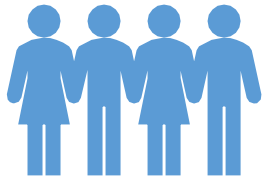
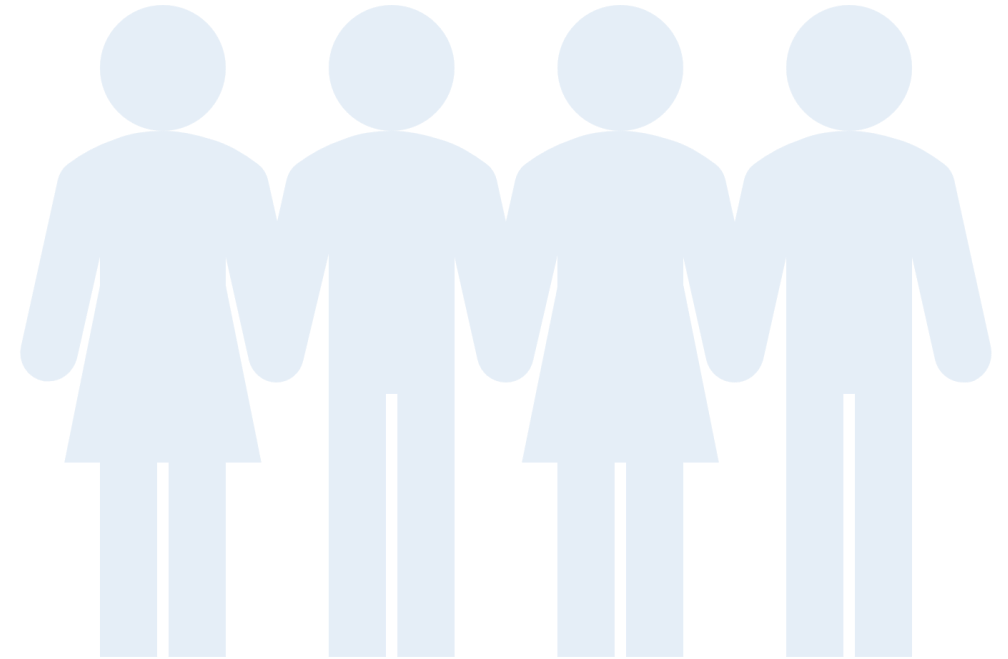


Group Case Study



Lending Club



Members:

1. Saurav Mishra (Group Facilitator)
2. Debasis Garabadu
3. Sandhita Agarwal
4. Senjuti Kar

EDA Methodology Followed

Business Objectives:

Driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of charged-off or default

Data Analysis Goals:

- Univariate Analysis
- Bivariate Analysis

Data Understanding:

Data Available:

- Loan data Set
- Loan Data Dictionary

Data Preparation and Analysis:

- Data Sourcing
- Data Cleaning
- Derived Metrics
- Univariate Analysis
 - Unordered Categorical
 - Ordered Categorical
 - Quantitative
- Segmented Univariate Analysis
- Bivariate Analysis
 - Categorical
 - Quantitative

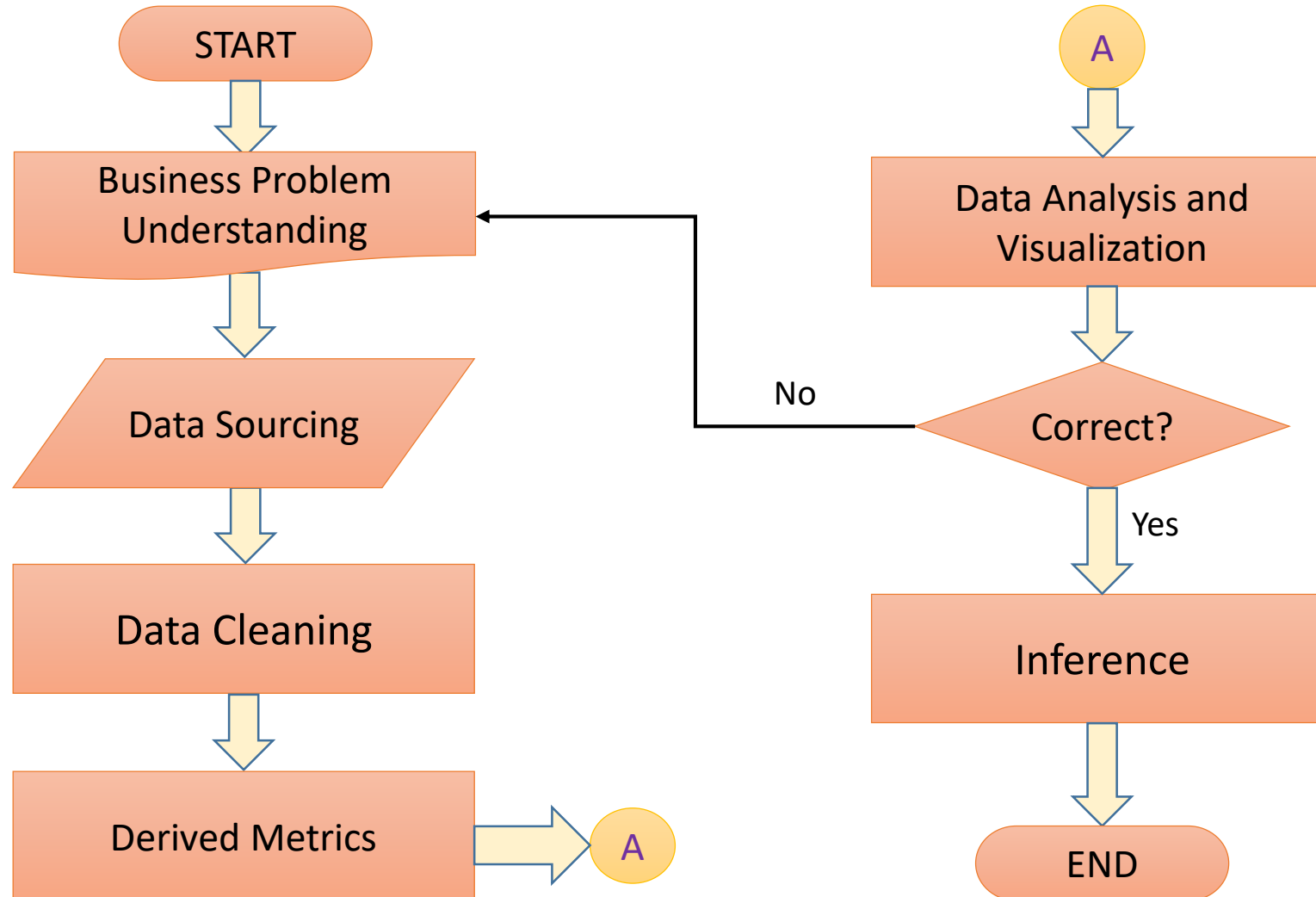
Data Visualization:

- Line Plots
- Bar Plots
- Distribution Plots
- Box Plots
- Joint / Combined Plots (Line + Bar)
- Cluster Map

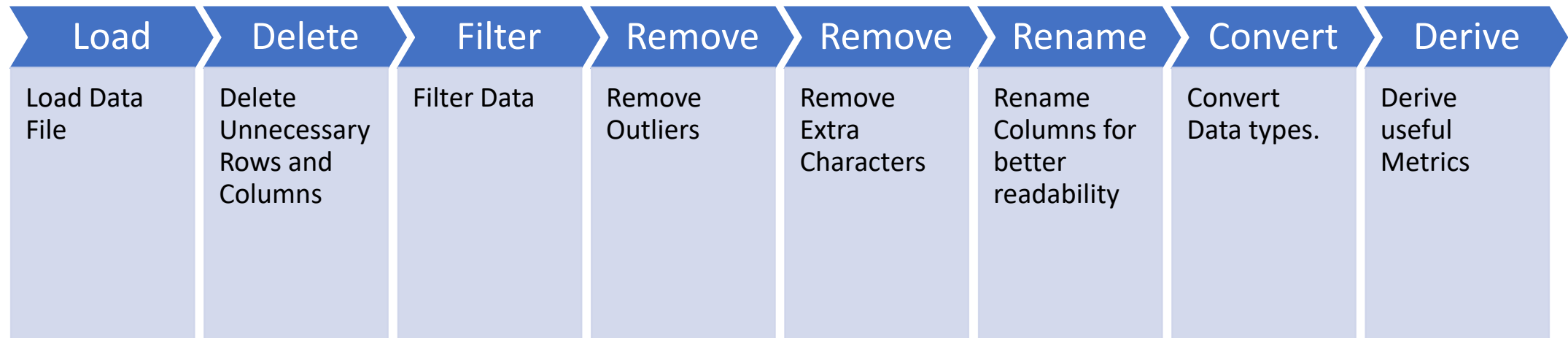
Recommendation

- Top 5 factors

Problem Solving Methodology



Data Sourcing, Cleaning & Derived Metrics



Univariate Analysis

- Unordered Categorical Variables

- Loan Status

- ✓ 5584 observations are Charged Off

- Home Ownership

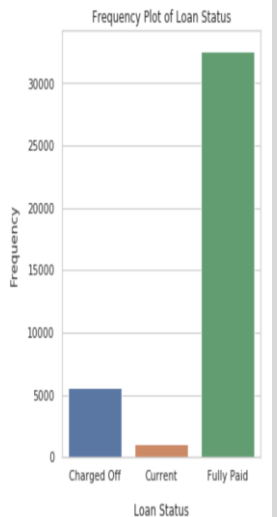
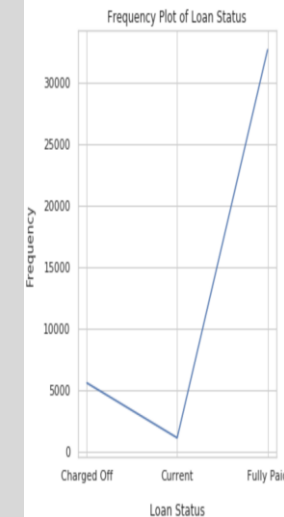
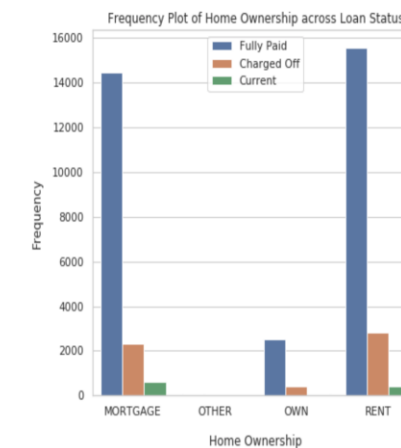
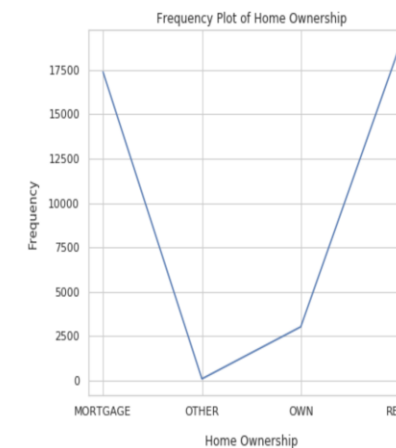
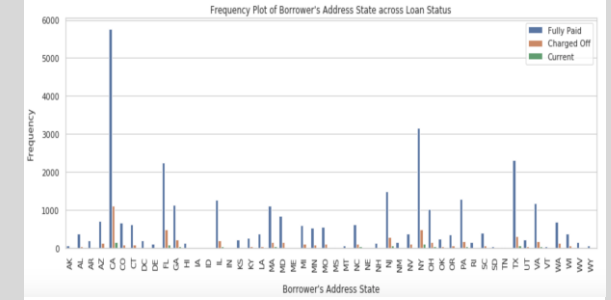
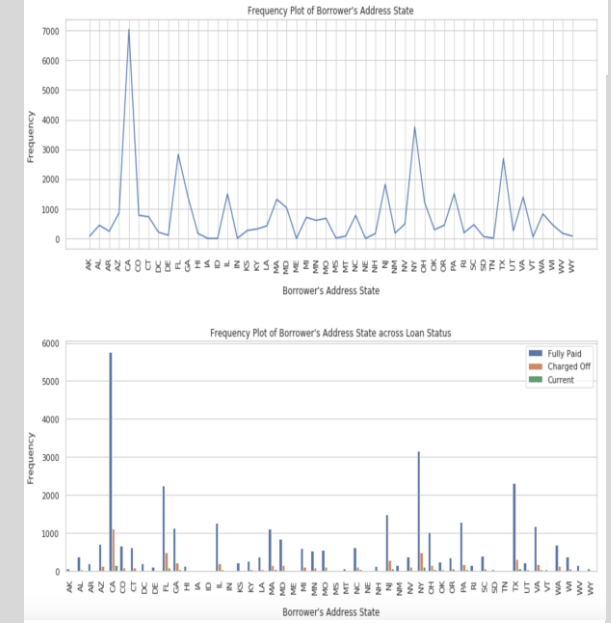
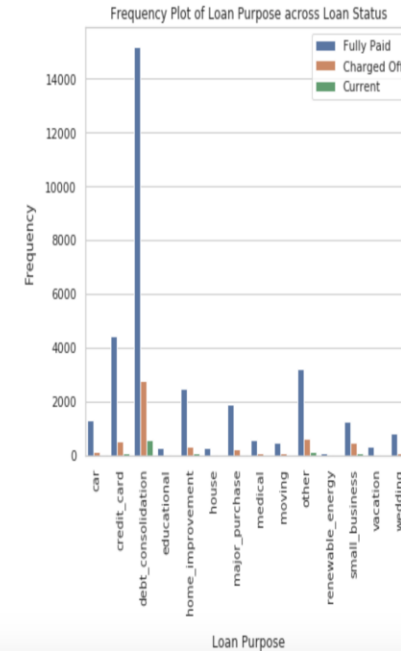
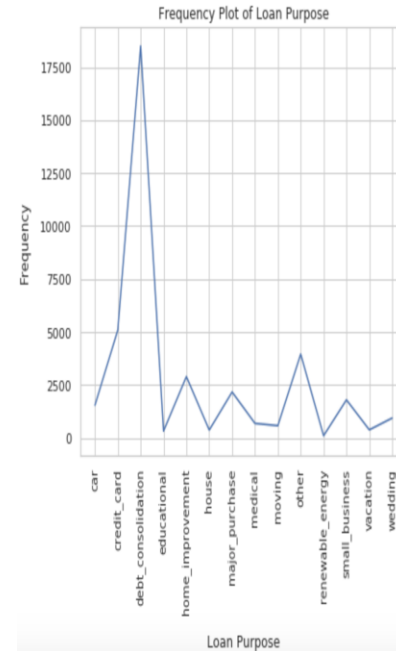
- ✓ Top 2 loan purpose of Charged off:
 - RENT
 - MORTGAGE

- Loan Purpose

- ✓ Top 3 loan purpose of Charged off:
 - debt_consolidation
 - other
 - credit_card

- Address State

- ✓ Top 3 Address States of Charged off:
 - CA
 - NY
 - FL



Univariate Analysis

- Ordered Categorical Variables

- Term

- ✓ Most of the Charged off applicants apply for 36 months Loan Term

- Grade & Sub-Grade

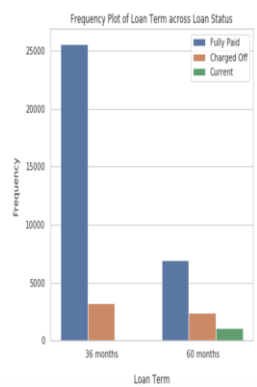
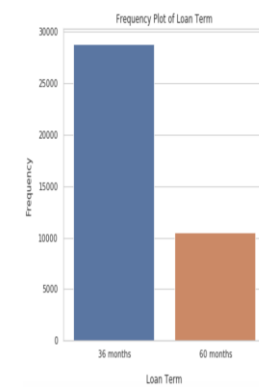
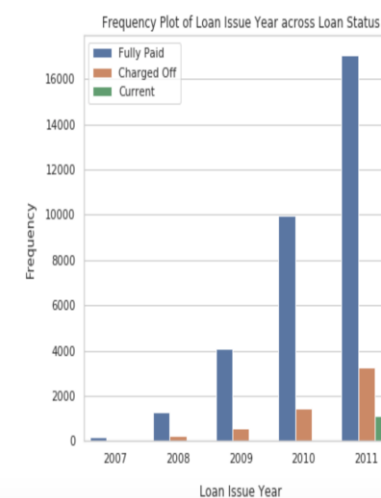
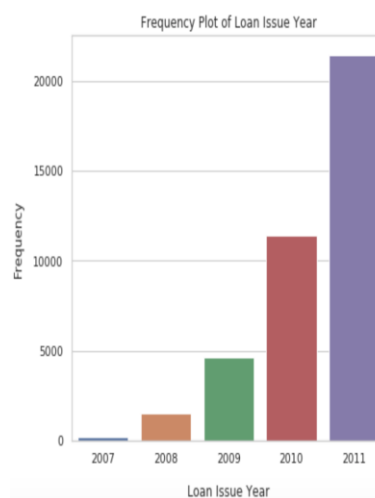
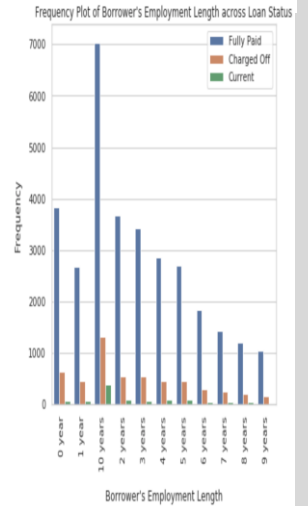
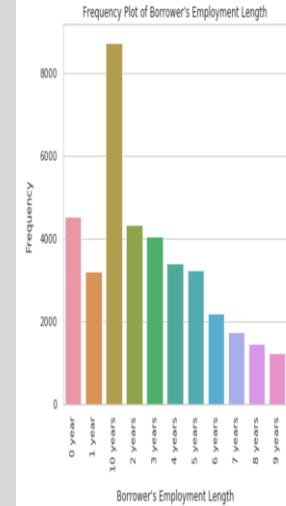
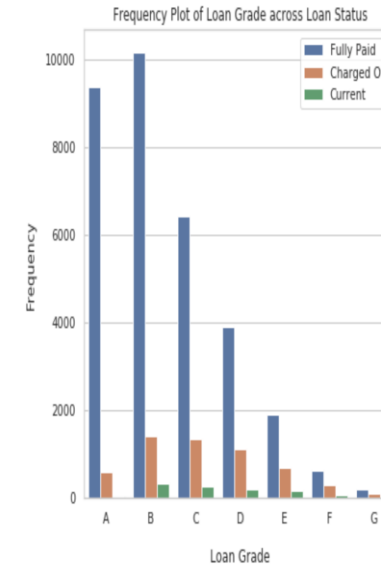
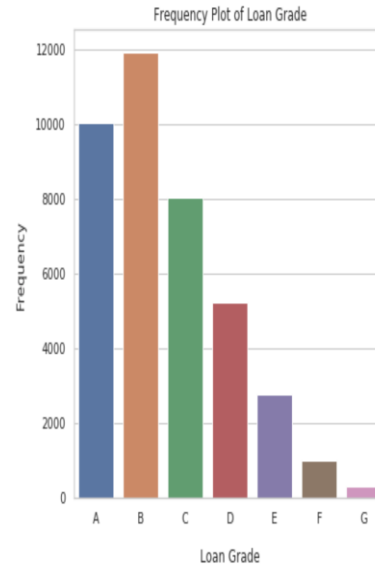
- ✓ Most of the Charged off applicants belong to Grade B, C and D

- Employment Length

- ✓ Top 3 for Charged off:
 - < 1 year
 - 1 year
 - 10+ years

- Issue Year

- ✓ Charged Off Applicants are more in the Loan issued year 2011.



Univariate Analysis

- Quantitative Variables

- Funded Amount

- ✓ Charged Off Applicants has a central tendency as 10,000 with minimum 25th percentile value as 5500 and maximum 75th percentile value as 16000.

- Interest Rate Percent

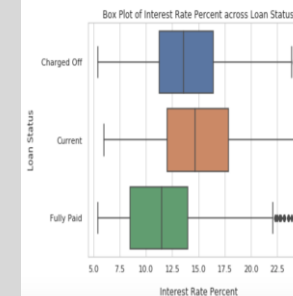
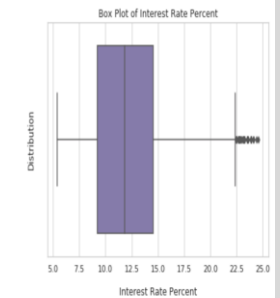
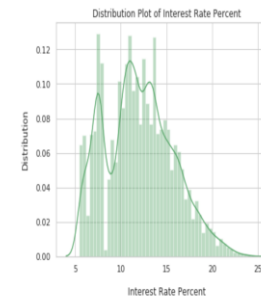
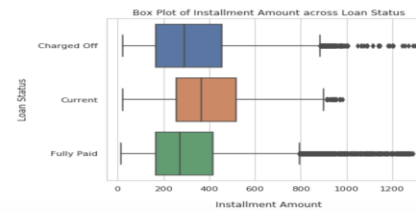
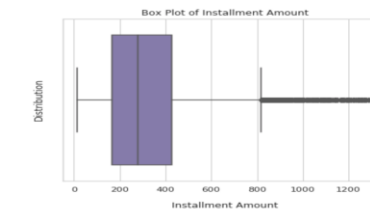
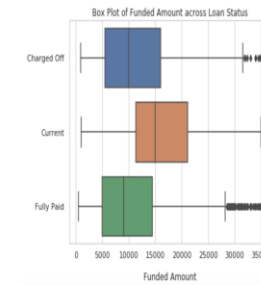
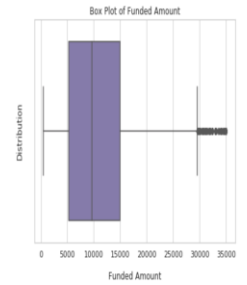
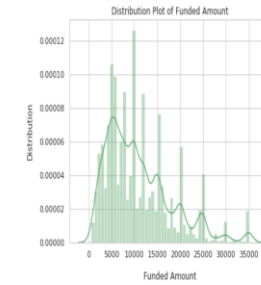
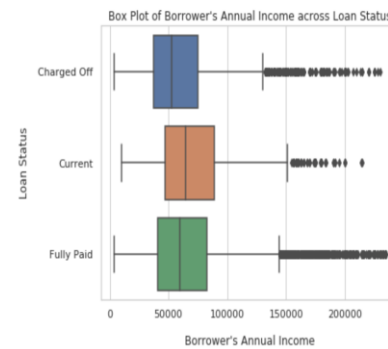
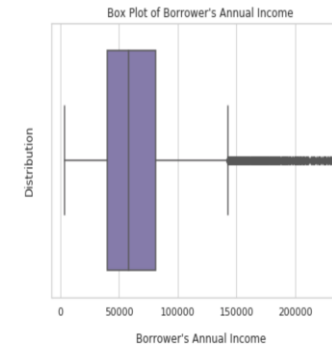
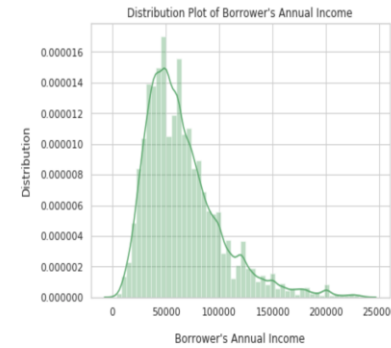
- ✓ Charged Off Applicants has a central tendency as 13.57% with minimum 25th percentile value as 11.28% and maximum 75th percentile value as 454.38

- Installment

- ✓ Charged Off Applicants has a central tendency as 292.04 with minimum 25th percentile value as 168.45 and maximum 75th percentile value as 454.38

- Annual Income

- ✓ Charged Off Applicants has a central tendency as 52,800 with minimum 25th percentile value as 37,000 and maximum 75th percentile value as 74,879.



Bivariate Analysis

• Categorical Variables

➤ Charged Off Rate - Top Loan Term

- ✓ 60 months = 22.70%

➤ Charged Off Rate - Grade & Sub-Grade

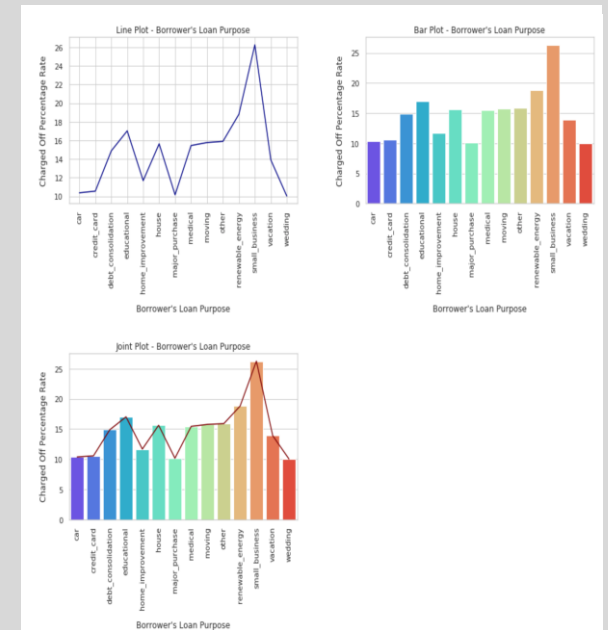
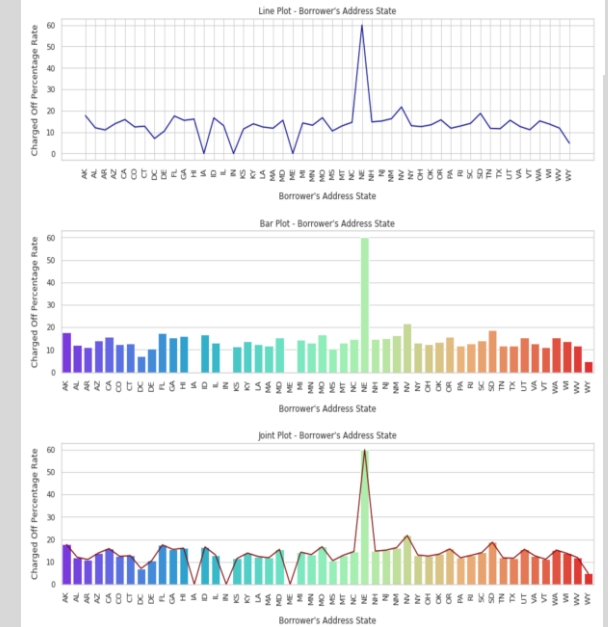
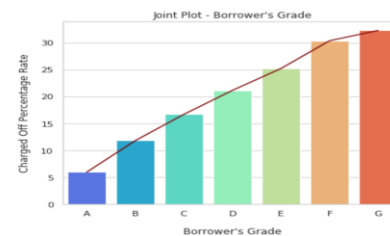
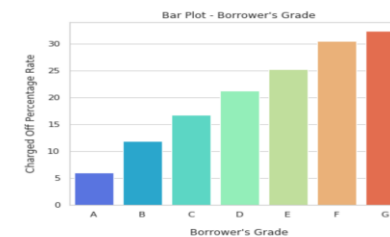
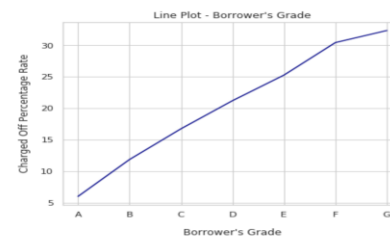
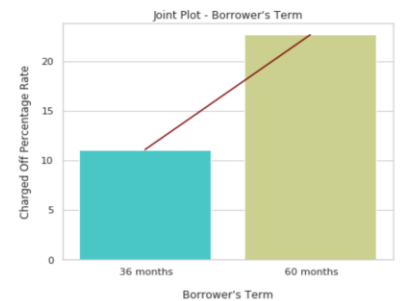
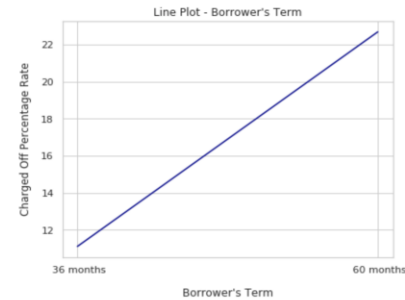
- ✓ As the Grade and the Sub Grades increases, Charged Off Rate increases

➤ Charged Off Rate - Top 5 Address State

- ✓ NE (Nebraska) = 60.00%
- ✓ NV (Nevada) = 21.75%
- ✓ SD (South Dakota) = 18.75%
- ✓ AK (Alaska) = 17.72%
- ✓ FL (Florida) = 17.54%

➤ Charged Off Rate - Top 3 Loan Purpose

- ✓ small_business = 26.27%
- ✓ renewable_energy = 18.81%
- ✓ educational = 17.03%



Bivariate Analysis

- Categorical Variables

- Charged Off Rate - Top Employment Length

- ✓ 1 year
- ✓ 0 year (< 1 year)
- ✓ 10 Years (> 10 years)

- Charged Off Rate – Annual Income

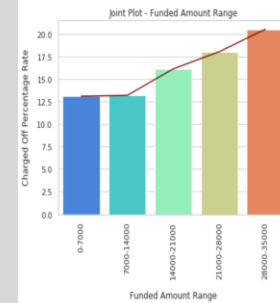
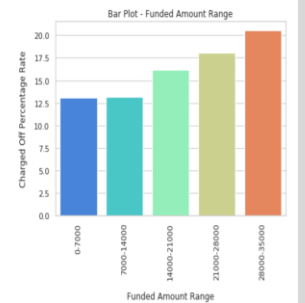
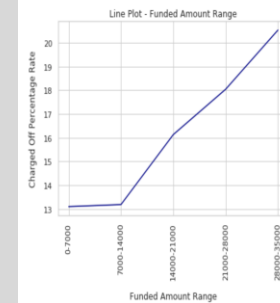
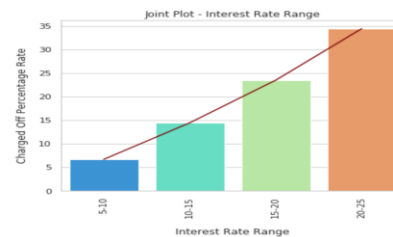
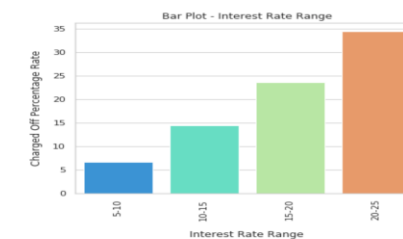
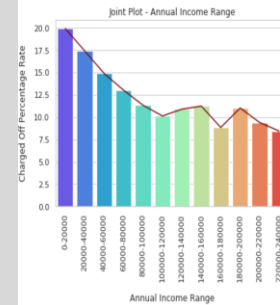
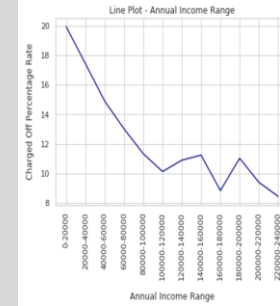
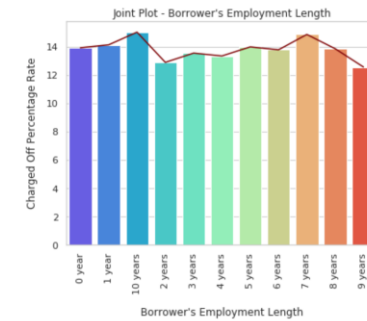
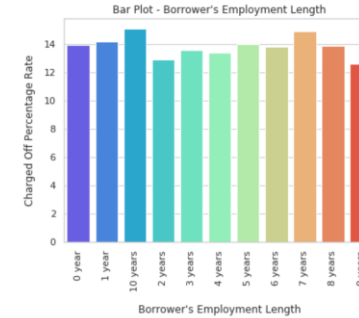
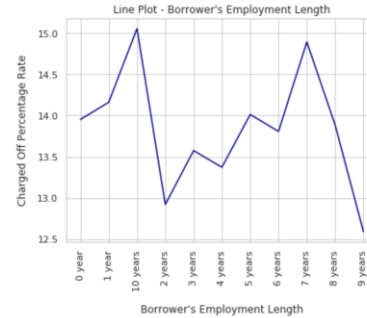
- ✓ As the Annual Income decreases, Charge Off Rate Increases

- Charged Off Rate – Interest Rate

- ✓ As the Interest Rate increases, Charge Off Rate Increases

- Charged Off Rate – Funded Amount

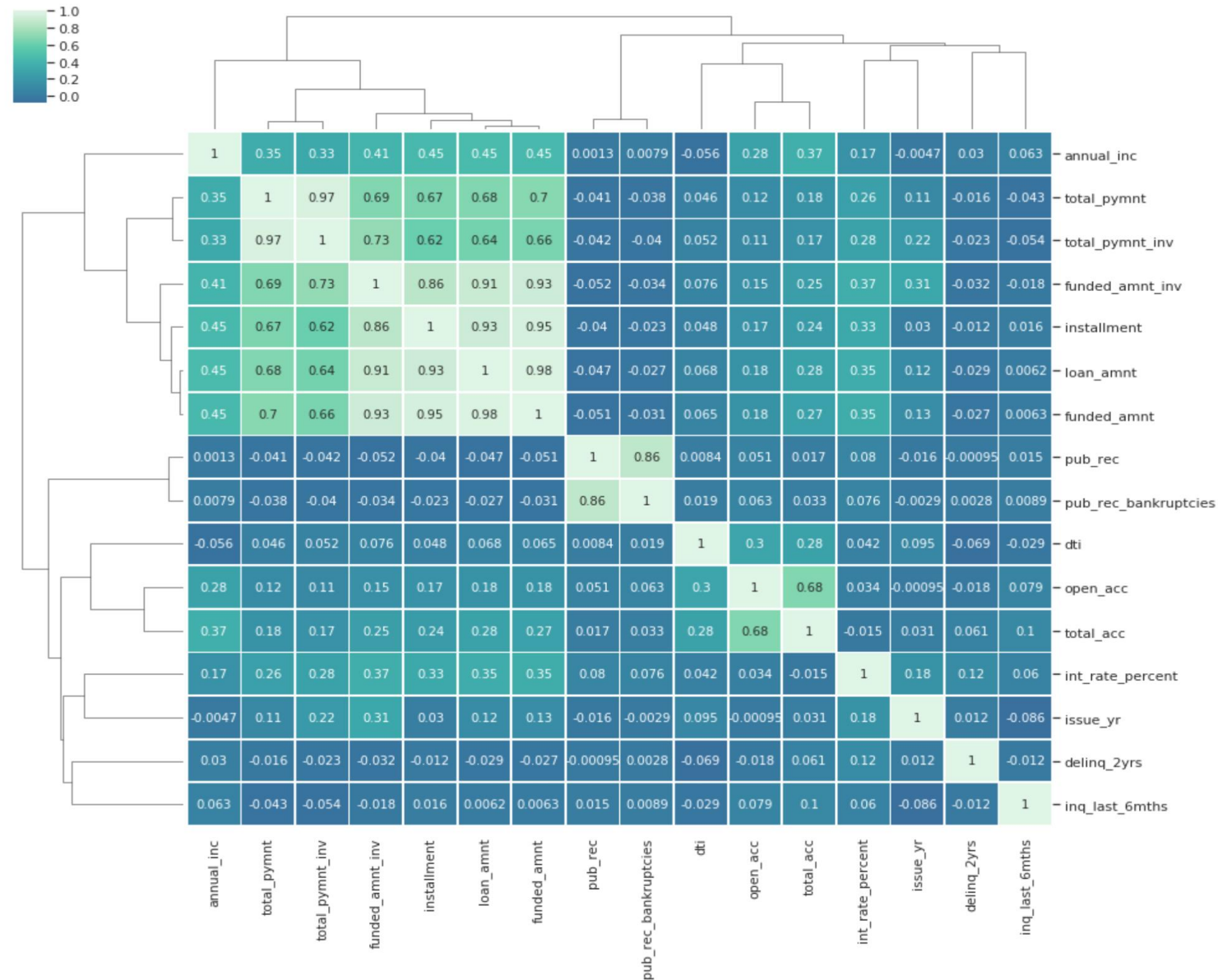
- ✓ As the Funded Amount increases, Charge Off Rate Increases



Bivariate Analysis

Quantitative Variables

- installment, funded_amnt, loan_amnt and funded_amnt_inv are highly correlated (positive) to each other. They form a cluster.
- dti (Debt-to-Income Ratio) is high correlated (positive) to issue_yr and vice-versa.
- int_rate_percent is negatively correlated to total_acc.



Conclusion

**Major
variables
to
consider
for loan
prediction**

Top Five Variables

- Loan Purpose
- Employment Length
- Interest Rate
- Grade & Sub-grade
- Term

Additionally the following
variables can also be
considered

- Annual Income
- Address State