

# Lending Club Case Study

- Sampatkuamr Akki

# Objective

The aim is to identify patterns in applicants profiles to determine if they are likely to default, which could lead to actions such as denying loans, reducing loan amounts, or adjusting interest rates.

When a person applies for a loan, the company can make two types of decisions:

Loan Accepted:

- Fully Paid: The loan is fully paid, including the principal and interest rate.
- Current: The applicant is in the process of paying instalments, and the loan tenure is not yet completed.
- Charged-Off: The applicant has defaulted by failing to make due payments for an extended period.

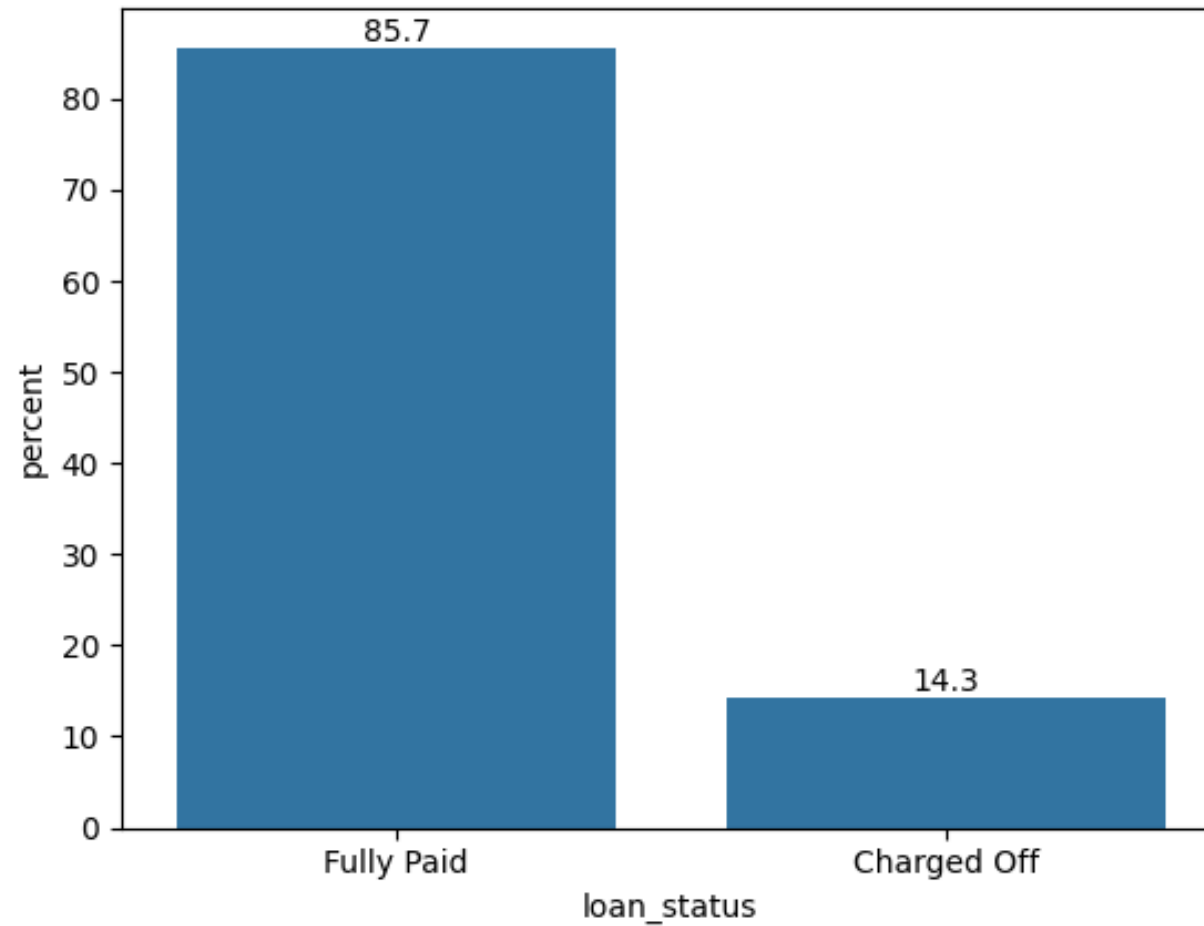
- Loan Rejected: The loan was denied due to the applicant not meeting the company's requirements, and no transactional history exists for these applicants.

Identifying these risky applicants can help reduce credit loss by preventing such loans from being approved. The company aims to understand the driving factors behind loan defaults to improve its risk assessment and portfolio management.

# Steps followed

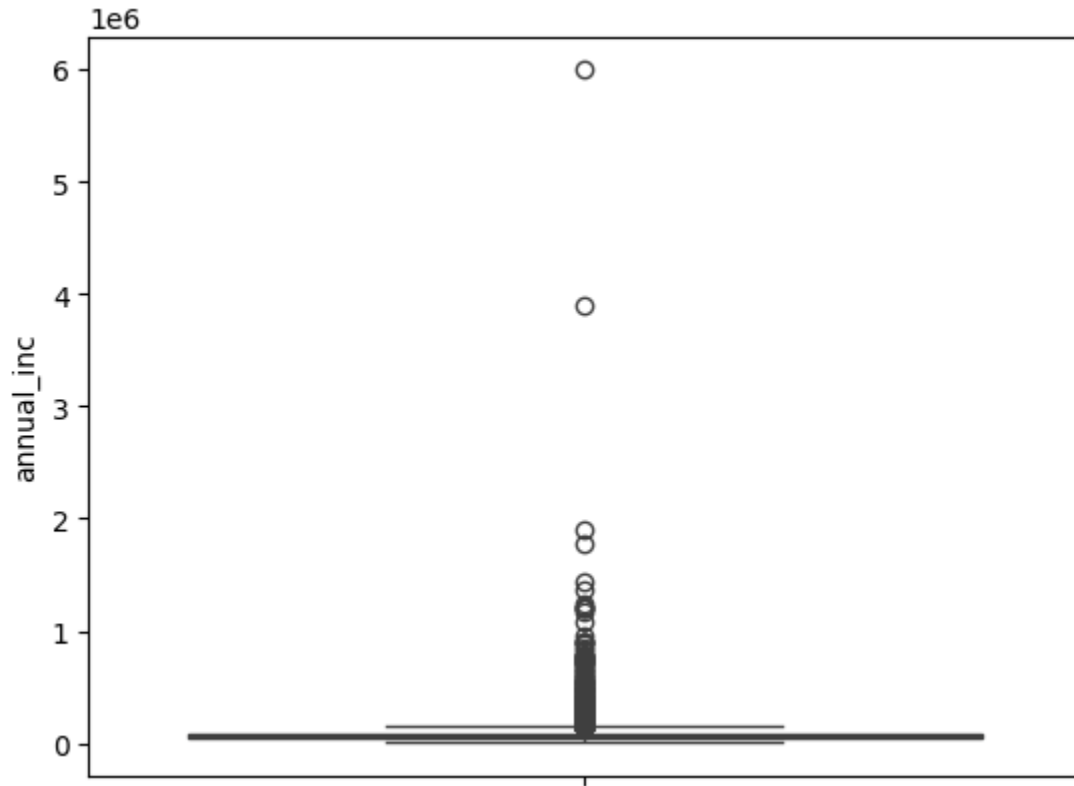
1. Loading the data
2. Data inspection
  - 2.1 `.head()`, `.info()`, `.describe()`, `.shape`
3. Data Cleaning
  - 3.1 Missing value treatment
  - 3.2 Checking for inconsistencies
  - 3.3 Outlier treatment using boxplots
4. Deriving new variables
  - 4.1 Binning of numerical variables
  - 4.2 Combining 2 or more columns using a formula
5. Exploratory data analysis
  - 5.1 Univariate analysis
    - Numerical: box plots, histograms, dist plots
    - Categorical: count plots, pie charts
  - 5.2 Segmented Univariate analysis
  - 5.3 Bivariate analysis
6. Multivariate analysis & heat map

# Percent of Fully Paid loan vs Percent of Charged off loan



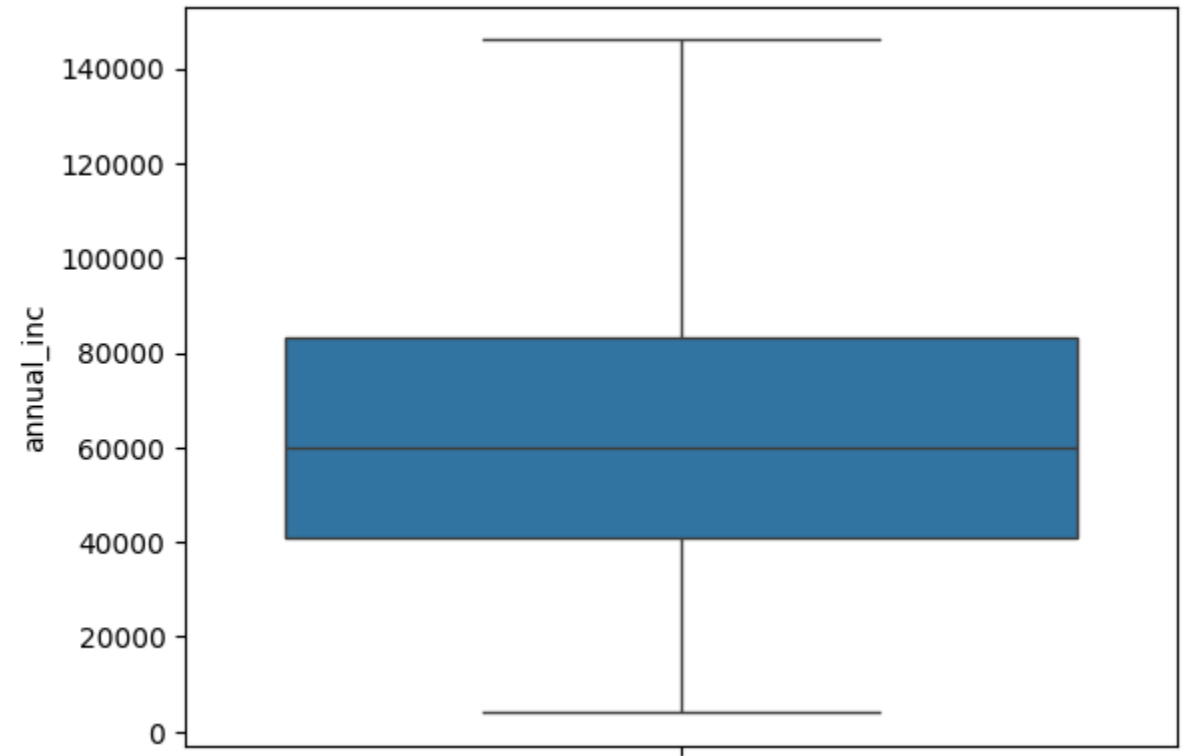
14.3% of the total loans are charged off

# Outliers treatment for Annual income



There are many outliers, hence we are fix this by-

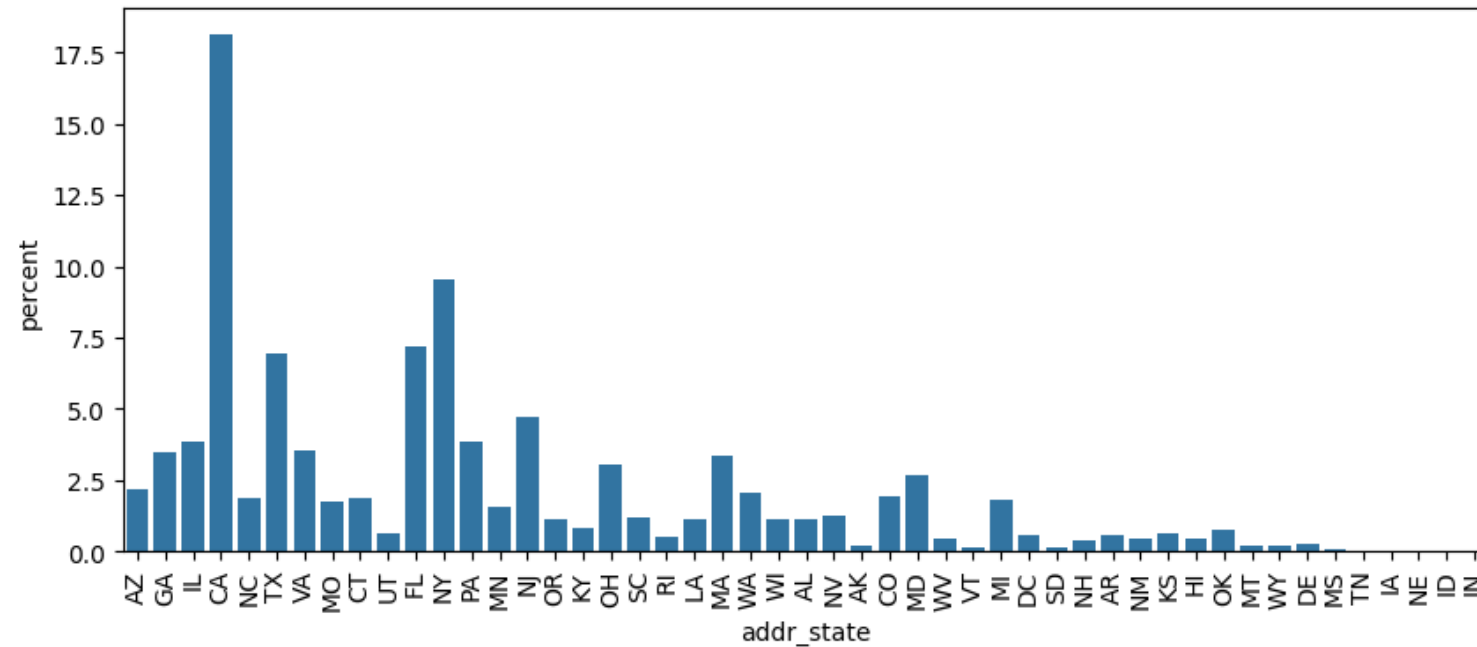
1. Removing the outliers
2. Capping the outliers



Here we have capped the outliers, hence we don't have outliers.

min	4000.000000
25%	41004.000000
50%	60000.000000
75%	83000.000000
max	145994.000000

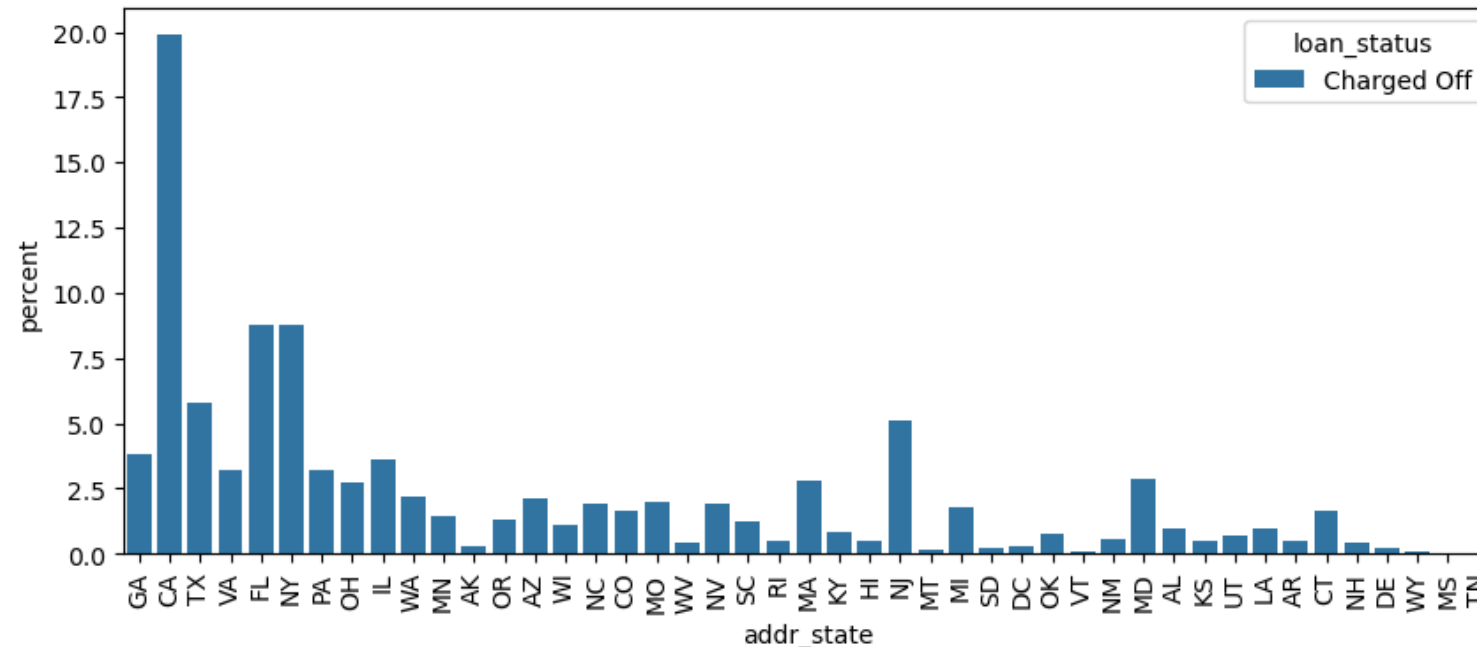
# Univariate : Percent of loan by each State



Percentage of total loan across states –

- CA state has about 18% of the total loans
- NY state has about 10% of the total loans
- FY state has about 7.5% of the total loans

Note - Considering both Fully pad & Charged off data

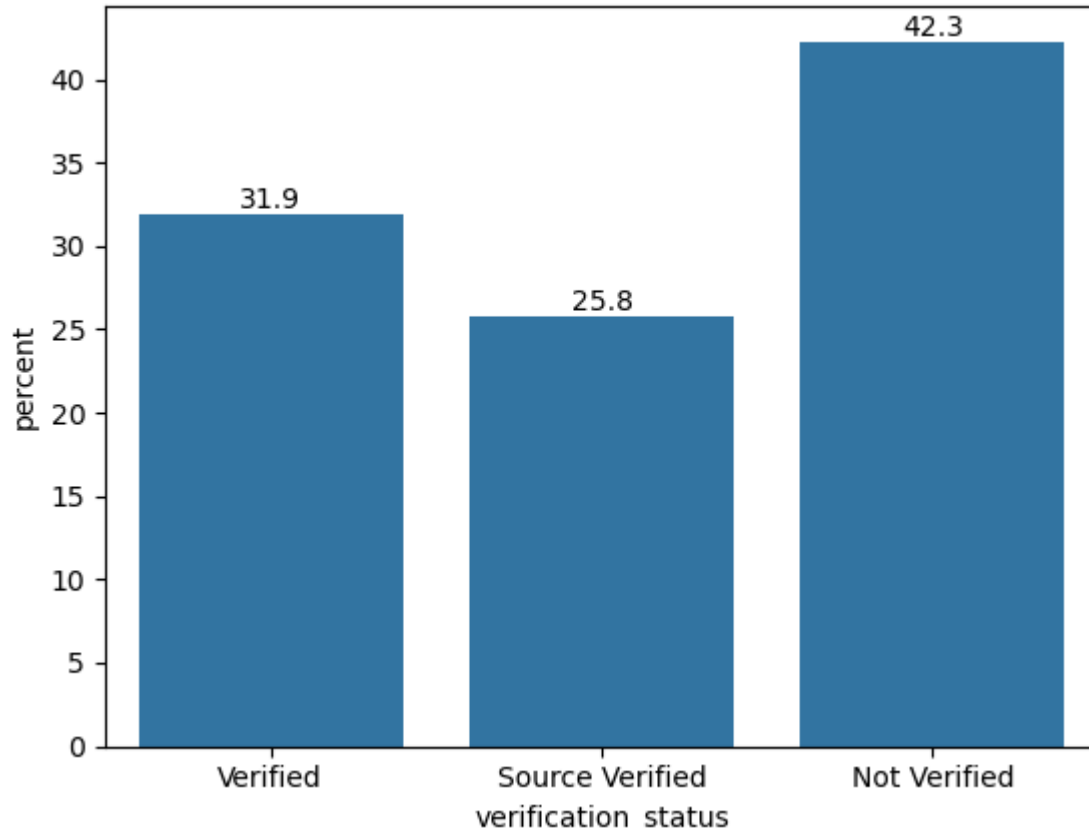


Percentage of total loan across states –

- CA has about 20% of the charged off loans
- FL & NY states have about 8.8% of the charged of loans

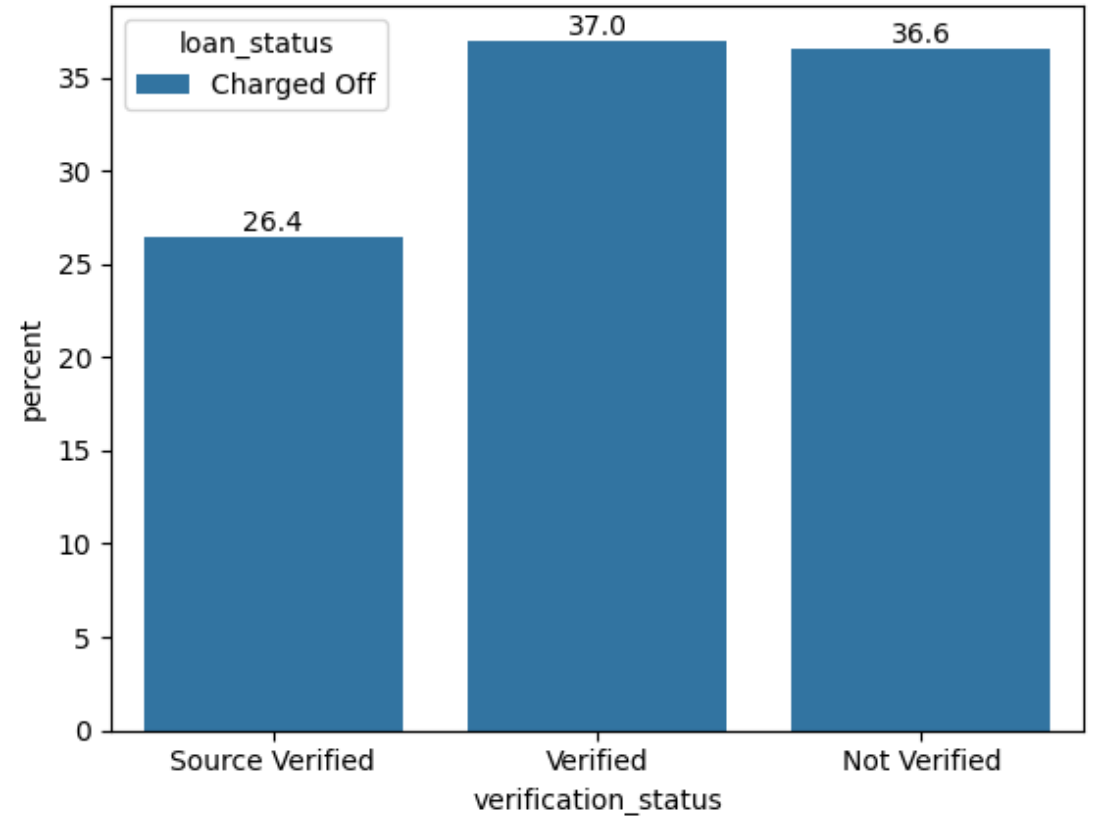
Note - Considering only Charged off data

# Univariate : Percent of loan by Verification status



Considering both Fully pad & Charged off data

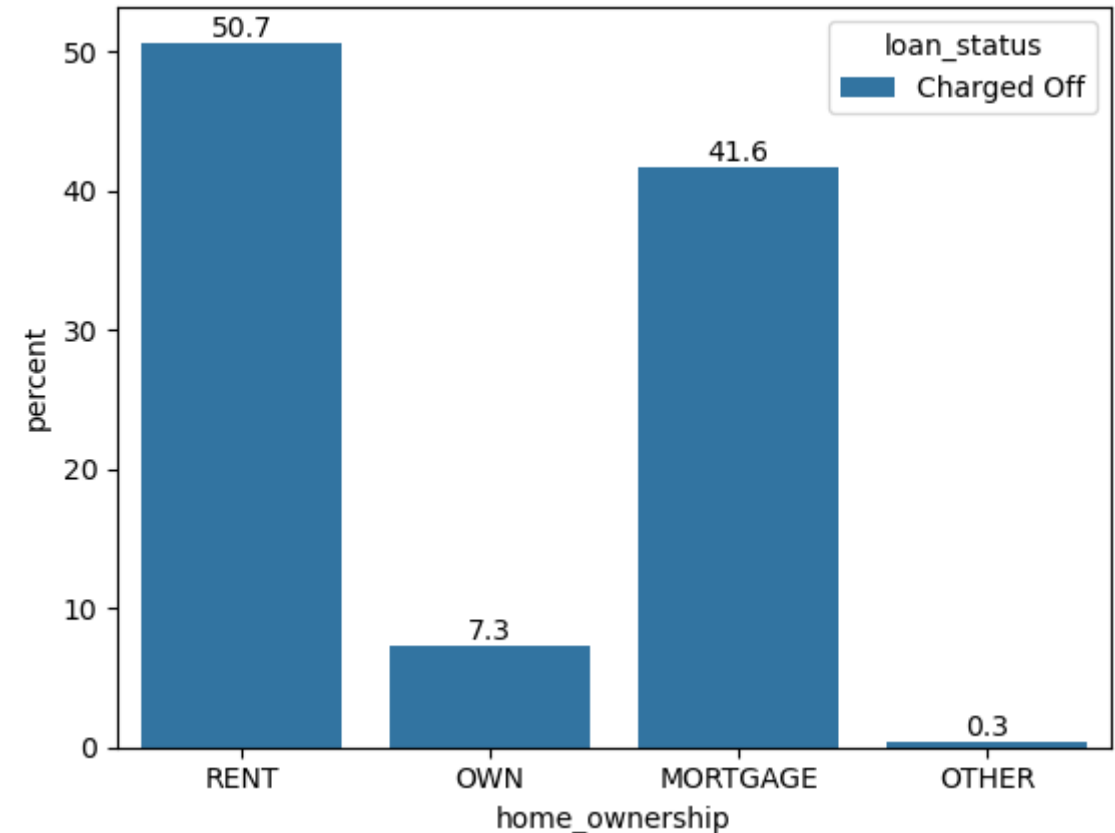
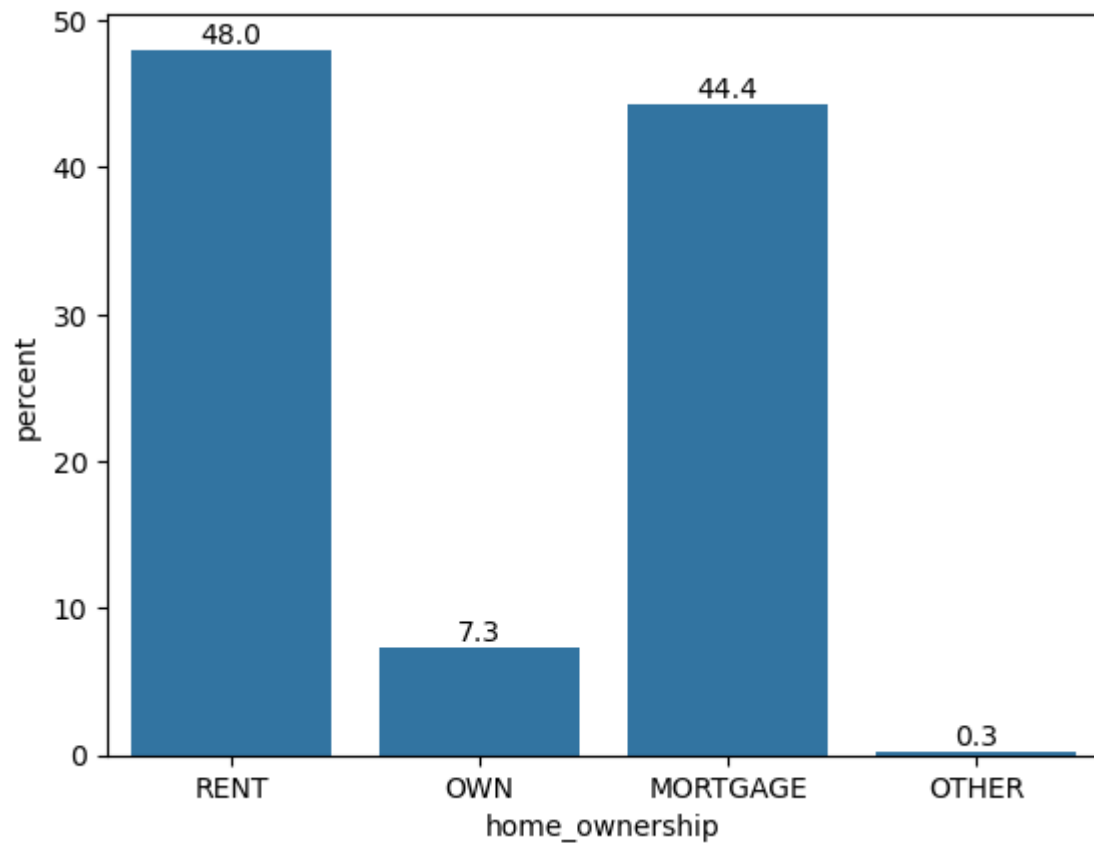
- 42.3% of the loan is given to Not Verified
- 31.9% of the loan is given to verified persons
- 25.7% of the loan is given to Source verified



Considering only Charged off data

- 37.0% of the Charged off loan are Verified
- 36.5% of the Charged off loan are Not Verified
- 25.7% of the Charged off loan are Source Verified

# Univariate : Percent of loan by Home Ownership



Considering both Fully pad & Charged off data

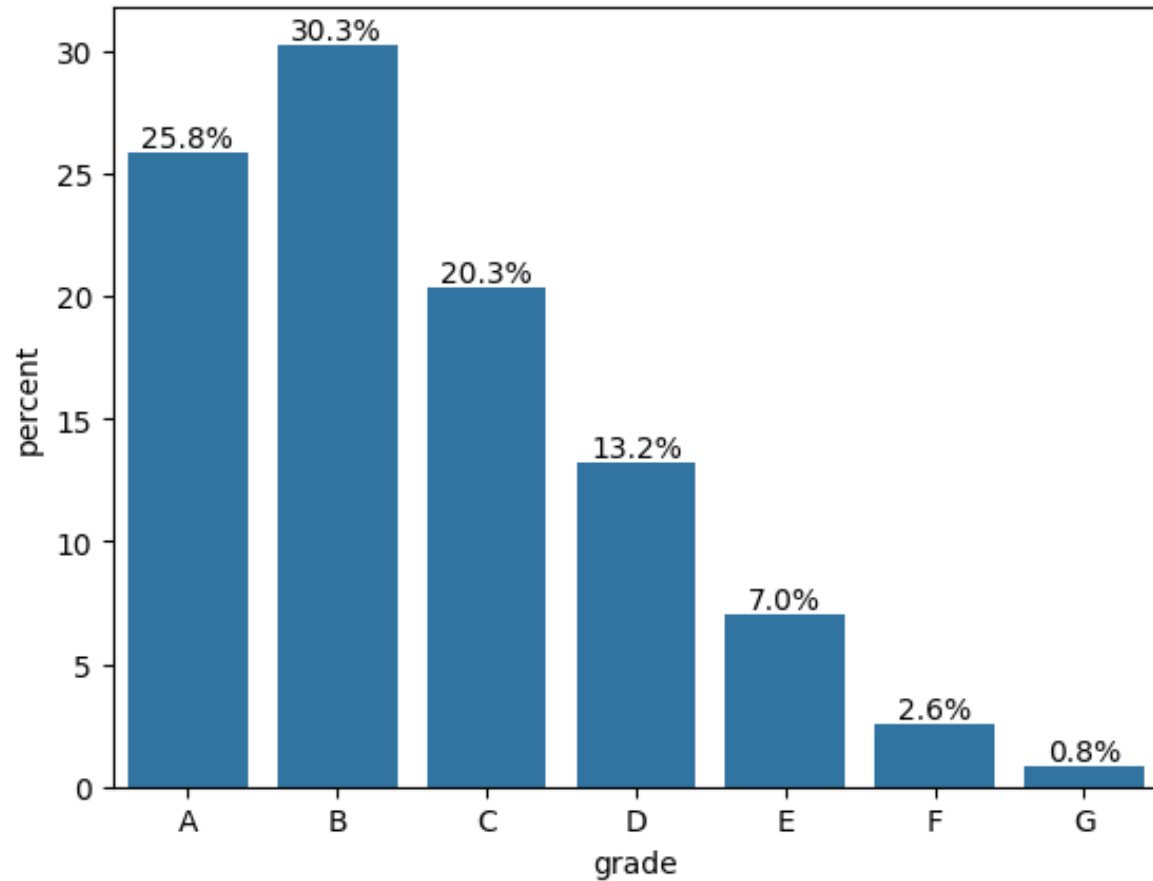
- 48.0% of the loan is given to Rented home
- 44.3% of the loan is given to Mortgage home
- 7.3% of the loan is given to Home owner

Considering only Charged data

- 50.7% of the Charged off loan is given to Rented home
- 41.6% of the Charged off loan is given to Mortgage home
- 7.3% of the Charged off loan is given to owner

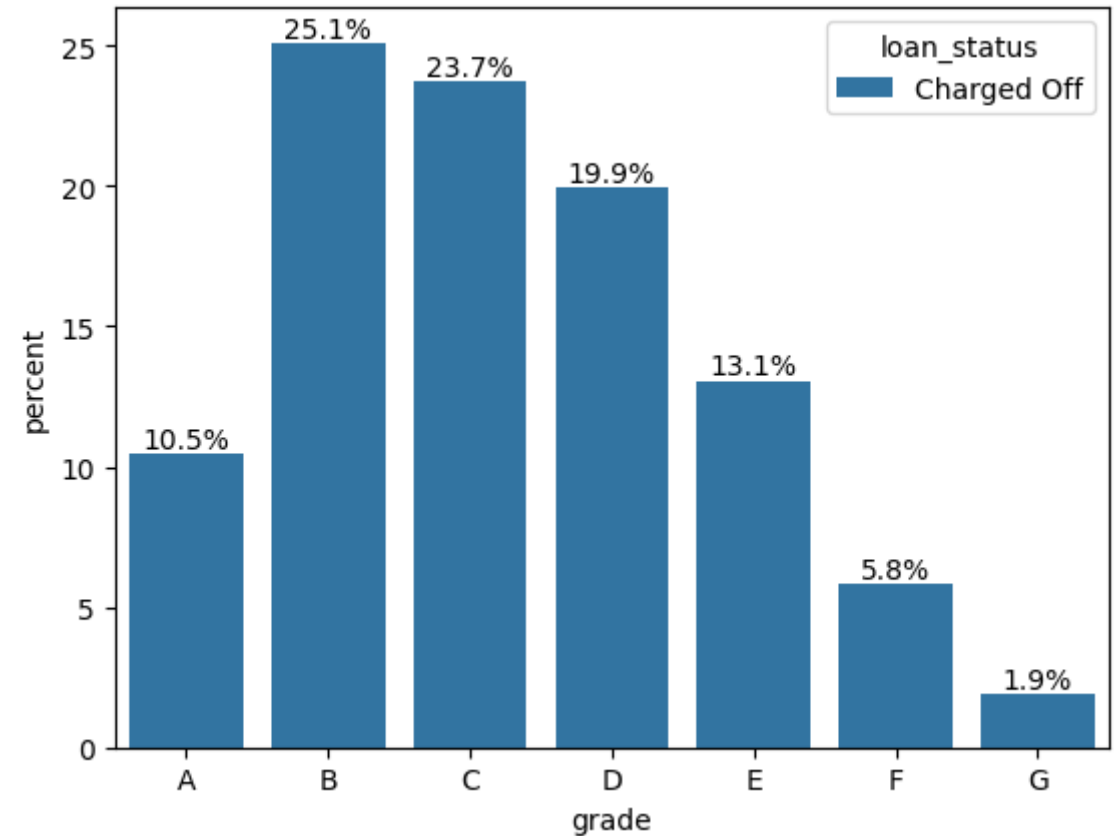


# Univariate : Percent of loan by Grade



Considering both Fully pad & Charged off data

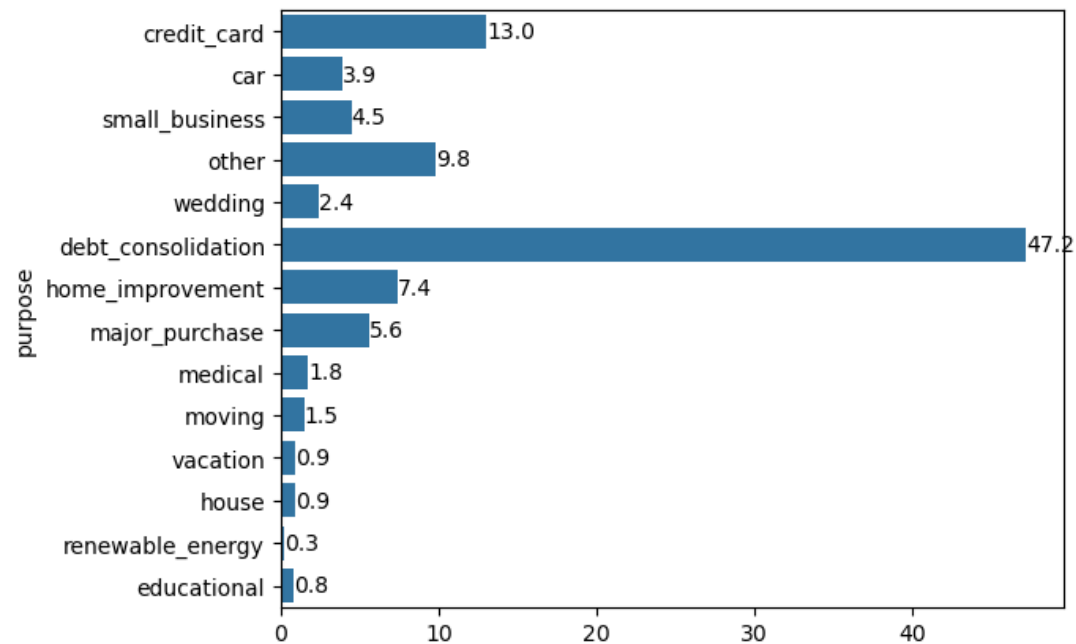
- About 30.2% of loan fall under grade B
- About 25.8% of loan fall under grade A



Considering only Charged off data

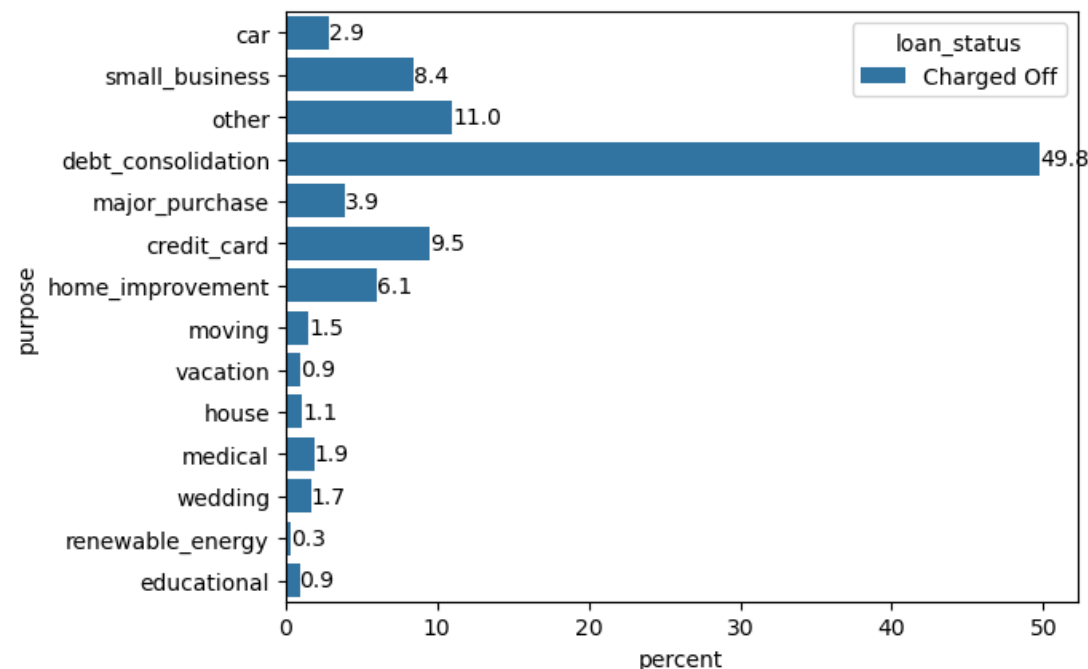
- About 25.1% of loan fall under grade B
- About 23.7% of loan fall under grade C

# Univariate : Percent of loan by purpose



Considering both Fully pad & Charged off data

- About 47.2% of loan is for Dept Consolidation
- About 13.0% of loan is for Credit card

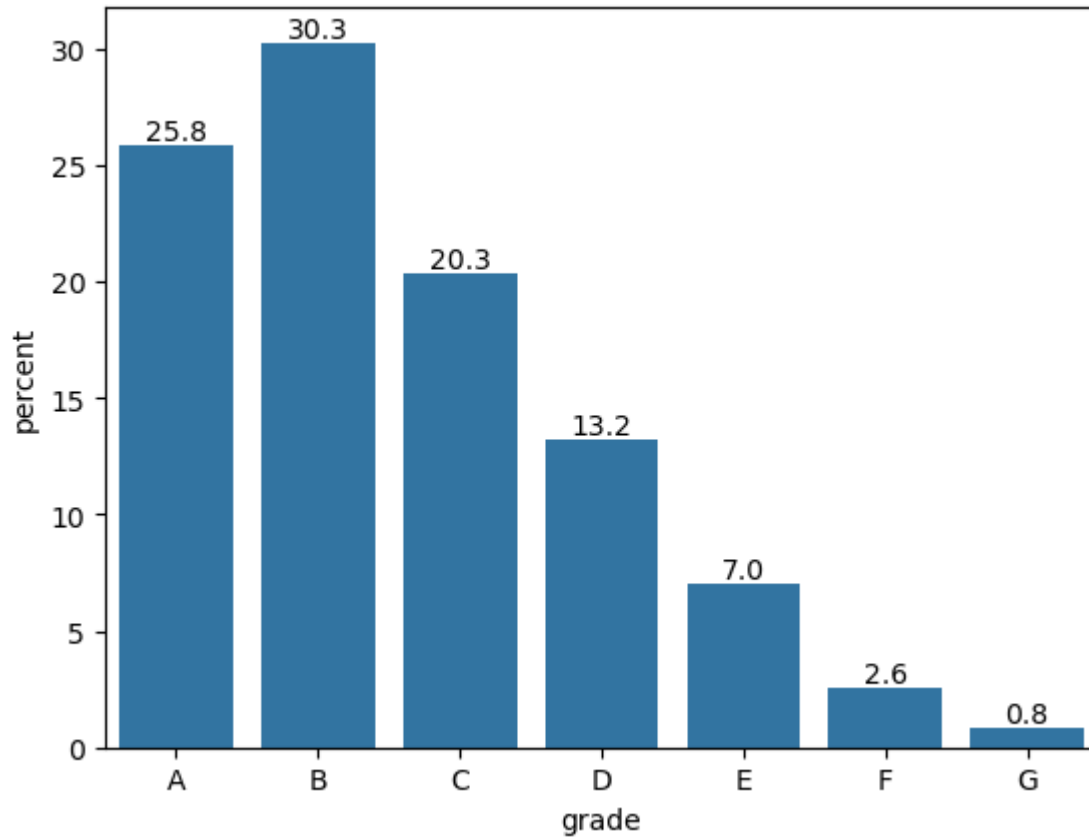


Considering only Charged off data

- About 49.7% of Charged off loan is for Dept Consolidation
- About 11.0% of Charged off loan is for Credit card

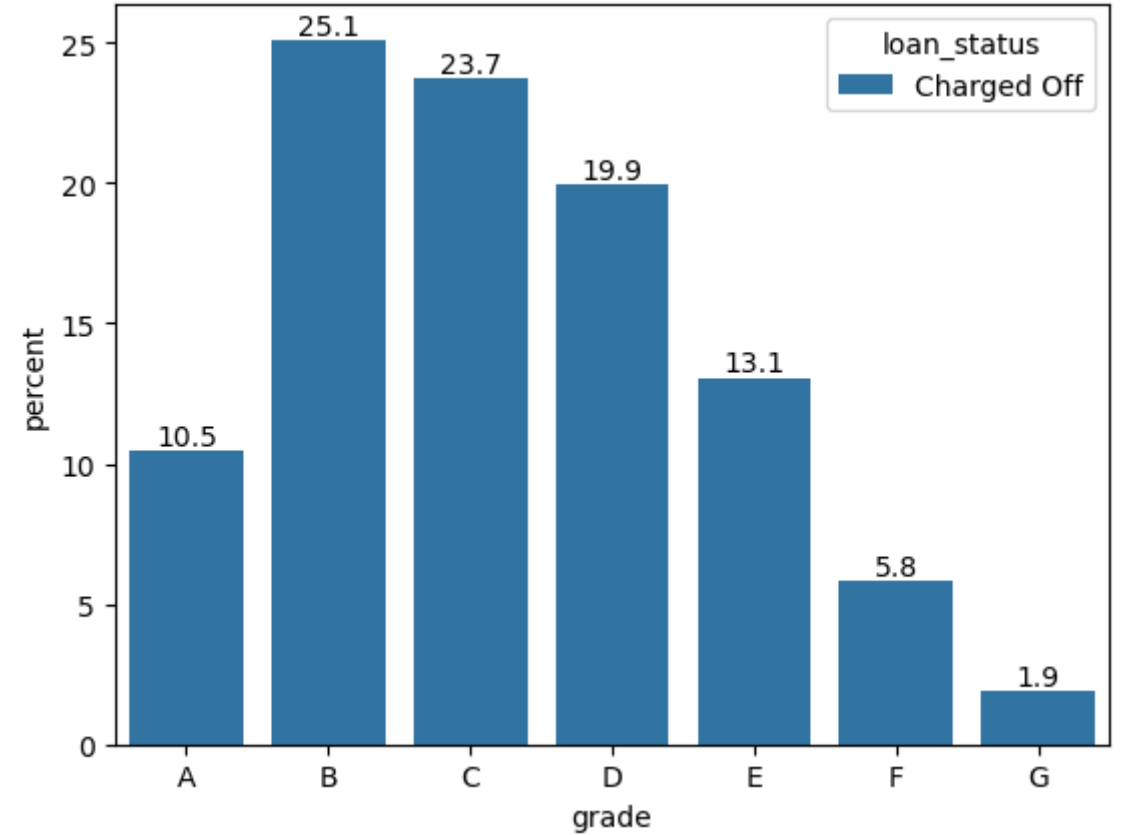
Loans given for Dept Consolidation & other are charged off most

# Univariate : Percent of loan by Employment length



Considering both Fully paid & Charged off data

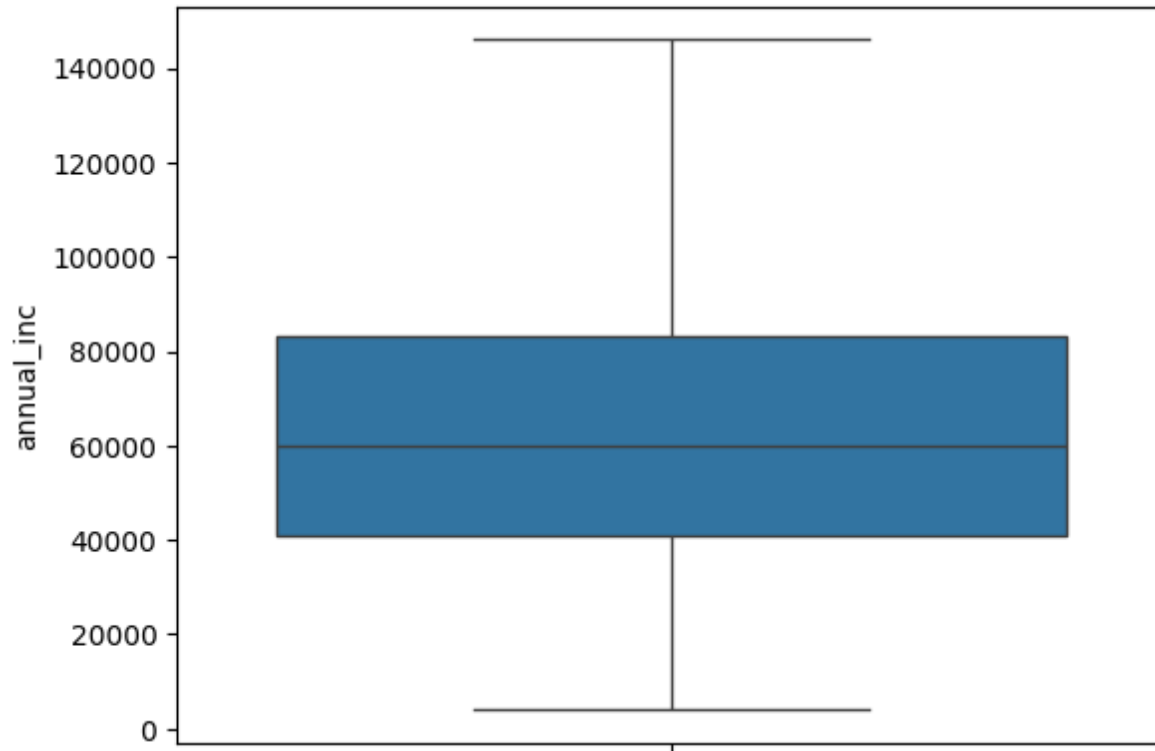
- About 30.2% of loan fall under grade B
- About 25.8% of loan fall under grade A



Considering only Charged off data

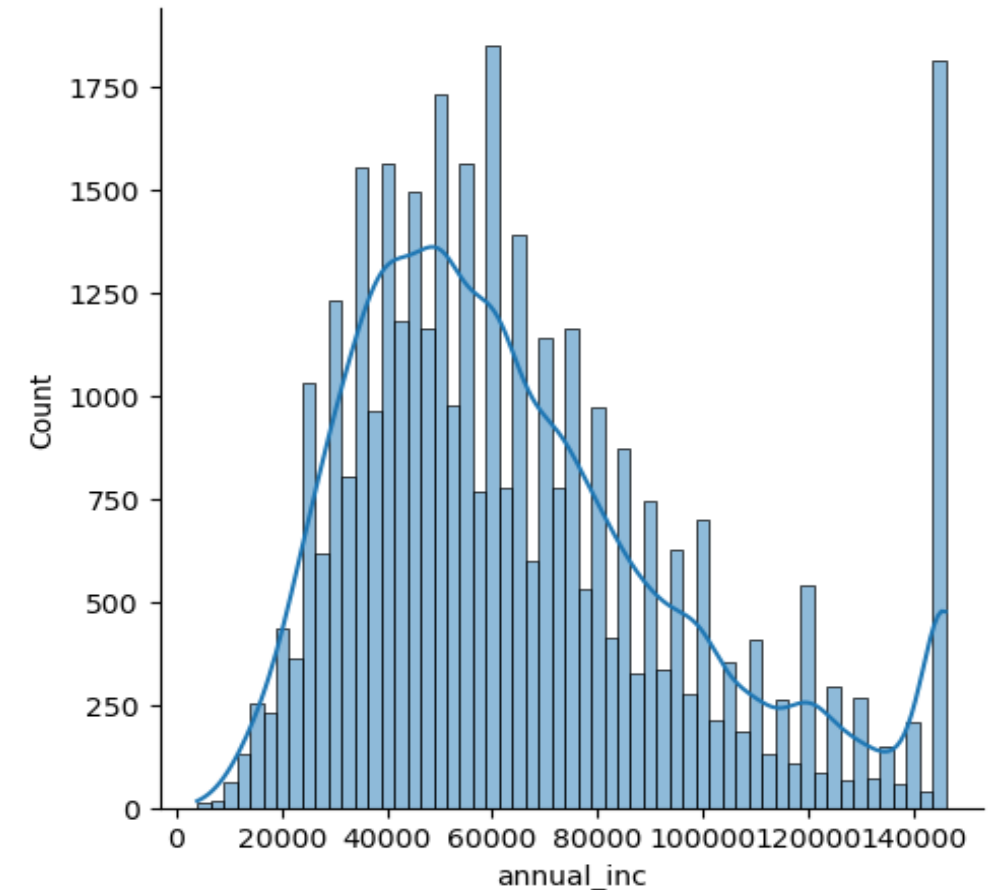
- About 25.1% of loan fall under grade B
- About 23.7% of loan fall under grade C

# Univariate : Analysis of annual income



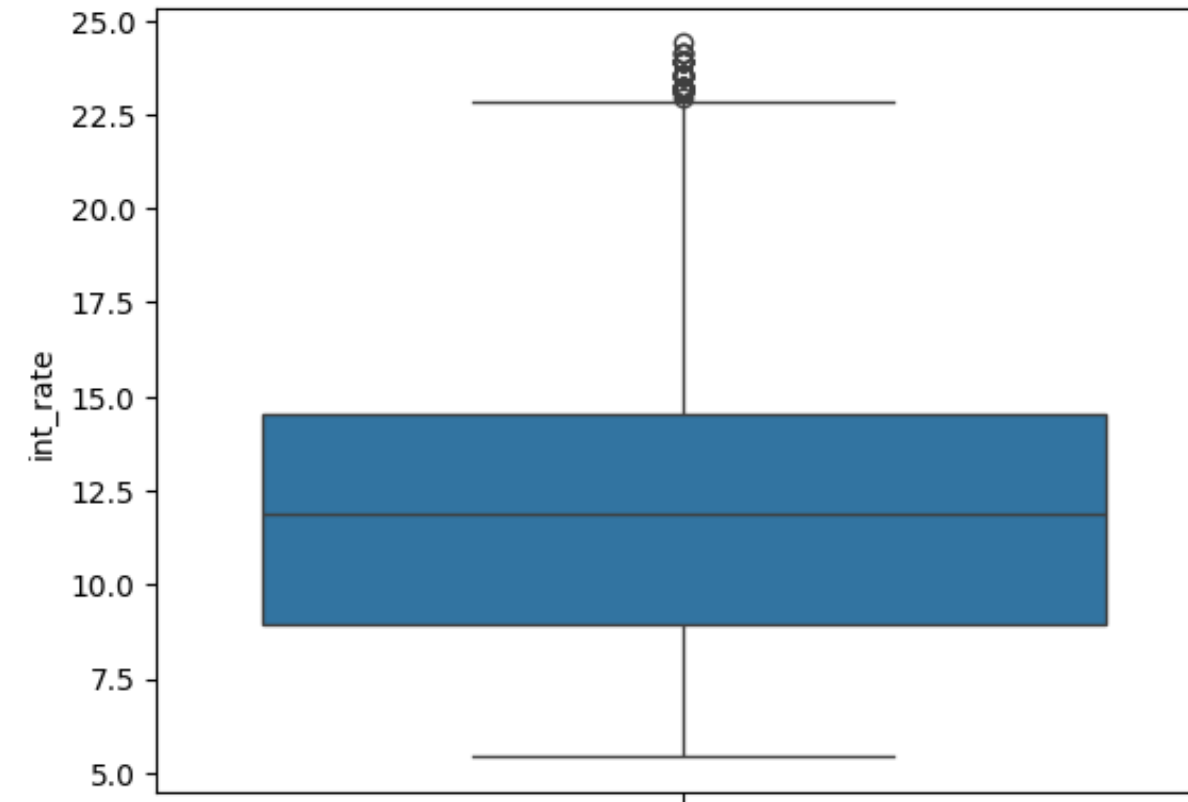
Considering both Fully paid & Charged off data

- about 50% of the customers have income between 41000 to 83000 (IQR)
- Average customers have income about 60000 (mean)



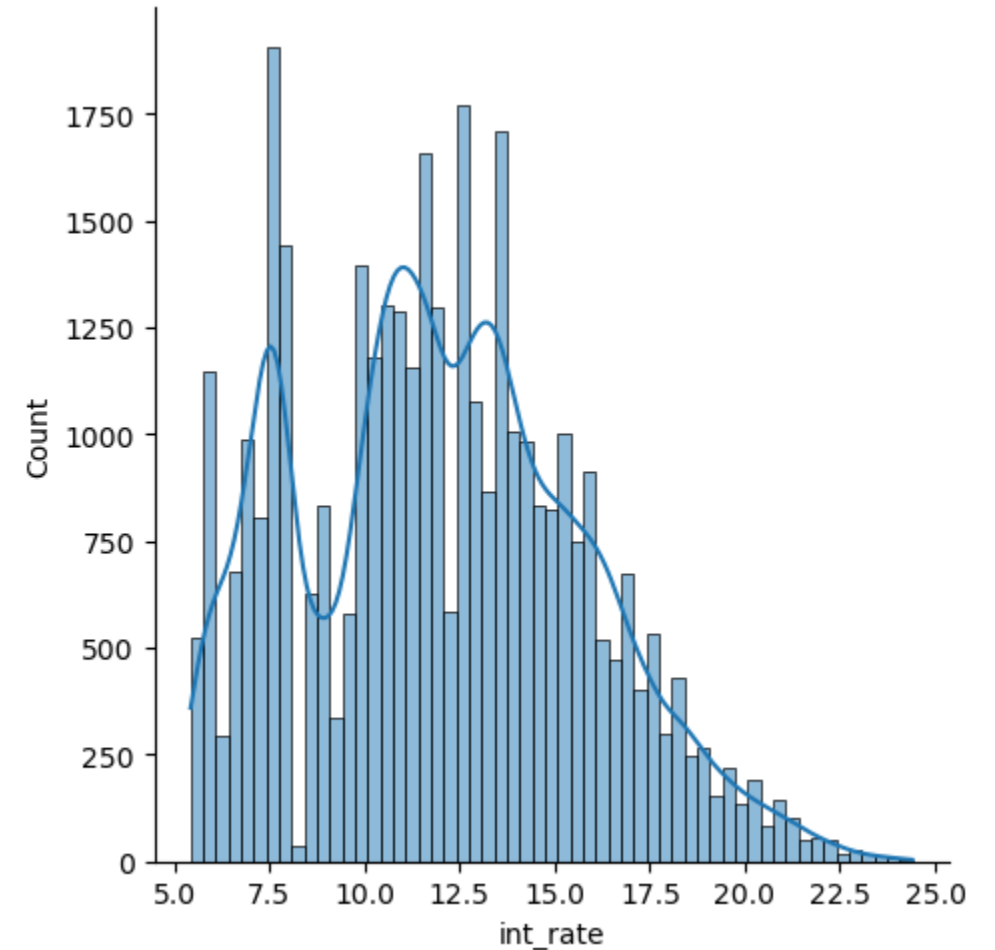
- the mode value of income is 60000
- we can see count increasing at the end because we have capped the outliers

# Univariate : Analysis of Interest rate



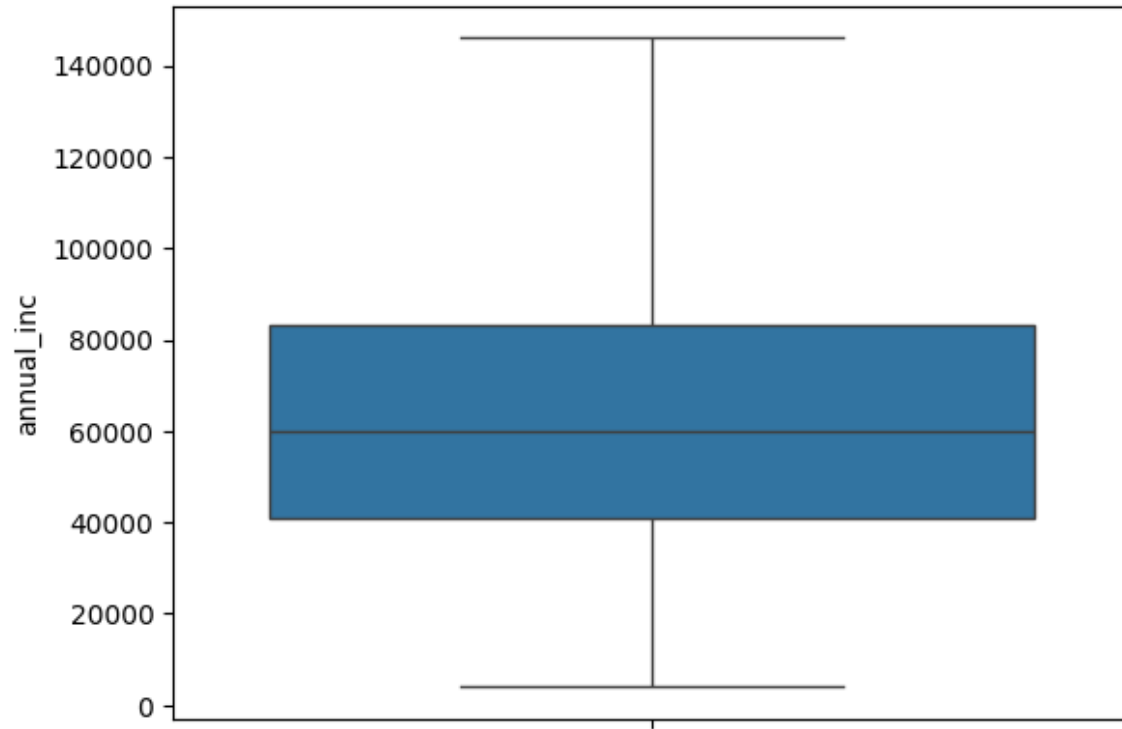
Considering both Fully pad & Charged off data

- about 50% of the loan have interest rate between 8.5% to 15% (IQR)
- Average loan have interest rate of 11.5%



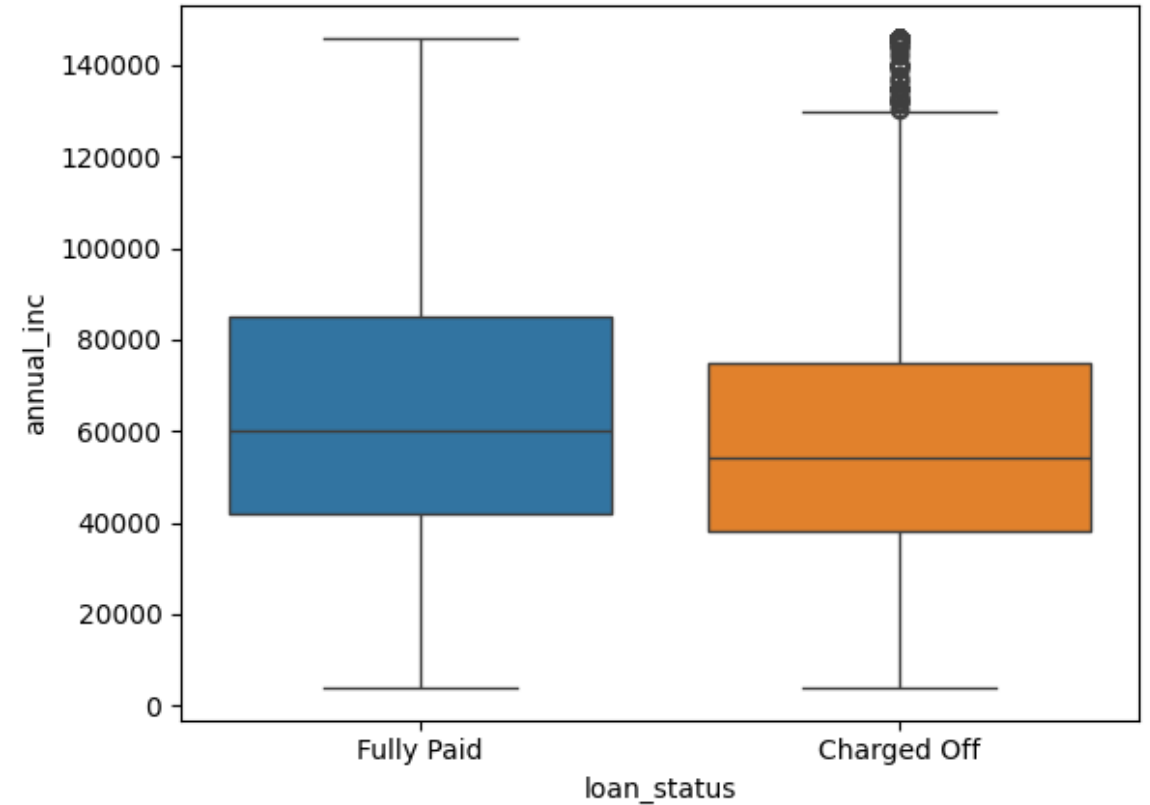
- most of the loan are given at interest rate of 7.5%

# Bi-variate : Analysis of annual income



Considering both Fully pad & Charged off data

- about 50% of the customers have income between 41000 to 83000 (IQR)
- Average customers have income about 60000 (mean)

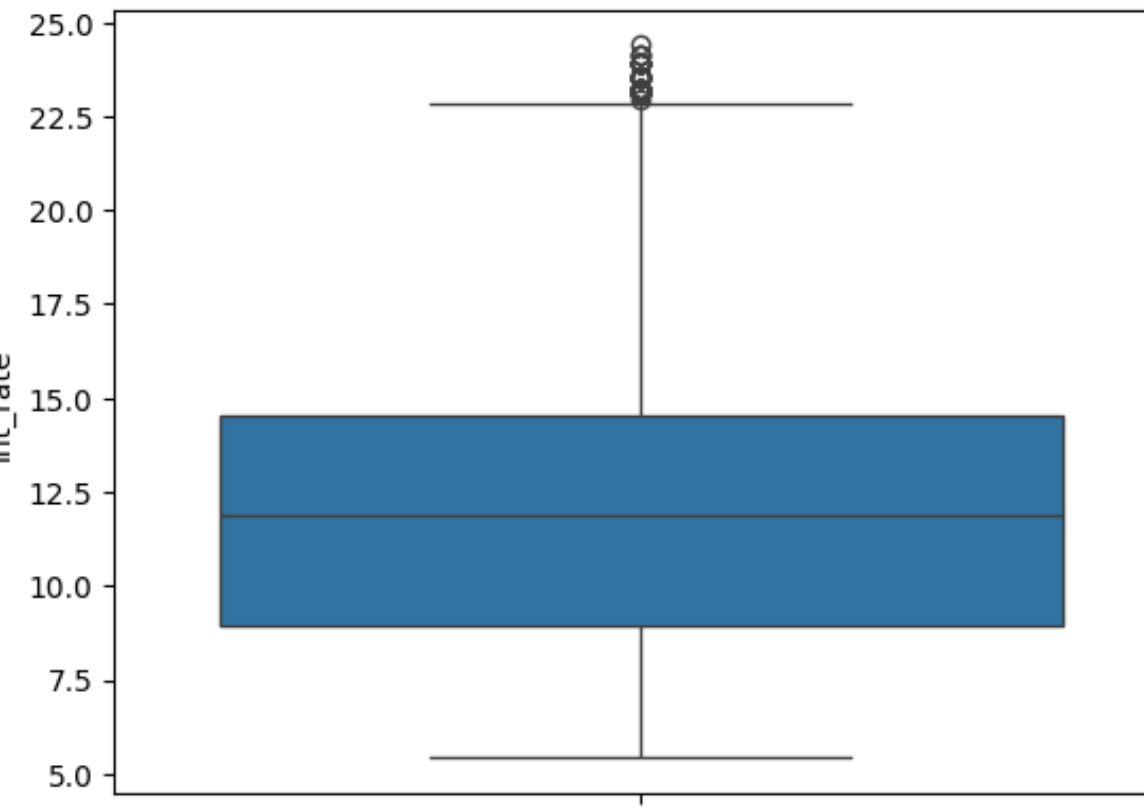


Considering only Charged off data

- about 50% of the customers have income between 40000 to 75000 (IQR)
- Average customers have income below 60000 (mean)

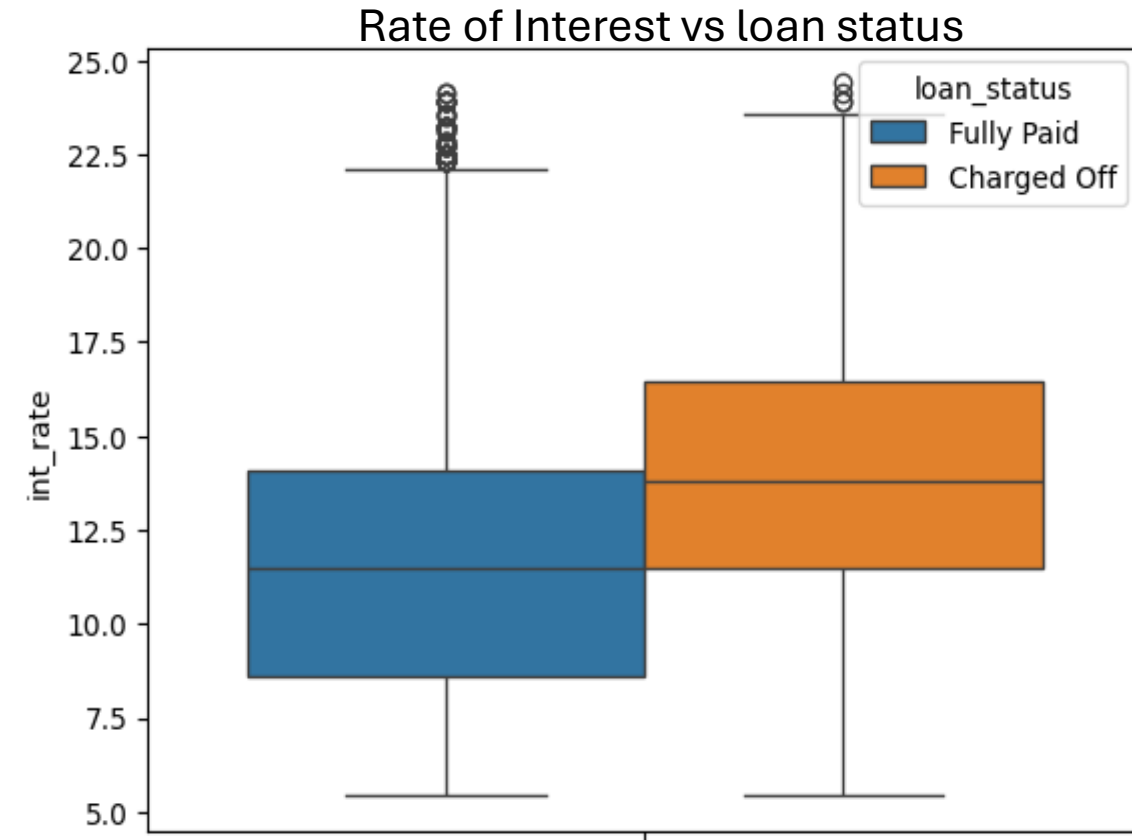
Persons with low annual income have charged off

# Bi-variate : Analysis of Interest rate



Considering both Fully pad & Charged off data

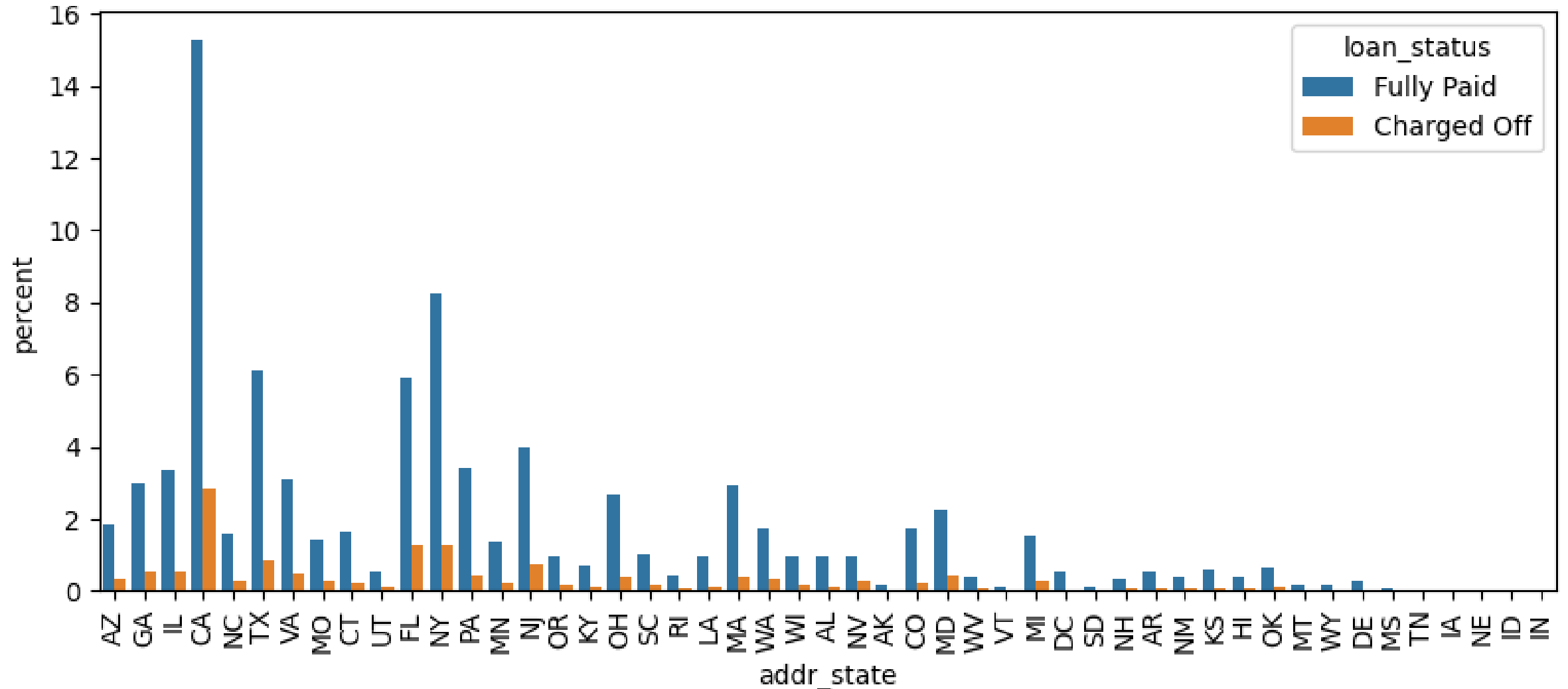
- about 50% of the loan have interest rate between 8.5% to 15% (IQR)
- Average loan have interest rate of 11.5%



Considering Fully pad & Charged off data separately

- Charged off loan have higher interest rate

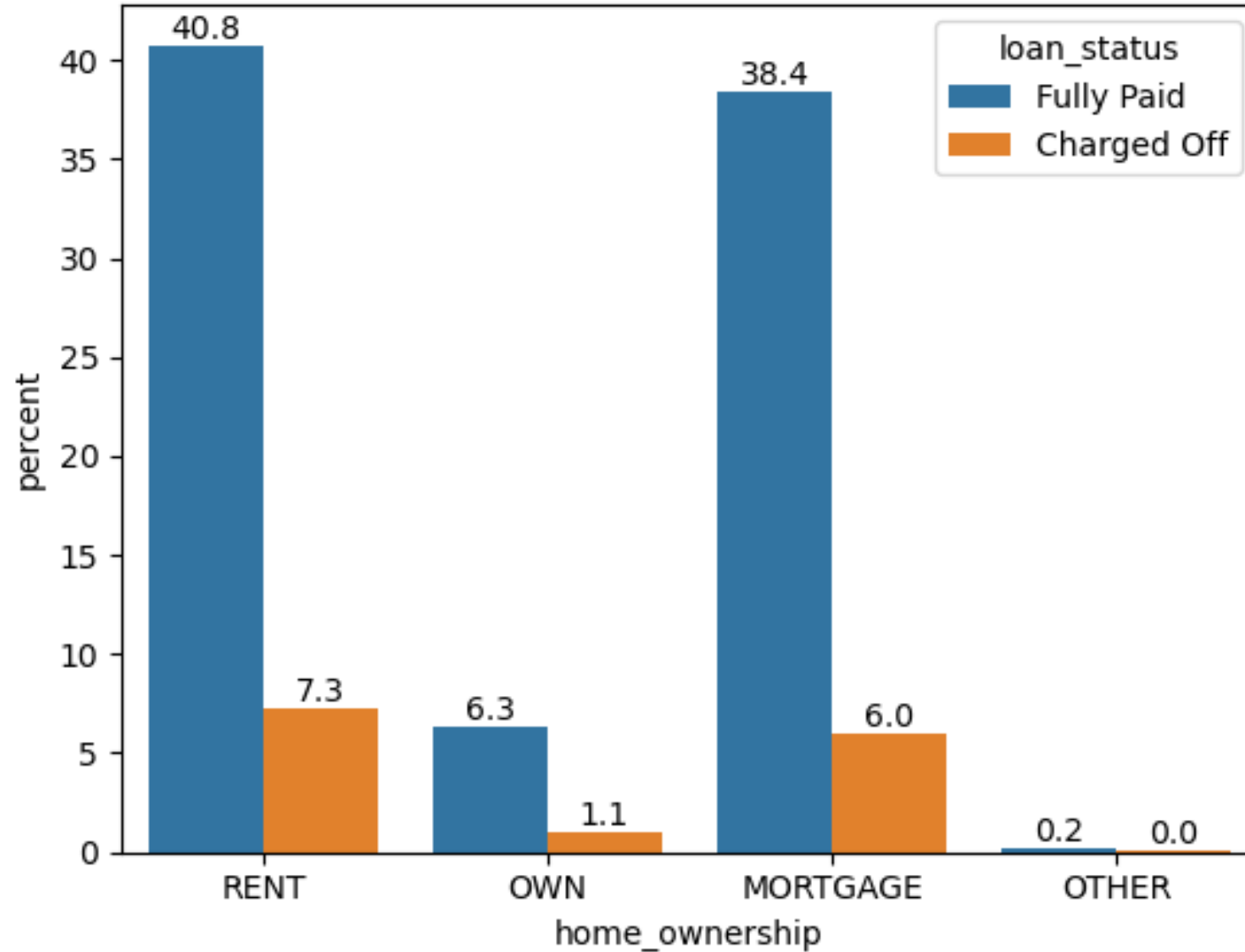
# Bi-variate : Percent of loan across State vs Loan status



Both Fully paid & Charged off curve follow same trend no much inference can be made

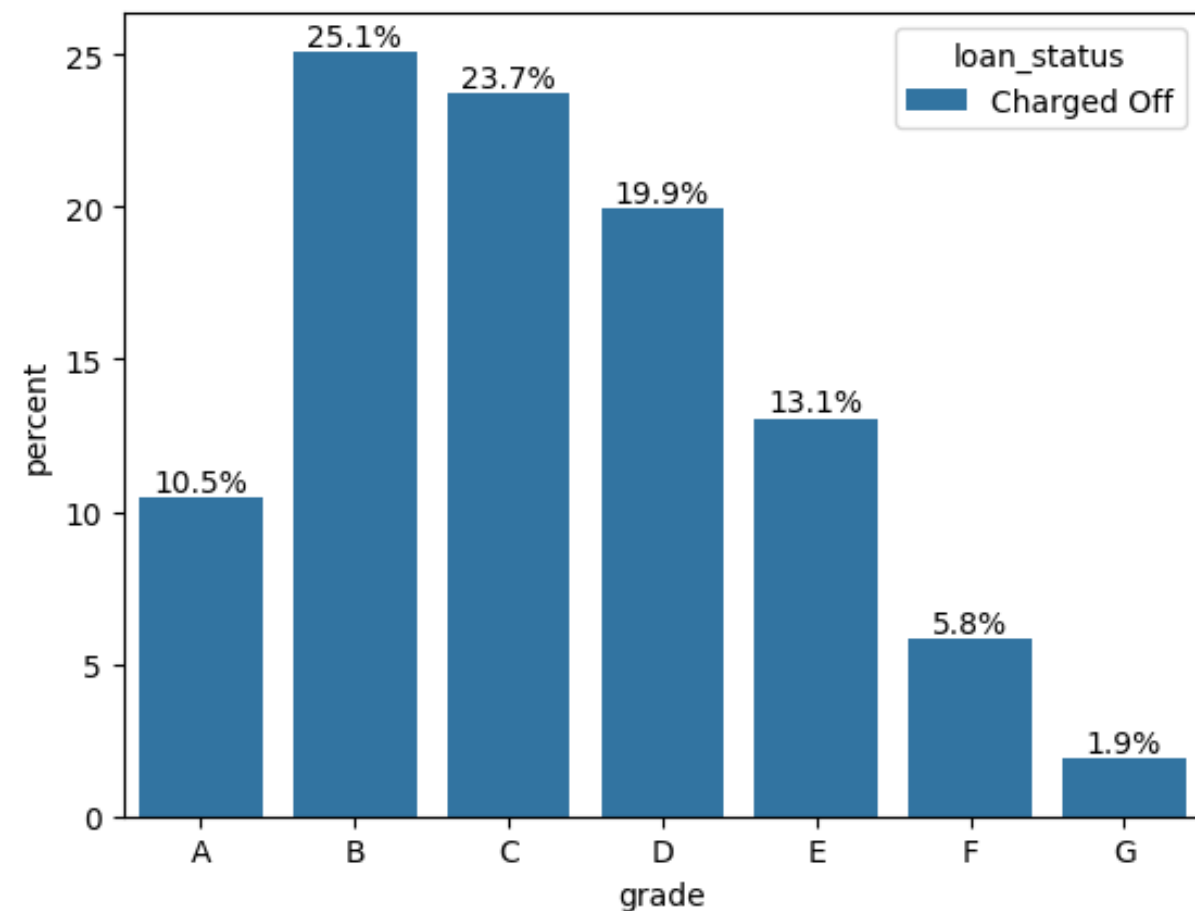
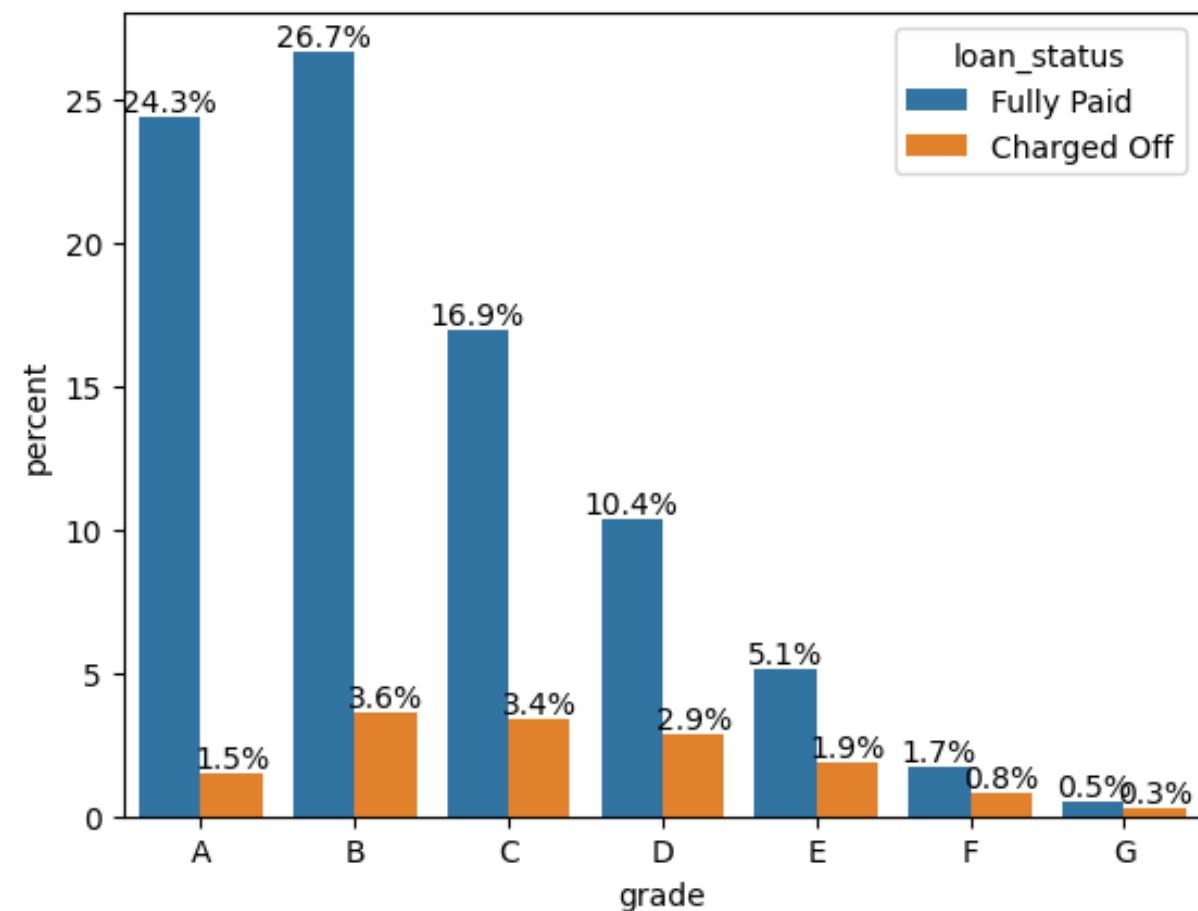


# Bi-variate : Percent of loan across home ownership vs Loan status



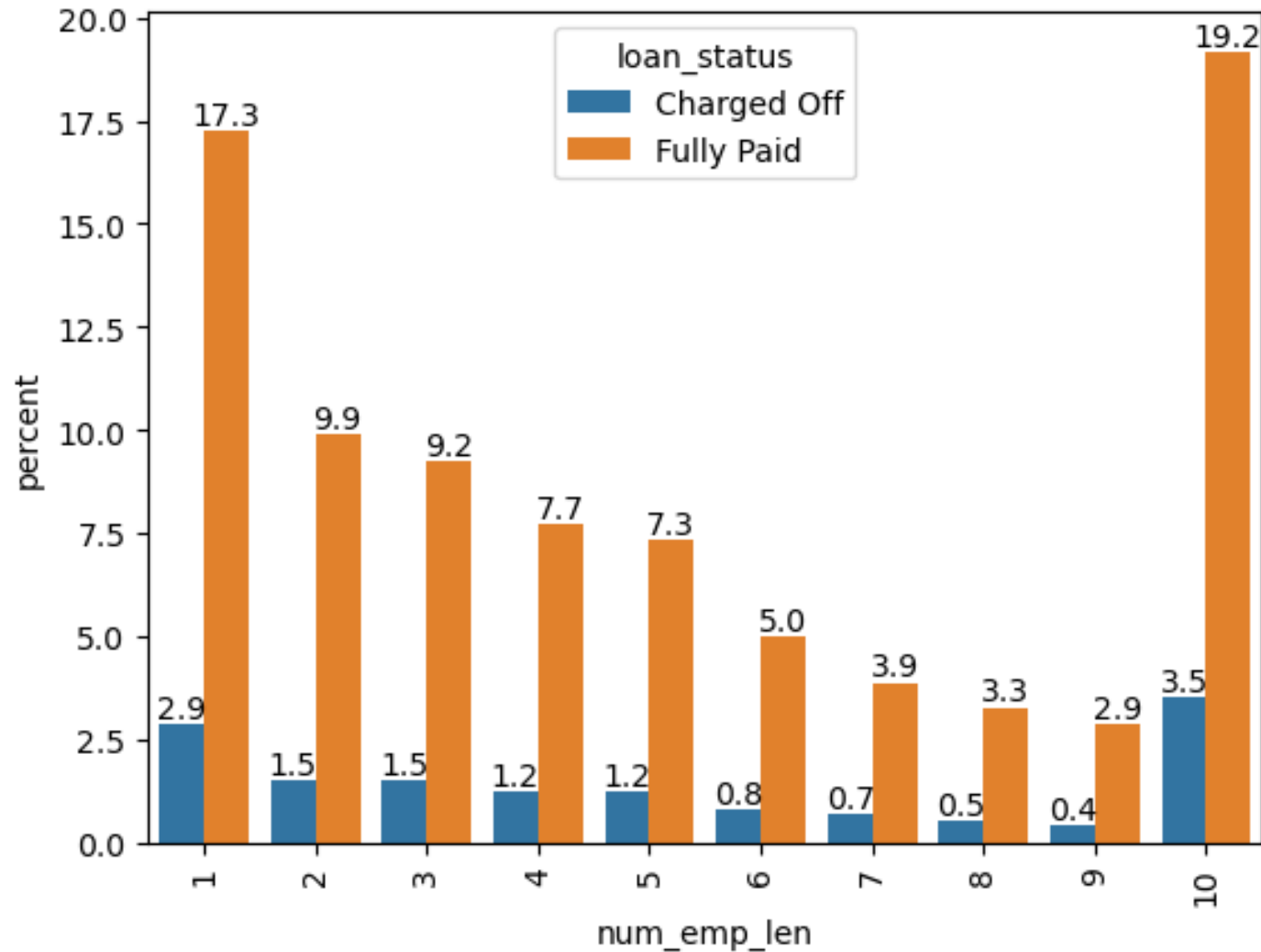
Both Fully paid & Charged off curve follow same trend no much inference can be made

# Bi-variate : Percent of loan across grades vs Loan status



Highest Fully paid grade is in the order B > A > C  
Highest Charged off grade is in the order B > C > D

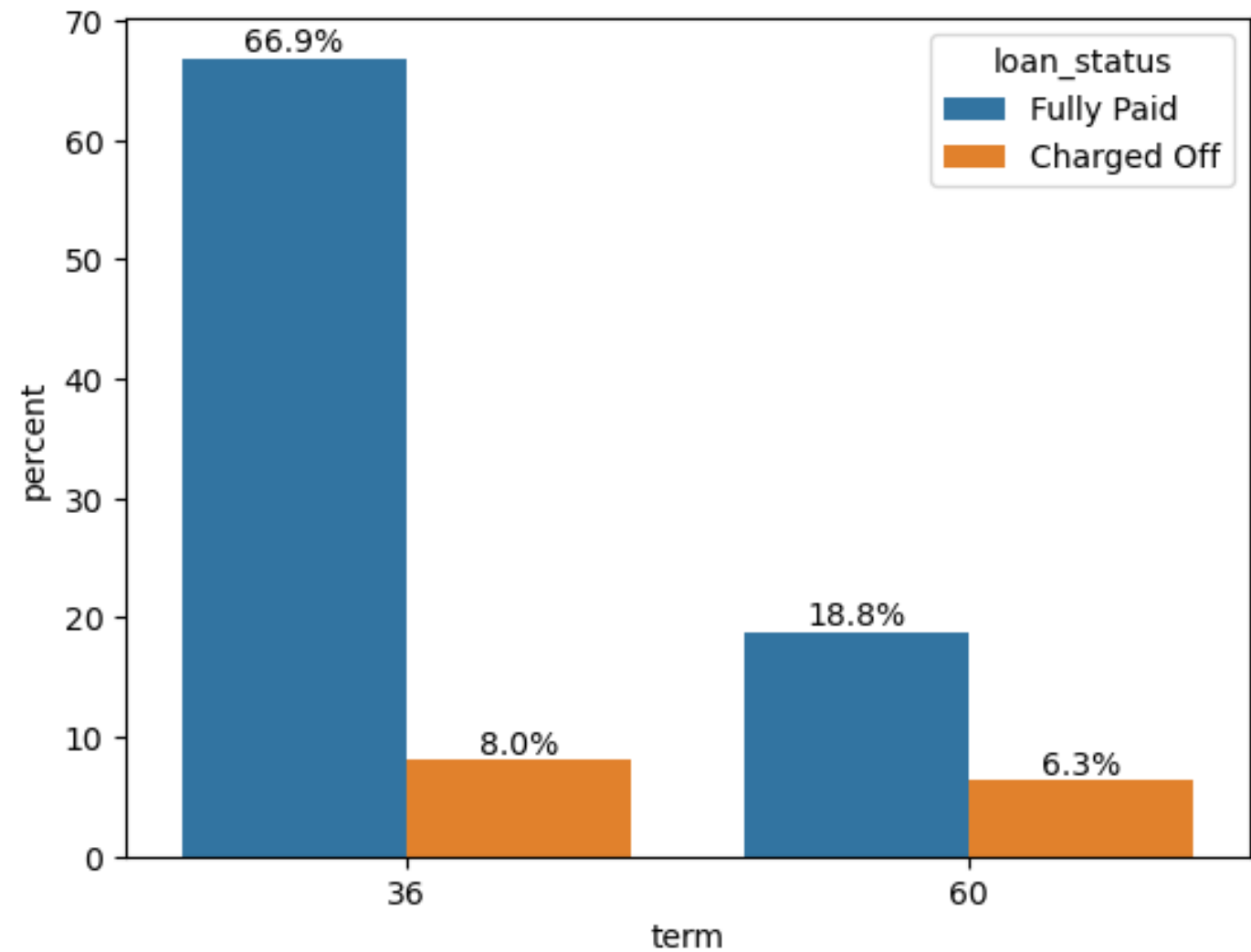
# Bi-variate : Percent of loan across employment length vs Loan status



Fully paid loan are high when the loan is given to 10+ years & <1 years experience

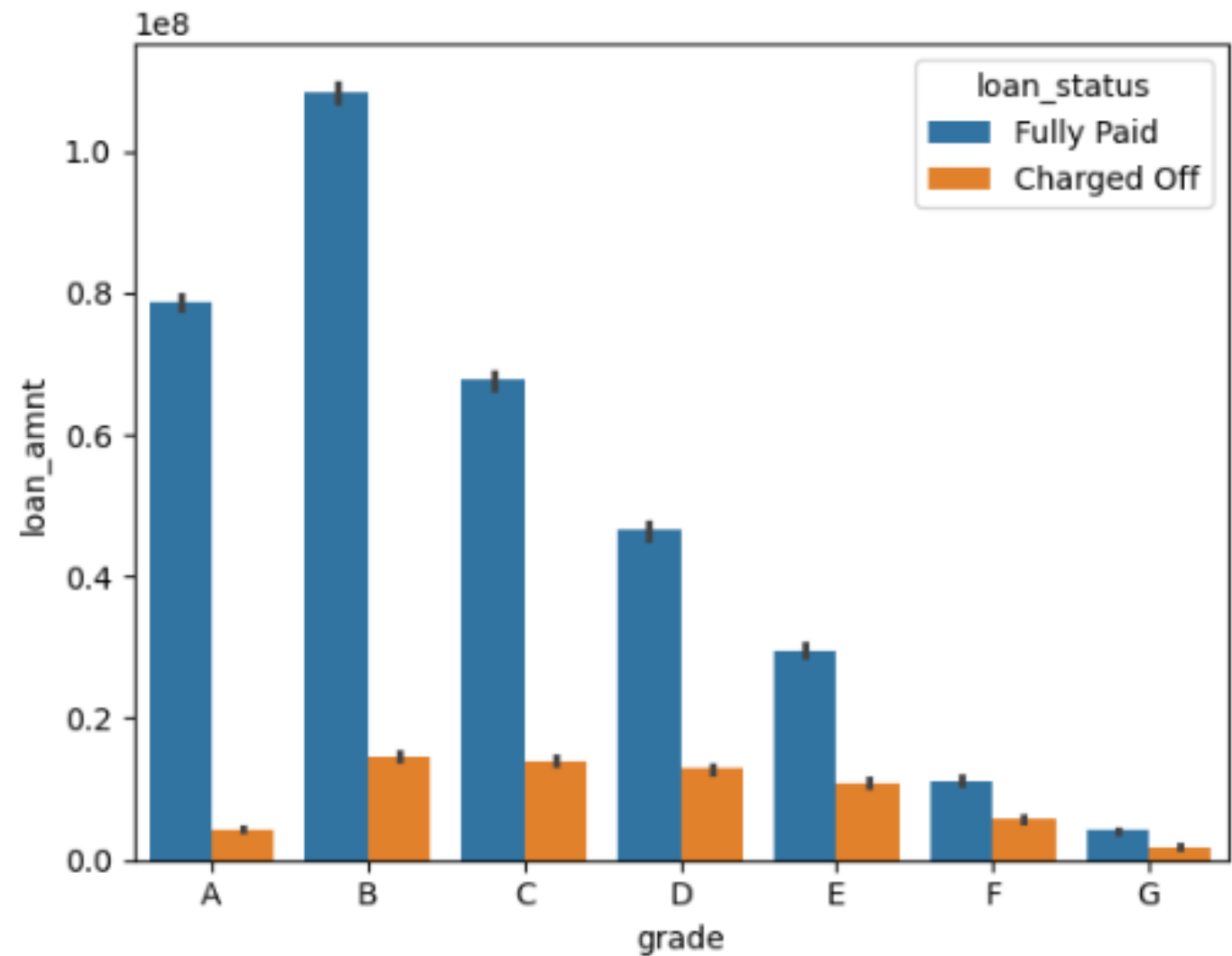
Similarly Charged off loan are high when the loan is given to 10+ years & <1 years experience

# Bi-variate : Percent of loan across term vs Loan status



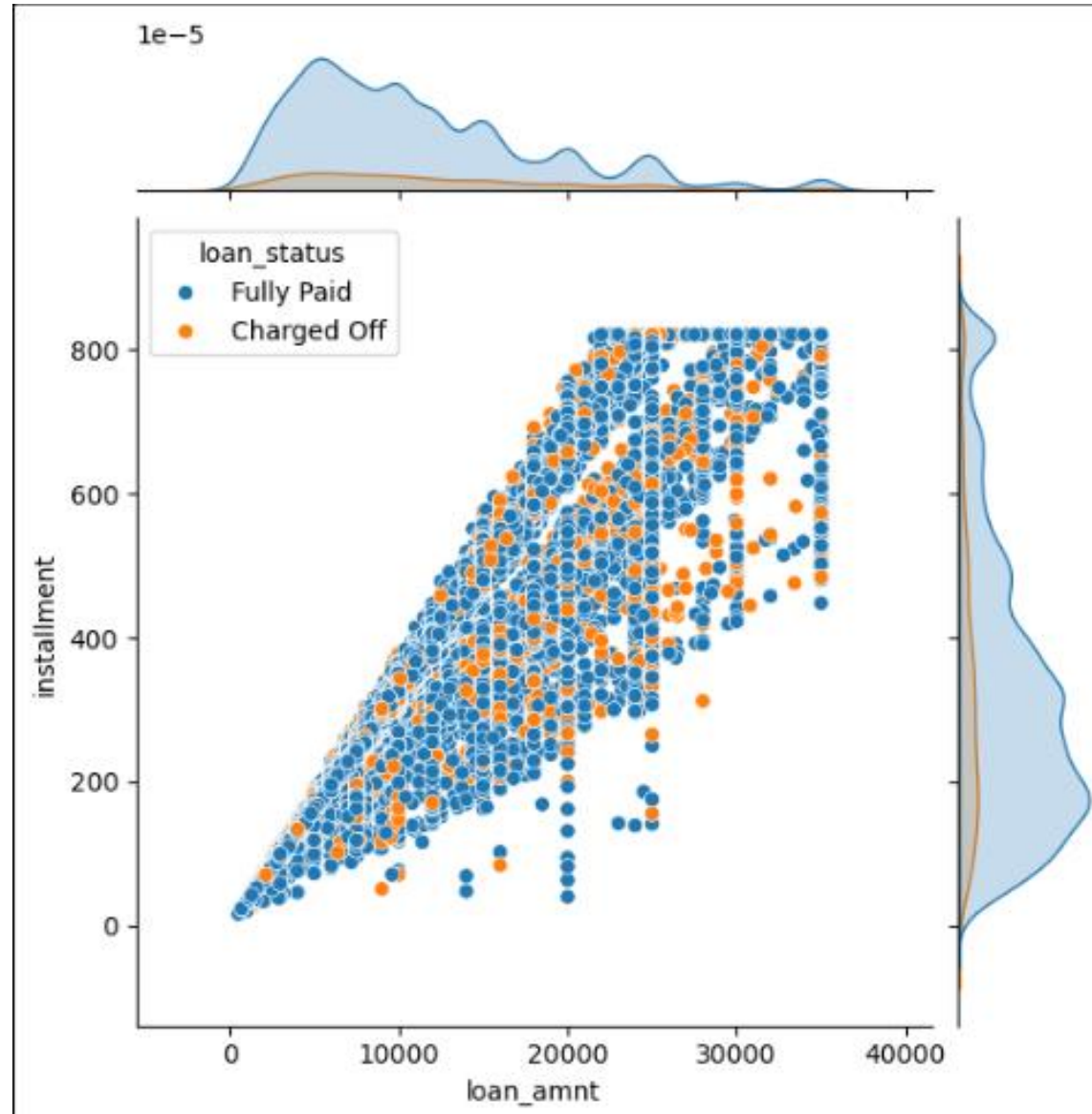
The loan given for 36 months has better chance of fully paid status (66.9% of the total loan given for short time is fully paid )

# Bi-variate : Loan for grade against loan status



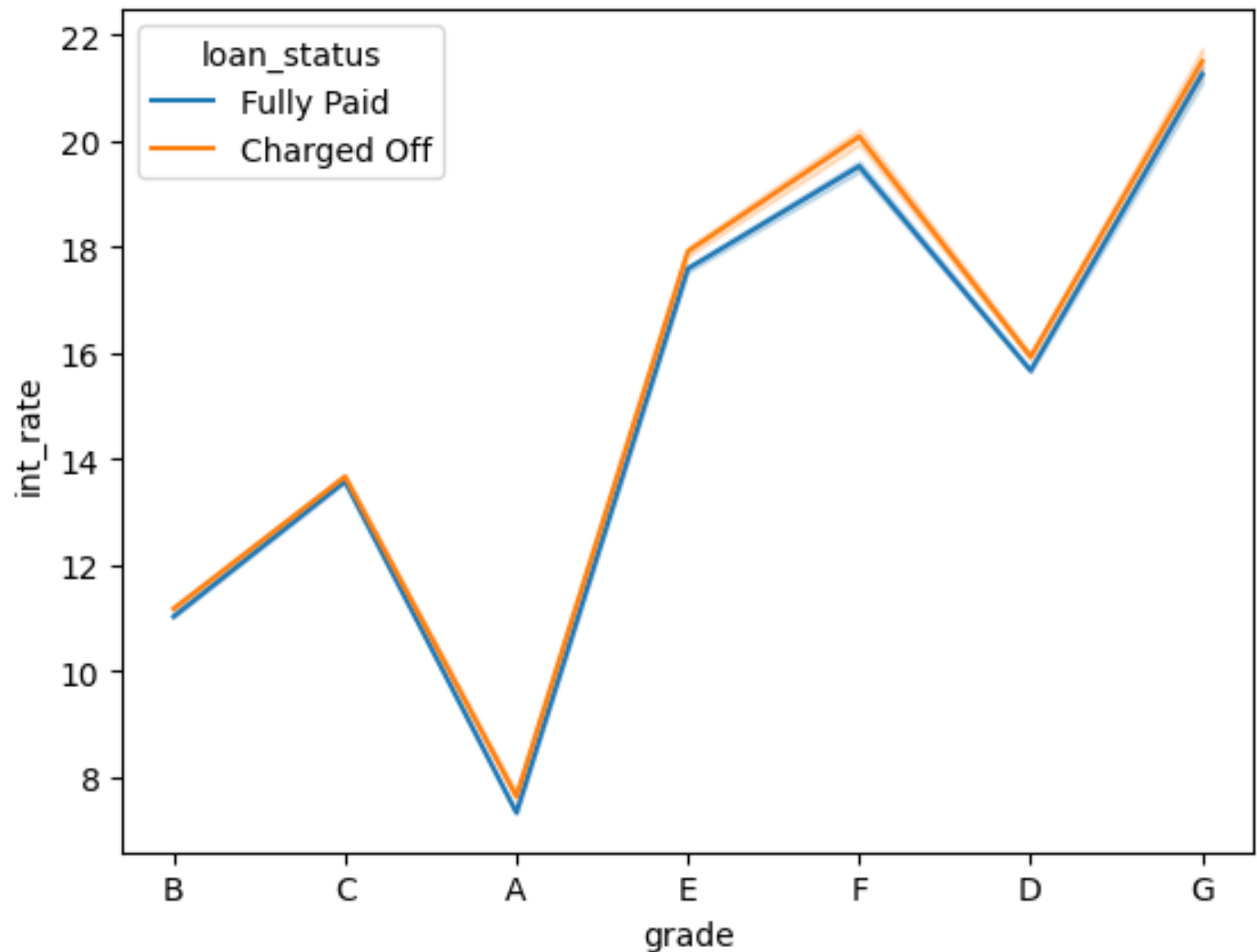
Highest amount of loan is given through grade B

# Multivariate : Loan amount vs Installment vs Loan status



As the loan amount is increasing the installment is also increasing

# Multivariate : Rate of interest vs grade vs Loan status



As the grade is increasing form A to G the rate of interest is also increasing

# Heat Map

## Corelation values -

- Charged 1.000000
- int\_rate 0.214772
- term 0.178445
- loan\_amnt 0.060773
- funded\_amnt 0.057563
- pub\_rec 0.049963
- pub\_rec\_bankruptcies 0.045625
- dti 0.042570
- installment 0.027164
- num\_emp\_len 0.016036
- annual\_inc -0.065247





# Conclusion -

## **Loan status Charged-off has correlation as follows:**

- 1) Strongly positive correlation** - `int_rate > term`
- 2) Medium positive correlation** - `loan_amnt > funded_amnt > pub_rec > pub_rec_bankruptcies > dti > installment > num_emp_len`
- 3) Medium negative correlation** - `annual_inc`