Lending Club Case Study
BY Veena and Amey

09-02-2022

- Problem Statement
- Data Analysis Approach during Case Study
- Understanding Columns
- Performing Categorization
- Visual Representation of Few Observations
- Recommendations

What is Lending Club?

Lending Club is a marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.

When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are 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

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.

In other words, the 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 utilize this knowledge for its portfolio and risk assessment.

LOAN DATASET

Loan Accepted → Default
Loan Accepted → Non-Default

Loan Rejected
(Not considered in dataset)
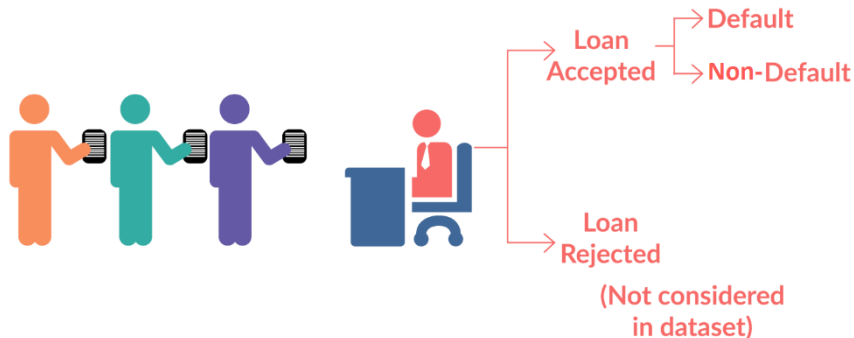
Fully paid: Applicant has fully paid the loan (the principal and the interest rate)

Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.

Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

Approach Followed During Case Study:

1. After having the final list of columns for analysis, each columns is analyzed to find out outliers if any.
2. Some columns had few incorrect data, which were dropped.
3. At this stage, the data is in proper shape for analysis. For analysis of single variable, boxplot and distplot are used.
4. To perform the analysis on multiple variables, few columns are grouped.
5. These grouped columns are used to create stacked bar charts.
6. Then, dataframe is separated for "Fully paid" and "Charged off" status.
7. For analysis of multilple variables for "Fully paid" and "Charged Off" dataframes, heatmaps are used.

List of Columns in final Data Frame

Columns
funded_amnt
term
int_rate
installment
grade
sub_grade
emp_length
home_ownership
annual_inc
verification_status
issue_d
loan_status
purpose
zip_code
addr_state
dti
earliest_cr_line
exp_level
loan_volume
income_group
debt_profile
state_region
int_rate_group
emi_type

Quantitative Variables:

funded_amnt

int_rate

installment

emp_length

annual_inc

dti

Unordered Categorical Variables:

home_ownership

loan_status

purpose

addr_state <==> state_region

Ordered Categorical Variables:

funded_amnt<==>loan_volume

term

int_rate<==>int_rate_group

installment<==>emi_type

grade

sub_grade

emp_length<==>exp_level

annual_inc<==>income_group

verification_status

issue_d

dti<==>debt_profile

We created below categories for ordered and unordered variables

Experience Levels
Freshers: 0 to 2 years
Intermediate: 2 to 5 years
Experienced: 5 to 9 years
Seniors: 10 years and more

Loan Volume
Small Loan: Loan up to USD5k
Medium Loan: Loan between USD5k and USD10k
Big Loan: Loan between USD10k and USD20k
Heavy Loan: Loan more than USD20k

Income Groups
Poor: Annual income up to USD30k
Lower Middle Clss: Annual income between USD30k and USD60k
Middle Class: Annual income between USD60k and USD90k
Upper Middle Class: Annual income between USD90k and USD120k
Rich: Annual income more than USD120k

Debt Profile
Very Low Debt: dti up to 5
Low Debt: dti between 5 and 10
Moderate Debt: dti between 10 and 15
High Debt: dti between 15 and 20
Very High Debt: dti more than 20

Interest Rate Groups
Low Int Rate: interest between 5 and 10
Average Int Rate: interest between 10 and 15
High Int Rate: interest between 15 and 20
Very High Rate: interest more than 20

EMI Level
Low EMI: Installment up to 100
Moderate EMI: Installment between 100 and 300
High EMI: Installment between 300 and 400
Very High EMI: Installment more than 400
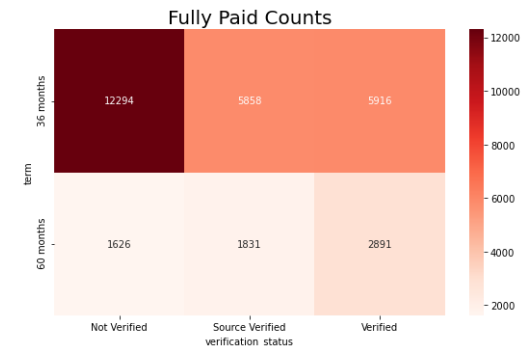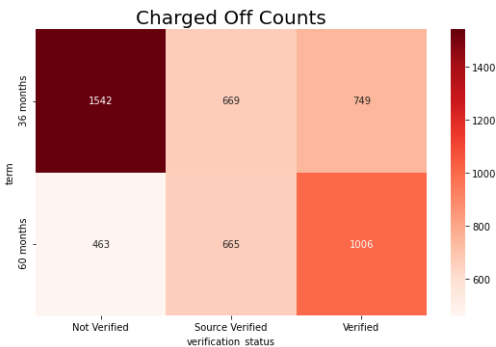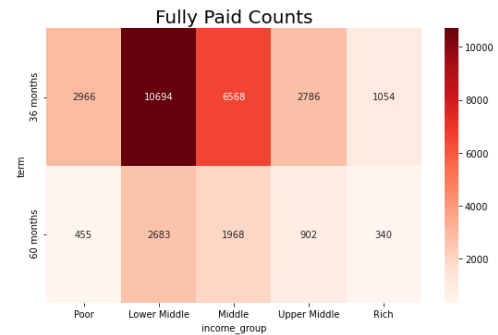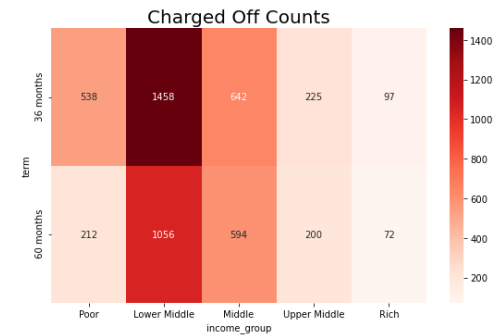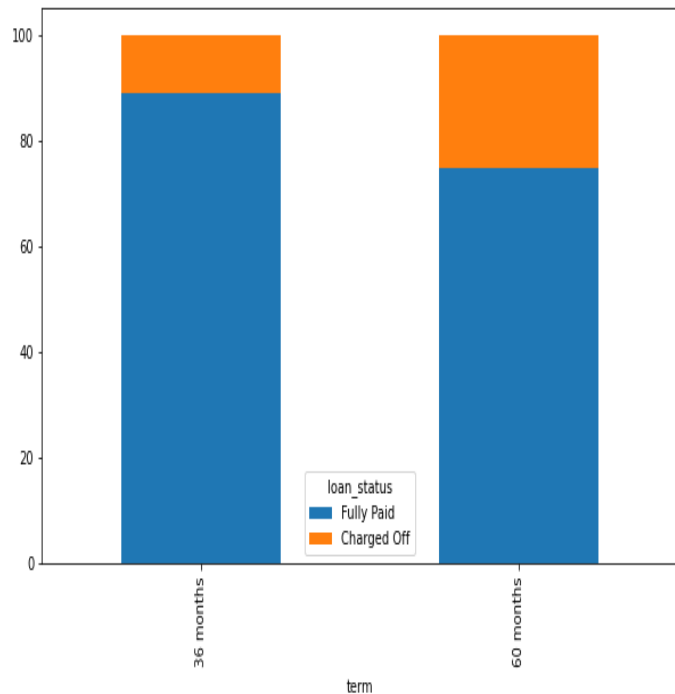
Region wise Categorization
Midwest: 'ND','MN','SD','IA','NE','KS','MO','WI','MI','OH','IN','IL
Northeast: 'PA', 'NY', 'NJ', 'CT', 'RI', 'MA', 'NH', 'VT', 'ME'
West: 'WA','OR', 'MT','ID','WY','NV','UT','CO','CA','AZ','NM', 'AK', 'HI'
South: 'TX','OK','AR','LA','MS','AL','TN','KY','GA','FL','SC','NC','VA','WV','DC','DE','MD'

In general, it is observed people with loan tenure of 5 years have high chances of default compared to loan tenure of 3 years

It is observed, people belonging to grades C,D,E & F have high tendency to default.
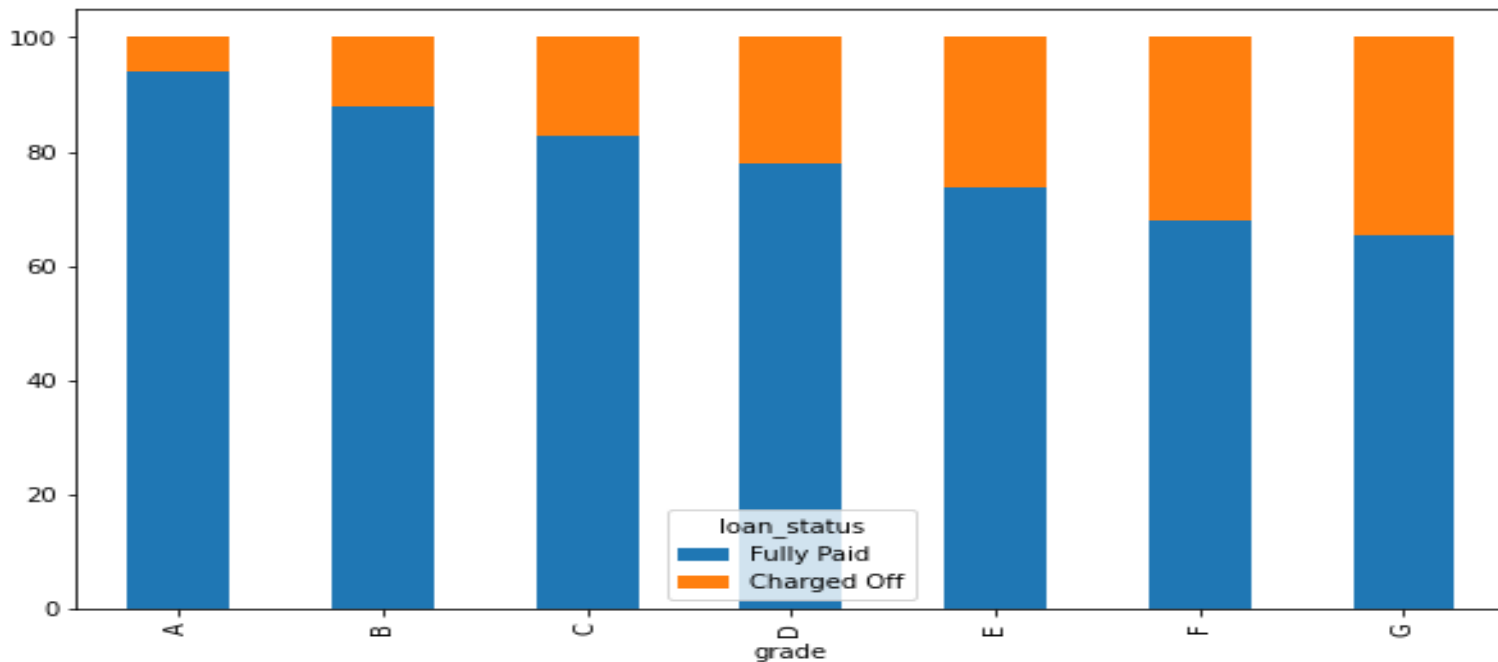
It is also observed that people for whom income source was not verified they tend to default

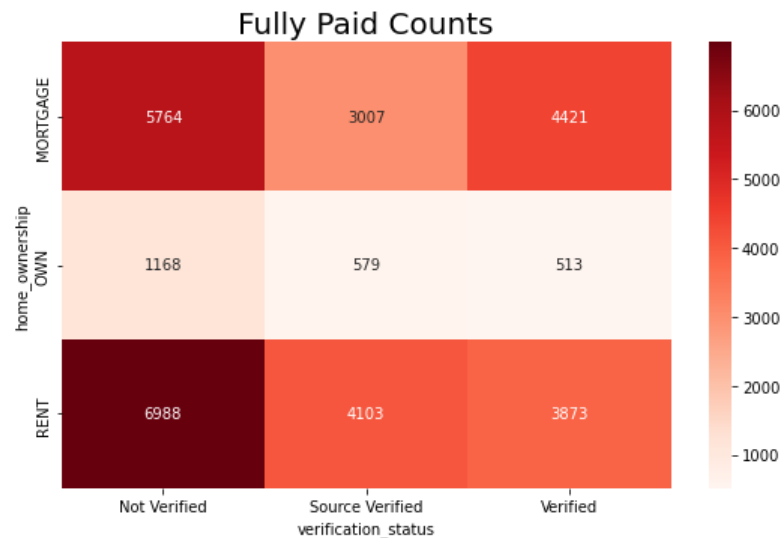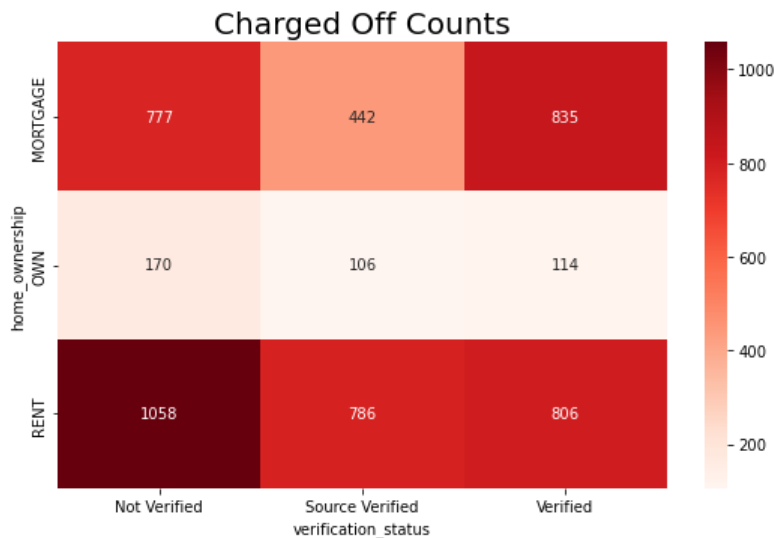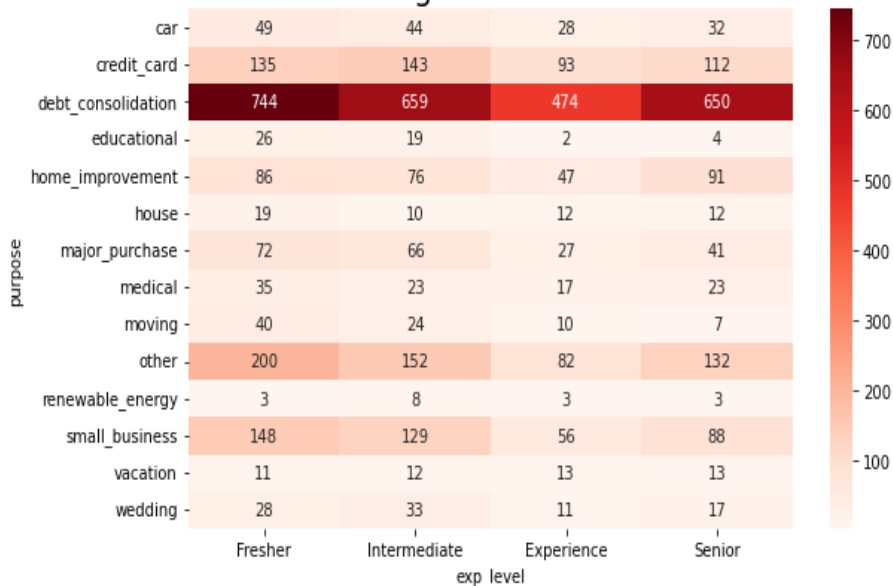In general, it is observed that loan taken for 'small businesses' are risky, followed by 'debt consolidation'

Charged Off Counts

purpose	Fresher	Intermediate	Experience	Senior
car	49	44	28	32
credit_card	135	143	93	112
debt_consolidation	744	659	474	650
educational	26	19	2	4
home_improvement	86	76	47	91
house	19	10	12	12
major_purchase	72	66	27	41
medical	35	23	17	23
moving	40	24	10	7
other	200	152	82	132
renewable_energy	3	8	3	3
small_business	148	129	56	88
vacation	11	12	13	13
wedding	28	33	11	17

exp_level

Fully Paid Counts

purpose	Fresher	Intermediate	Experience	Senior
car	409	344	240	267
credit_card	1407	1152	728	888
debt_consolidation	4494	4085	2581	3093
educational	130	59	32	31
home_improvement	506	588	417	680
house	90	91	40	51
major_purchase	592	541	288	356
medical	167	139	80	137
moving	209	126	56	58
other	1062	802	479	630
renewable_energy	15	13	20	21
small_business	413	320	193	209
vacation	68	98	45	80
wedding	307	285	119	85

exp_level

Recommendations

- In general, it is observed people with loan tenure of 60months have high chances of default compared to loan tenure of 36months.
- It is also observed, people belonging to grades C,D,E & F have high tendency to default.
- It is also observed that people for whom income source was not verified they tend to default.
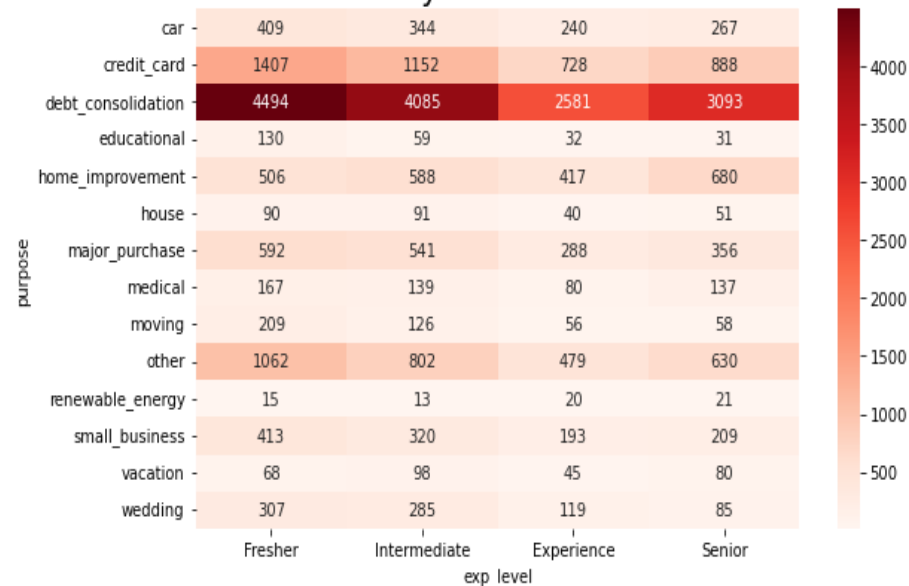- Other observation is that people who stay on rent or have their house as mortgage have a tendency to default.
- Also, people with dti between 15-20 are more likely to default.

THANK YOU