



Executive PG Programme in ML & AI



Lending Club Case Study

Submitted by -
Salil Chandan & Shubhankar Chahar

Objective:

In this case study, we are attempting to solve a real world business problem using Exploratory Data Science techniques. We will be understanding and solving a risk analytics problem in Banking and Financial Domain. We will be checking how data can be used effectively to solve business problems like defaulters prediction in Loan Lending club.

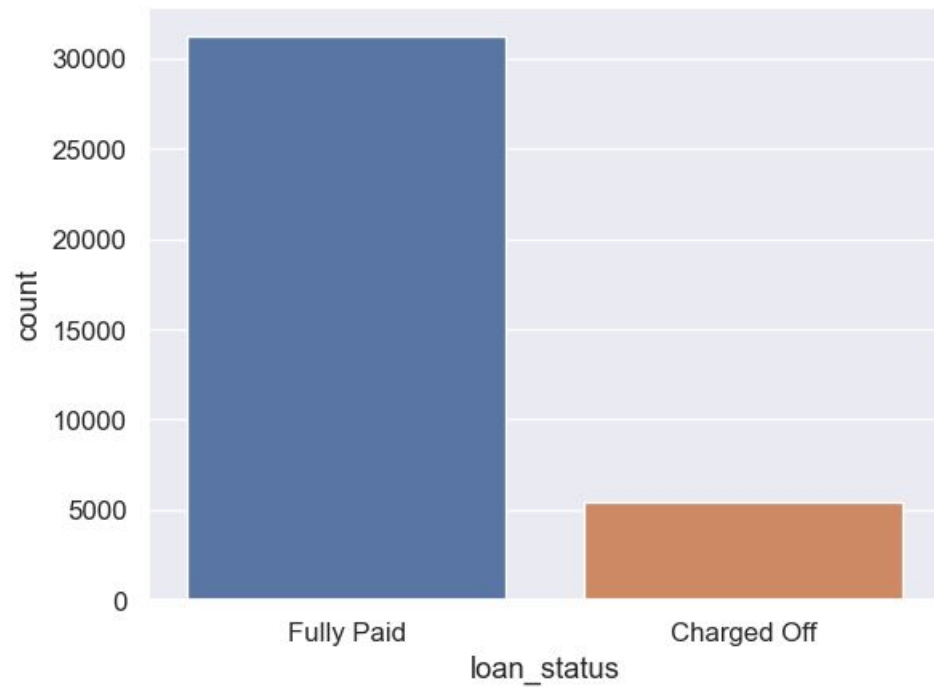
We need to identify variables which are strong indicators of default and potentially use the insights in approval /rejection decision making.

Types of variables:

- Customer (applicant) demographic
- Loan related information & characteristics
- Customer behaviour (if the loan is granted)

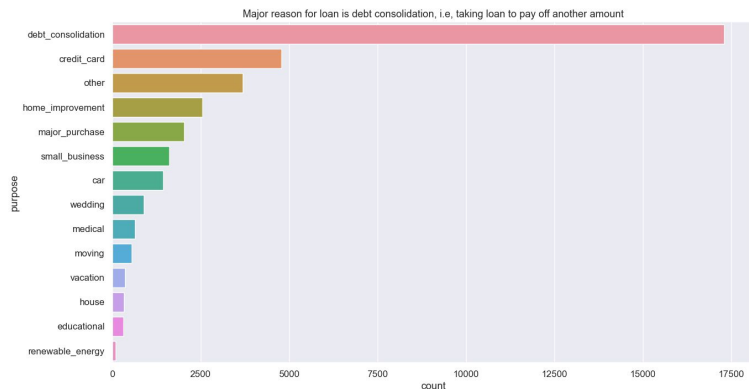
Customer demographic	Loan related information & characteristics	Customer behaviour
Employment Length	Loan Amount	Delinquency Year
Employment Title	Funded Amount	Earliest credit line
Annual Income	Interest Rate	Revolving balance
Zip Code	Loan Status	Recoveries
Description	Loan Grade	Loan Purpose

Understanding Data: Ratio of Defaulters as per historical data

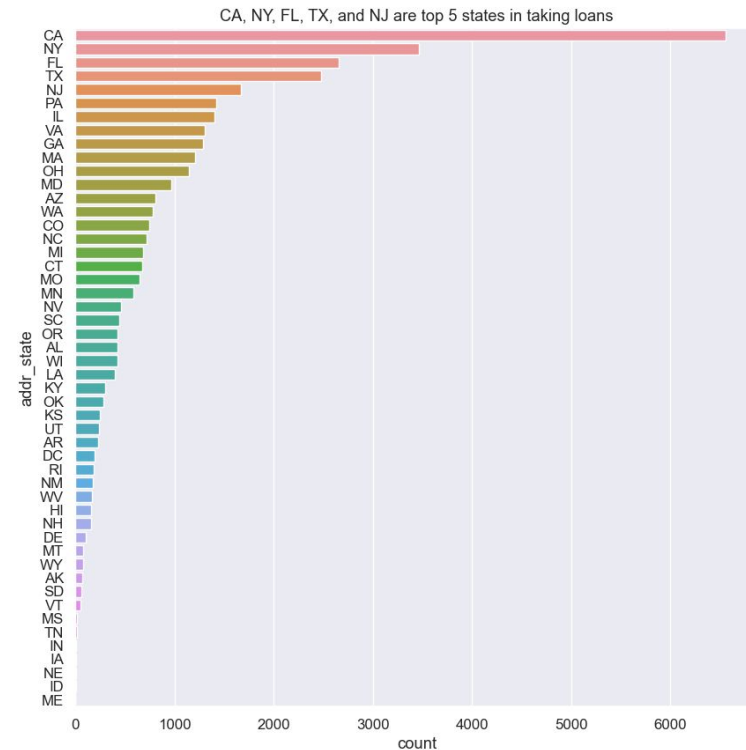


Insights to consider - Univariate Analysis

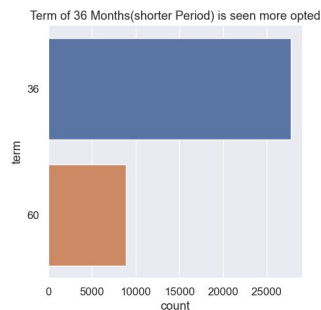
1) Debt Consolidation is major reason for taking loans



2) CA, NY, FL, TX, and NJ are top 5 states in taking loans



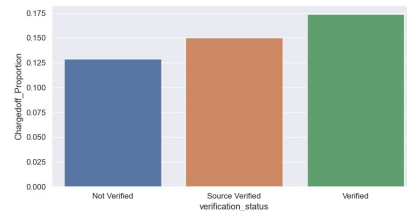
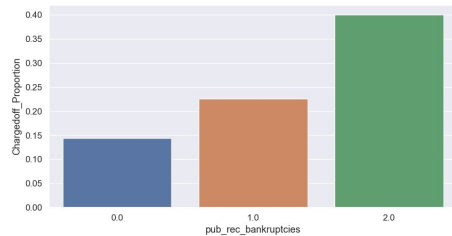
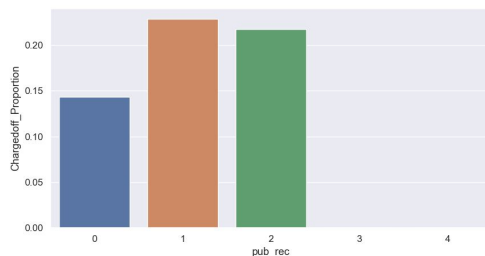
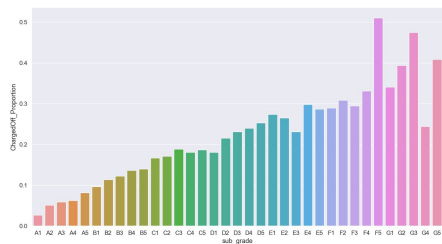
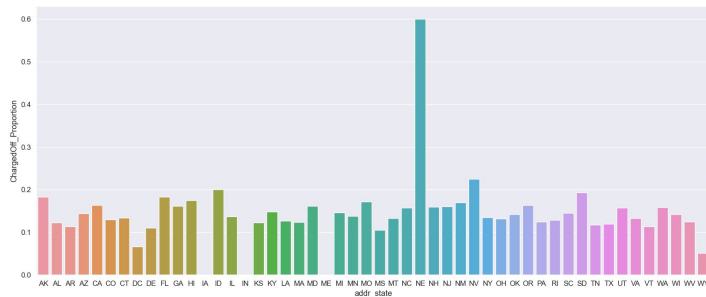
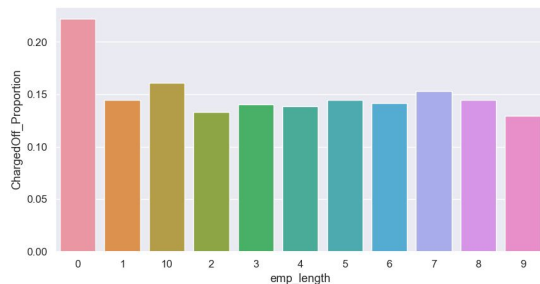
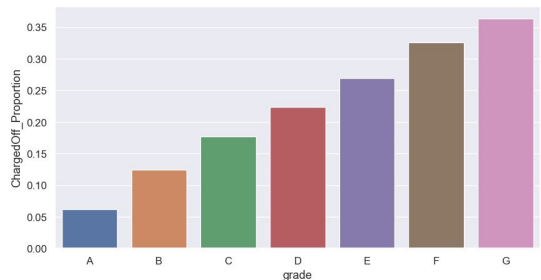
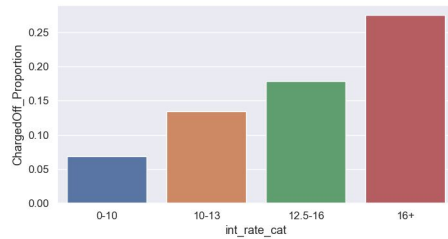
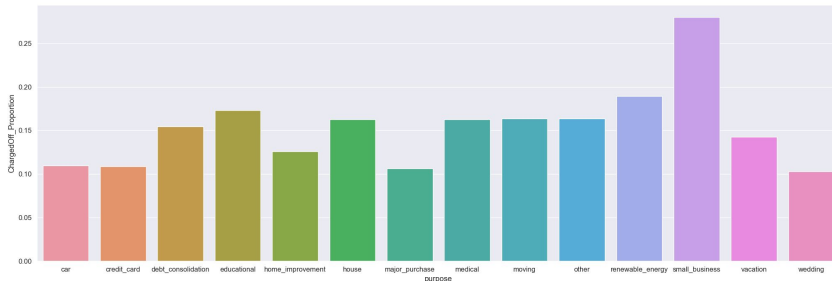
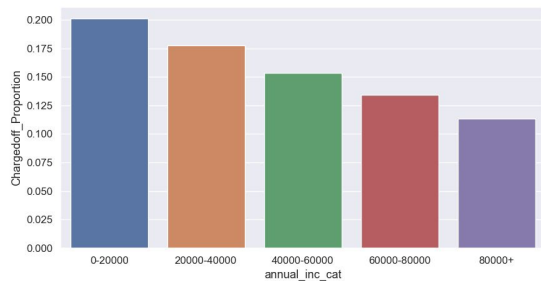
3) Term of 36 Months(shorter Period) is seen more opted



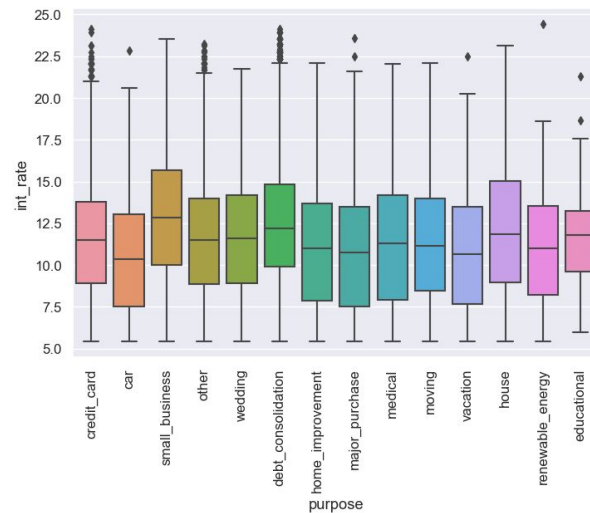
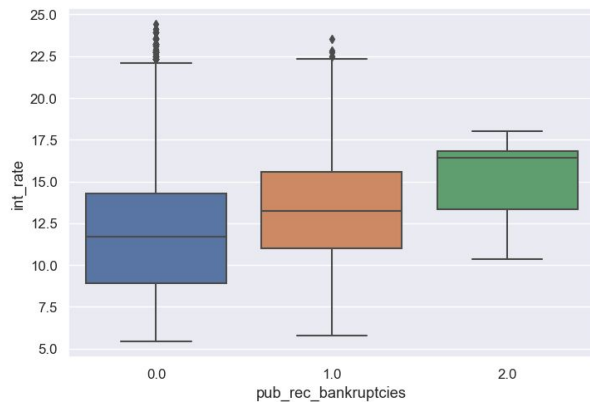
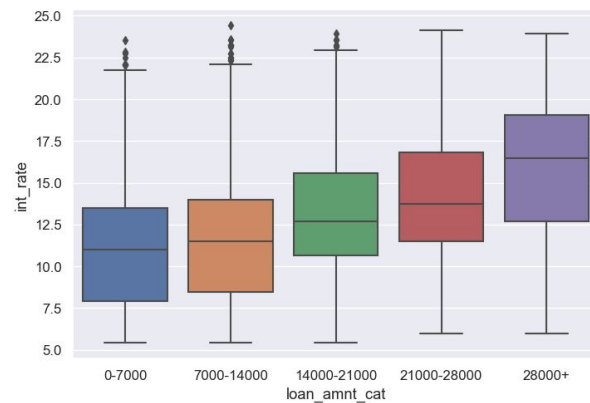
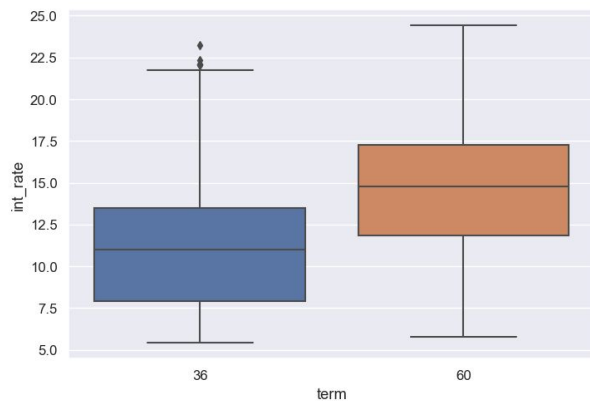
After Analysing important fields, we got these results and their relevance status in prediction of defaulters.

1. Income range 80000+ has less chances of charged off.
2. Income range 0-20000 has high chances of charged off.
3. Notice that with increase in annual income charged off proportion got decreased. So, they are inversely proportional.
4. Small Business applicants have high chances of getting charged off.
5. Renewable_energy where charged off proportion is better as compare to other categories.
6. Grade A has least chances of getting charged off
7. Grade F and Grade G have high chances of getting charged off
8. Chances of getting charged off is increasing with grades moving from A to G
9. Sub Grade A has least chances of getting charged off
10. Sub Grade F and Sub Grade G have high chances of getting charged off
11. Chances of getting charged off is increasing with Subgrades moving from A to G
12. Interest Rates which are less than 10% have very less chances of charged off.
13. Interest Rates greater than 16
14. Charged Off Proportion increases with higher interest rates
15. Those who are not working or have less than 1 year of work experience have high chances of getting charged off.
16. It makes sense as with less or no experience they don't have source of income to repay loan.
17. Rest of the applicants have more or less same chances of getting charged off.
18. States NE has very high chances of charged off but number of applications are too low to make any decisions.
19. States NV,CA and FL states shows good number of charged offs in good number of applications.

Charts of results on previous page



Analysing interest rate with other features



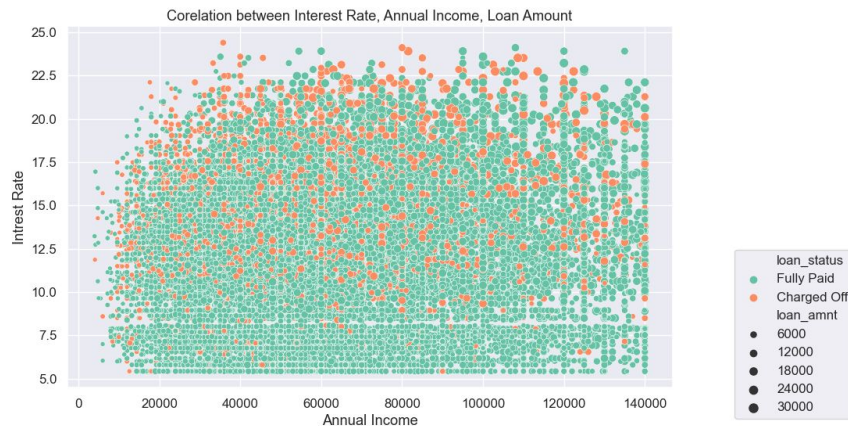
Interpretations from above analysis:

1. It is clear that average interest rate is higher for 60 months loan term.
2. Most of the loans issued for longer term had higher interest rates for repayment.
3. It is clear that interest rate is increasing with loan amount increase.
4. probably when loan amount is more it is taken for longer loan term, we saw earlier that longer the loan term more the interest rate.
5. Interest Rate increases with increasing number of public bankruptcies.
6. We earlier saw that small business as purpose contributed more to the defaulters. This can be explained with int_rate vs purpose graph as small_business has higher interest rate as compared to others.
7. Other than small_business, debt_consolidation as the purpose also has higher average interest rate with some extreme values (outliers).

Final Results after looking for correlation between interest rate and other features:

Final Observations

1. With increase in loan amount there is an increase in interest rate
2. Interest rate is increasing with loan amount & in increases is there in high charged off.
3. Higher the annual income higher the loan amount, provided interest rate is more.
4. Higher interest rate results in high Charged Off ratio yet numbers of defaulters.
5. For future reference, fields to consider are , Public Records, Purpose, Bankruptcies, etc.; analysis with insights of these columns are given above in the notebook cells.



Though there are no absolute clusters but we can see more green dots collected at the bottom and more reds on top-left corner side. This means that high interest rate and low income are major factors for defaults.

Also size of dots (showing loan amount size), is lower for people with lower income and rises as income rises. Overall, Loan Amount, Interest Rate, and Annual Income are positively correlated. However, for low income rates, we can observe an important insight that there're very few numbers of defaulters and small

However, for low interest rates, we can observe an important insight that there're very few numbers of defaulters and small increase in loan amount, making it a safe category with default status point of view.

Thank you