

LENDING CLUB CASE STUDY

Case Study Group:

- Akshaya Kumaar
- Anil Juneja

Problem Statement :

You work for a consumer finance company which 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

Objective:

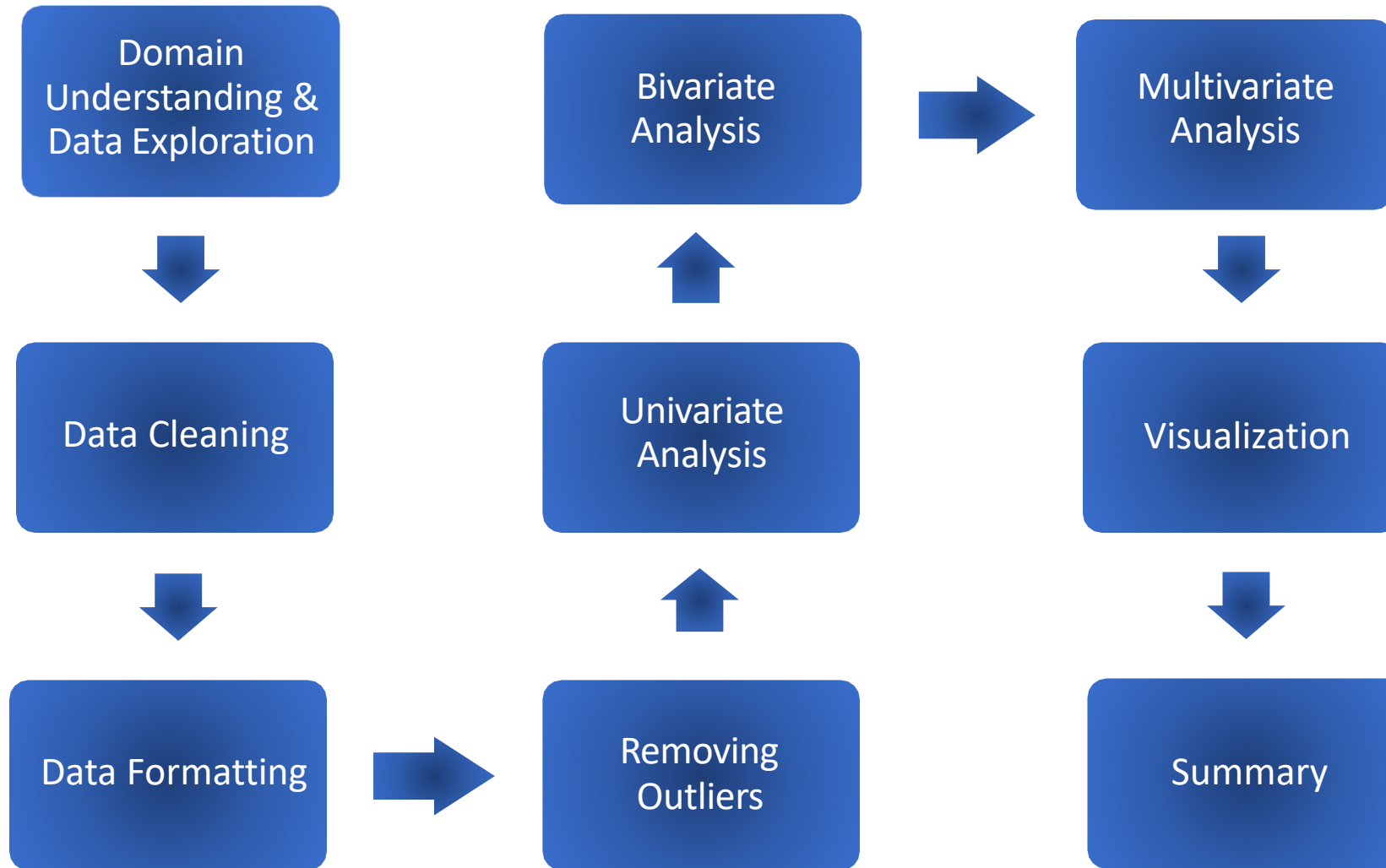
- Use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

Constraints:

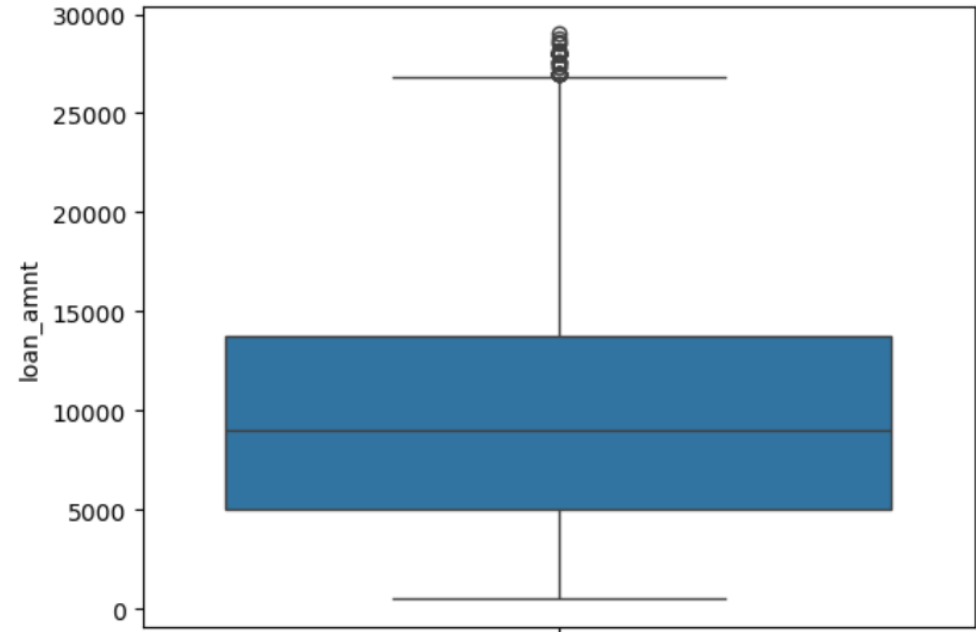
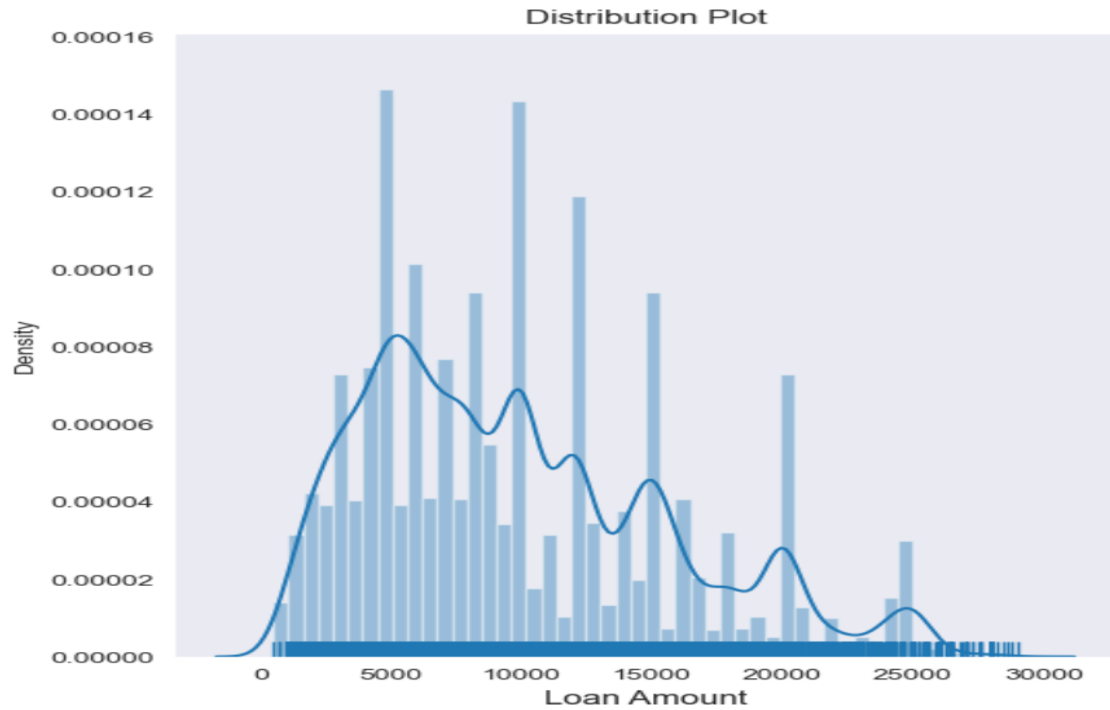
When a person applies for a loan, there are two types of decisions that could be taken by the company:

- Loan accepted: If the company approves the loan, there are 3 possible scenarios described below:
 - Fully paid: Applicant has fully paid the loan (the principal and the interest rate)
 - Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
 - Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan
- Loan rejected: The company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

Overall Approach of the Exploratory Data Analysis :



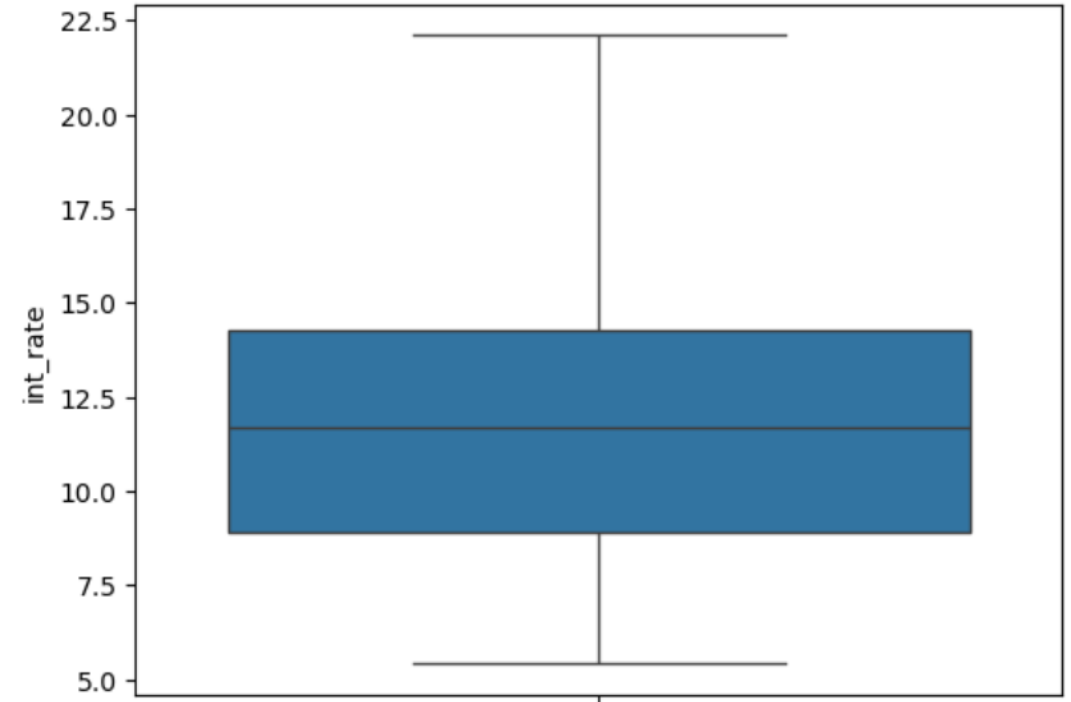
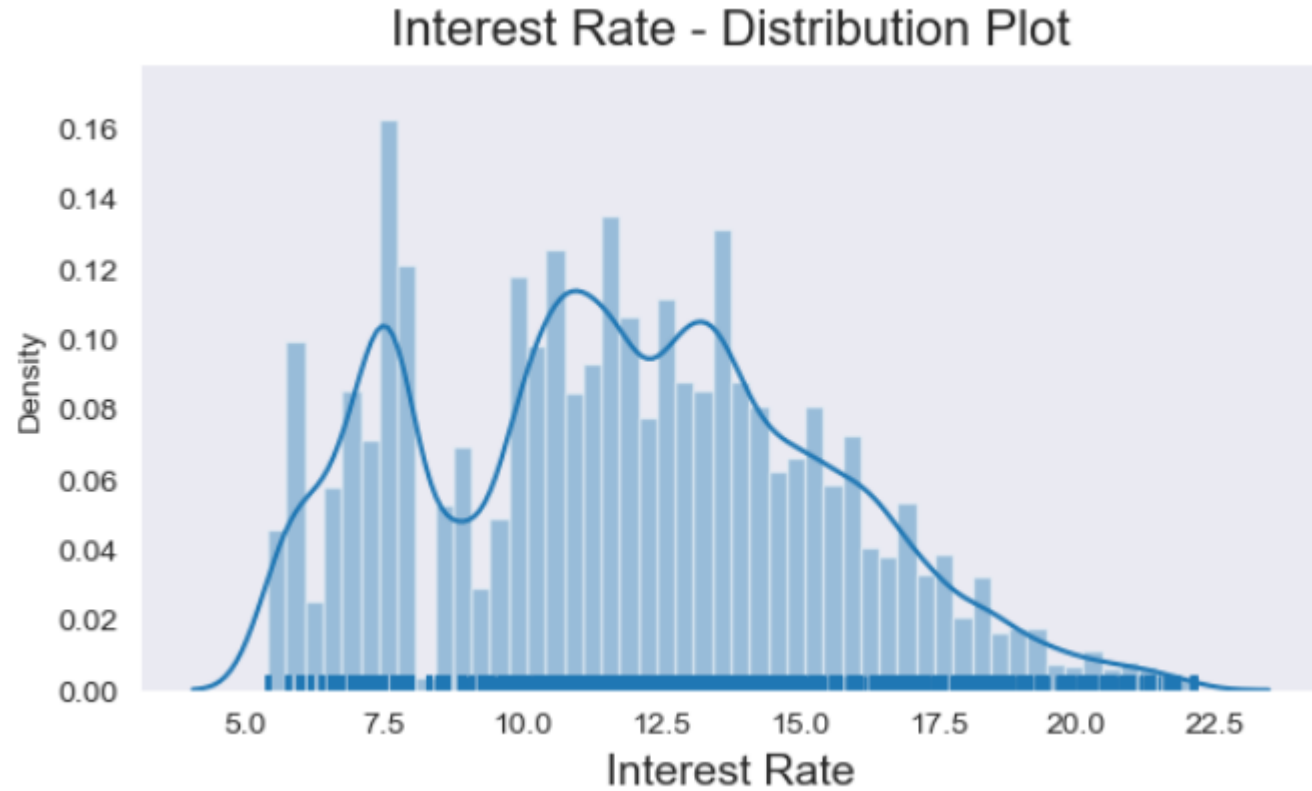
Loan Amount



Observation:

- Most of the loan amount applied for is in the range of 5k-14k with some outliers at the 20k mark.

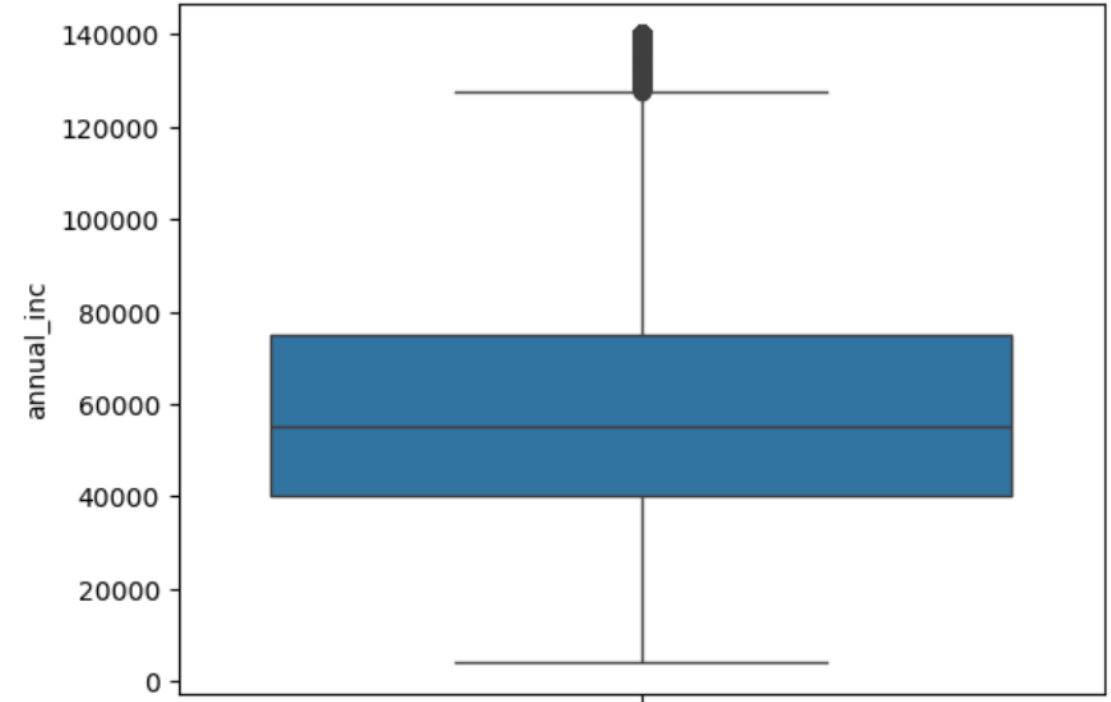
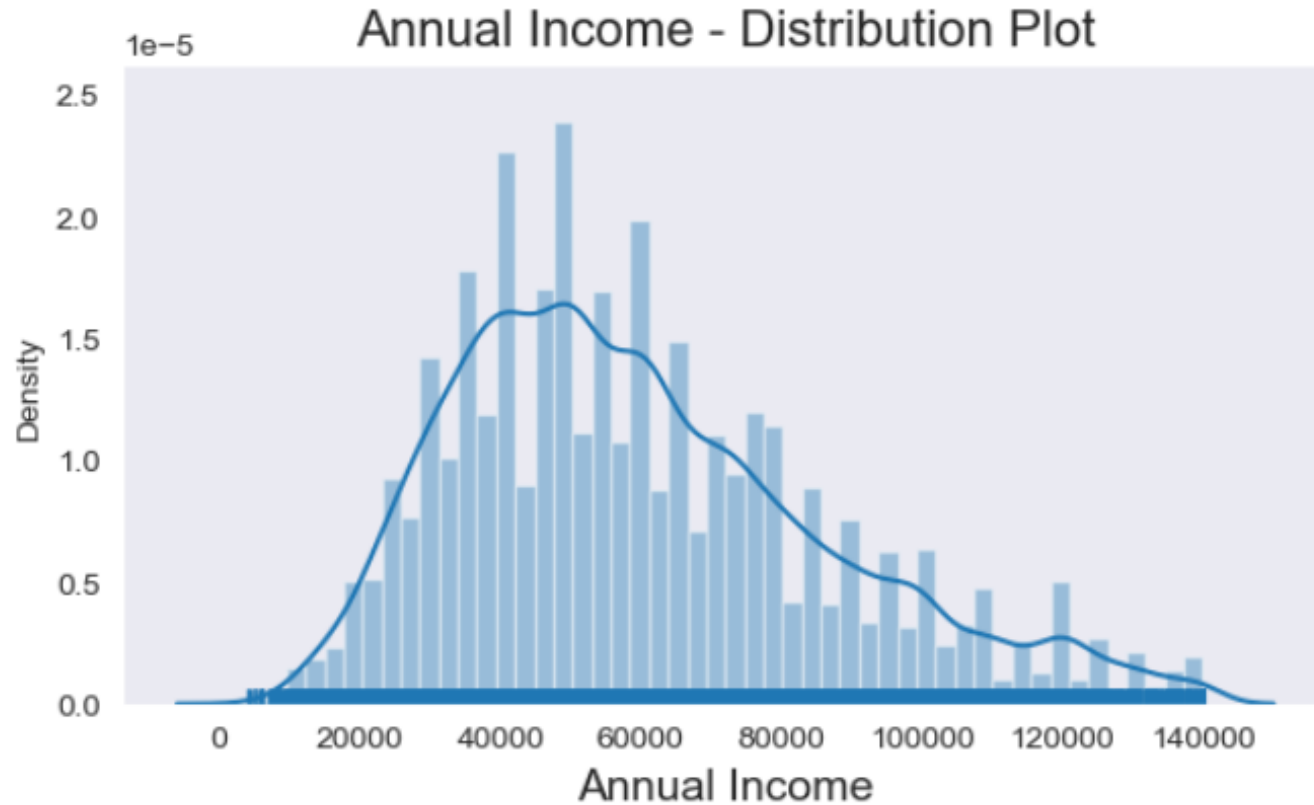
Interest Rate



Observation:

- Most of the applicant's rate of interest is between the range of 8%-14%.

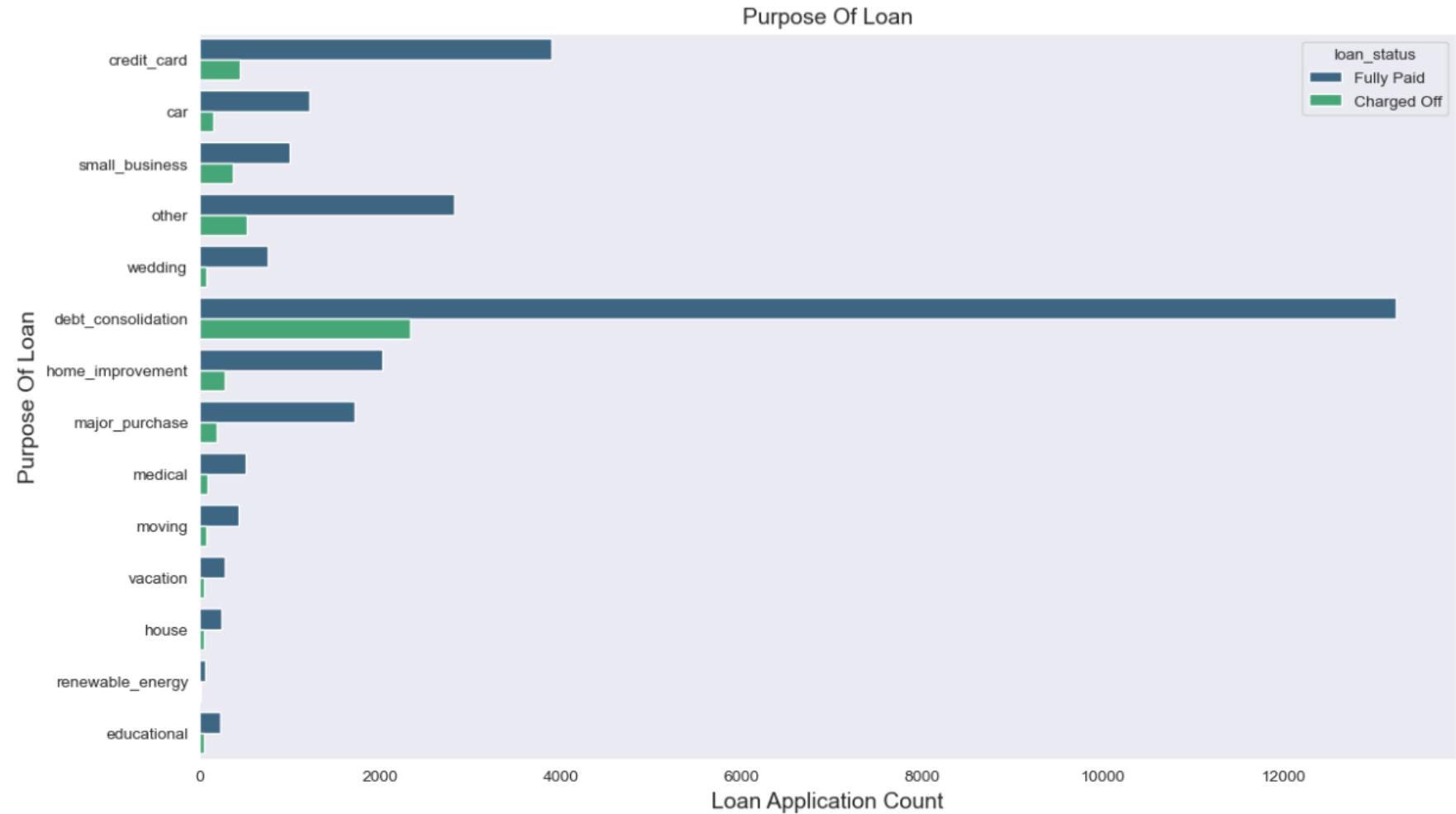
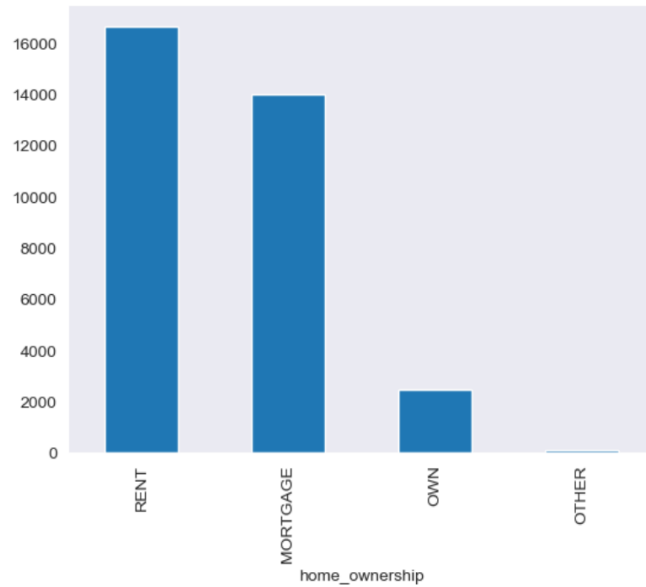
Annual Income



Observation:

- The Annual income of most of the applicants lies between the range of 40k-80k
- The highest peak in the distribution plot seems to be around at approx. 50k income.

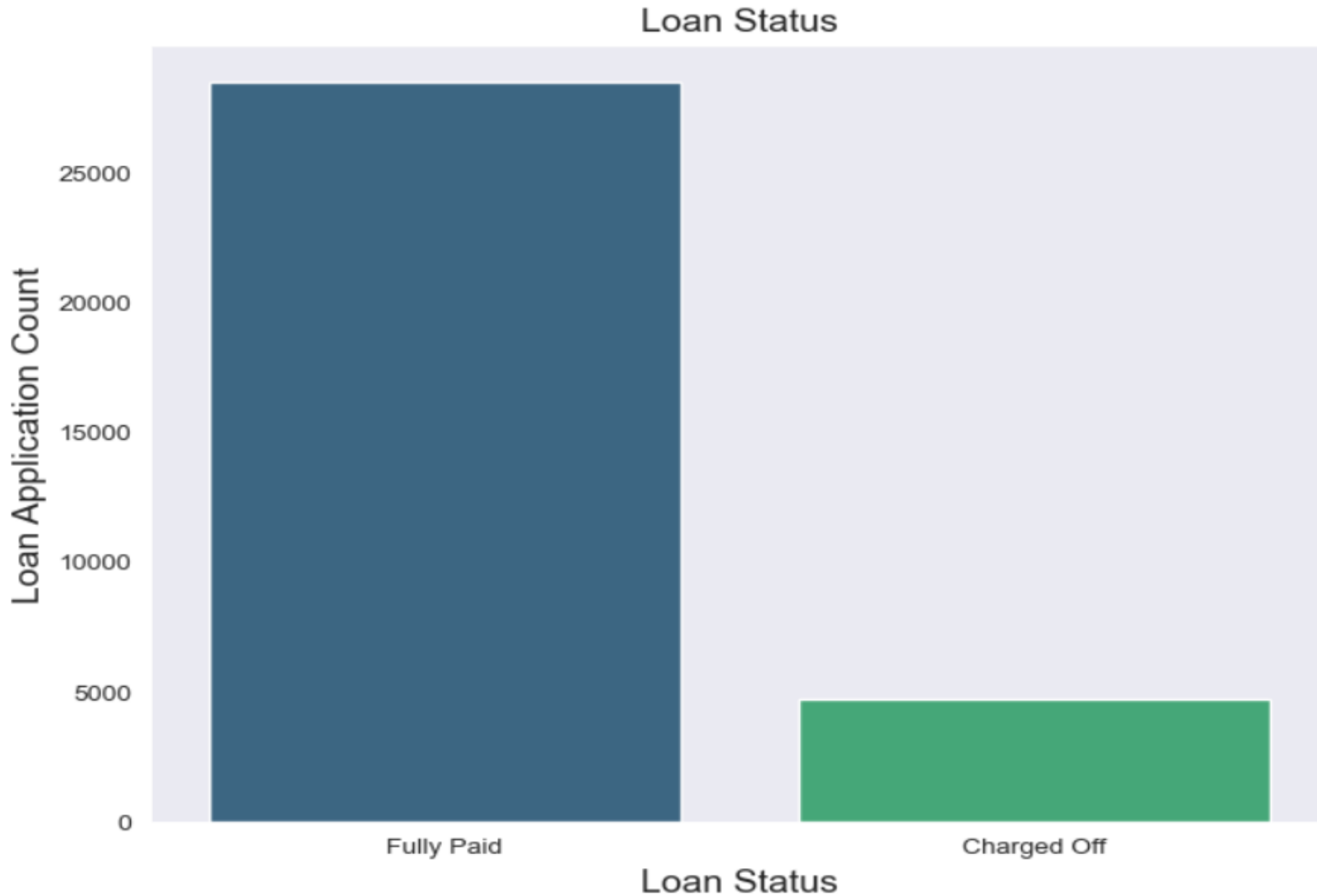
Purpose & Home Ownership



Observations:

- Majority of loan applicants fall under the Home Ownership category of either Rent or Mortgage
- Most of the loan applicants take out the loan firstly for pre-existing debt consolidations and secondly for credit card debt.

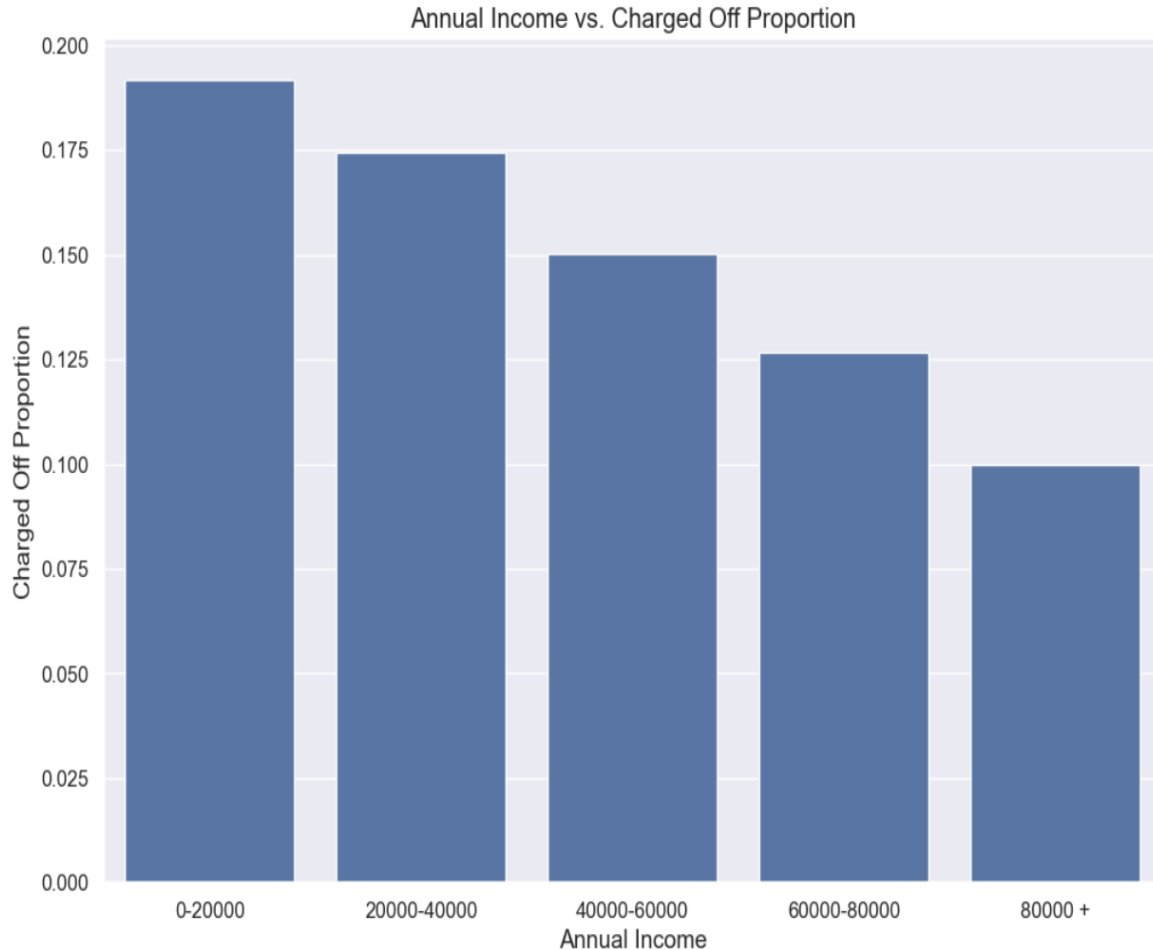
Loan Status



Observation:

- More than 80% of the total loans were fully paid out.
- Around 14% were Charged Off.

Annual Income vs Charged Off

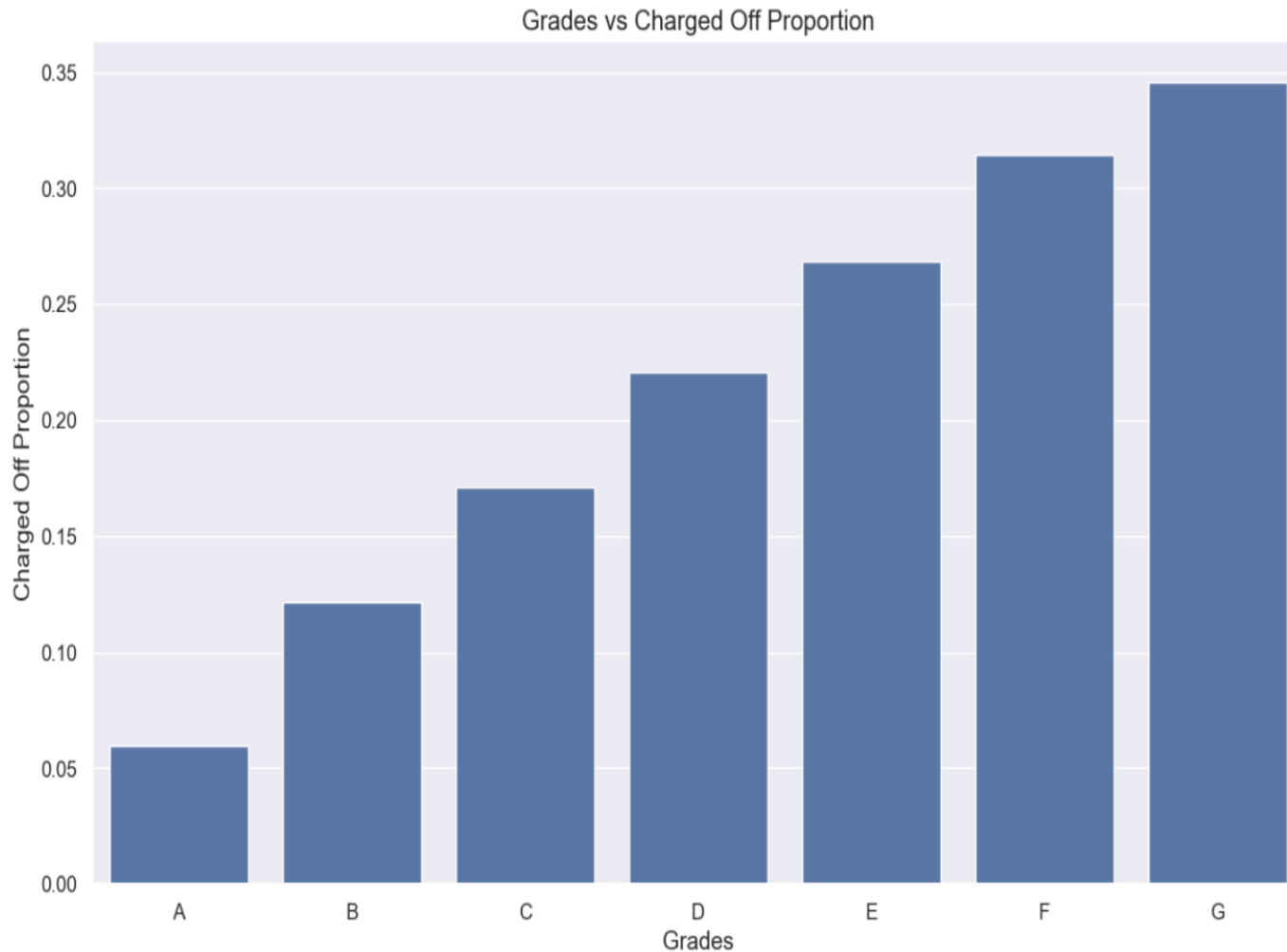


loan_status	annual_inc_bin	Charged Off	Fully Paid	Total	Charged_off_Proportion
0	0-20000	186	783	969	0.19
1	20000-40000	1384	6543	7927	0.17
2	40000-60000	1595	9028	10623	0.15
3	60000-80000	871	6010	6881	0.13
4	80000 +	678	6113	6791	0.10

Observations:

- Income range 0-20k has the highest probability of being charged off.
- Income range 80k+ has the lowest probability of being charged off.
- It can be easily observed that with increase in Annual Income, the probability of being Charged off gradually decreases.

Grades vs Charged Off



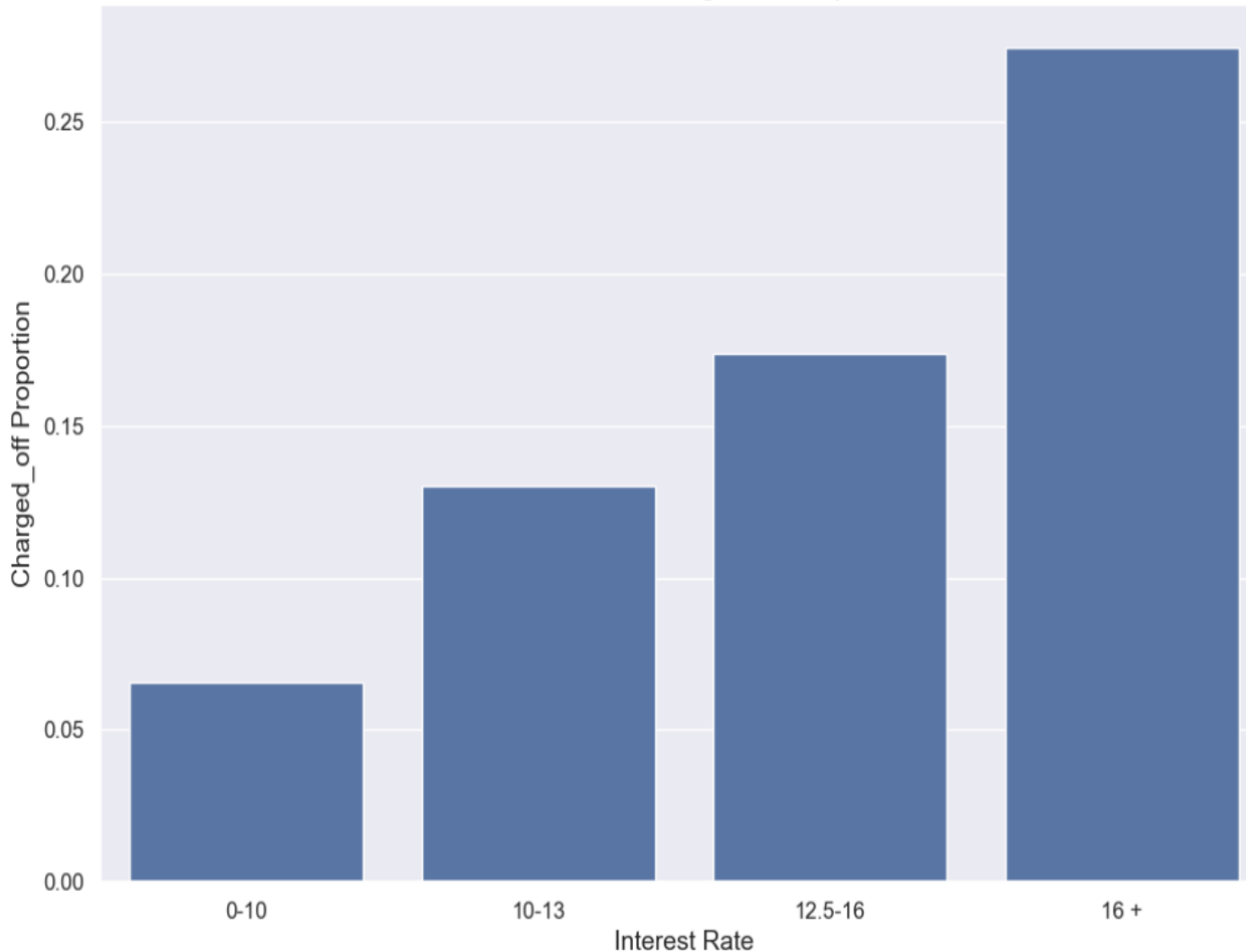
loan_status	grade	Charged Off	Fully Paid	Total	Charged_off_Proportion
6	G	55	104	159	0.35
5	F	220	480	700	0.31
4	E	557	1518	2075	0.27
3	D	941	3329	4270	0.22
2	C	1177	5702	6879	0.17
1	B	1225	8857	10082	0.12
0	A	539	8487	9026	0.06

Observations:

- Grade "A" has the least probability of being charged off.
- Grade "F" and "G" have very high probability of charged off, with "G" being the highest.
- Probability of being charged off increases while moving from grade "A" towards "G".

Interest Rate vs Charged Off

Interest Rate vs Charged Off Proportion

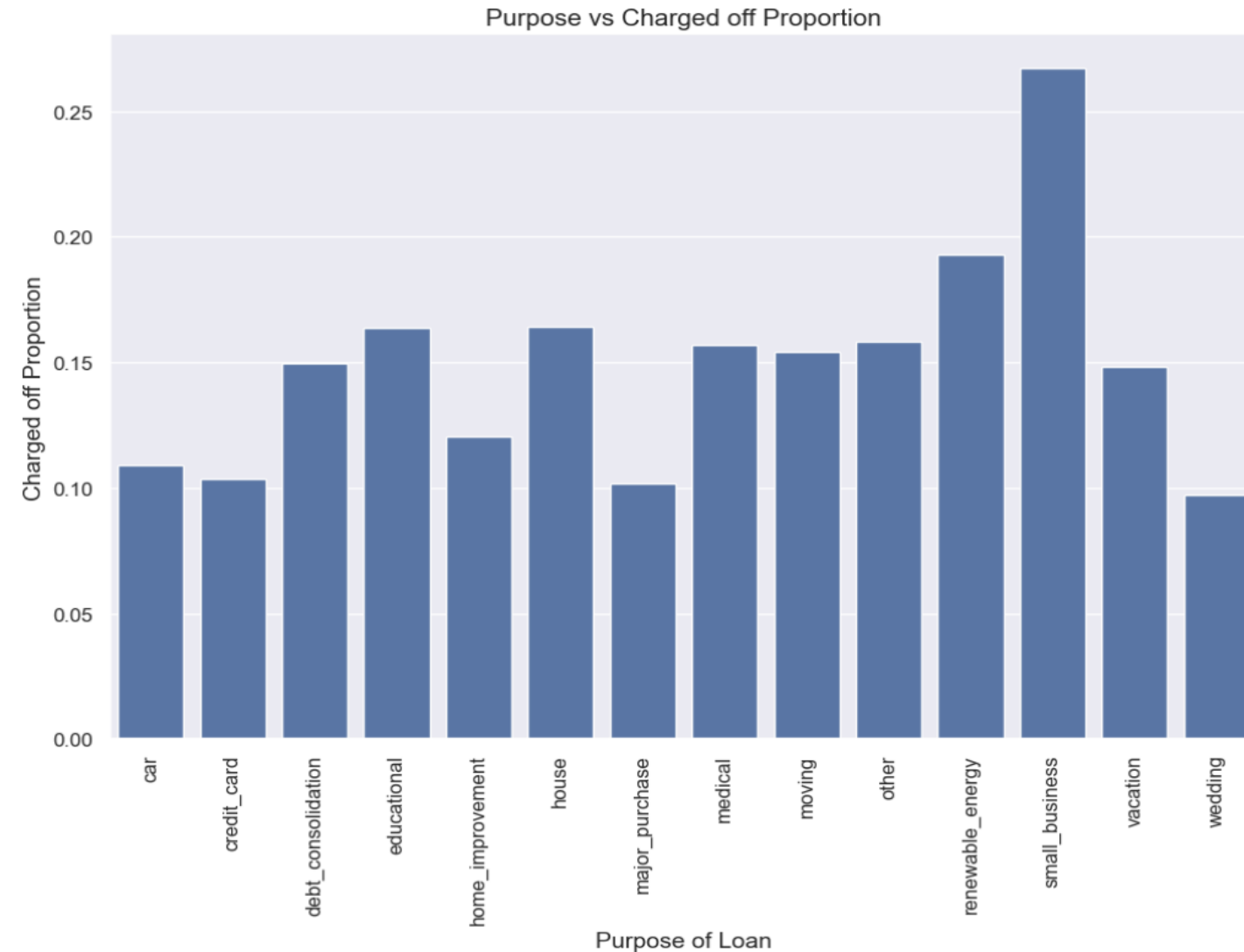


loan_status	int_rate_bin	Charged Off	Fully Paid	Total	Charged_off_Proportion
3	16 +	1026	2712	3738	0.27
2	12.5-16	1749	8318	10067	0.17
1	10-13	1045	6968	8013	0.13
0	0-10	716	10204	10920	0.07

Observations:

- Interest rate of $\leq 10\%$ has very low probability of being charged off.
- Interest rate $\geq 16\%$ has a higher probability of being charged off as compared to others.
- Charged off increases proportionately with increase in interest rate

Purpose vs Charged Off

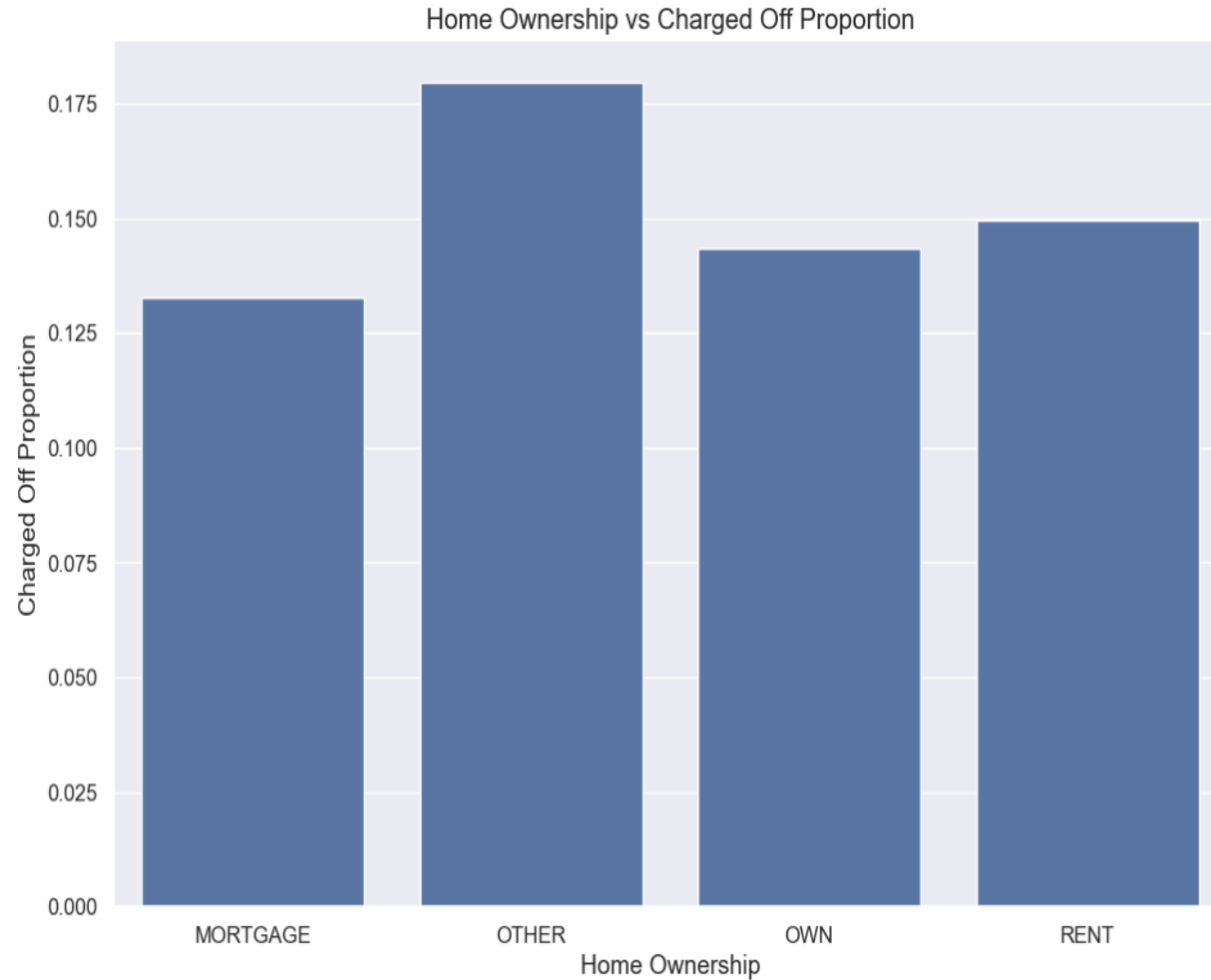


loan_status	purpose	Charged Off	Fully Paid	Total	Charged_off_Proportion
11	small_business	366	1003	1369	0.27
10	renewable_energy	16	67	83	0.19
5	house	49	249	298	0.16
3	educational	46	235	281	0.16
9	other	531	2823	3354	0.16
7	medical	95	510	605	0.16
8	moving	79	433	512	0.15
2	debt_consolidation	2329	13253	15582	0.15
12	vacation	49	281	330	0.15
4	home_improvement	277	2026	2303	0.12
0	car	150	1224	1374	0.11
1	credit_card	450	3894	4344	0.10
6	major_purchase	195	1719	1914	0.10
13	wedding	82	760	842	0.10

Observations:

- Small Business applicants have the highest probability of getting charged off.
- Renewable energy & Educational purpose have a higher probability of being charged off in comparison to other categories.

Home Ownership vs Charged Off

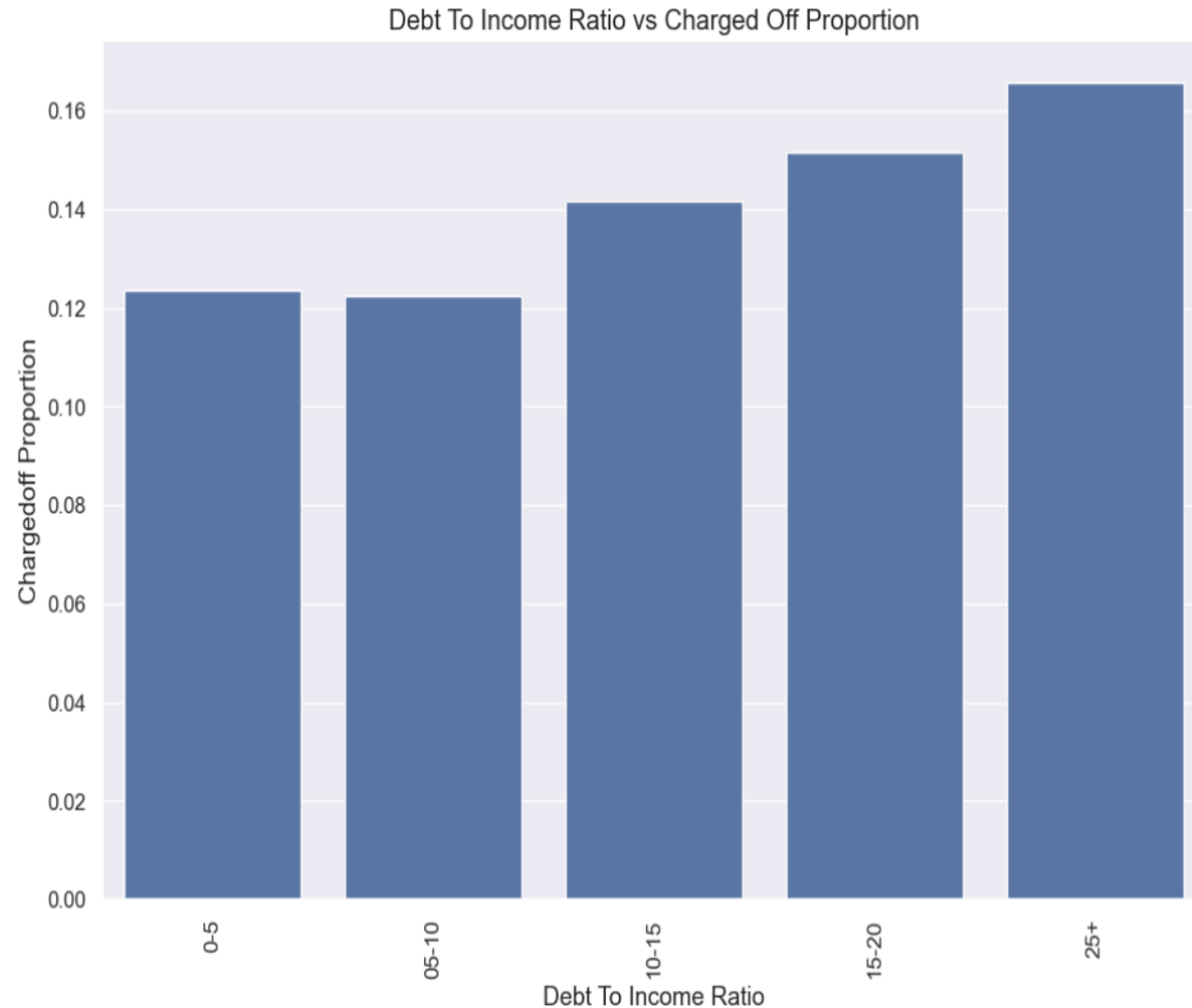


loan_status	home_ownership	Charged Off	Fully Paid	Total	Charged_off_Proportion
1	OTHER	16	73	89	0.18
3	RENT	2488	14156	16644	0.15
2	OWN	355	2121	2476	0.14
0	MORTGAGE	1855	12127	13982	0.13

Observations:

- Applicants who are in the Mortgage and Rent category have a higher probability of being charged off as the number of applicants are higher for these categories.
- Their proportion might seem lower but overall number of loan applications is higher thus they have a higher probability.

Debt to Income ratio vs Charged Off

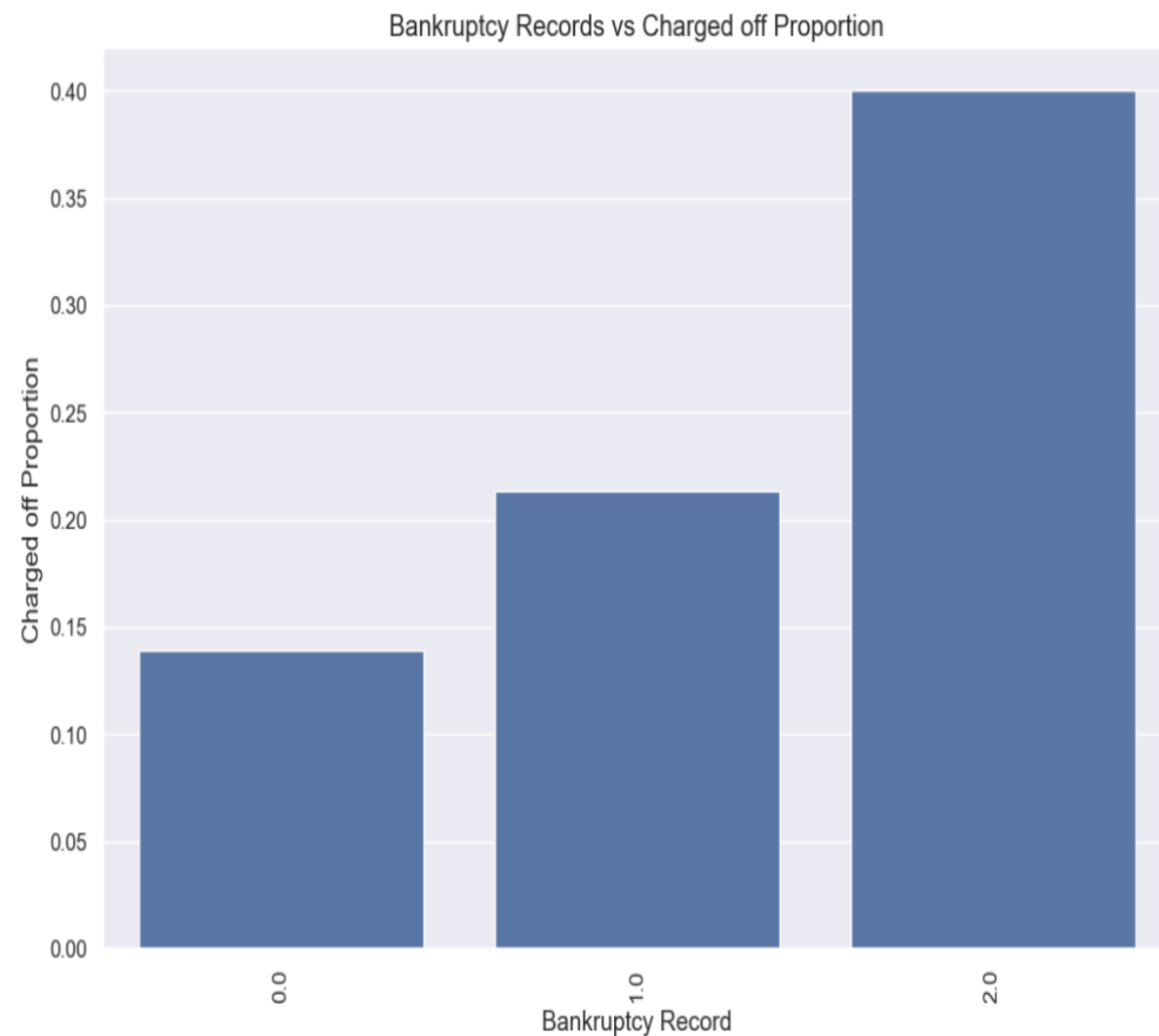


loan_status	dti_bin	Charged Off	Fully Paid	Total	Chargedoff_Proportion
4	25+	976	4910	5886	0.17
3	15-20	1166	6529	7695	0.15
2	10-15	1179	7151	8330	0.14
0	0-5	495	3508	4003	0.12
1	05-10	806	5781	6587	0.12

Observations:

- Higher the Debt to Income ratio, higher the probability of being charged off and vice versa
- Charged off proportion increases/decreases with increase/decrease in DTI.

Bankruptcy Records vs Charged Off

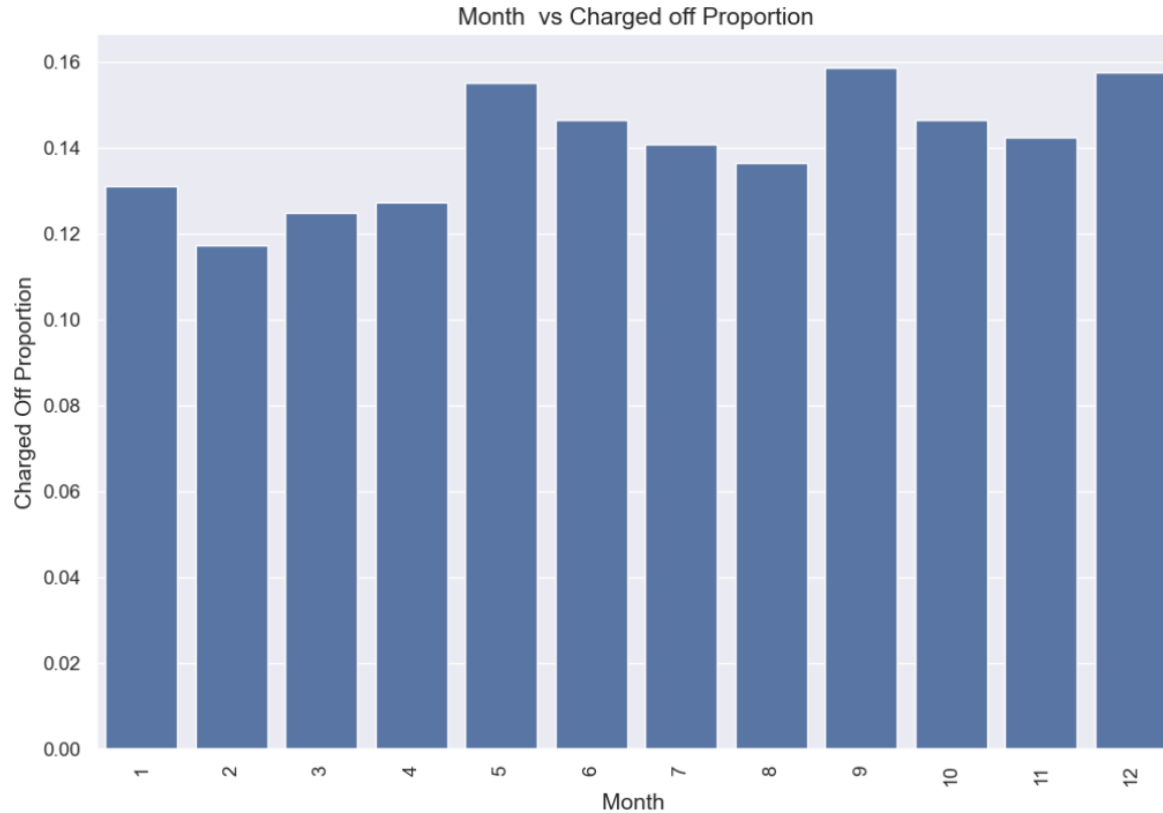


loan_status	pub_rec_bankruptcies	Charged Off	Fully Paid	Total	Chargedoff_Proportion
2	2.00	2	3	5	0.40
1	1.00	308	1137	1445	0.21
0	0.00	4404	27337	31741	0.14

Observations:

- Higher the public bankruptcy record greater the Charged Off proportion.

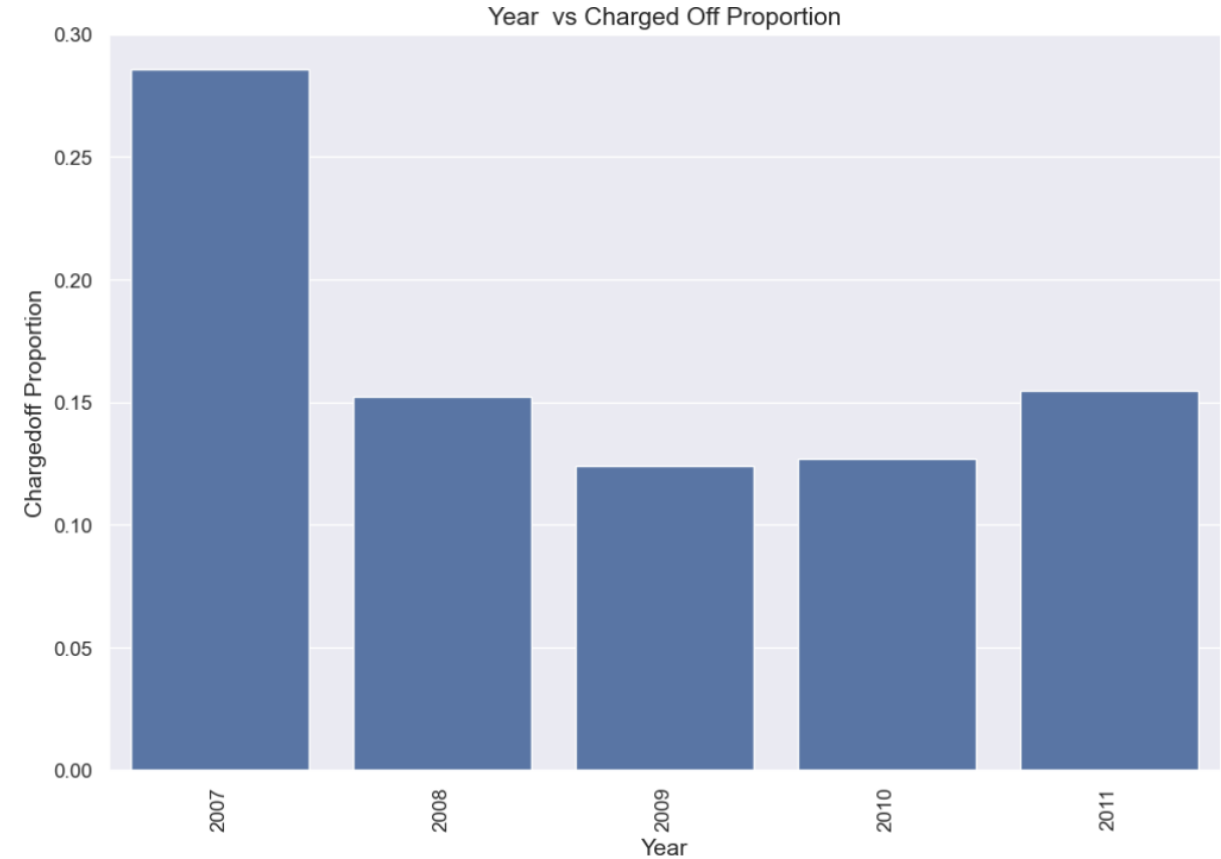
Month vs Charged Off



Observations:

- Loans approved in the months of May, September and December have higher probability of Charged Off.
- Loans approved in the month of February has the least probability of Charged Off.

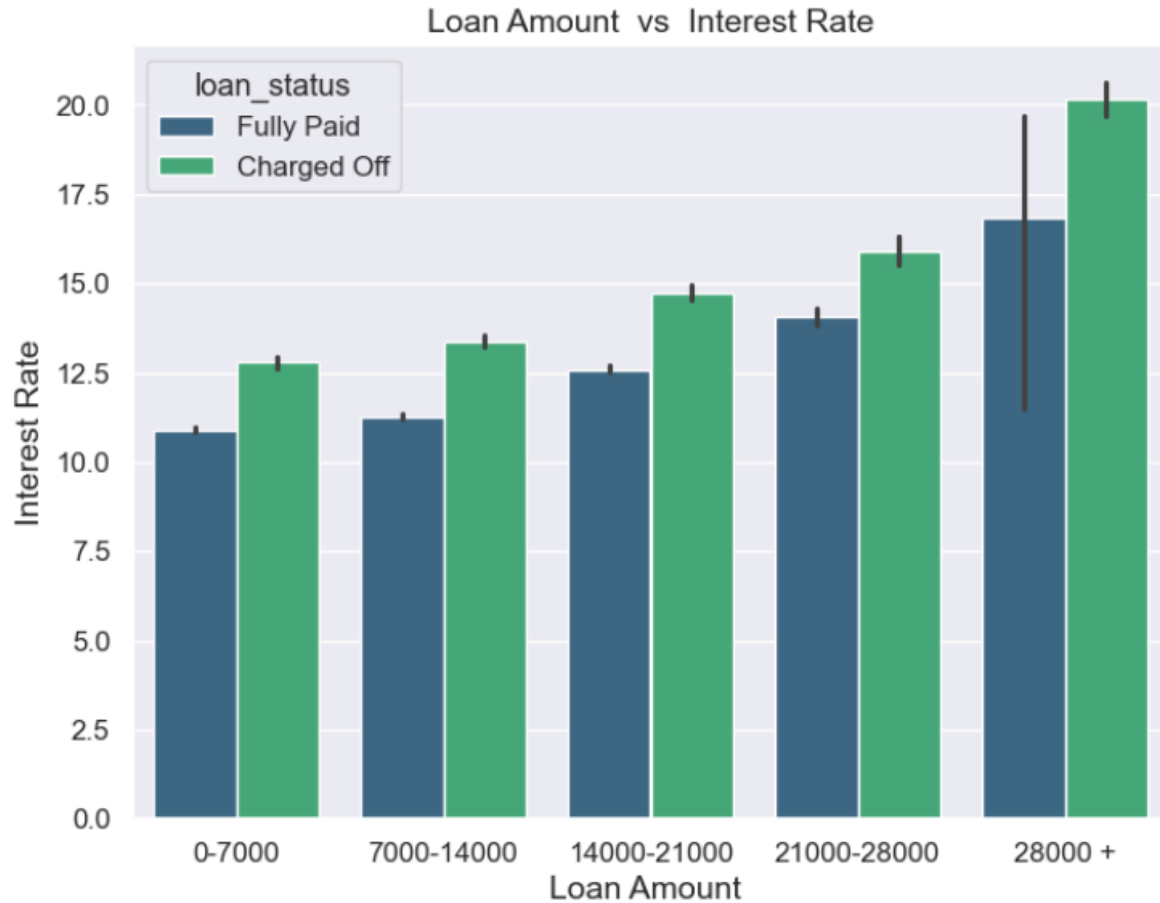
Year vs Charged Off



Observations:

- The year 2007 has the highest Charged off proportion.
- The year 2009 is having lowest Charged off proportion.
- The years 2008 & 2011 have almost equivalent Charged Off proportions.

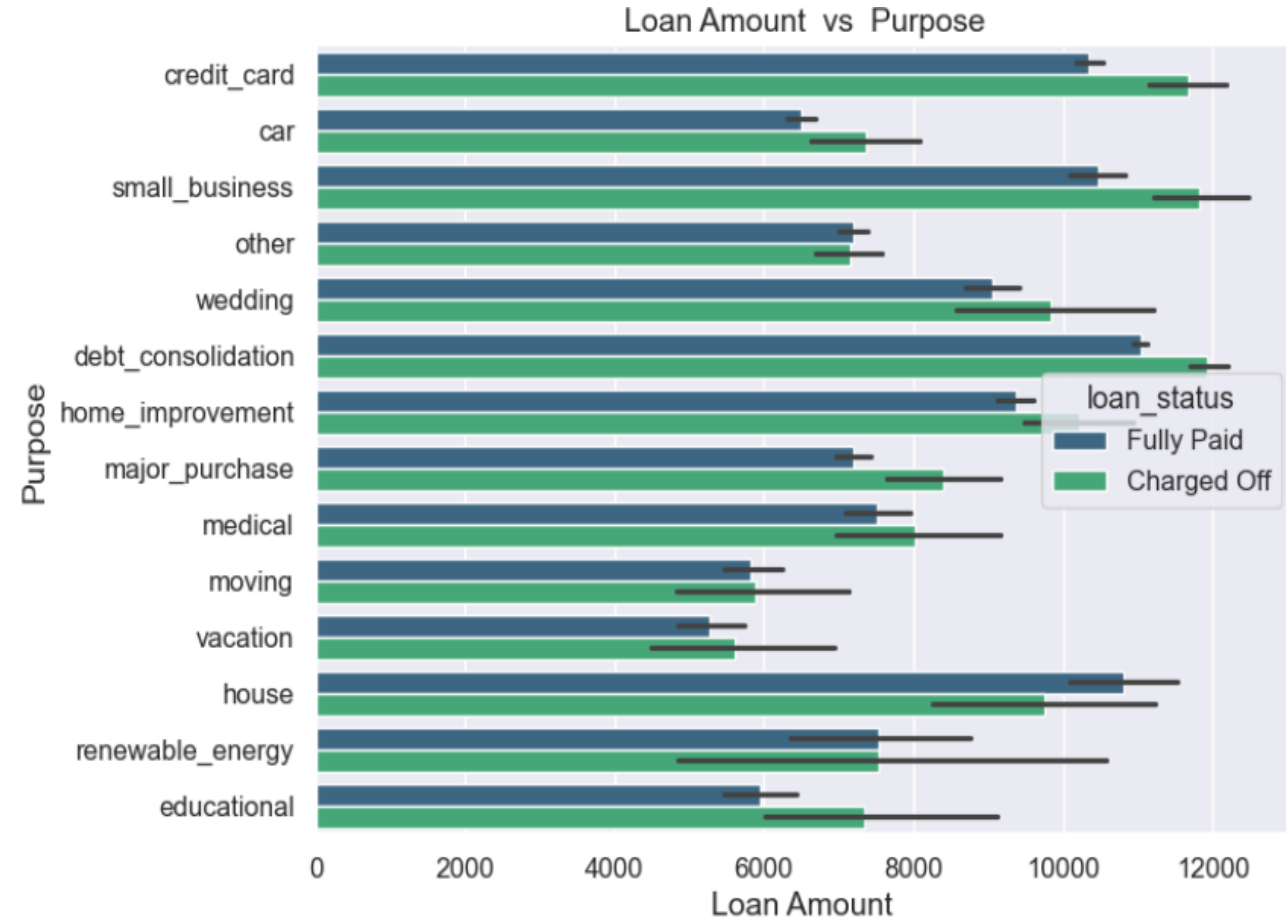
Loan Amount vs Interest Rate



Observations:

- Charged off proportion is highest for amount above 28k+ and have a higher interest rate
- Higher the interest rate & loan amount, higher the Charged off count.

Loan Amount vs Purpose



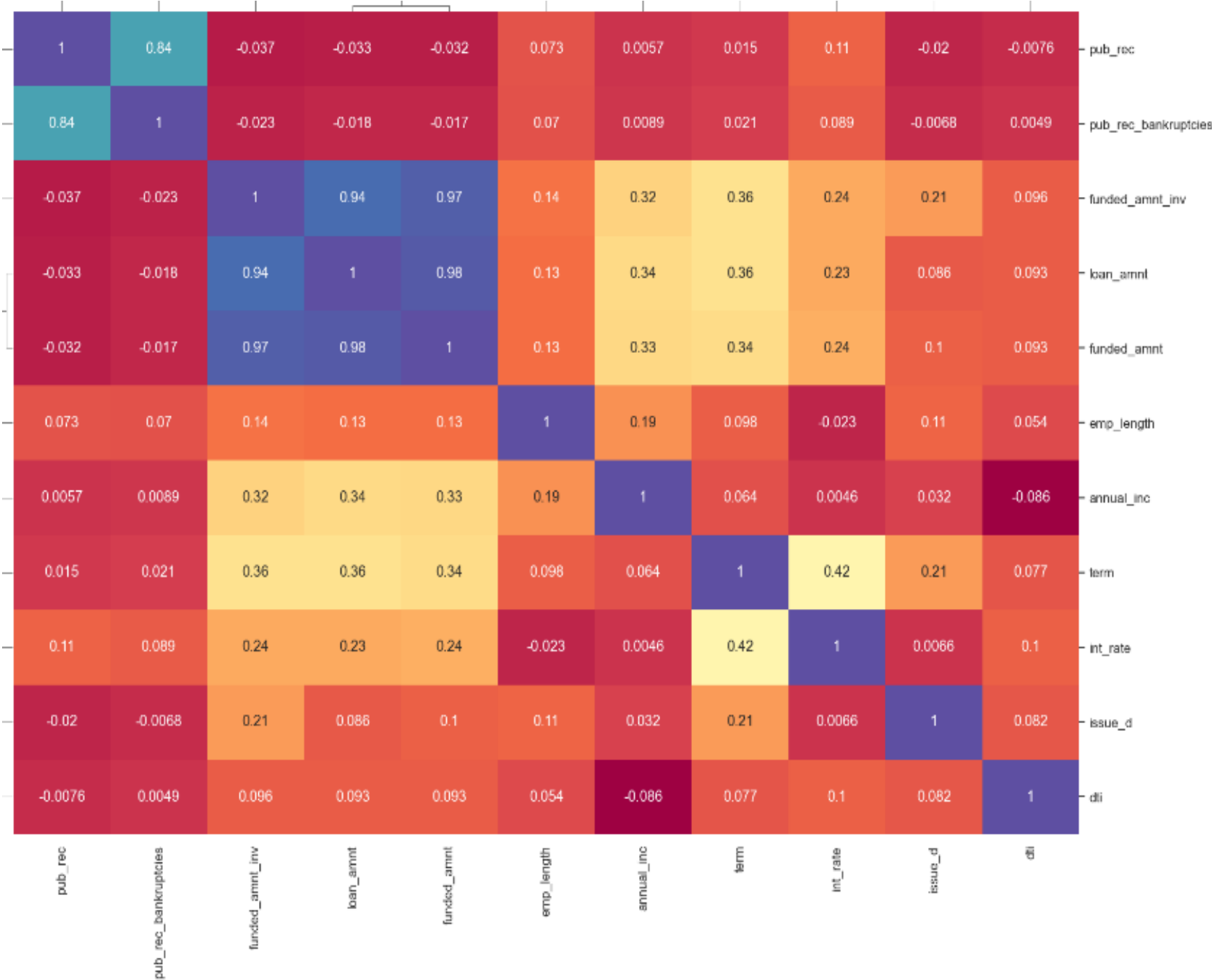
Observations:

- Highest charged off are from the debt consolidation column.
- Credit card and small business are a close second in being charged off.

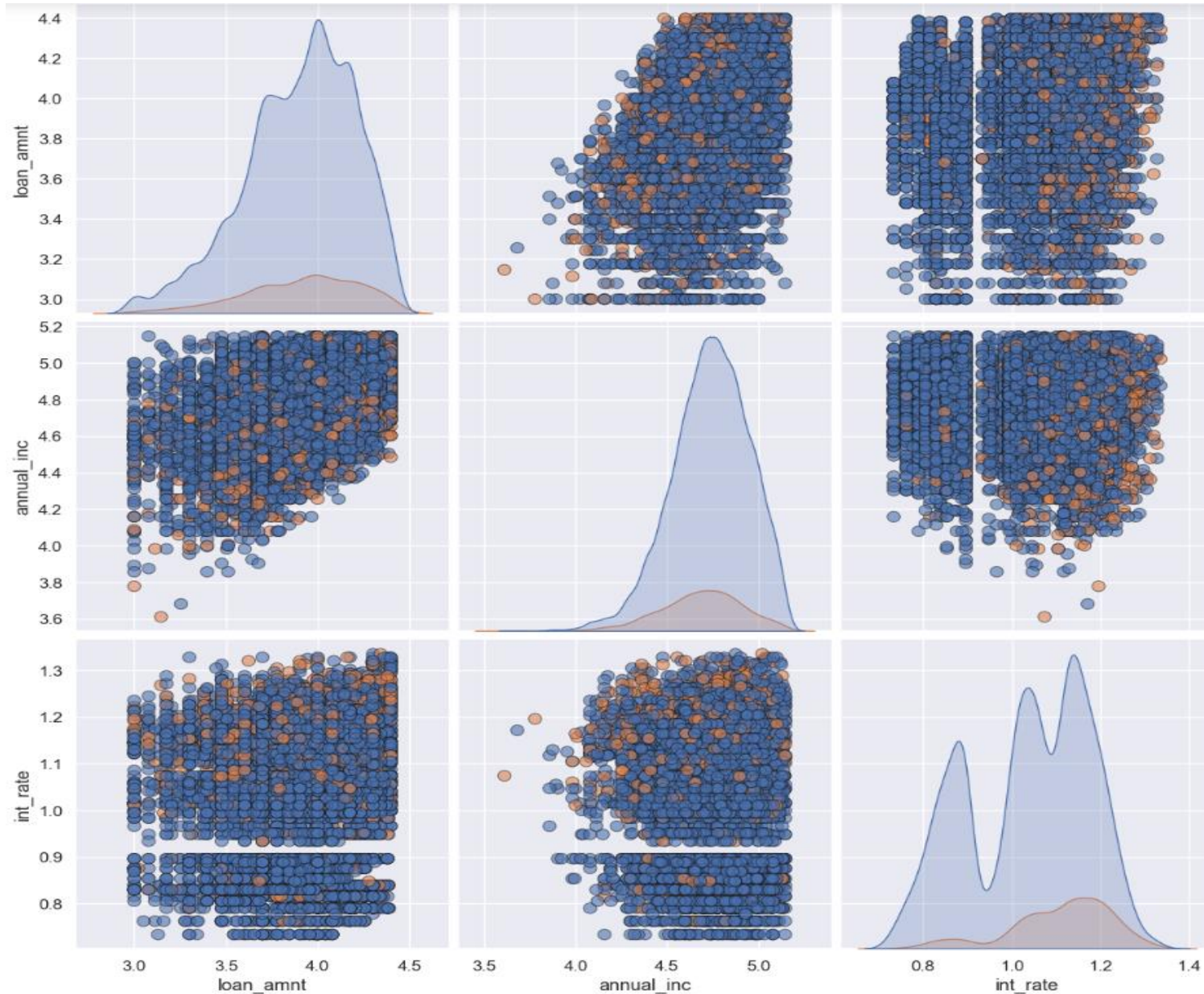
Correlation Matrix

Observations:

- Loan amount, Investor Funds, Funded amount are Positively correlated.
- Term has a Positive correlation with Interest Rate
- # There is a positive correlation between annual income and employment years.
- Annual income is negatively correlated with DTI
- Loan amount, Investor Funds, Funded amount have a Negative correlation with Pubic Record Bankruptcy
- Employment length is negatively correlated to Interest Rate



Multi Variate of Loan Amount, Annual Income, Year and Interest Rate :



Observations:

- Higher the interest rate , Higher charged off ratio
- Higher the annual income, Higher the loan amount slightly.
- Increase in number of charged off with increase in year.
- Interest rate is increasing with loan amount increase

Summary:

Based on the EDA done on the given Data set, we can conclude the following:

- The likelihood of being charged off rises as we progress from Grade “A” to “G”.
- The average interest rate for loans with a 60-month term is significantly higher than that for loans with a 36-month term.
- Charged off proportion increases as Interest Rate Increases.
- Individuals with negative public records are at a higher risk of being charged off than those without such records.
- Those applicants taking a loan to clear pre-existing Debt or are Small business owners, have a high probability of being defaulters.
- When the DTI is high(between 12+) the charged off risk increases.
- There is a significant rise in the Charged off count between the year 2010 and 2011. We also see that there is a consistent rise in the loan amount from the year 2008 onwards.
- Average interest rate is considerably higher for 60 months loan term than 36 months.
- Loan Defaulter are high when their Annual income ranges from 0 – 20k. As the annual income increases, the charged off proportion decreases.
- Loan applicants who have applied for a loan amount of more than 28k tend to have a high interest rate, which in turn can result in them becoming Charged off.
- Loans approved at certain months of the year seem to have a higher charged off rate than the rest of the months.
- Applicants who Rent a home or are paying Mortgage have a high probability of being loan defaulters.