LENDING CLUB CASE STUDY

Submitted by:

HariKrishna Kotha

Hemant R

Contents

- Problem Statement
- Data Description
- Data Understanding
- Data Cleaning & Pre-processing
- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
- Correlation Analysis
- Suggestions

Problem Statement

Lending Club, a Consumer Finance marketplace specializing in offering a variety of loans to urban customers, faces a critical challenge in managing its loan approval process. When evaluating loan applications, the company must make sound decisions to minimize financial losses, primarily stemming from loans extended to applicants who are considered **"Risky".**

- These financial losses, referred to as **Credit Losses**, occur when borrowers fail to repay their loans or default. In simpler terms, borrowers labelled as **"Charged-Off"** are the ones responsible for the most significant losses to the company.

- The primary objective of this exercise is to assist Lending Club in mitigating credit losses. This challenge arises from two potential scenarios:

  1. **Identifying applicants likely to repay their loans is crucial, as they can generate profits for the company through interest payments. Rejecting such applicants would result in a loss of potential business.**

  2. **On the other hand, approving loans for applicants not likely to repay and at risk of default can lead to substantial financial losses for the company.**

- The objective is to pinpoint applicants at risk of defaulting on loans, enabling a reduction in credit losses. This case study aims to achieve this goal through Exploratory Data Analysis (EDA) using the provided dataset.

- In essence, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

DataDescription

- Lending Club provided us with customer’s historical data. This dataset contained information pertaining to the borrower’s past credit history and ending Club loan information. The total dataset consisted of over 39717 records and 111 columns, which was sufficient for our team to conduct analysis. Variables present within the dataset provided an ample amount of information which we could use to identify relationships and gauge their effect upon the success or failure of a borrower fulfilling the terms of their loan agreement.
- There are lot of columns with no data available, so only considered the columns which have the data for analysis

Data Understanding

Dataset Facts:

Primary Fact:

Loan Status: The Principal Attribute of Interest (loan_status). This column consists of three distinct values:

- ✓ **Fully-Paid:** Signifies customers who have successfully repaid their loans.
- ✓ **Charged-Off:** Indicates customers who have been labeled as "Charged-Off" or have defaulted on their loans.
- ✓ **Current:** Represents customers whose loans are presently in progress and, thus, cannot provide conclusive evidence regarding future defaults.

For the purposes of this case study, rows with a "Current" status will be excluded from the analysis.

Decision Matrix:

Loan Acceptance Outcome - There are three potential scenarios:

Fully Paid - This category represents applicants who have successfully repaid both the principal and the interest rate of the loan.

Current - Applicants in this group are actively in the process of making loan installments; hence, the loan tenure has not yet concluded. These individuals are not categorized as 'defaulted.'

Charged-off - This classification pertains to applicants who have failed to make timely installments for an extended period, resulting in a 'default' on the loan.

Loan Rejection - In cases where the company has declined the loan application (usually due to the candidate not meeting their requirements), there is no transactional history available for these applicants. Consequently, this data is unavailable to the company and is not included in this dataset.

Data Understanding

Key Dimension of table analyzed and segregated by demography

Customer	Loan
Annual income	Loan Amount
Home ownership	Grade
Employment Length	Term
Debt To income	Loan Date
State	Purpose Of loan
	Verification Status
	Interest Rate
	Installment

Data Cleaning & Pre-processing

1. Loading data from loan CSV
2. Check for null values in the dataset
3. Check for unique values
4. Check for duplicated rows in data
5. Dropping Unnecessary Records & Columns
6. Common Functions
7. Data Conversion
8. Outlier Treatment
9. Imputing values in Columns

Data Cleaning & Pre-processing

1. **Loading data from loan CSV: After loading data into notebook identified some data types are not what expected.** So changed the data types of the columns
2. **Checking for null values in the dataset:** There're many columns with null values. Dropped them as they can't contribute any kind of analysis. Most of the columns in given data has null values.
3. **Checking for unique values:** If the column has only a single unique value, it does not make any sense to include it as part of our data analysis. We need to find out those columns and drop them from the dataset. 9 columns had such unique values and they were removed.
4. **Checking for duplicated rows in data:** No duplicate rows were found.
5. **Dropping Records and Columns:**
   1. Dropped records where **loan_status="Current"**. Current loan status applications will not be of interest to us as they are still in repayment mode.
   2. Dropping columns where missing data is **>=65%** as these columns will skew our data analysis and they need to be removed.
   3. Dropping extra columns containing text like **collection_recovery_fee, delinq_2yrs, desc, earliest_cr_line, emp_title, id, inq_last_6mths, last_credit_pull_d, last_pymnt_amnt, last_pymnt_d, member_id, open_acc, out_prncp, out_prncp_inv, pub_rec, recoveries, revol_bal, revol_util, title, total_acc, total_pymnt, total_pymnt_inv, total_rec_int, total_rec_late_fee, total_rec_prncp, url, zip_code** as these will not contribute to loan pass or fail.

Data Cleaning & Pre-processing

6. **Common Functions:** Common functions were created for repeating common operations like plotting bar graphs, box plots, histograms, countplots, binning etc.
7. **Data Conversion**: Converted columns like **debt to income (dti), funded amount (funded_amnt), funded amount investor (funded_amnt_inv) and loan amount (loan_amnt) to float** to match the data. Also converted **loan date (issue_d)** to **DateTime (format: yyyy-mm-dd).**
8. **Outlier Treatment:** Calculated the **Inter-Quartile Range (IQR)** and filtering out the outliers outside of lower and upper bound. During Outlier analysis the following observations were made
   - ✓ The annual income of most of the loan applicants is between 40K - 75K USD
   - ✓ The loan amount of most of the loan applicants is between 5K - 15K
   - ✓ The funded amount of most of the loan applicants is between 5K - 14K USD
   - ✓ The funded amount by investor for most of the loan applicants is between 5K - 14K USD
   - ✓ The interest rate on the loan is between 9% - 14%
   - ✓ The monthly installment amount on the loan is between 160 - 440
   - ✓ The debt to income ration is between 8 - 18

Data Cleaning & Pre-processing

9. **Imputing values in Columns:**

   - ✓ **Replaced missing values of annual_inc with the corresponding mode value of annual_inc of the emp_length annual_inc field:** They Employment length has **1015** missing values, which means either they are **not employed or self-employed (business owners).** Considering they have a decent average annual income, we have assumed that these are business owners and we have added their employment duration with the mode value of **emp_length** which is **10+ years**.
   - ✓ Mapped employment length with the respective number of years in int.
   - ✓ Imputed **NONE** values as **OTHER** for **home_ownership.**
   - ✓ Replaced the '**Source Verified**' values as '**Verified**' since both values mean the same thing i.e. the loan applicant has some source of income which is verified.
   - ✓ There are **660 null values** for **pub_rec_bankruptcies**. Dropped those rows as they cannot be imputed.

Post Data cleaning and Pre-processing of dataset, we were left with **36094** rows × **18** columns.

Univariate Analysis

- ✓ **Univariate analysis** is a statistical method used to analyze and summarize data sets consisting of **one variable**. It deals with the analysis of a single variable, rather than multiple variables, to understand its distribution, central tendency and dispersion.
- ✓ It was carried out for both **Categorical** and **Quantitative** Variables

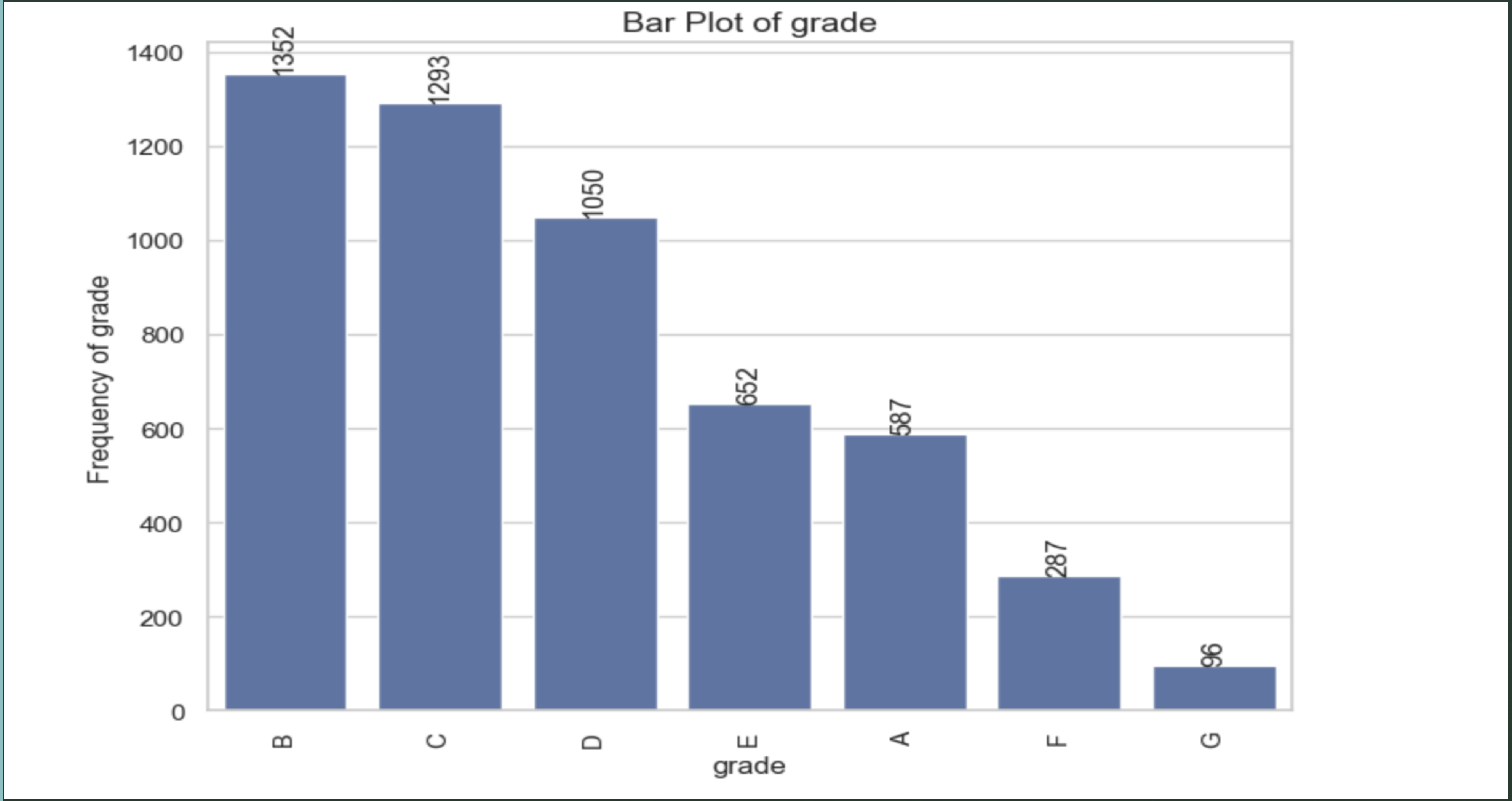
A. Categorical Variables:

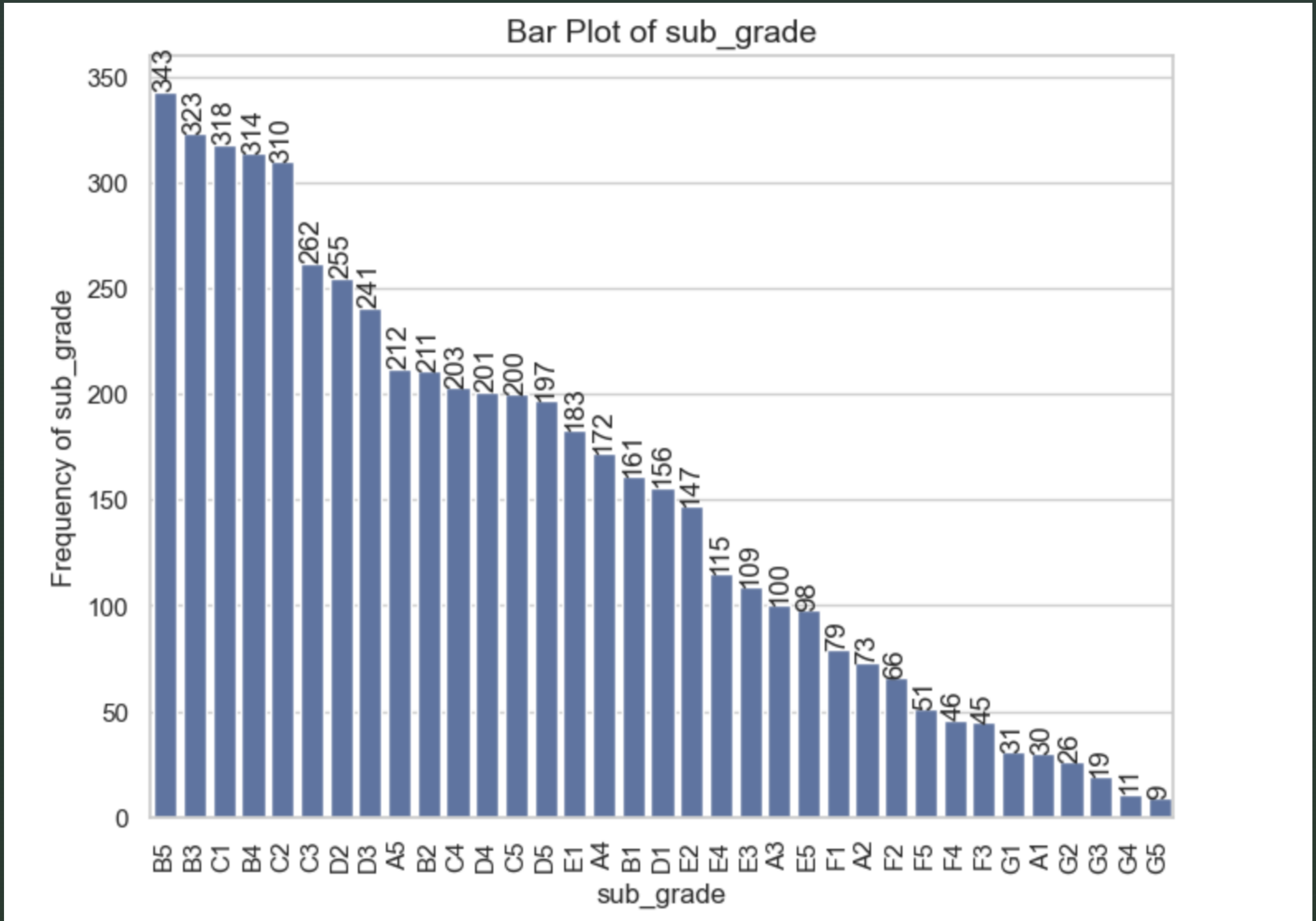
Ordered	Unordered
✓ Grade (grade) ✓ Sub grade (sub_grade) ✓ Term (36 / 60 months) (term) ✓ Employment length (emp_length) ✓ Issue year (issue_y) ✓ Issue month (issue_m) ✓ Issue quarter (issue_q)	✓ Address State (addr_state) ✓ Loan purpose (purpose) ✓ Home Ownership (home_ownership) ✓ Loan status (loan_status) ✓ Loan paid (loan_paid)

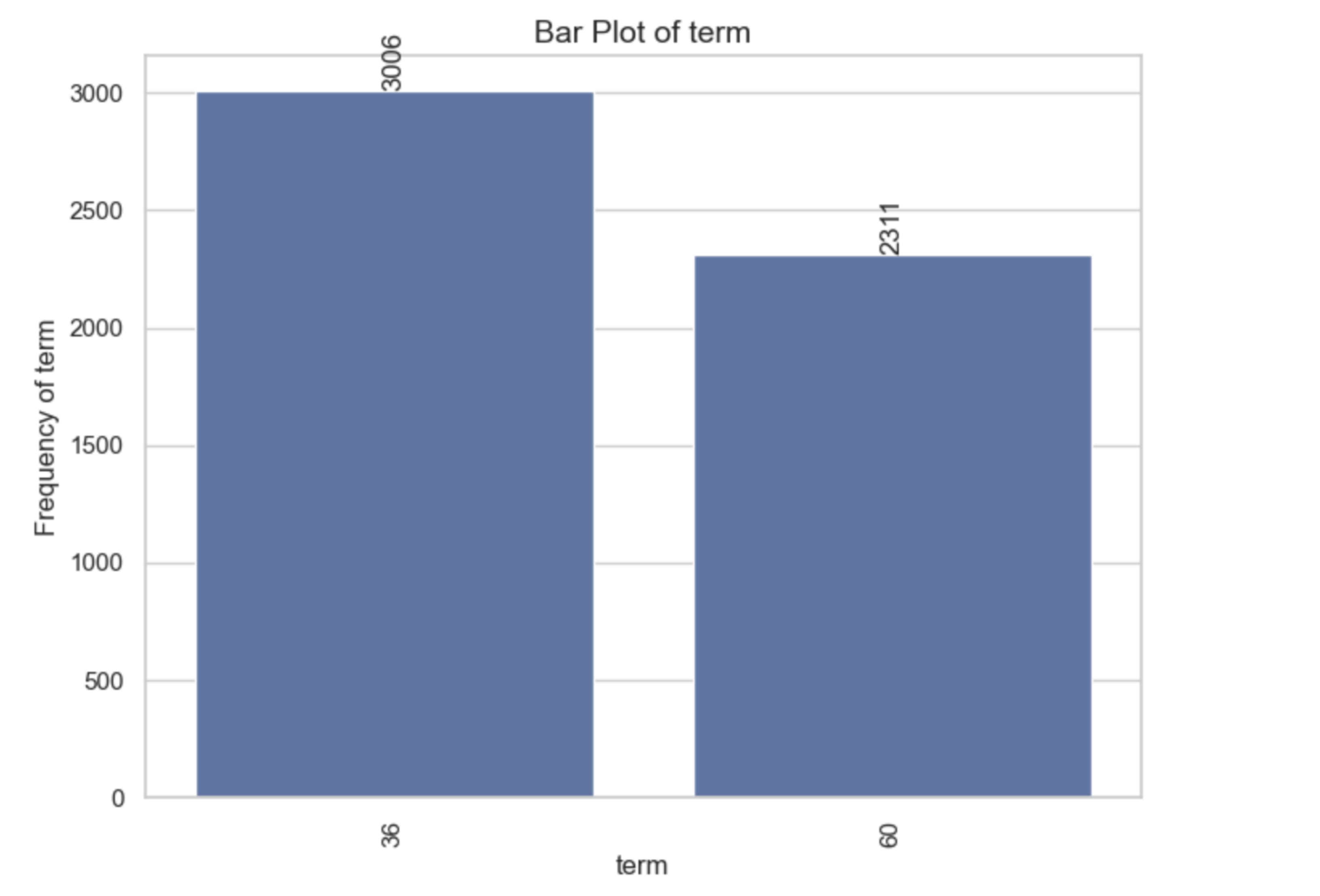
B. Quantitative Variables:

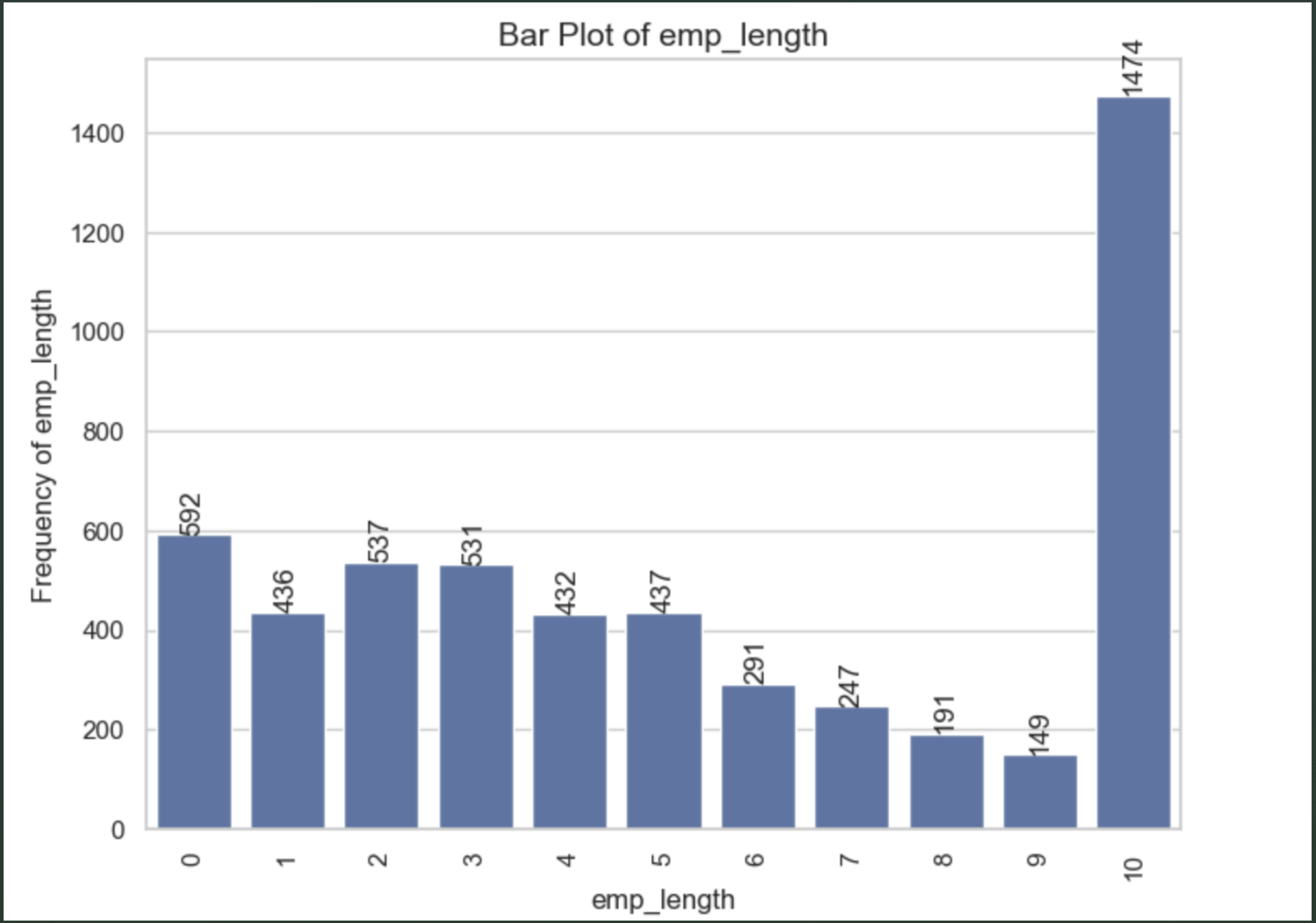
- ✓ Interest rate bucket (int_rate_bucket)
- ✓ Annual income bucket (annual_inc_bucket)
- ✓ Loan amount bucket (loan_amnt_bucket)
- ✓ Funded amount bucket (funded_amnt_bucket)
- ✓ Debt to Income Ratio (DTI) bucket (dti_bucket)
- ✓ Monthly Installment (installment)

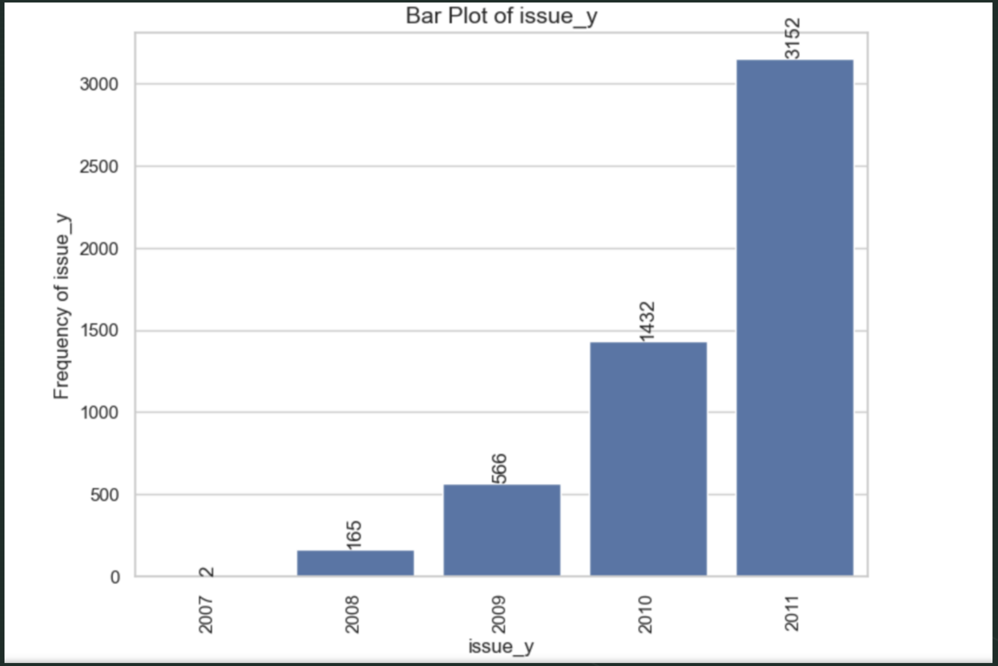
Univariate Analysis (Unordered Categorical)

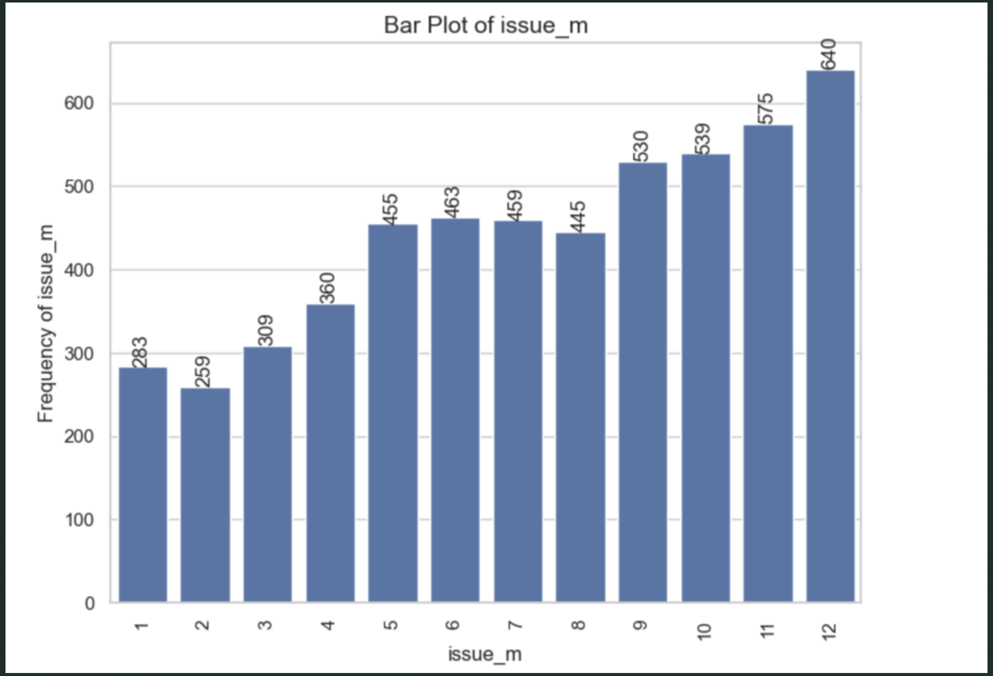


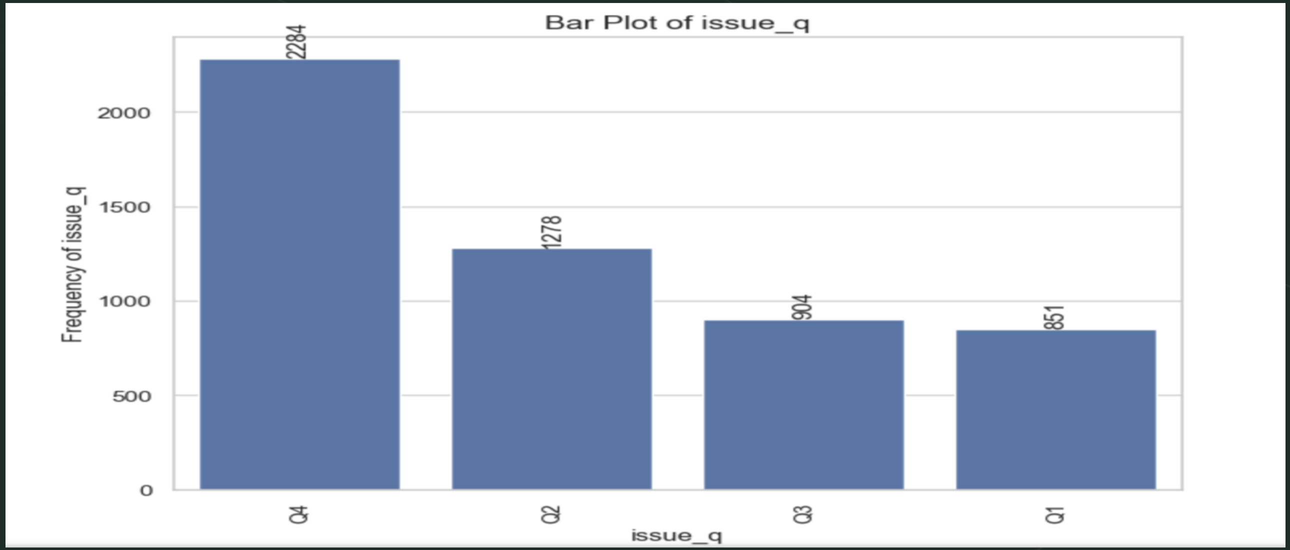

Bar Plot of sub_grade
Frequency of sub_grade
350
300
250
200
150
100
50
0
sub_grade
B5 343
B3 323
C1 318
B4 314
C2 310
C3 262
D2 255
D3 241
A5 212
B2 211
C4 203
D4 201
C5 200
D5 197
E1 183
A4 172
B1 161
D1 156
E2 147
E4 115
E3 109
A3 100
E5 98
F1 79
A2 73
F2 66
F5 51
F4 46
F3 45
G1 31
A1 30
G2 26
G3 19
G4 11
G5 9

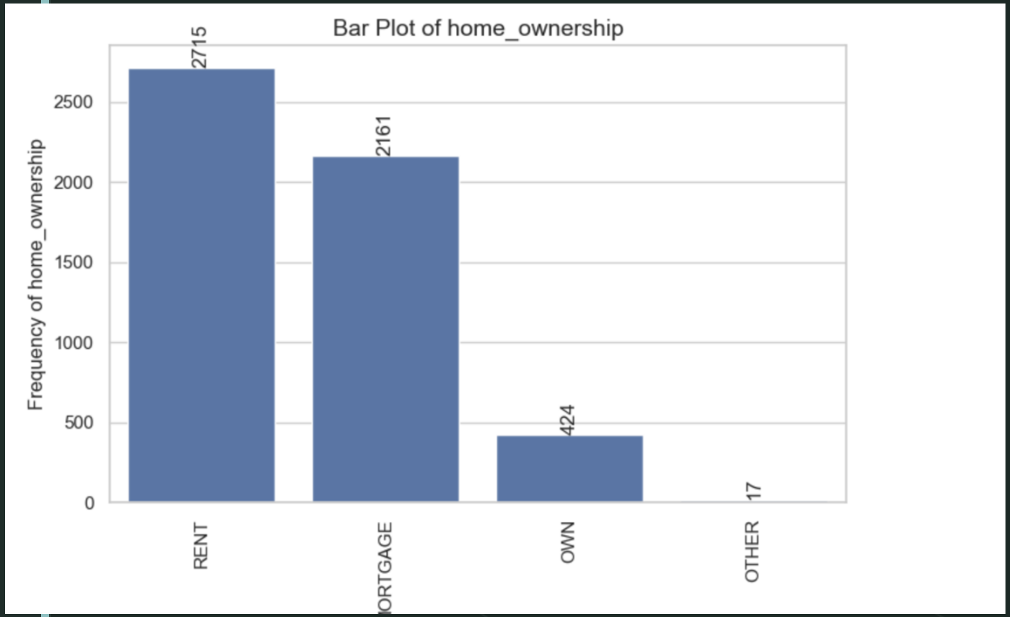

Bar Plot of term
Frequency of term
3000
2500
2000
1500
1000
500
0
3006
2311
36
60
term

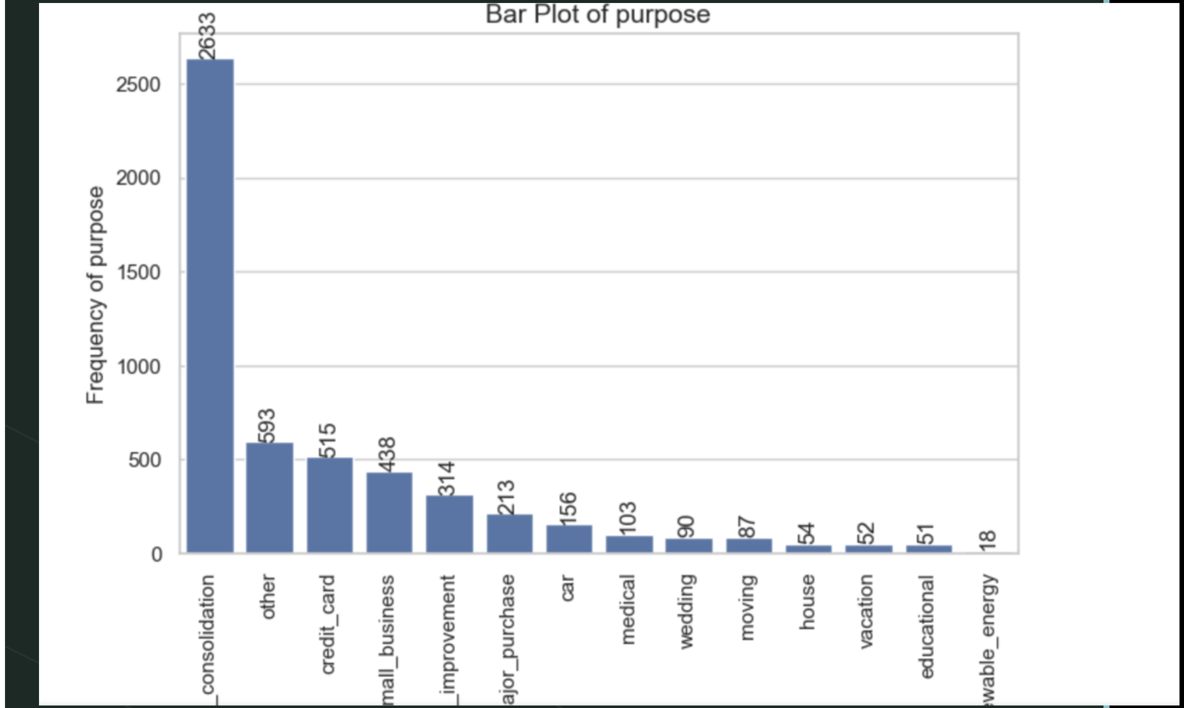

Bar Plot of emp_length
Frequency of emp_length
0
200
400
600
800
1000
1200
1400
592
436
537
531
432
437
291
247
191
149
1474
0
1
2
3
4
5
6
7
8
9
10
emp_length

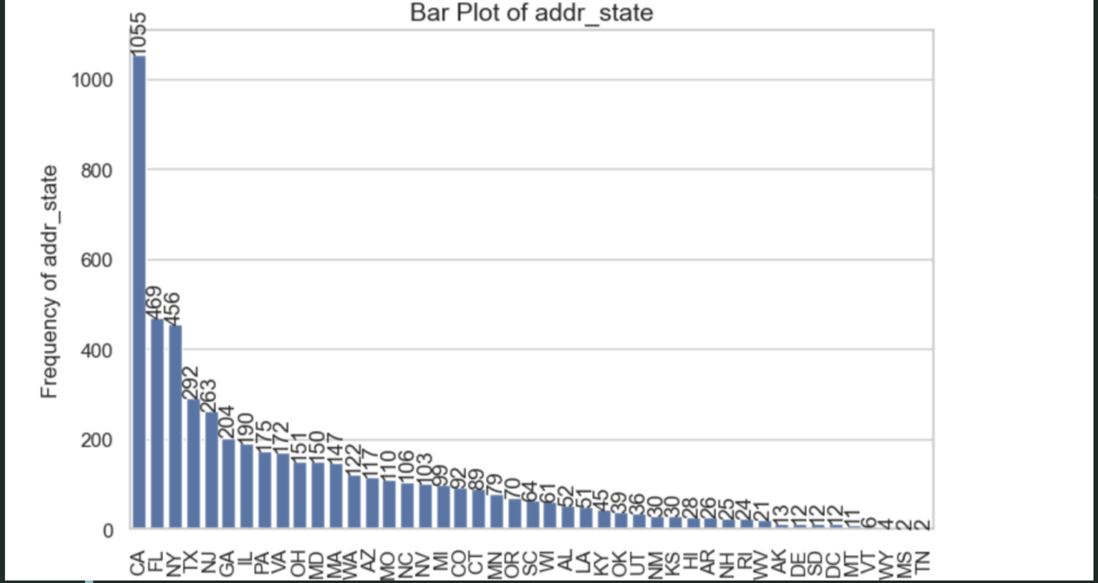

Bar Plot of issue_y
Frequency of issue_y
issue_y
0
500
1000
1500
2000
2500
3000
2007
2008
2009
2010
2011
2
165
566
1432
3152


Bar Plot of issue_m
Frequency of issue_m
issue_m
0
100
200
300
400
500
600
1
2
3
4
5
6
7
8
9
10
11
12
283
259
309
360
455
463
459
445
530
539
575
640


Bar Plot of issue_q
Frequency of issue_q
issue_q
0
500
1000
1500
2000
Q4
Q2
Q3
Q1
2284
1278
904
851


Bar Plot of home_ownership
Frequency of home_ownership
0
500
1000
1500
2000
2500
2715
2161
424
17
RENT
ORTGAGE
OWN
OTHER


Bar Plot of purpose
Frequency of purpose
0
500
1000
1500
2000
2500
2633
593
515
438
314
213
156
103
90
87
54
52
51
18
_consolidation
other
credit_card
mall_business
_improvement
ajor_purchase
car
medical
wedding
moving
house
vacation
educational
ewable_energy


Bar Plot of addr_state
Frequency of addr_state
0
200
400
600
800
1000
1055
469
456
292
263
204
190
175
172
151
150
147
122
117
110
106
103
99
92
89
79
70
64
61
52
51
45
39
36
30
30
28
26
25
24
21
13
12
12
12
11
6
4
2
2
CA
FL
NY
TX
NJ
GA
IL
PA
VA
OH
MD
MA
WA
AZ
MO
NC
NV
MI
CO
CT
MN
OR
SC
WI
AL
LA
KY
OK
UT
NM
KS
HI
AR
NH
RI
WV
AK
DE
SD
DC
MT
VT
WY
MS
TN

Observations & Inferences:

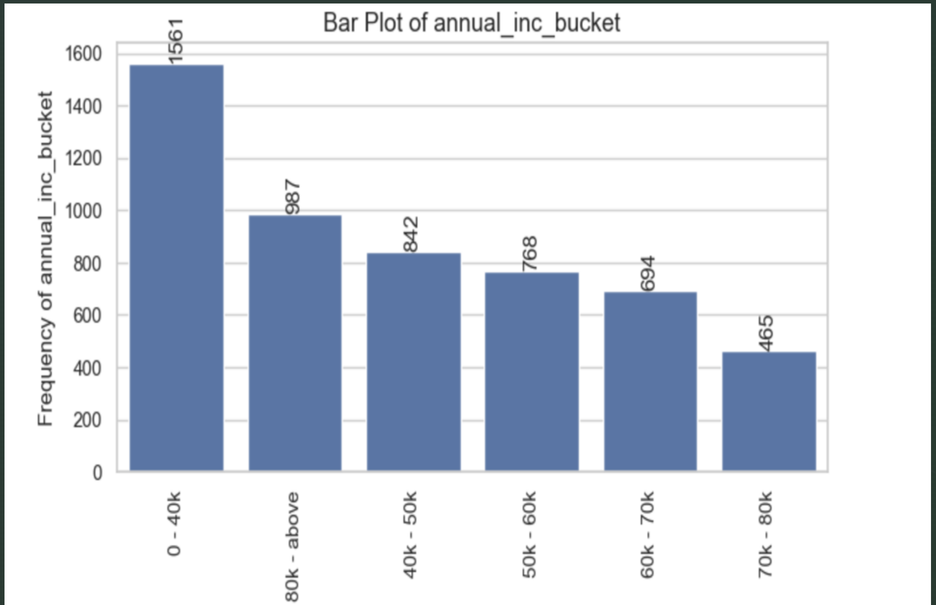
A. Ordered Categorical Variables:

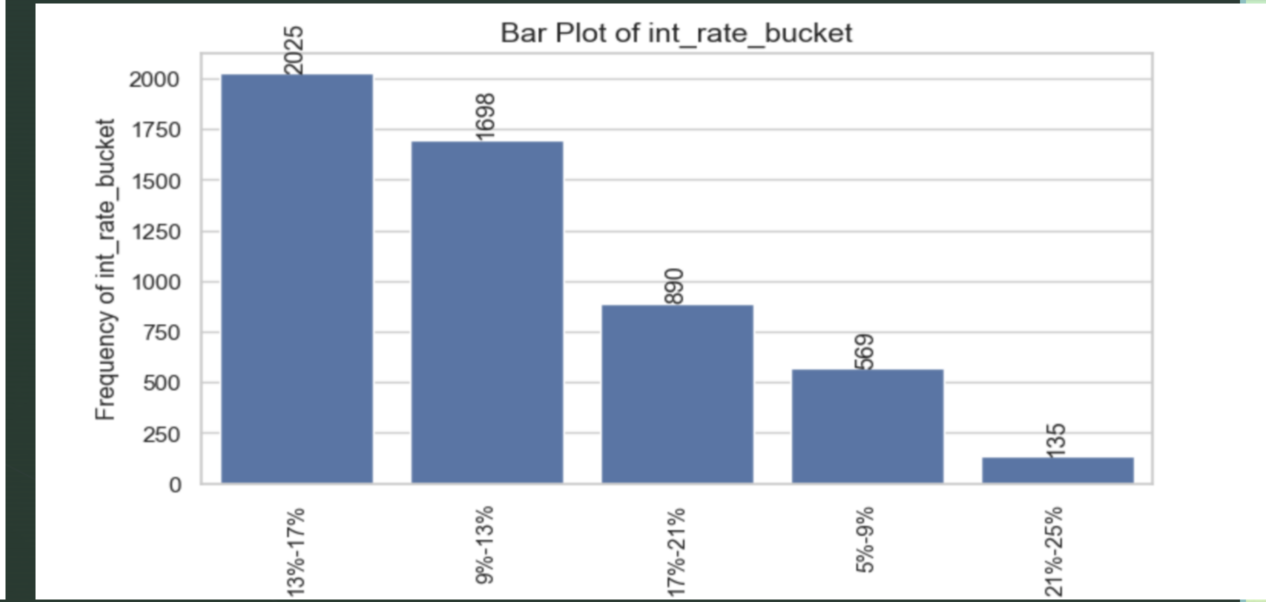
- Grade B had the highest number of "Charged off" loan applicants, with a total of 1,352 applicants, indicating that applicants with this credit grade faced challenges in repaying their loans.
- Short-term loans with a duration of 36 months were the most "Charged off" applicants, with 3,006 applications. This suggests that a significant portion of applicants who experienced loan default chose shorter repayment terms.
- Applicants who had been employed for more than 10 years accounted for the highest number of "Charged off" loans, totaling 1,474. This indicates that long-term employment history did not necessarily guarantee successful loan repayment.
- The year 2011 recorded the highest number of "Charged off" loan applications, totaling 3,152, signaling a positive trend in the number of applicants facing loan defaults over the years. This could be indicative of economic or financial challenges during that year.
- "Charged off" loans were predominantly taken during the 4th quarter, with 2,284 applications, primarily in December. This peak in loan applications during the holiday season might suggest that financial pressures during the holidays contributed to loan defaults.

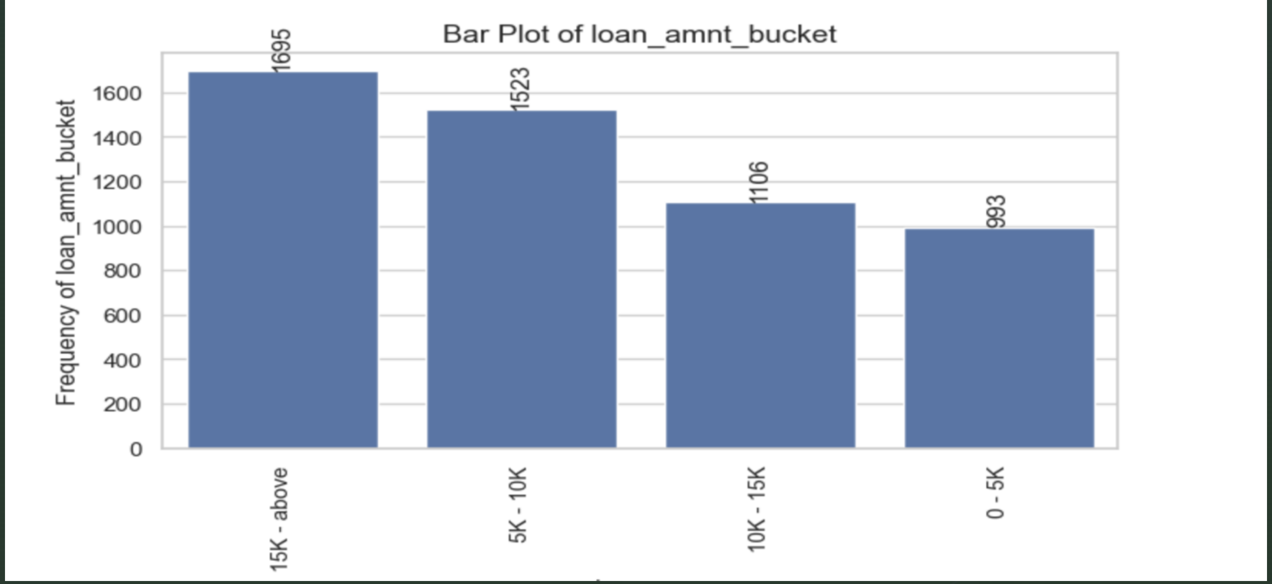
B. Unordered Categorical Variables:

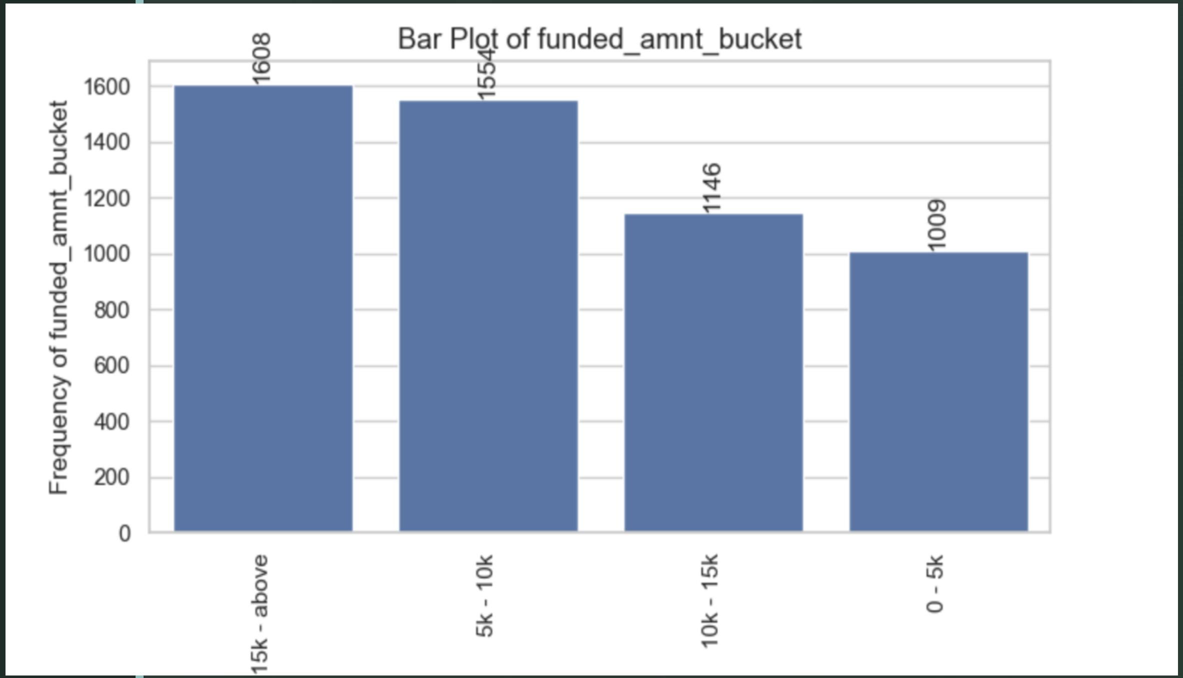
- California had the highest number of "Charged off" loan applicants, with 1,055 applicants. For such applicants, the lending company needs to implement stricter eligibility criteria or credit assessments due to a higher number of "Charged off" applicants from this state.
- Debt consolidation was the primary loan purpose for most "Charged off" loan applicants, with 2,633 applicants selecting this option. The lending company needs to exercise caution when approving loans for debt consolidation purposes, as it was the primary loan purpose for many "Charged off" applicants.
- The majority of "Charged off" loan participants, totaling 2,715 individuals, lived in rented houses. The lending company must assess the financial stability of applicants living in rented houses, as they may be more susceptible to economic fluctuations.
- A significant number of loan participants, specifically 5,317 individuals, were loan defaulters, unable to clear their loans. The lending company should enhance risk assessment practices, including stricter credit checks and lower loan-to-value ratios, for applicants with a history of loan defaults. They should offer financial education and support services to help borrowers manage their finances and improve loan repayment outcomes.

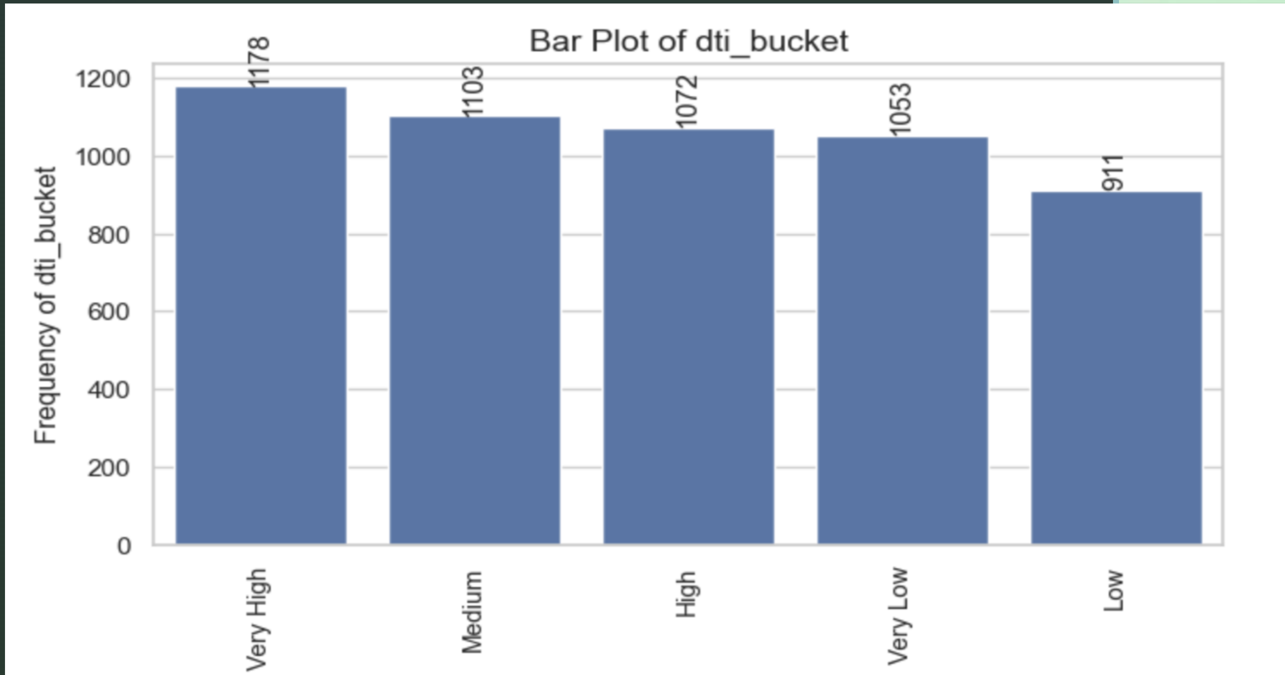
Univariate Analysis (Quantitative Variables)

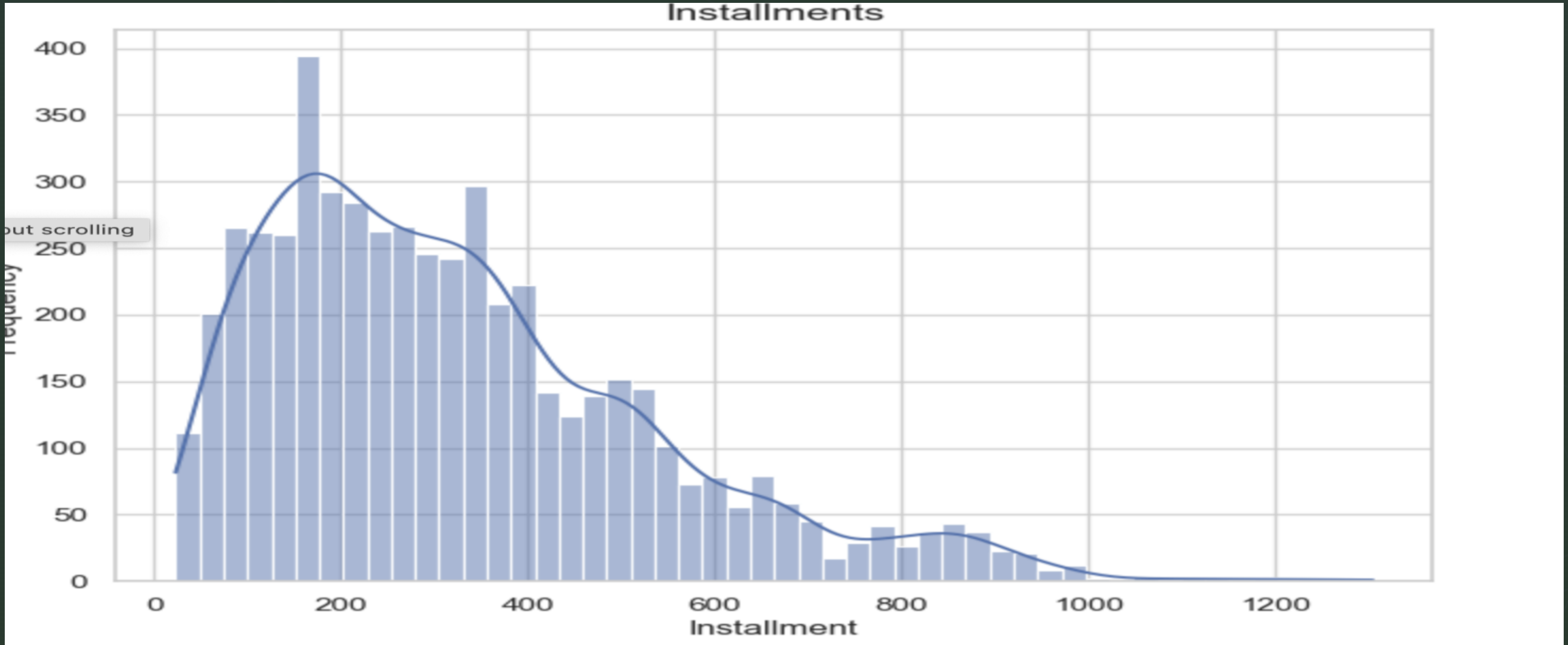







Bar Plot of funded_amnt_bucket
Frequency of funded_amnt_bucket
0
200
400
600
800
1000
1200
1400
1600
1608
1554
1146
1009
15k - above
5k - 10k
10k - 15k
0 - 5k


Bar Plot of dti_bucket
Frequency of dti_bucket
0
200
400
600
800
1000
1200
1178
1103
1072
1053
911
Very High
Medium
High
Very Low
Low


Installments
Frequency
0
50
100
150
200
250
300
350
400
0
200
400
600
800
1000
1200
Installment
out scrolling

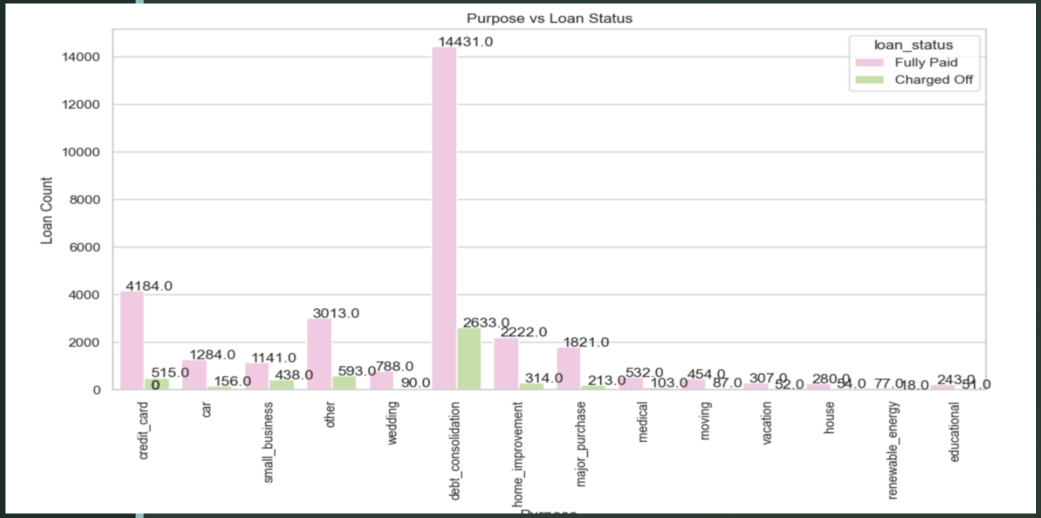
Observations

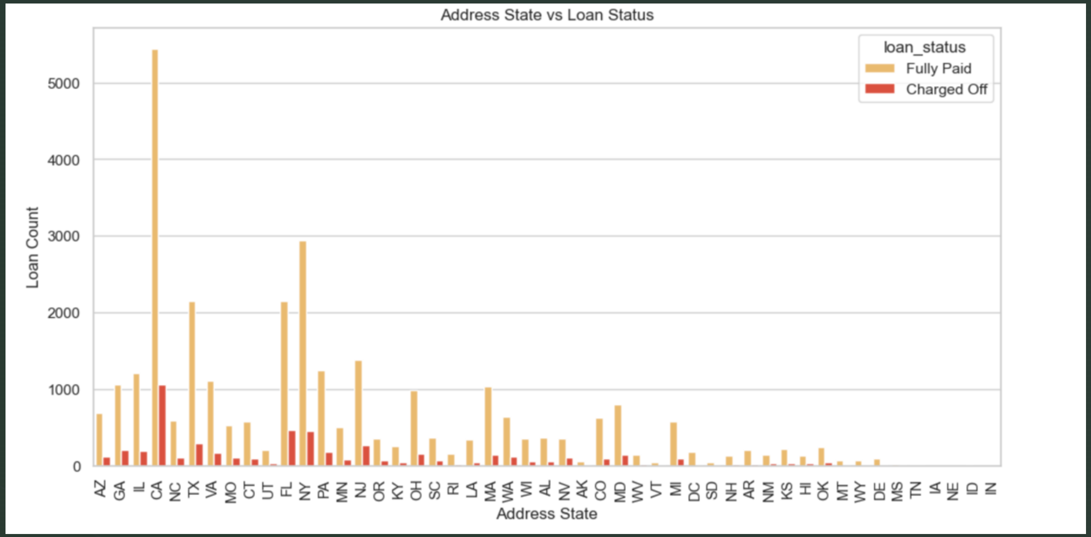
1. 1,561 loan applicants who charged off had annual salaries less than 40,000 USD. The lending company should exercise caution when lending to individuals with low annual salaries. They should implement rigorous income verification and assess repayment capacity more thoroughly for applicants in this income bracket.
2. Among loan participants who charged off (2,025), a considerable portion belonged to the interest rate bucket of 13%-17%.To reduce the risk of default, the lending company should consider offering loans at lower interest rates when possible.
3. 1,695 loan participants who charged off received loan amounts of 15,000 USD and above. The lending company should evaluate applicants seeking higher loan amounts carefully. They should ensure the applicants must have a strong credit history and repayment capability to handle larger loans.
4. 1,608 loan participants who charged off received funded amounts of 15,000 USD and above. The lending company should ensure that the funded amounts align with the borrower's financial capacity. They should conduct thorough credit assessments for larger loan requests.
5. Among loan participants who charged off, 1,178 loan applicants had very high debt-to-income ratios. The lending company
6. should implement strict debt-to-income ratio requirements to prevent lending to individuals with unsustainable levels of debt relative to their income.
7. Among loan participants who charged off, it's observed that the majority of them had monthly installment amounts falling within the range of 160-440 USD. The lending company should closely monitor and assess applicants with similar installment amounts to mitigate the risk of loan defaults.

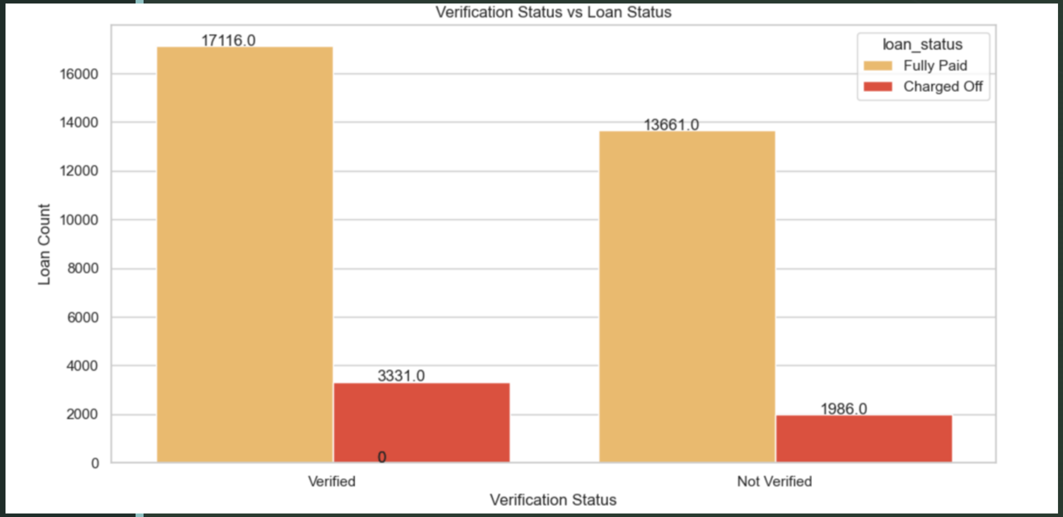
Bivariate Analysis

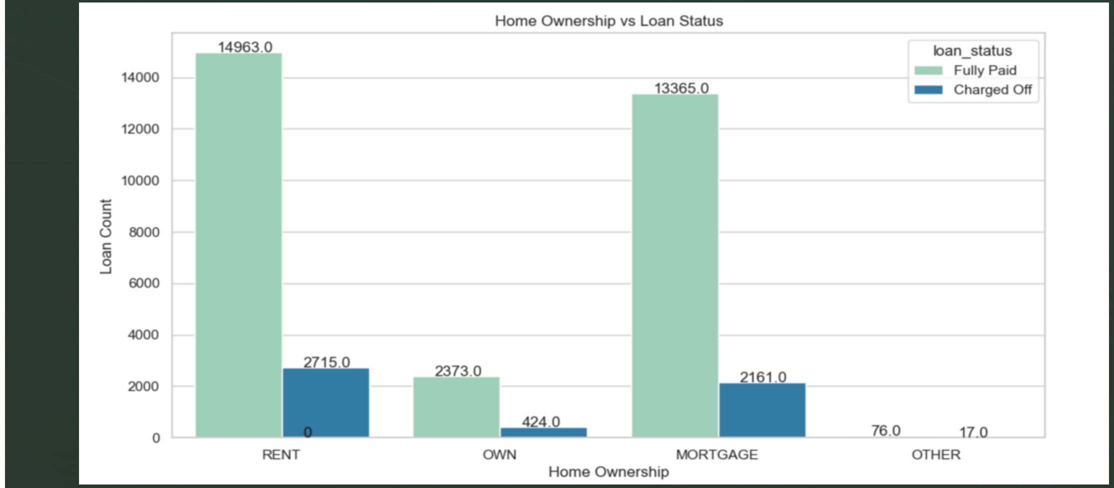
- Bivariate analysis is a statistical method that involves the simultaneous analysis of two variables (factors). It aims to determine the empirical relationship between them. The analysis can be used to test hypotheses, identify patterns, or explore relationships between the variables.
- It was carried out for both Categorical and Quantitative Variables

Bivariate Analysis (Unordered Categorical)

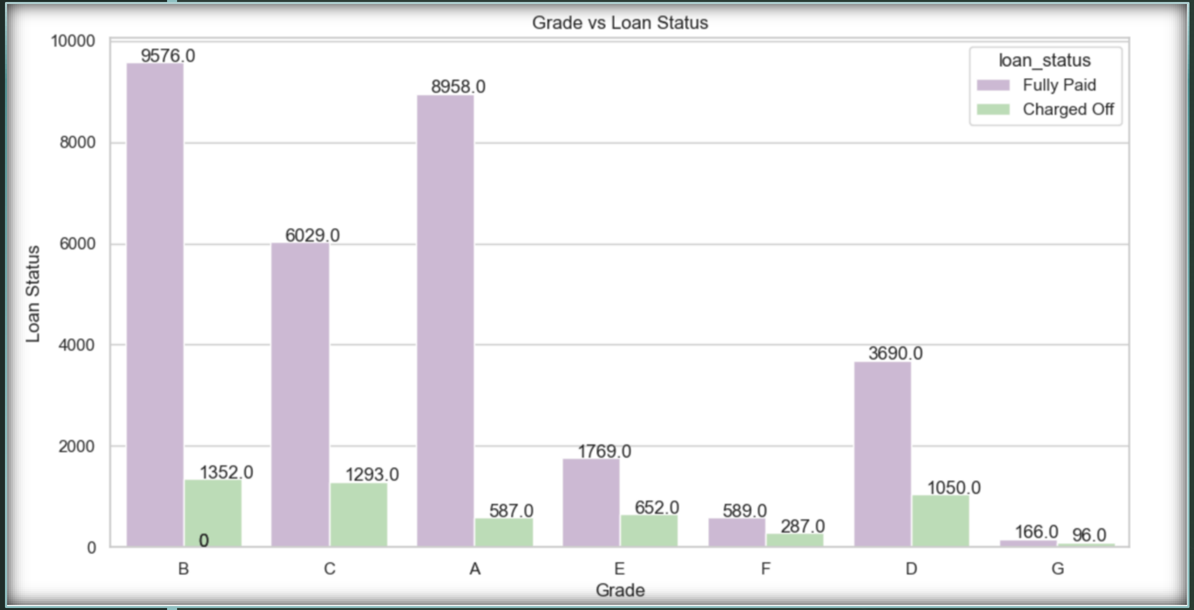

Purpose vs Loan Status

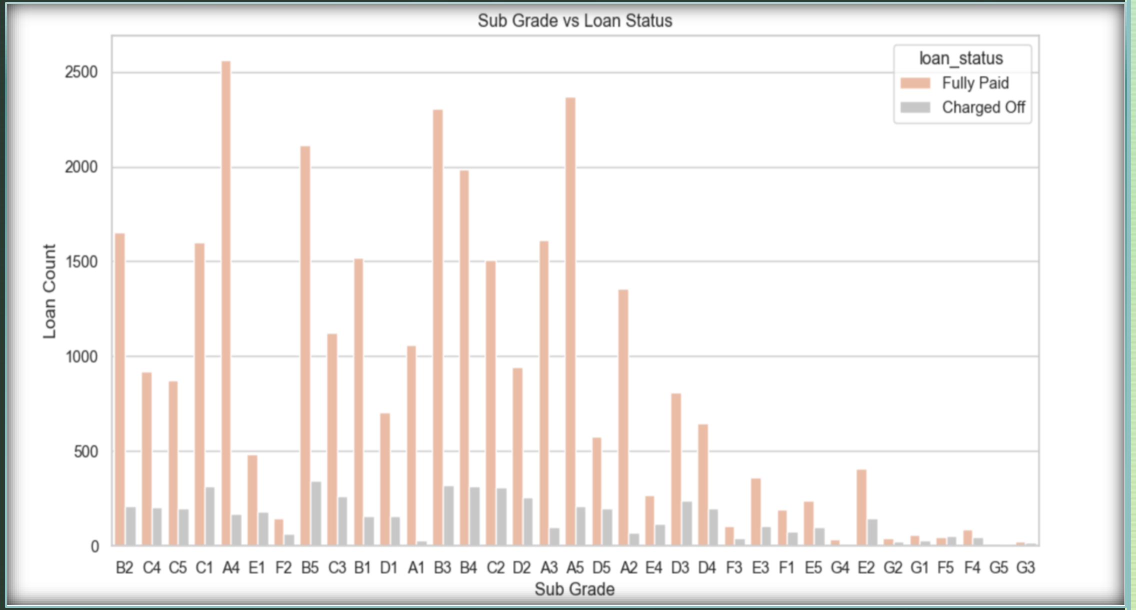

Address State vs Loan Status

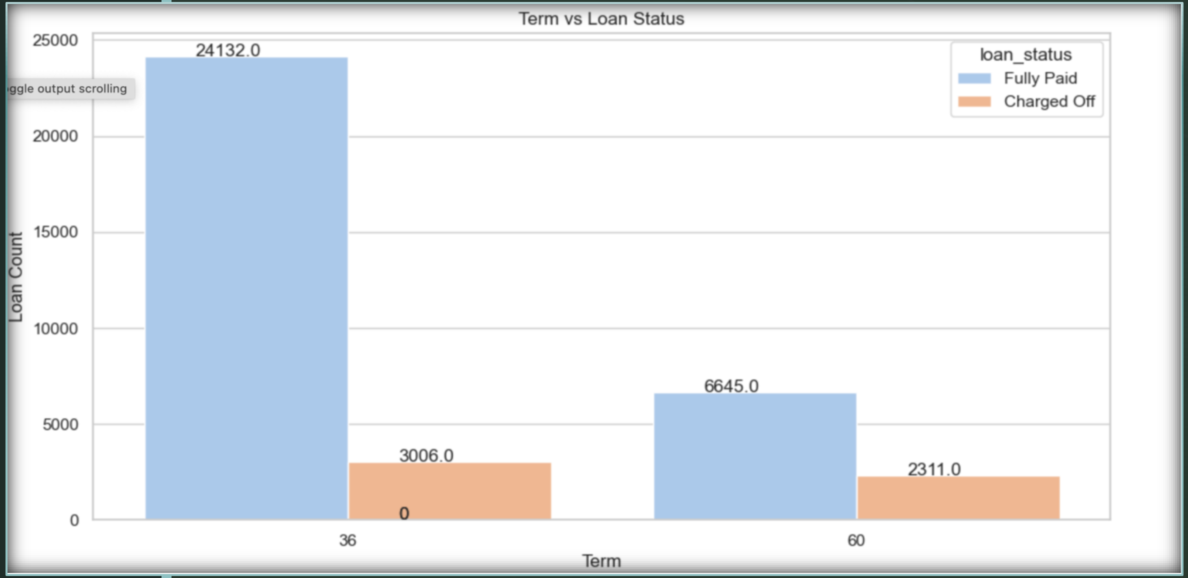

Verification Status vs Loan Status

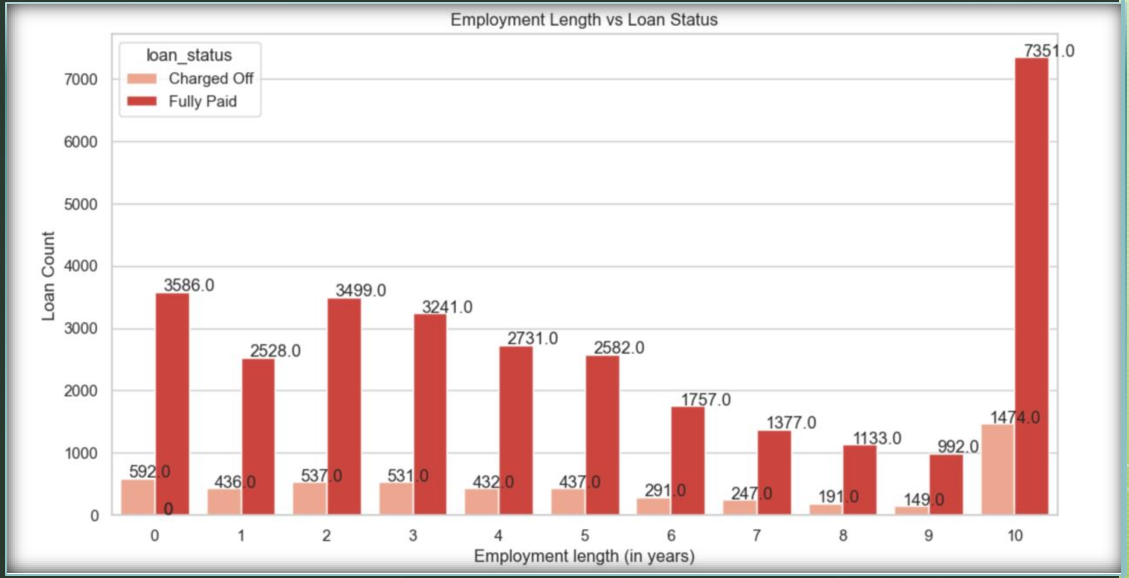

Home Ownership vs Loan Status

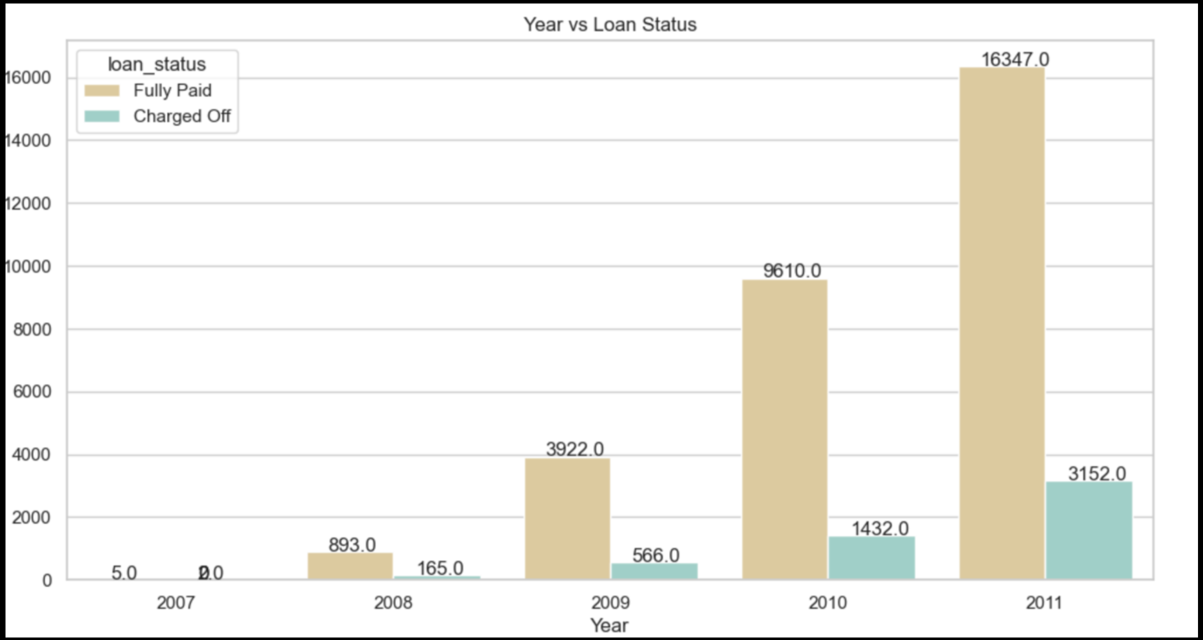
Bivariate Analysis (Ordered Categorical)

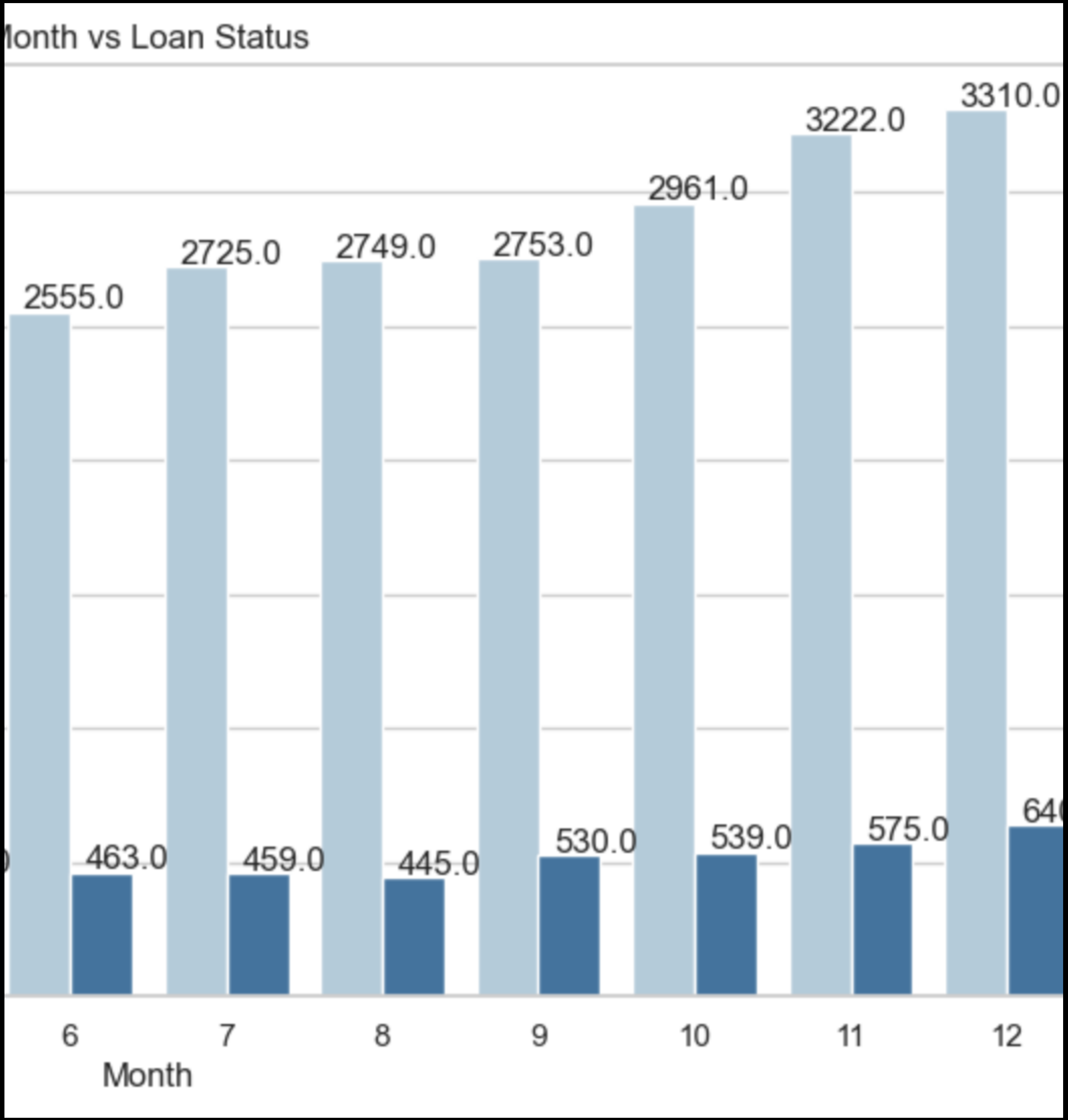


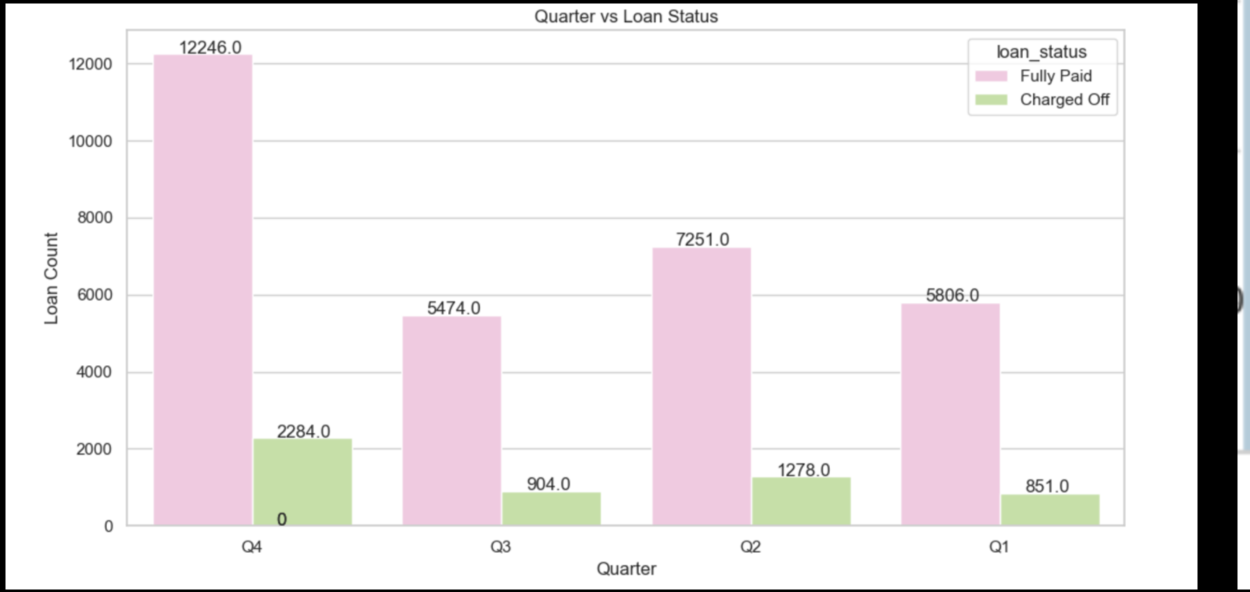







Year vs Loan Status
loan_status
Fully Paid
Charged Off
16000
14000
12000
10000
8000
6000
4000
2000
0
5.0
2.0
893.0
165.0
3922.0
566.0
9610.0
1432.0
16347.0
3152.0
2007
2008
2009
2010
2011
Year


Month vs Loan Status
2555.0
2725.0
2749.0
2753.0
2961.0
3222.0
3310.0
463.0
459.0
445.0
530.0
539.0
575.0
64
6
7
8
9
10
11
12
Month


Quarter vs Loan Status
loan_status
Fully Paid
Charged Off
Loan Count
12000
10000
8000
6000
4000
2000
0
12246.0
2284.0
0
5474.0
904.0
7251.0
1278.0
5806.0
851.0
Q4
Q3
Q2
Q1
Quarter

Inferences

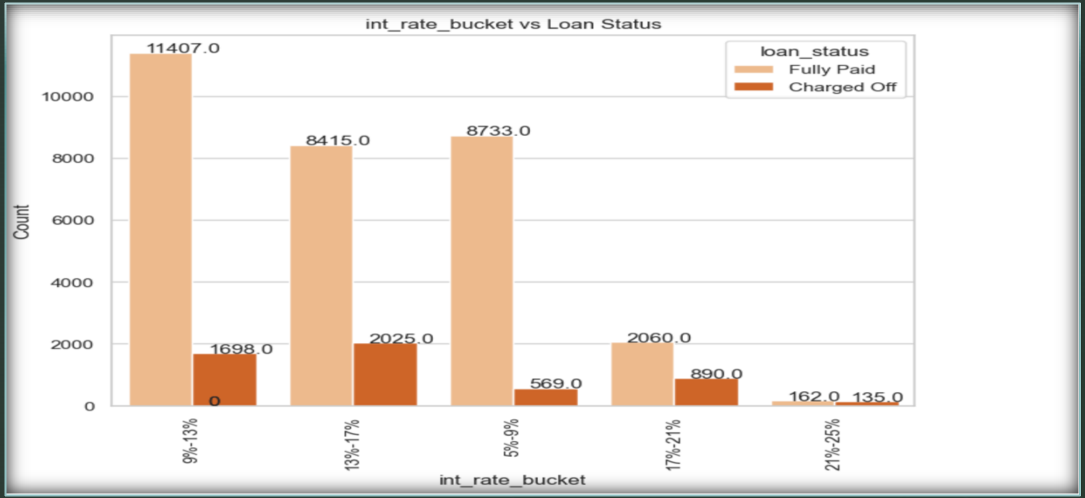
A. **Ordered Categorical Variables:**

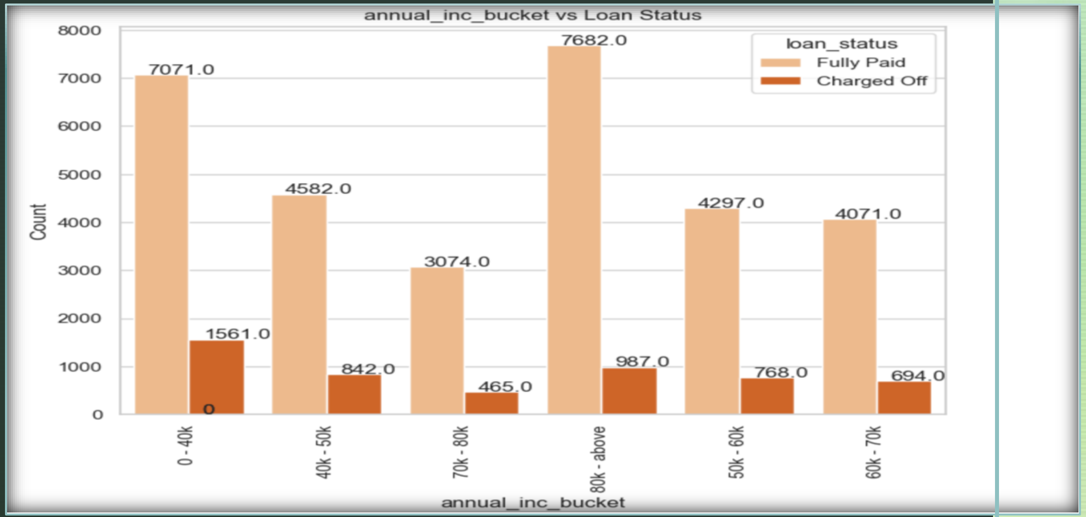
- Since loan applicants from Grades B, C, and D contribute to most of the "Charged Off" loans, the company should consider implementing stricter risk assessment and underwriting criteria for applicants falling into these grades.
- Pay special attention to applicants with Subgrades B3, B4, and B5, as they are more likely to charge off. Implementing additional risk mitigation measures or offering them lower loan amounts could be considered.
- Given that applicants opting for 60-month loans are more likely to default, the company should consider evaluating the risk associated with longer-term loans and potentially limiting the maximum term or adjusting interest rates accordingly.
- Loan applicants with ten or more years of experience are more likely to default. This suggests that experience alone may not be a reliable indicator of creditworthiness. The company should use a more comprehensive credit scoring system that factors in other risk-related attributes.
- The steady increase in the number of loan applicants from 2007 to 2011 indicates growth in the market. The company can capitalize on this trend by maintaining a competitive edge in the industry while keeping risk management practices robust.
- December and Q4 are peak periods for loan applications, likely due to the holiday season. The company should anticipate increased demand during these periods and ensure efficient processing to meet customer needs

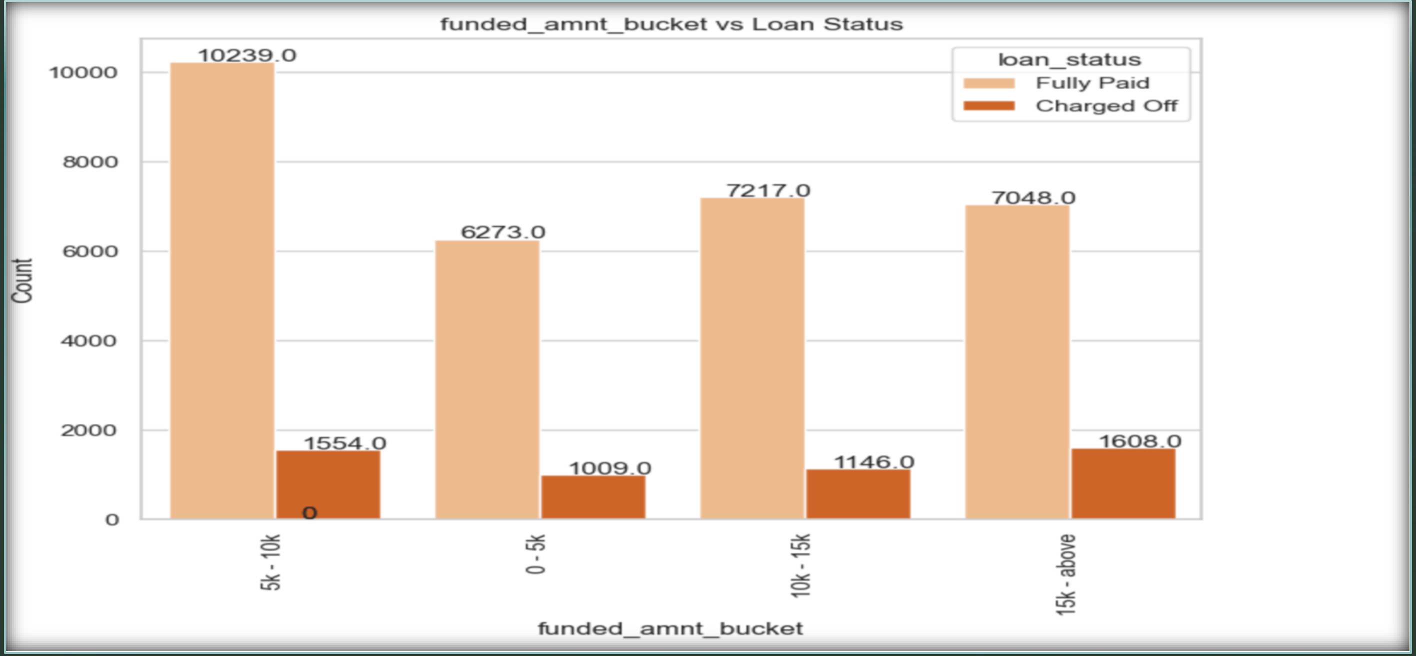
B. **Unordered Categorical Variables:**

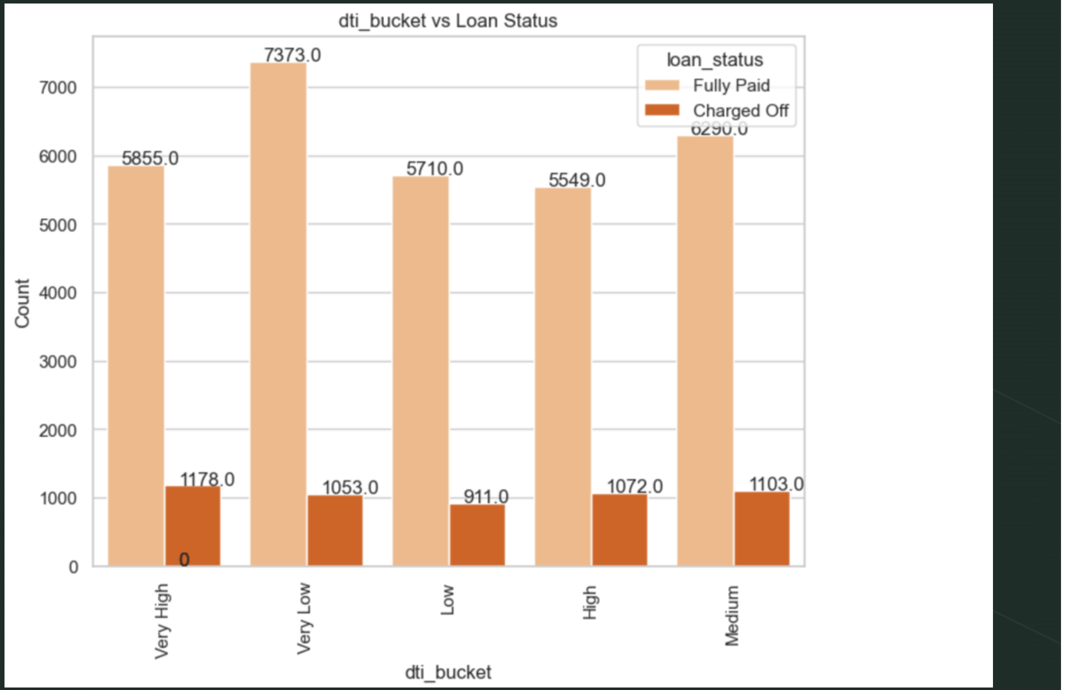
- Since debt consolidation is the category with the maximum number of loans and high default rates, the company should carefully evaluate applicants seeking debt consolidation loans and potentially adjust interest rates or offer financial counseling services.
- Applicants living in rented or mortgaged houses are more likely to default. This information can be considered in the underwriting process to assess housing stability and its impact on repayment ability.
- Verified loan applicants are defaulting more than those who are not verified. The company should review its verification process to ensure it effectively assesses applicant creditworthiness and consider improvements or adjustments.
- Loan applicants from states like California (CA), Florida (FL), and New York (NY) are more likely to default. The company should monitor regional risk trends and adjust lending strategies or rates accordingly in these areas.

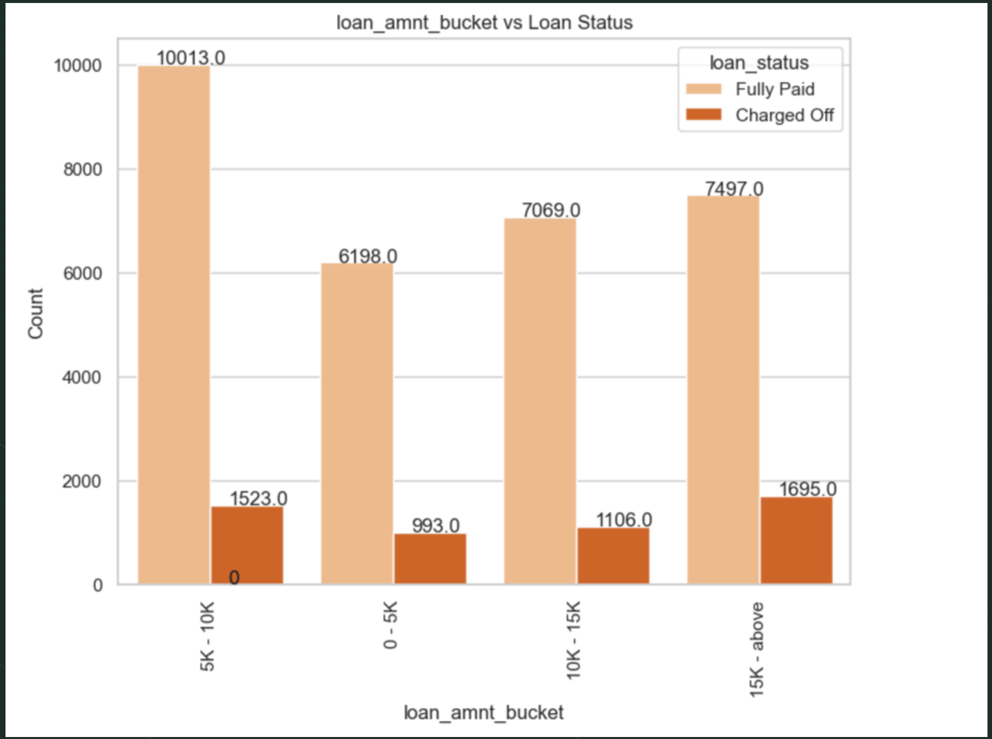
Bivariate Analysis (Quantitative Variables)








dti_bucket vs Loan Status
loan_status
Fully Paid
Charged Off
Count
0
1000
2000
3000
4000
5000
6000
7000
5855.0
1178.0
0
7373.0
1053.0
5710.0
911.0
5549.0
1072.0
6290.0
1103.0
Very High
Very Low
Low
High
Medium
dti_bucket


loan_amnt_bucket vs Loan Status
loan_status
Fully Paid
Charged Off
Count
0
2000
4000
6000
8000
10000
10013.0
1523.0
0
6198.0
993.0
7069.0
1106.0
7497.0
1695.0
5K - 10K
0 - 5K
10K - 15K
15K - above
loan_amnt_bucket

Observations:

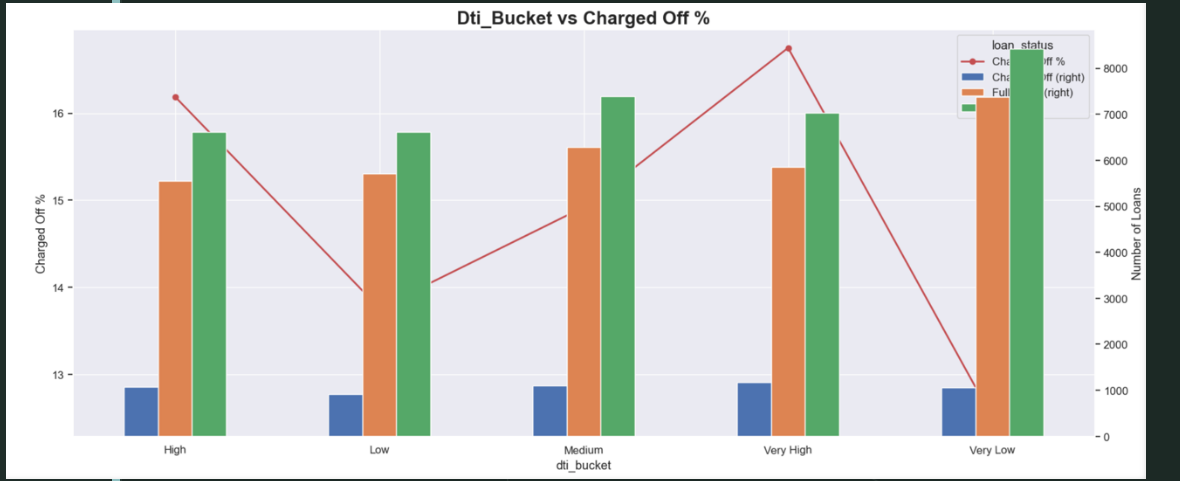
- ✓ A majority of the loan applicants who defaulted received loan amounts of $15,000 or higher.
- ✓ The majority of loan applicants who charged off had significantly high Debt-to-Income (DTI) ratios.
- ✓ A significant portion of loan applicants who defaulted received loans with interest rates falling within the range of 13% to 17%.
- ✓ A majority of the loan applicants who charged off reported an annual income of less than $40,000.

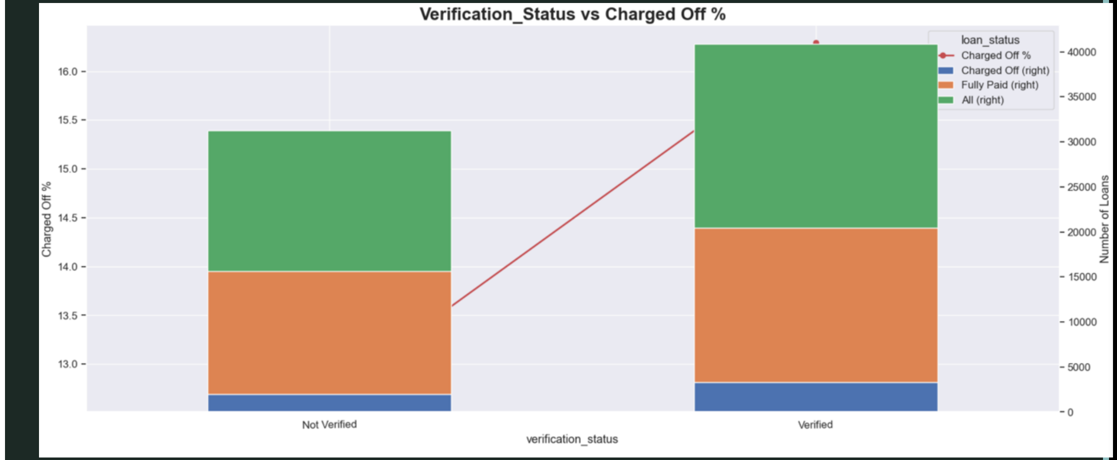
Inferences:

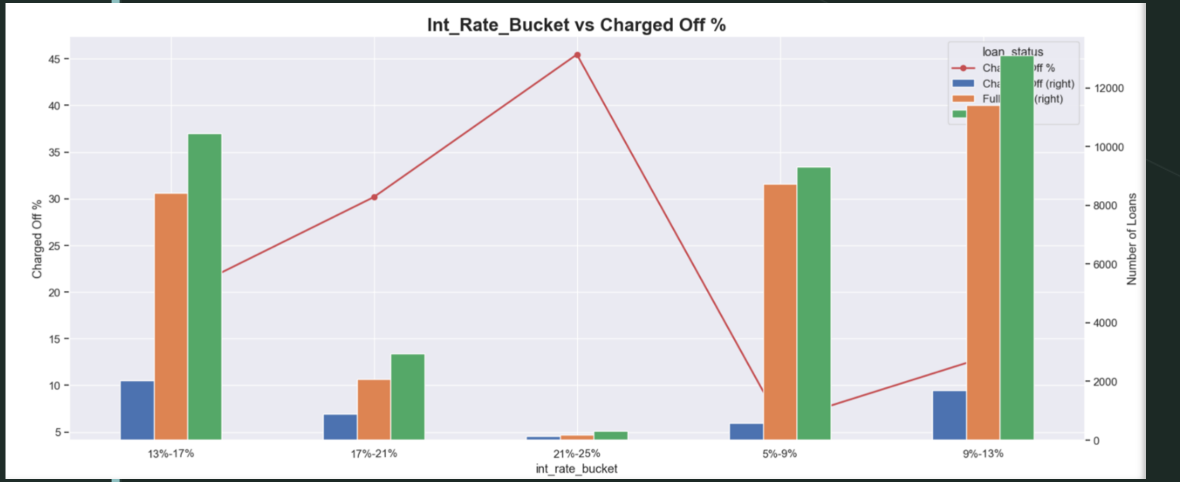
- ✓ Applicants receiving loan amounts of $15,000 or higher are more likely to default. The company can mitigate this risk by conducting more thorough assessments for larger loan requests and potentially capping loan amounts for higher-risk applicants.
- ✓ High Debt-to-Income (DTI) ratios and interest rates in the 13%-17% range are associated with defaults. The company should review its interest rate determination process and consider adjusting rates based on DTI ratios to better align with the borrower's ability to repay.
- ✓ Applicants with annual incomes less than $40,000 have a higher likelihood of defaulting. The company should consider offering financial education resources or setting maximum loan amounts based on income levels to ensure affordability for borrowers.

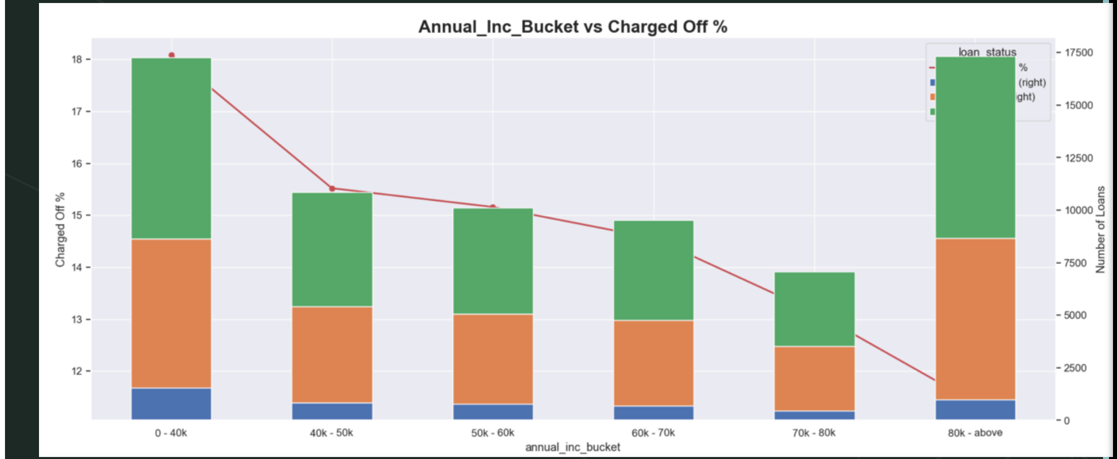
Multivariate Analysis

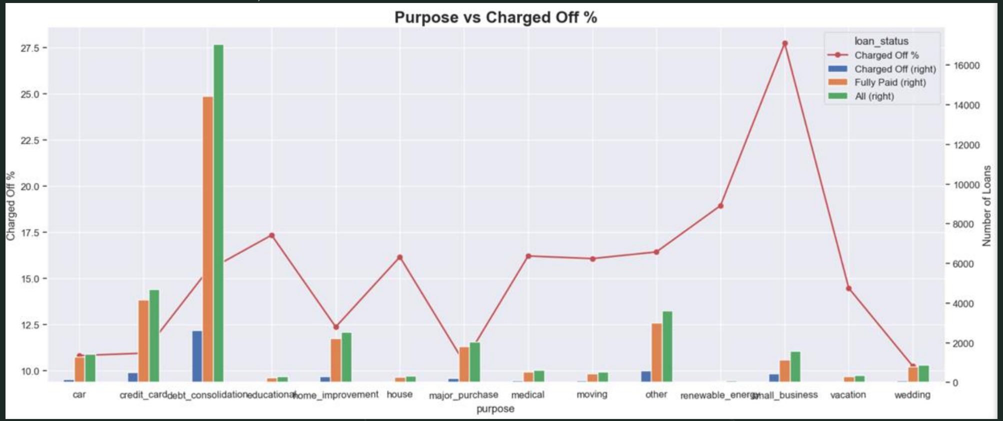
Multivariate analysis is a statistical technique used to analyze data that involves more than two variables.

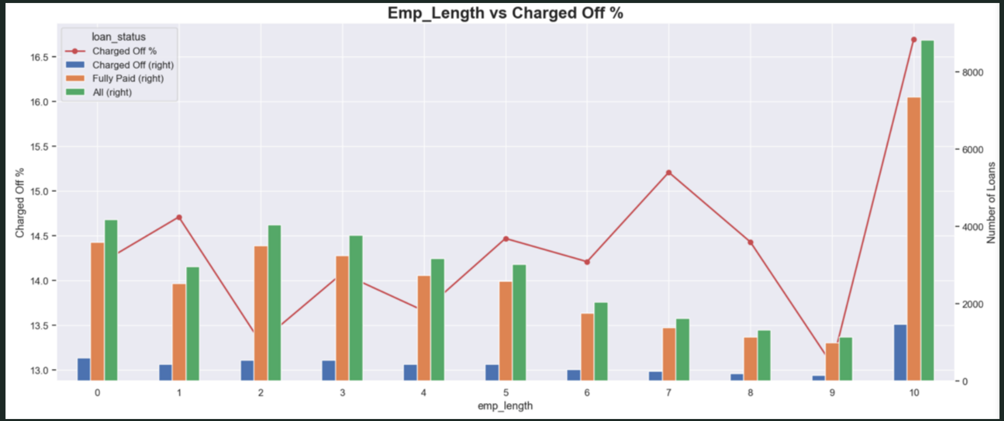

Dti_Bucket vs Charged Off %
loan_status
Cha ff %
Cha ff (right)
Full (right)
Charged Off %
Number of Loans
16
15
14
13
8000
7000
6000
5000
4000
3000
2000
1000
0
High
Low
Medium
Very High
Very Low
dti_bucket

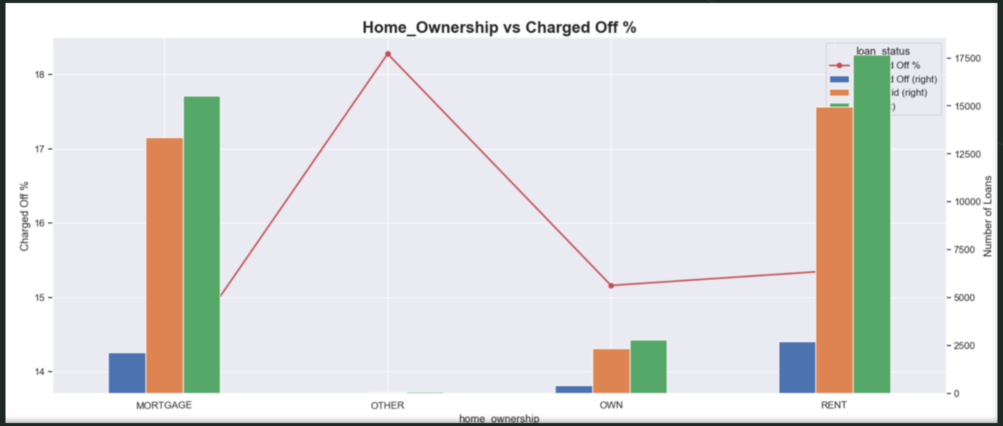

Verification_Status vs Charged Off %
loan_status
Charged Off %
Charged Off (right)
Fully Paid (right)
All (right)
Charged Off %
Number of Loans
16.0
15.5
15.0
14.5
14.0
13.5
13.0
40000
35000
30000
25000
20000
15000
10000
5000
0
Not Verified
Verified
verification_status

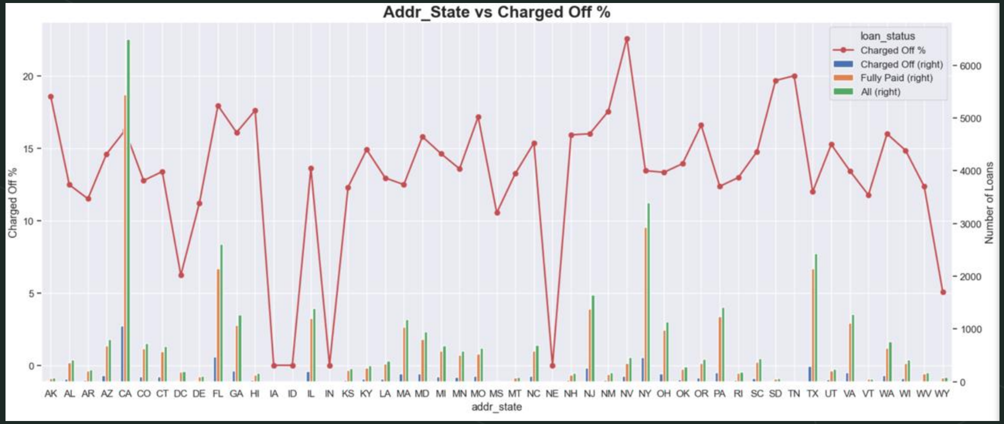

Int_Rate_Bucket vs Charged Off %
loan_status
Cha ff %
Cha ff (right)
Full (right)
Charged Off %
Number of Loans
45
40
35
30
25
20
15
10
5
12000
10000
8000
6000
4000
2000
0
13%-17%
17%-21%
21%-25%
5%-9%
9%-13%
int_rate_bucket


Annual_Inc_Bucket vs Charged Off %
loan_status
%
(right)
ght)
Charged Off %
Number of Loans
18
17
16
15
14
13
12
17500
15000
12500
10000
7500
5000
2500
0
0 - 40k
40k - 50k
50k - 60k
60k - 70k
70k - 80k
80k - above
annual_inc_bucket

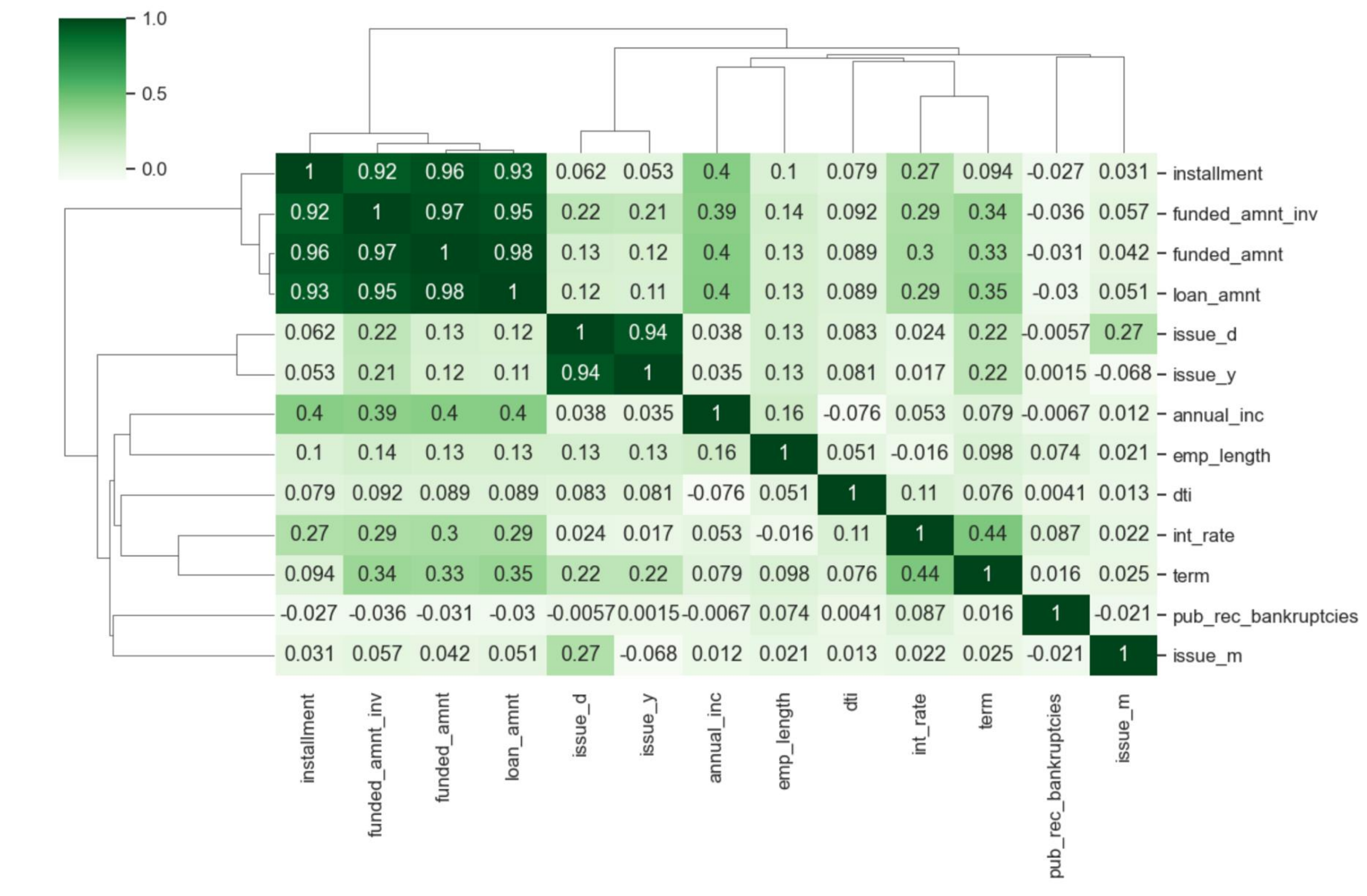

Purpose vs Charged Off %
loan_status
Charged Off %
Charged Off (right)
Fully Paid (right)
All (right)
Charged Off %
Number of Loans
27.5
25.0
22.5
20.0
17.5
15.0
12.5
10.0
16000
14000
12000
10000
8000
6000
4000
2000
0
car
credit_card
debt_consolidation
educational
home_improvement
house
major_purchase
medical
moving
other
renewable_energy
small_business
vacation
wedding
purpose


Emp_Length vs Charged Off %
loan_status
Charged Off %
Charged Off (right)
Fully Paid (right)
All (right)
Charged Off %
Number of Loans
16.5
16.0
15.5
15.0
14.5
14.0
13.5
13.0
8000
6000
4000
2000
0
0
1
2
3
4
5
6
7
8
9
10
emp_length


Home_Ownership vs Charged Off %
loan_status
d Off %
d Off (right)
id (right)
)
Charged Off %
Number of Loans
18
17
16
15
14
17500
15000
12500
10000
7500
5000
2500
0
MORTGAGE
OTHER
OWN
RENT
home_ownership


Addr_State vs Charged Off %
loan_status
Charged Off %
Charged Off (right)
Fully Paid (right)
All (right)
Charged Off %
Number of Loans
20
15
10
5
0
6000
5000
4000
3000
2000
1000
0
AK
AL
AR
AZ
CA
CO
CT
DC
DE
FL
GA
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
MI
MN
MO
MS
MT
NC
NE
NH
NJ
NM
NV
NY
OH
OK
OR
PA
RI
SC
SD
TN
TX
UT
VA
VT
WA
WI
WV
WY
addr_state

Correlation Analysis



	installment	funded_amnt_inv	funded_amnt	loan_amnt	issue_d	issue_y	annual_inc	emp_length	dti	int_rate	term	pub_rec_bankruptcies	issue_m
installment	1	0.92	0.96	0.93	0.062	0.053	0.4	0.1	0.079	0.27	0.094	-0.027	0.031
funded_amnt_inv	0.92	1	0.97	0.95	0.22	0.21	0.39	0.14	0.092	0.29	0.34	-0.036	0.057
funded_amnt	0.96	0.97	1	0.98	0.13	0.12	0.4	0.13	0.089	0.3	0.33	-0.031	0.042
loan_amnt	0.93	0.95	0.98	1	0.12	0.11	0.4	0.13	0.089	0.29	0.35	-0.03	0.051
issue_d	0.062	0.22	0.13	0.12	1	0.94	0.038	0.13	0.083	0.024	0.22	-0.0057	0.27
issue_y	0.053	0.21	0.12	0.11	0.94	1	0.035	0.13	0.081	0.017	0.22	0.0015	-0.068
annual_inc	0.4	0.39	0.4	0.4	0.038	0.035	1	0.16	-0.076	0.053	0.079	-0.0067	0.012
emp_length	0.1	0.14	0.13	0.13	0.13	0.13	0.16	1	0.051	-0.016	0.098	0.074	0.021
dti	0.079	0.092	0.089	0.089	0.083	0.081	-0.076	0.051	1	0.11	0.076	0.0041	0.013
int_rate	0.27	0.29	0.3	0.29	0.024	0.017	0.053	-0.016	0.11	1	0.44	0.087	0.022
term	0.094	0.34	0.33	0.35	0.22	0.22	0.079	0.098	0.076	0.44	1	0.016	0.025
pub_rec_bankruptcies	-0.027	-0.036	-0.031	-0.03	-0.0057	0.0015	-0.0067	0.074	0.0041	0.087	0.016	1	-0.021
issue_m	0.031	0.057	0.042	0.051	0.27	-0.068	0.012	0.021	0.013	0.022	0.025	-0.021	1

SUGGESTIONS

1. Grade B,C and D customers are more prone to loan default. Implement stricter guidelines when disbursing loan to mentioned grades
2. Sub grades B3, B4, and B5 are also more prone to loan default. Implement stricter guidelines when disbursing loan to mentioned grades
3. Longer term loans are more prone to loan default. Move with caution when dealing with them or charge high interest so that loss can be minimum.
4. Customers with more experience are prone to loan default. Which suggest that experience alone cannot trusted when disbursing loan consider other indicators also.
5. Q4 are peak periods for loan applications, likely due to the holiday season. The company should anticipate increased demand during these periods and ensure efficient processing to meet customer needs.
6. debt consolidation loan has more prone to loan default. Maintain caution when dealing with such type of loans
7. Customers living in rented or mortgaged are prone to loan default.
8. As expected Verified customers are less likely to default. So bring in more robust verification process and maximum try to verify all customers
9. Applicants receiving loan amounts of $15,000 or higher are more likely to default.
10. High Debt-to-Income (DTI) ratios and interest rates in the 13%-17% range are associated with defaults.
11. Applicants with annual incomes less than $40,000 have a higher likelihood of defaulting. The company should consider offering financial education resources or setting maximum loan amounts based on income levels to ensure affordability for borrowers.
12. The steady increase in the number of loan applicants from 2007 to 2011 indicates growth in the market. The company can capitalize on this trend by maintaining a competitive edge in the industry while keeping risk management practices robust.

Thank You