Loan Default Gramener Case Study

SUBMISSION

Team

- Krishnan Raghupathi
- Nanditha GN
- Srividya Ravichandran
- Abhijith N V

Key Objectives

- Identify driving factors that influence loan default
- Identify driver variables which are a strong indicators of default
- Ensure that the inferences are
  - Data driven
  - Explained through visualizations
  - Inferences are made based on correct trends

What Data we have

- Loan data for all loans issued through the time period 2007 to 2011from lending club
- Loans rejected by lending club data is not available in the data set
- Loans have 3 statuses
  - Fully Paid – Loans that are closed
  - Current – Ongoing loans
  - Charged Off – Loans that are defaulted by the customers
- Few key columns and our understanding of those columns are listed in the next slide

Column Name	Explanation
collections_12_mths_ex_med	Number of collections in 12 months excluding medical collections. Can be ignored, mostly NA/0 values
funded_amnt	Amount that was funded by the investors. In few cases this will be lesser than the Loan Amount, the remianing amount is problably lent by LC dirctly
int_rate	Interest Rate on the loan
last_pymnt_amnt	Last total payment amount received. This will be 0 for Charged off
last_pymnt_d	Last month payment was received
loan_status	Current status of the loan
mths_since_last_major_derog	Months since most recent 90-day or worse rating
mths_since_last_record	months since the last (derogatory) public records
open_acc	The number of open credit lines in the borrower's credit file. Ranges from 2 - 44
out_prncp_inv	Remaining outstanding principal for portion of total amount funded by investors
pub_rec	Number of derogatory public records.Another piece of data Lending Club provides is the number of public records on file. This is pulled from the credit report, so don’t confuse it with criminal public records. Credit report public records are things like bankruptcy, foreclosure, tax liens, and past-due child support.
recoveries	Recovery after defauls
revol_util	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
sub_grade	The majority of loans is either graded as B or C—together these correspond to more than 50% of the loan population.While there is a considerable amount of A graded or “prime” loans (~17%), there is a small amount of G graded, or “uncollectible” loans (~0,06%).
term	loan term. Only 2 term is available 36 months and 60 months
total_pymnt	Payments received to date for total amount funded
total_rec_int	Interest received to date
verification_status	Not verified, Verified (by LC) and Source verified are the options

Data cleaning

Column Name	Explanation
earliest_cr_line	Converted tot date time. Correcting dates in future to 1900 dates
issue_d	Converted tot date time
last_credit_pull	Converted tot date time
last_pymnt_d	Converted tot date time
next_pymnt_d	Converted tot date time
int_rate	Stripped % text and converted to float
revolv_util	Stripped % text and converted to float

Data we derived

Available column
- DTI
- Revolving Utilization
- Interest rate

Derived Columns
- DTI Bin
- Revolving Utilization
- Interest rate

Reason
- To understand patterns by specific bins

Available column
- Annual Income
- Loan Status

Derived Columns
- Annual Income Bin
- Loan Default

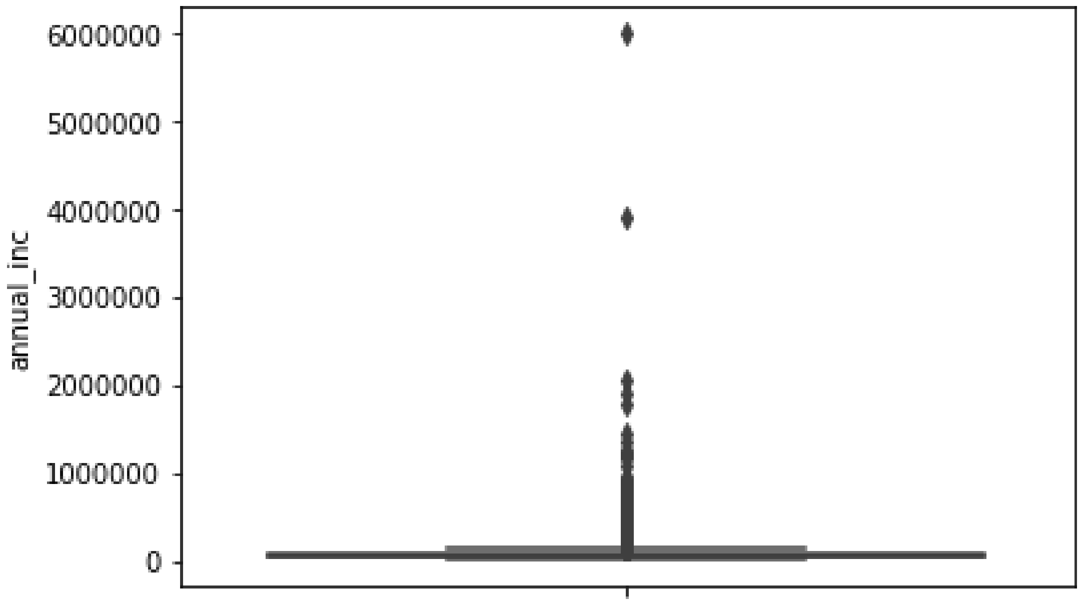
Reason
- Annual Income bins to understand patterns by specific bins
- Loan Default will have 1 for Charged Off and 0 for Fully Paid and Current. This will help in aggregating data

Analyses & Inferences

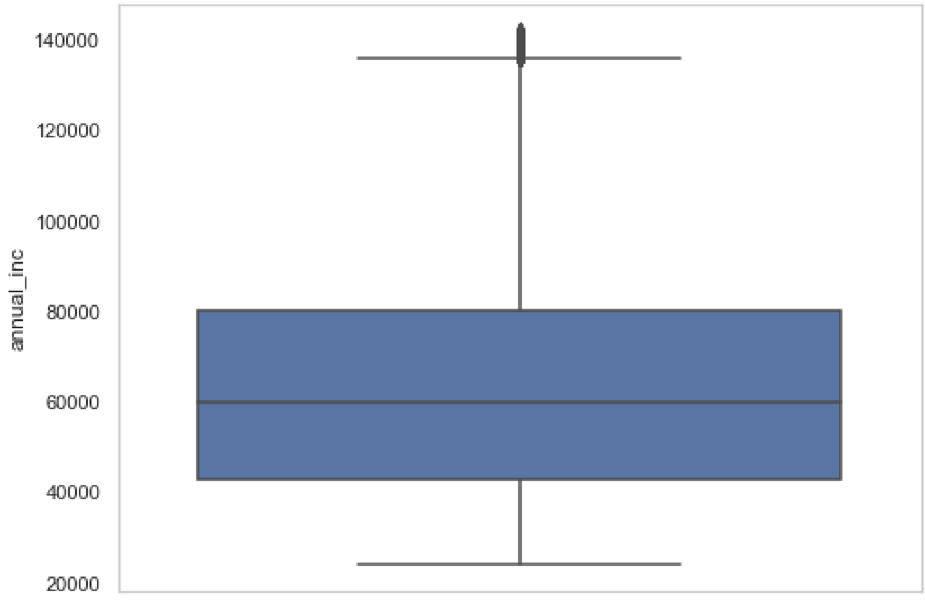
Univariate and outlier Analysis

UpGrad

Annual income distribution – With all data



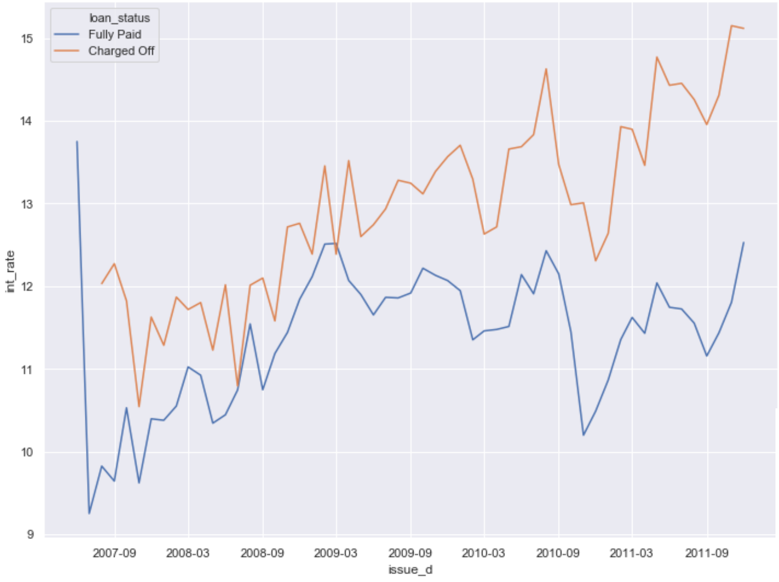
Annual income distribution – removing outliers

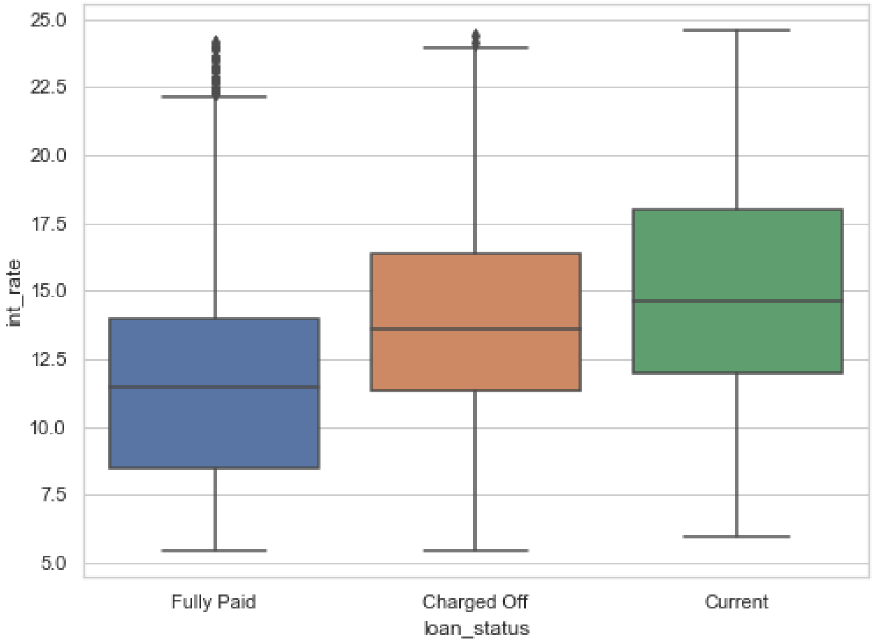


1) Univariate and Outlier analysis was performed on all important value columns and appropriate percentiles were taken to remove the outliers
2) Annual income graphs are shown abov
3) Similar exercise was done for
   1) Instalment
   2) DTI
   3) Revol_bal
   4) funded_amnt
   5) Interest rate
   6) Loan Amount

Microsoft Excel Worksheet

Rate of interest impact





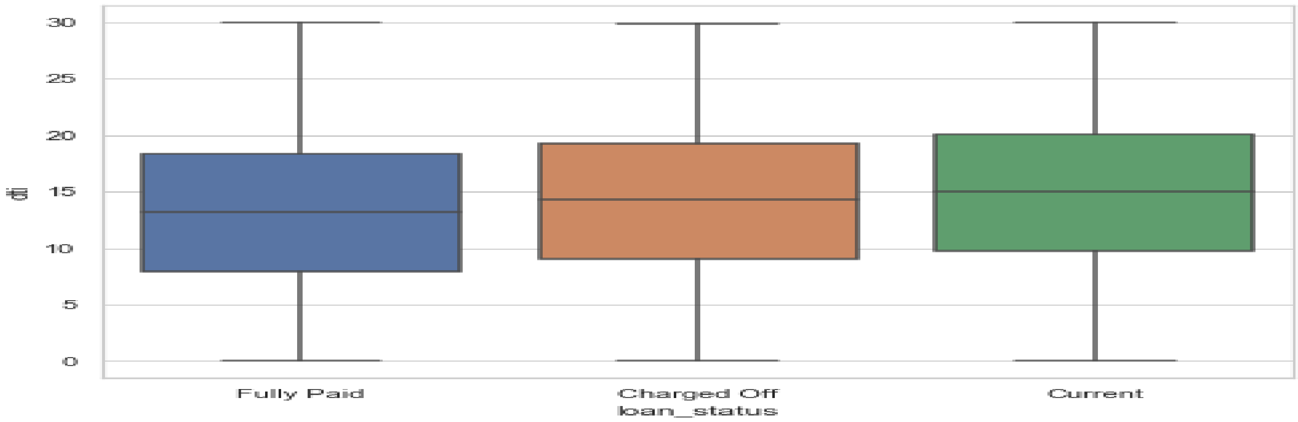
Observations:

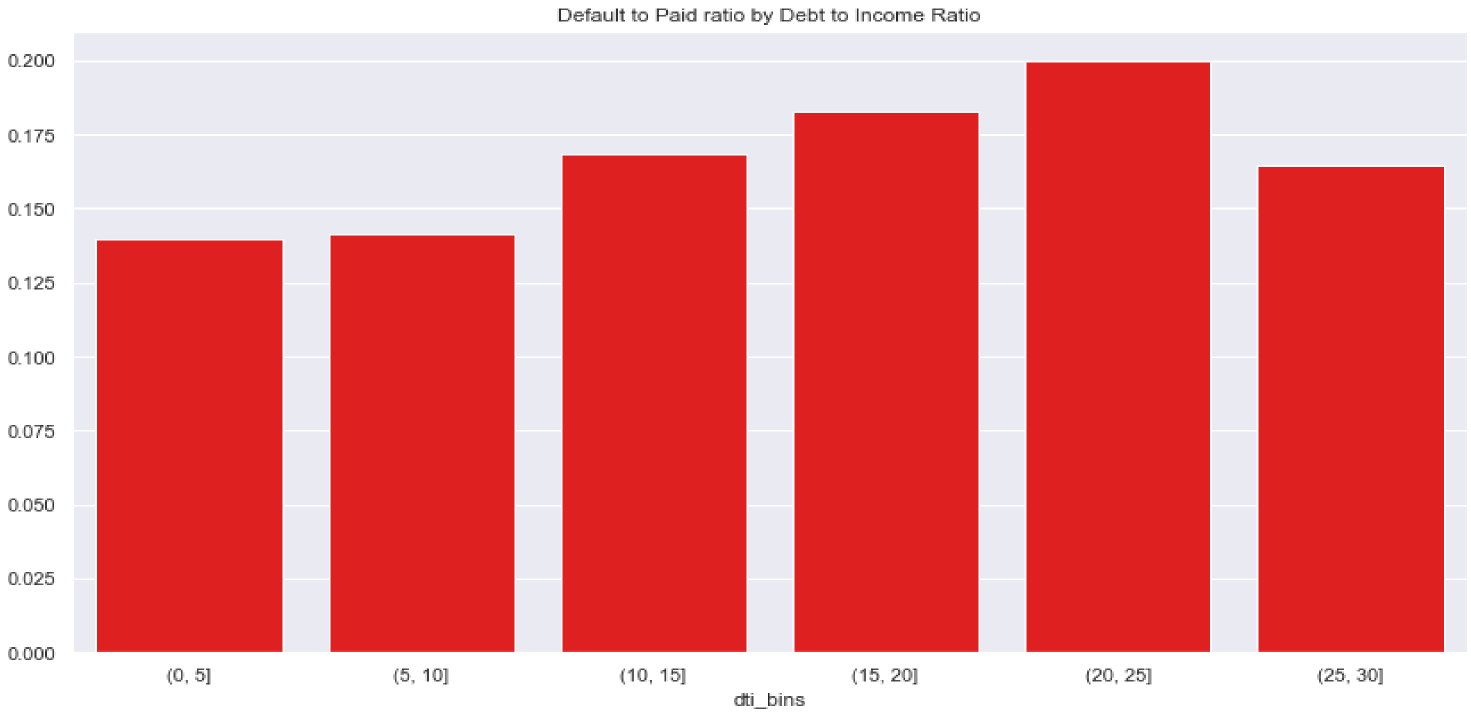
- Rate of interest for fully paid is consistently lower compared to charged off/default over the years
- Median of interest rate is higher for charged off compared to Fully paid
- Post 2010 the gap is significantly higher

Inferences:

- Higher rate of interest are guided by high risky loans
- People with lower grades tend to get higher rate of interest
- This risk factor leads to charged off

Debt to income ratio impact





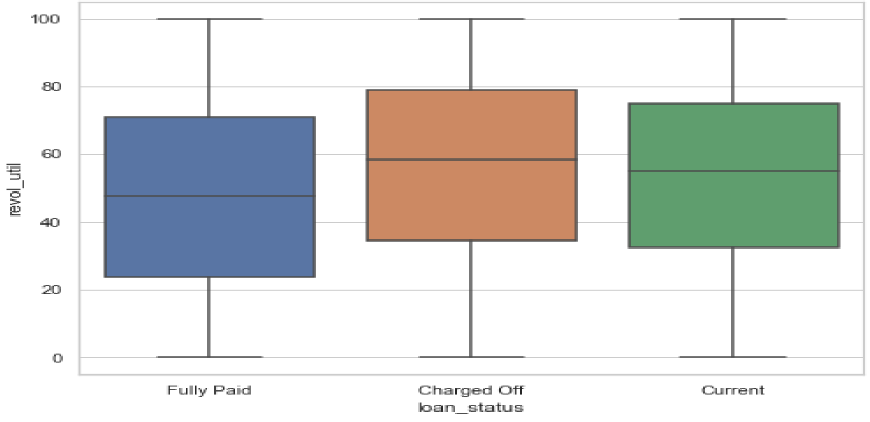
Observations:

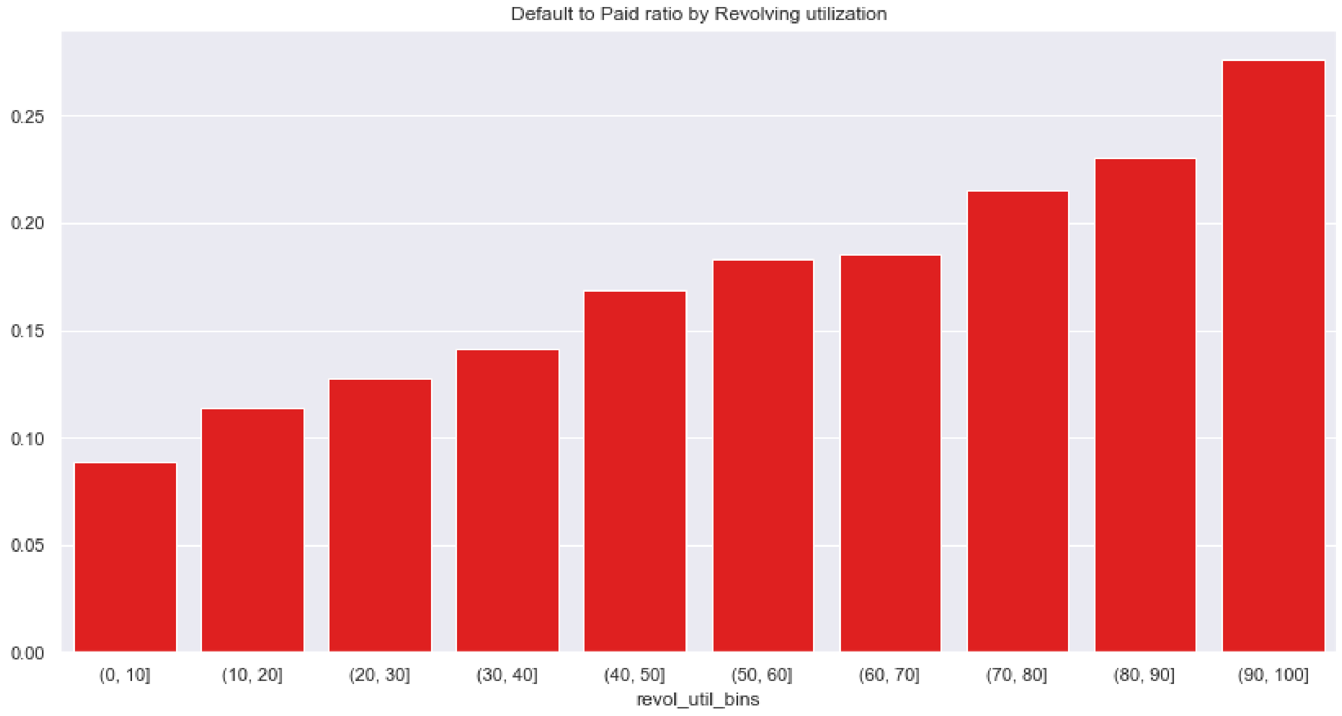
- DTI is higher for charged off compared to fully paid
- Charged off count to Fully paid count ratio is highest in the 20-25 DTI Bin

Inferences:

- Higher the debt to income ratio there is a high risk of people defaulting on the loan

Revolving utilization percentage impact





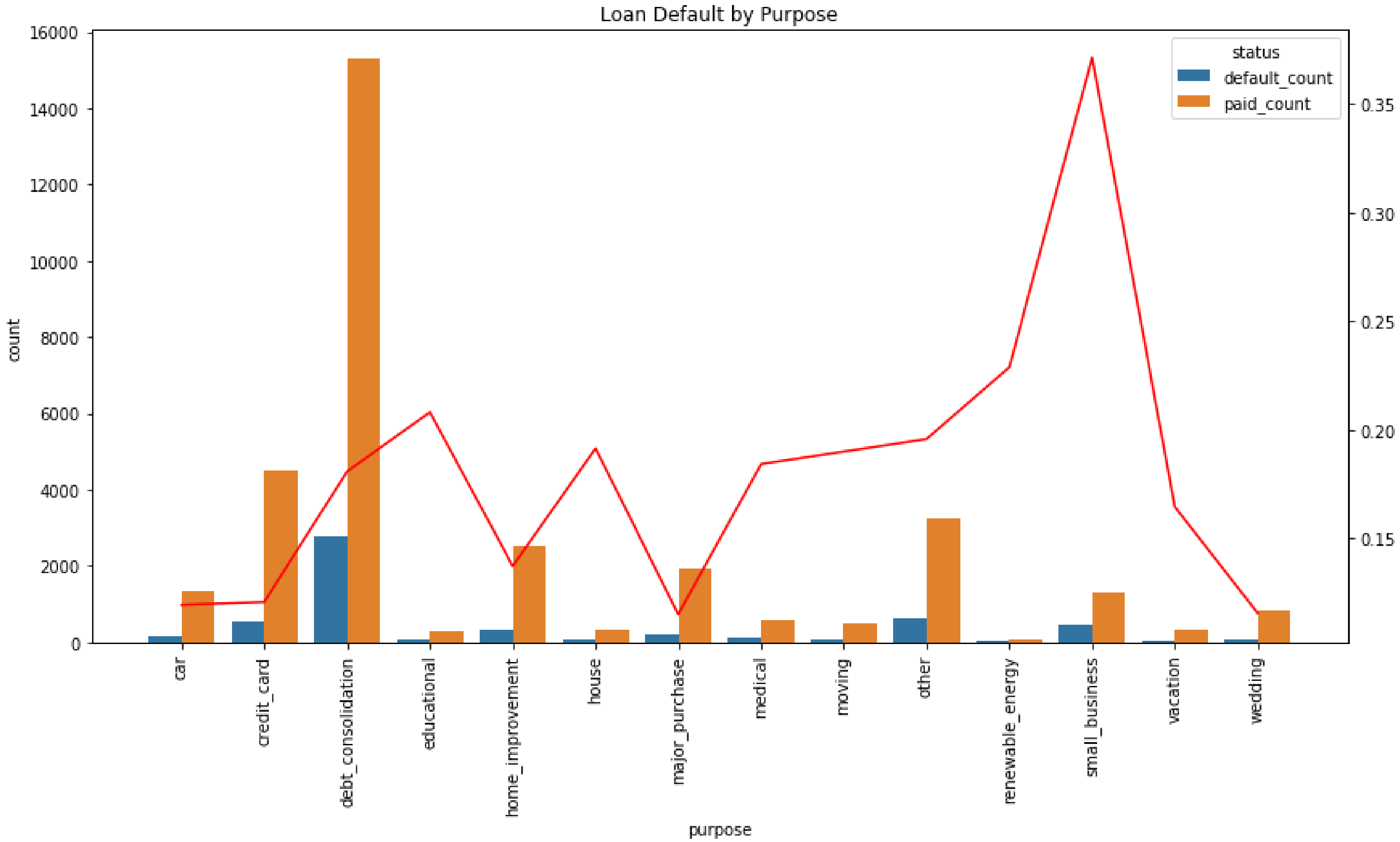
Observations:

- Revolving Utilization %ge is higher for charged off compared to fully paid
- Charged off count to Fully paid count ratio is increases as Revolving utilization %ge increases

Inferences:

- Customers having higher revolving utilizations have a high risk of defaulting on the loan

Loan purpose impact



Observations:

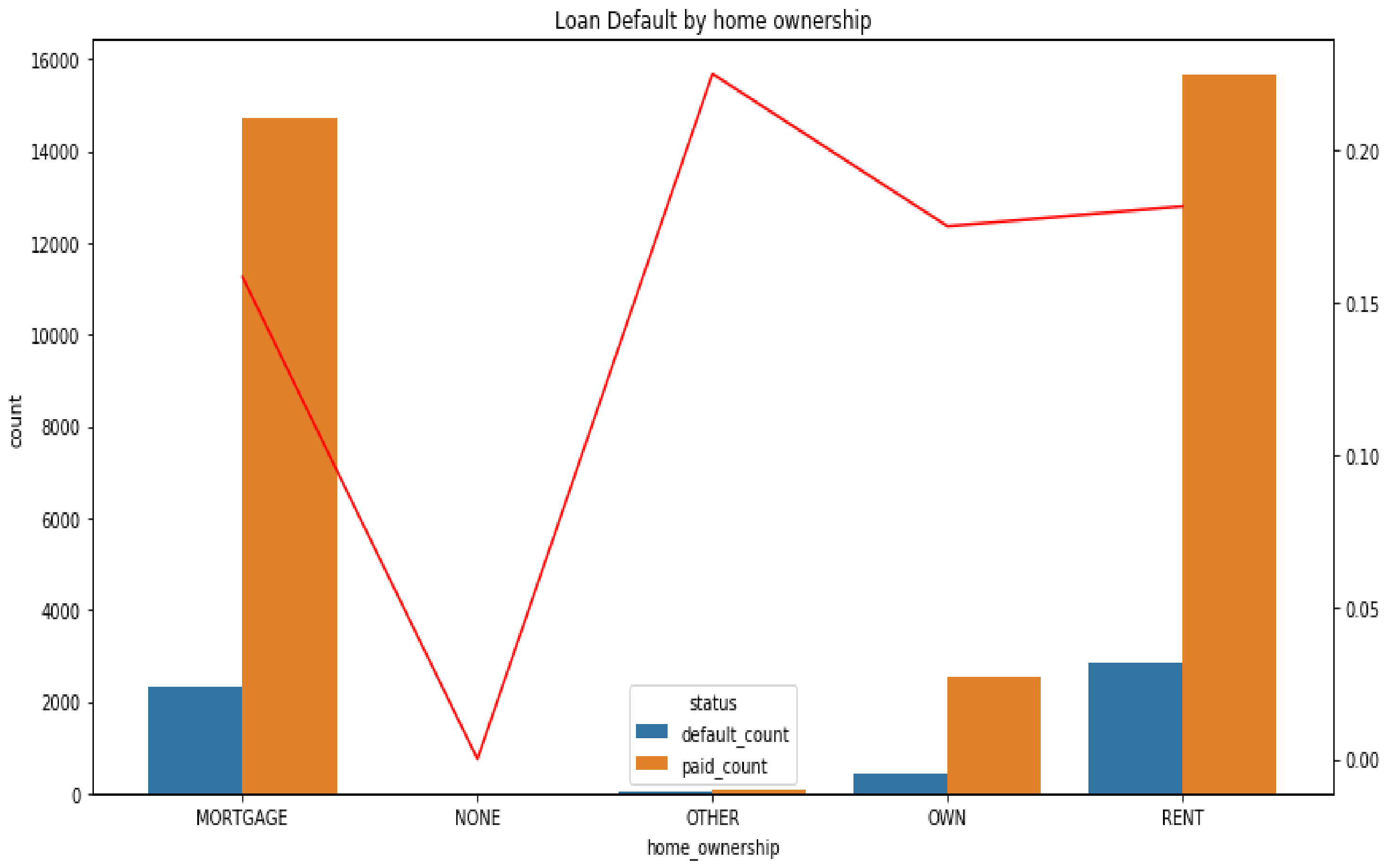
- ➢ Debt consolidation has the highest loan applications are for
- ➢ Small Business has the highest default percentage followed by renewal energy

Inferences:

- ➢ People take loan with better rate to consolidate their loans
- ➢ Small business have high risk of failure, hence loan default is also quite high

*** Red trend line shows the Charged off count to Fully paid count ratio. This is a good indicator of the likely hood of default**

Home ownership impact



Observations:

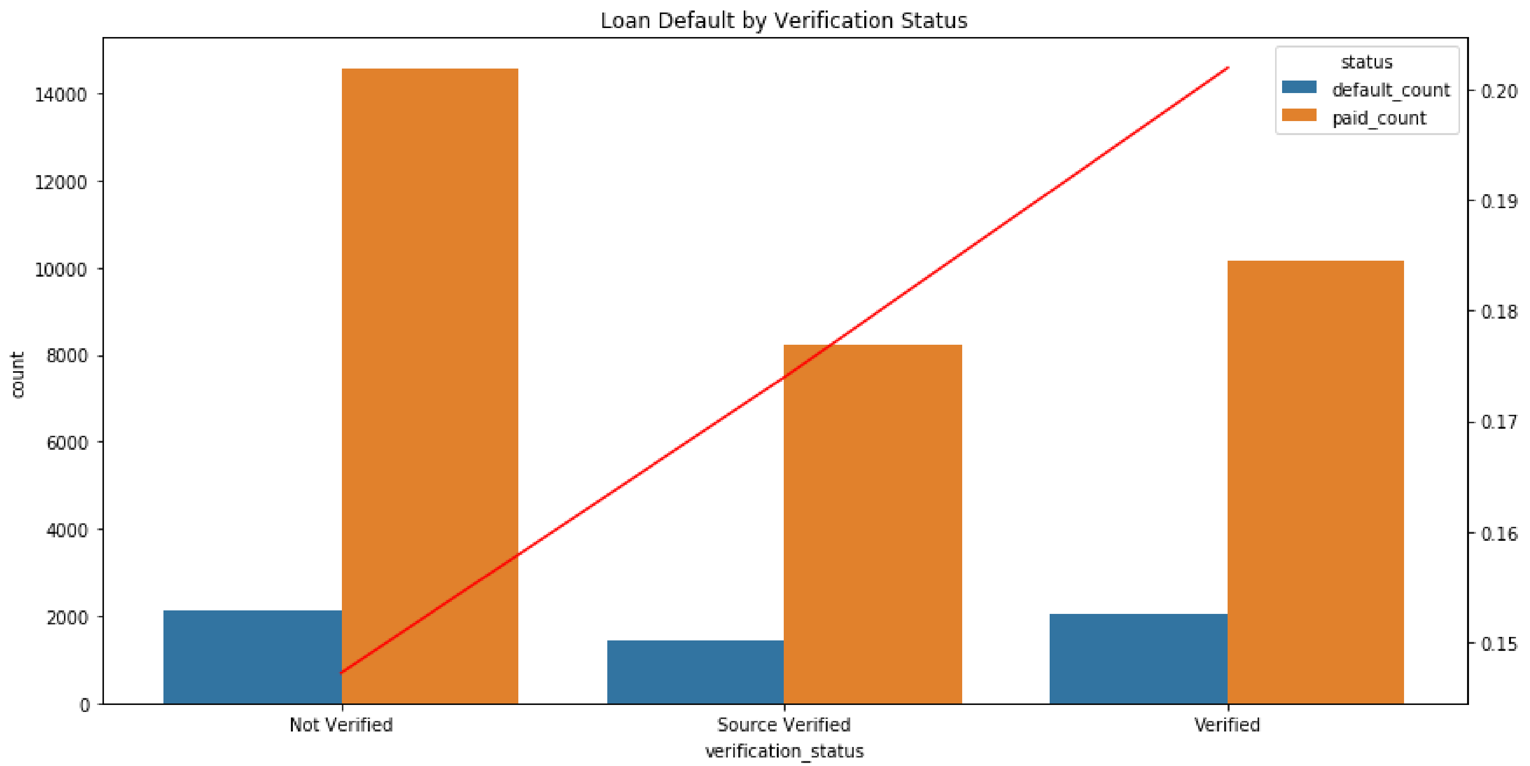
- Mortgage and rent have the highest loans given
- Between mortgage, rent and own, rented and own home owners tend to default more compared to Mortgage
- Although not many loans under others, it has a significant percentage of loan defaulting (>20%)

Inferences:

- Others is an undefined category, need more data to understand why it has more defaults
- Mortgage has lower percentage of default. Since mortgage is a major loan and any default on other loan will increase the RoI or loan category, hence Mortgage owners tends to pay back and have a less risk of defaulting

*** Red trend line shows the Charged off count to Fully paid count ratio. This is a good indicator of the likely hood of default**

Verification status impact



<u>Observations:</u>

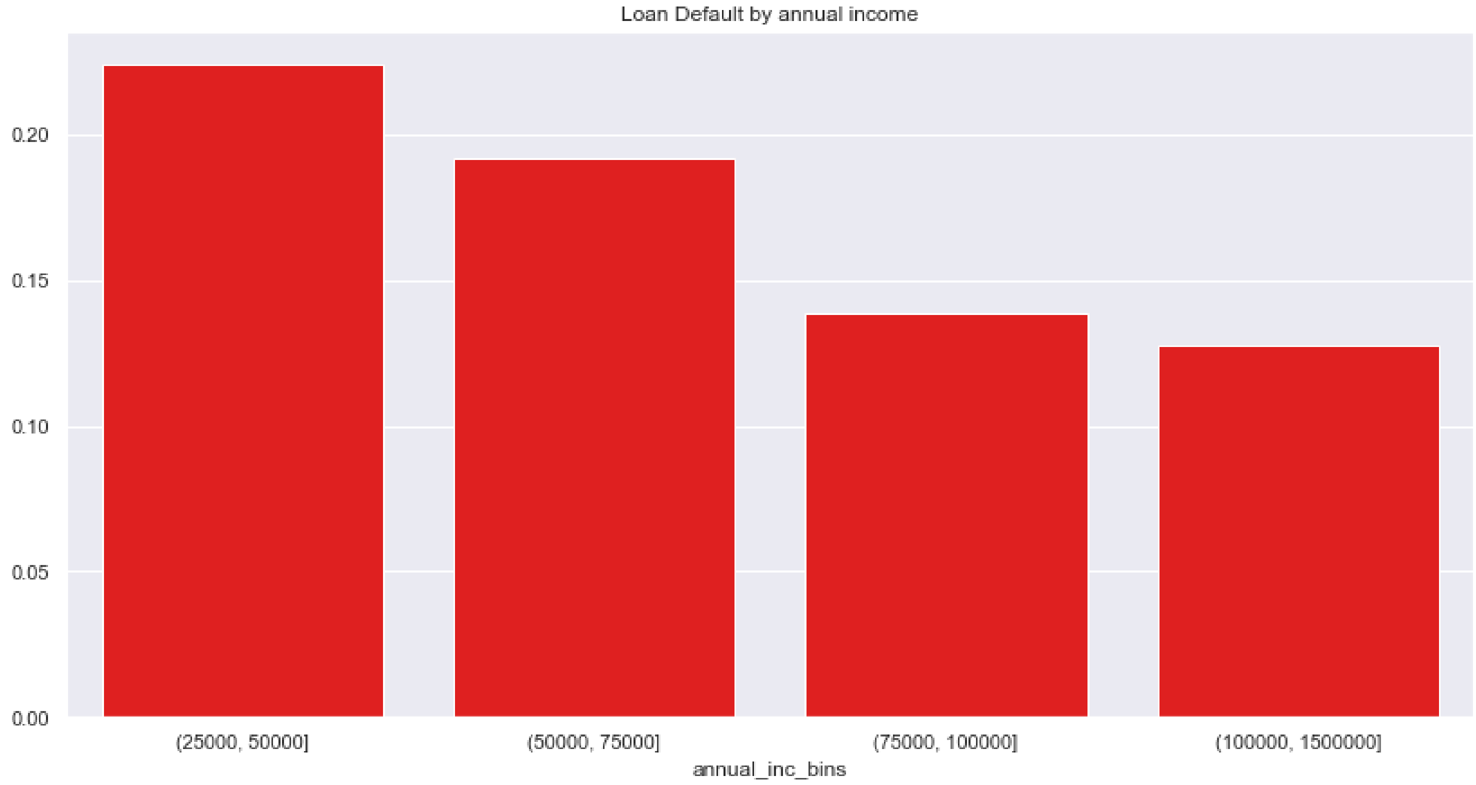
- Highest loans are given without any applications verification
- Verified loans have the highest default percentage

<u>Inferences:</u>

- Verification process need to be relooked into to see what are the reasons for default even after verification

*** Red trend line shows the Charged off count to Fully paid count ratio. This is a good indicator of the likely hood of default**

Annual income impact



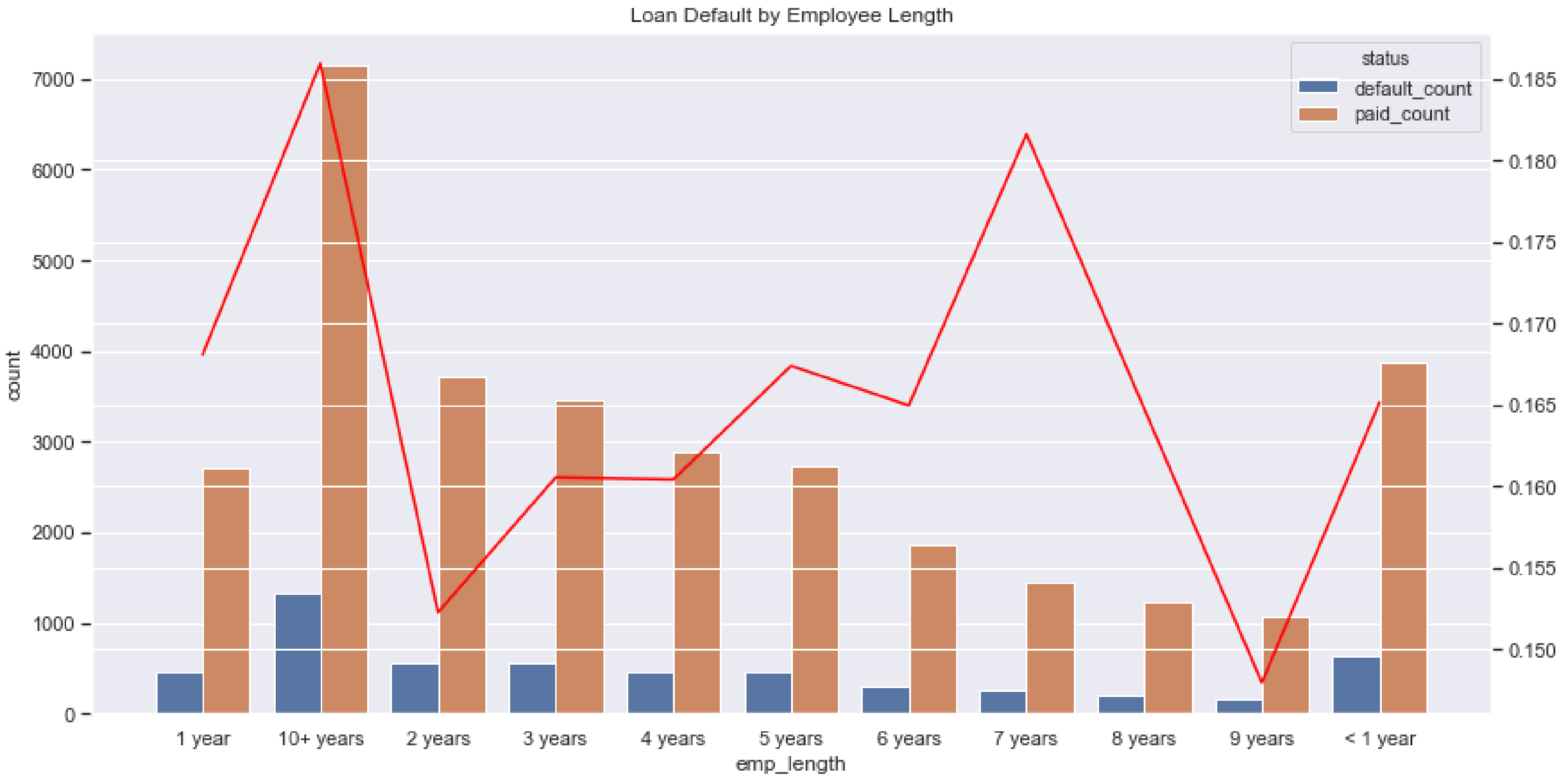
Observations:

- Higher the income rate lower the chance of default

Inferences:

- People on higher income bracket tends to default less

Employment length impact



Observations:

- Higher defaults are with people who have higher employment length 10+ years
- Less than 5 years tends to haave lesser default

Inferences:

* Red trend line shows the Charged off count to Fully paid count ratio. This is a good indicator of the likely hood of default

Conclusions – Driver Variables for Loan Default

Interest Rate	• Higher interest rates are having higher charged off • Since interest rates are derived by grades and subgrade, higher risk grades needs to be monitored for charged off
DTI : Debt to Income Ratio	• Higher the debt to income ratio there is a high risk of people defaulting on the loan
Revol Util: Revolving Utilization %	• Customers having higher revolving utilizations have a high risk of defaulting on the loan
Loan purpose	• People take loan with better rate to consolidate their loans • Small business have high risk of failure, hence loan default is also quite high
Home Ownership	• Others is an undefined category, need more data to understand why it has more defaults • Mortgage has lower percentage of default. Since mortgage is a major loan and any default on other loan will increase the RoI or loan category, hence Mortgage owners tends to pay back and have a less risk of defaulting
Verification Status	• Verification process need to be relooked into to see what are the reasons for default even after verification
Annual Income	• People with higher annual income have lesser default rate