Lending Club Case Study

Analyzing Loan Defaults and Risk Factors

(22/01/2025)

Kashish Sharma

Karthik Juluru

Problem Statement

Introduction

- In this case study, we applied the techniques of Exploratory Data Analysis (EDA) to gain insights into real business problems.
- Apart from using EDA techniques, we developed a basic understanding of risk analytics in banking and financial services, and learned how data is used to minimize the risk of losing money while lending to customers.

Problem Statement

- The dataset provided contains information about past loan applicants and whether they defaulted or not.
- Our aim was to identify patterns that indicate if a person is likely to default. These patterns can be used to make informed decisions such as denying the loan, reducing the loan amount, or lending to risky applicants at a higher interest rate.

Objectives

Business Objectives

- The company aims to identify risky loan applicants to reduce credit loss.

Analysis Goals

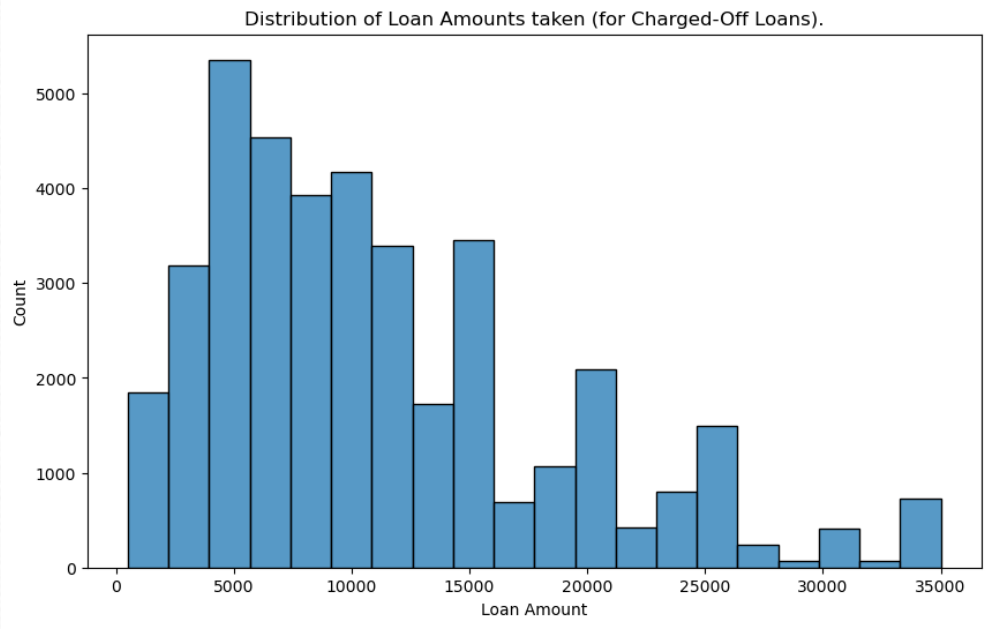
- The goal was to use EDA to understand how the loan attributes influence the tendency to default.
- Identify the driving factors behind loan default.
- Provide actionable recommendations to minimize financial risk.

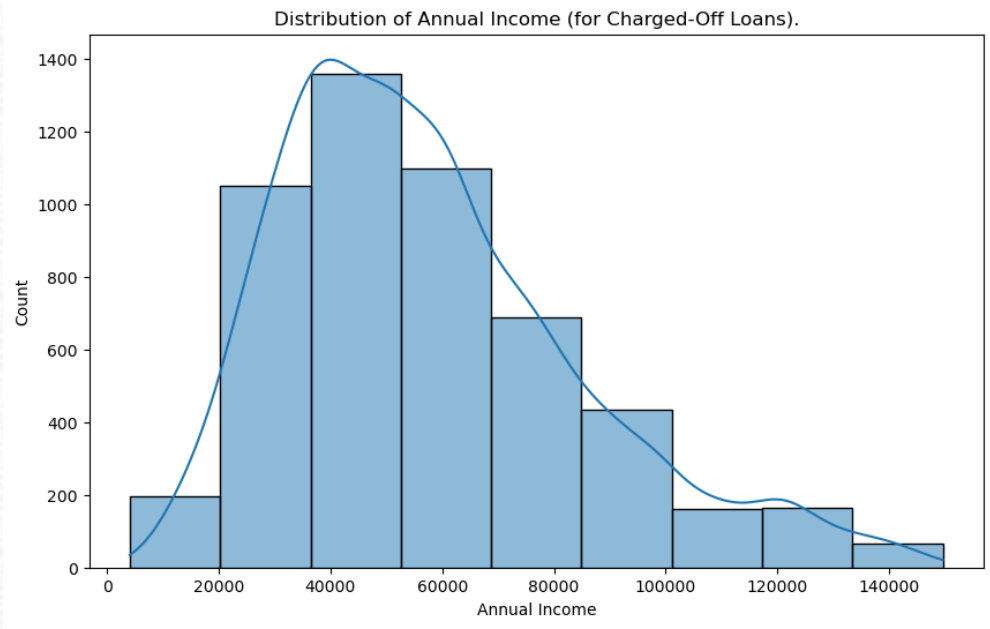
Data Cleaning

Cleaning Steps

- Out of 111 columns in loan.csv data set, we have considered 27 columns which were relevant to achieve our business objective.
- We handled missing values and outliers using python methods and various plots.
- Applied necessary transformations for data consistency, assigned good data types to make columns compatible with various operations and plots.

EDA - (Univariate Analysis)





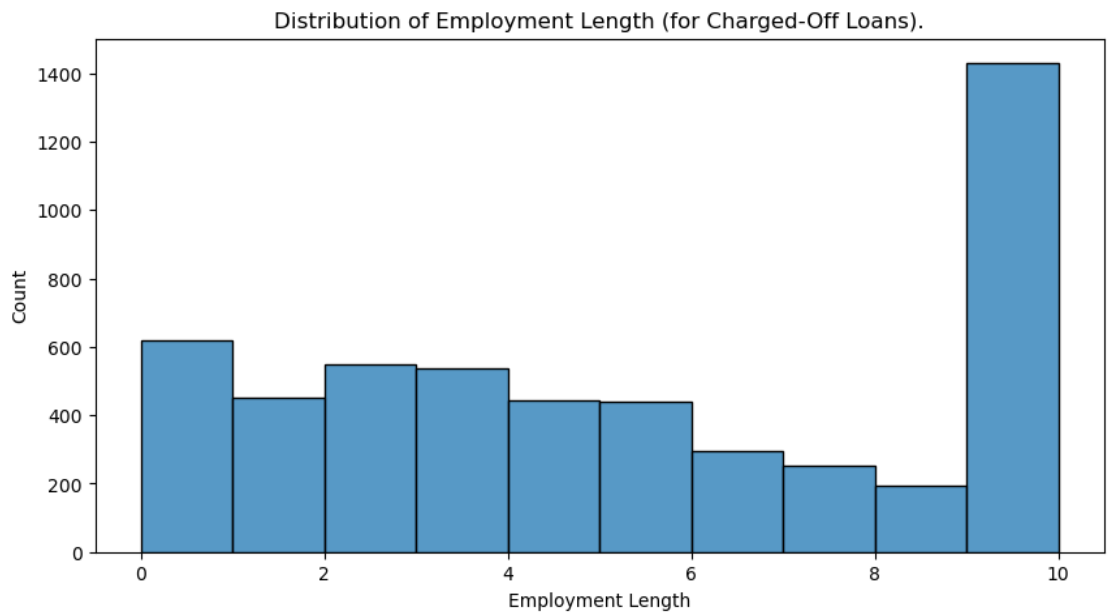
❖ **Inference**

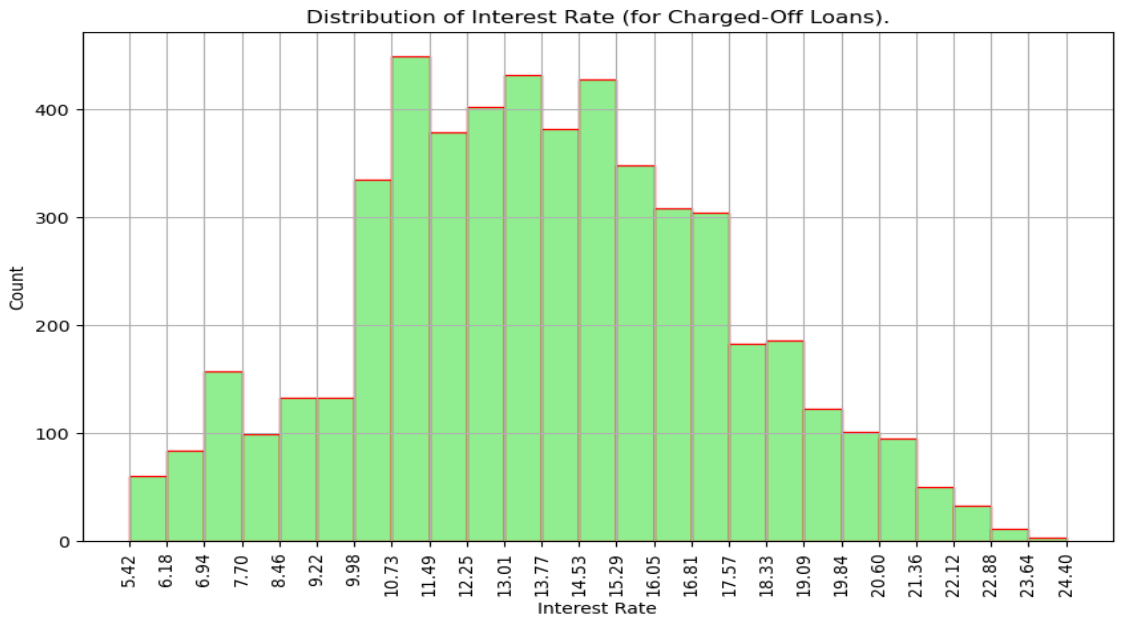
- The distribution of loan amounts is right-skewed, with most loans concentrated in the 0 to 15000 range, peaking at 5000 to 10000.
- Higher loan amounts above 30000 are rare, indicating limited demand or approval for large loans.

❖ **Inference**

- The distribution of annual income is again right-skewed, with most number of defaults done by individuals earning between 20000 to 70000, peaking around 40000.
- Individuals having higher income levels above 100000 are less likely to be defaulters.

EDA - (Univariate Analysis)





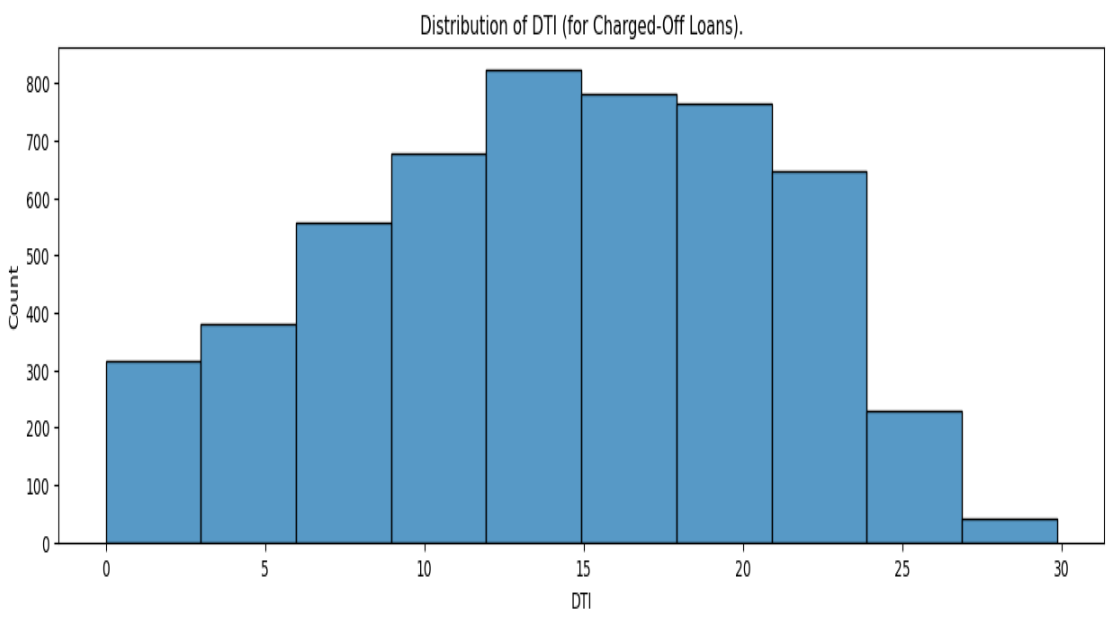
❖ **Inference**

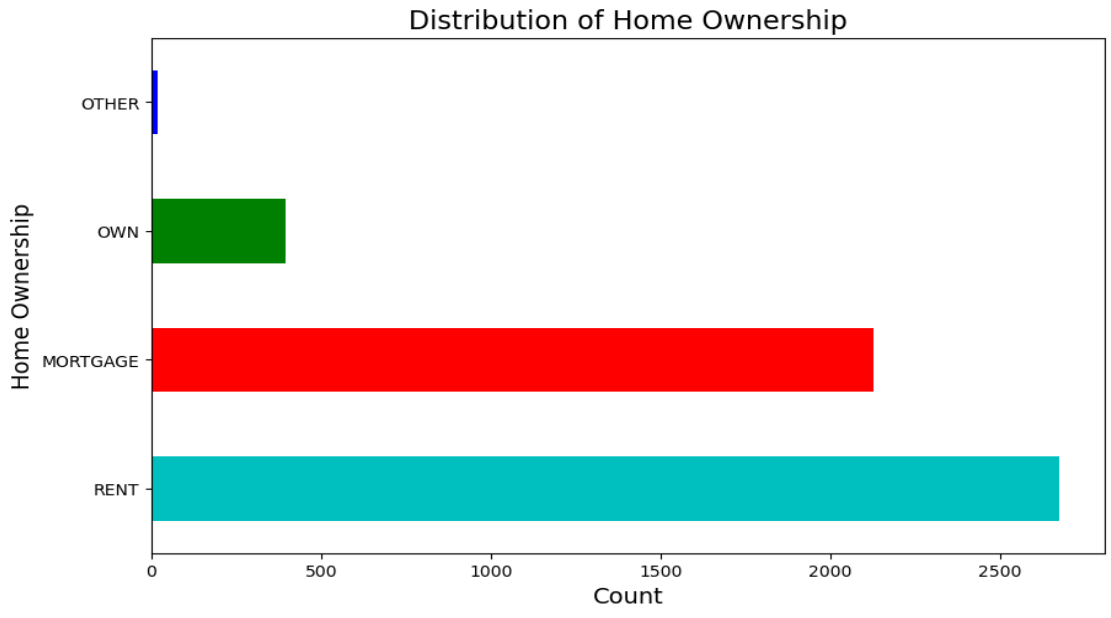
- Above chart indicates that the highest number of charged-off loans occur for individuals with 10+ years of employment.
- There is a relatively lower and more even distribution of charged-off loans for employment lengths from 0 to 9 years.

❖ **Inference**

- This histogram plot shows that charged-off loans are most frequent at interest rates between 10.73% and 15.29%.

EDA - (Univariate Analysis)





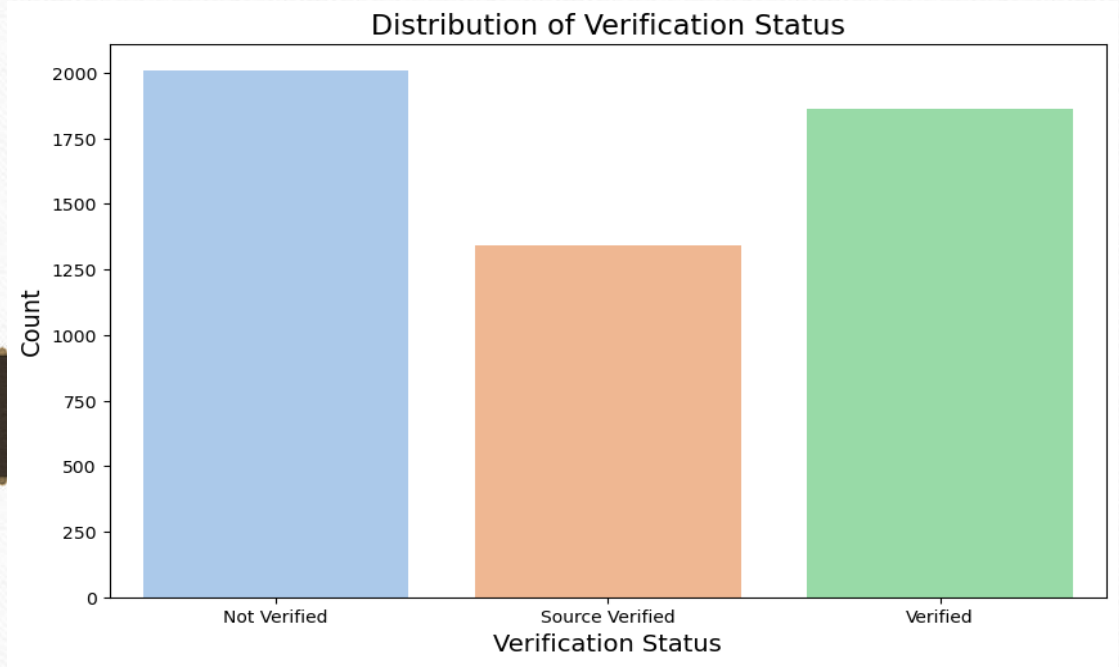
❖ **Inference**

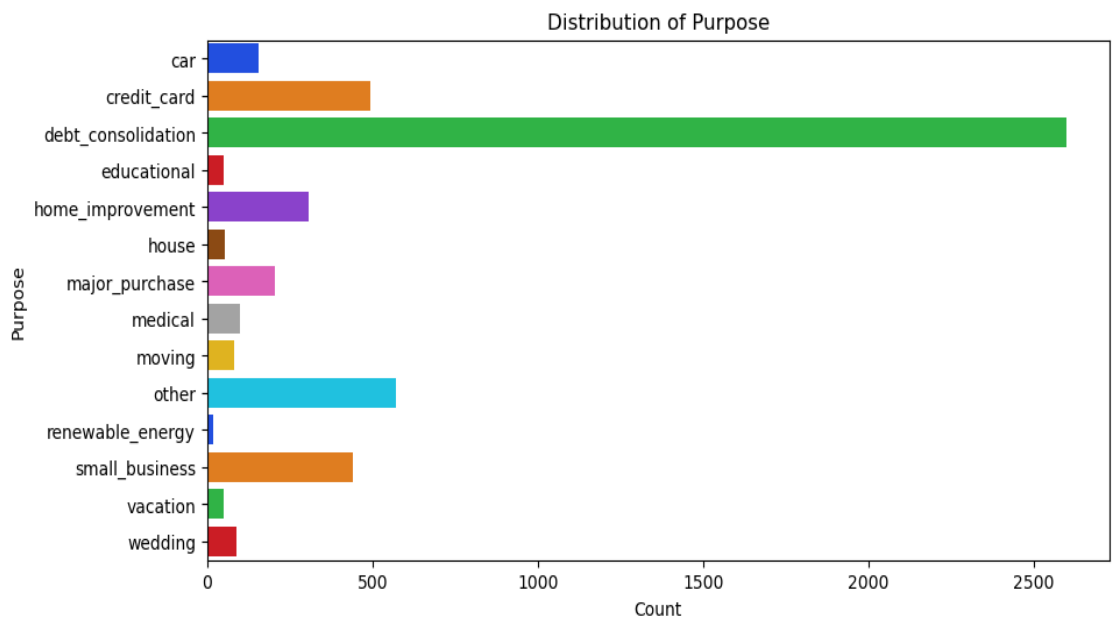
- Above plot indicates that the majority of these loans have a Debt-to-Income (DTI) ratios between 10 and 20, with the highest frequency around 15.
- This suggests that borrowers with a DTI ratio in this range are more likely to default on their loans. Managing DTI ratios could be a critical factor in reducing loan defaults..

❖ **Inference**

- This indicates that renters are at a higher risk of having charged-off loans compared to property owners and other types of ownership.

EDA - (Univariate Analysis)





❖ **Inference**

- This count plot indicates that loans that are not verified are at higher risk for charge-offs.

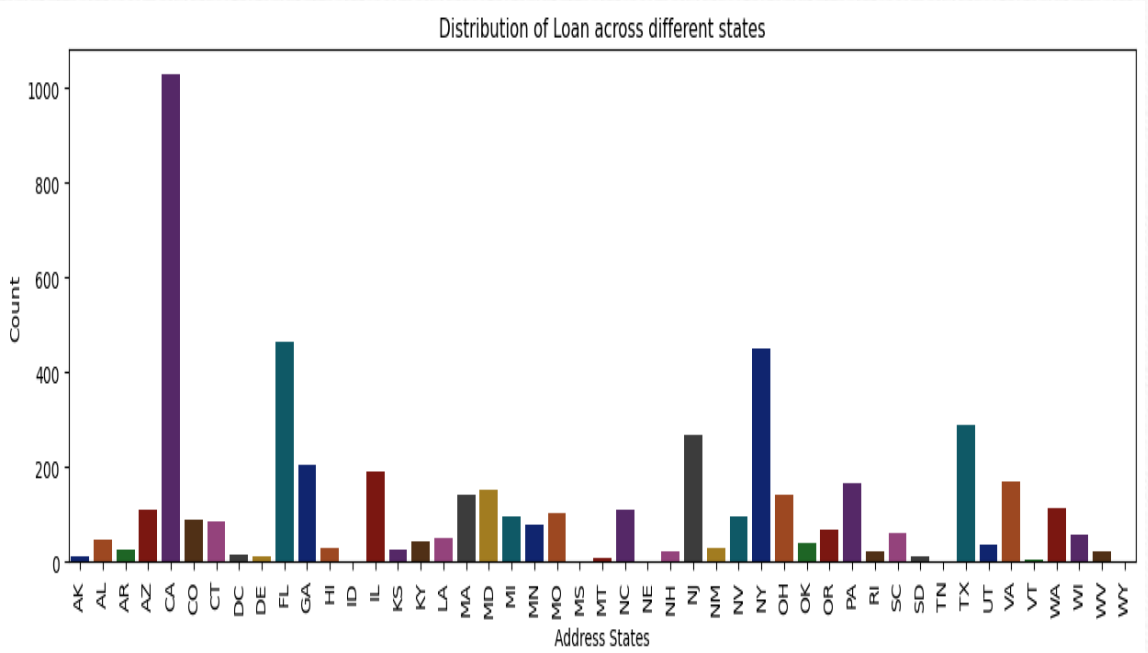
❖ **Observation**

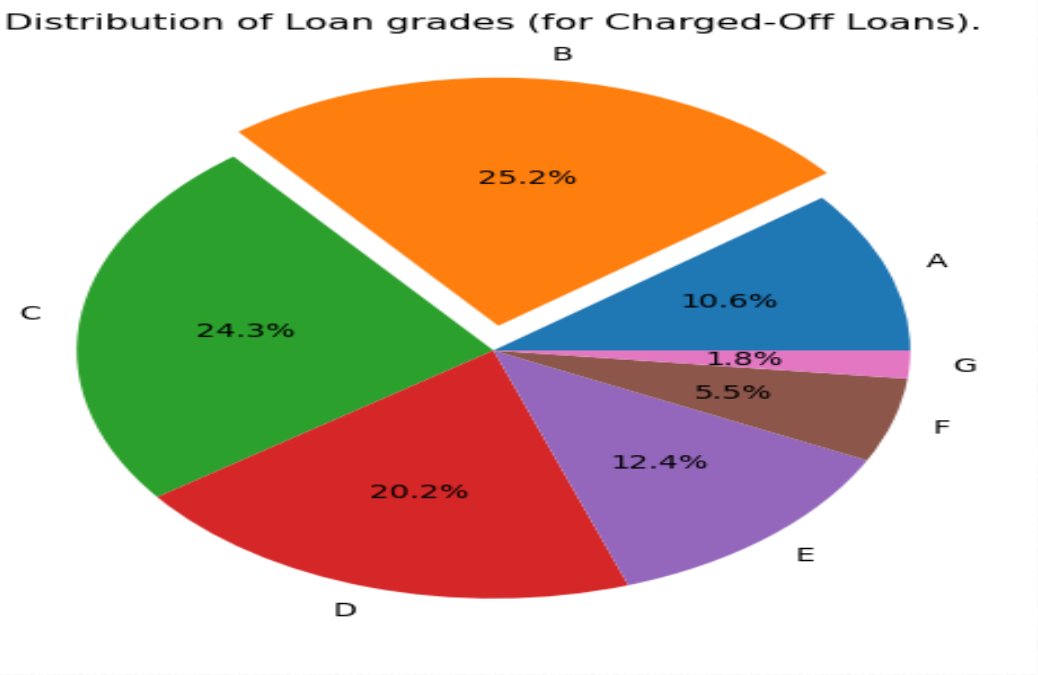
- Verified Loans also have a surprisingly high number of defaults.

❖ **Inference**

- Above bar chart shows that "Debt consolidation" has the highest number of charged-off loans.
- "Credit card" and "other" purposes also have significant number of charged-off loans. Categories like "renewable energy," "vacation," and "wedding" have the fewest charged-off loans.

EDA - (Univariate Analysis)





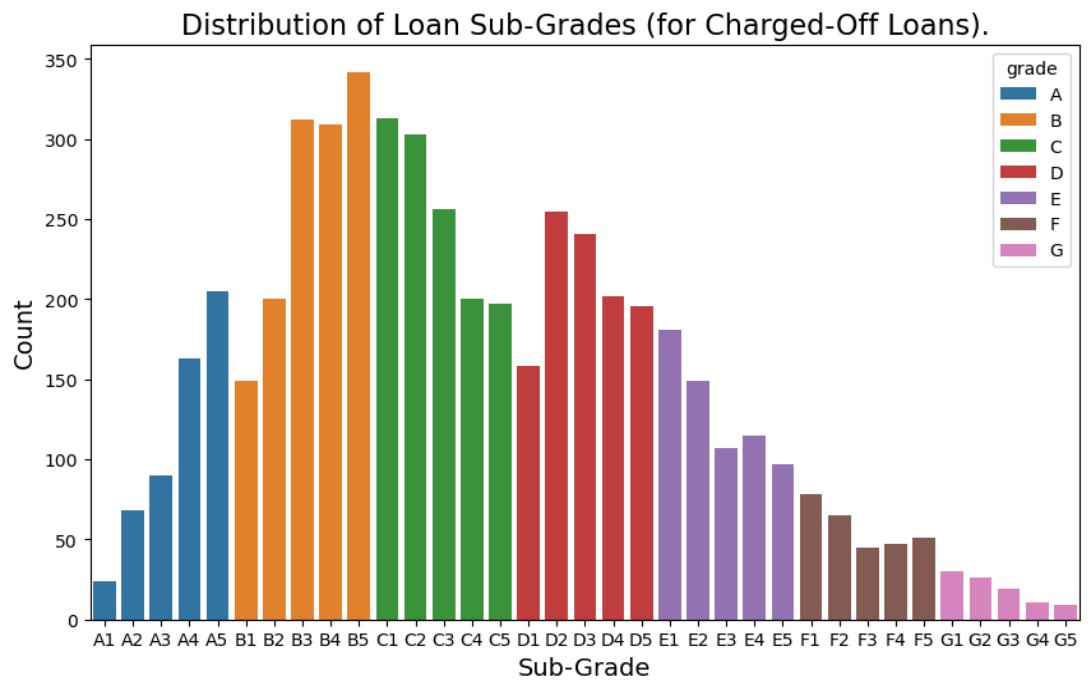
❖ **Inference**

- The bar chart shows that state CA has the most number of Charged-Off loans followed by NY and FL.

❖ **Inference**

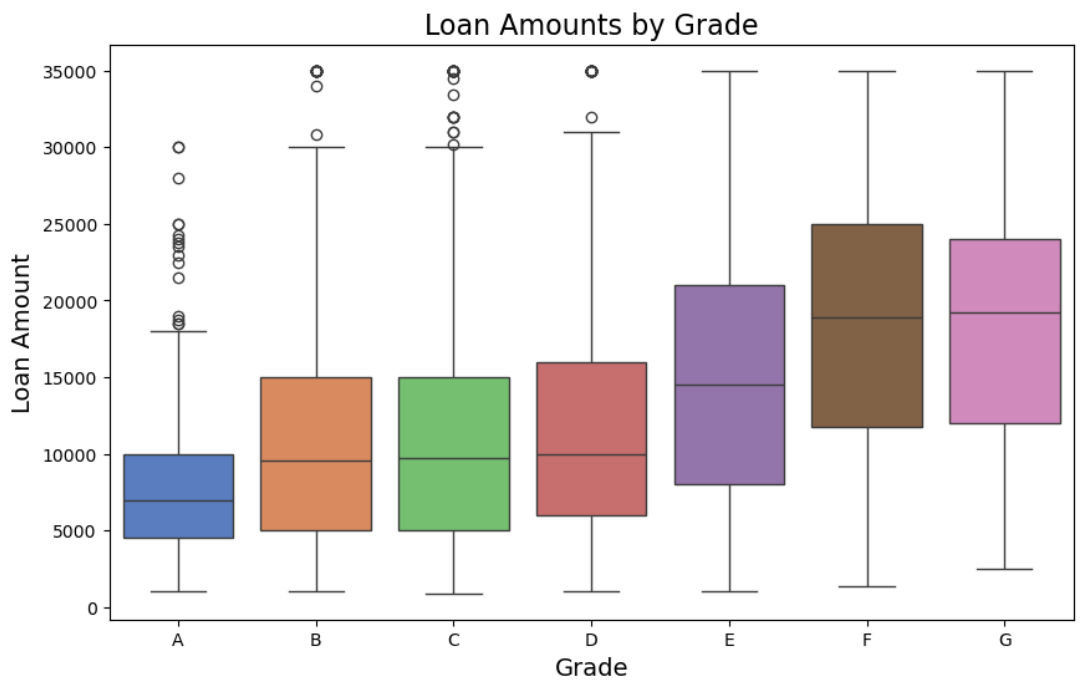
- The pie chart indicates that charged-off loans are most prevalent in grades B (25.2%) and C (24.3%). Combined these two grades constituting almost half of the charged-off loans, highlighting a higher risk in these categories.
- Grades F and G have the lowest percentage indicating fewer charged-off loans in these categories.

EDA - (Segmented Univariate Analysis)



❖ **Inference**

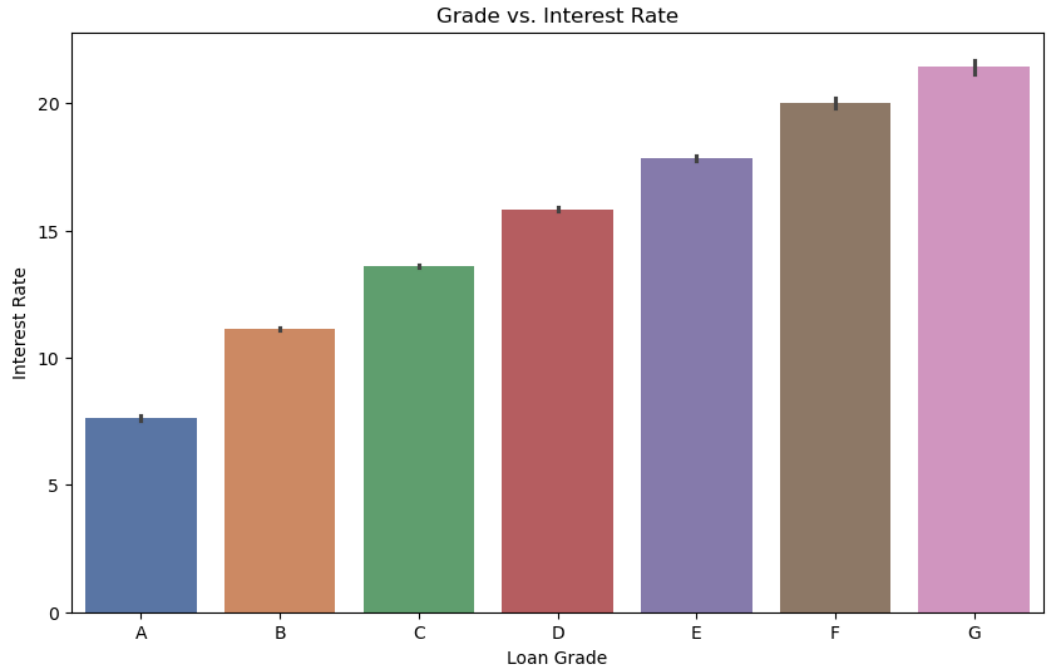
- The bar plot shows that the among highest B Grade shown in previous pie chart , B5 sub grade has the highest charged off loans.
- Most charged-off loans fall into sub-grades between B3 and C3.

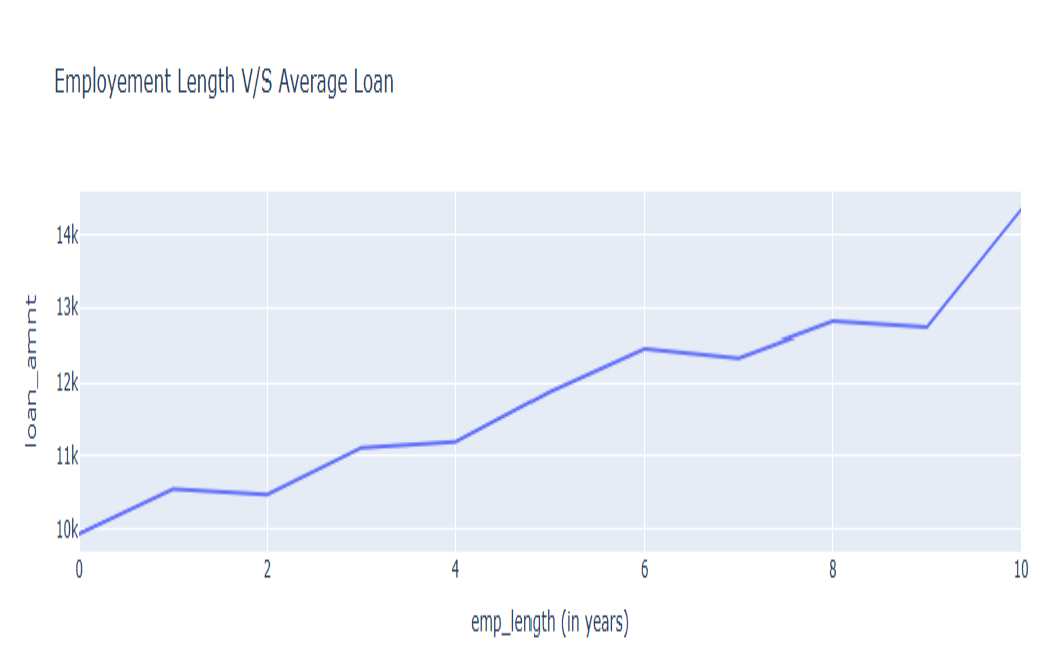


❖ **Inference**

- ***(Assumption: "A" being highest and "G" is Lowest)***
- The chart shows that loan amounts increase with lower grades, with borrowers in grades E, F, and G taking larger loans compared to those in grades A, B, and C.

EDA - (Bivariate Analysis)





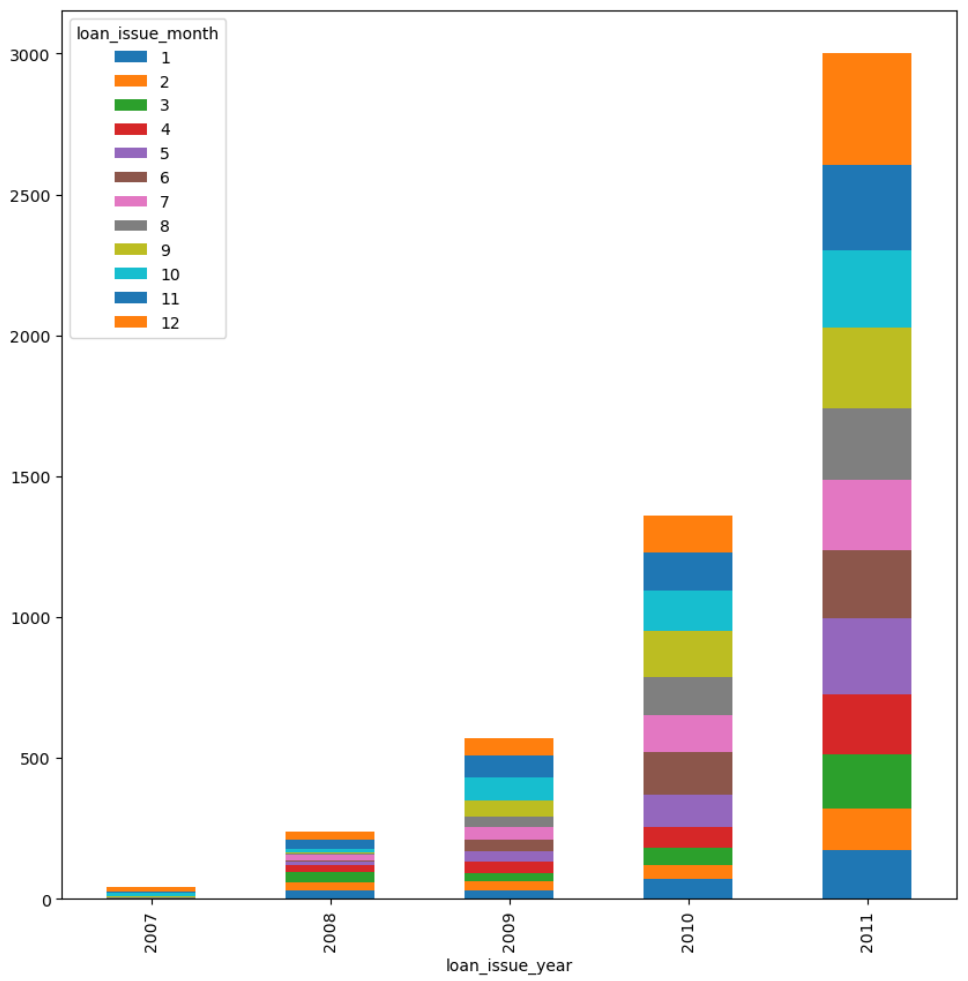
❖ **Inference**

- The bar chart shows that as the loan grade decreases from A to G, the interest rate increases.
- This suggests that lower loan grades are associated with higher interest rates, reflecting higher risk and loans with better grades have lower interest rates.

❖ **Inference**

- The line plot shows a positive correlation between employment length and the average loan amount. As the number of years of employment increases, the average loan amount also tends to increase.

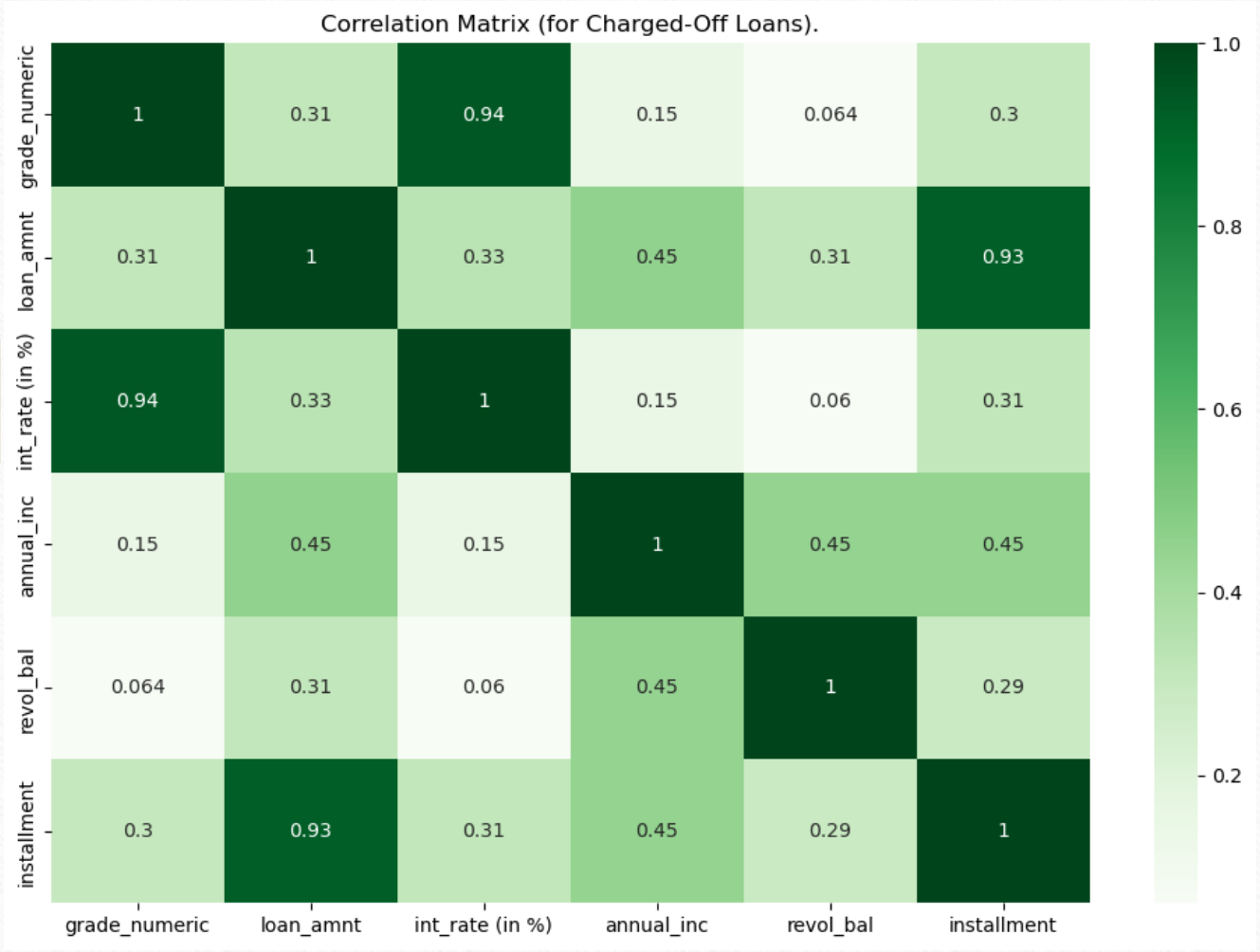
EDA - (Derived Metrics)



❖ **Inference**

- The bar chart shows the count of charged-off loans increasing each year from 2007 to 2011, peaking in 2011.
- Charged-off loans are spread evenly across all months, indicating a consistent issue throughout the year.

EDA - (Correlation Matrix)



❖ Inference

- The correlation matrix shows a **strong positive correlation between loan amount and installment (0.93)**, highlighting that larger loans result in higher monthly payments.
- **Grade_numeric has a strong correlation with interest rate (0.94),** indicating that lower credit grades are associated with higher interest rates.

Results

- **Key Findings from Analysis**

- **Loan Amount Distribution**: The distribution of loan amounts is right-skewed, with most loans concentrated in the 0 to 15000 range, peaking at 5000 to 10000. Higher loan amounts above 30000 are rare, indicating limited demand or approval for large loans.
- **Annual Income Distribution**: The distribution of annual income is right-skewed, with most defaults occurring among individuals earning between 20000 to 70000, peaking around 40000. Individuals with higher income levels above 100000 are less likely to default.
- **Employment Length**: The highest number of charged-off loans occur for individuals with 10+ years of employment. There is a relatively lower and more even distribution of charged-off loans for employment lengths from 0 to 9 years.
- **Interest Rates**: Charged-off loans are most frequent at interest rates between 10.73% and 15.29%.
- **Debt-to-Income (DTI) Ratios**: The majority of charged-off loans have DTI ratios between 10 and 20, with the highest frequency around 15. Managing DTI ratios could be critical in reducing loan defaults.
- **Home Ownership**: Renters are at a higher risk of having charged-off loans compared to property owners and other types of ownership.
- **Loan Verification**: Loans that are not verified are at higher risk for charge-offs. Surprisingly, verified loans also have a high number of defaults.
- **Loan Purpose**: "Debt consolidation" has the highest number of charged-off loans. "Credit card" and "other" purposes also have significant charged-off loans. Categories like "renewable energy," "vacation," and "wedding" have the fewest charged-off loans.

Results

- **Key Findings from Analysis**

- **State-wise Distribution**: The state of CA has the most charged-off loans, followed by NY and FL.
- **Loan Grades**: Charged-off loans are most prevalent in grades B (25.2%) and C (24.3%), highlighting higher risk in these categories. Grades F and G have the lowest percentage of charged-off loans.
- **Sub-Grade** Analysis: Among the B Grade, the B5 sub-grade has the highest charged-off loans. Most charged-off loans fall into sub-grades between B3 and C3.
- **Loan Amount by Grades**: Loan amounts increase with lower grades, with borrowers in grades E, F, and G taking larger loans compared to those in grades A, B, and C.
- **Interest Rates by Grades**: As loan grades decrease from A to G, the interest rate increases, reflecting higher risk associated with lower loan grades.
- **Employment Length and Loan Amount**: A positive correlation exists between employment length and average loan amount. As the number of years of employment increases, the average loan amount tends to increase.
- **Yearly Trends**: The count of charged-off loans increased each year from 2007 to 2011, peaking in 2011. Charged-off loans are spread evenly across all months, indicating a consistent issue throughout the year.
- **Correlation Matrix**: A strong positive correlation exists between loan amount and installment (0.93), indicating that larger loans result in higher monthly payments. Grade_numeric has a strong correlation with interest rate (0.94), indicating that lower credit grades are associated with higher interest rates

Conclusion

- **Recommendations**

- Focus on managing DTI ratios and interest rates to reduce loan defaults.
- Monitor applicants with 10+ years of employment more closely, as they have the highest number of charged-off loans. Implement additional verification for these individuals.
- Target interventions for high-risk groups, such as renters and those with certain loan purposes.
- Monitor trends and update risk models regularly to reflect changing patterns in loan defaults.
- Pay special attention to loans for "Debt consolidation," "Credit card," and "other" purposes, which have the highest number of charged-off loans. Develop specific strategies to mitigate risks for these purposes.
- Concentrate on states like CA, NY, and FL, which have the highest number of charged-off loans. Plan risk management strategies to these regions.
- Strengthen verification processes to reduce the higher risk of charge-offs associated with unverified loans, while still closely monitoring verified loans.