Lending Club Case Study

22–May–2024

By

1. Siddharth Shankar

2. Divyansh Modi

Lending Club Case Study

❑ Problem Statement and Objectives

❑ Data Preparation

❑ Data Analysis and Recommendations

Overview

Problem Statement

A consumer finance company, which specialises in lending various types of loans to urban customers, has to make a decision on the loan applications based on the applicant's profile and their probability to repay the loan.

▪ 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

Objectives

The company aims to identify patterns which indicate if a person is likely to default i.e. strong indicators / driving factors for default. The company can utilise this knowledge for its portfolio risk assessment, and for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

The objective of this case study is to use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

Data Preparation

Data Understanding

1. In the dataset, there are 39,717 entries of loan applications with 111 columns

2. The dataset contains information about past loan applicants and whether they 'defaulted' or not. It has three loan status values – Fully Paid, Charged off (defaulted) and Current

3. There are a number of columns with data quality issues such as
   ◦ missing values or single unique value
   ◦ integer or percentage values stored as strings
   ◦ outliers in some of the key columns used for analysis

Data Cleaning and Manipulation

The following steps were used to clean the data and prepare this into a format fit for data analysis.

1. Remove unnecessary columns (all blanks or single unique value or columns that are only available post loan approval which cannot be indicators if default by definition)

2. Remove unnecessary rows (rows with all missing values, current status loan applicants)

3. Deal with missing values

4. Standardising values

5. Deal with outliers

Data Analysis & Recommendations

Data Analysis Approach Overview

Univariate Analysis
1) Detailed examination of each key variable.
2) Individual slides dedicated to important variables.
*3) A summary slide highlighting key findings.*

Bi-variate or Multivariate Analysis
1) Analysis of interactions and relationships between variables i.e. understand which of these features influence the loan status and by how much
2) Adhering to the format established in the univariate analysis.
3) Emphasis on identifying patterns and correlations.
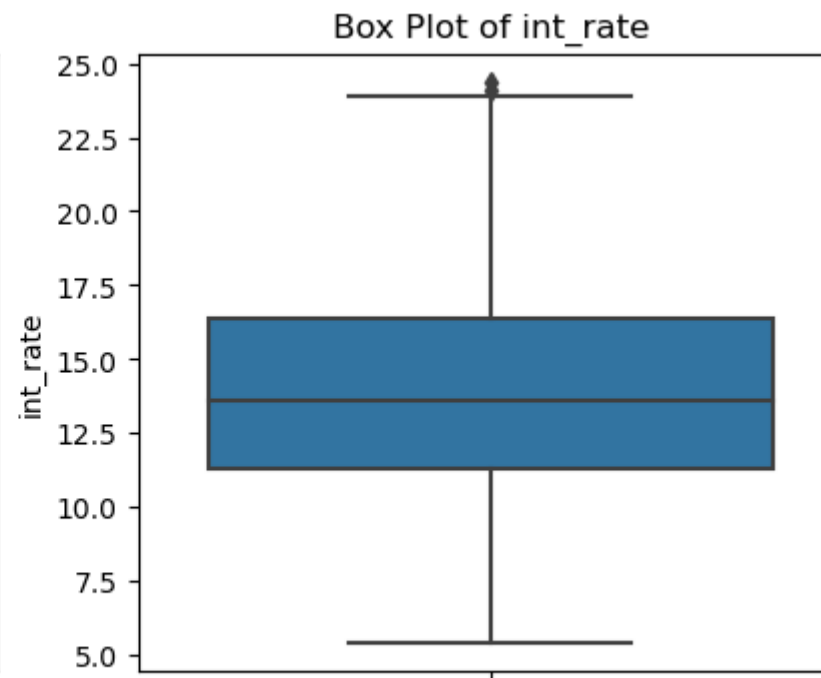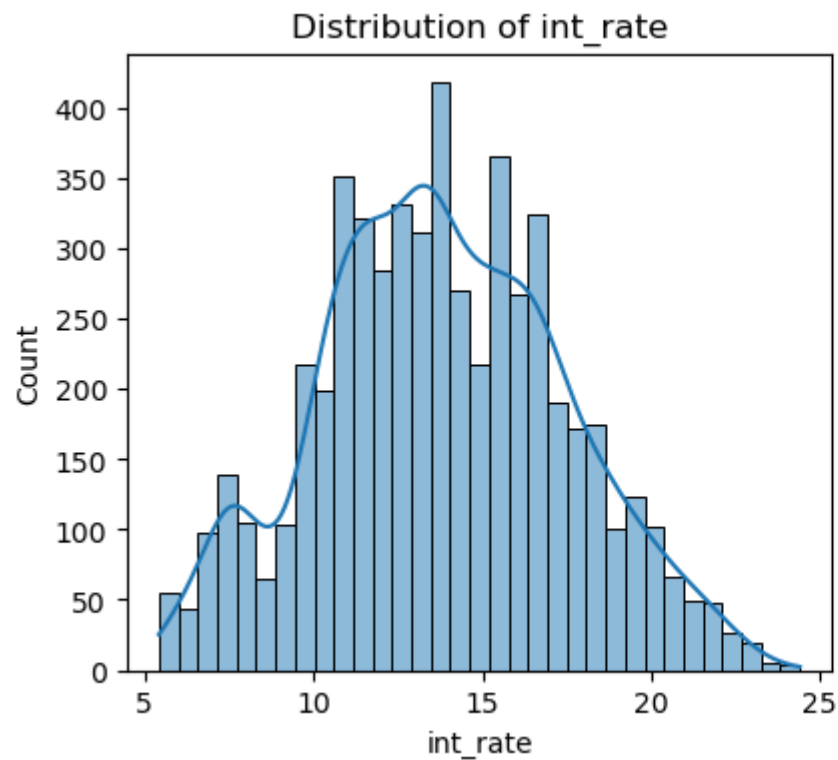*4) A summary slide highlighting key findings.*

Recommendations
Concluding with actionable insights and strategic recommendations derived from the analysis.
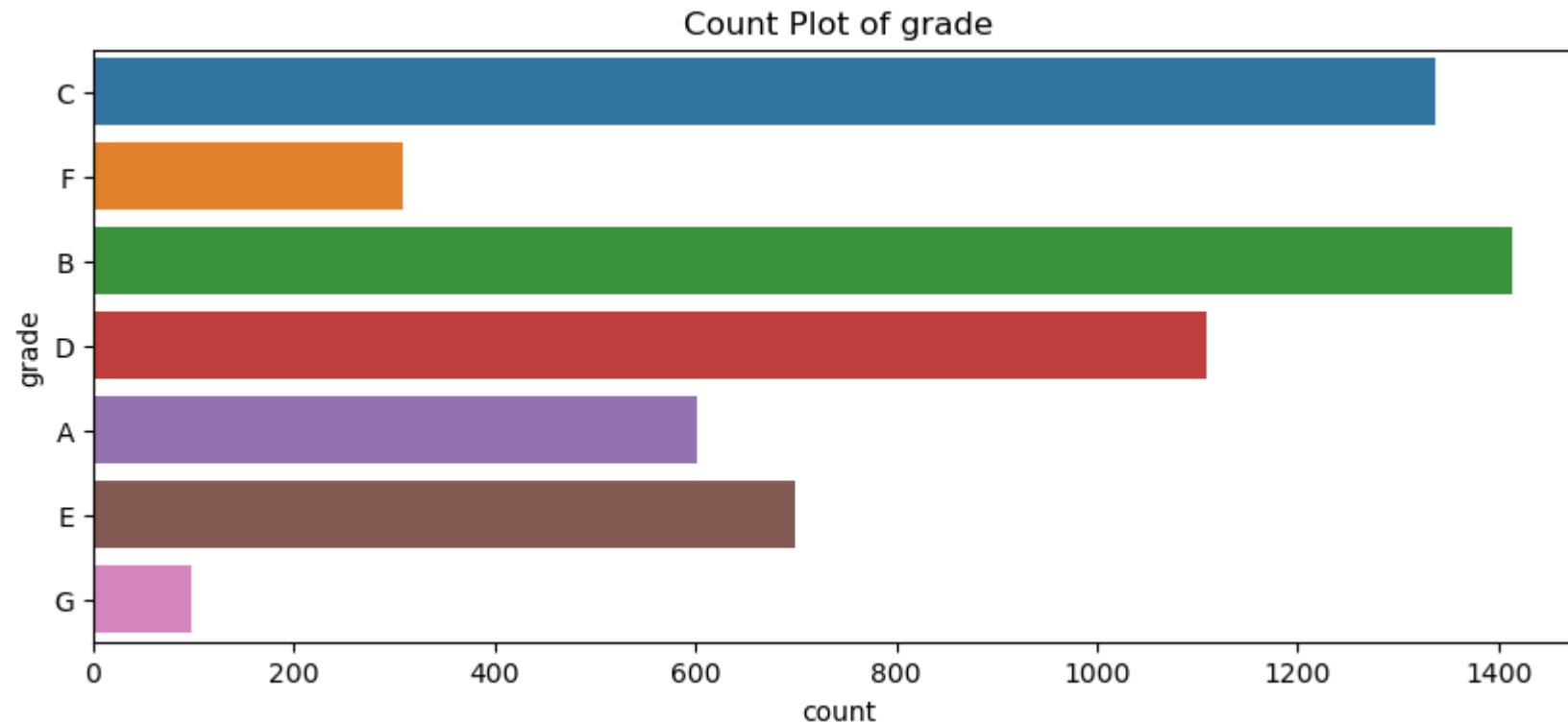
Loan Amt – Distribution within Charged Off applications
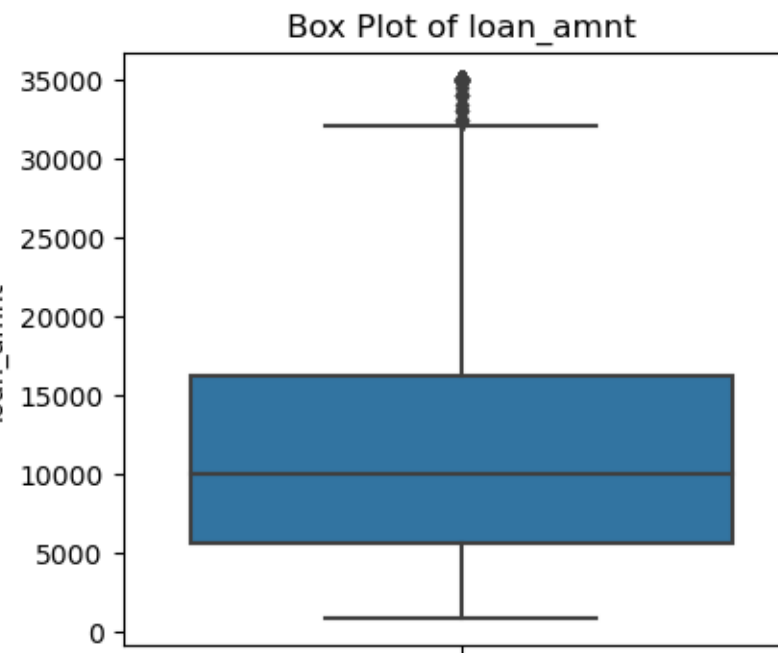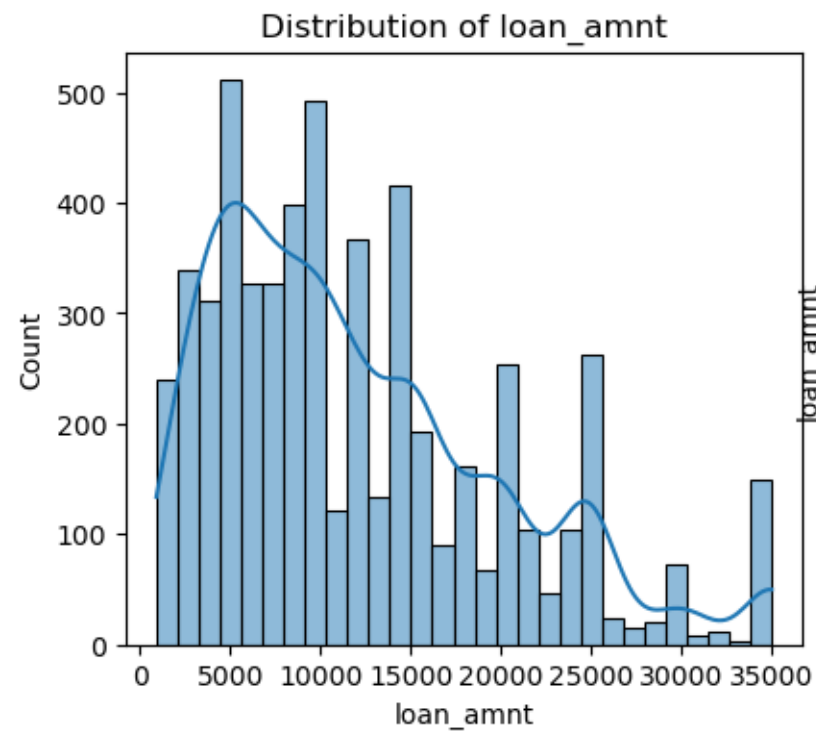
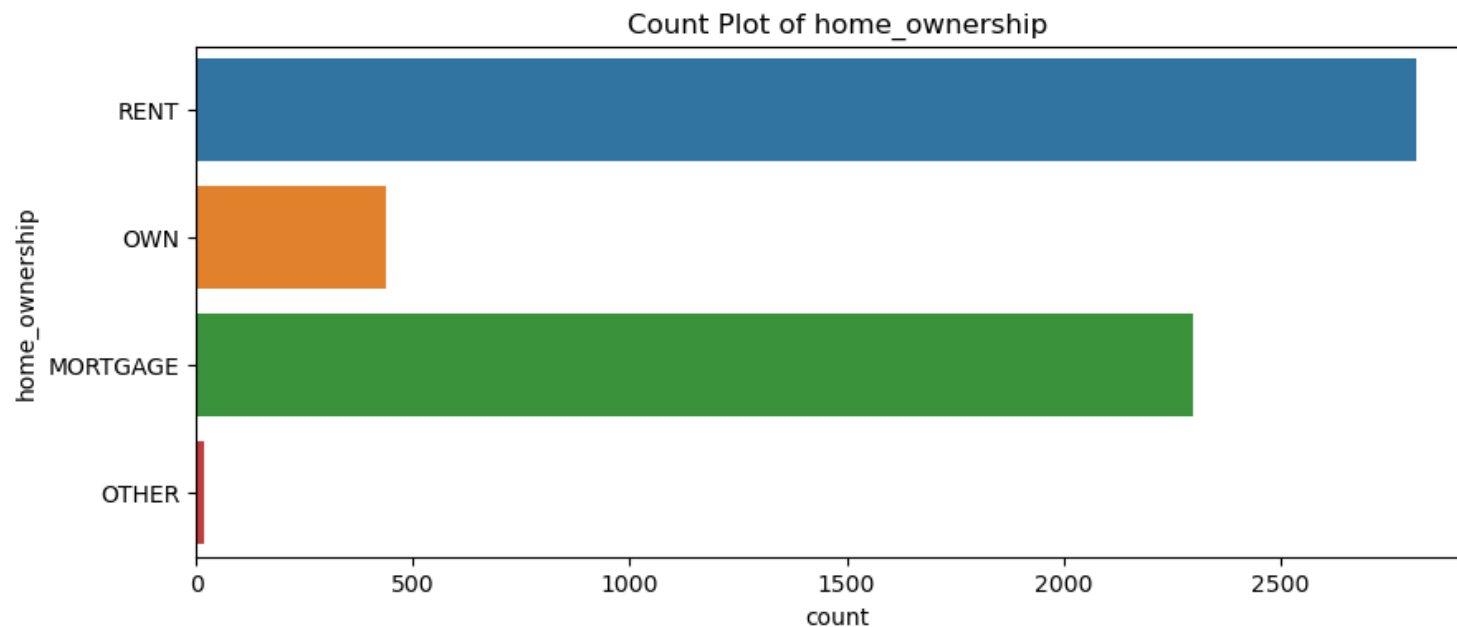# Interest Rate within Charged Off applications

Grade – Occurrence within Charged Off data

Count Plot of grade

Grade B has the highest number of charged off loan applicants

Home Ownership – Occurrence within Charged Off data

Count Plot of home_ownership

Home ownership of "Rent" has the highest number of charged off loan applicants

Employee Tenure – Occurrence within Charged Off data

Count Plot of emp_length
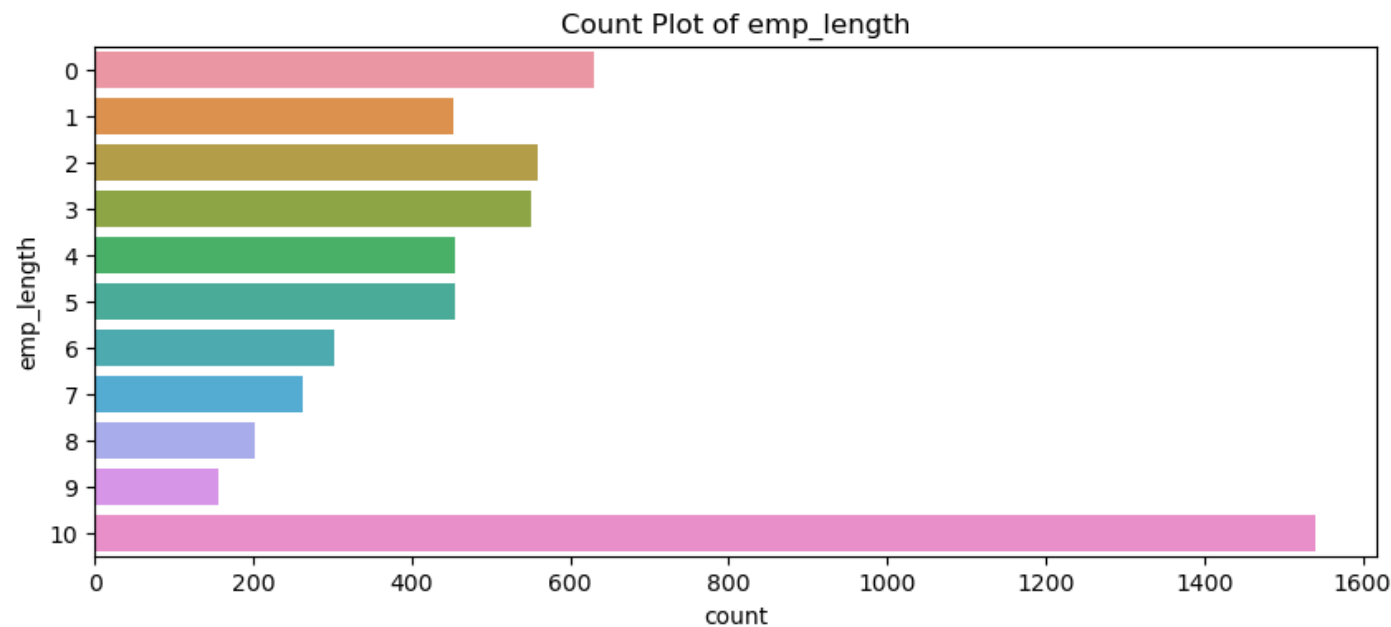
Employment length of 10 or more years has the highest number of charged off loan applicants

Purpose – Occurrence within Charged Off data

Count Plot of purpose
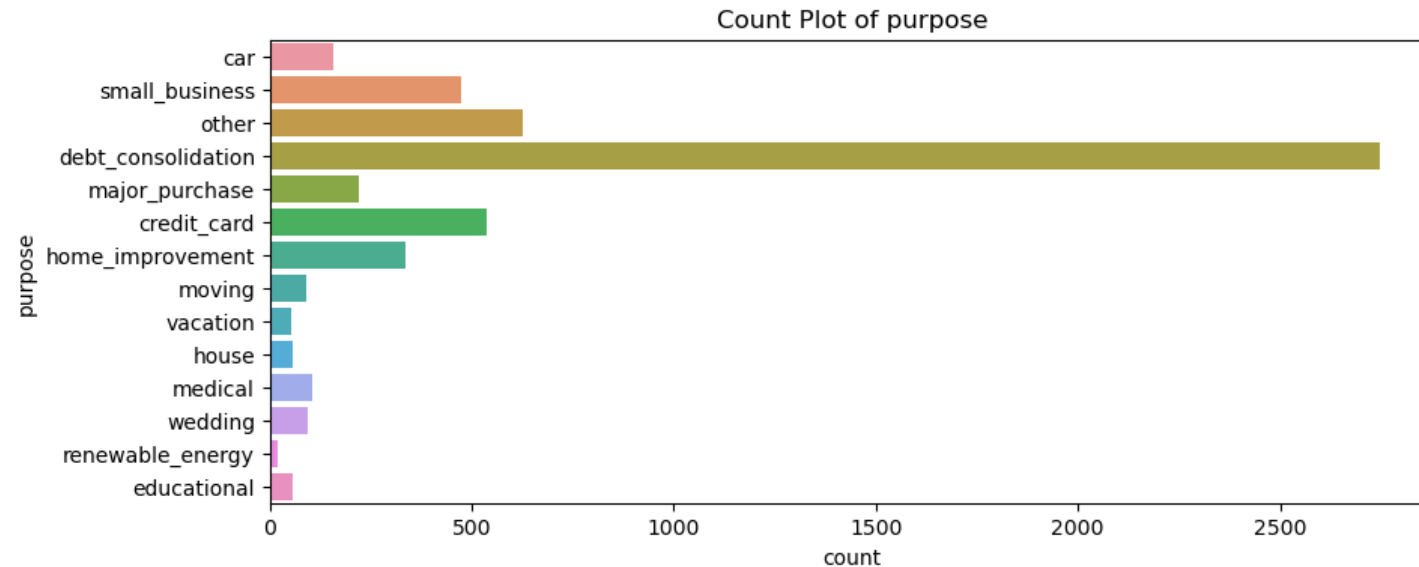
Debt consolidation, credit card, other and small business have the highest number of charged off loan applicants

Term – Occurrence within Charged Off data

Count Plot of term
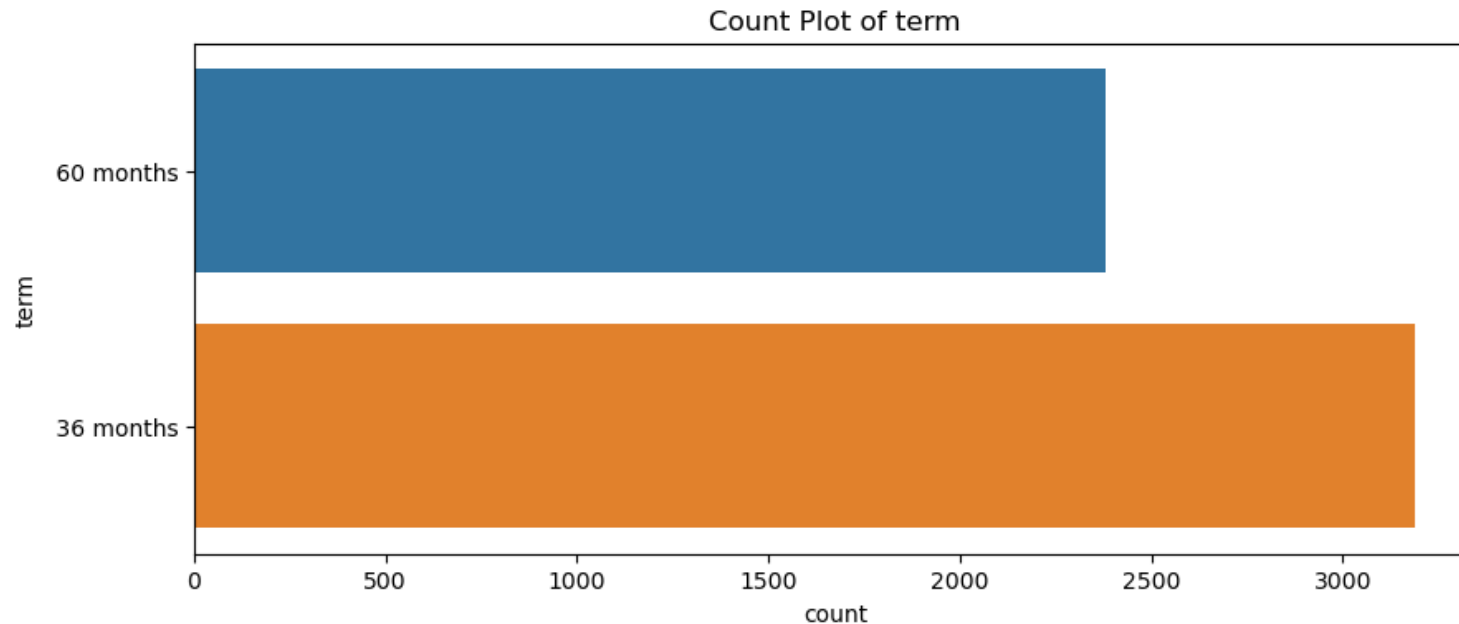
Term of 36 months has the highest number of charged off loan applicants

Issue Month Year – Occurrence within Charged Off data
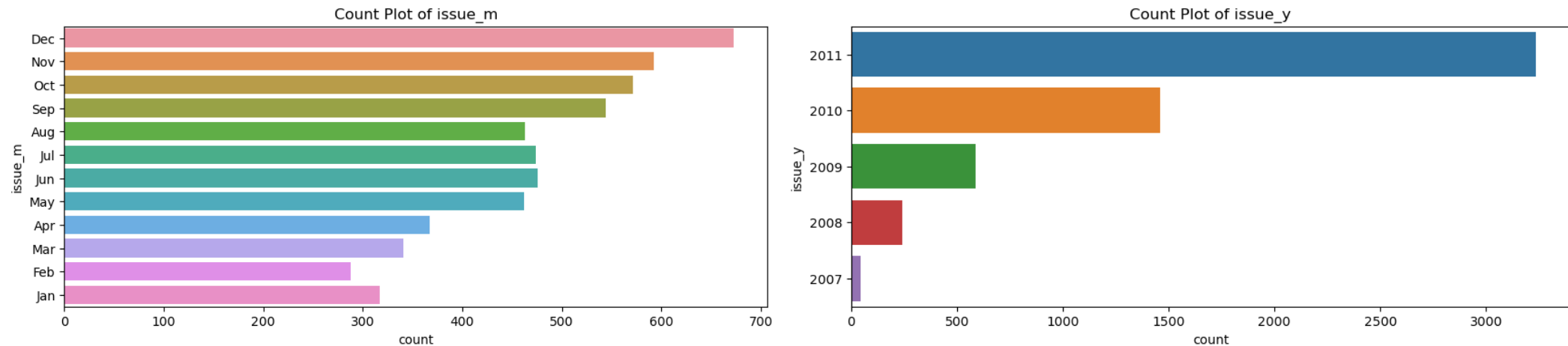
Loan applications issued in the second half of the year and in 2011 have the highest number of charged off loan applicants

Annual Income Range – Occurrence within Charged Off data

Count Plot of annual_inc_bins
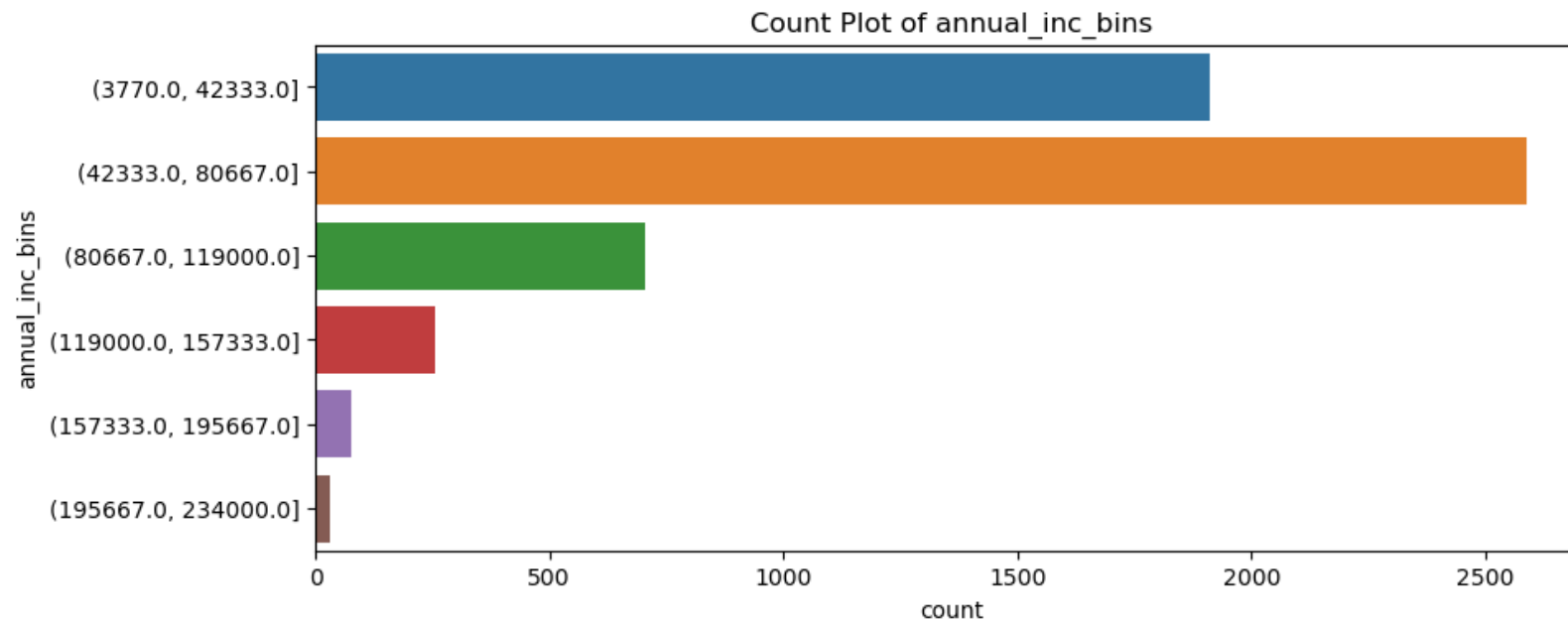
Applicants with an income range of 42K - 81K has the highest number of charged off loan applicants

Key Takeaways

These provide a strong indication of which features are prominent within the "charged off" set of loan applications

• Grade B has the highest number of charged off loan applicants

• Loan applicants with home ownership of "Rent"

• Loan applicants with employment length of 10 or more years

• Loan applicants with a purpose of Debt consolidation, credit card, other and small business

• Term of 36 months

• Loan applications issued in the second half of the year and in 2011

• Applicants with an income range of 42K - 81K

• Applicants with subgrade of B5

Loan Amts vs Income categories
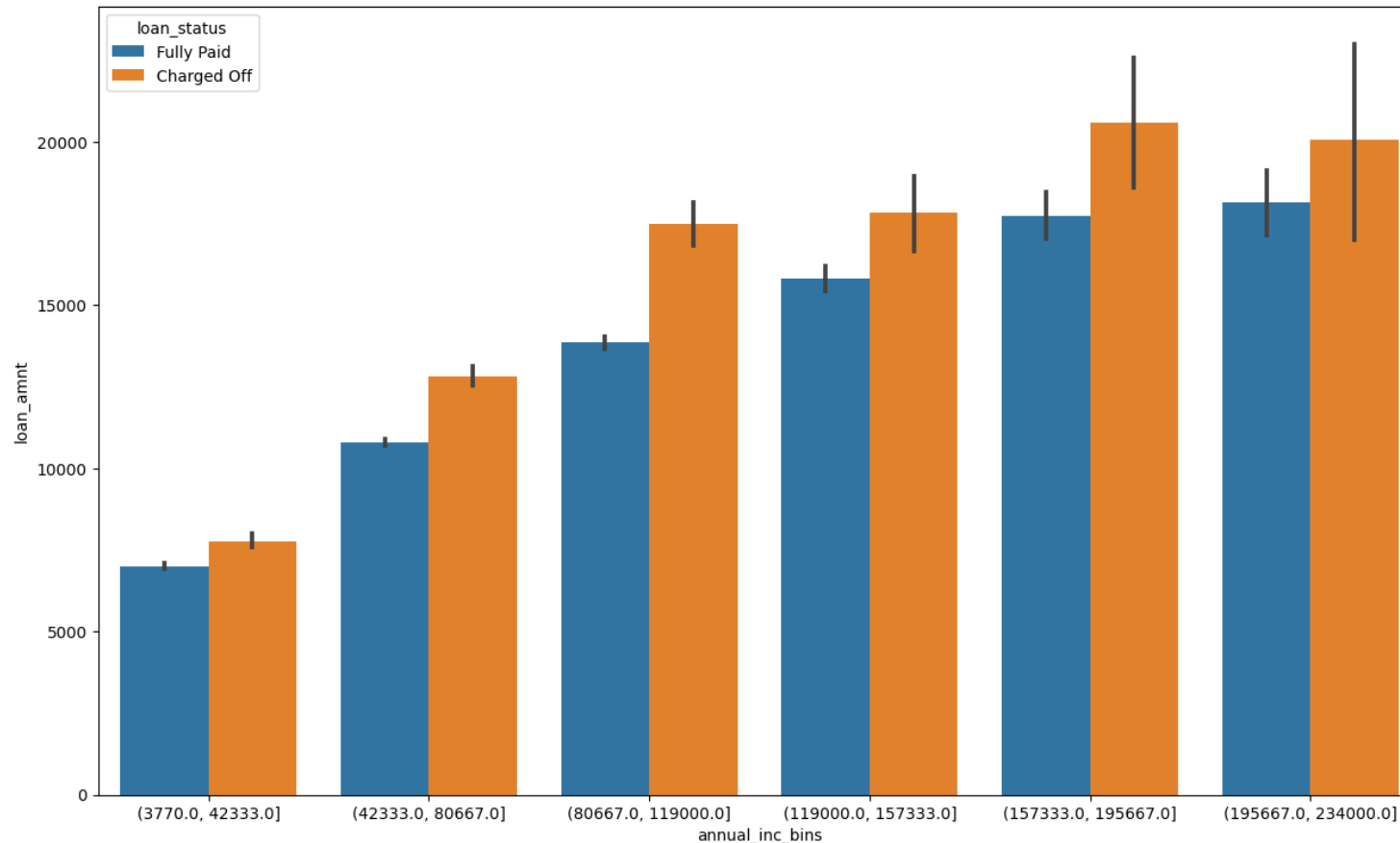
Charged off loan amounts are higher than fully paid loan amounts in all income categories

Loan Amts vs Employee lengths
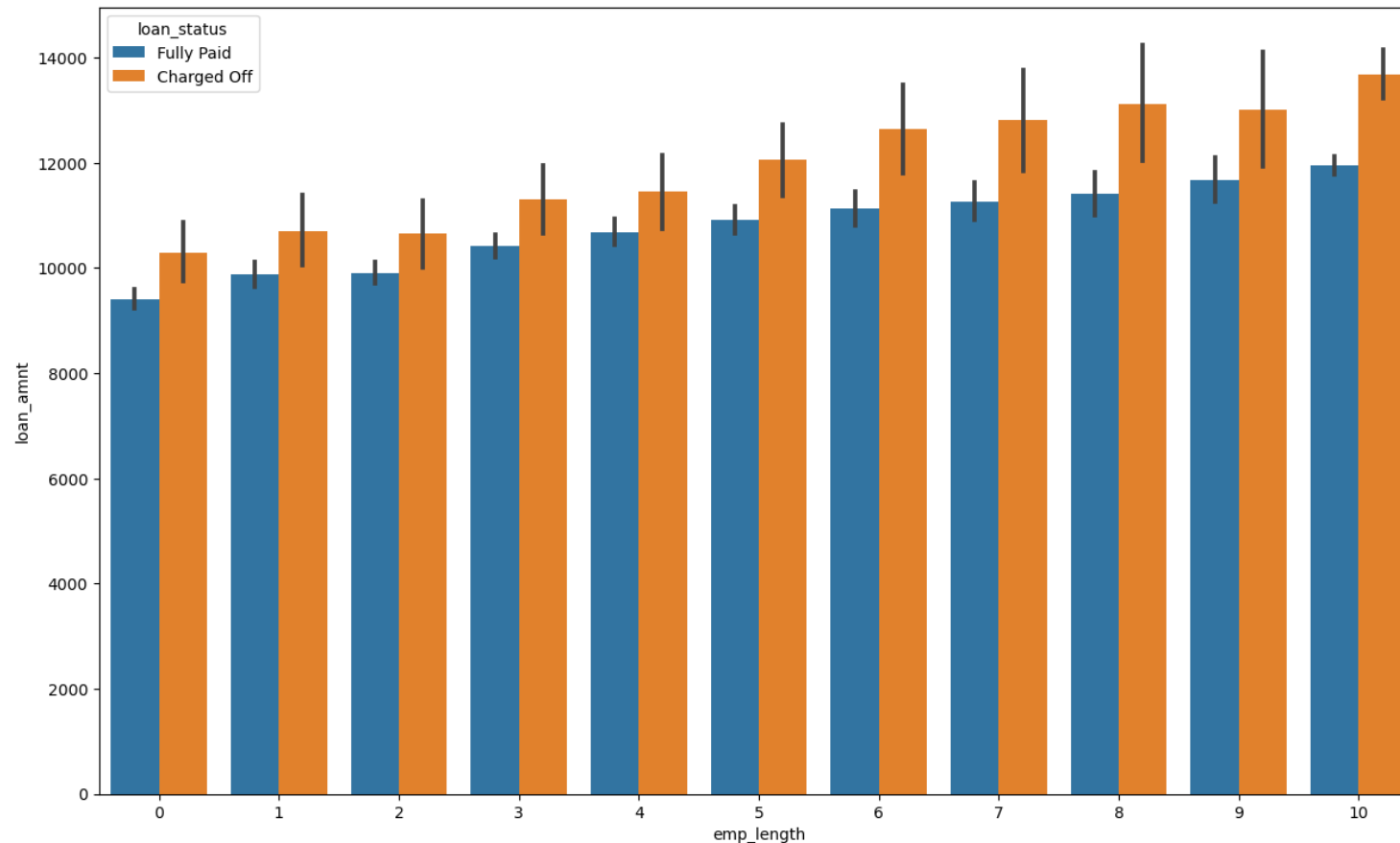
Charged off loan amounts are higher than fully paid loan amounts in all employee lengths

Grades

Grade G has the highest interest rate. Charged Off loan amounts are lower for Grade A and Grade G than Fully paid loan amounts.

Address State

This measure indicates the % of charged-off within this category vs total charged off across all categories. This indicates the **prominence** of the category or the number of incidences

This indicates the % of charged-off vs total closed loan applications (charged off + fully paid). This is the measure indicating **probability** of "charged off" for each of the categories.
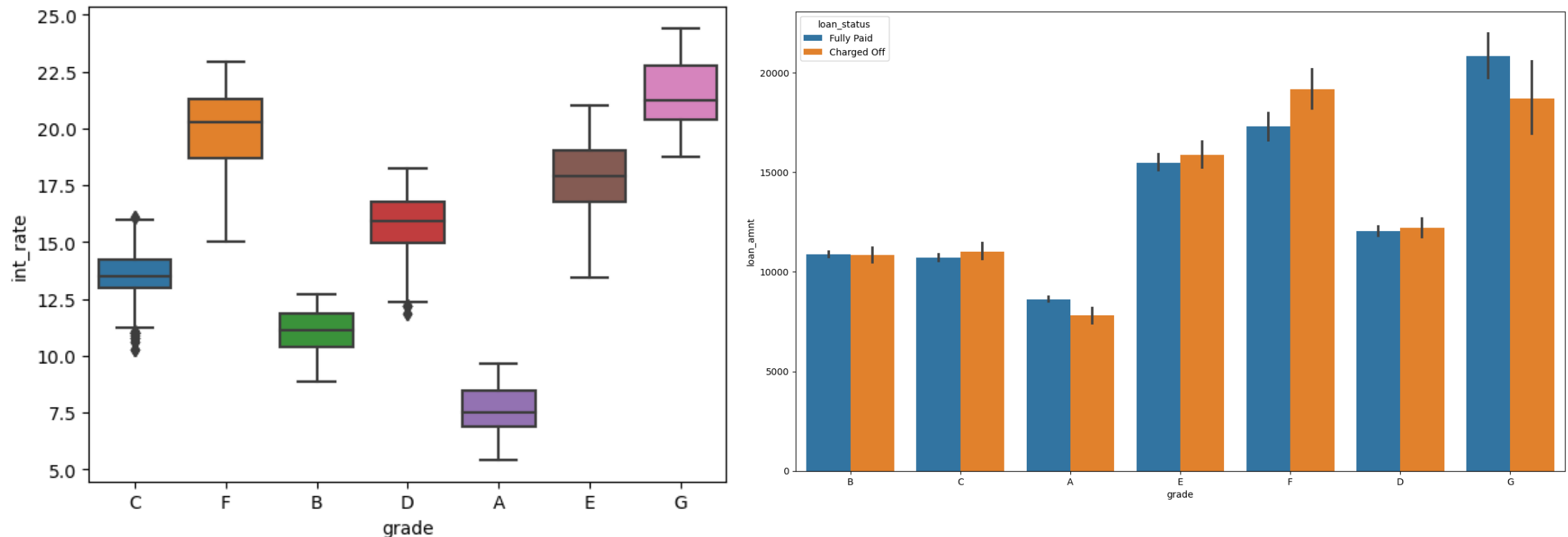
loan_status	Charged Off	Fully Paid	Charged_Off_perc_total	Fully_Paid_perc_total	Charged_Off_perc	Fully_Paid_perc
addr_state						
NE	3.00	2.00	0.05	0.01	60.00	40.00
NV	107.00	367.00	1.92	1.13	22.57	77.43
SD	12.00	49.00	0.22	0.15	19.67	80.33
AK	14.00	63.00	0.25	0.19	18.18	81.82
FL	496.00	2252.00	8.91	6.91	18.05	81.95
MO	113.00	551.00	2.03	1.69	17.02	82.98
HI	28.00	137.00	0.50	0.42	16.97	83.03
NM	30.00	149.00	0.54	0.46	16.76	83.24
ID	1.00	5.00	0.02	0.02	16.67	83.33
CA	1115.00	5760.00	20.03	17.68	16.22	83.78

State NE has 60% charge off to total loan ratio (high probability to default) but only has 0.05% of the number of incidences. State CA is the highest occurring state within charge off

Grades – Probability vs Incidences
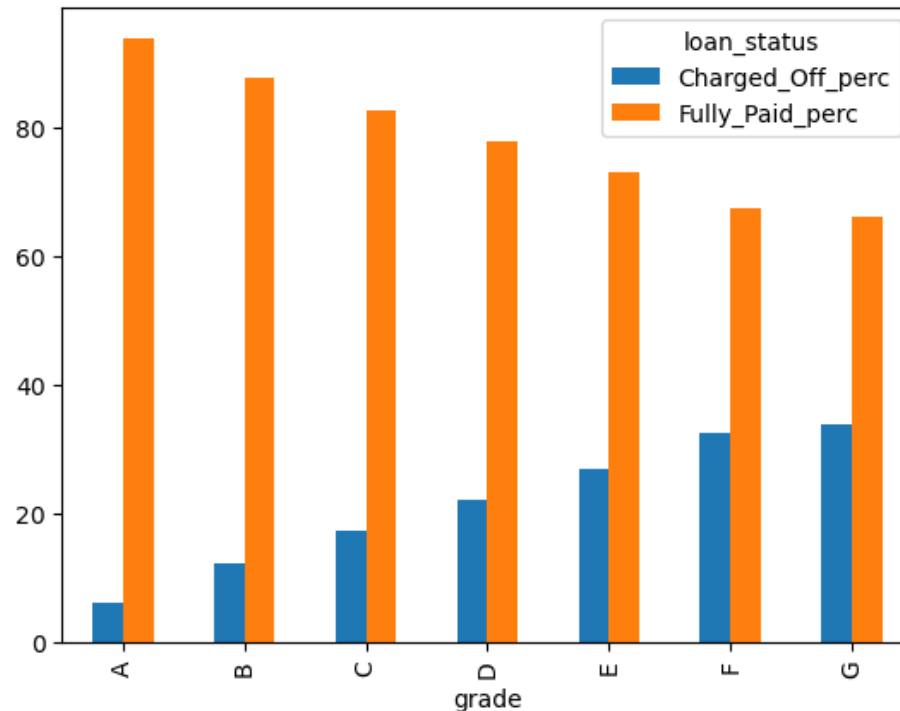
loan_status grade	Charged Off	Fully Paid	Charged_Off_perc_total	Fully_Paid_perc_total	Charged_Off_perc	Fully_Paid_perc
G	97	190	1.74	0.58	33.80	66.20
F	309	639	5.55	1.96	32.59	67.41
E	700	1899	12.57	5.83	26.93	73.07
D	1109	3903	19.92	11.98	22.13	77.87
C	1338	6401	24.03	19.65	17.29	82.71
B	1414	10166	25.40	31.21	12.21	87.79
A	601	9376	10.79	28.78	6.02	93.98

Grade G and F have the highest charge off % vs total loans (charge off + fully paid)
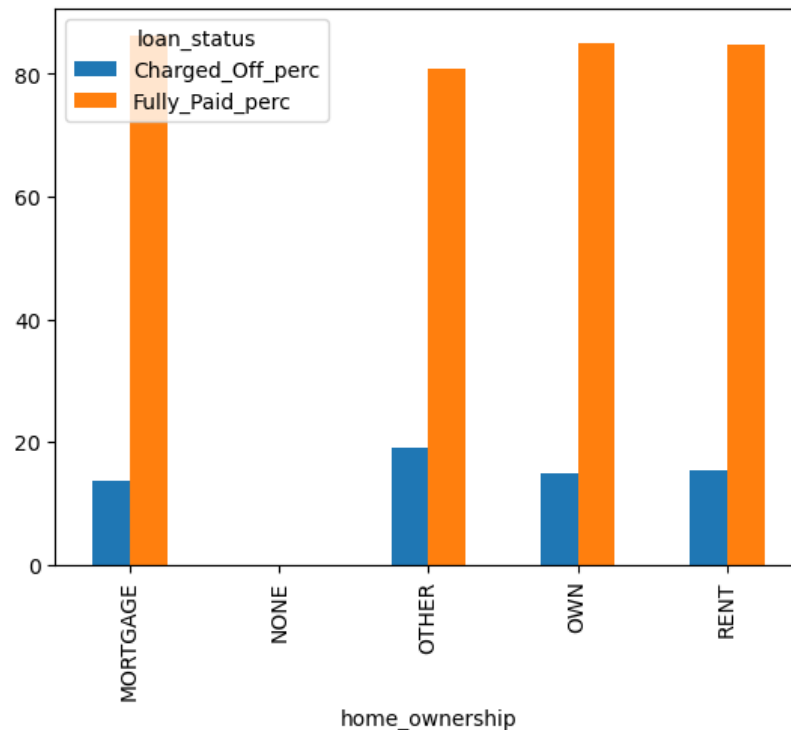
Home Ownership – Probability vs Incidences

loan_status	Charged Off	Fully Paid	Charged_Off_perc_total	Fully_Paid_perc_total	Charged_Off_perc	Fully_Paid_perc
home_ownership						
OTHER	18.00	76.00	0.32	0.23	19.15	80.85
RENT	2814.00	15551.00	50.54	47.74	15.32	84.68
OWN	438.00	2504.00	7.87	7.69	14.89	85.11
MORTGAGE	2298.00	14440.00	41.27	44.33	13.73	86.27

Home ownership Charge off % of 'Other' is 19% vs 15% for Rent. However Other makes up only 0.32% of the total charge off, while Rent makes up 51% of charge off

Employee Tenure – Probability vs Incidences

loan_status emp_length	Charged Off	Fully Paid	Charged_Off_perc_total	Fully_Paid_perc_total	Charged_Off_perc	Fully_Paid_perc
10	1541	7837	27.68	24.06	16.43	83.57
7	262	1432	4.71	4.40	15.47	84.53
1	452	2685	8.12	8.24	14.41	85.59
5	455	2709	8.17	8.32	14.38	85.62
8	202	1215	3.63	3.73	14.26	85.74
0	631	3836	11.33	11.78	14.13	85.87
6	303	1845	5.44	5.66	14.11	85.89
3	551	3422	9.90	10.51	13.87	86.13
4	454	2855	8.15	8.76	13.72	86.28
2	560	3680	10.06	11.30	13.21	86.79
9	157	1058	2.82	3.25	12.92	87.08

Employee tenure of 10 or more years has both the highest impact and probability of charge off. 0-1 years has the highest impact (% of total charge off), however the second highest probability of "charge off" (as indicated by charge off % vs total loan) is for 7 years at 15%
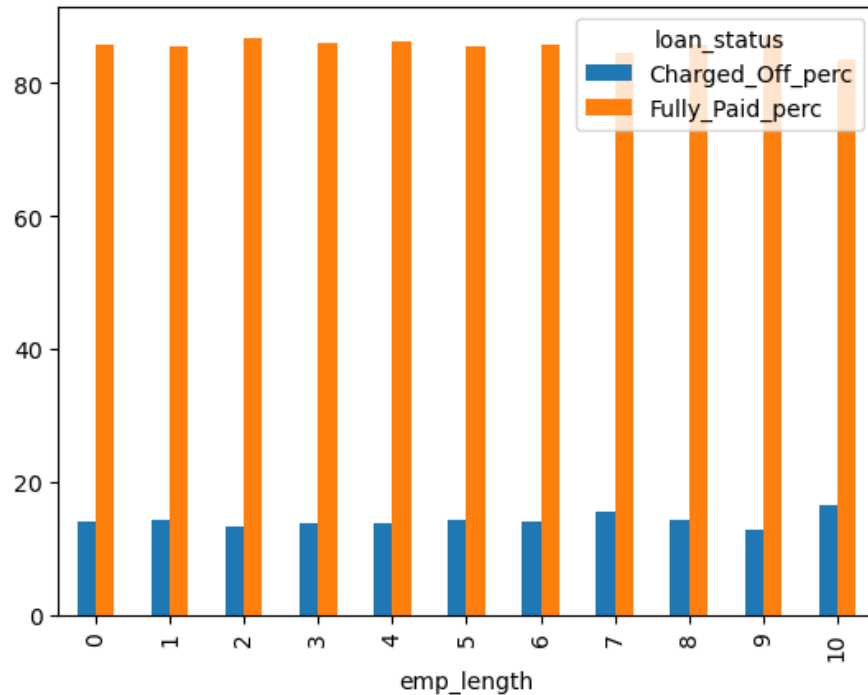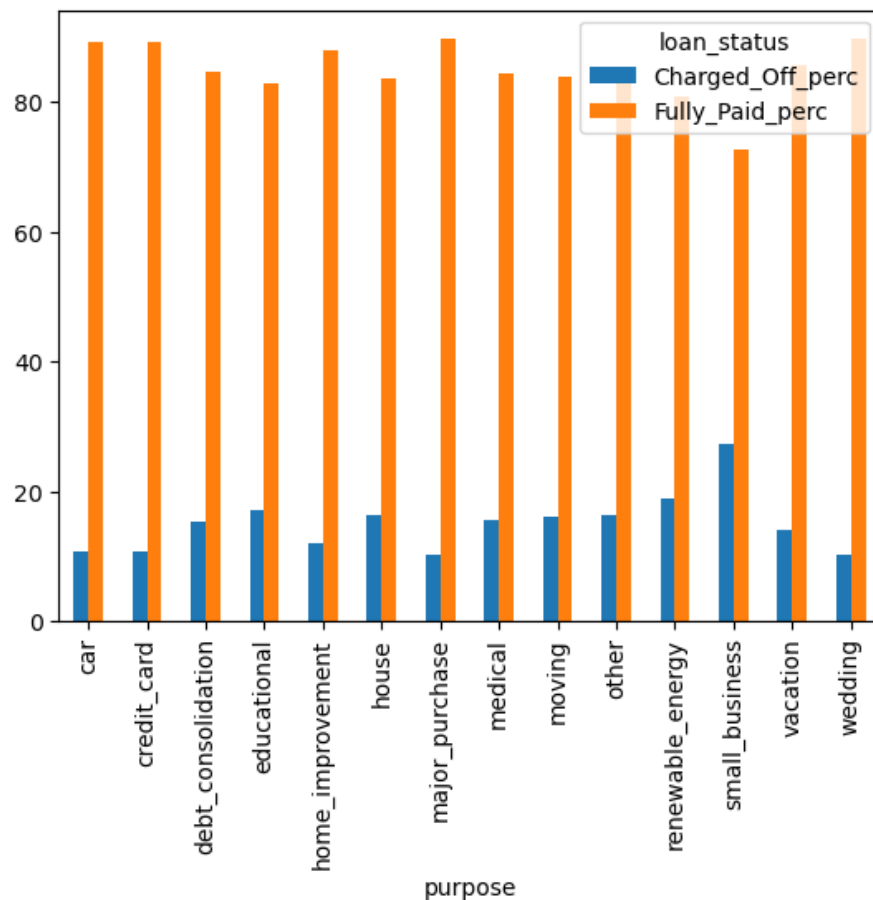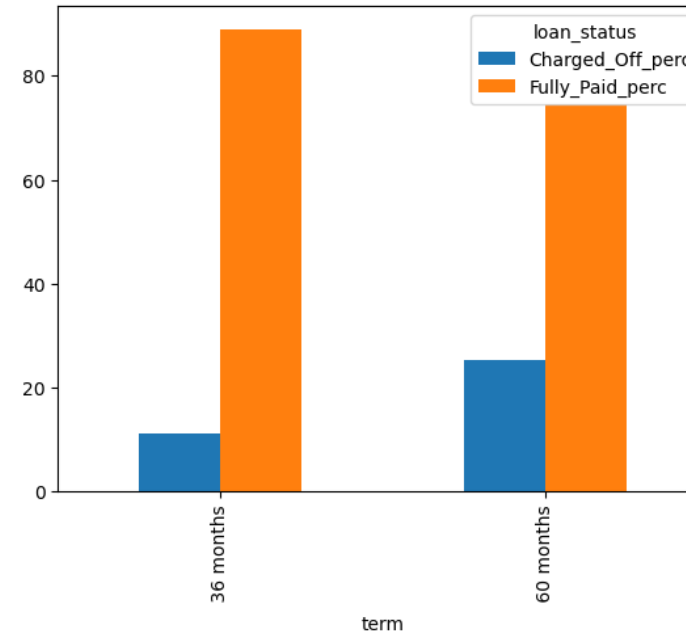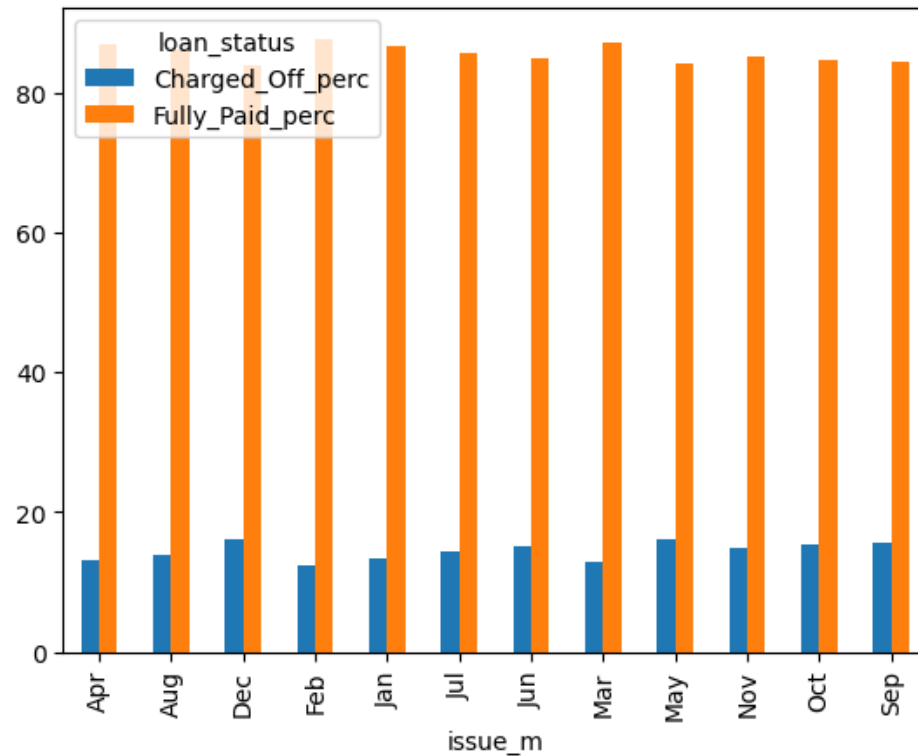
Purpose – Probability vs Incidences

loan_status purpose	Charged Off	Fully Paid	Charged_Off_perc_total	Fully_Paid_perc_total	Charged_Off_perc	Fully_Paid_perc
small_business	472	1250	8.48	3.84	27.41	72.59
renewable_energy	19	81	0.34	0.25	19.00	81.00
educational	55	265	0.99	0.81	17.19	82.81
other	625	3193	11.22	9.80	16.37	83.63
house	58	299	1.04	0.92	16.25	83.75
moving	91	477	1.63	1.46	16.02	83.98
medical	105	567	1.89	1.74	15.62	84.38
debt_consolidation	2747	15158	49.34	46.53	15.34	84.66
vacation	53	321	0.95	0.99	14.17	85.83
home_improvement	335	2457	6.02	7.54	12.00	88.00
credit_card	536	4449	9.63	13.66	10.75	89.25
car	159	1330	2.86	4.08	10.68	89.32
major_purchase	219	1905	3.93	5.85	10.31	89.69
wedding	94	822	1.69	2.52	10.26	89.74

"Small business" has the highest probability of "charge off" based on charge off % vs total loan, however debt consolidation occurs more frequently within the charge off dataset

Issue Month & Term – Probability vs Incidences

- Issue Mon - December is the highest in terms of both the probability and number of incidences within charge off.
- Term - 60 months has the highest probability of charge off

Income Group – Probability vs Incidences
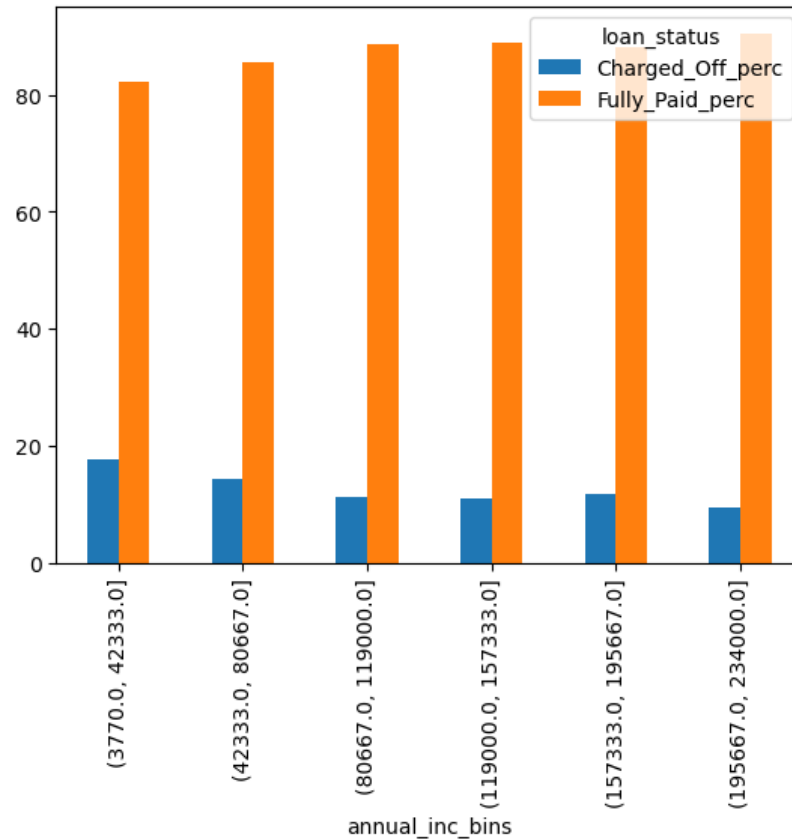
loan_status annual_inc_bins	Charged Off	Fully Paid	Charged_Off_perc_total	Fully_Paid_perc_total	Charged_Off_perc	Fully_Paid_perc
(3770.0, 42333.0]	1910	8845	34.30	27.15	17.76	82.24
(42333.0, 80667.0]	2589	15302	46.50	46.98	14.47	85.53
(157333.0, 195667.0]	77	576	1.38	1.77	11.79	88.21
(80667.0, 119000.0]	706	5518	12.68	16.94	11.34	88.66
(119000.0, 157333.0]	256	2048	4.60	6.29	11.11	88.89
(195667.0, 234000.0]	30	285	0.54	0.87	9.52	90.48

The lowest income group i.e. (3770.0, 42333.0] has the highest probability of charge off and makes up second highest proportion in the number of incidences within charge off (34%)

Sub Grade – Probability vs Incidences

| loan_status sub_grade	Charged Off	Fully Paid	Charged_Off_perc_total	Fully_Paid_perc_total	Charged_Off_perc	Fully_Paid_perc
F5	53	55	0.95	0.17	49.07	50.93
G3	19	25	0.34	0.08	43.18	56.82
G5	10	18	0.18	0.06	35.71	64.29
G2	26	47	0.47	0.14	35.62	64.38
F4	51	94	0.92	0.29	35.17	64.83
G1	30	61	0.54	0.19	32.97	67.03
E4	126	293	2.26	0.90	30.07	69.93
F3	50	118	0.90	0.36	29.76	70.24
F1	88	210	1.58	0.64	29.53	70.47
F2	67	162	1.20	0.50	29.26	70.74
E5	105	260	1.89	0.80	28.77	71.23
E1	195	517	3.50	1.59	27.39	72.61
E2	157	444	2.82	1.36	26.12	73.88
D5	209	615	3.75	1.89	25.36	74.64
D4	214	690	3.84	2.12	23.67	76.33
G4	12	39	0.22	0.12	23.53	76.47
E3	117	385	2.10	1.18	23.31	76.69
D3	252	848	4.53	2.60	22.91	77.09
D2	268	998	4.81	3.06	21.17	78.83
C3	269	1198	4.83	3.68	18.34	81.66
C5	207	928	3.72	2.85	18.24	81.76
D1	166	752	2.98	2.31	18.08	81.92
C4	211	987	3.79	3.03	17.61	82.39
C2	318	1591	5.71	4.88	16.66	83.34
C1	333	1697	5.98	5.21	16.40	83.60
B5	356	2244	6.39	6.89	13.69	86.31
B4	328	2096	5.89	6.43	13.53	86.47

F5 sub grade has the highest probability to charge off at 49% but only makes up <1% of the number of incidences

Key Takeaways

- Charged off loan amounts are higher than fully paid loan amounts in all income categories and employee lengths

- Charged Off loan amounts are lower for Grade A and Grade G than Fully paid loan amounts.

- Grade G has the highest interest rate and also the highest charge off % vs total loans (charge off + fully paid)

- State NE has 60% charge off to total loan ratio but only has 0.05% of the number of incidences. State CA is the highest occurring state within charge off

- Home ownership Charge off % of 'Other' is 19% vs 15% for Rent. However Other makes up only 0.32% of the total charge off, while Rent makes up 51% of charge off

- Employee tenure of 10 or more years has both the highest impact and probability of charge off. 0-1 years has the highest impact (% of total charge off), however the second highest probability of "charge off" (as indicated by charge off % vs total loan) is for 7 years at 15%

- Purpose - "small business" has the highest probability of "charge off" based on charge off % vs total loan, however debt consolidation occurs more frequently within the charge off dataset

- Issue Mon - December is the highest in terms of both the probability and number of incidences within charge off. 60 months has the highest probability of charge off

- The lowest income group i.e. (3770.0, 42333.0] has the highest probability of charge off and makes up second highest proportion in the number of incidences within charge off (34%)

- F5 sub grade has the highest probability to charge off at 49% but only makes up <1% of the number of incidences

Correlations Observations

- High correlation among loan amount, investor amount, installment amount and funding amount. Loan amount is a stronger direct indicator

- Annual income has negative correlation to DTI (debt to income ratio)

- Interest rate and loan amount are correlated
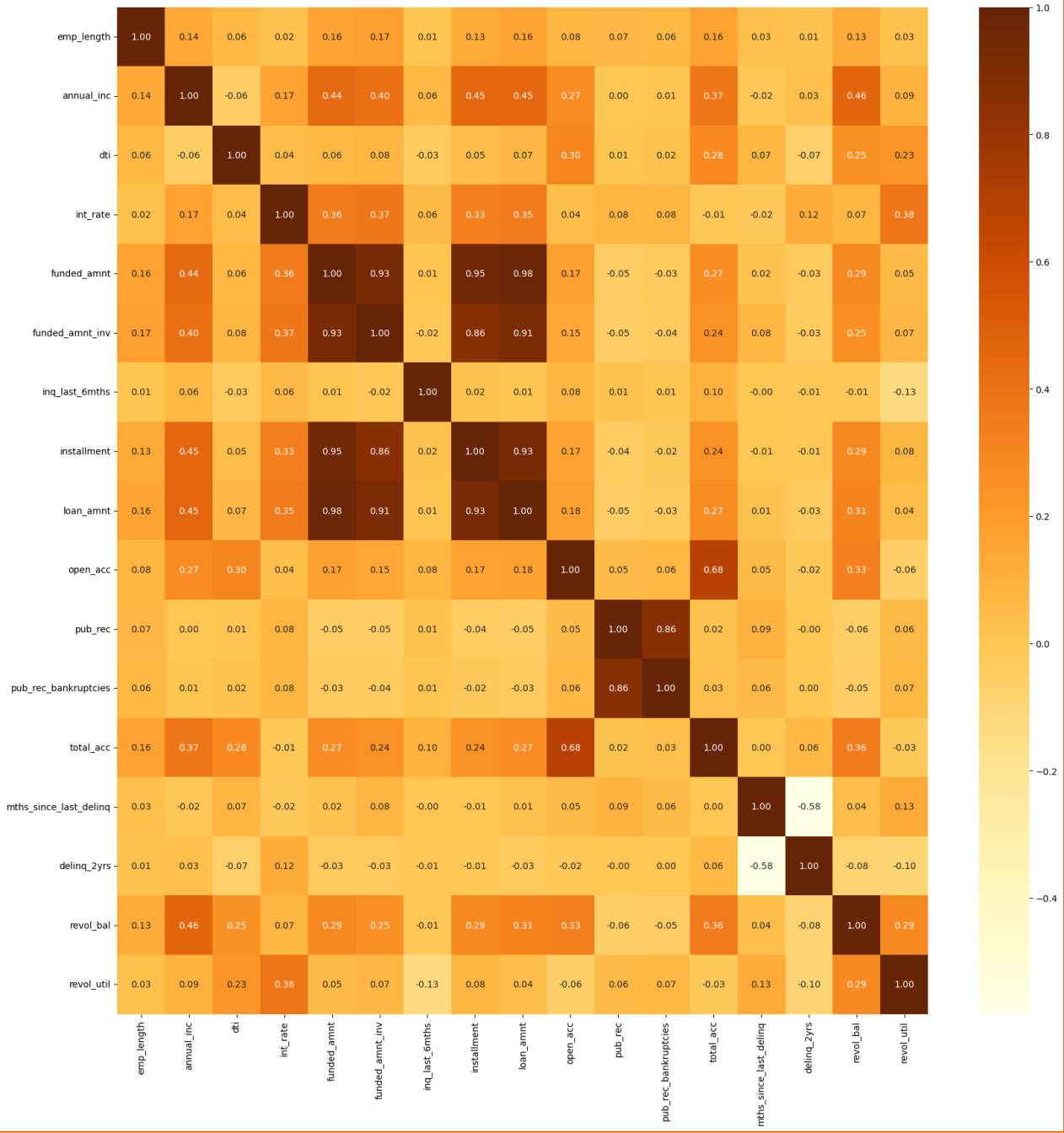
Correlations Observations

- High correlation among loan amount, investor amount, installment amount and funding amount. Loan amount is a stronger direct indicator

- Annual income has negative correlation to DTI (debt to income ratio)

- Interest rate and loan amount are correlated

Summary and Recommendations

The following features provide a set of strong indicators for the Lending club to perform a risk assessment on the loan applications.

1. Grades and higher interest rates are robust indicators to understand probability of defaulting. Lending club should have additional screening measures in place for grades such as G and high interest rates.

2. The probability for "charge off" is much higher in the following cases, therefore appropriate actions need to be taken to minimise the risk of loss.

    a. Applications with 'Purpose' of Small Business
    b. Applications with Home ownership of "Renting"
    c. Applications with Employee tenure of 10 years or more
    d. Applications with Lowest income group i.e. (3770.0, 42333.0]
    e. Applications from State CA etc
    f. Applications with Tenure of 60 months
    g. Public records and earlier bankruptcies