

Presentation

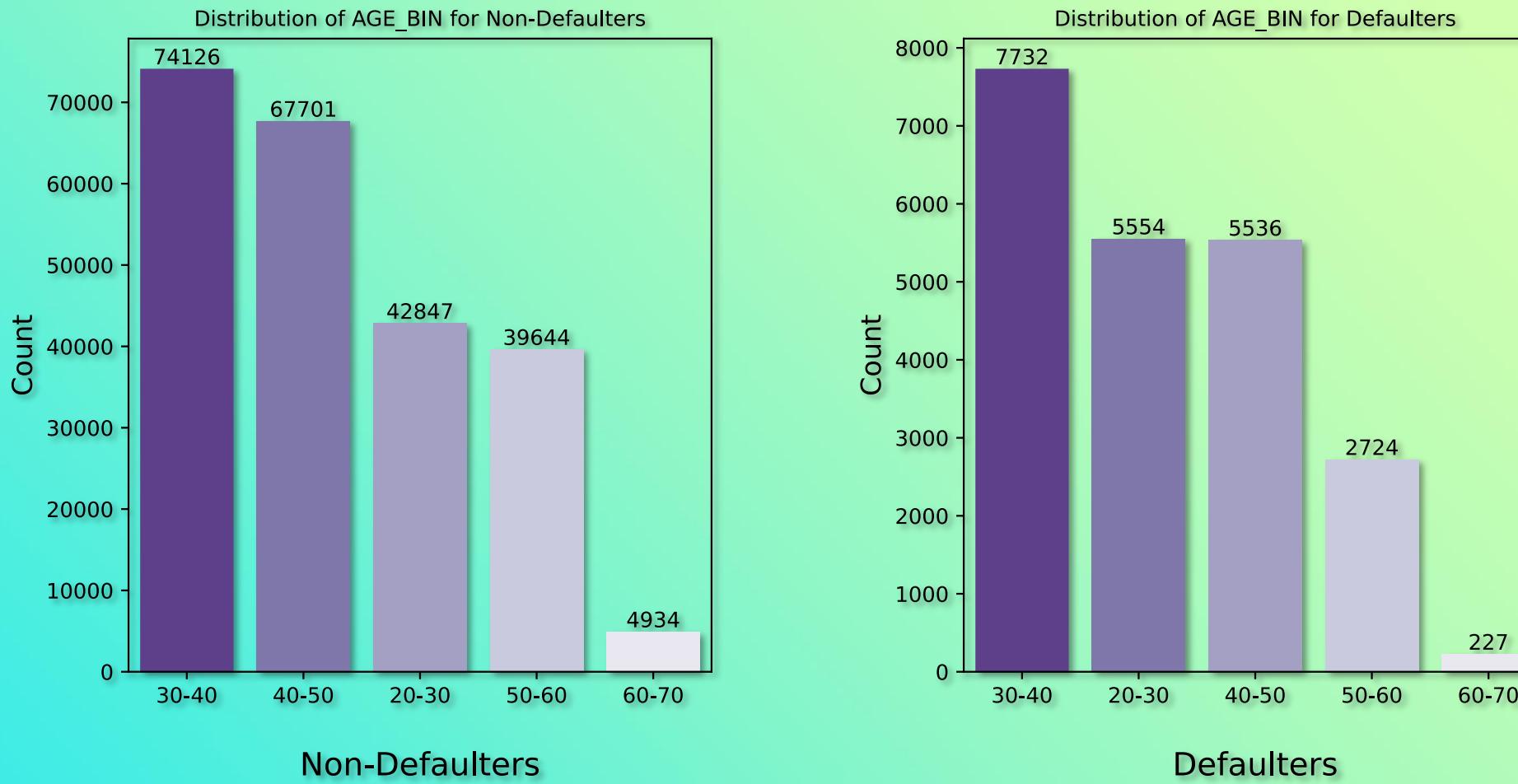
On Credit EDA Case Study

BY
DeePansh Jha

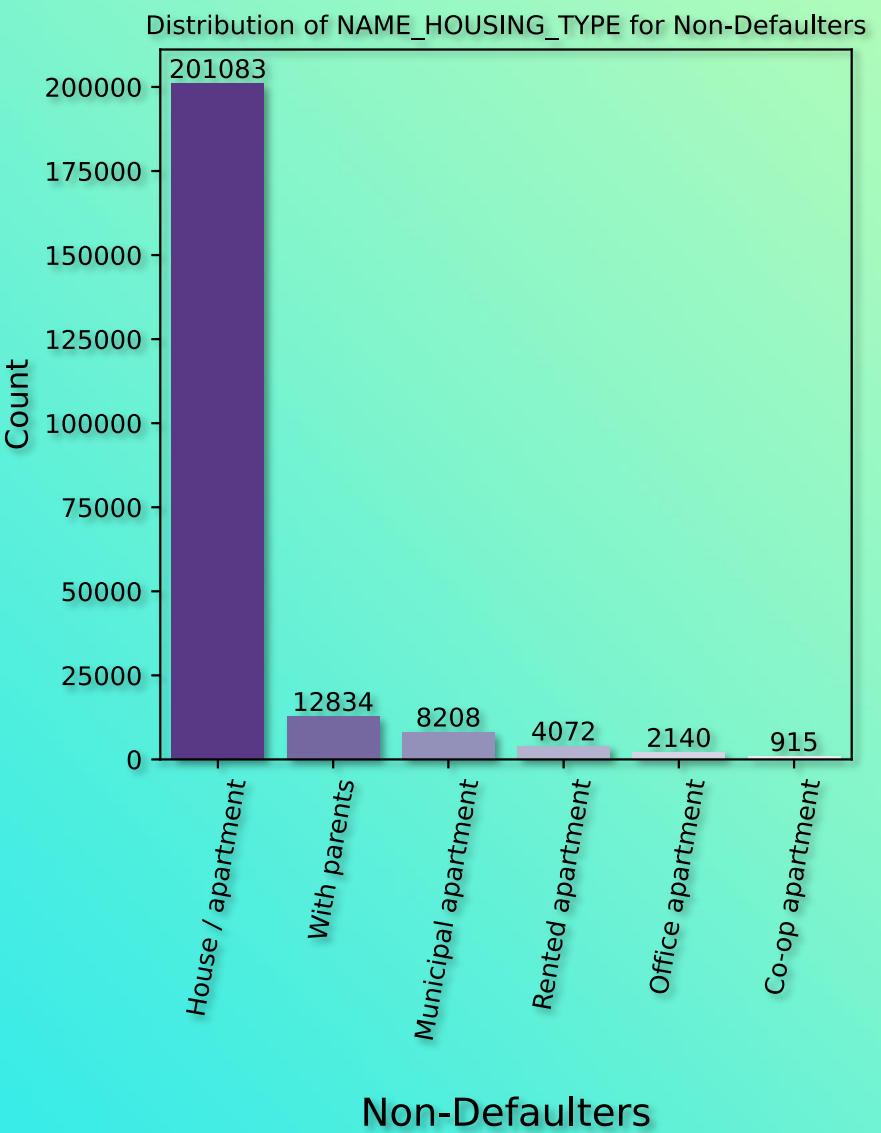
Steps

1. Data Understanding And Sourcing.
2. Check Data Quality And Data Cleaning.
3. Binning.
4. Check For Data Imbalance.
5. Data Analysis With Respect To 'Target' Column.
6. Merging Application Data With Previous Application.
7. Data Analysis With Respect To 'Contract Status' Column.
8. Recommendation.

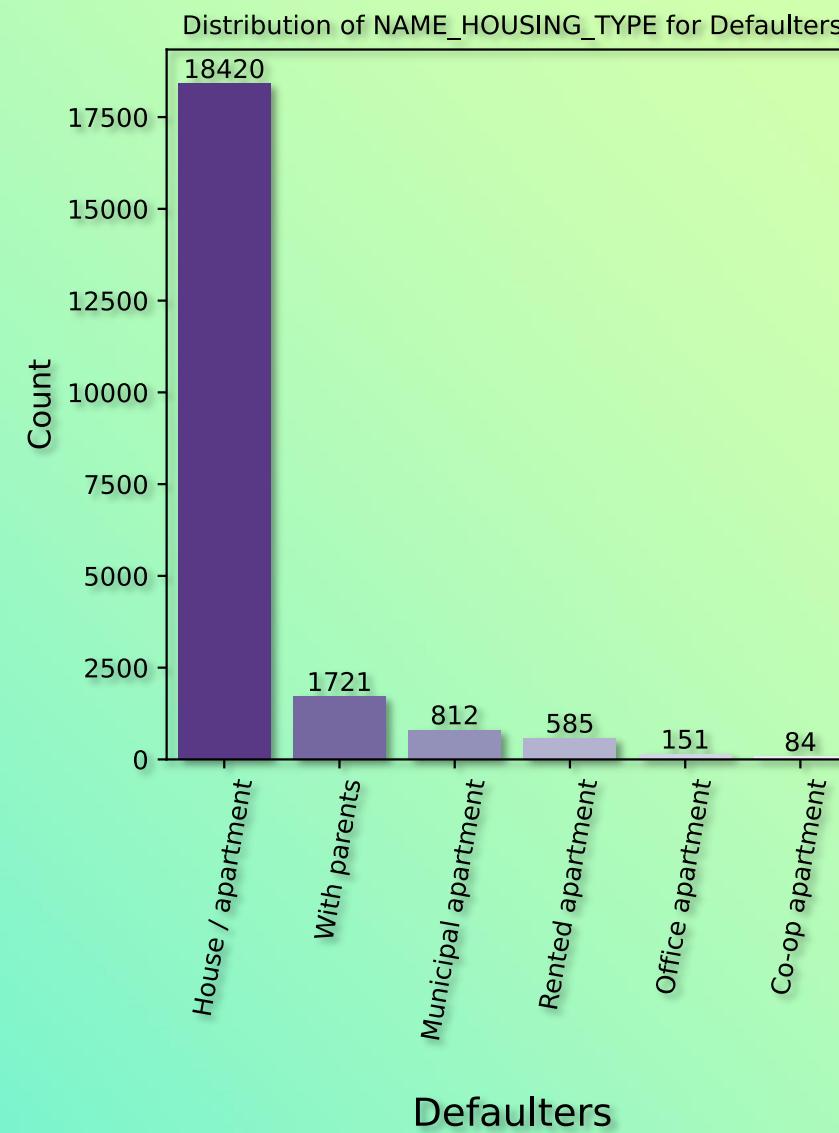
Univariate
Analysis



- There Is High Risk Of People With Age Group 20-30 To Be Defaulters.
- 50-70 Age Group People Are Less Defaulters.
- Age Group 30-40 Applied Most.

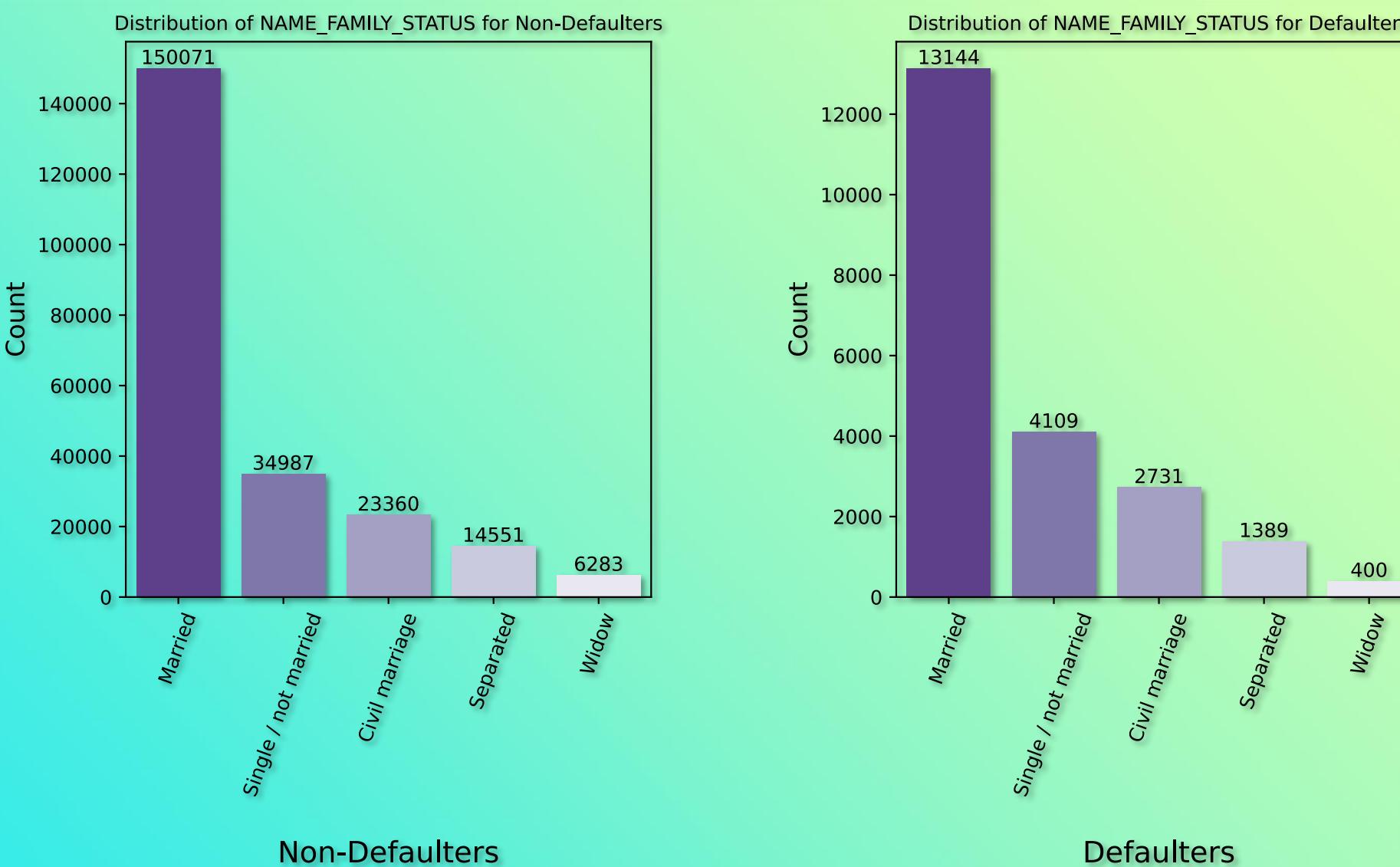


Non-Defaulters

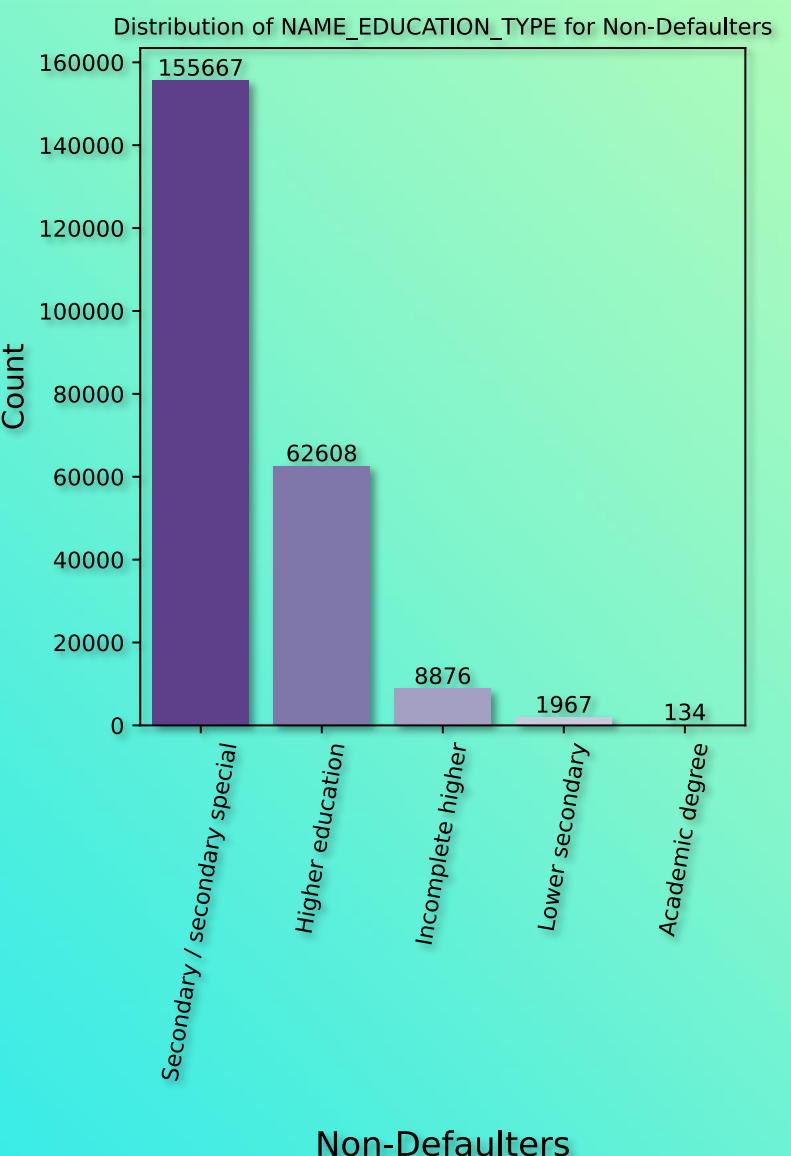


Defaulters

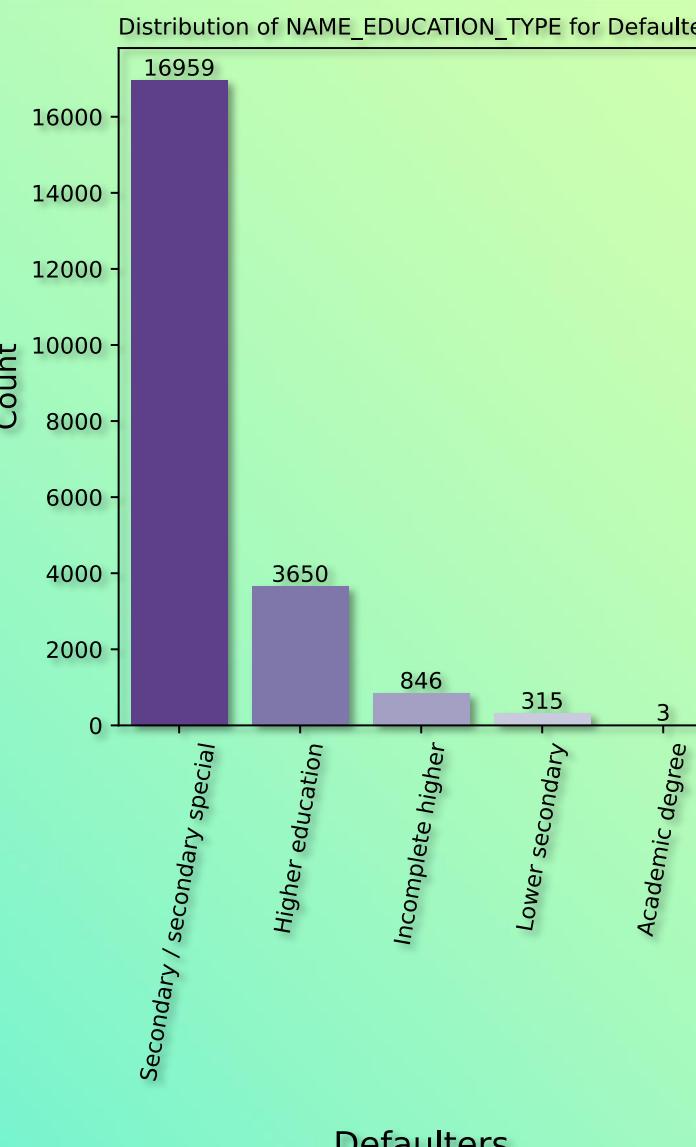
- People Can Be Defaulters Despite Having House
- People With Co-op Apartment Are Less Defaulters



- Most People Who Applied Are Married.
- Widows Are Less Defaulters, Singles Default More.

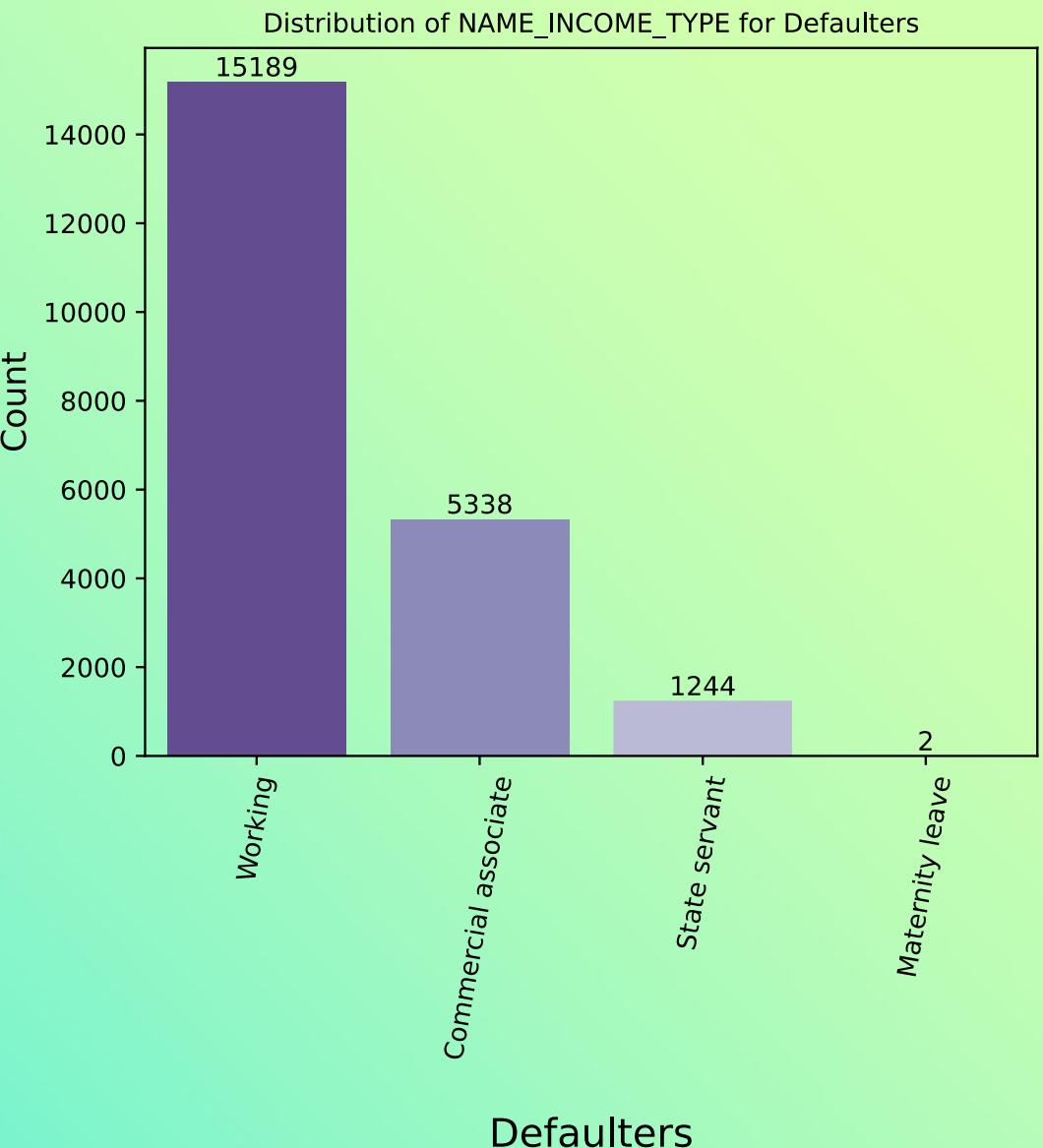
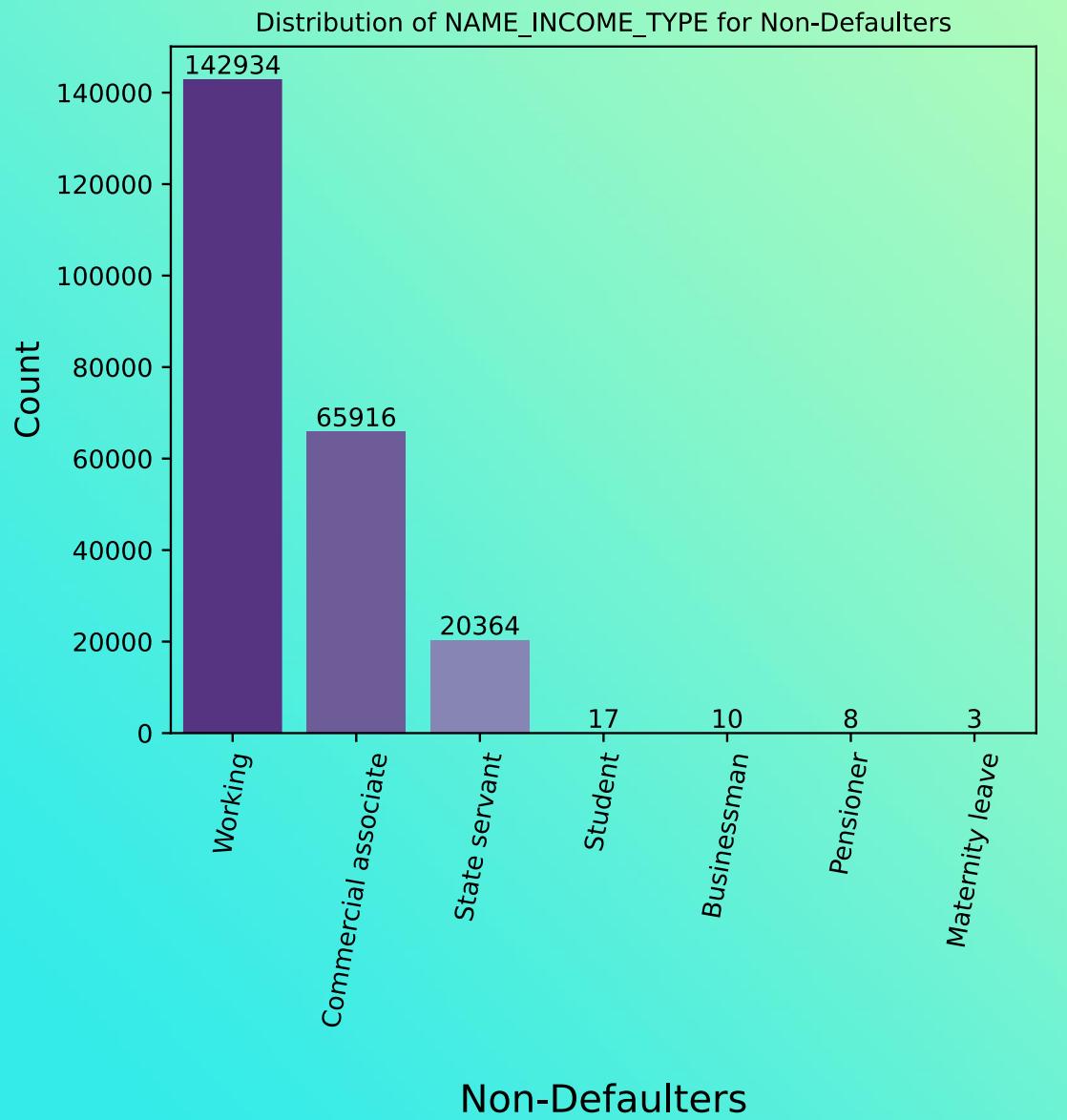


Non-Defaulters

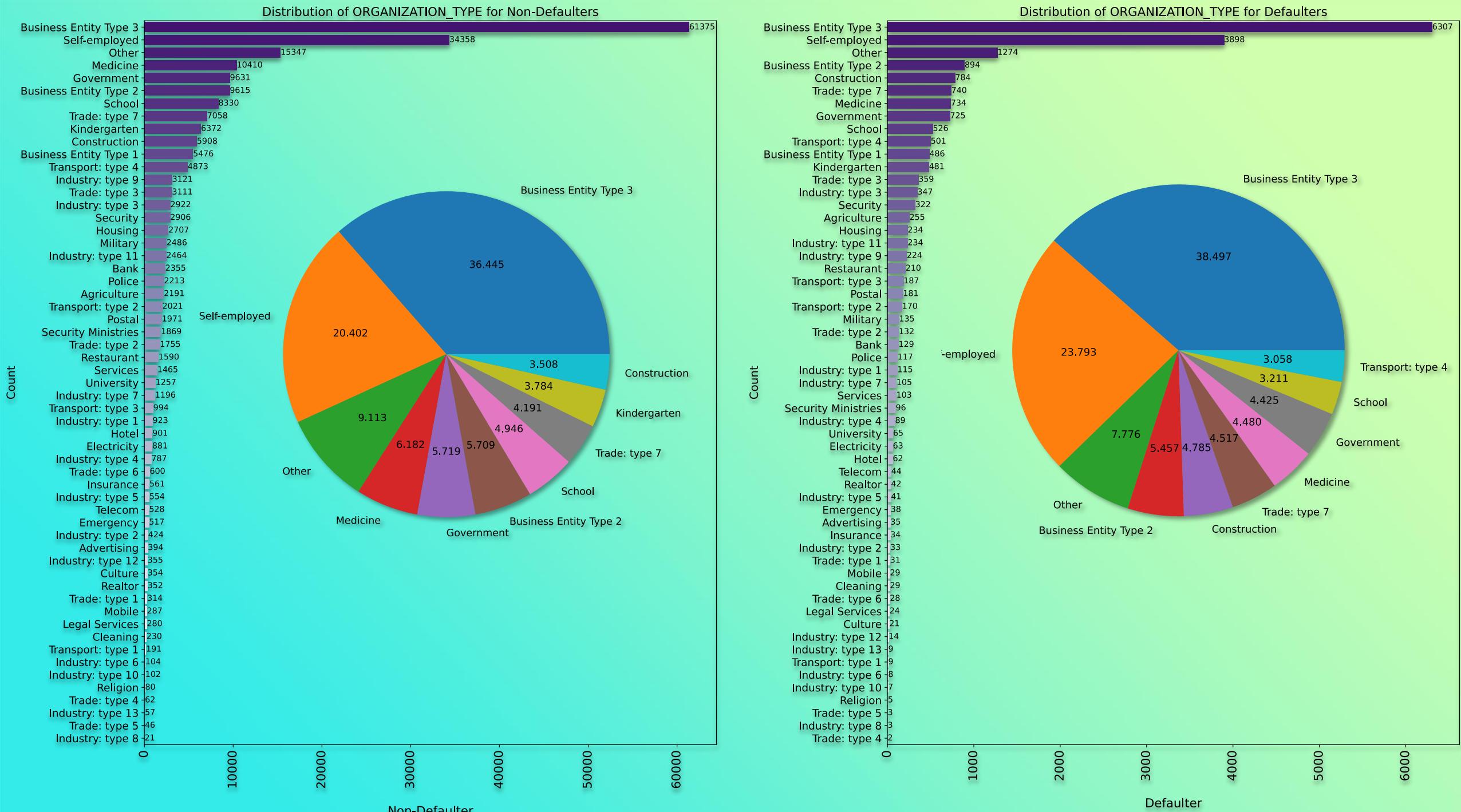


Defaulters

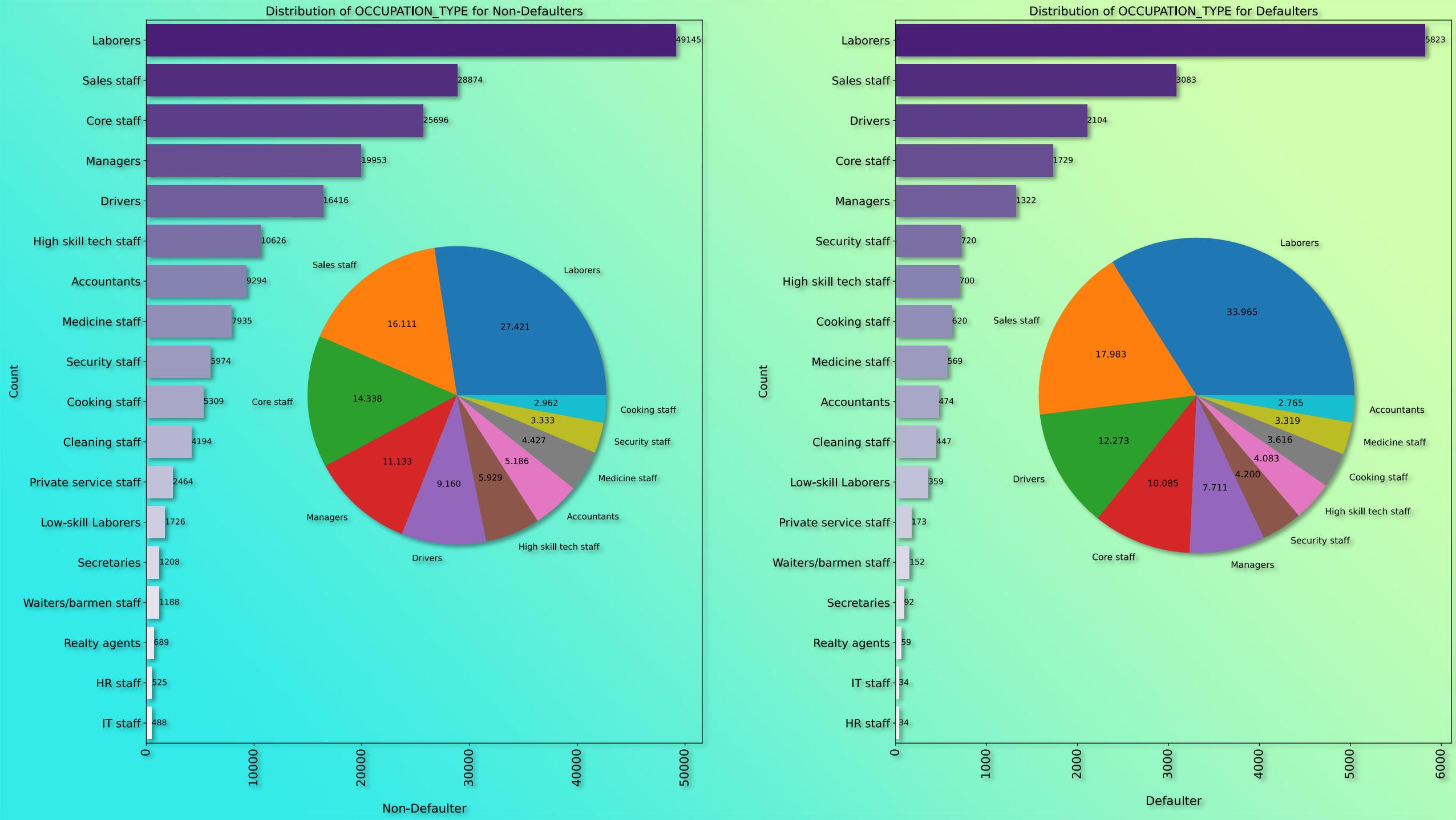
- Most People Applied Have Secondary Education.
- People With Higher Education Are Less Defaulters.



- Most People Applied Are Working Professional.
- People Who Are State Servant Are Less Defaulters.

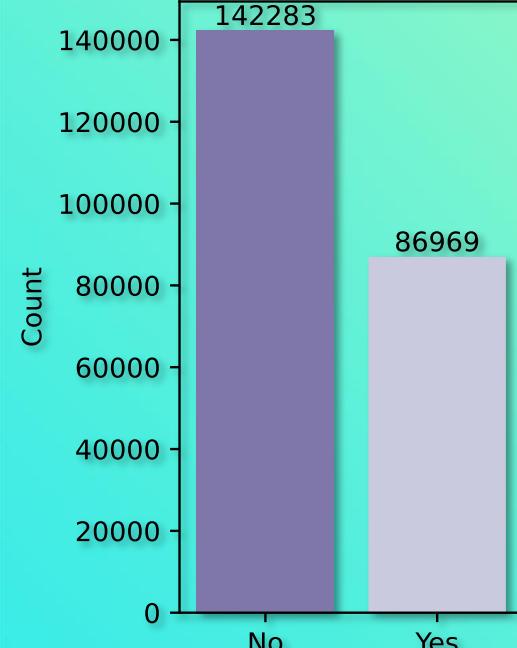


- Points to be concluded from the graph on the left (Non-Defaulters).
 - Clients which have applied for Loan are from most of the organization type ‘Business entity Type 3’ , ‘Self employed’ , ‘Other’ , ‘Medicine’ and ‘Government’.
 - Less clients are from ‘Industry type 8’,‘trade type 5’, ‘Industry type 13’, ‘trade type 4’, ‘religion’.
-
- Points to be concluded from the graph on the right (Defaulters).
 - Clients which have applied for Loan are from most of the organization type ‘Business entity Type 3’ , ‘Self employed’ , ‘Other’ , ‘Business entity Type 2’ and ‘Construction’.
 - Less clients are from ‘Trade type 4’,‘Industry type 8’,‘Trade type 5’, ‘religion’ ,‘Industry type 10, type 6’.



- Points to be concluded from the graph on the left (Non-Defaulters).
 - Clients which have applied for Loan are from most of the organization type ‘Laborers’ , ‘Sales Staff’ , ‘Core Staff’ , ‘Managers’ and ‘Drivers’.
 - Less clients are from ‘IT Staff’ , ‘HR Staff’ , ‘Realty Agents’ , ‘Waiters’ , ‘Secretaries’.
-
- Points to be concluded from the graph on the right (Defaulters).
 - Clients which have applied for Loan are from most of the organization type ‘Laborers’ , ‘Sales Staff’ , ‘Drivers’ , ‘Core Staff’ and ‘Managers’.
 - Less clients are from ‘HR Staff’ , ‘IT Staff’ , ‘Realty Agents’ , ‘Secretaries’ , ‘Waiters’.

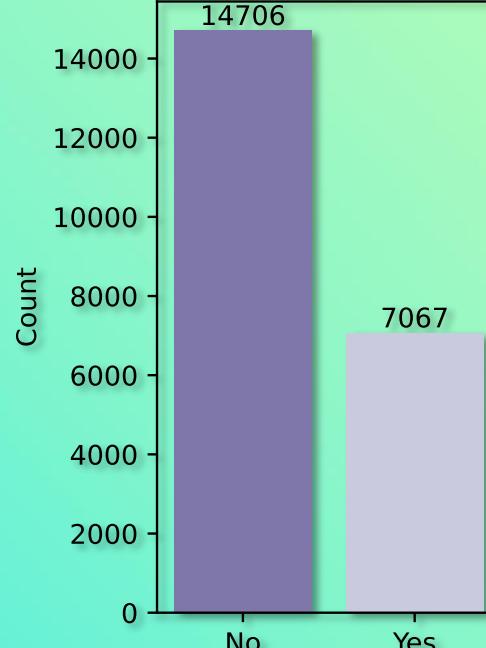
Distribution of OWN_CAR for Non-Defaulters



Non-Defaulters

- People Who Doesn't Own Realty Have Similar Ratio For Both Target.
- Most People Who Applied Own Realty.

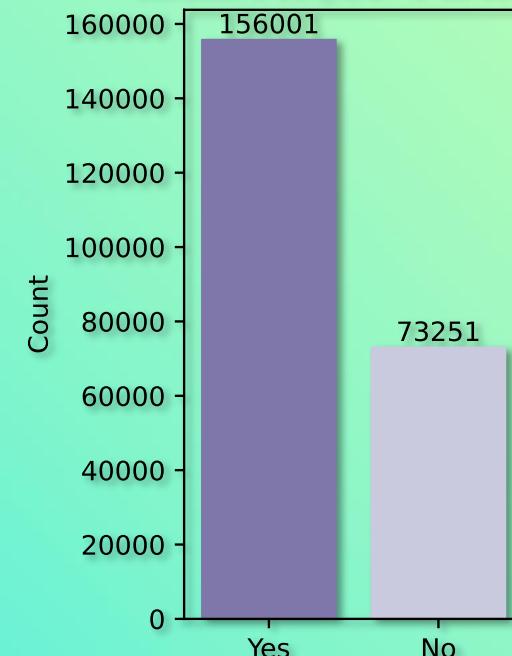
Distribution of OWN_CAR for Defaulters



Defaulters

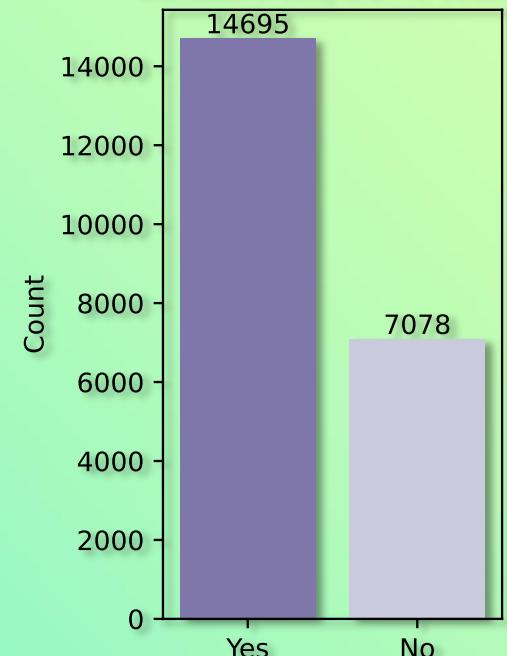
- People Who Own Car Defaults Less.
- Most People Who Applied Doesn't Own Car.

Distribution of OWN_REALTY for Non-Defaulters

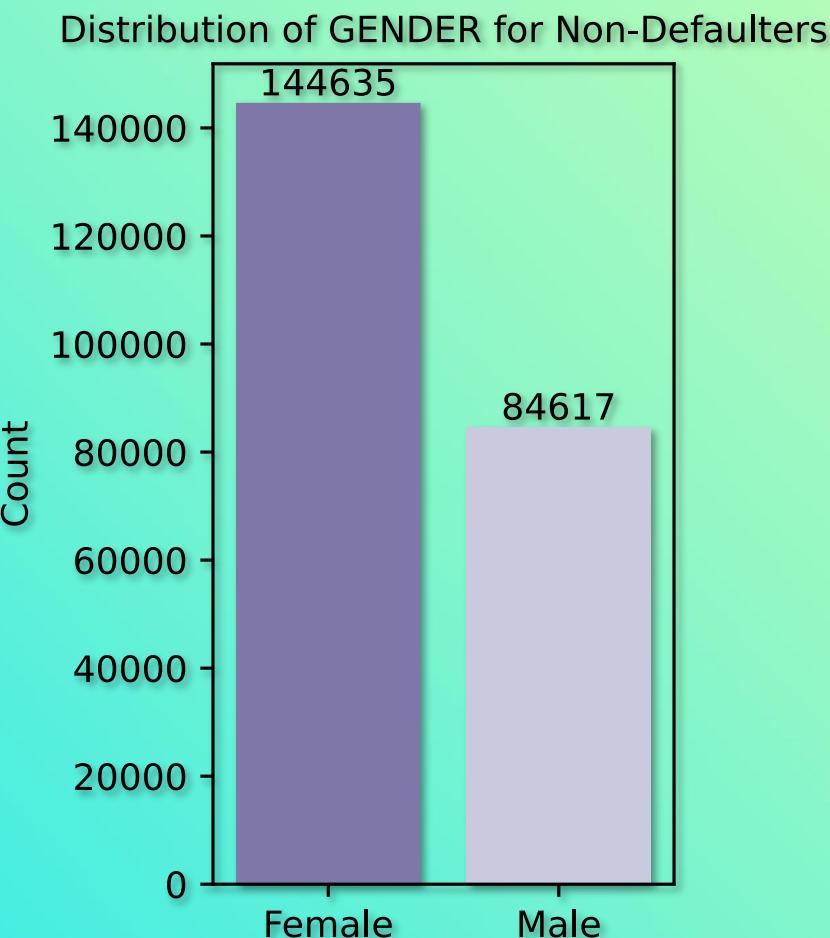


Non-Defaulters

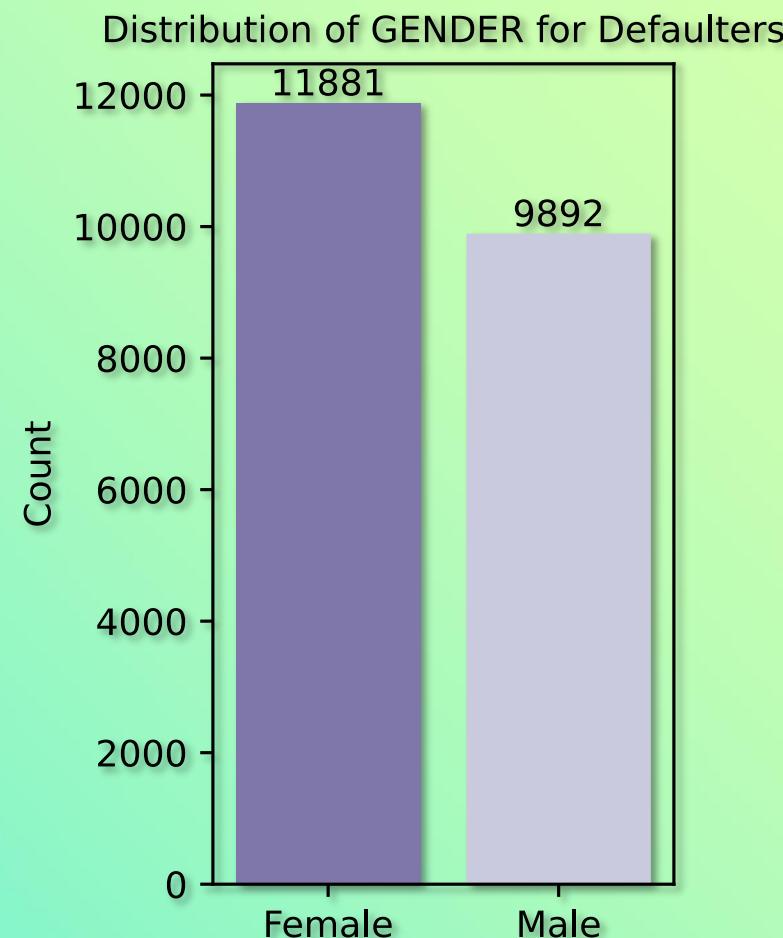
Distribution of OWN_REALTY for Defaulters



Defaulters

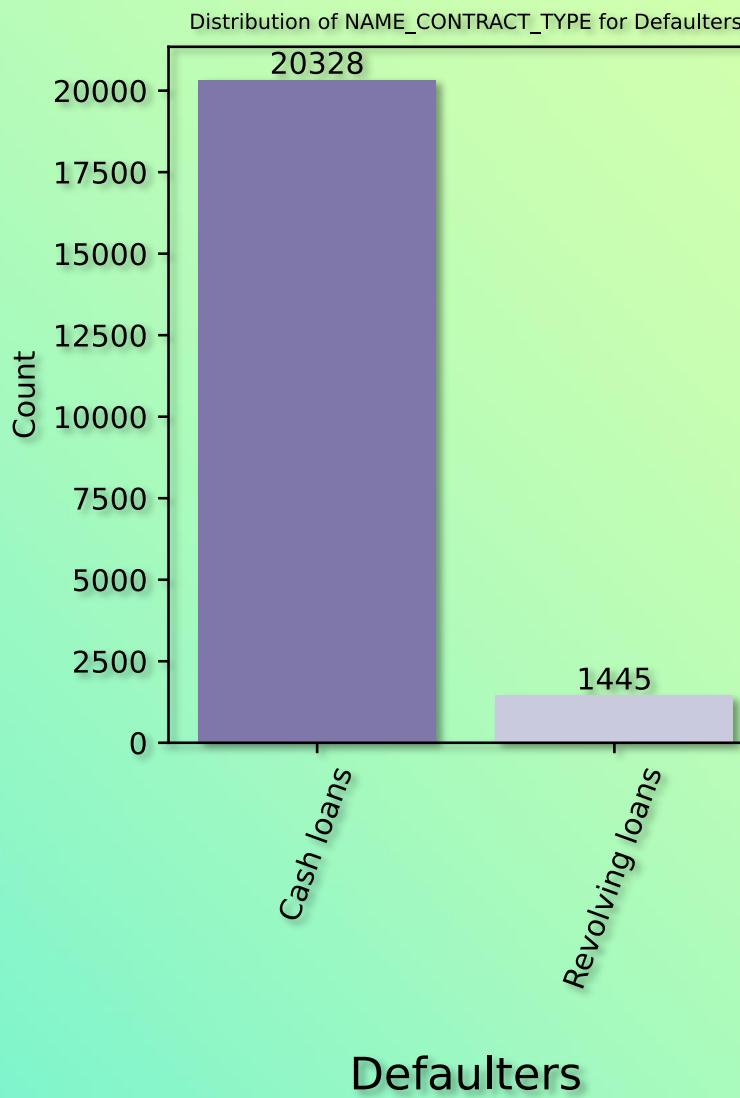
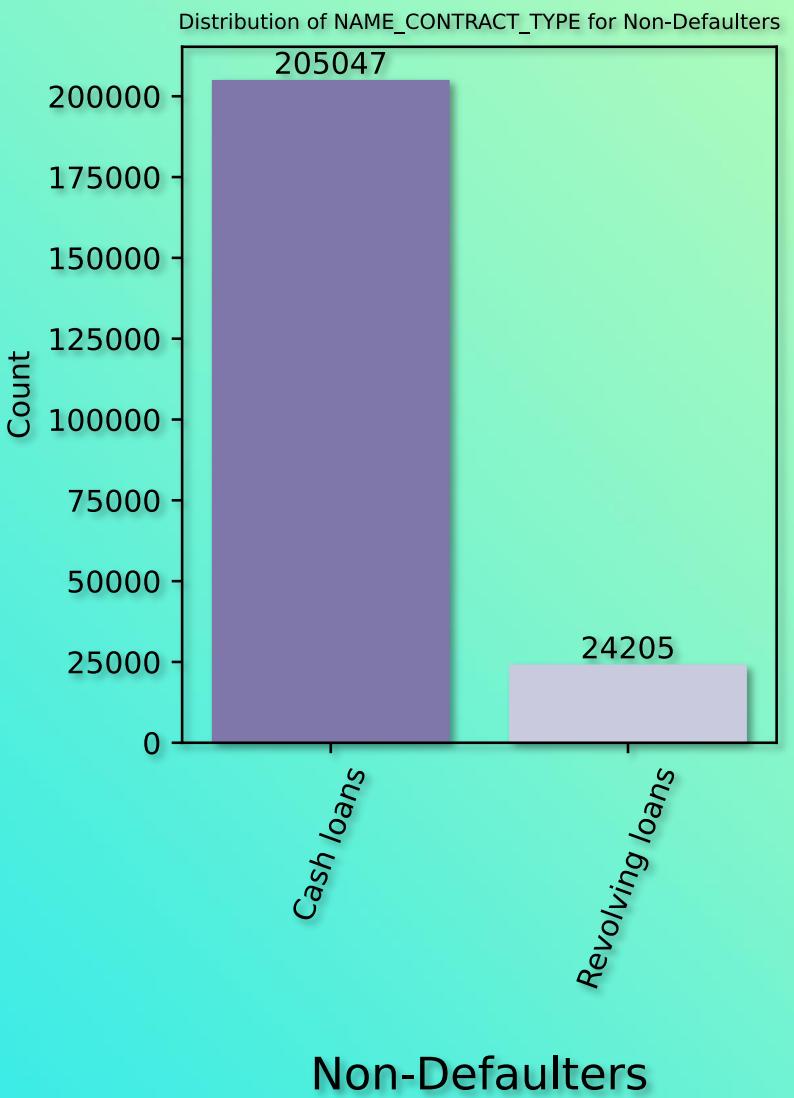


Non-Defaulters



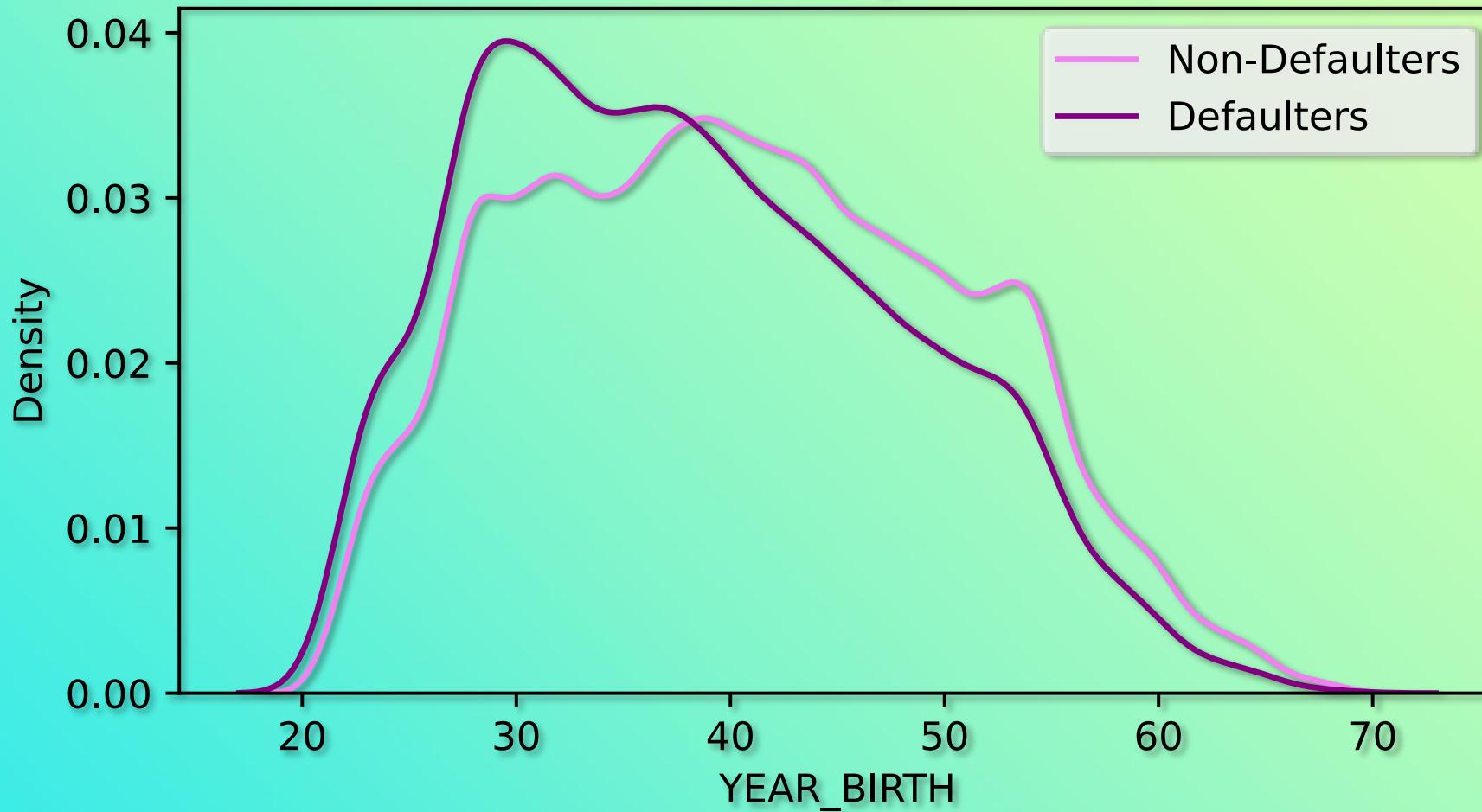
Defaulters

- Most Defaulters Are Males.
- Clients With Gender Female Applied Most.



- Less Defaulters Are From Revolving Loans.
- People With Cash Loans Applied Most.

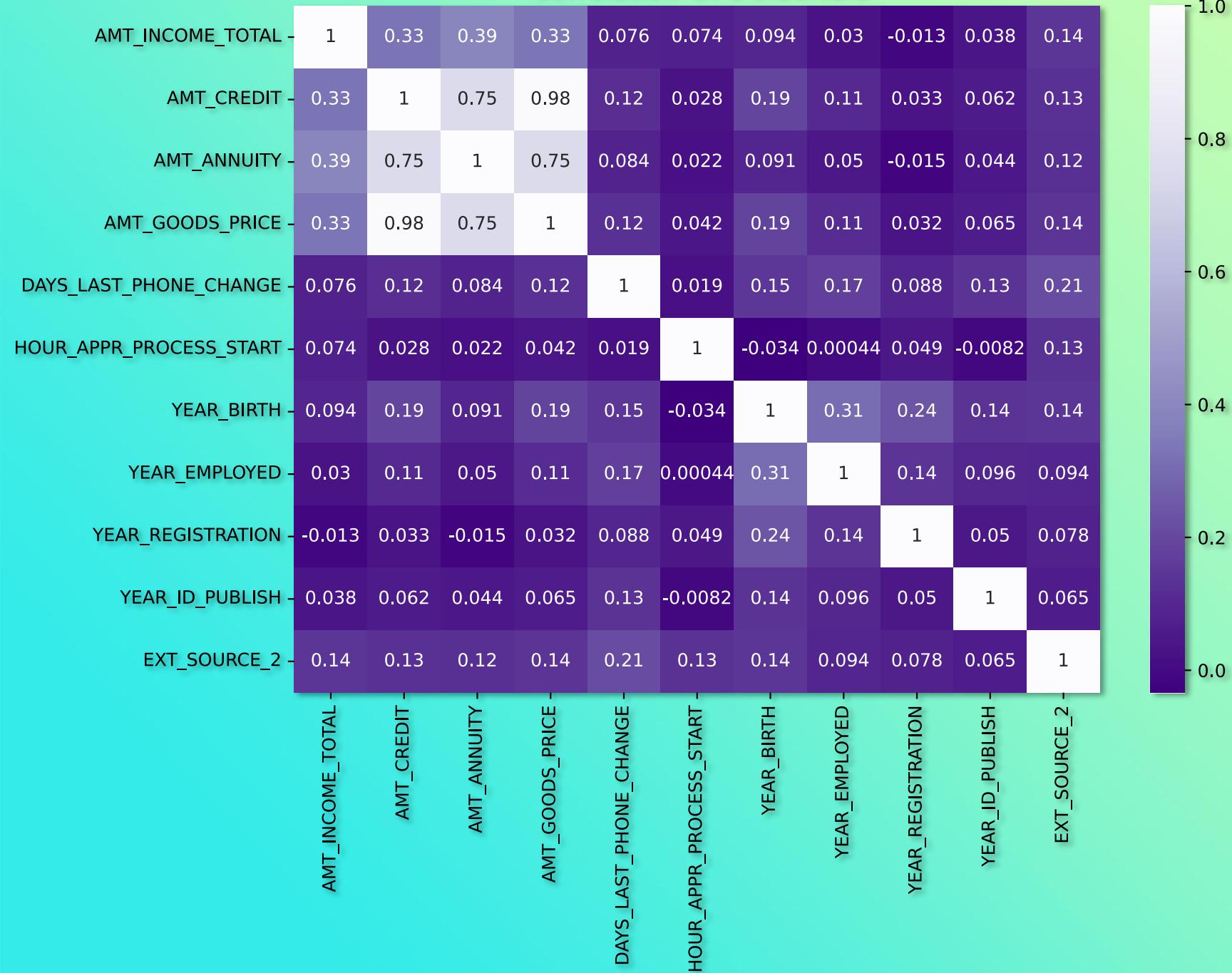
Distribution of YEAR_BIRTH In Target



- People With Age 28 To 38 Take More Loans And There Are More Defaulters With Age Around 28 To 30.

Correlation
TARGET Variable.

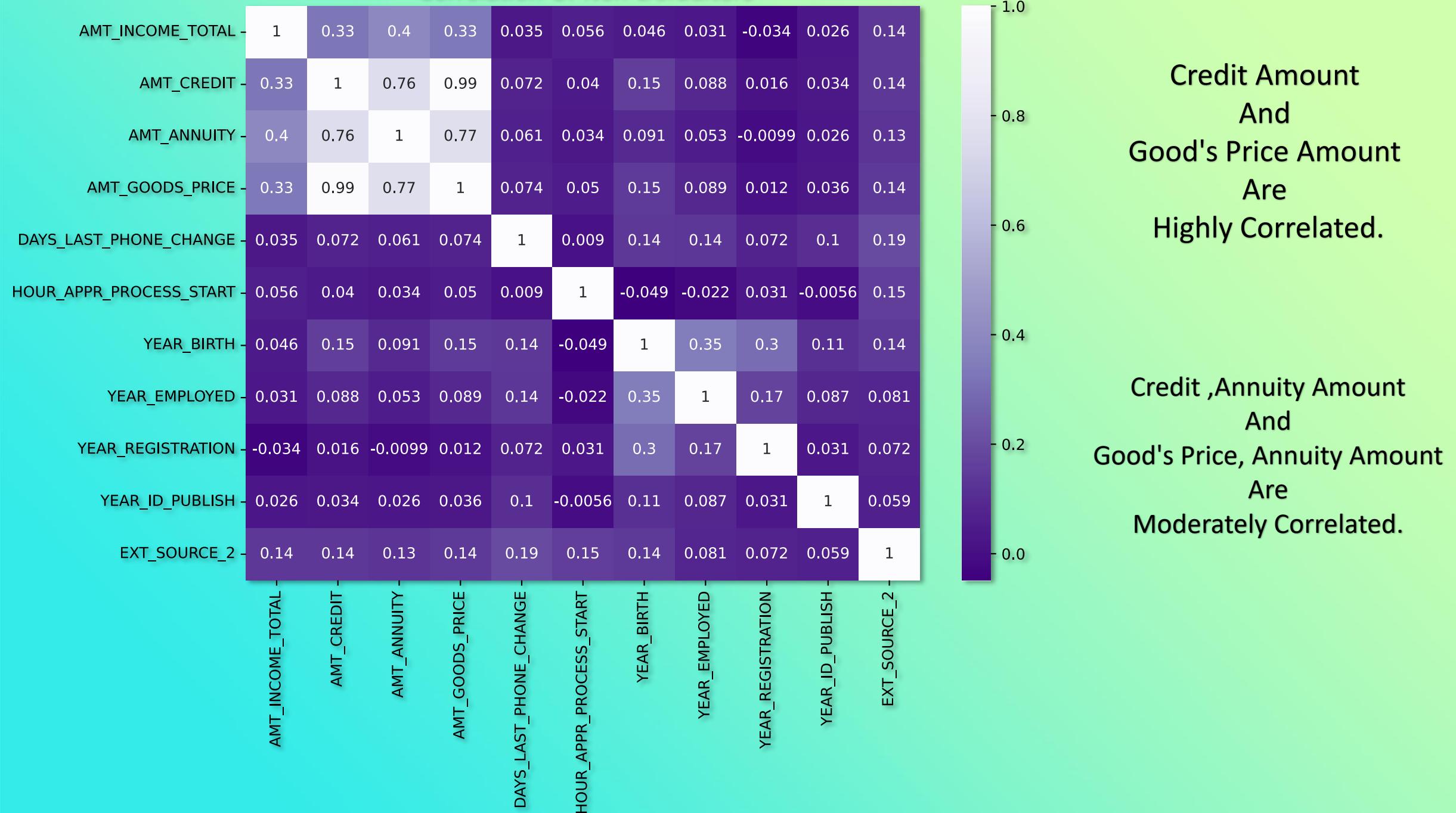
Correlation Of Defaulters



Credit Amount
And
Good's Price Amount
Are
Highly Correlated.

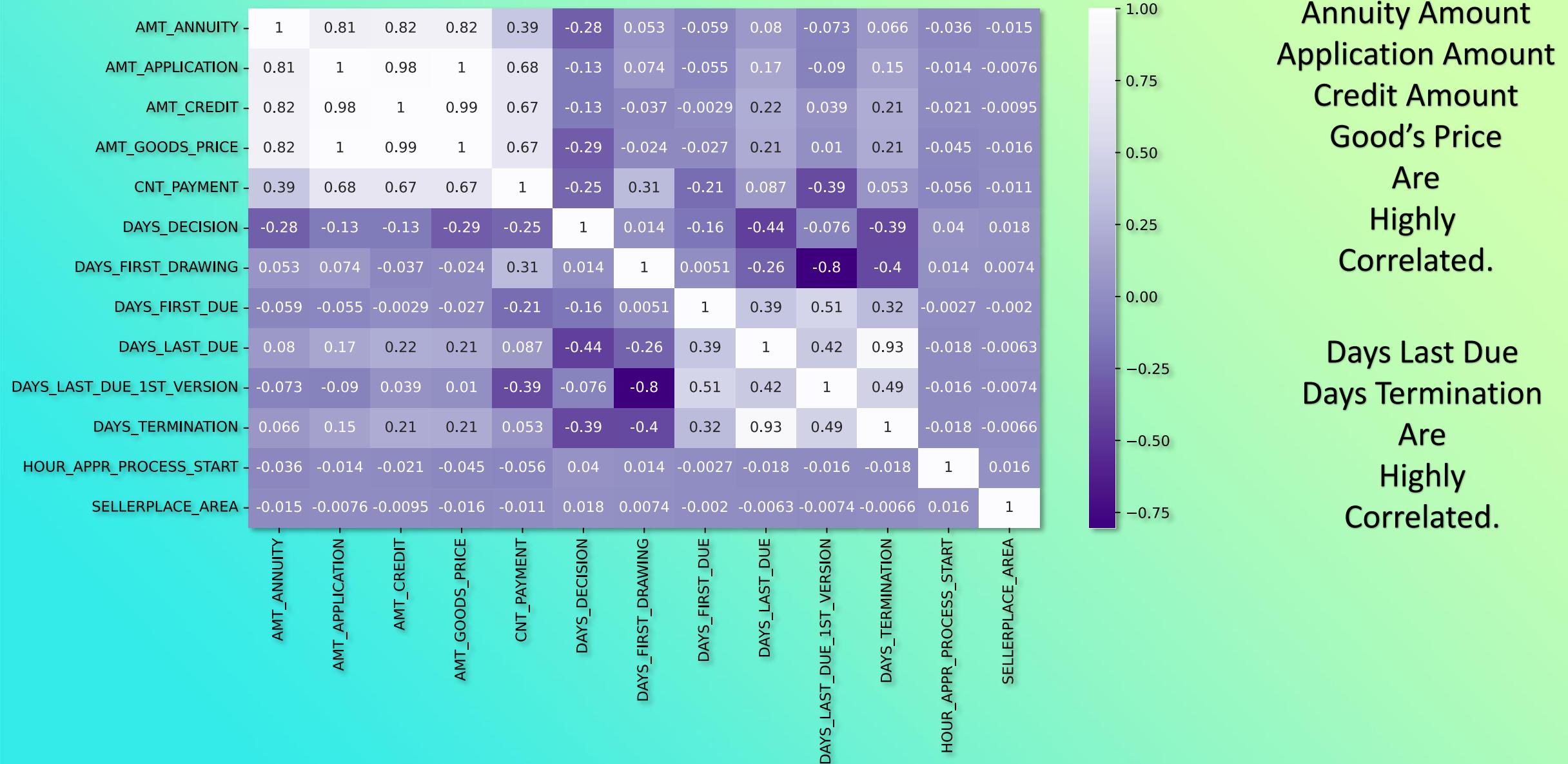
Credit ,Annuity Amount
And
Good's Price, Annuity Amount
Are
Moderately Correlated.

Correlation Of Non-Defaulters



Correlation
Previous APPLICATION.

Correlation Of Previous Application

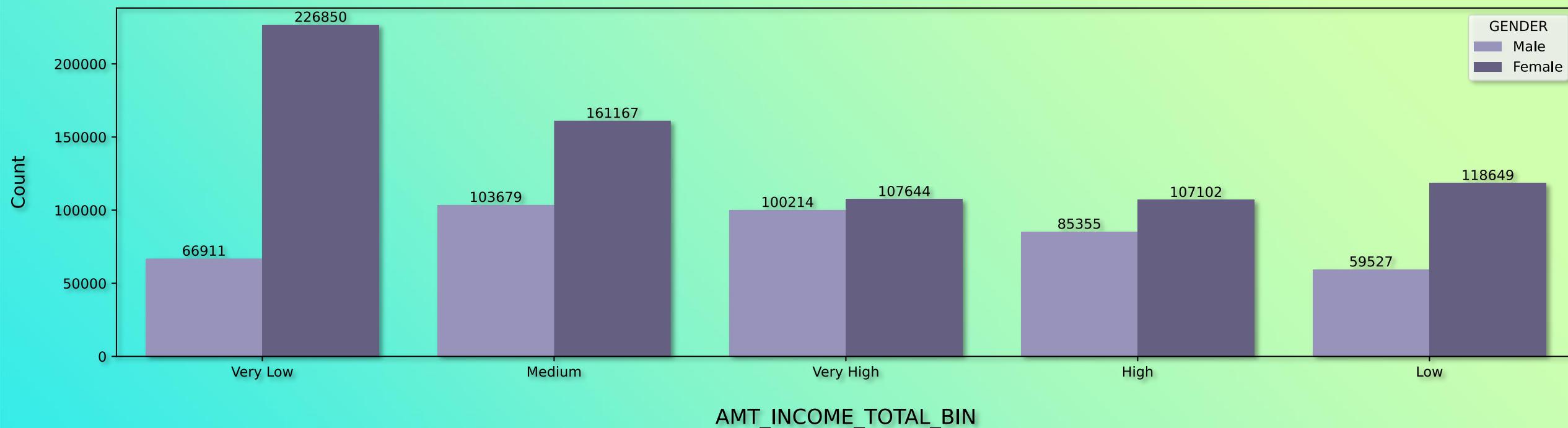


Annuity Amount
Application Amount
Credit Amount
Good's Price
Are
Highly
Correlated.

Days Last Due
Days Termination
Are
Highly
Correlated.

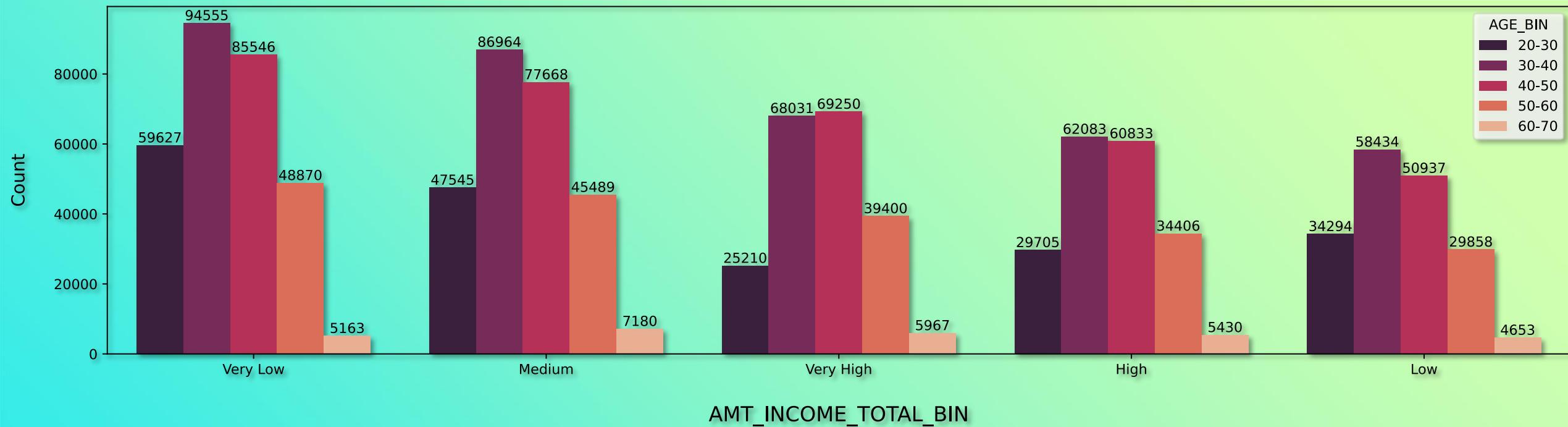
Bivariate Analysis

AMT_INCOME_TOTAL_BIN v/s GENDER



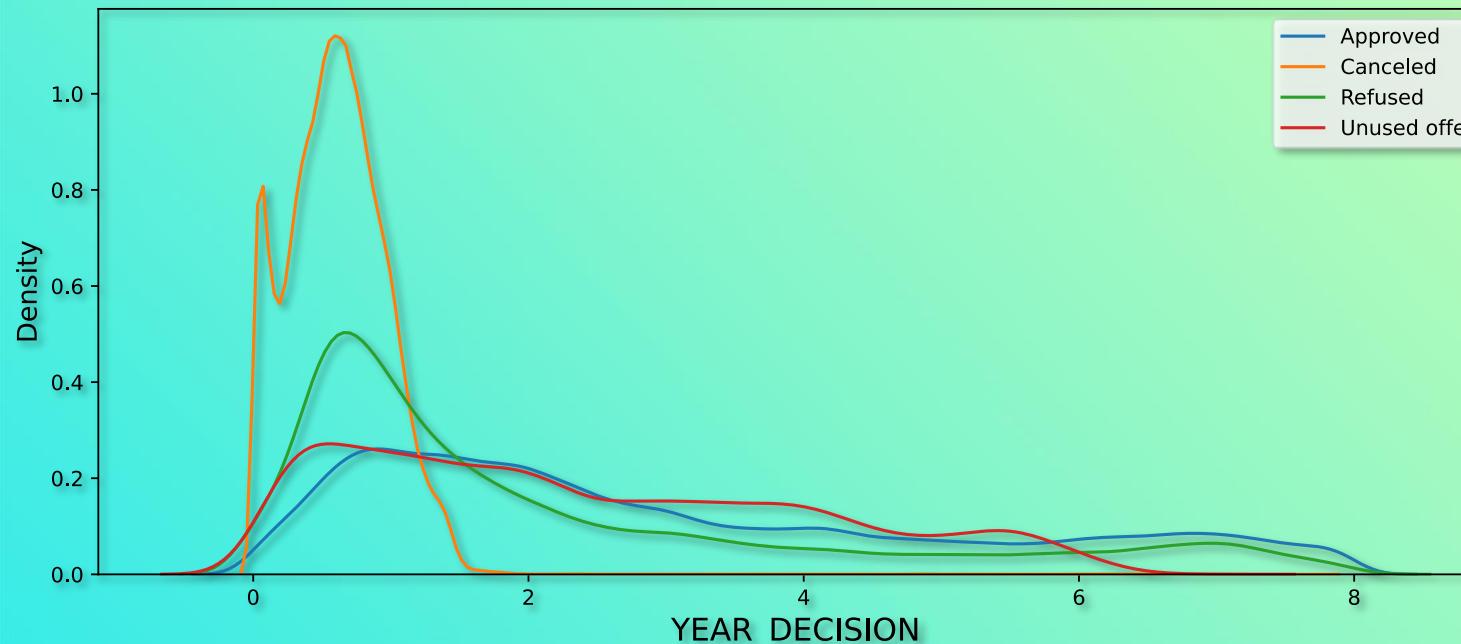
- Most People Applied For Loans Are Females.
- Most Females Are There With Very Low Total Income.

AMT_INCOME_TOTAL_BIN v/s AGE_BIN



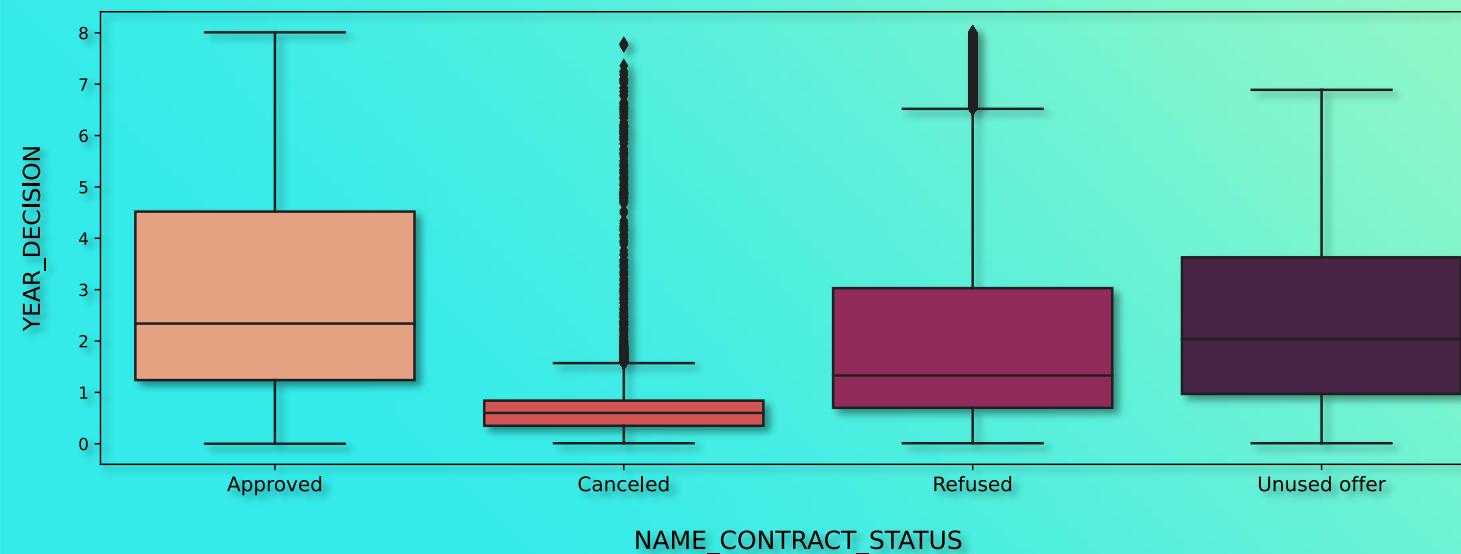
- Most People Are There With Age Group 30-40.
- Most People With Age Group 30-40 Has Very Low Total Income Amount.

Distribution Of NAME_CONTRACT_STATUS in YEAR_DECISION



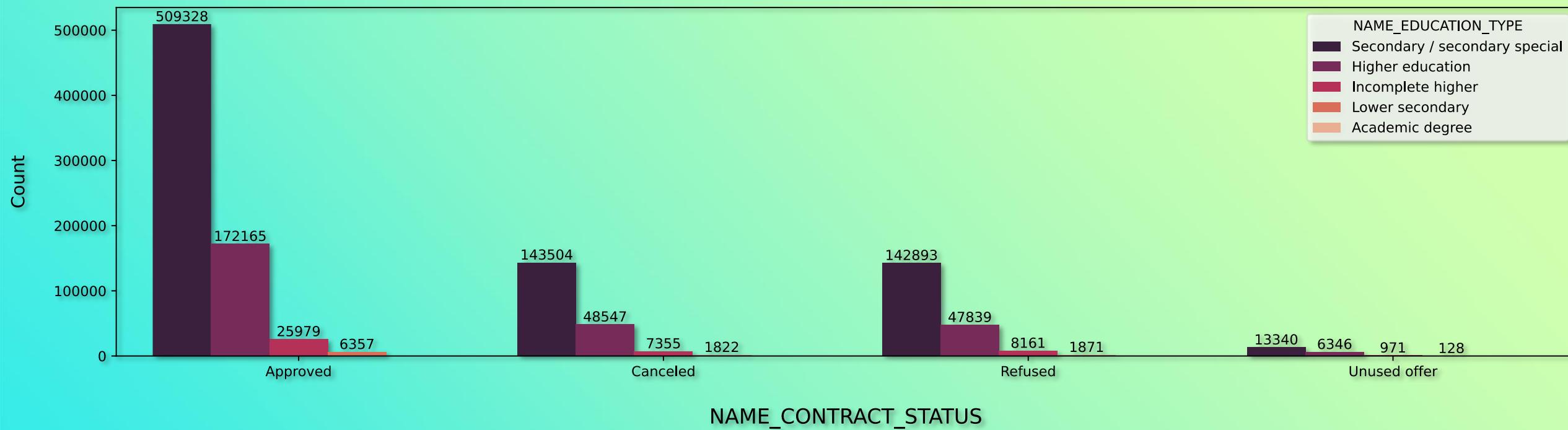
- People Who Made Their Decision Between 6 Months To 1 Years Got Their Application Canceled.

NAME_CONTRACT_STATUS v/s YEAR_DECISION



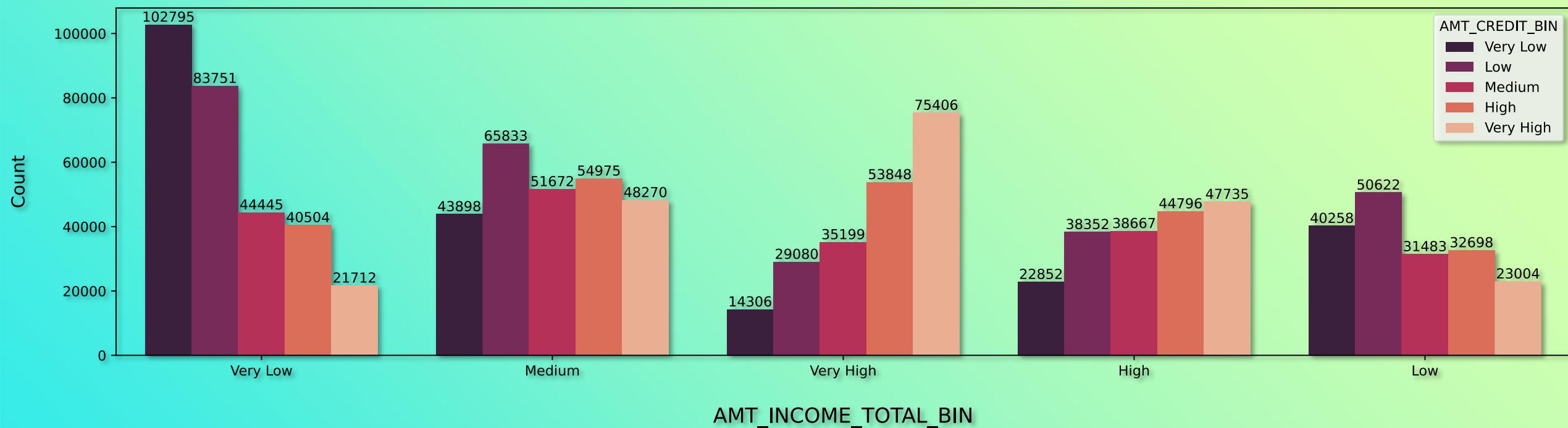
- People With Year Decision Between 1 To 4.5 Years Got Most Approvals.

NAME_CONTRACT_STATUS v/s NAME_EDUCATION_TYPE



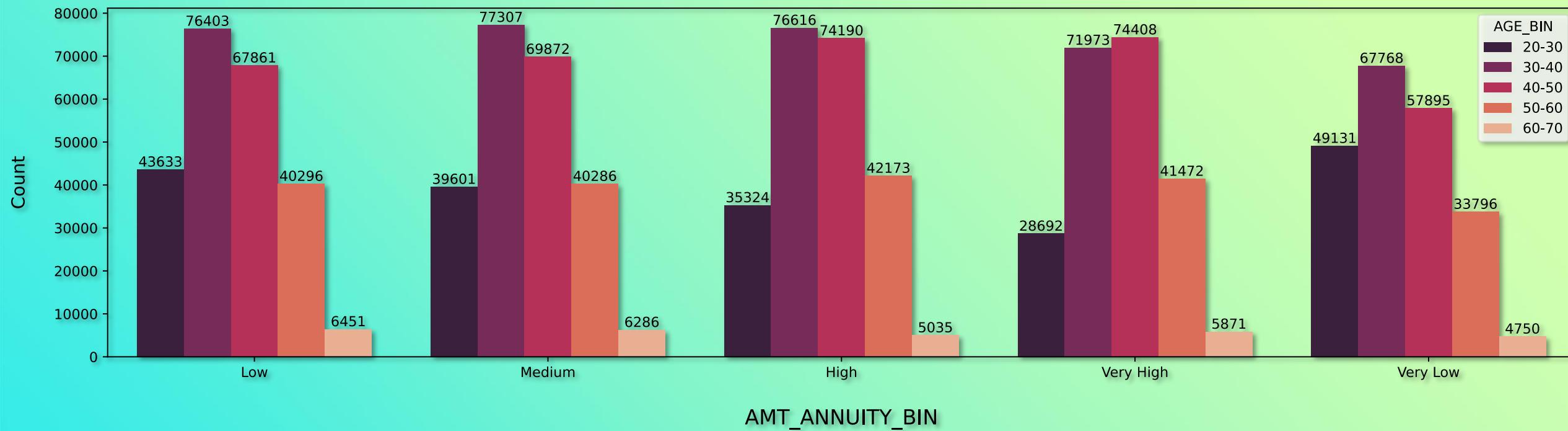
- People With Secondary/secondary Special Education Got Most Approvals.

AMT_INCOME_TOTAL_BIN v/s AMT_CREDIT_BIN



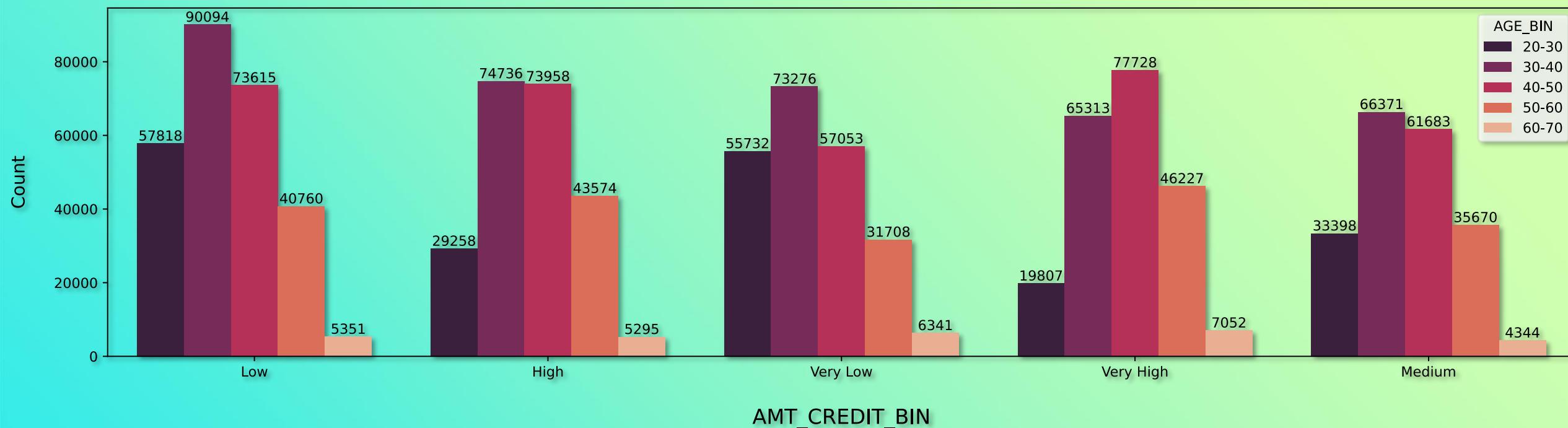
- People With Very Low Income Amount Has Very Low Credit Amount As well.
- People With Very Low Income Amount Applied Most.

AMT_ANNUITY_BIN v/s AGE_BIN



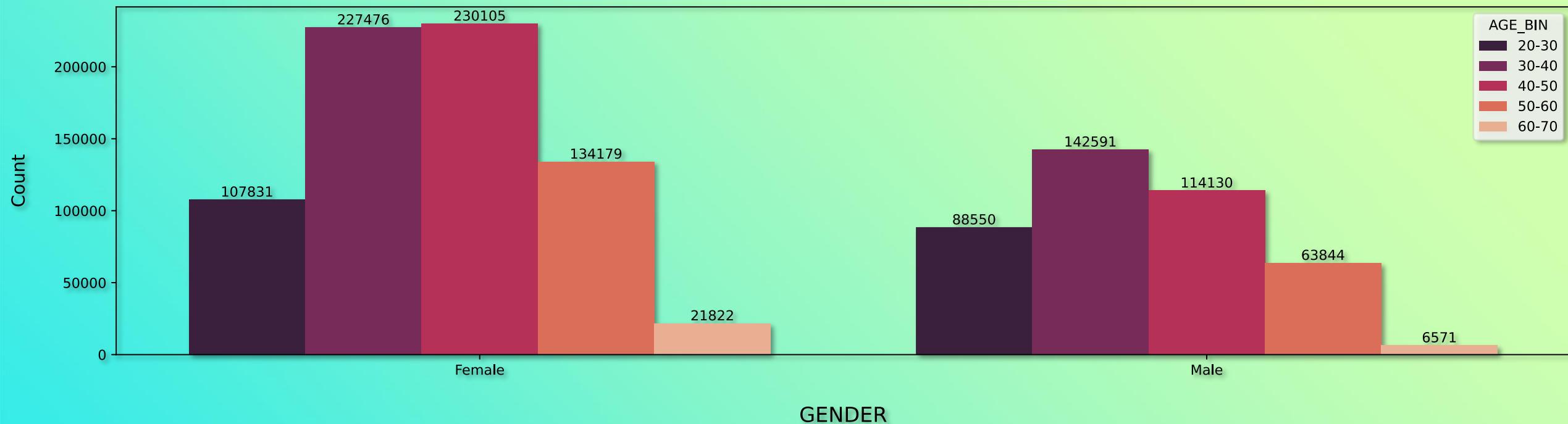
- Most People With Age Group 30-40 Has Low ,Medium ,High Annuity Amount.
- Most People With Age Group 40-50 Has Very High Annuity Amount.

AMT_CREDIT_BIN v/s AGE_BIN

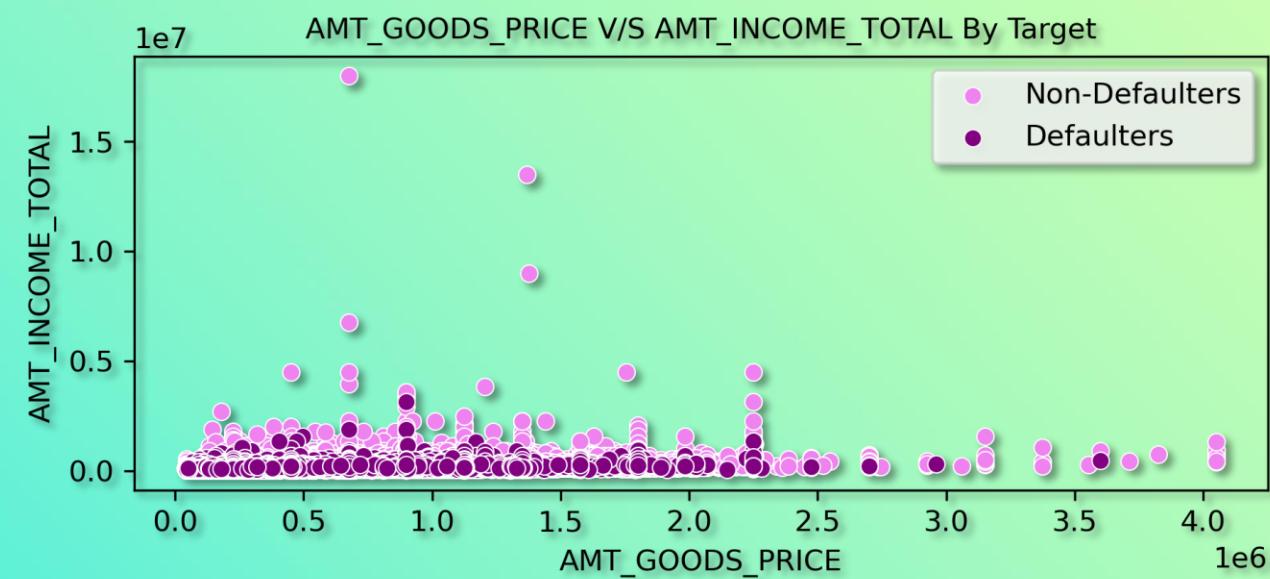
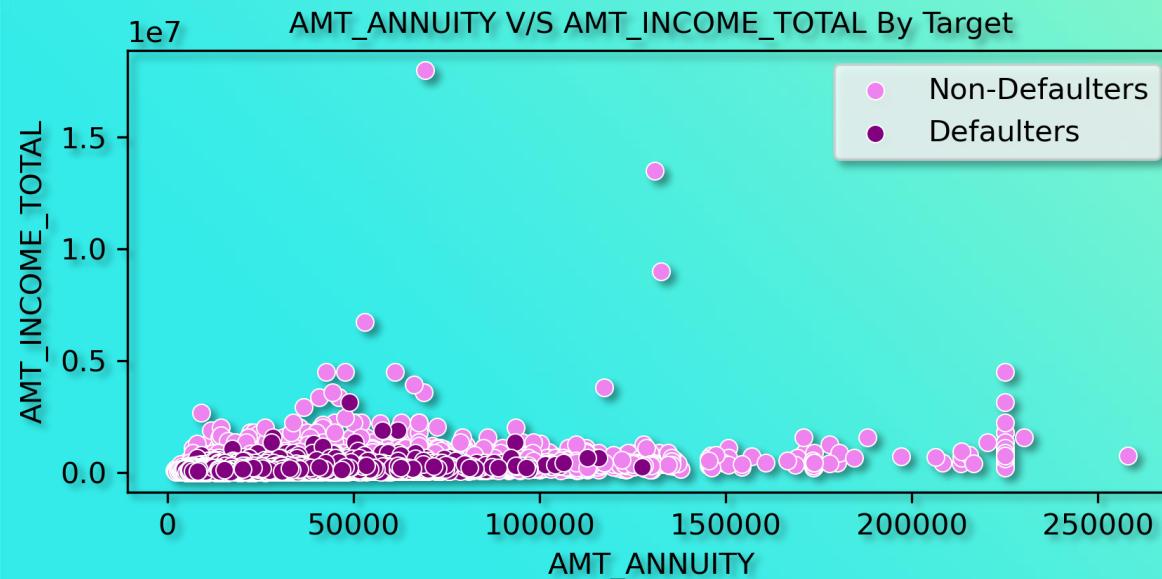
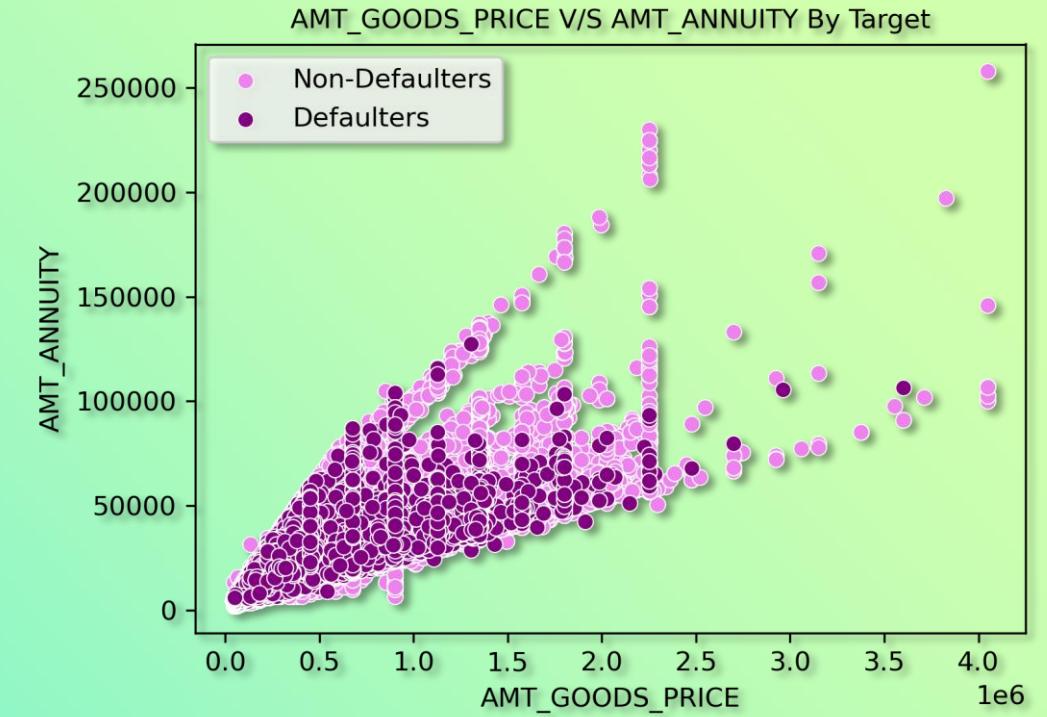
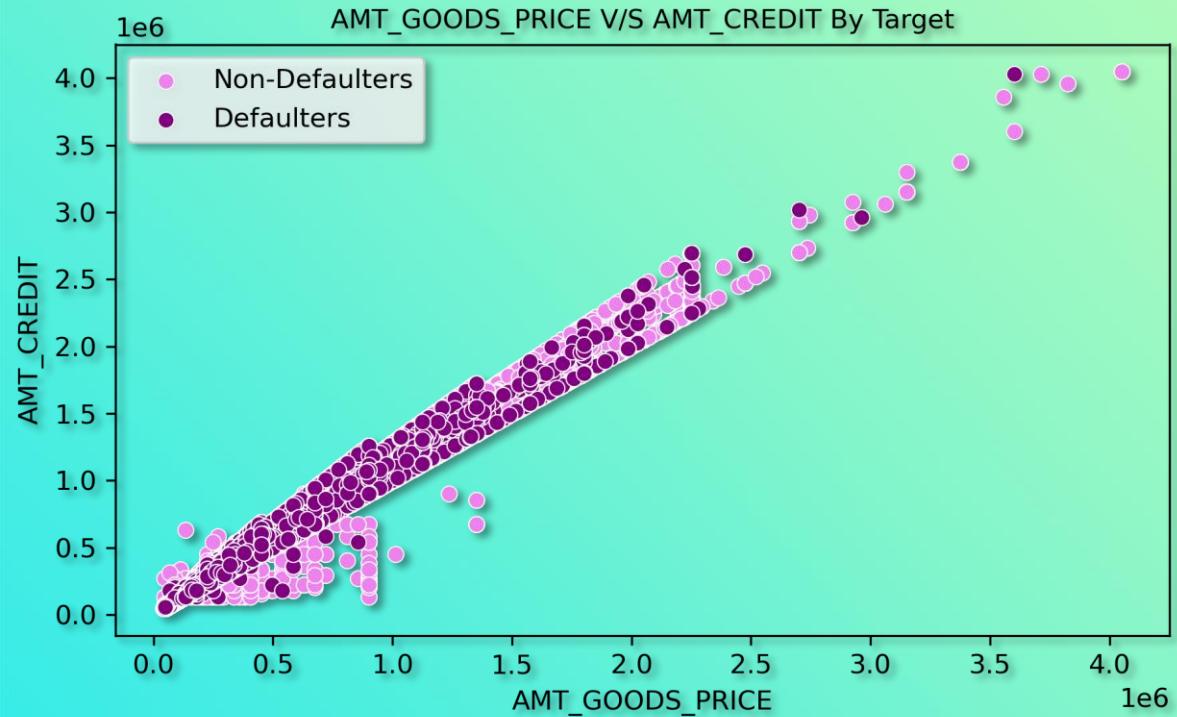


- Most People With Age Group 40-50 Has Very High Credit Amount.
- Most People With Age Group 30-40 Has Low Credit Amount.

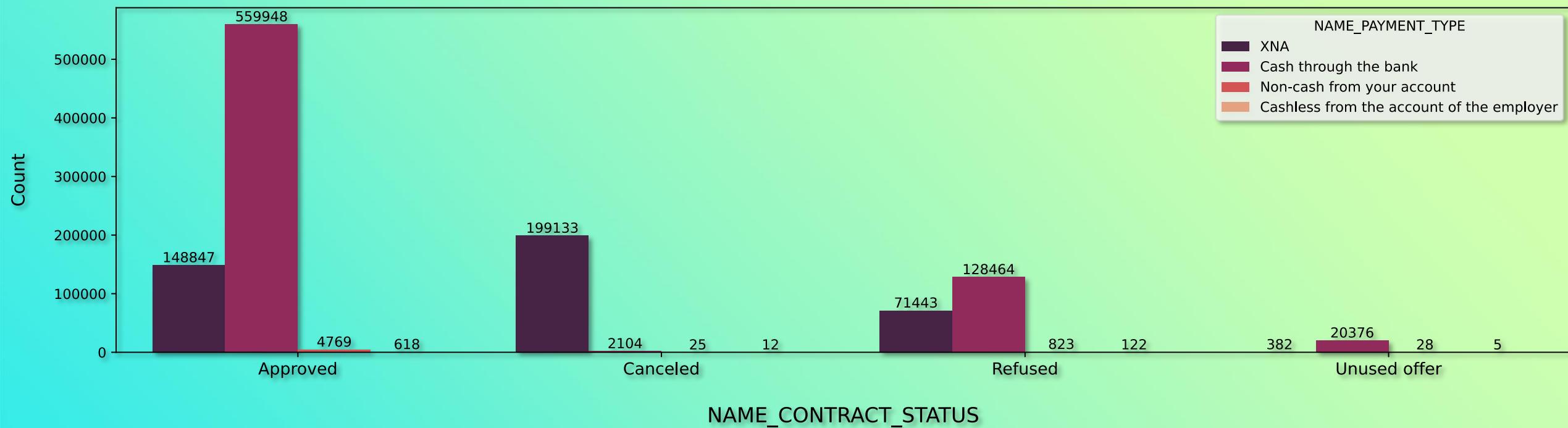
GENDER v/s AGE_BIN



- Most Numbers Of People Are Females With Age Group 30-50.

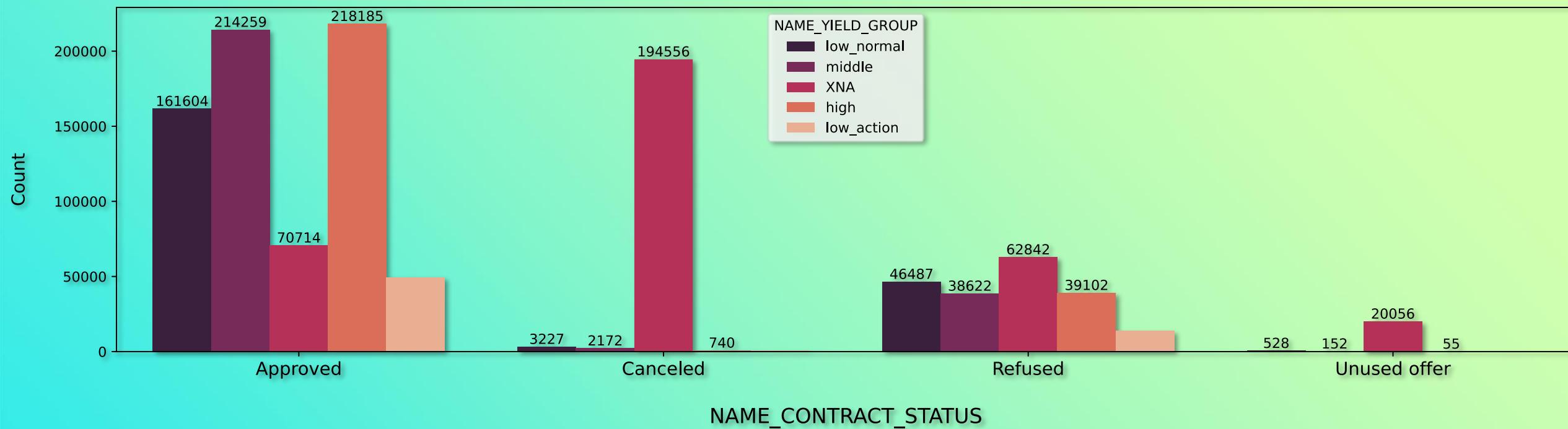


NAME_CONTRACT_STATUS v/s NAME_PAYMENT_TYPE



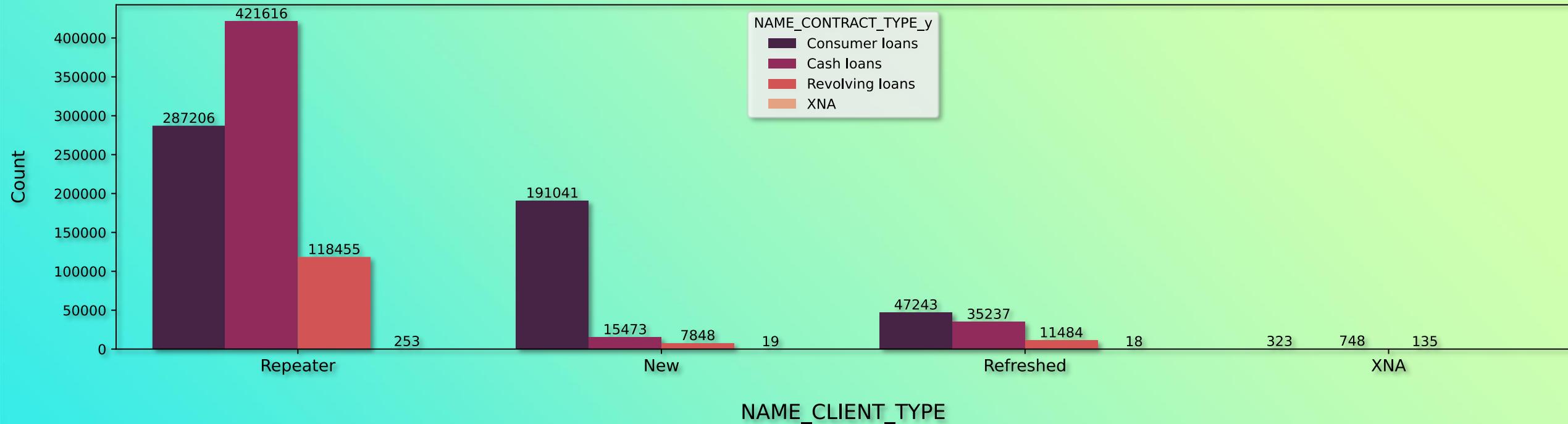
- People With Payment Type Cash Through Bank Got Most Approvals.
- People With Unknown Payment Type Got Most Cancellation.

NAME_CONTRACT_STATUS v/s NAME_YIELD_GROUP



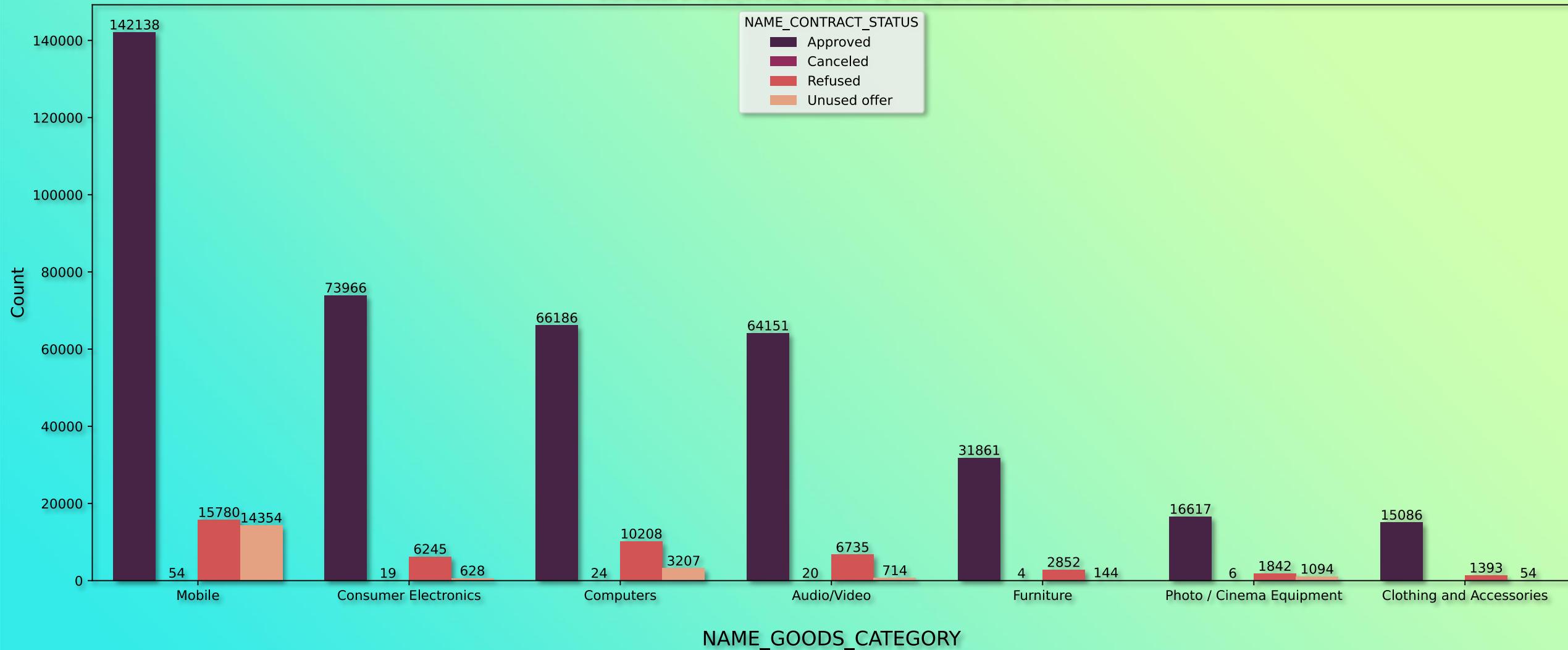
- People With Both High And Medium Interest Rate Got Most Approval.

NAME_CLIENT_TYPE v/s NAME_CONTRACT_TYPE_y



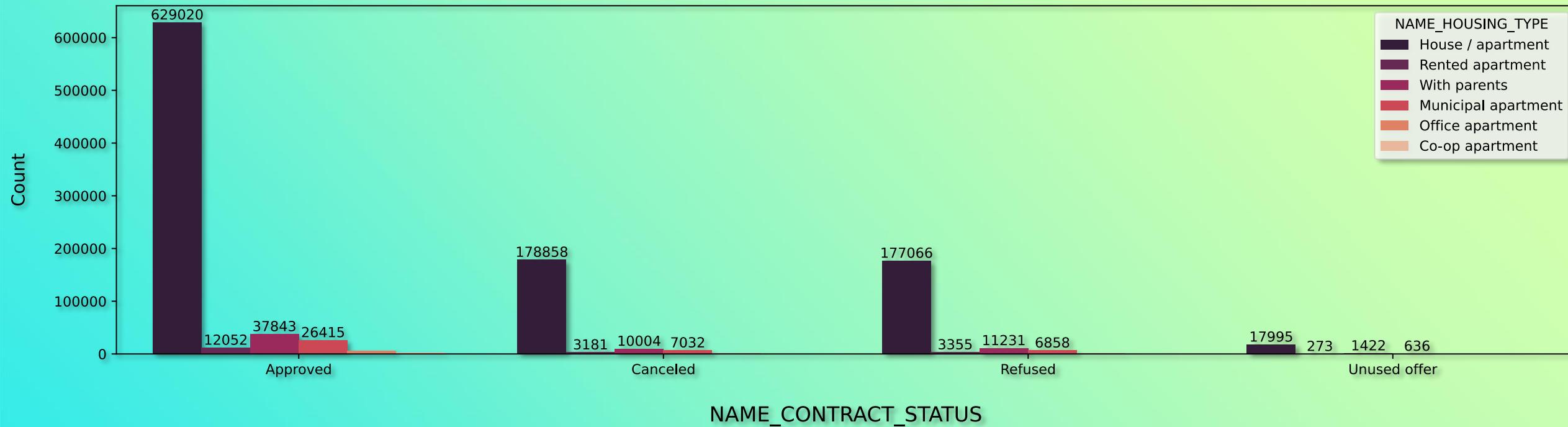
- People With Most Cash Loans Were Repeater Client.

Distribution of NAME_GOODS_CATEGORY By NAME_CONTRACT_STATUS



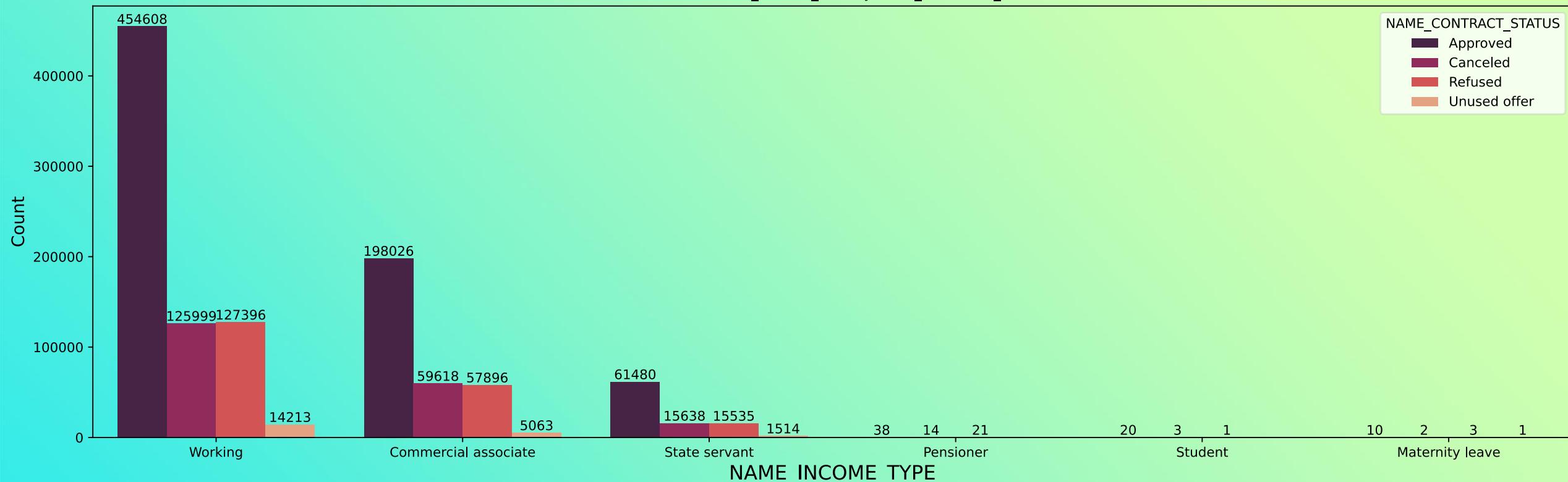
- People With Good's Category 'Mobile' Got Most Approvals.

NAME_CONTRACT_STATUS v/s NAME_HOUSING_TYPE



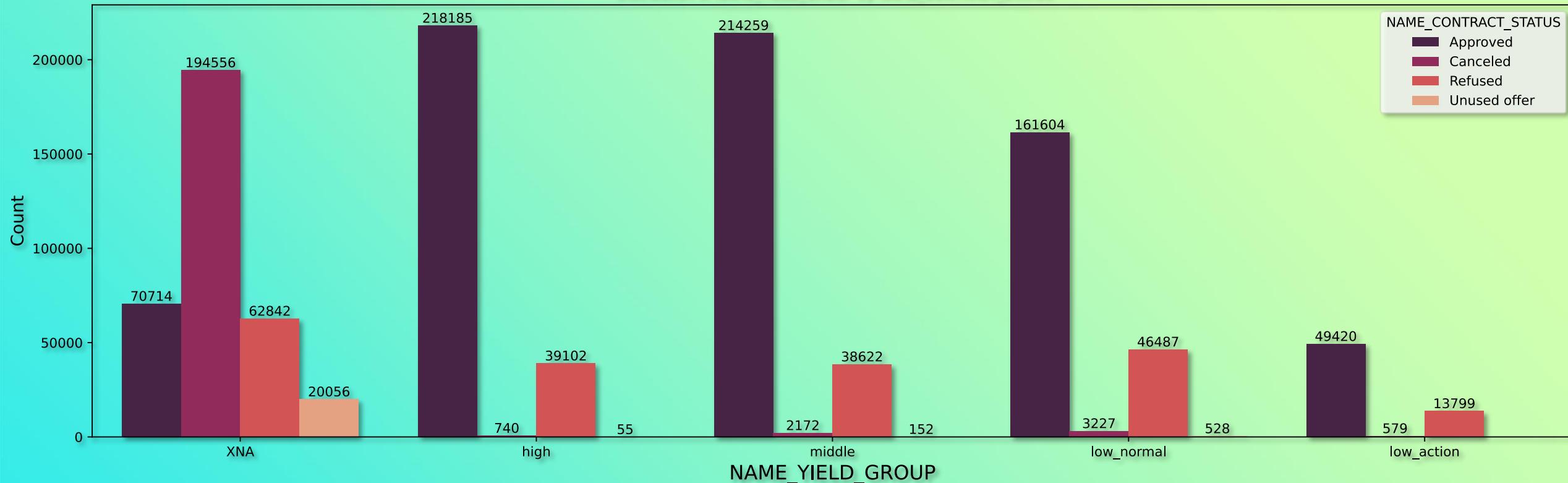
- People With Housing Type ‘House/Apartment’ Got Most Approvals.

Distribution of NAME_INCOME_TYPE By NAME_CONTRACT_STATUS

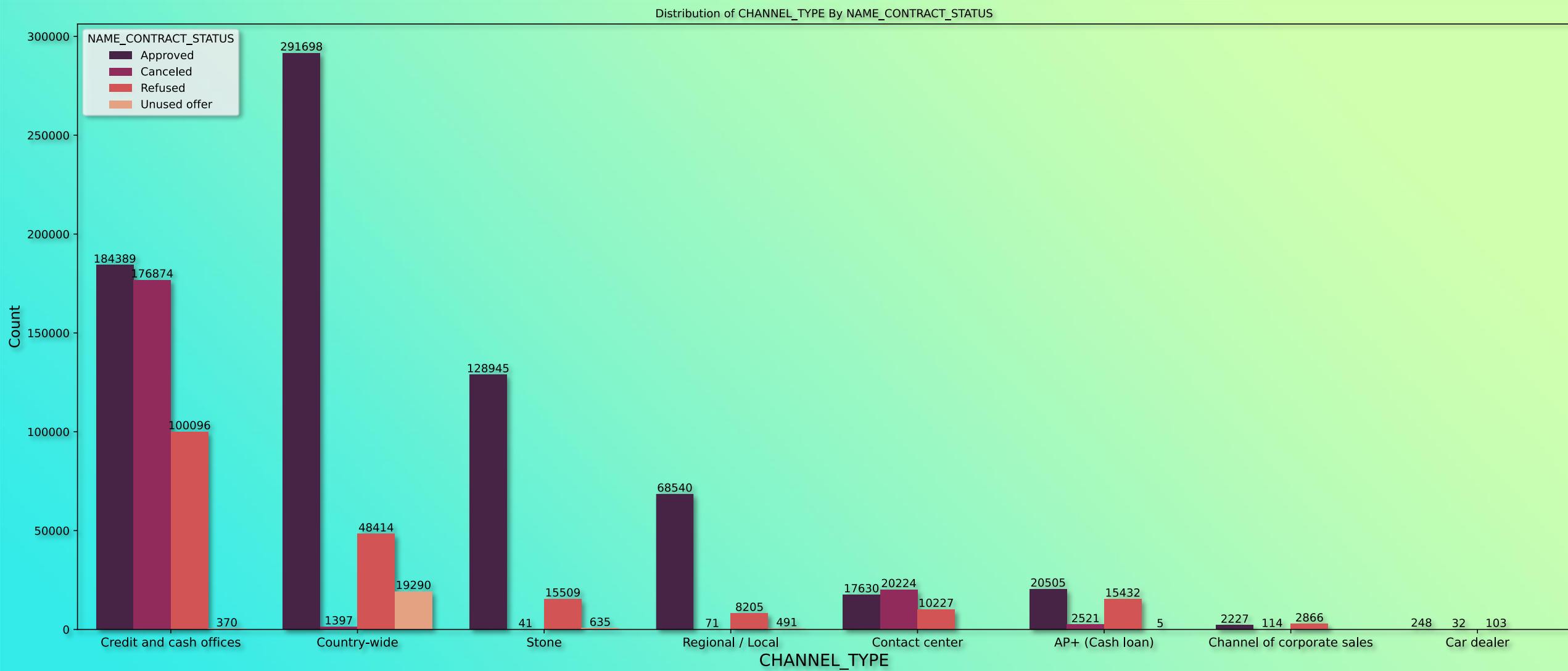


- People With Working Income Type Got Most Approvals.

Distribution of NAME_YIELD_GROUP By NAME_CONTRACT_STATUS

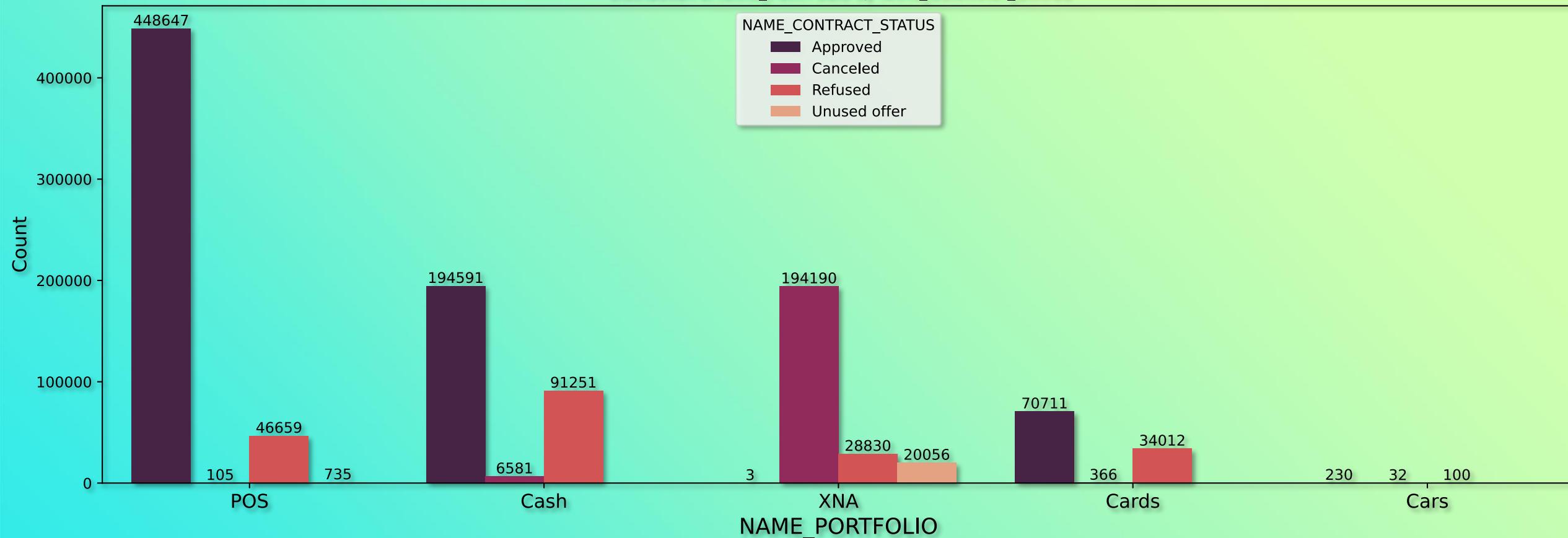


- People With ‘High’ And ‘Medium’ Interest Rate Got Most Approvals.
- People With ‘Unknown’ Yield Got Most Cancellations.



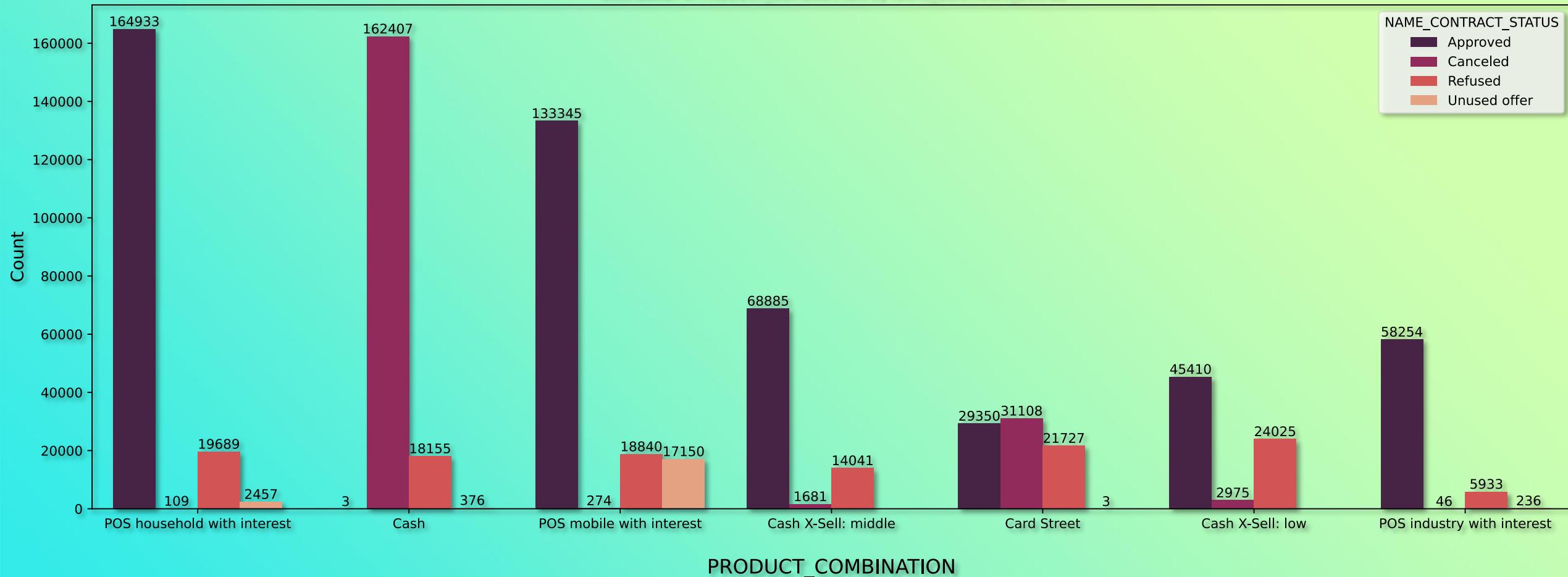
- People With Channel Type ‘Country-Wide’ Got Most Approvals.
- People With Channel Type ‘Credit And Cash Offices’ Have Most Cancellations.

Distribution of NAME_PORTFOLIO By NAME_CONTRACT_STATUS



- People With POS Portfolio Got Most Approvals.
- People With Cash Portfolio Got Most Refusals.

Distribution of PRODUCT_COMBINATION By NAME_CONTRACT_STATUS



- People With Product Combination POS Household With Interest Got Most Approvals.
- People With Product Combination Cash Have Most Cancellations.

Recommendation

- Bank Lends More Loans To Females.
- More Cash Loans Leads To Defaulters, Instead Revolving Loans Should Be Given.
- People With Higher Education Default Less.
- Married People Are Less Defaulter, Unmarried/Single People Defaults More.
- Older People Are Less Defaulters.
- People Working As State Servants Default Less.
- People With High Income Are Less Defaulters.
- Male Clients Are More Defaulters.
- Refused And Canceled Loan Status Are More Defaulters.
- People Who Take Higher Loans Are Less Defaulters.
- People With Product Combination POS Household With Interest Got Most Approvals.
- People With POS Portfolio Got Most Approvals.
- People With Channel Type ‘Country-Wide’ Got Most Approvals.
- People With ‘High’ And ‘Medium’ Interest Rate Got Most Approvals.
- People With Good’s Category ‘Mobile’ Got Most Approvals.
- People With Most Cash Loans Were Repeater Client.
- People With Payment Type Cash Through Bank Got Most Approvals.
- People With Year Decision Between 1 To 4.5 Years Got Most Approvals.