Lending Club case study

Analyst:
Sarva M Sumanth

Contents:

❑Business context

❑Problem Statement

❑Analysis Approach

❑Data Cleaning

❑Univariate analysis

❑Bivariate analysis

❑Multivariate analysis

❑Insights from analysis

❑Conclusion / Driver variables behind default
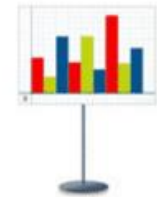
Business Context:

- **Consumer finance company** specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.

Borrowers apply for loans.
Investors open an account.

Borrowers get funded.
Investors build a portfolio.

Borrowers repay automatically.
Investors earn & reinvest.

- Types of risks are associated:
  - 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

- We need to calculate risk for each applicant and process accordingly in order to obtain profits and minimizing losses.

Problem Statement:

- Lending loans to 'risky' applicants is the largest source of financial loss (credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss.

- For every loan applied by the customer, bank can take either accept or reject the application:
  - **Loan accepted:** If the company approves the loan, there are 3 possible scenarios described below:
    1. **Fully paid**: Applicant has fully repaid the loan
    2. **Current**: Applicant is in the process of paying the instalments i.e., ongoing loan
    3. **Charged-off**: Applicant has not paid the due for a long period of time, i.e. he/she has defaulted on the loan
  - **Loan rejected**: Loan application is rejected because the candidate does not meet the qualifying criteria which is set before. Since the loan was rejected, there will be no details that has to be tracked for these applicants.

- **OBJECTIVE**: Finding the **driving factors** behind loan default, i.e. the variables which are strong indicators of default. Company applies these insights for its portfolio and risk assessment.

Analysis Approach:

We have been provided with complete loan data for all loans issued through the time period 2007 to 2011. We Apply EDA on the data and get insights to obtain our objective.

- Steps in our Analysis:
  - Data Sourcing.
  - Data Cleaning: We standardise values to obtain format that can be analysed on.
  - Univariate Analysis: Understanding single variables to understand their spread.
  - Bivariate Analysis: Understanding relation between two variables with help of plots.
  - Multivariate Analysis: Relation among more than two variables with pair plot.
  - Insights from the analysis.
  - Conclusion.

Data Cleaning:

- Lets see the shape of the loan_data table.

- We performed the following steps on the table:
  - Fix rows and columns
  - Fix missing values.
  - Standardize values.
  - Fix Invalid values.
  - Filter data.

- We arrive to this table shape after completing the above steps.

- The resultant table must not be prone to missing values and null columns. Having these in our table, largely affects the analysis.

```
#DataFrame shape
loan_data.shape
```

(39717, 111)

	int_rate	revol_util
0	10.65%	83.70%
1	15.27%	9.40%
2	15.96%	98.50%

	int_rate	revol_util
0	10.65	83.7
1	15.27	9.4
2	15.96	98.5

```
loan_data.shape
```

(38577, 18)

```
#print the columns after DataCleaning Process
loan_data.columns

Index(['loan_amnt', 'funded_amnt_inv', 'term_in_months', 'int_rate', 'grade',
       'sub_grade', 'emp_exp_in_years', 'home_ownership', 'annual_inc',
       'verification_status', 'loan_status', 'purpose', 'addr_state',
       'debt_to_income_ratio', 'credit_utilisation', 'pub_rec_bankruptcies',
       'issued_month', 'issued_year'],
      dtype='object')
```

Univariate Analysis:

Observations from univariate analysis:

• Funded amount to applicants follow binomial distribution and ranges from 5000 to 35000.

• Grade chart:
  • Most loans sanctioned are for Grade B, A applicants, followed by C.
  • Less percentage of loans are disbursed to grade G,F applicants.

• Purpose for taking loans is the most for 'debt-consolidation'.

• Rented house & mortgage house applicants are more in loan applicants.

• From address column we understand that most applicants are from 'CA`, `NY` and are driving majority of business to the banks.

Countplot for Grades of applicants

Countplot for home_ownership
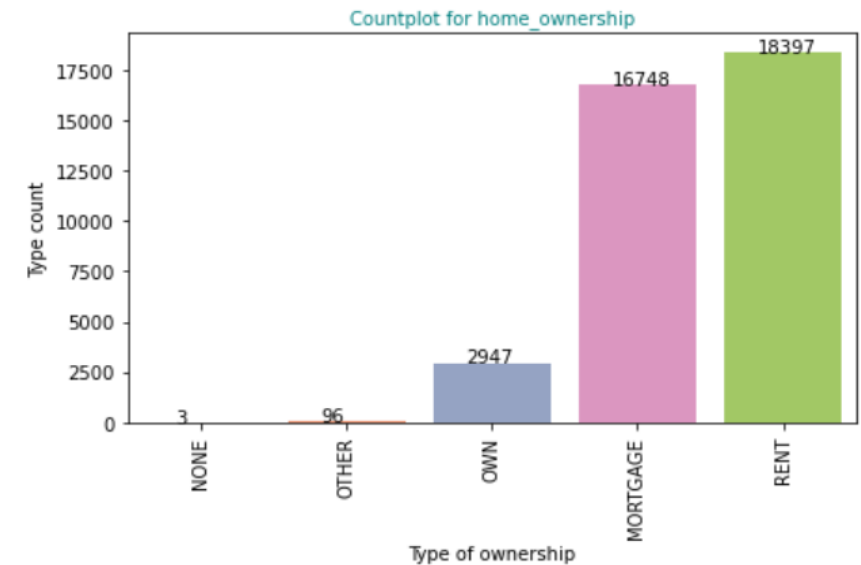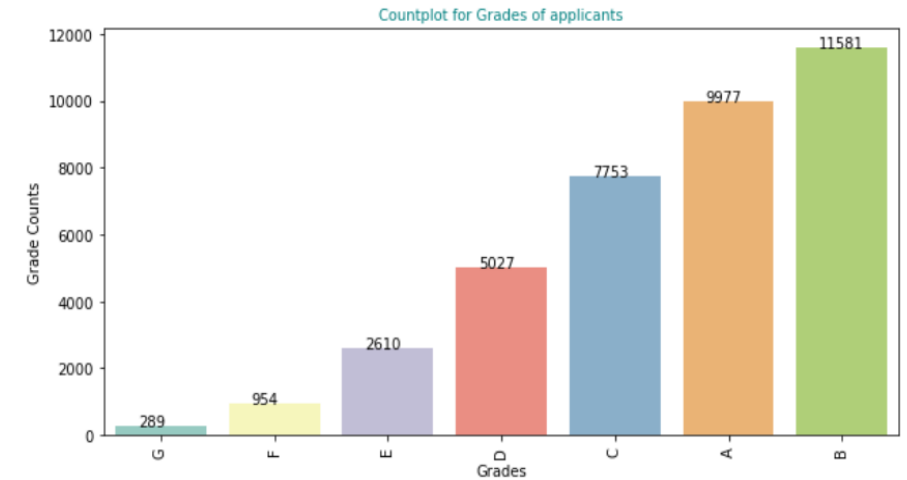
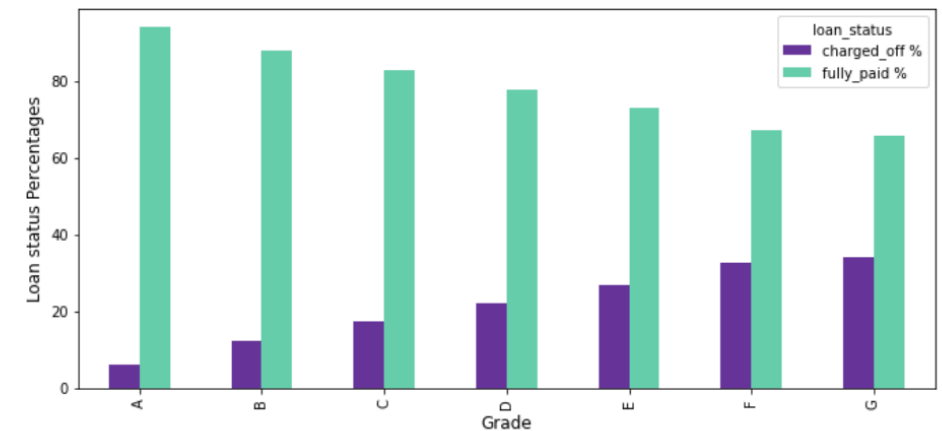# Bivariate analysis:

Observations from Bivariate analysis:

- Interest rate:
  - Smaller interest rates bring more healthy recovery ratios.
  - The Higher the interest rate the more the charged-off percentage.

- Grade variable:
  - The logic stands here, It is safe to lend to 'A' category applicant.
  - The charge-off ratio increases as the grade increases from A-B-C-D-E-F-G

- Credit_utilisation:
  - People who use less credit are safer to disburse money, as they have better repayment ratio.
  - Higher credit utilised applicants have higher charge-off ratio.


Int_Rate_Category vs Loan Status
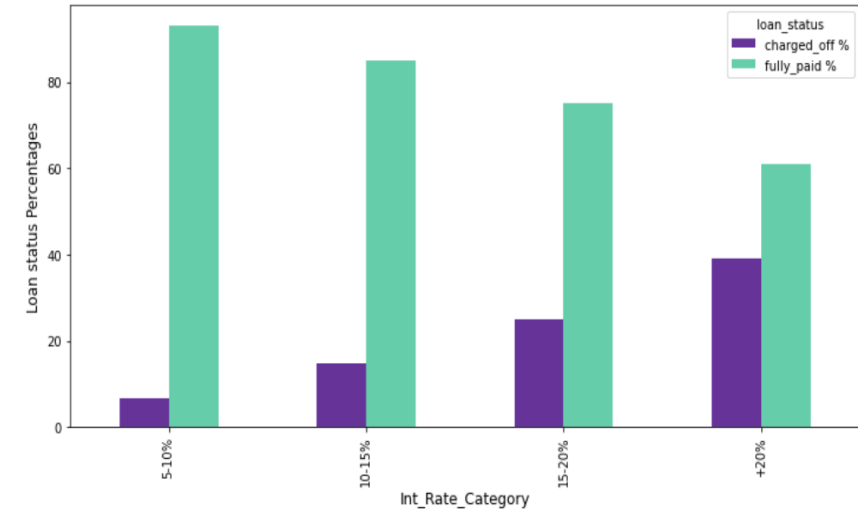
Grade vs Loan Status

Bivariate analysis:

Observations from Bivariate analysis:

- Term_period_in_months:
  - Charge off ratio is less if the term period is less i.e. up to 36 months.
  - Charge off ratio increases with increase in term period of the loan

- Purpose Variable:
  - People who take loan for 'small business' charged off more than others.
  - Wedding, major_purchase, car and credit card purchases has relatively low charge off ratio.

Term_In_Months vs Loan Status

Purpose vs Loan Status
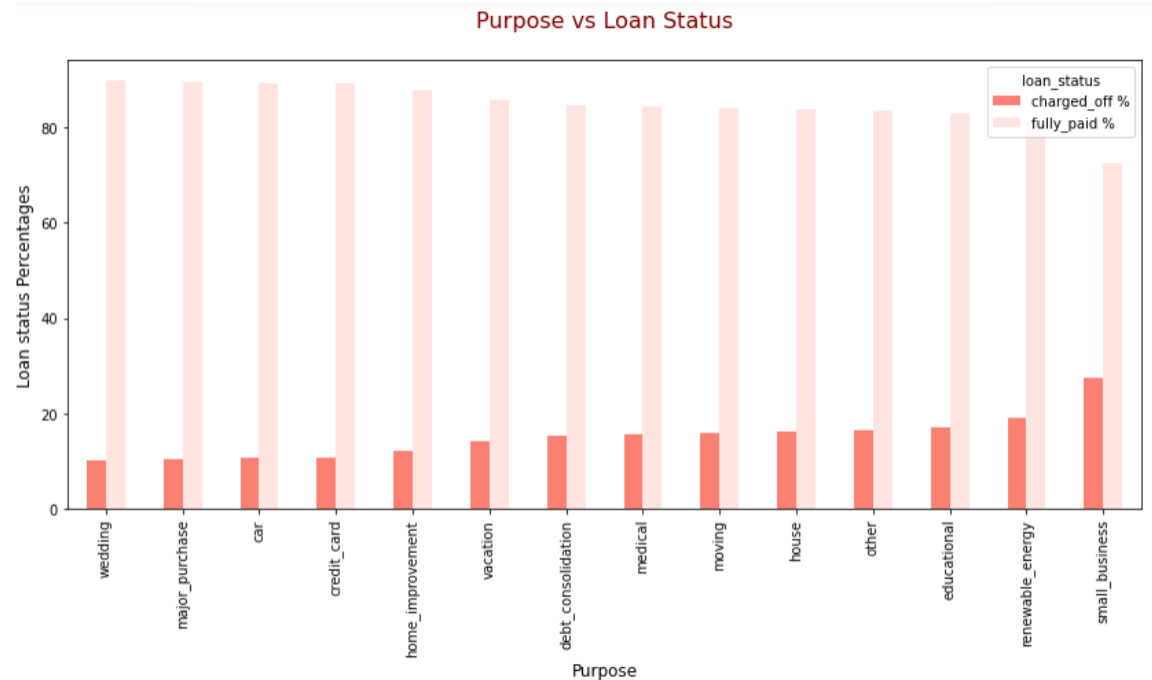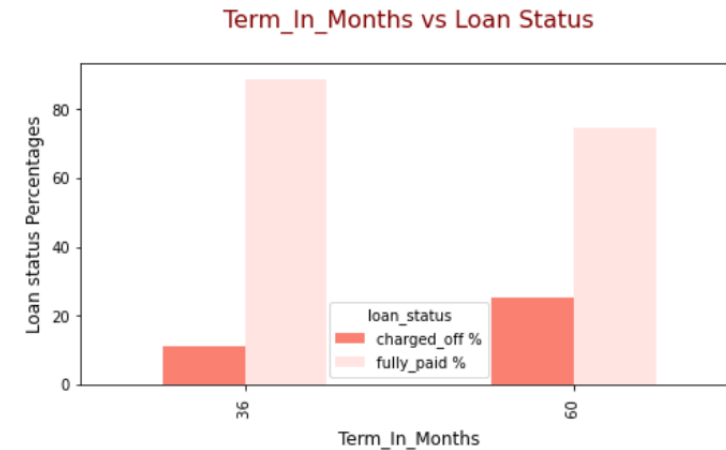
Bivariate analysis:

Observations from Bivariate analysis:

- Dti (Debt-income ratio):
  - It is observed that individuals with high 'debt_to_income_ratio' are defaulting more.

- Public record bankruptcies:
  - The distribution follows the below logic:
  - The applicant with 1/1+ public record bankruptcies has high default ratio.
  - We can infer that applicants with more bankruptcy count have higher probability to charge-off.

Debt_To_Income_Ratio_Category vs Loan Status
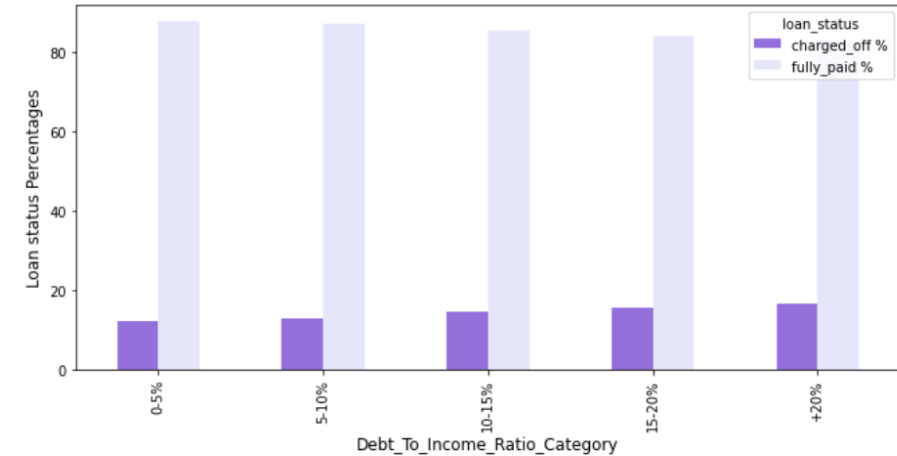
Pub_Rec_Bankruptcies vs Loan Status

Multivariate analysis:

Observations from Multivariate analysis:

- If interest rate and loan amount are high the charge-off increases and additionally if there is a public bankruptcy record then the charge is even higher.

- If debt-to-income ratio is higher then only people with low interest rates largely paid back the loan whereas people who are given loan with higher interest rates charged-off more.

- Lower annual income people with high interest rates charged off more than higher annual income people
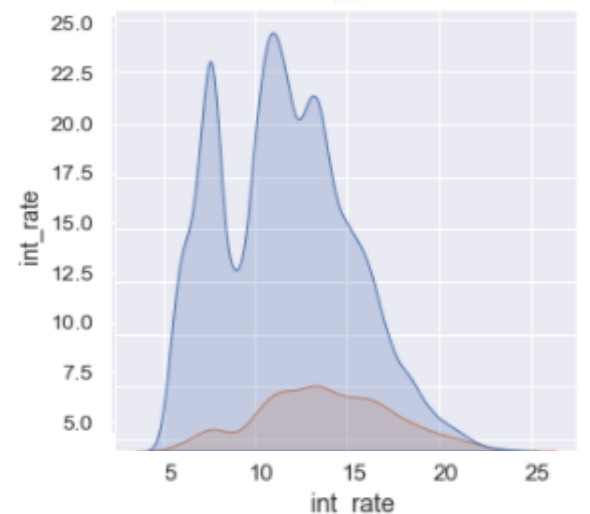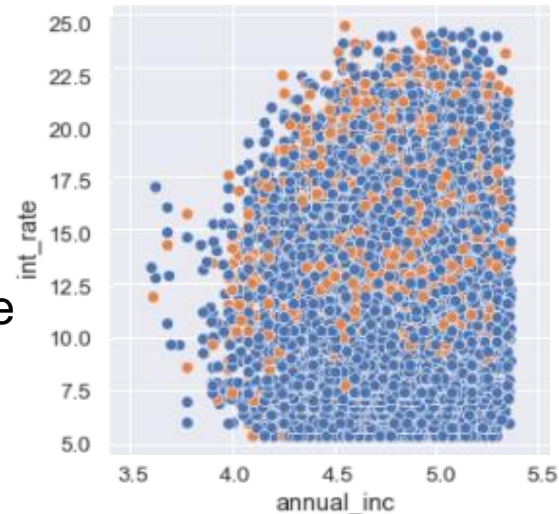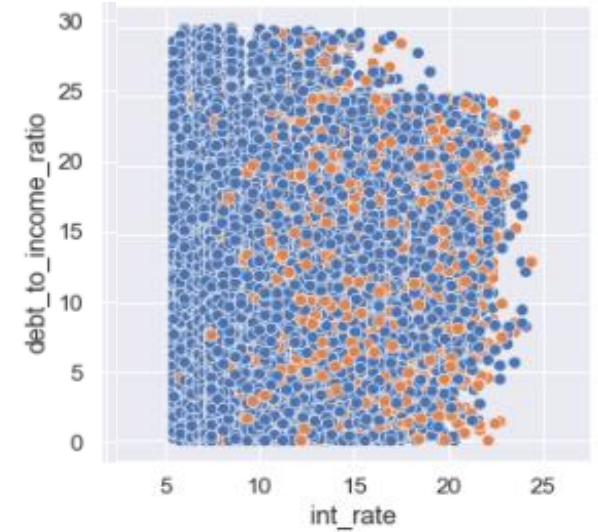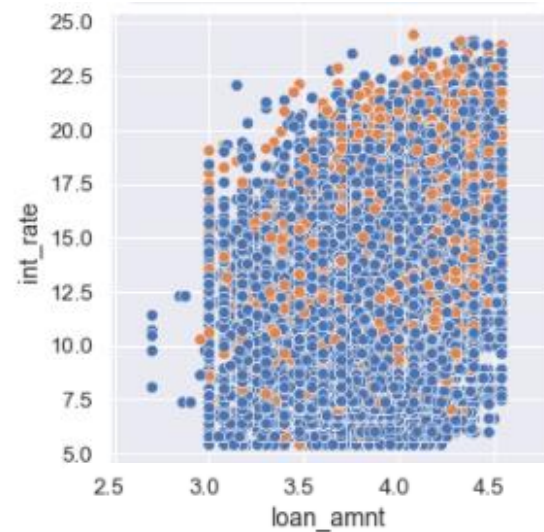
Driver variables behind loan default:

LOAN ATTRIBUTES	CONSUMER ATTRIBUTES
1. Interest rates. 2. Loan amount. 3. Term (Loan Tenure)	1. Purpose. 2. Public record bankruptcy. 3. Credit utilization. 4. Grade. 5. Annual Income.

Insights and recommendations from our analysis:

LOAN ATTRIBUTES	CONSUMER ATTRIBUTES

LOAN ATTRIBUTES

1. Lending with higher interest rates (above 14) is major factor of default, so those applicants should be lent with extra care.
2. Higher loan amount sanctioned is another variable that drives defaulting, this attribute should be studied carefully to avoid default.
3. High term period for the loan increases the chance of defaulting. So, lending at higher term should be supported by other positive attributes.

CONSUMER ATTRIBUTES

1. Purpose of loan is small-business or renewable energy then default percent is around whopping 20%. Lending to them is highly risky.
2. Public record bankruptcy of the individuals is another factor that is significant, person with 1 record has default percent of 20 and person with 2 or more records has default percent as high as 40. So such applicants can be rejected to avoid losses.
3. People with high credit utilization charge off more, lend if other positive attributes present.
4. F,G grade people tend to default more, not safe to disburse loan.
5. Lower income people tend to default more if interest rate is higher. Low interest rates are suggested to them.

Application with any of these qualities is risky and harmful and applications with combination of these qualities should be taken extra care in scrutiny, to avoid credit loss, as they carry multiplied risk.