EDA CASE STUDY

SUBMISSION

Group Name:

1. Alay Shah
2. Rutuja Mowade
3. Sabyasachi Tripathy
4. Rahul Rungta

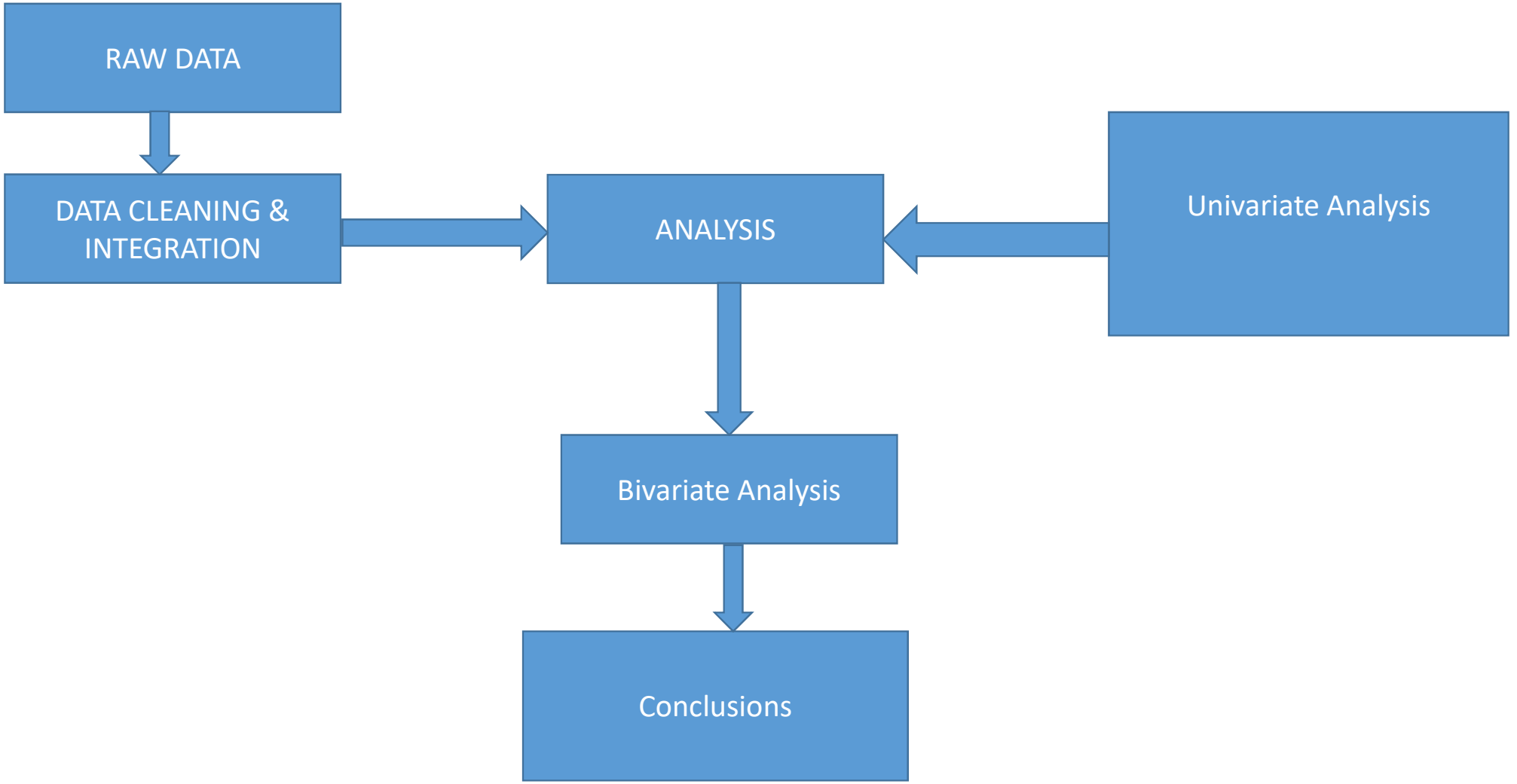
<Abstract>

Company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.

The company can utilise this knowledge for its portfolio and risk assessment.

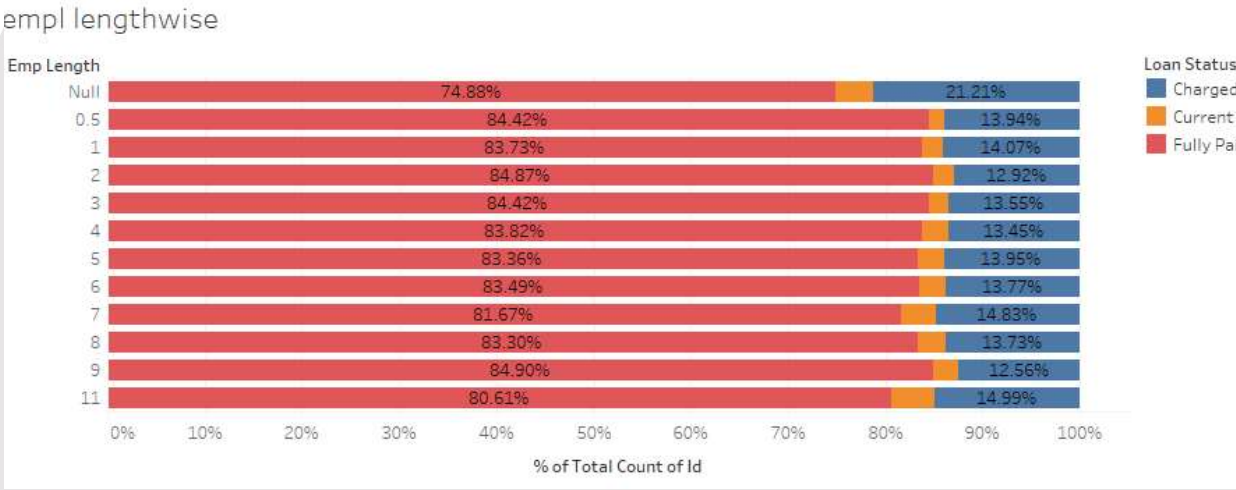
If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

Problem solving methodology: (FLOW CHART)

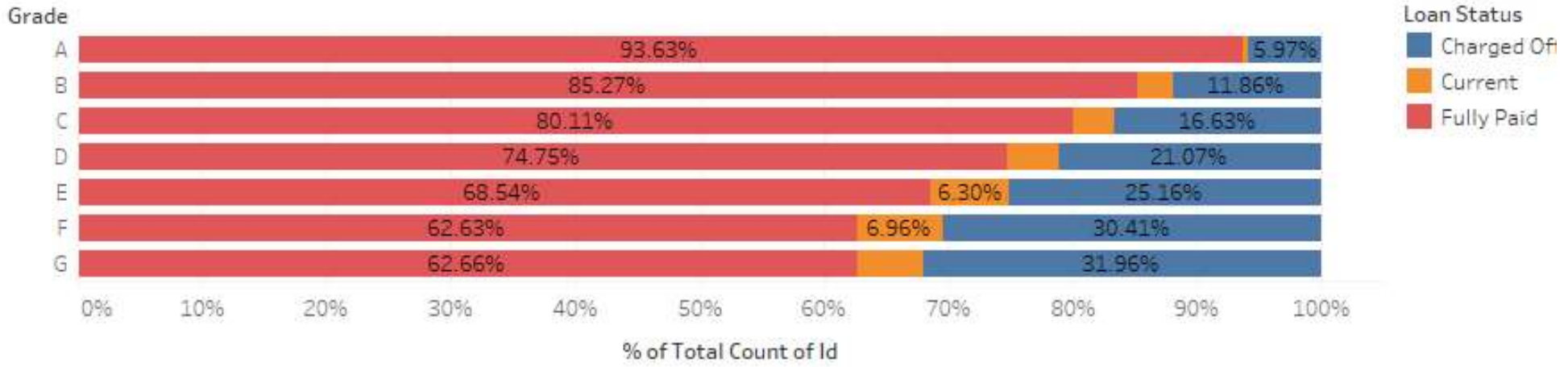


<Analysis>

- Univariate Analysis-I
- You see maximum defaulters where Employment length is not mentioned.

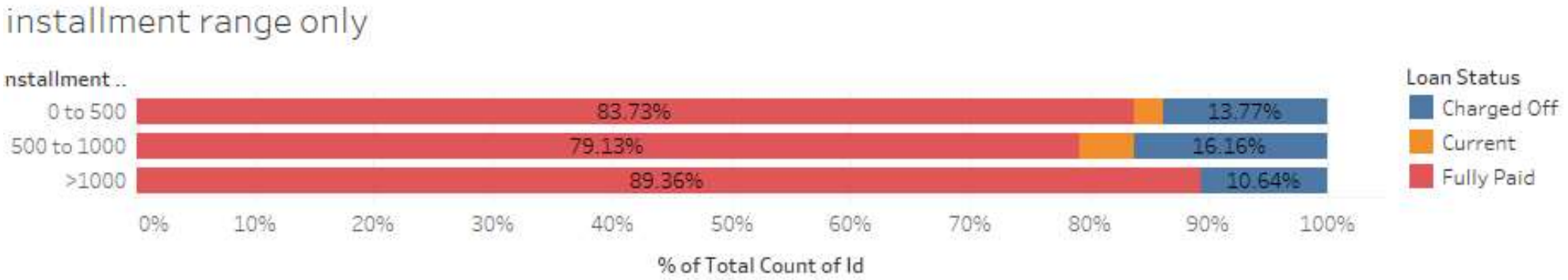


gradewise only

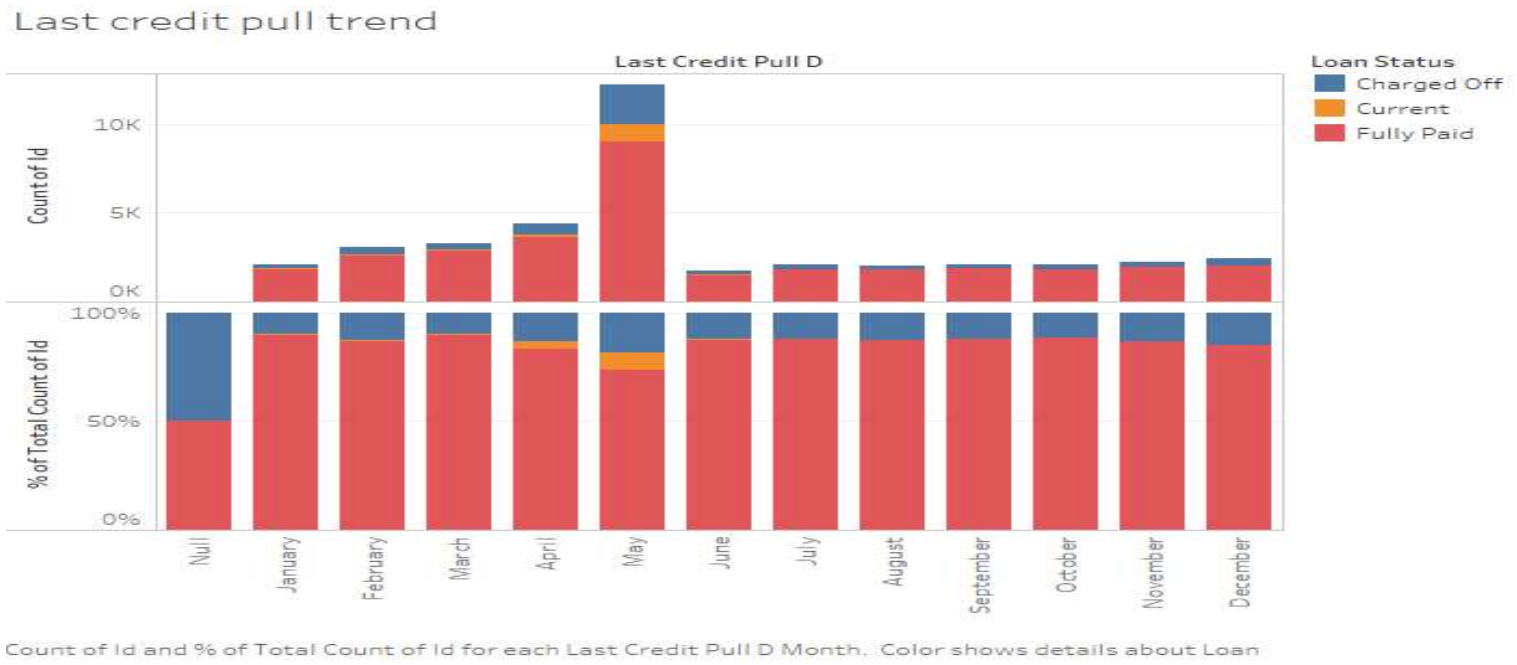


Univariate Analysis-II

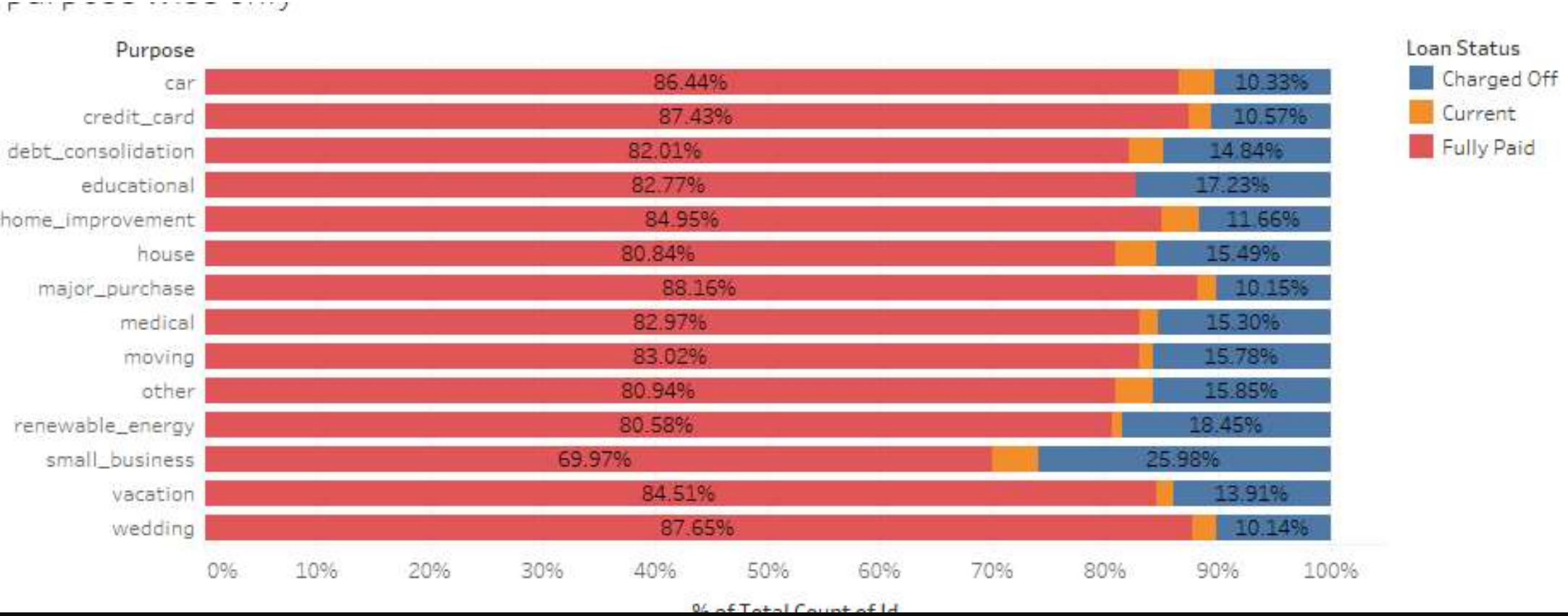
You see as Grade decreases from A to G % of defaulters increases.



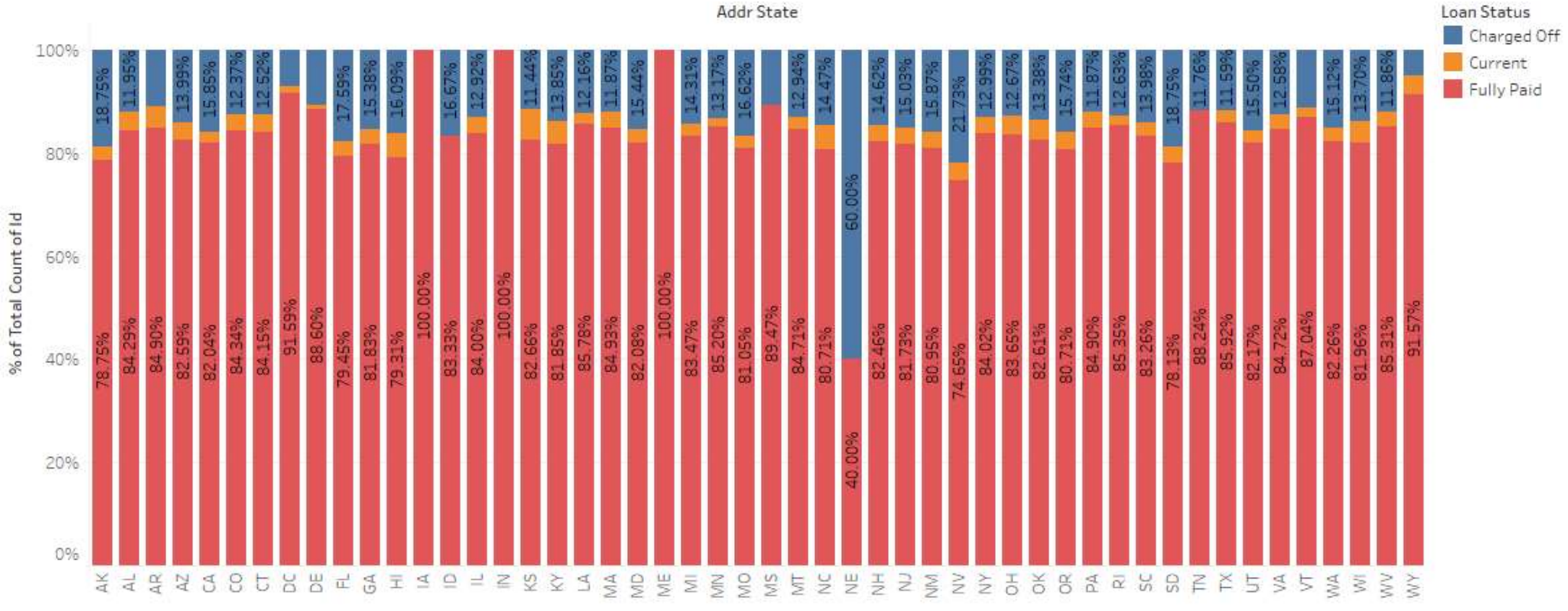
Installment Range Analysis



Last Credit Pull Trend:- Unusual spike in number of loans during month of May

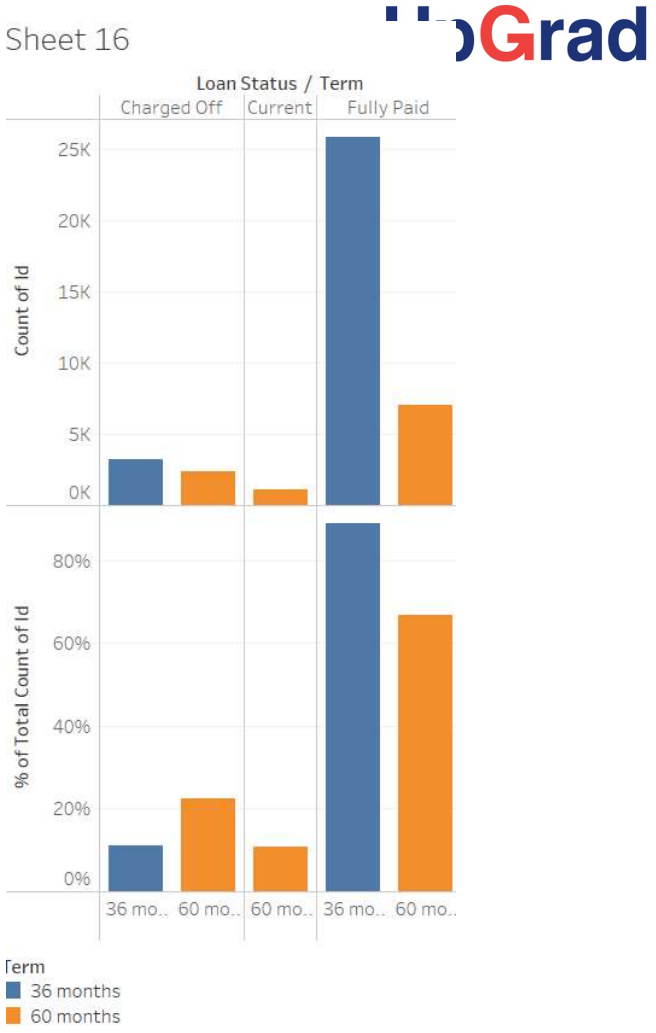


Looking at Purpose, you see maximum % of defaulters who took loan for business, banks may wants to look closely business plan, model and revenue capacity.

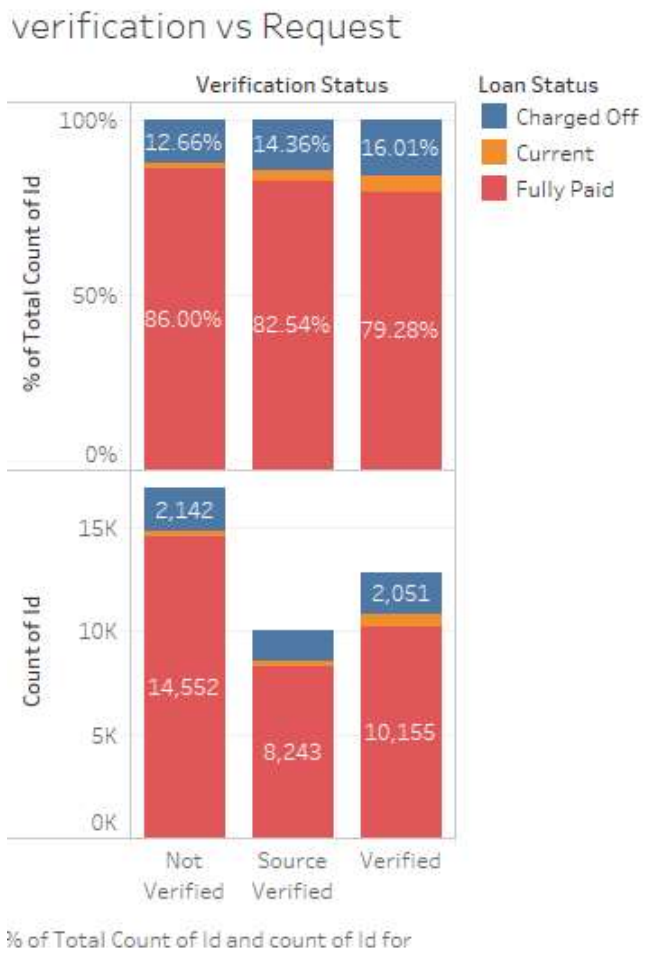


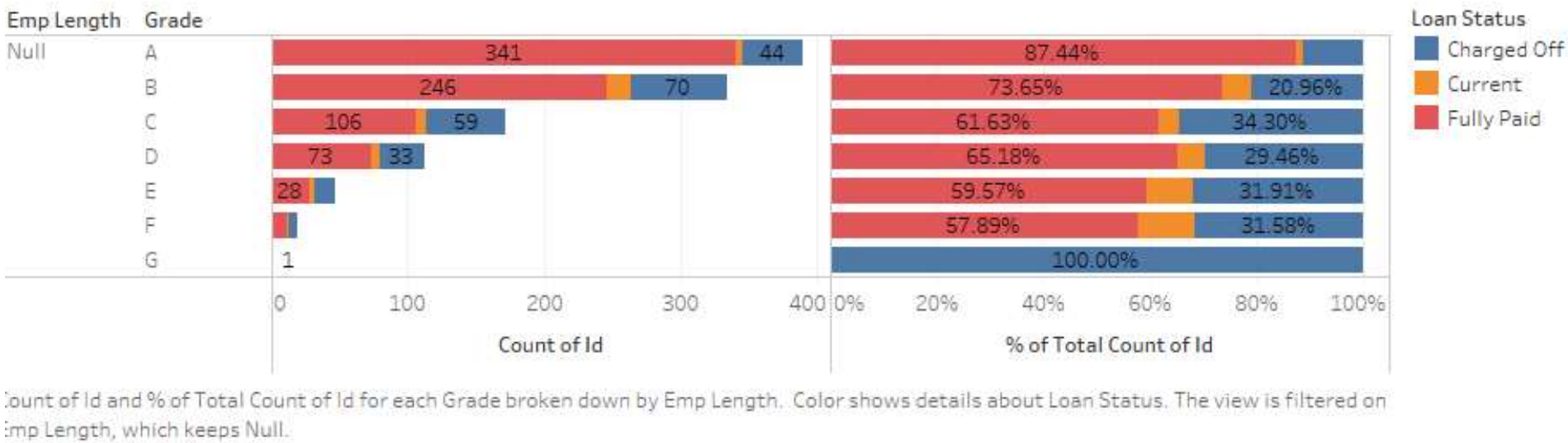
Maximum % of defaulters from Nevada, may be people using money for gambling in Vegas.

Term impact


Sheet 16
Loan Status / Term
Charged Off
Current
Fully Paid
Count of Id
25K
20K
15K
10K
5K
0K
% of Total Count of Id
80%
60%
40%
20%
0%
36 mo..
60 mo..
60 mo..
36 mo..
60 mo..
Term
36 months
60 months

Verification





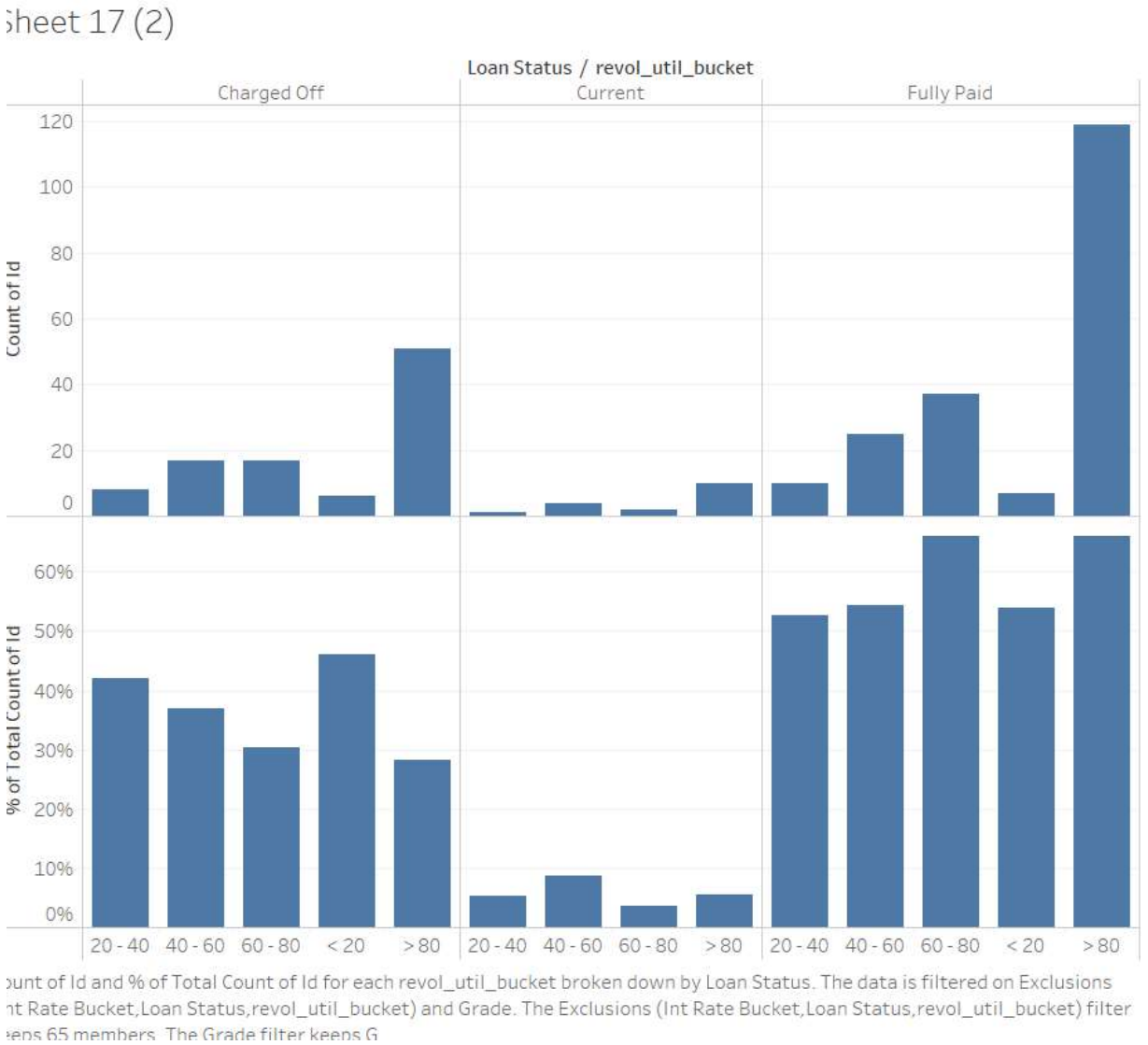
Bivariate Analysis-I
Employment length and grade

purpose vs grade

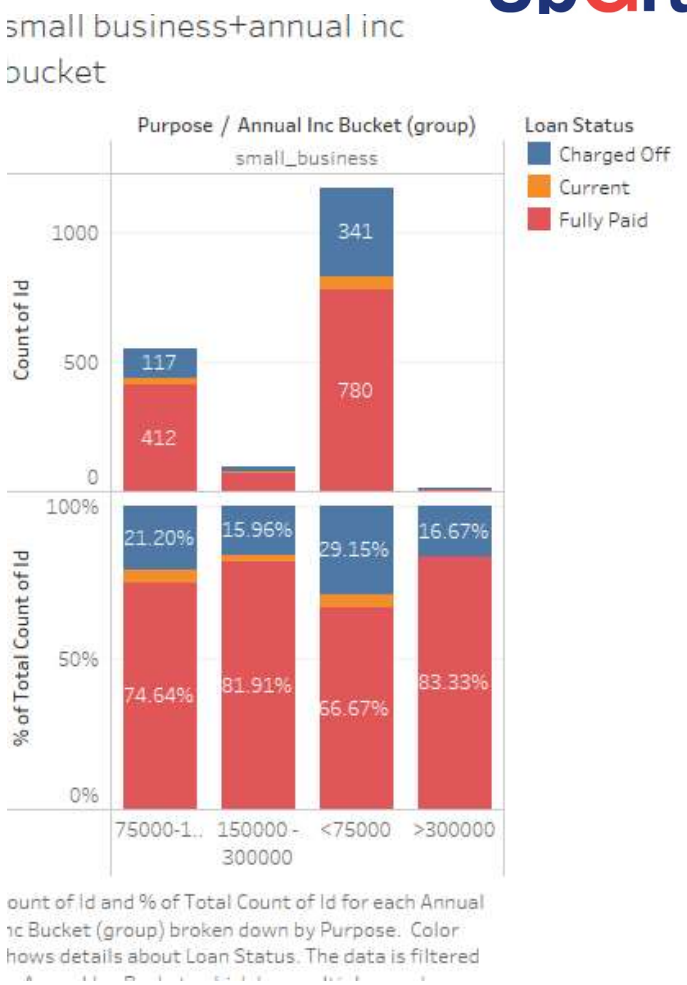
% of Total Count of Id and count of Id for each Grade broken down by Purpose. Color shows details about Loan Status. The view is filtered on Purpose, which keeps small_business.

Bivariate Analysis-II
Purpose and grade

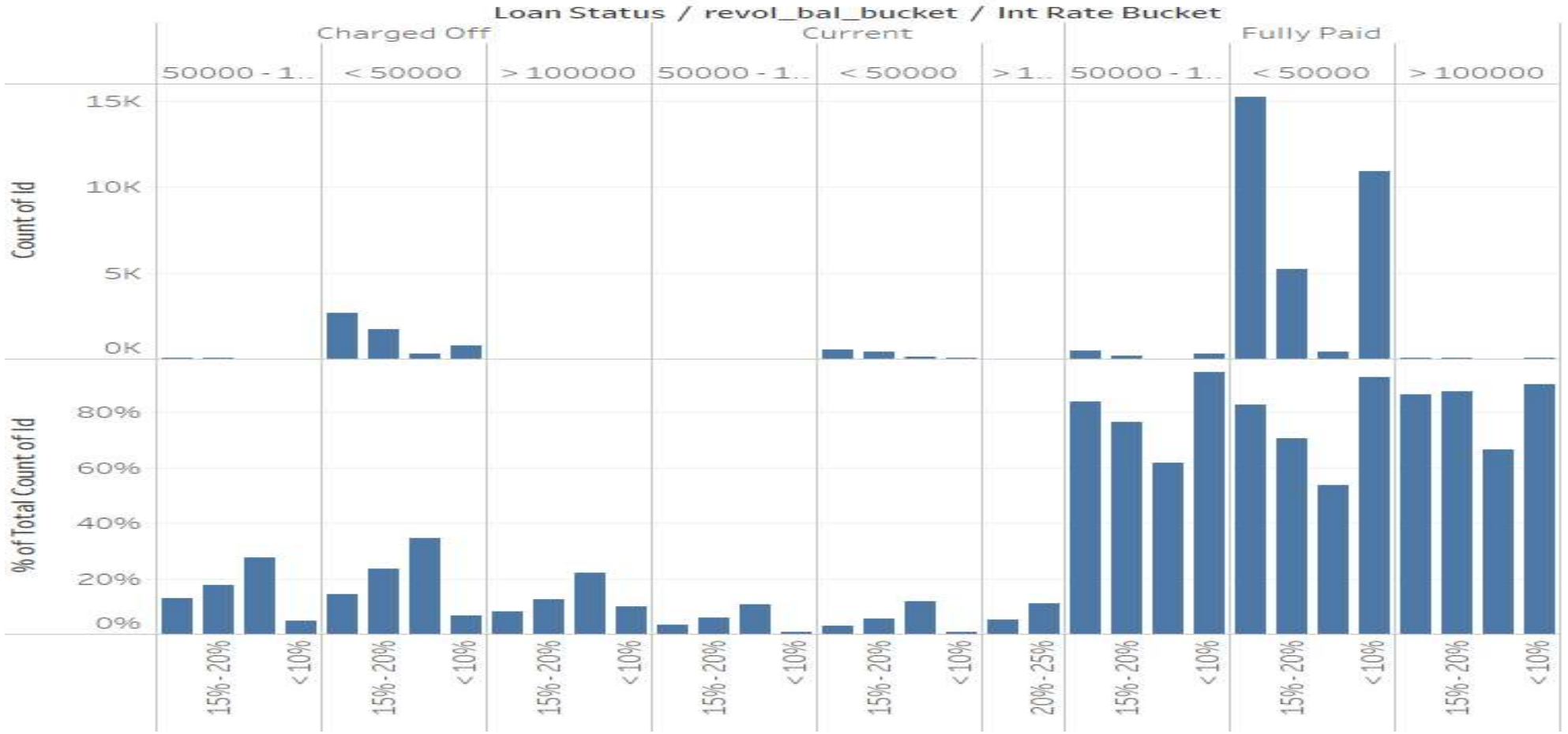
Bivariate Analysis-III Loan Status vs Revol bal bucket and Int Rate Bucket



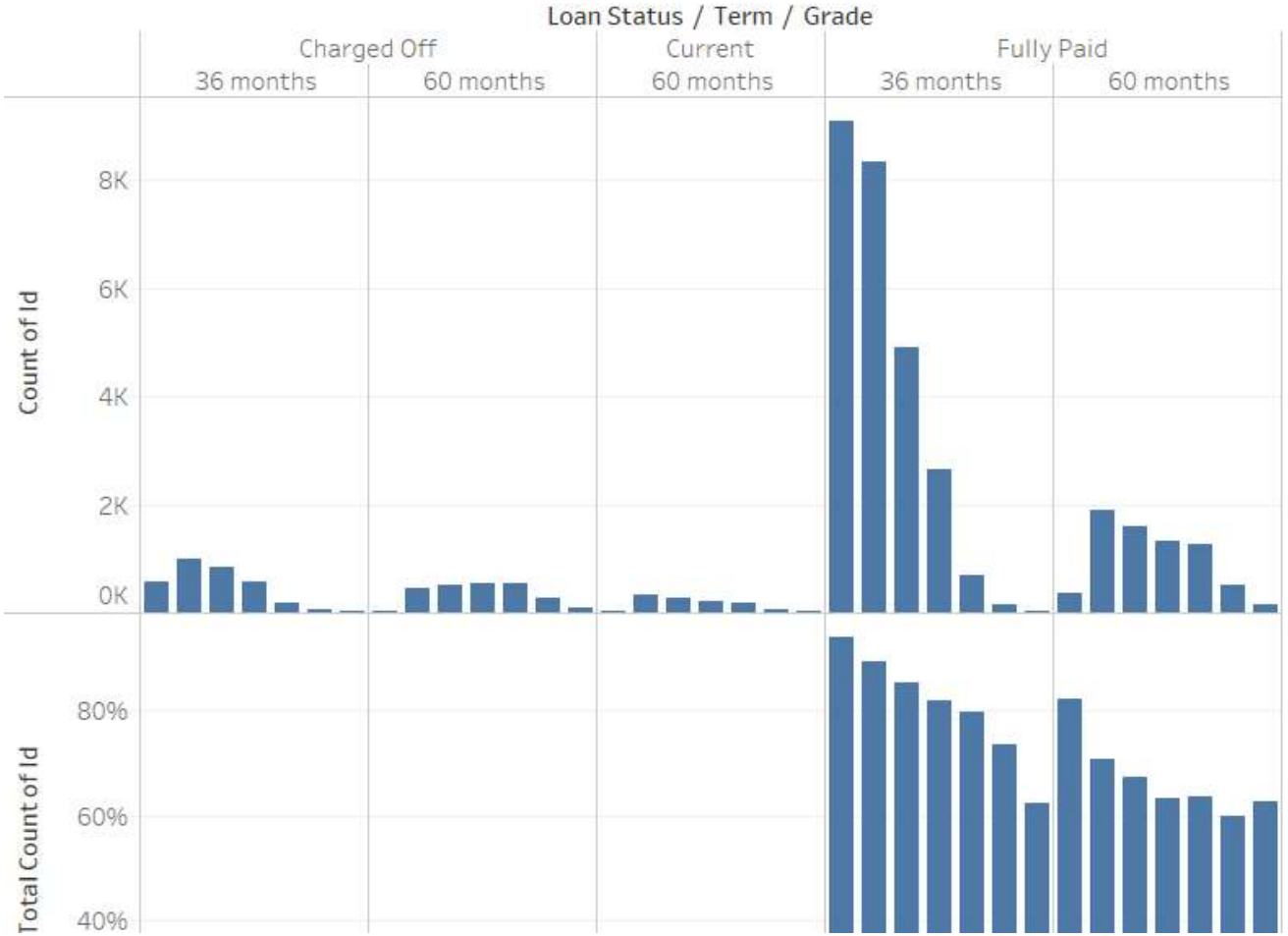
Bivariate Analysis-IV
Purpose along Annual Income bucket



Sheet 16


Loan Status / revol_bal_bucket / Int Rate Bucket
Charged Off
Current
Fully Paid
50000 - 1..
< 50000
> 100000
50000 - 1..
< 50000
> 1..
50000 - 1..
< 50000
> 100000
Count of Id
15K
10K
5K
0K
% of Total Count of Id
80%
60%
40%
20%
0%
15% - 20%
< 10%
20% - 25%

Sheet 21


Loan Status / Term / Grade
Charged Off
36 months
60 months
Current
60 months
Fully Paid
36 months
60 months
Count of Id
8K
6K
4K
2K
0K
Total Count of Id
80%
60%
40%

Conclusions

- As we have seen various plots, there are some very valuable insights we got, impact of different variables individually and also how other correlate with one another.
- Bankers can definitely use this insights to perform scrutiny on loan candidates, and avoid risk of loosing good customers and avoid defaulters.

Thanks