



CREDIT EDA CASE STUDY

BY:

RITESH PATIL

DIBYAJIT DHARA

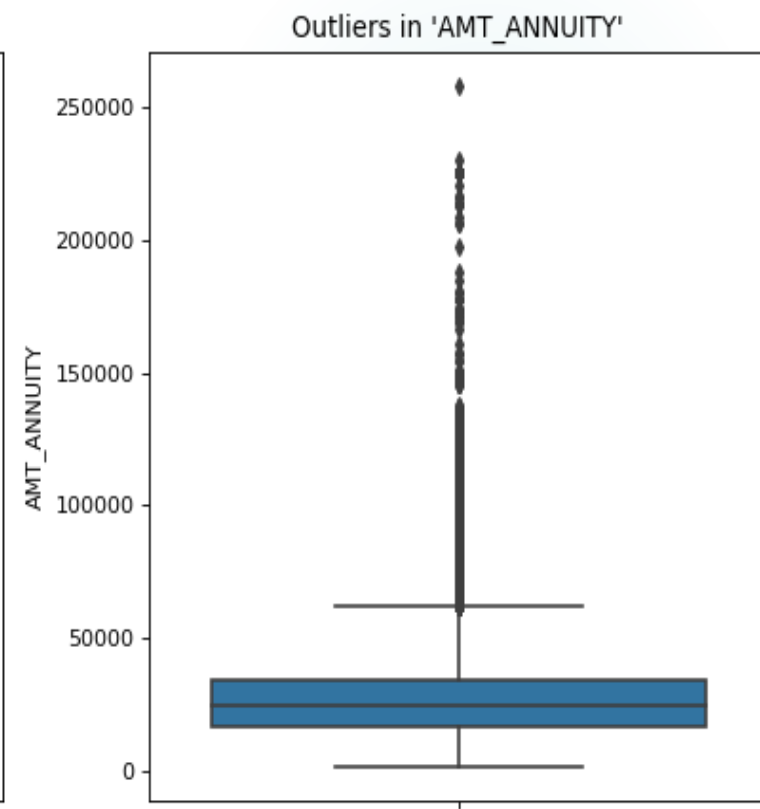
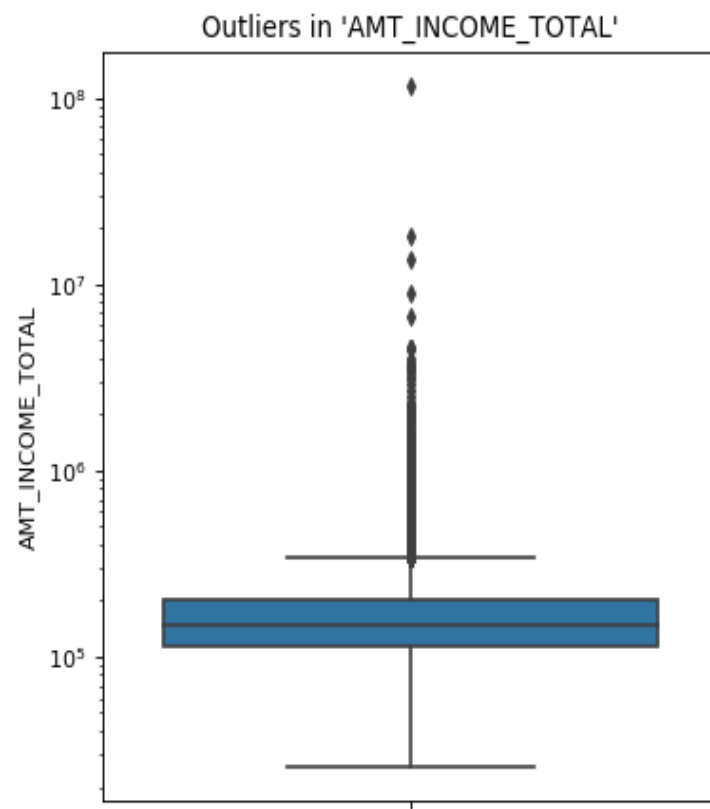
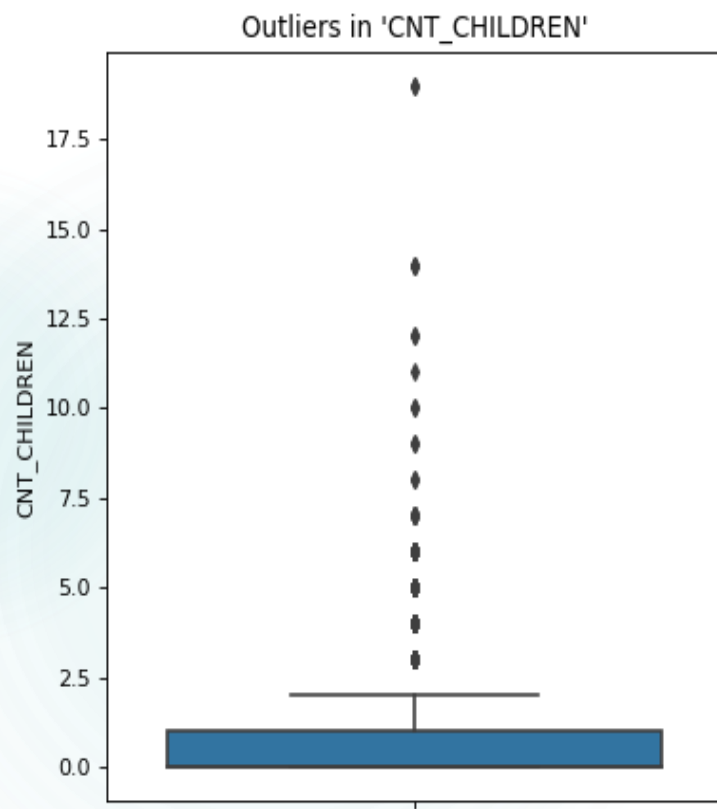
Steps Performed:

- ▶ Import all the required Libraries.
- ▶ Import the "Application Data file" from Local.
- ▶ Doing the Normal Routine Checks.
- ▶ Data Quality Check and Dealing with the Missing Values.
- ▶ Treating Outliers.
- ▶ Binning the columns.
- ▶ Analysis.
- ▶ Univariate Analysis for Categorical and Continuous Variables Based on Target 0 and Target 1.
- ▶ Correlation.
- ▶ Bivariate Analysis for Categorical and Continuous Variables Based on Target 0 and Target 1,
- ▶ Import the "Previous Application Data" file from local.
- ▶ Merge both the files of Application Data and Previous Application Data.
- ▶ Univariate Analysis on Merged Dataset for Categorical and Continuous Variables Based on NAME_CONTRACT_STATUS.
- ▶ Bivariate Analysis on Merged Dataset for Categorical and Continuous Variables Based on NAME_CONTRACT_STATUS.

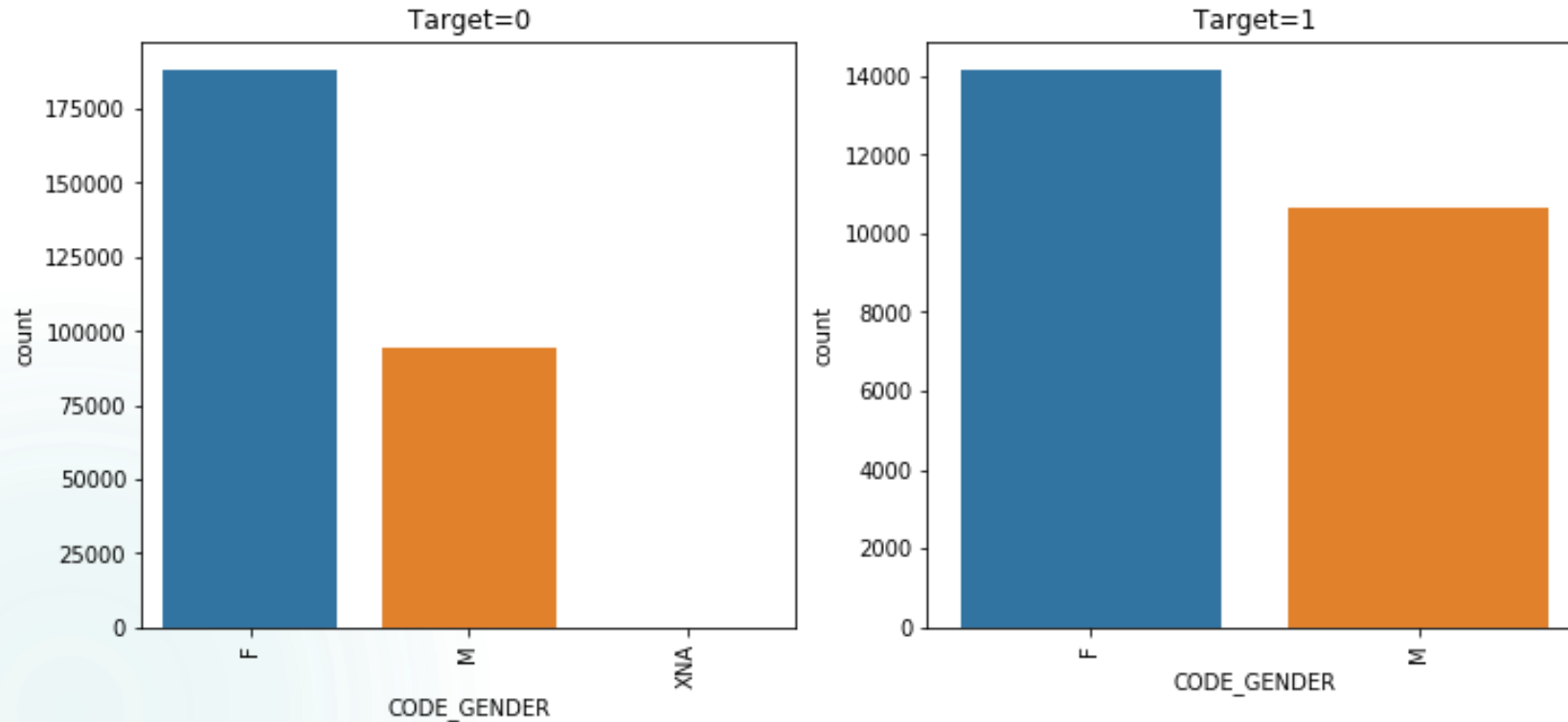
Data Quality Check and Dealing with the Missing Values:

- ▶ Finding the Percentage of Missing Values in all the Columns.
- ▶ Removing columns with more than 50% missing values.
- ▶ Filtering columns with missing percentage values less than 13% and then creating a new DataFrame "less_miss_df".

Treating Outliers



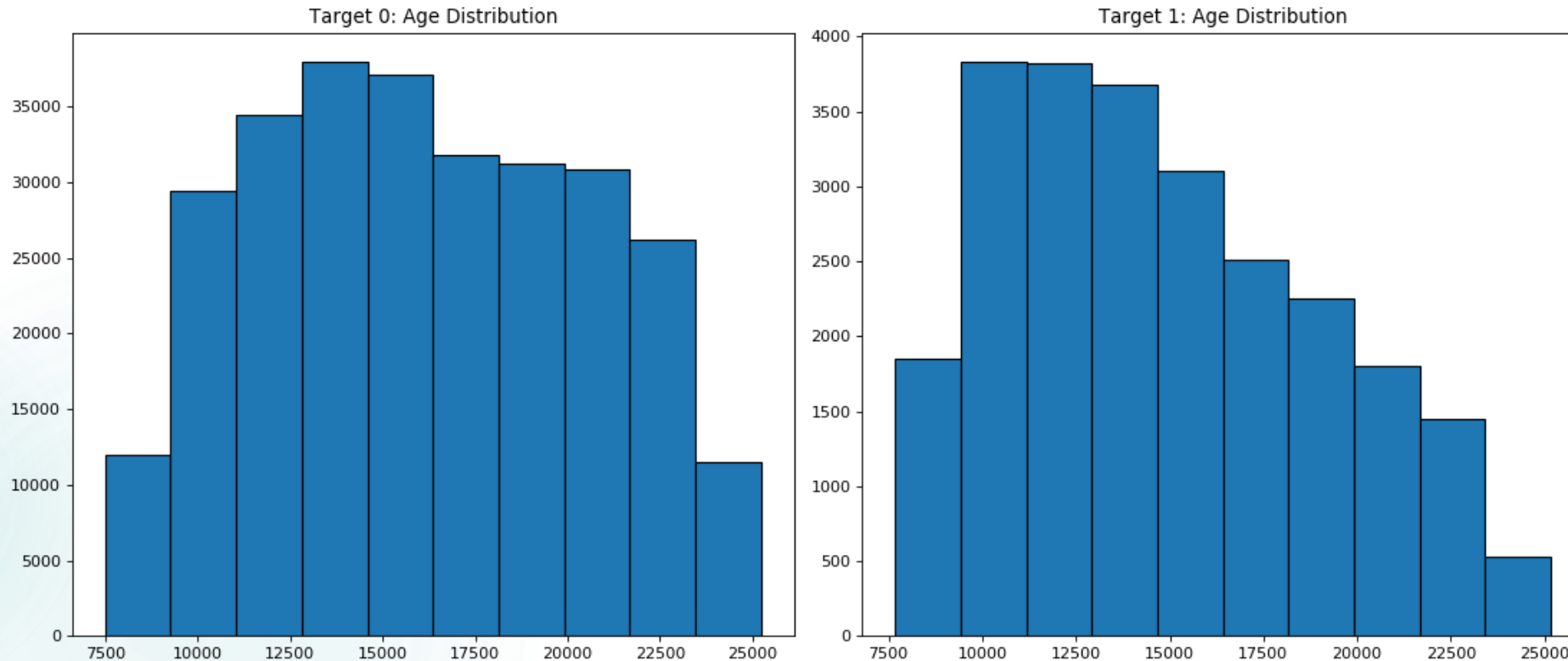
CODE_GENDER



Insight we got:

- ▶ Gender of applicants is one of the categorical variables selected for univariate analysis, which is a fairly important factor. The above plot clearly indicates that the number of Female applicants is higher than males in either case (i.e. Target 0 and Target 1).

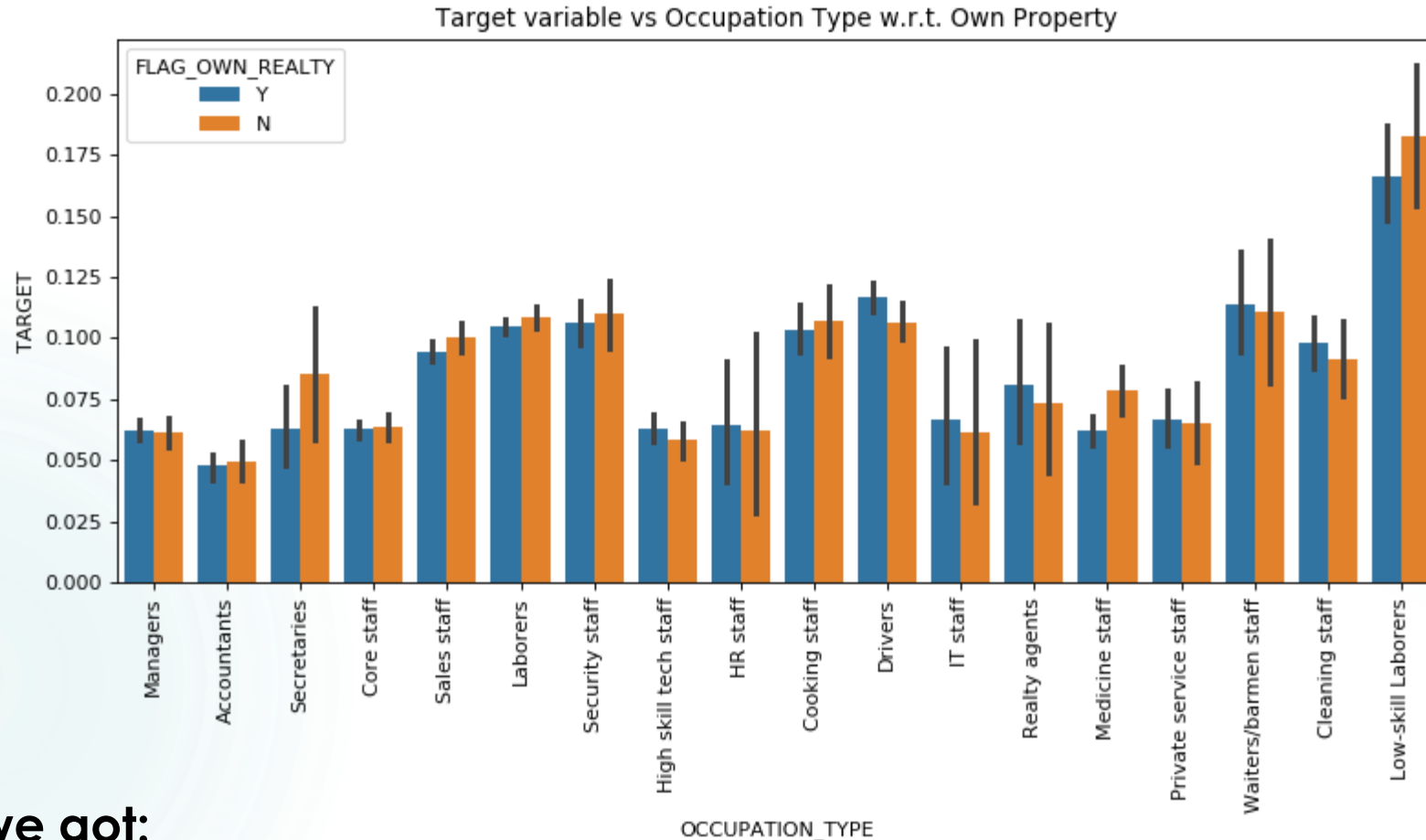
DAYS_BIRTH



Insight we got:

The age trend demonstrates quite simply that individuals between the ages of 27 – 35 (i.e. 10,000 to 12,500 days) earn the largest amount of loan approvals.

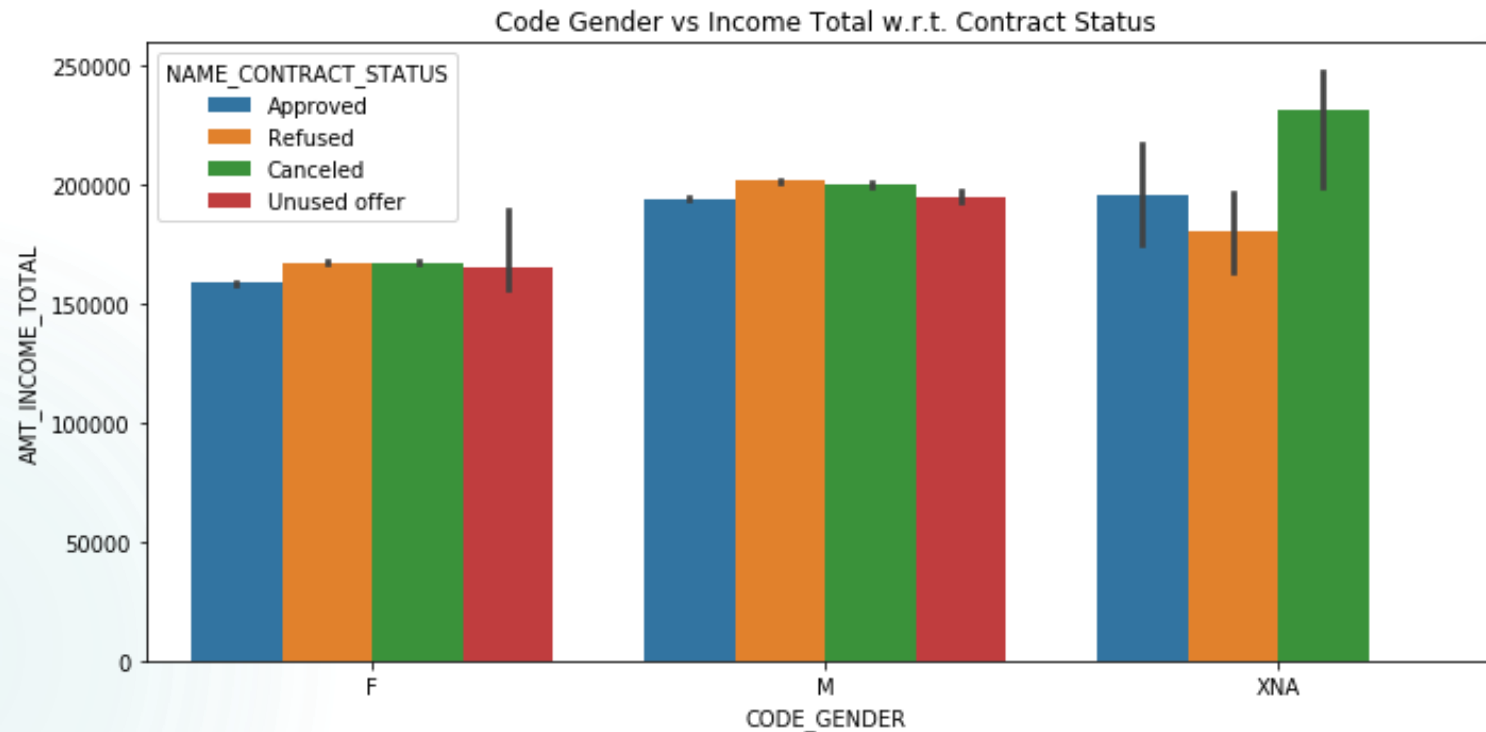
Target plot against OCCUPATION_TYPE



Insight we got:

- Bivariate target plot against type of client occupation tells us that low-skilled workers who do not own property tend to skip paying their loan on time.

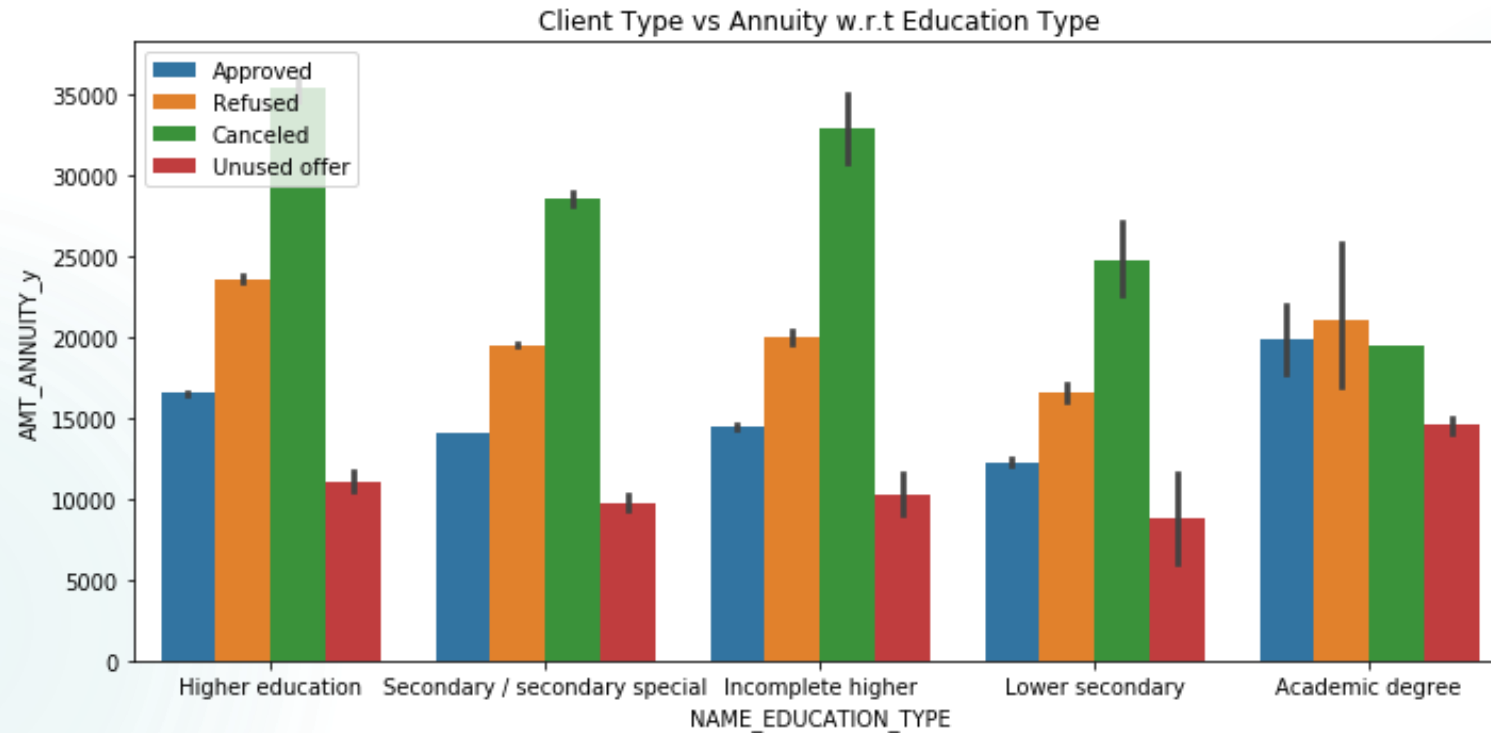
Contract Status vs Income Total w.r.t. Gender



Insight we got:

- ▶ Visualization reveals that all four contract statuses appear to level out around the 200,000 earnings point for male candidates.

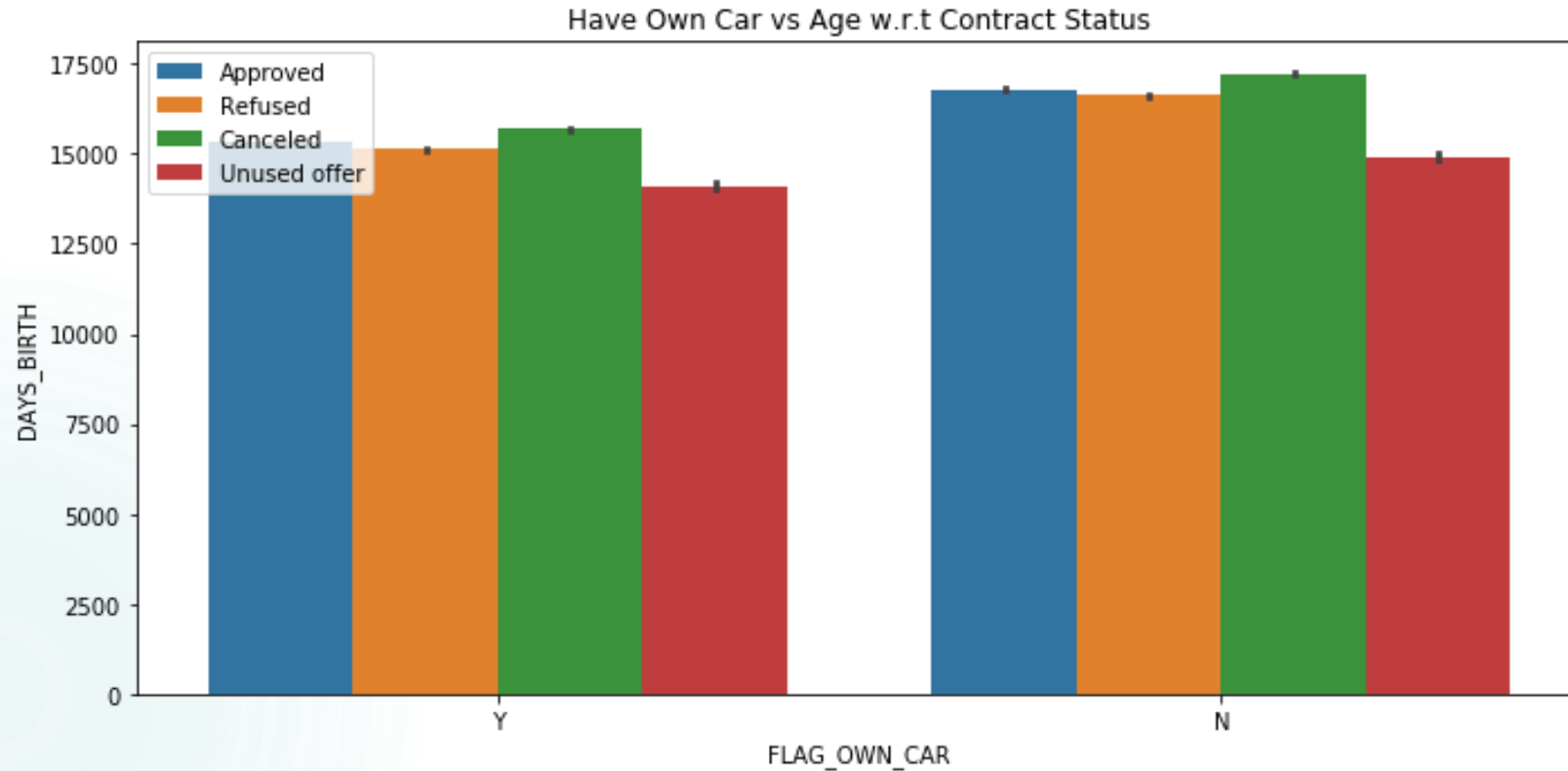
Client Type vs Annuity w.r.t Education Type:



Insight we got:

- ▶ Customers with an academic degree are more likely to get the approval of loan.

Insurance vs Age w.r.t Own Car:



Insight we got:

- ▶ Applicants who do not own a car tend to get approval for the loan.

Correlation For Target 0

VAR1	VAR2	CORR
AMT_GOODS_PRICE	AMT_CREDIT	0.986880
AMT_GOODS_PRICE	AMT_ANNUITY	0.776263
AMT_ANNUITY	AMT_CREDIT	0.771309
DAYS_EMPLOYED	DAYS_BIRTH	0.618048
AMT_ANNUITY	AMT_INCOME_TOTAL	0.418953
AMT_GOODS_PRICE	AMT_INCOME_TOTAL	0.349363
AMT_CREDIT	AMT_INCOME_TOTAL	0.342799
DAYS_REGISTRATION	DAYS_BIRTH	0.333151
DAYS_ID_PUBLISH	DAYS_EMPLOYED	0.274290
DAYS_ID_PUBLISH	DAYS_BIRTH	0.271314

Correlation for Target 1

VAR1	VAR2	CORR
AMT_GOODS_PRICE	AMT_CREDIT	0.982566
AMT_ANNUITY	AMT_CREDIT	0.752195
AMT_GOODS_PRICE	AMT_ANNUITY	0.752022
DAYS_EMPLOYED	DAYS_BIRTH	0.575097
DAYS_REGISTRATION	DAYS_BIRTH	0.289114
DAYS_ID_PUBLISH	DAYS_BIRTH	0.252863
DAYS_ID_PUBLISH	DAYS_EMPLOYED	0.226470
DAYS_REGISTRATION	DAYS_EMPLOYED	0.188929
DAYS_BIRTH	AMT_GOODS_PRICE	0.135754
DAYS_BIRTH	AMT_CREDIT	0.135316

conclusion

From all our analyses and visualizations so far, the following have emerged as variables that contribute most towards a borrower defaulting:

- ▶ Occupation Type.
- ▶ Education Type.
- ▶ Income Type.
- ▶ Age.
- ▶ Family Status.
- ▶ Days employed.
- ▶ Own Realty.
- ▶ Count of children.



Thank you!