Lending Club Case Study

Group Members:

Nasaraiah Vemula

Mehul Kumawat

Abstract

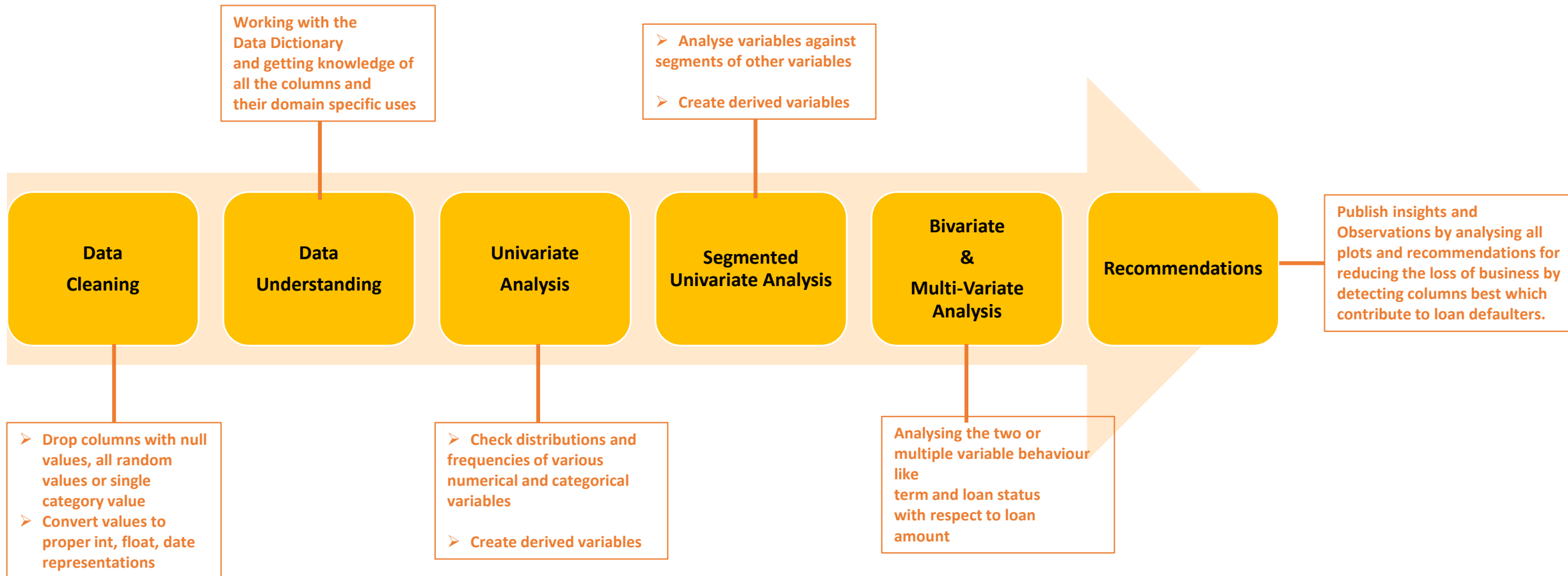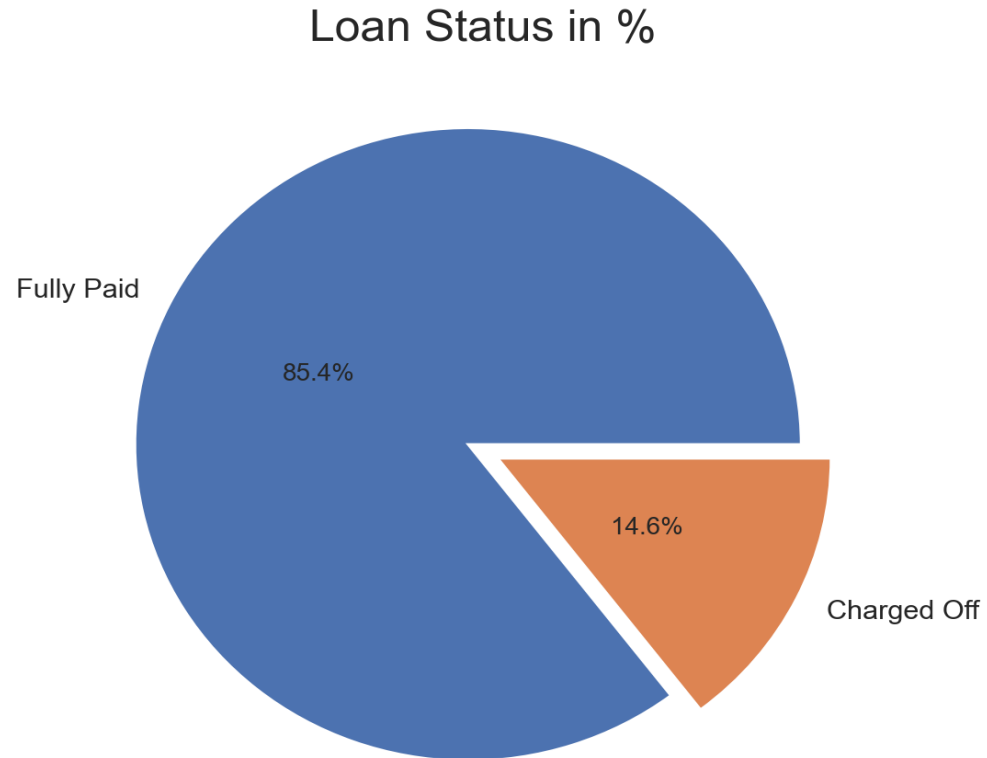| Company	Context	Problem statement
Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.	Lending Club wants to understand the **driving factors** behind loan default, i.e. the **driver variables** which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.	As a data scientist working for Lending Club analyse the dataset containing information about past loan applicants using EDA to understand how **consumer attributes** and **loan attributes** influence the tendency of default.

Analysis Approach

Working with the Data Dictionary and getting knowledge of all the columns and their domain specific uses

➤ **Analyse variables against segments of other variables**

➤ **Create derived variables**

Data Cleaning	Data Understanding	Univariate Analysis	Segmented Univariate Analysis	Bivariate & Multi-Variate Analysis	Recommendations

Publish insights and Observations by analysing all plots and recommendations for reducing the loss of business by detecting columns best which contribute to loan defaulters.

➤ **Drop columns with null values, all random values or single category value**
➤ **Convert values to proper int, float, date representations**

➤ **Check distributions and frequencies of various numerical and categorical variables**

➤ **Create derived variables**

Analysing the two or multiple variable behaviour like term and loan status with respect to loan amount

Analysis - Overall Loan Status in %

Loan Status in %

Observation :
Probability for the borrowers to get defaulted is **14.6%** in the over all loan data.

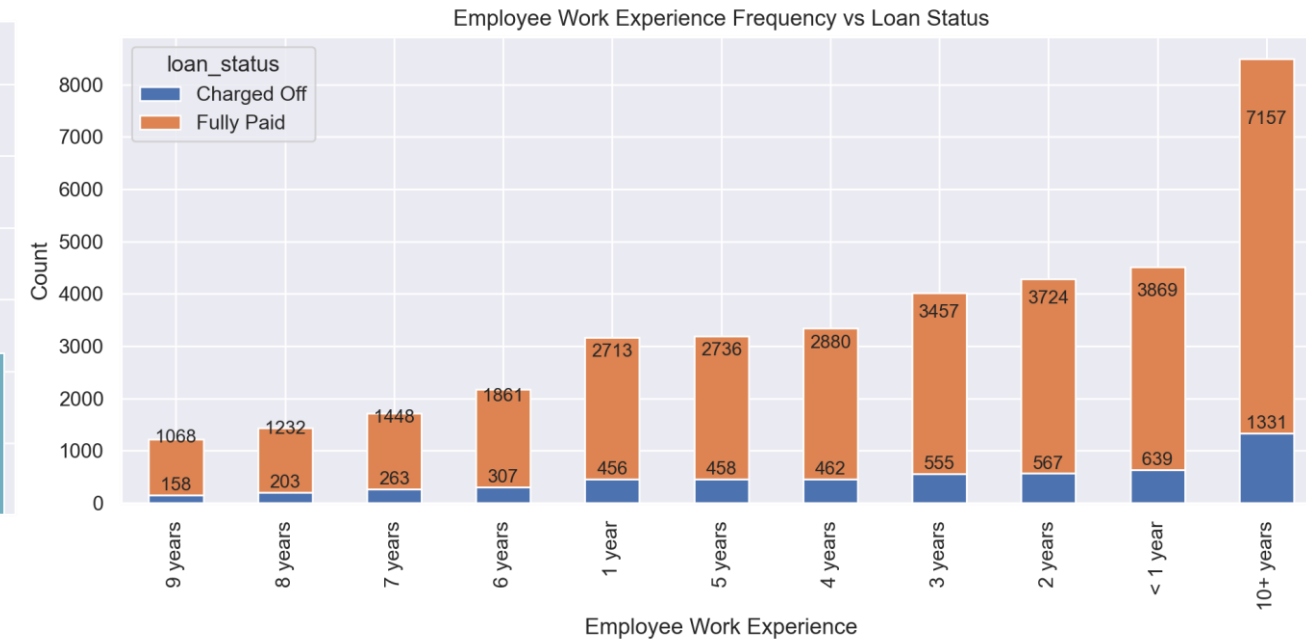# Analysis - Understanding Loans

Top 10 Employee Title
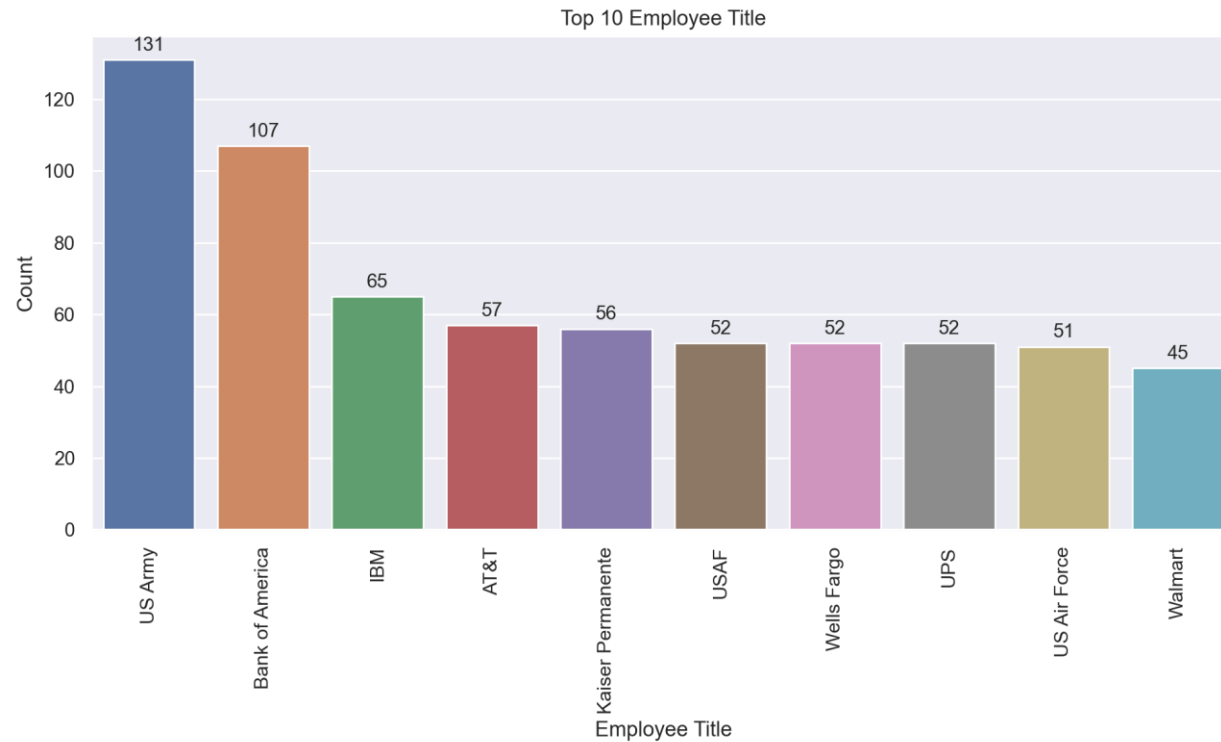
Observation:

These are top **5 employee title** who borrowed loan from Lending Club:
1. US Army
2. Bank of America
3. IBM
4. AT&T
5. Kaiser Permanents

Employee Work Experience Frequency vs Loan Status

Observation:

Analysing for all various work experiences, its observed that employee's who are having experiences **more than 10+ & less than one year** are more to borrow loans from Lending Club.

Similar proportion of **'Charged off' & 'Fully paid' Is observed** and their **standard deviation** is very low. The data gives a feel like we get better understanding on 'Further analysis'.

Analysis - Revolving line utilization rate & Credit Pulls

Revolving line utilization rate vs Loan Status
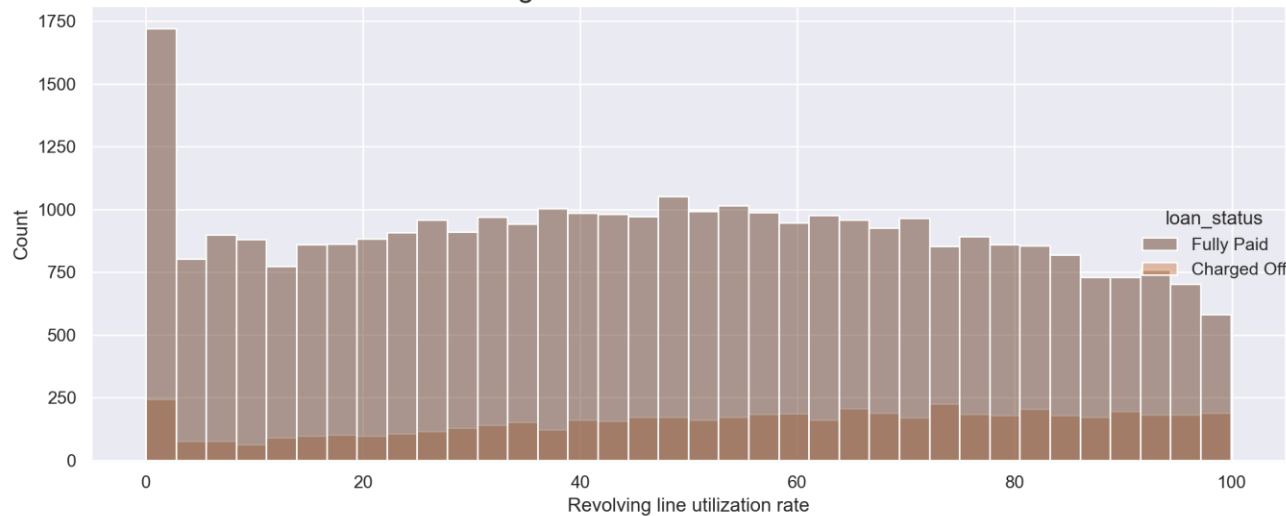
Observation:

As **Revolving line utilization rate** increases we notice that cases of **Charged Off** also increases in proportion.

This column plays crucial role to determine which applicant should be approved based on **Revolving line utilization rate**. If this is high the chances that applicant will default is high and vice versa. It will also help Lending Club to identify how much loan amount should be approved.

Credit-pull by LC in Months
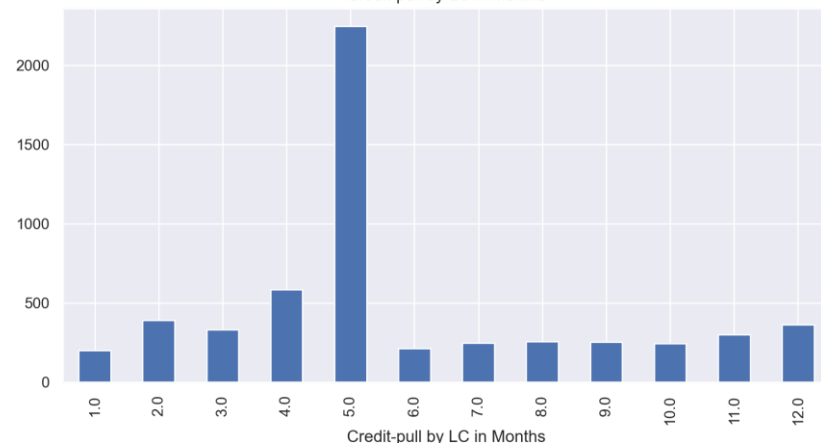

Credit-pull by LC in Years
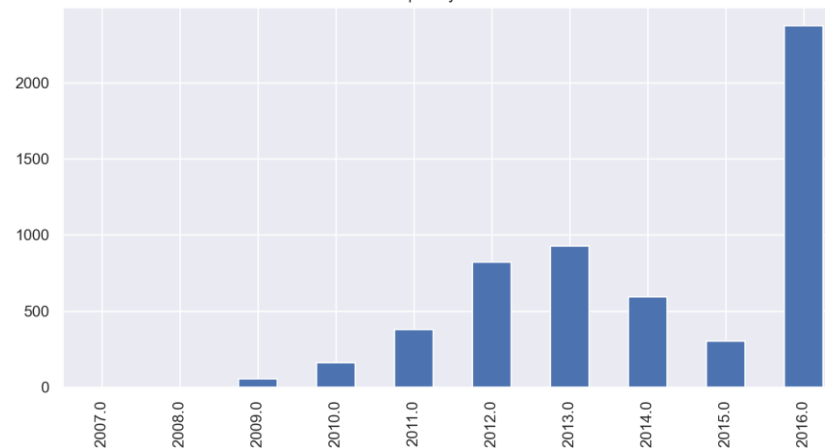
Interesting Titbits:

Based on the analysis above **no. of credit pulls** w.r.t **'months' & 'years',** it gives a clear picture to Lending club that they had done many credit-pulls during the month of **May** and in the years **2012,2013 and 2016**.

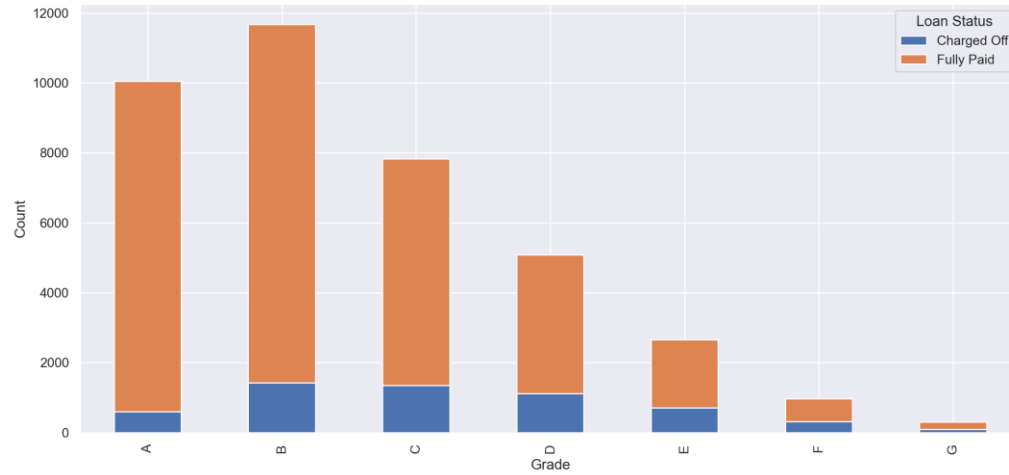# Analysis - Defaults by Borrower's Term & Interest Rate

Loan Term vs Loan Status in %

Histogram-Interest Rate vs Loan Status

Box Plot-Interest Rate vs Loan Status

loan_status	mean	median
Charged Off	14.0	14.0
Fully Paid	12.0	11.0
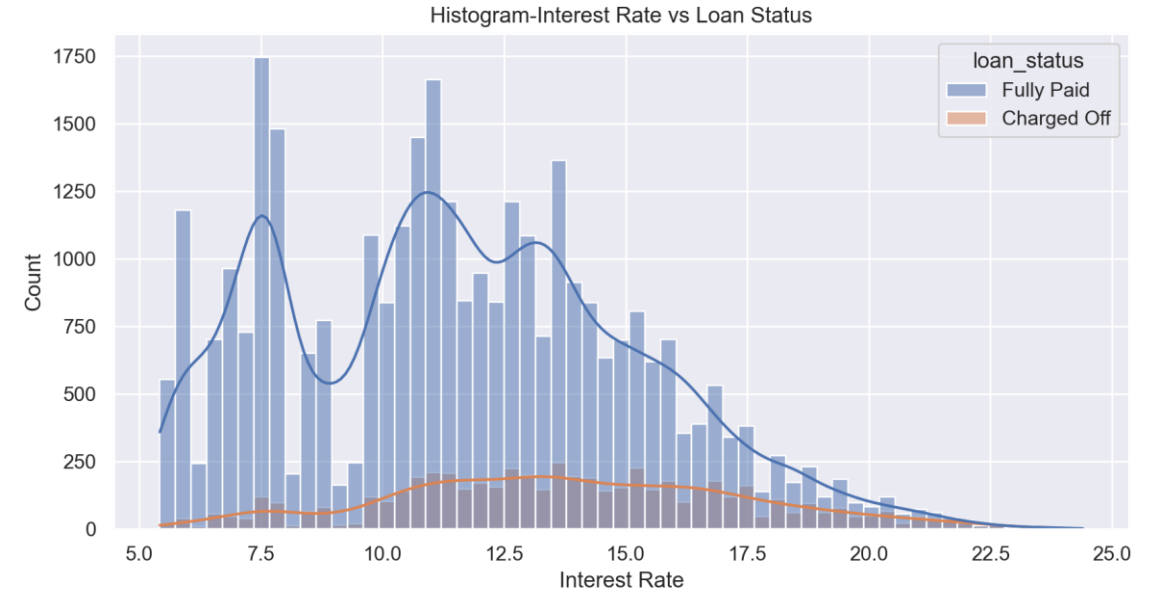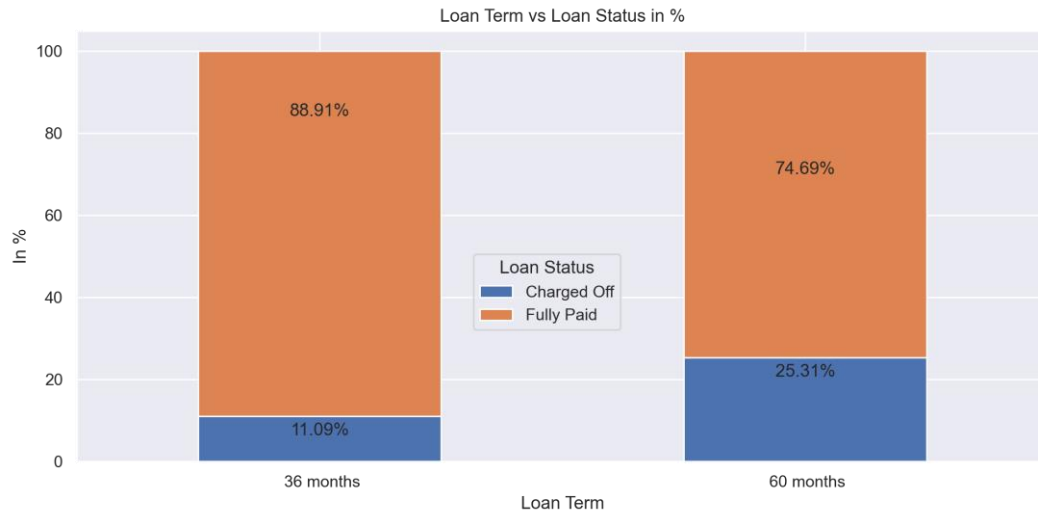
Observation:

- Its observed that there is higher probability of **25%** to be **Charged Off** if term is for **60 months**. In other words, because larger loan amount requires long term and in between borrowers are not able to re-pay the loan by any reason.

- As per the data the **Mean & Median** of Interest Rate column is slightly higher for Loan Status with **Charged Off** category. So we can conclude that Interest Rate column can be considered as a featured column. As the LC finds applicants records as risky then the interest rate of loan also increases in proportion.

- LC is able to identify the **Charged Off** behaviour and impose higher interest rate by average of **14%** which is **2%** higher than the **Fully Paid.**
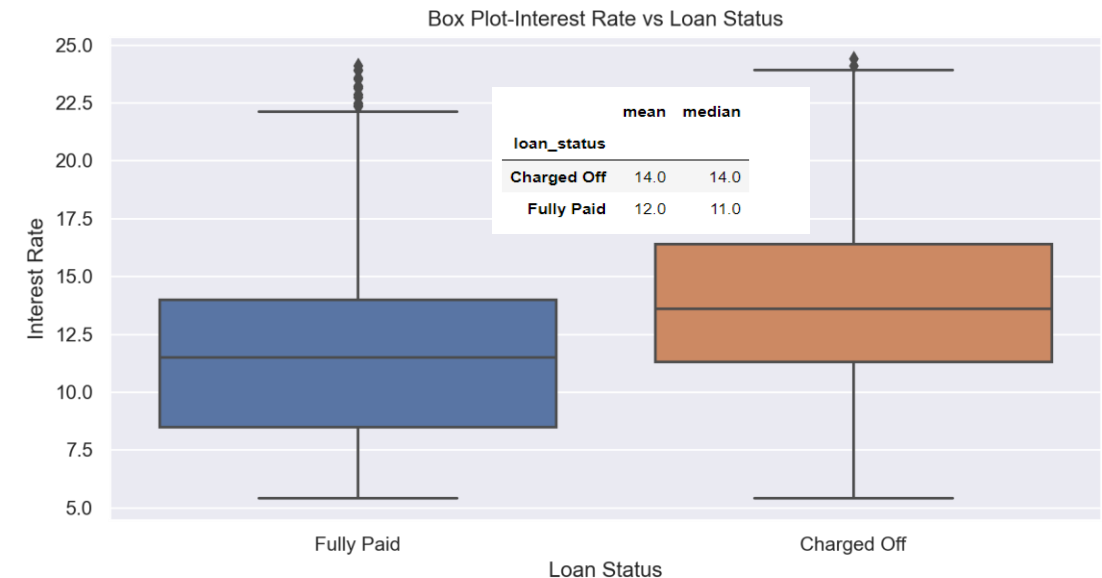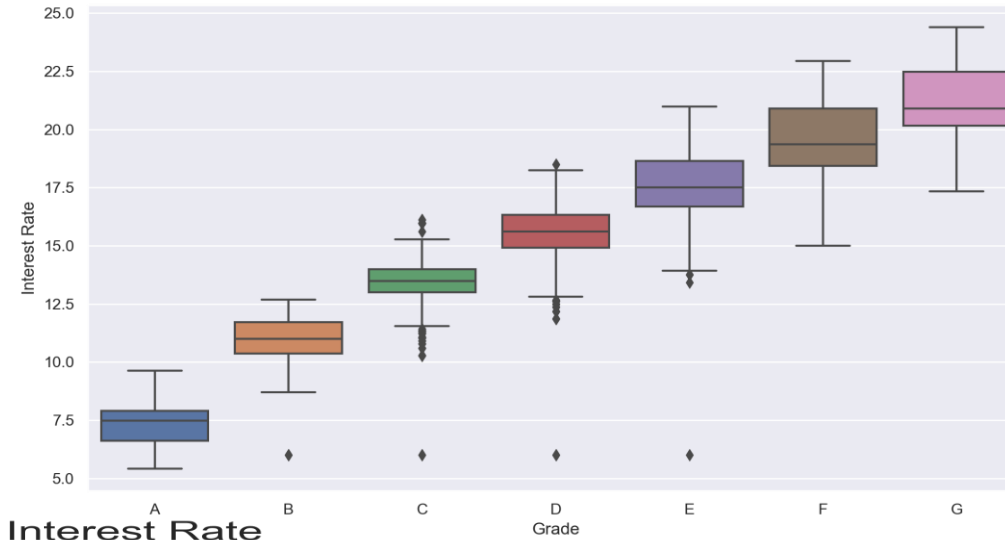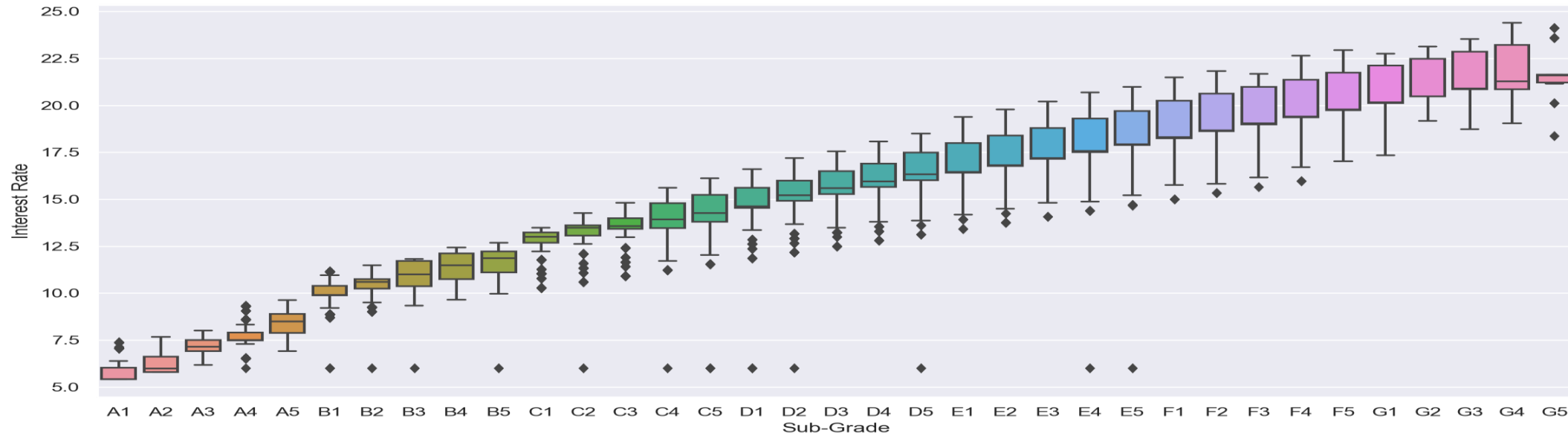
Analysis – Defaults by Grade & Sub Grade
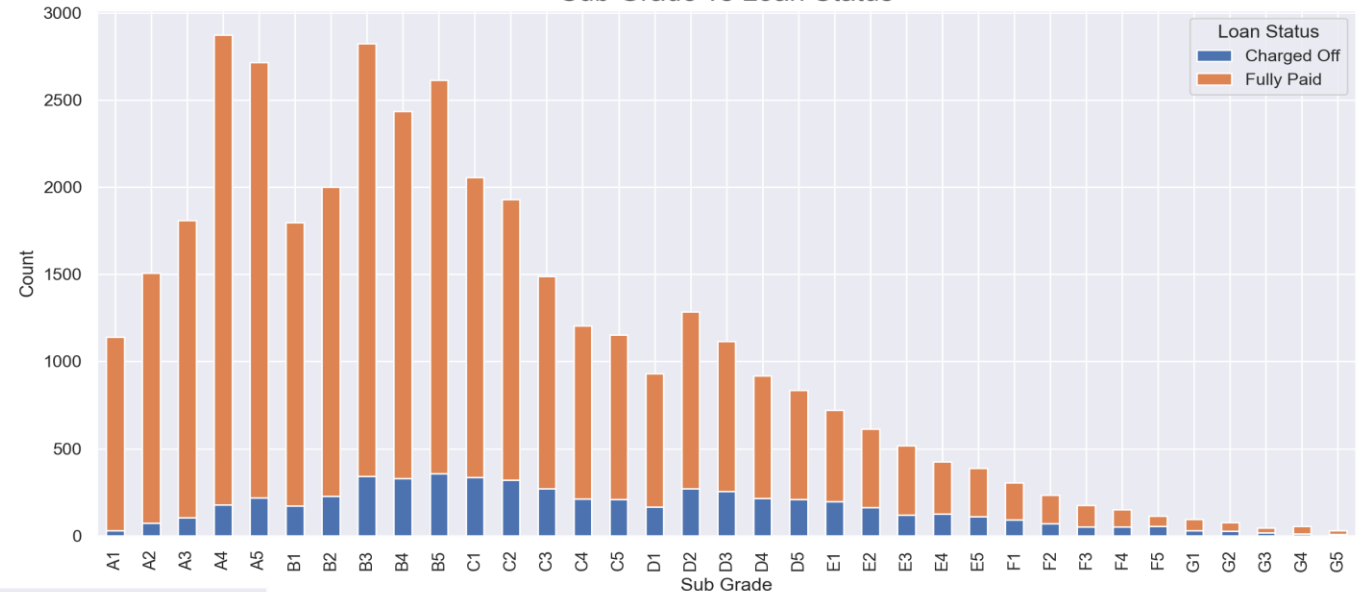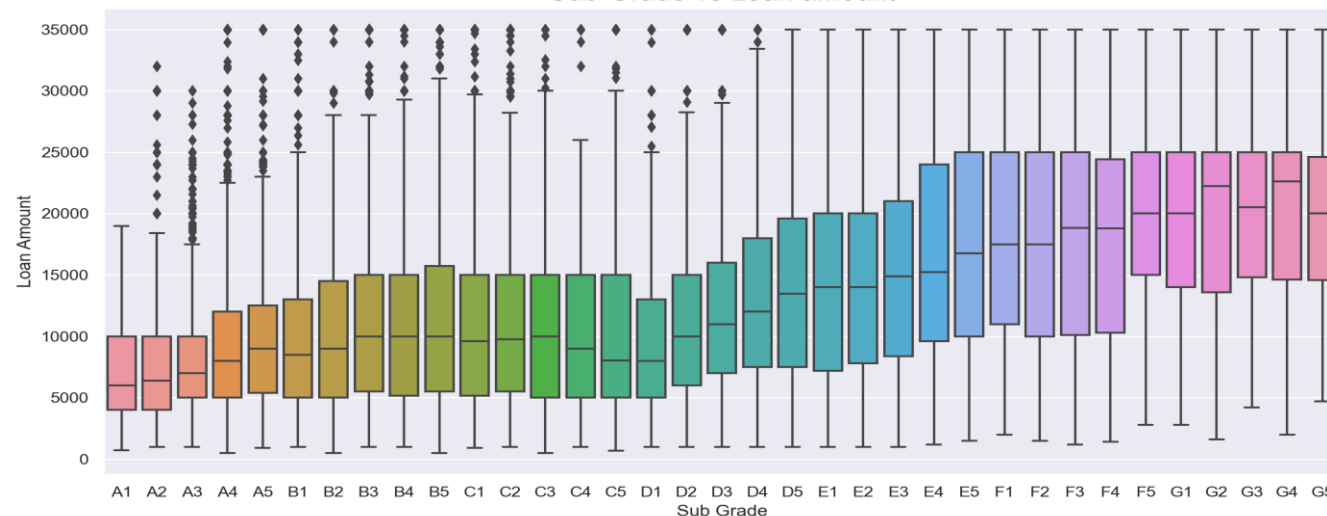
Analysis – Defaults by Grade & Sub Grade Cont.….
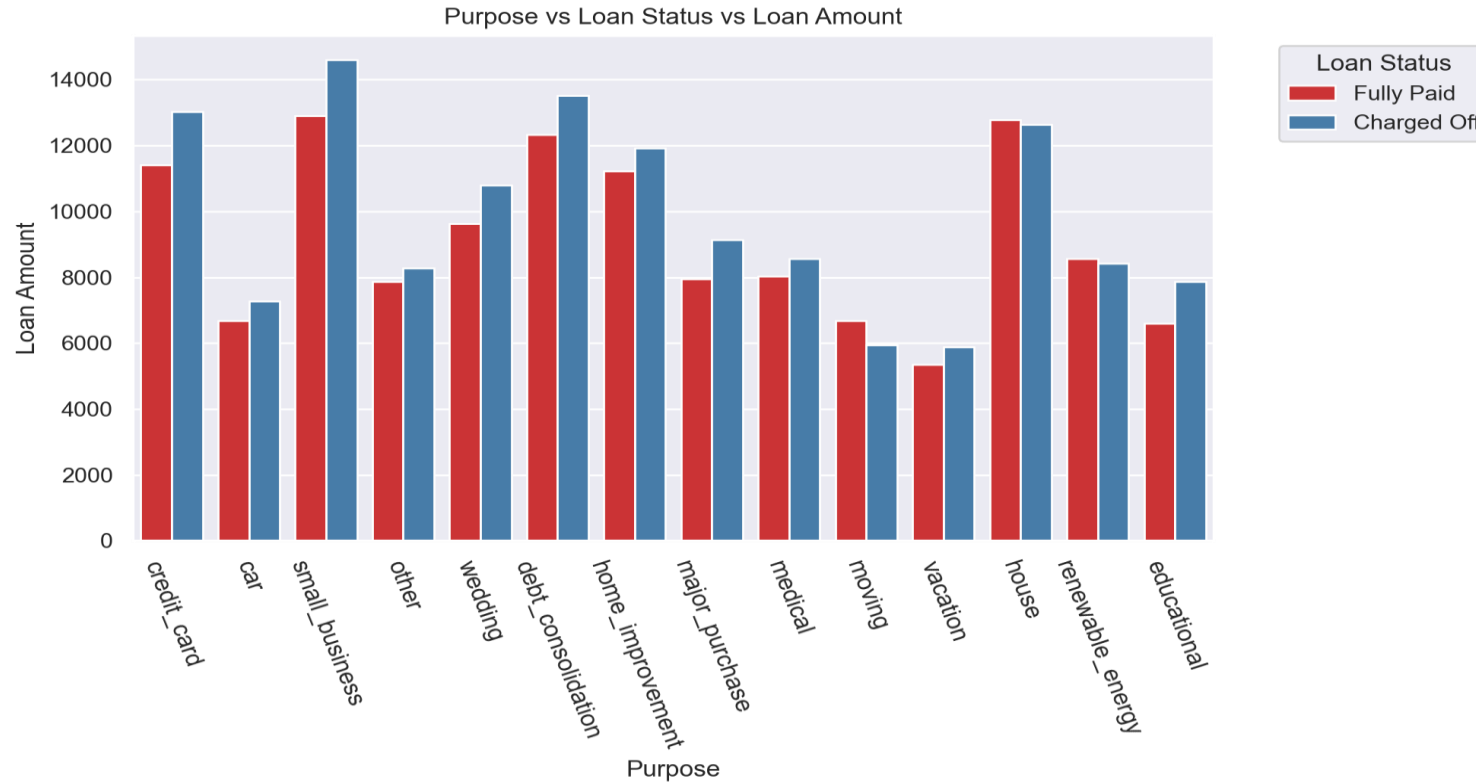

Sub Grade vs Loan Status

Sub-Grade vs Loan amount

Observation for Grade & Subgrade Column:

From the above analysis, LC considers Loan borrowers are much likely to **default** when their loan amounts are high and hence they give them loan at certain higher interest rate by deciding what grade and subgrade they are falling into.

Its observed that there are many **loan defaulters** from **B** and **C grade loan** as well. Hence as per the business objective they should be able to identify and categorize this applicants to lower grades so that they could be on safe side not running into losses.

Analysis – Defaults by Purpose

Purpose vs Loan Status vs Loan Amount

| loan_status	Charged Off	mean Fully Paid
purpose		
car	7264	6685
credit_card	13015	11397
debt_consolidation	13518	12320
educational	7859	6592
home_improvement	11914	11225
house	12640	12775
major_purchase	9135	7950
medical	8568	8041
moving	5936	6688
other	8285	7868
renewable_energy	8421	8565
small_business	14605	12905
vacation	5876	5342
wedding	10787	9637

Observation:

- Its observed from the above analysis that **'Charged Off'** cases are higher in almost all categories of Loan Purpose.

- Borrower's who took loans for **small business, debt consolidation, credit card, house & house improvement** purpose's have defaulted more.

Analysis – Defaults by Home Ownership

Home Ownership Type vs Loan Status in %

Home Ownership Type vs Debt to Income Ratio vs Loan Status

Home Ownership Type vs Loan amount vs Loan Status

Home Ownership Type vs Annual Income vs Loan Status

Observation:

- Borrowers with **Mortgage** home ownership are taking higher loans and **defaulting** the approved loans.

- Lending club should **stop** giving loans to this category when loan amount requested is **higher**.

- When **Annual Income** is high regardless of home ownership type, Borrowers are **much likely** to repay the loan.

- When **Debt to Income Ratio** is high regardless of Home Ownership Type, Borrowers are **less likely** to repay the loan.
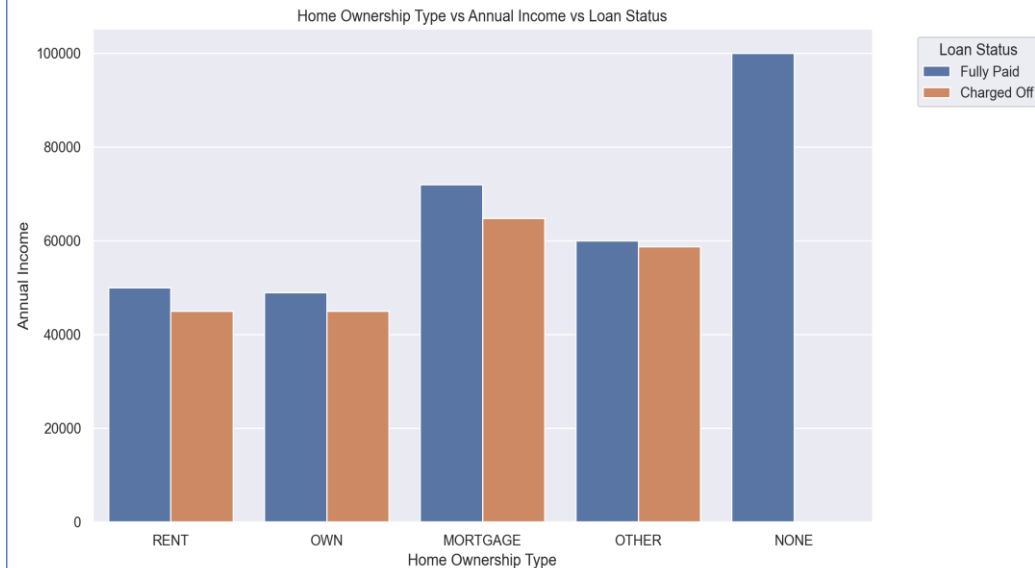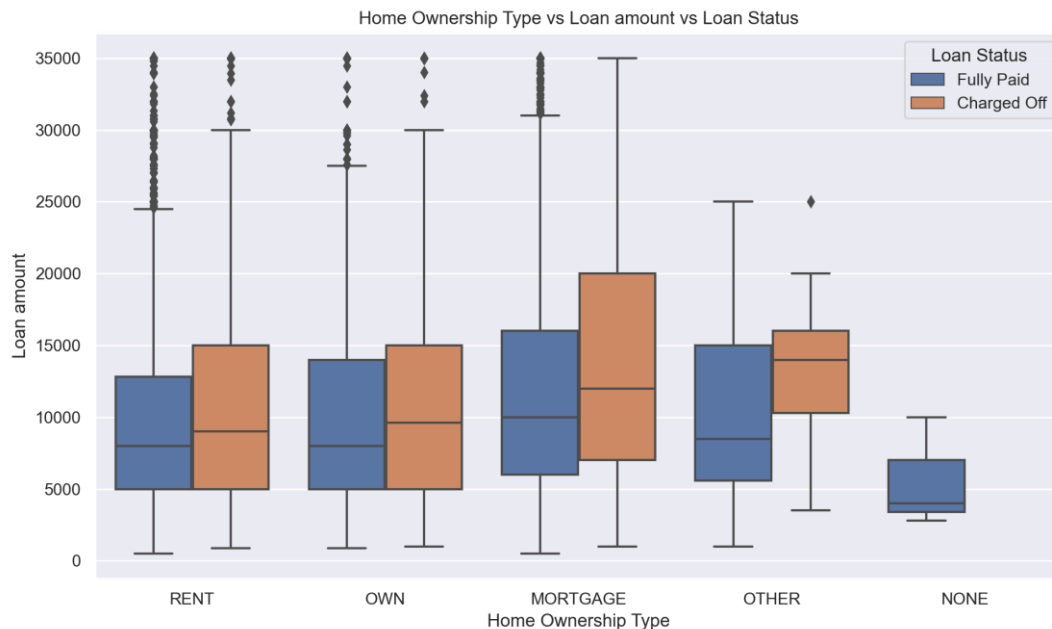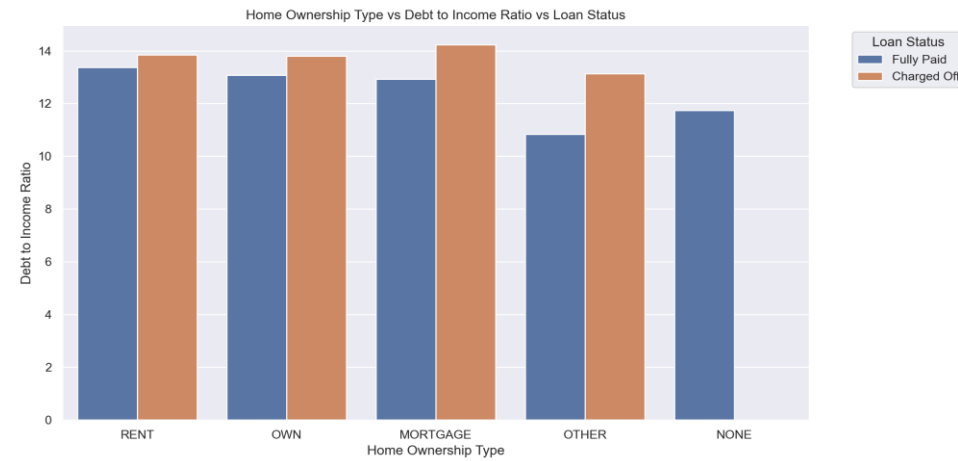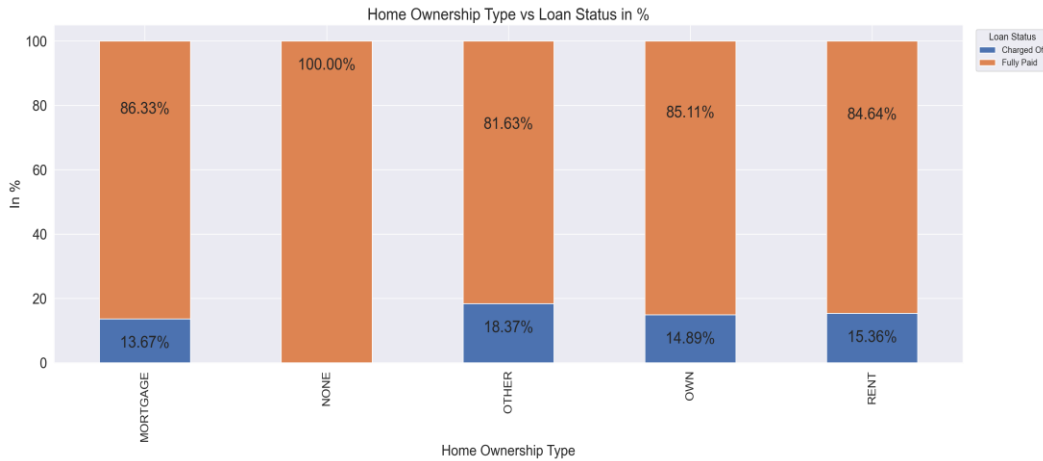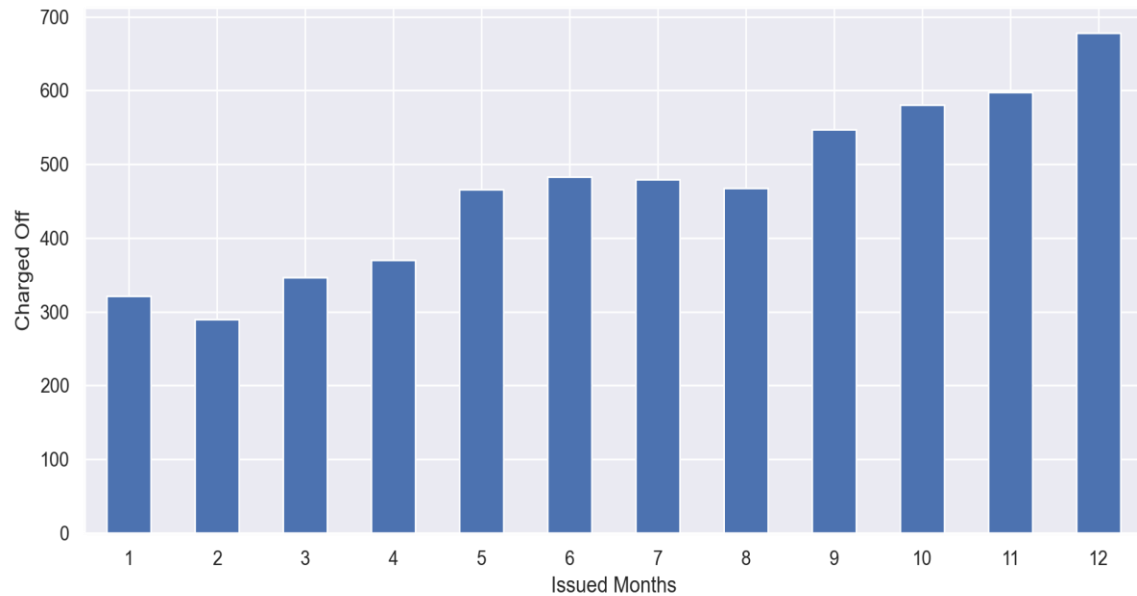
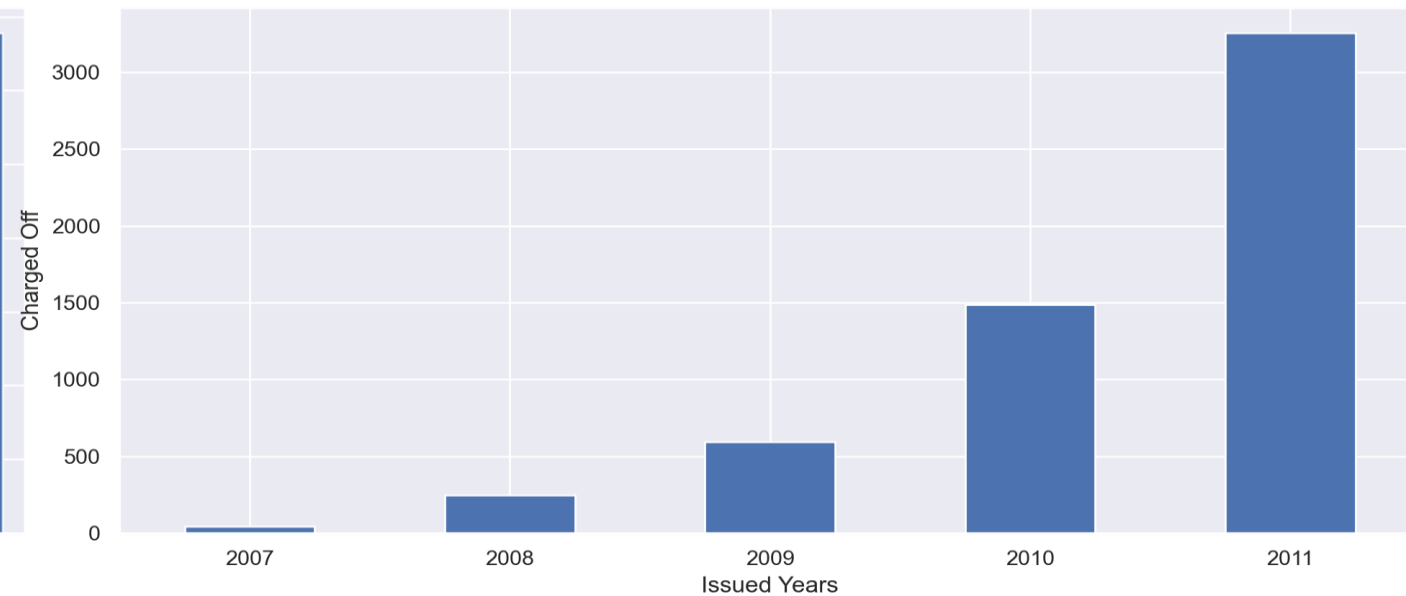# Analysis – Defaults by Verification Status
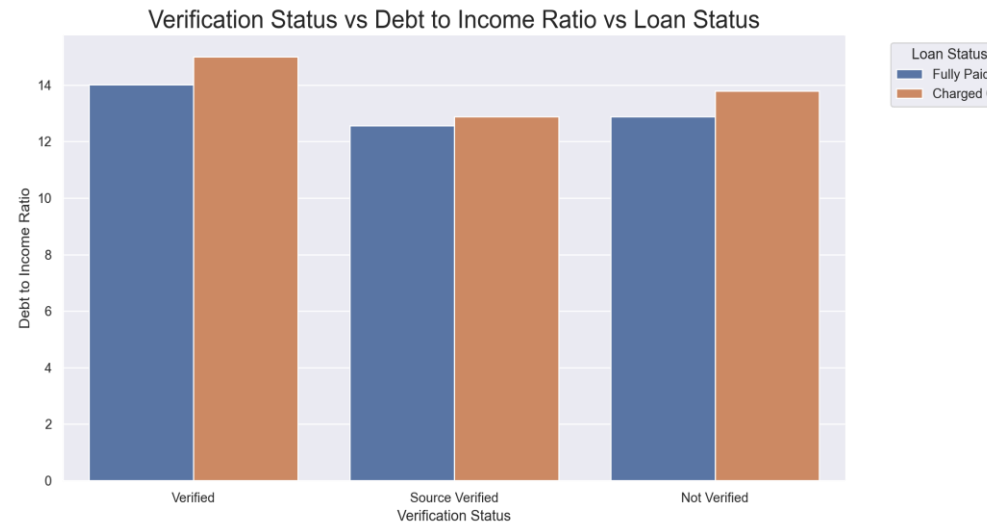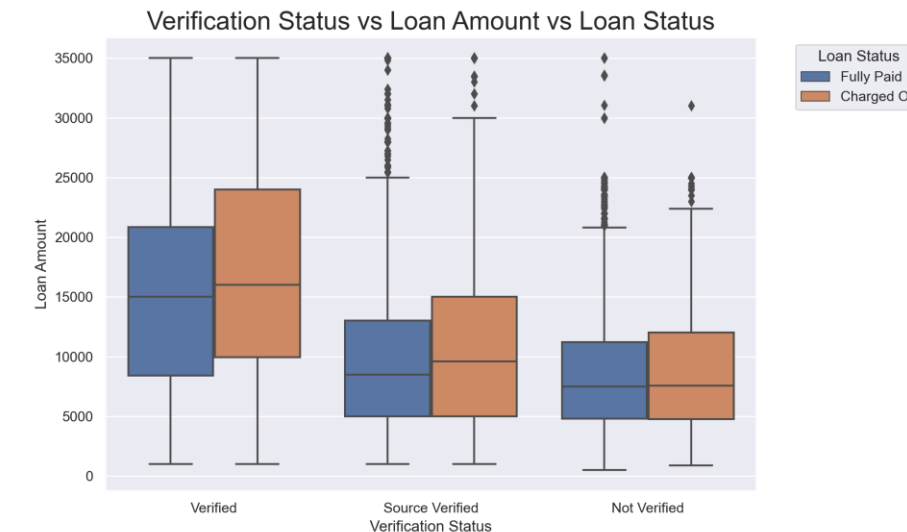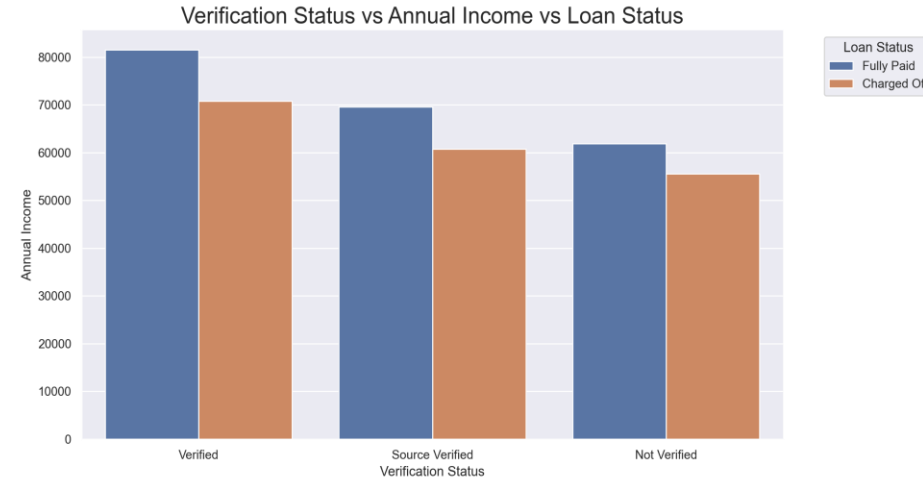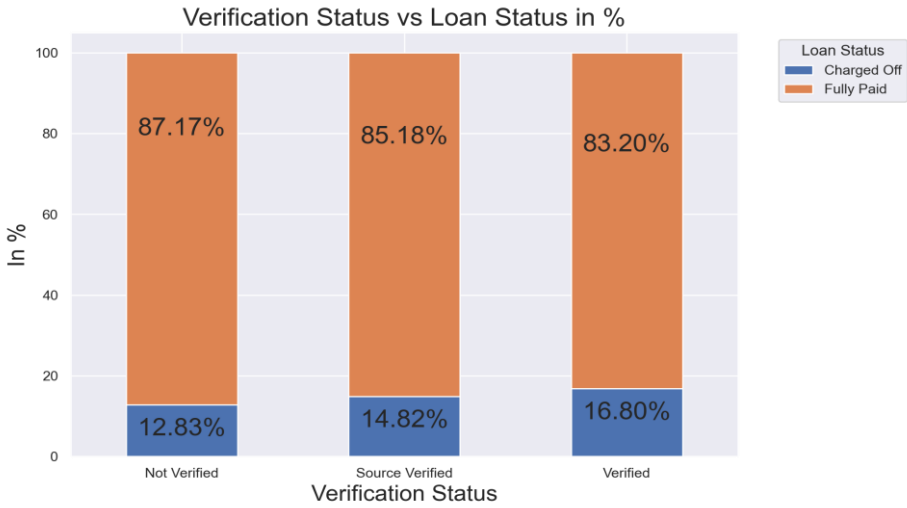
Observation:

LC gets a clear picture that their **Verification** team has poor mechanism in identifying the applicants based on **Debt to Income Ratio**. If **DTI ratio** is high then such applicant should be verified rigorously.

Analysis – Defaults by Issued Date

Observation:

There is increase in loan numbers as months get closer to the end of the year.
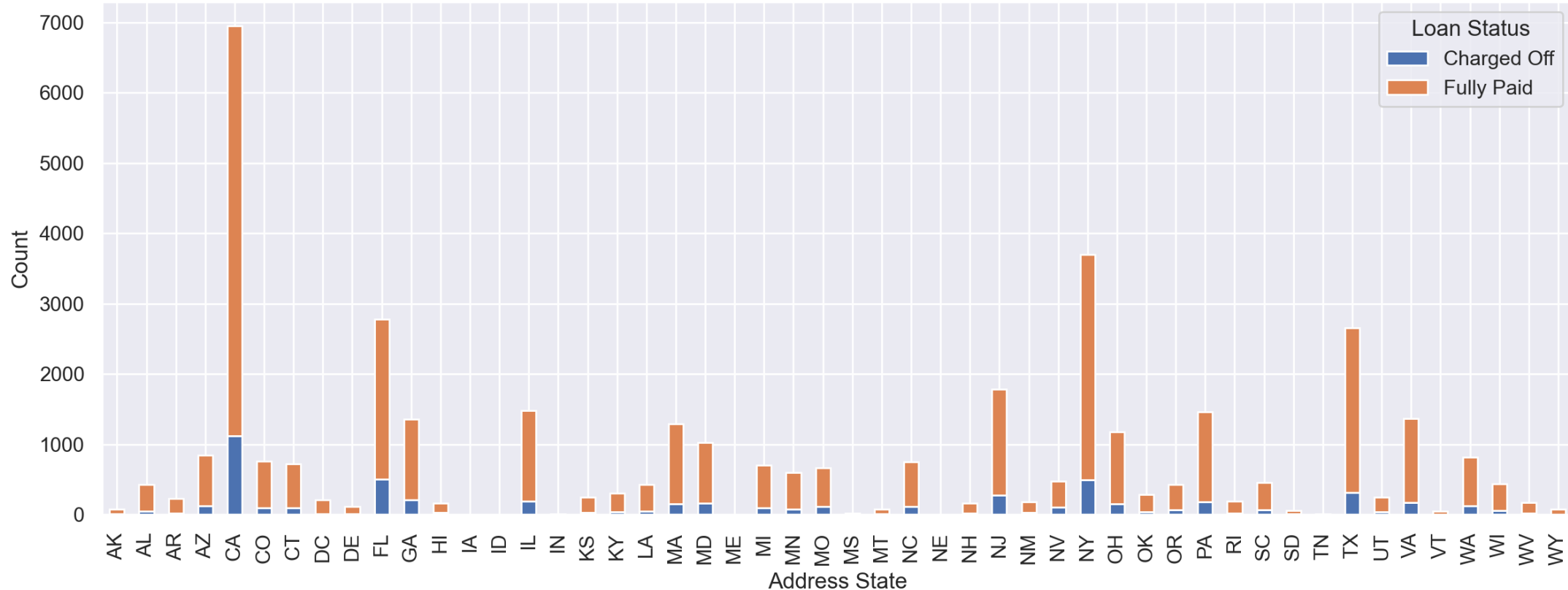
Financial Crisis
In the span of year 2007-11, its observed there is a spike in number of **charged off** cases. The one possible reason could be financial crisis happened in 2008 in US Markets.

Regarding the spike increase in charged off cases in 2011, please refer to the below link:
https://www.insidehighered.com/news/2012/10/01/two-year-default-rates-student-loans-increase-again
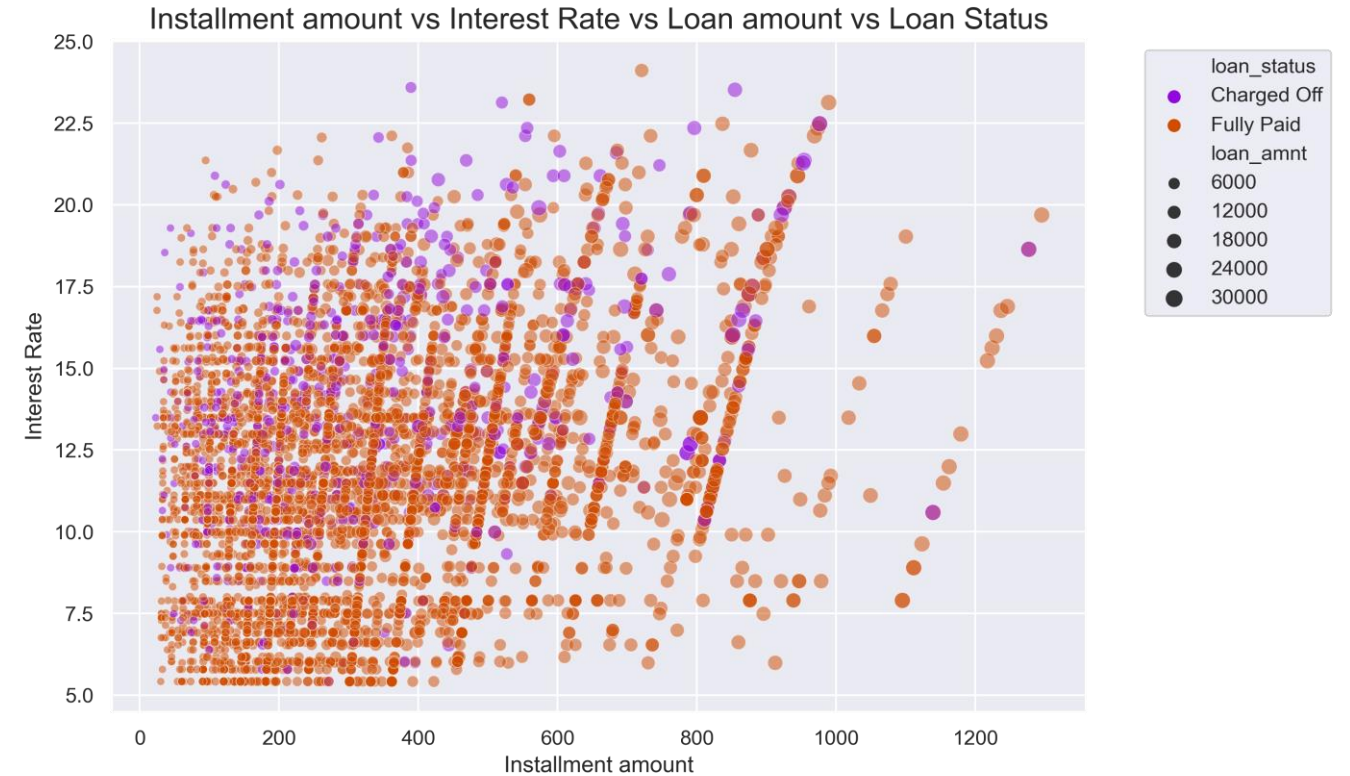
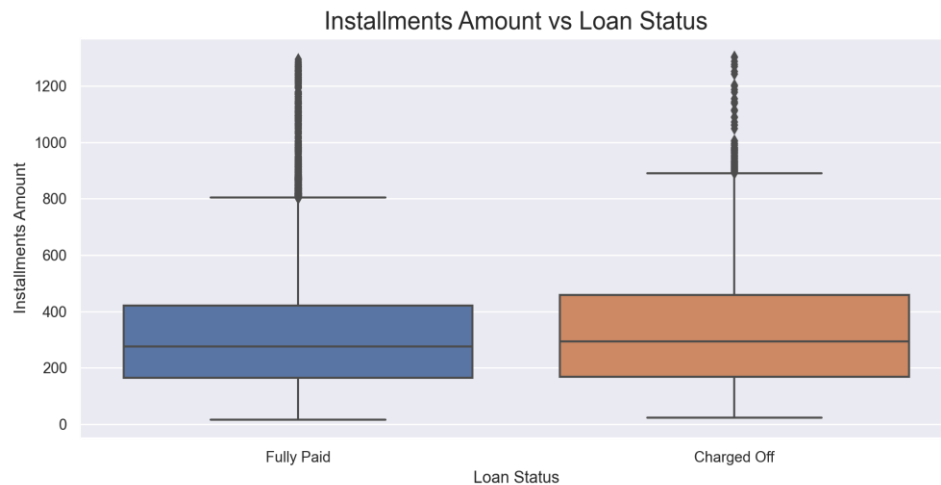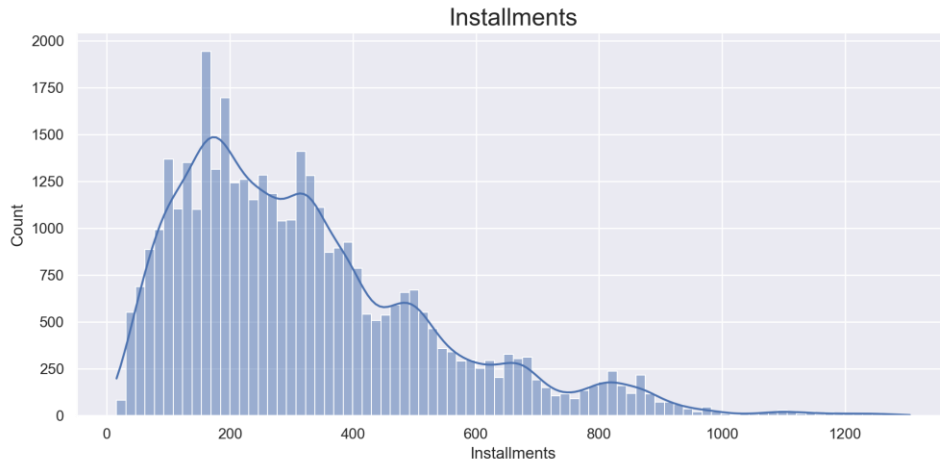# Analysis – Defaults by Address State

Address State vs Loan Status

Observation:
Maximum loans are in populous states, California, New York, Florida and Texas. These states also shows more number of charged offs in large number of applications.

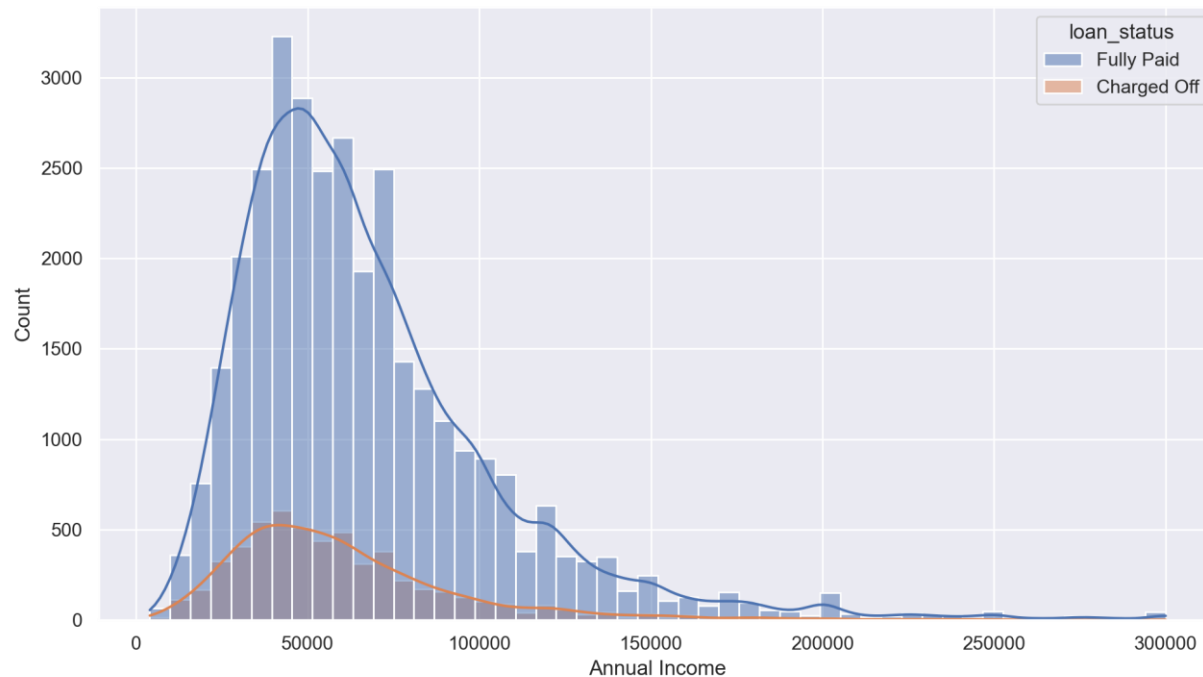# Analysis – Defaults by Installment Amounts

Observation:

LC finds that **higher loan amounts** are provided with **higher rate of interest** with **high installment amounts** and its noticed that many are **defaulted** during the case of **higher amounts.**
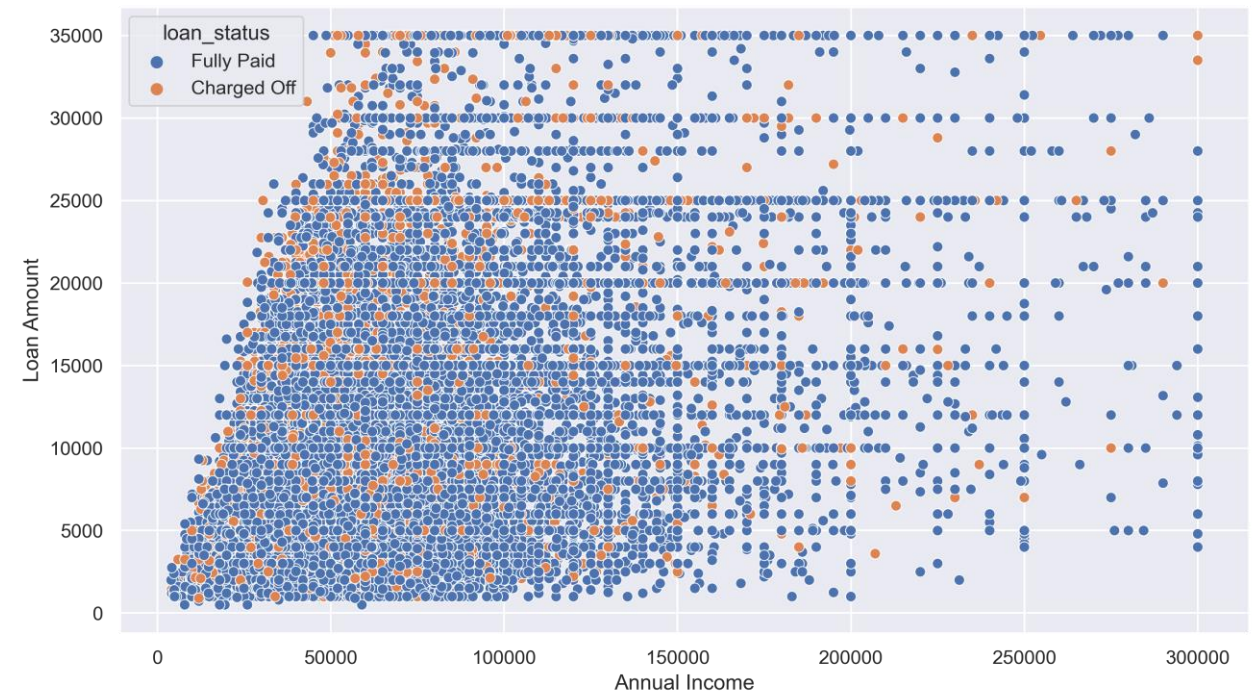
Analysis – Defaults by Annual Income

Annual Income vs Loan Status

Annual Income Ratio vs Loan Status vs Loan Amount

Observation:
Borrowers having annual income near to **50,000 USD** default on their loans at much higher rates. Loan default decreases with higher annual income. We do not find any relation of annual Income w.r.t Interest Rate, Installment Amounts and Debt to Income Ratio. However, We do noticed that there are loan borrowers who is having higher loan amount compare to their Annual Income. We see here that Lending Club has extended high-value loans to people with low income. There are many cases of people with income **50,000 USD** or less getting loans of **25,000 USD** or more. This practice should be curtailed.

Analysis – Defaults by Debt to Income Ratio

Debt to Income Ratio vs Loan Status

Debt to Income Ratio vs Loan Status vs Interest Rate
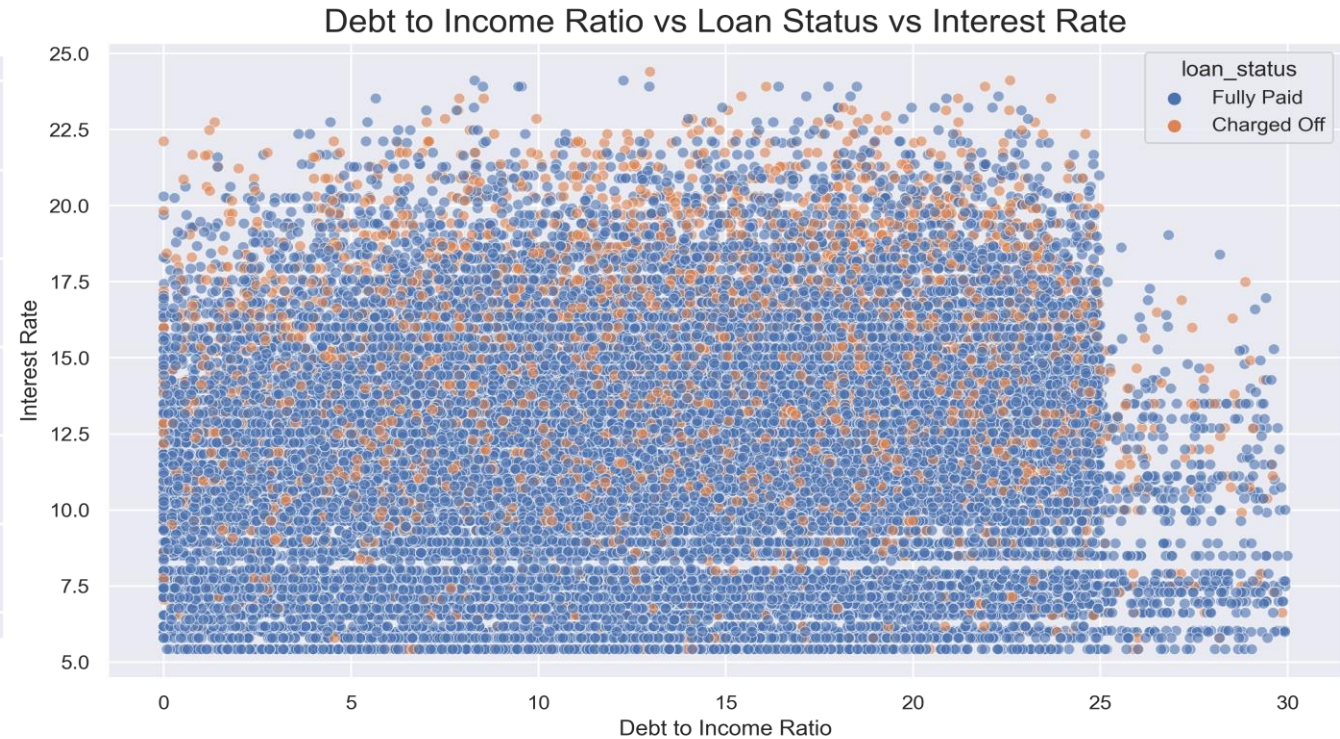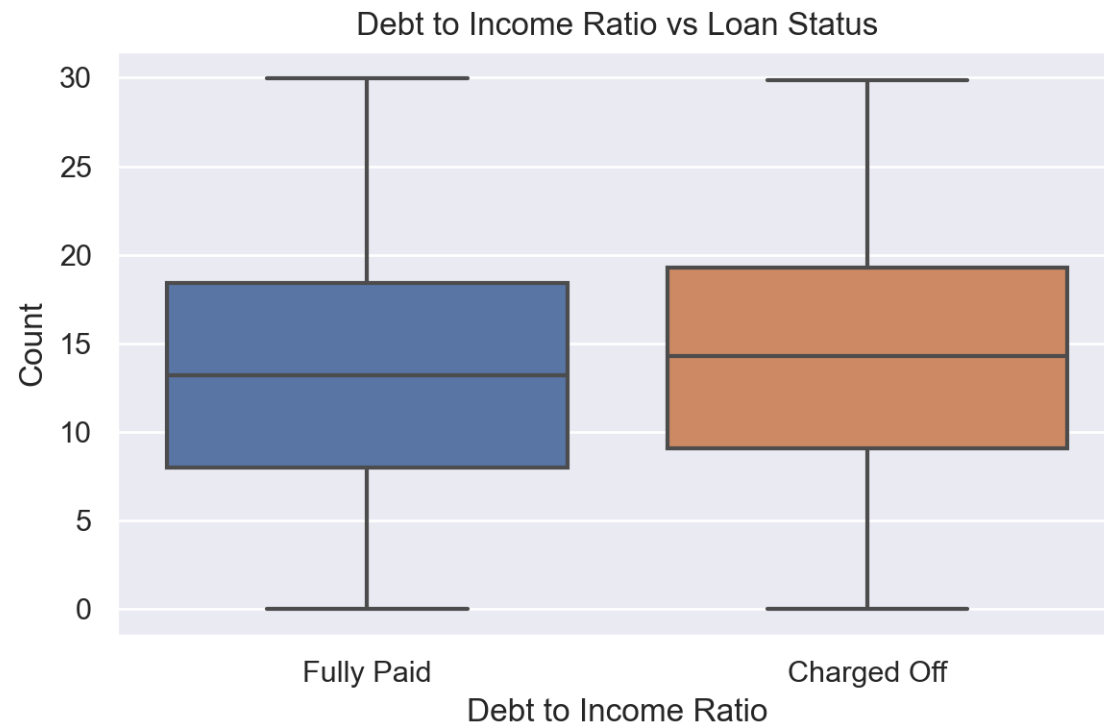
Observation:
- As the debt to income ratio increases we see increasing trend in number of charged off.
- Higher interest rates should be charged for higher DTI , but we see spread across all values.

Recommendations
Better Quality Borrowers

- **Reduce:**
  - Probability for the borrowers to get defaulted 14.6% should be reduced.

- **Strategy:**
  - LC should devise a strategy for the applicants having higher loan amount with a tenure of 60 months to reduce the loss.
  - Many loan defaulters are from B and C grade loan, hence as per the business objective they should be able to identify and categorize this applicants to lower grades so that they could be on safe side not running into losses.
  - LC's Verification team has poor mechanism in identifying the applicants based on Debt to Income Ratio.
  - If DTI ratio is high then such applicant should be verified rigorously.
  - Higher interest rates should be charged for higher Debt to income ratio.

- **Stop:**
  - Stop giving higher loan amounts to the applicants having higher revolving line utilization rate.
  - Borrowers with mortgage home ownership are taking higher loans and defaulting the approved loans.
  - Lending club should stop giving loans to this category when loan amount requested is more than 12000.

- **Cautious:**
  - LC should be cautious with the borrower's having purposes like small business, debt consolidation, credit card, house & house improvement as such ones requires higher loan amounts and chances of defaulting is also higher.
  - LC finds that higher loan amounts are provided with higher rate of interest with high installment amounts and its noticed that many are defaulted during the case of higher amounts.
  - LC has extended high value loans to people with low income. There are many cases of people with income 50,000 USD or less getting loans of 25,000 USD or more. This practice should be curtailed.