Lending Club Case Study

Javvaji Gokul

Krishnendra Nandi

Data Cleansing and Filtering

- Removed all behavioral columns as was given by facilitators as people who are applying the loans wont have that data
- Removed columns having all rows as null
- Unique data keys (id, member_id, url) have no bearing on analysis
- Emp_title & Title have no role in analysis
- Zip code is of no significance
- Removed all columns having null values at least 70%
- Removed all singleton value columns (columns having just 1 value)

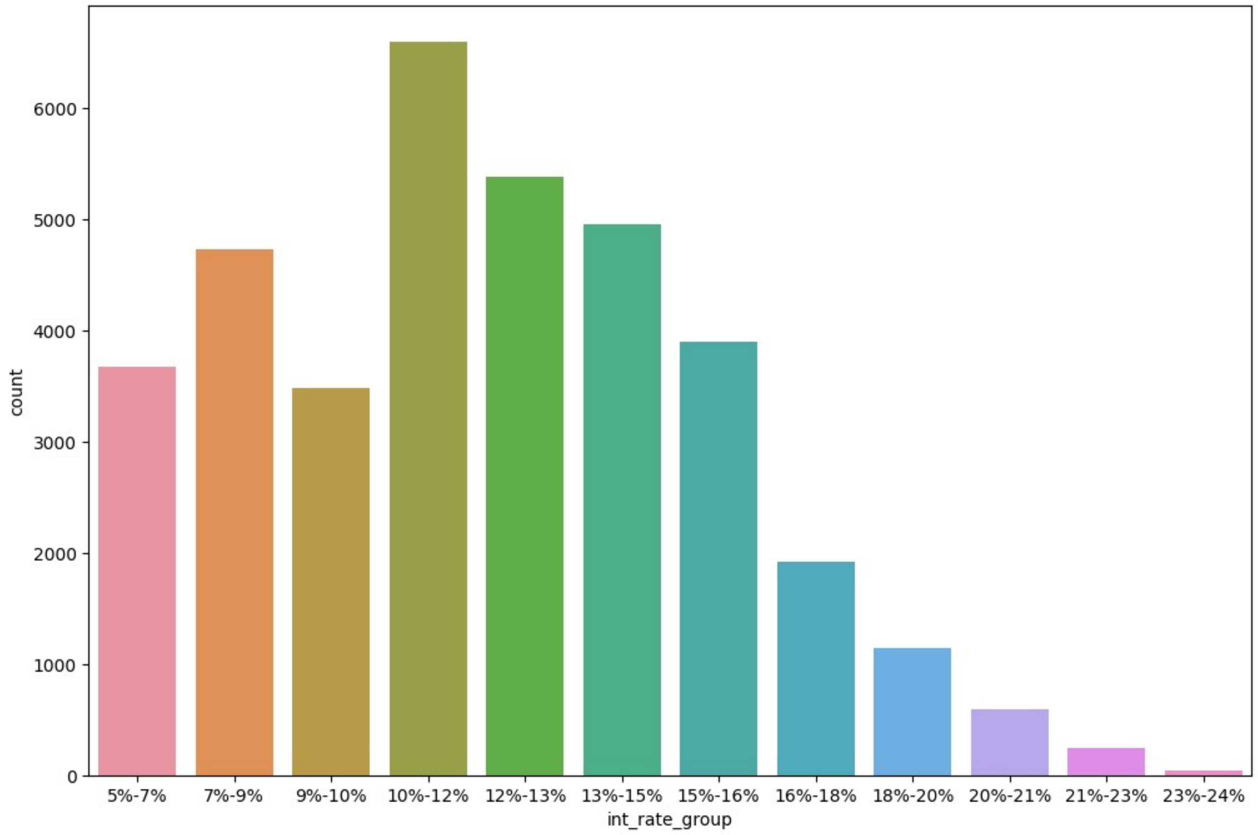
Data Assumptions

- Emp length not mentioned has been considered as -1 for visualization purposes
- Since we are considering defaulted loans , hence we didn’t consider “Current” as Loan status in our analysis
- Home ownership had “None” which was just very insignificant count (count=3) , hence removed it
- Removed outliers from Annual Income by considering 95 percentile data

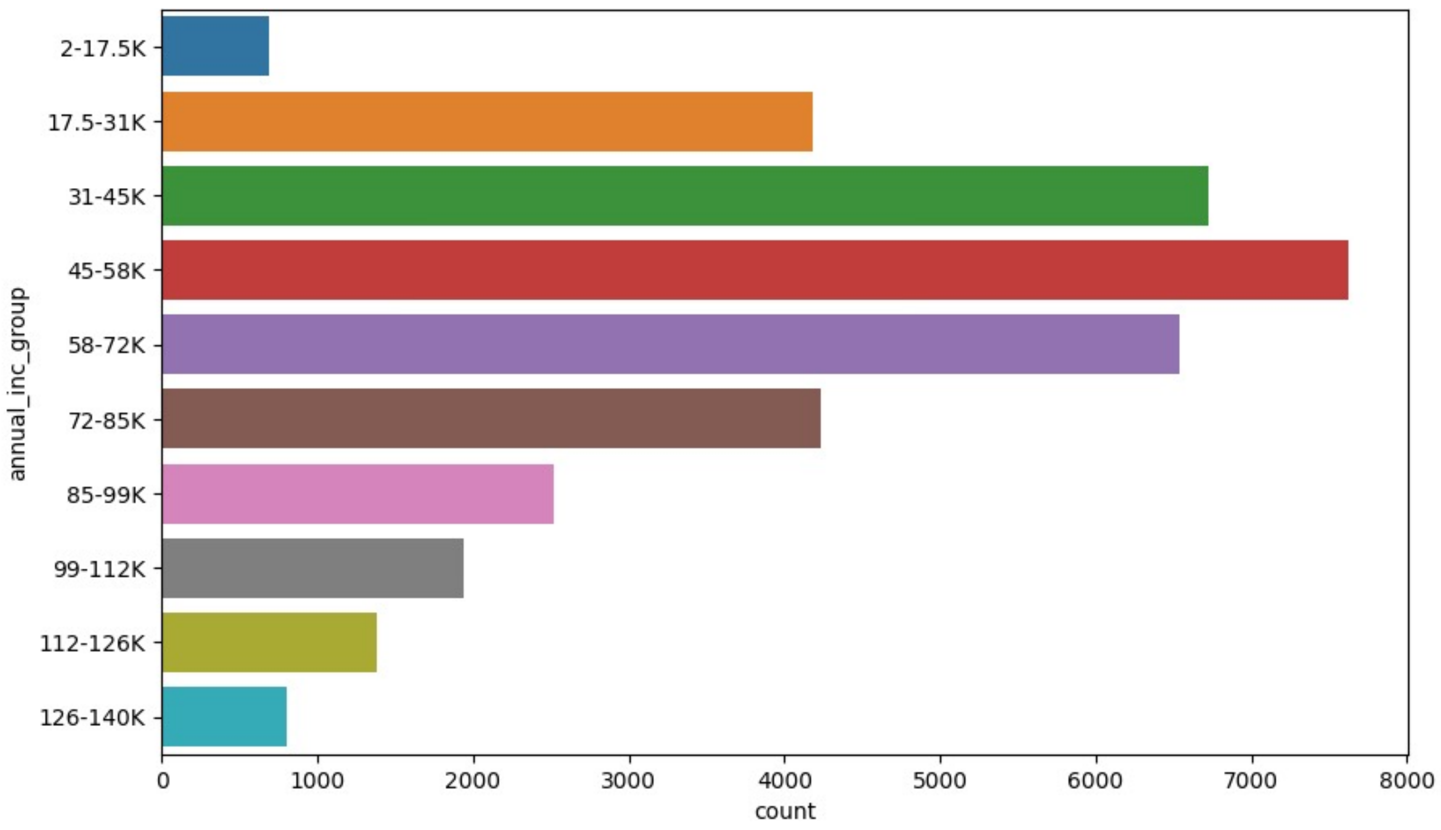
Data type Conversions

- Interest rate columns converted to numeric for analysis
- Term column in months was filtered by removing additional string and then converting to numeric
- Derived columns of year and month extracted from issue data
- Sub_Grade Column – Since we have grade column we converted the sub_grade column to numerical column for better analysis
- We created bins for interest rate and annual income columns

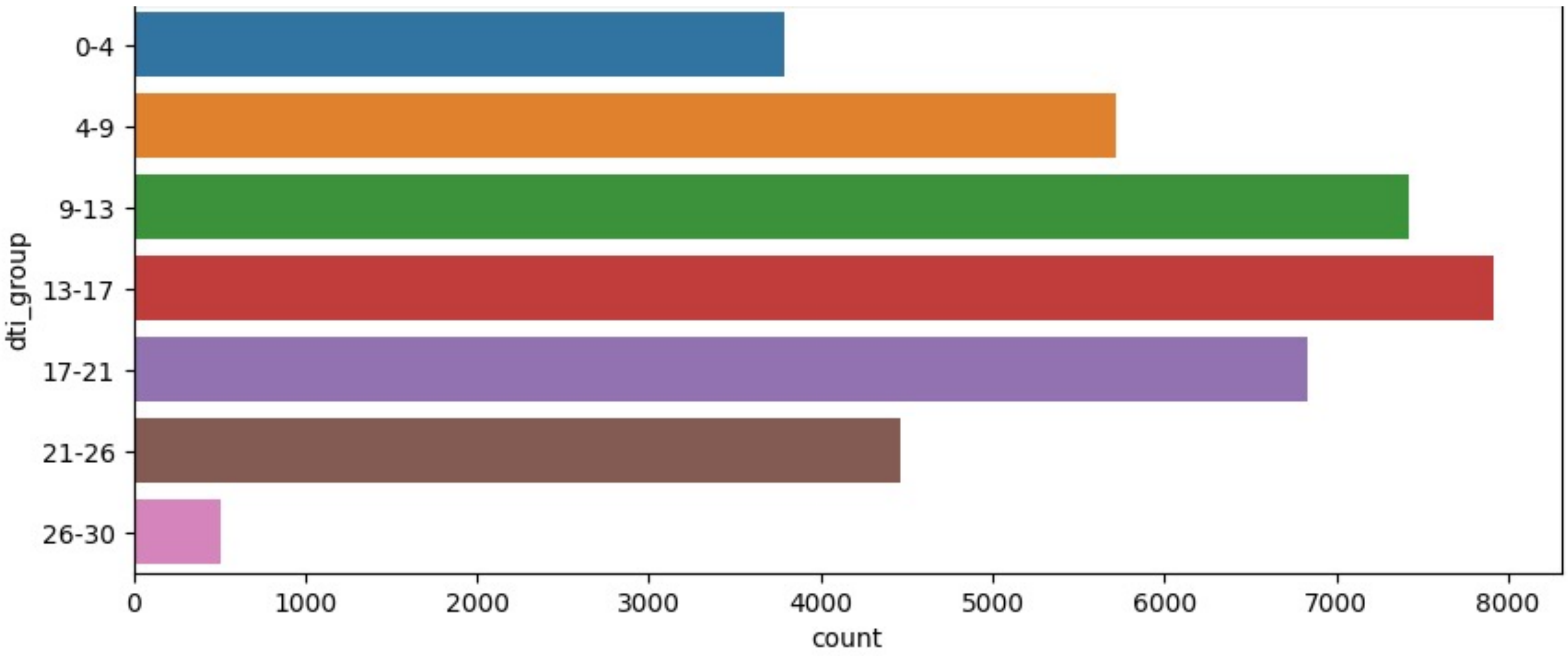
Interest Rate



Annual Income Group

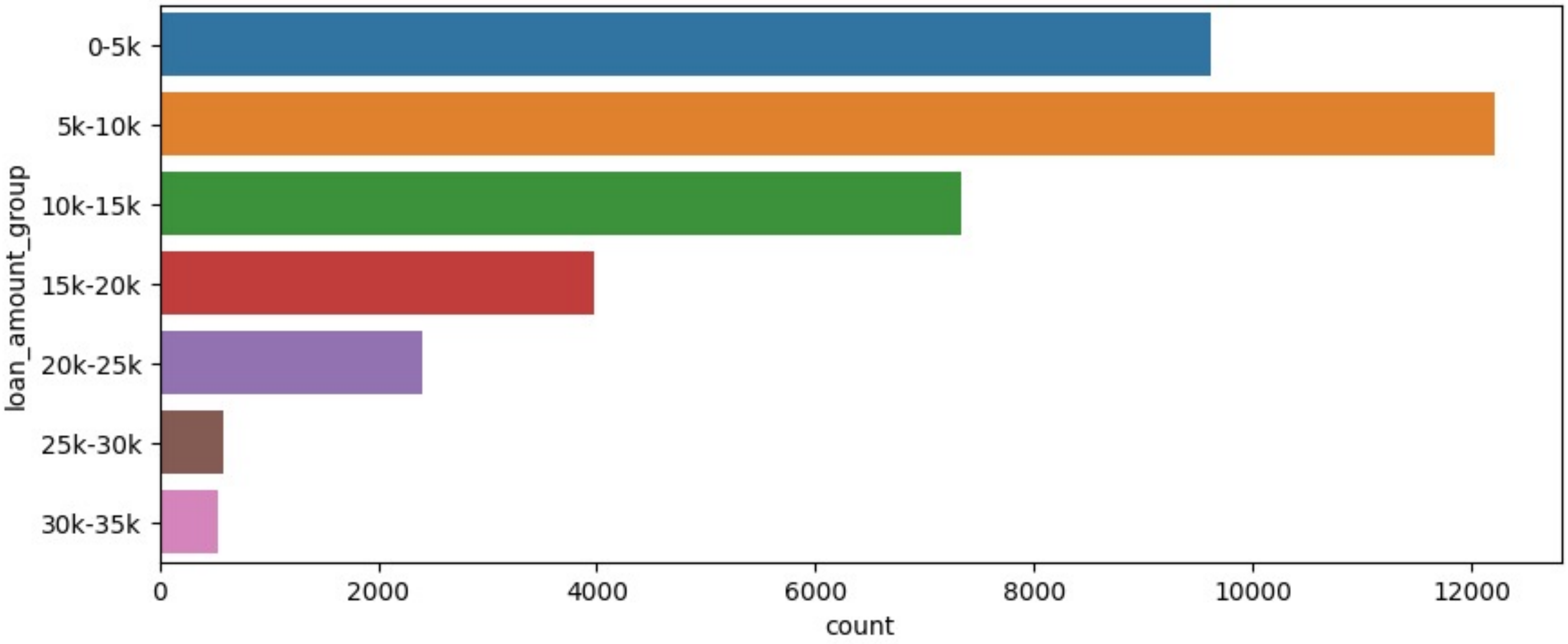


DTI Group

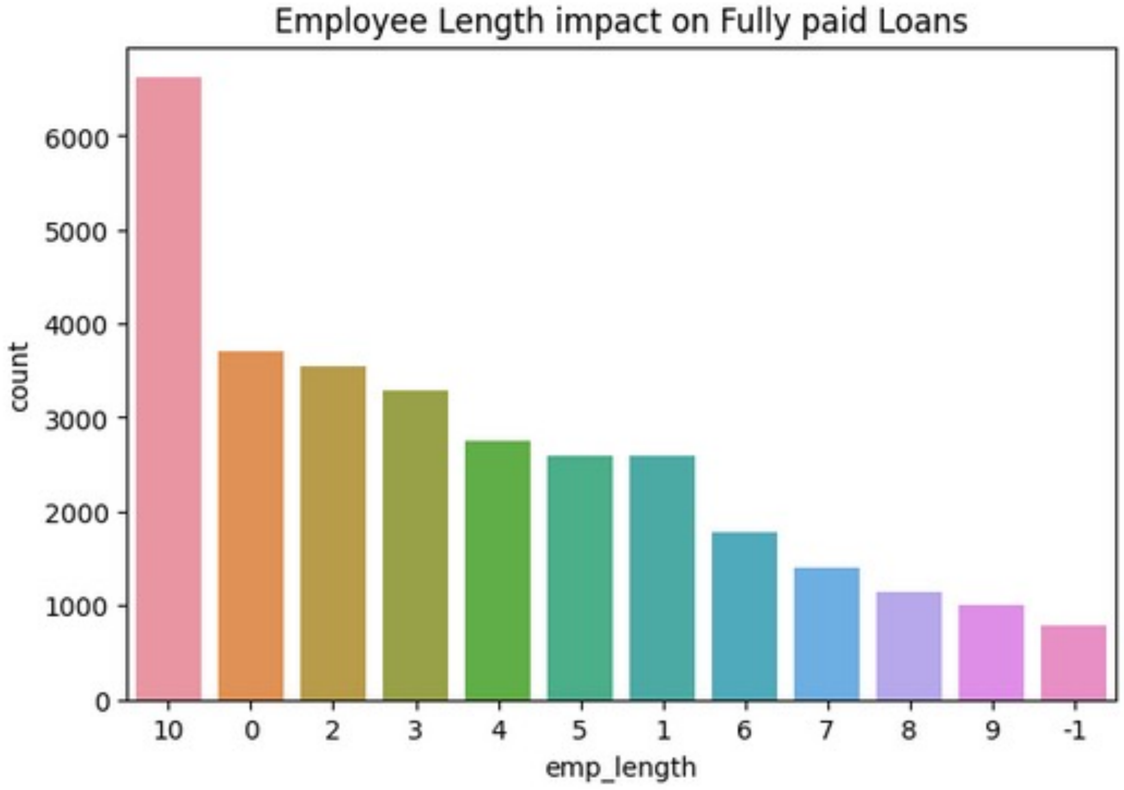

0-4
4-9
9-13
13-17
17-21
21-26
26-30
dti_group
0
1000
2000
3000
4000
5000
6000
7000
8000
count

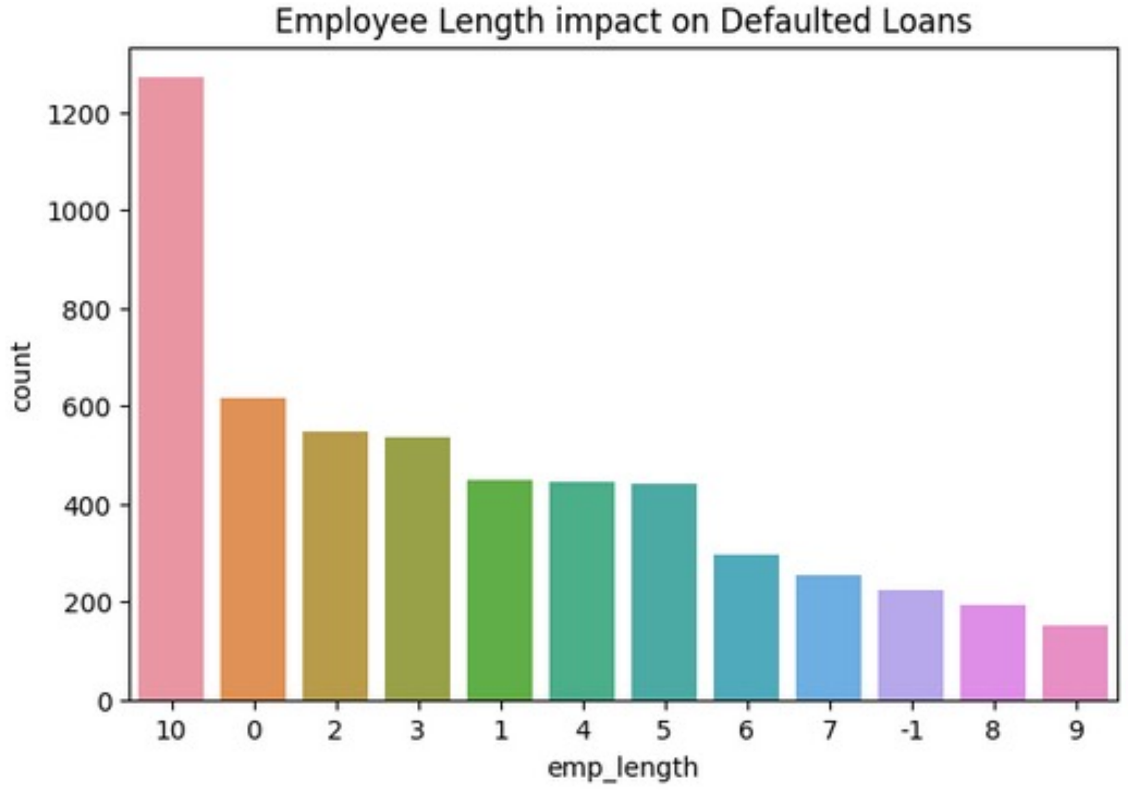
Internal

Loan Amount Group

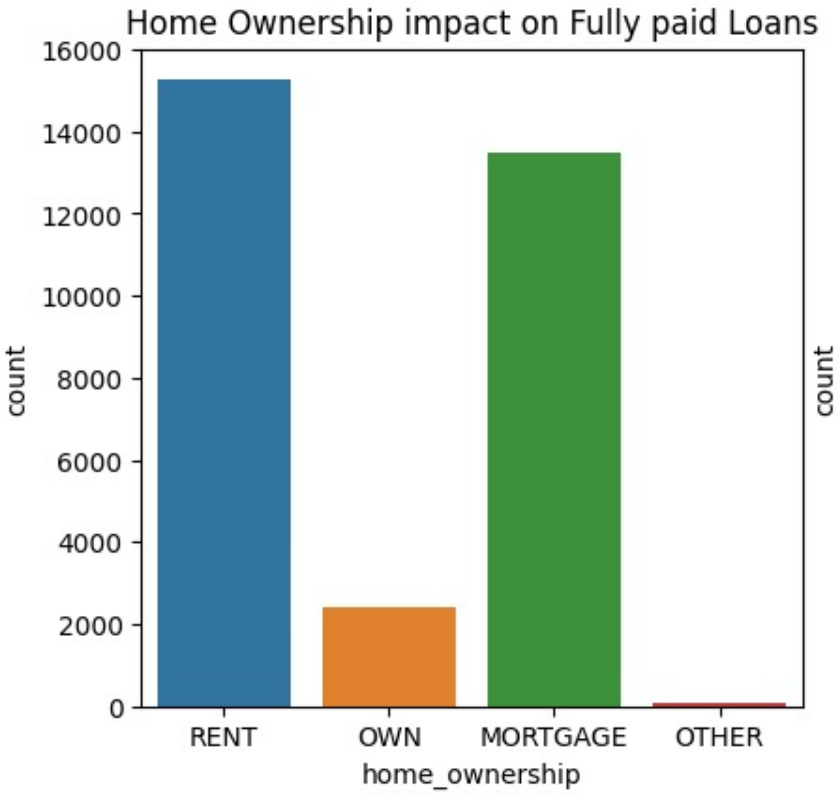


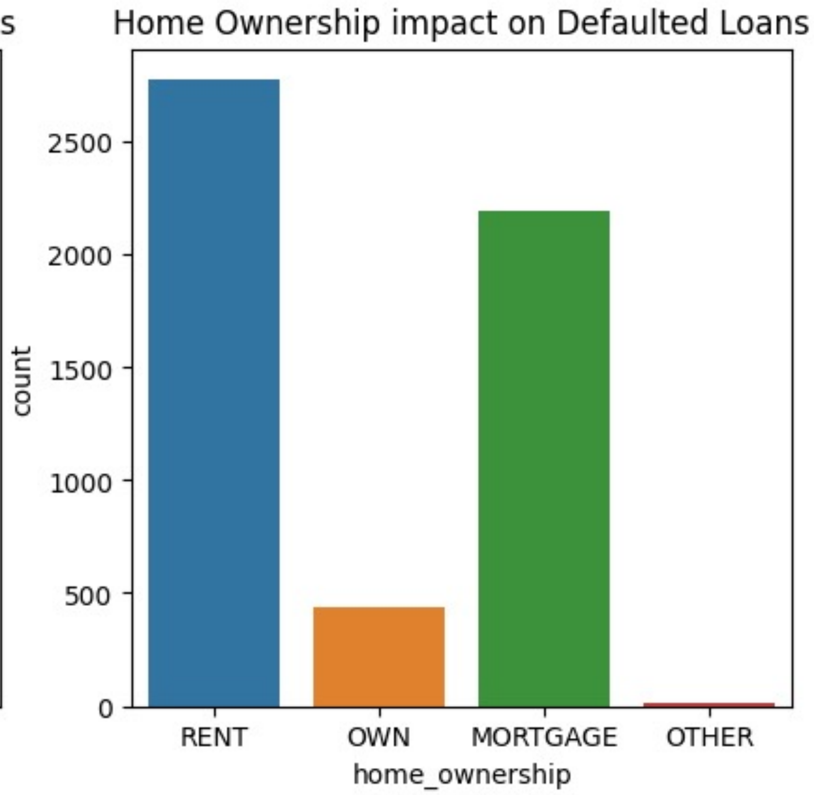
Employee Length Impact





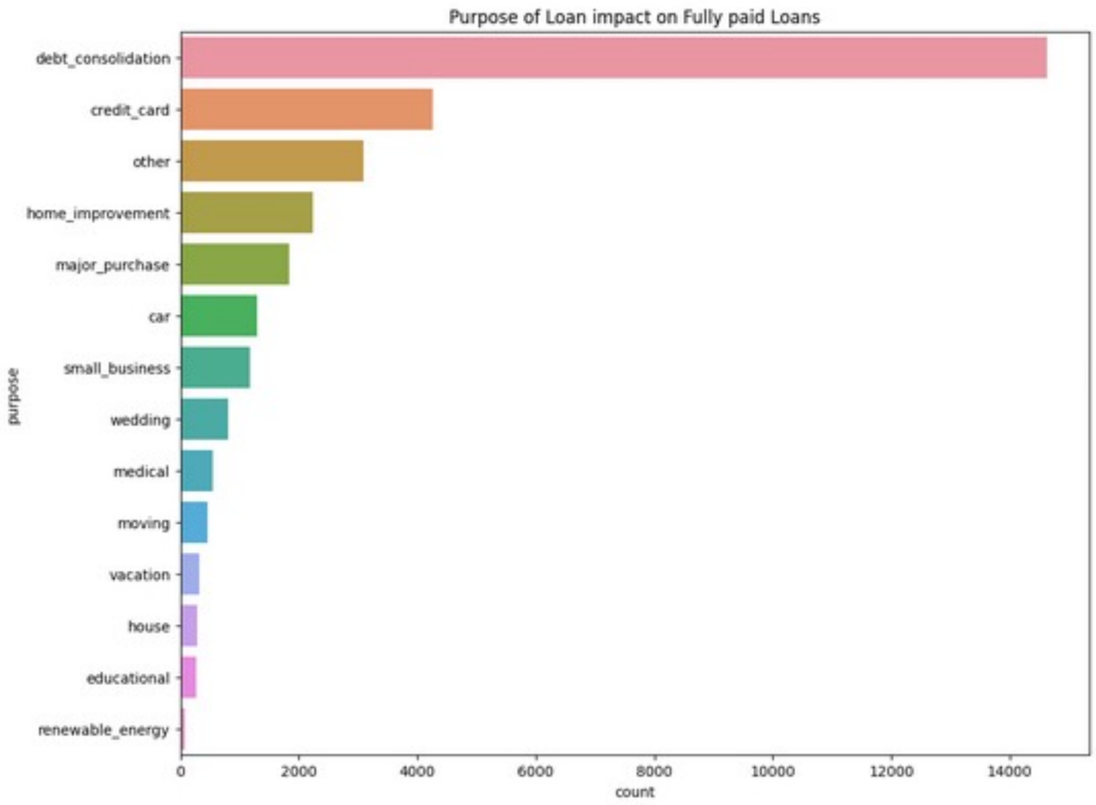
Home Ownership Impact

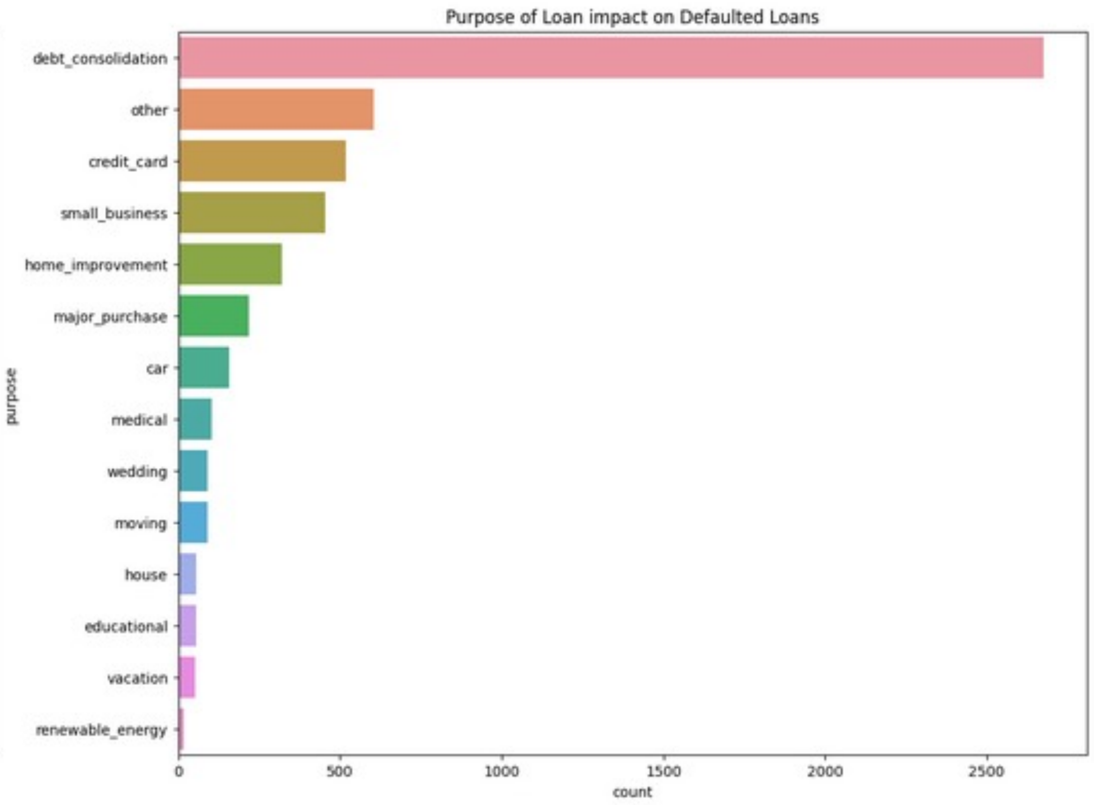




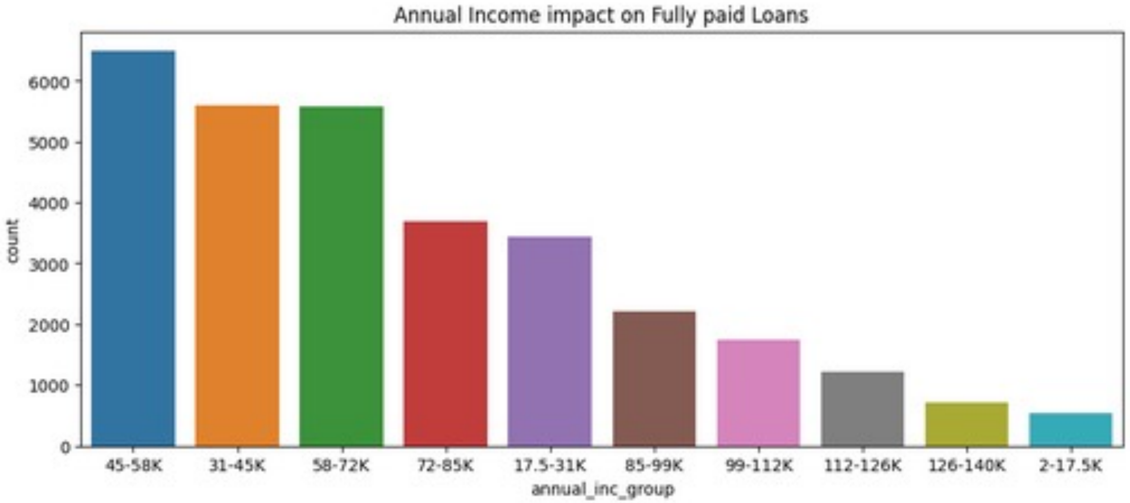
Internal

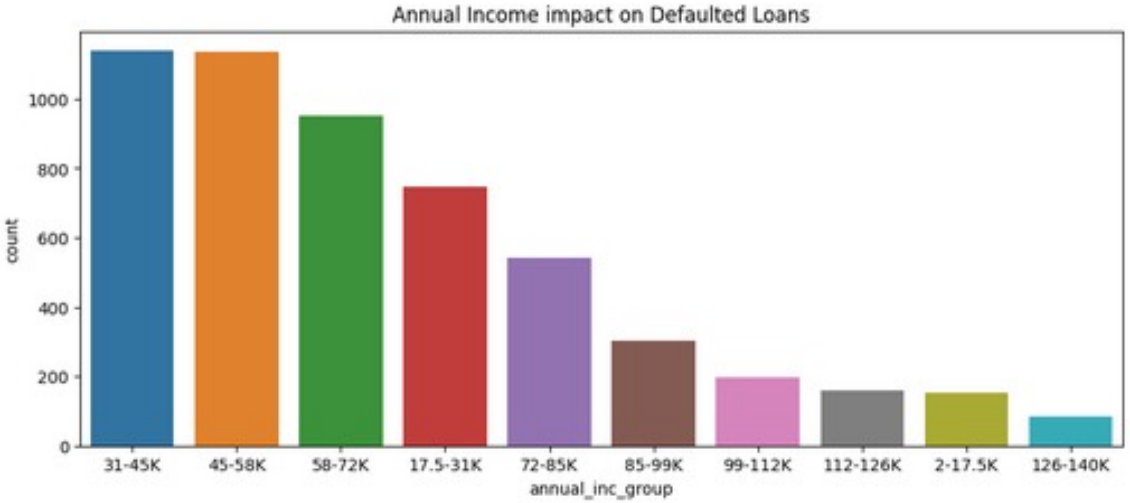
Purpose of Loan Impact



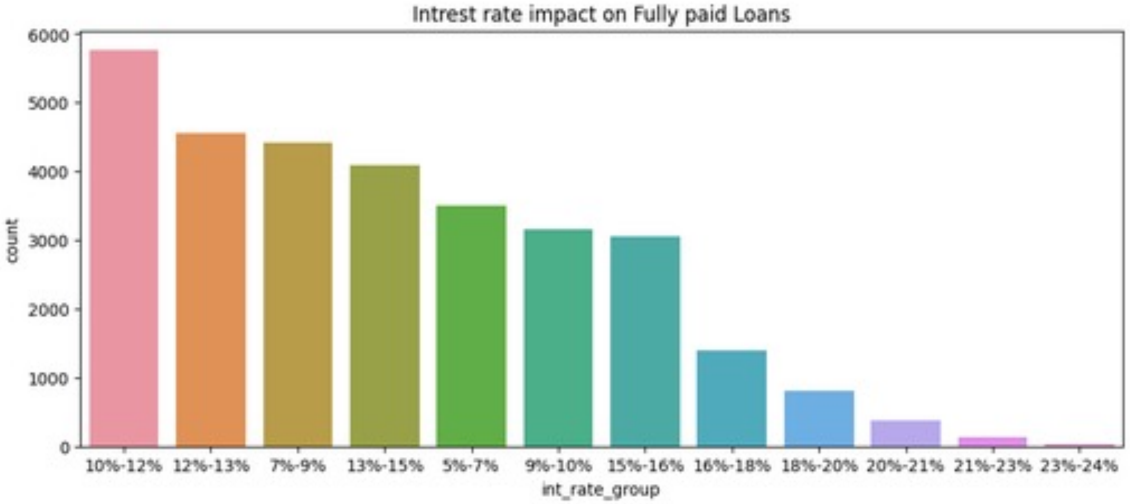


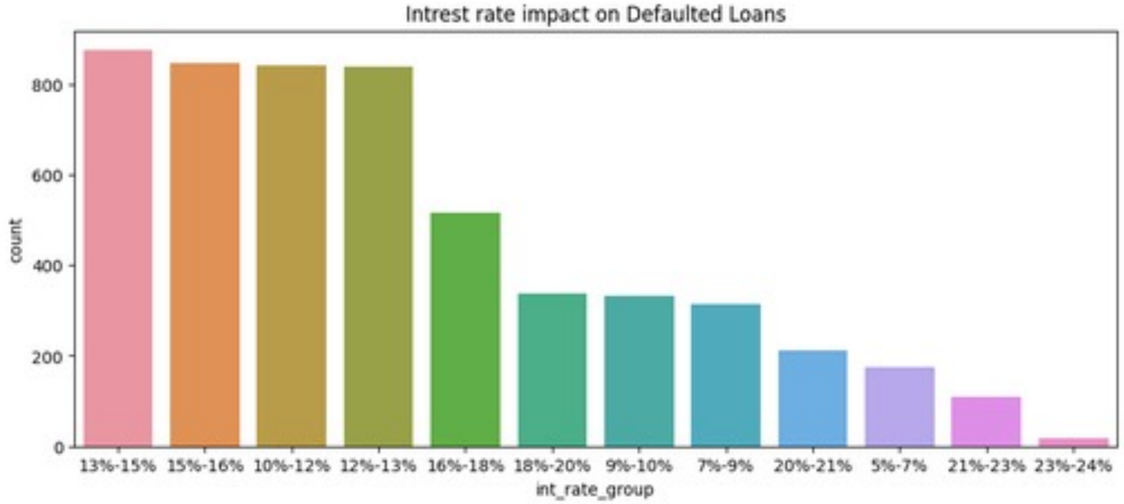
Annual Income Impact



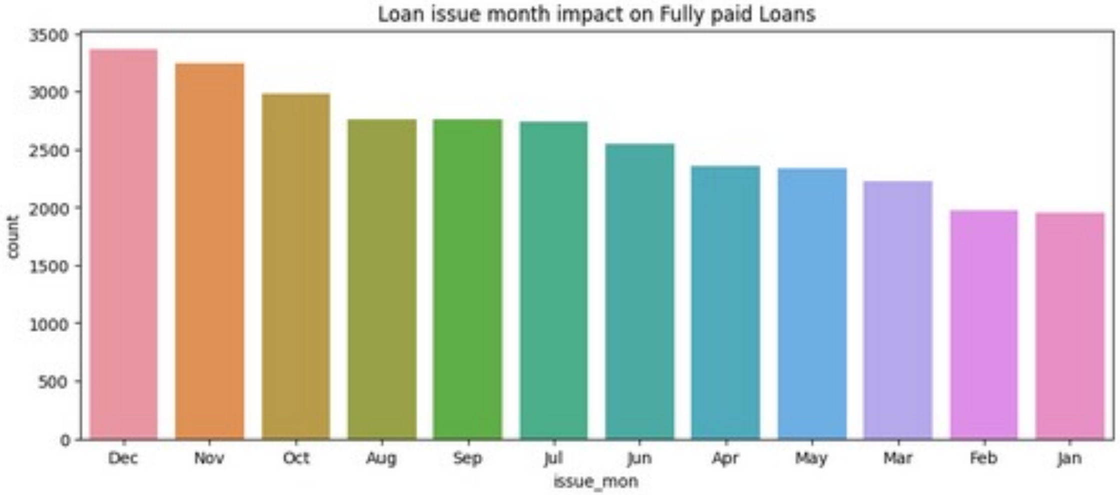


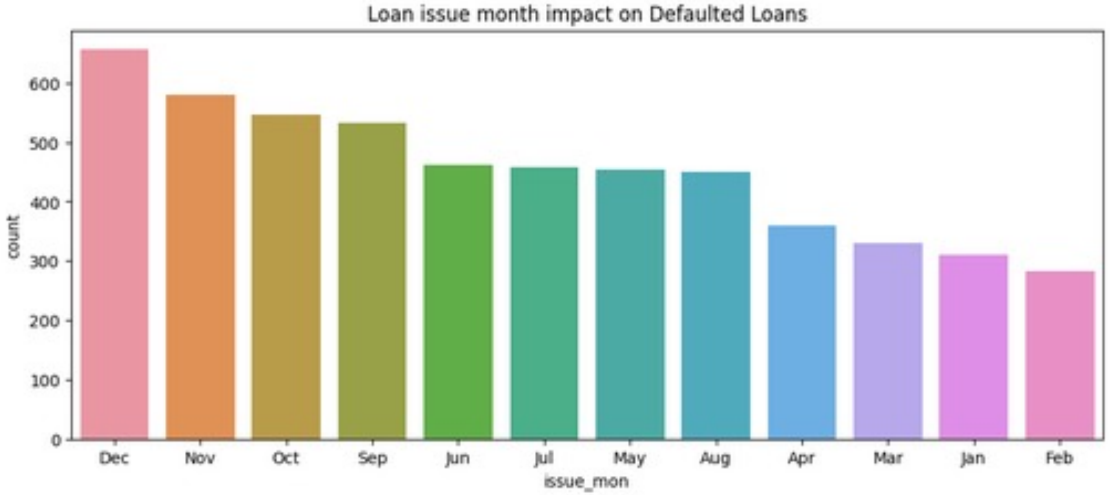
Interest Rate Impact



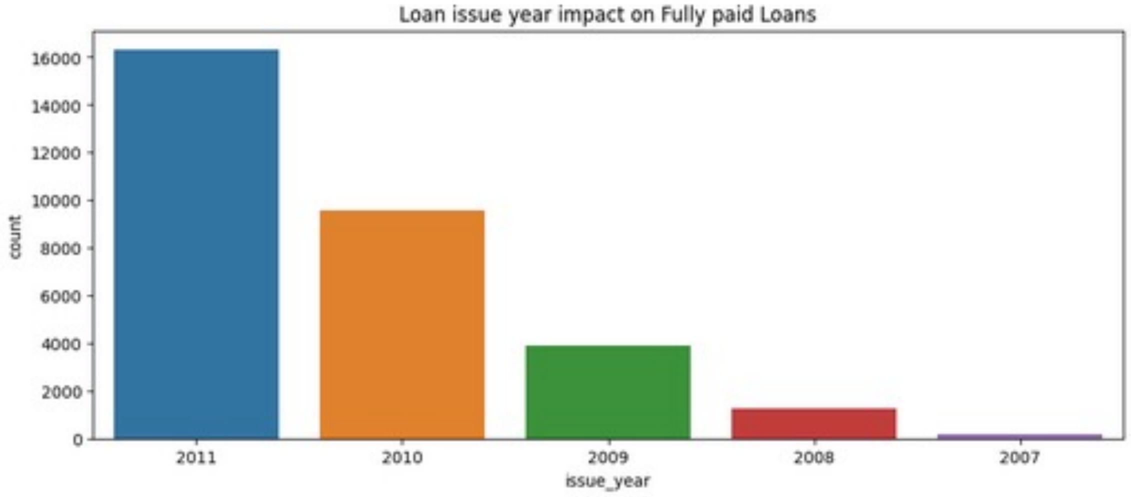


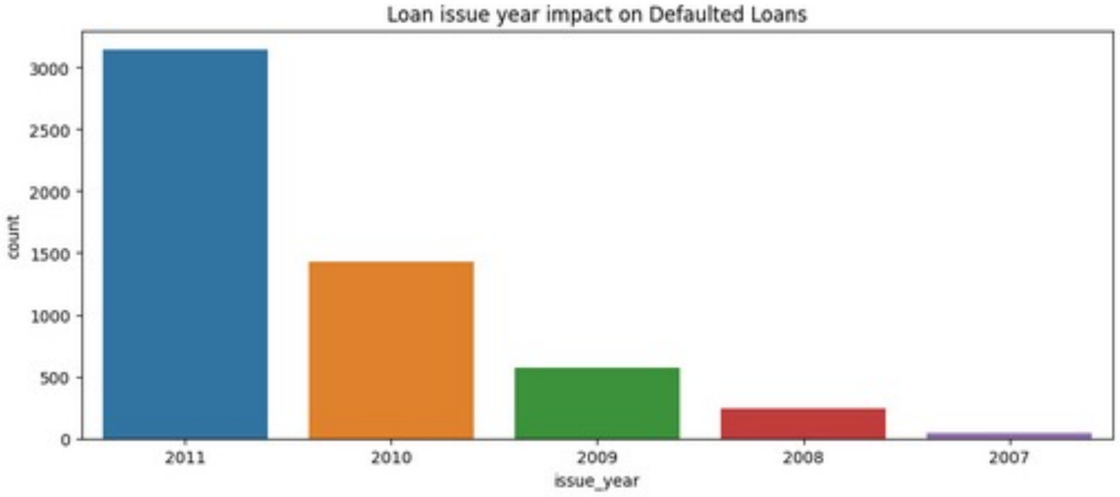
Issue Month Impact



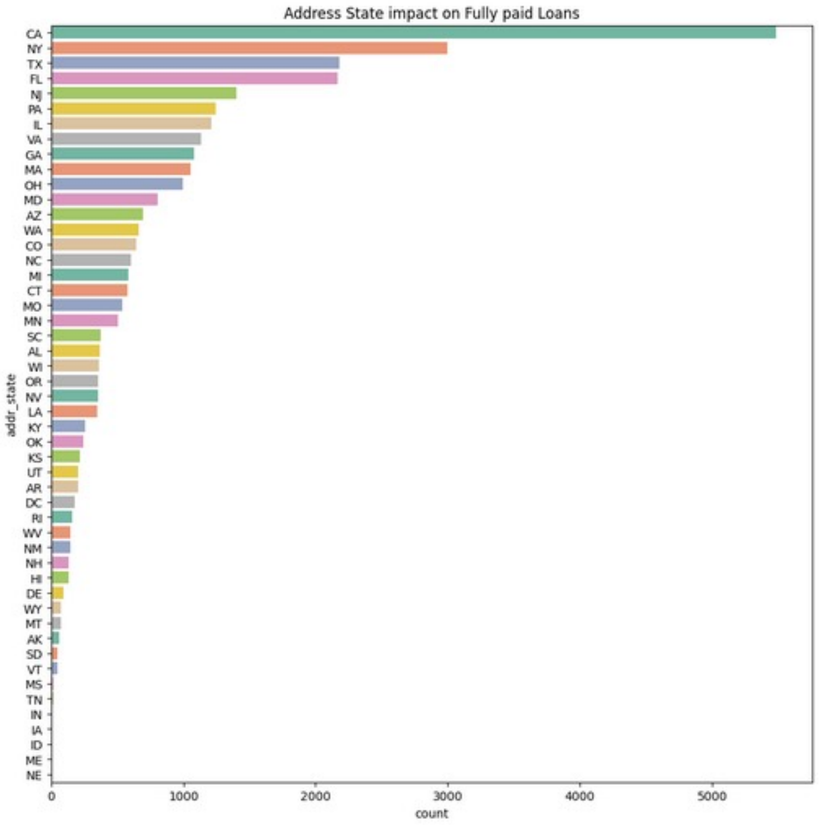


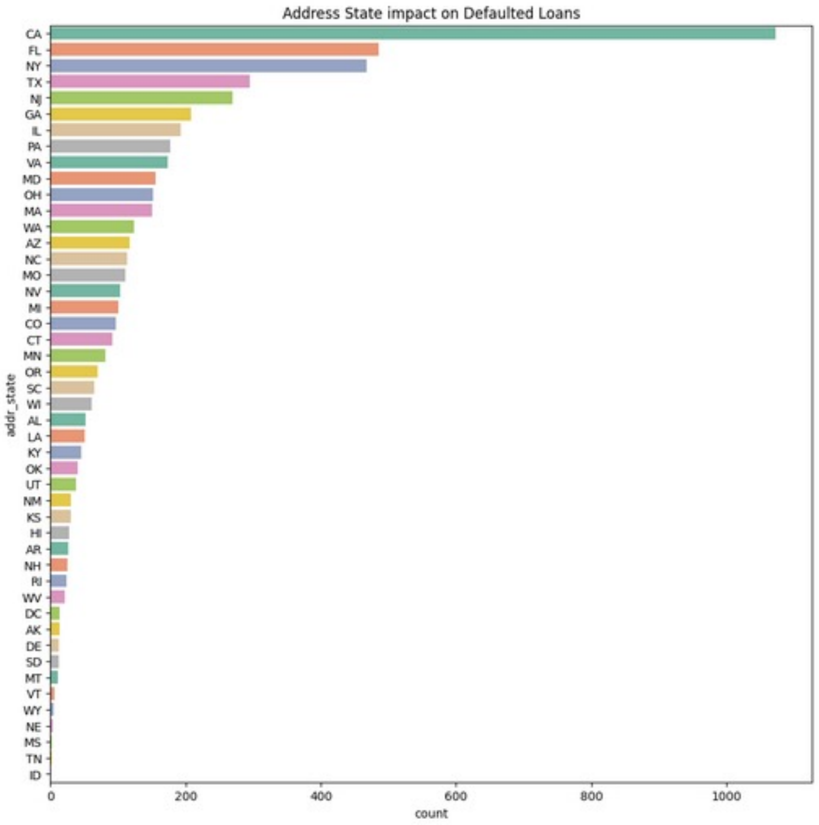
Issue Year Impact



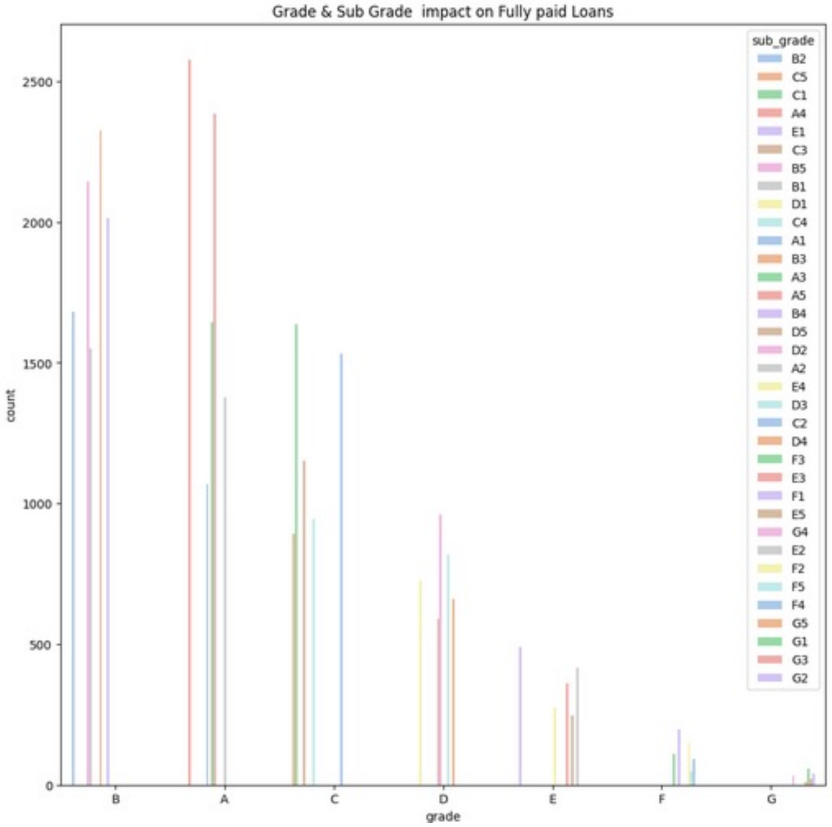


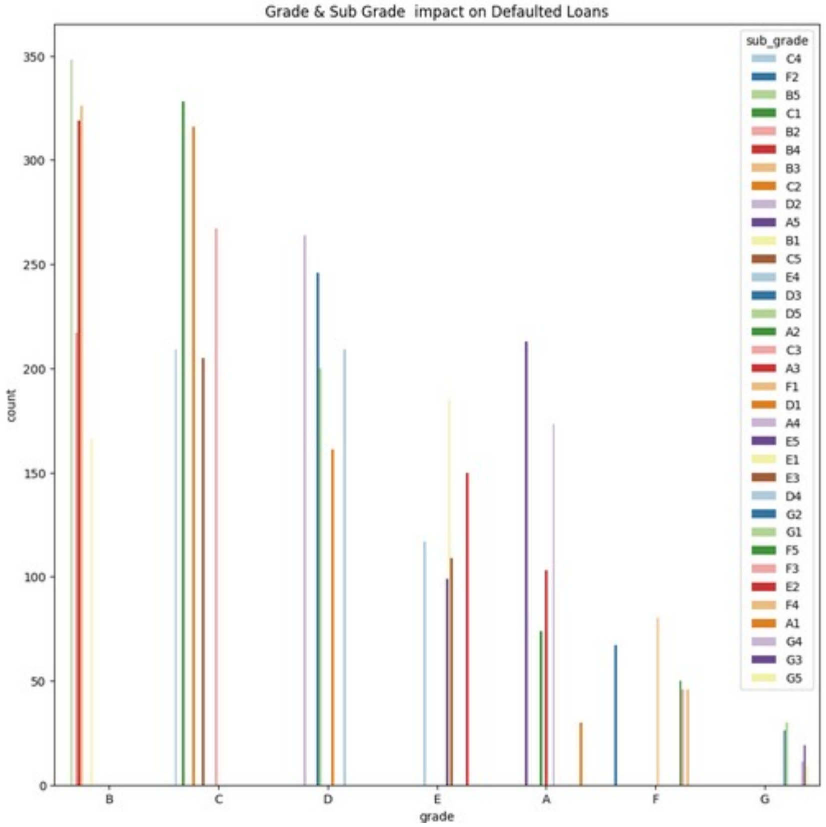
Address State Impact



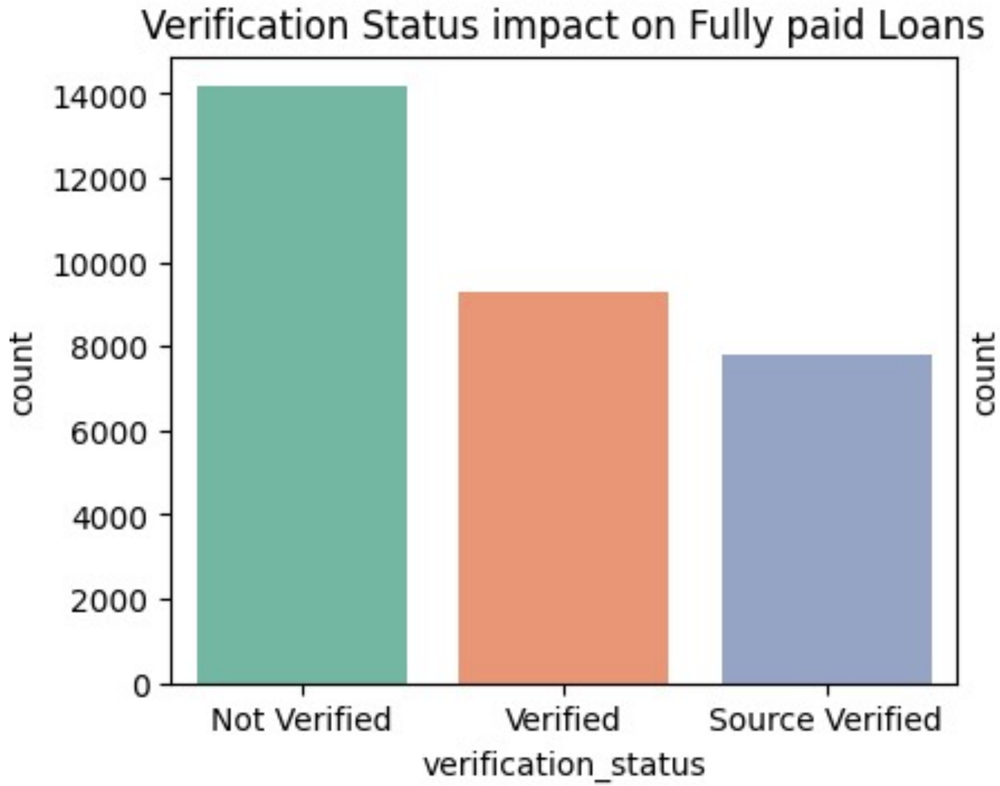


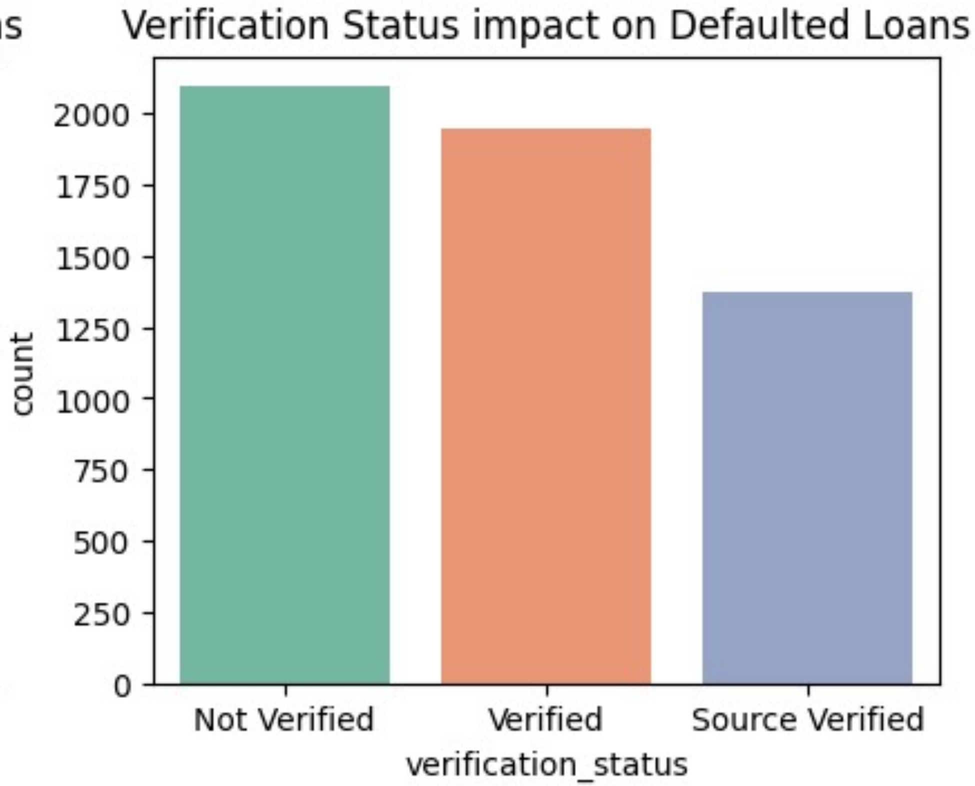
Grade and Sub Grade Impact



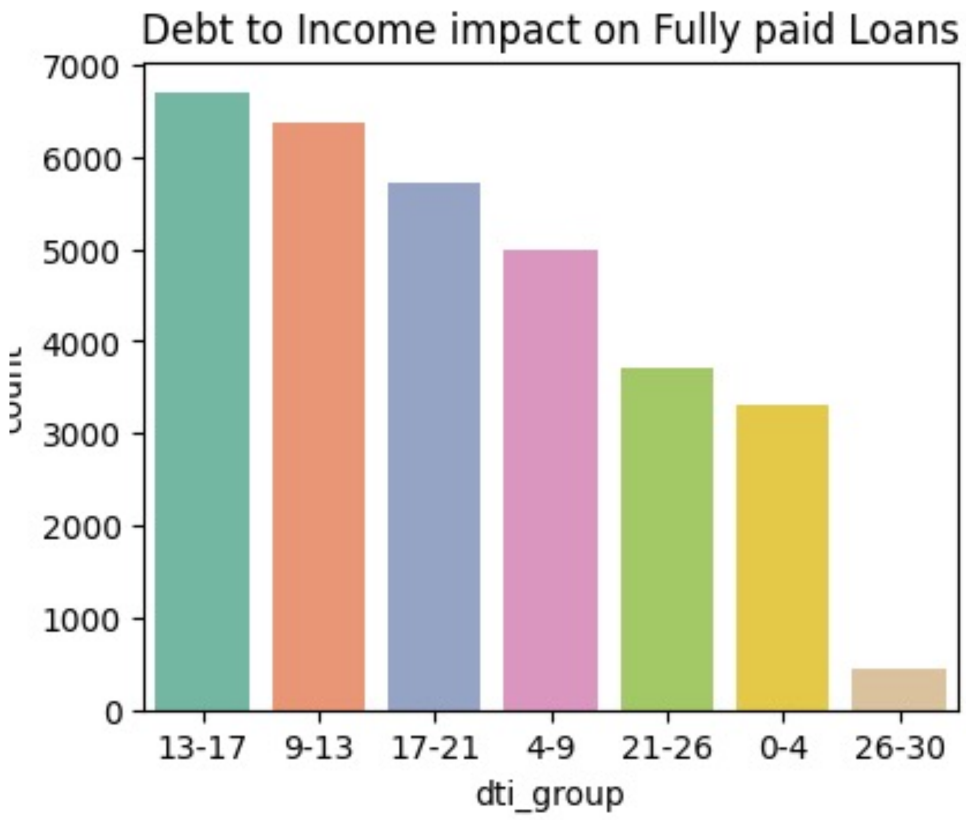


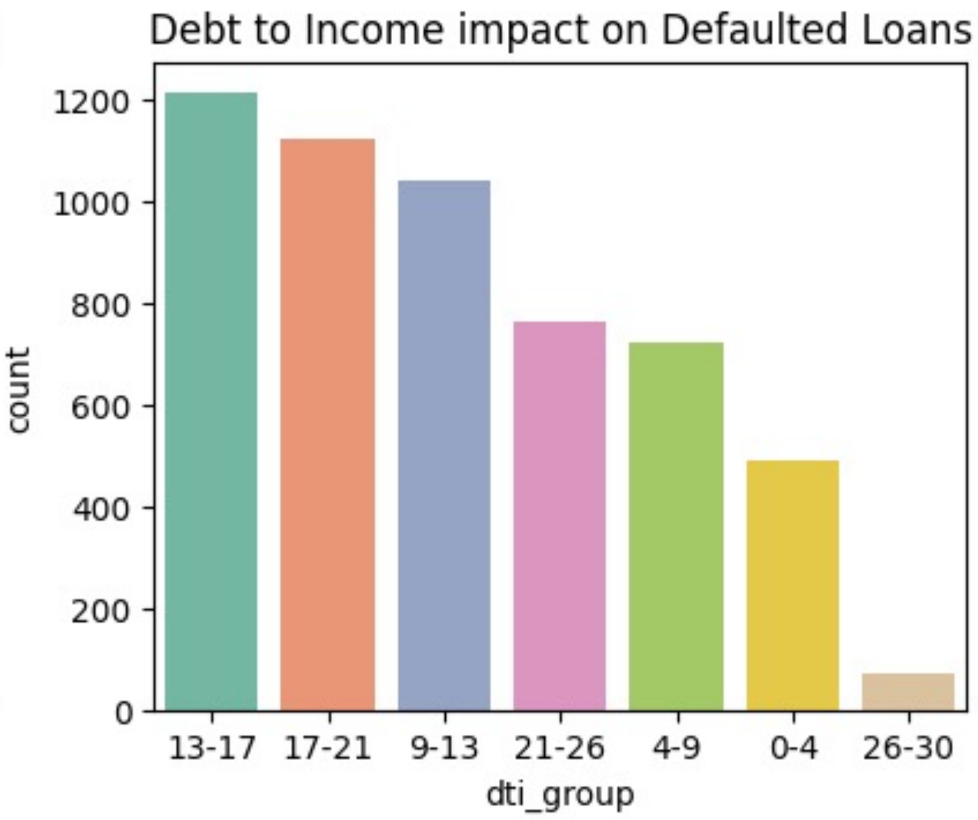
Verification Status Impact



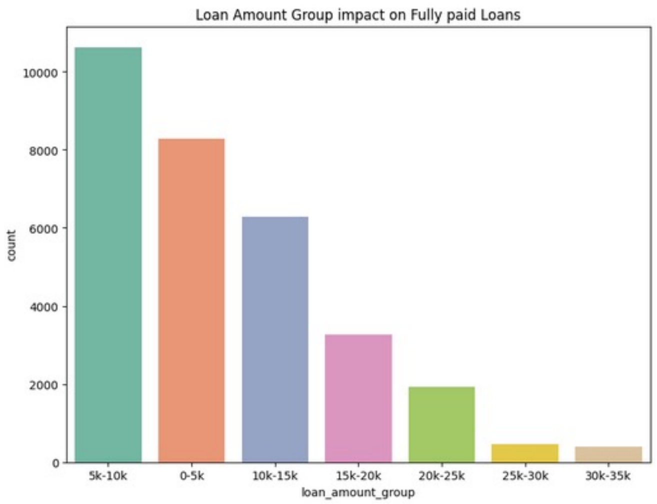


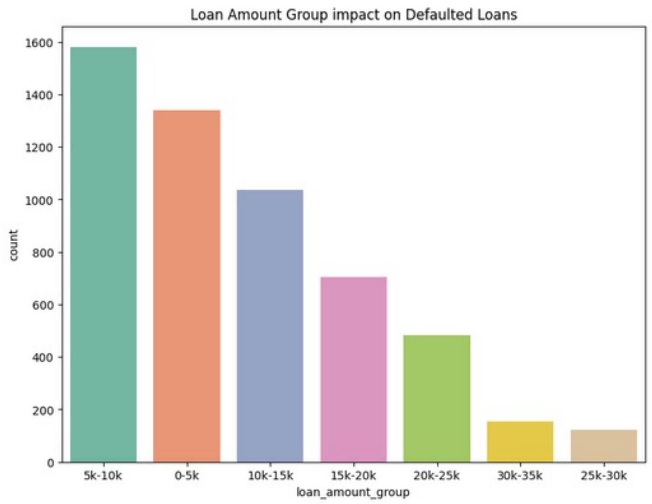
Debt to income Impact

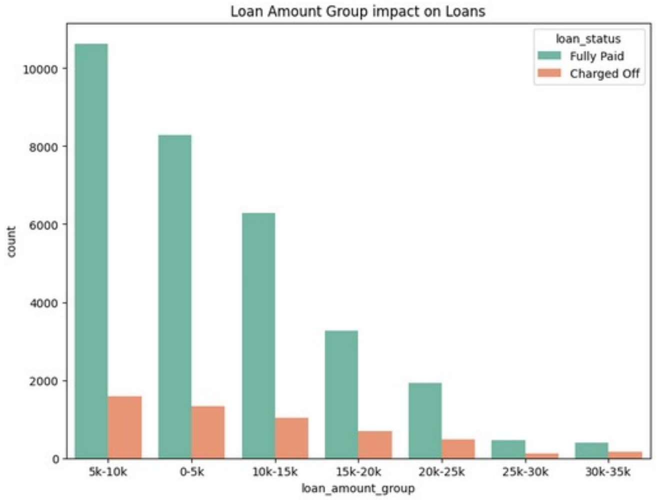




Loan Amount Group Impact







Some Observations so far

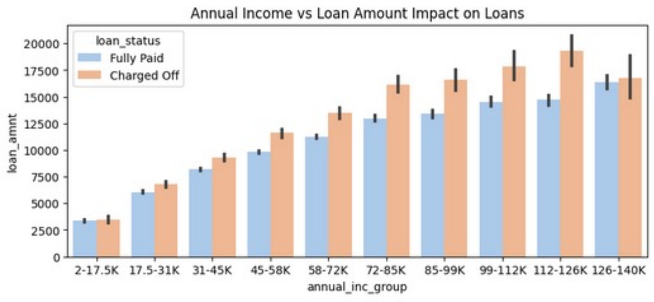
1. 10+ , <1 or No work experience Customers have mostly defaulted the loans.
2. Customers who are on rent house or mortgaged house are more likely to default than who own a house.
3. Debt consolidation is highest factor for taking loans and then defaulting the loan. So, while giving loan Approver we have to be mindful and check all necessary parameters for approving the loan.
4. 30K-70K(this fig is derived from by analysing 31-45K, 45-58K, 58-72K groups) range annual income customers are more likely to default loan.
5. 13-15% int rate loans are mostly defaulted customers. If we consolidate four buckets(13-15%, 15-16%, 10-12%, 12-13%) we can infer and say 10-16% int rate loans are more likely to default the loan.
6. Dec Month loans are most defaulted loans, Cusomers who take loan in Sep, Oct, Nov, Dec are more likely to default , may be they are taking loans to pay off other debts as they are nearing to year end.
7. Loan year pattern suggests us , year on year default loans are increasing and loan applications are increasing. At 2011 year more number of default loans occurred. Financial Crisis / Recission or other global factors might be reason for that.
8. Though California customers took more no of loans and then defaulted. Florida Customers have defaulted more loans if we consider the ratio of loans taken and defaulted.
9. B5, C1 grade loans are most defaulted loans.
10. Unverified loans(non BGV) loans are risky loans and defaulted loans. So, verification about customer helps in reducing the risk.
11. most of the customers who apply for loan falls in 13-17 dti range and they are also more likely to default the loans.
12. 5-10K loan are most defaulted as well as applied loans.

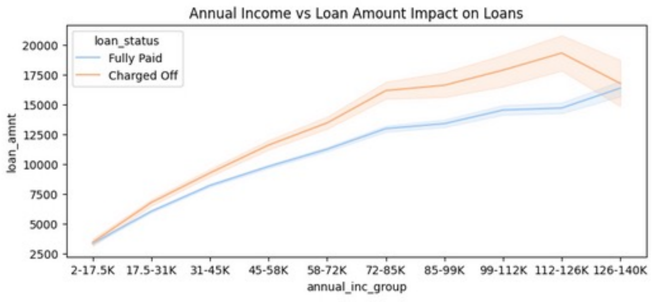
Some more Observations

1. Loans can be given to Mid Work Experience customers.
2. credit card loan purpose customers are mindful about their credit score, so loan issuing to them would reduce risk of loan.
3. A grade loans can be approved with less risk.

Bivariate Analysis

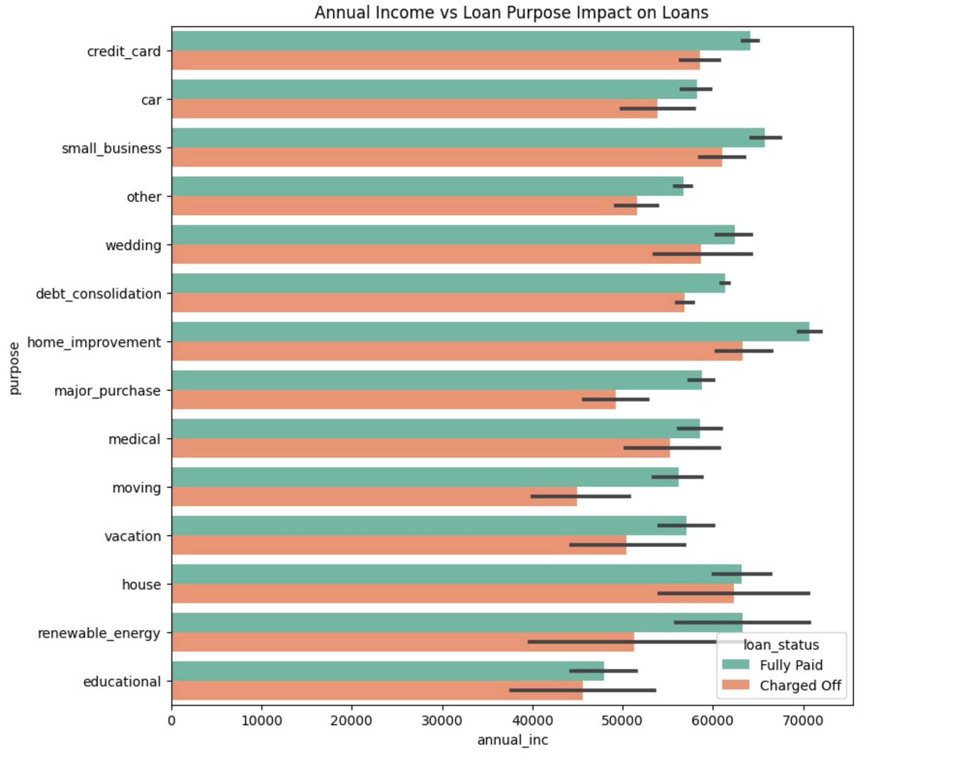
Annual Income vs Loan Amount





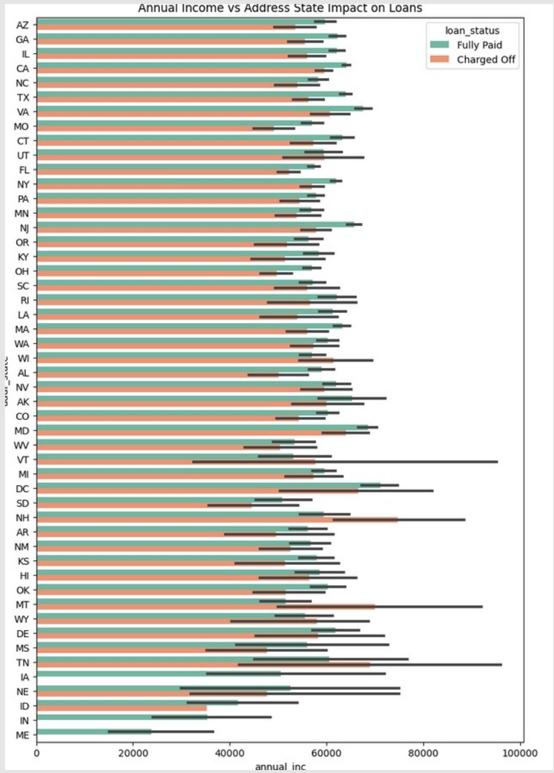
- ***Defaulted Loans have highest loan amount across all groups. 2. In case of Defaulted Loans, applied loan amount follows negative trend after 126K income range customers***

Annual Income vs Purpose



- ***Highest Annual Income Customers take loan on purpose of Home Improvement, house. We see Most defaulted loans also in these categories only***

Annual Income vs State

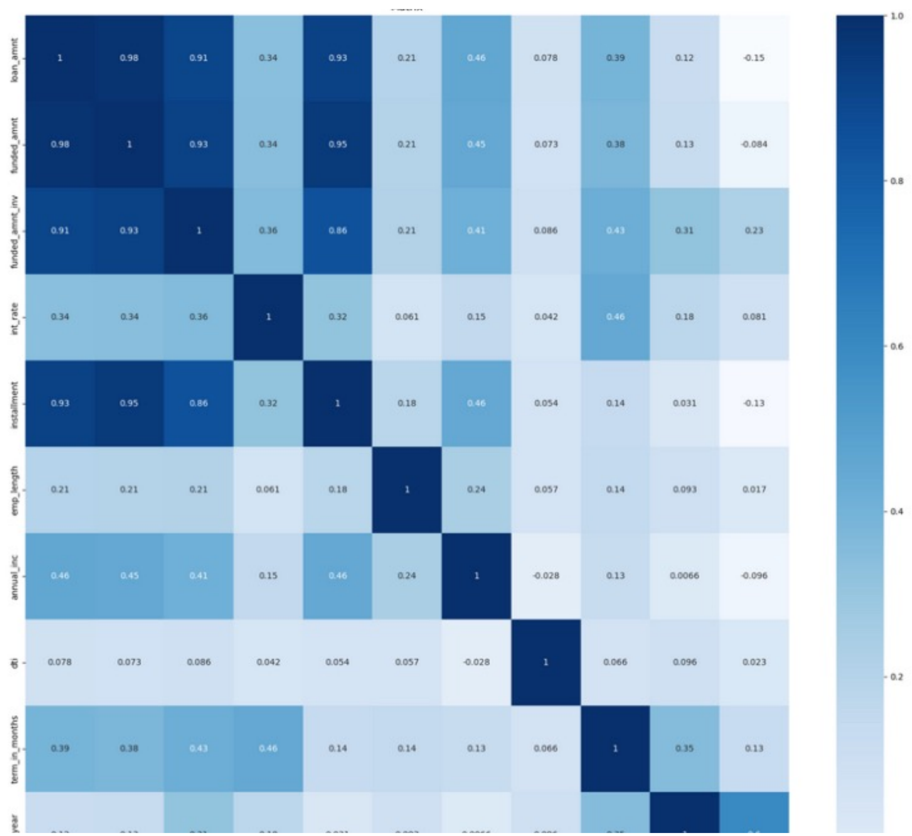


- ***There are no Defaulted Loans in IN, ME States***

Internal

Correlation Heat Map

- ***loan_ratio has positive correlation with issue_year(means a year must have been targetted for growth, or YOY targets have been revised and are upwards***



Thanks