GRAMENER

CASE STUDY SUBMISSION

1. Sarthak Dey
2. Ayon Sarkar
3. Abhishek Sinha
4. Prasenjit Deb

OBJECTIVE

The objective is to find the variables that correlate to the risk associated when approving loans to customers.

Risks are of two types:

- If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
- If the applicant is **not likely to repay the loan,** i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

We are finding the driving factors behind loan default of the customers.

DATA CLEANSING

- In this dataset 54 variables contain all the observations as NA so they are removed.
- Some columns are replaced with 0 in the place of NA.
- 4 columns from remaining has same data in column hence they were removed.
- Rounded of some numeric variables.
- Removed "months" from terms and "%" from some variables for visualization
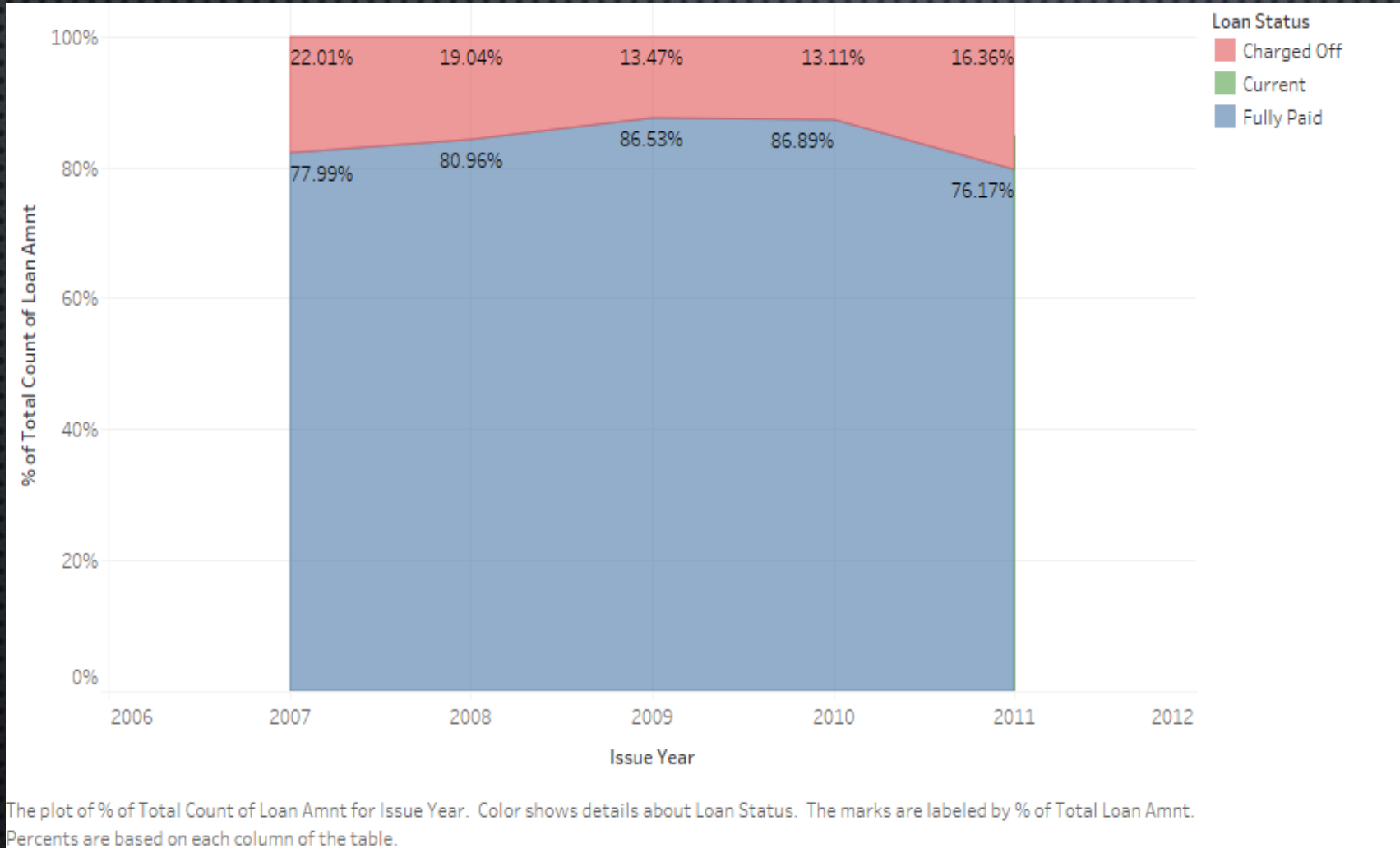
AMOUNT OF LOANS PER YEAR
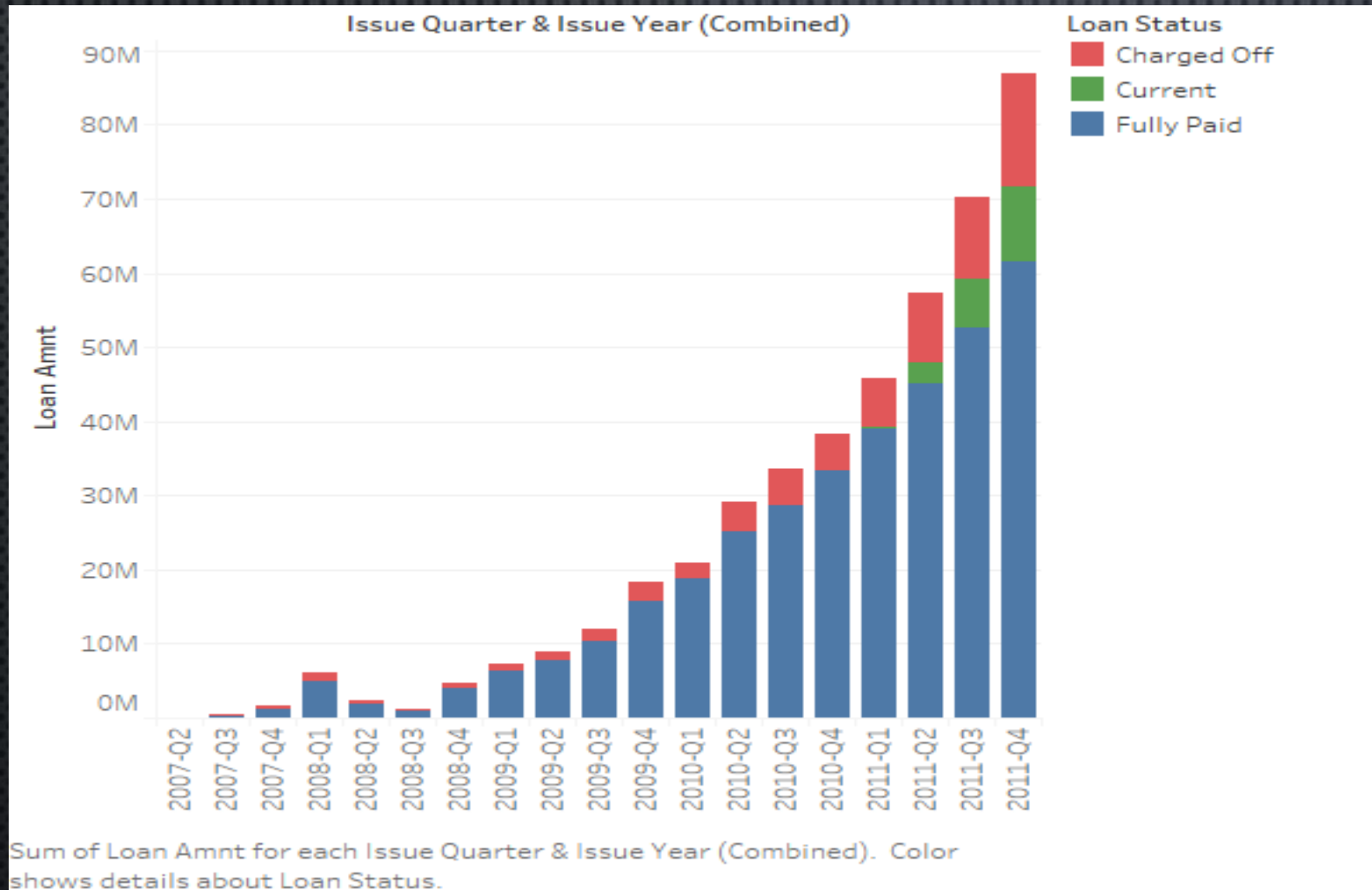
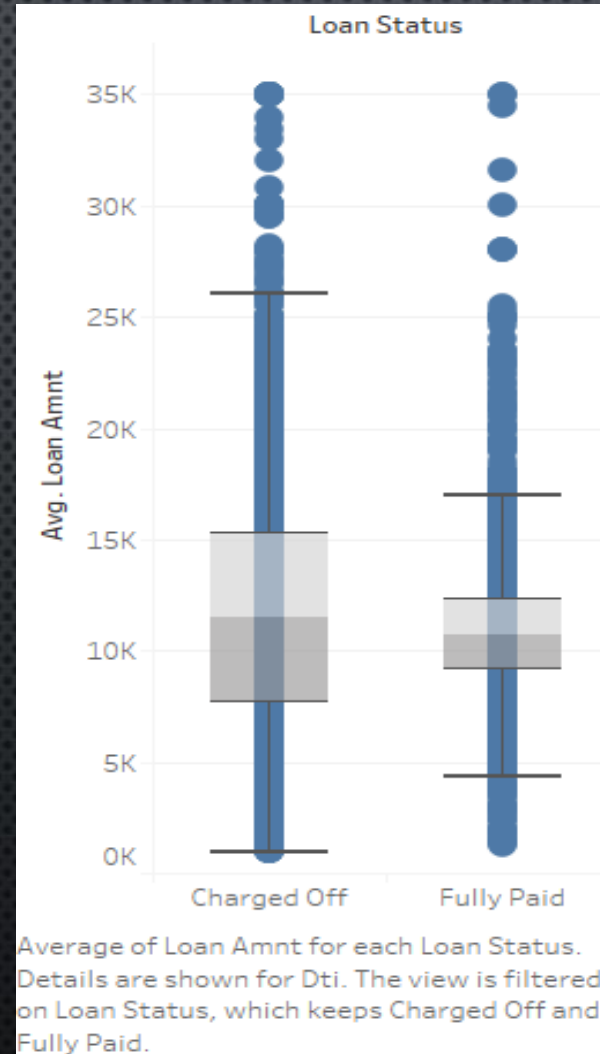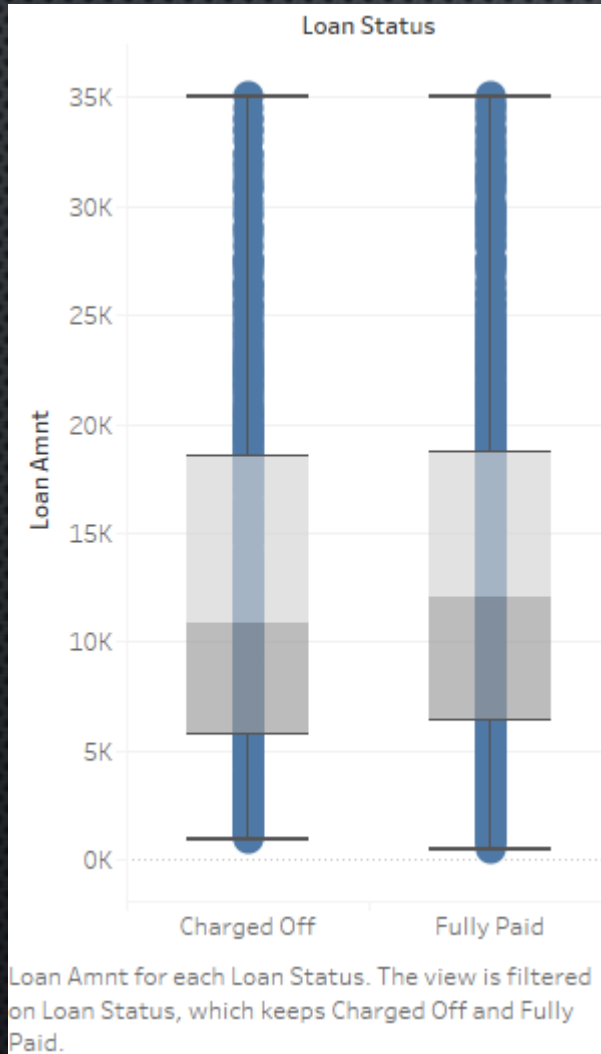The loan amount is on the rise every year as well as the defaulter amount.

AMOUNT OF DEFAULTER LOANS PER YEAR

The defaulter count percentage is lowest in 2009 and 2010.

LOANS VS STATUS

Loan Amnt for each Loan Status. The view is filtered on Loan Status, which keeps Charged Off and Fully Paid.

Average of Loan Amnt for each Loan Status. Details are shown for Dti. The view is filtered on Loan Status, which keeps Charged Off and Fully Paid.

The Median of loans for fully paid loans are slightly above than the charged off loans. The difference is not massive.

The DTI ratio median is higher for charged off loans than the fully paid loans.
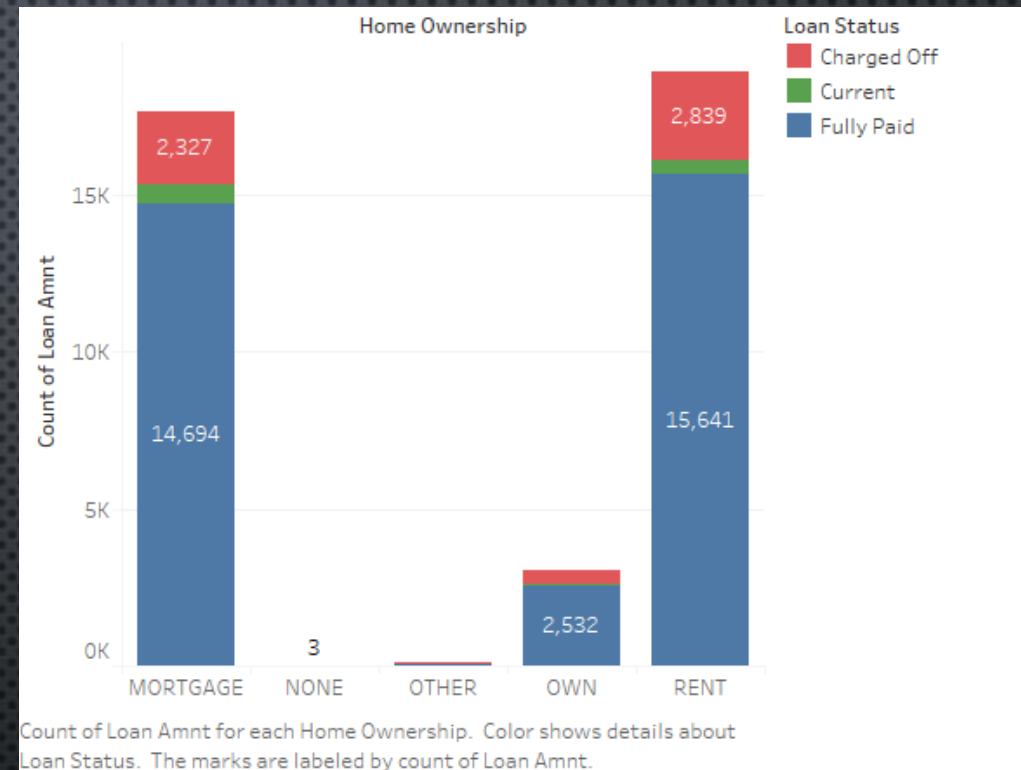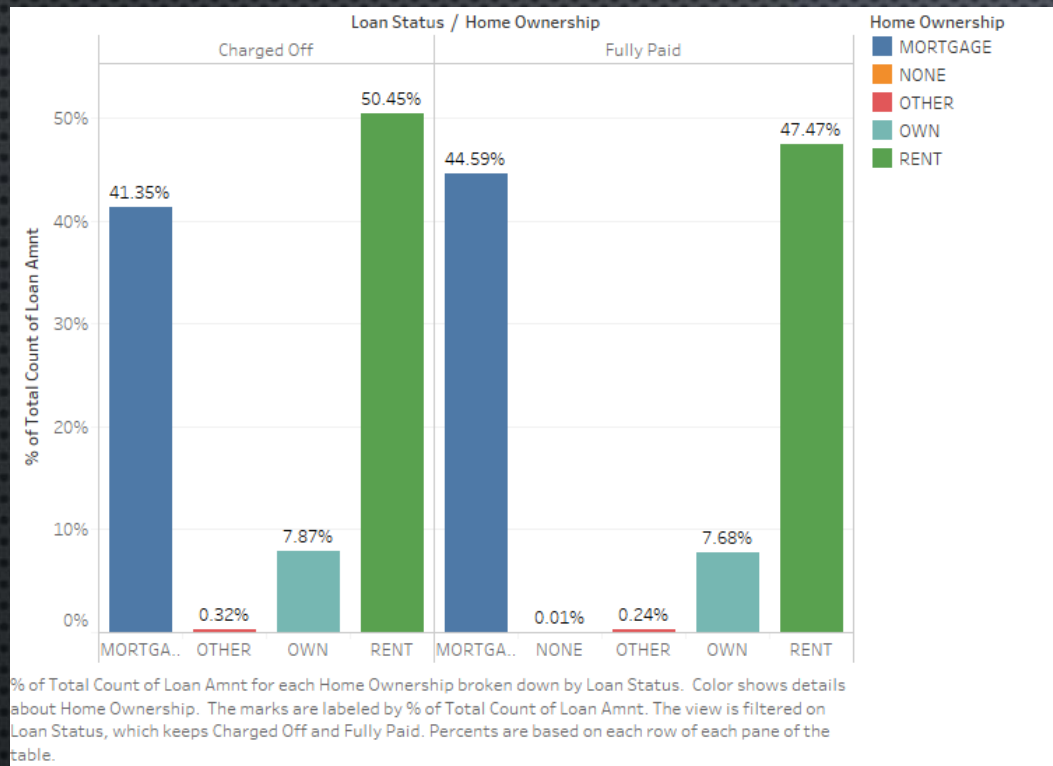
TERM VS LOAN STATUS
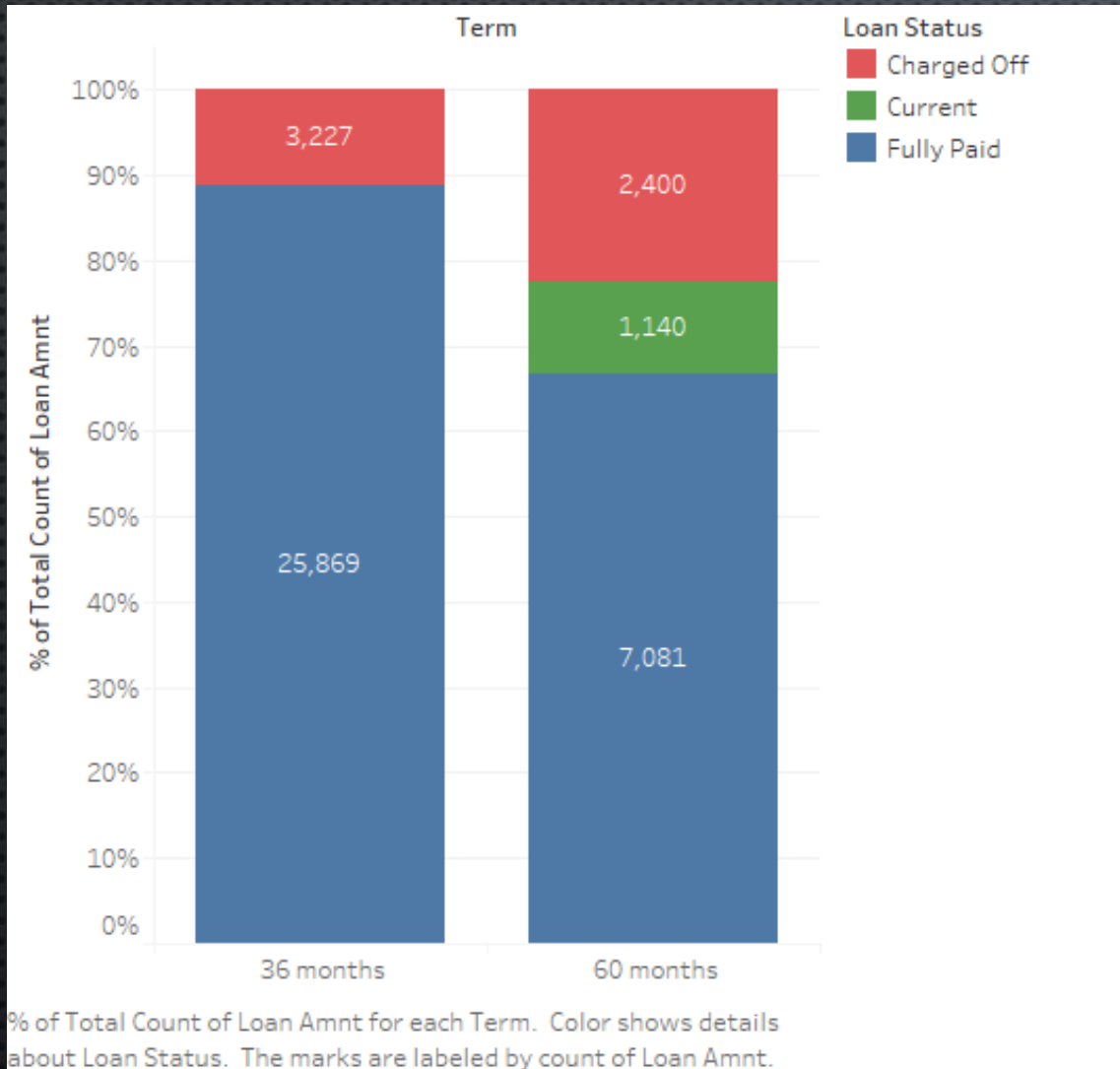
60 months term sees more defaulting of loans in percentage than 36 months.

However 36 months do have more number of defaulters.
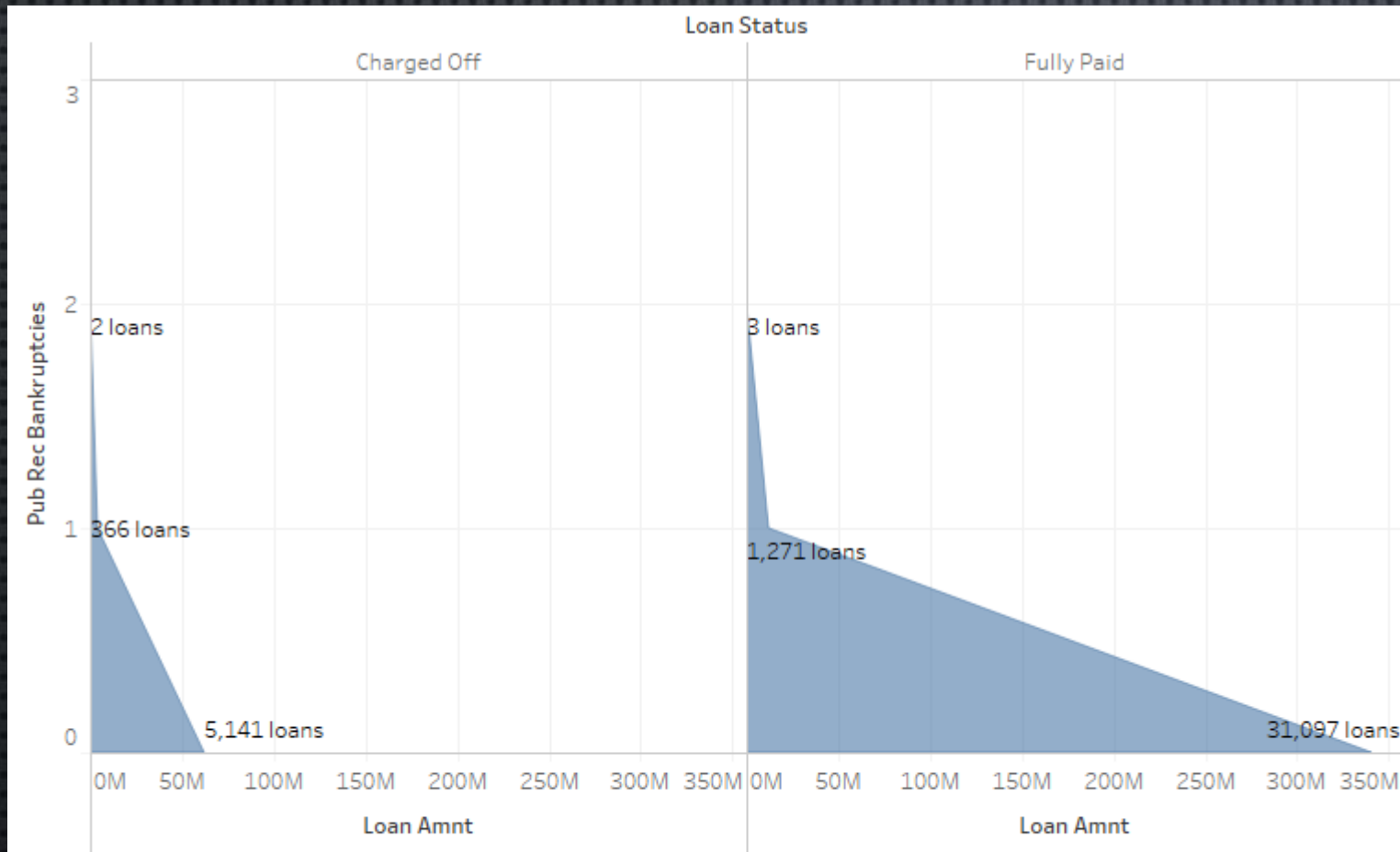
HOME OWNERSHIP VS LOAN STATUS

Plot shows percentage of charged off customers by home ownership.

Each pane is a separate graph showing the percentage of charged off and fully paid customer count by home ownership. Mortgage and Rented home owners are defaulting the most amount.

BANKRUPTCIES VS LOAN STATUS

The plot of sum of Loan Amnt for Pub Rec Bankruptcies broken down by Loan Status. The marks are labeled by count of Loan Amnt. The view is filtered on Loan Status and Pub Rec Bankruptcies. The Loan Status filter keeps Charged Off and Fully Paid. The Pub Rec Bankruptcies filter keeps non-Null values only.

Plot shows public recorded bankruptcies by loan amount for charged off and fully paid loan status.

Count of loans in the plot shows total count of loans for the number of bankruptcies.
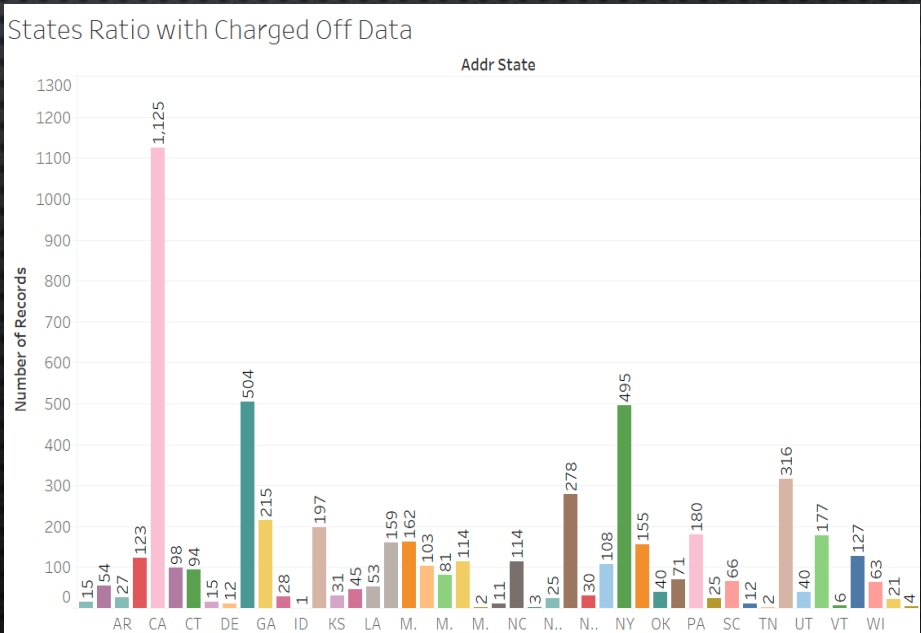
VERIFICATION VS LOAN STATUS
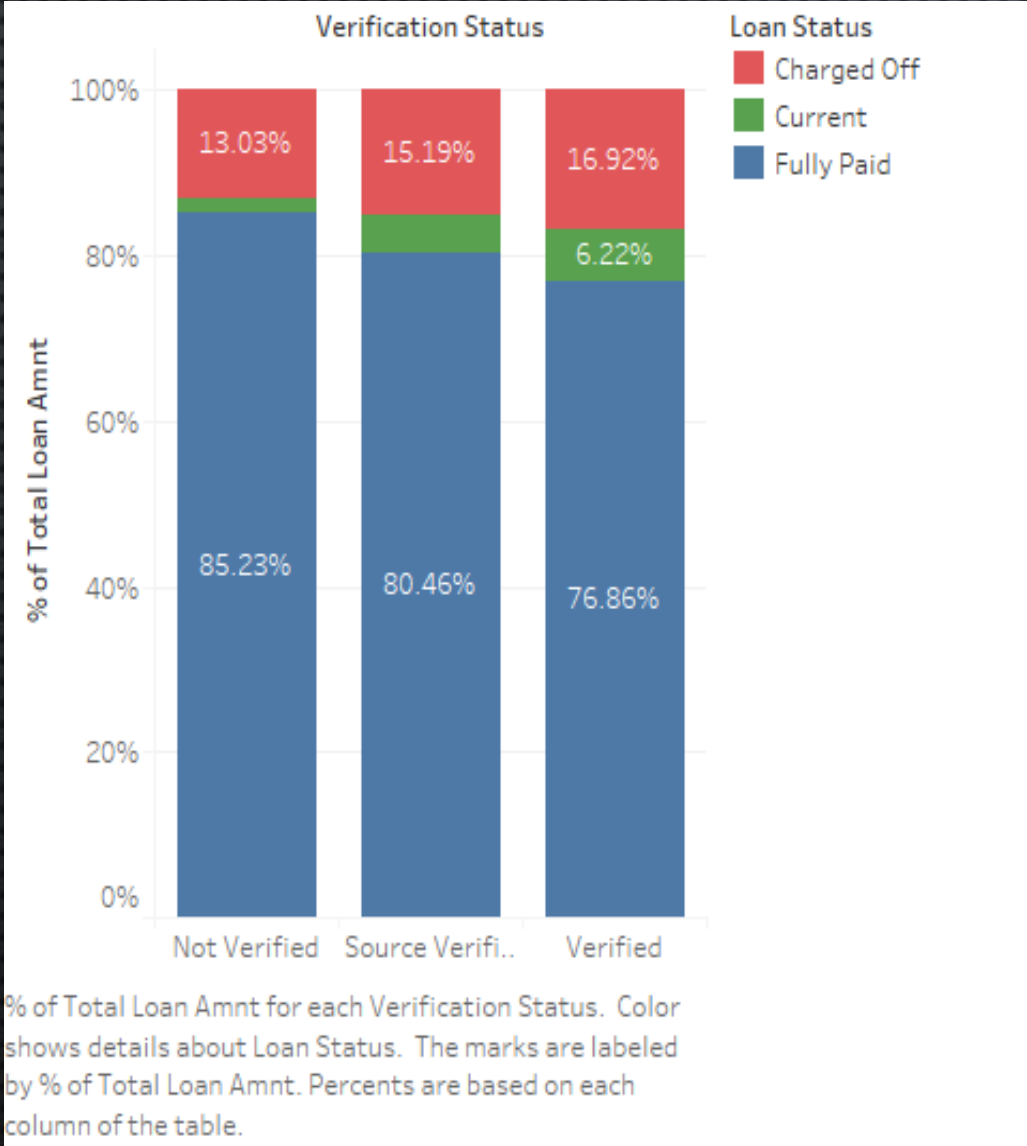
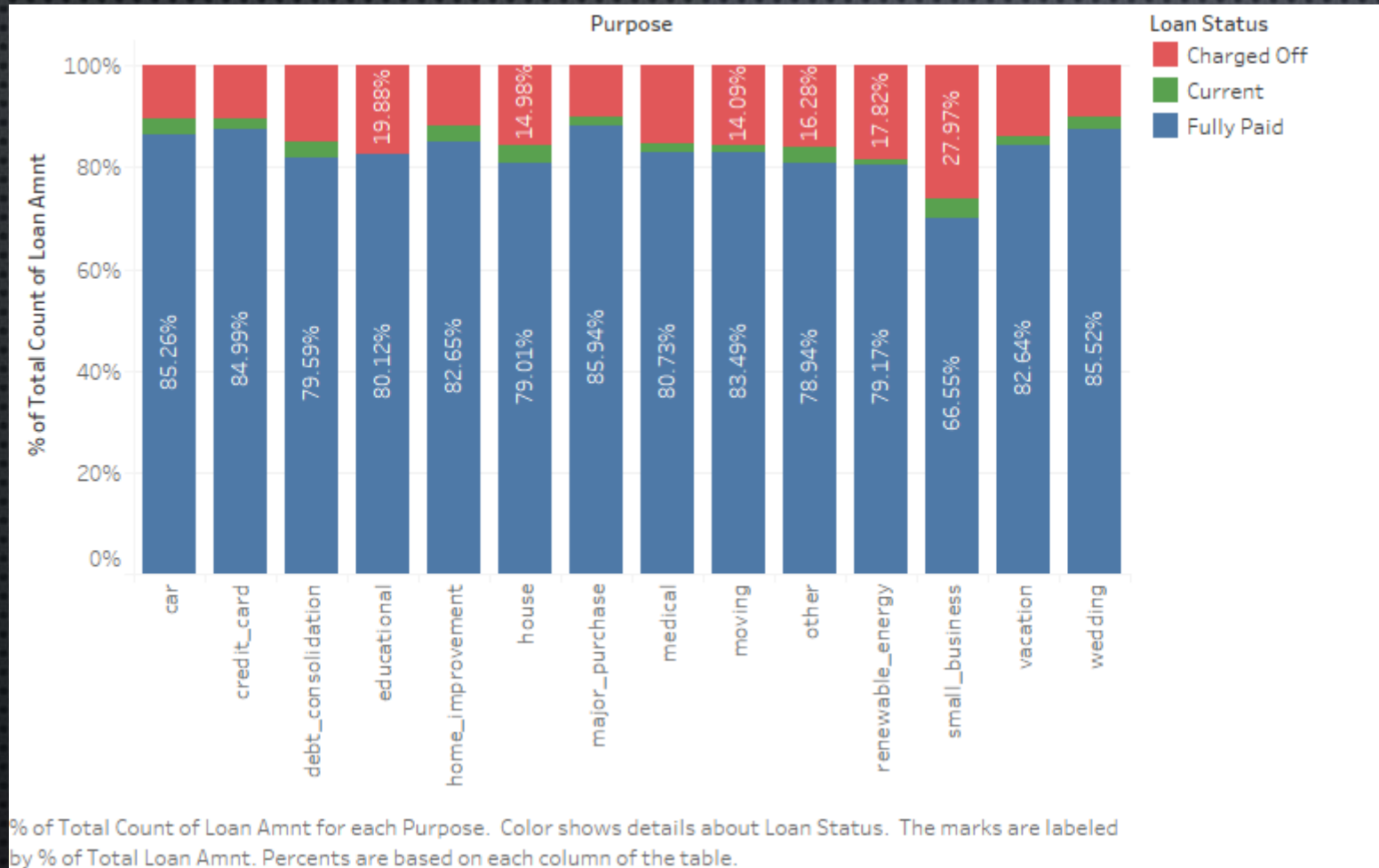Plot shows loan status for the verifications done for the customers.

Data shows that more percentage of verified loans are getting defaulted than unverified.

California State has most number of cases which are charged off

PURPOSE OF LOAN VS LOAN STATUS

% of Total Count of Loan Amnt for each Purpose. Color shows details about Loan Status. The marks are labeled by % of Total Loan Amnt. Percents are based on each column of the table.

Plot shows purpose of loan and the percentage of charged off customers.

Data shows that small businesses have the highest no of defaulters.
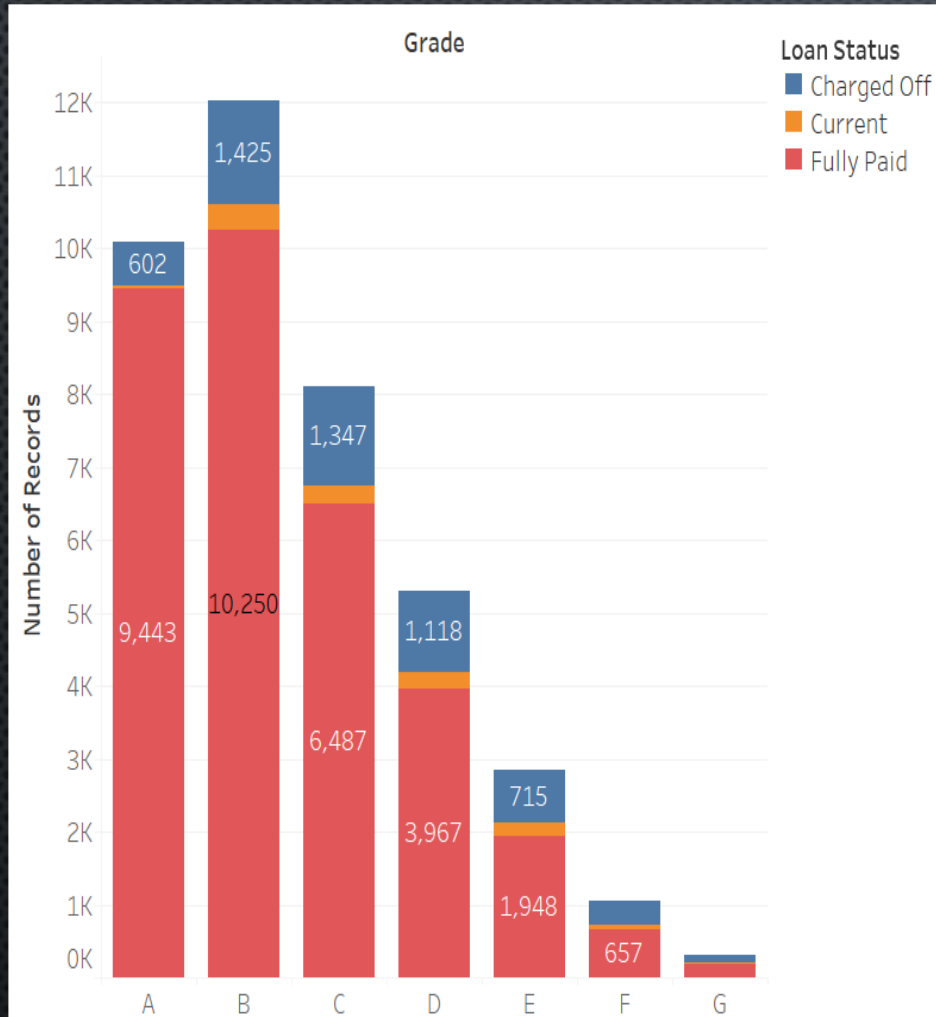
GRADE VS LOAN STATUS
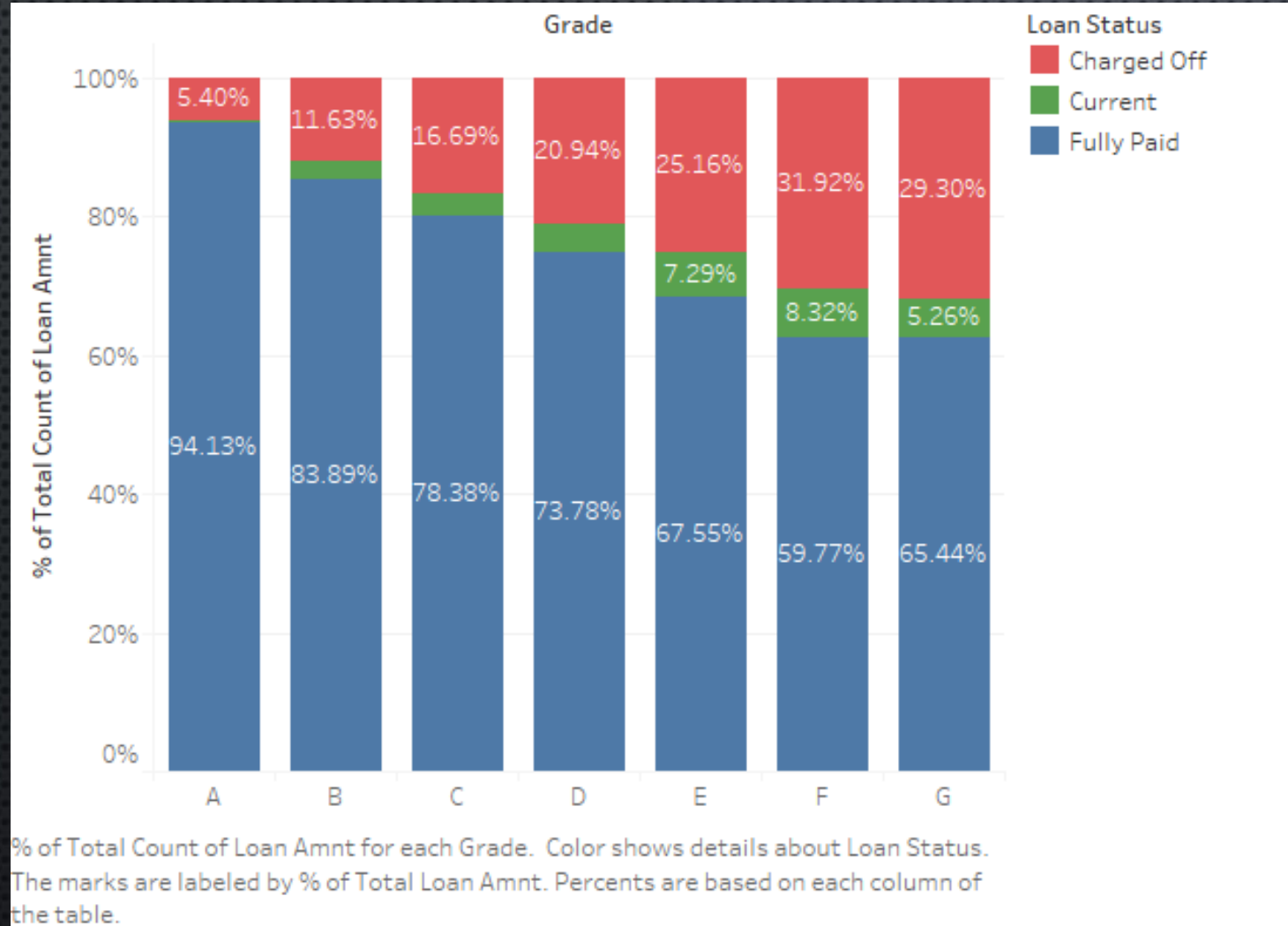
Plot percentage of charged of customers for each grade.

Data shows that lower the grade, the higher the percentage of charged of customers.
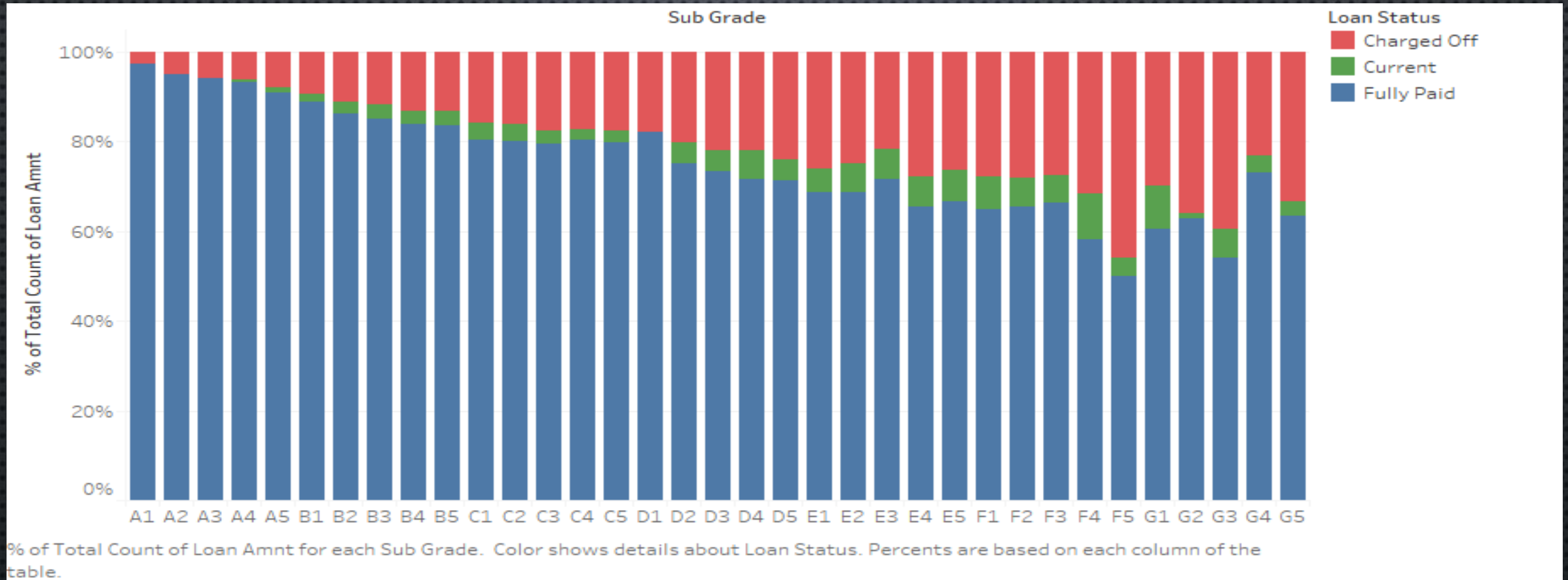
GRADE VS LOAN STATUS

B,C & D grades are showing the most no of defaulters.
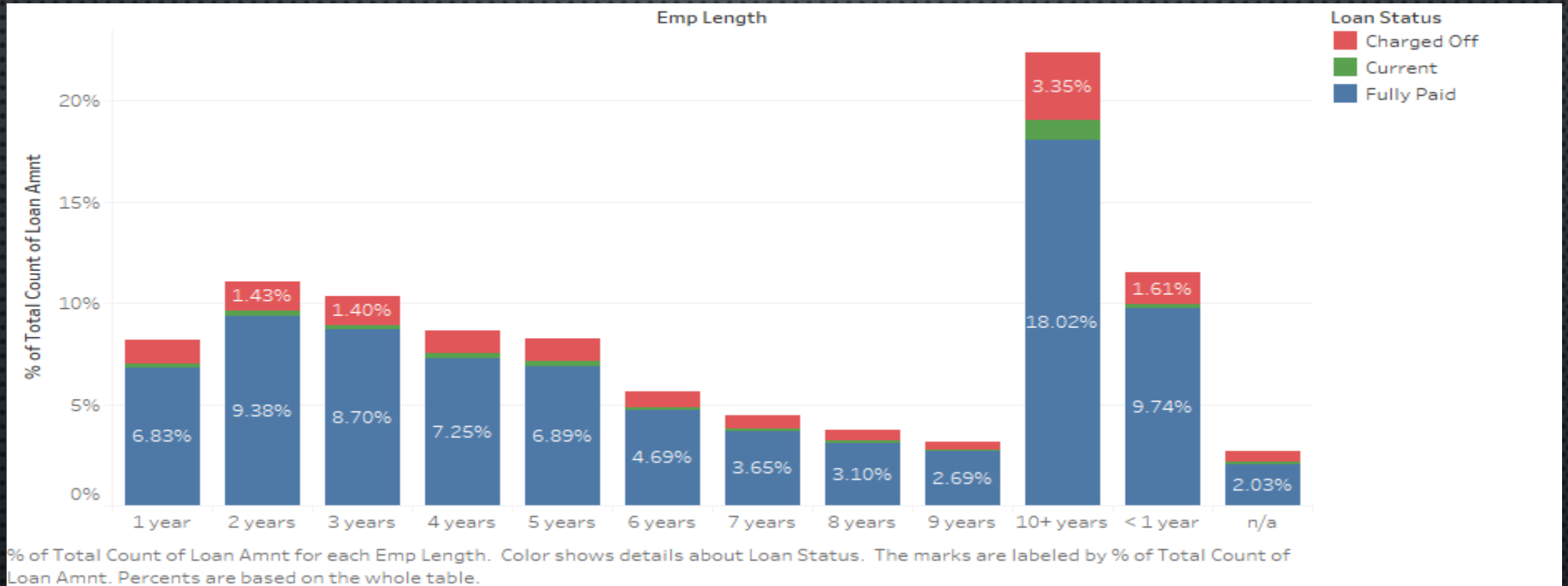
SUB GRADE VS LOAN STATUS

Plot percentage of charged of customers for each subgrade.

Data shows that generally lower the subgrade, the higher the percentage of charged of customers.

SUB GRADE VS LOAN STATUS

% of Total Count of Loan Amnt for each Emp Length. Color shows details about Loan Status. The marks are labeled by % of Total Count of Loan Amnt. Percents are based on the whole table.

Plot percentage of charged of customers for the years of their experience in their company. Percentage labels are calculated for the total data.
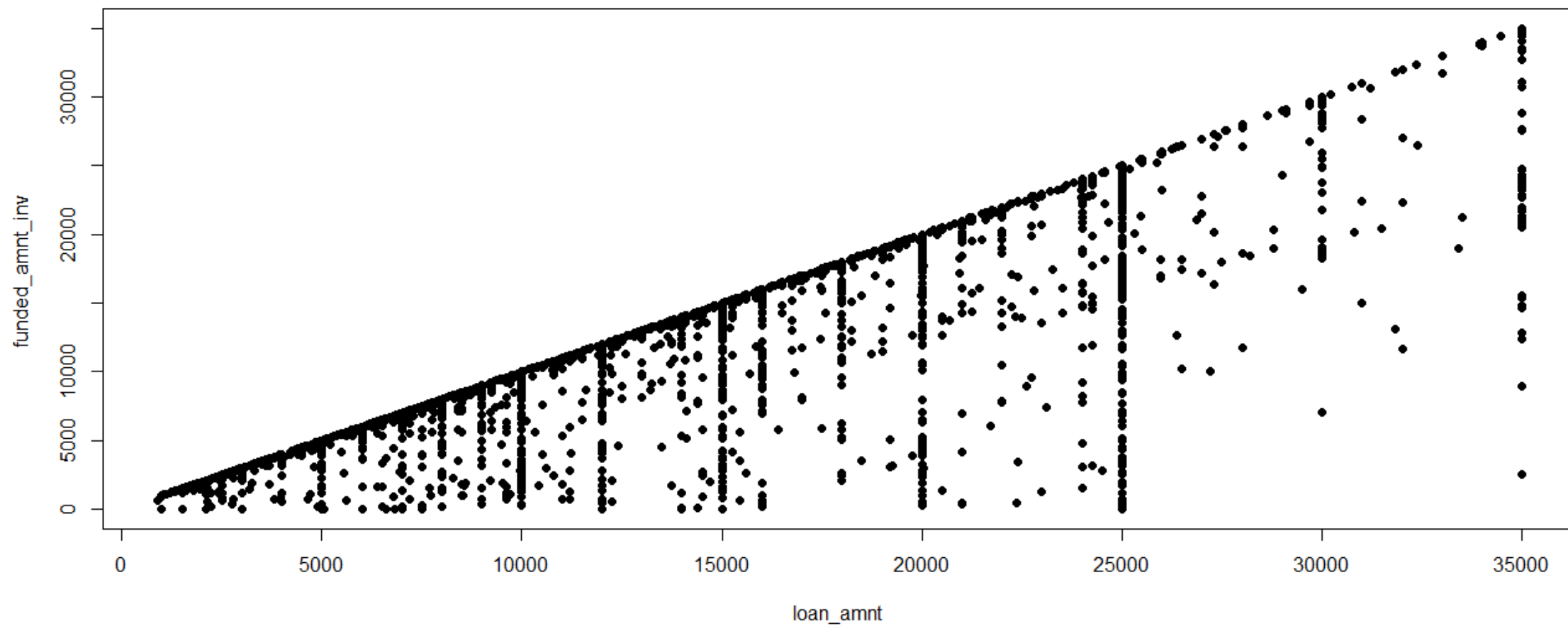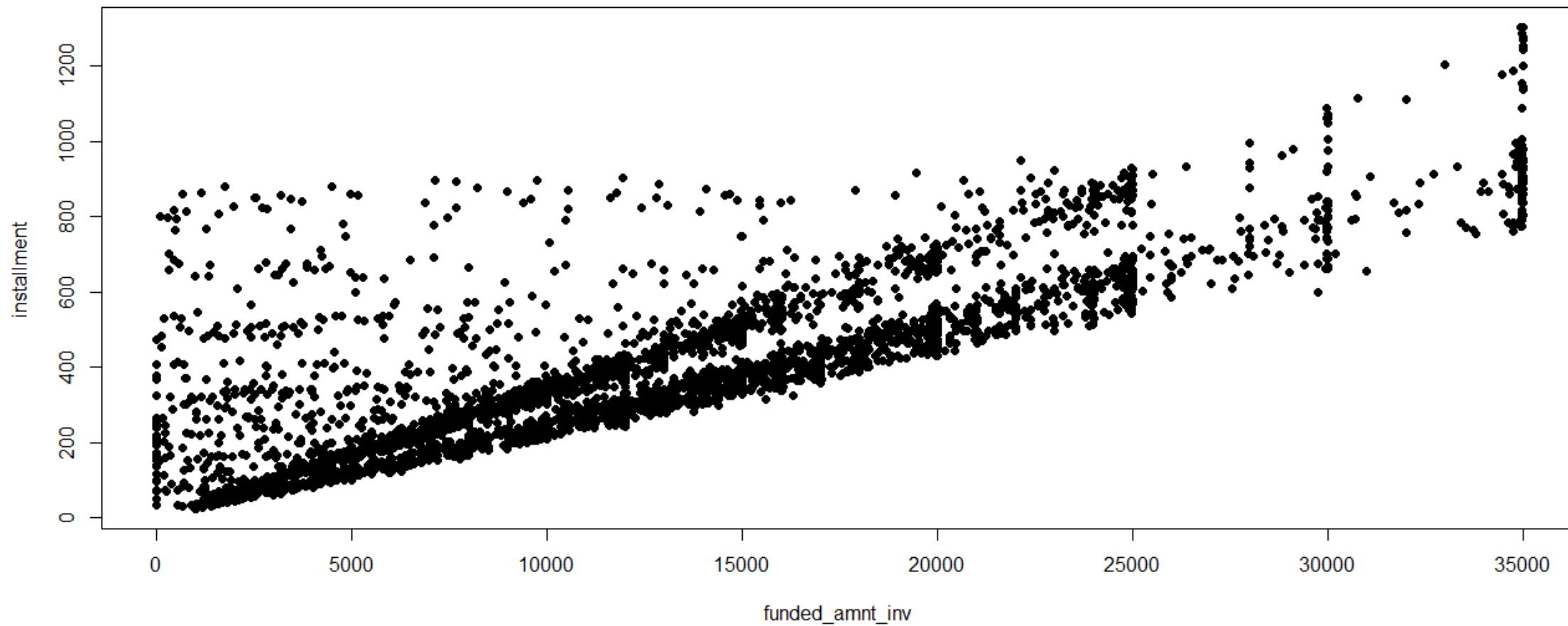
DELINQUENCY(2YEARS) VS LOAN STATUS
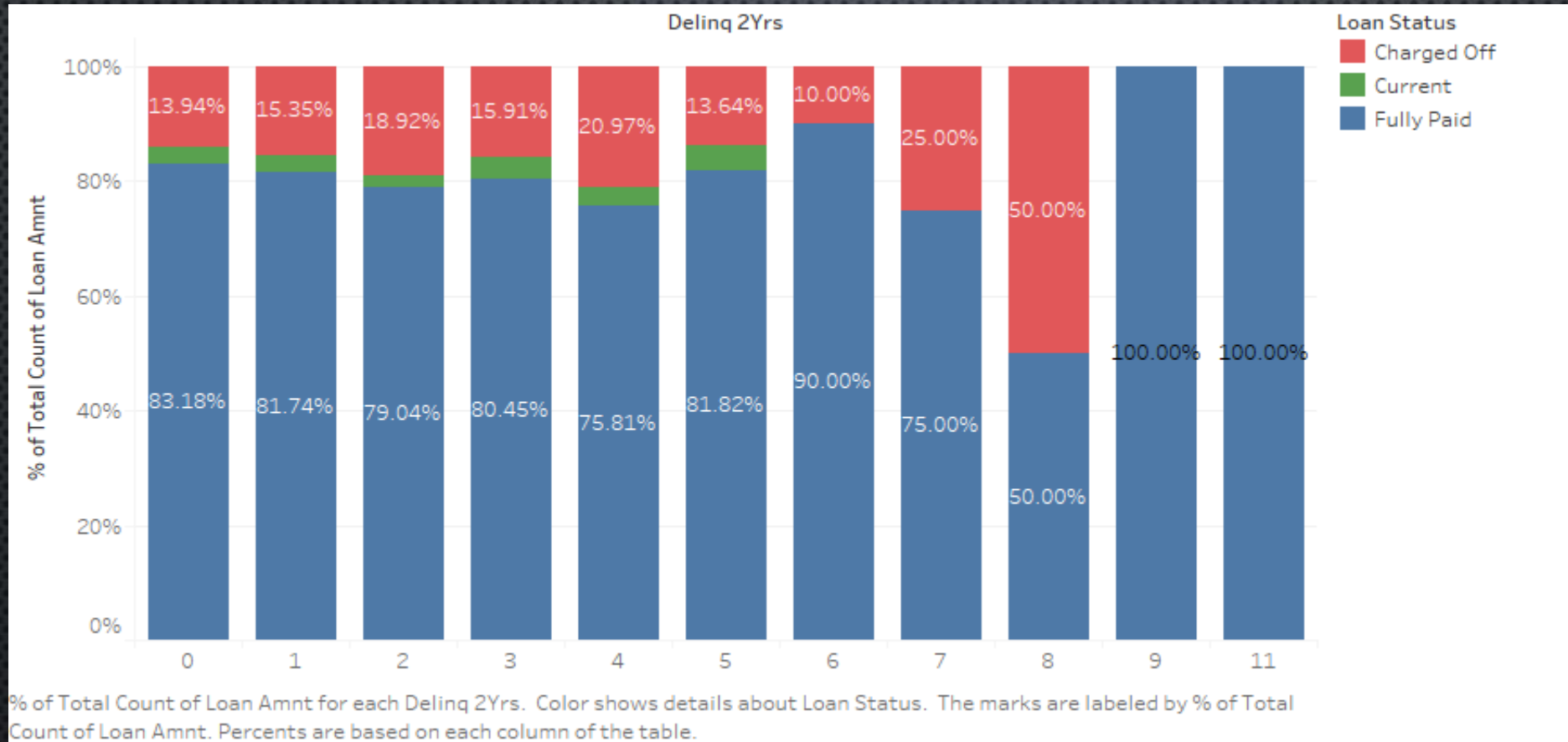
Plot percentage of charged of customers for the number of delinquency (30days) in the last 2 years. Percentage labels are calculated for the total data.

correlation between loan_amnt and funded_amnt_inv

correlation between funded_amnt_inv and installment

CREDIT CARD REVOLVING BALANCE