Lending Club Case Study

Analyze Loan Lending Risks

- Saravanan Ponmudi
- Sandeep Sonawane

Agenda

- About Case Study
- EDA Approach
- Technologies Involved
- Conclusive Summary

About Case Study

- Aiming to find out the applicants at risk of defaulting
- Enable reduction in credit losses

EDA Approach

- Data Understanding
- Data Cleaning & Manipulations
- Univariate & Segmented Univariate Analysis
- Bivariate Analysis

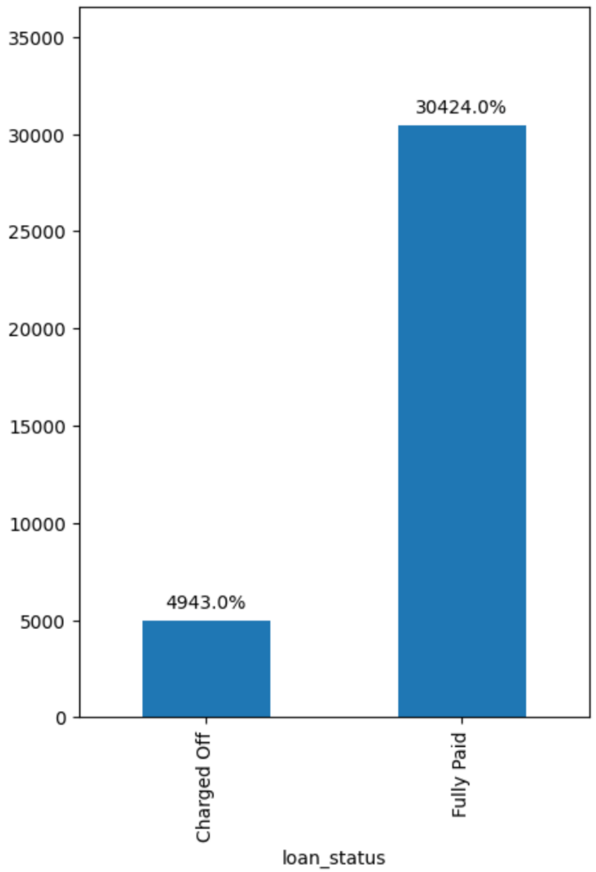
Data Understanding

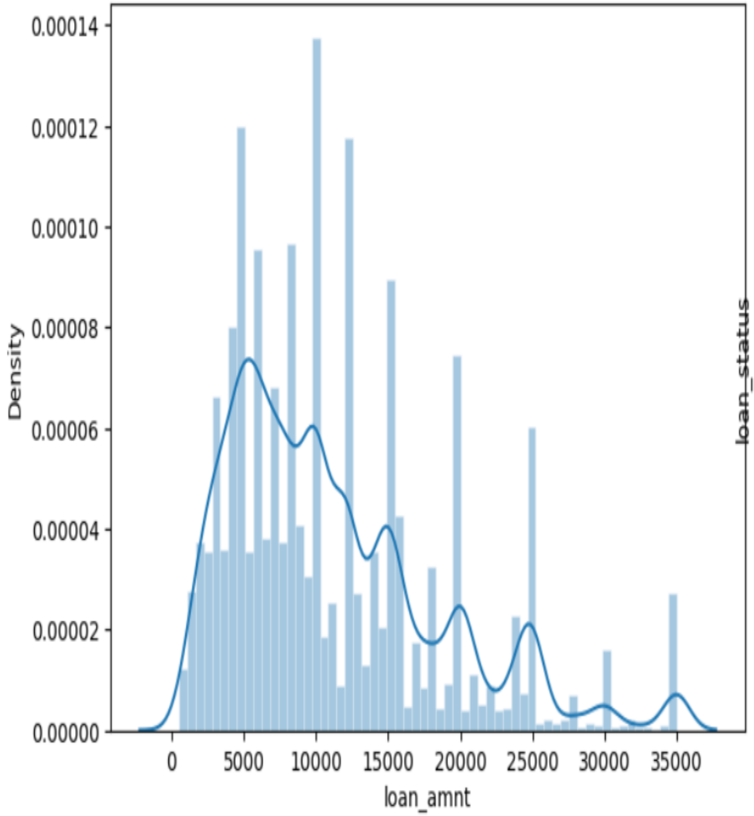
- Data loading & sanity
- Understanding data dictionary
- Knowing all columns
- Understanding domain uses of columns

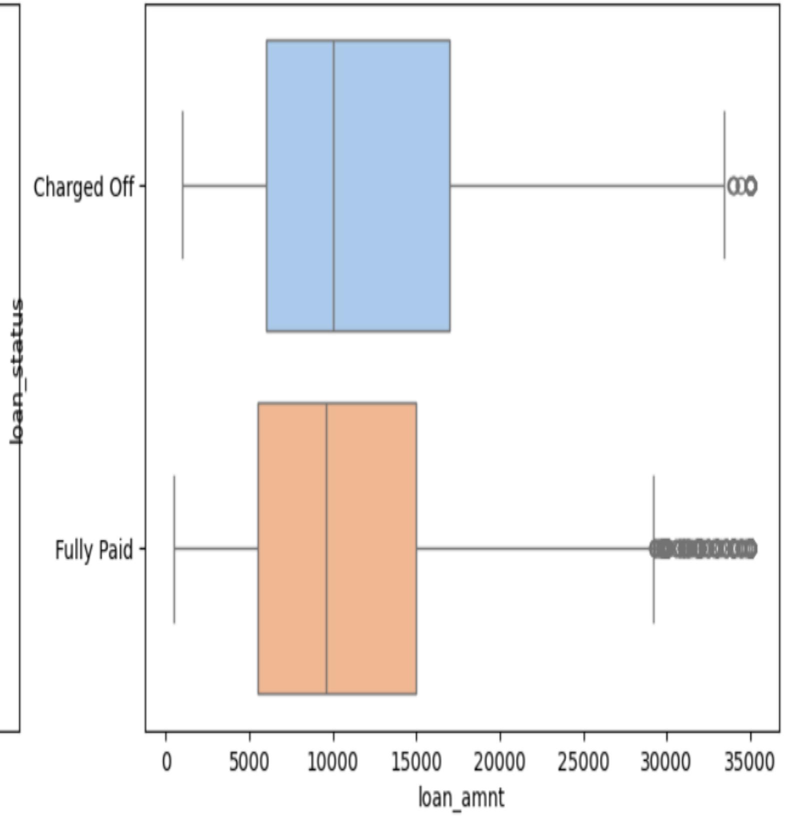
Data Cleaning & Manipulations

- Remove null valued columns
- Remove unique valued columns
- Remove unnecessary columns
- Understanding domain uses of columns
- Column format conversions
- Remove outliers
- Derive new attributes

Univariate & Segmented Univariate Analysis…



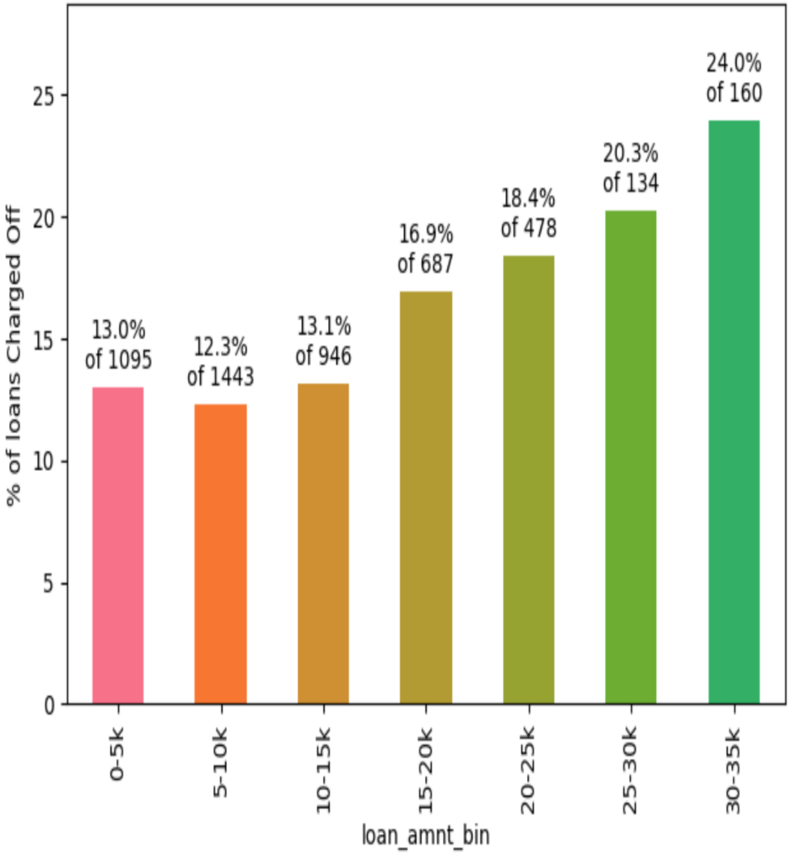


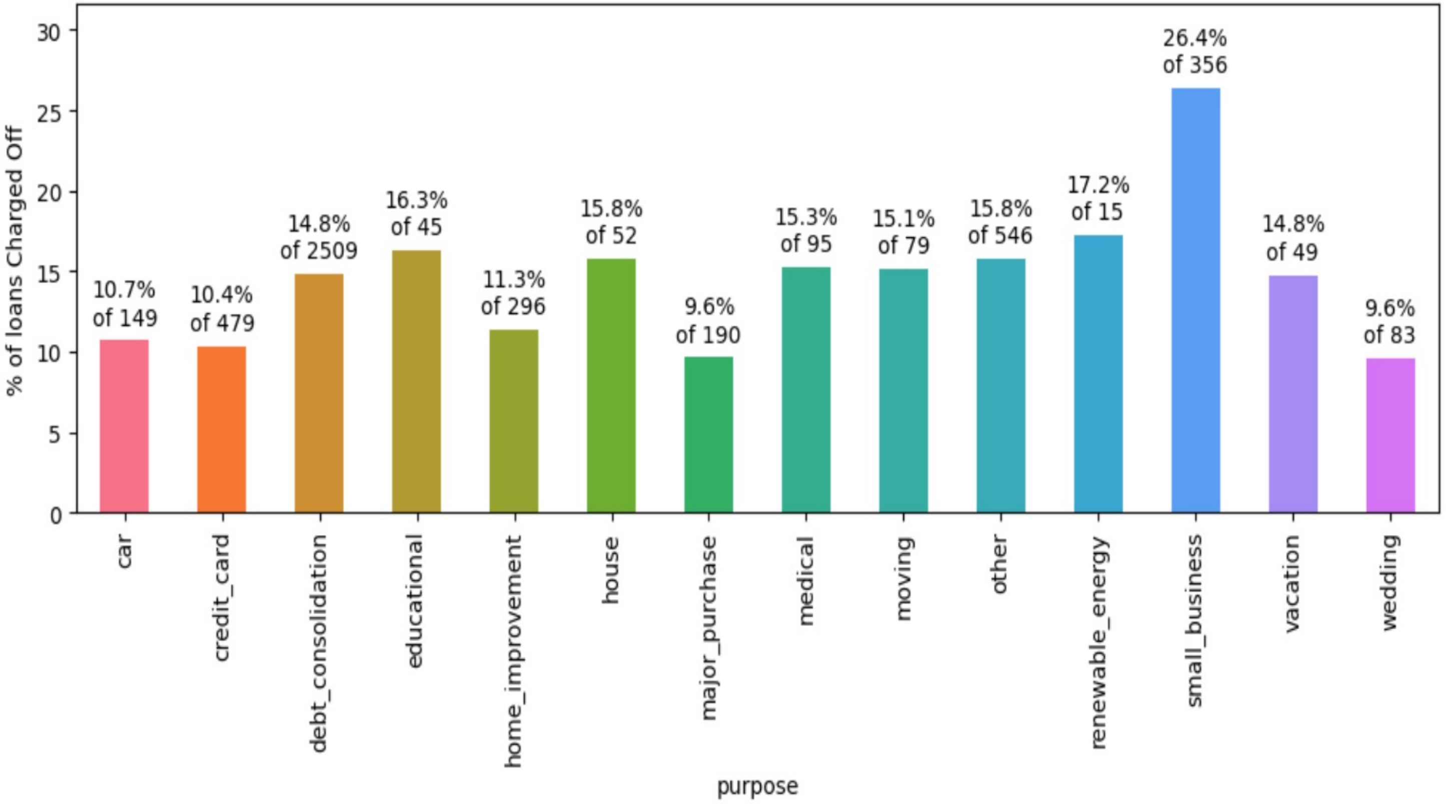


- Lending Club only recovers 57% of the loan amount when loans are defaulted. On fully paid-up loans, the company makes 17% profit.

- 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)
- Funding amounts see a spike around each 5000 boundary. We will use 5000 as bucket size for later analysis.
- Charged off loans are shifted towards higher average loan amount request.

Univariate & Segmented Univariate Analysis…

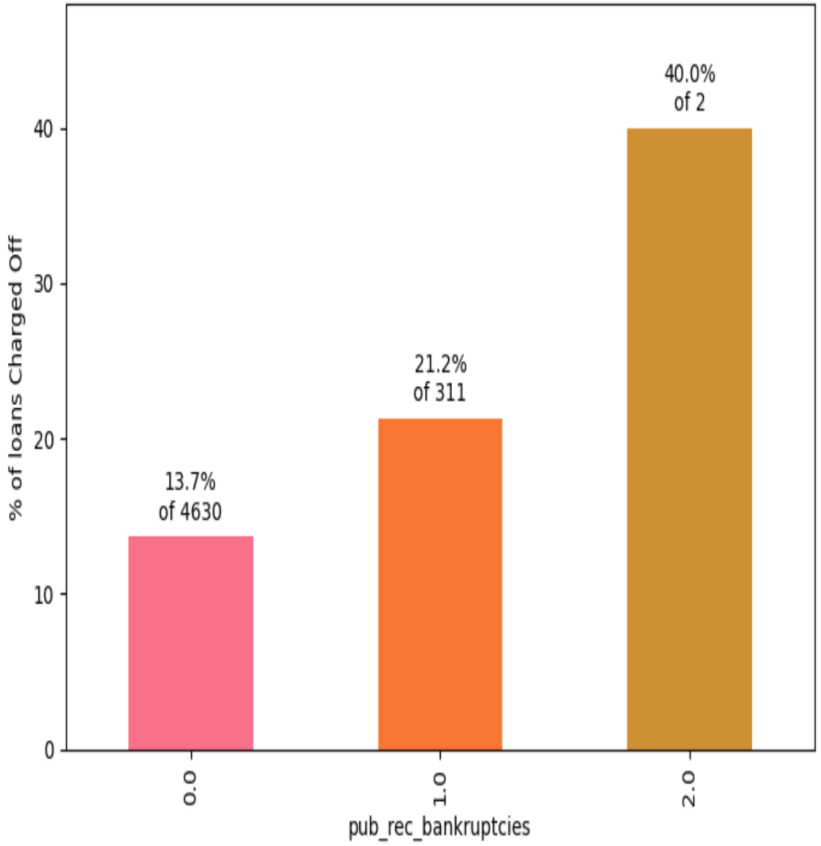


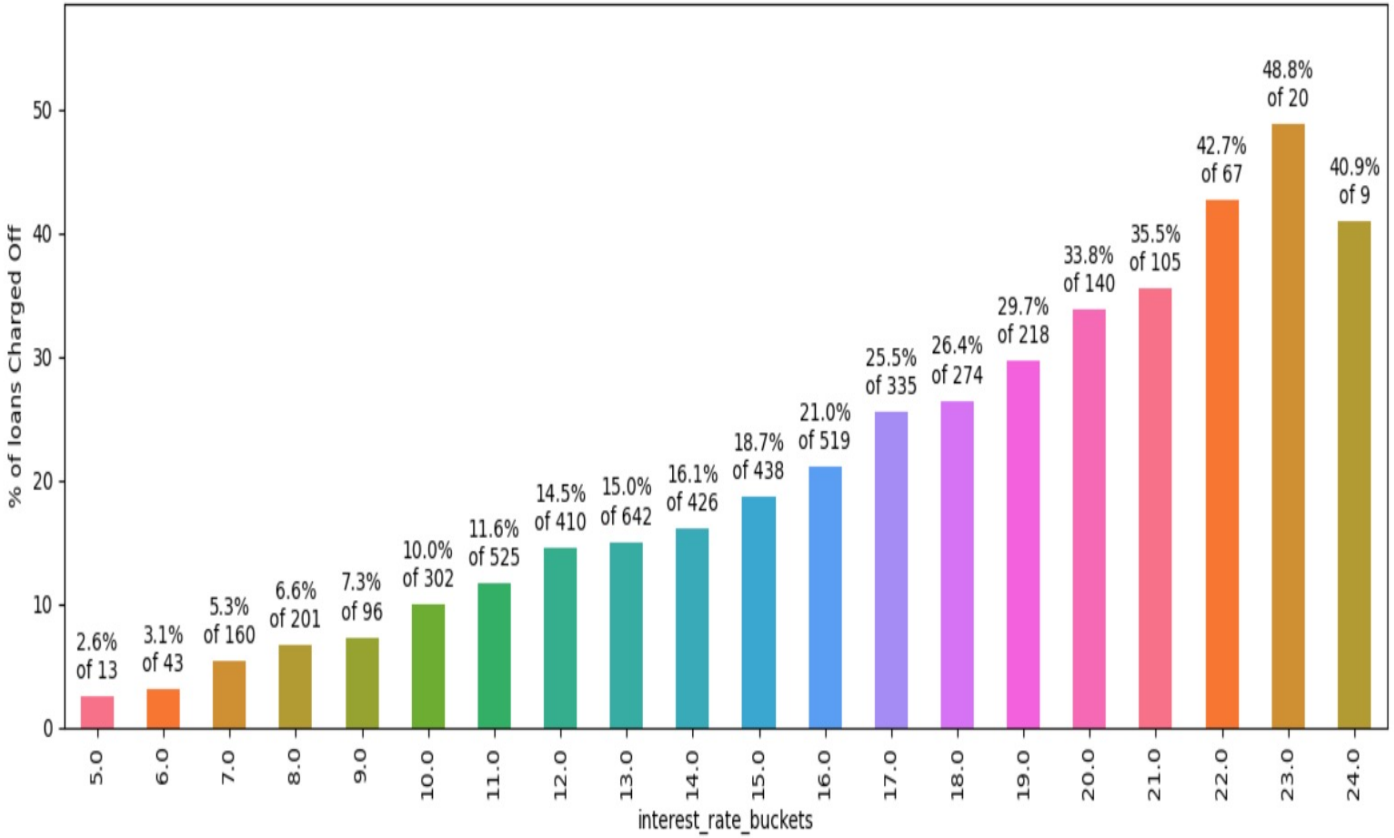


- **The % of charged off loans increases substantially as we go up the loan amount buckets. Most loans are below 20000 amount. The higher loans, though lesser in number, carry a substantially higher risk of default.**

- **26% of loans for small business are Charged Off. Making them the riskiest purpose.**
- **Approximately ~49% of the loans are issued for the purpose of dept consolidation.**
- **17% of the loans for renewable_enrgy are charged Off, but the number is too less to be of significance.**

Univariate & Segmented Univariate Analysis…

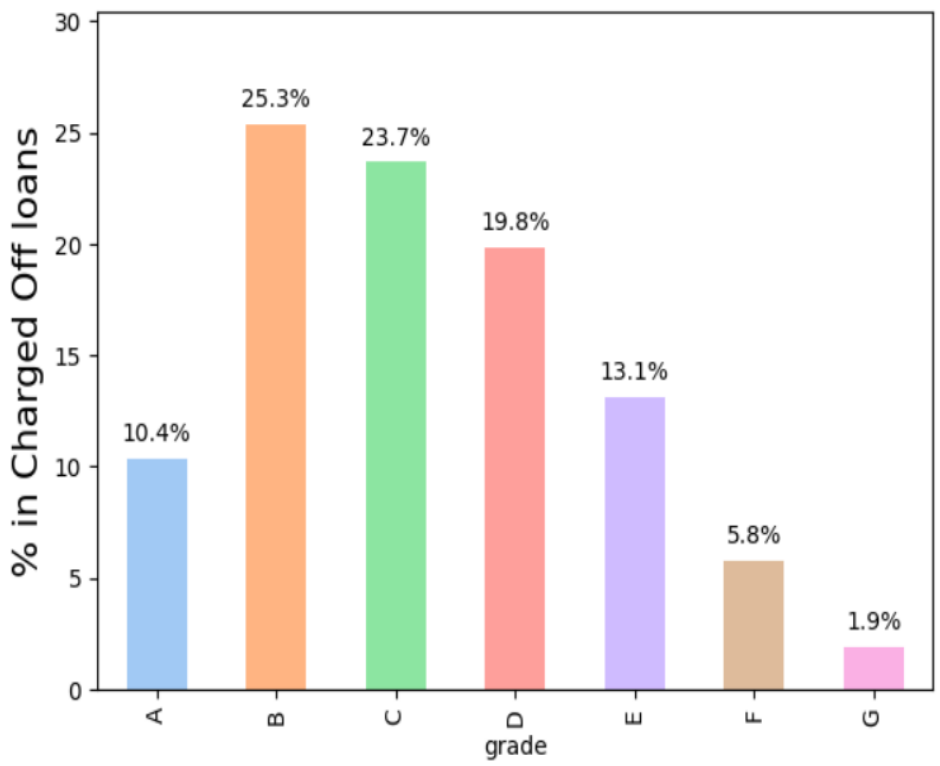


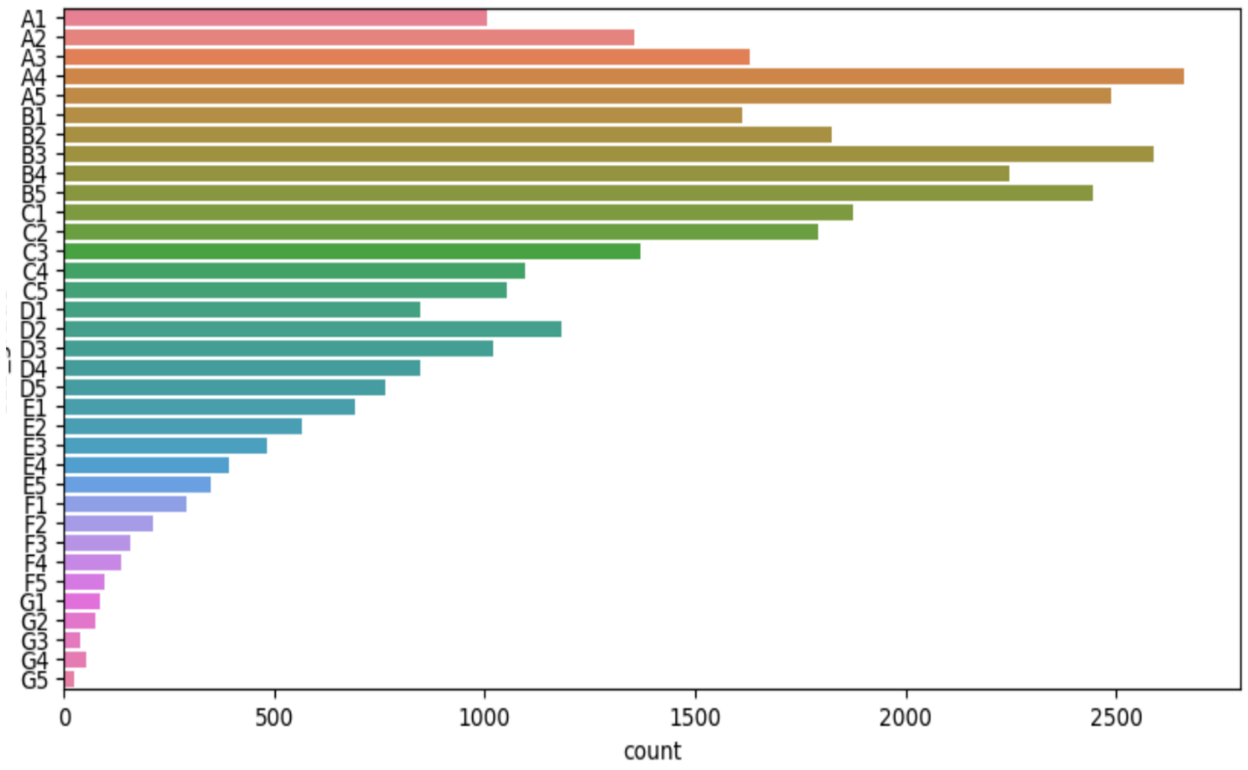


- The percentage of Charged Off loans is markedly higher when the borrower has a prior record of bankruptcy.

- The Charged Off loans are proportionate with interest rates

Univariate & Segmented Univariate Analysis…

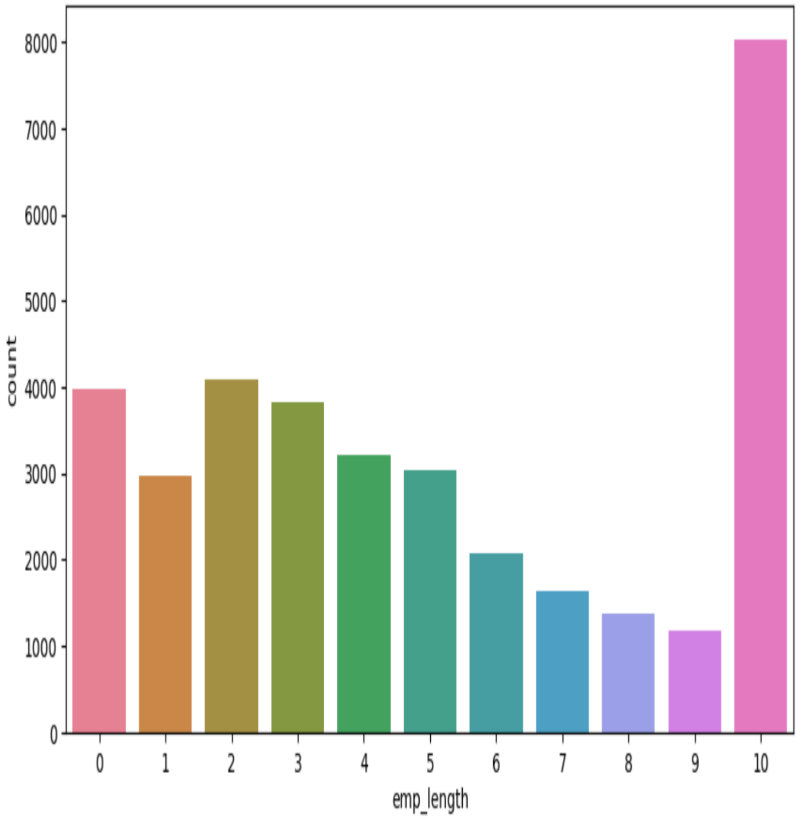


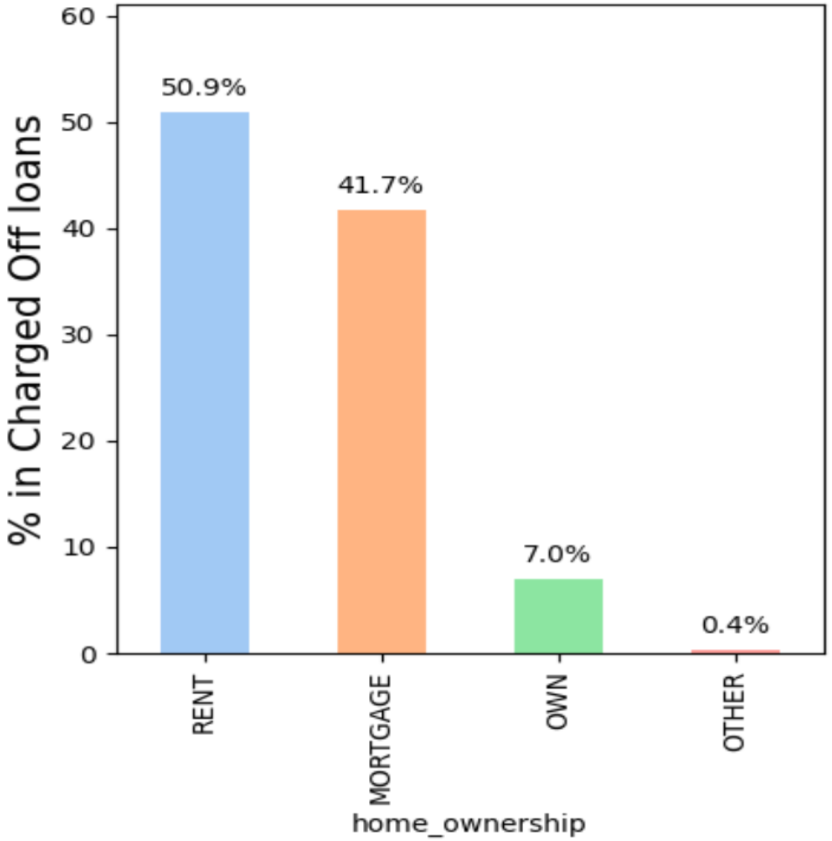


- Applications in Grade B,C & D contribute most to Charged-Off loans

- **Sub-Grade A3,A4,A5 & B3,B4,B5 loans most likely to default**

Univariate & Segmented Univariate Analysis…

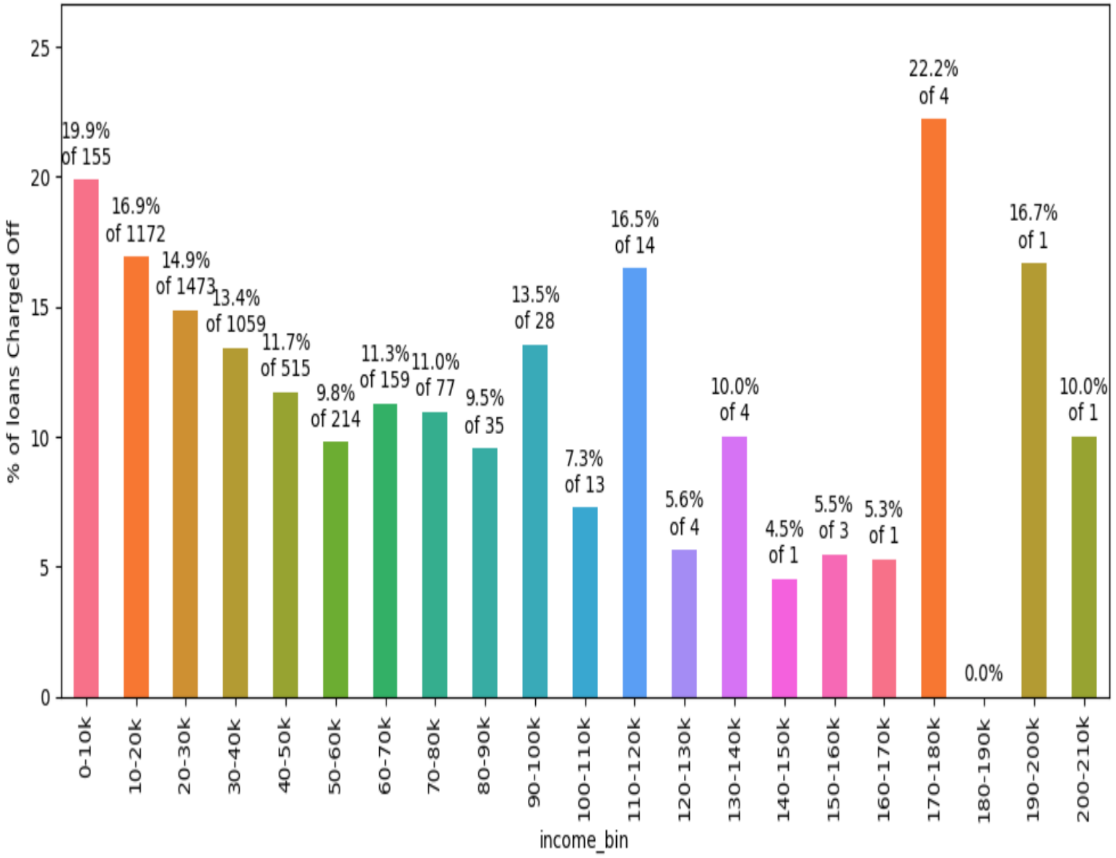


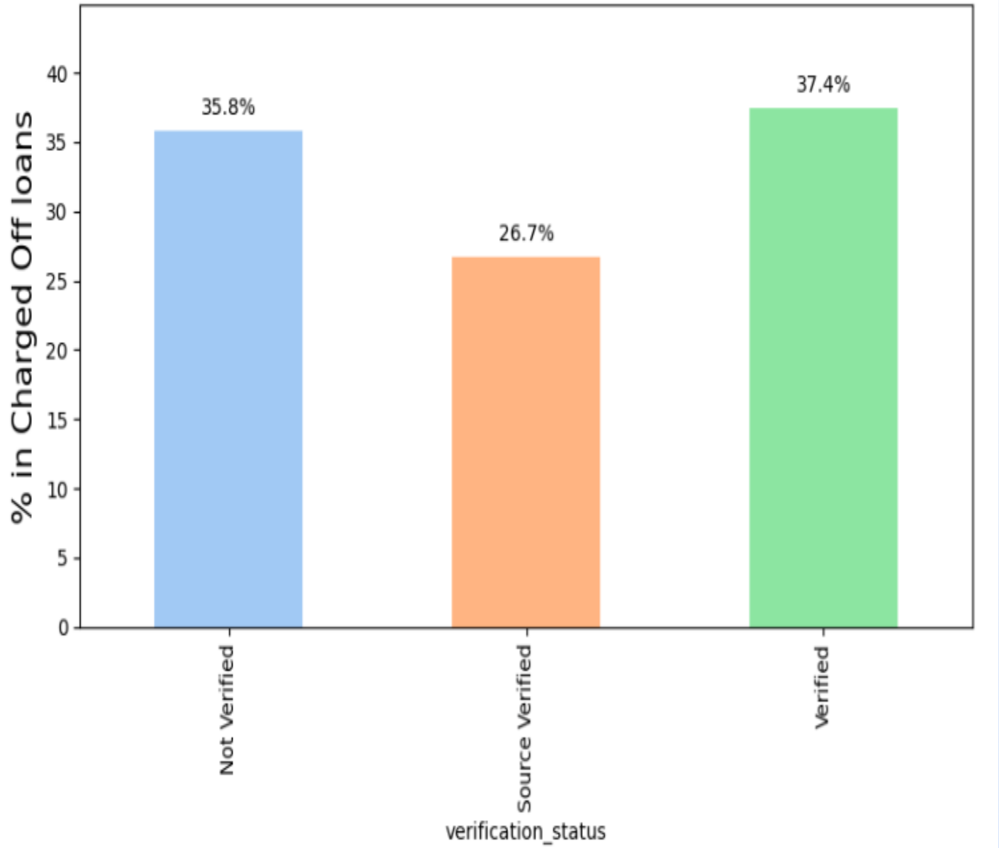


- High experience applicants i.e.10+ years trends most likely to default

- Applicants living in Rented or Mortgaged Houses most likely to default

Univariate & Segmented Univariate Analysis…

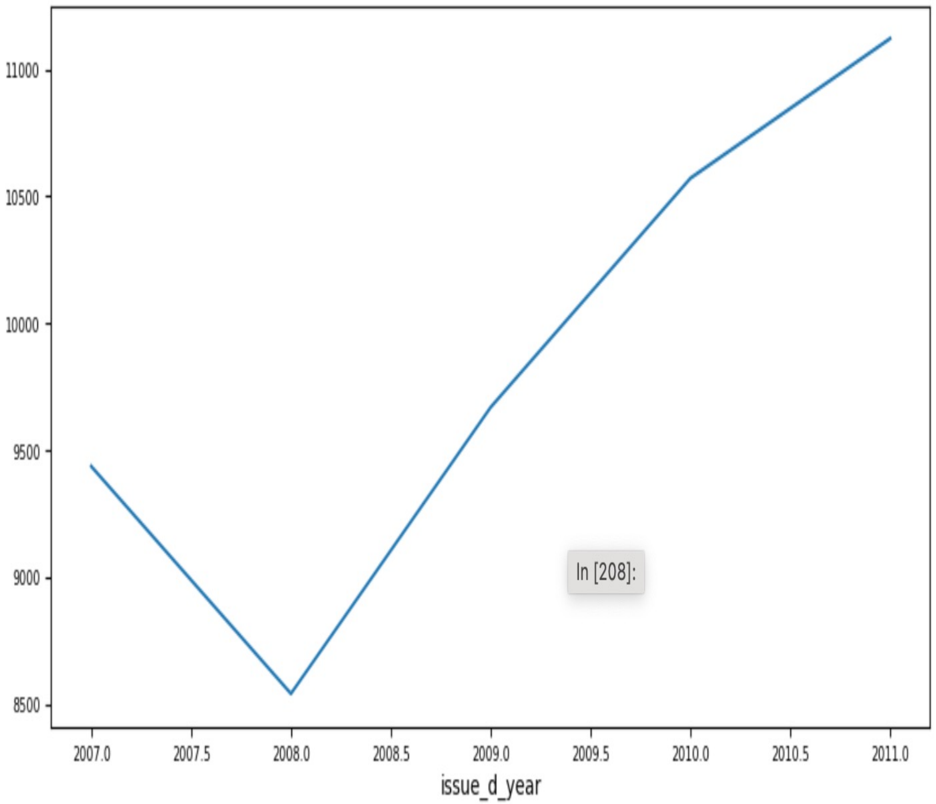


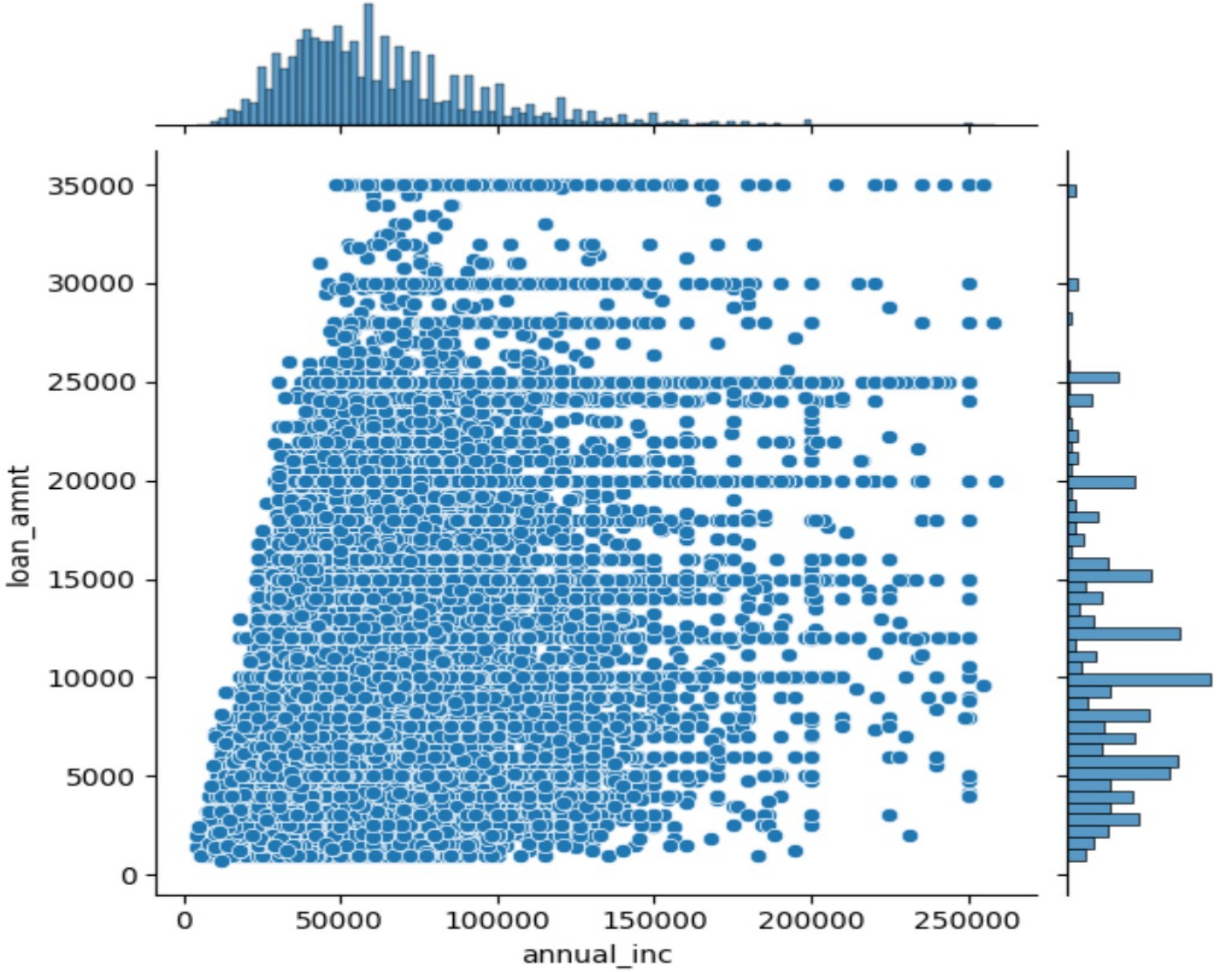


- Low annual incomes (< $40,000) shows defaulting trends

- Verified loans show more charged Off percentage as compared to Not verified loans.

Bivariate Analysis

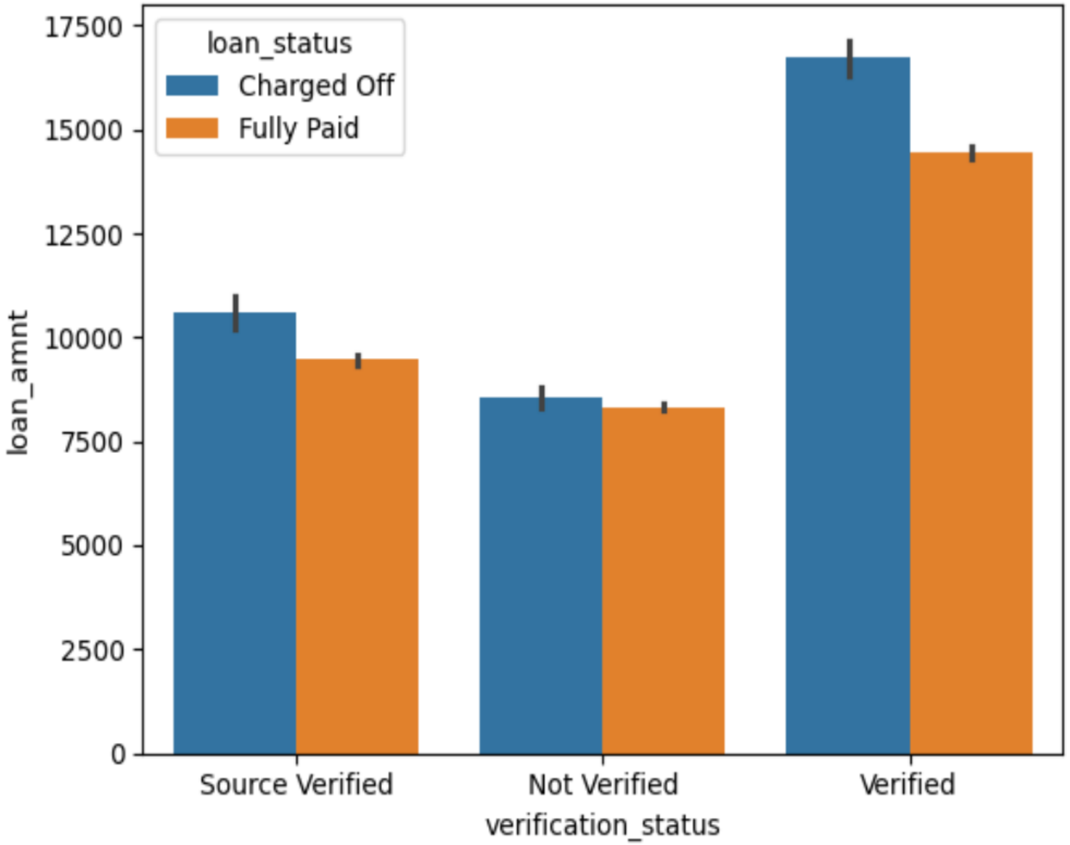


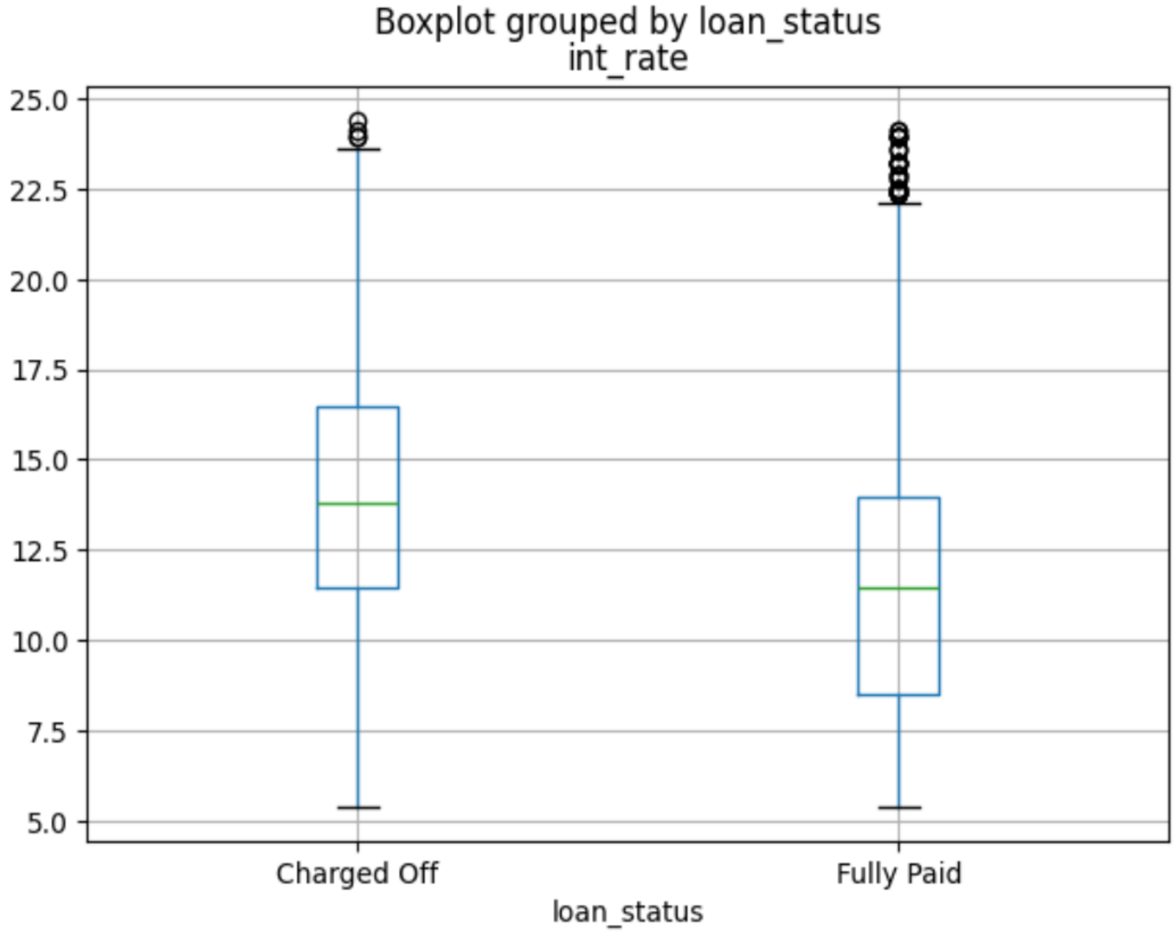


- Trend from 2008 to 2011 indicates growing loan pattern in market

- People with average income lower than 50000 taking loans of 25000 or higher, which is risk for loan lending

Bivariate Analysis





- **Higher loan amounts are Verified more often, but see higher charged-off rate**

- **Loans at a higher interest rate are more likely to be Charged Off.**

Technologies Involved

- Python 3x
- Jupytor Notebook
- Pandas
- Numpy
- Mathplotlib
- Seaborn

Conclusive Summary

- Charged-Off Observations
- Growth Trend
- Seasonal Trends
- Verification Process

Conclusive Summary...

- Charged-Off loan application observations
  - Applications in Grade **B**,**C** & **D** contribute most to Charged-Off loans
  - Sub-Grade **A3,A4,A5** & **B3,B4,B5** loans most likely to default
  - **60 months** term loans most likely to default
  - High experience **10+ years** trends most likely to default
  - Applicants living in **Rented** or **Mortgaged Houses** most likely to default
  - Applicants from **California (CA)**, **Florida (FL)** & **New York (NY)** shows default trend
  - Risk in offering loan for **$15,000 or more** amount
  - **High DTI ratios** with **interest rates in 13-17%** range are associated with defaults
  - **Low annual incomes (< $40,000)** shows defaulting trends
- Trend from **2008** to **2011** indicates **growing loan pattern** in market
- Seasonal Trend in December and Q4 shows peak in applications, may be due to holidays
- Verification Process shows less Charged-Off scenarios