

Lending Club Case Study

Anshul Sarswat

Ankit Gayen

LENDING CLUB ANALYSIS OVERVIEW

- IDENTIFYING BUSINESS PROBLEM
- DATA DESCRIPTION
- DATA PRESENTATION AND PROCESSING
- MODELS
- SUMMARY OF FINDINGS
- CONCLUSION

BUSINESS UNDERSTANDING

- We are working for Lending club a finance company which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:
 - If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
 - If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

OBJECTIVE

- The company wants to understand the driving factors (or driver variables) behind loan default (`loan_status = 'Charged Off'`), i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

DATA CLEANING

- DROP THE NULL/ADDITIONAL COLUMNS BASED ON ANALYSIS
- LEFT WOTH 44 COLUMNS
- CONVERT ALL VALUES INTO NUMETIC VALUES FOR BETTER CALCULATIONS

OBSERVATIONS

Fully Paid 82.96

Charged Off 14.17

Current 2.87

Name: loan_status, dtype: float64

DATA CLEANING

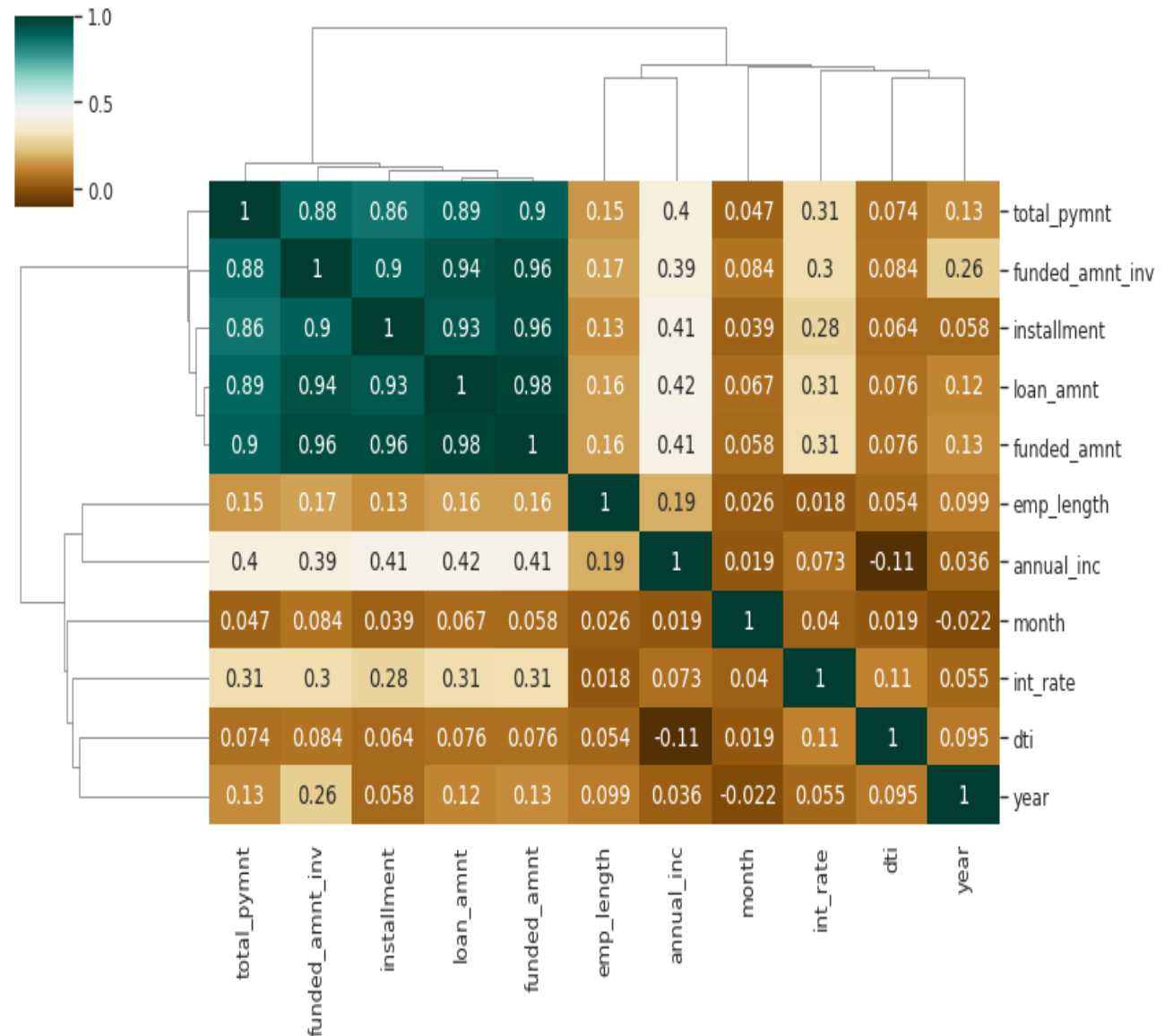
- ANALYSIS PF PURPOSE OF LOAN
 - Most of the loans taken for debt consolidation(47%) and Credit card bill payment.
 - Other is also one of the main area where loans were taken.
- **Derive Columns For Analysis**
 - Two new columns month and year were derived from issue date

UNIVARIATE ANALYSIS

- REMOVE OUTLIERS QUANTITILE 0.00 FROM ANNUAL INCOME
- ANALYSIS
 - AVERAGE % RATE IS 12%
 - AFTER 75% PERCENTILE INTEREST RATE ZOOMED TO 25% FROM 15%

BIVARIATE ANALYSIS

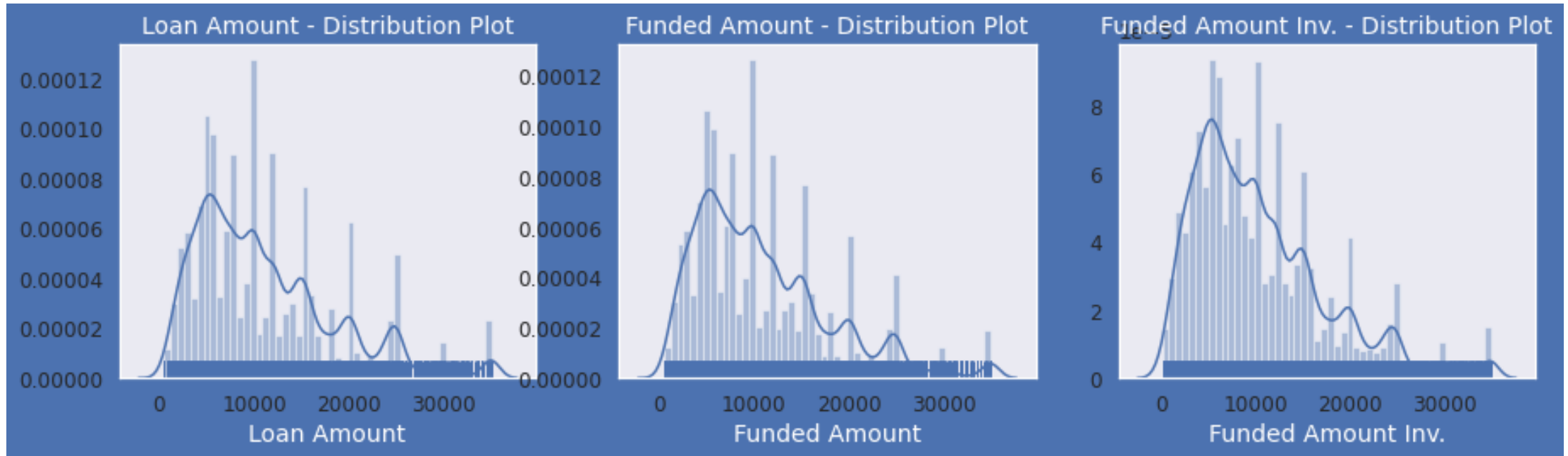
- Correlation Matrix- Quantitative Variable
- Heatmap with Dendrogram (clustermat) to show closeness among numerical variables.



BIVARIATE ANALYSIS

- Observation is that Loan amount, investor amount, funding amount are strongly correlated.
- Annual income with DTI(Debt-to-income ratio) is negatively correlated.
- Debt income ratio is the percentage of a consumer's monthly gross income that goes toward paying debts.
- That means when annual income is low DTI is high & vice versa.
- positive correlation between annual income and employment years. # That means income increases with work experience.

UNIVARIATE ANALYSIS

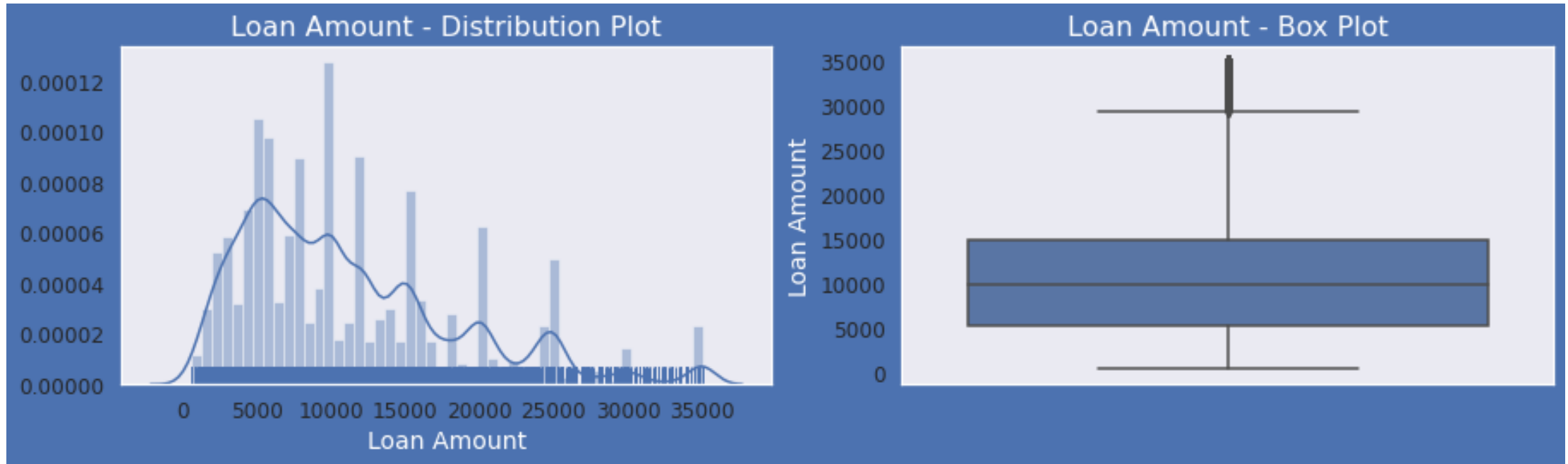


Observation:

Distribution of amounts for all three looks very much similar.

We will work with only loan amount column for rest of our analysis.

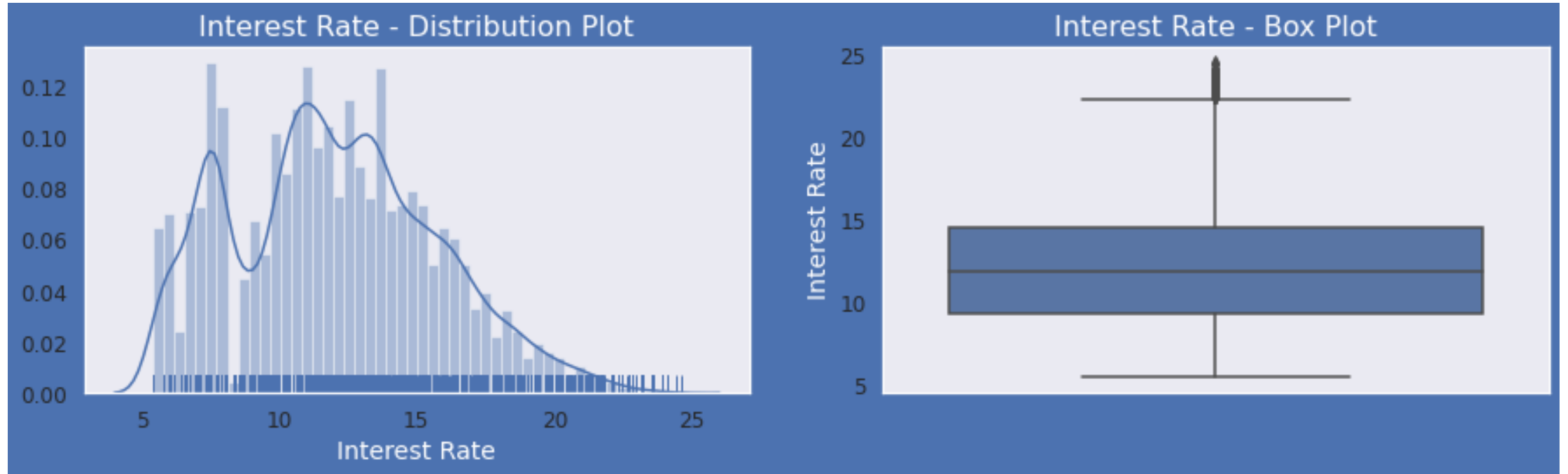
Univariate Analysis on Loan amount-Quantitative Variables



Observations :

Above plots show that most of the Loan amounts are in range of 5000 - 15000

Univariate Analysis on Interest Rate-Quantitative Variables

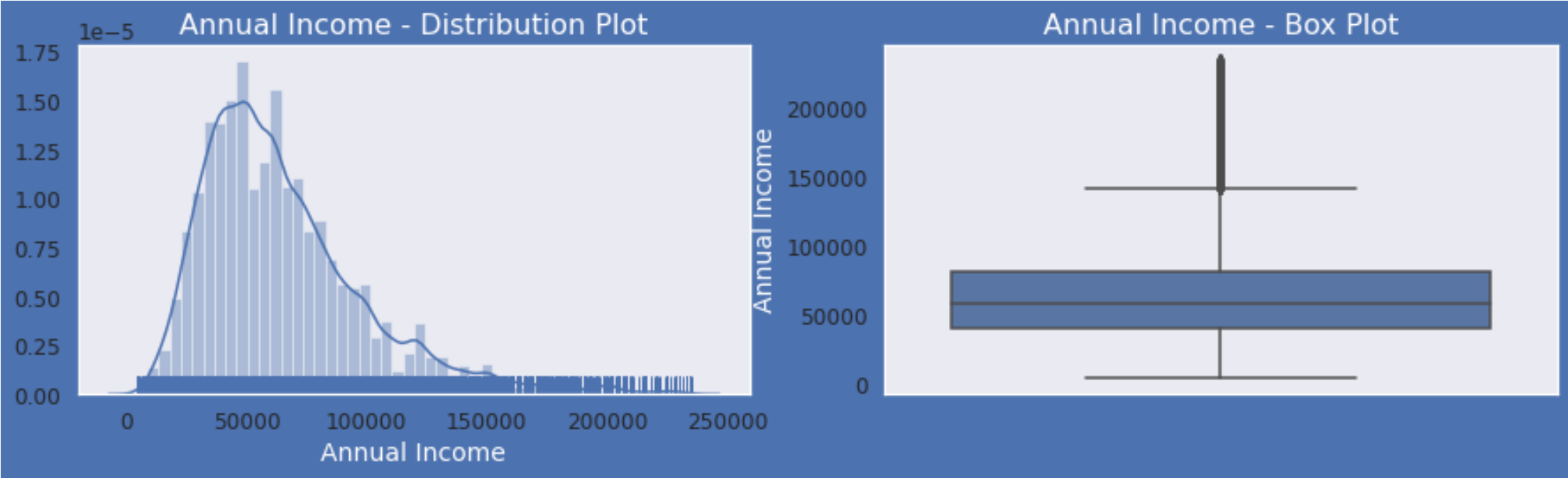


Observations :

Above plots show that most of the Interest Rates on loans are in range of 10% - 15%

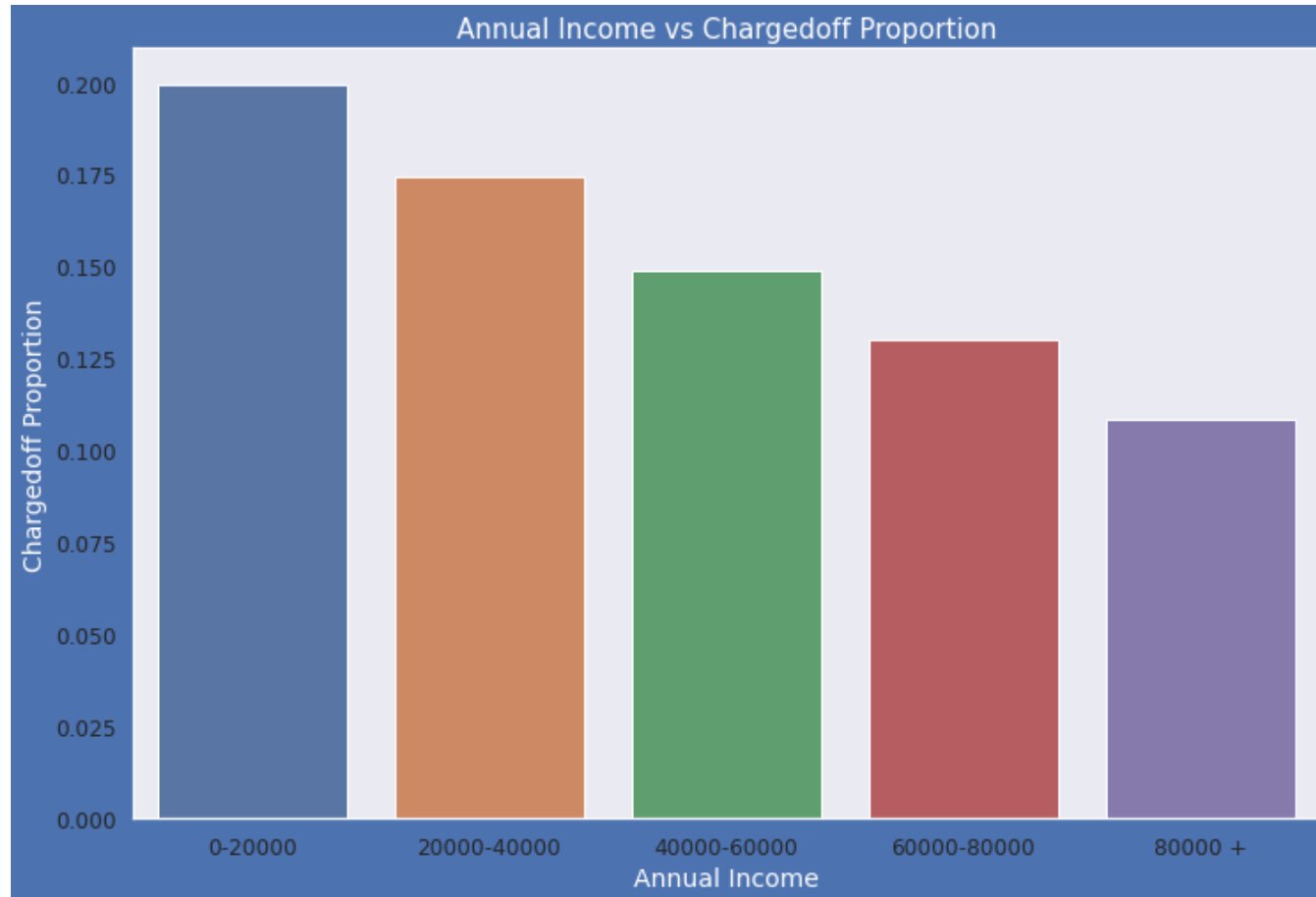
Observations : # Below plots show that most of the borrower's Annual incomes are in range of 40000- 80000

Univariate Analysis on Annual Income - Quantitative Varia



Observations :
Below plots show that most of the borrower's Annual incomes are in range of 40000- 80000

Bivariate Analysis : Part 1

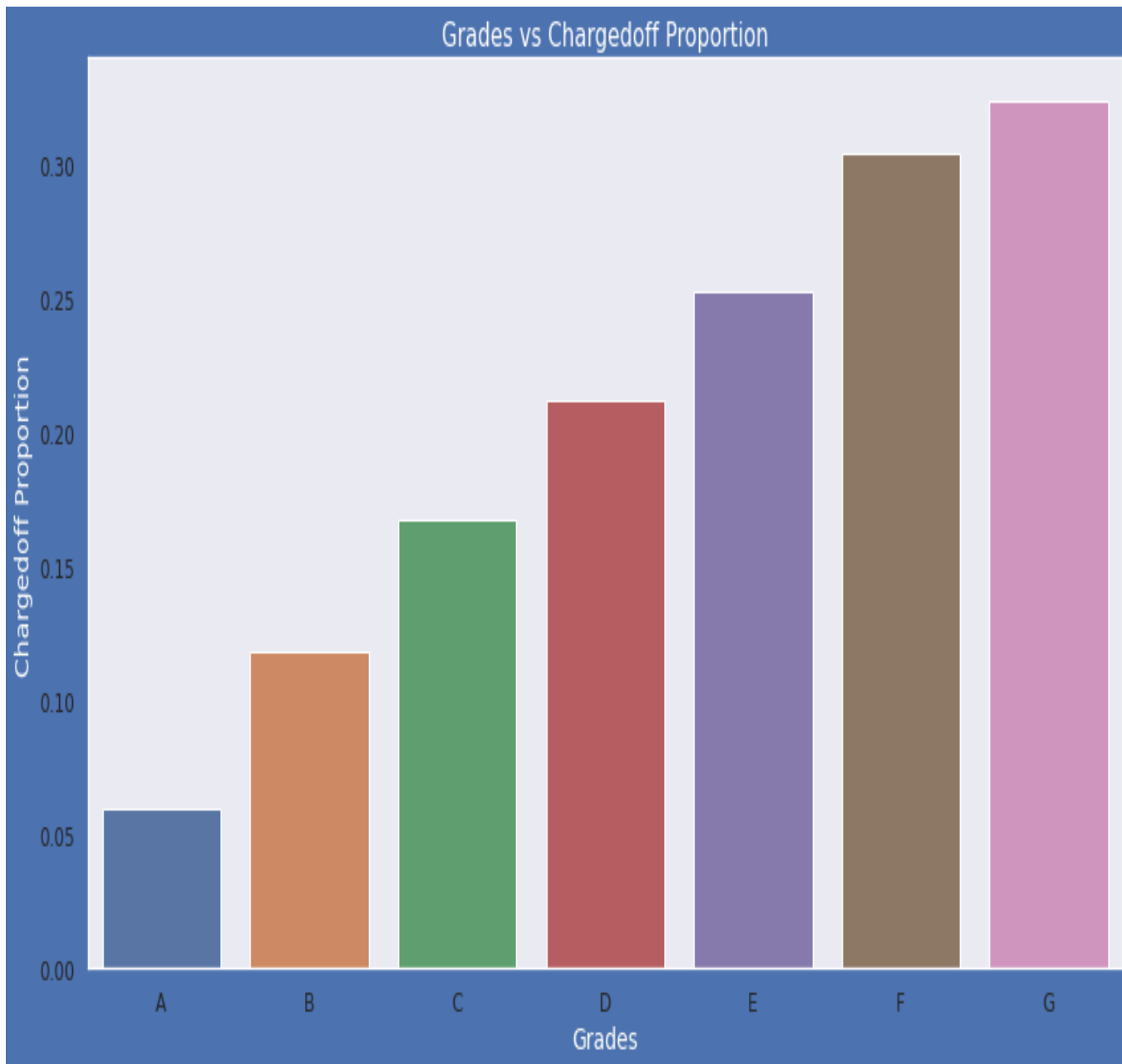


Observations:

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.



Observations:

Grade "A" has very less chances of charged off.

Grade "F" and "G" have very high chances of charged off.

Chances of charged of is increasing with grade moving from "A" towards "G"

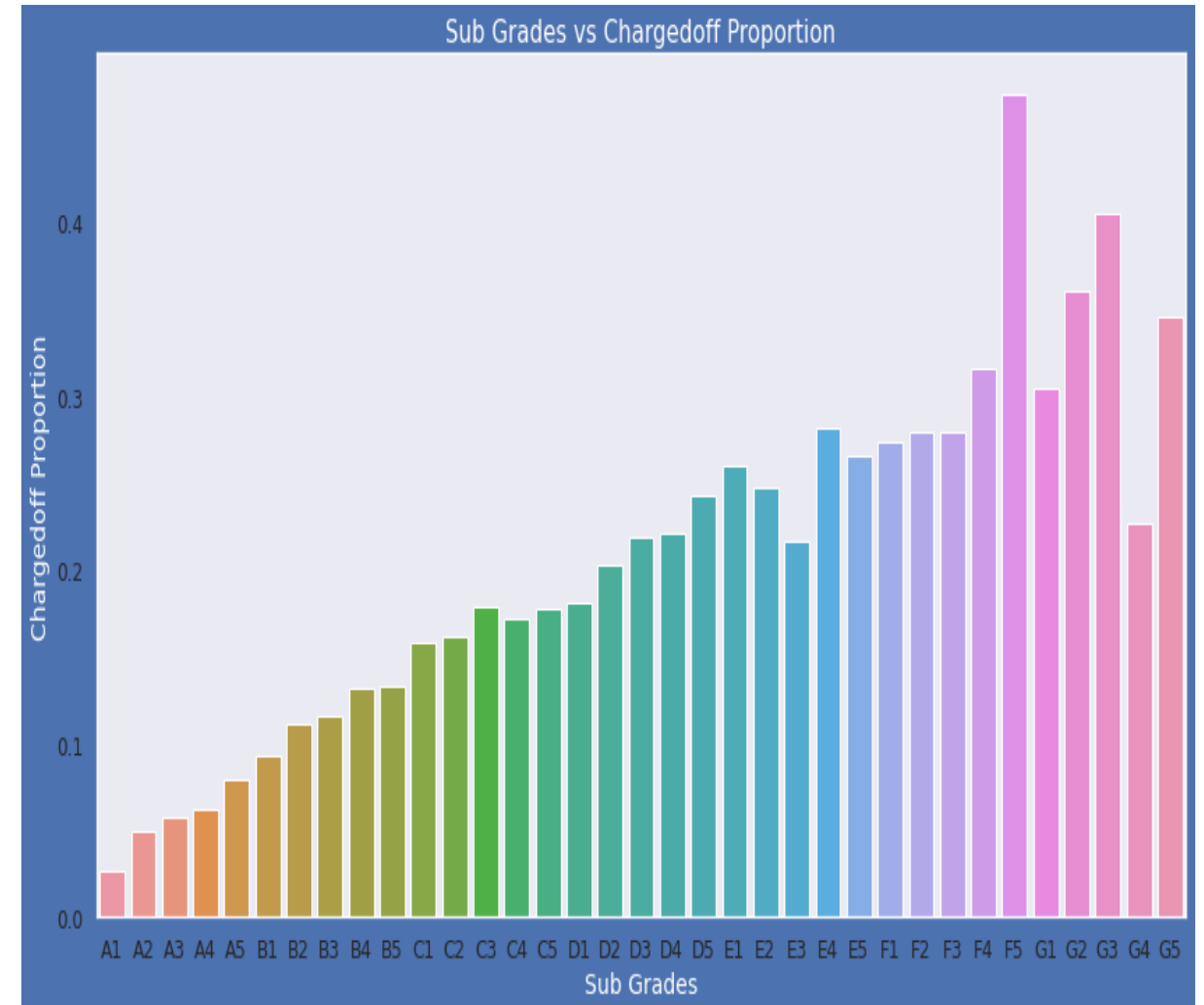
Bivariate Analysis on sub grade against Chargedoff_Proportion

Observations:

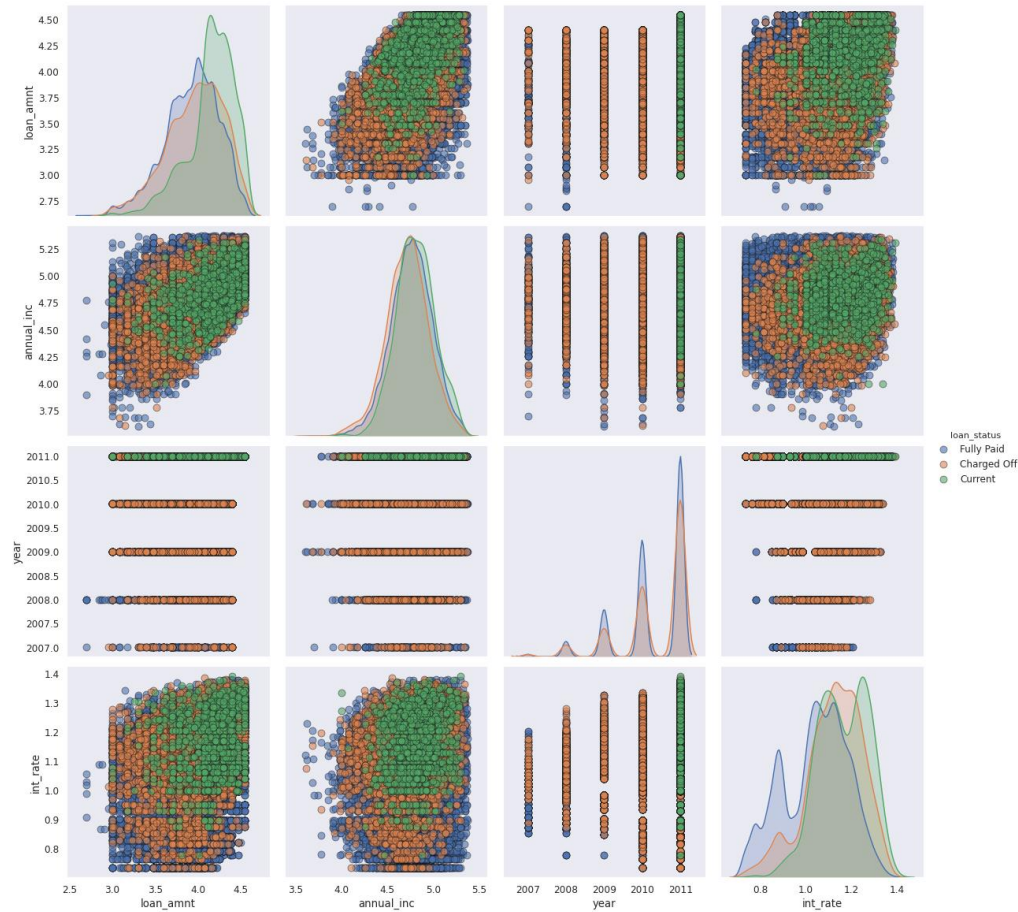
sub Grades of "A" has very less chances of charged off.

sub Grades of "F" and "G" have very high chances of charged off.

proportion of charged off is increasing with sub grades moving from sub grades of "A" towards sub grades of "G"



Multivariate Analysis - Pair Plots



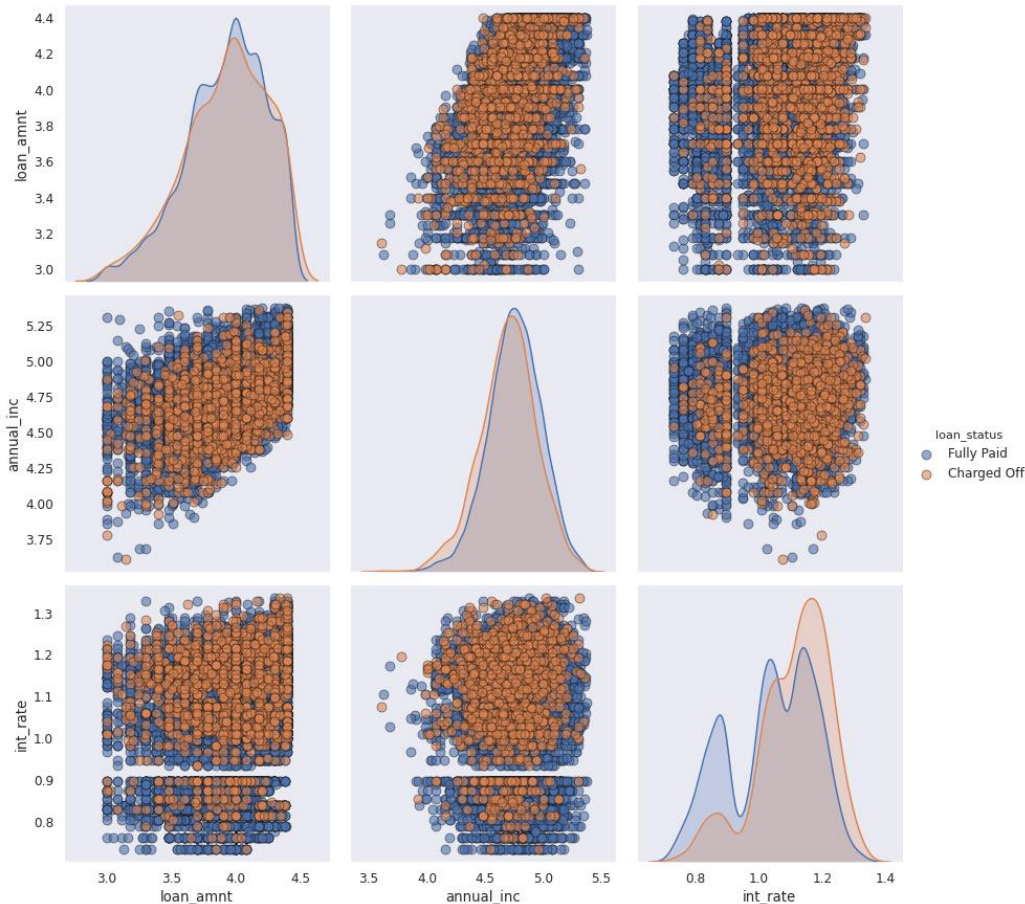
Observation:

Higher the interest rate higher charged off ratio #

Higher the annual income higher the loan amount slightly. # increase in number of charged off with increase in year.

interest rate is increasing with loan amount increase

Above plot is too dense so just pick year 2010 for getting some insights.



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 this results in high charged off.

OBSERVATIONS

- **Minor Impact**

- .Higher loan amount (above 16K)

- Higher installment amount (above 327)

- Lower annual income (below 37K)

- Higher debt to income ratio (above 15%)

- Applicant's address state (NV, SD, AK, FL, etc.)

- Loan issue month (Dec, May, Sep)

OBSERVATIONS

- **Heavy impact**
 - Higher interest rate (above 13%)
 - Higher revolving line utilization rate (above 58%)
 - Repayment term (5 years)
 - Loan grade & sub-grade (D to G)
 - Missing employment record
 - Loan purpose (small business, renewable energy, educational)
 - Derogatory public records (1 or 2)
 - Public bankruptcy records (1 or 2)

OBSERVATIONS

- **Combined impact**

- High loan amount & interest rate for lower income group
- High installment and longer repayment term
- Home ownership (other) and loan purpose (car, moving or small business)
- Residential state and loan purpose
- Income group and loan purpose