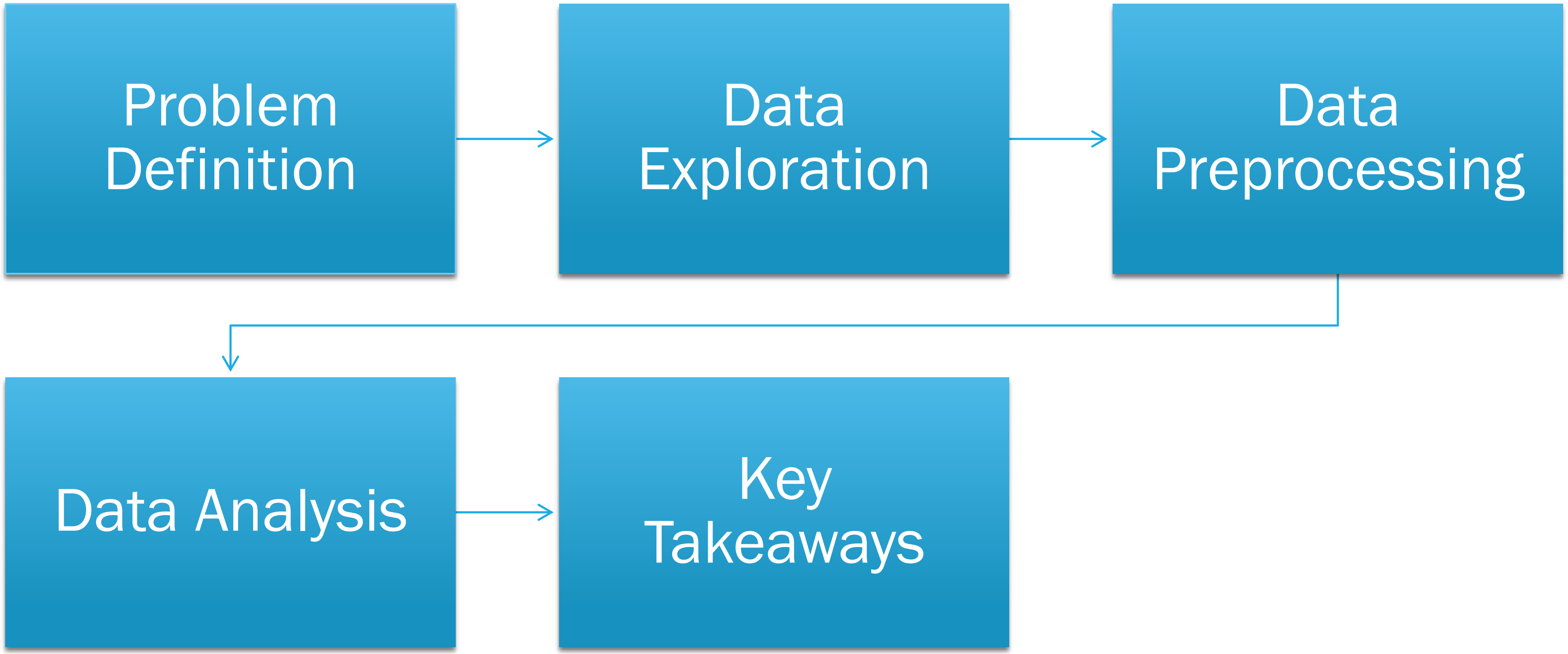
LENDING CLUB CASE STUDY

BY – Vinod Kumar / Bijay Kumar Mishra (Sept – 23 Batch)

PROJECT METHODOLOGY



PROBLEM DEFINITION

- To minimize credit losses by identifying risky loan applicants through Exploratory Data Analysis (EDA).
- The goal is to understand the key factors driving loan defaults, which will inform portfolio and risk assessment.

DATA EXPLORATION

- We have 2 files shared for the case study
  - loan.csv – That contains 39717 rows and 111 columns data for data analysis.
  - dictionary.xlsx – it contains all the definition of 111 columns used in loan.csv.

DATA CLEANING

- Identified columns
  - Having null values – criteria 40% or more null values
  - Single value across all the rows
  - Found 66 columns and dropped them
- Identified duplicate rows.
  - No duplicate entries were found.
- Identified non-essential columns
  - Dropped 21 columns
- Dropped all rows having null
- Removal of unnecessary characters from few columns
  - Removed '%', 'months', 'years' etc.

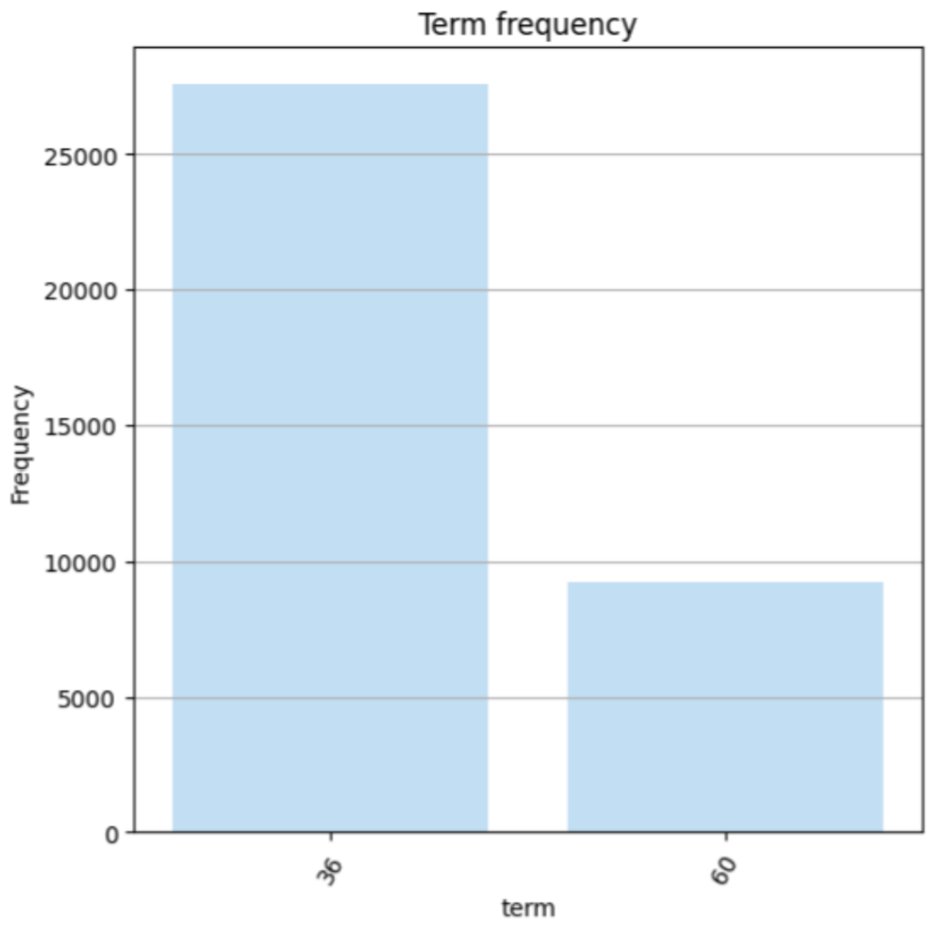
Final cleaned dataset has 37898 rows and 24 columns

DATA ANALYSIS

- Univariate Analysis
- Segmented Univariate Analysis
- Bivariate Analysis
- Derived Metrix

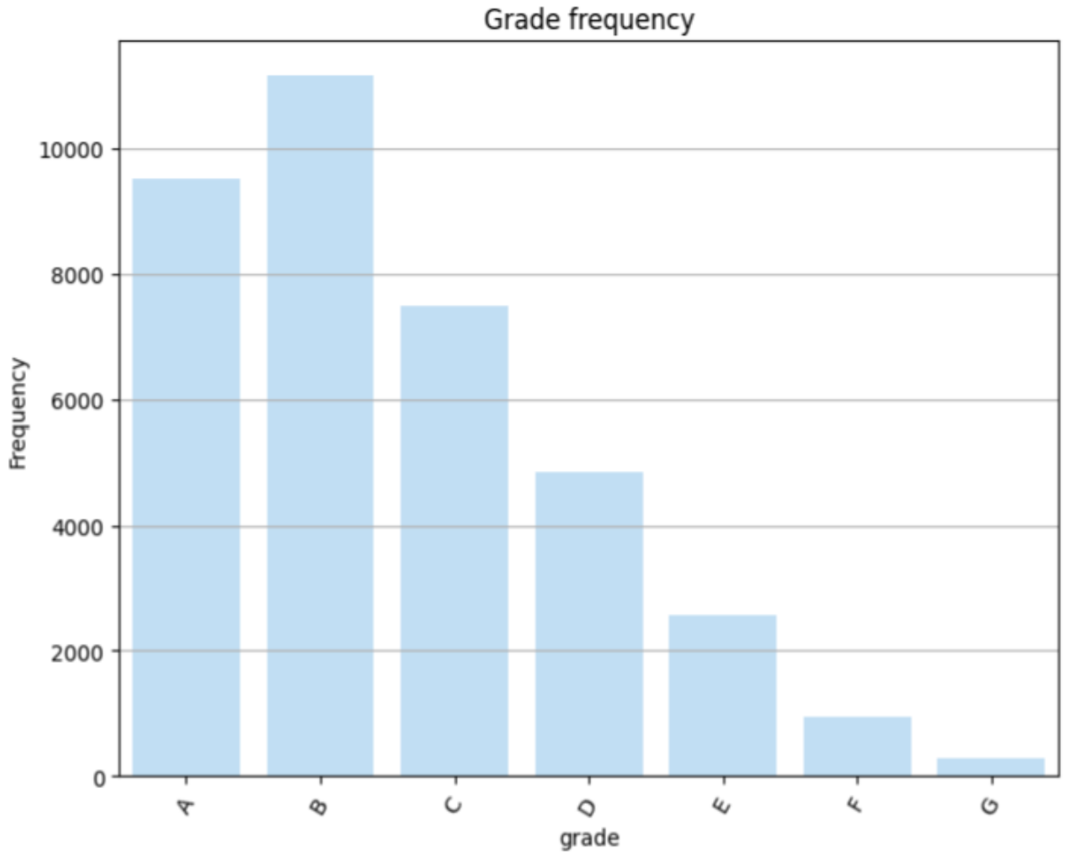
UNIVARIATE ANALYSIS

- Variable: Term
- Loan term of 36 Months: This is the most common loan term with 27,546 loans.
- Loan term of 60 Months: There are 9,254 loans with a term of 60 months.
- People prefer shorter term loans i.e. 36 Months.
- Demand for shorter term loans is almost 3 times of long term loans.



UNIVARIATE ANALYSIS

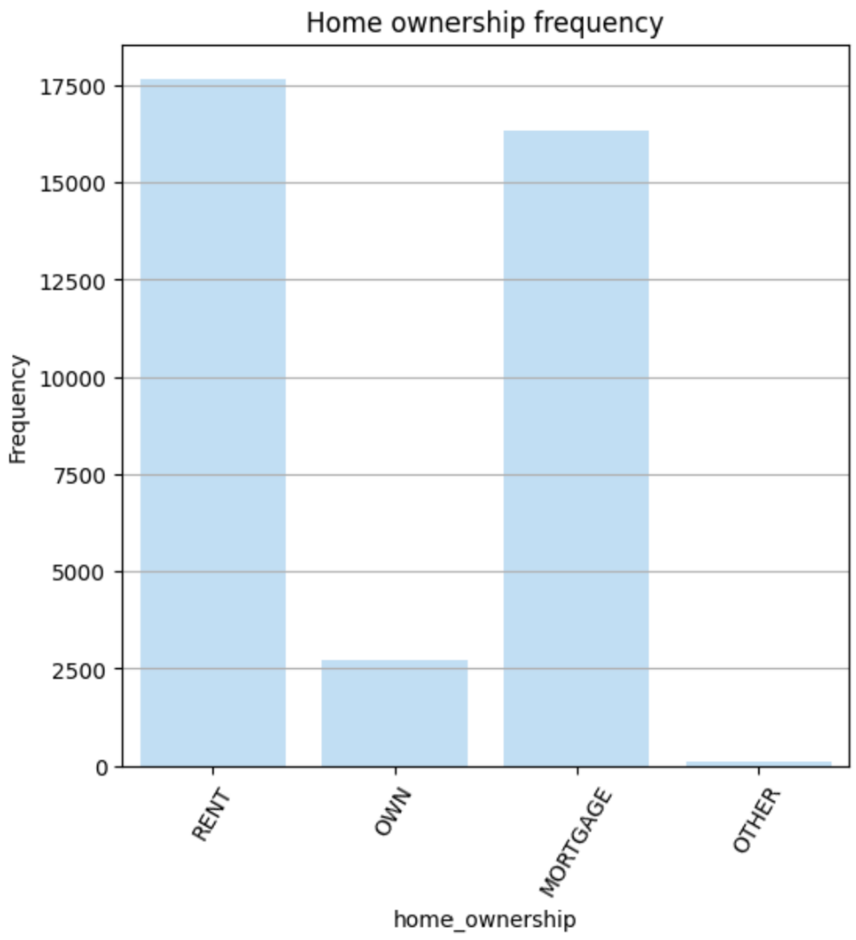
- Variable: Grade
- Grade A: There are 9,518 loans in this category.
- Grade B: This is the most common grade with 11,149 loans.
- Grade C: This category includes 7,474 loans.
- Grade G: This is the smallest category with 294 loans.
- Barring Grade ‘B’ employees demand is decreasing as grade is increasing



UNIVARIATE ANALYSIS

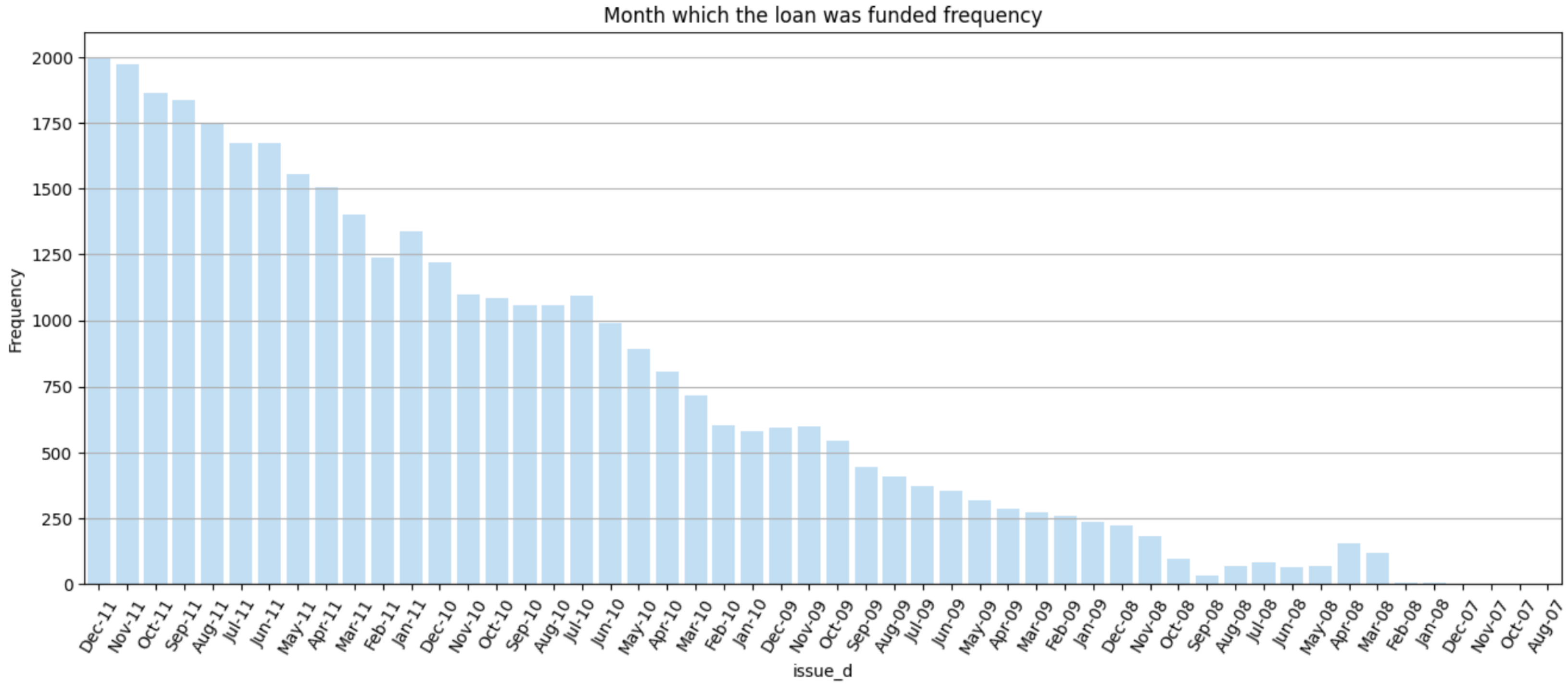
- Variable: Home ownership

- People having rented and mortgage homes have higher demand for loans compared to having own house.



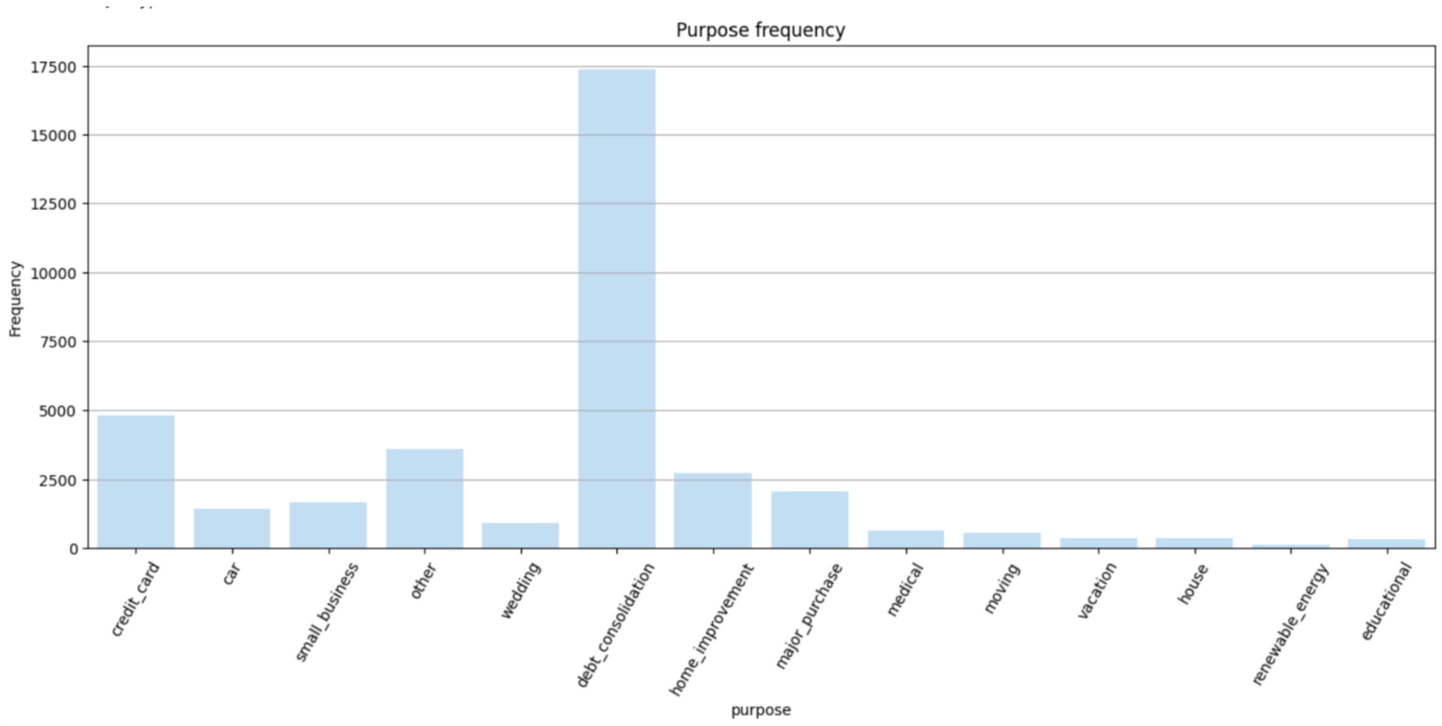
UNIVARIATE ANALYSIS

- Variable: The date of issue of loan
- The data shows a general increase in the number of loans issued over time, with fewer loans issued in the earlier years (2007-2009).



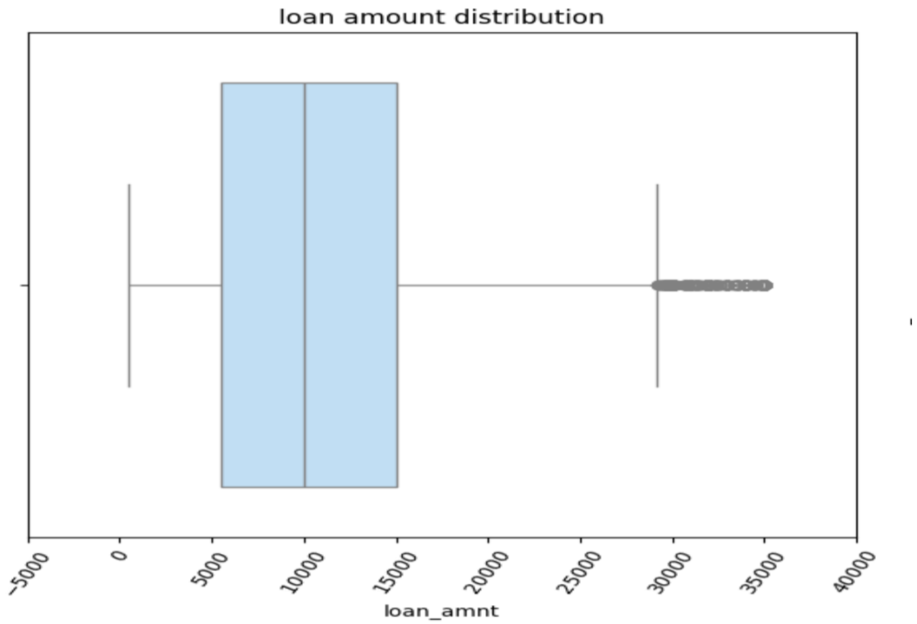
UNIVARIATE ANALYSIS

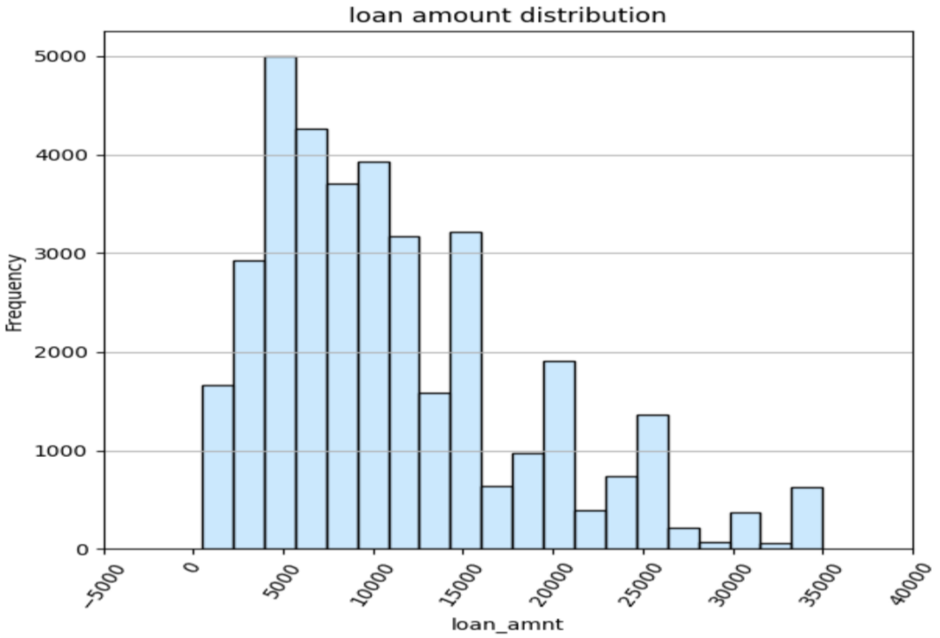
- Variable: Purpose
- The majority of loans are taken out for debt consolidation, with 17,373 instances.
- Other common purposes include credit card payments and home improvement.
- Purposes like education and renewable energy constitute a smaller portion of the loans.



UNIVARIATE ANALYSIS

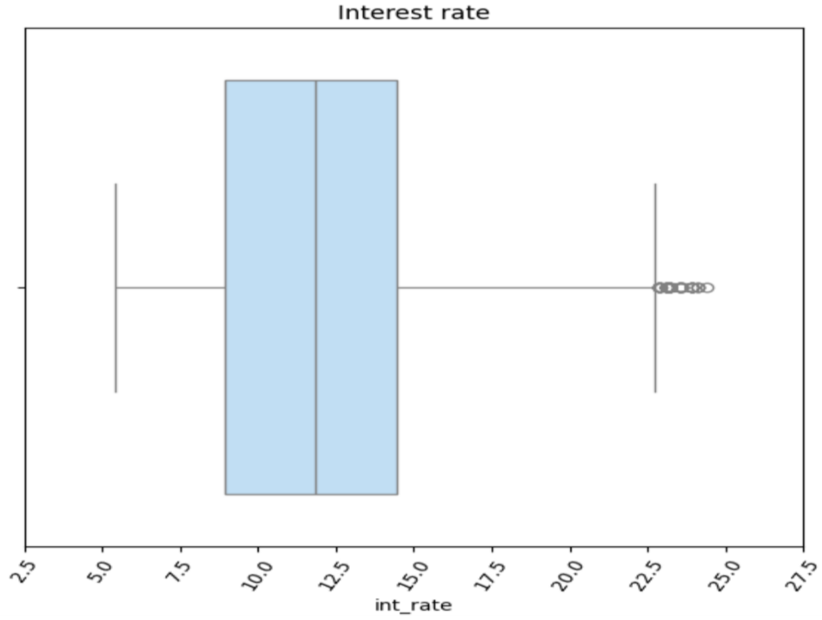
- Variable: Loan Amount
- **Mean:** The average loan amount is $11,149.54.
- **Standard Deviation:** The standard deviation is $7,369.86, indicating a wide dispersion of loan amounts.
- **Range:** Loan amounts range from $500 to $35,000.
- **Quartiles:** 25% of loans are $5,500 or less, 50% are $10,000 or less, and 75% are $15,000 or less.
- **Interpretation:** Most people borrow in the lower to middle end of the range, but there's a wide variety.

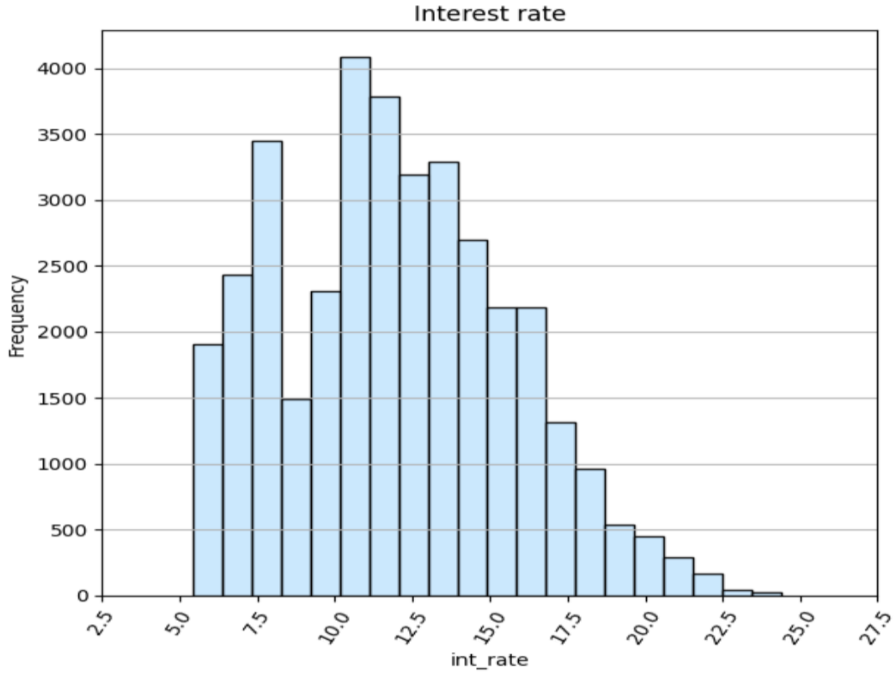




UNIVARIATE ANALYSIS

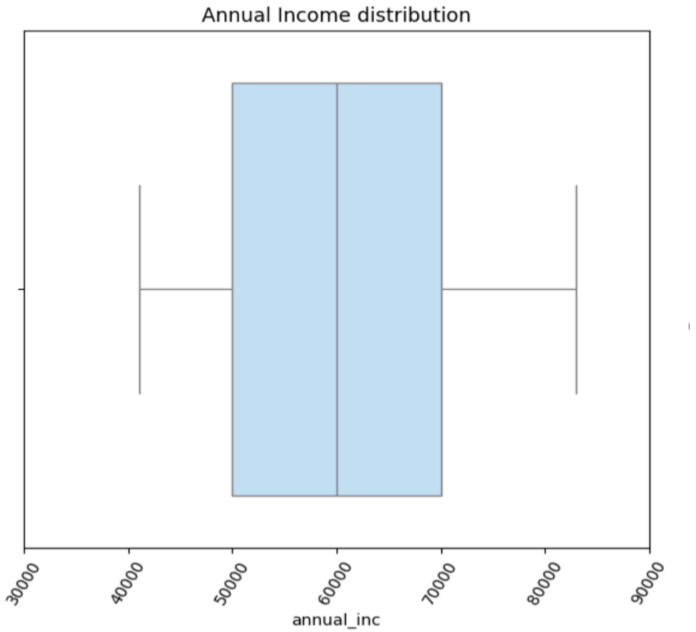
- Variable: Interest rate
- Mean: The average interest rate is 11.98%.
- Standard Deviation: The standard deviation is 3.70%, indicating that interest rates can vary quite a bit from the average.
- Range: Interest rates range from 5.42% to 24.40%, showing a wide spread in the rates charged.
- Median: The median interest rate is 11.83%, meaning half of the loans have interest rates below this and half are above.
- In summary, while there’s a broad range of interest rates, most of them hover around the 11-12% mark.

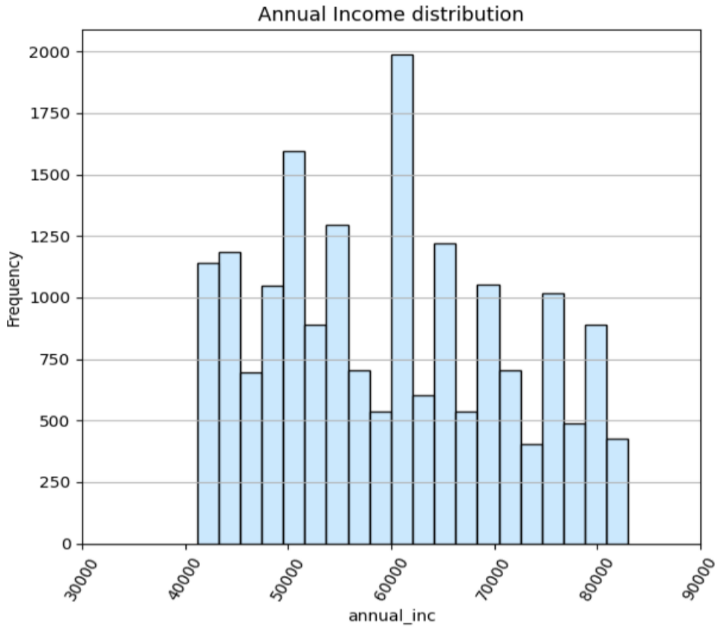




UNIVARIATE ANALYSIS

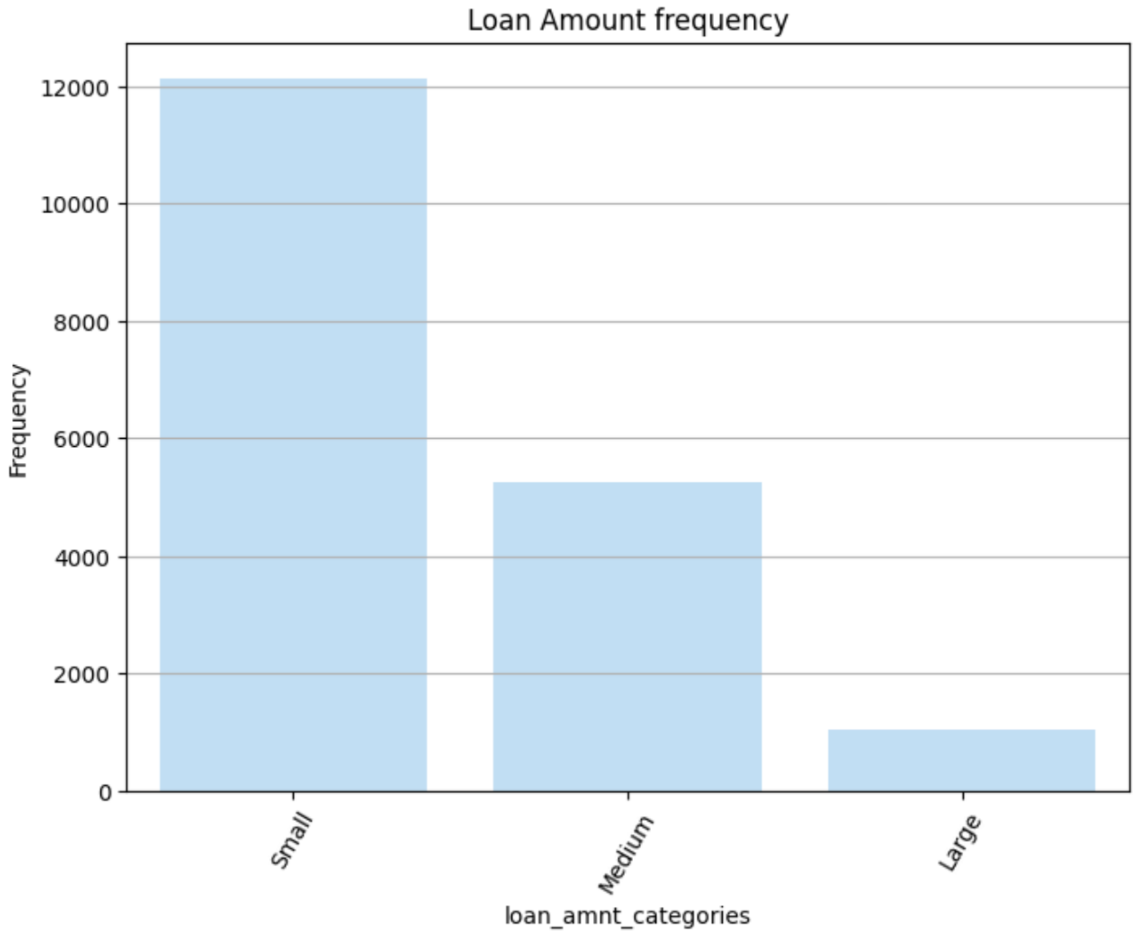
- Variable: Annual Income
- Found outliers. So, removed using IQR.
- After removal, there were 18,421 rows left.
- The average annual income of the 18,421 loan users is about $59,782.82, with a standard deviation of $11,534.64.
- Incomes range from $41,136 to $83,000. The median income is $60,000





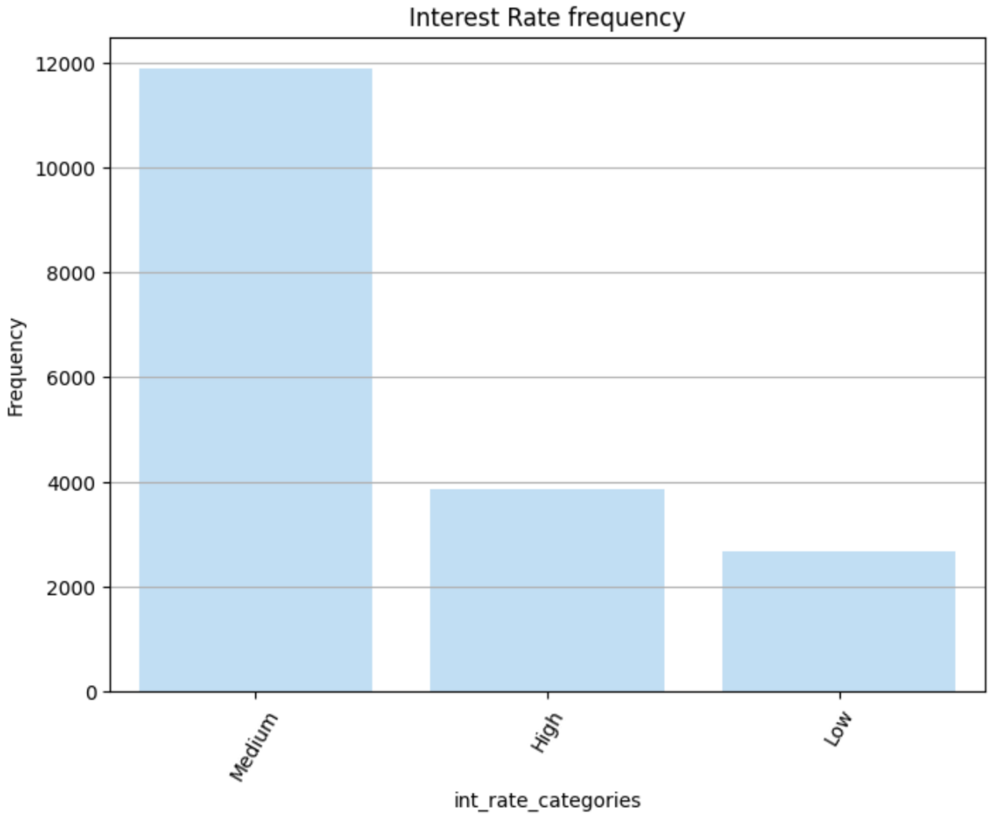
SEGMENTED UNIVARIATE ANALYSIS

- Segment: Loan amount categories
- Small Loans (<= $12,000): This is the most common category with 12,139 loans.
- Medium Loans ($12,000 - $24,000): There are 5,248 loans in this category.
- Large Loans (>$24,000): This is the smallest category with 1,034 loans.



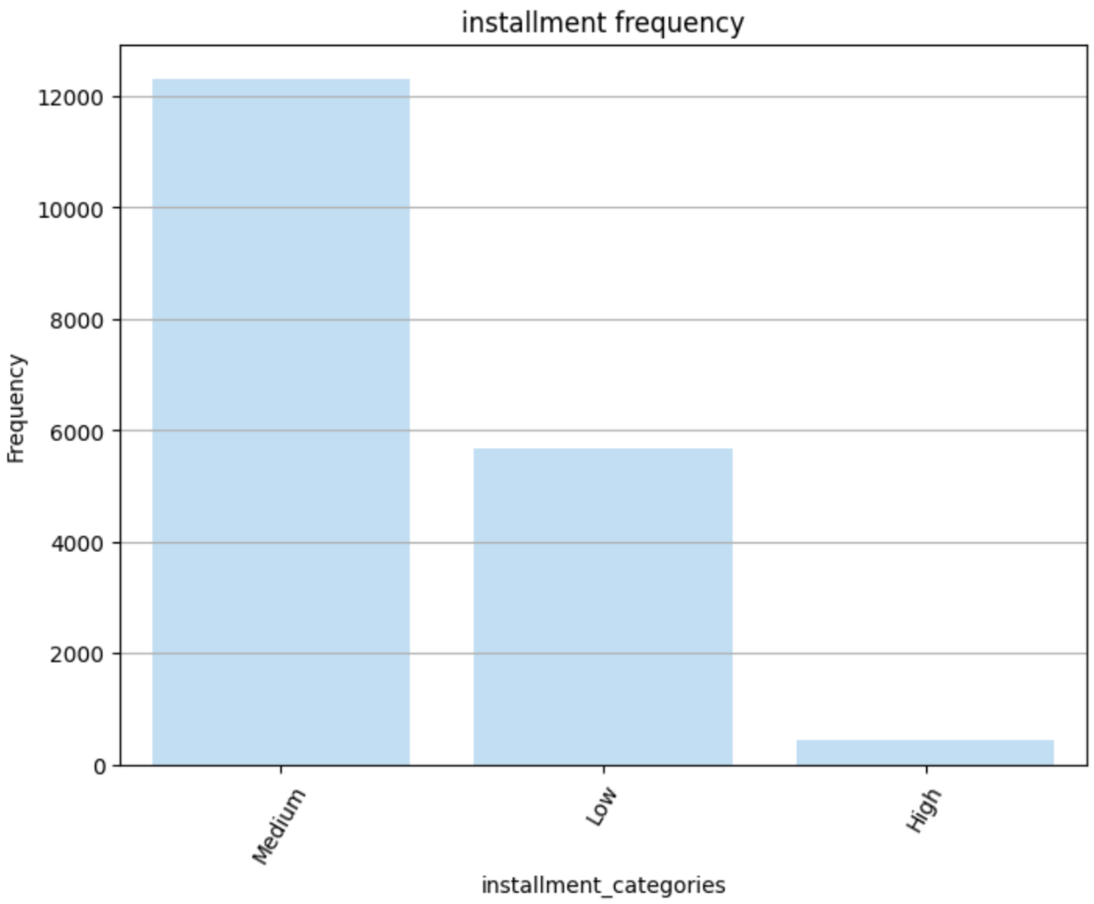
SEGMENTED UNIVARIATE ANALYSIS

- Segment: Interest rate categories
- Medium Interest Rates (<= 15%): This is the largest category with 11,893 loans.
- High Interest Rates (>15%): There are 3,861 loans in this category.
- Low Interest Rates (<= 7.5%): This is the smallest category with 2,667 loans.
- Loans with medium interest rates are in highest demand, followed by those with high rates, and finally, loans with low interest rates.



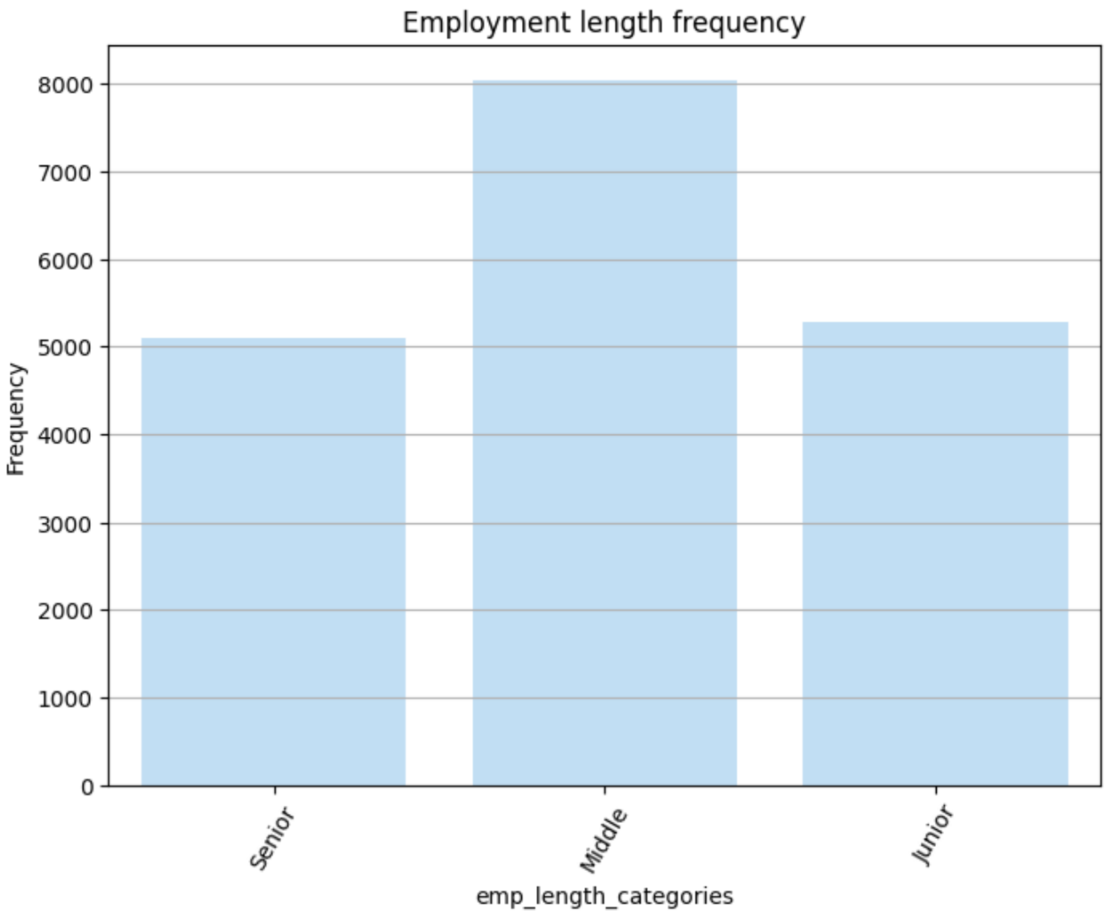
SEGMENTED UNIVARIATE ANALYSIS

- Segment: Installemnt amount categories
- Low Installments amount(<= $200): There are 5,657 loans in this category.
- Medium Installments amount (<= $800): This is the most common category with 12,321 loans.
- High Installments amount (>$800): This is the smallest category with 443 loans.
- Based on the amount of instalment, medium instalment amount loans are highest, followed by low amount and finally high.



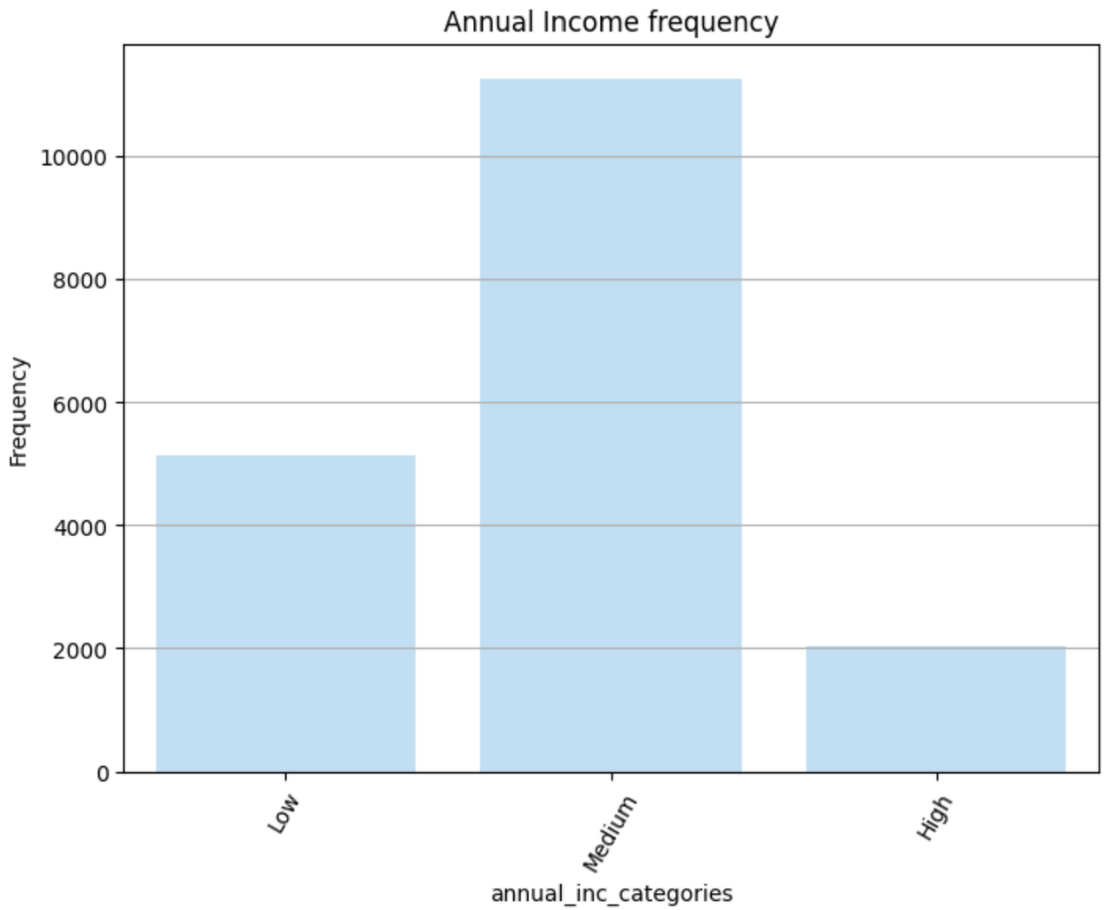
SEGMENTED UNIVARIATE ANALYSIS

- Segment: Employment experience categories
- Junior level (<= 2 years): There are 5,280 individuals in this category.
- Middle level (<= 8 years): This is the largest category with 8,048 people.
- Senior level (>8 years): This category includes 5,093 people.
- Based on number of years of experience, Middle-level experience employees have highest demand for loans, followed by junior and finally senior.



SEGMENTED UNIVARIATE ANALYSIS

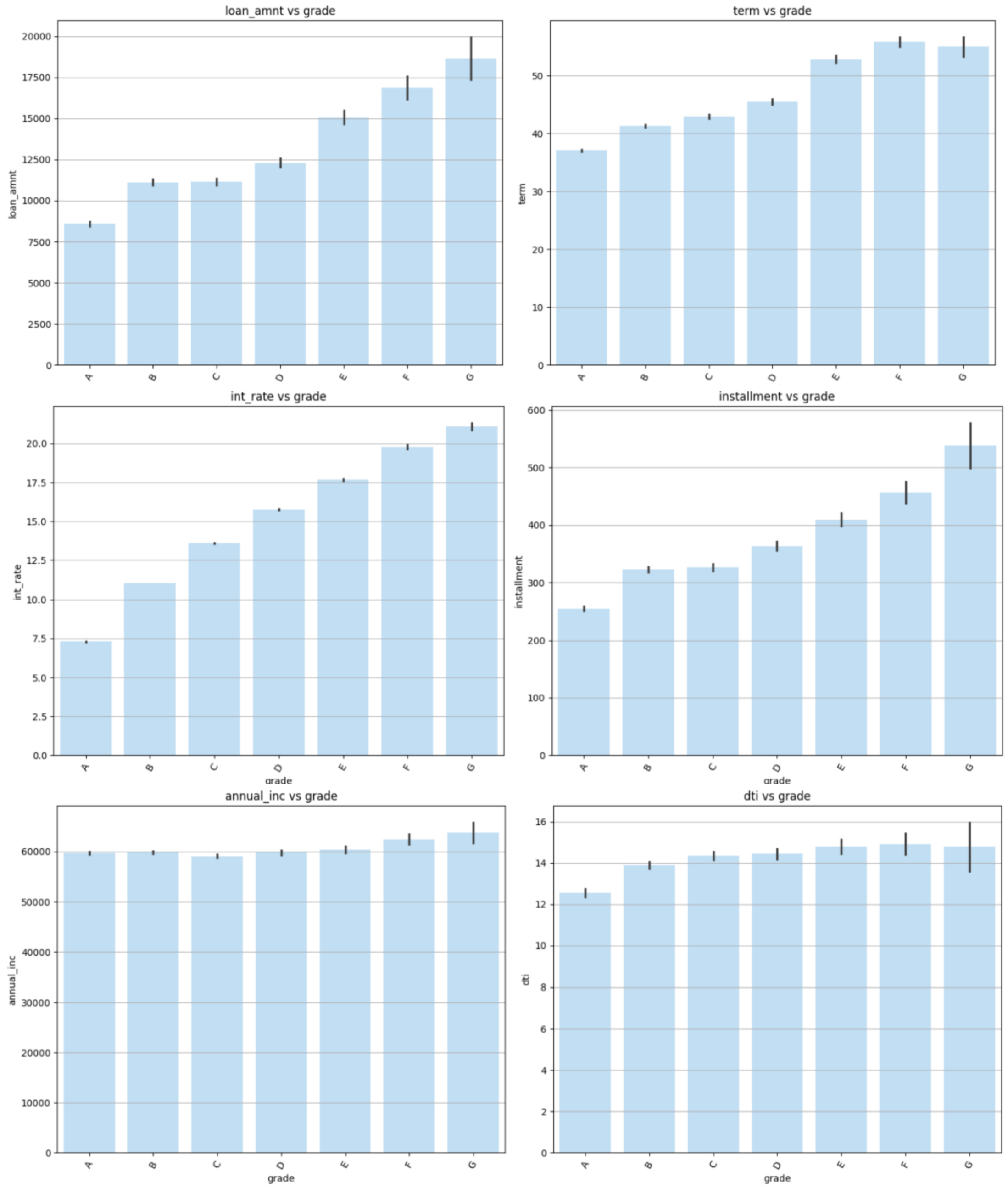
- Segment: Annual income categories
- Low Income (<= $50,000): There are 5,122 individuals in this category.
- Medium Income (<= $75,000): This is the largest category with 11,267 individuals. These individuals have a moderate level of income.
- High Income (>$75,000): This category includes 2,032 individuals.
- Based on annual income, Medium income group users have highest demand for loans, followed by Low income group and finally high income group



BIVARIATE ANALYSIS

- As the ‘Grade’ changes from ‘A’ to ‘G’, so do the loan amount, interest rate, and installment amounts.

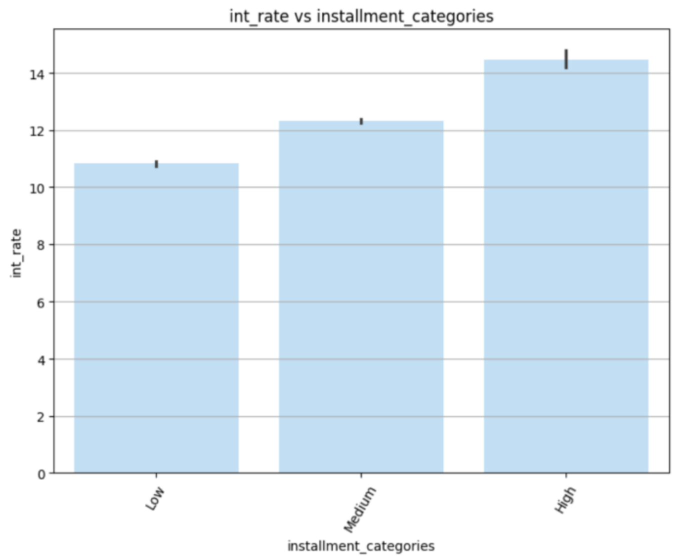
- This suggests that Grade ‘F’, ‘G’ levels are associated with higher risk, and therefore require higher interest rates and installments to compensate for this risk.

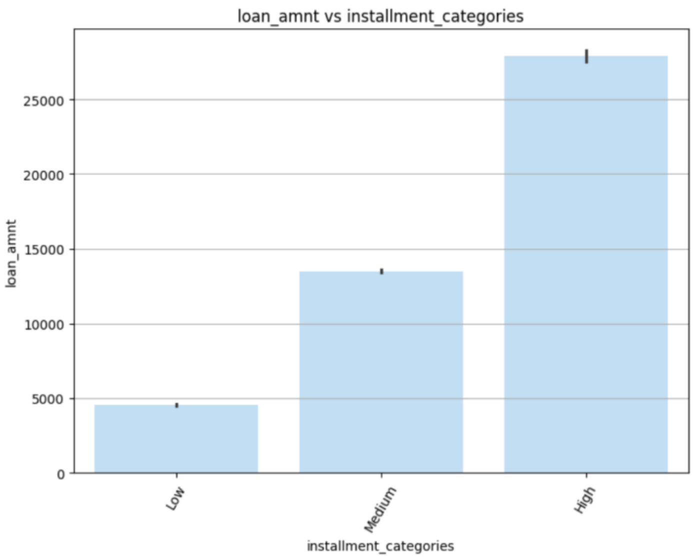


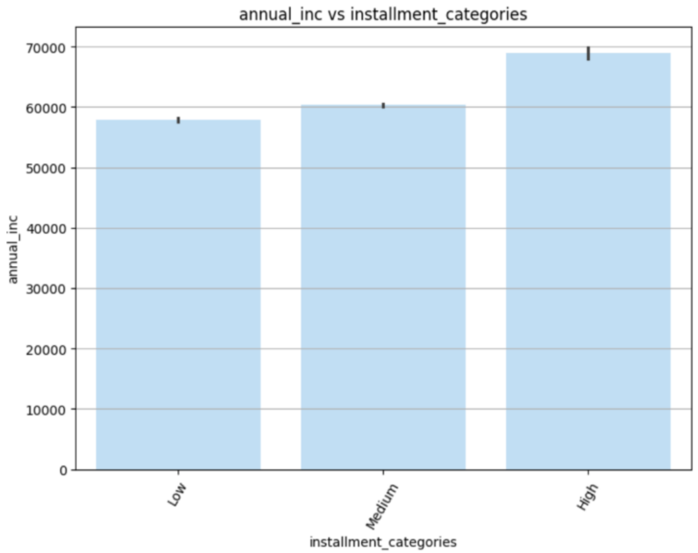
BIVARIATE ANALYSIS

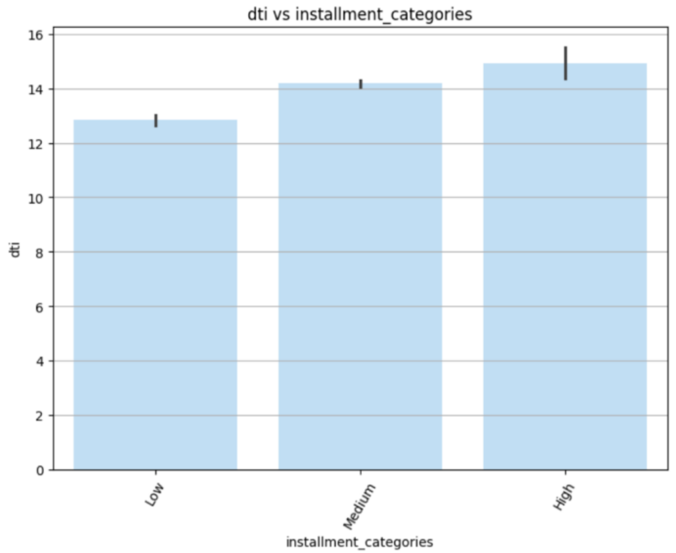
- Low amount instalments have less interest rate, followed by medium amount and finally, high amount instalments.

- As annual income or DTI increases, people tend to take higher installment loans.







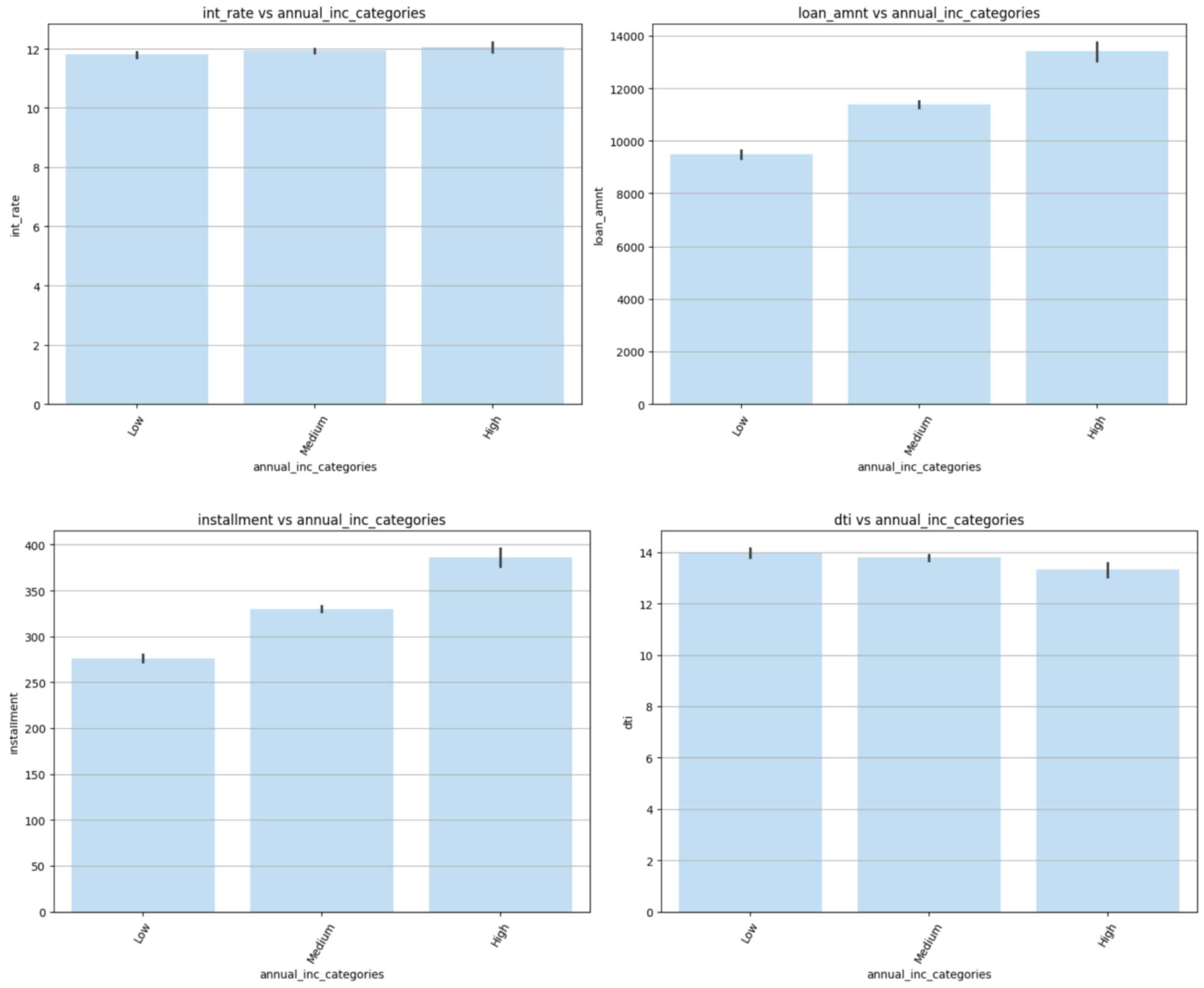


BIVARIATE ANALYSIS

- People with lower annual incomes tend to have smaller loan amounts and smaller installments, followed by those with medium annual incomes, and finally by those with high annual incomes.

- There seems to be no co-relation between annual income category and interest rate.

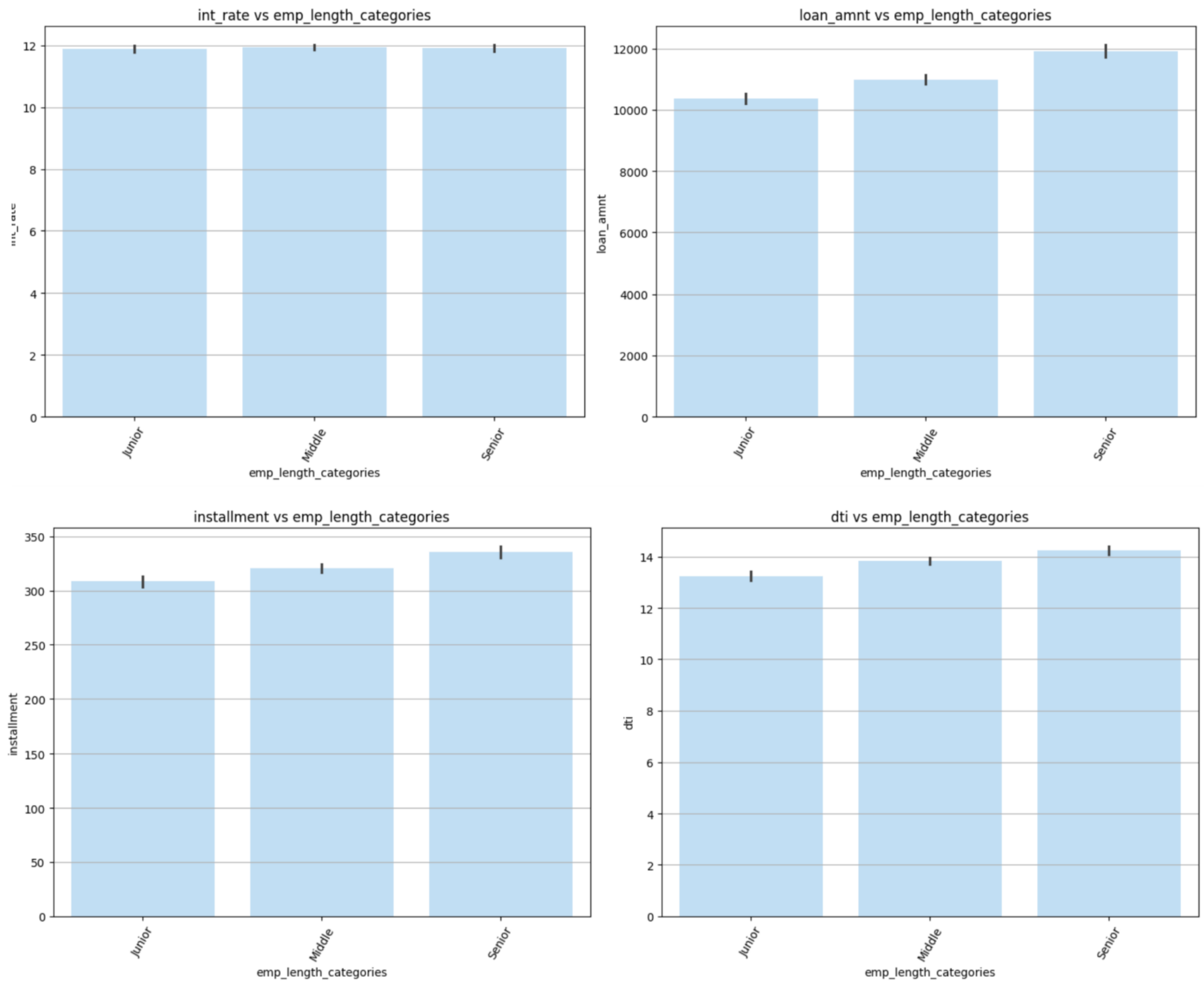
- DTI is slightly have negative co-relation.



BIVARIATE ANALYSIS

- Junior experience level employees have lower loan amount and instalment amount loan, followed by middle level and finally, by senior level.

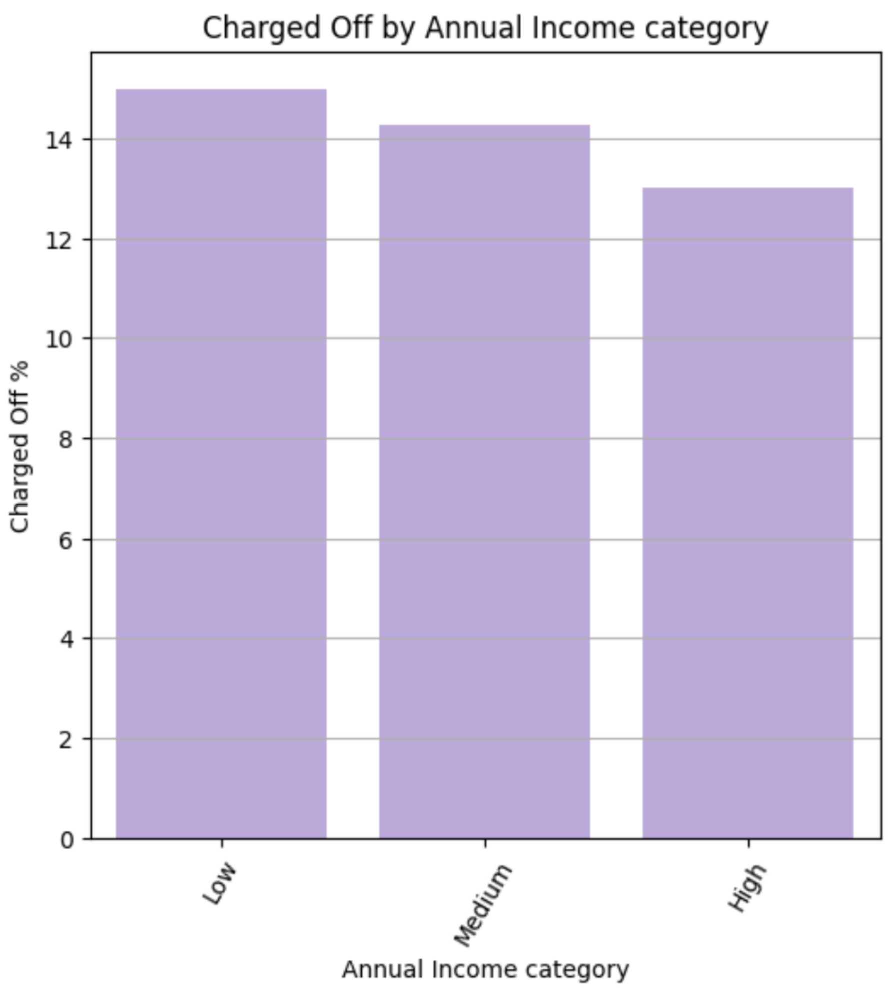
- Similarly, as junior experience level employees have lowest DTI, followed by middle experience and finally senior level.



DERIVED METRICS ANALYSIS

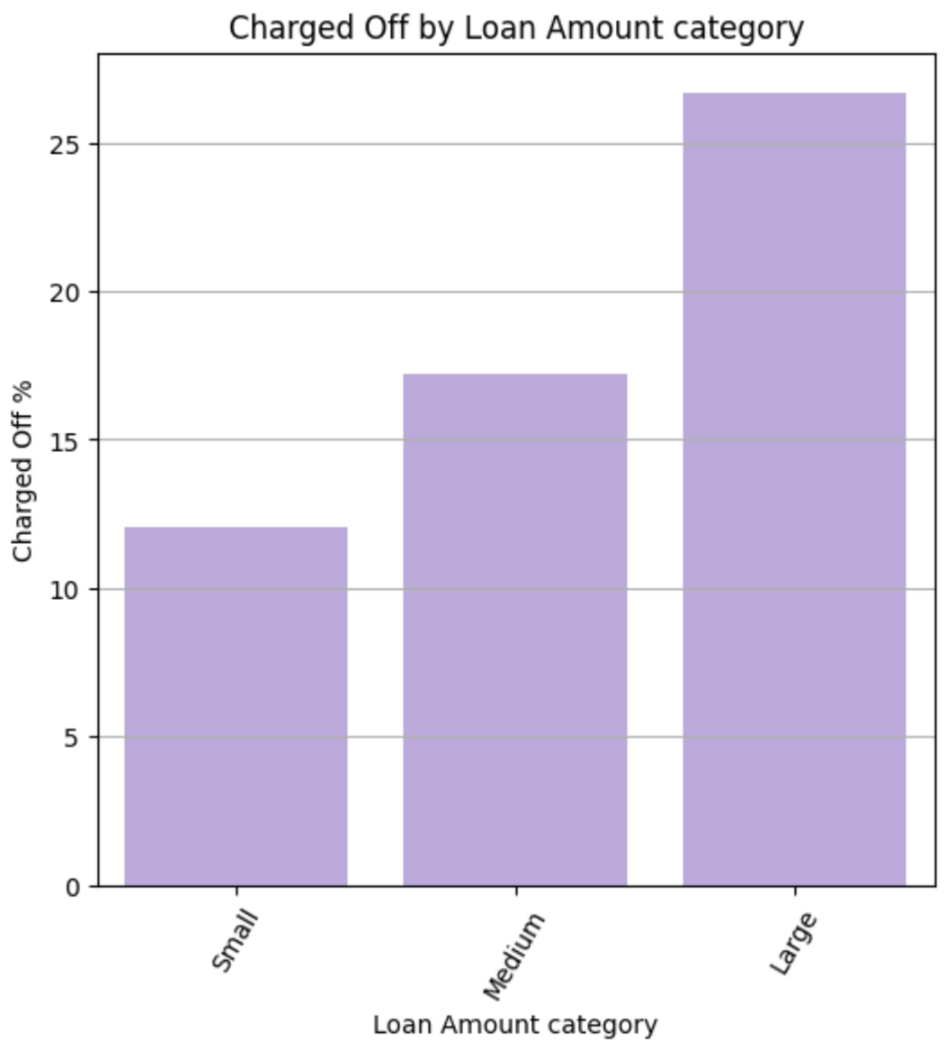
- Derived Metrics: Percentage of defaulters

- People in the low annual income range tend to default more often, followed by those in the medium income range, and finally those in the high income range.



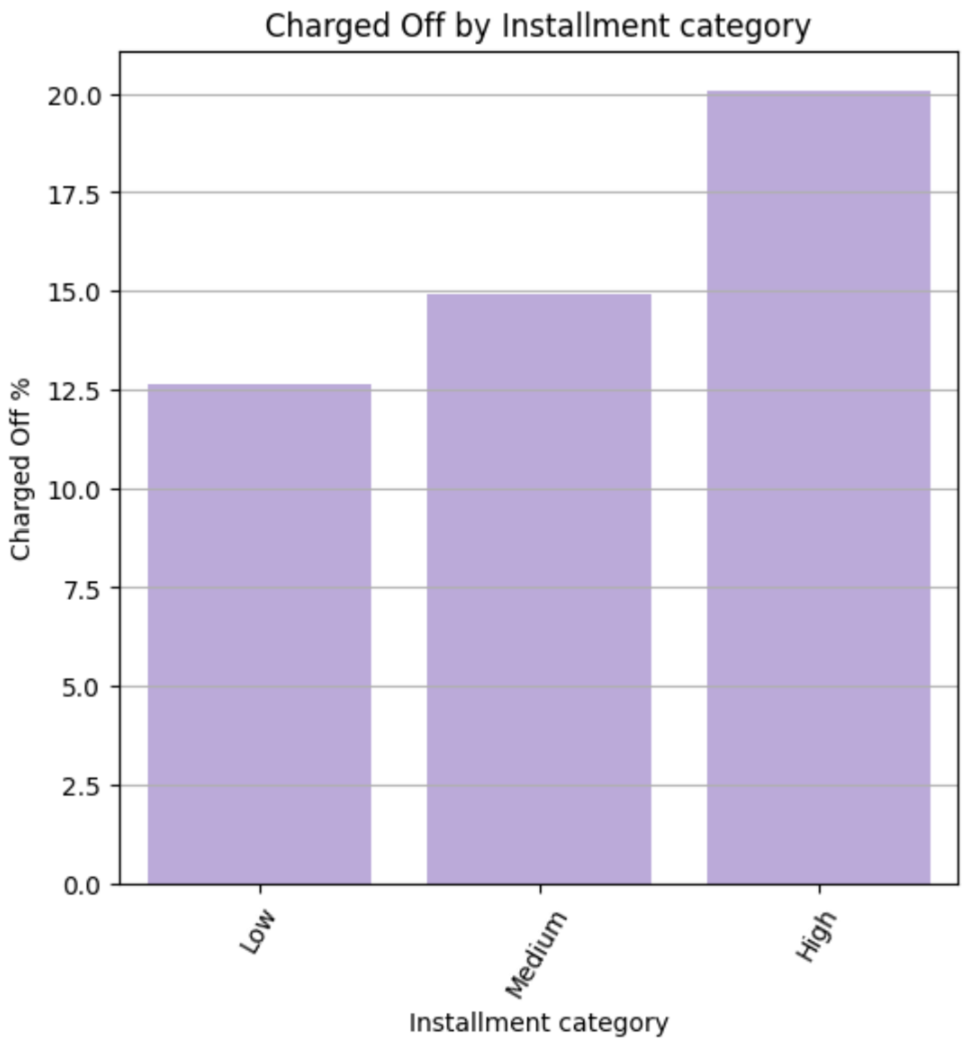
DERIVED METRICS ANALYSIS

- Derived Metrics: Percentage of defaulters
- As the loan amount increases, the rate of default also tends to increase.



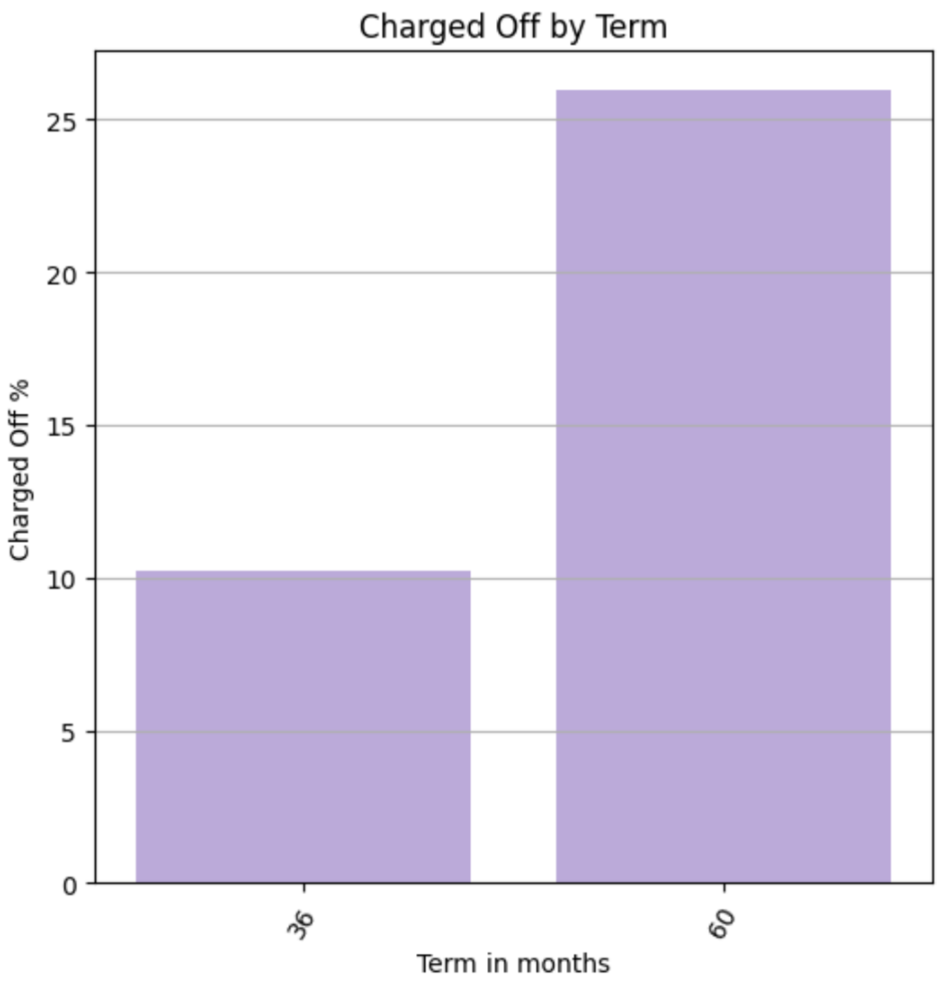
DERIVED METRICS ANALYSIS

- Derived Metrics: Percentage of defaulters
- As the installment amount increases, the rate of default also tends to increase.



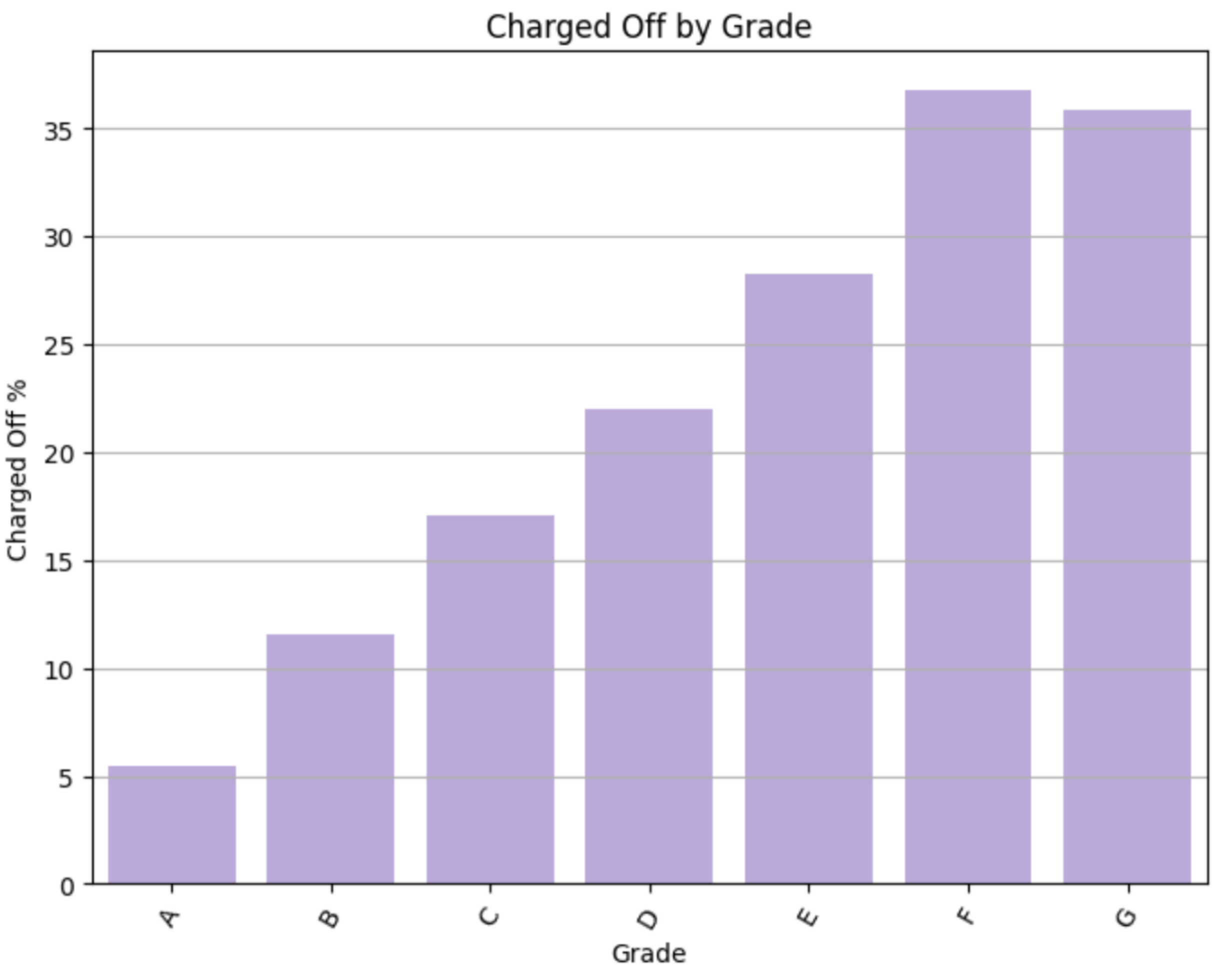
DERIVED METRICS ANALYSIS

- Derived Metrics: Percentage of defaulters
- Loans with longer terms tend to have a higher default rate.



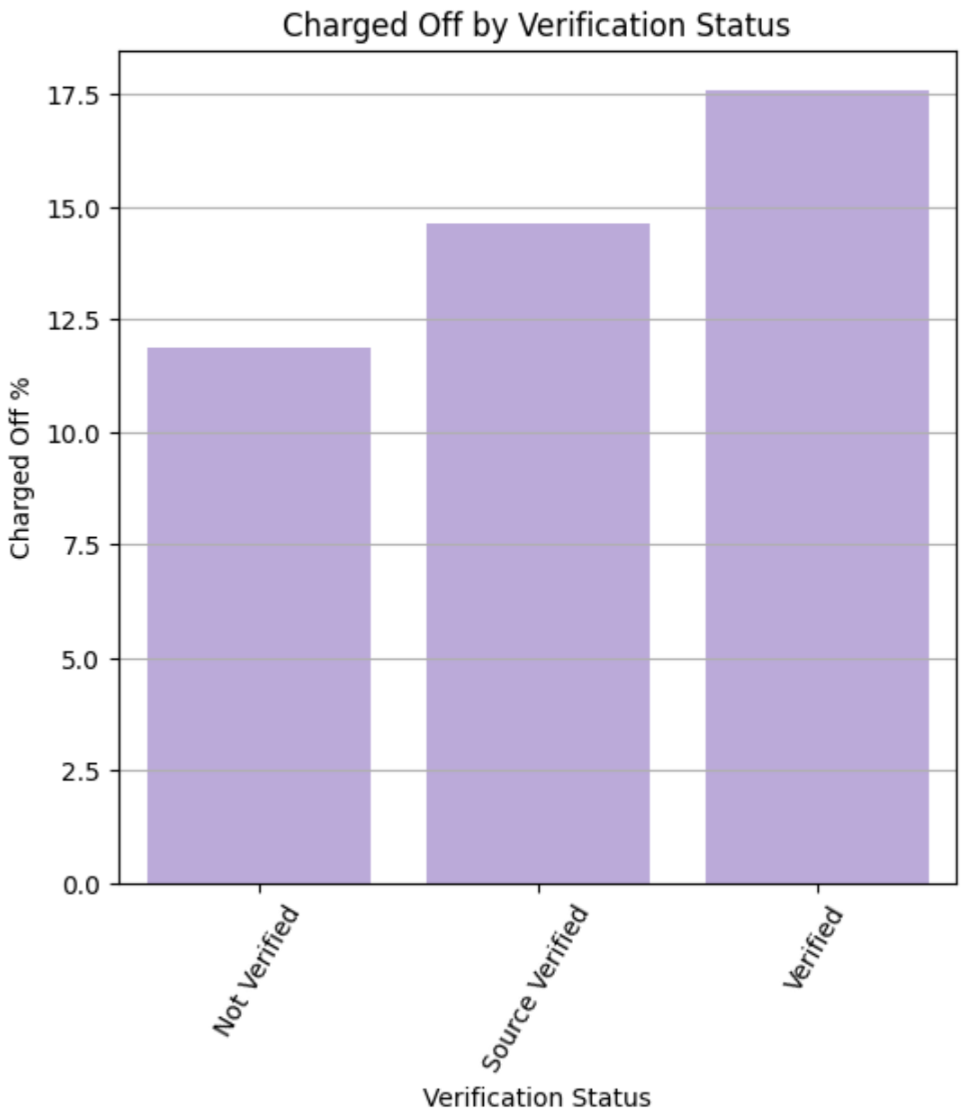
DERIVED METRICS ANALYSIS

- Derived Metrics: Percentage of defaulters
- As the grade changes defaulter rate is also increases.



DERIVED METRICS ANALYSIS

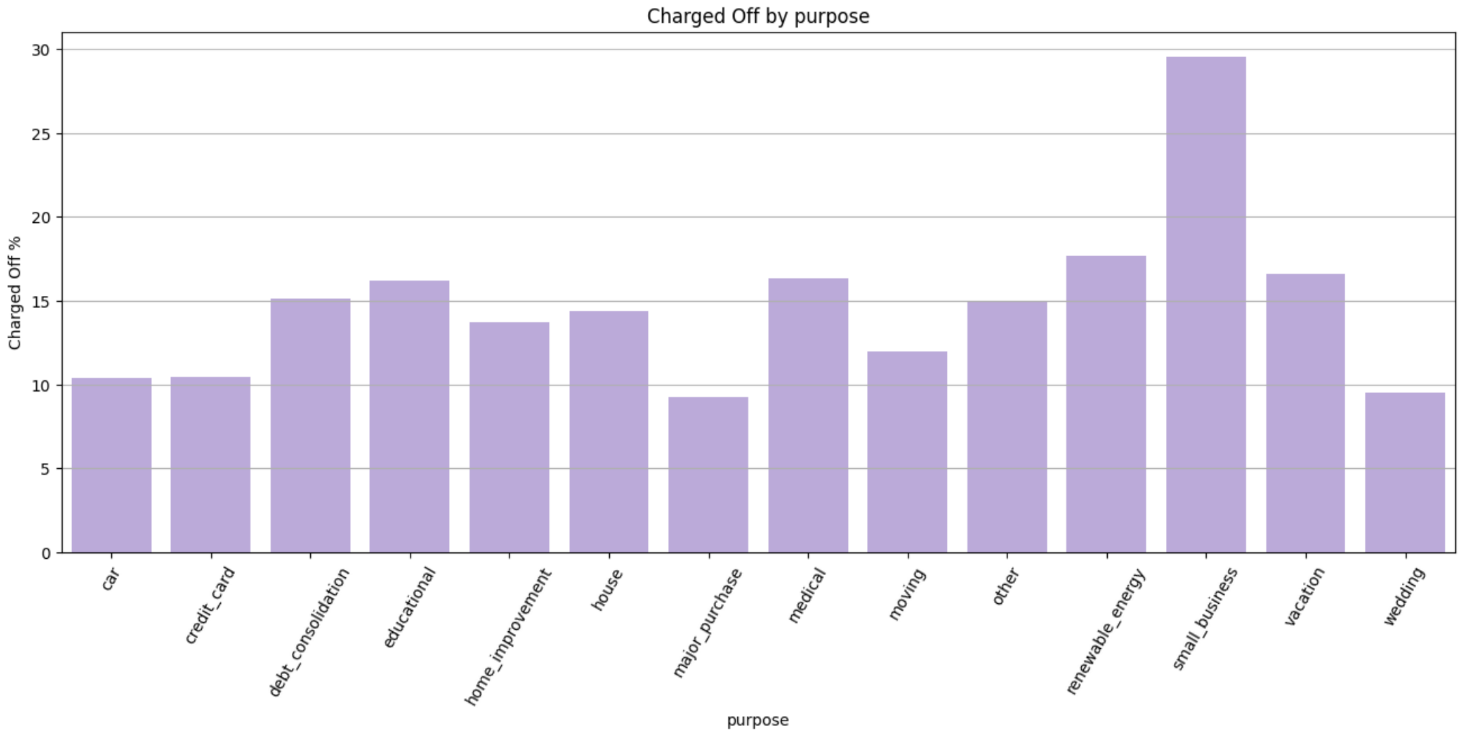
- Derived Metrics: Percentage of defaulters
- Surprisingly, loans that are not verified have a lower default rate compared to those that are verified.



DERIVED METRICS ANALYSIS

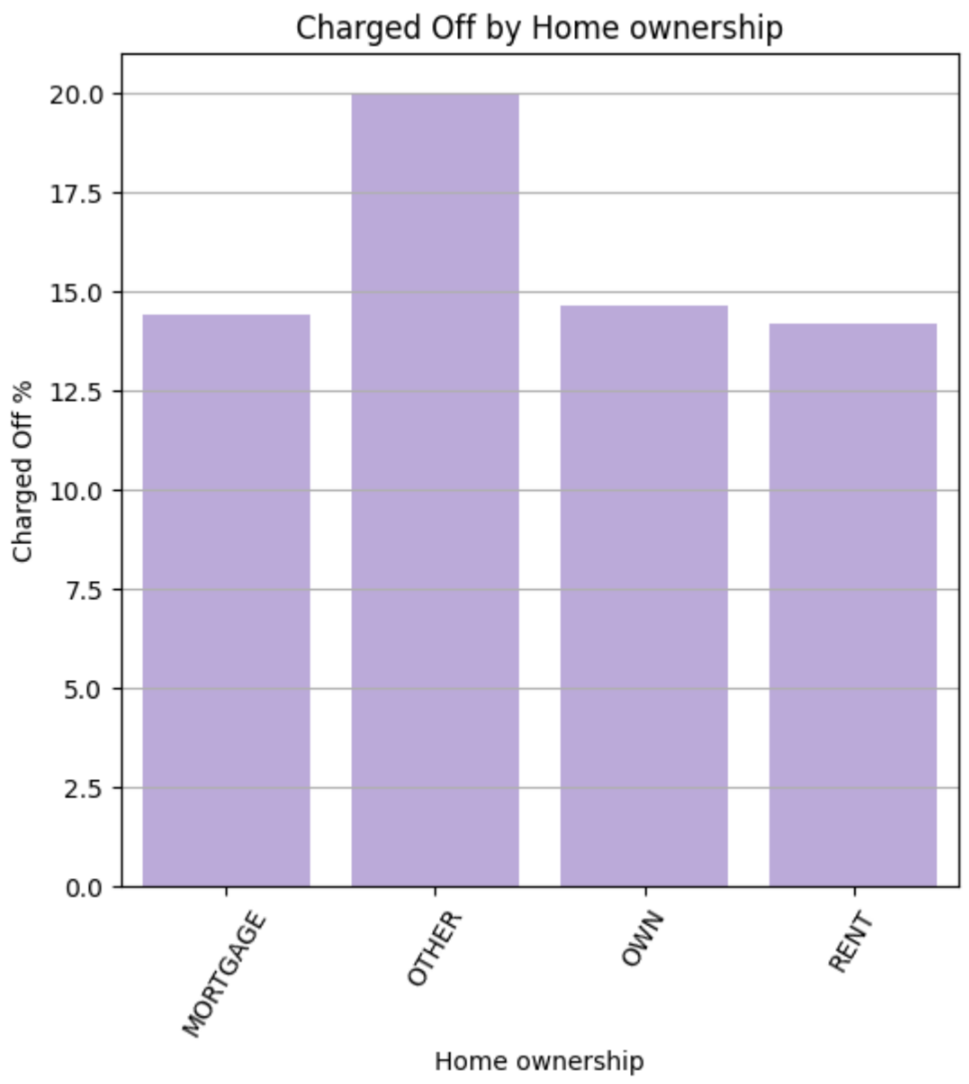
- Derived Metrics: Percentage of defaulters

- Loans taken for small business have highest defaulters rate. Least is for major purchases and wedding.



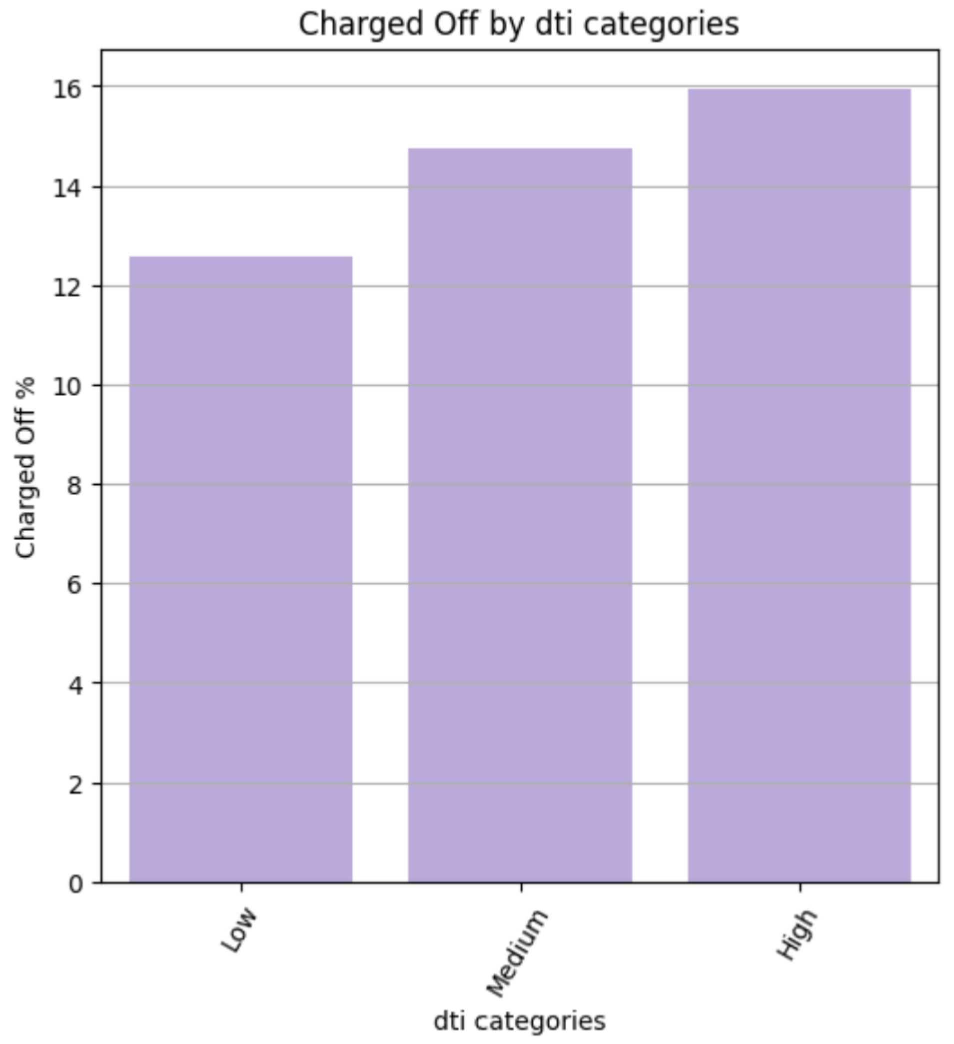
DERIVED METRICS ANALYSIS

- Derived Metrics: Percentage of defaulters
- People who have mentioned home ownership as Other have higher defaulter rate.



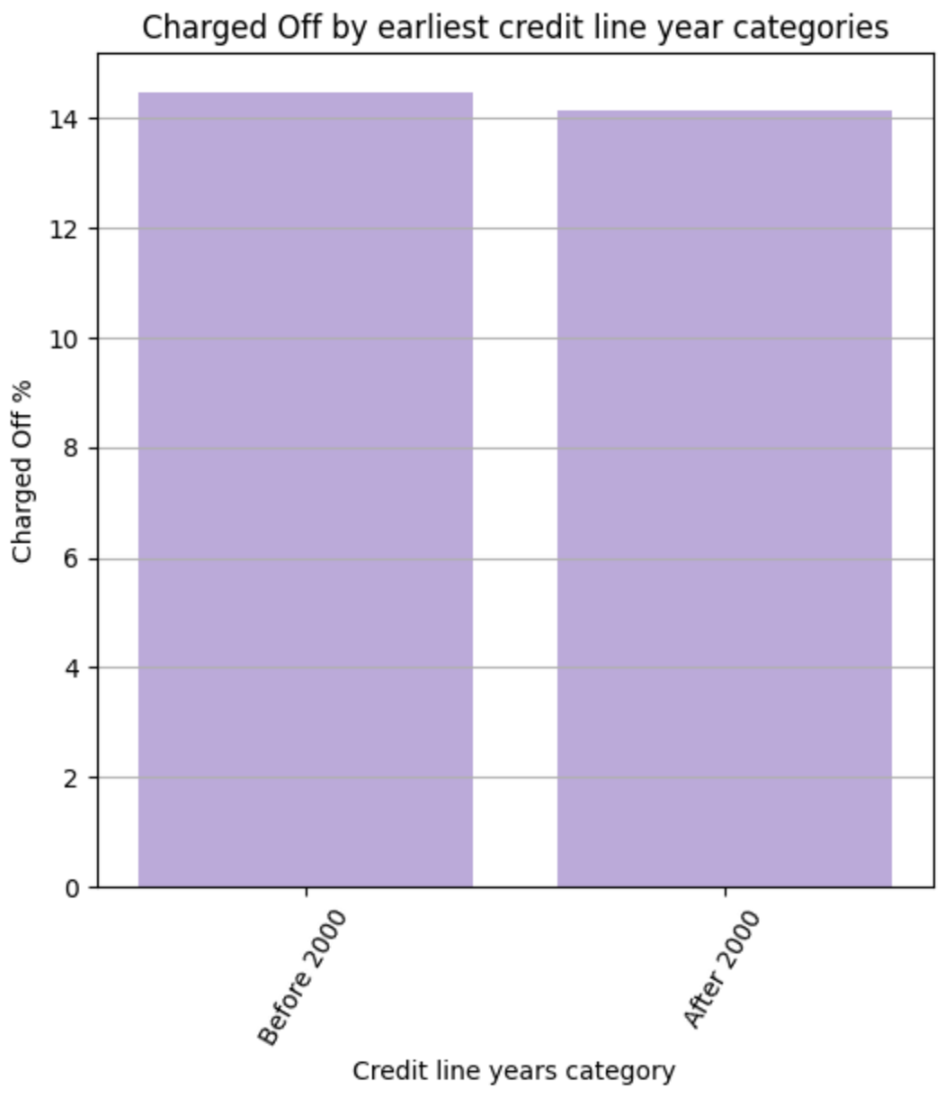
DERIVED METRICS ANALYSIS

- Derived Metrics: Percentage of defaulters
- People having low DTI have least defaulters percentage, followed by Medium DTI and finally by High DTI



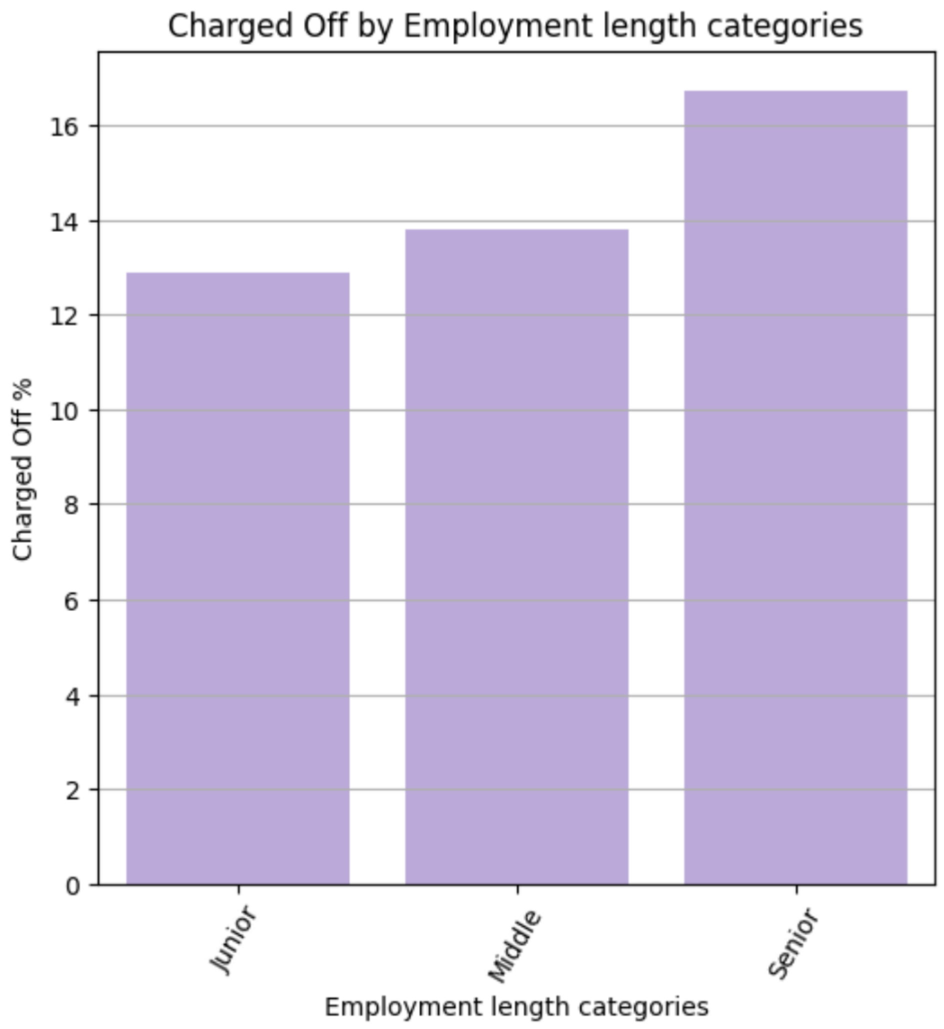
DERIVED METRICS ANALYSIS

- Derived Metrics: Percentage of defaulters
- Before 2000 percentage of defaulters was higher compared to after 2000



DERIVED METRICS ANALYSIS

- Derived Metrics: Percentage of defaulters
- Defaulters percentage is lowest in junior level employees, followed by middle level employees and finally, senior level employees.



KEY TAKEAWAYS

- **Driving Factors for Charge Offs:** Key factors include the loan grade, large loan amount, installment amount, term, and DTI.

- **Users Likely to Default:** These typically include
  - Users having high experience,
  - Users who have taken loans for small businesses,
  - Employees with ‘F’ and ‘G’ grades,
  - Users who have taken long-term loans
  - Users with high DTI.