LENDING CLUB CASE STUDY

PALASH BARANWAL

CONTENTS

- BUSINESS UNDERSTANDING
- OBJECTIVE
- DATA CLEANING
- UNIVARIATE ANALYSIS
- SEGMENTED UNIVARIATE ANALYSIS
- BIVARIATE ANALYSIS
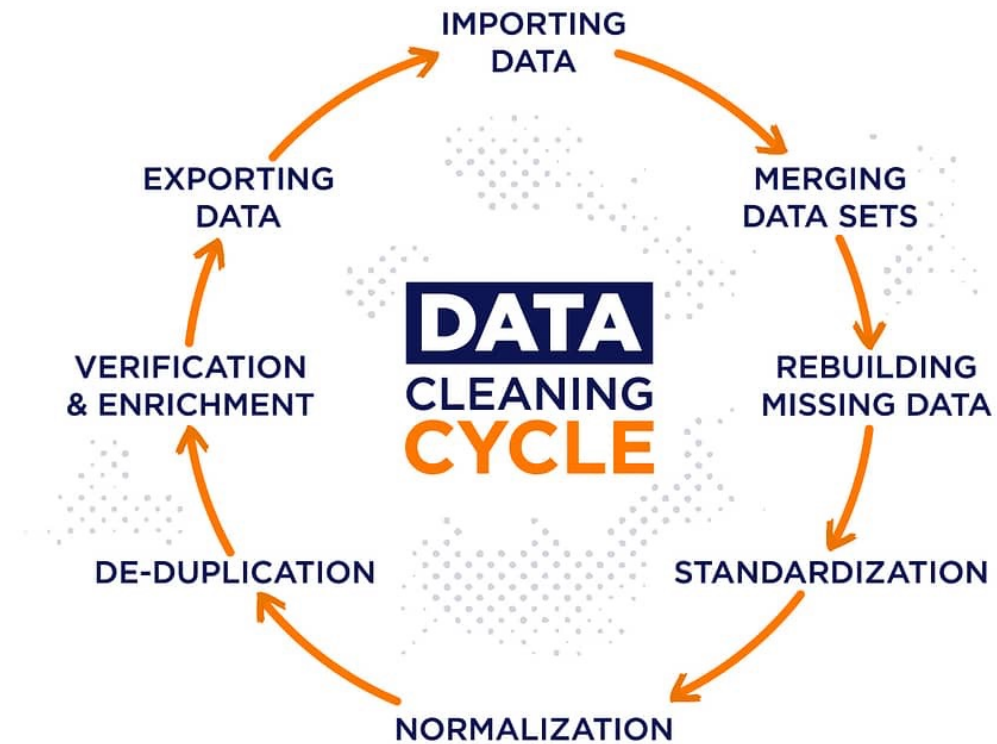- RECOMMENDATIONS

BUSINESS UNDERSTANDING

- A **Consumer finance company** which specialises 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. 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
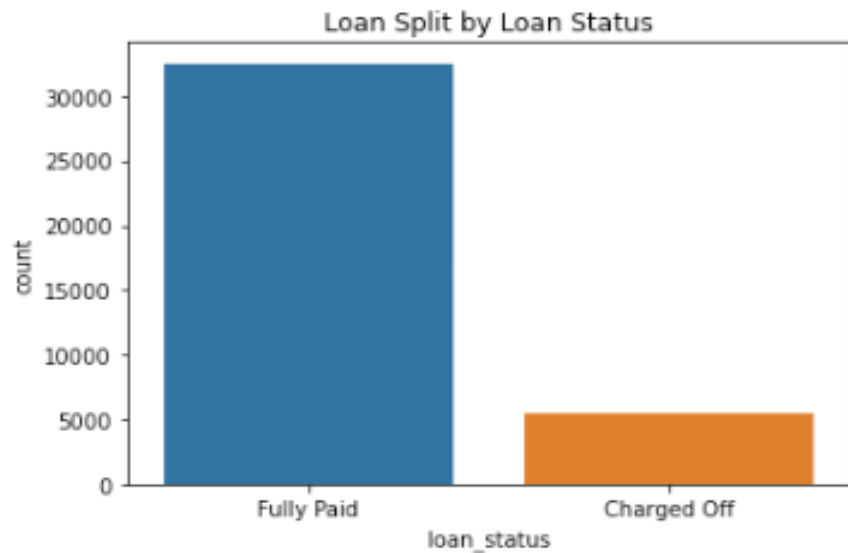
OBJECTIVE

- We will use EDA to understand how **consumer attributes** and **loan attributes** influence the tendency of default.

- We will also identify the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

DATA CLEANING

- Treating missing values
- Removing irrelevant columns
- Fixing incorrect data-types
- Removing outliers

IMPORTING DATA

MERGING DATA SETS

REBUILDING MISSING DATA

STANDARDIZATION

NORMALIZATION

DE-DUPLICATION

VERIFICATION & ENRICHMENT

EXPORTING DATA

DATA CLEANING CYCLE

UNIVARIATE ANALYSIS

Loan Split by Loan Status

- Most of the borrowers has fully paid the loans. However, ~17% of the loans are still charged-off.


Loan Split by Grades
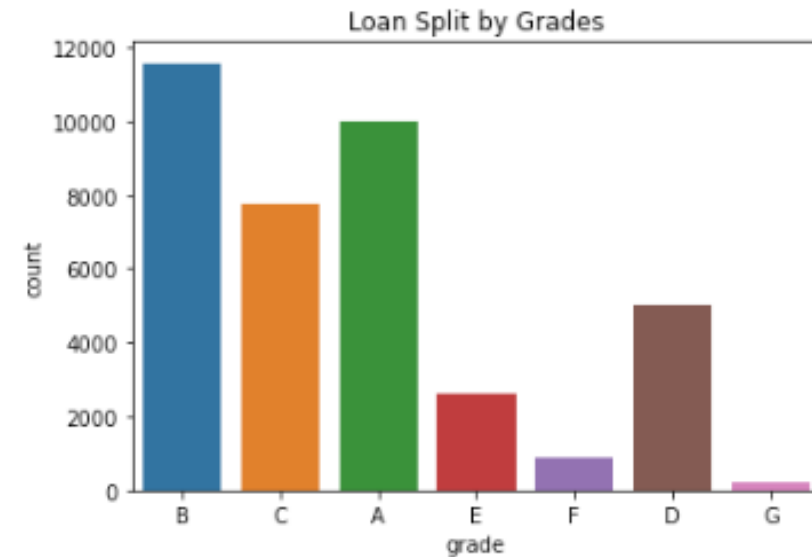
- Grades seem to have high correlation with interest rates, and are probably defined based on the categorization of interest rates. Based on #3, Higher the grade, higher the interest rate.
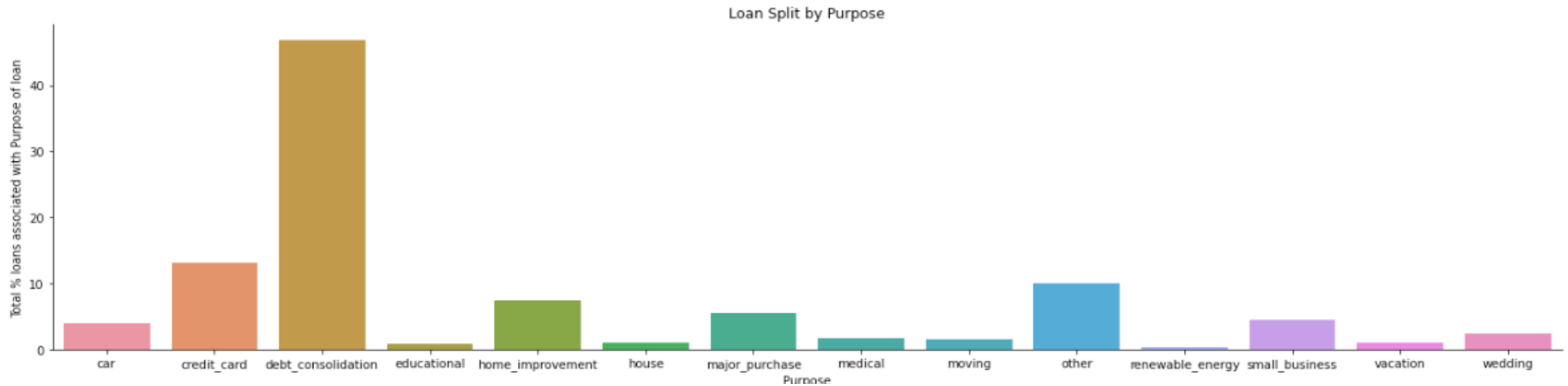
UNIVARIATE ANALYSIS
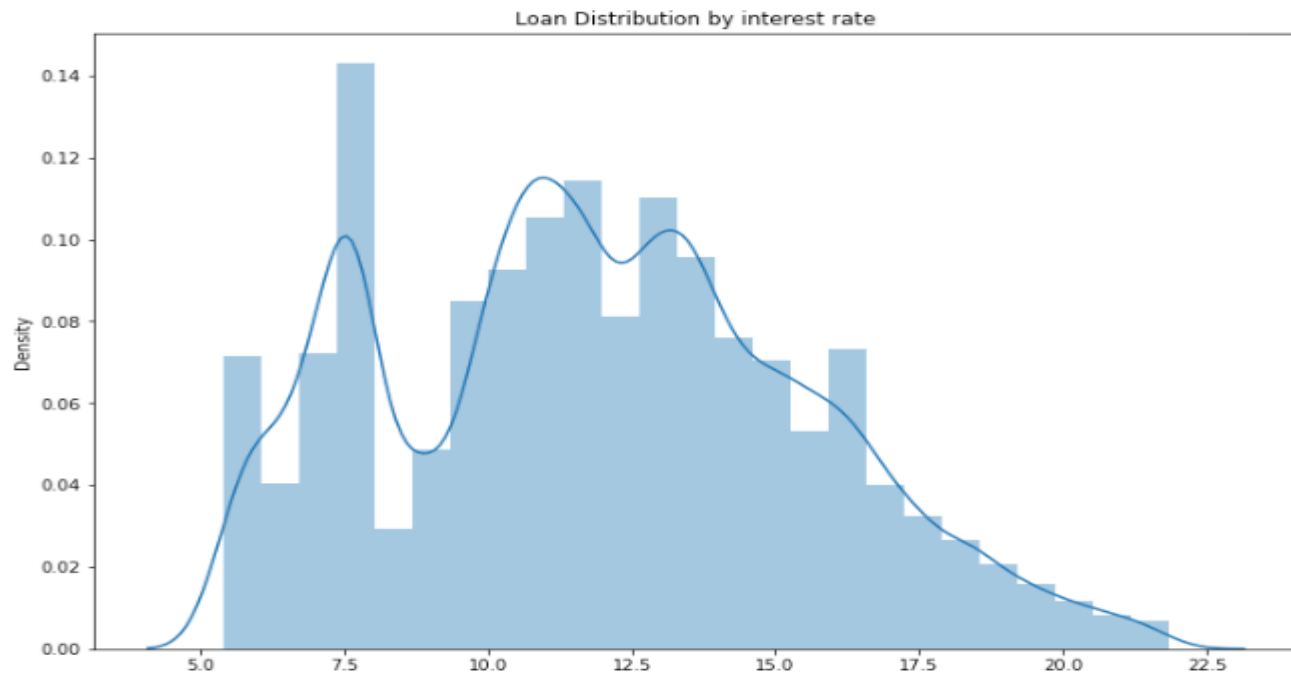
Loan Distribution by interest rate

- Loan interest varies from 5.42% to 21.82% The interest_rate of 10.99% is the most frequent interest rate, which is very close to median which is 11.71%. We also see a trend where as the loan_interest increases, the total count of loan decreases. That means, only a handful of loans being given has high interest rates.

UNIVARIATE ANALYSIS
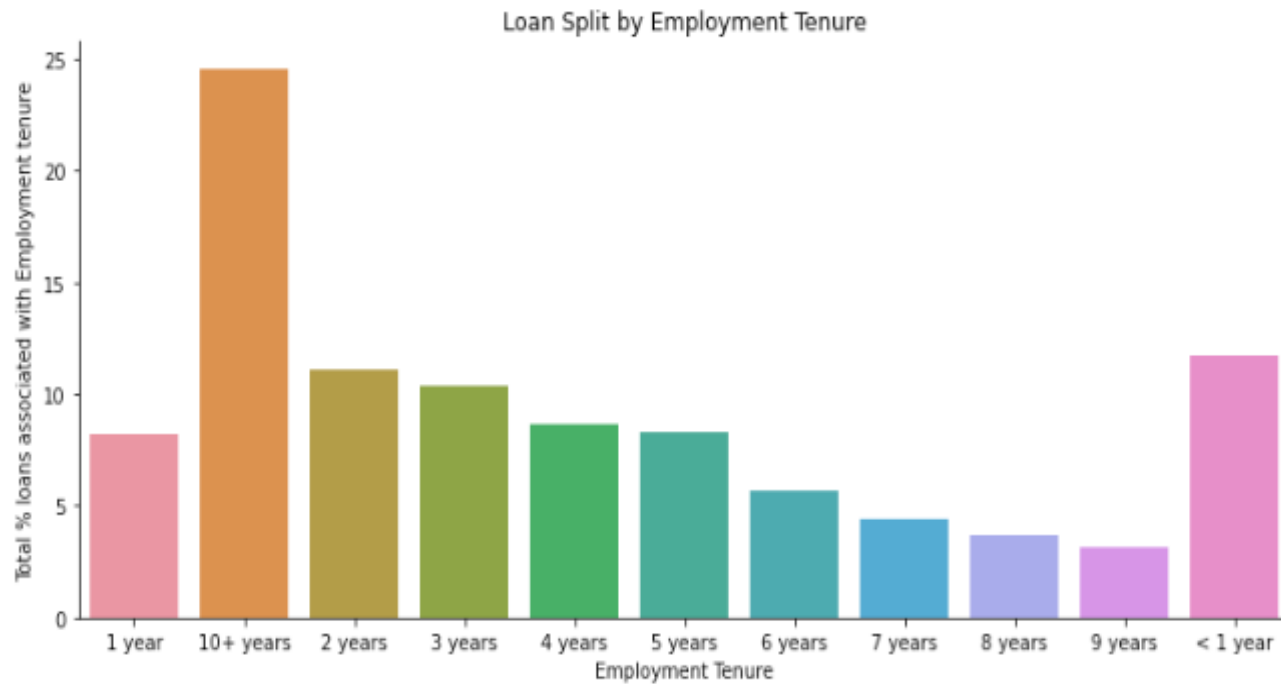
Loan Split by Employment Tenure

More than 25% of the loan is borrowed by applicants with 10+ years of experience. Also, freshers with < 1 year experience also contribute to close to 12% of total loans.

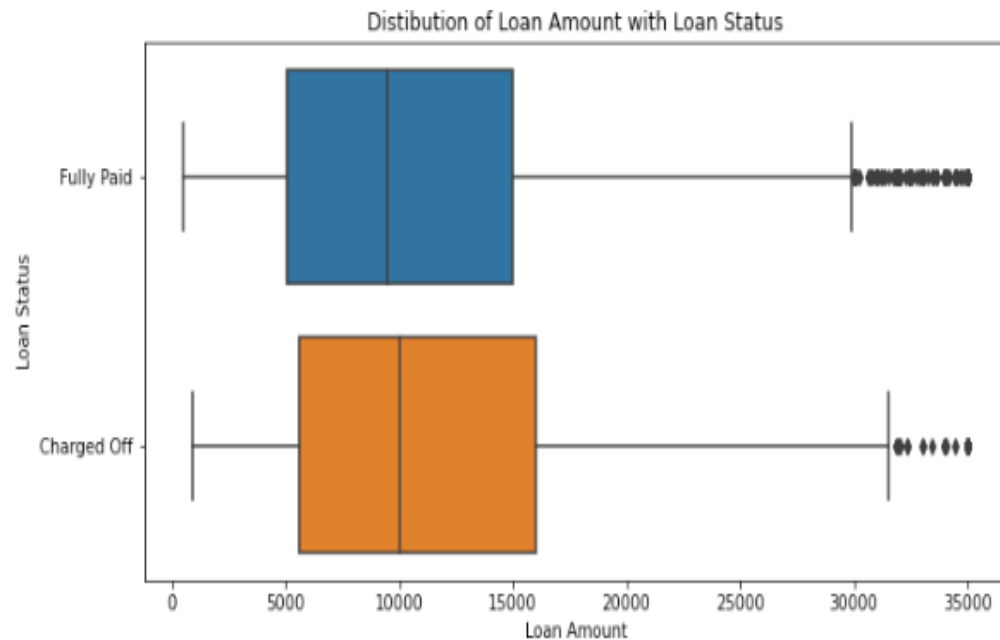# UNIVARIATE ANALYSIS

Loan Split by Purpose

- Close to 47% of loan is associated with debt-consolidation. While 13% is associated with credit-card payments

UNIVARIATE ANALYSIS - OBSERVATIONS

- Most of the borrowers has fully paid the loans. However, ~17% of the loans are still charged-off.

- Loan amount ranges from 500 to 35000. The amount of 10000 is the most frequent loan amount, which is very close to median which is 9600.

- Loan interest varies from 5.42% to 21.82% The interest_rate of 10.99% is the most frequent interest rate, which is very close to median which is 11.71%. We also see a trend where as the loan_interest increases, the total count of loan decreases. That means, only a handful of loans being given has high interest rates.

- Grades seem to have high correlation with interest rates, and are probably defined based on the categorization of interest rates. Based on #3, Higher the grade, higher the interest rate.

- More than 25% of the loan is borrowed by applicants with 10+ years of experience. Also, freshers with < 1 yeaf experience also contribute to close to 12% of total loans.

- 90% of loan is associated with people who have a Mortgage or pay rent. Mortgage and Rent have almosyt similar number of loans associated.

- 43% of loan is associated with borrowers who are not verified.

- Close to 47% of loan is associated with debt-consolidation. While 13% is associated with credit-card payments.

- We observe increase in number of loans issued throghtout 2007-2011. We can say that loan applications as well as loan approval rate is increasing.

- Number of loans increases towards end of month.

SEGMENTED UNIVARIATE ANALYSIS
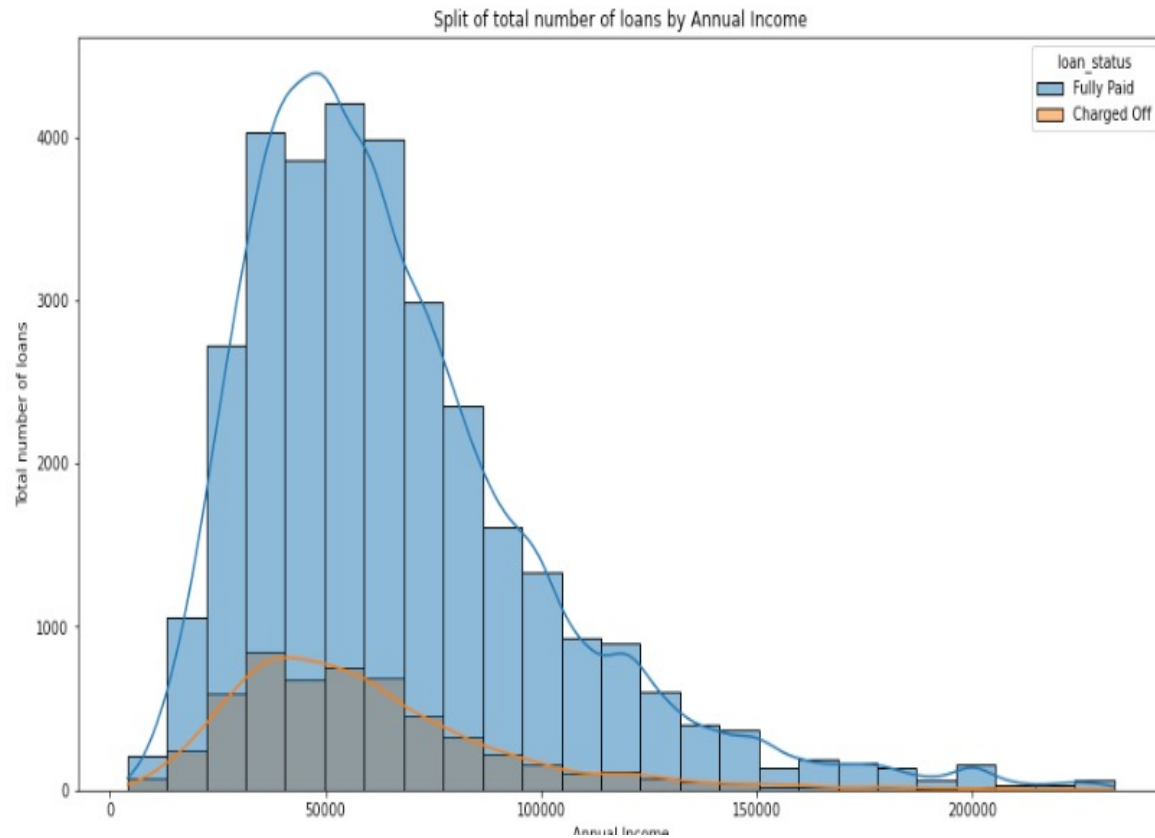
Distibution of Loan Amount with Loan Status

- Higher loan amounts have a higher chance of defaulting the loan. This is based on the 75th percentile numbers of the loan_amount.

SEGMENTED UNIVARIATE ANALYSIS
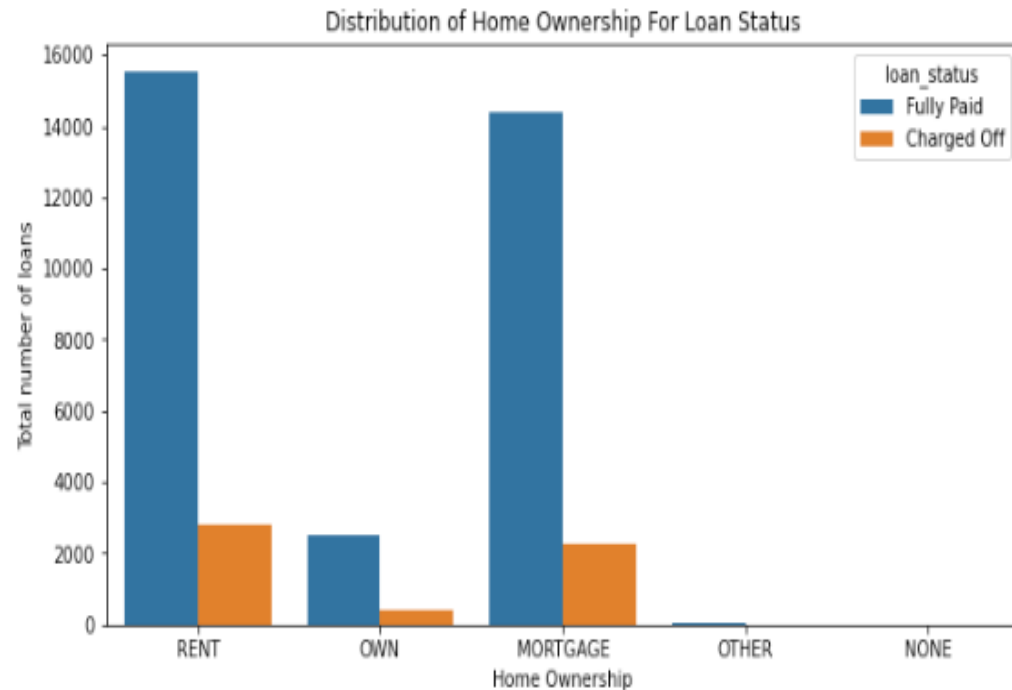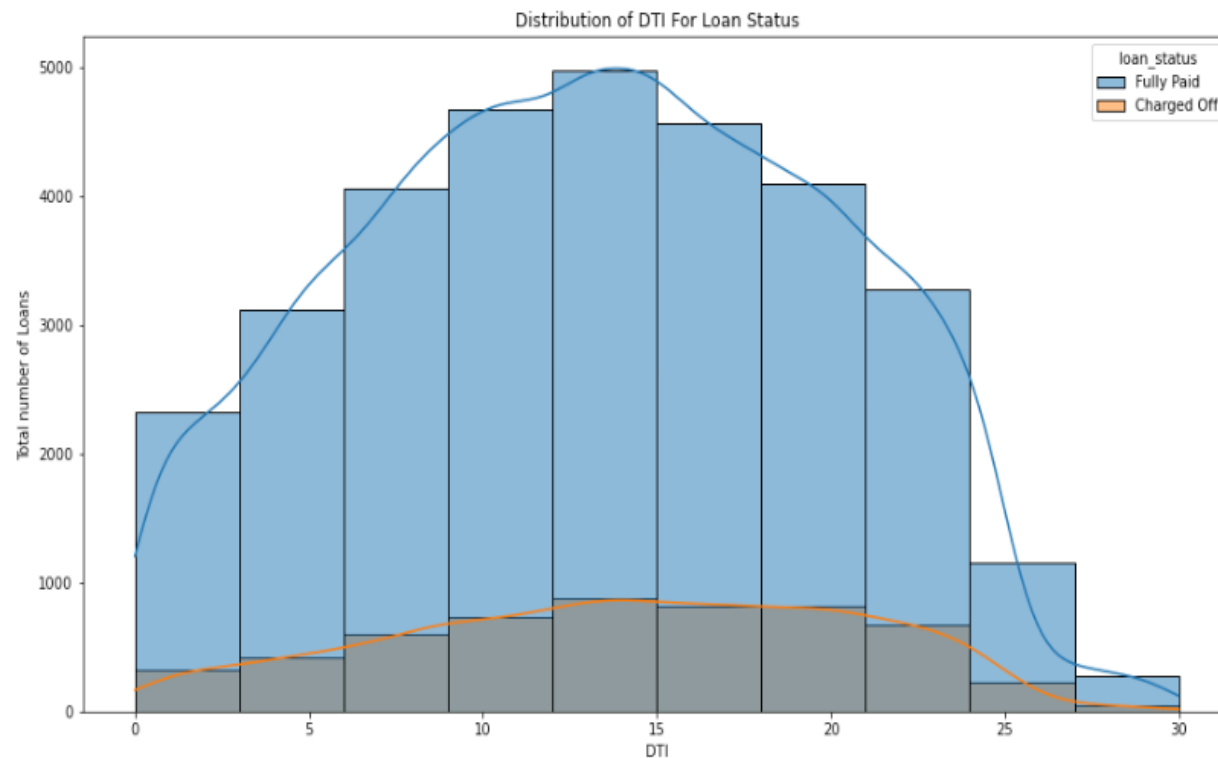

Distribution of Home Ownership For Loan Status

- Borrowers who Own a property have lesser chances of defaulting a loan. It is relatively safe to give to borrowers who Own a property.

SEGMENTED UNIVARIATE ANALYSIS

Split of total number of loans by Annual Income

Annual Income does not seem to play an important role in deciding if a loan will be defaulted. Loan Status seems to be uniformly distributed for "Full Paid" as well as "Charged off" loans

SEGMENTED UNIVARIATE ANALYSIS


Distribution of DTI For Loan Status

- Higher the DTI, higher the chances of a loan being defaulted. This is an important factor which should be considered while approving loan.
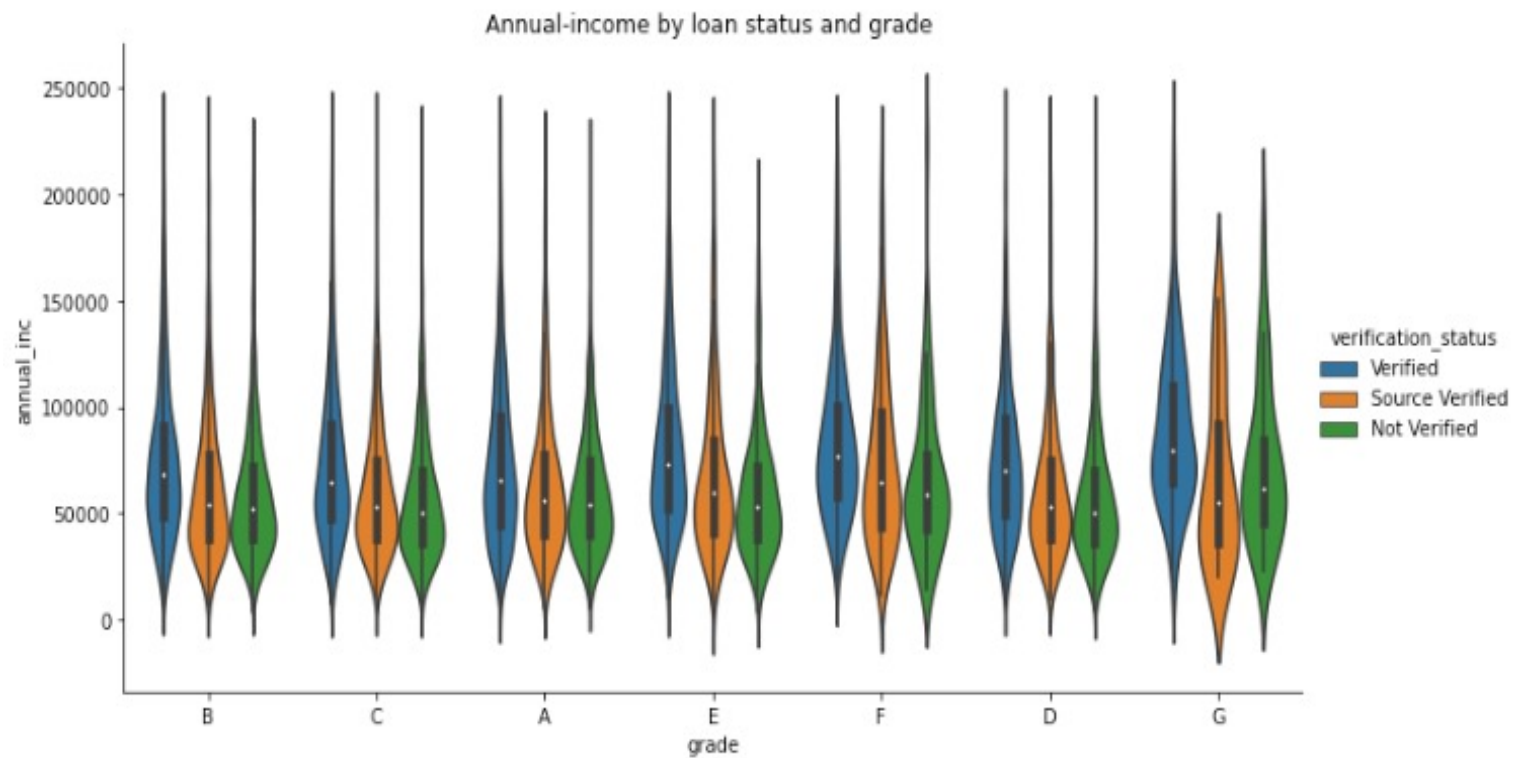
SEGMENTED UNIVARIATE ANALYSIS - OBSERVATIONS

- Higher loan amounts have a higher chance of defaulting the loan. This is based on the 75th percentile numbers of the loan_amount.

- Loan taken for a term of 60 months has higher chances of default. Close to 25% of the total loans taken for 60 months default.

- Borrowers who Own a property have lesser chances of defaulting a loan. It is relatively safe to give to borrowers who Own a property.

- Loan Status seems to be agnostic of the loan purpose and employment tenure. Because 47% of loan is to repay consolidated-debt, hence this bucket will have maximum number of defaulters and full-paid borrowers.

- Annual Income does not seem to play an important role in deciding if a loan will be defaulted. Loan Status seems to be uniformly distributed for "Full Paid" as well as "Charged off" loans

- With increase in interest_rates the percentage of loan defaults are increasing.

- Higher the DTI, higher the chances of a loan being defaulted. This is an important factor which should be considered while approving loan.

BIVARIATE ANALYSIS

Validate correlation between interest rate and grade
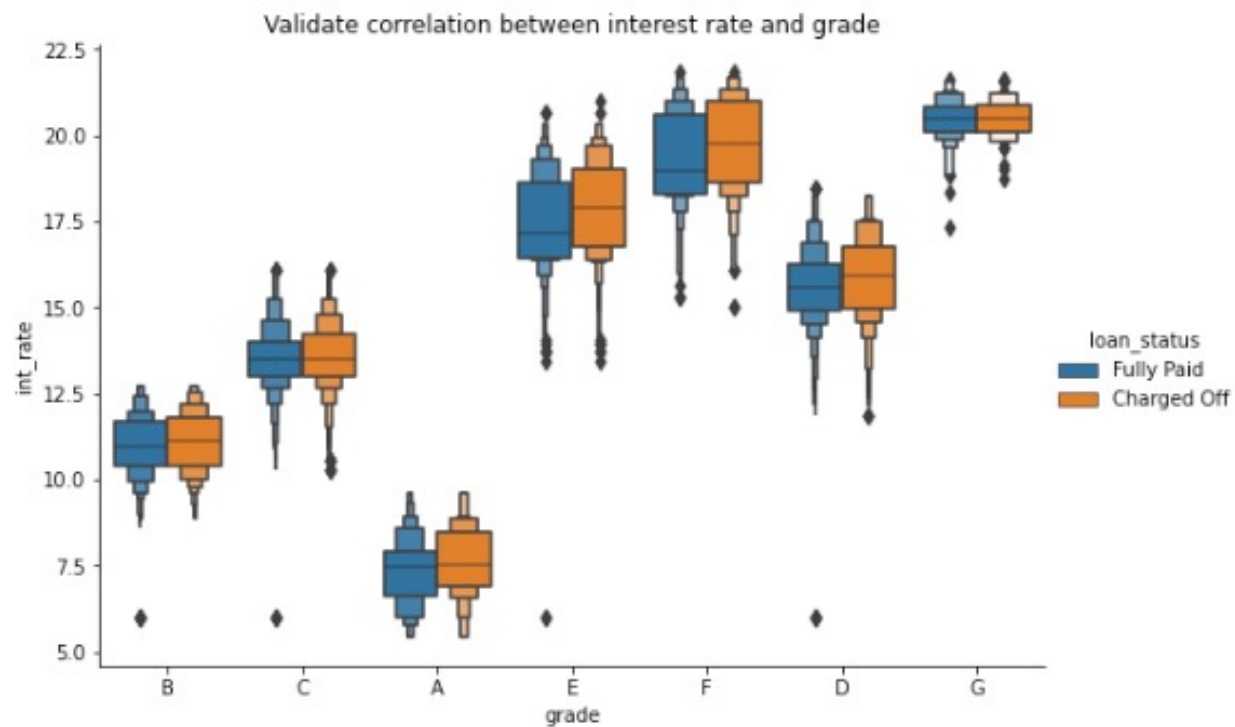
- Interest Rate and Grade are highly correlated with each other.

BIVARIATE ANALYSIS

Annual-income by loan status and grade

- Representation of annual income based on grade and verification status

BIVARIATE ANALYSIS - OBSERVATIONS

- We have a good correlation between installment and loan_amount.

- Interest Rate and Grade are highly correlated with each other.

- Borrowers with a rented property, have high occurrences of defaulting the loan.

- A positive public record bankruptcy denotes higher chances of a loan default.

DRIVING FACTORS BEHIND LOAN DEFAULT

- There are a number of attributes which have a direct impact on the loan being Charged-Off. Some of the key atributes are:

- Loan Amount - We observed that higher the loan amount, higher the chances of a loan being charged-off.

- Term - The loan term plays an important role. If loan is given for a term of 60 months, then it has higher chances of being charged-off.

- Home Ownership - Borrowers who are associated with a rented property have higher chances of defaulting a loan

- Public record bankruptcy - Borrowers with positive bankruptcy record have higher chances of defaulting a loan

- Grade - Borrowers with grades higher than F have higher chances of defaulting a loan

- Interest Rate - Higher interest rate is a strong factor of loan default. It is to be noted that grades and interest rates are closely correlated.

To summarize, for any loan application we should genuinely study below 5 attributes:

Loan Amount, Term of Loan, Home Ownership, Public bankruptcy records and Interest rates.