

Lending Club – Case Study



Problem Statement



Analysis Approach



Results of Analysis

Univariate Result

Bivariate Results



Conclusion

Context

Problem Statement



Lending Club is the largest online loan marketplace, facilitating various loans through a fast online interface.



As Loan Lending firm to understand the driving/influencing factor of defaulter loans to reduce the credit loss of business and increase the profitable loans



**EDA is done on Lending club
in below steps,**



Loading the data to DataFrame

Here we will read the data from csv file
and store it in a data frame object



Data type Conversions

In this step, we will check the column
data with the type of data

Based on requirement we have
changed the data type. E.g. even
though ID column is numeric, but as
per business perspective it's a unique
value for each customer, so keeping it
as a numeric will not give any benefit.
For this converting it to String to check
the unique values or not



Checking Business Conditions

Here we are checking the business
given condition e.g. id and member id
column should have all unique values,
term of the loan should be 36 & 60
only, etc

Analysis Approach



Data Cleaning

In this step, we removed all empty columns as well as columns with single value, e.g. **pymnt_plan** column only have value **n** for all the records, these kind of columns are not useful for analysis



Handling Missing values

Columns with more than 50 % missing values are dropped as imputing these values make affect on the analysis

For other missing values, following below approach

- Numeric column: imputing missing records with median of the column
- Categorical Column: Imputing with MODE value. But if the categorial column is an influential column, imputed with a new value



Cleaning Outliers

Outliers of the influential column like loan amount, annual income, etc. columns are dropped by taking the upper fence

Analysis Approach-Cont...

Analysis Approach- Cont...

Deriving Columns

- We have derived new columns based on exiting columns, like issue_d column
- We have also derived columns based on range of the values in a column

Analysis

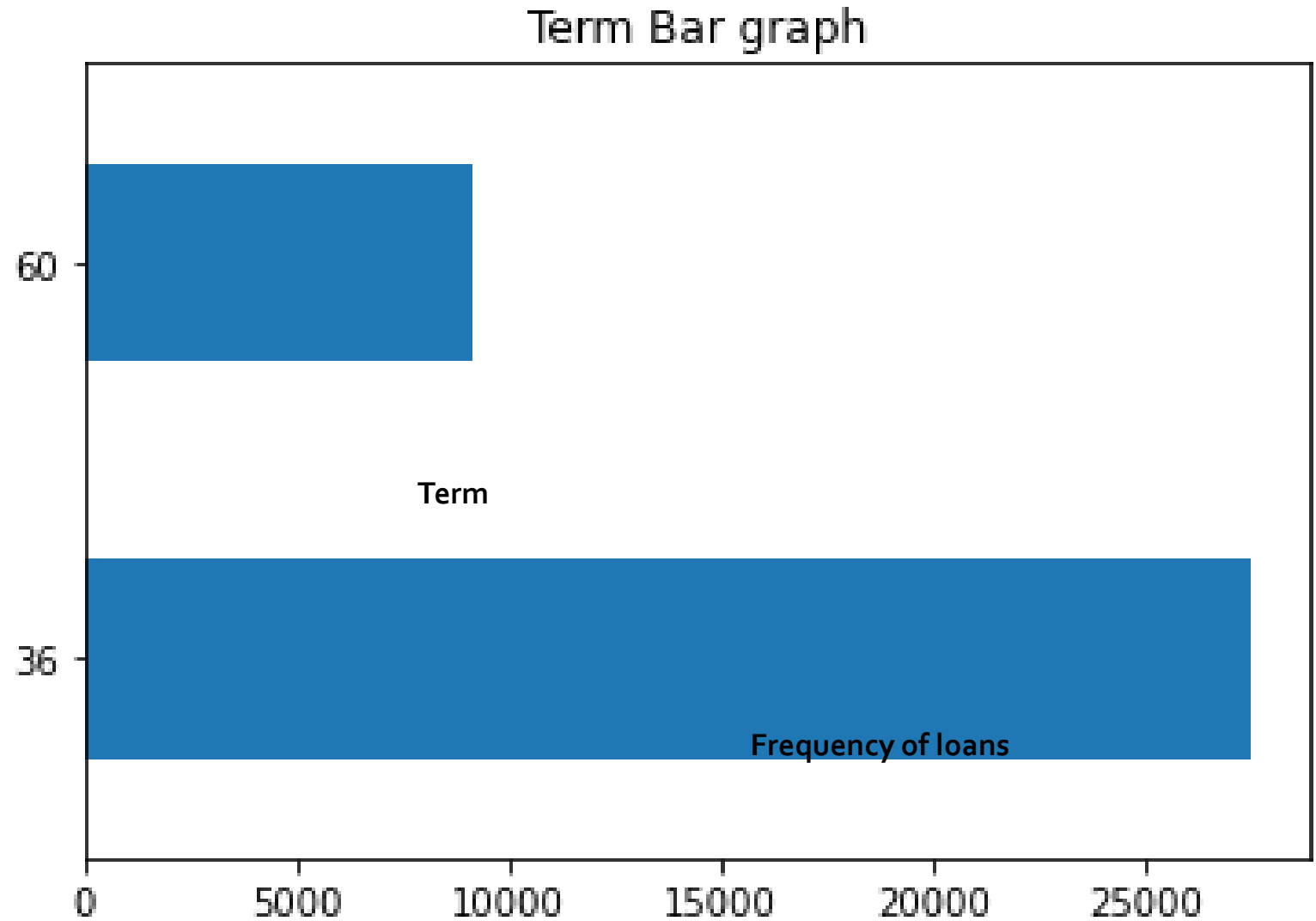
- We have done below analysis
 - Univariate
 - Bivariate
 - Multivariate



Results of Analysis – Univariate Results

Term

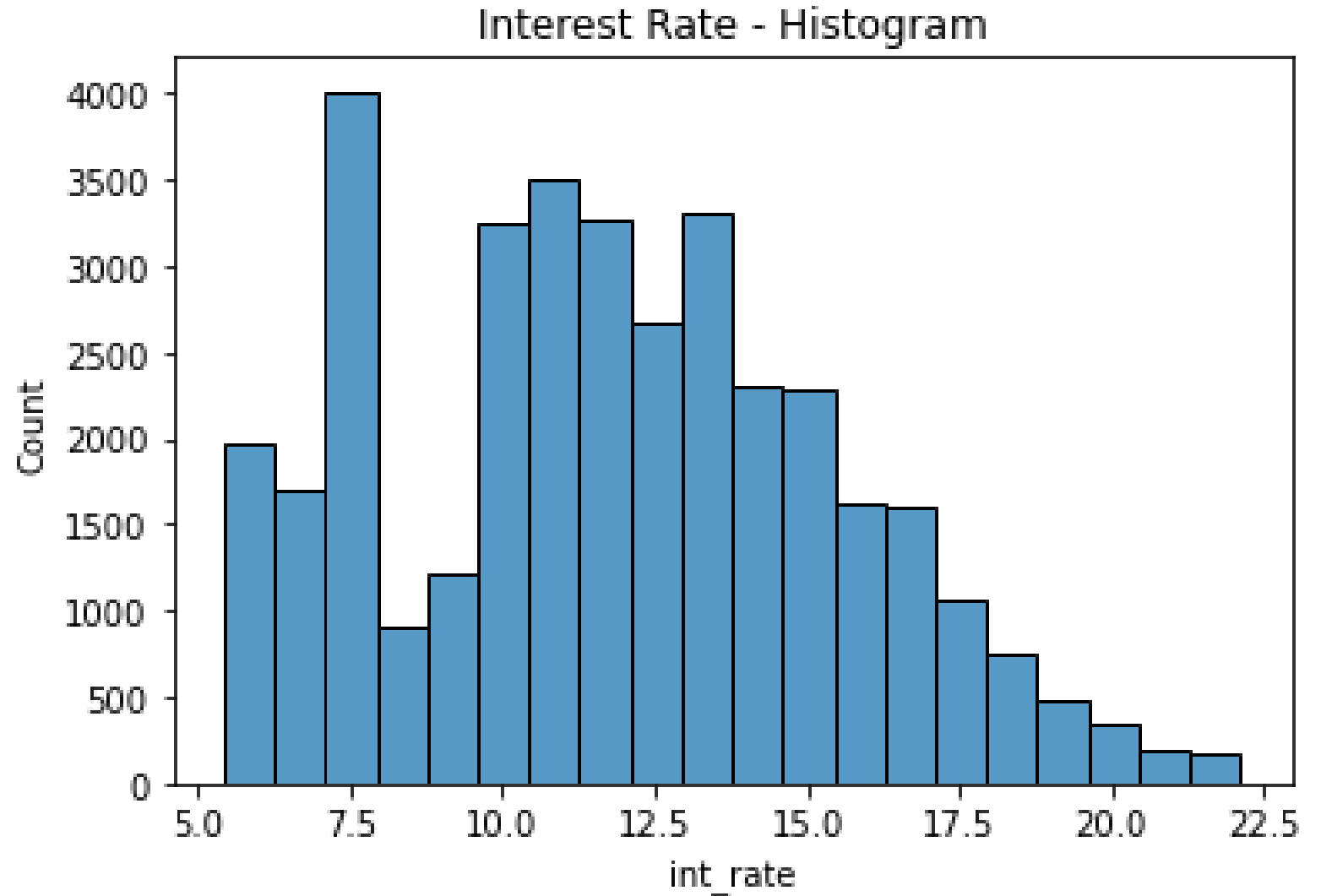
Most of the customers preferring 36 month term loan over 60 month, as 36 month term loan count is more than twice as 60 month term loan count



Interest Rate

From the above histogram, we can see that most of the loans are with 7.5 % interest loan

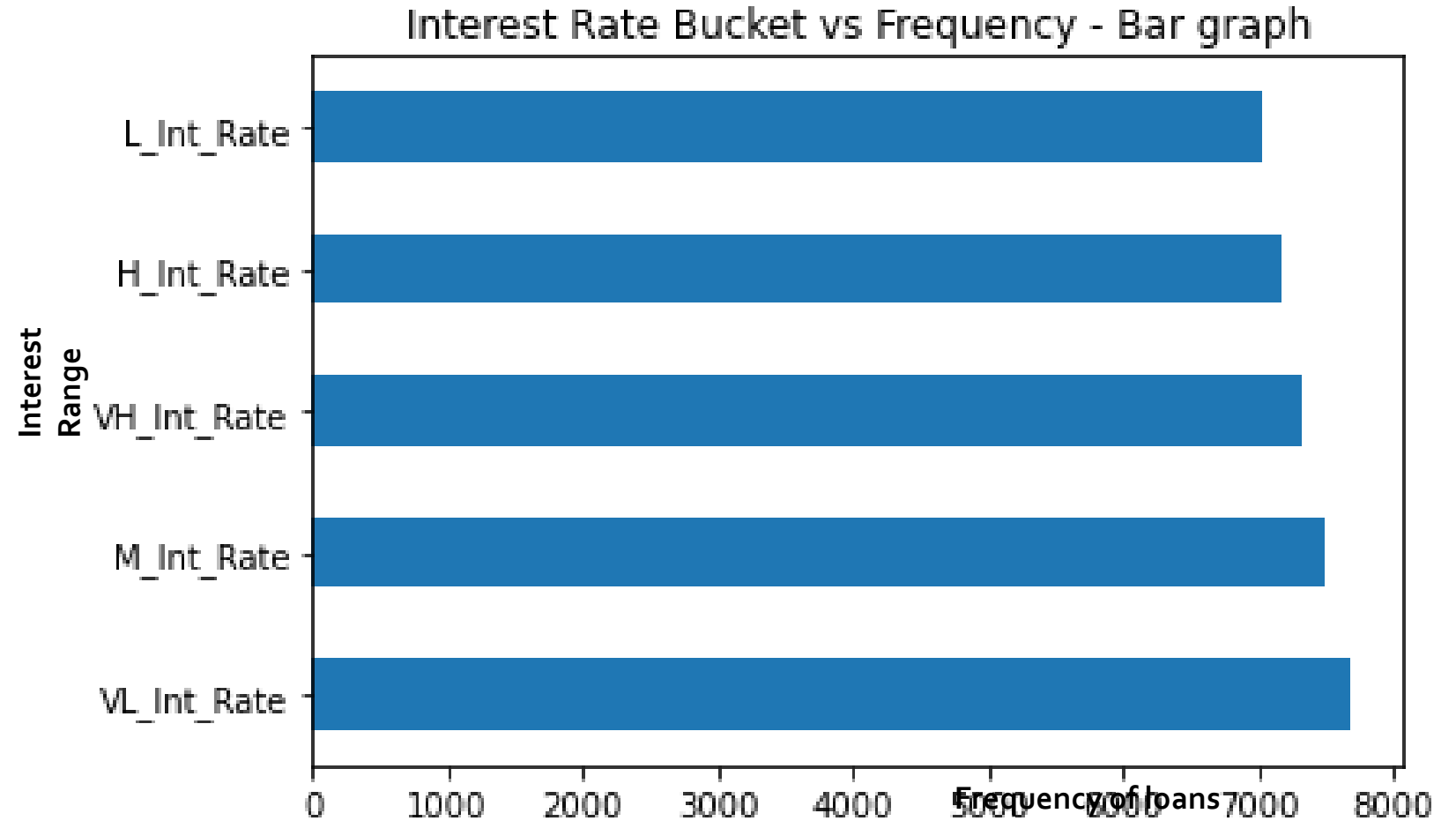
Number of loans are started to decrease from interest rate 15



Interest Rate Range

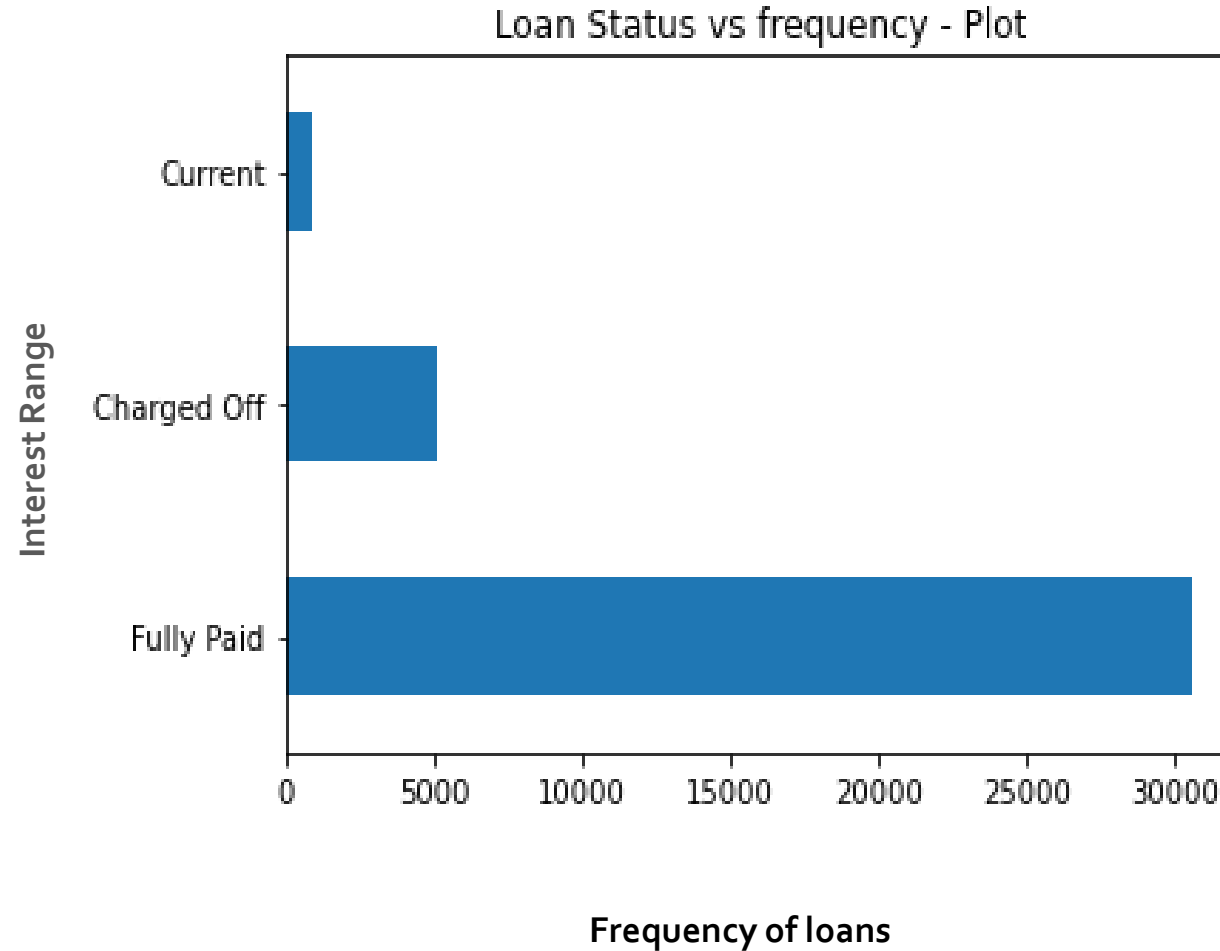
Now we can see that number of loans spread across the interest range are approximately equal

Very Low interest rate loan are have more loans in comparison



Loan Status

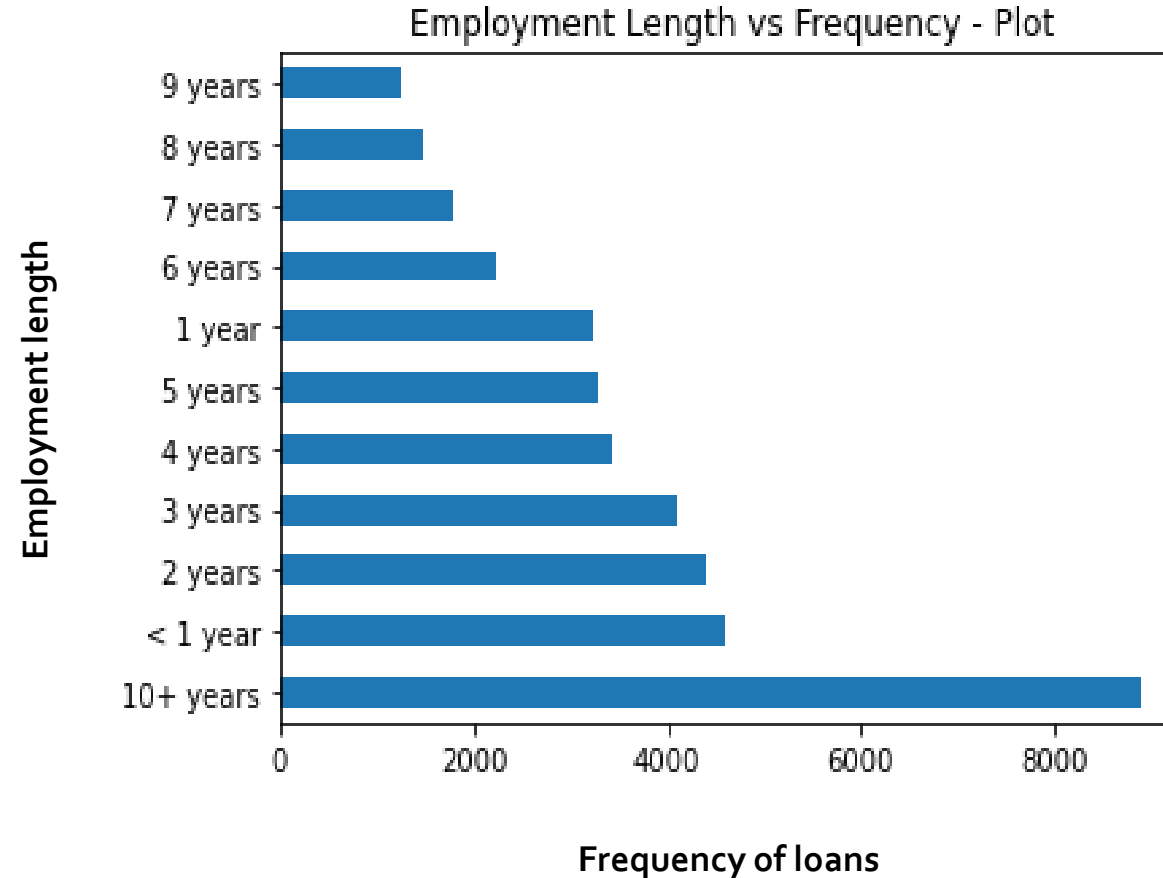
We can see that 30k+ loans are fully paid, 5K+ loans are Charged Off and the current loans are very less compare to others



Employment Length

We can observe that Customer/Borrower with 10+ years of experience had more than 8K loans and it's double the number of loans compares to other records

Except for 10+ years of experience customer, by experience increases number of customers taking loan is decrease (except for customer with 1 year experience)

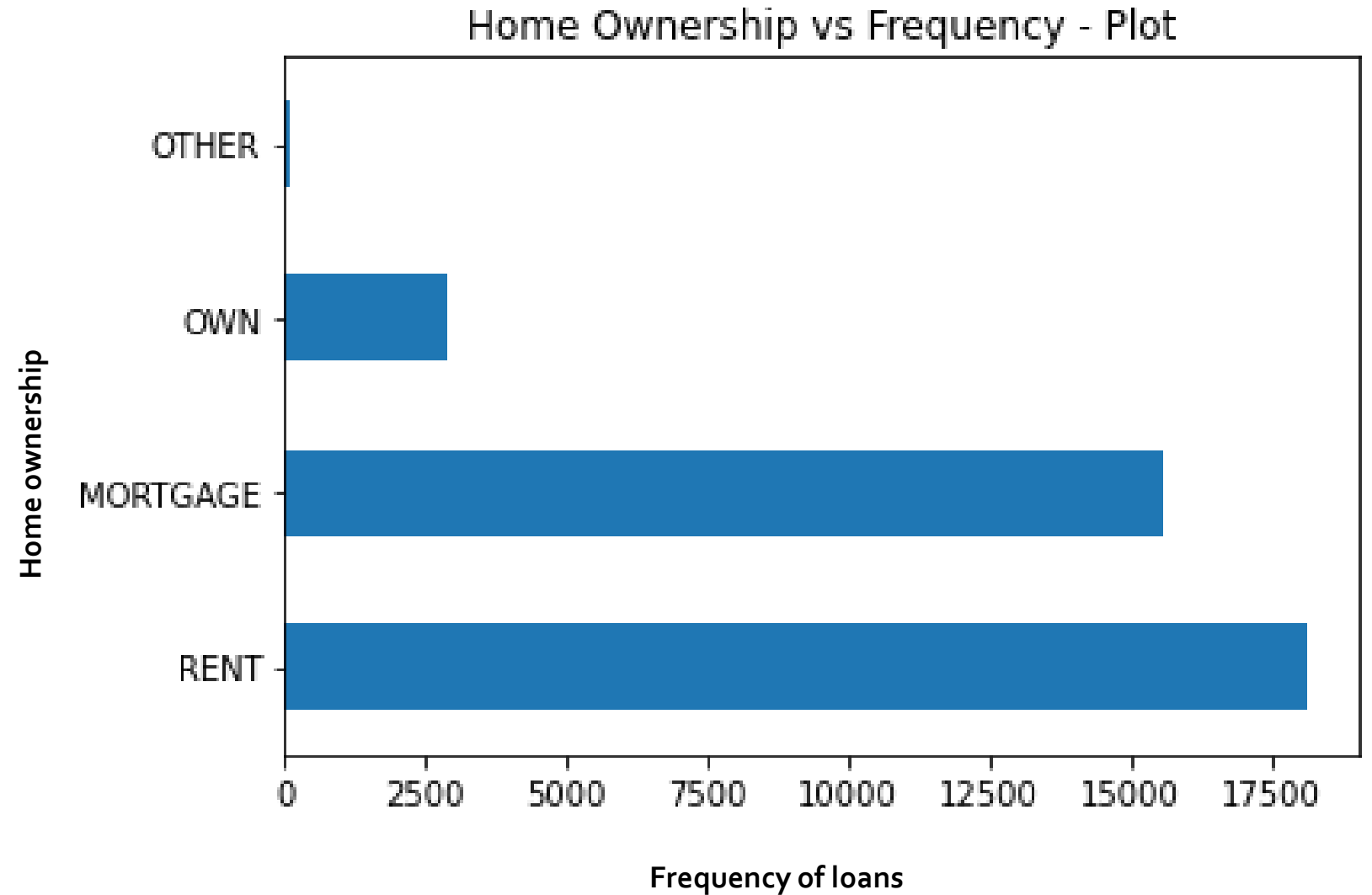


Home Ownership

Most of the customer are either in Rent or Mortgage

Only 2.5K+ customers are having own

Very negligible number of customers home ownership is Other



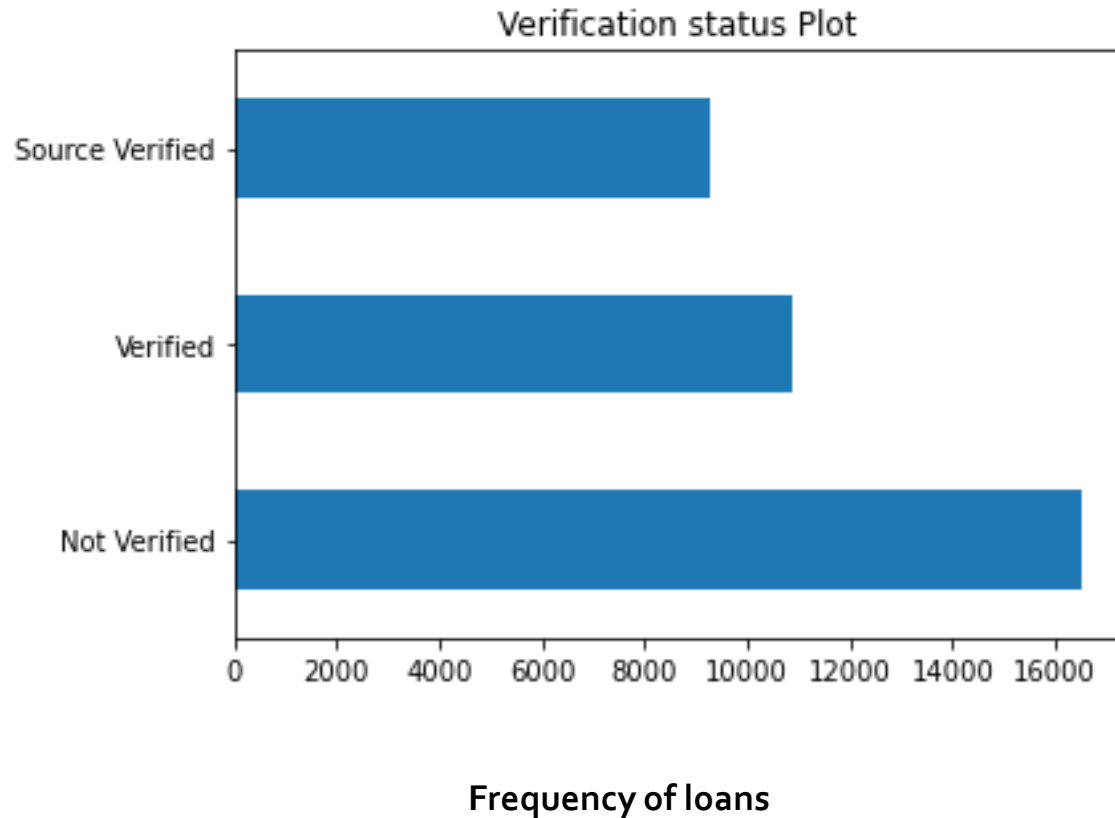
Verification Status

We can see most of the loan customers are Not verified compare individually with Verified and Source Verified customer

in total, verified customers (including both verified and source verified) are more than not verified.

Source Verified and Verified customers are approximately equal

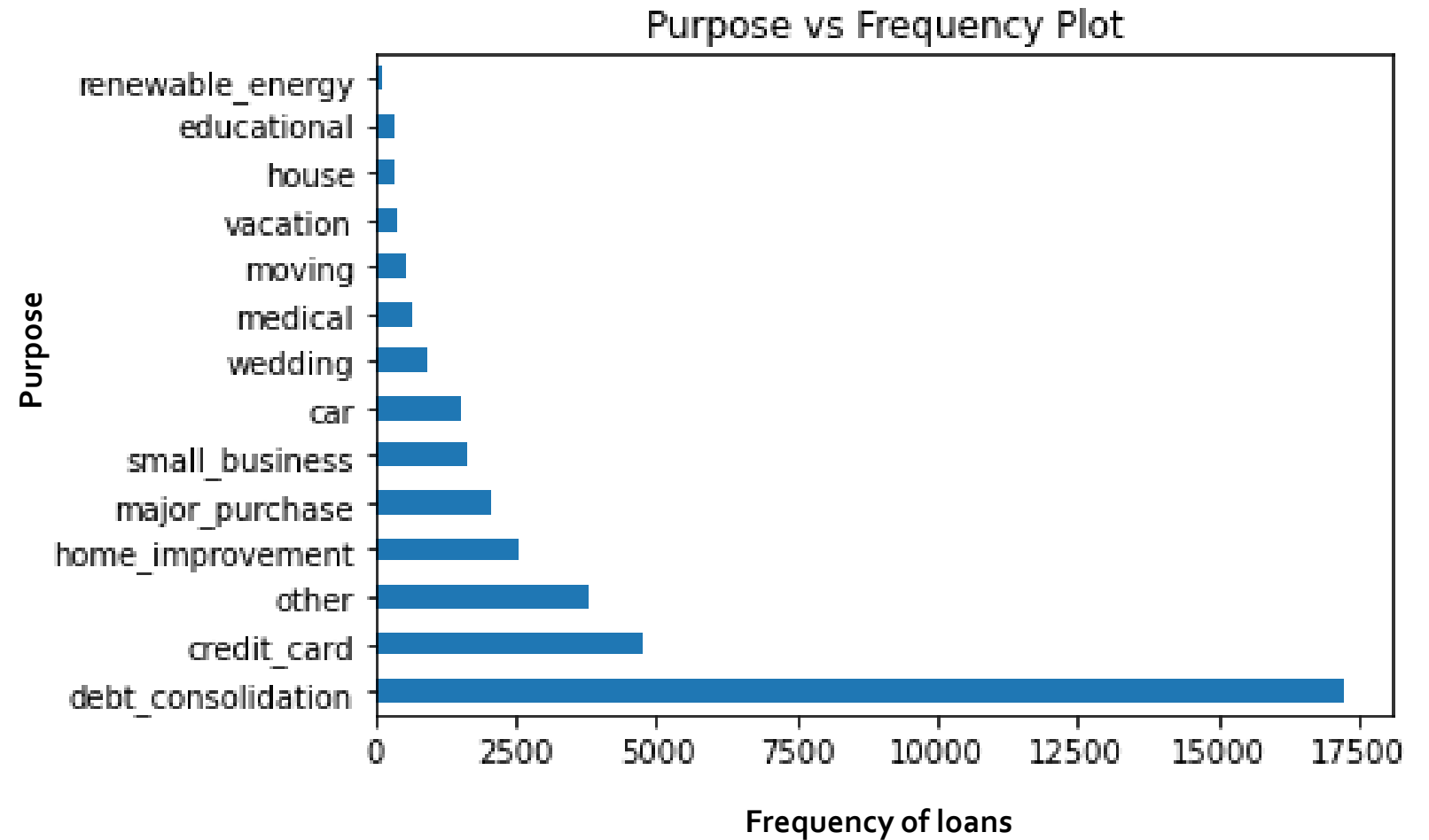
Verification Status



Purpose

Most of the customers (nearly half of them) having debt consolidation for purpose of loan

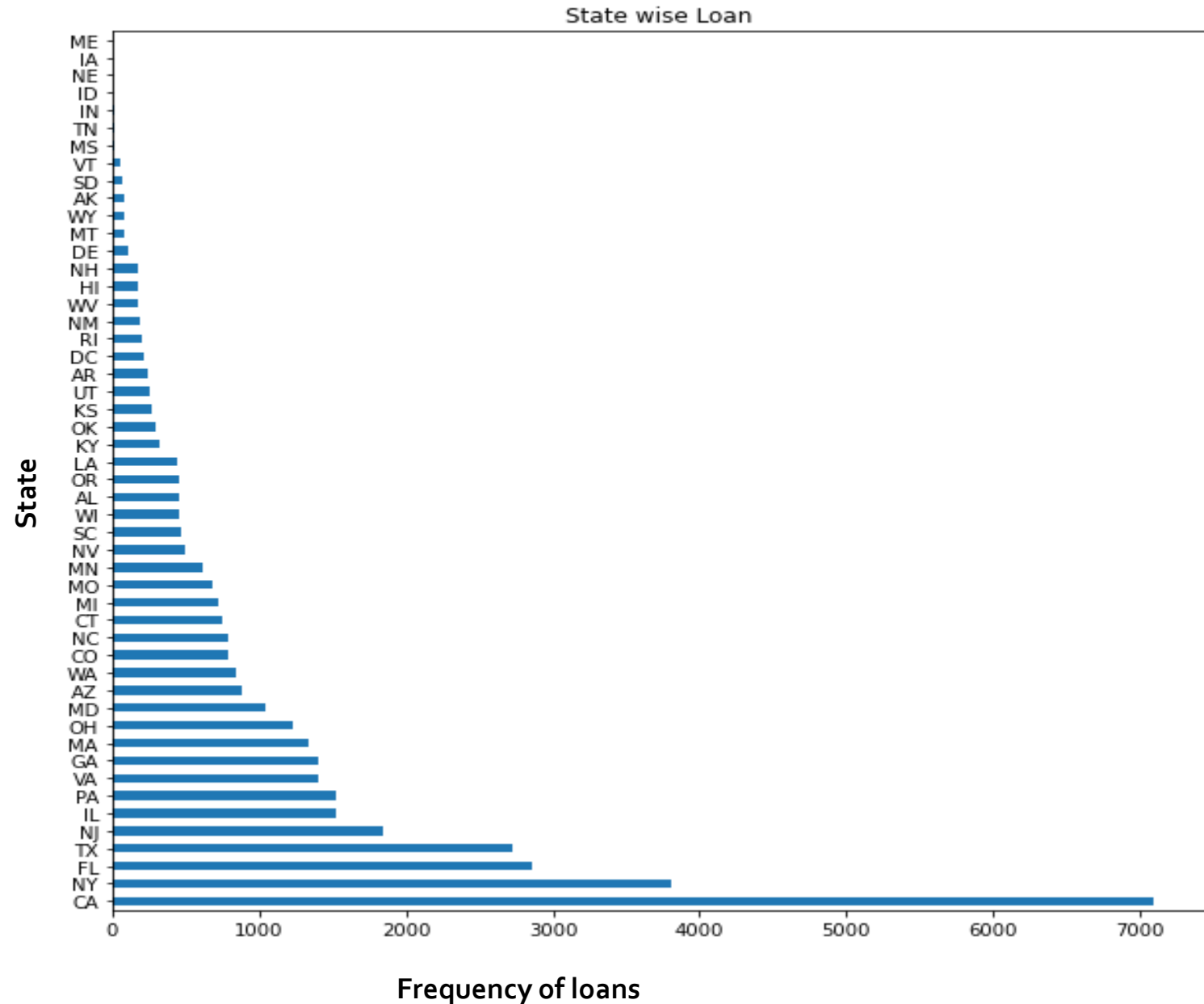
least amount of customers are for renewable energy



Address State

We can observe that California state has more number of loans compare to other state, almost double the amount of loans

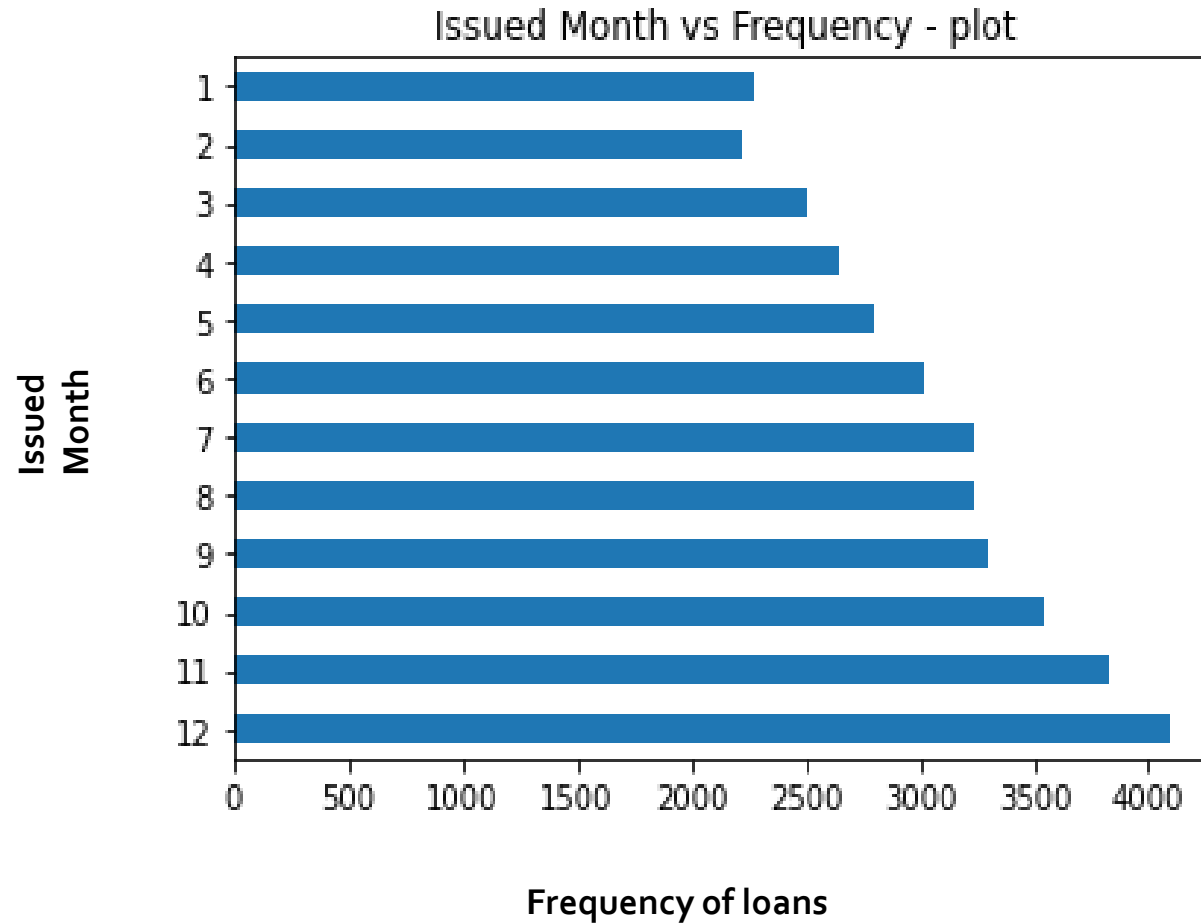
7 states(ME, NE, ID, IA, IN, TN & MS) are having very least loans



Issued Month

We can observe that there is a constant increase in number of loans issued by every month

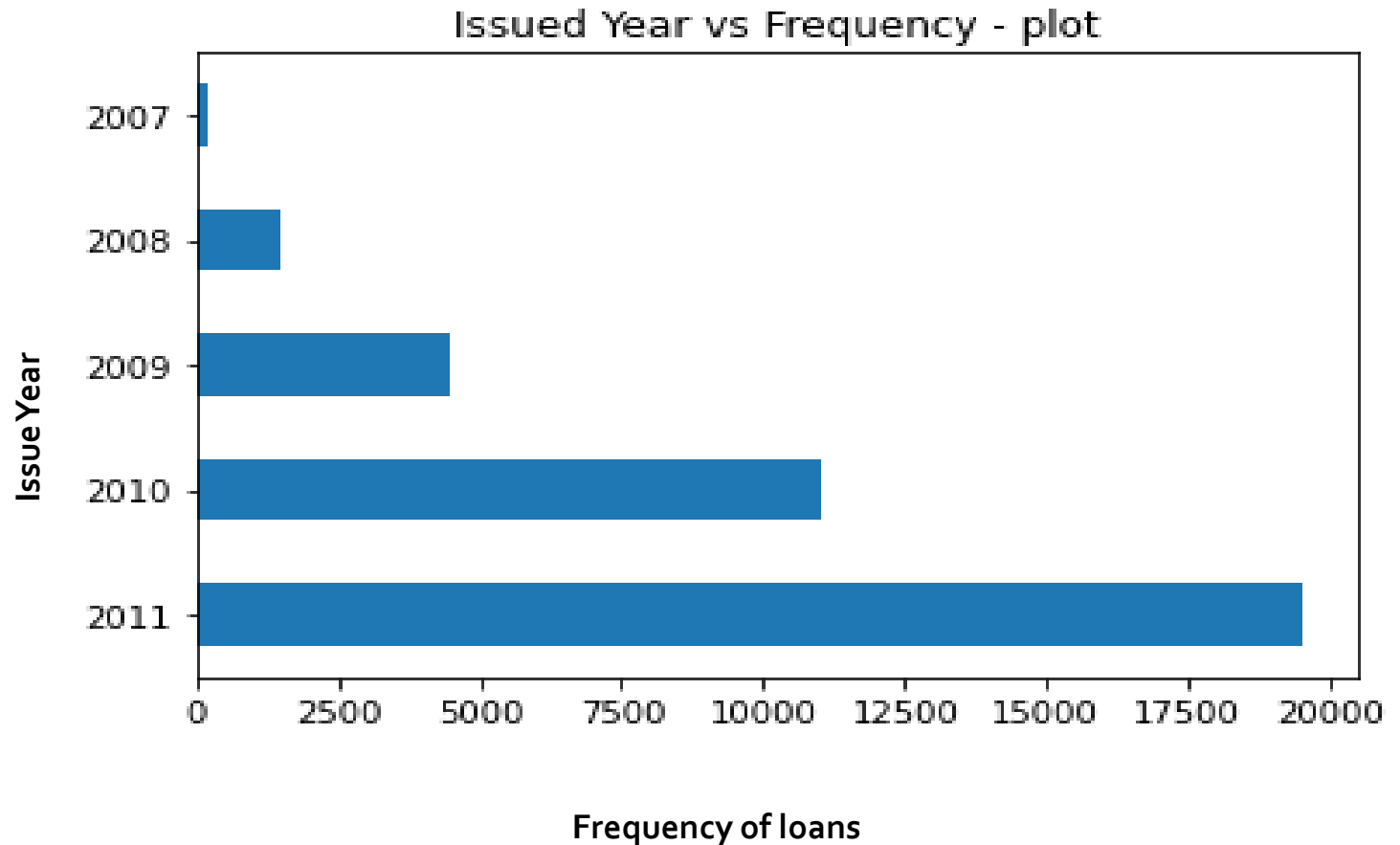
December (12) has the more number of loans



Issued Year

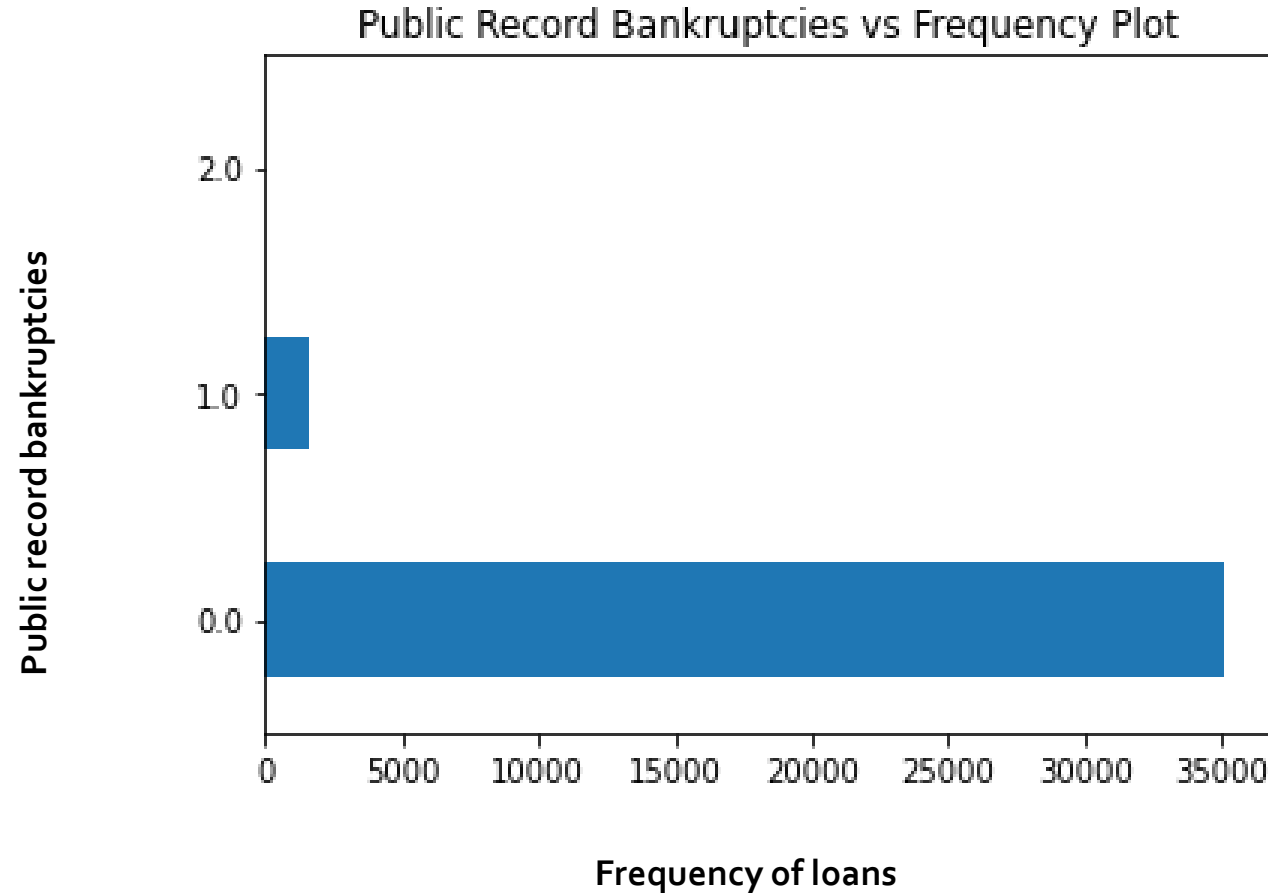
We can clearly see that number of loans issued are increased constantly (almost twice as previous year)

2011 has the more number of loans (around 50 % of the loans are from 2011)



Public Record Bankruptcies

Above 90 % of the customers have Zero bankruptcies and very few customer has a single record of bankruptcy and almost negligible amount of customers have 2 records of bankruptcies

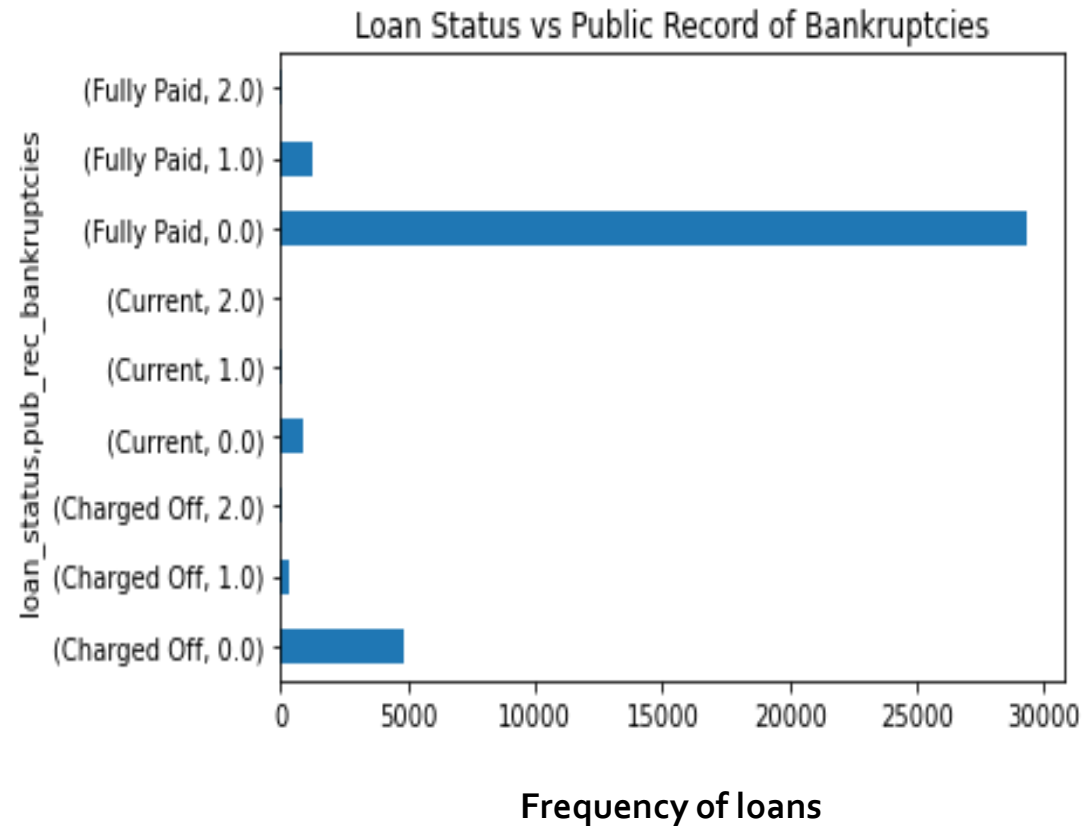


Results of Analysis – Univariate Results

Loan Status vs Public Record of Bankruptcies

We can see that very high number of loans are Fully paid and Zero public record bankruptcies

similar to that defaulted loans also high when Zero public record bankruptcies

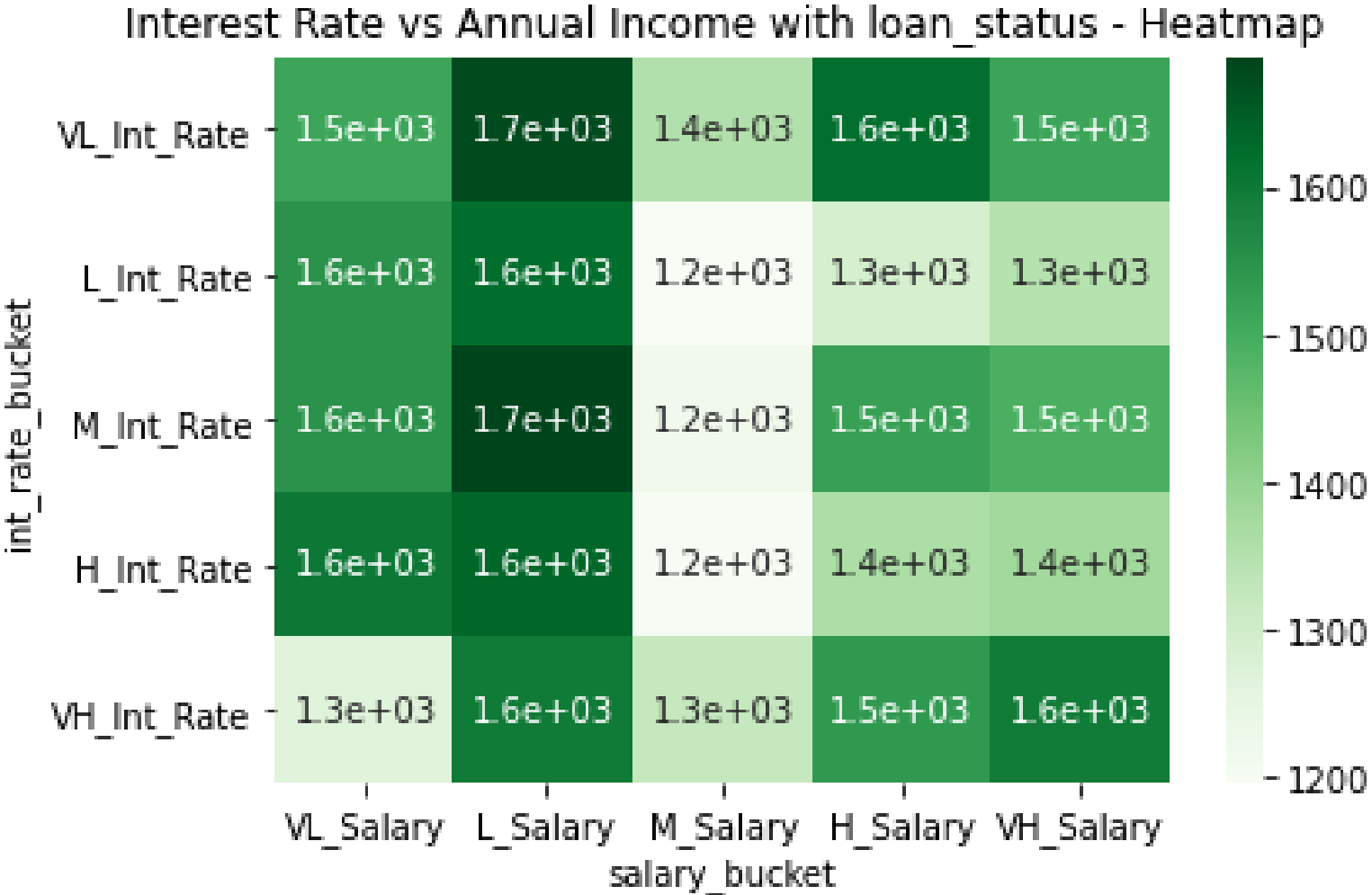


Interest Rate vs Annual Income

Very Low salary will increase the interest rate to high and very less relation with Very high rate of interest

Very High salary customer has more relation with very high and very low interest rate

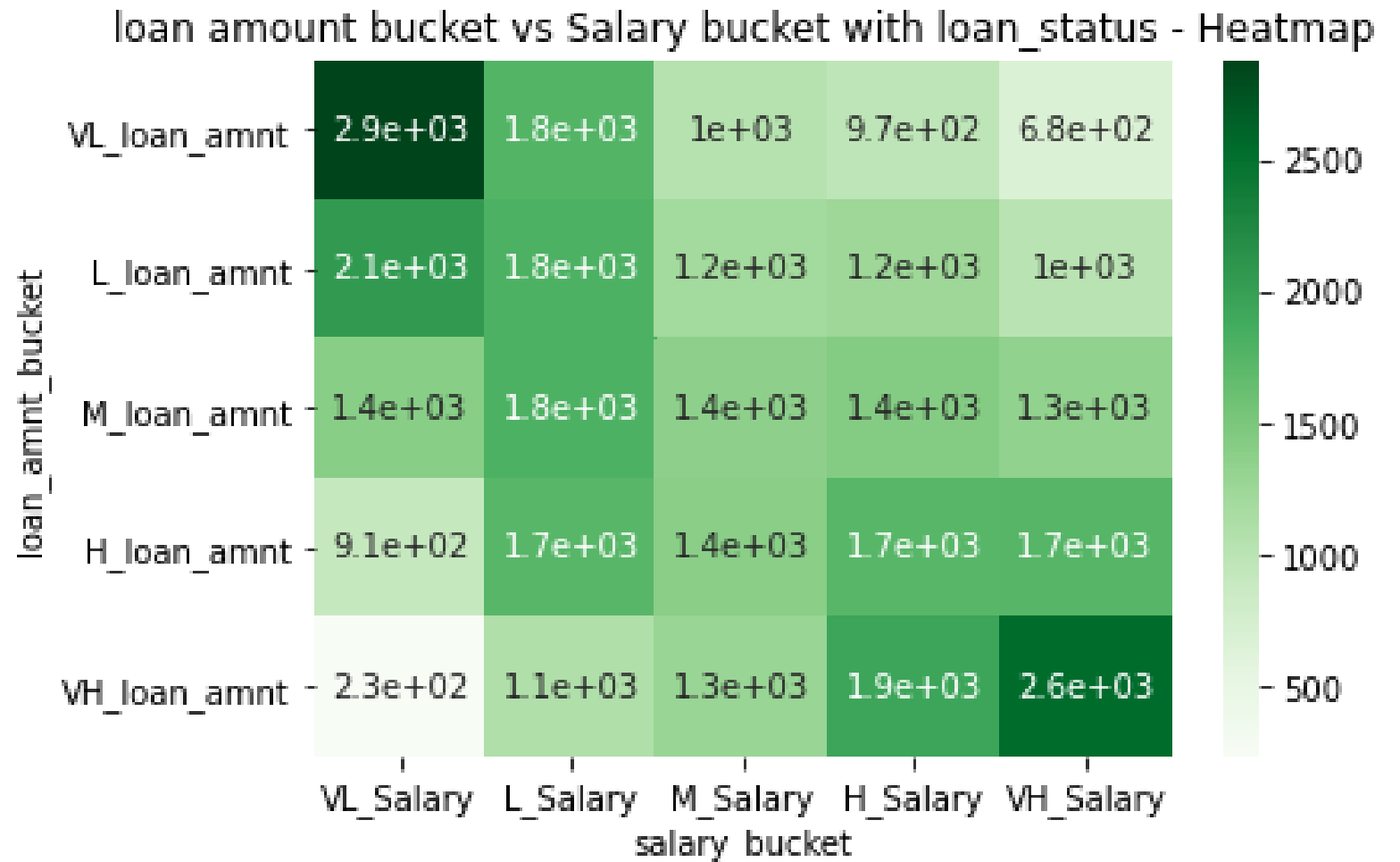
High salary package has the more change in very low interest rate



Loan Amount Bucket vs Salary Bucket

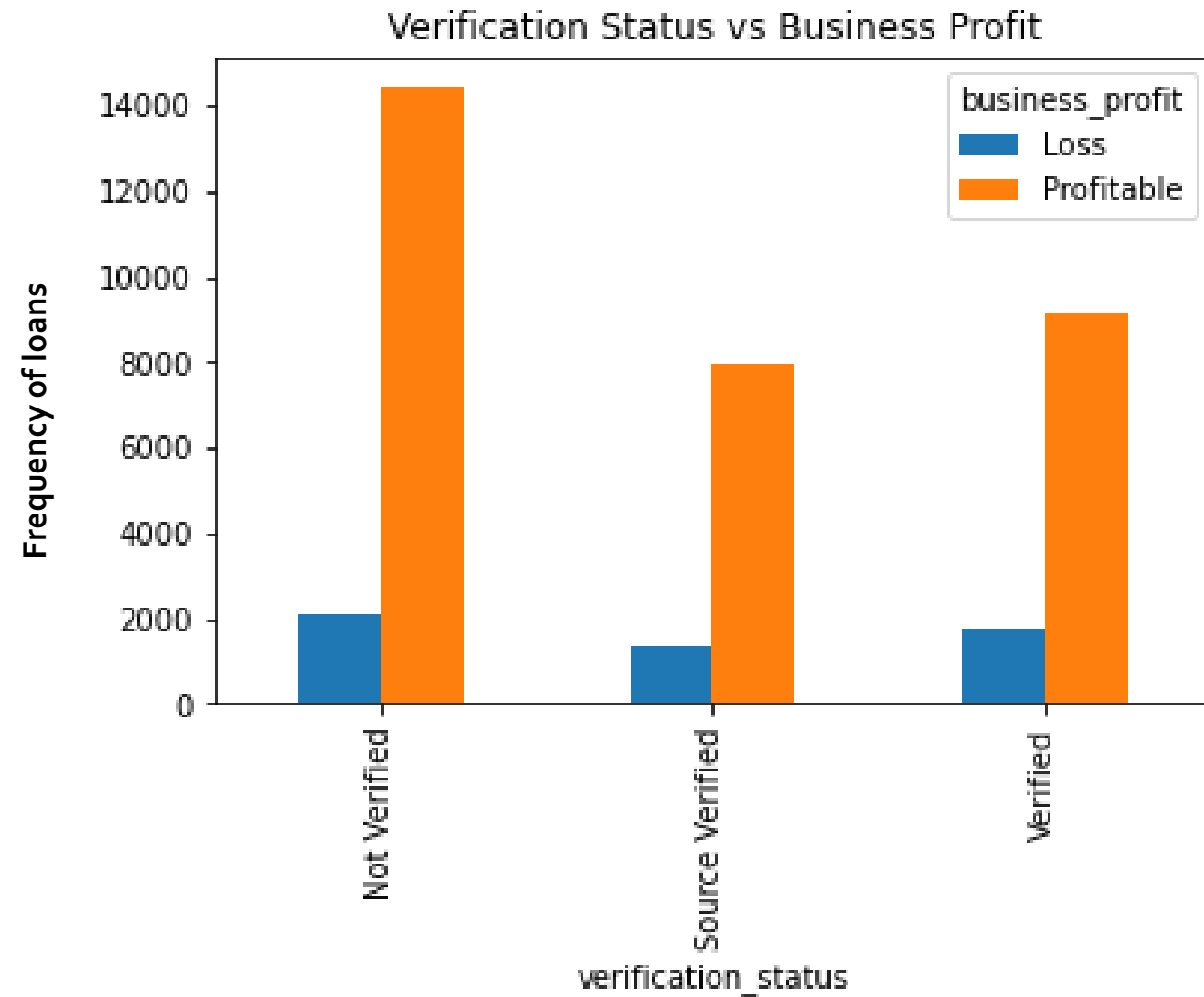
Very Low salary customer gets very low loan amount and less chances of very high loan

Very High salary customer has high chance of getting high and very high loan amount



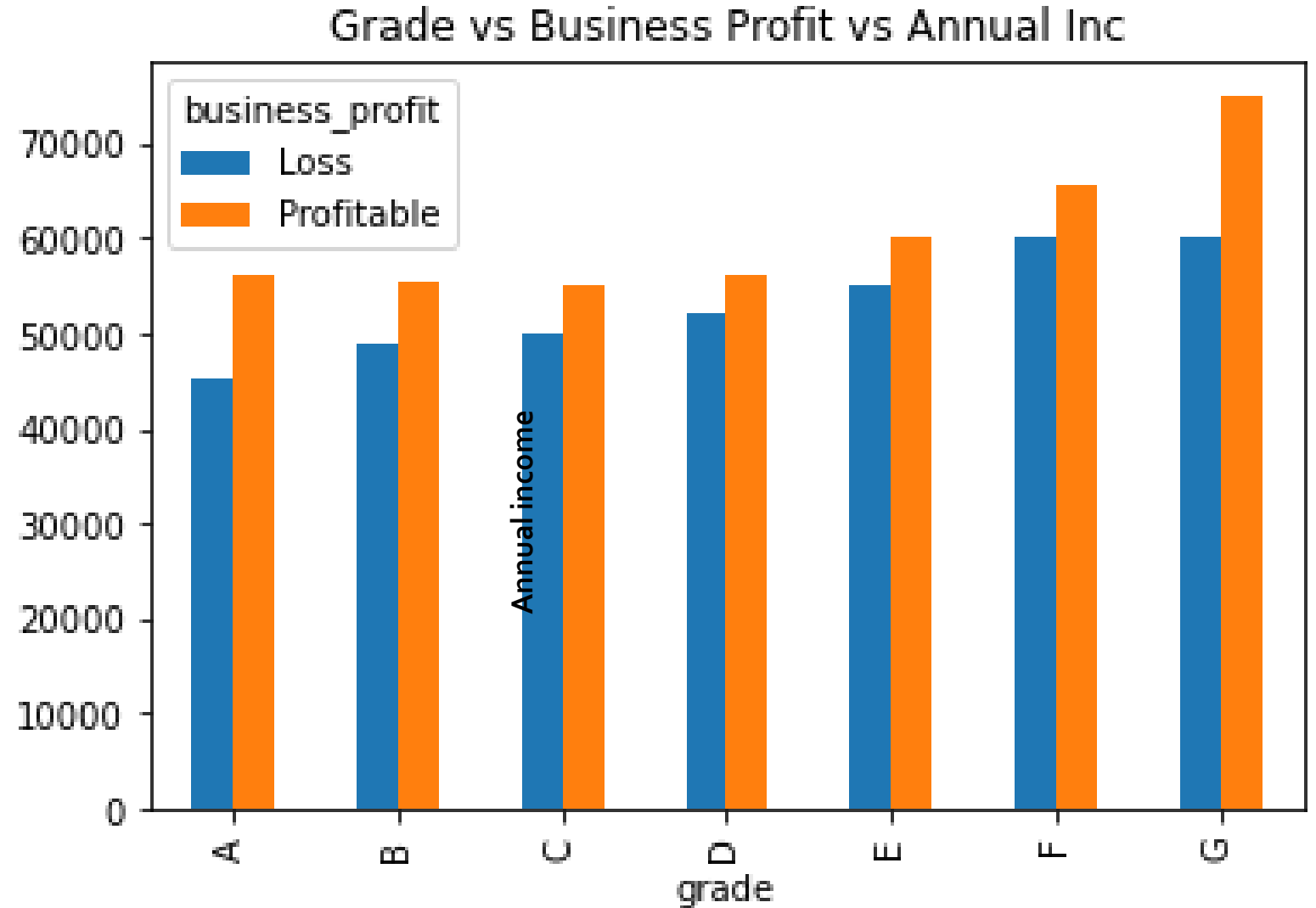
Verification Status vs Business Profit

We can observe that Most of the profit loans are of verified (sum of Source Verified and Verified)



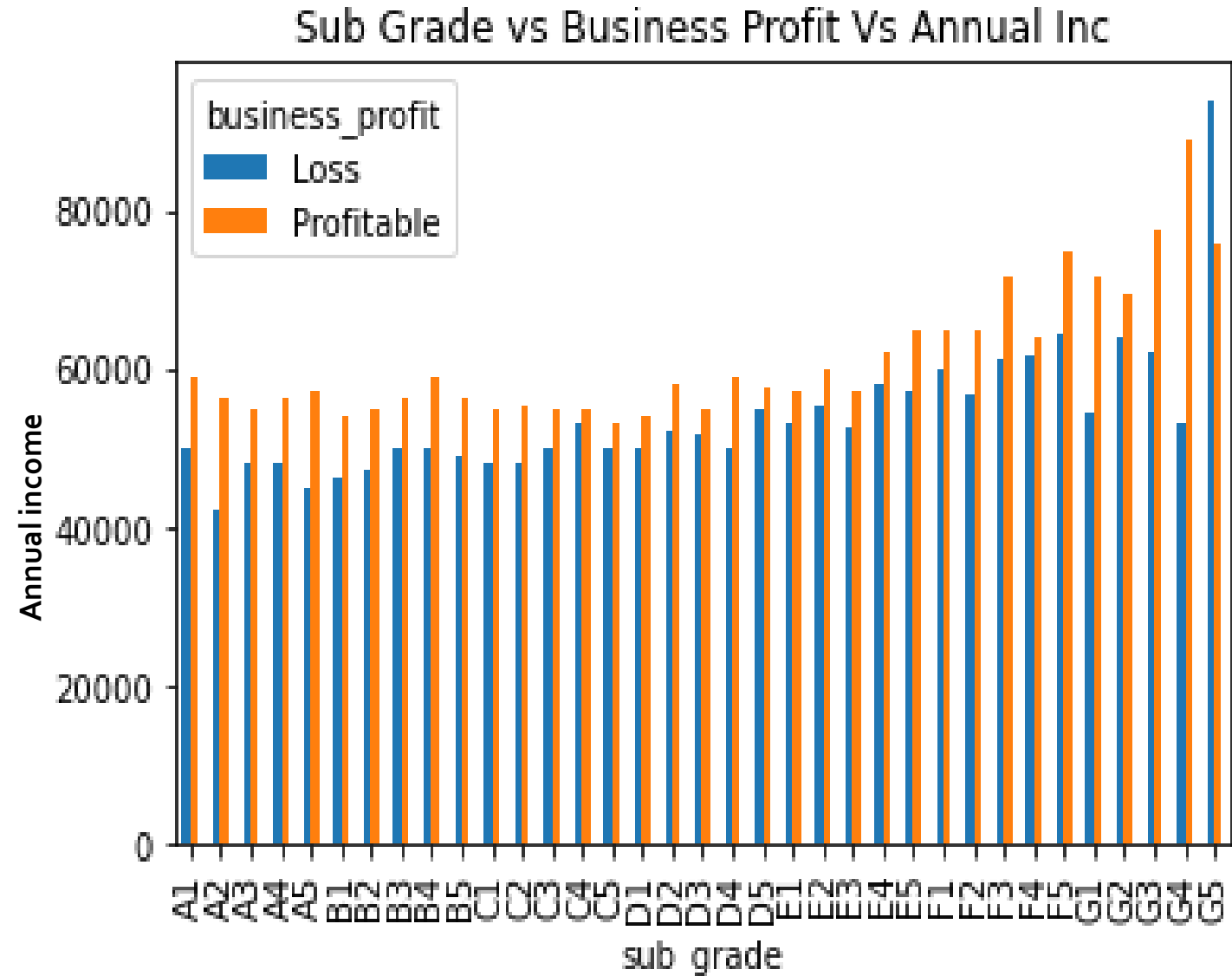
Grade vs Business Profit vs Annual Inc

There is a absolute dominance in profit compare to loss against the annual income, but even though profit is always higher compare to loss it's still Loss and Profit based on income is near values



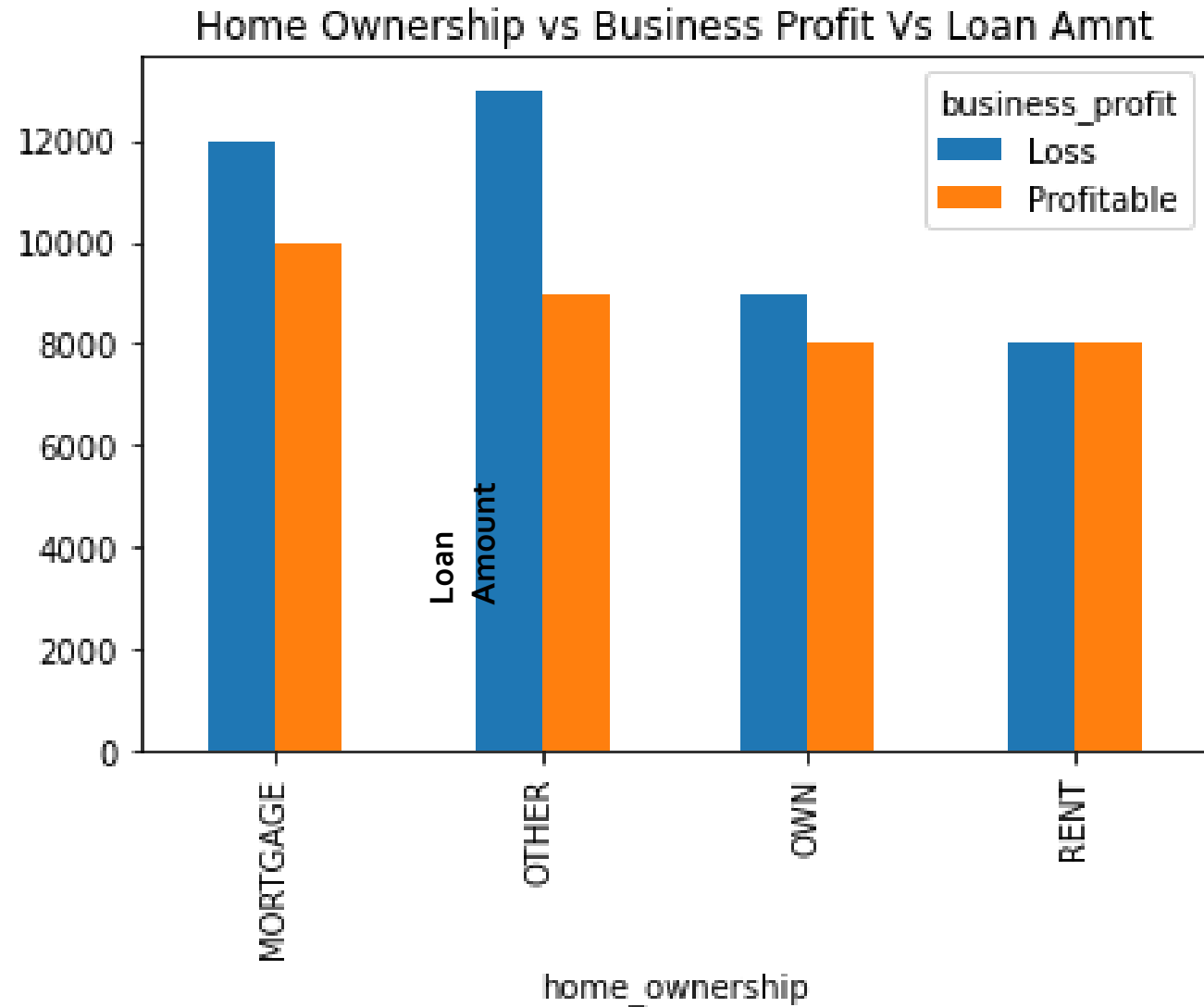
Sub Grade vs Business Profit

Here we can see that Loss is more for Sub Grade G5 with annual income more than 80K



Home Ownership vs Business Profit

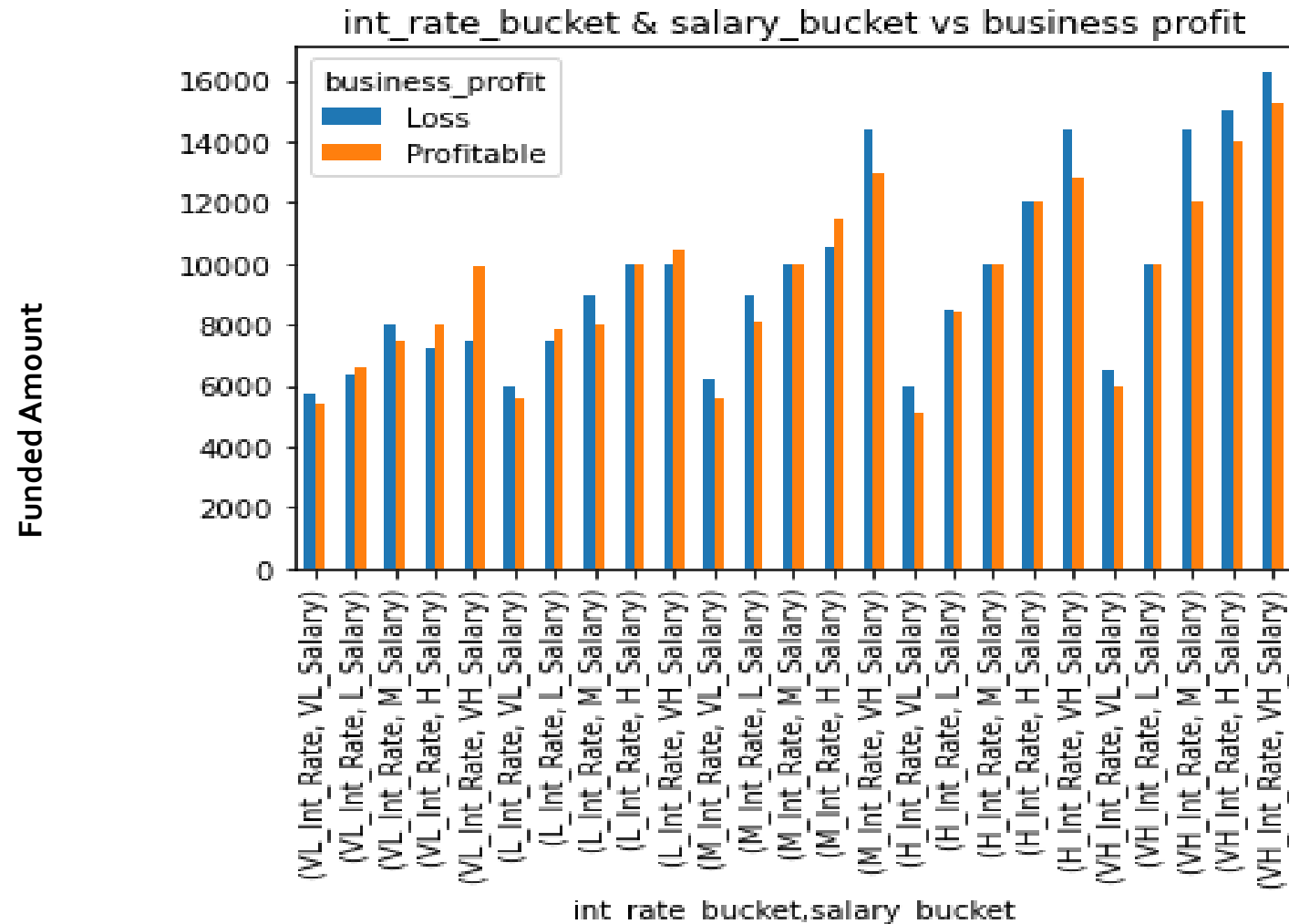
For above graph, now we can see that Loss is more than Profit when loan amount is more than 10K for Mortgage and Other ownership status



Int Rate Bucket & Salary Bucket vs Business Profit

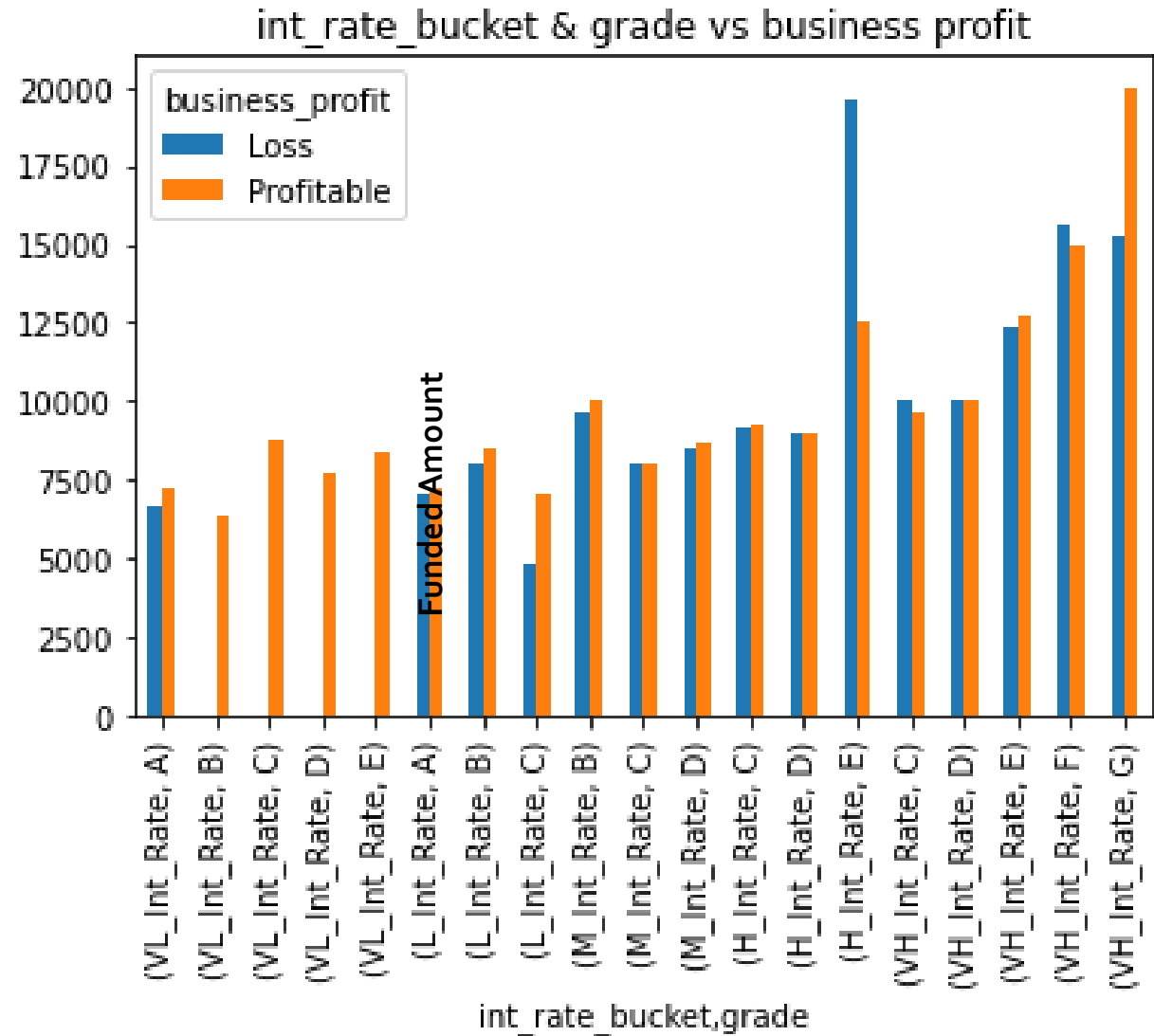
We can observe that Medium interest rate with Low salary, Medium salary, very high salary has the more chances of loss than profit

Very high interest rate with medium and high salary range customers with loan amount more than 12K loan amount are tends to become defaulter



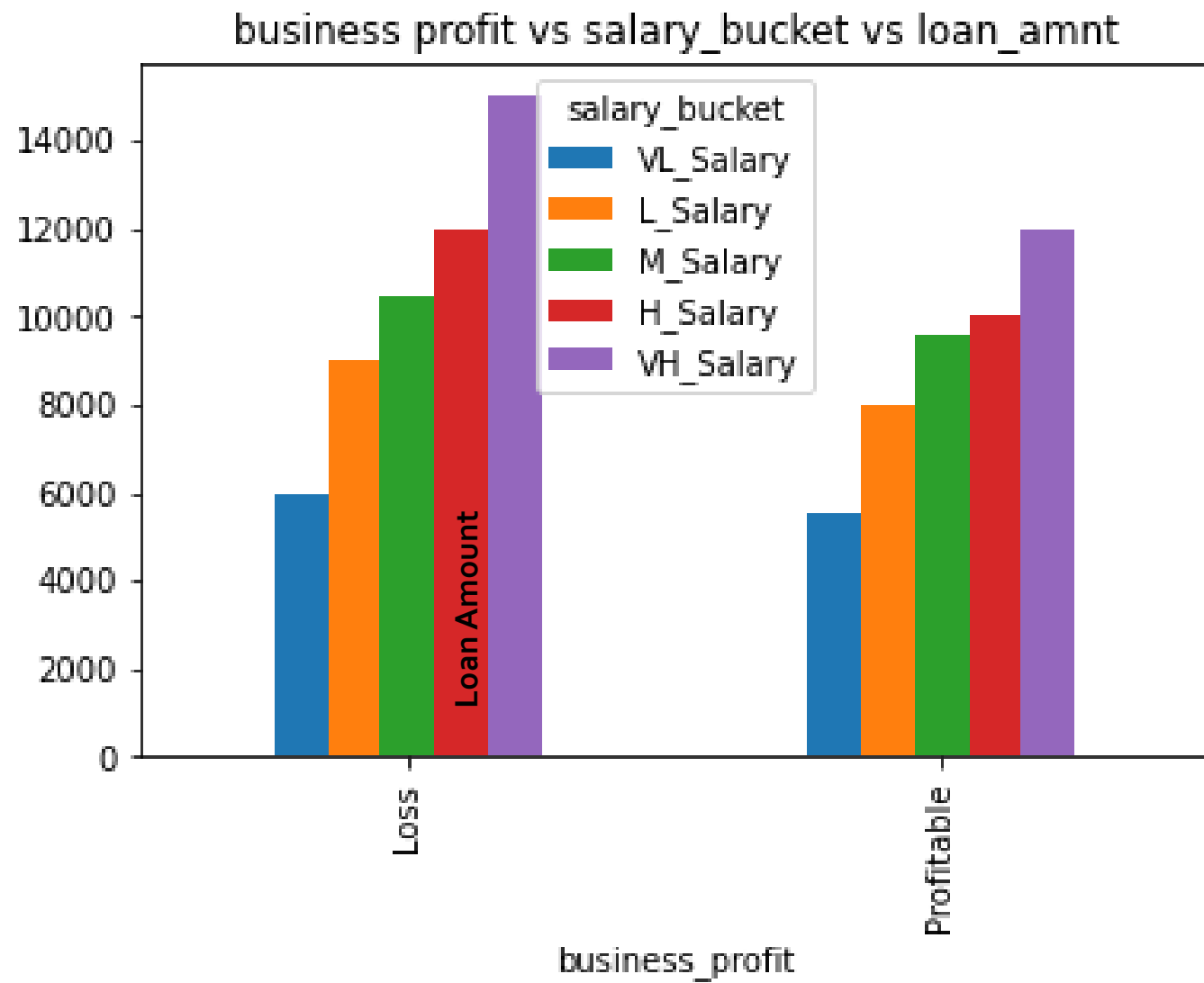
Int Rate Bucket & Grade vs Business Profit

We can clearly see there is almost zero loss for the loans with Very low interest rate for all grades except for Grade A



Business Profit vs Salary Bucket

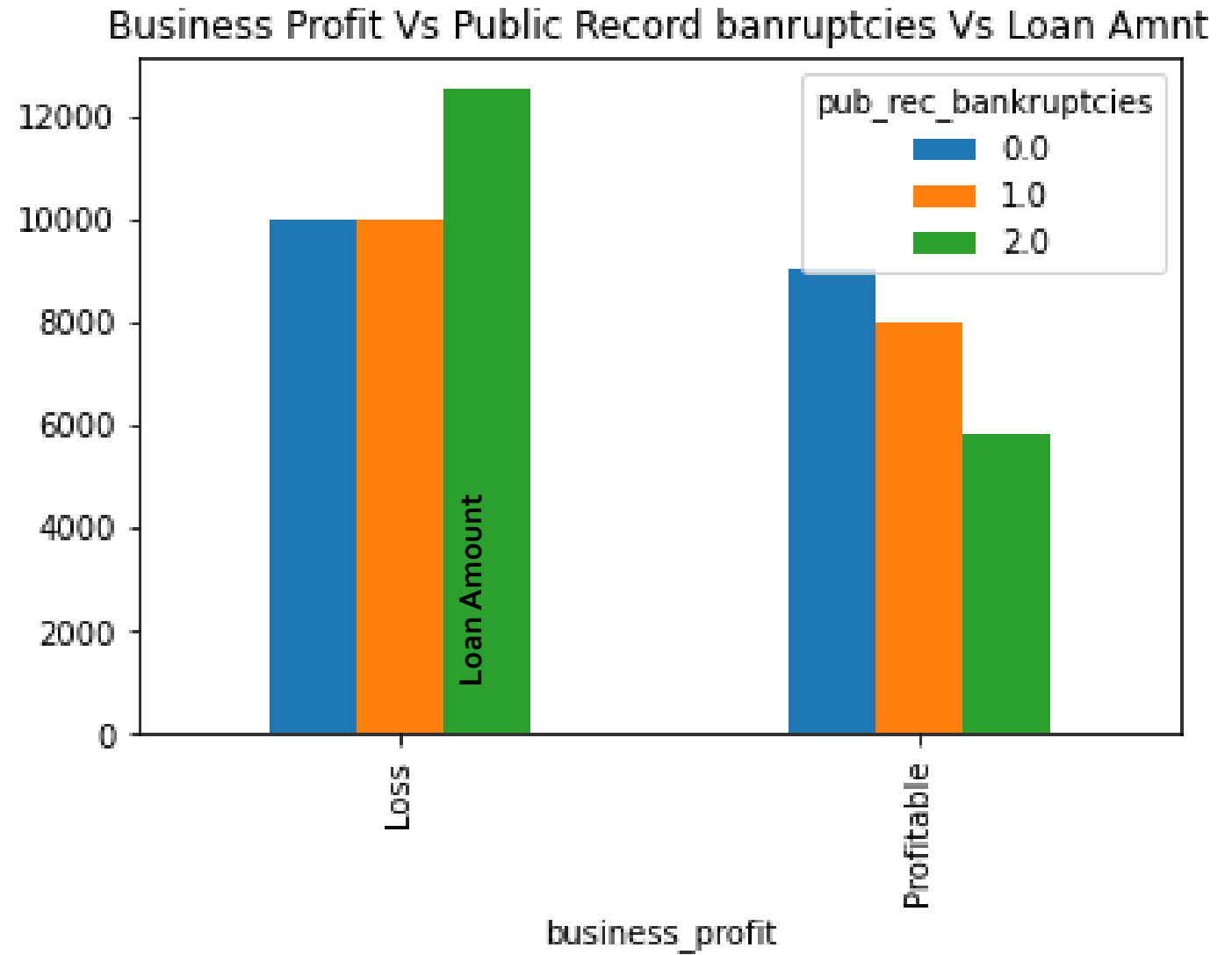
we can see that very high salary customer requesting 12K+ loan amount has the high change of Business Loss



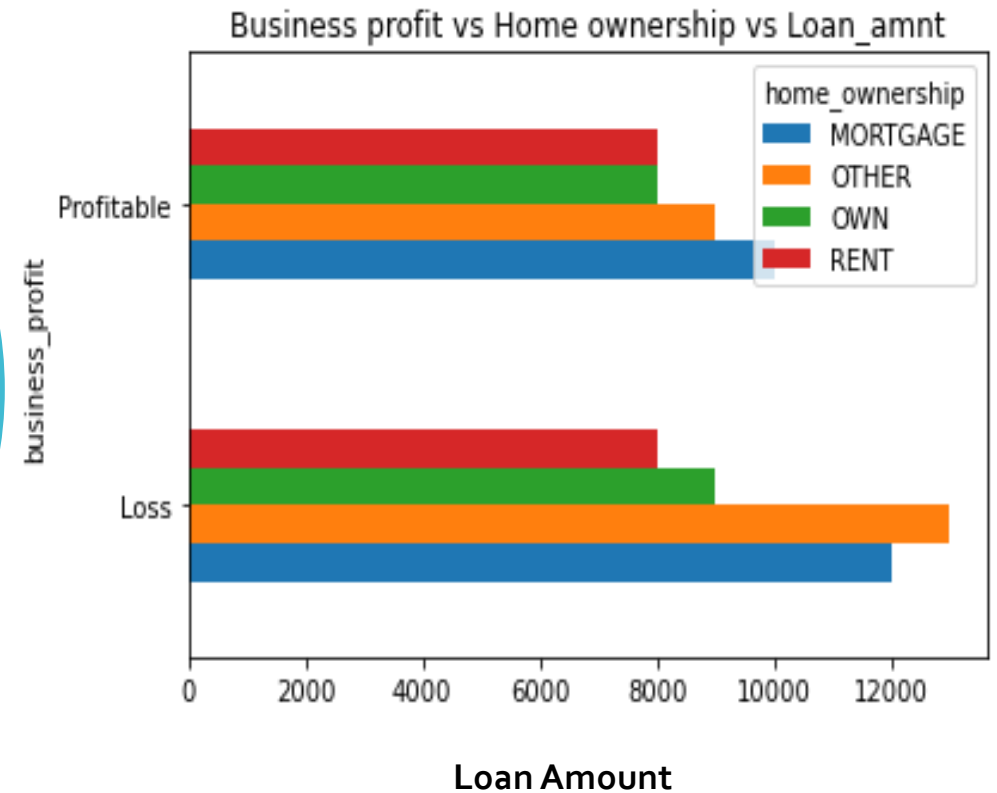
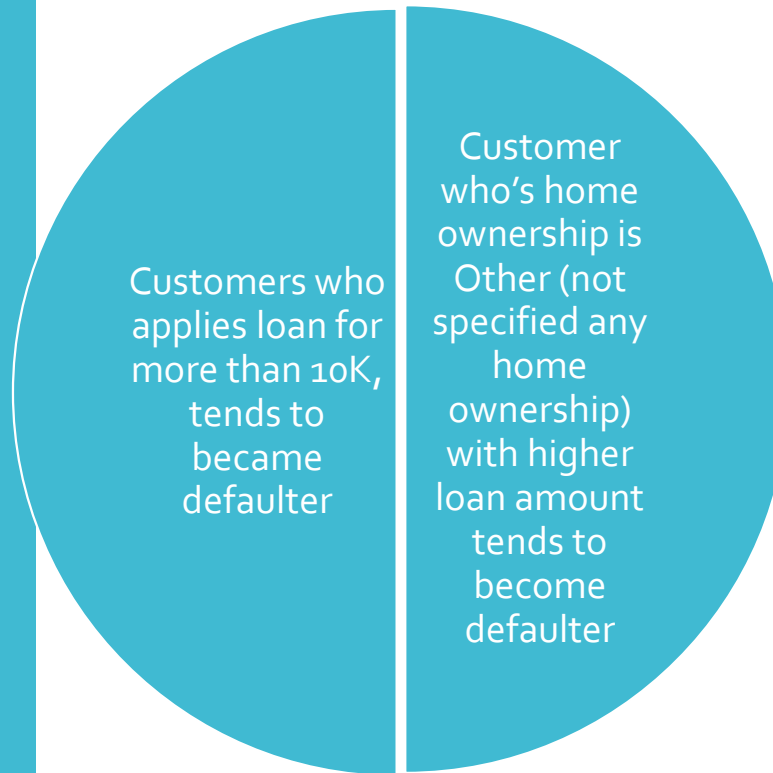
Business Profit Vs Public Record bankruptcies

From the graph we can see that a customer with 2 bankruptcies with loan amount more than 8k+ has more change of business loss or defaulter

The lower the loan amount the higher the chances in profits



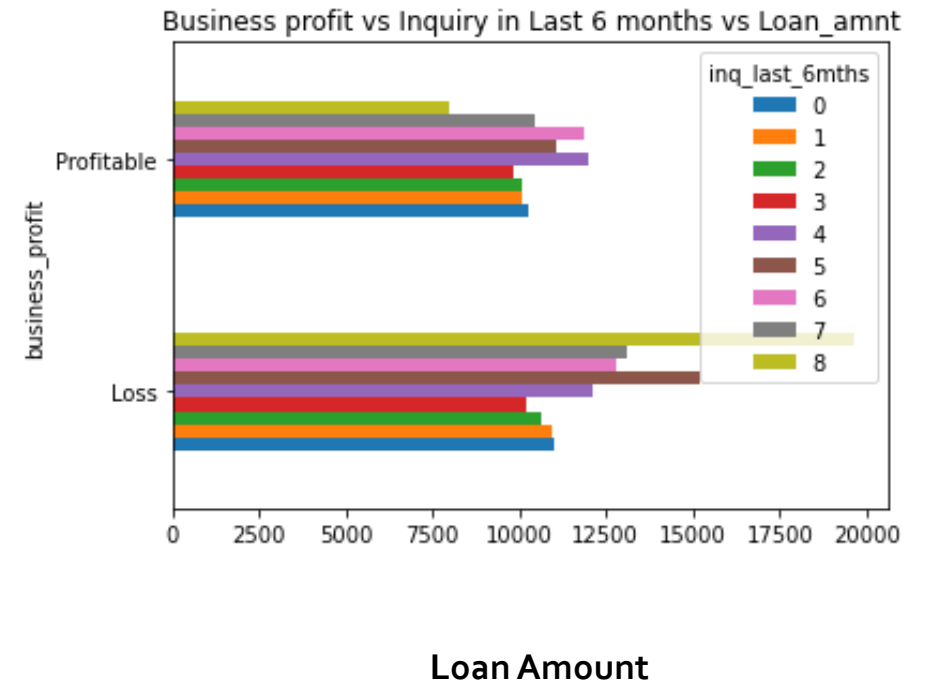
Business profit vs Home ownership



Business profit vs Inquiry in Last 6 months

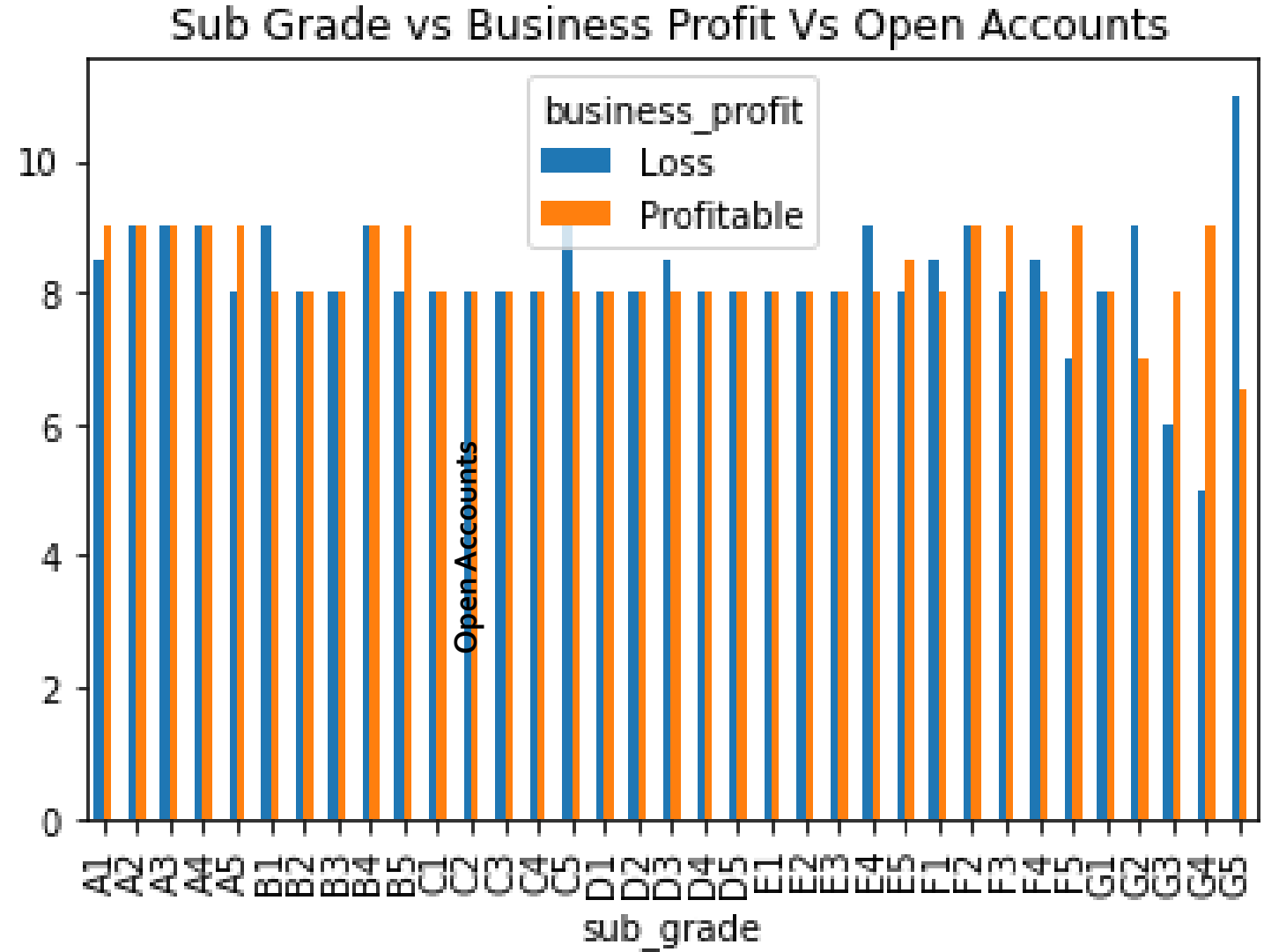
Another thing we can observe that customer with more than 5 inquiries in last 6 months has more chances of becoming defaulter

We can observe that customer with 8 inquiries in last 6 months has high chance of become defaulter, and the chance of becoming defaulter increases by higher loan amount



Sub Grade vs Business Profit Vs Open Accounts

Based on above graph we can observe that Customer with more than 8 account with Sub Grade G5, has more chance of becoming a Loss i.e. Defaulter



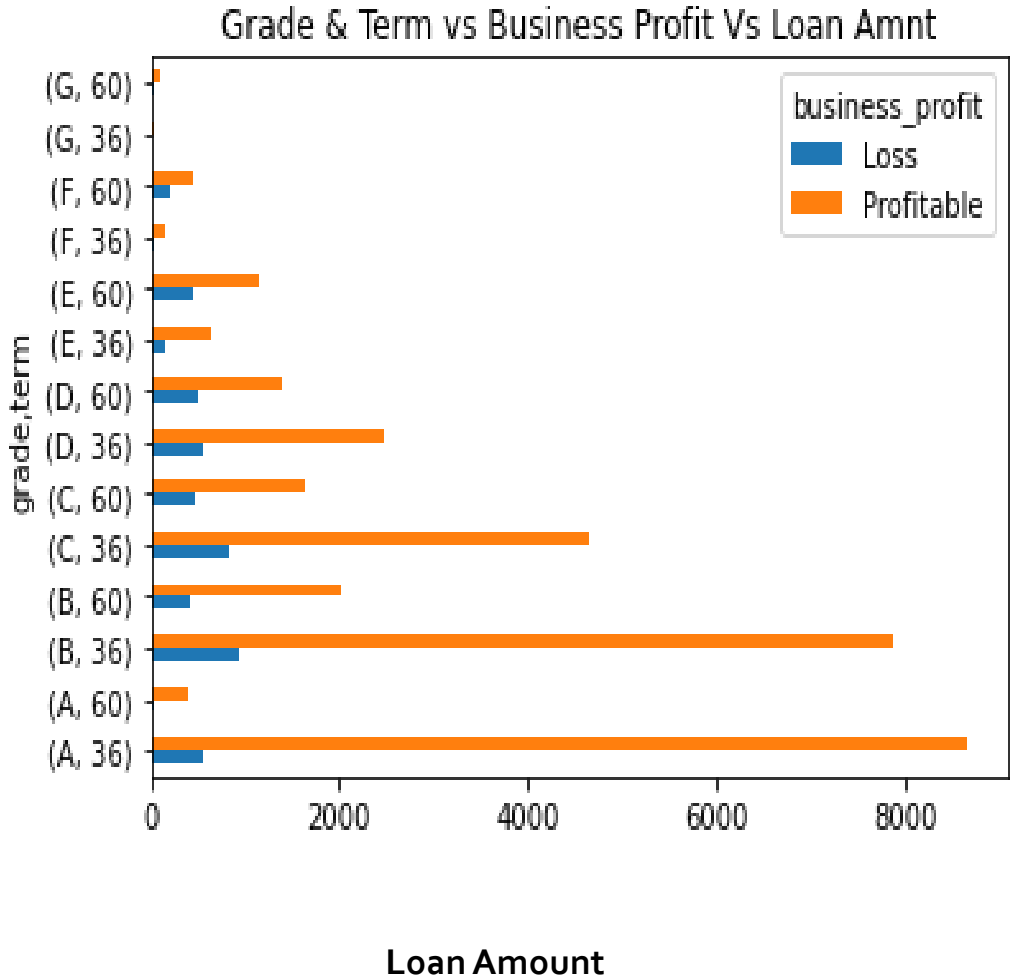
Sub Grade vs Business Profit Vs Open Accounts



FROM ABOVE PLOT WE CAN SEE THAT CUSTOMER WITH A GRADE AND 60 MONTH TERM PERIOD LOAN HAS NEGLIGIBLE LOSS AND WE CAN CONSIDER IT AS PROFIT WITHOUT LOSS



SAME FOR CUSTOMER WITH G GRADE WITH 60 MONTHS TERM & F GRADE WITH 36 MONTH TERM HAS NEGLIGIBLE LOSS





Customer with below condition are tend be become defaulter,

G5 grade with 8 Open accounts

8 inquiries in last 6 months

Home ownership status not given as well as Mortgage with 10K+ applied loan amount

2 Bankruptcies with loan amount more than 8k+

Very high salary with 12K+ applied loan amount



Customer with below condition are having high chance of profit to business,

Very low interest rate for all grades except for Grade A

A grade and 60 month term period loan

G grade with 60 months term & F Grade with 36 month term

Conclusion