Lending Club Case Study

Date: June 26 ‘2024

Team Members

- Anil Kumar Narayanan
- Sharwan Kumar

Problem Statement

- **Challenges in Loan Default Management**
  - Consumer finance companies face significant risks associated with loan defaults.
  - Deciding whether to approve or deny loans involves balancing potential loss of business versus financial risk.
- **Impact of Loan Defaults**
  - **Loss of Revenue**: Denying loans to potentially creditworthy applicants can lead to missed business opportunities.
  - **Financial Loss**: Approving loans to high-risk borrowers increases the risk of default, leading to financial losses.

Objective

1. **Identify Patterns**: Uncover key factors influencing loan default cases.
2. **Explore Data**: Analyze borrower profiles, loan attributes, and historical trends to understand risk drivers.
3. **Inform Loan Approvals**: Offer insights into which borrower characteristics and loan types pose higher risks.
4. **Risk Management:** Recommend strategies to minimize default rates while maximizing loan approval rates.
5. **Reduce Financial Loss**: Mitigate losses associated with defaulted loans by proactively identifying high-risk applicants.
6. **Enhance Profitability**: Ensure loans are approved with a balanced approach to business growth and risk exposure.

Data Overview

- **Dataset:** The dataset contains loan data from 2007 to 2011, detailing loan status (fully paid, current, charged-off), borrower attributes, and loan characteristics.
- **Data Dictionary:** Provides definitions and explanations of variables used in the dataset, aiding in understanding each feature's relevance to loan default risk.

Data Preprocessing

Cleaning and Preparation

Handling Missing Values

Step	Methodology	Result
Check for Completely Empty Rows	Identified and removed entirely empty rows.	No empty rows found.
Identify Missing Values	Detected missing values in columns and rows.	Identified and dropped rows, columns with more than 30% missing values. Dataset Dimensionality: 36432 rows x 45 columns
Identify and Remove Duplicate Records	Found and removed duplicate records.	No duplicate records found.
Handle anomalous data entries	Home Ownership Field: Tagging "NONE" to "OTHER".	To maintain categorical data consistency, updated 3 rows to "OTHER" in home_ownership field.

Data Preparation

Step	Methodology	Result
Convert 'term' Column to Integer	loan_ds["term"] = loan_ds["term"].str.replace(" months", "").astype("int64")	'term' column now contains integer values.
Convert Percentage Columns to Float	loan_ds['int_rate'] = loan_ds['int_rate'].str.rstrip('%').astype('float64').round(2) loan_ds['revol_util'] = loan_ds['revol_util'].str.rstrip('%').astype('float64').round(2) loan_ds['funded_amnt_inv'] = loan_ds['funded_amnt_inv'].round(2)	Percentage columns converted to float and rounded to two decimal places.
Round Specific Columns to Two Decimal Places	Rounded specified columns to two decimal places.	Columns rounded: 'total_pymnt', 'total_pymnt_inv', 'total_rec_prncp', 'total_rec_int', 'total_rec_late_fee', 'recoveries', 'collection_recovery_fee', 'last_pymnt_amnt'.
Clean 'emp_length' Column	loan_ds["emp_length"] = loan_ds["emp_length"].str.replace(r'\+ years', '', regex=True).str.replace(r' years', '', regex=True).str.replace(r' year', '', regex=True).str.replace(r'< 1', '0').str.strip()	'emp_length' column cleaned.
Split and Convert Date Columns	• Split 'issue_d' and other date columns into month and year parts. • Mapped month names to numeric values. • Converted short year format to full year format.	Date columns split and converted: 'issue_d', 'earliest_cr_line', 'last_pymnt_d', 'last_credit_pull_d'.
Tag 'NONE' to 'OTHER' in 'home_ownership' Field	loan_ds['home_ownership'] = loan_ds['home_ownership'].replace('NONE', 'OTHER')	'home_ownership' field cleaned.
Rename Columns	Renamed columns to represent revised values.	Columns renamed: 'term' to 'term_in_months', 'revol_util' to 'revol_util_in_percent', 'emp_length' to 'emp_length_in_years'.

Data Analysis

Univariate Analysis

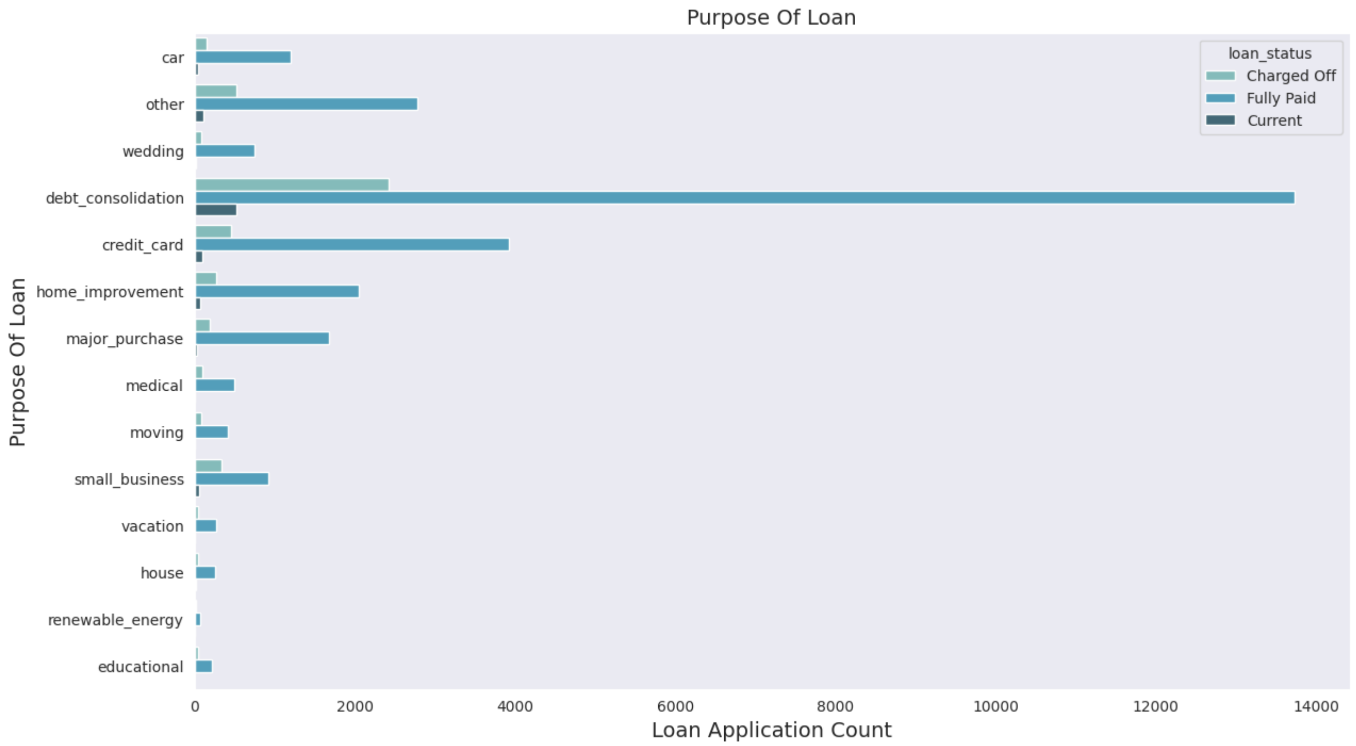
Risk Assessment 1: Purpose of Loan

High Volume of Debt Consolidation and Credit Card Loans:

- The majority of loans taken are for debt consolidation and paying credit card bills. Such borrowers may already have significant existing debt and might be under financial stress, relying on new loans to manage their debt load.

High Charge-Off Rates for Debt Consolidation and Credit Card Loans:

- The high charge-off rates indicate a higher risk of default for loans taken out for debt consolidation and credit card repayment. Borrowers in these categories might struggle to manage their overall debt levels, leading to increased likelihood of default. This can be due to various reasons, such as continued financial strain, lack of effective debt management strategies, or insufficient income to cover the new loan payments alongside existing obligations.



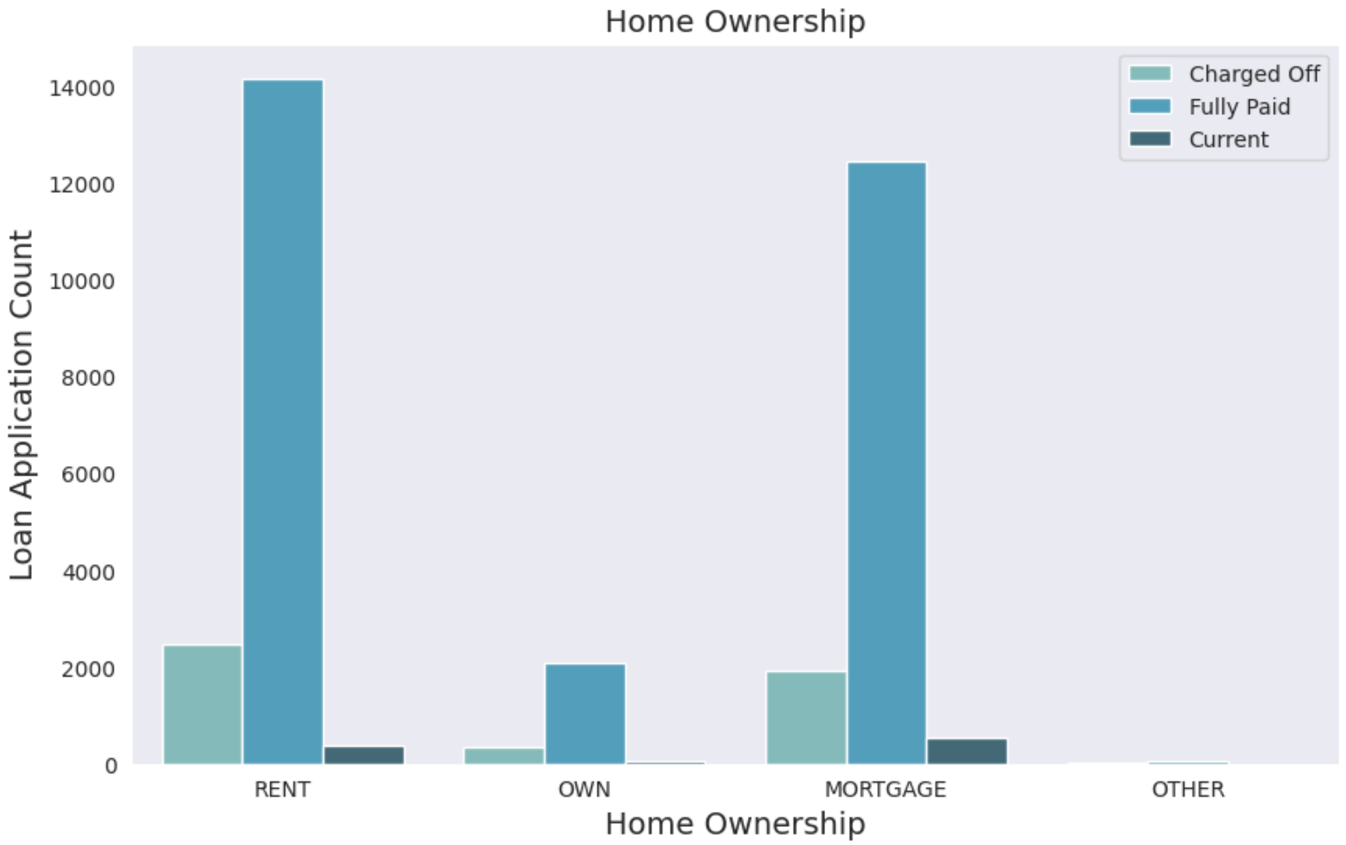
Risk Assessment 2: Home Ownership

High Volume of Renters and Mortgaged Homeowners:

- Most loan applicants are either renting their homes or have mortgaged homes. Borrowers with mortgages may have higher financial obligations or less financial stability compared to those who own their homes.

High Charge-Off Rates for Renters and Mortgaged Homeowners:

- There is a high number of charged-off loans among these categories. This suggests that renters and mortgaged homeowners may be more prone to financial distress, leading to a higher likelihood of default.



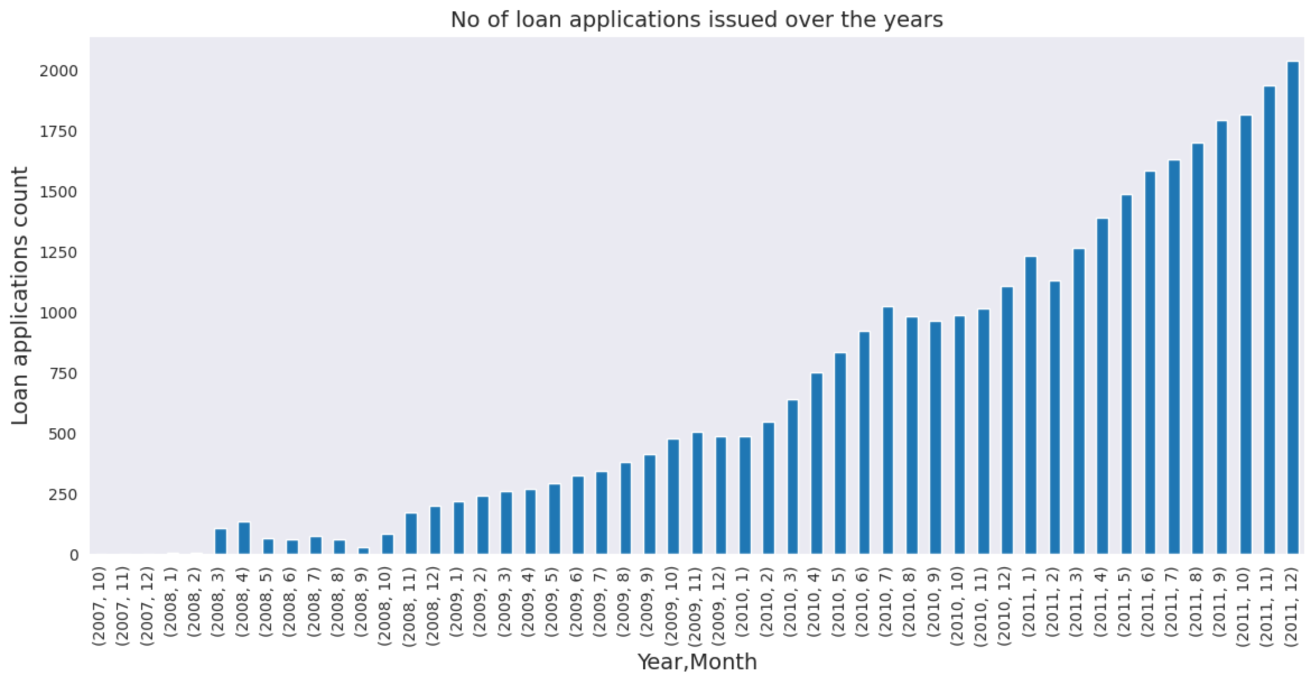
Risk Assessment 3: Number of Loan Application Issued

Increasing Loan Applications:

- The count of loan applications is increasing each year. As the number rises, there is an increase in the number of charged-off applications. This indicates a growing volume of lending, which, without proper risk management, could lead to a higher number of defaults.

Dip in Loans Issued in 2008:

- There was a dip in the number of loans issued between May and October 2008. The dip corresponds with the period of the global financial crisis, impacting loan issuance rates.

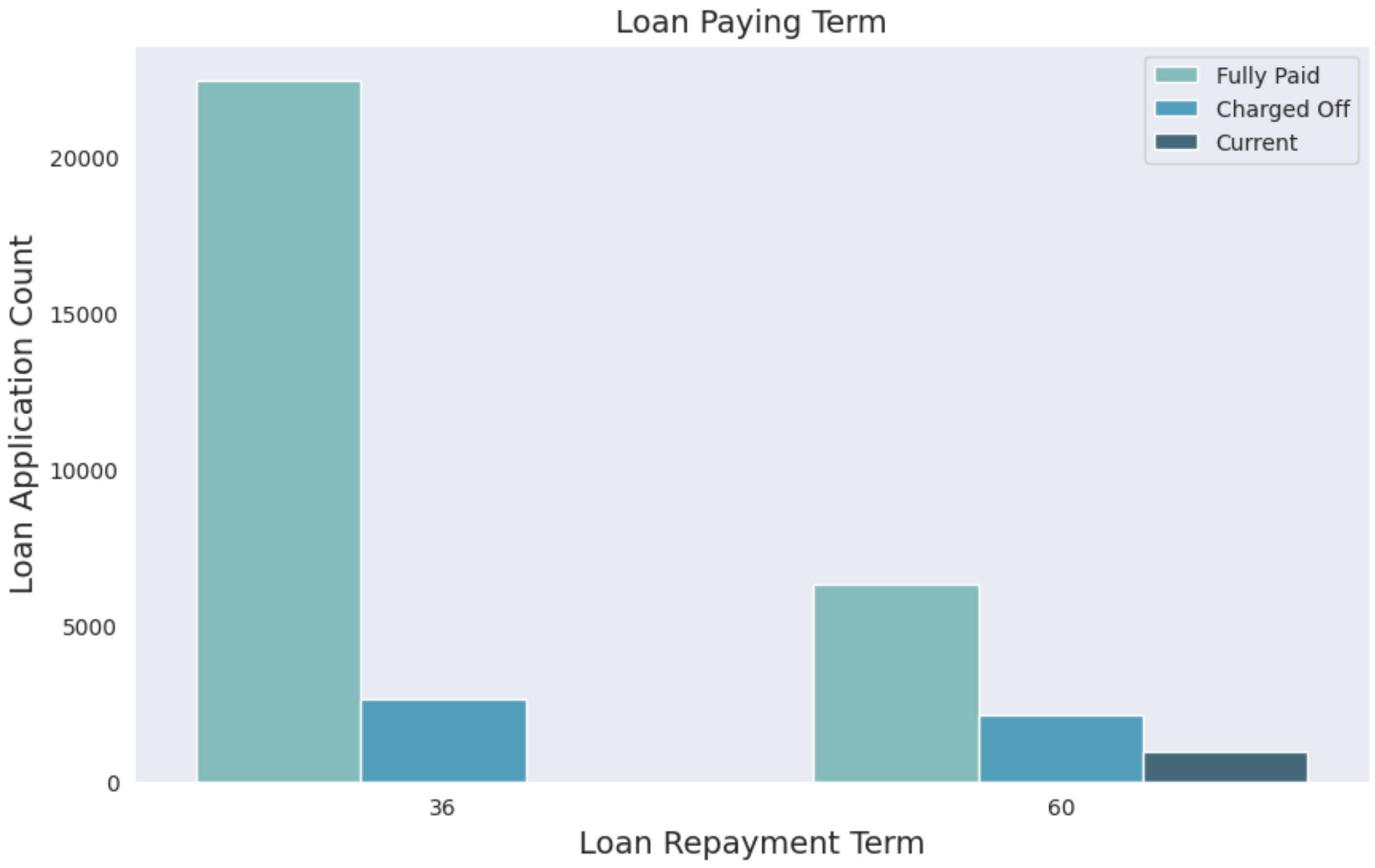


Risk Assessment 4: Loan Repayment Term

Higher Default Rates for longer duration loans:

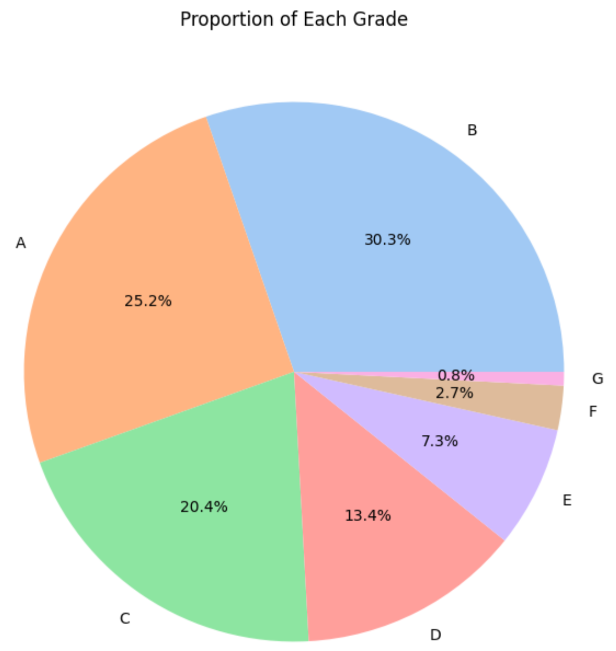
- The plot indicates that borrowers opting for 60-month loan terms have a higher percentage of defaults compared to those with 36-month loan terms.

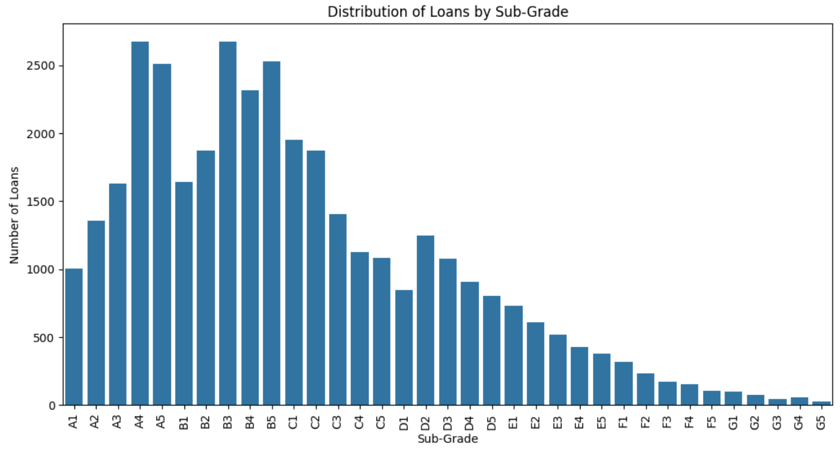
- This shows that longer the term, the higher is the probability that borrowers might face unforeseen financial challenges, leading to higher default rates.



Risk Assessment 5: Grade/Sub-Grade

- Loan grades are indicative of the borrower's creditworthiness and help assess the risk level associated with lending to different borrowers.
- In the frequency distribution of each grade. It is noticed that most of the loans are granted to the A and B grades totalling approximately 56%.
- A and B Grades represent borrowers with higher creditworthiness. These grades typically have lower interest rates and are considered lower risk.
- C and D Grades represent average creditworthiness. These grades have moderate interest rates and carry moderate risk.
- E, F, and G Grades: Represent lower creditworthiness. These grades have higher interest rates and are considered higher risk.





Bivariate Analysis

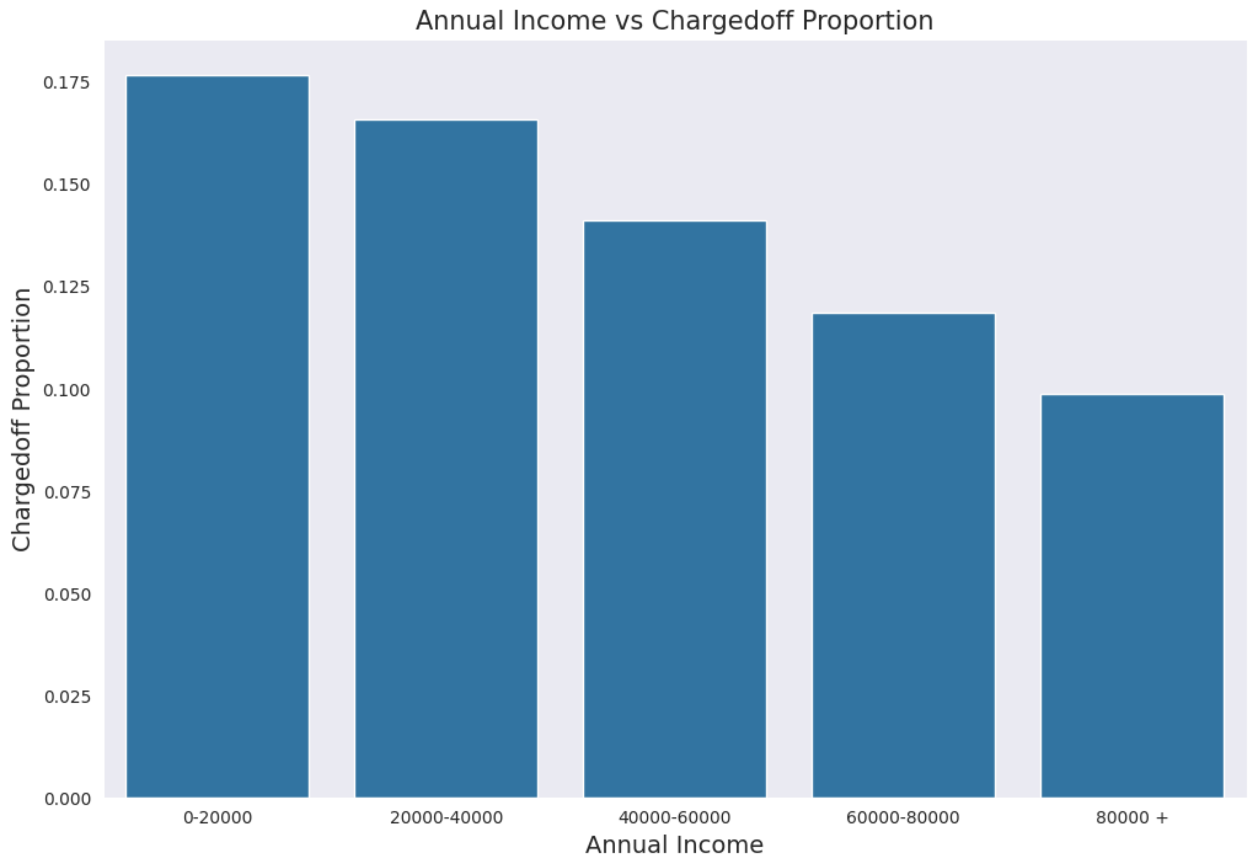
Risk Assessment 6:
Annual Income v/s Charged Off Proportion

Higher Income, Lower Default Rates:

- Borrowers with an annual income range of $80K and above have lower chances of being charged off. Higher income borrowers are less likely to default on their loans, suggesting better financial stability and lower risk

- Borrowers with an annual income range of $0 to $20K have higher chances of being charged off as they have a higher likelihood of defaulting, indicating higher financial distress.

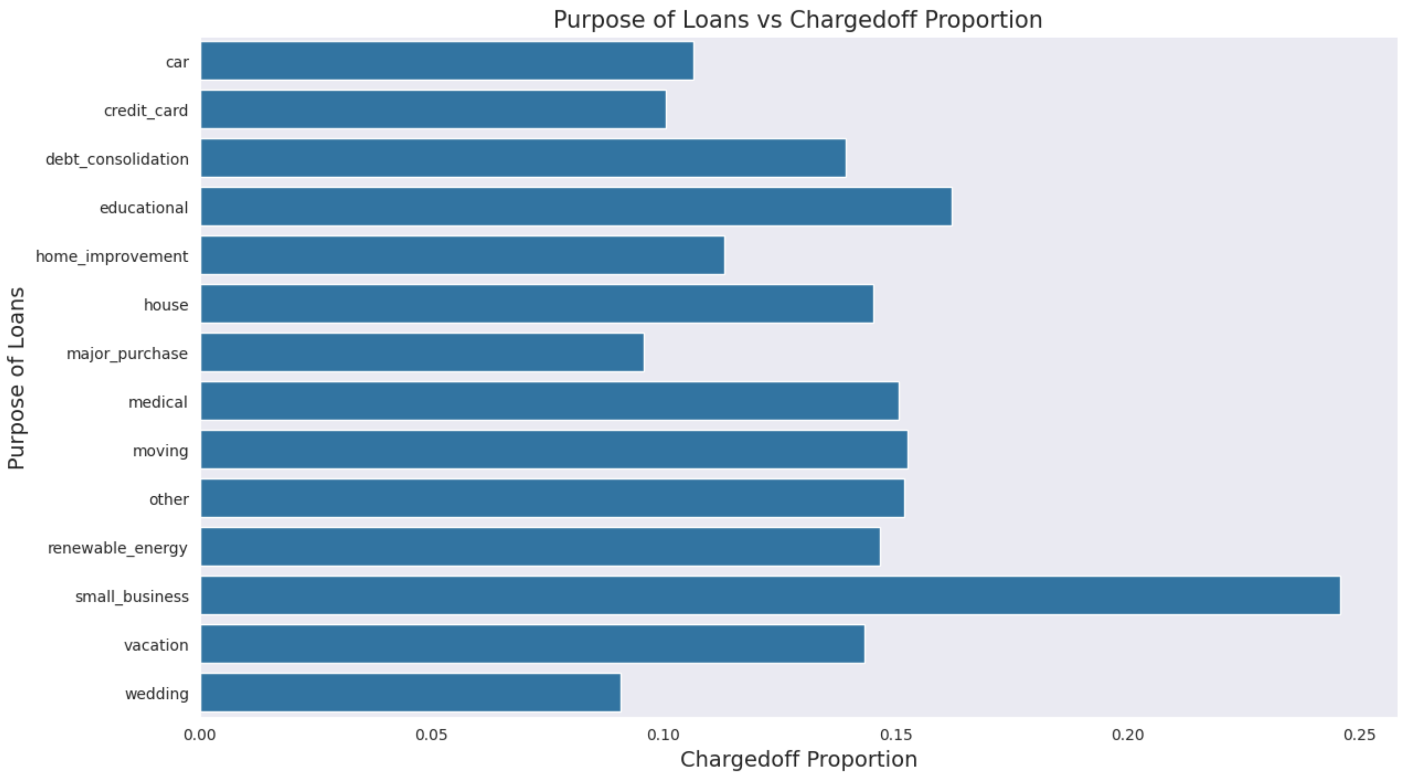
- As annual income increases, the proportion of charged-off loans decrease. This relationship suggests that income is a significant predictor of loan repayment ability.



Risk Assessment 7:
Purpose v/s Charged Off Proportion

- Small business loan applicants have high chances of getting charged off as they tend to have higher default rates, indicating a higher risk for lenders. This could be due to the volatile nature of small businesses and their revenue streams.

- Renewable energy loans have a better (lower) charged-off proportion compared to other categories which could be due to government incentives, steady income streams, or the borrowers' financial stability.



Risk Assessment 8:
Grades v/s Charged Off Proportion

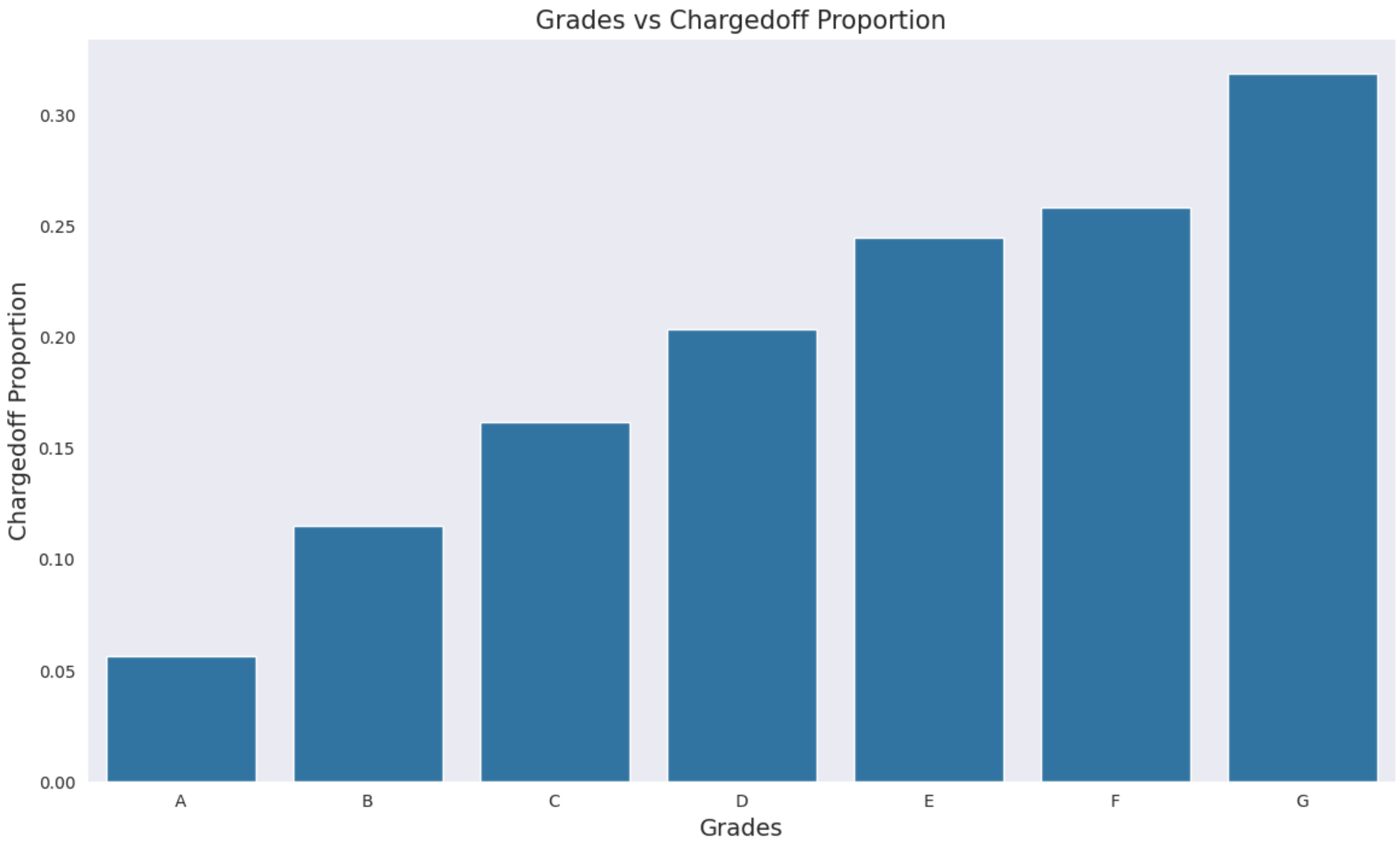
- Grade "A" Loans:

  These loans have very low chance of being charged off as they are of low risk, reflecting high creditworthiness and financial stability of the borrowers. These loans have a very low probability of default.

- Grades "F" and "G" Loans:

  Loans assigned Grades "F" and "G" have very high chances of being charged off as they are considered high risk, indicating poor creditworthiness and financial instability of the borrowers. These loans have a high probability of default.

The likelihood of a loan being charged off increases progressively as the grade moves from "A" towards "G".



Risk Assessment 9: Sub-Grades v/s Charged Off Proportion

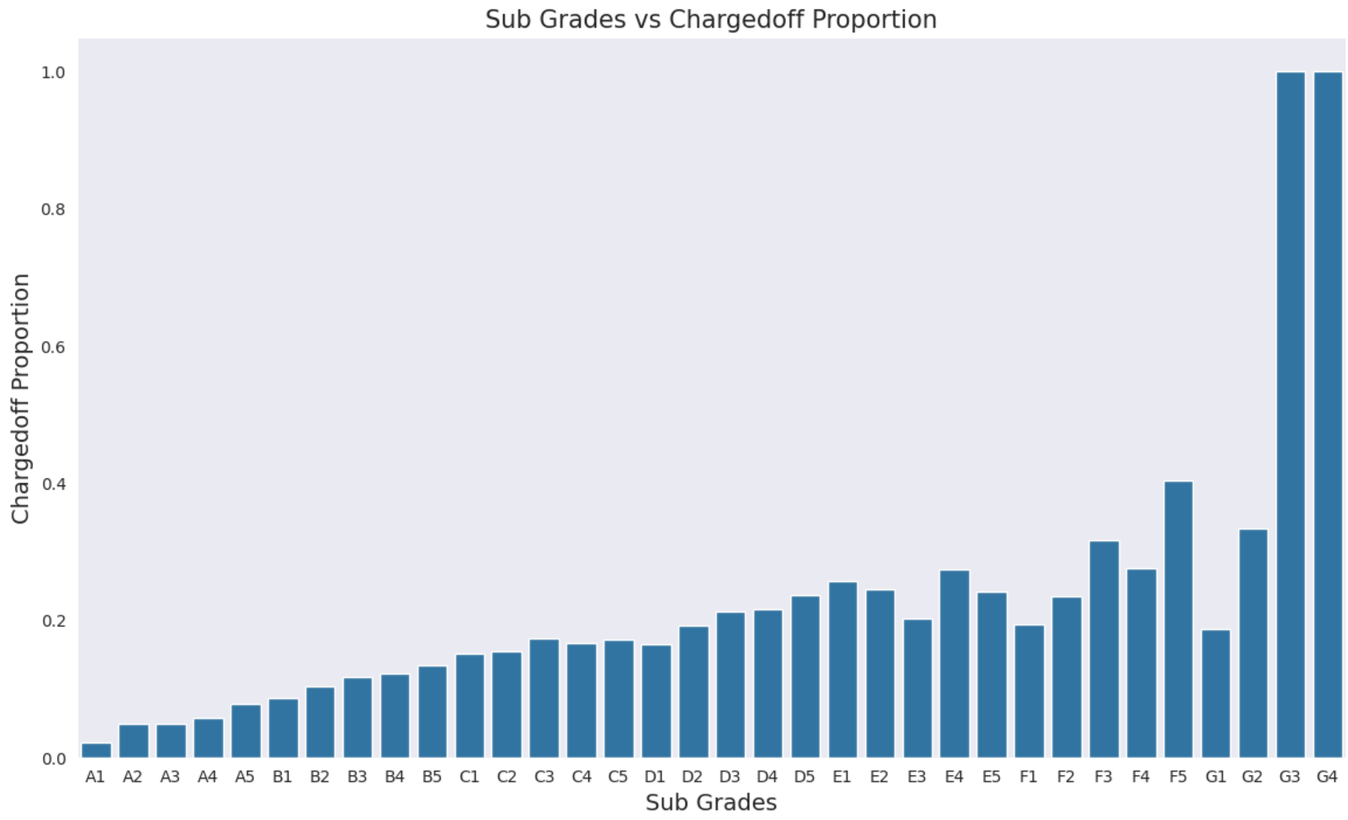
Subgrades of "A":

- Subgrades within the "A" category have very low probability of being charged off as borrowers with excellent creditworthiness and financial stability fall in this category resulting in a very low probability of default.

Subgrades of "F" and "G":

- Subgrades within the "F" and "G" categories have very high probability of being charged off as borrowers with poor credit histories and higher financial risks fall here leading to a significantly higher probability of default.

There is an increase in the proportion of charged-off loans as subgrades move from "A" towards "G" highlighting increase in credit risk and financial instability among borrowers, resulting in higher default rates.



Risk Assessment 10: Interest Rate v/s Charged Off Proportion

Interest Rates Less Than 10%:

- Loans with interest rates below 10% have very low chances of being charged off as these borrowers have better credit profiles and financial stability, leading to lower default probabilities.

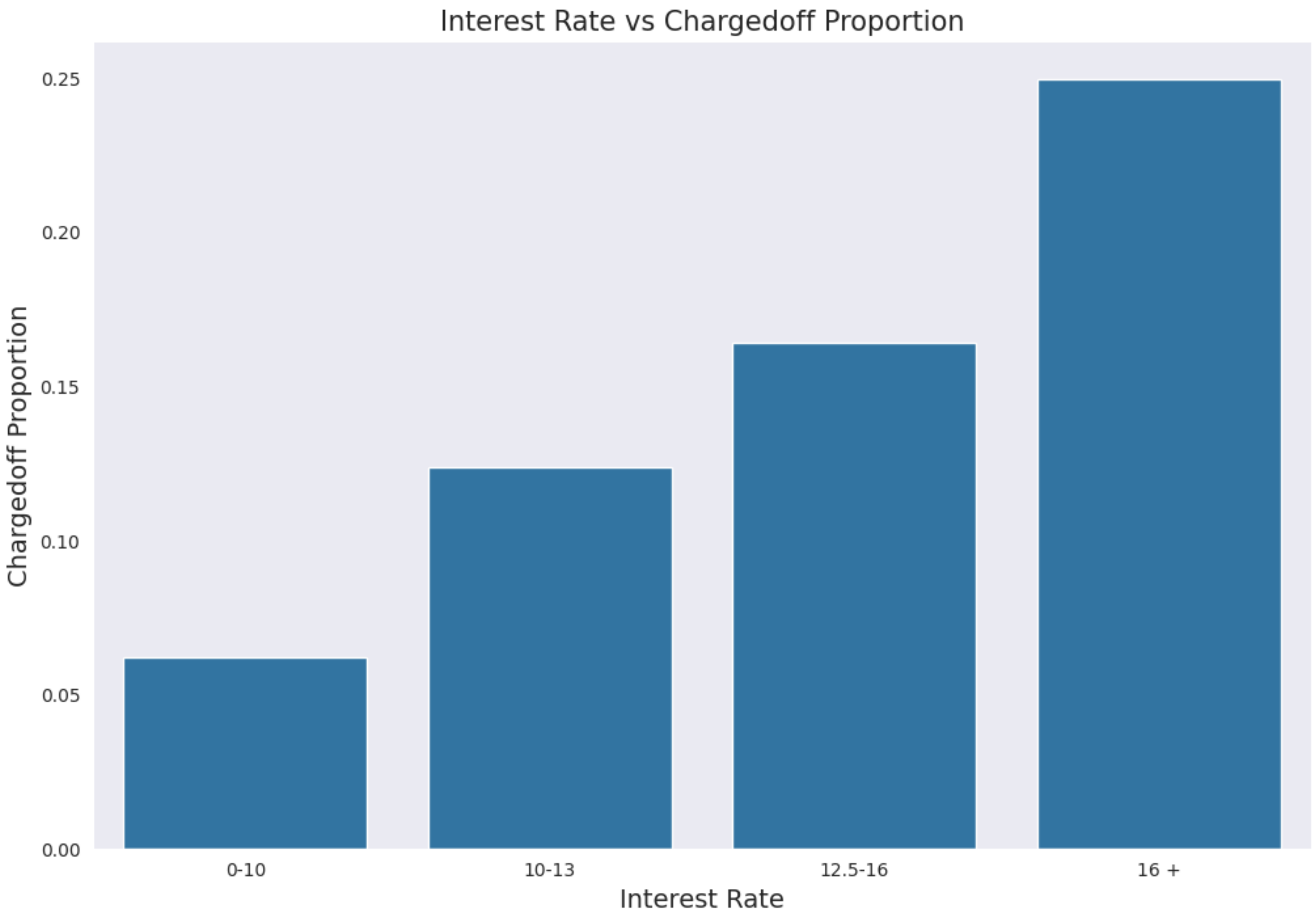
Interest Rates Between 10% and 16%:

- Loans with interest rates between 10% and 16% have moderate default rates as borrowers here have varying creditworthiness and financial stability, leading to moderate default risks.

Interest Rates More Than 16%:

- Loans with interest rates exceeding 16% show higher chances of being charged off as its associated with higher-risk borrowers or riskier loan terms, increasing the likelihood of default.

There is a clear trend where the proportion of charged-off loans increases with higher interest rates due to borrowers poor credit histories or higher debt-to-income ratios.



Risk Assessment 11:
Public Bankruptcy Records v/s Charged Off Proportion

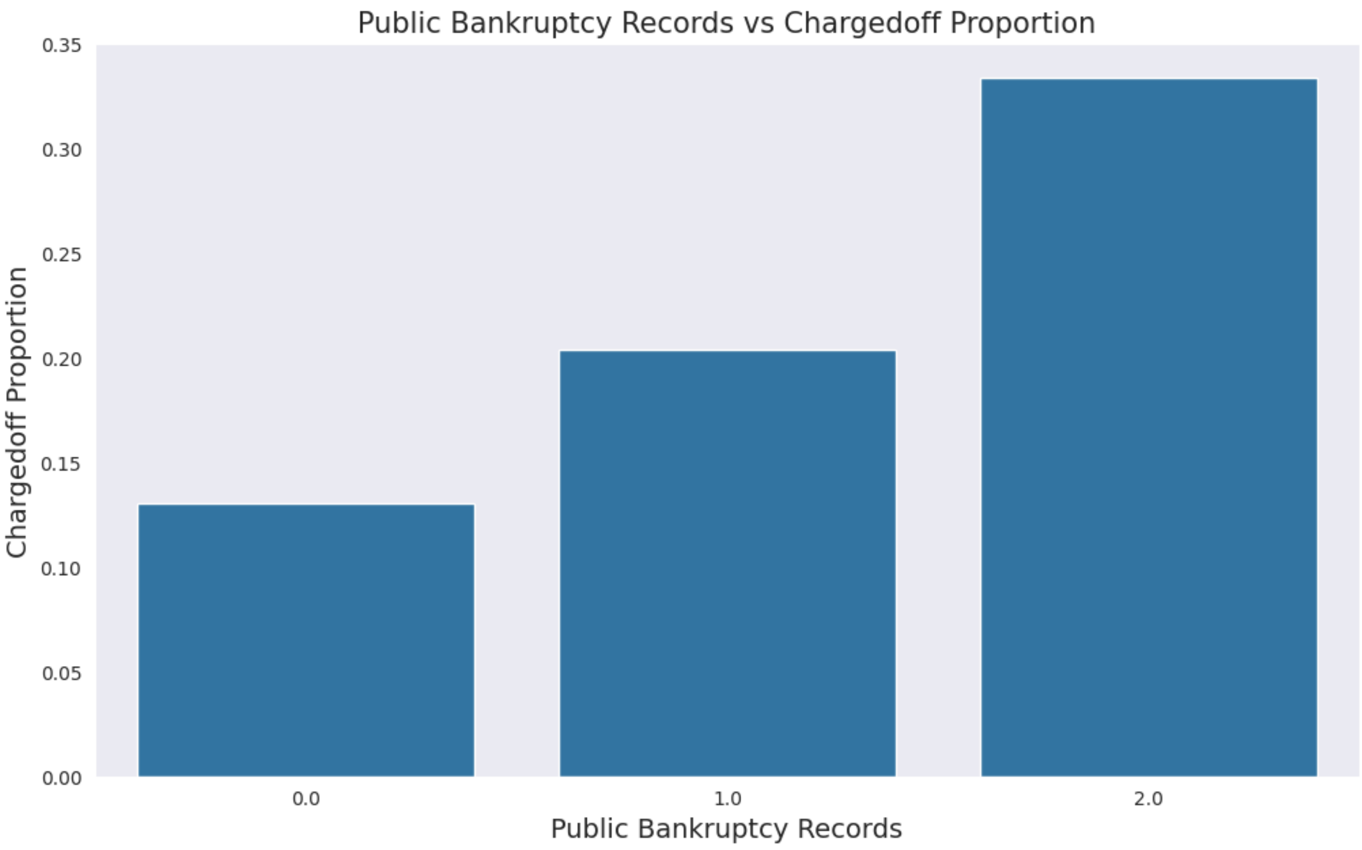
Borrowers with Public Record Bankruptcies:

- Those with value of 1 and above have a higher proportion of charged-off loans compared to those with no bankruptcies. This indicates financial distress and a higher likelihood of future payment defaults.

Unknown Values:

- While the exact reasons for unknown values are unclear, indicating potential incomplete data.

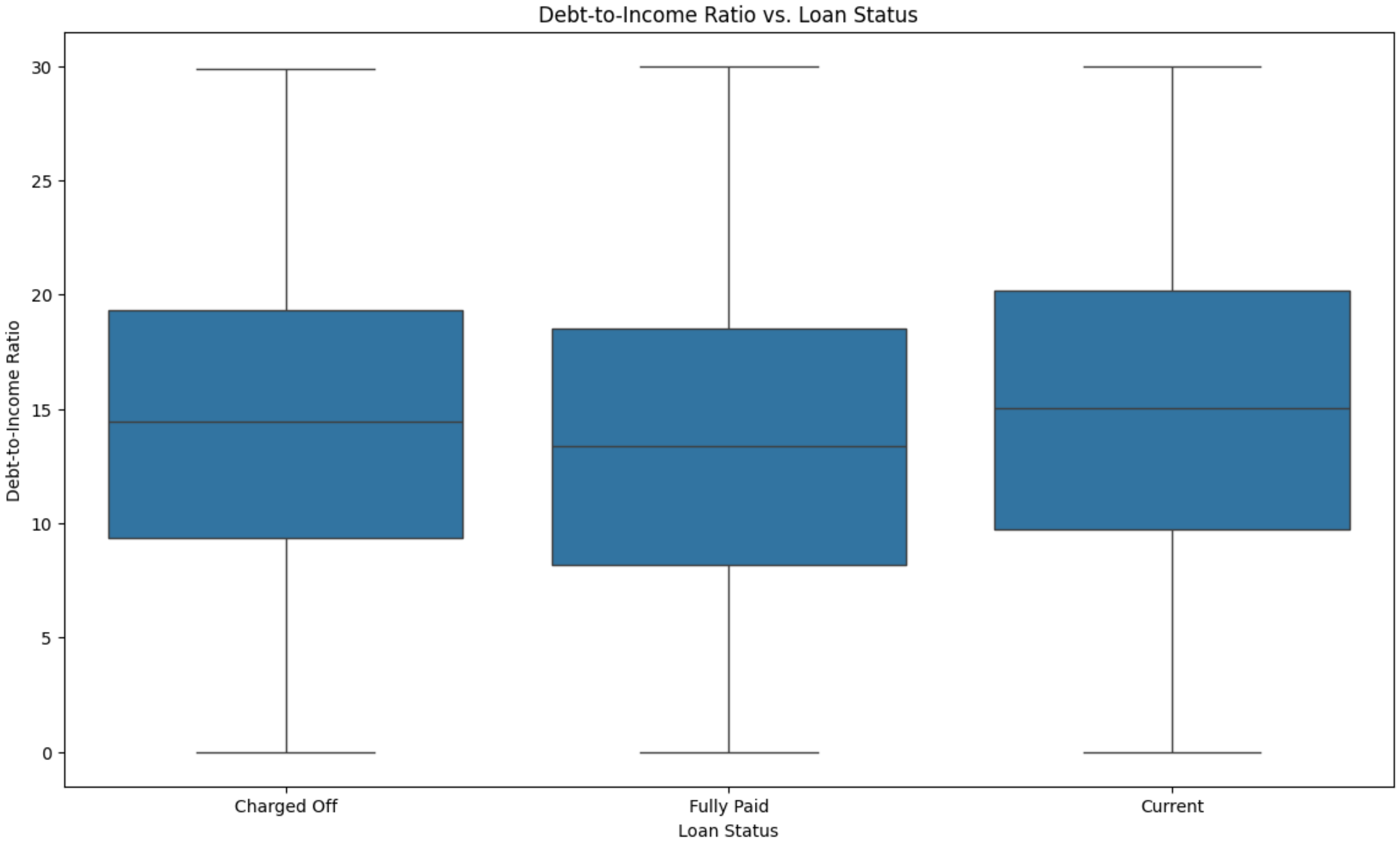
Borrowers with a history of past bankruptcies are more likely to default again.



Risk Assessment 12: Debt-to-Income Ratio vs Loan Status

- Borrowers with higher DTI ratios are more likely to face challenges in managing their debt obligations, leading to potential loan defaults

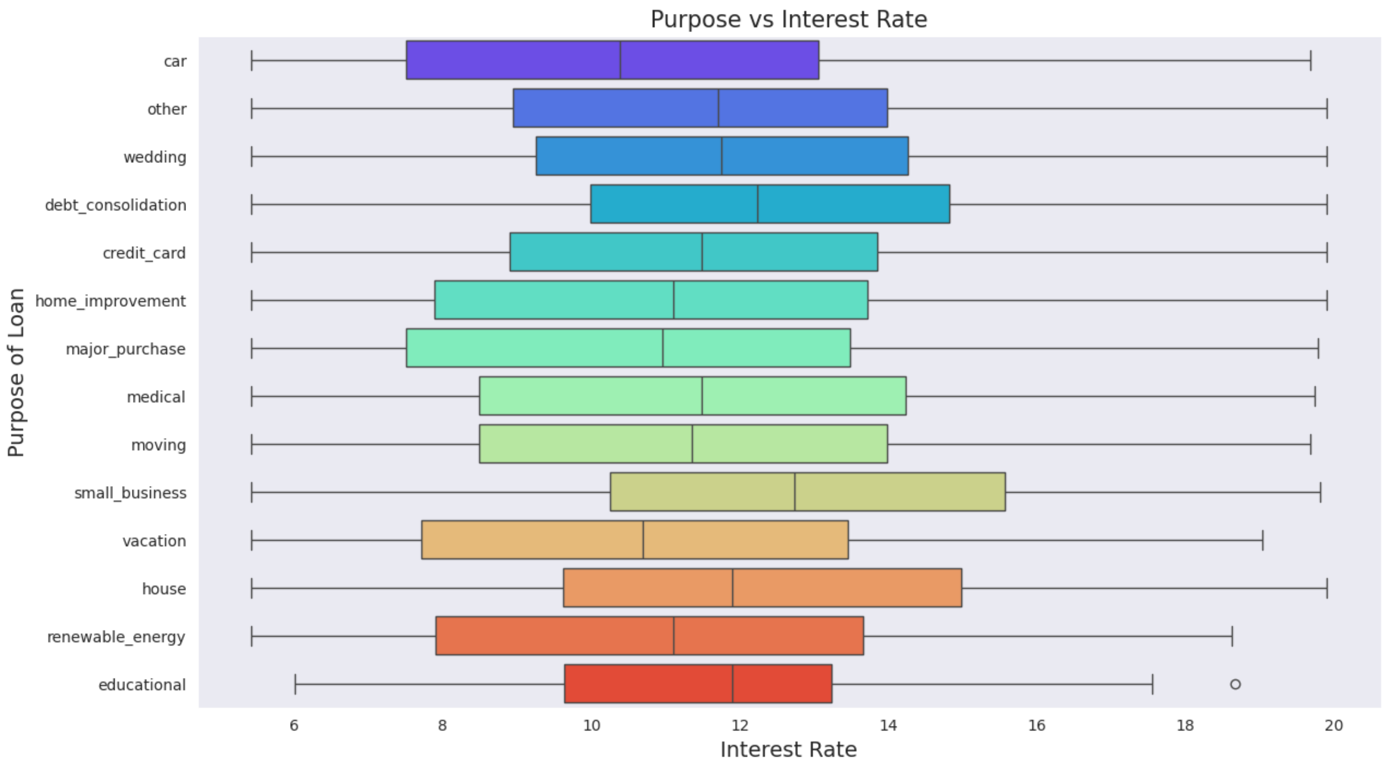
- The box plot distribution highlights that borrowers with higher DTI ratios are more likely to have their loans charged off, indicating a higher occurrence of default compared to borrowers with lower ratios.



Risk Assessment 13: Purpose v/s Interest Rate

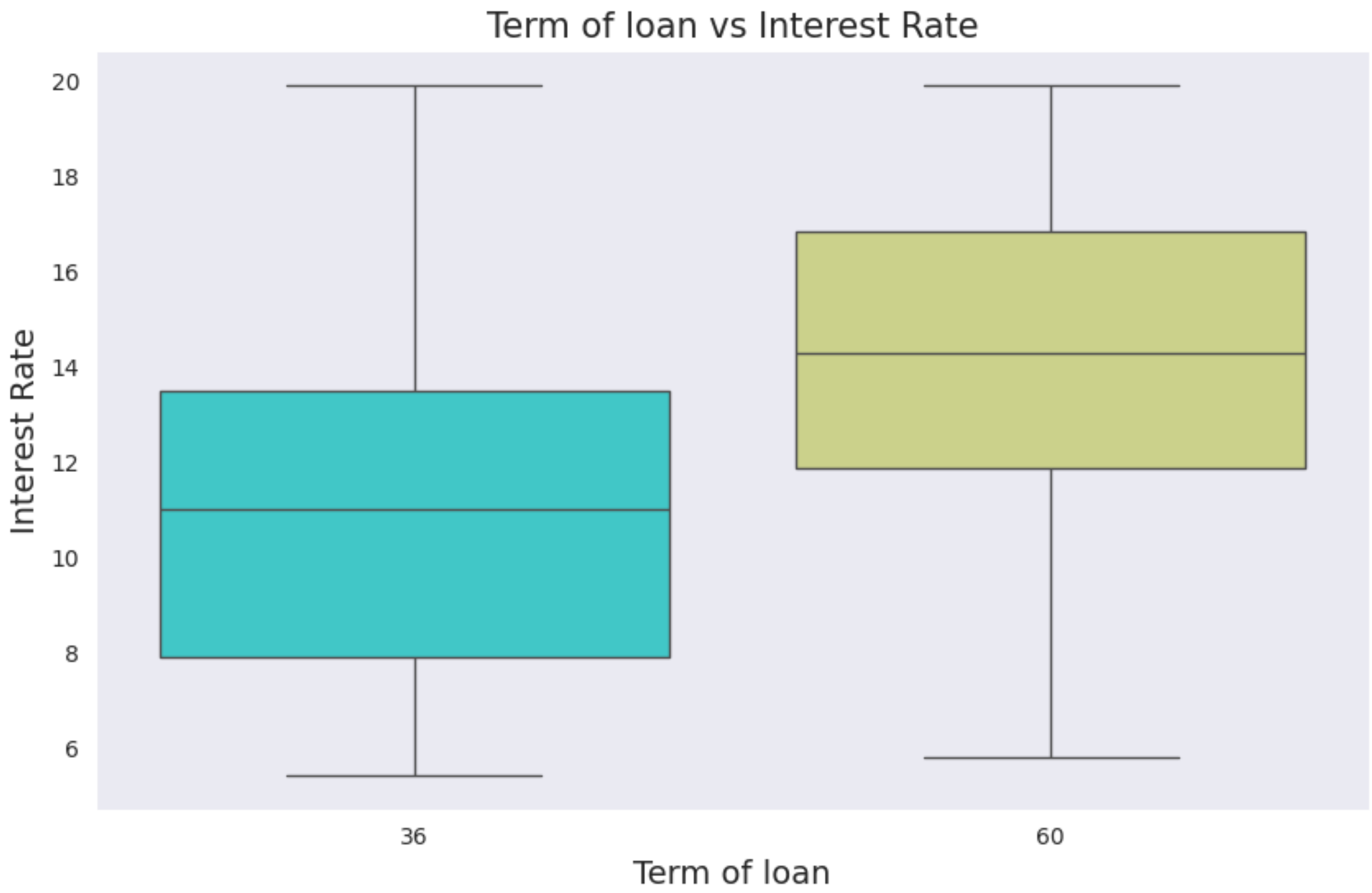
- Loans for small business purposes have the highest average interest rates among all loan categories due to higher risk and operational uncertainties.

- Debt consolidation loans rank second in terms of average interest rates possibly due to historical credit issues or the need for structured repayment plans.



Risk Assessment 14:
Term of Loan v/s Interest Rate

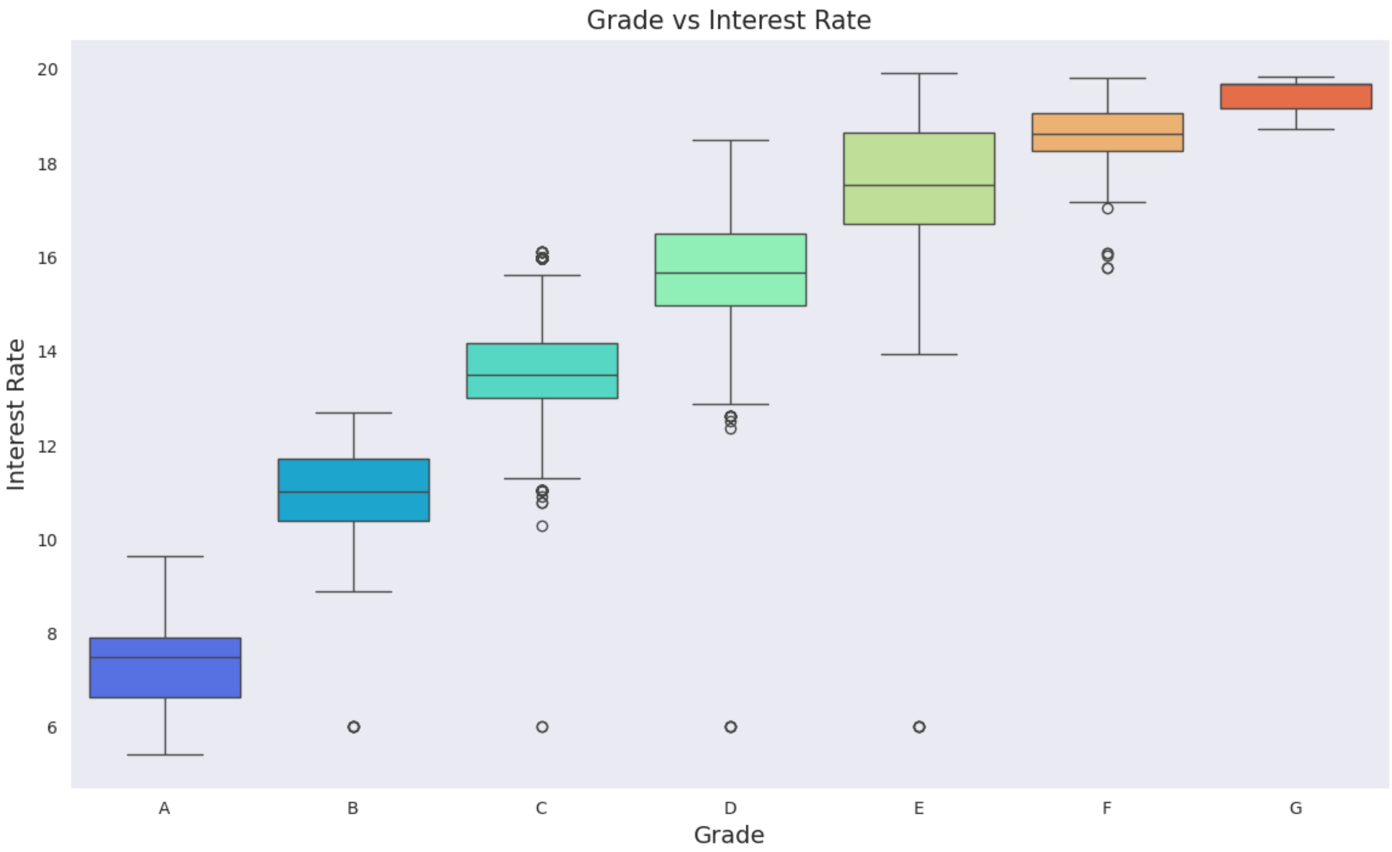
Loans with a 60-month term generally have higher average interest rates compared to those with a 36-month term. The longer period incurs higher interest rates due to increase risk to the lenders.



Risk Assessment 15: Grade v/s Interest Rate

- A-grade is considered the highest credit grade assigned to borrowers receiving lower interest rates on loans, reflecting lower perceived risk and higher creditworthiness.

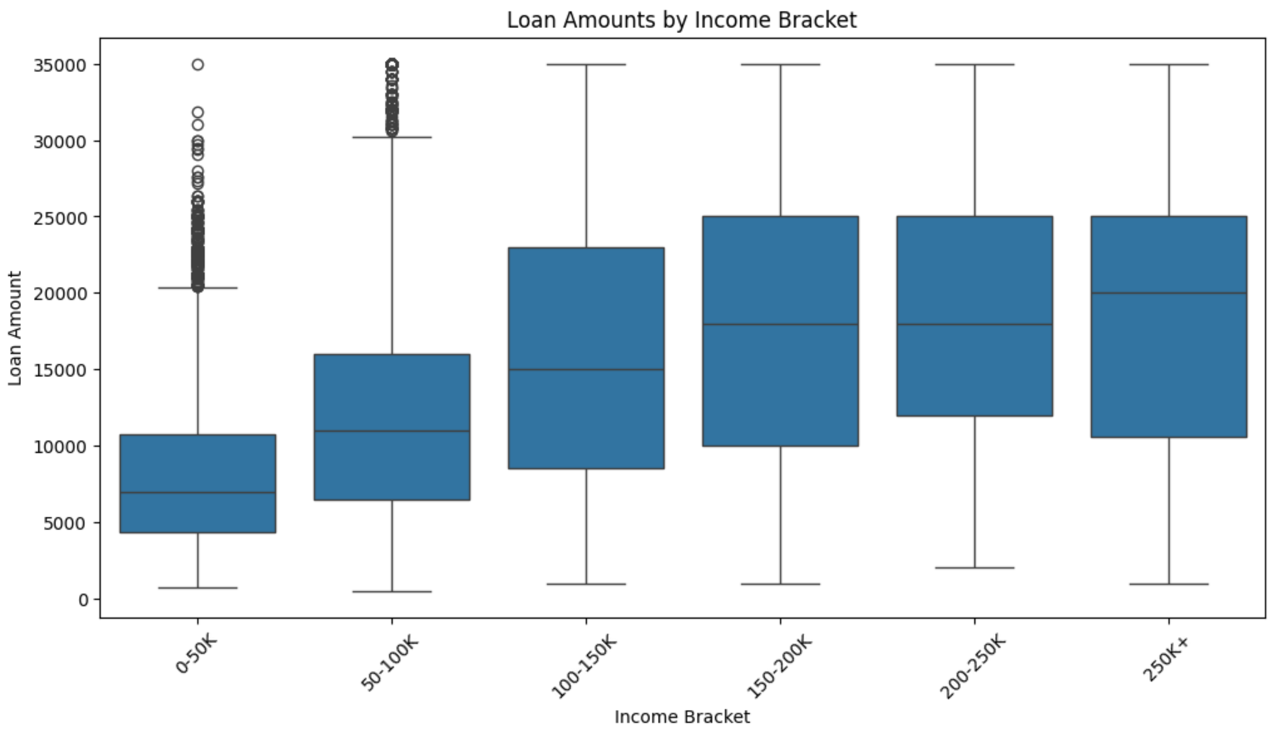
- Interest rates increase as loan grades move from A to F indicating higher risk to lenders, resulting in higher interest rates to compensate for potential defaults.



Risk Assessment 16: Loan Amounts v/s Income

- Annual income brackets are indicative of the borrower's financial capacity and ability to repay loans:
  - **Higher Income Brackets:** Borrowers in higher income brackets typically have more financial stability, higher creditworthiness, and lower debt-to-income ratios.
  - **Lower Income Brackets:** Borrowers in lower income brackets may have limited financial resources, higher debt-to-income ratios, and less stable income.

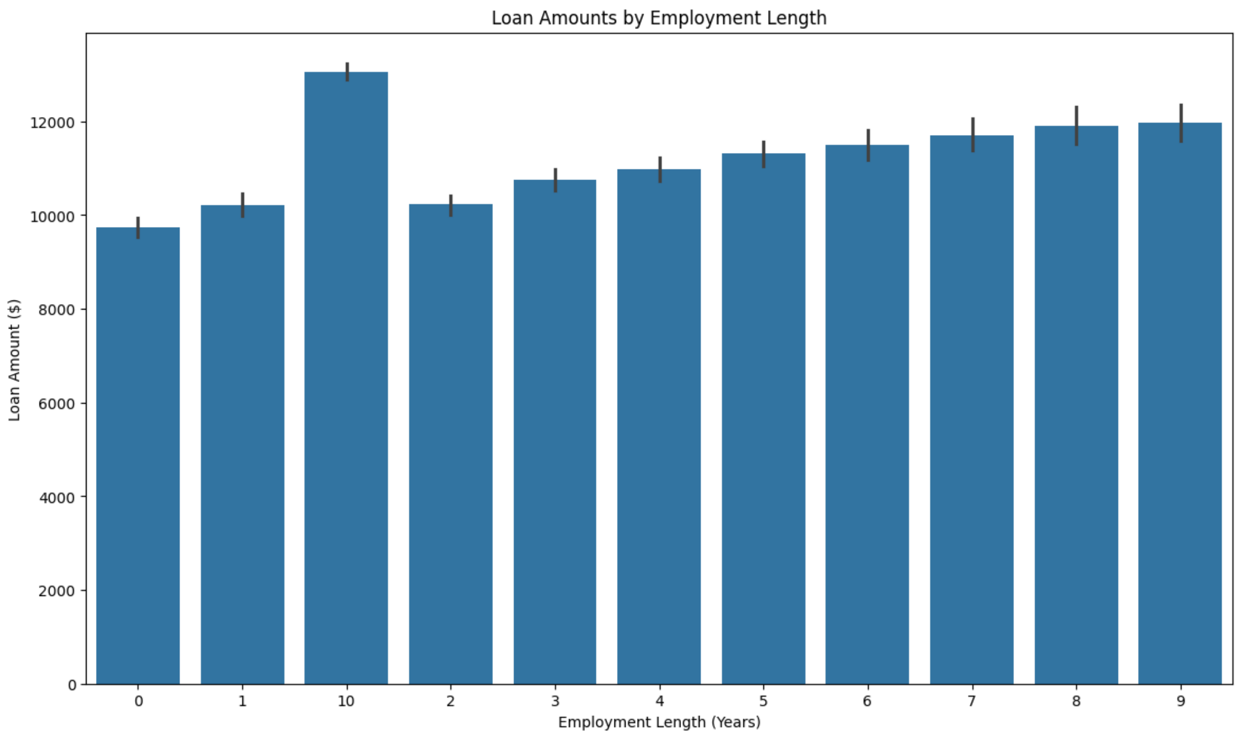
- The box plot clearly indicates borrowers falling within the higher income bracket are eligible to higher credit as against borrowers falling under the lower income bracket.



Risk Assessment 17: Loan Amounts v/s Employment Length

- Longer employment histories indicate greater job stability, reliable source of income, which can influence lenders' decisions to approve higher loan amounts:
  - **Job Stability:** Borrowers with longer employment histories are perceived as having a lower risk of income disruption or job loss. This stability reassures lenders that the borrower can sustain loan repayments over an extended period.
  - **Loan Eligibility:** Lenders may be more inclined to approve higher loan amounts for borrowers with longer employment histories due to their demonstrated ability to manage financial commitments and lower perceived risk of default.

- The histogram shows a trend where borrowers with longer employment histories receive higher loan amounts. This reflects borrower's financial stability, repayment capacity and lower risk.



Risk Mitigation Strategies to Minimize Loan Default

- Rigorous Loan Screening
- Loan Term Management
- Creditworthiness Assessment
- Income Verification
- Interest Rate Adjustment
- DTI Ratio Monitoring
- Employment Stability
- Targeted Risk Management Programs

Conclusion

The analysis of loan default risk using the provided Lending Club dataset reveals several critical insights into the factors influencing loan default. Key drivers of loan default include:

- **Loan Purpose**: Debt consolidation and credit card repayment loans exhibit high default rates. Borrowers using loans for these purposes are likely under financial stress.
- **Home Ownership**: Renters and mortgaged homeowners show higher default rates compared to homeowners without mortgages, indicating a correlation between financial obligations and default risk.
- **Loan Term**: Loans with longer terms (60 months) have higher default rates compared to shorter-term loans (36 months), suggesting that longer repayment periods increase the risk of financial instability.
- **Borrower Grades**: Loans graded "A" have low default rates, while those graded "F" and "G" have significantly higher default rates, indicating a strong relationship between borrower creditworthiness and default risk.
- **Income Levels**: Borrowers with higher annual incomes ($80K and above) show lower default rates, highlighting income as a significant predictor of repayment ability.
- **Interest Rates**: Higher interest rates are associated with increased default rates, reflecting the higher risk profiles of borrowers with such loans.
- **Debt-to-Income Ratio**: Higher DTI ratios are linked to higher default rates, highlighting the importance of managing debt relative to income.
- **Employment Length**: Longer employment histories correlate with higher loan amounts and lower default rates, suggesting job stability as a key factor in loan repayment.