Lending Club Case Study

Data Understanding

- The aim is to find the loan applicants that defaulted or not based on the past history . In this case we need to consider “loan_status”
- Dataset has around 40K entries with 111 features .We need to remove the columns which are not impacting loan status .
- Check the data types for better understanding .
- Check the data dictionary to understand the columns .
- Loan status has 3 values “Current”,”Fully paid” and “Charged off”, since we are not sure on the current status if they will pay up or not , we will ignore them and consider only full paid and charged off status.

Data cleaning

- Drop columns containing all null value, max null values ,unique value etc .
- Filter rows with missing values
- Change the data types for the columns whereever required eg : float64,int64 etc
- Convert certain column values to numeric variables for better analysis such as remove % or year etc .

Data Analysis

- Derived Columns
- Univariate Analysis
- Segmented Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis

Derived Metrics

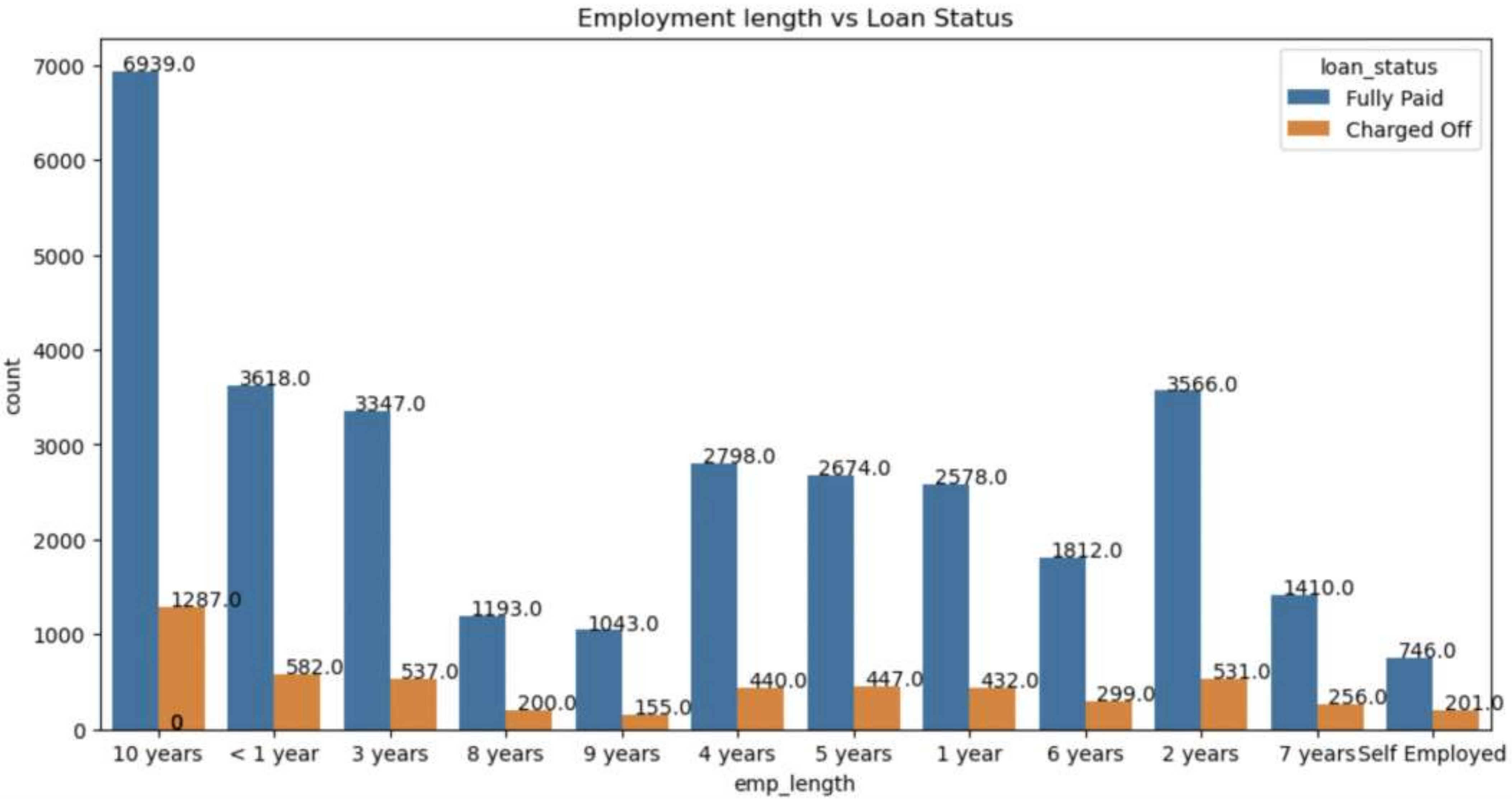
- Derive columns for Month and Year from column “issue_d”
- Derive column for “loan_amnt” to “annual_inc” ratio as ” ratio_loan_to_income”
- Create new column loan_status_code with 0 and 1 values based on loan_status column where 0=“Fully Paid” and 1=“Charged off”
- Creating different groups for interest rate
- Combine “Source Verified” and “Verified” into “Verified”
- Create Group annual_income

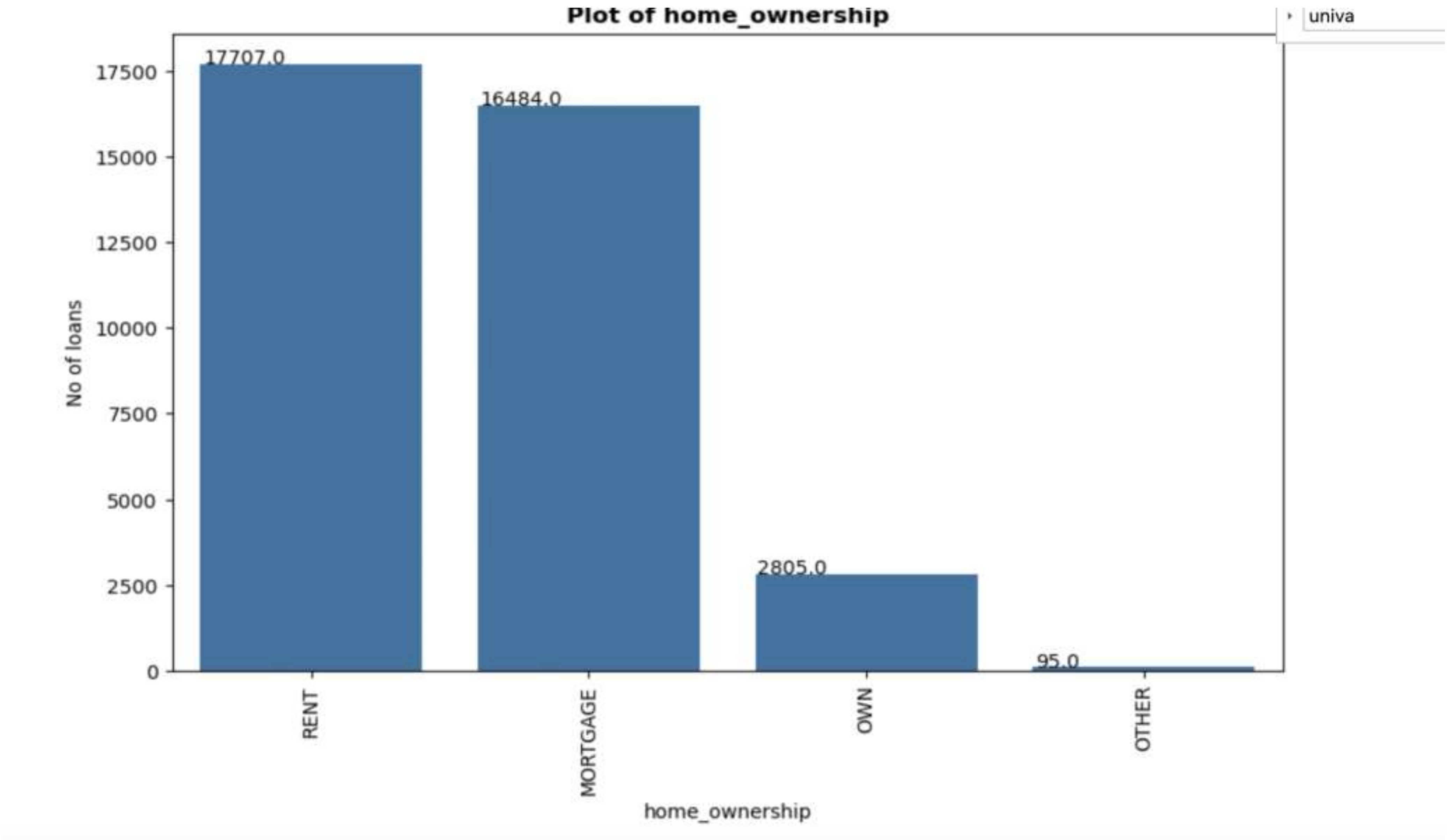
Univariate Analysis

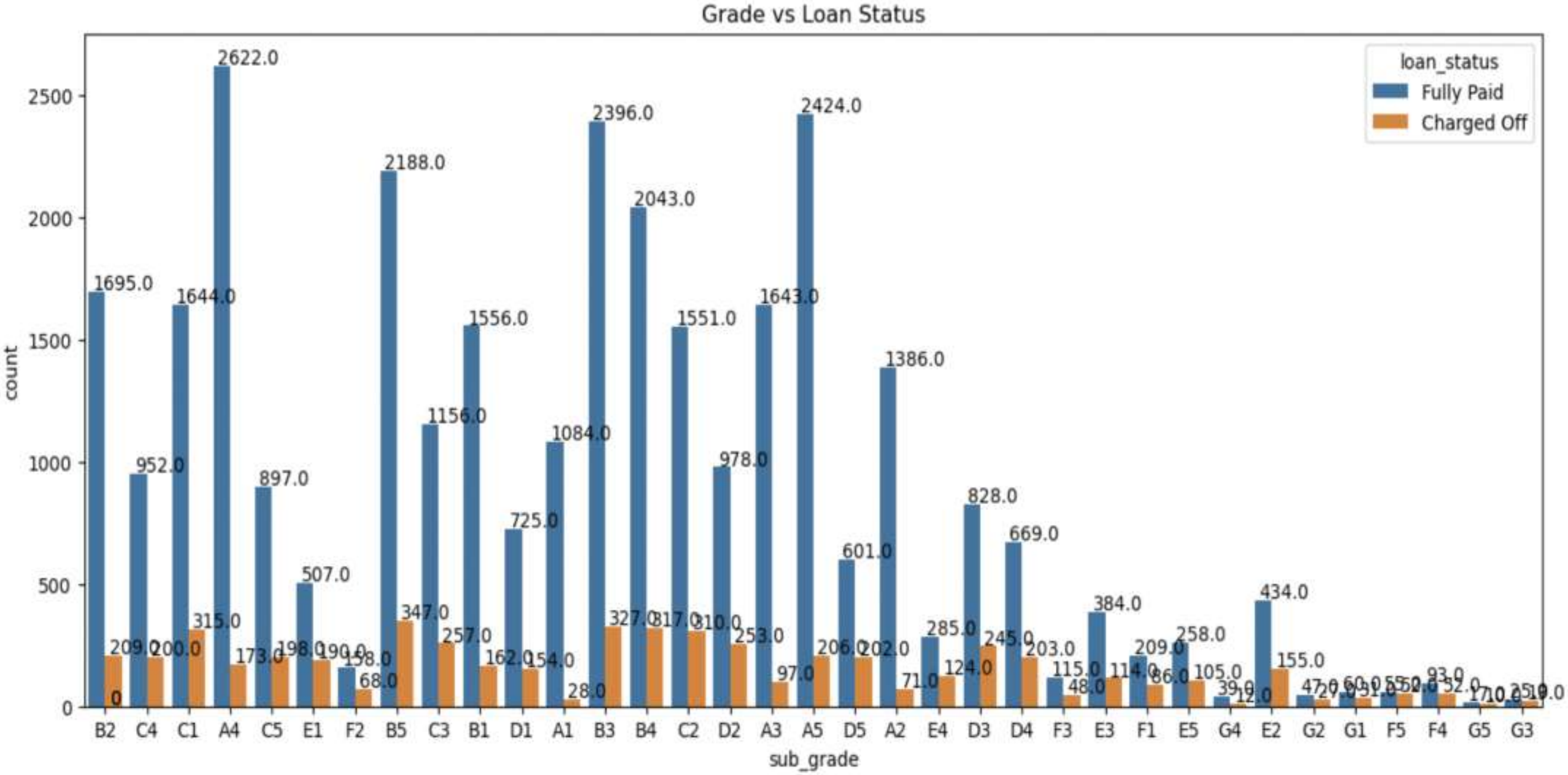
- Categorical variables
  - Ordered categorical data
    - Grade
    - Sub grade
    - Term (36 / 60 months)
    - Employment length
    - Loan issue year
    - Loan issue month
  - Un-ordered categorical data
    - State
    - Loan purpose
    - Home Ownership
    - Loan status
- Quantitative
  - Interest rate group
  - Annual income group
  - Loan amount
  - Funded amount
  - Loan amount to annual income ratio

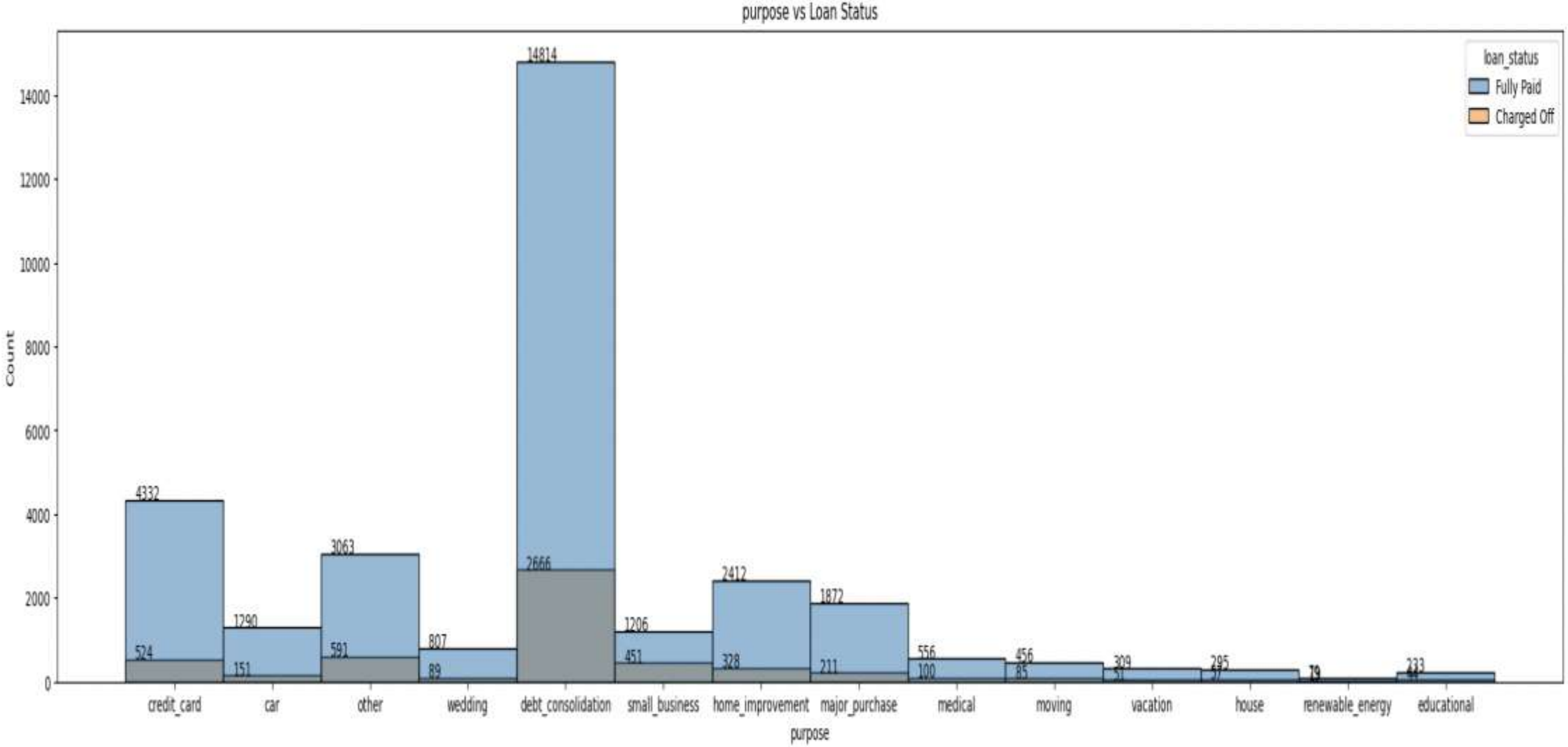
Univariate and Segmented Univariate analysis

- Grade A and B are given more loans compared to other grades
- Sub Grade A4, B3, A5, B5, B4 are given more loans compared to other grades
- term shows 36 months loans are issued more compared to 60 months loan
- employees with 10 years and above are given loan comapred with lesser experience
- year shows maximum loans were taken in the year 2011 and is in increasing trend since 2007
- Maximum loans were given in the month of Oct, Nov, Dec
- 14% of the total loans are charged off and 86% are fully paid
- States CA, NY, FL ,TX ,NJ are the top 5 states where maximum loans been issued .
- Maximum loans are given for debt consolidation .
- People who are in Rented house or Mortgate have availed maximum loans
- Funded amount is ranging from 500 to 35000 USD
- Installment amount is ranging from ~15 to 1300 USD
- The loan amount to income ratio mean is around 0.18
- Interest rate range 10 to 15 is the range where maximum loans have been issued
- 20 - 25% is the range where minimum loans have been issued


Employment length vs Loan Status
loan_status
Fully Paid
Charged Off
count
0
1000
2000
3000
4000
5000
6000
7000
6939.0
1287.0
0
3618.0
582.0
3347.0
537.0
1193.0
200.0
1043.0
155.0
2798.0
440.0
2674.0
447.0
2578.0
432.0
1812.0
299.0
3566.0
531.0
1410.0
256.0
746.0
201.0
10 years
< 1 year
3 years
8 years
9 years
4 years
5 years
1 year
6 years
2 years
7 years
Self Employed
emp_length


Plot of home_ownership
univa
No of loans
17500
15000
12500
10000
7500
5000
2500
0
17707.0
16484.0
2805.0
95.0
RENT
MORTGAGE
OWN
OTHER
home_ownership


Grade vs Loan Status
loan_status
Fully Paid
Charged Off
count
sub_grade
0
500
1000
1500
2000
2500
B2
C4
C1
A4
C5
E1
F2
B5
C3
B1
D1
A1
B3
B4
C2
D2
A3
A5
D5
A2
E4
D3
D4
F3
E3
F1
E5
G4
E2
G2
G1
F5
F4
G5
G3
1695.0
209.0
0
952.0
200.0
1644.0
315.0
2622.0
173.0
897.0
198.0
507.0
190.0
158.0
68.0
2188.0
347.0
1156.0
257.0
1556.0
162.0
725.0
154.0
1084.0
28.0
2396.0
327.0
2043.0
317.0
1551.0
310.0
978.0
253.0
1643.0
97.0
2424.0
206.0
601.0
202.0
1386.0
71.0
285.0
124.0
828.0
245.0
669.0
203.0
115.0
48.0
384.0
114.0
209.0
86.0
258.0
105.0
39.0
12.0
434.0
155.0
47.0
27.0
60.0
31.0
55.0
52.0
93.0
52.0
17.0
10.0
25.0
19.0

purpose vs Loan Status
loan_status
Fully Paid
Charged Off
Count
0
2000
4000
6000
8000
10000
12000
14000
4332
524
1290
151
3063
591
807
89
14814
2666
1206
451
2412
328
1872
211
556
100
456
85
309
51
295
37
79
19
233
credit_card
car
other
wedding
debt_consolidation
small_business
home_improvement
major_purchase
medical
moving
vacation
house
renewable_energy
educational
purpose

Summary of Bivariate Analysis

- Based on the counts, Grade B, C and D are top three in Charged Off
- A,B have have also very good numbers of paying it off fully
- Based on the counts sub grades C1,B5,C3,B3,B4,C2 and D2 have higher charged off count
- Maximum loans were issued to 10 or above category and the charged off is also highest in this category
- Defaulters are more in 36 months category
- In 2011 the defaulters are more
- Don’t find any conclusive answer looking at the numbers for months.
- The defaulters are more when the purpose is debt_consolidation .
- People of RENT or MORTGAGE failed to pay
- The verified loans are charged off more than Not verified
- Employment above 10 years have got more loans and more defaulters too
- States from CA,FL and NY have more defaulters .
- The median of 10 years and more is highest
- 6,7,8 and 9 having almost same median value .
- The median is more in case of defaulter .
- The IQR is more in case of charged off

Summary of Multivariate Analysis

- Higher the interest rate higher charged off ratio
- Higher the annual income higher the loan amount slightly.
- increase in number of charged off with increase in year.
- interest rate is increasing with loan amount increase