Lending Club Case Study: Analysing Loan Defaults

Overview

The company lends loans to urban customers, for the loan provided to the customer there are two probable outcomes:

- The loan is paid up fully
- The loan is charged off (customer unable to pay loan)

To optimize the decision of approving loan in order to :

- Not approve loans to the customers who are likely to default

Like most other lending companies, lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss).
This case study presents the analysis of historic loan data to present the factors which could lead to higher loan defaults.

Data Analyzed

Loan data set for all loans issued through the time period 2007 to 2011.

Data Preparation

Data Cleansing

- Data has **39717 rows** and **111 columns**
- For the sake of having a clearer view we set a threshold of **75% non-null values to keep** it in scope of analysis
- On deleting columns with more than 75% null values, **55 columns remain** in the scope of analysis

- On taking a deeper look at the columns **8 more columns can be dropped**,
  - ['pymnt_plan','initial_list_status','collections_12_mths_ex_med','policy_code','application_type','acc_now_delinq','chargeoff_within_12_mths', 'tax_liens’]
  - Because all of these columns **has one single value in all rows**.

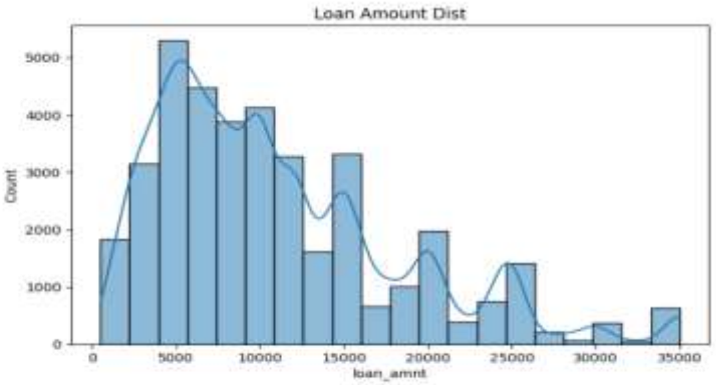
Null value/ Data Type processing also creating Derived Columns

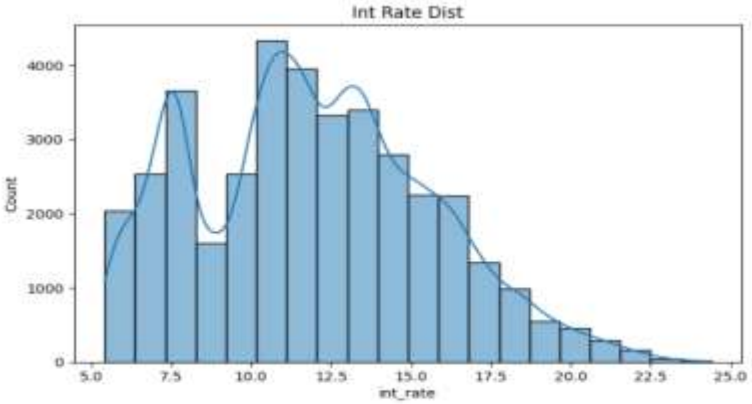
- Columns: term, int_rate, revol_util are object in data set and are be converted to numeric types int64 and float64
- emp_length_int (int64) is created from emp_length (object)
- earliest_cr_line is MON_YER (JAN-07) kind of date in data set, converted it to year
- pub_rec_bankruptcies null values are filled with 0

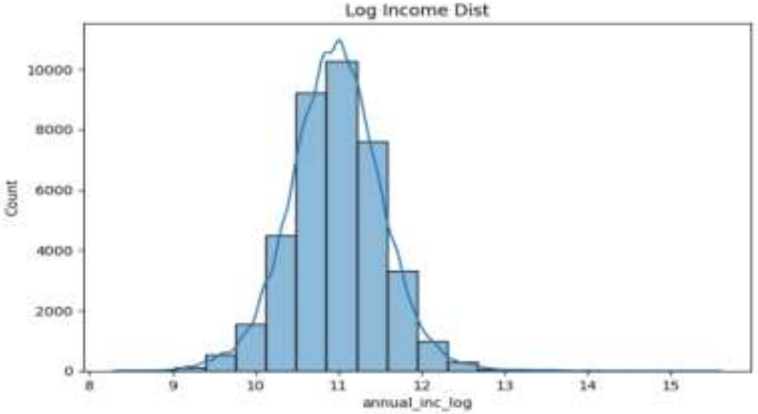
Univariate Analysis

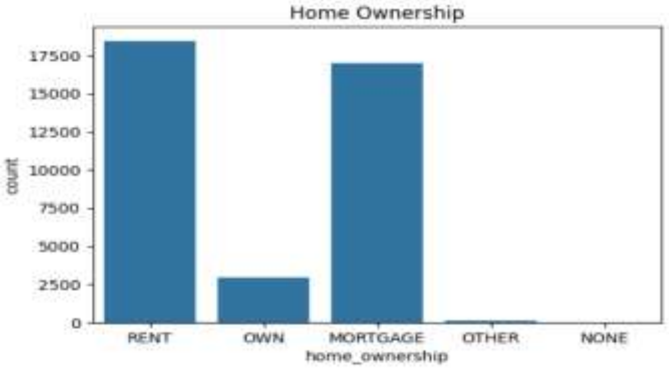
Looked at distributions of single variables.

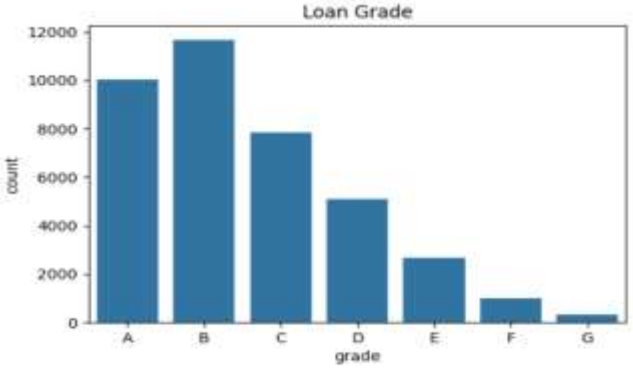
- Loan Amount Distribution (Histogram): Most loans are $5,000-$15,000, peaking at $10,000. Few above $30,000.
- Interest Rate Distribution (Histogram): Rates cluster 10-15%, some up to 25%. Shows varied risk pricing
- Log Annual Income Distribution (Histogram): Peaks around 11 (about $60,000), normal shape after log transform.
- Home Ownership (Count Plot): Renters highest (~18,000), then mortgage (~16,000), own (~3,000), others negligible.
- Loan Grade (Count Plot): B most common (~12,000), then A, C decreasing to G
- 26 columns are found to have outliers, summary of outliers is shown in python notebook







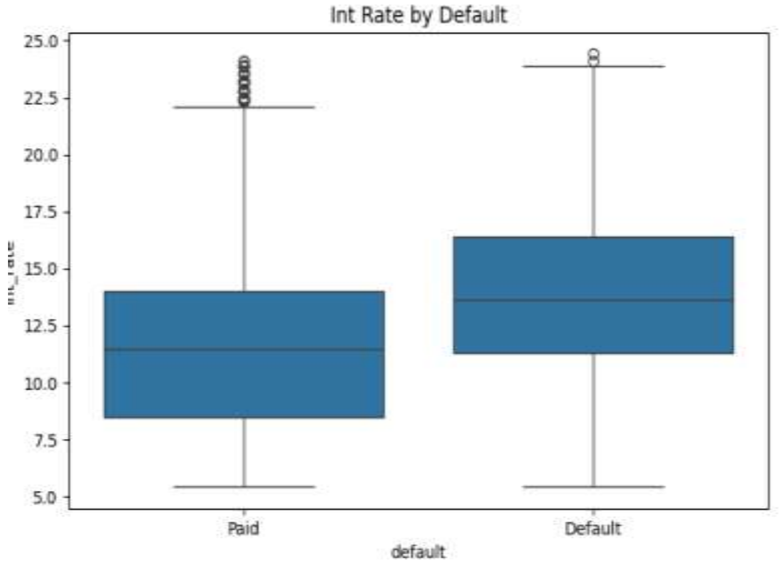


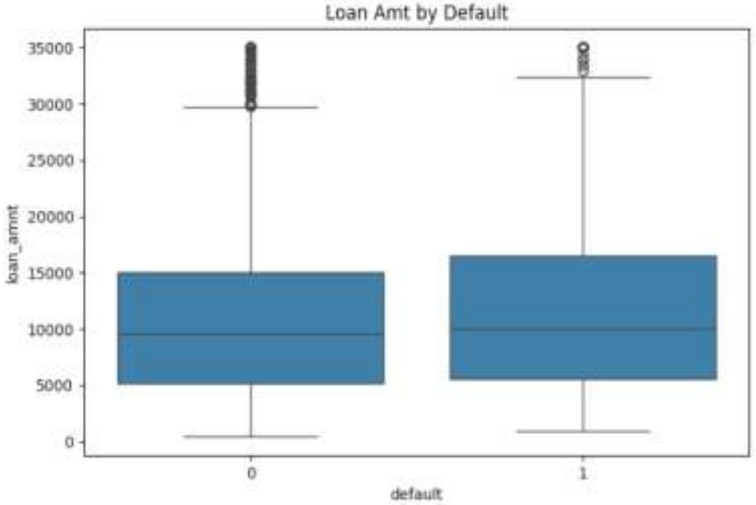


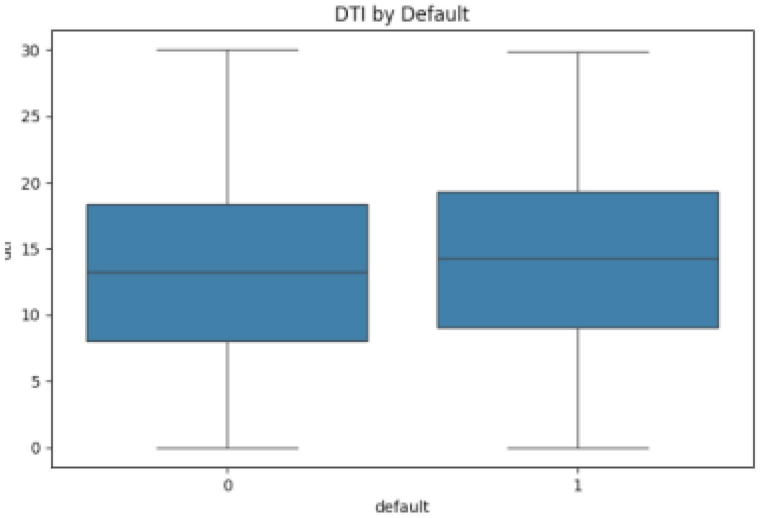
Segmented Univariate Analysis

Distributions split by groups like default status or grade.

- Loan Amount by Default (Box Plot): **Defaulters took larger loans (median ~$12,000)** than non-defaulters (~$10,000).
- DTI by Default (Box Plot): Defaulters show higher DTI (median ~17) vs ~15 for paid.
- Interest Rate by Default (Box Plot): Defaulters had higher rates (median ~15%) vs ~11% for paid.



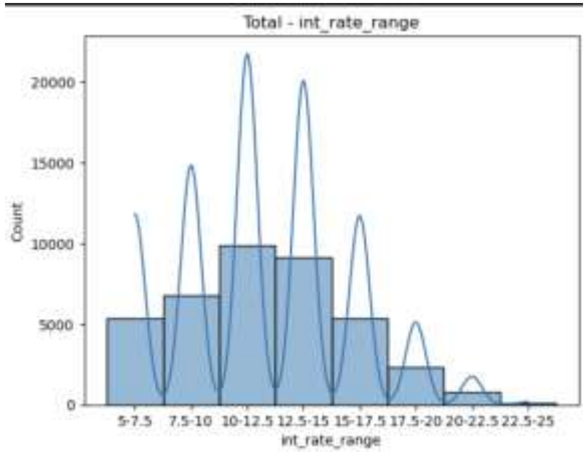


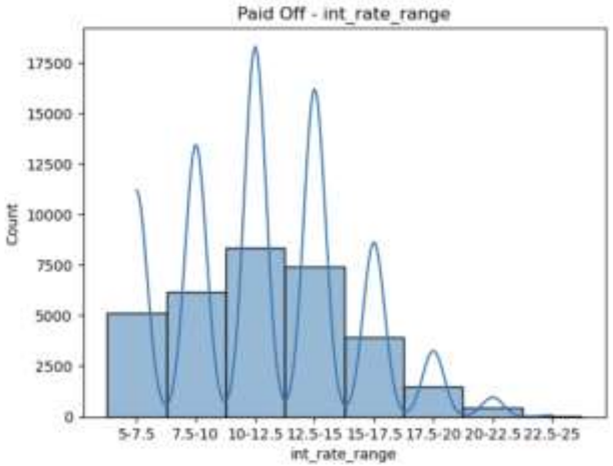


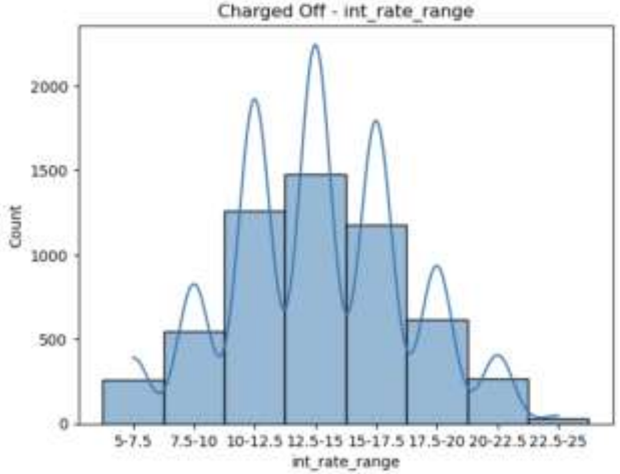
Bivariate Analysis

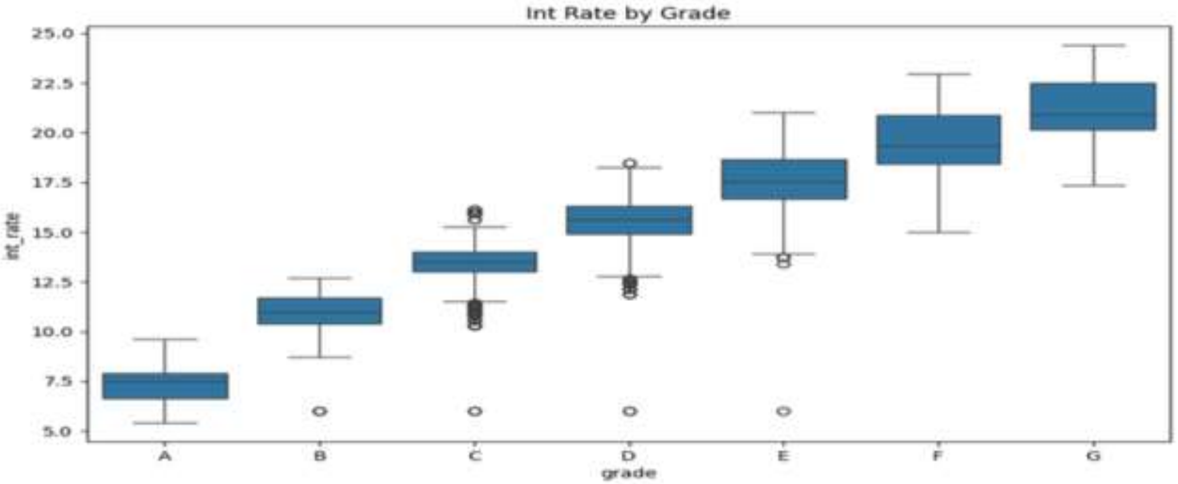
Quantitative Variables

- **'int_rate':**
  - Highest frequency of loans observed in interest range of 10-12.5 % interest rates
  - Highest loan defaults observed by interest rate **range 22.5-25(38%), 20-22.5(33%), 17.5-20 (26%), 15-17.5(22%)**
  - Loans from **15-25 % interest rates are most defaulted**

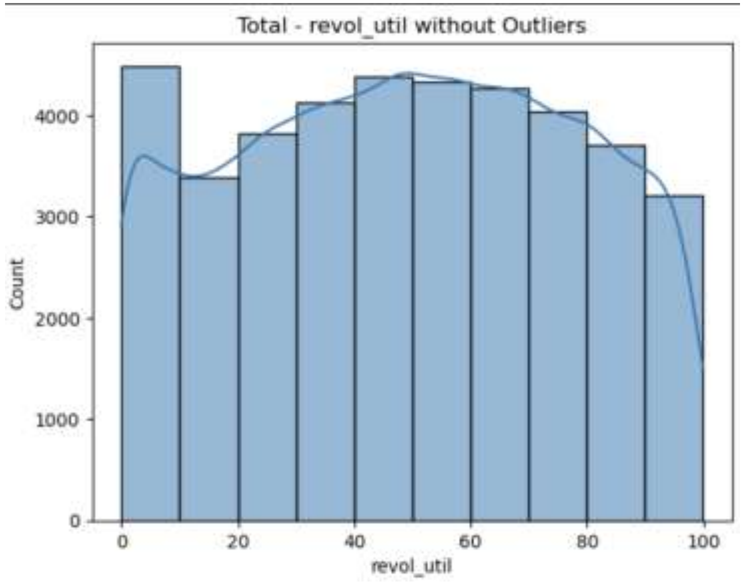


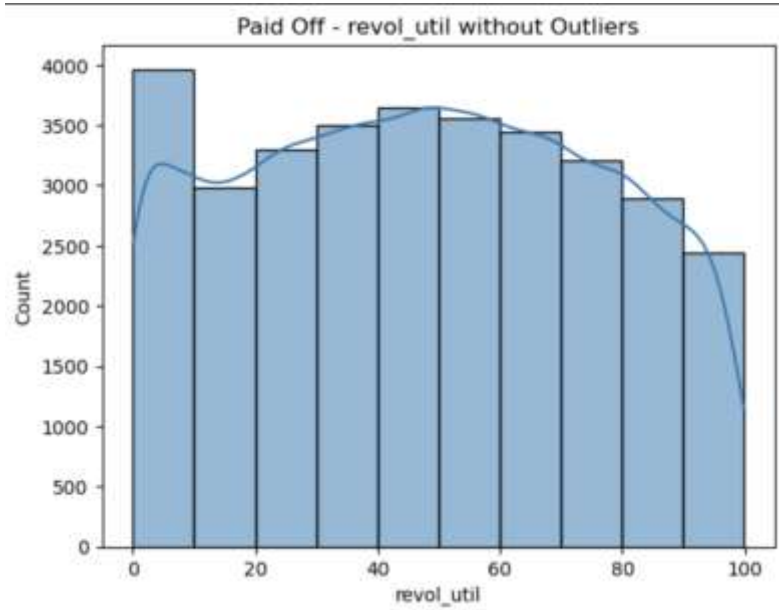


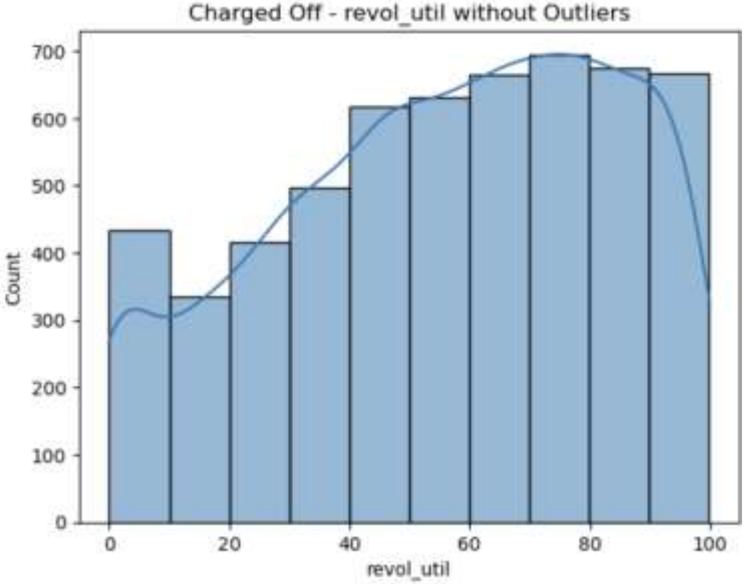




- **'revol_util':**
  - More charge offs(compared to general data pattern) is observed in **range of 50-95**
  - A high revolving balance on a loan means you are carrying over a large p debt from one month to the next on a revolving credit account, such as a credit card or a revolving line of credit. Instead of paying your balance in full, you are only making minimum payments
  - revol_util = revol_bal/total_credit







Categorical Variables

- **'term'**
  - 60 Month term loan is more charged off
  - General trend is people opt for 36 month loan more hence 36 month loans are more in number as well as more charged off
  - General trend is people opt for 36 month loan more hence 36 month loans are more in number as well as more charged off
  - But Although 60 month loans are opted less but whoever opts it has more chances of defaulting
  - in terms of approx ratio **36M - 3000/30000 (10%), 60M - 2500/10000 (25%)**
- **'grade'**
  - General trend of loans given is B > A > C > D > E > F > G
  - Charge off % by grade **G(32%) > F(30%) > E(25%) > D (21%) > C (16.6%) > B (11.8%) > A (6%)**
  - 80% grade G loans (worst performing) are of 60 months
  - mean revol_util for grade G loans is 71.4

- **'home_ownership'**
  - Most charge offs are done by **'OTHER' category with 18.4%** charge off
  - OTHER is a confusing category, there is also a NONE, this **needs to be clarified if it is conclusive enough.**
- **'purpose'**
  - Most charge offs done by loans with purpose**: small_business with 25% charge offs**

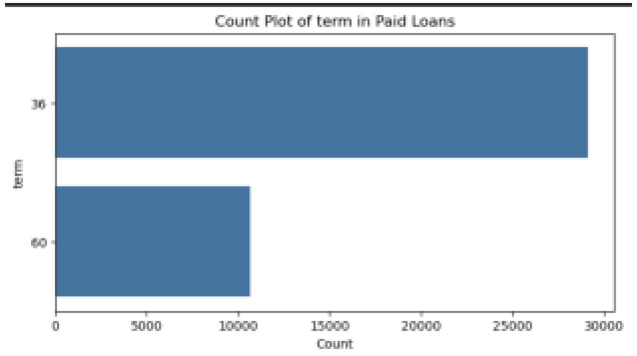
- **'addr_state'**
  - Most charege offs by **state - NV at 22%** deviation from general pattern fo 12-17 percent

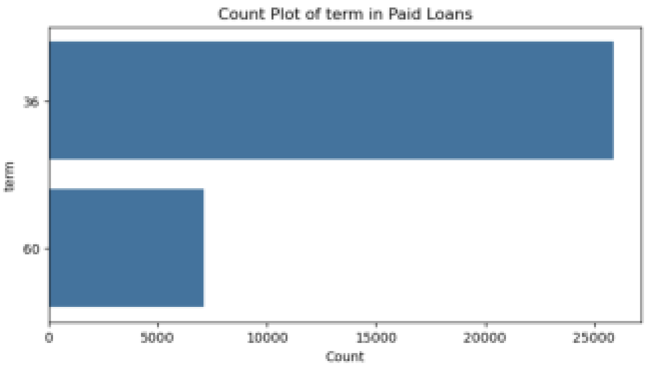
- **'earliest_cr_line'**
  - Most defaulting customer started their credit relations **in 2007 (20%), 2006(19%), 1973 (18%)**

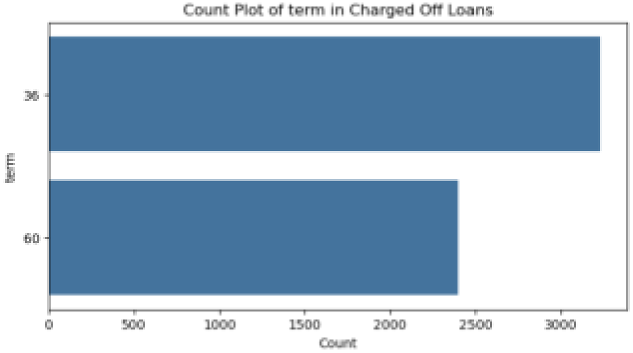
Graphs

Comparison of loan terms for

- total loans vs paid loans vs charged off
- 60M loans are charged of more

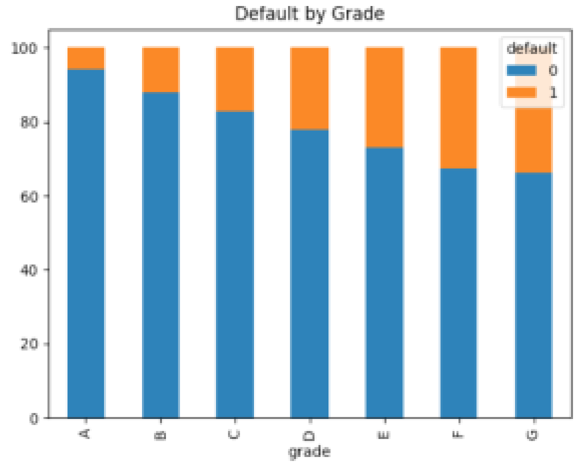




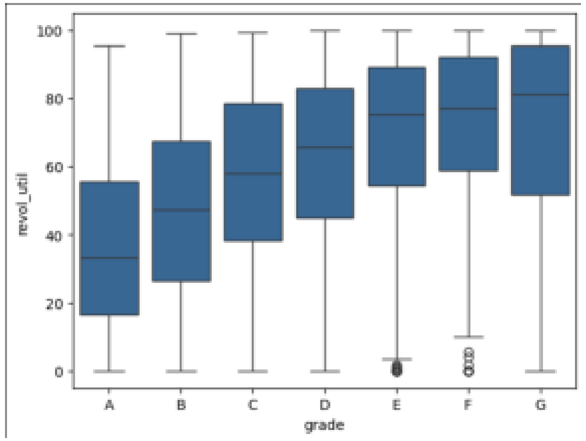


Grades

- Grades by most default
  - G>F>E>D>C>B>A

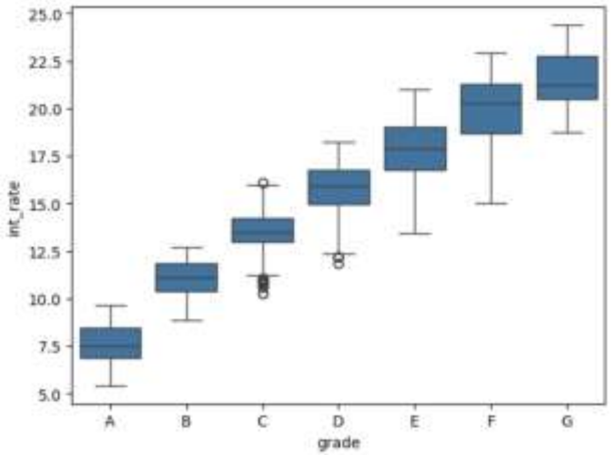


- Revolving utilization increases with grade A<B<C<D<E<F<G



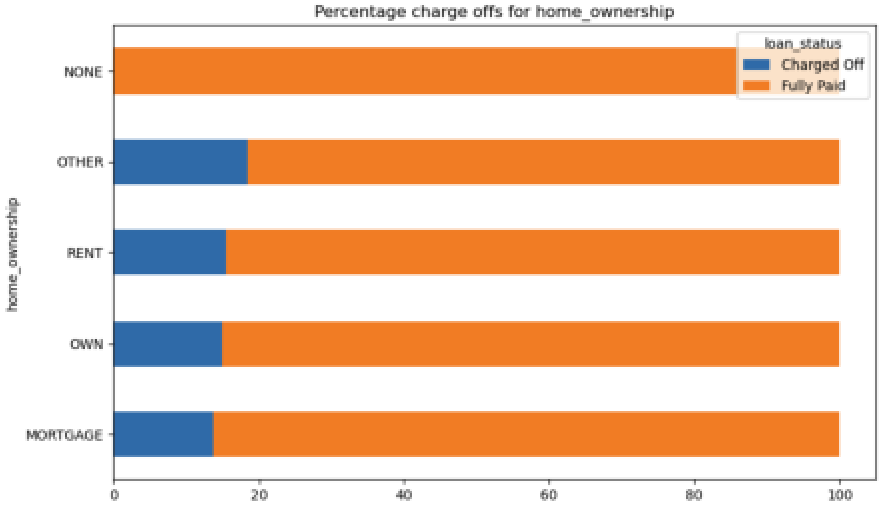
- Interest rates increase with grade A<B<C<D<E<F<G

- That should mean more defaults are happening with high interest rate and high revolving utilisation



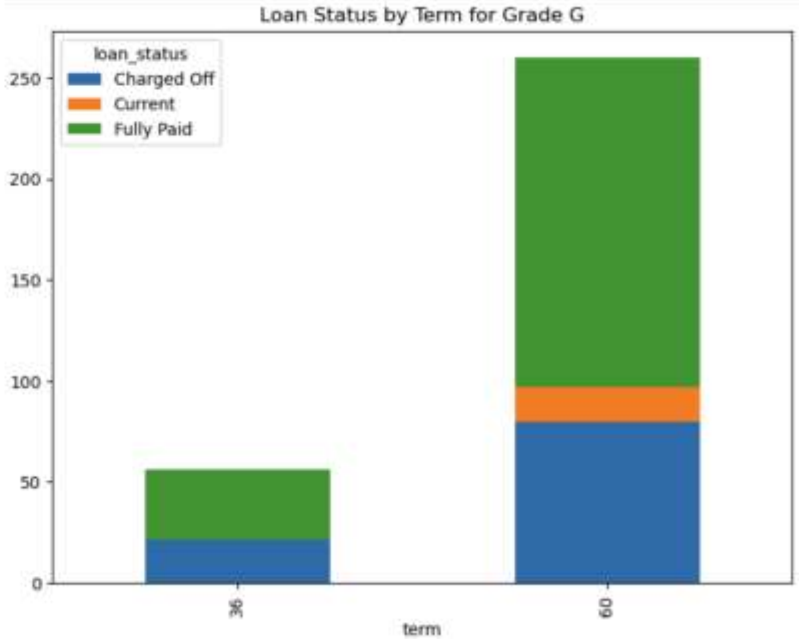
Home Ownership Split by loan status

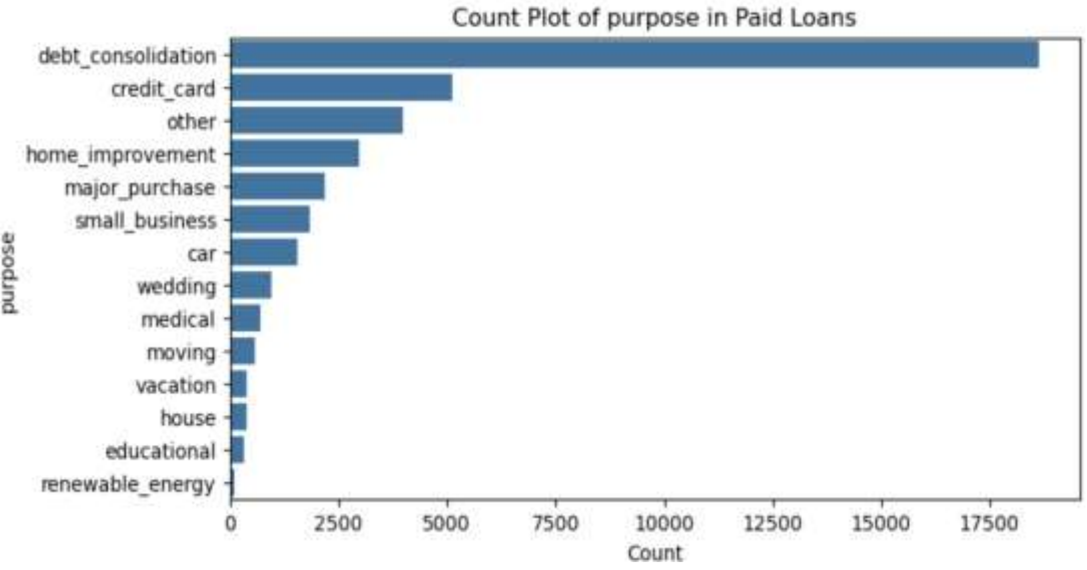
- Other Category has highest charge offs
- Rent is second highest charge off category
- Other is very less in quantity
- Conclusion would be to be more careful for RENT category while approving loan

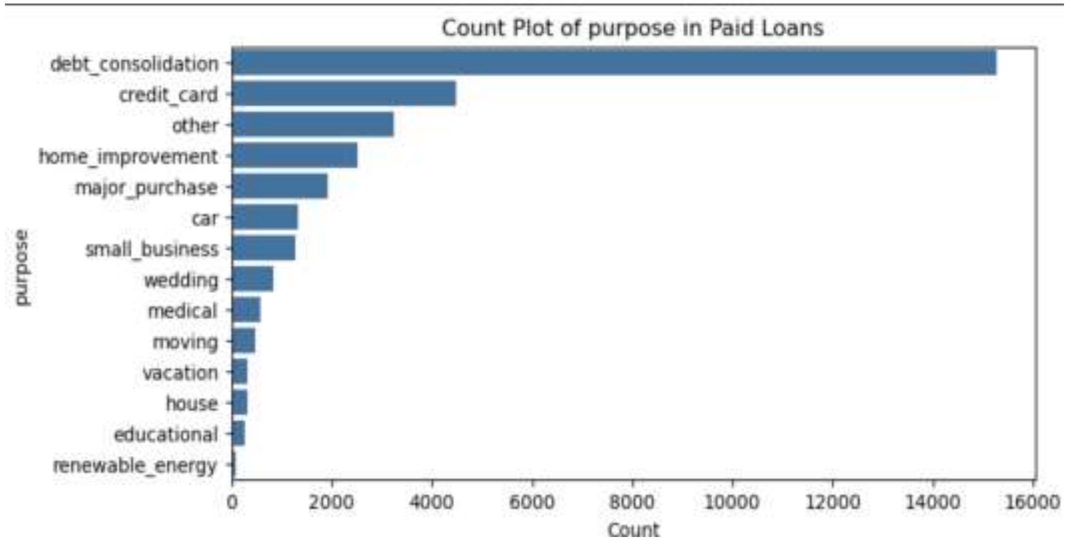


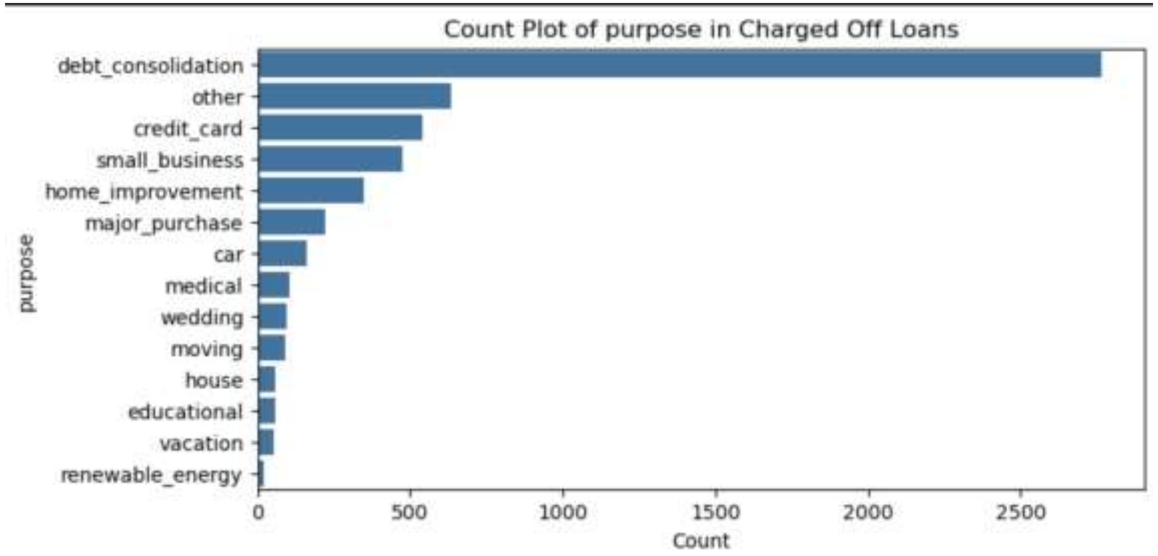
Loan Status by Term For Grade G

- More Charge offs observed in 60M category
- Grade G mostly have 60M term loans
- It is very clear that the financier wants to monetize more on Grade G with higher interest rates, term and loan amount
- But Grade G also comes with risk of more defaults



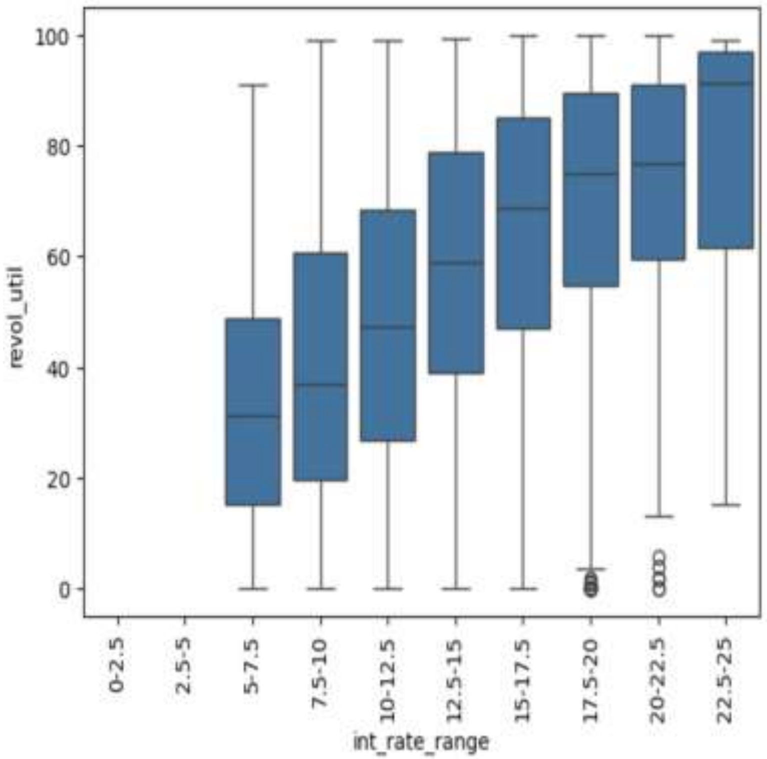


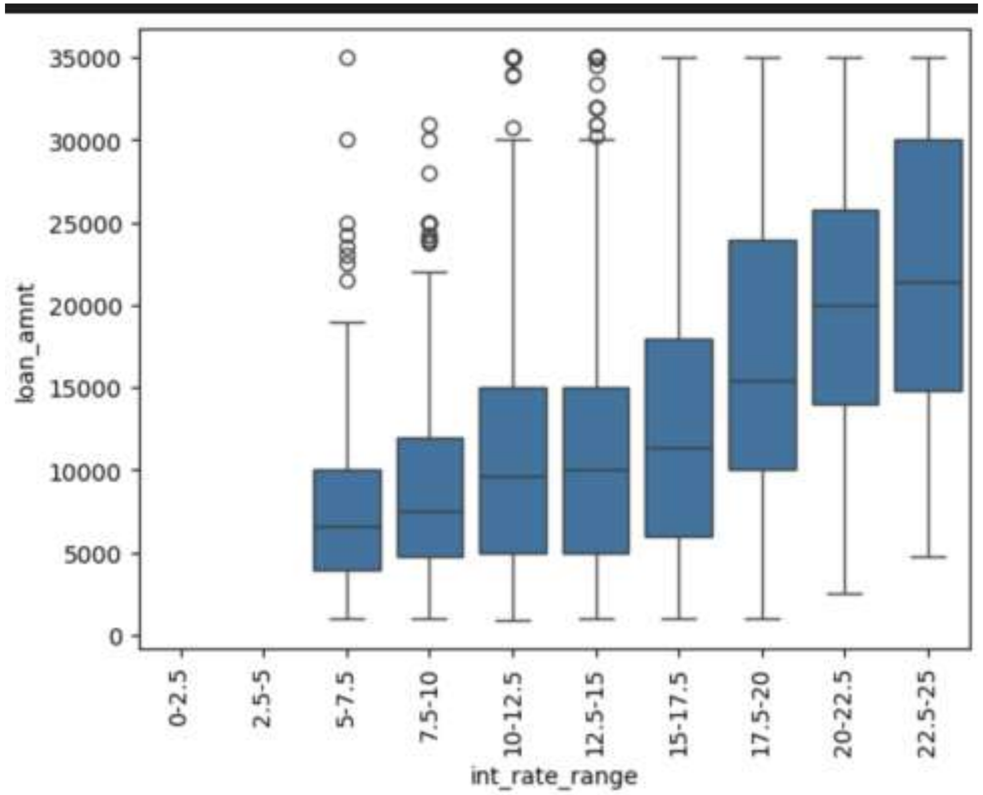




‘purpose’ small business is the riskiest category with 25% charge offs

- Interest rate ranges from, 15-25 has highest default ranges
- May be the interest rates are high due to revol_util of customer
- But high interest rates are also observed with high loan amounts
  - Loans at higher risk of default are exposing more money at risk





Conclusion

- Interest rate, revolving utilization, term, grade, home_ownership, purpose, address_state and earliest_cr_line seem to have direct impactful relationship with high 'Charge offs'

- Interest ranges between 15-25 have highest loan default percentage (charge off/total loan records)
  - Loan amounts are also increasing in range 15-25 with highest loan amounts being in 22.5-25% interest rate category
  - **Recommendation** would be to put a **cap loan amounts at higher interest rate to limit the losses**

- Loan term significantly impacts default rates. Although 36-month loans are more common**, 60-month loans have a much higher proportion of defaults**—about 25% of 60-month loans default compared to only 10% of 36-month loans. This indicates that longer-term loans are riskier and more likely to result in charge-offs.

- Credit grade is a strong indicator of loan default risk. **Lower credit grades (such as F and G) have significantly higher default rates**, with grade G experiencing defaults in over 30% of cases, compared to only 6% for grade

- High Revolving utilization for a customer is also an indicator of risk
  - If revolving utilization crosses **70-75 mark it is safer to reject loans**

- Loans with purpose of **small_business(25%)** and **renewable energy(18%)** has highest charge off rate

- 'earliest_cr_line ' seems to have interesting pattern where people starting their credit relations in **2007 (20%), 2006(19%), 1973 (18%)** has the most charge offs, it should be investigated what is common from new accounts policy perspective in these years

- Older credit relations are trusted with more loan amount earliest_cr_line with 1973 had an average loan amount of ~13000 whereas for 2006-07 average loan amount is ~7000-7500.