A STUDY ON LOAN DEFAULTS

POWERED BY LENDING CLUB DATA

Problem Statement

Identification of the risky loan applications by understanding the past history of charged-off borrowers so that the potential credit loss can be prevented by cutting down such applications.

ANALYSIS APPROACH

Steps:

- Extraction of data to python notebook
- Cleaning of data by handling missing values, assigning to the variables to correct data type, maintaining the format and keeping only the data which is relevant for analysis
- Creating new metrics by transforming the existing data as and when required to perform comprehensive assessment.
- Performing univariate/multivariate analysis and creating respective visualizations.
- Drawing insights from analysis

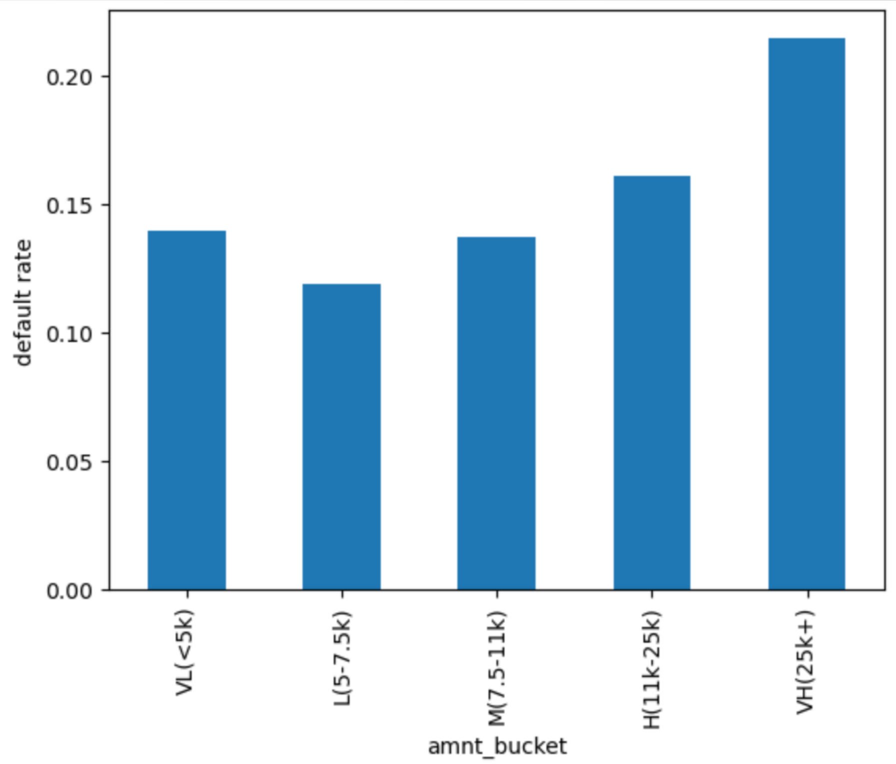
Assumptions:

- **Default rate or percentage calculated from the ratio of no of default case and total loan application in any segment.**
- **The currency of loan amount has been assumed to be USD.**
- **The applicants, who do not provide their employment length are assumed to be self employed.**
- **The other category in Home Ownership comprises of people who are homeless or they stay at their friends' or parents' place.**
- **The number of credit inquiries made, is excluding auto and mortgage inquiries**

INSIGHTS

FUNDED AMOUNT OF THE LOAN:

Overall the funded loan amount has no to minimal impact on default behaviour. However, if loan amount is more than 25K, the chances is defaulting is getting increased by at least 30% than the applications where funded amount is less. Same has been depicted in this plot, where the highest category of funded amount starts from 25K USD

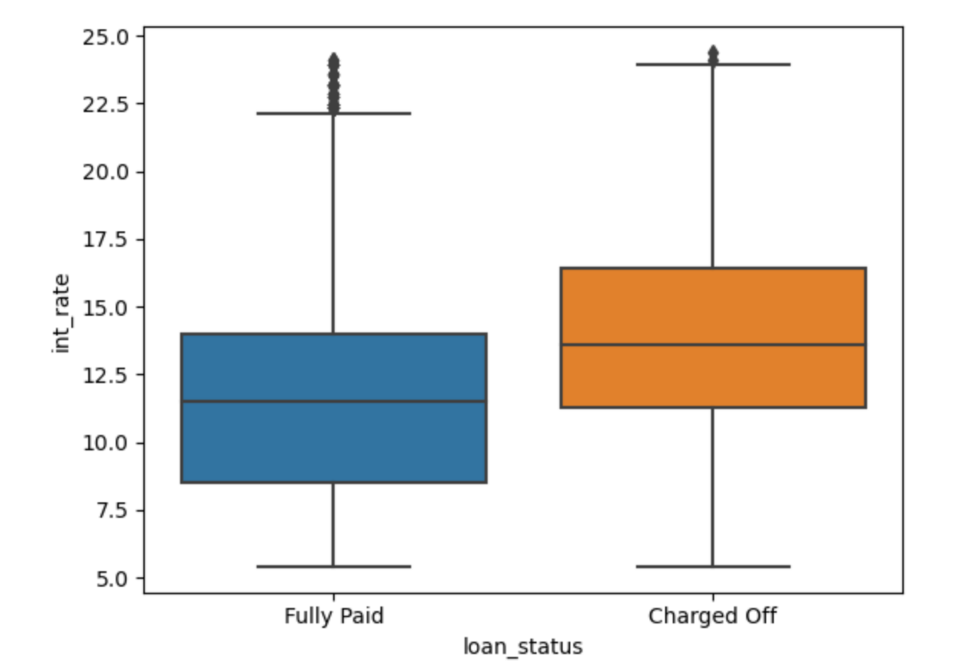


INSIGHTS

INTEREST RATE OF THE LOAN:

Borrowers accepting high interest of loan, tend to default more. The left plot is showing for loans which are defaulted, the median (50th Percentile of interest rate is) higher than that of fully paid loans. Below default rate for interest rates is also supporting same behaviour

loan_status	Charged Off	Fully Paid
int_bucket		
L(0-16)	0.122554	0.877446
M(6-20)	0.272962	0.727038
H(20-25)	0.384416	0.615584

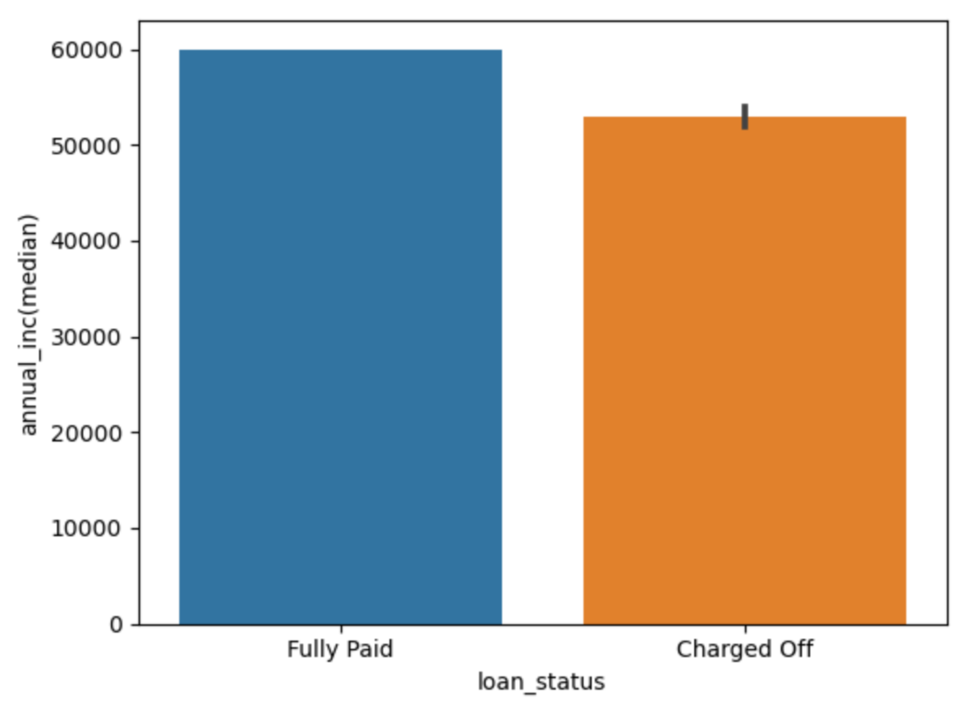


INSIGHTS

ANNUAL INCOME:

From this plot and values, it can be inferred that applicants who have low income, tend to default more. Applicants with low income have at least 36% more chances of default than their high income counterparts

loan_status income_bucket	Charged Off	Fully Paid
L(0-60k)	0.166023	0.833977
M(60-200k)	0.122433	0.877567
H(200-6000k)	0.105263	0.894737

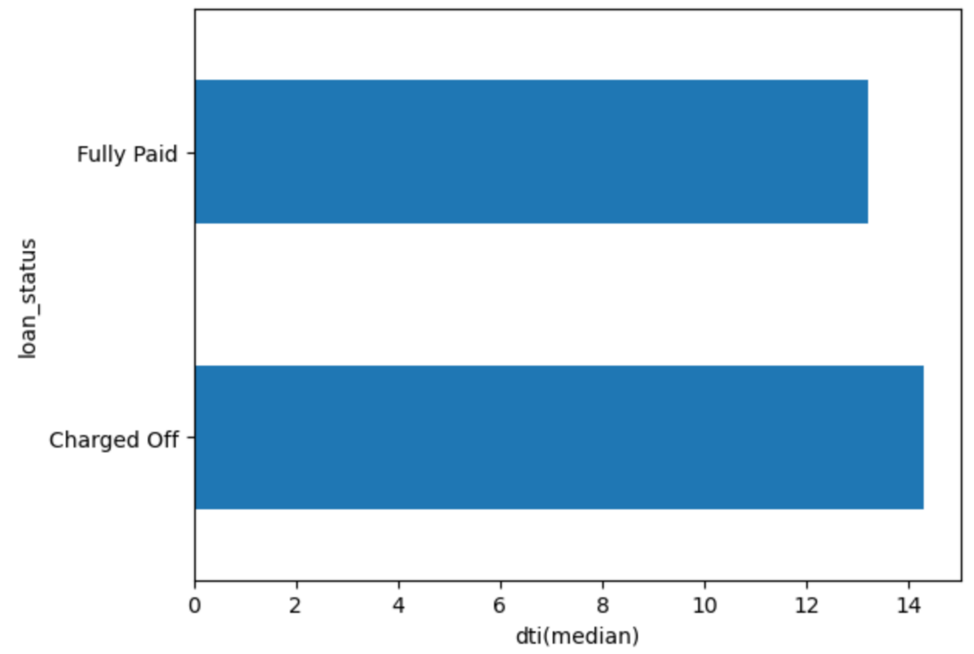


INSIGHTS

DATE TO INCOME RATIO:

Intuitively, applicants with highest monthly debt to income (dti) ratio have 37% of more chances to default than the applicants who have the lowest dti ratio.

loan_status	Charged Off	Fully Paid
dti_bucket		
L(0-8)	0.122536	0.877464
M(8-15)	0.142944	0.857056
H(15-20)	0.157644	0.842356
VH(20-32)	0.167338	0.832662

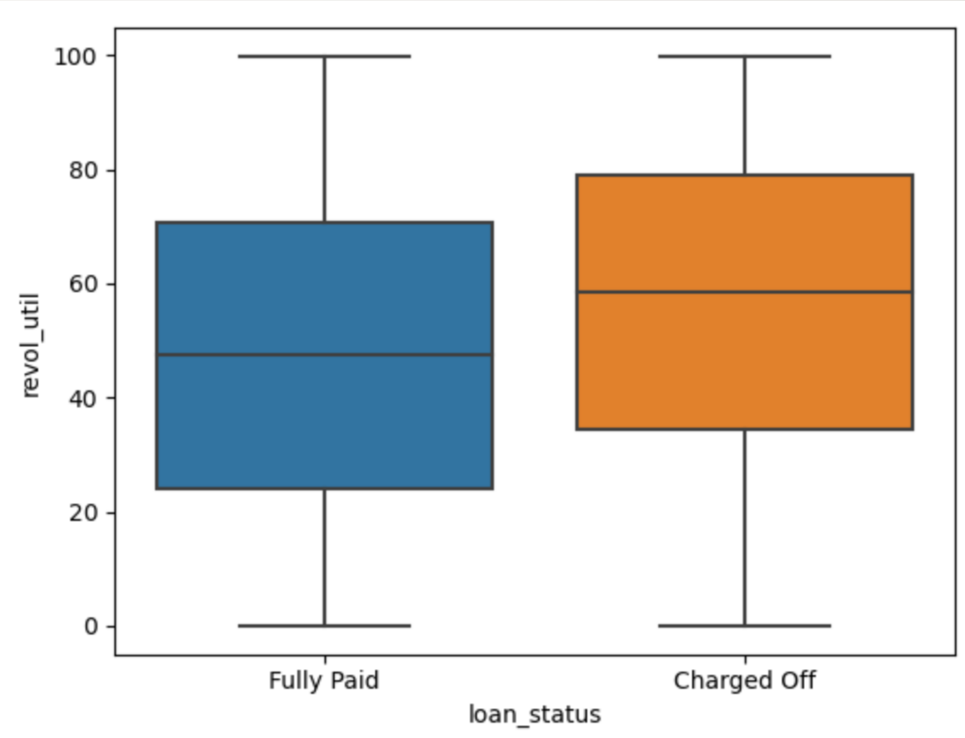


INSIGHTS

REVOLVING LINE UTILIZATION RATE:

Default rate is higher (>30%), when revolving credit line utilization is higher. This is inline with how high credit utilization impacts the credit score. A high credit utilization ratio (meaning applicant is close to maxing out all credit cards/lines) can often lower the credit score and lower credit score indicates a potential defaulter as shown in the plot and below chart of default rate.

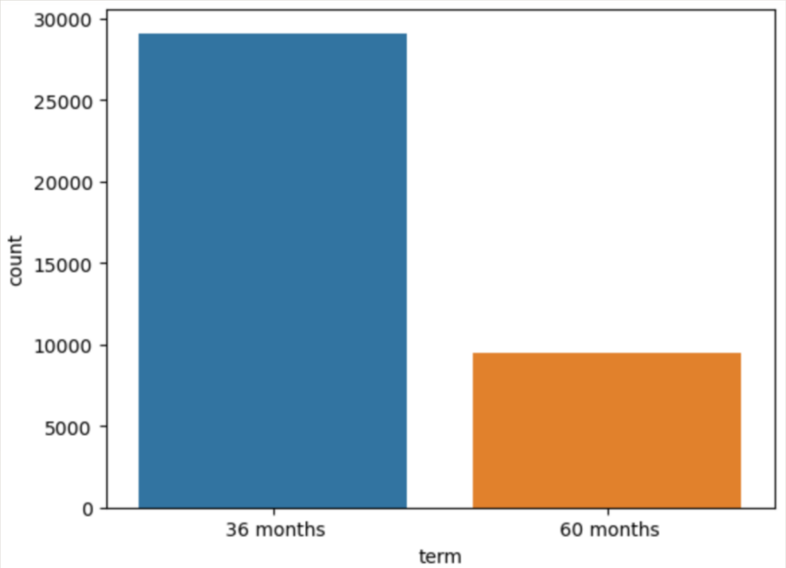
loan_status	Charged Off	Fully Paid
revutil_bucket		
L(0-3%)	0.122439	0.877561
M(3-30%)	0.099319	0.900681
H(30-100%)	0.163558	0.836442

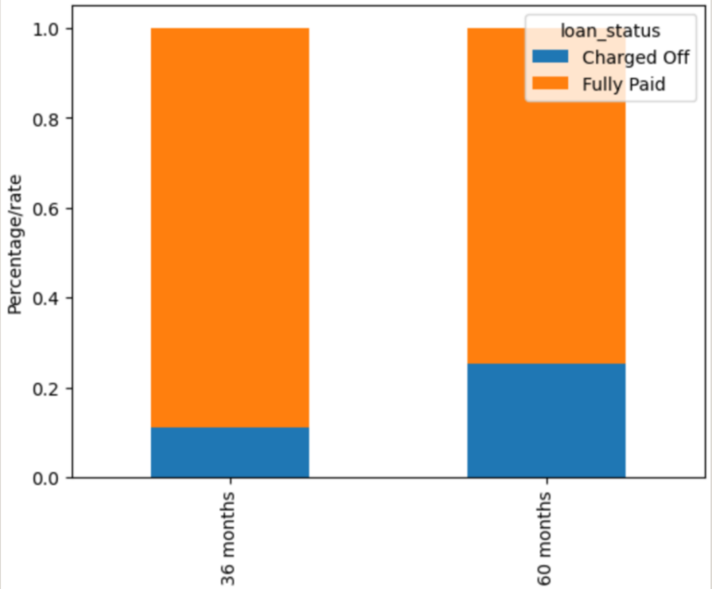


INSIGHTS

THE NUMBER OF PAYMENTS ON THE LOAN:

We have found that most of the loans are tenured for 36 months. Also, the charged-off/default rate (25%) is more than double for 60 months loan while compared to that of the 36 months loan



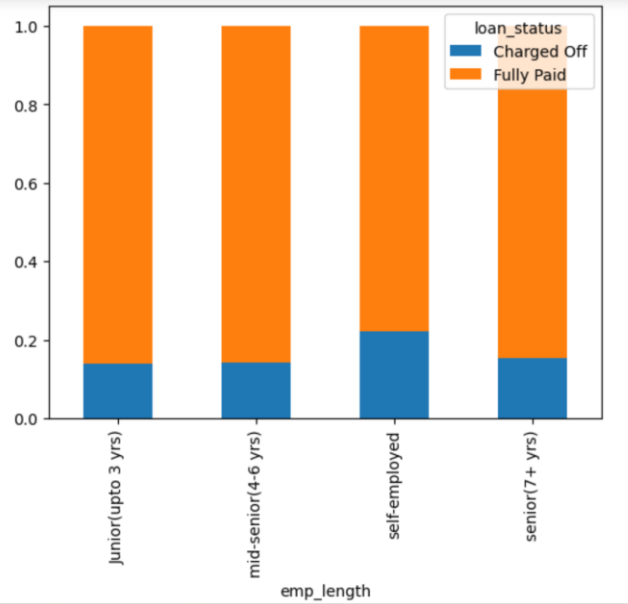


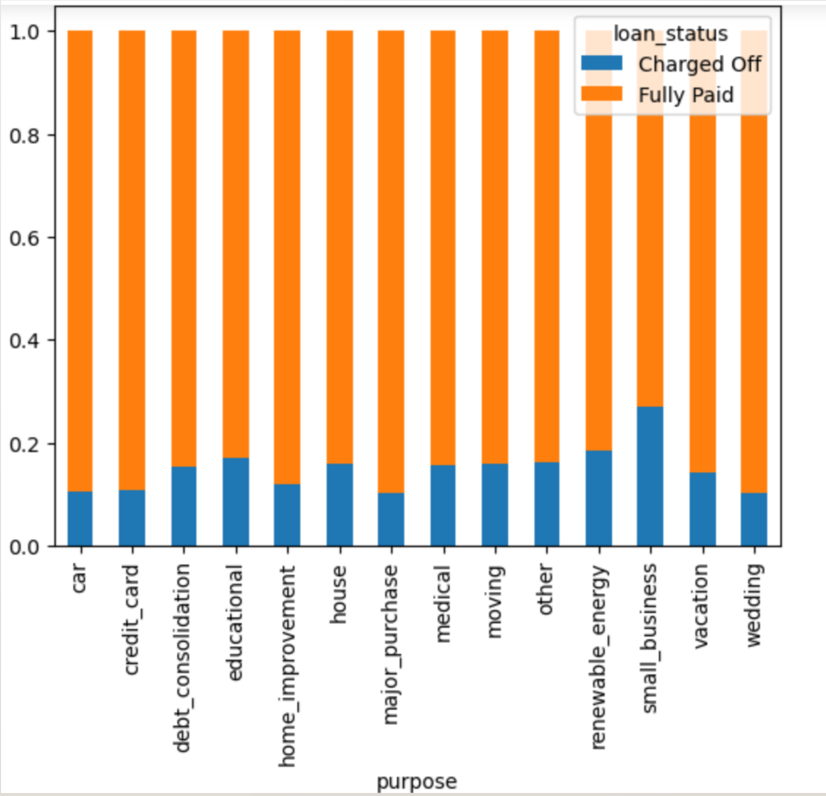
INSIGHTS

EMPLOYMENT LENGTH, PURPOSE OF THE LOAN

The borrowers, who took the loan to support small business, has at least 50% higher default rates than other purposes.

It has also been observed that the borrowers who are self-employed tend to default more.



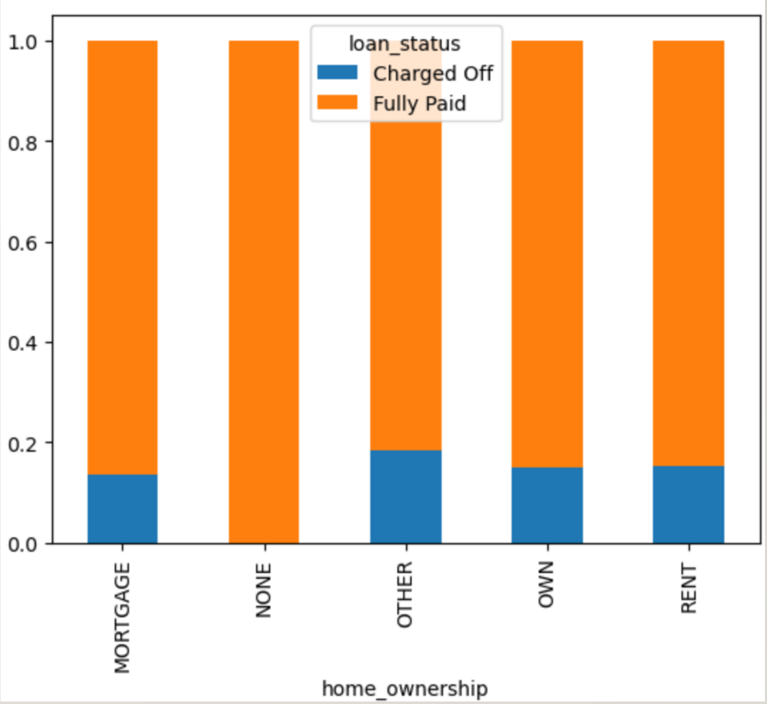


INSIGHTS

HOME OWNERSHIP

It has been observed that applicants who do not have home on their name by any means, tend to default more (at least by 20% than others). This can imply that people who are homeless or stay at their friends' or parents' place have financial struggles and find it difficult to the repay the loan.

loan_status	Charged Off	Fully Paid
home_ownership		
MORTGAGE	0.136713	0.863287
NONE	NaN	1.000000
OTHER	0.183673	0.816327
OWN	0.148908	0.851092
RENT	0.153626	0.846374

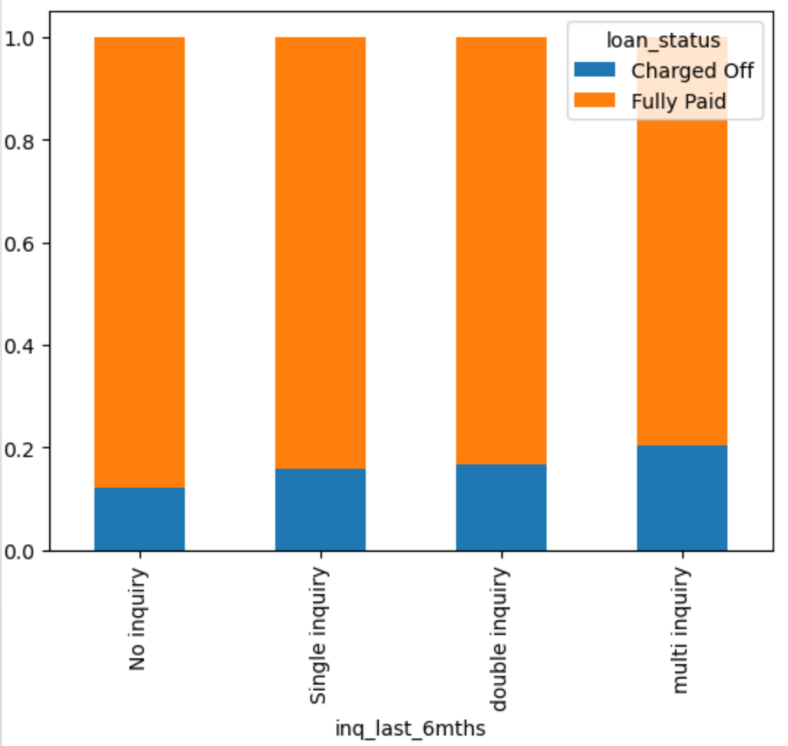


INSIGHTS

THE NUMBER OF INQUIRIES IN PAST 6 MONTHS

Applicants who made multiple(>2) inquiries, have higher default rate by 60% than that who never made any credit inquiry in past six months.

loan_status inq_last_6mths	Charged Off	Fully Paid
No inquiry	0.121866	0.878134
Single inquiry	0.157317	0.842683
double inquiry	0.166785	0.833215
multi inquiry	0.204558	0.795442

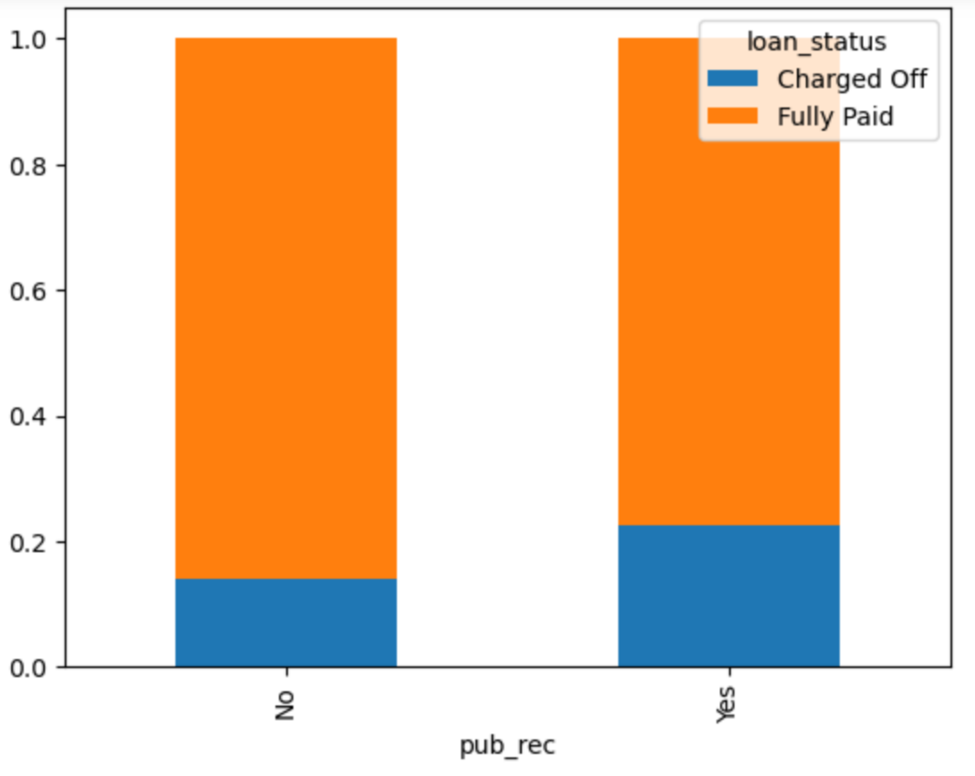


INSIGHTS

NUMBER OF DEROGATORY PUBLIC RECORDS

The borrowers who have derogatory public records (like bankruptcy, foreclosure etc.) have higher chances (by 57%) of defaulting than those who do not have any derogatory public records.

loan_status	Charged Off	Fully Paid
pub_rec		
No	0.141343	0.858657
Yes	0.225604	0.774396



RECOMMENDATIONS

Based on the analysis and interaction among different factors , we have identified the below driving factors which are strong indicator of defaults.

- Medium to high interest rate>6% are associated with high default rate. Applicants who are accepting such high rates should be investigated to avoid loan default. Special attention should be given when the loan has such interest rates with amount more than 5K USD.
- The revolving utilisation is an important parameter to identify potential defaulter as this accounts for individual's credit score also. Borrowers with revolving balance greater than 30% should be checked further as they depend too much on their available credit limit for their expenses.
- The applicants with low annual income(<60K) should be investigated further whether they would be able repay the loan. Special attention should be given if the applicant is self-employed.

RECOMMENDATIONS

- The borrowers who have derogatory public record and made any inquiry in past six months, should be investigated further to cut the default rate.
- Loan applied for small businesses should be checked further whether they have any collateral to avoid defaulting.
- Loans with tenure of 60 months have very high chance (almost double) to default across every purpose of loan category. Whenever such long terms loans are applied for any reason apart from car, home improvement, major purchase, moving and wedding, stringent policies should be in place to further investigate the applicants.

CONCLUSION

- The recommended attributes should be checked against the future loan applicants to identify the risky ones who will most likely default. Such application can rejected or amount can be lowered as per business decision.
- The analysis is done on data of year 2027 to 2011, recent trends can be found if analysis can be done on more recent data.
- It has been seen that the California has largest share of loan application being the most populous state in the US with a strong reputation for innovative businesses. However, Nebraska is showing highest rate of defaults although it's economy performed better than other states through the 2007-2009 great recession, this needs further investigation.