LENDING CLUB CASE STUDY

By:
Somasekhar Gangarapu
Ursina Sanderink

INTRODUCTION

Objective:

- This project aims to address the challenge of loan defaults in the lending industry, specifically for Lending Club, a major online loan marketplace.
- The focus is on improving loan approval processes by identifying patterns that predict the likelihood of applicants defaulting on their loans.

Background:

- This project is part of the AI & ML program @ UpGrad, specifically within a statistics course.

Dataset:

- The dataset used in this project `loan.csv` includes historical data on past loan applicants, covering their repayment history and loan outcomes.
- The key categories of loan outcomes are:
  - Fully Paid: Loans that have been repaid in full.
  - Current: Loans that are still being repaid.
  - Charged-Off: Loans where the applicant has defaulted.

HIGH LEVEL PROCESS

Data Preparation

- Load and inspect the dataset for completeness and integrity
- Understand the data columns by correlating with data dictionary

Data Cleaning

- Drop irrelevant, non-reliable, or non-informative columns
- Handle missing values
- Standardize data
- Handle outliers
- Sanity Checks

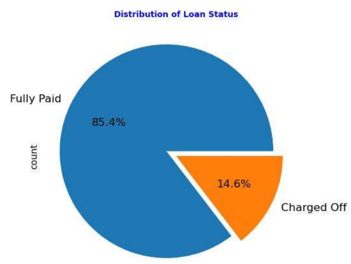
EDA

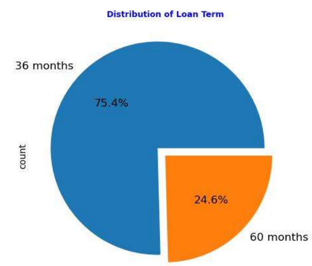
- Univariate Analysis
- Segmented Univariate Analysis
- Bivariate analysis

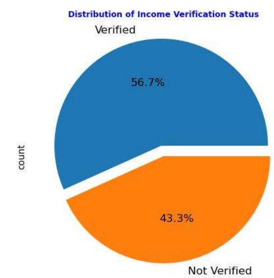
DATA PREPARATION & DATA CLEANING

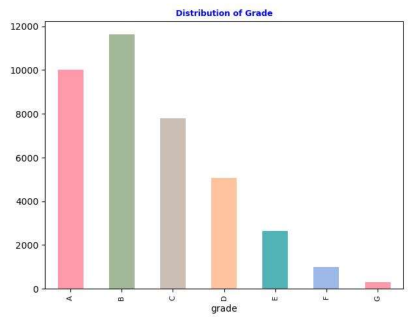
- Loaded dataset ‘loan.csv’ and noticed it have 39717 rows and 111 columns.
- Interpreted each column with data dictionary provided.
- Dropped columns with high irrelevant or missing data.
- Identified and filled missing values with appropriate statistical measures.
- Transformed data into consistent and standard format.
- Identified and removed outliers

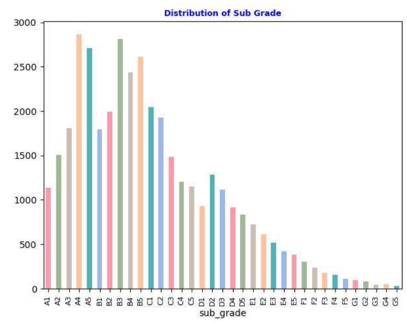
EDA – UNIVARIATE ANALYSIS

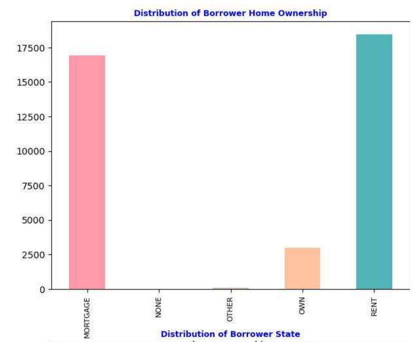


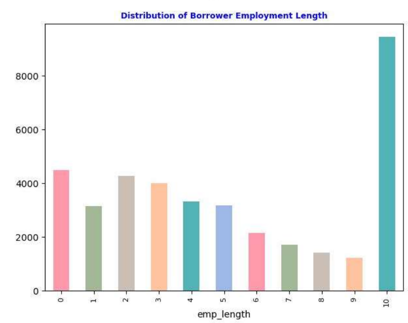


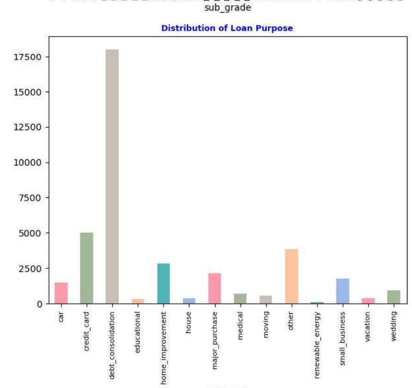


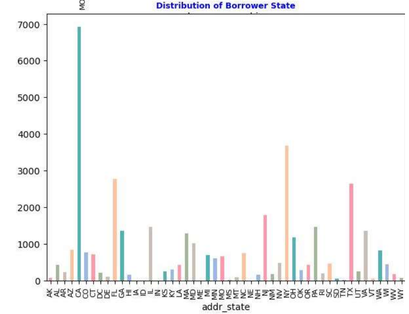












•Loan Status: The pie chart shows that 85.4% of loans are fully paid, while 14.6% are charged off. This indicates a *relatively low default rate*, which is positive for the lending platform.

•Loan Term: 75.4% of loans have a 36-month term, while 24.6% have a 60-month term. *Shorter-term loans are significantly more common*, possibly due to lower risk or borrower preference.

•Income Verification Status: 56.7% of borrowers have verified income, while 43.3% are not verified. This suggests that the platform accepts a *substantial number of loans without income verification*.

•Grade: Loans are graded from A to G, with B being the most common, followed by A and C. This *indicates a concentration in higher-quality loans* according to the platform's risk assessment.

•Sub Grade: The sub-grades show a more granular risk assessment, with A4, A5, B3, B4, and B5 being the most common. This *reflects a nuanced approach to risk categorization*.

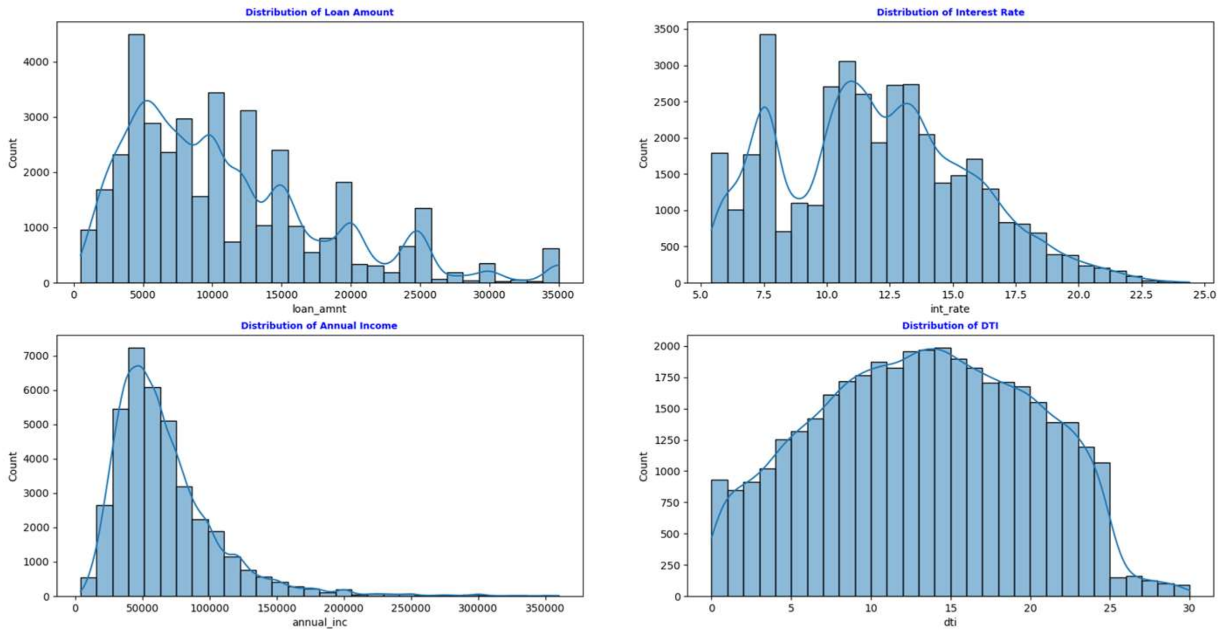
•Borrower Home Ownership: Most borrowers are renters, followed by those who own homes. Mortgage holders are the smallest group, suggesting a *diverse range of borrower housing situations*.

•Borrower Employment Length: There's a significant spike at the "10+" years category, indicating many long-term employed borrowers. *Other categories are more evenly distributed*.

•Loan Purpose: Debt consolidation is overwhelmingly the most common loan purpose, followed by credit card payoff. This suggests *many borrowers are using loans to manage existing debt*.

•Borrower State: California (CA) has the highest number of borrowers, followed by New York (NY) and Texas (TX). This *likely reflects population density and economic activity in these states*.

EDA – UNIVARIATE ANALYSIS



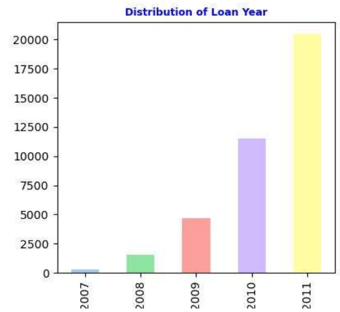
•**Loan Amount:** The majority of loan amounts applied for by borrowers are concentrated between 4K and 10K dollars, with a notable peak around $5,000. This indicates that *most borrowers prefer moderate loan amounts* rather than very high or low amounts.

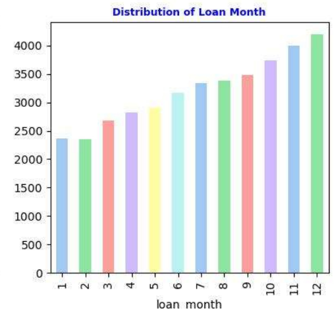
•**Interest Rate:** Interest rates exhibit a roughly normal distribution with peaks around 11%. This suggests that most loans are offered within this interest rate range, which could be a common range set by the lending policies.

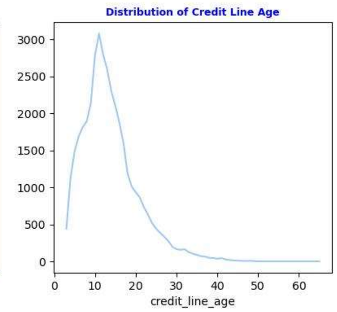
•**Annual Income:** Annual incomes are right-skewed, indicating that a *significant portion of the borrowers report incomes below $100,000*. Higher incomes are less common, which might reflect the target demographic of the lending club focusing on middle-income earners.

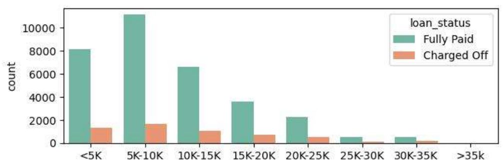
•**DTI (Debt-to-Income Ratio):**The DTI ratio is evenly distributed with a peak around 15%, suggesting that *many borrowers have a moderate level of debt compared to their income*. This ratio is crucial for assessing borrowers' ability to manage additional debt obligations.

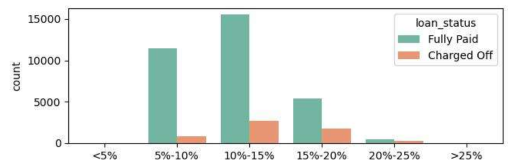
EDA – UNIVARIATE ANALYSIS – DERIVED & SEGMENTED

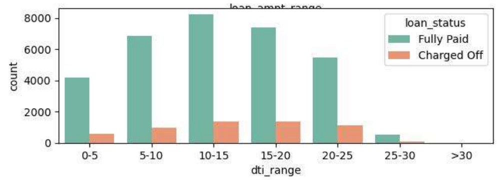


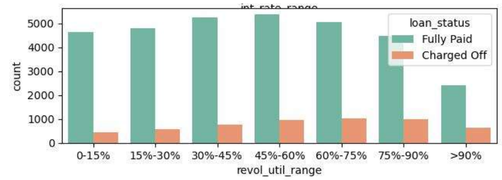












• Significant increase in the number of loans issued from 2007 to 2011, with 2011 having the highest number of loans by far. This trend *indicates rapid growth in loan issuance over these years*, possibly reflecting an expansion in market reach or changes in lending policies that made more capital available or more borrowers eligible.

• Clear seasonal pattern throughout the year. Loan numbers are lowest in January and February, then gradually increase with the highest number of loans occurring in December (month 12).

• The distribution of credit line age displays a pronounced right-skewed pattern with a peak occurring sharply between 10 and 15 years.

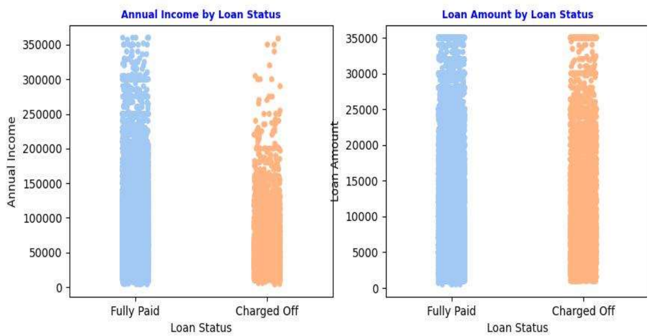
• Majority of loans issued for amount ranging $5K-10K.

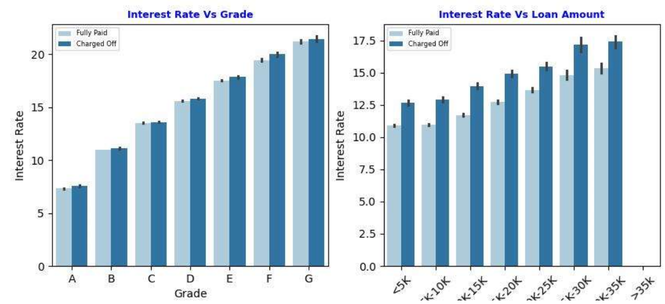
• The 10%-15% interest rate range is most prevalent, followed by 5%-10%.

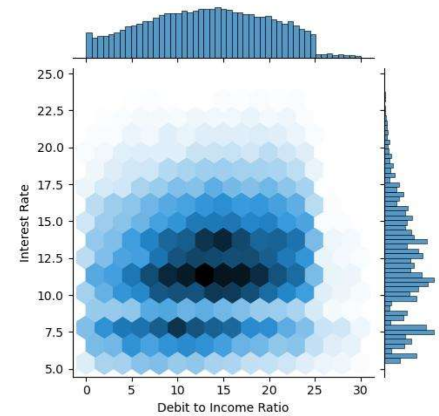
•Debit to Income Ratio distribution peaks in the 10%-15% range, with a fairly even spread across 0%-25%.

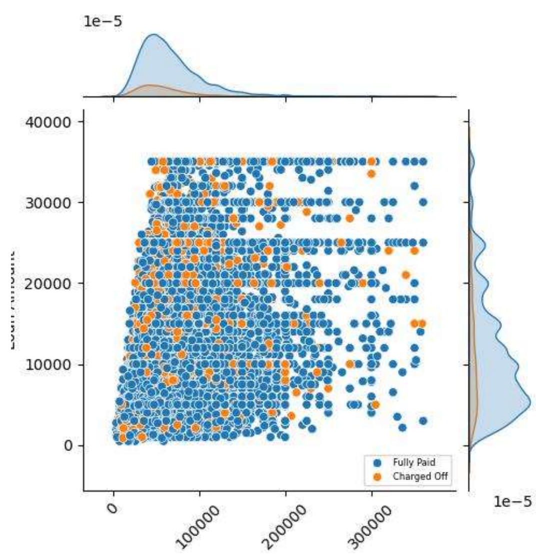
•Revolving Utilization distribution is relatively even across 0-60% ranges, dropping for ranges above 60.

EDA – BIVARIATE ANALYSIS









• Higher-income borrowers are more likely to fully repay loans.

• Slight tendency for higher default rates on larger loans.

• Clear positive correlation between loan grade and interest rate, with rates increasing as the grade moves from A to G.

• Charged-off loans consistently have slightly higher interest rates than fully paid loans across all grades, indicating that higher-risk borrowers are charged more.

• General trend of increasing interest rates as loan amounts increase.

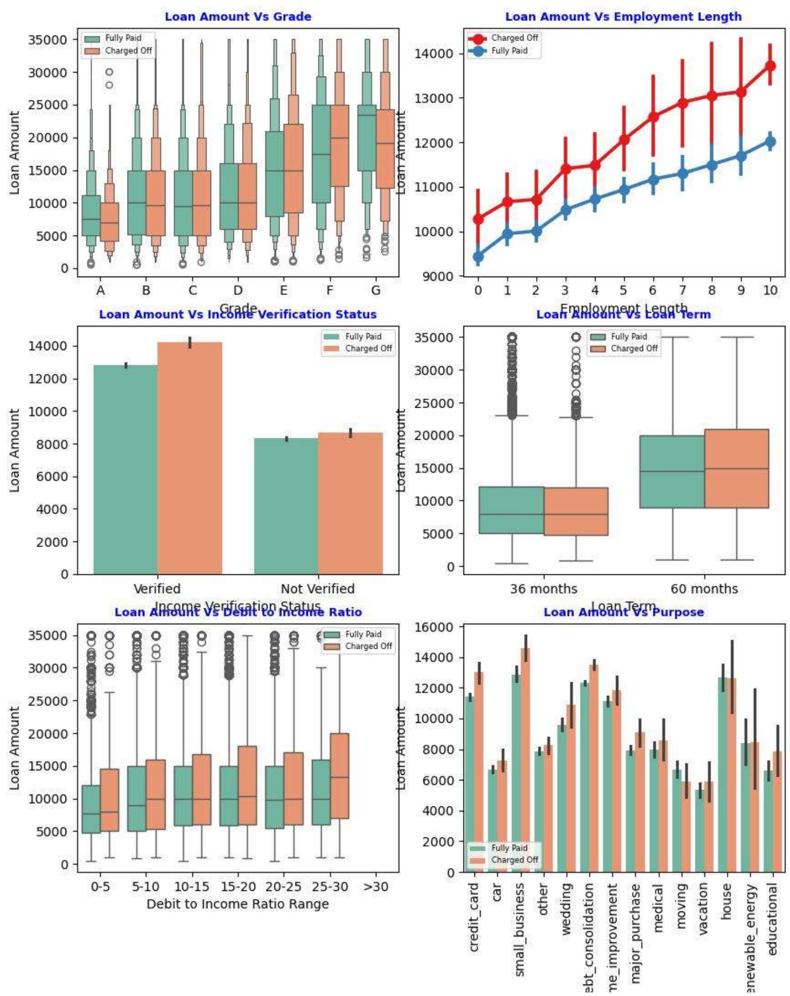
• The gap between interest rates for fully paid and charged-off loans remains fairly consistent across loan amounts.

• The concentration of loans in the 10-20 DTI range with 10-14% interest rates likely represents Lending Club's current "safe zone" for loan approvals.

• Majority were issued to borrowers with income around 50K dollars and for loan amount 5K dollars.

• Higher charged off noticed in loans issued to low income borrowers.

EDA – BIVARIATE ANALYSIS



• The distribution of loan amounts for both fully paid and charged-off indicates that loans under grade F have high risk of getting charged off especially higher amount loans.

• There's a clear positive correlation between employment length and loan amount for both fully paid and charged-off loans, with charged-off loans consistently higher across all employment lengths.

.

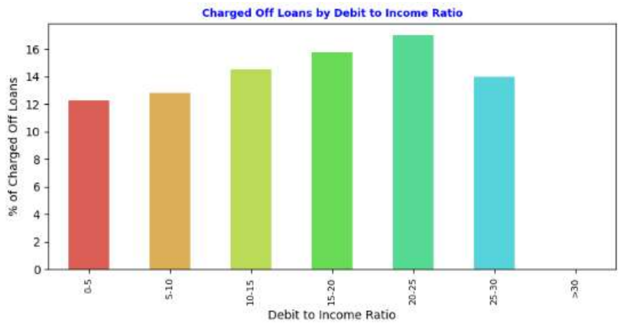
•Verified incomes correspond to significantly higher loan amounts for both fully paid and charged-off loans. This counterintuitive result suggests that income verification alone may not be sufficient for assessing repayment capability.

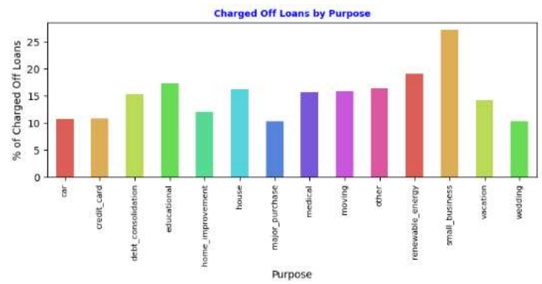
•The higher upper quartile for charged-off 60-month loans is particularly noteworthy, indicating that a significant portion of larger long-term loans are at higher risk of default.

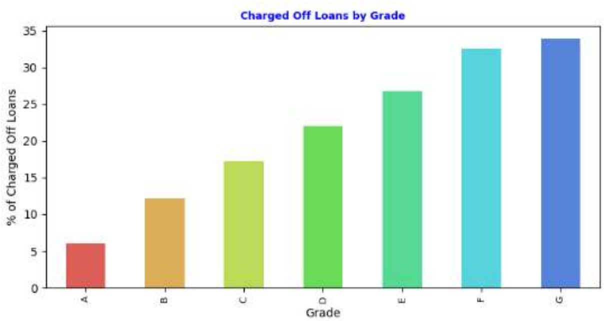
•The graph shows a slight increase in loan amounts as DTI increases, with charged-off loans generally having higher amounts across all DTI ranges.

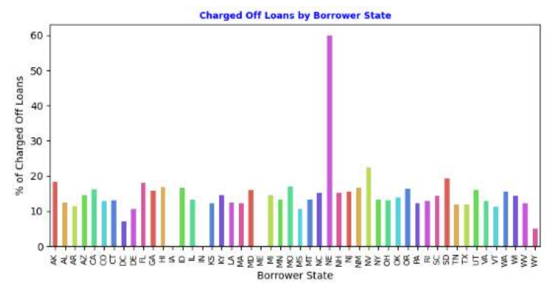
•Small business loans stand out with the highest average amounts and the largest gap between fully paid and charged-off loans. Debt consolidation, house and credit card refinancing also show high loan amounts..

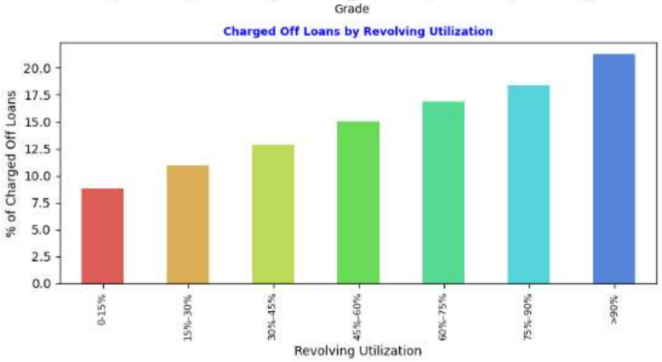
EDA – BIVARIATE ANALYSIS

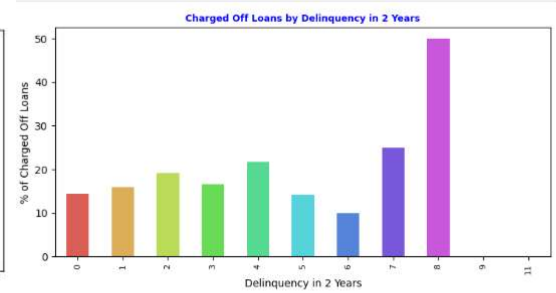






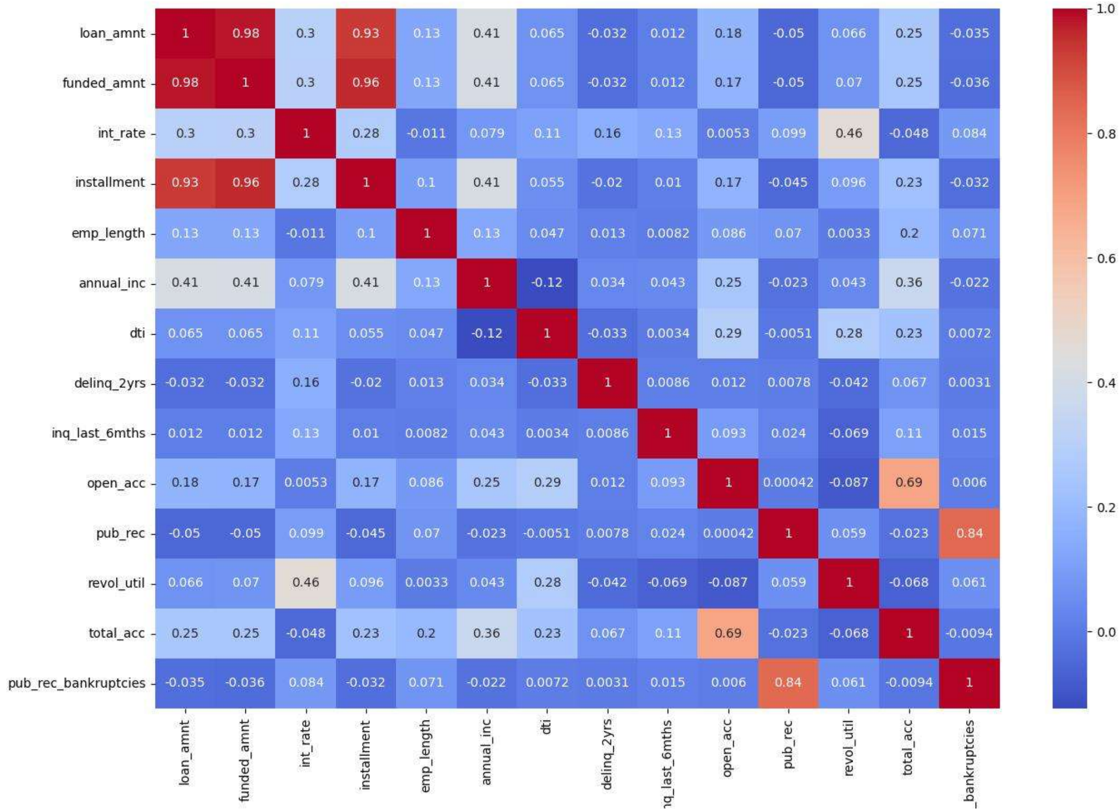






- Percentage of loans charged off significantly increases as the DTI ratio climbs, peaking distinctly in the 21-25% range where over 16% of the loans end up charged off.
- Percentage of charged-off loans increases progressively from grade A through G.
- The percentage of charged-off loans increases significantly as revolving utilization grows.
- Loans issued for 'Small business' have the highest charge-off rates (27%).
- The percentage of charged-off loans varies notably by state. While Nebraska (NE) shows the highest default rate at 60%, the extremely low volume of only 5 loans requires cautious interpretation. Following NE, the states with the highest default rates and more substantial data are Nevada (NV) at 22.43% with 477 loans, South Dakota (SD) at 19.35% with 62 loans, Alaska (AK) at 18.18% with 77 loans, and Florida (FL) at 18.02% with 2769 loans..
- Delinquency within two years demonstrates that higher delinquencies correlate with increased charge-off rates, peaking at eight delinquencies where the loans charged off reach 50%.
- The percentage of charged-off loans increases with the number of public record bankruptcies, from 14.24% with no bankruptcies to 40.00% with two bankruptcies.
- Charged-off rates increase with the presence of derogatory public records, particularly noticeable for borrowers with one record, reflecting a 22.74% charge-off rate compared to 14.14% for those with none

EDA – BIVARIATE ANALYSIS



- Loan amount is correlated highly with Annual Income, moderately with total accounts
- Interest rate in correlated highly with Revolving Utilization.
- Interest rate is moderately with loan amount and slightly with delinquency count in 2 years, Inquires in 6 months and DTI (Debit to Income)
- As expected,
  - Public derogatory records highly correlated with public records
  - Installment is highly correlated to Loan amount
  - Annual Income is correlated with employment length

CONCLUSIONS

- ❑ **Loan Purpose:** Debt consolidation and credit card payoff are the most common purposes, indicating a prevalent use of these loans for managing existing debt.
- ❑ **Seasonal Trends:** There is a clear seasonal pattern in loan issuance, peaking in December, which may reflect borrowers' financial planning trends at year-end.
- ❑ **Income Verification and Default Rates:** The status of income verification does not consistently correlate with lower default rates, suggesting that it should not be overly relied upon, especially for larger loans.
- ❑ **Account Number Risks:** The number of open accounts presents a U-shaped risk profile, where very low and very high counts are associated with increased default risks, highlighting the complexity of financial management among borrowers with extreme numbers of credit lines.

Loans with below criteria are highly contributing to charged off:

- ✓ Loans with higher interest rate (>12%)
- ✓ Loans with grade 'F' and loan amount > 20K
- ✓ Higher amount loans (>13K) for small business, debt consolidation or credit card
- ✓ Borrower Debit to income ratio > 25% and loan amount > 15K
- ✓ Borrower revolving utilization > 60%
- ✓ Loans from states like AK, FL, NE, NV, SD
- ✓ Delinquency in 2 years > 2
- ✓ Borrower derogatory public records > 0
- ✓ Borrower annual income <50K and loan amount > 5K

RECOMMENDATIONS

Recommendation	Description
Implement Risk-Based Pricing	Develop a nuanced interest rate model that reflects diverse risk factors including DTI, credit utilization, and loan purpose, ensuring rates are commensurate with potential risks
Geographic Risk Management	Establish tailored lending criteria for different regions, especially in states with historically higher default rates, to mitigate location-based risks.
Specialized Debt Consolidation Programs	Offer tailored financial products for debt consolidation with accompanying advisory services to help borrowers manage their debts more effectively.
Credit Utilization Strategy	Tighten approval criteria for borrowers with high credit utilization rates (>60%), recognizing this as a significant predictor of potential default
Seasonal Lending Adjustments	Plan for seasonal fluctuations in loan demand, optimizing capital allocation and marketing strategies to match the observed end-of-year surge in borrowing.
Stringent Criteria for Small Business Loans	Increase scrutiny of small business loan applications, possibly requiring detailed business assessments or additional security to offset the higher observed risk.
Conservative Approach to Derogatory Records	Maintain stringent lending standards for applicants with derogatory public records or bankruptcies, potentially requiring additional assurances such as guarantors or collateral
Stricter DTI Thresholds	Enforce more rigorous reviews for loans where the DTI exceeds 20-25%, identifying this range as a critical risk threshold
Refined Credit History Assessment	Employ a detailed approach to credit history evaluation, wary of borrowers at the extremes of credit account numbers
Robust Income Verification	Strengthen the income verification process, especially for larger loan amounts, to ensure that reported incomes are accurate and reliable.
Adjust Loan Terms Based on Amount	Tailor the terms of loans, particularly the maximum amount and duration, to better manage the risk profile of longer-term, larger loans.

THANK YOU