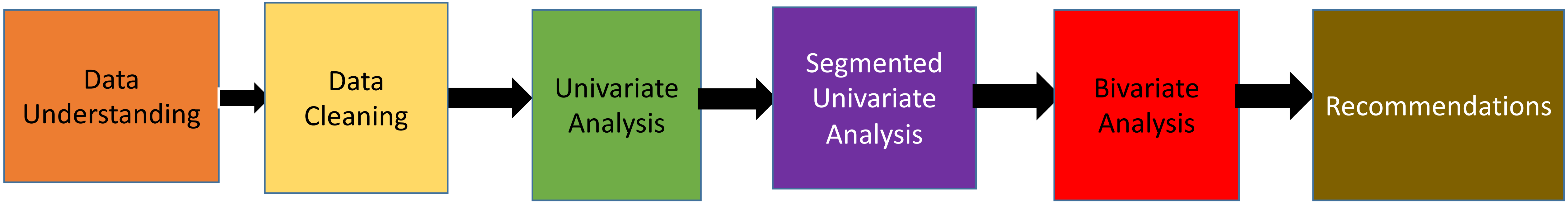
LENDING CLUB CASE STUDY

Pooja Ravi Shankar

Background

- Lending club is a consumer finance company connecting borrowers to lenders, facilitating personal loans, business loans, credit card loans, debt consolidation loans etc
- Borrowers can easily access lower interest rate loans through the online interface.
- The business objective is to identify the driver variables which are strong indicators of defaulting of loan and potentially use the insights in the decision making (approval or rejection of the loan application)

Workflow


Data Understanding
Data Cleaning
Univariate Analysis
Segmented Univariate Analysis
Bivariate Analysis
Recommendations

Problem solving Methodology

Data Understanding → **Data Cleaning** → **Univariate Analysis**

Understanding the different types of data variables: demographic, Loan characteristics and customer Behavior variables. Understand the shape: #columns and #rows, data type of each column, and how the values are distributed in each column

Remove columns having more than 90 percent missing/null values, remove rows having majority null values. Removing the unnecessary columns which don't affect the defaulting of loan(customer behavior variables, zip code, desc etc.)

Analyzing default rates for each of the categorical variables (grade, term, verification status etc.) and numerical variables(loan amount, annual income, interest rate etc) by binning them into low/medium/high categories.

Problem solving Methodology

Segmented Univariate Analysis → Bivariate Analysis → Recommendations

Segmented Univariate Analysis

Analyzing default rates across loan amount, annual income , interest rate all segmented by purpose.

Bivariate Analysis

Analyzing the correlation between different numerical variables (loan amount, dti, int rate, installment, funded amount etc.) for a given loan status.

Recommendations

Recommending the top 5 driver variables/ predictors which affect the default rate the most or which defaults the most.

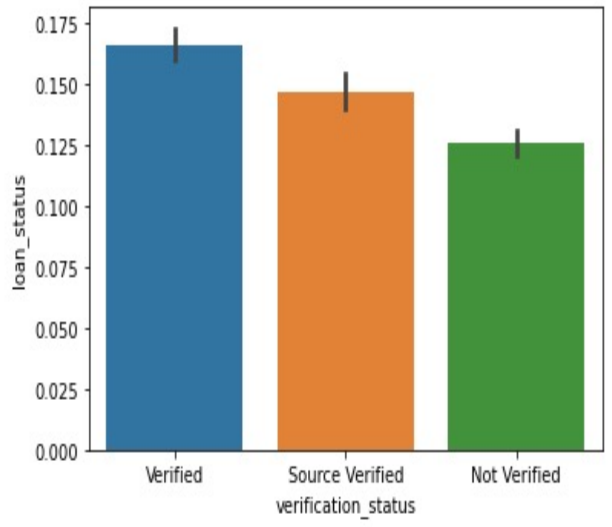
Data Understanding

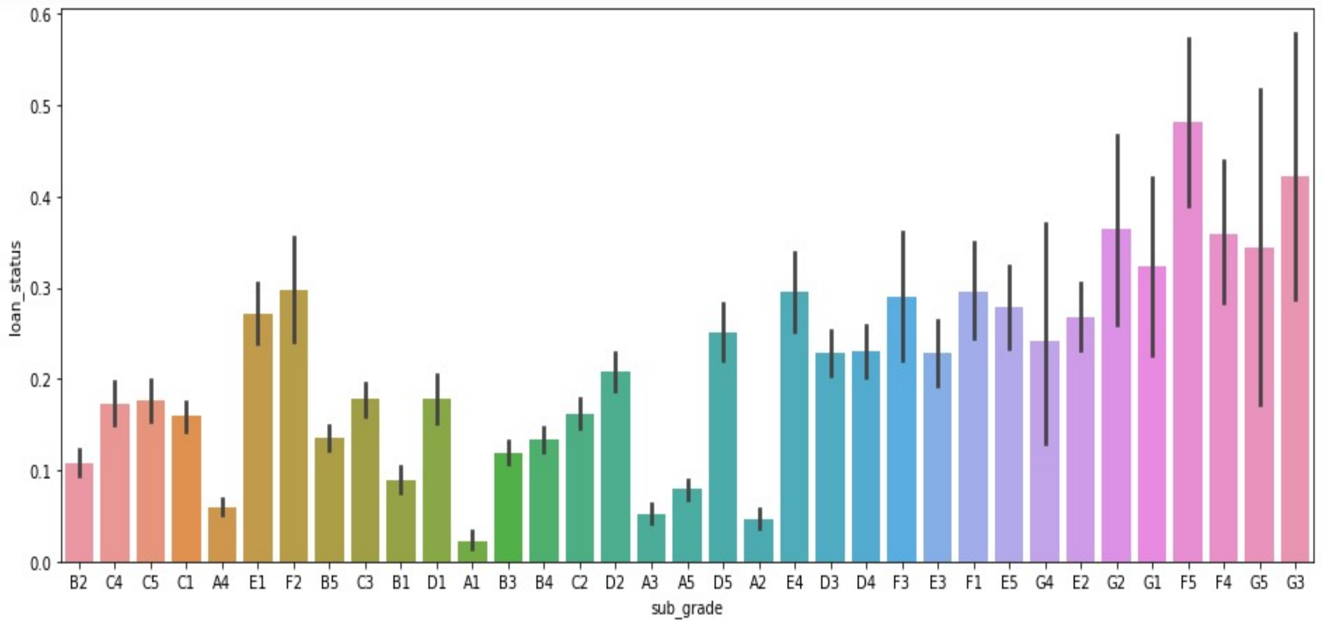
Types of variables:

- Customer Demographic
- Loan characteristics
- Customer behaviour (Generated after the loan is approved)

Customer Demographics	Loan Characteristics	Customer Behavior variables
Employment title	Loan amount	Recoveries
Employment length	Funded amount	Revolving Balance
Annual Incomes	Funded amount investment	Earliest credit deadline
Description	Interest Rate	Delinquency year-2
Zip code	Loan Status	Application type
	Loan Grade	Collection recovery fee

Data Analysis



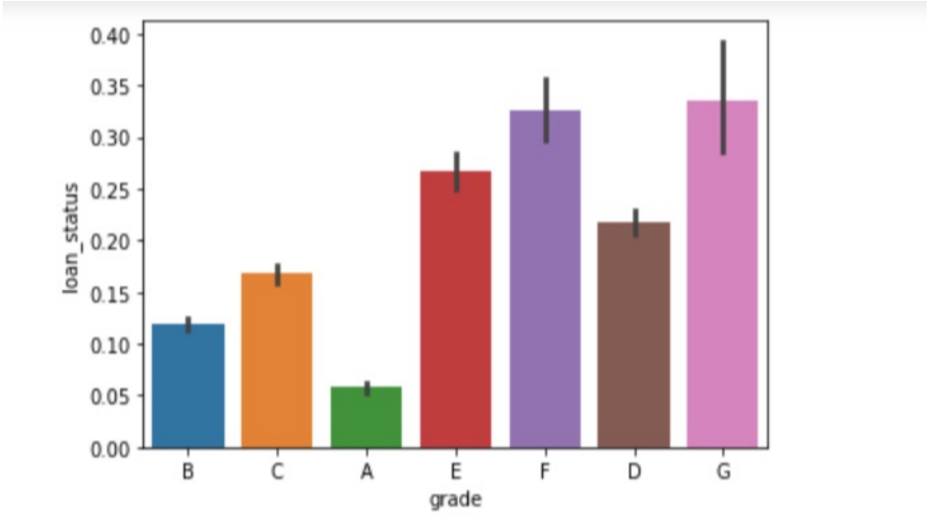


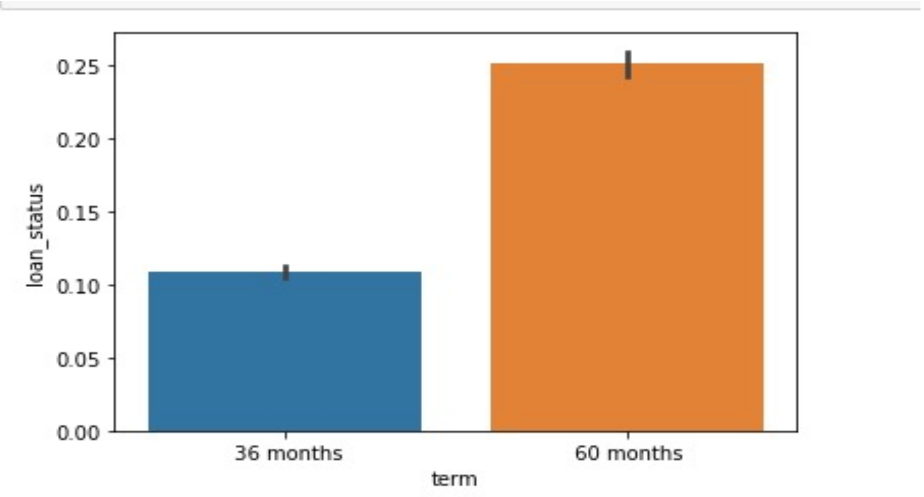
On analyzing the default rates for different categories of verification status, we realize that, Surprisingly the verified loans default more than the not verified ones.

On analyzing the default rates for various subgrade categories, we realize that, A1 defaults less than A2 and A2 defaults less than A3 and so on.

Data Analysis

- Overall Default rate is 14percent.

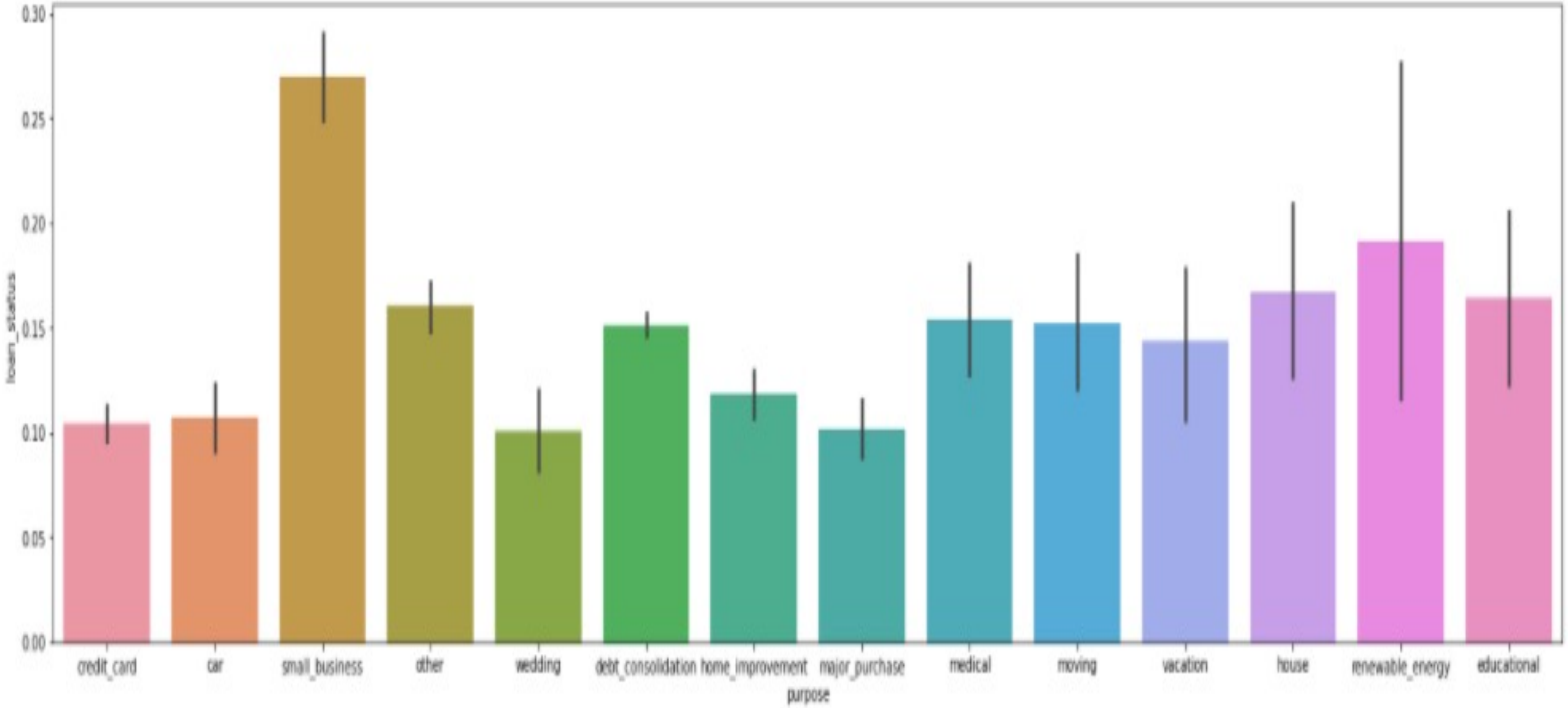


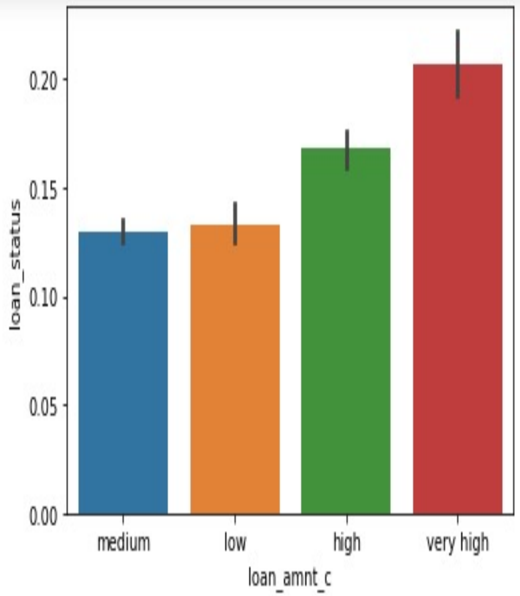


On analyzing the default rates for various grade categories, we realize that the Default rate is higher for grade G compared to other grades. Clearly as the grade of the loan moves from A to G, the default rate increases. This is expected because grade is decided by the Lending club based on the riskiness of the loan.

On analyzing the default rates for the term categories, we realize that, There are only 2 loan terms : 36 months and 60 months. Default rate is higher for 60months than 36months. 60 months loans defaults more than 36months.

Data Analysis

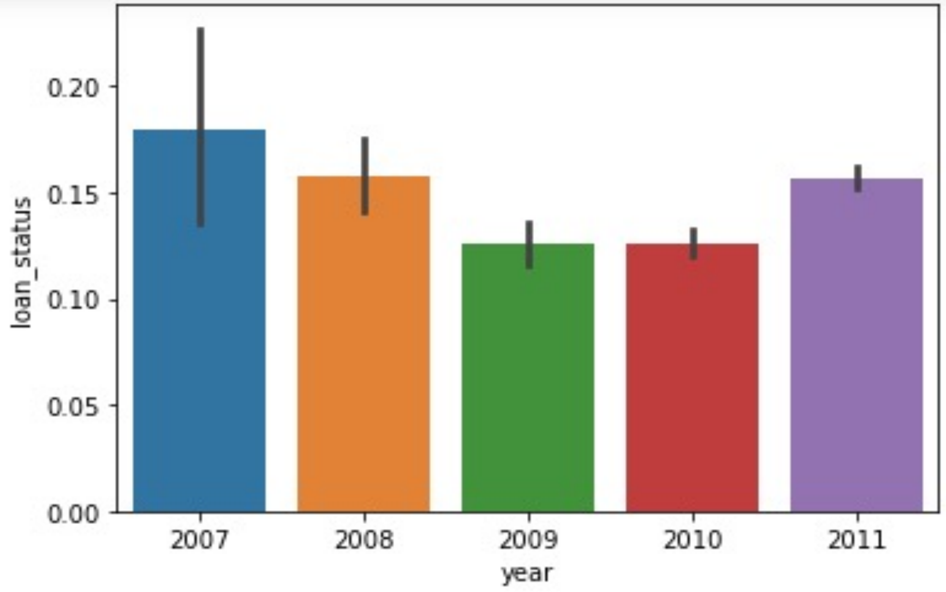


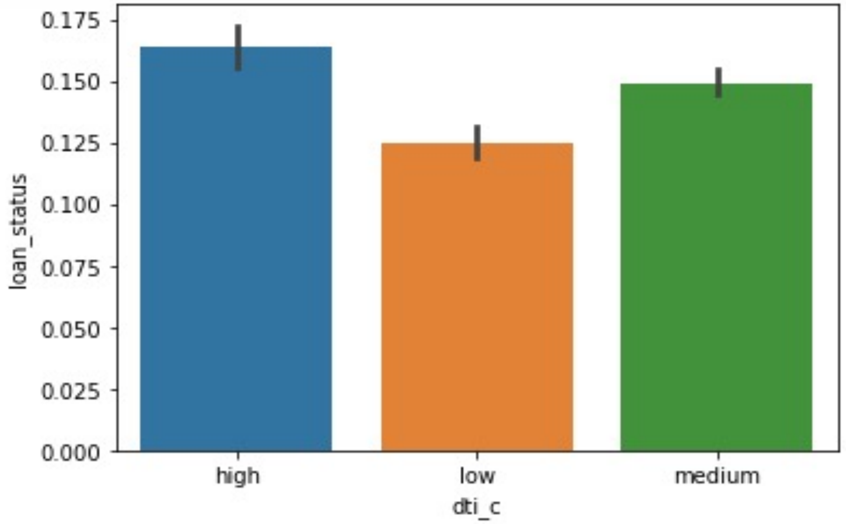


On analyzing the default rates for various purposes of applying the loan, we realize that the Small Business loans defaults the most and then renewable energy.

On analyzing the default rates for the different category of loan amounts, we realize that, 'Very High' loan amounts defaults the most.

Data Analysis

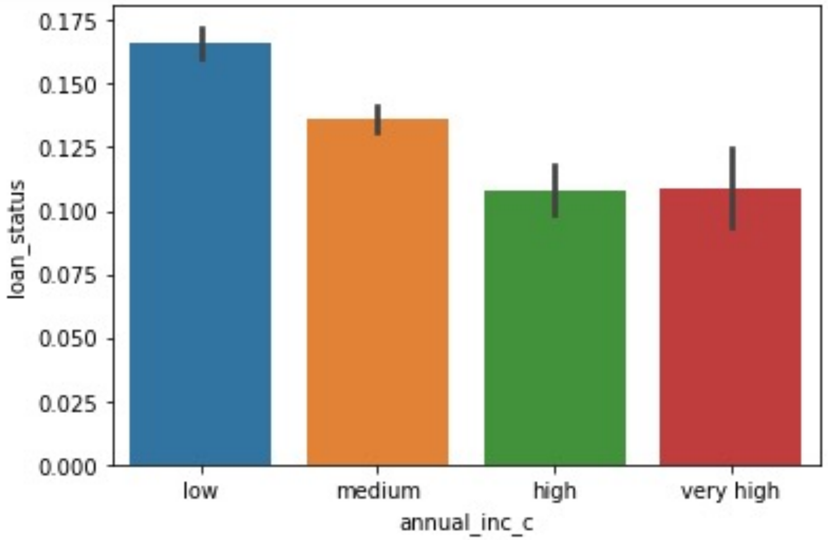


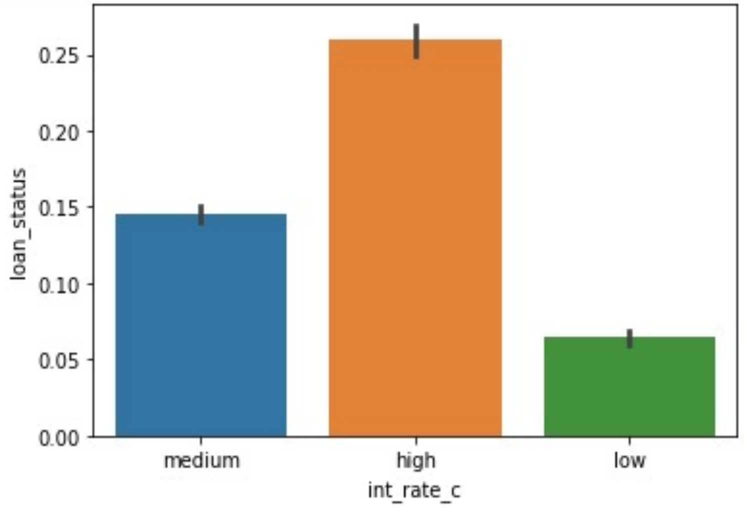


On analyzing the default rates year wise, we realize that the Default rate had suddenly increased in 2011, despite reducing from 2008 to 2010.

On analyzing the default rates for different DTI categories, we realize that High DTI (Debt to income ratio) translates to higher default rates.

Data Analysis

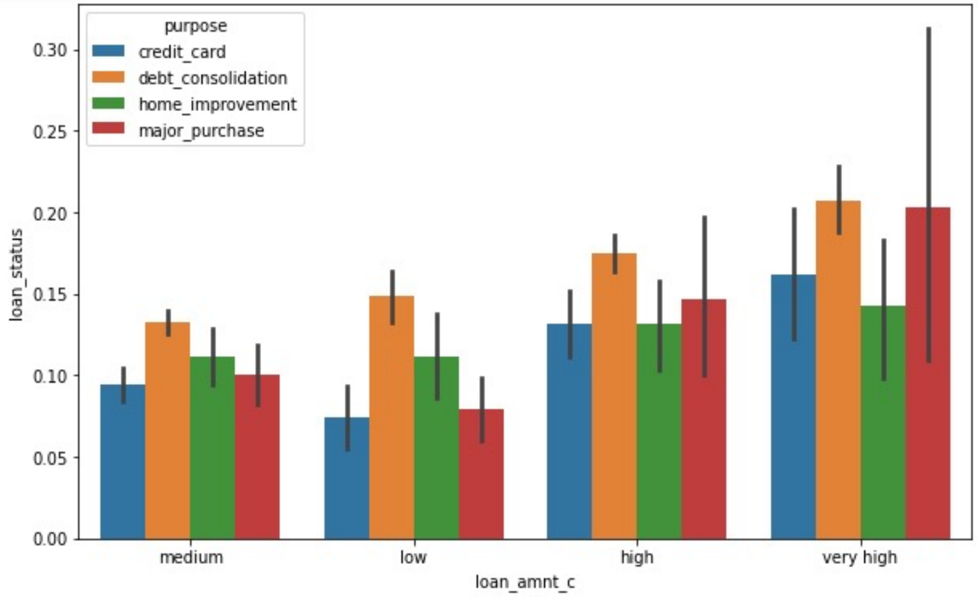


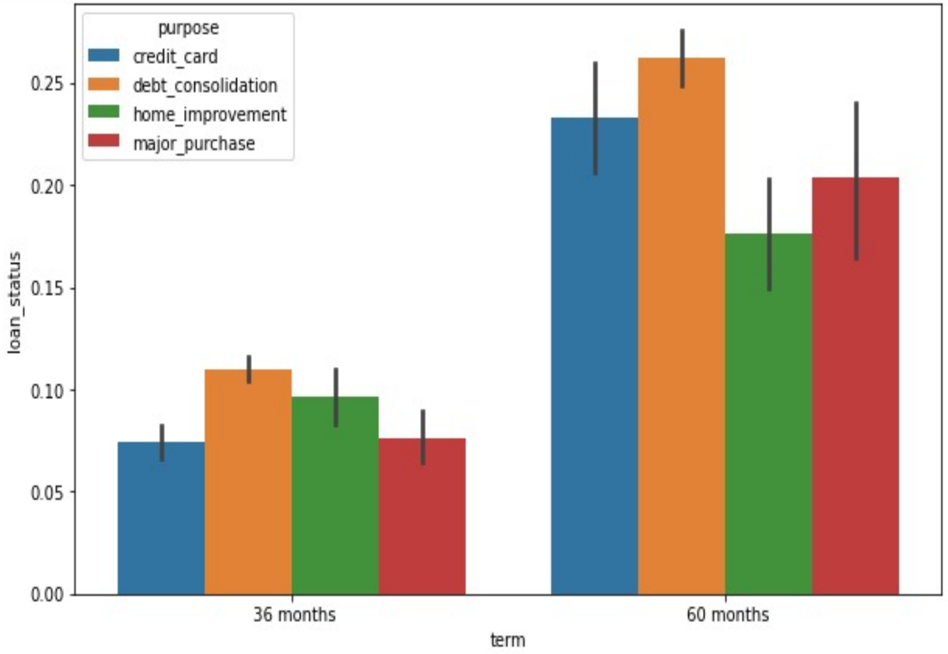


On analyzing the default rates for various annual income categories, we realize that the lower annual income category defaults the most

On analyzing the default rates for different interest rate categories, we realize that Higher interest rate category defaults the most.

Data Analysis

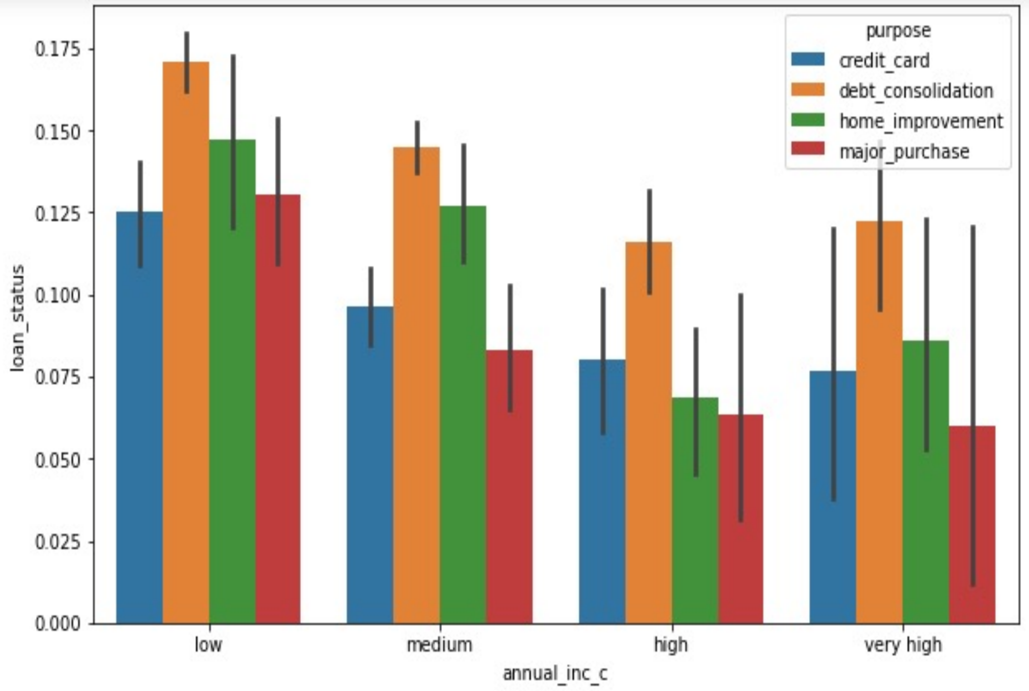


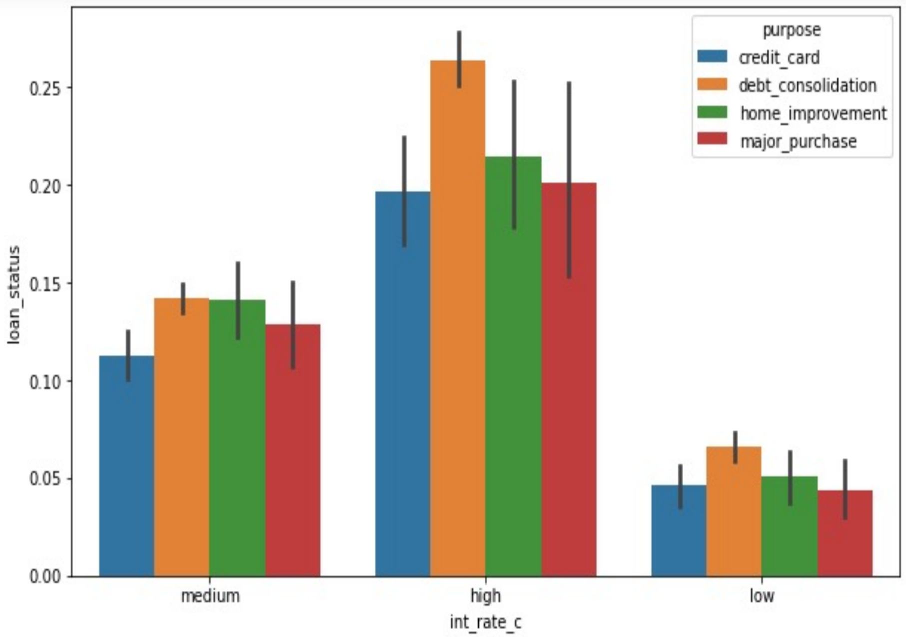


On plotting default rates across different loan amount categories, Segmented by purpose types, we realize that debt consolidation loans defaults the most.

On plotting default rates across different Term categories, Segmented by purpose types, we realize that debt consolidation loans defaults the most.

Data Analysis

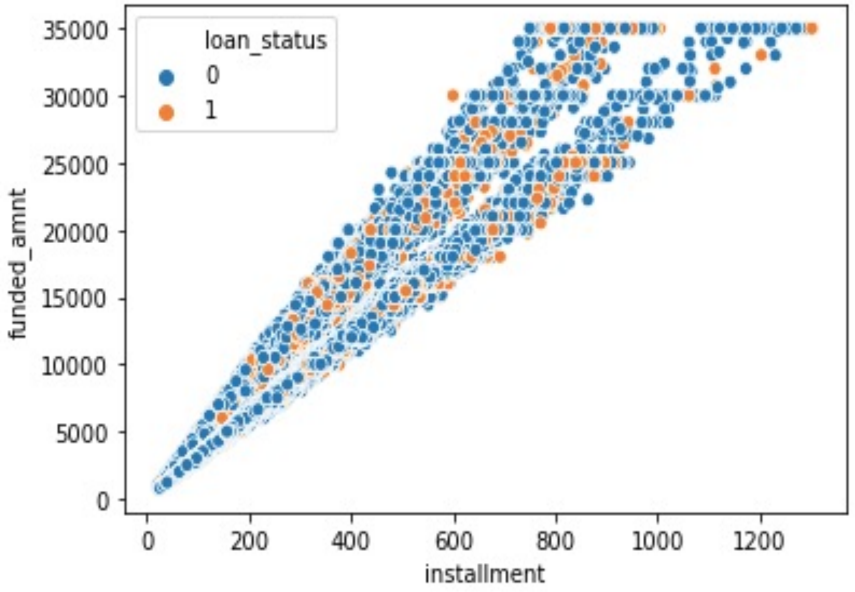


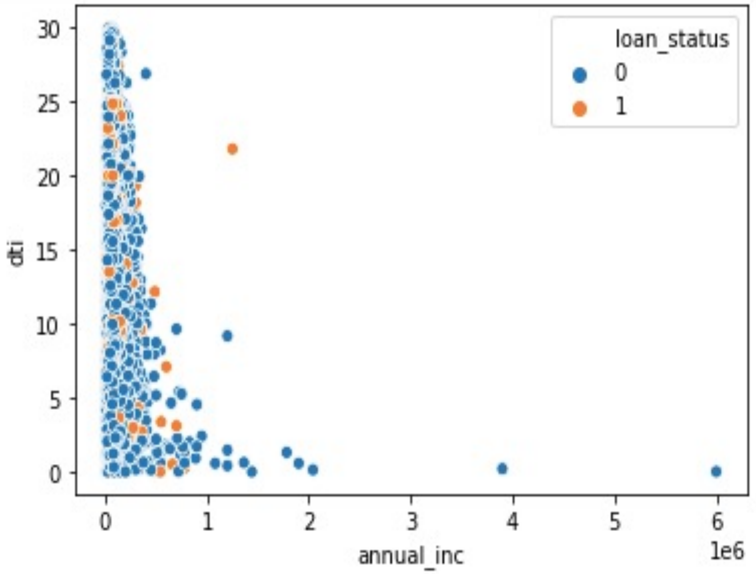


On plotting default rates across different Annual income categories, Segmented by purpose types, we realize that debt consolidation loans defaults the most.

On plotting default rates across different Interest rate categories, Segmented by purpose types, we realize that debt consolidation loans defaults the most.

Data Analysis

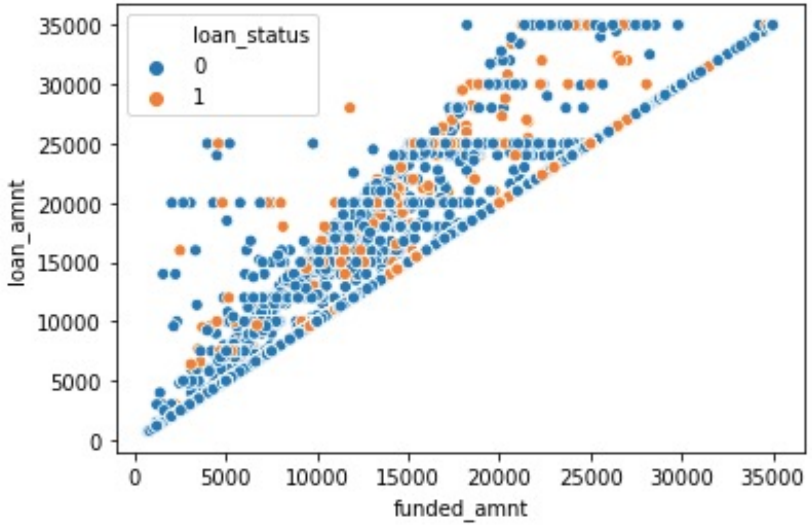


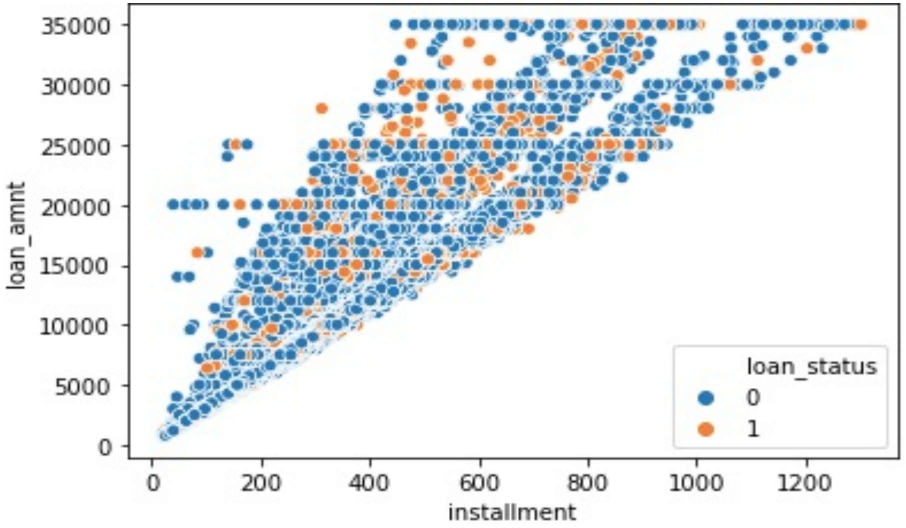


On analyzing the relationship between the numerical variables : funded amount and installment for given loan status, we realize that they are positively correlated.

On analyzing the relationship between the numerical variables : dti and annual income for given loan status, we realize that they are negatively correlated.

Data Analysis

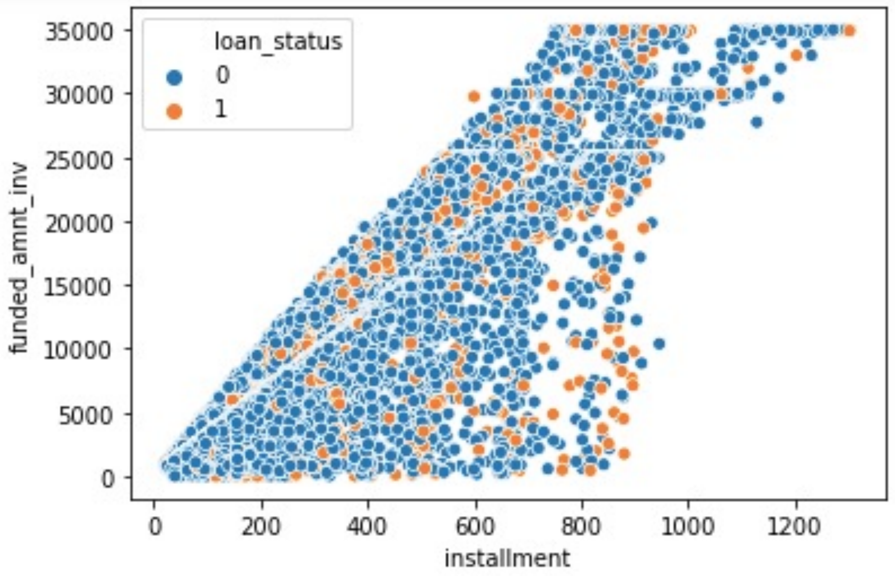


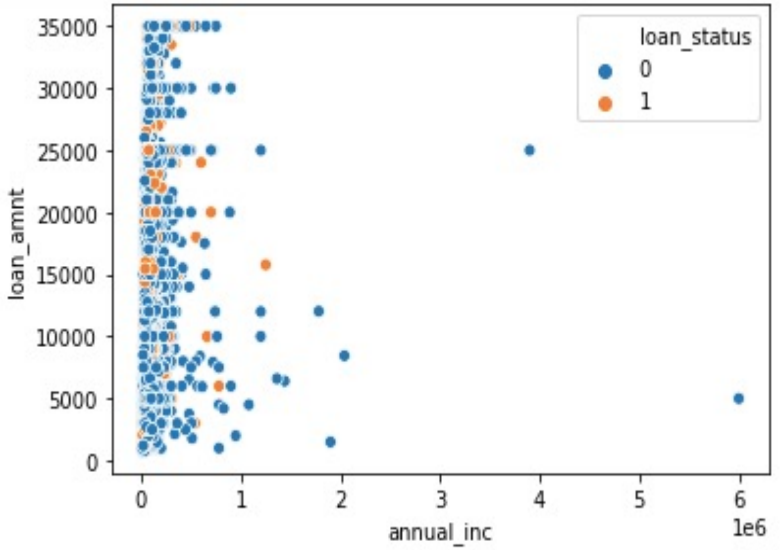


On analyzing the relationship between the numerical variables : loan amount and funded amount for given loan status, we realize that they are positively correlated.

On analyzing the relationship between the numerical variables : loan amount and installment for given loan status, we realize that they are positively correlated.

Data Analysis





On analyzing the relationship between the numerical variables : funded amount investment and installment for given loan status, we realize that they are positively correlated.

On analyzing the relationship between the numerical variables : loan amount and Annual income for given loan status, we realize that they are not correlated.

Recommendations

Top 5 driver variables/predictors that affect the default rate the most :

- Employee title (emp_title)
- Sub grade (sub_grade)
- Grade (grade)
- Interest rate (Categorized into High/medium/low) (int_rate_c)
- Home ownership (home_ownership)