Task 14: Loan Case Study

• **Project Description:** The project is about analyzing the Loan case study dataset. The Project involves 2 data sets. Where 1st dataset represents the previous application of the customers which has a detailed information about the customers like education, job title, housing type etc. It also includes purpose of loan and loan status whether approved, rejected, cancelled or unused offer. The 2nd dataset represents new loan application which also include detailed information about the customers like education, job title, housing type etc. It also includes a target column which represents whether the customer earlier applied was a defaulter or not.

• **Approach:** First step is to understand the dataset and most important factors which are considered when a loan is to be approved. From a set of questions around those factors which must be clear in order to make a decision of loan approved or rejection. Import the dataset and inspect the dataset. Find the outliers which may affect the solutions of the questions. Find the null value column wise as well as row wise. Find the columns having null values more than 40% and drop those columns. Find the percentage of null values and try and fill out nearest possible value if the percentage of null values are low. Drop out unnecessary columns. Identify the outlier and describe according why it was considered as an outlier. Create Bins for the continuous values like loan amount and salary. Find the imbalance ratio. Then perform variant analysis and find top 10 correlation for the client. Plot some graph and try to find out useful insights.

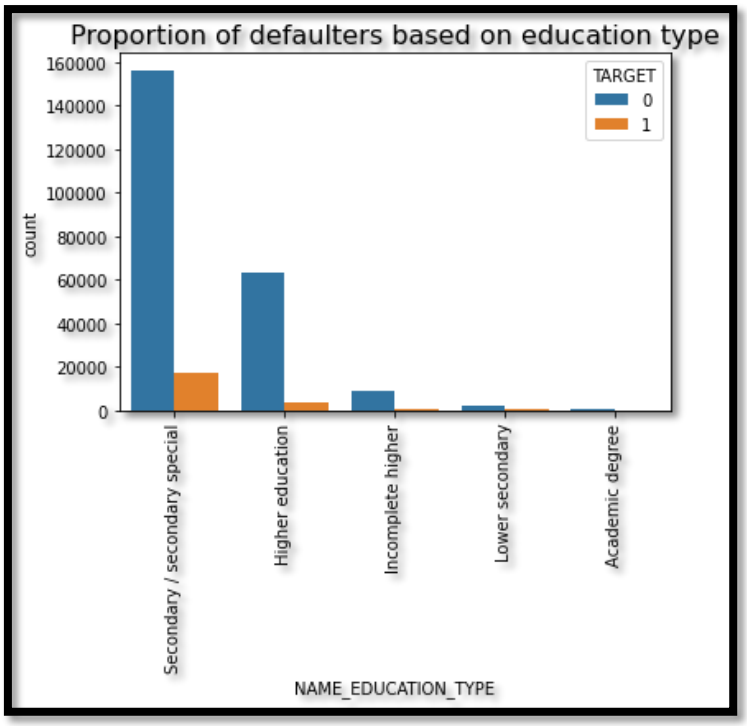
• Tech-Stack Used:

Software used: - Google Colab
Version: - Python 3
Why: - Hosted by jupyter Notebook free and simple to use requires no setup.

• Insights:

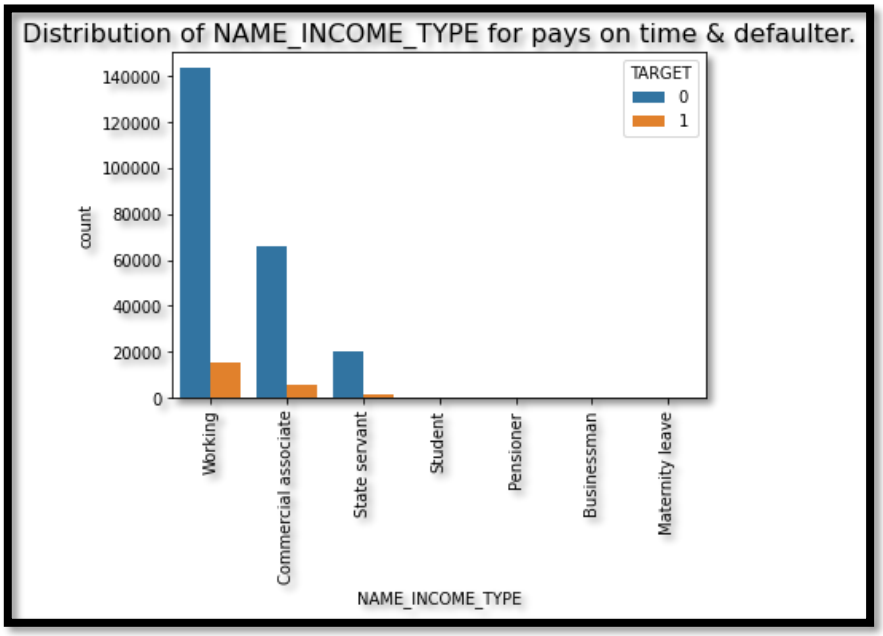
Proportion of defaulters based on education type

- Higher the education less the default.



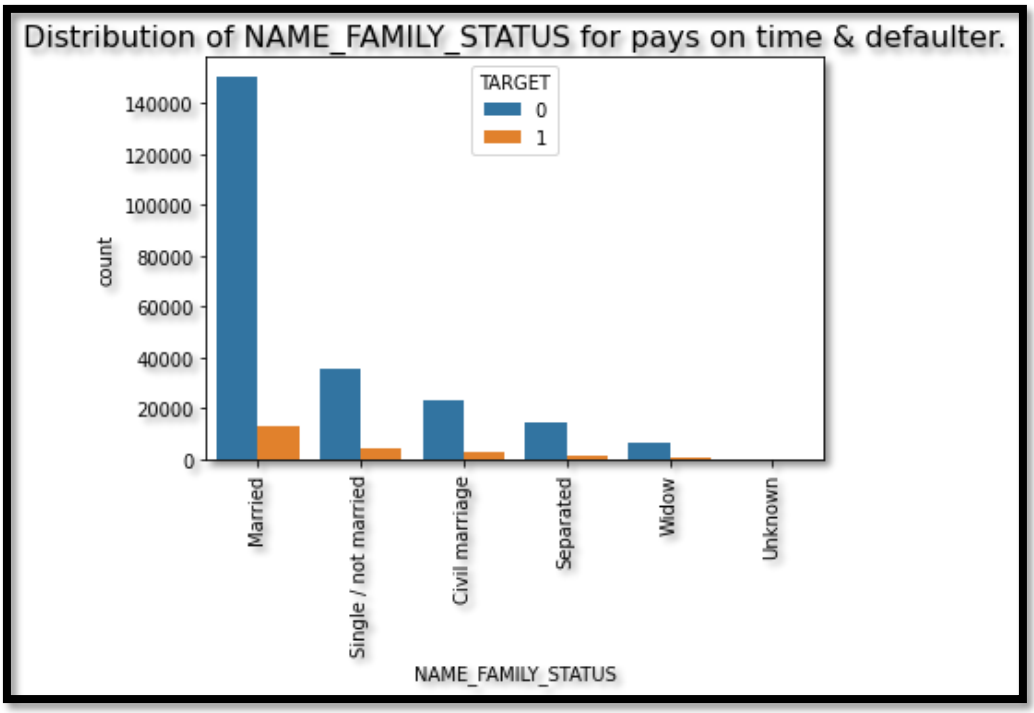
Proportion of defaulters based on Income Type

- Most the customer who applied for the loan belong to working income type.
- Student, Pensioner, Businessman must give more loan has they have maximum probability paying back on time
- Better returns of loan as income increases.



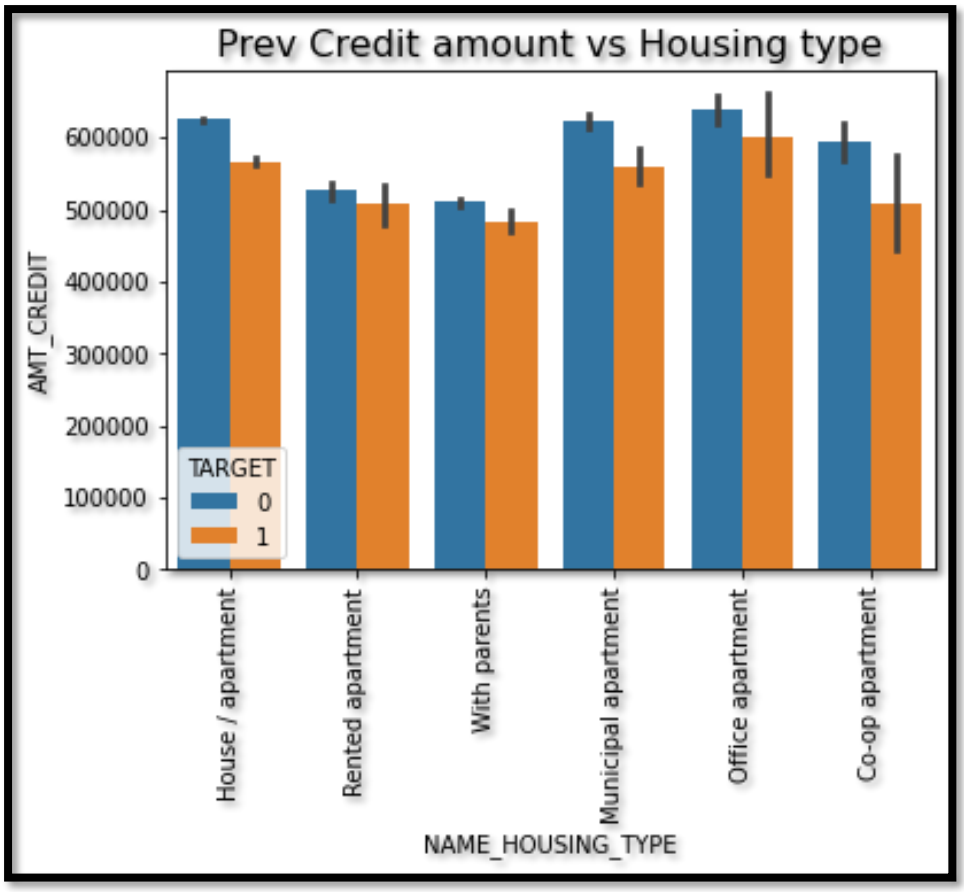
Proportion of defaulters based on Family Status

- Married people have less possibility of not able to pay back loans.
- Single people default more.



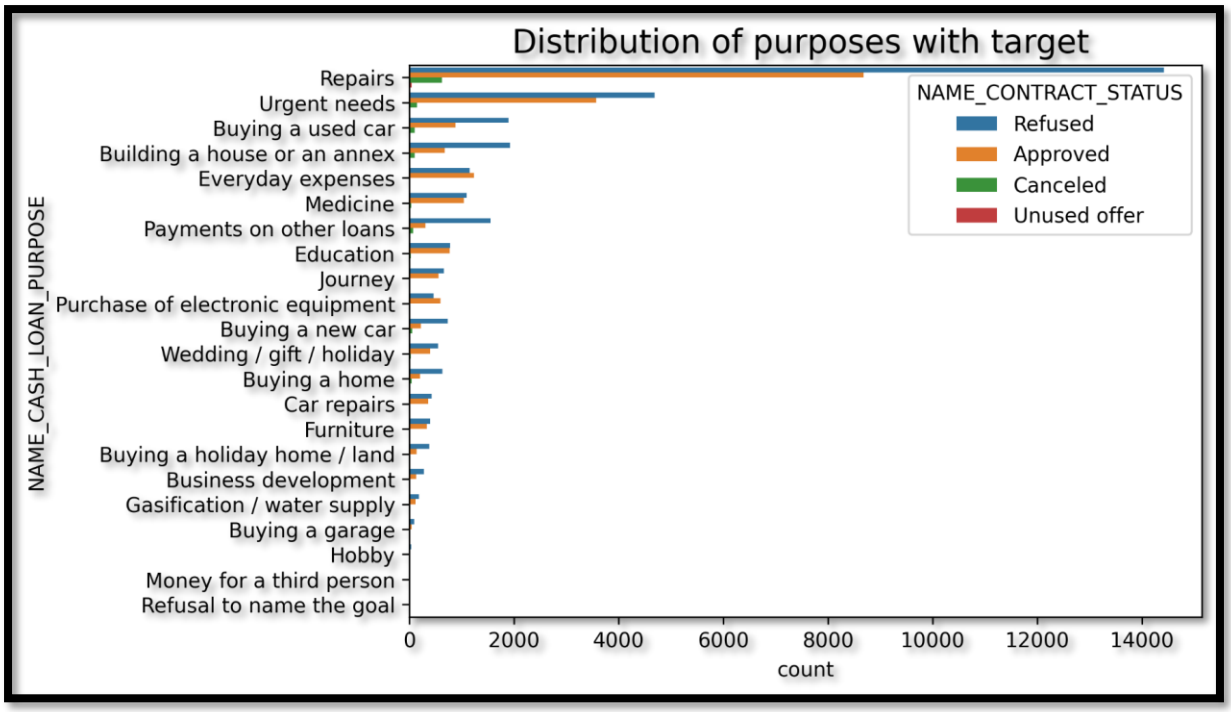
Proportion of defaulters based on Previous amount credit and housing type.

- Customer having Name_housing_type as with parents have paid the loan on time followed by House/apartment and Municipal Apartment.
- Most of Co apartment customers were unable to pay back the loan.



Distribution of purpose based on target

- Maximum difference for approval and rejection was found for buying a new car.
- Maximum Refused & Approved loan belongs to repair purpose.



Conclusion

- The proportion of defaulter is 9.48%.
- Maximum customer applied & approved for the loan were females.
- Most of the cash loans are in default so bank must provide more resolving loans.
- Higher the education less the default.
- Higher the income lesser the default.
- Customer having Name_housing_type as with parents have less default.
- Most of Co apartment housing type customers are defaulters.
- Giving loans to married customer is safer as it has lower default rate.
- Maximum difference for approval and rejection was found for buying a new car.
- Maximum Refused & Approved loan belongs to repair purpose.

• **Result:** This project has helped me strengthen the basic concepts. While project was challenging enough to test basic understanding & helped me learn new things. This one project touched every concept of data analytics and learnt new things which things would be considered loan.

• **Drive Link:**
https://drive.google.com/drive/folders/1p_XF29nYUtosY7ZqzBpTf9DSC_yajLVj?usp=sharing