# LENDING CLUB CASE STUDY

Facilitator: Veena Ashok

Team Member : Namrata Thakur

## Business Objective:

Company which provides online loan marketplace wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default to minimize credit loss.

Borrowers who default cause the largest amount of loss to the lenders. By identifying these risky loan applicants, such loans can be reduced thereby cutting down the amount of credit loss.

# Problem solving methodology

**Data Sourcing and Cleaning**

- Standardise values in relevant columns of available datasets
- Identify and treat missing values, if necessary
- Filter out duplicates in the dataset

**Understand the Dataset**

- Identify the target columns
- Identify columns which are not relevant to the case study and drop them
- Standardize the column content

**Data Filtering**

- Filter out dataset to contain records of only 'Fully Paid' and 'charged_off' loans.
- Bin continuous variables for better analysis

**Analysis**

- Perform univariate and segmented univariate to understand each column
- Perform Bivariate to understand the relation between 2 variables and how it drives default of loans
- Perform multivariate analysis if needed

**Conclusions**

- Analyse the results and draw conclusions

# Dataset and Target Parameter Details

**Dataset details:** Dataset contains 39717 rows and 111 columns

**Dataset contains:**

- geographic details,
- loan characteristics
- customer behaviour variables

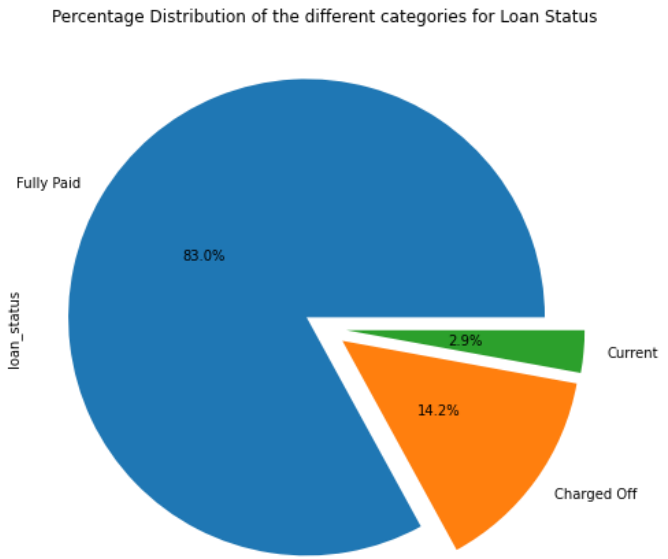
**Target Variable:** loan_status- it indicates the status of loan application.

Values can be:

- Fully Paid
- Charged_off
- Current

**Columns dropped :**

- Columns with 100% null values : 54 columns
- Columns with rate of missing value > 60%: 3 columns
- Columns which are not relevant for the objective : 38 columns




- We see that 83% of loans that were approved were fully paid off.
- 14.2% of total loans are defaulted.

We are trying to find pointers/variables to detect defaulters.

# Data Cleaning

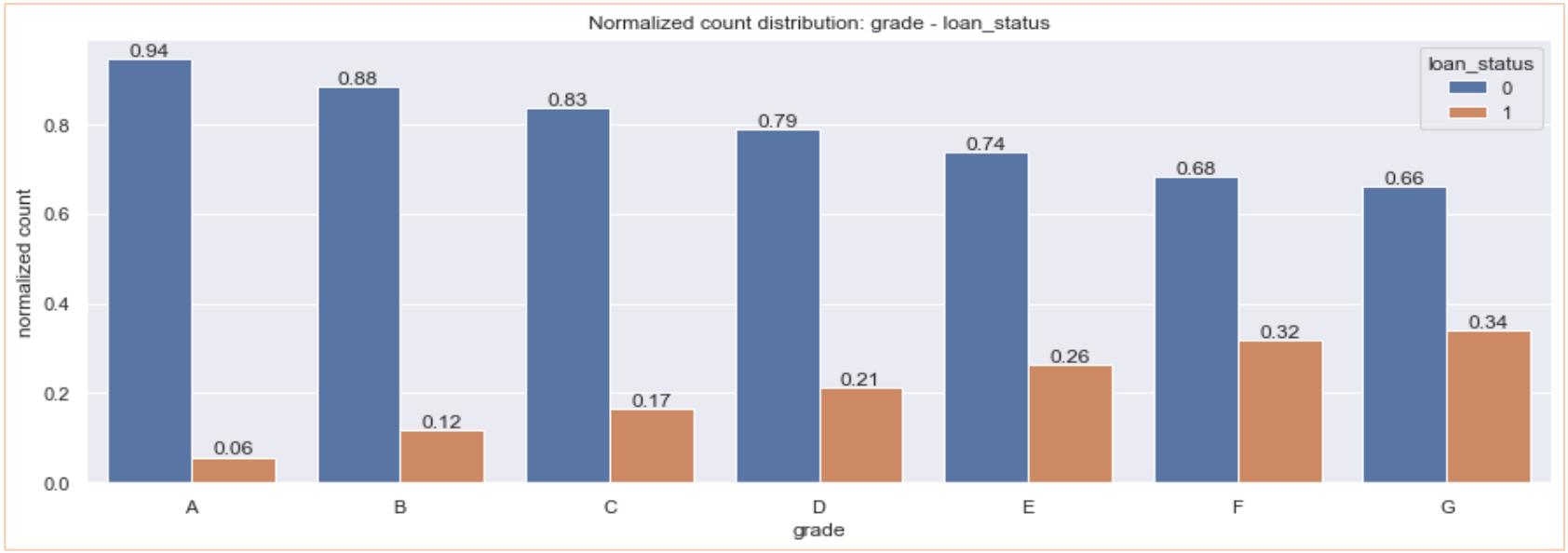
**Assumptions:**

- Customer behaviour variables are not available at the time of loan application. Hence, they can be removed from dataset.
- Loan records having loan status as ‘current’ can de dropped as they do not contribute to identifying driving factors of loan default.
- Assume ‘10+years’ of employment length as 10 and ‘<1 year’ as 0.5 year.
- Standardize column content loan_status:
  - 0 represents fully paid loans
  - 1 represents charged off loans

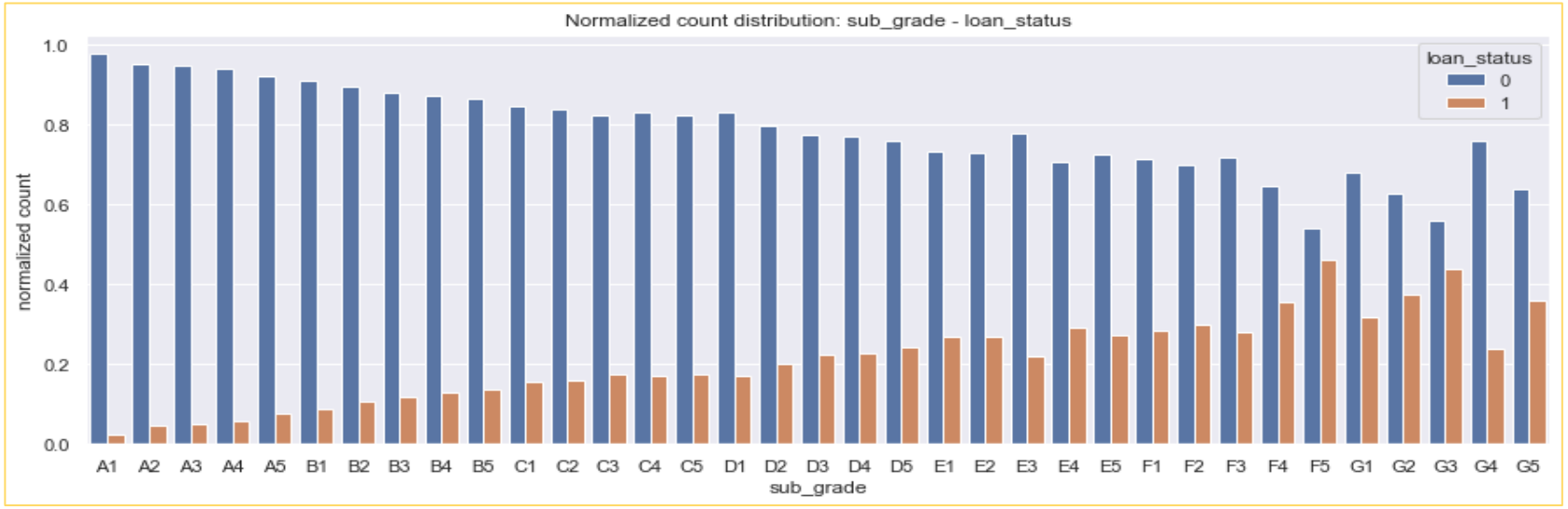
**Important columns for the Analysis:**

loan_amnt, term, int_rate, grade, sub_grade, emp_length, home_ownership, annual_inc, verification_status, issue_d, loan_status, purpose, addr_state, dti, pub_rec_bankruptcies

# Univariate Analysis

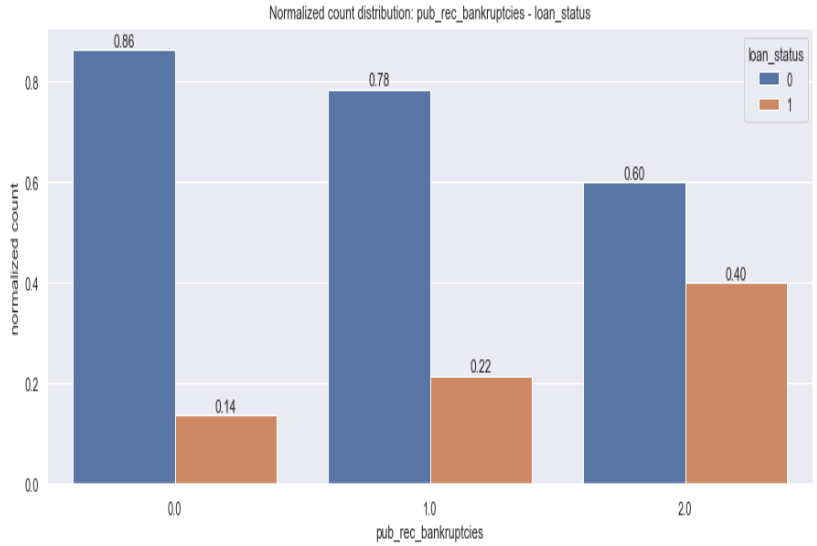



Chances of charged_off cases increase as the applicants fall in the lower grades

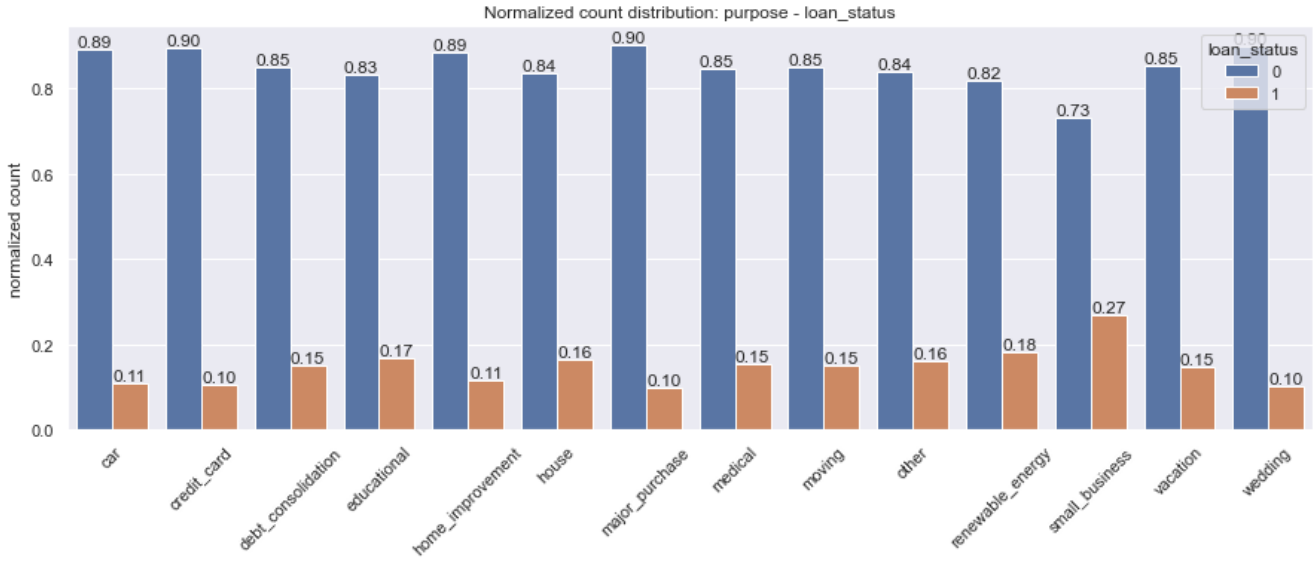


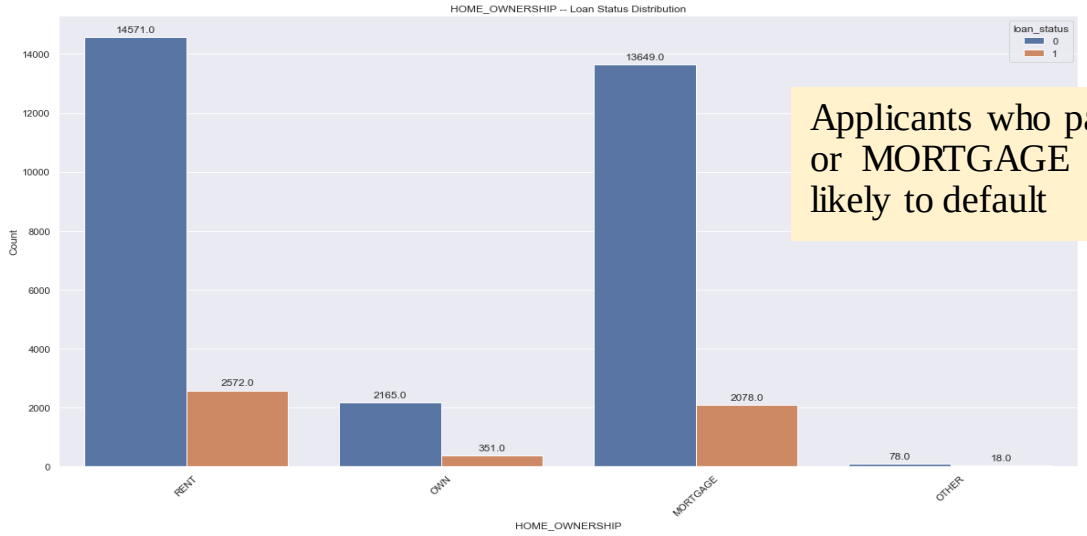

A general trend can be observed within the subgrade of applicants and charged off loans. Within a grade, the chances of charge off are high if the applicant belongs to lower sub grade

Applicants with more records of bankruptcies are more likely to default and result in charge off

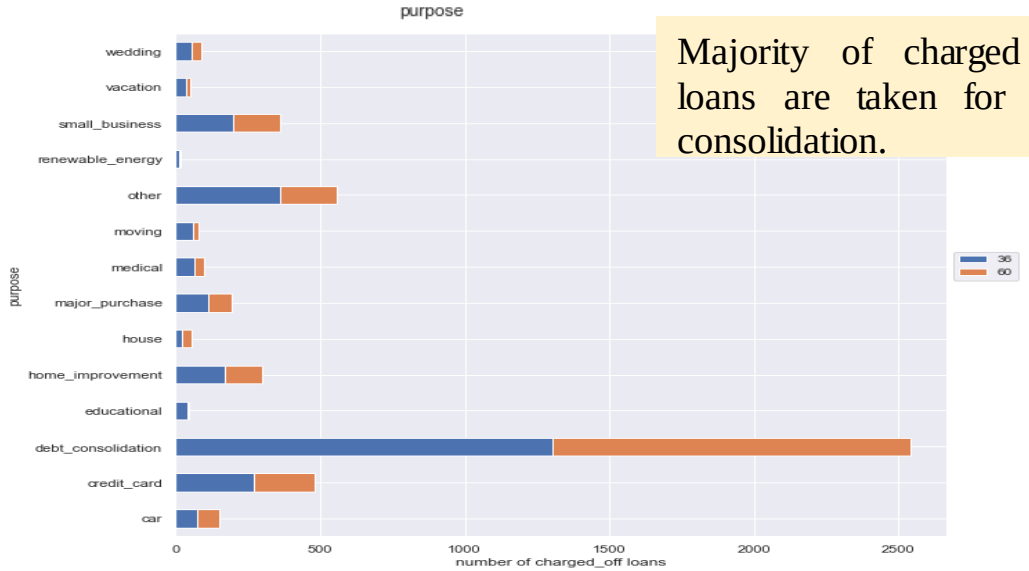



Applicants who have taken loans for small business are more likely to default and result in loan charge off



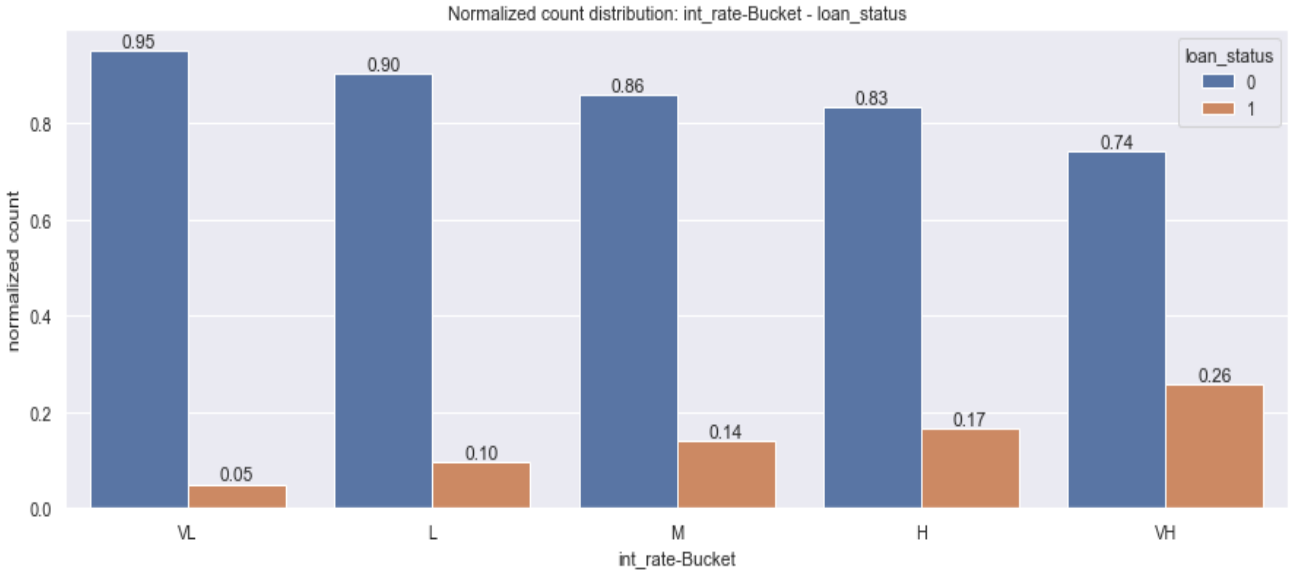




Applicants who pay RENT or MORTGAGE are more likely to default

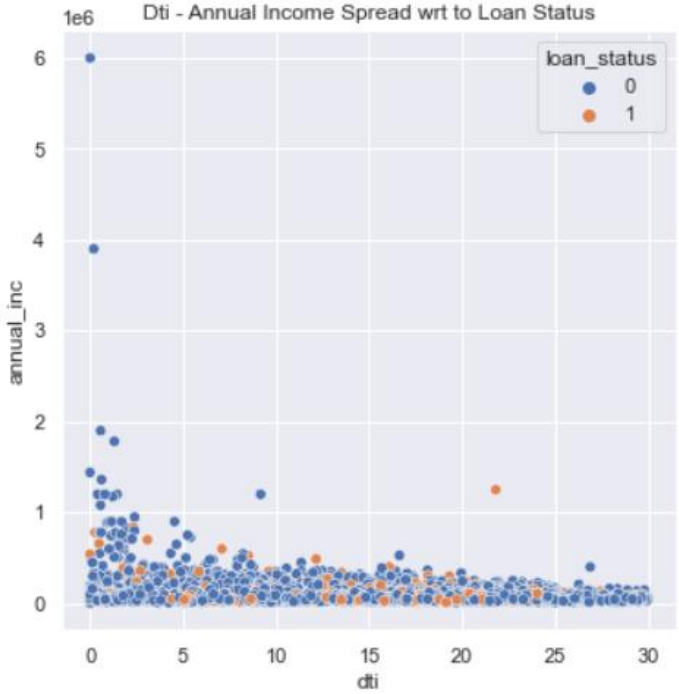



Majority of charged off loans are taken for debt consolidation.

Chances of charged-off cases increases as interest range goes from very low to very high.




As annual income increases debt to income ratio decreases



## Plot charged off cases against month and year of loan issue:

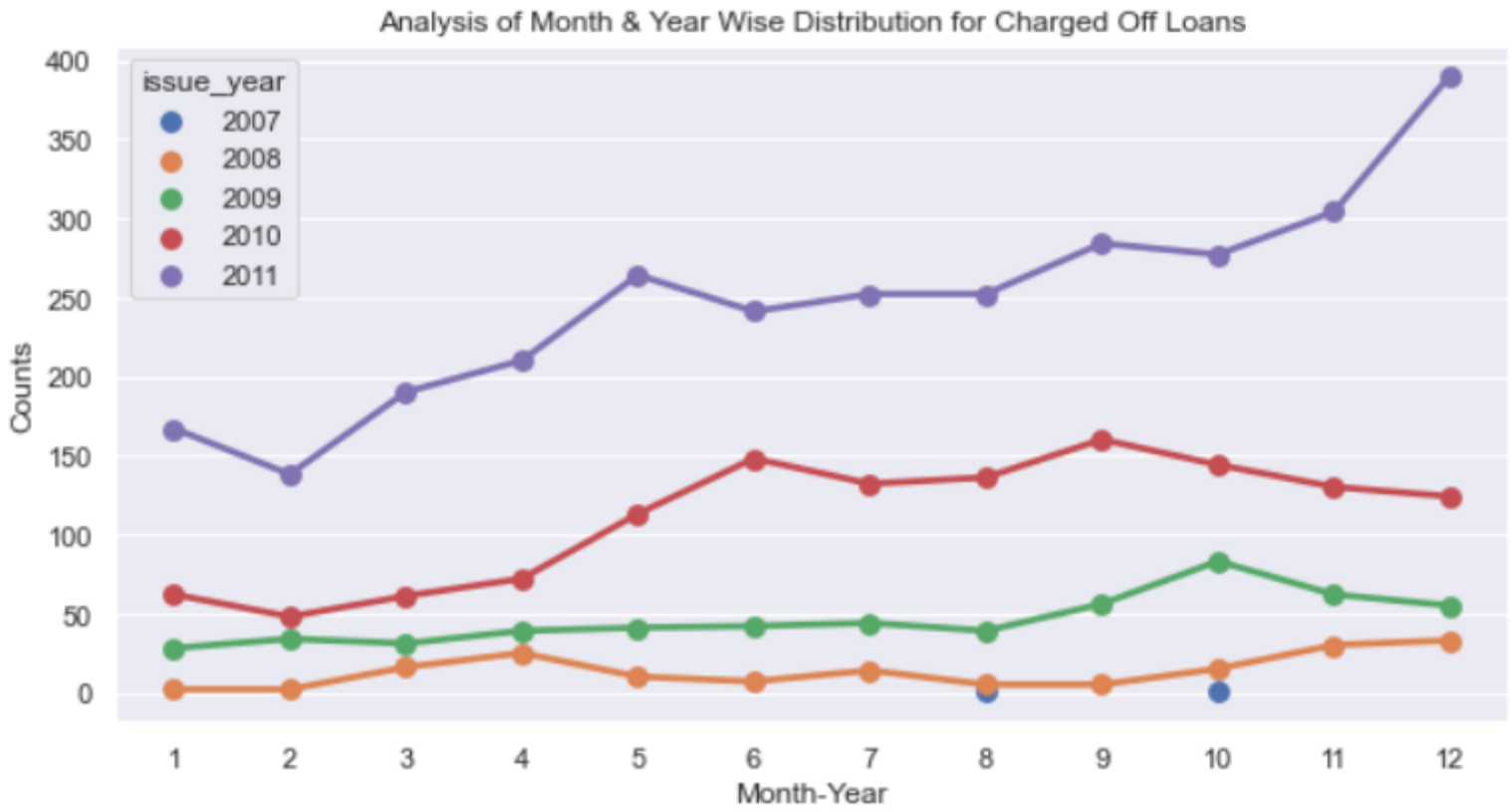
**Observations**:

- We saw a major hike in count of defaulters for the month of 'May (05)' and 'Dec (12)' in the year 2011.

- For the year 2010, the hike was at 'June (06)'.

- For the year 2009, the hike was at 'Oct (10)'.

So we can conclude that there is no obvious trend in the month of defaults but May, June, Oct and Dec seems to be somewhat risky.



# Relation among quantitative variables

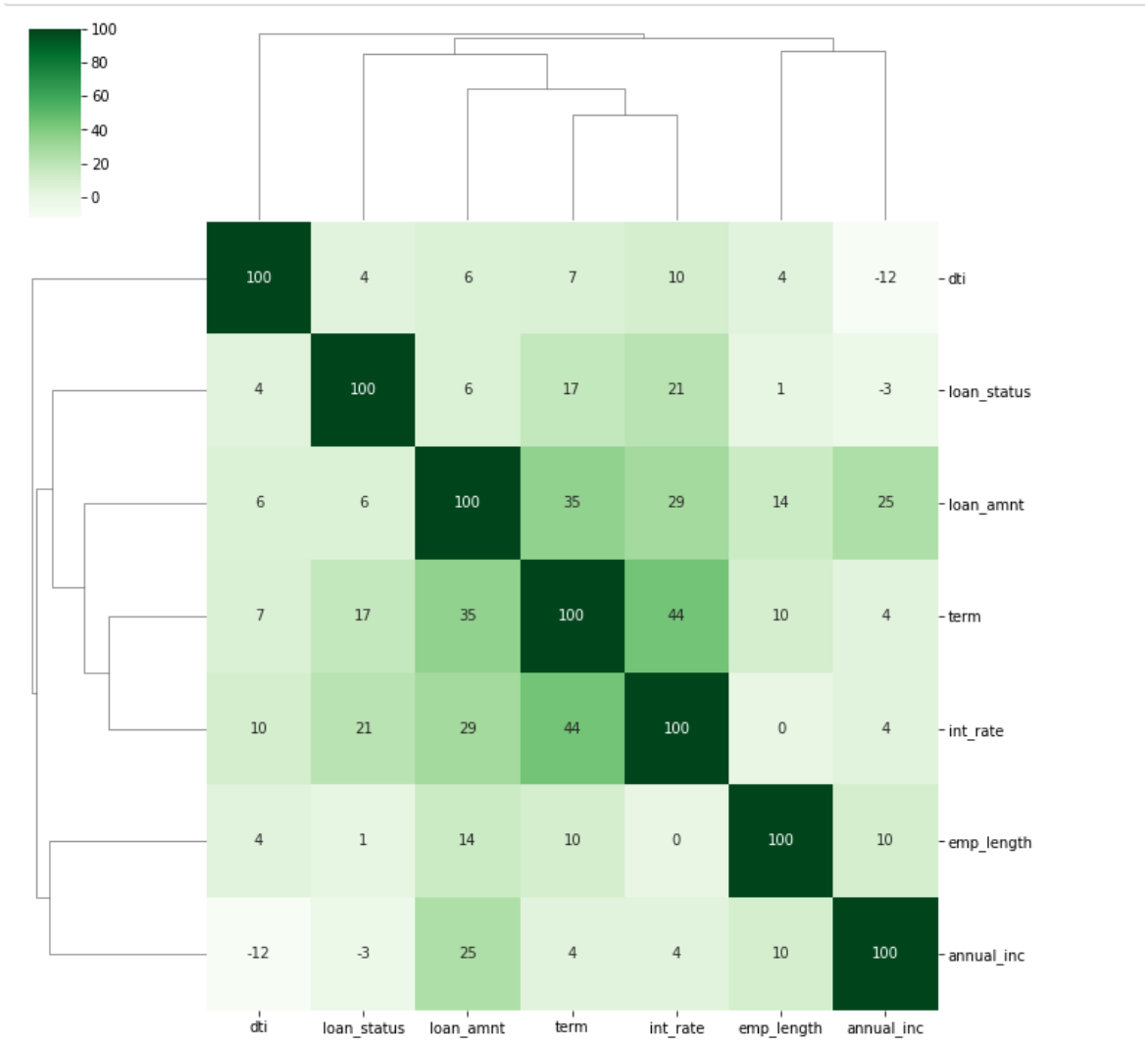
**Observations**:

We see that the cluster 'loan_amnt', 'term', 'int_rate' are quite correlated with one another.

'Annual Income' and 'DTI' are negatively correlated meaning the rise of one decreases the other.

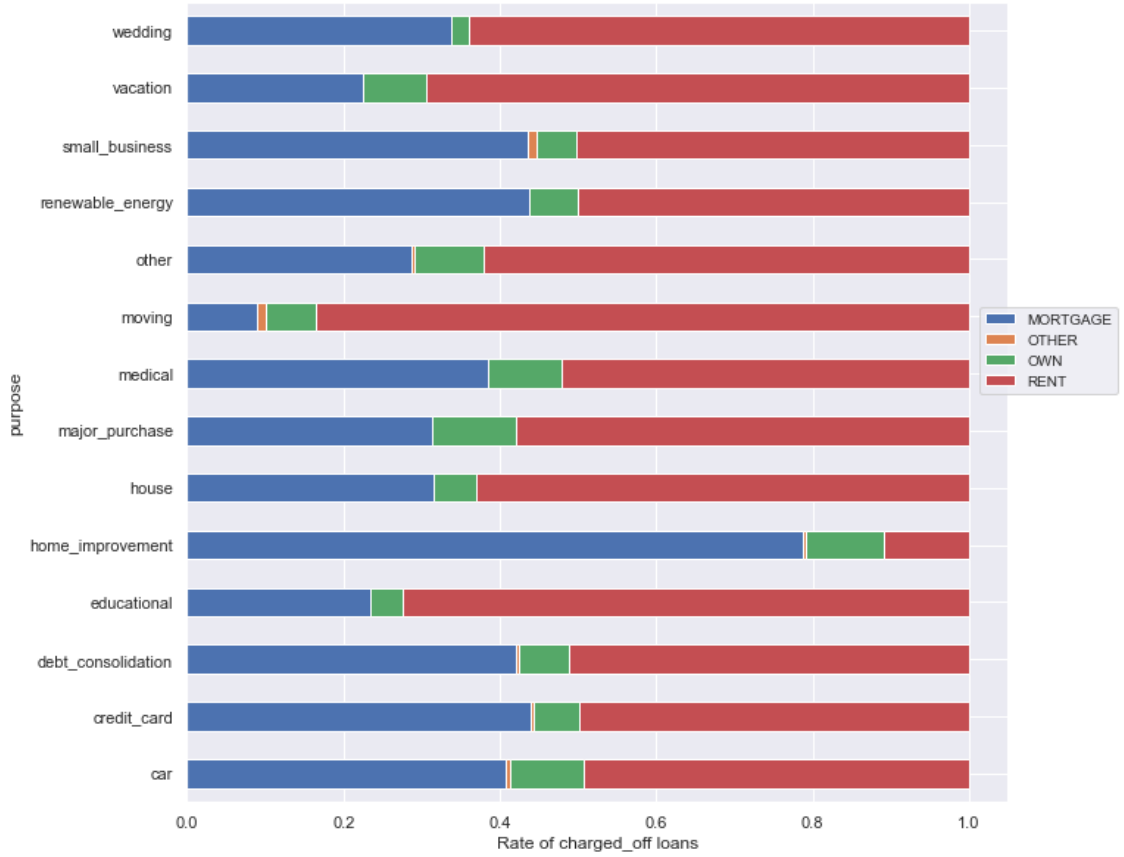
'loan_status' is more positvely correlated with 'term' and 'int_rate' than with others.

'loan_status' is negatively correlated with 'annual_inc'. That means the rise of one decreases the other.

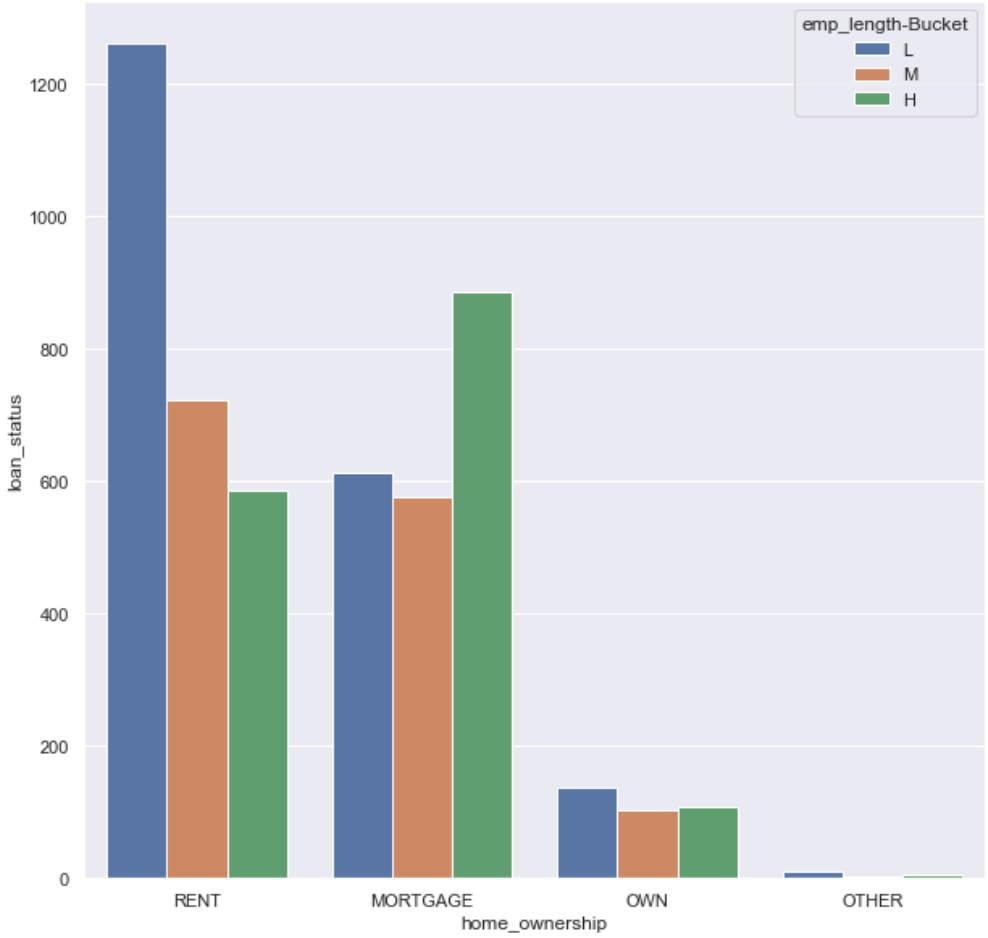


# Bivariate Analysis

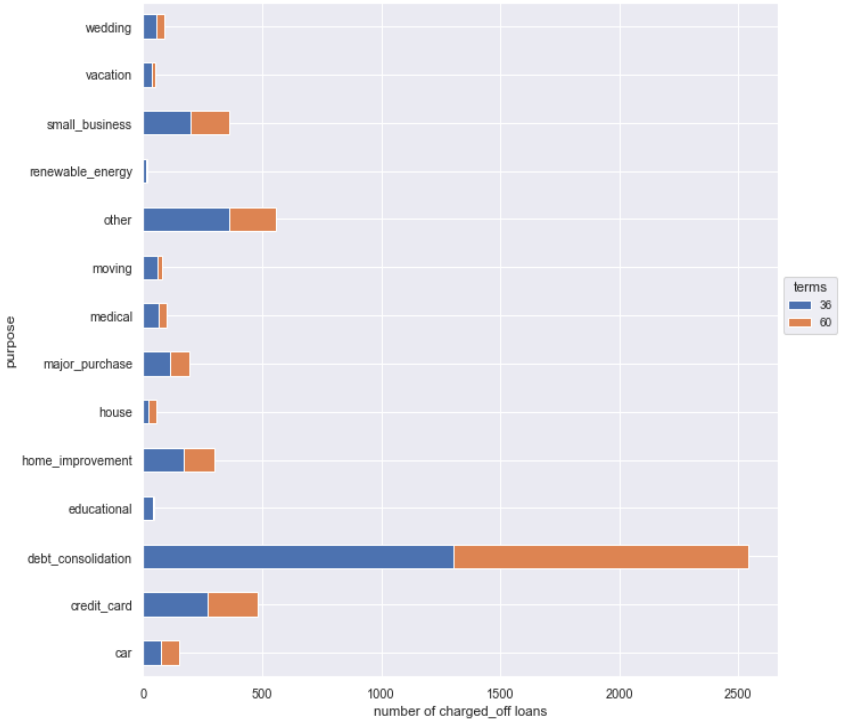
Applicants who pay RENT or MORTGAGE are more likely to default across different loan purposes as well



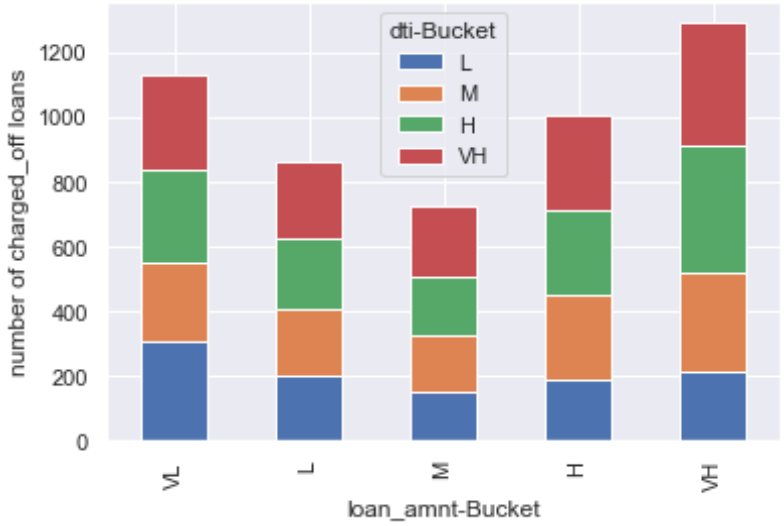

Those who pay rent and has less employment are very much likely to default loan



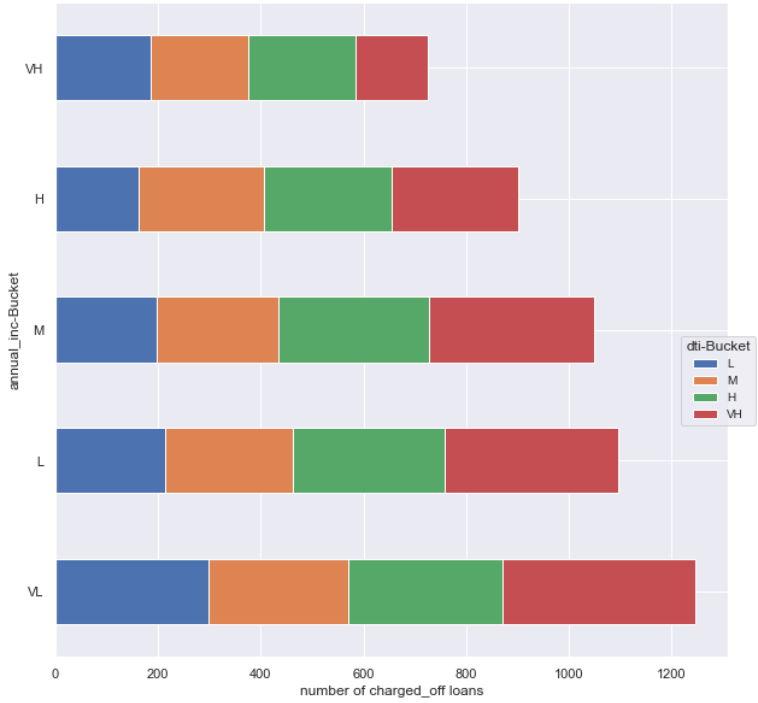
Majority of charged_off loans are taken for debt consolidation




As loan amount increases to medium, dti decreases, as well as charged off cases. Whereas, beyond the medium bucket, dti and charged off are gradually increasing



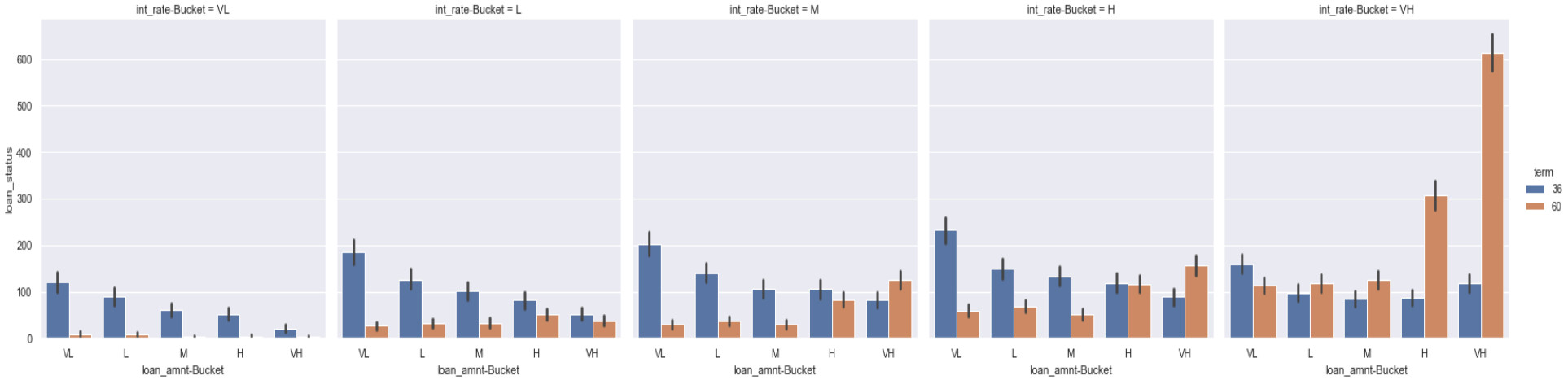

As annual income increases, less number of loans end up charged off. Also dti reduces with increase in income.



# Multivariant Analysis

Applicants with higher employment duration become defaulters, when they have mortgage to pay off.

Applicants with high experience while paying off mortgage took huge amount of loan resulting in loan default.


If an applicant has applied for a loan amount greater than medium and the associated interest rate is very high, then it is very much likely that the person will become a defaulter if the chosen payment plan is of 60 months.

# Analysis Results :

**Conclusions / Recommendations:**

1. For Loan term as 60, we see a higher chance of default with ~ 25% of loans defaulting.
2. As the Grade/Subgrade of the loan changes from 'A' to 'G' the chance of default also increases.
3. Grade E , F , G seems to be shows more defaulters than any other Grades.
4. Loans for which the income is Verified have marginally more defaults.
5. Maximum number of defaults happen in “small businesses” ~ 27% of loans are defaulted.
6. Higher loan amounts with higher term and high interest rates are more likely to default.
7. As interest rate increases, the chances of default also increases with ~26% of loans are defaulting when interest rate is very high.
8. As bankruptcies record increases from 0.0 to 1.0 , the chances of default also increases with ~ 22% of loans are defaulting having bankruptcies record as 1.0.
9. For High and Very High values of DTI, the defaulting chances increases.
10. For low annual income, we see a marginally more percentage of defaulters with ~18% of loans with very low annual income defaulting.

**Recommended variables to detect defaulters are**:

1.Interest Rate

2.Grade / Sub Grade

3.Purpose

4.Term

5.DTI

6.Loan Amount

7.Public bankruptcy records

8.Annual Income

9.Verification Status