# Lending Club Case Study

Collaborators –
Prashant & Manasa

# Problem Statement

Urban consumer finance company approves loans based on applicant's profile. 2 risks involved in decision-making

If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

If the applicant is not likely to repay the loan, i. e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

Data: past loan applicants + default history. Goal: identify default patterns to inform loan decisions (approve, reduce, or increase interest rate)

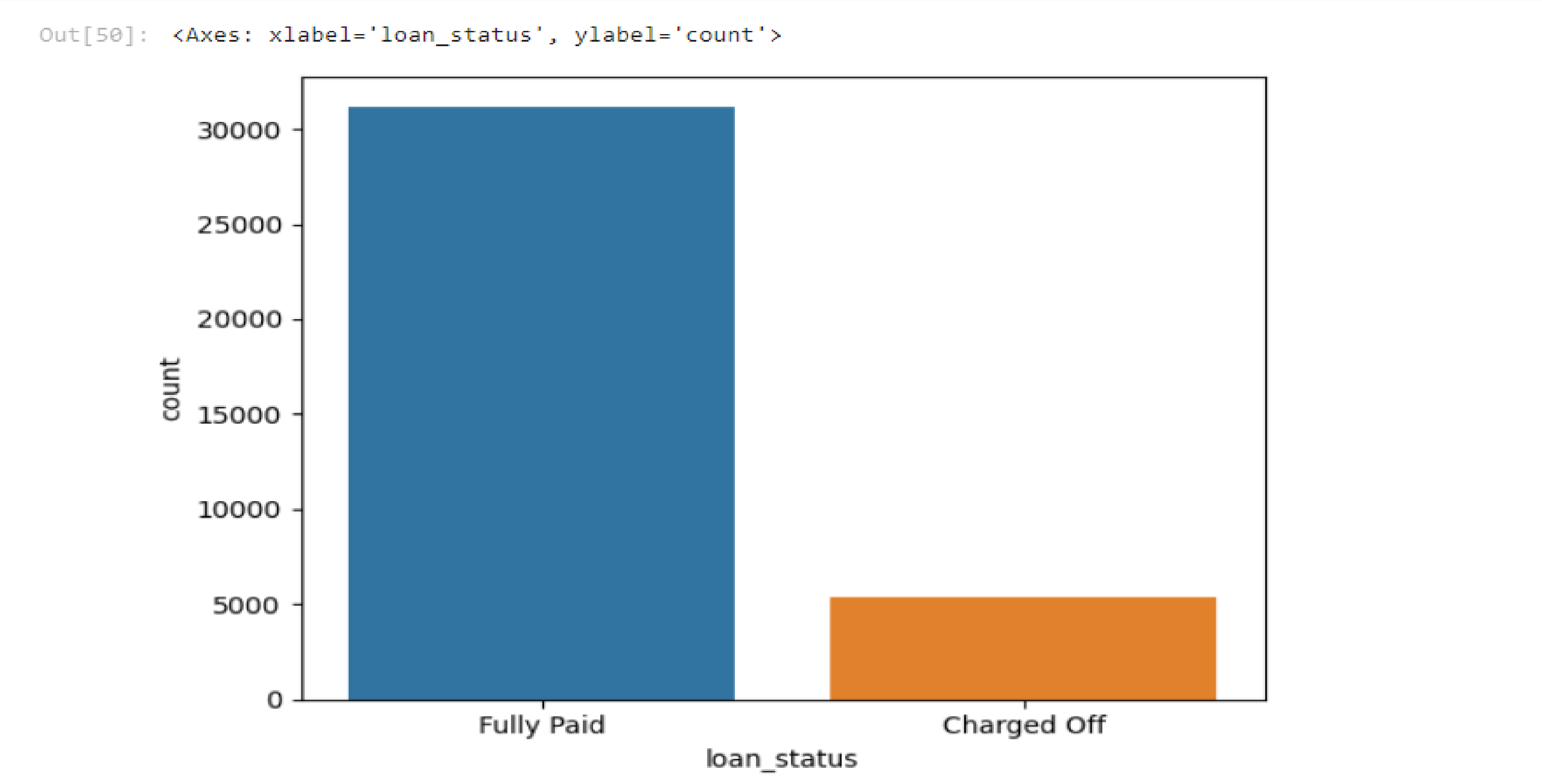
In this case study, you will use EDA to understand how consumer attributes and loan attributes influence the tendency of default. default

# Let's Start with refining our data…

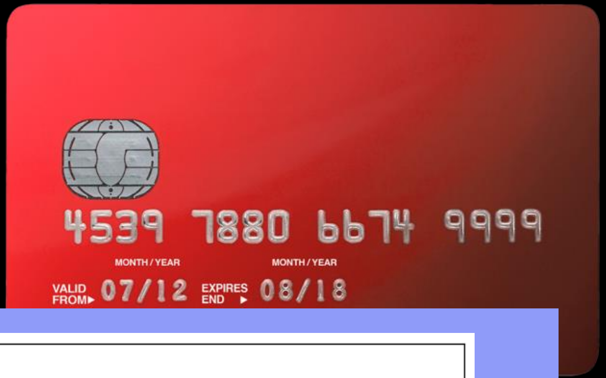
- Removing Null Values rows
- Cleaning up data by removing columns with no data
- Refilling null values in specific columns with relevant data
- Dropping unnecessary fields which won't help with our analysis
- Removing outliers from specific fields

# Next, we will look at a few Visualizations we made from the dataset.

1. Loan status count difference between Fully Paid and Charged Off loans

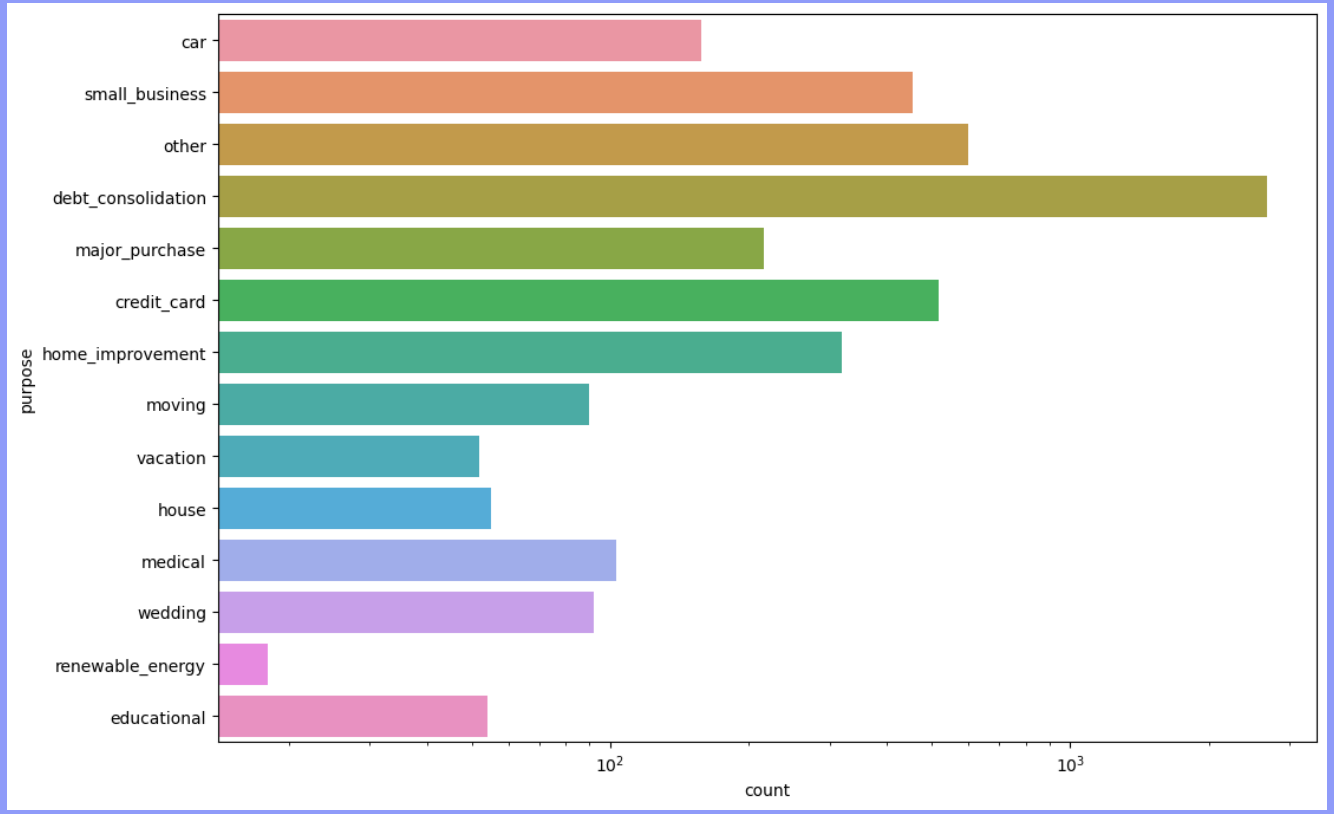


# 2. Purpose of Loan given




We can analyze this chart and see the count of all the defaulted loans taken, and the purpose for which they were taken.

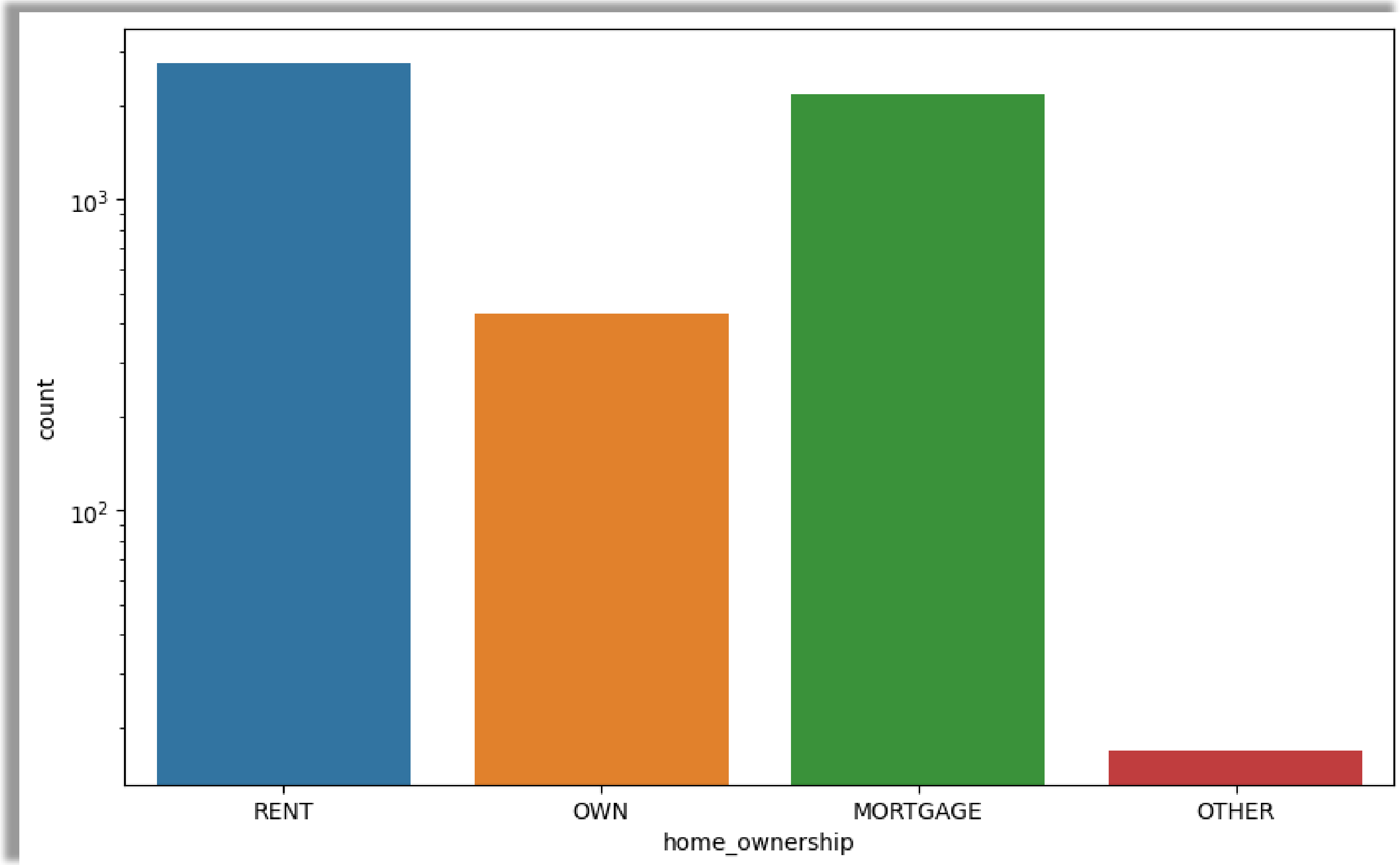
Most loans were taken for debt-consolidation i.e. loan to pay other debt.



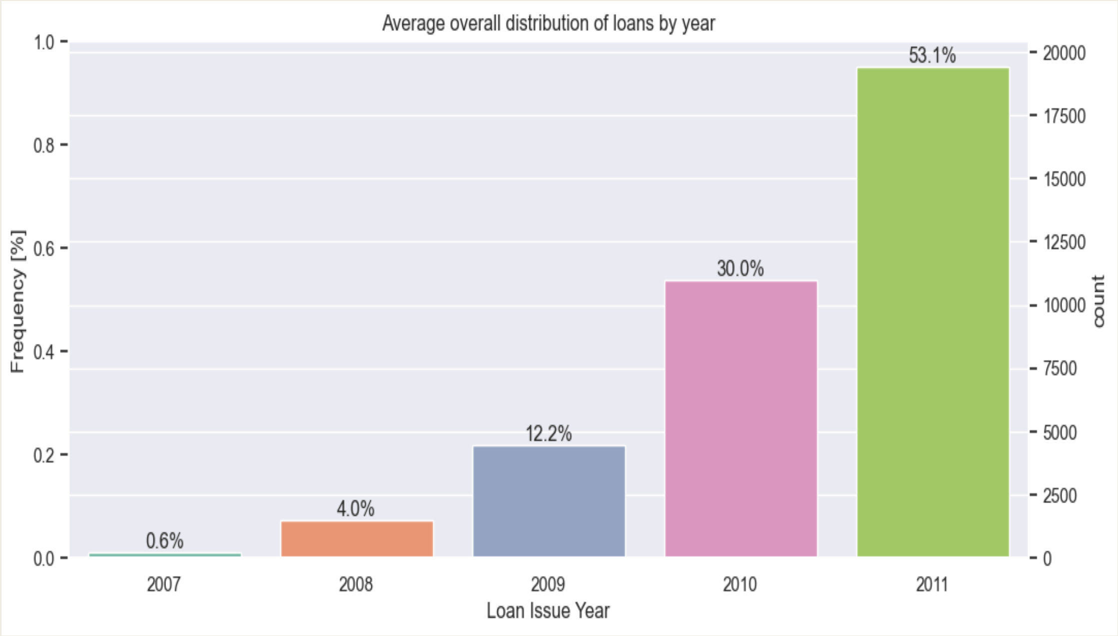
# 3. Analyzing the home ownership metrics

From this graph we can clearly see that most defaulting customers are those who either Rent or Mortgage their homes.

Customers who own a home are less likely to default.



# 4. Average overall distribution of loans by Year

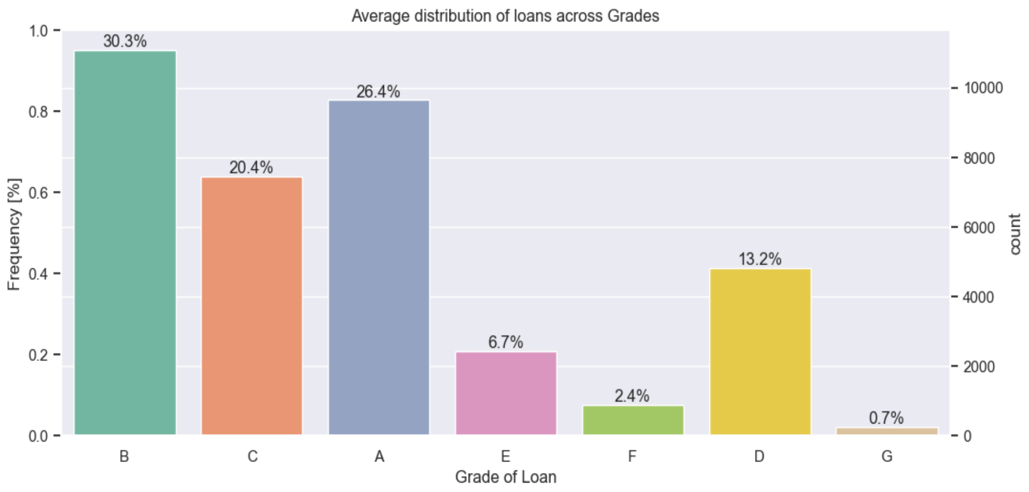



Most number of loans that defaulted were sanctioned in the year 2011.

This is an increasing trend that might be attributed to more number of people opting for debt post the 2007 US Stock market collapse, and unable to repay back the loans.

## 5. Avg Distribution of Loan across grades

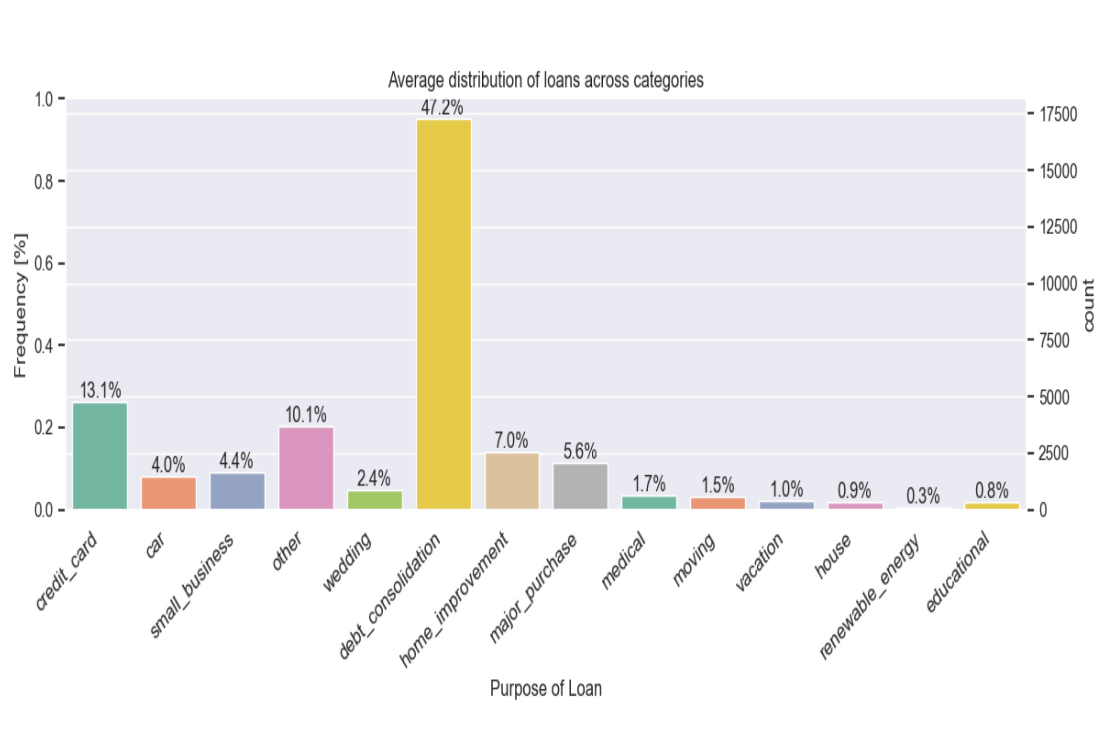
Loans of Grade B had highest probability of default.




Distribution of Loans

## 6. Plotting between Purpose of Loan and Percentage count

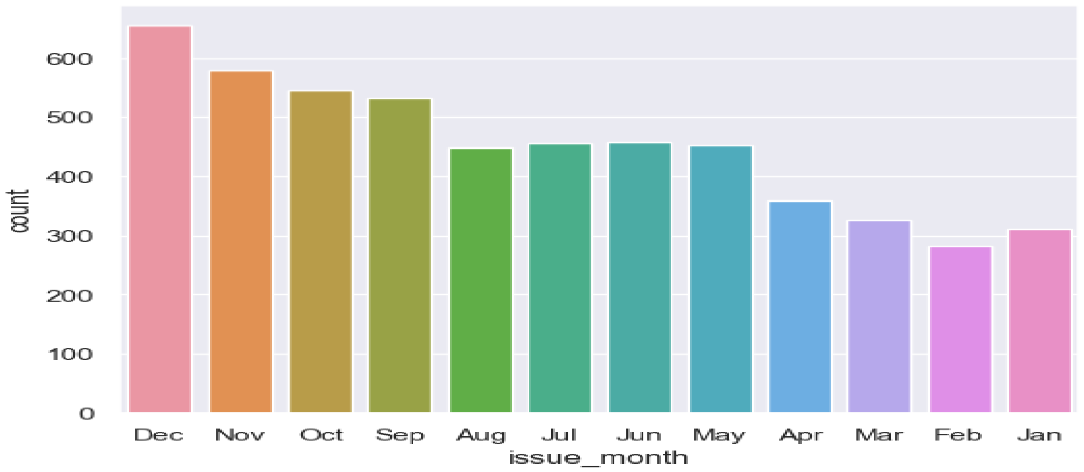
Debt Consolidation was the most notorious purpose as it resulted in maximum number of defaults.

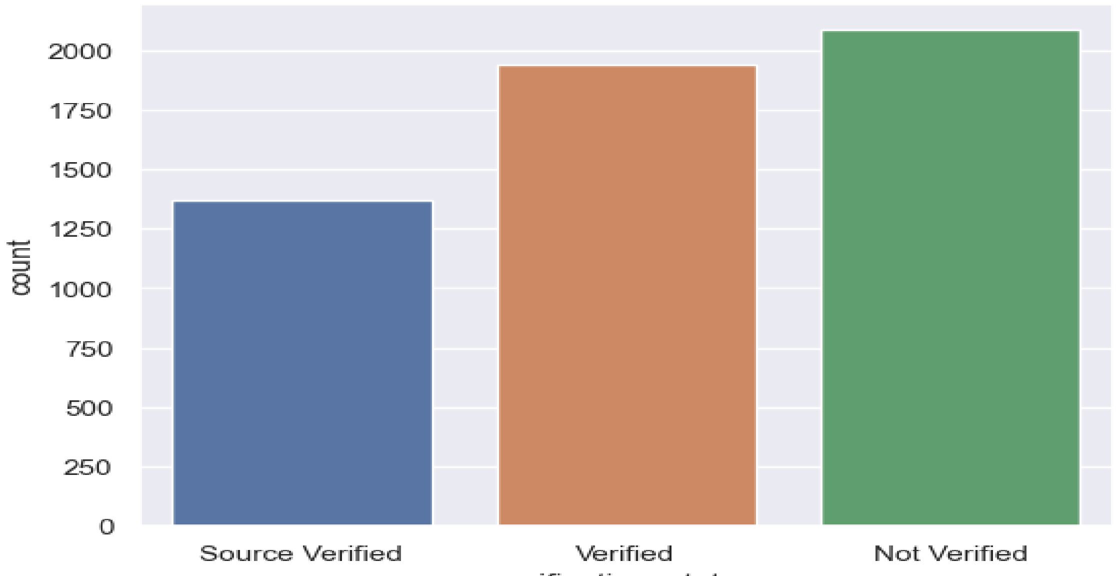


# 7. Loan verification status and time frame plotting

In these two charts we can see the following trends:

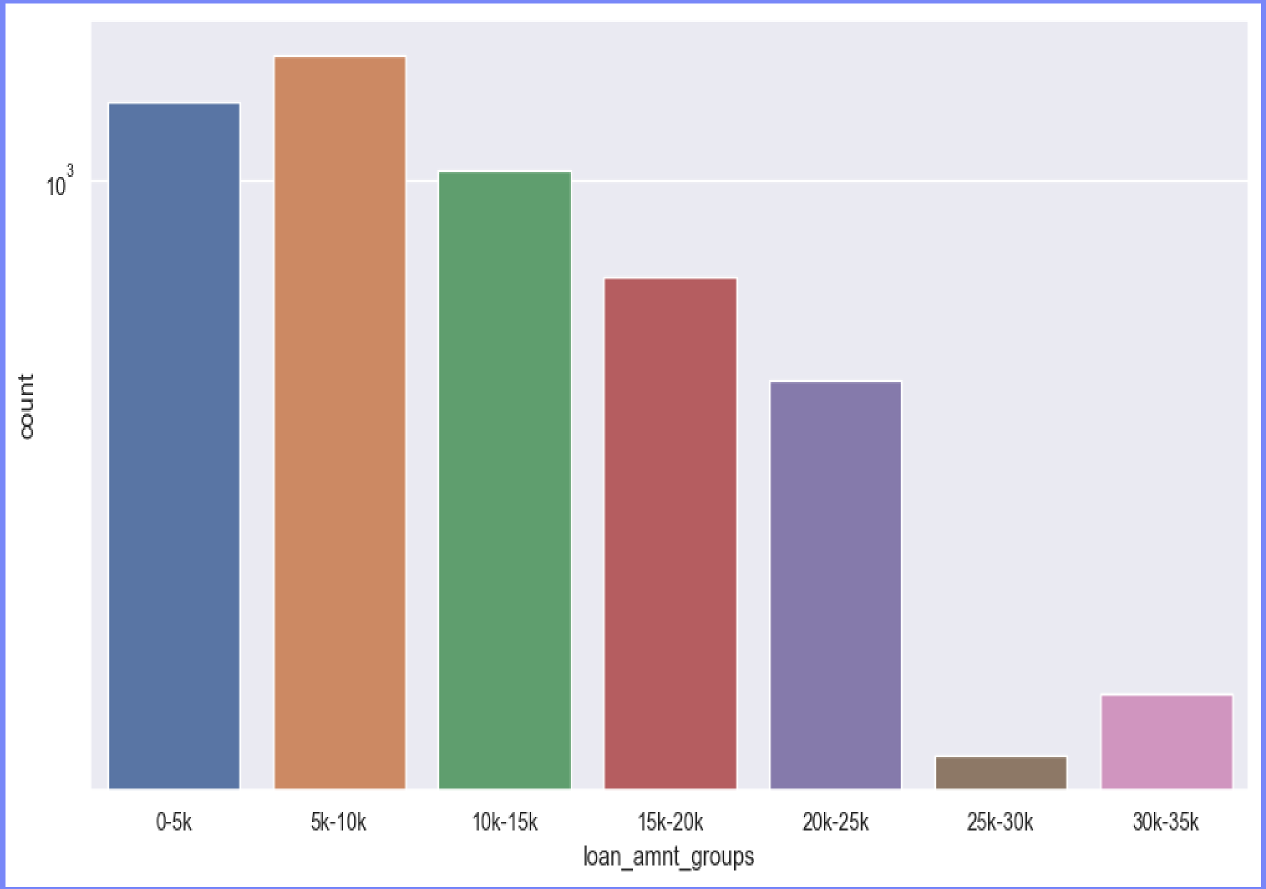
1. December is the month when most issued loans defaulted.
2. Verified and Not Verified loans were more prone to default compared to source verified loans.



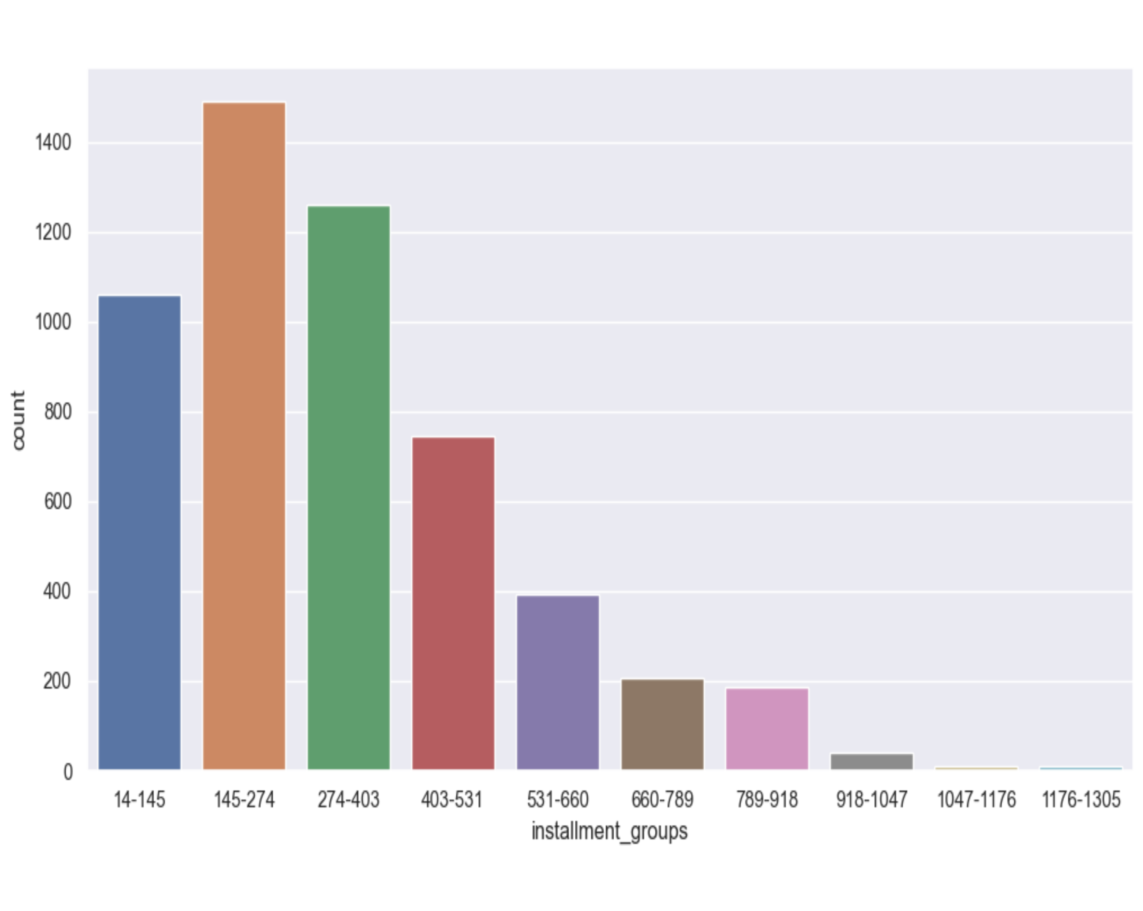



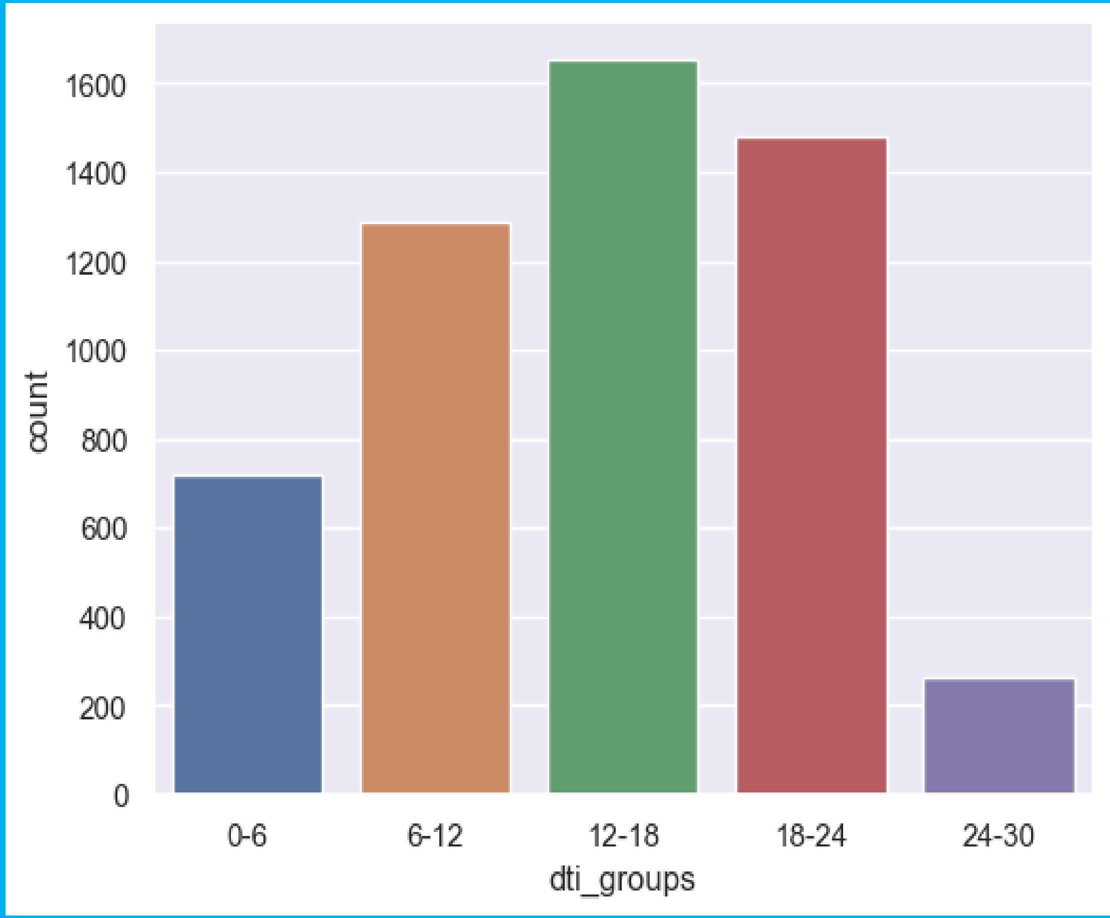
# 8. Count of defaulted loans by sanctioned loan amount

Most number of loans that defaulted were smaller loans in the range of 0k-25k, peaking at 5k-10k range.

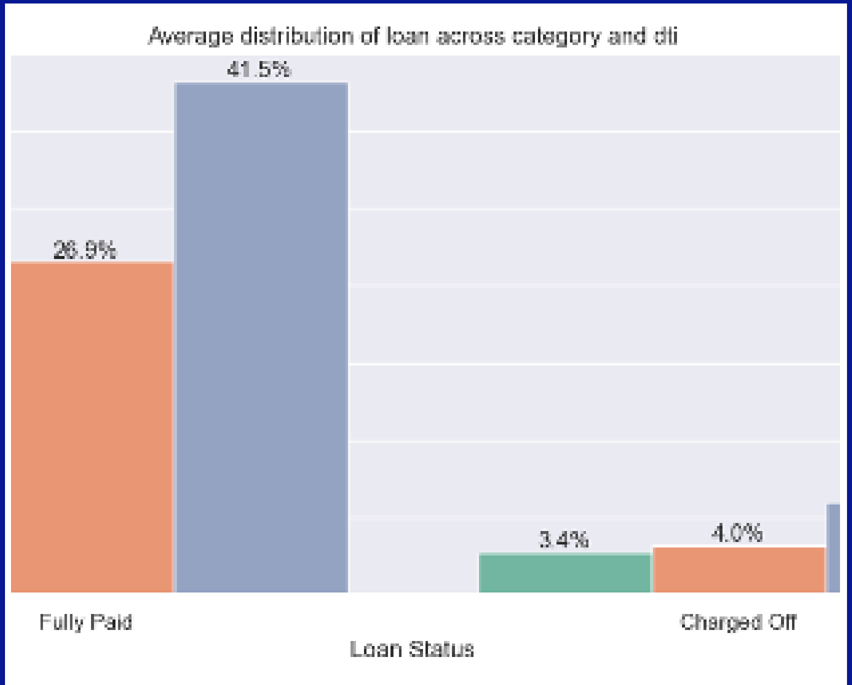


# 9. A few more charts containing Categorical analysis on other fields



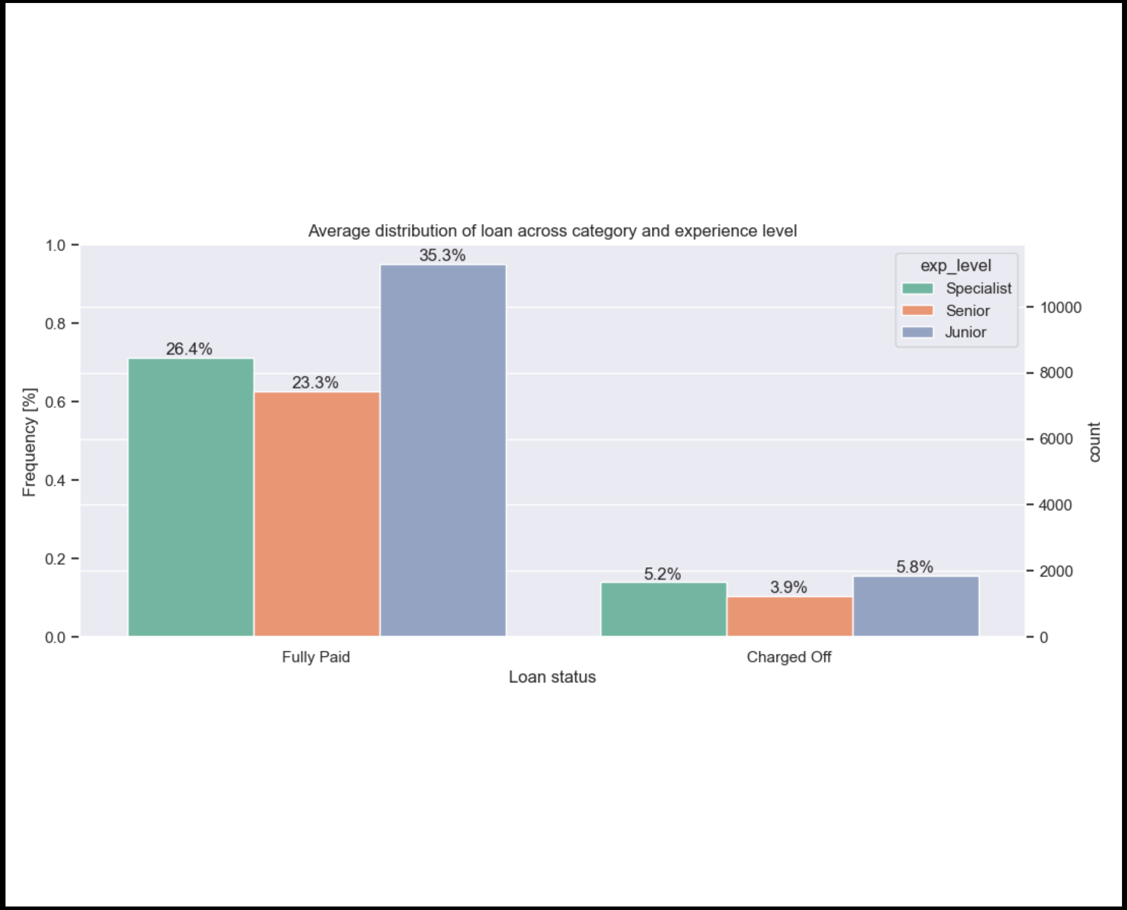



# We analyzed the dataset further by bifurcating these into simpler categories




Loan Category Vs DTI

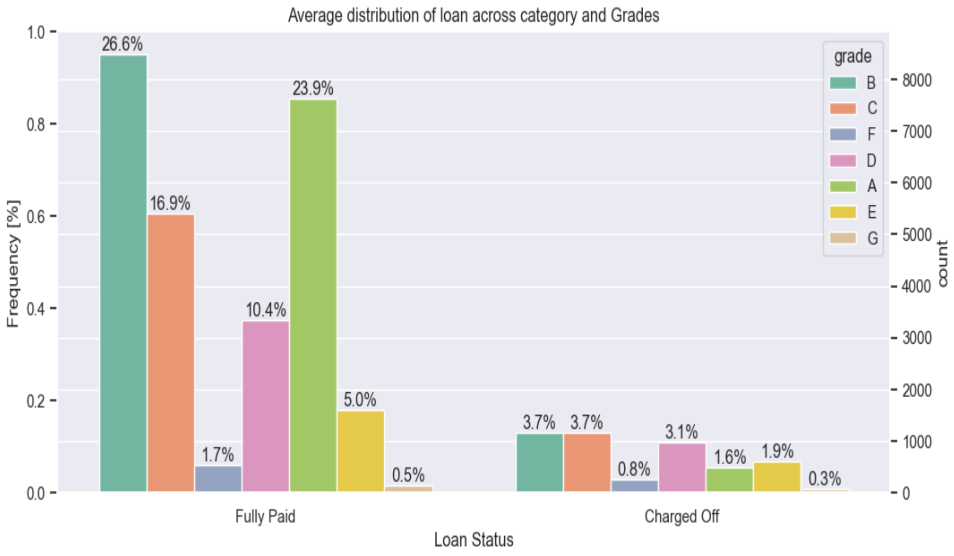
Loan category vs
Work Experience Level



# Loan Category Vs Grades

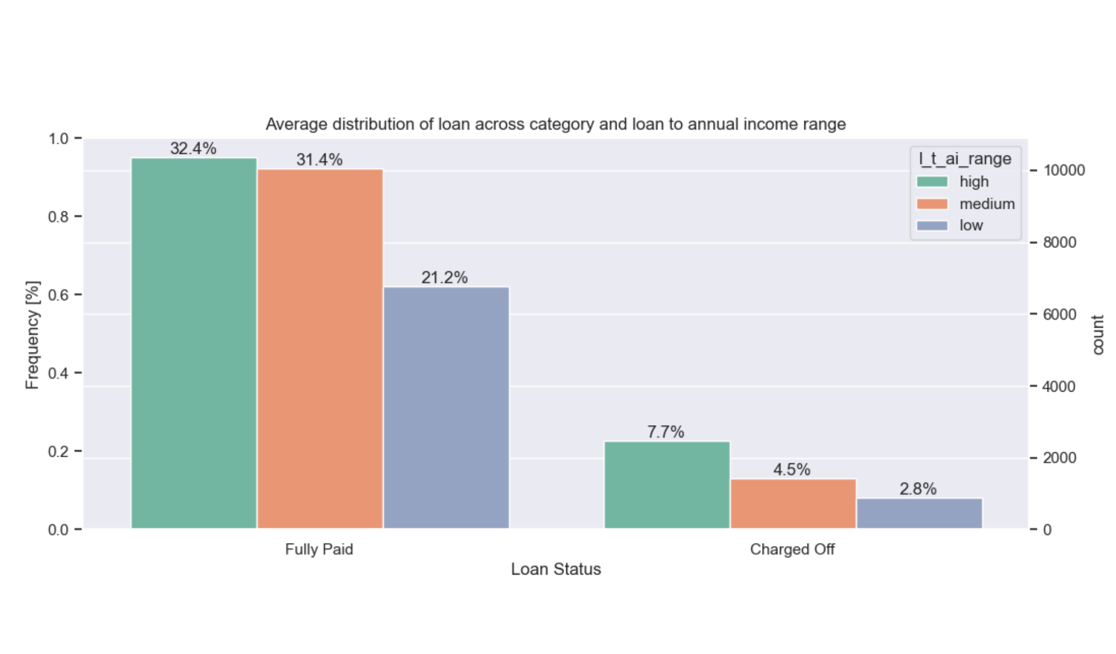
In this bivariate analysis we can see that most fully paid loans were also Grade B.

Grade E loans showed a skew towards defaulting loans.




# Loan Category Vs Loan to Annual income ratio

Lower loan to annual income ratio meant that the customer was more likely to fully pay their loan.



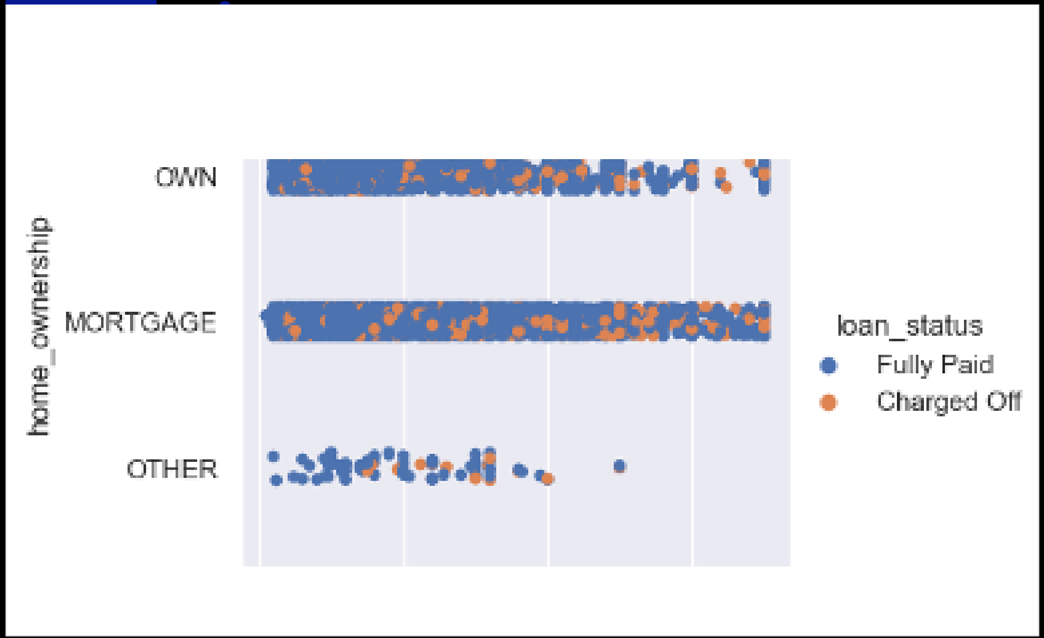
# From our analysis we reached at these defaulters' characteristics

- Customers who have an income of range 31201-58402
- Customers that paid back at interest rate of 13-17%
- Customers who are on 'Rent'
- Customers with 20-37 open_acc
- If the loan status is "Not verified"
- Customers employed for 10 years
- Grade is B
- Customers who took loan to clear prior debts
- Customers who took loan in December, or the year 2011

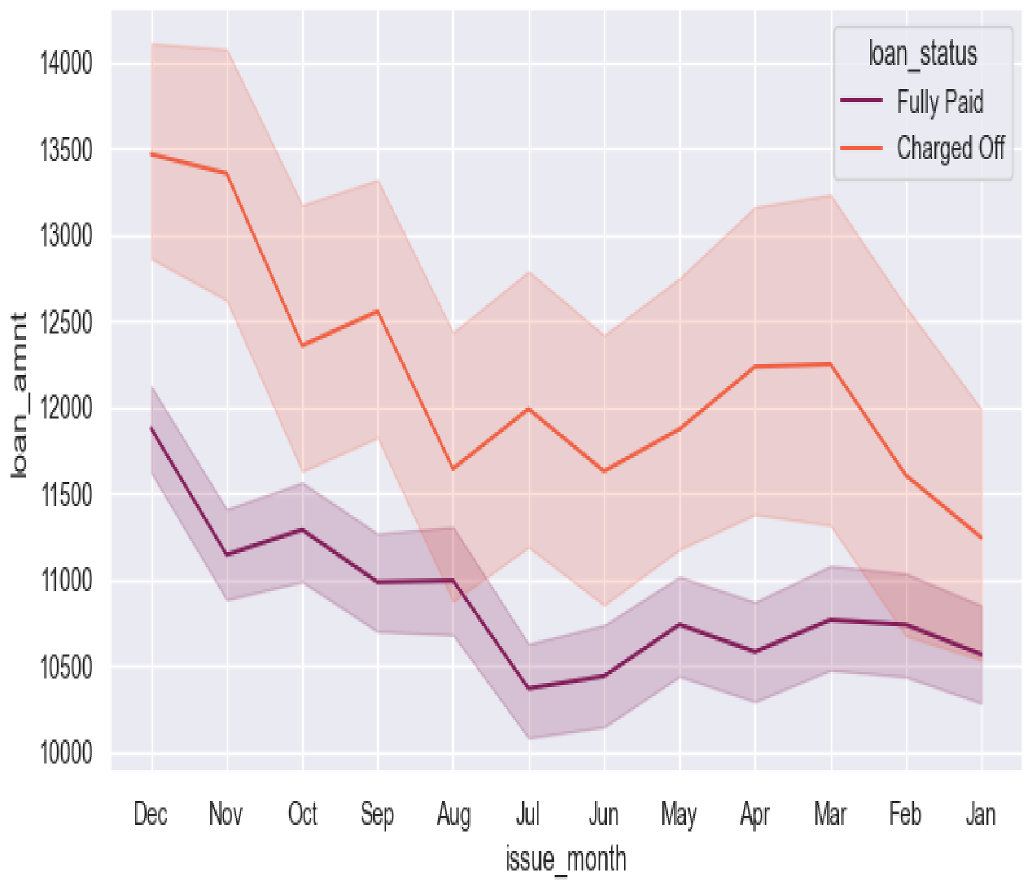
# Let's do further analysis on multiple metrics

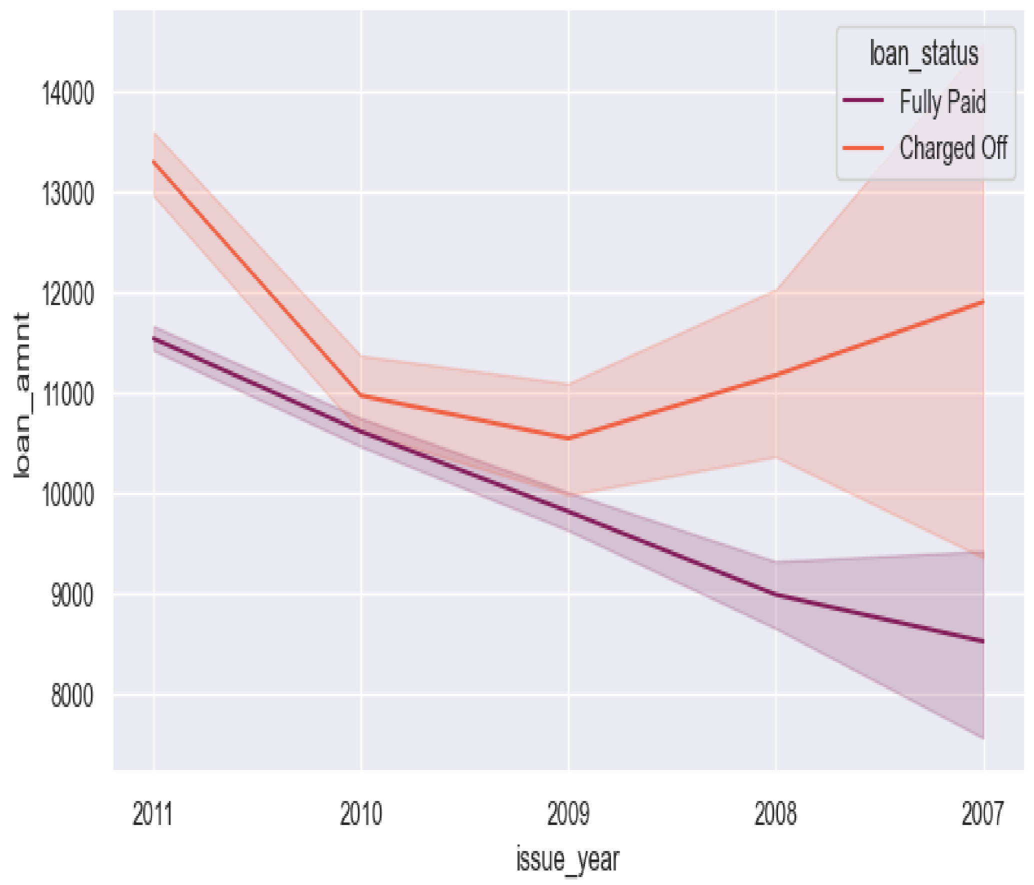
**Compared two dimensions:**
**home ownership vs. loan status**

Found that maximum number of fully paid loans were given to customers who owned their home.



# Loan issued Month & Year against loan status



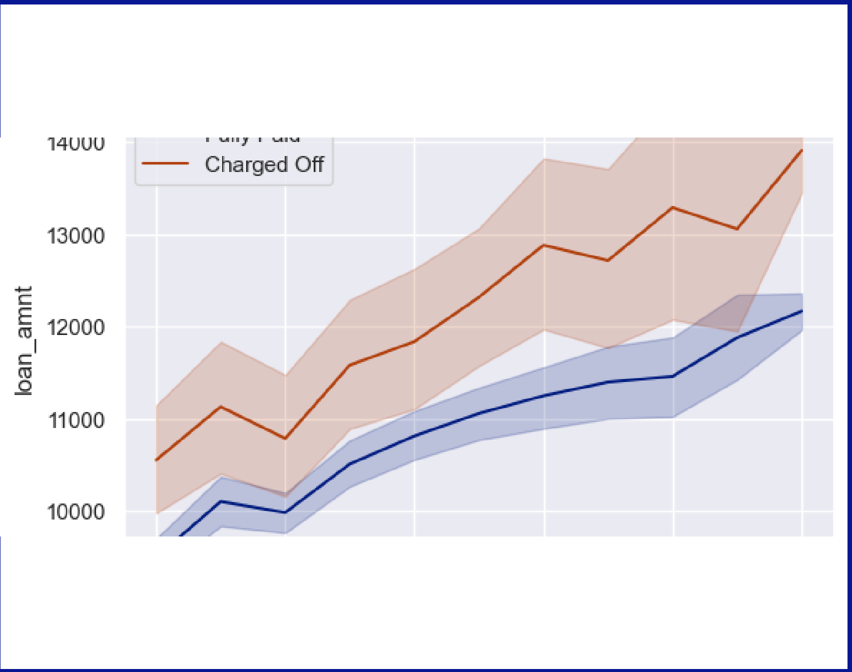



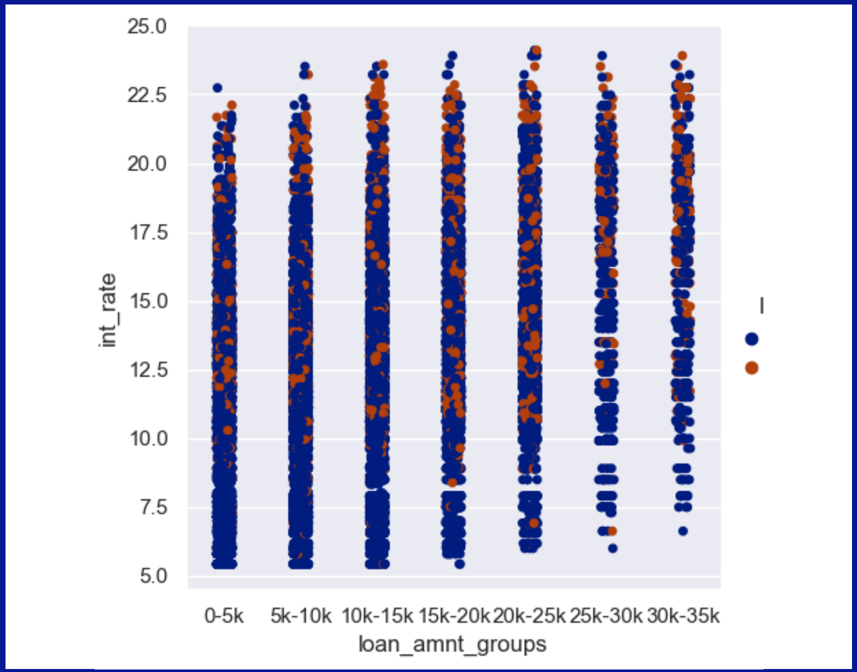
# Your title goes here

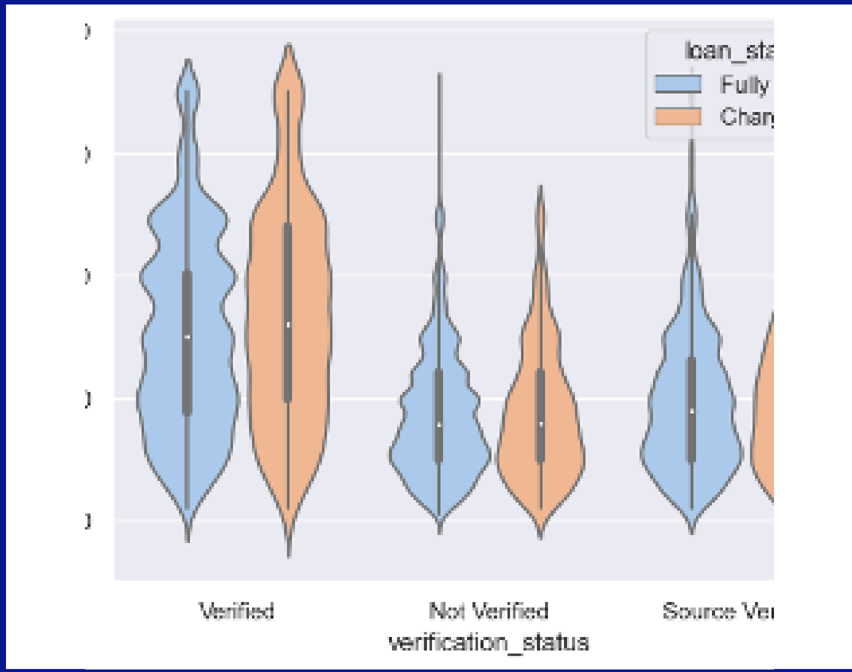
1.Verification status

2.Loan amount Groups

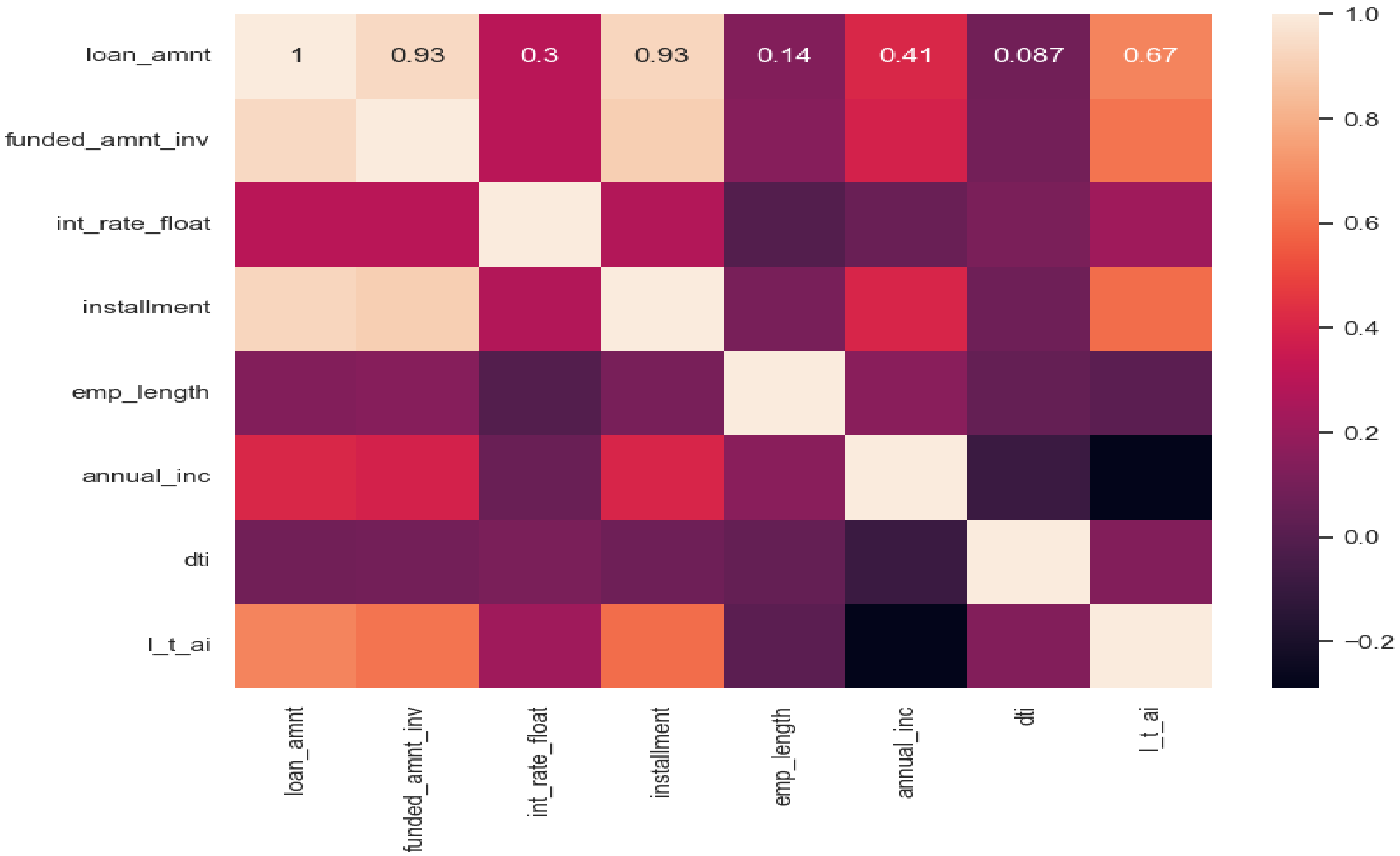
3.Loan Status
vs Employment length



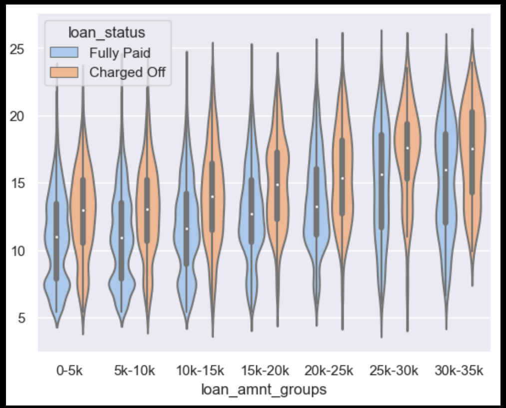



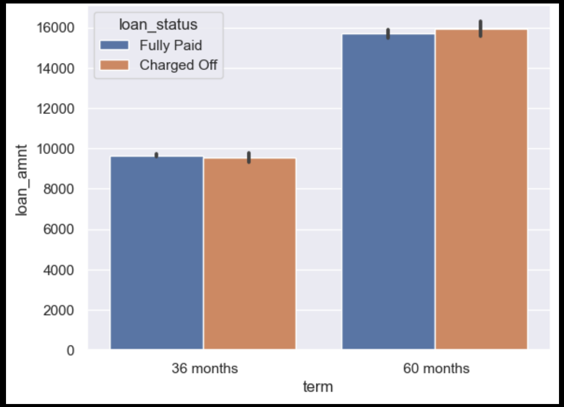



# Heat map to analyze correlation between different columns



# Plots about
# Interest rates Vs Loan Rates
# &
# Loan amount vs Term




We can also see correlation between fields like annual income and dti, funded amount and employment tenure etc

Strong correlation: interest rates & loan defaulters — Fully paid loans = lower interest rates, Charged Off = higher interest rates

No major difference was seen when comparing status of loan with the loan term

# Final Observations

From this Exploratory Data Analysis, we have observed these trends in defaulting customers:

- ✓ Those who receive interest at the rate of 21-24% and have an income in the range of 70k-80k.
- ✓ If employment tenure is 10 yrs and loan amount is 12k-14k.
- ✓ If it is a verified loan for an amount ranging over 16k.
- ✓ When the loan is for 60k-70k and the applicant took it for "Home Improvement".
- ✓ Customers who take a loan for 60k-70k and don't own a home, i.e. either Rent or Mortgage.
- ✓ When the grade is F and loan amount is between 15k-20k.

We also observed a few good traits for people who fully paid their loans:

- ✓ If home is self owned, such customers are more likely to fully pay.
- ✓ Lower interest rates were an important factor in fully paid loans.
- ✓ Lower sanctioned amounts between 5k-10k were mostly fully paid.
- ✓ Lower employment tenure between 3-6 years showed a good trend.
- ✓ Grade B loans were generally fully paid.
- ✓ Lower ratio of loan amount to annual income meant that these customers were less likely to default.

# Thank You