

# Lending Club Case Study

Program : Executive PG Programme in Machine Learning

By:  
Narendra Singh Shekhawat  
Chintan Bhavsar

# Overview

- Problem Statement
- Study Objective
- Data Preparation
- Data Processing
- Data Analysis Univariate and Segmented-Univariate
- Data Analysis : Bivariate
- Conclusion

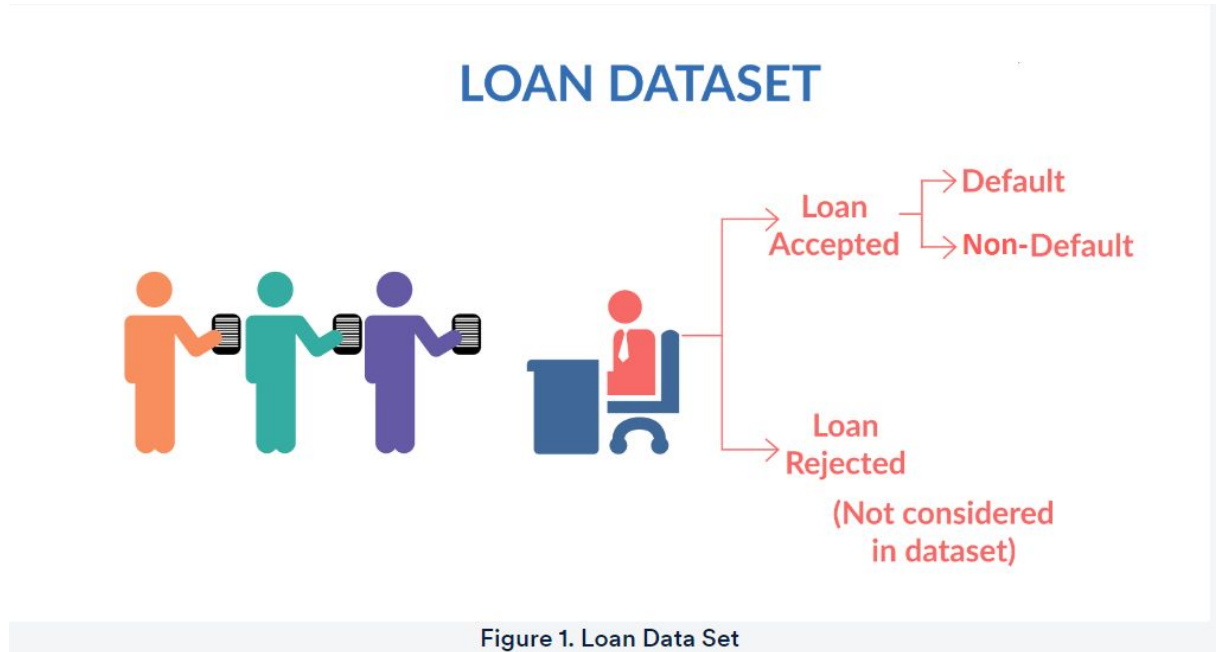
# Problem Statement

The objective of this case study is to identify risky loan applicants and the driving factors behind loan defaults in order to reduce the amount of credit loss experienced by the company.

This company is the largest online loan marketplace that provides personal loans, business loans, and financing for medical procedures, with a fast online interface that allows borrowers to access lower interest rates. Defaulting borrowers, labeled as "charged-off," are responsible for the largest amount of loss to the company.

# Study Objective

The company aims to use exploratory data analysis to identify strong indicators of default, which can then be used for portfolio and risk assessment purposes.



# Data Processing

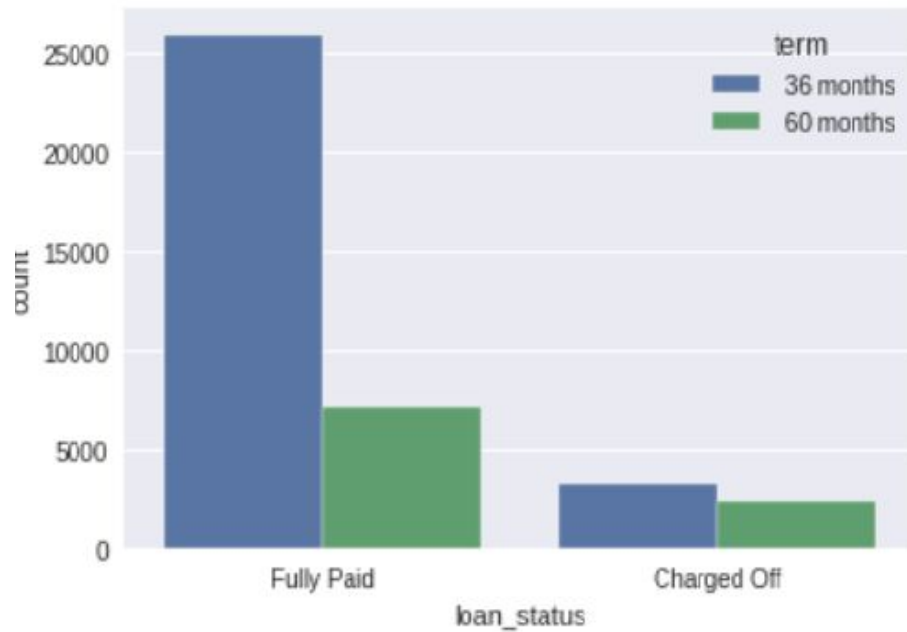
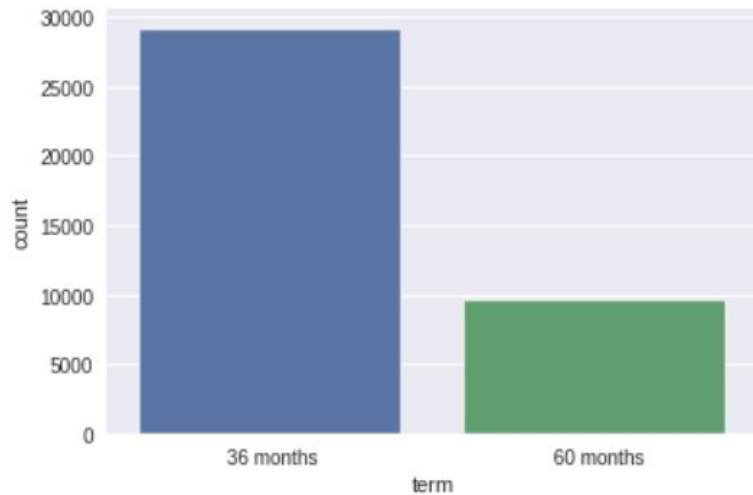
1. Data Collection : Company transformed all paper based data into excel file for doing advanced data analysis.
2. Data Preparation : In this process we removed any inconsistencies or errors, and transforming it into a format that can be analyzed.
3. Data Analysis : In this process we analyzed prepared data using various techniques such as statistical methods, machine learning algorithms, and data visualization tools. The goal is to extract insights and useful information from the data.
4. Data Dissemination : We are disseminating to stakeholders with our results so they can use over model in their business.

# Data Preparation

- There are 111 variables and 39716 application present in the given dataset.
- Any columns in the dataset that lacked more than 30% of the data were eliminated.
- Removing unnecessary columns which has “NA” values in the dataset.
- After performing above two operations we have 15 variables into the current dataset.
- We do have some missing values and we are filling these missing values from mode values which is present into the dataset.
- Checking outlier and treatment of the outlier with help of percentile method.
- Checking and remove the skewness using power transform method.

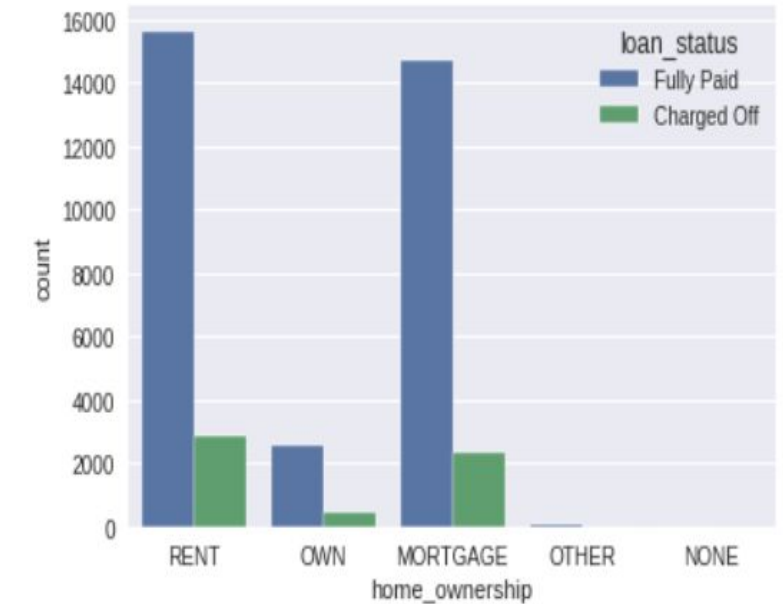
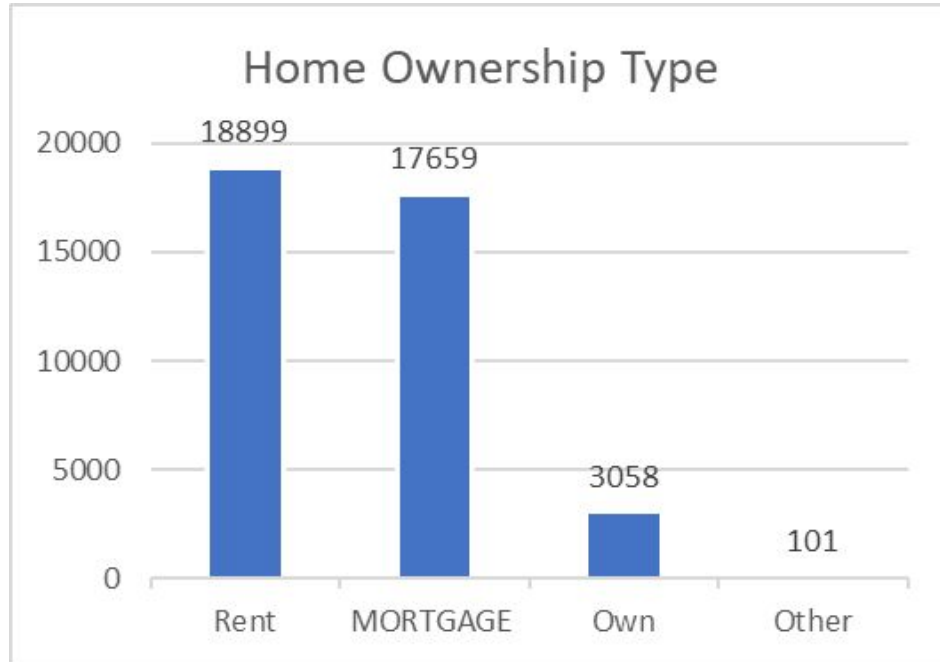
# Data Analysis - Univariate , Segmented Univariate , And Bivariat

# Loan Term

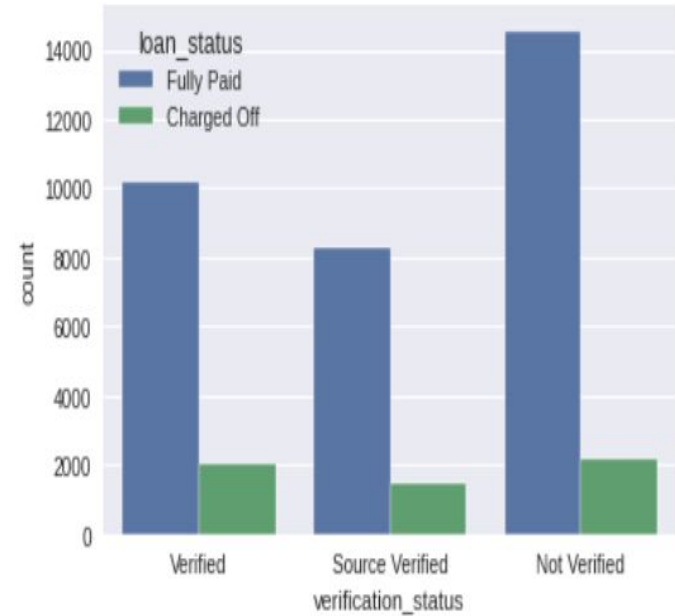
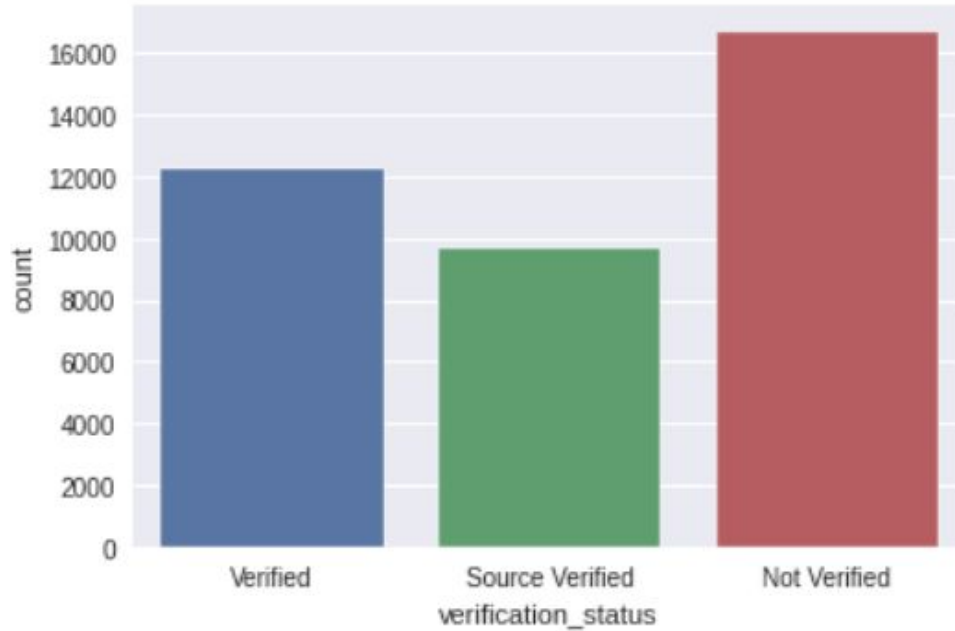




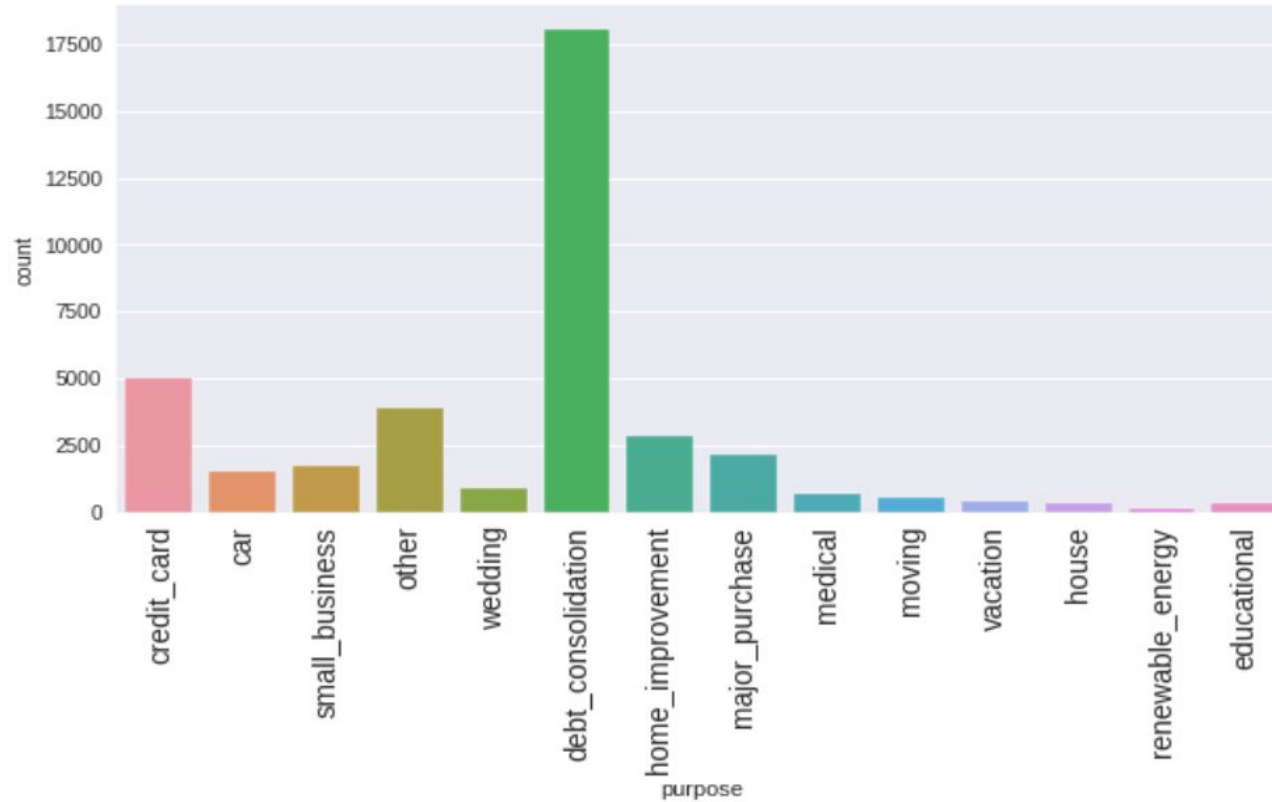
# Home Ownership Type who took loan



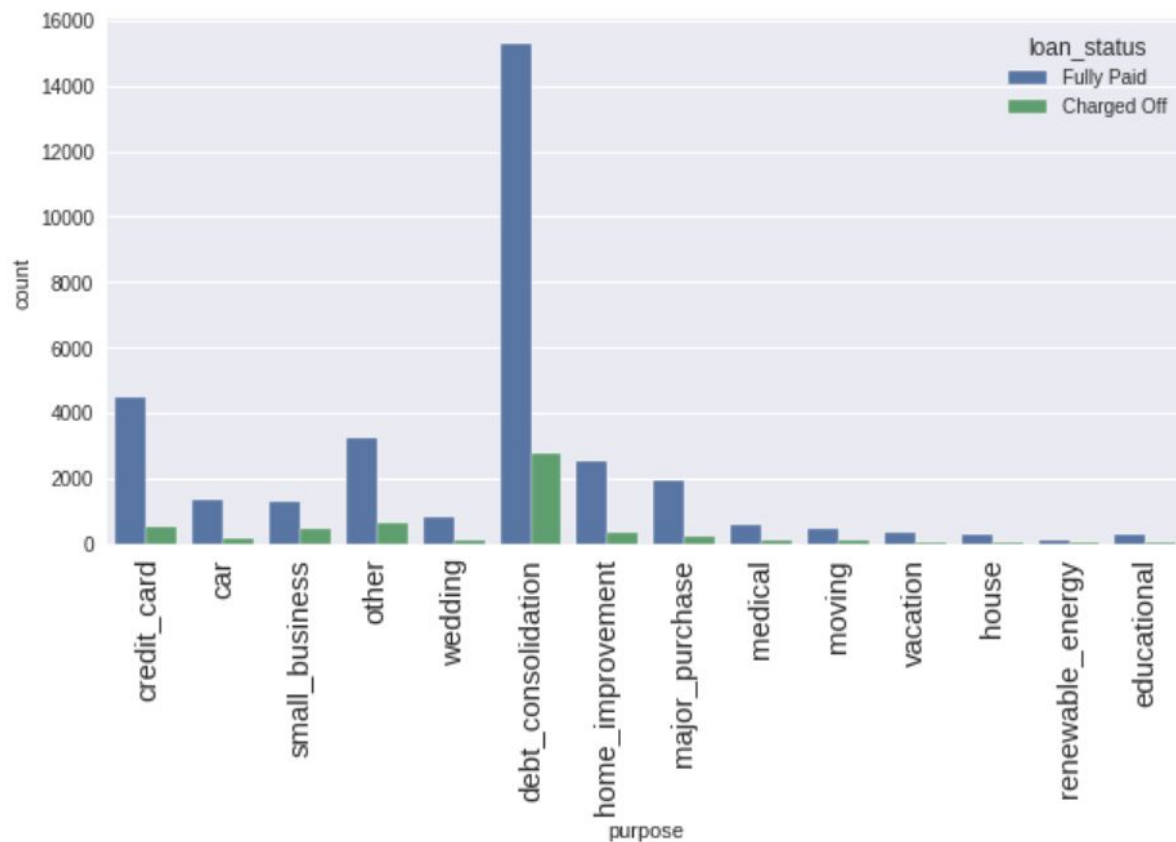
# Loan application verification



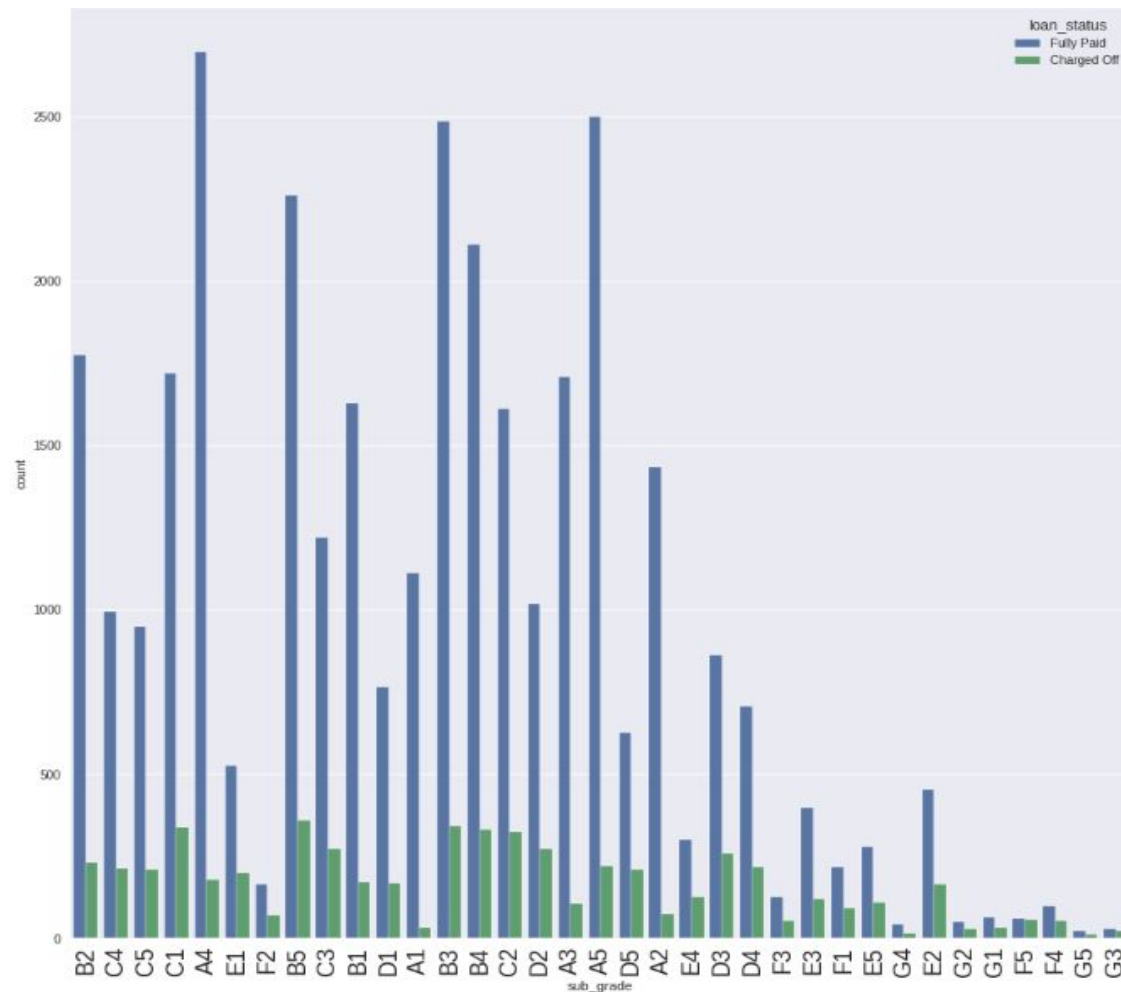
# Loan purpose



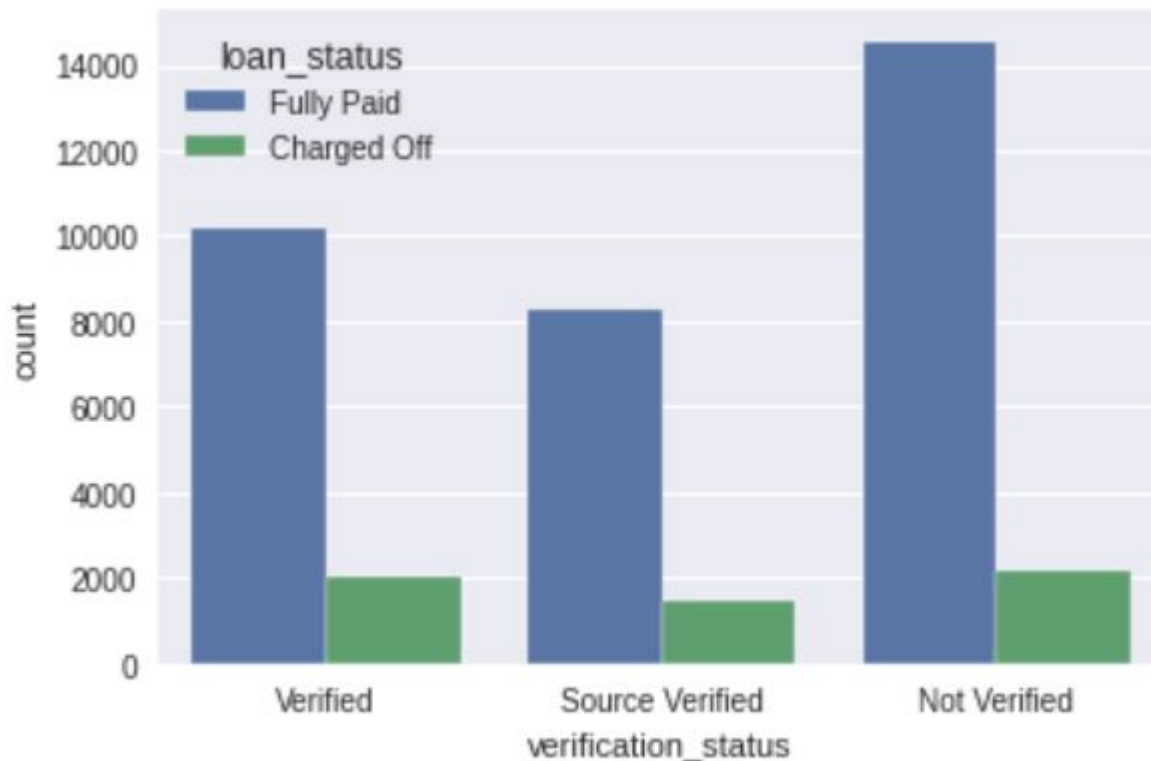
# Loan Purpose Segmented by Loan Status



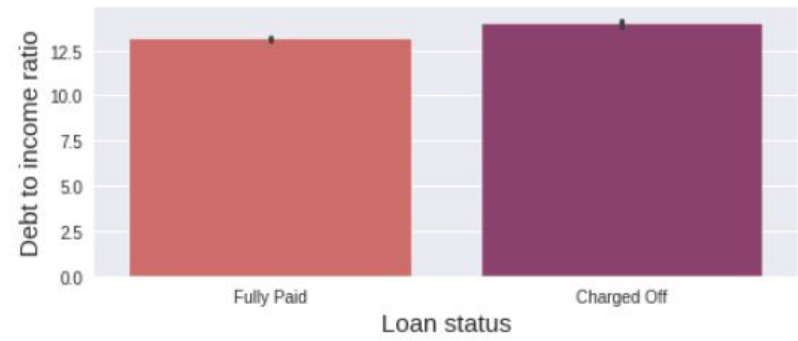
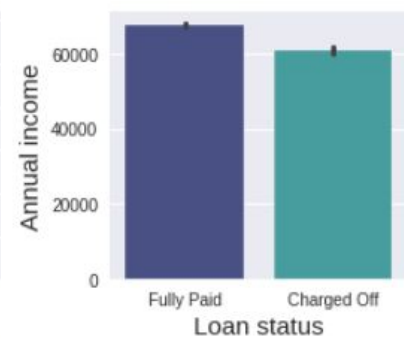
# Subgrade segmented by loan status



# Loan Verification status with respect to loan status



# Data Analysis Bivariate



Correlation Matrix in loan amount , funded amount , funded amount invested, annual income and dti

	loan_amnt	funded_amnt	funded_amnt_inv	annual_inc	dti
loan_amnt	1.000000	0.981790	0.937922	0.268999	0.062436
funded_amnt	0.981790	1.000000	0.956172	0.264798	0.062194
funded_amnt_inv	0.937922	0.956172	1.000000	0.251981	0.070663
annual_inc	0.268999	0.264798	0.251981	1.000000	-0.121530
dti	0.062436	0.062194	0.070663	-0.121530	1.000000



# Conclusion

1. It can be observed that loans with a 36-month term are much more prevalent compared to loans with a 60-month term. Additionally, when looking at the loan status, it was found that customers who have defaulted have nearly an equal ratio of both term lengths. On the other hand, clients who have fully paid their loans are predominantly associated with the 36-month term. Thus, it is likely that borrowers with a 60-month term loan are at a higher risk of defaulting.
2. The number of loan borrowers who own their home is very low, whereas those who are renting or have a mortgage are much higher in number.

## Conclusion - to be continue...

3. The verification status is a significant factor, as it is apparent that a large number of applicants are not verified. Furthermore, the proportion of charged-off applicants is similar in both the verified and not-verified categories. Therefore, modifications may be necessary in the verification process.
4. The purpose of debt consolidation accounts for almost 46% of all loans, with credit cards being the next most common purpose. Given that a significant portion of loan defaults occur in this category, it is advisable to implement more rigorous verification procedures for borrowers seeking debt consolidation loans. This can help mitigate the risk of default and improve the overall loan portfolio performance.

## Conclusion - to be continue...

5. In terms of interest rates, it has been observed that charged-off borrowers tend to have an interest rate of 9.99% or higher. Therefore, borrowers with an interest rate above this threshold are at a greater risk of being charged off.
6. Regarding loan amount, there is a symmetrical reduction in the amount of loans granted as the amount increases from 10,000 to 30,000 , with the largest number of loans being granted at 10,000. Furthermore, borrowers who have applied for loans beyond 10,000 are more likely to default on their payments.
7. Around 70% of installments paid by clients fall within the range of 100 to 400 dollars, with a sharp decline observed after the 400\$. However, it is worth noting that there is a sudden surge in installment payments around the 900\$.

## Conclusion - to be continue...

8. The debt-to-income ratio exhibits a normal distribution, ranging from 0 to 25. However, it is worth noting that charged-off borrowers tend to have a higher DTI ratio compared to fully paid borrowers.
9. The loan amount, funded amount, and funded amount by investors exhibit a strong correlation, which is evident from the correlation matrix.
10. The average annual income of charged-off borrowers is lower than that of fully paid borrowers.
11. Borrowers who are falling into category B5,B4,B3,C1,C2,D3,D3, and D4 are having high probability for charging off their loans.

# Driver Variables

- Loan term
- Home ownership
- Loan amount
- Loan Rate of Interest
- Loan application verification
- Loan purpose
- dti

# Recommendation

1. To reduce the likelihood of loans being charged off, banks should consider limiting their interest rates as there is a positive correlation between interest rates and the probability of default.
2. Borrowers who took out loans for debt consolidation have a significantly higher probability of being charged off on their loans.
3. Debt-to-Income ratio (DTI) is low, the loan verification process becomes more stringent, with lenders scrutinizing every aspect of the borrower's financial history to ensure a higher likelihood of loan repayment.
4. We need to strengthen loan application verification process.
5. Lending club needs to be rethink of how they can reframe the approach of deciding the rate of interest on loan.

# Software/ Technologies Used

- Python 3.X
- Google Colab
- numpy - 1.22.4
- pandas - 1.3.5
- missingno - 0.5.2
- seaborn - 0.11.2
- matplotlib - 3.5.3
- Windows 10
- VSCode -1.76.0

# Thanks