

# EDA CASE STUDY

Pieush Vyas (DS-27)

# AIM

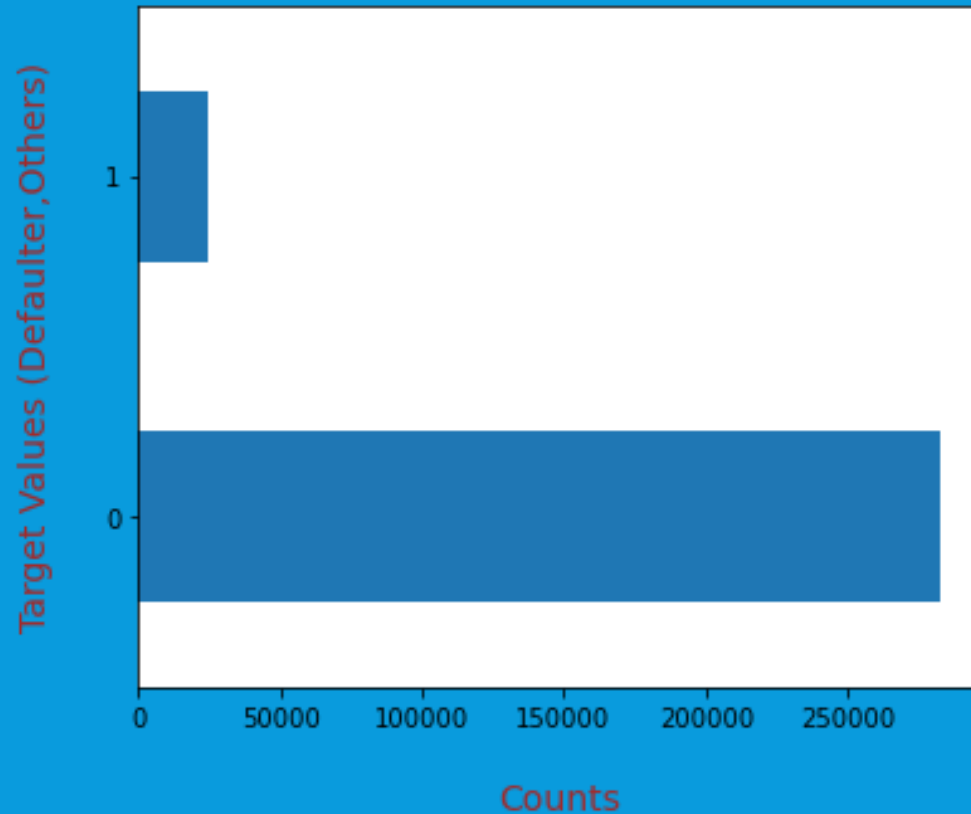
- Analyze the patterns available in the given datasets which helps to understand about the client behavior and type with payment difficulties or not.
- With the analysis of another dataset one can understand the previous history of clients whether loan were Approved/ Cancelled/ Refused/ Offer were unused.
- All such analysis allows to take fare business decisions further to give loan with zero or very less financial loss.

# APPROACH

- Use the concept of EDA for cleaning, preparing, examining and for data Visualizing.
- During this process identify useful row and columns, replace or delete the unwanted/ null values from various rows/ columns.
- Plot various graphs to visualize the data which helps to take further decision in terms of Target data or loan status.
- For the above purpose we do Univariate, Bi-Variate or Multi-Variate analysis.
- During all the Visualization we use various Bar-Graphs, Box-plot, and Heat-maps which helps to do better analysis, comparison, and to find out correlation among various numerical variables.

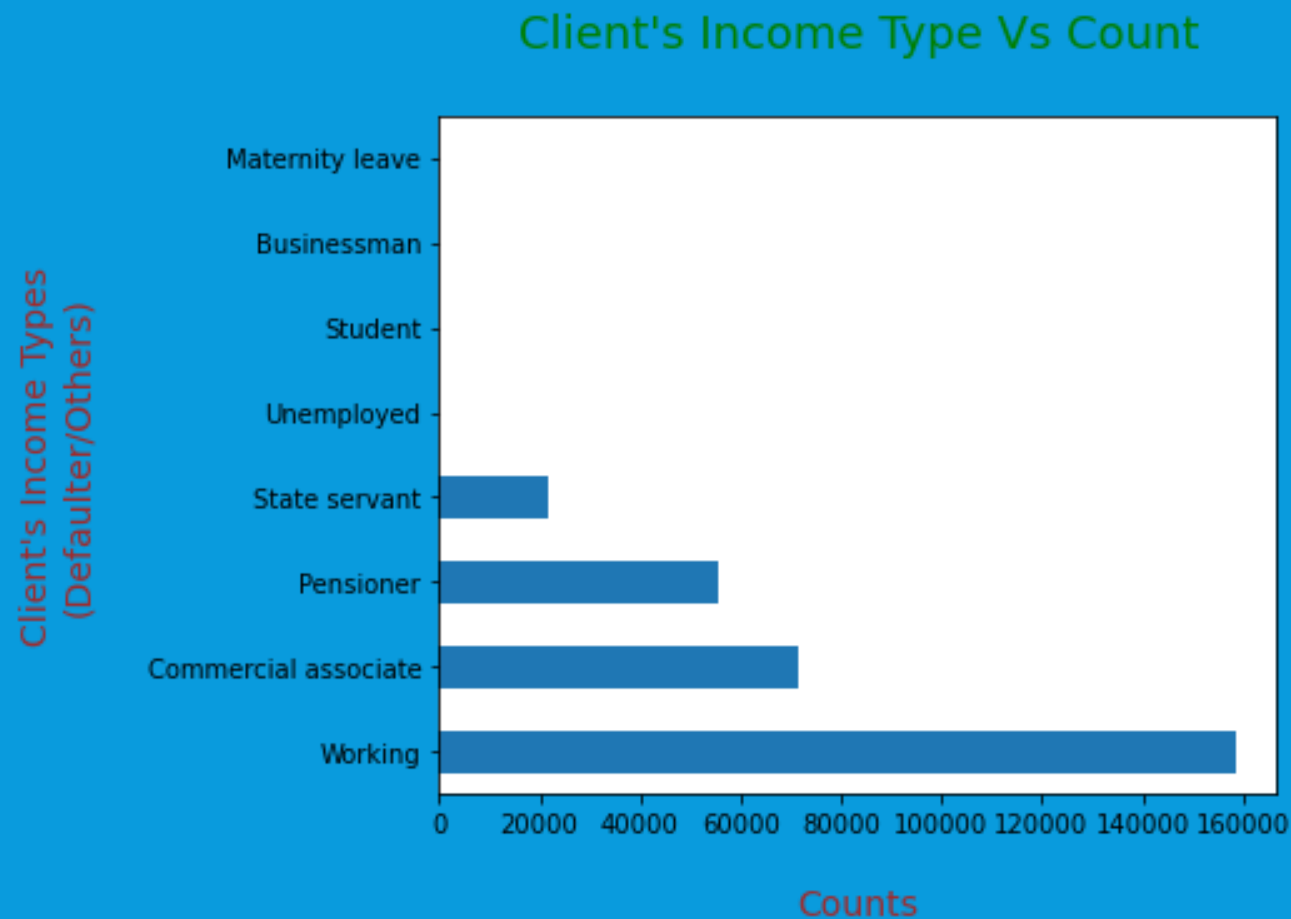
# ANALYSIS FOR APPLICATION DATA

Defaulters/delayed payment, Other Payments Vs Count



This Bar-graph shows that loan defaulters are very less than all other cases.

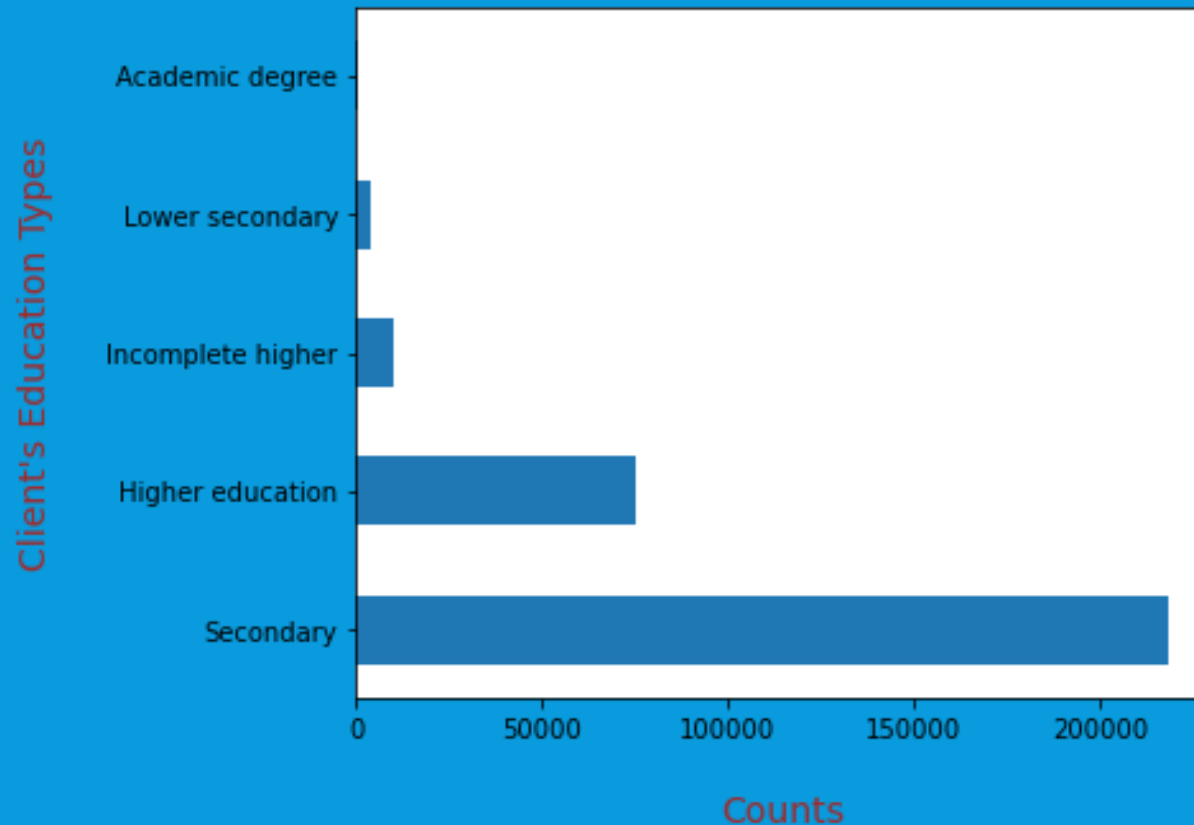
# ANALYSIS FOR APPLICATION DATA



This Bar-Graph shows that maximum people's source of income is Job/working and others are Commercial Associate, Pensioner, and State Servant whereas very few are Businessman, Students, Unemployed and are on Meternity leave.

# ANALYSIS FOR APPLICATION DATA

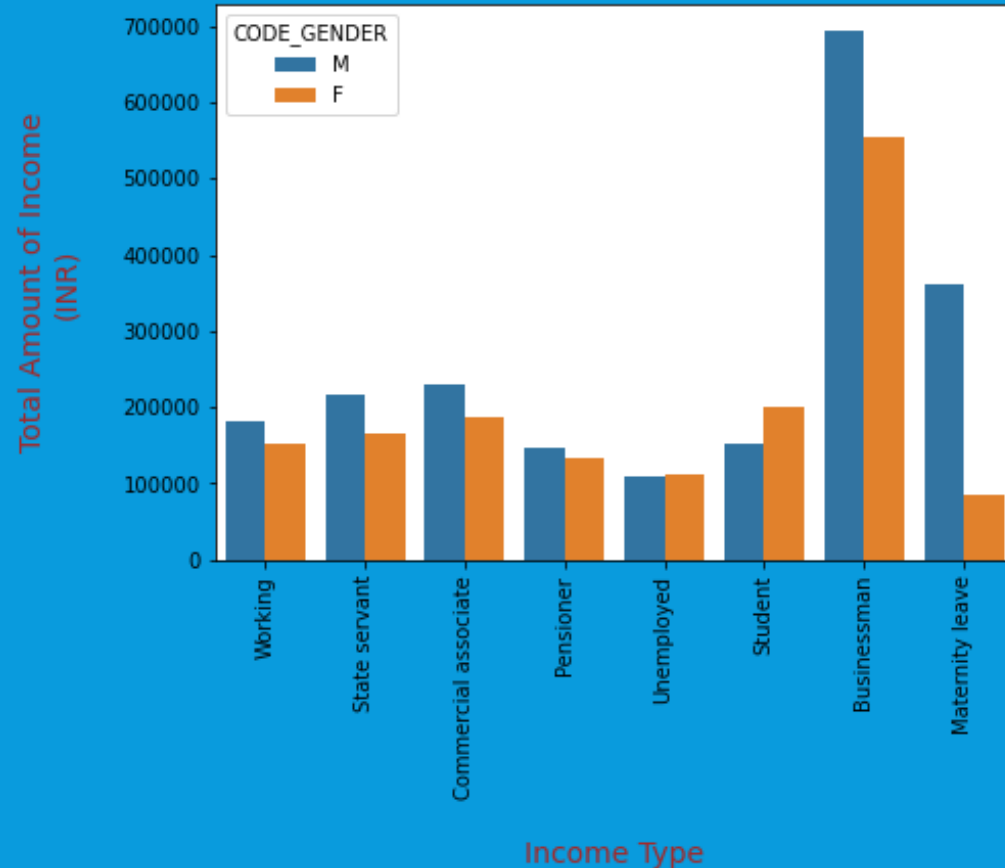
Education Type Vs Count



According to this Bar-Graph we can say that maximum people have done up-to Secondary education, whereas very less have done Academic degree.

# ANALYSIS FOR APPLICATION DATA

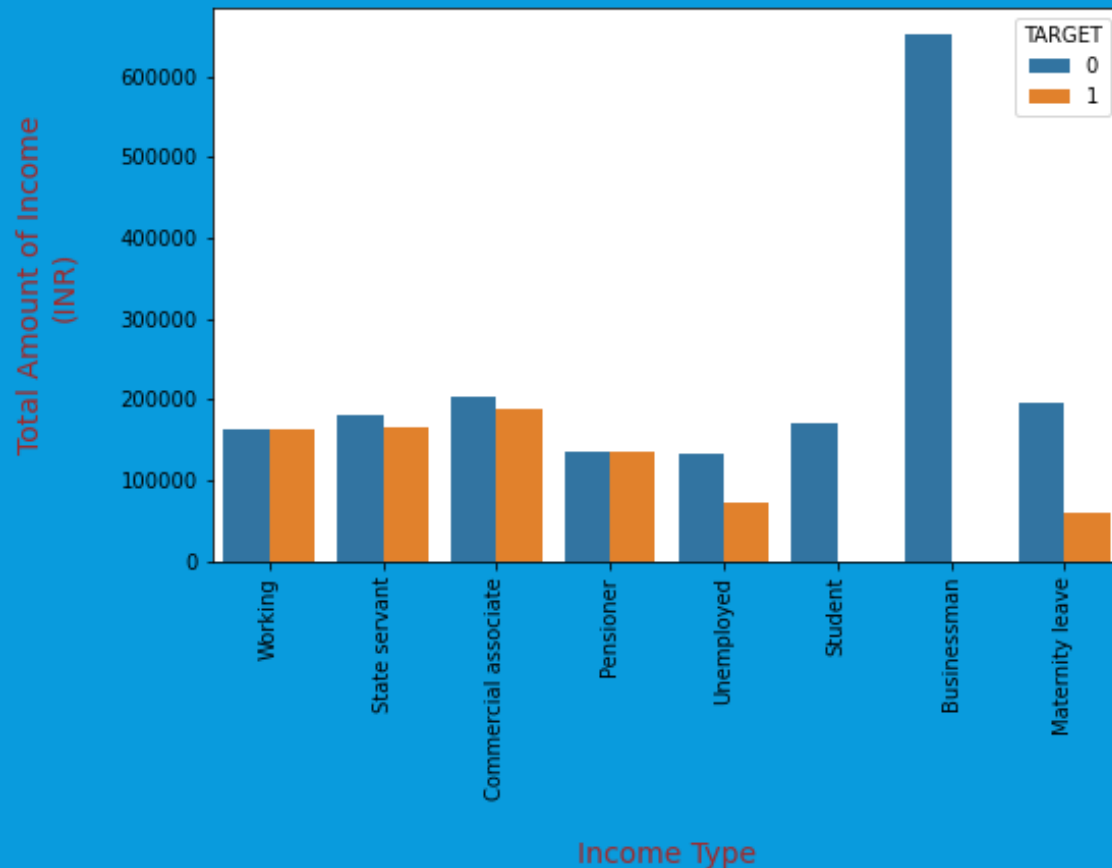
Income Type Vs Total Income in terms of Gender



According to this Bar-Graph it is clear that among all the categories of Income Type (accept 'Unemployed and 'Student') 'Male' has higher Total income than 'Female'.

# ANALYSIS FOR APPLICATION DATA

Income Type Vs Total Income in terms of Target



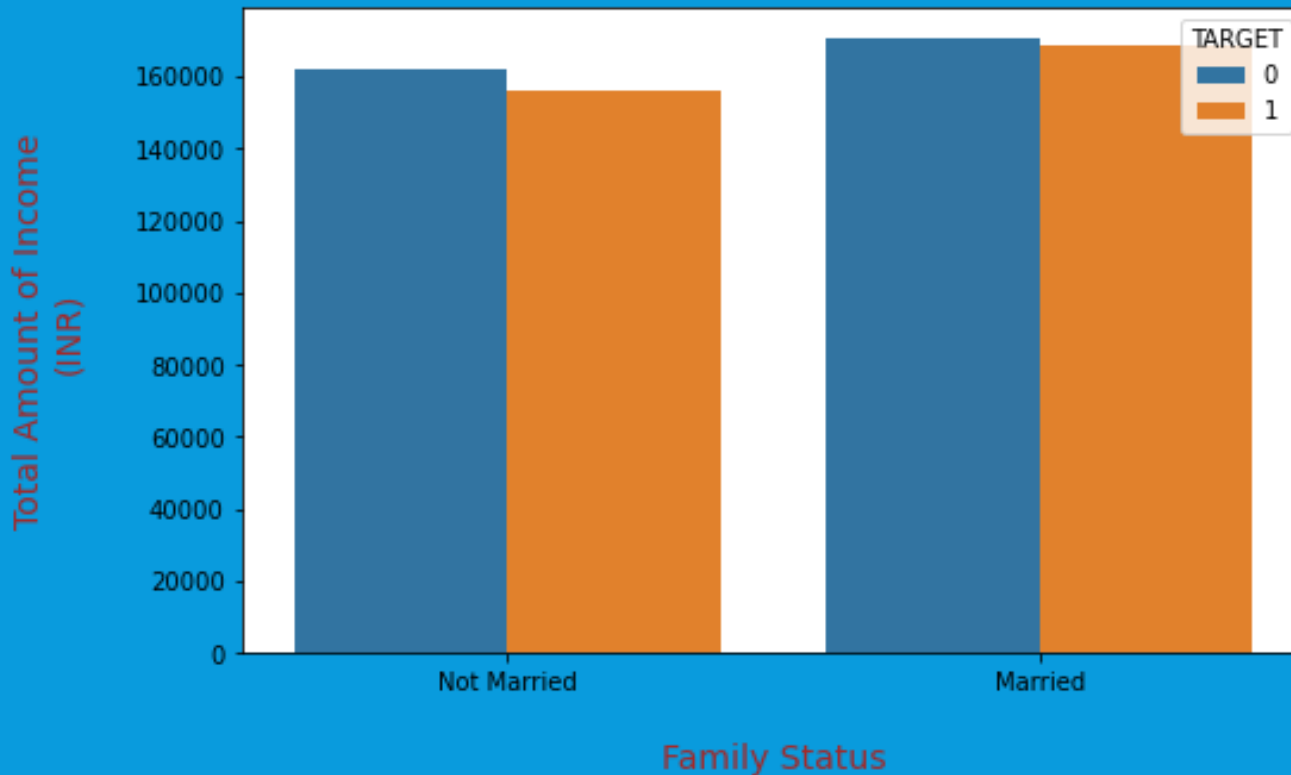
Inferences 1: With the help of graph it can be seen that there are no people who are businessman and having average income more than INR 600000 are not the loan defaulter / did delayed payment, whereas the people who are Working/Pensioner has equal amount of loan defaulter and other cases.

Inferences 2: It means Businessman are less Defaulters / did delayed payment.



# ANALYSIS FOR APPLICATION DATA

Family Status Vs Total Income in terms of Target



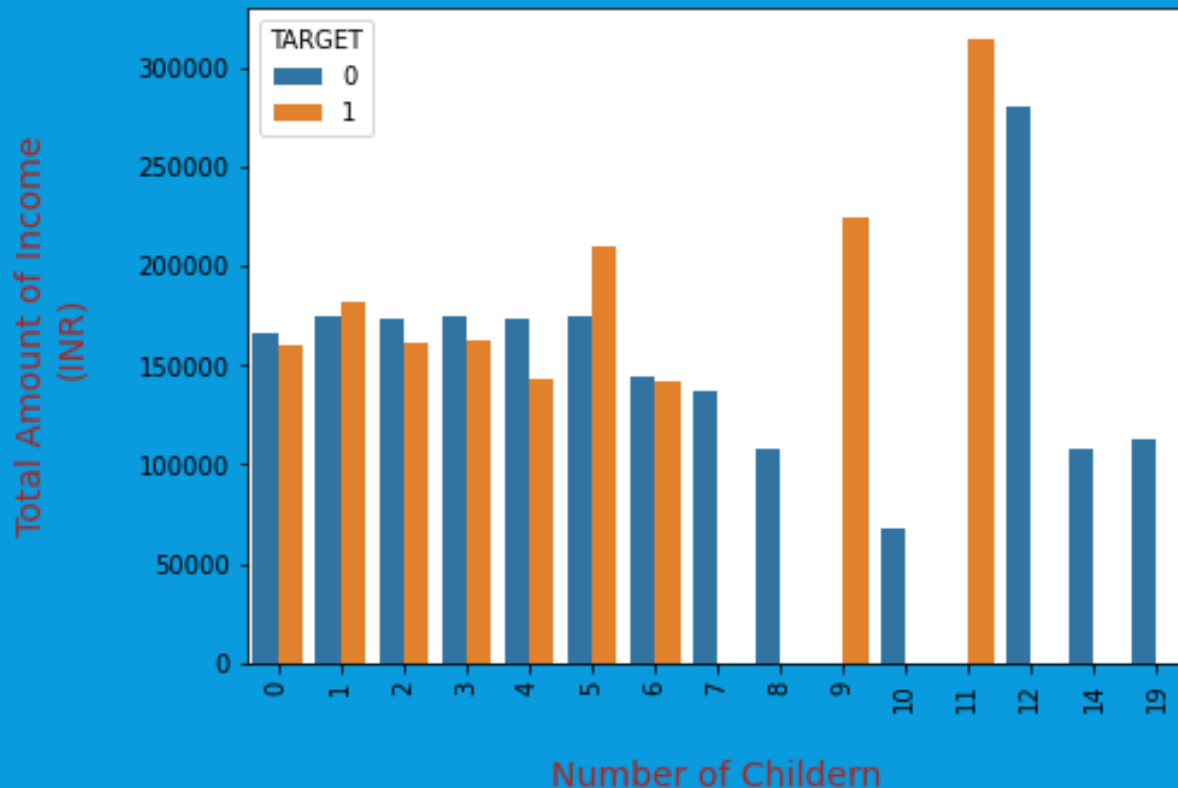
Inference 1: The above Bar Graph is showing that 'Married' people who has good income having less marginal difference between loan defaulter and all other cases.

Inference 2: People who are 'Not Married' and having higher income has little less defaulter than Married people.

Inferences 3: It means 'Not Married' people can be preferred over 'Married' people during loan approvals.

# ANALYSIS FOR APPLICATION DATA

Number of Children Vs Total Income in terms of Target



Inference 1: In most of the cases as number of children increases loan defaulters/delayed payment increases.

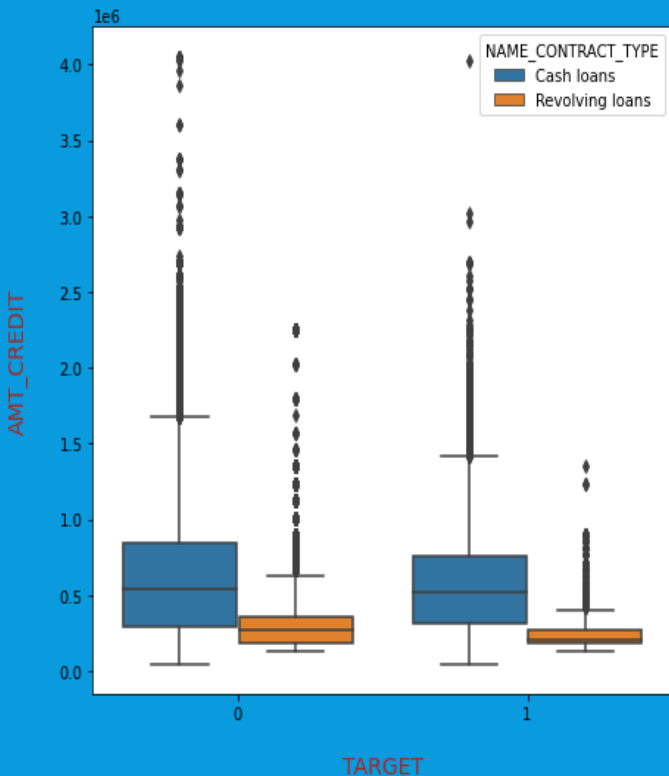
Inference 2: As Total Income increases number of children increases.

Inferences 3: Therefore, if number of Children are more loan defaulters/ delayed payment are more, hence for up to 4 Children loan defaulters are less.

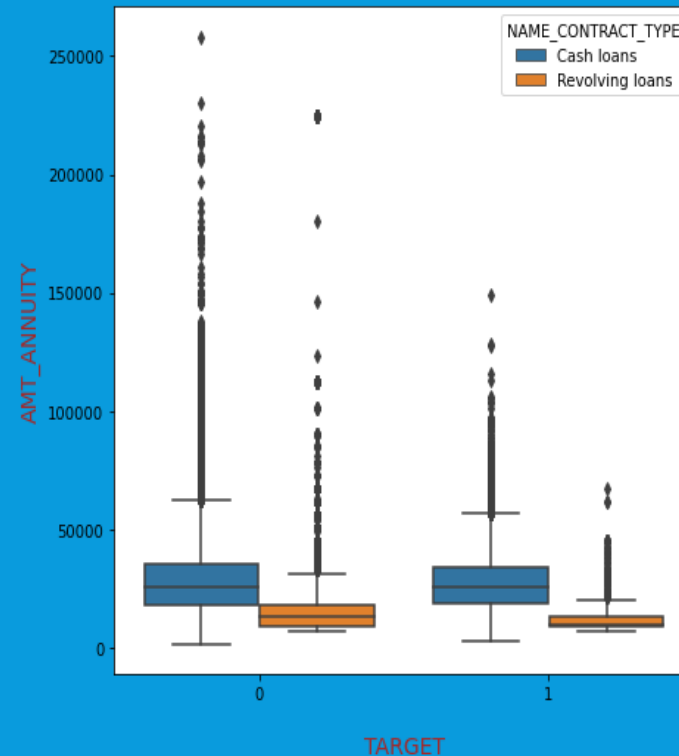
Inferences 4: Bank should avoid giving loan to the people who having more Children.

# ANALYSIS FOR APPLICATION DATA

Amount Credit Vs Target



Amount Annuity Vs Target (Defaulter/Others)



Inference 1: For the client whether they are loan defaulter/ did delayed payment or for all other cases having higher Amount Annuity for Cash loans than Revolving loans.

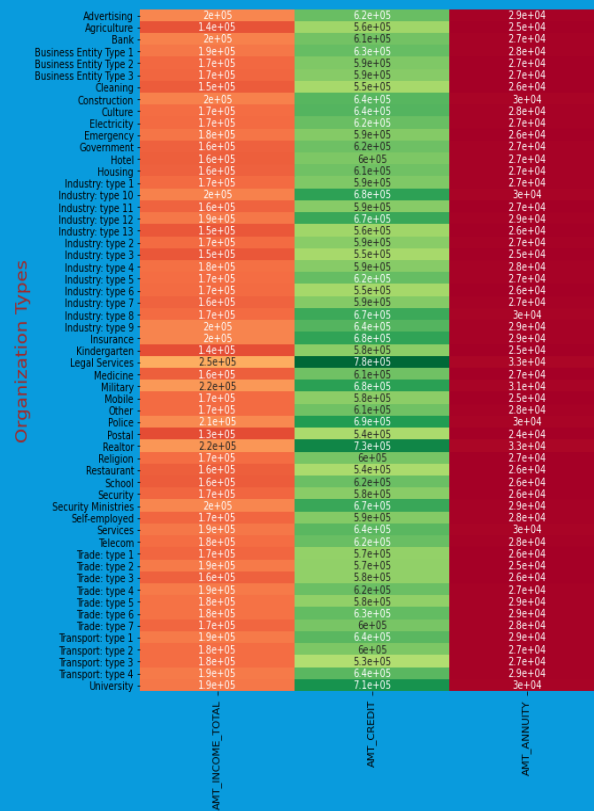
Inference 2: Amount Annuity is equally distributed for Cash loans whether it is loan defaulter/ did delayed payment or for all other cases.

Inference 3: For both the Targets Revolving loans maximum is less than the minimum of Cash loans.

Inference 4: For both the cases Credit amount and Amount Annuity is equally divided for Cash loans whereas, Revolving loan's upper half is higher in both the Targets.

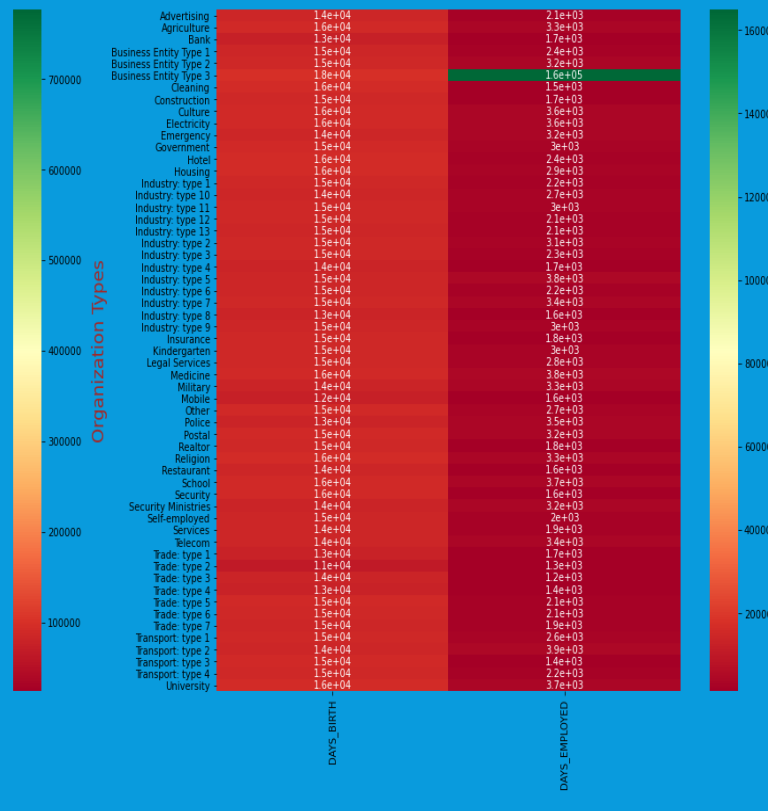
# ANALYSIS FOR APPLICATION DATA

Organization Types Vs Different Amounts



Amount Totals

Organization Types Vs Different Days



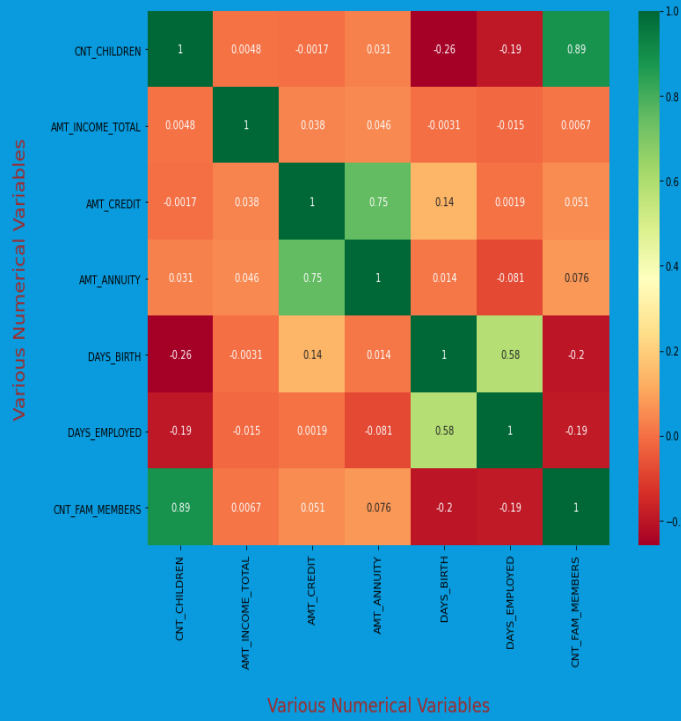
Number of Days

Inference 1: Although Maximum Organization Type for the 'legal services' is in medicore range still maximum Credit amount of the loan is for the same Organization Type ('Legal Services') whereas Amount Annuity is less than 100000 for all the Organization Types.

Inference 2: Maximum number of days employed at the time of loan are for the Organization type 'Business Entity Type 3'.

# ANALYSIS FOR APPLICATION DATA

Correlation among Various Numerical Data for loan defaulters



Correlation among Various Numerical Data for All other cases



Inferences 1: No. of Children and total family members count are proportionate with each other for both the cases, therefore more no. of Children means more no. of family members and less total income.

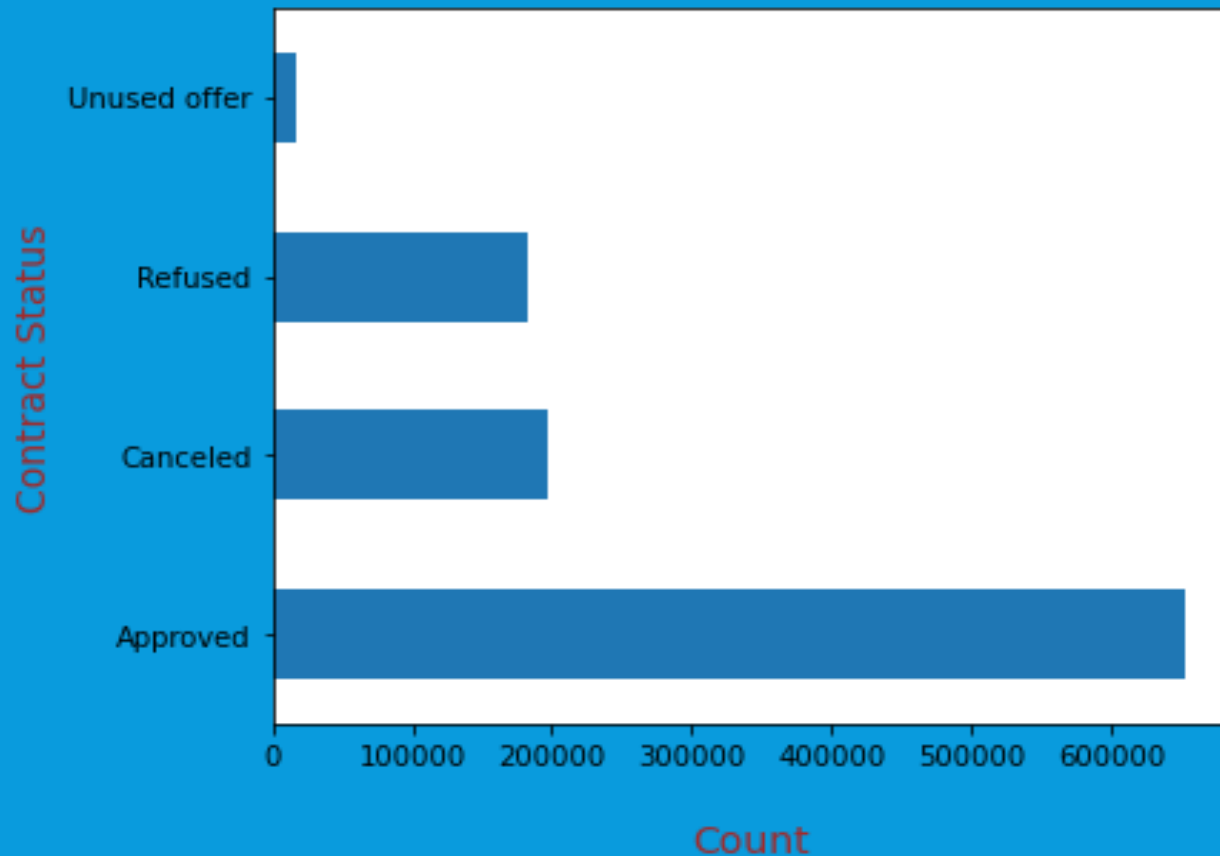
Inferences 2: Total income amount and Credit income is inversely proportional to no. of children, therefore less total income, credit amount for more number of children but this ratio is little better for 'Target 0' instead of 'Target 1'.

Inferences 3: For both the cases 'Amount Annuity' is inversely proportional to Days Employed, therefore Amount Annuity is less for more no. of days employed.

Inferences 4: For maximum variable like Total Income, Credit Amount, Amount Annuity value is higher for 'Target0' with respect to 'Target1'.

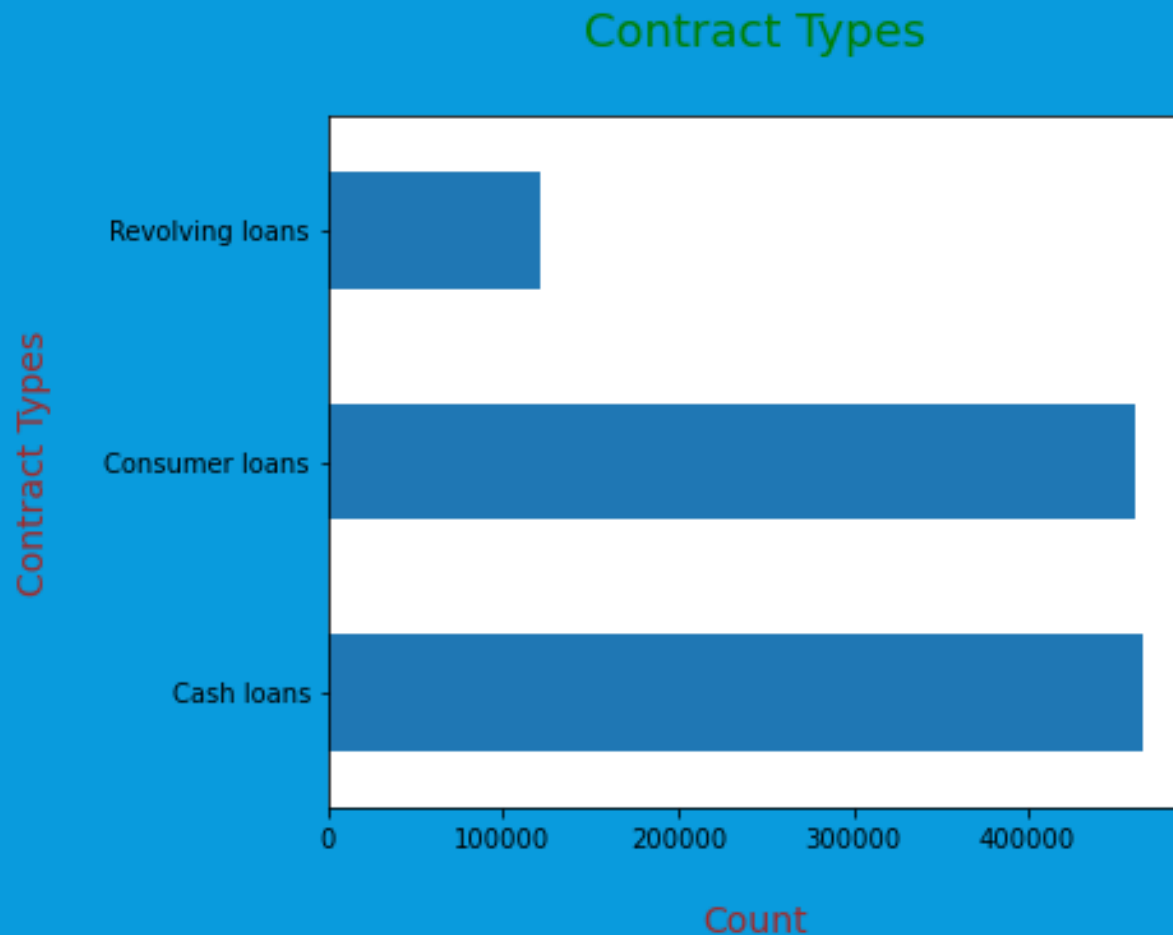
# ANALYSIS FOR PREVIOUS APPLICATION DATA

Various Contract Status Counts



Inferences 1: According to the above Bar-Graph 'Approved' loan has highest number among all the four Contract Status whereas 'Unused offer' is very less.

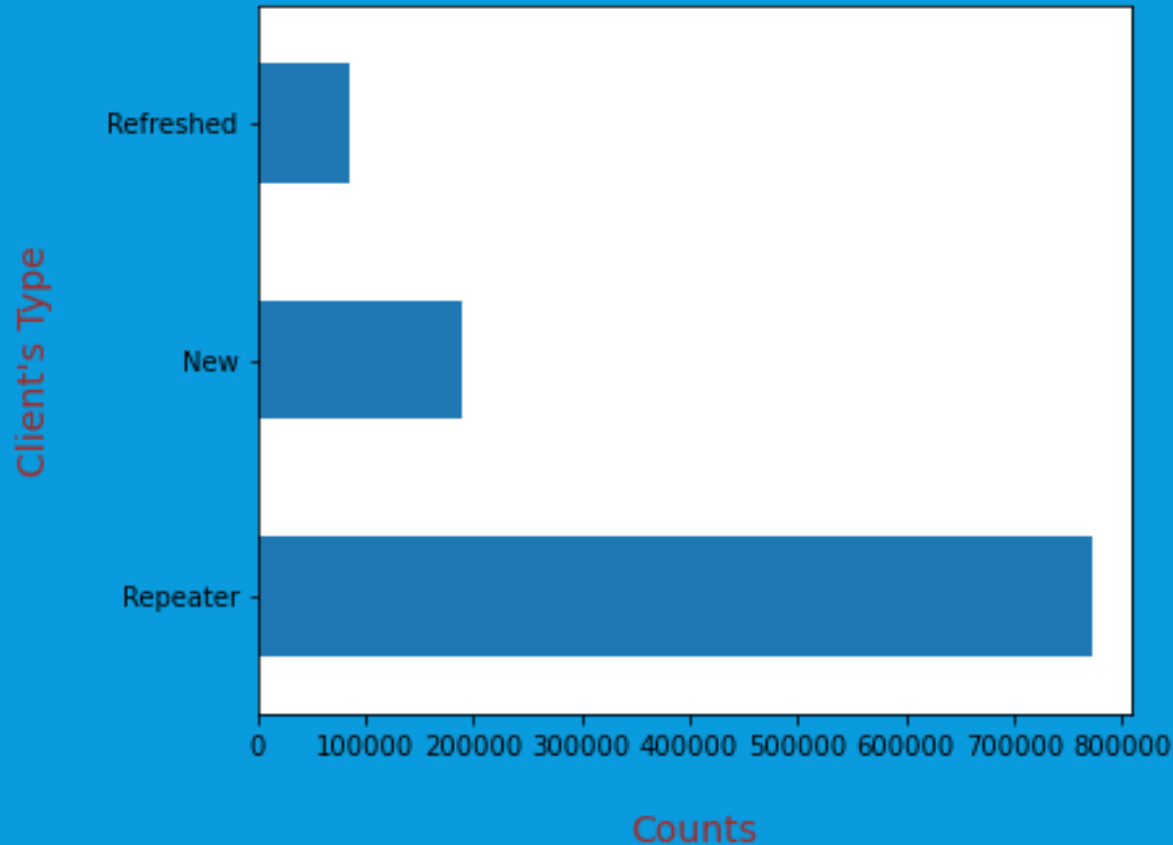
# ANALYSIS FOR PREVIOUS APPLICATION DATA



Inferences 1: According to the above Bar-Graph maximum Cash loan and Consumer loan having negligible difference and are highest in number, whereas 'Revolving loans' are very less.

# ANALYSIS FOR PREVIOUS APPLICATION DATA

Clinet Types for loan

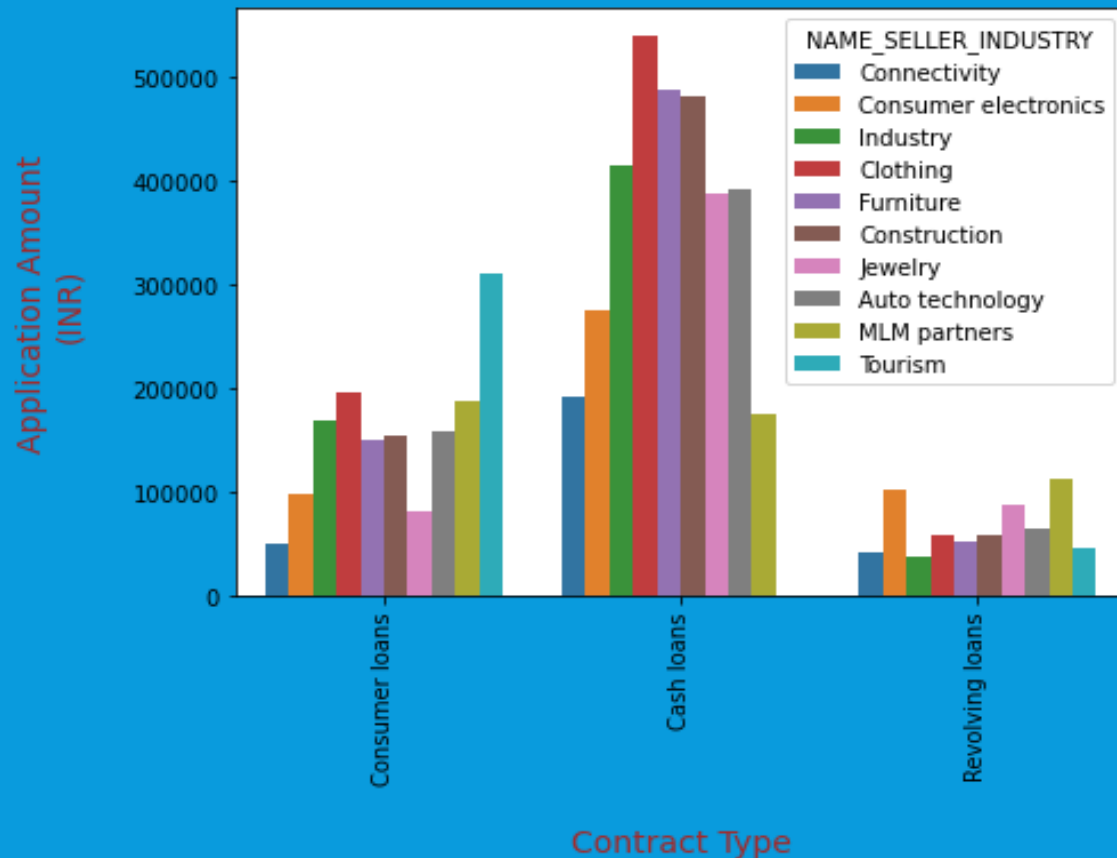


Inferences 1: According to the above Bar-Graph maximum client type is 'Repeater for loan, whereas 'Refreshed' are very less.



# ANALYSIS FOR PREVIOUS APPLICATION DATA

Contract Type Vs Application Amount in terms of Seller Industry



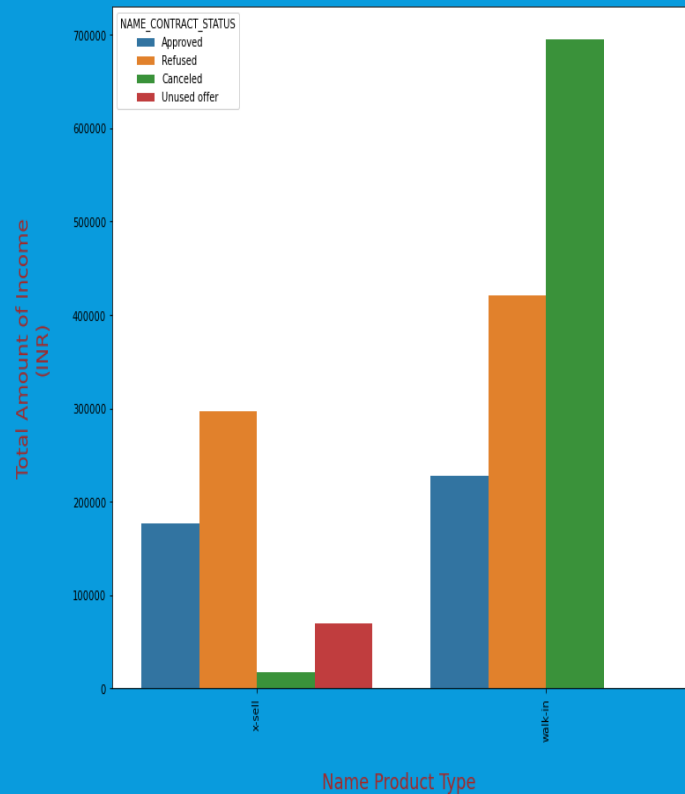
## Inferences

Inferences 1: From above Bar-Graph it is clear that among all the categories of Seller Industry Clothing has highest Application Amount for Cash loan Type Contract.

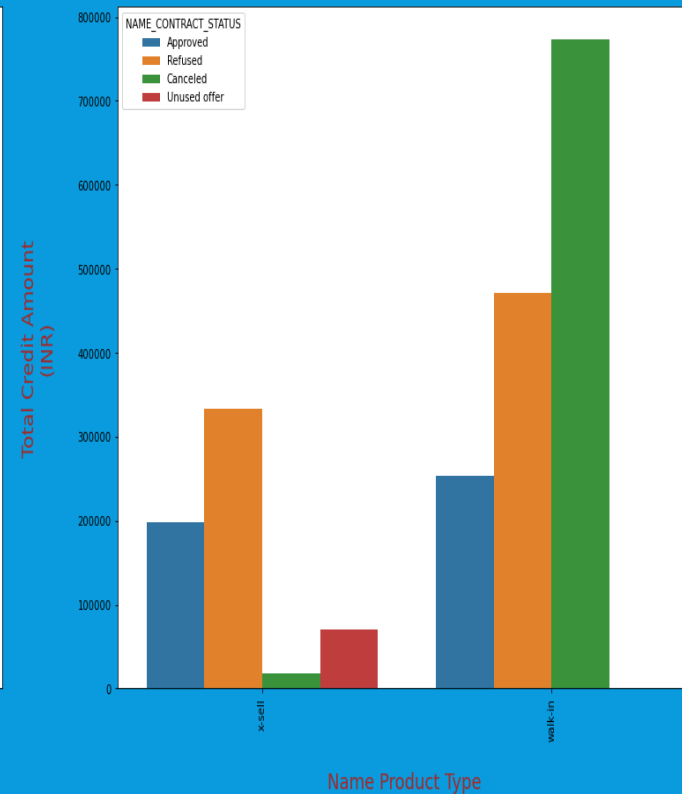
Inferences 2: Consumer loans for Tourism Seller Industry has highest Application Amount whereas it is less for Revolving loans.

# ANALYSIS FOR PREVIOUS APPLICATION DATA

Product Type Vs Application Amount



Product Type Vs Credit Amount



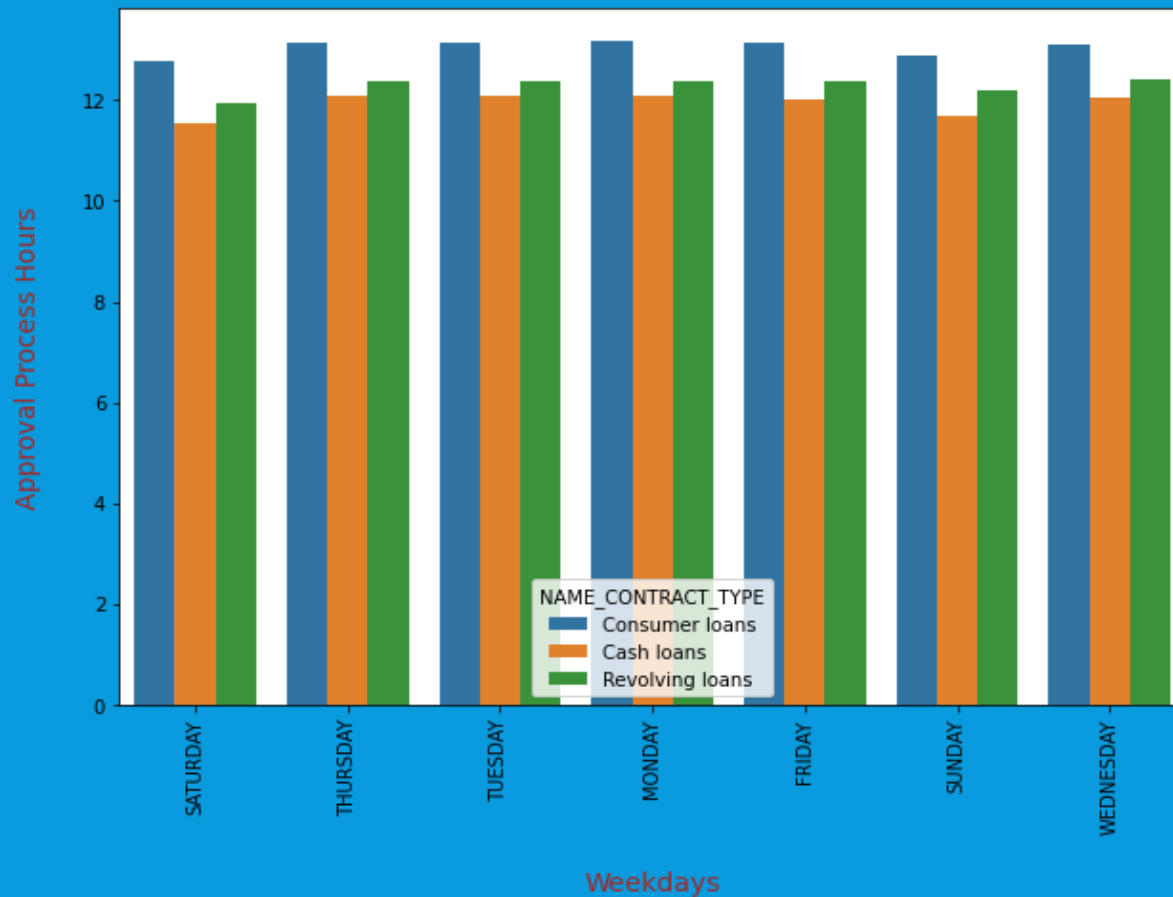
## Inferences

Inferences 1: Canceled Contracts for 'Walk-in' Product Type is highest for higher Income Amount and Credit Amount, whereas it is very less for X-Sell Product Type.

Inferences 2: Contract Status as a Unused Offer for 'Walk-in' Product Type is lowest or almost nil for both the amounts.

# ANALYSIS FOR PREVIOUS APPLICATION DATA

Weekdays Vs Hours in terms of Contract Type

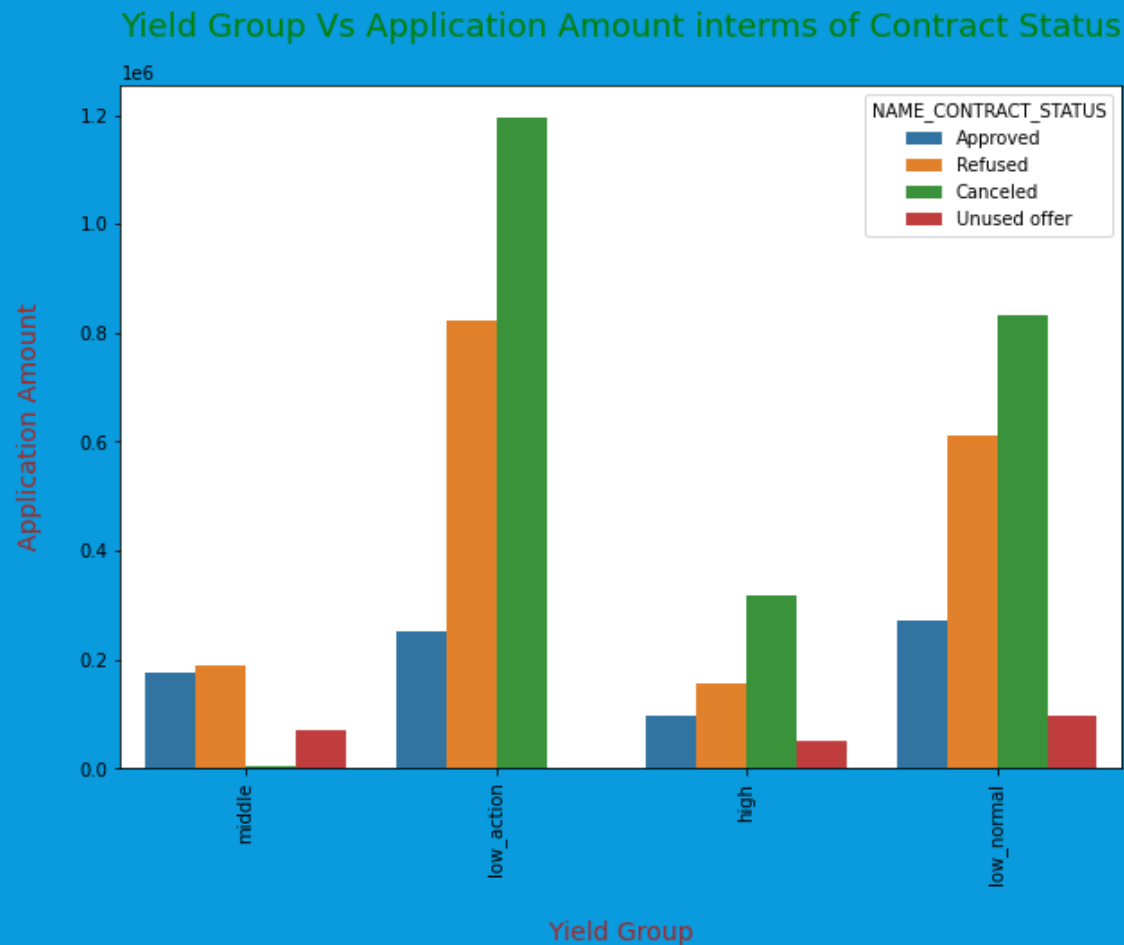


## Inferences

Inferences 1: Approval Process Hours are highest for each day for all kind of Contract Types whether it is Consumer loans, Cash loans or Revolving loans.

Inferences 2: Approval Process Hours are lowest for Cash loans for any day of the week.

# ANALYSIS FOR PREVIOUS APPLICATION DATA



## Inferences

Inferences 1: Cancelled Application are higher for Higher Application amount for the three different Yield Groups i.e. 'Low\_Action', 'Low\_Normal', and 'High' whereas for Middle type of Yield Group it is very-2 less.

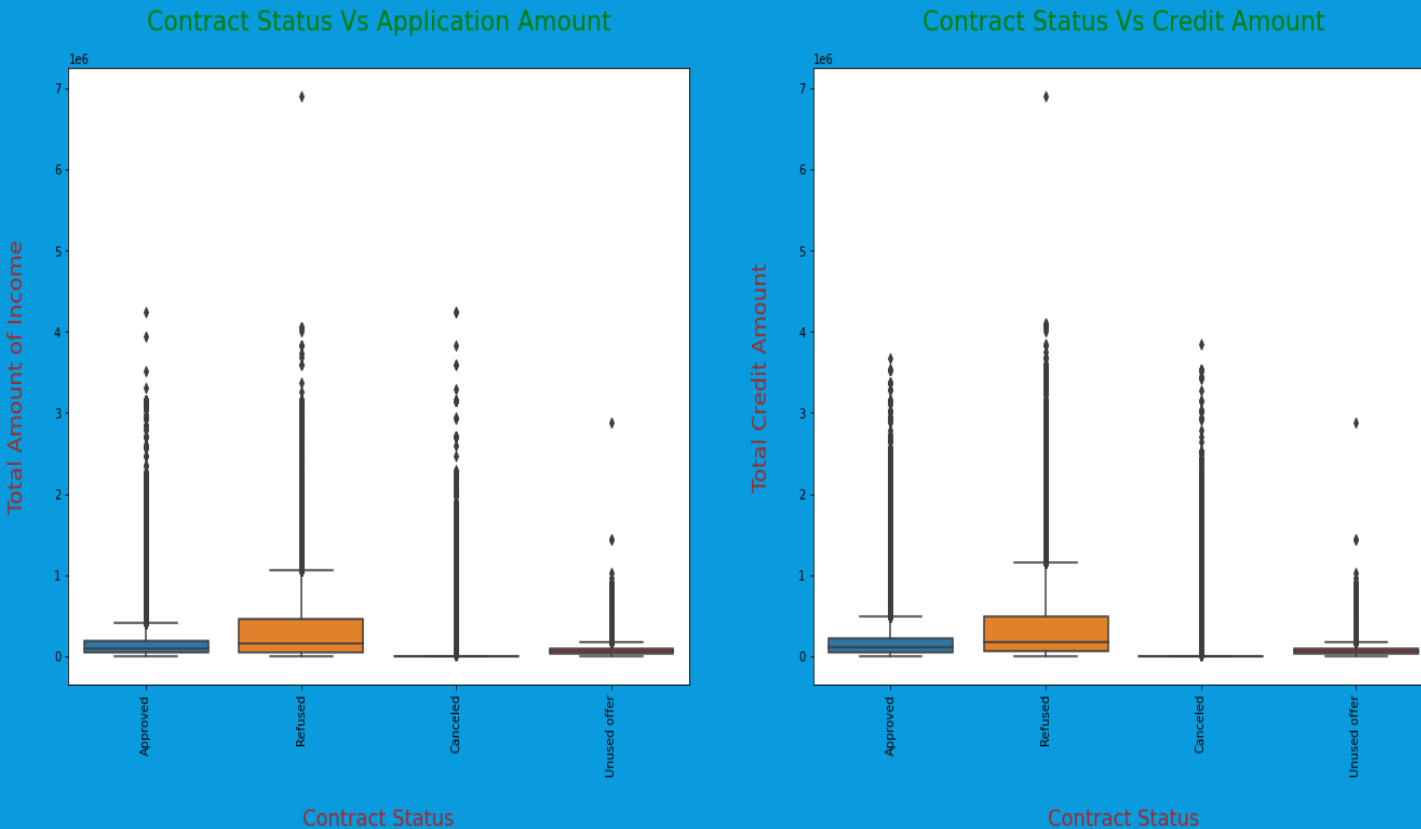
Inferences 2: It means Cancel rate is very less minimum amount in Middle Group.

# ANALYSIS FOR PREVIOUS APPLICATION DATA

## Inferences

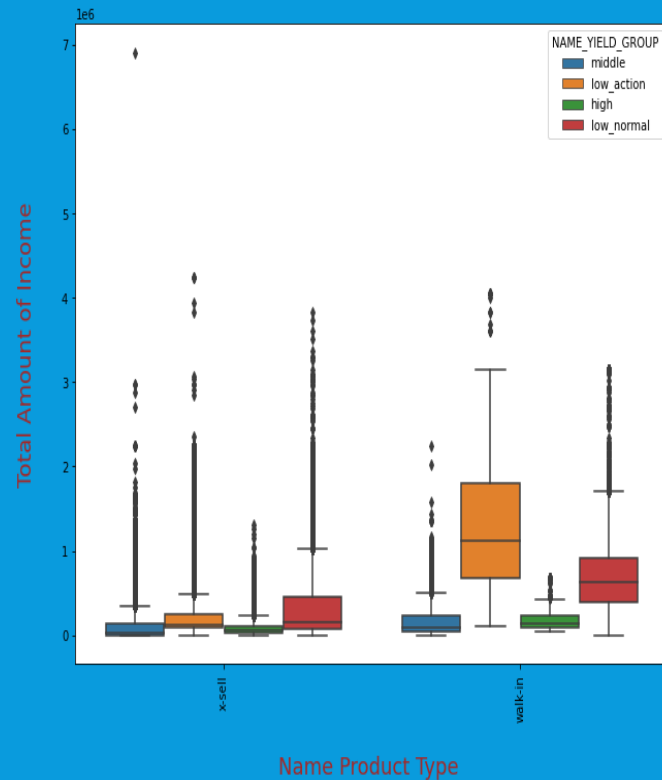
Inferences 1: Distribution of values for all kind of Contract Status has almost similar kind of distribution for both the Amounts whether it is Total Amount or Credit Amount.

Inferences 2: For Refused Contract distribution is unequal for both the amount whether it is Total Income Amount or Credit Amount.

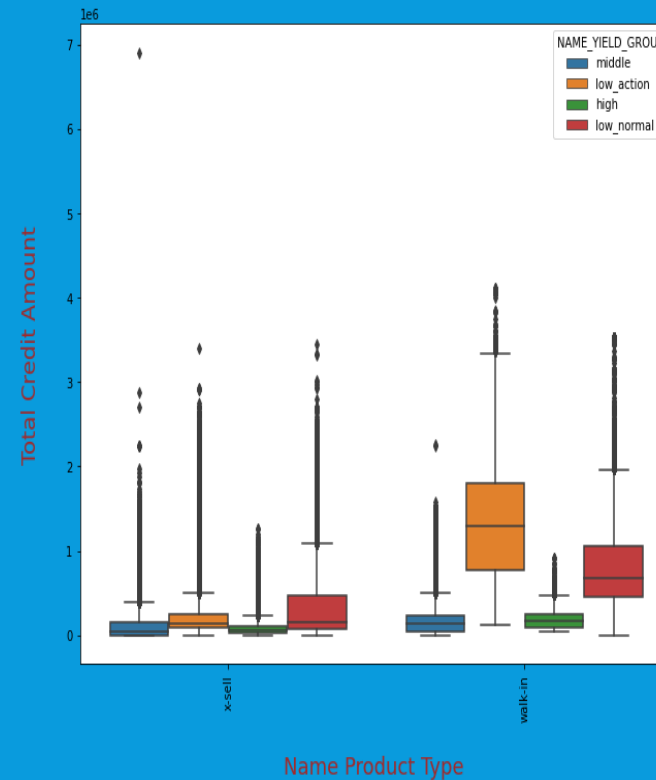


# ANALYSIS FOR PREVIOUS APPLICATION DATA

Product Type Vs Application Amount



Product Type Vs Credit Amount



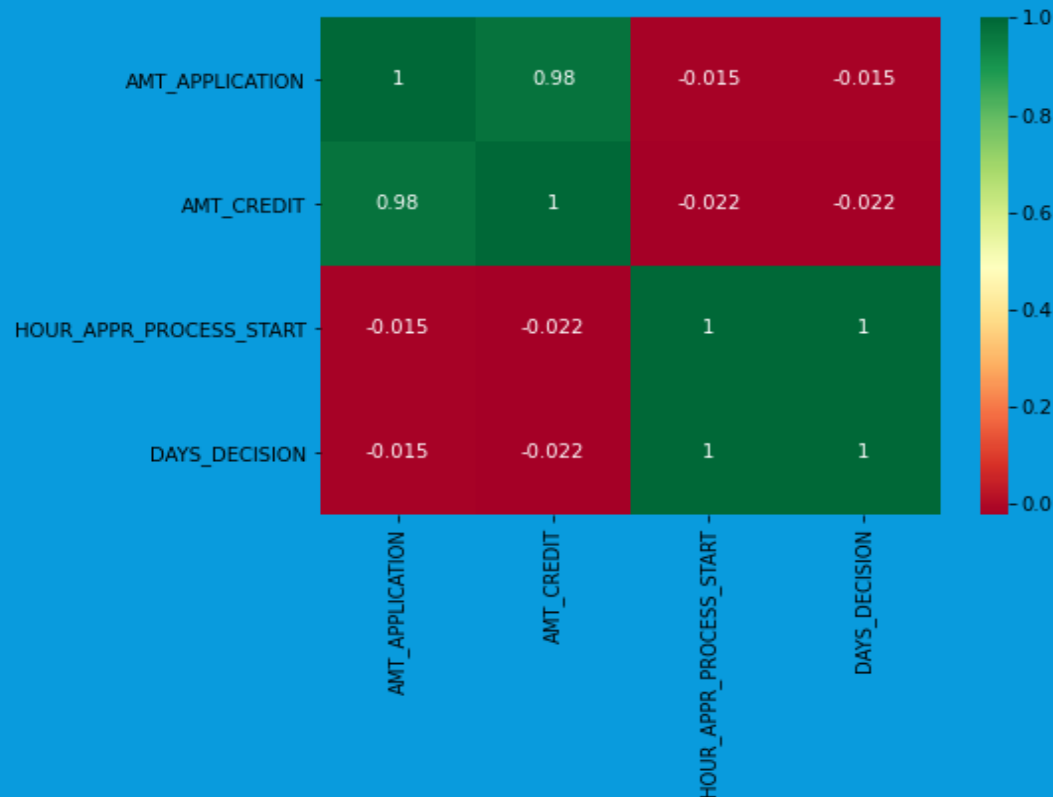
## Inferences

Inferences 1: Distribution of values for both the Product Types (X-Sell, Walk-in) has almost similar for all kind of Yield Groups whether it is Total Amount or Credit Amount.

Inferences 2: For low\_action Yield Group distribution is equal for product type 'Walk-in' whether it is Total income Amount or Total Credit Amount.

# ANALYSIS FOR PREVIOUS APPLICATION DATA

Correlation between Various Amounts & Durations



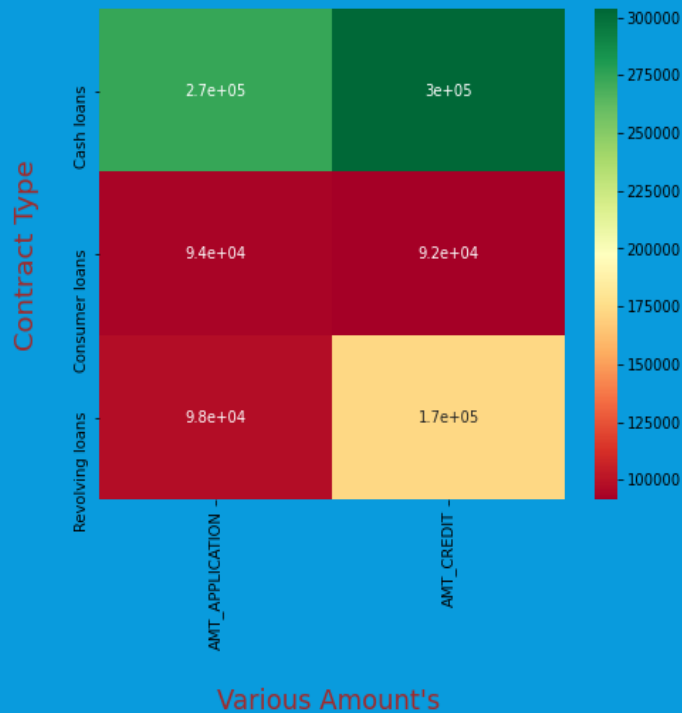
## Inferences

**Inference 1:** Application Amount is inversely proportional to the no. of Days Decision and required total Approval hours. It means higher Application Amount is sanctioned in less no. of days. and Application Process start in less no. of hours.

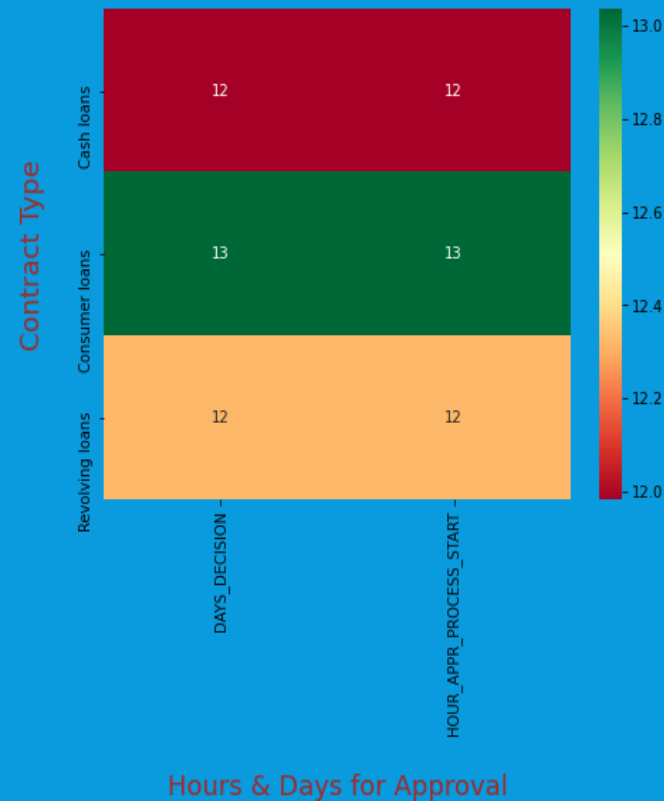
**Inference 2:** Credit Amount is also inversely proportional to the no. of Days Decision and required total Approval hours. It means higher Credited Amount is sanctioned in less no. of days. and Application Process start in less no. of hours.

# ANALYSIS FOR PREVIOUS APPLICATION DATA

Contract Type Vs Various Amounts



Contract Type Vs Approval Hours, Days



## Inferences

**Inference 1:** Cash loan are highest for Amount Credited and Application Amount.

**Inference 2:** Revolving loans has very less value for Amount Credited.

**Inference 3:** Revolving loans approved in very less no. of days and it's Process also started in minimum hours, whereas Cash loans are taking more no. of days to approve and more hours to start the process.

**Inference 4:** Overall Revolving loans are taking less hours, minimum no. of days in decision but their loan amount and Credited amount is also less.



THANKS VERY MUCH FOR YOUR  
PATIENCE!