LENDING CLUB

EDA CASE STUDY

upGrad and IIITB Machine Learning and AI Program

Group - Niharika Patne & Sandhya Puru Shar

PROBLEM STATEMENT: The objective is to analyze historical loan applicant data to identify patterns of default risk, enabling decisions on loan approval, modification, or denial based on the applicant's willingness to repay.

PURPOSE OF THE CASE STUDY: Performing a comprehensive EDA on the Loan Data set, and analyzing correlations between all the independent variables.

- Identify characteristics of borrowers who are more likely to default.
- Analyze the impact of loan amount, income level, and other financial metrics on loan performance.
- Provide actionable insights that could help in refining lending practices and risk assessment methodologies.

DATASET OVERVIEW:

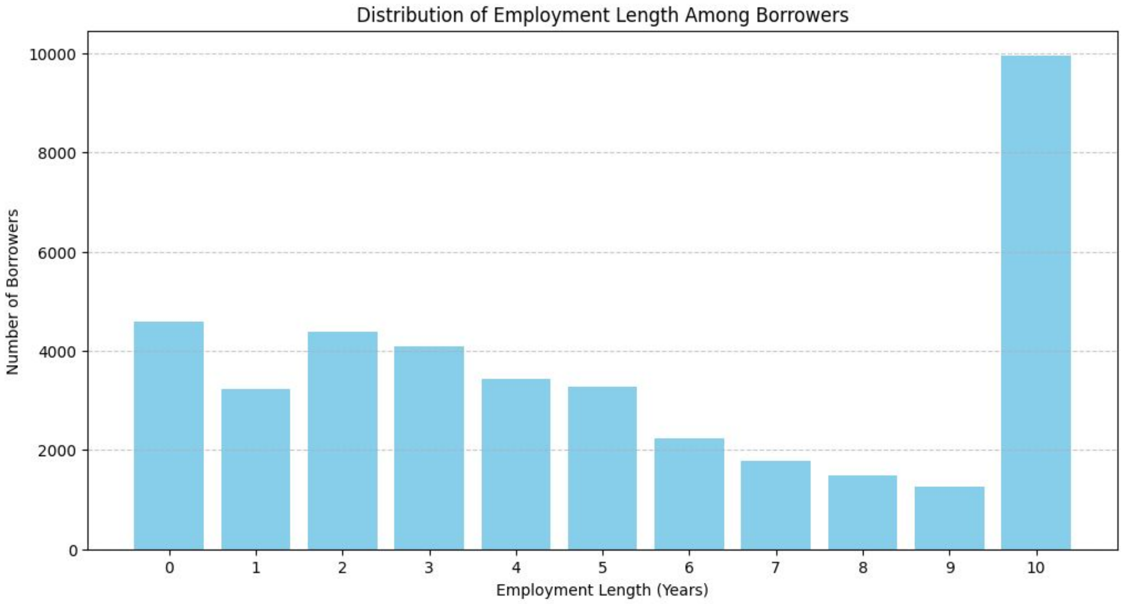
- Target Variable: Loan_status - where the customer has Fully Paid, Charged off, Current paying the loans.
- Important variables: Installment, annual income, DTI, loan status, grade, demographic variables of the customer, etc.

Demographics of Loan Paying Customers

1.The 10-year employment length category has the highest number of borrowers, nearly reaching 10,000 individuals.

2. Employment lengths of "0" years (less than one year) and "2 years" also exhibit high frequencies, indicating a significant presence of relatively new workers in the borrower pool.

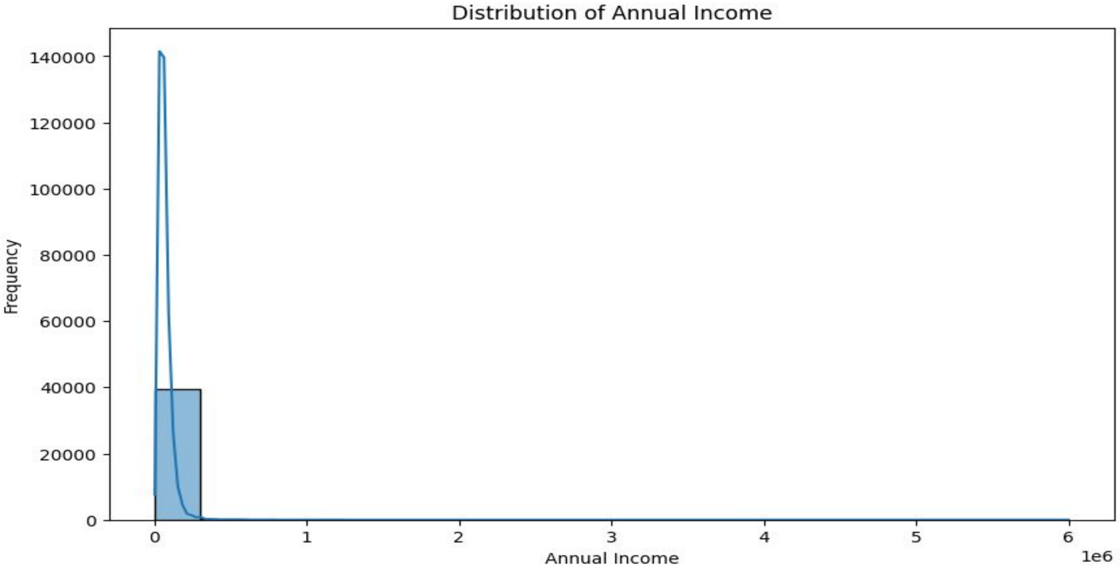
3.Between 6 and 9 years are less common among borrowers, highlighting a lower representation of mid-range employment stability within the dataset.



Understanding Loan Payers Ability to Pay

1.The frequency of borrowers with lower annual incomes is significantly higher. The highest frequency is observed at the lowest income bracket, which rapidly decreases as income increases.

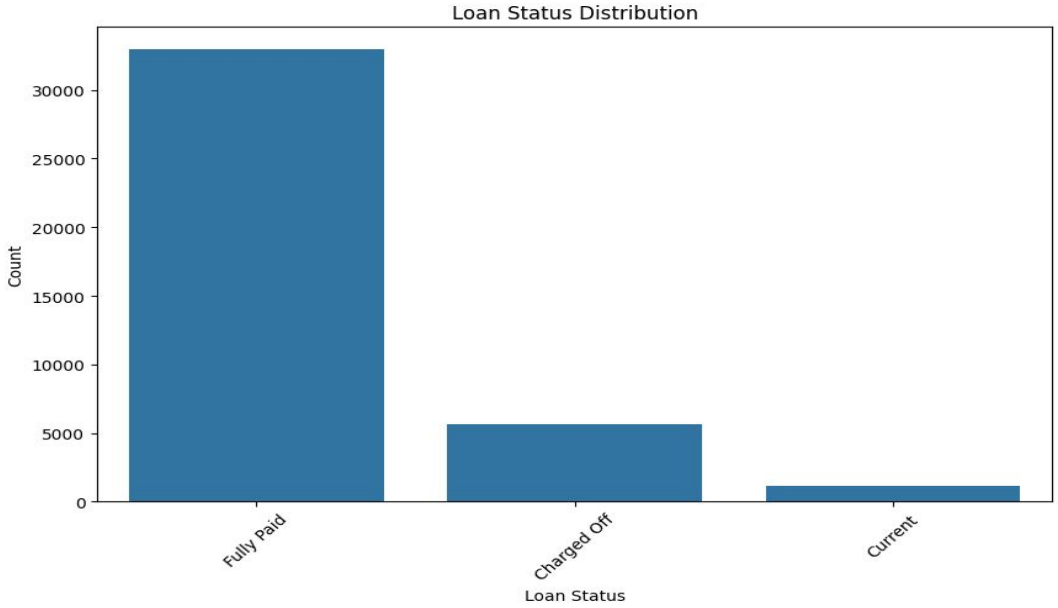
2.The presence of high-income borrowers is minimal, which might reflect the platform's user base or possibly the general accessibility and appeal of such financial products to lower-income individuals.



Understanding the Target Variable

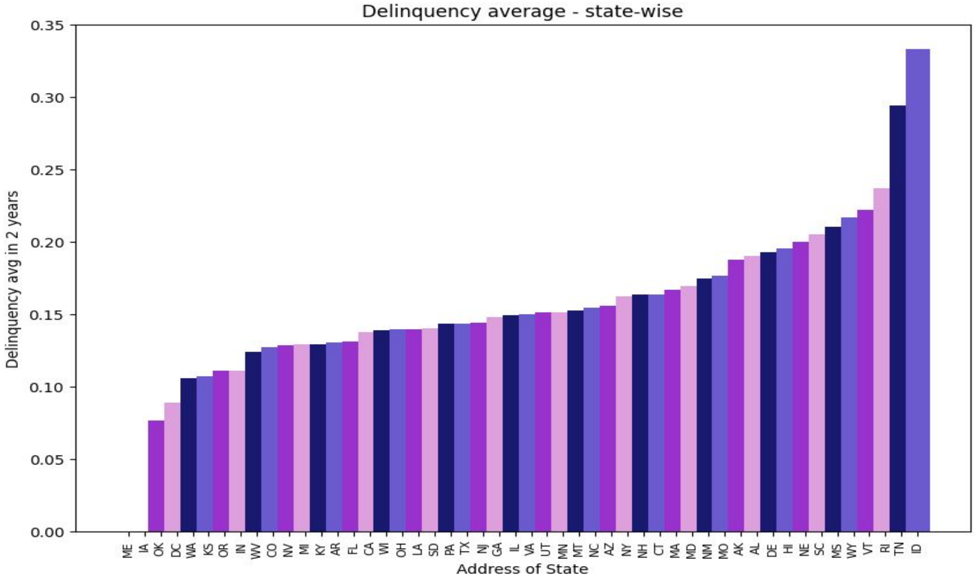
1.Loan Performance: The high volume of "Fully Paid" loans indicates good performance or successful selection of borrowers who are capable of repaying their debts.

2.Risk Management: The existence of "Charged Off" loans highlights the inherent risks in lending and underscores the need for effective risk assessment and management strategies.



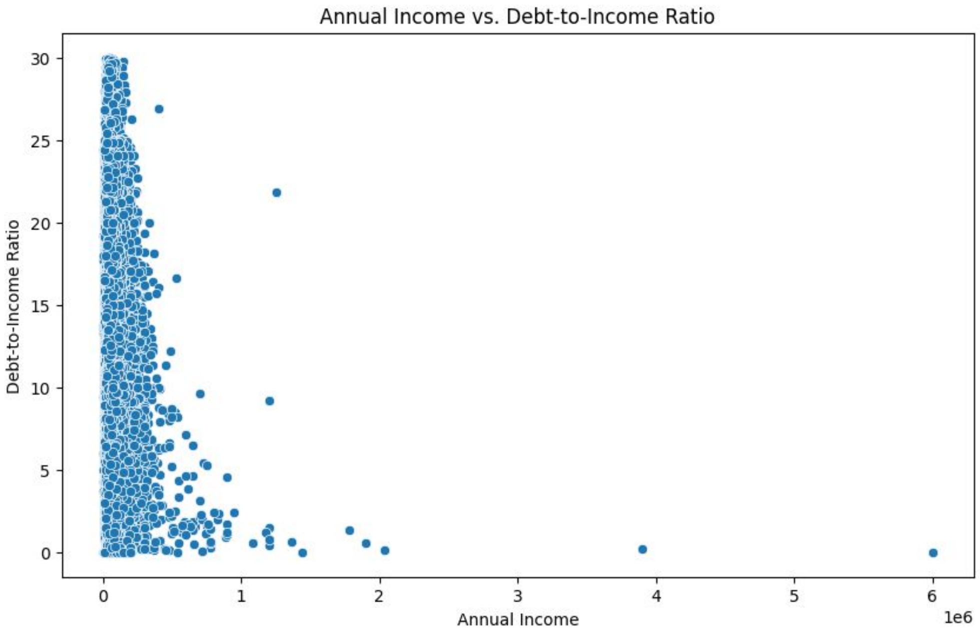
Delinquency average vs State

Loan-takers prevailing from states Tennessee and Idaho, have a higher delinquency as compared to other states



Annual Income vs. DTI

As income increases, the distribution of DTI ratios tends to compress and lower, suggesting that higher-income earners generally manage lower DTI ratios.
This is likely because their higher income provides more financial leeway in managing debt.

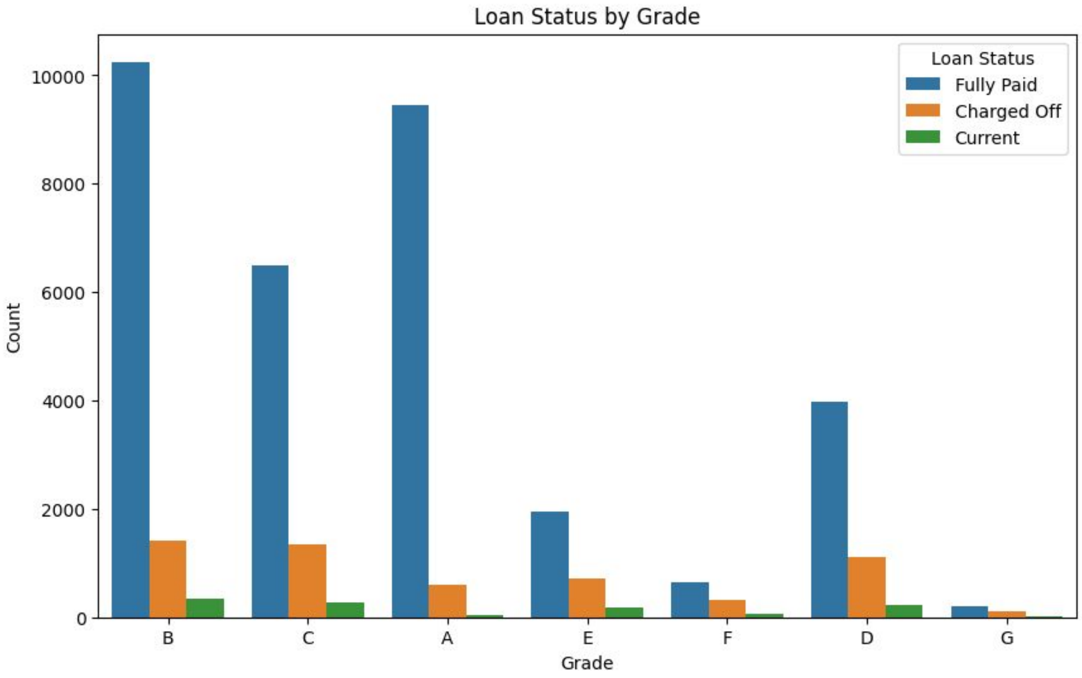


Loan status by grade

1.The frequency of charged off loans increases as the grade worsens, with the highest numbers seen in grades D through G. This trend highlights the increased risk associated with these grades.

2.Grade G, despite its lower volume of loans, has a significant proportion of charged off loans, underscoring the high risk associated with lending to borrowers in this category.

3.Charge offs for Grade A are comparatively low with the amount of Fully paid loans.



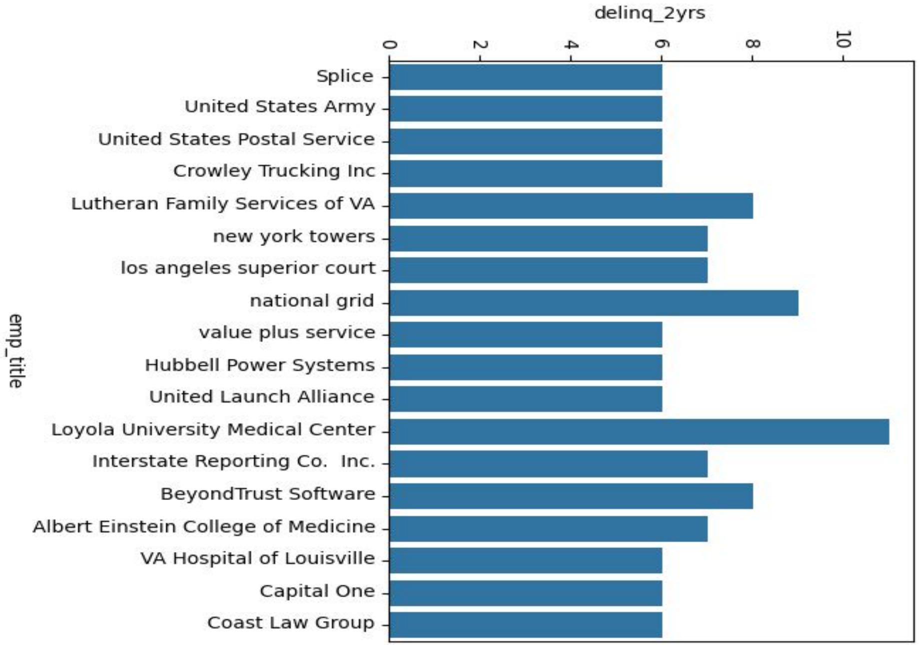
Applicants with emp_titles vs Delinquency

Applicants with emp_titles :

1. 'Loyola University Medical Center', 'national grid', 'Lutheran Family Services of VA', and 'BeyondTrust Software' and few such are more delinquent as compared to employees with other titles.

 2.This is an indicator to the lending club to raise a risk, scrutinize pay-backs and perform continuous follow-ups for loans taken by employees of this title.

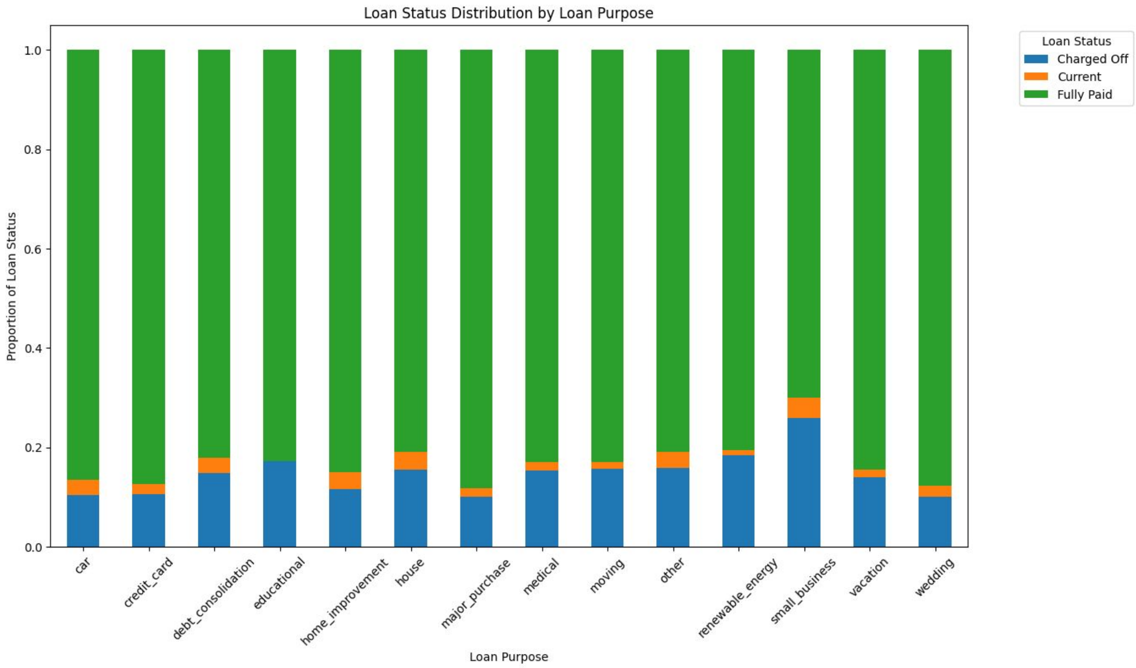
3.And also, while issuing new loans to employees of such titles,he lending club can perform thorough background checks to verify the pay-back capability.



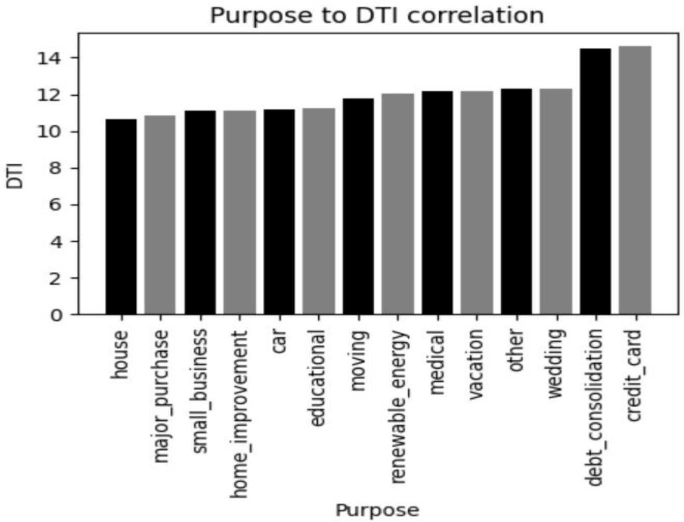
Loan status across loan purposes

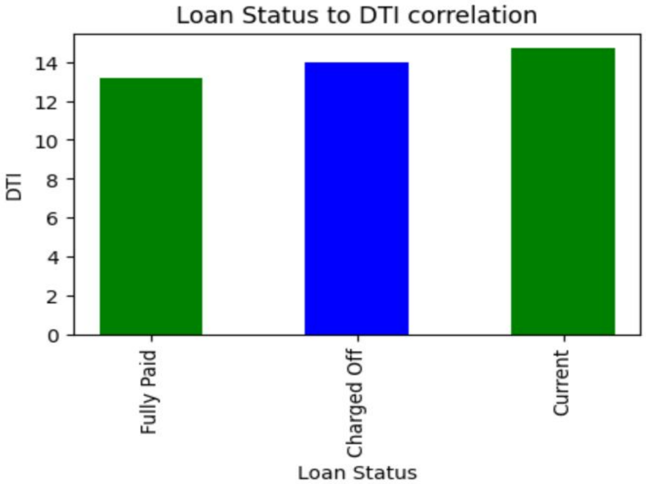
Certain loan purposes like small_business and renewable_energy show a relatively higher proportion of loans being charged off compared to others.

This suggests these types of loans carry a higher risk of default.



Multivariate Analysis





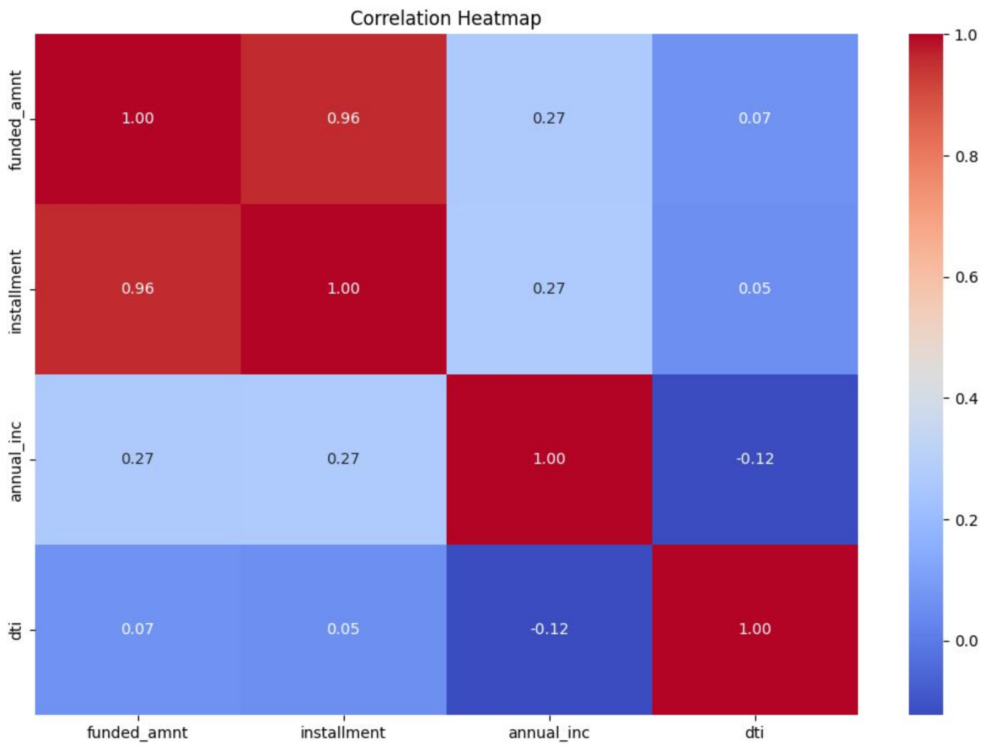
It is depicted that the Purposes for loan owing to 'debt_consolidation' and 'credit_card' are the ones with loan takers having higher DTI values. In the second plot, on an average the higher DTI maps to the loans being Charged Off. This concludes that when the Lending Club comes across Purposes listed above, probably the loan should not be approved, or the interest rates should be high. It also indicates high risk for the Current loans falling under these Purposes.

Correlation heatmap

1.'funded_amnt', 'installment' - Highly correlated indicating that as the loan amount increases, the installment payments increase almost proportionally.
funded_amnt and installment have a moderate positive correlation (0.27) with annual_inc.

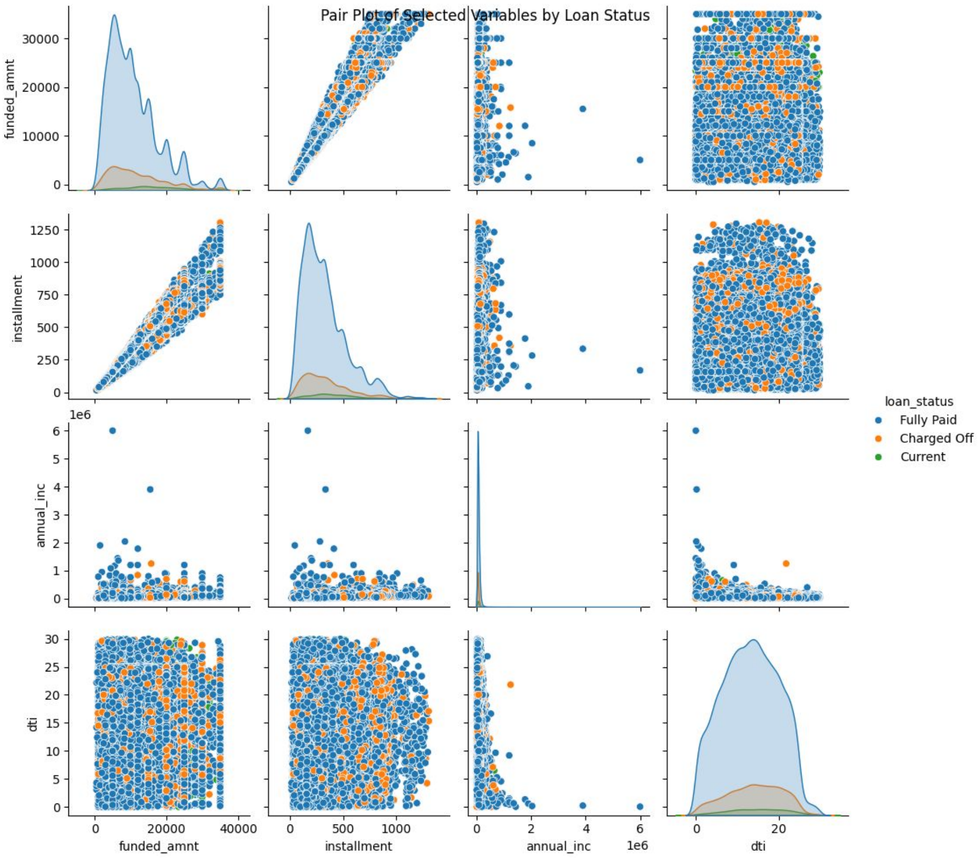
2.This suggests that borrowers with higher incomes tend to take out larger loans, which subsequently have higher installments.

3.Dti shows lower correlation with all the three variables.



Correlation between Important Variables

1.Loan Amount and Installments: The direct correlation suggests that as the loan amount increases, so does the financial burden on the borrower, which can impact the likelihood of repayment.
Income Insights: Higher incomes do not necessarily correlate with higher loan amounts, which might suggest a conservative borrowing behavior among higher earners or possibly that they are less represented in the data.
2.Risk Factors: Higher DTIs and higher loan amounts seem to correlate with a greater likelihood of a loan being charged off. Lenders might use this information to adjust credit risk assessments or loan terms.
3.Debt Management: The broad distribution of DTIs and its relative independence from income suggests that debt management and financial planning are critical across all income levels.



Scope of Case Study

Summarize Key Findings:

- Highlight the identified risk factors associated with different grades, employment titles, states, and loan purposes.
- Emphasize the impact of DTI and income levels on loan performance.

Future Directions:

- Suggest continuous data monitoring and analysis to update lending strategies as borrower behaviors and economic conditions evolve.
- Recommend exploring advanced analytics techniques and machine learning models to predict default more accurately.

Call to Action:

- Encourage implementation of the recommendations to improve loan performance.
- Propose setting up a task force to review and integrate these insights into current lending practices.

Conclusion: Optimizing Loan Performance and Mitigating Risks

Lending Club can enhance its ability to manage risks effectively, leading to a more stable loan portfolio with improved repayment rates and reduced defaults:

1. Intensify risk assessments for loan grades D through G, known for higher default rates, by tightening lending criteria and possibly adjusting interest rates or loan terms to reflect the associated risk.
2. Exercise increased caution for loans intended for high-risk purposes such as small businesses and renewable energy, applying stringent review processes and adjusting financial terms to manage these risks appropriately.
3. Leverage correlation insights between loan amounts, installments, and annual incomes to tailor loan offerings effectively, ensuring that loans are sustainable for borrowers' financial situations and closely monitoring larger loans to prevent borrower over-leverage.
4. Sharpen focus on debt-to-income ratios, particularly for debt consolidation and credit card loans, setting stricter limits and requiring solid proof of income to prevent financial overstretch.