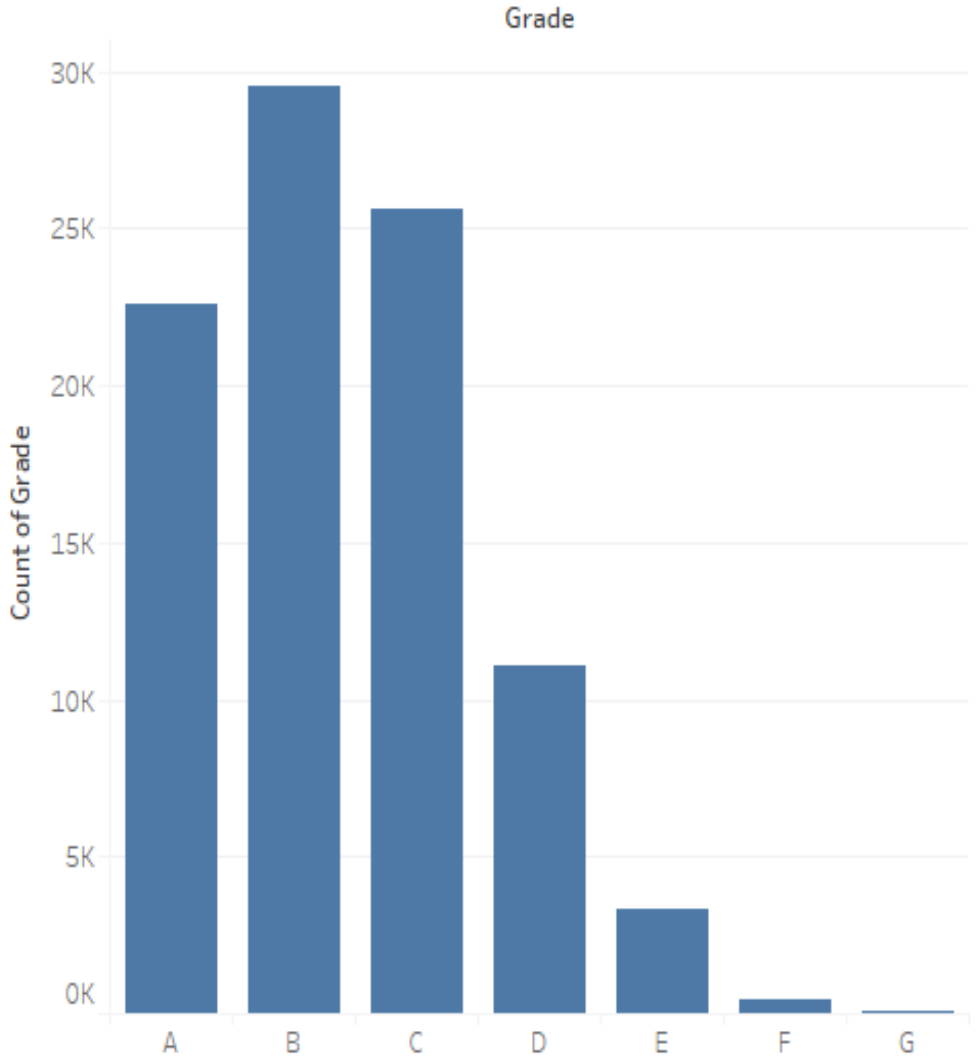
About Lending Club

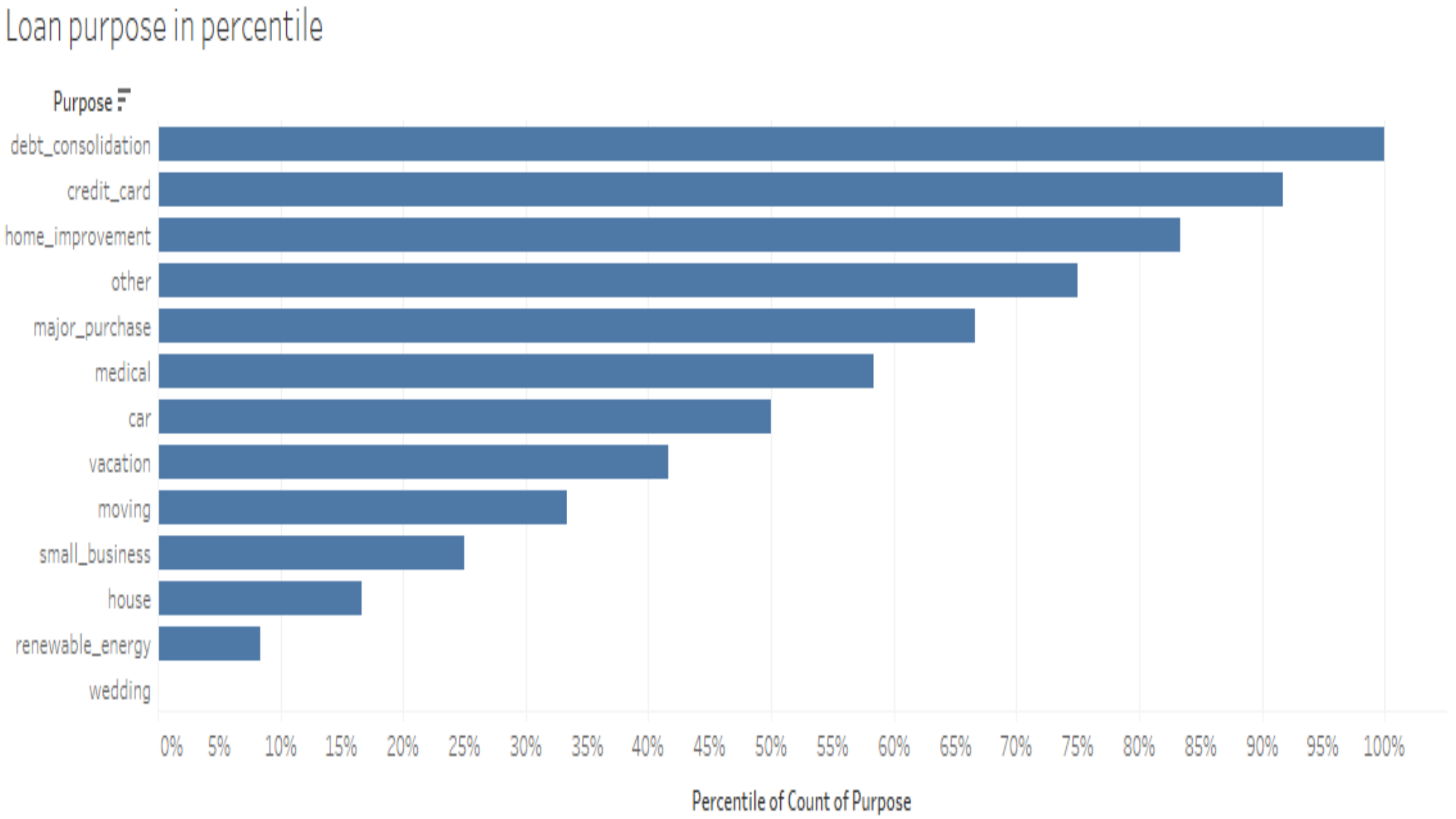
- LendingClub is the largest personal loans provider in the **U.S**. and a marketplace
- It is essentially a peer to peer lending platform.
- **Peer-to-peer** (P2P) lending, is the practice of lending money to individuals or businesses through online services that match lenders with borrowers.

No.of Loan by Grade

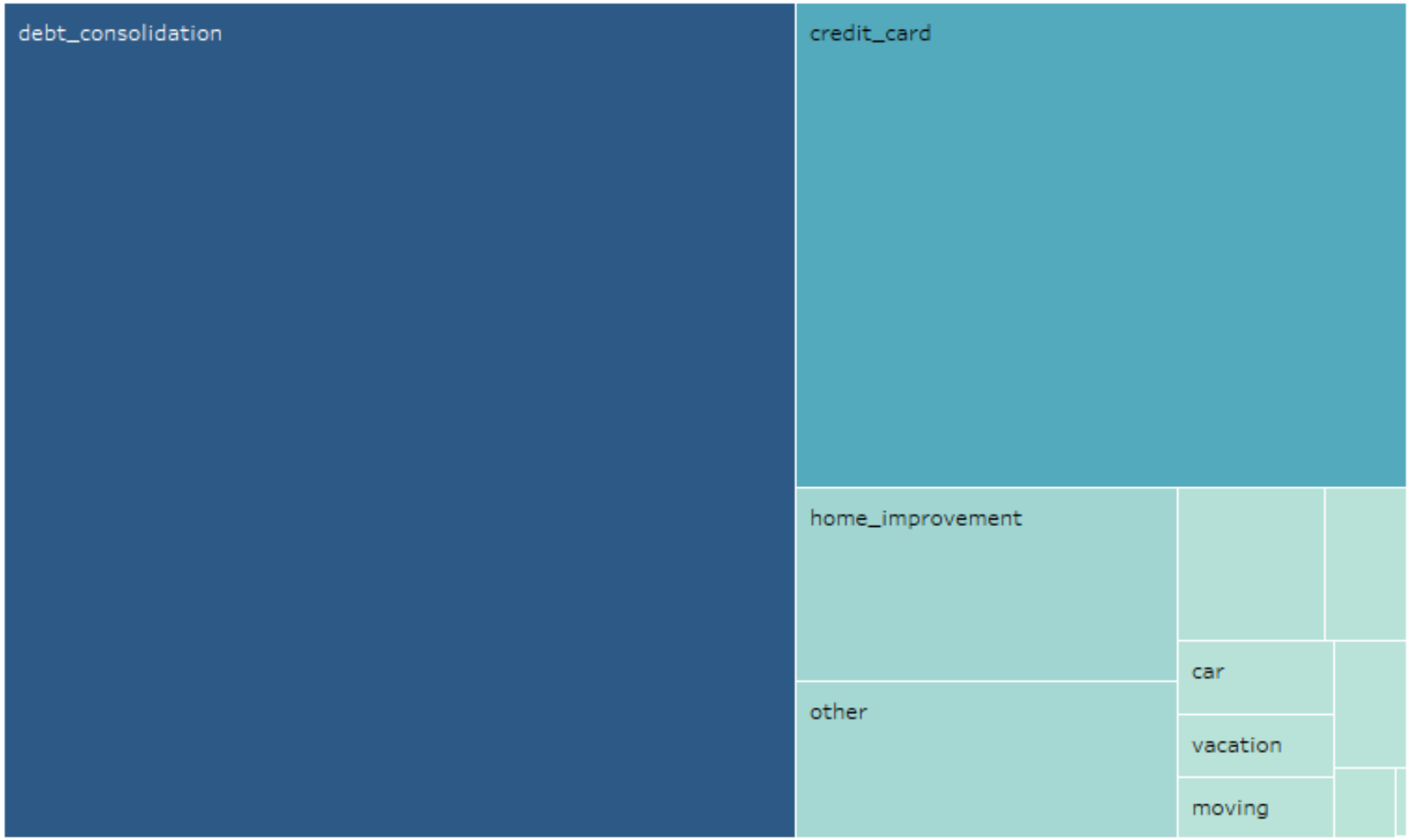


Maximum loans is observed to be falling under the category of B and C.

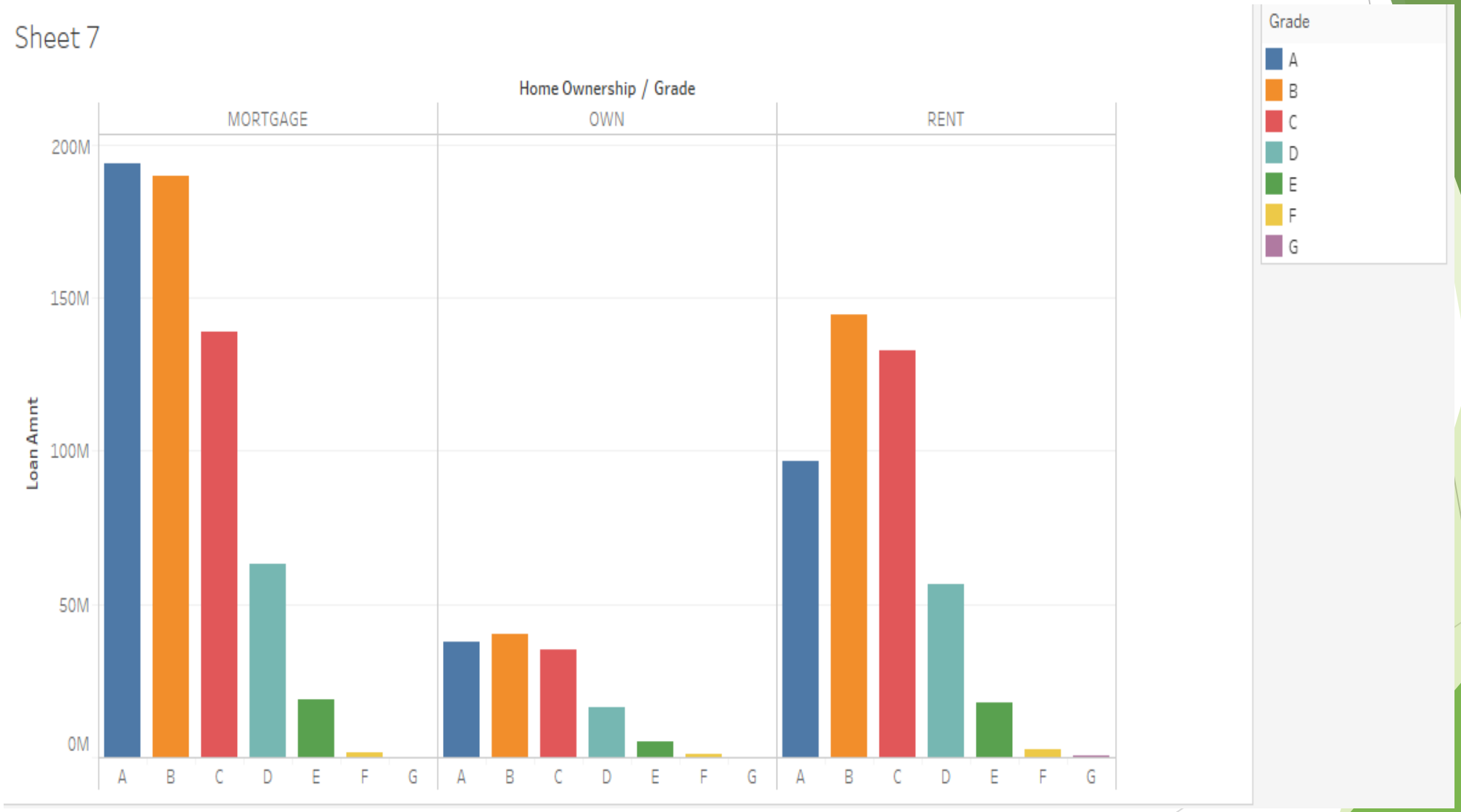
Major categories of loan purpose

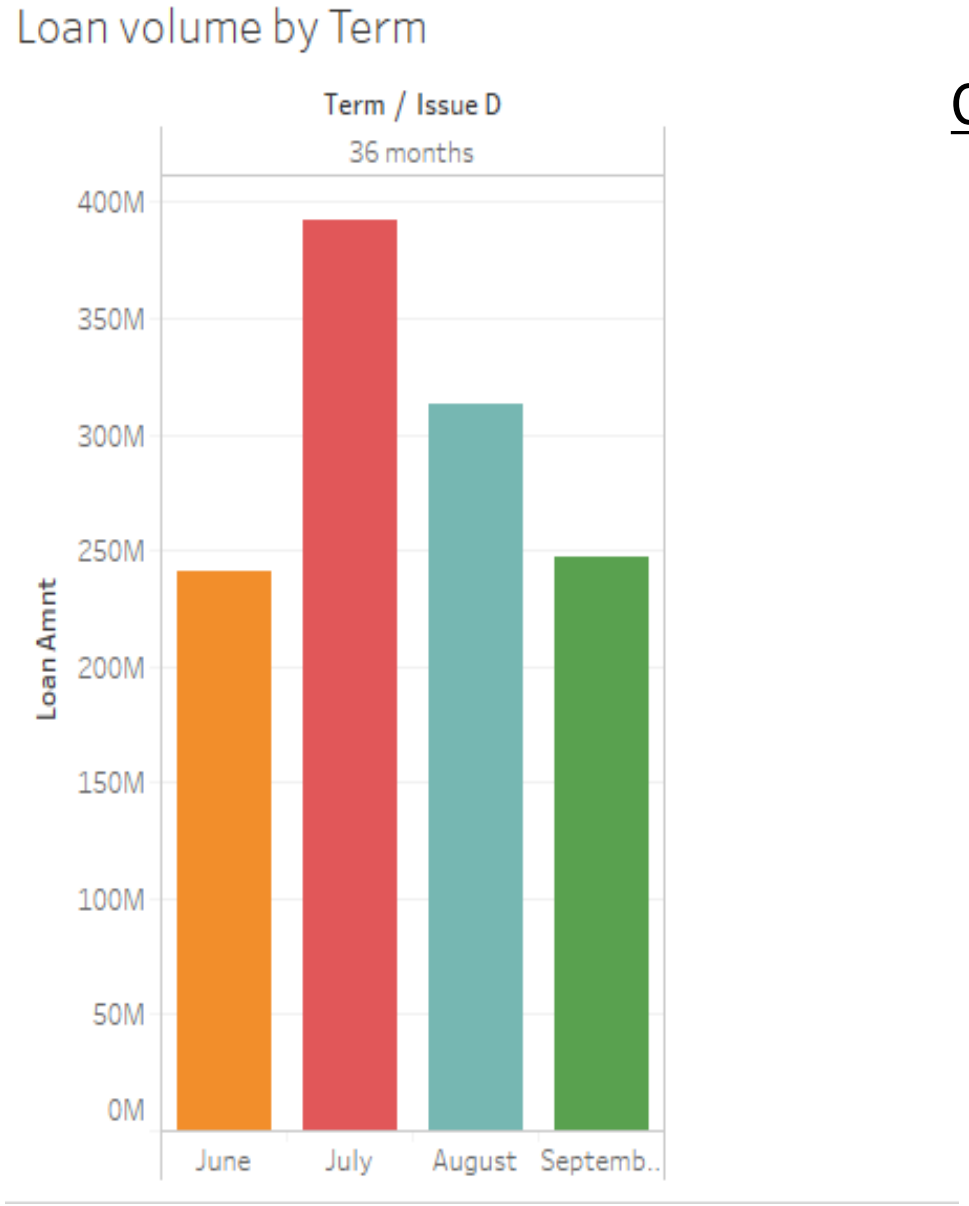


Tree Map showing Loan Purpose



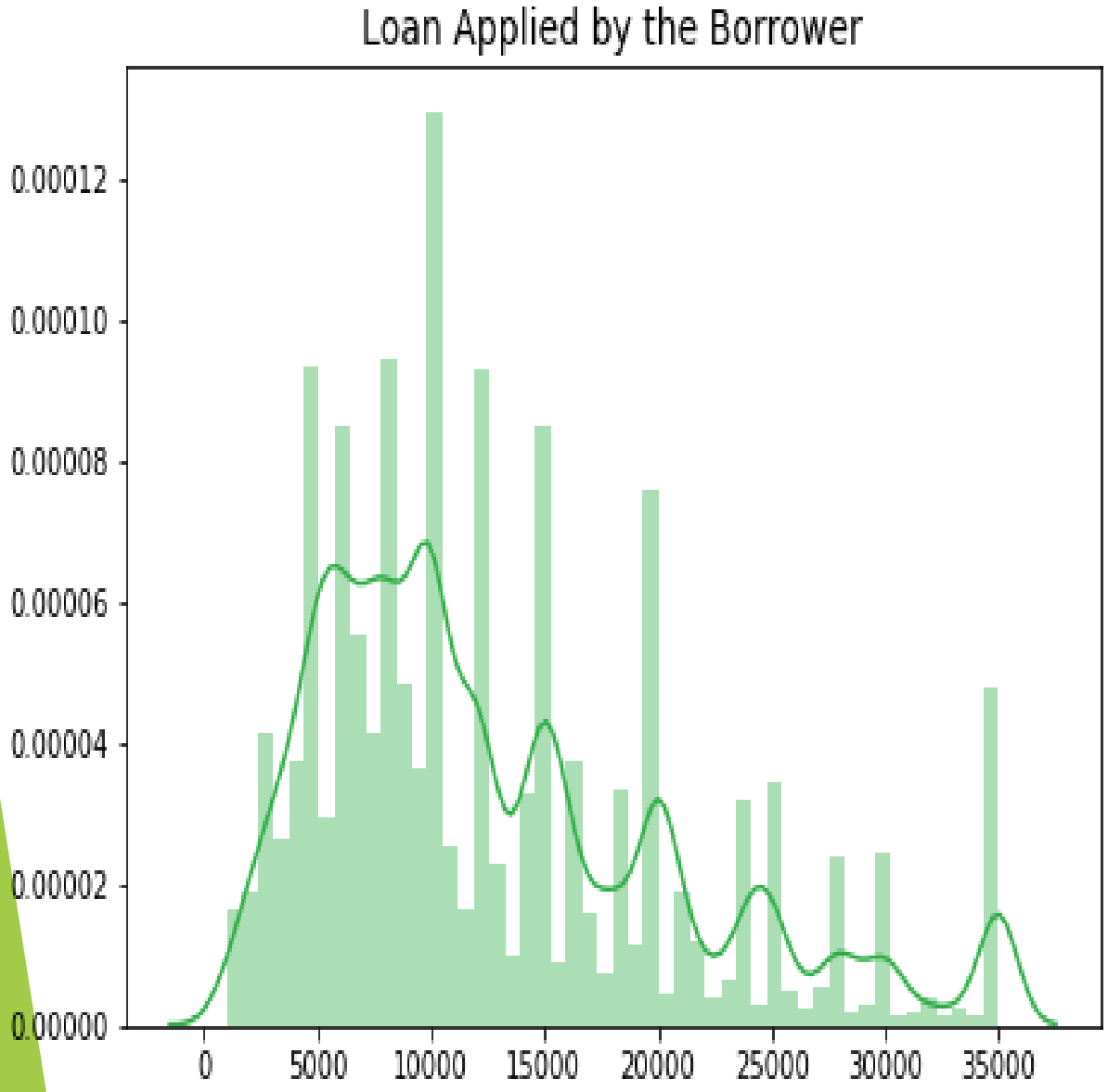
Distribution of loans based on Homeownership

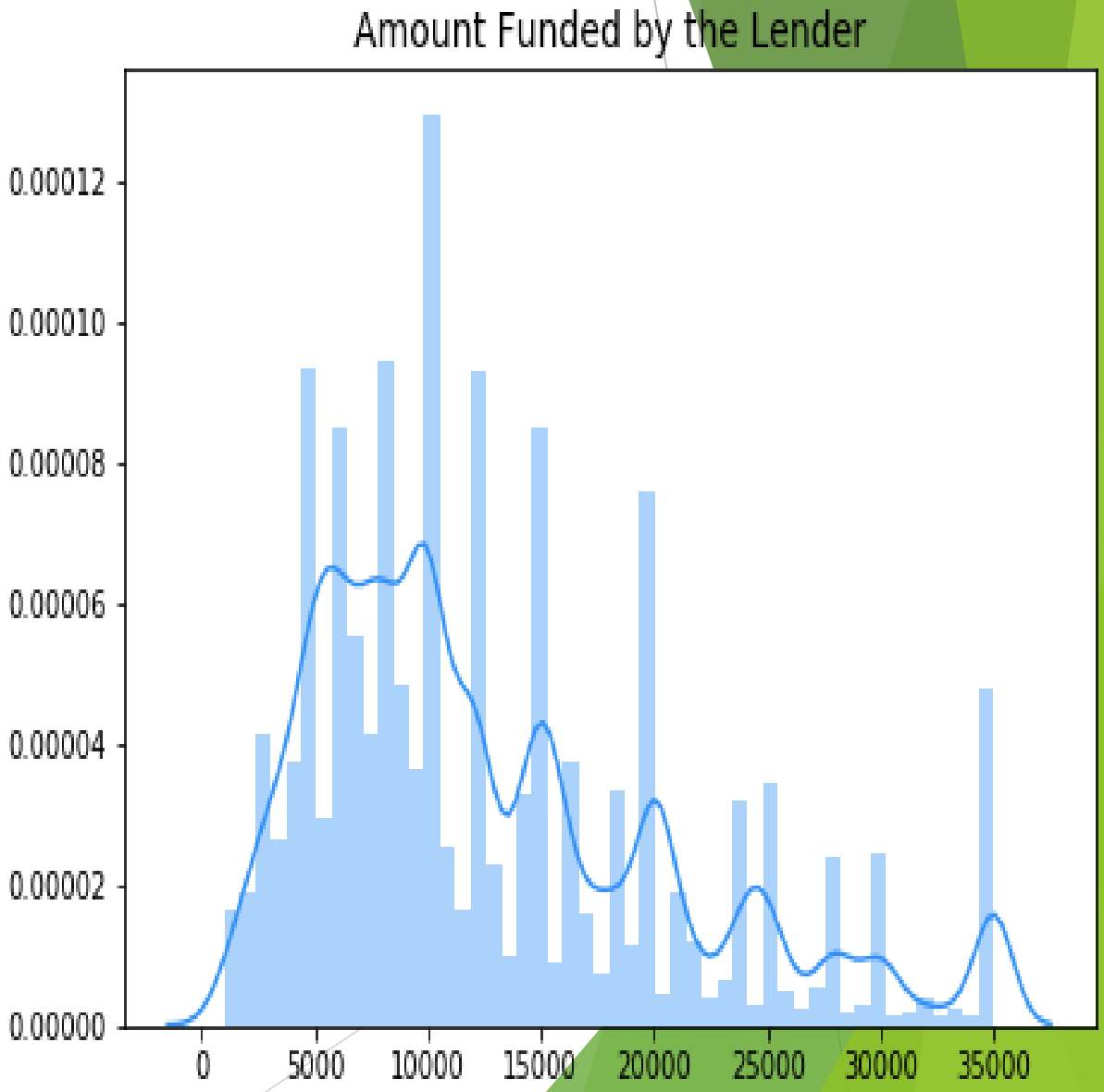


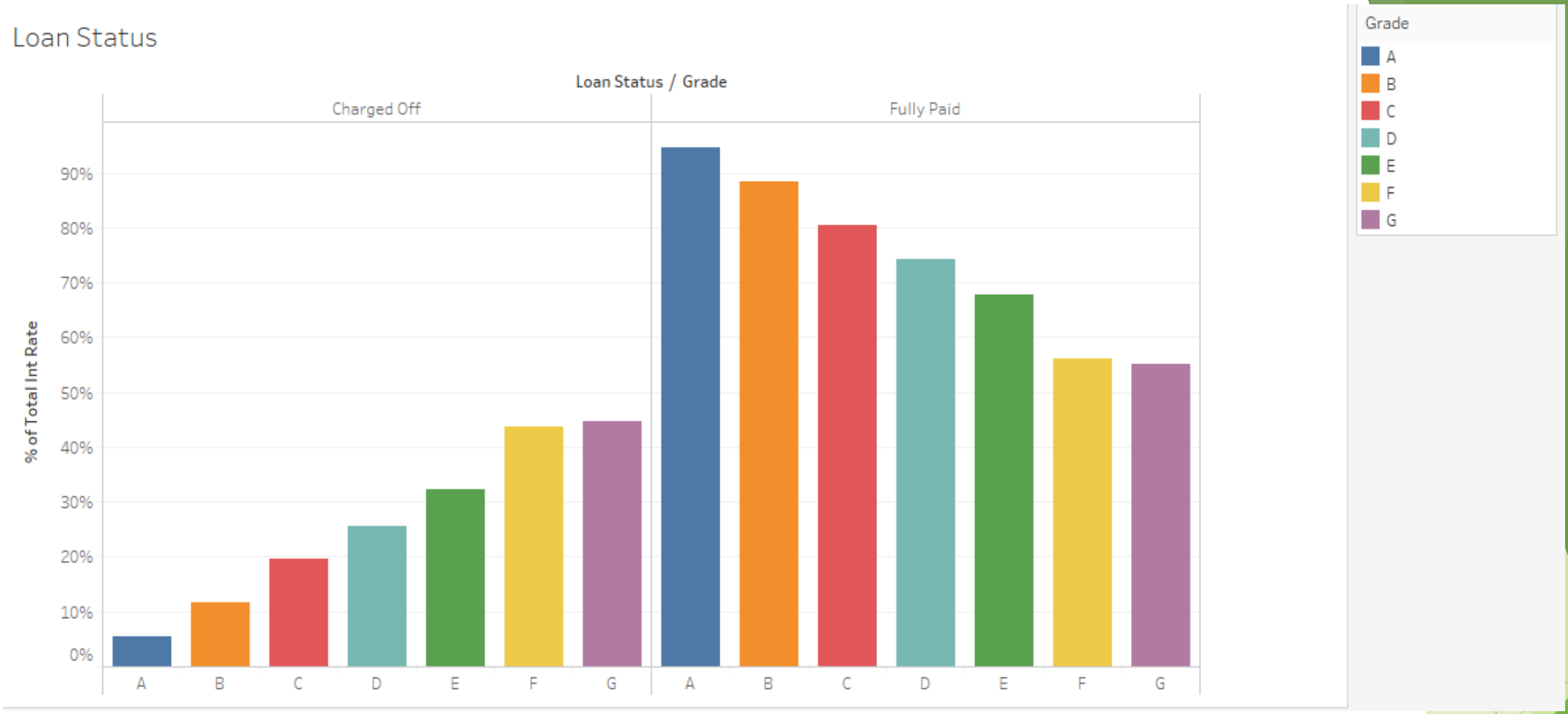


Overall loan monthly wise

Proportion of Supply and Demand

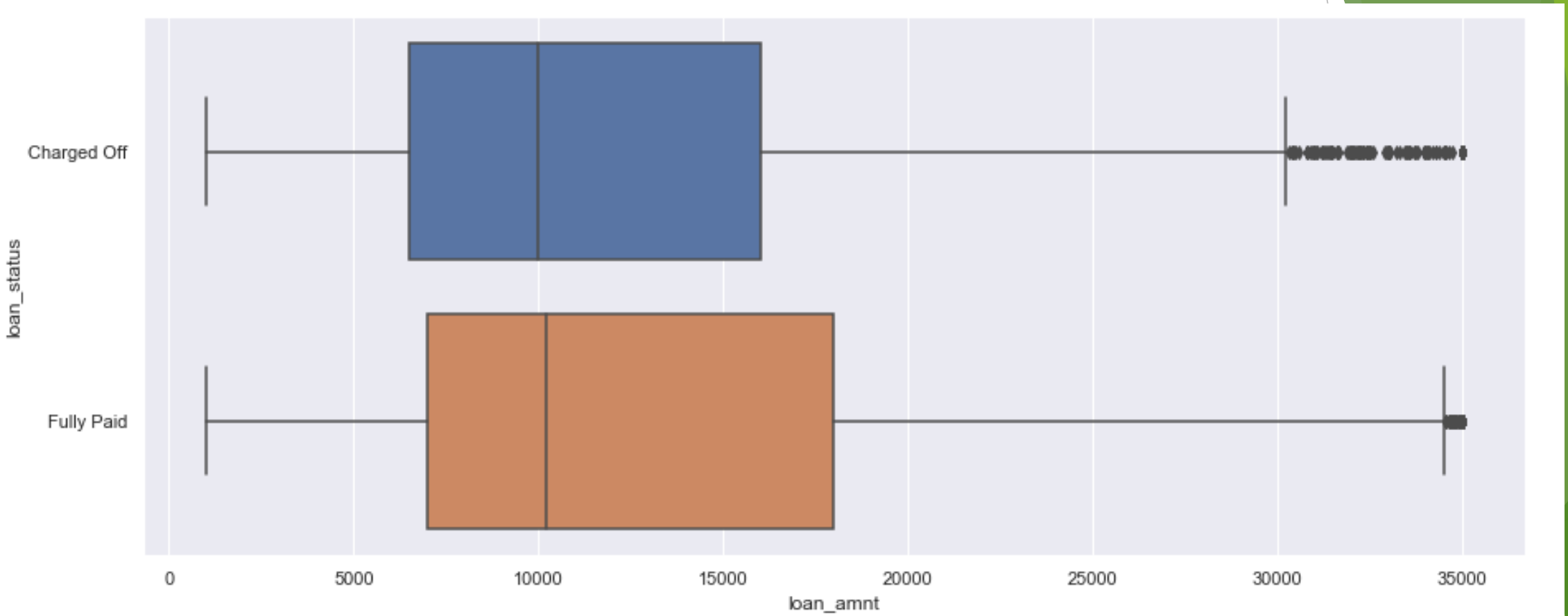




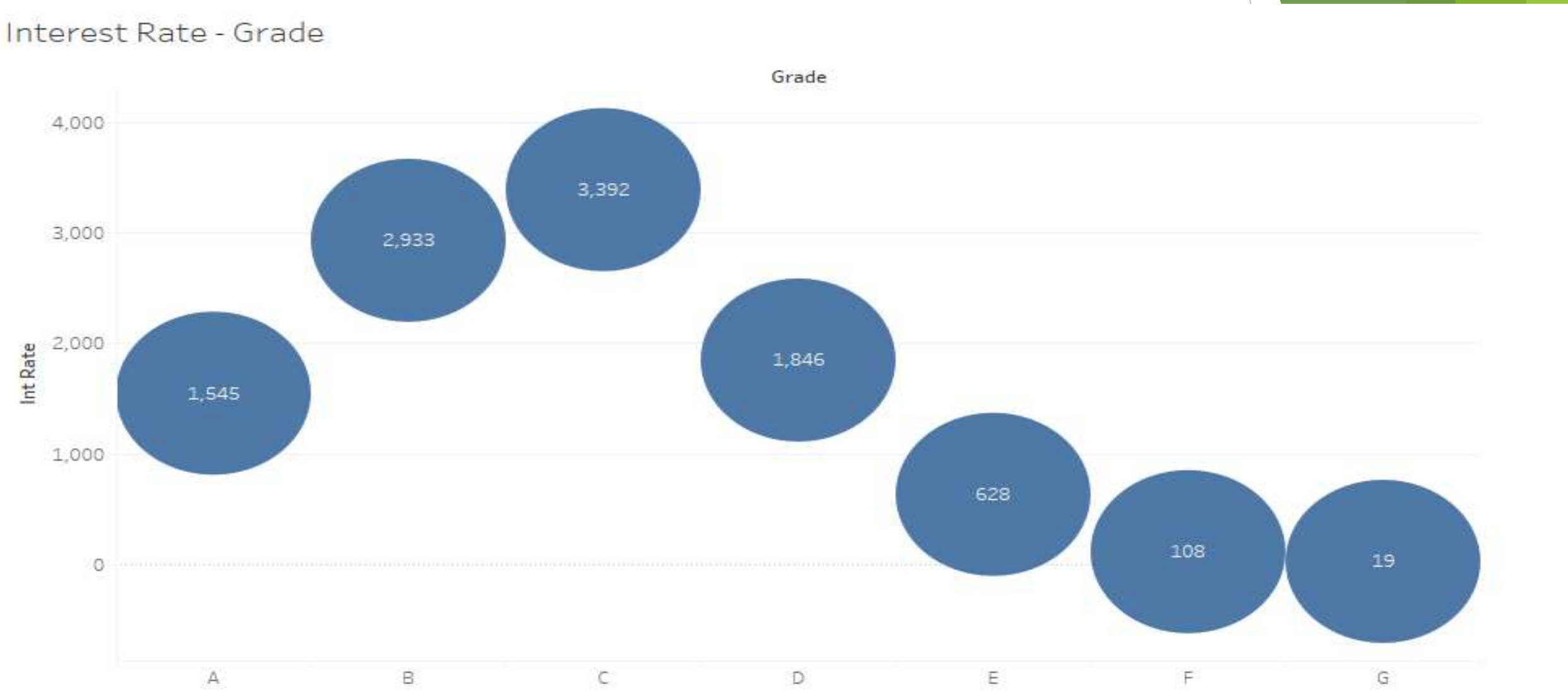


Loan Status categorised as Fully Paid and Charged off loan

1. Based on Grade
2. Based on interest Rate


Charged Off
Fully Paid
loan_status
0
5000
10000
15000
20000
25000
30000
35000
loan_amnt

Interest rate distribution over Grade



Problem Statement

The task is the development of the appropriate ML model which can be used to predict '**Fully Paid" vs " Charged Off**" loans in the Lending Club platform. In this task, you need to develop models to predict which loans are likely to be paid off and which will default. You also have to judge the interrelationship among the variables and determine an effective investment decisions based on your predictive models – we need examine this in details here. You need to focus on parameter tuning, and reliable performance estimates through resampling and cross-validation.

Data Cleaning

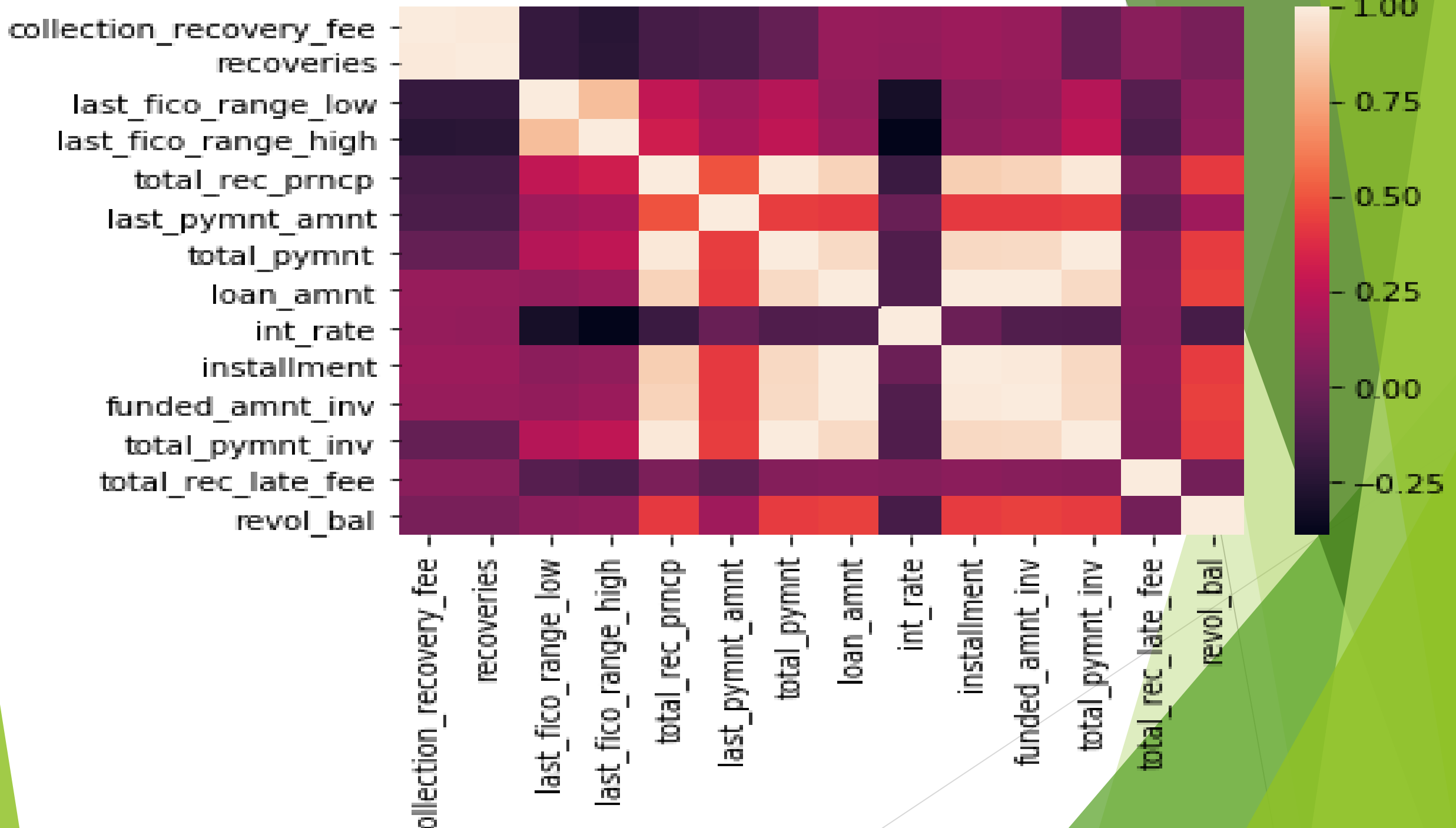
- Deleting columns with no values
- Excluding Univariate columns
- Removal of attributes such as url ,zipcode etc based on description

Data consists of 92624 rows and 79 attributes.

Tools Used : Tableau

Programming language : Python

Models are based on Machine Learning Algorithms.


collection_recovery_fee
recoveries
last_fico_range_low
last_fico_range_high
total_rec_prncp
last_pymnt_amnt
total_pymnt
loan_amnt
int_rate
installment
funded_amnt_inv
total_pymnt_inv
total_rec_late_fee
revol_bal
1.00
0.75
0.50
0.25
0.00
−0.25

Numerical variable Feature Selection

13 attributes were extracted after decision tree modelling

Model	Random Forest	Decision Tree
Accuracy Score	96.53	96.56

Categorical variable

Pi-square test has been applied on all varaibles and accordingly , attributes of importance is sorted out

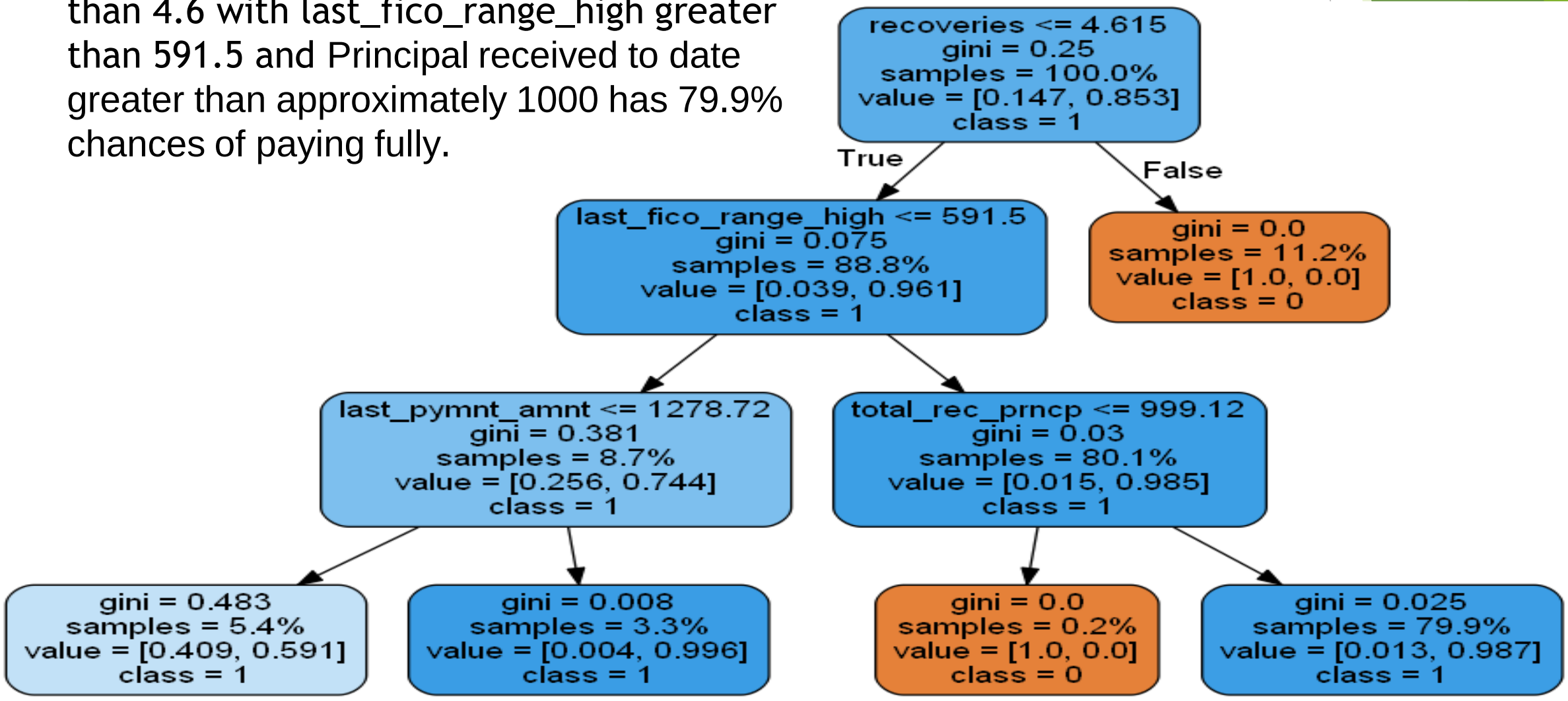
Overall Performance of Models on Loan Status

Models were applied on data obtained after cleaning

MODEL	DECISION TREE	BAGGING	RANDOM FOREST	ADA BOOSTING	GRADIENT BOOSTING
ACCURACY SCORE	**95.99%**	**99.62%**	**99.48%**	**99.28%**	**99.26%**

Loan Status interpretation

Given post charge off gross recovery less than 4.6 with last_fico_range_high greater than 591.5 and Principal received to date greater than approximately 1000 has 79.9% chances of paying fully.



Factors affecting loan_status

- **Loan amount** :The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value.
- **Funded_amnt_inv** : The total amount committed by investors for that loan at that point in time.
- **Purpose_credit_card** : A category provided by the borrower for the loan request.
- **Home_ownership** : The home ownership status provided by the borrower during registration. Our values are: **RENT**, OWN, **MORTGAGE**, OTHER.
- **Total_rec_prncp** : Principal received to date
- **Int_rate** : Interest Rate on the loan
- **Total_pymnt** : Payments received to date for total amount funded
- **Title_Debt consolidation** :The loan title provided by the borrower
- **Emp_length_10+ years** : Employment length in years.