GRAMENER CASE STUDY

SUBMISSION

BY

Srinivasan Narayanaswamy

Vishnu Vivek Ravindran

Sri Yogesh M E

Srinithi Sampath

Objective & Problem Statement

- Few loan applicants tend to default on paying back, so the banks are forced to Charge off the loan.
- Gramener wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default.
- The company can utilize this knowledge for its portfolio and risk assessment.

Financial Risk Analysis

- This is a presentation which helps consumer finance company to assess the risk of lending loans to applicants thereby cutting down the amount of credit loss.

- There are two types of decision could be made
  - Loan Accepted
    - Fully Paid
    - Current
    - Charged-Off (Defaulted)
  - Rejected

- If an applicant is likely to repay the loan and if company rejected the loan, it's a loss of business to the company.
- If an applicant is NOT likely to repay the loan and if company approves the loan it leads to financial loss to the company.
- We are expected to analyse the risk involved in proving loan to the applicants in-order to minimize credit loss to the company. To be precise, now we are interested in Charged-off applicants whom have defaulted their loans. Let's take a look at the Data set.

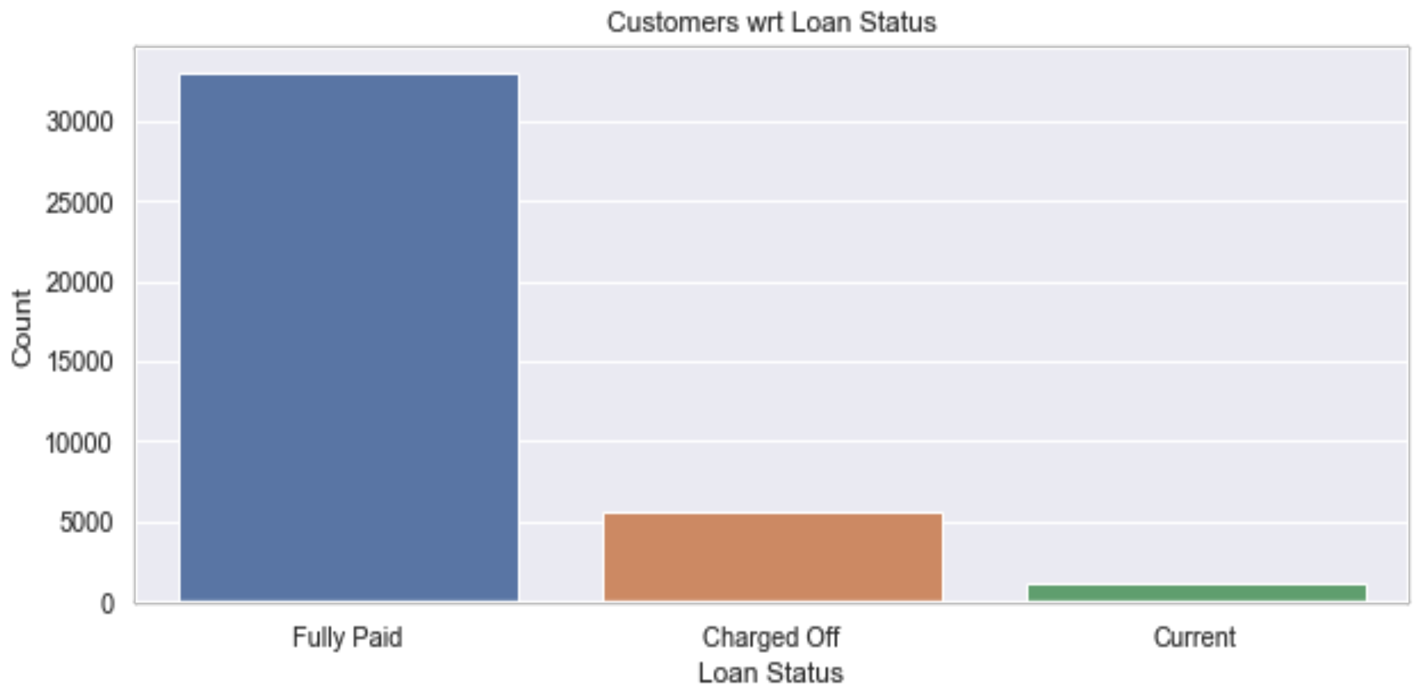
Data Cleaning and Handling Missing data

- While importing the loan.csv, use a date parser to import date in the right format.
- There are many insignificant columns we need to remove before we create insights from the dataset.
- id is the column which uniquely identifies the loan applicant and there seems not any duplicates.
- There are 54 columns with 100% missing values are to be removed, also many other insignificant columns which are not needed for our analysis. Removing those leaves us with 18 columns.
- Strip off percentage from the below columns and convert them to float
  - int_rate
  - revol_util

Data Cleaning and Handling Missing data (Contd.)

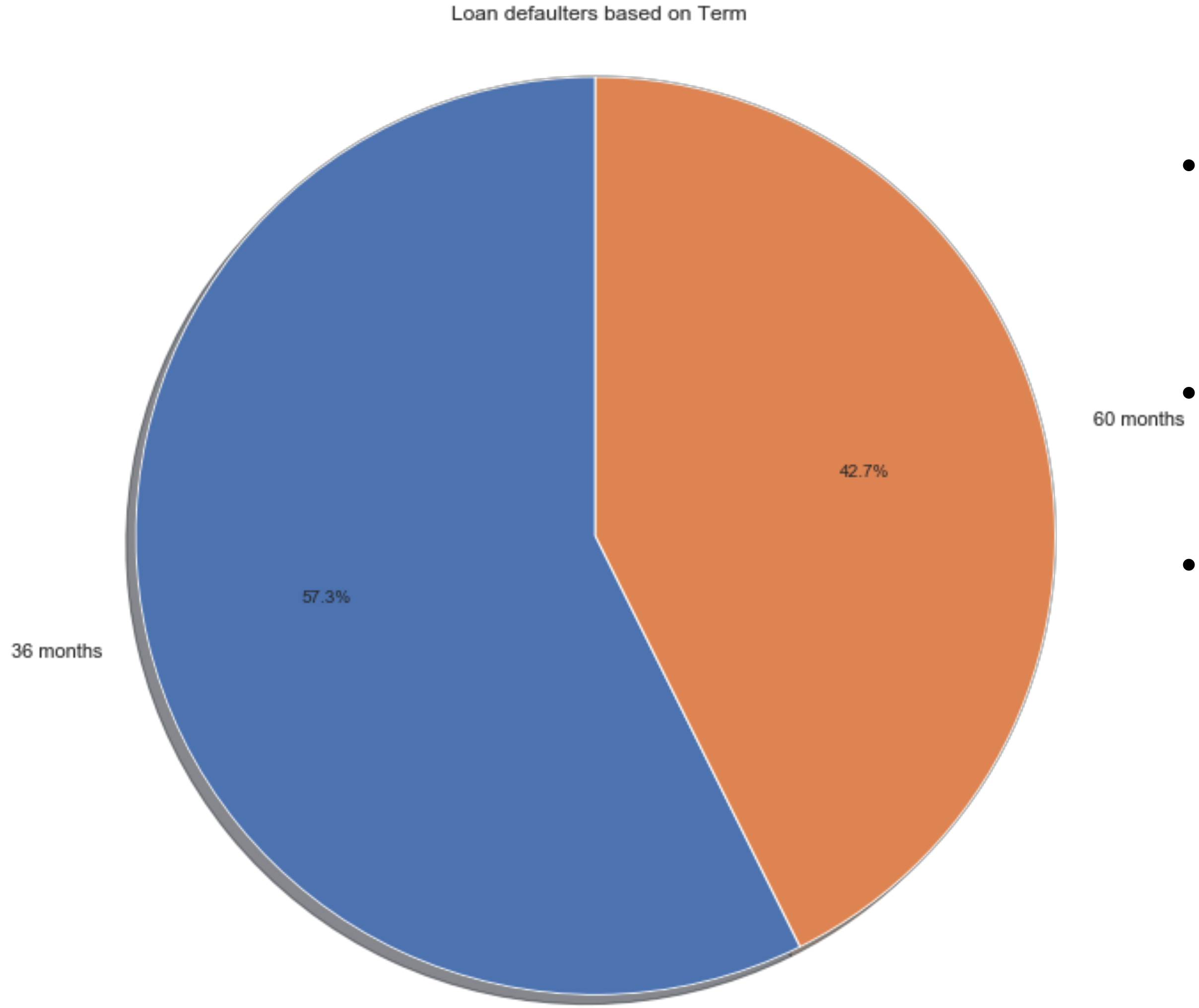
- After removing those columns, there are not many missing values in rows. So we do not delete or impute rows.
- Some of the rows have emp_length as NaN, which can be imputed with the mode of emp_length which is 10.
- Converting emp_length column as integer helps in analyzing the data.
- Using issue_d we can derive the below columns which can help in analyzing the data.
  - Year
  - Month
  - Quarter

Univariate analysis of Loan Status



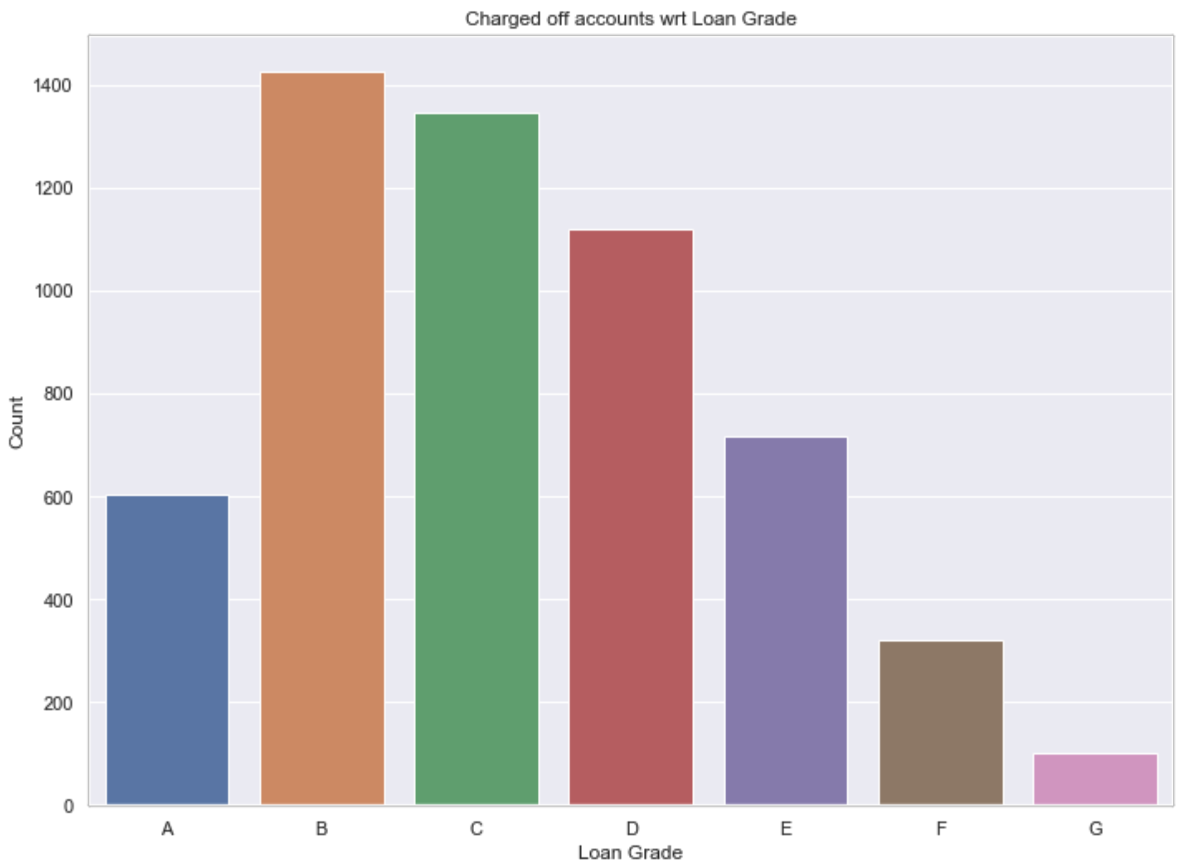
Almost 14.17 % of the total loans are Charged Off, which indicates that the banks have to take stringent checks before approving a loan.

Loan defaulters based on Term



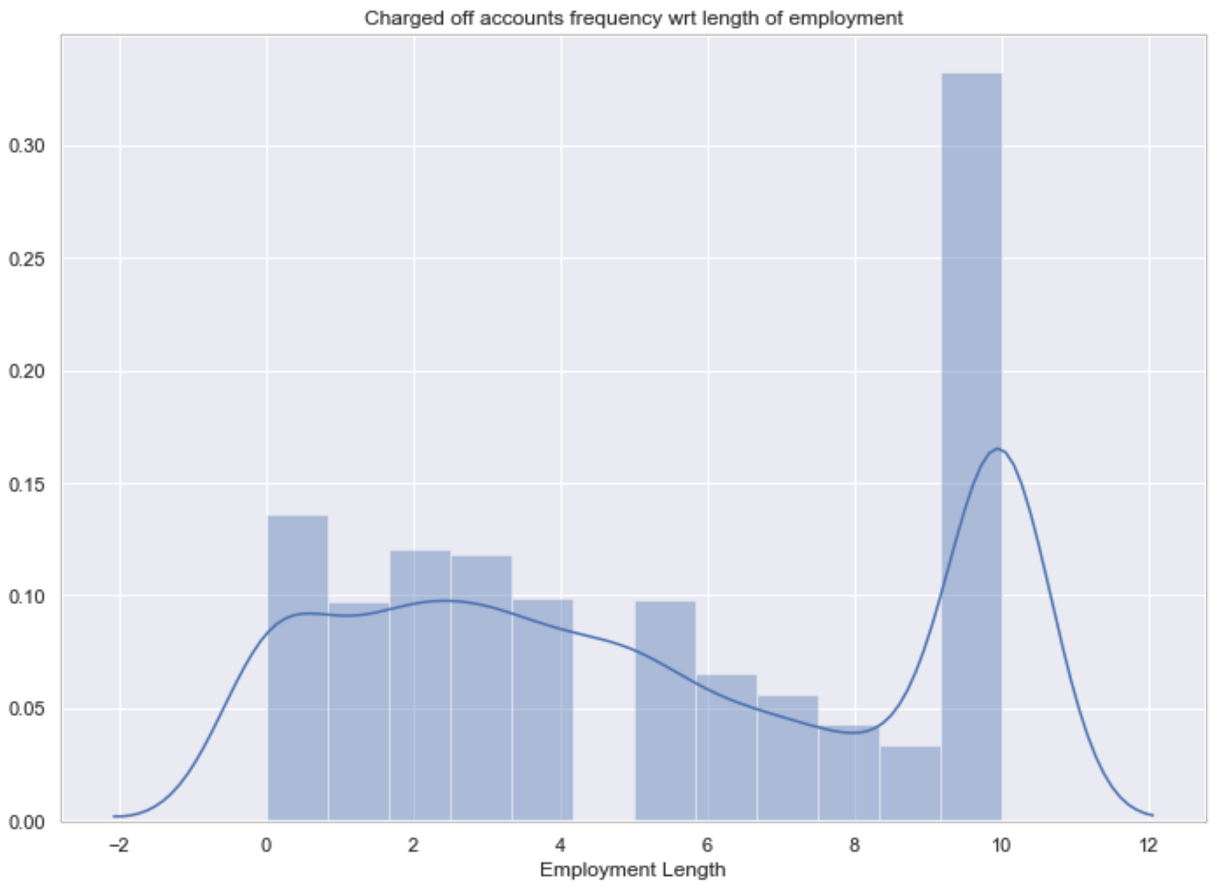
- Out of the total loan defaulters, 57.3% had the loan of term 36 months
- Overall 42.7% of the defaulters had the term of 60 months.
- So, giving a long term loan poses a slightly lower financial risk.

Charged off accounts based on Loan Grade



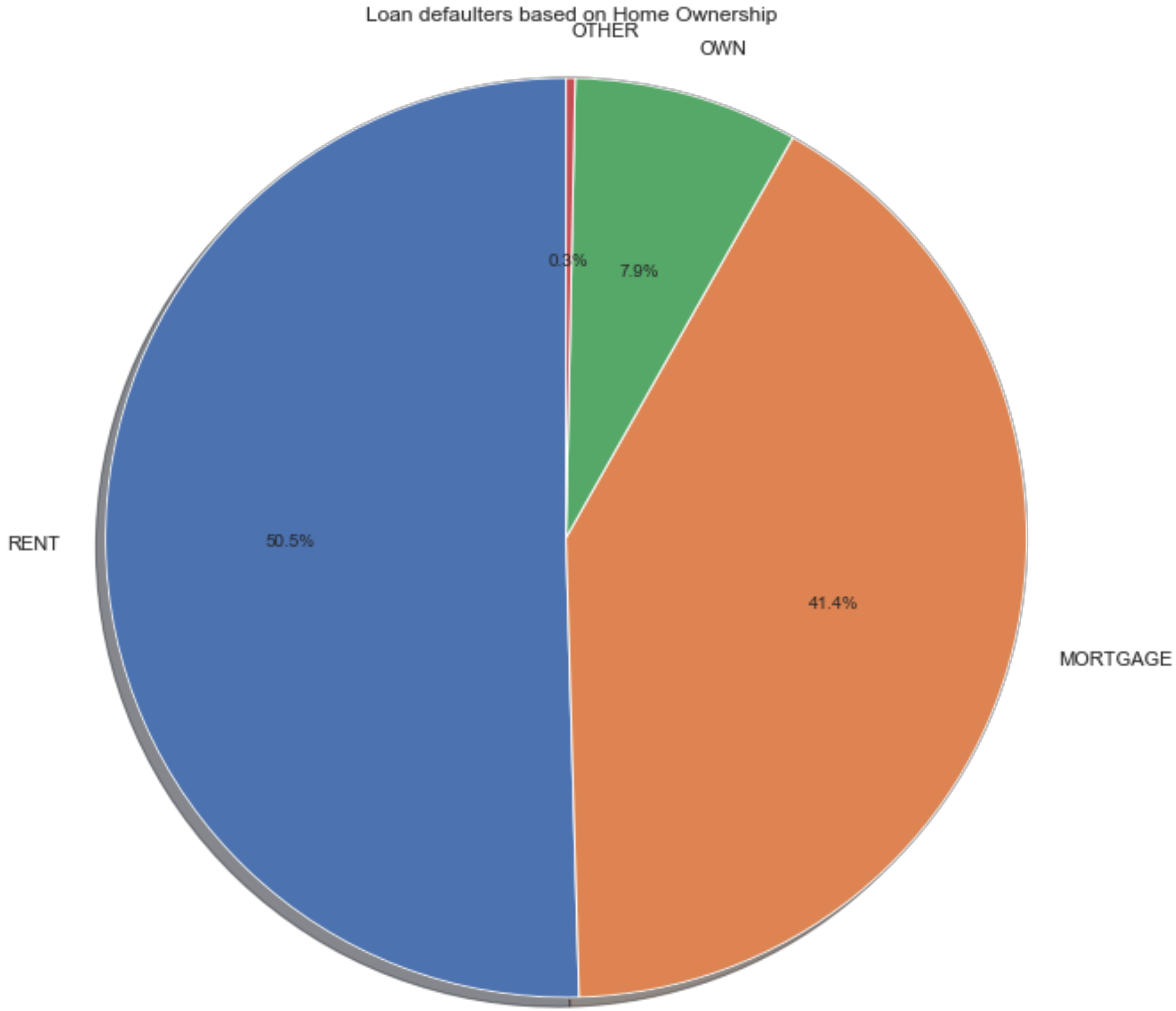
- Grades B,C,D have comparatively higher number of loan defaulters than grades E,F,G.
- So, loans corresponding to grades B,C,D are at a higher financial risk.

Charged off accounts frequency based on work experience



Loan defaulters are much higher for employees with more than 10 years of experience

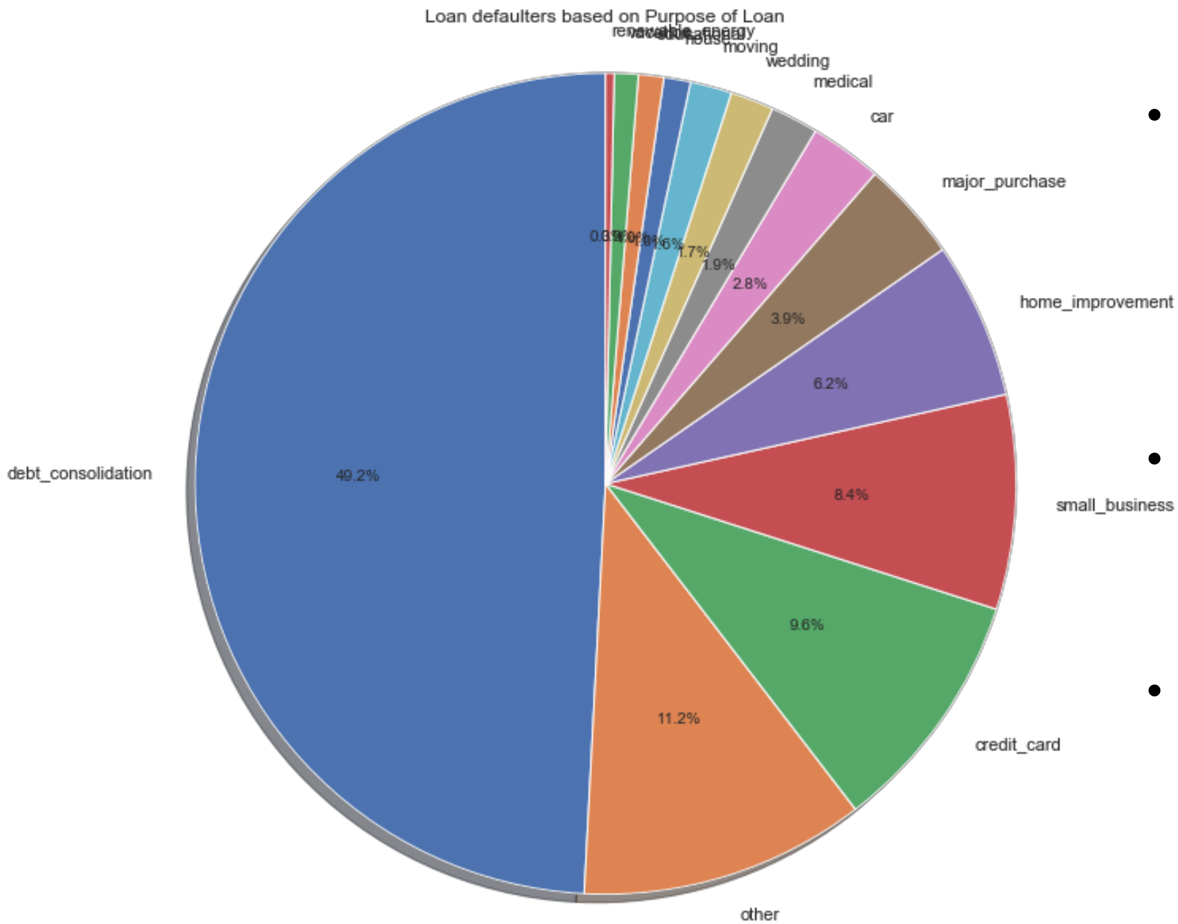
Loan defaulters based on Home Ownership



- People who rent homes are more likely to default on loans than the people who mortgage

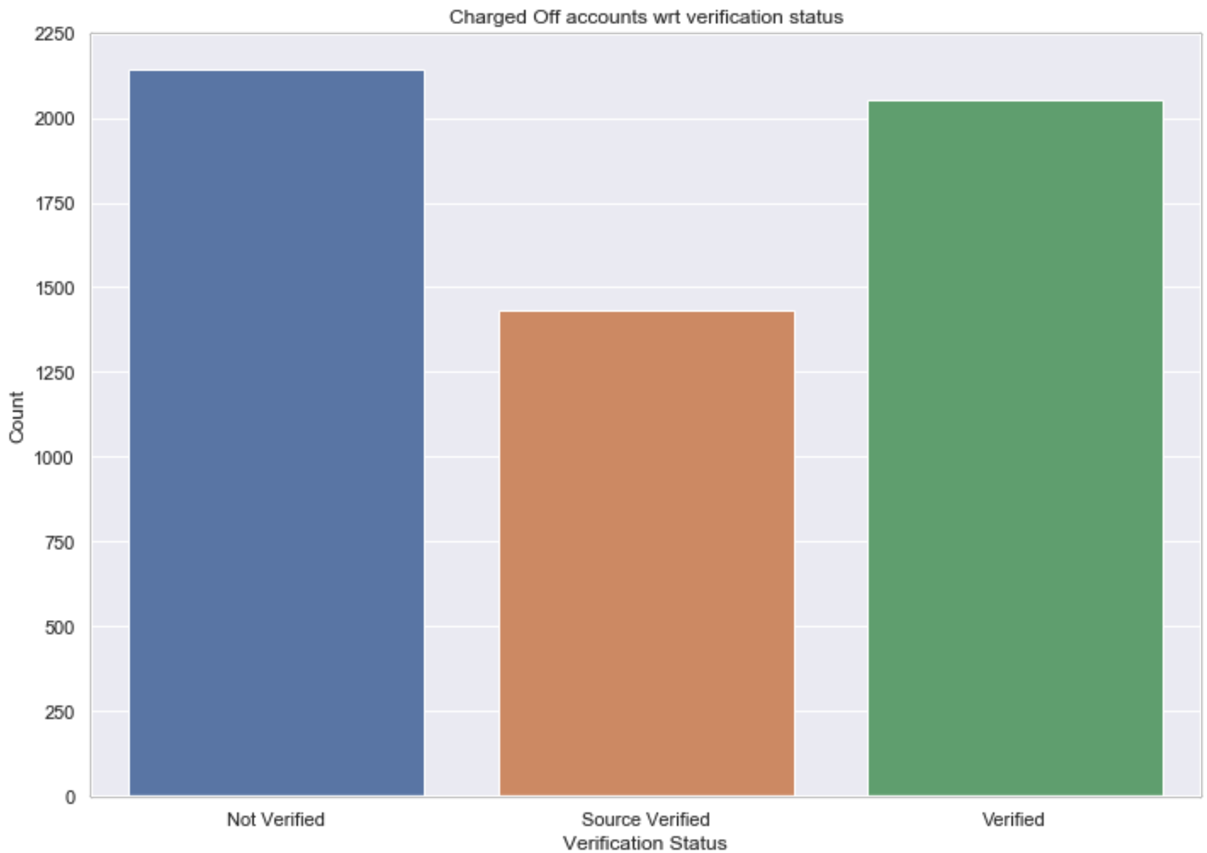
- People who own a house are less likely to default on a loan payment

Loan defaulters based on Purpose of Loan



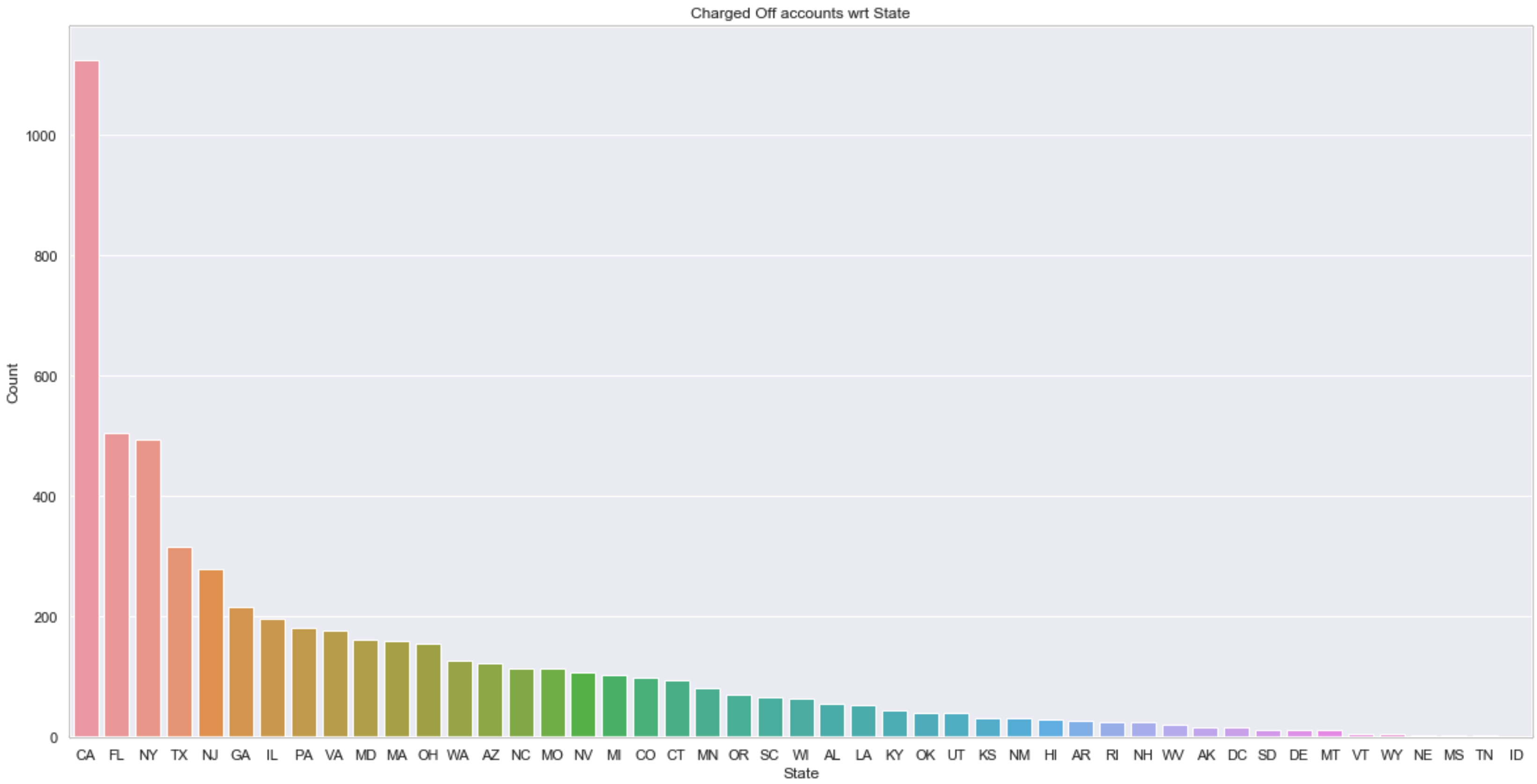
- Debt consolidation means taking out a new loan to pay off a number of liabilities and consumer debts, generally unsecured ones.
- People who take loan for Debt Consolidation, are more likely to default on payment.
- The next obvious thing is credit card payment, where most customers default on payment.

Analyzing Charged Off accounts based on Verification status



- It is interesting to observe that even after the verification of the loan application's background, the loan is being approved to them.
- Due to the carelessness of the bank employees, this leads to a significant financial loss.

Charged Off accounts based on State

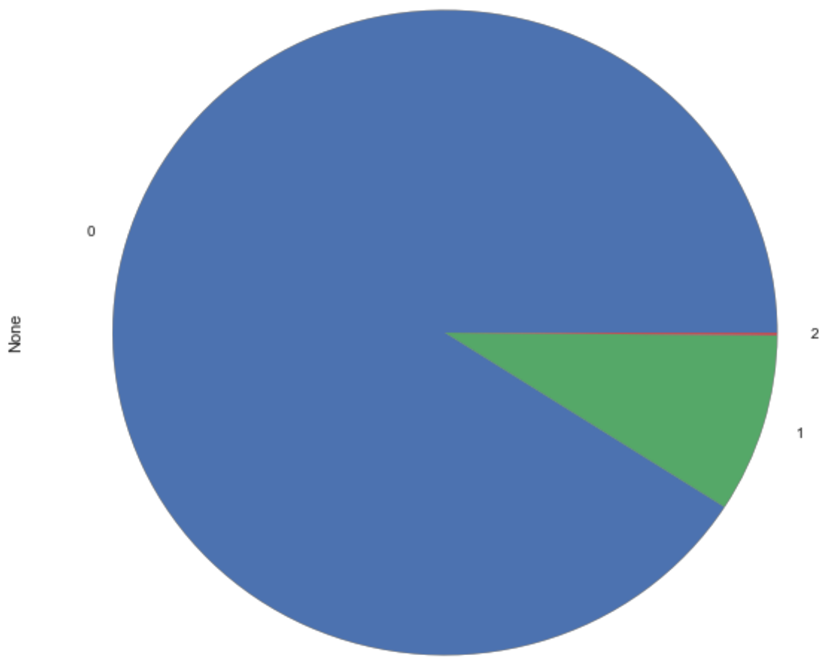


- Apparently California, Florida, New York are leading in loan defaulters

Insights on Applicants Experience (>10)

- We see the applicants who 10 and more than that are likely to default their loans, lets see some interesting insights.
- Applicants with experience 10 more than that in below top 5 states are likely to default their loans.
- Applicants with above criteria who has ZERO public derogatory public records are more likely default their loans.
- Applicants with above criteria who defaults their loans in the year 2011,2010,2009 top 5 state wise.

Defaulters emp_exp > 10 vs derogatory public records

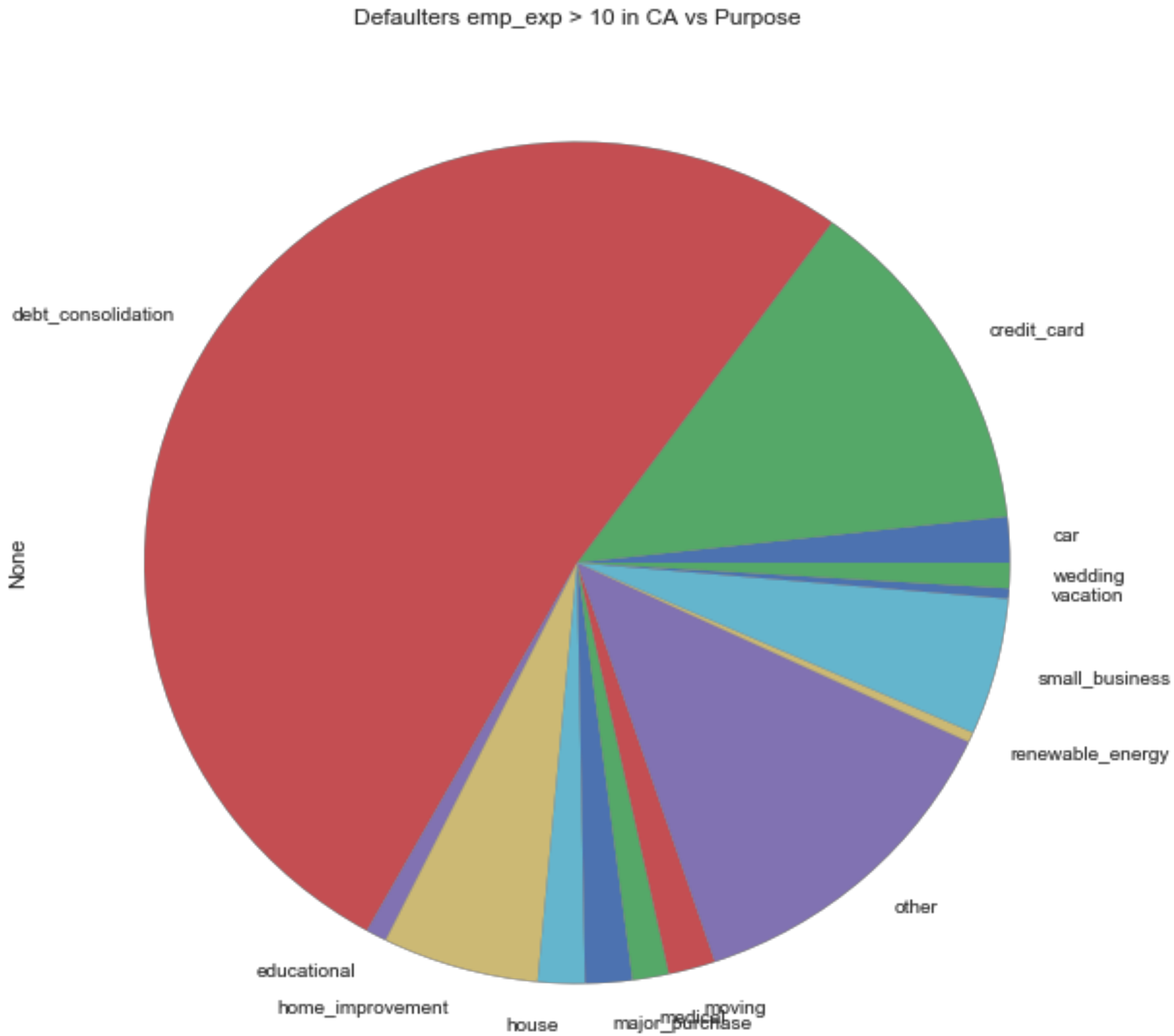


Issue_yr	addr_state	count
2010	CA	78
2010	NY	40
2010	FL	37
2010	TX	19
2010	NJ	15

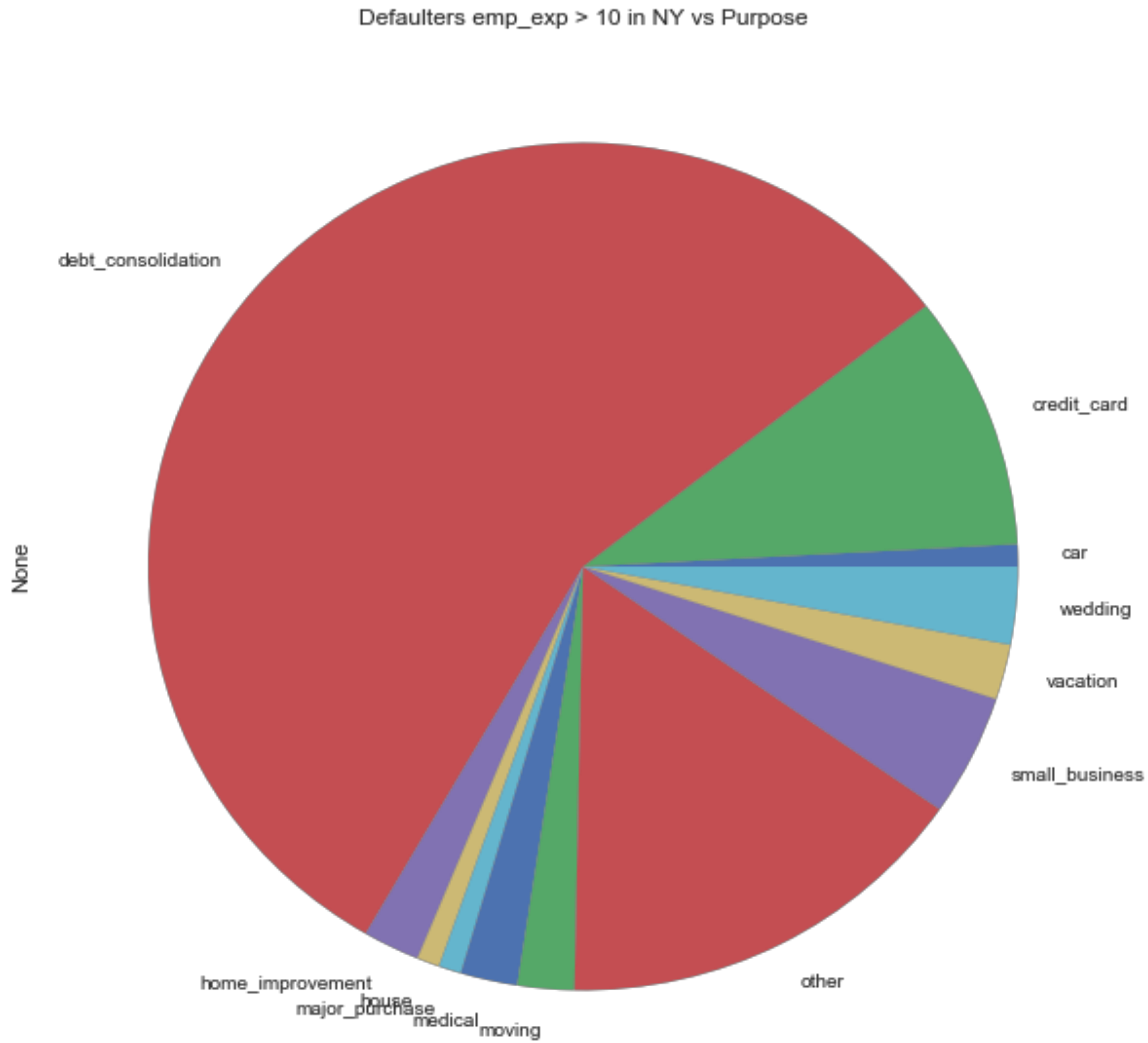
Issue_yr	addr_state	count
2009	CA	20
2009	NY	6
2009	TX	6
2009	FL	6
2009	NJ	5

Issue_yr	addr_state	count
2011	CA	190
2011	NY	88
2011	FL	82
2011	TX	52
2011	NJ	43

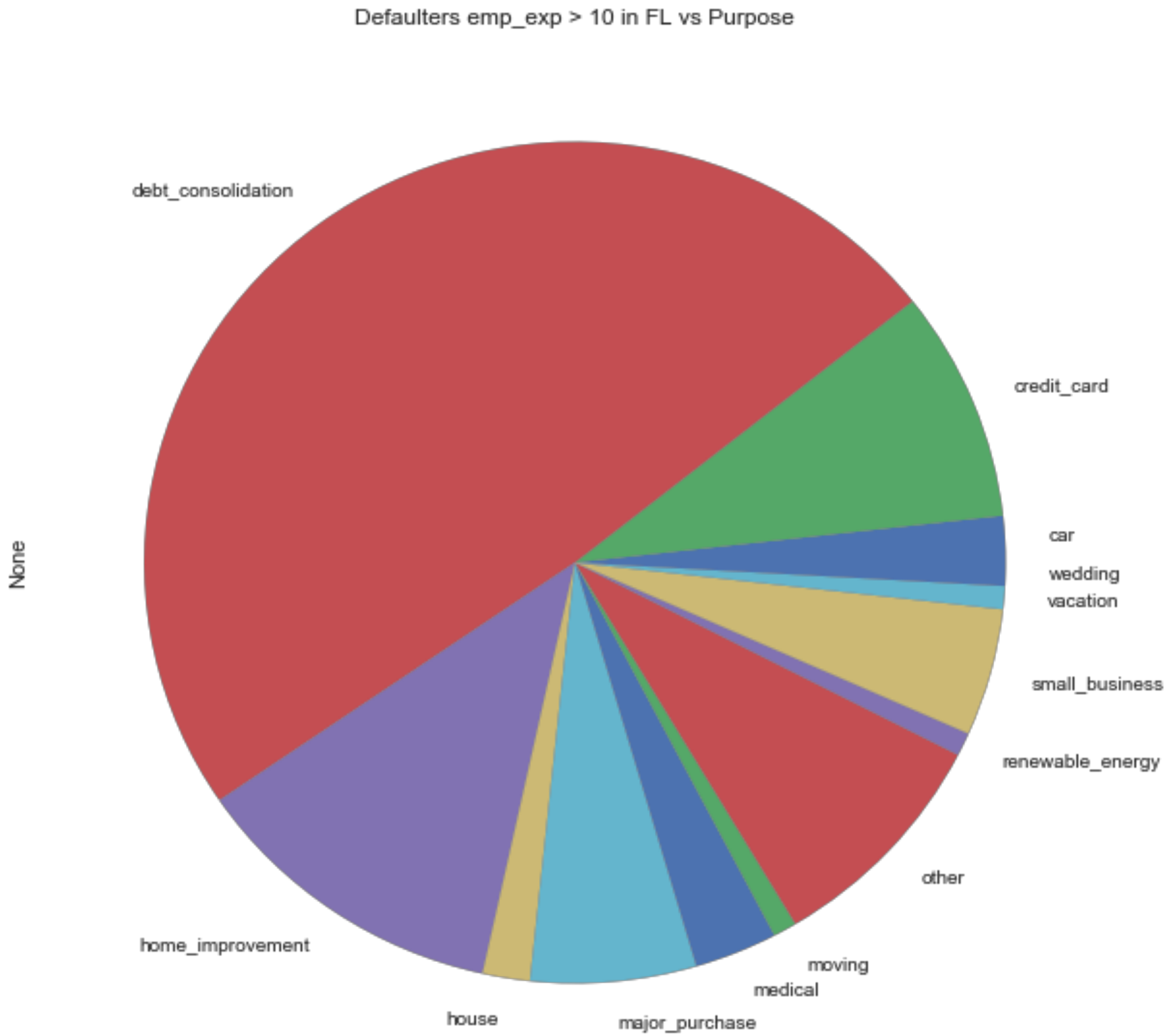
Defaulters : Employee Experience > 10 in CA vs Purpose



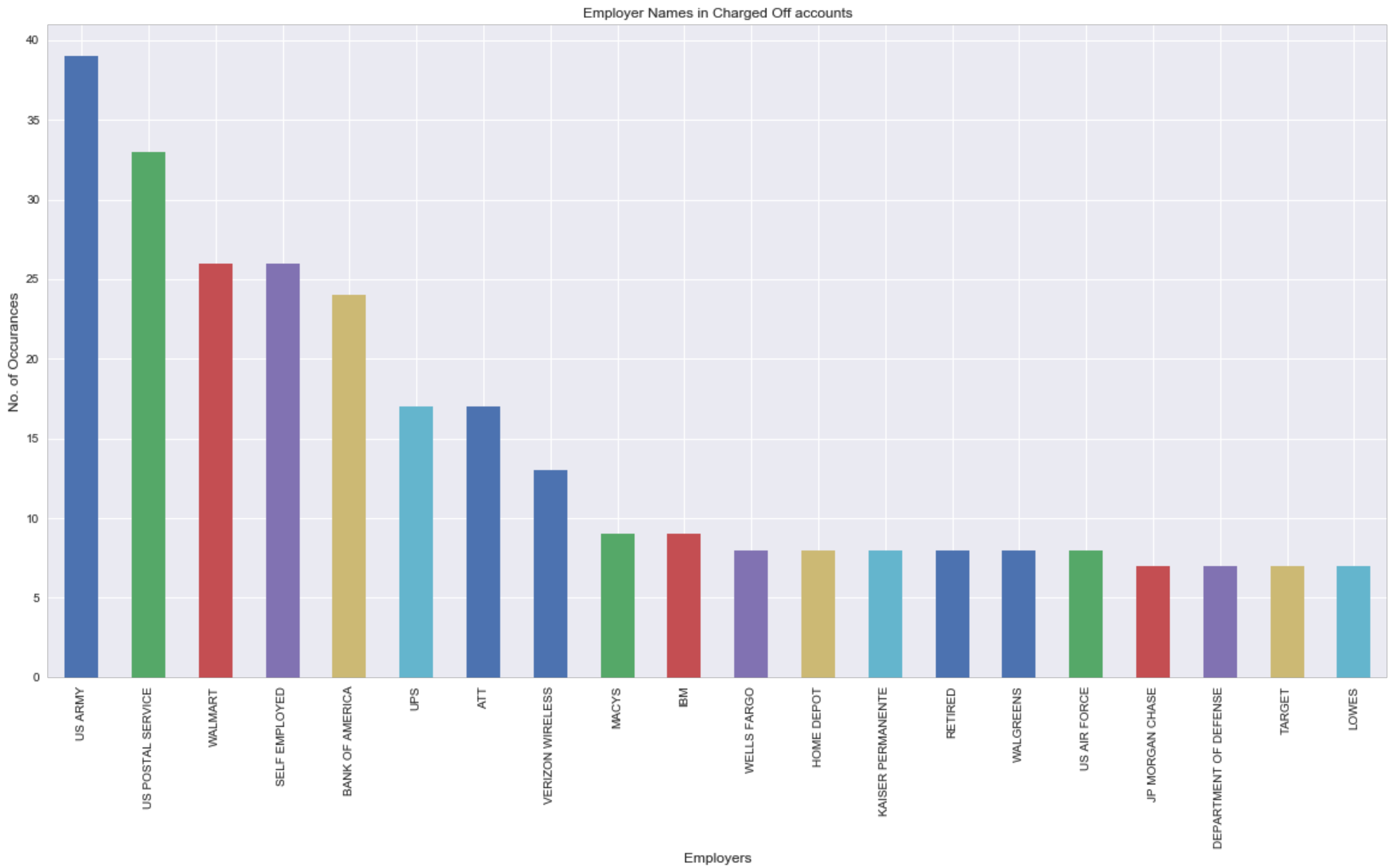
Defaulters : Employee Experience > 10 in NY vs Purpose



Defaulters : Employee Experience > 10 in FL vs Purpose

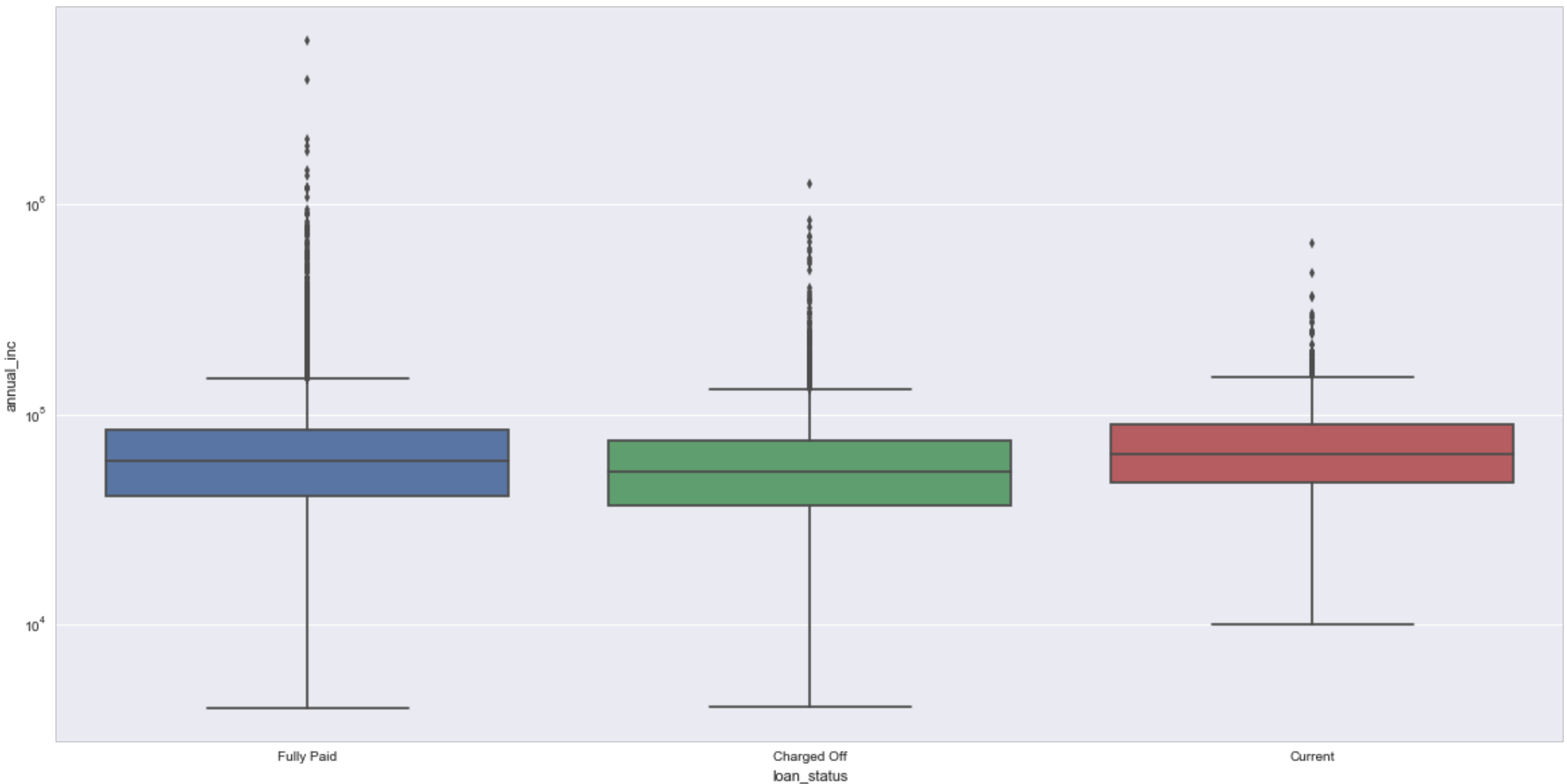


Employer Names in Charged Off accounts



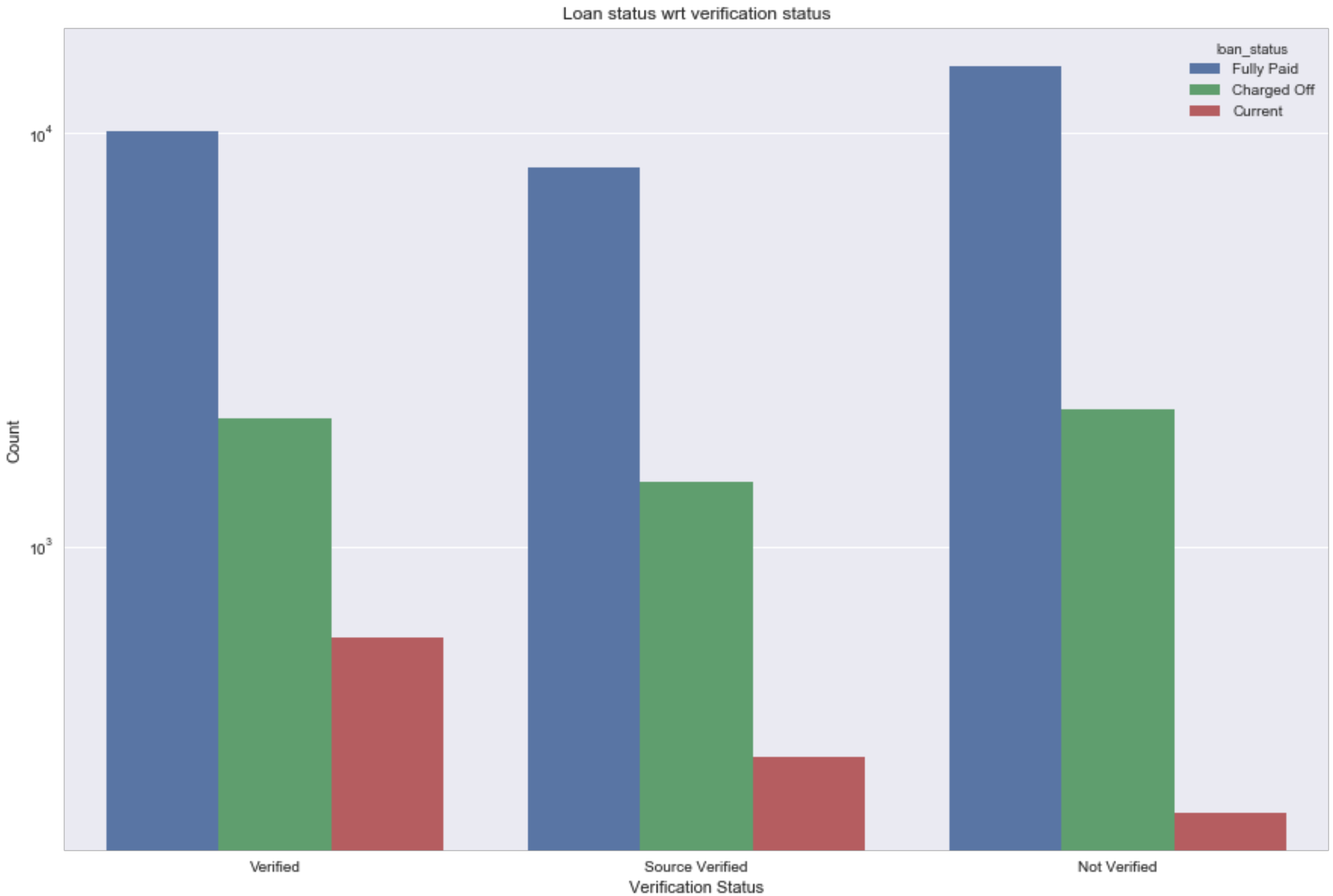
- It can be observed that loan applicants belonging to US ARMY, US POSTAL SERVICE, WALMART and those who are SELF-EMPLOYED are more likely to default on loans.

Analysis of Annual Income with respect to Charged Off accounts



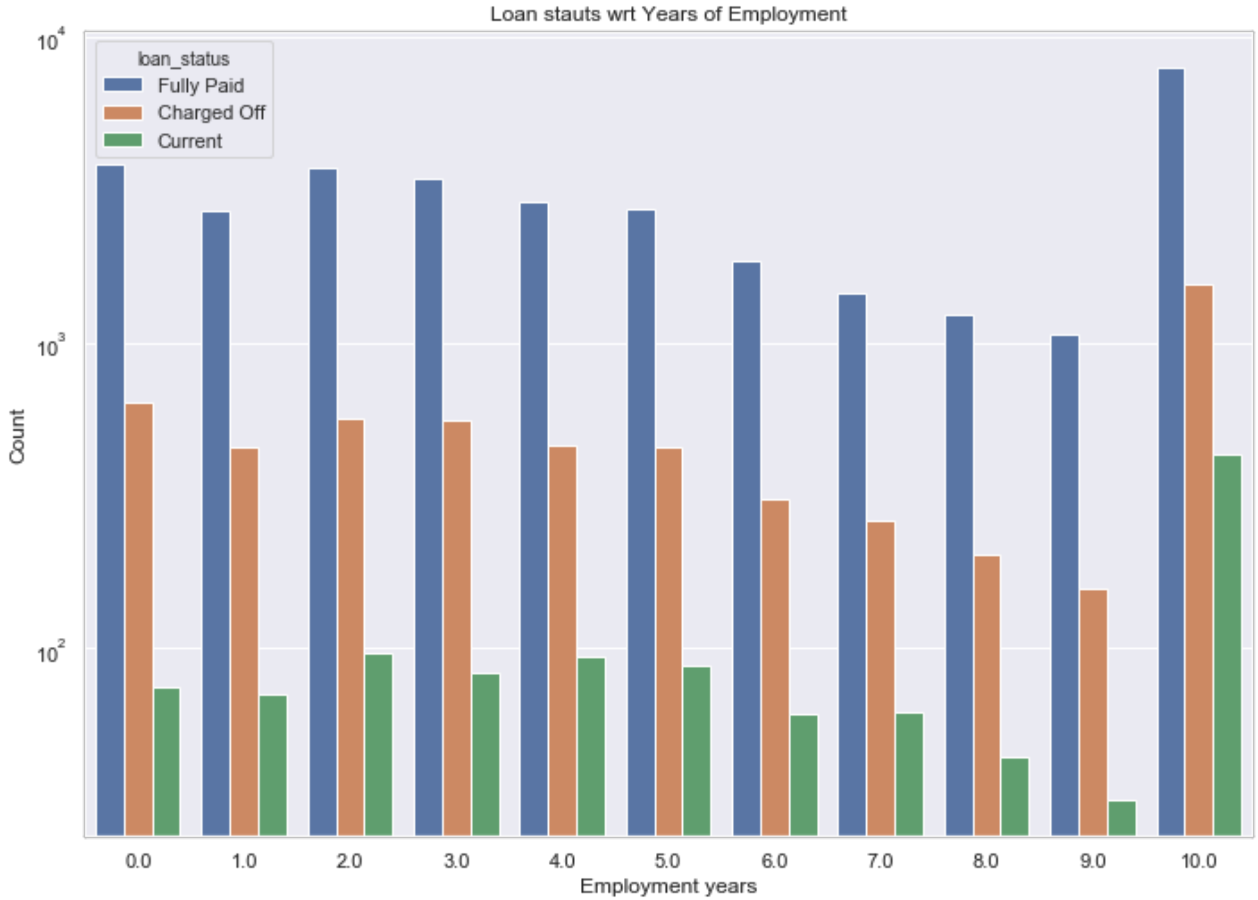
- Considering that the Y-axis is taken on a log scale, the mean annual income of the loan defaulters tend to be lower than the mean annual income of the one who paid off.

Loan status wrt verification status



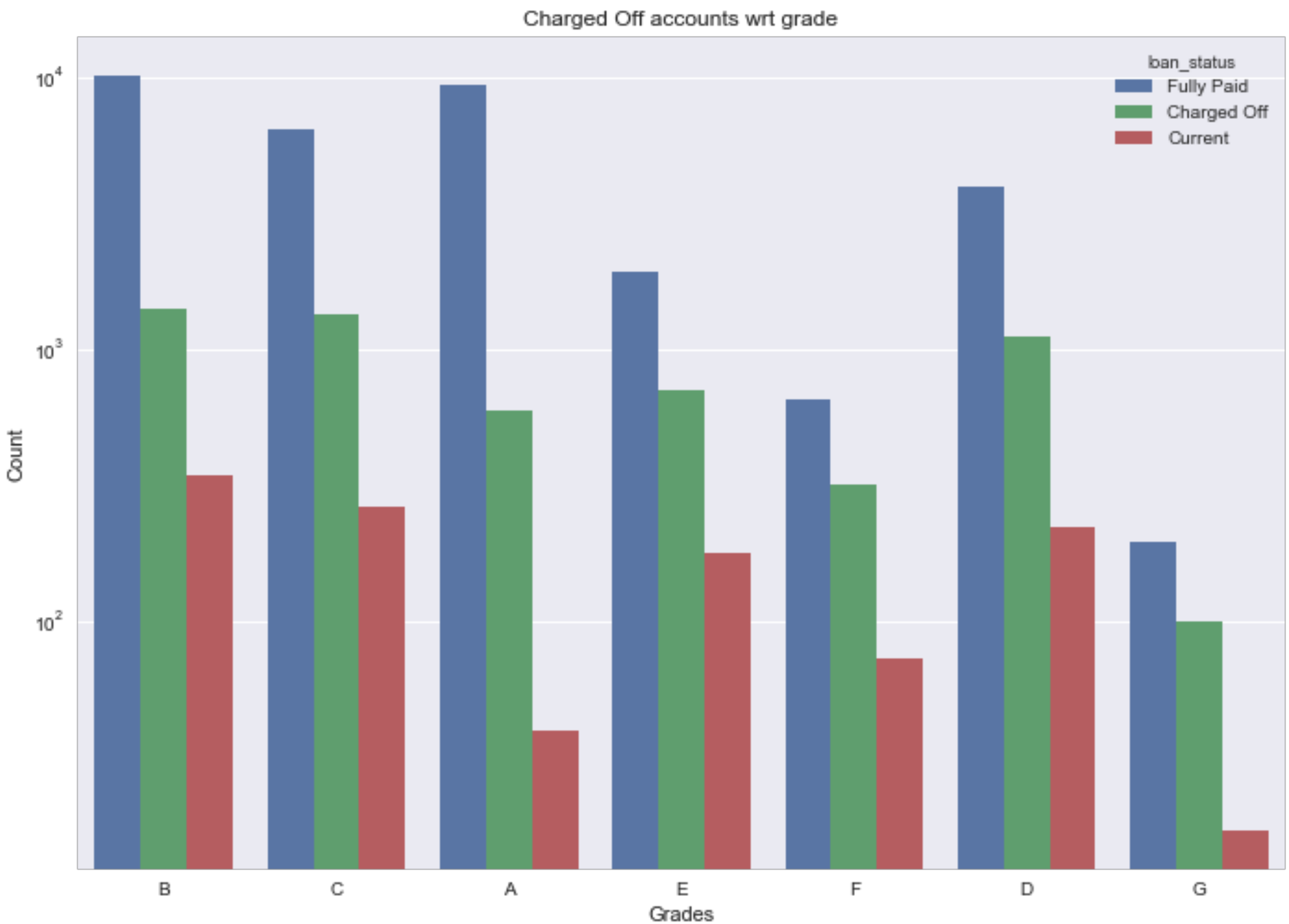
- It can be observed that the non-verified loan accounts are more likely to charge off than the verified ones. So the banks have to adopt stringent verification norms.

Segmented Univariate Analysis: Loan status wrt Years of Employment

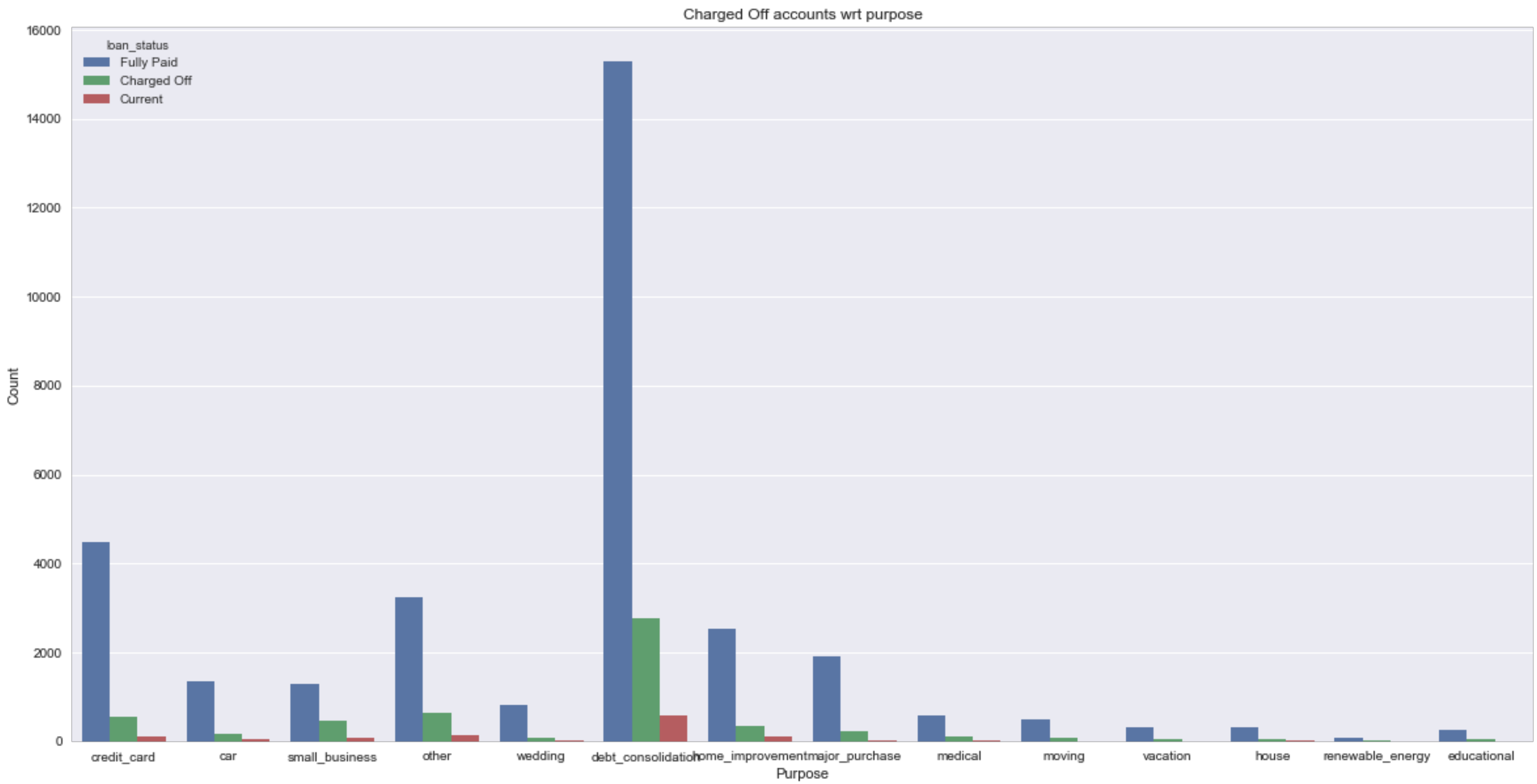


- Here, inference is that people with more than 10 years of experience are likely to default on loan payment. But it has to be considered that the 10 here implies the total aggregation of all the loan applicants who have more than 10 years of experience

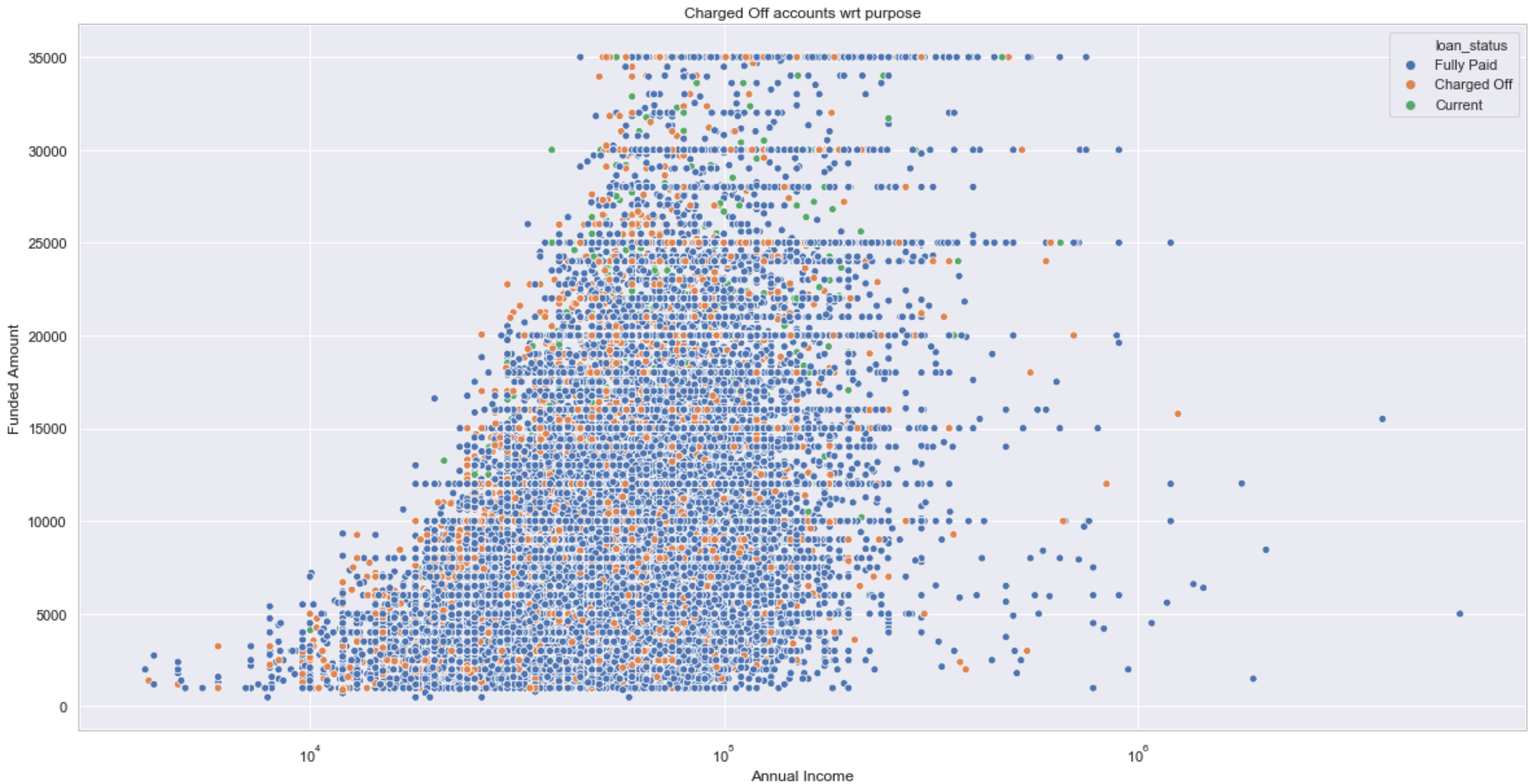
Charged Off accounts wrt grade



Charged Off accounts wrt purpose

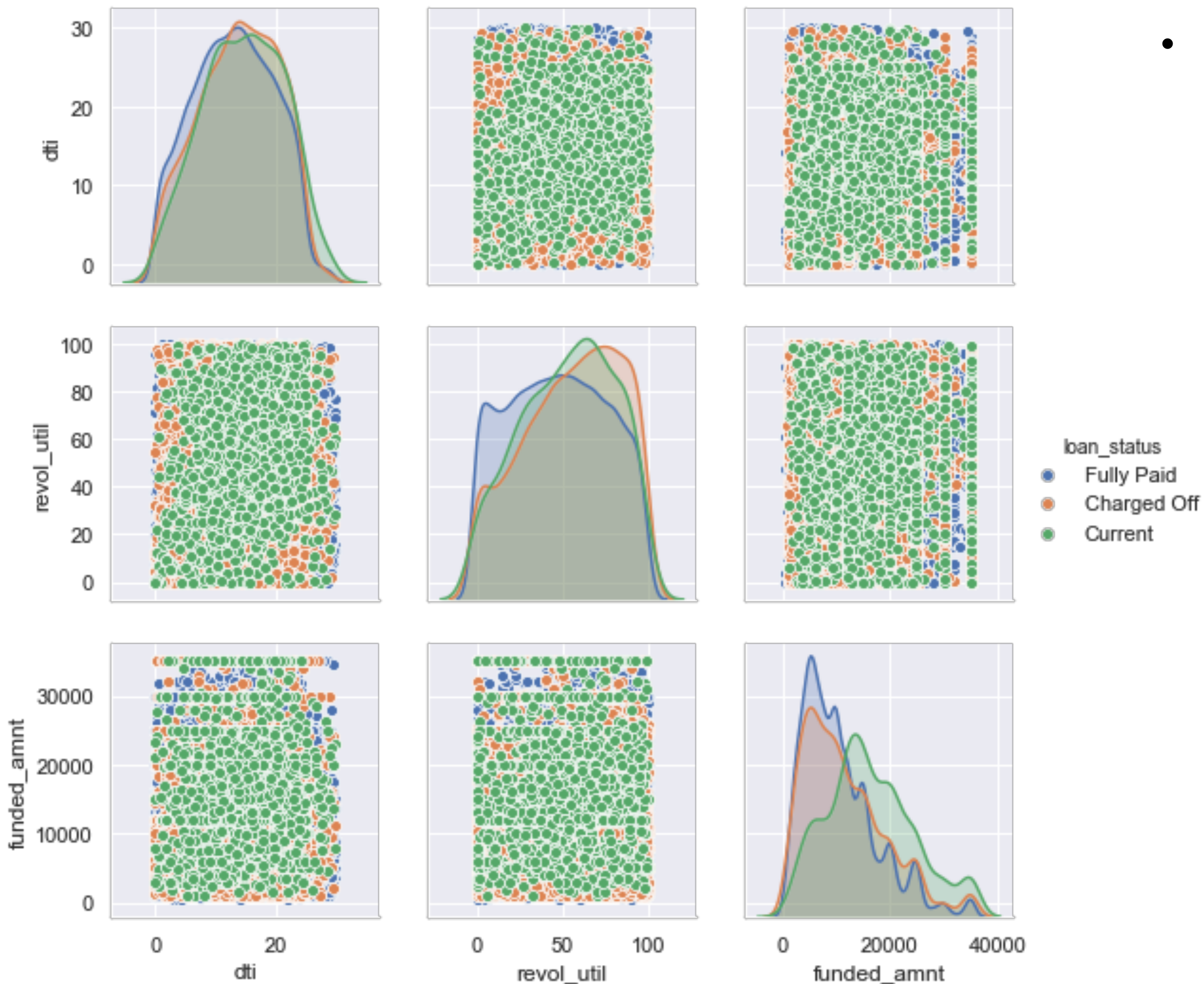


Bivariate analysis of Annual Income Vs Funding Amount



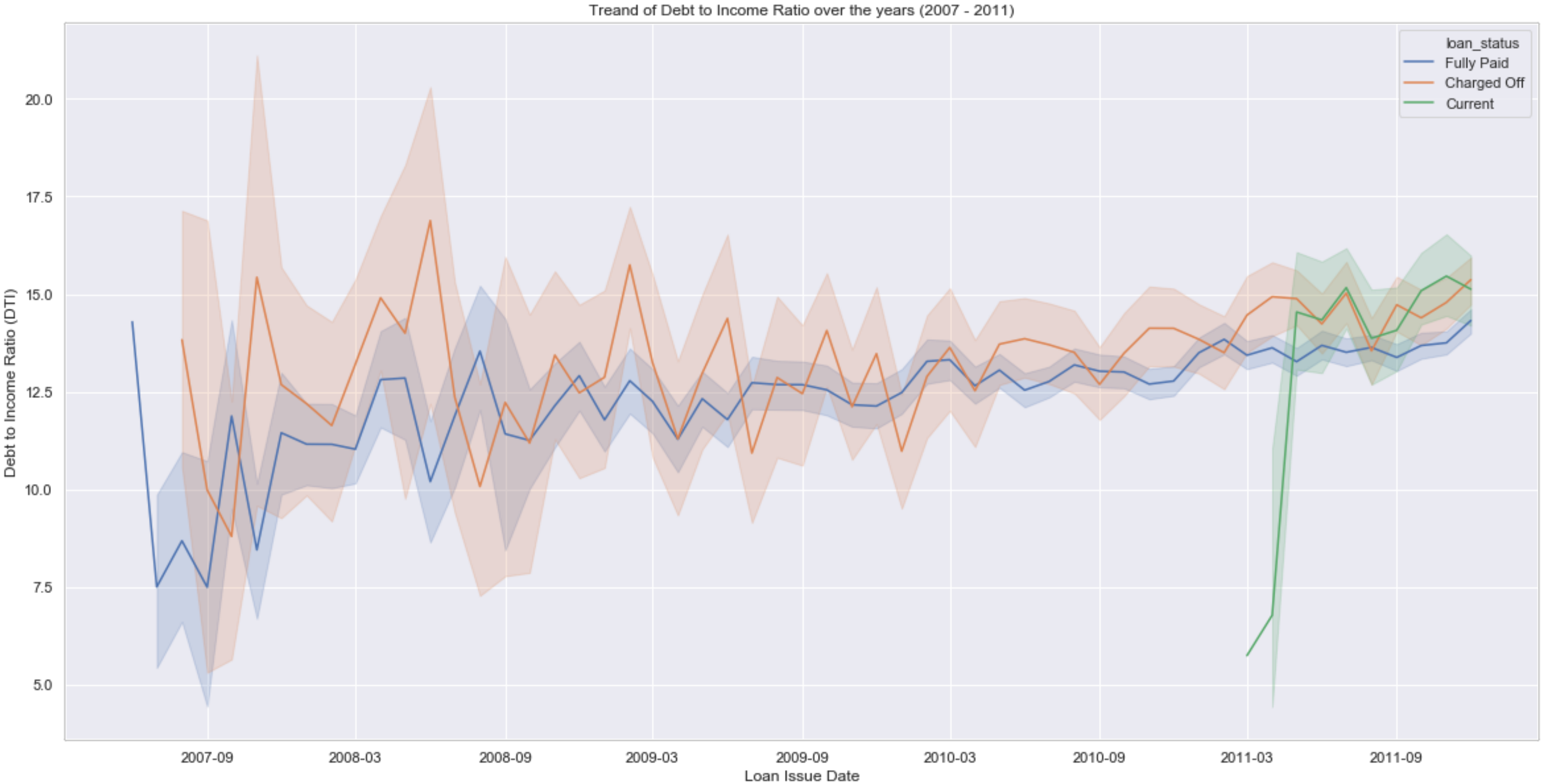
- Here we can observe a perfect correlation between the annual income and the funding amount, which highlights the fact that banks require certain annual income in order to approve a funding of the corresponding loan amount.

Correlation analysis between DTI, Revolving line utilization rate and Fund amount



- In the plot, we can observe a certain level of co-relation between the DTI, Revolving Rate and the Funding Amount.

Treand of Debt to Income Ratio over the years (2007 - 2011)



From this plot, we can conclude that the DTI has been gradually increasing with time. This means that the people are more inclined to spend more. The orange curve shows that more and more loans are getting Charged off through the years, and so the banks have to adopt stringent policies over approving loans

Thank You