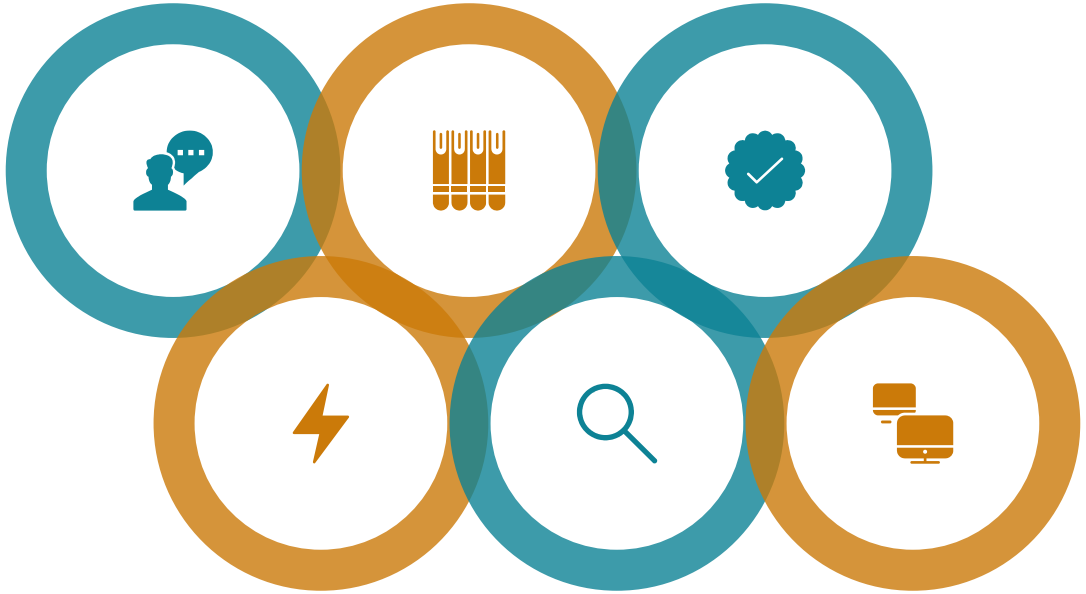
Lending Club Case Study

Background &Problem Statement

This company 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.

Lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss).

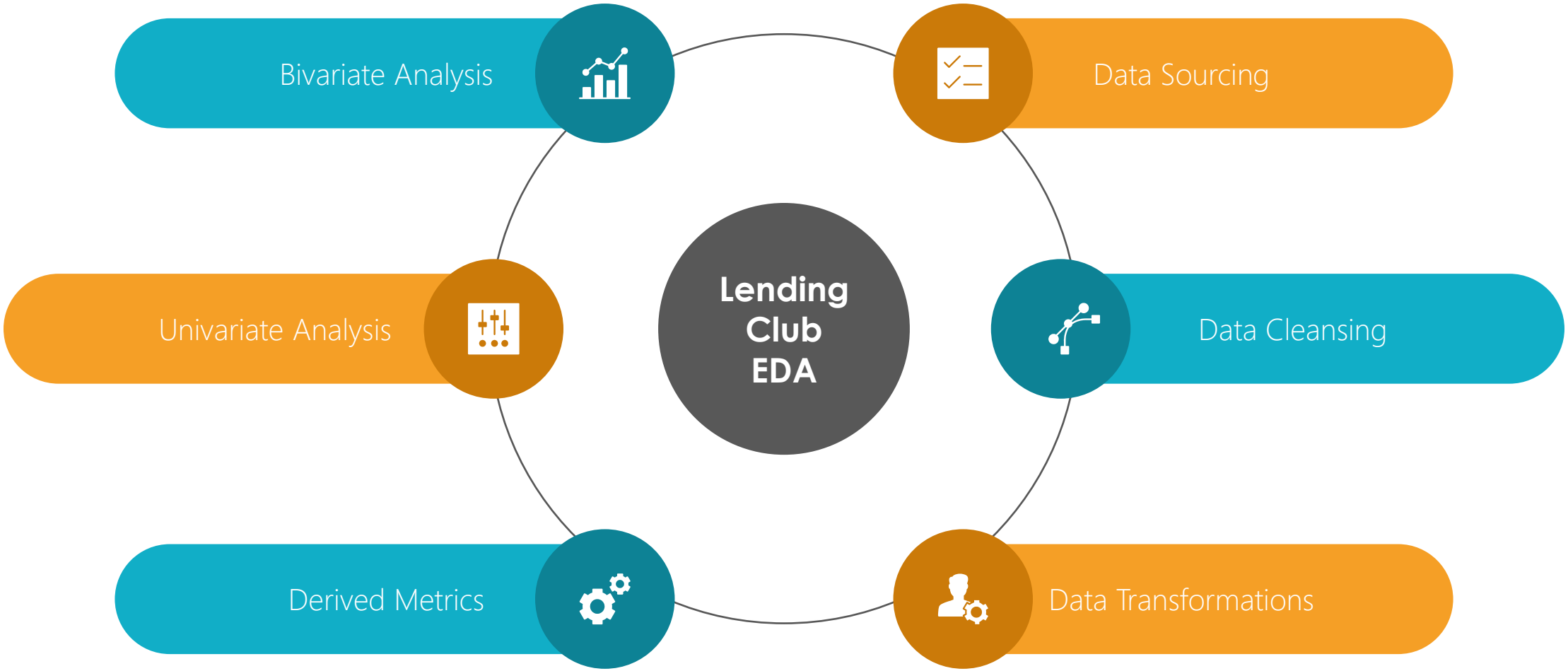


Identification of risky loan applicants in advance can reduce credit loss.

Identification of risky applicants using EDA is the aim of this case study.

Company wants to understand the driving factors (or driver variables) behind loan default.

Solution Approach



Data Cleansing

All Null Values

Removed columns with all null values. There are around 54 columns with all values as NA

Unique value count

If the count is very less or huge it may not be very useful for analysis. So these columns have been removed

Remove out of scope columns

All IDs can be removed as it is unique and will not give any useful insights. Also there are certain features which captures information after awarding the loan.

Final Column Count

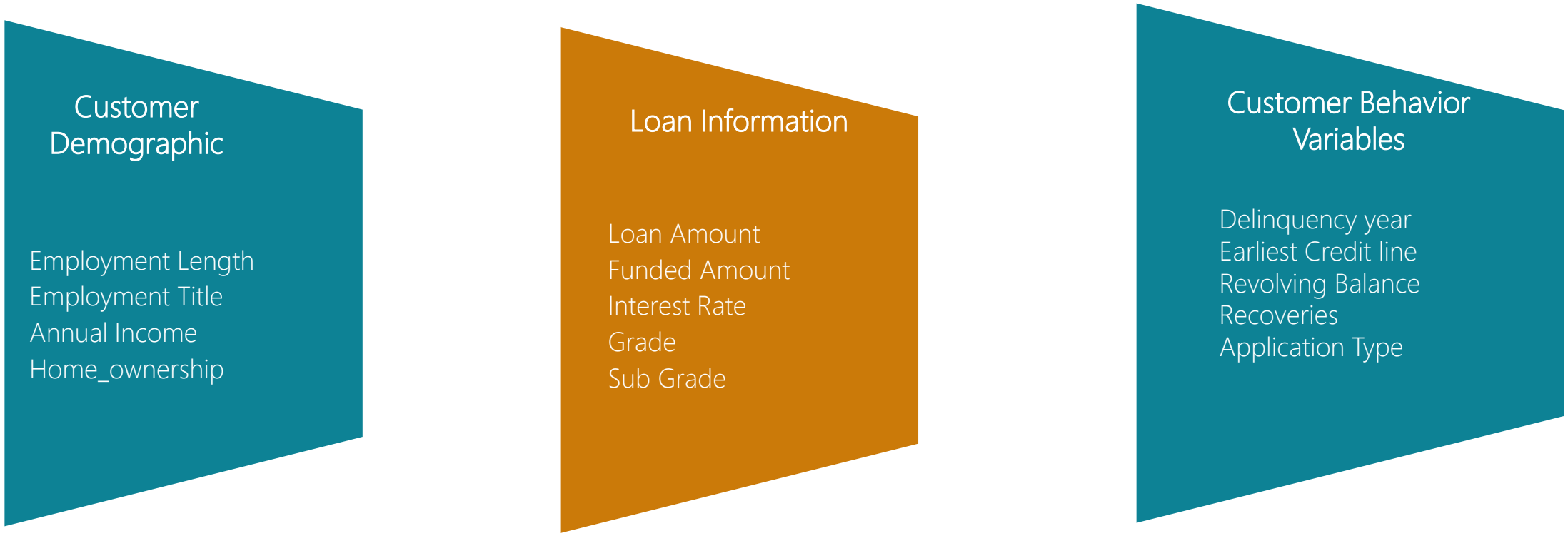
Our final data set is having 20 columns and it consists of customer attributes while applying loan and also loan attributes.

Dependent Variable filter

Business objective is to find out the likely defaulters or people who pay back full amount.
So we need to focus on observations having loan_status as "Fully paid" or "Charged off".
From a business objective standpoint loan_status value "Current" is not having any significance

Variable Types

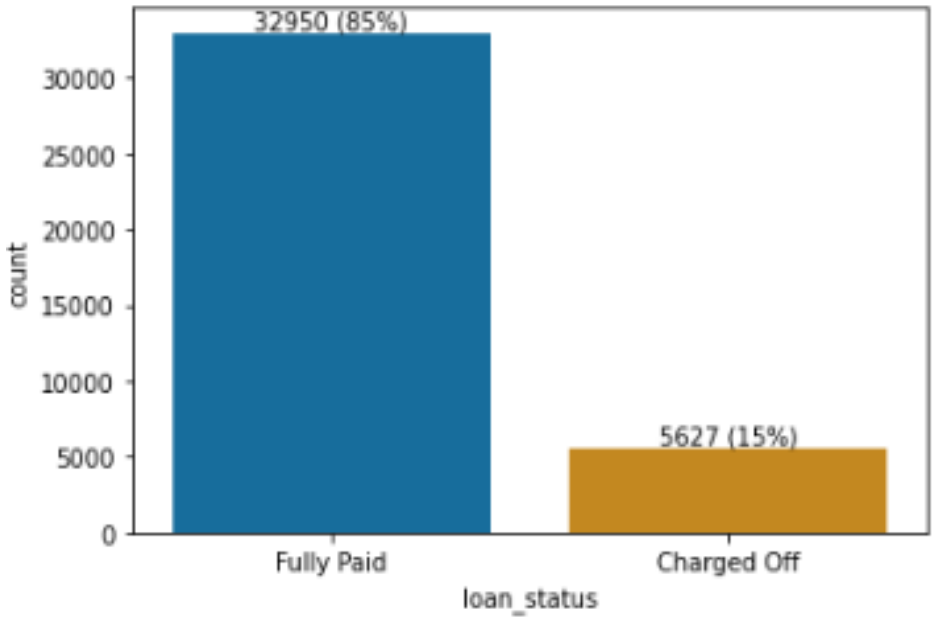
All variables (Independent Variables) can be classified into 3 categories. Some example fields against each categories are captured below



Customer behavior variables are not available at the time of loan application, and thus they cannot be used as predictors for credit approval. Hence we have removed Customer behavior variables from our data.

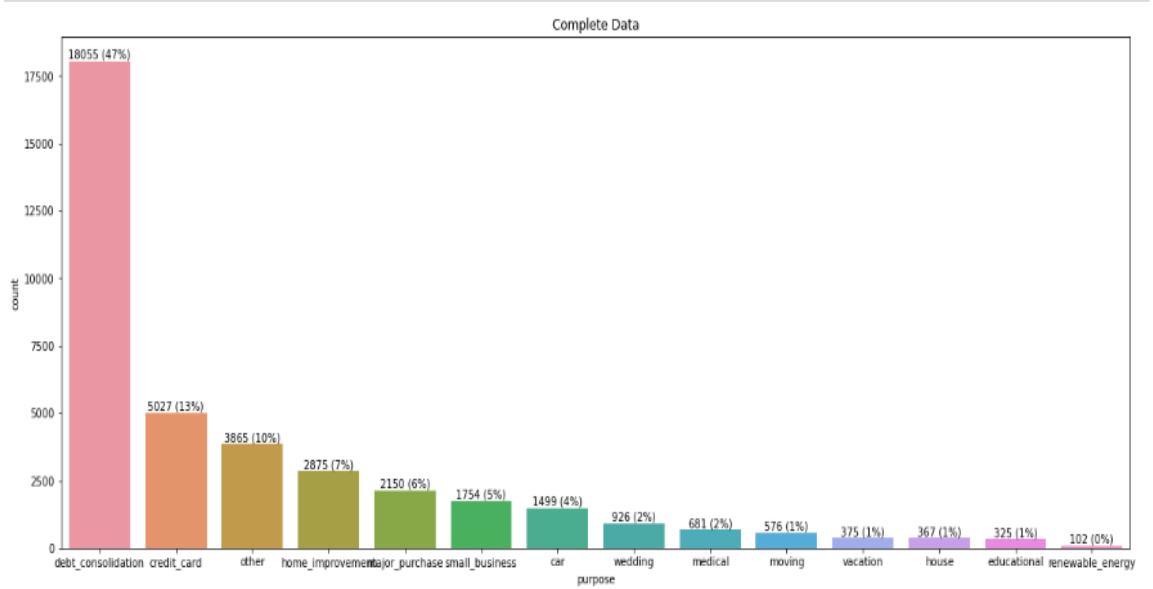
Univariate Analysis

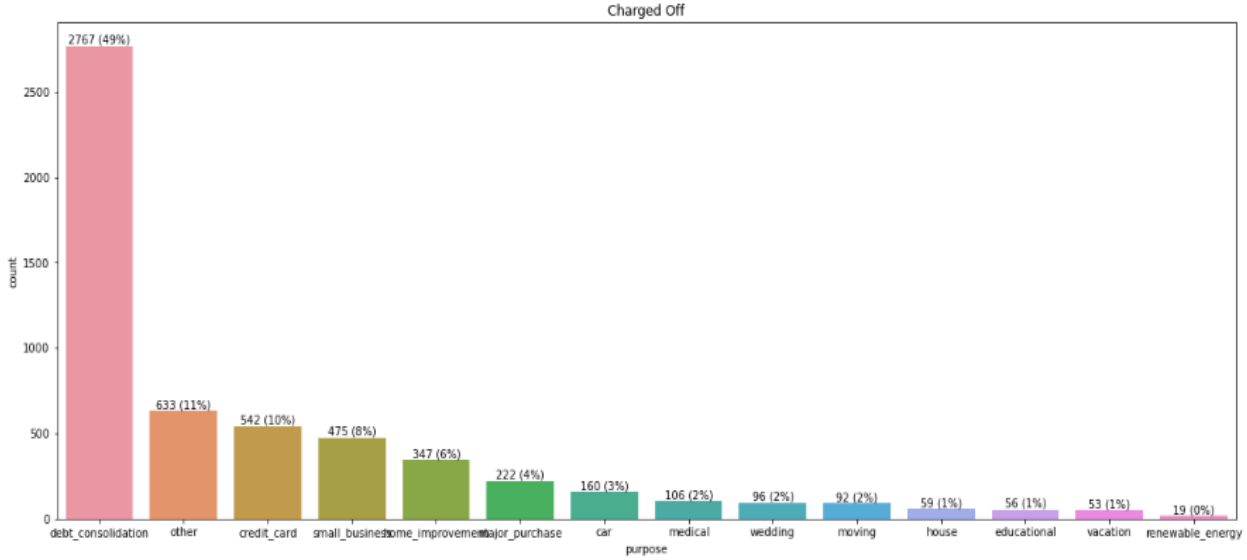
Loan Status



Average default rate is around 15 %

Loan Purpose

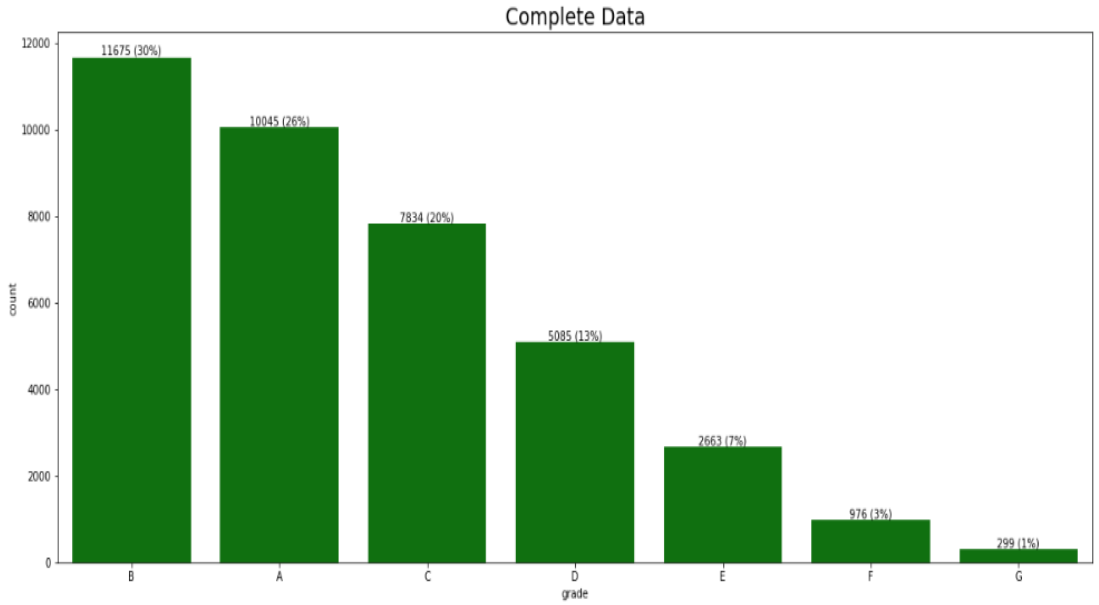


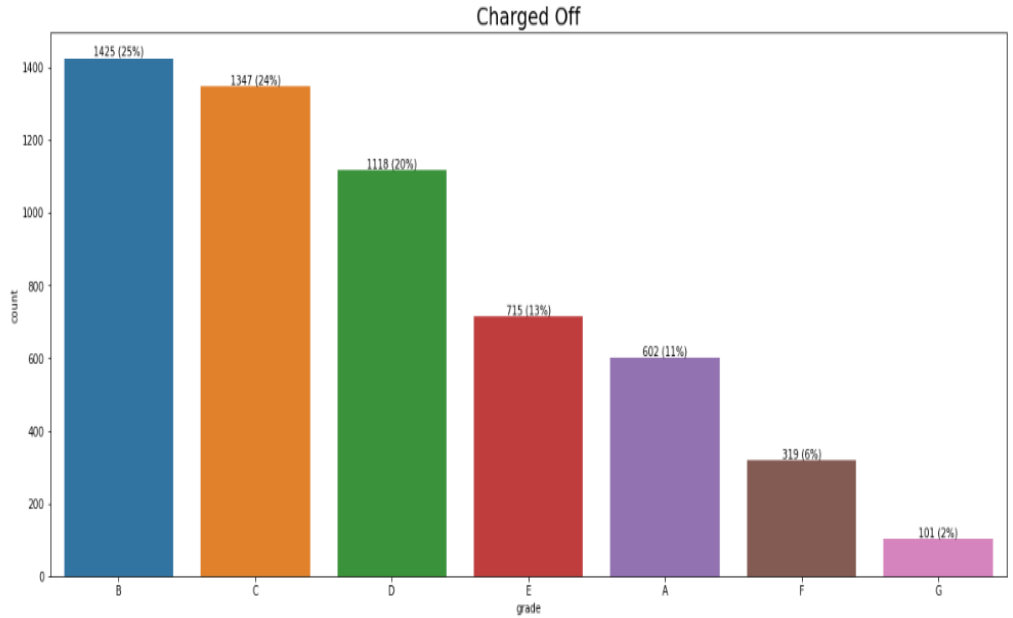


Major purpose of the loan are ::
debt_consolidation, credit_card, other, home_improvement, major_purchase, small_business, car
Order changed slightly for only Charged off cases ::
debt_consolidation, other, credit_card, small_business , home_improvement, major_purchase, car
small_business, debt consolidation, other categories percentage increased when we restrict the data set to "charged off".

So closer inspection is required before providing the loan for these purposes.

Grade

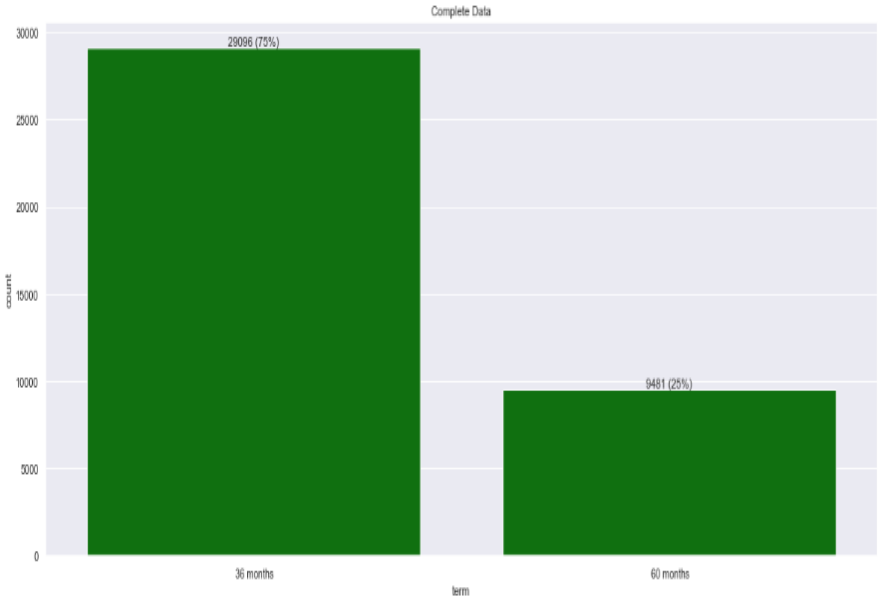


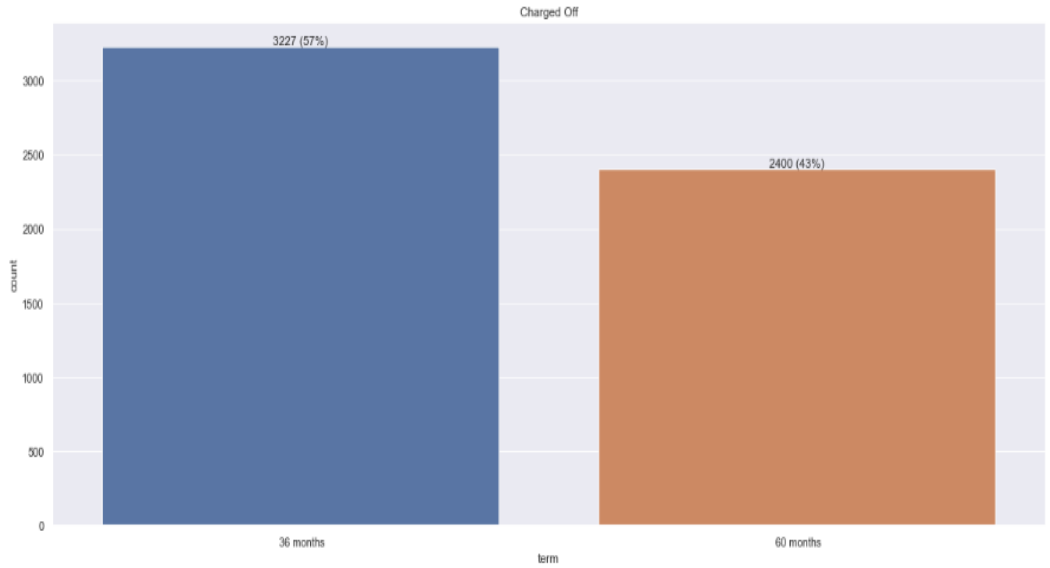


Almost 90% of loans are in Grades B, A, C, D

% of defaulters with respect to total number are less in Grade A compared to other Grades

Term

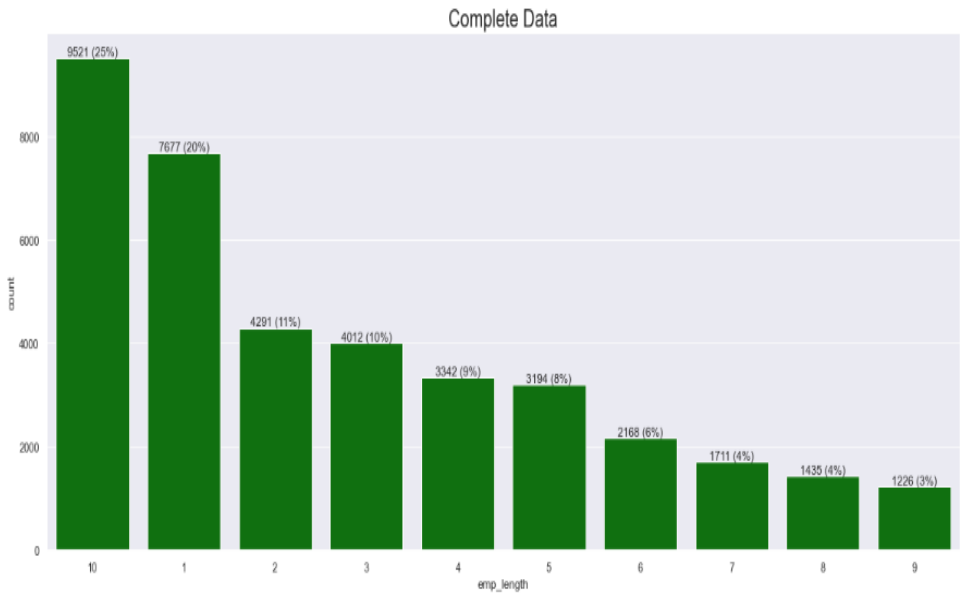


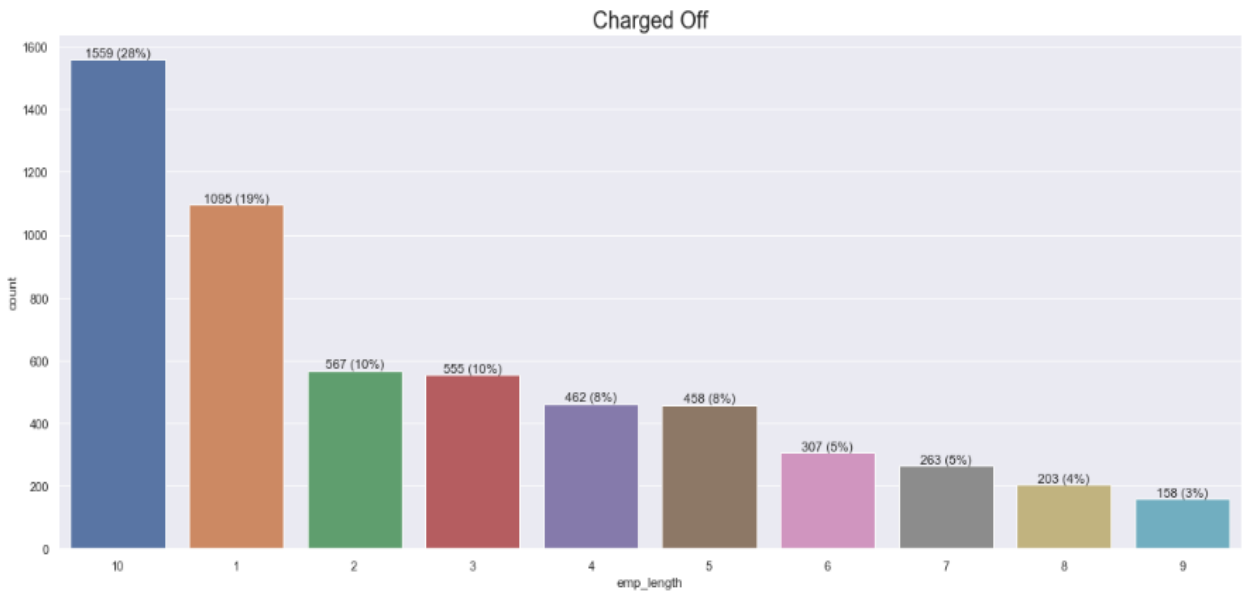


60 month terms are 25% of the total records but when it comes to defaulters the percentage increased to 43%.

Company should be more careful while awarding 60 months loan term.

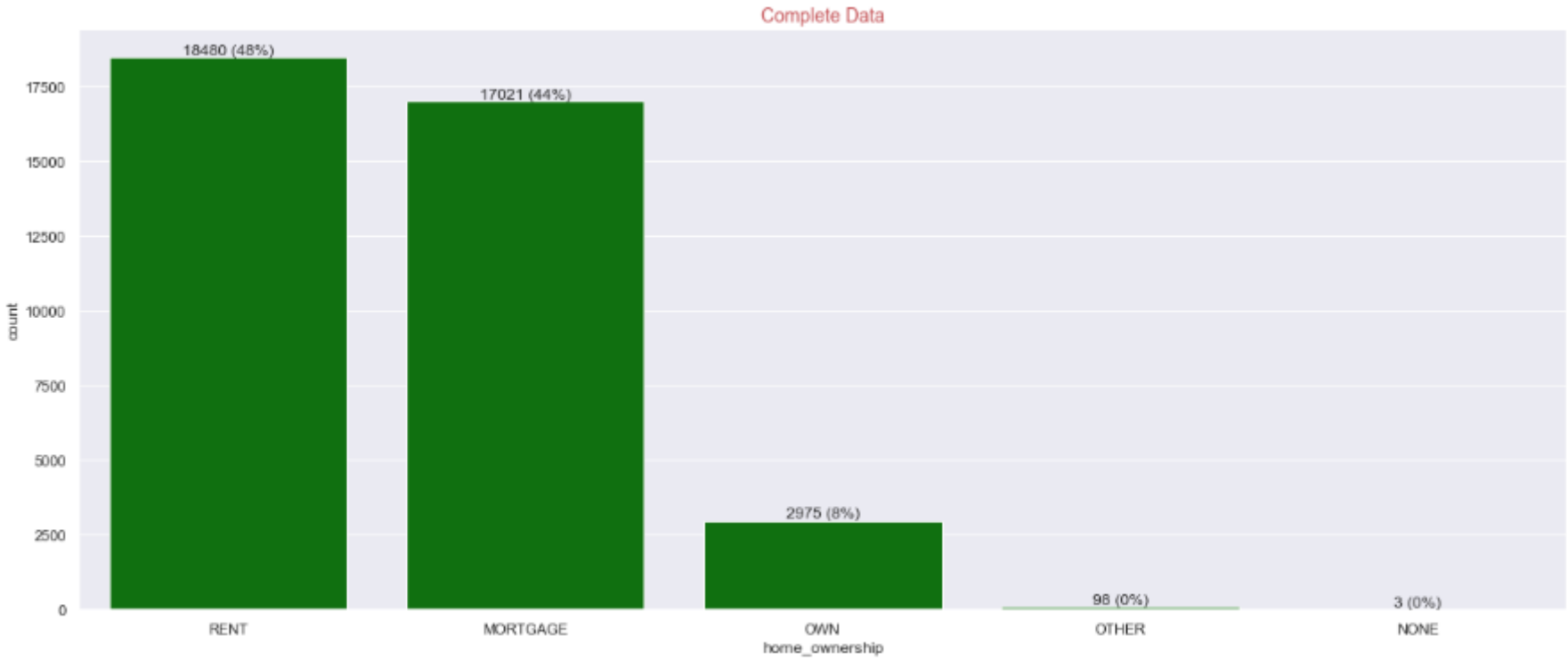
Employment Length





People with 1 year and 10 year experience are taking loans more frequently and they are the most defaulters also

Home Ownership

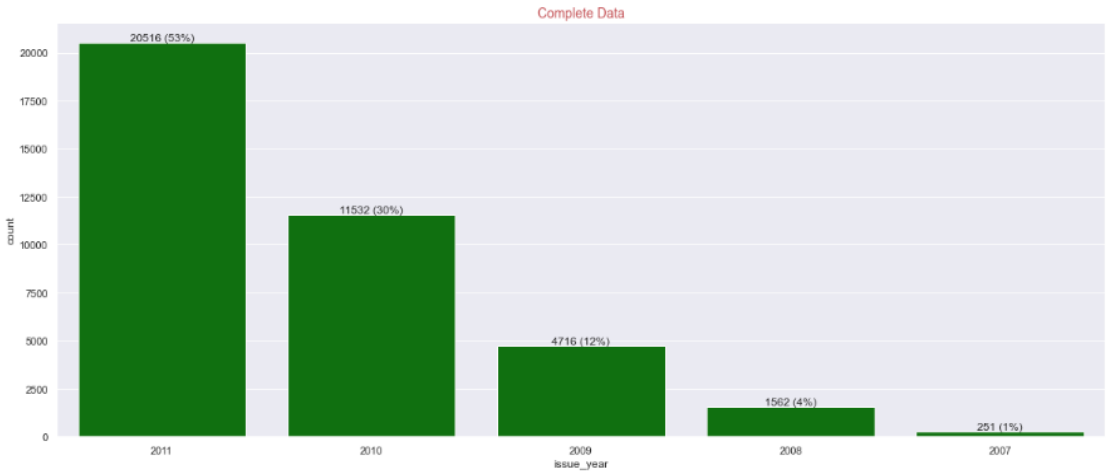


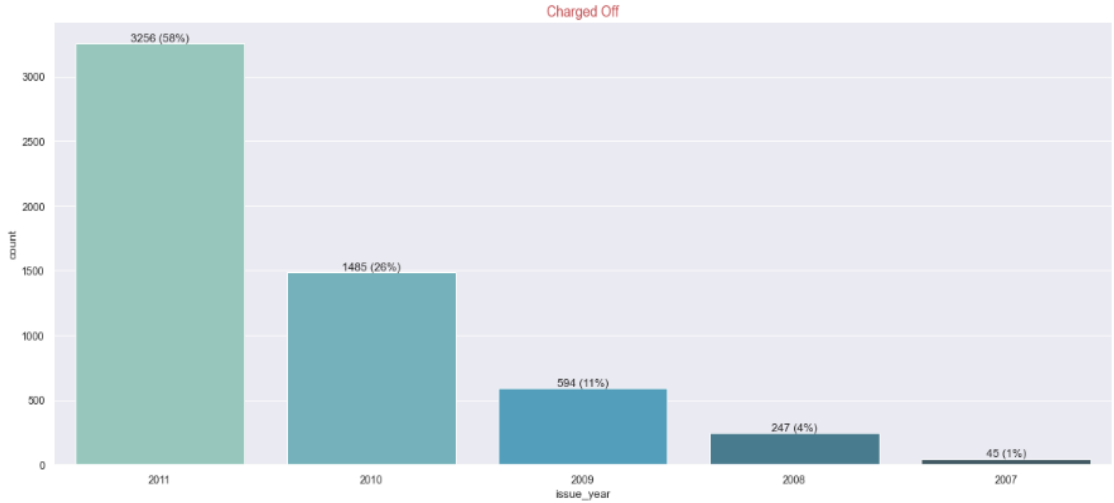
Loan applicants from people who owns a home are comparatively less

Verification Status

43% Loans are being issued even if the income is not verified

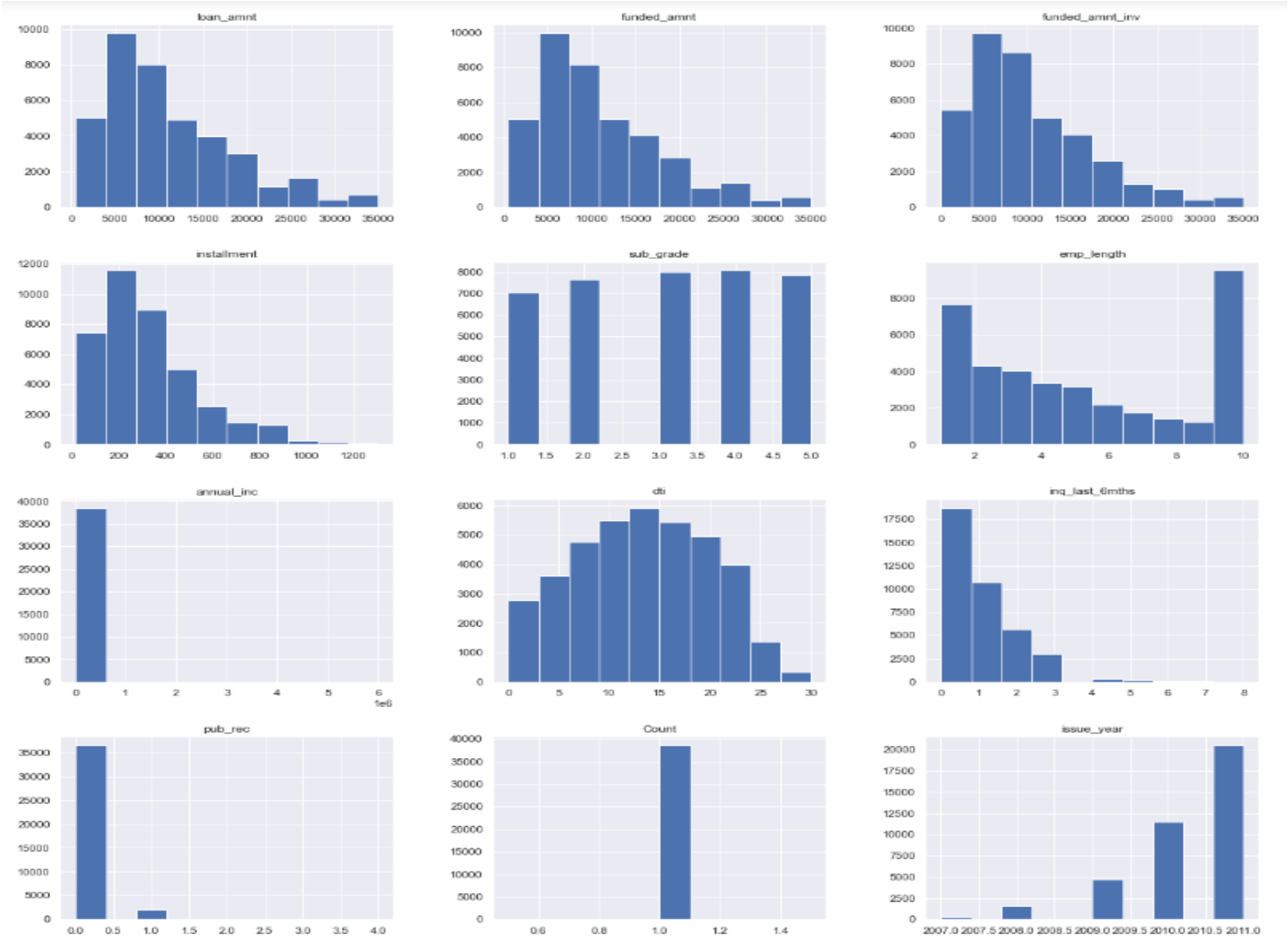
Loan Issue Year





There is a consistent increase in loan applicants each year. Number of defaulters are increasing every year more than double in comparison with previous year

Numeric Variable Distribution



- **75 % of loan amount request are for less than 15000**
- **dti is almost normally distributed**
- **It is very evident that there are some outliers in Annual income**

Derived Metrics

Derived Metrics

- **Since we have Loan Amount and Annual income fields , we can find out a new ratio Loan to Annual Income which will be helpful in our further analysis**

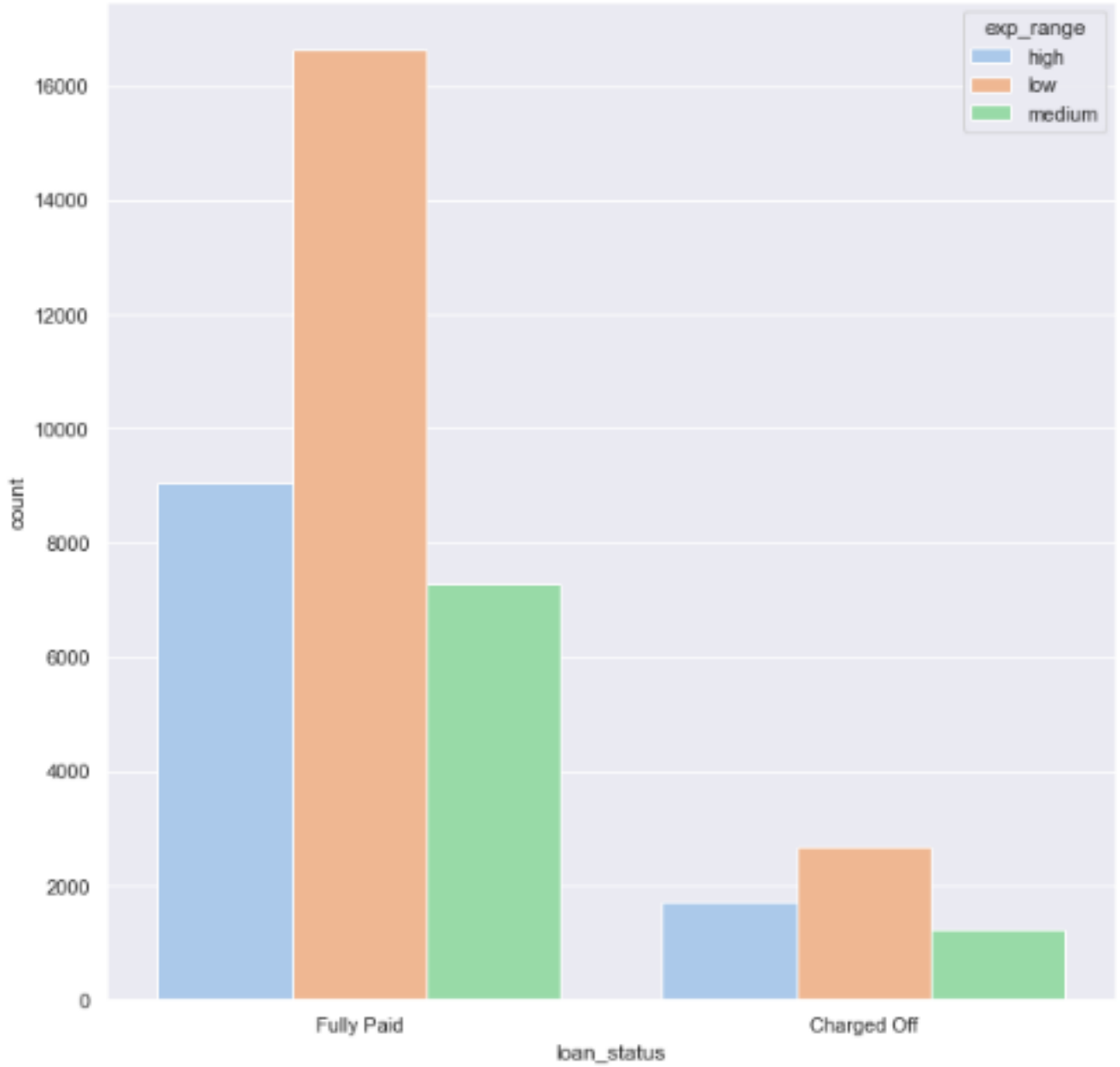
- **Created a new column captures dti range**
  low – Less than or equal to 10
  medium – Greater than 10 and less than or equal to 20
  high – Greater than 20

- **Loan to annual income range has been created**
  low – Less than or equal to 0.15
  medium – Greater than 0.15 and less than or equal to 0.25
  high – Greater than 0.25

- **Employee length range has been created**
  low – Less than or equal to 4
  medium – Greater than 4 and less than or equal to 8
  high – Greater than 8

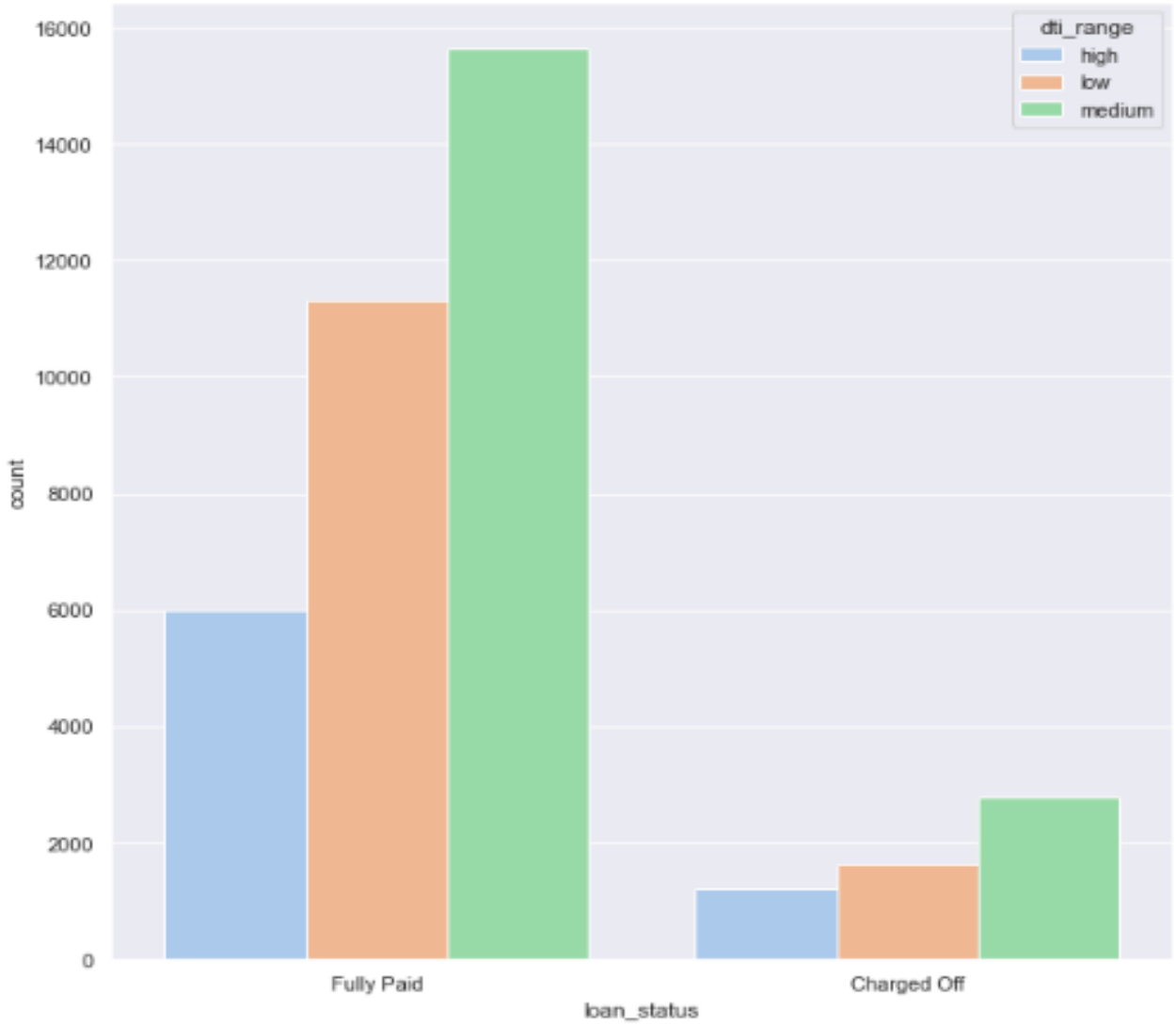
Bivariate Analysis

Loan Status vs Experience



- **people with less experience are having high chance of default**

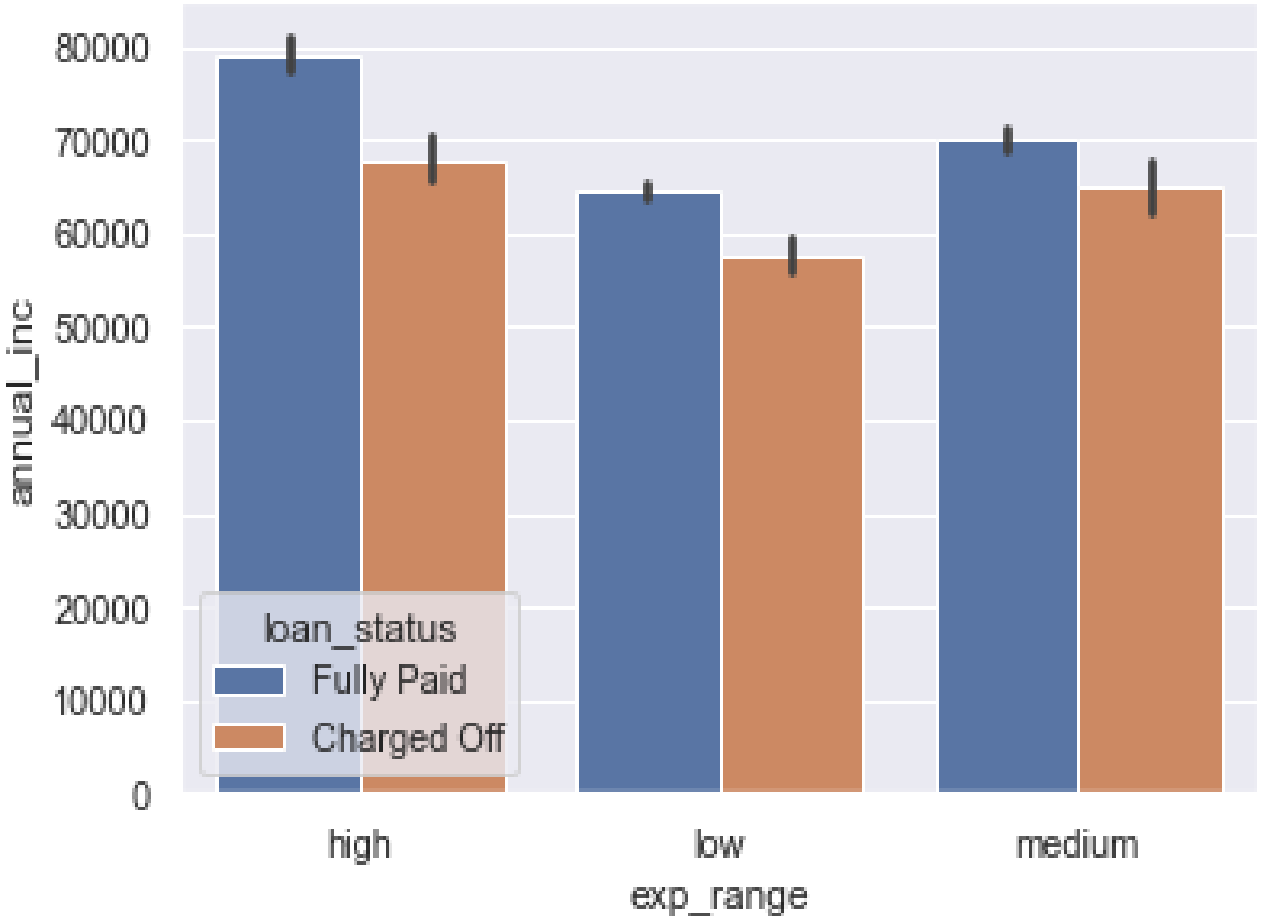
Loan Status vs dti Range



- **People in Medium dti range is taking more loans and proportionately default rate is also high in medium range**

Annual Income vs Experience

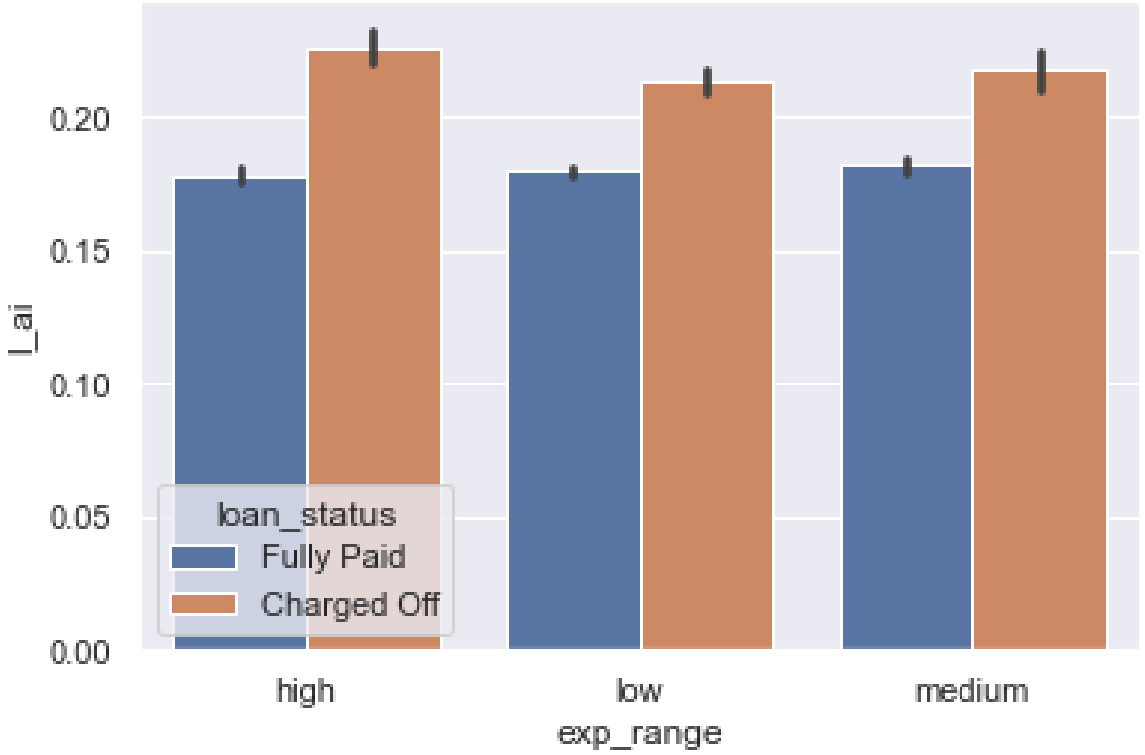
We compared Annual Income and experience within the purview of loan status.



- **Average Annual income of Fully paid applicants are always higher across experience range**

Loan to Annual Income Vs Experience

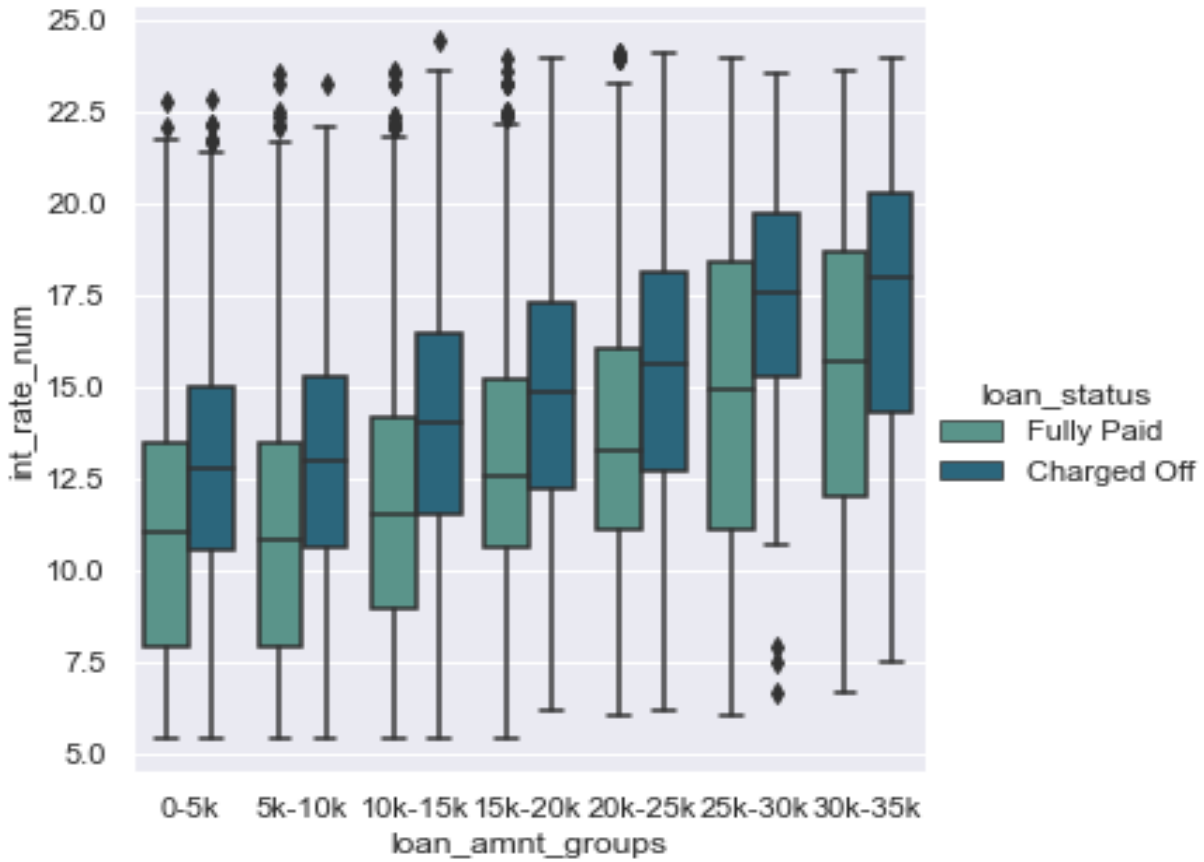
We compared Loan to Annual Income Ratio and experience within the purview of loan status.



- **Average Loan to annual income ratio is always higher in charge off applicants across experience range**

Interest Rate Vs Loan Amount Group

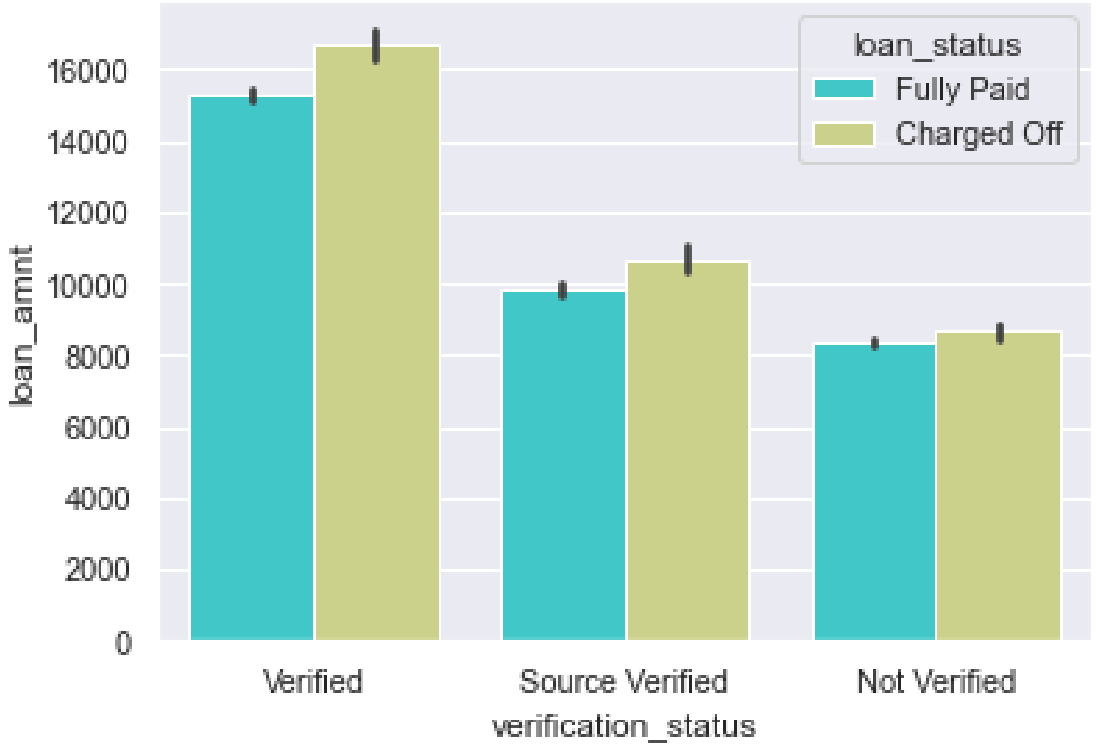
We compared Interest Rate and Loan Amount Group within the purview of loan status.



- **Interest rate distribution of Charged off loans are consistently higher across loan amount groups**

Loan Amount vs Verification

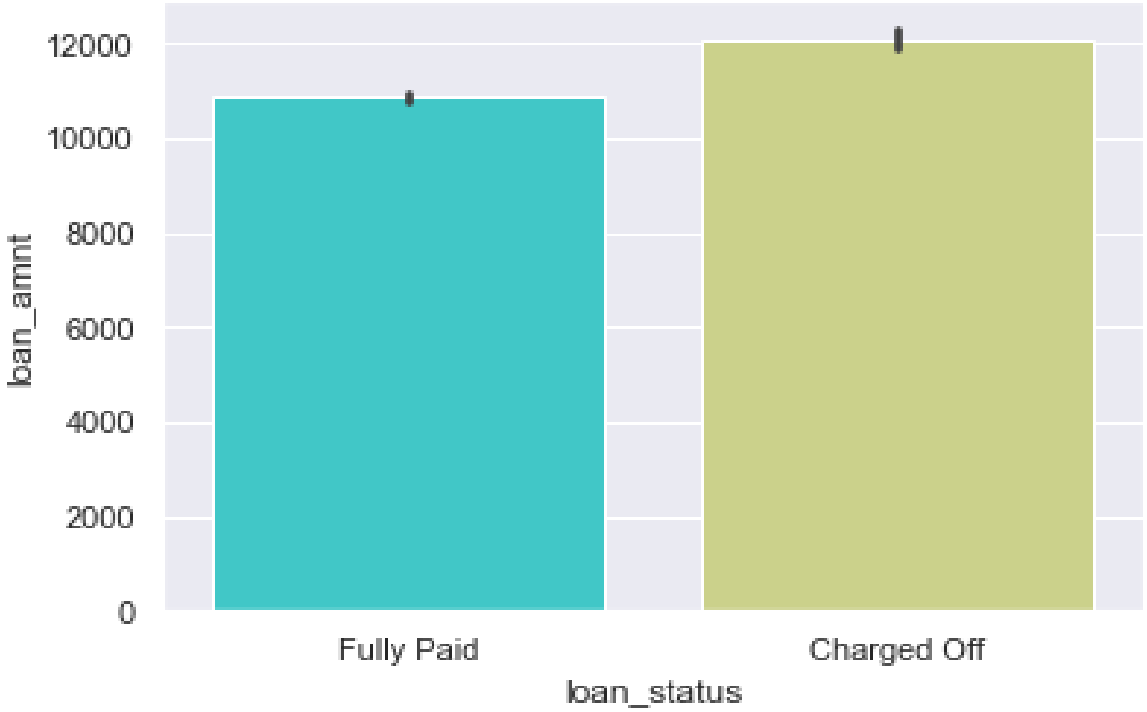
We compared Loan Amount and Verification status within the purview of loan status.



- **Loan amount of verified sources are higher. Either company is approving higher loan amounts after verification or Priority verification is being done on higher amount loans.**

Loan Amount vs Loan Status

We compared Loan Amount and Loan status.



- **Average loan amount of Charged off group is higher than Fully paid. This means tightening of loan amount eligibility scrutiny is required.**

Deductions from EDA

Inferences & Recommendations

- Average default rate is around 15%
- Loan default percentage increase for purposes small_business, debt consolidation, other categories when we restrict the data set to "charged off". So closer inspection is required before providing the loan for this purpose
- 60 month terms are 25% of the total records but when it comes to defaulters the percentage increased to 43%. Company should be more careful while awarding 60 months loan term
- % of defaulters with respect to total number are less in Grade A compared to other Grades
- People with 1 year and 10 year experience are taking loans more frequently and they are the most defaulters also.
- Loan applicants from people who owns a home are comparatively less. So it is less risky to award loan to a home owner.
- 43% Loans are being issued even if the income is not verified
- There is a consistent increase in loan applicants each year. Number of defaulters are increasing every year more than double in comparison with previous year
- 75 % of loan amount request are for less than 15000
- People with less experience are having high chance of default
- People in Medium dti range is taking more loans and proportionately default rate is also high in medium range
- Average Annual income of Fully paid applicants are always higher across experience range. It is comparatively less risky to give loan to people having high annual income.
- Average Loan Amount to annual income ratio is always higher in charge off applicants across experience range.

Thank You

24Slides