Lending Club

Exploratory Data Analysis – Case Study

Devi Murugappan

1/4/2022

Agenda

- Problem Statement
- Objective
- Approach
- Input
- Univariate Analysis
- BiVariate Analysis
- Segmented UniVariate Analysis
- Conclusion

Problem Statement

Lending Club Case study is based on the dataset from Fintech Domain. Digital Lending organizations will have to determine the profile of Loan borrower to assess the risk associated.

Two Types of Risks

- 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,** then approving the loan may lead to a **financial loss** for the company

Objective

- Identify the **driving factors behind loan default** in FinTech Domain.

- **Derive data point** which will help the organization to make
  - **Risk assessment** for processing loan applications
  - Aid to **build customized loan portfolio**

Approach

- Load the data
- Data Cleaning
- Handle Missing Data
- Handle Outlier
- Sanity Check
- Generate Derived metrics
- UniVariate Analysis
- BiVariate Analysis
- Segmented Univariate Analysis
- Conclusion

Input

- Dataset　　　　　: Loan.csv
- Data Dictionary　: Data_Dictionary.xlsx

Technology Used

- Jupiter Notebook
- Programming : Python 3.x
- Libraries : NumPy, Pandas, Matplotlib, Seaborn, Datetime

UniVariate Analysis

- Debt to Income Ration Indicator determines the ability to repay. Sample contains borrowers whose dti frequency is in the range of $5 - 25$

- Interest Rate Distribution shows the average rate of interest across different borrowers falls into the bucket of 10 - 15
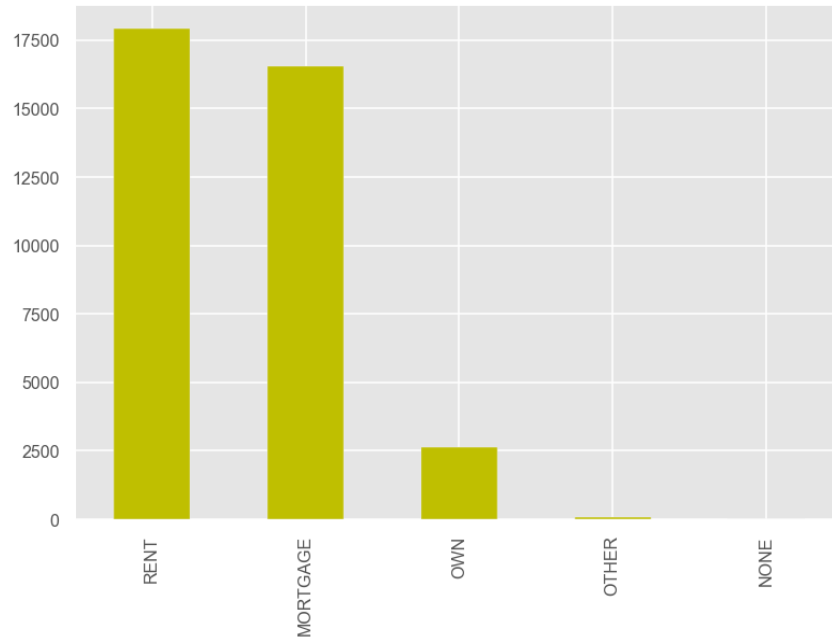

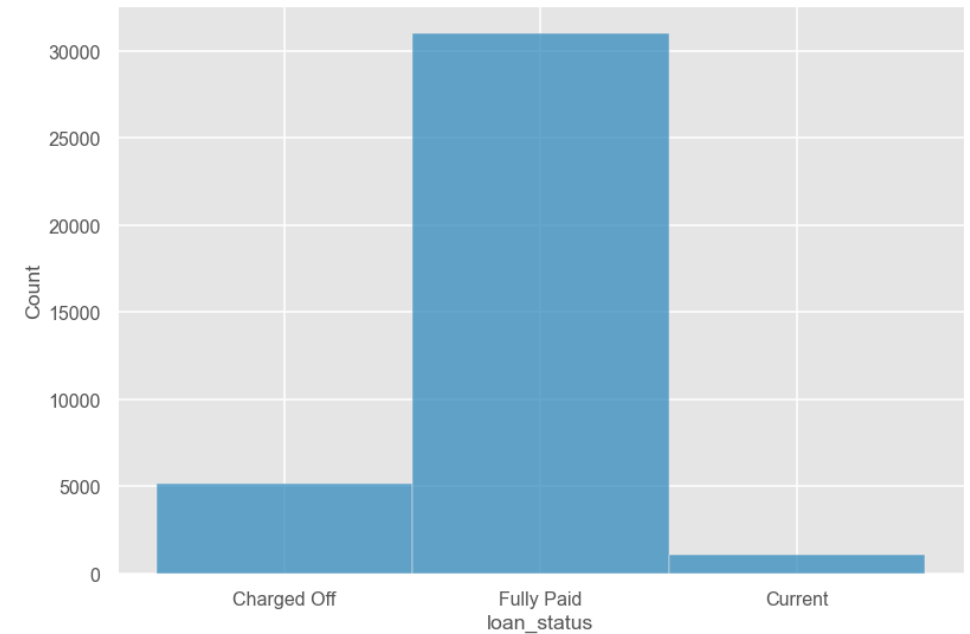Debt to Income Ratio Indicator
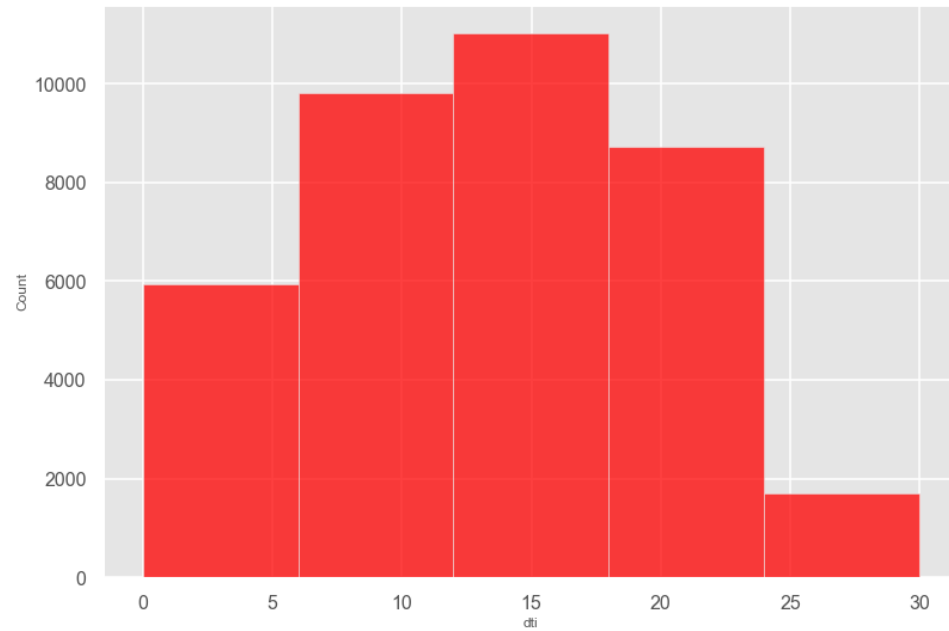
Interest Rate Distribution
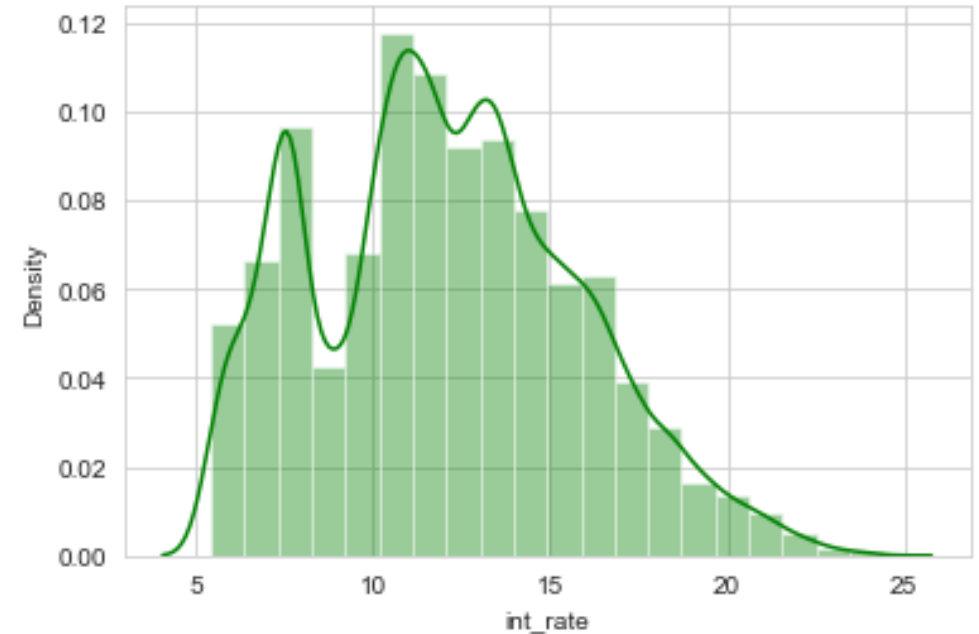
UniVariate Analysis

- Home Ownership Vs Loan Prospects Graph is an indicator of borrower's majority who opt for Loan also shows the financial support to repay

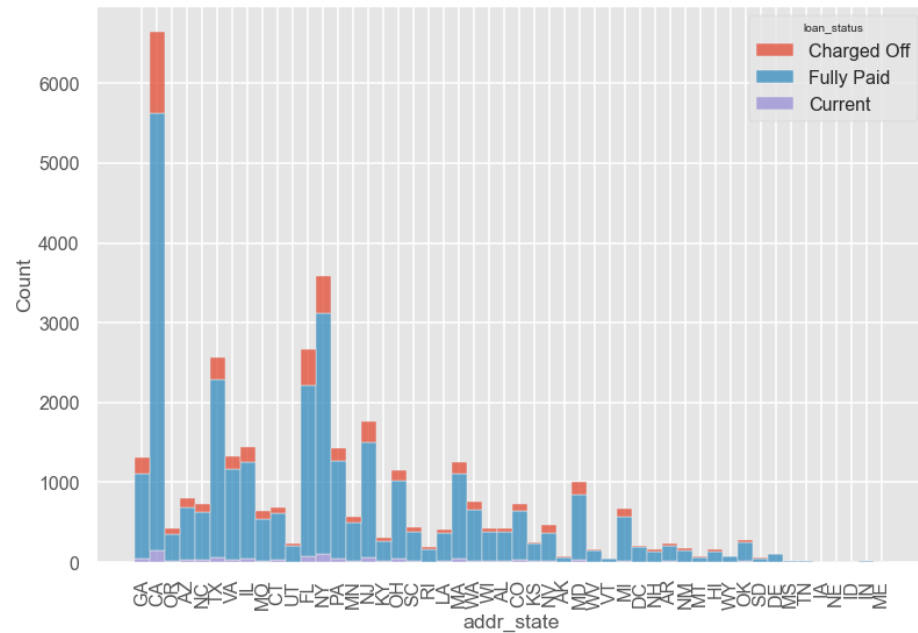- Outstanding Loan Status provides insight into the distribution of loan ,and the number of charged off lending for the financial institution
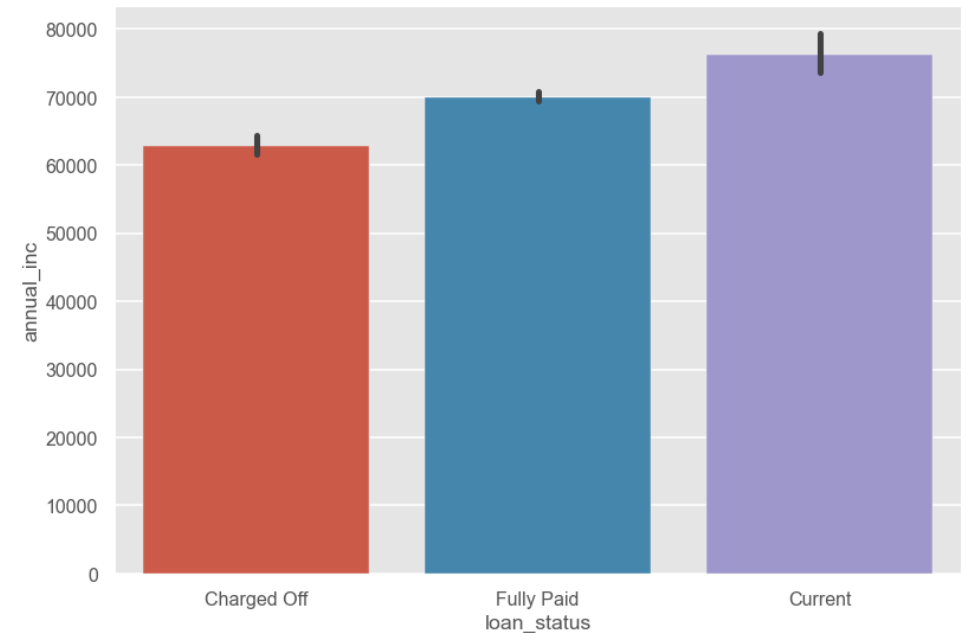
Bi-Variate Analysis

- Distribution of Loan by State shows from the dataset CA state with high number of borrowing and charged off loans
- Impact of Annual Income on Loan Status depicts borrowers with average annual income highly run into repay trouble
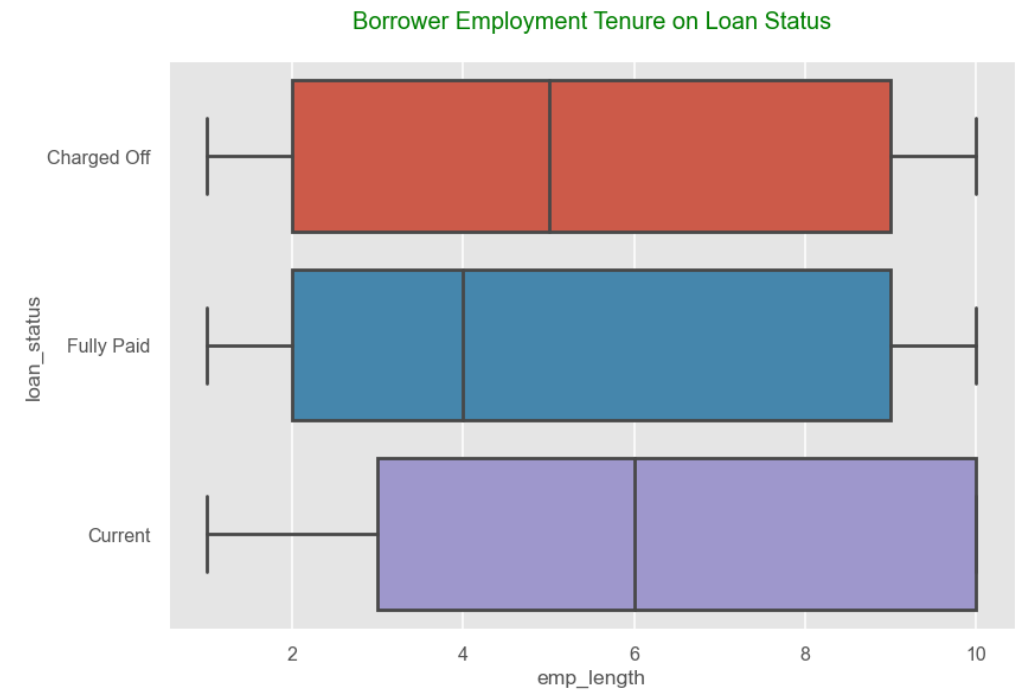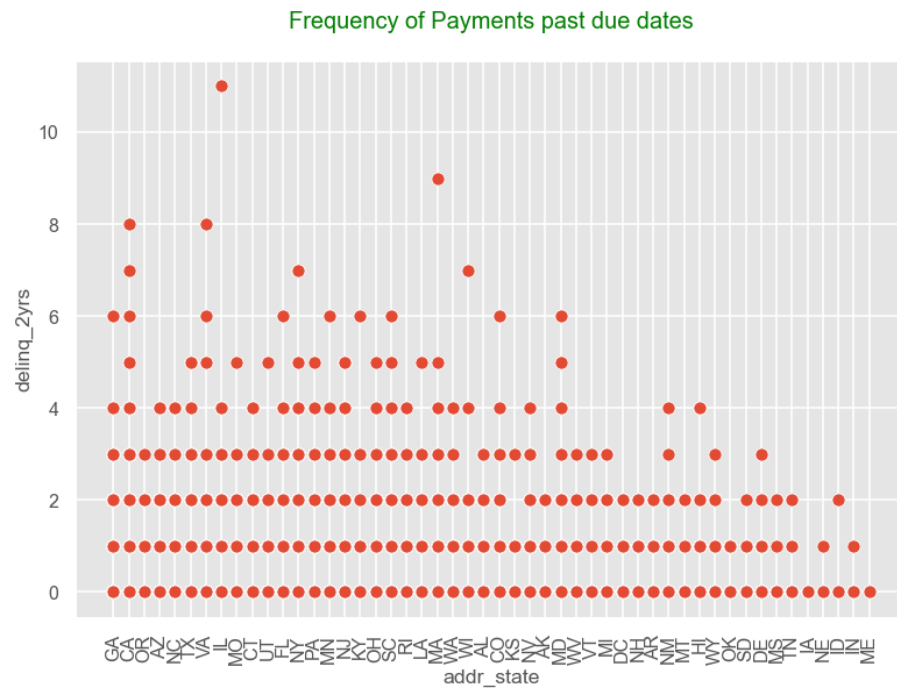
Bi Variate Analysis

- Frequency of Payments past due date shows the states where the delinq_2yrs above 4 have high probability of running into charged off loans

- Borrower Employment Tenure on Loan Status shows charged off loans have higher median compared to fully paid
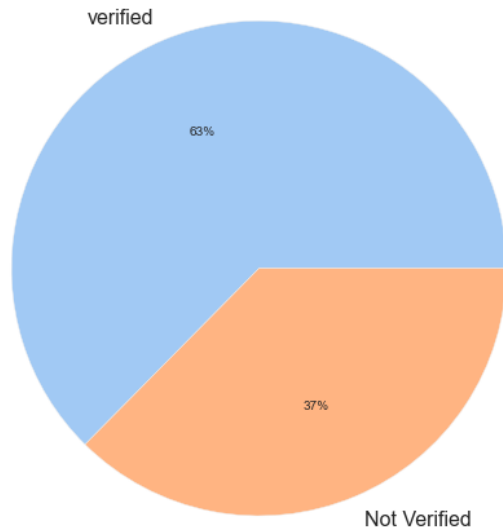
Bi Variate Analysis

- Varying Interest by Year shows the median interest between 4 yrs. (2007 – 2011) has been between 10 – 12.5%


Varying Interest Rate by Year
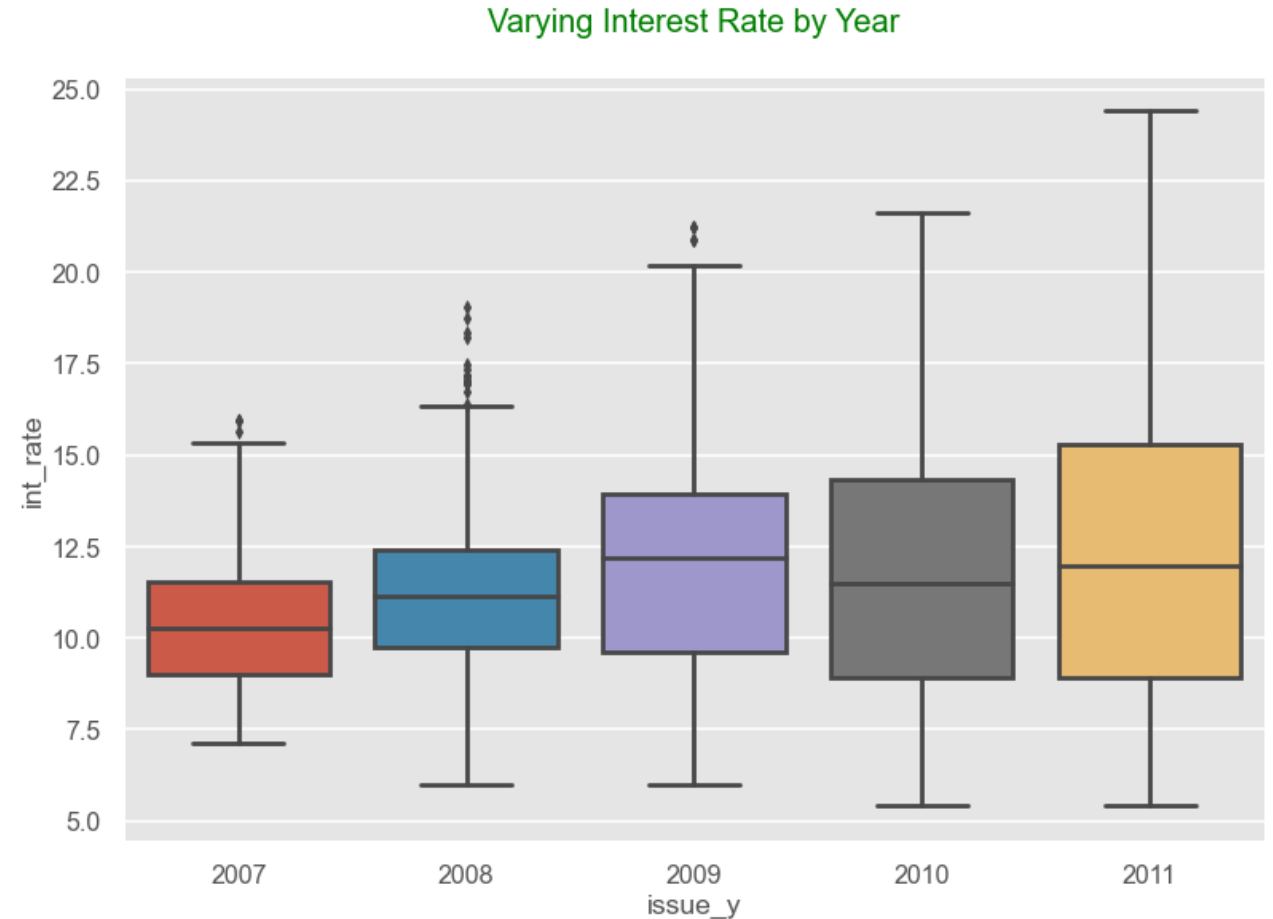
Segmented Analysis

- Distribution of Charged off Loans by Status of Verification of loan application shows that 37% of the default loans can be avoided if it was verified

- Charged off Loans varying Interest Rate and Funded amount across the years shows that interest rate has been on a gradual increase whereas the funded amount shows spike around later years
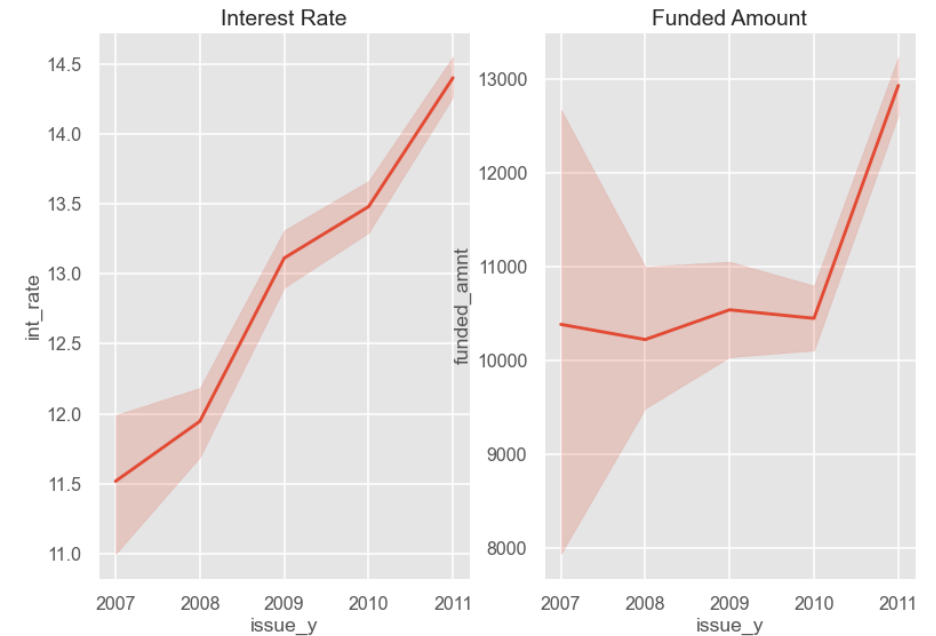
Segmented Analysis

- Pair Plot shows the behavior of Charged off loans against the funded amount, interest rate, issue year, dti and annual income
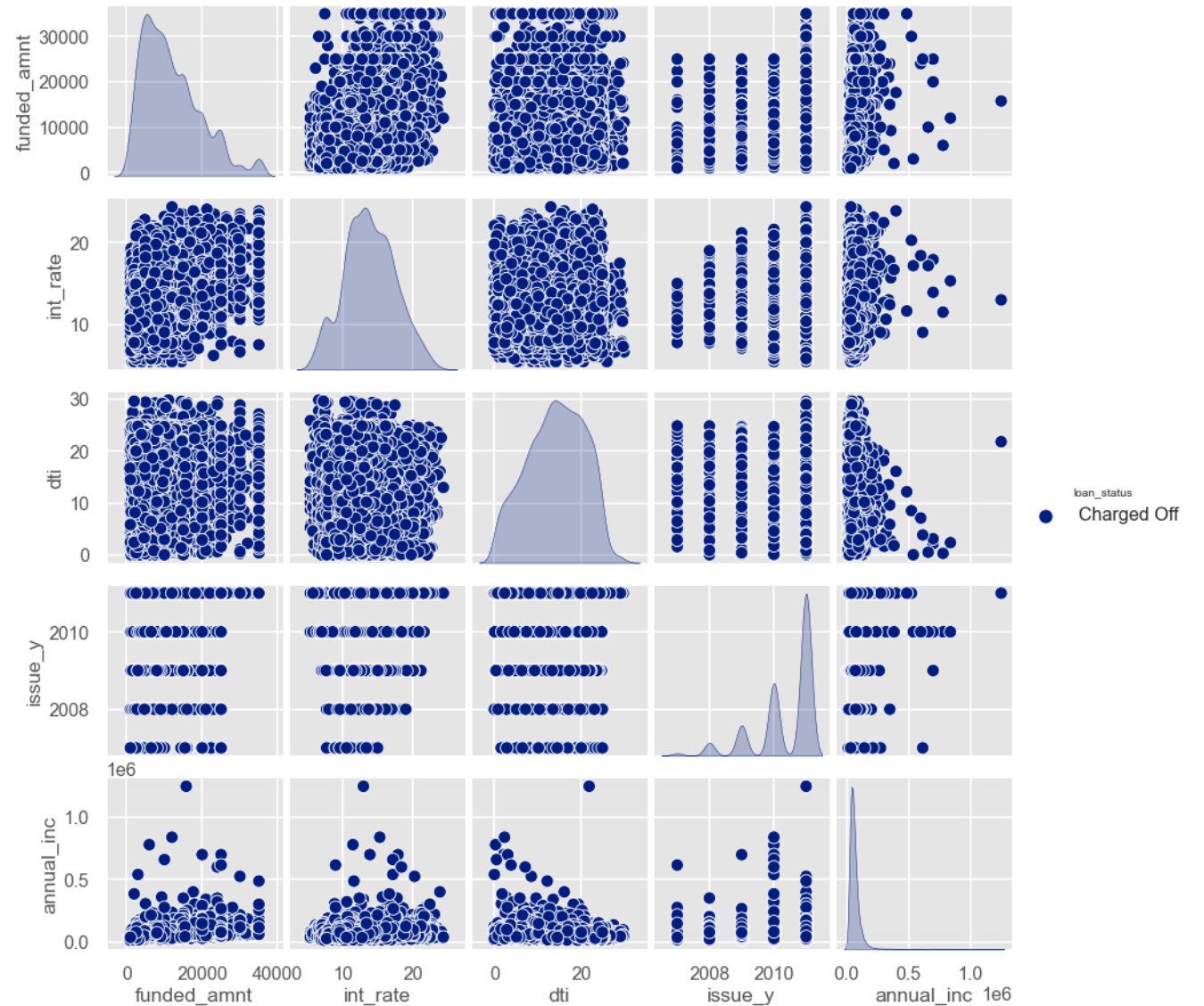
Conclusion

- Analysis of the Loan data set provides insights on how Exploratory Data Analysis can be used to study the different variables in the dataset, their characteristics, relationship with other variables.

- It also programmatically shows how to derive meaningful insights from the data and how data visualization influences business decision making