Lending Club Case Study

Exploratory data analysis.

Data Cleaning

1.Few feature in loan dataset has most of the values to be null.

These features removed for analysis

2. For our analysis loan_status ,only full paid and charged off considered as we analysis only default and non default applicant,current paying loan ampont is not considered .hence its filtered.

Univariate Quantitative(Numeric) feature analysis

- Interest Rate variable

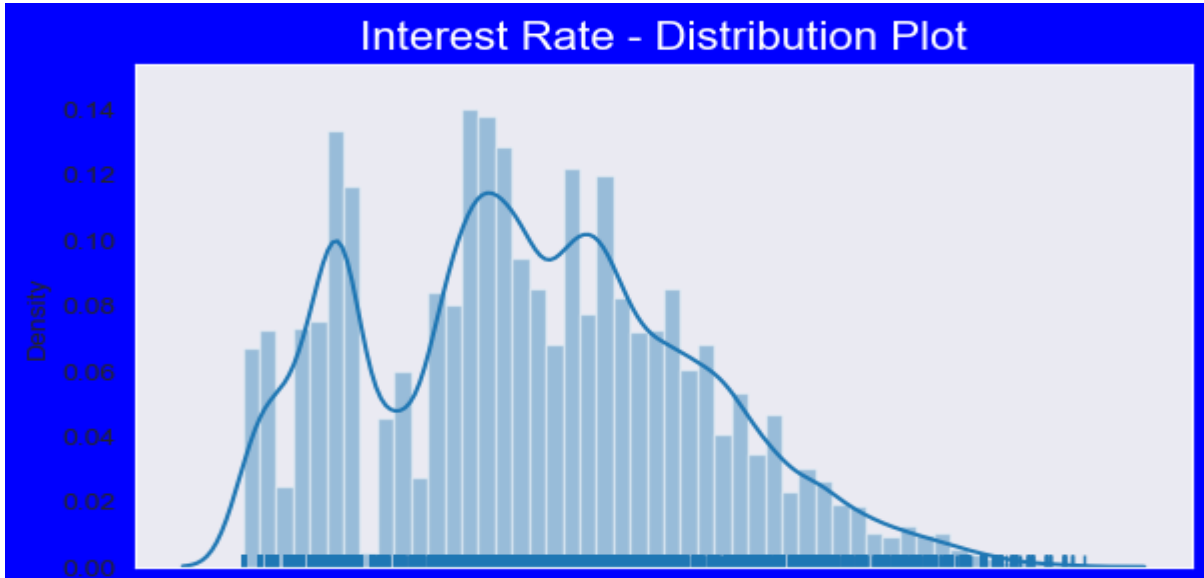
Observations :32% of loan amount taken by applicant are in interest rate 0f 0-10

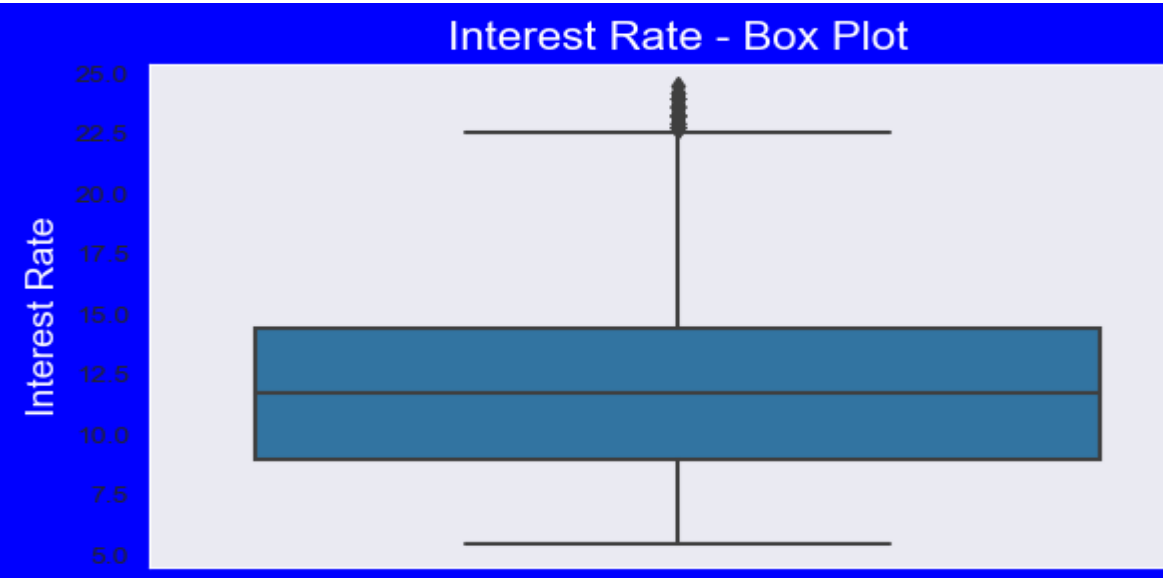
30% of loan amount taken by applicant are in interest rate 0f 12.5-16.

Below plot shows that most of interest rate on loan is 10 to 15%

maximum interest rate is 24.59%

mean interest rate is 11.93%





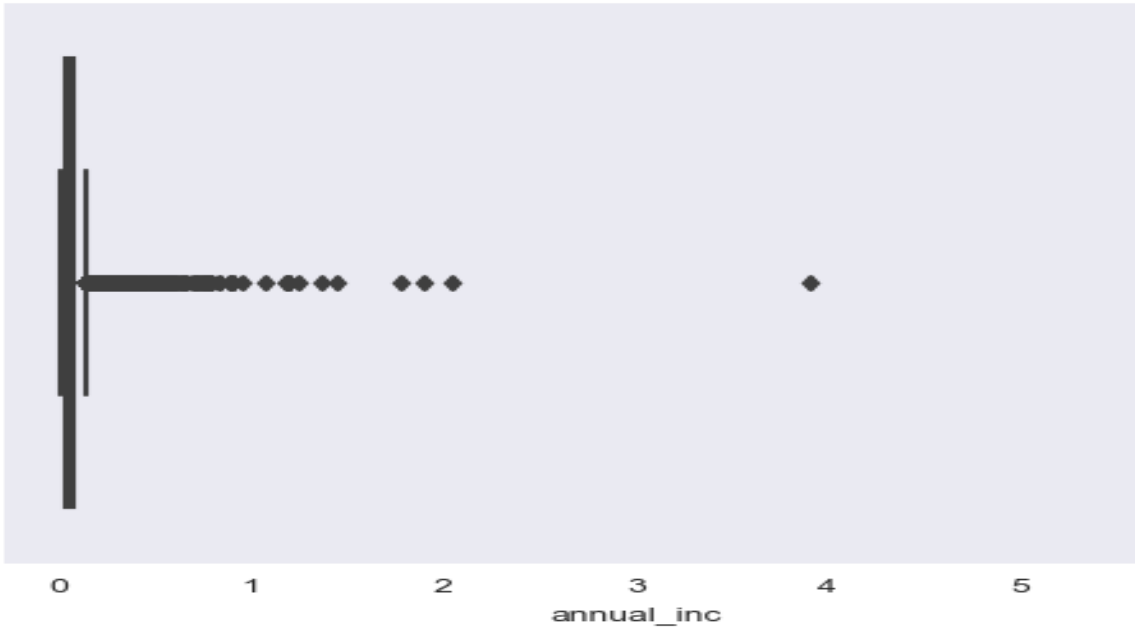
Univariate Quantitative(Numeric) feature analysis

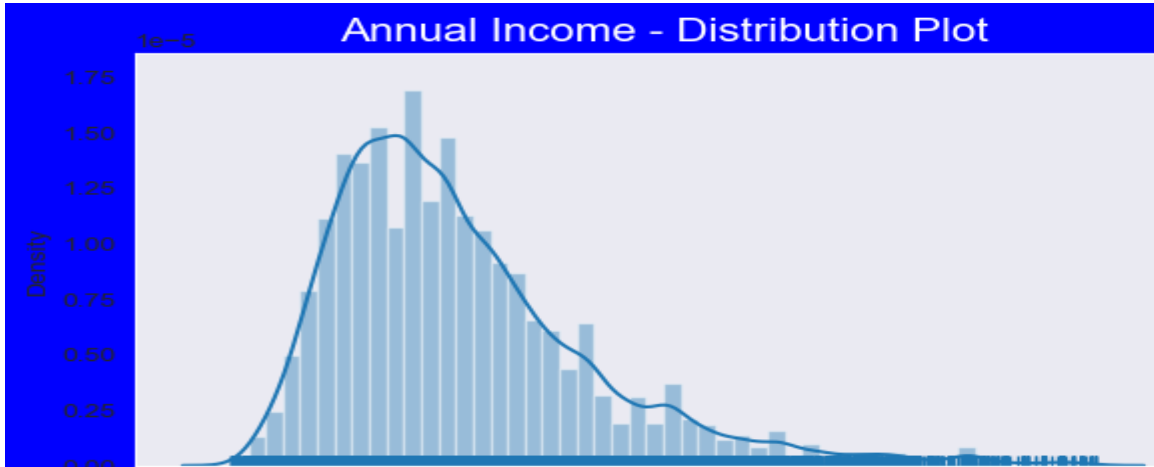
- Annual Income

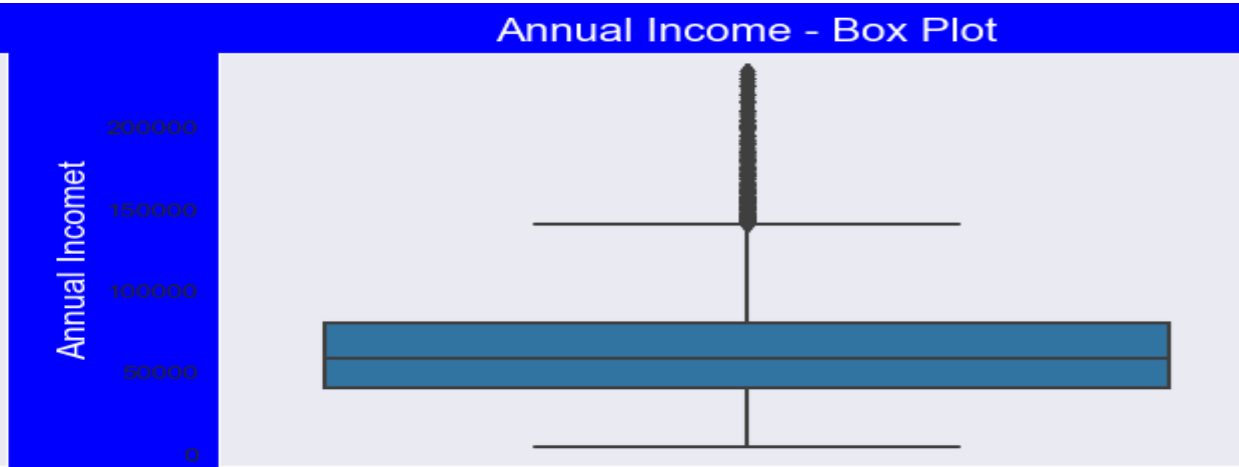
- observation:there is outier in annul income variable.
- These Outlier is filtered for further analysis.
- After filtering outliers:
- 29% applicant whose annual income is 40000-60000 has taken loan

25% applicant whose annual income is 80000+ has taken loan

- # only 3% applicant whose annual income is 0-20000 has taken loan.
- Below plot shows that annual income of whose has taken loan is range of 40000-60000

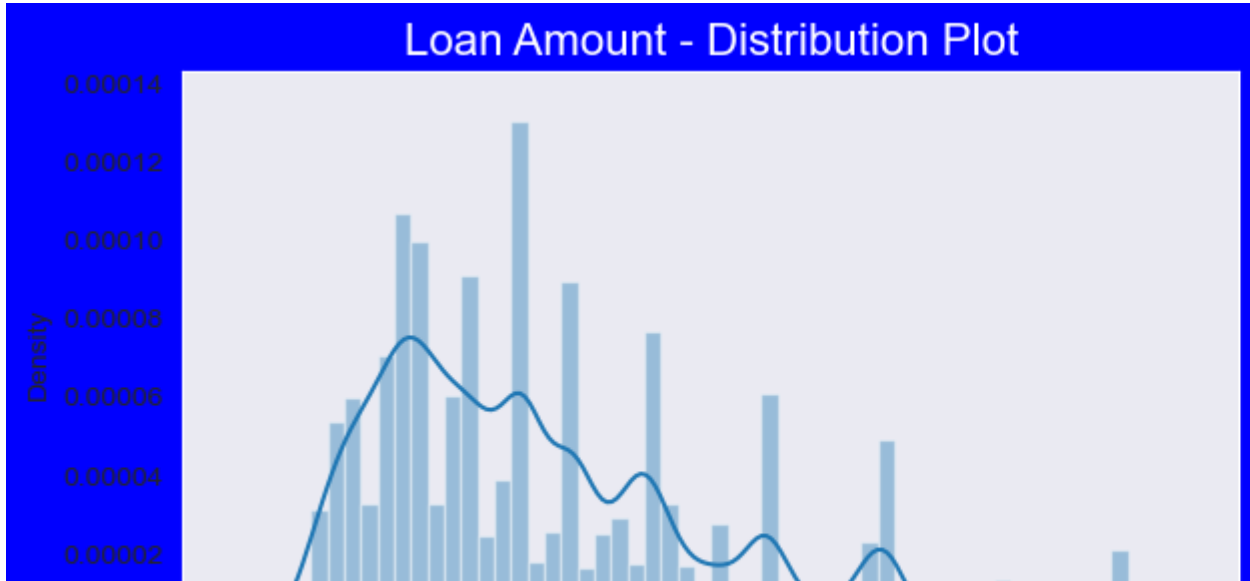


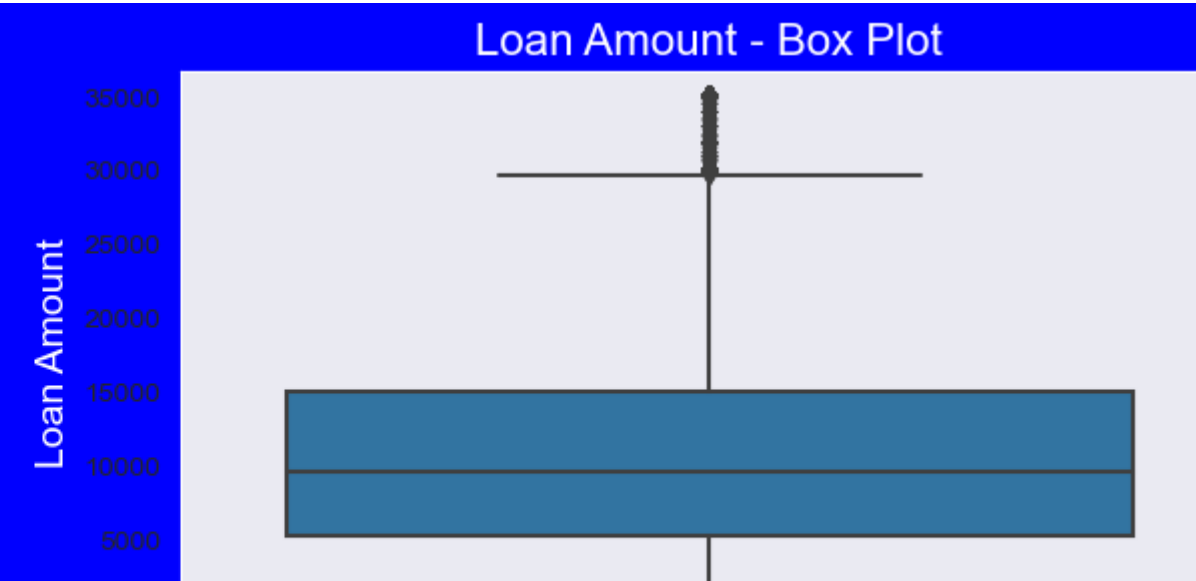




Univariate Quantitative(Numeric) feature analysis

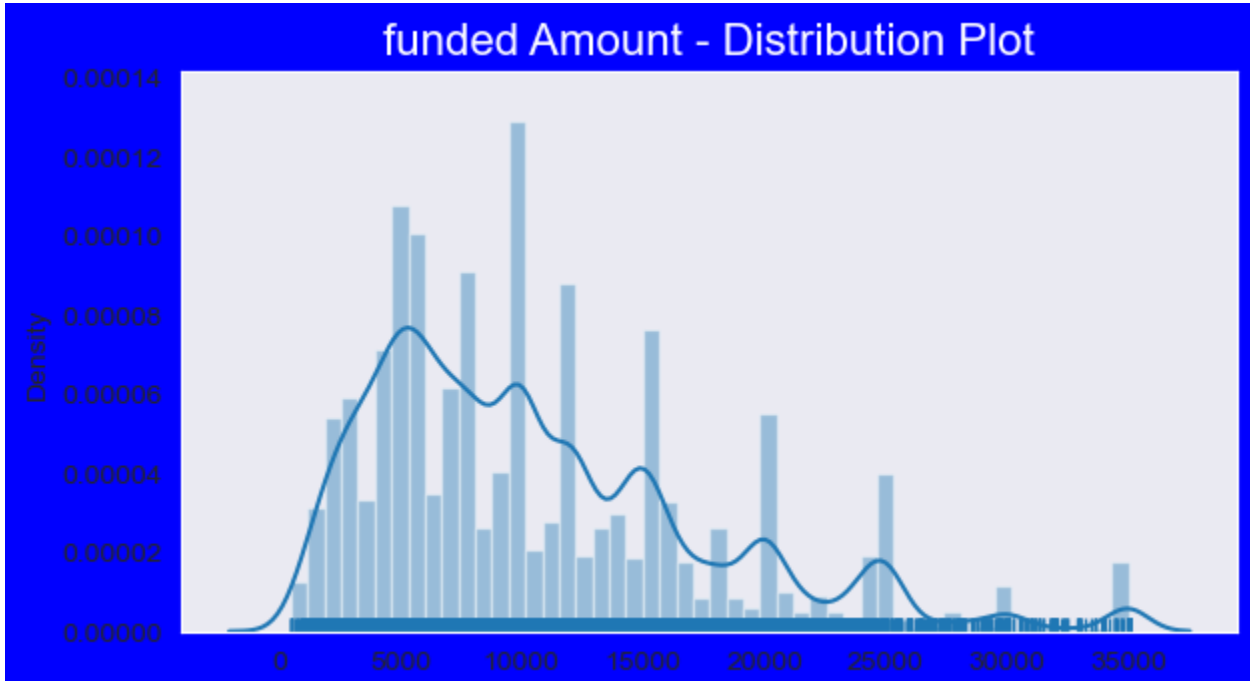
- Loan amount
  - Observations :37% of loan amount in the range of 0-7000
  - 34% of loan amount in the range of 7000-14000
  - only 2% loan amount in the range of 280000+.
  - Below plots shows that most of loan amount between 5000 -15000
  - maximum loan amount is 350000 and mean is 10962

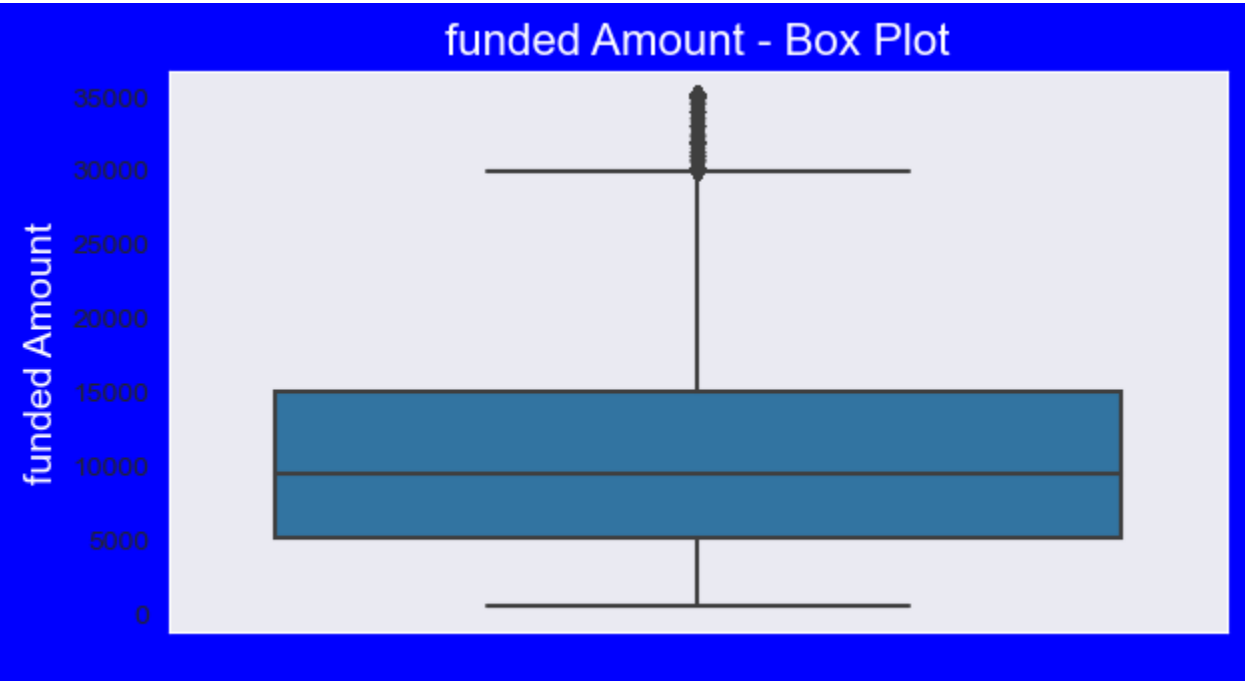




Univariate Quantitative(Numeric) feature analysis

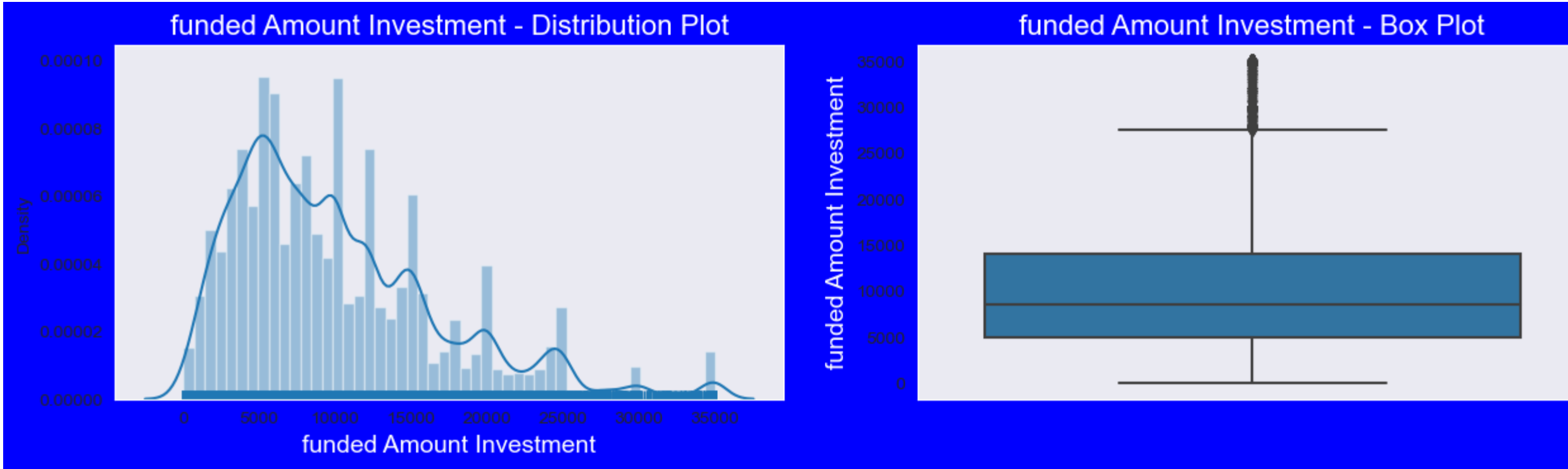
- funded_amnt
  - 37% of funded amount in the range of 0-7000
  - 34% of funded amount in the range of 7000-14000
  - only 2% funded amount in the range of 280000+
  - Below plots shows that most of funded amount is in range of 5000-15000





Univariate Quantitative(Numeric) feature analysis

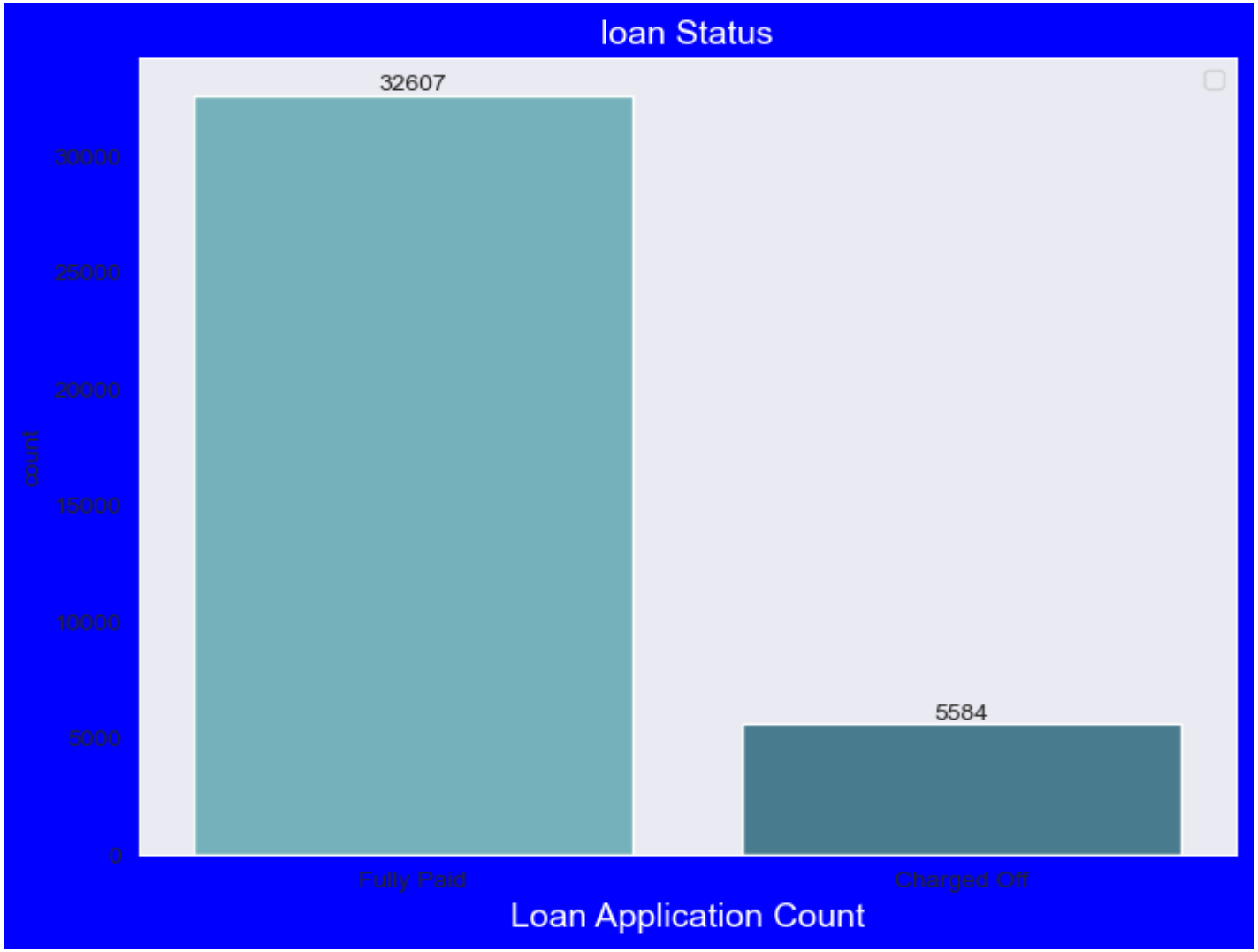
- Funded amount investment
  - 37% of funded amount in the range of 0-7000
  - 34% of funded amount in the range of 7000-14000
  - only 2% funded amount in the range of 280000+
  - Below plot shows that most of funded amount investment is in range of 5000-15000



Univariate Orderd feature analysisanalysis

- Loan Status
  - observation:85% of loans also full paid

14% loans are not paid or charged off.

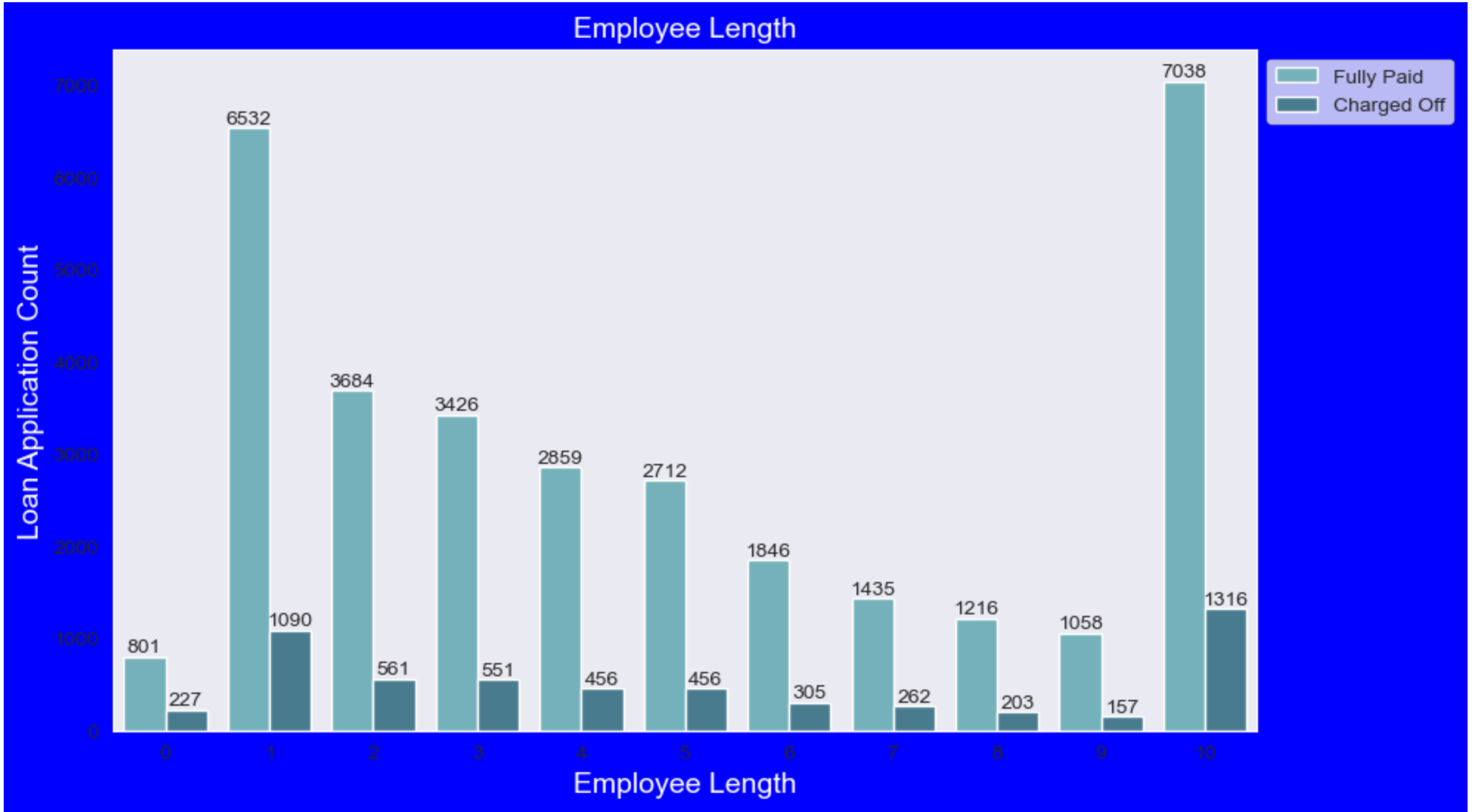


Univariate Orderd feature analysis

- Employee length

- observation:22% of people of 10+ experience have taken loan
- 19% loans taken by 1 year experienced people
- 2% loans taken by freshers or no experienced people
- 10+ year and 1 experienced applicant have most full paid the loan amount

Univariate Orderd feature analysis



Univariate Orderd feature analysis

- Term

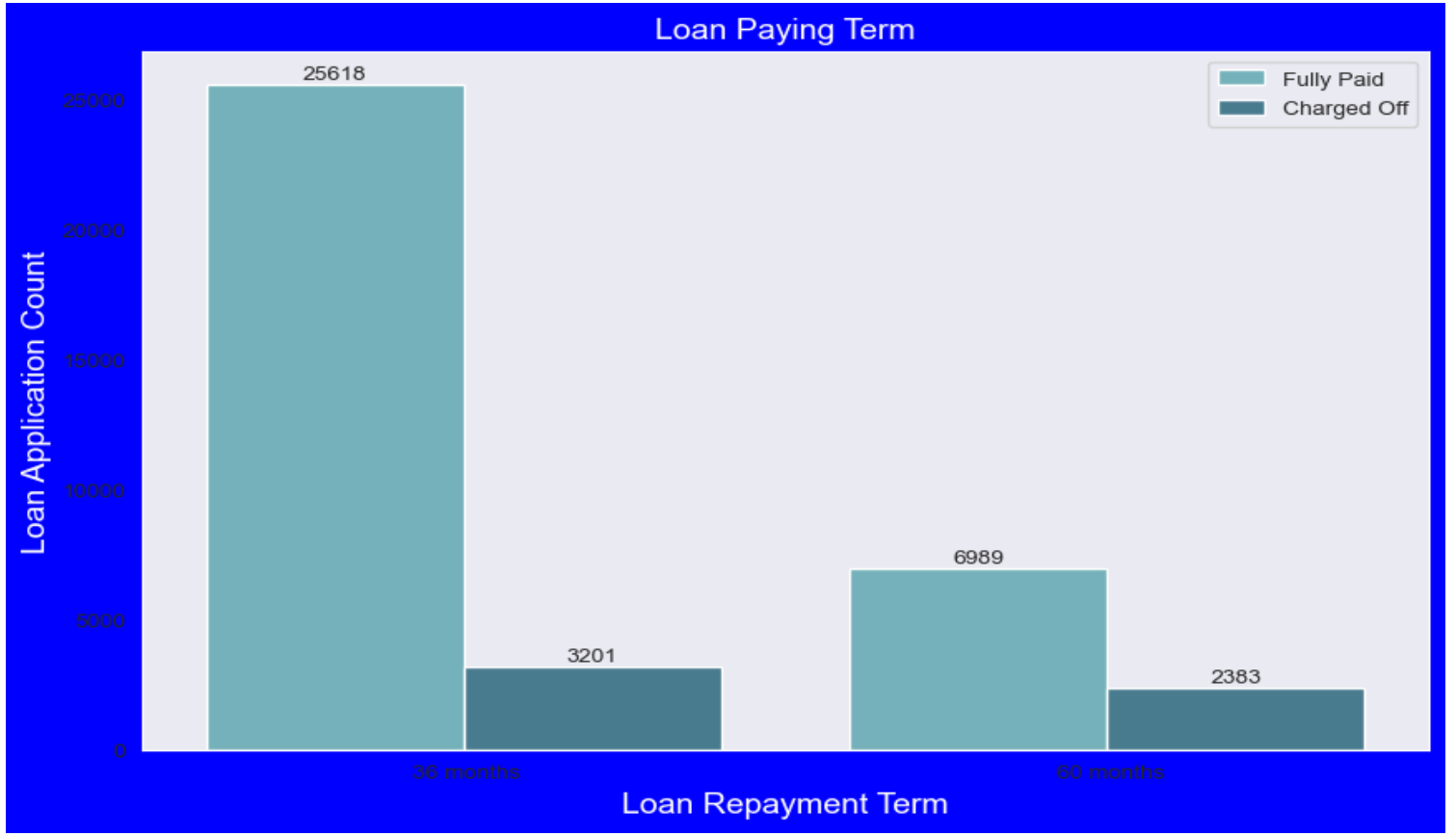
- observation:75% loans taken fro 36 months tenure
- # 25% loans taken for 60 months.
- Below plot shows that those

who had taken loan to repay in 60 months

had more % of number of applicants getting

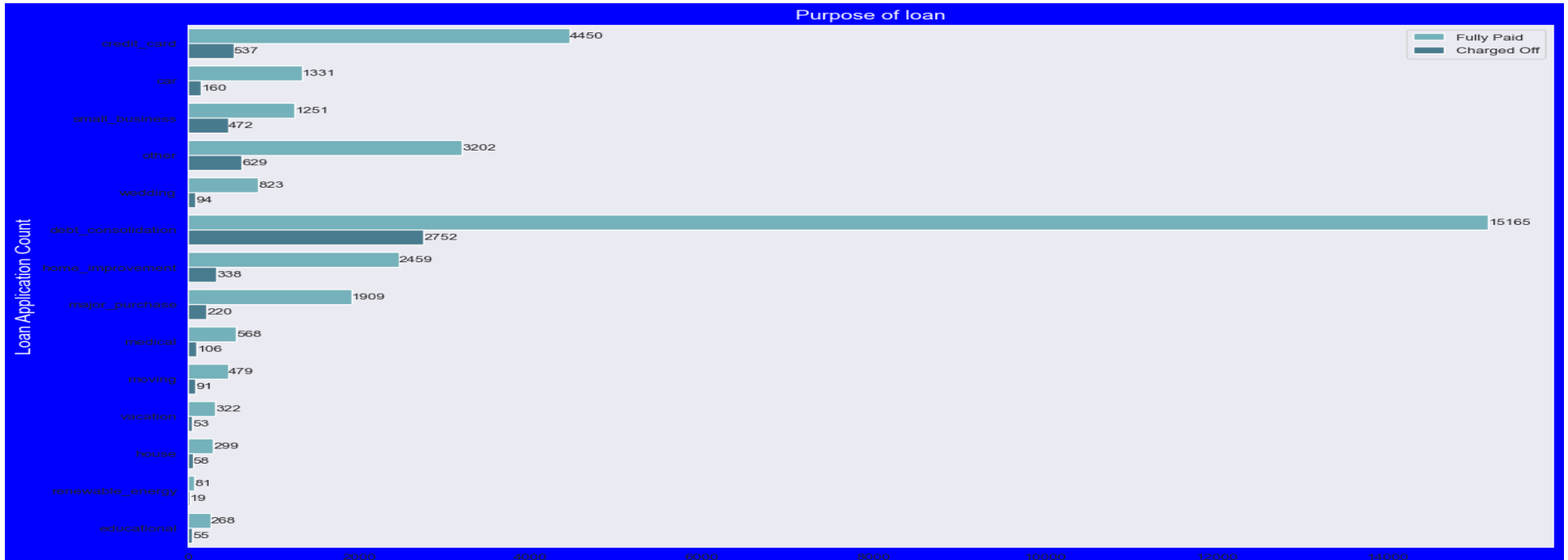
charged off as compared to applicants

who had taken loan for 36 months.



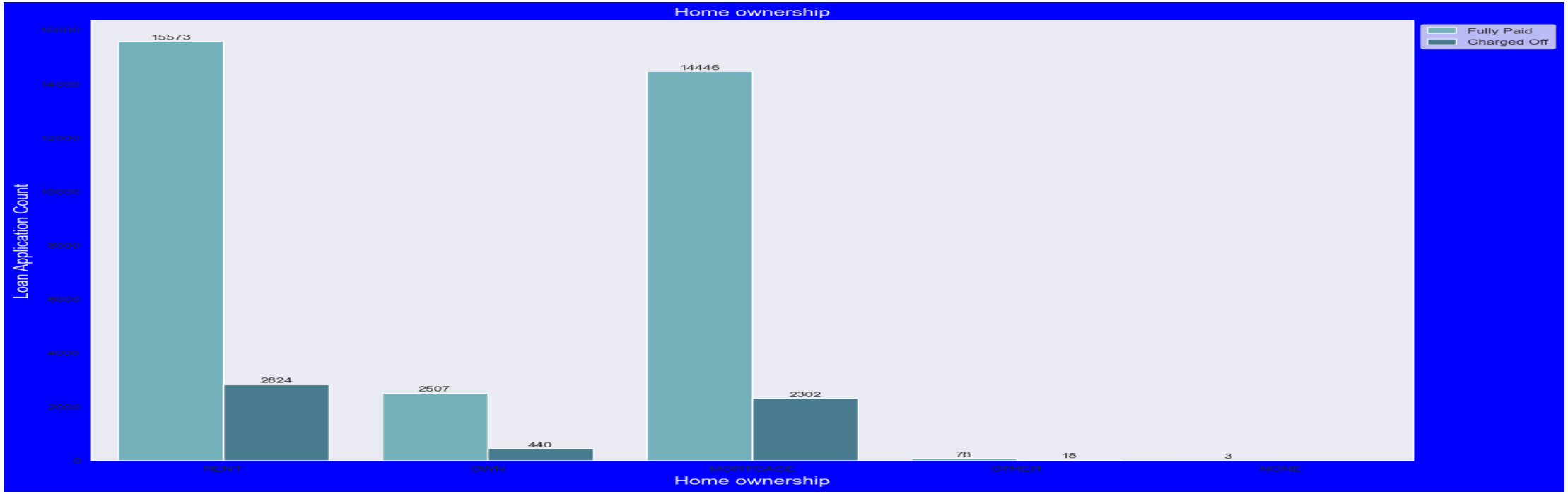
Univariate UnOrderd feature analysis

- Purpose
  - observation:46% of loans purpose is debt_consolidation
  - 12% loans purpose is credit_card
  - 10% loans taken for other purpose.
  - Below plot shows that most of the loans were taken for the purpose of debt consolidation & paying credit card bill has paid full
  - Loan amt.



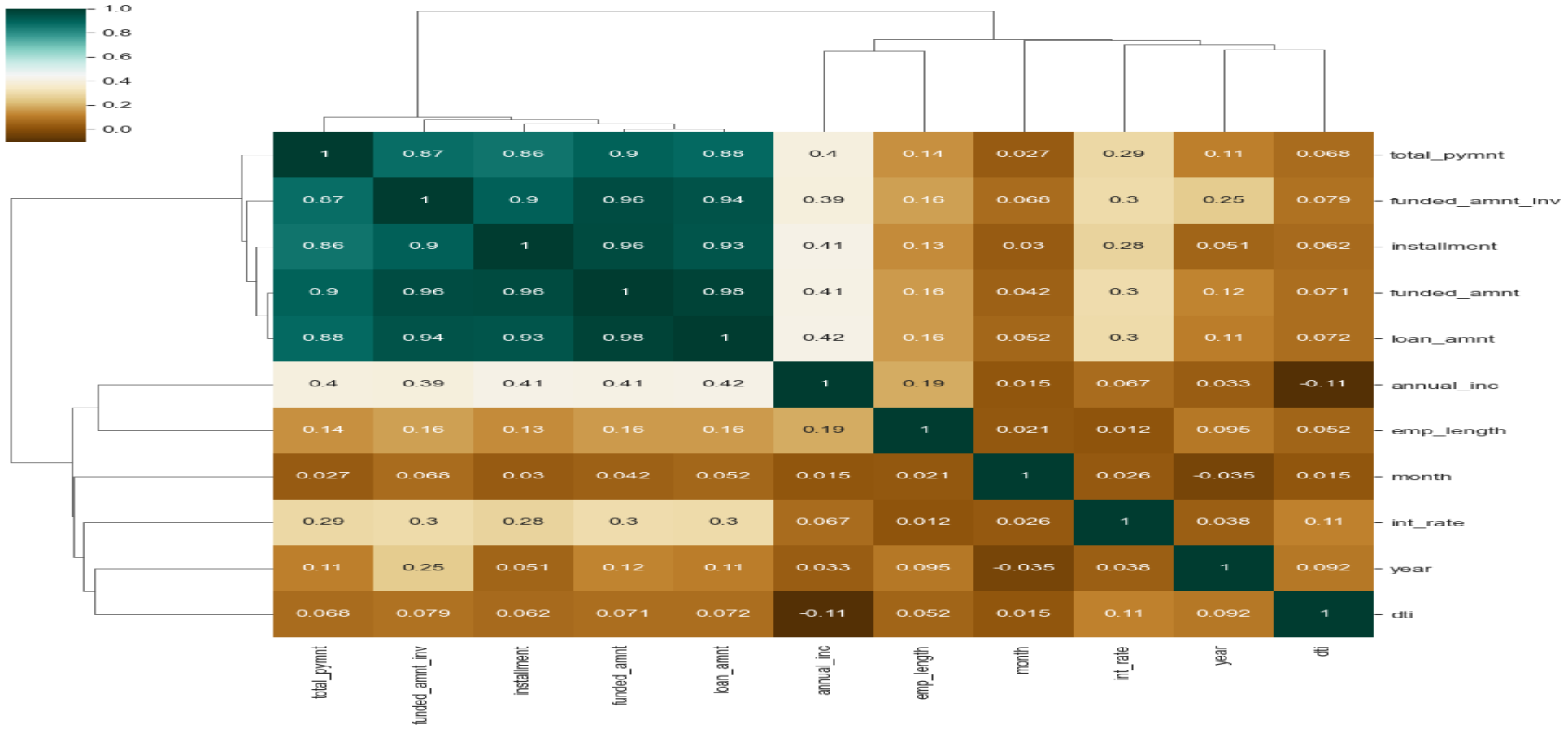
Univariate UnOrderd feature analysis

- Home Ownership:
  - observation:48% of peoples who are in rent have taken loan
  - 44% of peoples have taken loan by mortage
  - only 7% of peoples who are in own home have taken loan



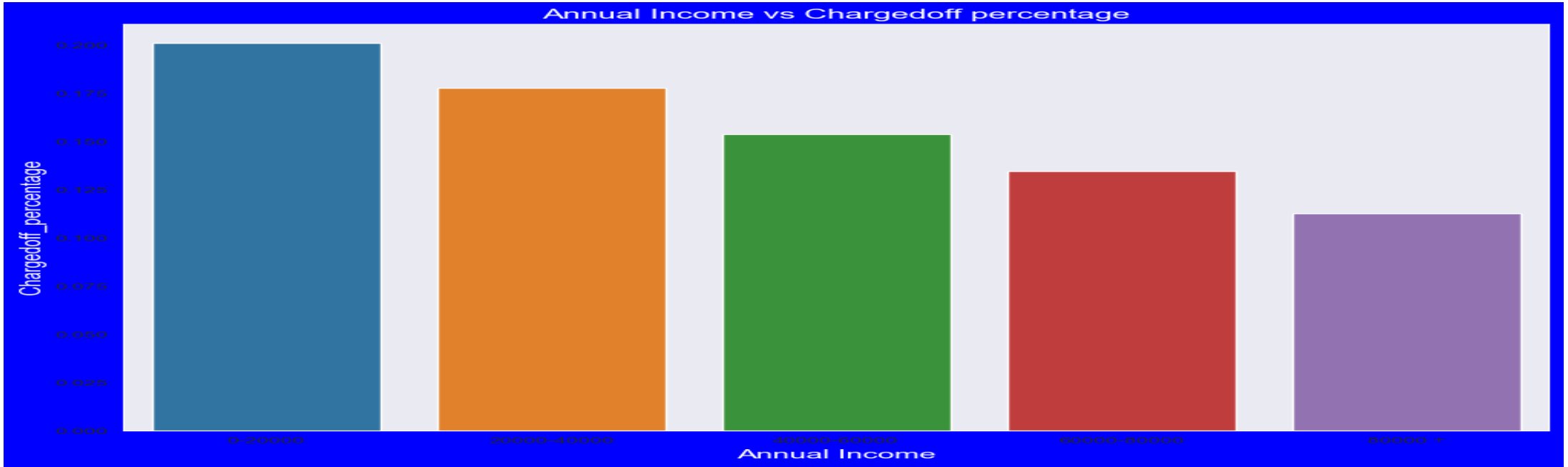
Correlation Matrix - Quantitative Variables

- observation:total_pymnt,loan_amnt,funded_amnt,funded_amnt_inv,installment are highly correlate
- anuul_income and dti are negatively correlated.as annual_income increases dti decreases



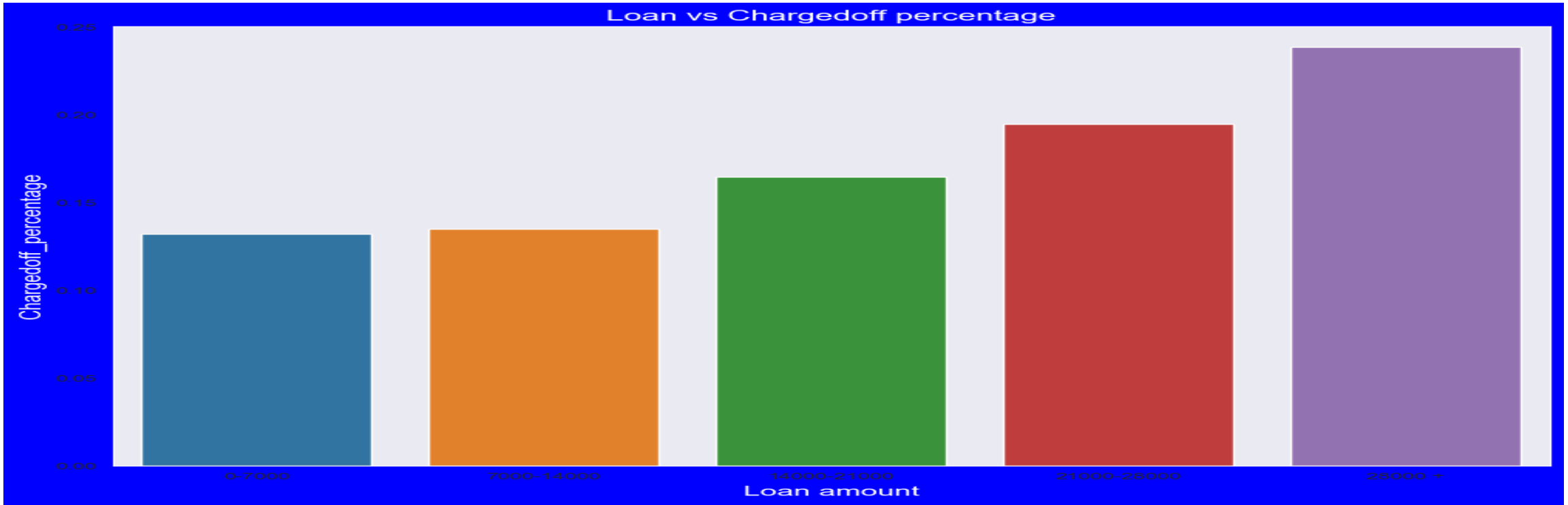
Bivariate Analysis

- Bivariate Analysis on annual income against Chargedoff_percentage.
- Observations:
- # Income range 80000+ has less chances of charged off.
- # Income range 0-20000 has high chances of charged off.
- # Notice that with increase in annual income charged off proportion got decreased



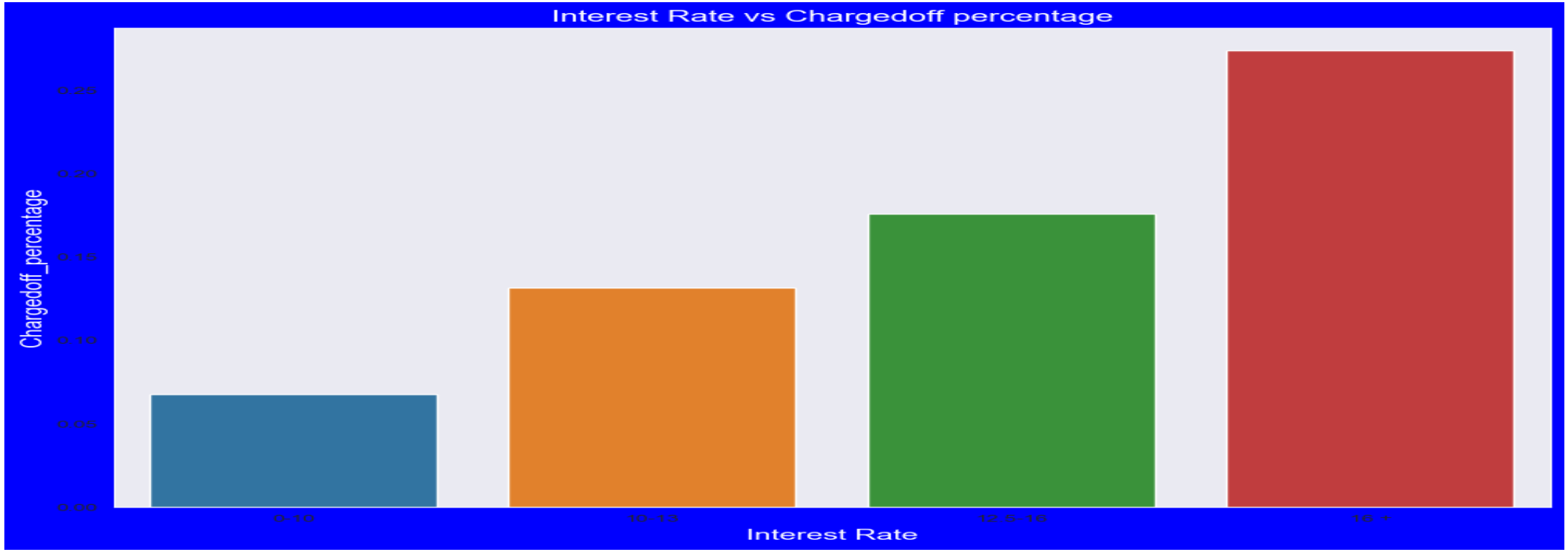
Bivariate Analysis

- Bivariate Analysis on loan amount against Chargedoff_percentage.
- Observations: loan amount 0-7000 has less chances of charged off.
- # loan amount 280000+ has high chances of charged off.
- # Notice that with increase in loan amount charged off proportion got increased



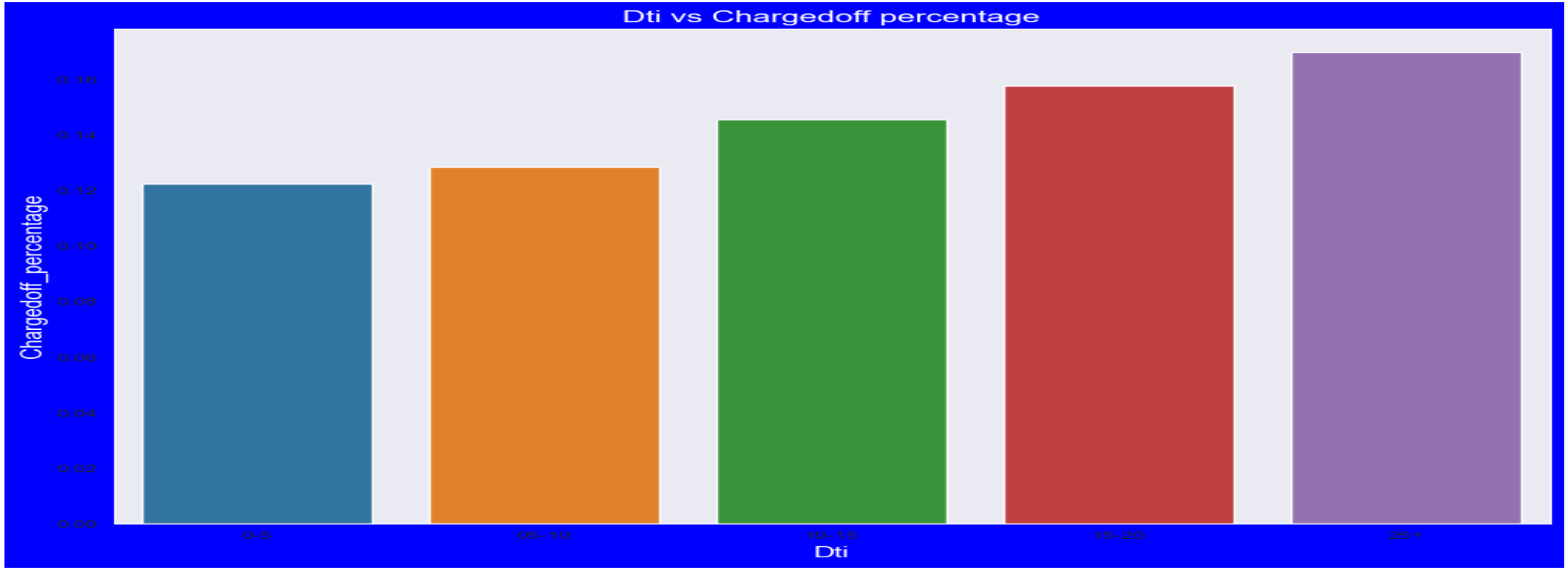
Bivariate Analysis.

- Bivariate Analysis on interest rate against Chargedoff_percentage
- Observations:Interest Rate 16+% has more chances of charged off.
- # Interest Rate 0 -10% has less chances of charged off.
- # Notice that with increase in Interest Rate charged off proportion got increase.



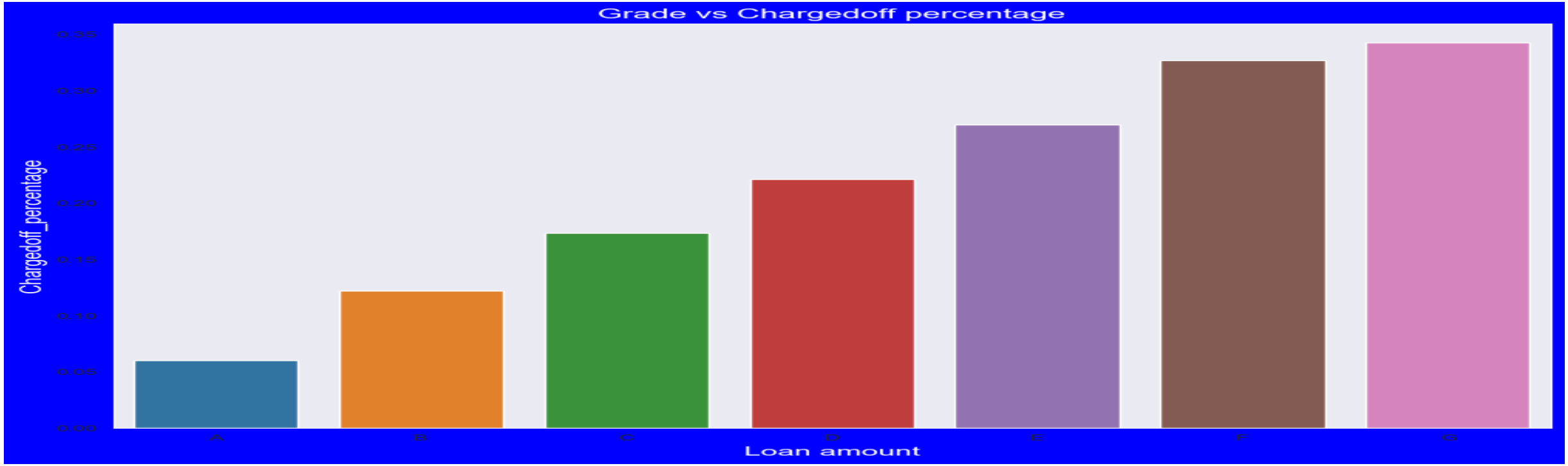
Bivariate Analysis

- Bivariate Analysis on dti against Chargedoff_percentage.
- Observations:Dti of applicant 25+ has more chances of charged off.
- # Dti of applicant 0-5 has less chances of charged off.
- # Notice that with increase in Dti charged off proportion got increased.



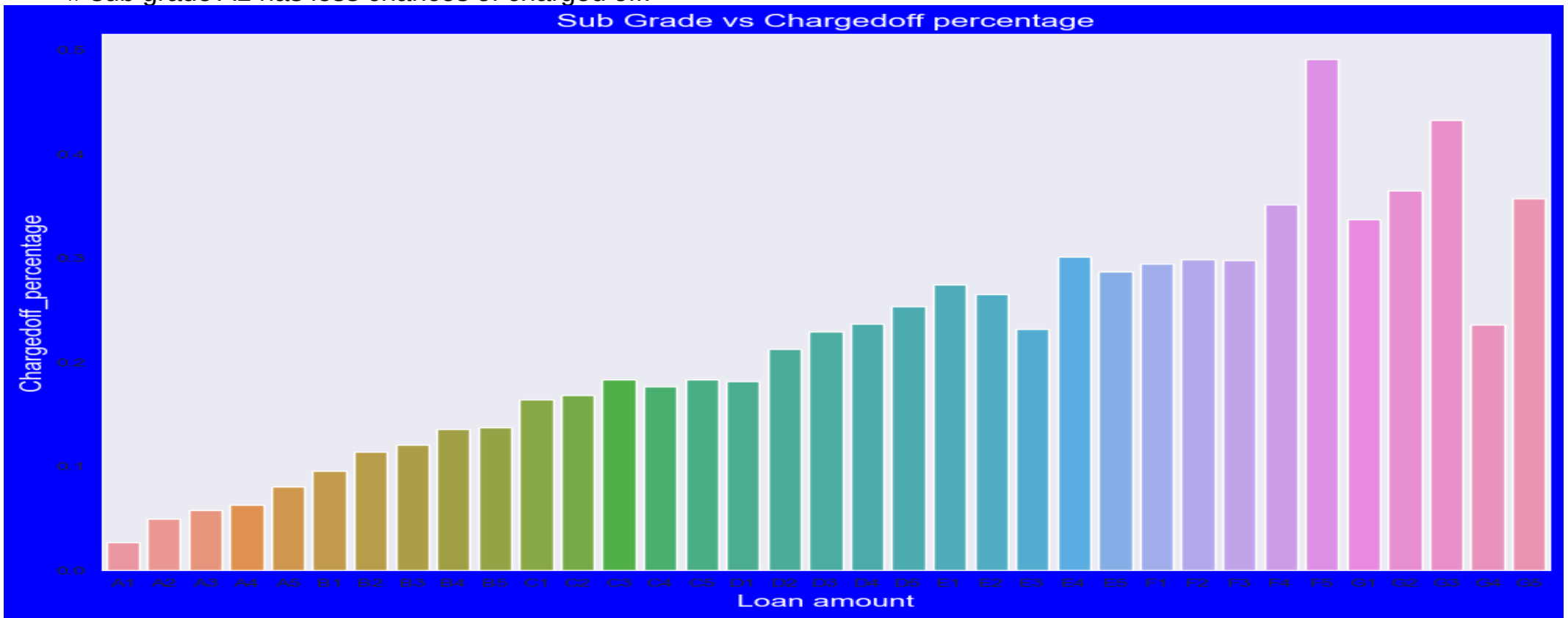
Bivariate Analysis

- Bivariate Analysis on grade against Charged off_percentage.
- Observations: Grade G has more chances of charged off.
- # Grade A has less chances of charged off.
- # charged off inceases as grade moves from A to G.



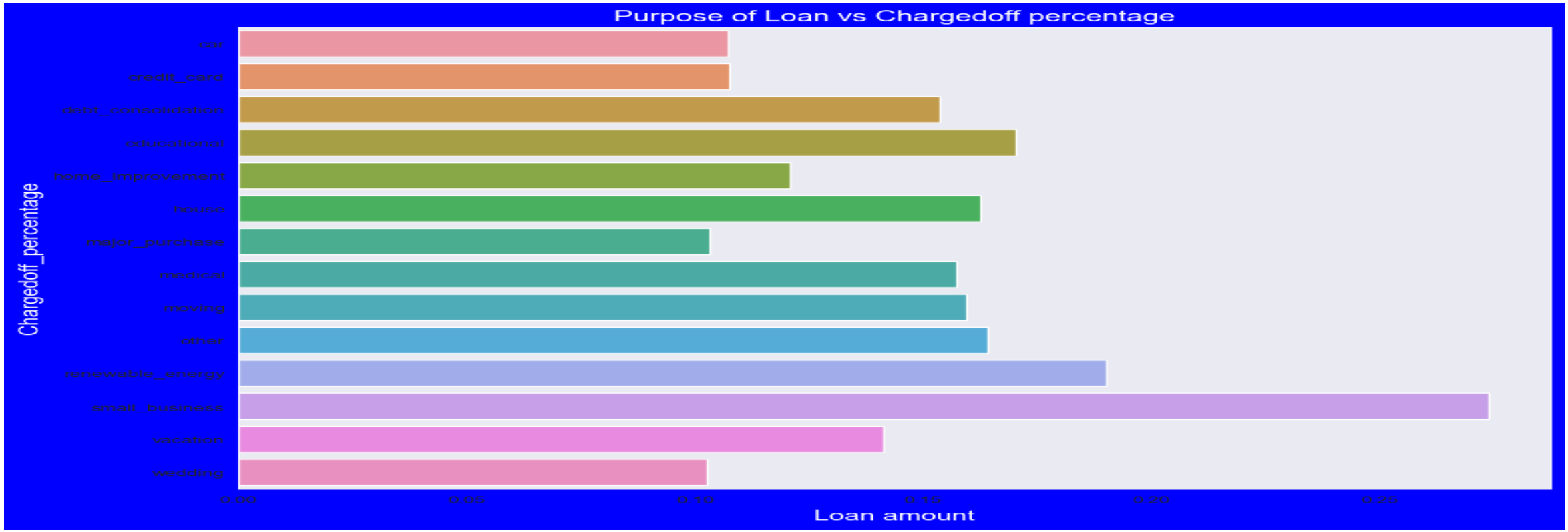
Bivariate Analysis

- Bivariate Analysis on sub grade against Chargedoff_percentage.
- # Observations:sub grade F5 has more chances of charged off.
- # sub grade A1 has less chances of charged off.



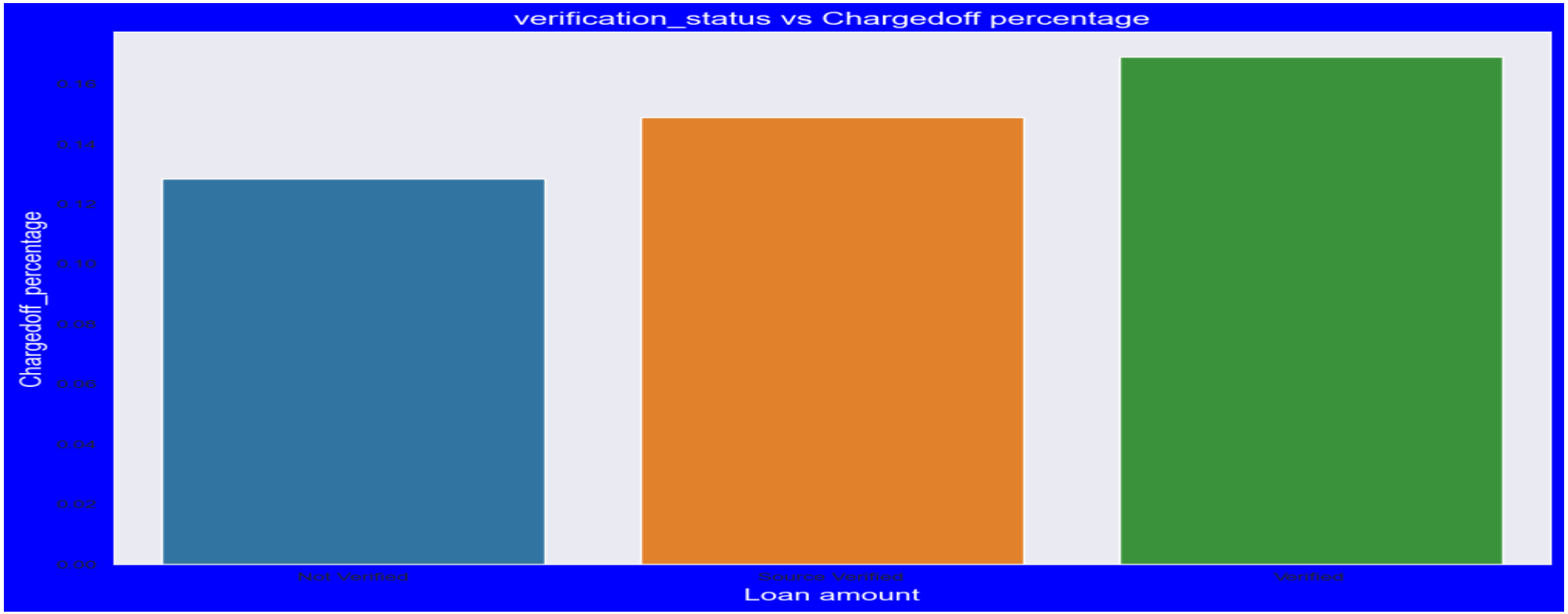
Bivariate Analysis

- Bivariate Analysis on purpose of loan against Chargedoff_percentage
- Observations:Loan which is taken for small business has more chances of charged off.
- # Loan which is taken for wedding,credit_card and car has less chances of charged off.



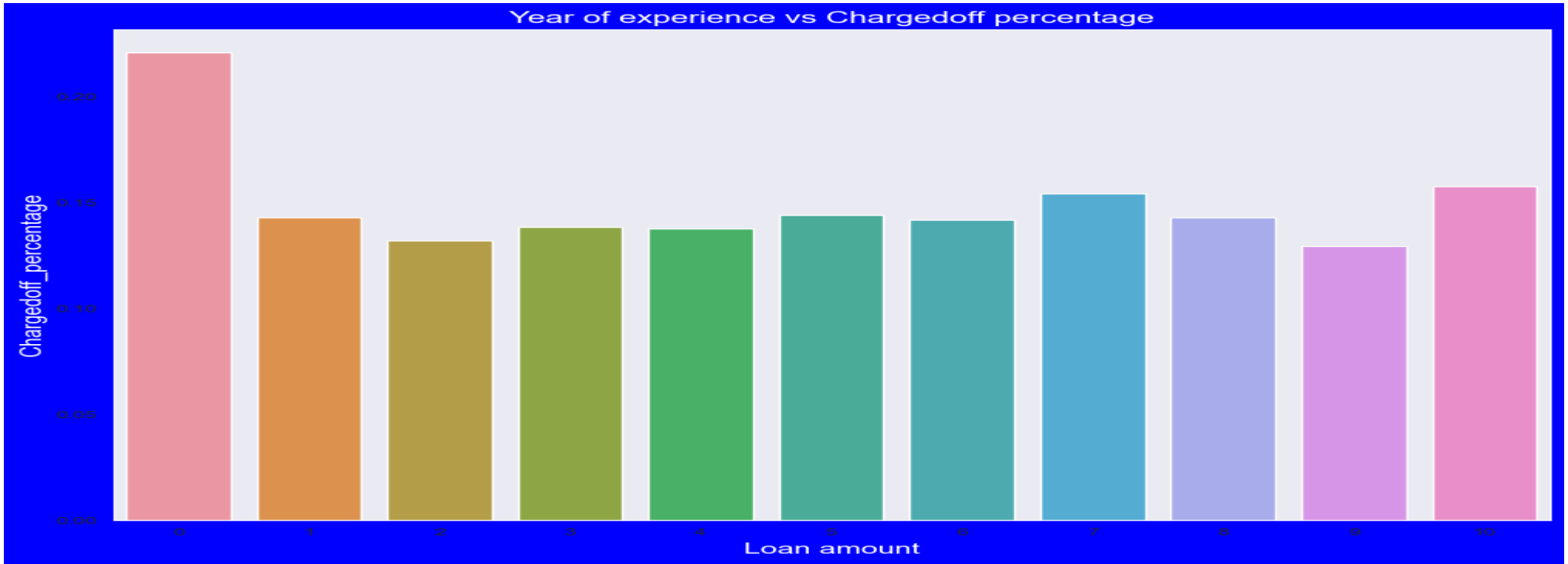
Bivariate Analysis

- Bivariate Analysis on verification_status against Chargedoff_percentage.
- # Observations: There is not much difference in charged off proportion.
- # This variable doesn't provide any insights for charged off.



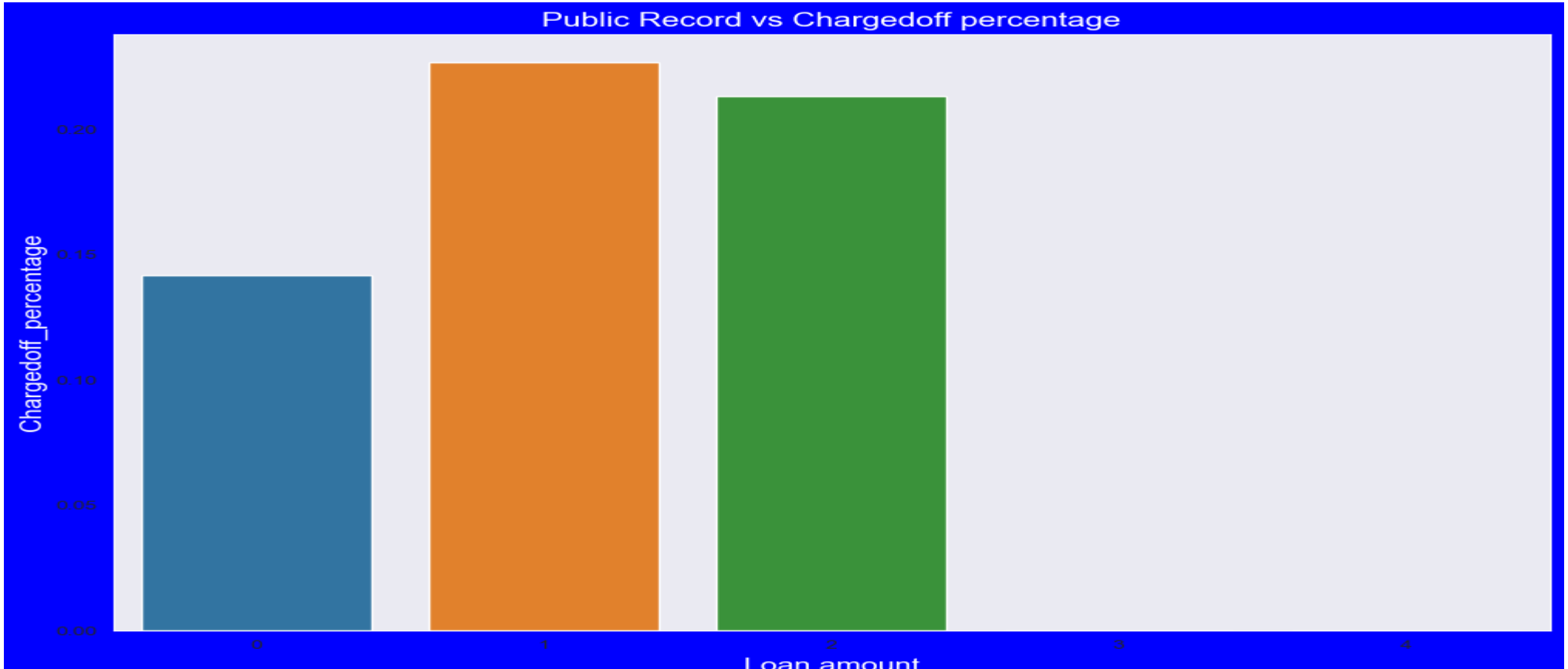
Bivariate Analysis

- Bivariate Analysis on employment length against Charged off_percentage
- Observations: Employee experience of 10 years has less chances of charged off.
- # No Employee experience has high chances of charged off.



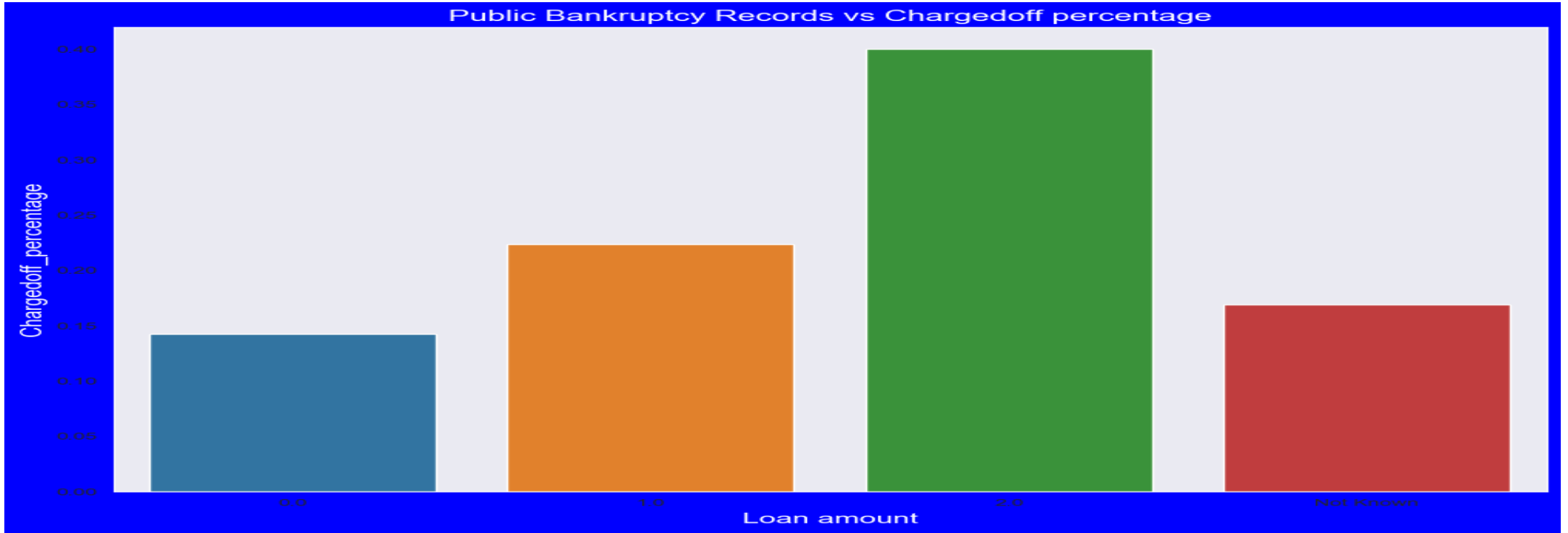
Bivariate Analysis

- Bivariate Analysis on public record against Chargedoff_percentage
- # Observations :Public record range 3-4 has no charged_off
- # public record range 0-2 has more chances of charged_off



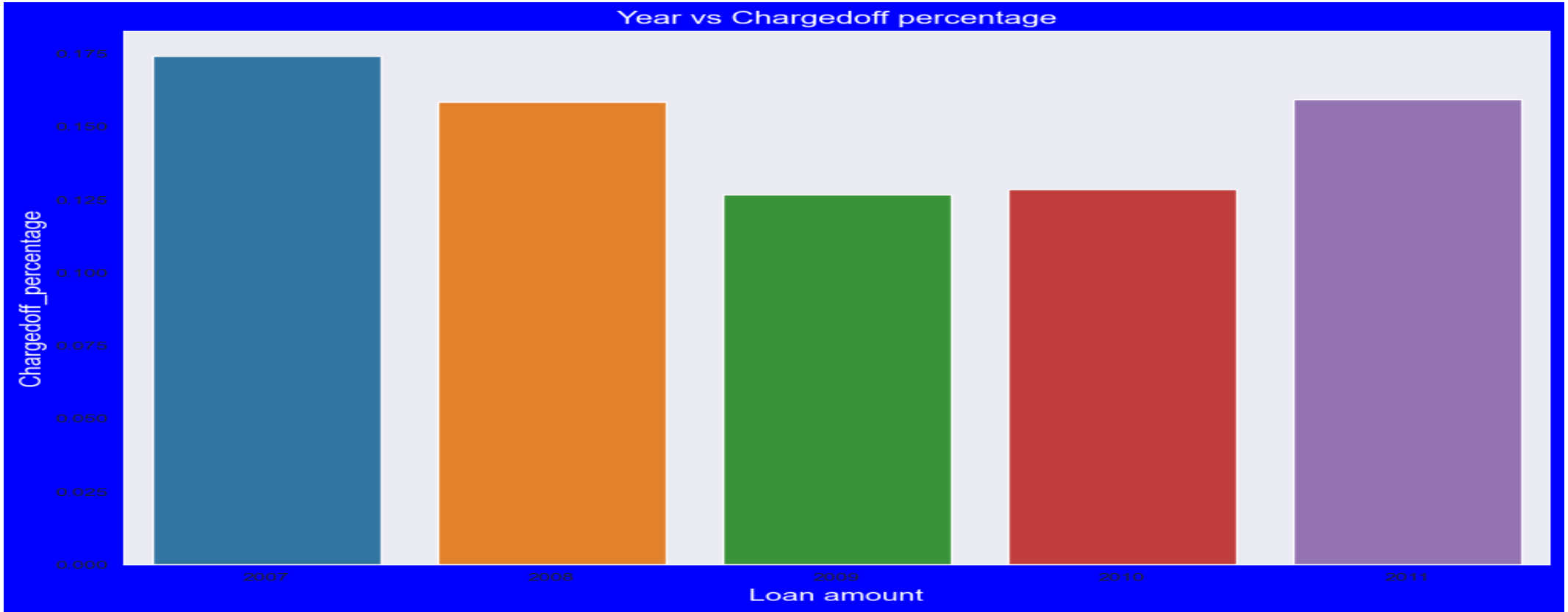
Bivariate Analysis

- Bivariate Analysis on Public Bankruptcy Records against Chargedoff_percentage
- # Observations: Those who already have pub_rec_bankruptcies value 2, have charged off proportion higher than who have no pub_rec_bankruptcies.
- Not known is the column for which we don't have any information about borrower.
- This also makes sense that who has defaulted before has chances of dafaulting in future as well.



Bivariate Analysis

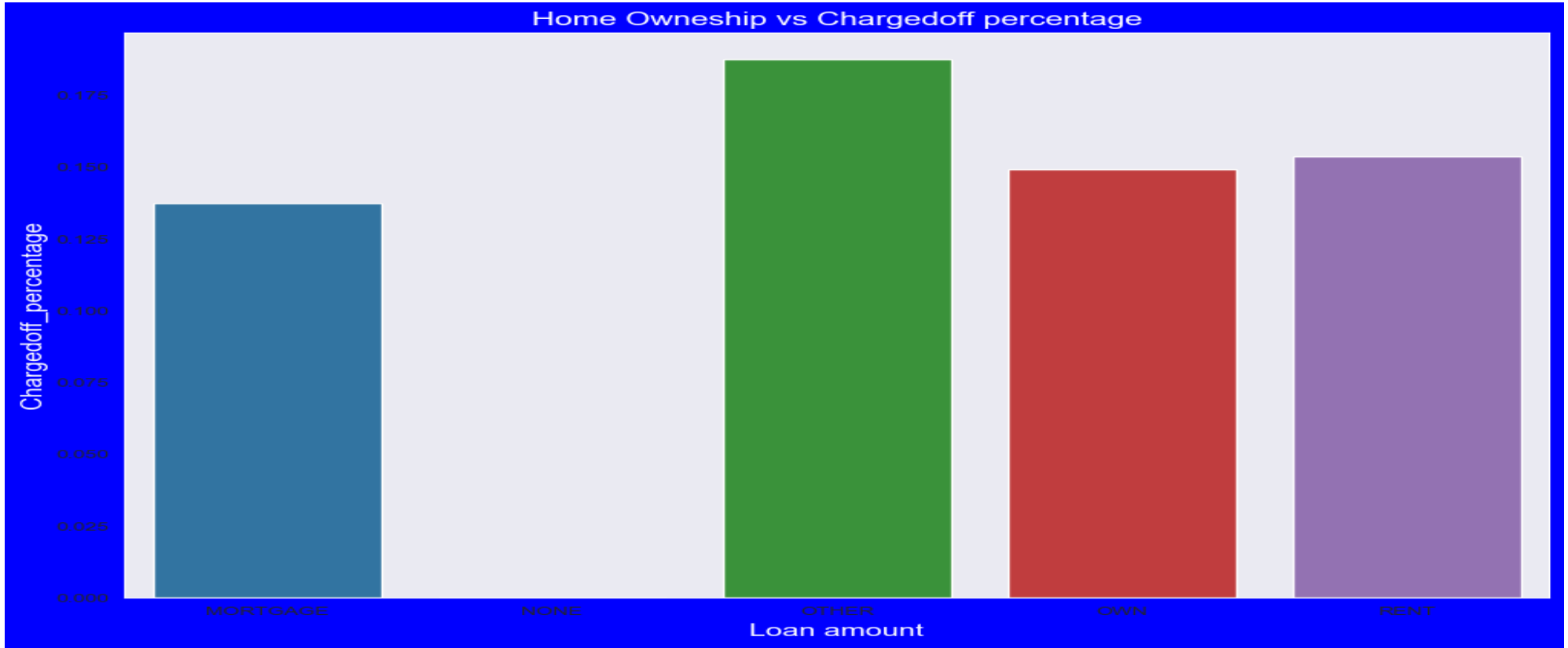
- Bivariate Analysis on Year against Chargedoff_percentage.
- Observations: Highest percentage of charged off(17%) happened in the year 2007
- Next Highest percentage of charged off(15%) happened in the year 2011 and 2008



Bivariate Analysis

Bivariate Analysis on Home Ownership Records against Chargedoff_percentage

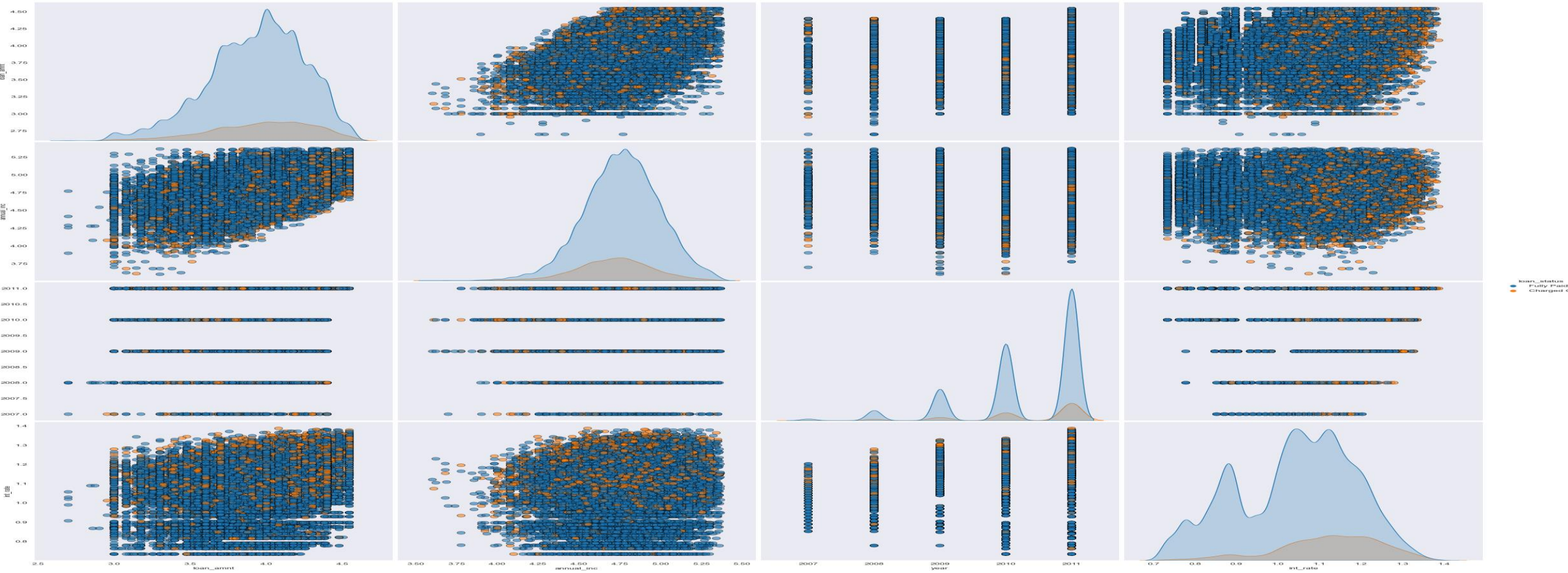
Observations: The percentage of charged off for having home rent,own or mortage remains almost same others having 18% of charged off



Multivariate Analysis

- Observation: Higher the interest rate higher charged off ratio
- Higher the annual income higher the loan amount slightly.
- interest rate is increasing with loan amount increase

Multivariate Analysis



Conclusion:

1.Data Cleaning:

1.Few feature in loan dataset has most of the values to be null.

These features removed for analysis

2. For our analysis loan_status ,only full paid and charged off considered as we analysis only default and non default applicant,current paying loan ampont is not considered .hence its filtered

Conclusion:

- Giving Loan to following cases may causes business loss to the company

1. Applicant whose annual Income range 0-20000 has high chances of charged off. Notice that with increase in annual income charged off proportion got decreased.

2. loan amount 0-7000 has less chances of charged off. loan amount 280000+ has high chances of charged off. Notice that with increase in loan amount charged off proportion got increased.

3. Rate 16+% has more chances of charged off. Interest Rate 0 -10% has less chances of charged off.Notice that with increase in Interest Rate charged off proportion got increase.

4. Dti of applicant 25+ has more chances of charged off. Dti 0-5 has less chances of charged off. Notice that with increase in Dti charged off proportion got increased.

5. The percentage of charged off for having home rent,own or mortage remains almost same others having 18% of charged off.

6. Highest percentage of charged off(17%) happened in the year 2007.Next Highest percentage of charged off(15%) happened in the year 2011 and 2008.

7. Those who already have pub_rec_bankruptcies value 2, have charged off proportion higher than who have no pub_rec_bankruptcies.

Not known is the column for which we don't have any information about borrower.This also makes sense that who has defaulted before has chances of dafaulting in future as well.

Conclusion:

8. Public record range 3-4 has no charged_off.and public record range 0-2 has more chances of charged_off.

9. Employee experience of 10 years has less chances of charged off. No Employee experience has high chances of charged off.

10. Loan which is taken for small business has more chances of charged off. Loan which is taken for wedding,credit_card and car has less chances of charged off.

11. Grade G has more chances of charged off. Grade A has less chances of charged off. charged off inceases as grade moves from A to G.