BANK LOAN CASE STUDY

Project description:

- Task is to use Exploratory Data Analysis (EDA) to analyze patterns in the data and ensure that capable applicants are not rejected.
- **When a customer applies for a loan, your company faces two risks:**

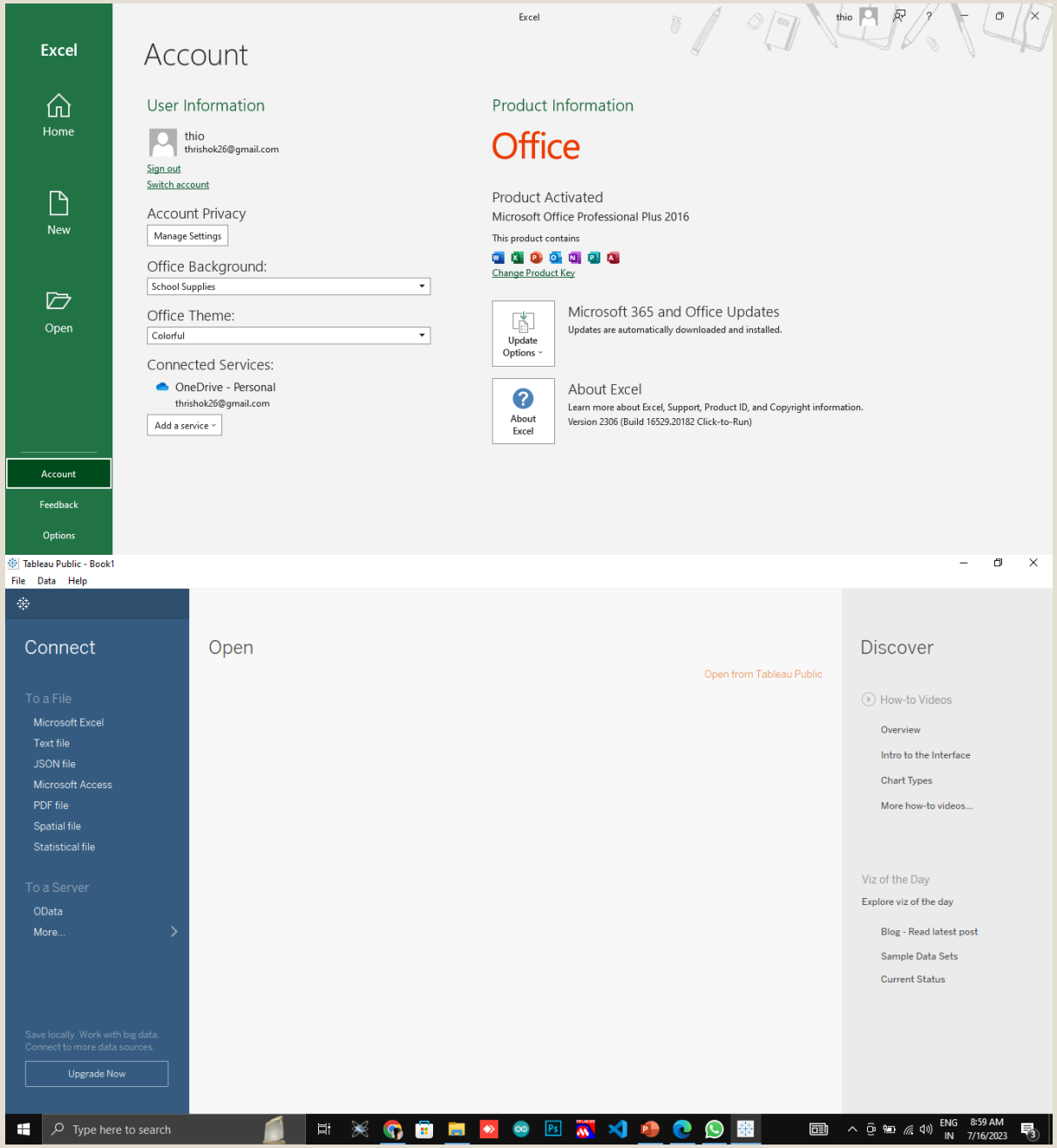
1. If the applicant can repay the loan but is not approved, the company loses business.
2. If the applicant cannot repay the loan and is approved, the company faces a financial loss.

Approach:

- Identify Missing Data and Deal with it.
- Identify Outliers in the Dataset.
- Analyze Data Imbalance.
- Perform Univariate, Segmented Univariate, and Bivariate Analysis.
- Identify Top Correlations for Different Scenarios.

Tech-Stack used:

- I have used Microsoft Office Professional Pro 2016 and tableau to achieve output for the given data

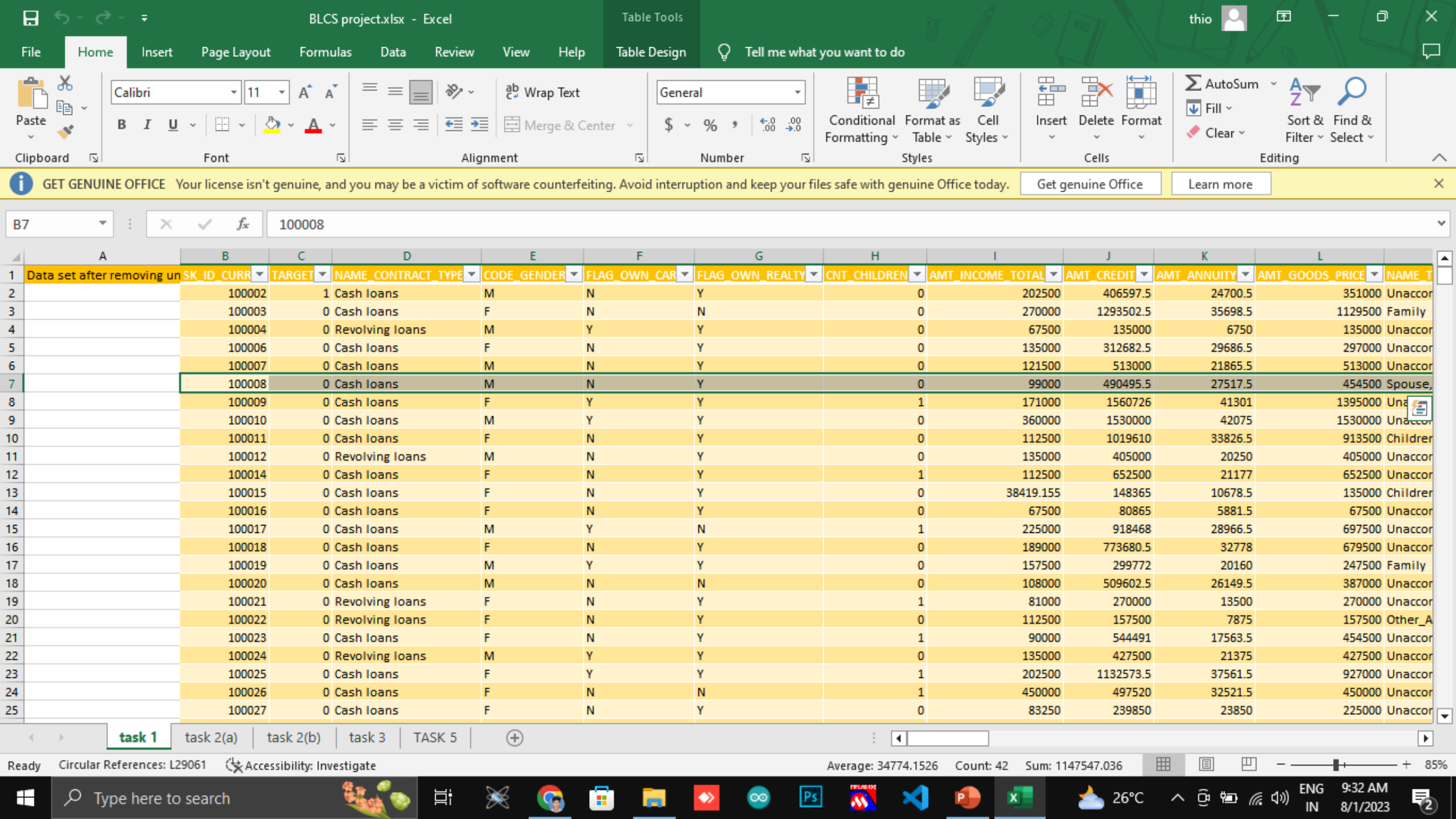


Insights:

- Through this project I have learned Exploratory Data Analysis (EDA).
- Dealing with missing values.
- Finding outliers and Performing Univariate, Segmented Univariate, and Bivariate Analysis.
- Top Correlations for Different Scenarios.

Task 1: Identify Missing Data and Deal with it Appropriately:

- ->removed duplicate values based on SK_ID_CURR.
- ->removing columns having more than 50% null values by using this formula "=100-((C2/50000)*100)". This helps to find percentage of null values.
- ->removed columns having more than 50% of null values.
- ->filling missing cells with average of previous 10 values and next 10 values.
- ->filling missing values with average of the column.

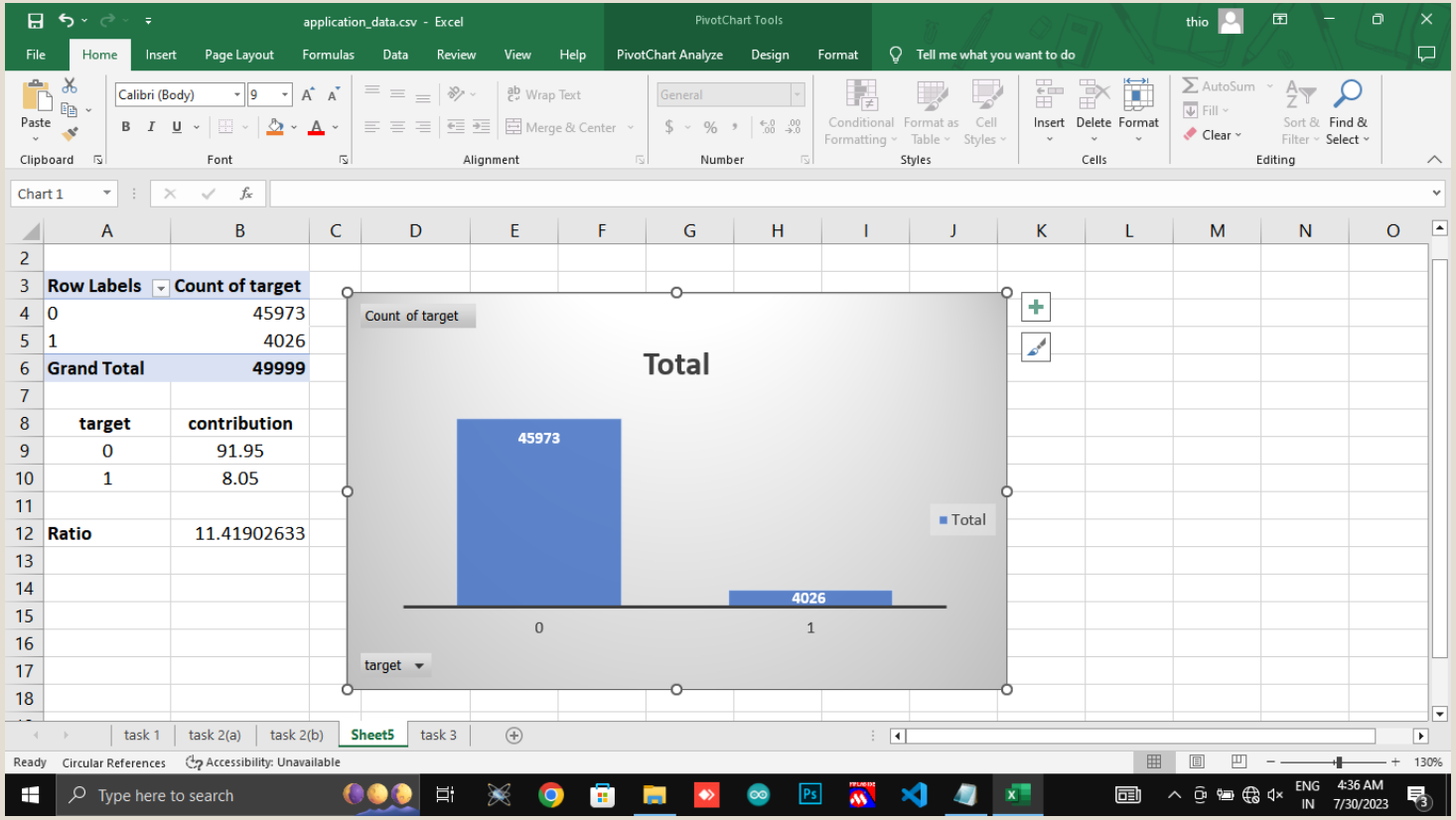

BLCS project.xlsx - Excel
Table Tools
thio
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Help
Table Design
Tell me what you want to do
Paste
Calibri
11
Wrap Text
General
Conditional Formatting
Format as Table
Cell Styles
Insert
Delete
Format
AutoSum
Fill
Clear
Sort & Filter
Find & Select
Clipboard
Font
Alignment
Number
Styles
Cells
Editing
GET GENUINE OFFICE Your license isn't genuine, and you may be a victim of software counterfeiting. Avoid interruption and keep your files safe with genuine Office today.
Get genuine Office
Learn more
B7
100008
Data set after removing un
SK_ID_CURR
TARGET
NAME_CONTRACT_TYPE
CODE_GENDER
FLAG_OWN_CAR
FLAG_OWN_REALTY
CNT_CHILDREN
AMT_INCOME_TOTAL
AMT_CREDIT
AMT_ANNUITY
AMT_GOODS_PRICE
NAME_T
100002 1 Cash loans M N Y 0 202500 406597.5 24700.5 351000 Unaccor
100003 0 Cash loans F N N 0 270000 1293502.5 35698.5 1129500 Family
100004 0 Revolving loans M Y Y 0 67500 135000 6750 135000 Unaccor
100006 0 Cash loans F N Y 0 135000 312682.5 29686.5 297000 Unaccor
100007 0 Cash loans M N Y 0 121500 513000 21865.5 513000 Unaccor
100008 0 Cash loans M N Y 0 99000 490495.5 27517.5 454500 Spouse,
100009 0 Cash loans F Y Y 1 171000 1560726 41301 1395000 Una
100010 0 Cash loans M Y Y 0 360000 1530000 42075 1530000 Unaccor
100011 0 Cash loans F N Y 0 112500 1019610 33826.5 913500 Childrer
100012 0 Revolving loans M N Y 0 135000 405000 20250 405000 Unaccor
100014 0 Cash loans F N Y 1 112500 652500 21177 652500 Unaccor
100015 0 Cash loans F N Y 0 38419.155 148365 10678.5 135000 Childrer
100016 0 Cash loans F N Y 0 67500 80865 5881.5 67500 Unaccor
100017 0 Cash loans M Y N 1 225000 918468 28966.5 697500 Unaccor
100018 0 Cash loans F N Y 0 189000 773680.5 32778 679500 Unaccor
100019 0 Cash loans M Y Y 0 157500 299772 20160 247500 Family
100020 0 Cash loans M N N 0 108000 509602.5 26149.5 387000 Unaccor
100021 0 Revolving loans F N Y 1 81000 270000 13500 270000 Unaccor
100022 0 Revolving loans F N Y 0 112500 157500 7875 157500 Other_A
100023 0 Cash loans F N Y 1 90000 544491 17563.5 454500 Unaccor
100024 0 Revolving loans M Y Y 0 135000 427500 21375 427500 Unaccor
100025 0 Cash loans F Y Y 1 202500 1132573.5 37561.5 927000 Unaccor
100026 0 Cash loans F N N 1 450000 497520 32521.5 450000 Unaccor
100027 0 Cash loans F N Y 0 83250 239850 23850 225000 Unaccor
task 1
task 2(a)
task 2(b)
task 3
TASK 5
Ready
Circular References: L29061
Accessibility: Investigate
Average: 34774.1526
Count: 42
Sum: 1147547.036
85%
Type here to search
26°C
ENG
IN
9:32 AM
8/1/2023

Task 2: Identify Outliers in the Dataset:

- ->finding 1st quartile,3rd quartile,Inter Quartile Range,upper limit and lower limit.
- ->creating scatter plot to find outlier using target and amount total income.
- ->creating scatter plot to find outlier using target and CNT_children.
- Outlier1
- Outlier2
- Outlier3

Task 3: Analyze Data Imbalance:

- ->Finding data imbalance by counting 0's and 1's in Target
- ->Plotting bar graph and Calculating their contribution



Task 4: Perform Univariate, Segmented Univariate, and Bivariate Analysis:

- Finding univariate and Bivariate analysis by creating income bins ,target,credit bins,applicants and average of amount credit.

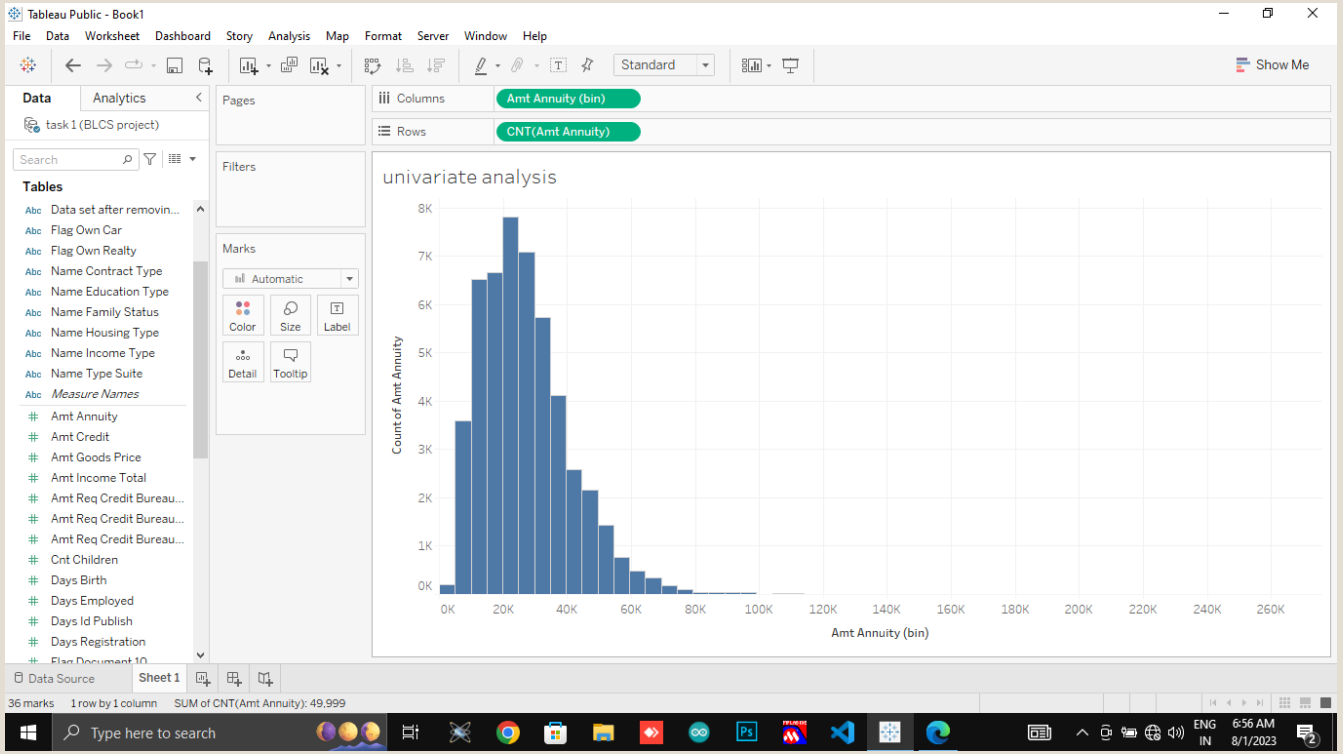


Tableau Public - AI vs T

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help

Show Me

Dashboard Layout

Default
Pho...
Device Pre...

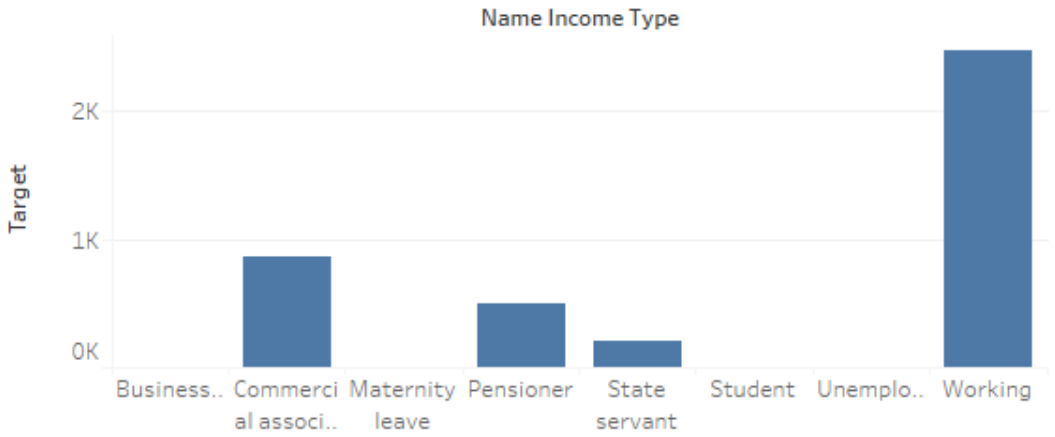
Size
Automatic

Sheets
Sheet 1
Sheet 2
Sheet 3
Sheet 4

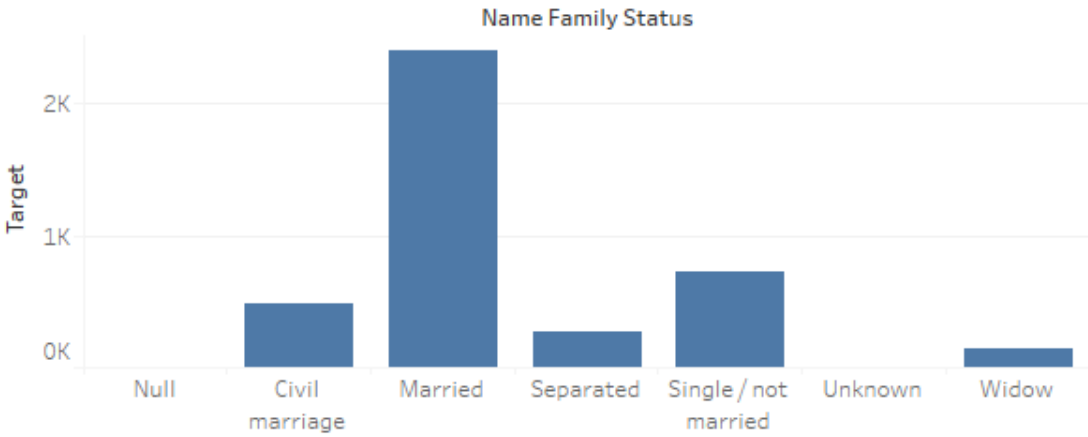
Objects
Horizontal C...
Vertical Con...
Text
Extension
Ask Data
Data Story
Image
Blank
Workflow

Tiled Floating

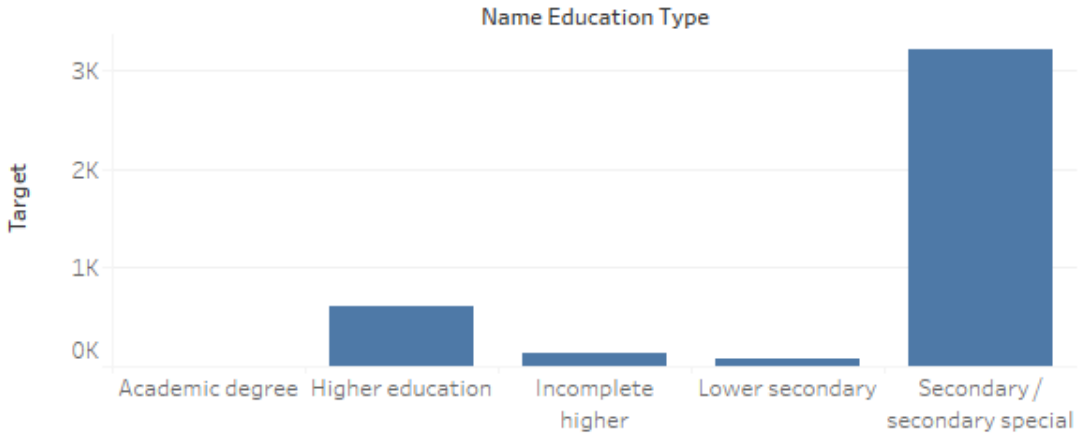
Income type vs Target



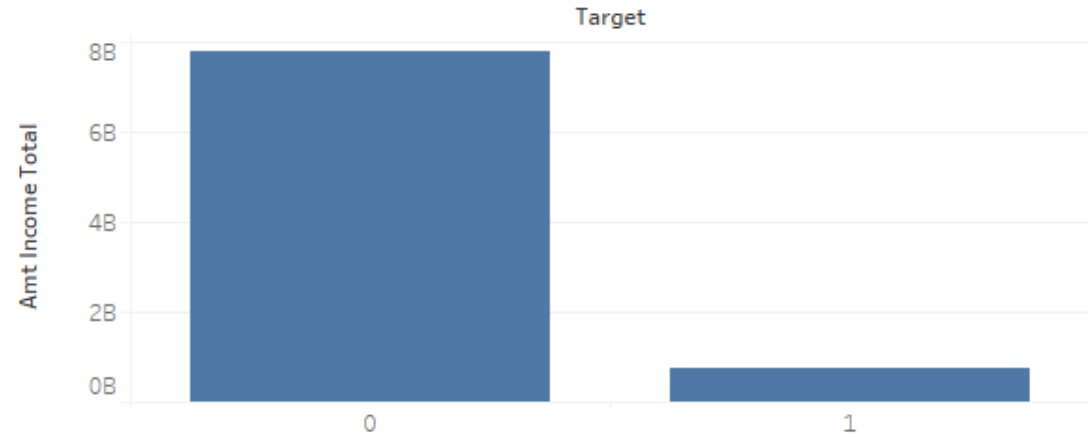
Family member VS Target



Education type VS Target



amount income vs target



Data Source | Sheet 1 | Sheet 2 | Sheet 3 | Sheet 4 | Dashboard 1

Thrishok S

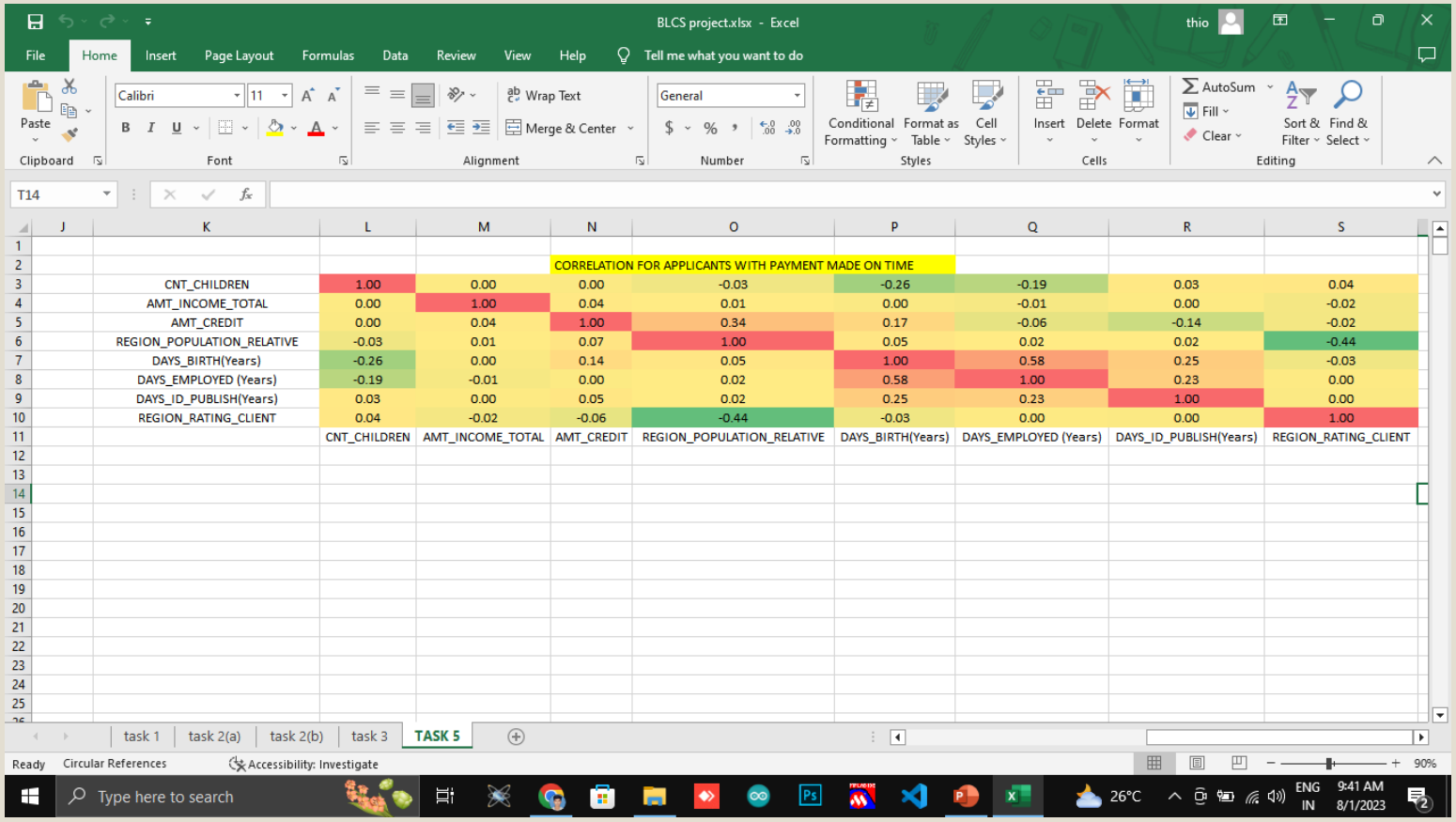
Type here to search

ENG IN 8:39 AM 8/1/2023

Task 5: Identify Top Correlations for Different Scenarios:

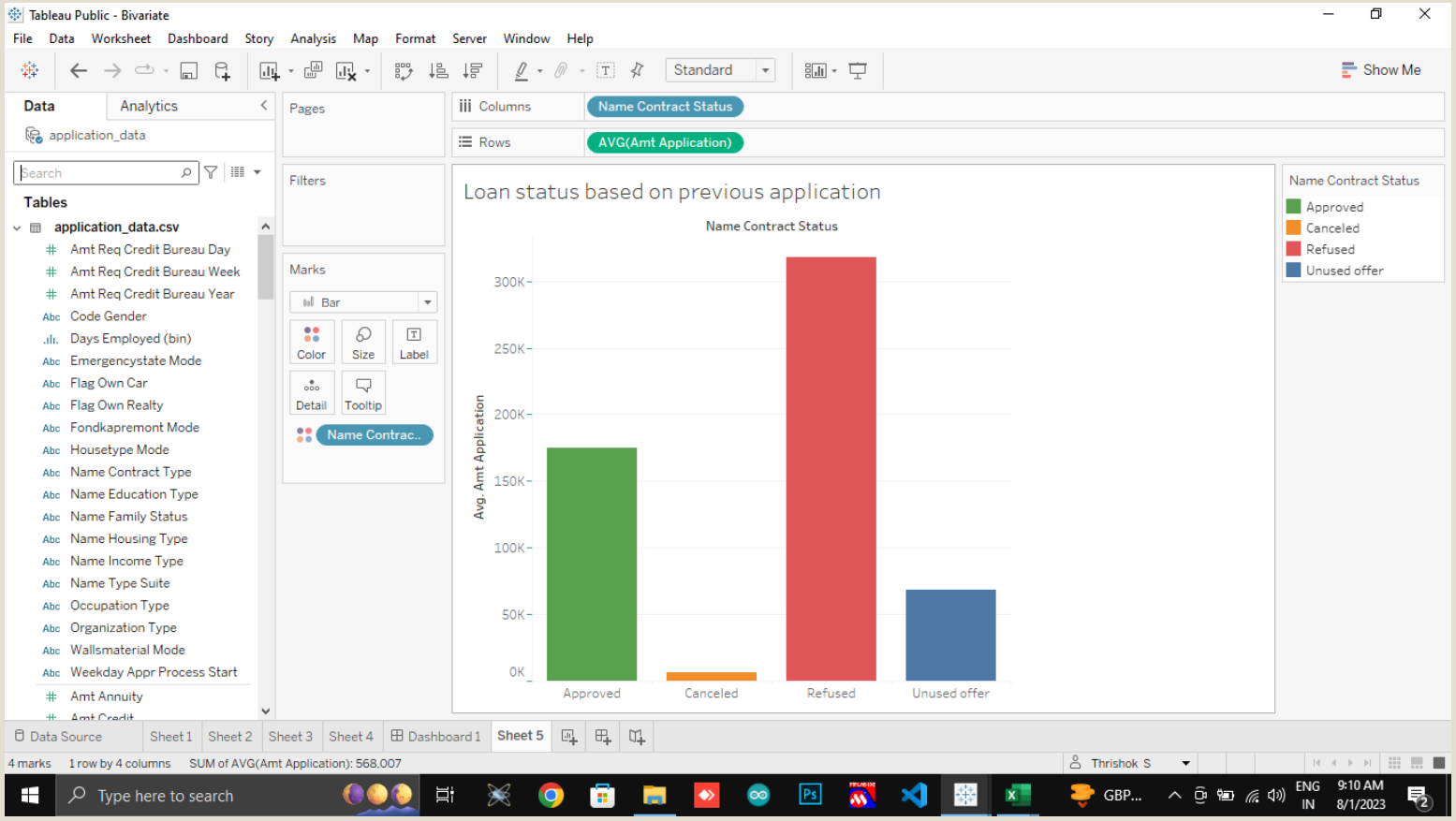
- Top ten reasons for loan cancellation and refusal

1.Amount Application
2.Cash loan Purpose
3.Goods Category
4.Product Combination
5.Product type
6.Channel type
7.Months Decision
8.Contract type
9.Client type
10.Payment type



Result:

- I completed each activity in the risk analytics procedure in turn.
- The following are the project results: The bank's problem statement calls for identifying the principal causes of bank loan default. The corporation will use the information to assess risks. Here, we provide two very large data sets.
- All of the client's data at the time of application is contained in the file "application data.csv." The data relates to whether or not a client is experiencing financial difficulties.
- The client's prior loans are detailed in the file "previous application.csv." It states if the previous application was approved, rejected, cancelled, or not used.



Click here for xlsx file