#LifeKoKaroLift

Lending Club Case Study

Godwin Paul Vincent

Lending Club: Business Understanding

upGrad

What is Lending Club?

Lending Club is a *consumer finance company* which specializes in lending various types of loans to urban customers, matching borrowers with investors using an online marketplace.

What is its loan approval process?

When the company receives a loan application, company must decide for loan approval:

- Loan approval is based on the applicant's profile, with the goal of minimizing credit loss
- **Credit loss** is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed
- '**Charged-off**' customers are the defaulters who cause the largest amount of loss to the lenders

What are the types of risks?

Two types of risks are identified:

- Rejecting applicants who are likely to repay the loan will result in loss of business
- Approving loans to applicants not likely to repay may lead to financial loss for the company.

Lending Club: Objective

upGrad

What is the business objective?

- The primary objective is aimed at minimizing credit loss to the company. Like most other lending companies, lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss).
- The goal is to identify these risky loan applicants, If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss.

What is the case-study exercise objective?

- 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.
- Identification of such factors for risky applicants **using EDA of the given dataset**, is the aim of this case study.

Lending Club: Approach

upGrad

Data Cleaning & Manipulation Checklist

- Dropping Rows - where loan_status = "Current"
- Dropping Columns with N/A values
- Convert the data types
- Identify columns with blank values which need to be imputed
- Analysis of the dataset post cleanup
- Outlier Treatment.

Analysis

- Univariate
- Bivariate
- Derived Metrics Analysis

Conclusions

- Inferences and Recommendations

Lending Club: Data Understanding

upGrad

The following are the key important attribute groups in the given data set

Customer attributes: The attributes describing the 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 demographically.

Lending Club: Data Understanding

upGrad

Loan attributes: The attributes related to the loan.

- **Loan Amount (loan_amt) –** Borrower requested loan amount
- **Funded Amount(funded_amnt) -** The recommended amount based on history/risk
- **Funded Amount Invested(funded_amnt_inv) -** The final amount offered as loan by the investor
- **Loan Status (loan_status)** - Key Leading Attribute the column has three distinct values
    - **Fully-Paid** - The customer has successfuly paid the loan
    - **Charged-Off** - The customer is "Charged-Off" ir 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
- **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 publically available for the customer. Higher the value, lower is the success rate.
- **Grade (grade)**
- **Term (term)**
- **Loan Date (issue_date)**
- **Purpose of Loan (purpose)**
- **Verification Status (verification_status)**
- **Interest Rate (int_rate)**
- **Installment (installment)**

Lending Club: Data Understanding

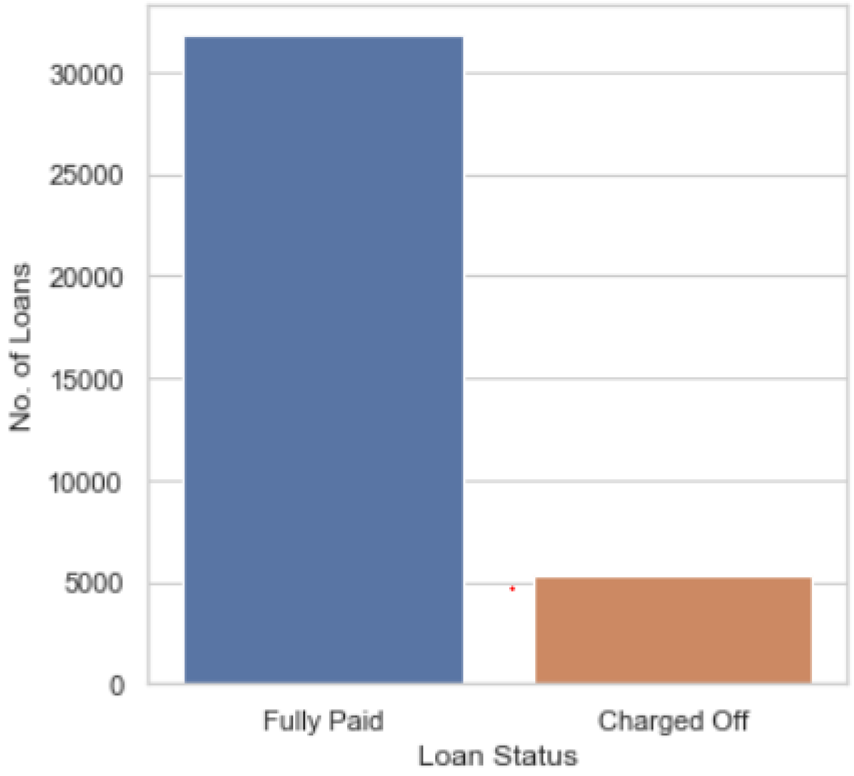
upGrad

Other Analysis:

- **Customer Behaviour Columns** - Columns which describes customer behaviour will not contribute to the analysis. The current analysis is at the time of loan application, but the customer behaviour variables generate post the approval of loan applications.
  - (delinq_2yrs, earliest_cr_line, inq_last_6mths, open_acc, pub_rec, revol_bal, revol_util, total_acc, out_prncp, out_prncp_inv, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int, total_rec_late_fee, recoveries, collection_recovery_fee, last_pymnt_d, last_pymnt_amnt, last_credit_pull_d, application_type)
- There are multiple columns with **NA values** only (*evaluated after dropping rows with loan_status is Current)*. The **columns will be dropped**
- There are multiple columns where the **values are only zero**, the **columns will be dropped**
- There are no summary, header or footer rows, column number found in the dataset.

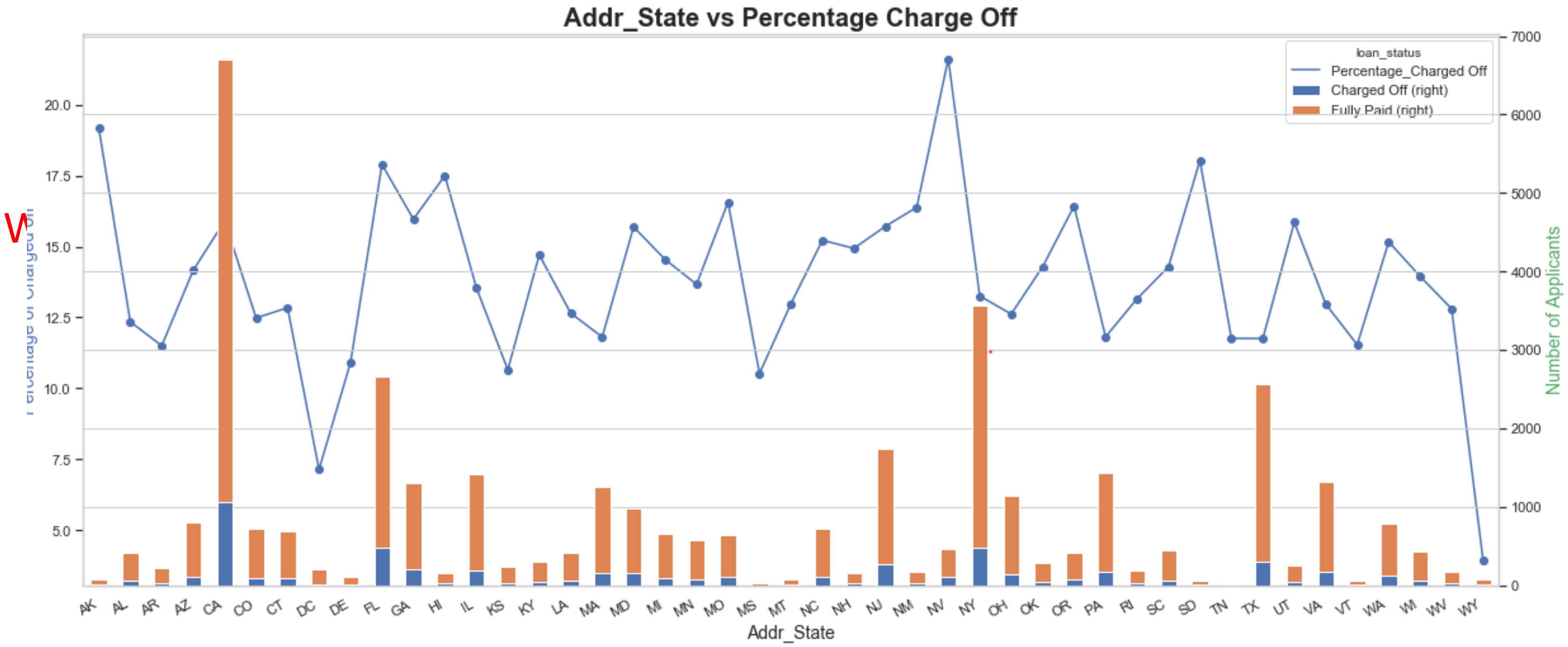
Loan Fully paid vs Charged Off

- More than 30000 loans are fully paid.
- Around 5000 loans are charged off.



Loan Analysis by State

- Maximum no. of loans are applied by people from CA.
- Maximum no. of loans are charged off for people of NV.

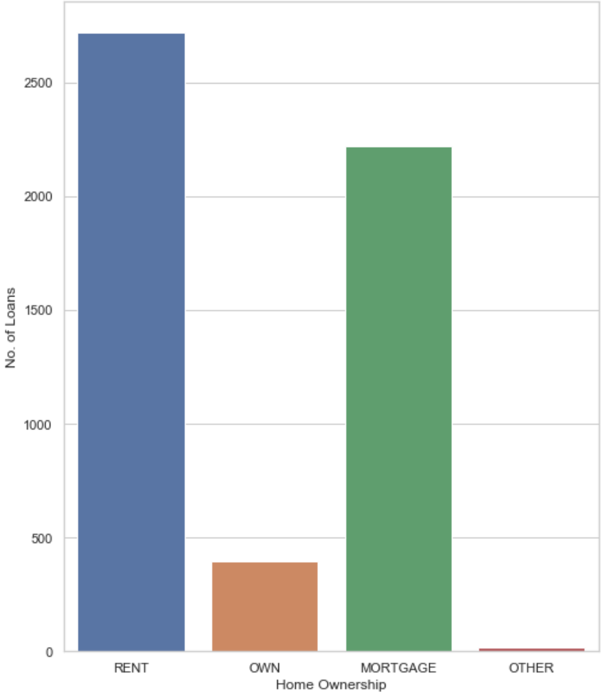


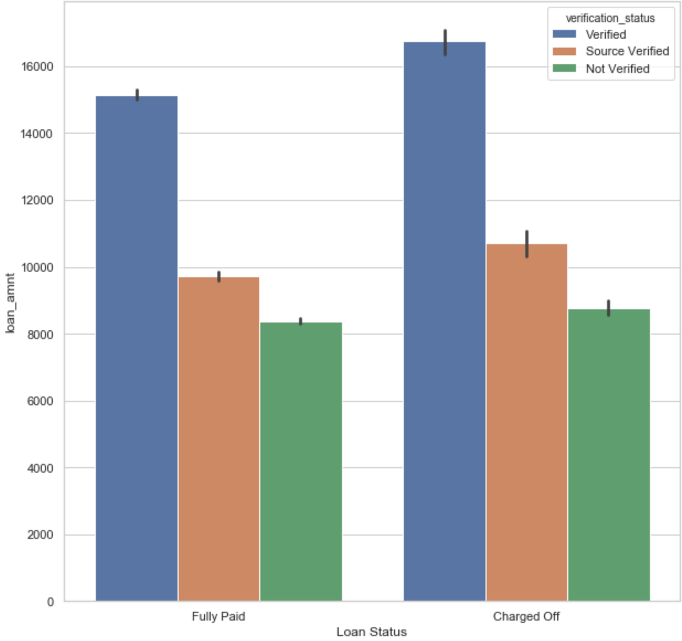
Lending Club: Data Analysis

upGrad

Loan Status and its Relationship with Verification Status and Home Ownership.

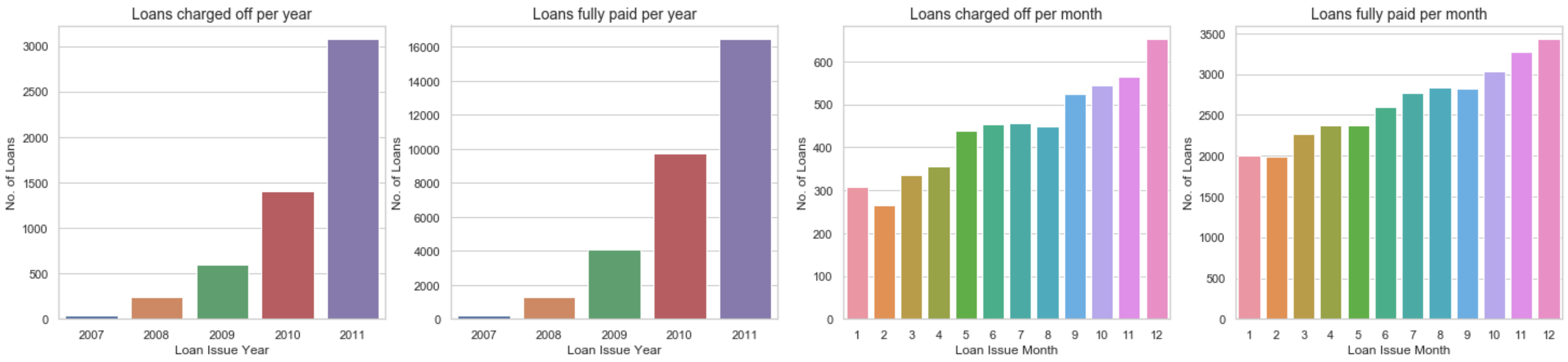
- The highest number of charged-off loans corresponds to those with a Verified verification status.
- Individuals who live in rented accommodations have the highest number of charged-off loans





Trends in Loans by Year and Month

- The number of loans, both fully paid and charged off, has been consistently increasing year-over-year, with a peak in 2011. This positive trend indicates a growing demand for loans, which is good news for Lending Club.
- Looking at the monthly trends, it becomes clear that most loans are either fully paid or charged off towards the end of the year, likely due to a higher number of loan applications during that time.

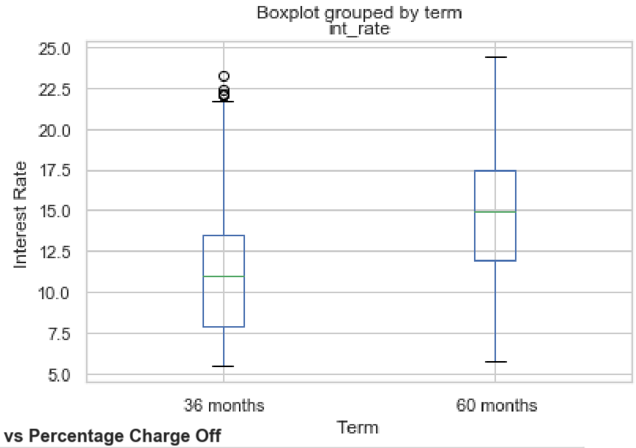


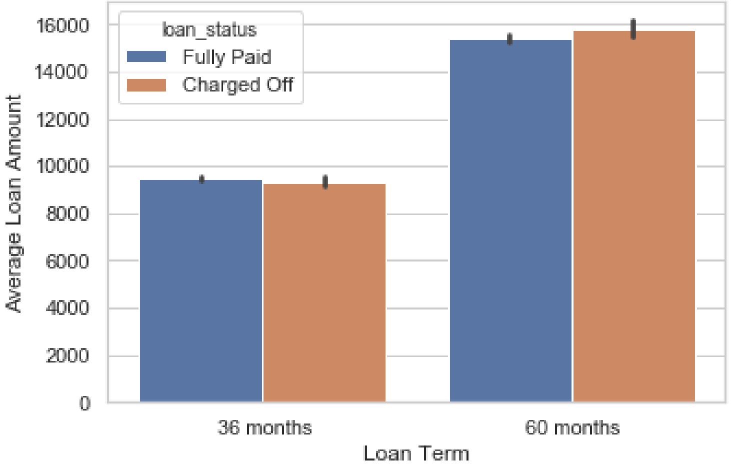
Lending Club: Data Analysis

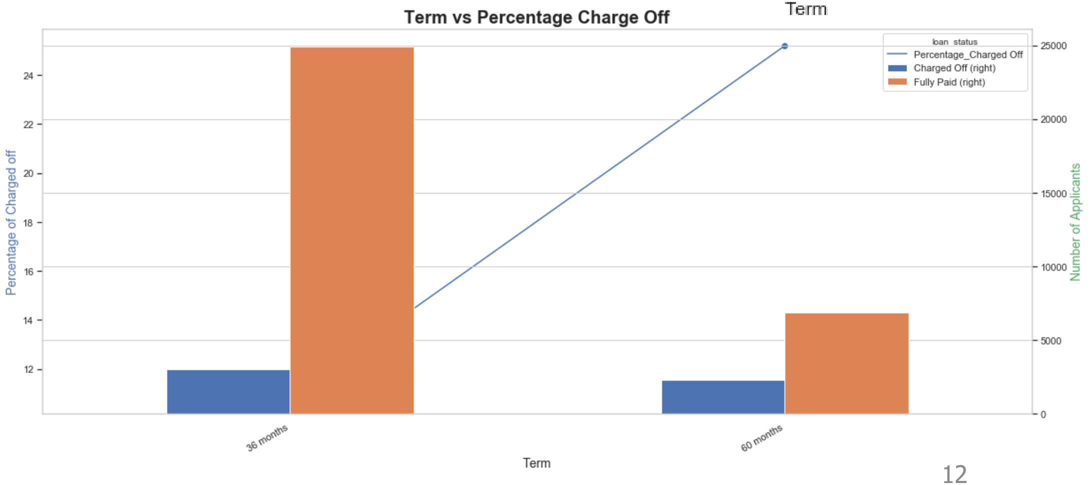
upGrad

Analysis by Loan Term

- Interest rates are higher for 60-month loans.
- Loan amounts are similar for both terms and loan statuses.
- Average loan amount is higher for 60-month loans.

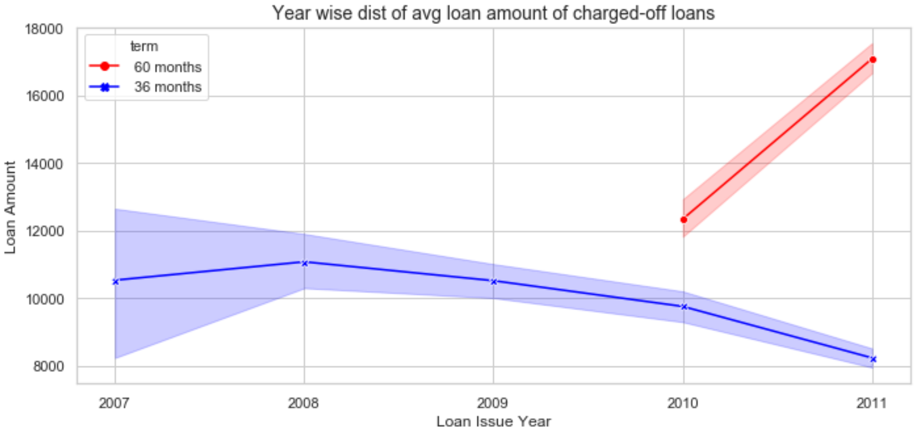


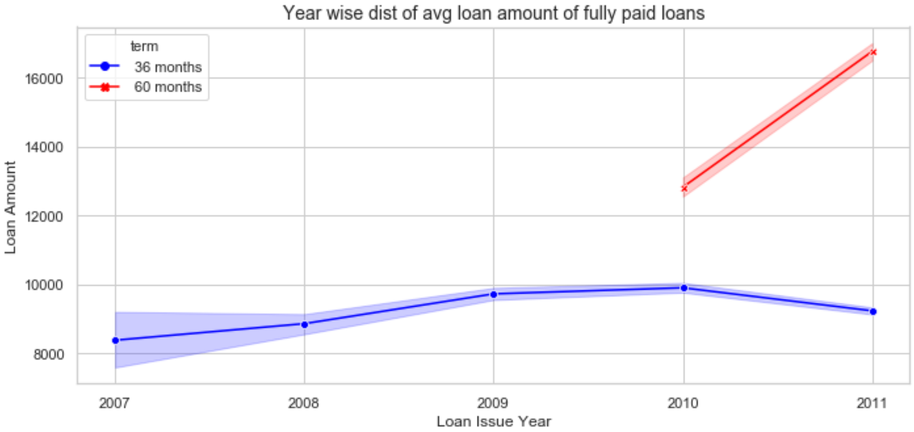




Analysis by Loan Term (Continued)

- Average interest rates for both charged off and fully paid loans in the 36-month term have been increasing yearly.
- The 60-month loan term was introduced in 2010, coinciding with a significant decrease in average interest rates for both charged off and fully paid loans.
- The average interest rate for 60-month loans is higher compared to the 36-month term.

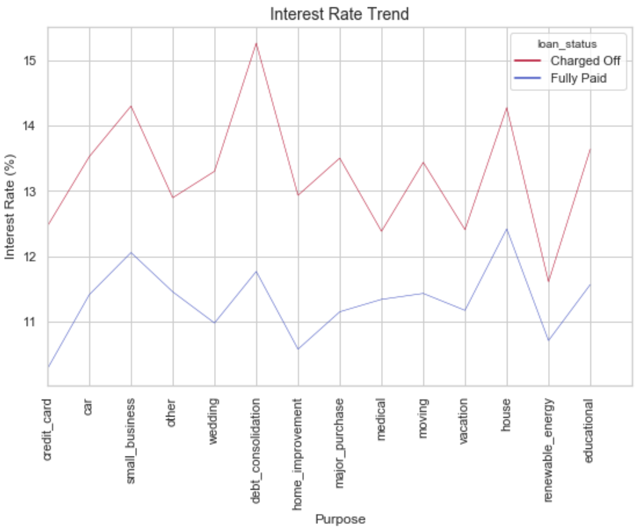


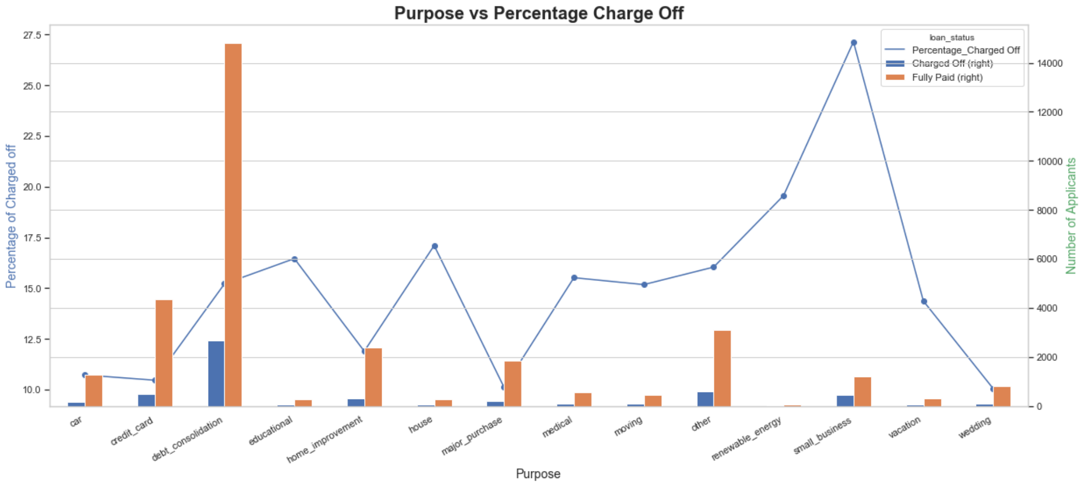


Lending Club: Data Analysis

Impact of Loan Purpose on Charged Off Percentage

- Regardless of the loan purpose, charged off loans generally have higher interest rates..



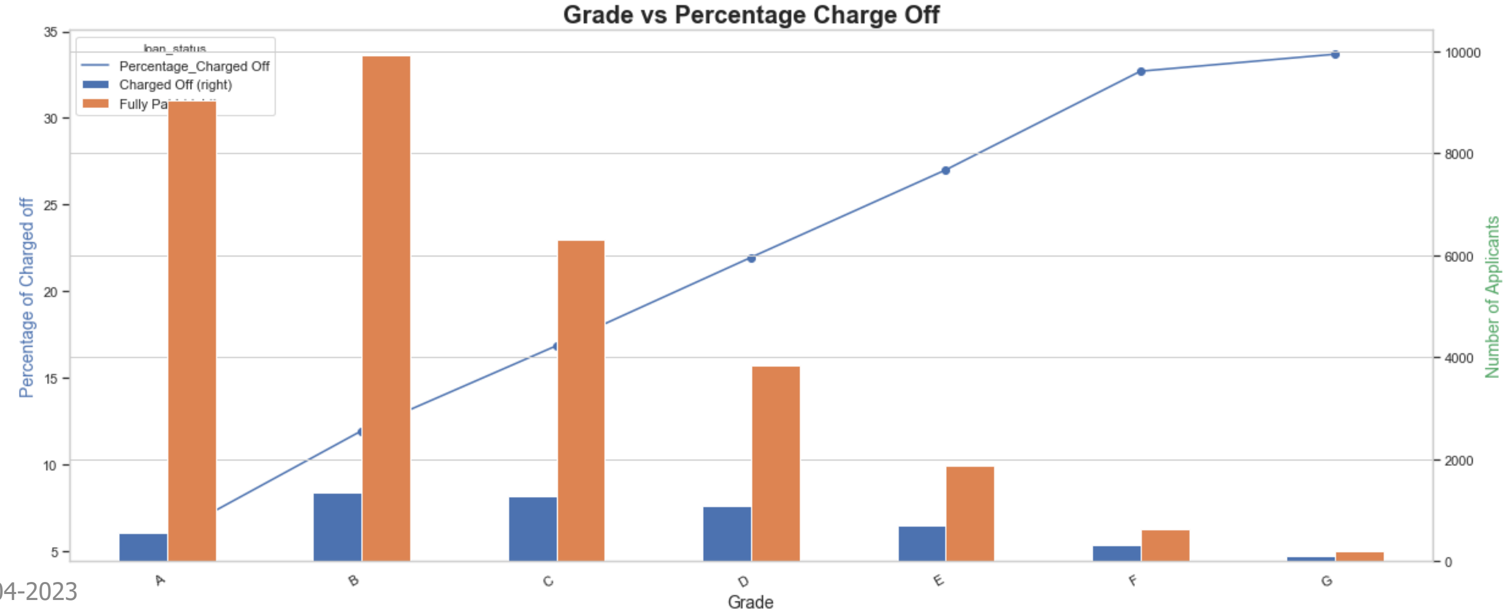


Lending Club: Data Analysis

upGrad

Impact of Loan Grade on Charged Off Percentage

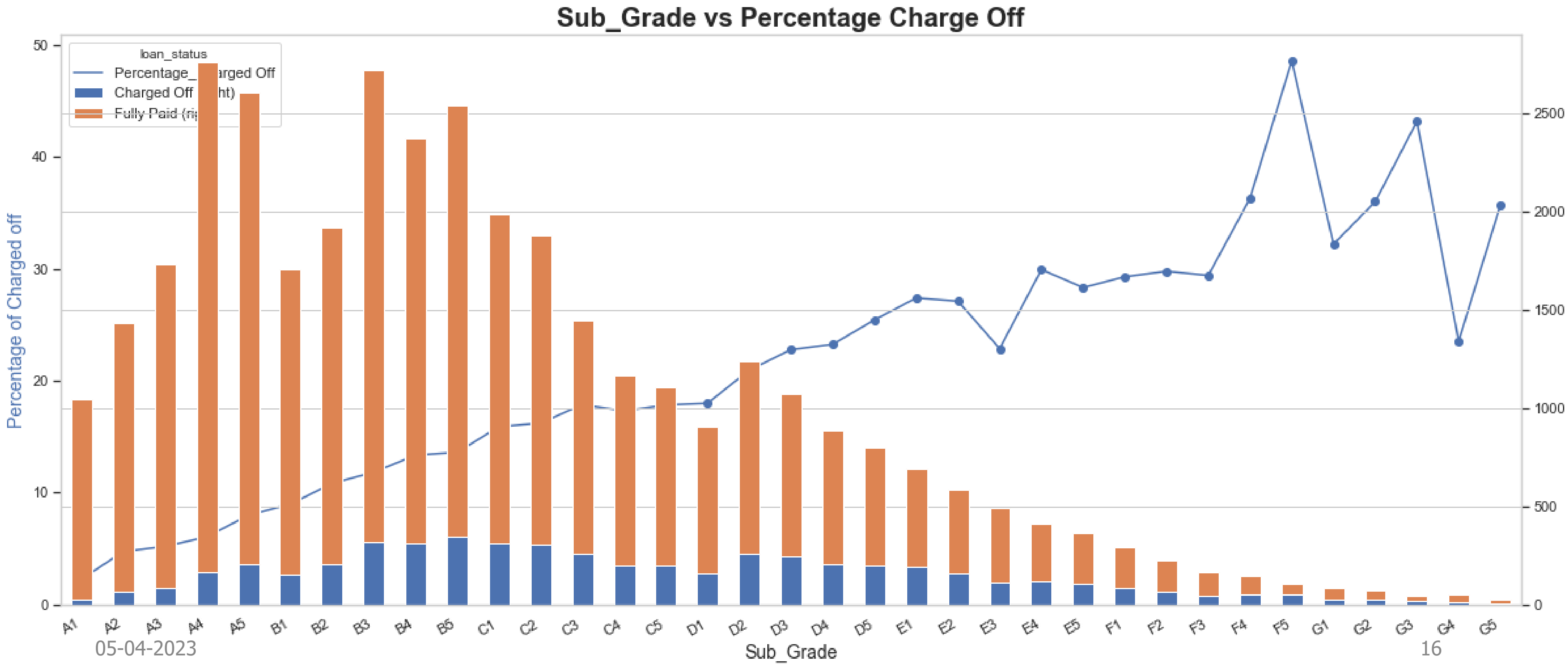
- Grades A and B are considered safe loans, with the sub-grades A4 and B3 having the highest number of applicants.
- In contrast, Grades E, F, and G are considered less safe compared to others.



Lending Club: Data Analysis

upGrad

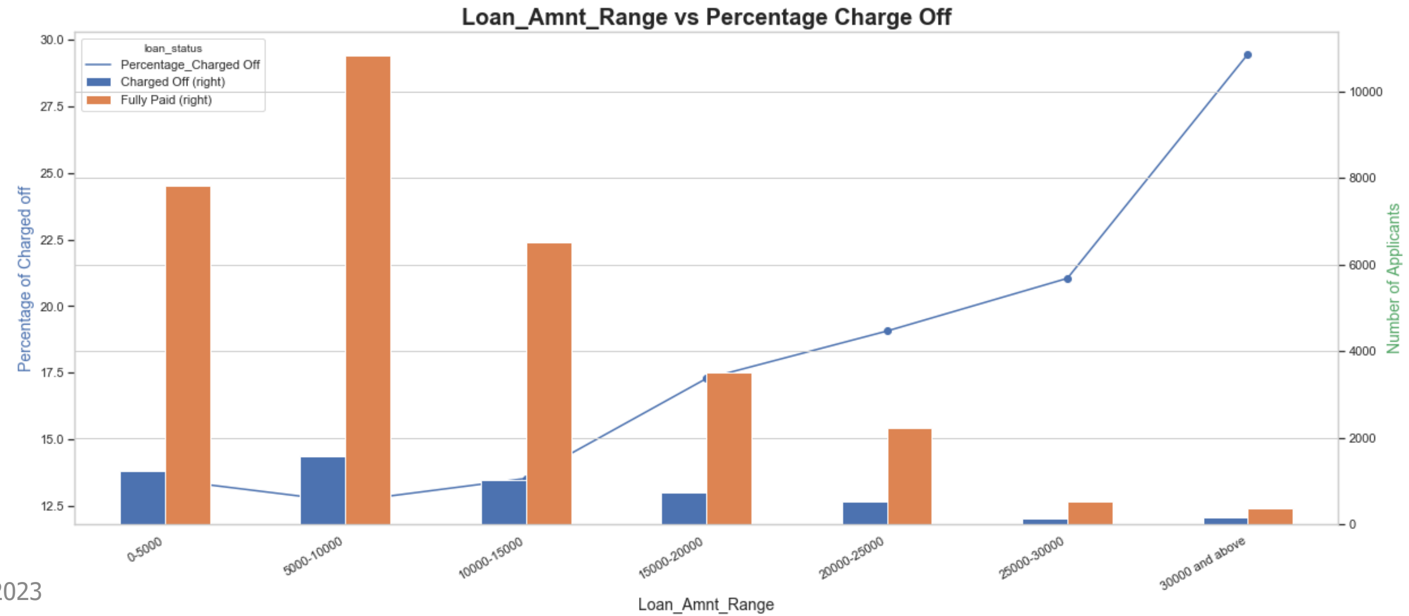
Sub Grade vs Charged Off Percentage



upGrad

Impact of Loan Amount on Charged Off Percentage

- Loans with higher amounts are more likely to be charged off. The largest charged off percentage (29.47%) is observed for loans with a value of 30,000 or more.
- Meanwhile, the majority of loans (12,410) have amounts between 5,000 and 10,000.

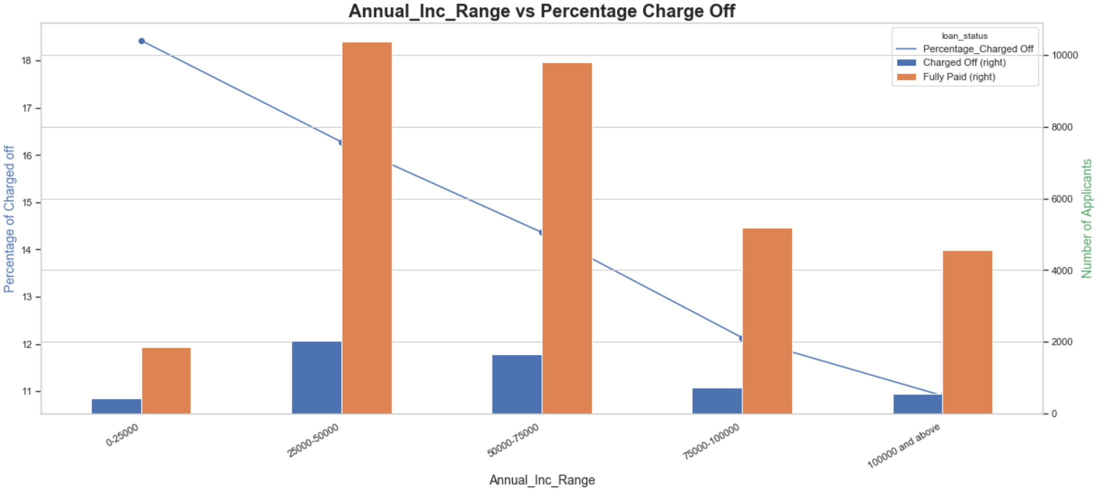


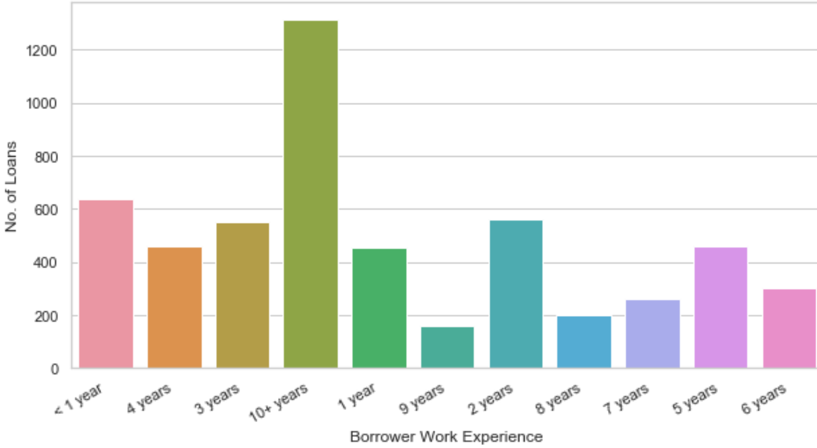
Lending Club: Data Analysis

upGrad

Impact of Annual Income on Charged Off Percentage

- As annual income increases, the percentage of charged off loans decreases.
- The lowest charged off percentage is observed for incomes greater than 100,000, while the highest is for incomes below 25,000.
- The greatest number of loans are applied by applicants with annual incomes ranging from 25,000 to 75,000.



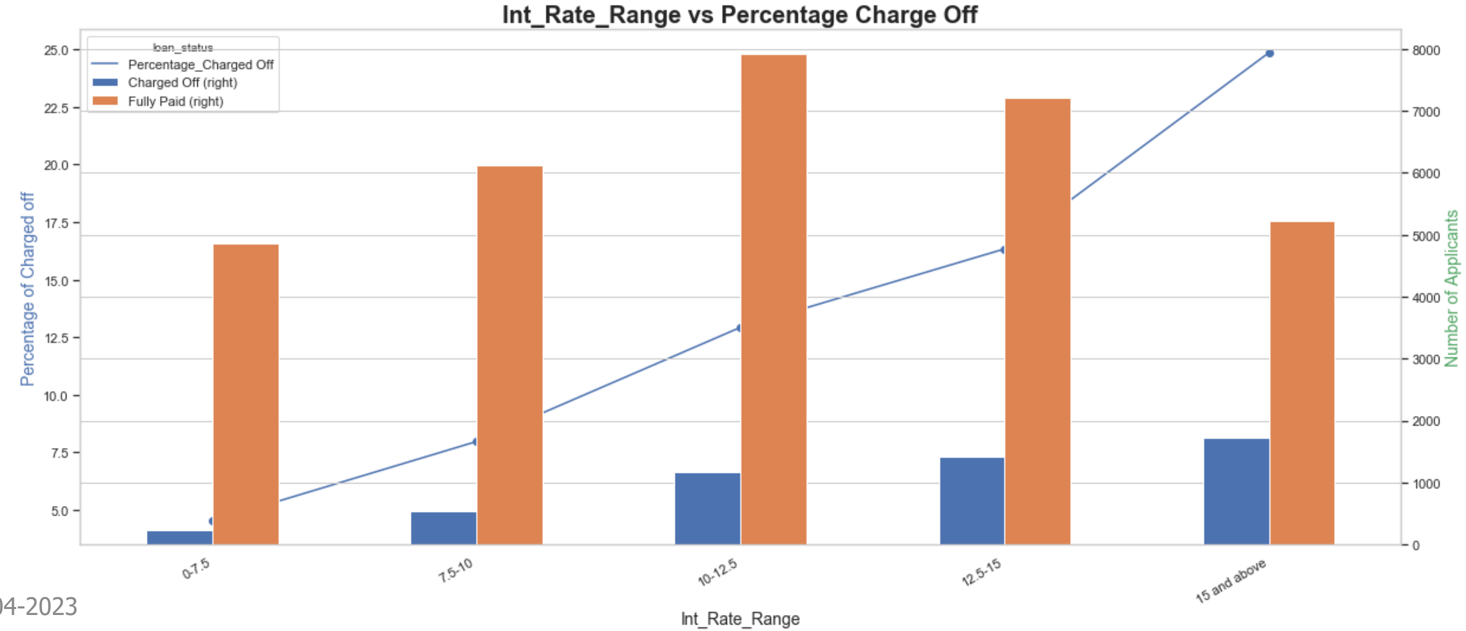


Lending Club: Data Analysis

upGrad

Impact of Loan Interest Rate on Charged Off Percentage

- Higher interest rates are associated with higher charged off percentages.
- The lowest charged off percentage is observed for interest rates below 7.5%, while the most loans are applied by applicants with interest rates ranging from 10% to 15%.

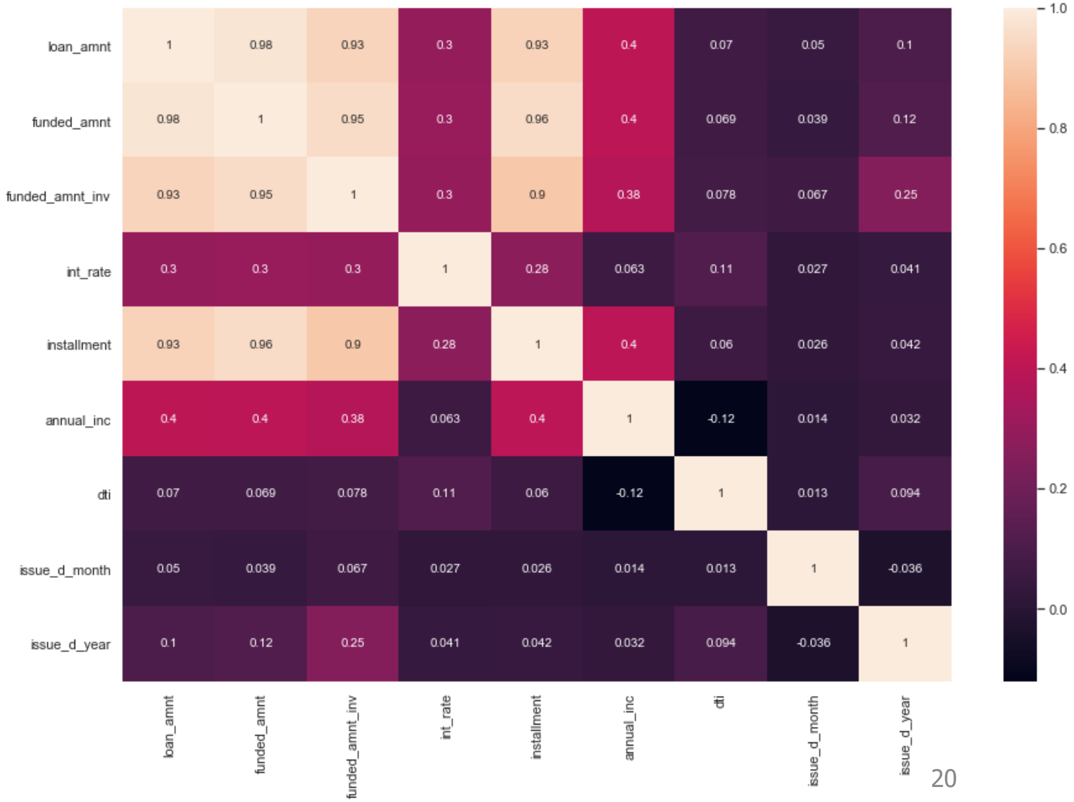


Lending Club: Data Analysis

upGrad

Correlation Heat-Map

- The loan amount, funded amount, funded amount invested, and installment show strong positive correlations with each other.
- Meanwhile, annual income exhibits a negative correlation with DTI, suggesting that as income increases, the DTI decreases.



Lending Club: Conclusion

upGrad

Drivers for Risky Loan (Rejection):

- **Loans for small businesses:** Small business loans may be considered riskier.
- **Loan amounts of 30,000 USD or more:** Higher loan amounts may be a sign of greater risk.
- **Low annual income (less than 25,000 USD):** Low-income levels may indicate a higher likelihood of default.
- **High interest rates (above 15%)**
- **Grades E, F, and G:** These loan grades are riskier
- **Rent or mortgage status:** Renters or those with high mortgage payments are at a higher risk of default.
- **60-month loan terms:** Longer loan terms may increase the likelihood of default.