BANK LOAN CASE STUDY

ANALYSIS

INTRODUCTION

A finance compnay provides loans, including to those without enough credit history resulting in defualters and delinquents.

Goal is to analyze application patterns to minimize risks:

- Rejecting able applicants
- Approving incapable applicant

Dataset has applications with:

- Payment difficulties (Target 1)
- On-time payments (Target 0)

4 outcome when applying:

- Approved
- Cancelled
- Refused
- Unused offer

EDA used to understand how different attributes influence target which will help make informed decisions and reduce risk

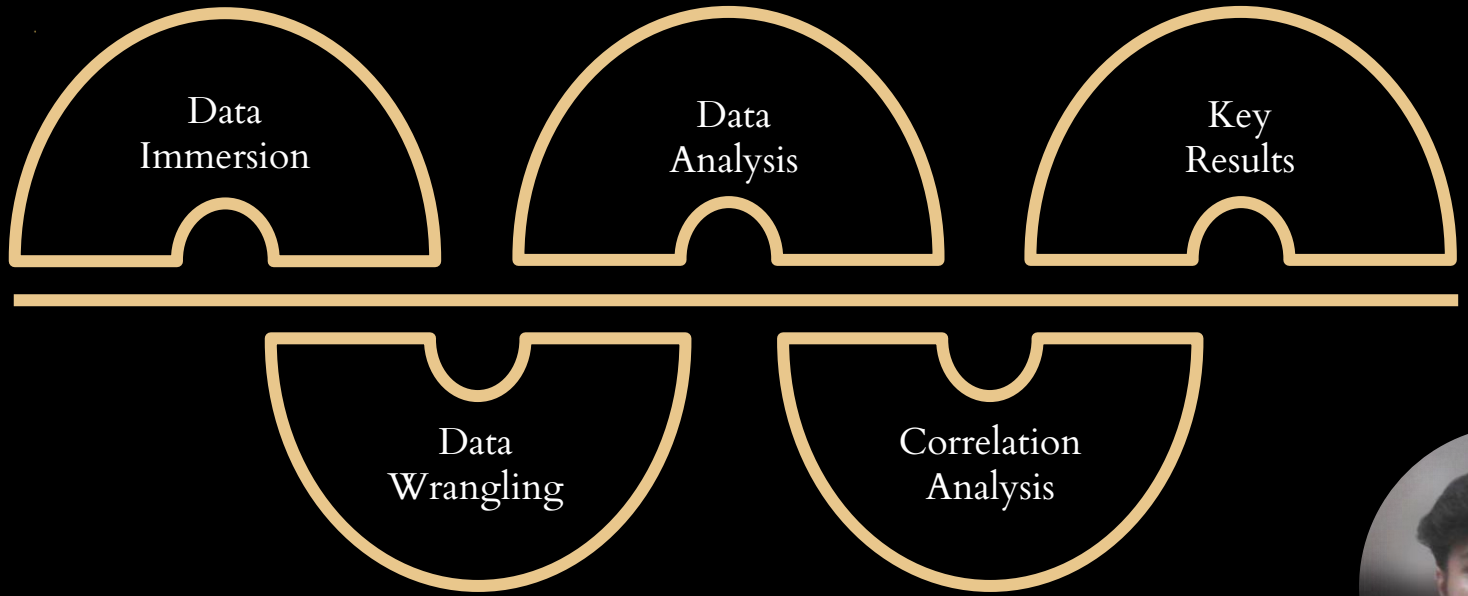
TECH STACK

MS-Excel 365

Link to all Excel files (Ctrl+click)

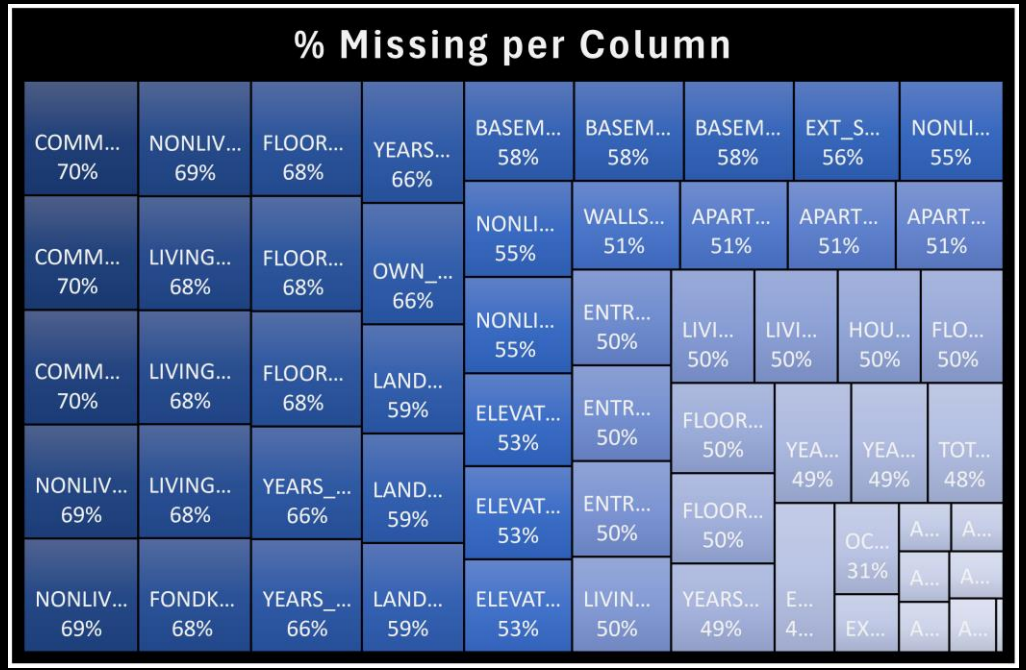
- Outlier and missing values – contains analysis for outlier detection & missing values
- Analysis – contains all the univariate and bivariate analysis
- Correlation – contains previous and current application merged data and correlation analysis

APPROACH

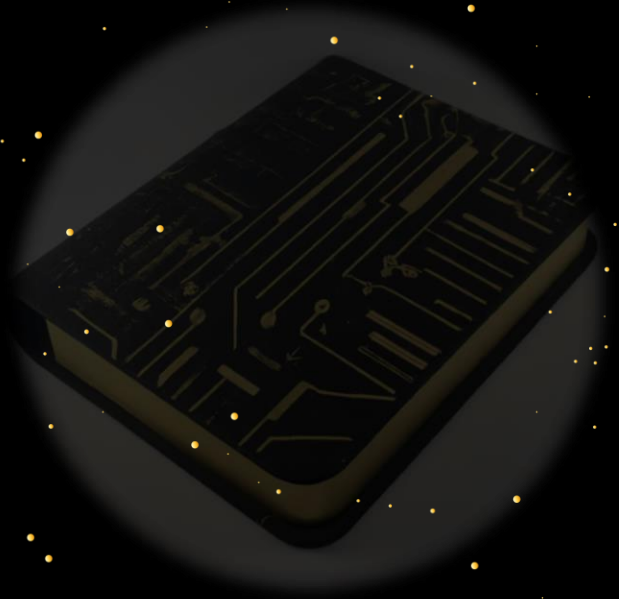

Data Immersion
Data Analysis
Key Results
Data Wrangling
Correlation Analysis

DATA IMMERSION

Current Application Data
124 Features
49,999 Records



DATA WRANGLING



- Columns with more than 45% null dropped (49)
- 18 columns to impute

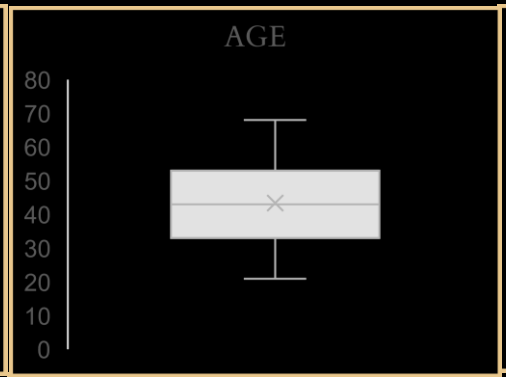
Columns to drop	Columns to impute
COMMONAREA_AVG	OCCUPATION_TYPE
COMMONAREA_MODE	EXT_SOURCE_3
COMMONAREA_MEDI	AMT_REQ_CREDIT_BUREAU_HOUR
NONLIVINGAPARTMENTS_AVG	AMT_REQ_CREDIT_BUREAU_DAY
NONLIVINGAPARTMENTS_MODE	AMT_REQ_CREDIT_BUREAU_WEEK
NONLIVINGAPARTMENTS_MEDI	AMT_REQ_CREDIT_BUREAU_MON
LIVINGAPARTMENTS_AVG	AMT_REQ_CREDIT_BUREAU_QRT
LIVINGAPARTMENTS_MODE	AMT_REQ_CREDIT_BUREAU_YEAR
LIVINGAPARTMENTS_MEDI	NAME_TYPE_SUITE
FONDKAPREMONT_MODE	OBS_30_CNT_SOCIAL_CIRCLE
FLOORSMIN_AVG	DEF_30_CNT_SOCIAL_CIRCLE
FLOORSMIN_MODE	OBS_60_CNT_SOCIAL_CIRCLE
FLOORSMIN_MEDI	DEF_60_CNT_SOCIAL_CIRCLE
YEARS_BUILD_AVG	EXT_SOURCE_2
YEARS_BUILD_MODE	AMT_GOODS_PRICE
YEARS_BUILD_MEDI	AMT_ANNUITY
OWN_CAR_AGE	CNT_FAM_MEMBERS
LANDAREA_AVG	DAYS_LAST_PHONE_CHANGE
LANDAREA_MODE	
LANDAREA_MEDI	
BASEMENTAREA_AVG	
BASEMENTAREA_MODE	
BASEMENTAREA_MEDI	
EXT_SOURCE_1	
NONLIVINGAREA_AVG	
NONLIVINGAREA_MODE	
NONLIVINGAREA_MEDI	
ELEVATORS_AVG	
ELEVATORS_MODE	
ELEVATORS_MEDI	
WALLSMATERIAL_MODE	
APARTMENTS_AVG	
APARTMENTS_MODE	
APARTMENTS_MEDI	
ENTRANCES_AVG	
ENTRANCES_MODE	
ENTRANCES_MEDI	
LIVINGAREA_AVG	
LIVINGAREA_MODE	
LIVINGAREA_MEDI	

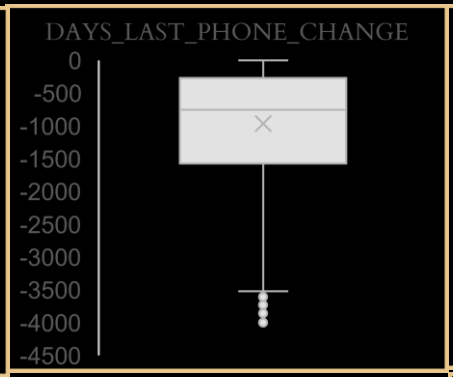
DATA WRANGLING

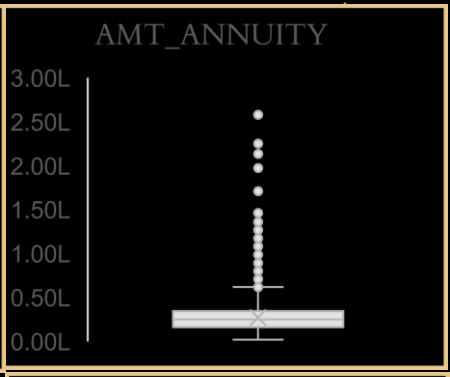
Column	Imputation Value
AMT_ANNUITY	24939.00
AMT_GOODS_PRICE	450000.00
CNT_FAM_MEMBERS	2.00
DAY_LAST_PHONE_CHANGE	-755.00
EXT_SOURCE_2	0.51
EXT_SOURCE_3	0.51
AMT_REQ_CREDIT_BUREAU_HOUR	0.00
AMT_REQ_CREDIT_BUREAU_DAY	0.00
AMT_REQ_CREDIT_BUREAU_WEEK	0.00
AMT_REQ_CREDIT_BUREAU_MON	0.00
AMT_REQ_CREDIT_BUREAU_QRT	0.00
AMT_REQ_CREDIT_BUREAU_YEAR	1.00
OBS_30_CNT_SOCIAL_CIRCLE	0.00
OBS_60_CNT_SOCIAL_CIRCLE	0.00
DEF_30_CNT_SOCIAL_CIRCLE	0.00
DEF_60_CNT_SOCIAL_CIRCLE	0.00
OCCUPATION_TYPE	Unknown
NAME_TYPE_SUITE	Unaccompanied

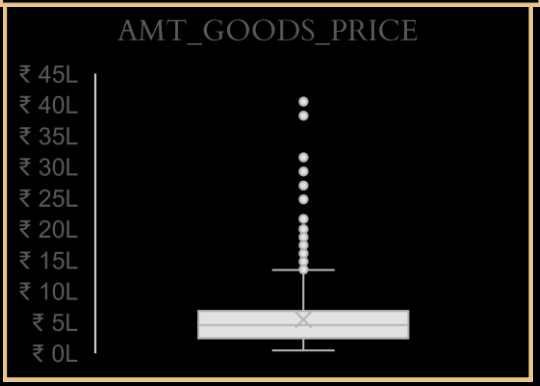
OUTLIERS

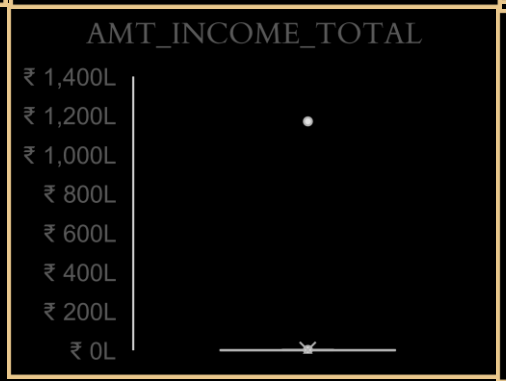
YEARS_EMPLOYED
200.00
0.00
-200.00
-400.00
-600.00
-800.00
-1000.00
-1200.00

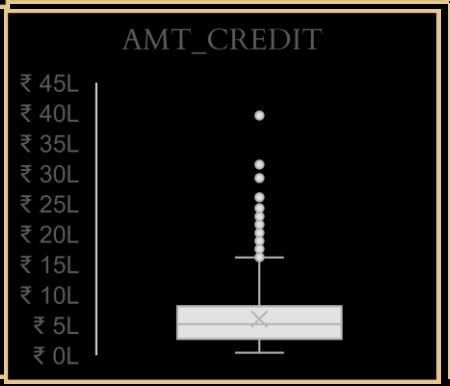

AGE
80
70
60
50
40
30
20
10
0

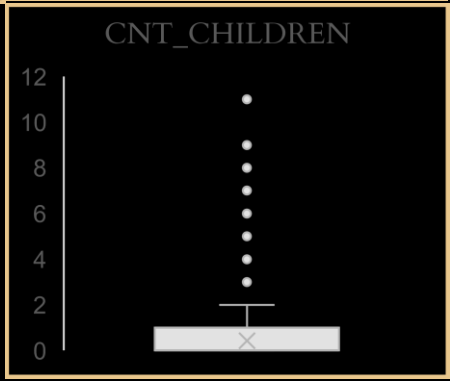

DAYS_LAST_PHONE_CHANGE
0
-500
-1000
-1500
-2000
-2500
-3000
-3500
-4000
-4500


AMT_ANNUITY
3.00L
2.50L
2.00L
1.50L
1.00L
0.50L
0.00L

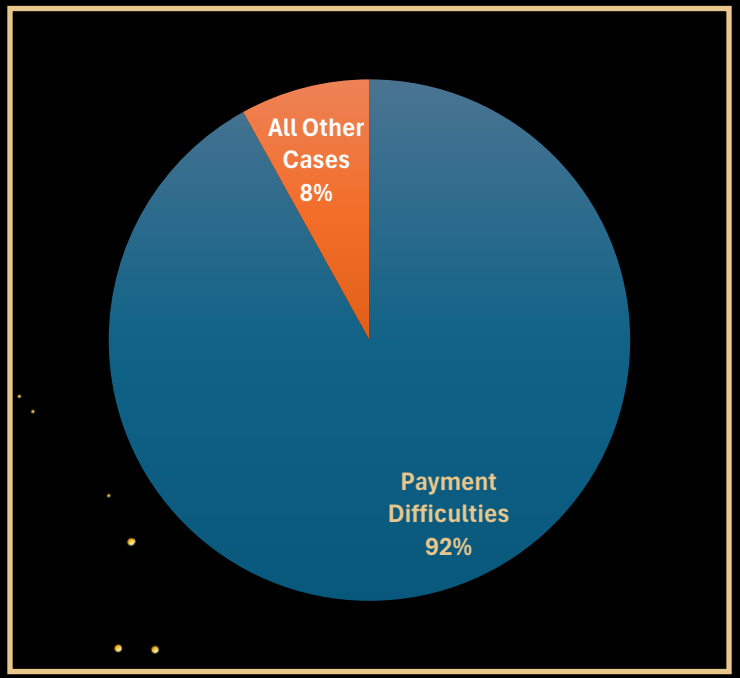

AMT_GOODS_PRICE
₹ 45L
₹ 40L
₹ 35L
₹ 30L
₹ 25L
₹ 20L
₹ 15L
₹ 10L
₹ 5L
₹ 0L


AMT_INCOME_TOTAL
₹ 1,400L
₹ 1,200L
₹ 1,000L
₹ 800L
₹ 600L
₹ 400L
₹ 200L
₹ 0L


AMT_CREDIT
₹ 45L
₹ 40L
₹ 35L
₹ 30L
₹ 25L
₹ 20L
₹ 15L
₹ 10L
₹ 5L
₹ 0L


CNT_CHILDREN
12
10
8
6
4
2
0

DATA IMBALANCE



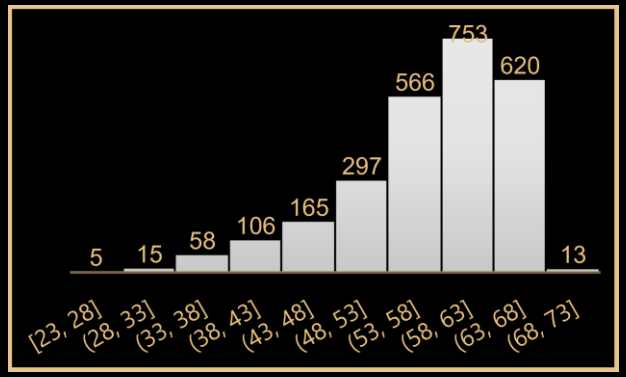
- 92% payment difficulties
- 8% other cases

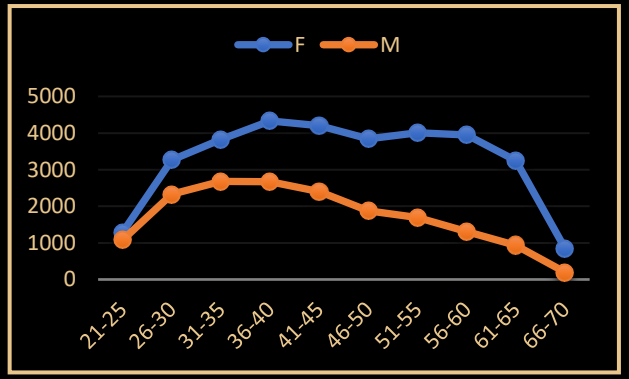
Imbalance ratio = 11.42

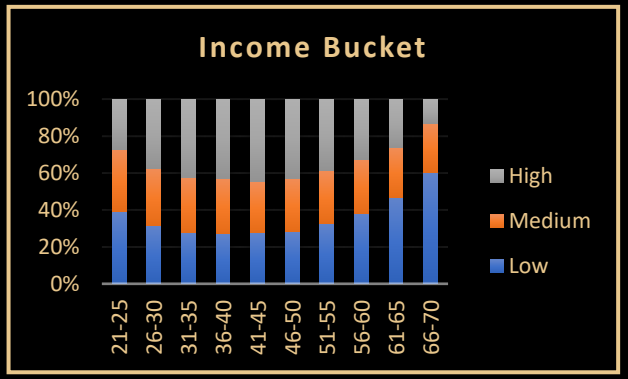
For every customer falling under "All other cases" there are ~11.5 customers with payment difficulties

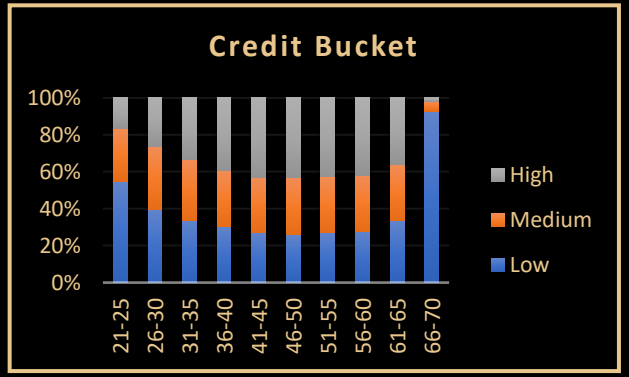
UNIVARIATE ANALYSIS

Age univariate/segmented univariate



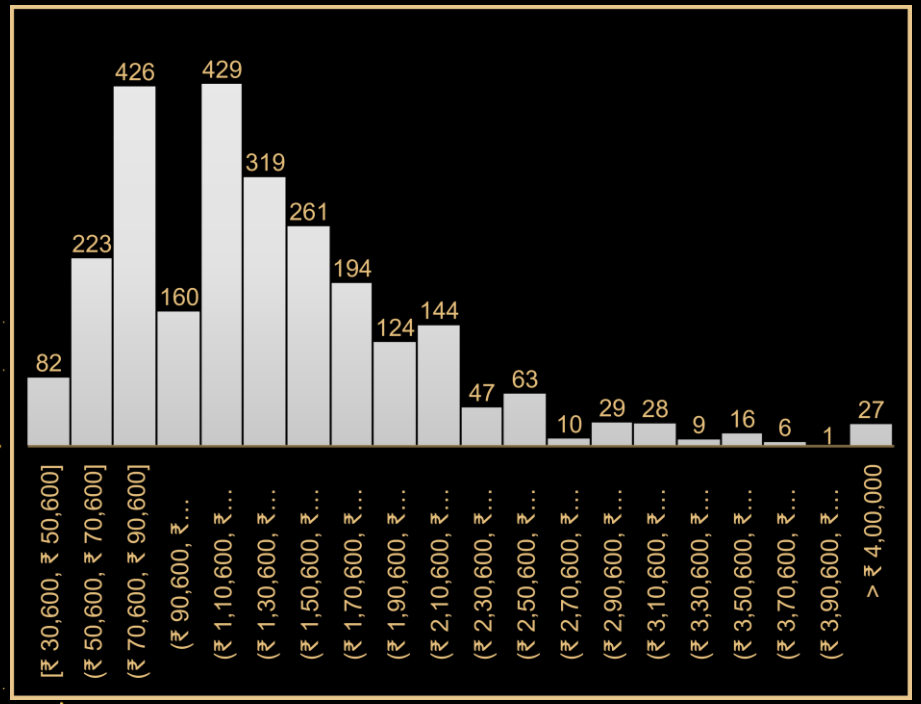


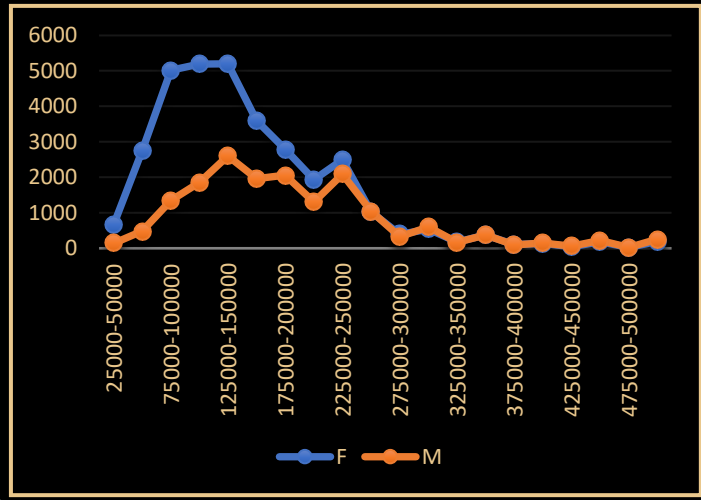




UNIVARIATE ANALYSIS

Income univariate/segmented univariate

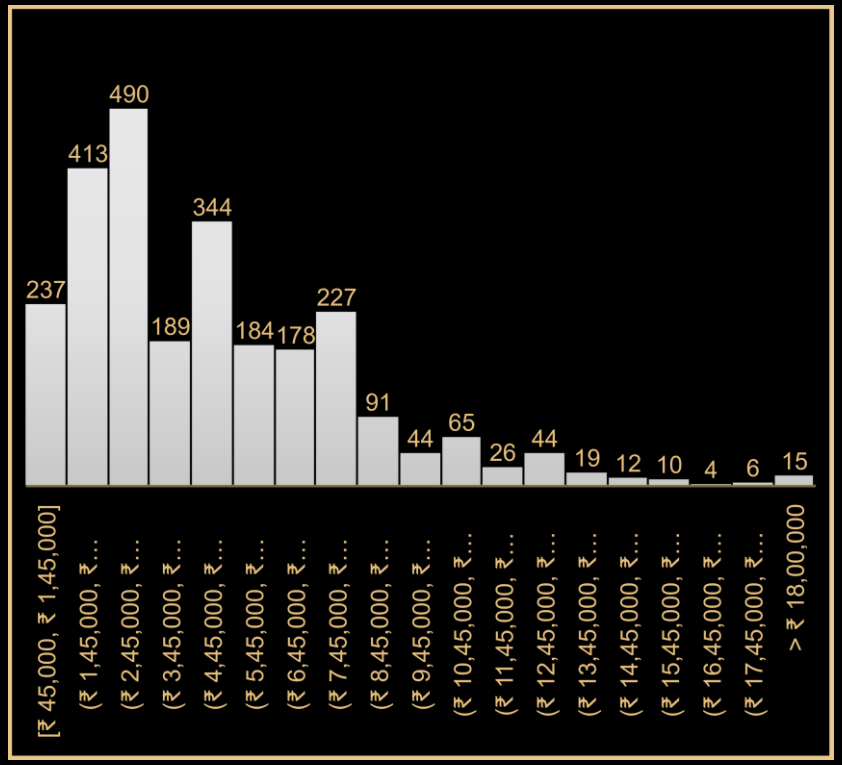


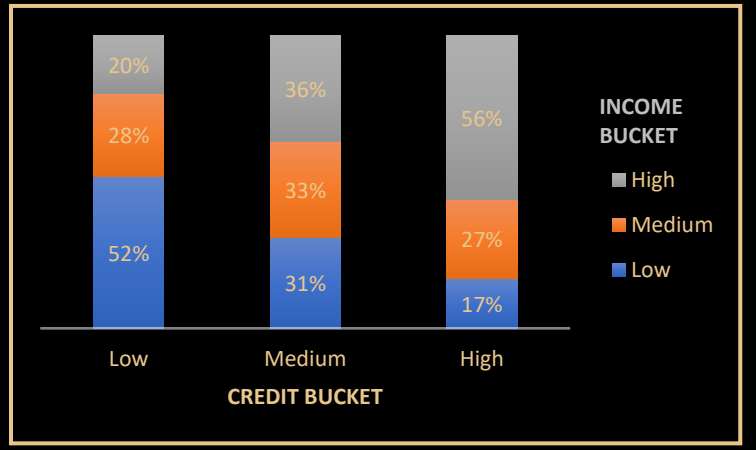


Min	₹ 25,650
Q1	₹ 1,12,500
Median	₹ 1,45,800
Q3	₹ 2,02,500
Max	₹ 11,70,00,000
Std Dev	₹ 5,31,819
Mean	₹ 1,70,768

UNIVARIATE ANALYSIS

Credit univariate/segmented univariate

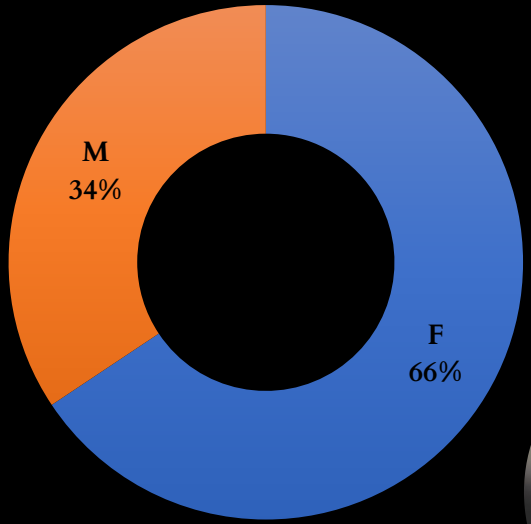


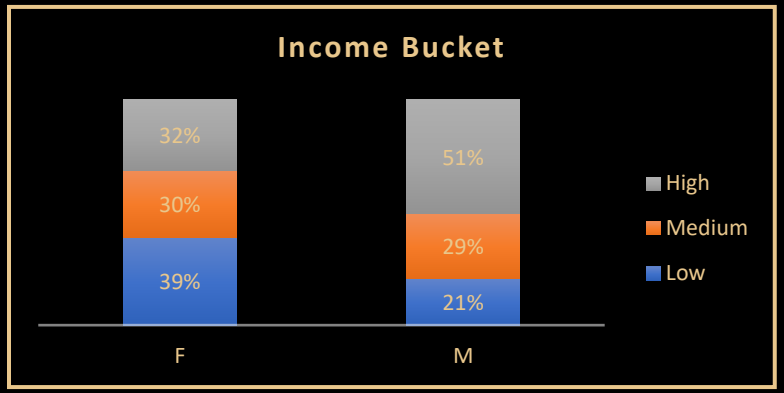


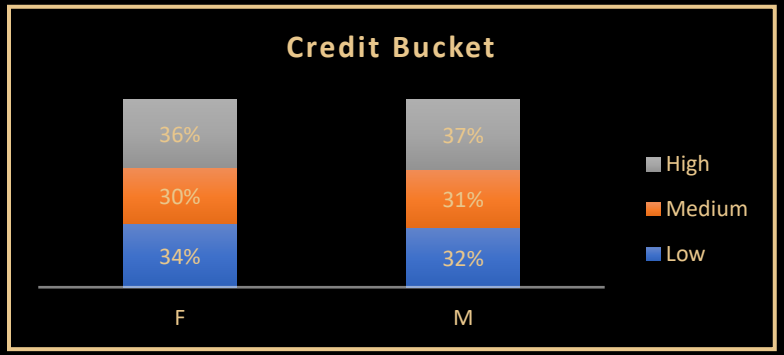
Min	₹ 45,000
Q1	₹ 2,70,000
Median	₹ 5,14,778
Q3	₹ 8,08,650
Max	₹ 40,50,000
Std Dev	₹ 4,02,415
Mean	₹ 5,99,701

UNIVARIATE ANALYSIS

Gender univariate/segmented univariate

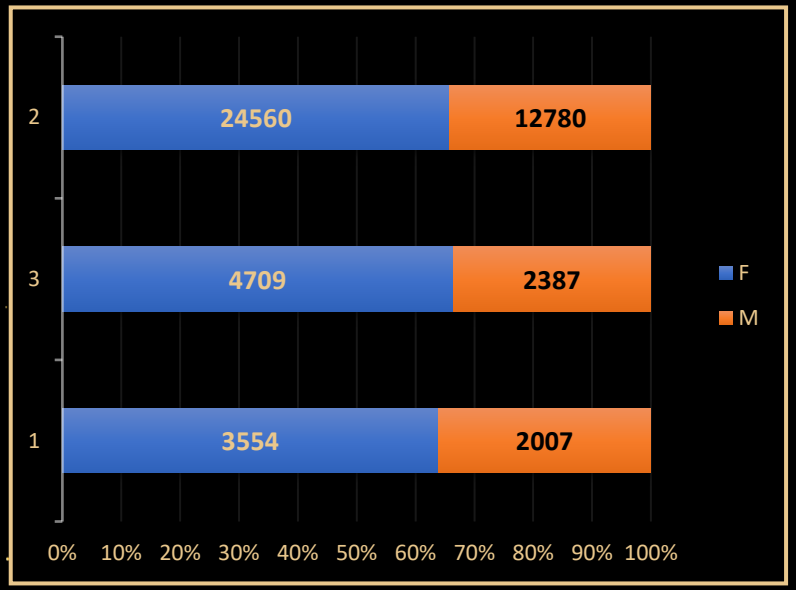


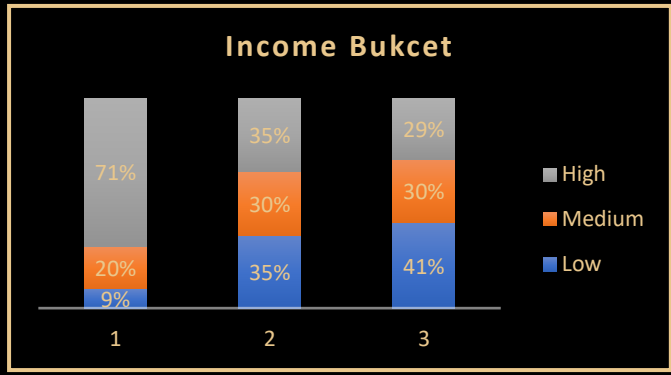


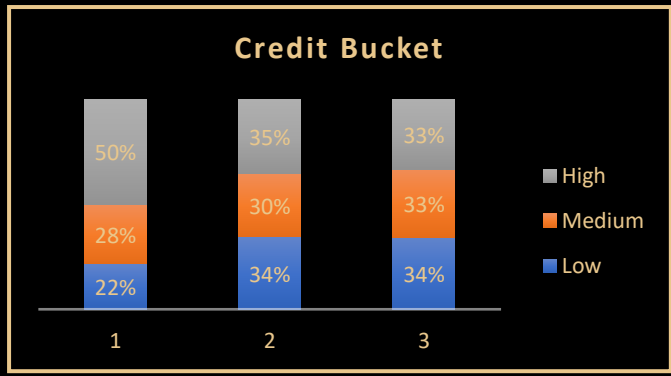


UNIVARIATE ANALYSIS

Region rating with city univariate/segmented univariate

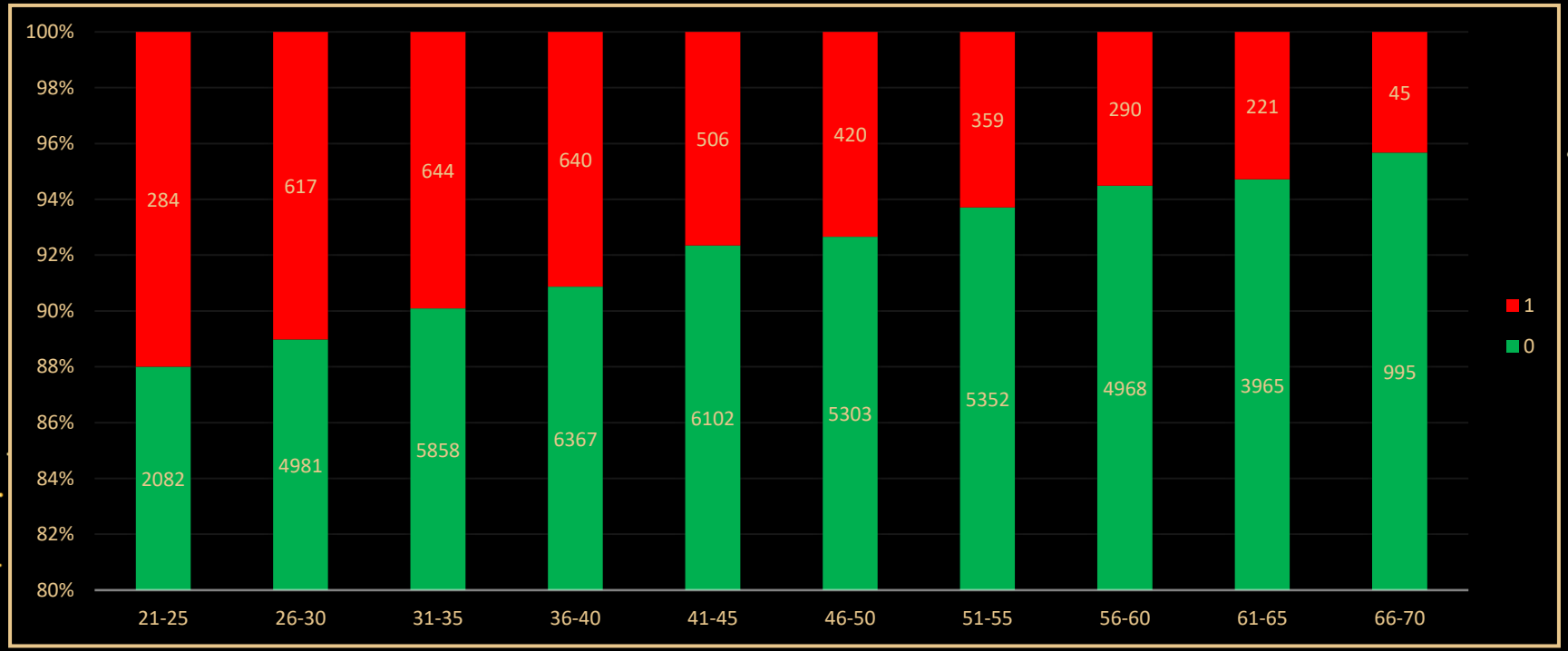






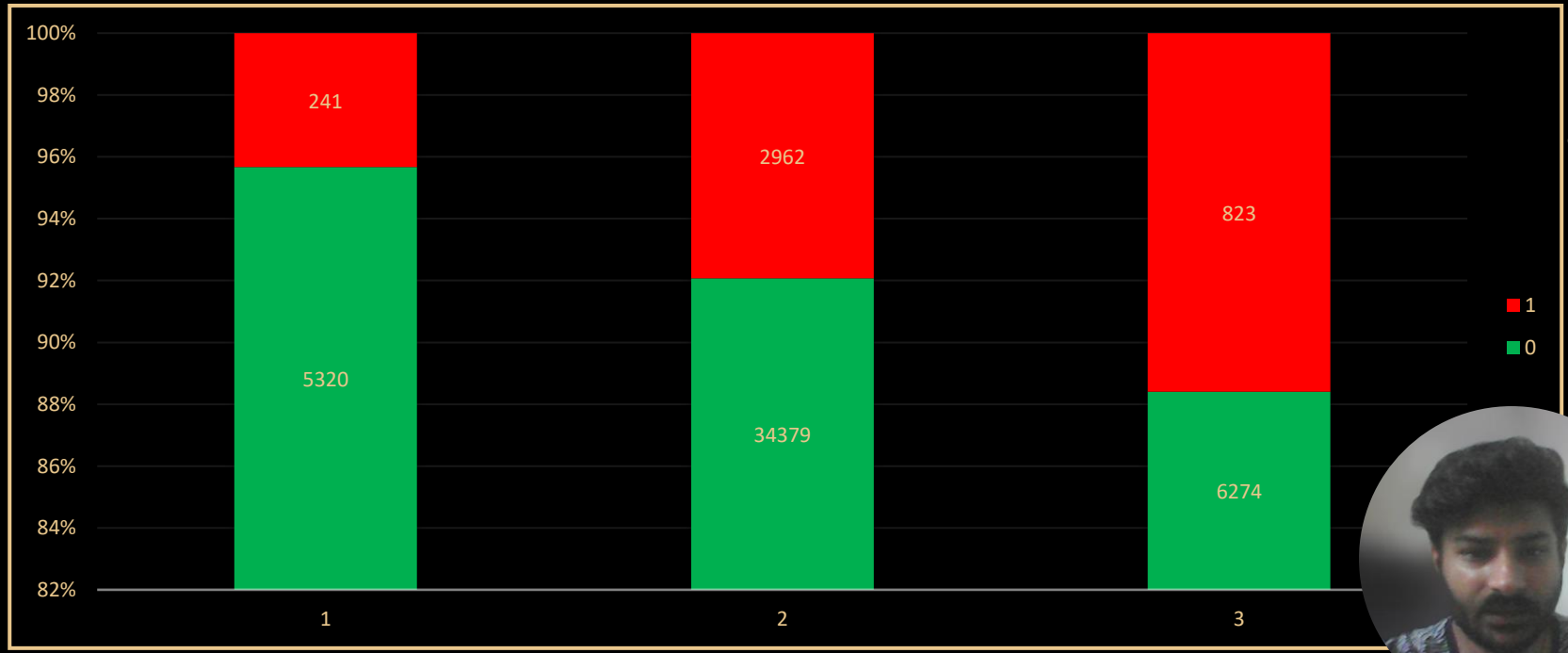
BIVARIATE ANALYSIS

Age vs. Target



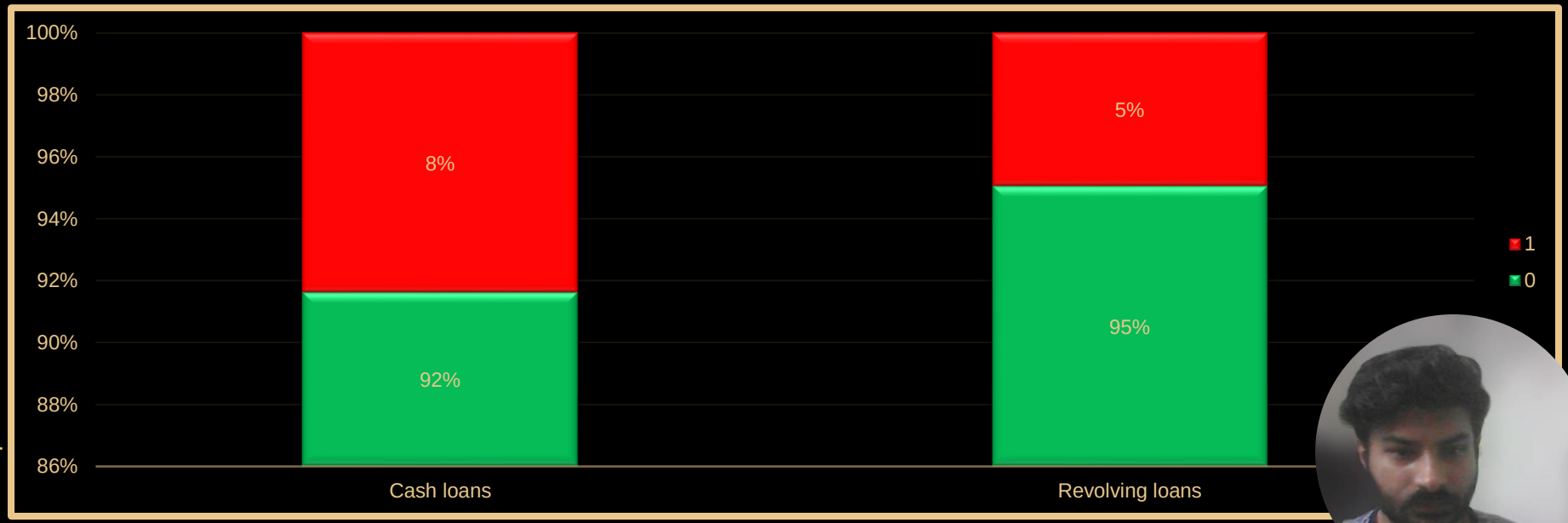
BIVARIATE ANALYSIS

Region rating w/ city vs. Target



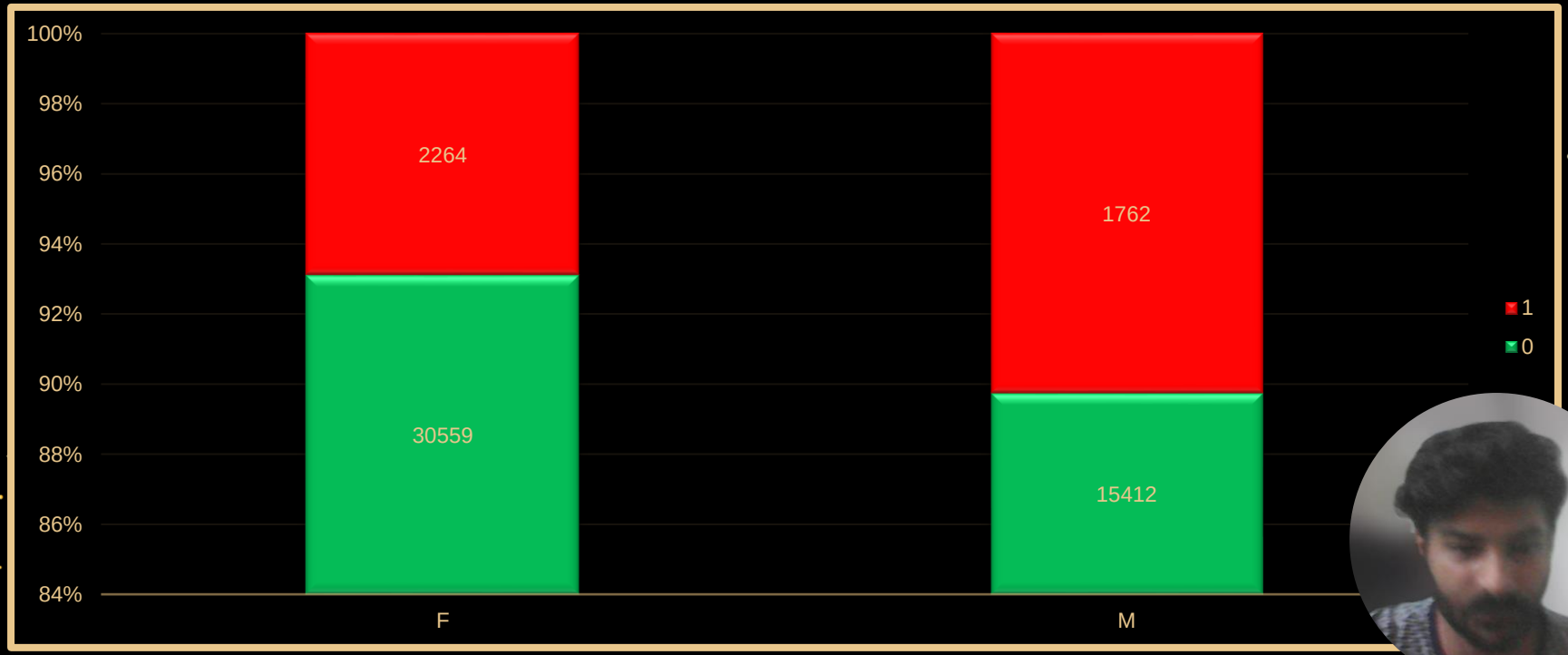
BIVARIATE ANALYSIS

Loan type vs. Target



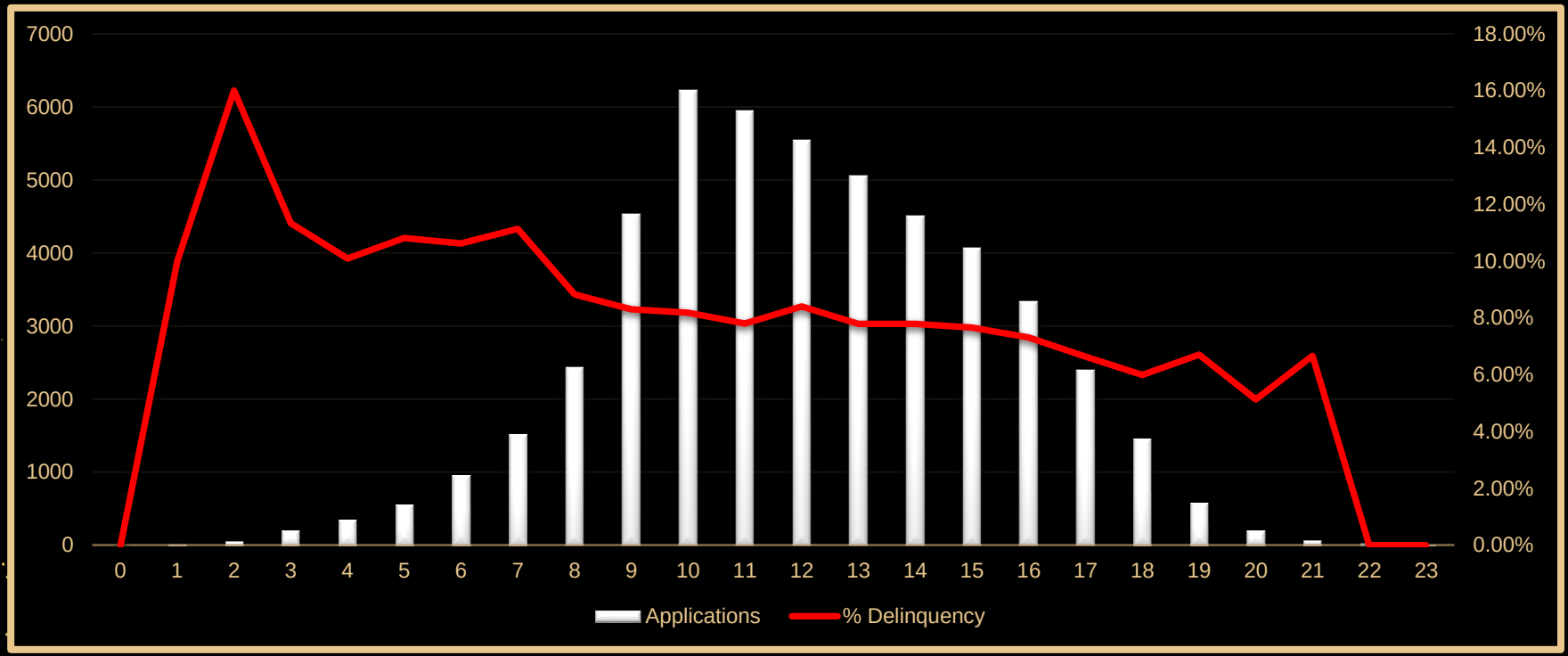
BIVARIATE ANALYSIS

Gender vs. Target



BIVARIATE ANALYSIS

Hour of Day vs. Target



TOP 10 CORRELATIONS

Delinquents (payment difficulties)

Absolute value	Value	Features
0.999967116	0.999967116	YEARS_BUILD_AVG - YEARS_BUILD_MEDI
0.999954856	0.999954856	NONLIVINGAPARTMENTS_AVG - NONLIVINGAPARTMENTS_MEDI
0.999902027	-0.99990203	DAYS_EMPLOYED - FLAG_EMP_PHONE
0.999890172	0.999890172	YEARS_BEGINEXPLUATATION_AVG - YEARS_BEGINEXPLUATATION_MEDI
0.999870952	0.999870952	YEARS_BEGINEXPLUATATION_MODE - YEARS_BEGINEXPLUATATION_MEDI
0.999762114	0.999762114	YEARS_BEGINEXPLUATATION_AVG - YEARS_BEGINEXPLUATATION_MODE
0.999623632	0.999623632	YEARS_BUILD_AVG - YEARS_BUILD_MODE
0.999592688	0.999592688	YEARS_BUILD_MODE - YEARS_BUILD_MEDI
0.999495894	0.999495894	FLOORSMIN_AVG - FLOORSMIN_MEDI
0.999366414	0.999366414	LIVINGAPARTMENTS_AVG - LIVINGAPARTMENTS_MEDI

TOP 10 CORRELATIONS

Re-payers

Absolute value	Value	Features
0.999962	0.999962399	YEARS_BEGINEXPLUATATION_AVG - YEARS_BEGINEXPLUATATION_MEDI
0.999921	0.999921238	YEARS_BUILD_AVG - YEARS_BUILD_MEDI
0.999775	0.999775113	YEARS_BEGINEXPLUATATION_AVG - YEARS_BEGINEXPLUATATION_MODE
0.999742	-0.999742029	DAYS_EMPLOYED - FLAG_EMP_PHONE
0.999727	0.999726951	YEARS_BEGINEXPLUATATION_MODE - YEARS_BEGINEXPLUATATION_MEDI
0.999644	0.999644139	YEARS_BUILD_MODE - YEARS_BUILD_MEDI
0.9996	0.999599734	YEARS_BUILD_AVG - YEARS_BUILD_MODE
0.998946	0.998946151	LIVINGAPARTMENTS_AVG - LIVINGAPARTMENTS_MEDI
0.998785	0.998785029	FLOORSMIN_AVG - FLOORSMIN_MEDI
0.998592	0.998592178	FLOORSMAX_AVG - FLOORSMAX_MEDI

KEY INSIGHTS

- Massive data imbalance - 92% of customers have payment difficulties
- Males have higher delinquency rate (10%) than females (7%)
- Region 1 customers most affluent - 71% high income, 50% high credit
- Delinquency rate drops with age - 12% for 21-25 vs. 4% for 66-70
- Region 3 has highest delinquency, followed by Region 2 and 1
- Cash loans slightly riskier (8% delinquency) than revolving loans (5%)

RESULTS

- Managing large dataset in Excel was challenging - provided learning opportunity
- Handling outliers and missing values using Power Query
- Learned when to impute vs. drop missing values
- Joined tables using Power Query left join
- Used Analysis ToolPak for correlations - new skill