

LENDING CLUB CASE STUDY

- Rahul Gupta
- Kanhaiya Choudhary
- ML C39 Batch

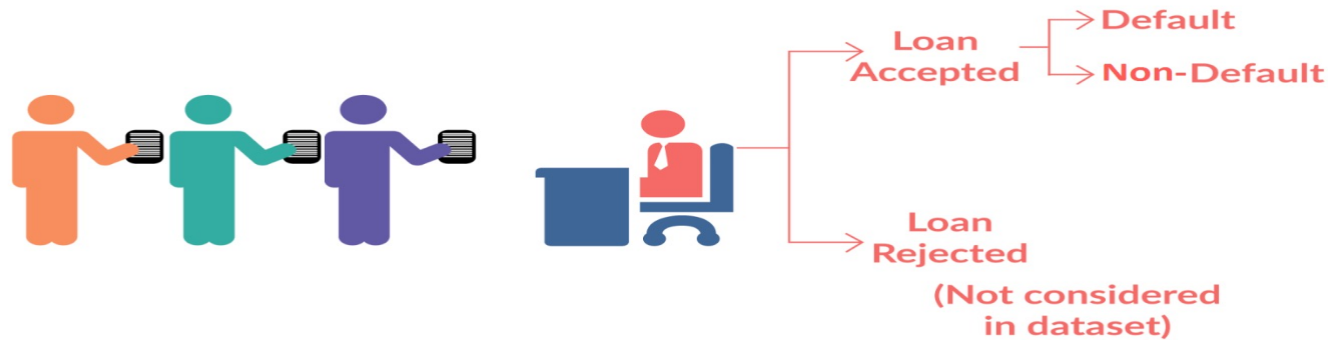


PROBLEM STATEMENT

- To understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of becoming a default.
- The company can utilise this knowledge for its portfolio and risk assessment.



LOAN DATASET



- **Fully paid:** Applicant has fully paid the loan (the principal and the interest rate)
- **Current:** Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
- **Charged-off:** Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan



PROBLEM SOLVING METHODOLOGY

- To understand how **consumer attributes** and **loan attributes** influence the tendency of default.
- Using EDA for this analysis
- Steps –
 - Data Cleaning
 - Data Understanding
 - Univariate Analysis
 - Segmented Univariate Analysis
 - Bivariate Analysis
 - Recommendation



DATA CLEANING

- Check the percentage of missing values
- Remove all those with very high missing percentage
- For columns with less missing percentage: perform Imputations
- Drop rows where the missing percentage is quite high
- Drop features/columns which are not needed for analysis

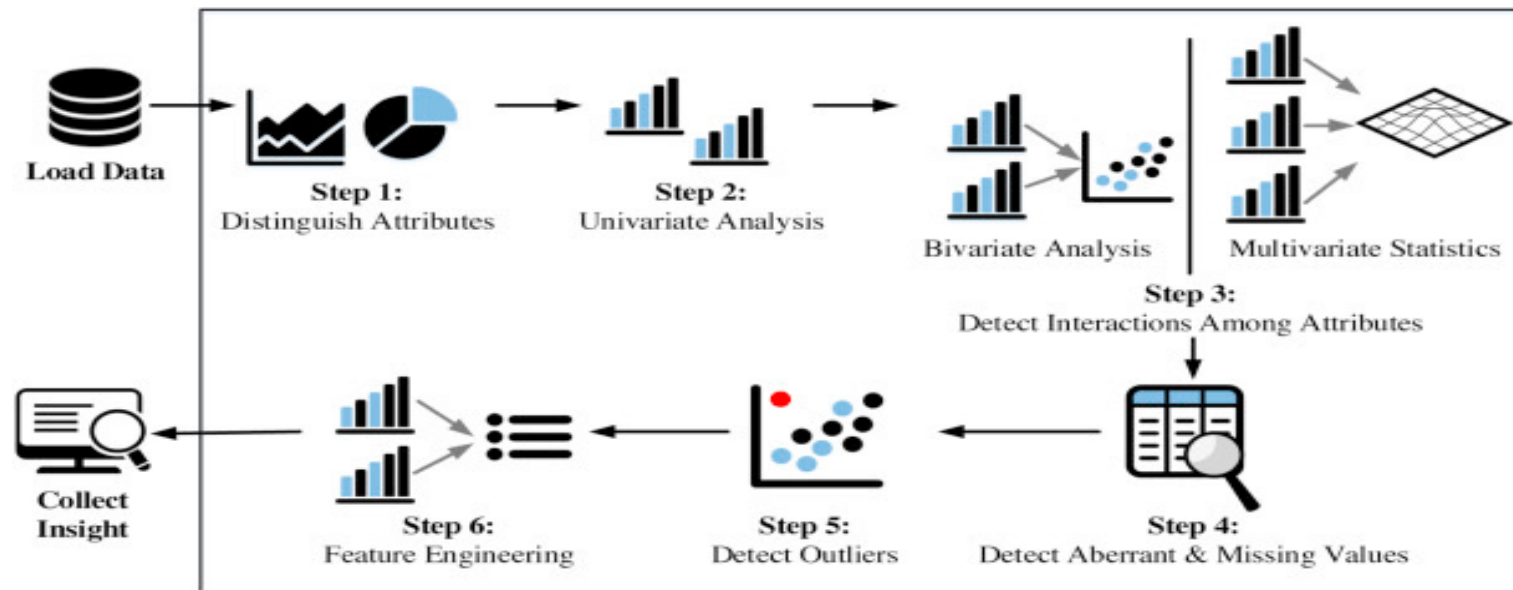


DATA ANALYSIS

- The objective is to identify predictors of default so that at the time of loan application, we can use those variables for approval/rejection of the loan.
 - There are broadly three types of variables –
 - those which are related to the applicant
 - Loan characteristics
 - Customer behavior variables
- Now, the customer behavior variables are not available at the time of loan application, and thus they cannot be used as predictors for credit approval.

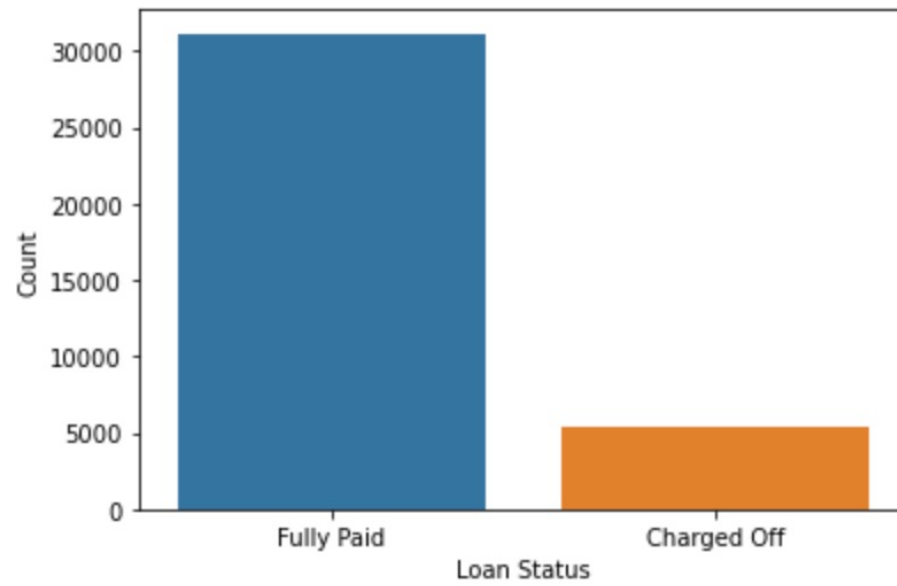


EDA PROCESS



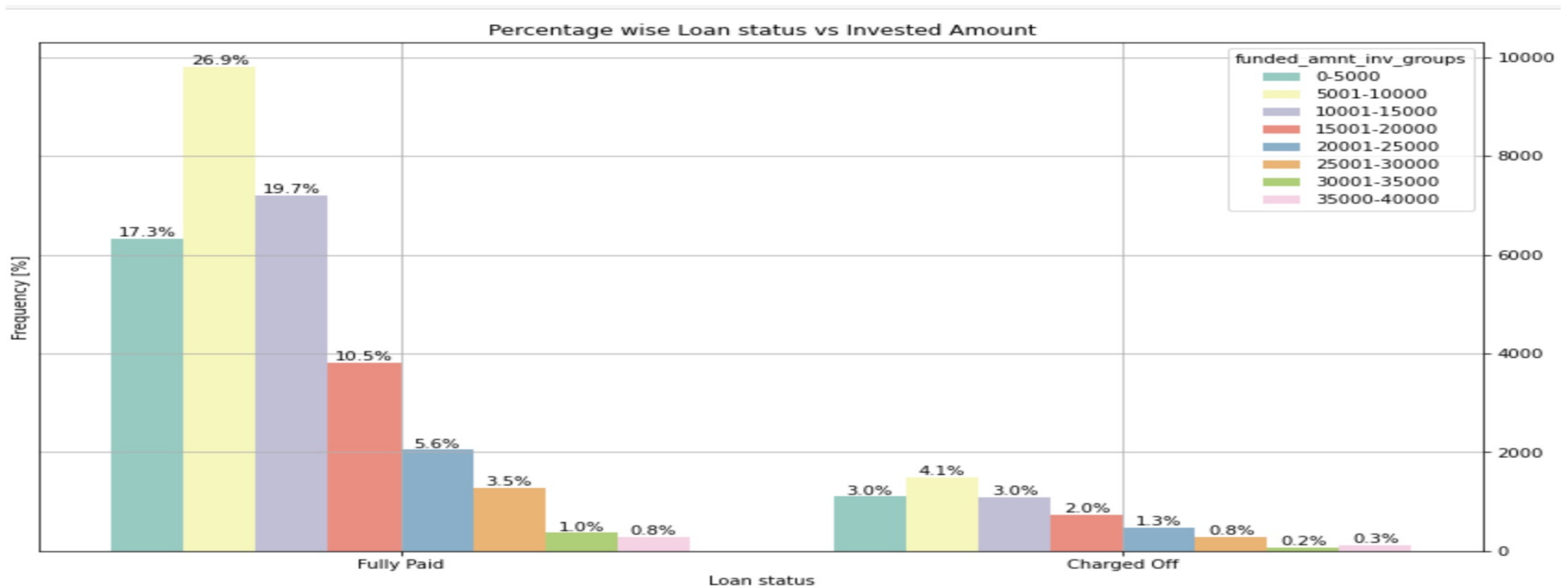
ANALYSIS





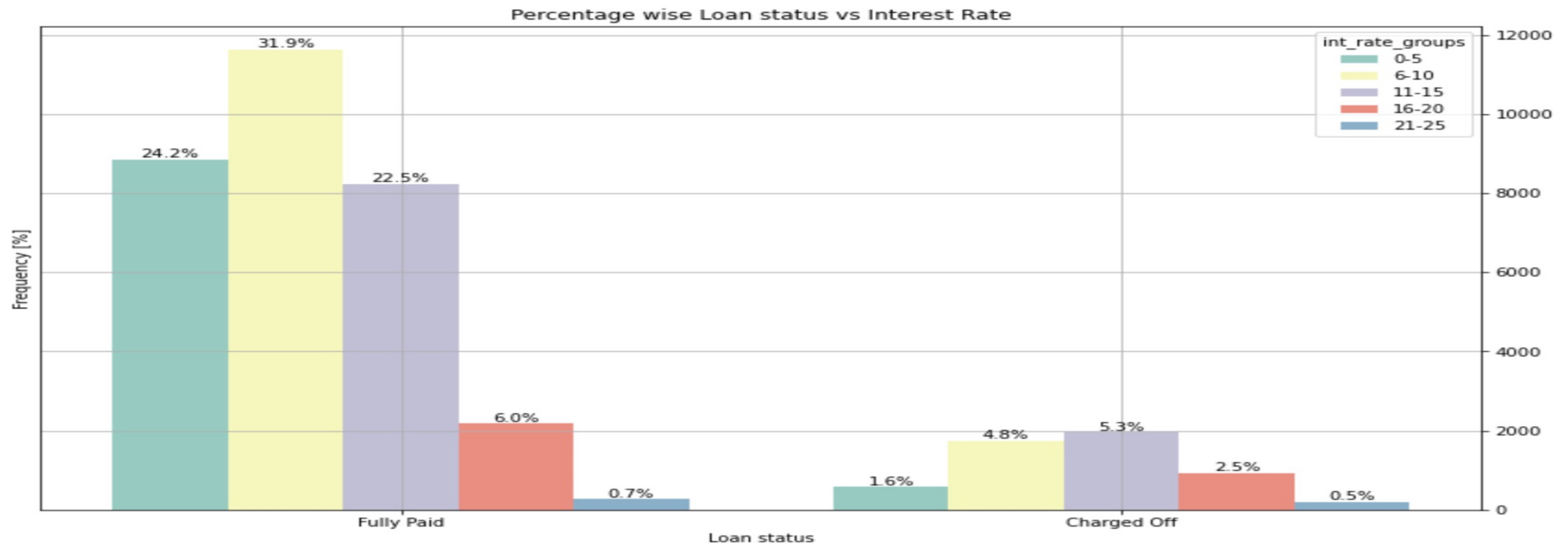
- Split of Fully Paid vs Charged Off (Defaulters)





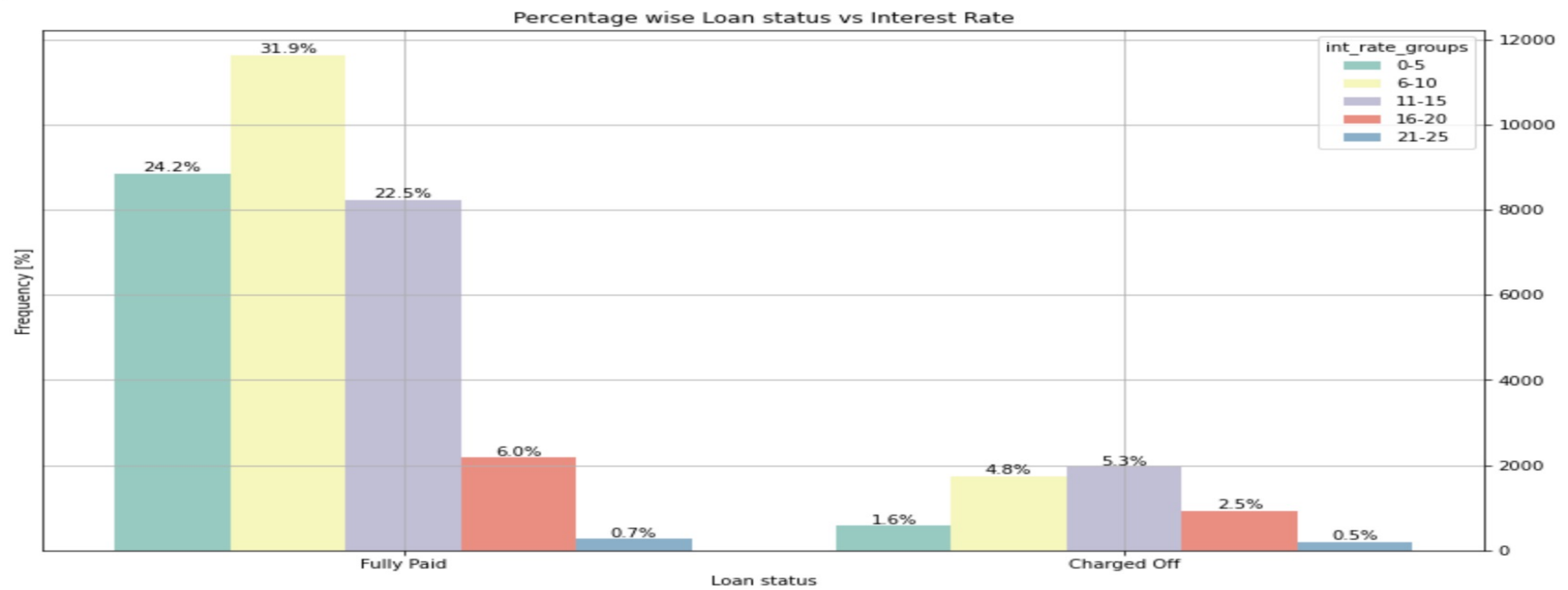
- We see funded amount invested is positively correlated with both Charged off and Fully Paid.
- We see that Employees who have exp 10+, 0, 2 & 3 years are likely to become Defaulters.





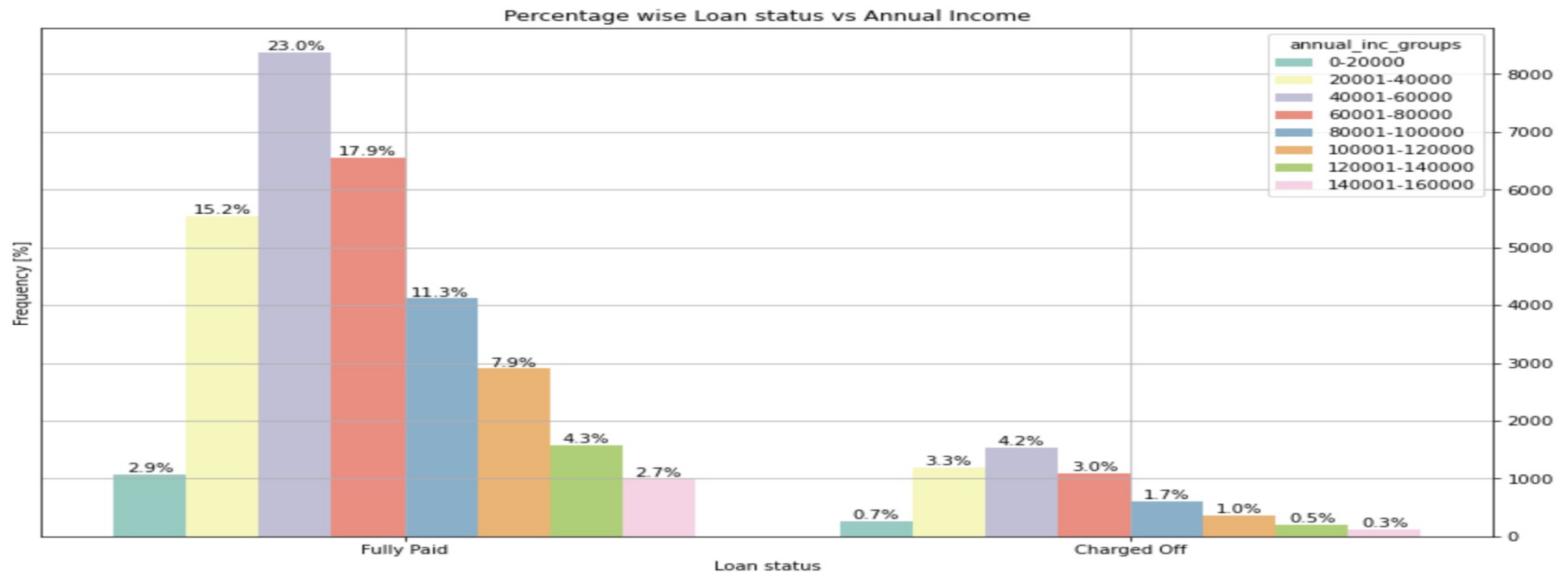
- *We see annual income is positively correlated with both Charged off and Fully Paid.*
- *We also see if annual income is in range of 40001 to 60000 then chances of becoming defaulter is higher.*
- *Looking at the ratio of, say charged off vs Fully Paid say for 0-20000, has higher chances of becoming defaulter.*





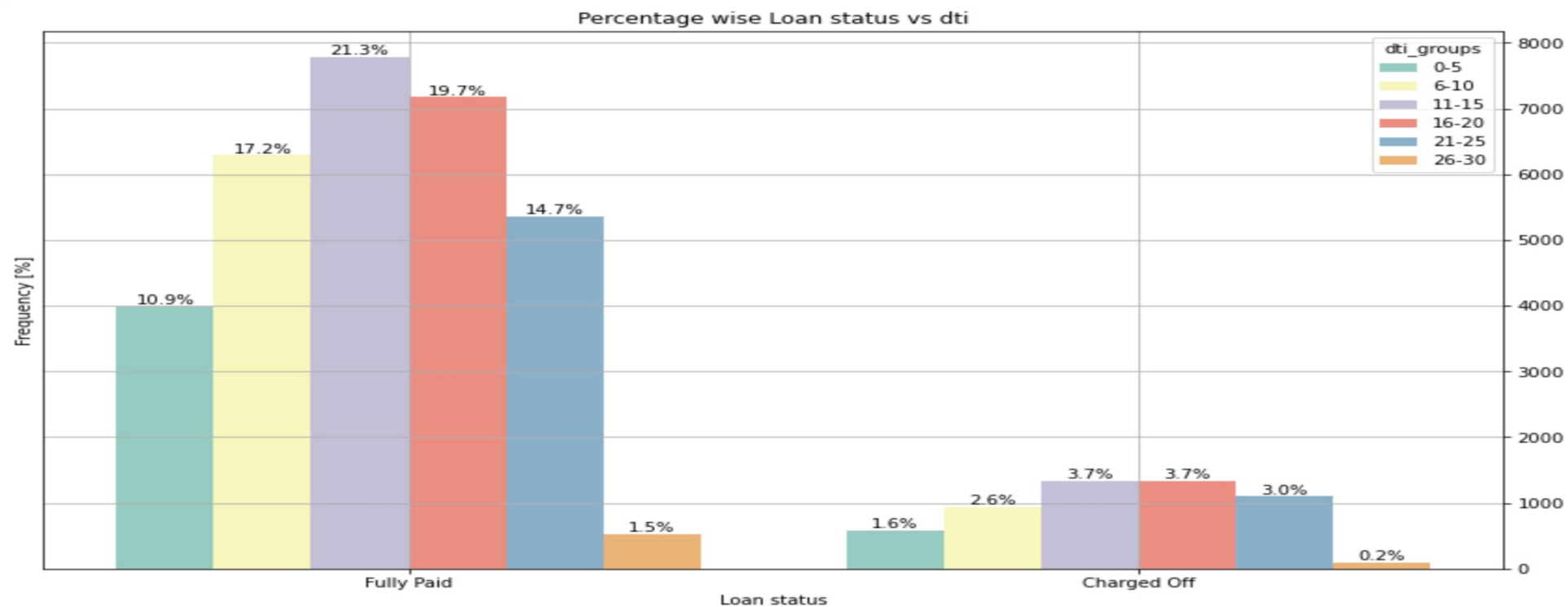
- *We see interest rate is positively correlated with both Charged off and Fully Paid.*
- *We also see if interest rate rate is in range of 11% to 15% then chances of becoming defaulter is higher.*
- *But ratio wise 21-25 % intereset rate group has higher chances of becoming defaulter.*





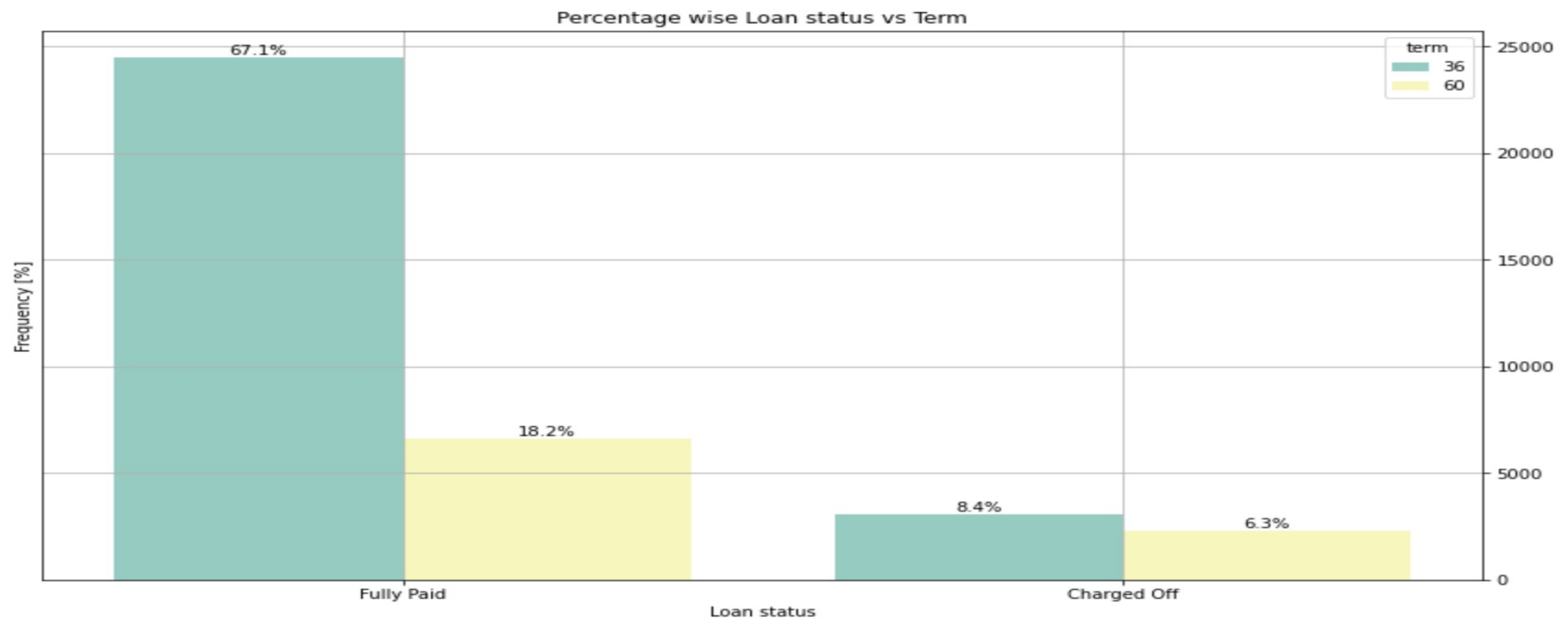
- We see annual income is positively correlated with both Charged off and Fully Paid.
- We also see if annual income is in range of 40001 to 60000 then chances of becoming defaulter is higher.
- Looking at the ratio of say charged off/ Fully Paid , say for the range 0-20000 has higher chances of becoming defaulter.





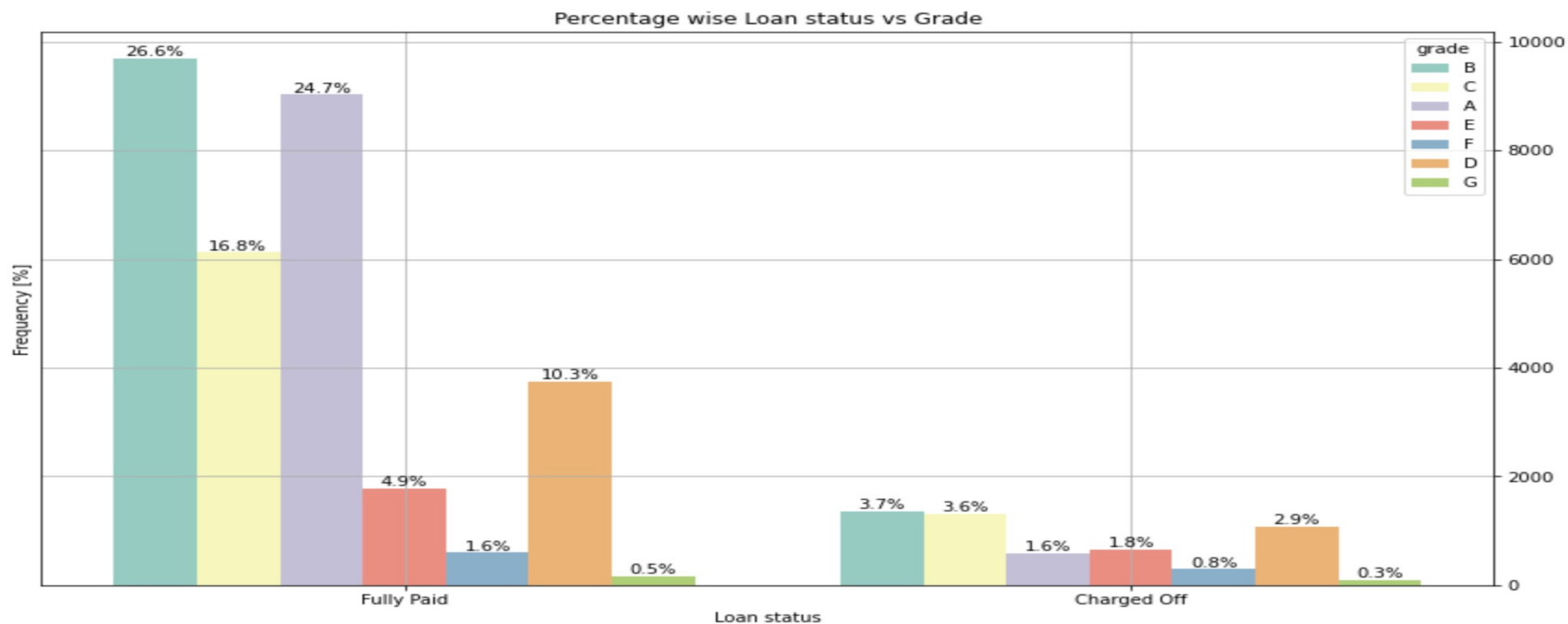
- We see *dti* is positively correlated with both Charged off and Fully Paid.
- We also see if *dti* is in range of either 11 to 15 or 16-20 then chances of becoming defaulter is higher.





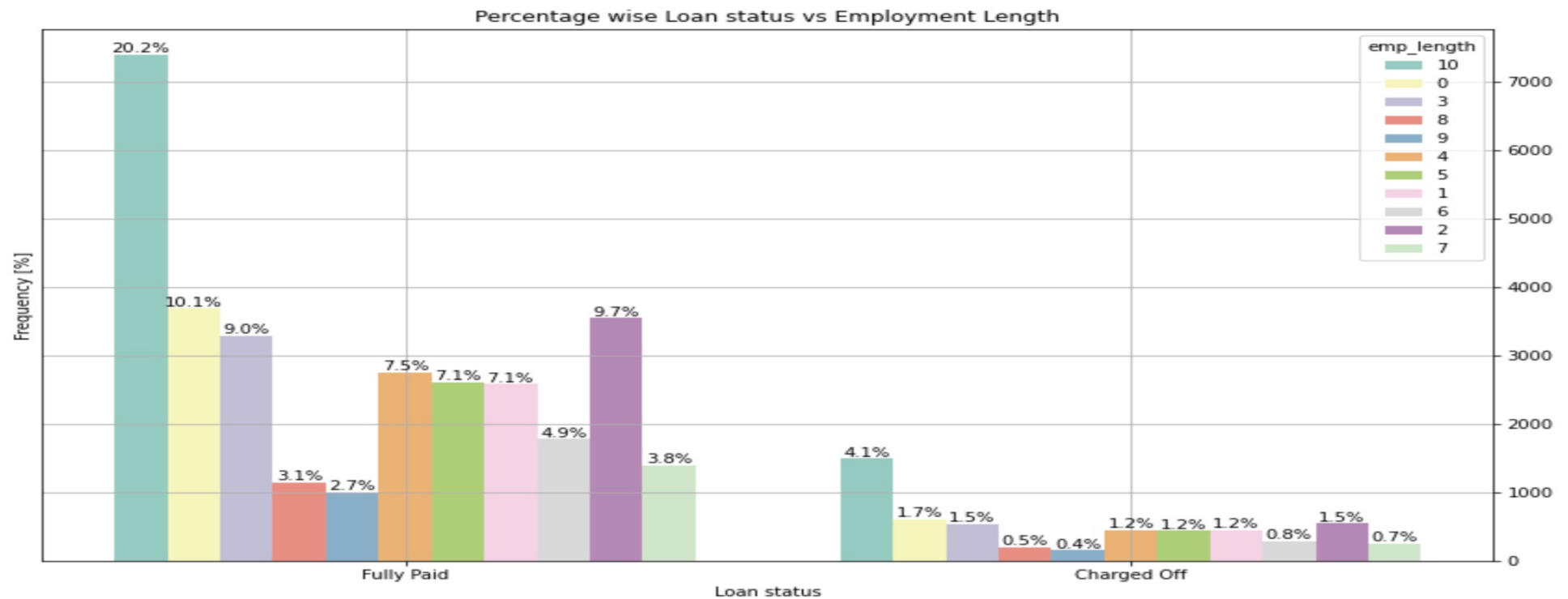
- *We see loan term is positively correlated with both Charged off and Fully Paid.*
- *We also see if term is 36 months then chances of becoming a defaulter is higher.*
- *But ratio wise 60 months has much higher number and hence higher chances of becoming a defaulter.*





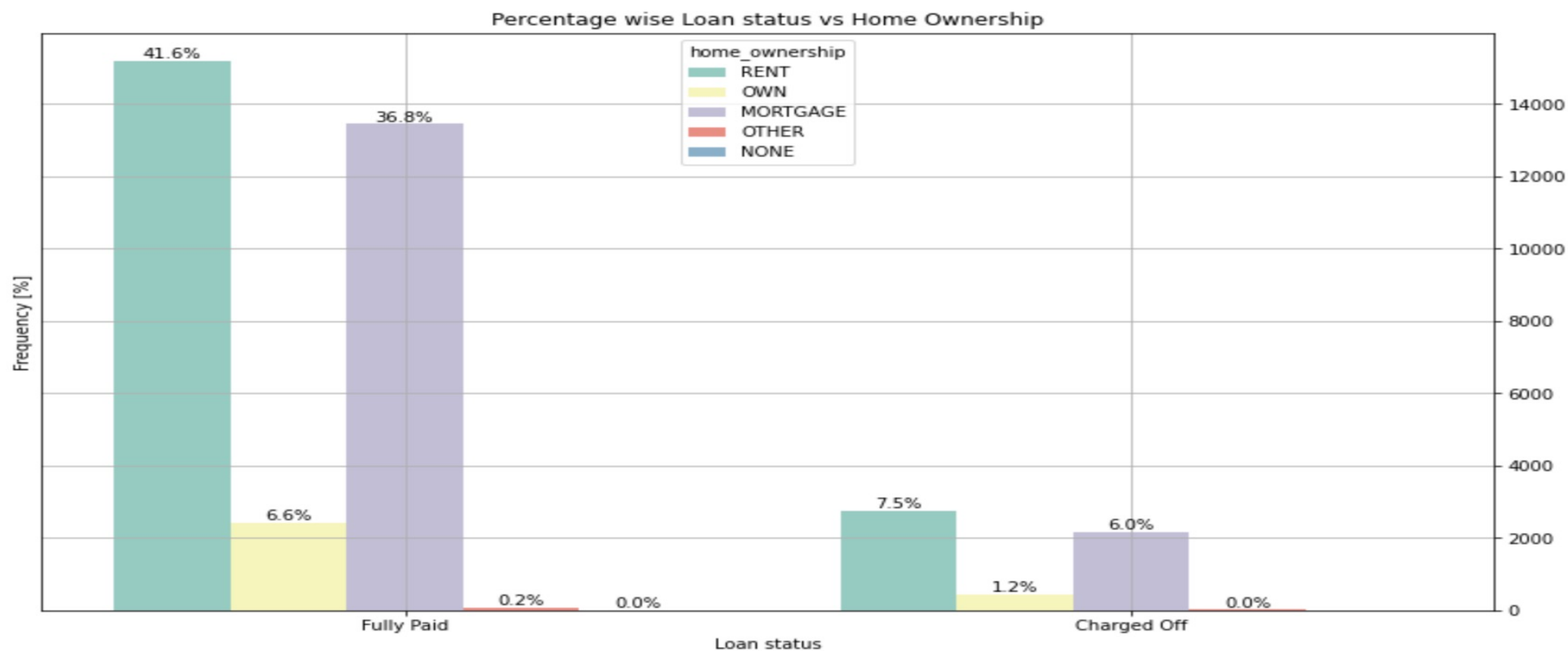
- We see grade is positively correlated with both Charged off and Fully Paid.
- We also see if grade is B, then chances of becoming defaulter is higher.
- But ratio wise G has higher chances of being defaulter.





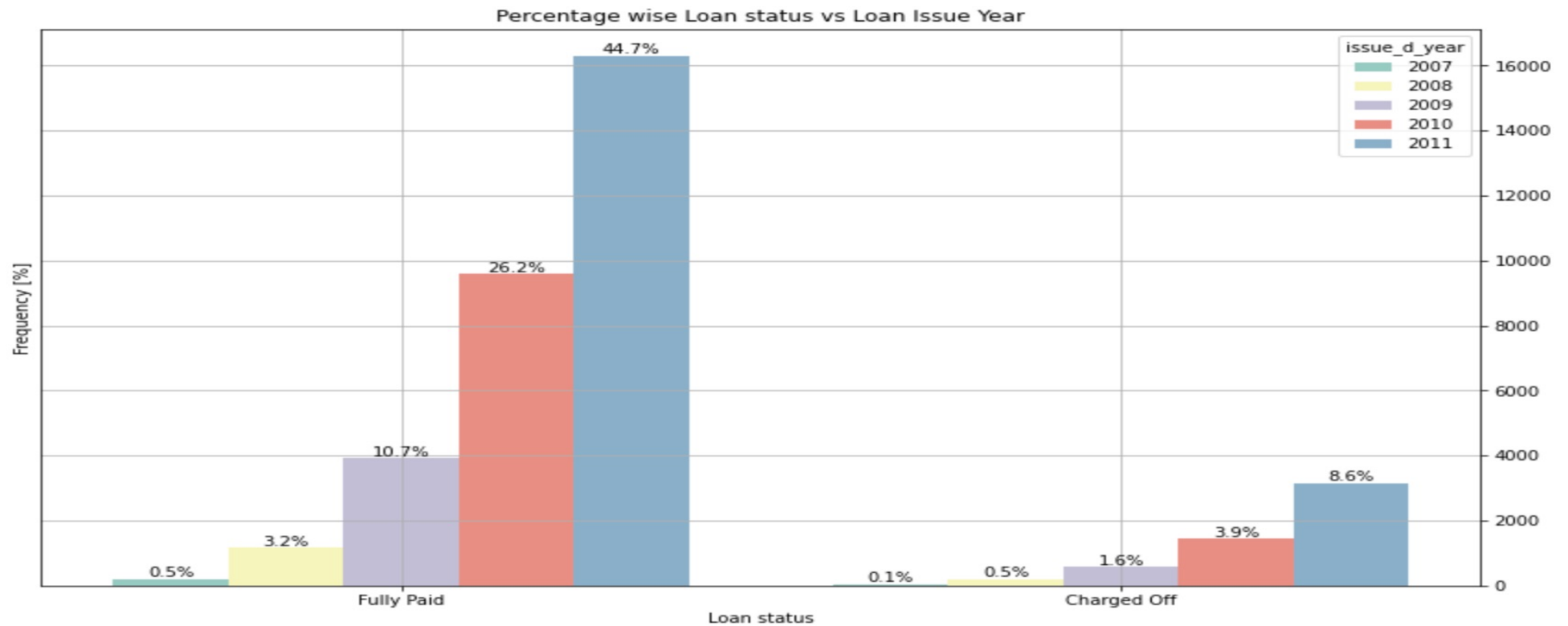
- We see employment length is positively correlated with both Charged off and Fully Paid.
- Basically, employment length may not be a good criteria to detect Defaulters.
- Also, we see that Employees who have exp 10+, 0, 2 & 3 years are likely to become Defaulters.





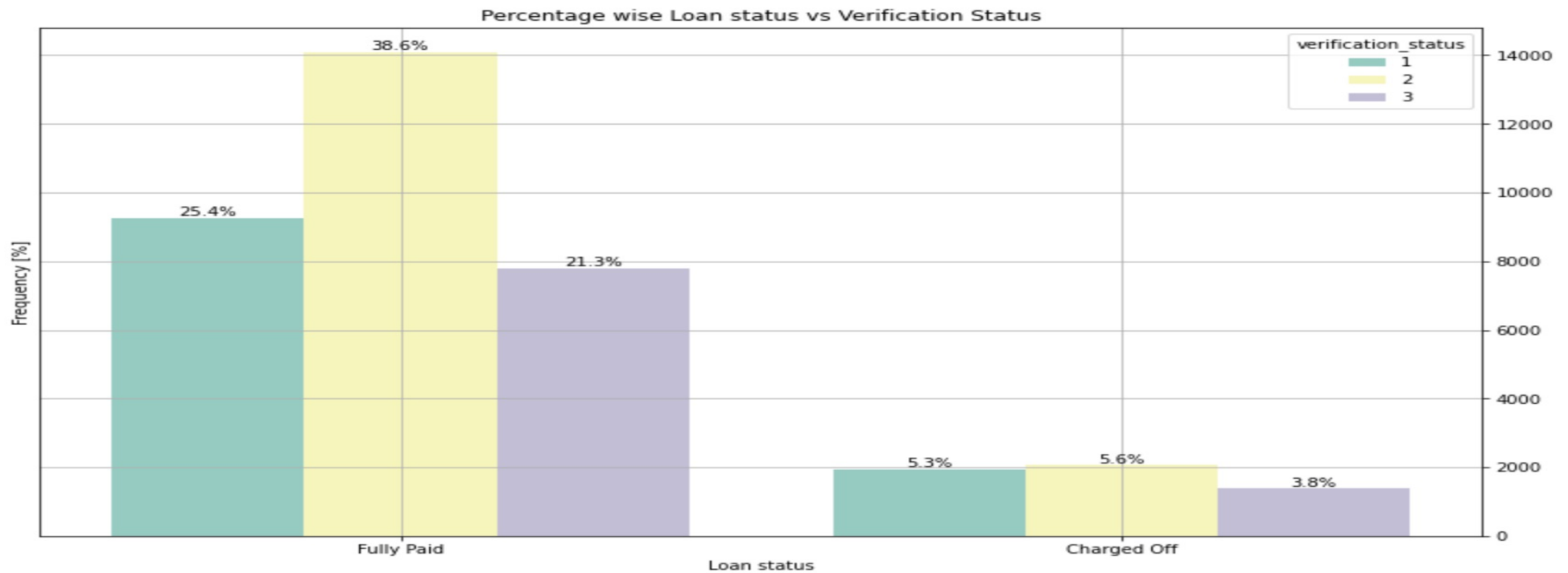
- We see home ownership is positively correlated with both Charged off and Fully Paid.
- We also see if home ownership is either RENT or MORTGAGE, then chances of becoming defaulter is higher.





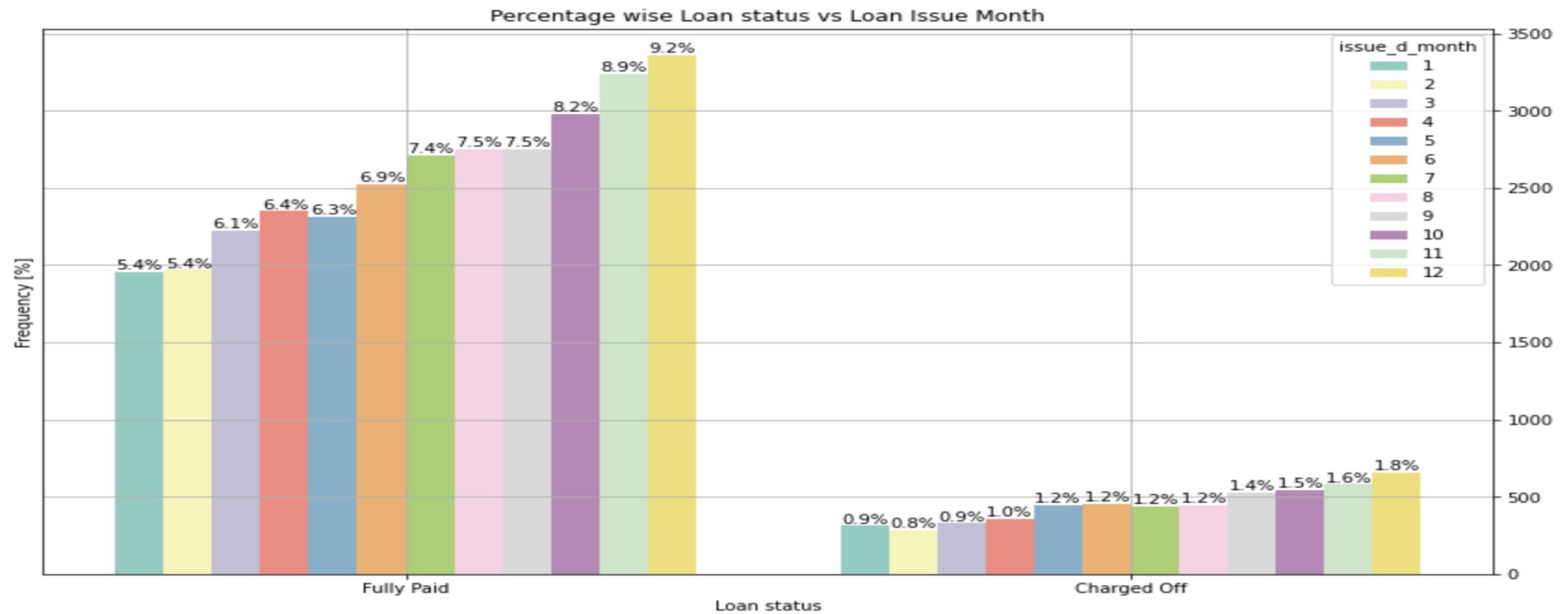
- *We see loan issue year is positively correlated with both Charged off and Fully Paid.*
- *We also see if loan issue year is 2011 then chances of becoming defaulter is higher.*





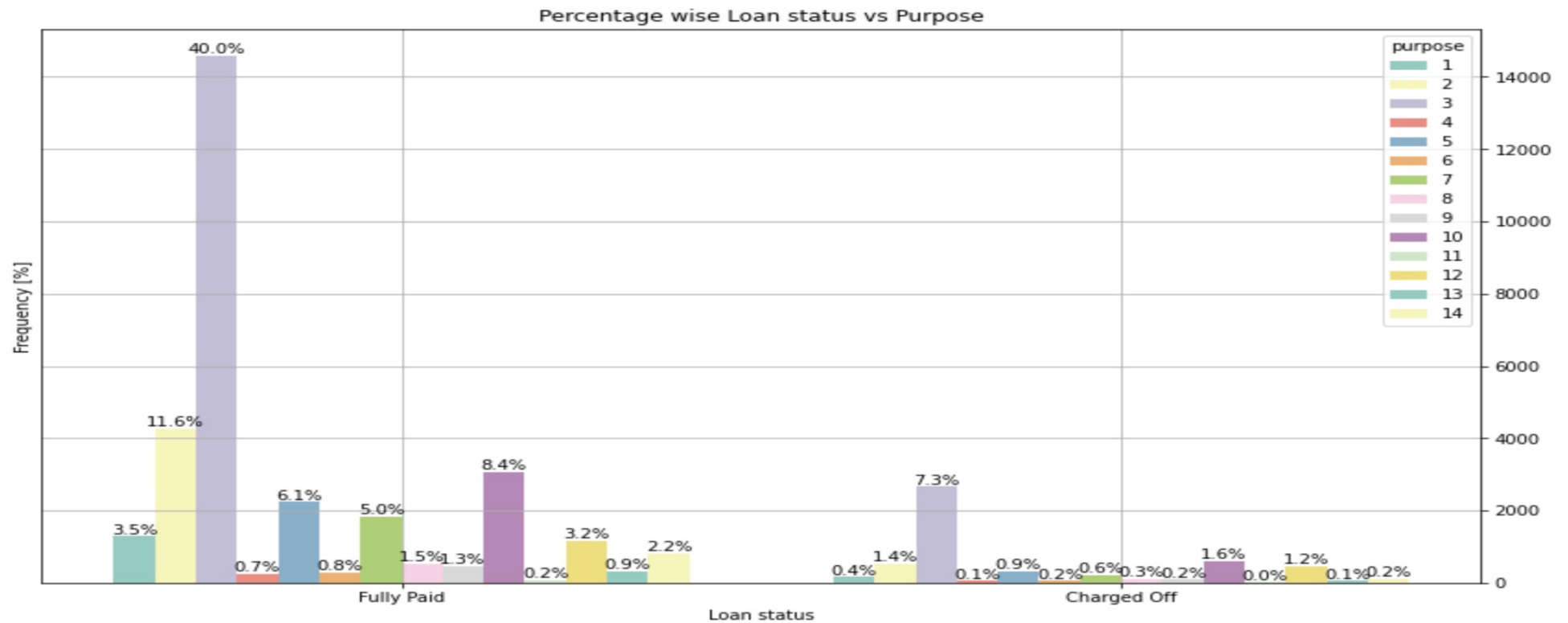
- *We see verification status is positively correlated with both Charged off and Fully Paid.*
- *We also see if verification status is not verified then chances of becoming defaulter is higher.*
- *But percentage wise Verified also has higher chances of being defaulter.*





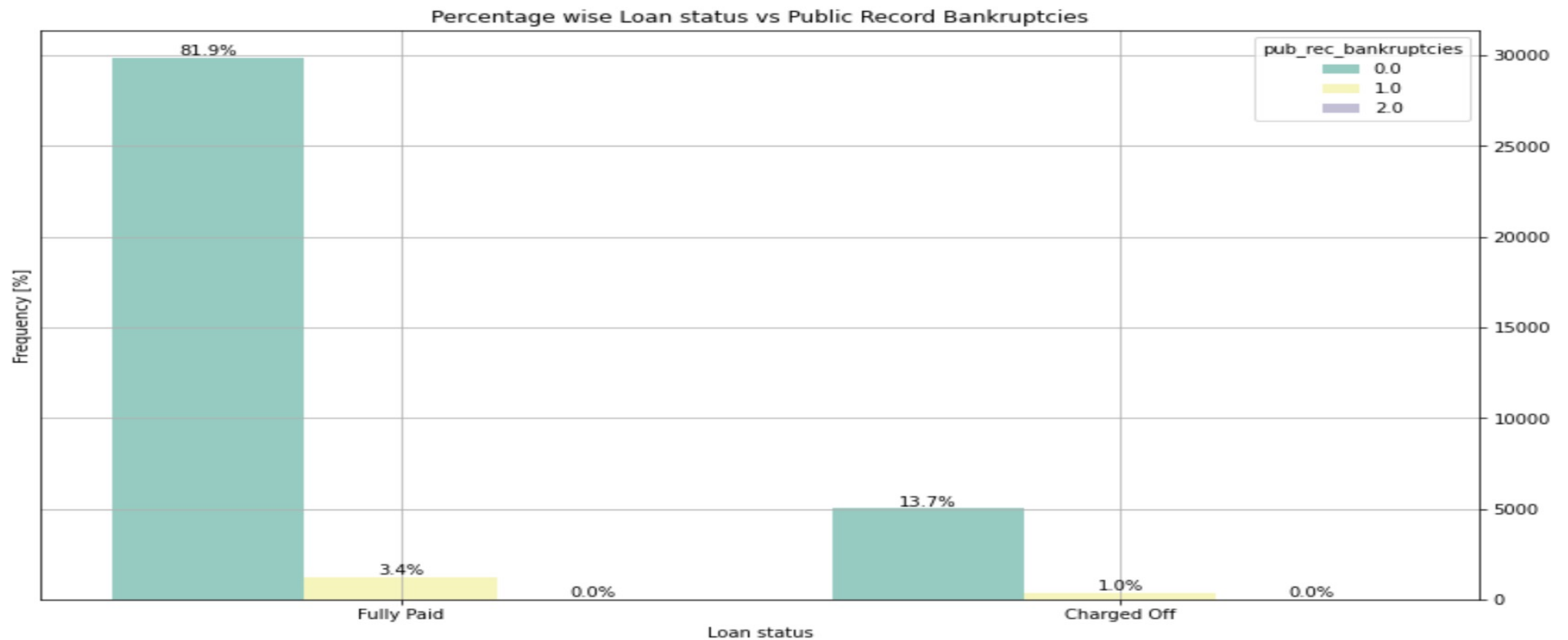
- *We see Loan Issue month is positively correlated with both Charged off and Fully Paid.*
- *We also see if loan issue month is either Dec or Nov, then chances of becoming defaulter is higher.*





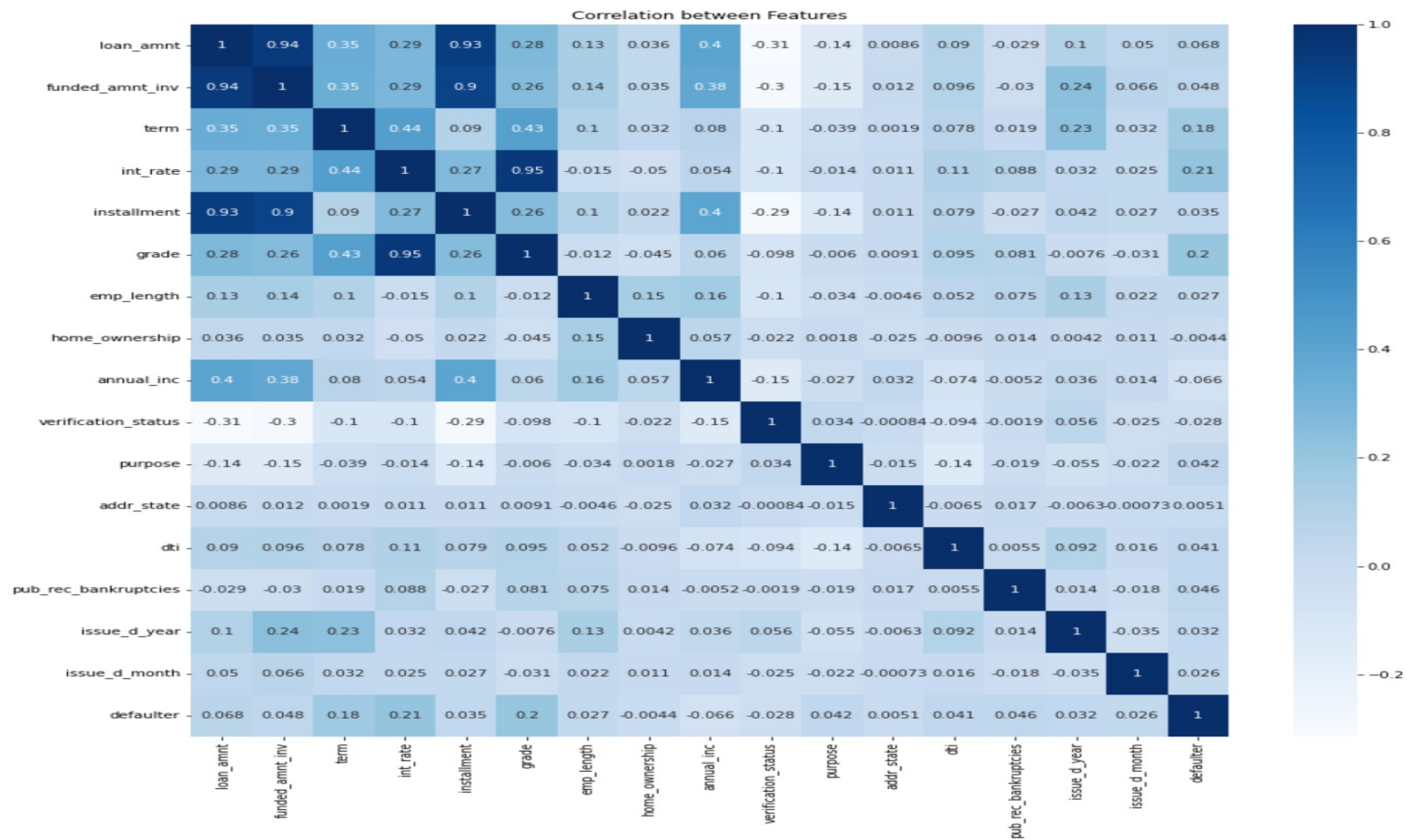
- We see purpose is positively correlated with both Charged off and Fully Paid.
- We also see if purpose is debt_consolidation, then chances of becoming defaulter is higher.





- We see number of public record bankruptcies is positively correlated with both Charged off and Fully Paid.
- We also see if number of public record bankruptcies is 0 then chances of becoming a defaulter is higher.





We plotted this plot/heatmap to show correlation of different features.



- **Summary based on the correlation heatmap:**

- Negatively correlated features

- Home ownership, borrower's annual income and verification status are negatively correlated with defaulter.
 - a. If home ownership is "RENT" or "Mortgage", then chances of becoming defaulter is higher.
 - b. If verification status is "Verified", then chances of becoming defaulter is higher.
 - c. For Annual income , looking at the ratio of say charged off/ Fully Paid, say for 0-20000 range, has higher chances of becoming defaulter. Heatmap is also confirming this.

- Positively correlated features

- Key features which are positively correlated are term, interest rate and grade.
- Other remaining features are also positively correlated, but the extent of correlation is less as compared to term, interest rate and grade.
 - a. So higher the term, greater will be chances of becoming defaulter. Hence 60 months is the likely term with more percentage of defaulters.
 - b. Higher the interest rate means higher the chances of becoming defaulter. Hence likely range with higher percent of defaulters would be 21-25%
 - c. Higher the grade means higher the chances of becoming defaulter. Hence grade G is likely range with higher percent of defaulters.

