

Lending Club Case Study

Shyam Sundar Kothandaraman
Pooja Kumari

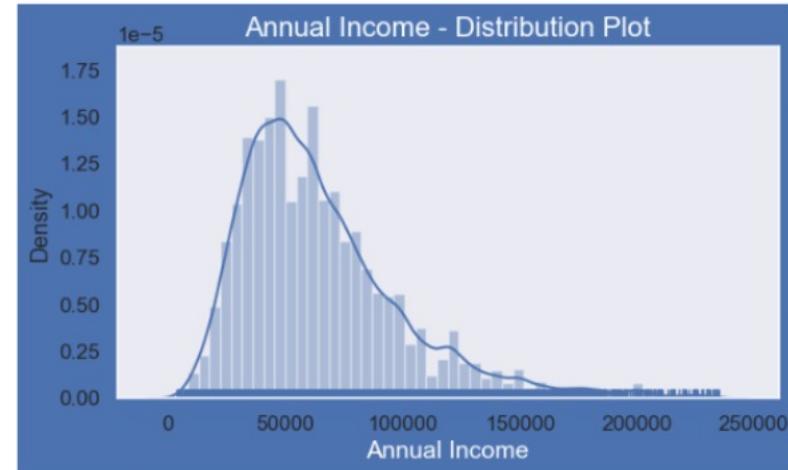
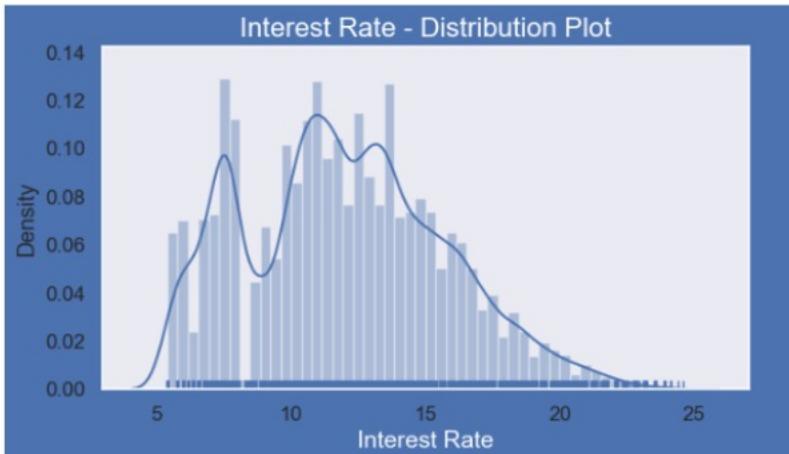
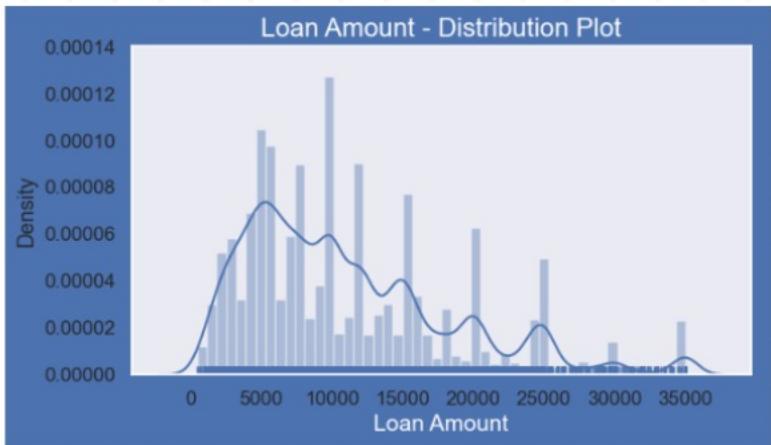
Business Objective:

- ❖ Identify and evaluate robust indicators of loan default.
- ❖ Leverage data analysis to quantify the influence of these indicators.
- ❖ Develop and employ advanced risk assessment models.
- ❖ Continuously adapt models for dynamic economic conditions and borrower behavior.

Data Cleaning and Reduction

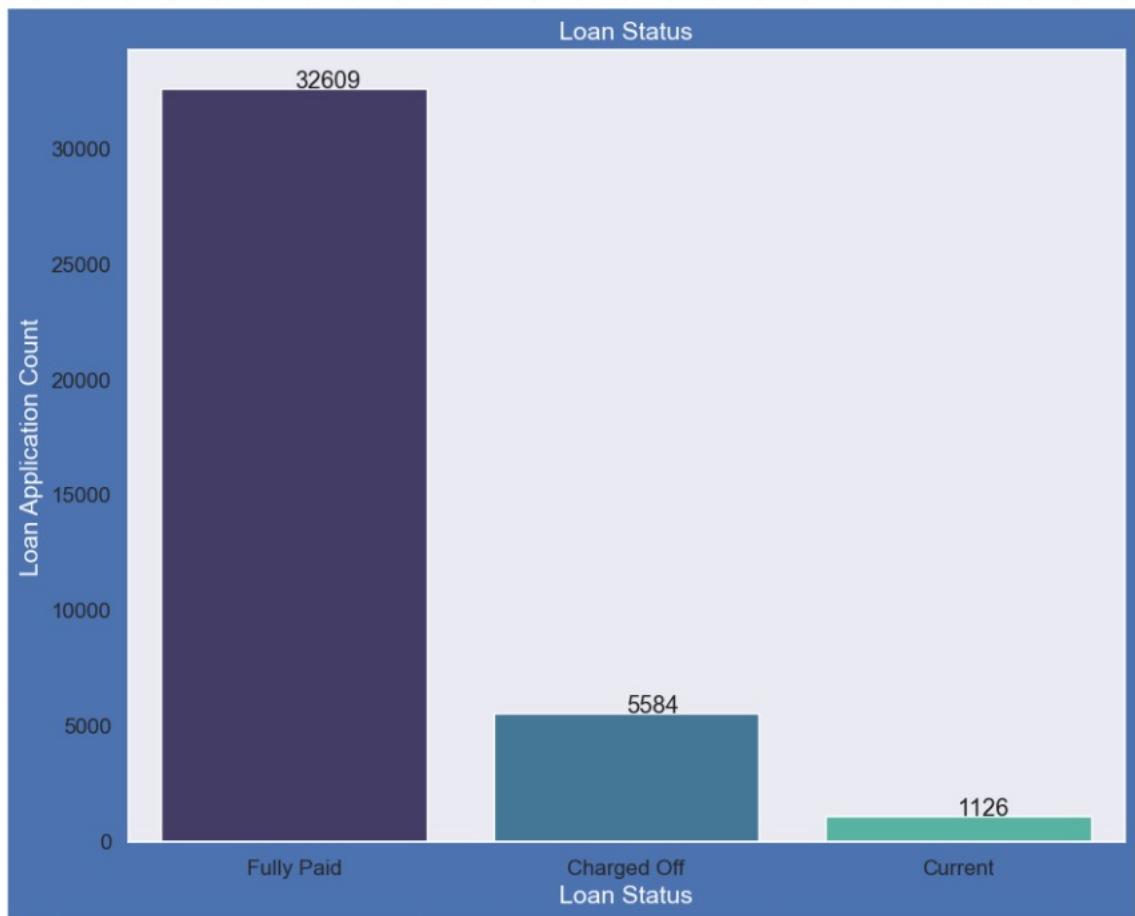
- ❖ **Original Data:** 39,717 rows and 111 columns.
- ❖ **Data Reduction:**
 - ❖ Dropped 54 Empty Columns.
 - ❖ Eliminated 9 columns with a single value throughout.
 - ❖ Removed 3 columns with unique values for all rows.
 - ❖ Excluded 2 fields with over 90% missing values.
 - ❖ Discarded 4 fields where most values were 0 or outliers.
 - ❖ Eliminated 14 unnecessary fields (e.g., "title," "zip_code").
- ❖ **Final Data Set:** Streamlined to 25 relevant columns for analysis.

Univariate Analysis



- ❖ The majority of loan amounts in the dataset are concentrated in the range of 5000 to 15000.
- ❖ The plots reveal that the predominant interest rates on loans fall within the 10% to 15% range.
- ❖ The plots demonstrate that the majority of borrowers have annual incomes within the 40,000 to 80,000 range.

Loan Status:



The single plot illustrates that approximately 14% of the total loans issued have been charged off.

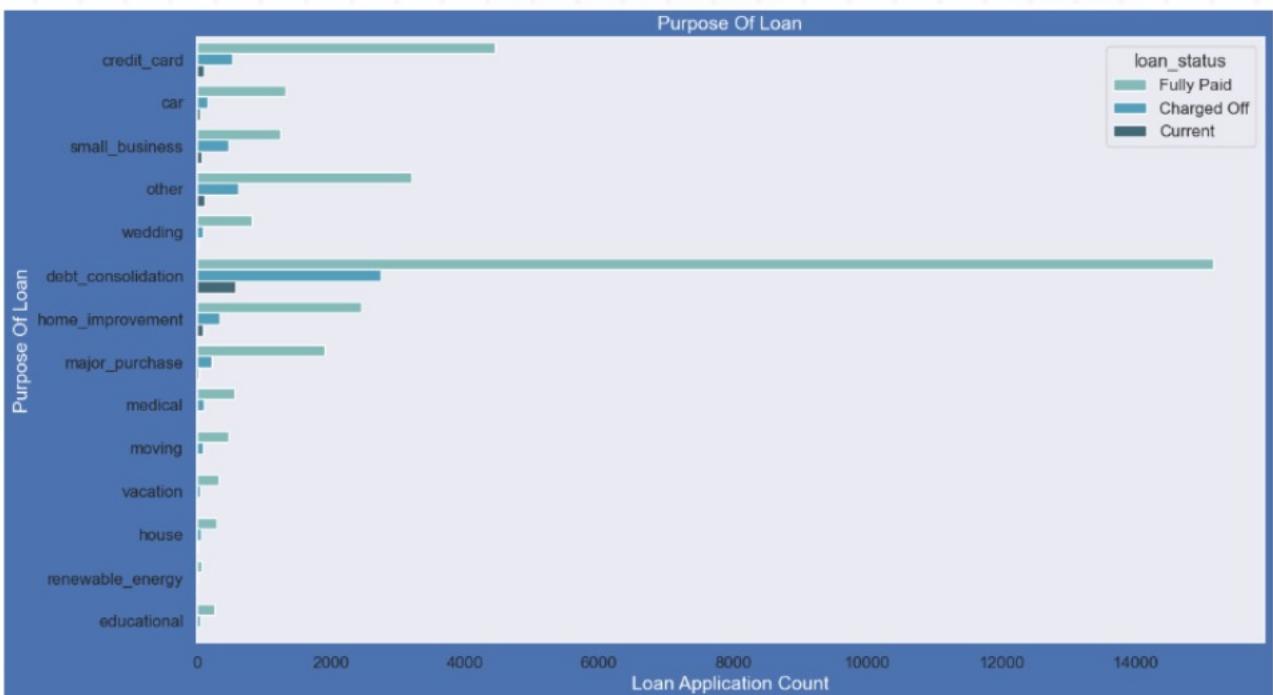
Purpose Of Loan:

Common Loan Purpose and Default Rate:

- ❖ Debt consolidation is the most prevalent loan purpose, accounting for 47% of loans, with a default rate of 14.8%.
- ❖ Loans for small businesses have the highest default rate at 26%.

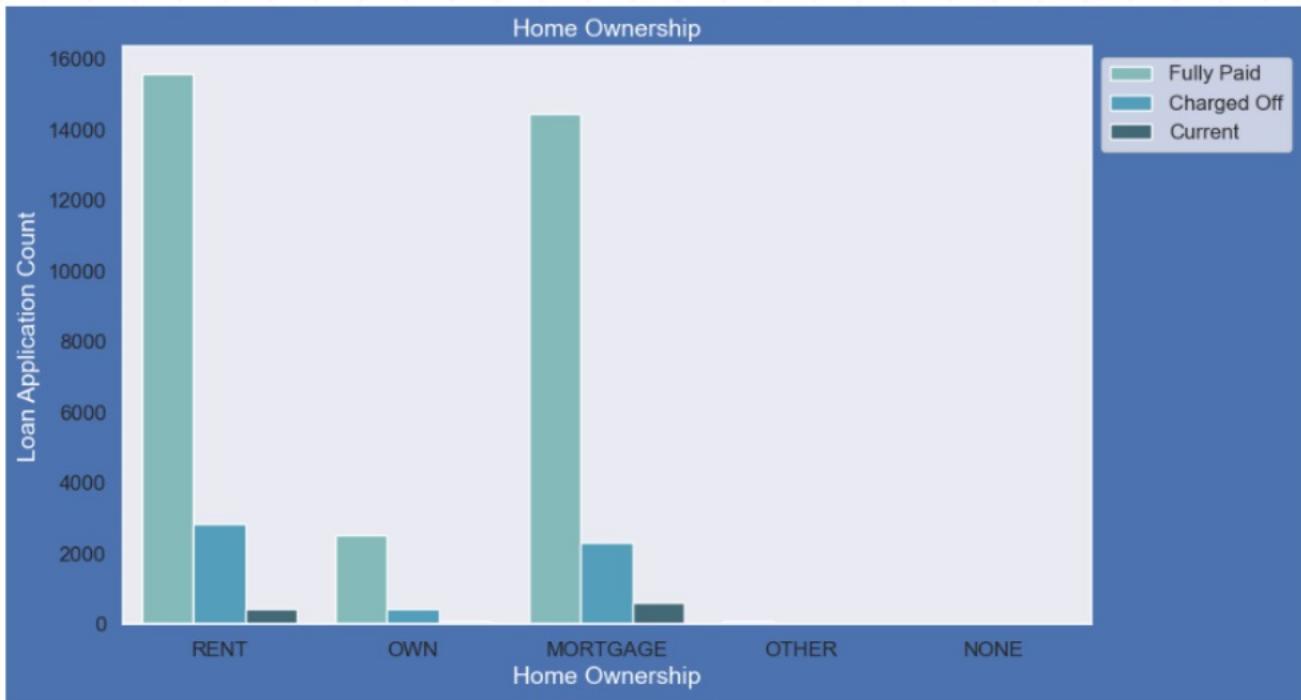
Low Default Rate for Specific Purposes:

- ❖ Loans for wedding, car, or major purchases have notably low default rates, with only a 10% chance of default.



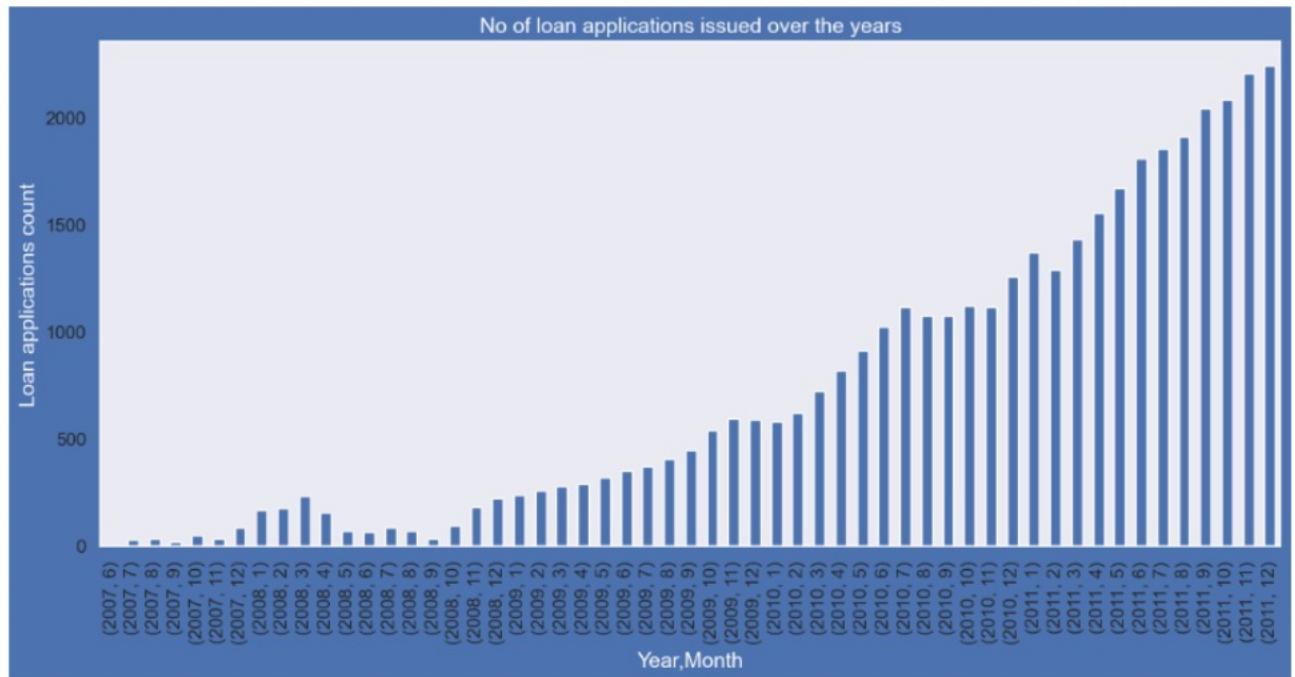
Home Ownership:

- ❖ The majority of applicants live in rented homes or have mortgaged their homes.
- ❖ Due to the high number of applicants from these categories, charged-off loans are also relatively high.
- ❖ This format effectively conveys the relationship between housing status and charged-off loans.



Ordered Categorical Variables - Number of loans which were approved every year/month:

- ❖ Year-over-year, there is a consistent rise in the number of loan applications.
- ❖ The increasing number of loan applications correlates with a higher count of charged-off applications.
- ❖ Notably, in 2008, there was a significant drop in the number of loans issued, particularly from May to October, likely attributed to the economic recession.



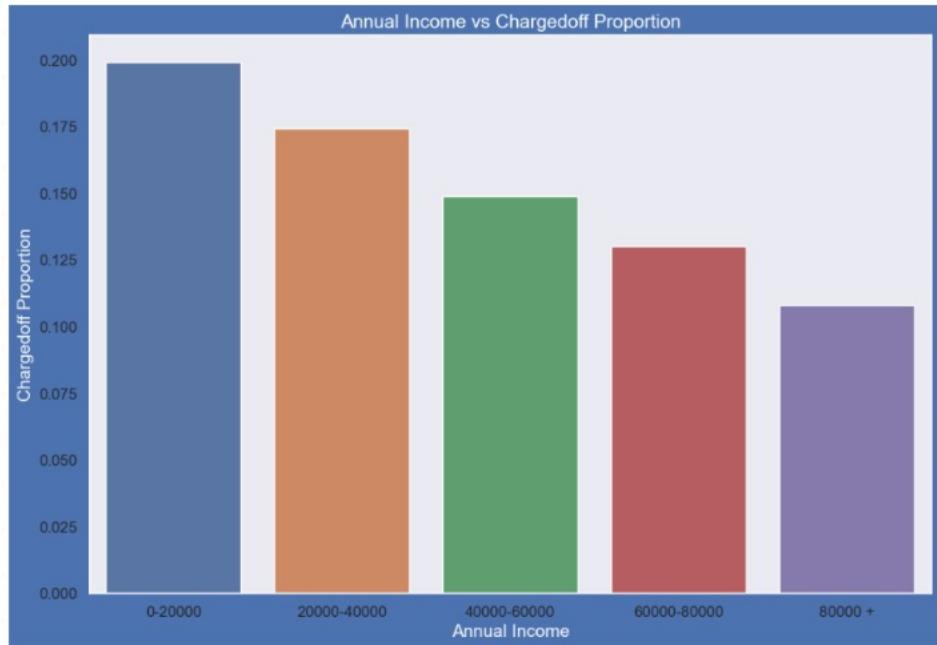
Ordered Categorical Variables- Loan Paying Term:



- ❖ Applicants who opted for a 60-month loan repayment term had a higher percentage of charged-off applications compared to those with a 36-month loan term.

Bivariate Analysis

Annual income vs Chargedoff_Proportion :



nnual_inc_cats	Charged Off	Current	Fully Paid	Total	Chargedoff_Pro
0-20000	237	9	943	1189	
20000-40000	1514	170	7004	8688	
40000-60000	1729	345	9534	11608	
60000-80000	1024	240	6597	7861	
80000 +	1080	362	8531	9973	

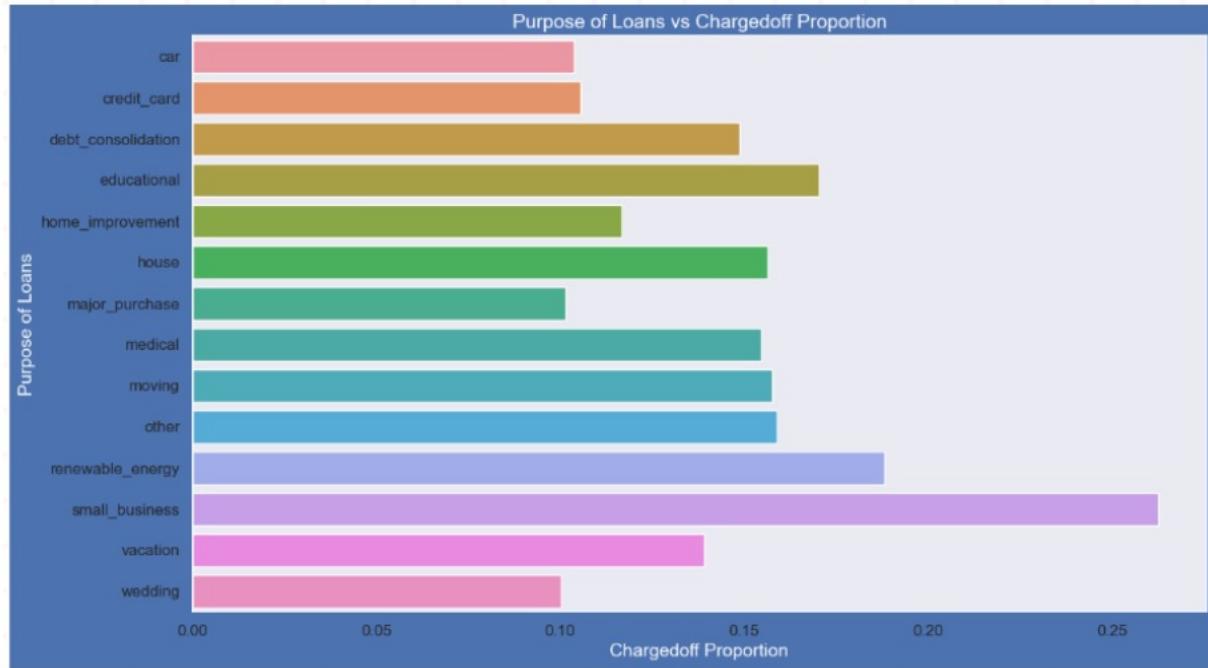
Income Range and Charged-Off Chances:

- Applicants with an income range of 0-20,000 have a higher likelihood of charged-off loans.
- Conversely, those in the 80,000+ income range are less likely to experience charge-offs.

Income and Charge-Off Relationship:

- Notably, as annual income increases, the proportion of charged-off loans decreases, indicating a strong correlation between income level and loan repayment.

Purpose of Loan vs Chargedoff_Proportion:



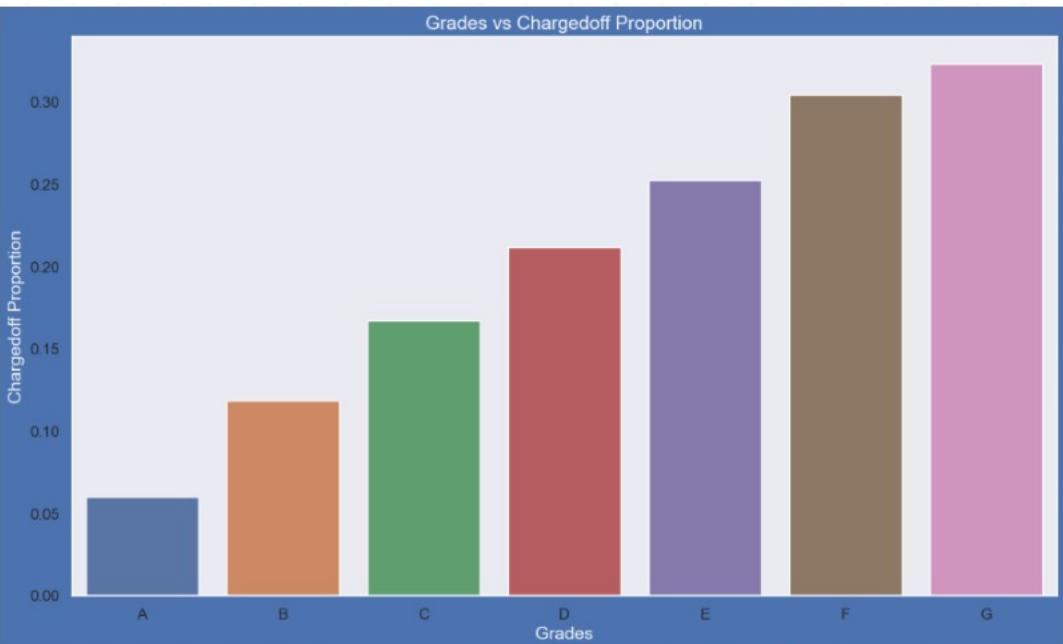
loan_status	purpose	Charged Off	Current	Fully Paid	Total	Chargedoff_Proportion
11	small_business	472.00	74.00	1251.00	1797.00	0.26
10	renewable_energy	19.00	1.00	81.00	101.00	0.19
3	educational	55.00	0.00	268.00	323.00	0.17
9	other	629.00	126.00	3203.00	3958.00	0.16
8	moving	91.00	7.00	479.00	577.00	0.16
5	house	58.00	14.00	299.00	371.00	0.16
7	medical	106.00	12.00	568.00	686.00	0.15
2	debt_consolidation	2752.00	580.00	15165.00	18497.00	0.15
12	vacation	53.00	6.00	322.00	381.00	0.14
4	home_improvement	338.00	97.00	2459.00	2894.00	0.12
1	credit_card	537.00	102.00	4450.00	5089.00	0.11
0	car	160.00	50.00	1332.00	1542.00	0.10
6	major_purchase	220.00	37.00	1909.00	2166.00	0.10
13	wedding	94.00	20.00	823.00	937.00	0.10

- ❖ Small business applicants have a notably higher risk of experiencing loan charge-offs.
- ❖ In contrast, the "renewable_energy" category exhibits a lower proportion of charged-off loans compared to other loan purposes.
- ❖ This format effectively highlights the charge-off risk differences between small business applicants and the renewable energy category.

Grade vs Chargedoff_Proportion:

- Grade "A" indicates low chances of loan default and reflects strong creditworthiness.
- Grades "F" and "G" are associated with high chances of loan default and indicate substantial default risk.
- Grade Trend demonstrates a clear pattern where the likelihood of default increases from "A" to "G" grades.
- Grade Trend highlights the reliability of loan grades as predictors of default risk.

loan_status	grade	Charged Off	Current	Fully Paid	Total	Chargedoff_Proportion
6	G	99	17	190	306	0.32
5	F	312	71	642	1025	0.30
4	E	704	176	1906	2786	0.25
3	D	1112	216	3916	5244	0.21
2	C	1342	262	6411	8015	0.17
1	B	1414	344	10167	11925	0.12
0	A	601	40	9377	10018	0.06



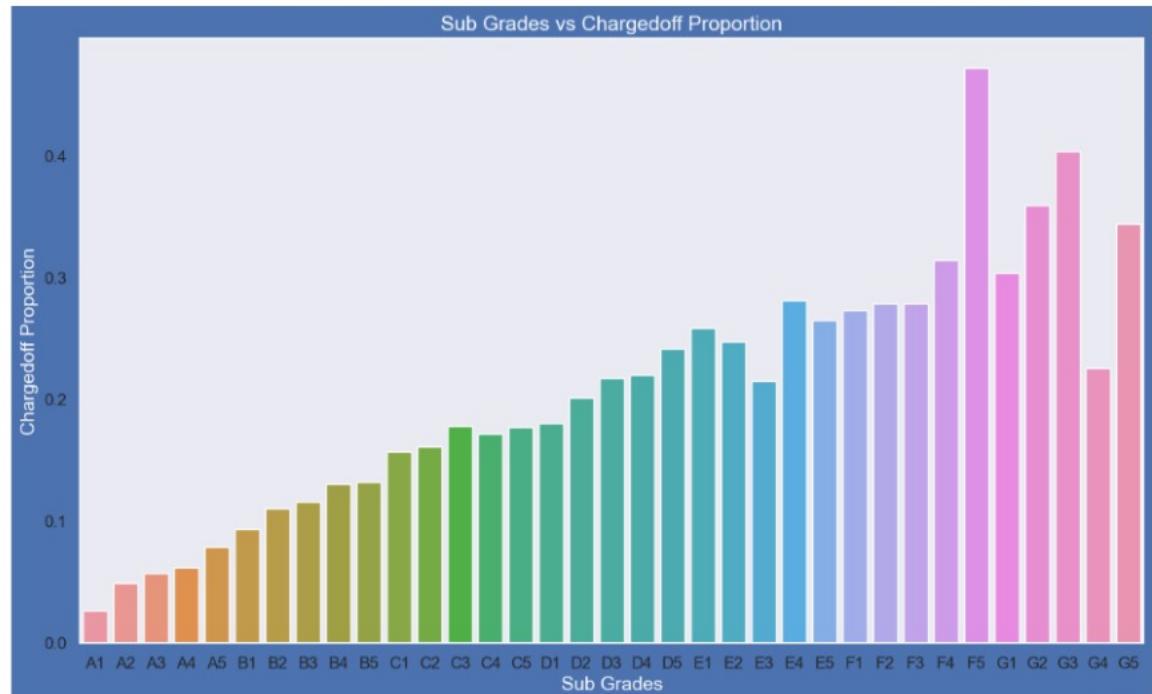
Sub-grade vs Chargedoff_Proportion:

Sub-grade within "A":

- ❖ Exhibit very low chances of loan charge-offs.
- ❖ Indicate strong creditworthiness within the "A" grade.

Sub-grade within "F" and "G":

- ❖ Are associated with significantly higher chances of loan charge-offs.
- ❖ Show a proportionally increasing risk of charge-offs as sub-grade move from "A" to "G" within these categories.
- ❖ This format effectively conveys the differences in charge-off risk among sub-grades within the "A," "F," and "G" categories.



Interest rate vs Chargedoff_Proportion:

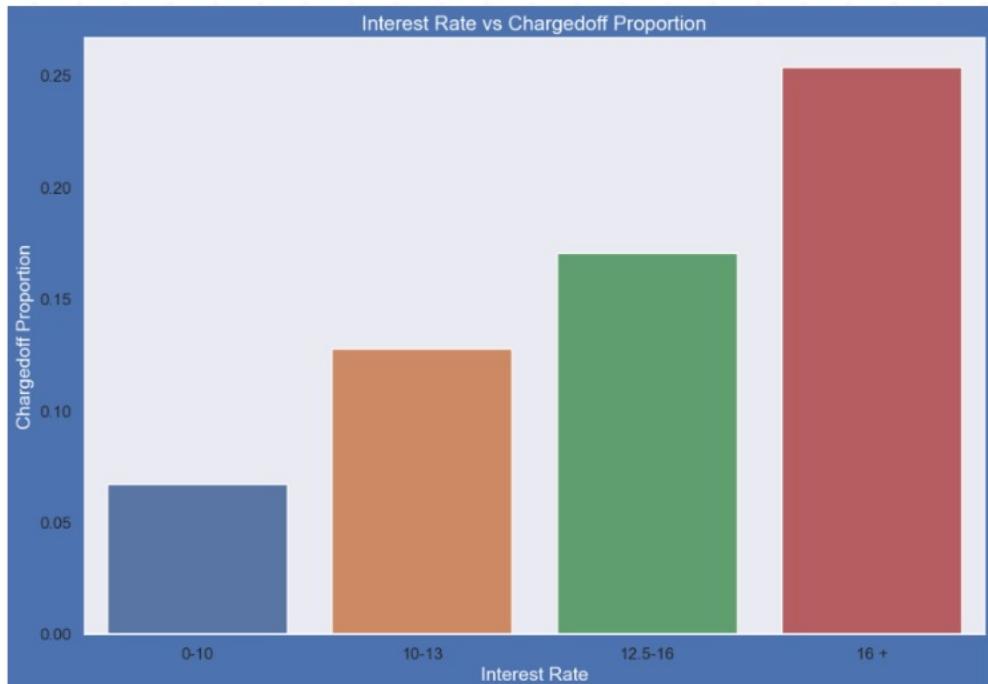
Interest Rate Less Than 10%:

- Loans with interest rates below 10% are associated with very low chances of being charged off.
- Interest rates start from a minimum of 5%.

Interest Rate More Than 16%:

- Loans with interest rates exceeding 16% have a higher likelihood of being charged off compared to other interest rate categories.
- Charged-off proportions increase with higher interest rates.

loan_status	int_rate_cats	Charged Off	Current	Fully Paid	Total	Chargedoff_Proportion
3	16 +	1250	351	3317	4918	0.25
2	12.5-16	1995	329	9354	11678	0.17
1	10-13	1224	269	8083	9576	0.13
0	0-10	825	77	11403	12305	0.07



Employment length vs Chargedoff_Proportion:

Less Than 1 Year of Work Experience or Not Employed:

- ❖ Individuals in this category have a higher likelihood of loan defaults.
- ❖ This is logical as they may lack a stable source of income to repay the loan.

Other Applicants:

- ❖ Applicants with more work experience or stable employment status have similar chances of loan default.
- ❖ Their employment status may not be as significant a factor in loan default rates.
- ❖ These points effectively convey the impact of employment history and income stability on loan default risk.



State vs Chargedoff_Proportion:

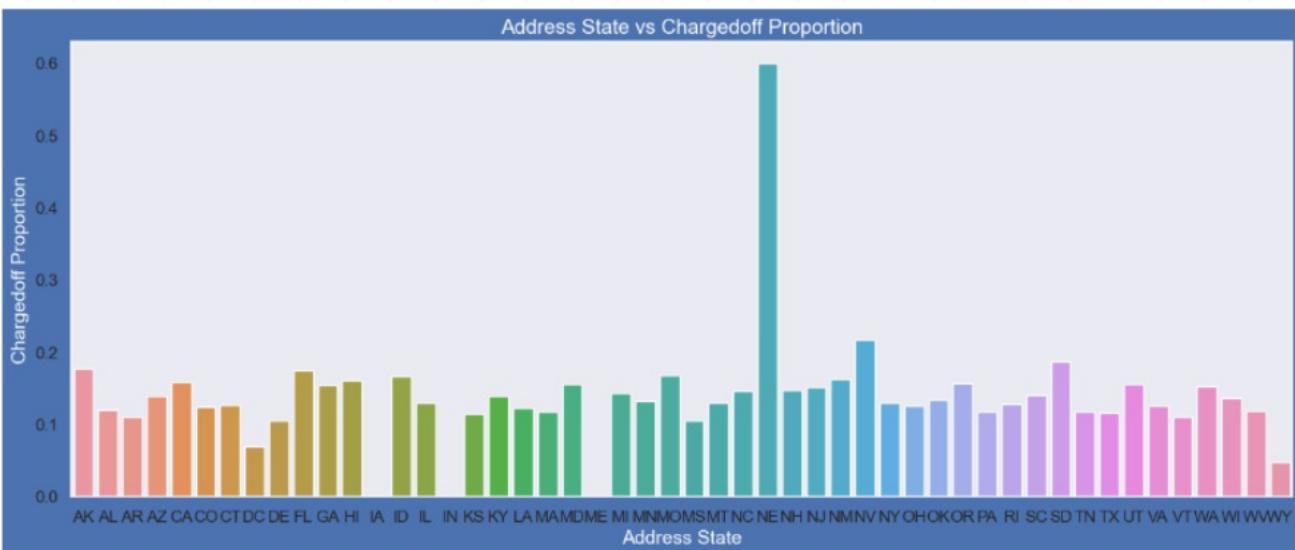
State NE (Nebraska):

- ❖ Nebraska indicates a high likelihood of loan defaults, but the sample size is limited for definitive conclusions.

States NV (Nevada), CA (California), and FL (Florida):

- ❖ Nevada, California, and Florida exhibit a significant number of loan defaults, with a substantial impact on loan default rates.

This format effectively communicates the regional variations in loan default rates and acknowledges the potential impact of sample size.



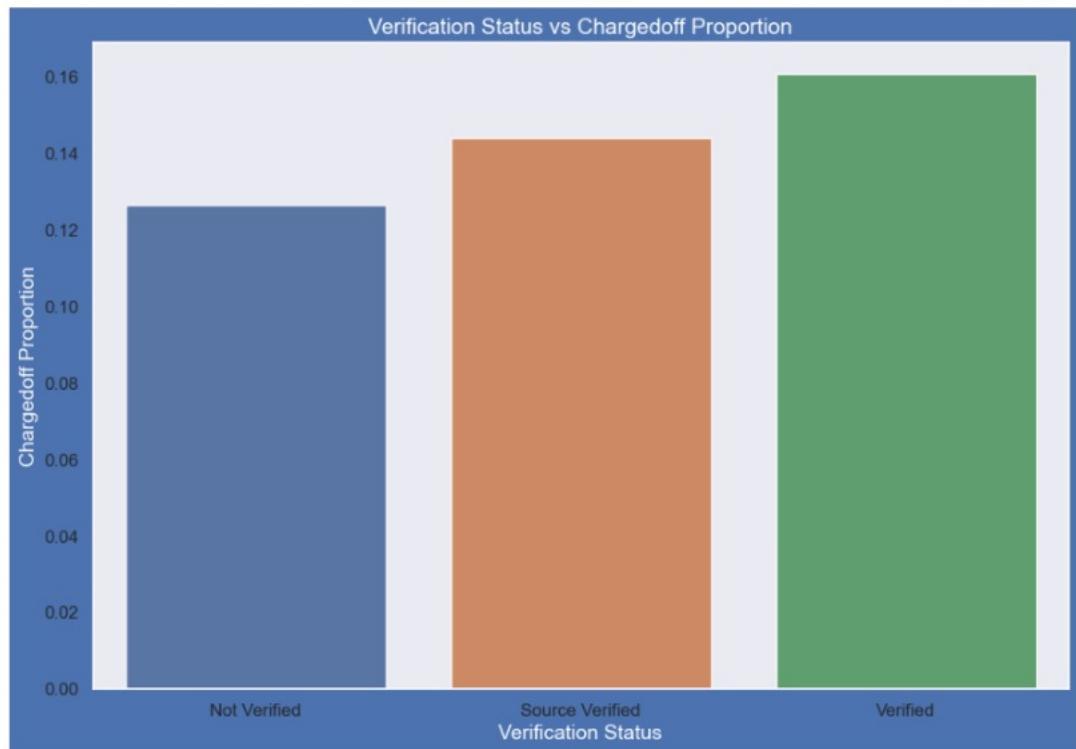
Verification_Status vs Chargedoff_Proportion:

Income Verification Risk:

- ❖ Approximately 42% of borrowers' income is 'not verified,' which poses a higher risk.
- ❖ Loans without income verification have a 12% chance of default.

Funding Comparison:

- ❖ The total amount funded by the company to non-verified loans is lower when compared to verified loans.



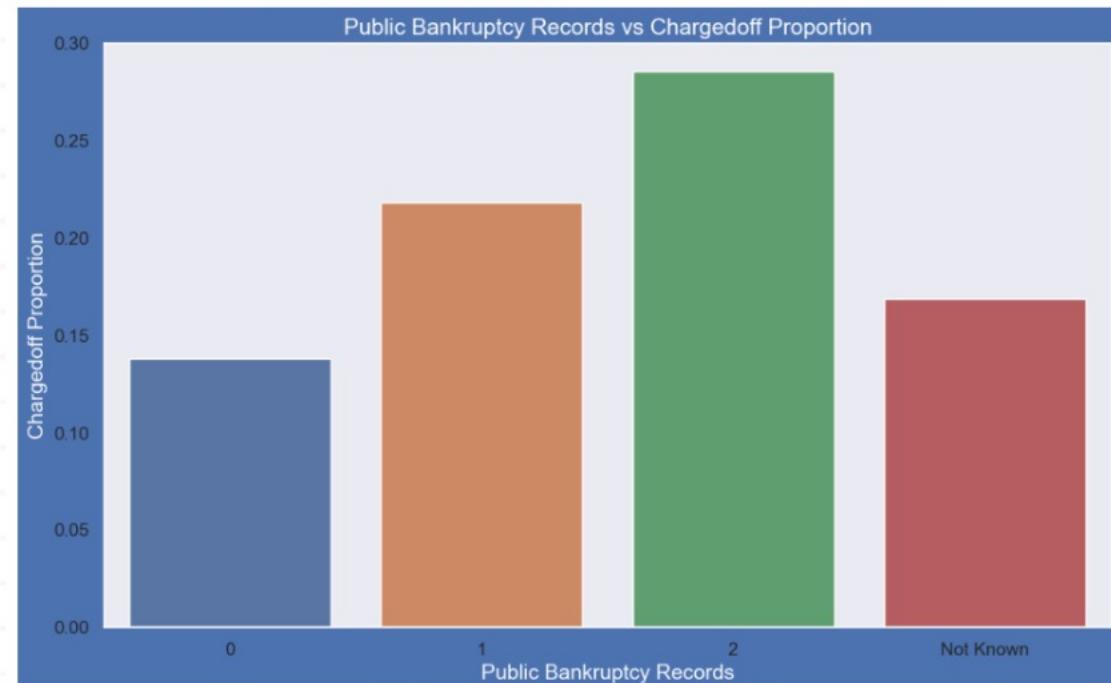
Pub_rec_bankruptcies against Chargedoff_Proportion:

Pub_rec_bankruptcies:

- ✿ Borrowers with a pub_rec_bankruptcies value of 1 have a higher charged-off proportion compared to those with no pub_rec_bankruptcies.

Recurrent Default Risk:

- ✿ Borrowers who have defaulted in the past are at a higher risk of defaulting in the future, reflecting the importance of credit history in assessing default risk.



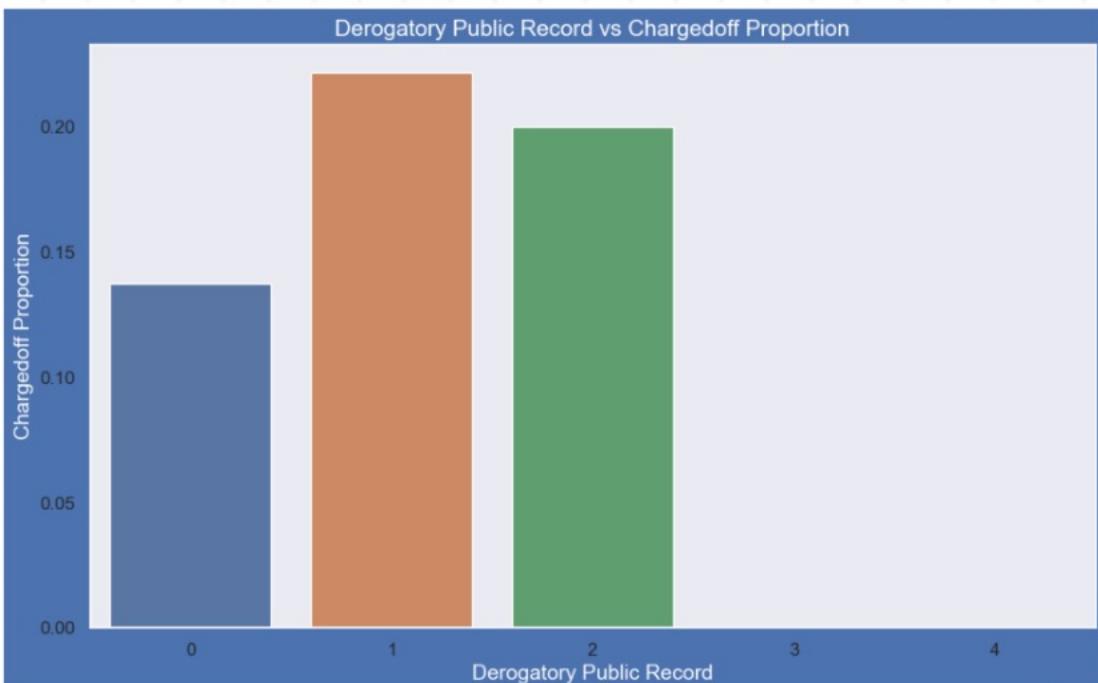
Derogatory Public Record vs Chargedoff_Proportion:

Derogatory Items Definition:

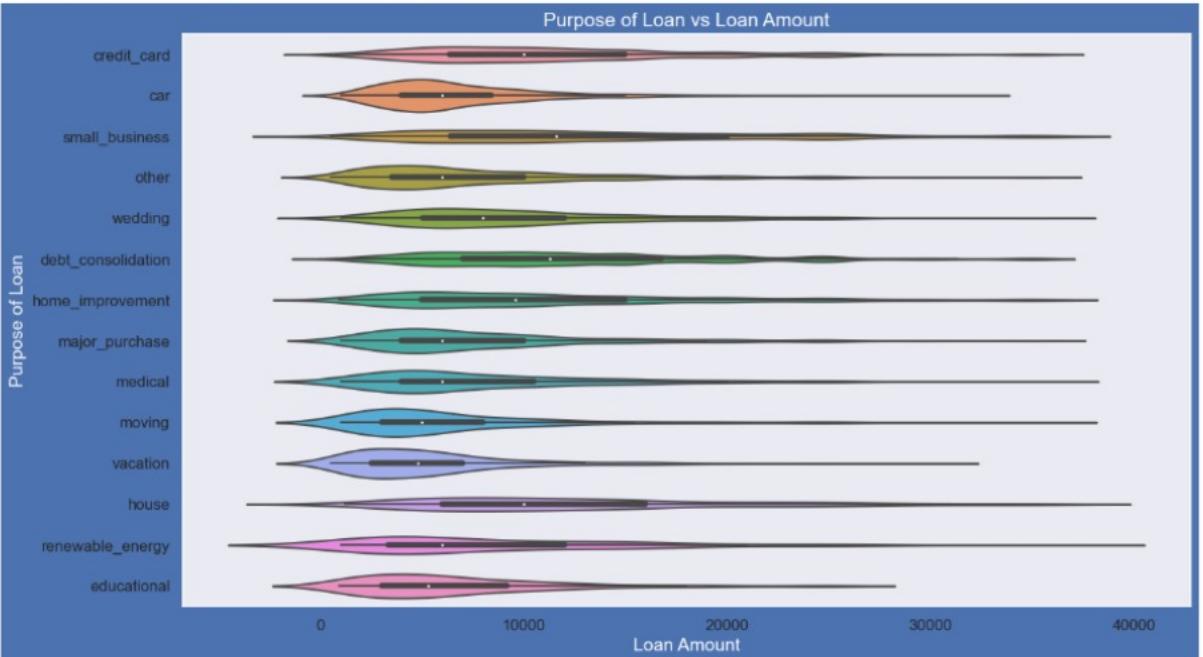
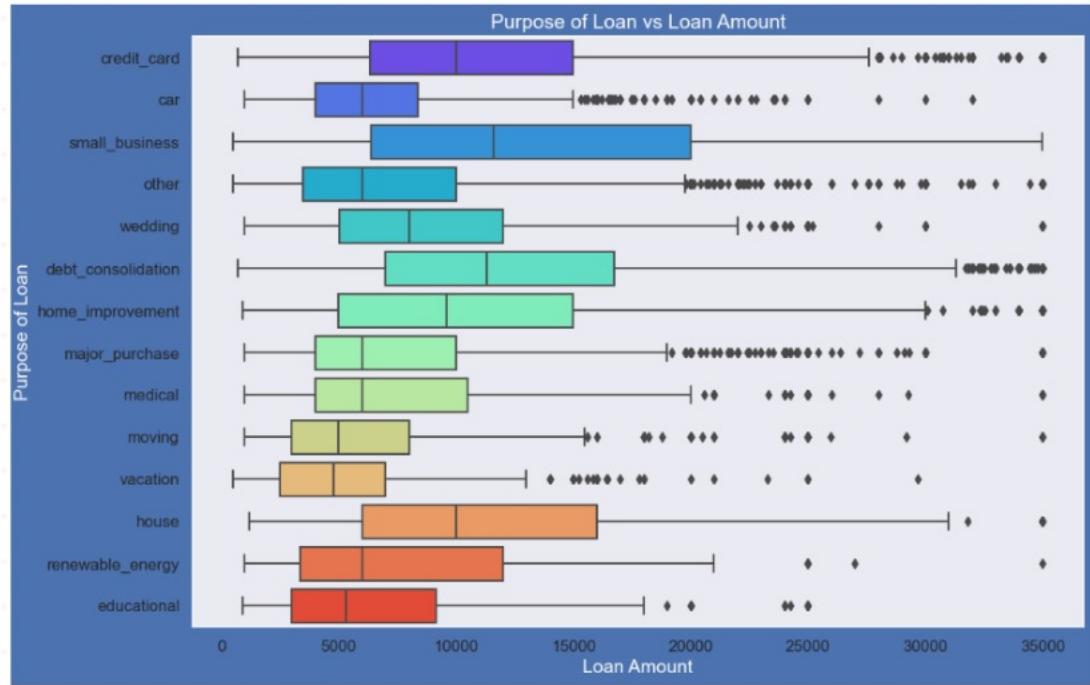
- ❖ Derogatory items are entries that lenders consider negative, indicating financial risk and impacting credit qualifications. Public records and collections fall into this category as they reflect unpaid financial obligations.

Impact of Pub_rec on Charged-Off Loans:

- ❖ Borrowers with a pub_rec value of 1 or 2, indicating a Derogatory Public Record, have a higher likelihood of loan charge-offs compared to those with no derogatory record.
- ❖ For pub_rec counts of 3-4, where the numbers are limited, it is not possible to draw definitive conclusions regarding charged-off chances.

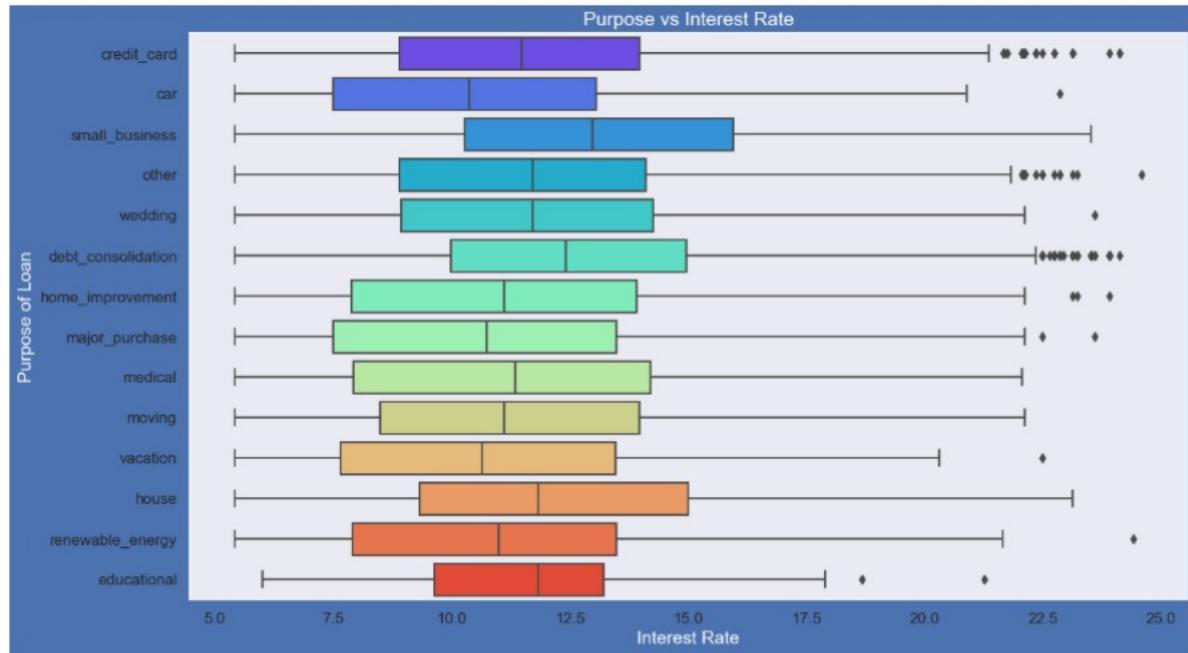


Purpose of loan vs Loan amount:

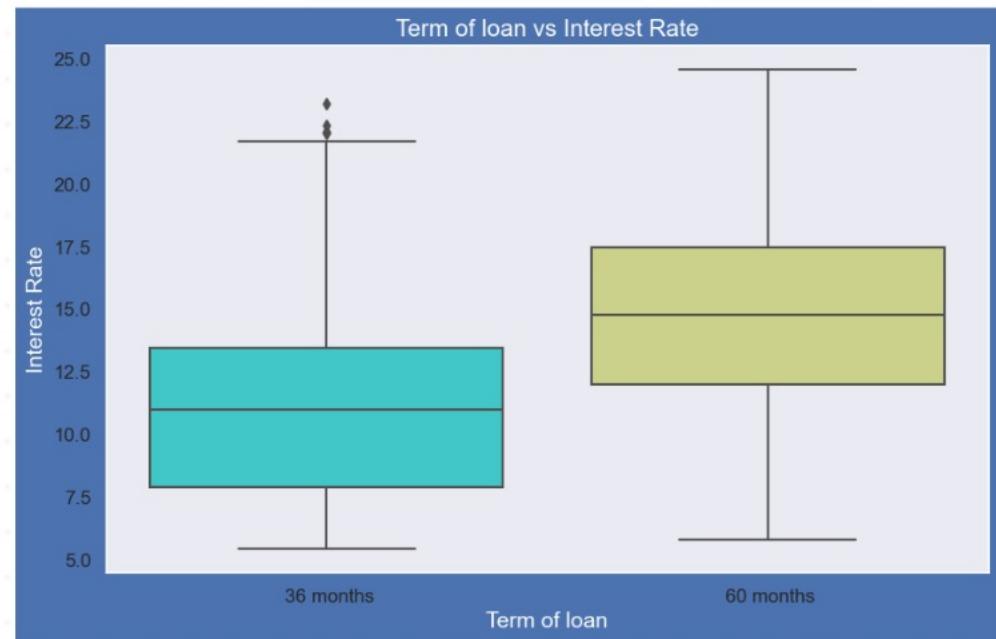


- ❖ The median, 95th percentile, and 75th percentile of loan amounts are highest for loans taken for small business purposes, followed by debt consolidation and credit card loans.
- ❖ Small business loans exhibit the highest loan amount distribution, with debt consolidation in the second position and credit card loans as the third-highest.

Purpose vs Interest Rate:



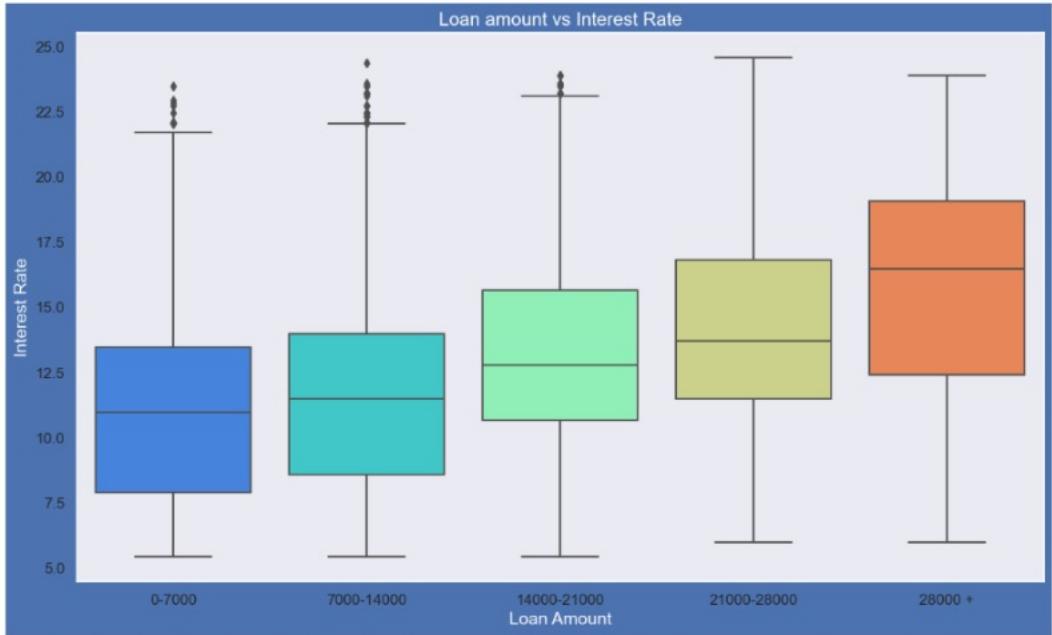
Term of loan vs Interest Rate:



- ❖ Small business loans have the highest average interest rates, indicating that borrowers taking loans for small business purposes typically face higher interest costs.
- ❖ Debt consolidation ranks second in terms of higher average interest rates, suggesting that borrowers using loans for consolidating debt also face relatively higher interest expenses.

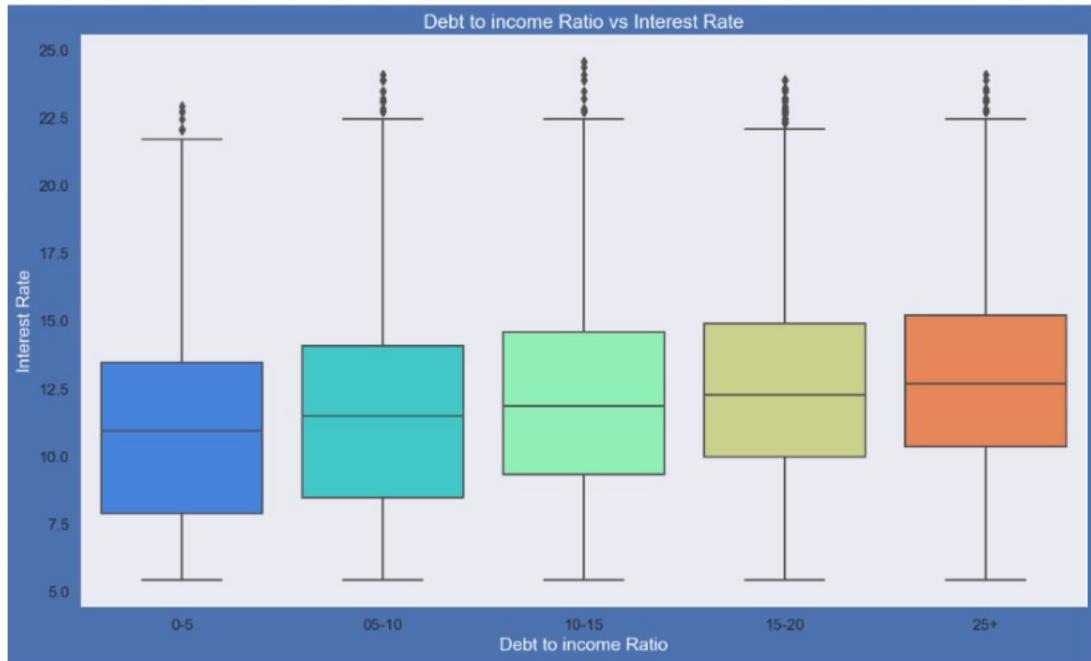
- ❖ The average interest rate is notably higher for 60-month loan terms, indicating that most loans issued for longer terms come with higher interest rates for repayment.

Loan Amount vs Interest Rate:



- ❖ It is clear that interest rate is increasing with loan amount increase.
- ❖ probably when loan amount is more it is taken for longer loan term, we saw earlier that longer the loan term more the interest rate.

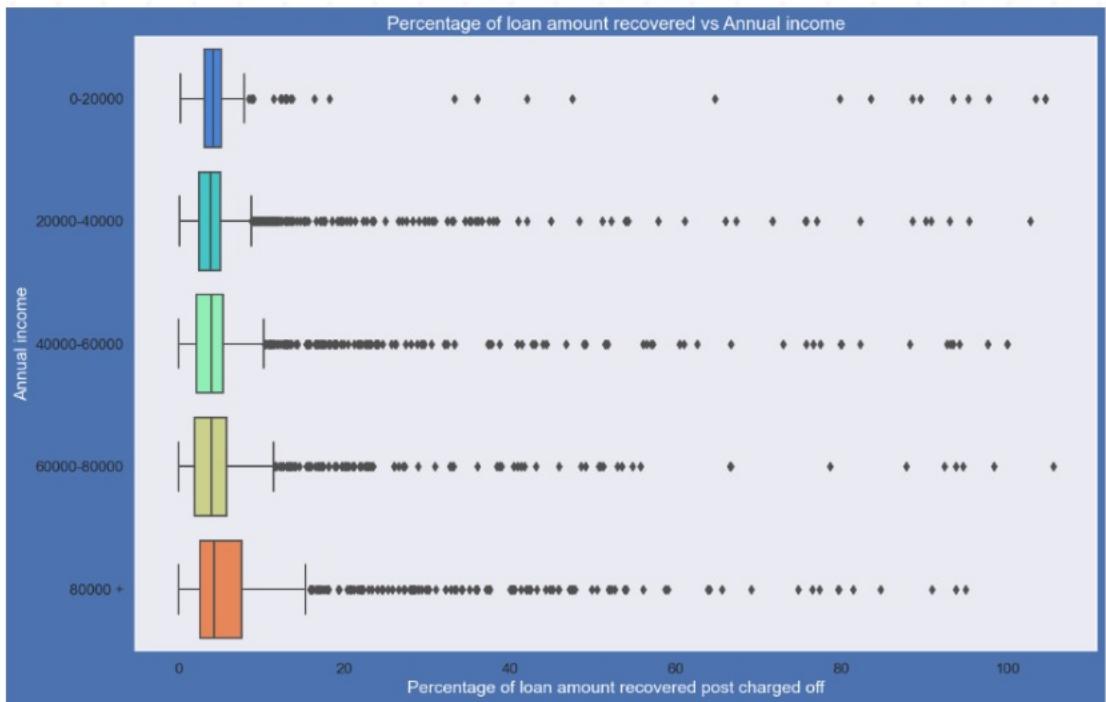
DTI vs Interest Rate:



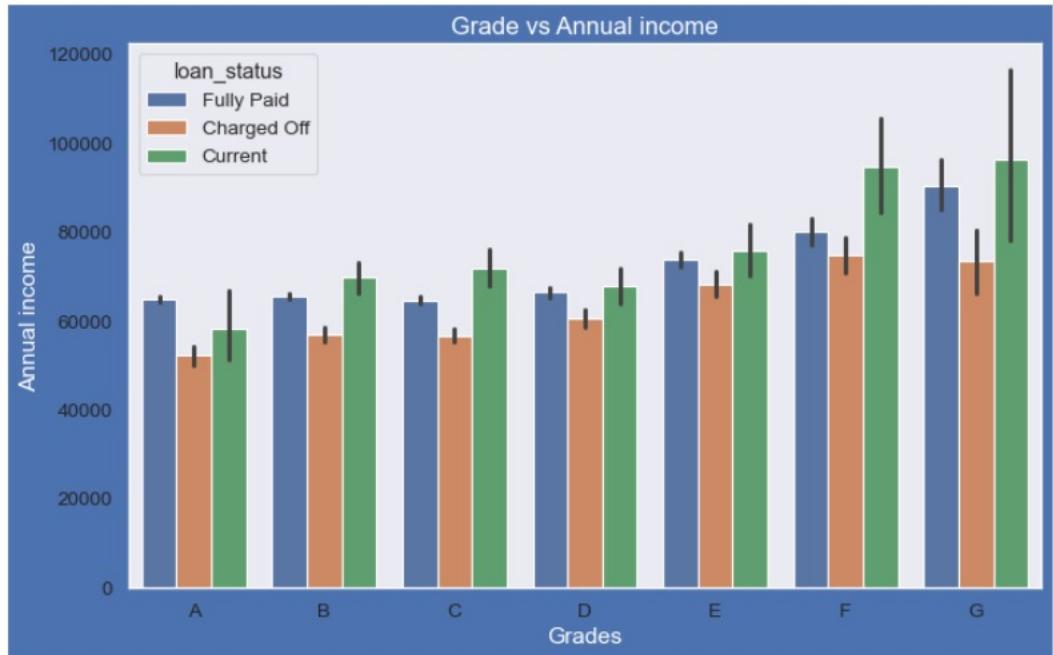
- ❖ While the plot indicates no significant variation, there is a slight increase in interest rates with higher DTI (Debt-to-Income) ratios, suggesting that borrowers with lower DTI may have the potential for lower interest rates.

Recoveries vs Annual income:

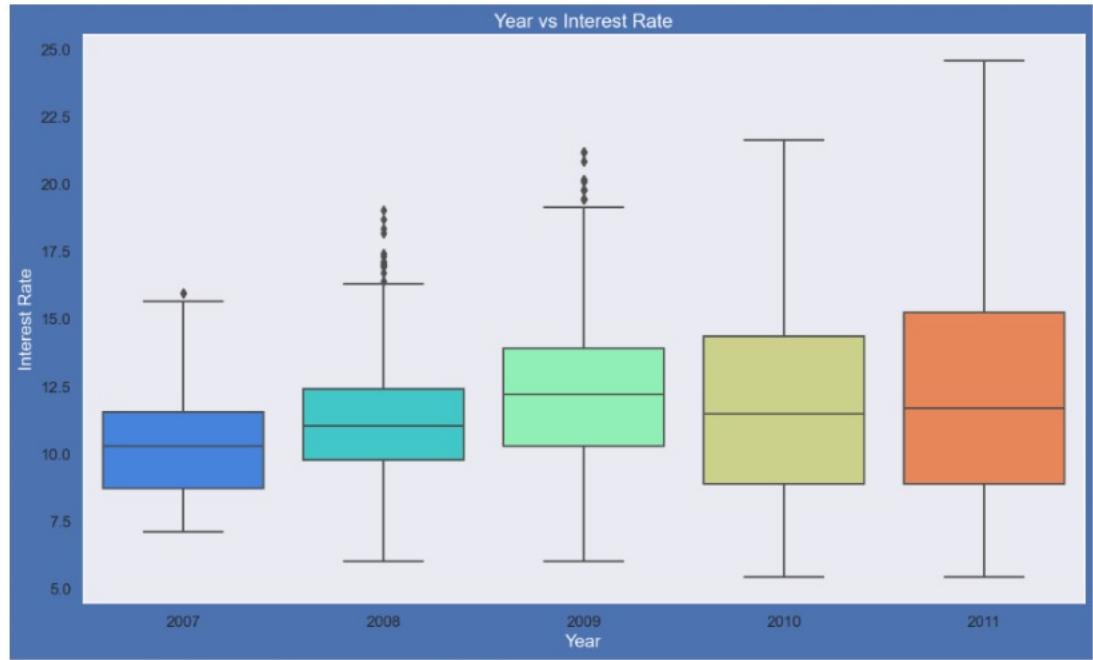
- ❖ Higher percentage of loan amount is recovered when annual income is high.
- ❖ Plot shows no significant variation but there is slight increase in recovery percentage with increase in annual income.



Grade vs Annual Income:



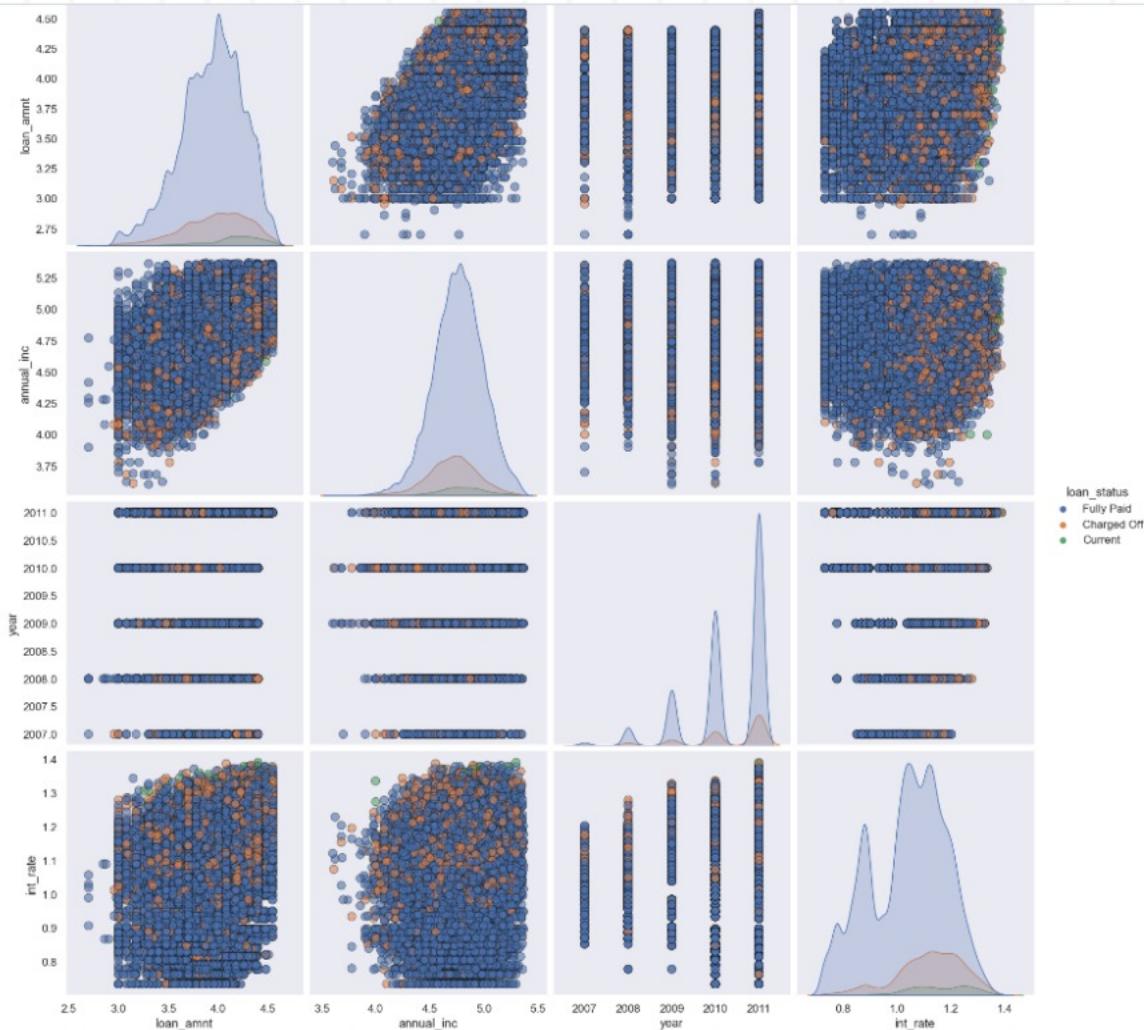
Year vs Interest Rate:



- ◆ Borrowers who are "charged off" tend to have lower annual incomes compared to those who have "paid fully" across all grades, considering loans within the same interest rate range.

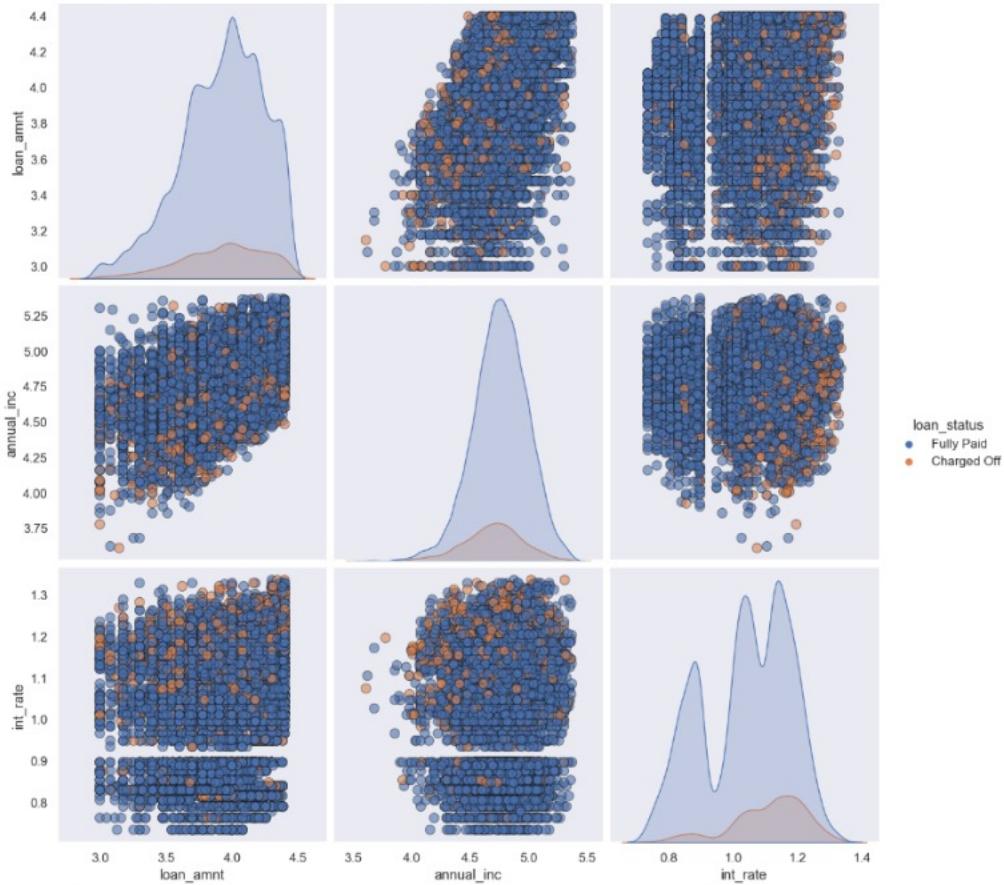
- ◆ The plot demonstrates a gradual increase in interest rates over the years.

Multivariate Analysis - Pair Plots:



- ❖ A higher interest rate is associated with a higher charged-off ratio.
- ❖ Borrowers with higher annual incomes tend to secure slightly larger loan amounts.
- ❖ The number of charged-off loans tends to increase over the years.
- ❖ The interest rate generally increases with higher loan amounts, potentially contributing to a higher charged-off ratio.

Multivariate Analysis - Pair Plots:



Interest Rate Impact:

❖ There is a correlation between higher interest rates and a higher charged-off ratio, suggesting that loans with higher interest rates are associated with a greater risk of default.

Income and Loan Amount:

❖ A higher annual income is generally linked to a slightly higher loan amount, indicating a positive relationship between income and borrowing capacity.

Interest Rate and Loan Amount:

❖ Interest rates tend to increase with the loan amount, resulting in a higher charged-off rate, likely due to the additional financial burden for borrowers.

Conclusion:

Default Rate: On average, 14% of loans result in defaults.

Key Factors:

- ❖ Lengthy tenure loans (60 months) have double the default rate compared to shorter-term loans (11%).
- ❖ Higher-grade loans with higher interest rates are associated with an increased risk of default.
- ❖ High credit utilization (Revol_util) is linked to a higher risk of loan repayment issues.

Inferences:

- ❖ Loans with high amounts are often given for 60-month, upper-grade, and small business loans, which have a higher chance of default.
- ❖ Borrowers with over 10 years of employment and verified income experience a higher default rate.
- ❖ Defaults on upper-grade loans result in substantial losses due to the high loan amounts involved.
- ❖ Interest rates for 60-month loans are generally higher than those for 36-month loans.

「

Thank You

」