

# Credit EDA Case Study

**Anitha Ammati**

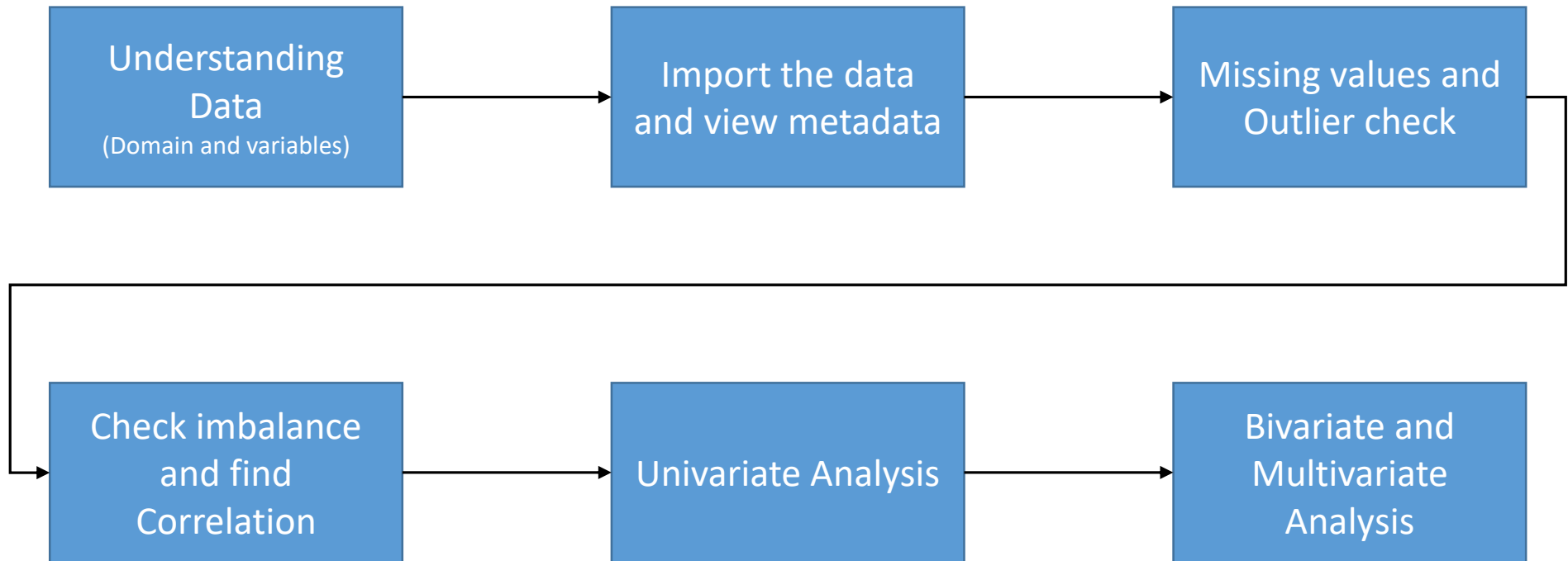
**Apoorva Srinivas**

May 2021

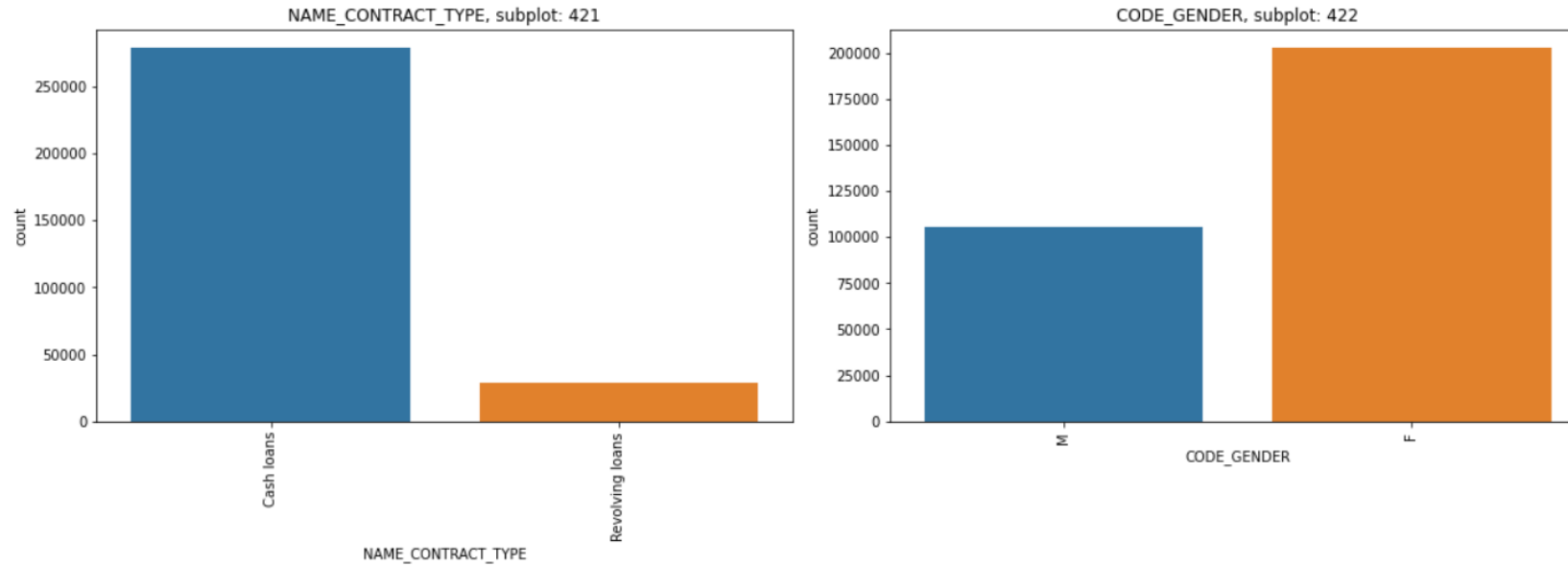
# PROBLEM STATEMENT

- A Finance Company is performing a risk analysis on the clients loan applications.
- A data set containing the applicants current and previous application data is given.
- As analysts, we perform EDA on the given data set and gain insights.
- Our objective is to analyze the data and identify:
  - Clients who are likely to default their loan payments.
  - Clients who are most unlikely to default their loan payments.

# Our Analysis Approach

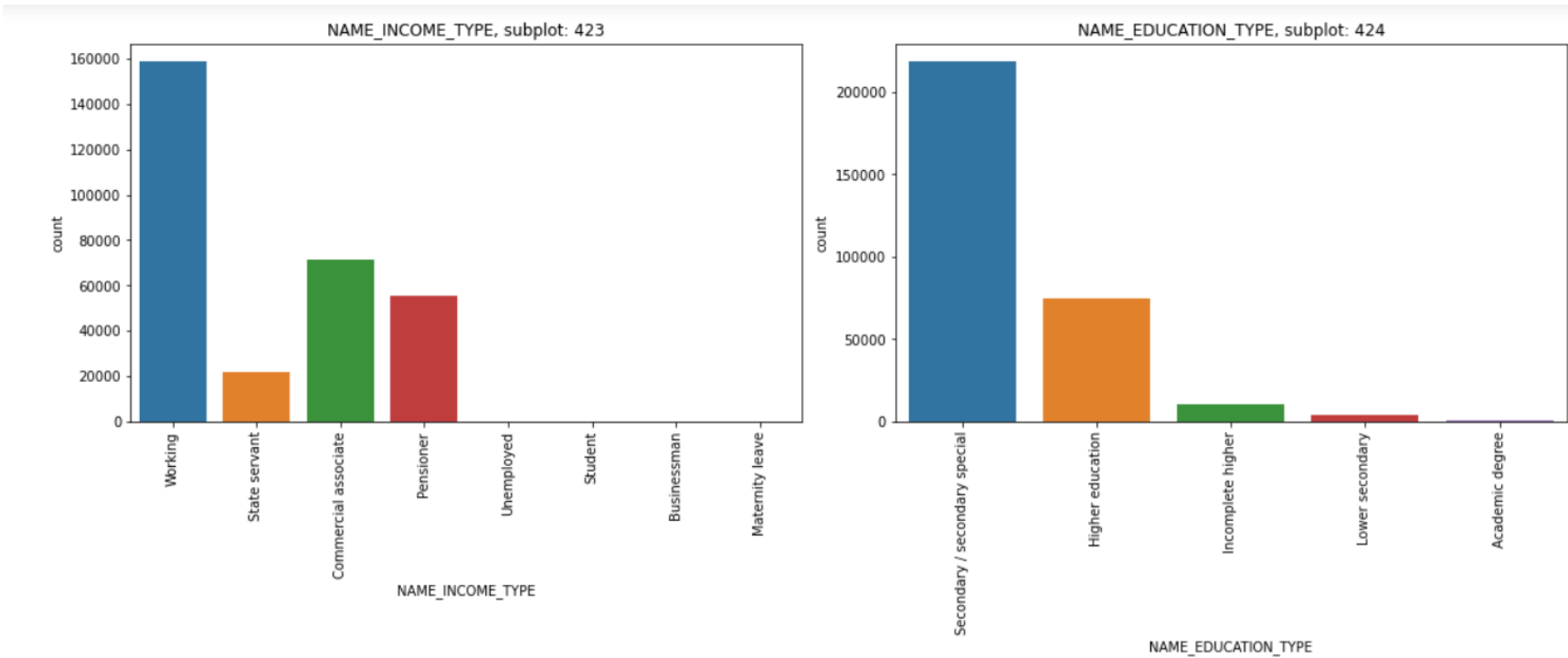


# INFERENCES ON LOAN TYPE & GENDER



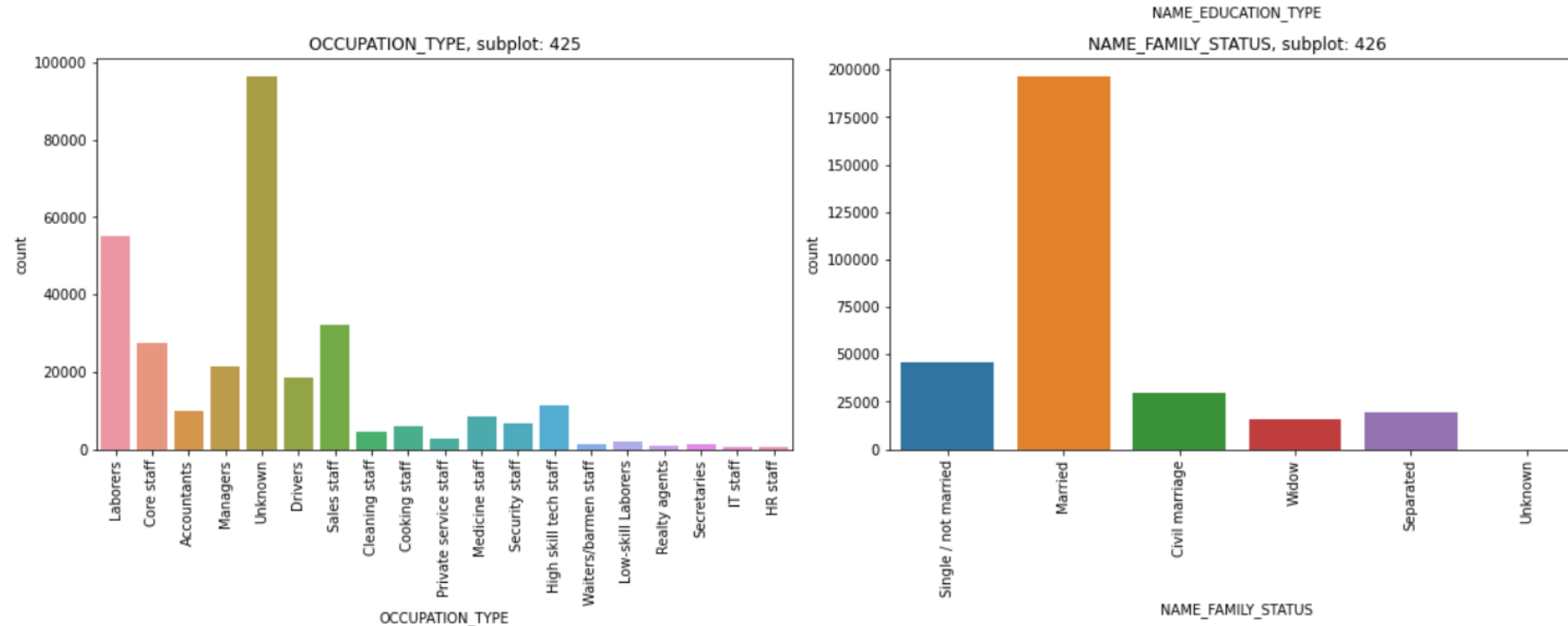
- The total number of applicants who have defaulted is 9% and others who are unlikely to default is 91%.
- Overall, the applications for cash loans (90%) are far greater than revolving loans(10%).
- There are more number female applicants compared to males.
- The data set is highly imbalanced.

# INFERENCES ON INCOME & EDUCATION



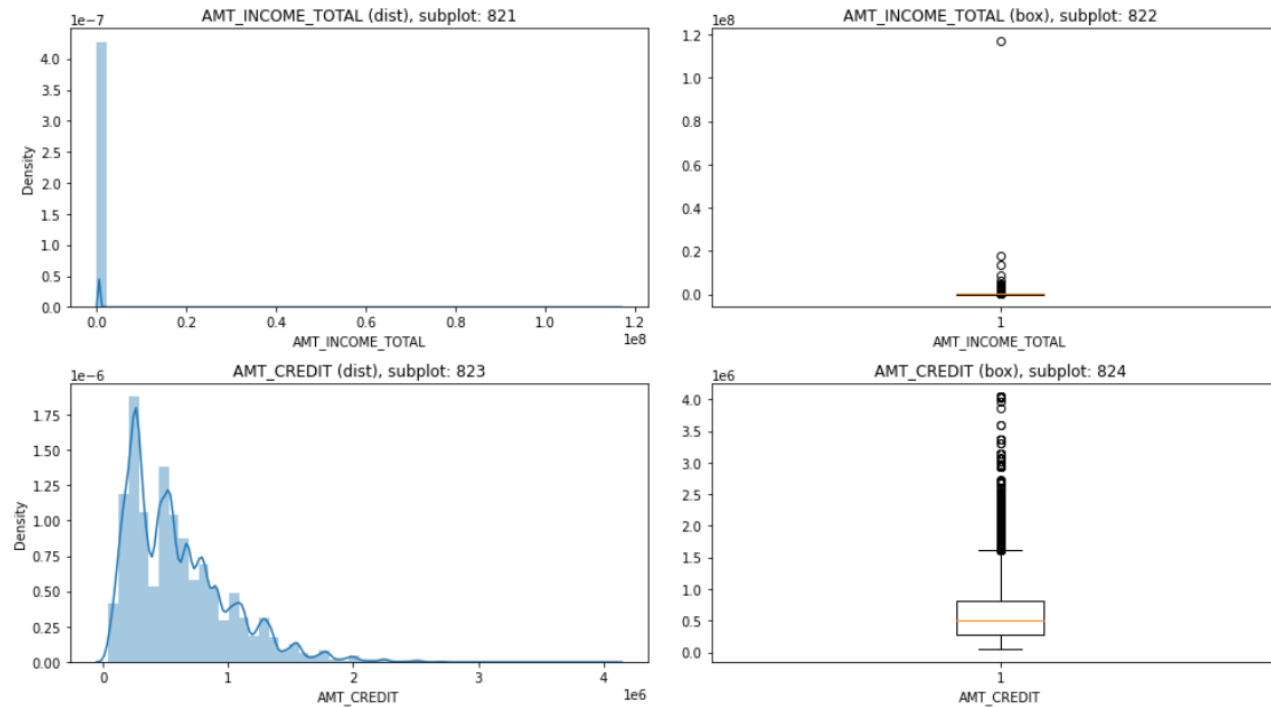
- The majority of the applicants come from the working category of people having secondary type of education.

# INFERENCES ON OCCUPATION & FAMILY



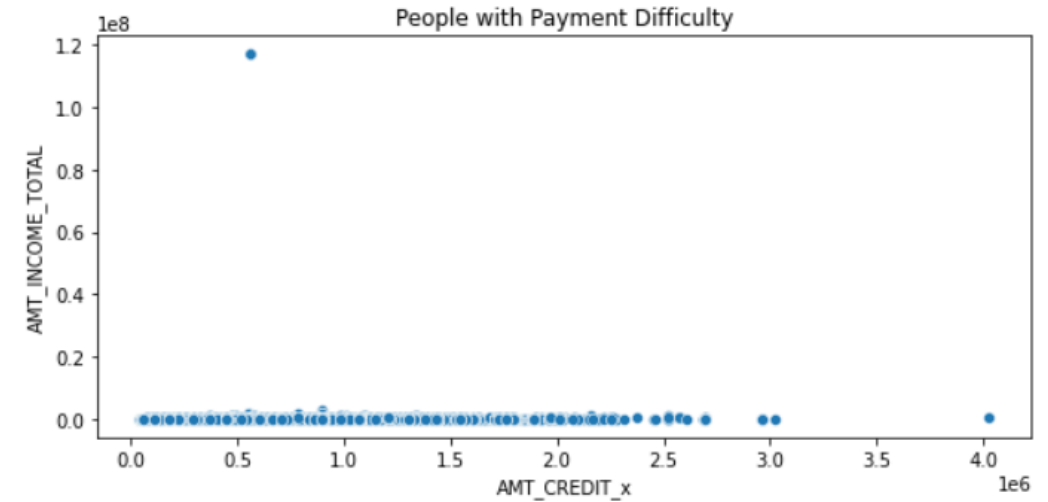
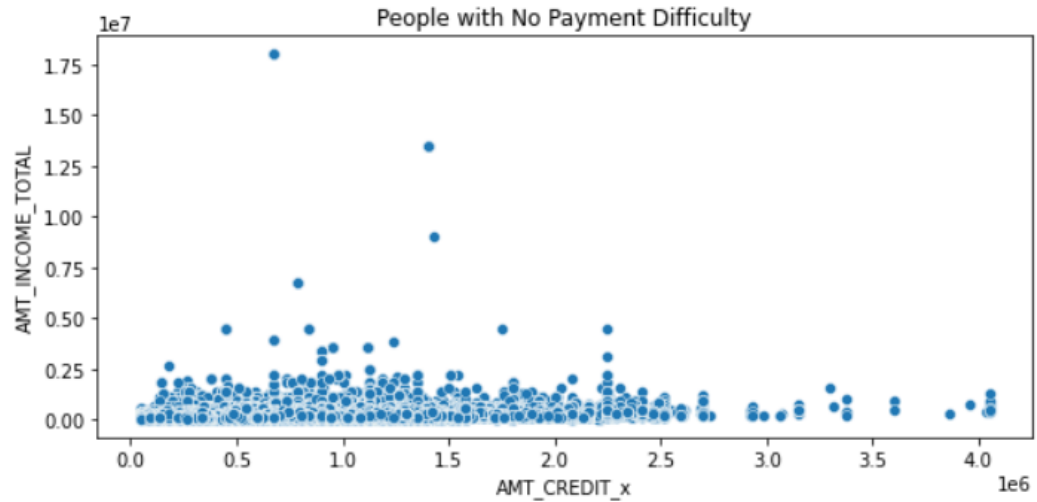
- ▶ The majority of the applicants' occupation is unknown, which means that either the clients have not disclosed their occupation or belong to business, maternity leave, student, pensioner or unemployed categories.
- ▶ The graph clearly shows that most applications belong to Laborers category.

# INFERENCES ON INCOME & CREDIT



- The income of most of the applicants and hence their credit amount lie in the lower range since most of the clients are laborers and their income is in the lower bracket.
- There are outliers in the income total field. These maybe high income businessmen, managers or realtors and we cannot ignore them from our analysis. Therefore, their credit amount is more and hence we can observe outliers in the credit amount field.

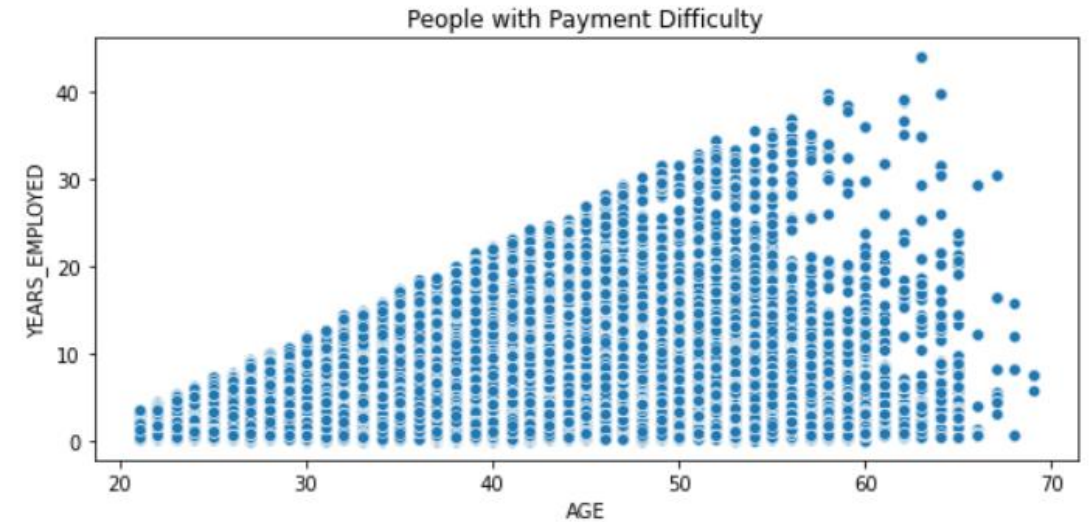
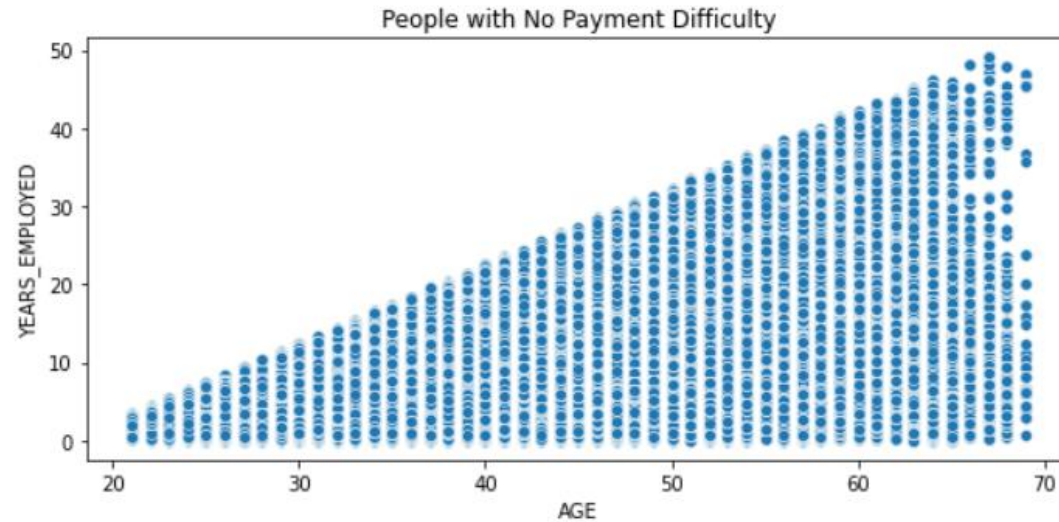
# INFERENCES INCOME AND CREDIT



- ➡ Though the income is in the lower bracket, clients have applied for varied credit amount.
- ➡ Since majority of the applicants are laborers and do not have much savings, they rely on loans to fulfill their needs and desires.
- ➡ People who earn a steady and higher income invest in savings and they are better equipped to handle their finances and hence they do not require large loans.

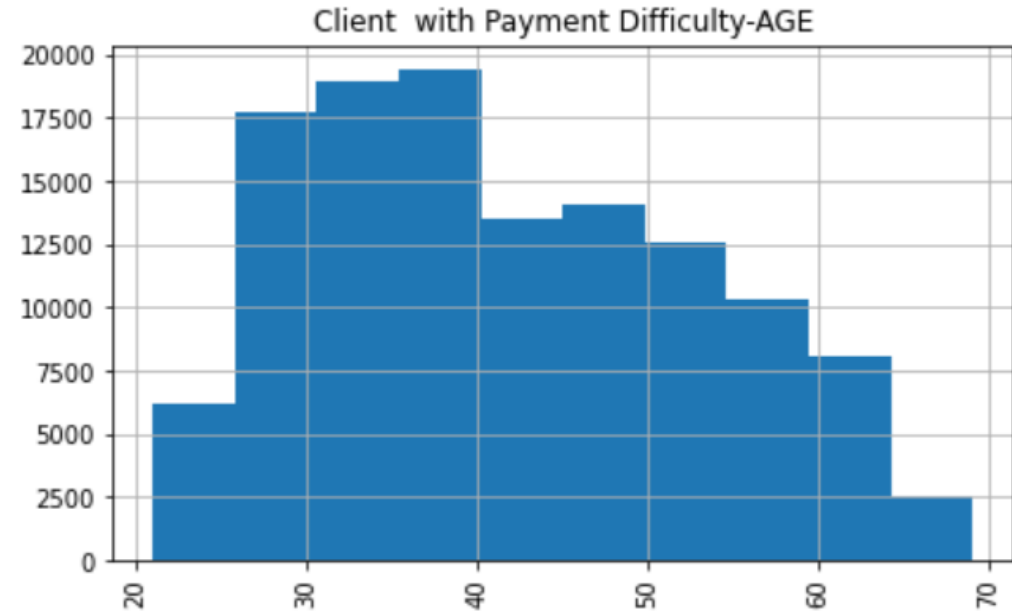
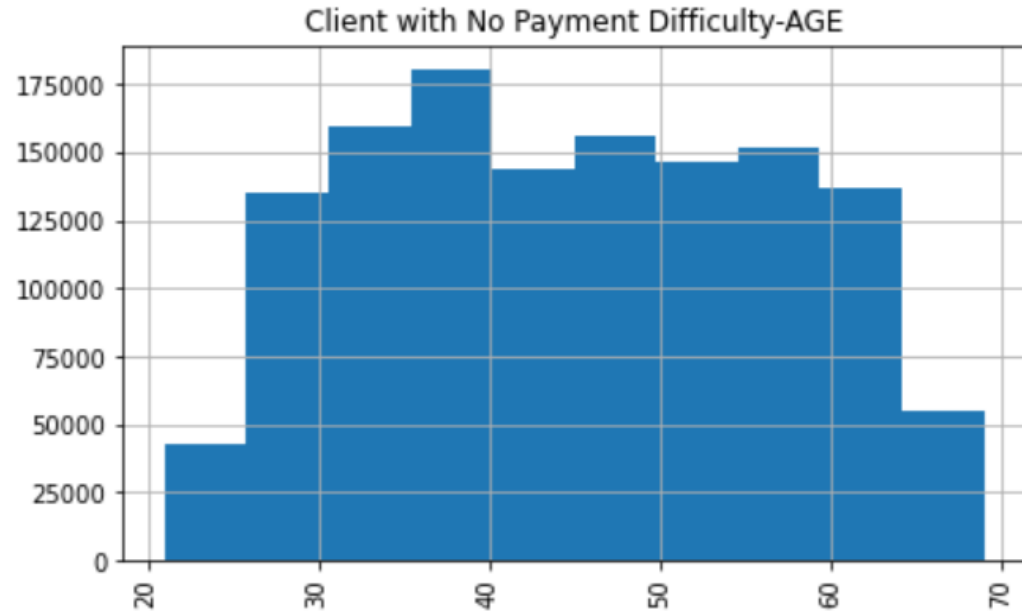


# INFERENCES ON AGE & WORK EXPERIENCE



- Clients in the age group between 55 and 70, who had less years of work experience had difficulty in repaying the loan.

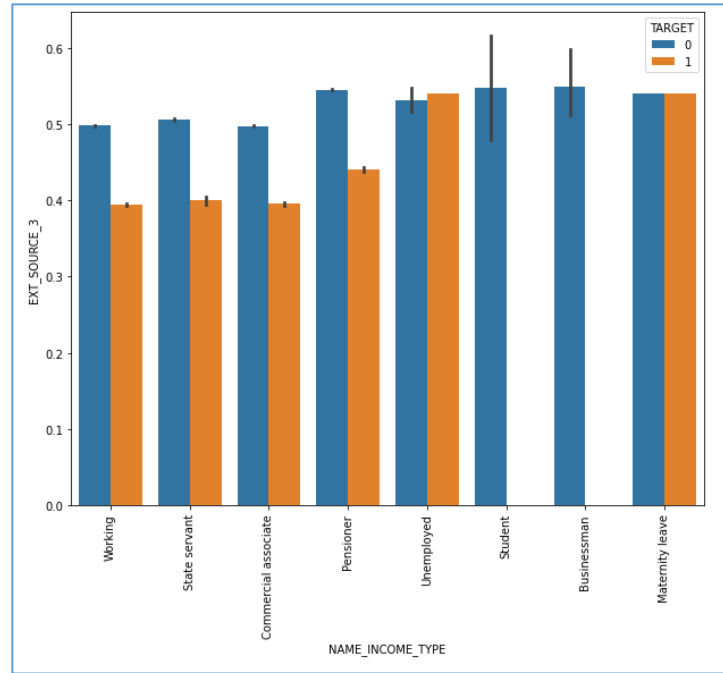
# INFERENCES ON AGE



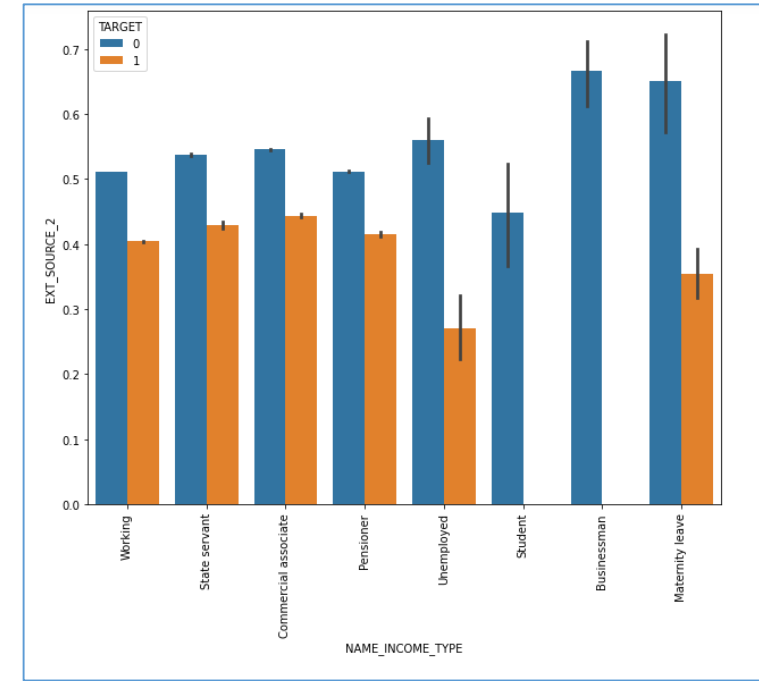
- Majority of the defaulters lie in the age group 30-40.
- As the age increases, we see a majority in the age group (50–70) are likely to repay the loan.

# INFERENCES ON EXTERNAL SOURCES

1

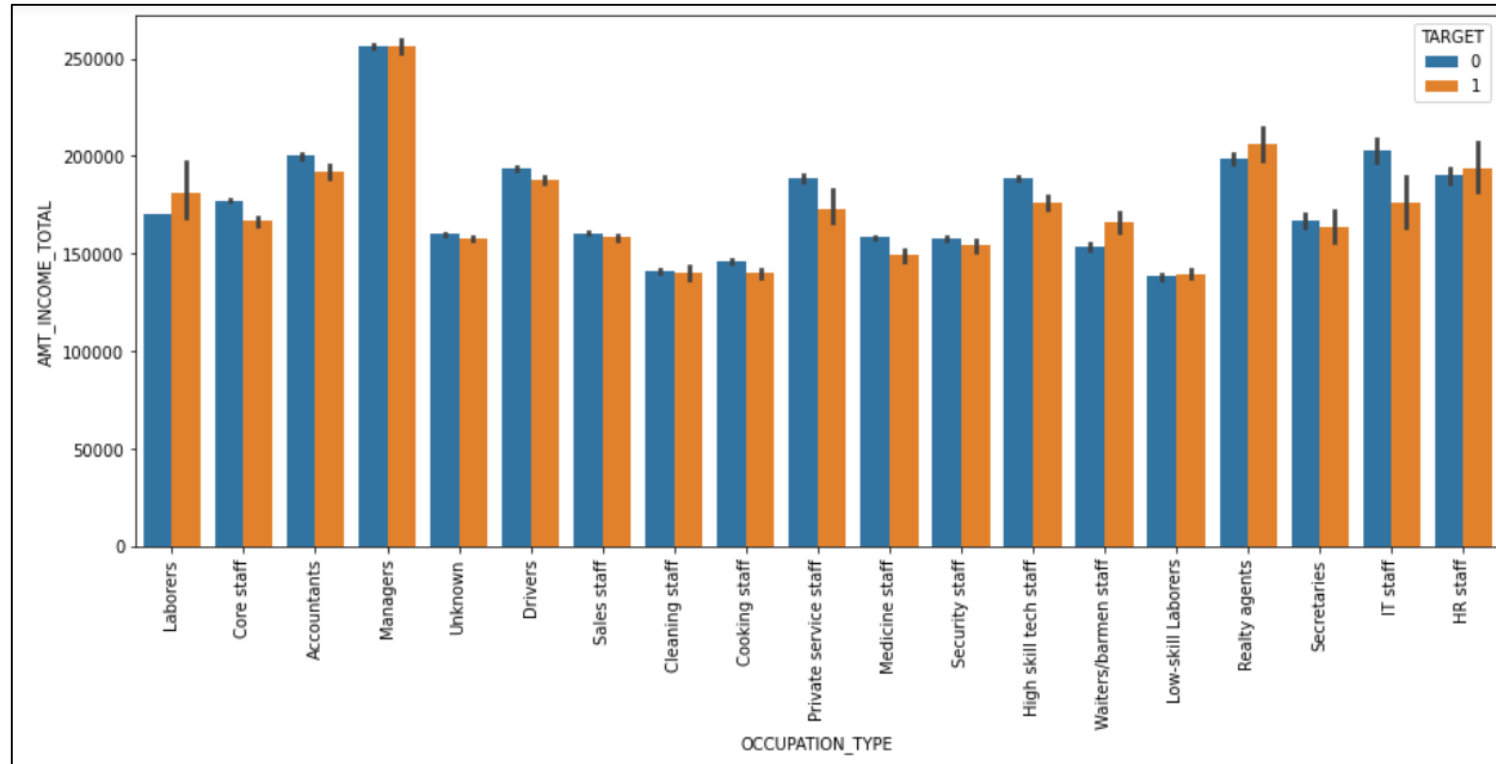


2



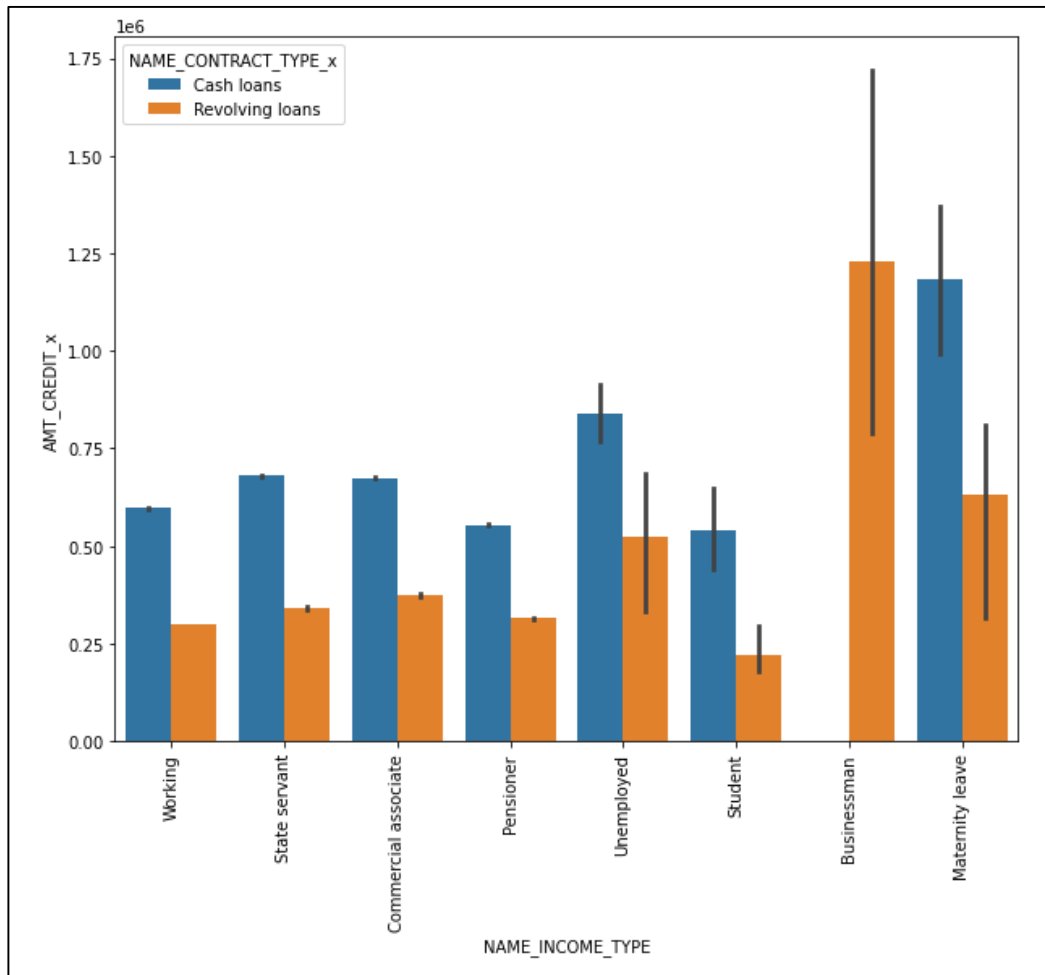
- The score from external sources are reliable to identify the clients who have a tendency to default.
- As shown in plot 1 and 2, businessmen, student and pensioners with a high EXT\_SOURCE\_3 and EXT\_SOURCE\_2 scores are highly unlikely to default.
- Unemployed clients with a high EXT\_SOURCE\_2 score are unlikely to default.

# INFERENCES ON INCOME w.r.t OCCUPATION



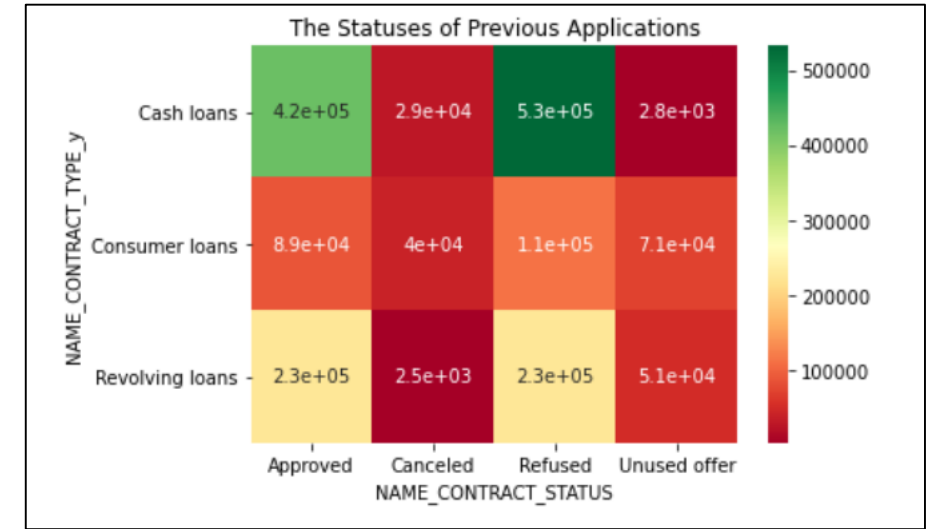
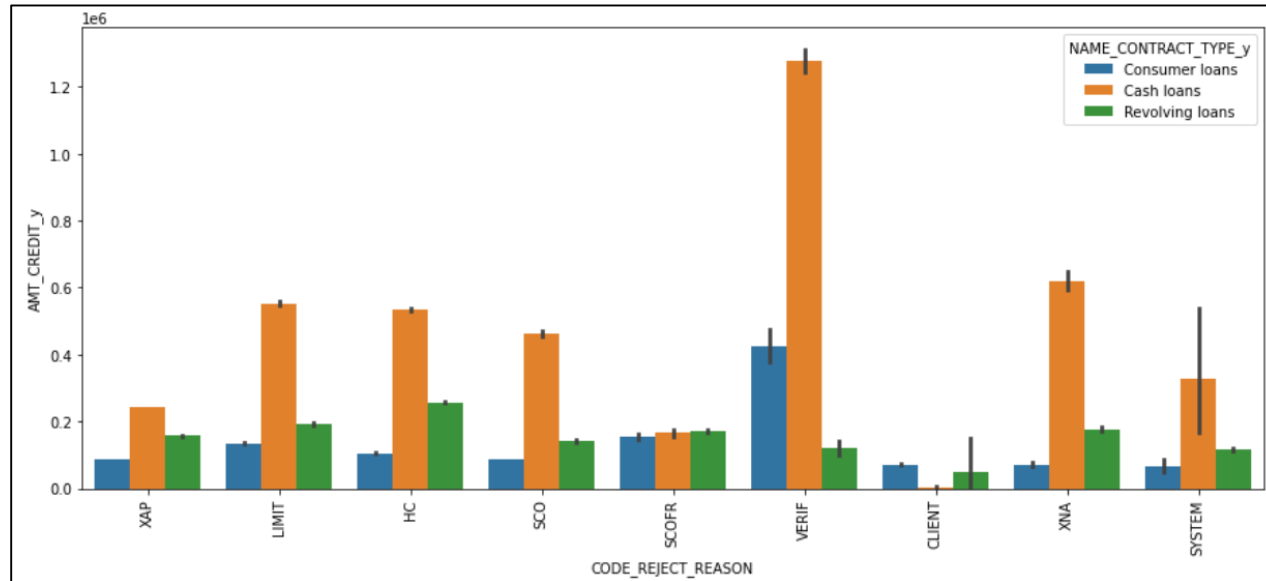
- Here, we would like to focus on the laborer category who are a majority of the applicants.
- Laborers who earn more income compared to other laborers are likely to default.
- Although Managers earn a higher income, there is an equal chance that they will default the loan payment
- Hence, we cannot consider the total income as a parameter to identify defaulters.

# INFERENCES ON INCOME & CREDIT



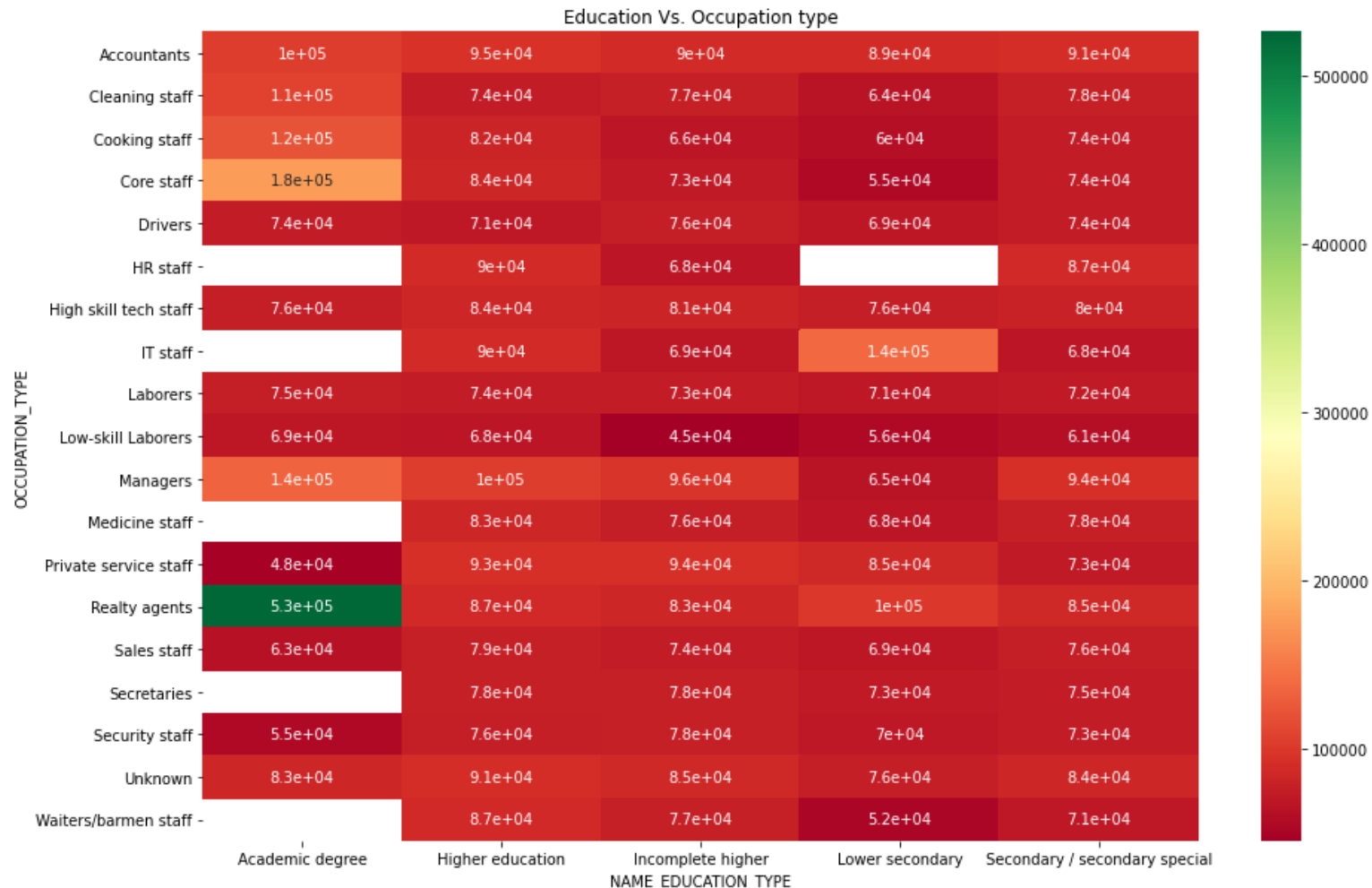
- Businessmen apply for the highest credit revolving loans.
- Clients on maternity leave apply for highest credit cash loan.

# INFERENCES ON STATUS & REJECTION



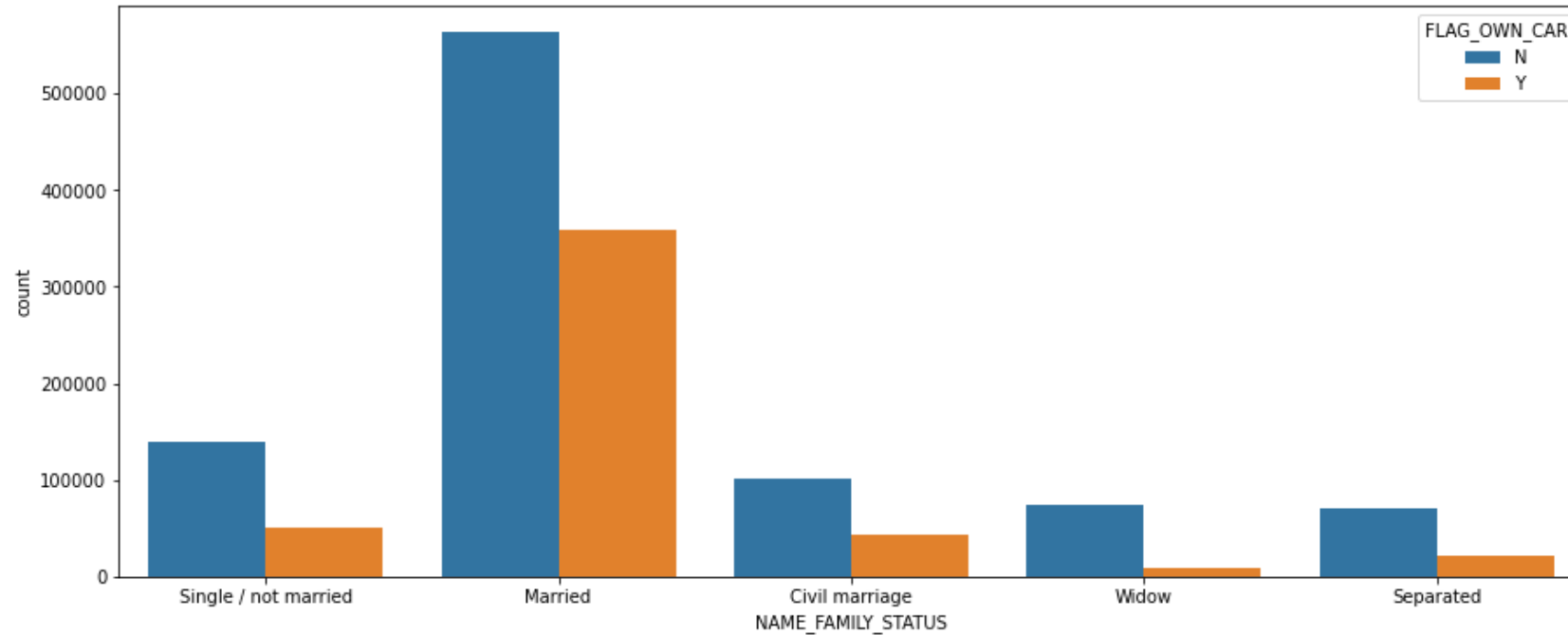
- To understand the trend in previous applications, we plotted a bar graph of reason of rejection against the credit amount.
- A heatmap is plotted to understand the relation between the credit amount and the previous application status.
- The graph clearly indicates that previous cash loan applications with very high credit amount are bound for rejection on the basis of verification.

# INFERENCES ON OCCUPATION & EDUCATION



- The heatmap shows the median credit amount of a client with a particular category of occupation and education type.
- As evident from the heatmap, realty agents with an academic degree have applied for loans with high credit.

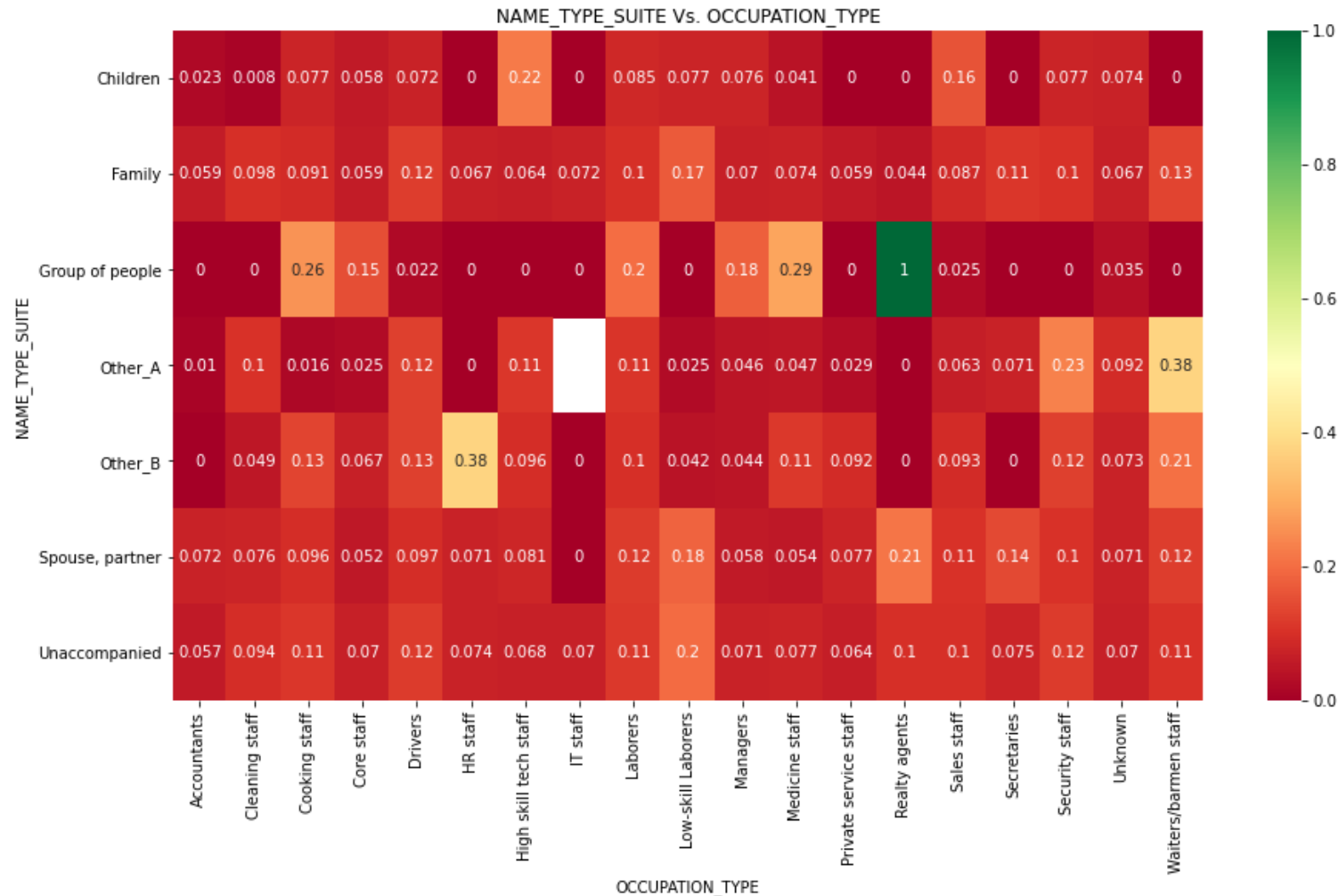
# INFERENCES ON FAMILY STATUS



- We infer from this chart that married applicants who do not own a car are applying for cash loans.
- We can frame a hypothesis that mostly female laborers who are married are applying for cash loans to buy a car.

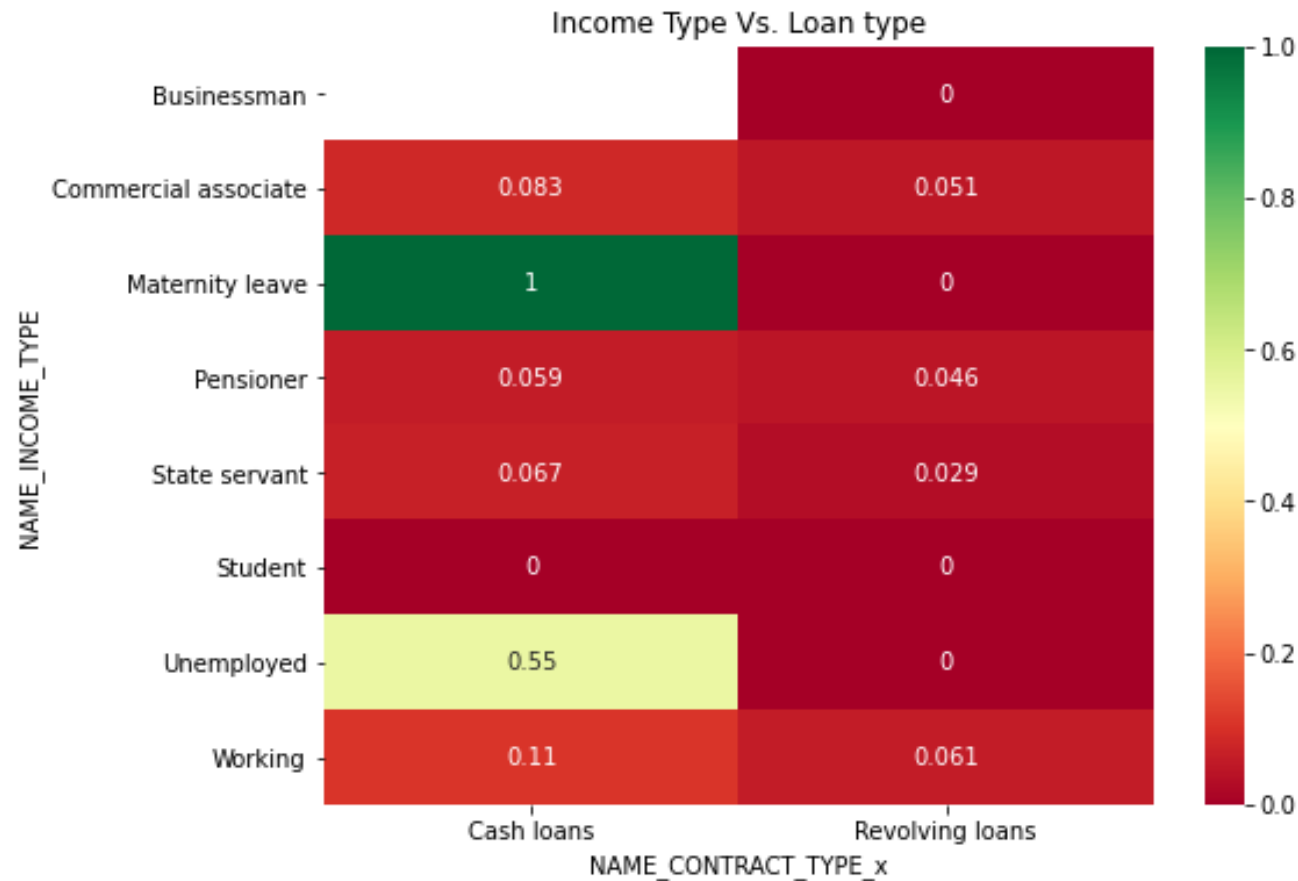


# INFERENCES ON ACCOMPANY & RISK



- We analyzed if the person accompanying the applicant can give us any insight regarding the tendency to default.
- As evident from the heatmap, realty agents accompanied by a group of people will default.

# INFERENCES ON INCOME & LOAN TYPE



- The heatmap displays the average value of whether an applicant from a particular income type is likely to default.
- Applicants who are on maternity leave and have applied for cash loans will default.
- Unemployed applicants who have applied for cash loans have a 0.55 probability of being defaulters.

# CONCLUSION

- Businessmen apply for loans with high credit and are less likely to default. Therefore, their applications can be approved.
- Majority of applications are received from female laborers. Although they are likely to default, they tend to repay the loans in the long term. Hence, their loans can be approved with lower credit amount after verification of their income.
- Applicants in the working category who have a normalized score of 0.5 and above from the various external sources can be approved for cash loans.
- Clients with a high income do not apply for higher credit loans and it is safe to assume they do not have a tendency to default.
- Clients in the age group between 30-40 had no difficulty in repayment of loan.