Lending Club

Case Study

Submitted by:

K.M Kadir Koushik
&
Shine Sivadasan

<Abstract>

This presentation explains the insights extracted from the key features in the data and also supported by recommendations which can help Lending Club to understand the essence of those features for loan approval of an applicant.

There are two types of risks associated with the bank's decision:

- If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company.
- If the applicant is **not likely to repay the loan**, i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company.

EDA is done to understand how consumer attributes and loan attributes influence the tendency of default.

<Problem solving methodology>

Programming Language: Python 3

Jupyter Notebook

Libraries used: Pandas, Numpy, Matplotlib & Seaborn.

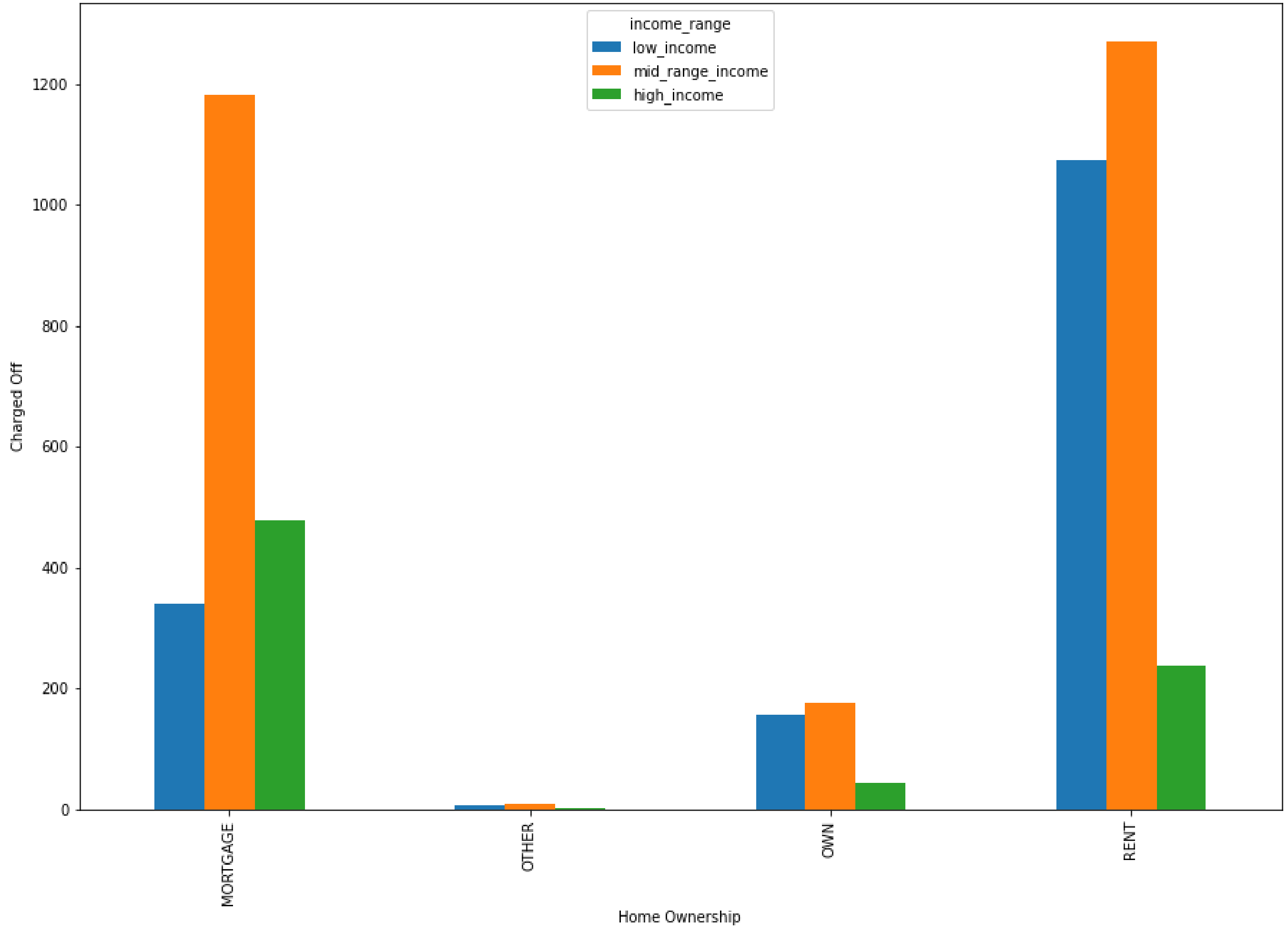
File used: ‘loan.csv’

< Annual Income vs Home Ownership >

Note: The values in the vertical axis of the graph is total number of "Charged-Off" cases.

Insight: Here we see that customers who own their home has very low tendency to default especially the ones who are in high income group.
Customers who pays rent for their residence, are very likely to default especially the mid range income earners, followed by low income earners.
Customers who has mortgage, among them the mid range income earners are most likely to default.

Recommendation:
LC can design favorable loan schemes for those who pays mortgage or rent for residence to reduce business loss and simultaneously favorable loan schemes, to attract borrowers who owns their residence and other.



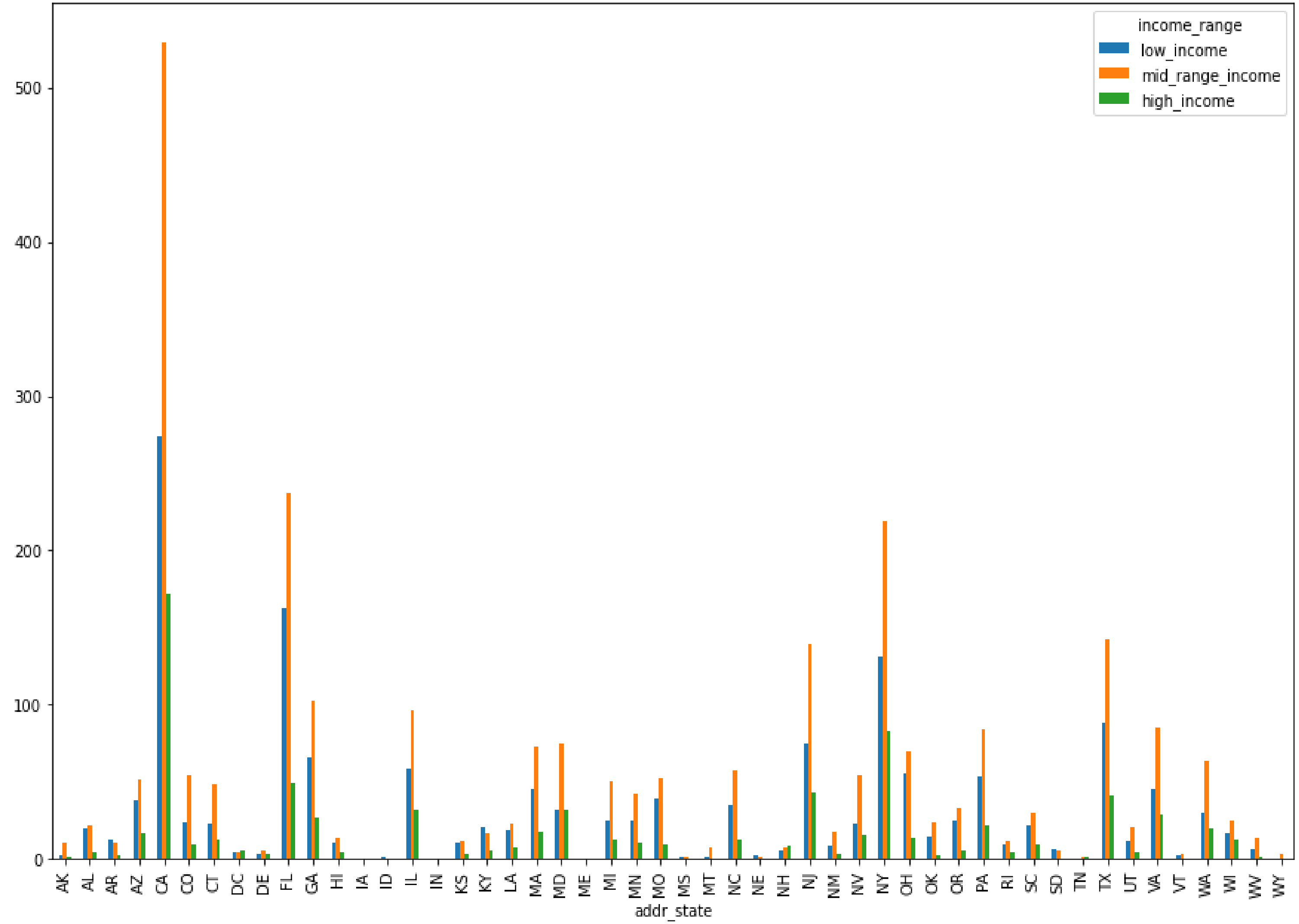
Note:
Three annual income basket:
basket 1:
High income (4000 - 46000)
Basket 2:
Mid range income (46000 - 92000)
Basket 3:
Low income (92000 – 142000)

<Annual Income vs Address State >

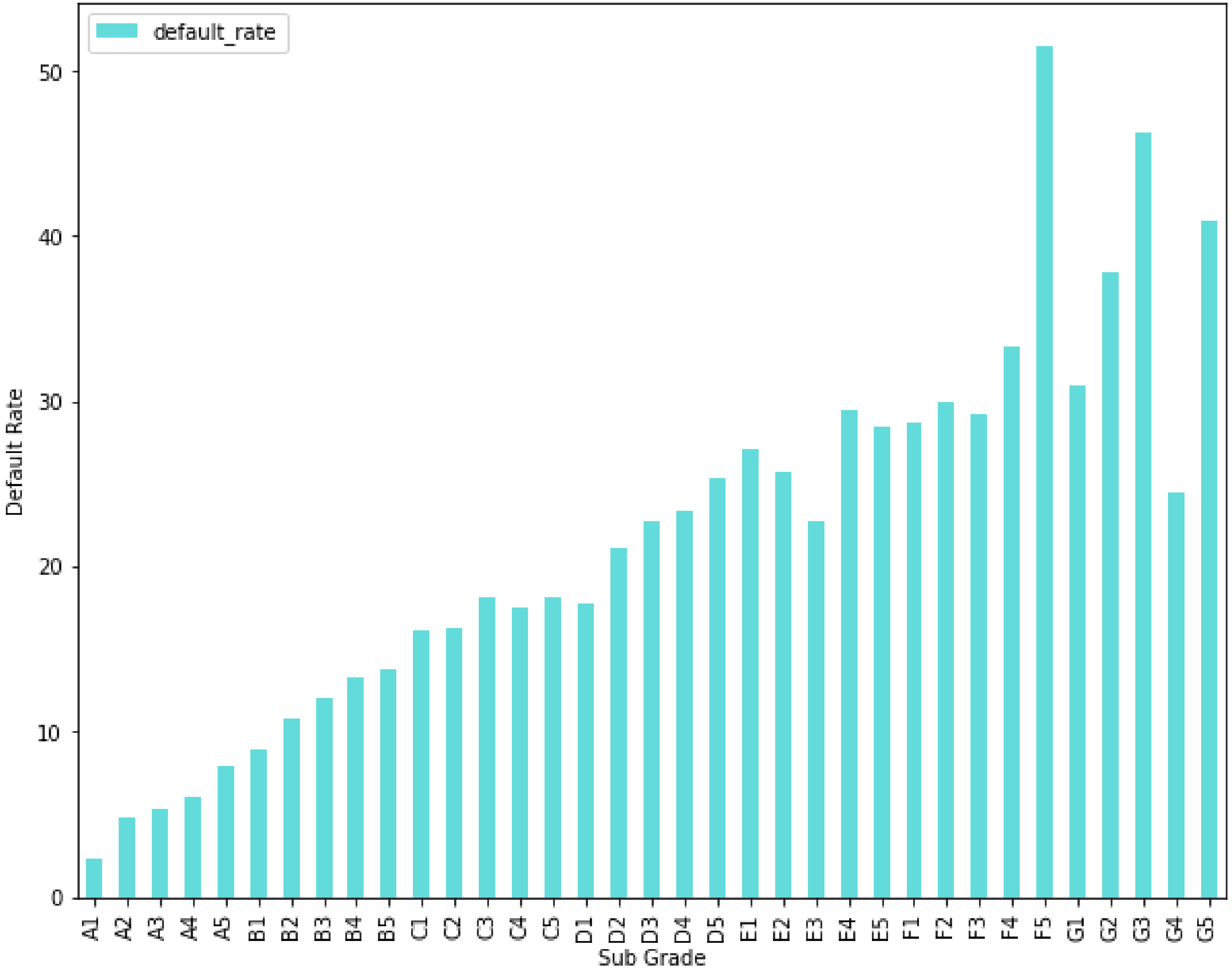
Note: The values in the vertical axis of the graph is total number of "Charged-Off" cases.

Insight:
We see that population from mid range income basket yields most charged off cases; moreover, California, New York, Florida, Texas being top 4 states with very high charged off cases.

Recommendation: LC can configure different measures upon different states to minimize the frequency.



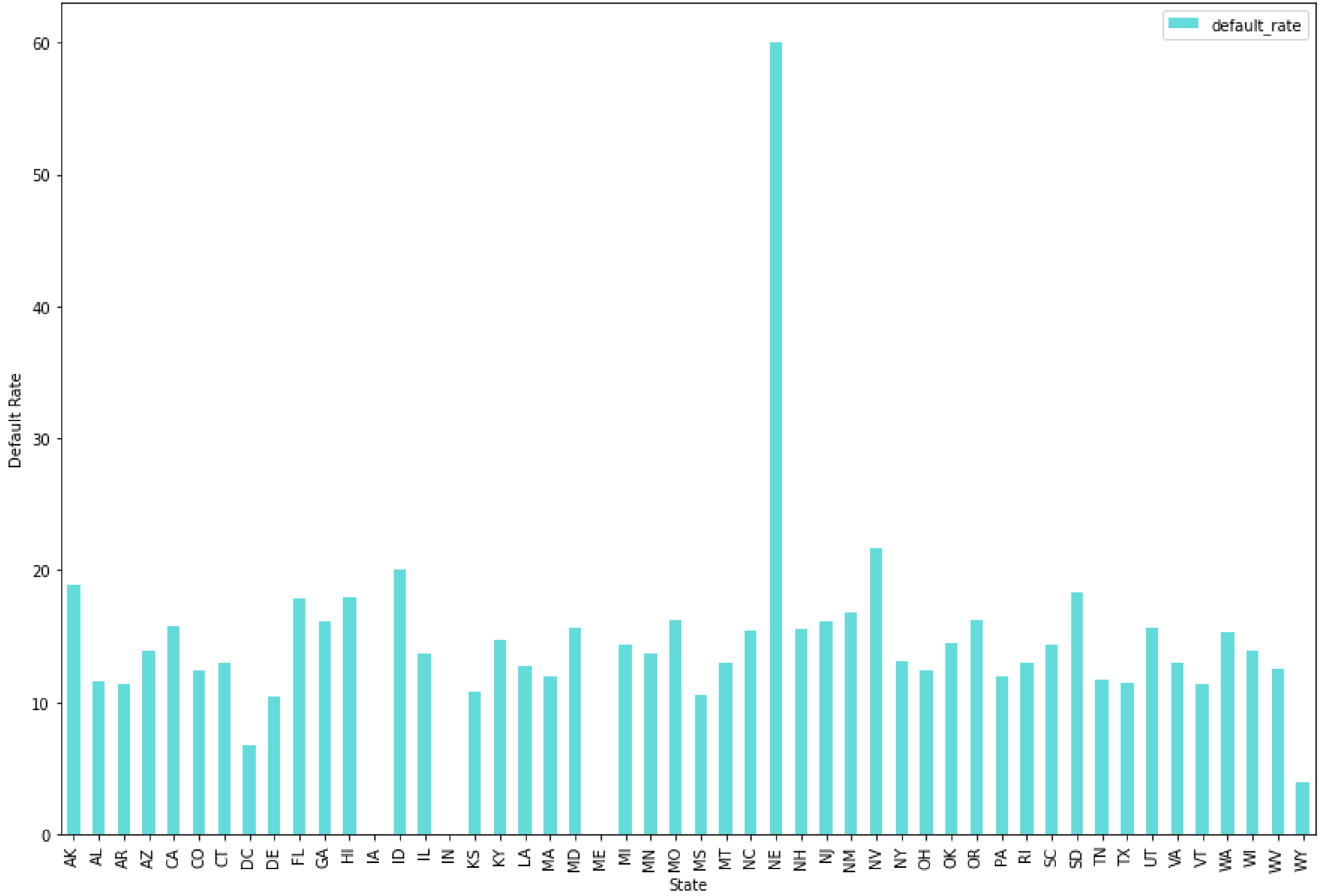
<Loan Sub Grade vs Default Rate>



Insight: Loan sub-grades F5, G2, G3, G5 have high default rate.

Recommendation: Since these loan sub grades reflects risky loan so LC management can consider limiting loan amount or devise policies that can make borrower of these sub grades subject to meeting conditions in order to sanction the loan.

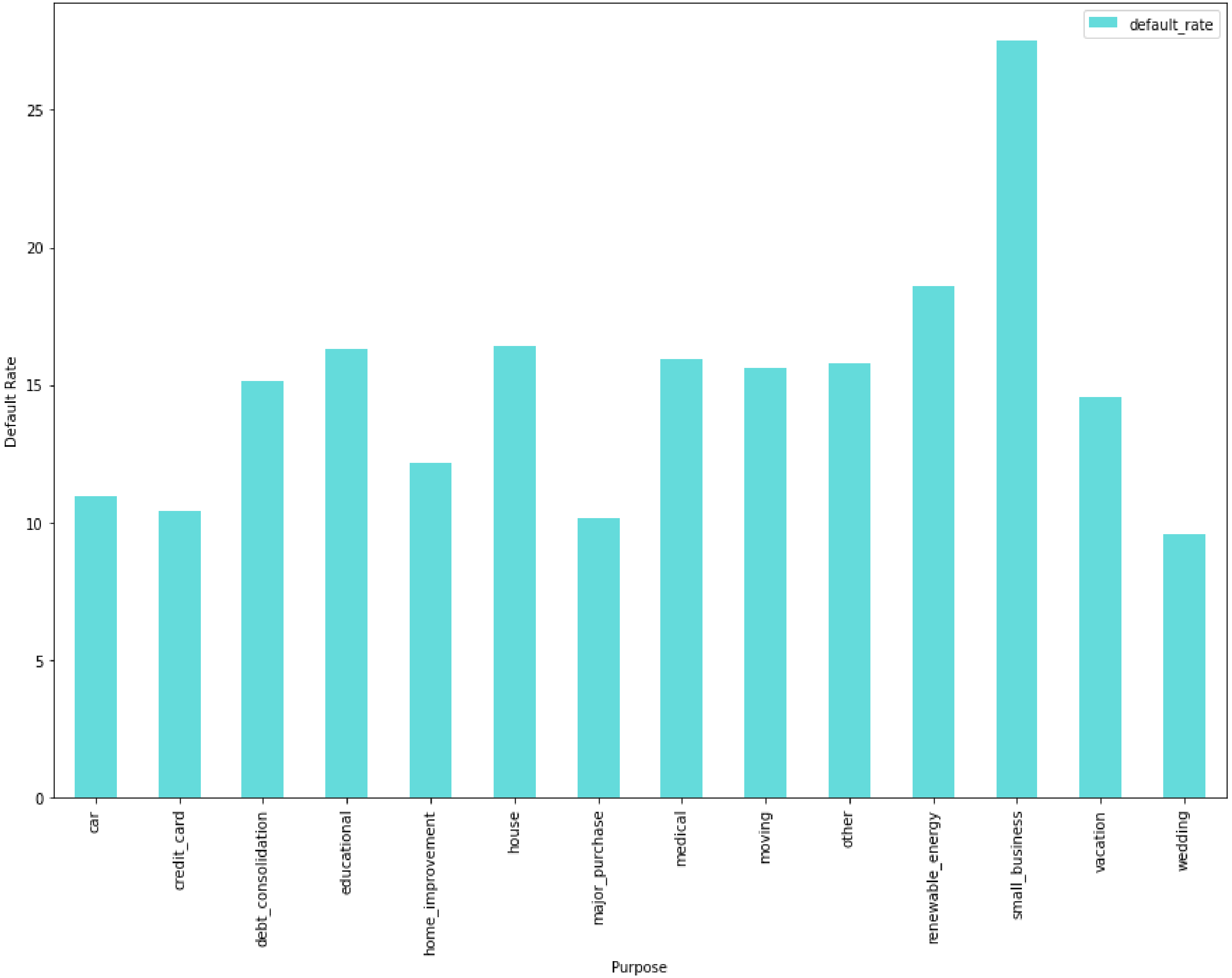
< Address State vs Loan Sub grade >



Insight: Nebraska (NE) has quite a significant rate of default cases.

Recommendation: LC can first analyze why borrower from Nebraska has such high tendency to default then formulate credit constraint measures.

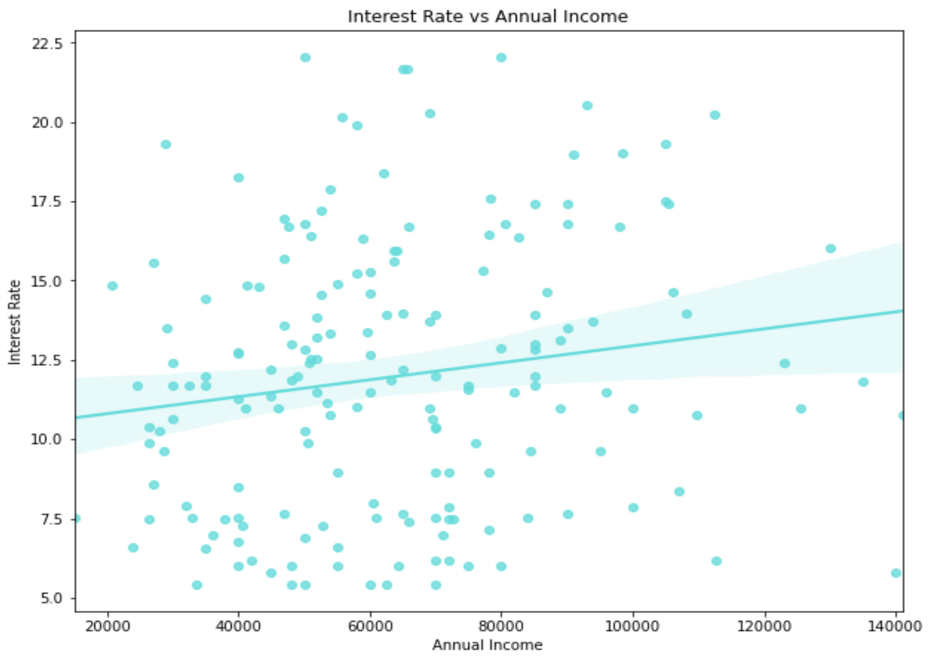
< Purpose vs Default Rate >

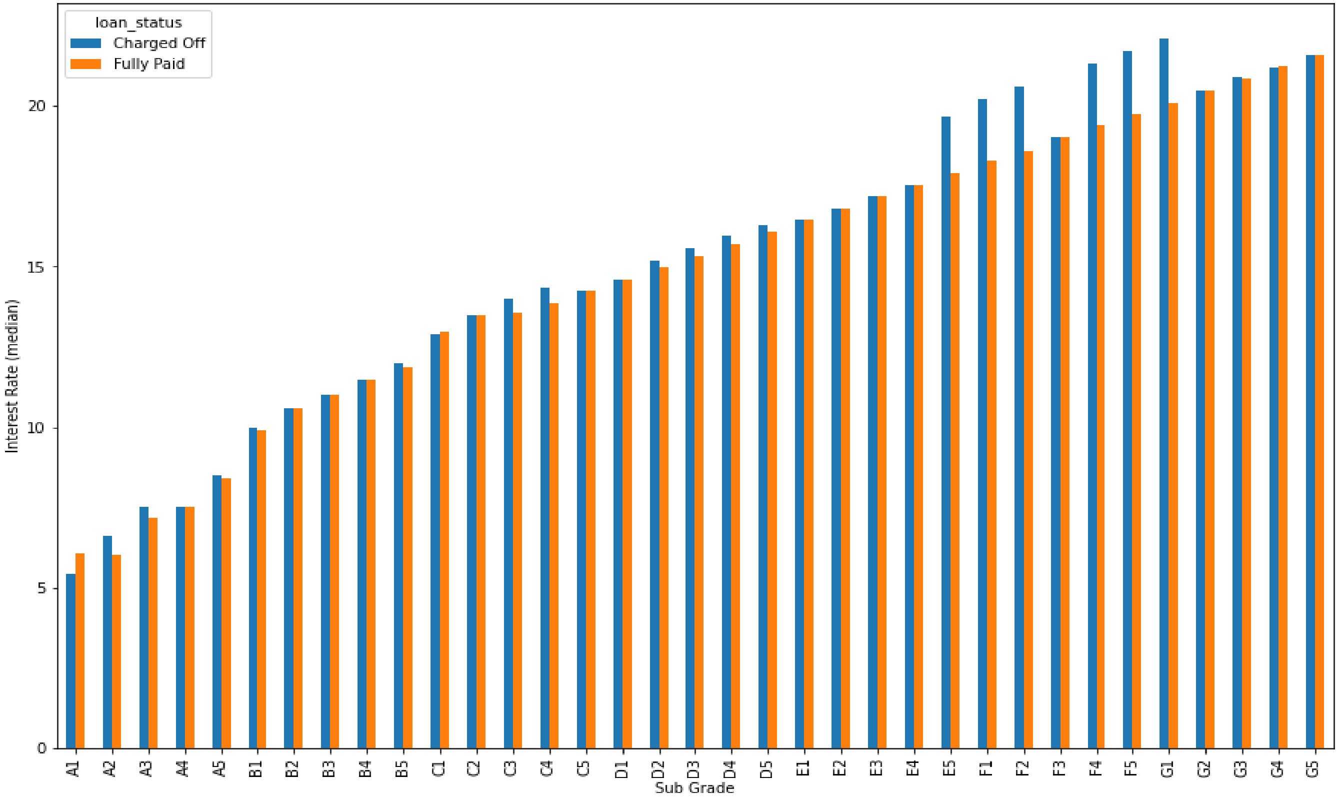


Insight: Borrowers for ‘Small Business’ purpose are most likely to default.

Recommendation: LC can consider taking soft measures to potential entrepreneurs of small business or startups
depending on borrowers’ plan to utilize the credit and repay.

Interest Rate vs Income & Loan Sub Grade





There seems to have slight positive correlation between interest rate and annual income.
Suspicion: high income earners are expected to be good borrowers and often charged reasonable rates unless they have tough credit record issues.

Recall *Annual Income vs Home Ownership* analysis in slide 4; high value loans are also being taken by mid range earners yet risky loan grade.

Insight: Notice that there are spikes in the median interest rate charged on the loan sub-grade D5, E5, F1, F2, F4, F5, G1.

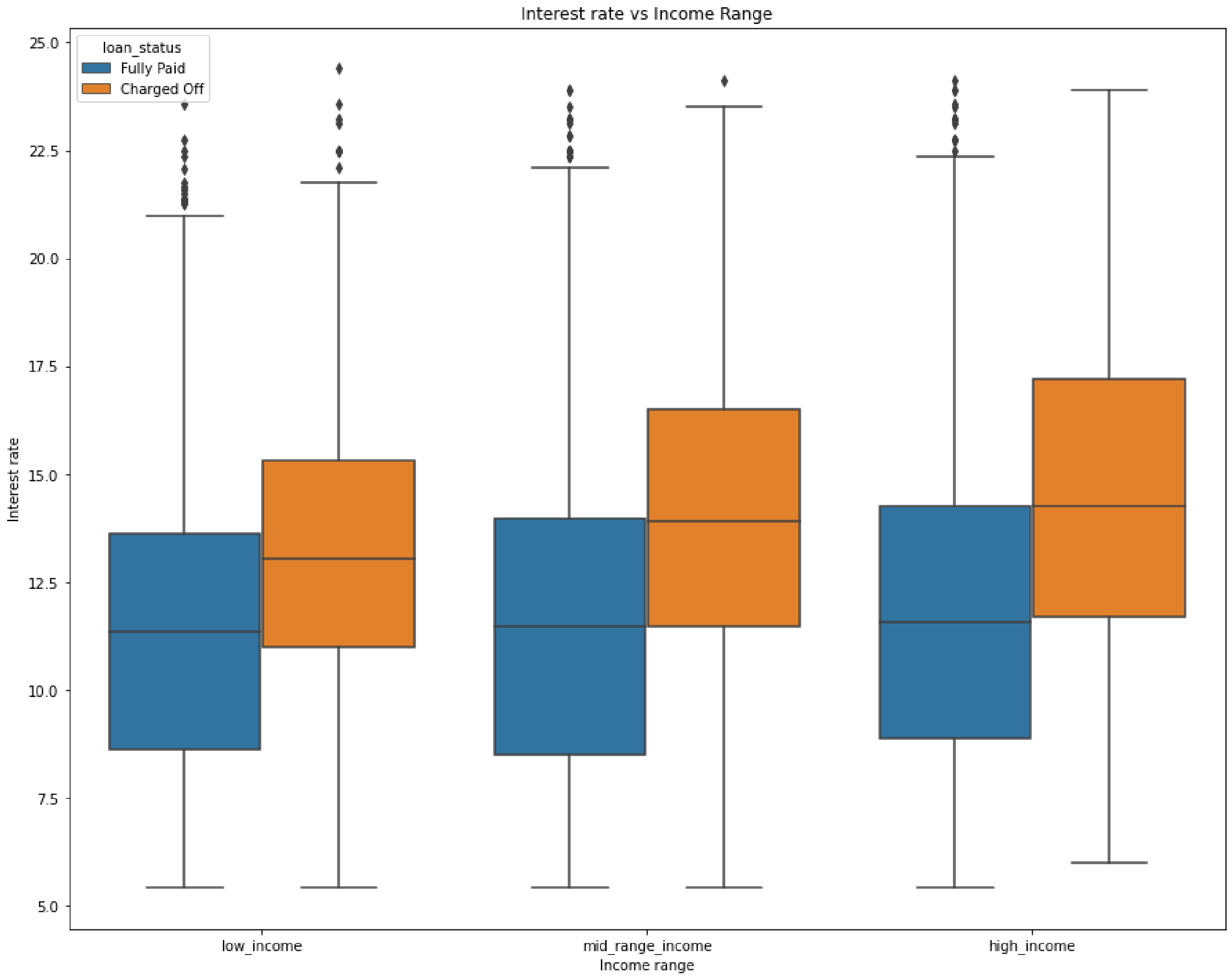
Recommendation: LC can consider determining new range of interest rate to be charged on these loan sub grade.

Interest Rate vs Income Range

Note:
Three annual income basket:
basket 1:
High income (4000 - 46000)
Basket 2:
Mid range income (46000 - 92000)
Basket 3:
Low income (92000 – 142000)

Insight: In both ‘high income’ and ‘mid range income’ basket, the median interest rate borne by charged-off group is equivalent to the upper limit (75th percentile) interest rate borne by the group ‘fully-paid’

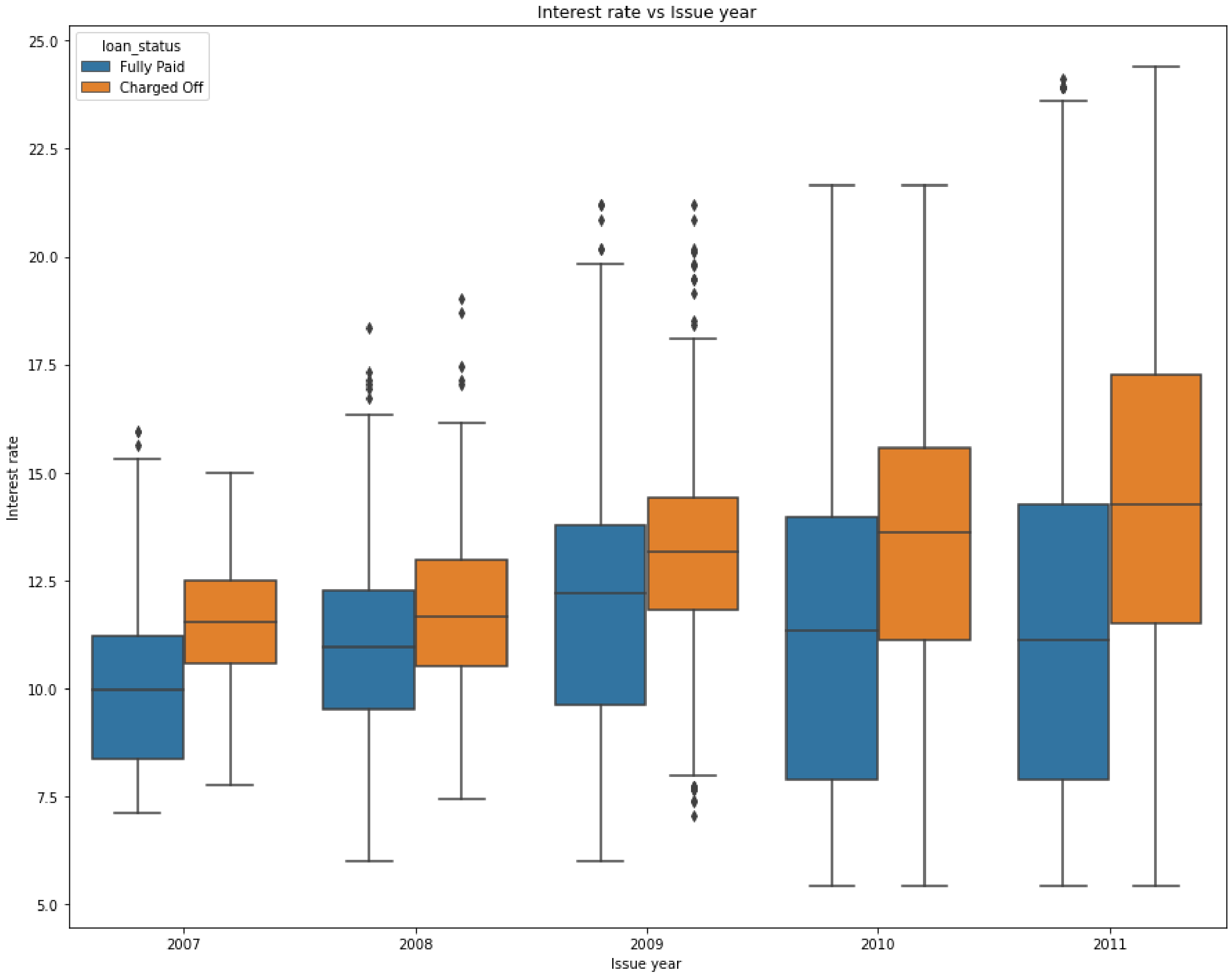
Recommendation: LC can decide upon flexing interest rate to balance the weight upon different income groups in order to minimize default.



Interest rate vs Issue Year

Insight: The Charged-Off group has been experiencing higher median interest, this says interest is one major issue! The number of charge-off case increasing as well as their median interest rate.

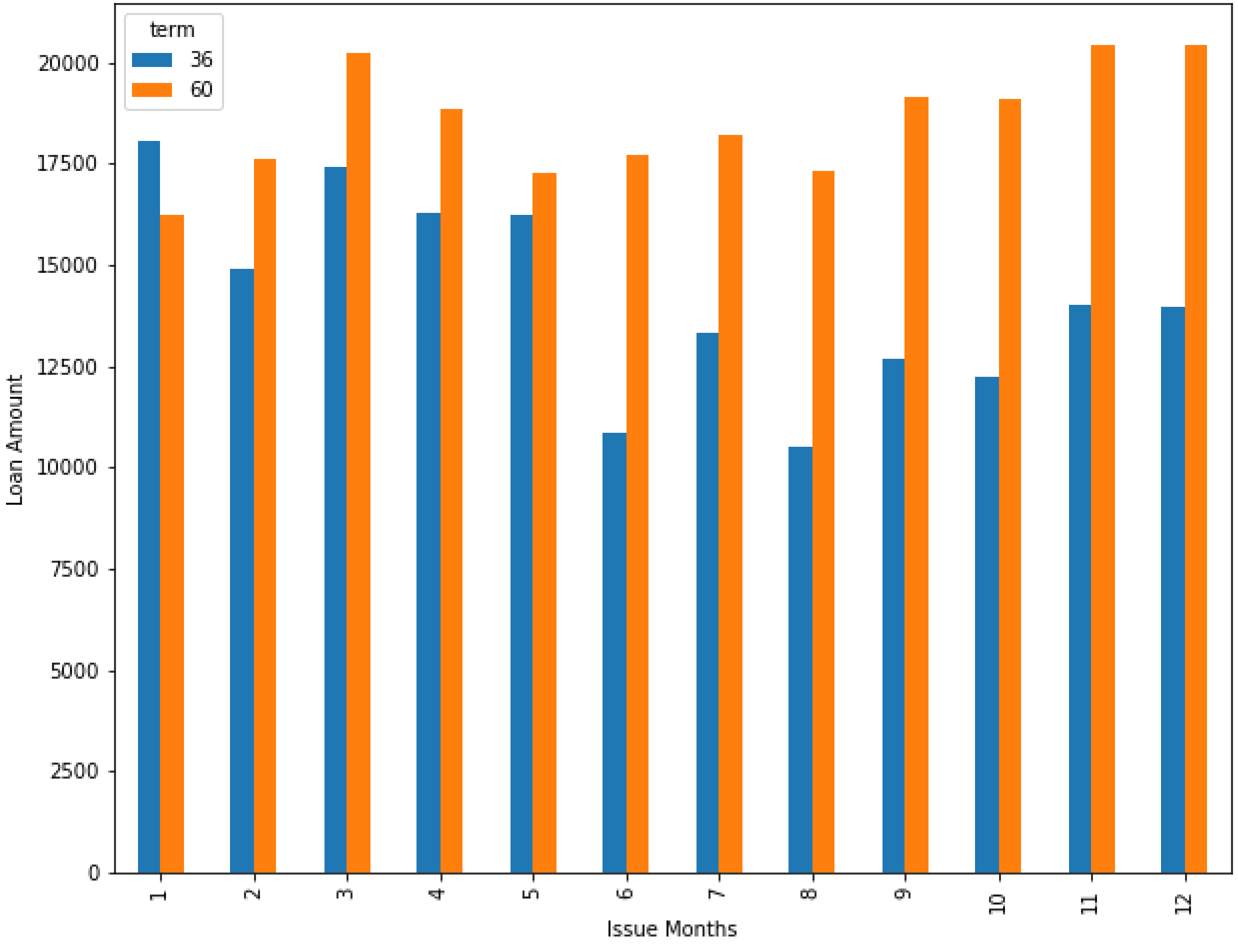
Recommendation: LC can consider determining better calculation of interest rate



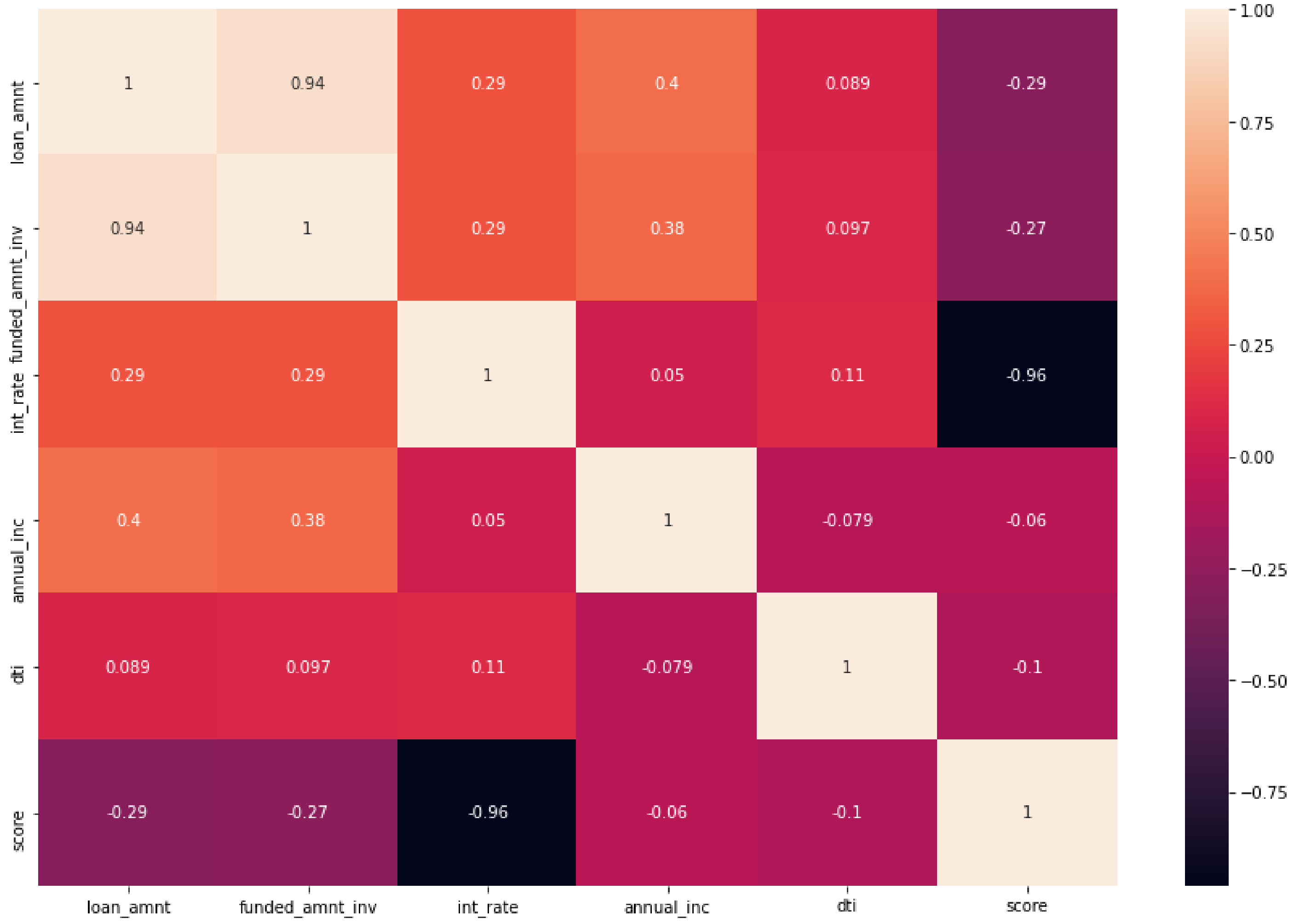
Loan Terms vs Issue Month

Insight: The demand for 36-Month Term loans seems to decrease in the second and third quarter of the year.

Recommendation: Lending Club can prioritize more in availing funds for 36-month term loans in **first** and **last quarter** of the year; and focus gathering funds for 60-month term loans in **second** and **third** quarter of the year.



Loan Sub Grade Analysis

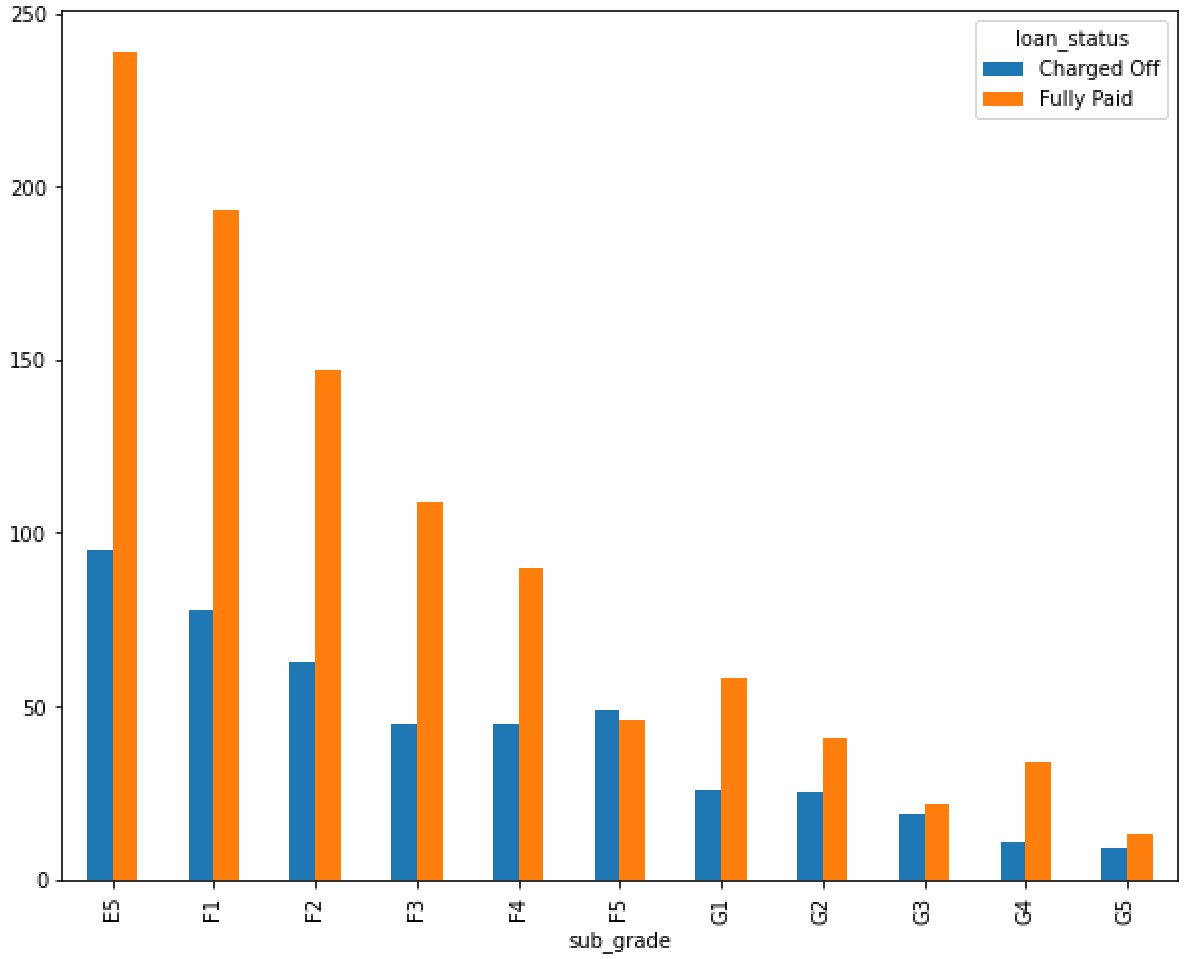


Note: Scores are assigned between the range 30%-100% with variation of 2% for individual loan subgrades from G5 to A1.

Insight: Watching the strong negative correlation between score and interest rate, I can say that a person with a good score (i.e: good loan sub-grade) has a strong chance of obtaining very favorable interest rate. However we can not say that person is as likely to apply for large amount.

Recommendation: This is where LC can shed focus and figure out opportunity of marketing their loan facilities to attract potential borrowers in order to improve this correlation in future.

<Loan Sub Grade Analysis>



#Note the vertical value here is COUNT of the scores

Insight: Below 5th percentile of data, the cost of 10 good customer is loss from approx 5 bad customers (ratio 1:2).

Recommendation: Even though this ratio seems to diminish with quantity, however this information can be vital in calculating marginal cost /benefit or financial risk diversification.

Top 6 Driver Variables!

So the top 6 important driver variables we can strongly suggest:

1) Annual Income – As different mid range income earner are most defaulting especially those who pay rent or mortgage for residence.

2) Address State – As a particular state (e.g: Nebraska in our case) can show high rate of defaulting.

3) Interest Rate – As population with Charged-Off loan status has been experiencing rising interest rate every year.

4) Loan Terms- As demand for the 2 different loan terms vary across different quarters of the year.

5) Purpose – As particular category (e.g: small business in our case) may have significantly high rate of defaulting.

6) Loan Sub Grade – As particular loan sub grades tend to default significantly higher.