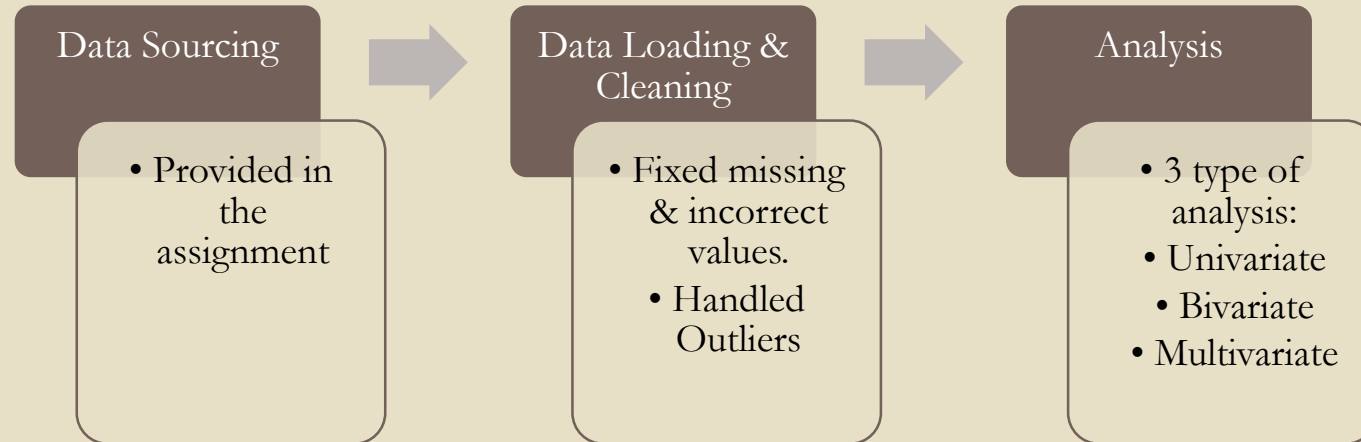# CREDIT EDA: CASE STUDY

By RISHIKA BANSAL

Problem Statement

- To identify clients as defaulter or approved based upon their history & capability to repay the loans on time.

- To identify and understand the strong driving factors of defaulters.

Purpose & Objective

◦ This EDA analysis will help the company to make decisions on loan approval, i.e. people who cannot repay the loan installments can be considered for taking actions and the one who can repay can be accepted.

◦ This will help them control losses in business.

Approach

Data Sourcing

- Provided in the assignment

Data Loading & Cleaning

- Fixed missing & incorrect values.
- Handled Outliers

Analysis

- 3 type of analysis:
- Univariate
- Bivariate
- Multivariate

Methodology: Data Cleaning

- Columns with more than 40% missing values were dropped (around ~49 columns)

- Not removing few missing value columns to maintain biasness of the data

- Imputing some missing values using mode and median to handle outliers.

- 5 Outliers were handled separately

- Correcting incorrect values like negative days data into positive, etc.

Methodology: Data Analysis

- Data was analyzed on 3 sets of data, i.e.
    - Application data
    - Previous application data
    - Merged Data
- Univariate, Bivariate and Multivariate Analysis was carried out on all 3 sets of data separately
- Correlation among continuous variables of defaulter & non-defaulters data was found and represented using heat maps.

Key Indicators of Default

○ Previously refused loans: Analysis shows the previously rejected clients find it difficult to pay the loans and according to numbers ~12% are refused loans.

○ Client's age: Loan approval for age 17-35 years is not advisable. 55+ is advisable according to the analysis.

○ Client's Family status: The least payment difficulties is from widows than unmarried & civil married clients.

○ Contract Type: Cash loans %age for payment difficulties is ~10% while ~6% for revolving loans

○ Client's gender: ~12% of males were found to have payment difficulties, whereas females just had ~8%.

○ Client's education level: ~12% of people who finished education at lower secondary have difficulty in paying loans.

○ Loan amount: ~11% of clients face difficulties when amount range is in between 4 to 6 lacs.
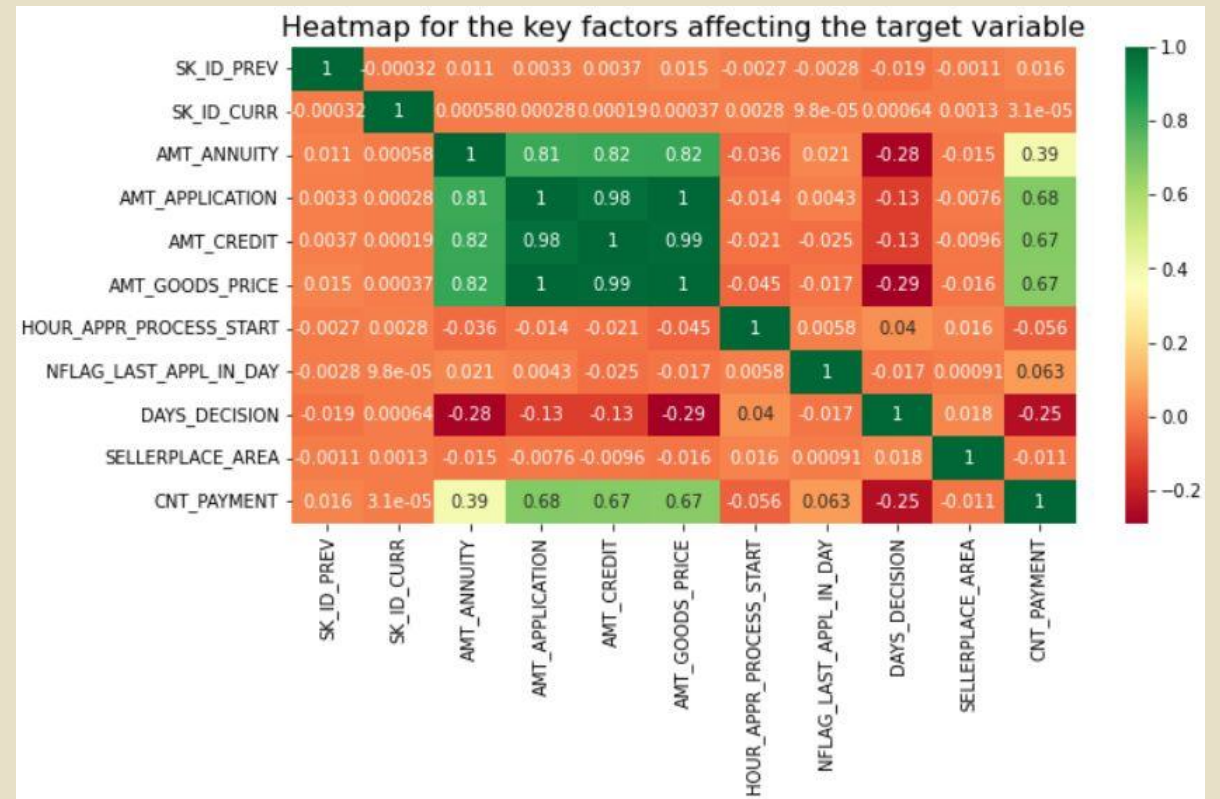
Observations: Correlation of Application data

- Comparing top 10 correlations b/w Payment difficulties and On-Time Payments

- Highest correlation is between GOOD_PRICE AND AMT_CREDIT

- 0.987 is for on time payment correlation

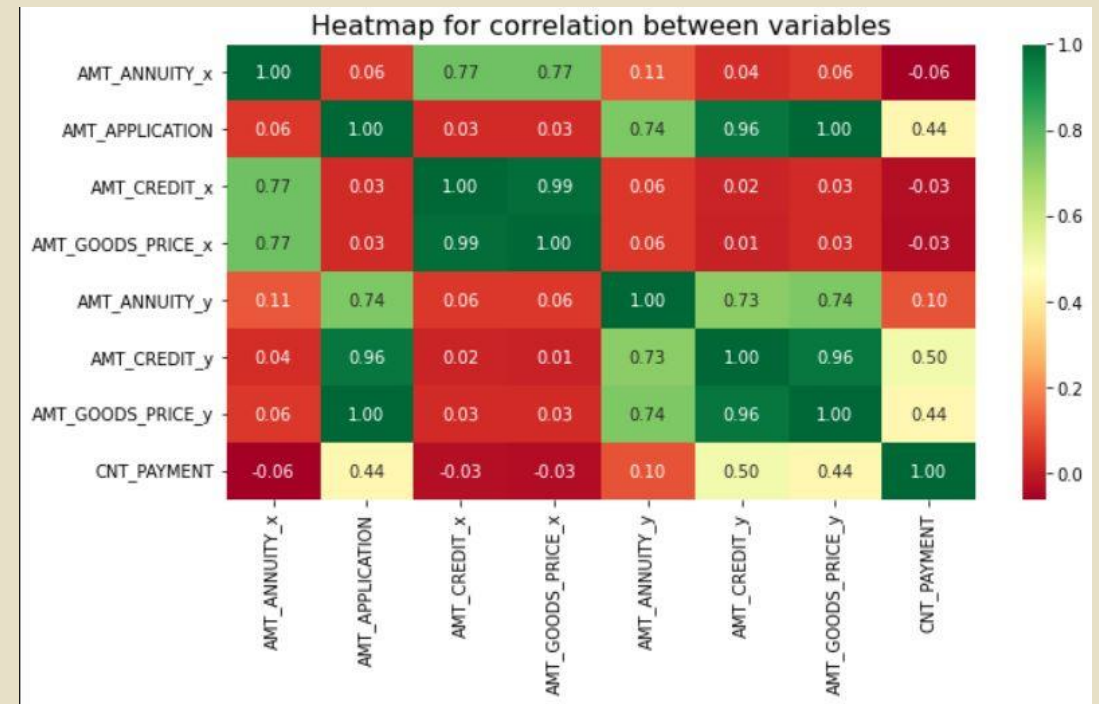- 0.982 is for payment difficulties correlation
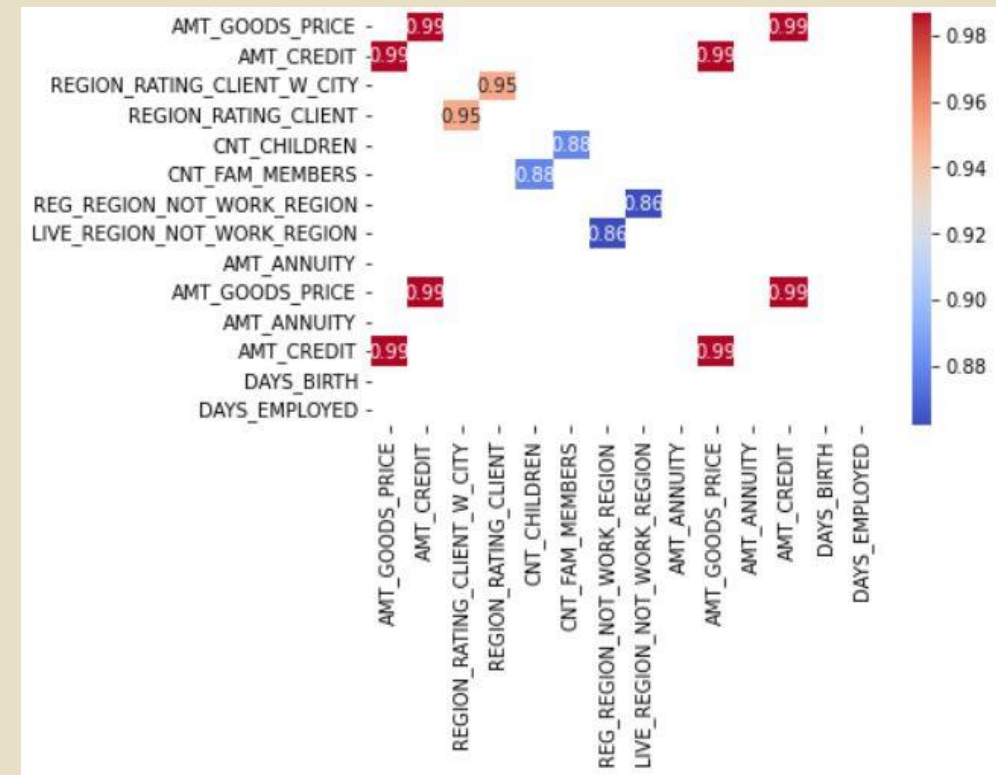
Observations: Correlation of Previous Application data

- Goods, credit and number of application are highly correlated with amount annuity.

- Credit amount is highly correlated with amount application.

- Goods price and days of decision is negatively correlated.

Heatmap for the key factors affecting the target variable

Observations: Correlation of merged data

○ AMT_APPLICATION has a high correlation with AMT_ANNUITY_y,AMT_CREDIT_y,AMT_GOODS_PRICE_y and decent correlation with CNT_PAYMENT

○ AMT_GOODS_PRICE_y has a high correlation with AMT_ANNUITY_y,AMT_CREDIT_y,AMT_APPLICATION and decent correlation with CNT_PAYMENT

○ AMT_CREDIT_y has a high correlation with AMT_GOODS_PRICE_y and decent correlation with CNT_PAYMENT

○ AMT_ANNUITY_x has a high correlation with AMT_GOODS_PRICE_y,AMT_CREDIT_y

○ AMT_ANNUITY_x has a high correlation with AMT_GOODS_PRICE_x,AMT_CREDIT_x

○ AMT_CREDIT_x has a high correlation with AMT_GOODS_PRICE_x

Heatmap for correlation between variables

	AMT_ANNUITY_x	AMT_APPLICATION	AMT_CREDIT_x	AMT_GOODS_PRICE_x	AMT_ANNUITY_y	AMT_CREDIT_y	AMT_GOODS_PRICE_y	CNT_PAYMENT
AMT_ANNUITY_x	1.00	0.06	0.77	0.77	0.11	0.04	0.06	-0.06
AMT_APPLICATION	0.06	1.00	0.03	0.03	0.74	0.96	1.00	0.44
AMT_CREDIT_x	0.77	0.03	1.00	0.99	0.06	0.02	0.03	-0.03
AMT_GOODS_PRICE_x	0.77	0.03	0.99	1.00	0.06	0.01	0.03	-0.03
AMT_ANNUITY_y	0.11	0.74	0.06	0.06	1.00	0.73	0.74	0.10
AMT_CREDIT_y	0.04	0.96	0.02	0.01	0.73	1.00	0.96	0.50
AMT_GOODS_PRICE_y	0.06	1.00	0.03	0.03	0.74	0.96	1.00	0.44
CNT_PAYMENT	-0.06	0.44	-0.03	-0.03	0.10	0.50	0.44	1.00

Recommendations

- Avoid sanctioning loan to the clients whose application have been rejected before.

- Within the age of 17 to 35, loan sanctioning should be avoided and should be sanctioned to age group of 55+.

- Widows should be sanctioned loans and unmarried & civil married clients should be avoided.

- There is a lower risk w.r.t. female clients than male clients, so loan sanctioning should be preferred to them.

- Clients having lower education level should be avoided because of bad risk.

- Loan amounts in between 4 to 6 lacs should be avoided as it is the riskiest range.

- Sanctioning revolving loans are better than cash loans as banks risk rates are less.

Conclusions

After the analysis also, the banks will have to keep a close look on the above mentioned key factors. This will ensure better business opportunity & growth, preventing banks to take 'default' decisions.

Thank you!