

Lending Club Case Study

ANUSHA R I

Problem Statement

Lending Club is a peer-to-peer lending platform that connects borrowers with individual and institutional investors. The company's mission is to transform the banking industry by offering lower interest rates on loans and better returns on investment.

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

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

Data Set

- ▶ The data set has 111 columns and 39717 rows.
- ▶ The data can be categorized broadly into 3 categories –
 - ▶ Customer behavior parameters – Employment length, Title, Home Ownership etc
 - ▶ Loan Characteristics – loan amount, funded amount, interest rate etc
 - ▶ Loan Repayment parameters - revolving balance, next payment date , delinq etc

Methodology



Data Cleansing – Steps done

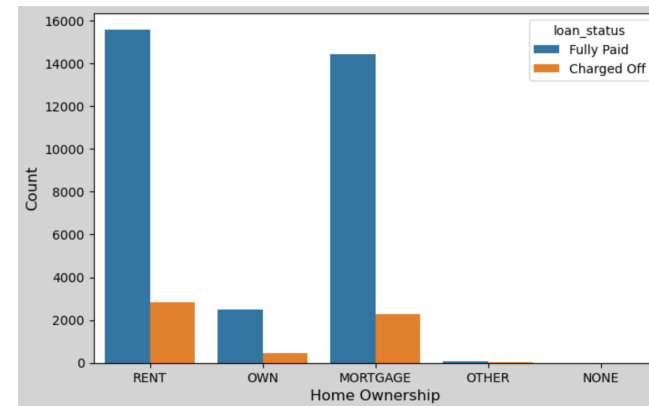
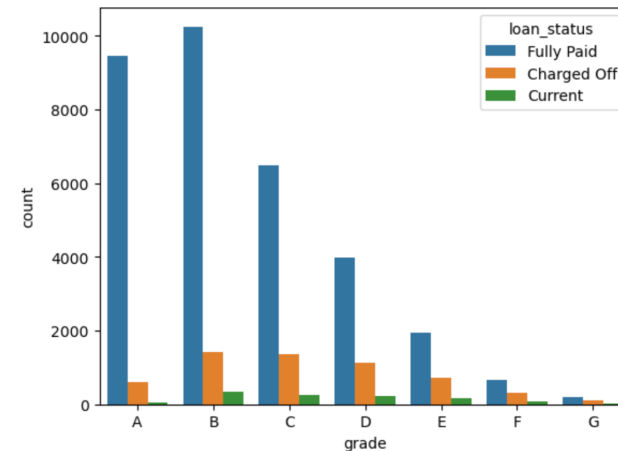
- ▶ Cleaning up columns with many null values
- ▶ Removing all insignificant columns
- ▶ Removing Rows with loan status as Current since we wanted to analyse the probability of a new customer repayment, we discarded all customers data with an existing loan in progress.
- ▶ Cleaning the columns values. For eg: removing % from int_rate, Years from employment length etc

Data Enrichment – Derived columns

- ▶ Added columns issue month and year from the loan issue date to get more insights with loans issued and the status

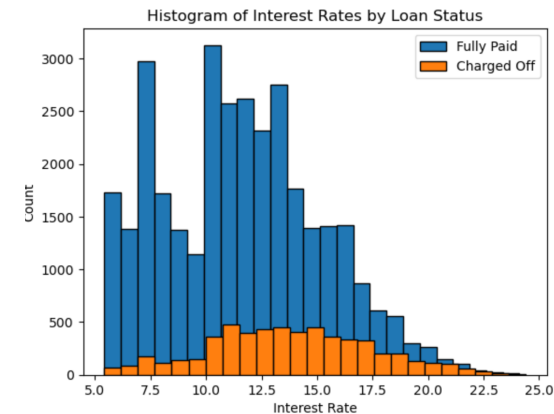
Univariate Analysis - Summary

- ▶ Customers who have taken the loan in Grade B are more likely to default.
- ▶ Customers who seem to have a rented house are more probable to default.

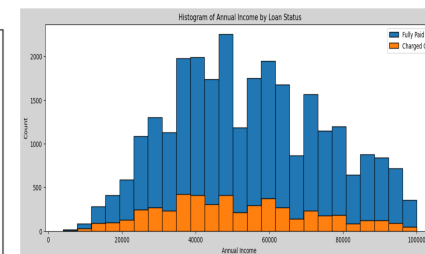
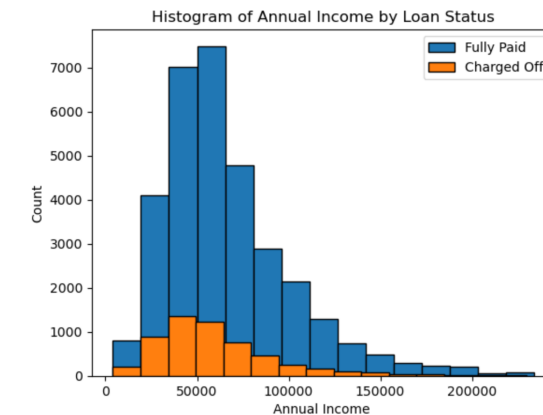


Univariate Analysis - Summary

- ▶ Customers whose loan is in the range of 10-17.5% are likely to default .

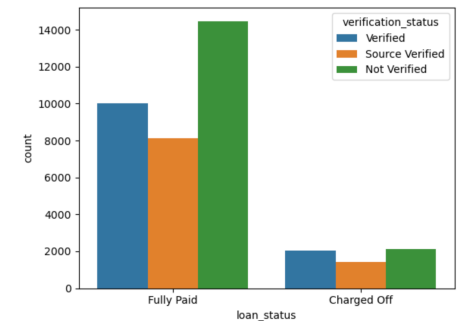
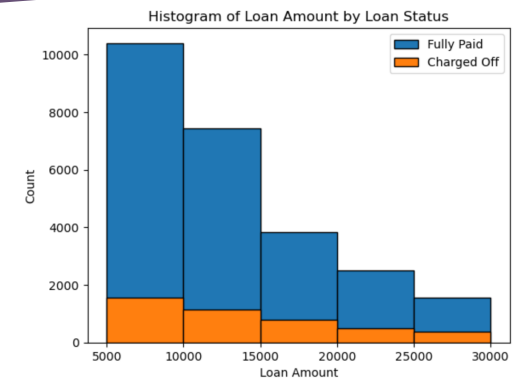


- ▶ Customers whose income range is between 35-60K have a higher chance of defaulting.



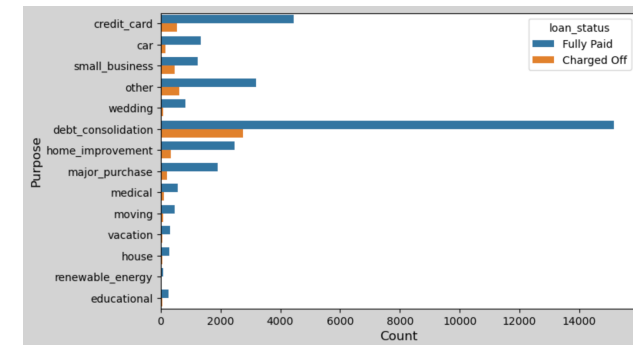
Univariate Analysis - Summary

- ▶ Customers with lower loan amount are charged off more, but also the total customers are high compared to higher loan amount.
- ▶ Customers have minimal difference in Verified and Non verified customers with respect to charge off. But source verified customers charge off seem to be lesser.

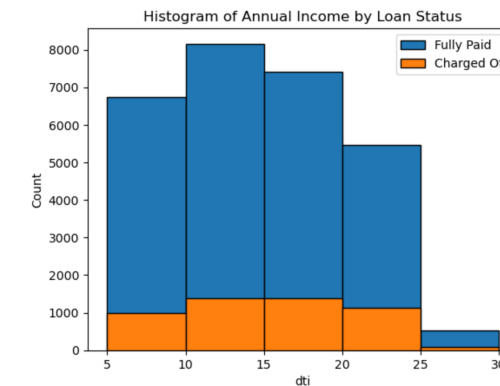


Univariate Analysis - Summary

- ▶ Customers who have taken loan for debt consolidation are high in number and so is the fully paid and charged off customers.

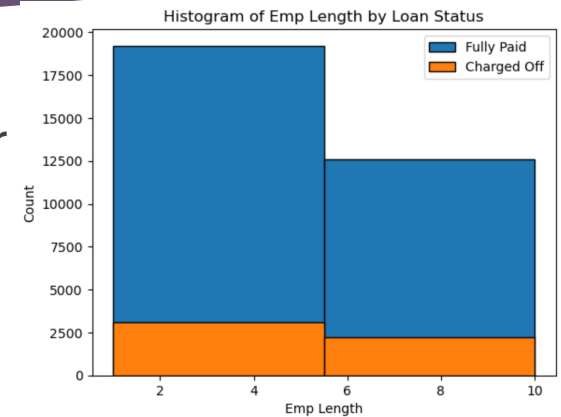


- ▶ Customers who have a dti between 10-20 default the most.

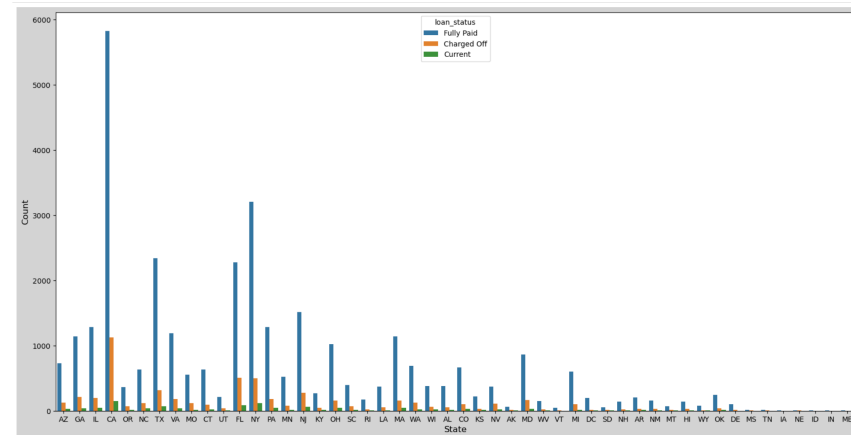


Univariate Analysis - Summary

- Customers with lower employee tenure seem to default compared to higher tenure.

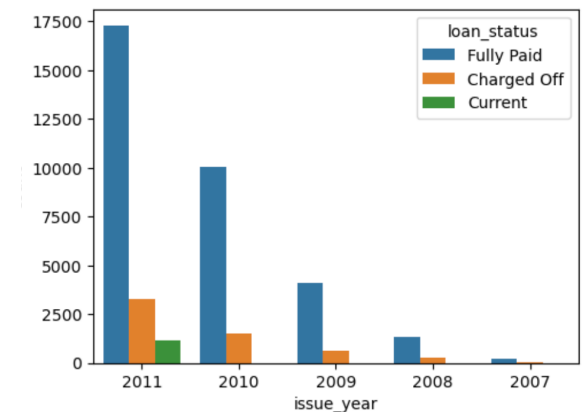
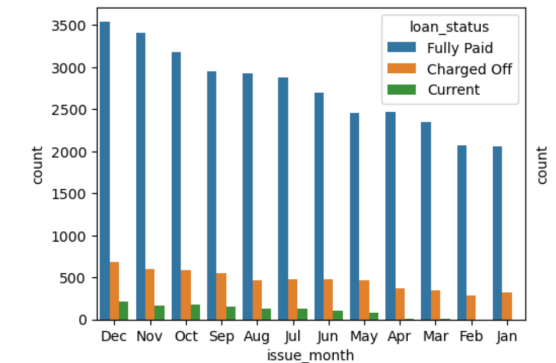


- Customer from CA tend to default more.



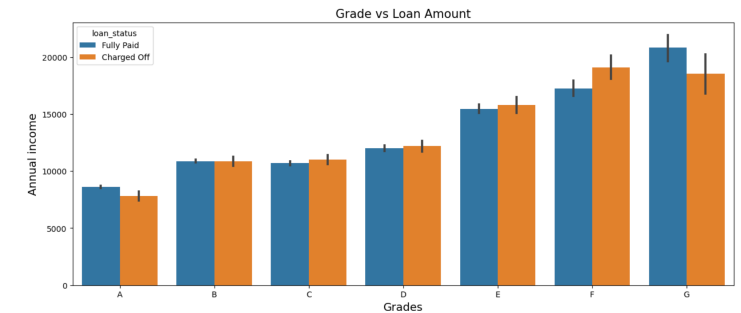
Univariate Analysis - Summary

- ▶ Customers who have taken loan in 2011 are more in number and have also charged off more.
- ▶ Customers have taken more loans in December seem to have defaulted more.

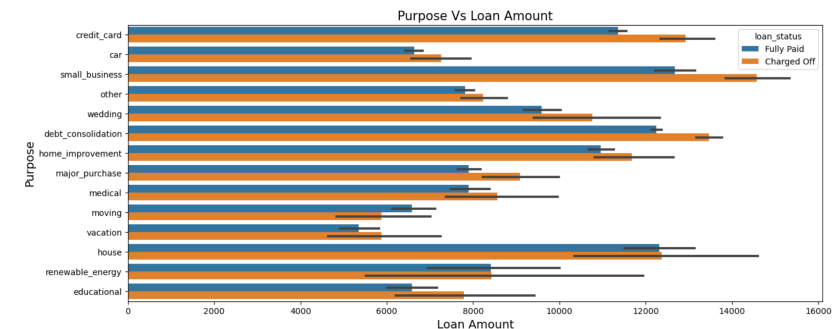


Bivariate Analysis - Summary

- ▶ Customers with loan grades F and have an annual income range 15K-20K default more.

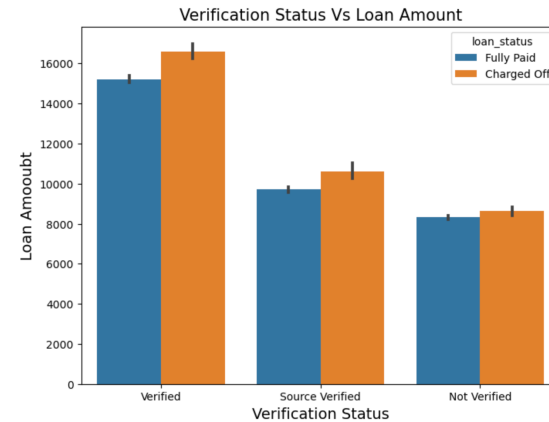


- ▶ Customers who have taken a loan for small business/ credit card /debt consolidation and the loan amount is in the range 12k-16k default more often.

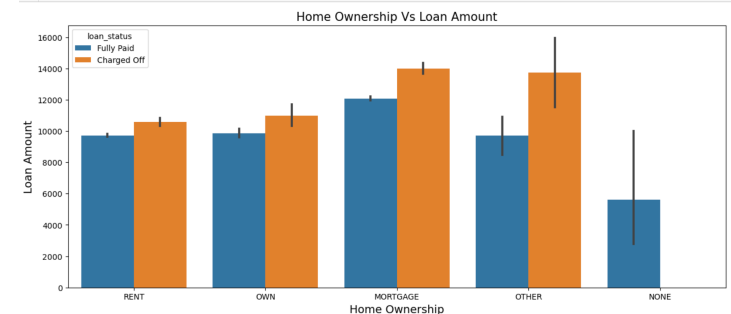


Bivariate Analysis - Summary

- ▶ Customer with Verified status and have higher loan amounts greater than 16K have higher chances to default.

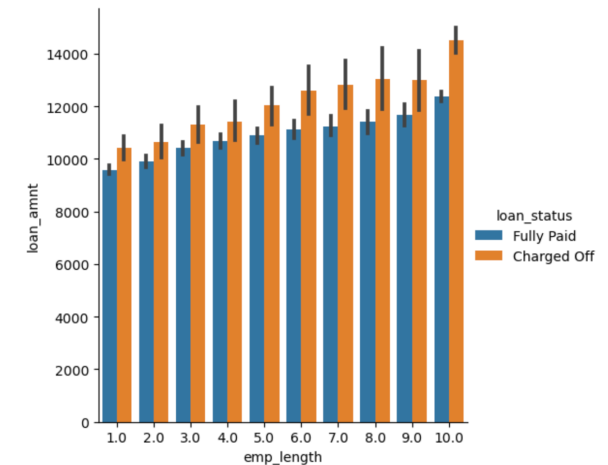
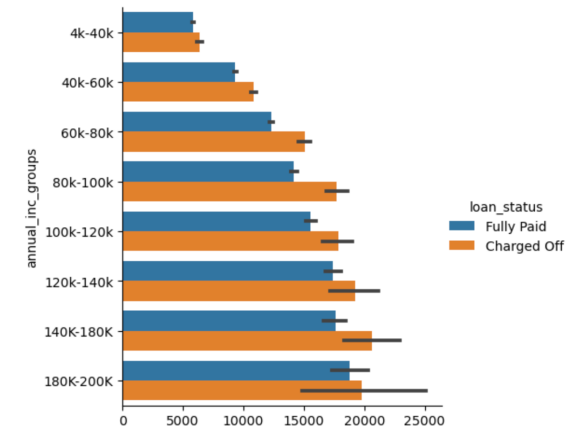


- ▶ Customers with a mortgaged home and has an income around 14K-16K default more.



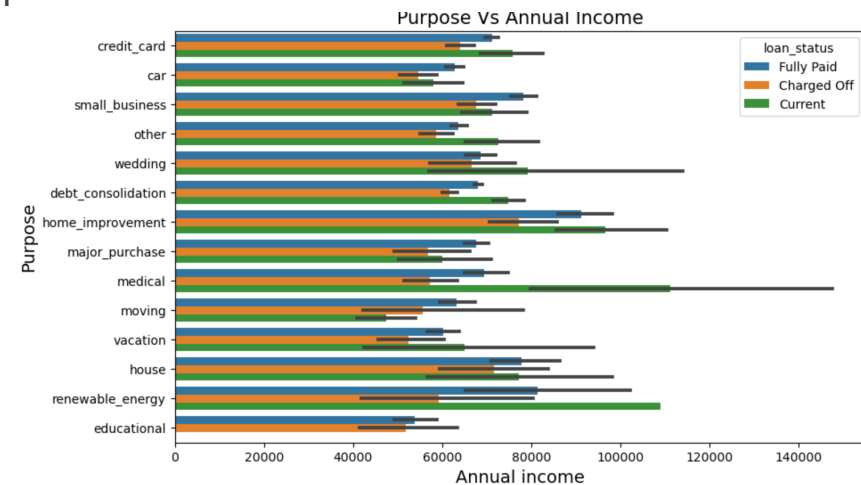
Bivariate Analysis - Summary

- ▶ Customers with higher income above 140K and higher loan amount default more.
- ▶ Customers with loan amount greater than 12K with employment tenure greater than 6 years seem to have defaulted more often.



Bivariate Analysis - Summary

- Customers taking loan for home improvement and have income of 60k -70k are more probable to default.





Thank You