Bank Loan Case Study

By – Yuganter Pratap

Data Analytics Tasks

1. Identify Missing Data and Deal with it Appropriately
2. Identify Outliers in the Dataset
3. Analyze Data Imbalance
4. Perform Univariate, Segmented Univariate, and Bivariate Analysis
5. Identify Top Correlations for Different Scenarios

Description

Company faces a challenge: some customers who don't have a sufficient credit history take advantage of this and default on their loans. Your 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.

The dataset you'll be working with contains information about loan applications. It includes two types of scenarios:
1. Customers with payment difficulties: These are customers who had a late payment of more than X days on at least one of the first Y installments of the loan.
2. All other cases: These are cases where the payment was made on time.

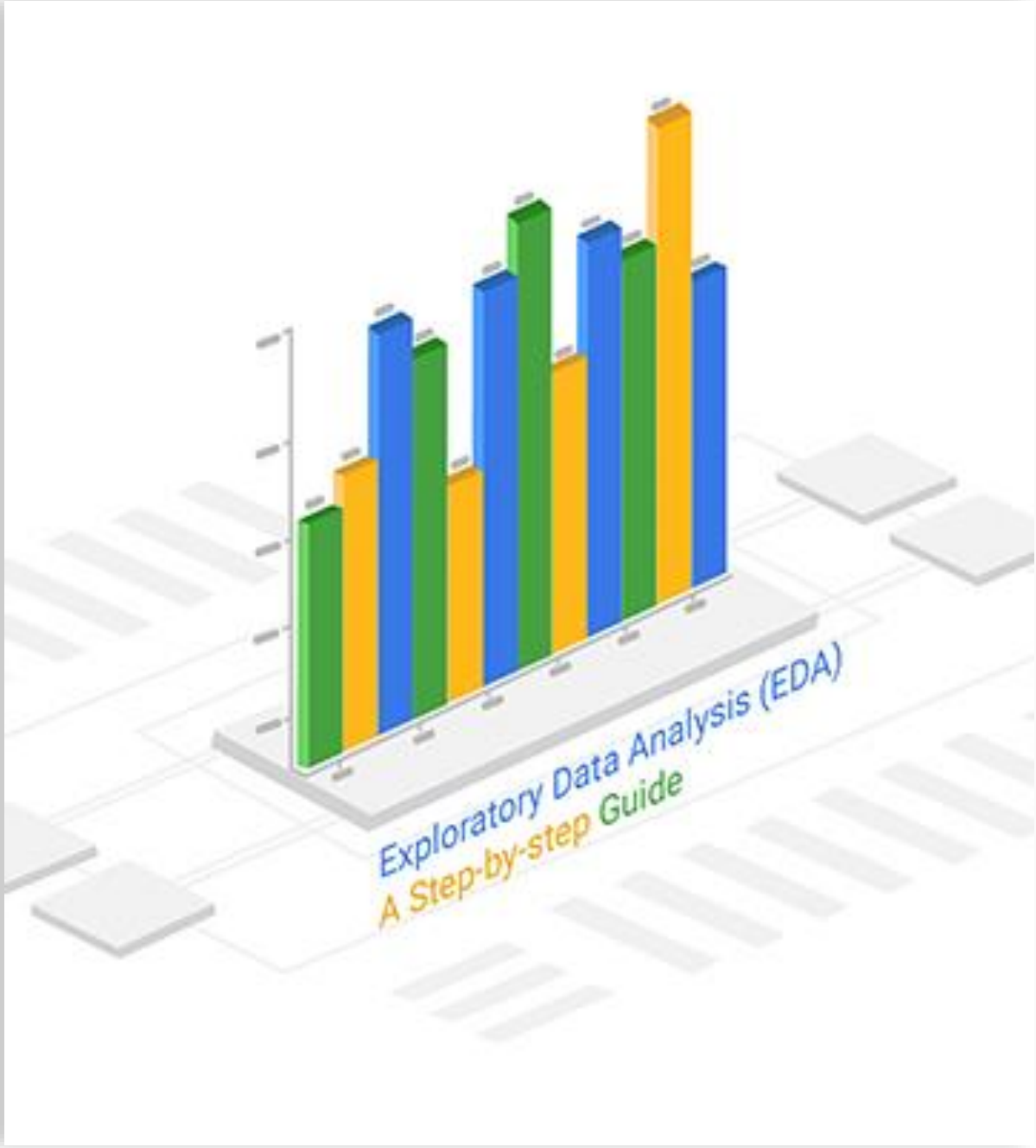
When a customer applies for a loan, there are four possible outcomes:
1. Approved: The company has approved the loan application.
2. Cancelled: The customer cancelled the application during the approval process.
3. Refused: The company rejected the loan.
4. Unused Offer: The loan was approved but the customer did not use it.

Your goal in this project is to use EDA to understand how customer attributes and loan attributes influence the likelihood of default.

Approach

As the data is quite large, we'll be using **jupyter notebook on vscode** for the data cleaning and **excel** for the analysis.

Our approach towards the project involved firstly inspecting the data, looking for missing values, handling the outliers to prevent them from hindering in our analysis, then analyzing the data and finally visualizing our finding in the form of various graphs and plots.



Tech-Stack Used

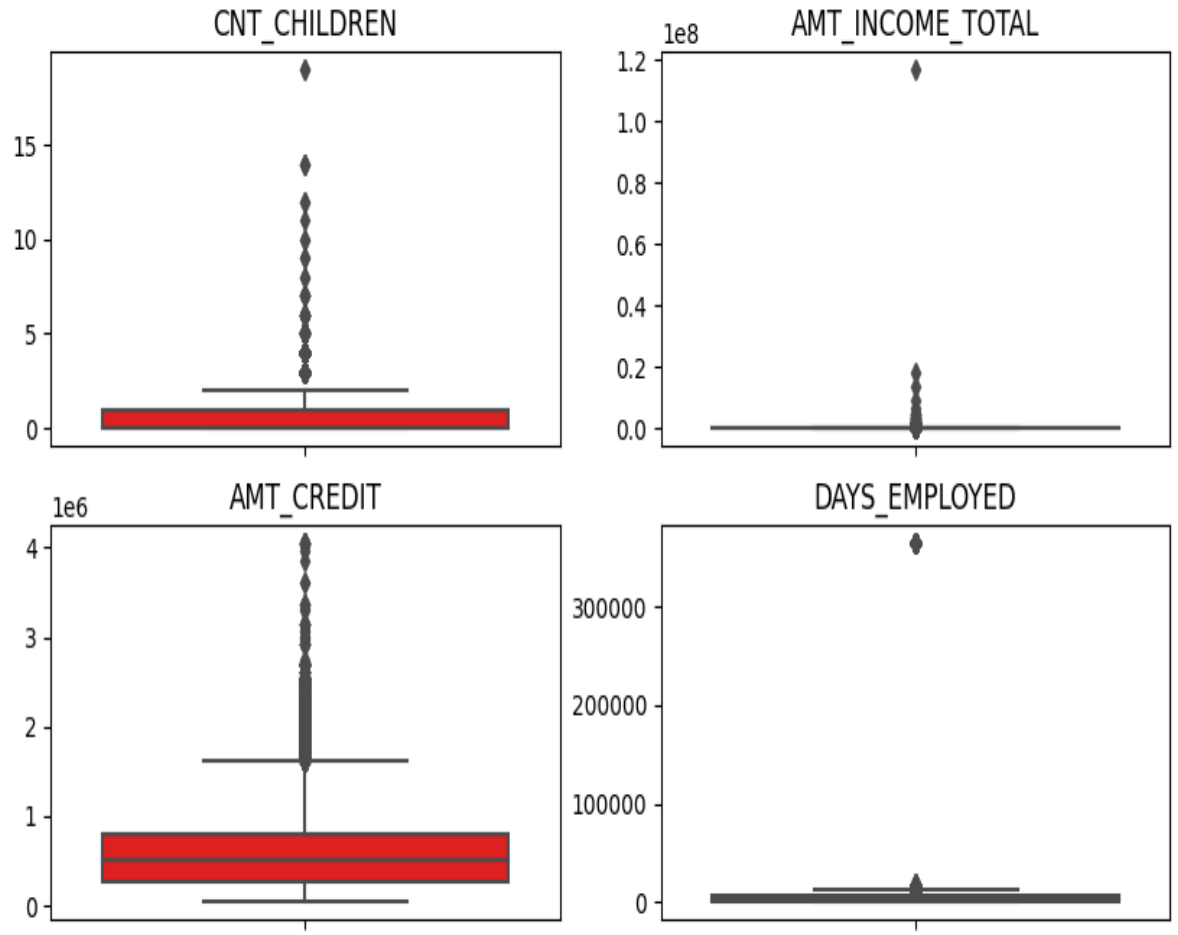
Python

Excel

Exploratory Data Analysis

- Click here for the jupyter notebook to see the process of data cleaning.
- Excel file
- Outliers:

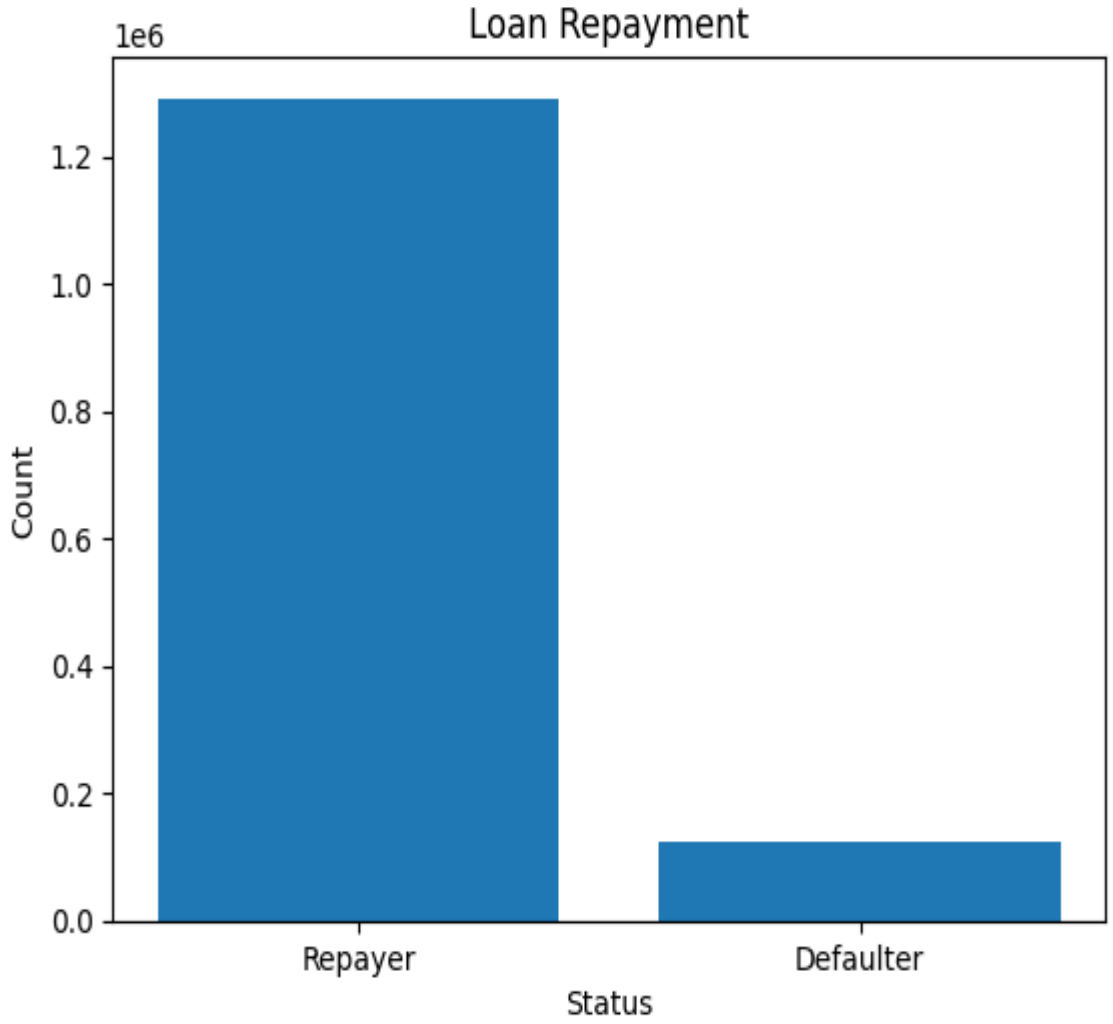
1. There is an instance where the no.of children are 18.
2. Total Income close to 12 crore is out of the norm for someone in need of loan
3. Days of employment more than 1000 years is not possible
4. In Amount Credited you can see a significant amount of outlies from the mean value of the entire dataset.



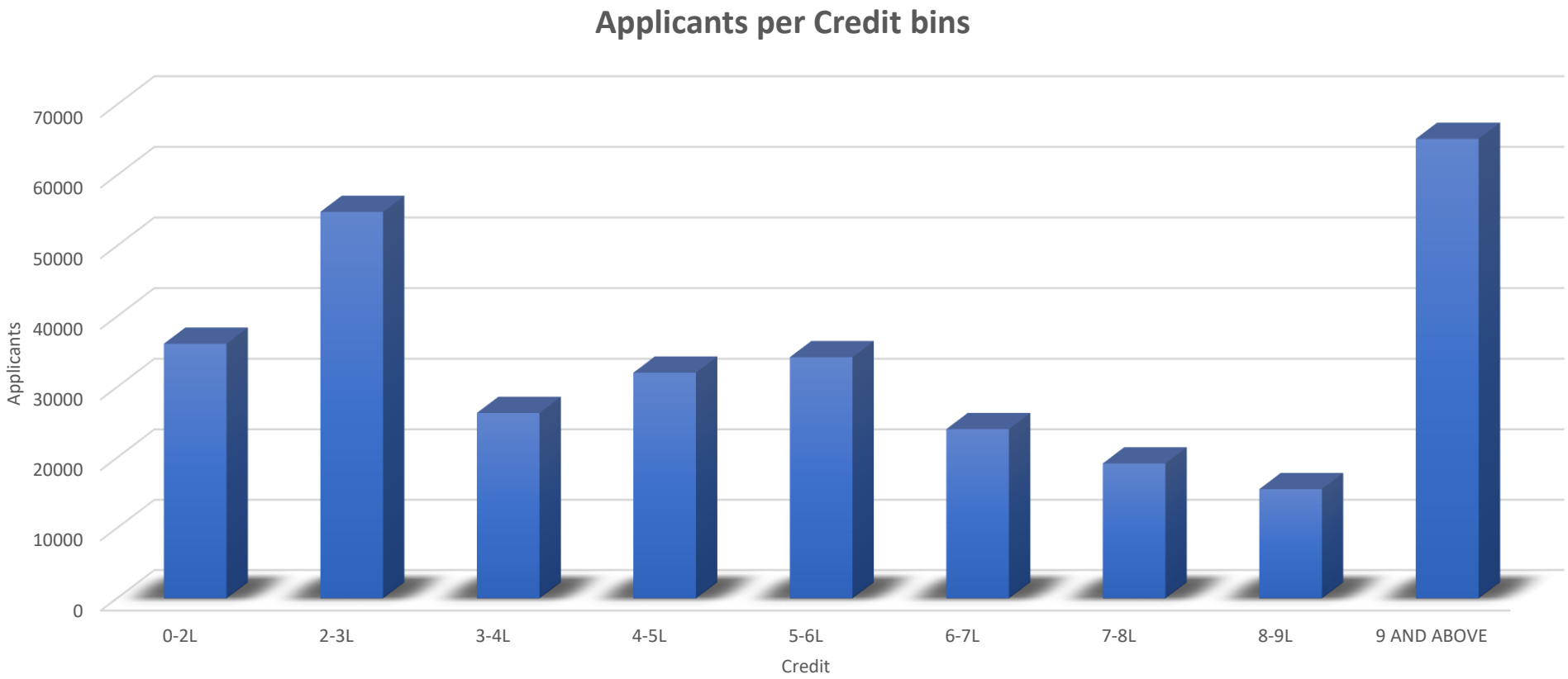
Analyze Data Imbalance

After Merging the 2 cleaned datasets, this was the loan repayment status. 12 lakh repayors and less than a lakh defaulters.

This may seem an insignificant number at first but as of now we don't know how much credit these defaulters account for. Its crucial to assess and avoid these defaulters.

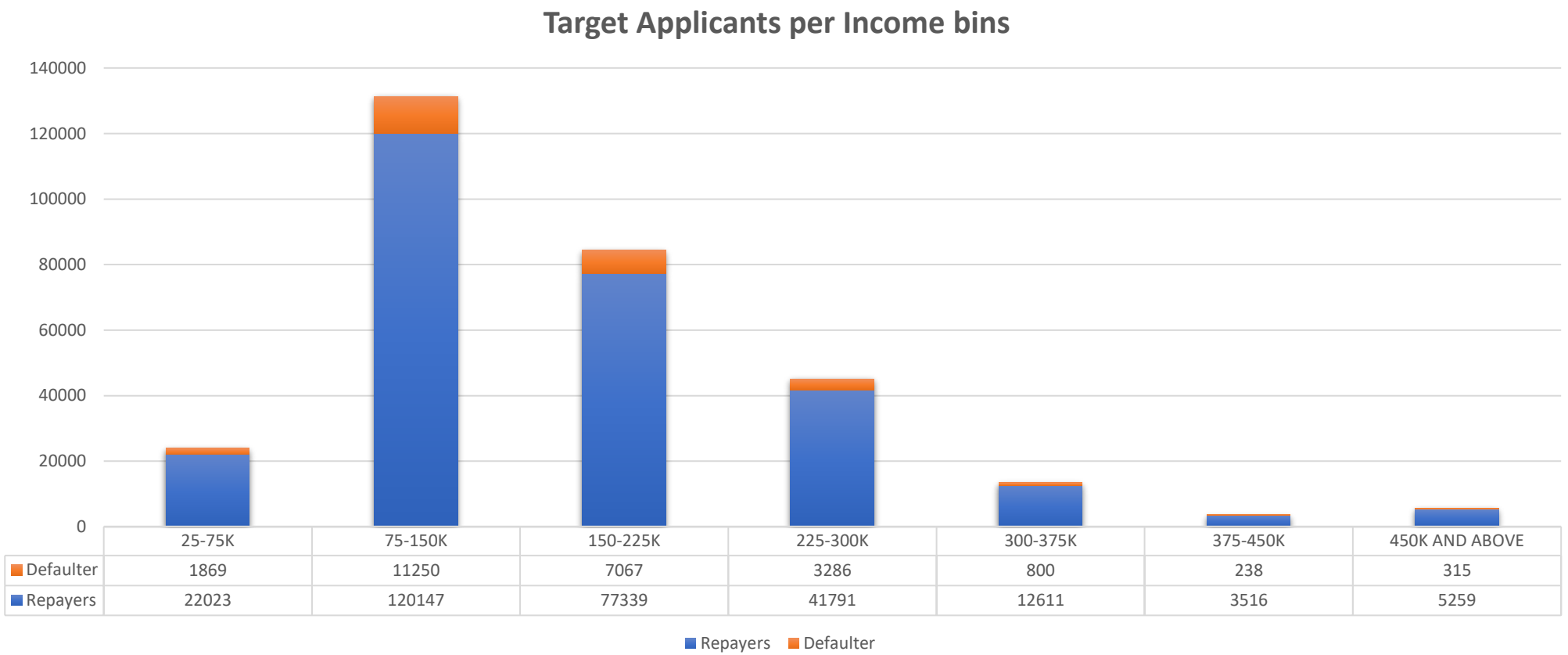


Univariate, Segmented Univariate, and Bivariate Analysis



This Univariate Analysis showcases the number of applicants per credit bins. Most applications are for loan of more than 9 lakhs

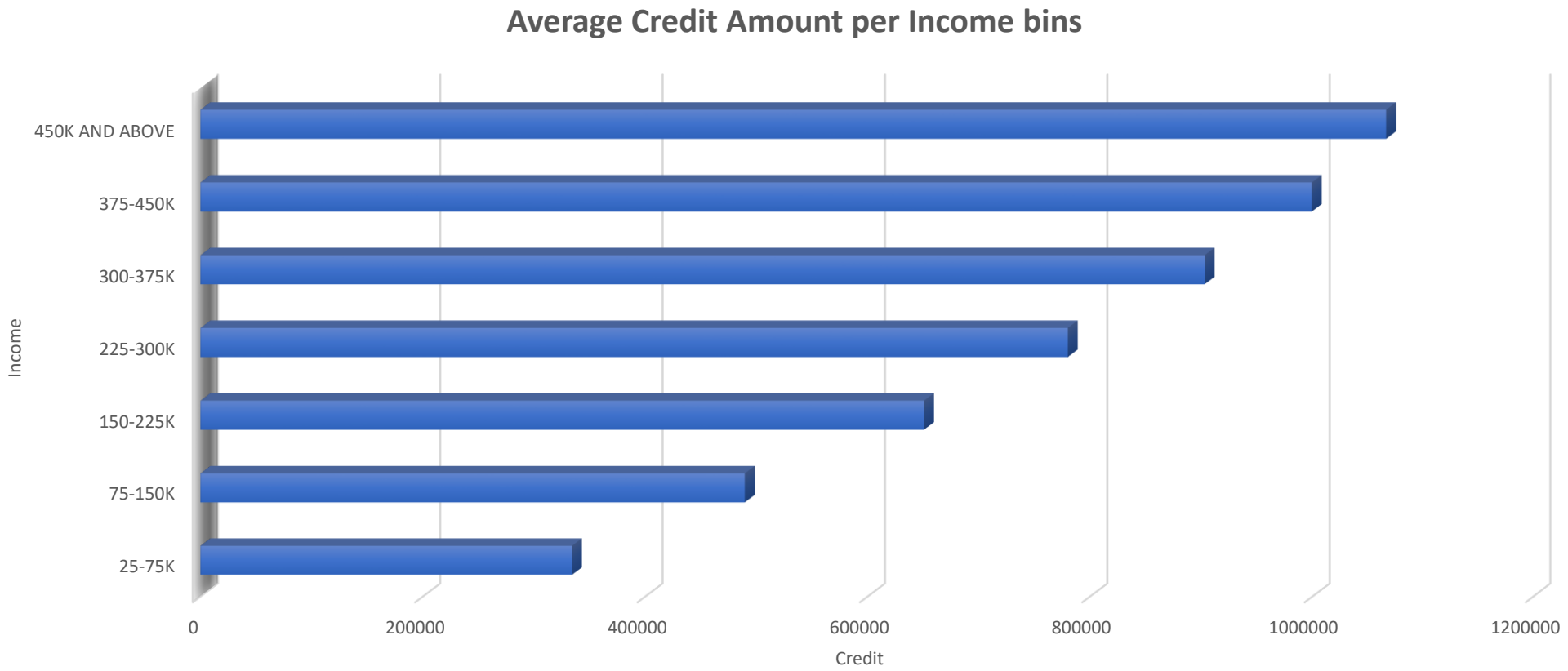
Univariate, Segmented Univariate, and Bivariate Analysis



	25-75K	75-150K	150-225K	225-300K	300-375K	375-450K	450K AND ABOVE
Defaulter	1869	11250	7067	3286	800	238	315
Repayers	22023	120147	77339	41791	12611	3516	5259

This Segmented Univariate Analysis showcases the number of target applicants per income bins. People in the income slab of 75-150k generate the most defaulters. Higher income slabs have less defaulters.

Univariate, Segmented Univariate, and Bivariate Analysis



This Bivariate Analysis showcases the Average credit amount per income bins. Income slab of 75k-150k having the highest number of defaulters have an average loan of 490K approx. Higher income slabs have higher average credit amount.

Analysis of Region rating and Gender

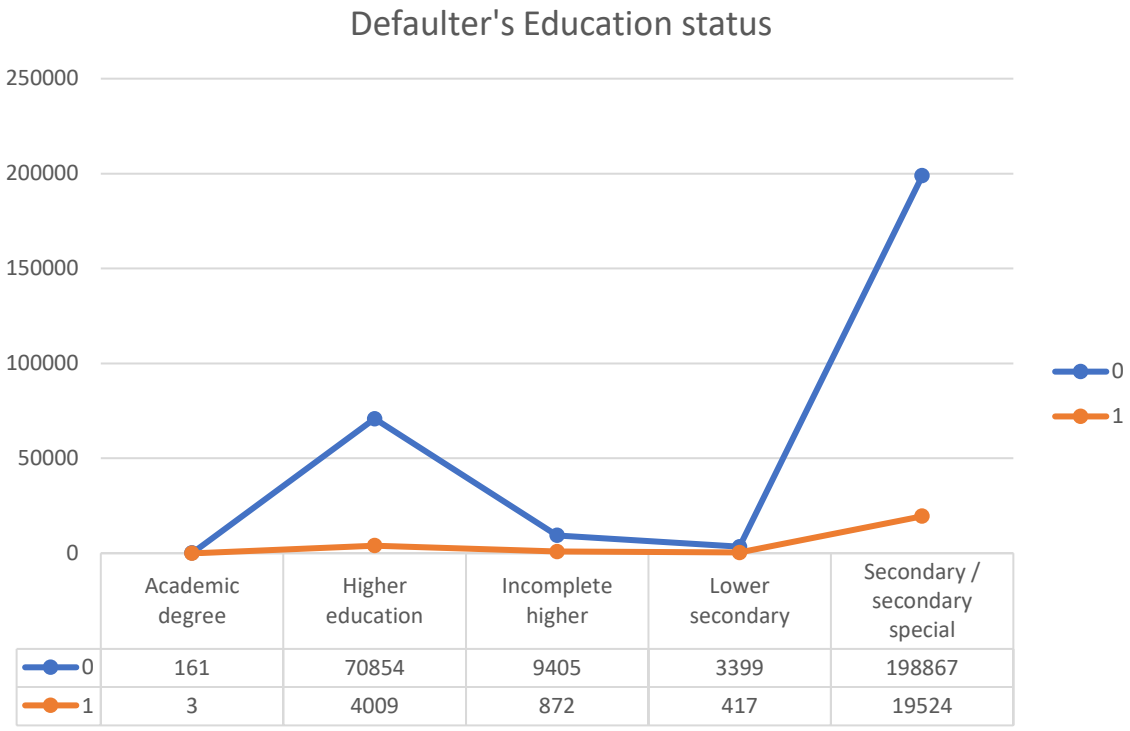
- RATING 1 is safer.
- People who live in Rating 2 has highest defaults.

Count of TARGET	Column Labels		
Region Rating	0	1	Grand Total
1	30645	1552	32197
2	209077	17907	226984
3	42964	5366	48330
Grand Total	282686	24825	307511

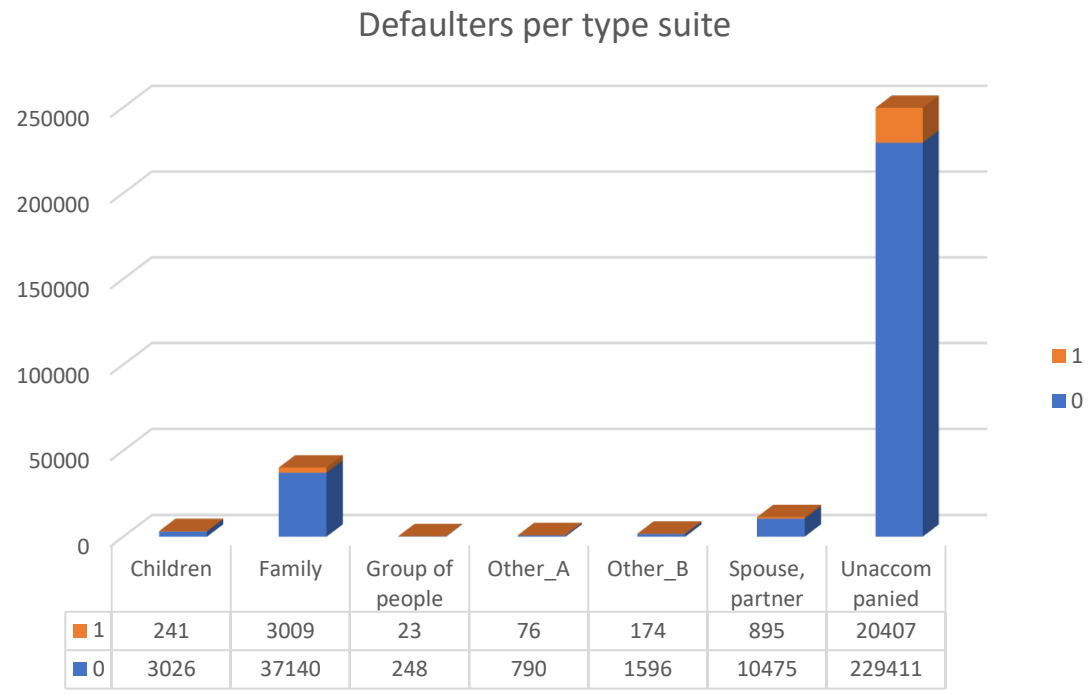
- Men are at relatively higher default rate by percent, but raw numbers show that women have a lot more defaulters.

Count of TARGET	Column Labels			
Gender	0	1	Grand Total	Defaulter percentage
F	188278	14170	202448	7.00%
M	94404	10655	105059	10.14%
Grand Total	282682	24825	307507	

Analysis of Education Status and Suite Type



	Academic degree	Higher education	Incomplete higher	Lower secondary	Secondary / secondary special
0	161	70854	9405	3399	198867
1	3	4009	872	417	19524

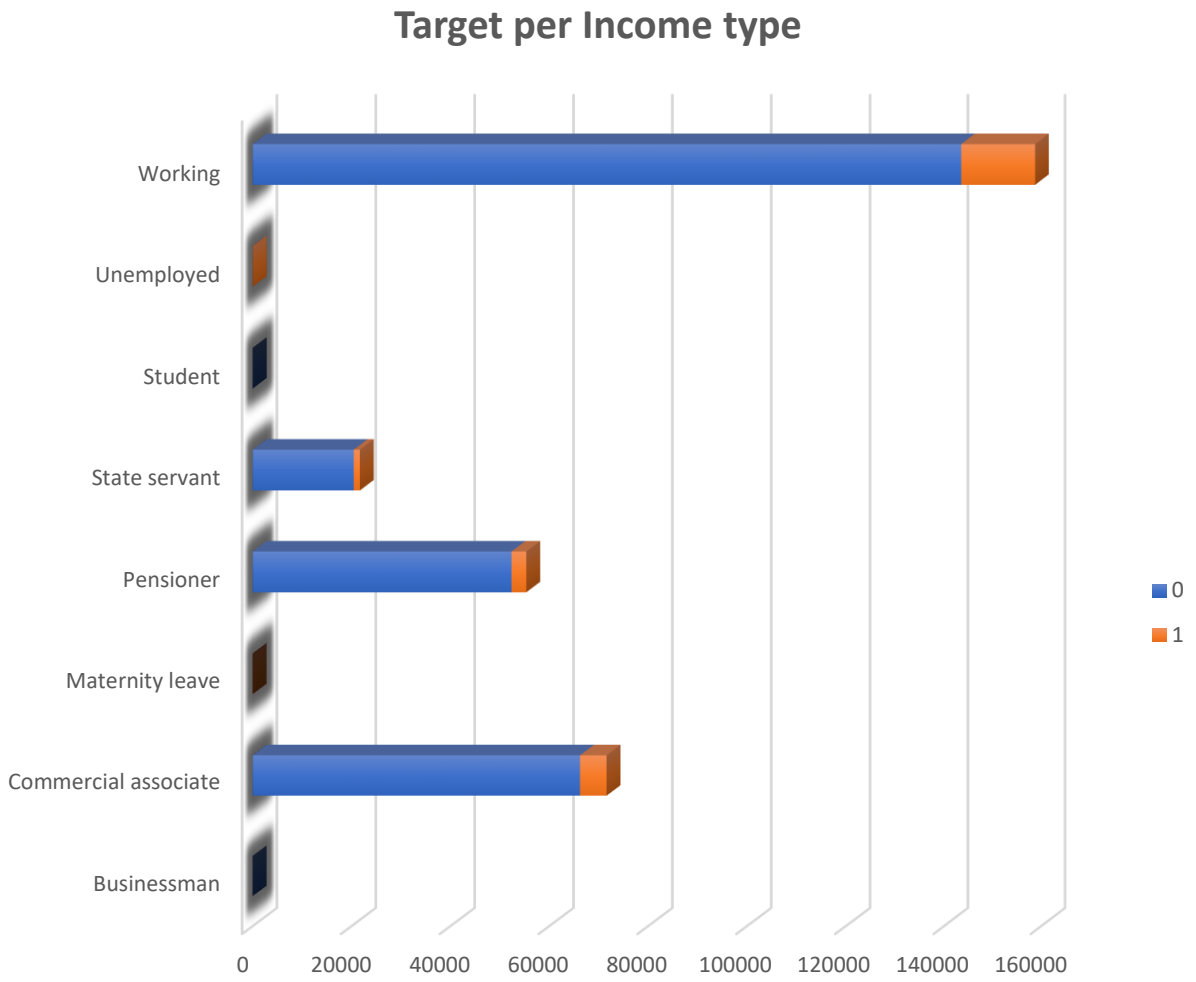


	Children	Family	Group of people	Other_A	Other_B	Spouse, partner	Unaccompanied
1	241	3009	23	76	174	895	20407
0	3026	37140	248	790	1596	10475	229411

- Individuals with a stronger educational background are more prone to loan repayment compared to those with lower educational qualifications.
- Academic degree has less defaults.
- Individuals who are on their own(Unaccompanied) encounter challenges when it comes to loan repayment.

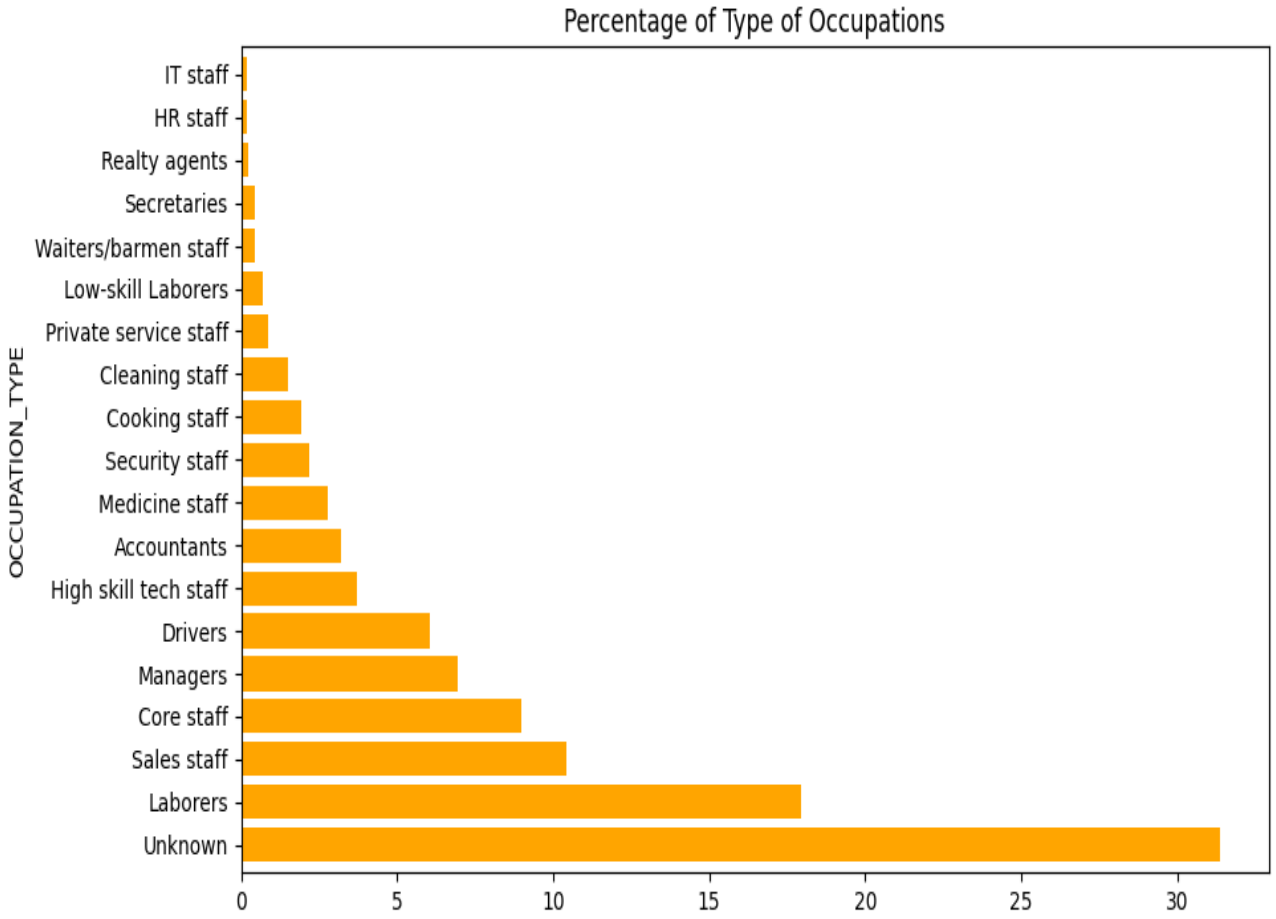
Analysis of Income type

- Student and Businessmen have no defaults.
- The primary defaulters are individuals categorized as part of the working class or involved in business partnerships as Commercial associates. While the unemployed exhibit a lower default rate in terms of numbers, it's important to acknowledge that a significant proportion of unemployed individuals face loan approval challenges.



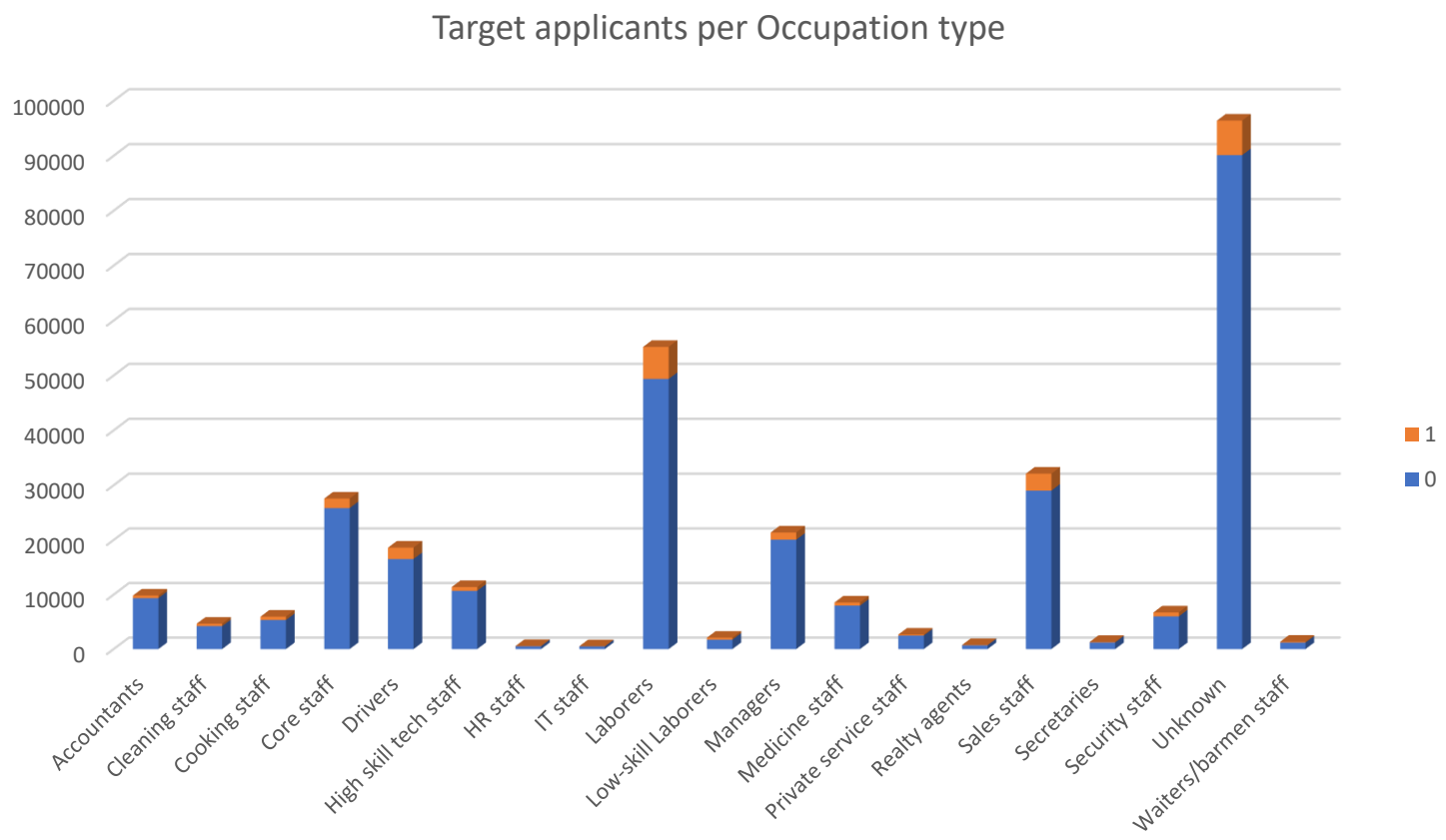
	Businessman	Commercial associate	Maternity leave	Pensioner	State servant	Student	Unemployed	Working
0	10	66257	3	52380	20454	18	14	143550
1		5360	2	2982	1249		8	15224

Analysis of Occupation type



- We lack information about the occupation of over 30% of our applicants.
- The largest volume of loan requests comes from laborers, followed by individuals employed in sales roles.
- Let's take a look at how these factors effect the loan repayment status

Analysis of Occupation type

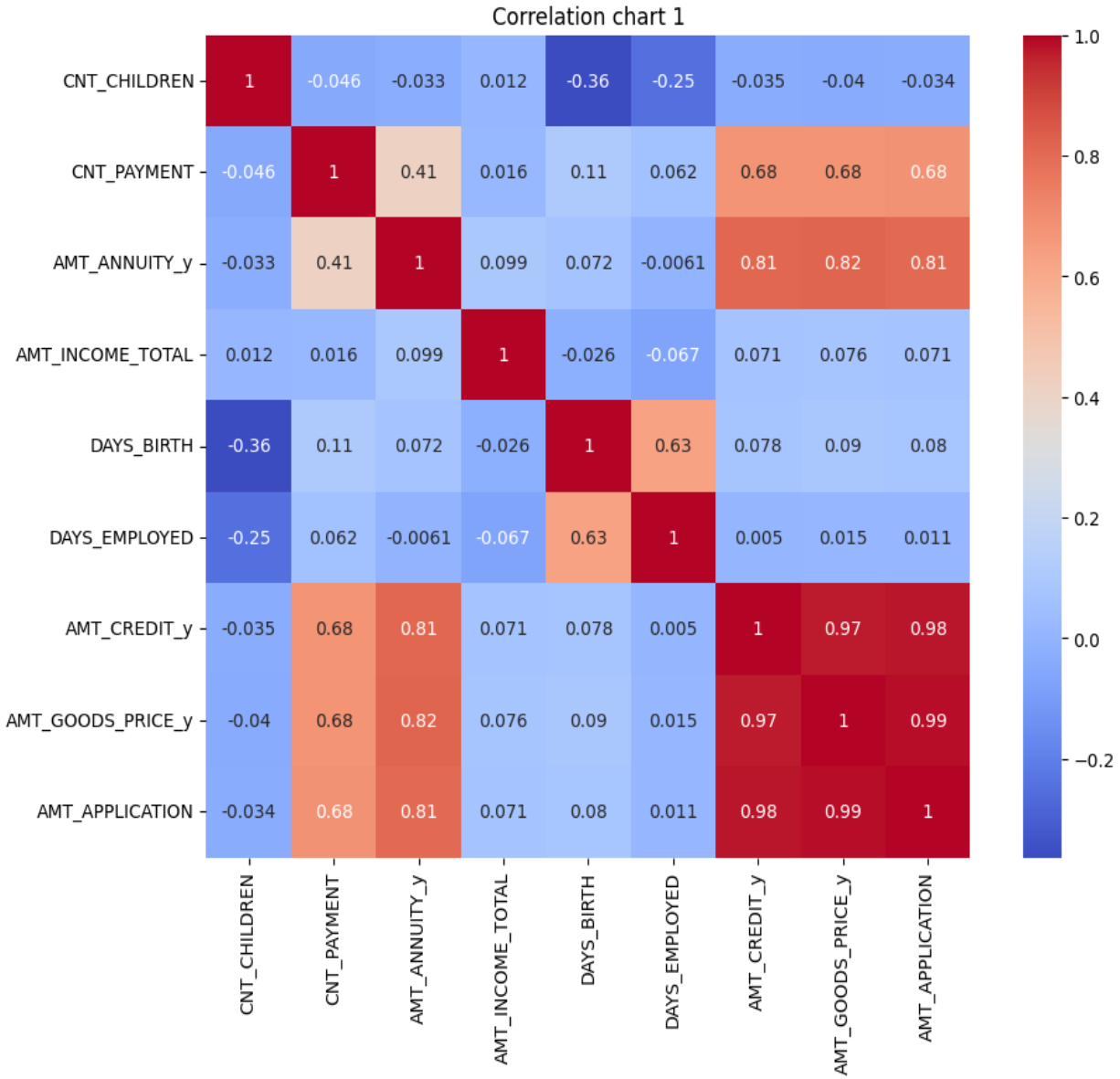


- A significant portion of defaulters have undisclosed occupations, a concerning issue that requires immediate rectification for future cases. It is essential to accurately determine an individual's employment status before approving a loan, especially to mitigate the risk of lending to the unemployed.
- Additionally, laborers exhibit a notably elevated number of instances where loan repayment defaults occur.

Top Correlations

We can see strong correlation at the bottom right side of this correlation chart between Credit, Goods Price and Amount Application.

Amount Annuity shares a very strong correlation with these columns as well.



Insights

Repayors

- 90.6% of loan recipients demonstrate a commendable ability to meet their repayment obligations.
- Higher income slabs have less defaulters.
- Region Rating of 1 is safer to give loan to.
- Individuals with a stronger educational background are more prone to loan repayment.
- Student and Businessmen have no defaults.
- People with zero to two children tend to repay the loans.
- People living with families mostly payback the loan.

Defaulters

- While the remaining 9.4% represent defaulters.
- Most number of defaulters are from income bins of 75-150k,150-225k,225-300k
- Region rating 2 having many defaulters comparatively.
- People with Lower Secondary & Secondary education.
- The primary defaulters are individuals categorized as part of the working class or involved in business partnerships as Commercial associates
- A significant portion of defaulters have undisclosed occupations.
- laborers exhibit a notably elevated number of instances where loan repayment defaults occur.
- Individuals who are on their own(Unaccompanied) encounter challenges when it comes to loan repayment.

Suggestions

- A significant portion of defaulters have undisclosed occupations, a concerning issue that requires immediate rectification for future cases. It is essential to accurately determine an individual's employment status before approving a loan, especially to mitigate the risk of lending to the unemployed.
- As we can see that many defaulters are laborer's ,therefore bank should either avoid their applications or give them special privilege and loan on less interest.

THANK YOU!

CONTACT INFO : VANSHTHAKUR0425@GMAIL.COM