EDA - Lending Club Case Study

Group Members:

Abhishek Das

Ashutosh Kulkarni (Group Facilitator)

Case Study

- Perform the EDA on the data to identify the driving factors / variable that results in the loan defaulters.
- Loan Data Set - Loan data for all loans issued through the time period 2007 t0 2011.

    -Loans Accepted

        - Fully paid Loans
        - Current ongoing Loans
        - Defaulted / Charged off loans

    - Loans Rejected

Objective – Using EDA identify how **consumer attributes** and **loan attributes** influence the tendency of default.

EDA Step 1: Data Sourcing

- Read data from ‘loan.csv’ file

```
# 1.1 Read the loan csv file.
loan_data = pd.read_csv('loan.csv', encoding = 'ISO-8859-1', low_memory=False)

# 1.2 Review data
loan_data.head()
```

id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership
1077501	1296599	5000	5000	4975.0000	36 months	10.65%	162.8700	B	B2	NaN	10+ years	RENT
1077430	1314167	2500	2500	2500.0000	60 months	15.27%	59.8300	C	C4	Ryder	< 1 year	RENT
1077175	1313524	2400	2400	2400.0000	36 months	15.96%	84.3300	C	C5	NaN	10+ years	RENT

EDA Step 2: Data Cleaning

- Initial Data frame Shape

```
# 2.1 Identify the initial DataFrame Shape
loan_data.shape
```

```
(39717, 111)
```

- Steps for data cleaning:
  - Fix missing values
  - Fix rows and columns
  - Standardise the values

Data Cleaning: Fix Missing Values

- Identify the NULL or NaN values percentage for each column

```
# 2.2 Identify the NULL or NaN values Percentage for each column
missing_loan_data = round(100*(loan_data.isnull().sum()/len(loan_data.id)), 2)
missing_loan_data.loc[missing_loan_data > 0]
```

Observations :

- Many columns has 100% NULL or NaN values
- Other columns has percentages varying between 2% to 97% NULL or NaN values

- After reviewing all the columns, all the columns which has NULL values percentage above and equal to 30% are dropped.

Data Cleaning: Fix Missing Values

- Review and handling for remaining columns with missing values

```
emp_title                    6.1900
emp_length                   2.7100
title                        0.0300
revol_util                   0.1300
last_pymnt_d                 0.1800
last_credit_pull_d           0.0100
collections_12_mths_ex_med   0.1400
chargeoff_within_12_mths     0.1400
pub_rec_bankruptcies         1.7500
tax_liens                    0.1000
dtype: float64
```

Observations :

- Columns `'emp_title'`, `'emp_length'` and `'title'` has very low missing values percentage
- Customer behaviour variables `'revol_util'`, `'last_pymnt_d'`, `'collections_12_mths_ex_med'`, `'chargeoff_within_12_mths'` and `'tax_liens'` are not available at the time of loan application
- Column `'pub_rec_bankruptcies'` has very low missing value percentage. The column values are 0, 1, 2, and nan only.

- For columns `'emp_title'`, `'emp_length'`, `'title'`, and `'pub_rec_bankruptcies'` the missing values rows are removed
- Columns variables `'revol_util'`, `'last_pymnt_d'`, `'collections_12_mths_ex_med'`, `'chargeoff_within_12_mths'` and `'tax_liens'` are dropped

Data Cleaning: Fix Missing Values

- Review if any remaining columns for missing values

```
# Review the remaining columns for missing values
missing_loan_data = round(100*(loan_data.isnull().sum()/len(loan_data.id)), 2)
missing_loan_data[missing_loan_data != 0]
```

```
Series([], dtype: float64)
```

- No columns has missing values.
- Final data frame shape after data cleaning

```
# Review the final dataFrame shape
loan_data.shape
```

```
(36539, 48)
```

Data Cleaning: Fix Rows and Columns

- Drop any insignificant columns that do not add value for the analysis
- Drop columns with either single unique value or all unique values
- Below columns are dropped

```
'id',                   'member_id',                'pymnt_plan'
'url'                   'zip_code'                  'initial_list_status',
'policy_code',          'application_type',         'acc_now_delinq'
'delinq_amnt',          'delinq_2yrs',              'earliest_cr_line',
'inq_last_6mths',       'open_acc',                 'pub_rec',
'revol_bal',            'total_acc',                'out_prncp',
'out_prncp_inv',        'total_pymnt',              'total_pymnt_inv',
'total_rec_prncp'       ,'total_rec_int',           'total_rec_late_fee',
'recoveries',           'collection_recovery_fee',  'last_pymnt_amnt',
'last_credit_pull_d'
```

Data Cleaning: Fix Rows and Columns

- Final data frame shape after data cleaning

```
# Review the final dataFrame shape
loan_data.shape
```

```
(36539, 20)
```

Data Cleaning: Standardise Values

- `'issue_d'` columns type changed to datetime64
- `'emp_length'` column values updated to have only numeric values and renamed to `'emp_length_years'`
- `'int_rate'` column type changed to float64 and removed '%' symbol
- `'int_rate'` renamed to `'int_rate_percent'`
- Split the column `'issue_d'` into month and year column as `'issue_d_month'` and `'issue_d_year'` columns

Before Standardising

```
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   issue_d     36539 non-null  object
 1   emp_length  36539 non-null  object
 2   int_rate    36539 non-null  object
 3   issue_d     36539 non-null  object
dtypes: object(4)
```

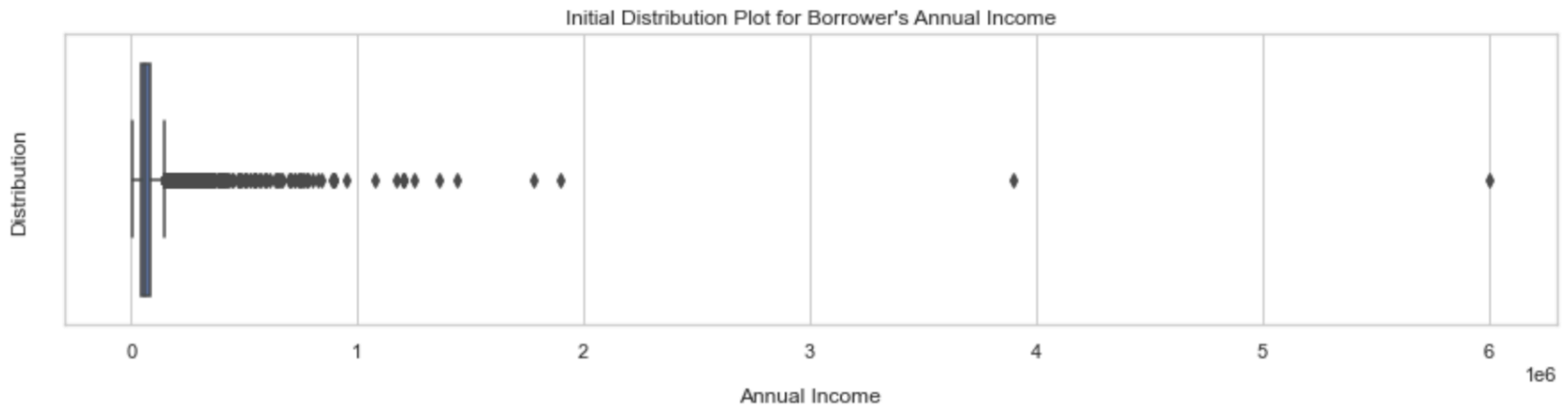
After Standardising

```
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   issue_d           36539 non-null  datetime64[ns]
 1   emp_length_years  36539 non-null  int64
 2   int_rate_percent  36539 non-null  float64
 3   issue_d_month     36539 non-null  int64
 4   issue_d_year      36539 non-null  int64
dtypes: datetime64[ns](1), float64(1), int64(3)
```

Data Cleaning: Standardise Values

- Outliers identification for column `'annual_inc'`

```
count      36539.0000
mean       69240.4414
std        63509.7472
min         4000.0000
25%        41902.0000
50%        60000.0000
75%        83000.0000
max      6000000.0000
Name: annual_inc, dtype: float64
```



Observations:

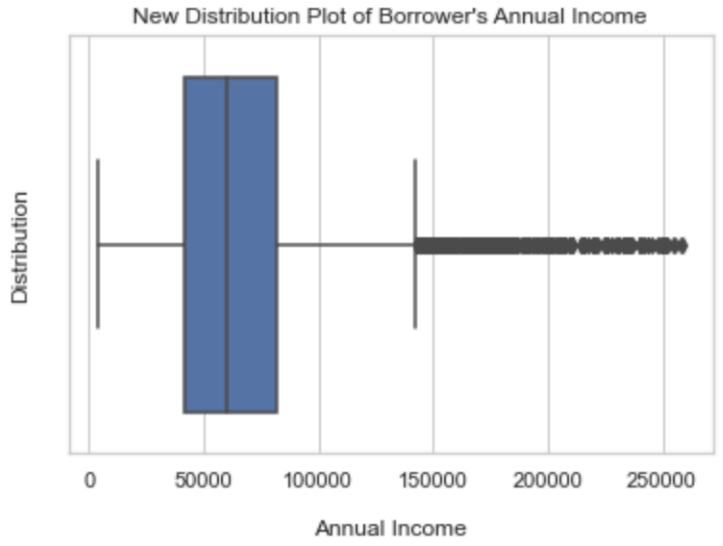
- The annual income reported by the borrowers range from min of 4,000 to max of 6,000,000.
- The median annual income is around 60,000
- Most people have an annual income less than 115000
- The above box plot shows that there is a high amount of outliers present.

- Conclusion: Keep the annual income that is within +3 to -3 standard deviations to get rid of outliers (Nelson rule).

Data Cleaning: Standardise Values

- Outliers removal for column `'annual_inc'`

```
count     36278.0000
mean      66376.4811
std       35725.9562
min        4000.0000
25%       41500.0000
50%       60000.0000
75%       82000.0000
max      259000.0000
Name: annual_inc, dtype: float64
```



Observations:

- The max annual income reduced to 259,000 .
- The median annual income remains same at 60,000

```
# Review the final dataFrame shape after removing outliers
loan_data.shape
```

```
(36278, 22)
```

Variable Analysis

- Key variable for analysis `'loan_status'`

```
## Key column for analysis - 'loan_status'
loan_data.loan_status.value_counts()
```

```
Fully Paid     30228
Charged Off     4993
Current         1057
Name: loan_status, dtype: int64
```

Observations:

- Loans with status as ‘Current’ will not help for the analysis
- Only loans with status as ‘Fully Paid’ and ‘Charged Off’ are considered for analysis.

```
# Review the dataframe shape after removing the current loan status rows

loan_data.shape
```

```
(35221, 22)
```

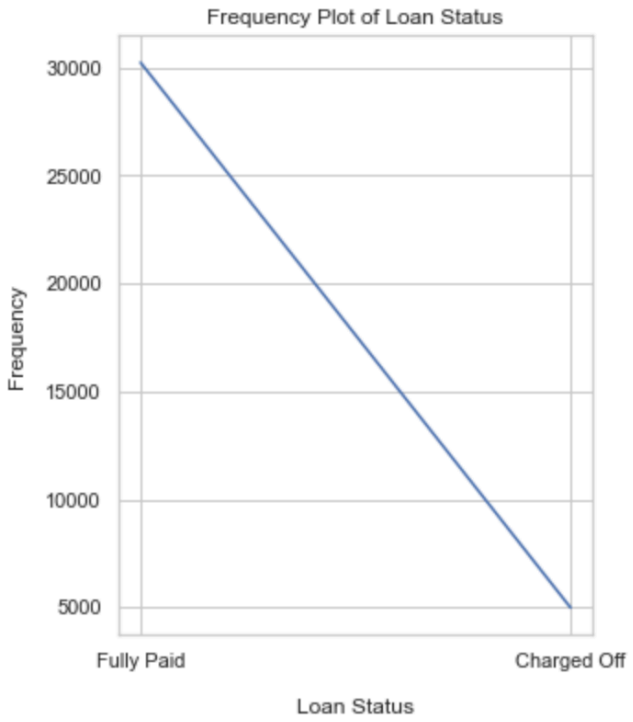
EDA Step 3: Univariate Analysis: Unordered Categorical Variables

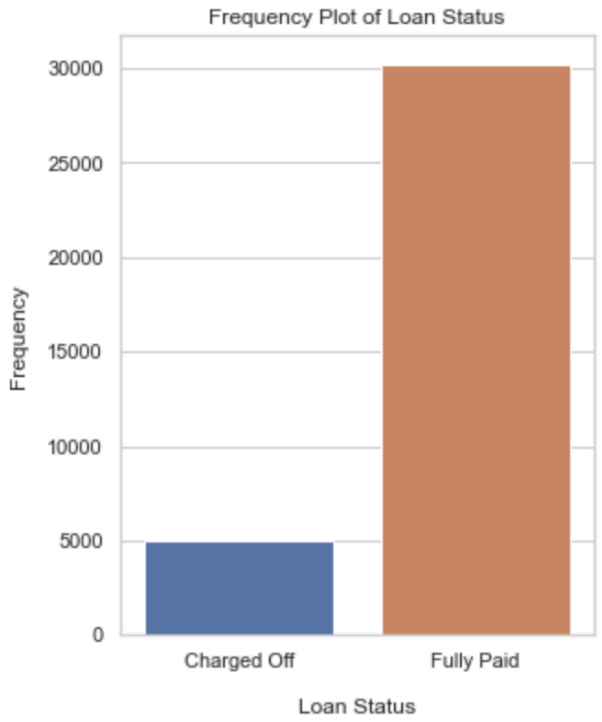
- Unordered Categorical Variables are:
  - `'loan_status'`
  - `'home_ownership'`
  - `'purpose'`
  - `'addr_state'`

Unordered Categorical Variable: `loan_status`

- Rank-Frequency Plot of Unordered Categorical Variable: `'loan_status'`

```
Charged Off     4993
Fully Paid     30228
Name: loan_status, dtype: int64
```



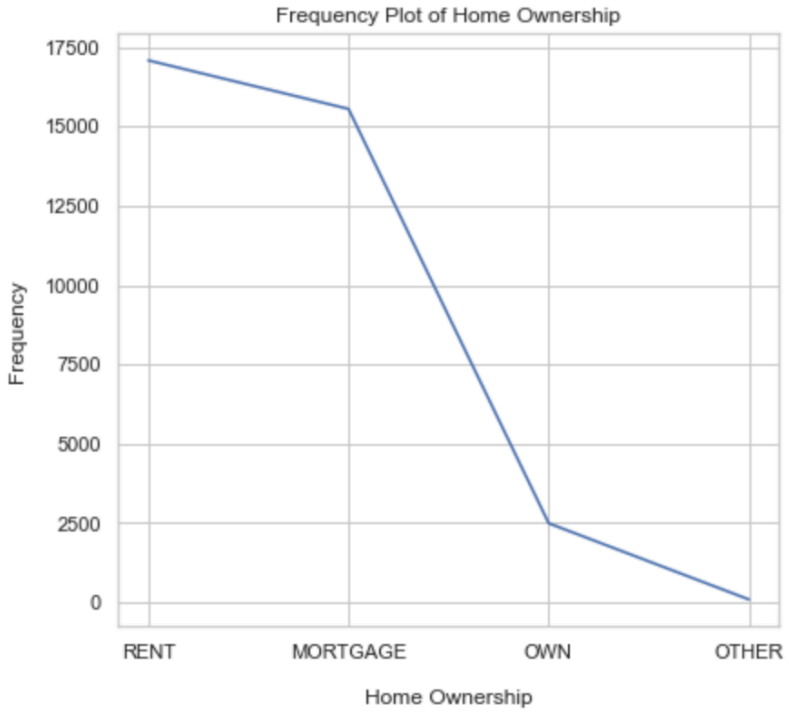


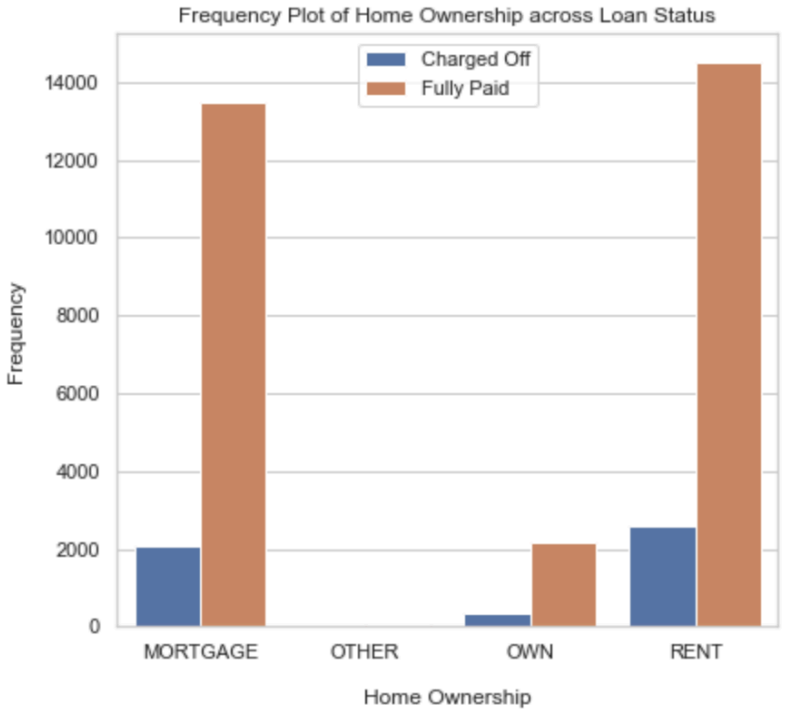
Observation:

- Defaulted Loan Percentage 14.18 %

Unordered Categorical Variable: `home_ownership`

- Rank-Frequency Plot of Unordered Categorical Variable: `'home_ownership'`






```
MORTGAGE     15551
OTHER           95
OWN           2495
RENT         17080
Name: home_ownership, dtype: int64
```

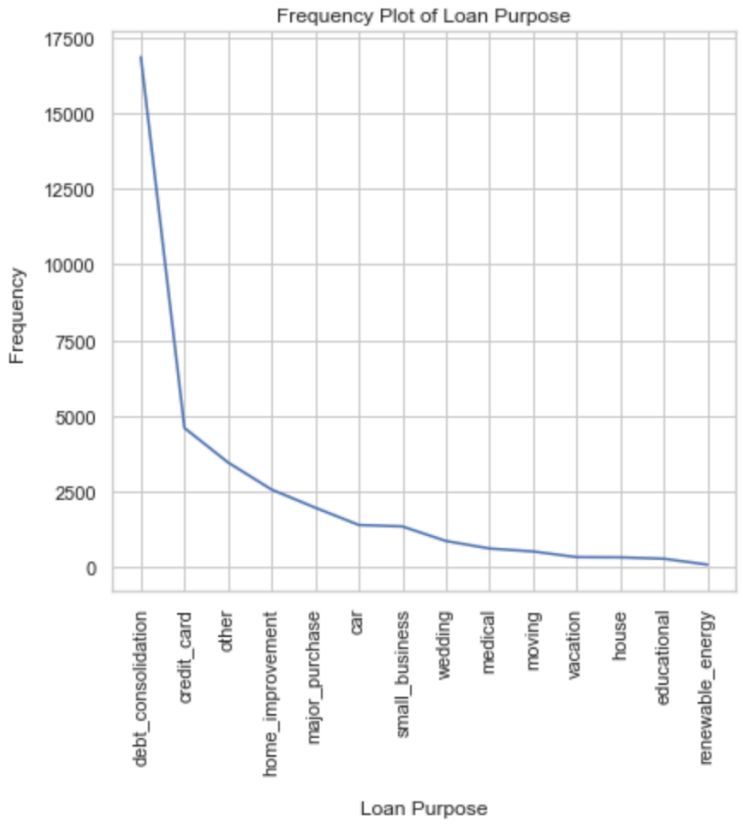
```
loan_status  home_ownership
Charged Off  MORTGAGE             2064
             OTHER                 18
             OWN                  349
             RENT                2562
Fully Paid   MORTGAGE            13487
             OTHER                 77
             OWN                 2146
             RENT               14518
Name: home_ownership, dtype: int64
```

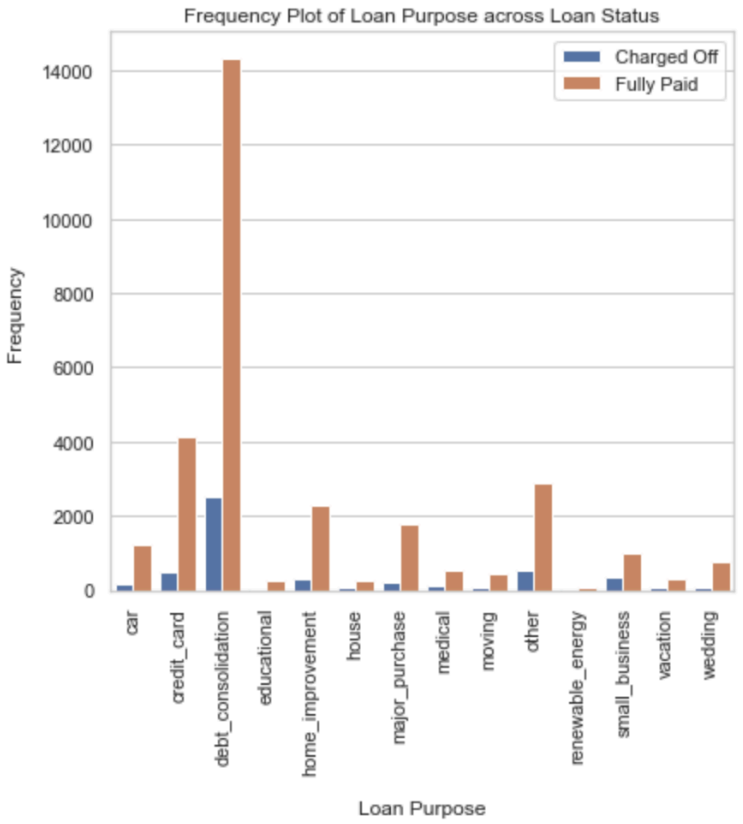
Observation:

- The initial analysis of `'home_ownership'` variable shows that the frequency of charged off loans is **high** for **RENT** and **MORTGAGE** home ownership types.

Unordered Categorical Variable: `purpose`

- Rank-Frequency Plot of Unordered Categorical Variable: `'purpose'`



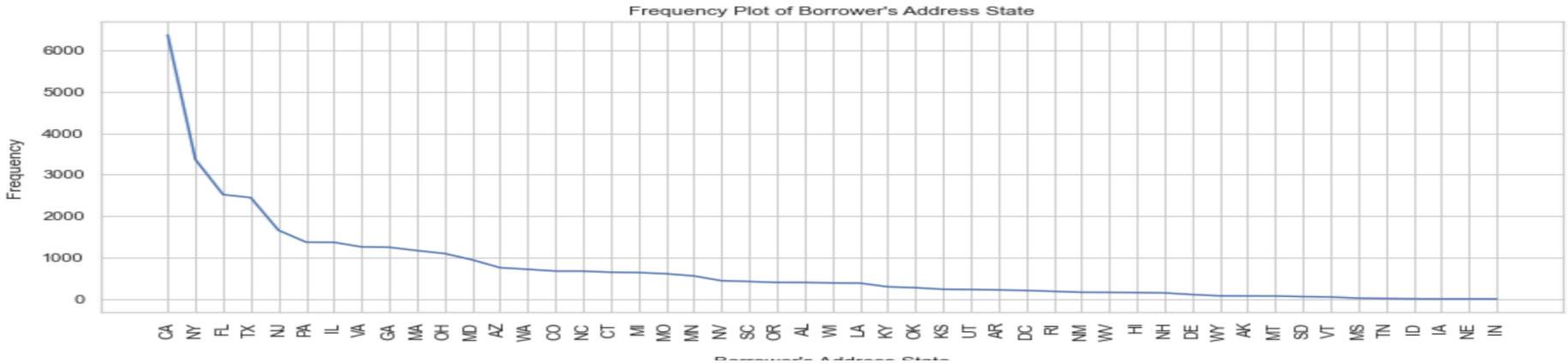


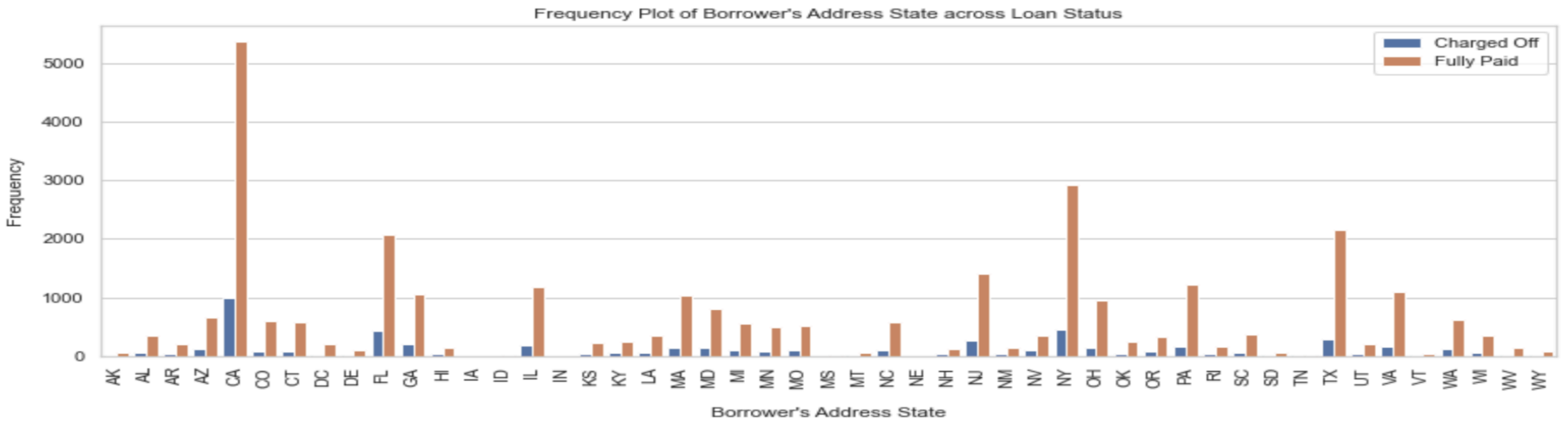
Observation:

- The analysis of the '`purpose`' variable shows that the highest loans are defaulted for the purpose of '`debt_consolidation`' among the other purpose categories.

Unordered Categorical Variable: `addr_state`

- Rank-Frequency Plot of Unordered Categorical Variable: `'addr_state'`





Observation:

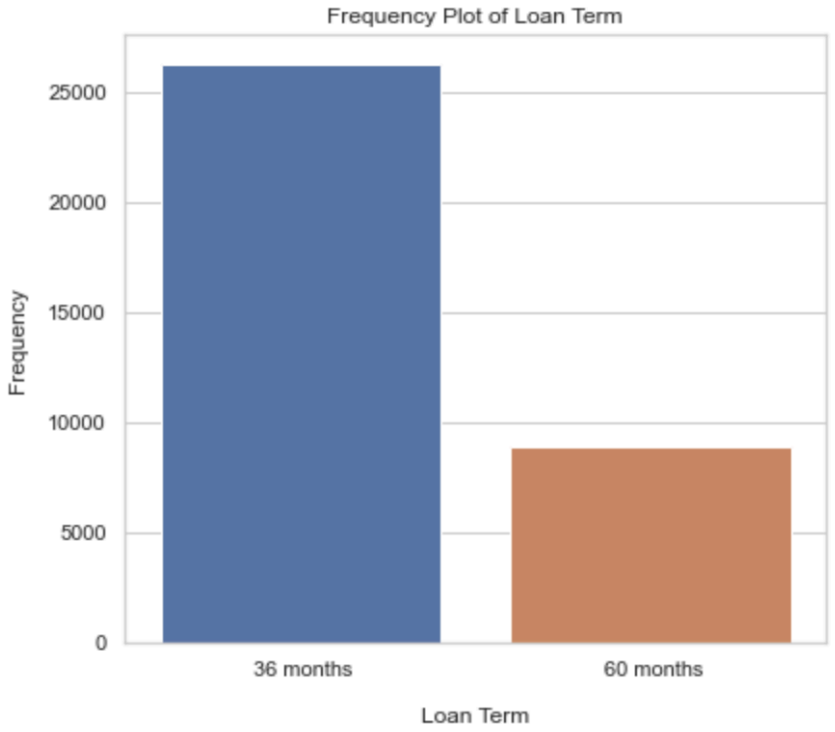
- The analysis of the `'addr_state'` variable shows that the highest loan defaulted are from states `CA`, `NY` and `FL`

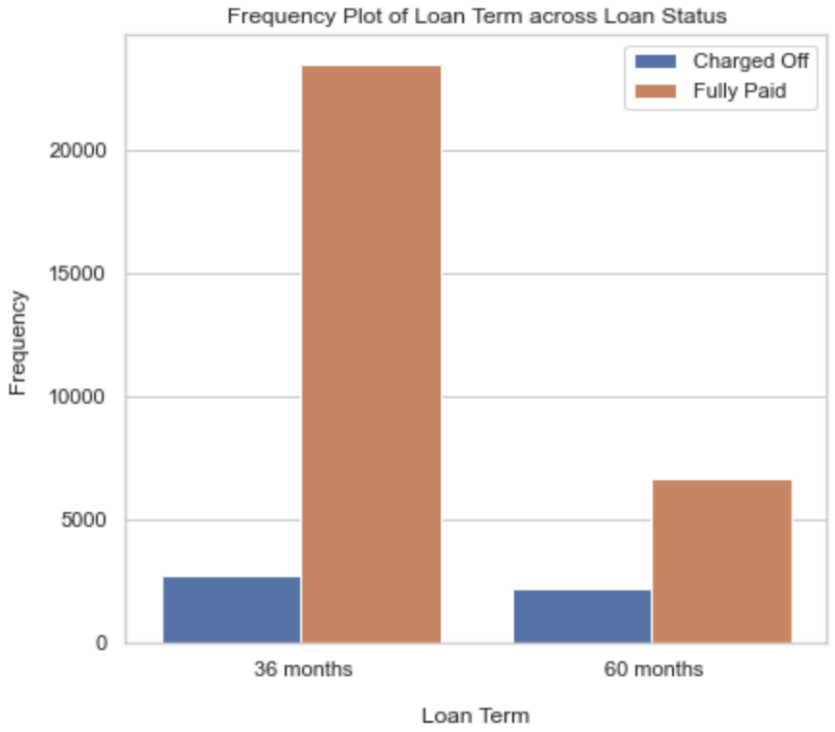
Univariate Analysis: Ordered Categorical Variables

- Ordered Categorical Variables are:
  - `'term'`
  - `'grade'`
  - `'sub_grade'`
  - `'emp_length'`
  - `'issue_yr'`

Ordered Categorical Variable: `term`

- Rank-Frequency Plot of Unordered Categorical Variable: `'term'`





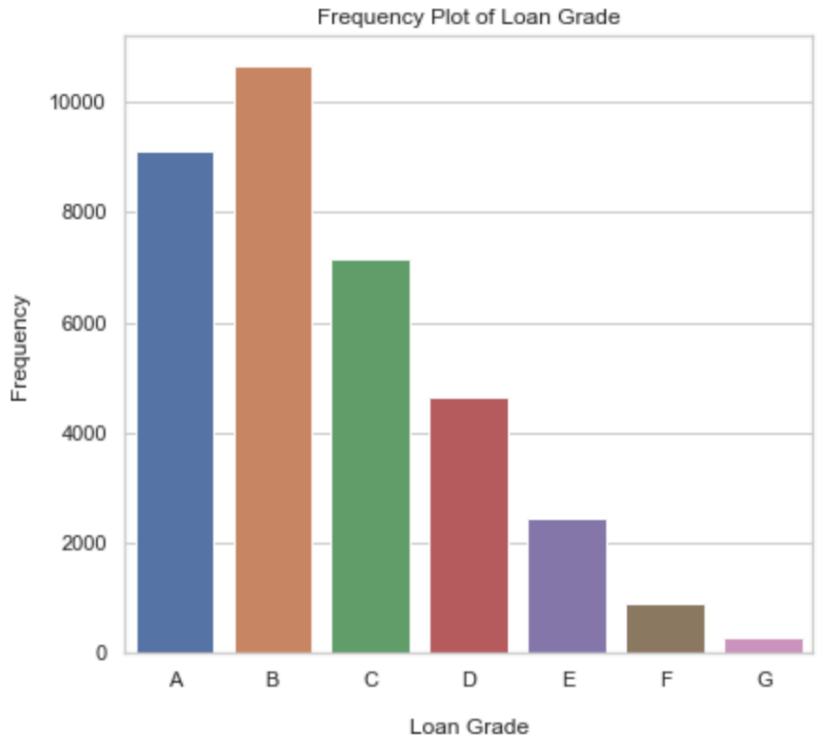

```
 36 months    26290
 60 months     8931
Name: term, dtype: int64


loan_status  term
Charged Off    36 months     2774
               60 months     2219
Fully Paid     36 months    23516
               60 months     6712
Name: term, dtype: int64
```

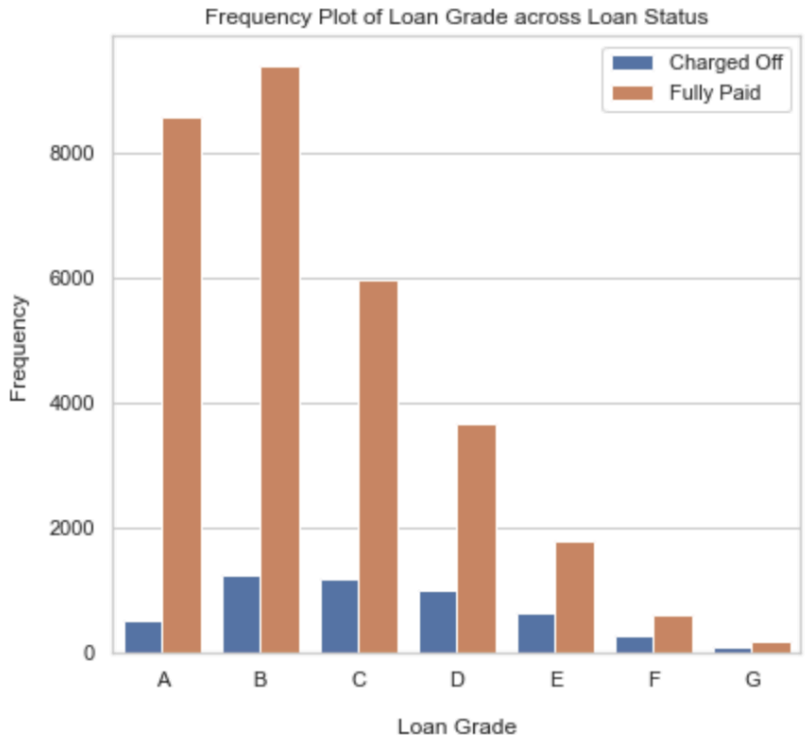
Observation:

- The analysis of the '`term`' variable shows that the highest number of loan defaulters are for 36 months term.

Ordered Categorical Variable: `grade`

- Rank-Frequency Plot of Unordered Categorical Variable: `'grade'`





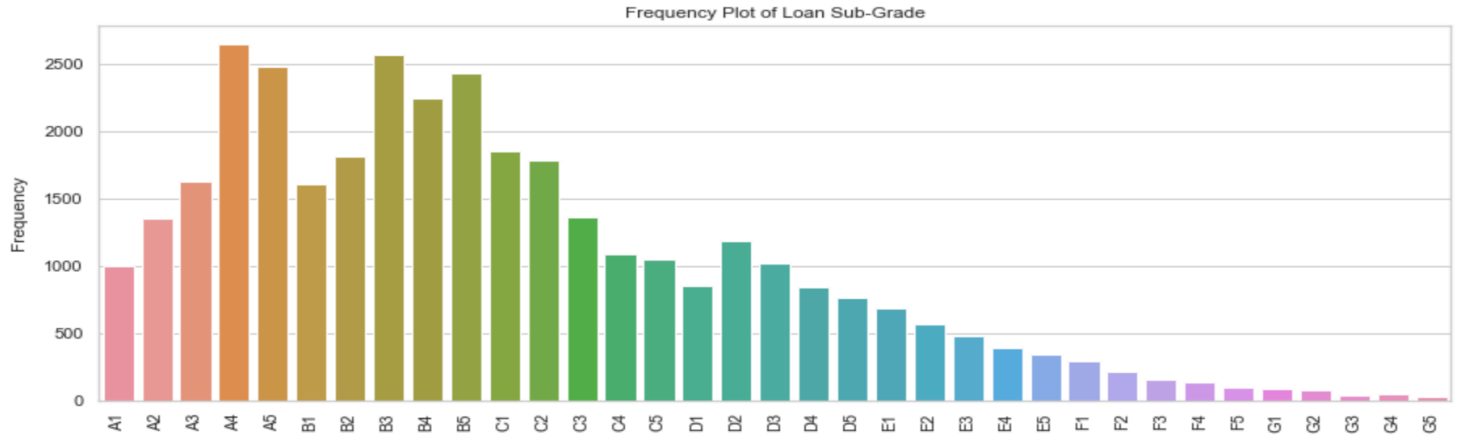

```
loan_status  grade
Charged Off  A          519
             B         1261
             C         1188
             D          994
             E          651
             F          285
             G           95
Fully Paid   A         8591
             B         9405
             C         5969
             D         3664
             E         1809
             F          608
             G          182
Name: grade, dtype: int64
```

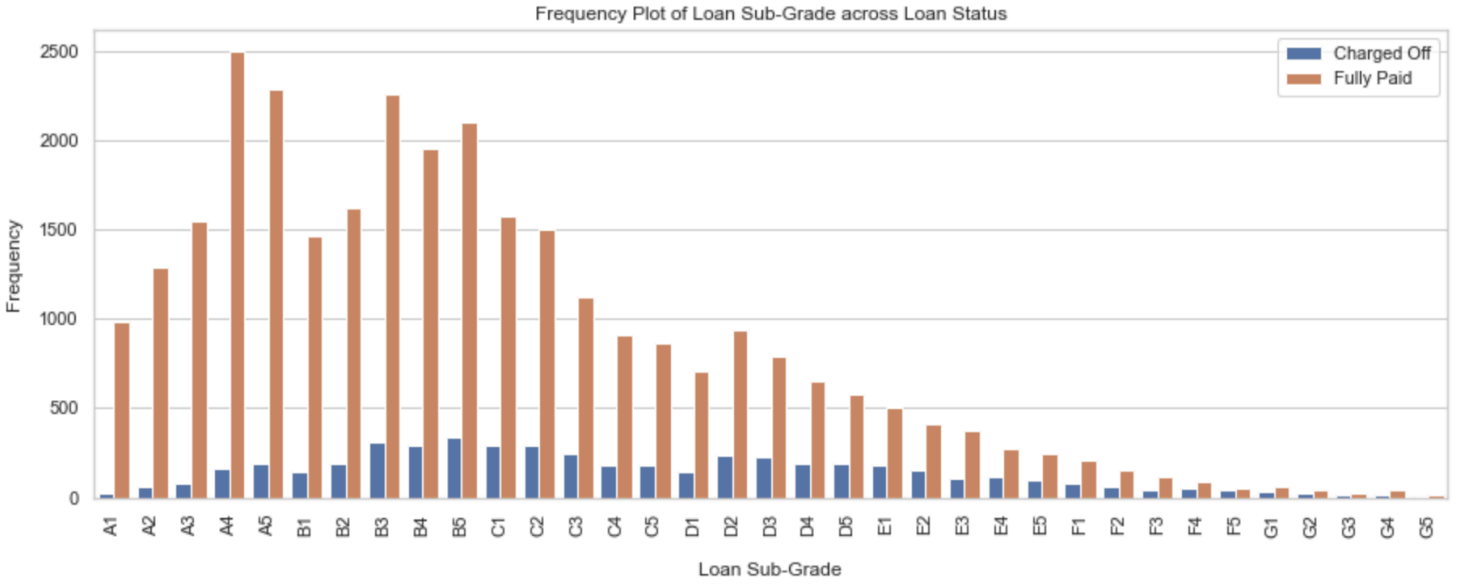
Observation:

- The analysis of the `'grade'` variable shows that the most of the loans charged off for the grades `'B'`,`'C'` and `'D'`

Ordered Categorical Variable: `sub_grade`

- Rank-Frequency Plot of Unordered Categorical Variable: `'sub_grade'`



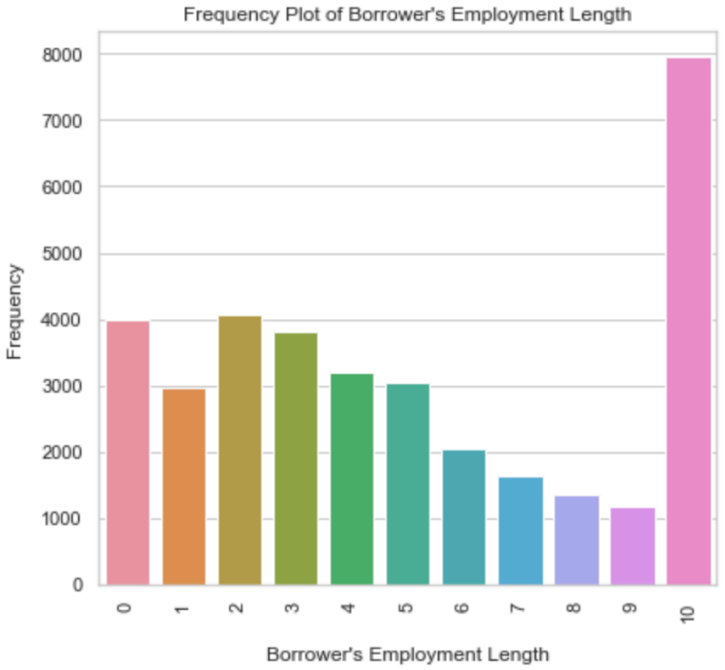


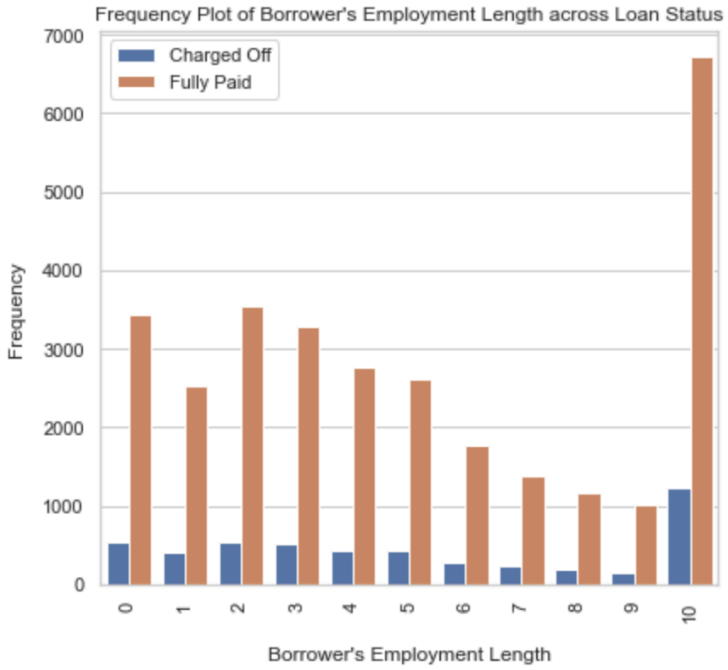
Observation:

- The distribution shows that, of the Grade B, C and D, highest Charged off applicants belong to Sub Grades:
- `Grade B => B3, B4, B5`
- `Grade C => C1, C2, C3`
- `Grade D => D2, D3, D5`

Ordered Categorical Variable: `emp_length_years`

- Rank-Frequency Plot of Unordered Categorical Variable: `'sub_grade'`





loan_status	emp_length_years	
Charged Off	0	540
	1	419
	2	534
	3	525
	4	432
	5	430
	6	288
	7	248
	8	195
	9	150
	10	1232
Fully Paid	0	3438
	1	2536
	2	3537
	3	3293
	4	2770
	5	2608
	6	1766
	7	1382
	8	1161
	9	1023
	10	6714

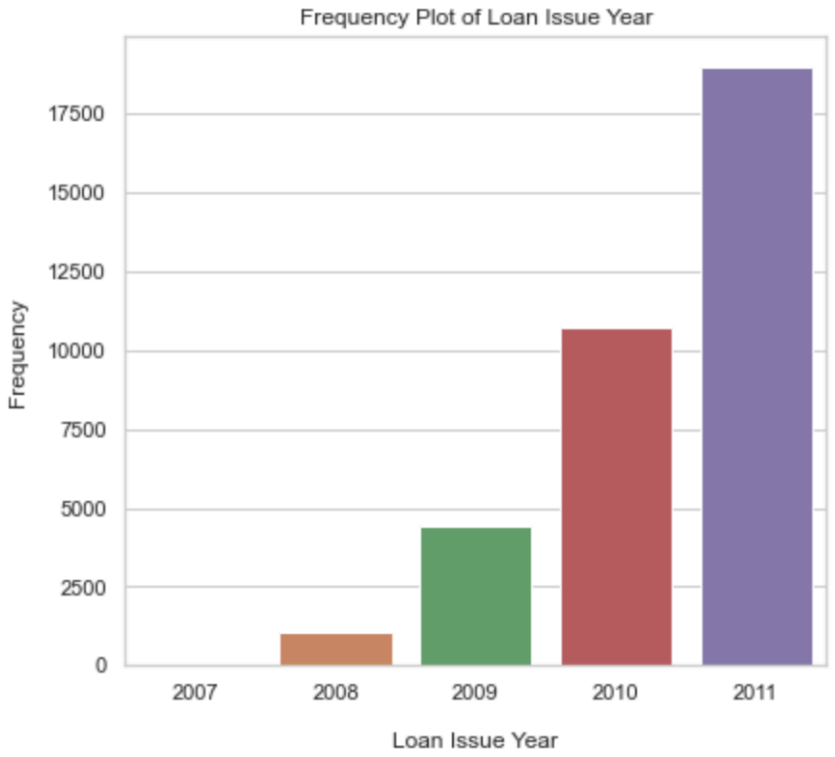
Name: emp_length_years, dtype: int64

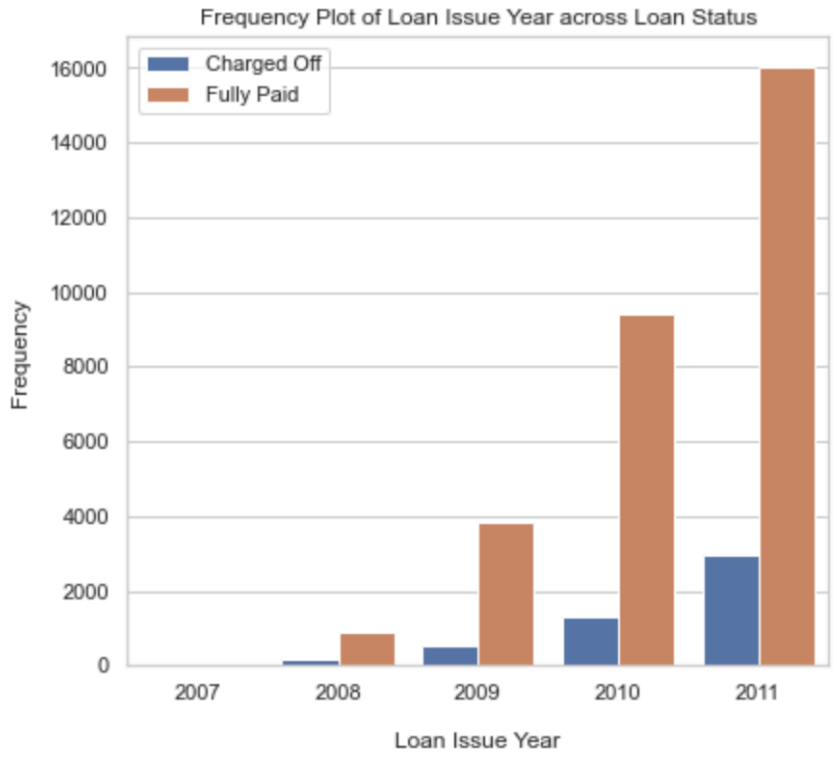
Observation:

- The analysis shows that most defaulted loan bowers has employment length as 10+ years

Ordered Categorical Variable: `issue_d_year`

- Rank-Frequency Plot of Unordered Categorical Variable: `'issue_d_year'`






```
loan_status  issue_d_year
Charged Off  2007                 2
             2008               163
             2009               552
             2010              1317
             2011              2959
Fully Paid   2007                 4
             2008               896
             2009              3860
             2010              9426
             2011             16042
Name: issue_d_year, dtype: int64
```

Observation:

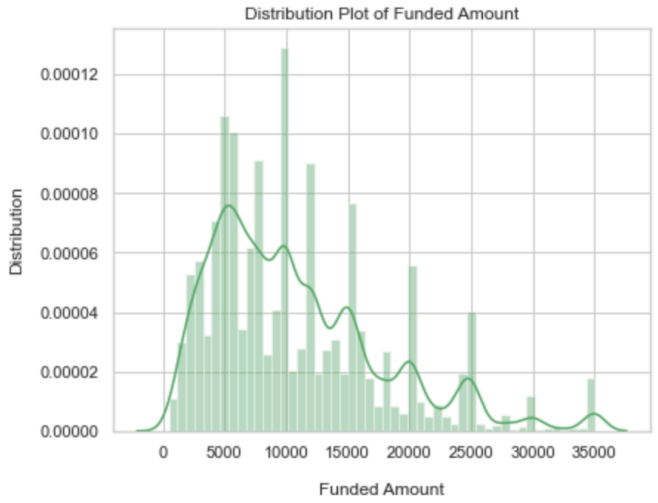
- The analysis shows that that loans issued in 2011 have the highest charged off loans.

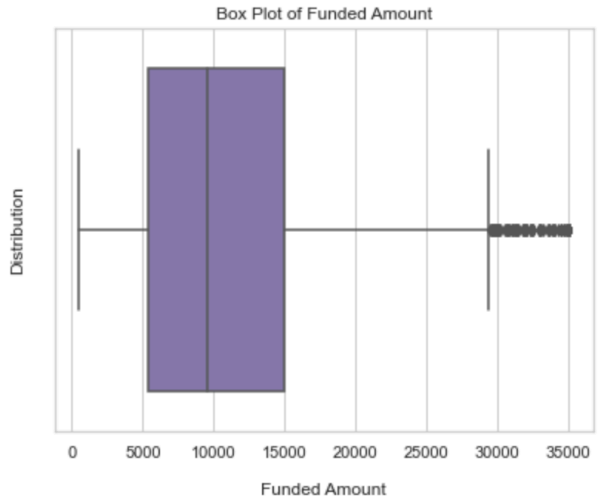
Univariate Analysis: Quantitative Variables

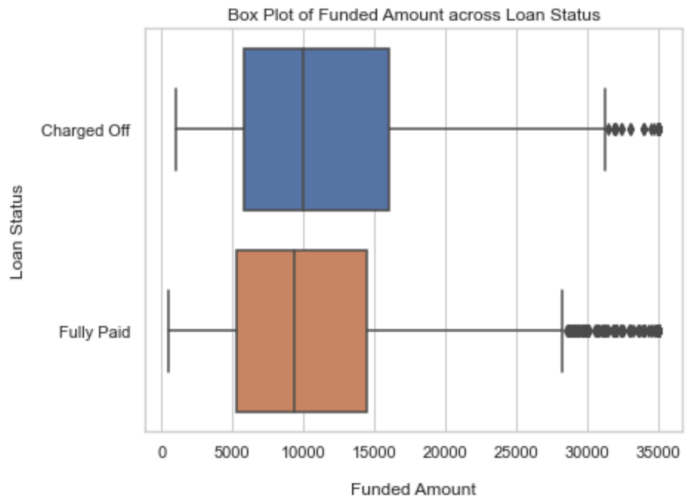
- Quantitative Variables are:
  - `'funded_amnt'`
  - `'int_rate_percent'`
  - `'installment'`
  - `'annual_inc'`
  - `'dti'`

Quantitative Variable: 'funded_amnt'

- Distribution and Box Plots of Quantitative Variable: 'funded_amnt'







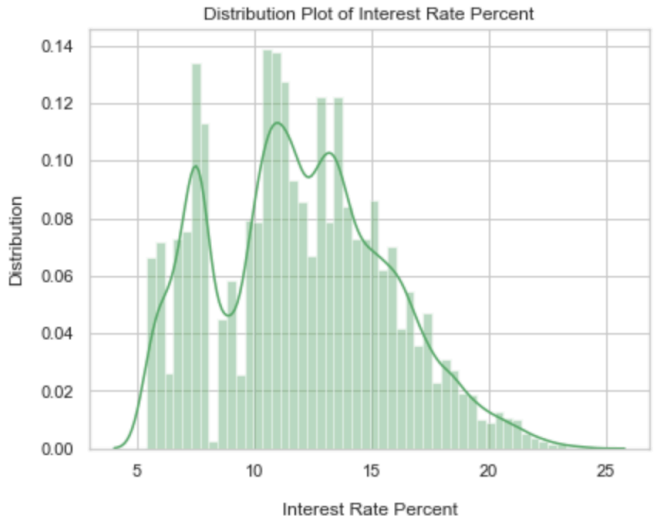
loan_status	count	mean	std	min	25%	50%	75%	max
Charged Off	4993.0000	11827.5035	7710.3750	1000.0000	5800.0000	10000.0000	16000.0000	35000.0000
Fully Paid	30228.0000	10628.8201	6882.4007	500.0000	5300.0000	9400.0000	14500.0000	35000.0000

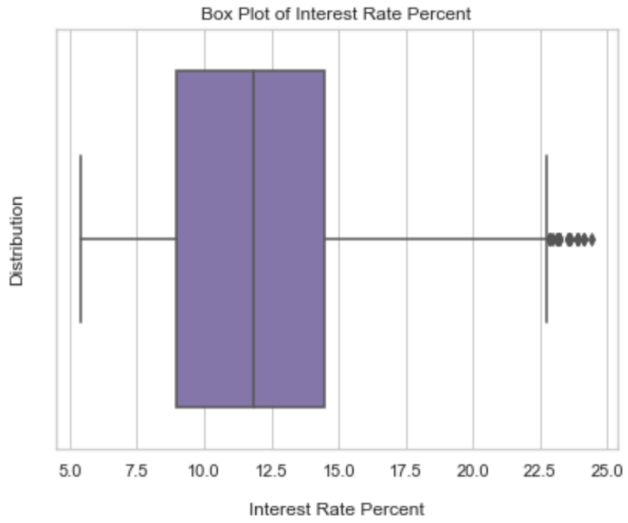
Observations:

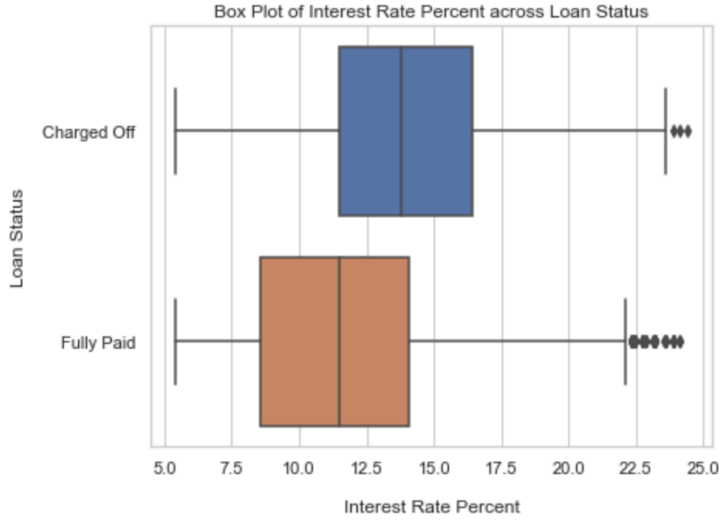
- Overall, the applied loan amount distribution is slightly right-skewed with mean greater than the median. Most of the loans granted are below 15000 (75 percentile value)
- Charged off loans are shifted towards higher average loan amount request.

Quantitative Variable: 'int_rate_percent'

- Distribution and Box Plots of Quantitative Variable: 'int_rate_percent'







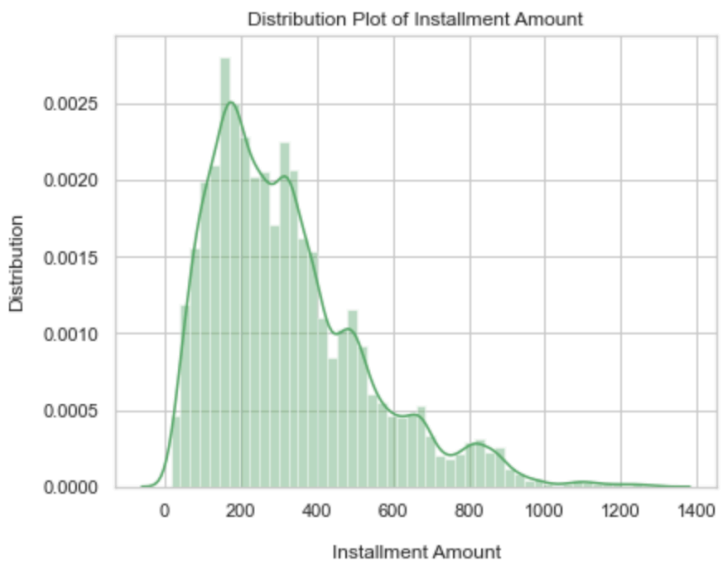
loan_status	count	mean	std	min	25%	50%	75%	max
Charged Off	4993.0000	13.9268	3.6404	5.4200	11.4900	13.7900	16.4500	24.4000
Fully Paid	30228.0000	11.6576	3.6082	5.4200	8.5900	11.4900	14.0900	24.1100

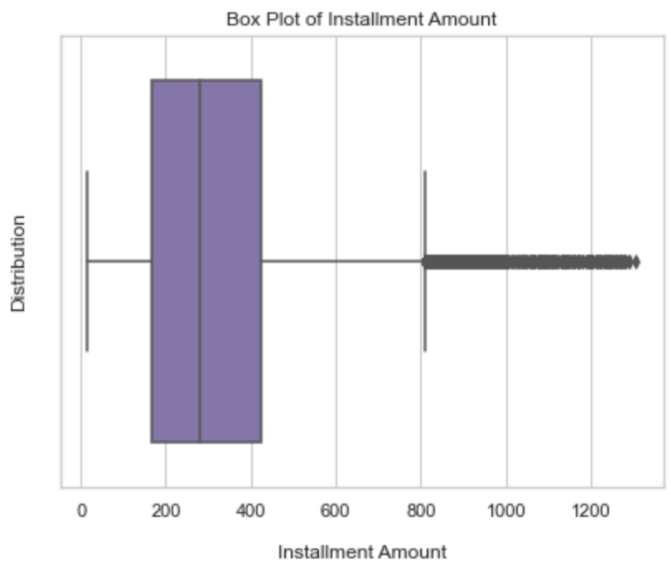
Observations:

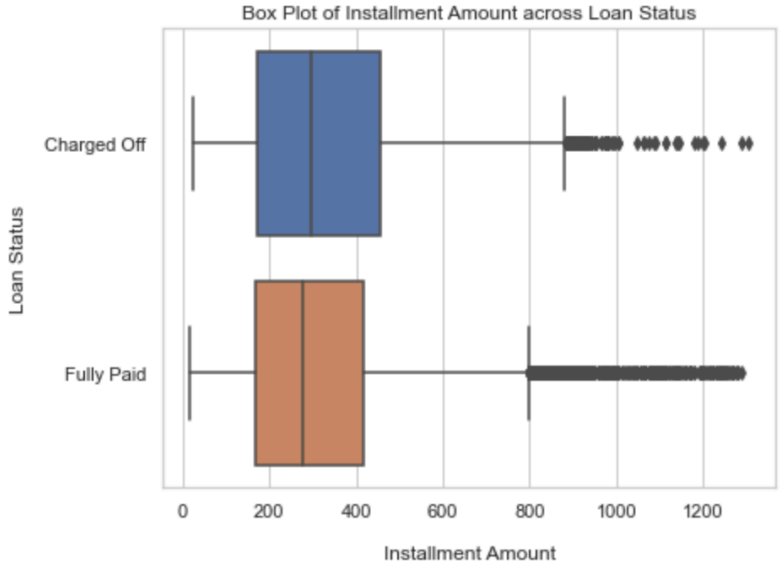
- Overall, the interest rate varies from 5.42% to 24.4% with an average interest rate of 11.8%.
- The interest rate for Charged Off loans appear to be higher than for Fully paid. This is expected as the risk increases and the rate of interest imposed on the loan also increases.

Quantitative Variable: 'installment'

- Distribution and Box Plots of Quantitative Variable: 'installment'







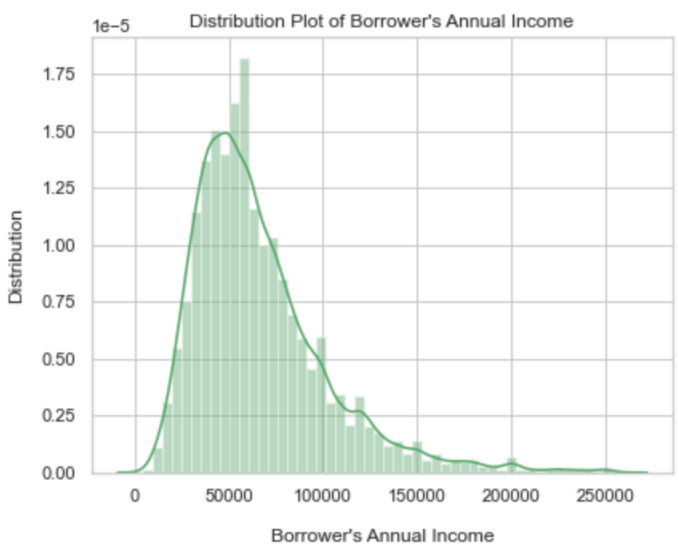
	count	mean	std	min	25%	50%	75%	max
loan_status								
Charged Off	4993.0000	336.1748	213.3655	22.7900	171.4300	296.7200	455.6300	1305.1900
Fully Paid	30228.0000	319.6301	203.9156	16.0800	166.6175	276.9000	418.9150	1288.1000

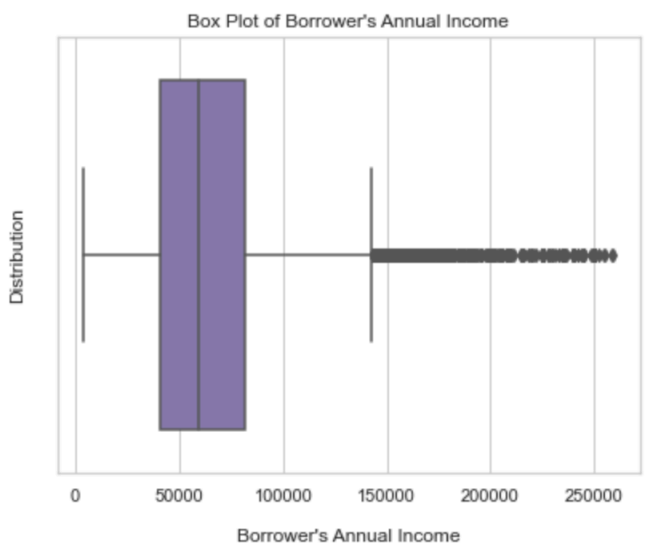
Observations:

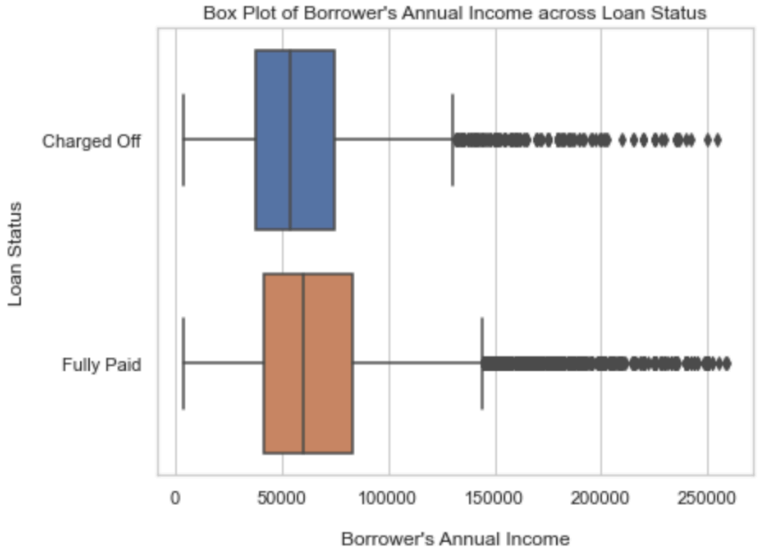
- The Charged Off loan Applicants has a central tendency of instalment amount as 296.72 with minimum 25th percentile value as 171.43 and maximum 75th percentile value as 455.63.
- The Charged Off loans have high instalment on an average.

Quantitative Variable: 'annual_inc'

- Distribution and Box Plots of Quantitative Variable: 'annual_inc'







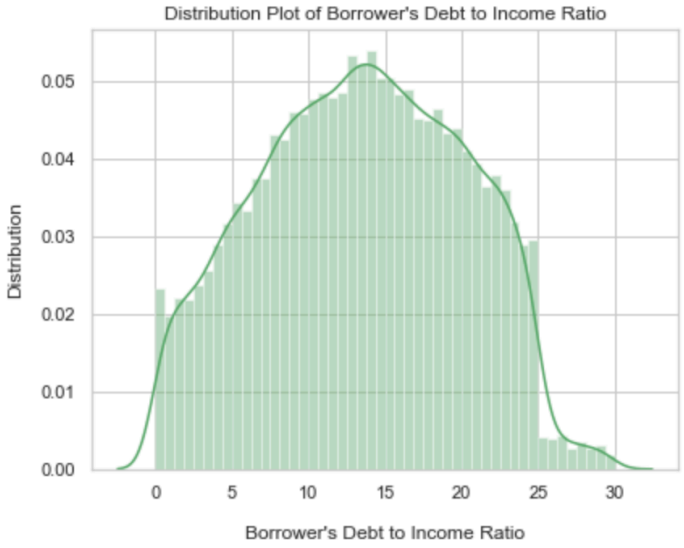
loan_status	count	mean	std	min	25%	50%	75%	max
Charged Off	4993.0000	60681.8622	33296.8313	4080.0000	38004.0000	54000.0000	75000.0000	254616.0000
Fully Paid	30228.0000	67057.5930	35953.1757	4000.0000	42000.0000	60000.0000	83004.0000	259000.0000

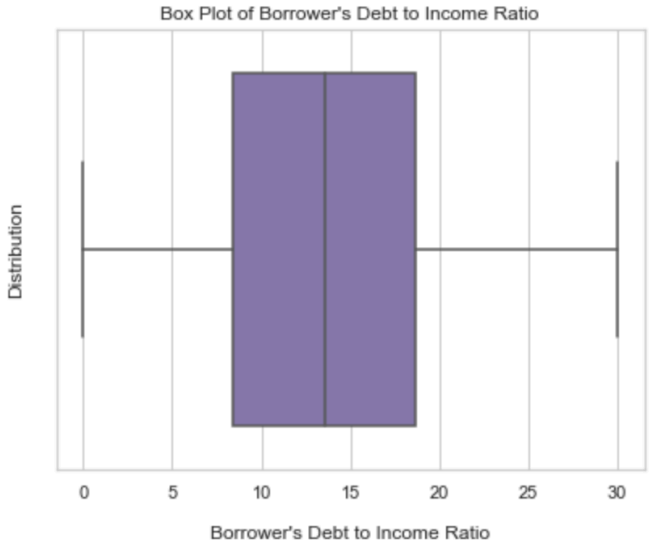
Observations:

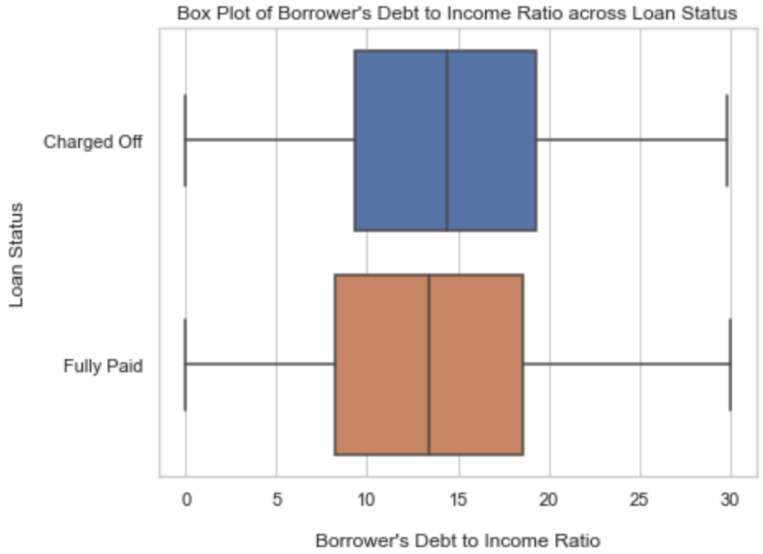
- The Charged Off loan applicants have a central tendency of annual income as 54000.00 with minimum 25th percentile value as 4080.00 and maximum 75th percentile value as 75000.00

Quantitative Variable: 'dti'

- Distribution and Box Plots of Quantitative Variable: 'dti'







loan_status	count	mean	std	min	25%	50%	75%	max
Charged Off	4993.0000	14.1622	6.5166	0.0000	9.3500	14.4400	19.3400	29.8500
Fully Paid	30228.0000	13.3271	6.6342	0.0000	8.2200	13.4000	18.5600	29.9900

Observations:

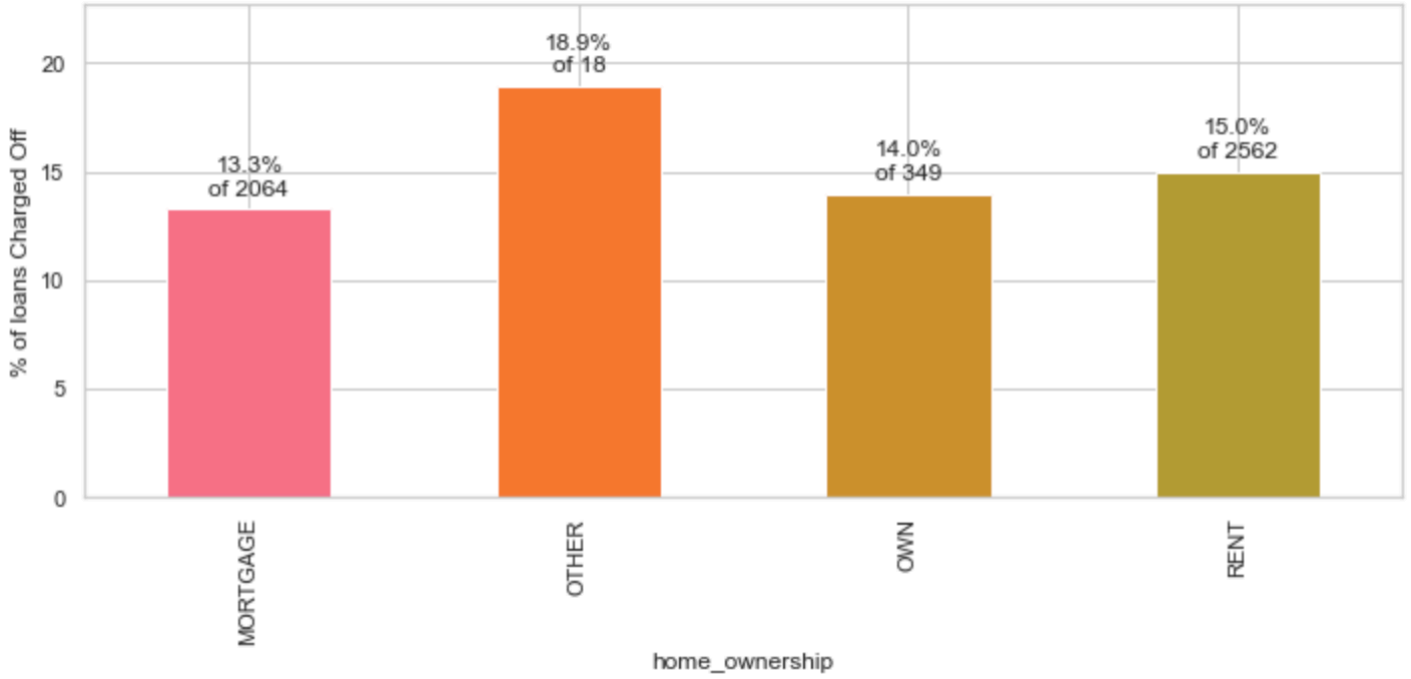
- The Charged Off loan applicants has a central tendency of 'dti' ratio as 14.44 with minimum 25th percentile value as 9.34 and maximum 75th percentile value as 19.35

EDA Step 4: Bivariate Analysis: Categorical Variable

- Variables are:
  - 'home_ownership'
  - 'purpose'
  - 'addr_state'
  - 'term'
  - 'grade'
  - 'sub_grade'
  - 'emp_length'
  - 'issue_d_year' and 'issue_d_month'
  - 'funded_amnt'
  - 'int_rate_percent'
  - 'installment'
  - 'annual_inc'
  - 'dti'

Categorical Variables: 'home_ownership'

- Categorical bivariable Analysis: 'home_ownership' against Charged Off Percentage Rate

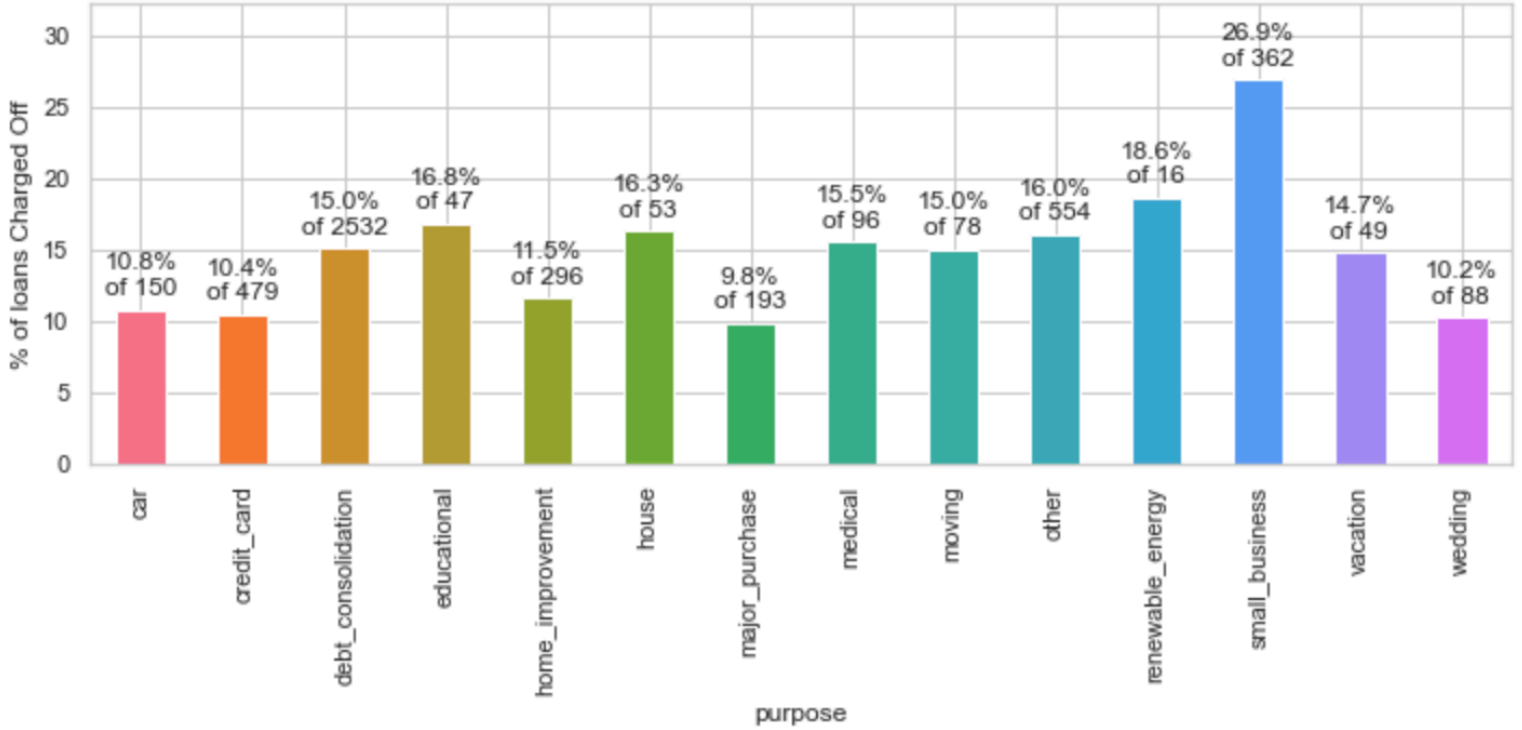


Observations:

- There is no substantial impact of 'home_ownership' on the charged off loans.

Categorical Variables: 'purpose'

- Categorical bivariable Analysis: 'purpose' against Charged Off Percentage Rate

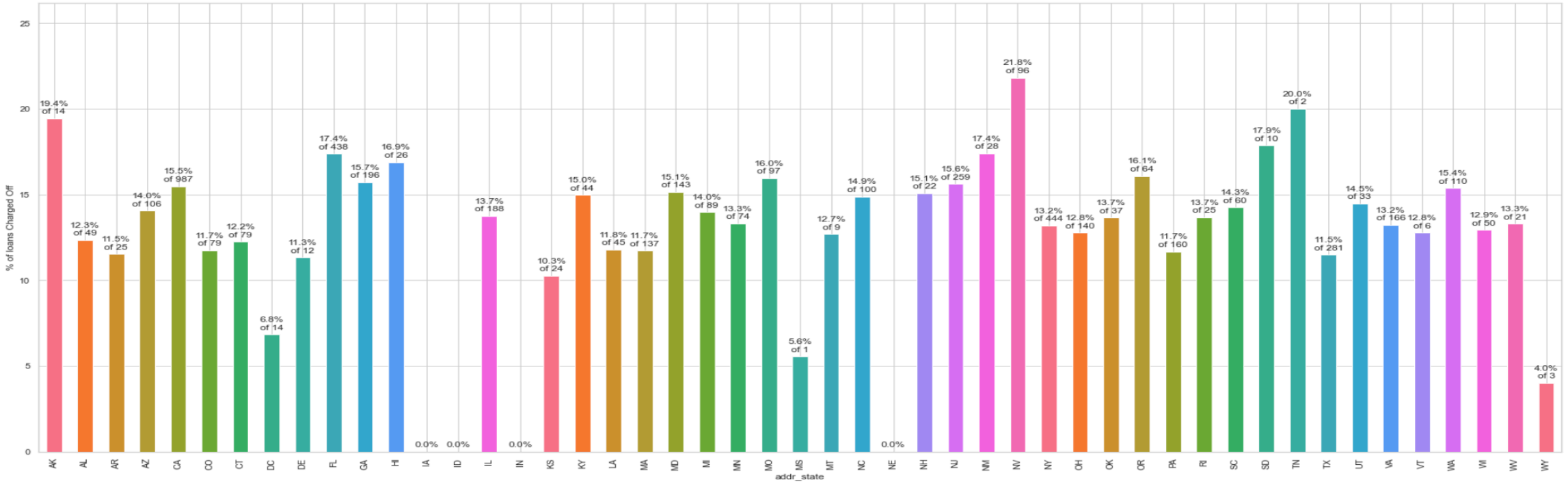


Observations:

- 26% of loans for small businesses are Charged Off. Making them the most risky purpose.
- Approximately ~49% of the loans are issued for the purpose of debt consolidation.
- 17% of the loans for renewable energy are charged Off, but the number is too less to be of significance.

Categorical Variables: 'addr_state'

- Categorical bivariable Analysis: 'addr_state' against Charged Off Percentage Rate

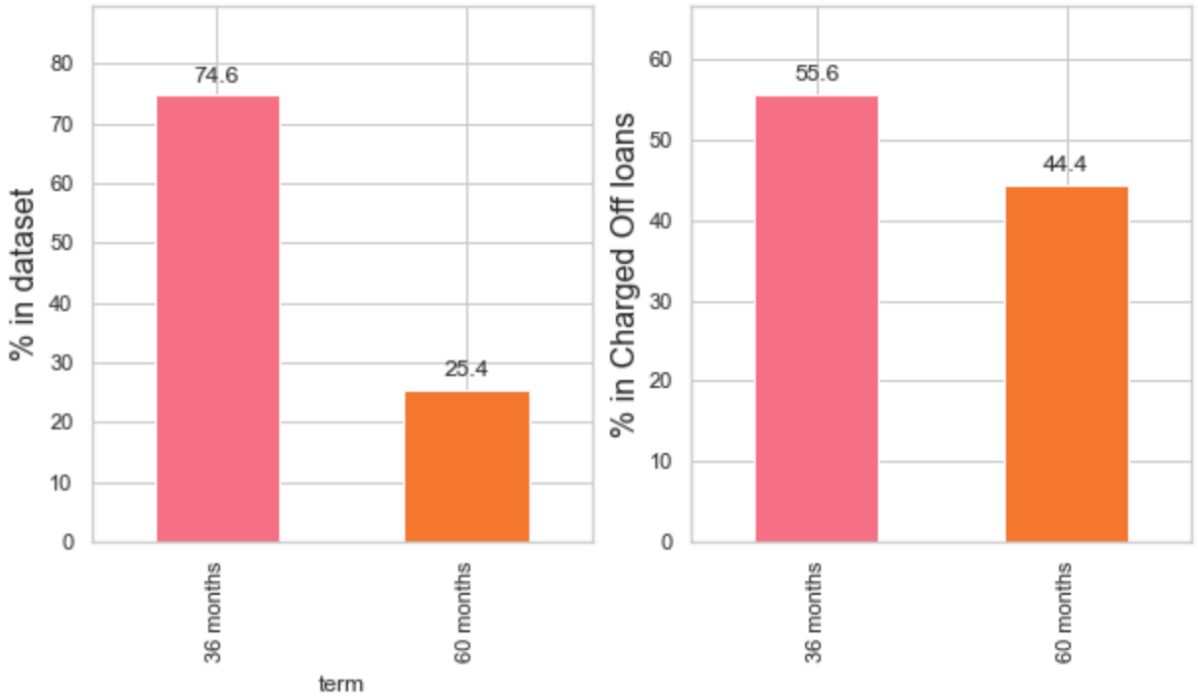


Observations:

- Maximum loans are from California state followed by New York, Florida and Texas state.
- States with higher Charge Off rates have very low numbers of loans. The percentage is therefore NOT significant and should be ignored. Overall, this variable has no impact on loan defaulting.

Categorical Variables: 'term'

- Categorical bivariable Analysis: 'term' against Charged Off Percentage Rate

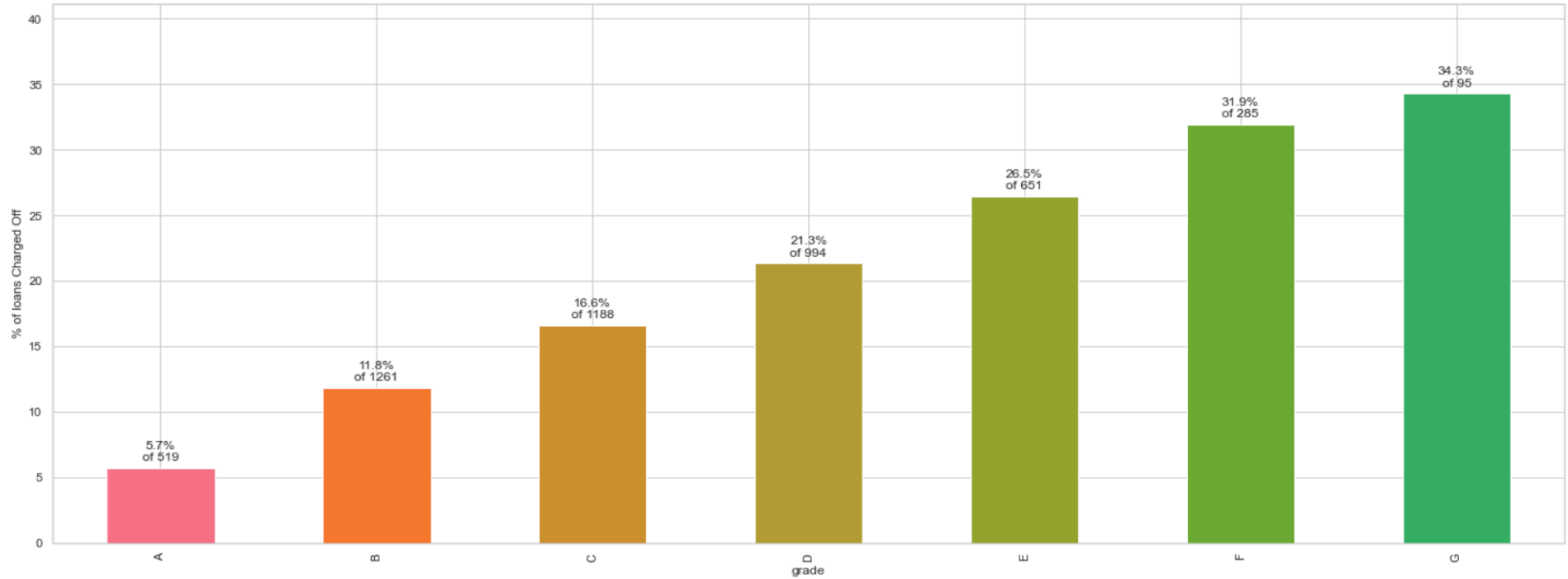


Observations:

- Among Charged Off loans, percentage of term 60 months rises to 45% as compared to total dataset.
- The higher term loans have a higher chances of default.

Categorical Variables: 'grade'

- Categorical bivariable Analysis: 'grade' against Charged Off Percentage Rate

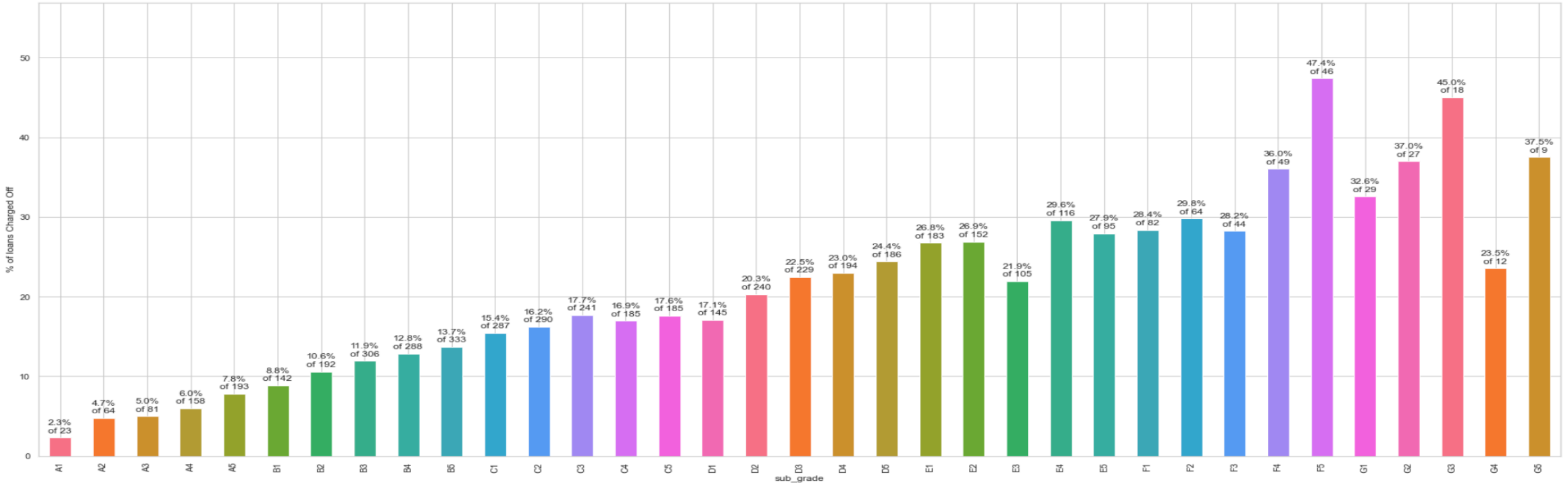


Observations:

- Nearly 30% of loans in Grades F and G see a default.
- Loans Grade E onwards have a tendency to be charged off more.

Categorical Variables: 'sub_grade'

- Categorical bivariable Analysis: 'sub_grade' against Charged Off Percentage Rate

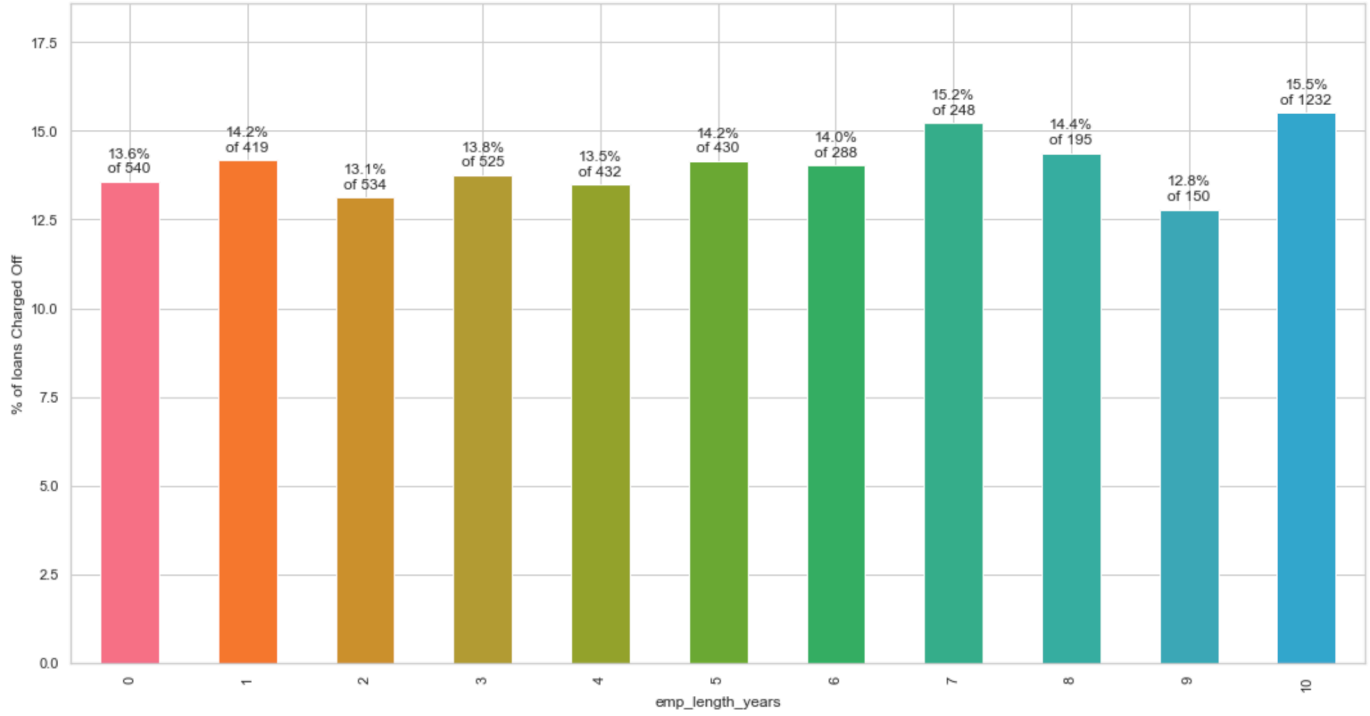


Observations:

- As the grade and subgrade increased the percentage of loans charged off also increased.

Categorical Variables: 'emp_length_years'

- Categorical bivariable Analysis: 'emp_length_years' against Charged Off Percentage Rate

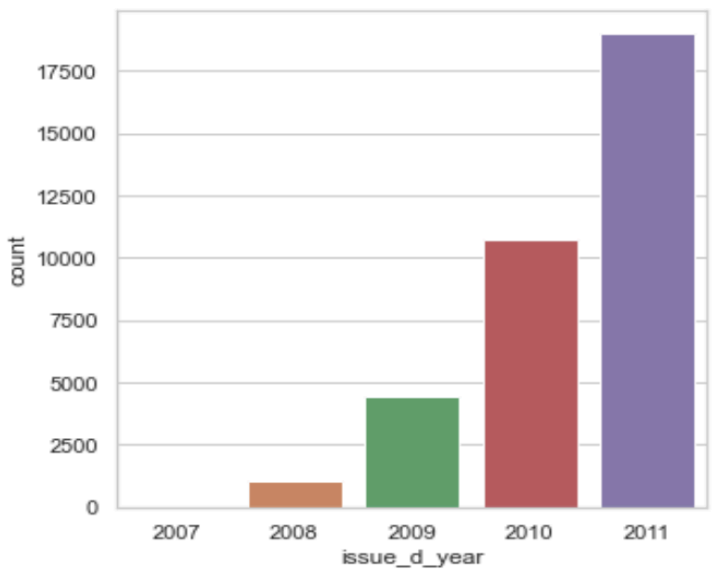


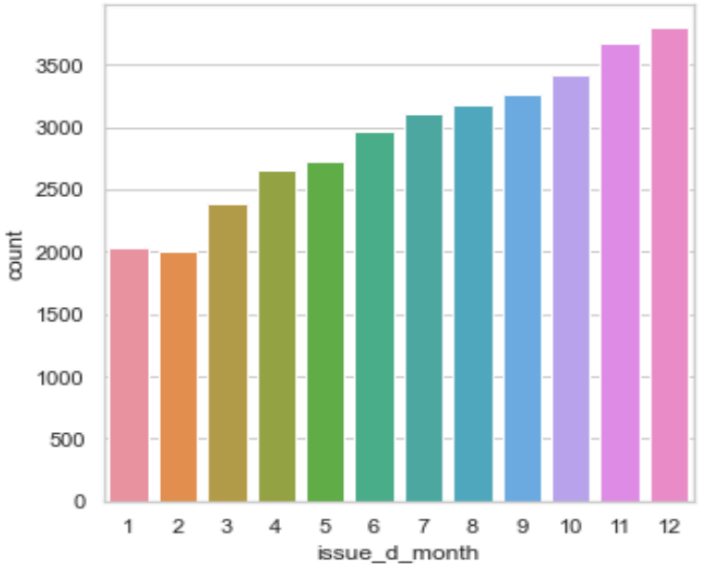
Observation:

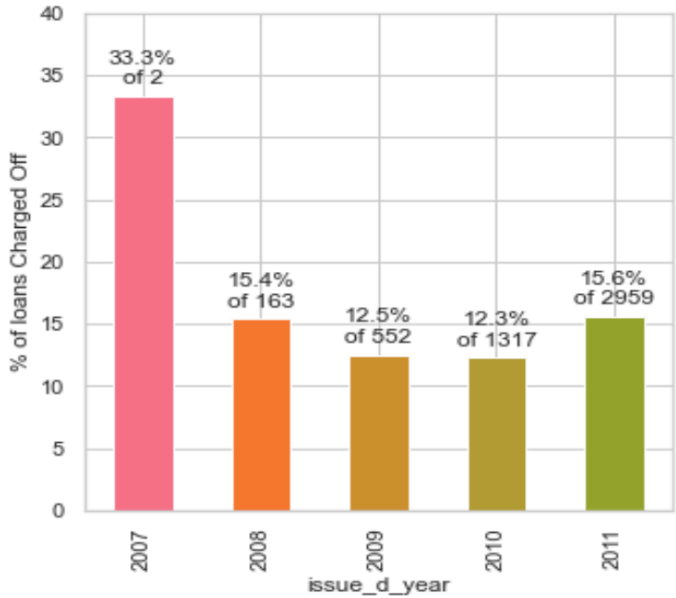
- 15.5% of loans charged off for the applicants with 10 years or more employment length.

Categorical Variables: 'issue_d_year' and 'issue_d_month'

- Categorical bivariable Analysis: 'issue_d_year' and 'issue_d_month' against Charged Off Percentage Rate





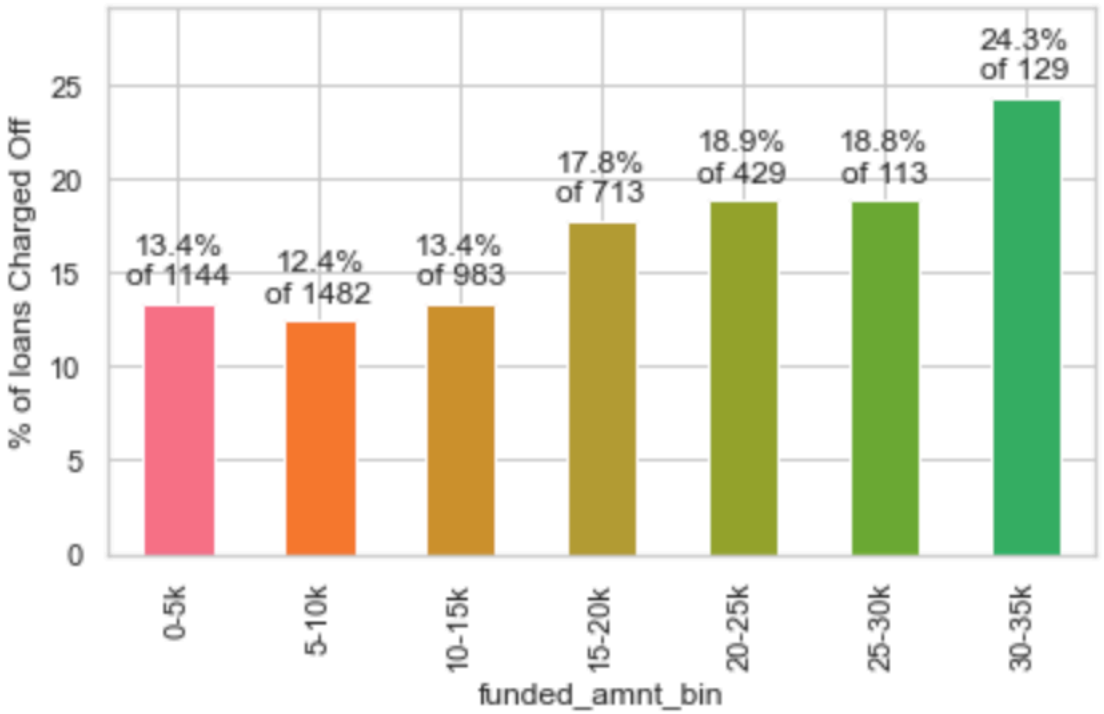


Observations:

- Over the years, LC has given more loans. From 2007 to 2011 the loans issued have significantly increased.
- The number of loans issued increases from Jan to Dec. In the month of December the maximum number of loans were issued.
- Year of loan has no significant impact on loans being charged off.

Categorical Variables: 'funded_amnt'

- Categorical bivariable Analysis: 'funded_amnt' against Charged Off Percentage Rate
- Funded loan amount has values ranging from 0 to 35000.
- Create a derived categorical variable 'funded_amnt_bin' using 5000 as bucket size.

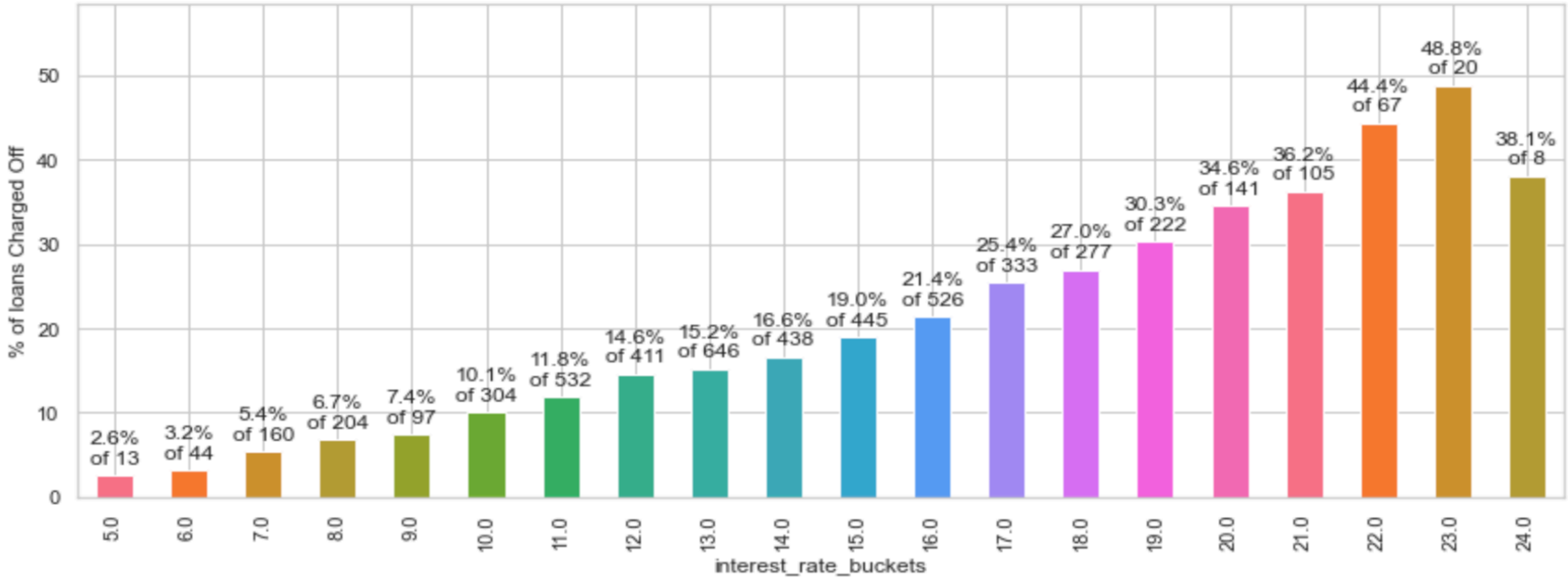


Observations:

- The % of charged off loans increases substantially as the funded loan amount increases.
- The majority of loans are below 20000.
- It shows that higher the funded loan amount higher risk of the default.

Categorical Variables:'int_rate_percent'

- Categorical bivariable Analysis: 'int_rate_percent' against Charged Off Percentage Rate
- Interest Rate has many of small values ranging from 5% to 25%.
- Create a derived categorical variable 'int_rate_percent_buckets' using 1 unit as bucket size.

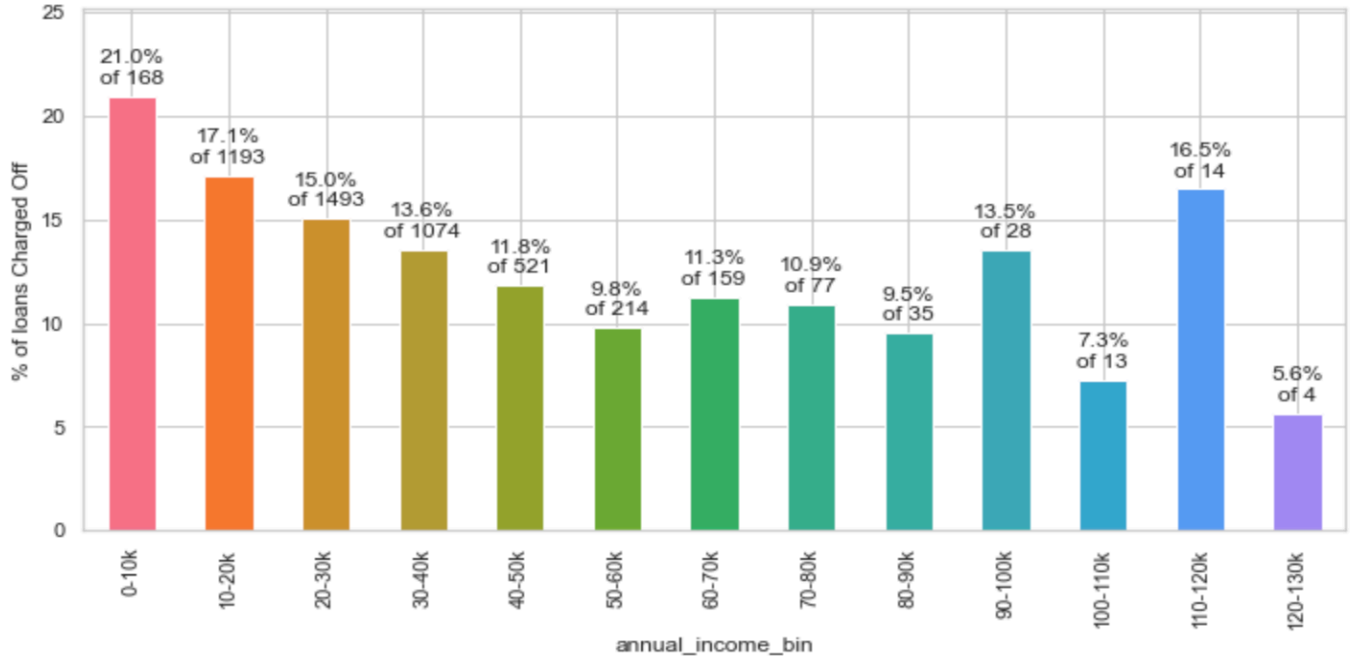


Observation:

- As the interest rate increases the Percentage of charged off loans also increases.

Categorical Variables: 'annual_inc'

- Categorical bivariable Analysis: 'annual_inc' against Charged Off Percentage Rate
- Borrower's Annual Income has values ranging from 4000 to 234996.
- Create a derived categorical variable 'annual_income_bin' using 10000 as bin size.

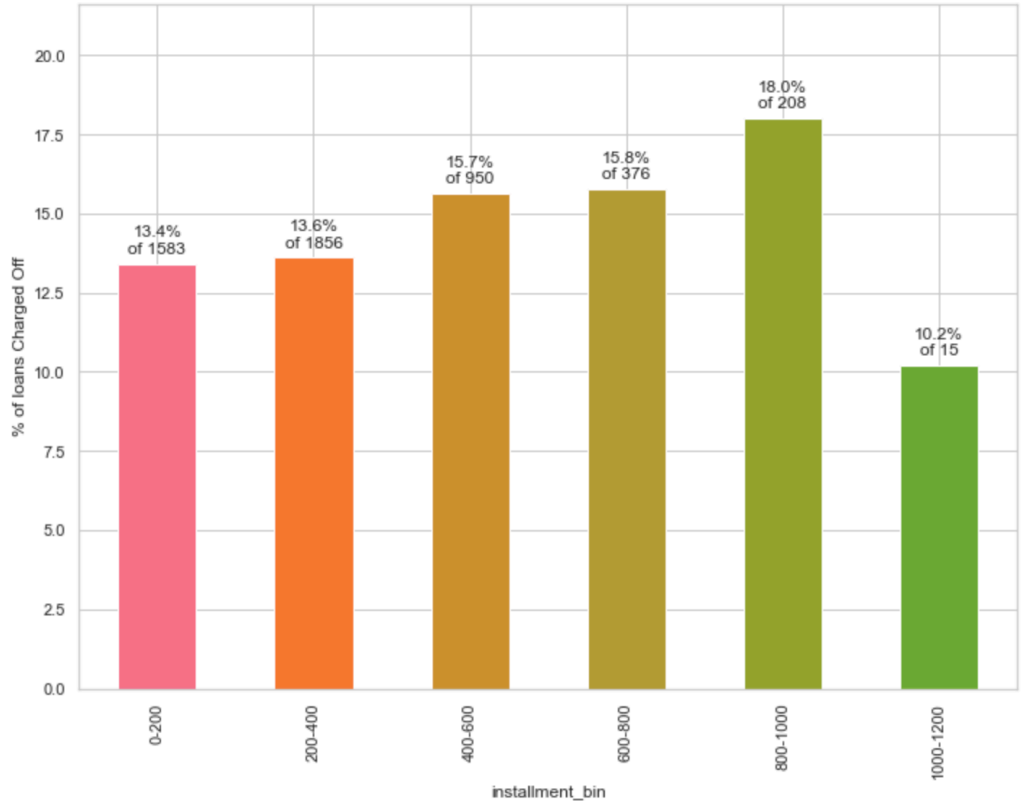


Observation:

- Loan defaults are higher for lower income and progressively reduce as incomes go up.

Categorical Variables:'installment'

- Categorical bivariable Analysis: 'installment' against Charged Off Percentage Rate
- Borrower's Instalment has values ranging from 16 to 1400.
- Create a derived categorical variable 'installment_bin' using 200 as bin size.

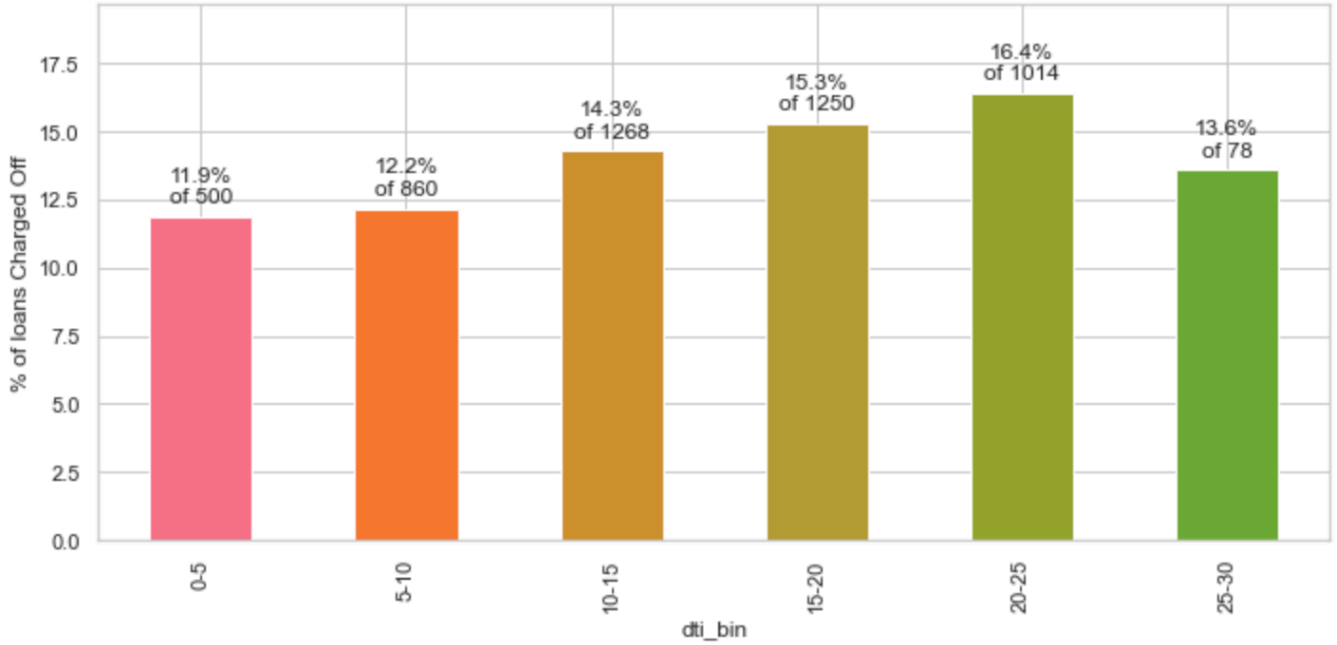


Observation:

- The installment amount increases as the percentage charged off loan increases.

Categorical Variables:'dti'

- Categorical bivariable Analysis: 'dti' against Charged Off Percentage Rate
- The borrower's Debt-to-Income Ratio has values ranging from 0 to 30.
- Create a derived categorical variable 'dti_bin' using 5 as bin size.

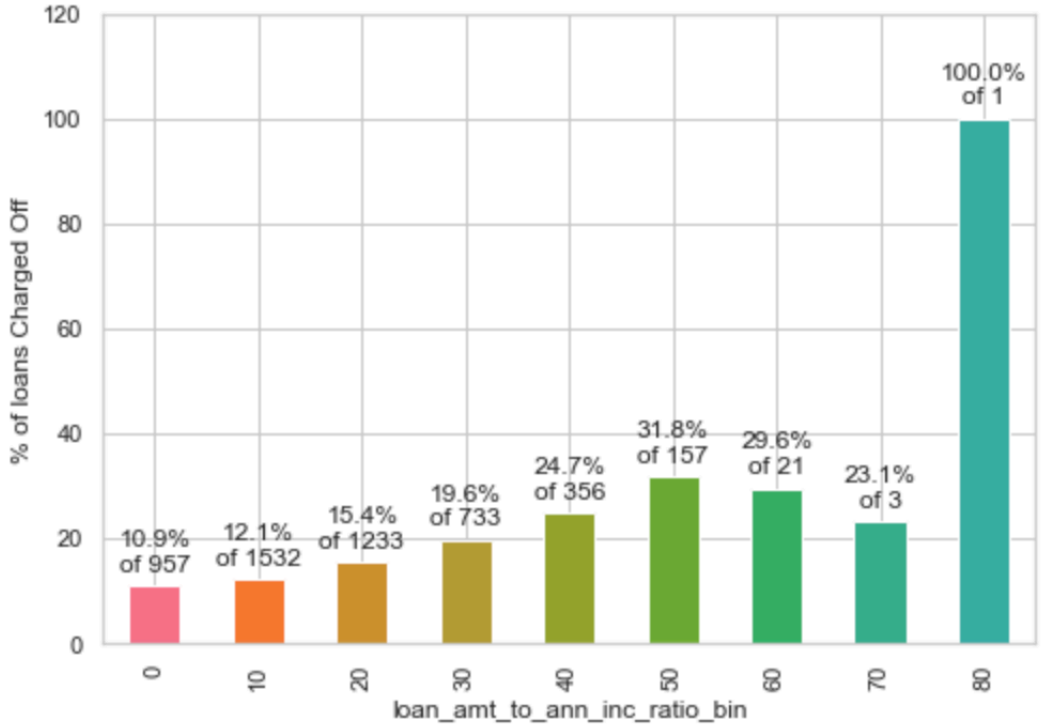


Observations:

- When the 'dti' ratio is higher than 15, higher percentage of loans are Charged Off
- Higher the 'dti' ratio higher will be the chances of loan getting defaulted

Derived Variables:'loan_amnt'To 'annual_inc'

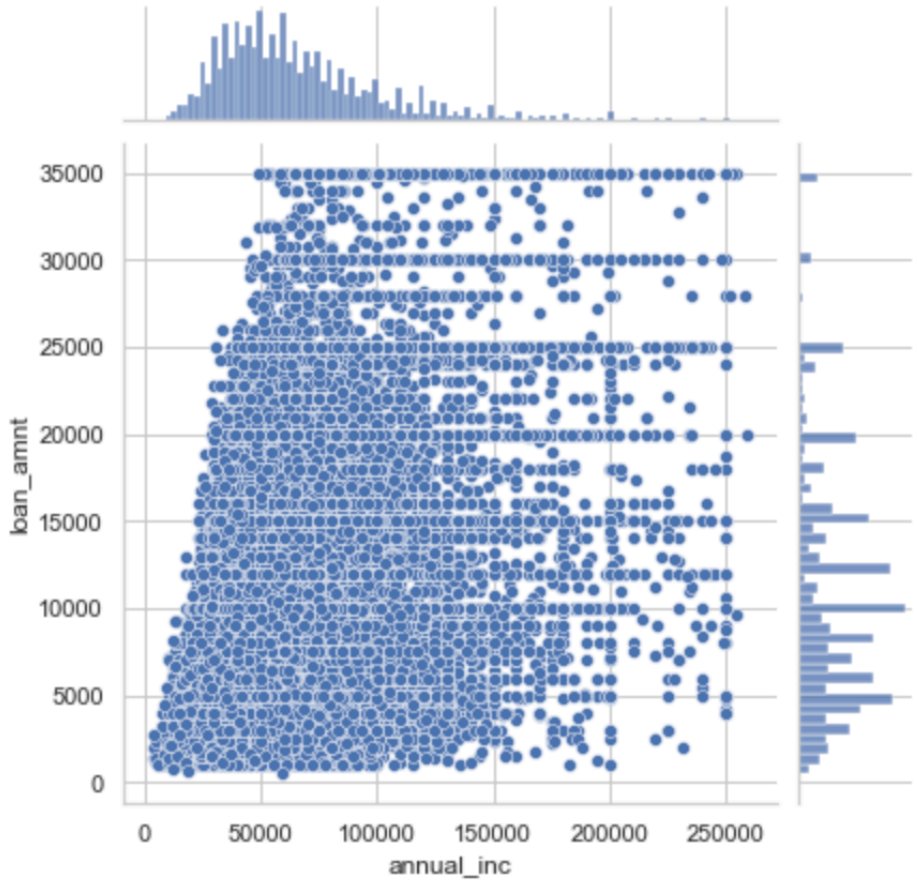
- New derived variable defined as `'loan_amt_to_ann_inc_ratio'`
- Create a derived categorical variable `'loan_amt_to_ann_inc_ratio_bin'` using 10 as bin size.



Observations:

- As long as the loan amount is less than 20% of annual income, defaults are low.
- Loan amounts of 30% of annual income or higher see a high rate of default.

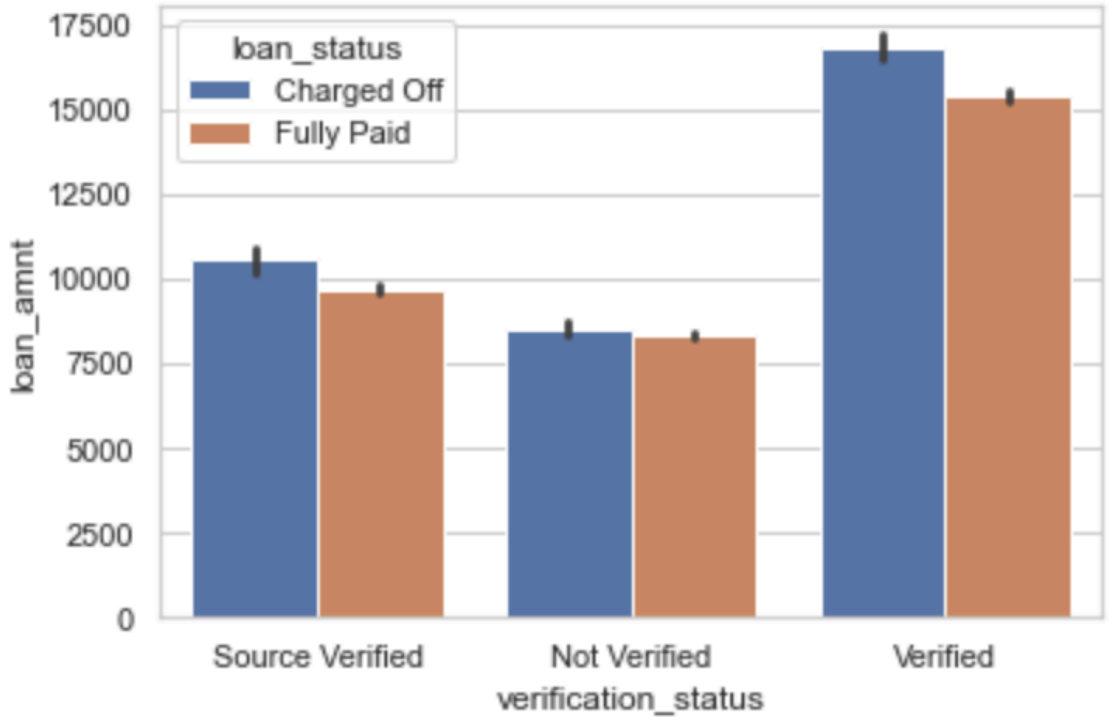
Bivariate Analysis:'annual_inc' Vs 'loan_amnt'



Observation:

- The borrowers with average income lower than 50000 taking loans of 25000 or higher. This could lead to charged off loans.

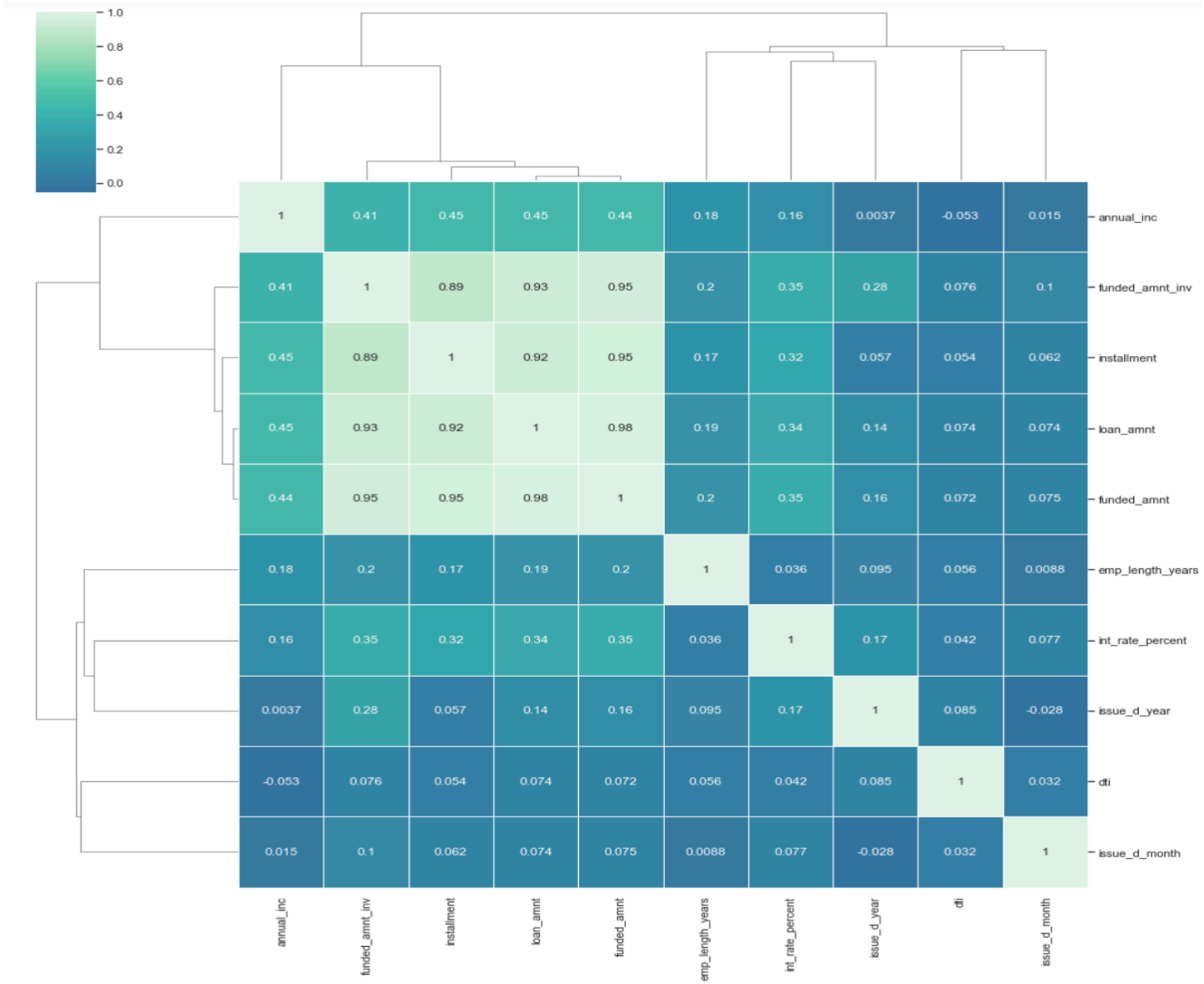
Bivariate Analysis:'loan_amnt' Vs 'verification_status' Vs 'loan_status'



Observations:

- Higher loan amounts are Verified more often.
- The higher amount loans have a higher default rate.
- The higher loan amounts are verified more often by Lending Club.

Bivariate Analysis: Correlation Metric



Observations:

- `'installment'`, `'funded_amnt'`, `'loan_amnt'` and `'funded_amnt_inv'` are highly correlated to each other. They form a cluster.
- `'dti'` (Debt-to-Income Ratio) is negatively correlated to `'annual_inc'`

EDA : Conclusion

As per EDA below **consumer** and **loan** attributes are identified which can influence the tendency of default.

- Annual Income - Borrowers with lower income range may result in loan defaulters.
- Employment Length - Borrowers with employment tenure as 10 years or more have resulted in more charged off loans. Additional data needs to be sourced and analysed to identify the reasoning.
- Loan Purpose - Loans for 'small business' and 'debt consolidation' purposes have resulted in a high number of charged off loans.
- Address State - Borrowers from specific states such as 'California', 'New York', 'Florida' and 'Texas' states are defaulted more as compared to other state residents.
- Debt-to-Income Ratio - Borrowers with higher debt as compared to their income resulted in more charged off loans.
- Loan Term - Higher loan terms show a higher rate of charged off loans.
- Grades - Loans grade 'E' onwards have a tendency to be charged off more. As the grade and subgrade increased the percentage of loans charged off also increased.
- Funded Loan Amount - Higher funded loan amount results in higher risk of the loan defaulting. The risk of loan defaulting increases, if the loan amount is more than 30% of Borrower's annual income.
- Interest Rate - Higher interest Rate results in higher loan default rates as well.
- Instalment Amount - Loans with higher Instalment amount have tendency to default.

EDA : Recommendations

As per EDA below recommendations are suggested

- Annual Income – Prefer borrowers with high income over low-income borrowers.
- Loan Purpose –
  - If the loan applied for 'small business', the perform the details analysis of financials of the small business and their plans to utilise the loan amount.
  - Preferably avoid any loan applied for the purpose of 'debt consolidation'
- Debt-to-Income Ratio – Avoid borrowers with high debt as compared to their income.
- Loan Term – Always associate high term loans with higher interest rates.
- Grades – Avoid loans graded as 'E' level onwards.
- Funded Loan Amount – Always verify the borrows source of income and approve loans amount not more than 5-10 times of borrower's annual income.
- Instalment Amount – Avoid any loan approvals if the Instalment amount is more than the 40% of borrower's monthly income.
- Borrowers Verification - Always conduct a detailed background verification of borrowers for their employment details, residential details and financial details.