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”
- 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 statuses
- We will also consider “funded_amt” variable to be impacted by “loan_status”(funded_amt_inv can be ignored since its an individual choice)
- Drop columns like id,member id , null values,

Data cleaning

- Drop columns containing max null value, single value, which doesn’t add any value
- Filter rows with missing values
- Clean and convert entire column to correct data type eg : date,int64 etc
- Convert certain column values to numeric variables for better analysis

Data Analysis

- Derived Metrics
- Univariate Analysis
- Segmented Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis

Derived Metrics

- Derive metrics for Month and Year from column “issue_d”
- Derive metrics for “loan_amnt” to “annual_inc” ratio
- 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

Summary of univariate analysis

- Grade A and B are given more loans compared to other grades
- 36 months loans are issued more compared to 60 months loan
- Grade A4, B3, A5, B5, B4 are given more loans compared to other grades
- Maximum loans were taken in the year 2011. The trend is increasing with the increase in the year
- Employees with 10 years and above are given loan compared with lesser experience
- There is increasing trend in number of loans with increase in the months. Maximum loans were given in the month of Oct, Nov, Dec.
- 15 % of the total loans are charged off
- States CA, NY, FL and TX are the states for which maximum loans have been issued
- Maximum loans are given for debt consolidation, paying off Credit card and 'other' reasons
- Education and renewable energy is the least category where loans have been given
- People who are in Rented house or Mortgage have availed maximum of the loans
- Funded amount is ranging from 5000 to 15000 USD
- Installment amount is ranging from 200 to 400 USD
- The amount to income ratio median is around 0.18; this is a good indicator
- Interest rate range 9 to 13 is the range where maximum loans have been issued
- 21 - 25% is the range where minimum loans have been issued

Bivariate Analysis

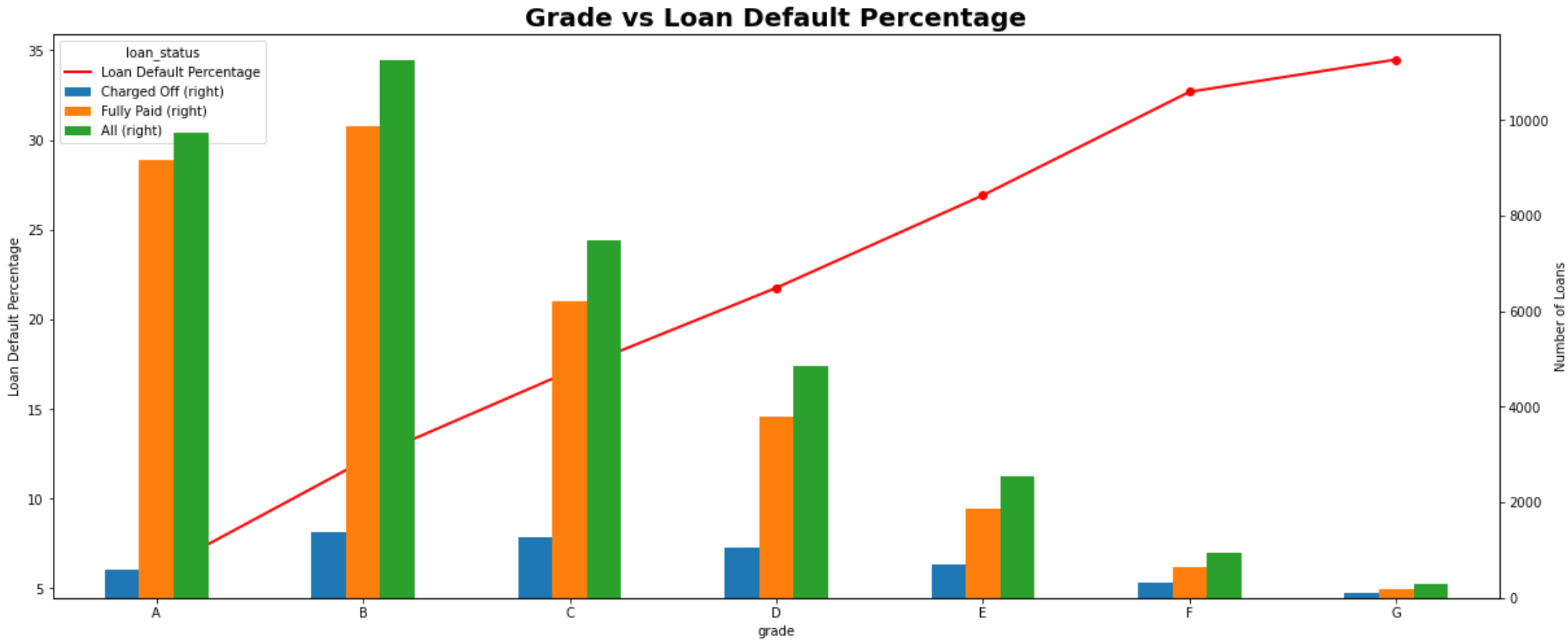
- Ordered categorical data
  - Grade
  - Sub grade
  - Term (36 / 60 months)
  - Employment length
  - Year
  - Month
- Un-ordered categorical data
  - State
  - Loan purpose
  - Home Ownership
  - Verified status
- Quantitative
  - Interest rate
  - Annual income
  - DTI
  - Loan amount to annual income ratio

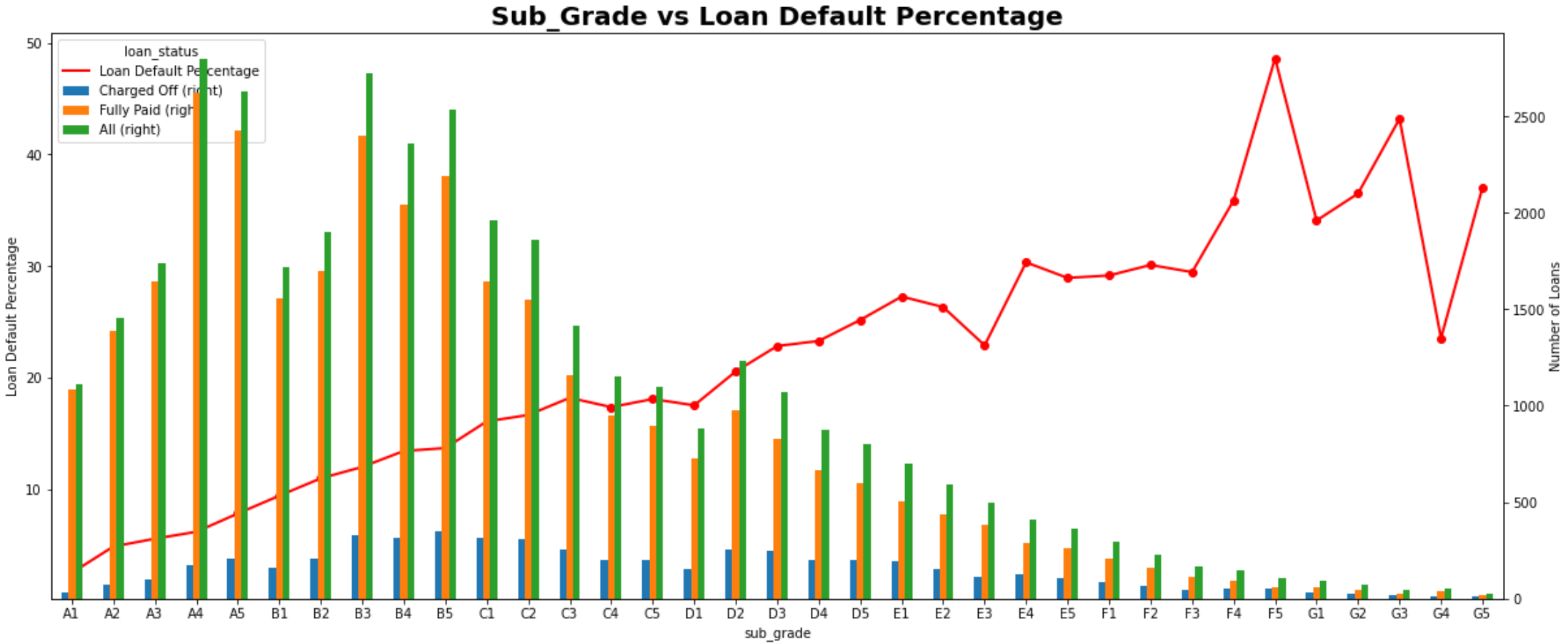
Summary of Bivariate Analysis

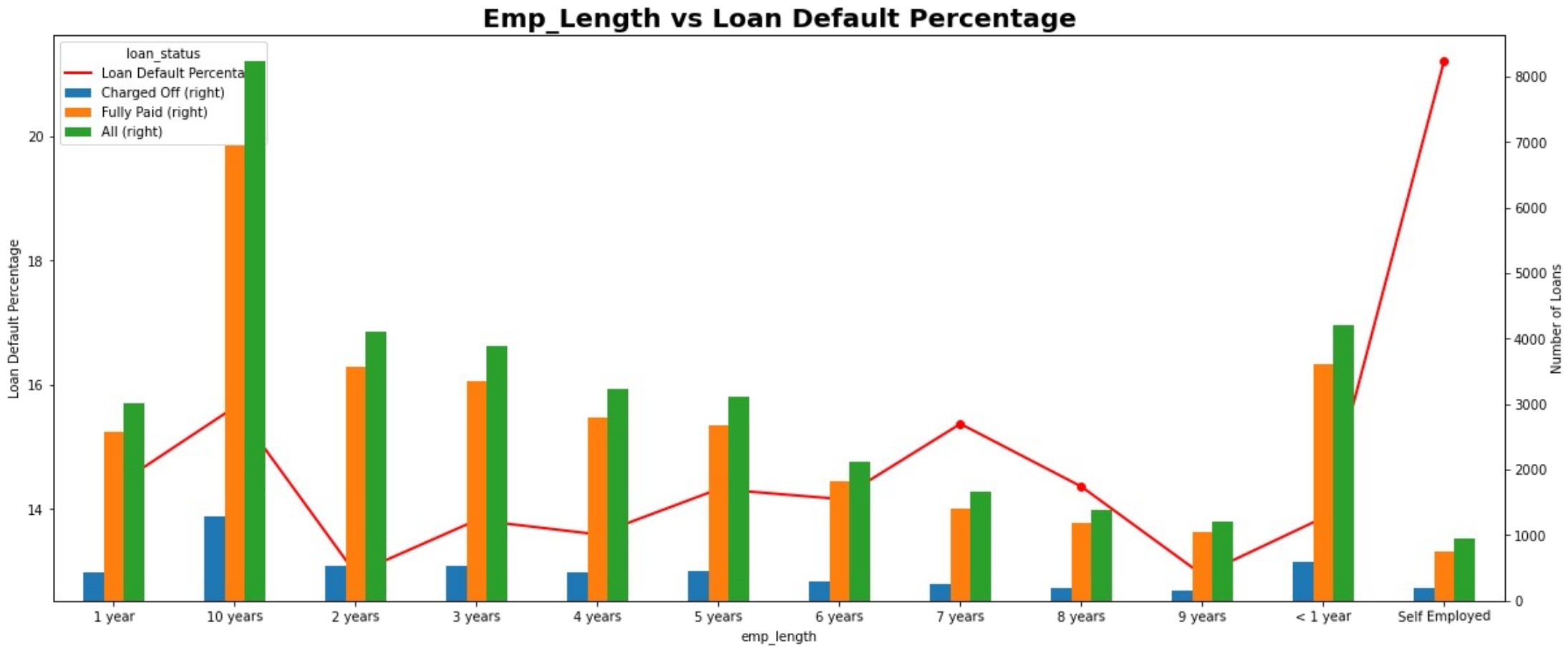
- 36 month loan default is more compared to 60 month
- Maximum loans are issued for category 10 years and they are the maximum defaulters
- Based on the counts, Grade B, C and D are top three in Charged Off
- Based on the counts, Grade B3,B4,B5, C1,C2, D3, D4 top sub grades in Charged Off
- Plot of loan issue year shows maximum loans were taken in the year 2011
- Also high loans are being Charged Off in 2011
- Plot of the loan issue month shows maximum loans were given in the month of Oct, Nov, Dec.
- Also high loans are being Charged Off for the loans issued in Sep - Dec months
- Loans with purpose debt consolidation, other, credit card and home improvement categories have failed to pay the loan compared with education / renewable energy
- Also debt consolidation is the category where maximum loans are issued.
- People who are in Rent or Mortgage have failed to pay the loan compared with people in Own house
- The amount of Verified loans which are Charged Off is more compared to Not Verified
- Employment length of 10 years got more loans and they are the maximum defaulters
- Followed by year 6, 7, 8 and 9 where the medians are almost close
- Grade F, G and E are the three category which are top three in Charged off
- Grade F and G have the median around 20k and Q3 at 25k
- Grade A is the least with median at 7.5k
- DTI doesn't seem to be contributing because the median is close to each other for Fully Paid and Charged Off
- “Amnt_to_int_ratio” is contributing for bad loans

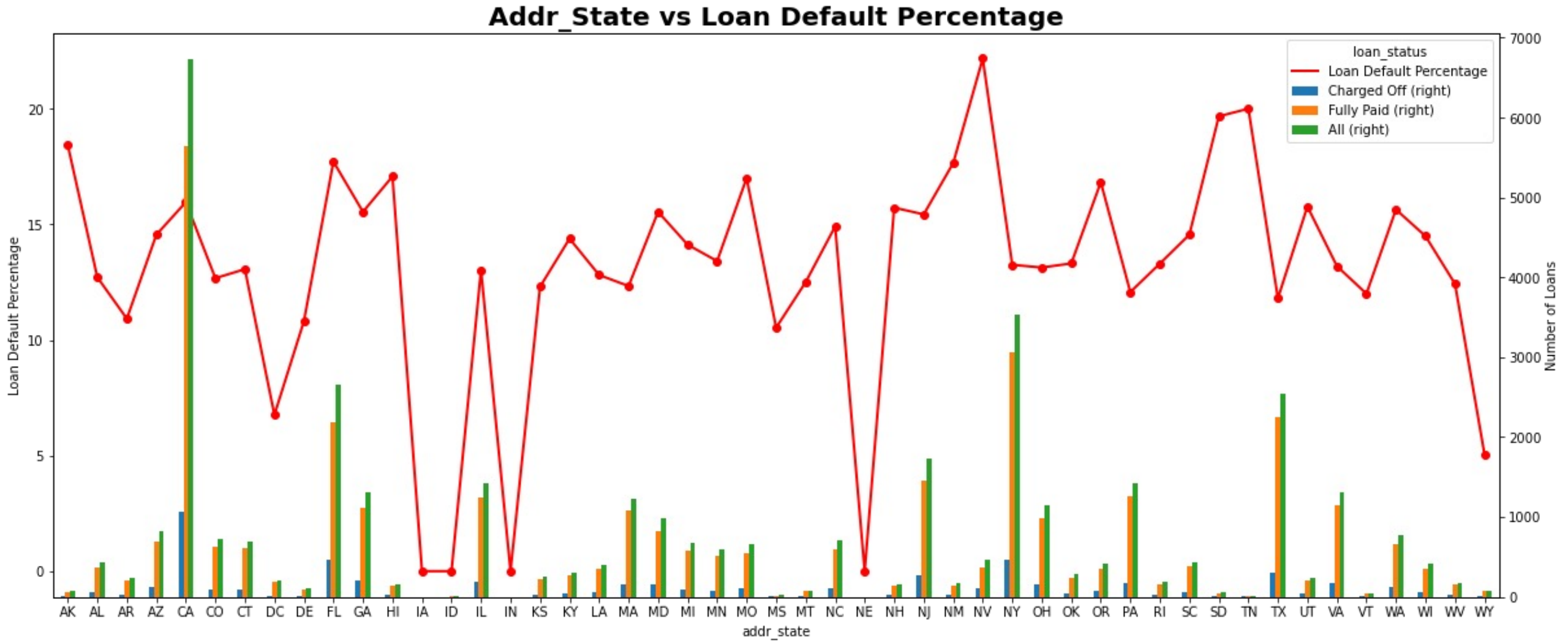
Summary of Multivariate Analysis

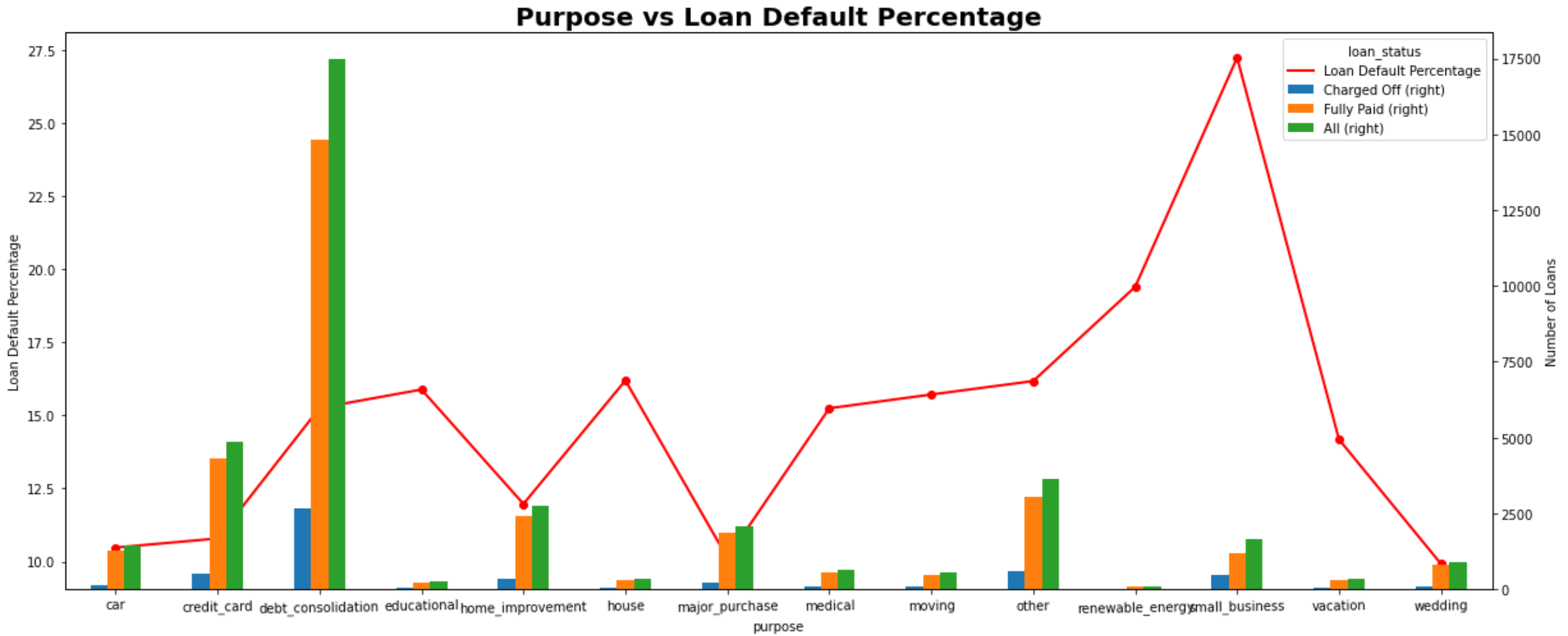
- Borrowers from sub grade F5, G3 and G5 have maximum tendency to default.
- Self employed, 10 years and 7 years categories has maximum tendency to default the loan.
- Tendency to default the loan is increasing from Grade A to Grade G
- Borrowers from states NV, TN, SD, AK, FL and HI have maximum tendency to default the loan.
- Small business category has the highest chances to default the loan
- Borrowers from Other category have highest tendency to default the loan.
- Borrowers from Other category have highest tendency to default the loan.
- The borrowers who are in lower income groups have maximum tendency to default the loan and it generally decreases with the increase in the annual income.
- The tendency to default the loan is increasing with increase in the interest rate.
- Correlation graph shows variables annaul_inc, installment, issue_m, amnt_to_inc_ratio, loan_amt, funded_amt, issue_y have negative impact on the loan_status variable

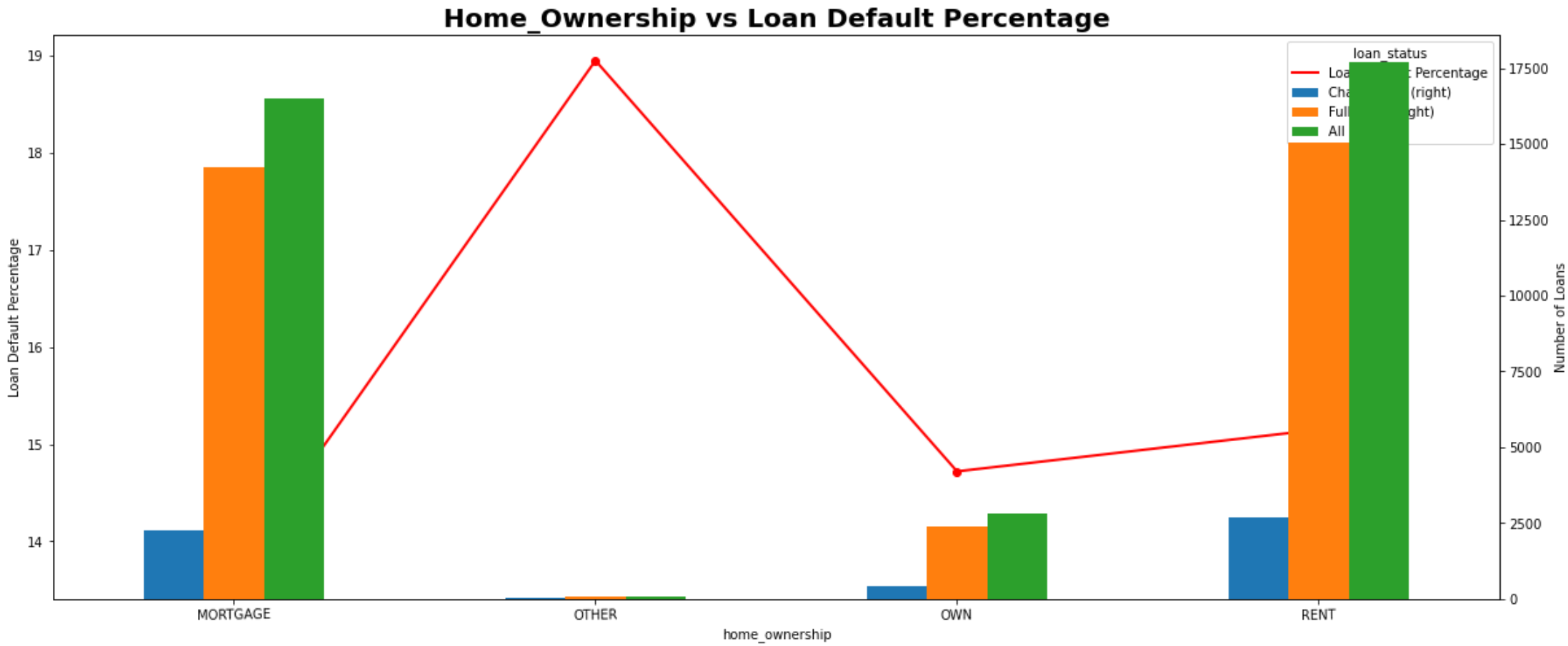

Grade vs Loan Default Percentage
loan_status
Loan Default Percentage
Charged Off (right)
Fully Paid (right)
All (right)
Loan Default Percentage
Number of Loans
grade
A
B
C
D
E
F
G
5
10
15
20
25
30
35
0
2000
4000
6000
8000
10000

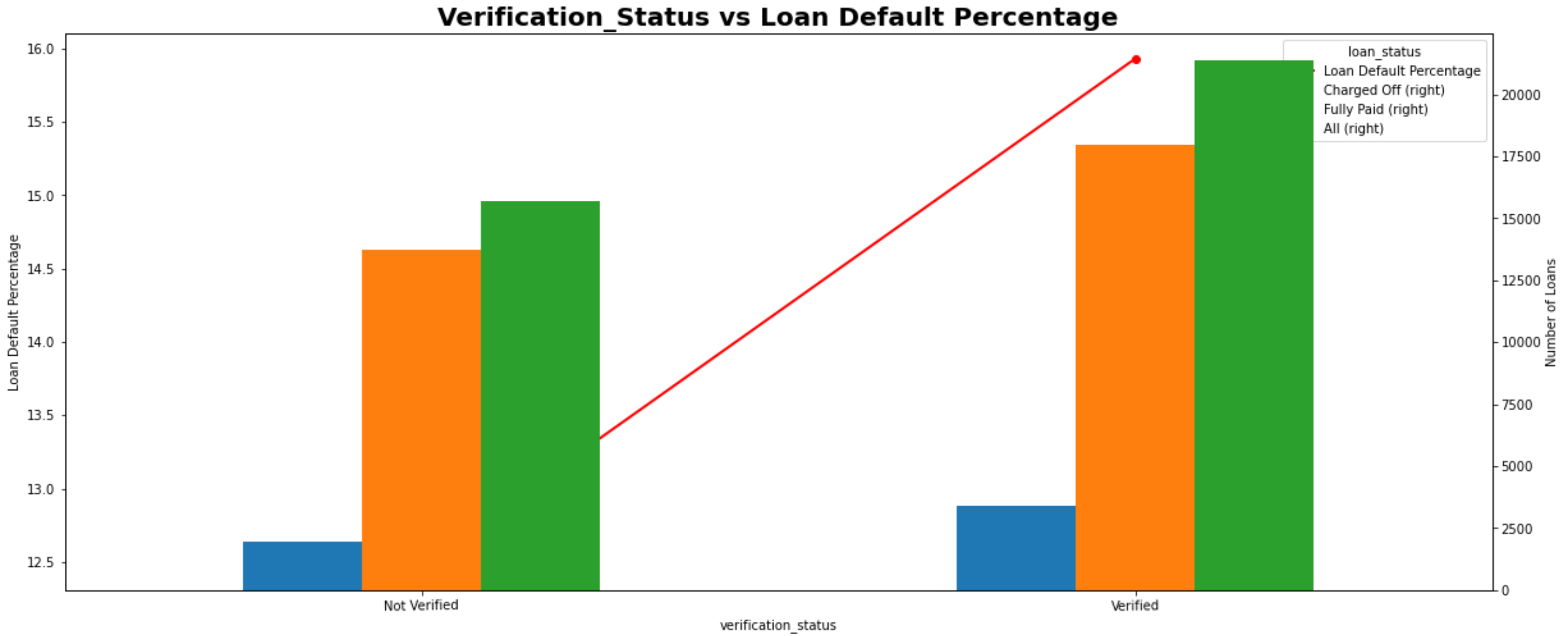

Sub_Grade vs Loan Default Percentage
loan_status
Loan Default Percentage
Charged Off (right)
Fully Paid (right)
All (right)
Loan Default Percentage
Number of Loans
sub_grade
A1 A2 A3 A4 A5 B1 B2 B3 B4 B5 C1 C2 C3 C4 C5 D1 D2 D3 D4 D5 E1 E2 E3 E4 E5 F1 F2 F3 F4 F5 G1 G2 G3 G4 G5
10 20 30 40 50
0 500 1000 1500 2000 2500

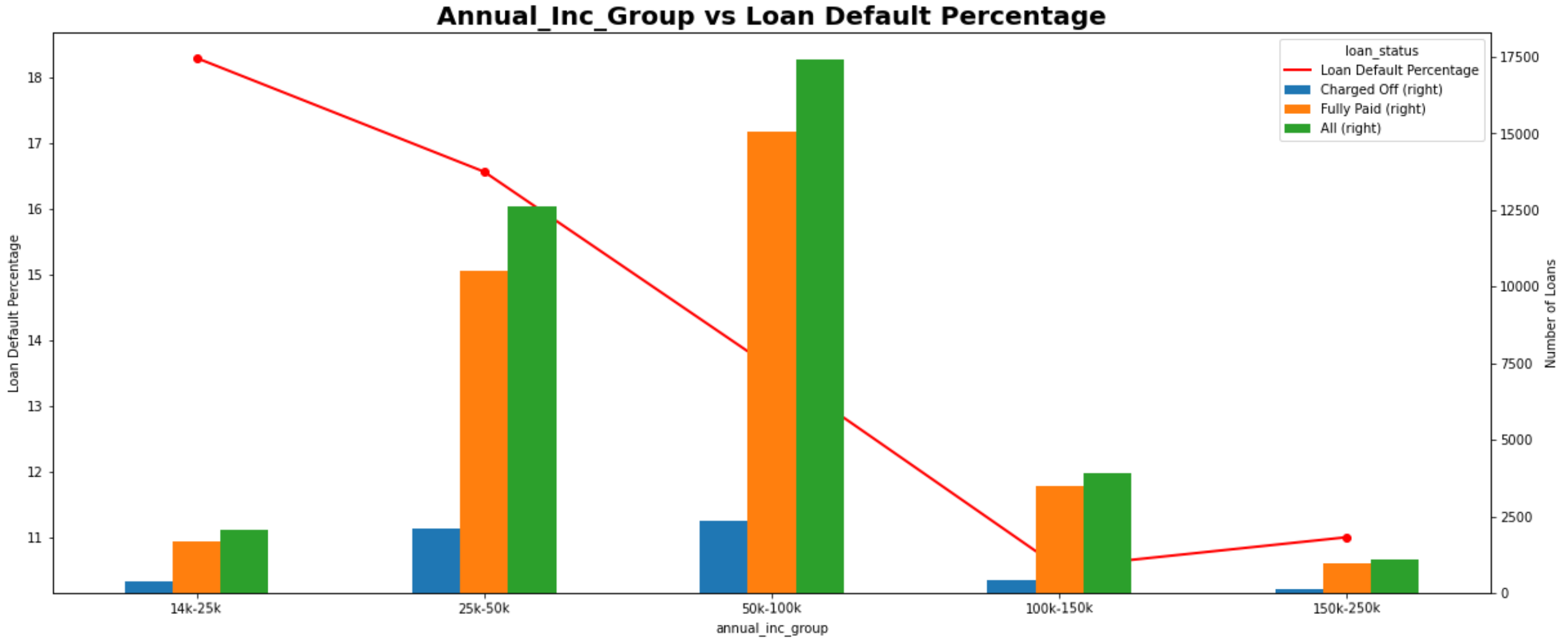

Emp_Length vs Loan Default Percentage
loan_status
Loan Default Percenta
Charged Off (right)
Fully Paid (right)
All (right)
Loan Default Percentage
Number of Loans
20
18
16
14
8000
7000
6000
5000
4000
3000
2000
1000
0
1 year
10 years
2 years
3 years
4 years
5 years
6 years
7 years
8 years
9 years
< 1 year
Self Employed
emp_length

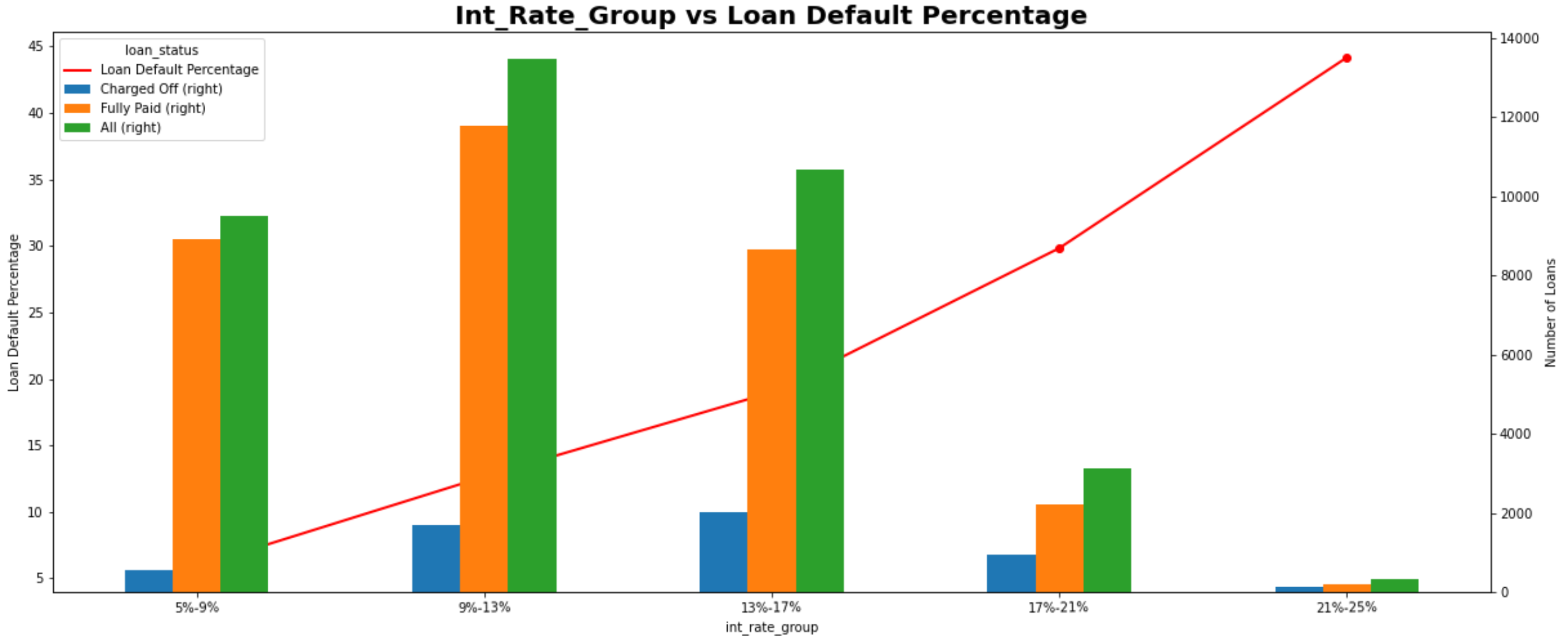

Addr_State vs Loan Default Percentage
loan_status
Loan Default Percentage
Charged Off (right)
Fully Paid (right)
All (right)
Loan Default Percentage
Number of Loans
addr_state
AK AL AR AZ CA CO CT DC DE FL GA HI IA ID IL IN KS KY LA MA MD MI MN MO MS MT NC NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY
0 5 10 15 20
0 1000 2000 3000 4000 5000 6000 7000


Purpose vs Loan Default Percentage
loan_status
Loan Default Percentage
Charged Off (right)
Fully Paid (right)
All (right)
Loan Default Percentage
Number of Loans
27.5
25.0
22.5
20.0
17.5
15.0
12.5
10.0
17500
15000
12500
10000
7500
5000
2500
0
car
credit_card
debt_consolidation
educational
home_improvement
house
major_purchase
medical
moving
other
renewable_energy
small_business
vacation
wedding
purpose


Home_Ownership vs Loan Default Percentage
Loan Default Percentage
Number of Loans
home_ownership
MORTGAGE
OTHER
OWN
RENT
loan_status
19
18
17
16
15
14
17500
15000
12500
10000
7500
5000
2500
0


Verification_Status vs Loan Default Percentage
loan_status
Loan Default Percentage
Charged Off (right)
Fully Paid (right)
All (right)
Loan Default Percentage
Number of Loans
16.0
15.5
15.0
14.5
14.0
13.5
13.0
12.5
20000
17500
15000
12500
10000
7500
5000
2500
0
Not Verified
Verified
verification_status


Annual_Inc_Group vs Loan Default Percentage
loan_status
Loan Default Percentage
Charged Off (right)
Fully Paid (right)
All (right)
Loan Default Percentage
Number of Loans
annual_inc_group
14k-25k
25k-50k
50k-100k
100k-150k
150k-250k
18
17
16
15
14
13
12
11
17500
15000
12500
10000
7500
5000
2500
0


Int_Rate_Group vs Loan Default Percentage
loan_status
Loan Default Percentage
Charged Off (right)
Fully Paid (right)
All (right)
Loan Default Percentage
Number of Loans
int_rate_group
5%-9%
9%-13%
13%-17%
17%-21%
21%-25%
45
40
35
30
25
20
15
10
5
14000
12000
10000
8000
6000
4000
2000
0