

Landing Club Case Study

Loan Portfolio Data Analysis

22nd Sep 2024



Case Study Objective

Business	Context	Problem statement
<p>Consumer Finance company deals in various types of Loan facilities to urban customers pertaining to different purposes.</p>	<p>At loan approval stage, business carries significant risks</p> <p>False Negative Applicants likely to repay loan but not approving loan.</p> <p>False Positive Applicants not likely to repay loan but approving loan.</p>	<p>Based on historical data trend, company is looking to improve the decision making process leveraging driver variables to mitigate risks resulting in either business loss or financial loss to the institution.</p>

Data Understanding, Cleaning & Manipulation

Source Loan Dataset
(2007 - 2011)

Features : 111
Records: 39717

Inference : Loan Status

Data Wrangling Considerations

Feature Level

- There are quite a few columns with Null values. Dropped all columns with missing 2/3rd of data.
- Remove single-valued columns.
- Removed dependent variables using correlation metric.
- Columns formatting with data type correction (Interest Rate, Revolving utilization rate)
- Handle Outliers (Annual Income)
- Identified significant columns based of business knowledge
- Derived new columns (Loan_issue_month, Loan_issue_year, installment_to_monthly_income_ratio)

Records Level

- There are none rows with all Null values.
- There are no duplicate rows.
- Segmented dataset for closed loans (load status includes 'Fully Paid' & 'Charged Off')

Final Shortlisted Features with Assumptions

Feature Name	Type	Description
addr_state	Consumer	The state provided by the borrower in the loan application
annual_inc	Consumer	The self-reported annual income provided by the borrower during registration.
delinq_2yrs	Consumer	The number of 30+ days past-due incidences of delinquency in the past 2 years
dti	Consumer	A ratio of total monthly debt payments divided by the borrower’s self-reported monthly income.
emp_length	Consumer	Employment length in years. Possible values are between 0 and 10 or more years.
home_ownership	Consumer	The home ownership status (RENT, OWN, MORTGAGE, OTHER)
purpose	Consumer	A category provided by the borrower for the loan request.
verification_status	Consumer	Indicates if income was verified by LC, not verified, or if the income source was verified
issue_d	Loan	The month which the loan was funded
grade	Loan	LC assigned loan grade
installment	Loan	The monthly payment owed by the borrower if the loan originates.
int_rate	Loan	Interest Rate on the loan
loan_amnt	Loan	The listed amount of the loan applied for by the borrower.
revol_util	Loan	Revolving line utilization rate.
sub_grade	Loan	LC assigned loan subgrade

Final Shortlisted Features with Assumptions

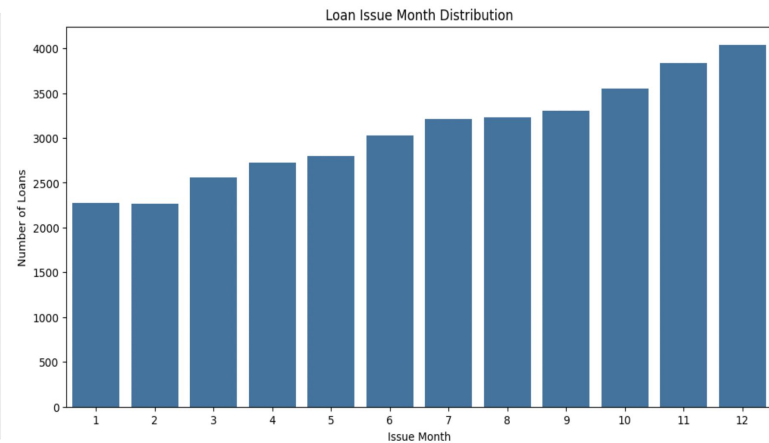
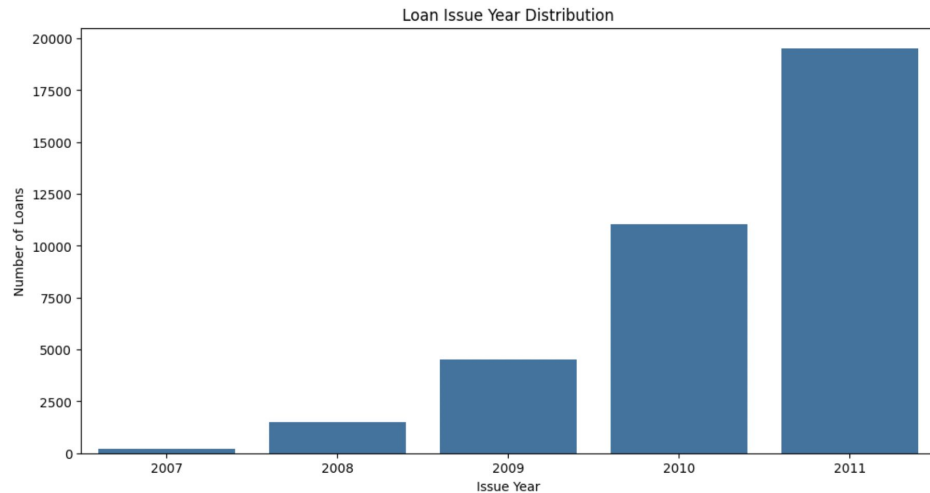
Feature Name	Type	Description
term	Loan	The number of payments on the loan. Values are in months and can be either 36 or 60.
Loan_Issue_Year	Derived	Loan Issue Year derived from issue_d
Loan_Issue_Month	Derived	Loan Issue Month derived from issue_d
install_to_income	Derived	Ratio of Loan Monthly Installment and Income of Applicant
loan_status	Target	Current status of the loan

Assumption

We have made an assumption that grade and subgrade was available at the time of loan evaluation.

Data Analysis

Univariate Analysis (Loan Issue Month & Year)



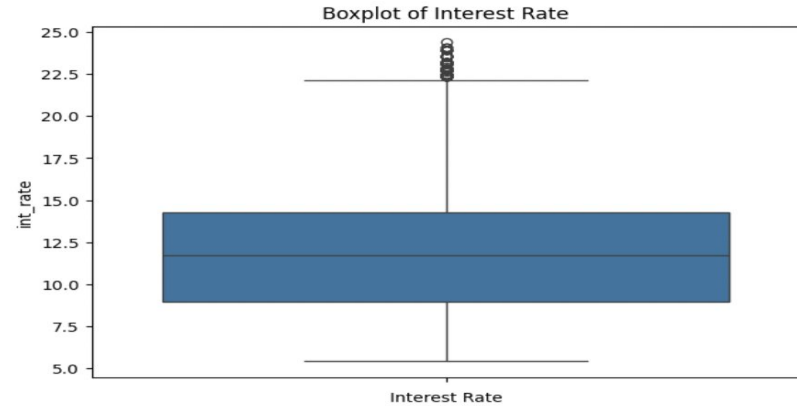
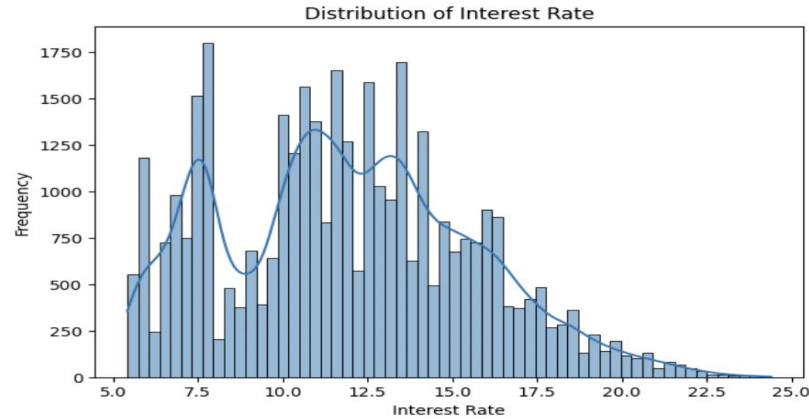
Observations

- Based on derived variable **Loan Issue Year vs Loan count**, it is clear that loan lending business have increased multifold from 2007 till 2011.
- This inorganic growth may be due to **US Great Recession** and hence lending business during this time period [Reference link](#)
- Based on derived variable **Loan Issue Month vs Loan count**, there is also **seasonality** in lending business. It is high during Year end for festive period (Christmas, New Year).

Inference

- Lending business grew multifold during US recession from 2007 till 2011.
- It is seasonal business with spikes during holiday and Festive season.

Univariate Analysis (Interest Rate)



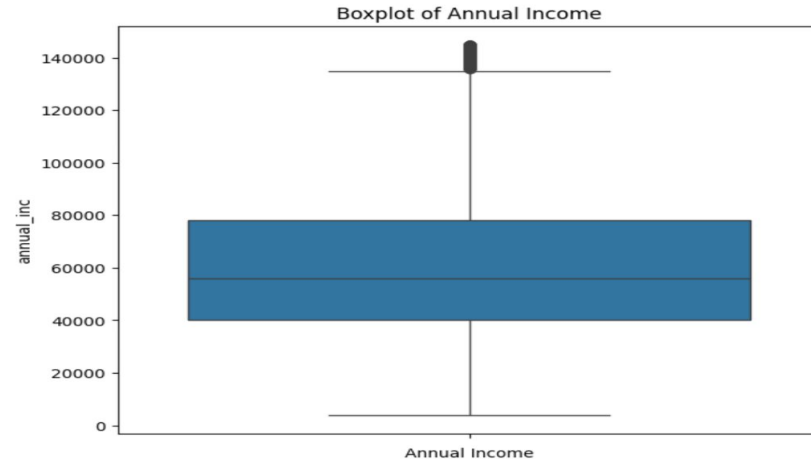
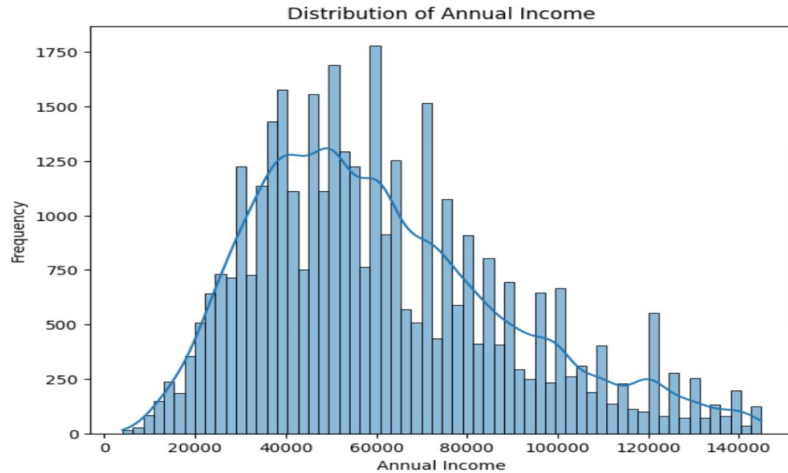
Observations

- Based on **Interest Rates vs Frequency**, it is clear that interest rates have **IQR 8.9% - 14.2 %** with **mean 11.8% & median 11.7%**
- Countplot shows right skewed distribution and same is visible from outliers in boxplot

Inference

- Though Interest rates are mostly focussed b/w **11% - 12 % range** with some outliers making it **right skewed** distribution.

Univariate Analysis (Annual Income)



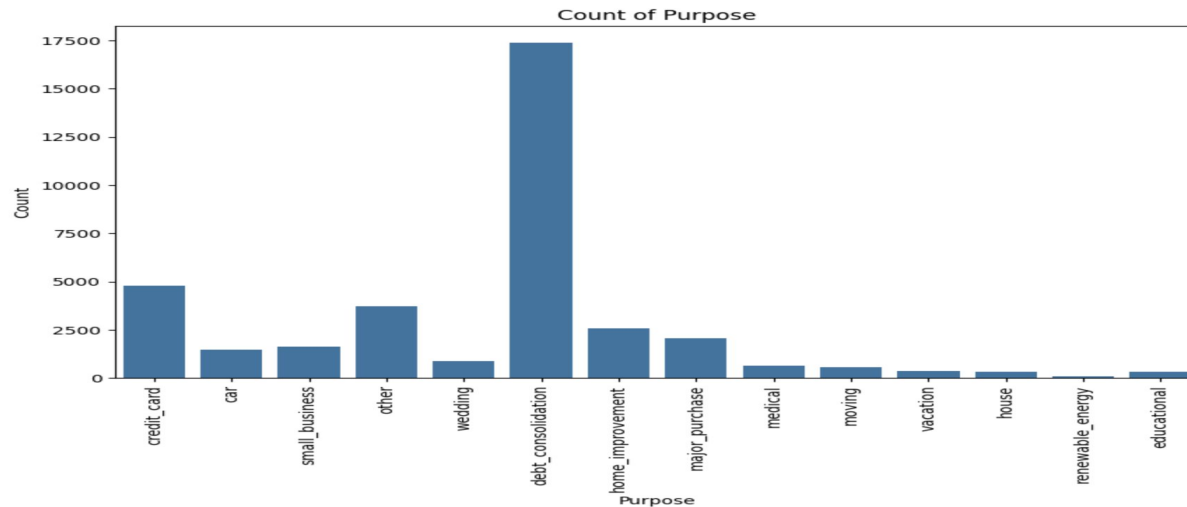
Observations

- The distribution is right-skewed with peak of the distribution is around \$60,000.
- The IQR, represented by the box, is relatively narrow, suggesting that the majority of incomes are clustered around the median.

Inference

- The right-skewed distribution and the presence of outliers suggest a significant income inequality in the dataset.
- A small number of individuals earn significantly higher incomes than the majority.
- The majority of individuals have incomes clustered around \$60,000, indicating a concentration of incomes at this level.

Univariate Analysis (Loan Purpose)



Observations

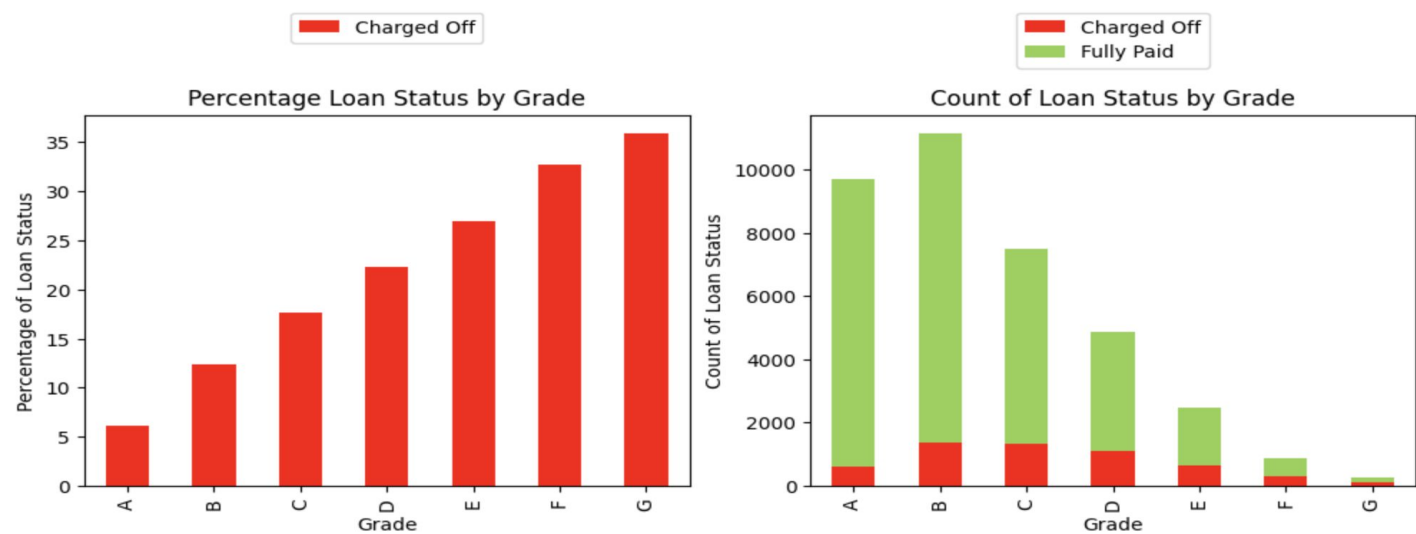
- The most common loan purpose is Debt Consolidation, followed by Credit Card.
- Car, Small Business, and Wedding also have relatively high counts.

Inference

- Large Number of debt consolidation loans indicates that many individuals are struggling to manage their overall existing debt.
- Consumer Spending: The popularity of credit cards and car loans suggests that consumer spending is a significant driver of loan demand.

Bivariate Analysis (Loan Status vs Grade)

14



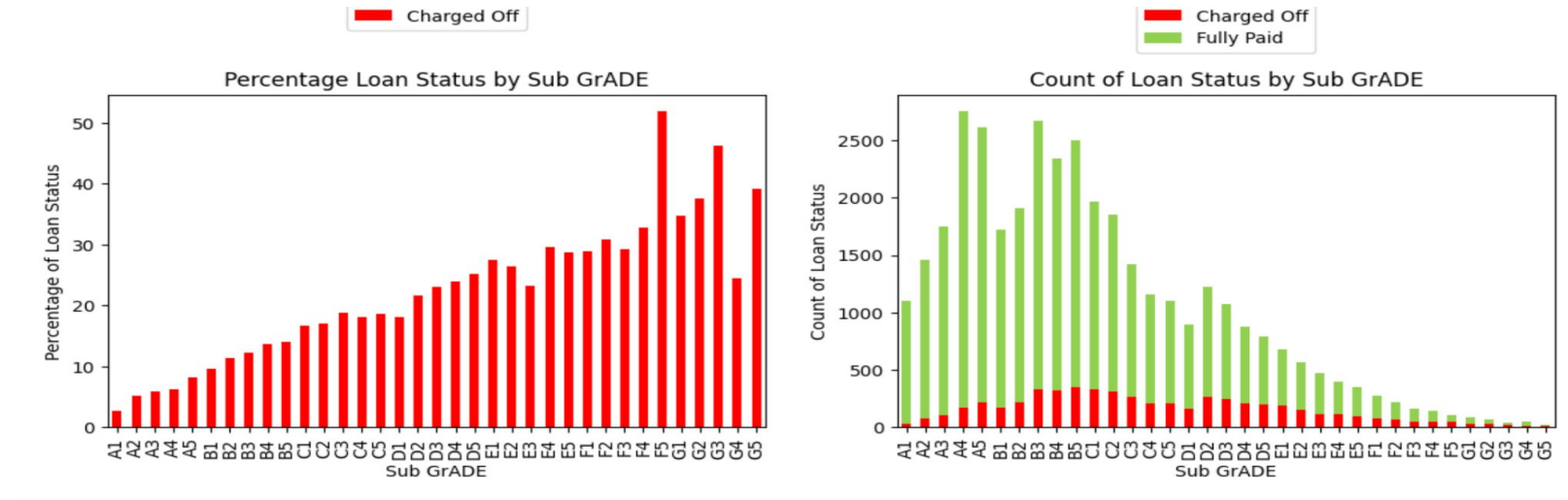
Observations

- There's a clear upward trend in default rates as we progress from Grade A to Grade G.
- This increase in default rates is linear
- While Grades F and G have fewer data points, the other grades have sufficient sample sizes.

Inference

- The loan grade is a strong driving factor of defaulted Loan.

Bivariate Analysis (Loan Status vs Sub-Grade)



Observations

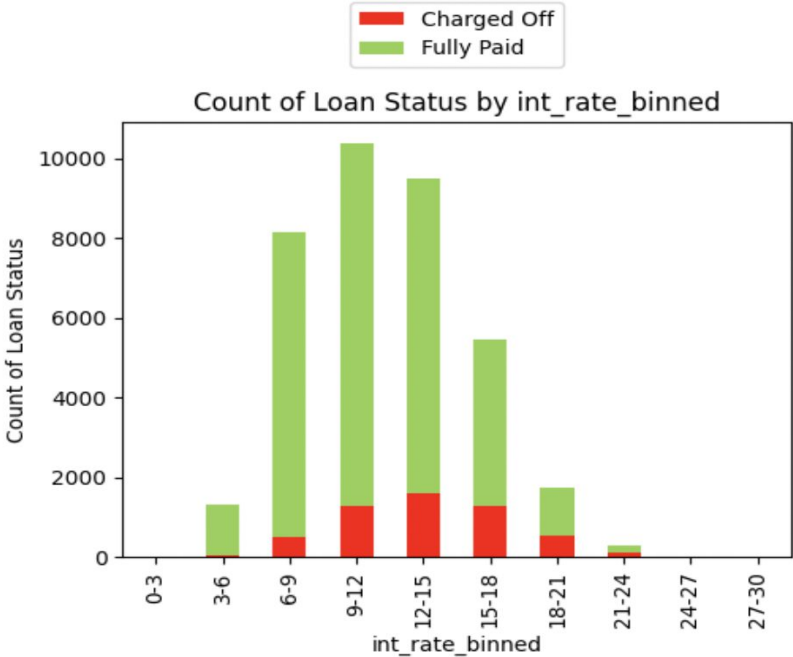
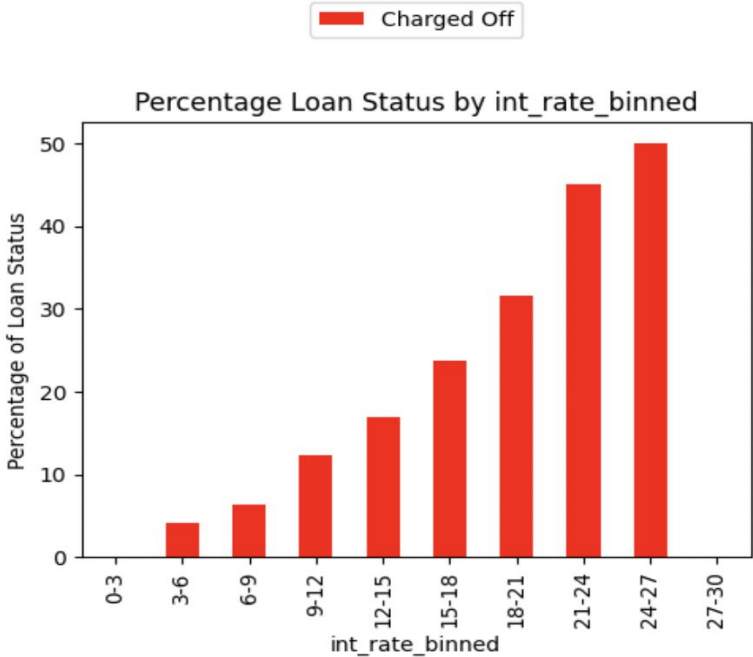
With higher **installment to monthly income** ratio, it is clear that loan defaults have been increasing.

- There's a clear upward trend in default rates as we progress from Grade A to Grade G and within the grades from sub grade 1 to 5.
- This increase in default rates is roughly linear, with few exceptions around C3 to C4 and E2 to E3 and few spikes in F5a and G3
- While Grades F and G have fewer data points, the other grades have sufficient sample sizes.

Inference

The loan Sub-grade is a strong driving factor of defaulted Loan.

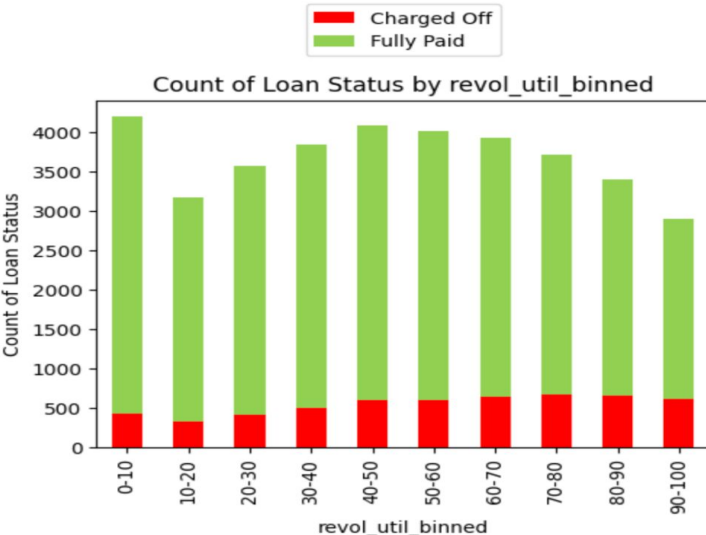
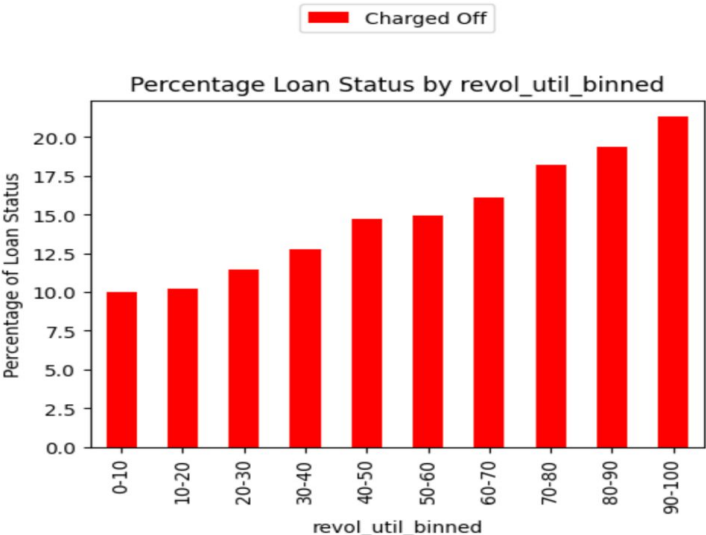
Bivariate Analysis (Loan Status vs Interest Rate)



Inference

There is clear impact of interest rates on Loan Defaults. With higher interest rates, loan defaults have been increasing.

Bivariate Analysis (Loan Status vs Revolving Credit Utilization)



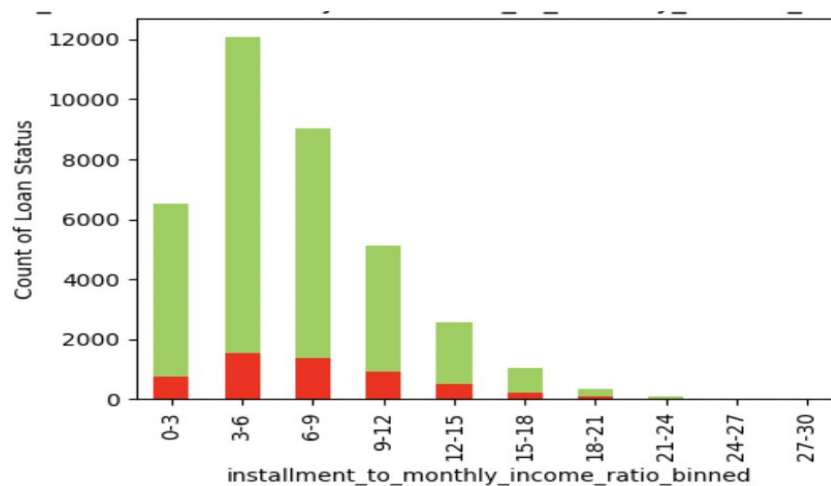
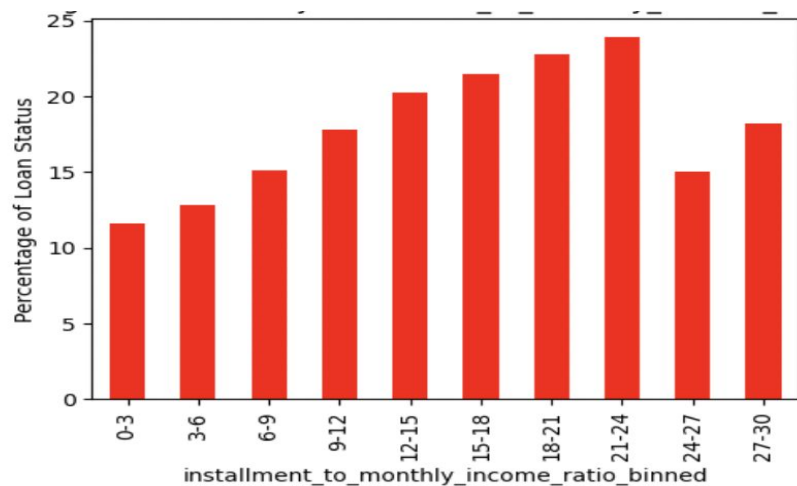
Observations

- The count of charged-off loans is relatively high across all revolving utilization bins, indicating that a significant number of borrowers, regardless of revolving utilization level, are at risk of defaulting on their loans.
- The percentage of charged-off loans generally increases as the revolving utilization binned increases. This suggests that borrowers with higher revolving utilization are more likely to default on their loans.

Inference

Strong relationship relationship exists between revolving utilization and loan status.

Bivariate Analysis (Loan Status vs Installment To Monthly Income)



Observations

With higher **installment to monthly income** ratio, it is clear that loan defaults have been increasing.

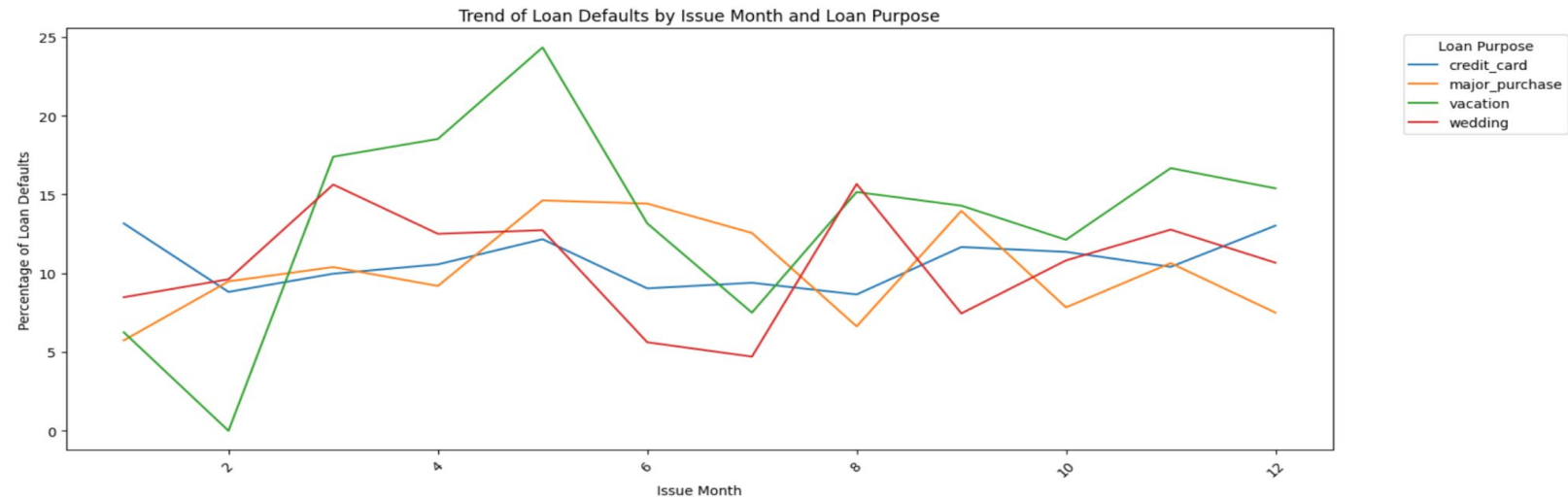
It must be due to the fact applicant is not able to pay back due to constrained monthly income as compared to monthly installment value.

There is some decrease towards the extreme right bins, should be due to fact that they are very less number of loans on that bin category.

Inference

It is better to keep **loans for higher terms** (60 months) to keep installment to monthly income ratio in control.

Multivariate Analysis(Loan Defaults by Issue Month and Loan Purpose)



Observations

Seasonality: There are noticeable seasonal patterns in loan defaults across different loan purposes:

- Credit Card: Defaults peak in November and December, likely due to holiday spending and increased consumer debt.
- Major Purchase: Defaults also show a peak in May and June, suggesting that larger purchases made during the holiday season may contribute to financial strain and increased default risk.
- Vacation: Defaults tend to be lower during the holiday season but may increase in the summer months (May-June and September) due to vacation expenses.

Loan Purpose Comparison: Across all issue months, credit card and major purchase loans consistently have higher default rates than vacation and wedding loans.

Inference

Loan defaults are influenced by both seasonal factors and loan purpose.

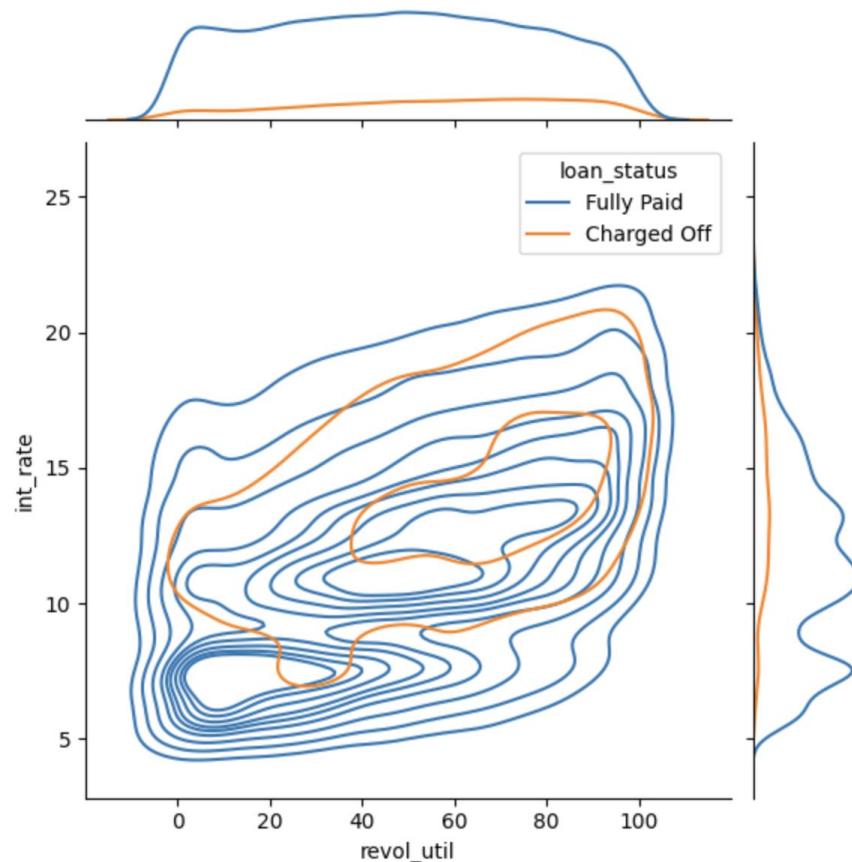
Multivariate Analysis(Loan Status vs Interest Rate vs Revolving Credit Utilization)

Observations

- The marginal distribution of revolving utilization is shown along the bottom axis. It appears to be right-skewed, with a larger number of borrowers having lower revolving utilization.
- The marginal distribution of interest rate is shown along the right axis. It is also right-skewed, with a larger number of borrowers having lower interest rates.

Inference

- Borrowers with higher revolving utilization might tend to have higher interest rates, especially among those who have been charged off.
- joint density plot suggests that there is a relationship between revolving utilization, interest rate, and loan status.



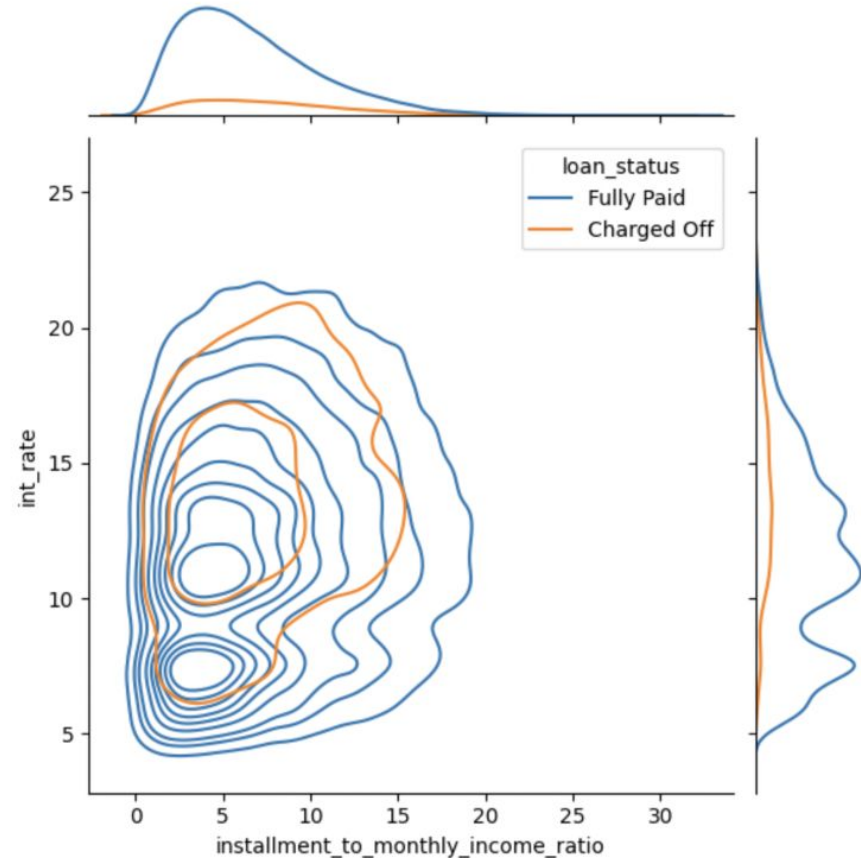
Multivariate Analysis(Loan Status vs Interest Rate vs Installment To Monthly Income)

Observations

- Graph shows Fully Paid loans are highly concentrated at Bottom Left quadrant.
- Charged Off loans are increasing towards top right quadrant.

Inference

- With Higher interest rates and Installment to monthly income ratio, probability to default increases.
- However, this trend is not very strong and may be influenced by other factors.



Multivariate Analysis (Correlation Metric & Heat Map)

Observations Strong Positive Correlations with loan_status_encoded:

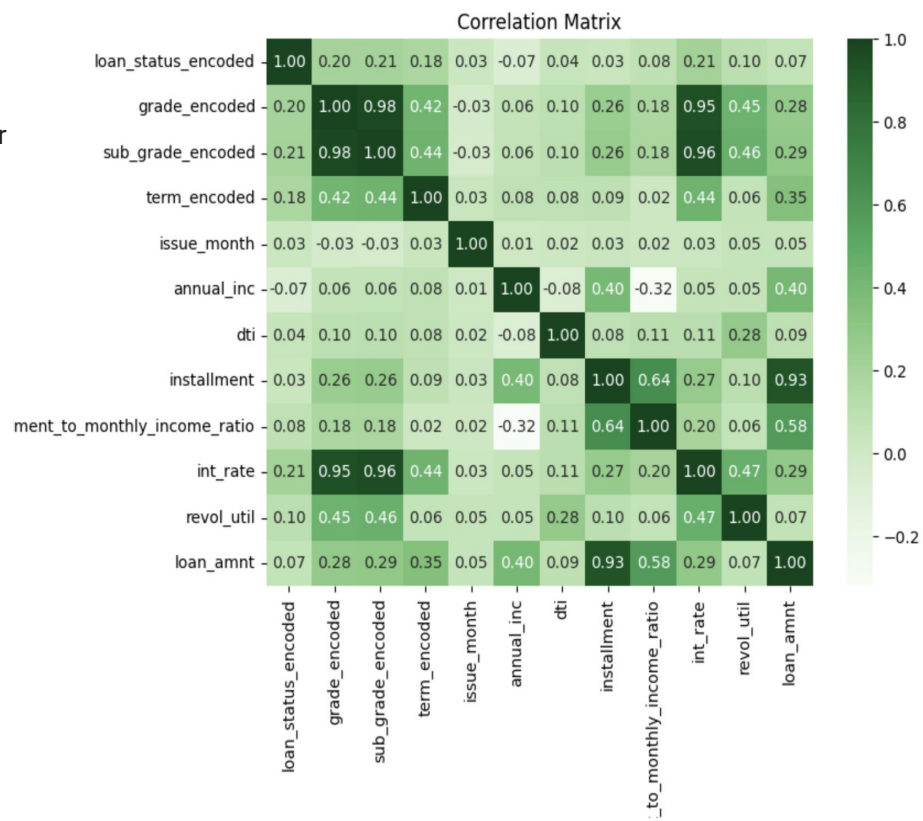
- Grade and Sub_grade have strong positive correlations with Loan Status. This suggests that borrowers with lower grades and sub-grades are more likely to default on their loans.
- Interest Rate has a strong negative correlation with Loan Status. This indicates that borrowers with higher interest rates are less likely to default on their loans.

Moderate Positive Correlations with loan_status_encoded:

- Revolving Credit Utilization has a moderate positive correlation with Loan Status. This suggests that borrowers with higher revolving utilization are more likely to default on their loans.
- Installment to Monthly Income Ratio has a moderate positive correlation with Loan Status. This indicates that borrowers with higher Installment to Monthly Income Ratio ratios are more likely to default on their loans.

Weak or No Correlations with loan_status_encoded:

- Term, Issue Month, Annual Income, Debt to Income Ratio, and Loan Amount have weak or no correlations with Loan Status.





Driving Factors And Recommendations








Driving Factors

Based on data analysis, below are the significant variables that are driving factors for Loan default.

The variables are listed in decreasing weightage of impact they have have. i.e. highest impact driving factor is listed first.

1. **Grade**
2. **Sub Grade**
3. **Interest Rate**
4. **Revolving Credit Utilization**
5. **Installment to Monthly Income Ratio**
6. **Term**
7. **Delinquency Count**
8. **Issue Month**
9. **Installment**
10. **Annual Income**

Recommendations

-  Increase focus on higher term loans (60 months).
-  With exponential business growth, default loan rate has increased significantly. Business shall strengthen and enhance loan risk assessment process with expansion goals.
-  Applicants who are delinquent in last 2 years carries more risk to default.
-  Business should continue doing a good job while determining loan grades/Sub grades.. Lower loan grades clearly trend to higher defaults rates.
-  Additional preference should be given to applicants eligible for lower interest rate with higher installment to monthly income ratio who are less likely to default.
-  Introduce more flexible loan repayment plans, especially around the holiday season to reduce defaults.
-  Introduce stringent controls for high-risk loan purposes such as debt consolidation.



Thank You