

Lending Club - Case Study

...

February 5, 2023

Team:

Pradeep
Kaushal

Background

Lending Club is the largest online peer-to-peer 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.

Some popular products are credit card loans, debt consolidation loans, house loans, car loans etc.

Objective

Business Objective is to identify variables which are strong indicators of default and potentially use the insights in approval / rejection decision making. So that the company can utilise this knowledge for its portfolio and risk assessment.

Identification of these indicators & such applicants using EDA is the aim of this case study.

Approach

01

Data Cleaning

- Data Importing
- Null Handling & Data Imputation
- Keeping Relevant columns
- Correcting Data Types
- Outlier Removal

02

Univariate Analysis

- Understand the data
- Visualise the data using Histograms, Box plots
- Evaluate central tendency and skewness

03

Bivariate Analysis

- Understand the data after selecting relevant variables
- Visualise the relationship using plots (scatter, line, heatmap)
- Calculate correlation

04

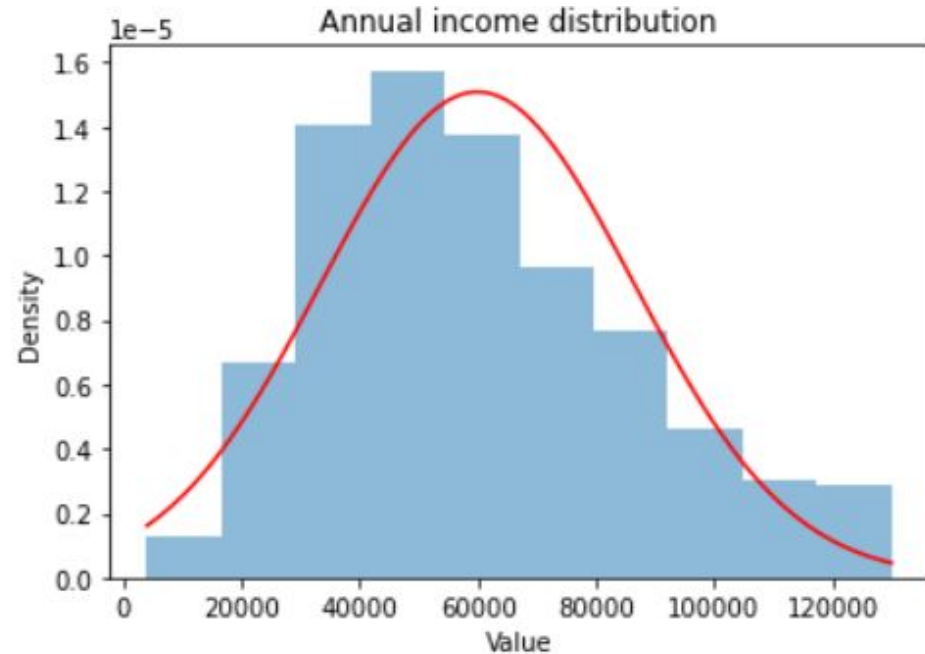
Conclusion

- Draw inferences & Actionable Insights

Univariate Analysis

Annual Income

- Distribution is skewed, peaking at an average annual salary of 55400
- Q1: 40000, Q3: 76000
- IQR: 36000
- Lower bound: -14000.0
- Upper bound: 130000.0

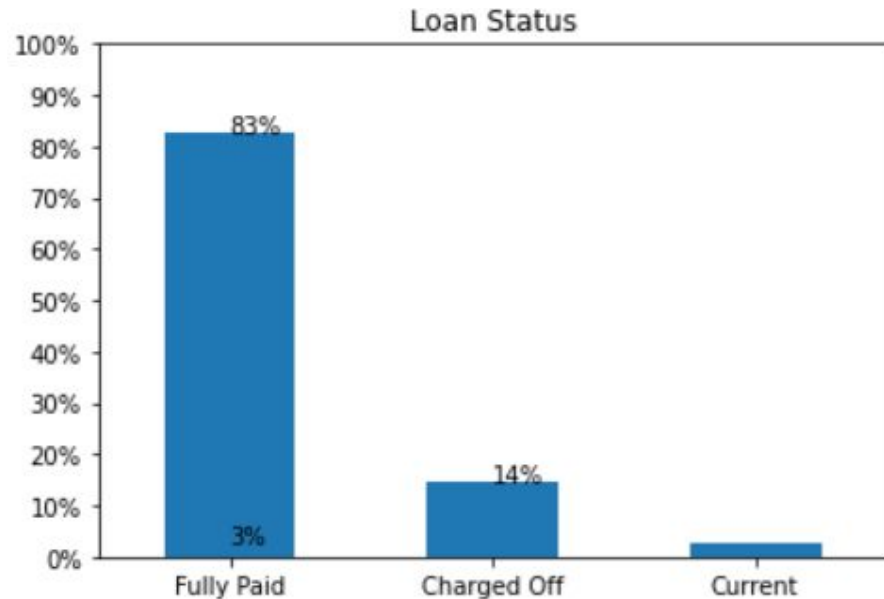


Q1: 40000.0, Q3: 76000.0, IQR: 36000.0,
Median: 55400, Mean: 59946,
Lower bound: -14000.0, Upper bound: 130000.0

Univariate Analysis

Loan Status

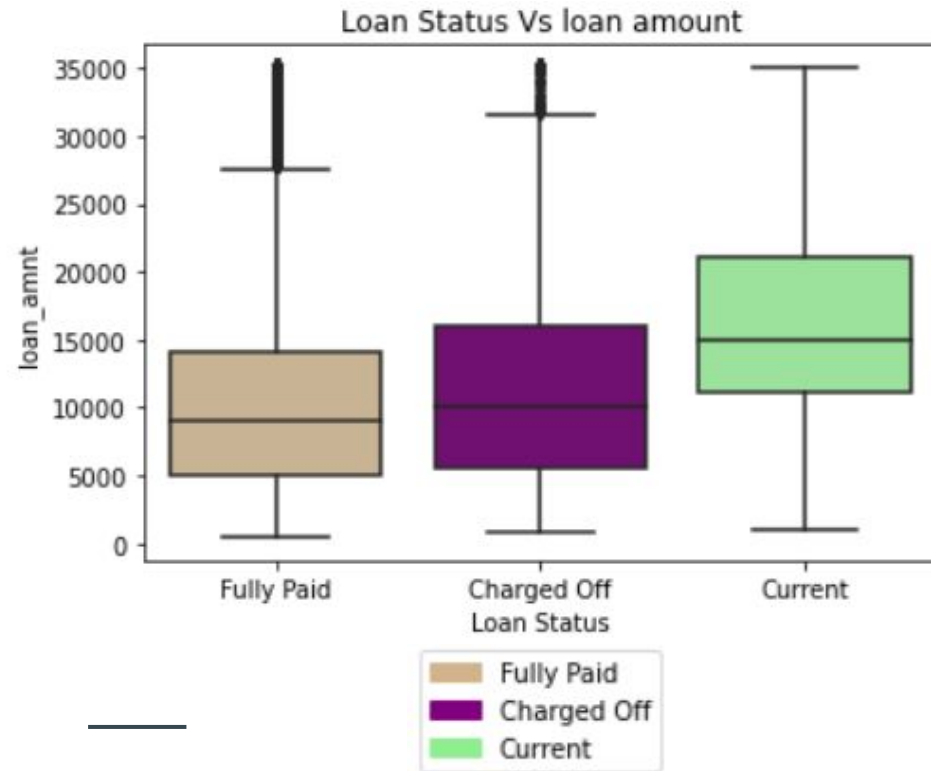
- 83% Loans are fully paid
- 14% Loans have been defaulted by the debtors
- Only 3% loans are currently running
- Total Loans: 37919
- Fully Paid: 30796
- Charged Off: 5359
- Current: 1036



Bivariate Analysis

Loan Status vs Loan Amount

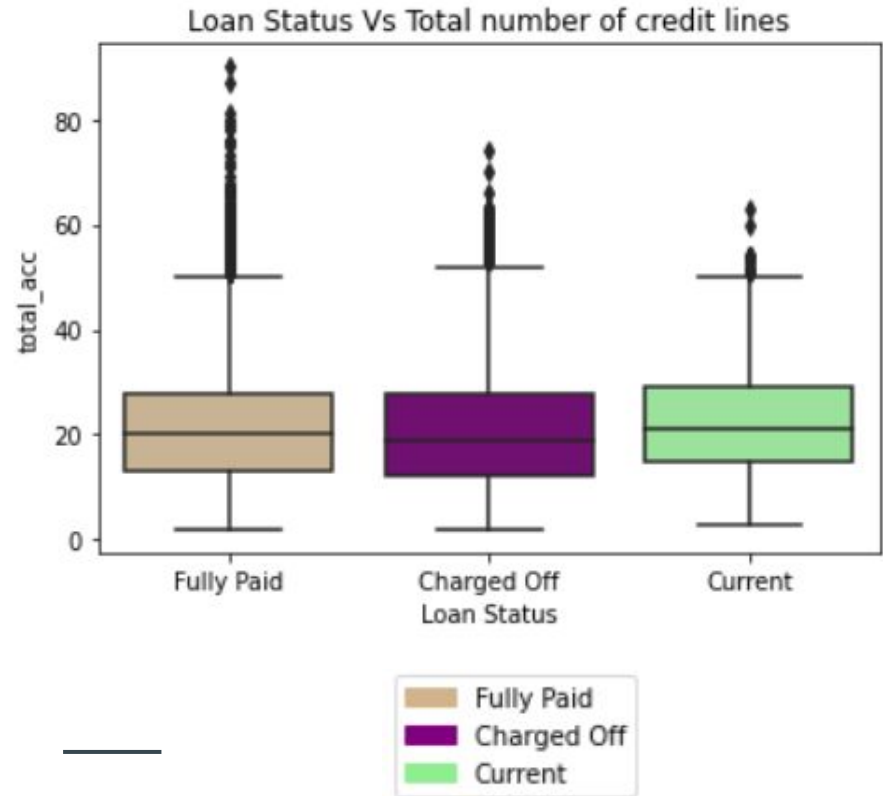
- Fully Paid and Charged off loans have almost similar distributions of Loan amount
- On the other hand, disburseals made for Current Loans seems to be on very higher loan amounts



Bivariate Analysis

Loan Status vs Total Credit Lines

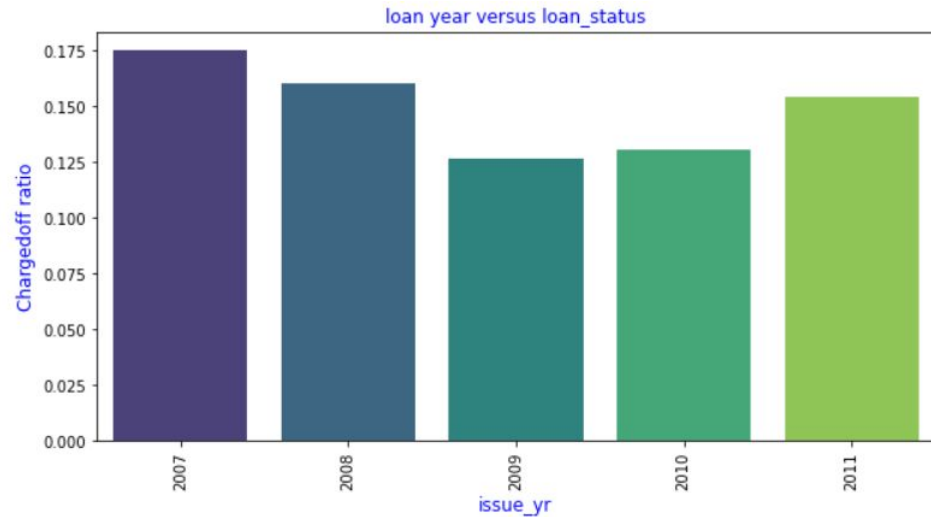
- Similar distribution across all the three categories of loan status
- It can be inferred that number of credit lines of a borrower is not an important factor.



Bivariate Analysis

Loan Status vs Loan Year

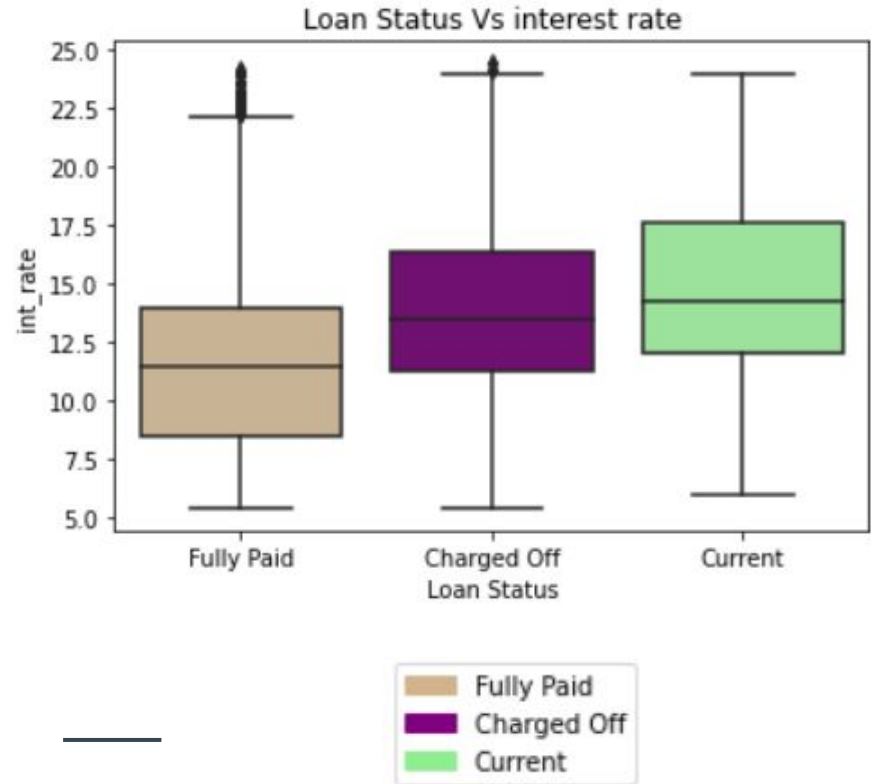
- Year 2007 has most loan defaults at 17.5% with a decreasing trend till 2009 (12.5%)
- It is worth noting and revisit changes in strategy, if any, in year 2010 as a sharp increase in loan defaults is observed in 2011 (15%)



Bivariate Analysis

Loan Status vs Interest Rate

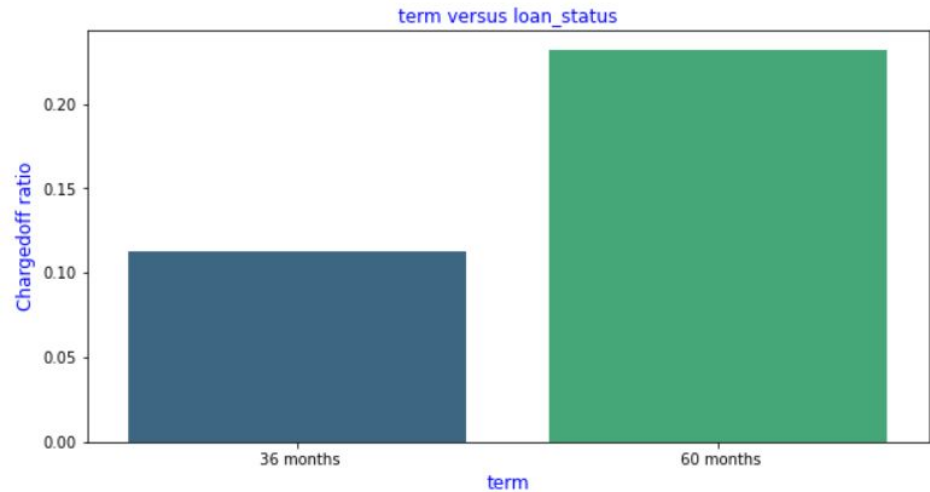
- Interestingly, debtors to whom loans are disbursed at a higher Interest rate have a greater tendency to default a loan
- Building on this fact, a lot of cases seems to be in a riskier cohort given the interest rate of current is high.



Bivariate Analysis

Loan Status vs Term

- Interestingly, loans which are disbursed for a higher term, i.e., 60 months have 23% defaulters as compared to 11% for 36 months

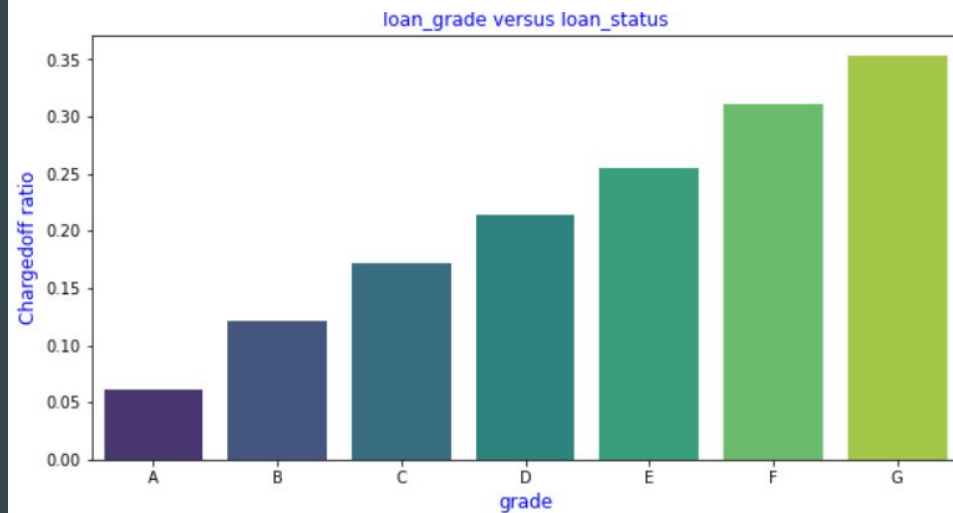


row_id	term	Charged Off	Current	Fully Paid	Total	Chargedoff_percent
1	60 months	2276	1036	6501	9813	0.231937
0	36 months	3083	0	24295	27378	0.112609

Bivariate Analysis

Loan Status vs Loan Grade (subgrade)

- Linear relationship between Loan Grade & Loan defaulter (Charged off ratio) is observed. As shown in graph it increases linearly from 6% for A to 35% for G
- It can be inferred that as the Loan grade degrades the risk of loan default increases
- It is also worth noting that similar trend is observed for Loan subgrade

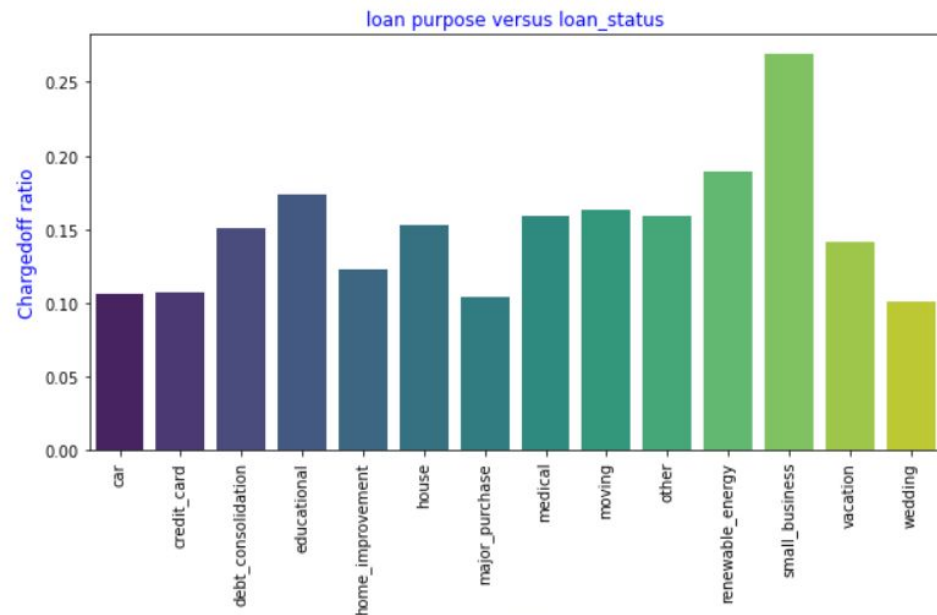


grade	Charged Off	Current	Fully Paid	Total	Chargedoff_percent
G	94	14	158	266	0.353383
F	287	57	578	922	0.311280
E	648	154	1743	2545	0.254617
D	1066	204	3701	4971	0.214444
C	1312	245	6076	7633	0.171885
B	1366	323	9599	11288	0.121013
A	586	39	8941	9566	0.061259

Bivariate Analysis

Loan Status vs Purpose

- Lending loans to Small business has the highest risk of being charged off followed by renewable energy & education
- Loans taken for 'Debt Consolidation' are defaulted the most followed by credit card

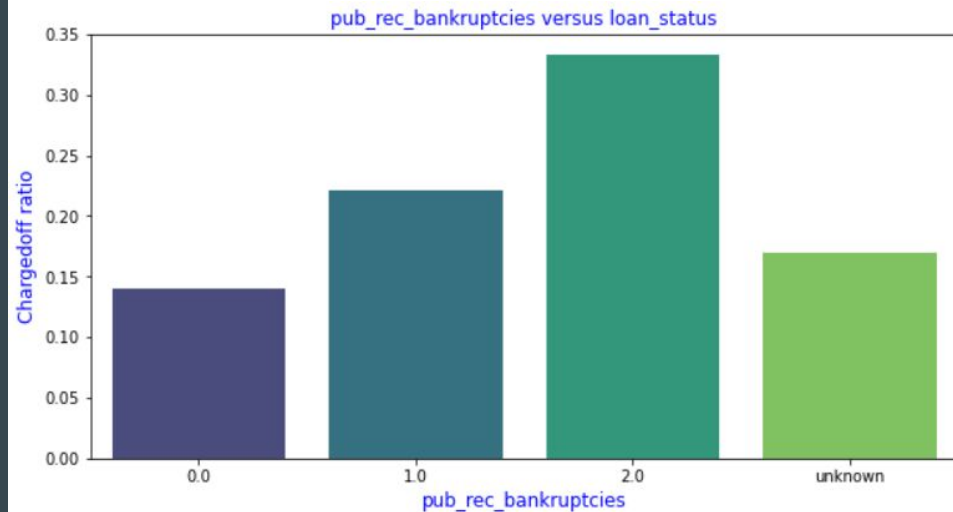


purpose	Charged Off	Current	Fully Paid	Total	Chargedoff_percent
small_business	447	70	1145	1662	0.268953
renewable_energy	18	1	76	95	0.189474
educational	54	0	257	311	0.173633
moving	89	7	448	544	0.163603
medical	102	7	533	642	0.158879
other	600	115	3070	3785	0.158520
house	53	12	281	346	0.153179
debt_consolidation	2652	540	14434	17626	0.150460
vacation	52	5	311	368	0.141304
home_improvement	315	81	2178	2574	0.122378
credit_card	515	94	4193	4802	0.107247
car	158	50	1279	1487	0.106254
major_purchase	214	35	1803	2052	0.104288
wedding	90	19	788	897	0.100334

Bivariate Analysis

Loan Status vs Public Record Bankruptcies

- As the count of public record bankruptcies increases, the contribution of loan defaulters increases (2x)
- Higher the public record bankruptcies higher is the risk of default

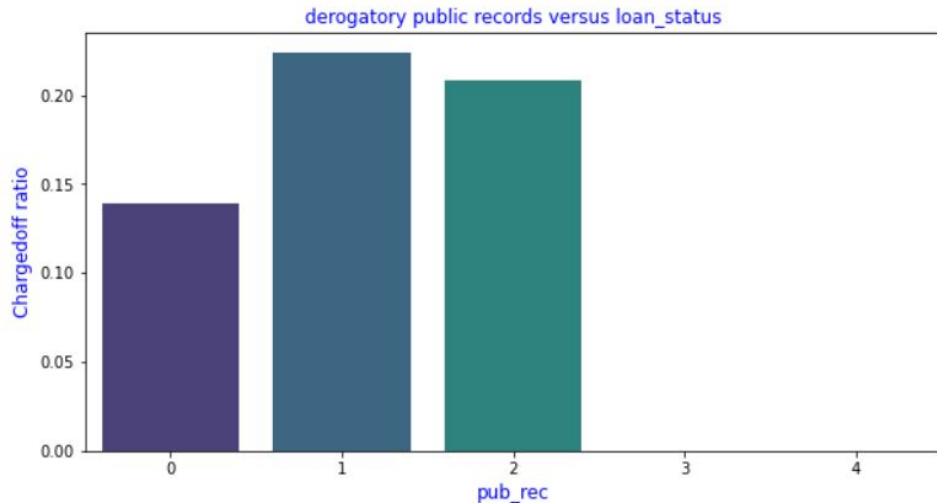


pub_rec_bankruptcies	Charged Off	Current	Fully Paid	Total	Chargedoff_percent
2.0	2	1	3	6	0.333333
1.0	358	35	1229	1622	0.220715
unknown	111	0	542	653	0.169985
0.0	4888	1000	29022	34910	0.140017

Bivariate Analysis

Loan Status vs Derogatory Public Record

- As the count of public record bankruptcies increases, the contribution of loan defaulters increases (2x)
- Higher the public record bankruptcies higher is the risk of default

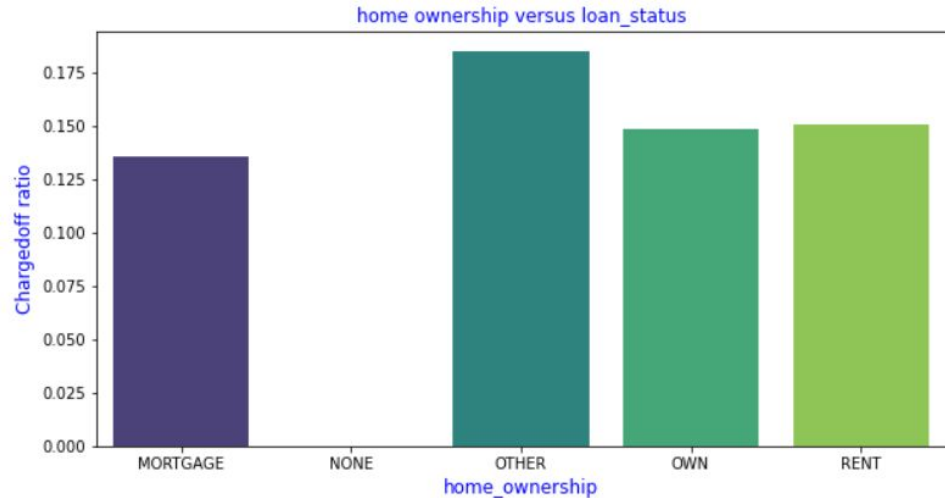


row_id	pub_rec	Charged Off	Current	Fully Paid	Total	Chargedoff_percent
1	1	446	41	1504	1991	0.224008
2	2	10	2	36	48	0.208333
0	0	4903	993	29248	35144	0.139512
3	3	0	0	6	6	0.000000
4	4	0	0	2	2	0.000000

Bivariate Analysis

Loan Status vs Home Ownership

- No major difference is observed in case of Mortgage, Rent or Own
- Ignore None and Other as cases are very low



home_ownership	Charged Off	Current	Fully Paid	Total	Chargedoff_percent
OTHER	17	0	75	92	0.184783
RENT	2760	399	15157	18316	0.150688
OWN	431	77	2396	2904	0.148416
MORTGAGE	2151	560	13165	15876	0.135488
NONE	0	0	3	3	0.000000