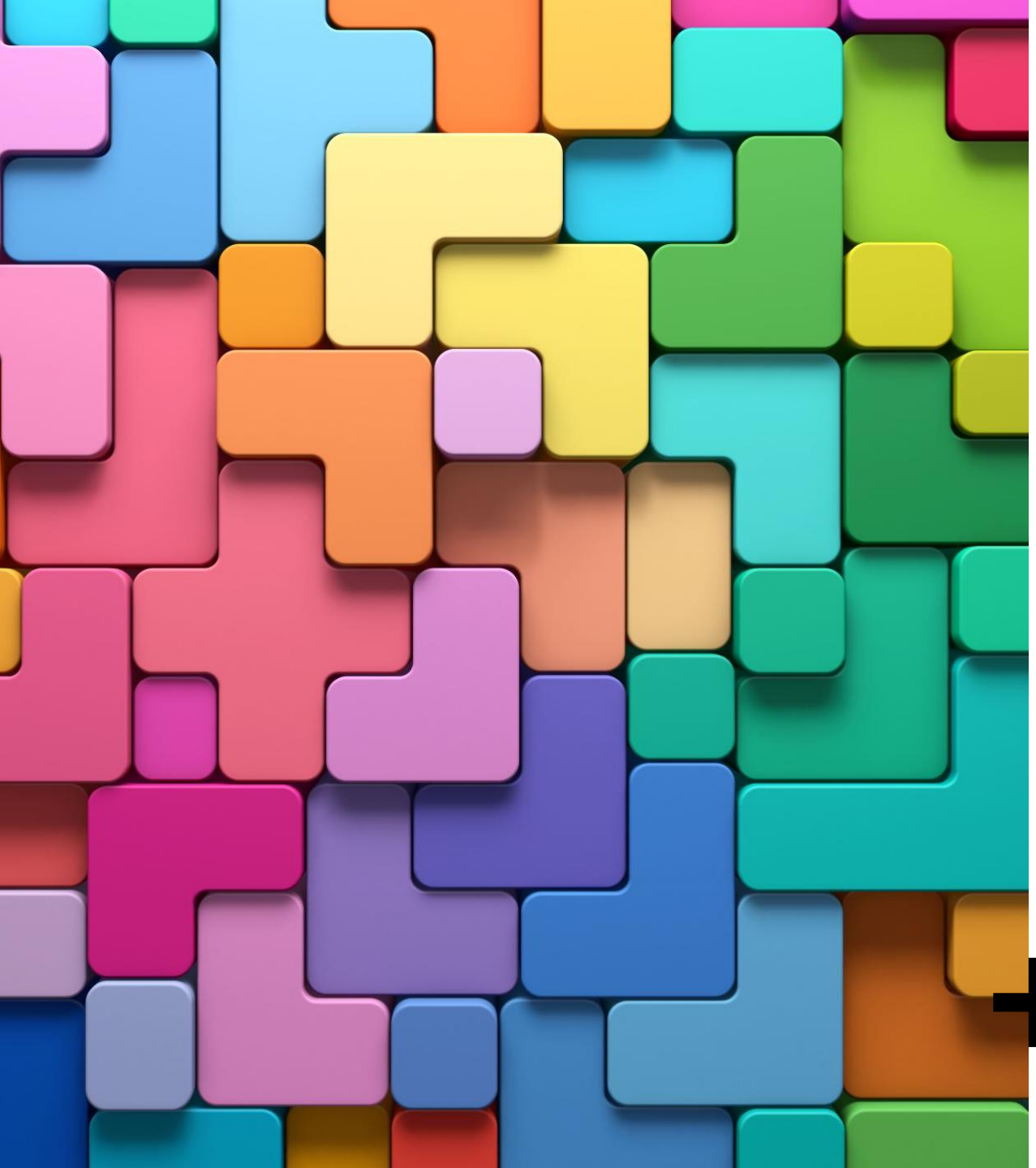


Lending Club Case Study

Milton Nepoli A

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 utilize this knowledge for its portfolio and risk assessment

Loan Dataset

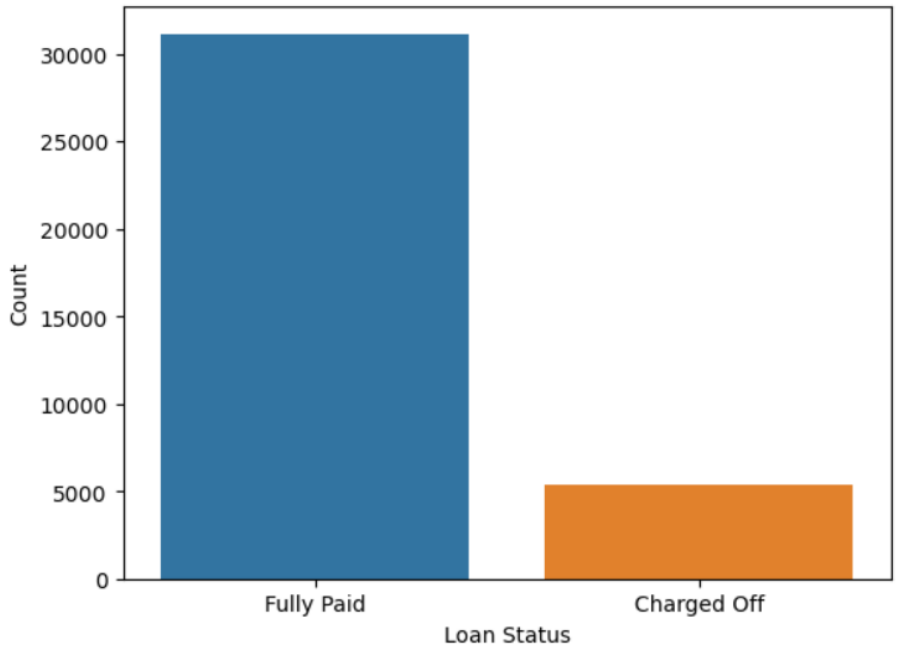
- Loan Accepted: if company approves the loan, there are 3 possible scenarios described below:
  - Fully paid: Applicant has fully paid the loan (the principal and the interest rate)
  - Current: Applicant is in the process of paying the instalments, i.e., the tenure of the loan is not yet completed. 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
- Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

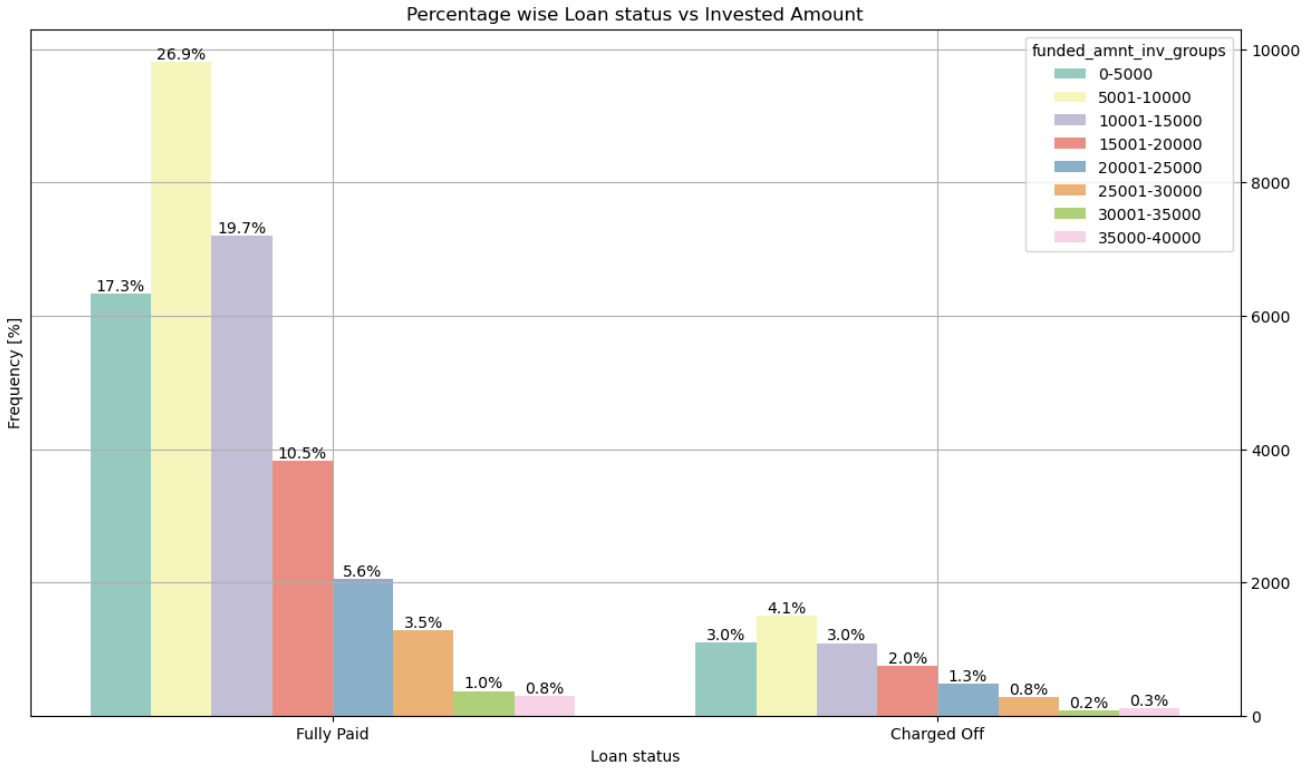
Problem Solving Methodology

- How consumer attributes and loan attributes influence the tendency of defaulter
- Using EDA for this analysis
  - Data Cleaning
  - Data Understanding
  - Univariate Analysis
  - Segmented Analysis
  - Bivariate Analysis

Analysis

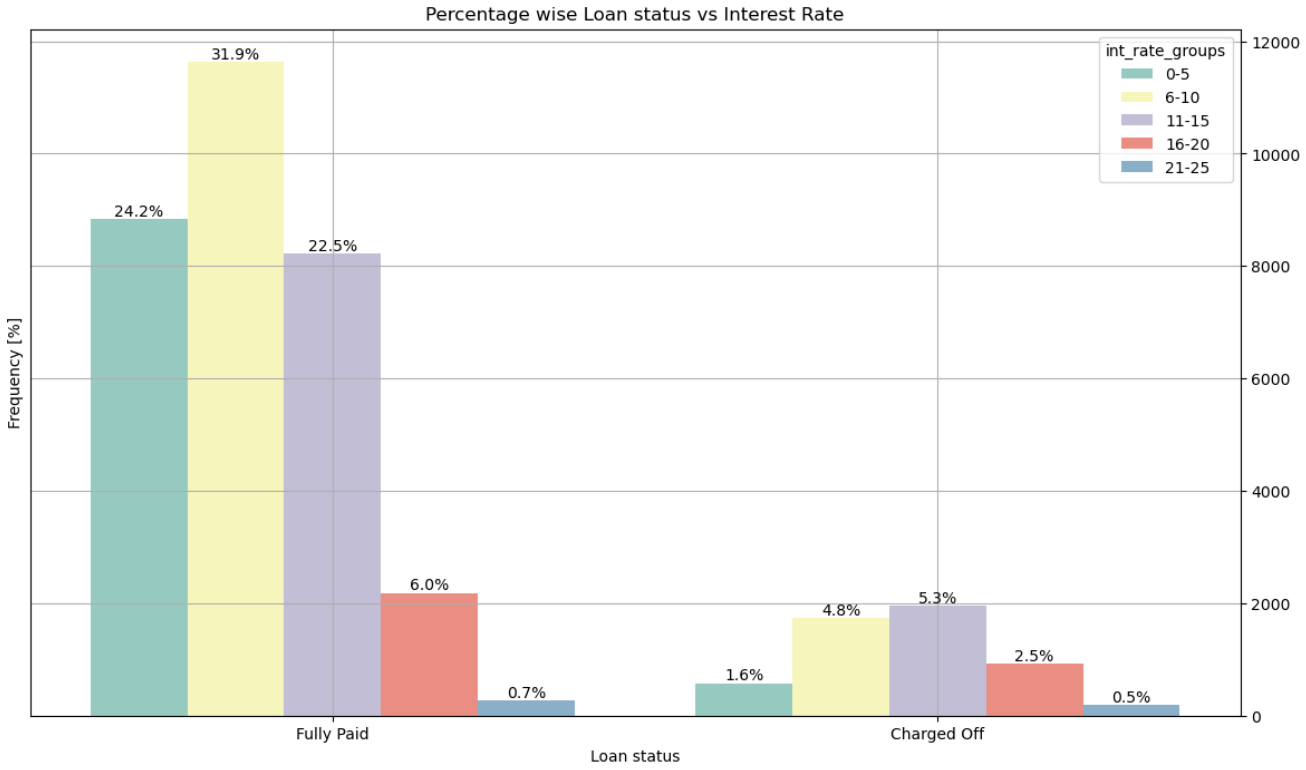
Split of Fully Paid vs Charged Off





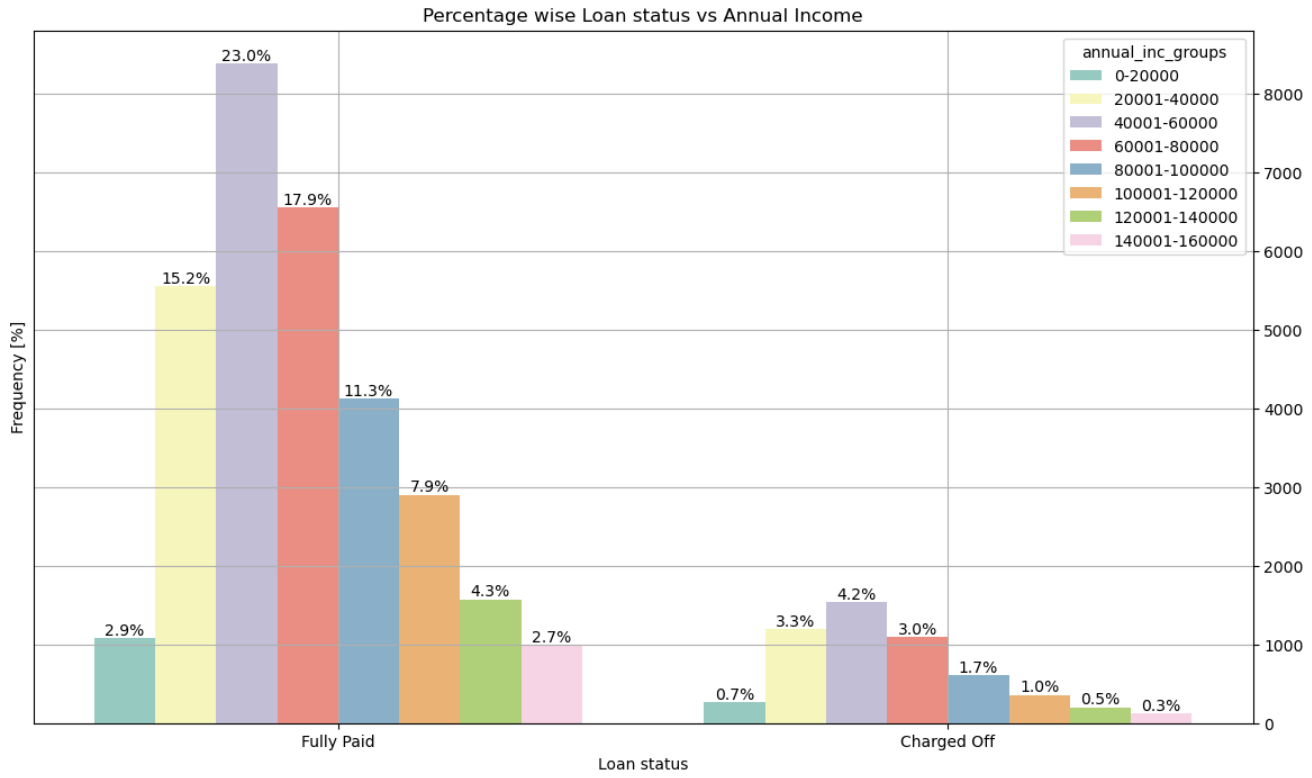
Loan Status vs Invested Amount

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



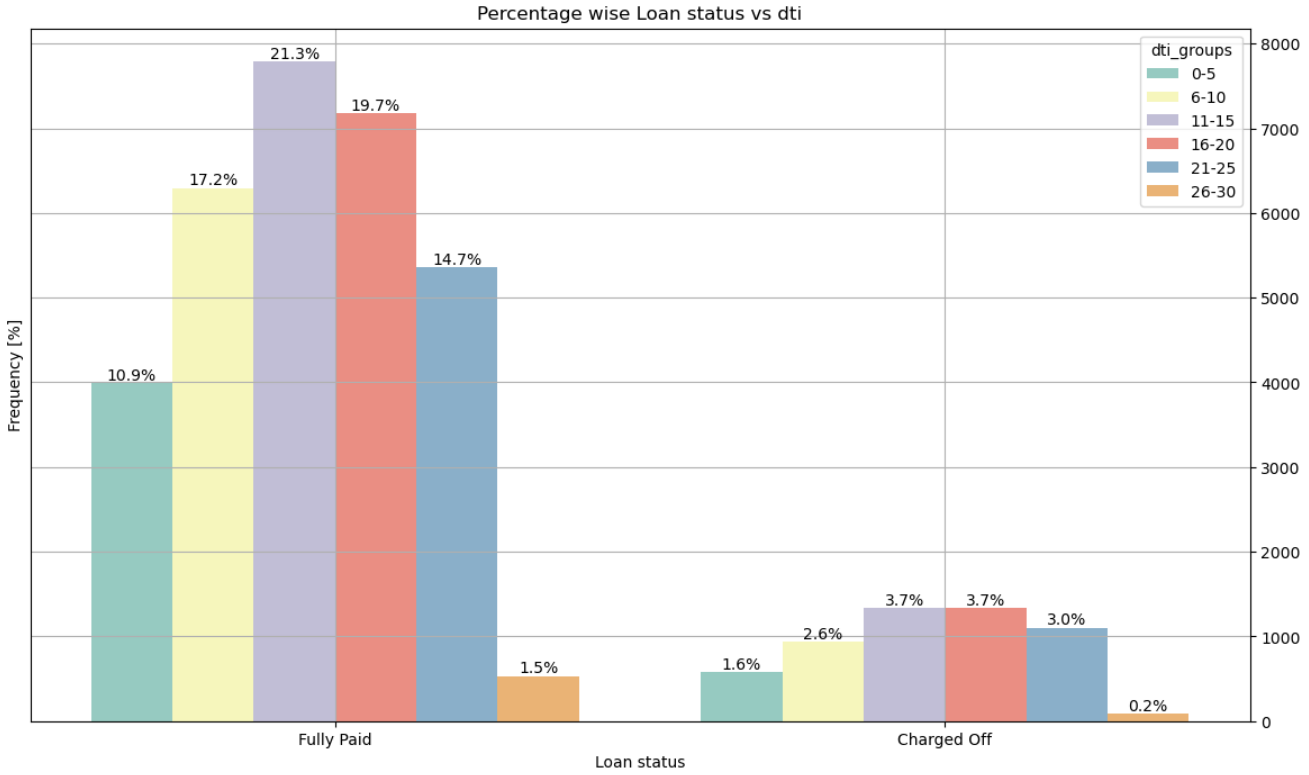
Loan Status vs Interest Rate

- I see annual income is positively correlated with both Charged off and Fully Paid
- I 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



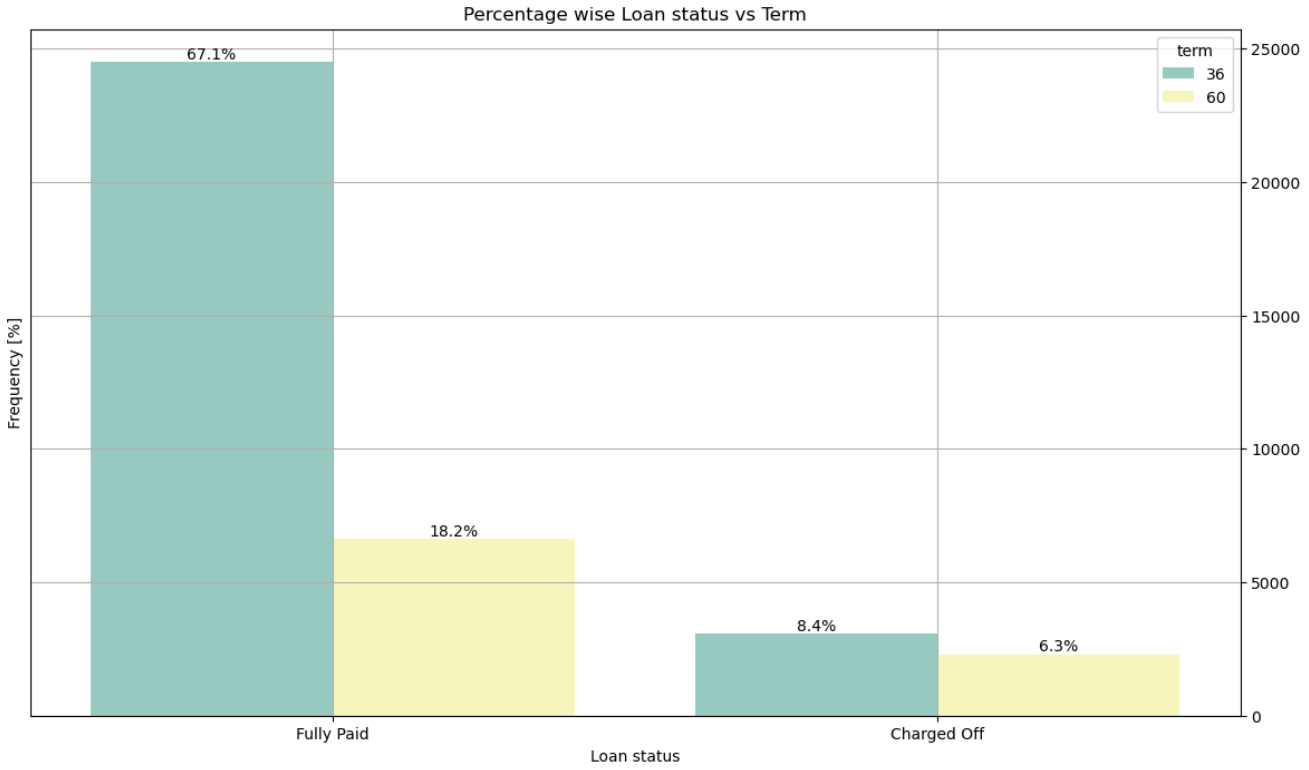
Loan Status vs Annual Income

- I see annual income is positively correlated with both Charged off and Fully Paid
- I 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 0-20000, 0.7/2.9 nearly = 24.13 has higher chances of becoming defaulter



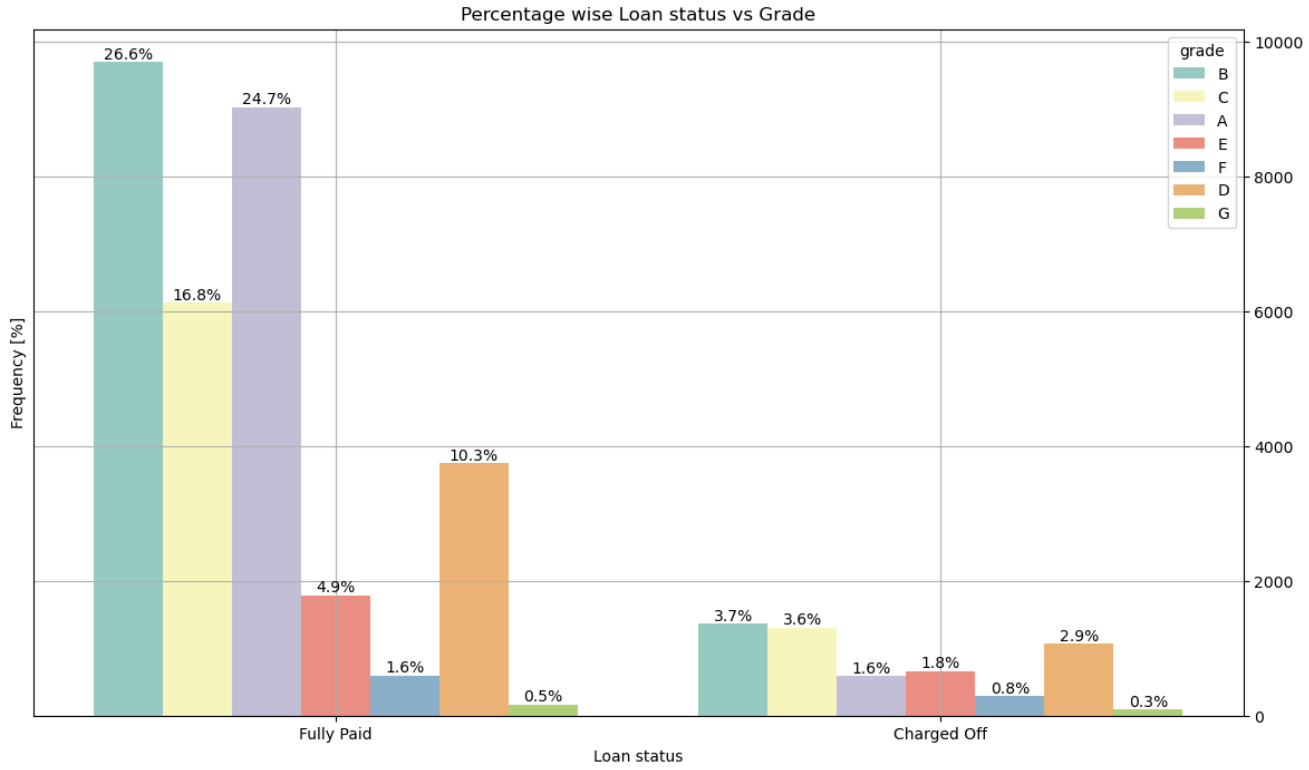
Loan Status vs DTI

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



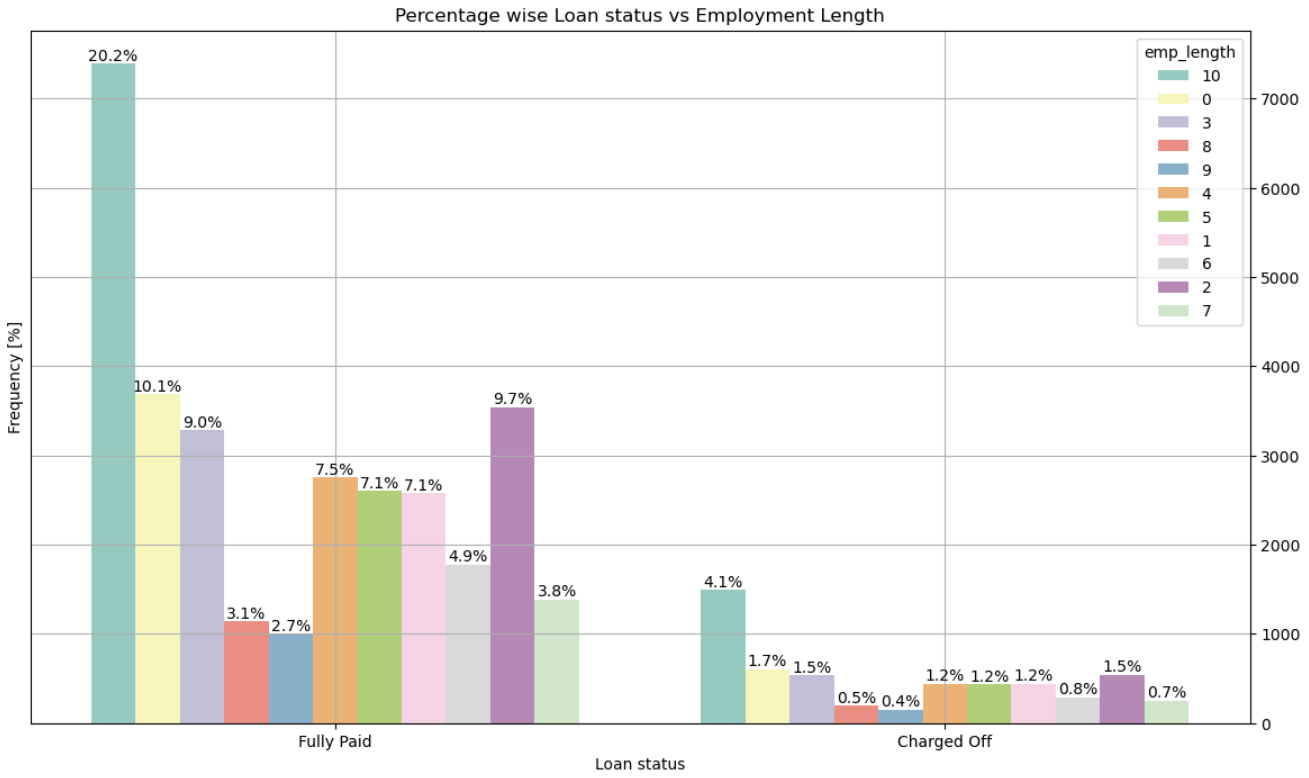
Loan Status vs Term

- I see term is positively correlated with both Charged off and Fully Paid
- I also see if term is 36 months, then chances of becoming defaulter is higher
- But ratio wise 60 months has much higher number



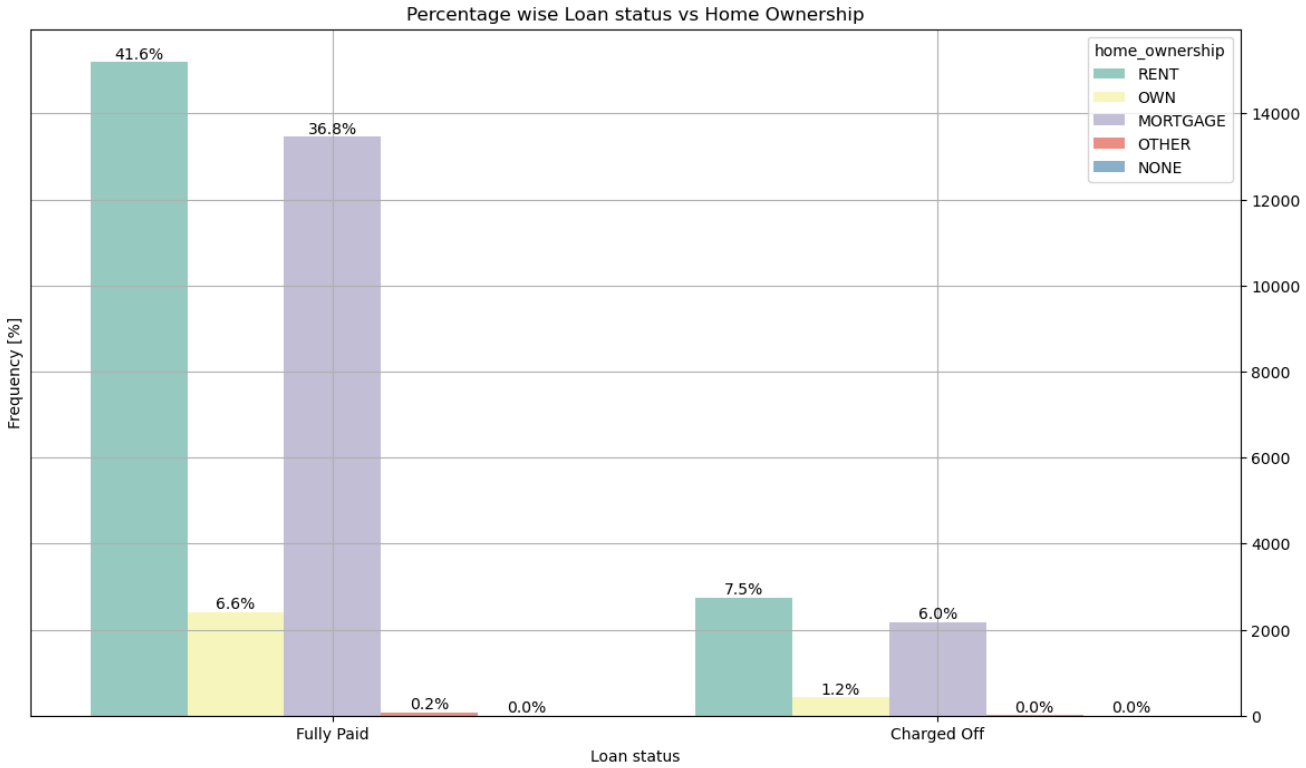
Loan Status vs Grade

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



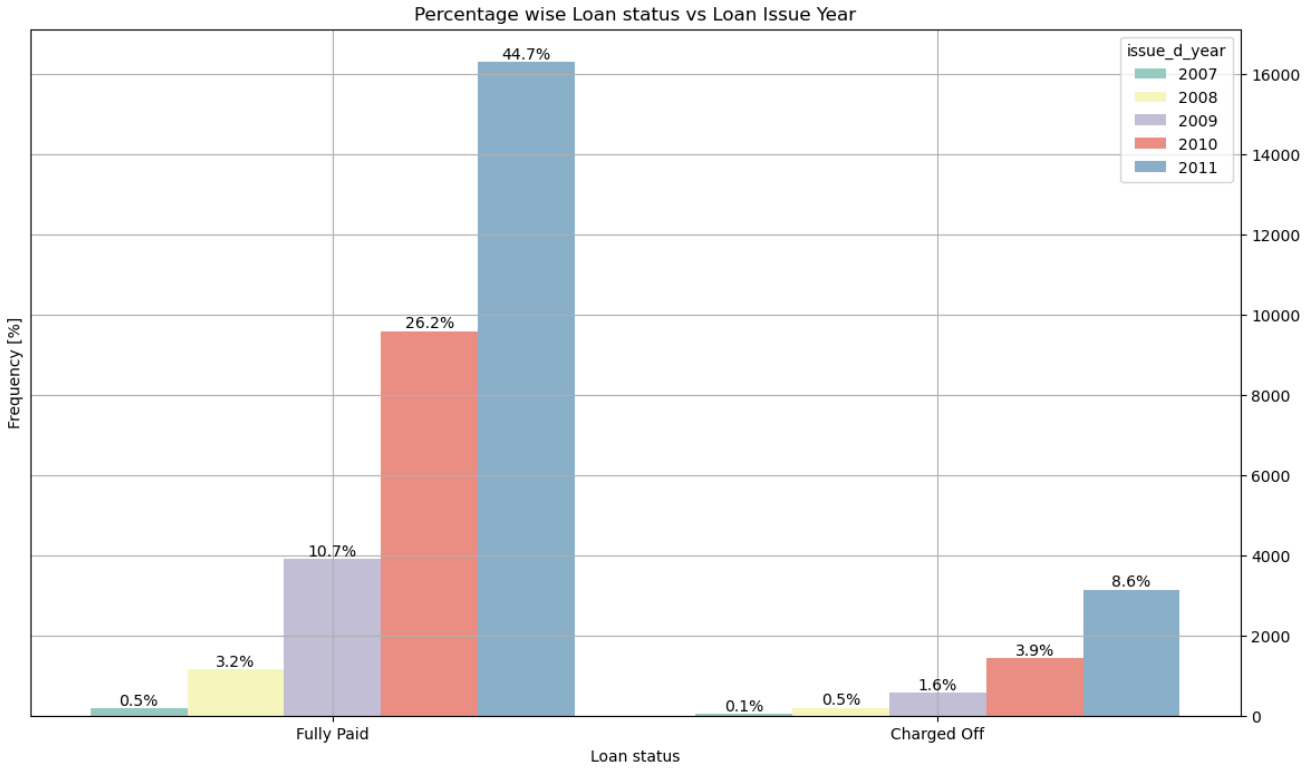
Loan Status vs Employment Length

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



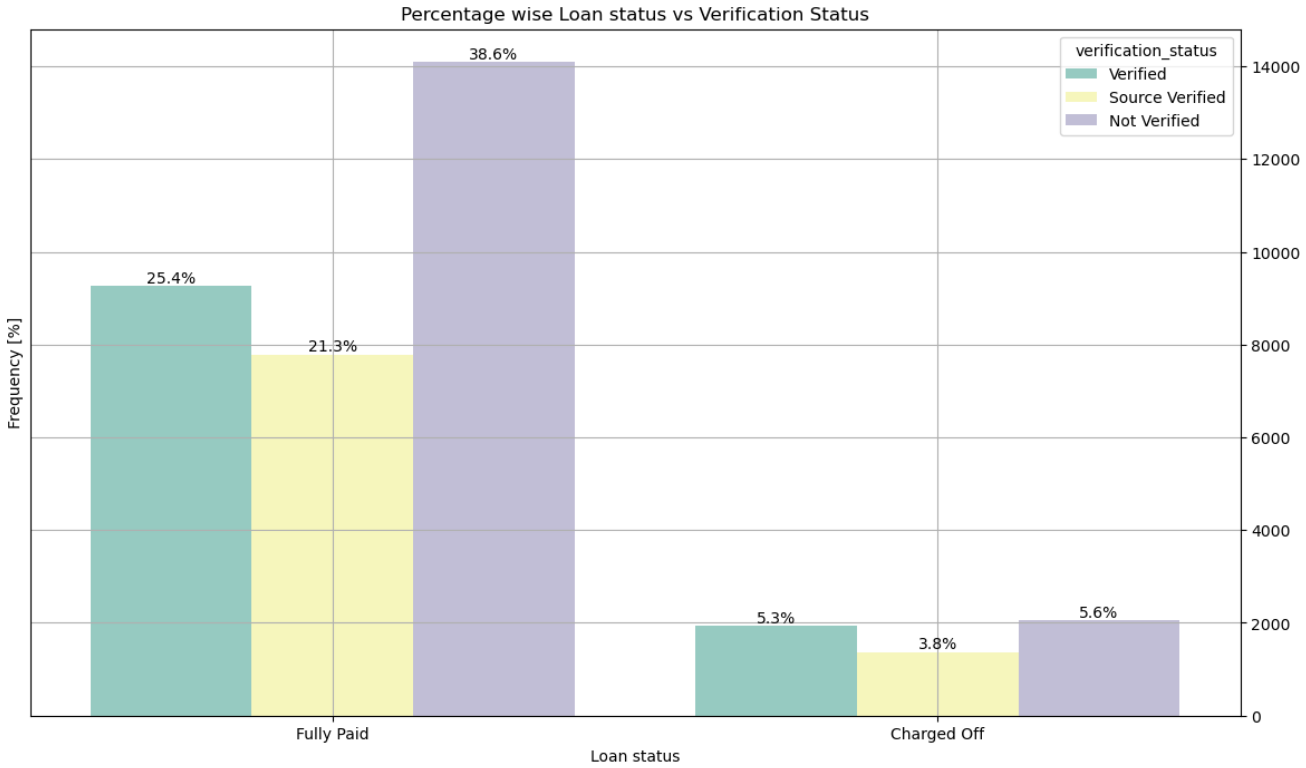
Loan Status vs Home Ownership

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



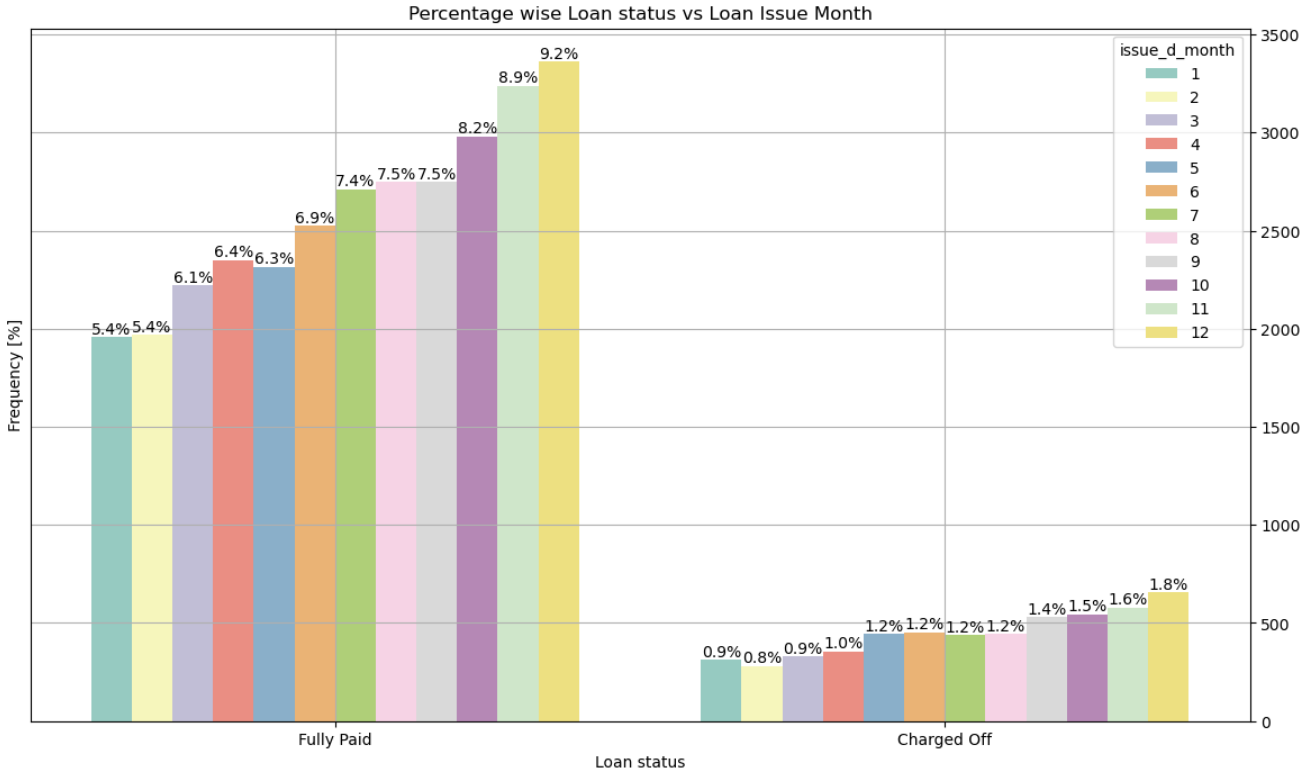
Loan Status vs Loan Issue Year

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



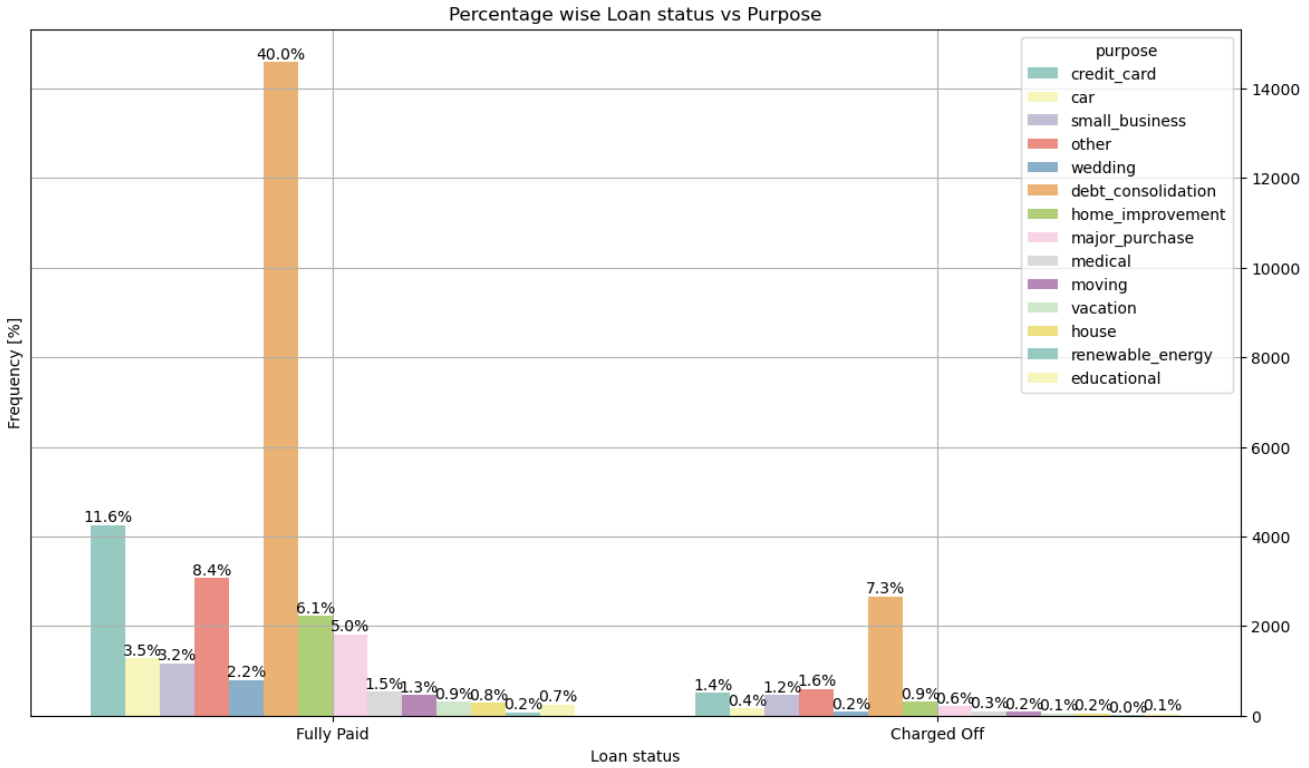
Loan Status vs Verification Status

- I see verification status is positively correlated with both Charged off and Fully Paid
- I 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



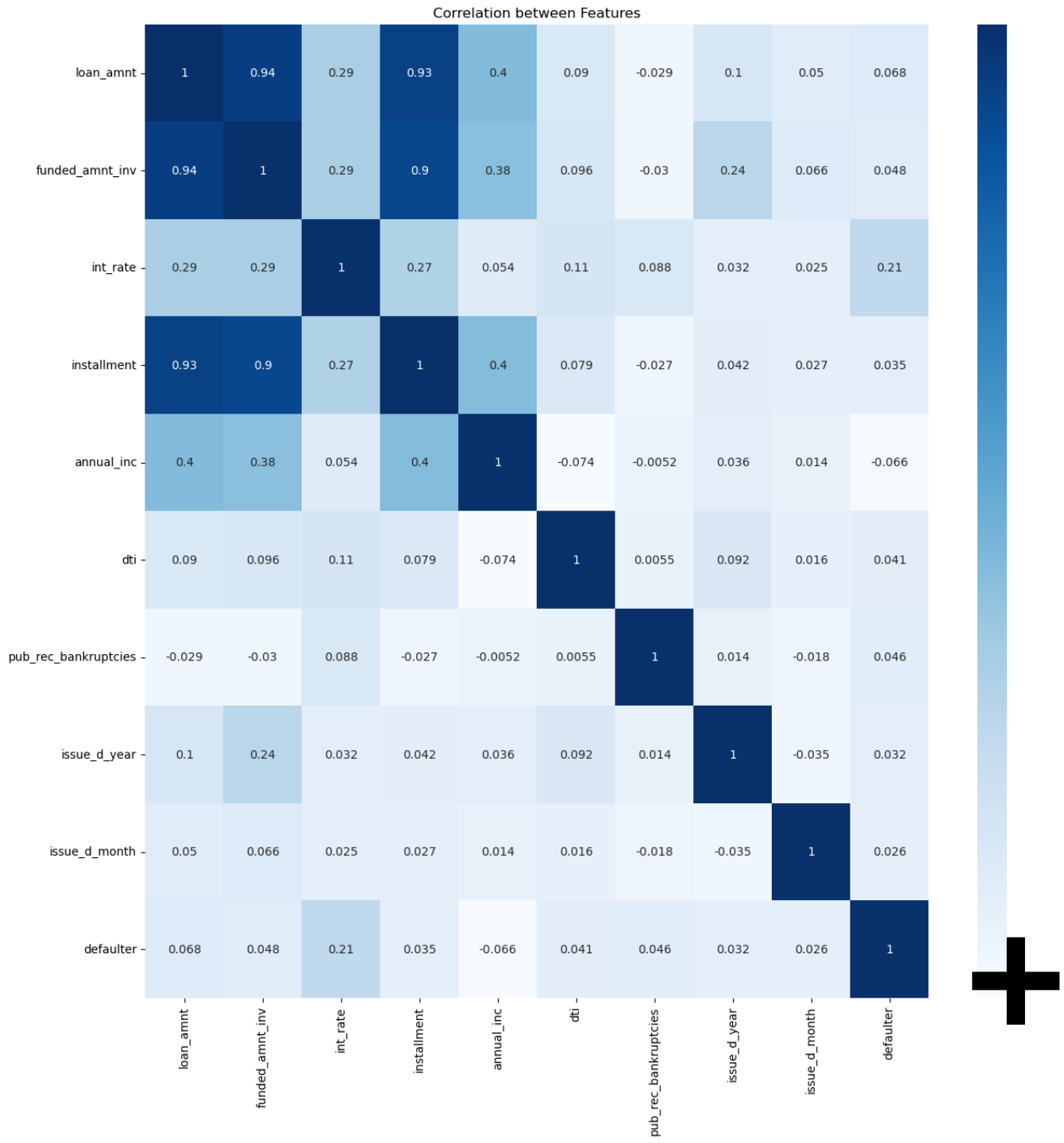
Loan Status vs Loan Issue Month

- I see issue month is positively correlated with both Charged off and Fully Paid
- # I also see if issue month is either Dec or Nov then chances of becoming defaulter is higher

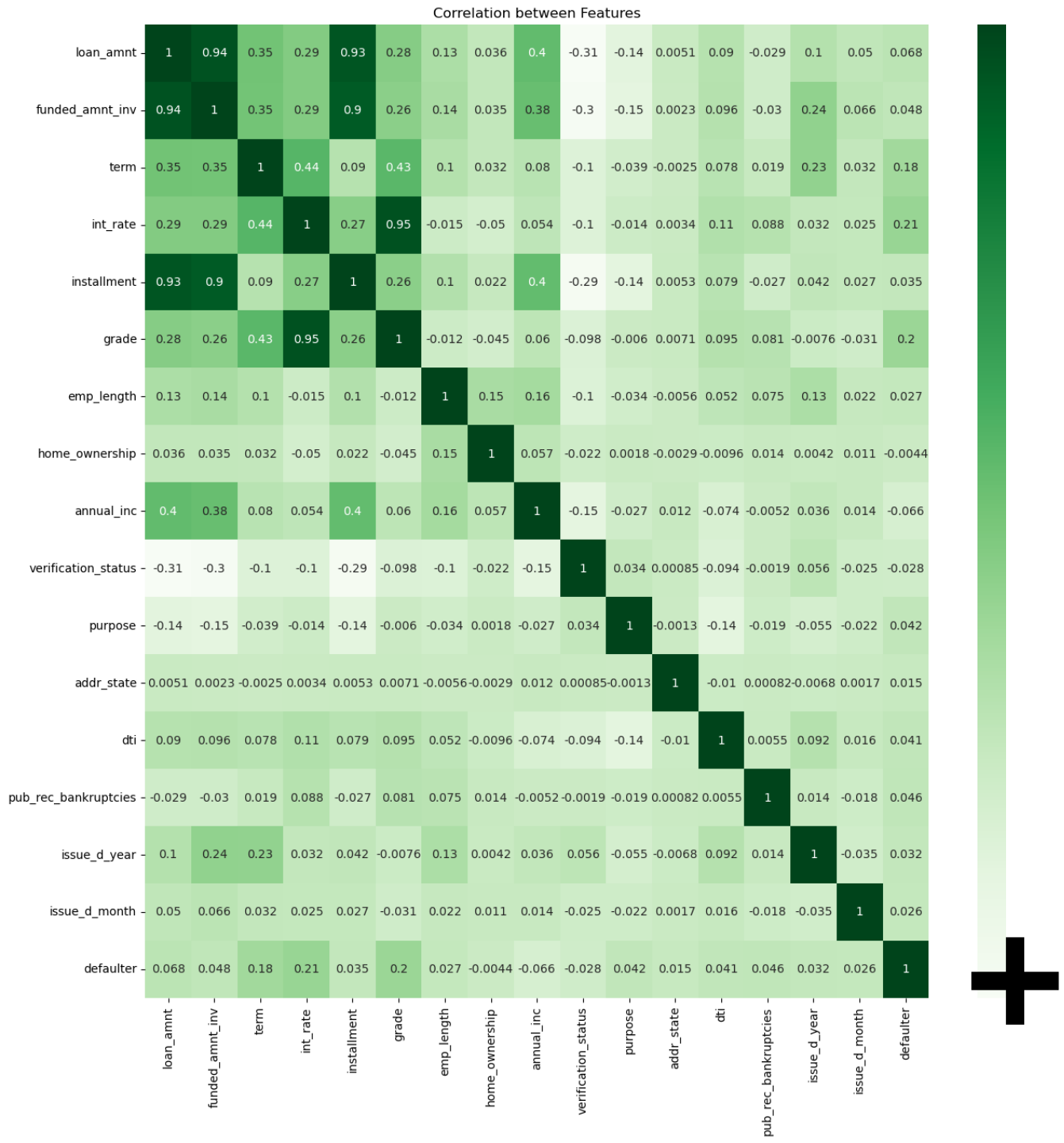


Loan Status vs Purpose

- I see purpose is positively correlated with both Charged off and Fully Paid
- I also see if purpose is debt consolidation then chances of becoming defaulter is higher



Correlation



Correlation between Features

Conclusion

Based on heatmap it is observed that home_ownership, annual_inc and verification_status with defaulter are negatively correlated,

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, 0.7/2.9 nearly = 24.13 has higher chances of becoming defaulter. Heatmap is also confirming this

Term, int_rate, grade are positively correlated,

A. So higher the term greater will be chances of becoming defaulter - 60 months

B. Higher the int_rate means higher the chances of becoming defaulter - 21-25%

C. Higher the grade means higher the chances of becoming defaulter - G