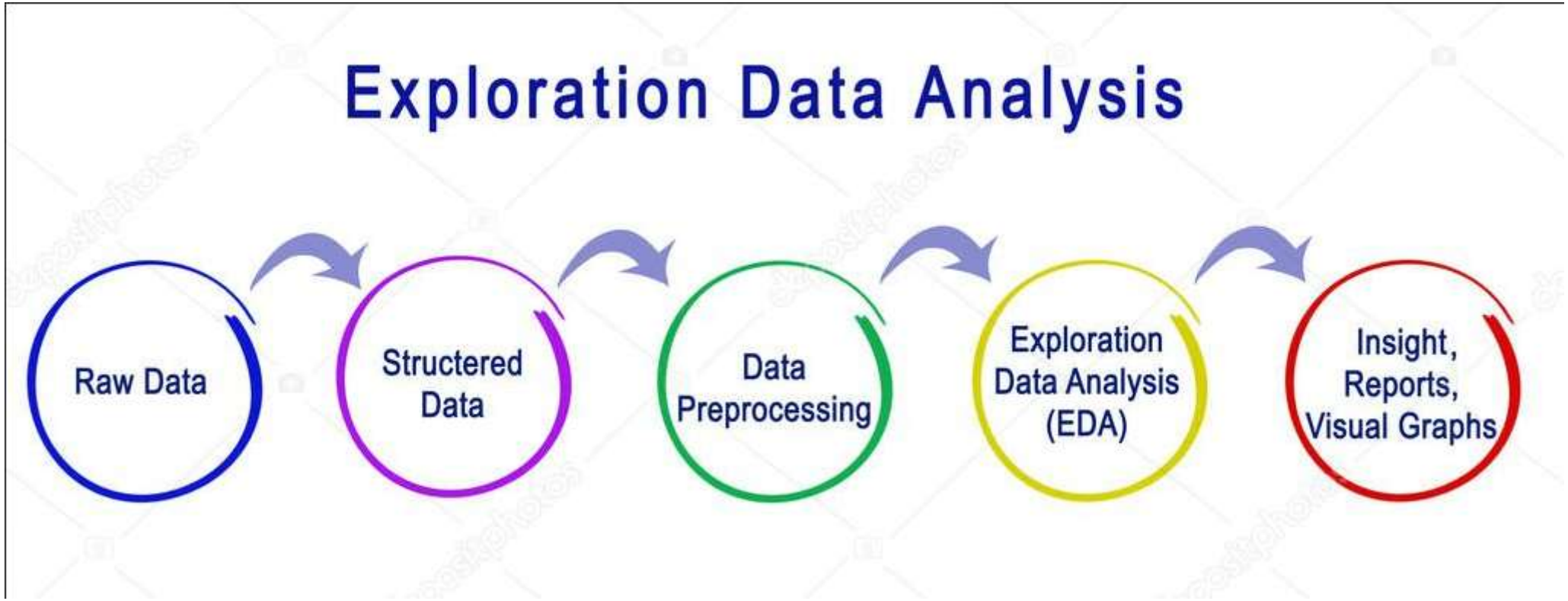
Analyzing loan Data application

Problem statement:

A "Consumer Finance Company" is a business that specializes in giving loans to clients. They want to comprehend the trends among clients who have trouble making instalment payments. These characteristics can assist in recognizing these loan applications, which can then result in loan denial, loan reduction, lending (at a higher interest rate), and, ultimately, an improvement in analysing risk assessment.


Exploration Data Analysis
Raw Data
Structered Data
Data Preprocessing
Exploration Data Analysis (EDA)
Insight, Reports, Visual Graphs

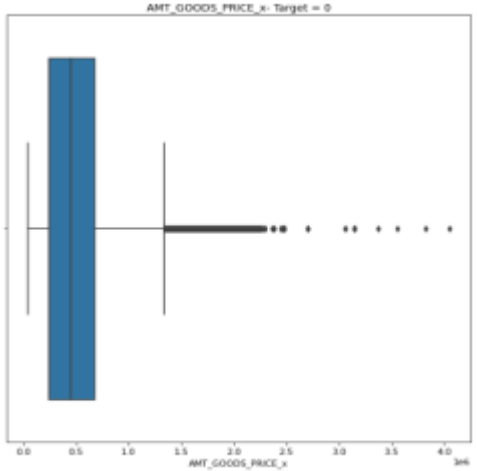
Steps to analyse:

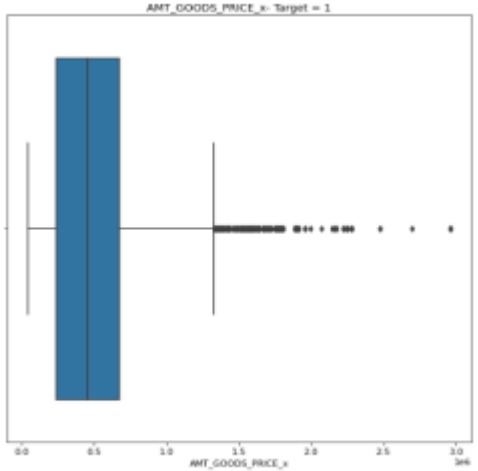
- Exploratory Data analysis for understanding Data set ,cleaning the Data set (Deleting unnecessary columns and columns having most of the null values),Treating outliers, Imputation.
- Identify the key columns and perform Univariate, Bivariate for both numerical and categorical variables.
- Important variables in data set are:

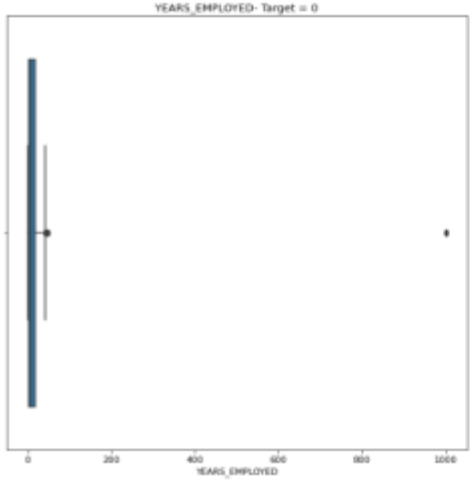
- *TARGET:*client with payment difficulties: he/she had late payment more than X days on at least one of the first Y installments of the loan in the sample.
- *NAME_CONTRACT_STATUS:* Contract status (approved, cancelled, ...) of previous application

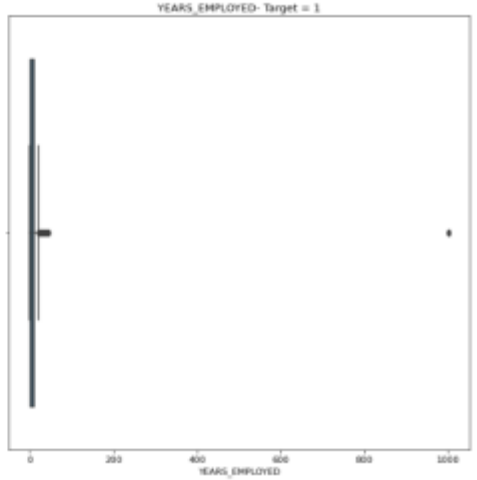
Analysing categorical and Numerical columns with Target variable:

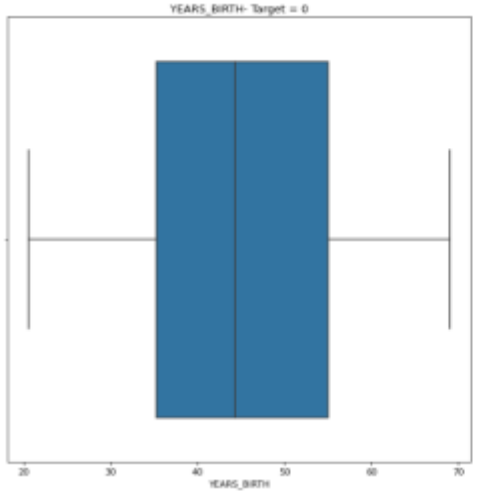
Univariate Analysis:

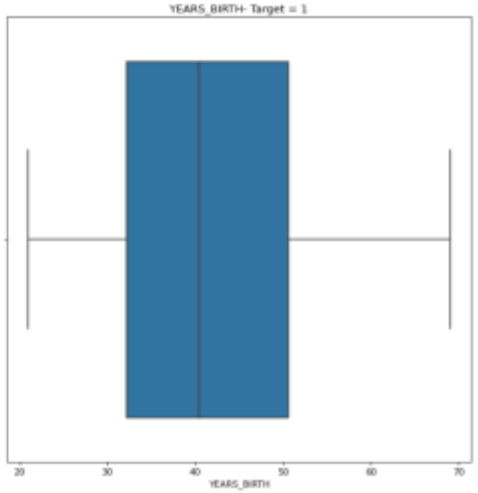


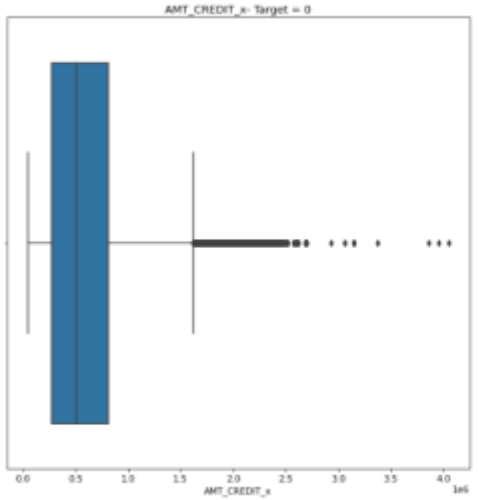


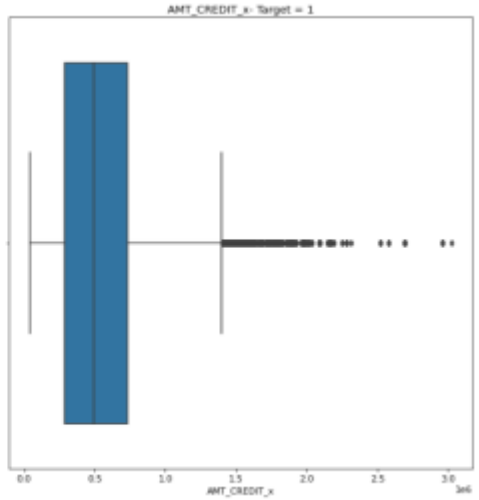






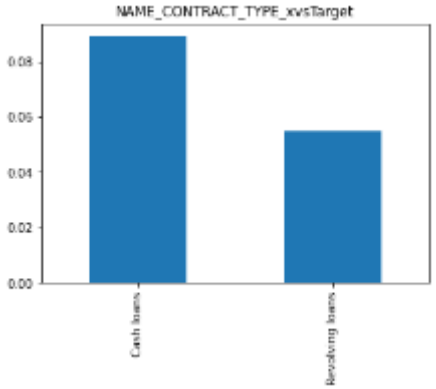


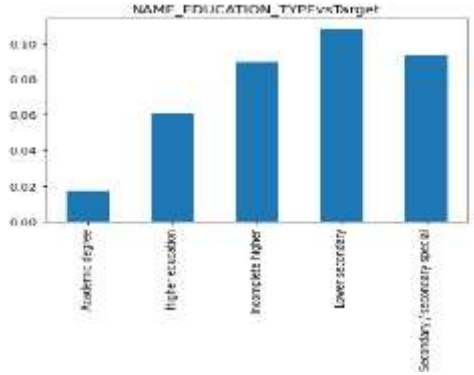


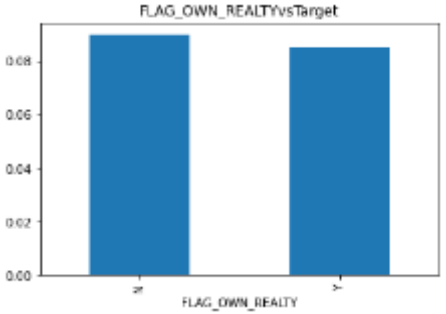


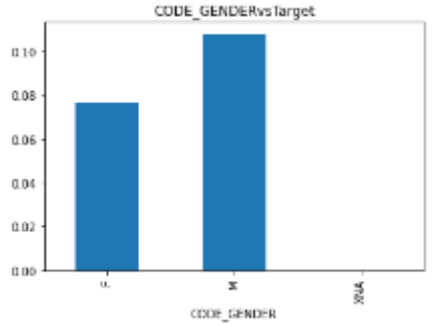
Bivariate analysis:

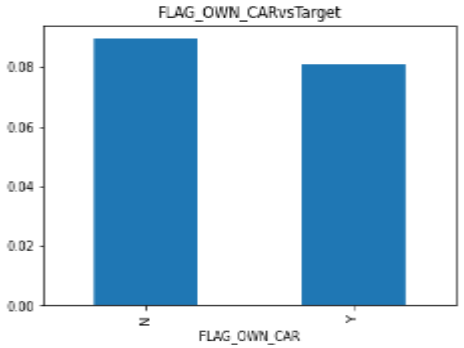
Analysing which category of variable having high default rate.

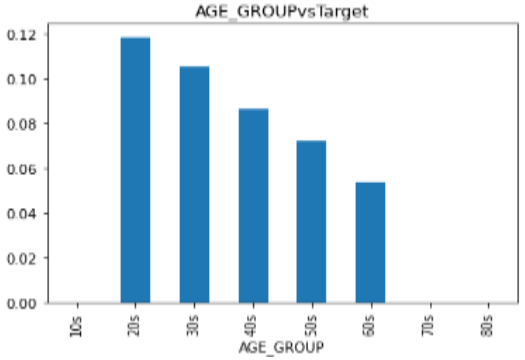


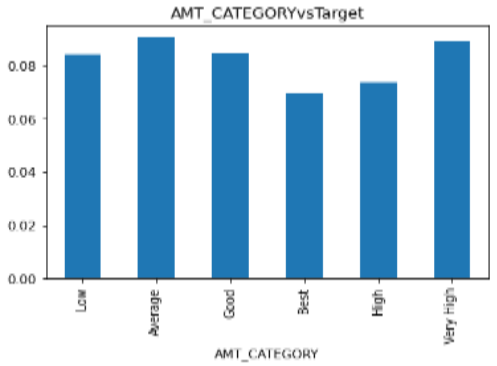












Correlation for Numerical columns:

To analyse which variables are having more correlation

Correlation analysis for Target value 0

	VAR1	VAR2	Correlation	Correlation_abs
1216	AMT_GOODS_PRICE_y	AMT_APPLICATION	1.00	1.00
502	OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	1.00	1.00
250	AMT_GOODS_PRICE_x	AMT_CREDIT_x	0.99	0.99
1217	AMT_GOODS_PRICE_y	AMT_CREDIT_y	0.99	0.99
1175	AMT_CREDIT_y	AMT_APPLICATION	0.98	0.98
377	REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.94	0.94
1637	DAYS_TERMINATION	DAYS_LAST_DUE	0.93	0.93
289	CNT_FAM_MEMBERS	CNT_CHILDREN	0.88	0.88
544	DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.86	0.86
1174	AMT_CREDIT_y	AMT_ANNUITY_y	0.82	0.82

We explore data sets for target =0 to check if top 10 correlated pair of variables are same across both the data sets

Correlation analysis for Target value 1

	VAR1	VAR2	Correlation	Correlation_abs
1216	AMT_GOODS_PRICE_y	AMT_APPLICATION	1.00	1.00
502	OBS_60_CNT_SOCIAL_CIRCLE	OBS_30_CNT_SOCIAL_CIRCLE	1.00	1.00
1217	AMT_GOODS_PRICE_y	AMT_CREDIT_y	0.99	0.99
1175	AMT_CREDIT_y	AMT_APPLICATION	0.98	0.98
250	AMT_GOODS_PRICE_x	AMT_CREDIT_x	0.98	0.98
377	REGION_RATING_CLIENT_W_CITY	REGION_RATING_CLIENT	0.96	0.96
1637	DAYS_TERMINATION	DAYS_LAST_DUE	0.94	0.94
1552	DAYS_LAST_DUE_1ST_VERSION	DAYS_FIRST_DRAWING	-0.89	0.89
289	CNT_FAM_MEMBERS	CNT_CHILDREN	0.89	0.89
544	DEF_60_CNT_SOCIAL_CIRCLE	DEF_30_CNT_SOCIAL_CIRCLE	0.86	0.86

We explore data sets for target =1 to check if top 10 correlated pair of variables are same across both the data sets

Conclusion:

As per the analysis customers likely to default are:

- NAME_CONTRACT_TYPE - Cash loans are more likely to default
- CODE_GENDER – Male customers are more likely to default
- NAME_EDUCATION_TYPE – Customers with Academic degree are less likely to default
- AGE_GROUP – Customers of age between 19-29 are more likely to default
- If AMOUNT_CREDIT is more than the AMOUNT_ANNUITY then check twice before giving the loan
- Compare AMOUNT_GOODS_PRICE and AMOUNT_CREDIT before approving the loan.