



Loan Default Risk Analysis EDA

SHASHANK

Approach



Goal of EDA

This EDA aims at dealing with two types of risk-

- Reduce the risk of giving loan to defaulter and thus saving bank money.
- Reduce the loss to the bank by not refusing the loan to the Non defaulter.

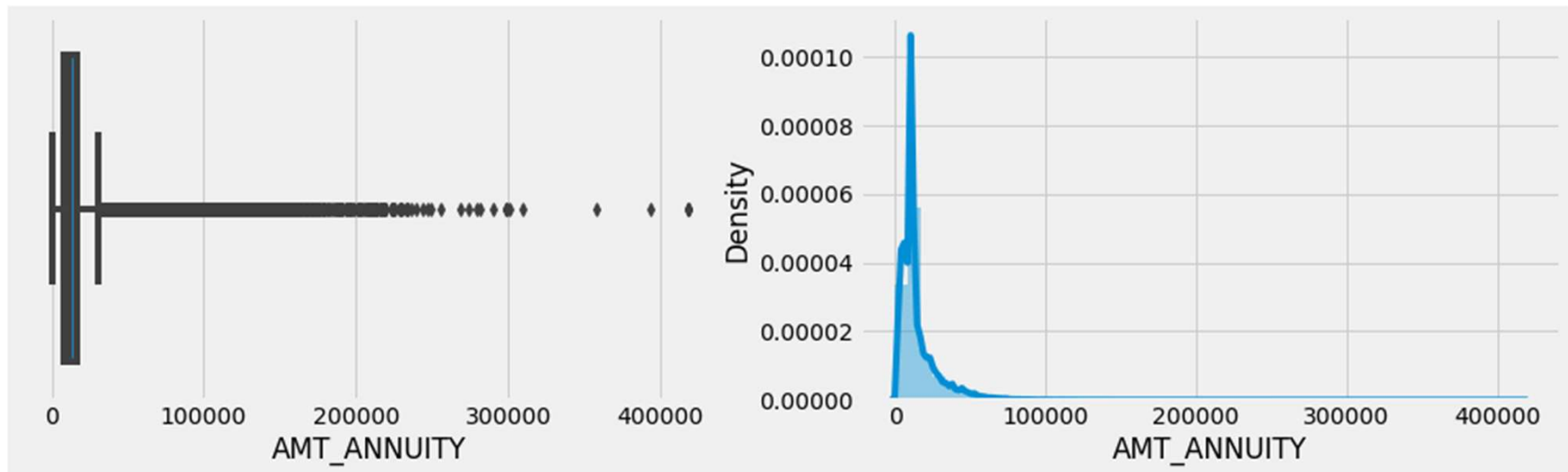
💡 Dropping/Imputing Approach

- The columns with more than 40% NaN values have been dropped as its major data is missing.
- Dropped the rows with NaN value less than 1% in respective columns..
- And the other NaN values in the columns are imputed with the median in case of Categorical variable and Continuous Variable are imputed with their mode.

Note

I have not changed the dtype of categorical columns having values in the form of 0 or 1 to object type as in numerical form taking the mean would directly give the response which would not be possible after conversion of dtype.

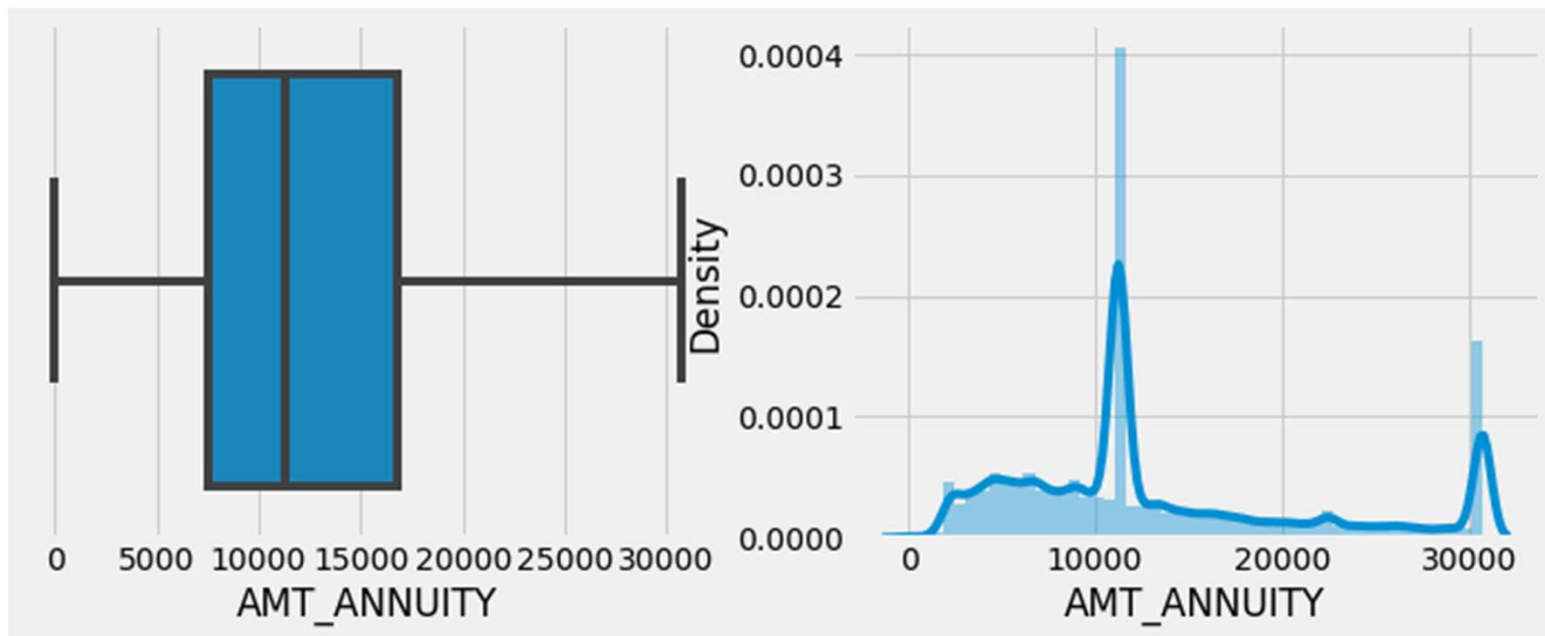
💡 Outlier presence!



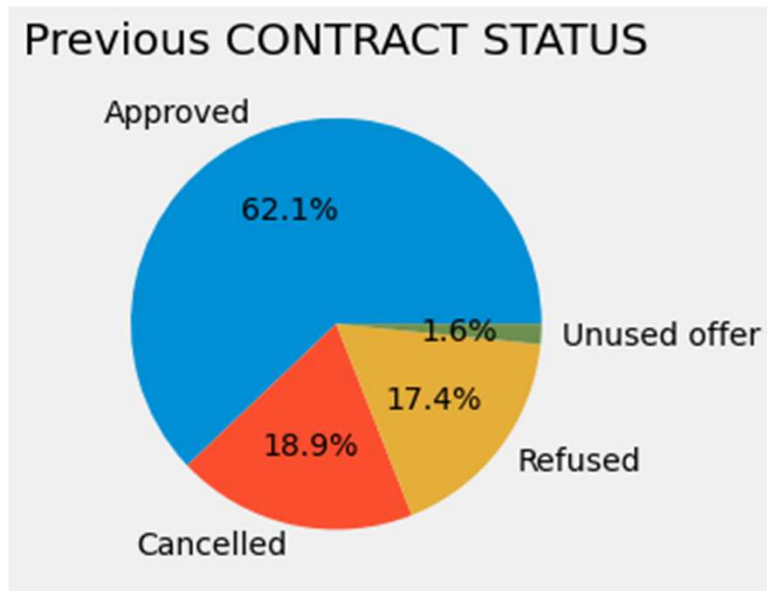
Approach to deal with outlier

I have imputed the Outlier with their respective nearest whisker value i.e., replacing the lower boundary outliers with the lower whisker value and upper boundary outliers with the upper whisker value.

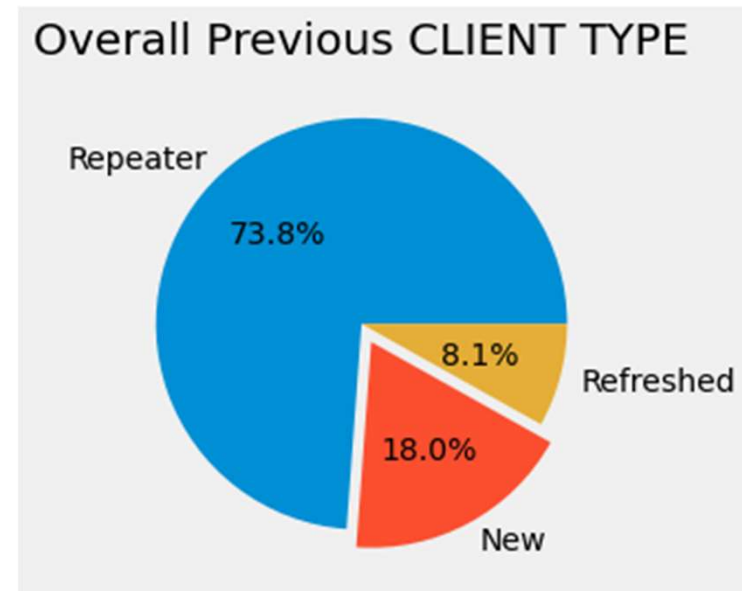
💡 Outlier after treatment



💡 Status of previous application



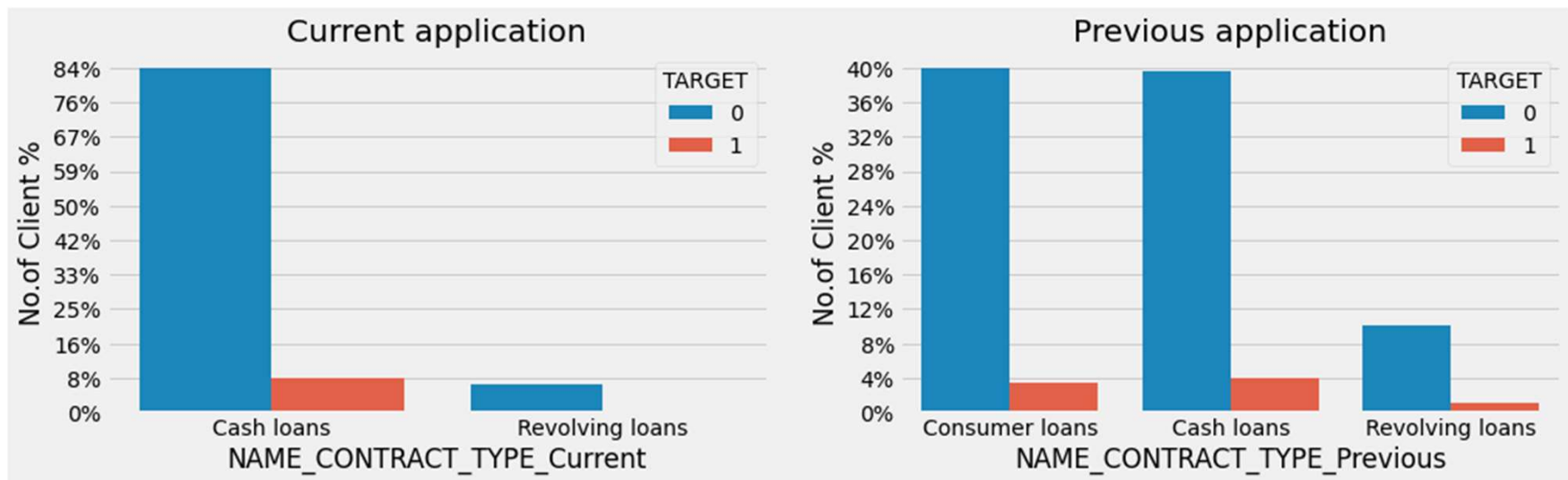
💡 Previous Clients Type



💡 Which loan type is in demand?

0= Non Defaulter

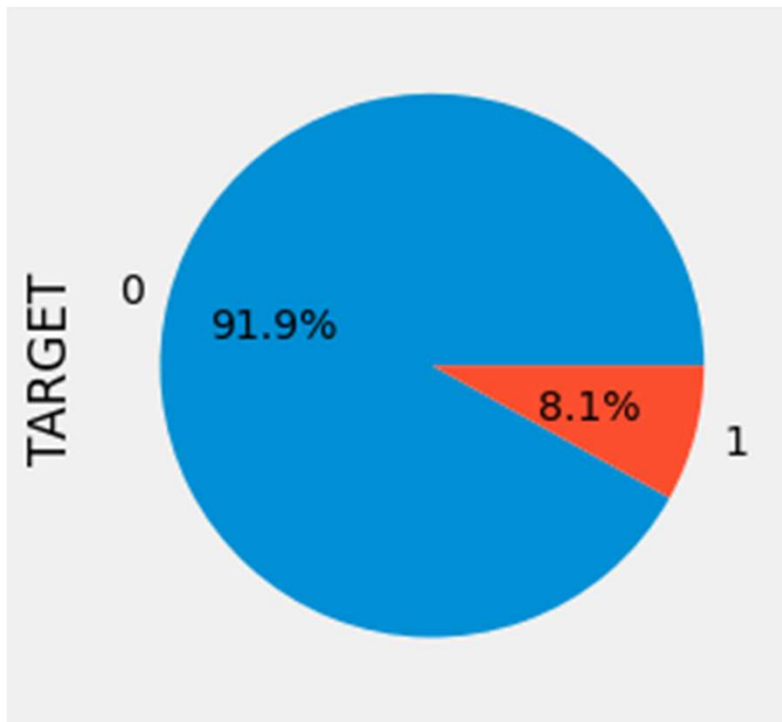
1= Defaulter



Conclusion

- The demand for cash loan has increased abruptly from 44% to 92% compare to the previous data set while there is no demand for consumer loan now

💡 Data Imbalance in Target Column



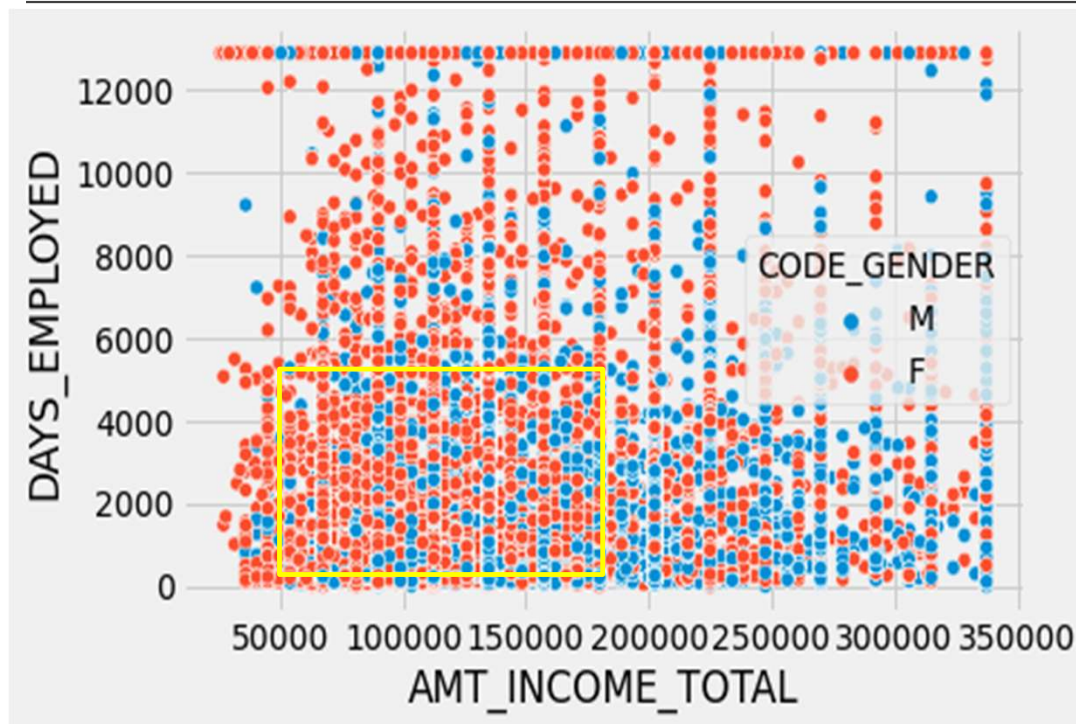
Good news for Bank

Non defaulters are very high in number ie., most of the clients pay back the loan timely.

0= Non Defaulter

1= Defaulter

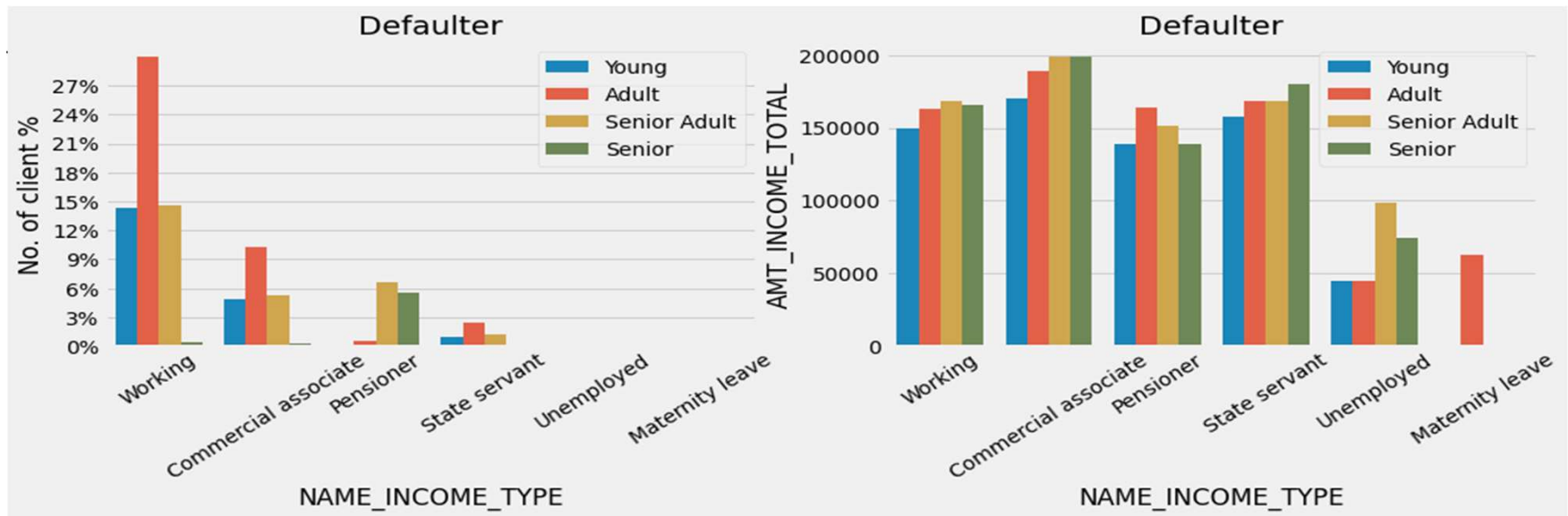
💡 For how long & how much defaulters have been earning ?



Conclusion

- Most of the defaulters income total is less than 2.5 lakh and most of them are working for about 10 years.
- Most of the female defaulter has income less than 2 lakh.

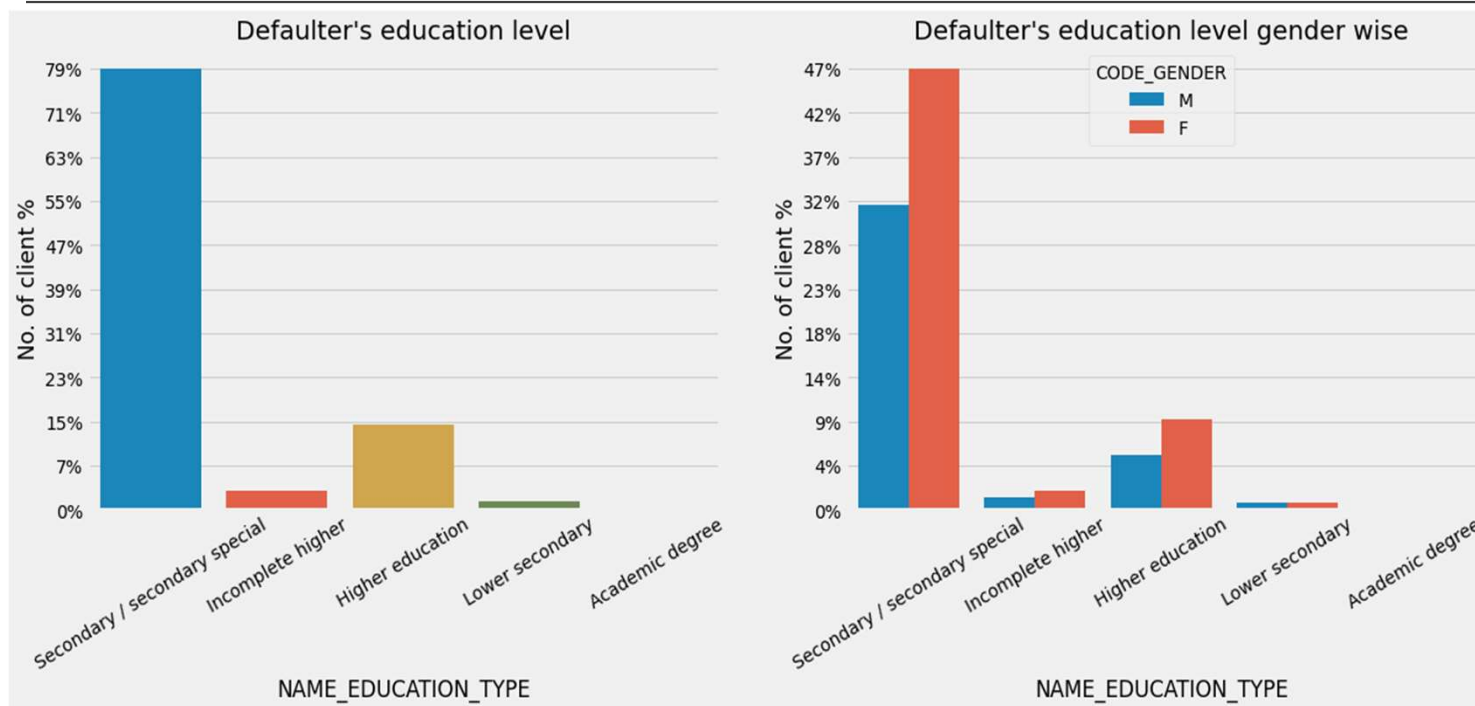
💡 Which Income type based on age seems to default more?



Conclusion

- Adults seem to have more default case and in that too Working adults are most defaulter .
- Average income of Commercial associate are comparatively greater than others.
- Among defaulters senior are having higher average income.

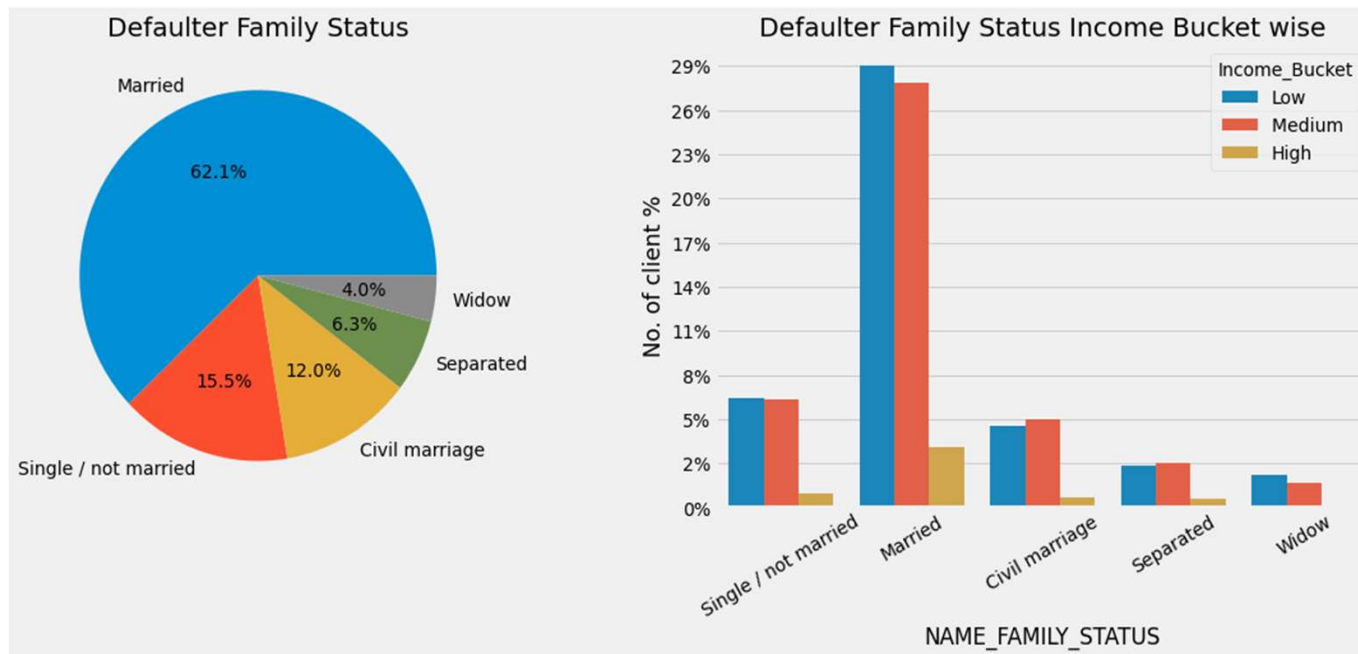
💡 What is the education level of most of the defaulter?



Conclusion

- Defaulter's count decreases with increase in their education level.
- Even from Secondary level 47 % are female.

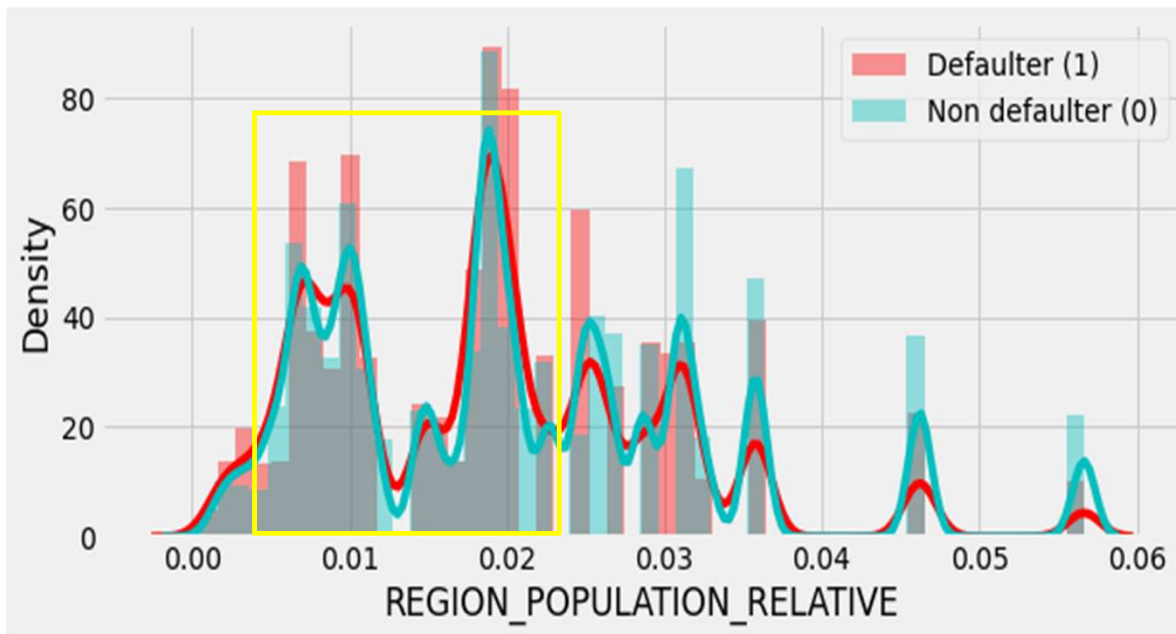
💡 What is the family status of most of the defaulter?



Conclusion

- Married clients seem to do more default and most of them have either low or medium income.

💡 Which region seems to be having more default case?



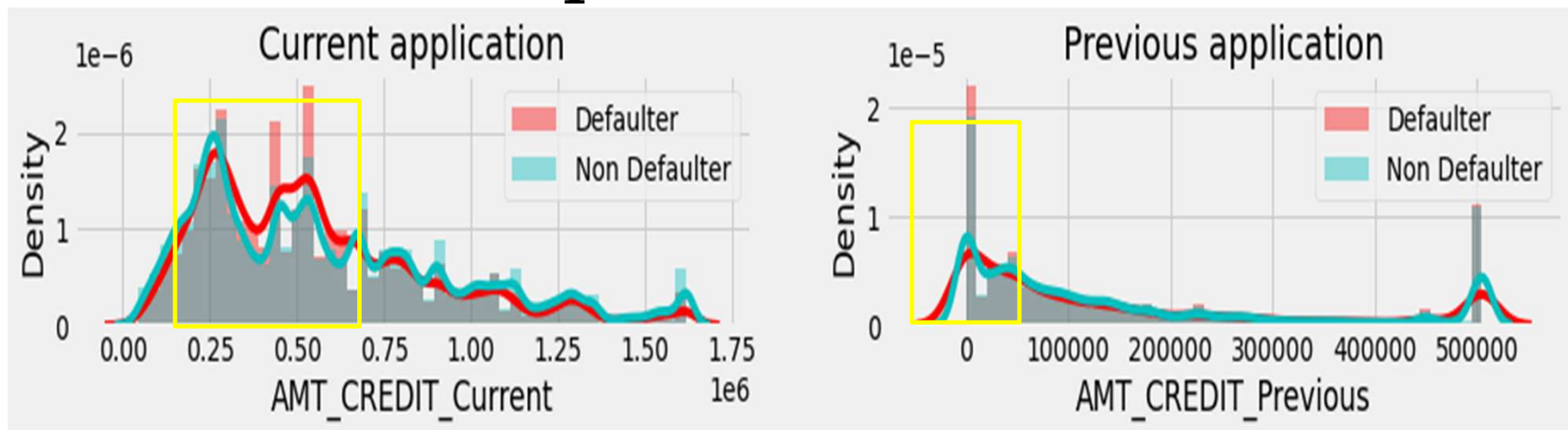
Assumption

- Cities have High population density
- Villages have low population density

Conclusion

- Though the pattern for both defaulter and non defaulter are same but point here is ,maximum loan is taken by area with low population i.e., by village people as cities' have high population density.

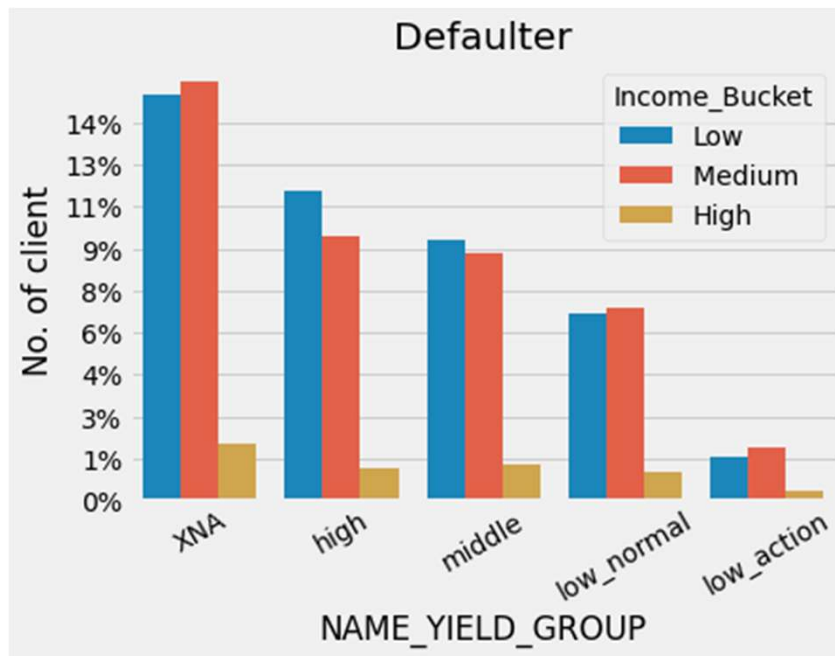
💡 What is the spread of credit amount given by bank to it's client in compare to the previous credit?



Conclusion

- The distribution of current credit amount has increased greatly.
- Earlier maximum credit given was around 5 lakh but this time it is around 17.5 lakh.
- Earlier defaulters' credit amount were less than 1 lakh but this time most of the defaulters' lies in the range of 0-7.5 lakh

💡 What is the effect of the interest rate on the defaulters?



Assumption

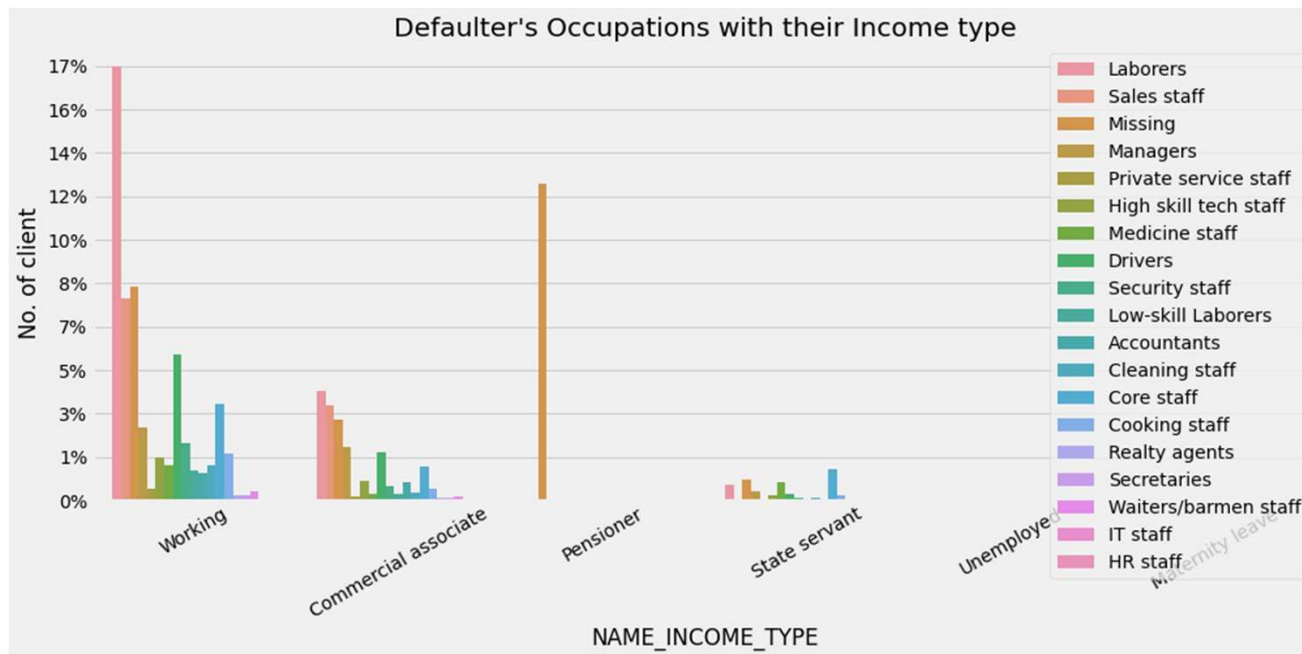
I have not imputed the XNA in the above plot as it's meaning is not clearly mentioned in the data dictionary, also it is huge in number so imputing it without having its proper information will not bring justice to my analysis.

Conclusion

- The loan with greater interest seems to default more so we can draw a conclusion that greater the interest rate higher is the risk



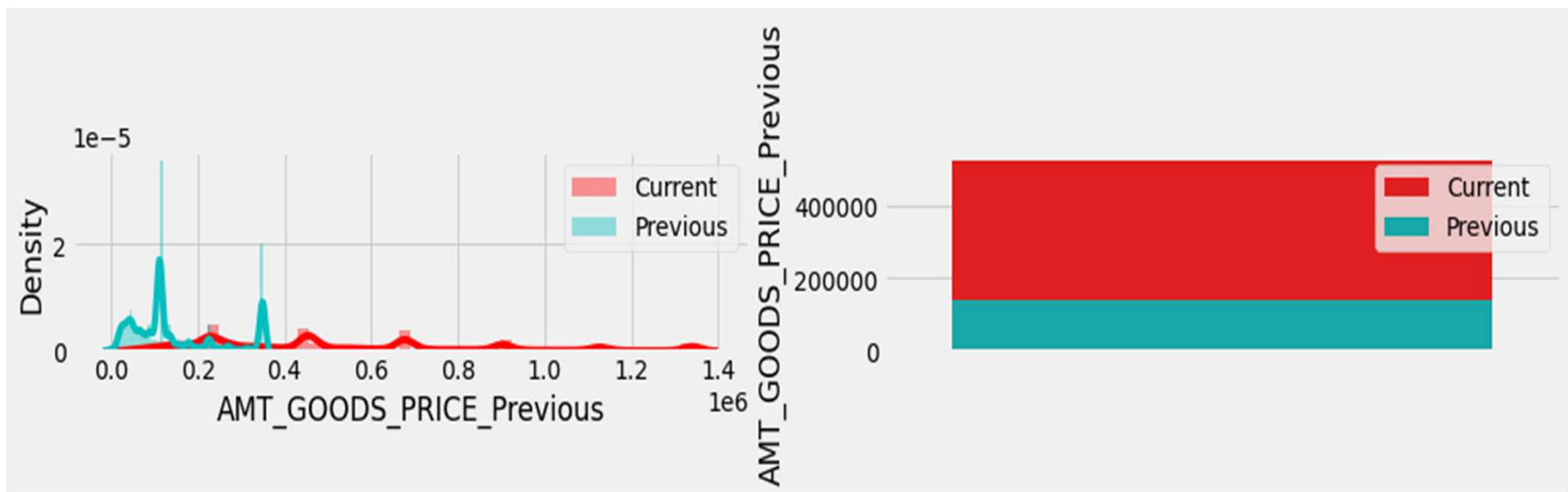
Occupation with their Income type Pattern



Conclusion

- Working defaulter are the majority and among them Labourers are the highest defaulters.
- Pensioner also consists of 12 % among defaulters.

💡 What can be said about GOODS LOAN?



Conclusion

- The demand for goods loan is on increase both number and amount wise.

Conclusion (Defaulter Pattern)

Bank should pay extra attention while dealing with people falling under below mentioned criteria while giving loan -

- Loan Type - Cash loan
- Gender - "Female" clients
- Education Type - "Secondary level" of education (Higher the Education lesser the default)
- Occupation Type - "WORKING" and also little focus on "Commercial associates/Pensioner"
- INCOME Type - "LABOURERS"
- Income Bucket - "LOW/MEDIUM"
- Age Bucket - "Adult"

Conclusion (Defaulter Pattern)

- Interest charged by bank - Higher the rate charged greater the risk
- Credit Amount - Most of the defaulters' lies in the range of 0-7.5 lakh
- Family status - "Married"
- Organization Type - Business Type 3 /Self Employed

Thank you
