

# Presentation On Credit EDA Case study

DONE BY SANTHOSHI SUBHASYA

# Purpose

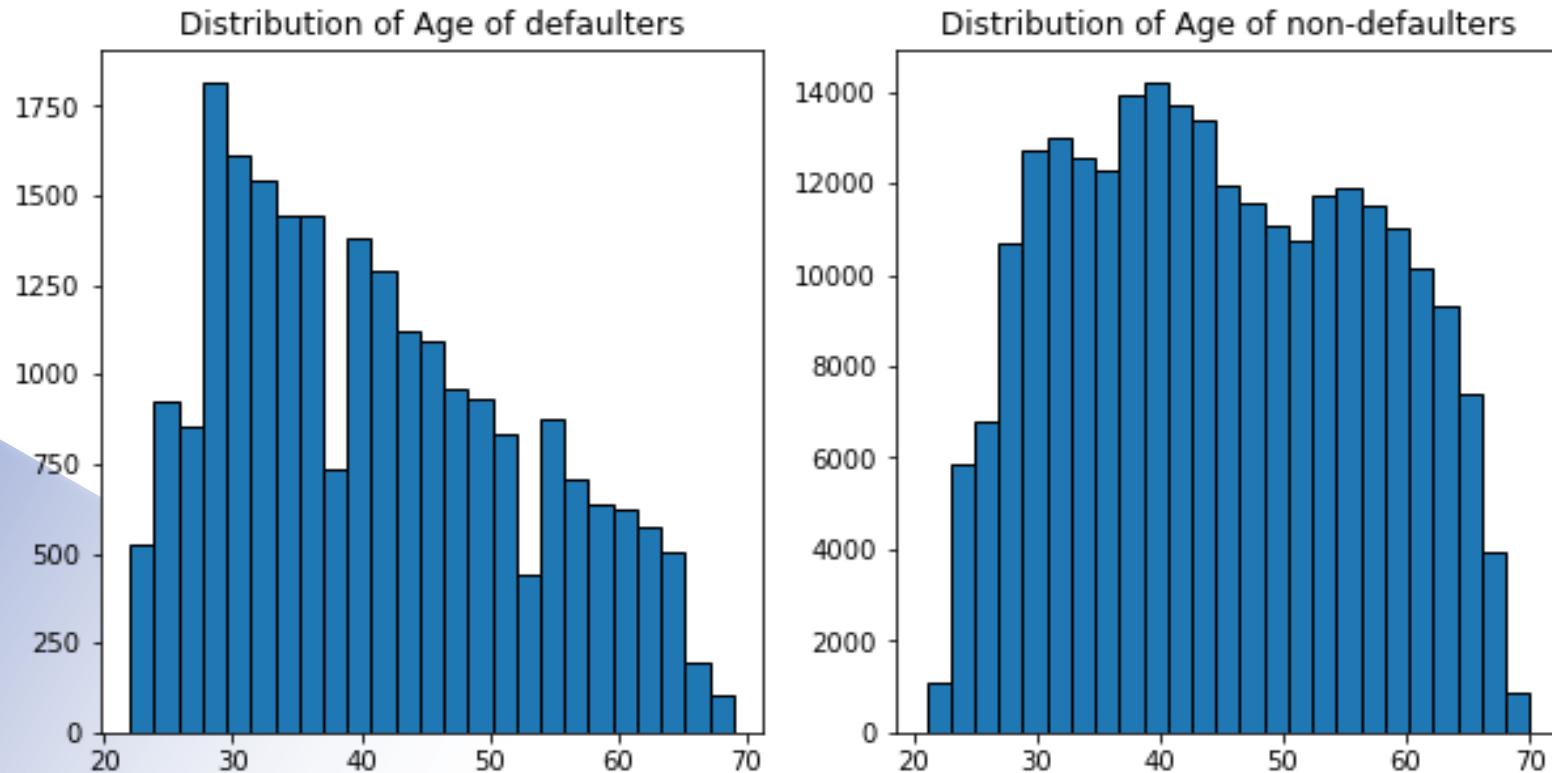
Credit risk analysis will help the company to make a decision for loan approval based on the applicant's profile.

Which controls loss of business to the company and avoid financial loss for the company.

# Steps:

1. Data understanding and sourcing
2. check for Data quality issues and Binning
3. check for Data imbalance and univariate, segmented univariate & Bivariate analysis, correlation
4. Merging of application data with previous application data
5. Data analysis by univariate, segmented univariate , Bivariate analysis and correlation
6. Recommendations and Risks

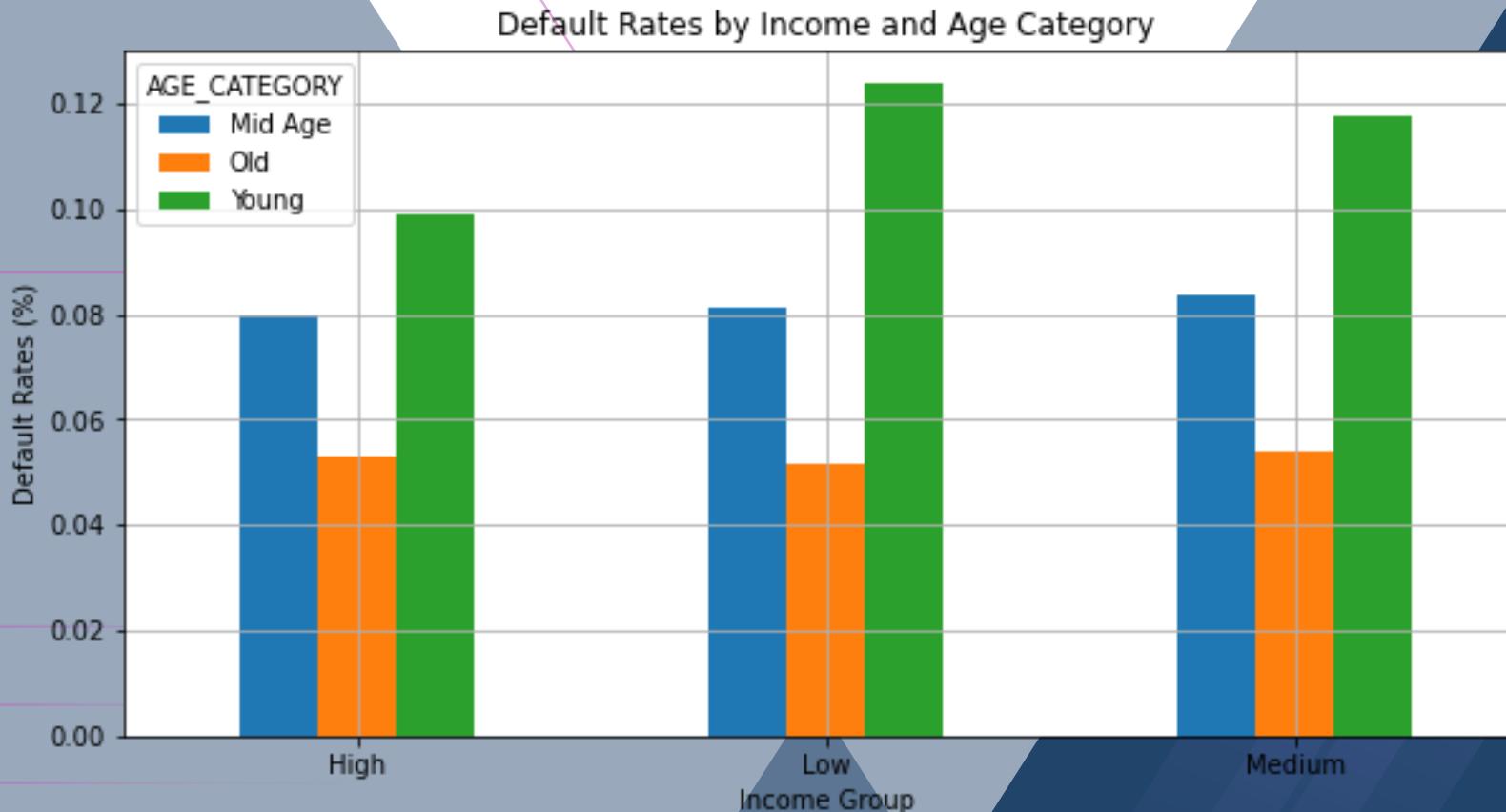
# Univariate Analysis of continuous data in Application data



Around 29 years to 40 years people are more defaulters. There is high chance to be defaulted of the young people. Non-defaulted people are almost equally distributed

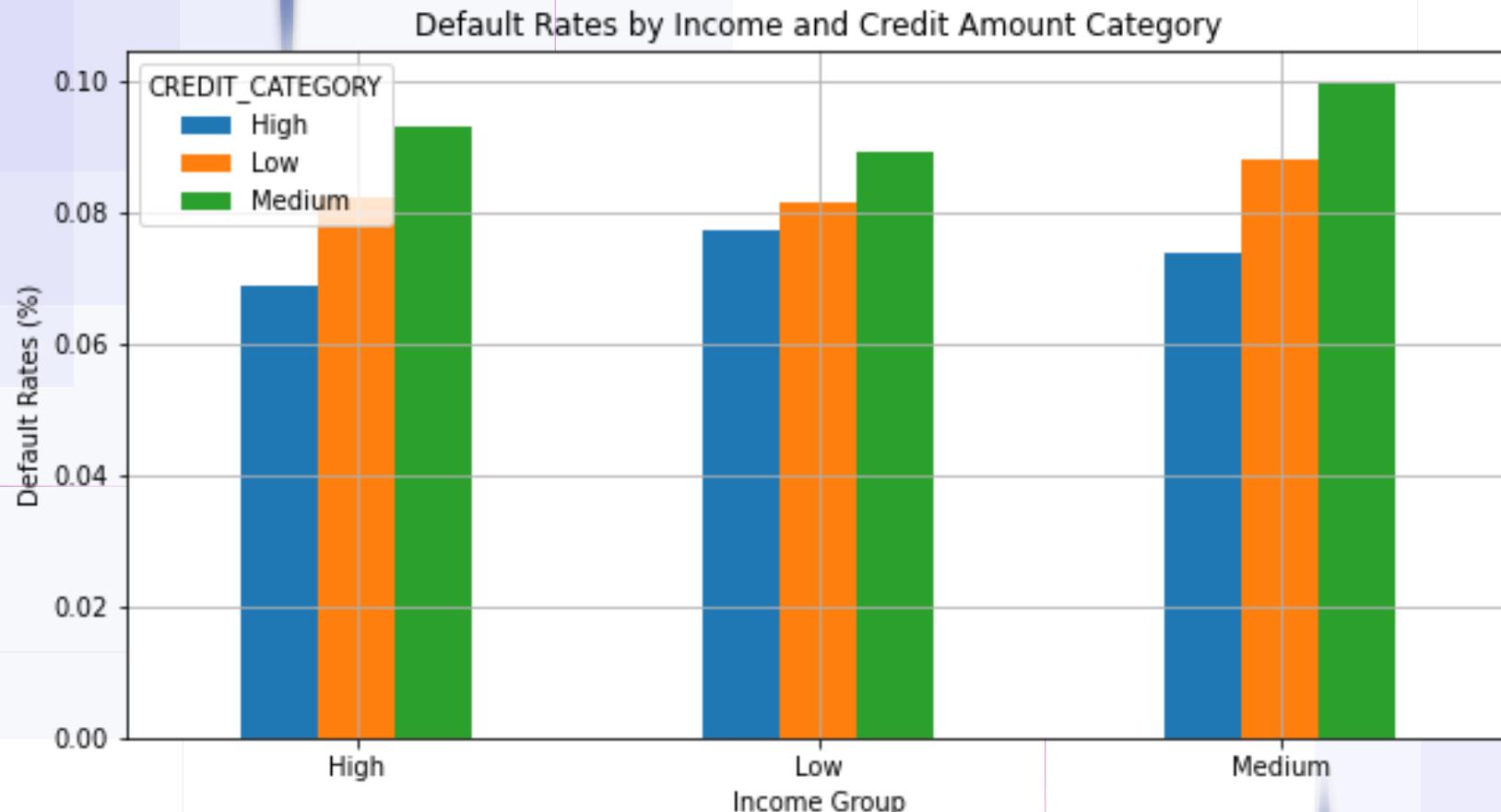
# Segmented univariate analysis

## Age and Income Category Segmented two variables



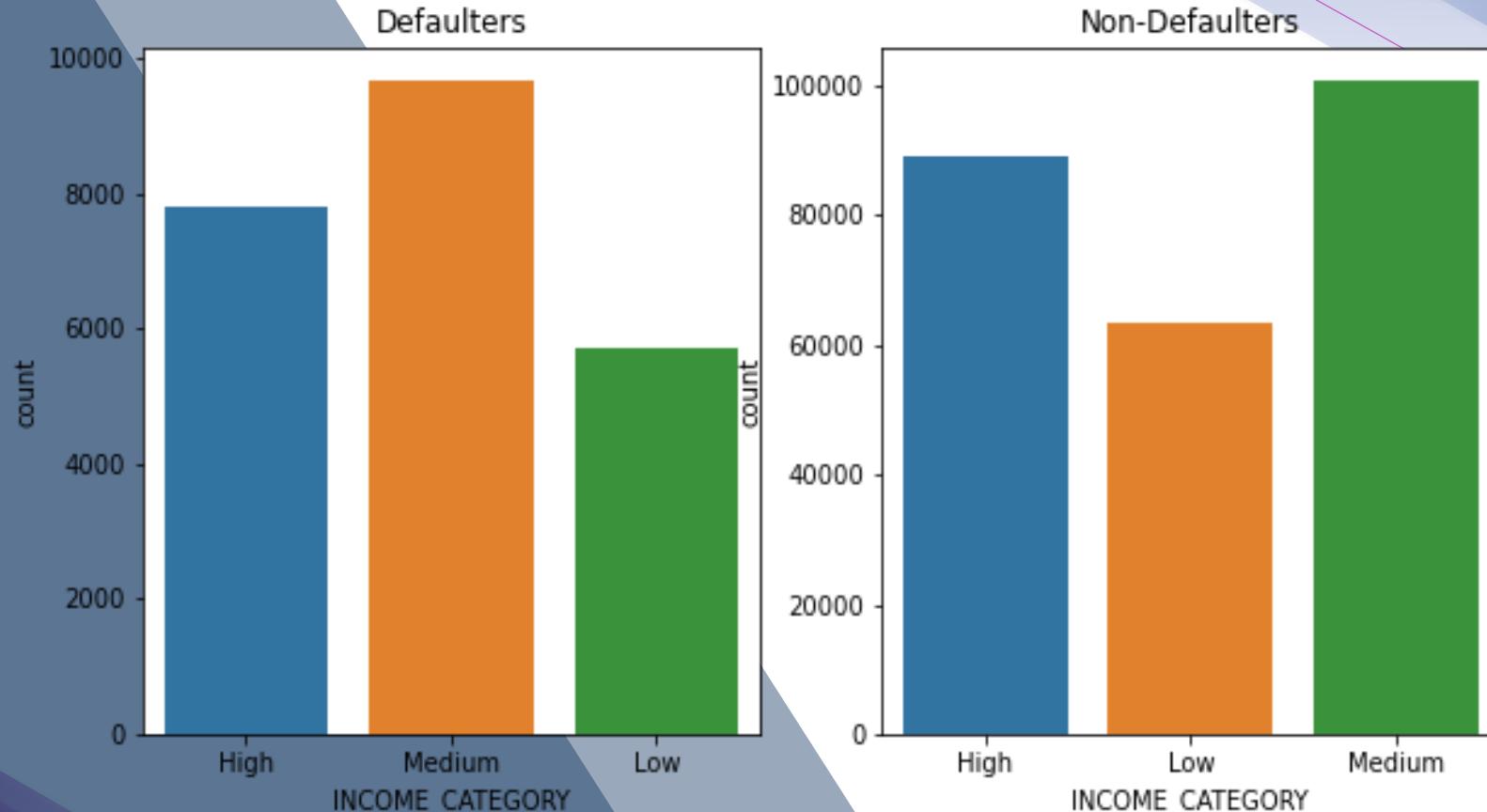
Illustrates that irrespective of the income groups, the chances of default decreases as the age of the applicants increases.

# Income and Credit amount category two segmented



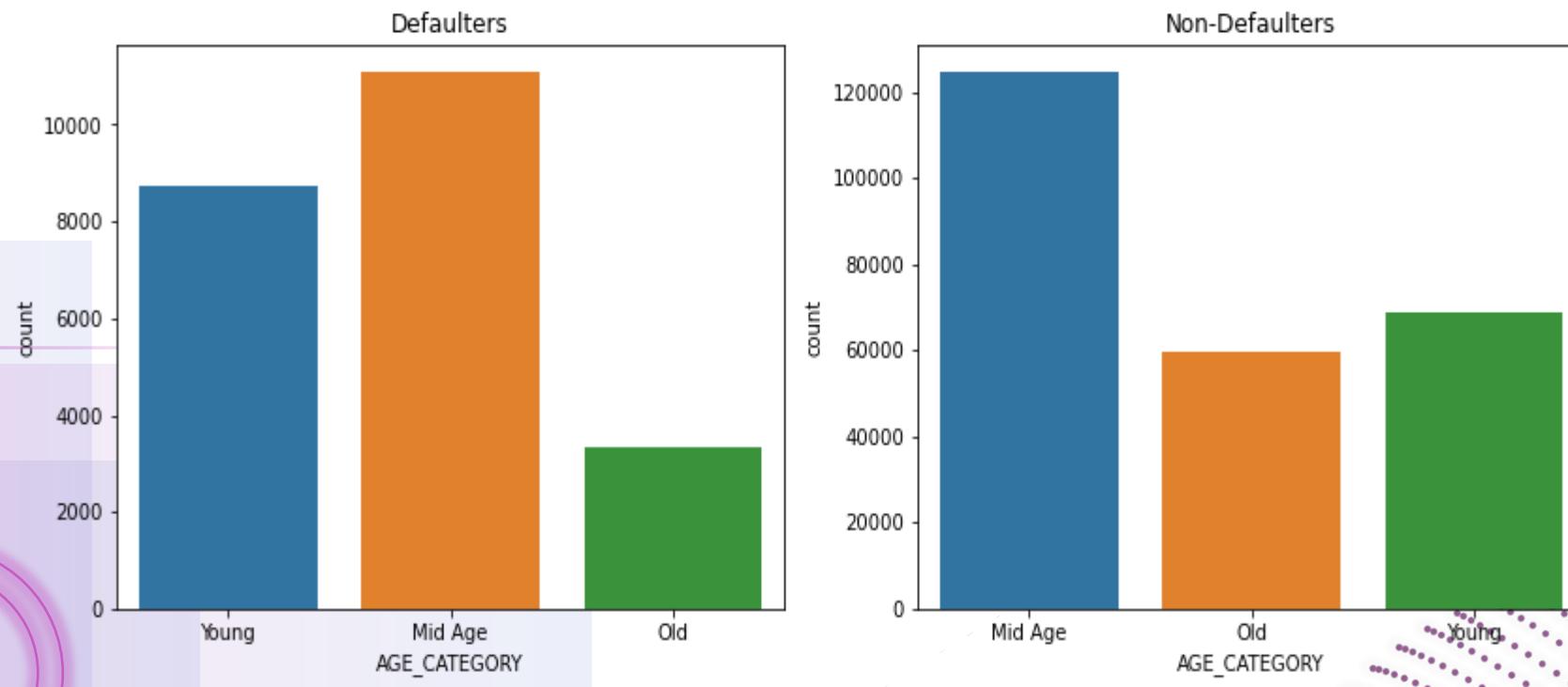
Depicts that irrespective of the income group, the chances of default increases as the credit amount increases. Also if we compare credit amount categories by different income groups, then the default rates for all the three credit amount categories are lower in the high income group relative to the medium and low income groups

# Income Category Segments For Both The Default And Non-default

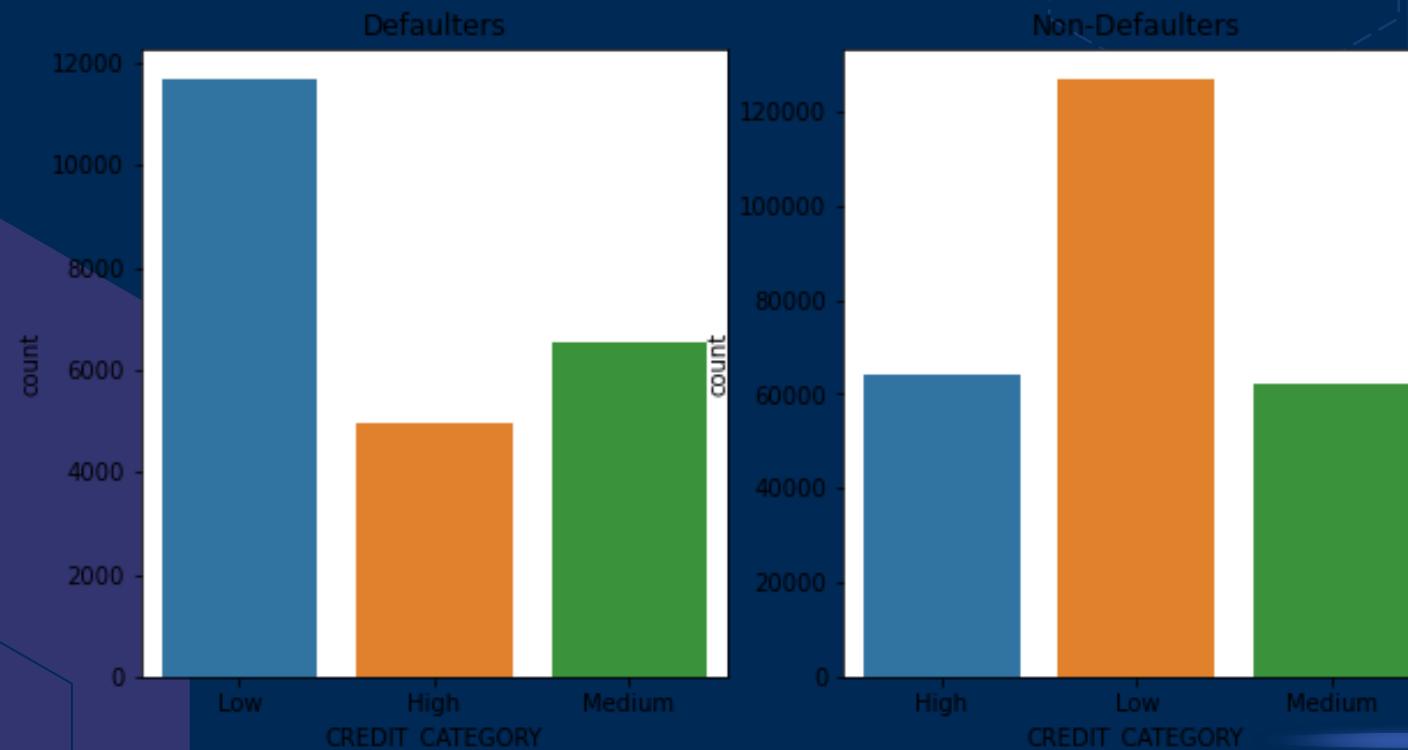


Low income group has more defaulter followed by high income group.

# AGE\_CATEGORY segments for both the default and non-default

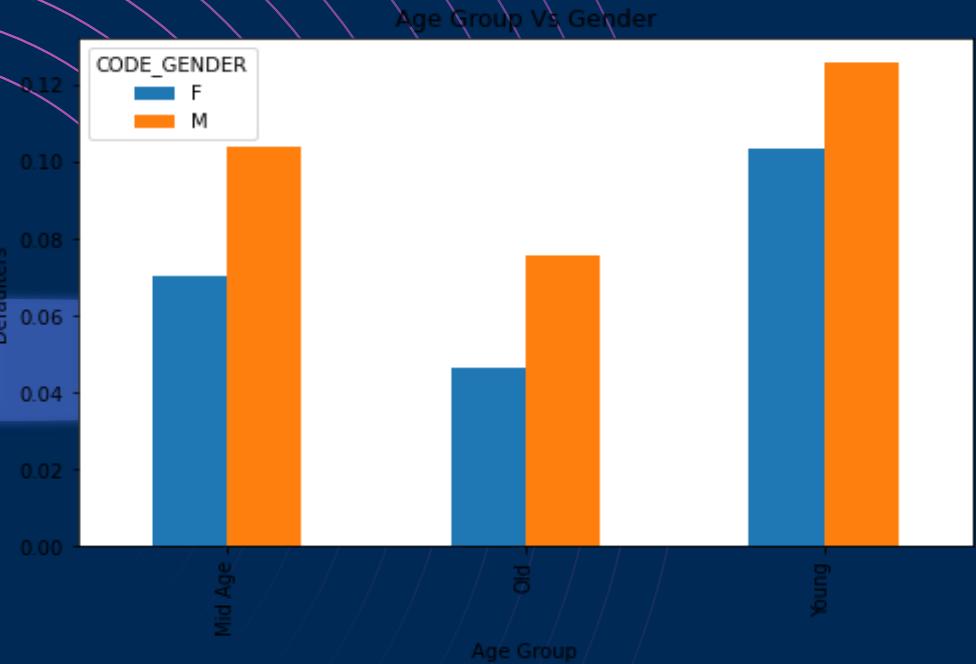
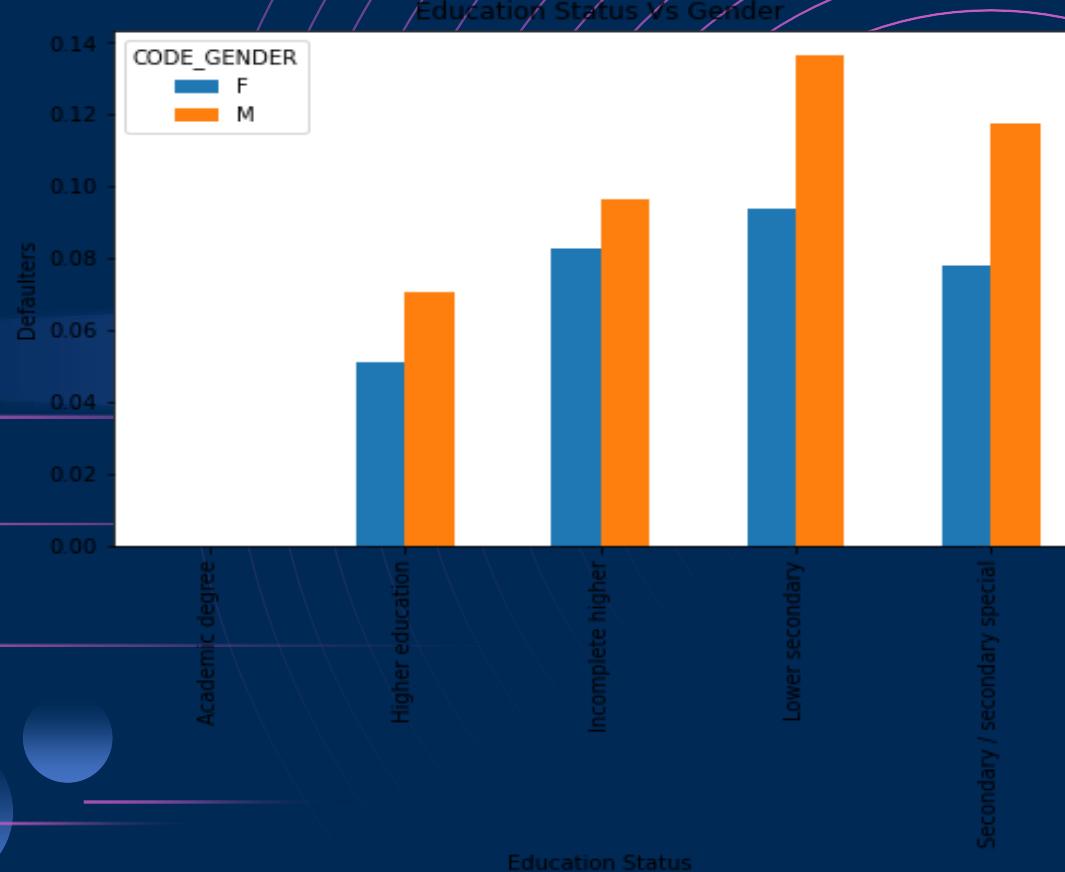


# CREDIT\_CATEGORY segments for both the default and non-default



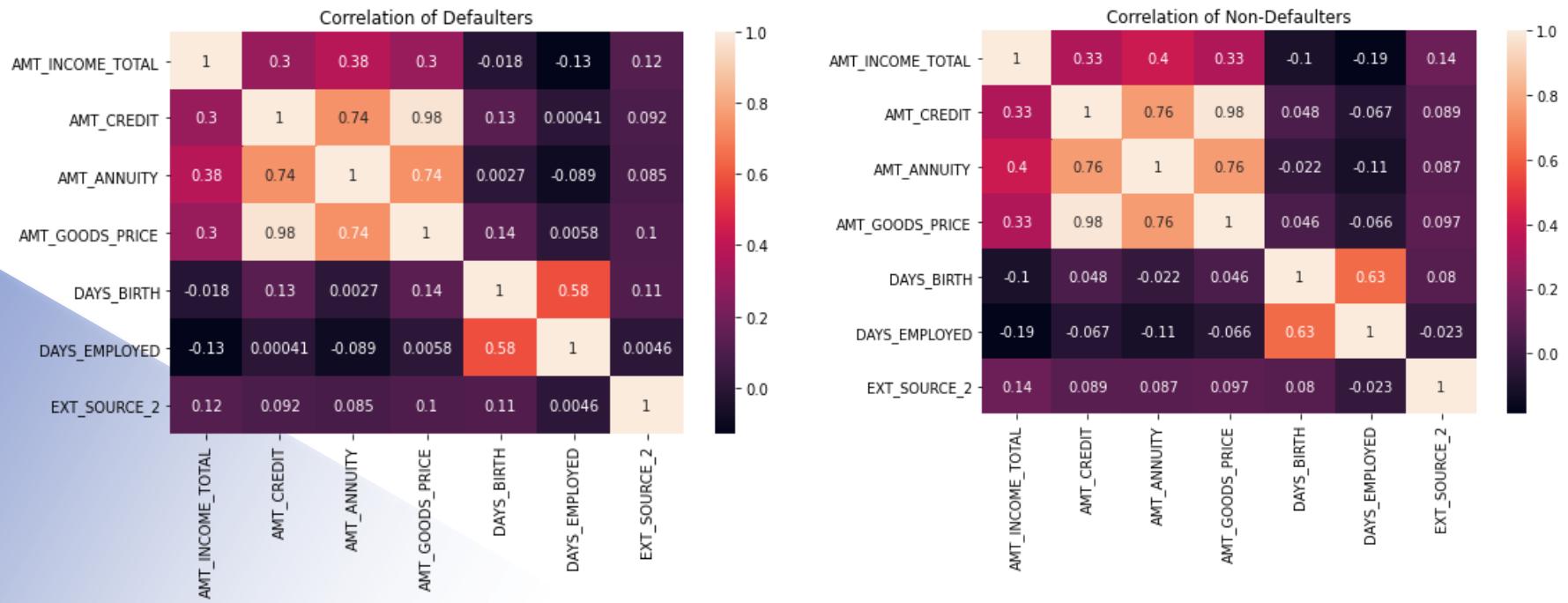
# Education Status Vs Gender wise defaulters

## Age Group Vs Gender



# Correlation of the continuous variables of Defaulter data frame

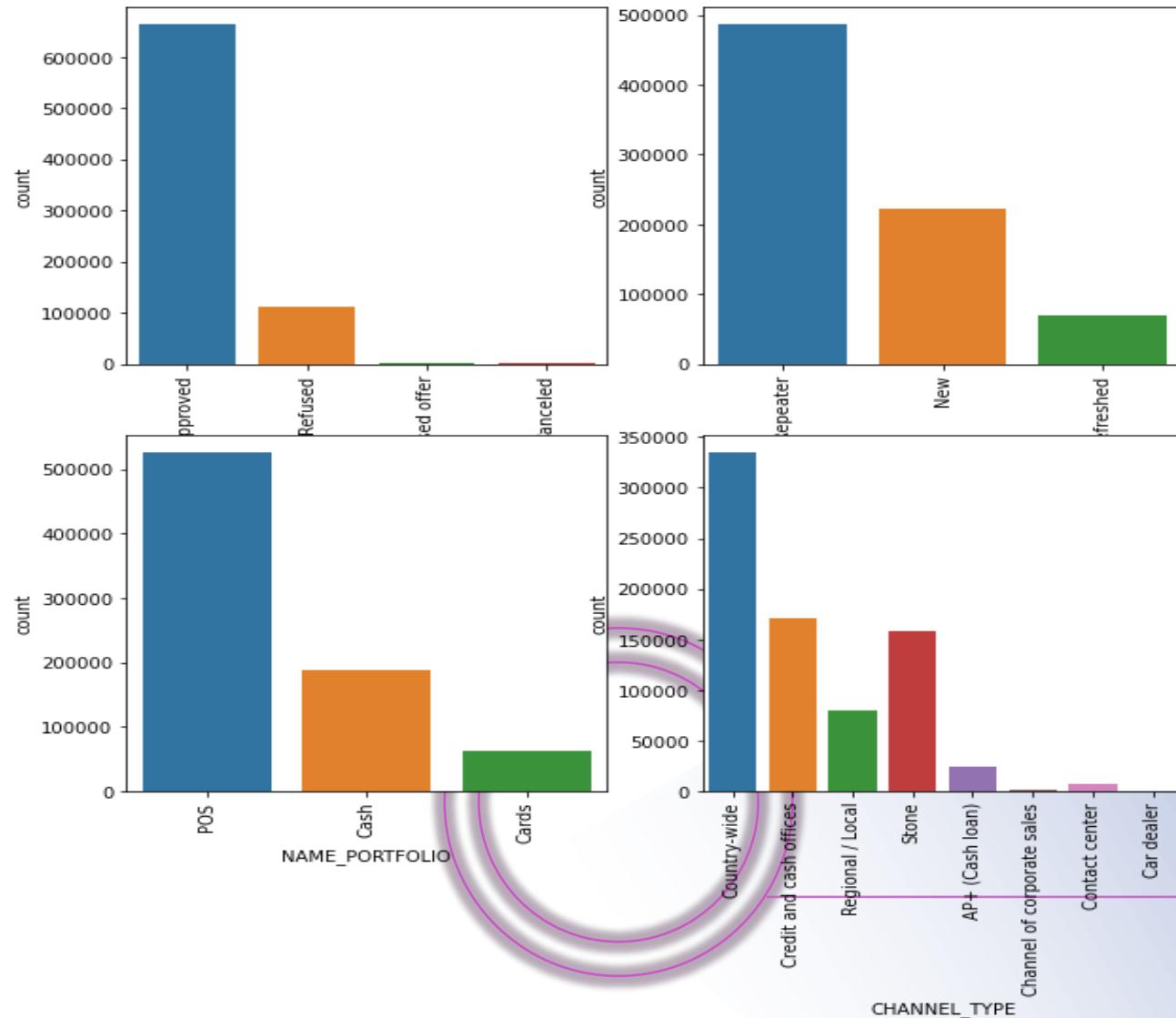
## Correlation of the continuous variables of Non-Defaulter data frame



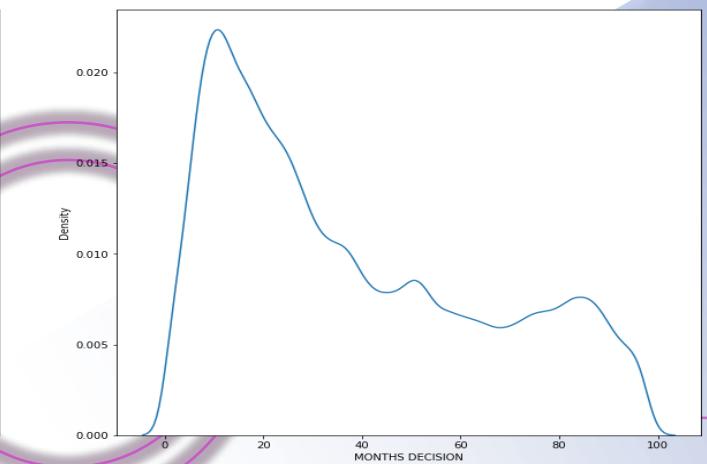
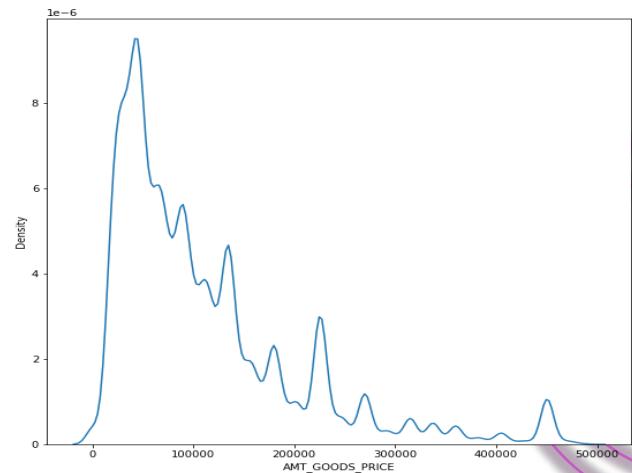
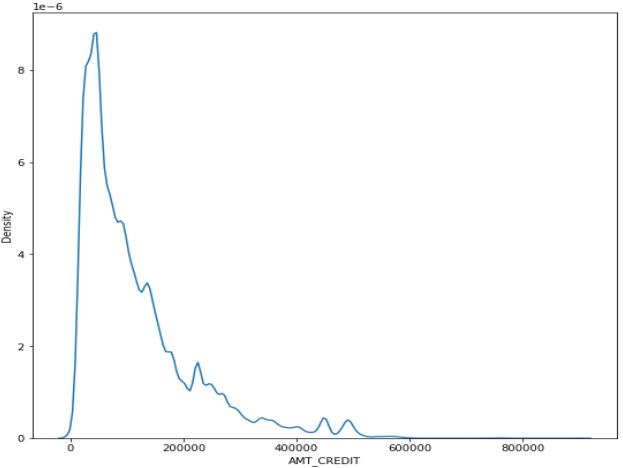
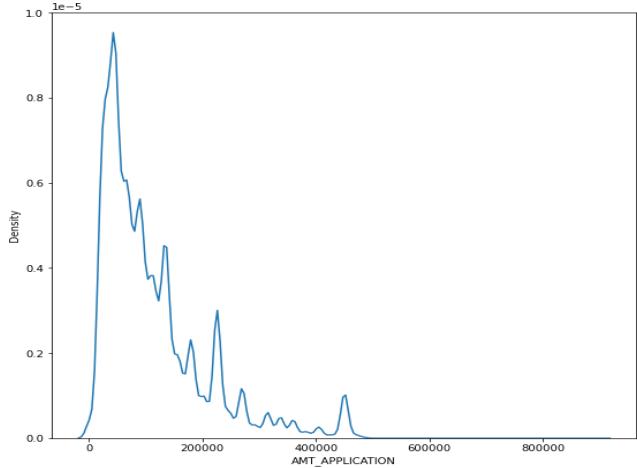
**GOODS\_PRICE** and **AMT\_CREDIT**, **AMT\_ANNUITY** and **AMT\_AMT\_CREDIT** are highly correlated. External Rating is highly correlated with all **DAYS\_BIRTH**(Age), **GOODS\_PRICE**, **AMT\_CREDIT**.

**GOODS\_PRICE** and **AMT\_CREDIT**, **AMT\_ANNUITY** and **AMT\_AMT\_CREDIT** are moderately correlated with each other. External Rating is highly correlated with all **DAYS\_BIRTH**(Age), **GOODS\_PRICE**, **AMT\_CREDIT**

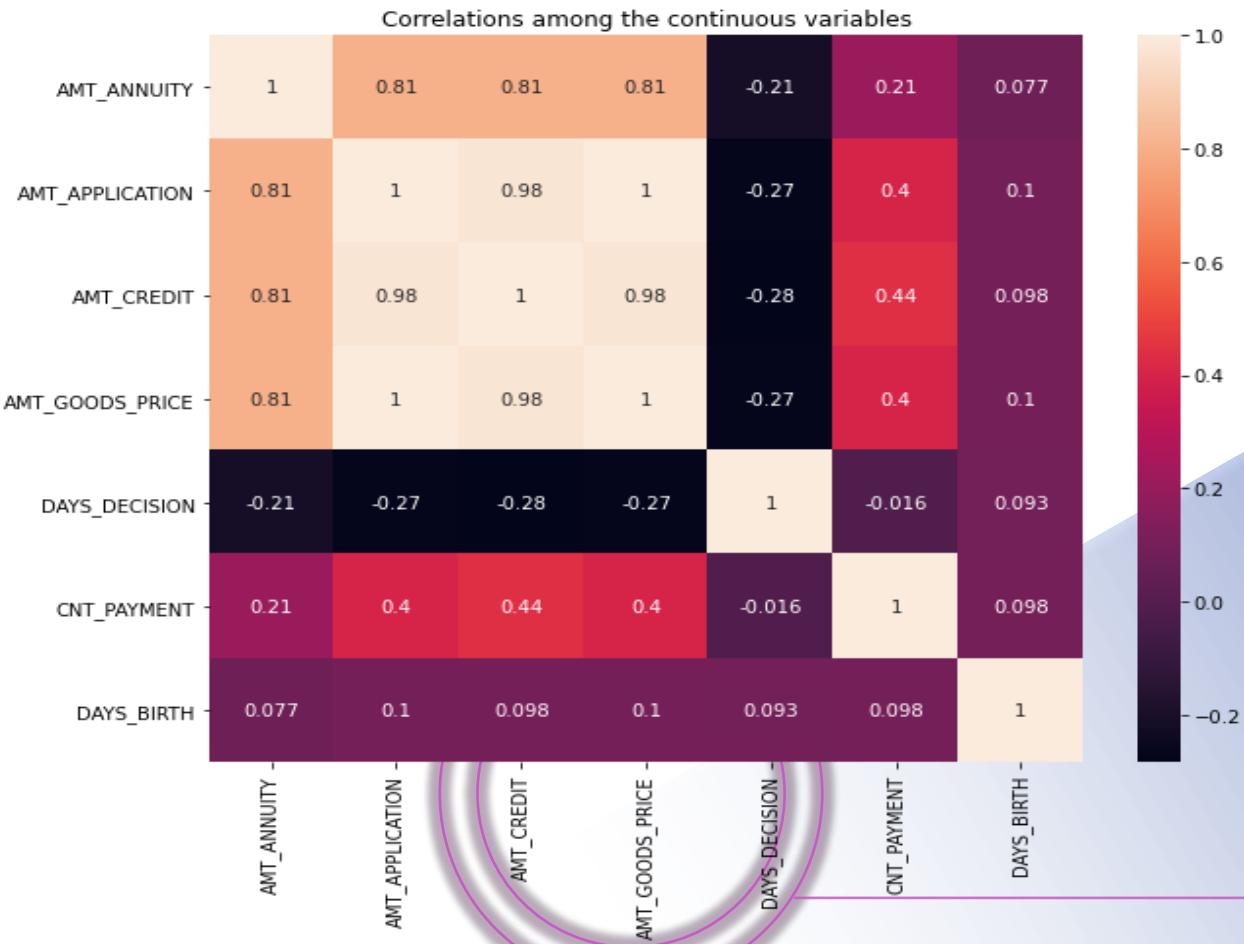
# Univariate Analysis on categorical columns



# Univariate analysis on continuous columns



# Correlations among the continuous variables



# Recommended group where loan can be credited (Less chance to be defaulter)

- Clients who are working as a state servant.
- Old people of any income group.
- Client with high income category.
- Old female client.
- Client with higher education (female).
- Any client who's previous loan was approved.
- Widow who has unused previous loan status.
- Refreshed client who has unused loan status previously.

## Risky Group

- Lower secondary educated clients are the most in number to be defaulted when their previous loans were cancelled or refused.
- Male clients with civil marriage.
- Previously refused loan status group

# THANK YOU