

# Lending Club Case Study

Submitted by: Gaurav Emani and Govind Banura

---

# Table of Content

A thick yellow horizontal bar spans the width of the slide, with a vertical yellow bar on the right side.

- Case Synopsis
- Recommended Approach
- Data Exploration and Data Cleaning
- Exploratory Data Analysis (EDA)
  - Univariate Analysis
  - Bivariate Analysis
- Actions Recommended for the Bank



# Case Synopsis

- The case study revolves around a consumer finance company specializing in lending various types of loans to urban customers. When assessing loan applications, the company faces two primary risks:
  - approving a loan to an applicant who may default, resulting in financial loss, or
  - denying a loan to a reliable applicant, leading to a loss of business opportunity.
- The company aims to leverage Exploratory Data Analysis (EDA) to understand how consumer and loan attributes influence the likelihood of loan default.
- By effectively leveraging data analysis and machine learning techniques to identify patterns, the company aims to mitigate risks associated with lending, optimize portfolio management, and enhance decision-making processes.

.

---



# Recommended Approach

We have identified the following steps in this approach:

1. **Data Exploration:** Understand the dataset structure, features, and distribution of defaulters vs. non-defaulters.
  2. **Data Cleaning and Pre-processing:** Handle missing values, outliers, and pre-process data for analysis.
  3. **Exploratory Data Analysis (EDA):** Analyse relationships between variables (using Univariate and Bivariate Analysis), uncover trends, and assess the impact of attributes on loan default.
  4. **Interpretation and Insights:** Interpret model results to identify key indicators of loan default and provide actionable recommendations.
-



# DATA EXPLORATION AND DATA CLEANING

# DATA EXPLORATION & DATA CLEANING

- **Conclusion:** After pre-processing, the dataset now contains 45 columns and 36,847 rows, which are in a cleaner and more structured format, making it ready for further exploratory data analysis and modelling.
- Overall, these pre-processing and cleaning steps contribute to ensuring that the dataset is well-prepared for subsequent analytical tasks, enabling better insights and decision-making.

ACTION	ACTIONS PERFORMED	OUTCOME OF ACTION PERFORMED
Dropping Rows	<ol style="list-style-type: none"><li>1. Rows where the loan status is "Current" are removed as they are still in progress.</li><li>2. Duplicate rows are checked for but not found.</li></ol>	The final dataset after dropping rows is (38577, 111)
Dropping Columns	<ol style="list-style-type: none"><li>1. Columns with 100% missing values, unique values, and constant values are identified and dropped.</li><li>2. Columns with more than 60% missing values are removed</li></ol>	The final dataset after dropping columns is (38577, 37).
Data Conversion	<ol style="list-style-type: none"><li>1. The term column is converted to integer values by extracting only the month values.</li><li>2. The % sign is removed from the interest rate column and converted to float.</li><li>3. The issue_d column is converted from string object to DateTime format, and additional columns for year, month, and quarter are created.</li><li>4. The emp_length column is converted to integer values.</li><li>5. Rows/columns with NaN values in specific columns are removed.</li><li>6. Certain numeric columns are binned into different buckets for analysis.</li></ol>	The final dataset after conversion is (36847, 43).

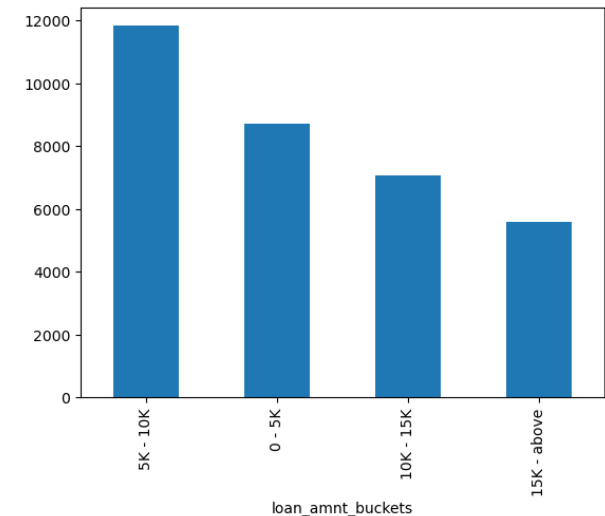
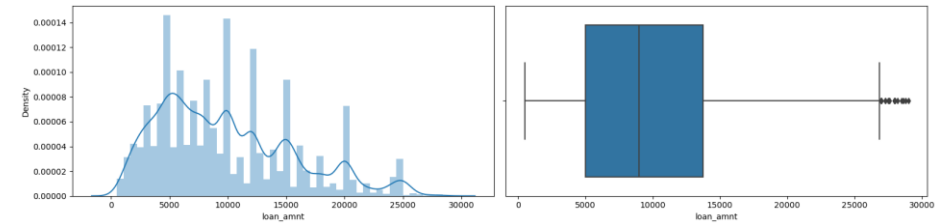
# EXPLORATORY DATA ANALYSIS (EDA)

## I. UNIVARIATE ANALYSIS

# Univariate Analysis

We have performed a thorough univariate analysis on various features of a dataset related to loans utilizing box plots, distribution plots, and bar plots to understand the distributions and characteristics of different variables. Here's a summary of our analysis:

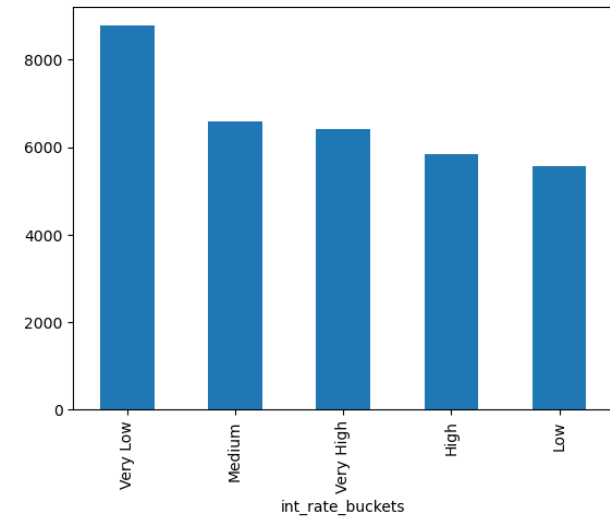
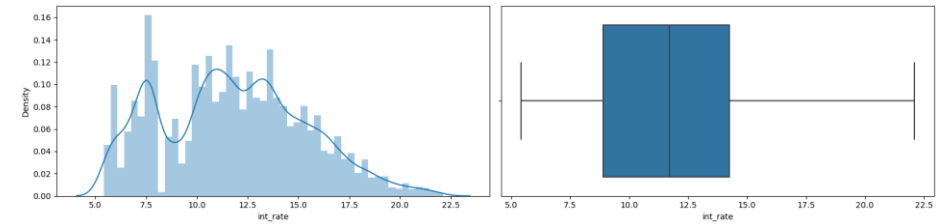
- 1. Loan Amount:** By the boxplot, we can see that most of the loan applications are for 5k to 14k amount. After creating multiple buckets for the Loan Amount, we can see that most of the loan are below 10 k .





# Univariate Analysis

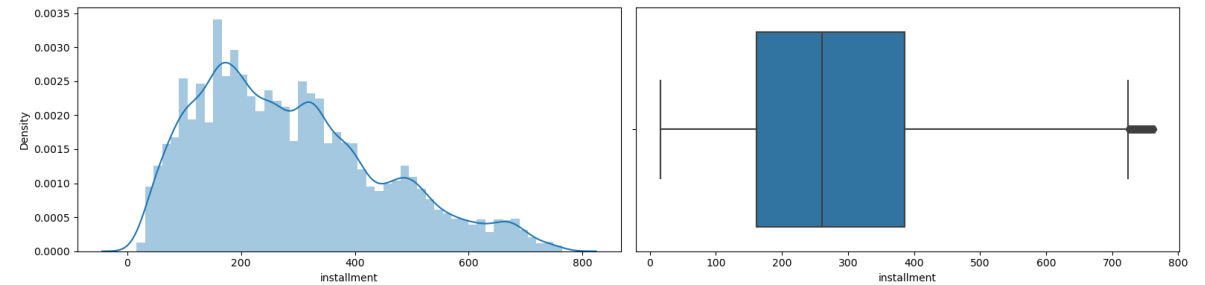
2. **Interest Rate:** Interest rates are starting from 5% and going till 22%, where most of the applications are in the range of 8.5% to 14.5%



# Univariate Analysis

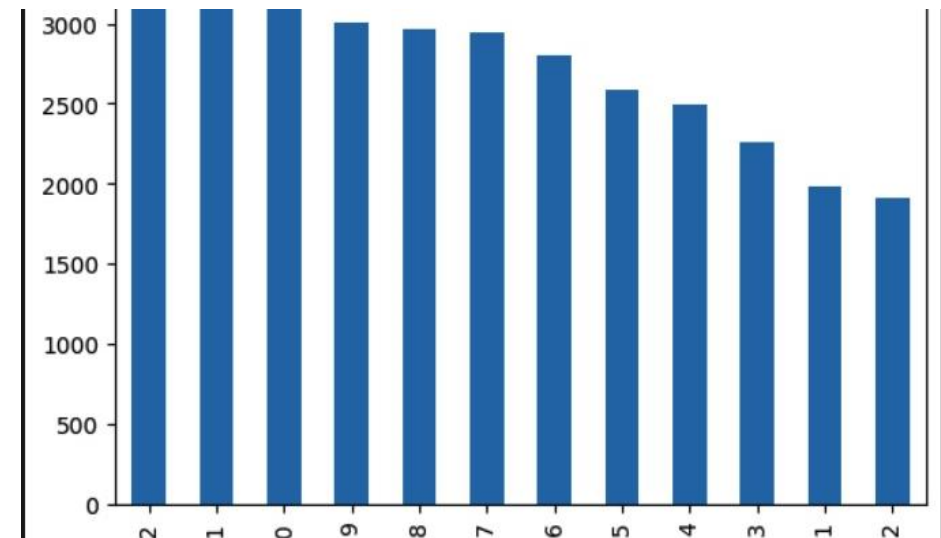
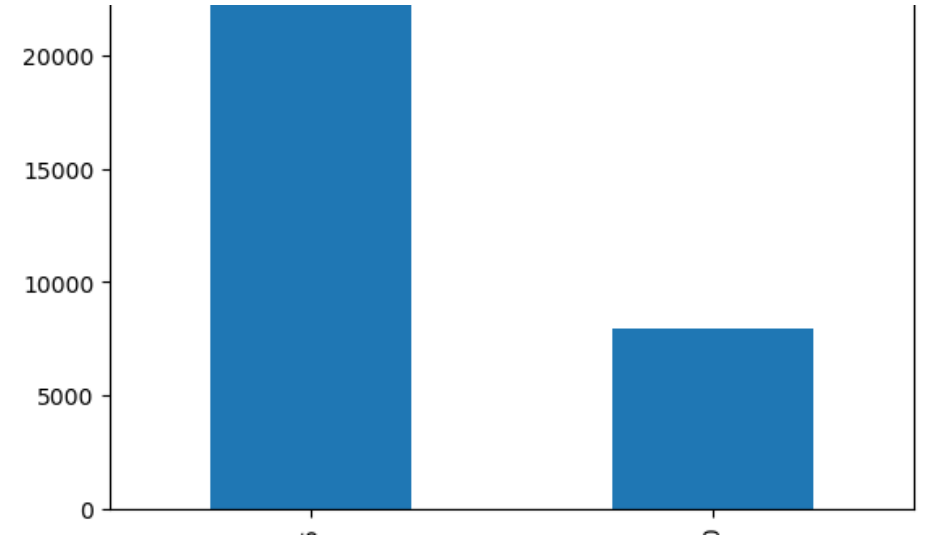
---

- 3. Installments:** Installments are going from 20 to 800 but majority lies between 160 to 400.



# Univariate Analysis

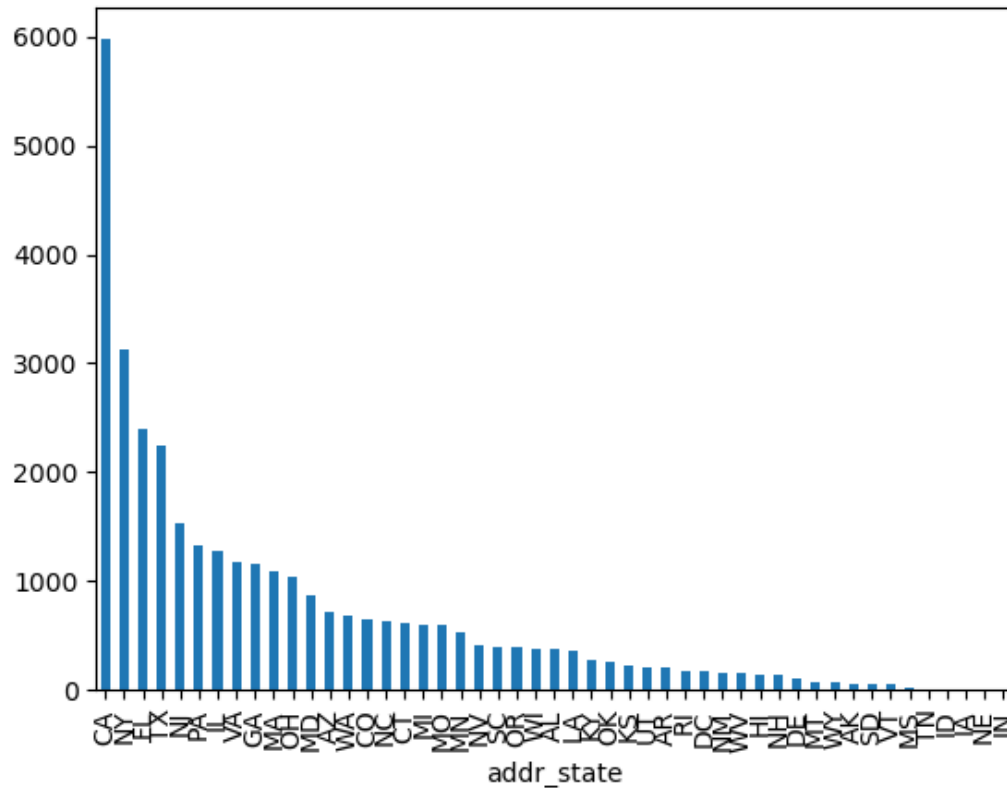
4. **Loan Term:** There are more application towards the year end for the loans 60% loans having a term of 36 months.



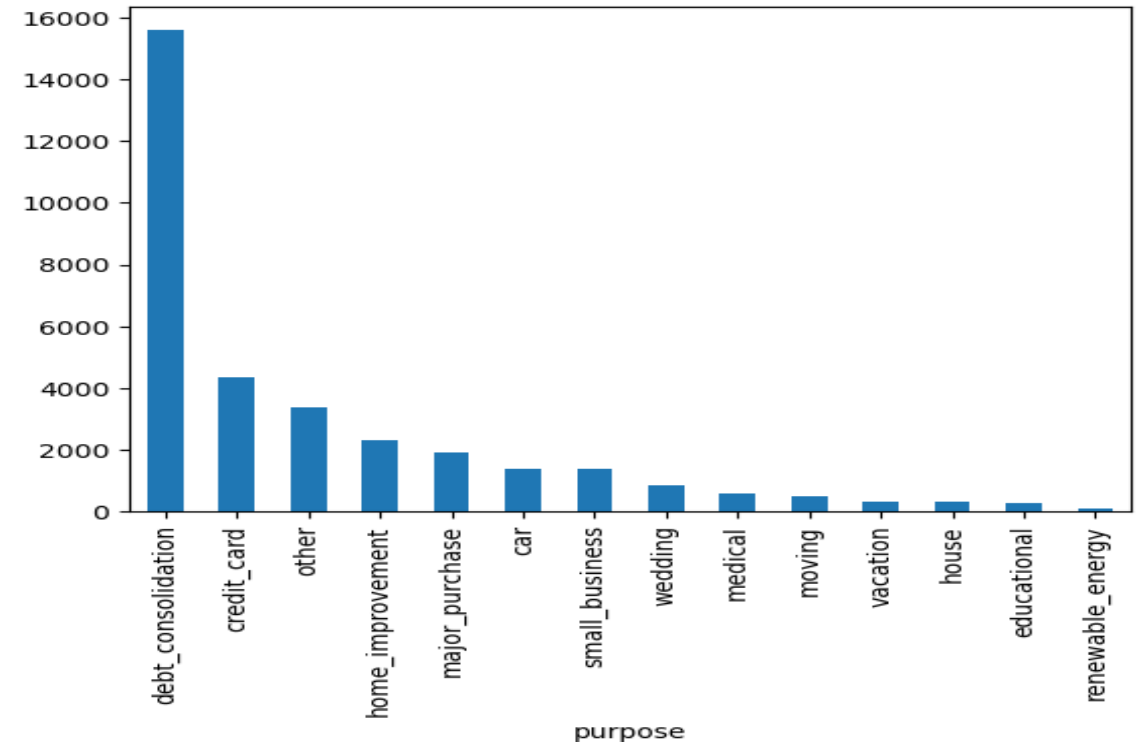
# Univariate Analysis



5. **State:** California have highest numbers of applicants; This means that the bank have to put more scrutiny on the applications.



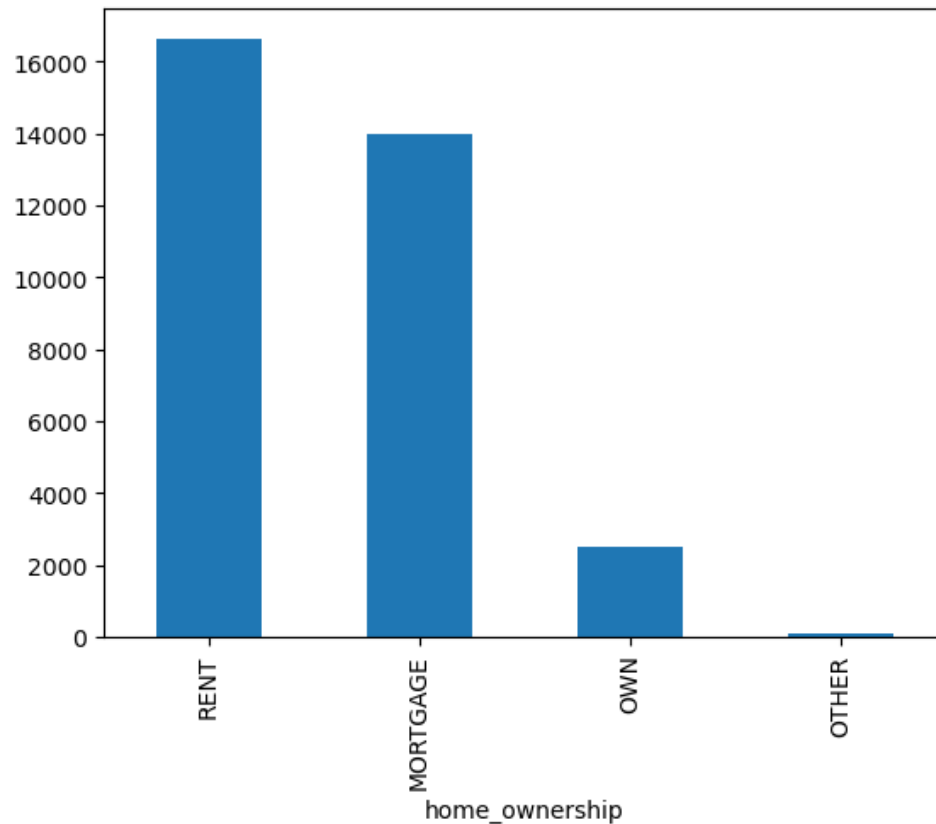
6. **Purpose:** Majority of the application of the loan is for DEBT\_CONSOLIDATION



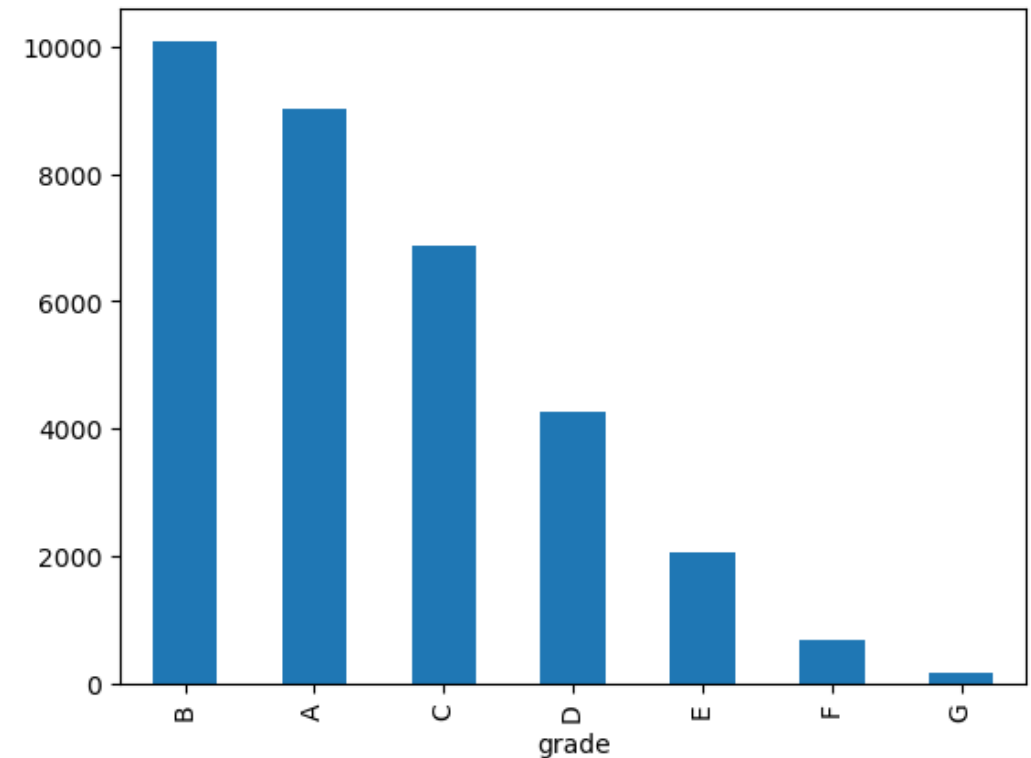
# Univariate Analysis



**7. Home Ownership:** Most of the applicants don't have their own house.



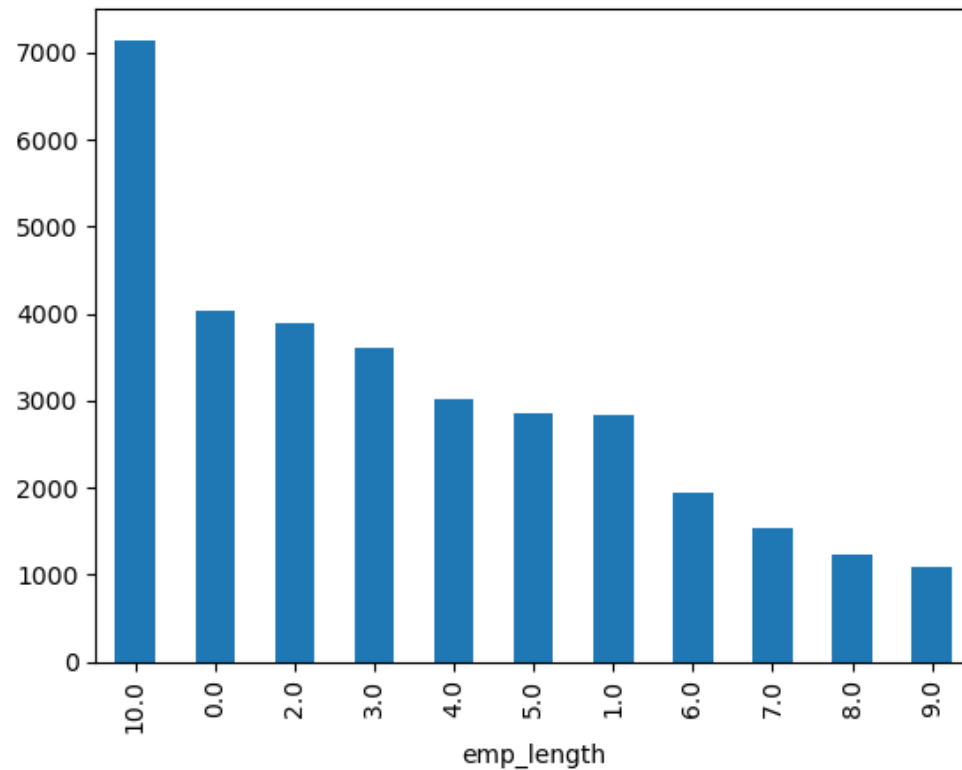
**8. Grade:** Majority of the applications are from Grade B, Second highest is grade A.



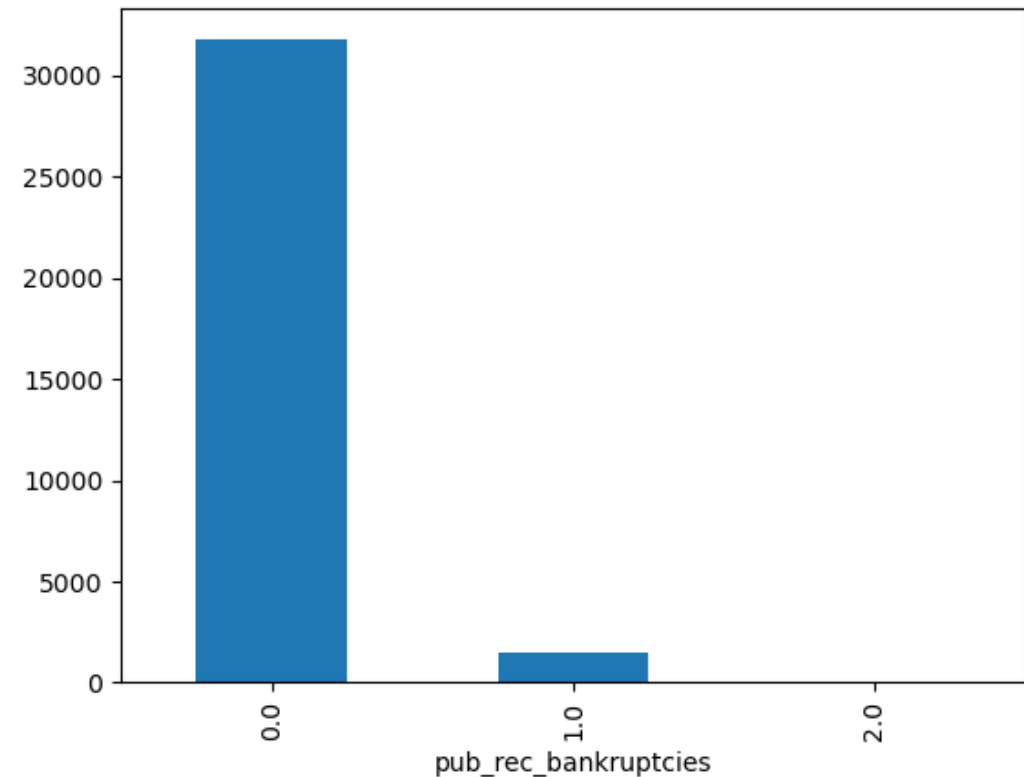
# Univariate Analysis



**9. Employment Length:** Most of the employment length is 10+ years.



**10. Public Records Bankruptcies:** 90% of the applicants are not bankrupt





11. **DTI (Debt-to-Income Ratio):** DTI Ranging from 0 to 800, with 50% of the data lies between 160 to 400.
12. **Year:** Loan applications are increasing every year.
13. **Month:** People apply for the loan towards end of the year, maybe due to holiday sessions.
14. **Loan Amount, Funded Amount, Funded Amount Inv:** These variables are highly correlated, with loan amount being the most informative. Hence, dropping funded amount related columns.
15. **Interest Rate Buckets:** Most loan applications have very low-interest rates.

---

## Univariate Analysis

These are additional factors as a part of univariate analysis:

# EXPLORATORY DATA ANALYSIS (EDA)

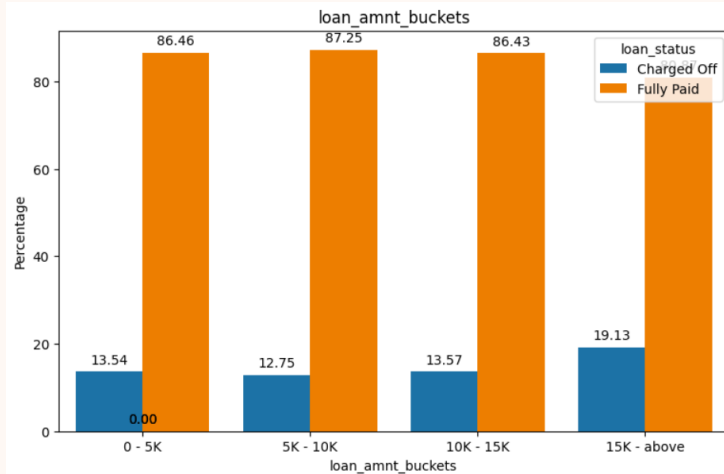
## II. BIVARIATE ANALYSIS



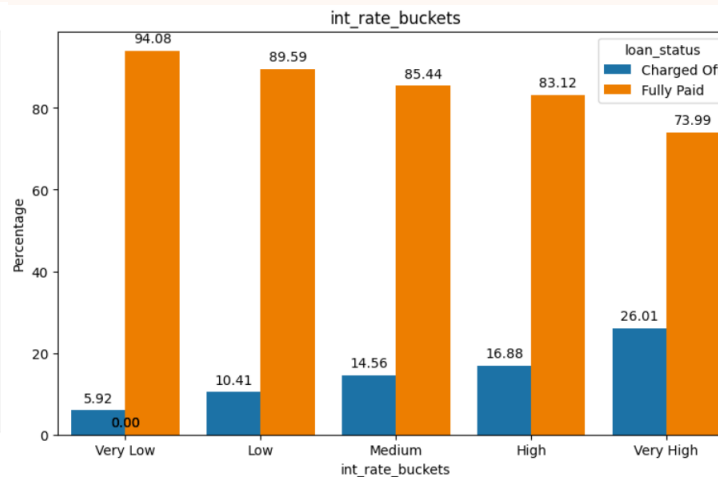
# Bivariate Analysis | Factors impacting Defaulters

Here's a summary of univariate analysis using bar plots to understand the distributions and characteristics of different variables.

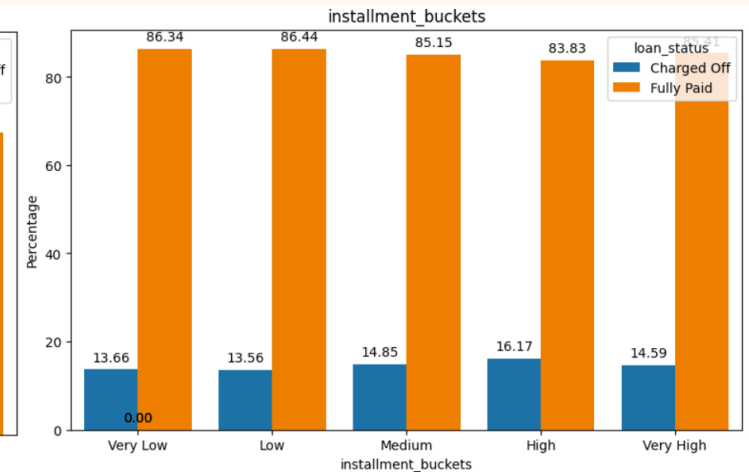
- 1. Loan Status Vs Loan Amount:** The graph shows that if loan amount is high, it is more likely to be charged off



- 2. Interest rate Vs Loan Status:** The graph shows that there is high chance of the loan to get default if the interest rate is high

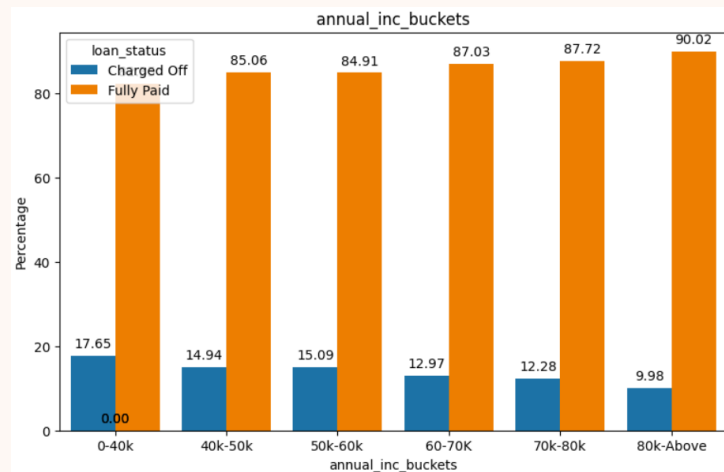


- 3. Instalment Bucket rate Vs Loan Status:** high loan amount can lead to higher number of defaulters

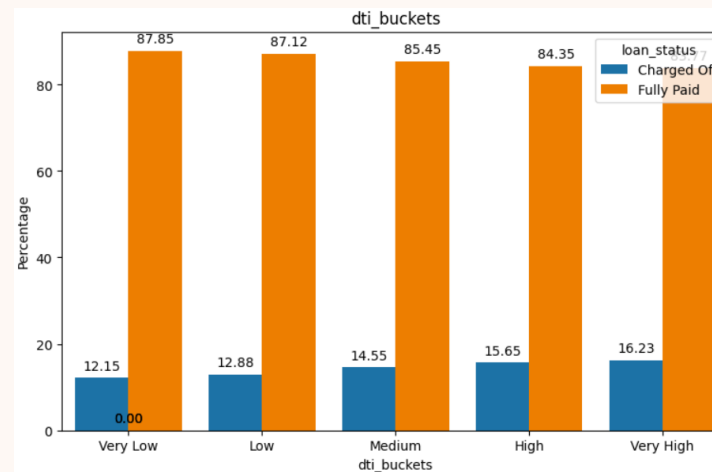


# Bivariate Analysis | Factors impacting Defaulters

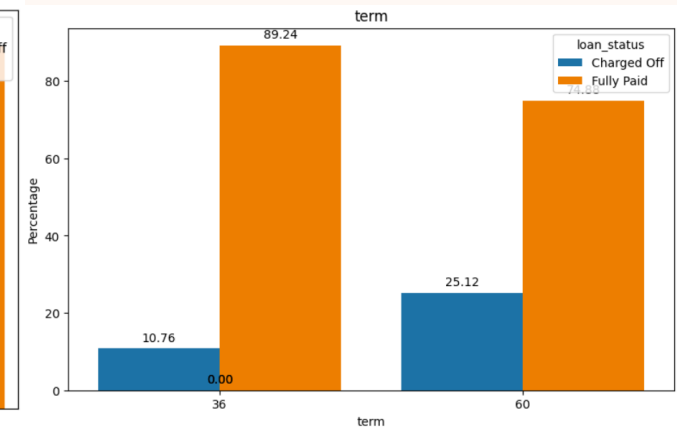
4. **Annual income Vs Loan Status:** High annual income tends to have less loan defaulters



5. **DTI buckets Vs Loan Status:** High Debt to income ratio can lead to high defaulters

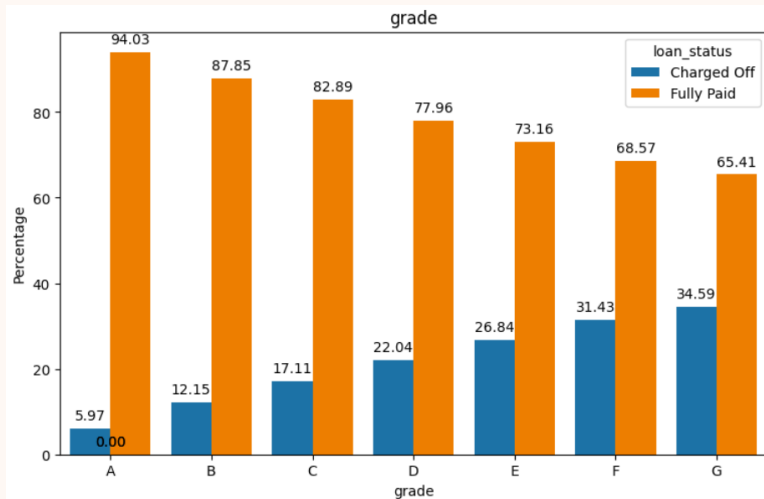


6. **Loan Term rate Vs Loan Status:** 5 years loan term duration have high defaulters

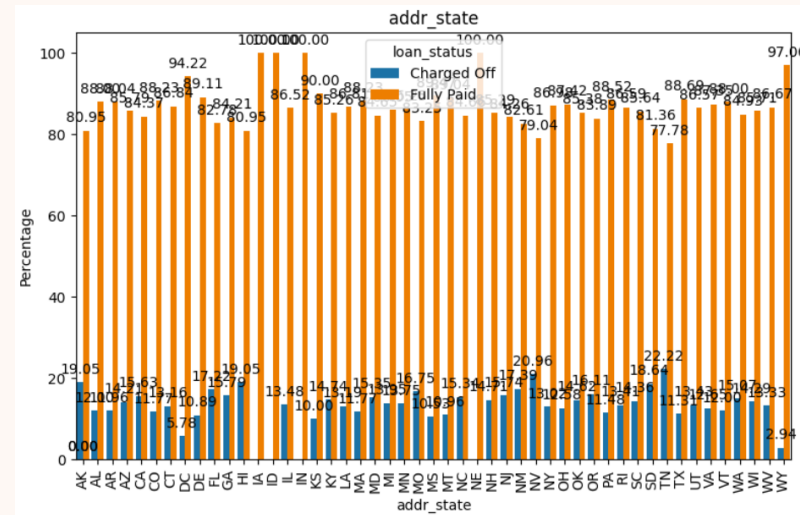


# Bivariate Analysis | Factors impacting Defaulters

**7. Grade Vs Loan Status:** Loan grades having highest default percentage. High grade have higher default rate



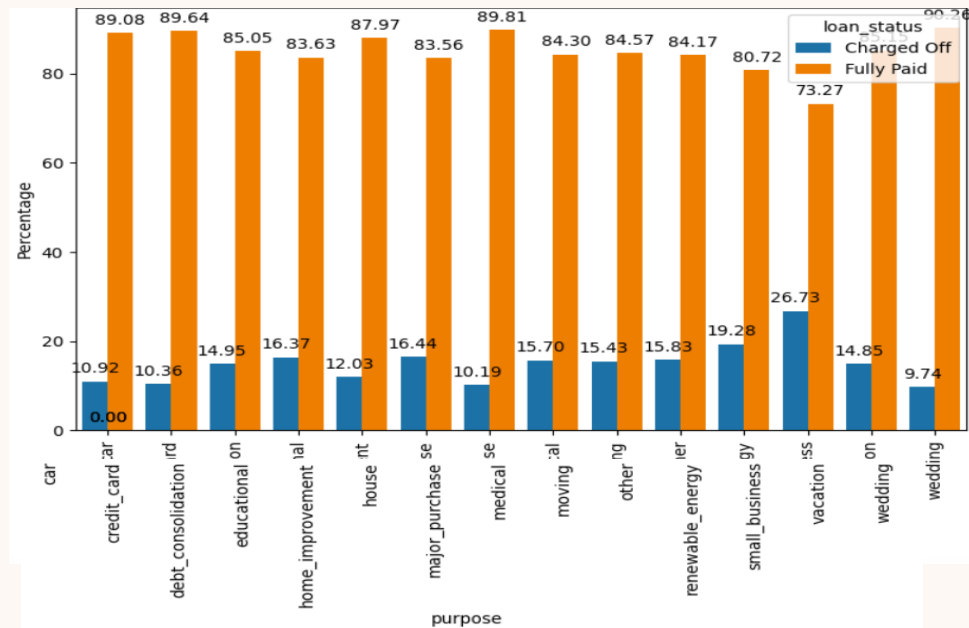
**8. Addr\_State Vs Loan Status:** Applications from TN, NV, AK, HI, SD are the most risky



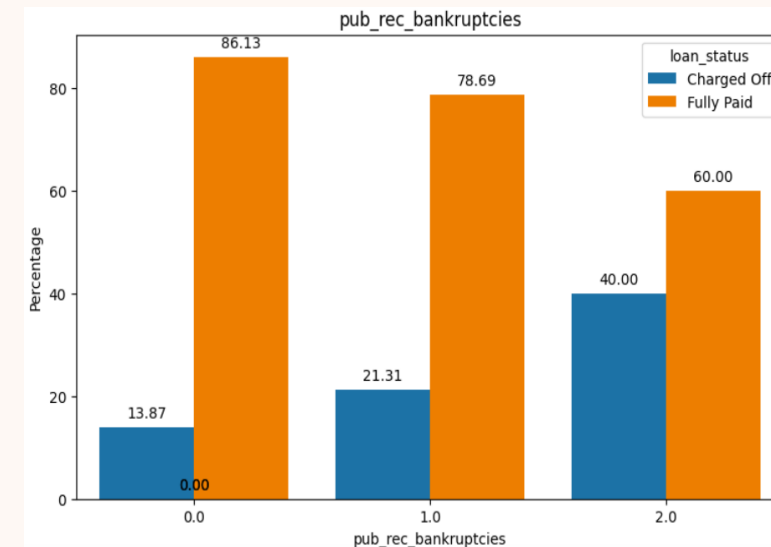
addr_state	loan_status	Count	Total_Count	Percentage
TN	Charged Off	2	9	22.222222
NV	Charged Off	87	415	20.963855
AK	Charged Off	12	63	19.047619
HI	Charged Off	28	147	19.047619
SD	Charged Off	11	59	18.644068

# Bivariate Analysis | Factors impacting Defaulters

**9. Loan Purpose Vs Loan Status:** Small Business, educational loans and renewable energy have high number of defaulters



**10. Public Record Bankruptcies Vs Loan Status:** People with high rate of bankruptcies are more defaulters



# Actions Recommended for the Bank

Bank should consider the below factors identified in Bivariate Analysis before approving a loan

1. **Loan Amount:** Higher loan amounts are more likely to be charged off. This suggests that the bank may need to reassess its risk assessment strategies for larger loans.
2. **Interest Rate:** Loans with higher interest rates have a higher chance of default. The bank should consider adjusting interest rates based on risk assessment to mitigate defaults.
3. **Instalment Amount:** Higher instalment amounts correlate with higher default rates. The bank might want to review its loan terms and conditions to ensure borrowers can comfortably afford repayments.
4. **Annual Income:** Borrowers with higher annual incomes tend to default less. The bank may consider setting income thresholds for loan eligibility or adjusting loan terms based on income levels.
5. **Debt-to-Income Ratio (DTI):** Higher DTI ratios are associated with higher default rates. The bank should carefully evaluate borrowers' DTI ratios when assessing loan applications.
6. **Loan Term:** Loans with longer terms, particularly 5-year terms, have higher default rates. The bank may need to adjust its lending criteria for longer-term loans or closely monitor these loans for potential defaults.

# Actions Recommended for the Bank

Overall, the bank should use these insights to refine its risk assessment models, adjust lending policies, and tailor loan terms to minimize default risk and maintain a healthy loan portfolio.

7. **Loan Grade:** Higher loan grades show higher default percentages. The bank should revisit its grading system and potentially tighten lending criteria for higher grade loans
8. **Purpose of Loan:** Loans for small businesses, renewable energy, and education have higher default rates. The bank might want to reassess its risk assessment for these specific loan purposes.
9. **State of Residence:** Applications from certain states like TN, NV, AK, HI, and SD are riskier. The bank should consider regional economic factors and adjust its lending policies accordingly.
10. **Public Record Bankruptcies:** Individuals with a history of bankruptcies are more likely to default. The bank should factor in past financial history when assessing loan applications and consider additional measures for applicants with bankruptcy records.