Consumer Finance Company Case Study

Team : Amit Goyal and Vamshi Raghu Guntha

Date : 21-08-2024

Problem Statement

Consumer finance company specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.

Two types of risks are associated with the bank's decision:

If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

Business Objectives

how **consumer attributes** and **loan attributes** influence the tendency of default.

Loan Accepted	The company approves the loan.
Loan rejected	The company had rejected the loan

Data Cleaning

There are 39717 rows and 111 columns present in the dataset.

Data Loading

Build data frame by reading CSV

Data Understanding

Understand the data with help of data dictionary

Dropping Columns

Dropped the columns which are having more than 80% of invalid data or when it has less than 2 unique values

Missing Value Imputation

Filling employee title with 'NaN' as 'Unknown'
Filling Homeownership with 'NONE' as 'OTHER'.

Result: Post cleaning there are 39717 rows and 44 columns are present

Data Analysis

Analyzing the driving factor for defaulting the loan applicants:

Identify non-driving factor

Storing unique values like Id, URL

Convert Data Type

Convert all percentages to float type

Date Time Format

Convert all time-series columns to Date-Time format.

9

Categorical columns

19

Numerical columns

2

Date Time columns

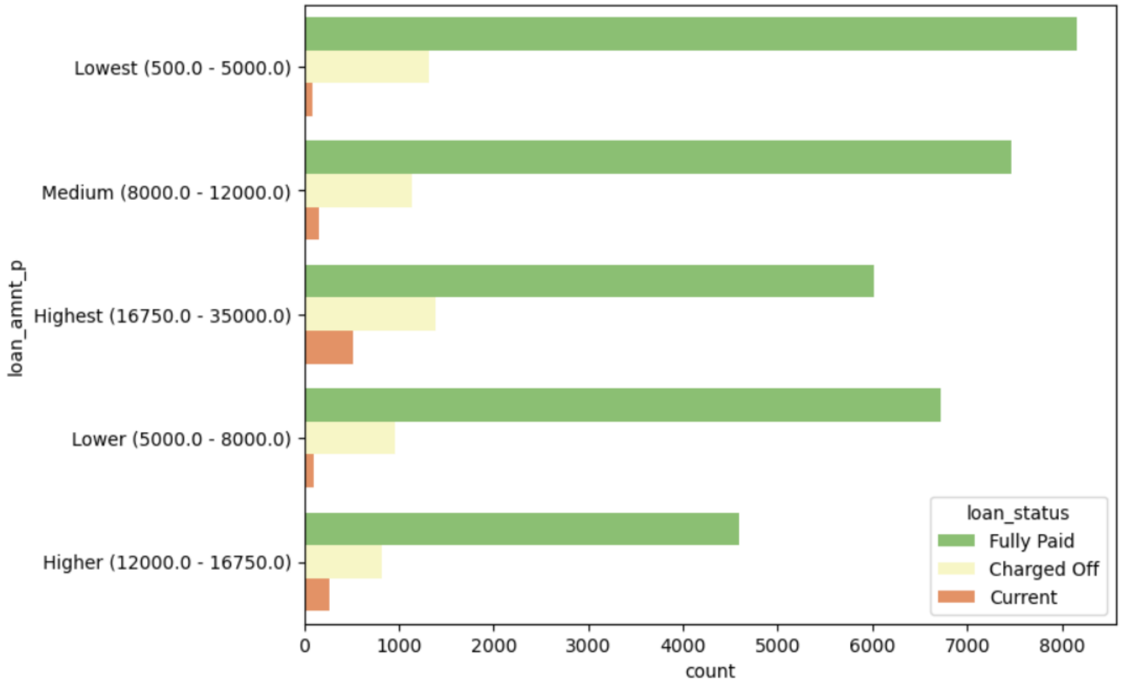
	Column	Dtype	Unique Count
1	loan_amnt	int64	885
2	funded_amnt	int64	1041
3	funded_amnt_inv	float64	8205
4	term	object	2
5	int_rate	float64	371
6	installment	float64	15383
7	grade	object	7
8	sub_grade	int32	5
9	emp_title	object	28821
10	emp_length	object	11
11	home_ownership	object	4
12	annual_inc	float64	5318
13	verification_status	object	3
14	issue_d	datetime64[ns]	55
15	loan_status	object	3
16	purpose	object	14
17	addr_state	object	50
18	dti	float64	2868
19	delinq_2yrs	int64	11
20	earliest_cr_line	datetime64[ns]	526
21	inq_last_6mths	int64	9
22	open_acc	int64	40
23	pub_rec	int64	5
24	revol_bal	int64	21711
25	revol_util	float64	1089
26	total_acc	int64	82
27	pub_rec_bankruptcies	float64	3
28	int_rate_p	float64	371
29	revol_util_p	float64	1089
30	loan_status_code	int64	2

Consumer Finance Company Case Study

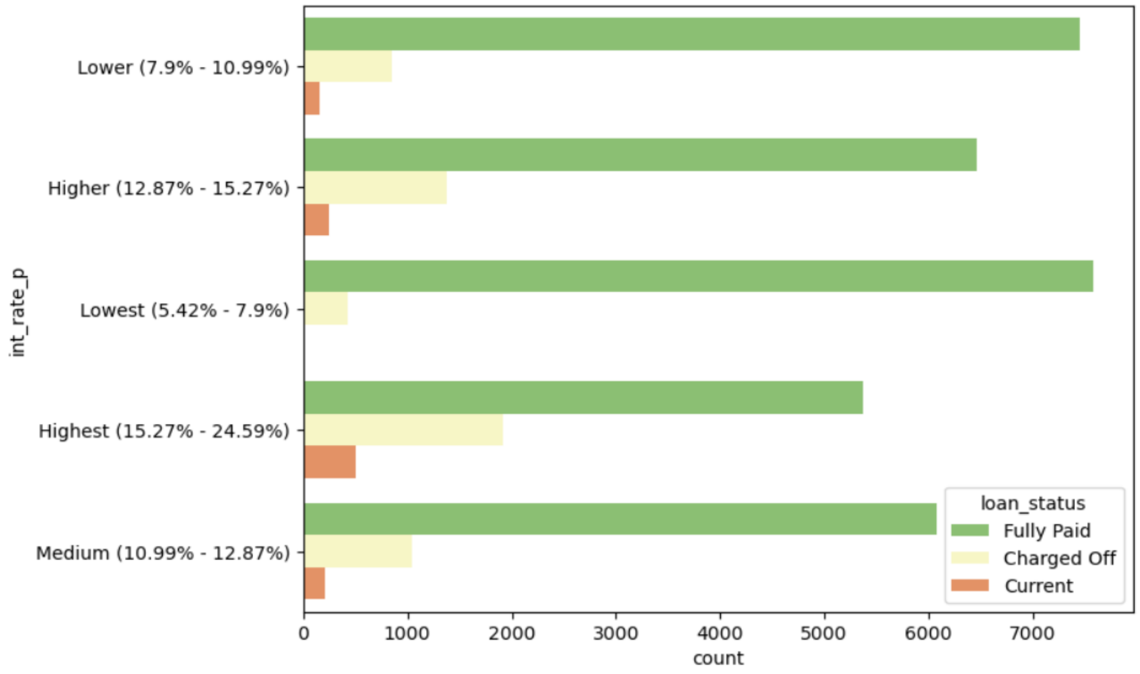
Univariate Analysis

The data has only one variable (Univariate).

Higher the loan amount, greater the chance of the loan getting default.



Higher the interest rate leads to higher charged off%

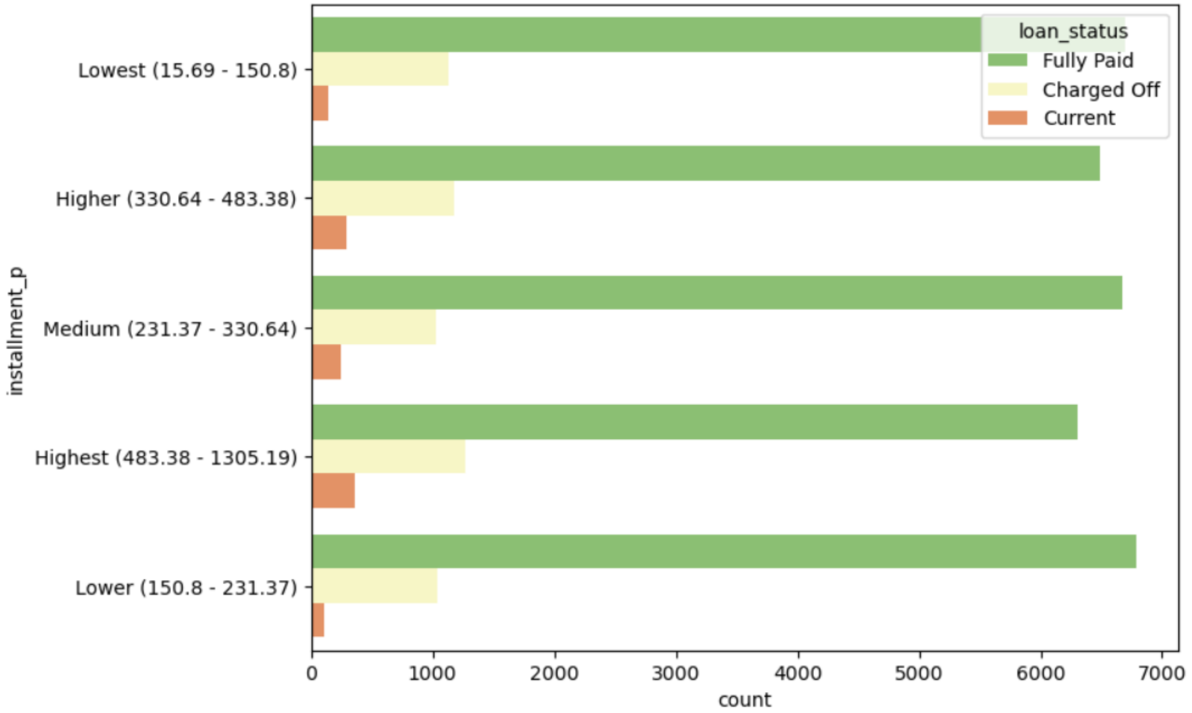


Consumer Finance Company Case Study

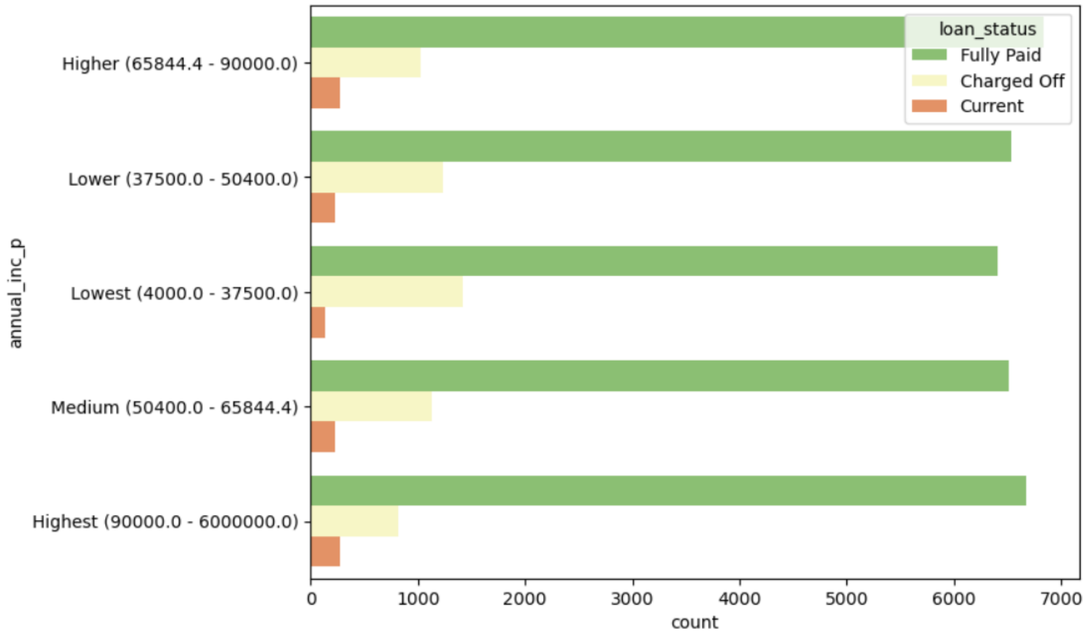
Univariate Analysis

The data has only one variable (Univariate).

Higher installment amounts shows higher default percentages.



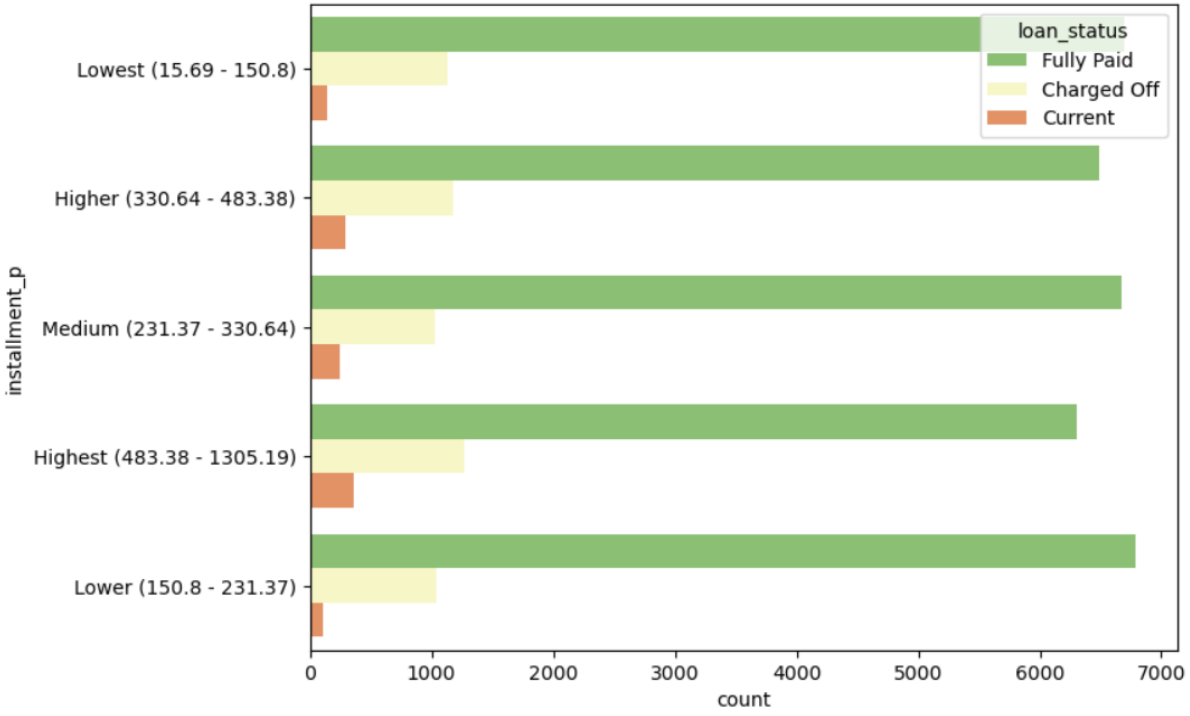
Higher the income higher the repayment %



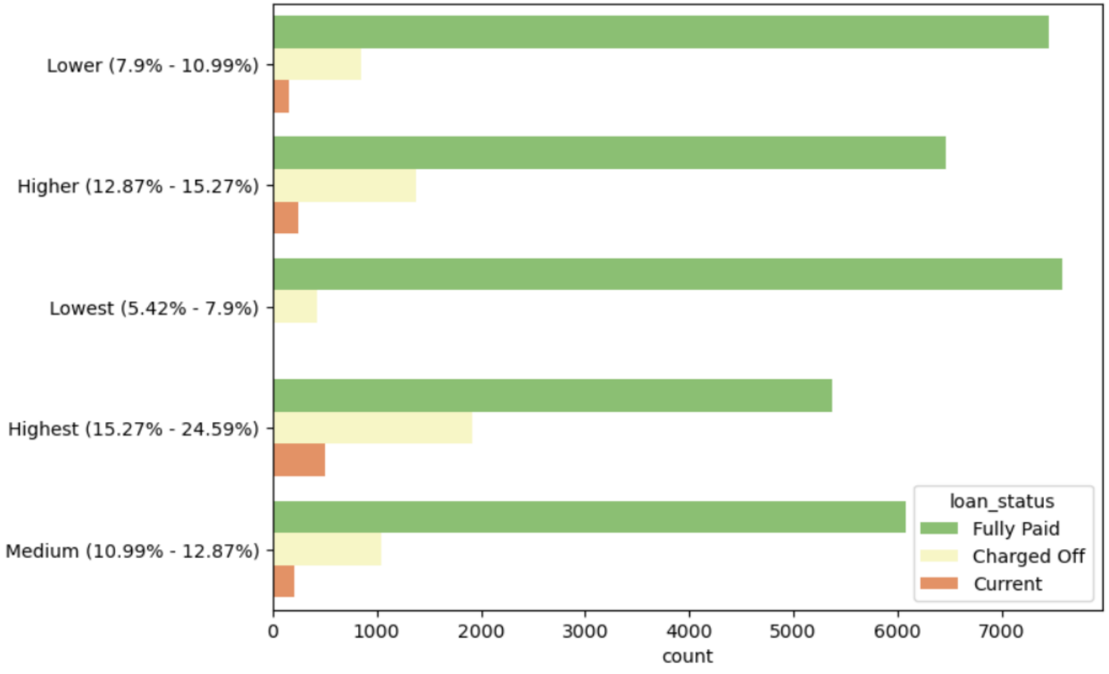
Univariate Analysis

The data has only one variable (Univariate).

Higher installment amounts shows higher default percentages.



Higher the interest rate leads to higher charged off%

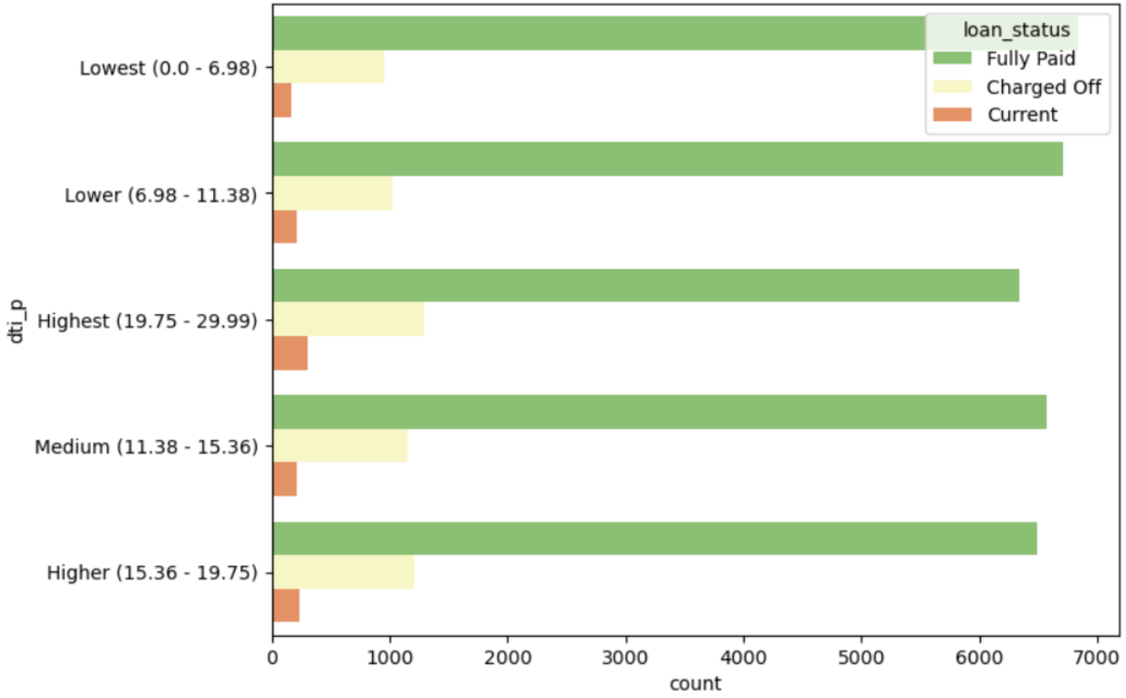


Consumer Finance Company Case Study

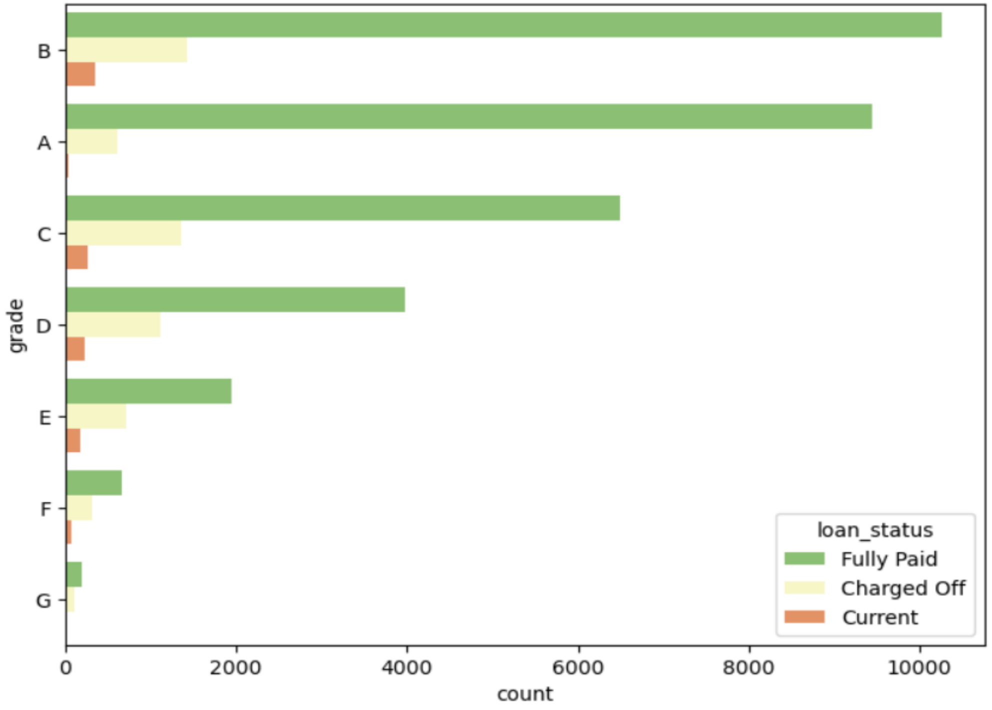
Univariate Analysis

The data has only one variable (Univariate).

Higher DTI (debt to income ratio) will lead to higher charged off %



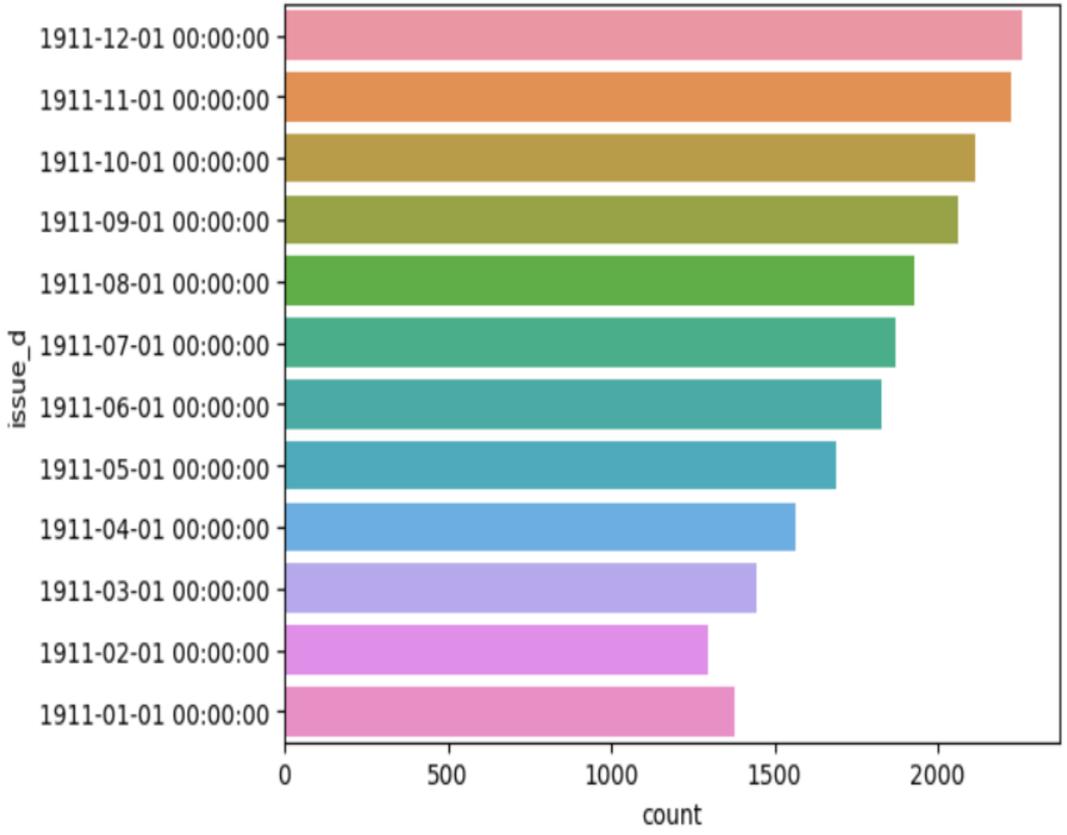
Loan grades having highest default percentages.



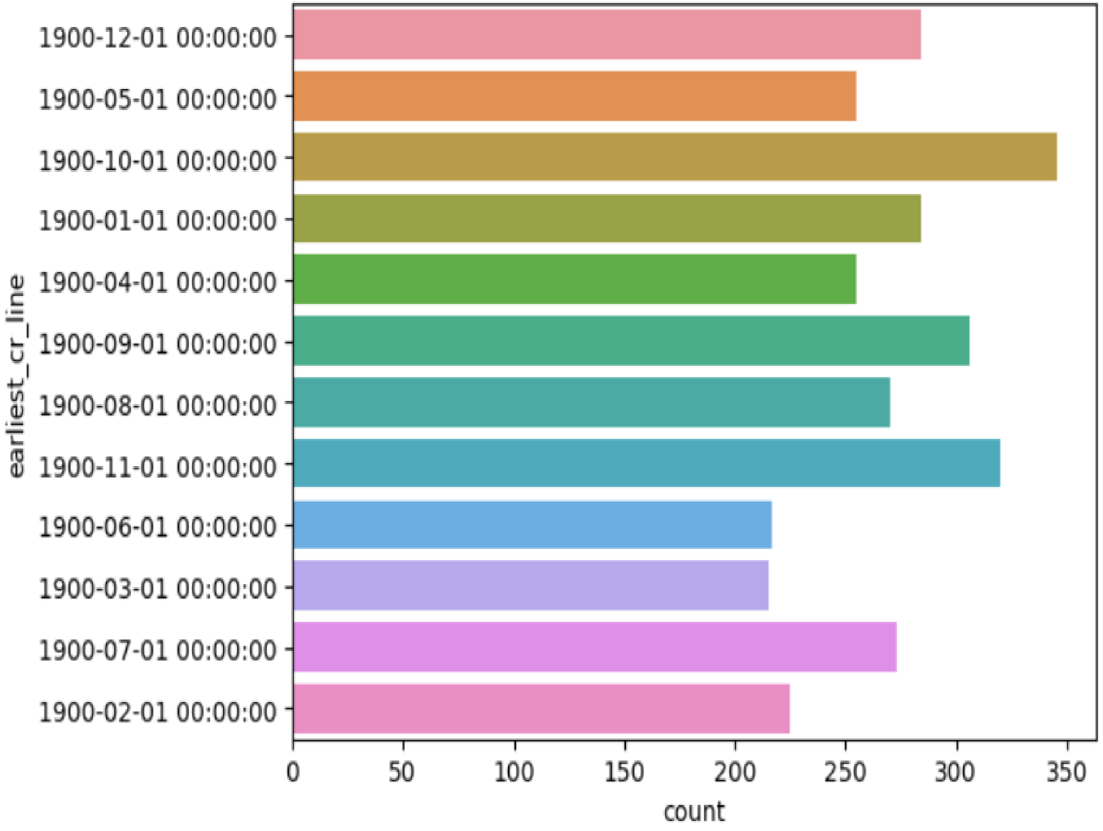
Univariate Analysis

The data has only one variable (Univariate).

Higher chance of defaulted on month 12 of year 1911 on issue_d column



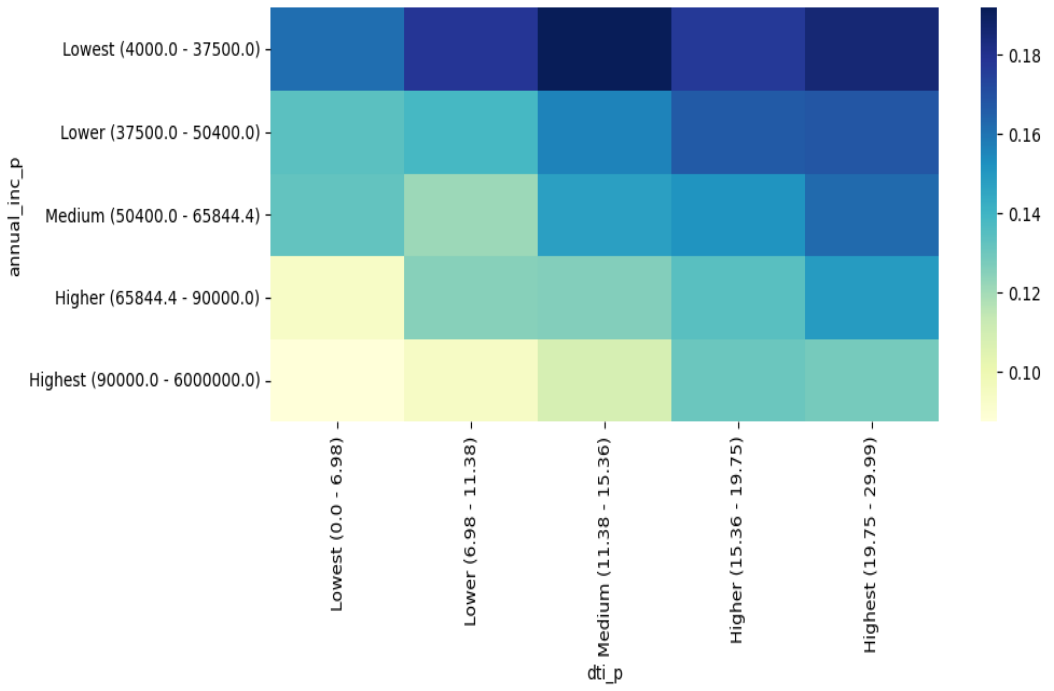
Higher chance of defaulted on month 10 of year 1900 on earliest_cr_line column



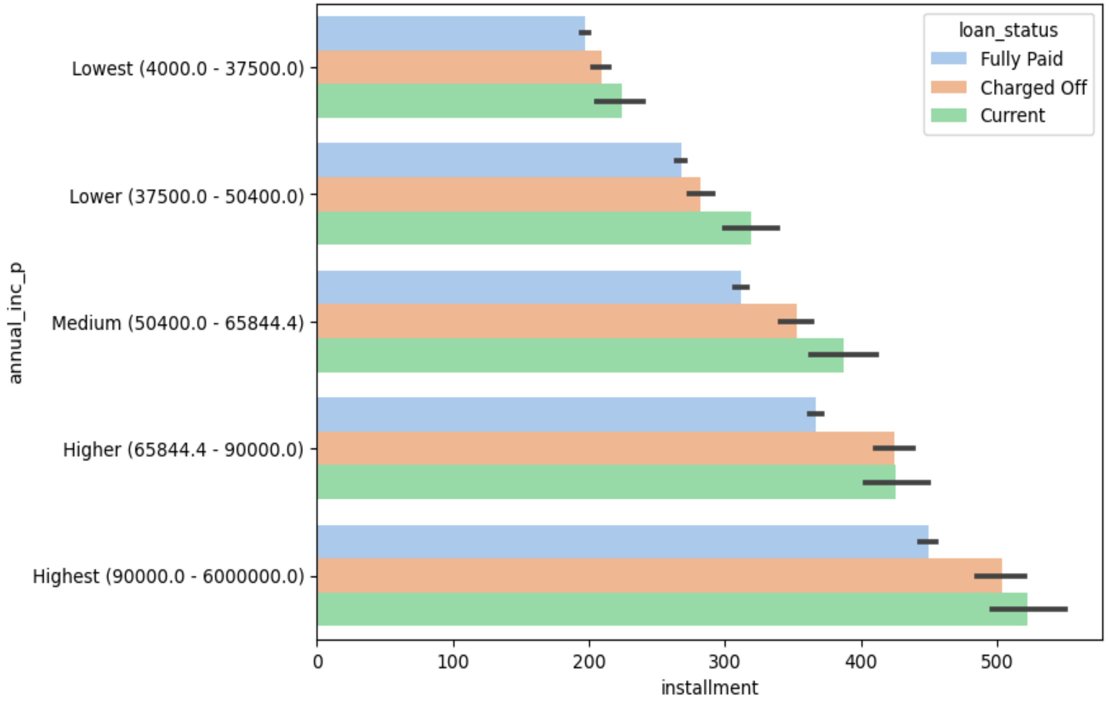
Bivariate Analysis

The data has only two variables

Medium debt-to-income group in the lowest income range is the most risky when it comes to loan repayment.

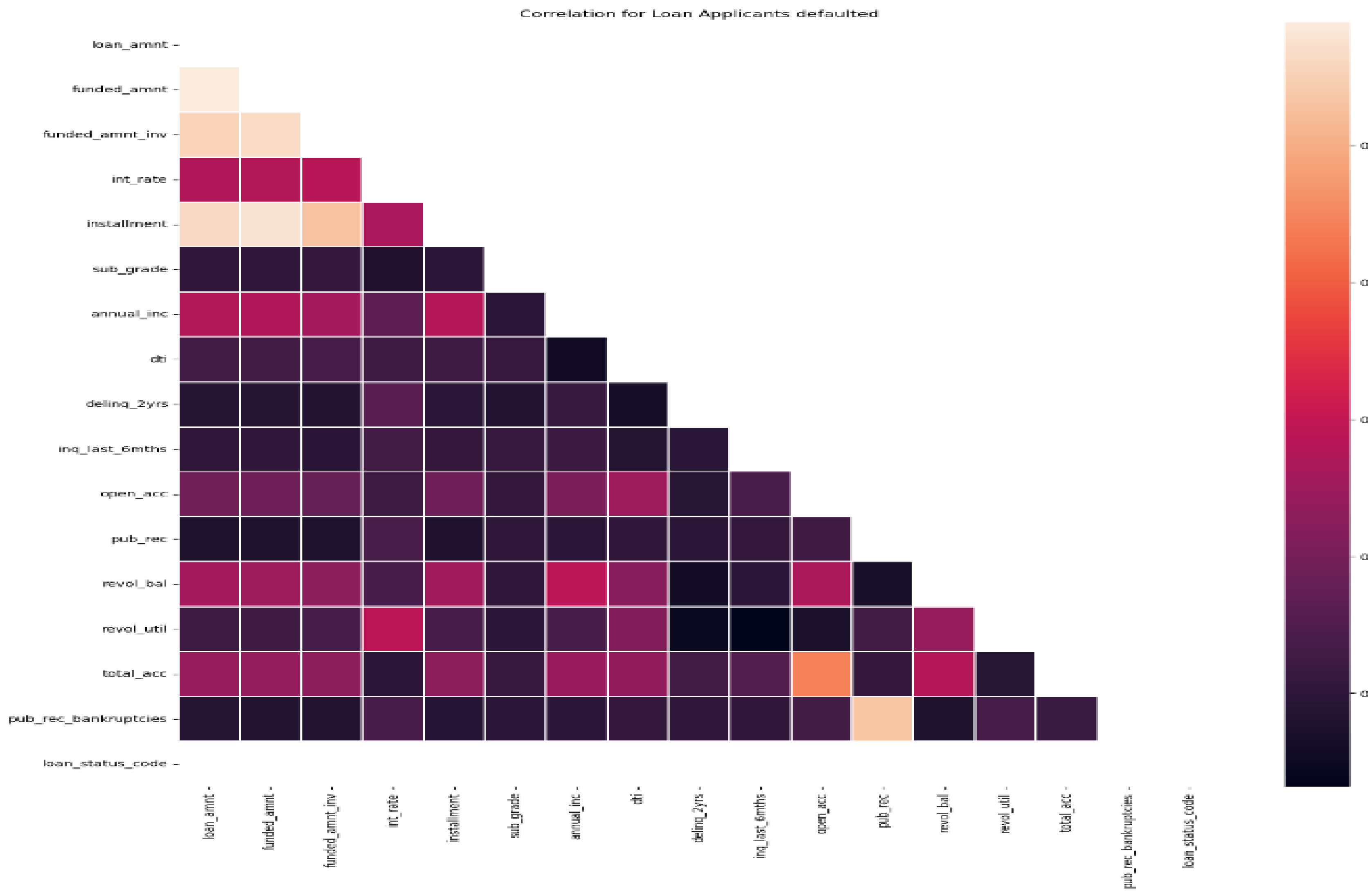


Higher installments for any income group have more number of defaults.



Bivariate Analysis

The data has only two variables



Summary

Apart from the ones highlighted below, I am sure there will be multiple others too; however, according to me, these are the most impactful ones.

Minor Impact

- ❑ Higher loan amount (above 16K)
- ❑ Higher installment amount (above 327)
- ❑ Lower annual income (below 37K)
- ❑ Higher debt to income ratio (above 15%)
- ❑ Applicant's address state (NV, SD, AK, FL, etc.)
- ❑ Loan issue month (Dec, May, Sep)

Heavy impact

- ❑ Higher interest rate (above 13%)
- ❑ Higher revolving line utilization rate (above 58%)
- ❑ Repayment term (5 years)
- ❑ Loan grade & sub-grade (D to G)
- ❑ Missing employment record
- ❑ Loan purpose (small business, renewable energy, educational)
- ❑ Derogatory public records (1 or 2)
- ❑ Public bankruptcy records (1 or 2)

Combined impact

- ❑ High loan amount & interest rate for lower income group
- ❑ High installment and longer repayment term
- ❑ Home ownership (other) and loan purpose (car, moving or small business)
- ❑ Residential state and loan purpose
- ❑ Income group and loan purpose.