BANK LOAN CASE STUDY

BY VISHAL SINGH SANGRAL

APPROACH

- Data cleaning
- Removing duplicates and imputing missing values
- Data modelling
- Data visualizing
- Creation of report
- Presentation video

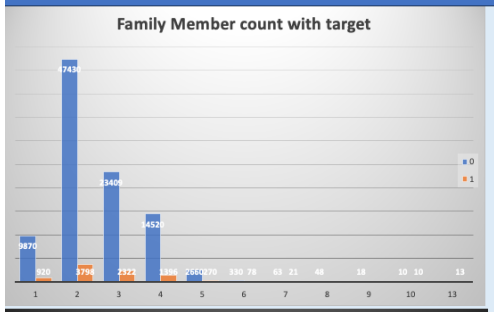
TECH STACK USED

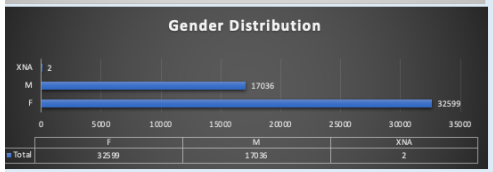
- Excel desktop 365
- Microsoft power point 365

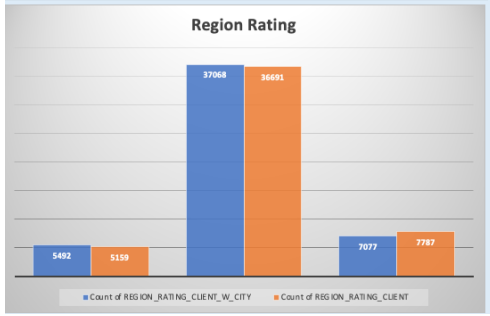
- A lot of unwanted columns and missing values in the data set
- Prescence of outlier
- Positive correlation has been scene between AMt_ANNUITY, AMT_CREDIT AND AMT_TOTAL_INCOME
- High class imbalance in the dataset is visible

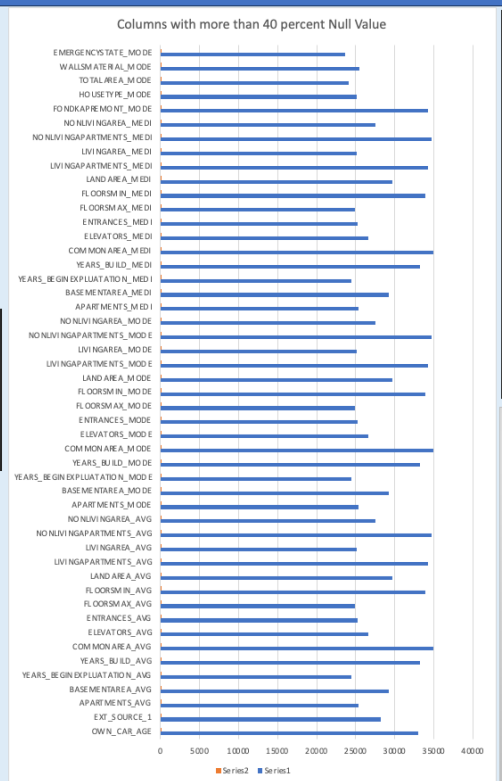
INSIGHTS

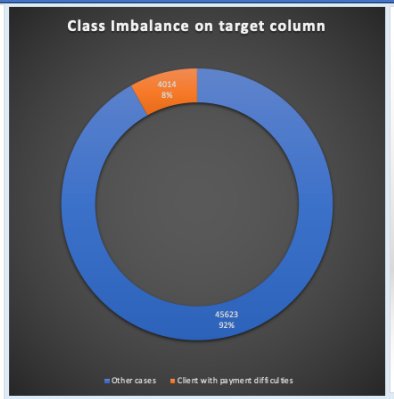
BANK LOAN CASE STUDY DASHBOARD

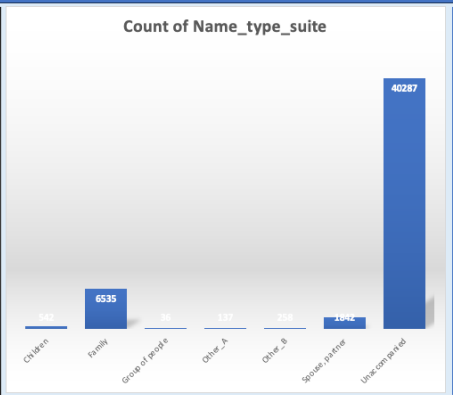


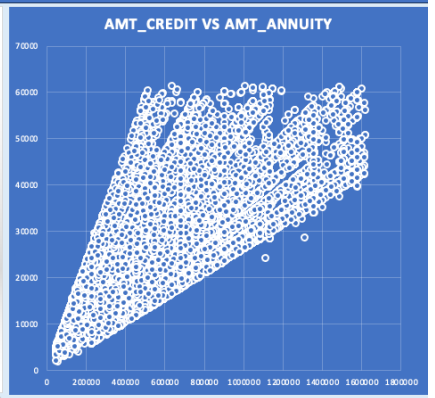


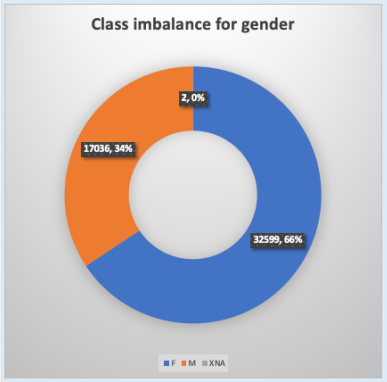


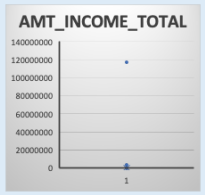


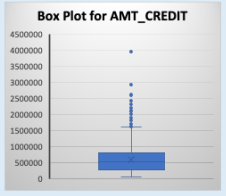


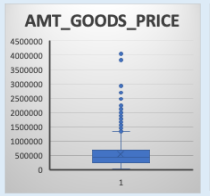


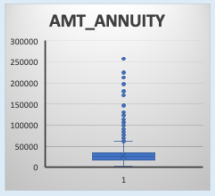


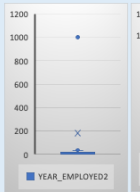


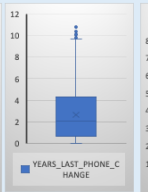


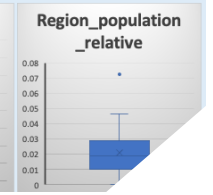


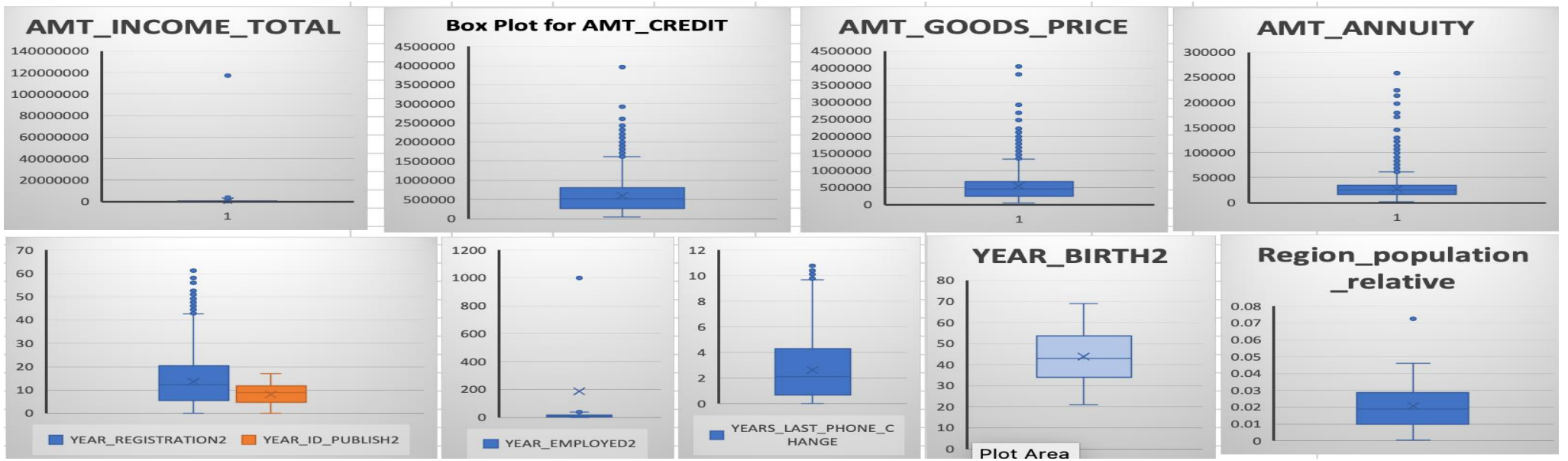








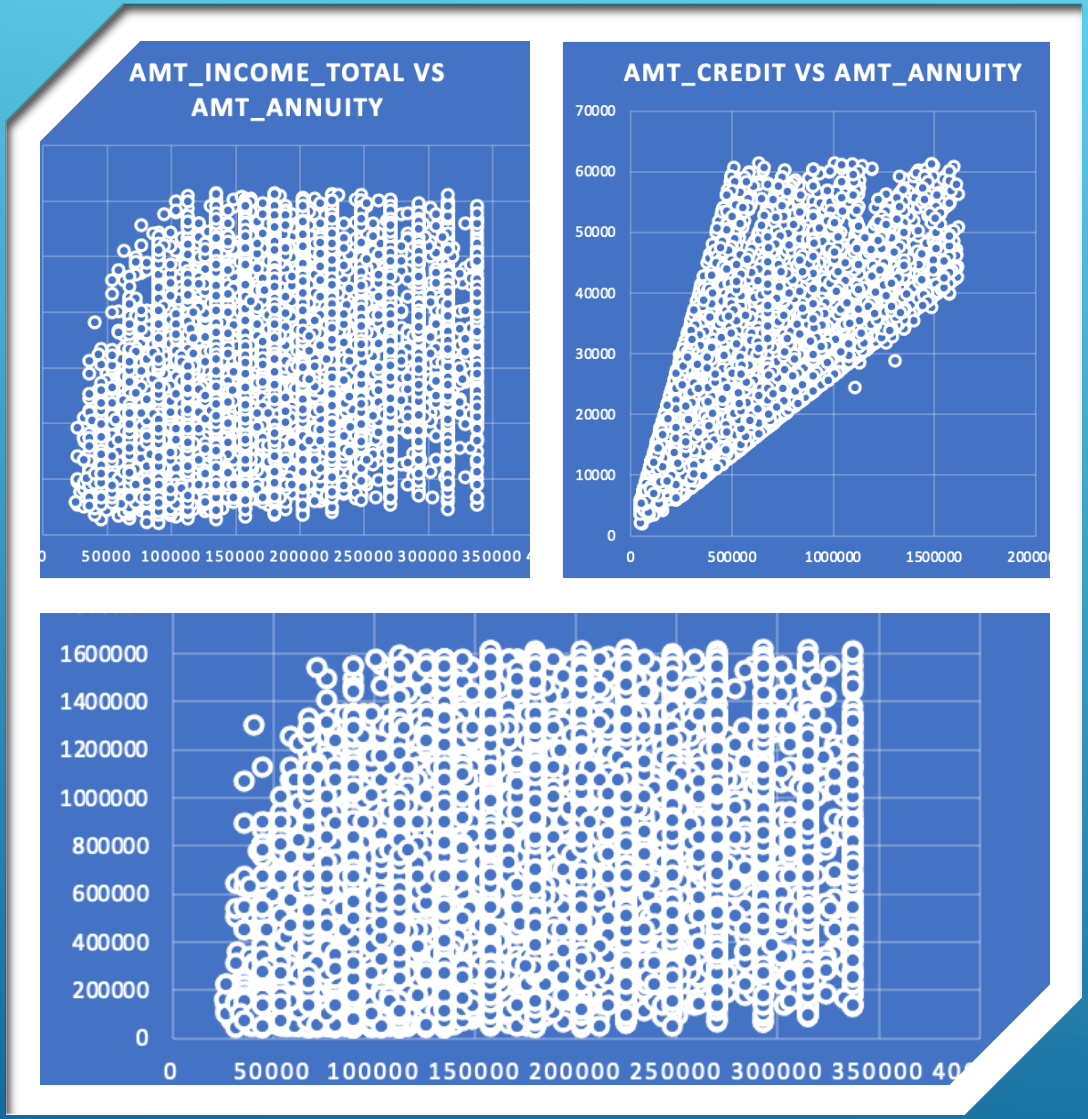




PRESCENCE OF OUTLIER

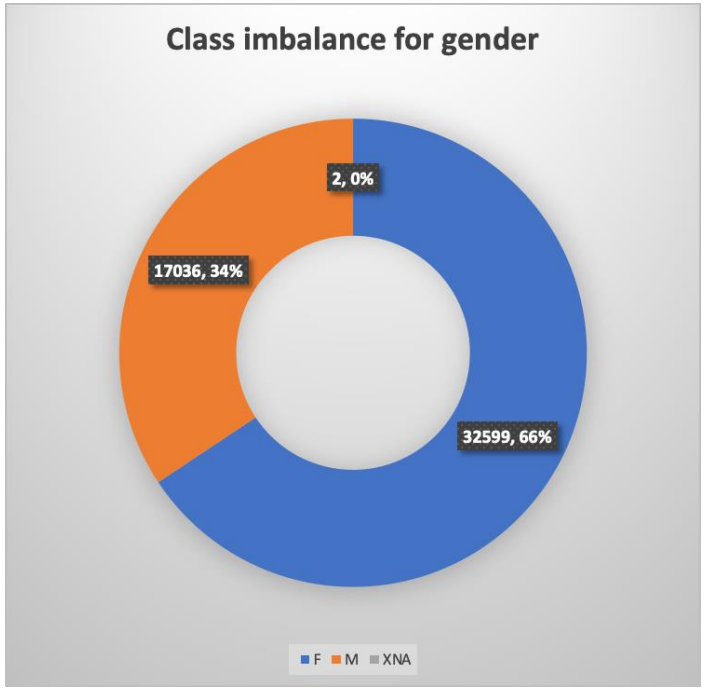
SAMPLE COLUMNS WITH MORE THAN 40 PERCENT OF MISSING VALUE

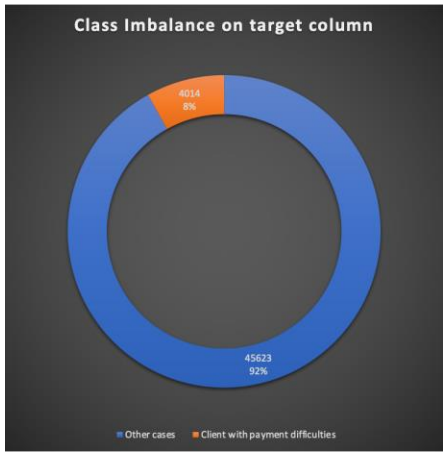
COLUMN TO DROP AS THE NULL VALUE PERCENTAGE IS GREATER THAN 40%		
column Name	**Null values counts**	**Null Value Percentage**
OWN_CAR_AGE	32950	65.90%
EXT_SOURCE_1	28172	56.35%
APARTMENTS_AVG	25385	50.77%
BASEMENTAREA_AVG	29199	58.40%
YEARS_BEGINEXPLUATATION_AVG	24394	48.79%
YEARS_BUILD_AVG	33239	66.48%
COMMONAREA_AVG	34960	69.92%
ELEVATORS_AVG	26651	53.30%
ENTRANCES_AVG	25195	50.39%
FLOORSMAX_AVG	24875	49.75%
FLOORSMIN_AVG	33894	67.79%
LANDAREA_AVG	29721	59.44%
LIVINGAPARTMENTS_AVG	34226	68.45%
LIVINGAREA_AVG	25137	50.28%
NONLIVINGAPARTMENTS_AVG	34714	69.43%



COLUMNS WITH MODERATE TO HIGH POSITIVE CORRELATION

COLUMNS WITH HIGH IMBALANCE STRUCTURE





RESULTS

- After analyzing the dataset, I've observed a correlation between individuals with higher incomes and their ability to repay loans. Additionally, there's a notable relationship between clients with larger credit amounts and their annuity repayment amounts, suggesting a higher monetary circulation in the market among these clients.

DRIVE LINK

https://drive.google.com/drive/folders/1wYnMayzE_KCTWfEGgqdd8grz_OCZDbx1?usp=sharing

THANK YOU