



# Lending Club Case Study

Nitish Sharma

Chetna Priyadarshini



# Problem Statement

- Help a finance company identify risky loan applicants
- Understand consumer attributes and loan attributes that are likely to cause the account to be defaulted
- To serve above
  - Clean data removing null columns, replace NA with 0 where applicable, identify and remove rows that do not have any important information
  - Study distribution of data on various rows and perform outlier treatment
  - Understand correlation between different variables in people who defaulted on the loan and identify a common pattern
- Combine the rows together to arrive at a conclusion on risk factors for loan defaulters

# Data Cleaning

Initial data: 39717 rows and 111 columns

Removed null columns: 39717 rows and 57 columns

Removed columns with only one value: 39717 rows and 51 columns

Removed columns which have more than 20% null values: 39717 rows and 47 columns

Removed columns with id/membership\_id(all unique vals): 39717 rows and 45 columns

Removing columns not interesting for analysis: 39717 rows, 23 columns

Replace remaining null values with 0

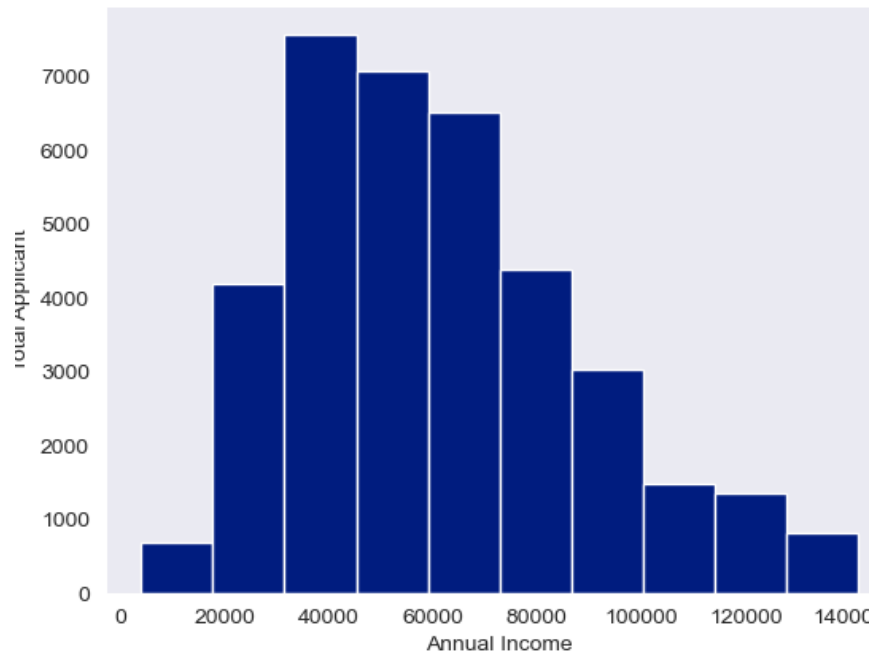
Standardize column values removing trailing string of %, months, x, converting to numeric where possible

Remove outliers from Loan Amount, Annual Income, Debt to Income ratio columns

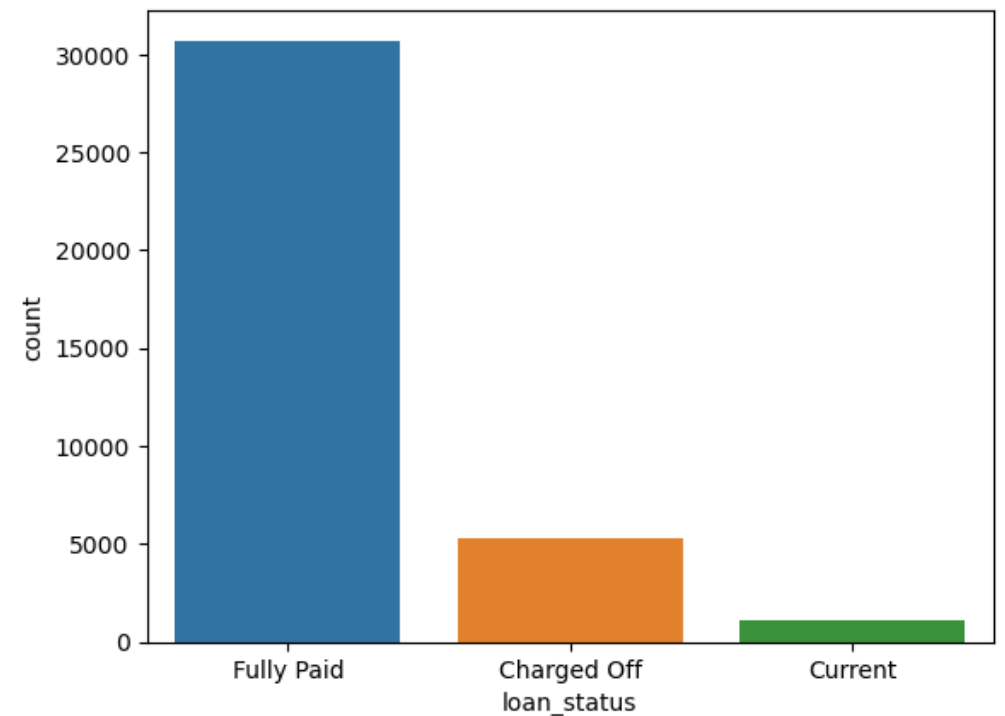
# Univariate Analysis: Loan Status & Annual Income

- We can see that most of the users are in Fully paid category
- maximum number of loan applicants have annual income of 30000 - 40000.
- 4500 people have salary less than 30000.

Annual Income of the Applicants

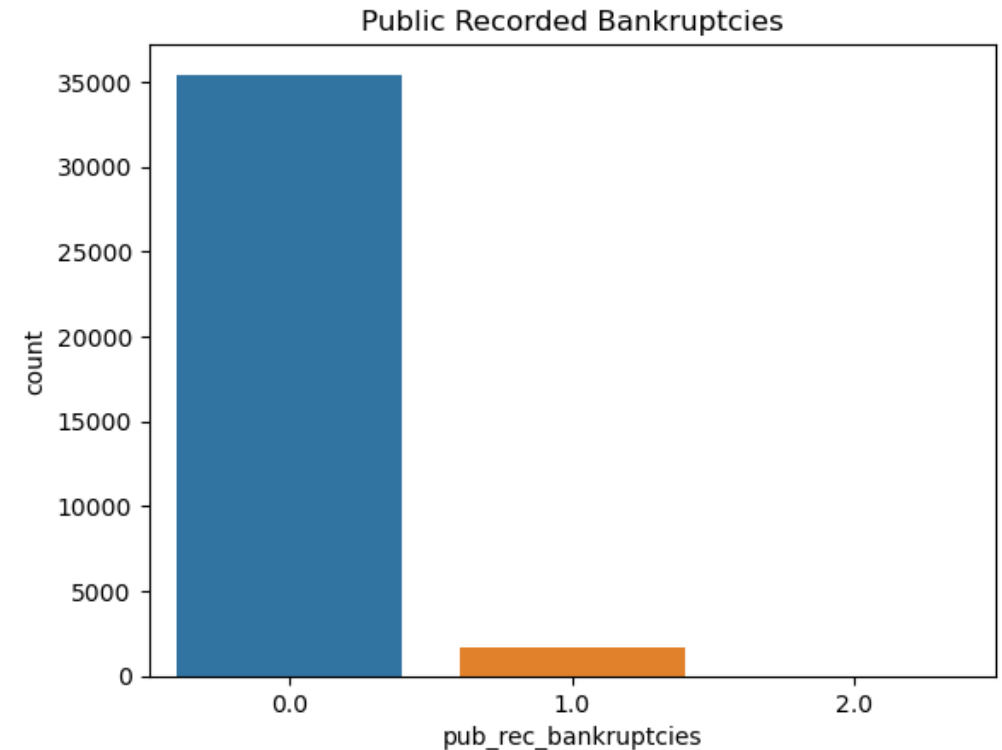
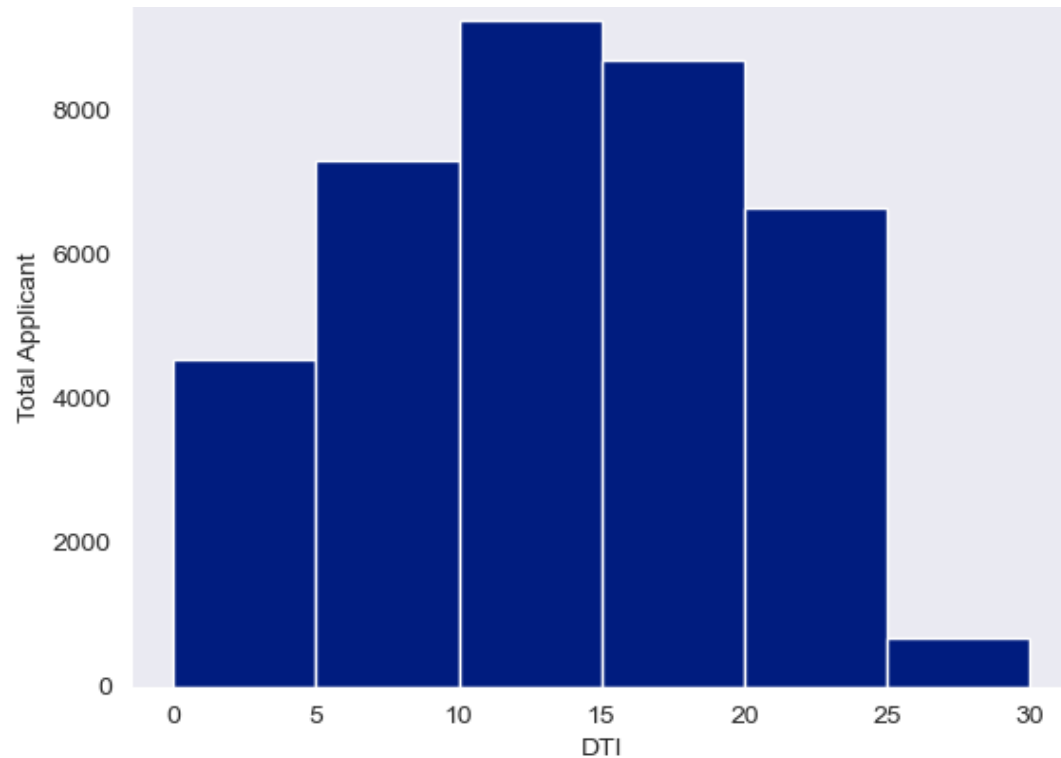


Loan Status



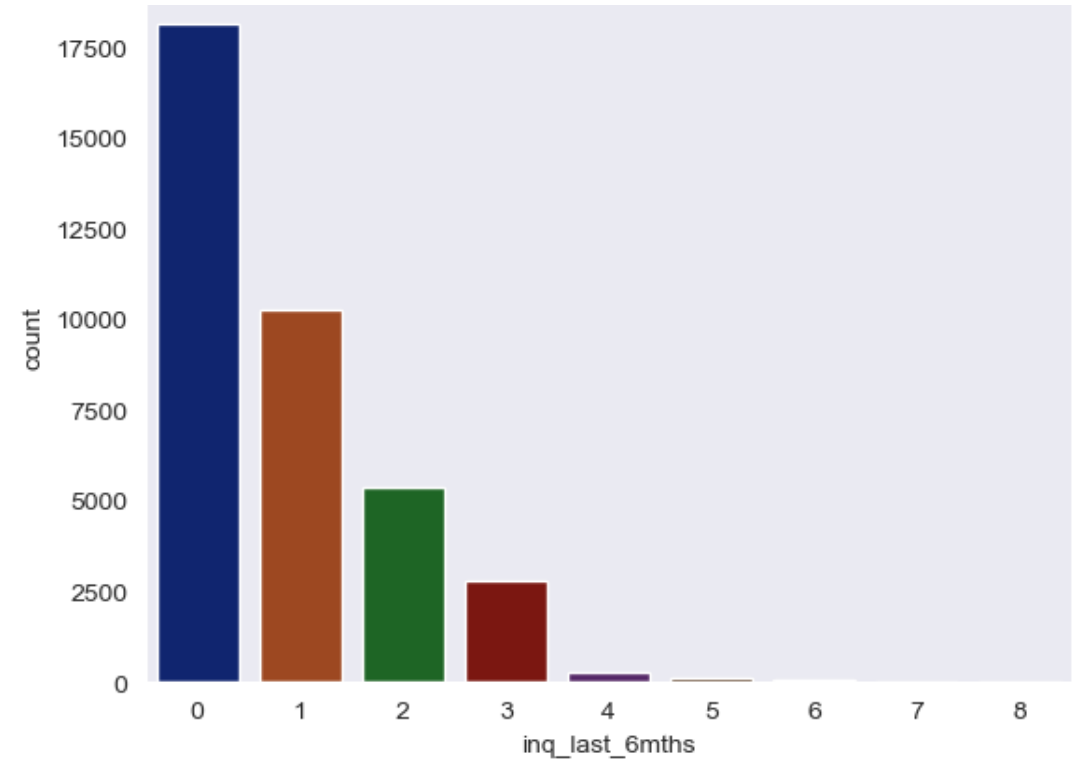
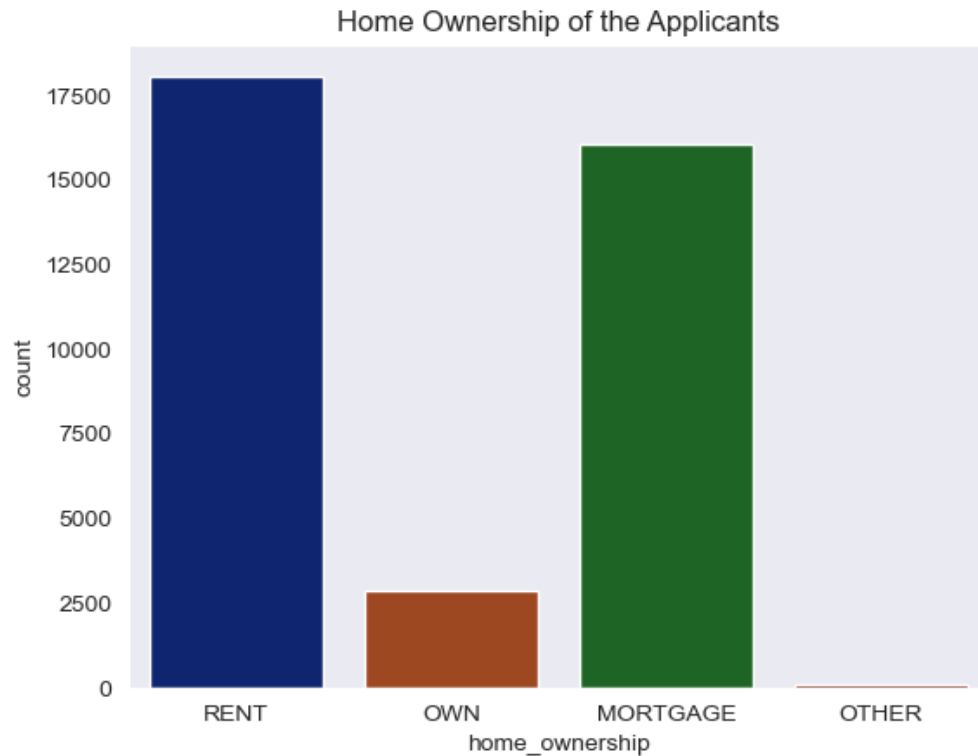
# Univariate Analysis: Debt to Income & Public recorded bankruptcies

- Composition of highest applicants with top 3 DTI falls under the categories as 10-15, 15-20 and 5-10
- 1634 applicants with public recorded bankruptcies



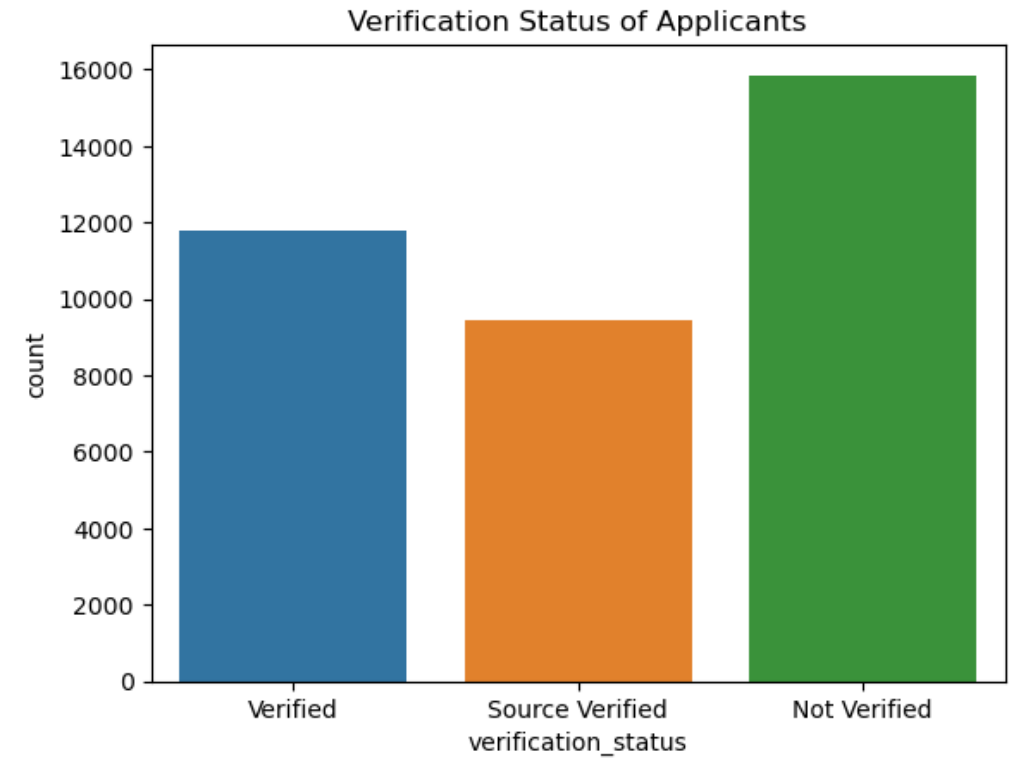
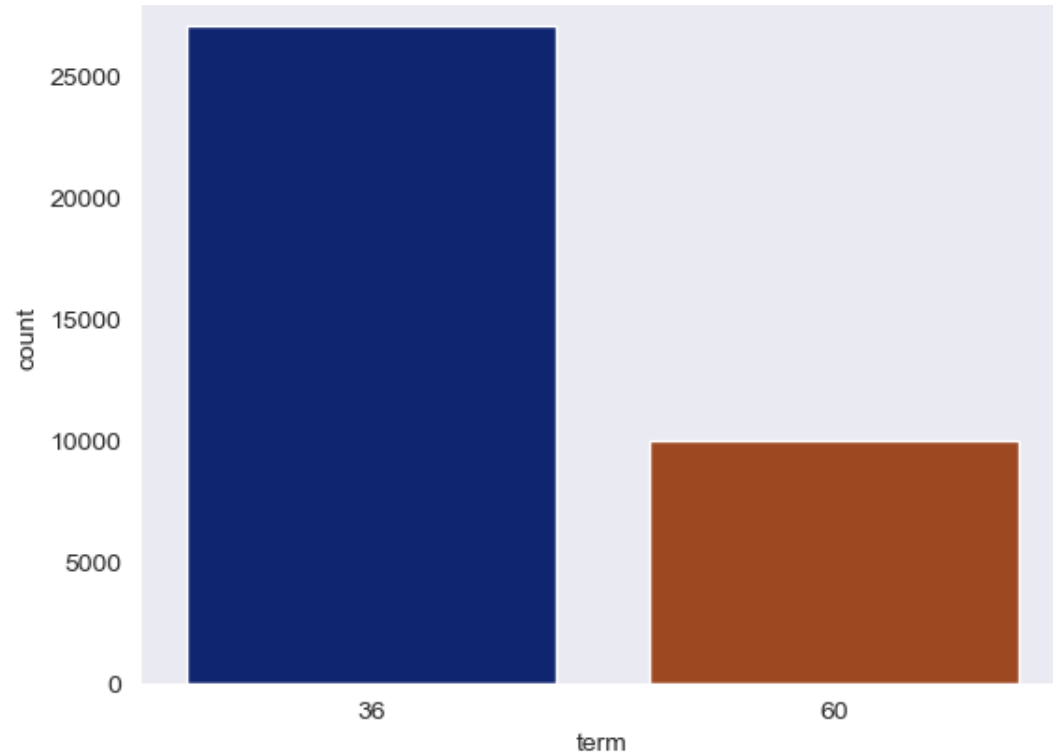
# Univariate Analysis: Home Ownership & Loan enquiry in last 6 months

- Most of the applicants are either living on rent or property is mortgage.
- Most of the applicants have 0 to 3 enquiries.
- Some applicants with enquiry between 4 to 8



# Univariate Analysis : Loan Term & Verification Status

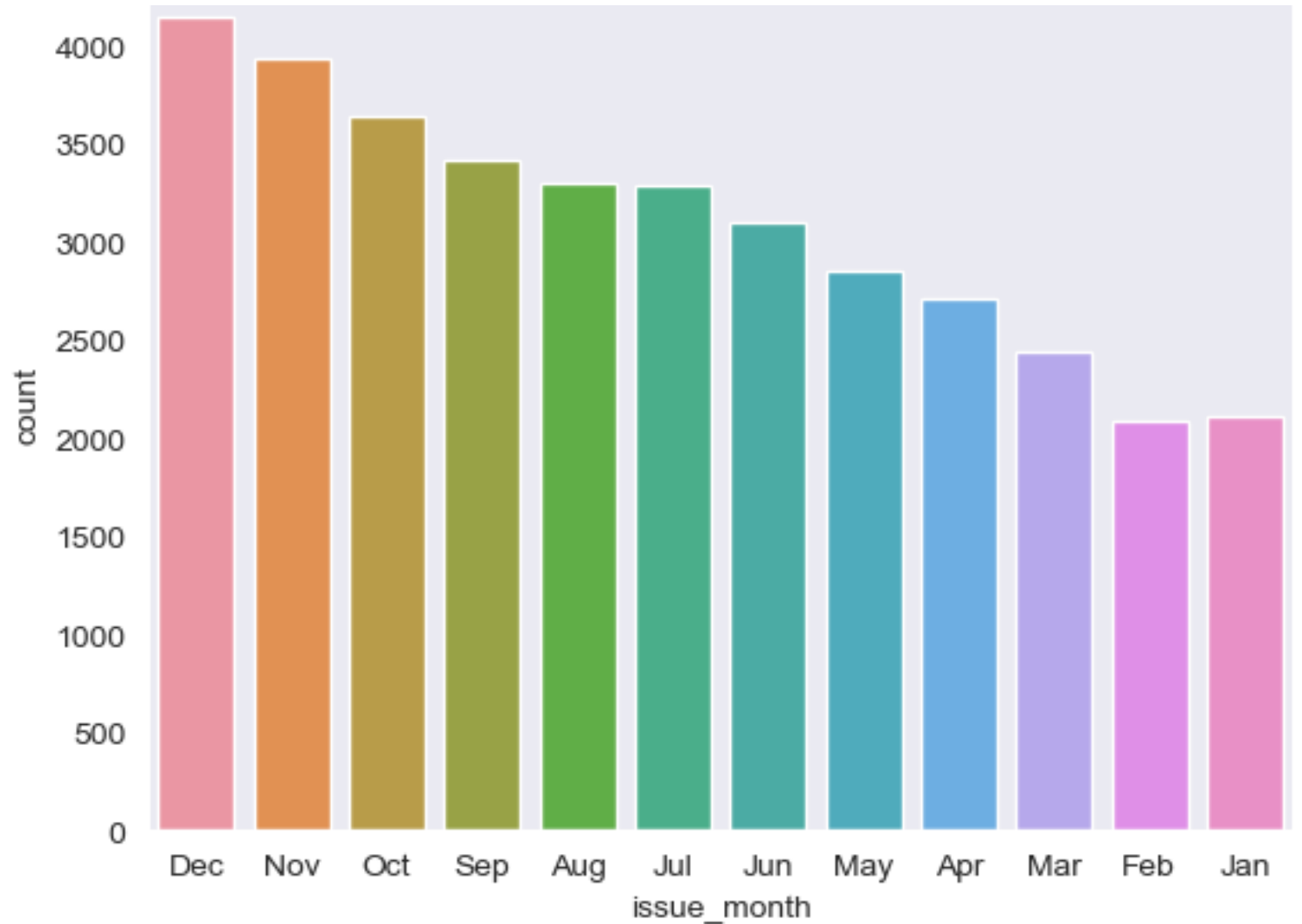
- maximum applicants have a Loan term of 36 months.
- This represents high number of applicants as not verified. This is followed by the count of source verified and verified are nearly same.



# Univariate Analysis: Issue of Loan(Month)

---

Maximum number of  
loans were issued in  
December





# Segmented Analysis

1

Break down loan amount data in 0k-7k, 7k-14k, 14k-21k, 21k-28k, 28k +

2

Break down annual income data 0k-2k, 20k-40k, 40k-60k, 60k-80k, 80k+

3

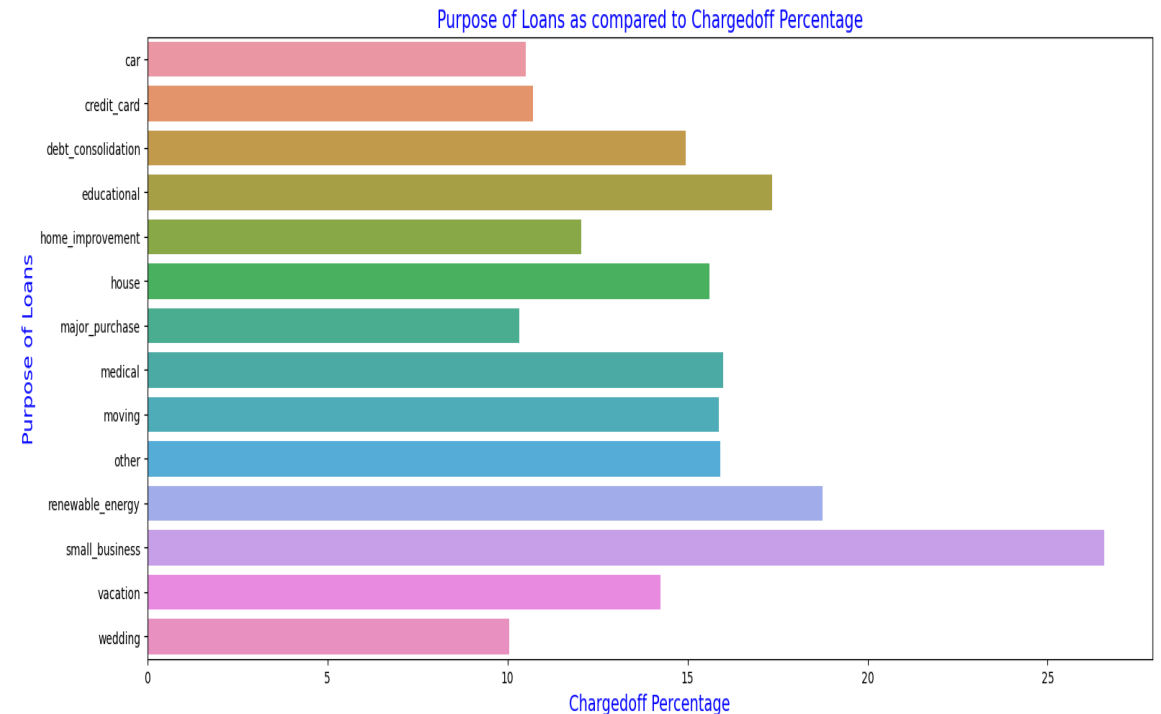
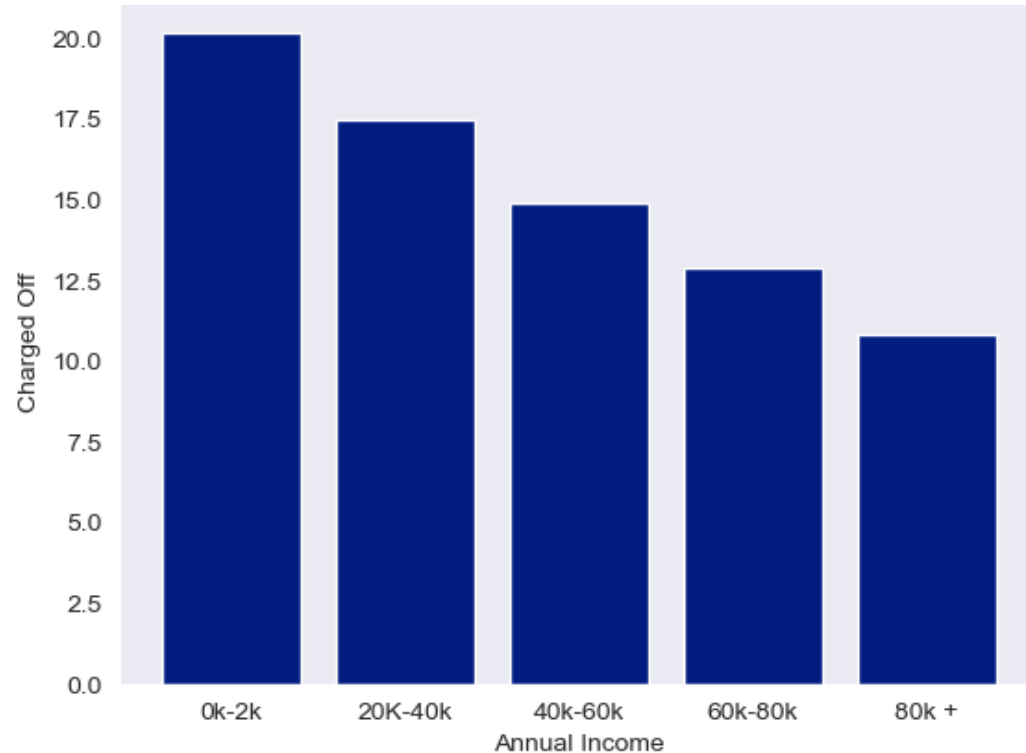
Break down interest rate in group 0-10, 10-13, 13-16, 16+

4

Break down debt to income ratio in groups 0-5, 5-10, 10-15, 15-20, 20+

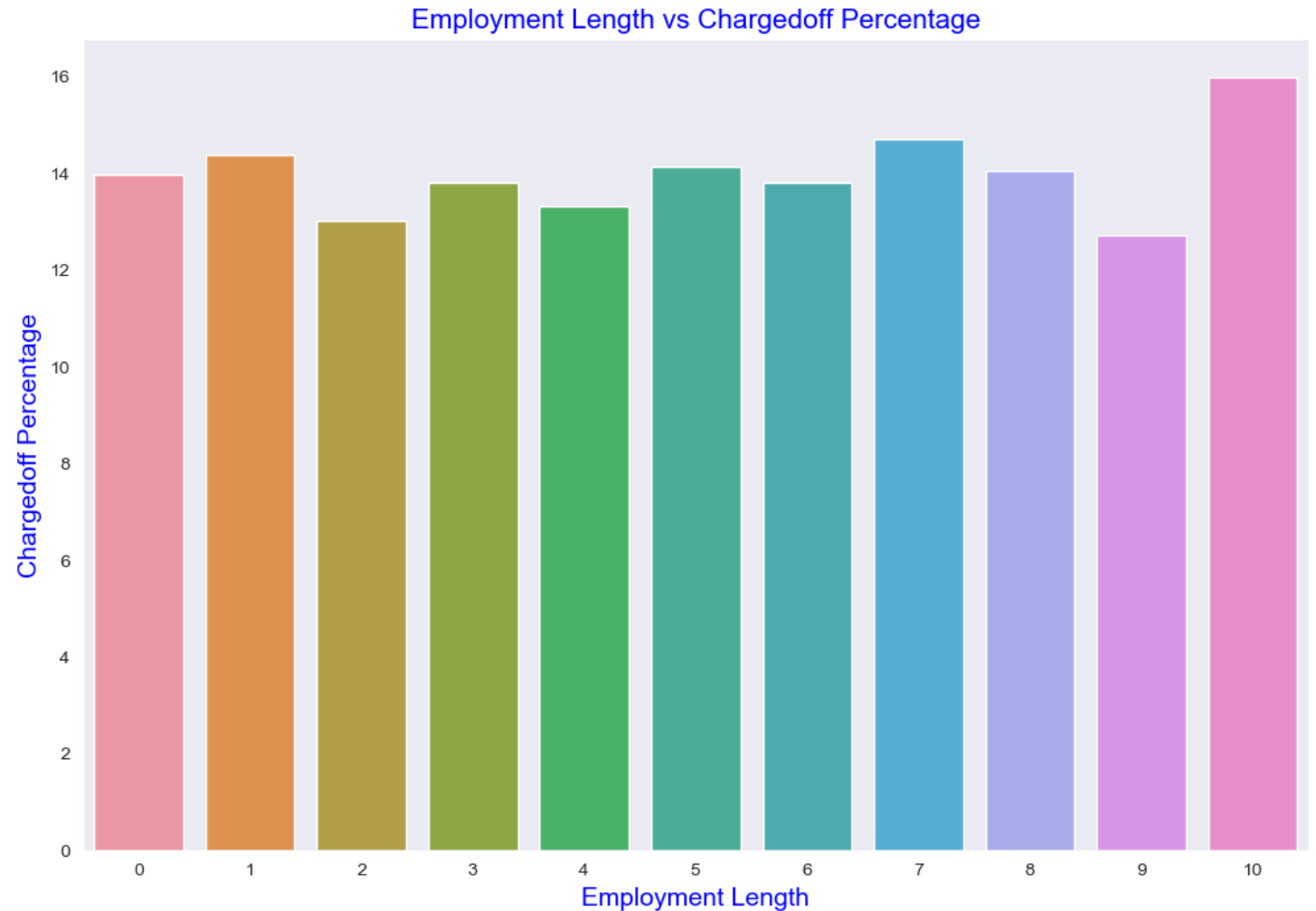
# Bivariate Analysis: Annual Inc & Purpose of Loan vs Charged Off Percentage

- Income range 0-20000 has high chances of charged off.
- Income 80000+ has less chances of charged off.
- Notice that with increase in annual income, the charged off percentage is decreasing.
- Small Business applicants have high chances of getting charged off.
- Major purchase has charged off percentage is better compared to other categories.



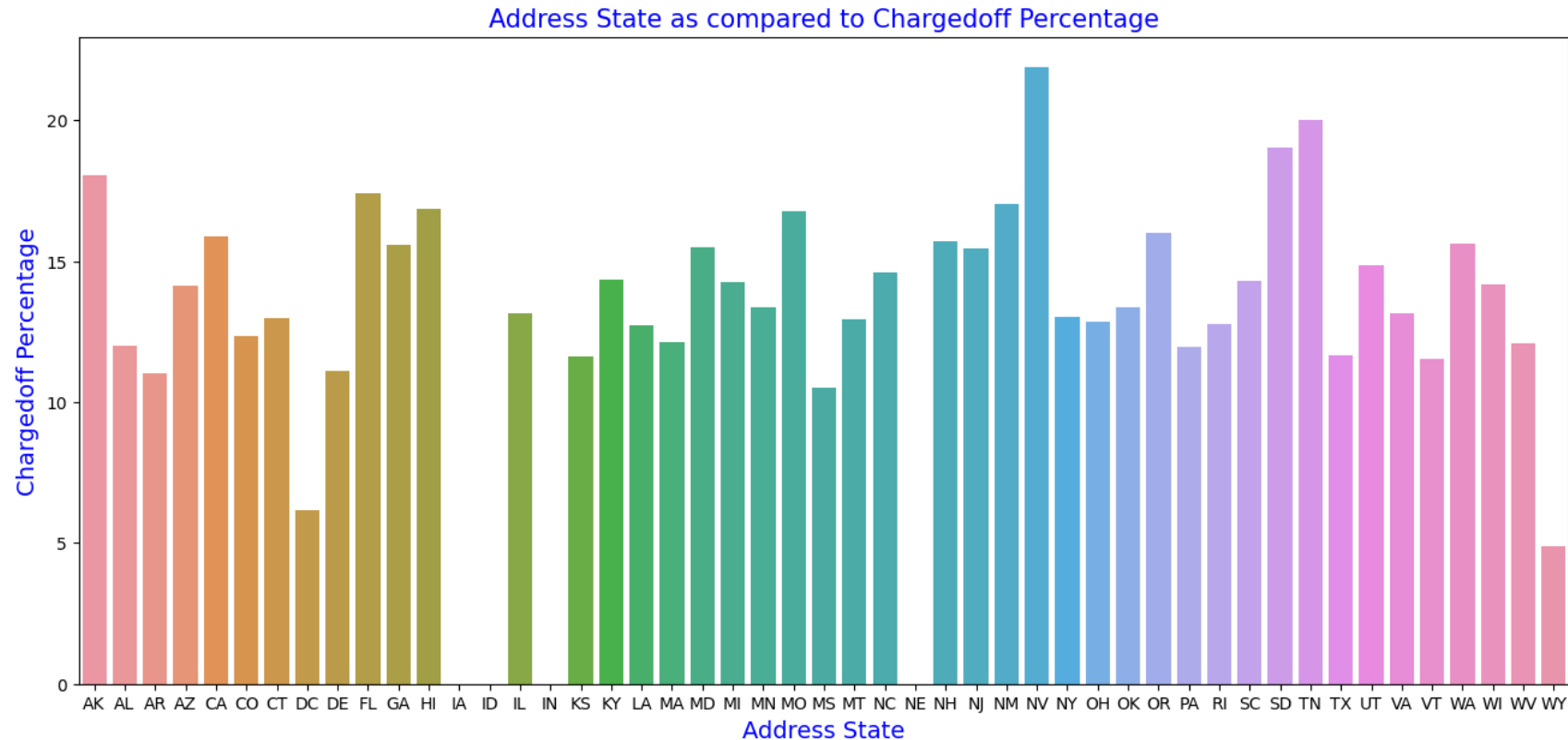
## Bivariate Analysis: Employment Length vs Charged Off

- Those who have employment length more than 10 years have the highest chance of charge off.
- People with employment might have higher debts leading to higher rates of charge off
- Those with work experience less than 1 year have high chances of getting charged off.
- It makes sense as with less or no experience they don't have source of income to repay loan.



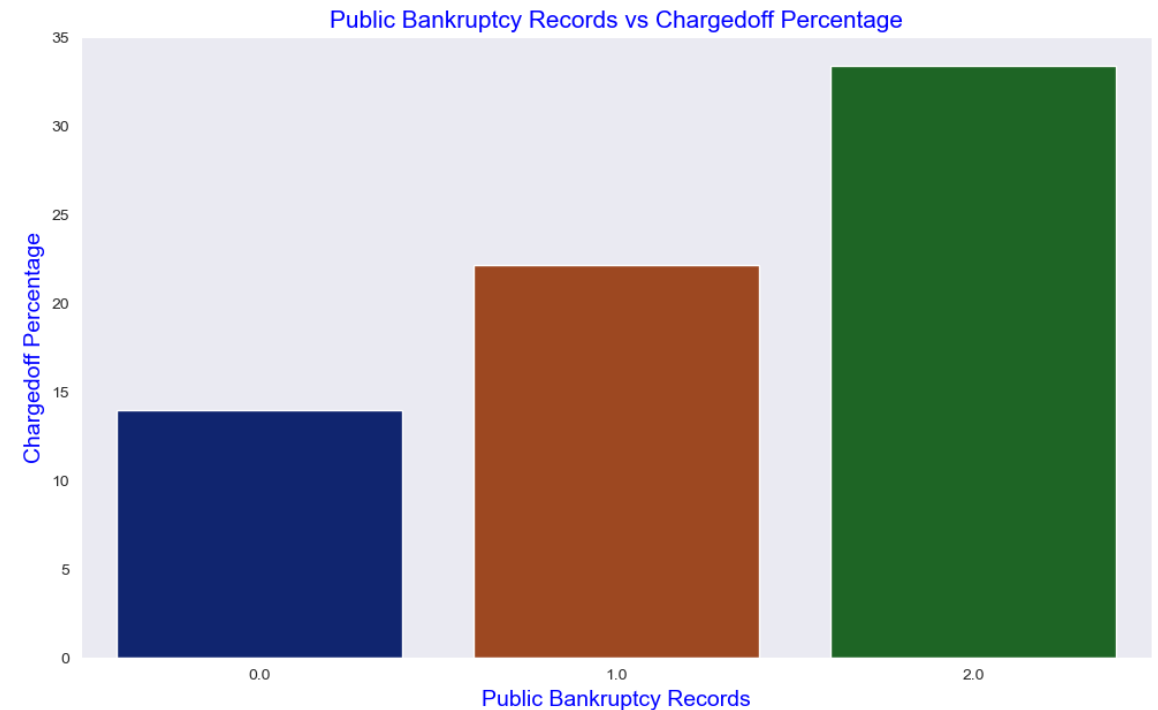
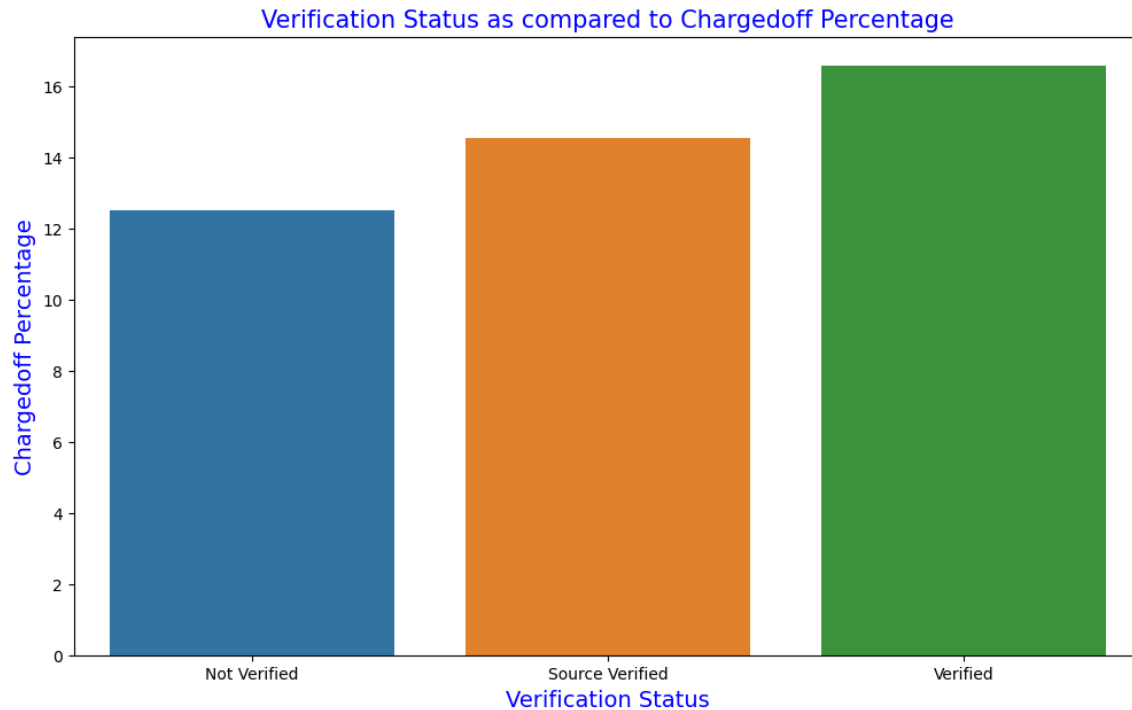
# Bivariate Analysis: Address state vs Charged Off Percentage

- The State Address 'NV' has the highest rate of charge off.
- This is followed by 'TN' and 'SD' respectively.
- Seems like 'IA' and 'IN' has no charge off(maybe there is no data from that place)
- 'WY' and 'DC' has the lowest rate of charge off.



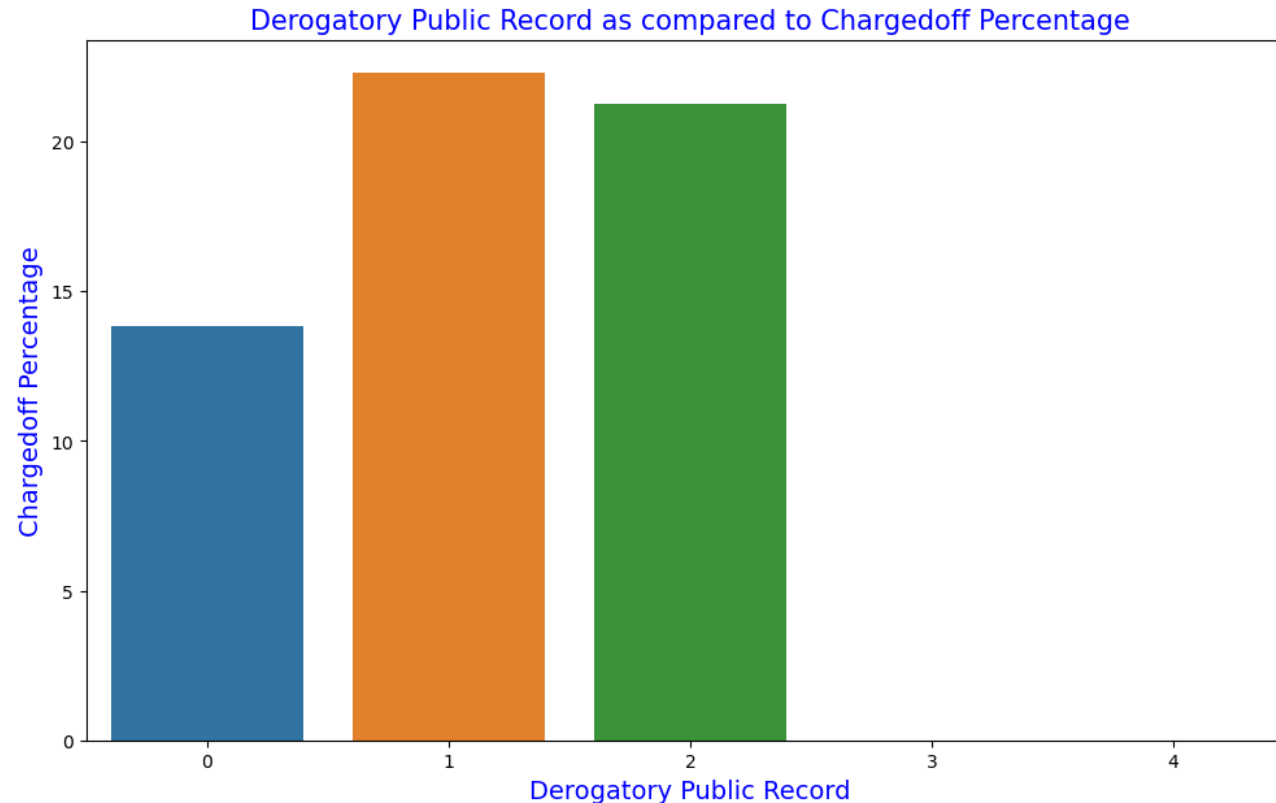
# Bivariate Analysis: Public bankruptcy records & Verification vs Charged off percentage

- The highest rate of charge off is with Bankruptcies as 2.
- Charge off rate is directly proportional to the rise in Public bankruptcies records.
- There is not much difference in charged off proportion for verification status



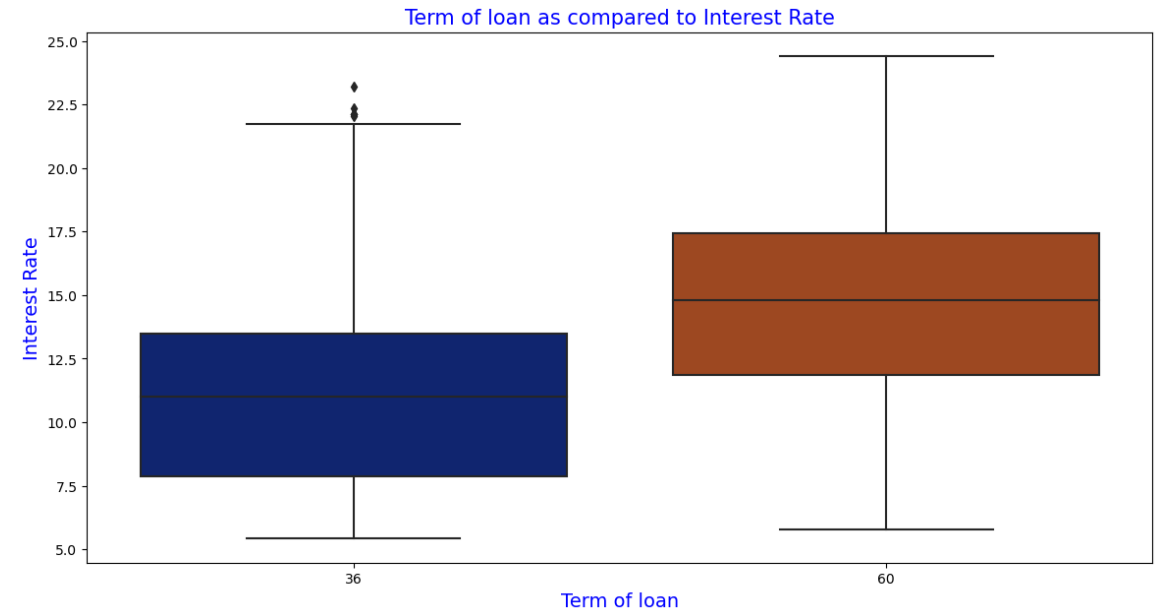
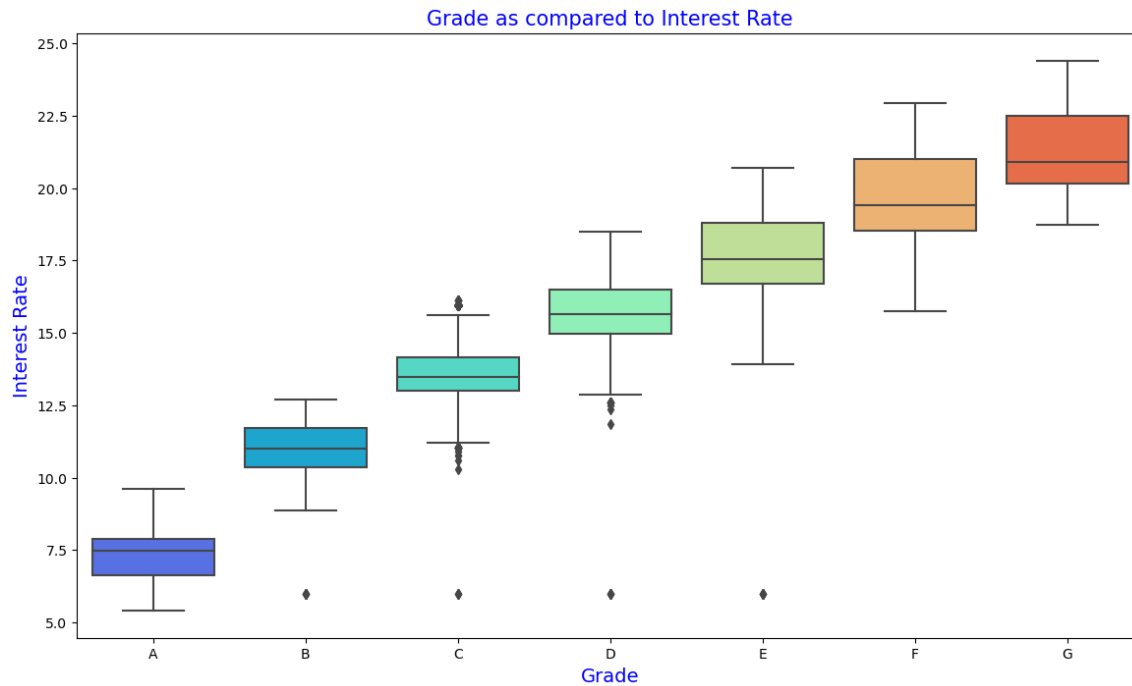
# Bivariate Analysis: Derogatory public record vs Charged off percentage

- The highest rate of charge off is with Bankruptcies as 2.
- Charge off rate is directly proportional to the rise in Public bankruptcies records.



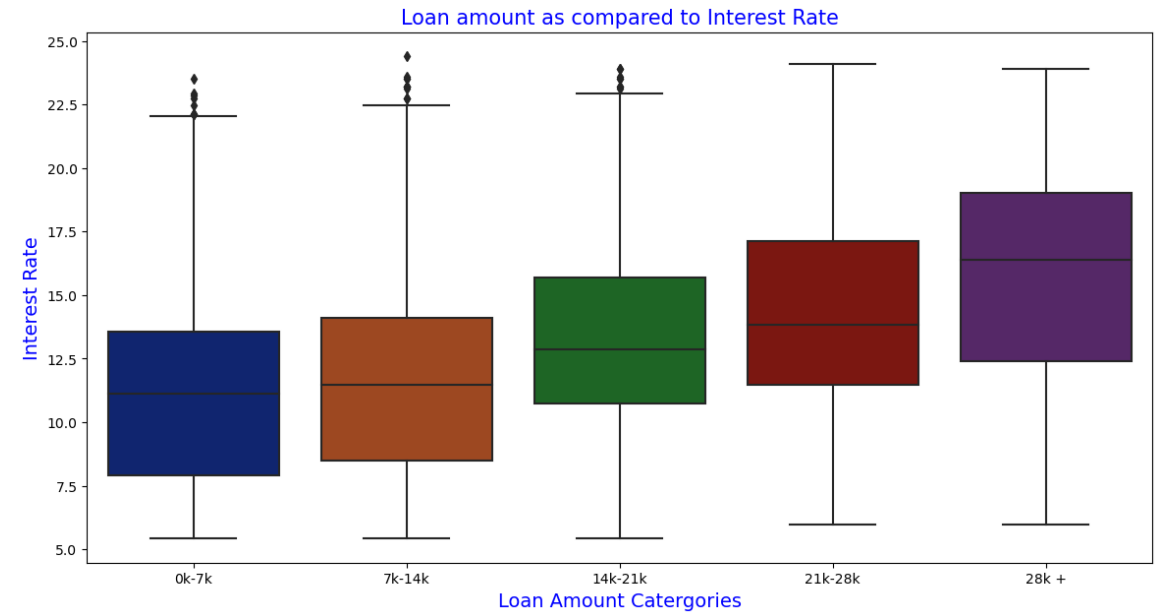
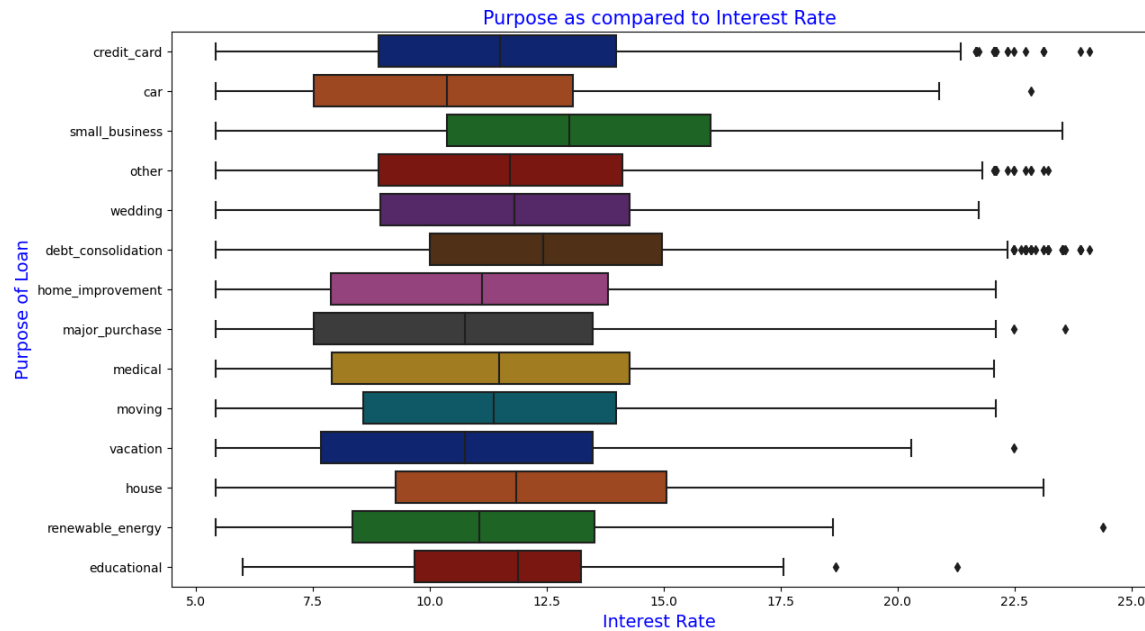
# Term of loan & Grade vs Interest Rate

- Average interest rate is higher for loan term of 60 months.
- Median of applicants with term 36 months pay less, around 11% compared to median of ones with 60 months term of around 15% .
- -Grade A has the lowest interest rates.  
-Grade G has the highest interest rates.  
-Interest rates are increments in proportion to growth of grades in ascending order



# Purpose of loan & Loan Amount vs Interest Rate

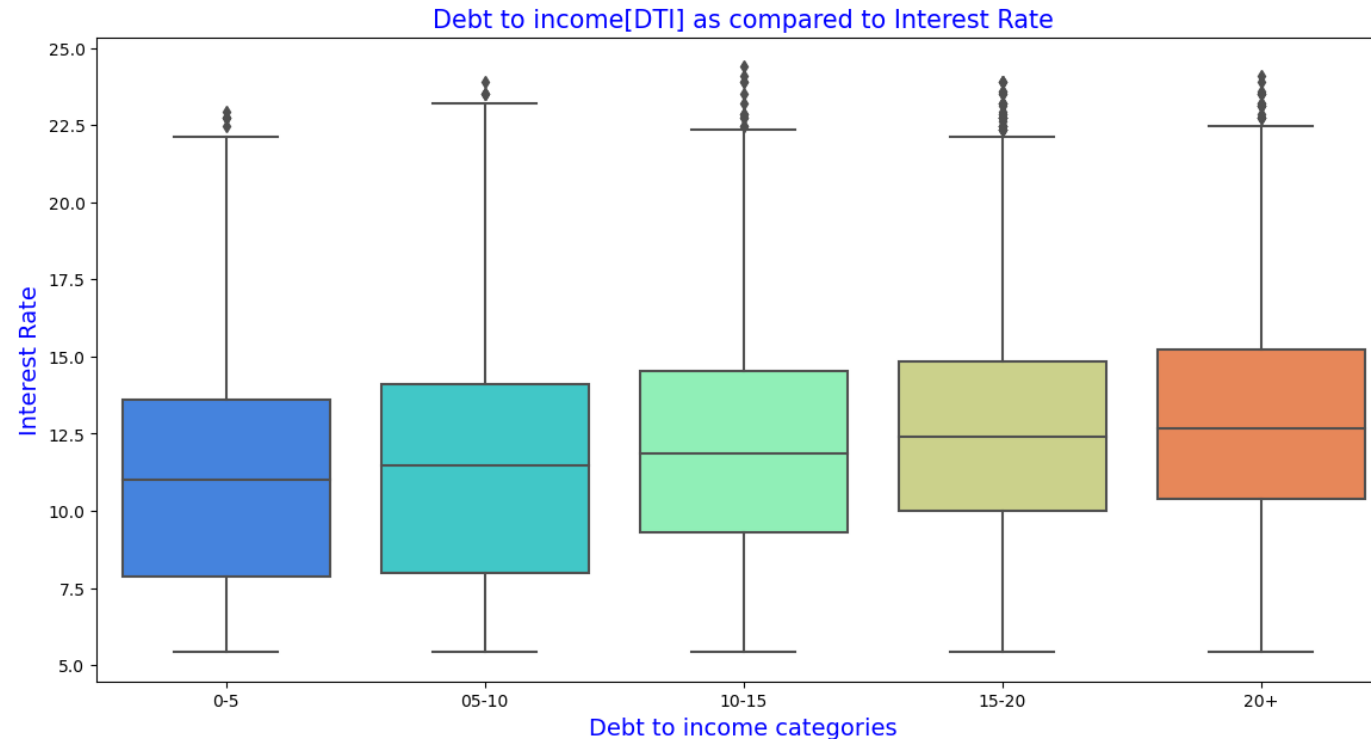
- -Increase in Loan amount is directly proportional to increase in interest rates.
- Interest rate is highest for loan amounting to 28K+, whereas it's the lowest for loan amount below 14K[0-7K, 7K-14K]
- Average interest rate is highest for small business purpose. -Loans taken for small business purposes had to repay the loan with more interest rate as compared to other.
- Debt consolidation is 2nd where borrowers had to pay more interest rate.





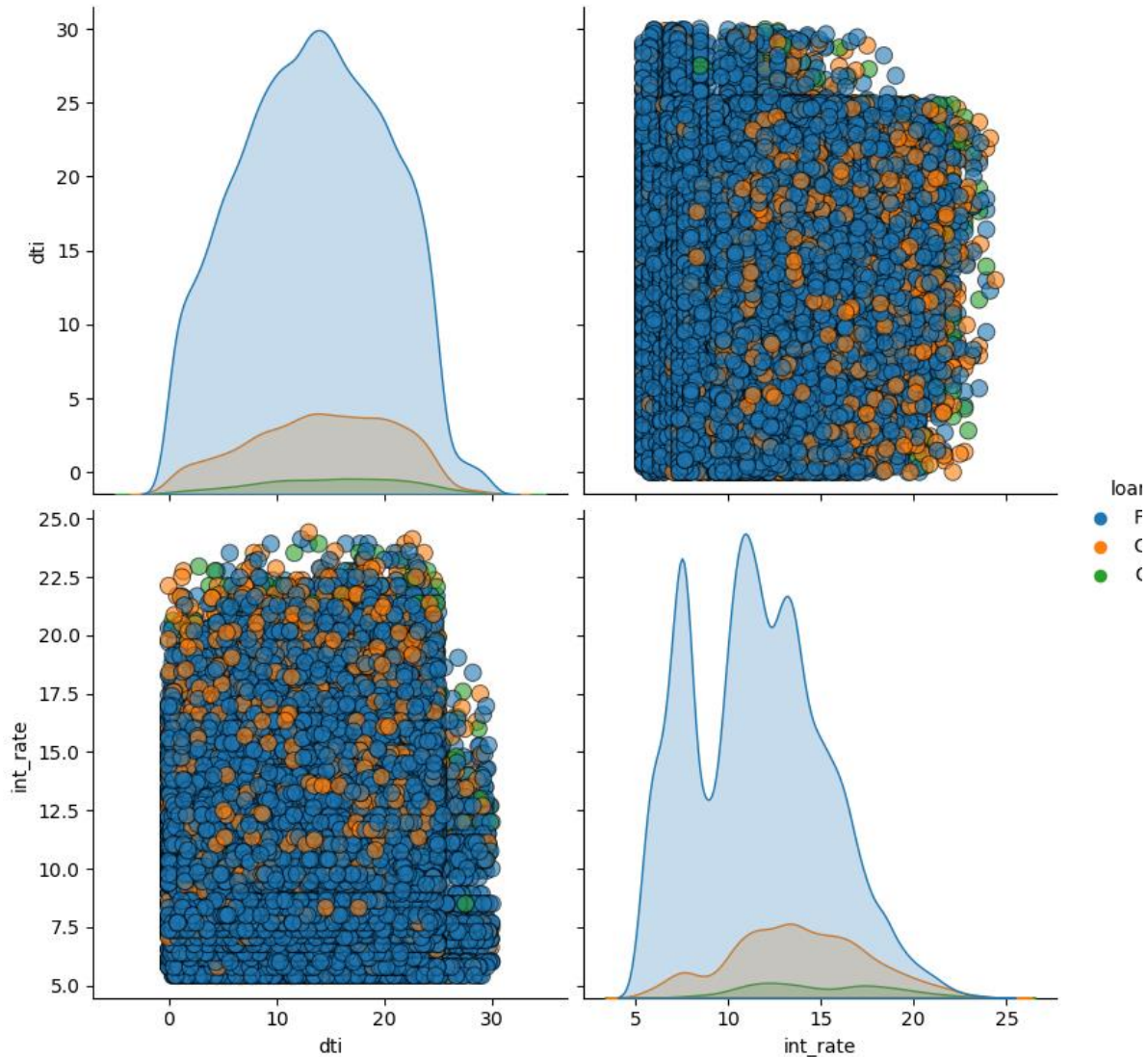
# Debt to income[DTI] vs Interest Rate

There seems to be no or negligible difference in interest rates on loan as compared to DTI

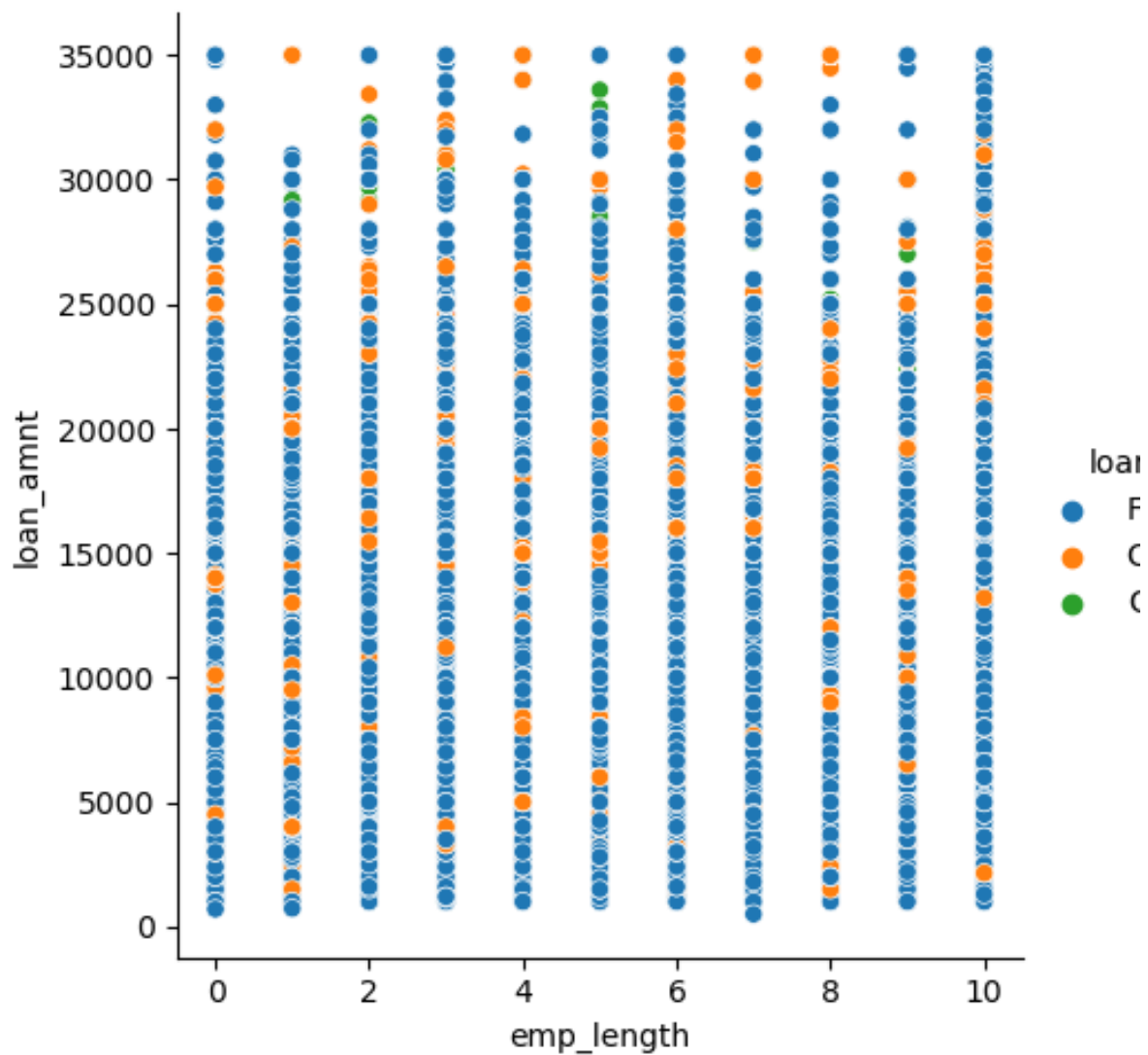


# Multivariate Analysis: DTI, Interest Rate, Loan Status

---



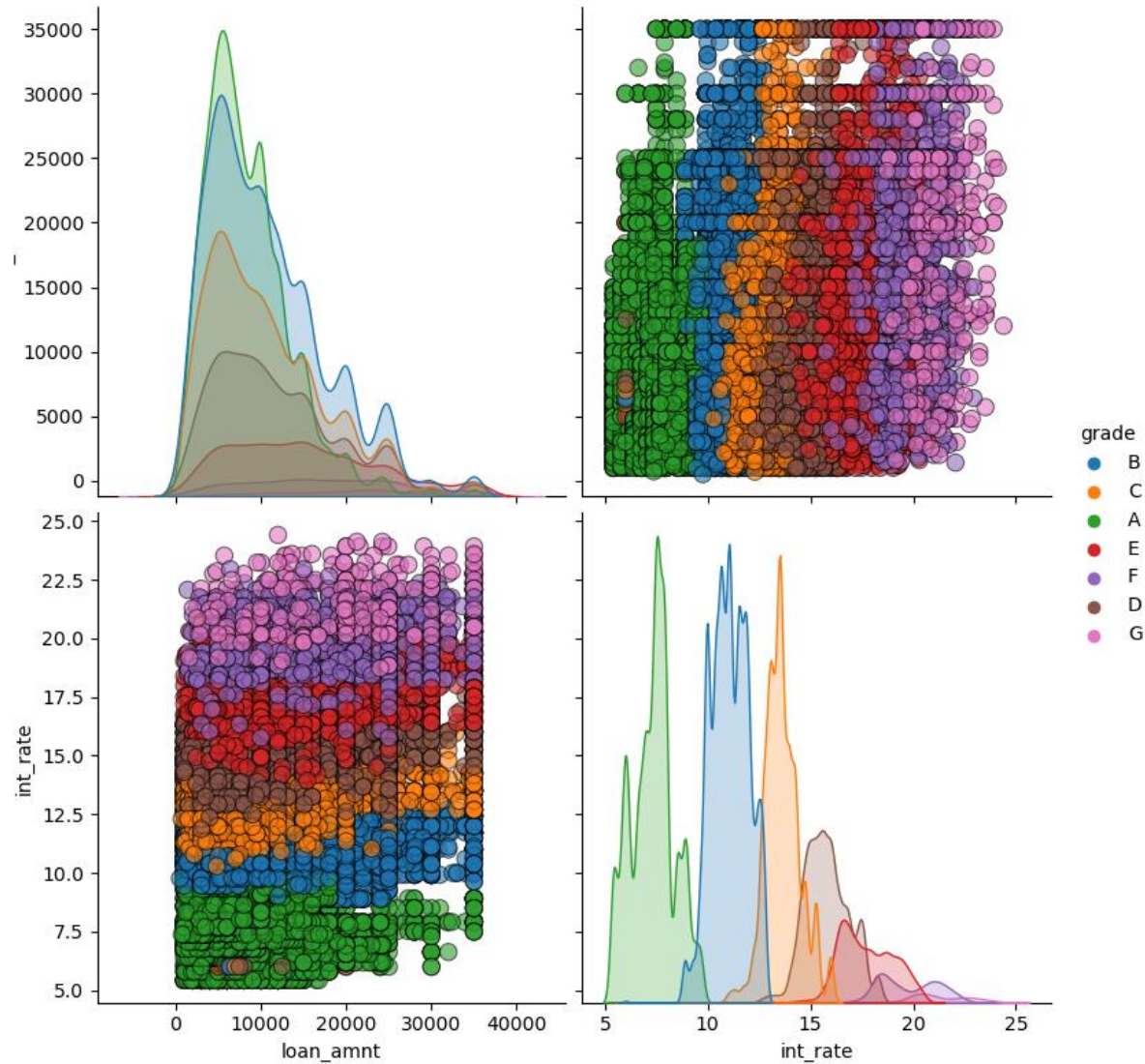
- High number of charge off can be related to high number of interest rate. (Specially above 16% holds most charged off.)



## Multivariate Analysis: Loan Amount, Employment Length, Loan Status

---

- High loan amount combined with high employment length is high risk of default



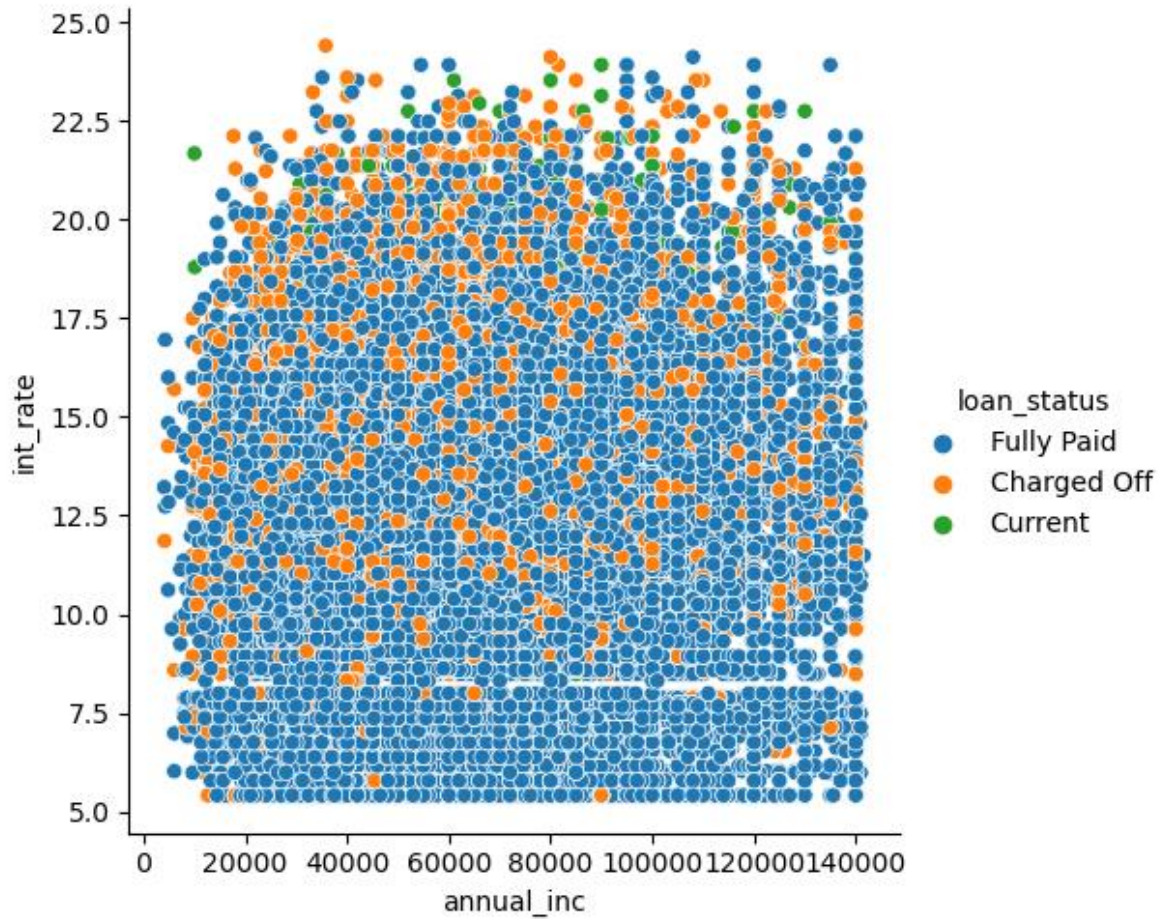
## Multivariate Analysis: Loan Amount, Loan Grade, Interest Rate

- Seems like Grade are directly related to high interest amount. We can see as rise with the grade in ascending order, there is rise orderly rise in interest rate

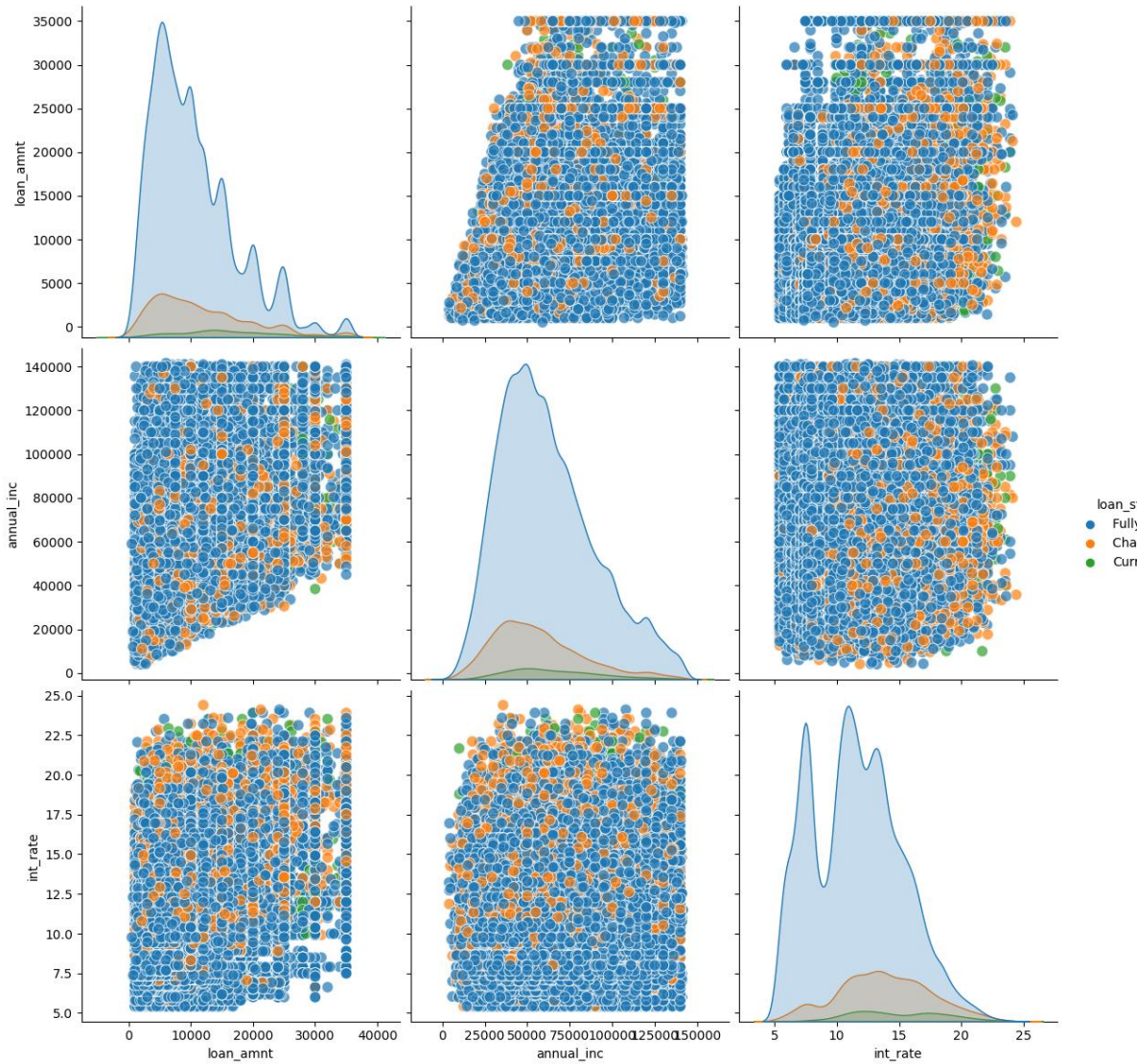


## Multivariate Analysis: Annual Income, Interest Rate, Loan Status

---

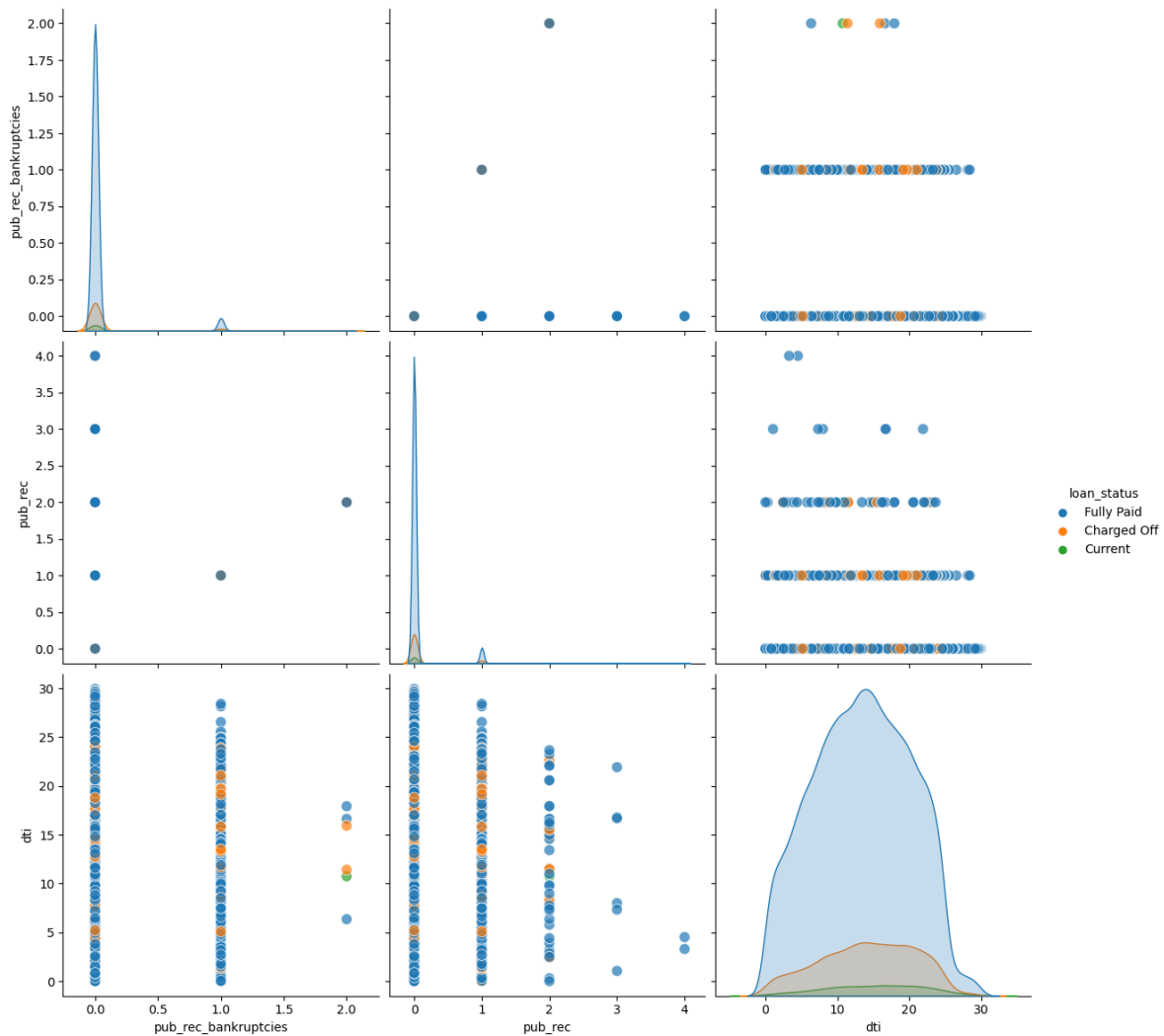


- Charged off Applicant seems to be high for applicants with less salary band (specifically below 90000)



## Multivariate Analysis: Loan Amount, Annual Income, Interest Rate, Loan Status

- High interest rates (16+%) and high loan amount is related to charge off.
- Charge off is more for applicants with annual income < 90000.



## Multivariate Analysis: Public Recorded bankruptcies, Public Record, Interest Rate, DTI, Loan Status

- Charge off is high for applicants with bankruptcies and derogatory public records than ones without the records.





## Conclusion

- People with employment length more than 10 years or less than 1 year are likely to default
- Lower income group(up to 20000) is likely to default.
- Small business loan are likely to default
- People with public reported bankruptcies are highly likely to default