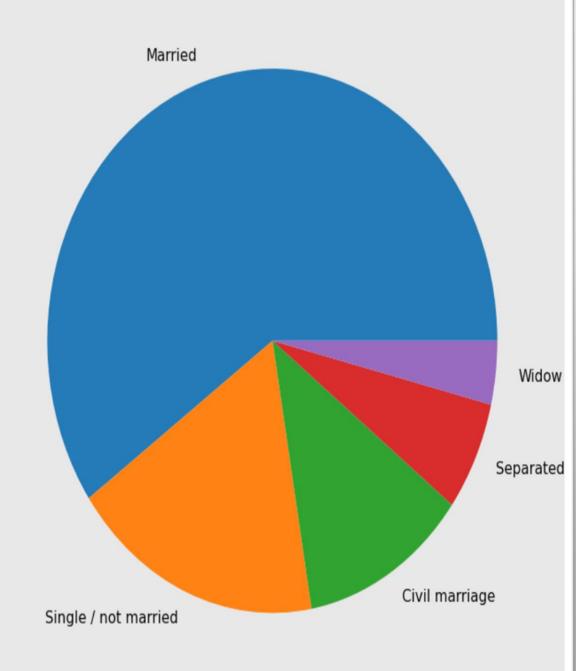
Type Of Loans for Defaulters

DISTRIBUTION FOR LOAN TYPE

- WE CAN OBSERVE THAT CASH LOANS ARE MOSTLY PREFERRED BY BOTH DEFAULTERS AND NON-DEFAULTERS. THOUGH THERE IS A SIGNIFICANT DECREASE IN DEFAULTERS WHO GO FOR REVOLVING LOANS

Family Status of Non - Defaulters

Family Status of Defaulters

DISTRIBUTION OF FAMILY STATUS

- WE OBSERVE AN INCREASE IN DEFAULTERS FOR BOTH SINGLE/NOT MARRIED AND CIVIL MARRIAGE, WHERE AS THERE IS A DECREASE IN WIDOW DEFAULTERS WHEN COMPARED WITH BOTH DEFAULTERS AND NON-DEFAULTERS
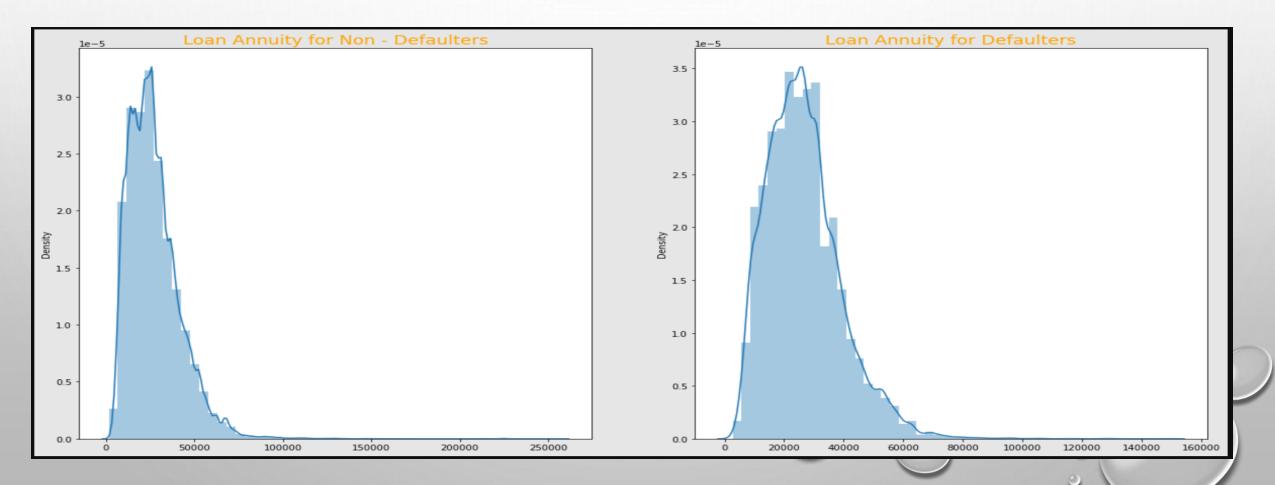
TARGET ANALYSIS FOR NUMERICAL VARIABLES
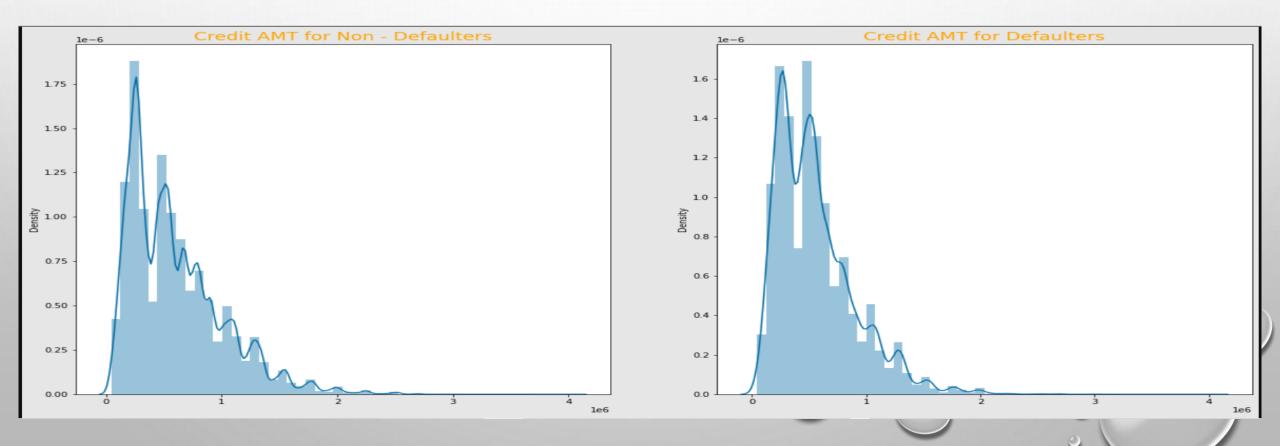
LOAN ANNUNITY

- WE CAN OBSERVE THAT THE DISTRIBUTION CURVE IS QUITE NORMAL FOR BOTH DEFAULTERS AND NON-DEFAULTERS.

CREDIT AMOUNT

- WE CAN OBSERVE THAT THE DISTRIBUTION CURVE IS NOT NORMAL FOR BOTH DEFAULTERS AND NON-DEFAULTERS. WE CAN SAY THAT THE CREDIT AMOUNT OF THE LOAN VARIES.
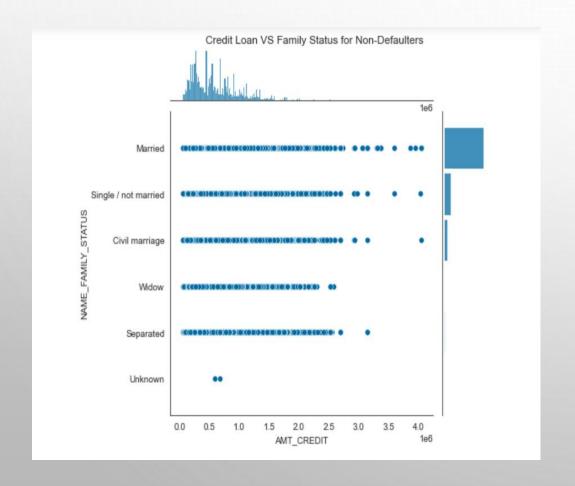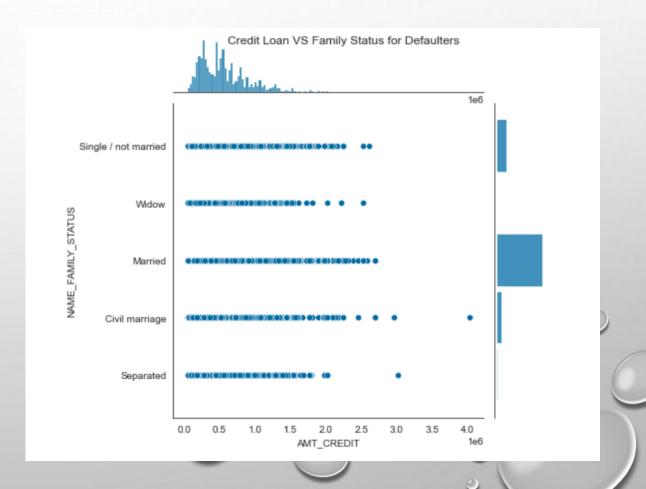
BIVARIATE ANALYSIS FOR NEW APPLICATION DATASET
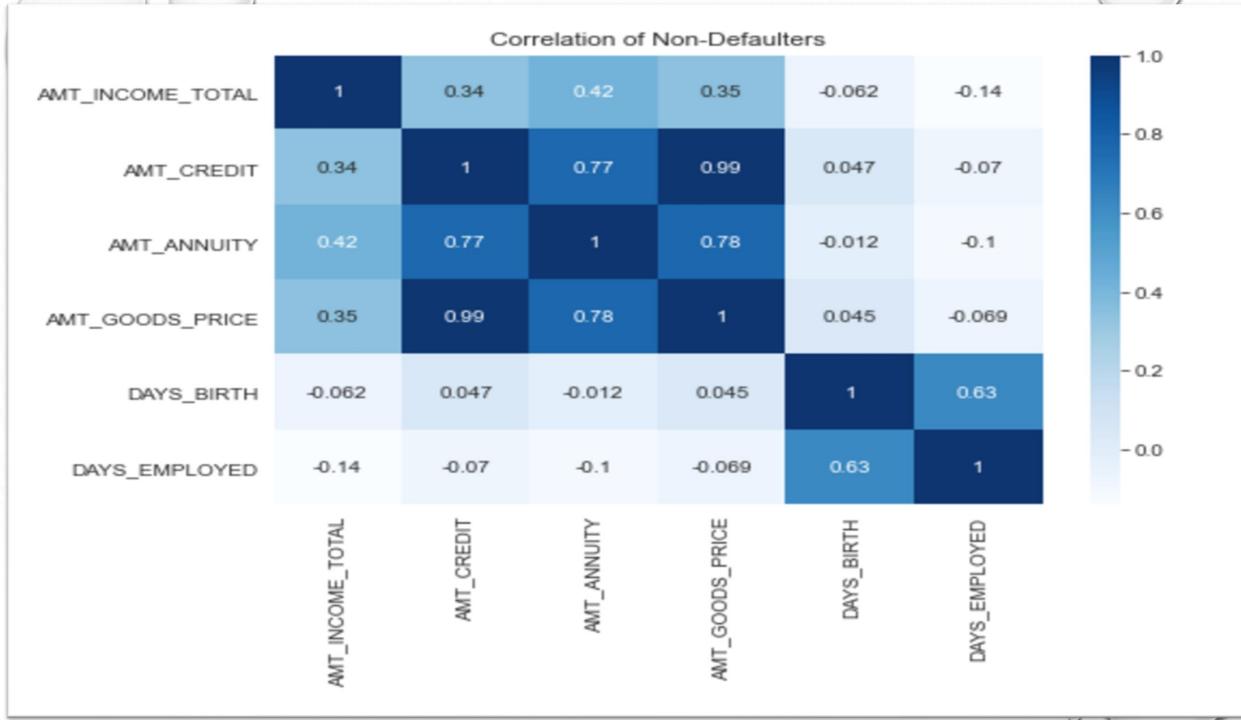
AMT_CREDIT VS NAME_FAMILY_STATUS

We can observe that people having Single/not married and Married Family Status who are non-defaulters have more credit loan amount as compared to defaulters.
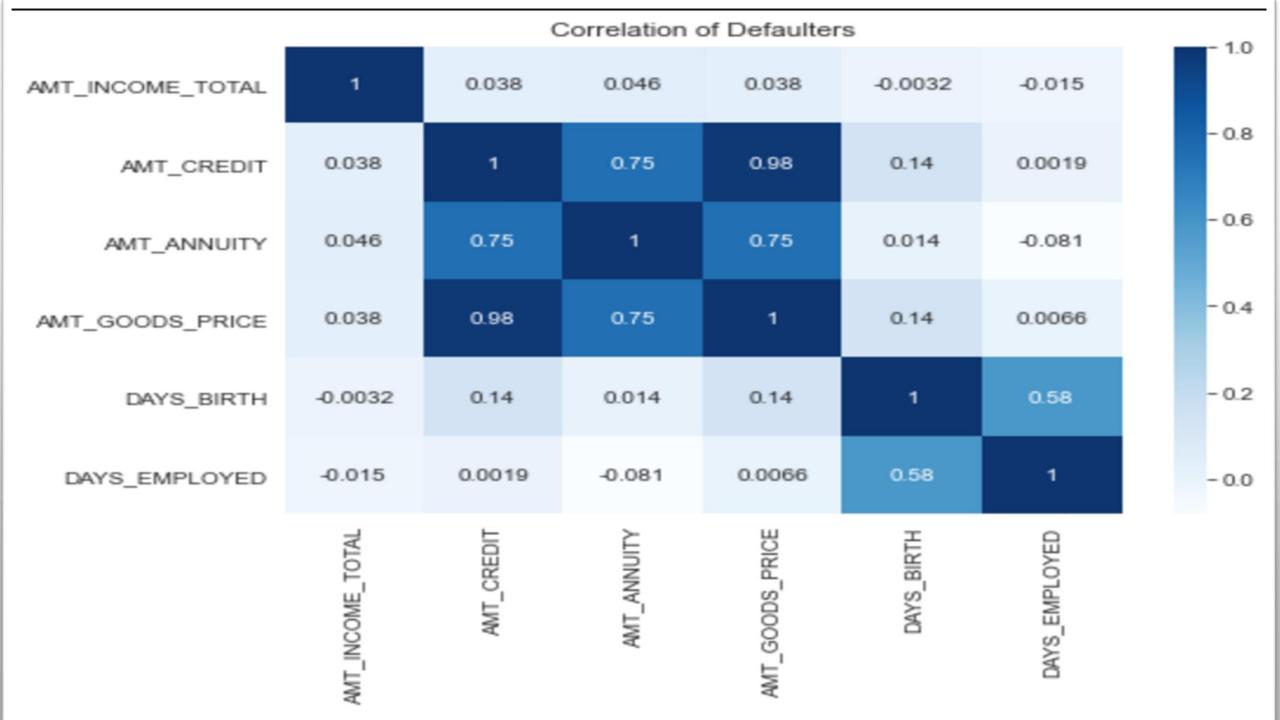
CORRELATION OF TARGET 0 (NON - DEFAULTER) AND TARGET 1 (DEFAULTER)

Correlation of Non-Defaulters

Correlation of Defaulters

INSIGHTS FOR CORRELATION OF TARGET 0 AND TARGET 1

- CREDIT AMOUNT IS INVERSELY PROPORTIONAL TO THE DATE OF BIRTH, WHICH MEANS CREDIT AMOUNT IS HIGHER FOR LOW AGE AND VICE-VERSA.

- WE OBSERVE THAT THERE IS A HIGH CORRELATION BETWEEN GOODS PRICE AND CREDIT AMOUNTS.

- ALSO WE OBSERVE THAT THE GOODS PRICE WITH LOW AMOUNT HAVE HIGH LOAN ANNUITY IN DEFAULTERS.

- IN NON-DEFAULTERS, PEOPLE WITH ALL AGE GROUPS TEND TO TAKE HIGHER CREDIT LOAN AS COMPARED TO NON-DEFAULTERS.

TOP 10 CORRELATION FOR DEFAULTERS

|	VAR_1	VAR_2	CORRELATION	CORRELATION_ABS
42	AMT_CREDIT	AMT_GOODS_PRICE	0.987024	0.987024
14	AMT_ANNUITY	AMT_GOODS_PRICE	0.776433	0.776433
44	AMT_CREDIT	AMT_ANNUITY	0.771308	0.771308
31	DAYS_BIRTH	DAYS_EMPLOYED	0.626030	0.626030
15	AMT_ANNUITY	AMT_INCOME_TOTAL	0.418954	0.418954
7	AMT_INCOME_TOTAL	AMT_GOODS_PRICE	0.349425	0.349425
43	AMT_CREDIT	AMT_INCOME_TOTAL	0.342801	0.342801
39	DAYS_REGISTRATION	DAYS_BIRTH	0.333017	0.333017
38	DAYS_REGISTRATION	DAYS_EMPLOYED	0.214511	0.214511
22	DAYS_EMPLOYED	AMT_INCOME_TOTAL	-0.140392	0.140392
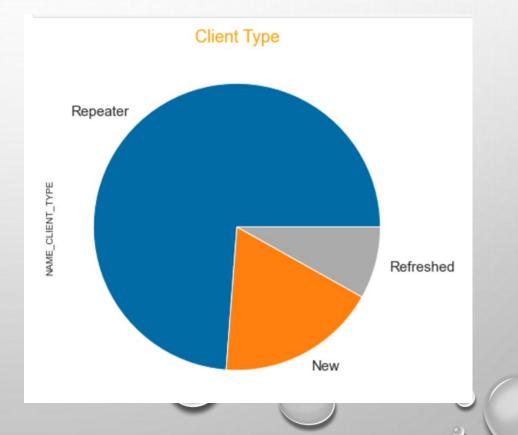
DATA ANALYSIS FOR PREVIOUS APPLICATION DATASET

INSIGHTS FOR NAME_CLIENT_TYPE

- WE CAN OBSERVE THAT THERE IS A HUGE AMOUNT OF CLIENTS WHO ARE REPEATERS.
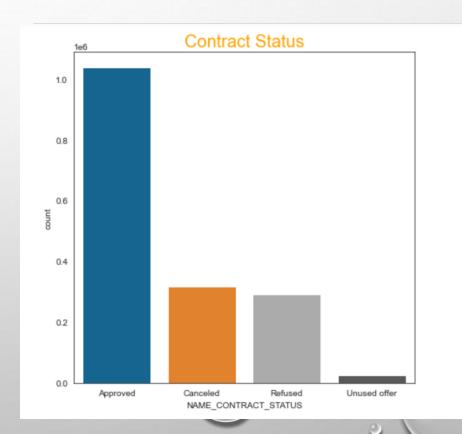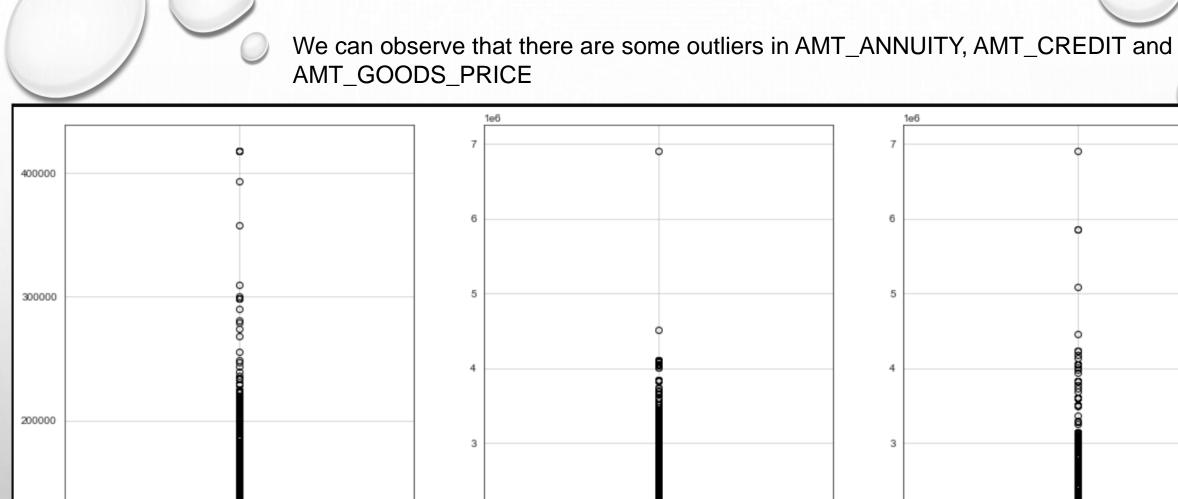
INSIGHTS FOR NAME_CONTRACT_STATUS

- WE CAN OBSERVE THAT MAJORITY LOANS ARE APPROVED. ALSO, CANCELED AND REFUSED LOANS ARE ALMOST EQUAL.
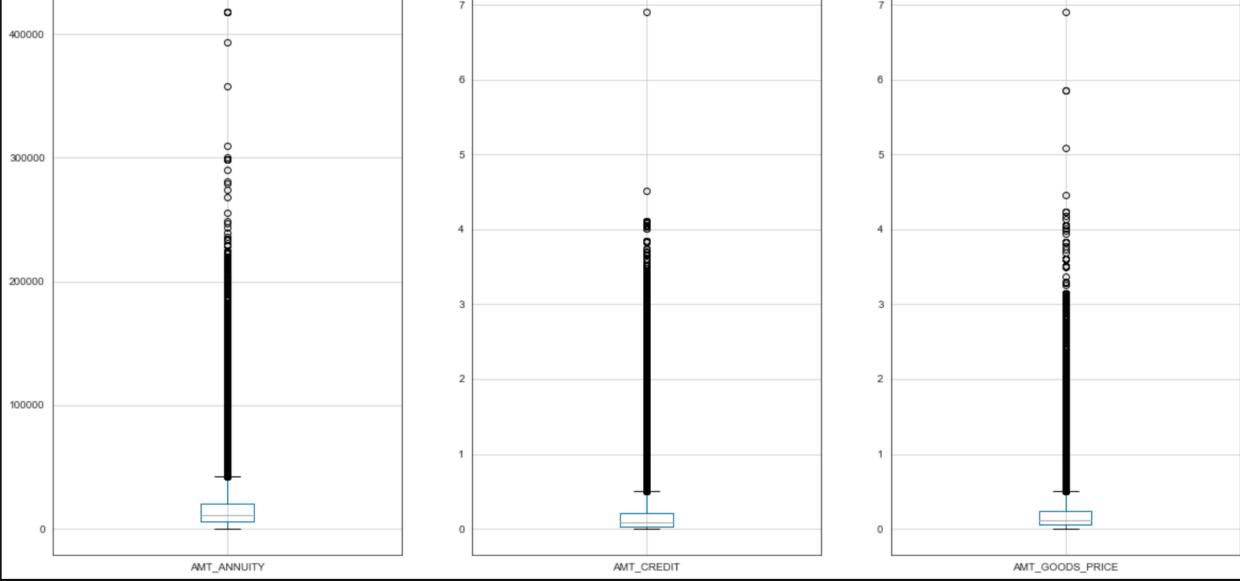
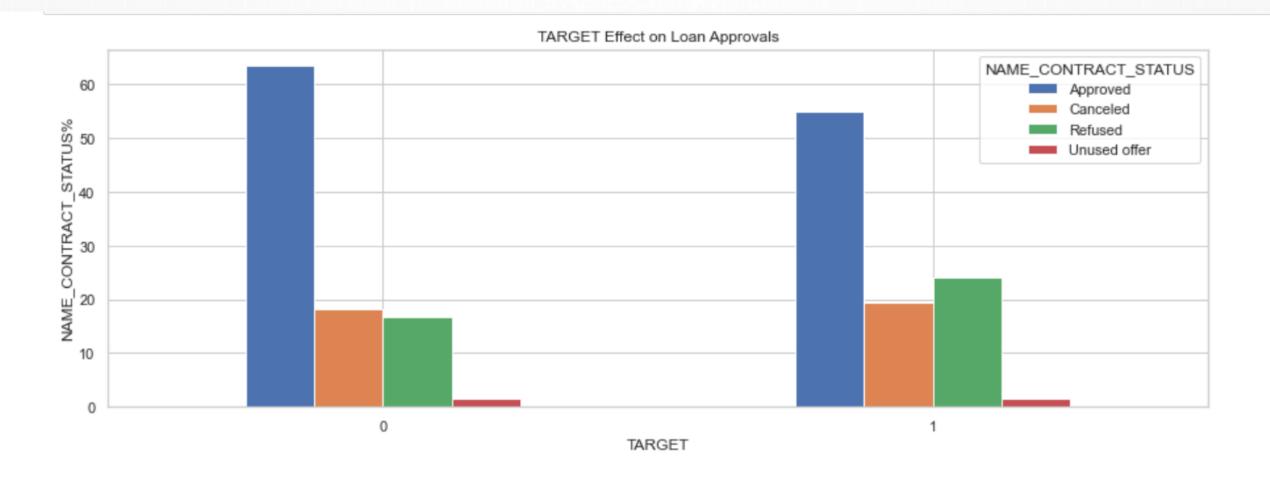# UNIVARIATE ANALAYSIS FOR AMT_ANNUITY, AMT_CREDIT AND AMT_GOODS_PRICE

We can observe that there are some outliers in AMT_ANNUITY, AMT_CREDIT and AMT_GOODS_PRICE

BIVARIATE ANALYSIS FOR PREVIOUS APPLICATION DATASET

TARGET VS NAME_CONTRACT_STATUS

TARGET Effect on Loan Approvals
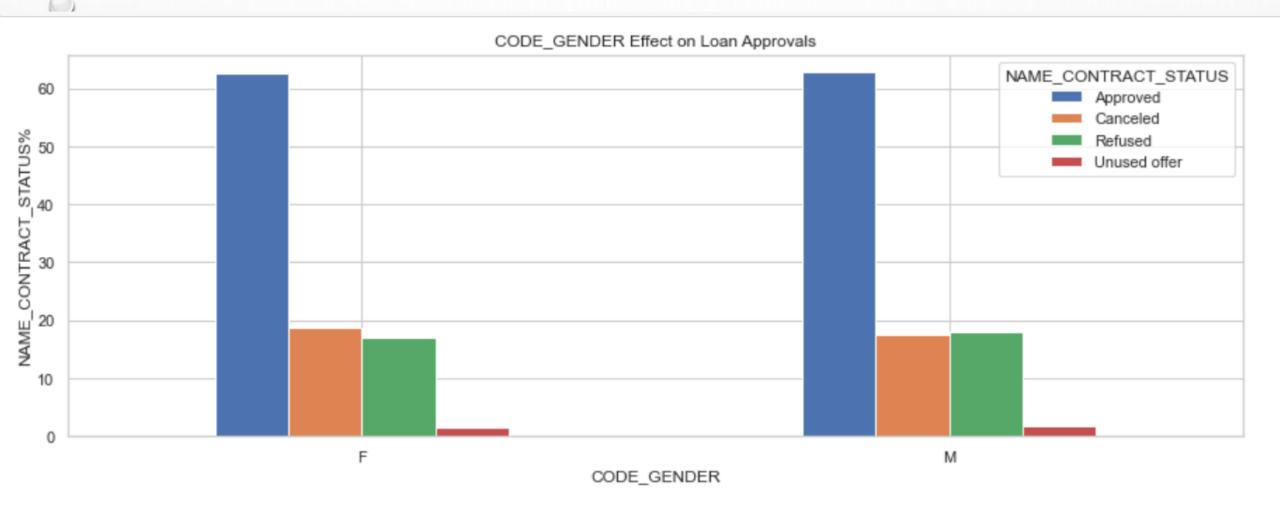
We observe that there are more defaulters who had refused the loan offer previously as compared to non-defaulters.

CODE_GENDER VS NAME_CONTRACT_STATUS

CODE_GENDER Effect on Loan Approvals

We can observe that there is no much difference on the Contract Status with respect to Gender.

CONCLUSION

- BANKS SHOULD CONSIDER INCOME TYPE OF STUDENT , PENSIONER AND BUSINESSMAN FOR SUCCESSFUL PAYMENTS.

- BANKS SHOULD FOCUS LESS ON INCOME TYPE **WORKING** AS THEY ARE HAVING MOST NUMBER OF UNSUCCESSFUL PAYMENTS.

- BANKS SHOULD CONSIDER CLIENTS FROM HOUSING TYPE WITH PARENTS AS TOP PRIORITY AS THEY ARE HAVING LEAST NUMBER OF UNSUCCESSFUL PAYMENTS.

- FEMALE CLIENTS WITH HIGHER EDUCATION HAVE LESS CHANCES OF BEING A DEFAULTER.

- CLIENTS WITH HIGH INCOME CATEGORY CAN BE PREFERRED FOR BANK LOANS.

- GROUP WHOSE PREVIOUS LOAN STATUS WAS REFUSED HAVE SIGNIFICANT CHANCE OF DEFAULTING.

THANK YOU !!