Loan Club Case Study

26-06-2024

By Group Members
Ekansh Chaturvedi
Faizan Sharieff

Table of Content

Data Understanding

Leading Attribute

Loan Status - Key Leading Attribute (loan_status). The column has three distinct values

- Fully-Paid - The customer has successfuly paid the loan
- Charged-Off - The customer is "Charged-Off" has "Defaulted"
- Current - These customers, the loan is currently in progress and cannot contribute to conclusive evidence if the customer will default of pay in future

For the given case study, "Current" status rows will be ignored

Decision Matrix

1. Loan Accepted - Three Scenarios

- 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

2. Loan Rejected - The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

Important Columns

The given columns are leading attributes, or predictors. These attributes are available at the time of the loan application and strongly helps in prediction of loan pass or rejection. Key attributes Some of these columns may get dropped due to empty data in the dataset

Customer Demographics

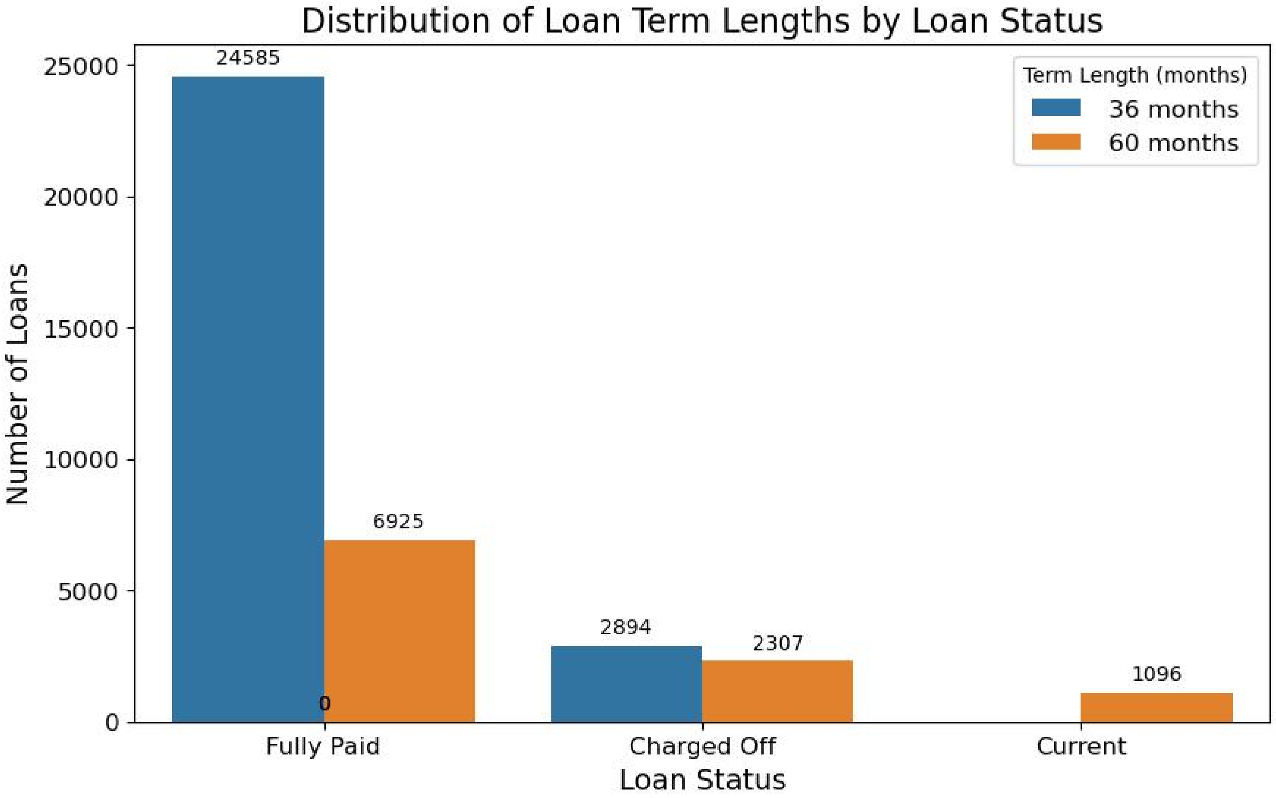
- Annual Income (annual_inc) - Annual income of the customer. Generally higher the income, more chances of loan pass
- Home Ownership (home_ownership) - Wether the customer owns a home or stays rented. Owning a home adds a collateral which increases the chances of loan pass.
- Employment Length (emp_length) - Employment tenure of a customer (this is overall tenure). Higher the tenure, more financial stablity, thus higher chances of loan pass
- Debt to Income (dti) - The percentage of the salary which goes towards paying loan. Lower DTI, higher the chances of a loan pass.
- State (addr_state) - Location of the customer. Can be used to create a generic demographic analysis. There could be higher delinquency or defaulters demographicaly.

Loan Attributes

- Loan Ammount (loan_amt)
- Grade (grade)
- Term (term)
- Loan Date (issue_date)
- Purpose of Loan (purpose)
- Verification Status (verification_status)
- Interest Rate (int_rate)
- Installment (installment)
- Public Records (public_rec) - Derogatory Public Records. The value adds to the risk to the loan. Higher the value, lower the success rate.
- Public Records Bankruptcy (public_rec_bankruptcy) - Number of bankruptcy records publocally available for the customer. Higher the value, lower is the success rate.

Univariate Analysis: Insights

upGrad

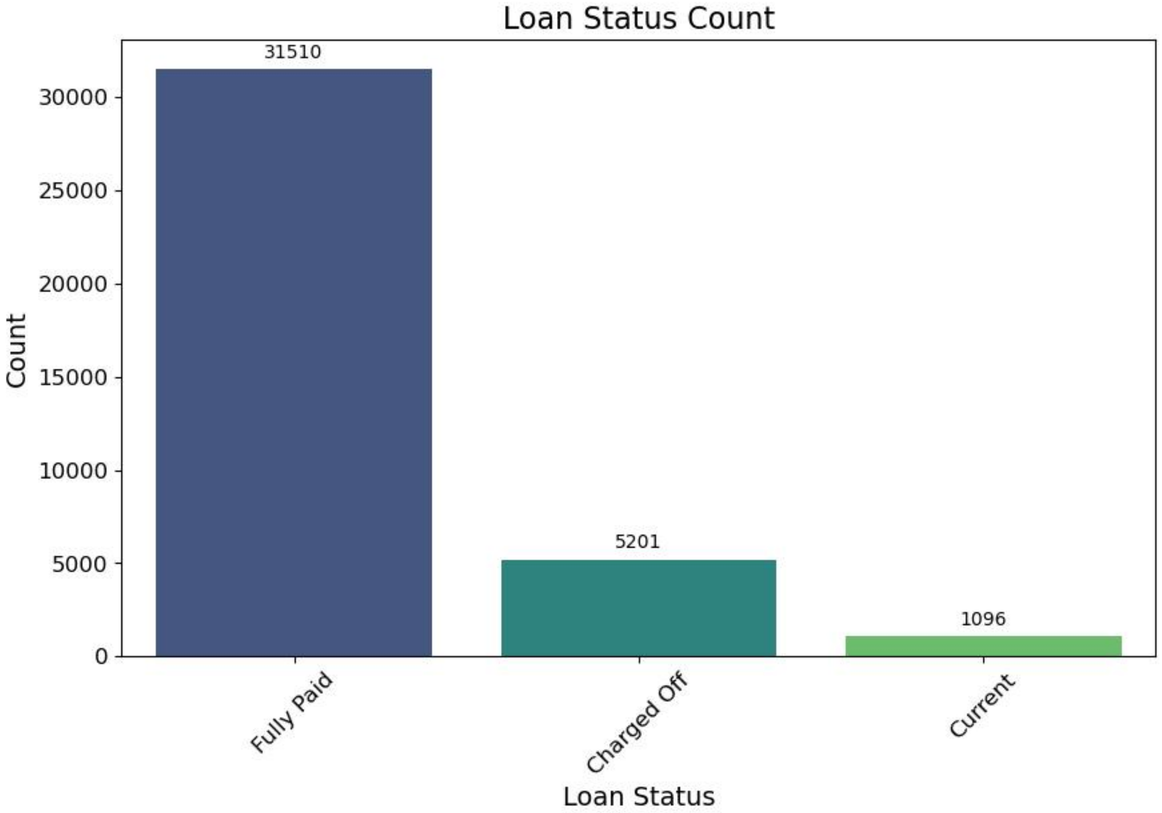


Analysis of Loan Term Lengths and Loan Status
In the above plot, we can see that the dataset contains customers with two types of loan terms: **36 months and 60 months**. More than 30,000 customers have fully paid off their loans, with around 25,000 of them taking loans for a term of 36 months, and only around 7,000 taking loans for a term of 60 months. This indicates that a majority of customers opted for the 36-month loan category..

However, it's interesting to note that the number of customers who were charged off is roughly the same for both terms, at around 2,500 each category. This suggests a higher percentage of customers with 60-month loans are unable to repay their loans compared to those with 36-month loans.

Univariate Analysis: Insights

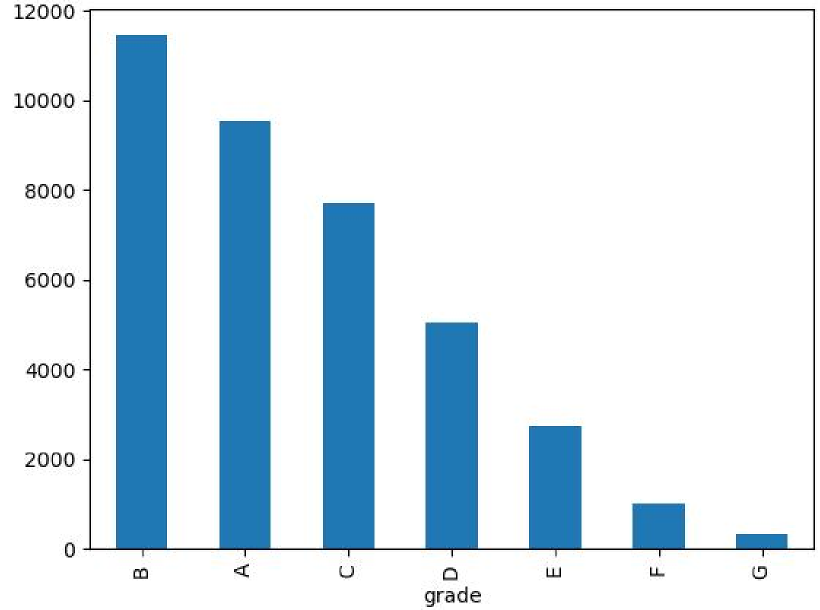
upGrad



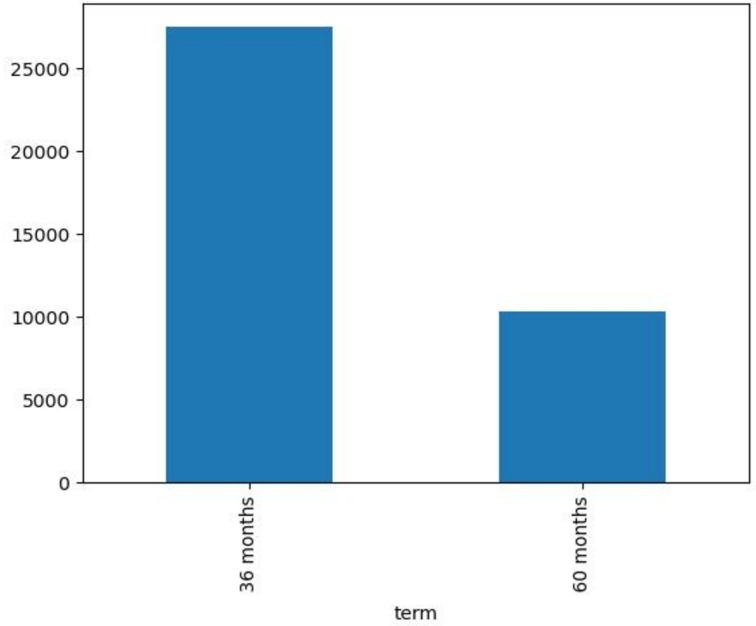
Loan Status of Members for each category(Fully Paid, Charged Off and Current)

31510 members have paid the loan in full, **5201** members have been Charged off, **1096** members are currently still paying

Univariate Analysis Ordered Categorical: Insights



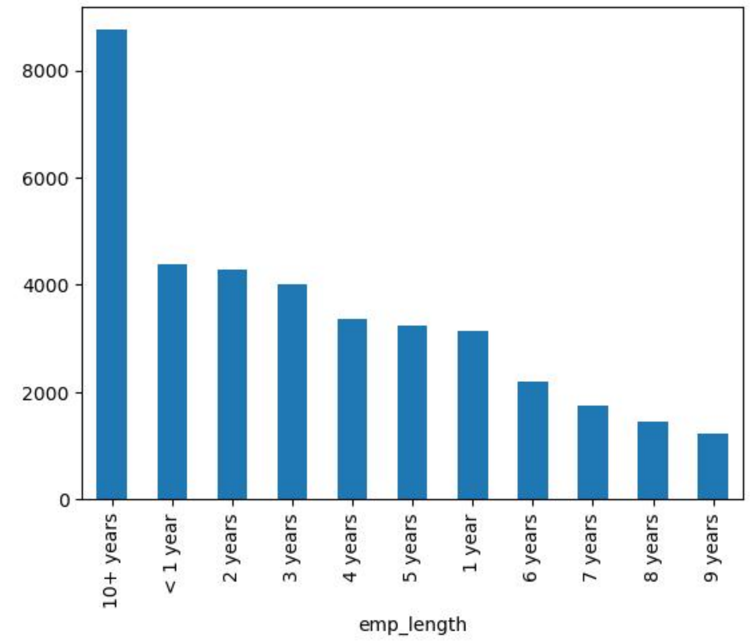
Majority of the Loan applicants fall under the category of Grade B



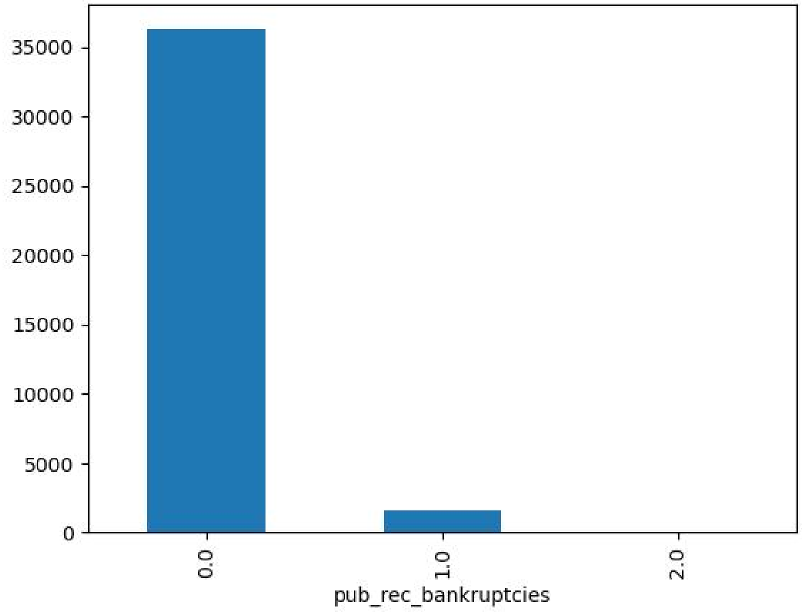
Majority of the loan applications are for 36 months

Univariate Analysis Ordered Categorical: Insights

upGrad



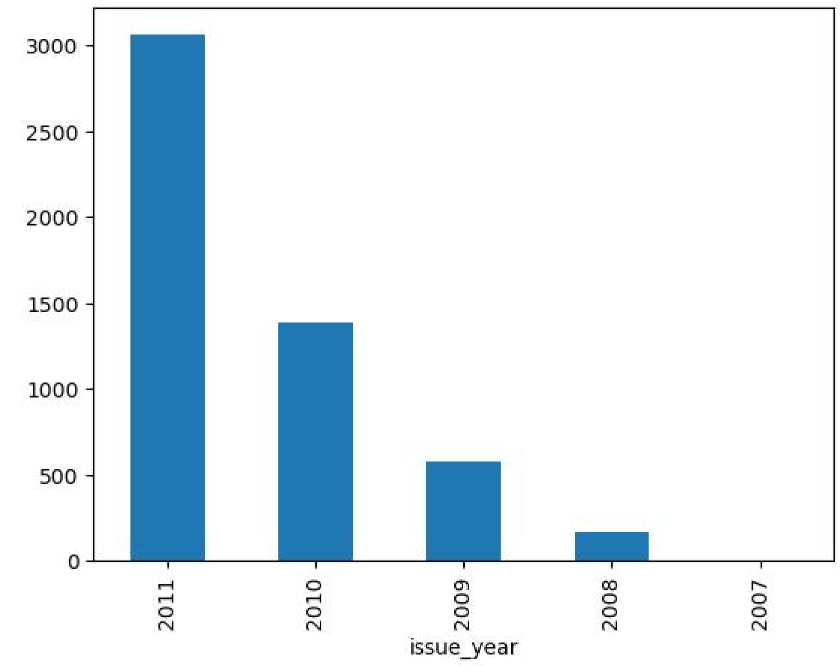
Majority of the loan applications are having more than 10 years of experience



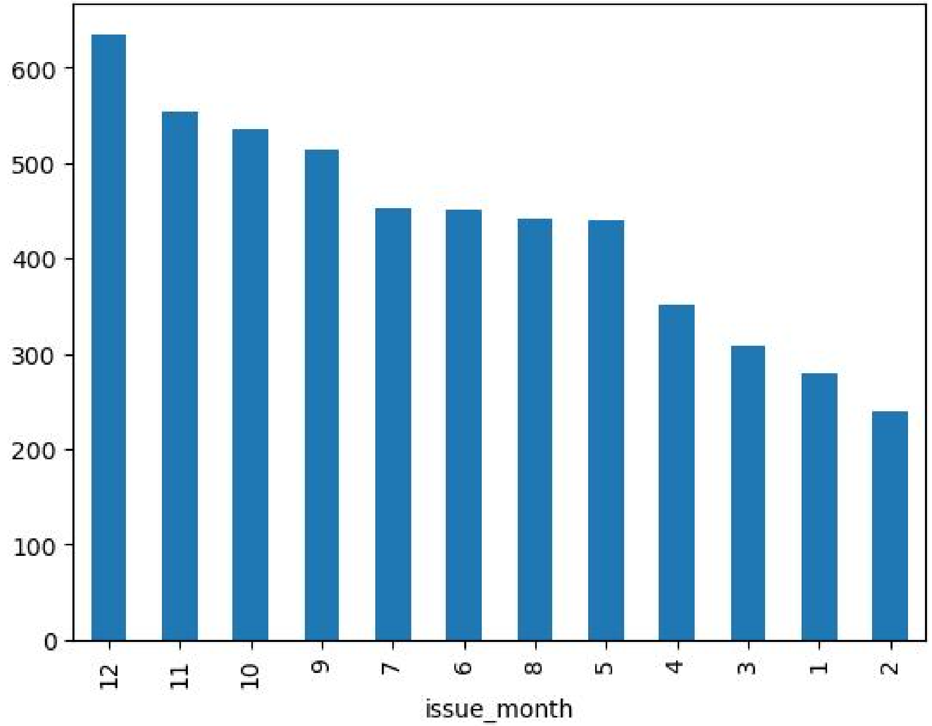
Majority of the loan applications are not having public record of bankruptcies

Univariate Analysis Derived Variables:
Insights

upGrad

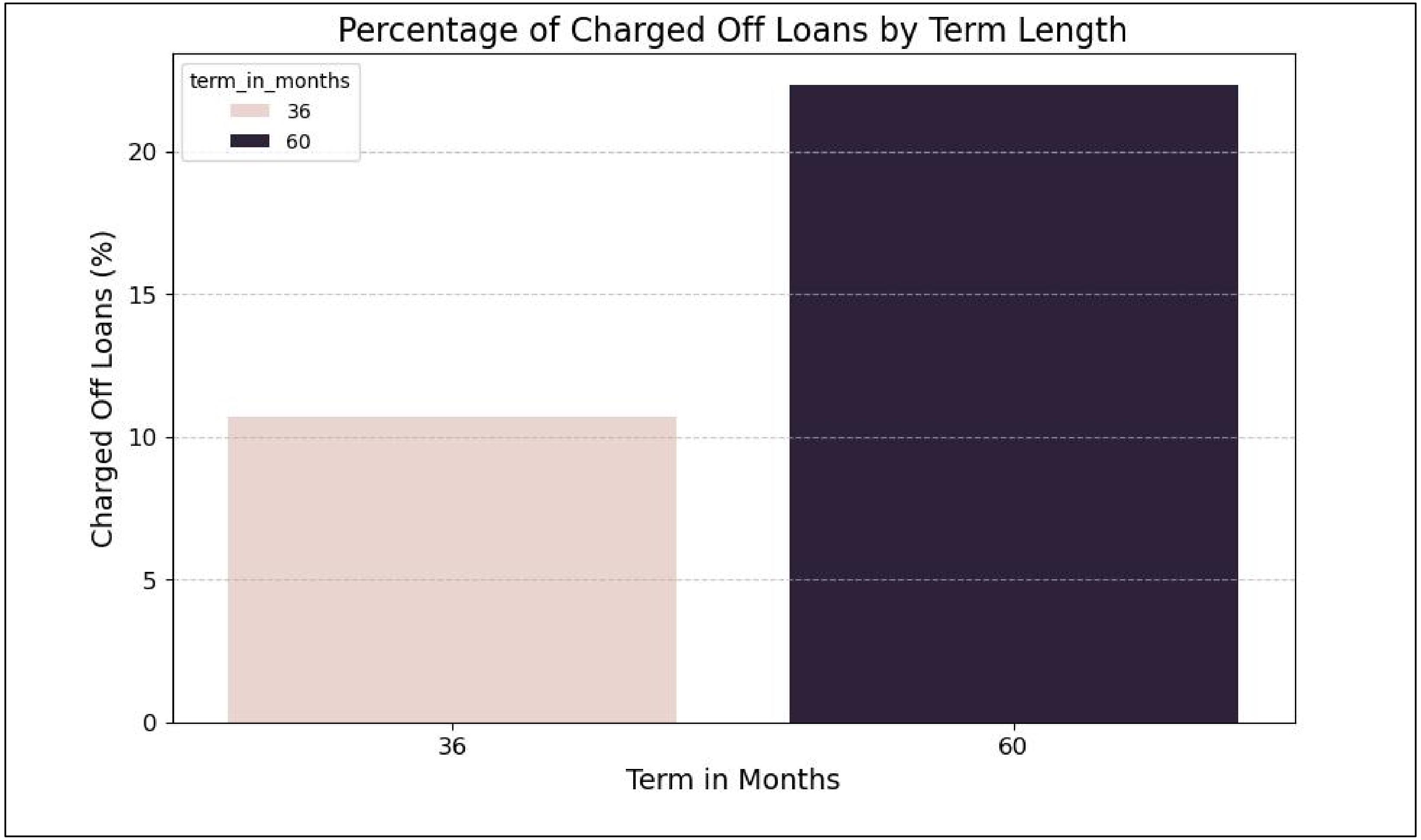


Every year risk is increasing as more write off loans are seen every year



Loan approval team must be cautious during the months of October, November and specially December where a lot of borrowers need loans and will be a risk to the company

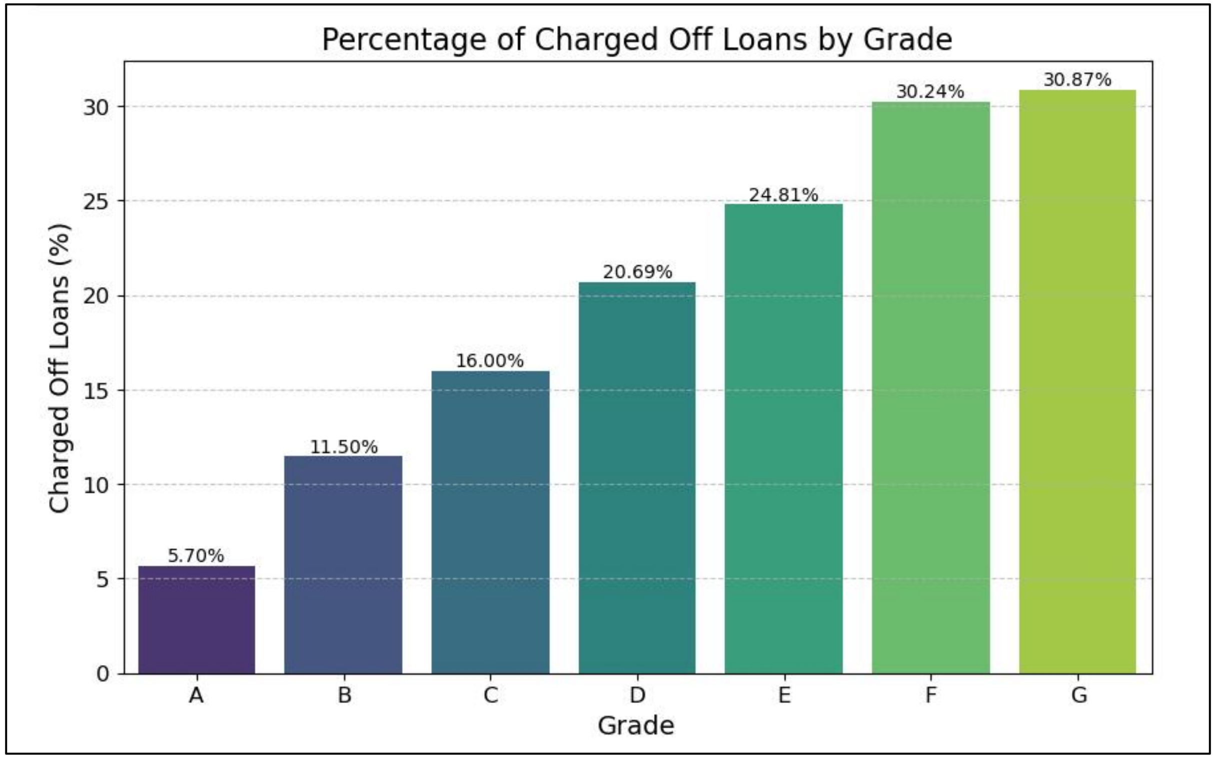
Bivariate Derived Variables: Insights



Analysis of Charged-Off Loans with Respect to 36 and 60 Months Terms

We can visualize that around 11 percent of customers with charged-off loans belong to the 36 months term category, while approximately 23 percent of customers with charged-off loans belong to the 60 months term category.

Bivariate Derived Variables: Insights



Percentage of Charged off loans by Grade
The analysis of the percentage of charged-off loans by grade provides valuable insights into the credit risk associated with different grades. Here are some key points you can include in your business report:

Grade G has the highest percentage of charged-off loans at 30.87%, indicating that borrowers in this category are more likely to default on their loans. This suggests that loans issued to Grade G borrowers should be carefully evaluated and monitored.

Grades E and F also have high percentages of charged-off loans, at 24.81% and 30.24% respectively. This indicates a significant risk associated with these borrower categories, requiring closer scrutiny and potentially higher interest rates or stricter approval criteria.

Lower grade loans (D, E, F, and G) show a clear trend of higher default rates, suggesting that these borrowers may require more stringent credit evaluation and risk management practices.

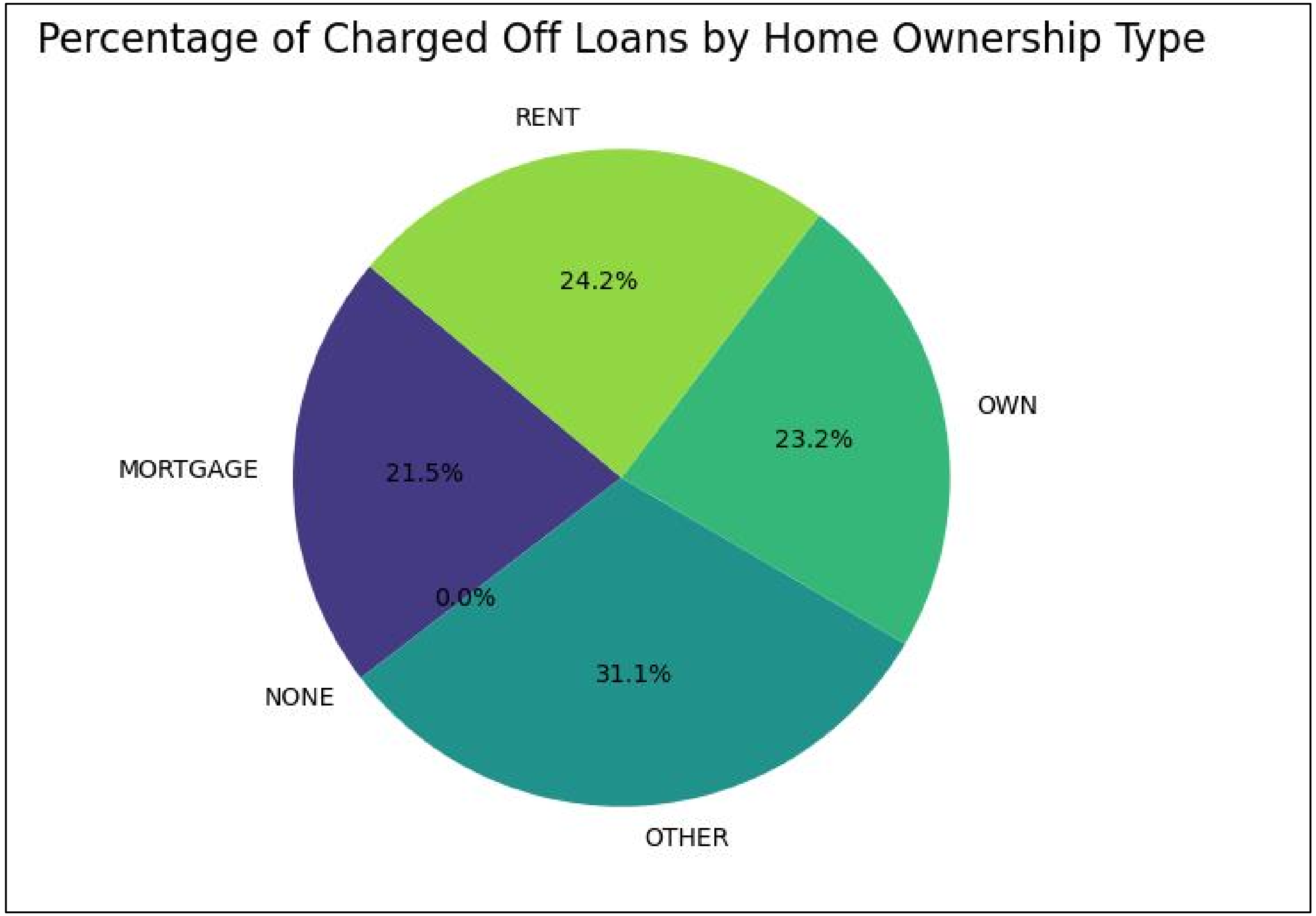
Higher grade loans (A and B) have comparatively lower default rates, at 5.7% and 11.5% respectively. This indicates that borrowers in these categories are more creditworthy and may be considered lower risk.

Grade C falls in the middle, with a default rate of 16%. This grade represents a moderate level of risk compared to the higher and lower grade categories.

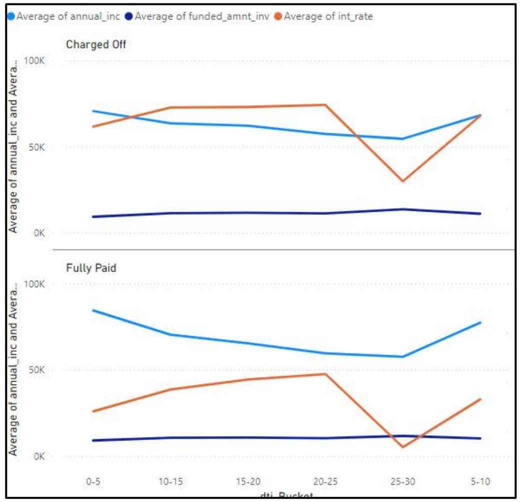
Overall, the data underscores the importance of proper risk assessment and loan pricing based on borrower grades. It highlights the need for lenders to carefully evaluate borrower profiles and implement effective risk management strategies to mitigate default risk.

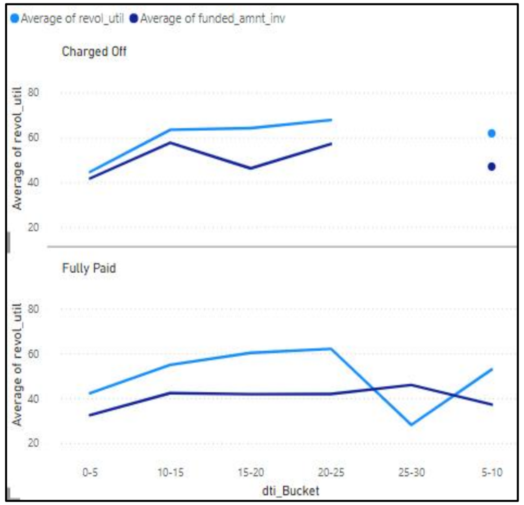
In conclusion, the analysis of charged-off loans by grade provides valuable insights that can help lenders make more informed decisions regarding loan approval, pricing, and risk management.

Bivariate Derived Variables: Insights



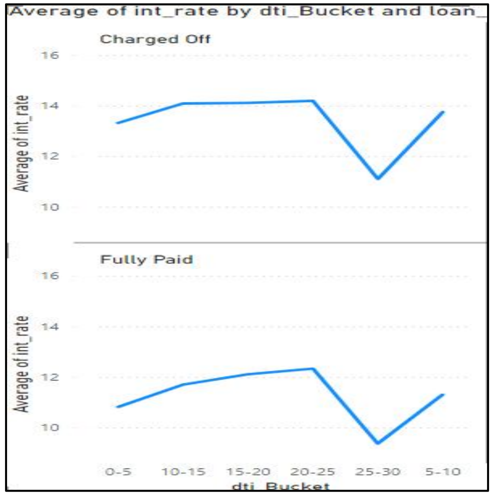
Bivariate Derived Variables: Insights

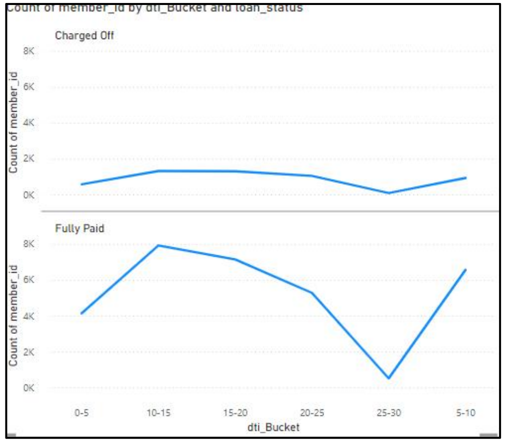




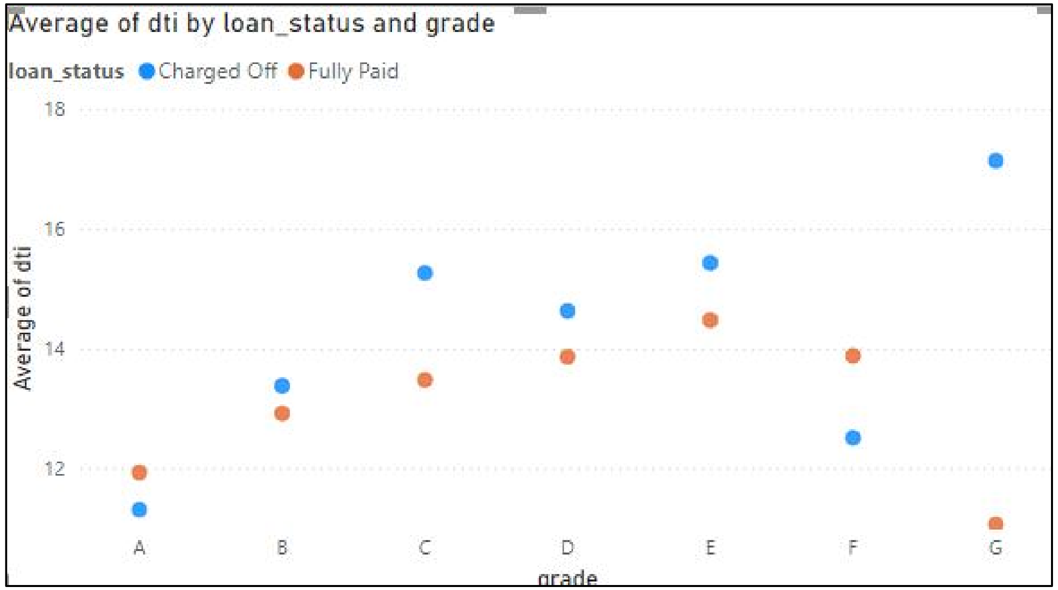
Income to Interest Ratio of Charged Off Accounts shows :

- DTI - with 25-30 are more likely to default if interest rate is above 10%
- DTI- 0-25 - are more likely to default if interest rate is above 12.5%
- DTI – with 16-25 are more likely to default if interest rate is above 12%
- Revol_utilization with Greater than 60% are more likely to default with DTI under 20.

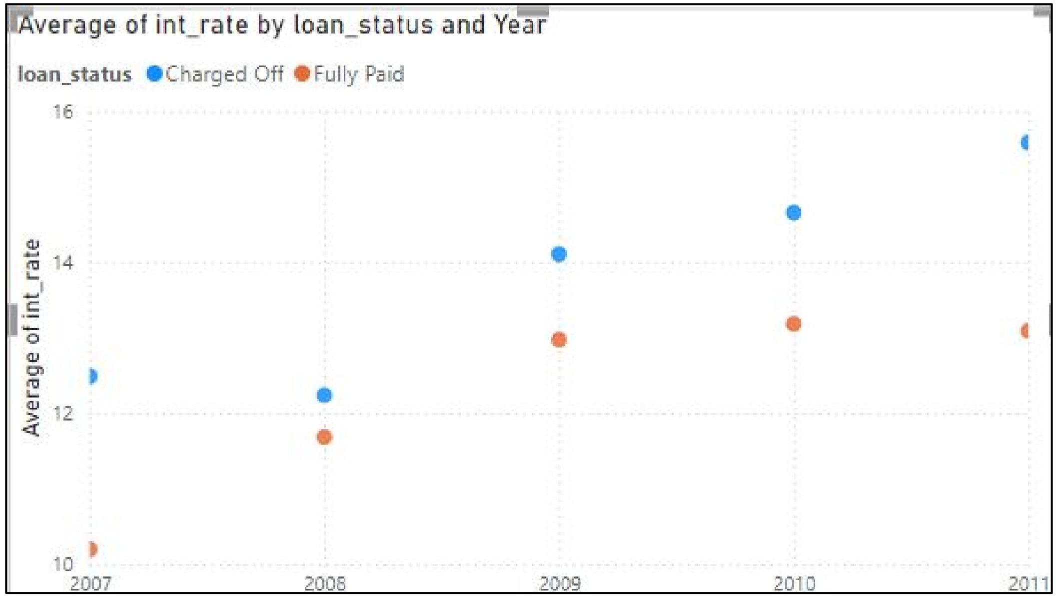




Bivariate Derived Variables: Insights

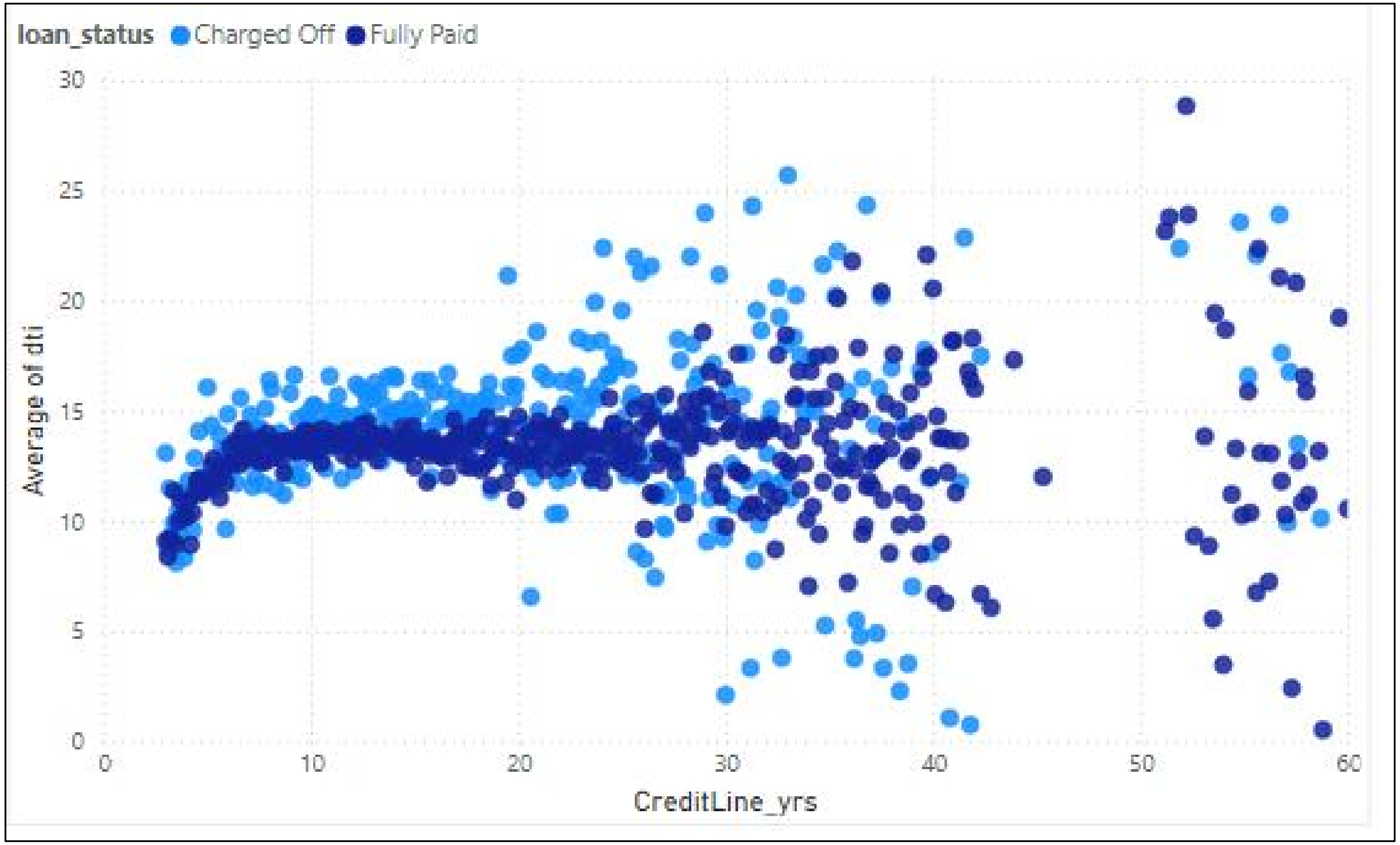


Graph on the left shows average_dti from all grade of Charged_off Accounts were higher than Fully-Paid then A and F grade as exception to trend.



Graph on the left shows the avg interest rate of Charged_off Accounts were higher than Fully-Paid as trend.

Bivariate Derived Variables: Insights



A relation on dti > 14% is likely to default.
Credit Line more than 50 creates liklihood to default

Multivariate Variables: Insights

upGrad

loan_status	car	credit_card	debt_consolidation	educational	home_improvement	house	major_purchase	medical	moving	other	renewable_energy	small_business	vacation	wedding	Grand Total
Charged Off	1112900	6569350	35517100	391450	3826175	710450	1845525	864850	501525	4833900	155100	6380425	285350	944050	63938150
Fully Paid	8624425	49381300	183086175	1613225	27331525	3753925	14716225	4366450	3098500	24063275	663650	15592800	1604175	7792350	345688000
Grand Total	9737325	55950650	218603275	2004675	31157700	4464375	16561750	5231300	3600025	28897175	818750	21973225	1889525	8736400	409626150
Charge Off % sahre	13%	13%	19%	24%	14%	19%	13%	20%	16%	20%	23%	41%	18%	12%	18%

Charged off Ratio is higher in Small Business, educational, Debt-consolidation & medical.

Emp_length	AVG Funded Amount
10+ years	13504
3 years	10570
< 1 year	9343
2 years	9816
5 years	11303
4 years	10607
1 year	9160
6 years	11461
7 years	12102
8 years	11919
9 years	11803
Grand Total	11241

Average of prn_rcv_percent	Column Labels			
Row Labels	Charged Off	Current	Fully Paid	Grand Total
car	43%	92%	100%	94%
educational	41%		100%	91%
vacation	39%	91%	100%	91%
credit_card	39%	88%	100%	94%
moving	39%	93%	100%	91%
renewable_energy	38%	93%	100%	89%
home_improvement	37%	91%	100%	93%
major_purchase	36%	92%	100%	94%
other	36%	90%	100%	90%
debt_consolidation	35%	89%	100%	90%
wedding	35%	91%	100%	94%
house	34%	88%	100%	89%
medical	32%	90%	100%	90%
small_business	31%	90%	100%	82%
Grand Total	36%	90%	100%	91%

Of All accounts , highlighted are more prone to losses with writeoff

THANK YOU