

The background is a light blue gradient with several realistic water droplets of various sizes scattered across the surface. The droplets have highlights and shadows, giving them a three-dimensional appearance.

EDA Assignment Summary by **Susmit Chakraborty**

Problem Statement

The Bank find it hard to give loans to the people due to their insufficient and non-existent credit history. Due to this reason, some consumers take advantage by becoming a defaulter. Considering myself as an employee in that bank and have to use EDA to analyze the data present in the datasets. This will ensure that the applicants capable of repaying the loan are not rejected.

When the bank receives a loan application, the bank has to decide for loan approval based on the applicant's profile.

There are two types of risks associated with the bank's decision:

1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the bank.
2. If the applicant is not likely to repay the loan, then approving the loan results in a financial loss for the bank.

Assumptions

1. Some of the columns are considered to be **unused**, therefore they are **dropped** from both the databases.
2. Some columns have more than 30 % data missing, those are **dropped** before analysis.
3. **Missing values** are handled using either **median or mode**.
4. **Outliers** are capped into a upper limit like **0.75 quantile** or **0.9 quantile** value.

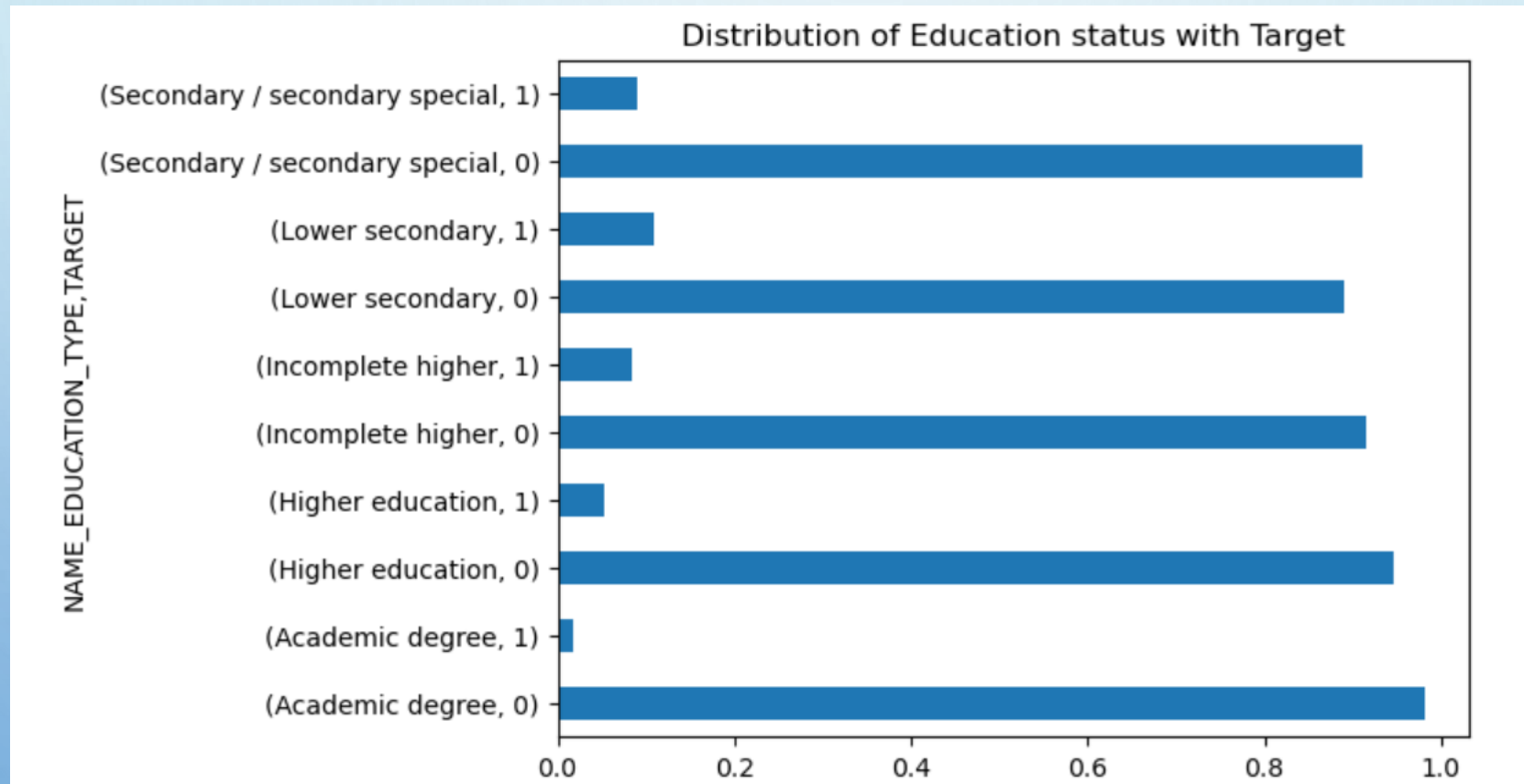
Approach & Methodology

1. Data **importing** and understanding
2. Checking **Sanity** for different columns
3. **Removing** the unimportant columns
4. Checking and handling the **missing values** with either mode/median
5. Checking for **outliers** and handling them with either binning/capping
6. **Univariate** analysis using Bar-plot, Barh-plot, Histogram, Pie-Chart, Boxplot, etc.
7. **Bivariate** analysis using Bar-plot, Barh-plot, , Scatterplot, Pair-plot, etc.
8. **Multivariate** analysis using Correlation matrix, Heatmap, Boxplot, etc.
9. **Insight** generation from all above analysis.

Graphs and Insights

INSIIGH 1 :

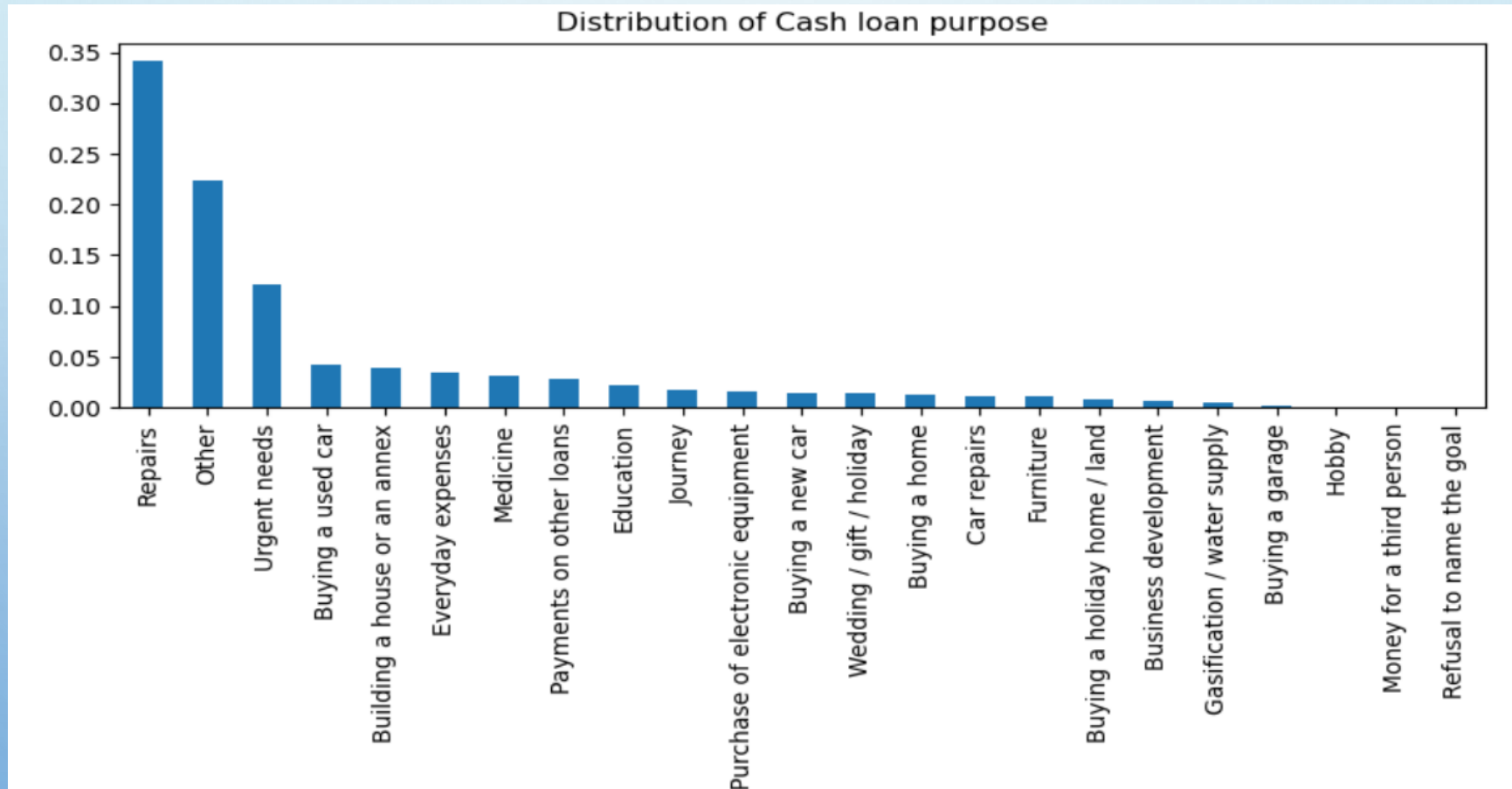
Persons with Secondary / secondary special type of education are the best choice for targeting.



Graphs and Insights

INSIIGH 2:

Those who are applying loan for repairing should be treated as most successful persons.

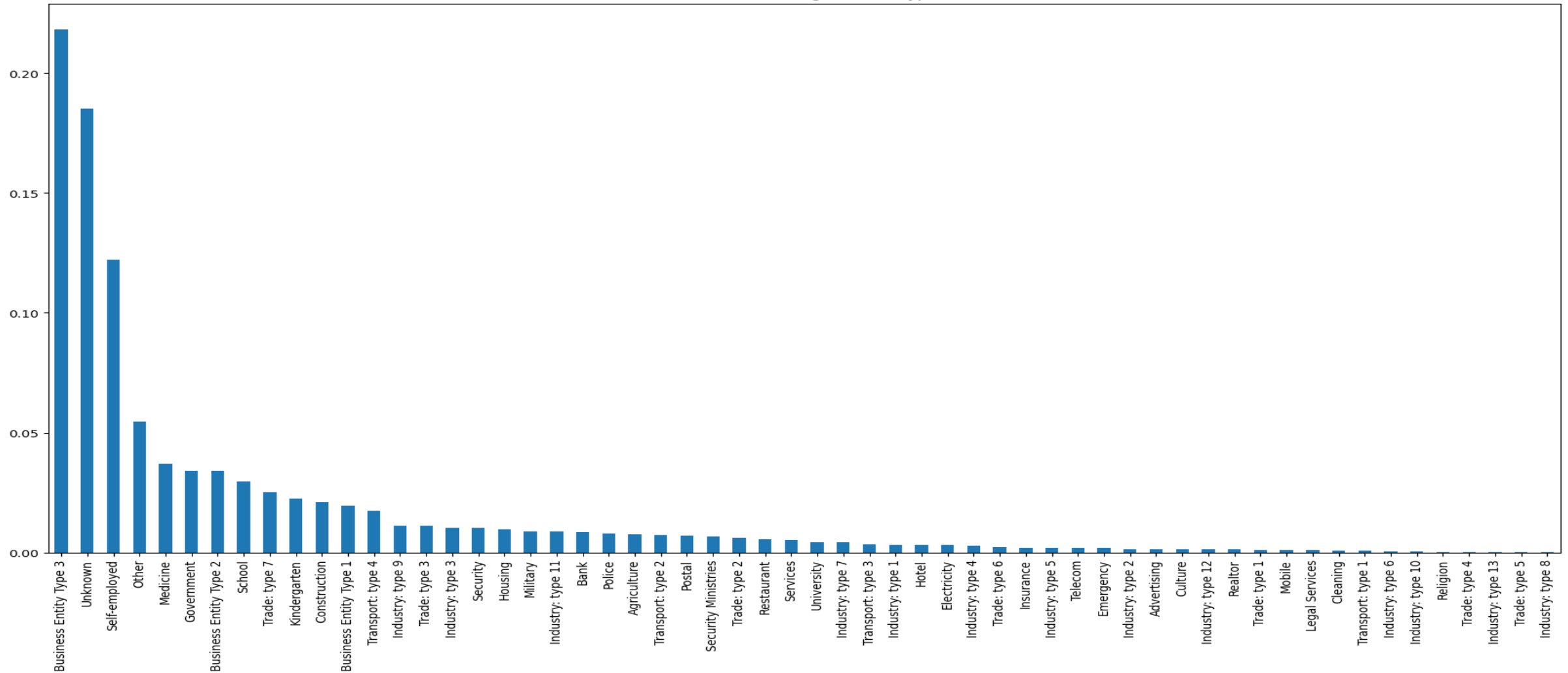


Graphs and Insights

INSIIGH 3:

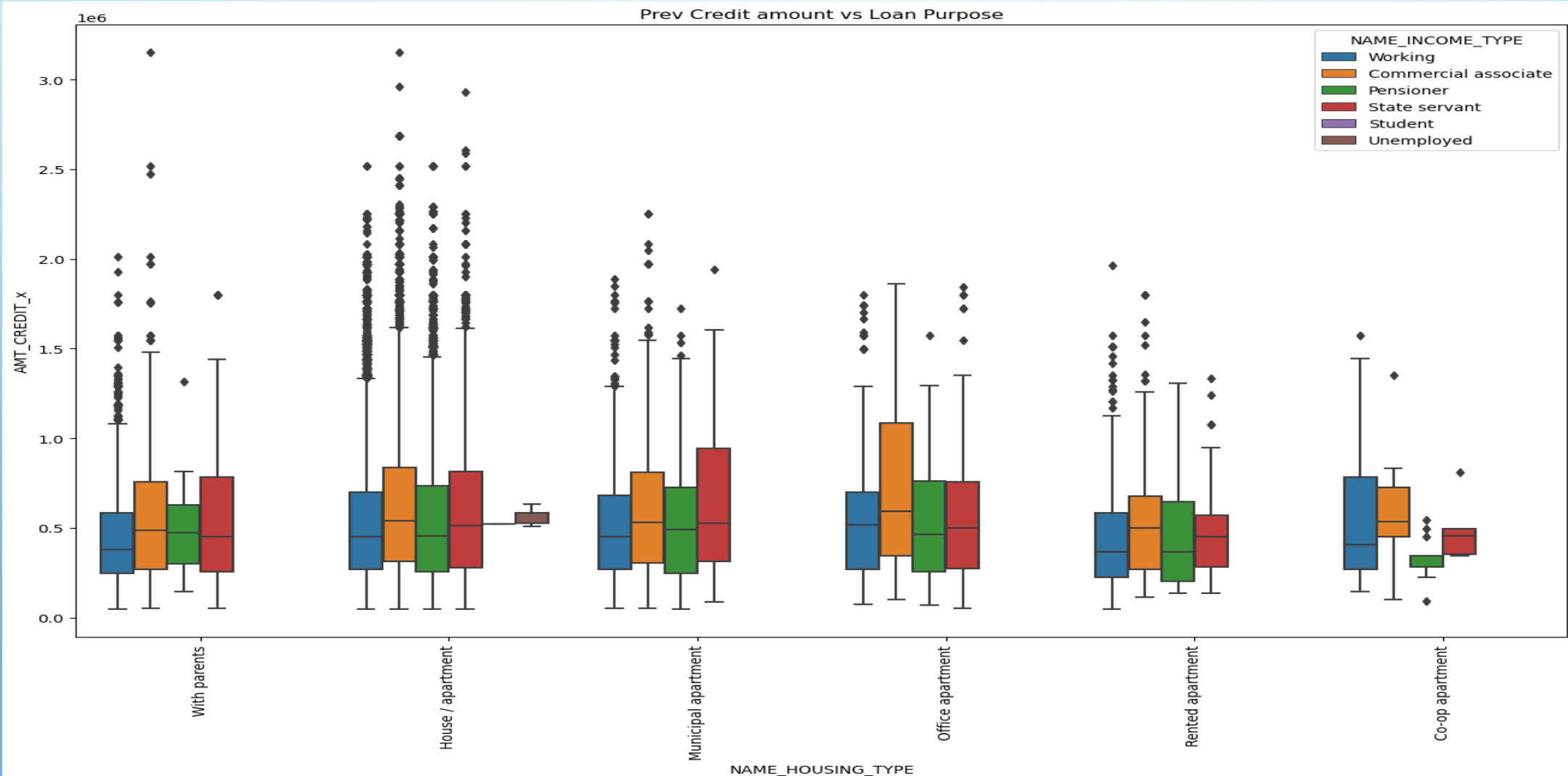
Bank should focus on 'Business Entity Type 3' & 'Self-employed' as organization type for successful payment.

Distribution of Organization type



INSIIGH 4: Graphs and Insights

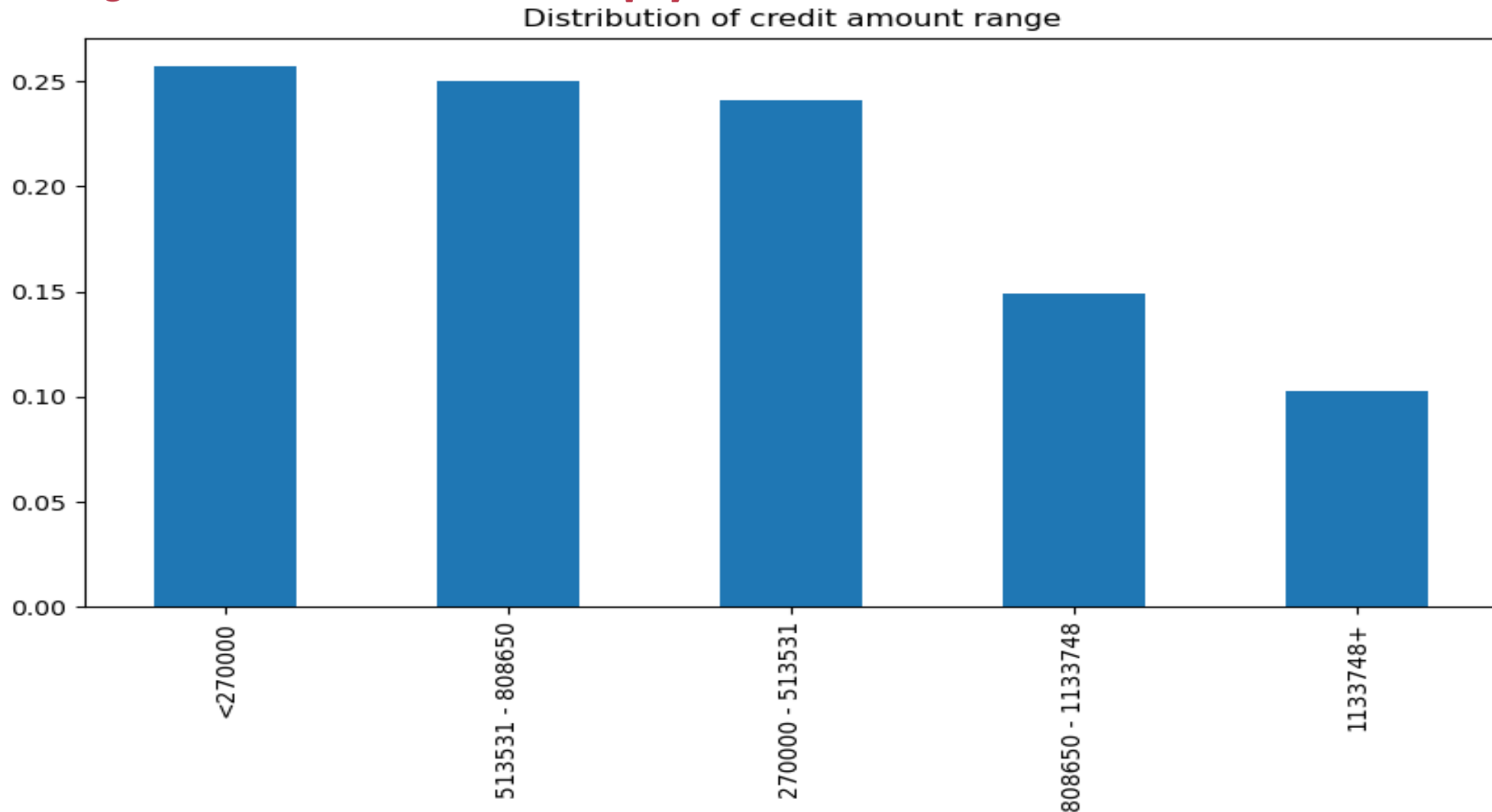
Bank should focus on 'students', 'Businessmen' and 'pensioners' with 'Office apartment' , 'Municipal apartment' & 'House/apartment' as housing.



Graphs and Insights

INSIIGH 5 :

Higher credit generates risk of unsuccessful payments thus amount for the loan should be less than Rs.500000



Conclusion

After univariate, Bivariate and multivariate analysis it can be concluded that there are lots of insights in the database, but only important are prescribed the previous slides.

Besides the previous insights, here all are given below:

1. Females are the better option to approve the loan
2. 'Business Entity Type 3' type persons are the best fitted organization for targeting
3. Person with Secondary / secondary special type of education are the best choice for targeting
4. More than 50 % people of the targeted person are working
5. Credit amount for the loan should be less than Rs.500000
6. Loan should be approved for those people who have income ranging between Rs.180000- Rs.210000
7. Loan Annuity between Rs.30000 - Rs.35000 is the best fitted case among all

As per the loan amount is concerned, the below conclusions are made:

1. 'NAME_INCOME_TYPE' = 'Businessman'
2. 'CODE_GENDER' = 'male'
3. 'NAME_EDUCATION_TYPE' = 'Higher education' & 'Academic Degree'
4. 'NAME_FAMILY_STATUS' = 'Married'
5. 'ORGANIZATION_TYPE' = 'Industry: type 8' , 'Legal Service' , 'Realtor' & 'University'

Conclusion (Cont.)

After analysis the 'TARGET' = 1 category type data, the following insights can be stated:

1. 57 % of the person those are in 'TARGET' = 1, are female but as per the counting concerns Females are the better choice than male.
2. Though more than one fourth persons in 'TARGET' = 1 are belong to Business Entity Type 3, it remain the best choice of organization.
3. 78 % have education qualification is Secondary/ secondary special in 'TARGET' = 0 database, but it remains the best choice.
4. Counting of different income type gives working persons are the best choice in different income categories.
5. lower credit reduces the risk.
6. 18000 - 21000 range of income gives the maximum risk among all ranges, but it remain the best choice for approving the loan.
7. Similarly 30000 - 35000 remain the best choice of annuity range.
8. Correlation matrix and heatmap results very much similar insights as the insights generated for correlation matrix of 'TARGET' = 0 data.

Based on the above analysis made on Previous application data. the following insights can be generated:

1. All the contracts were of Cash loan type
2. 34 % of previous loan due to Repairing
3. 91 % payment was done by cash through the bank
4. Maximum credit amount range was between 100000 to 200000
5. More than 40 % of the loan was with the high rate
6. Loan for repairing gave the maximum interest

Conclusion (Cont.)

Conclusions made after analyzing the merged database:

1. Repairing is the most common reason for applying for Loan
2. The credit amount of Loan for the following reasons is higher:
'Buying a home', 'Buying a land', 'Buying a new car' and 'Building a house'.
3. Income type of state servants have a very high credit amount.
4. Office apartment type housing has the most TARGET 0 and Co-op apartment type of housing has the most TARGET 1.

THANK YOU