Gramener

EDA Case Study

Shubham Agarwal
Vaibhav Jain
Keshav Yadav
Alaparthi Vijaya Lakshmi Sai

Background:

- The company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.
- Borrowers can easily access lower interest rate loans through a fast online interface.

Objective:

- To find driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.
- This knowledge can be utilized for company’s portfolio and risk assessment.

Analysis Approach

Go through the Data Set and Understand Data Set and meaning of different factors

Remove Columns which are not important

Find/Drive Important variables(Columns) on which data can be segregated

Analysis of defaulters vs Total Loans(Defaulters and Full Paid) based on single variable

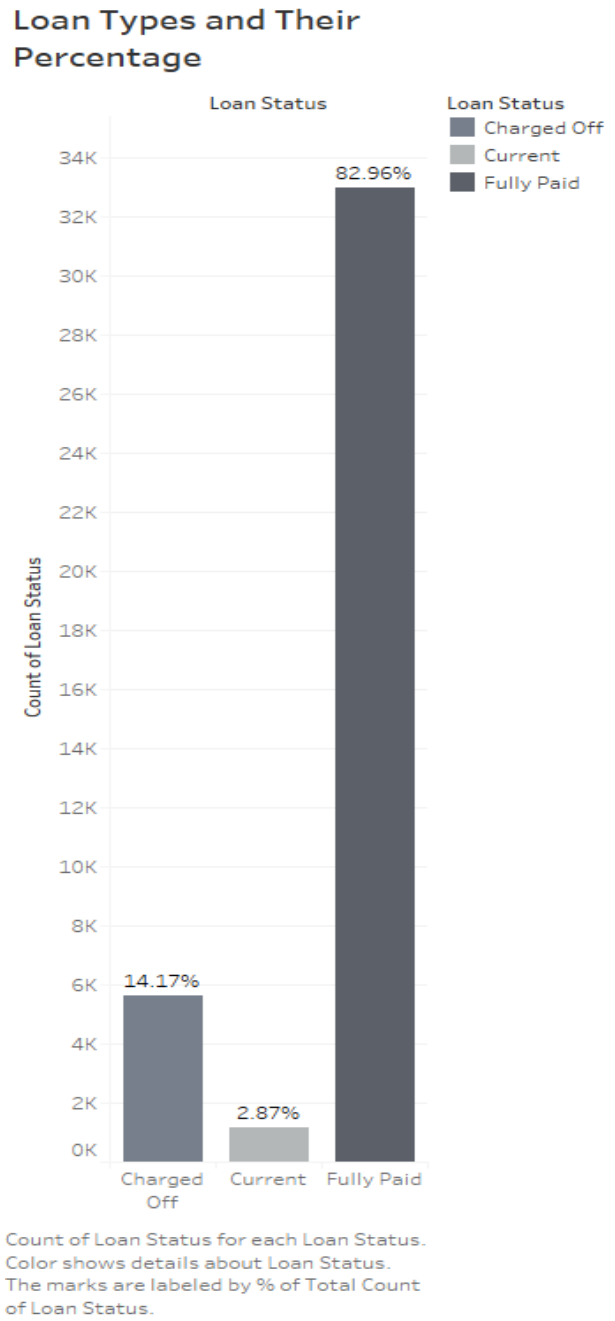
Analysis of defaulters vs Total Loans(Defaulters and Full Paid) based on different variable

Identify at least the 5 important driver variables (i.e. variables) which are strong indicators of default.

Understanding Data

Category	Loan Approved	Percentage
Charged Off	5627	14.17%
Current	1140	2.87%
Fully Paid	32950	82.96%
Total	39717	100

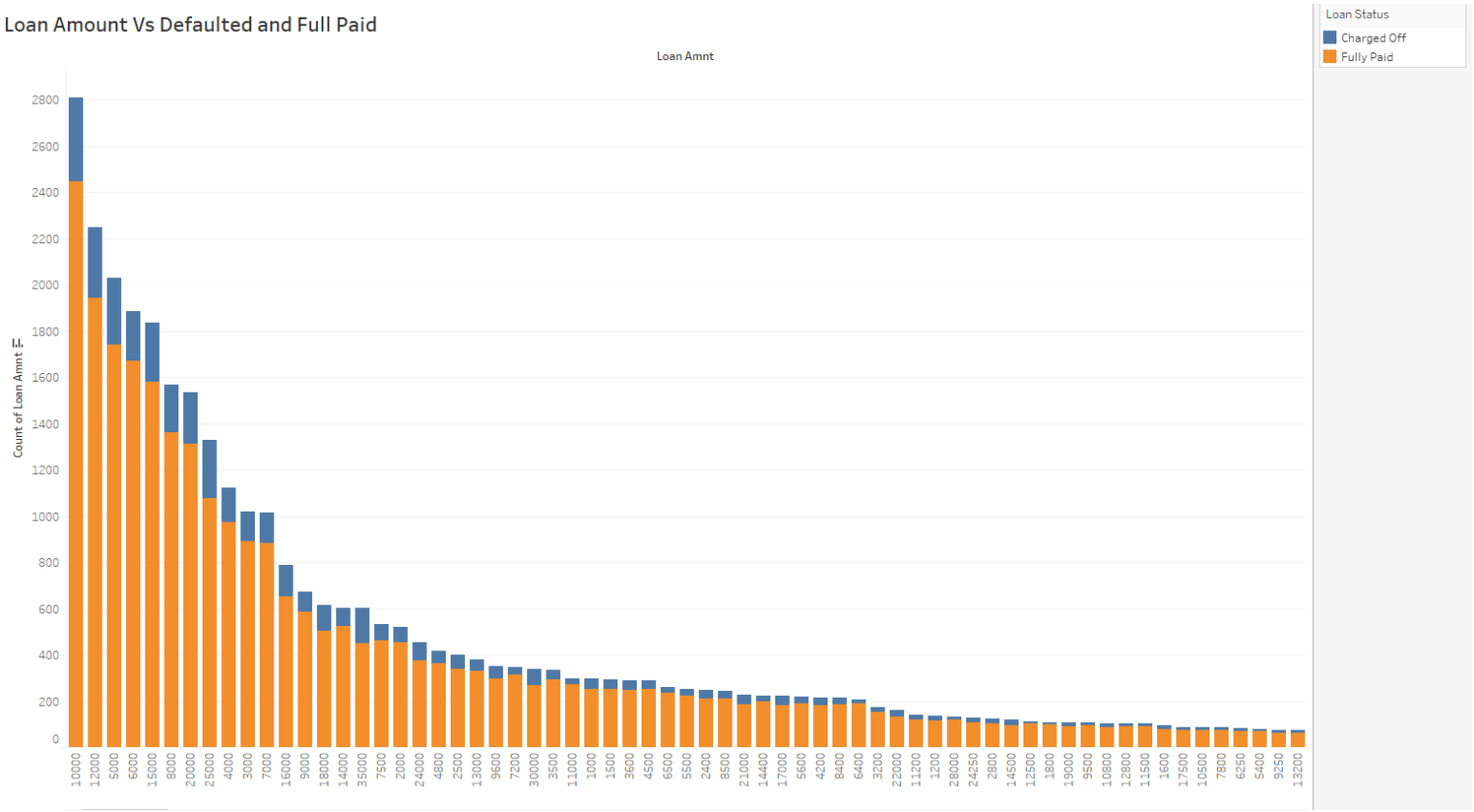
Thus there is around 14.17% loan is defaulted.



Univariate and Bivariate Analysis

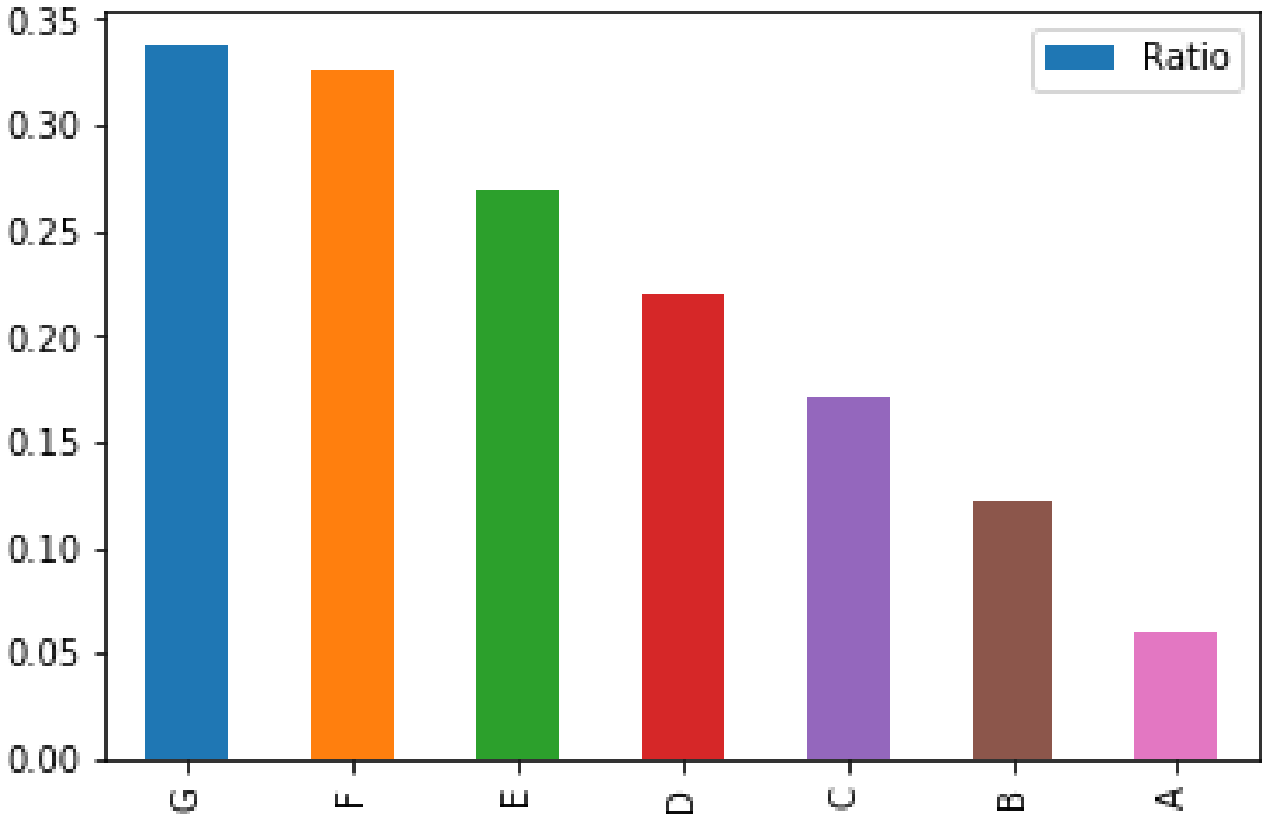
- Now we try to measure effect of each variable (column) on the loan status. Thus we check if any individual factor works as driving factor for loan default.

- We did analysis of ratio of count of defaulted loans to sum of defaulters and paid off loans. Thus we got clear idea about how much of total loans is defaulted.



- First we did univariate analysis of Loan Amount(loan_amnt). We found that its clear that :
- Ration of defaulters to total loans is higher for following Loan amounts: 25500, 100500, 1550 but we also noticed that these are the same amounts for which highest numbers of loans are granted.
- So, we could not find any significant co-relation between Loan amount and defaulted loans.

Univariate Analysis with Loan Amount



Graph showing Defaulted Loans Vs All Loans(Defaulted + Fully Paid)

- Grade is an important factor while driving loan resultant to be default.
- Ration of defaulters to total loans is higher for following grades:

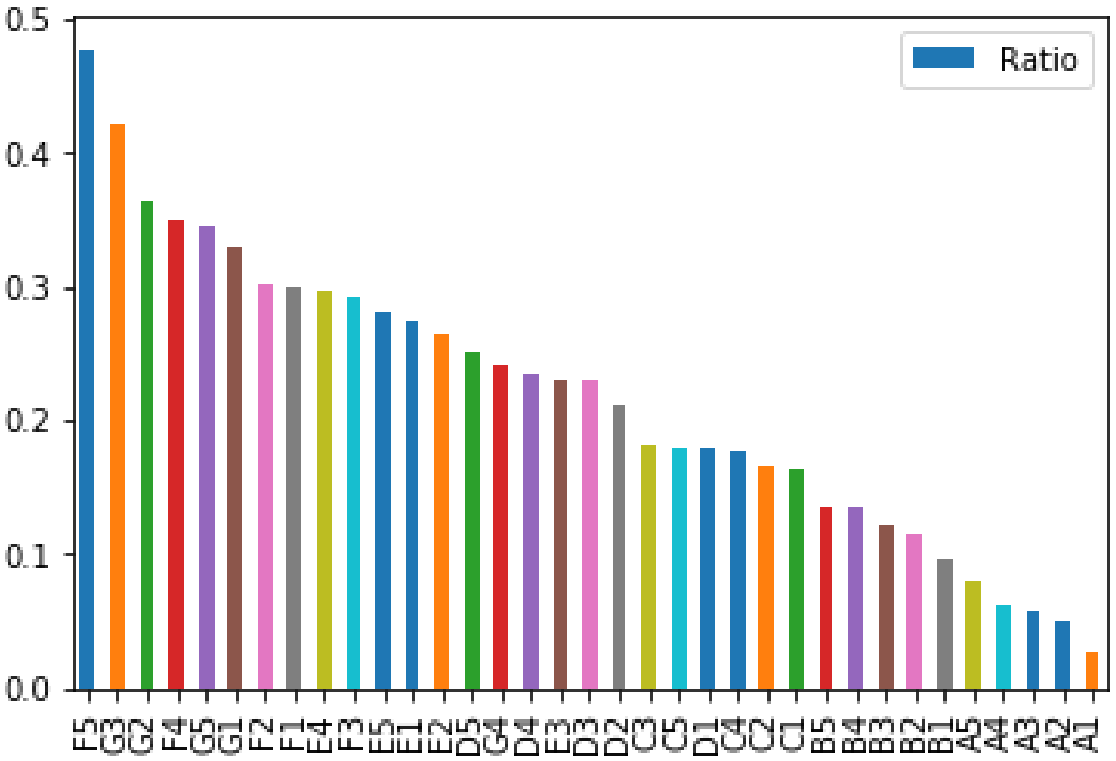
➢ Grade G : 34% Loan Given to these grade get defaulted

➢ Grade F : 33% Loan Given to these grade get defaulted

➢ Also we should note that preference should be given to those who have Grade A and Grade B.

- So, we fond significant co-relation between Grade and defaulted loans.

Conclusion from Sub-grades graphs:

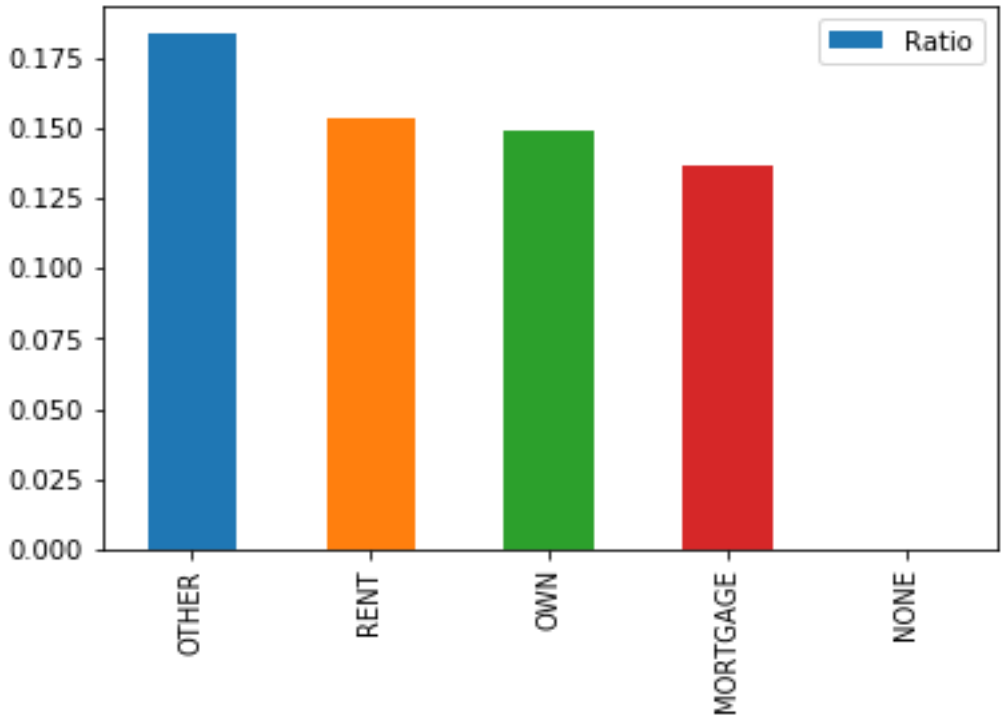


Graph showing Descending Order of Sub-Grades to which Loans are issued and resulted default.

From Analysis of Sub-grades it became clear that top defaulters exist in **Fx** and **Gx** sub-grades

Thus **sub-grade** variable **(i.e. sub_grade column)** is also a driving factor.

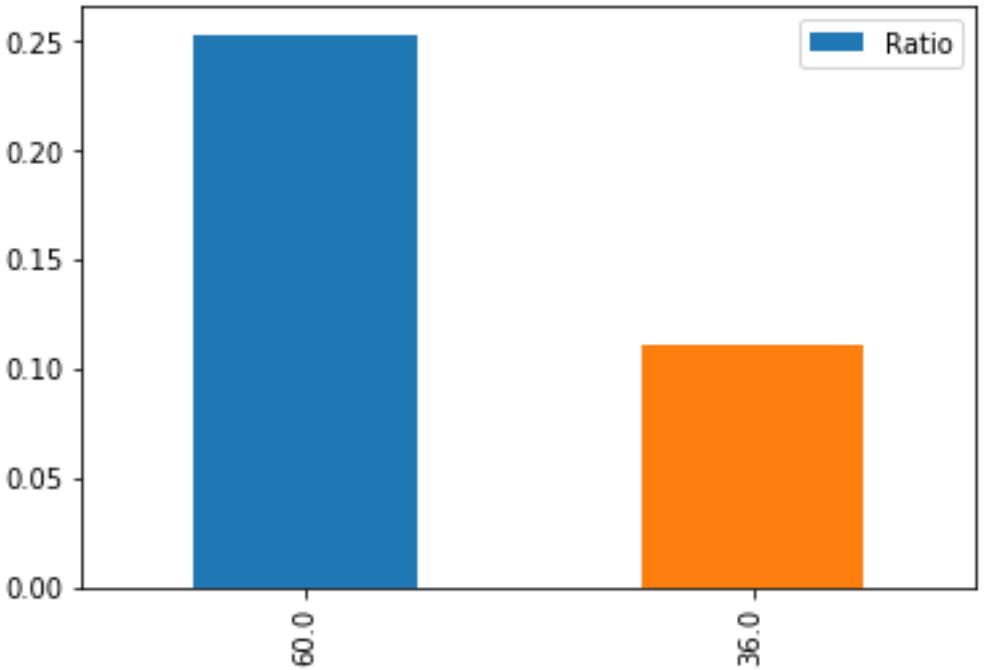
Conclusion from Home Ownership graphs:



Graph showing Descending Order of Home Ownership to which Loans are issued and resulted default.

- From Analysis of Home Ownership it became clear that yet most defaulters exist in “OTHERS” category.
- Thus Home Ownership variable (i.e. **home_ownership** column) is also a driving factor.

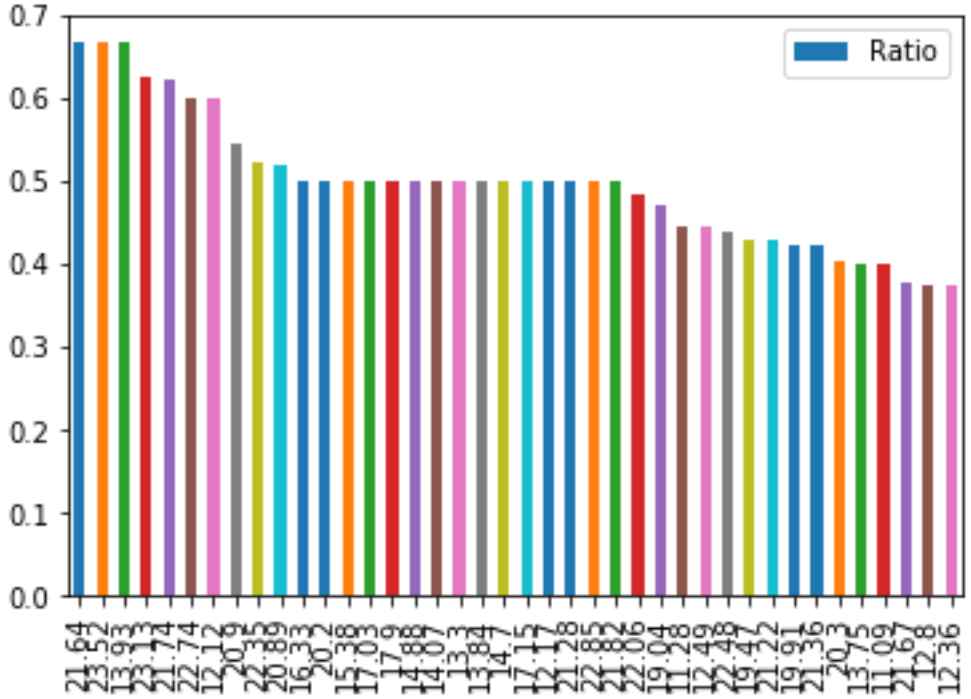
Conclusion from Loan Term graphs:



Graph Showing Loan Term(Loan Period) to which Loans are issued and resulted default.

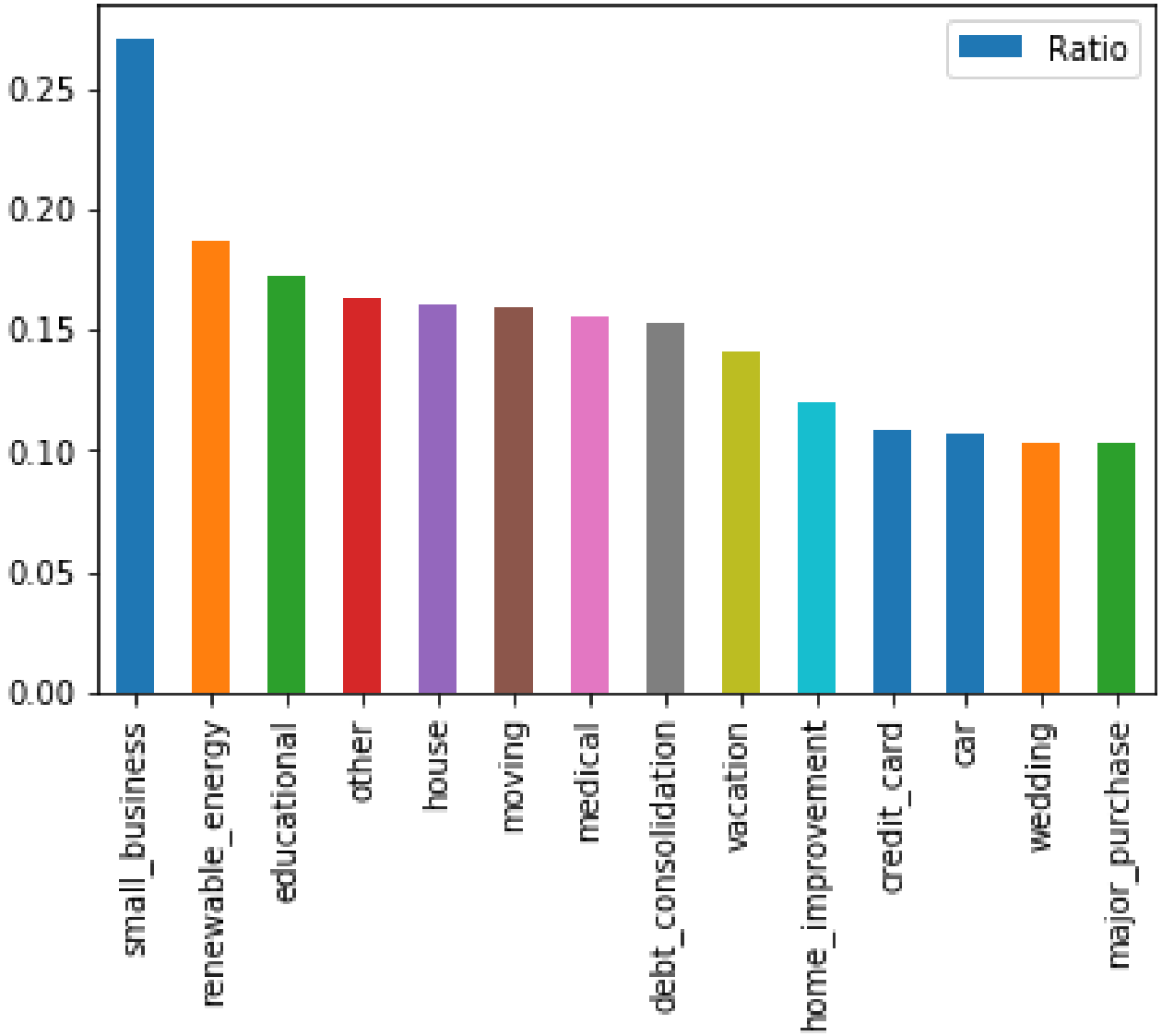
- From Analysis of Loan Term it is clear that top defaulters exist in “60 months” term and the difference is significant.
- Thus Loan Term variable (i.e. **term** column) is also a driving factor.

Conclusion from Loan Term graphs:



Graph Showing Interest Rate which Loans are issued and resulted default.

- From Analysis of Interest Rate graph it is clear that top defaulters exist in some portion like above 15% and for some percentage like 21.64 defaulters are around 65% of people who got loan at that rate.
- Thus Interest Rate variable (i.e. **int_rate** column) is also a driving factor.



- From Analysis of Loan Purpose it became clear that it has most significant driving feature.
- Most Loan defaulters exist in small_business, renewable_energy
- Thus Loan Purpose variable (i.e. **purpose** column) is also a driving factor.

Sector	No of Loans Distributed
debt_consolidation	18641
credit_card	5130
other	3993
home_improvement	2976
major_purchase	2187
small_business	1828
car	1549
wedding	947
medical	693
moving	583
vacation	381
house	381
educational	325
renewable_energy	103

- Till now we found driving factors for defaulted loans using single variable i.e. univariate analysis. These factors can be much dependent over some other significant factor that multiply this significance level.

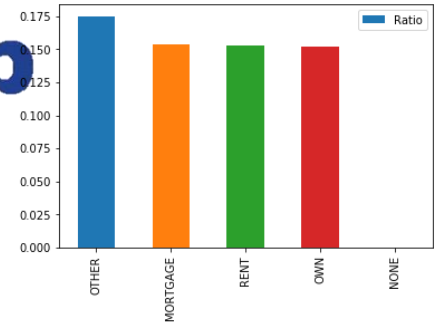
- We found that company performed most of its business in 5-6 fields(excluding others) of loans. Therefore we can extend our findings in these field.

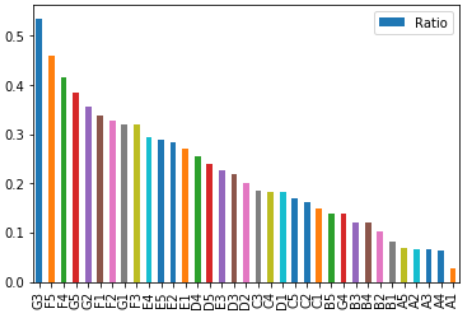
- Therefore we go for Bivariate analysis.

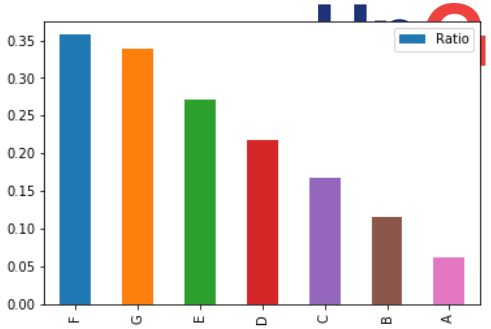
- For that we first find those sectors in which company distributed most of its loans.

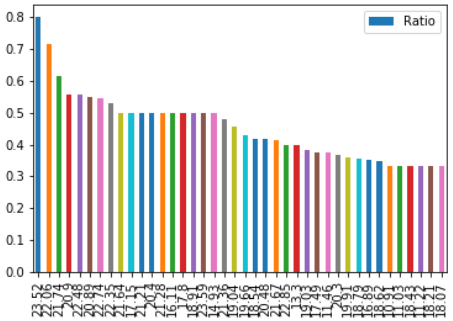
- We find following sectors in which most loans are distributed:
  - debt_consolidation (46.93%)
  - credit_card(12.91%)
  - home_improvement (7.5)
  - major_purchase (5.5)
  - small_business (4.6)

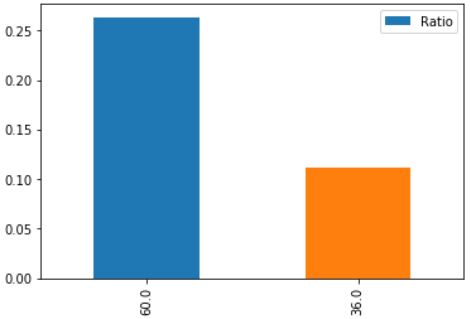
- We would check following driving factors on these.
  - grade
  - sub_grade
  - home_ownership
  - term
  - int_rate





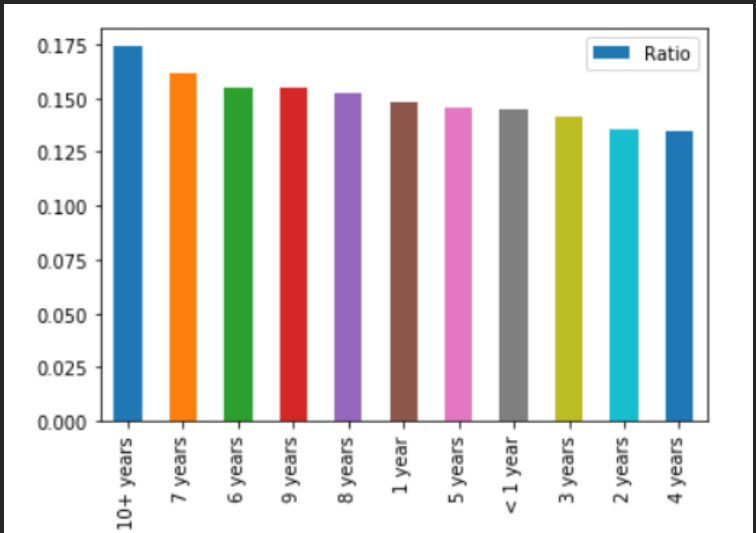


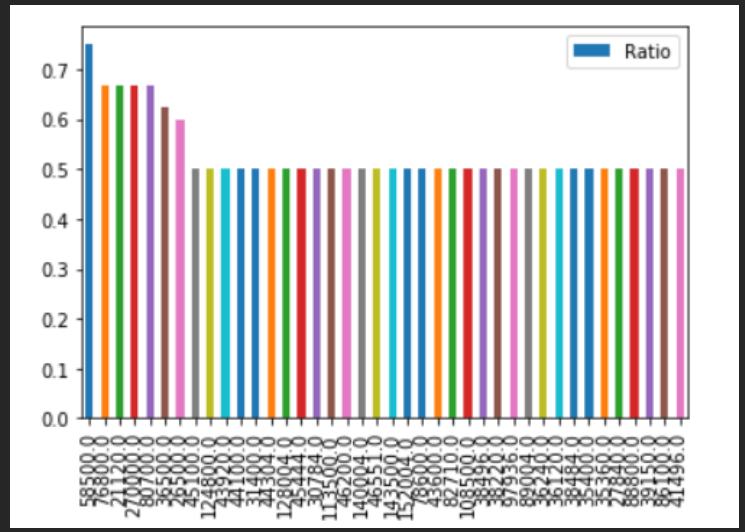


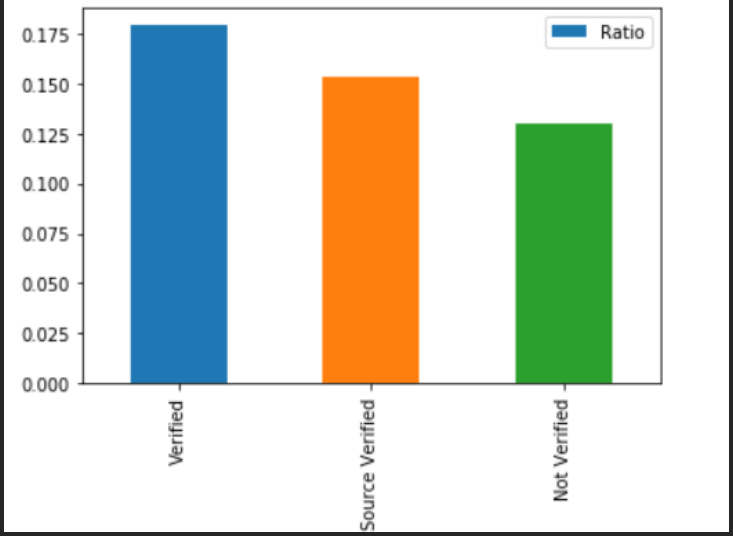


- Graphs of **debt_consolidation** vs

'grade'
'sub_grade'
'home_ownership'
'int_rate'
'term'
'annual_inc'
'Verification status'
'emp_length'







- These plots represents that ‘Validation status’,’annual_inc’,’emp_length’ are not the driving variables for the analysis as there is **not much** variance in the plots.

  Where as 'grade’
  'sub_grade’
  'home_ownership’
  'int_rate'
  ‘term’ show **more** variance in plot hence taken as driving variables.

Conclusion

We have following 5 factors as driving factors which can play major role while granting any loan in major business fields.

grade – Prefer A,B. Avoid G and F

sub_grade – Prefer Ax, Bx, and avoid Fx

home_ownership – Close scrutiny to users with “OTHER” status

Term – “60 Months” loan cause default.

int_rate – Interest rates is driving factor.