Lending Club Case Study

A submission assignment from upGrad and IIITB Machine Learning and AI Program
Team: Aishwarya Gobburi and Girish Kolhe
Nov 2024

Agenda

- Executive Summary
- Problem Statement
- Exploratory Data Analysis (EDA) Approach
- Data Cleaning and Manipulation
- Data Profiling & Statistics
- Key Findings and Analysis
- Recommendations and Conclusion

Executive Summary – Lending Club case study

Potential Impact	
↓	**Defaulted accounts**
↑	**Revenue Increase**

Business Problems

The primary challenge for a consumer finance company is to mitigate financial losses caused by 'risky' loan applicants who default, referred to as 'charged-off' customers.

It is expected that the case study identifies patterns which indicate if a person is likely to default, enabling the company to refine its loan approval process

This case study aims to leverage Exploratory Data Analysis (EDA)

Key Findings

0.17 is Default-to-fully paid ratio

14.5% accounts are contributing to Charged-Off accounts

42% of accounts are not verified

92% of defaulted borrowers does not own house

69% of charged off accounts are from B, C and D grade

54% of defaulted accounts have open account between 5 to 10

Summary of key findings is on 52nd slide

Key Recommendations

- For the loan purpose for debt consolidation and credit card, implement stricter verification process.
- Prefer Source verification
- If the loan tenure is less then verify the DTI ratio.
- Own house status can be considered to lower the risk of default
- Additionally consider income stability such as DTI, average current balance of all accounts and other indicators (credit or debit) while analyzing purpose

Detailed recommendation are on 53rd slide

Problem Statement

Background

- A consumer finance company which specializes in lending various types of loans to urban customers.
- This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures
- Borrowers can easily access lower interest rate loans through a fast online interface.
- Like most other lending companies, lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss)
- Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed
- Borrowers who **default,** cause the largest amount of loss to the lenders
- This case study is aim to *identify these risky loan applicants to minimize the credit loss*

Objective

- This case study aims at identifying pattern and different ratios or variables directly or indirectly suggests the defaulters by virtue of identifying attributes such as employee length, title, geographic location or spending patterns
- Primarily EDA will be leveraged to reveal these information.

Important Areas

- Analyze different type of data available about loan applicants
- Establish correlations and dependency between available information to find out pattern

Exploratory Data Analysis (EDA) Approach

	Data Cleaning & Manipulation	Data Profiling & Statistics	Key Findings & Analysis	Recommendations
Key Activities	▪ **Data Sourcing:** access and understand the data ▪ **Data Cleaning:** Handle missing values, remove duplicates ▪ **Data Manipulation and Transformation**: correct dtypes, add new columns as necessary	▪ **Descriptive Statistics:** Calculate necessary stats such as IQR ▪ **Data Visualization:** Use data visualization techniques to plot patterns ▪ Explore relationship between various attributes	▪ **Business driven**: Understand various attributes contribution to charged-off and fully-paid accounts ▪ Analyze individual variables ▪ Explore relationship between multiple variables	▪ **Key Insights**: Identify most significant pattern ▪ **Reporting**: Summarize the pattern and insights ▪ Recommendations
Key Outcome	▪ Access to data ▪ Data sanitization	▪ individual attribute impact analysis ▪ Various patterns and percentage break-up	▪ Business metrics ▪ Univariate, Bivariate and Multi-variate analysis	▪ Executive summary ▪ Recommendation ▪ Conclusion

Data Cleaning and Manipulation

Data understanding - Key Attributes

- *Out of 111 columns (attributes) & from available data below attributes are important for analysis*
- *A quick glance through the data dictionary reveals below possible categorization.*

Borrower's Demographic

- home_ownership
- zip_code
- addr_state

Loan Information

- loan_amnt
- funded_amnt
- funded_amnt_inv
- term
- int_rate
- installment
- grade and sub_grade
- purpose

Borrower's Employment

- emp_title (employer title)
- emp_length

Borrower's Financial

- annual_inc
- dti
- home_ownership

Loan Status Information

- issue_d
- loan_status
- pymnt_plan
- desc (description)
- verification_status

Borrower's Credit

- delinq_2yrs
- earliest_cr_line
- inq_last_6mths
- mths_since_last_delinq
- mths_since_last_record
- open_acc
- pub_rec
- pub_rec_bankruptcies

Data definition for key attributes

Attribute	Definition
home_ownership	Home ownership type (Owned, Rent, Mortgage)
Grade & Sub-grade	Risk based rating
purpose	Category provided by the borrower for the loan request
emp_title (employer title)	Job Title or Employer title
emp_length	Employment length
dti (Debt to income)	Ratio of borrower's income already in used
issue_d	Loan funding month
loan_status	Key attribute comprises of values Fully Paid: Paid by borrower Charged-Off: Defaulter Current: Loan is in-progress
verification_status	Income or income source verification
open_acc	open credit lines in the borrower's credit file
pub_rec	# of derogatory public records
pub_rec_bankruptcies	Number of public record bankruptcies

*Refer **Data dictionary** for more details about this data*

Data Cleaning and Manipulation

Data understanding - Other Attributes

- *87 columns (attributes) out of 111, mentioned below are not useful in this analysis as per remark captured in table*

Attributes	Remark about content
'verification_status_joint', 'annual_inc_joint', 'mo_sin_old_il_acct', 'bc_util', 'bc_open_to_buy', 'avg_cur_bal', 'acc_open_past_24mths', 'inq_last_12m', 'total_cu_tl', 'inq_fi', 'total_rev_hi_lim', 'all_util', 'max_bal_bc', 'open_rv_24m', 'open_rv_12m', 'il_util', 'total_bal_il', 'mths_since_rcnt_il', 'open_il_24m', 'open_il_12m', 'open_il_6m', 'open_acc_6m', 'tot_cur_bal', 'tot_coll_amt', 'mo_sin_rcnt_rev_tl_op', 'mo_sin_rcnt_tl', 'mort_acc', 'num_rev_tl_bal_gt_0', 'total_bc_limit', 'total_bal_ex_mort', 'tot_hi_cred_lim', 'percent_bc_gt_75', 'pct_tl_nvr_dlq', 'num_tl_op_past_12m', 'num_tl_90g_dpd_24m', 'num_tl_30dpd', 'num_tl_120dpd_2m', 'num_sats', 'num_rev_accts', 'mths_since_recent_bc', 'num_op_rev_tl', 'num_il_tl', 'num_bc_tl', 'num_bc_sats', 'num_actv_rev_tl', 'num_actv_bc_tl', 'num_accts_ever_120_pd', 'mths_since_recent_revol_delinq', 'mths_since_recent_inq', 'mths_since_recent_bc_dlq', 'dti_joint', 'total_il_high_credit_limit', 'mths_since_last_major_derog','mo_sin_old_rev_tl_op'	100% data from these columns is NA
'next_pymnt_d', 'mths_since_last_record', 'mths_since_last_delinq'	Data from these columns is mostly NA (97%, 92%, 64% respectively)
'desc','title','collection_recovery_fee', 'delinq_2yrs', 'earliest_cr_line','id', 'last_pymnt_d', 'member_id', 'out_prncp', 'out_prncp_inv', 'recoveries', 'revol_bal', 'revol_util', 'total_pymnt', 'total_pymnt_inv', 'total_rec_int', 'total_rec_late_fee', 'total_rec_prncp', 'last_pymnt_amnt','funded_amnt_inv','url'	Data from these columns is not much of a help for this analysis
'pymnt_plan', 'initial_list_status', 'collections_12_mths_ex_med', 'policy_code', 'application_type', 'acc_now_delinq', 'chargeoff_within_12_mths', 'delinq_amnt', 'tax_liens'	Data from these columns has only 1 value

Pre-Analysis

Load data and drop columns

1. Load data and read 0^{th} row as header
2. Understand Percentage of NA values for each column
3. Drop Columns with 100% NA values
4. Drop Columns having NA values more than 50%
5. Drop Columns which are not useful for analysis
6. Dataset (39717, 24)

Data Transformation

1. Clean 'term' column, removed 'months' word
2. Clean 'int_rate' column, removed '%'
3. Converted 'issue_d' to month year format
4. Derive data -> added columns with segments for month from 'issue_d' and loan amount
5. Derive data -> added columns with segments for dti and int_rate
6. Derive data -> added columns with segments for zip code and annual income

A

Continued on next slide

Pre-Analysis

Data Impute and row(s) clean-up

A → Separated fully-paid and charged-off dataset for further analysis → Drop 'emp_title' as it is not fit to impute data → Impute 'emp_length' based on 'grade' column → Drop 'pub_rec_bankruptcies' as it is not fit to impute data → Duplicate rows (zero found) → Removed loan status which is in-progress → B

```python
# Employee legth can be imputed as a function of grade
for item in df['grade'].unique():
    # Get the mode for the current grade
    mode_value = df[df['grade'] == item]['emp_length'].mode()[0]
    df.loc[(df['grade'] == item) & (df['emp_length'].isna()), 'emp_length'] = mode_value
```

Outlier and Outlier treatment

B → Checked 'loan_amt', Outlier treatment not required → Checked 'funded_amnt, Outlier treatment not required → Checked 'int_rate', Outlier treatment not required → Checked 'dti', Outlier treatment not required → Checked 'installment', Outlier treatment not required → Checked 'annual_inc', carried out Outlier treatment, also checked z score → Checked 'open_acc', carried out Outlier treatment → **Final Dataset (37880, 29)**

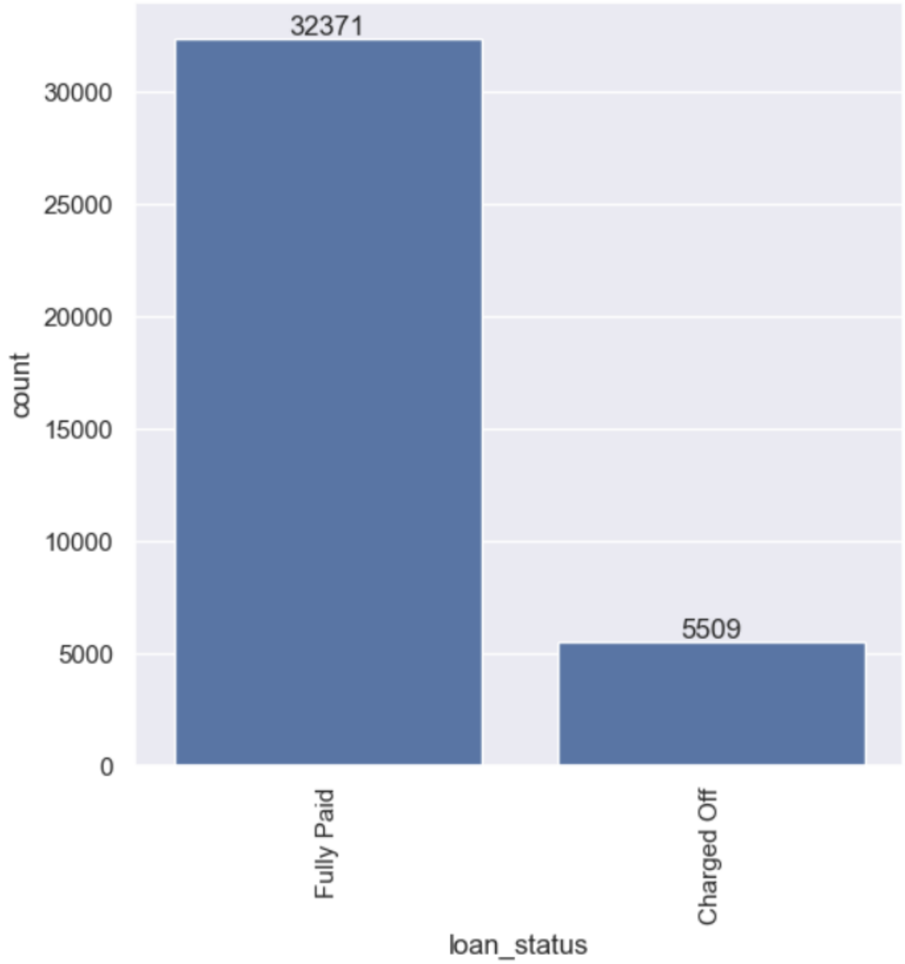
Pre-Analysis

Pre-analysis data pointers

1) Annual Income IQR is 38000.0. 25% to 75% of loan applicants have income between 40K to 78K ($)
2) Loan amount IQR is 10K. 25% to 75% of loan applicants have loan amount between 5K to 15K ($)
3) Funded amount IQR is 10K. 25% to 75% of loan applicants have funded amount between 5K to 15K ($)
4) Interest rate IQR is 5%. 25% to 75% of loan applicants have interest rate between 9% to 14%
5) Debt to Income (DTI) IQR is 19. 25% to 75% of loan applicants have DTI from 8% to 18%
6) Monthly installment amount IQR is 280. 25% to 75% of loan applicants have installments from 160 to 440 ($)
7) Typically 6 to 12 open credit lines are observed for loan applicants
8) 75% of loan applicants prefer 36 months duration

Data Visualization and Analysis

High level Summary - Fully Paid against Charged off (Defaulted) Accounts



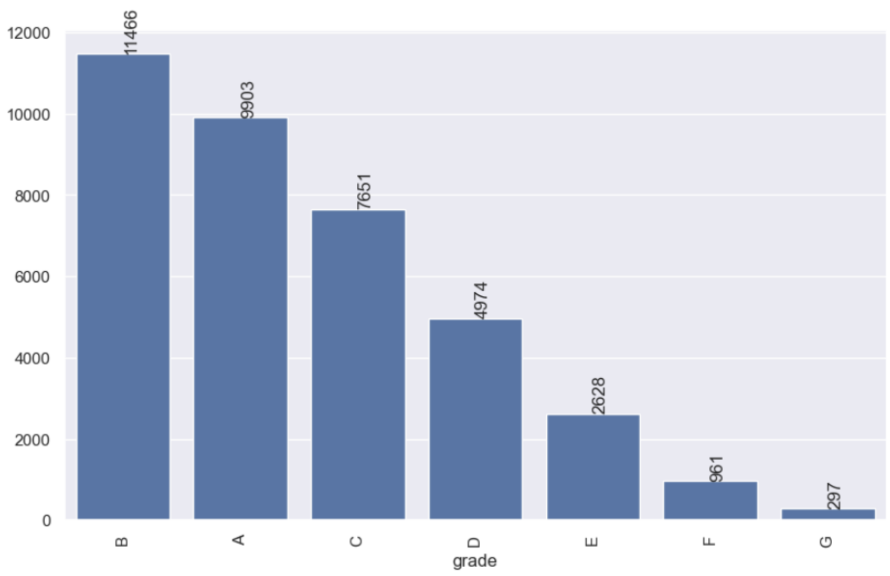
Loan Status	Percentage (Countwise)
Fully Paid	85.5 %
Charged Off	14.5 %

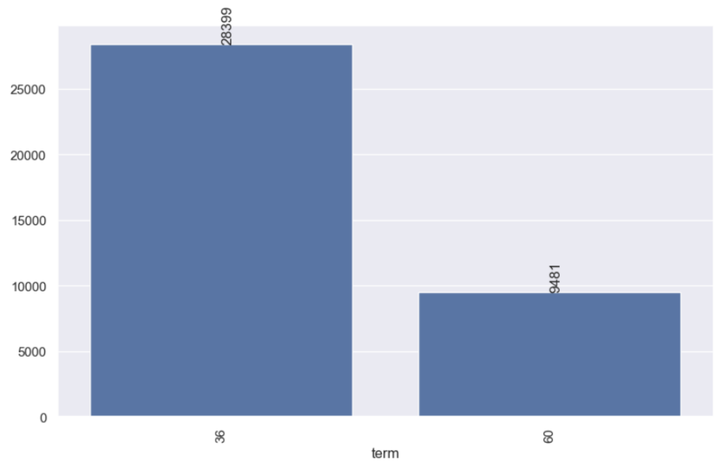
- It's important to note that default rates will vary based on the creditworthiness of borrowers and the economic environment.
- Lending Club should continue refining policies based on recent data.
- Market analysis shows Lending Club, like many peer-to-peer lending platforms, typically experiences default rates in the range of 15% to 30% for bad debt

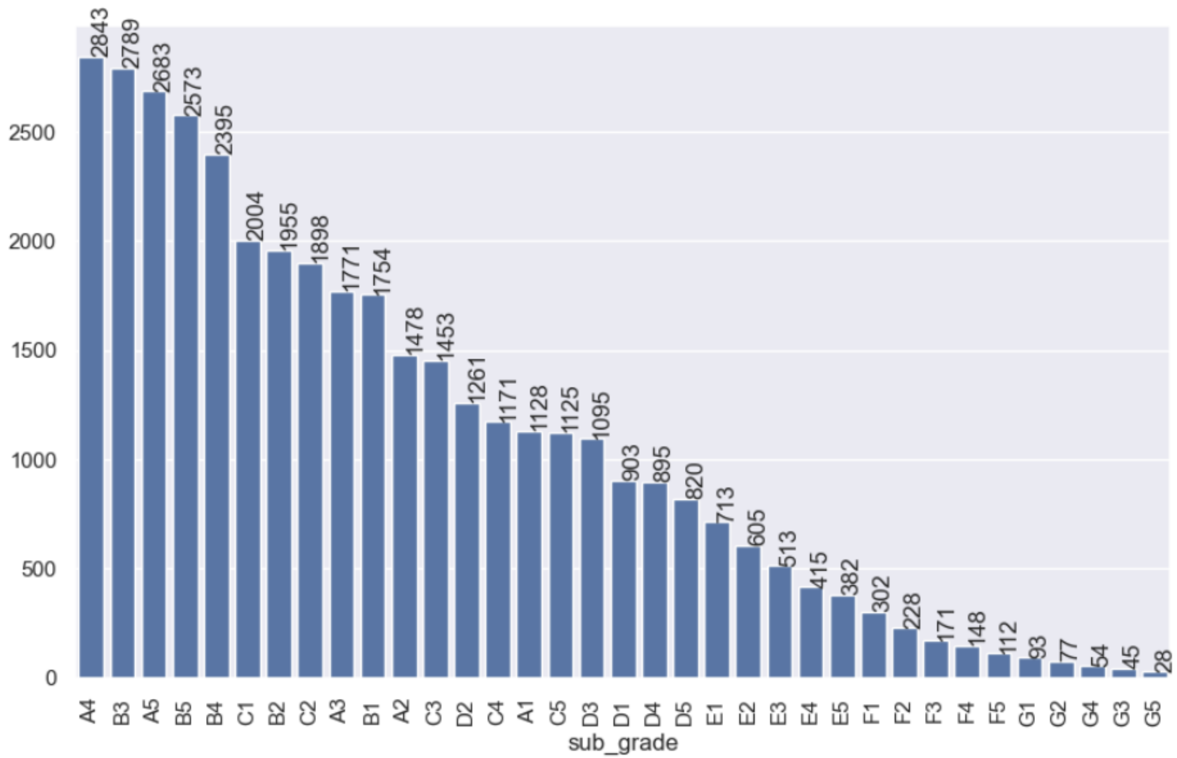
Univariate – Ordered categorical Analysis and Findings

Data Visualization and Analysis

Univariate – Ordered categorical Analysis (on complete dataset)

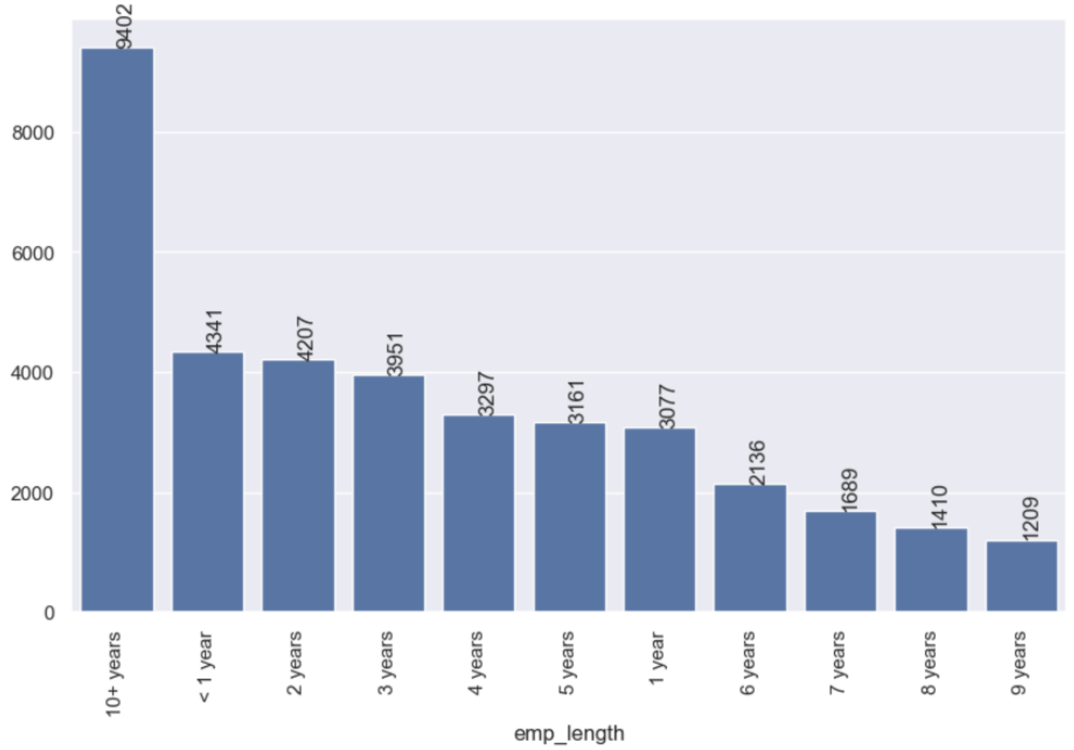


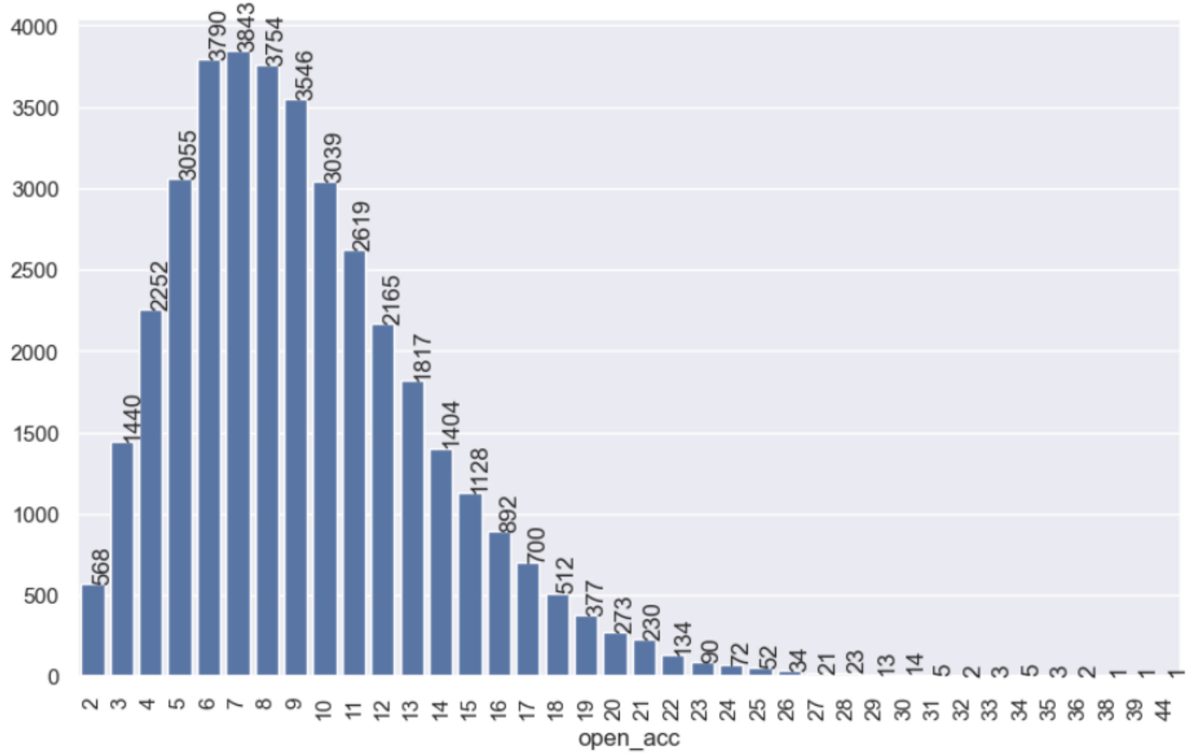




Data Visualization and Analysis

Univariate – Ordered categorical Analysis (on complete dataset)





Data Visualization and Analysis

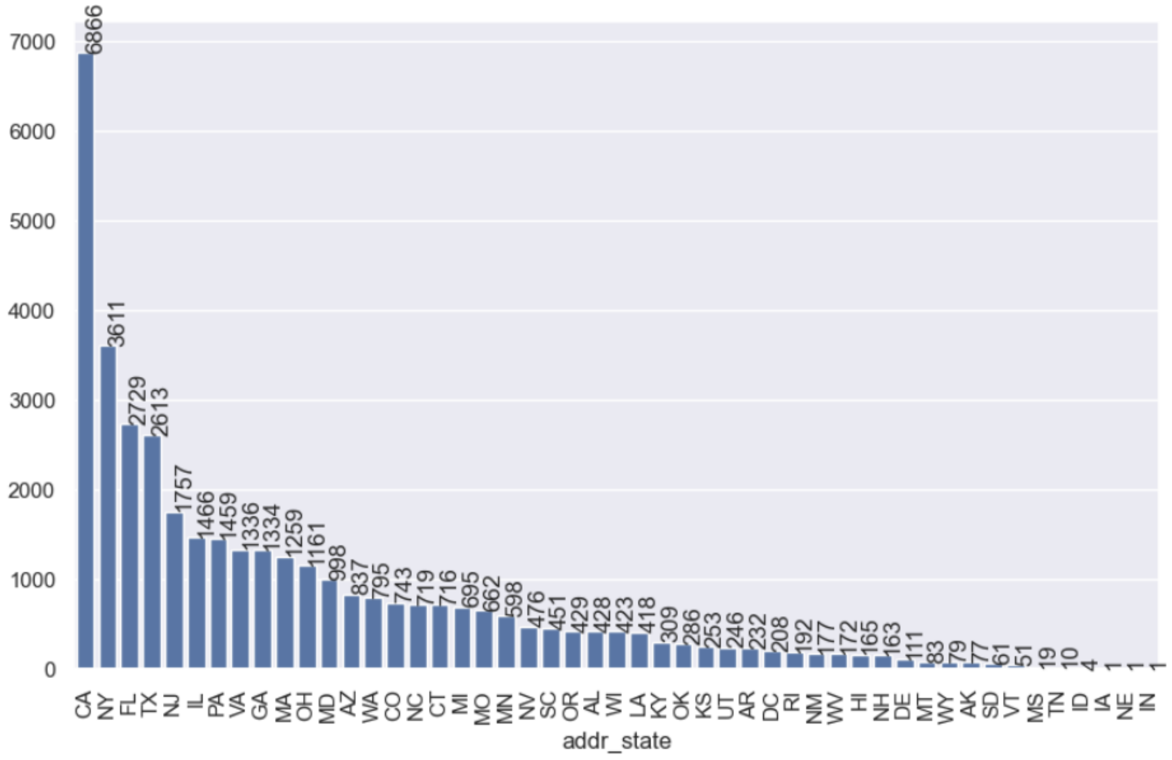
Univariate – Ordered categorical Analysis (on complete dataset) - Key Findings

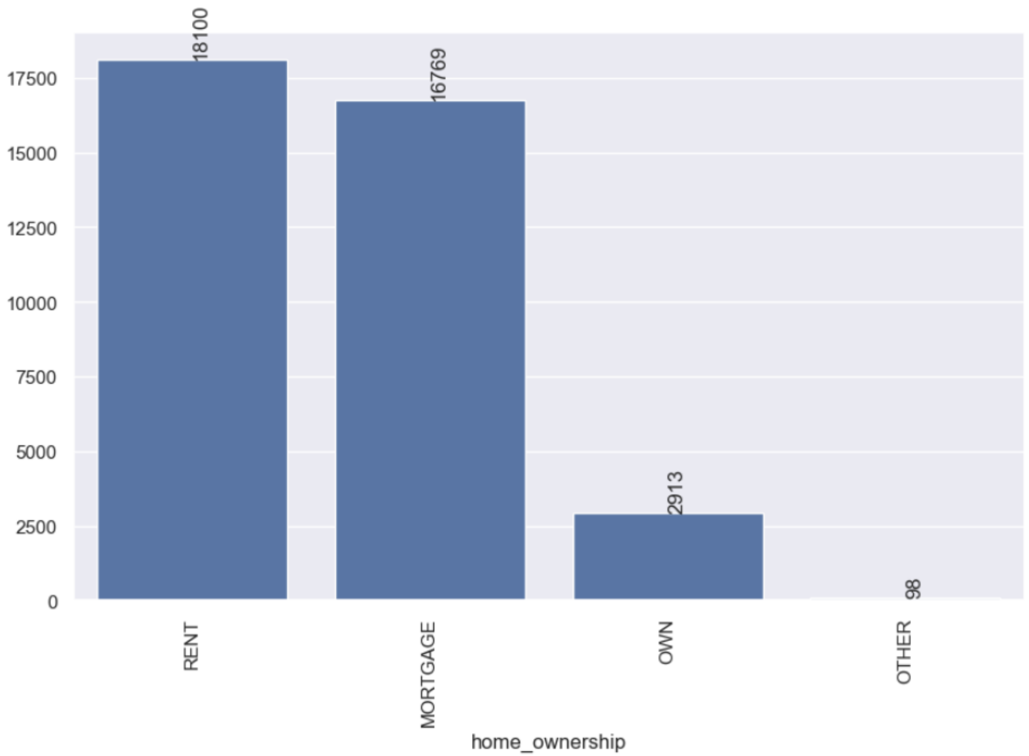
- Around 77% of accounts are from B, C and D grades. Out of which, B grade contributes the most of the loan accounts i.e., 30%
- Short term loans (tenure = 36 months) were highest in all loan accounts.
- Applicants who are employed for 10+ years appear to be contributing to highest loan accounts.
- 54% of loan accounts have open account between 5 to 10

Univariate – Unordered categorical Analysis and Findings

Data Visualization and Analysis

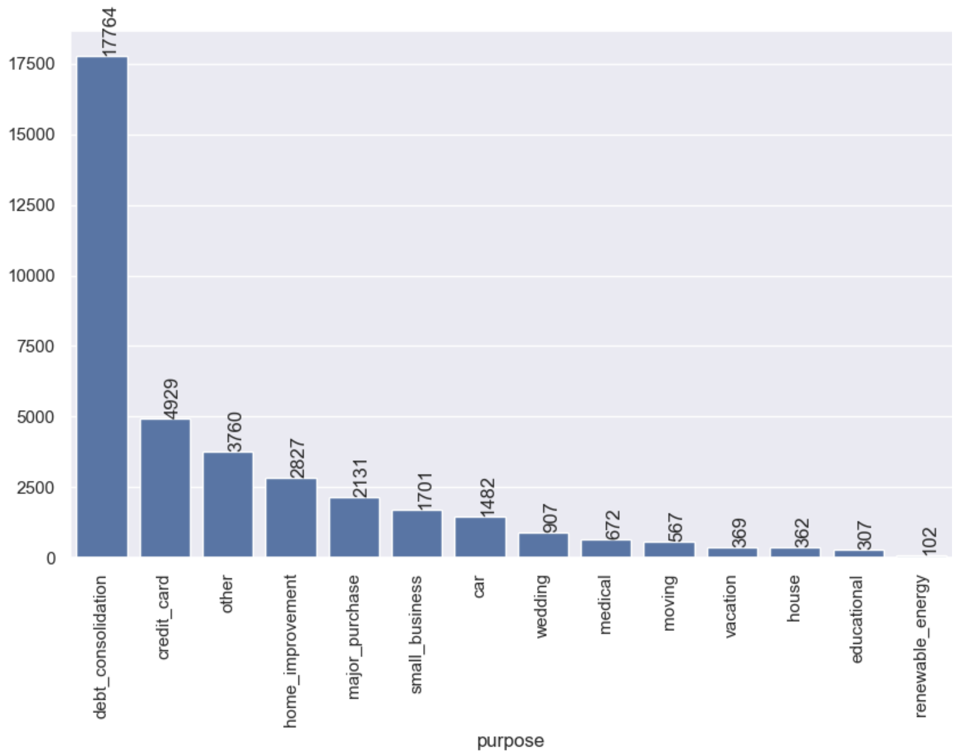
Univariate – Unordered categorical Analysis (on complete dataset)

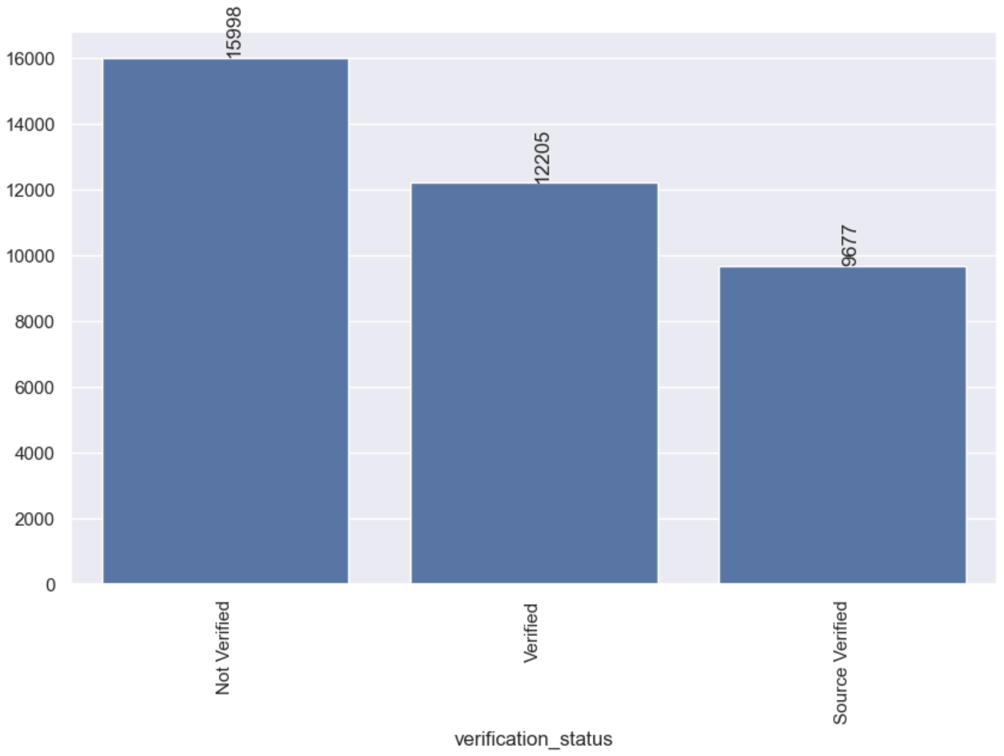




Data Visualization and Analysis

Univariate – Unordered categorical Analysis (on complete dataset)





Data Visualization and Analysis

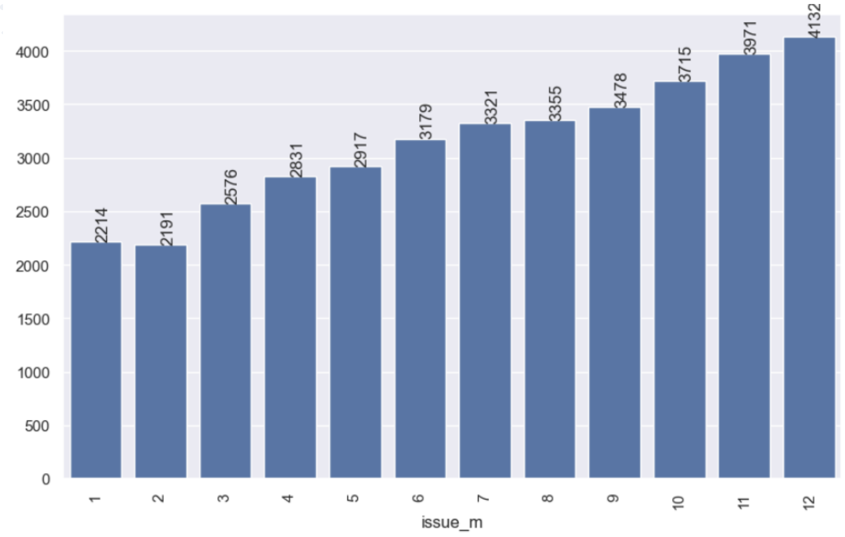
Univariate – Unordered categorical Analysis (on complete dataset) - Key Findings

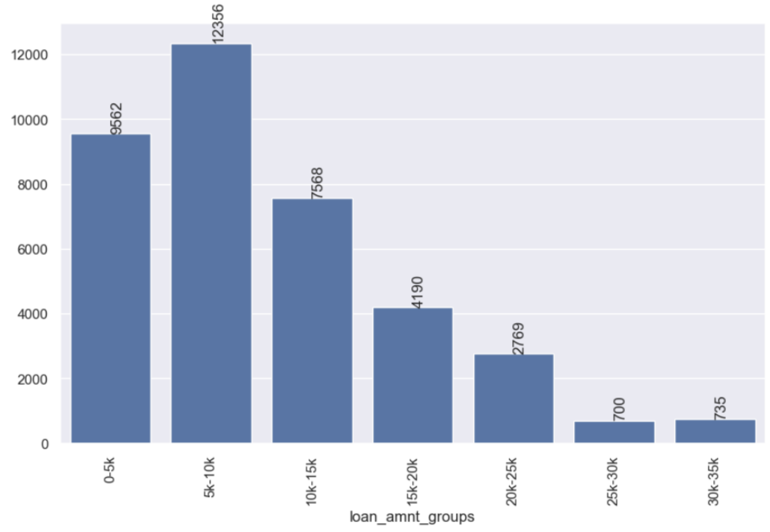
- California has the highest number applicants which is contributing 18.1% to total percent. And also California has highest number of charged off applicants as well. Hence the lending/finance company has to implement stricter rules for eligibility criteria due the higher defaulters
- Most of the applicants live in rented house contributing to 47.8%. And also the highest defaulters are from rented house category. The company must assess the financial status of rented applicants as they are more suspectable to be defaulters.
- A significant number of applicants are loan defaulters which is around 14%. The lending company has to assess more on the reasons why applicants are unable to repay the loan. They have to concentrate more on risk factor assessments etc.,
- Majority of the applicants are who had applied the loan for debt_consolidation which is contributing to 47% of the total percent. The lending company has to take strict measures while approving loan for the purpose of debt consolidation as we could observe defaulters are also high under the debt consolidation.
- Applicants who are not verified contribute to 42.2 percent to the total of verification status. It is always recommended to verify the applicants before processing the loan.

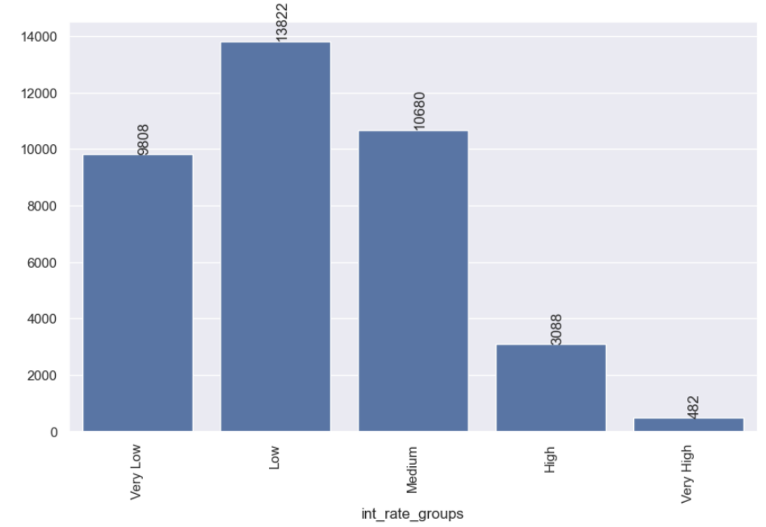
Segmented Analysis and Findings

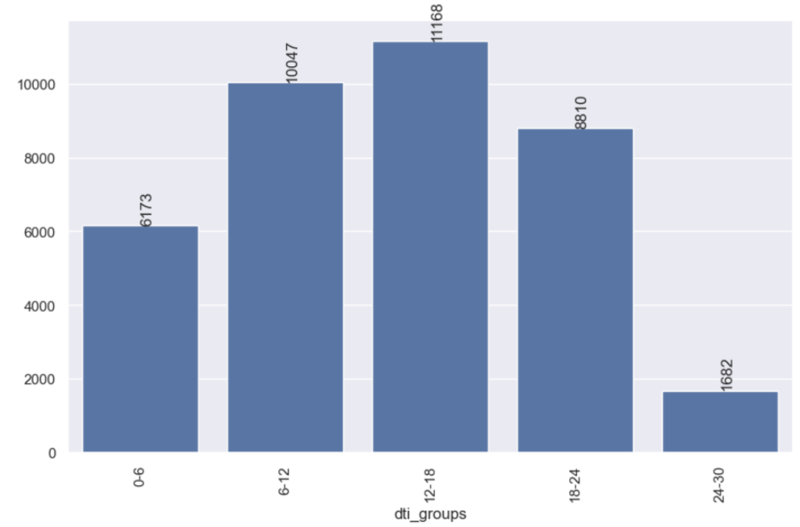
Data Visualization and Analysis

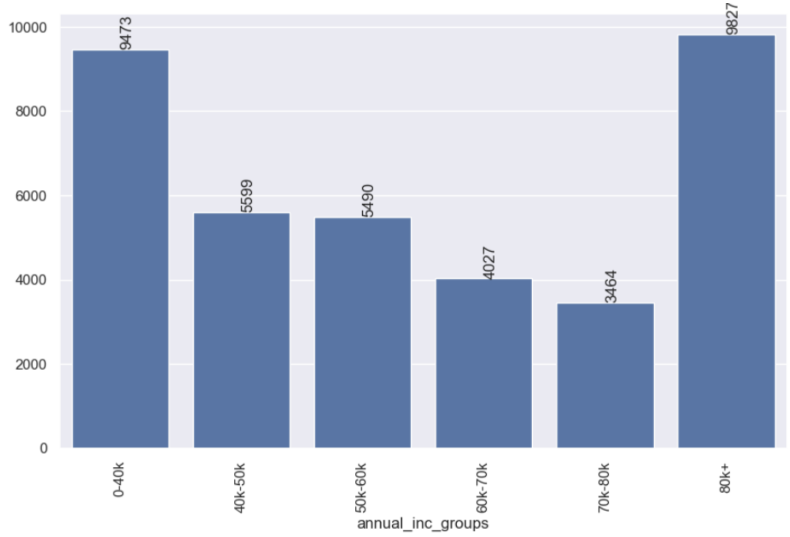
Segmented Analysis (on complete dataset)

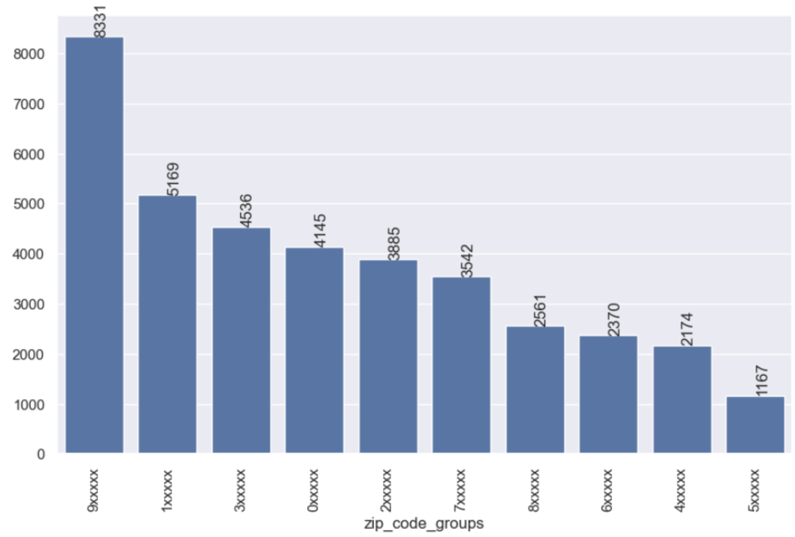












Data Visualization and Analysis

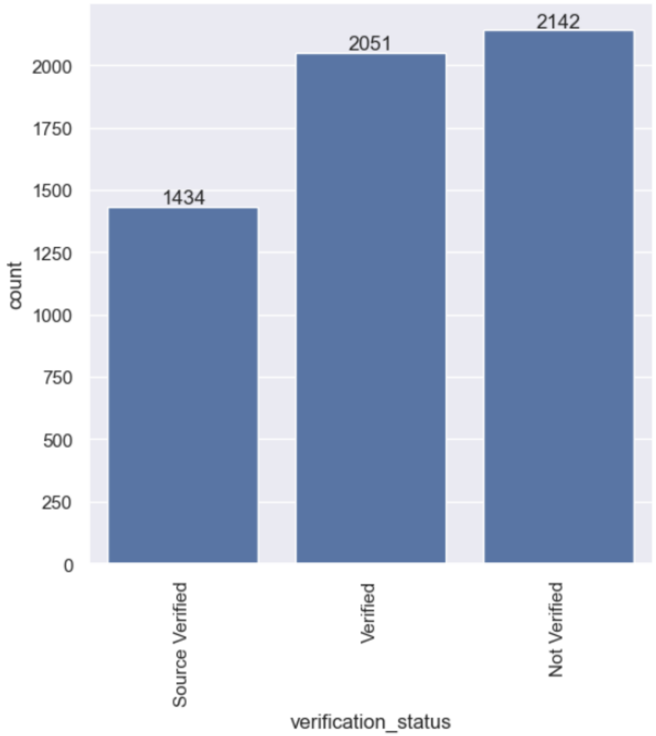
Segmented Analysis (on complete dataset) - Key Findings

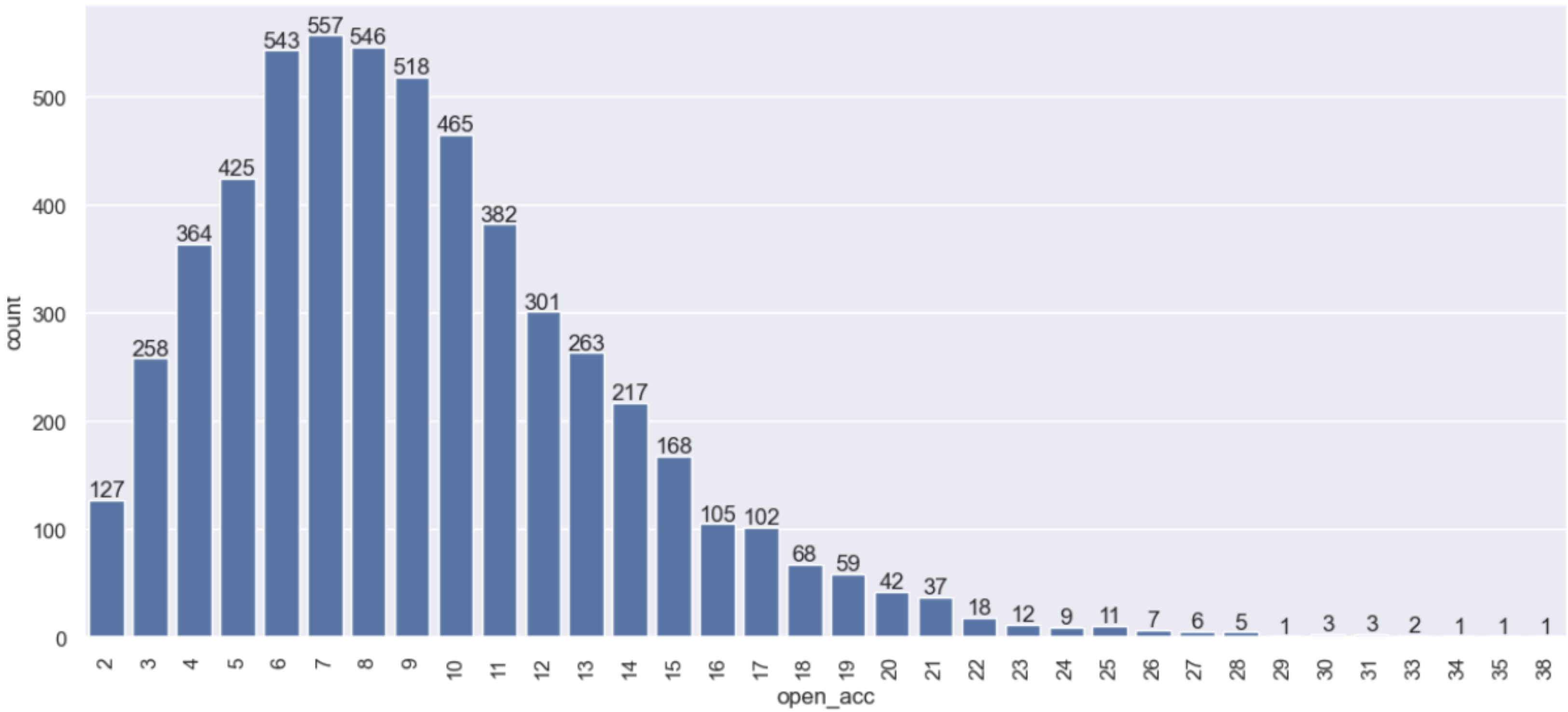
- Most of the loans are issued in the second half of the year. December is the highest month where loans are issued which contributing to 10.9%.
- loan amount in between 5k to 10k contribute to 32.6% of the total loan amount.
- Among loan participants who charged off, most of the loan applications have the high DTI ration. The lending company should implement strict DTI ratio requirements to prevent lending to individuals with unsustainable levels of debt relative to their income.
- 45% of loan applicants had annual salaries of 0 to 40K and 80k+. Out of which the charged off applicants are under the income group of 0 to 40k. Hence it is always recommended to calculate the repayment ability of the applicant
- 36.5 % of loan applicants belong to low interest rate groups. It is also observed that where interest rate is medium are having more defaulters. It is advisable to reduce the interest rate in order to get the loan retrieved

Deep dive - Charged off (Defaulted) and Fully-Paid Accounts Analysis

Data Visualization and Analysis

Charged off (Defaulted) Accounts Analysis



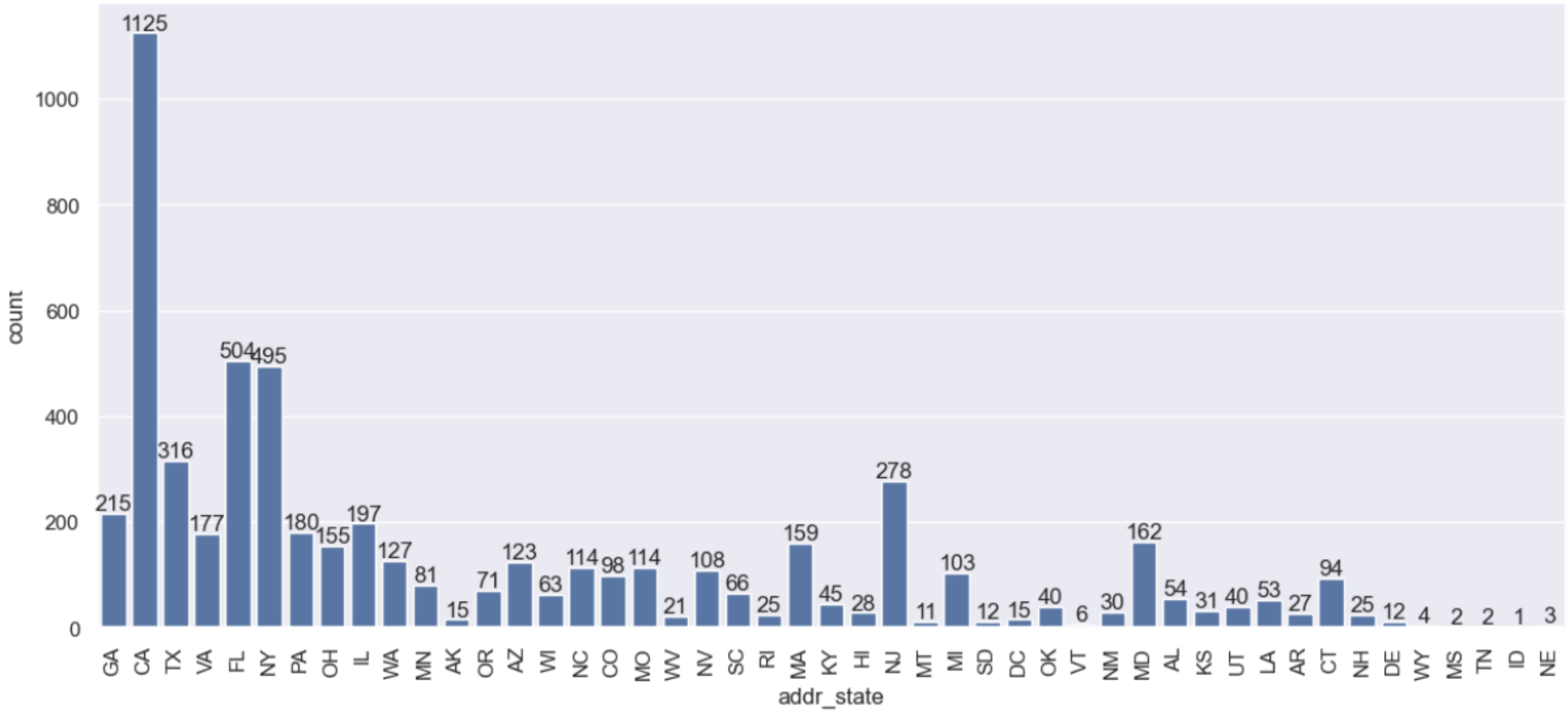


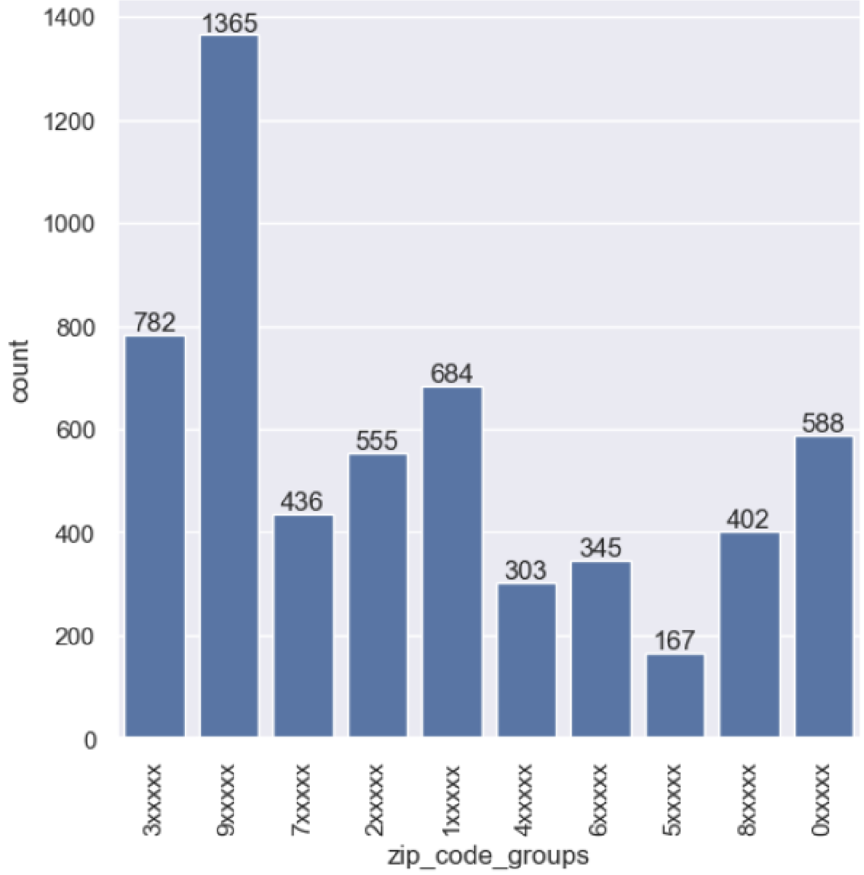
Key Findings

1. Defaulters % for Source verified (25%) is lesser than verified and non-verified (74%)
2. Almost 74% of defaulted borrowers are from category verified and non-verified
3. 54% of defaulted accounts have open account between 5 to 10
4. Open accounts less than 5 and more than 15 shows less defaulting tendency

Data Visualization and Analysis

Charged off (Defaulted) Accounts Analysis



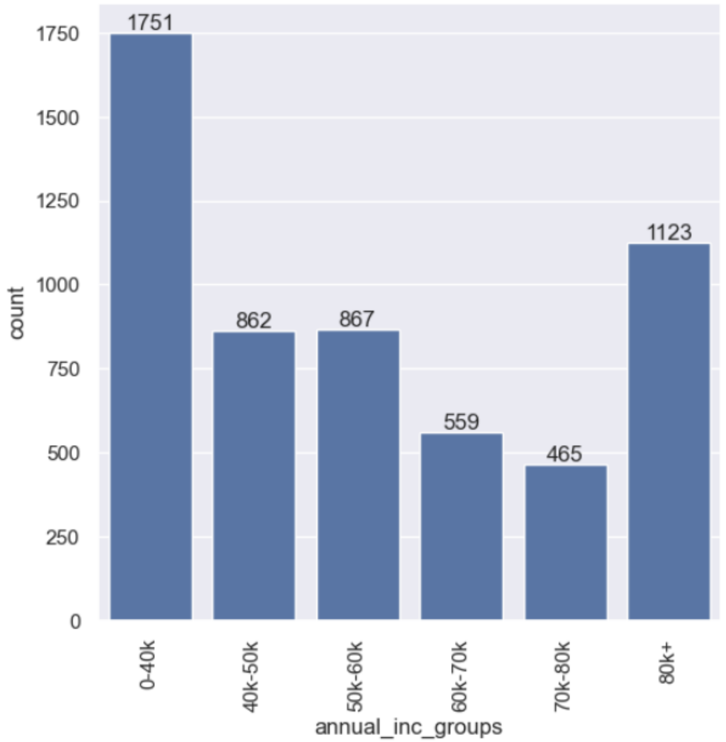


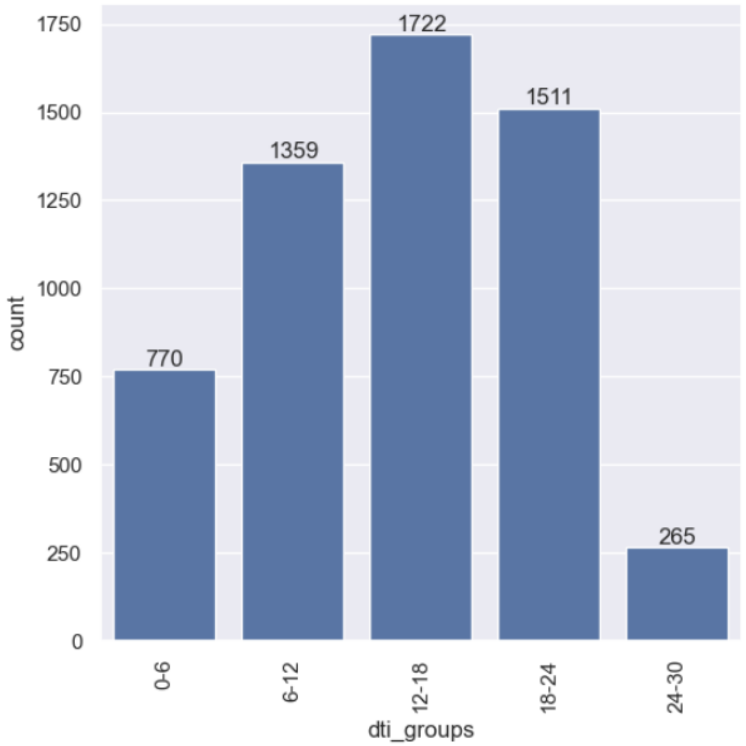
Key Findings

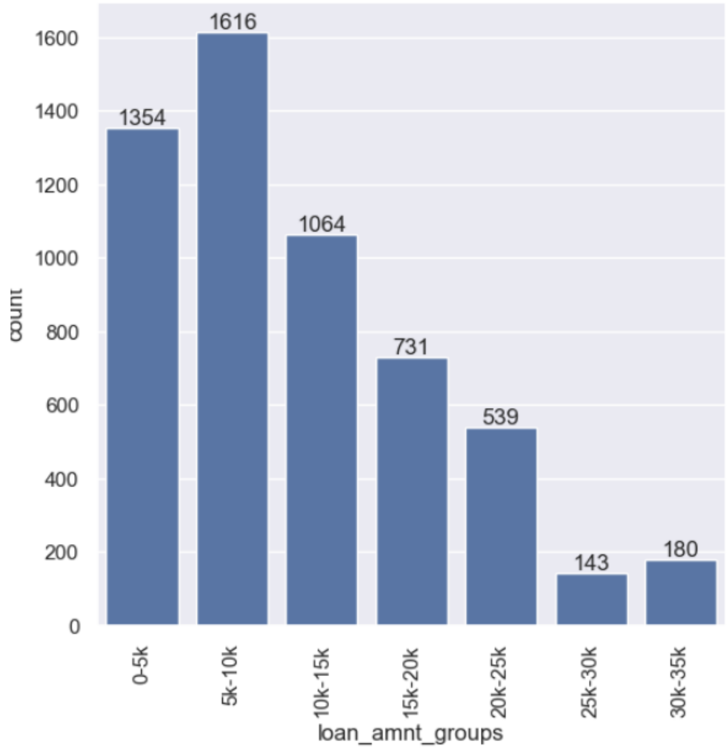
1. States CA, FL, NY, TX, NJ contributes ~48% of defaulters
2. Zip Code starts 9xxxxx, 3xxxxx, 1xxxxx contributes ~50% of defaulters

Data Visualization and Analysis

Charged off (Defaulted) Accounts Analysis





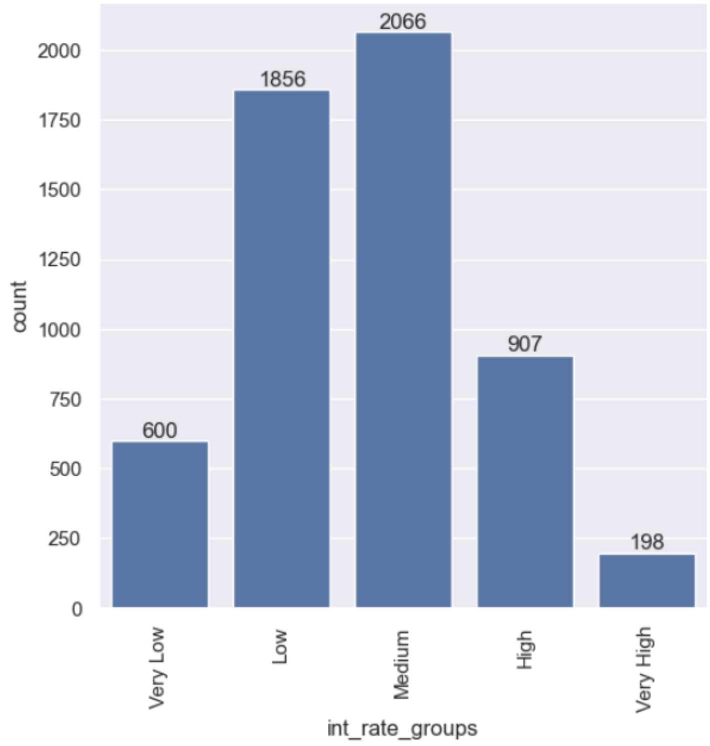


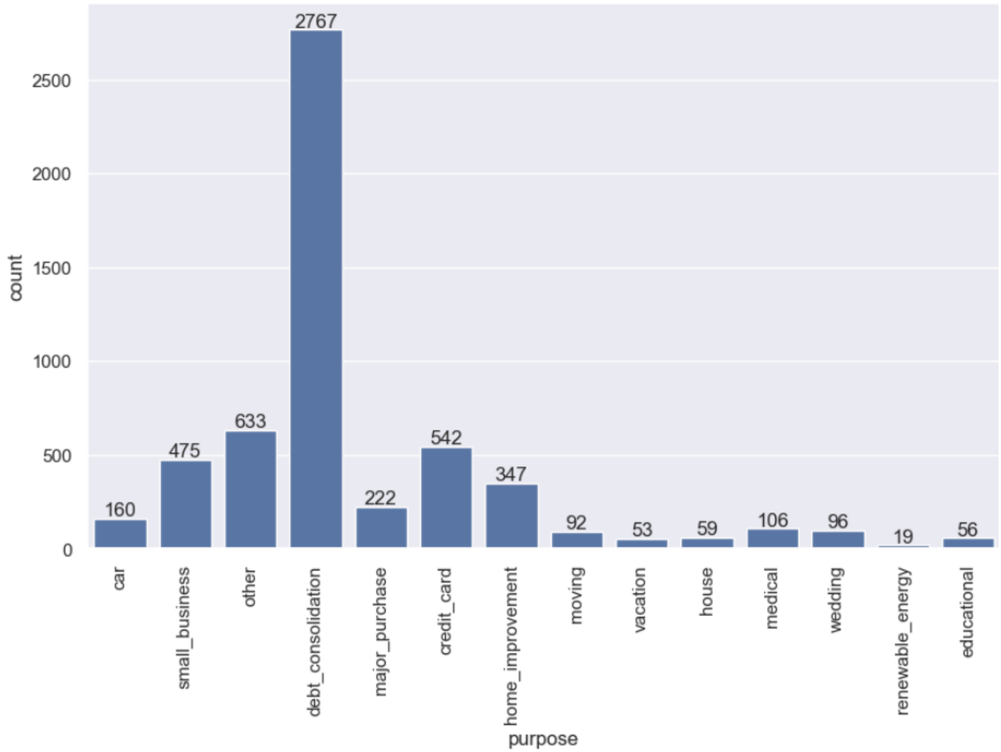
Key Findings

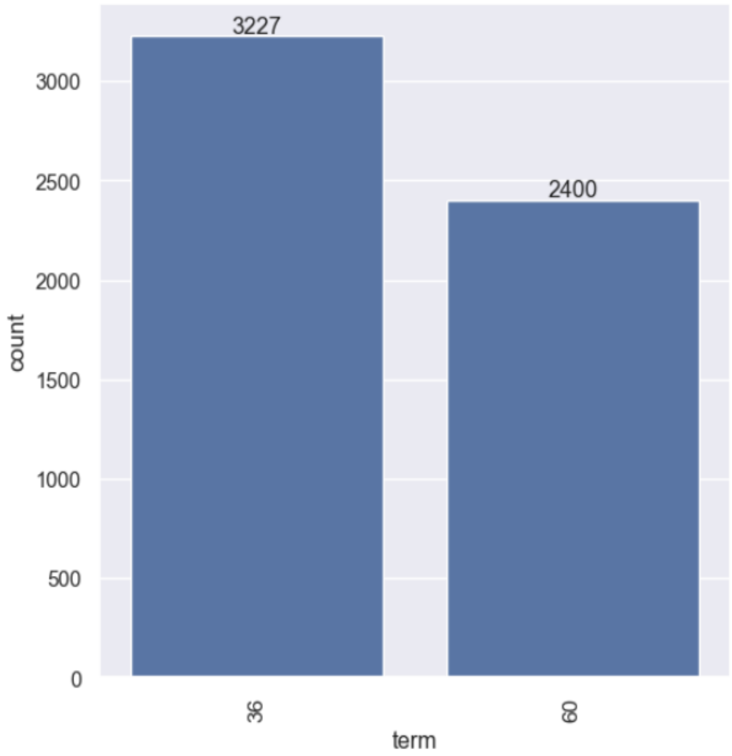
1. ~62% of defaulted loan applicants had income till 60K
2. ~38% of defaulted loan applicants had DTI with 12%
3. ~72% of defaulted borrowers have loan amount till 15K

Data Visualization and Analysis

Charged off (Defaulted) Accounts Analysis





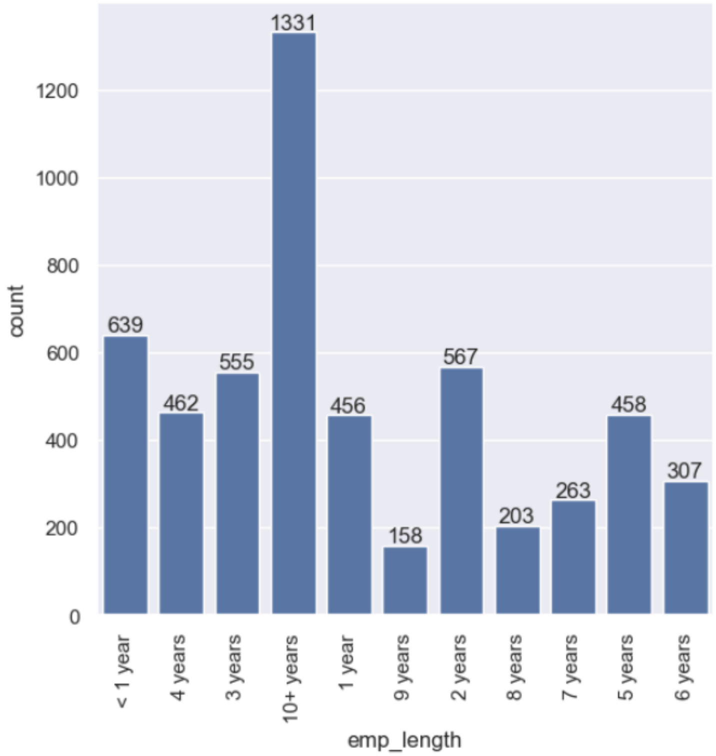


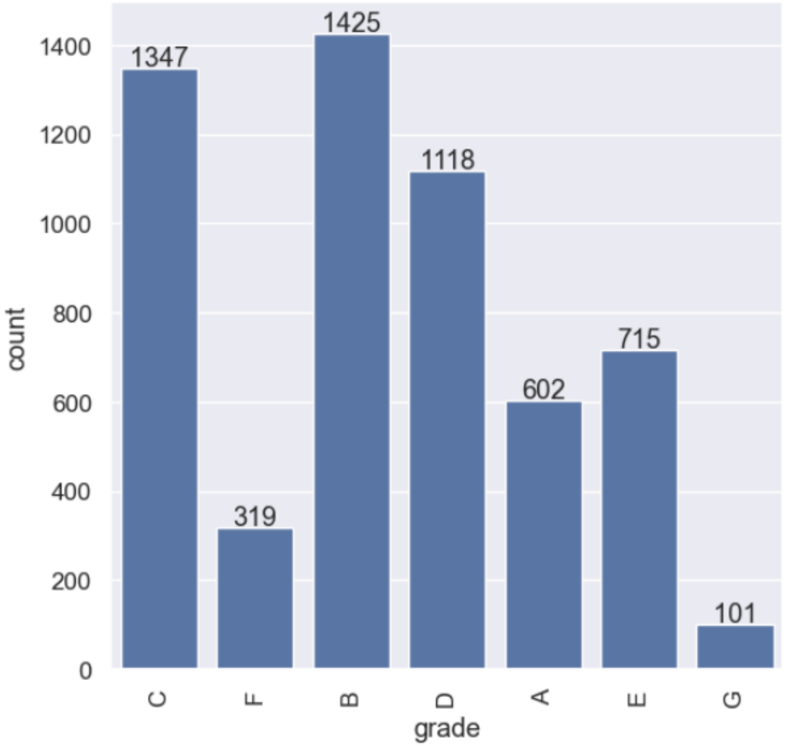
Key Findings

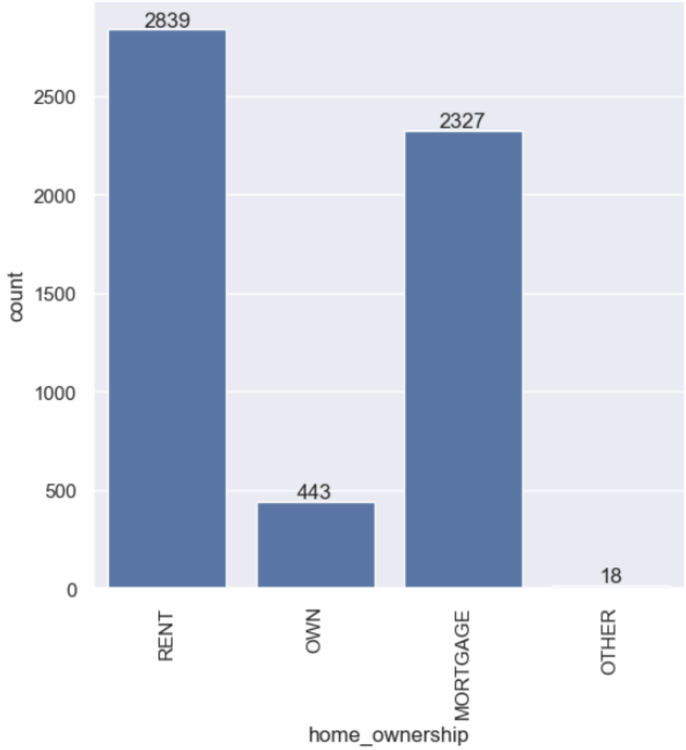
1. 67% defaulted loan applicants had interest rate between Low and Medium
2. 59% of defaulted loan applicants borrowed to debt_consolidation and for credit cards bills
3. 57% of locant applicants having 36 months term are defaulting

Data Visualization and Analysis

Charged off (Defaulted) Accounts Analysis







Key Findings

1. 36% of defaulters have working experience more than 7 years
2. 69% of defaulted accounts are from B, C and D grade
3. 50% of defaulters does not own house.

Data Visualization and Analysis

Charged off (Defaulted) Accounts Analysis

Top 15 Employers for defaulters

```
top_15_defaults
```

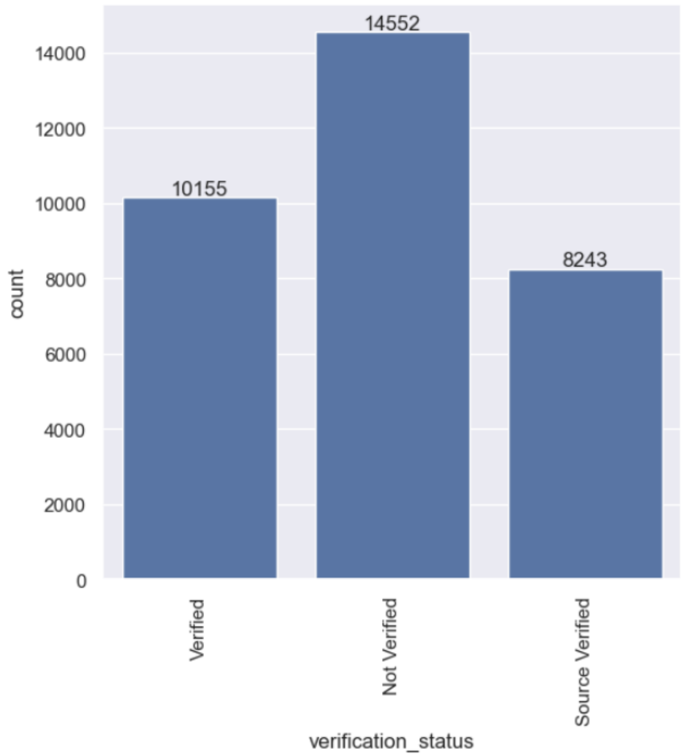
	emp_title	Default_Count
0	Bank of America	20
1	US Army	18
2	Walmart	14
3	UPS	12
4	AT&T	12
5	IBM	9
6	Kaiser Permanente	8
7	U.S. Army	8
8	Target	7
9	USPS	7
10	US Postal Service	7
11	verizon wireless	6
12	US Air Force	6
13	Home Depot	6
14	JP Morgan Chase	6

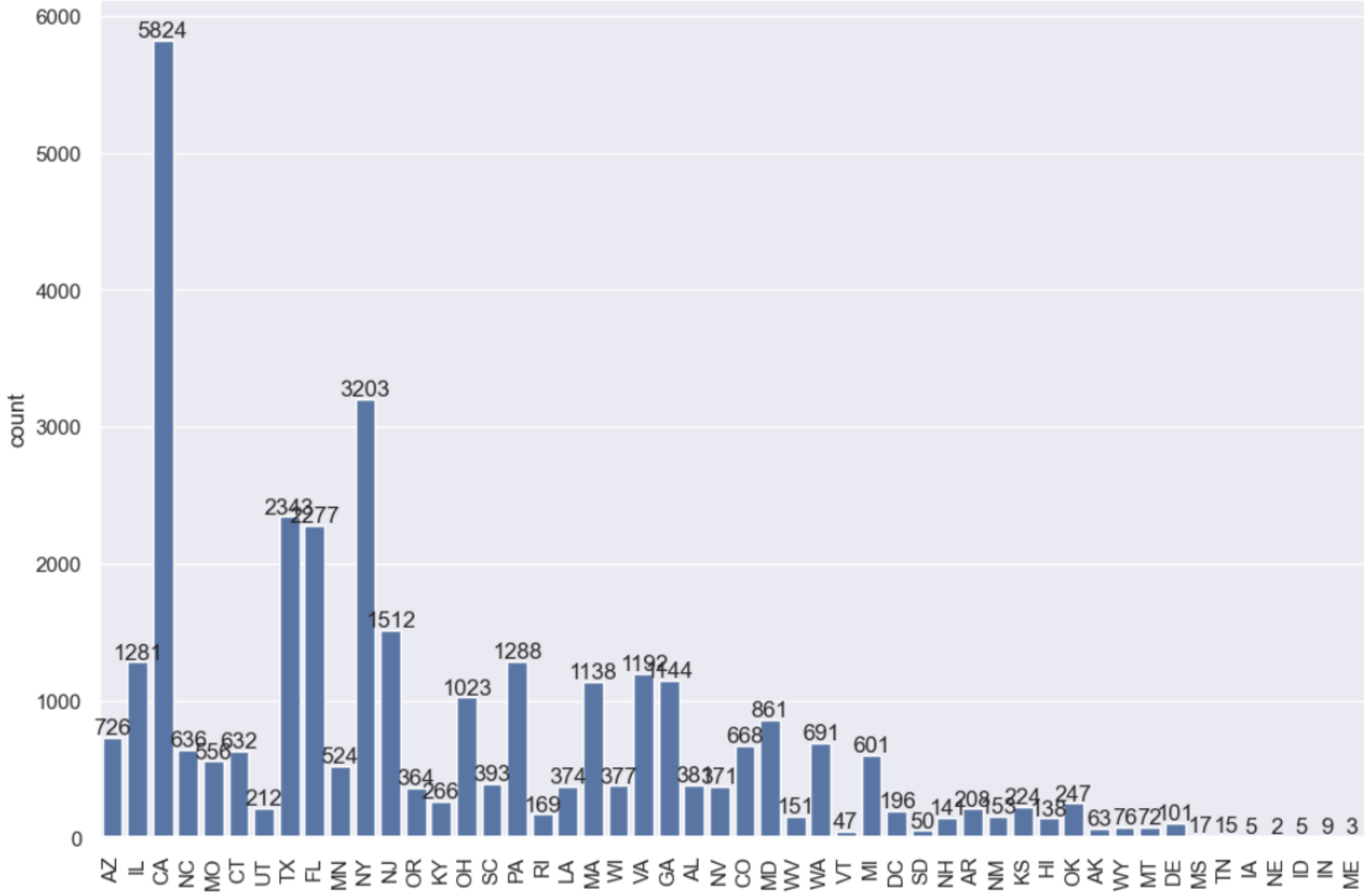
Key Findings

1. Top 15 employers contributes 146 defaulters out of 5509 defaulters. Nothing conclusive on employer title

Data Visualization and Analysis

Fully-Paid Accounts Analysis



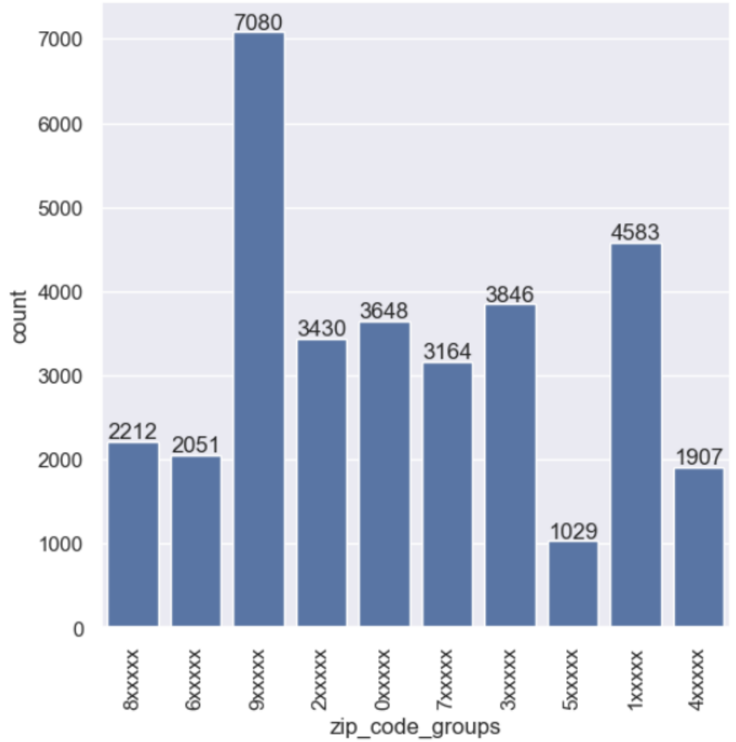


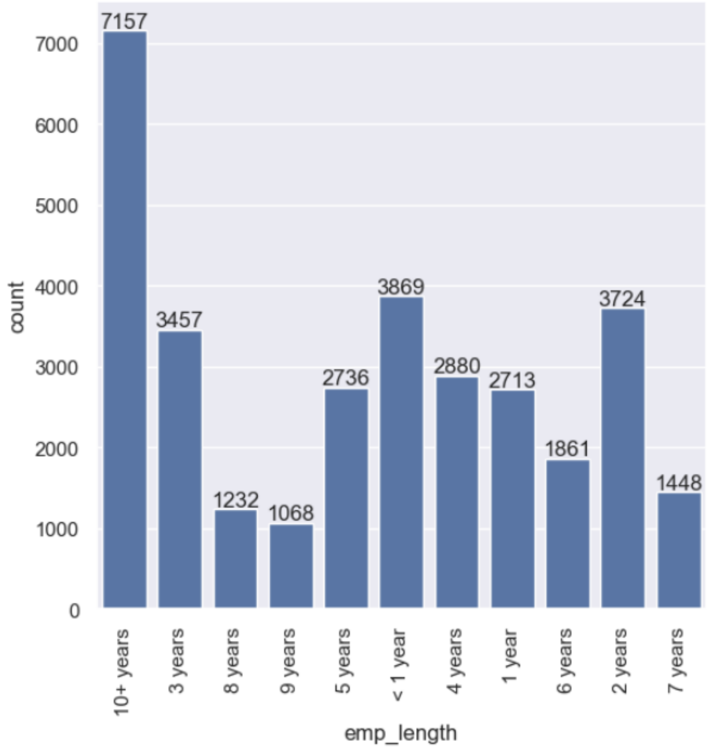
Key Findings

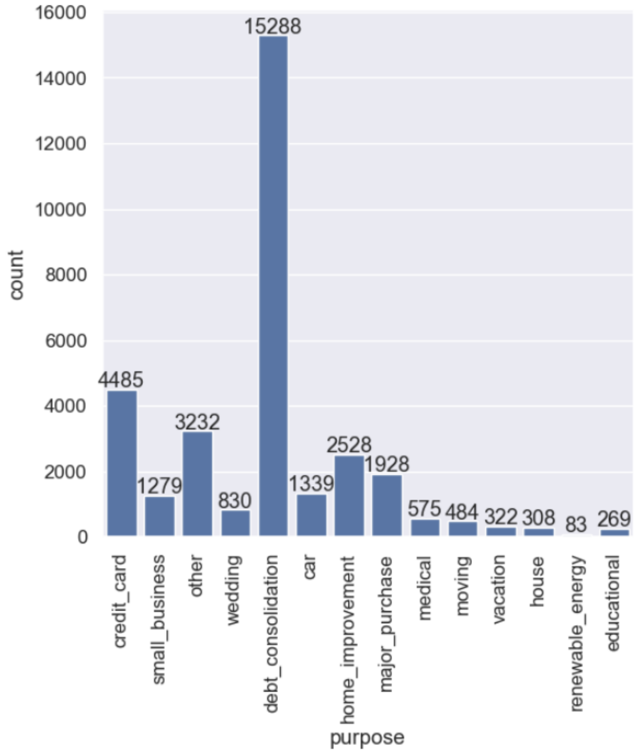
1. 56% of fully paid borrowers were verified
2. 46% of fully paid borrowers are from states CA, FL, NY, TX, NJ

Data Visualization and Analysis

Fully-Paid Accounts Analysis





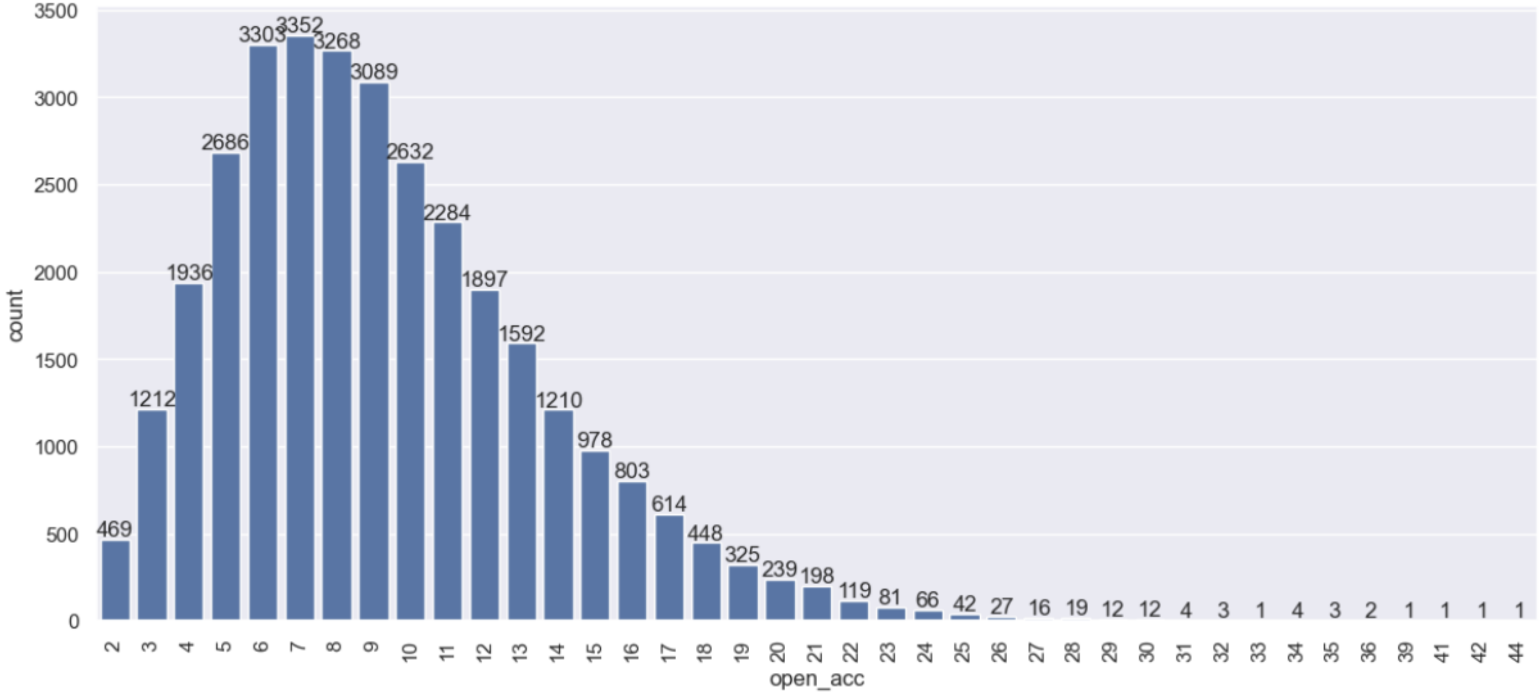


Key Findings

1. 47% of fully paid borrowers are from Zip Code starts with 9xxxxx, 3xxxxx, 1xxxxx
2. 34% of fully paid borrowers are having working experience more than 7 years
3. 60% of fully paid borrowers, borrowed for the purpose for debt_consolidation and for credit cards bills

Data Visualization and Analysis

Fully-Paid Accounts Analysis



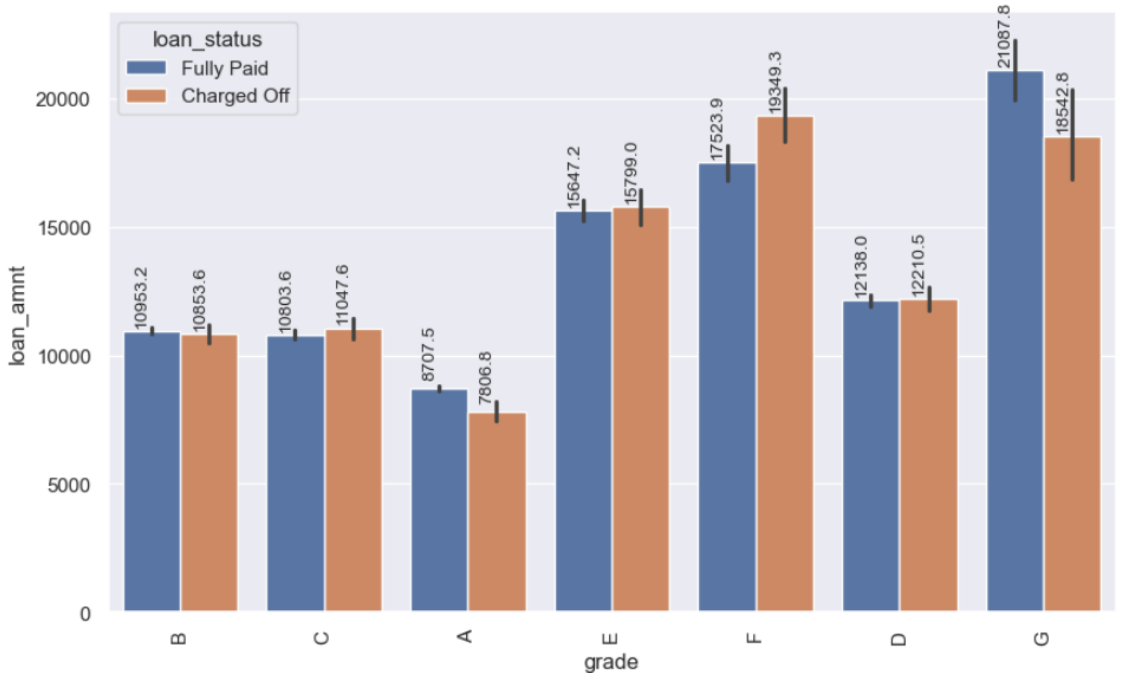
Key Findings

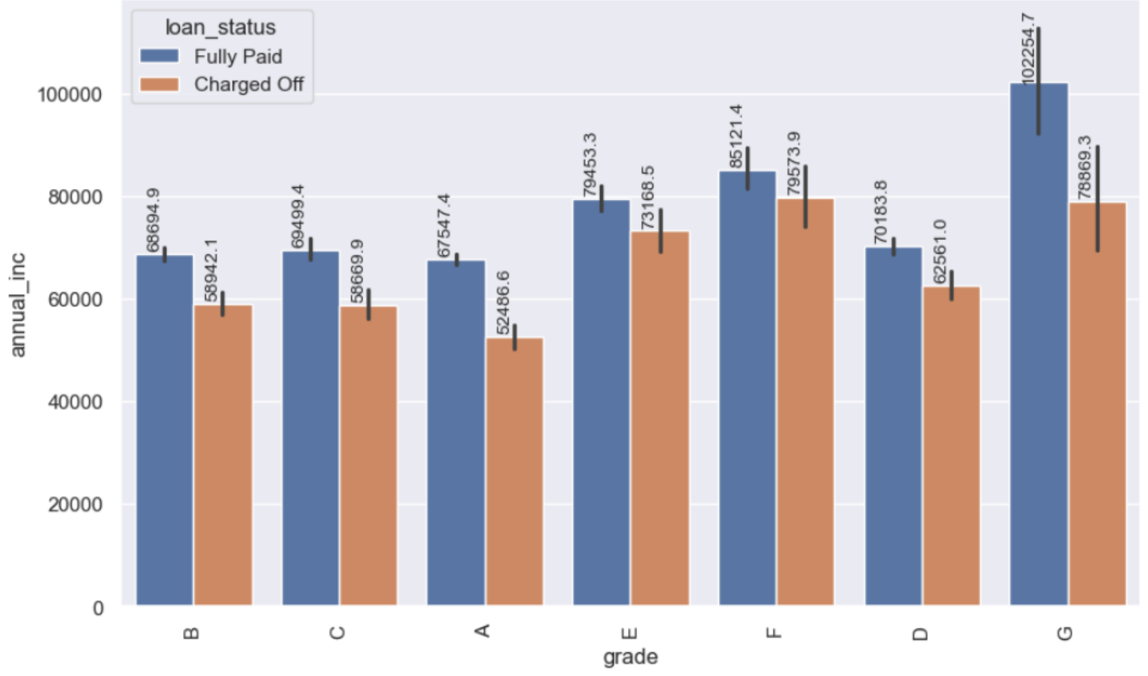
1. 56% of fully paid borrowers had open account between 5 to 10

Bivariate Analysis

Data Visualization and Analysis

Bivariate Analysis (on complete dataset)



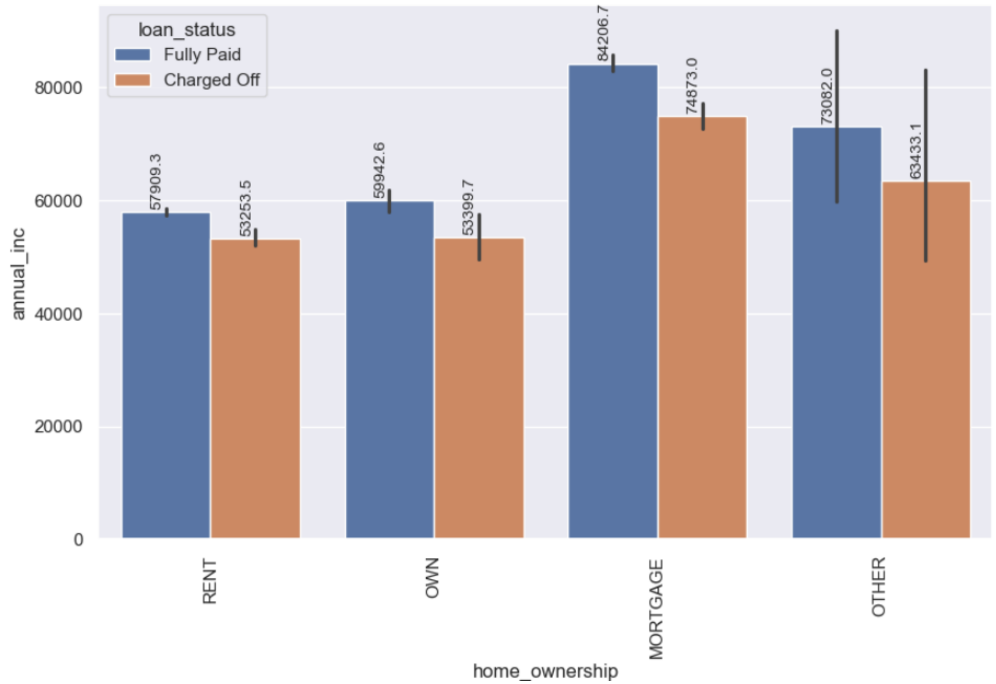


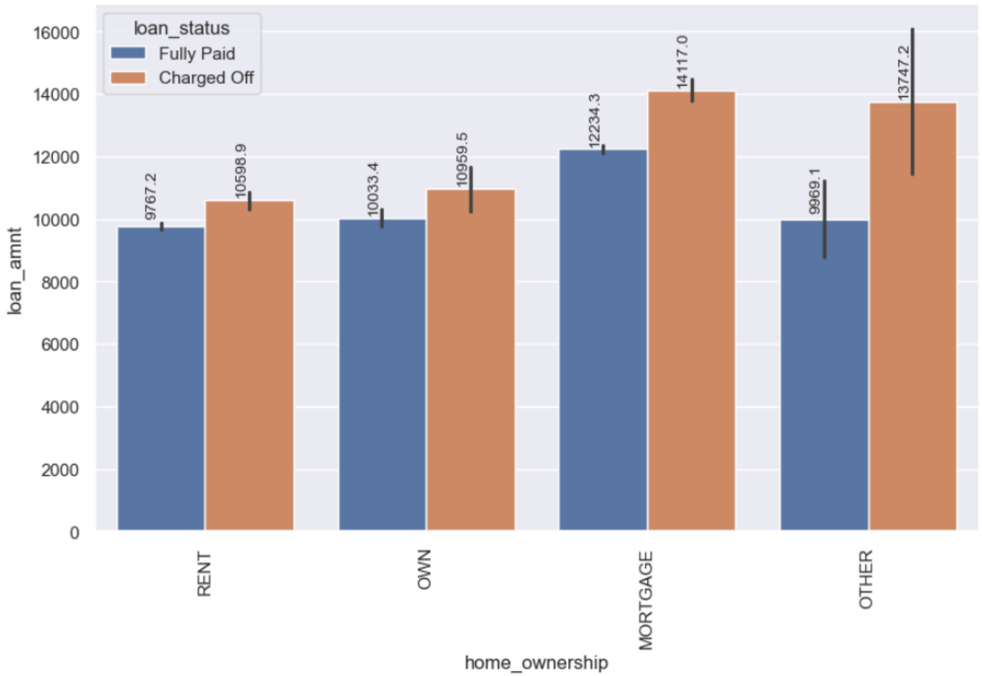
Key Findings

1. Across all the grades "Charged off" loans are almost equal to "Full paid" loans. Therefore we may need to look into additional factors to get more insights
2. For each grade, the annual income of the applicants with "Fully paid" loans is generally higher than those retrieved with charged off loans. This shows us that higher income applicants are more likely to repay their loans fully.

Data Visualization and Analysis

Bivariate Analysis (on complete dataset)



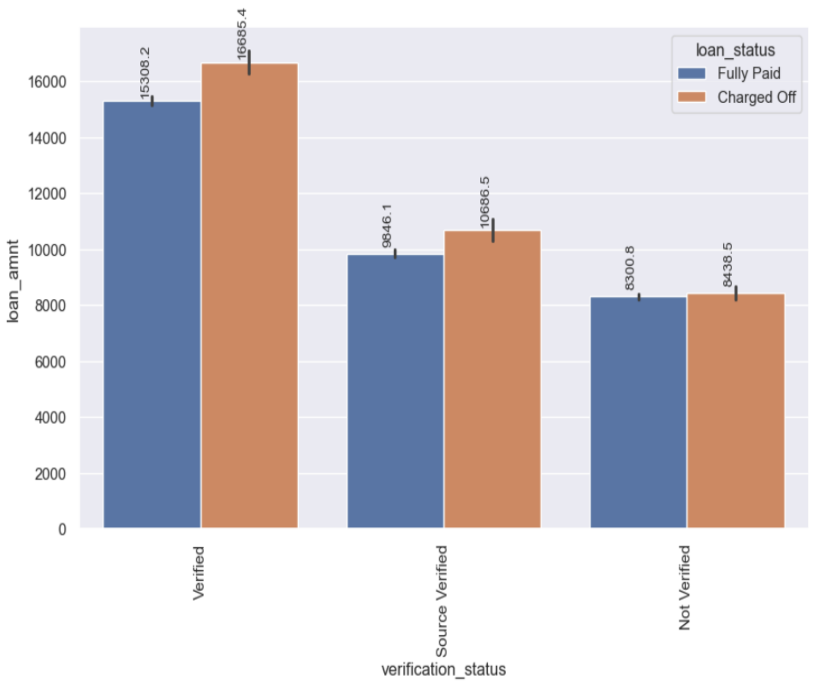


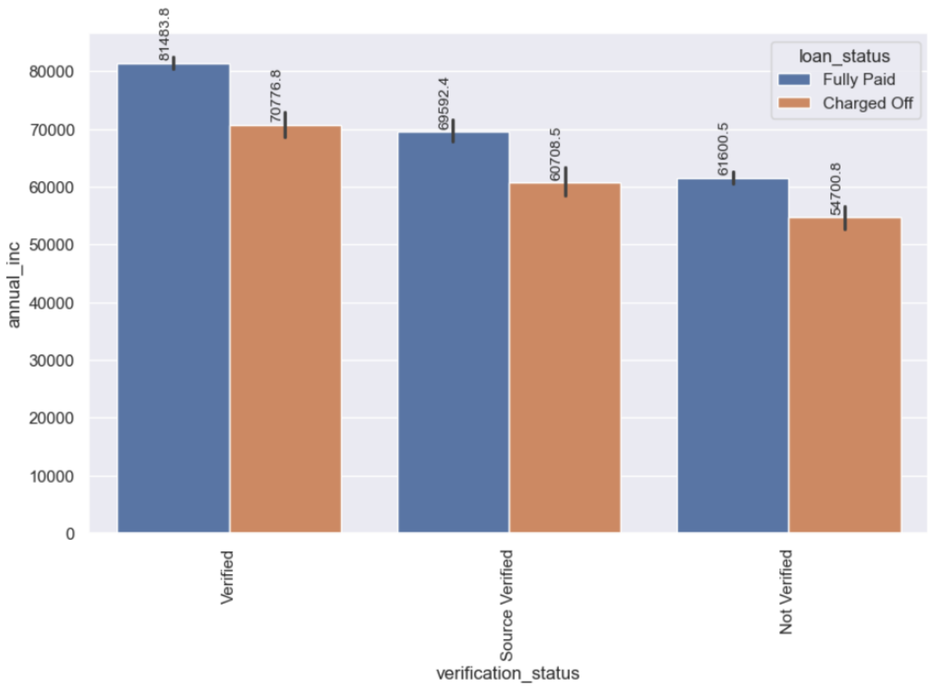
Key Findings

1. Fully paid applicants have higher annual income when compared to charged off employees.
2. Applicants under RENT have the lowest annual income.
3. The significant difference in annual income can be clearly seen for certain categories like Mortgage and Rent.
4. Charged off loans generally have higher loan amounts compared to Fully paid loans
5. Applicants under Rent category have lowest average loan amounts, regardless of loan status.
6. Applicants under Mortgage and Other category tend to take larger loans which may increase in defaulters.

Data Visualization and Analysis

Bivariate Analysis (on complete dataset)



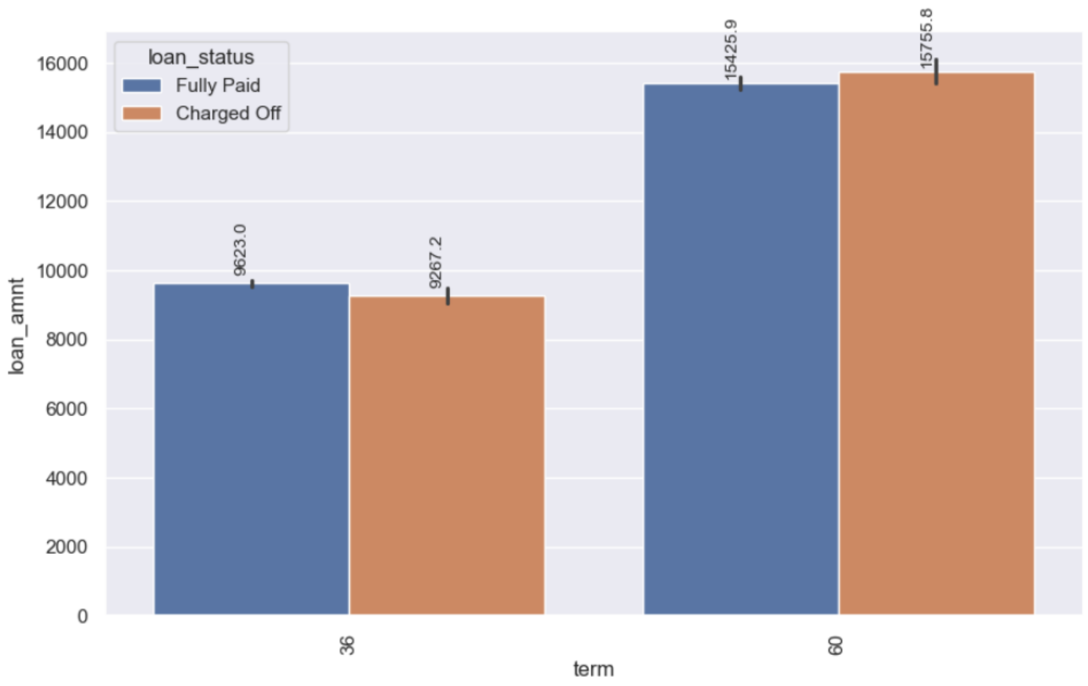


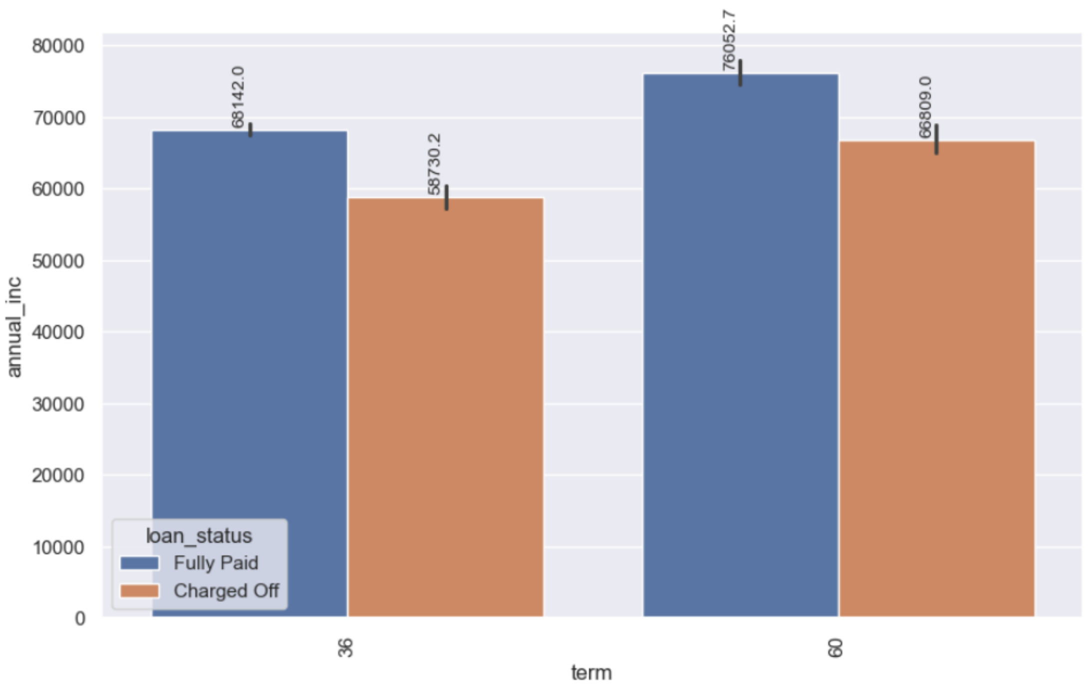
Key Findings

1. Among all the verification status, charged off loans has higher loan amounts when compared to fully paid loans.
2. Applicants with verification status "Verified" tend have higher loan amounts
3. Applicants with verification status "Not Verified" have lower loan amounts may be due to limitations by lending company.
4. Loans with higher annual income are likely to be repaid more when compared to other status.
5. Applicants with Not Verified status have highest proportion with charged off loans.

Data Visualization and Analysis

Bivariate Analysis (on complete dataset)



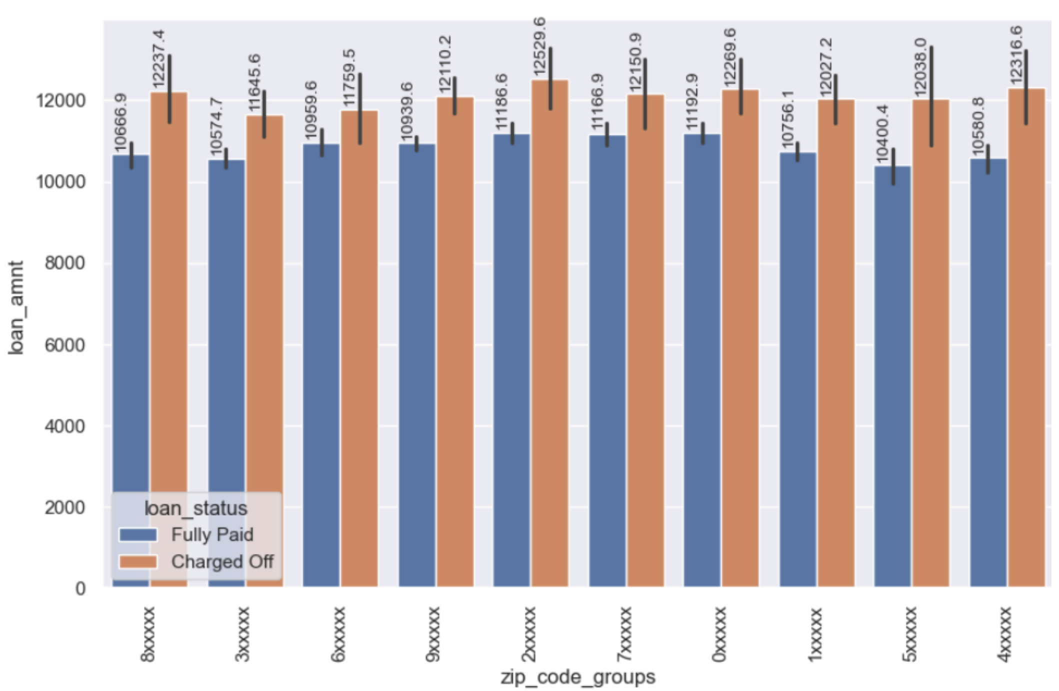


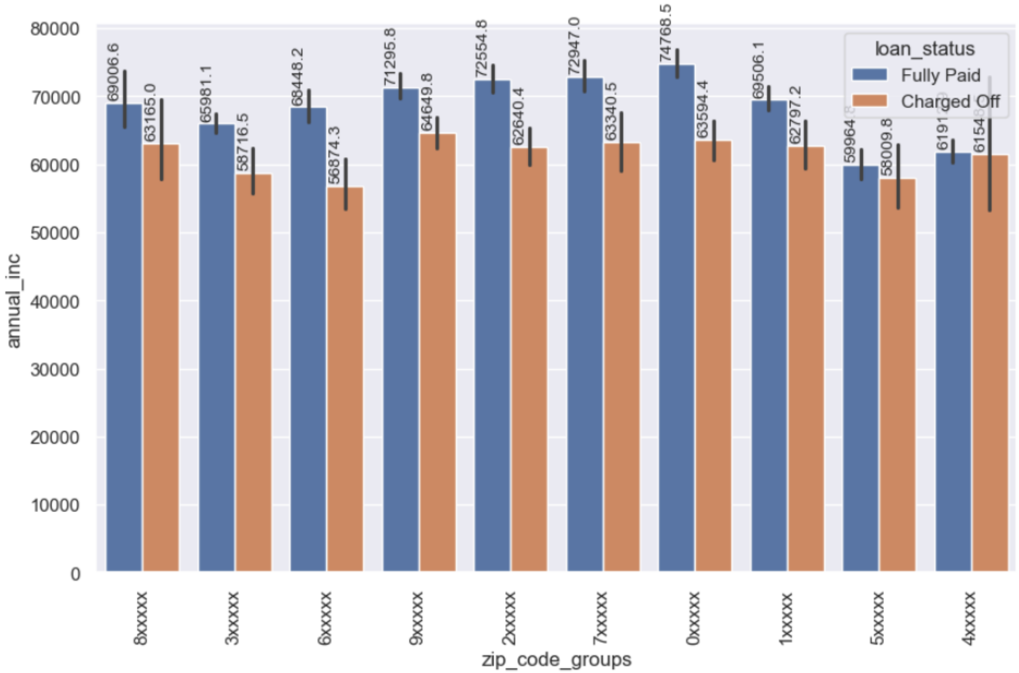
Key Findings

1. The longer loan terms are associated with higher loan amounts and higher risk of repayment.
2. Across both the terms applicants with higher annual income appear to repay the loans in higher numbers where Applicants with lower annual income tend to be defaulters.

Data Visualization and Analysis

Bivariate Analysis (on complete dataset)



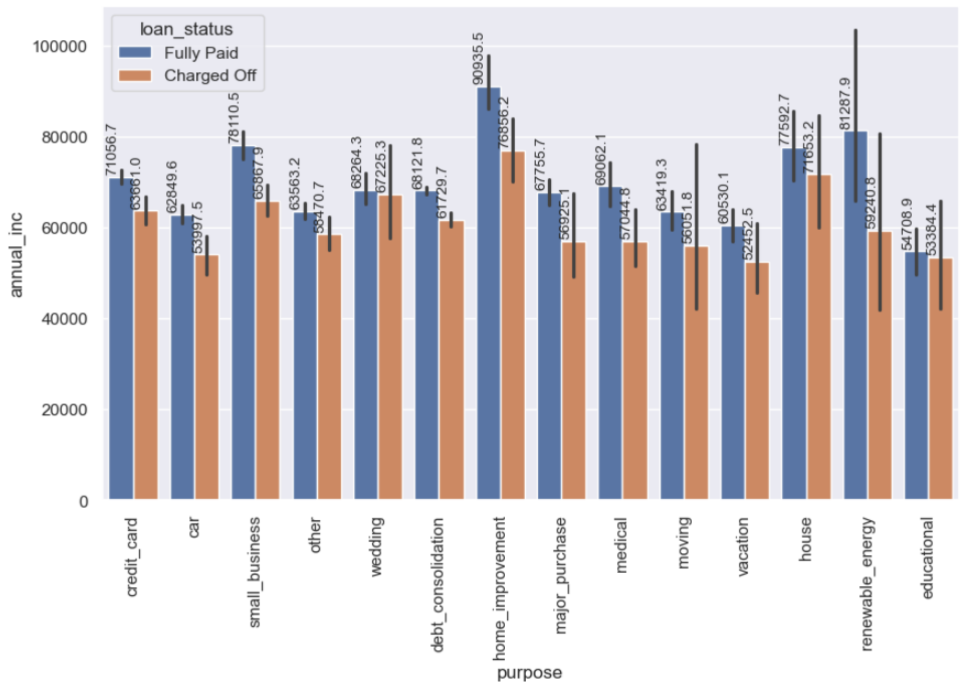


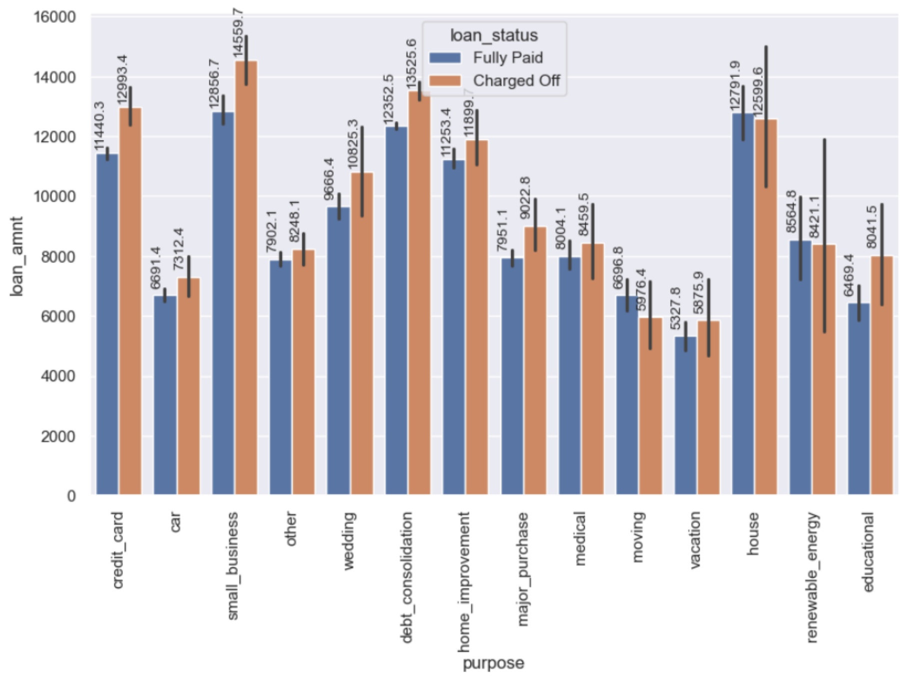
Key Findings

1. There is no significant difference in average loan amounts between two categories in most of the zip code groups
2. The loan amount distribution seems to be slightly same among all the zip codes. So it suggests loan amount distribution is not impacted by zip_code_groups
3. Applicants categorized as "Full paid" seems to have higher annual income when compared to "Charged off".
4. Only some zip code groups show higher Annual income. But, most of the Annual incomes seems to be same across all the zip_code_groups
5. Employees whose income is higher tend to pay the loans on time

Data Visualization and Analysis

Bivariate Analysis (on complete dataset)





Key Findings

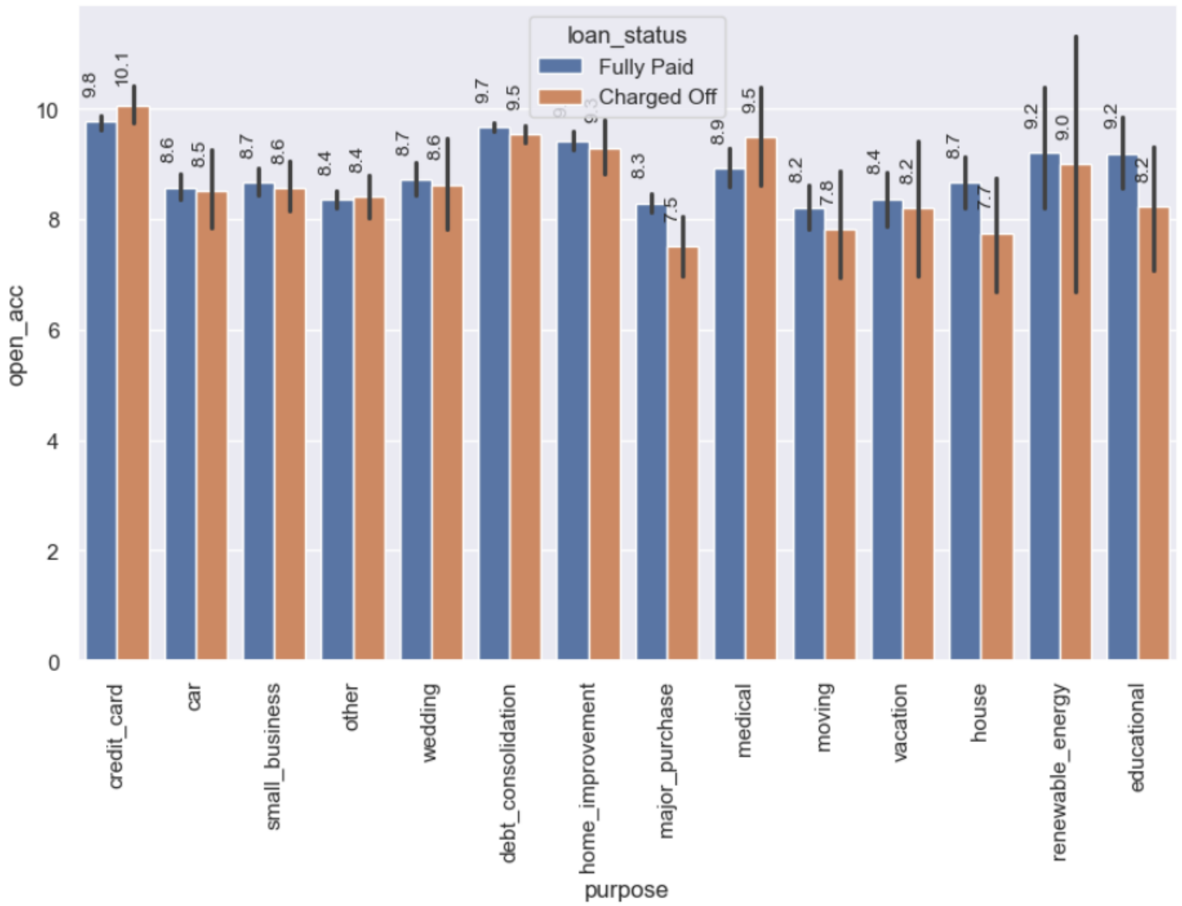
1. Higher Annual income correlates with higher chances of reducing the defaulters
2. Purpose like small_business and medical show the notable differences, where charged off loans are linked to income
3. Credit_card loans have highest charged off rate when compared to others
4. Vacation loans have lesser charged off rate.

Recommendation:

1. Stricter rules has to be implemented while approving the loans for credit_cards

Data Visualization and Analysis

Bivariate Analysis (on complete dataset)

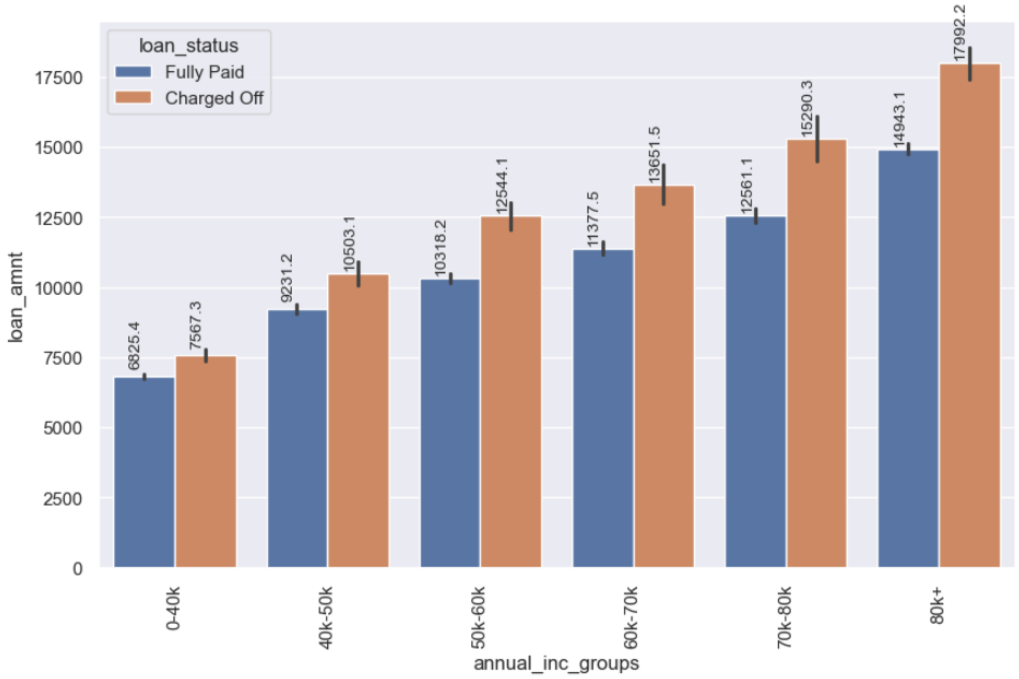


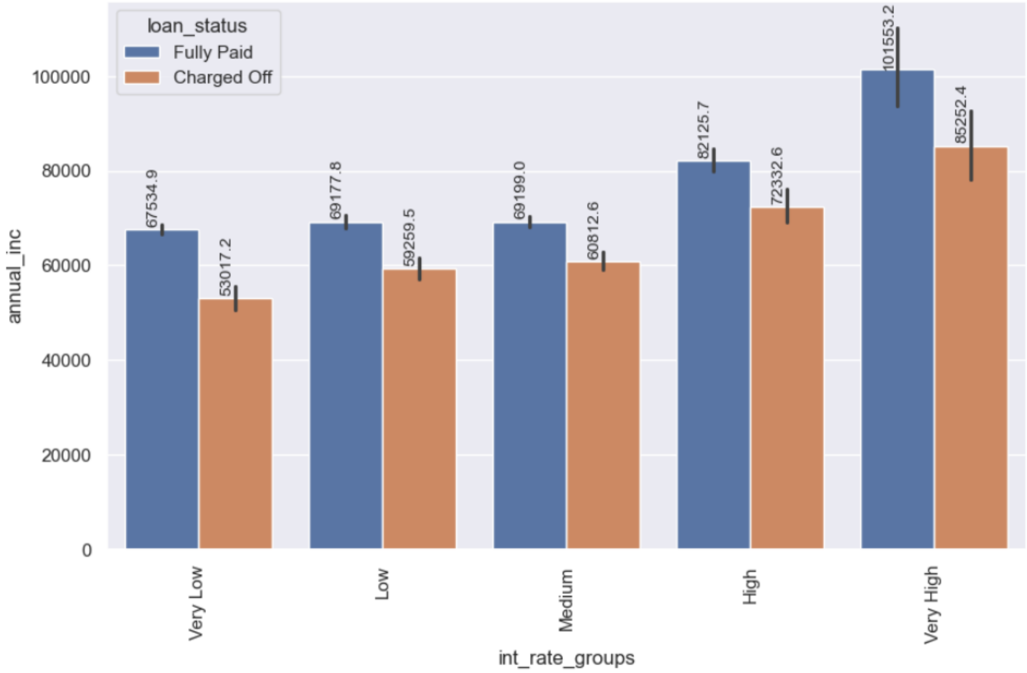
Key Findings

1. Majority portion of loans have higher "Charged off" loans across all segments and credit card stands higher among them again.

Data Visualization and Analysis

Bivariate Analysis (on complete dataset)





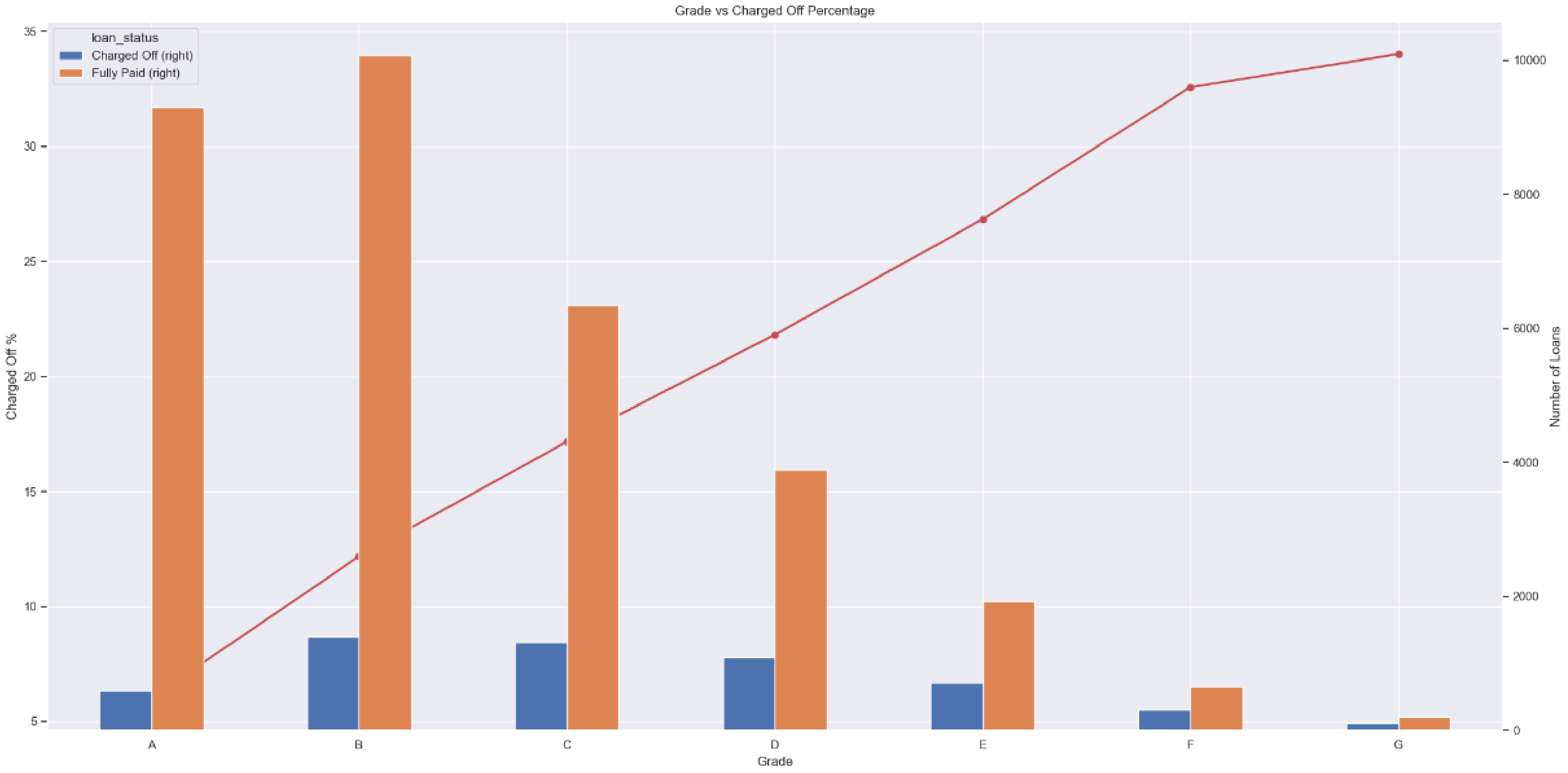
Key Findings

1. Plot shows us clearly that , "the higher the annual income, the higher the loan amount“, so is the risk of default
2. Applicants with higher annual income tend to qualify for loans with higher interests.
3. As the interest rate increases the proportion of "charged off" loans also increases

Multi-variate Analysis

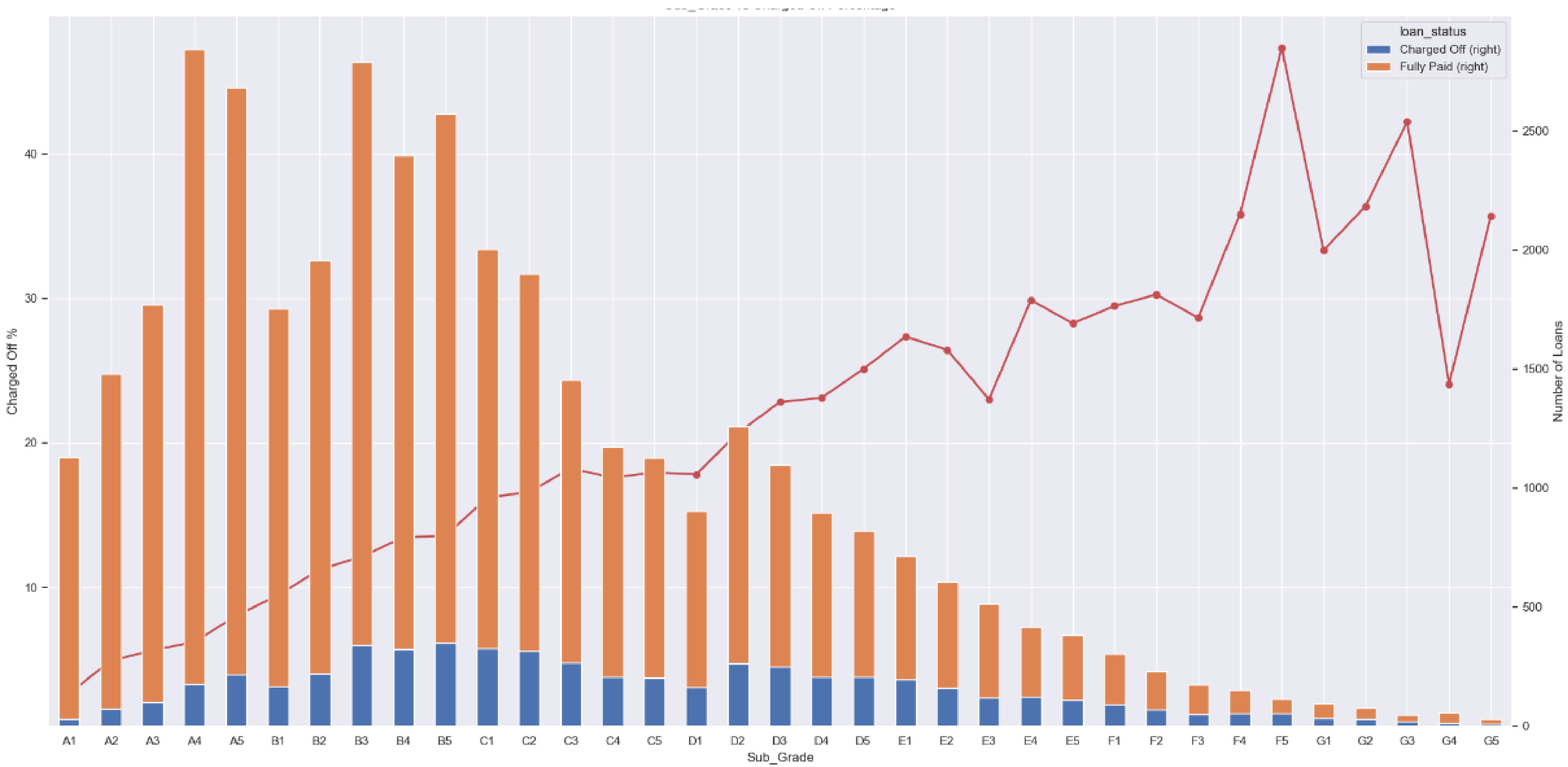
Data Visualization and Analysis

Multi-variate Analysis (on complete dataset)



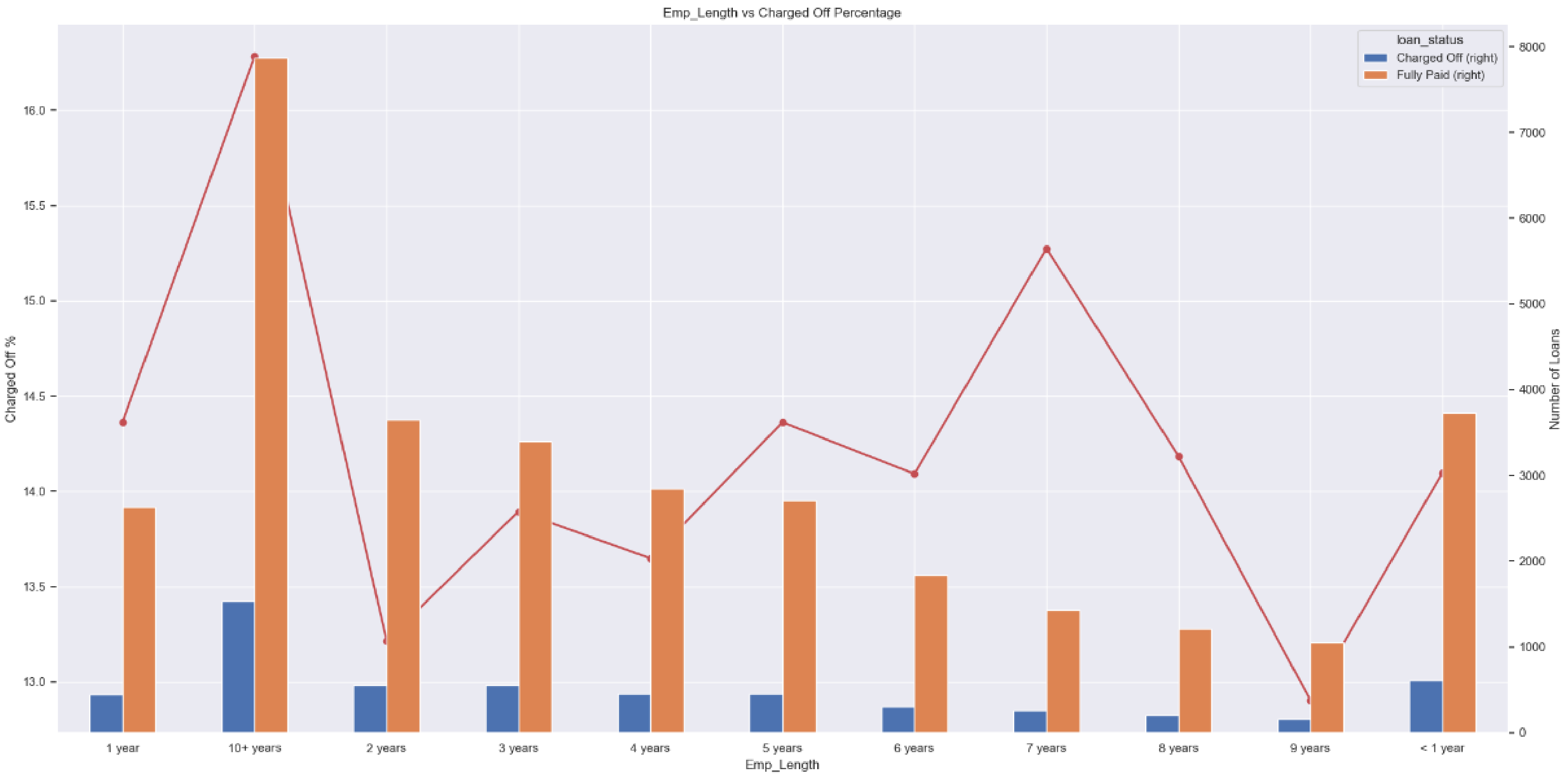
Data Visualization and Analysis

Multi-variate Analysis (on complete dataset)



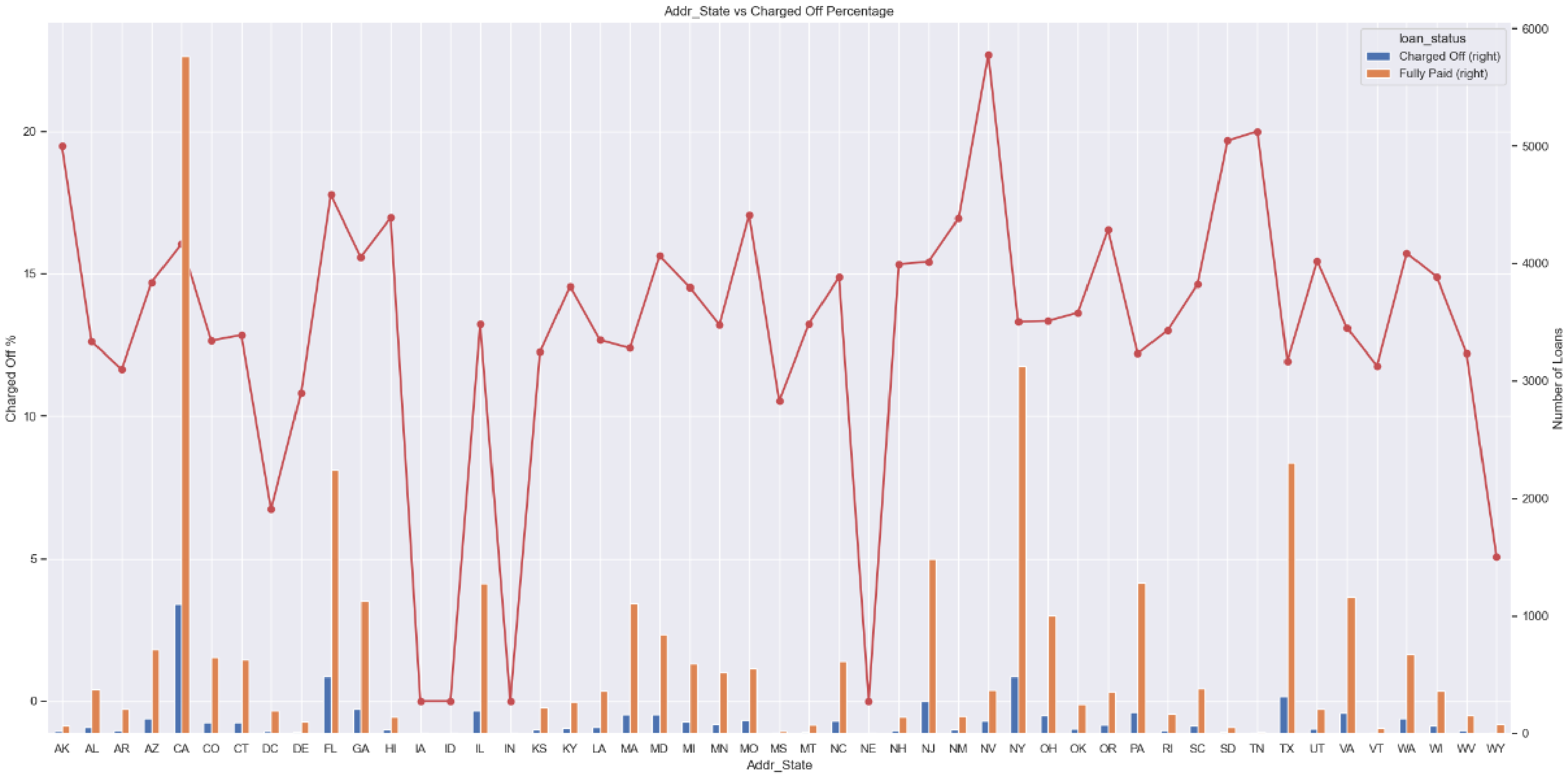
Data Visualization and Analysis

Multi-variate Analysis (on complete dataset)



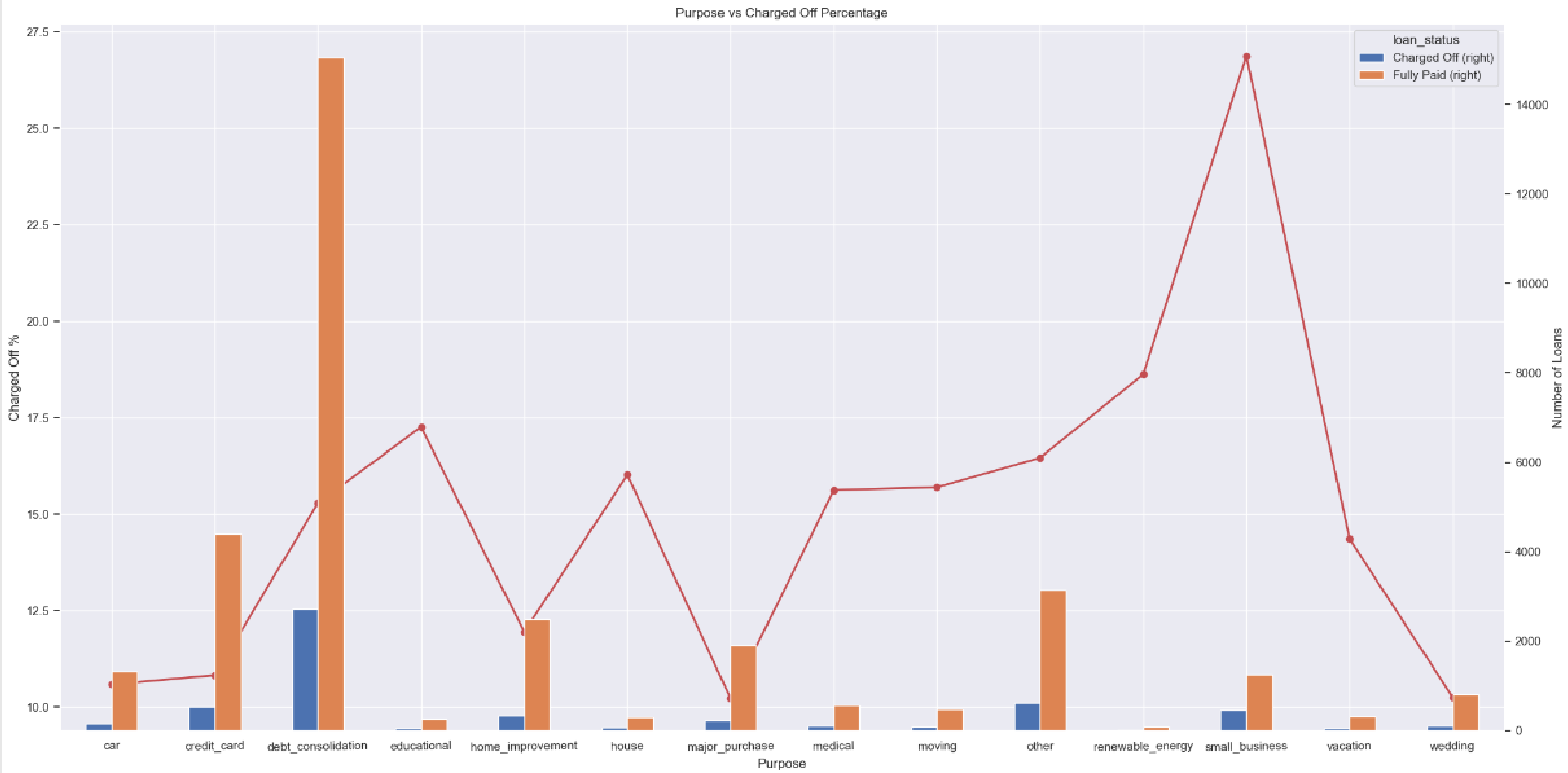
Data Visualization and Analysis

Multi-variate Analysis (on complete dataset)



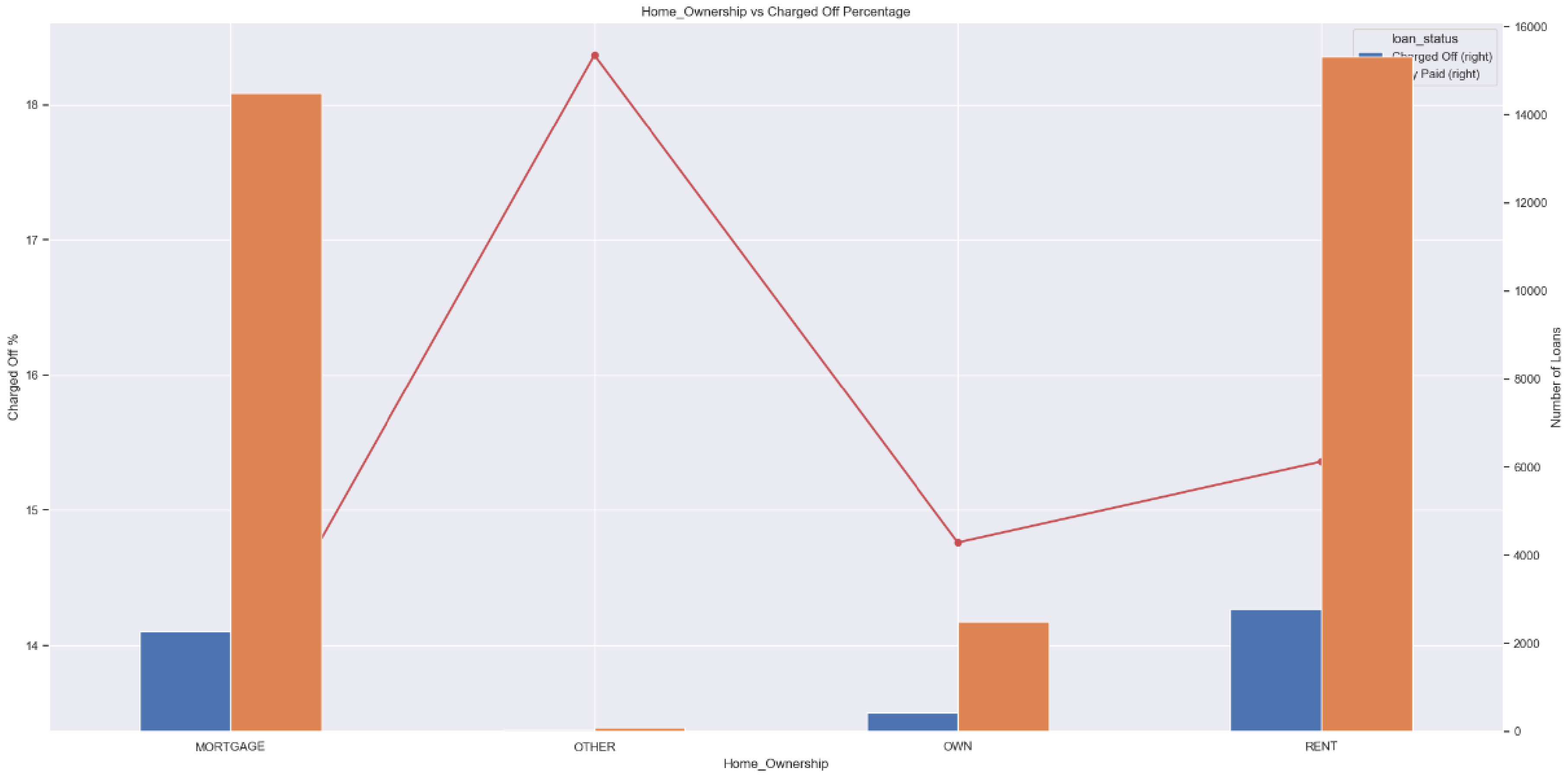
Data Visualization and Analysis

Multi-variate Analysis (on complete dataset)



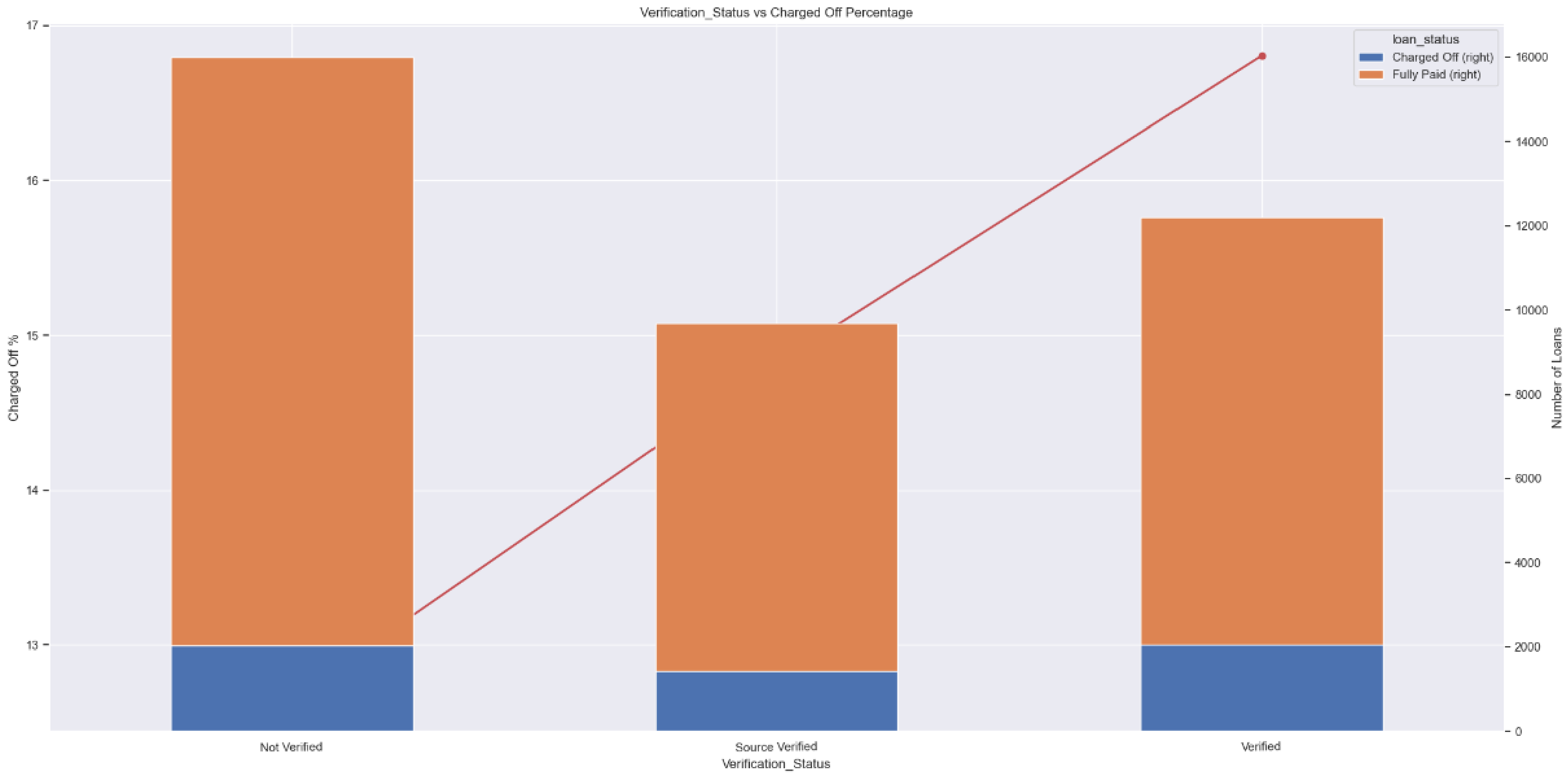
Data Visualization and Analysis

Multi-variate Analysis (on complete dataset)



Data Visualization and Analysis

Multi-variate Analysis (on complete dataset)



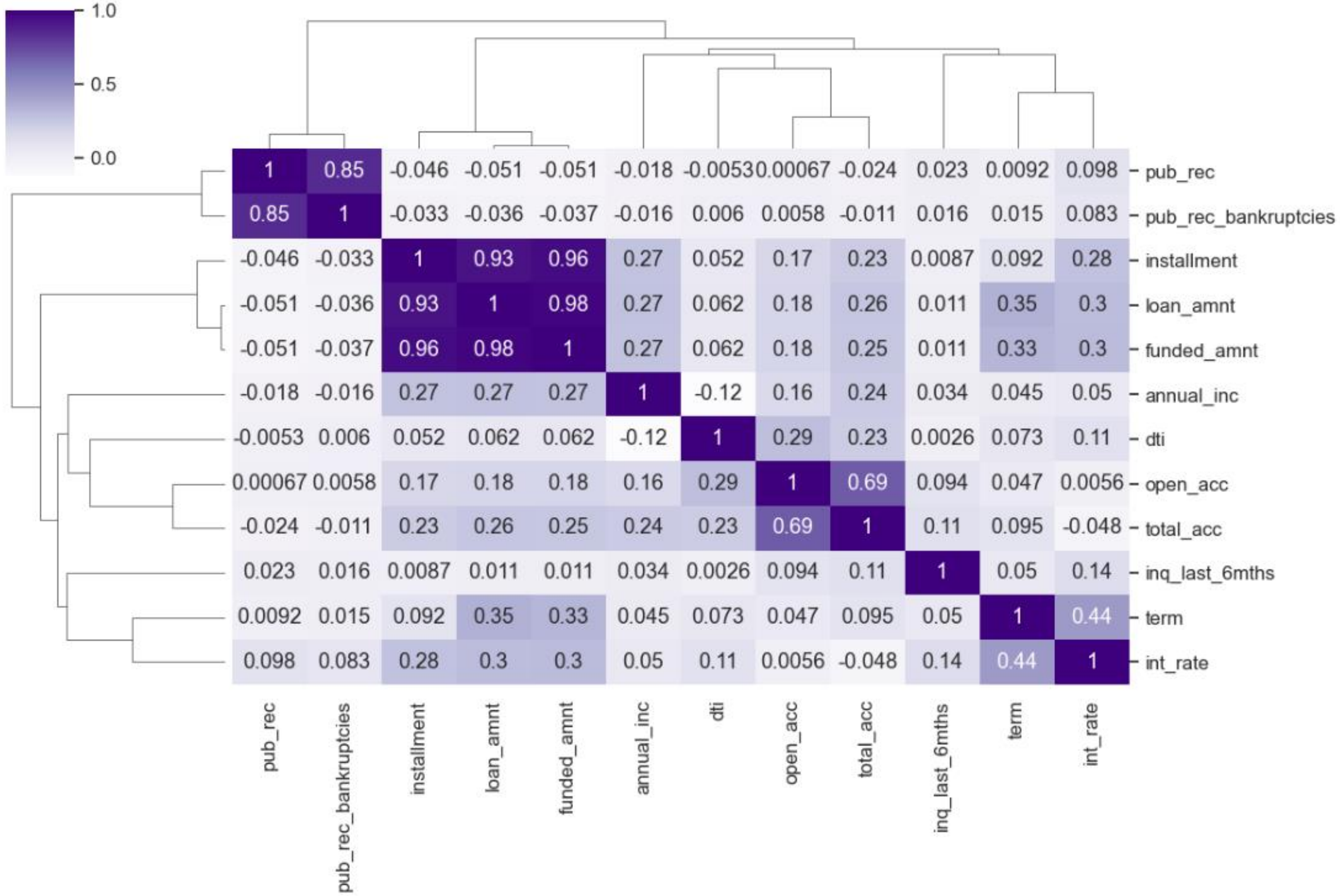
Data Visualization and Analysis

Multi-variate Analysis (on complete dataset) - Key Findings

1. Tendency of defaulters increase in the B, C, D grades
2. Subgrades B3, B4, B5 have high defaulters
3. Employees with 10+ years of service tend to more defaulters
4. Applicants from states 'CA’, 'FL' and 'NJ' have more number of defaulters.
5. Applicants for debt consolidation have high tendency to default loan.
6. Applicants from Rented House Ownership have highest tendency to default the loan.
7. Applicants with Verified and Not Verified status has high tendency to default.

Data Visualization and Analysis

Correlation Analysis (on complete dataset)



Strong correlation:

1. 'installment' has strong correlation with loan_amnt, funded_amnt
2. interest rate with term.
3. annual income with loan amount

Negative correlation:

1. pub_rec_bankruptcies and loan amount
2. annual_inc and dti

Data Visualization and Analysis

Summary of key findings

Attributes	Findings
Default rate	14%
Grade and Sub-grade	▪ 77% of accounts are from B (B3, B4, B5), C and D ▪ B grade contributes to ~30%
Term	Based on the count, 36-month loans have a higher number of defaulters, whereas 60-month loans tend to have higher loan amounts among defaulters.
Employee Length	10+ years appear to be contributing to highest charged off percent
Open accounts	▪ 54% of defaulted accounts have open account between 5 to 10 ▪ Open accounts less than 5 and more than 15 shows less defaulting tendency
State and Zip code	▪ California has the highest number applicants (17.7%) and defaulters (20%). ▪ Applicants from states CA, FL, NY, TX, NJ contributes ~48% of defaulters ▪ Zip Code starts 9xxxxx, 3xxxxx, 1xxxxx contributes ~50% of defaulters
Home Ownership	▪ Loan applicants with rented house contributing to 47.8%. ▪ The highest defaulters are from rented house category 50.5%. ▪ Applicants under RENT have the lowest annual income.
Purpose	▪ debt_consolidation and credit cards (~60%) contributes to highest purpose. ▪ defaulters are also high under the debt consolidation (49.2%). ▪ small_business and medical show the notable differences, where charged off loans are linked to income
Verification status	▪ Applicants who are not verified contribute to 42.2%. ▪ Applicants with Not Verified status have highest proportion (38.1%) with charged off loans. ▪ Defaulters % for Source verified (25%) is lesser than verified and non-verified (74%).
DTI	▪ ~38% of defaulted loan applicants had DTI with 12%. 30.6% of defaulted applicants is between 12 to 18% ▪ Among loan participants who charged off, most of the loan applications have the high DTI ratio.
Loan amount	~72% of defaulted borrowers have loan amount till 15K.
Interest rate	▪ 67% defaulted loan applicants had interest rate between Low and Medium groups ▪ It is also observed that where interest rate is medium are having more defaulters.
Annual Income	▪ ~62% of defaulted loan applicants had income till 60K. ▪ Higher Annual income correlates with higher chances of reducing the defaulters ▪ Fully paid applicants have higher annual income when compared to charged off employees.
Loan processing volume	Most of the loans are issued in the second half of the year. December month is the highest month (10.9%)

Back to Executive Summary

Recommendation and Conclusion

Recommendation

1. For the loan purpose for debt consolidation and credit card, implement stricter verification process.
2. Prefer Source verification
3. If the loan tenure is less then verify the DTI ratio.
4. Own house status can be considered to lower the risk of default
5. Higher annual income or high employee length does not guarantee of full payment of loan. Additionally consider income stability such as DTI, average current balance of all accounts and other indicators while analyzing purpose
6. In case if a borrower’s DTI is high and eligibility is 36 month, move the borrower to 60 month term to reduce the installment amount

Conclusion

1. While lending platforms, typically experiences default rates in the range of 15% to 30% for bad debt, any bad debt is credit loss for lending business. Hence every attempt should be made to curtail it down.
2. Default rates depends on creditworthiness of borrowers and the economic environment, hence constant efforts has to be ensured to fine tune the policy
3. Exploratory Data Analysis is an ongoing effort. We should continuously implement recommendations and revisit EDA to ensure the best insights and improvements are derived

Back to Executive Summary

Thank you!