# CREDIT EDA CASE STUDY

By Richa jha and Nikita Aggarwal

# LIBRARIES USED

NUMPY – **It is the fundamental package for scientific computing**

PANDAS – It is fast, powerful, flexible and easy to use open source data analysis and manipulation tool

SEABORN - It is a library for making statistical graphics in Python.

ITERTOOLS – It gives Functions creating iterators for efficient looping

# BEFORE CLEANSING

- In Application data
  - Total Rows : 307511
  - Total Columns : 122

- In Previous Data
  - Total Rows: 1670214
  - Total Column : 37

# AFTER CLEANSING

- In Application data
  - Total Rows : 307511
  - Total Columns : 35

- In Previous Data
  - Total Rows: 69635
  - Total Column : 22

# OCCUPATION TYPE

- Least people applying for loan where from IT , HR and Reality Agents
- Maximum people applying for loan has not known category and hence considered as unknown
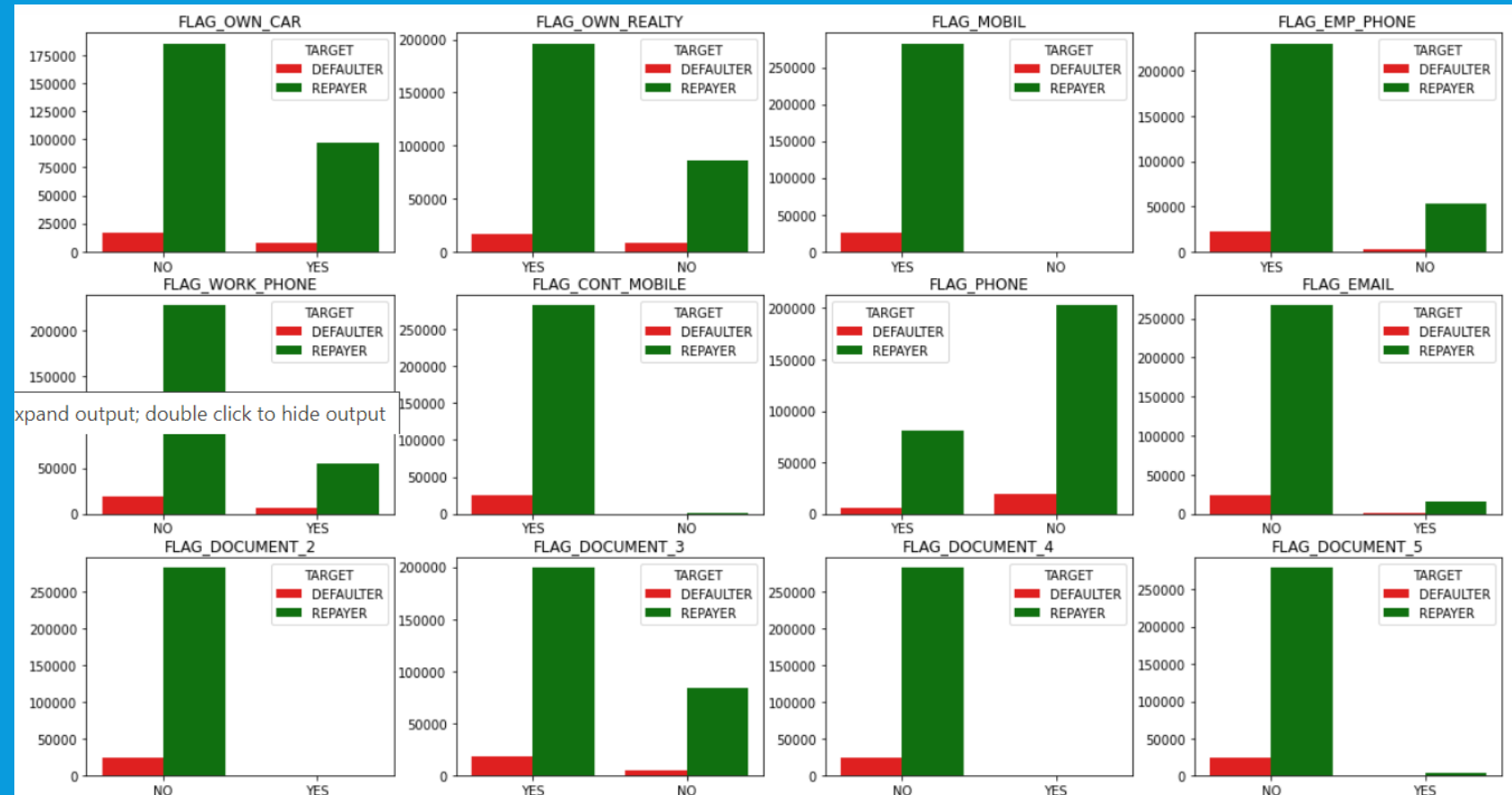


Occupation type graph
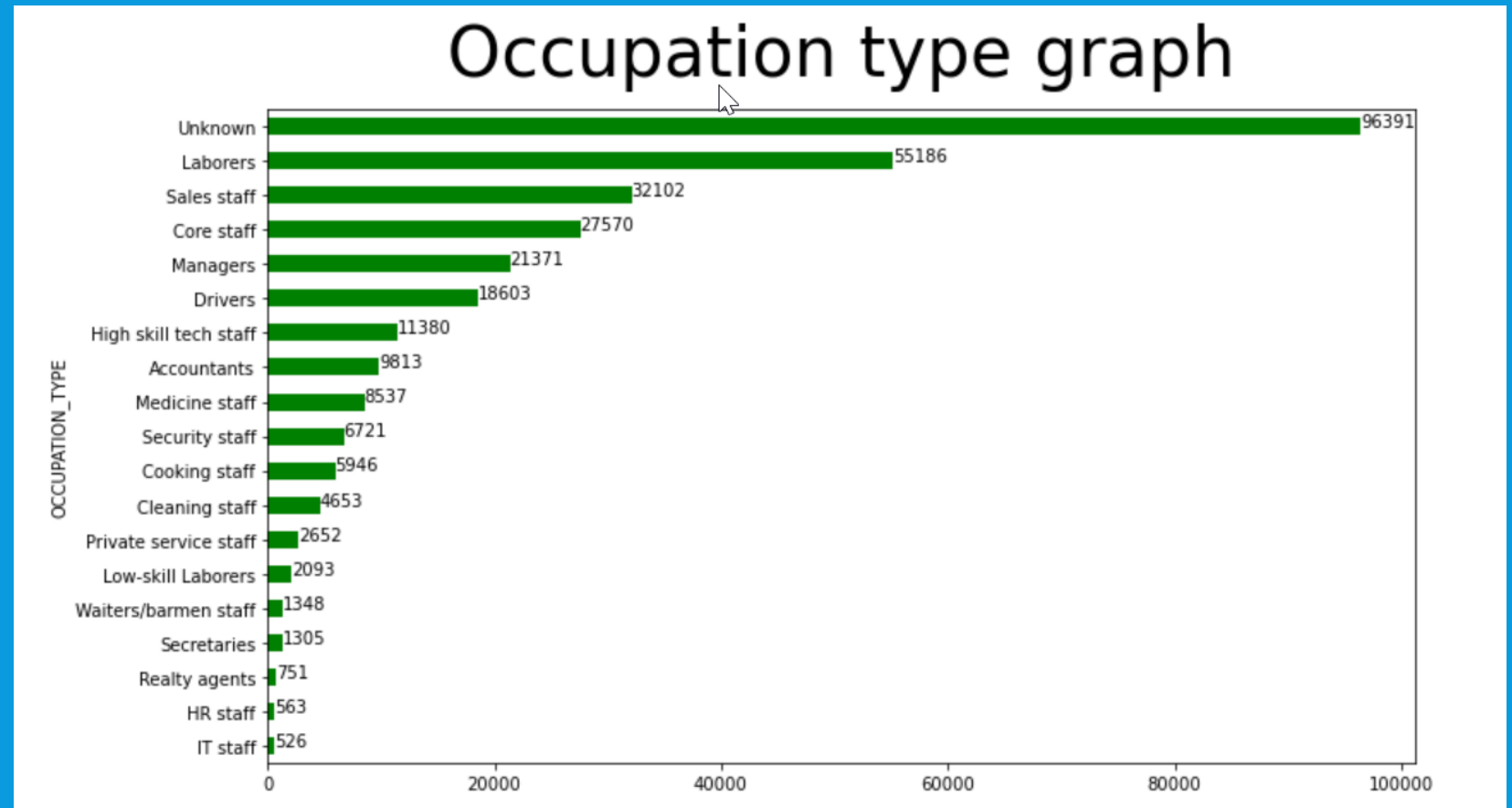
# NORMALIZED COLUMN

- No correlation we can find between EXT_SOURCE_2 and EXT_SOURCE_3
- Both columns are delete



Correlation between External data source 2 and 3
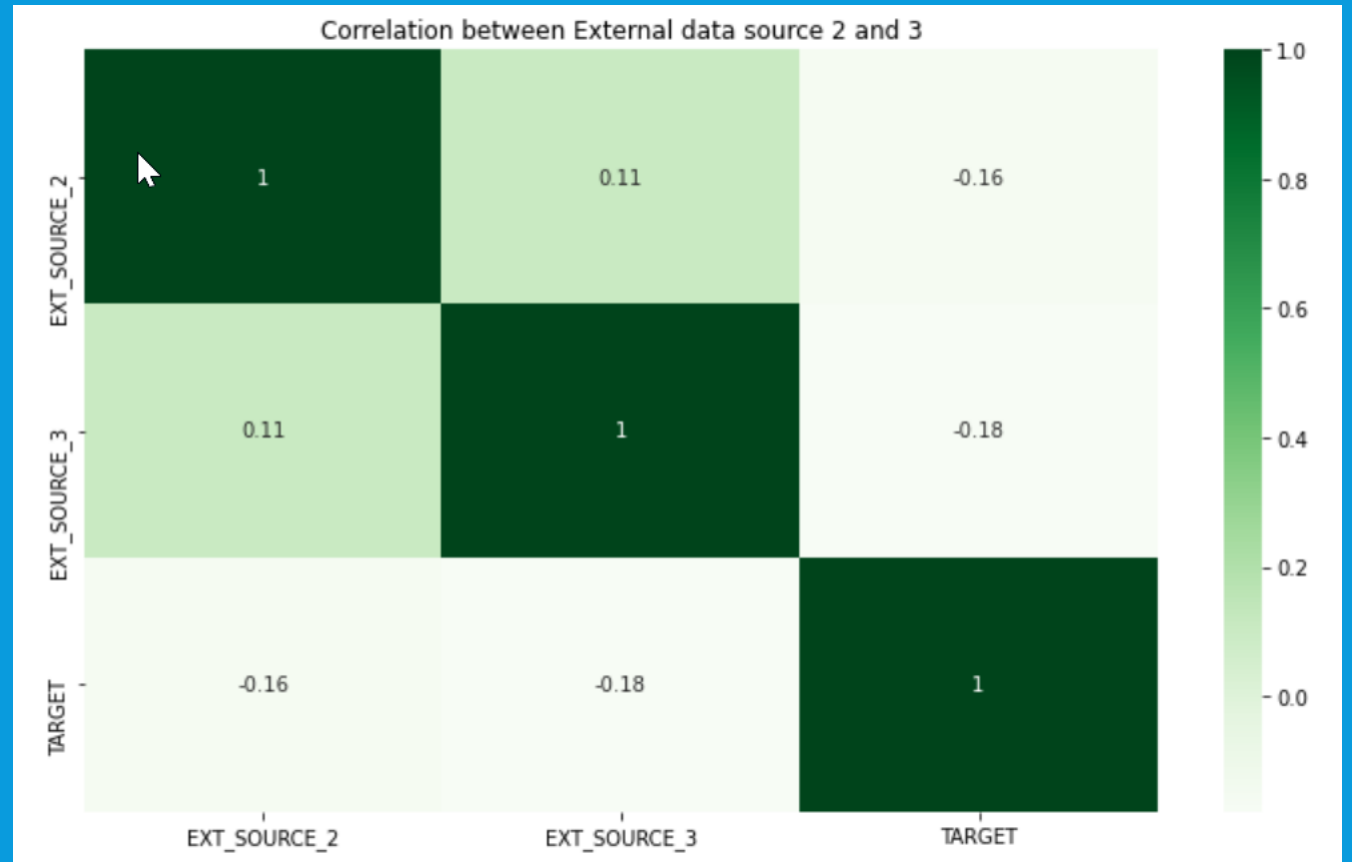
# FLAG GRAPH

'FLAG_OWN_REALTY','FLAG_MOBIL','FLAG_EMP_PHONE','FLAG_CONT_MOBILE', and 'FLAG_DOCUMENT_3' have more repayor.

# INCOME RANGE FOR EACH GENDER

1. Women are getting more Income than men
2. People with income range form 1L to 2L are highest.
3. On an average we can observe that female earns more than male
4. 5 Lakh and above there are very less people and almost and female are equal



Income Range by Gender in each bin

# CREDIT AMOUNT APPLIED BY GENDER

- Women has more credit than men in all the ranges
- The highest create amount requested is between 2L-3L
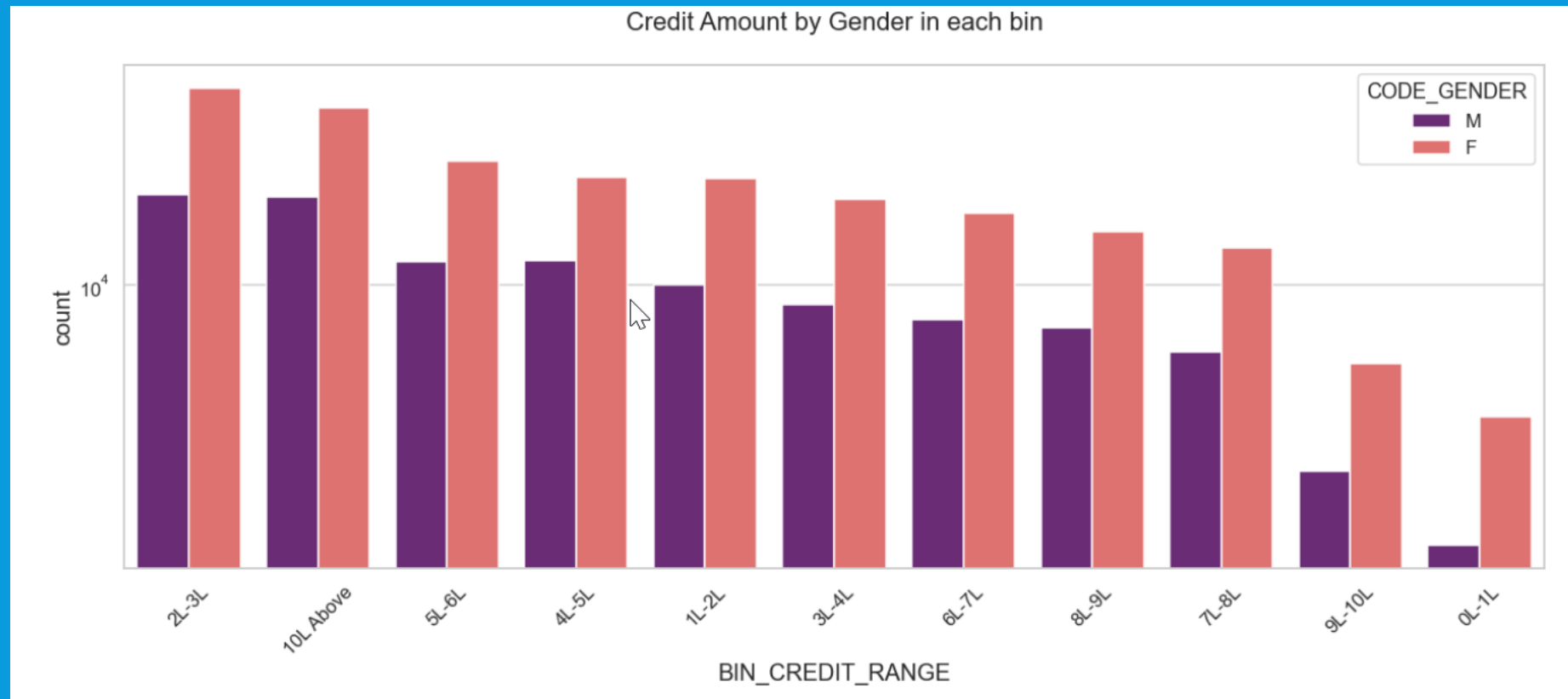
# LOAN APPLIED FOR PURCHASE OF GOODS

- Women asked for more goods loan than men
- In all the category women is on top



Price of goods for which Loan is requested by Gender in each bin
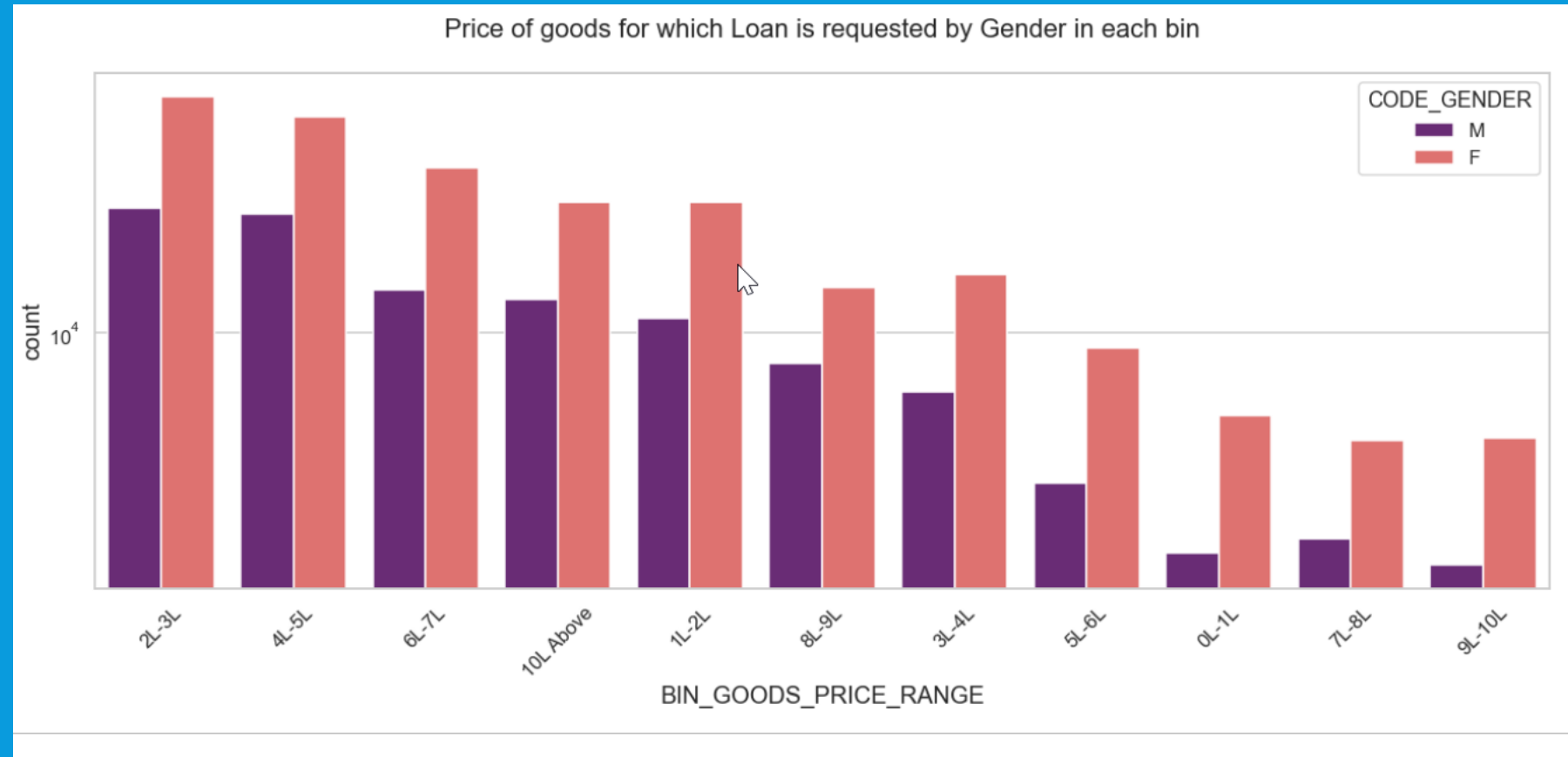
# TYPE OF INCOME GENDER WISE

- Working Professional and Commercial associate lead the ladder
- Females lead even in this category
- Person on Maternity leave or Businessman are less
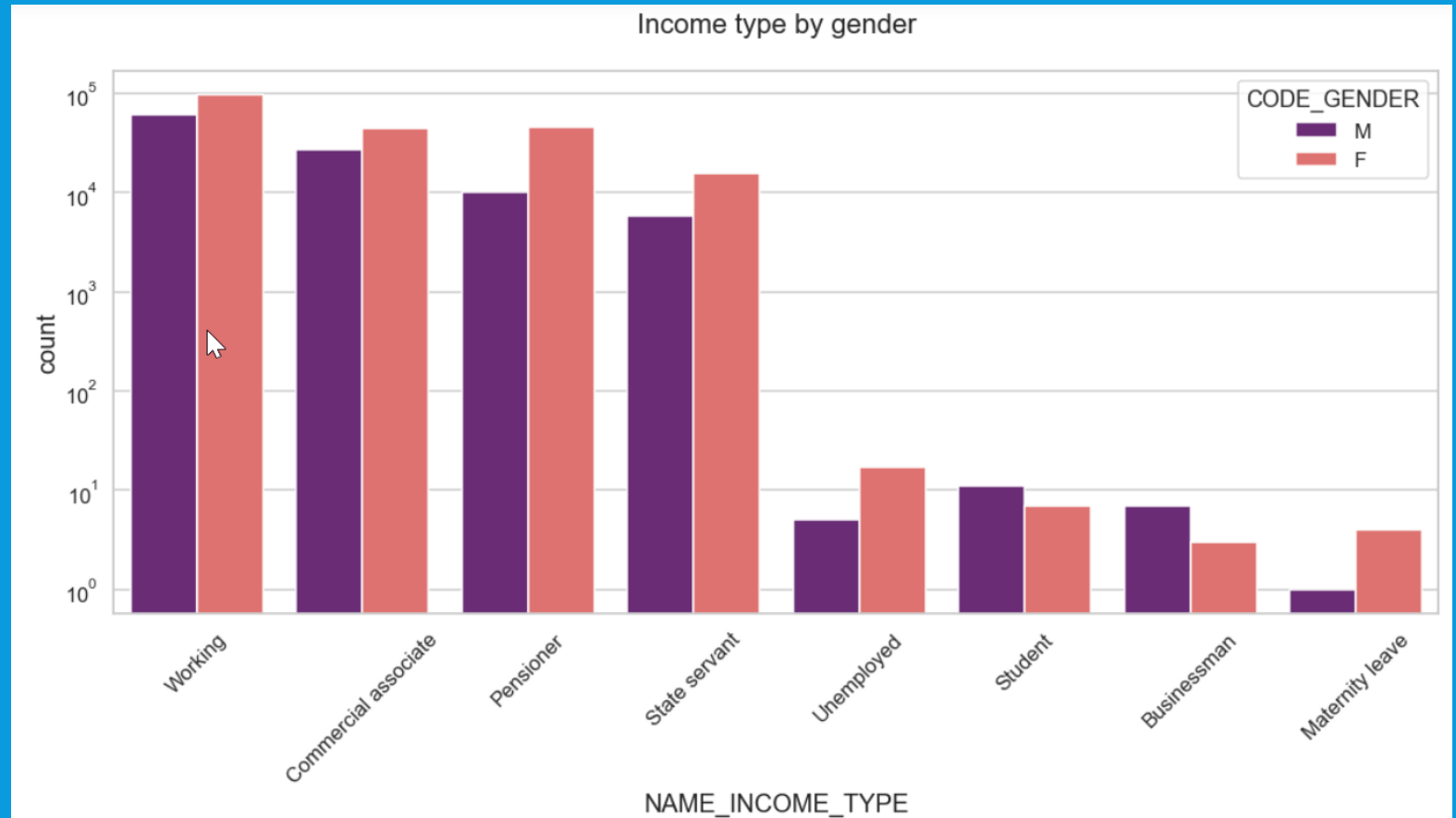
# OUTLIERS

- AMT_ANNUITY, AMT_CREDIT, AMT_GOODS_PRICE,CNT_CHIL DREN have outliers.
- AMT_INCOME_TOTAL has huge number of outliers , by which we can say that few loan application has high income than others.
- DAYS_BIRTH has no outliers which means the data available is reliable.
- DAYS_EMPLOYED has outlier values around 350000(days) which is around 958 years. Which clearly indicate that is wrong entry
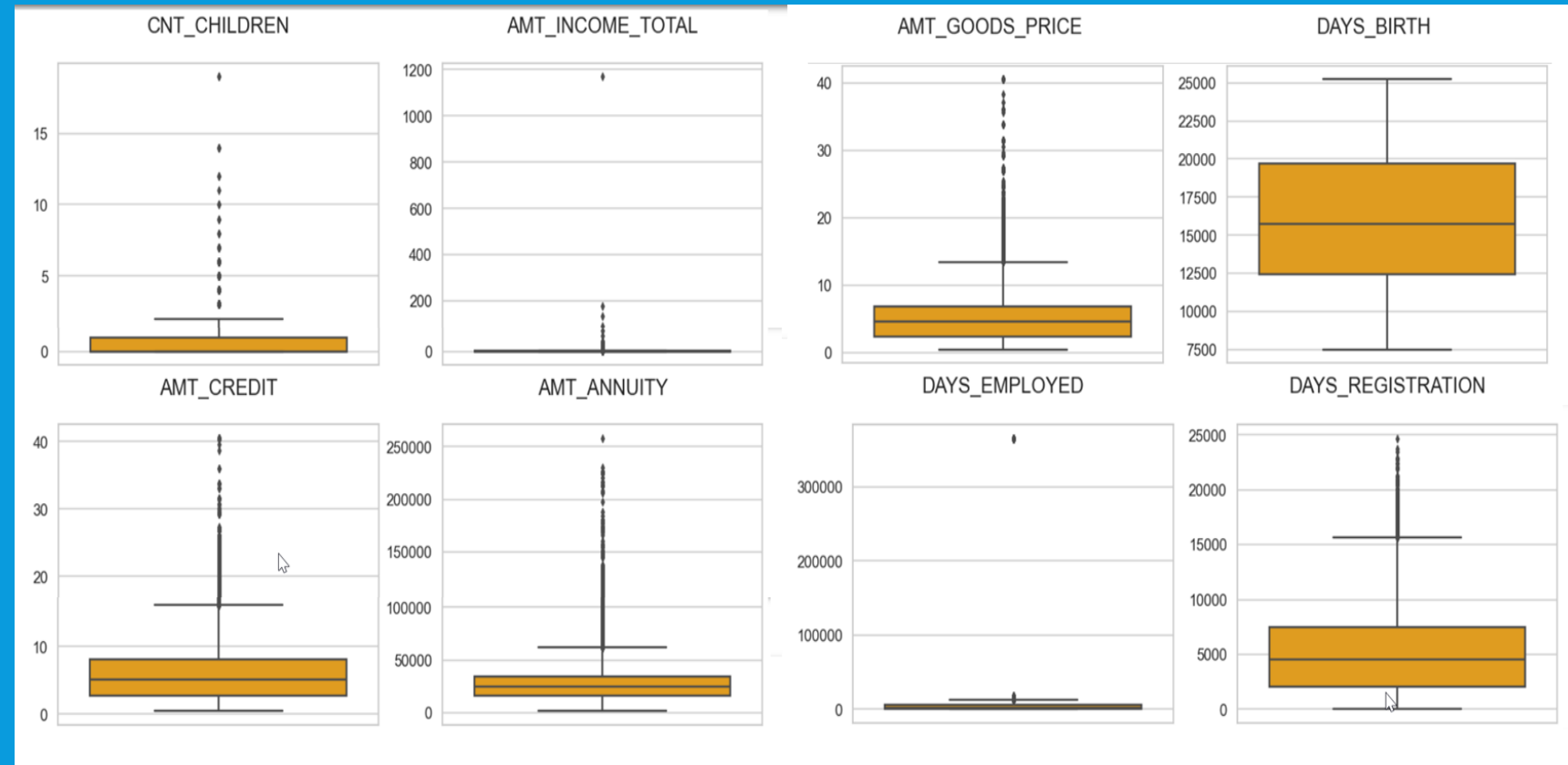
# DEFAULTER CORRELATION

- Credit amount is inversely proportional to the date of birth
- Credit amount is inversely proportional to the number of children client have
- Income amount is inversely proportional to the number of children client have
- less children client have in densely populated area.
- Credit amount is higher to densely populated area.
- The income is also higher in densely populated area.



Correlation for Defaulters
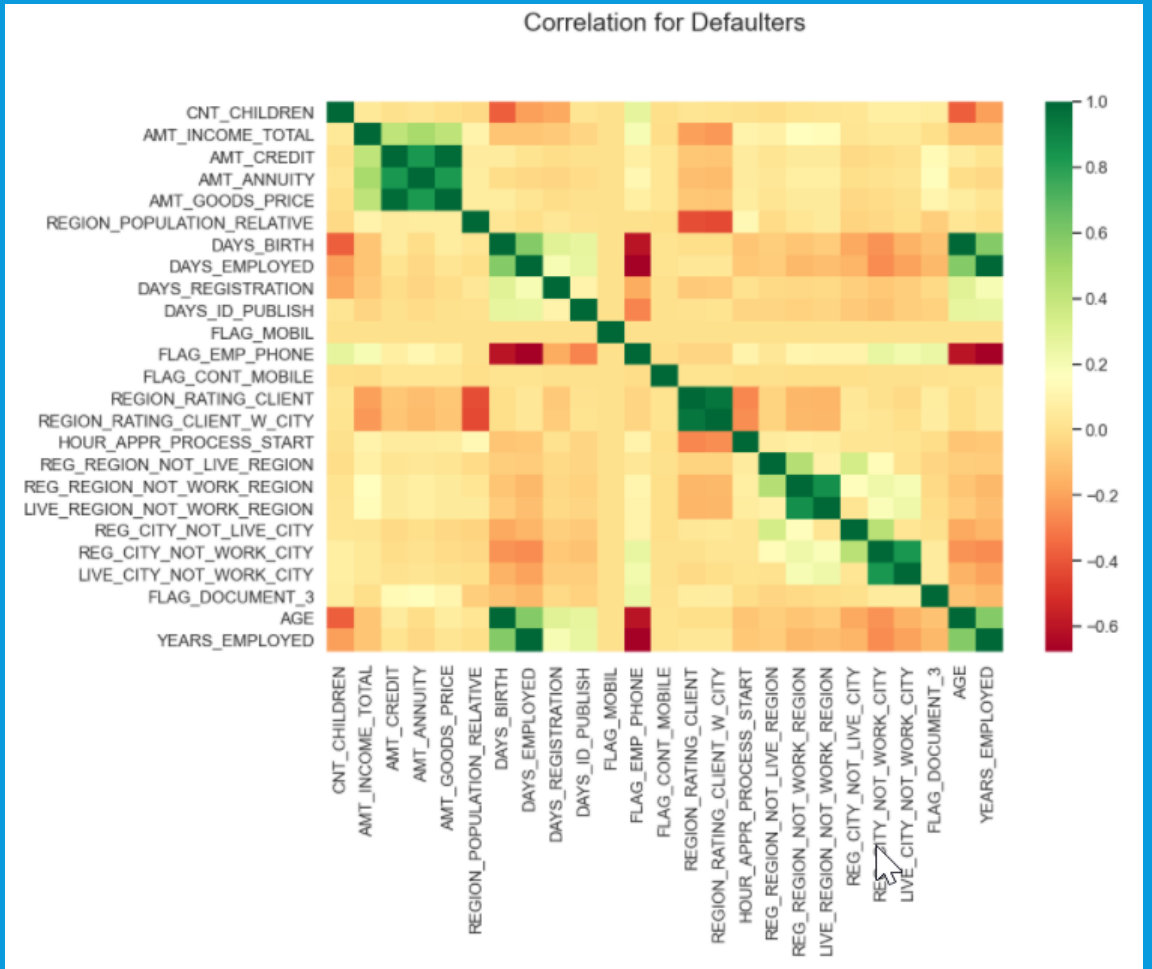
# NON DEFAULTER CORRELATION

- The client's permanent address does not match contact address are having less children
- The client's permanent address does not match work address are having less children



Correlation for Non Defaulters
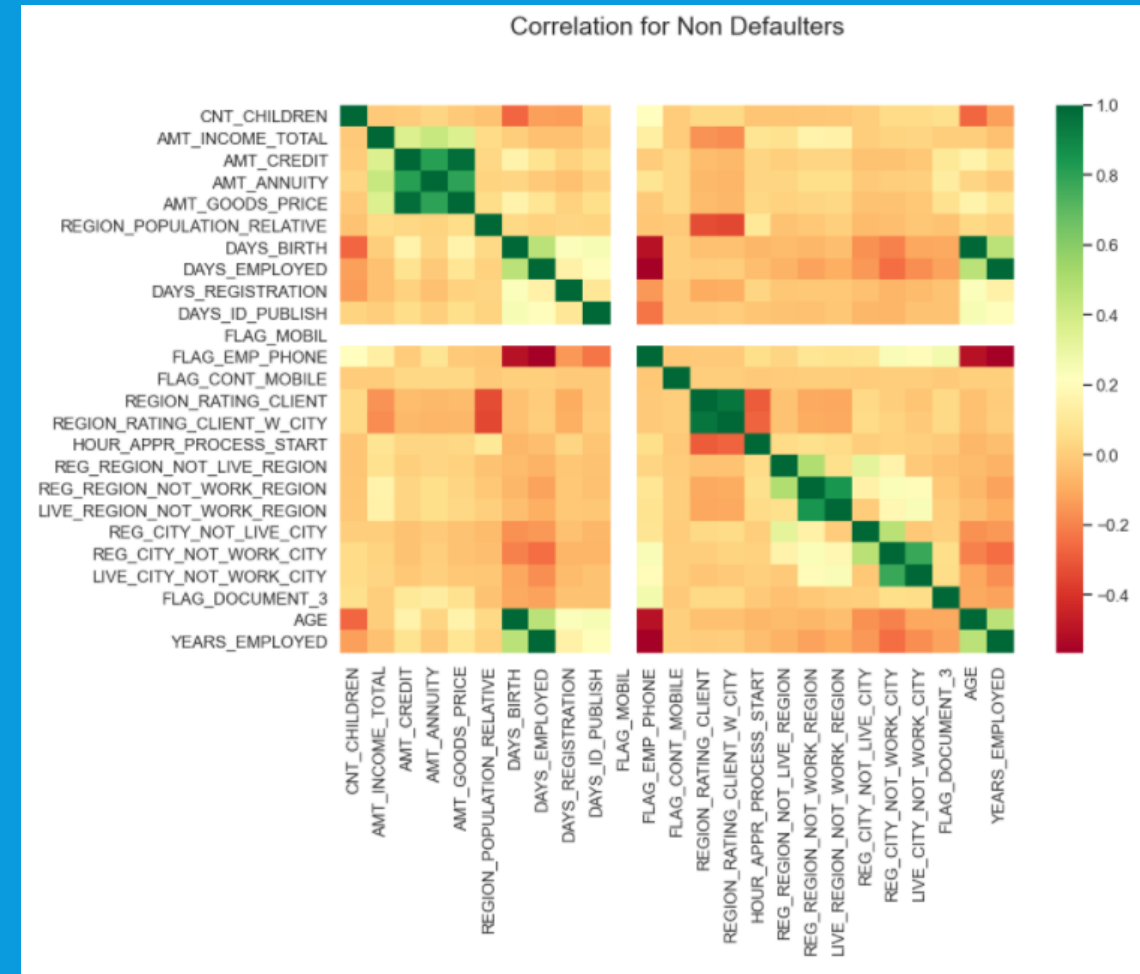
# PREVIOUS DATASET

# UNIVARIABLE ANALYSIS

- Most rejection of loans came from purpose 'repairs'.
- For education purposes we have equal number of approves and rejection
- Paying other loans and buying a new car is having significant higher rejection than approvals.
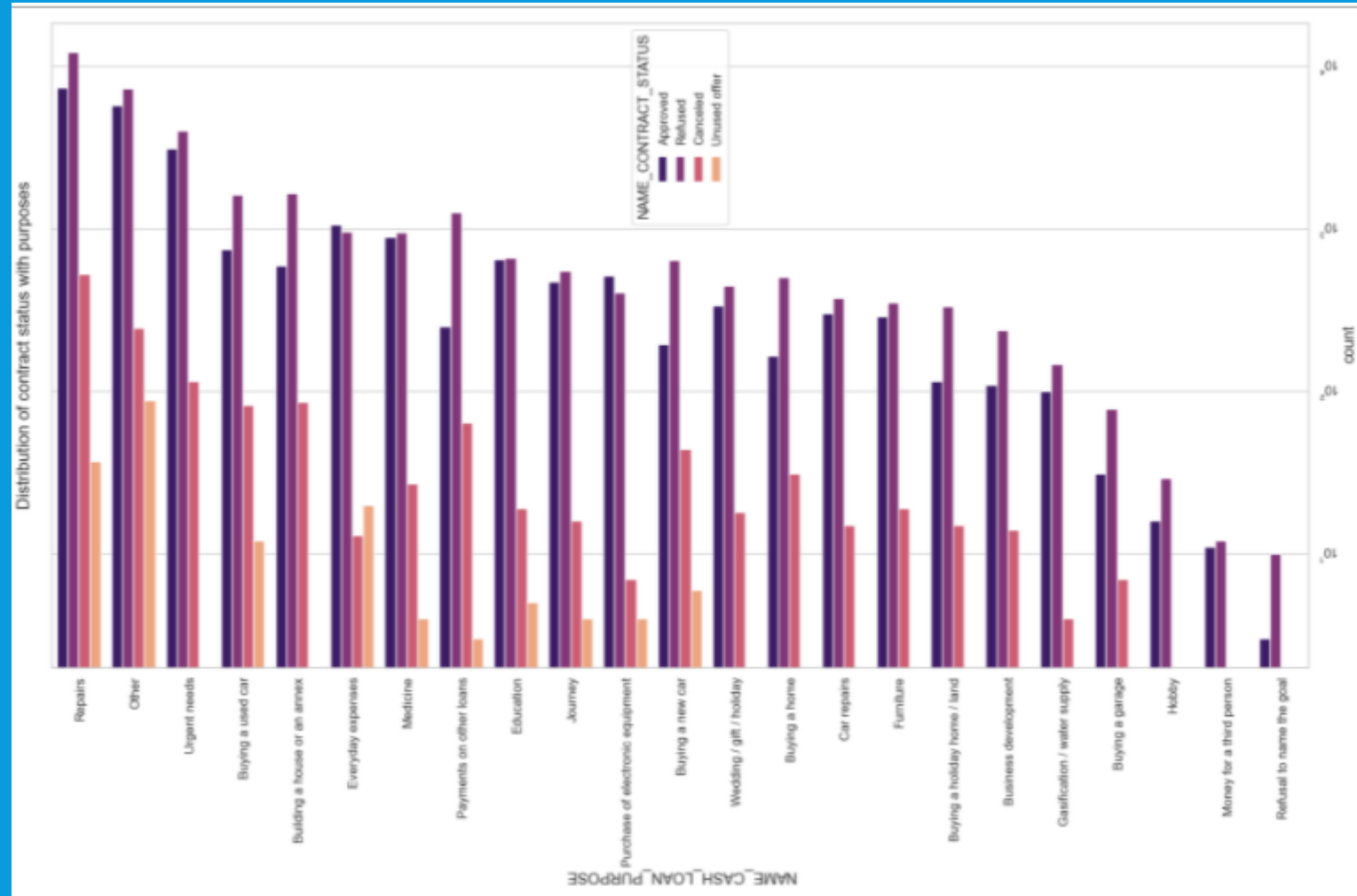
# PURPOSE WITH TARGET

- Loan purposes with 'Repairs' are facing more difficulties in payment on time.
- There are few places where loan payment is significant higher than facing difficulties. They are 'Buying a garage', 'Business development', 'Buying land', 'Buying a new car' and 'Education' Hence we can focus on these purposes for which the client is having for minimal payment difficulties.
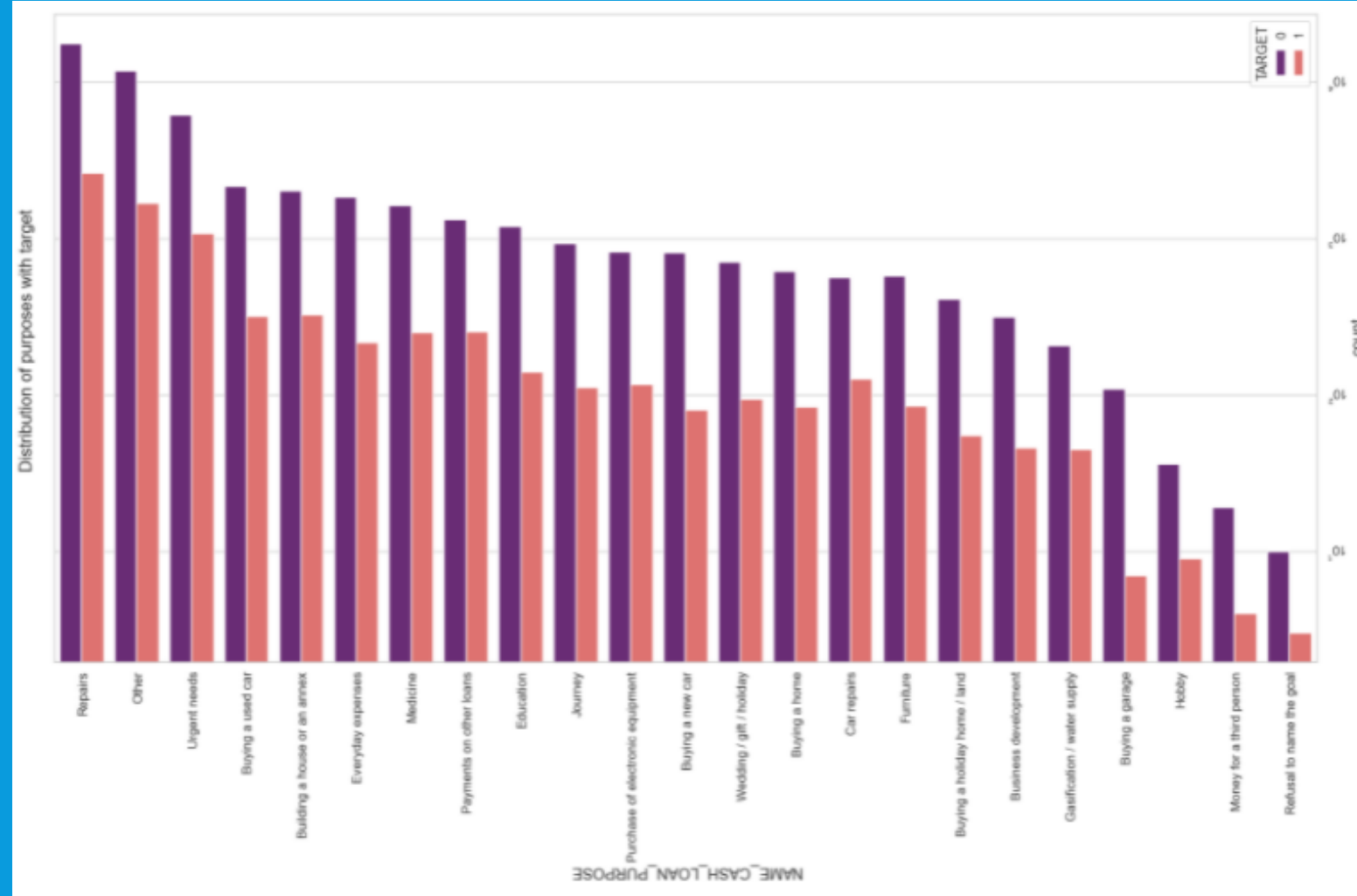
# COMPARISON OF PREVIOUS CREDIT V/S LOAN PURPOSE WITH AMOUNT CREDIT

- The credit amount of Loan purposes like 'Buying a home', 'Buying a land', 'Buying a new car' and 'Building a house' is higher.
- Income type of state servants have a significant amount of credit applied
- Money for third person or a Hobby is having less credits applied for.



Prev Credit amount vs Loan Purpose
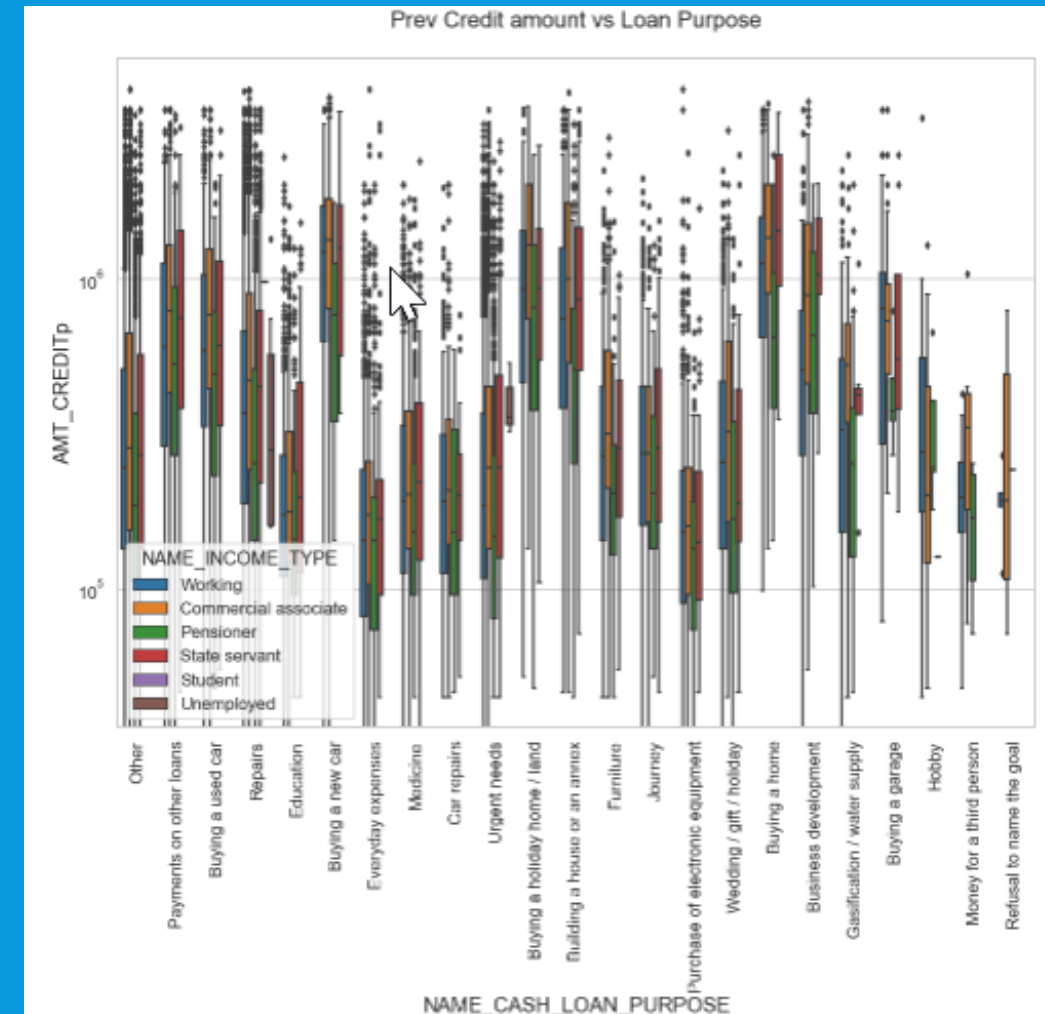
# PREVIOUS CREDIT AMOUNT V/S TYPE OF HOME THEY OWN

- Here for Housing type, office apartment is having higher credit of target o and co-op apartment is having higher credit of target .
- So, we can conclude that bank should avoid giving loans to the housing type of co-op apartment as they are having difficulties in payment.
- Bank can focus mostly on housing type with parents or House\apartment or municipal apartment for successful payments.


Prev Credit amount vs Housing type
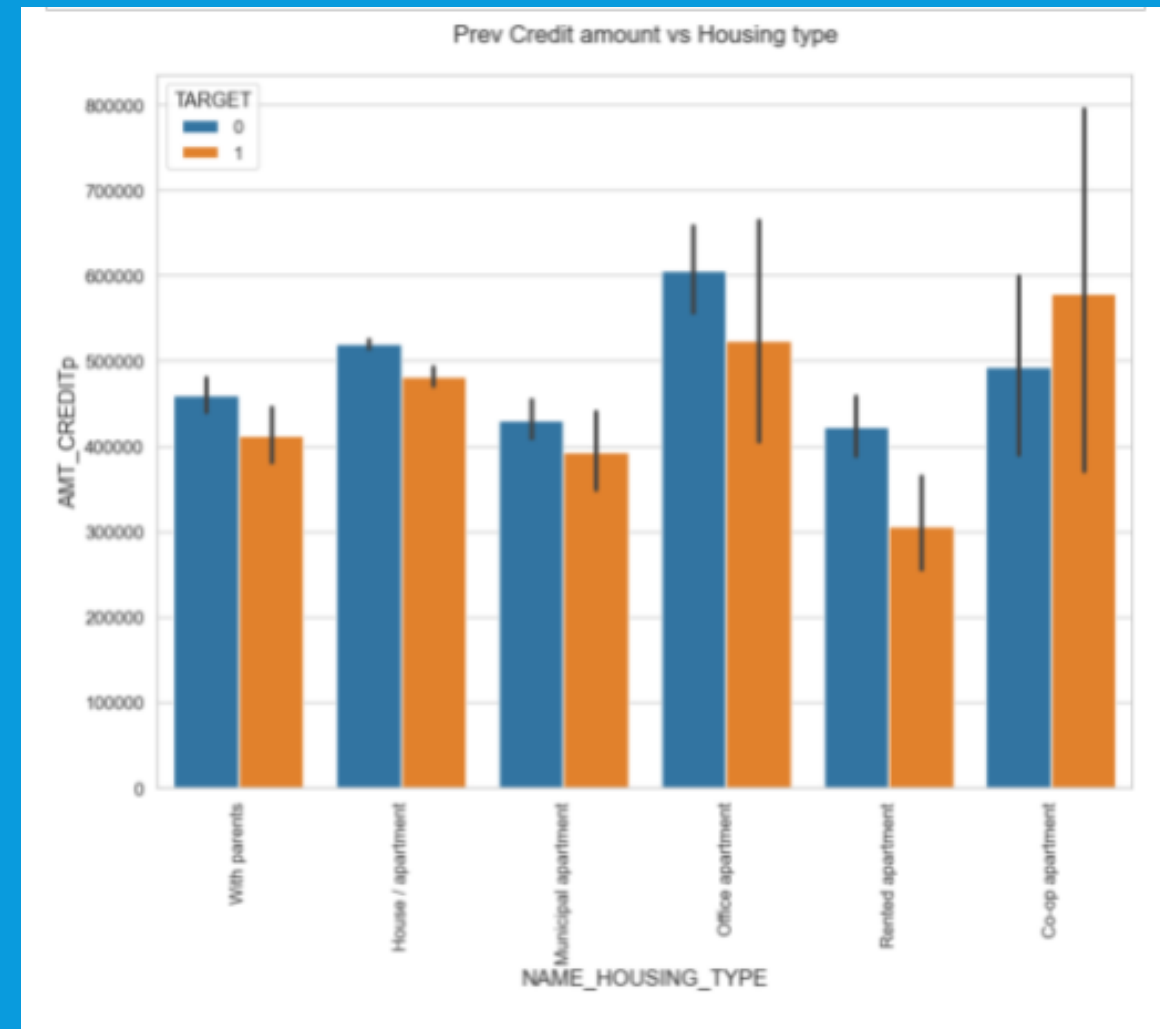
# CONCLUSION

1. Banks should focus more on contract type 'Student' ,'pensioner' and 'Businessman' with housing 'type other than 'Co-op apartment' for successful payments.

2. Banks should focus less on income type 'Working' as they are having most number of unsuccessful payments.

3. Also with loan purpose 'Repair' is having higher number of unsuccessful payments on time.

4. Get as much as clients from housing type 'With parents' as they are having least number of unsuccessful payments.¶