Lending Club Case Study

Submission

By
Md Merajul Islam
&
Bishal Sharma

The problem

Company

Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.

Borrowers can easily access lower interest rate loans through a fast online interface.

Context

Lending Club wants to understand the **driving factors** behind loan default, i.e. the **driver variables** which are strong indicators of default.

The company can utilise this knowledge for its portfolio and risk assessment.

Problem statement

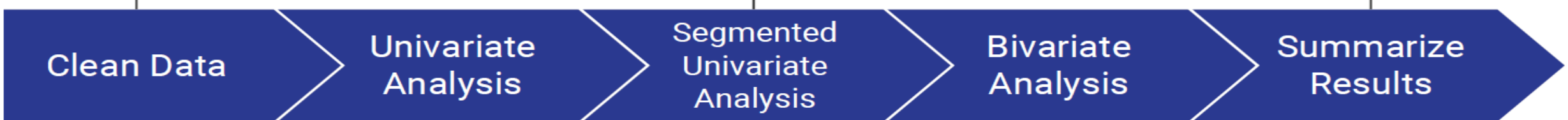
As a data scientist working for Lending Club analyze the dataset containing information about past loan applicants using EDA to understand how *consumer attributes* and *loan attributes* influence the tendency of default

Analysis Approach

> Drop columns with null values, all random values or single category value
> Convert values to proper int, float, date representations

> Analyze variables against segments of other variables
> Create derived variables

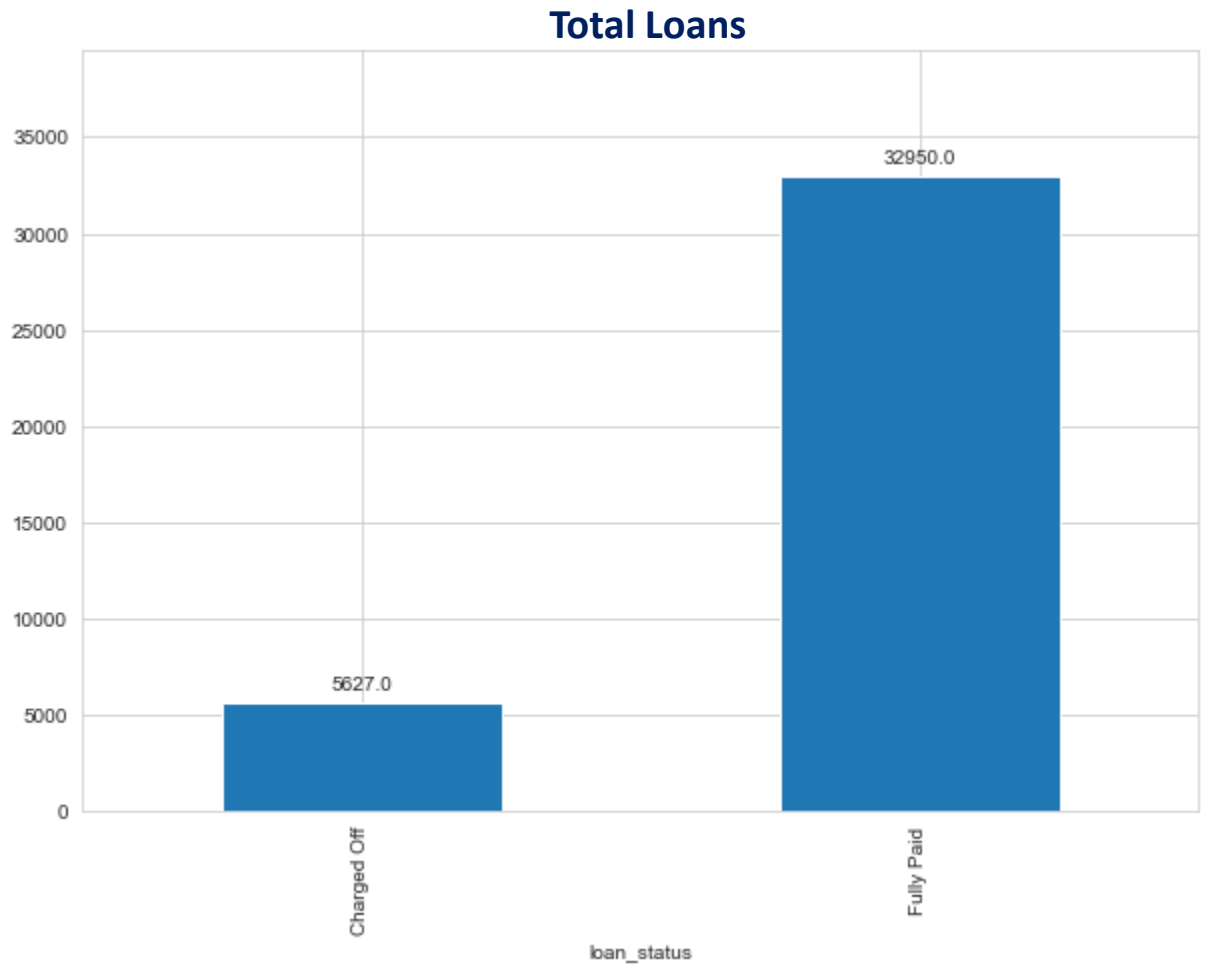
Publish insights and observations

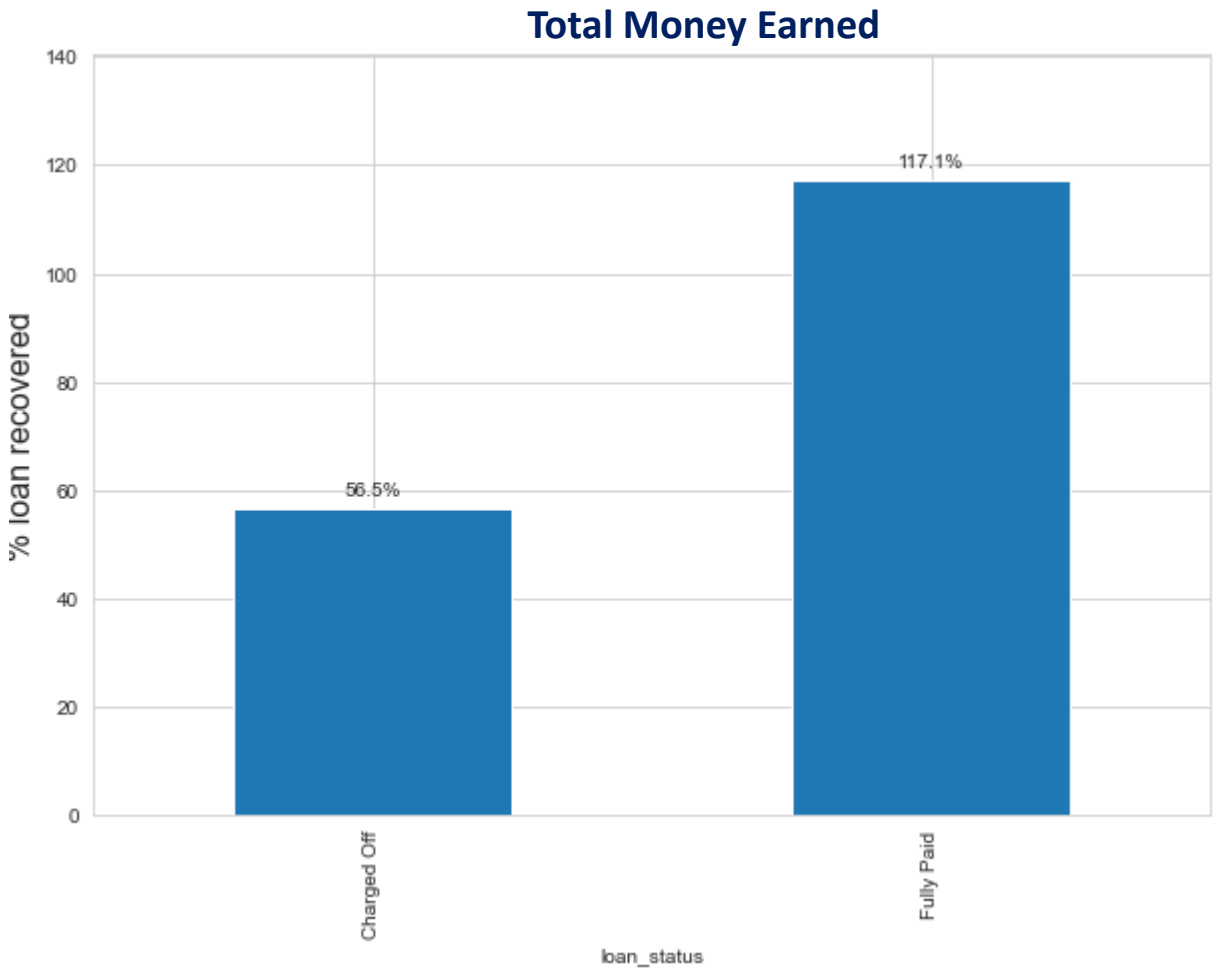


> Check distributions and frequencies of various numerical and categorical variables
> Create derived variables

> Do correlation analysis
Check how two variables affect each other or a third variable
> Analyze joint distributions

Overall Loan Status



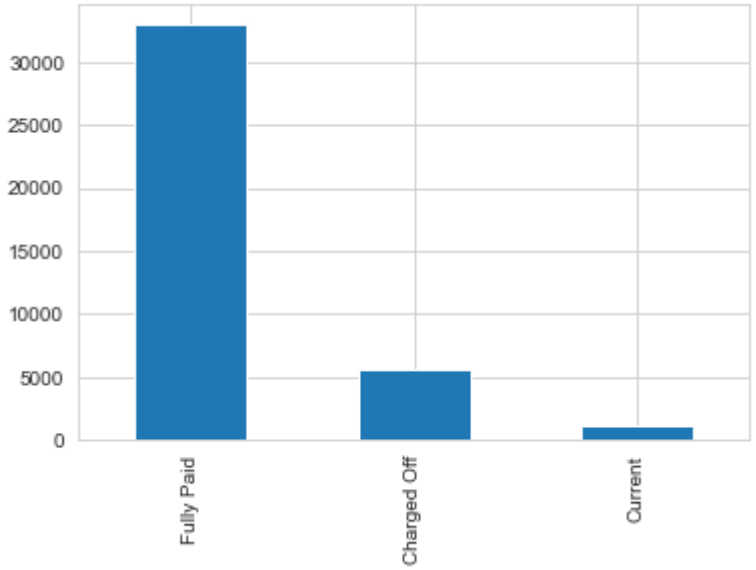


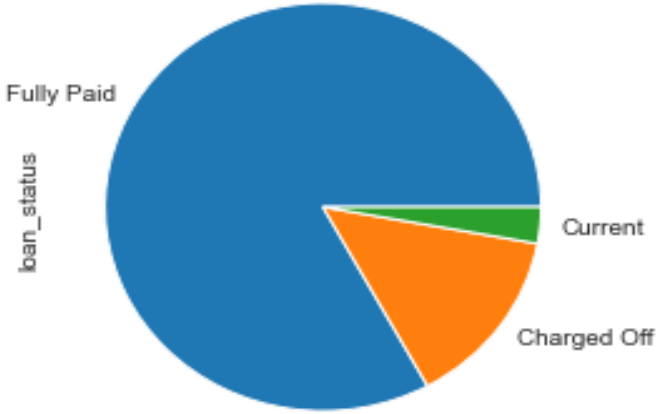
Approximately 14% of loans are defaulted.
Any variable that increases percentages of default to higher than 16.5 should be considered a business risk.

Lending Club only recovers 56.5% of the loan amount when loans are defaulted.

Only fully paid up loans, the company makes 17% profit.

Analysis – Understanding Loans

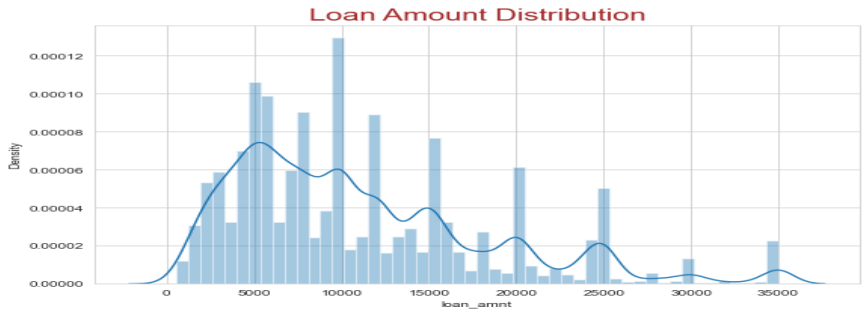


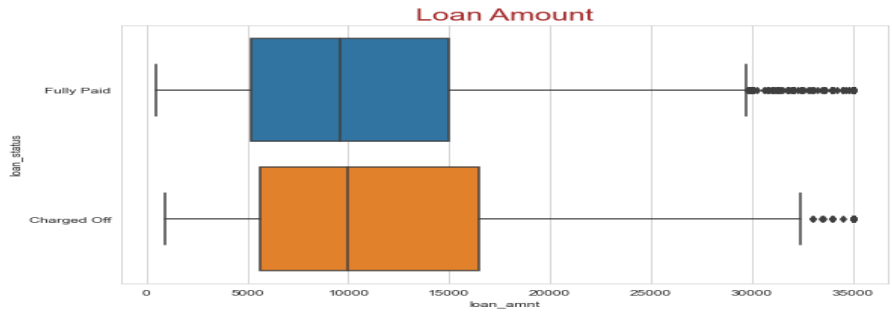


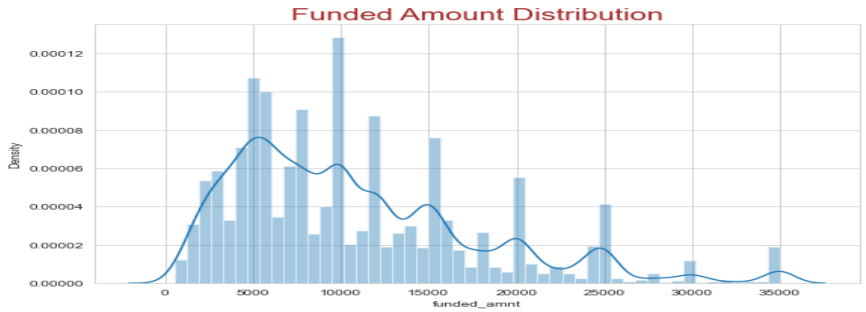
Observations:

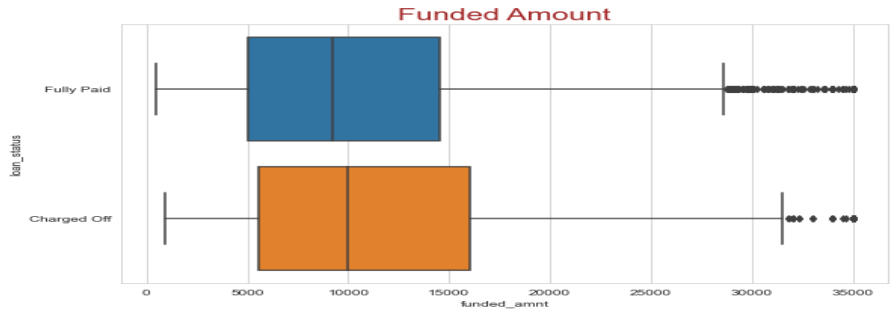
1.Most of the Loans are Fully Paid
2.A vary small portion of the data is for current loan, which is irrelevant for the current analysis so we can remove them
3.The ratio of Charged Off data and Fully Paid data is huge which can cause bias in the analysis.

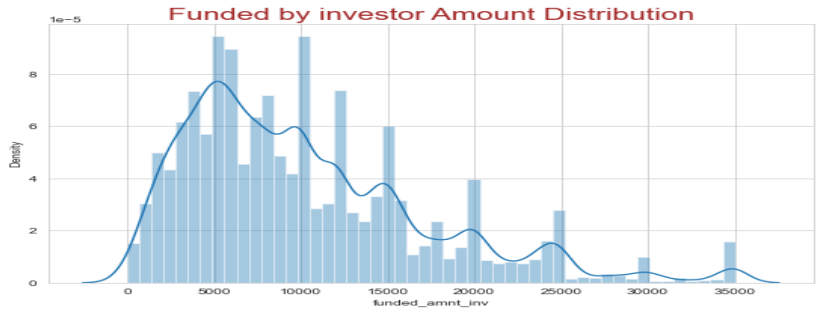
Analysis – Understanding Loans. Continue..

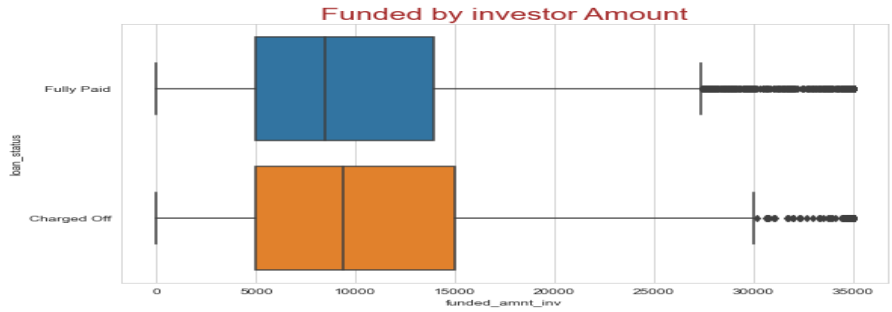












Understanding:

•Most of the loan amounts are in the range of 0-15K dollars

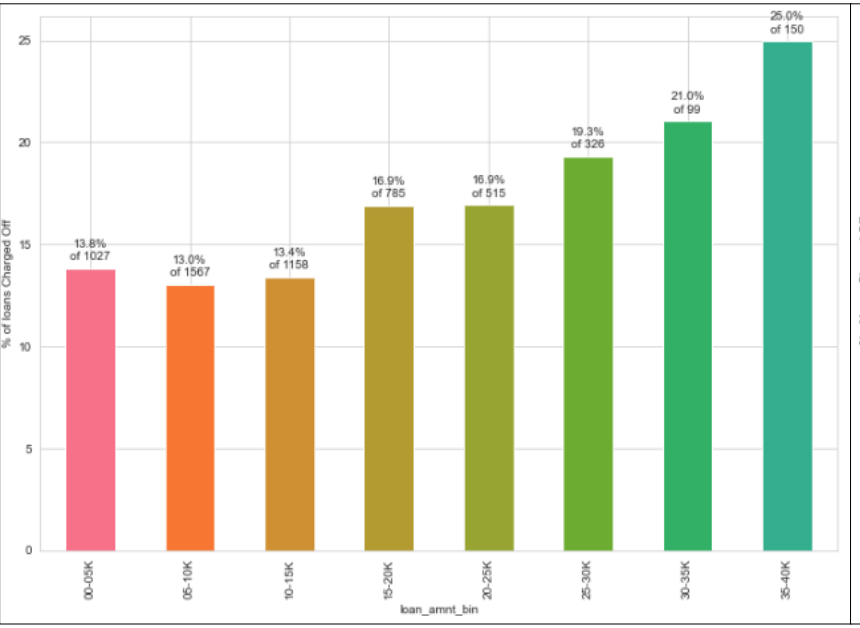
•data is little left skewed

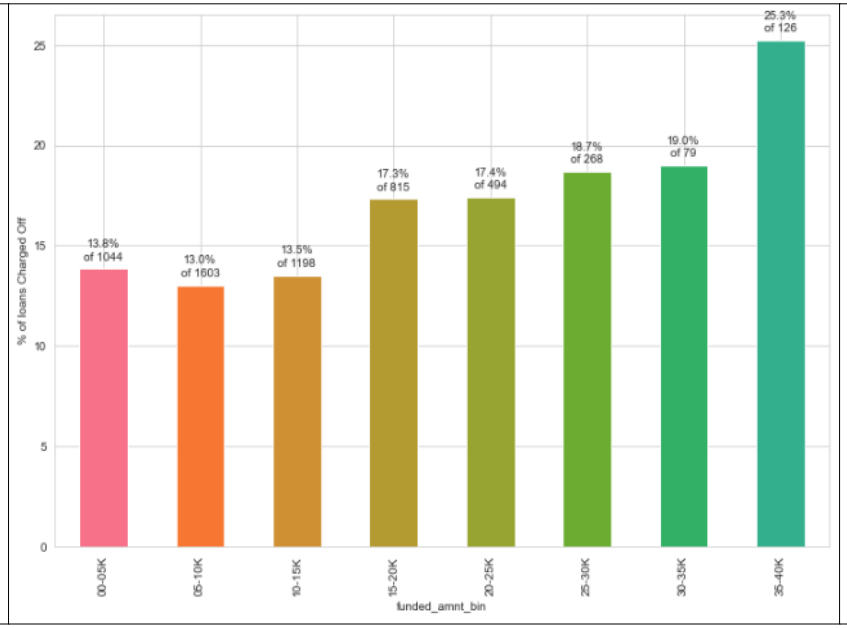
Observation:

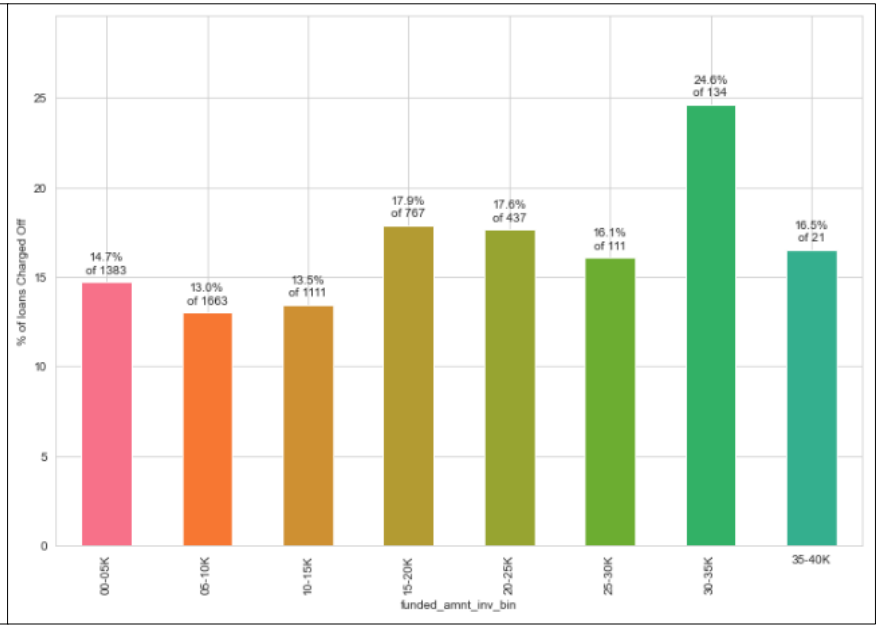
•Overall, the applied loan amount distribution is slightly right-skewed with mean greater than the median. Most of the loans granted are below 15000 (75 percentile value)

•Funding amounts see a spike around each 5000 boundary. We will use 5000 as bucket size for later analysis.

Analysis – Understanding Loans. Continue..



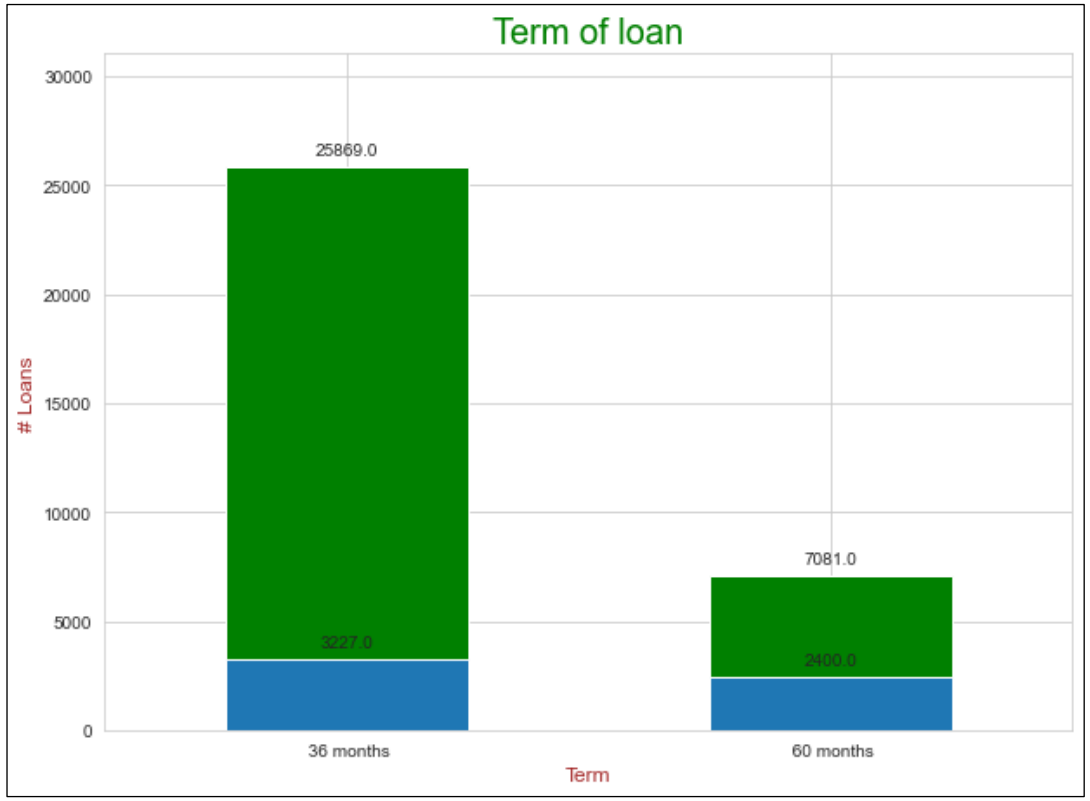


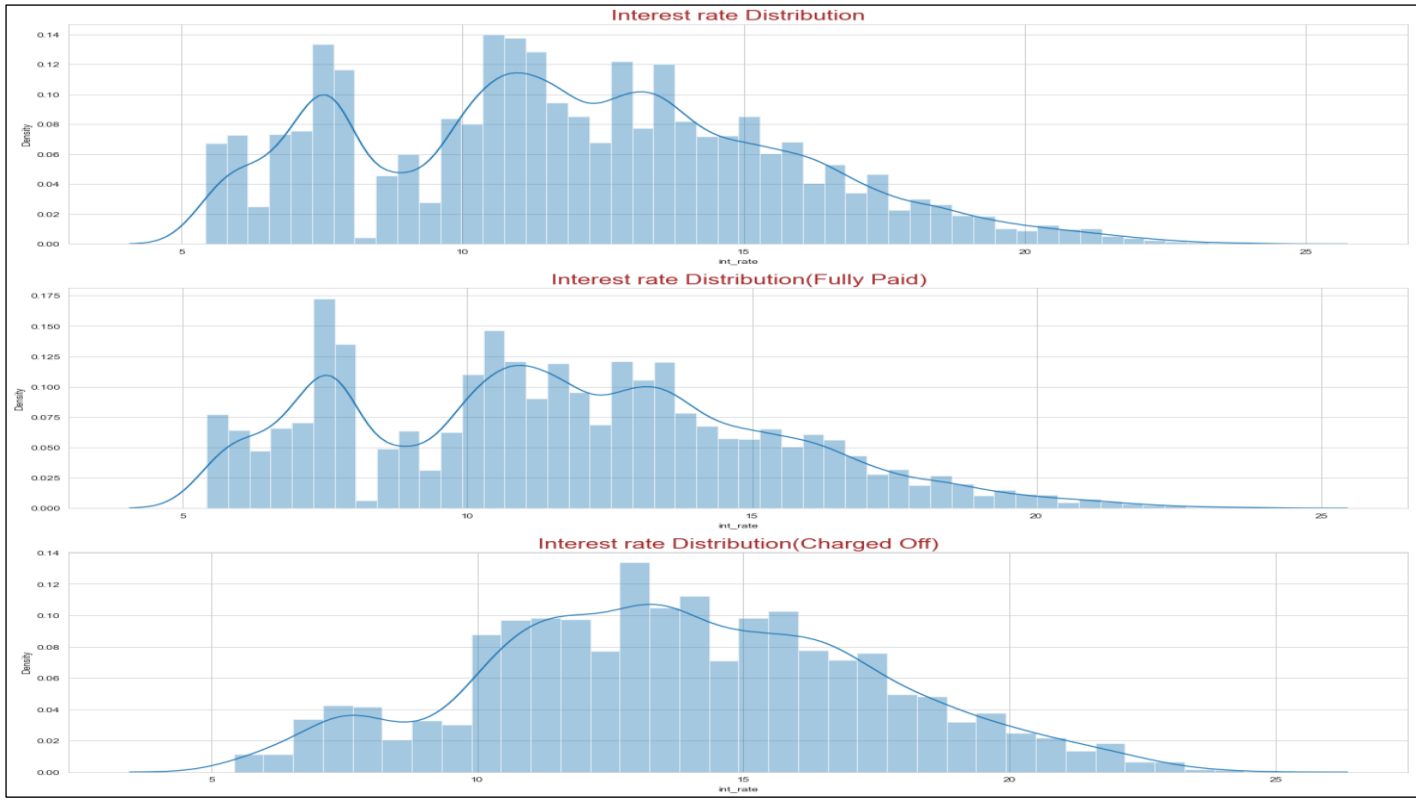


Observation:

The % of charged off loans increases substantially as we go up the loan amount buckets. Most loans are below 20000 amount. The higher loans, though lesser in number, carry a substantially higher risk of default.

Analysis – Understanding Loans. Continue..





Observation

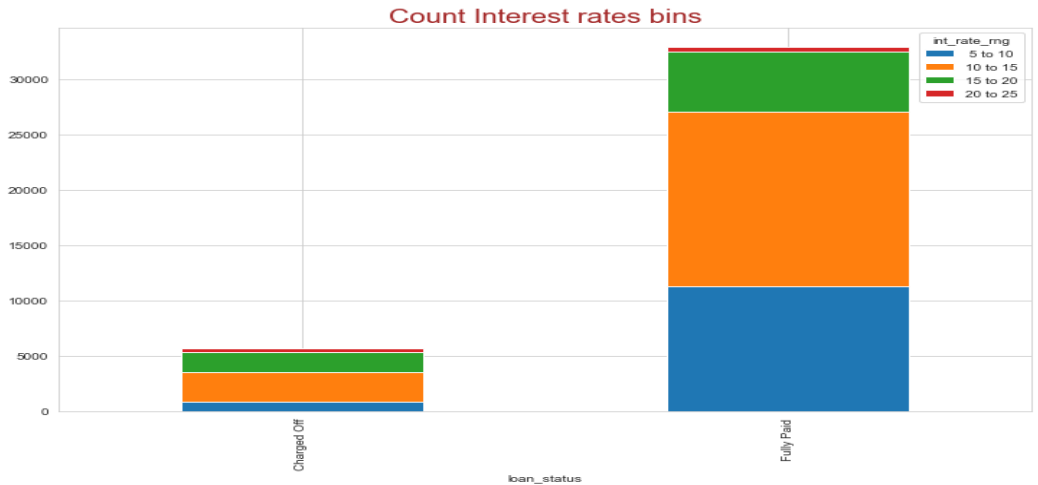
1.For 36 months loan the Charged off percent is less as compared to 60 months loan.

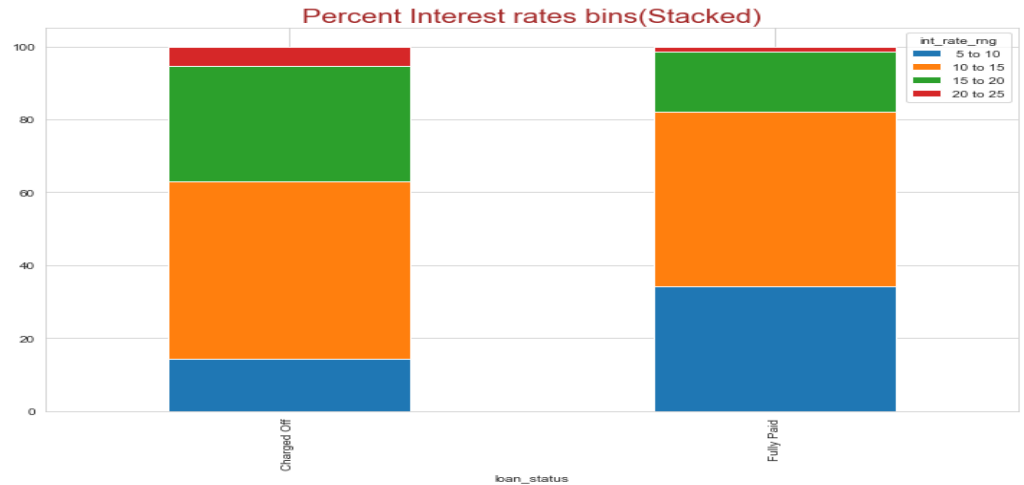
2.Longer tenure loans are associated with higher risk.

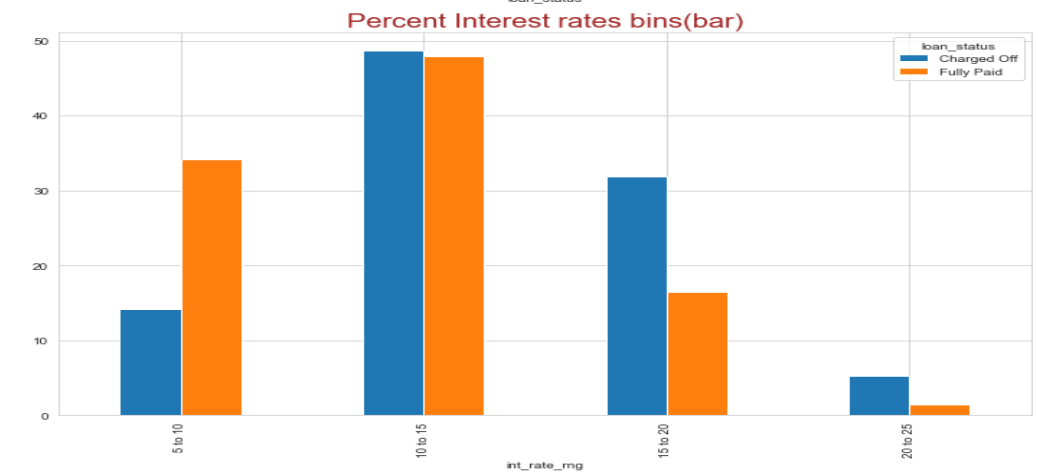
Observation

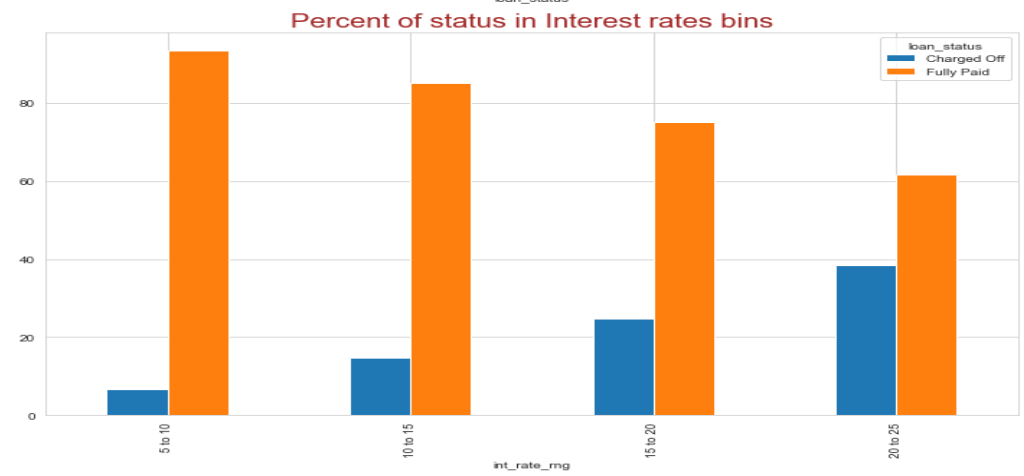
1.The Interest rate distribution across the data and Fully paid data is similar. This can be due to the fact that the ratio of fully paid data is more.
2.The Distribution of Fully paid interest rate is skewed towards left, Indicating the chances of default is low when the interest rate is low.
3.The Distribution of interest rate for charged off loans are relatively normally distributed.

Analysis – Understanding Loans. Continue..





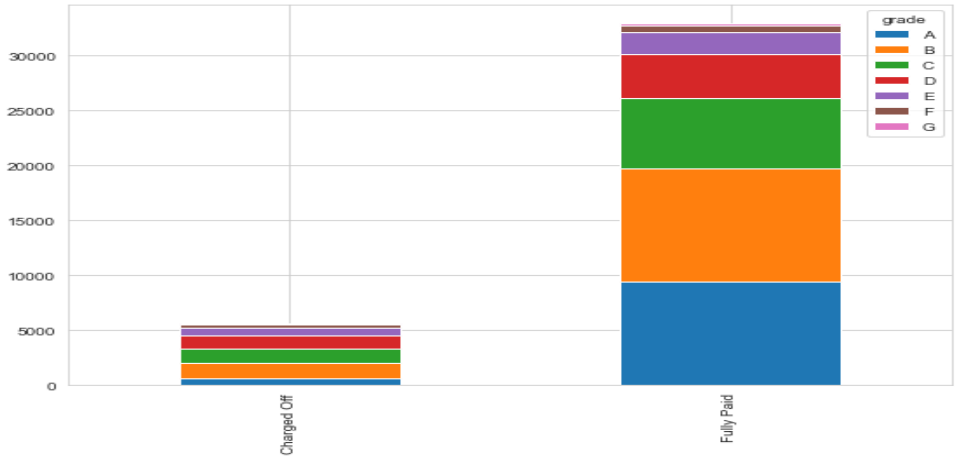


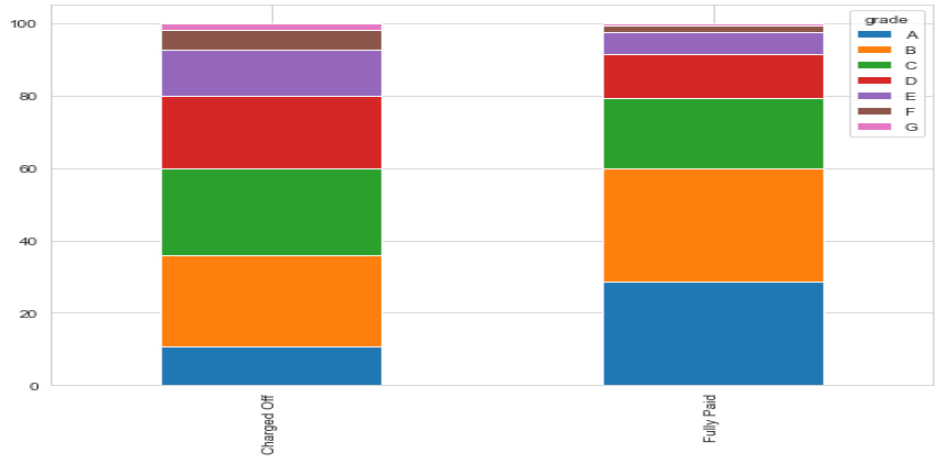


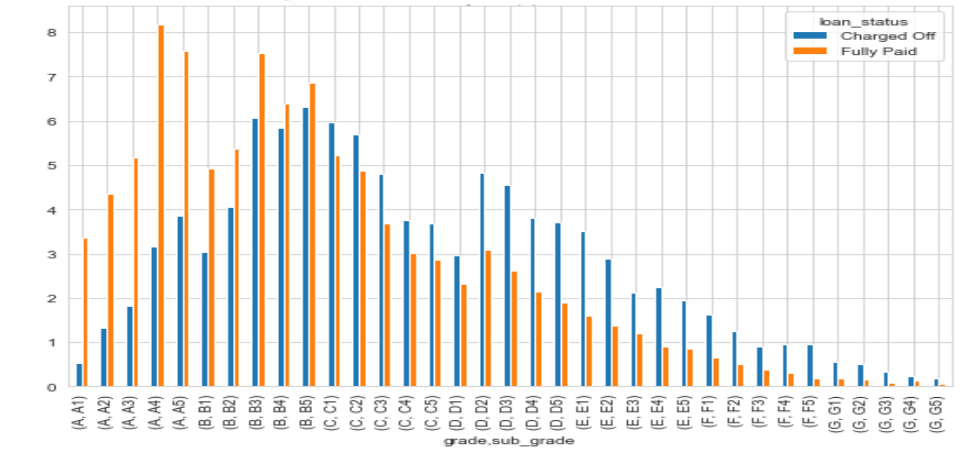
Observation:

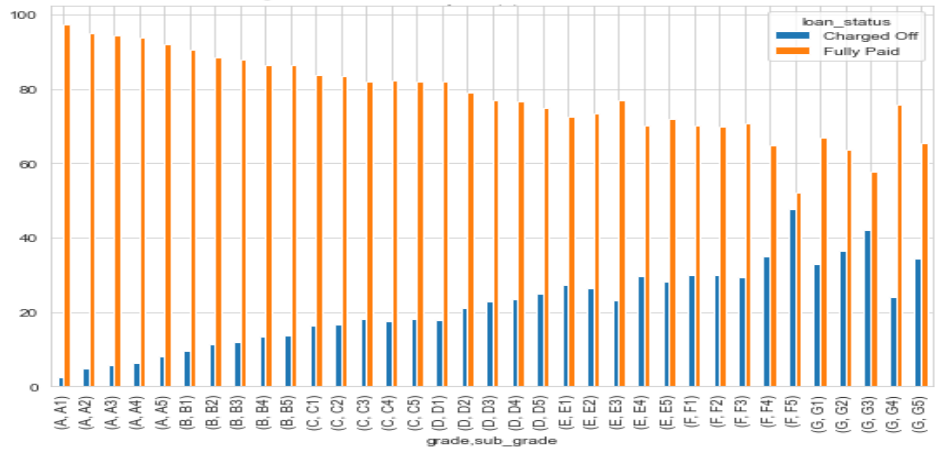
From the above we can see that higher interest rates contribute more to Charge off, increasing the risk..

Analysis – Understanding Loans. Continue..





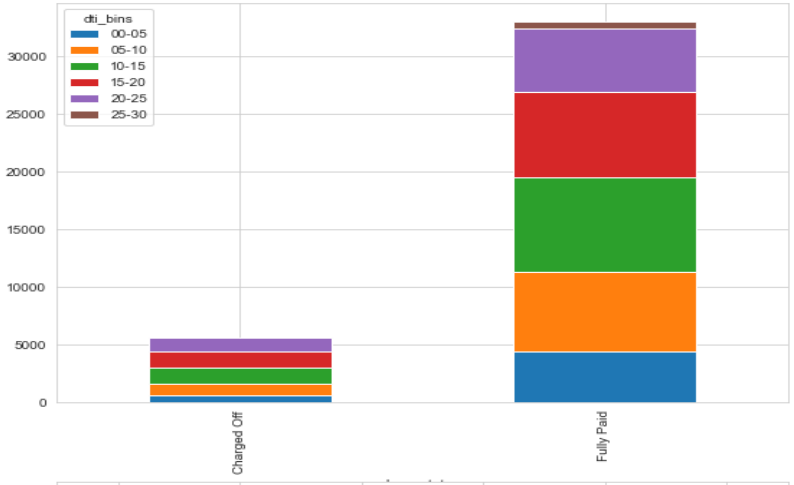


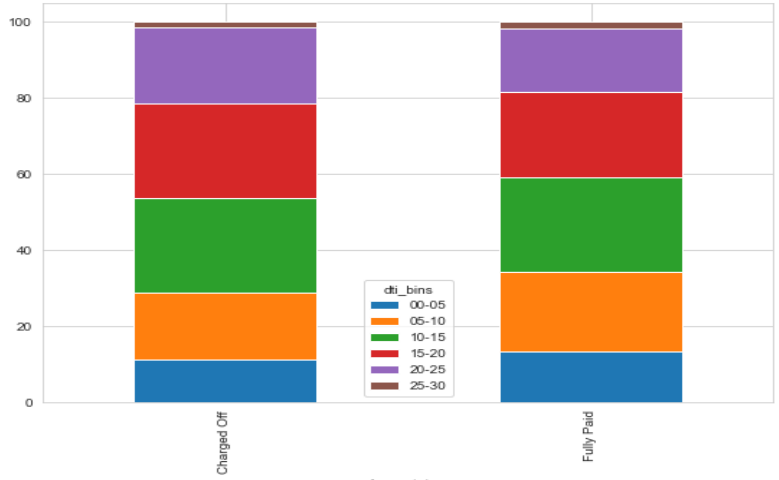


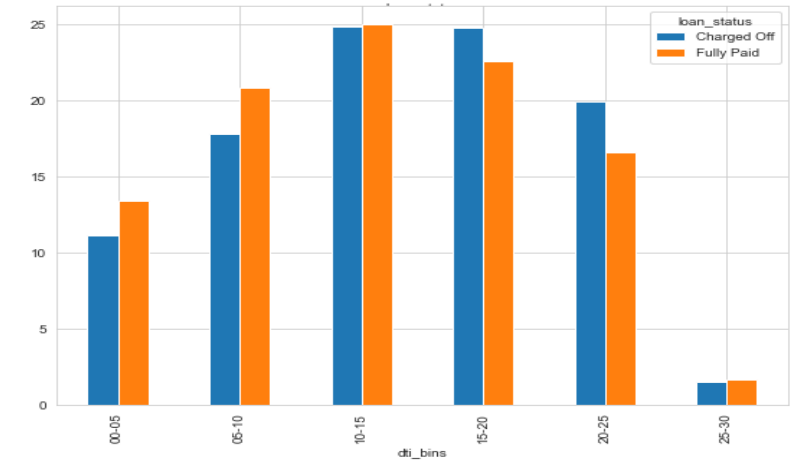
Observation:

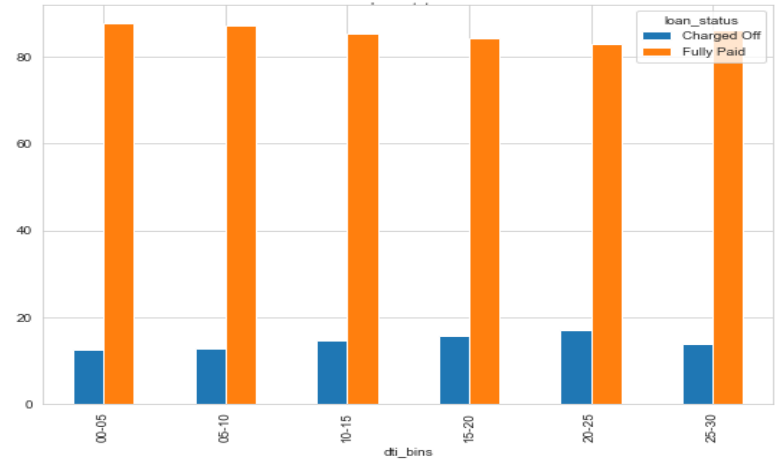
- The Higher-Grade loans are approved more often as compared to lower grade loans (A being higher and G lower)
- The loans with lower grade has a higher probability of defaults.

Analysis – Understanding Loans. Continue..





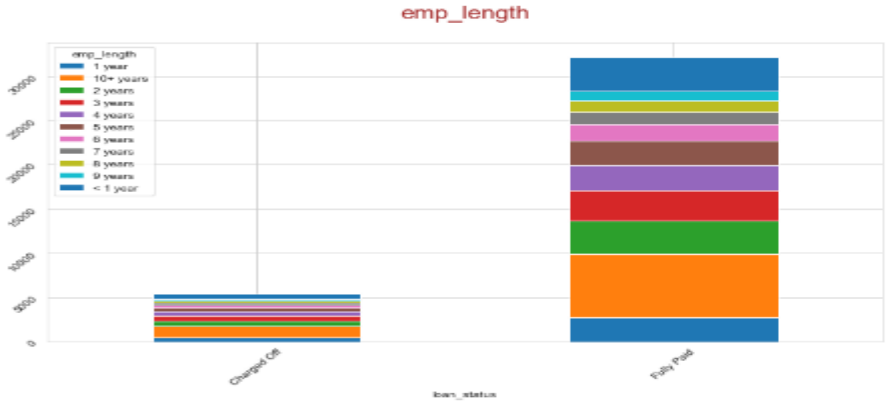


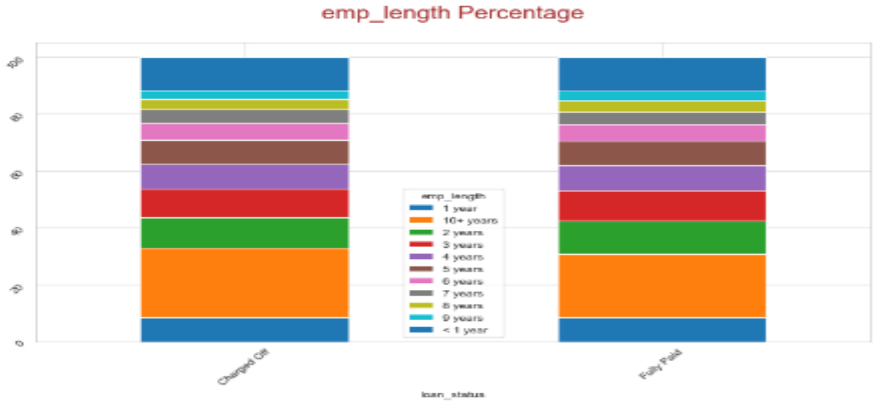


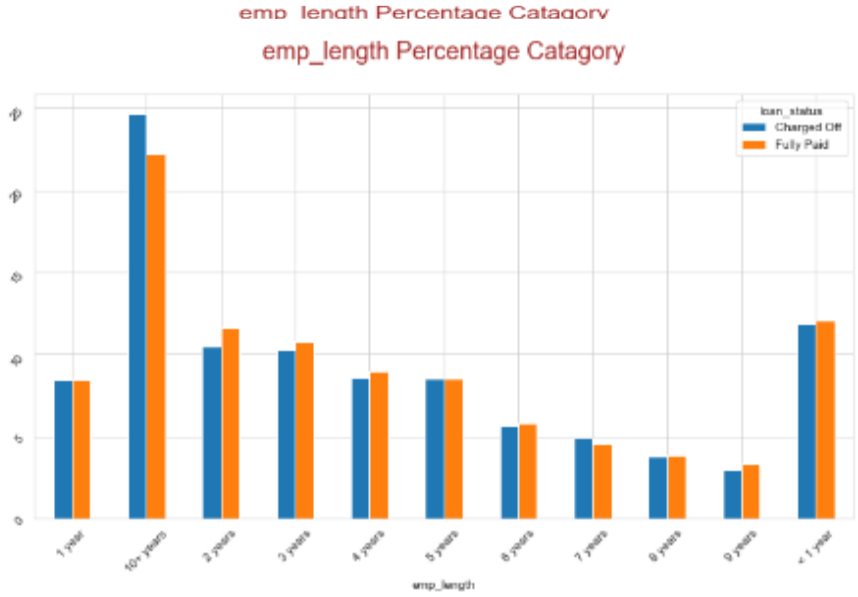
Observation:

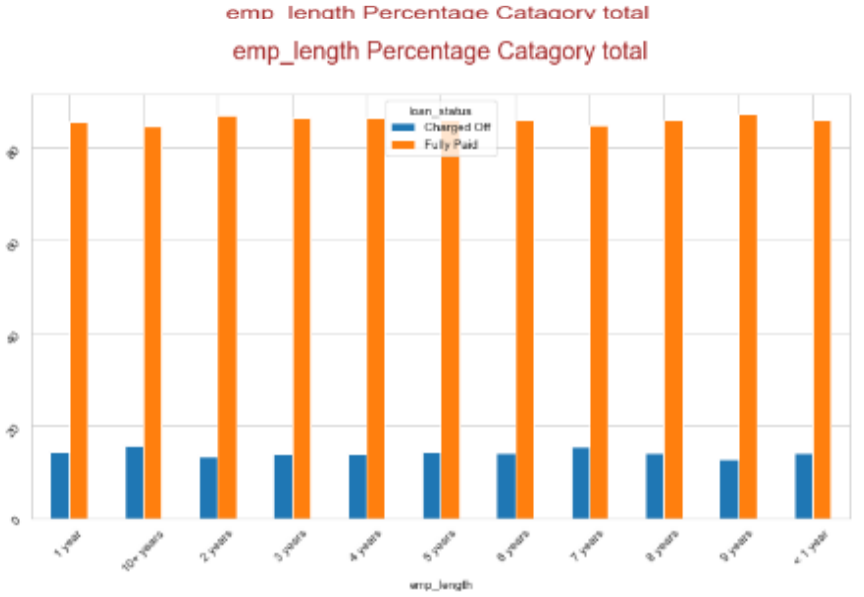
- DTI for defaulters is skewed towards right, i.e. for higher DTI the probability of Charge Off is high.
- The percentage of defaulters in each bin of DTI is in the range of 10-18, and from the distribution of data we can see that Lending Club i

Analysis – Understanding Loans. Continue..



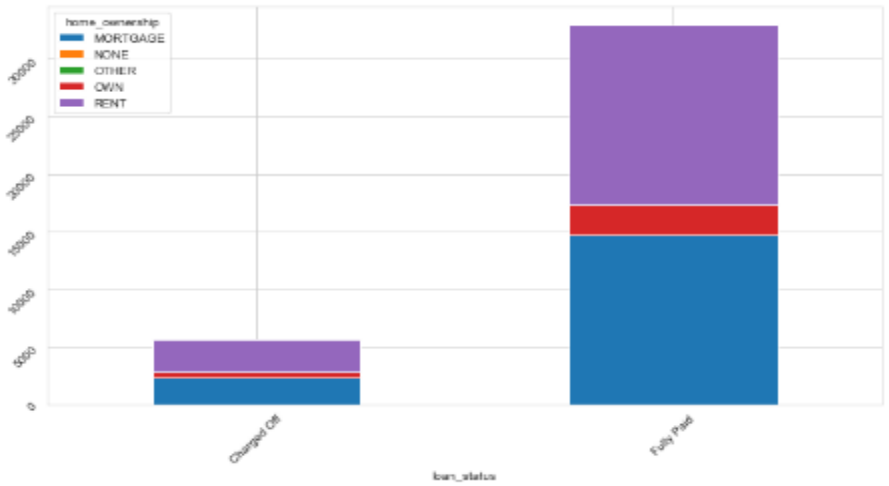




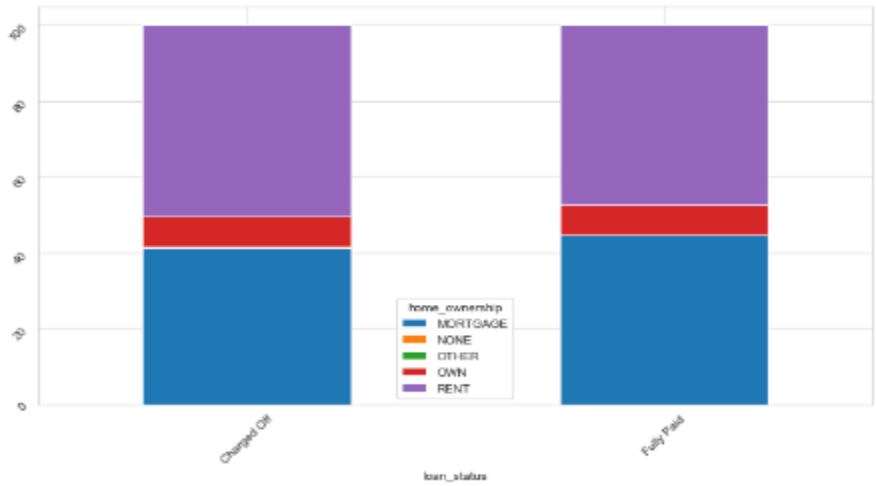


Analysis – Understanding Loans. Continue..

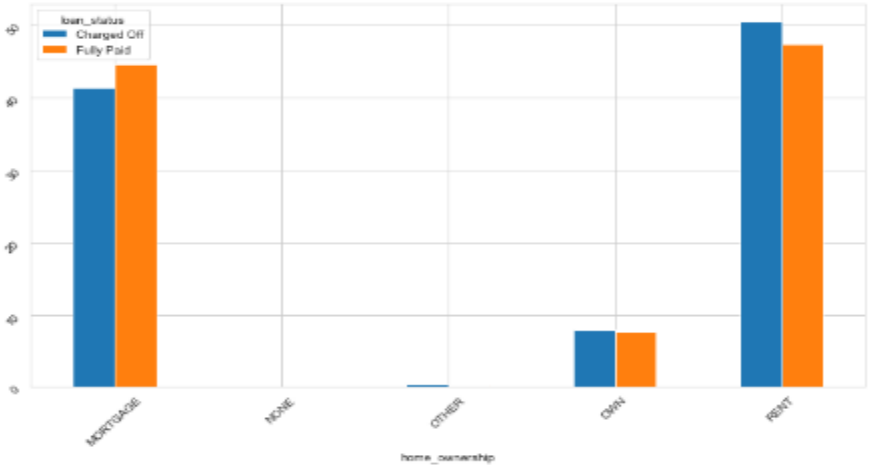
home_ownership



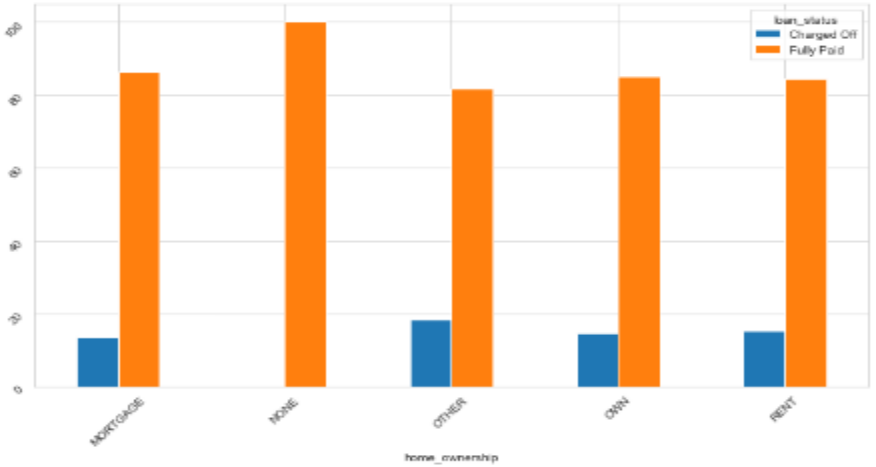
home_ownership Percentage



home_ownership Percentage Catagory

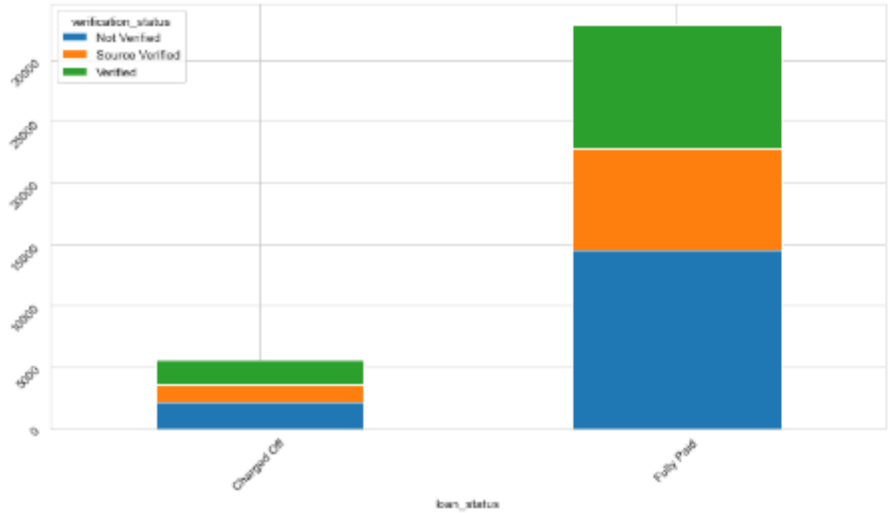


home_ownership Percentage Catagory total

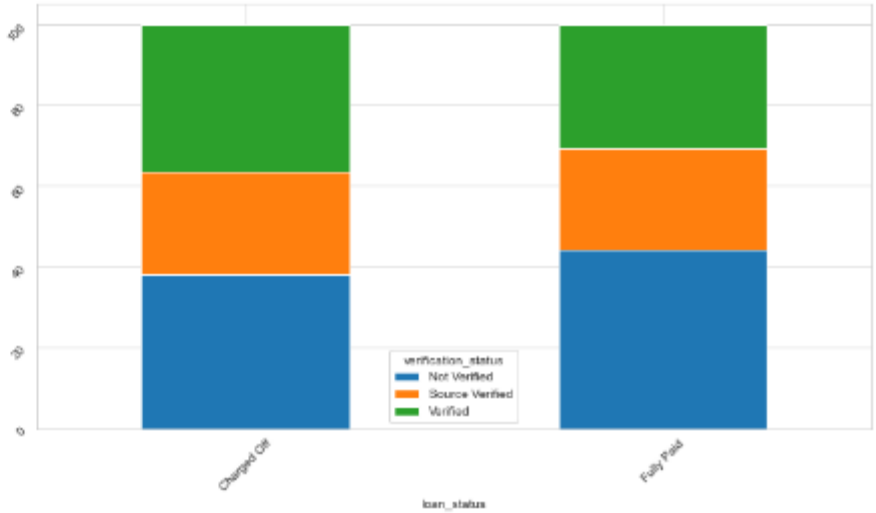


Analysis – Understanding Loans. Continue..

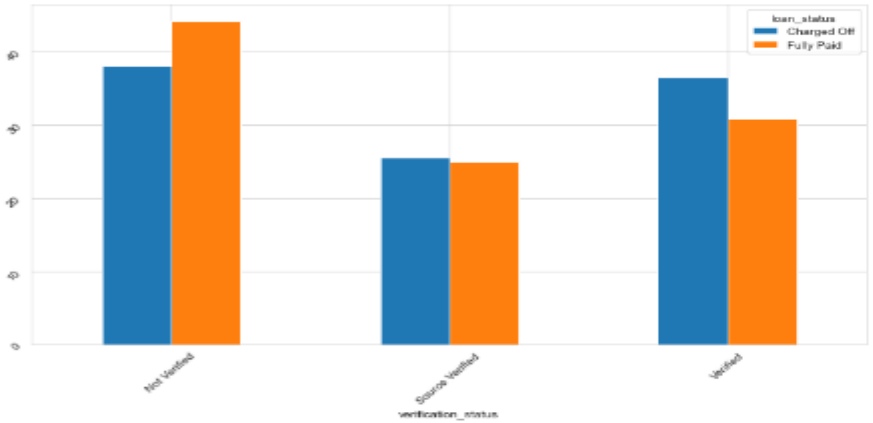
verification_status



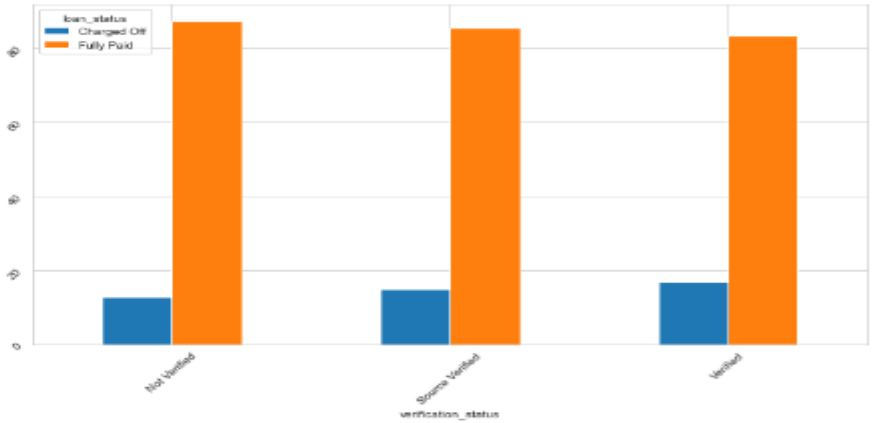
verification_status Percentage



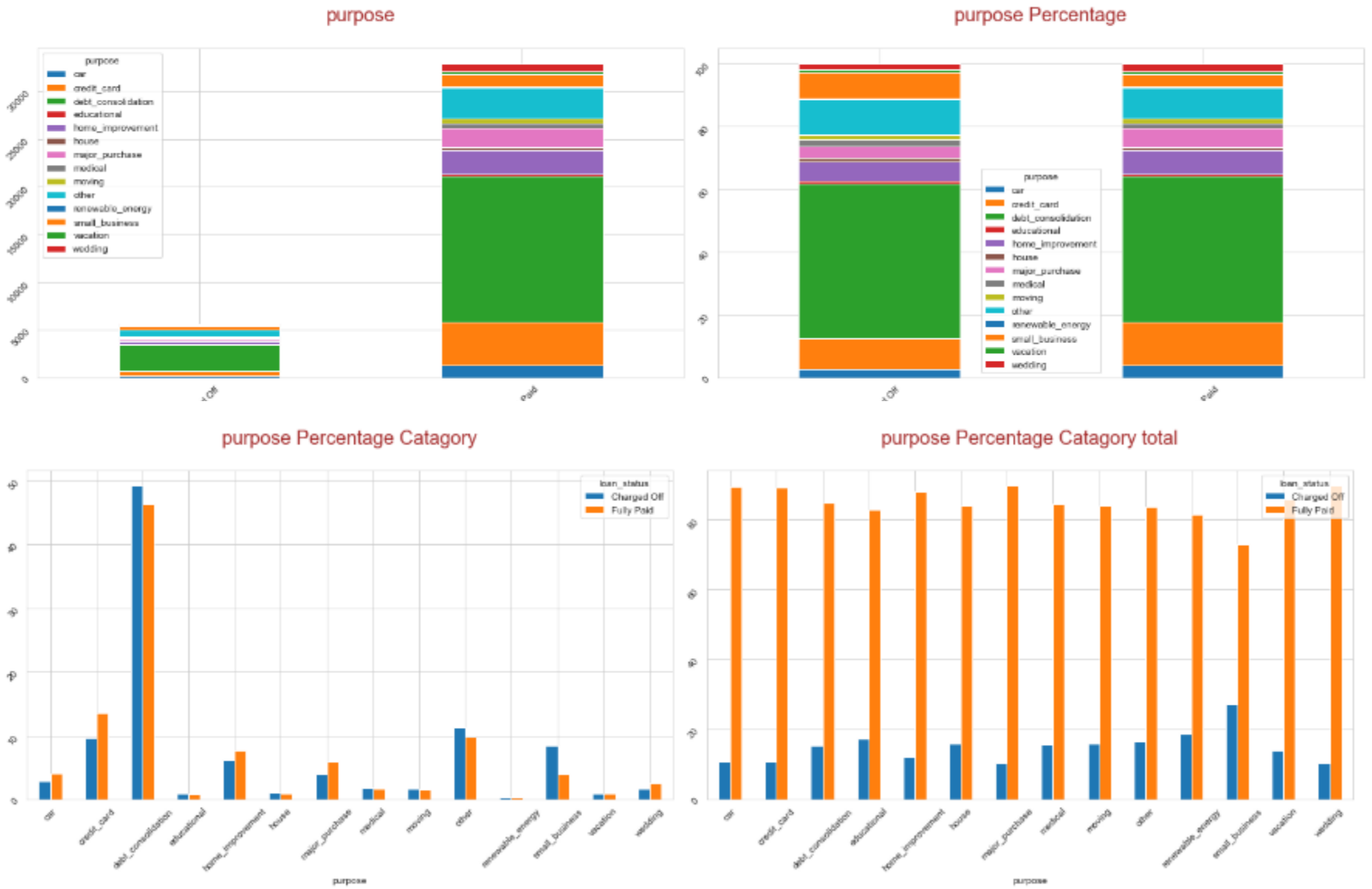
verification_status Percentage Catagory



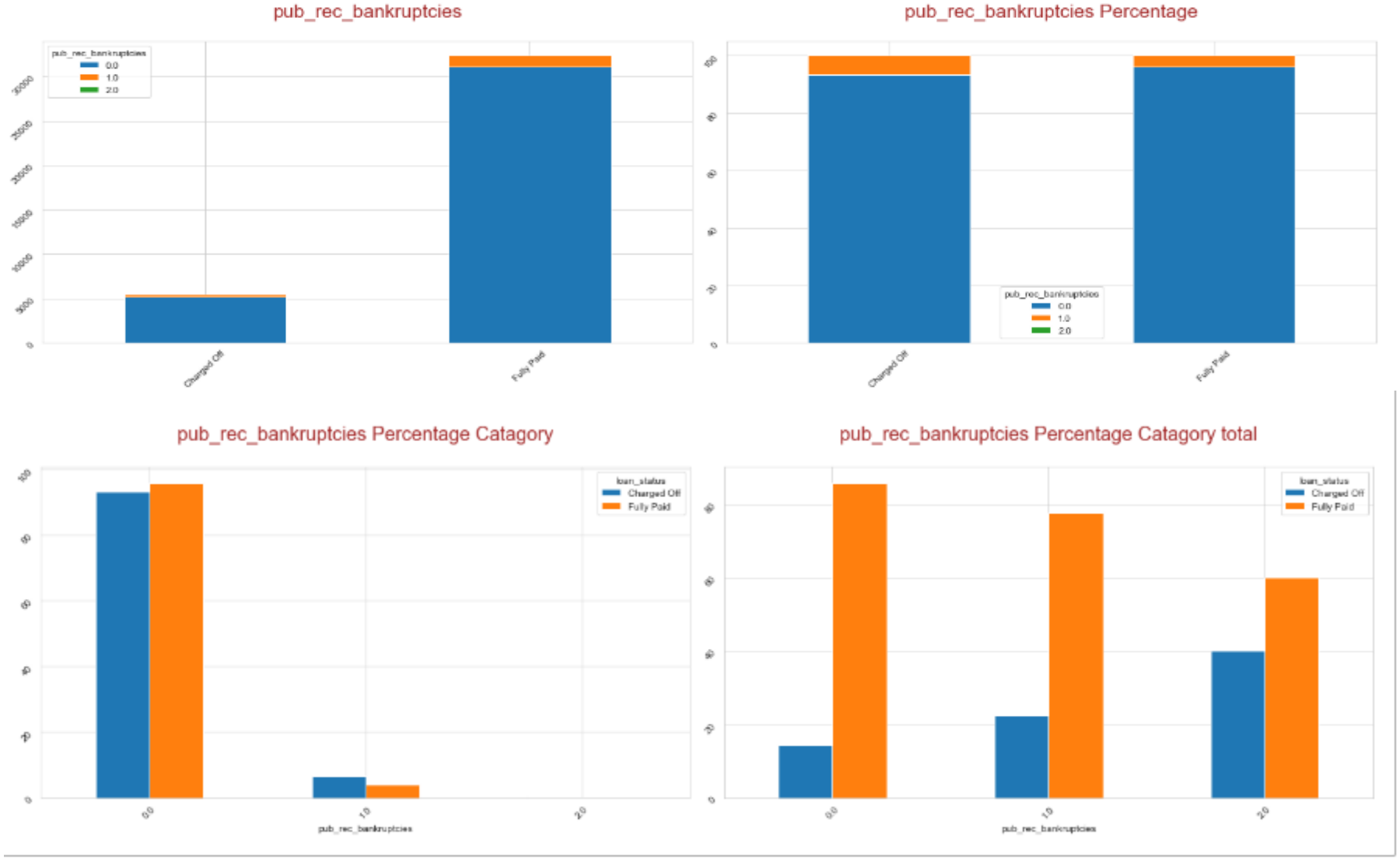
verification_status Percentage Catagory total



Analysis – Understanding Loans. Continue..



Analysis – Understanding Loans. Continue..



Analysis – Understanding Loans. Continue..

Observation:

1.DTI for defaulters is skewed towards right, i.e. for higher DTI the probability of Charge Off is high. emp_length
    1. There is no significant difference in fully paid and charged off for different employment length

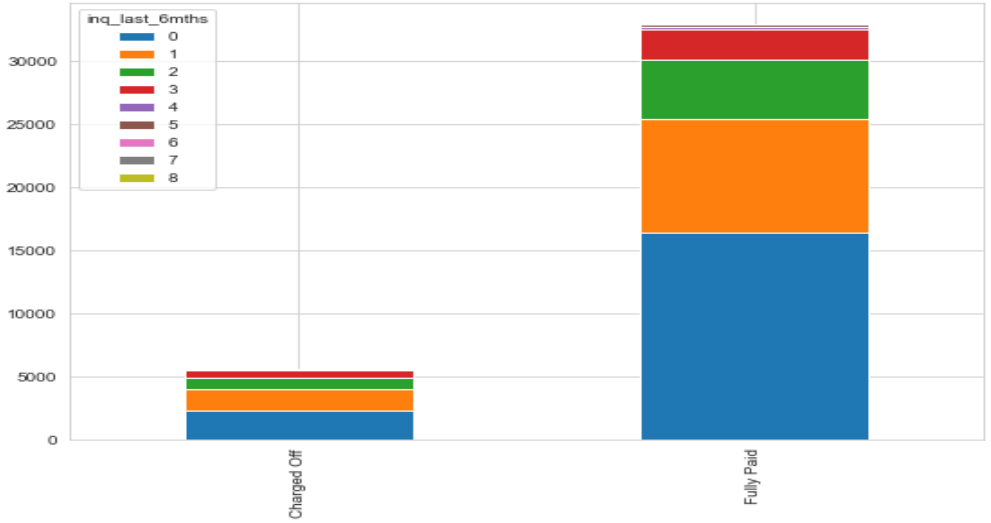
2.homeownership
    1. There is no significant trend difference between fully paid and defaulters in relation to home ownership

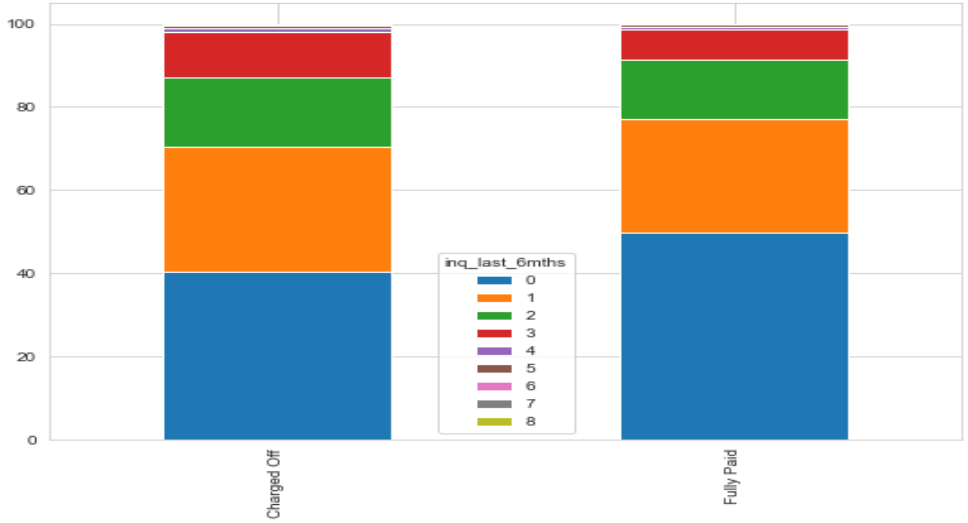
3.verification status
    1. There is a counter intuitive trend visible here, the verified customers have a higher chance of default as compared to unverified borrower

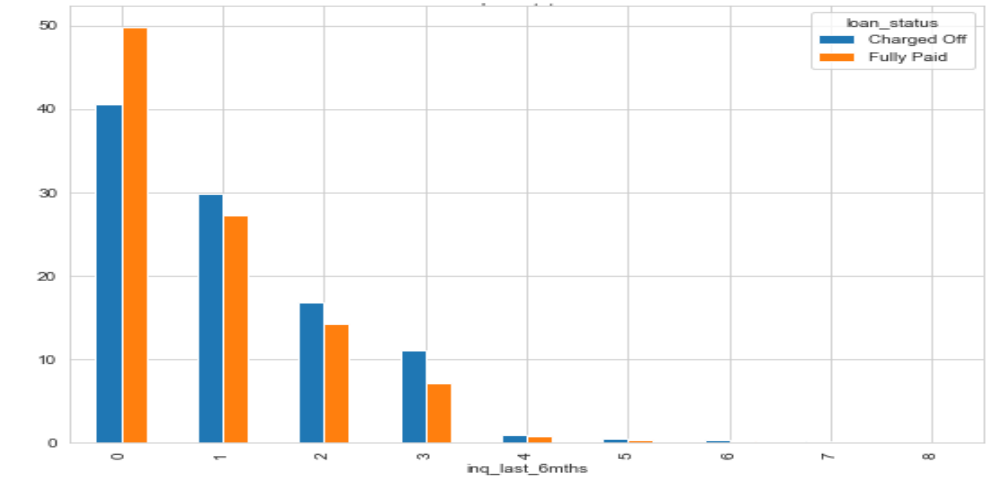
4.purpose
    1. Dept consolidation is the main reason for the loans
    2. Small business have a higher chance of getting charged off

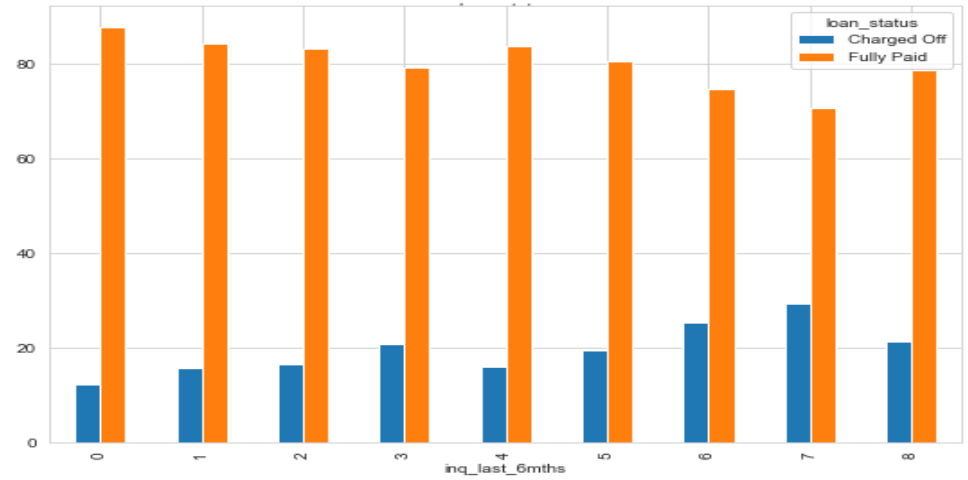
5.pub_rec_bankruptcies
    1. The number of loans provided to borrower with public bankruptcy record is very less, still it follows the expected trend of high defaults with high number of bankruptcies

Analysis – Understanding Loans. Continue..





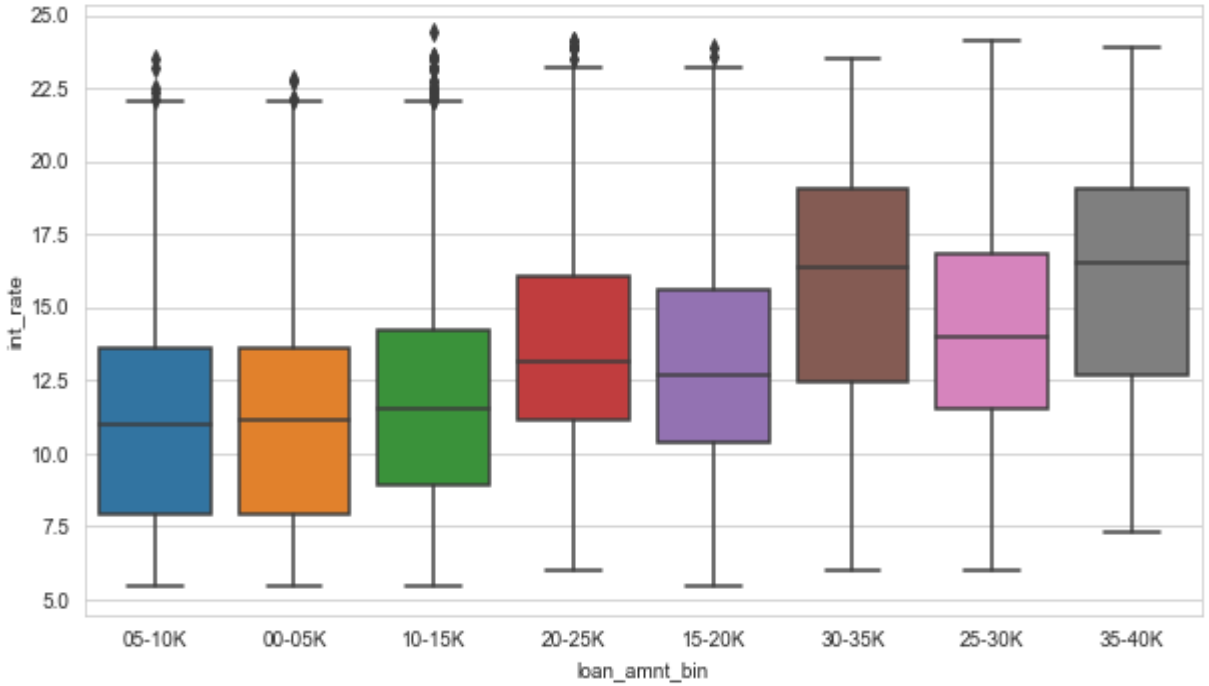


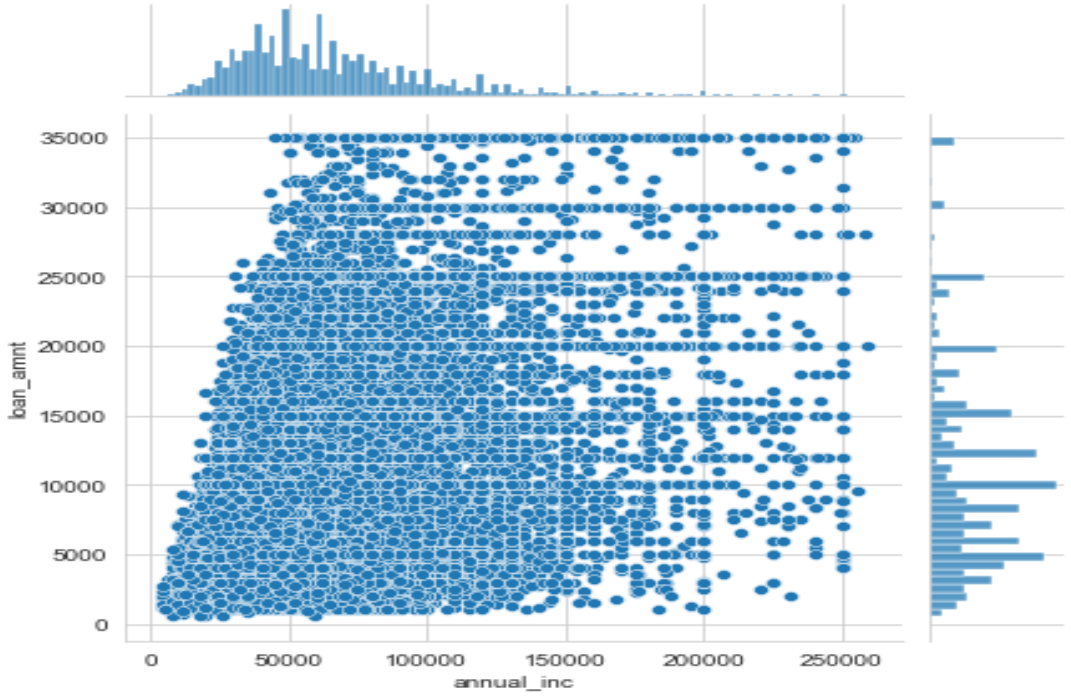


Observation:

The number of inquiry has effect on the charge off, loans having more number of inquiry have a higher chance of default.

Analysis – Understanding Loans. Continue..

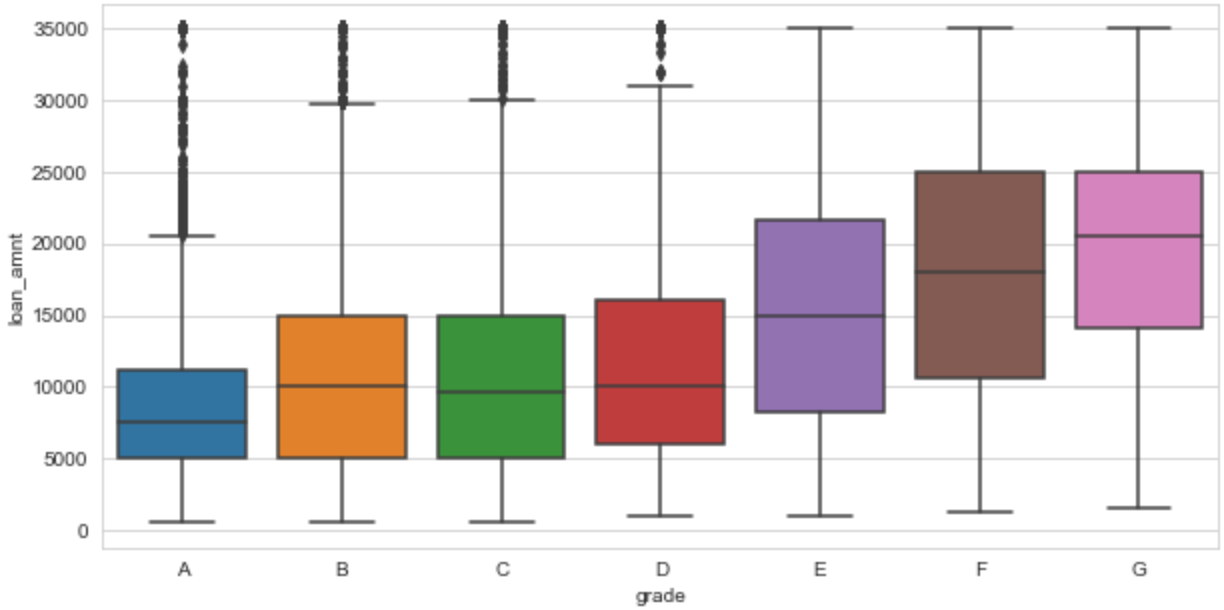


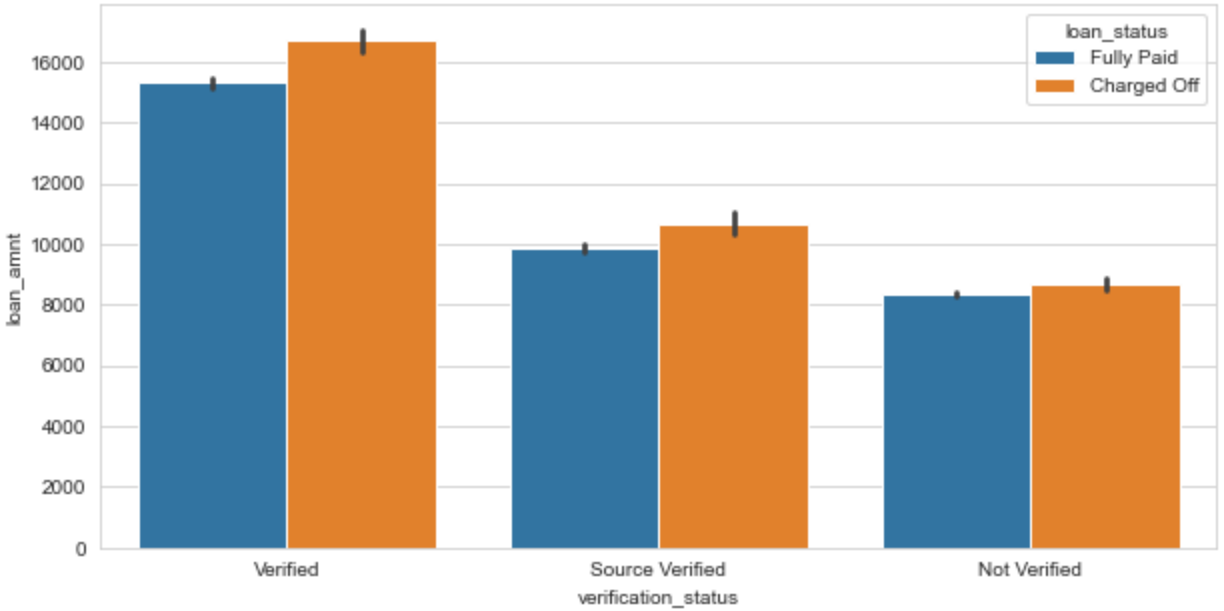


Observation:

There are people with average income lower than 50000 taking loans of 25000 or higher. These would be risky loans.

Analysis – Understanding Loans. Continue..





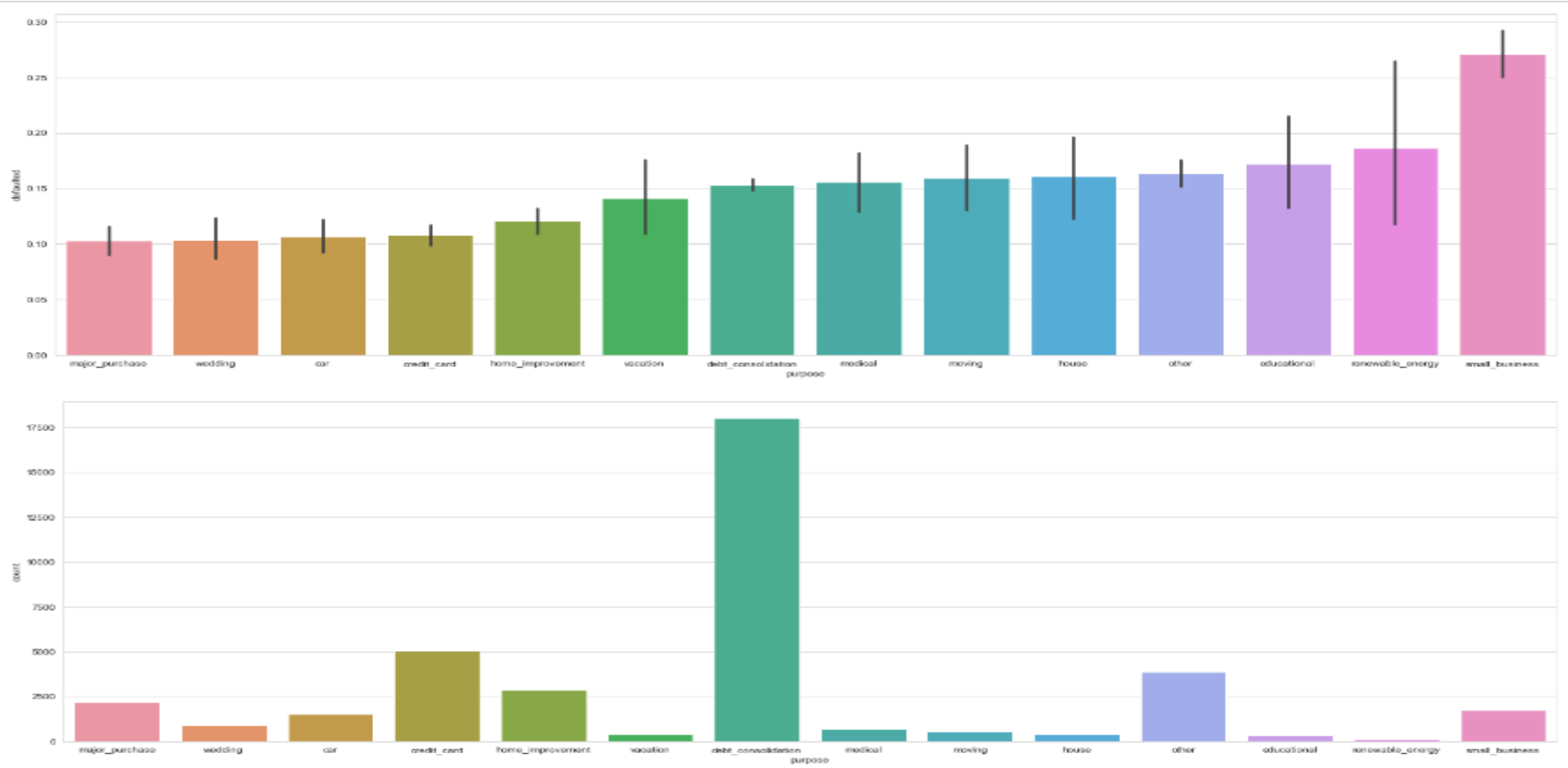
Observation

•Larger loans generally appear to be given a lower grade, with the median loan amount for a grade G loan being almost 10000 higher than that of a grade A, B, or C loan.

Observation

•Higher loan amounts are Verified more often.
•We already know that larger loans are less in number but see a higher charge off rate.
•This, combined with previous observation, explains why verified loans see a higher rate of default. It's not the verified status per se, it's the fact that higher loan amounts are riskier and are also verified more often by Lending Club.

Analysis – Understanding Loans. Continue..



Observation:

•If you see the trend of the defaulters there are more people from educational, renewable energy & small business people present.

•And with the reason as Other there is a huge no of people and there is a high chance of rejecting the loan if it's from the other purpose.

Conclusion:

Below are the top variables that are impacting the defaulters:

- **Grade, SubGrade** -- As grade changes there are defaulters changing, G being the highest
- **Purpose of the loan** -- Small business has more defaulters
- **Interest rate** -- As interest rate increases defaulters are moe
- **Term** -- There are more defaulters with 60 months term than 36 months
- **Annual income** -- If the annual income is more then high possibility that he pays the loan
- **Debt to Income** -- Percentage of default rises with dti ratio. As the dti ratio rises above 20, the loans become risky

Recommendations:

1. **Stop** – approving loans where amount/income is higher than 30%.
2. **Reduce** – number of approvals where purpose is small business.
3. **Stop** – approving high-value loans when revolving line utilization rate greater than 75%.
4. **Stop** – approving loans to people with prior bad record. Or at least stop approving high-value loans.
5. **Start** – charging higher interest rates for loans with dti greater than 20.

Thank You