LENDING CLUB ASSIGNMENT

SUBMISSION

Team members: 1. Ian Lagat
       2. Poonam Nakhava (Group Facilitator)

Abstract

- The main goal of this analysis is to find the driving factors of defaulters for a consumer finance company.
- This will help the company in identifying patterns which indicate if a person is likely to default or not.
- This information may be used for taking actions such as:
    1. denying the loan
    2. reducing the amount of loan
    3. lending (to risky applicants) at a higher interest rate
   4. And so on...

Problem solving methodology

Analysis - I

For the figure in right, the columns *loan_amnt* and *funded_amnt* have a linear relationship, hence we can take any one of these two for analysis.

For the bottom figure, we can see that most of the applicants were not verified and hence it could be a factor for the company facing credit losses at the hands of these risky borrowers.
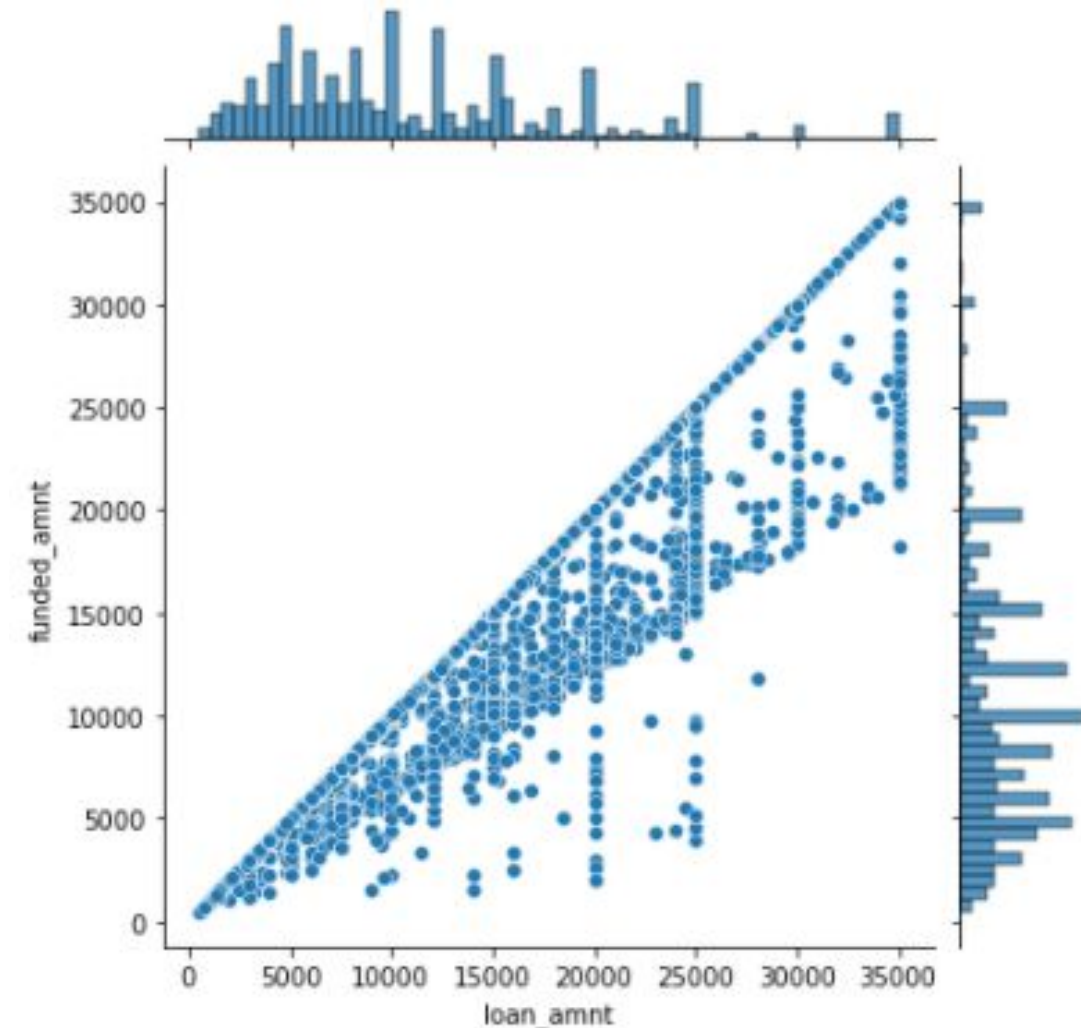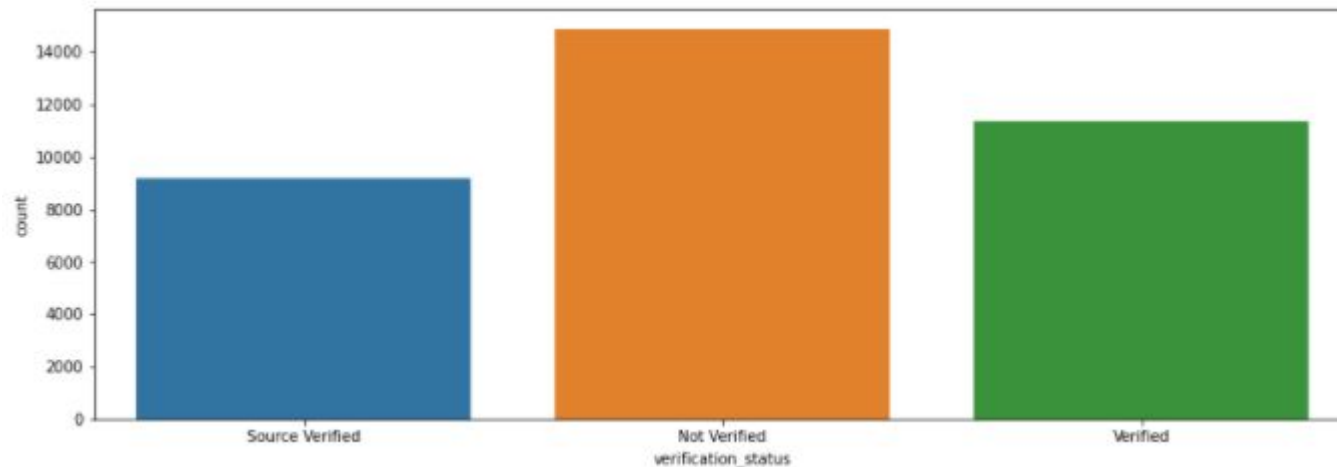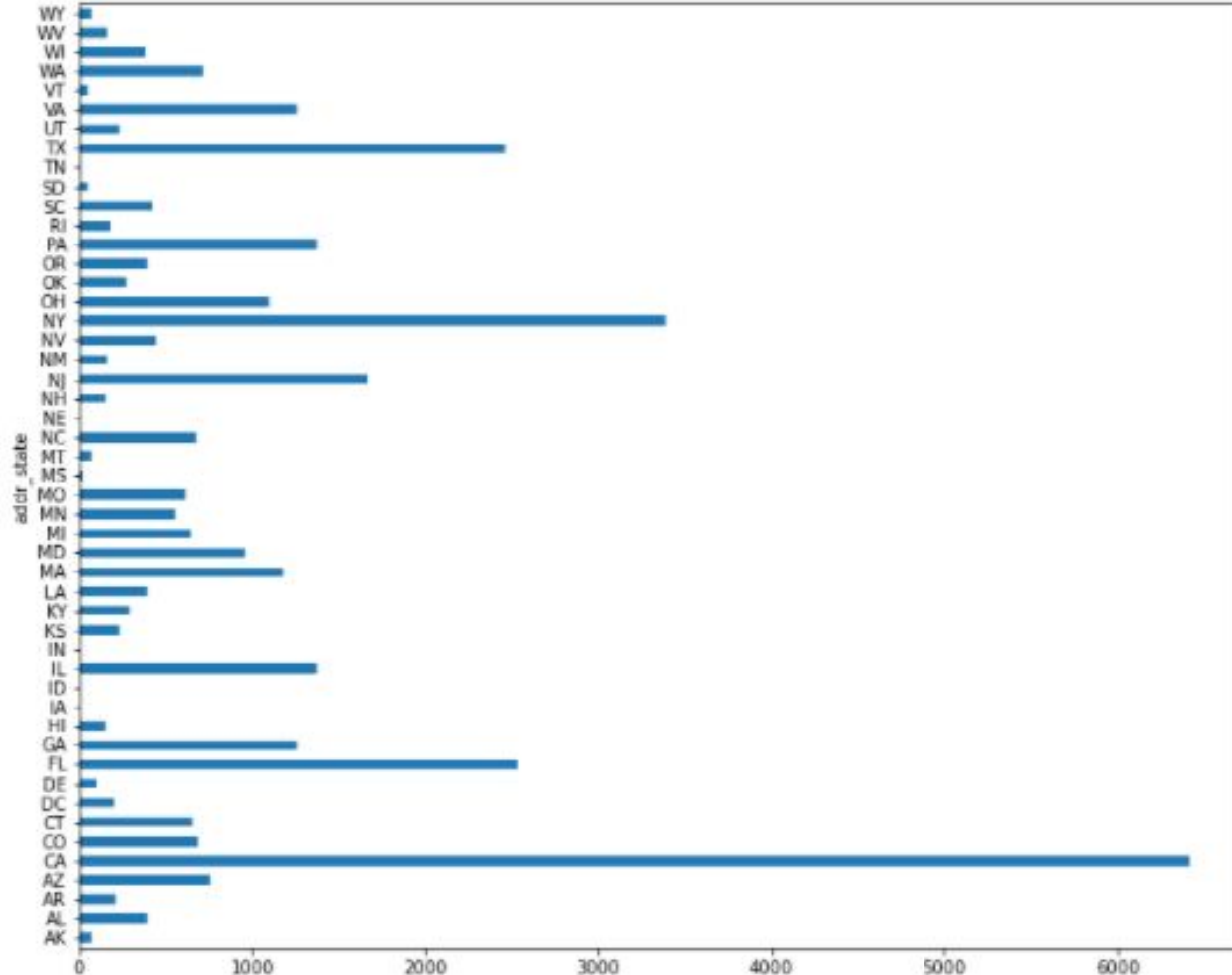
Analysis - II
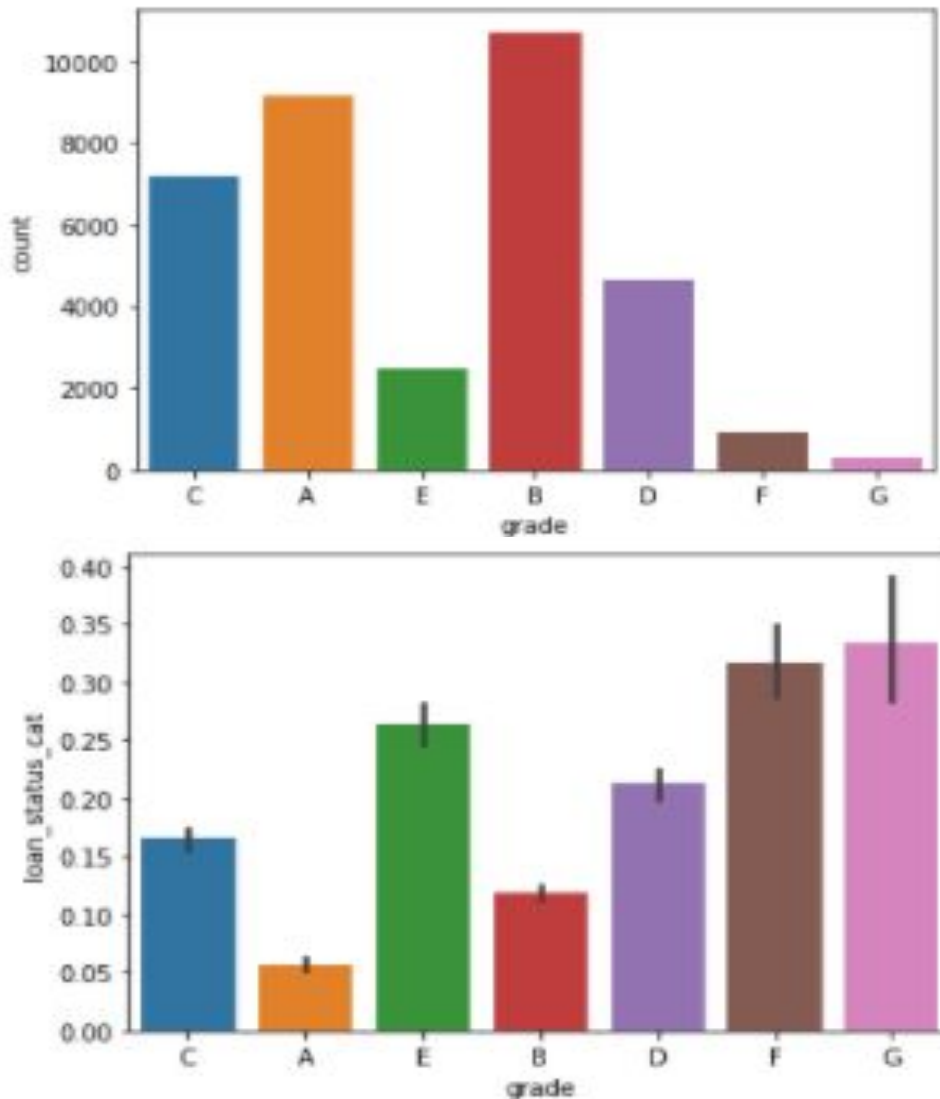
On analysing the impact of grade variable on the *loan_status* category we found that, *grade* 'B' customers have taken the highest amount of loan and grades 'F' and 'G' have taken the lowest amount of loan.
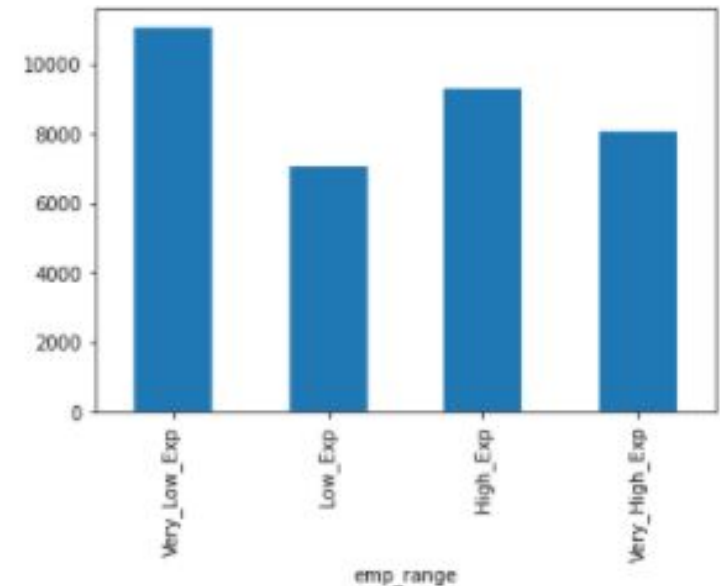
But when we analysed the same variable for the defaulters, we found the most defaulters to be from the grades 'F' and 'G'.

Analysis - III

We can find the top 3 cities who have taken the highest number of loans to be California (CA), New York (NY), Florida(FL) and Texas (TX)
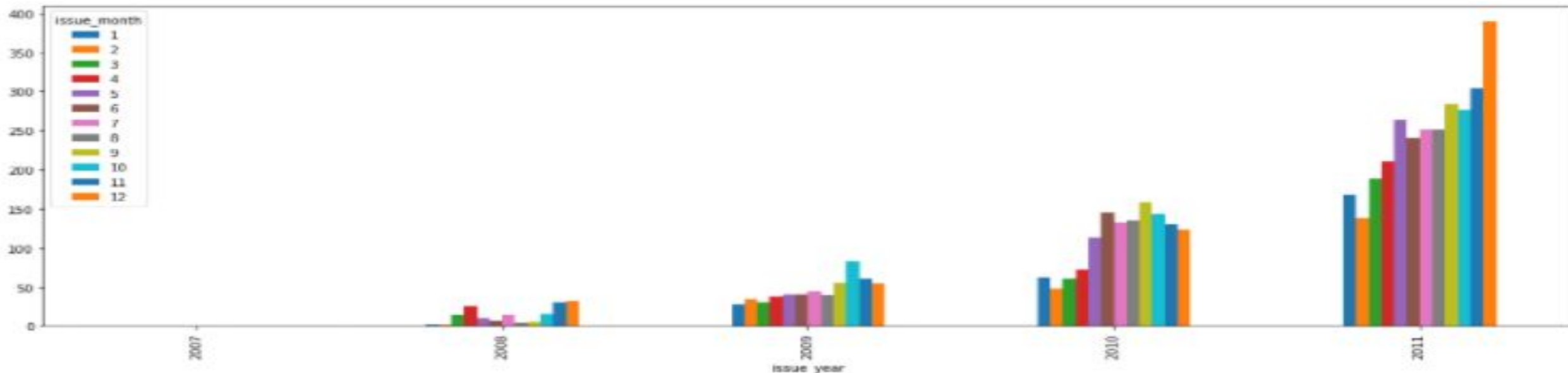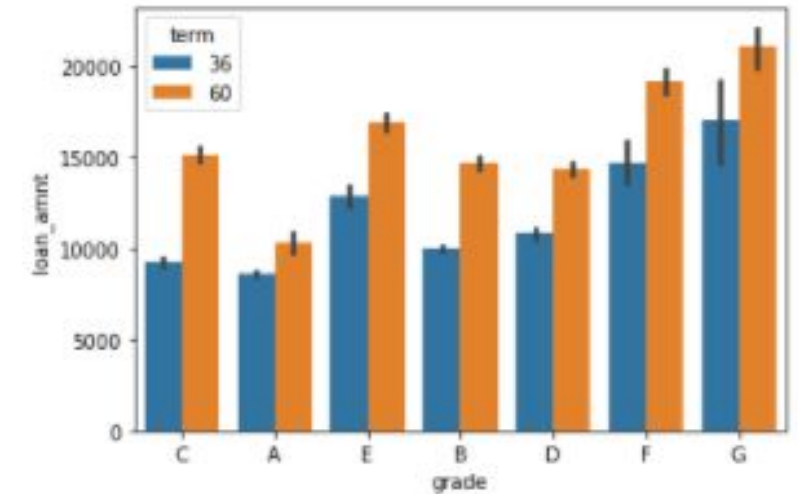
Also, the customers having less employment experience to be higher defaulters.

Analysis - IV

The highest loan amount was borrowed by the grades 'F' and 'G' for a term of 60 months.

The company issuing loans to the customers went on increasing yearly, with the most defaulters found in December, 2011

Analysis - V

The maximum amount of loan was borrowed with the purpose of debt consolidation, credit card and other.
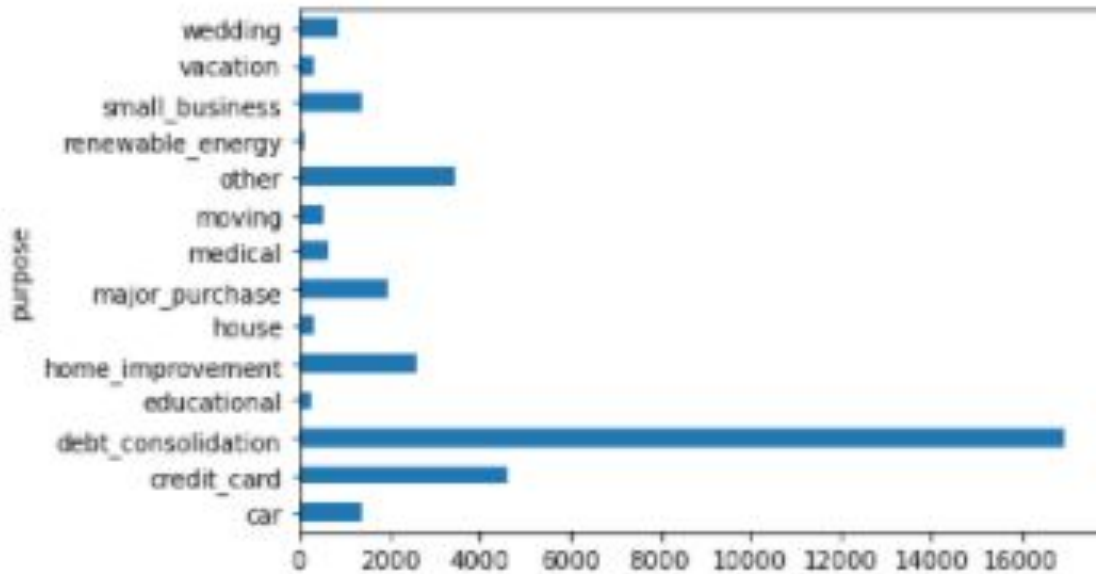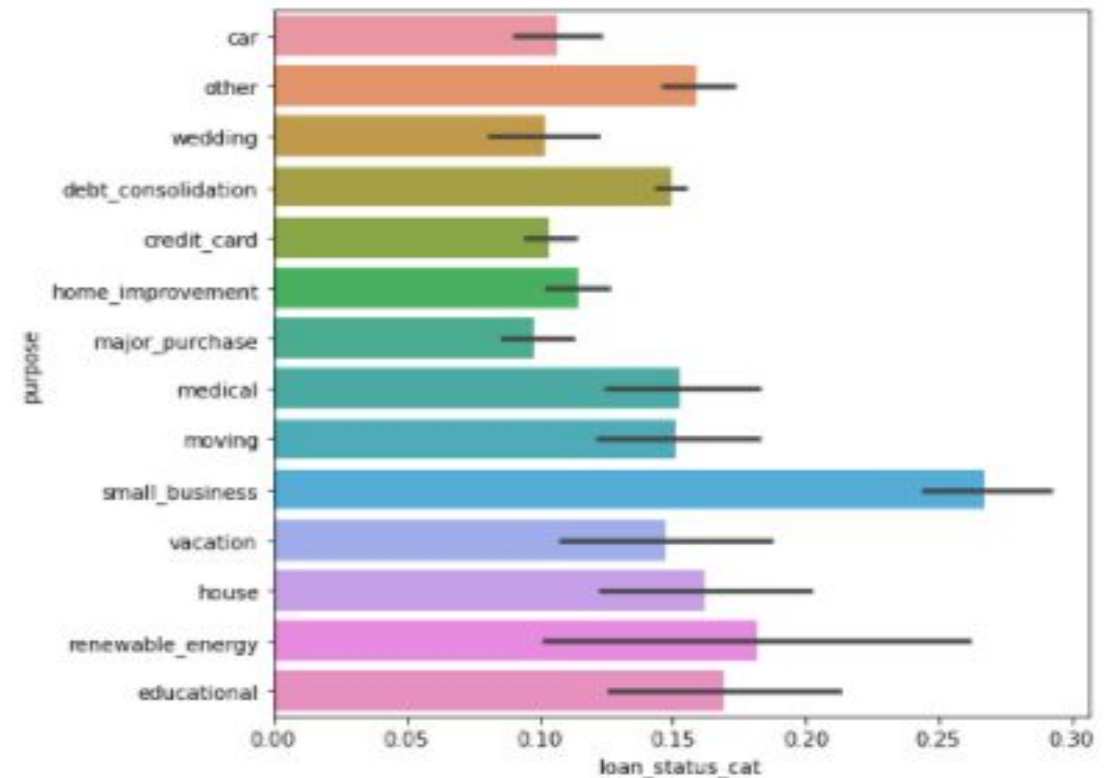

Whereas the maximum defaulters were the customers who had opted for the loan for the purpose of small business, renewable energy and education

Results

The plots clearly show the relation of each variable with *loan_status*, which signifies if a customers a defaulter and has been charged-off or if he has been able to completely repay his loan.

The greater bar range for charged-off customers is evident for the columns *pub_rec, term* and *pub_rec_bankruptcies*.

Whereas the *grade* graph shows the grades 'F' and 'G' being in the higher range of defaulters.

For the *home_ownership,* it is shows the value 'Others' having most defaulters.
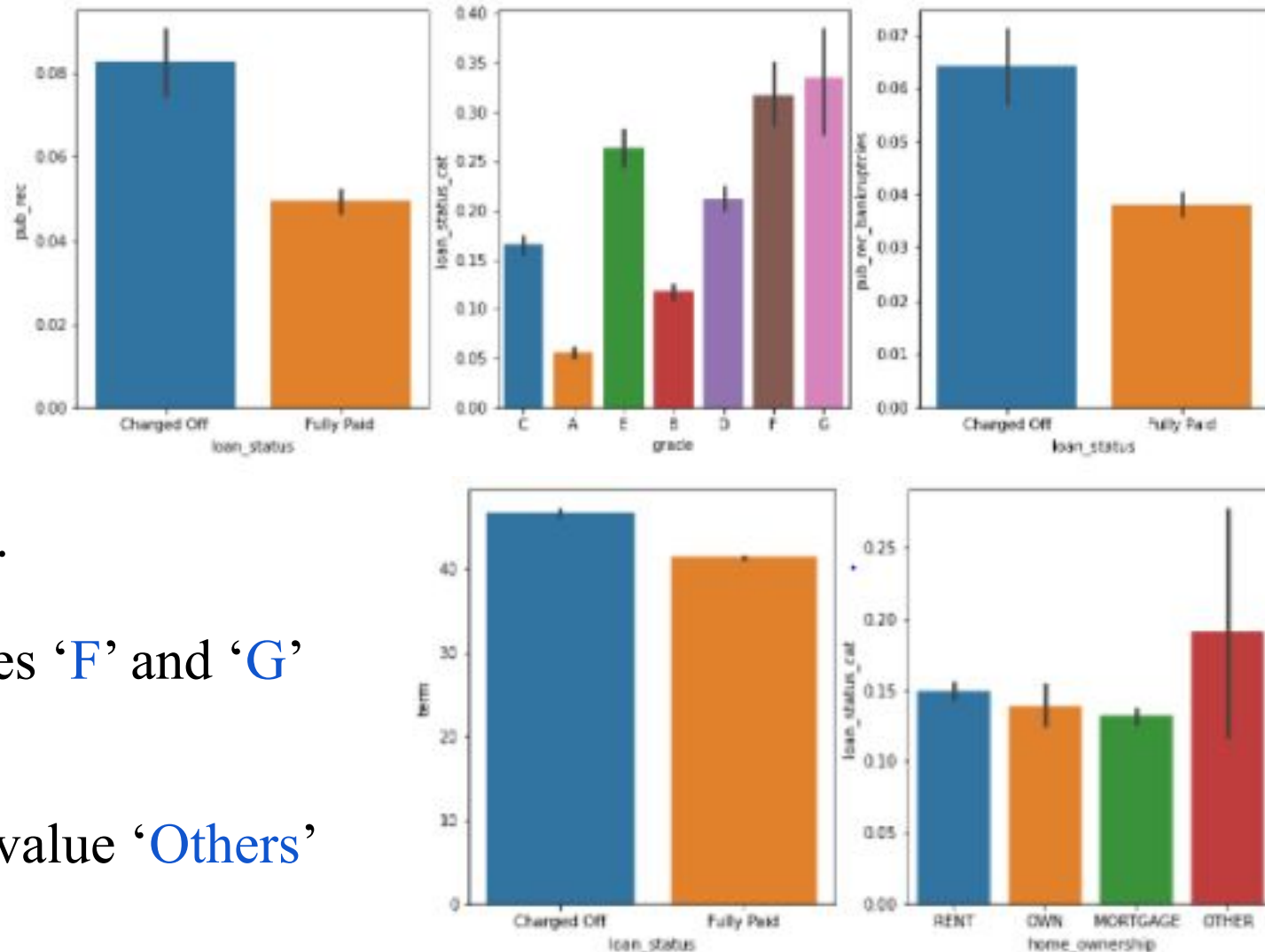
Conclusions

From the analysis made so far, we can deduce that the most strong driver variables for a customer being charged-off are as follows:

1. *pub_rec*: It clearly shows that if a person has a number of derogatory public records, then there is high possibility of him being charged-off.

2. *grade*: It shows that even if grades 'F' & 'G' were lowest loan borrowers, they were still the highest defaulters.

3. *pub_rec_bankruptcies*: Customers who have higher records of bankruptcies are bound to be charged-off more.

4. *term*: The people who have opted for a 60 months term period for their loan repayment are the highest defaulters.

5. *home_ownership*: Customers who don't seem to have a fixed home ownership are likely to default more.