Lending Club Case Study

Kriti Pawar, Pooja Sonkar

Business Objective

- LendingClub company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

- Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who **default** cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

- If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

- In other words, the company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Data Understanding

- Loan data has **111** variables.
- Number of rows are **39717.**
- Overall Missing cell percentage is **51.3%.**
- Total Numeric variables are **25**
- Total Categorical variables are **84**
- Total Boolean variables are **2**

Data Cleaning

- **Removed** all the variables which has **more than 10%** missing values.
- **Updated** the values of few features by removing str literals i.e i**nterest_rate**, **dti** etc.
- **Derived new features** from the existing ones i.e from **issued_d** extracted **year** and **month** etc
- **Removed** few **customer behaviour variables**.

Data Analysis

Univariate Data Analysis

Purpose
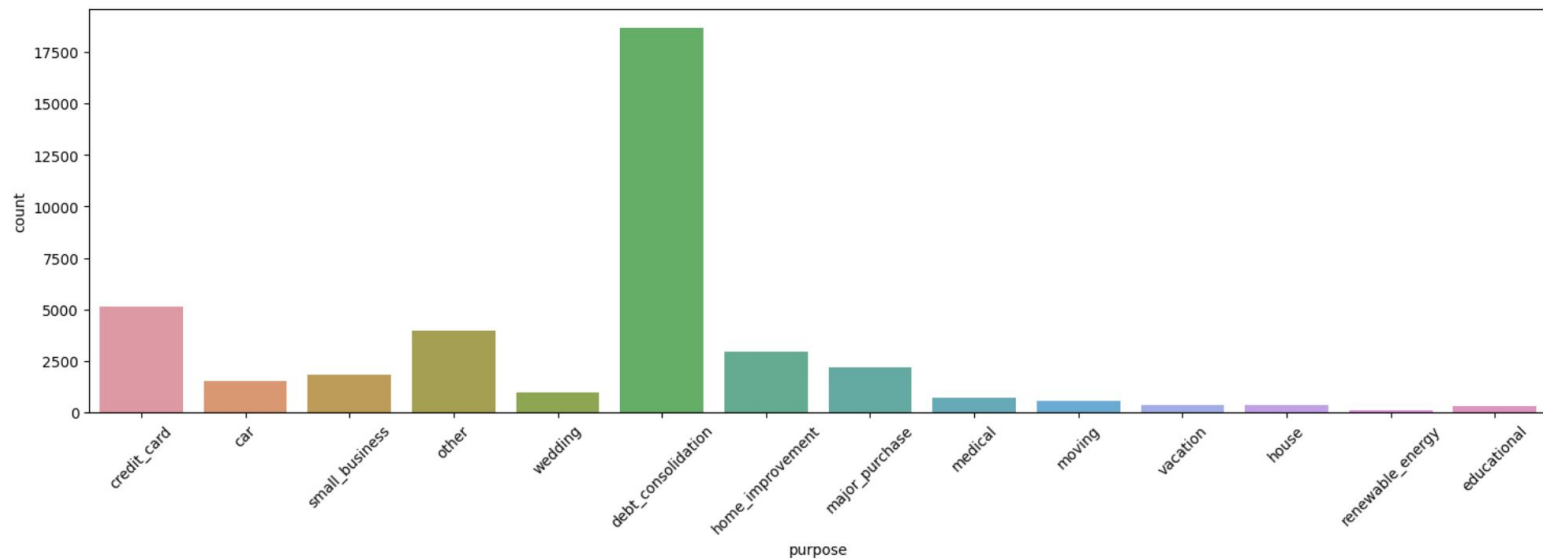
Most of the user apply for loan with the purpose of consolidating their previous debt.

Loan Status
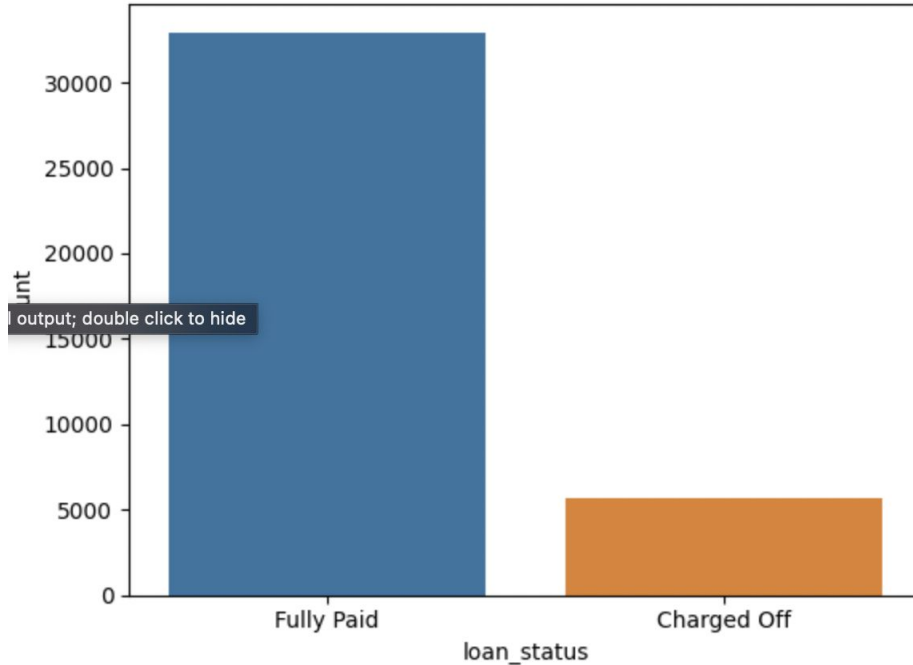

```
Fully Paid        0.829620
Charged Off       0.141677
Current           0.028703
```

Observation:

This graph of loan_status show the distribution of **Fully paid** and **Charged-Off** borrowers.

83% of borrowers pay their loan in time,
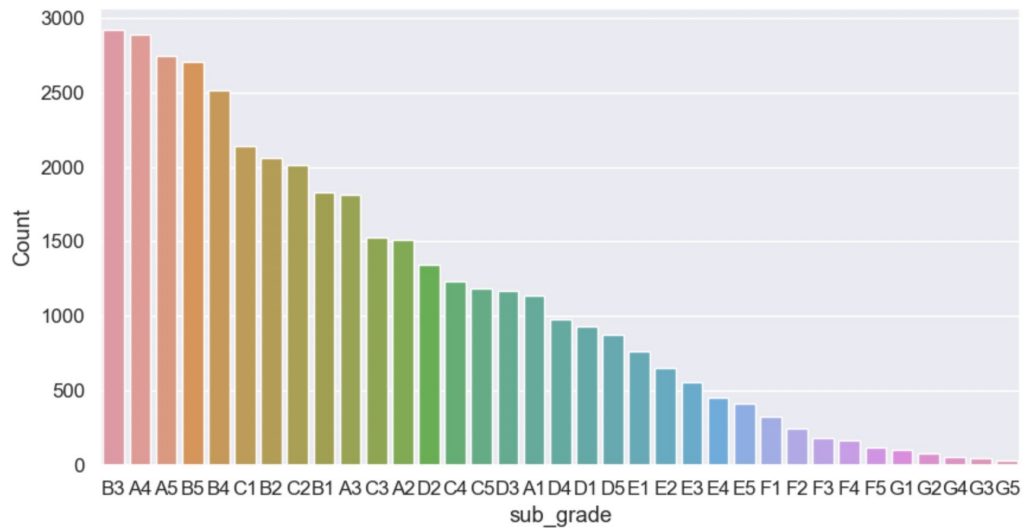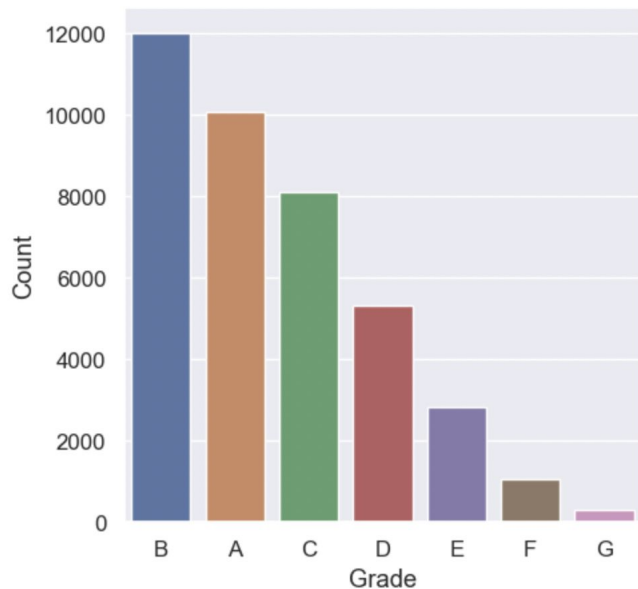
Data is **imbalance**
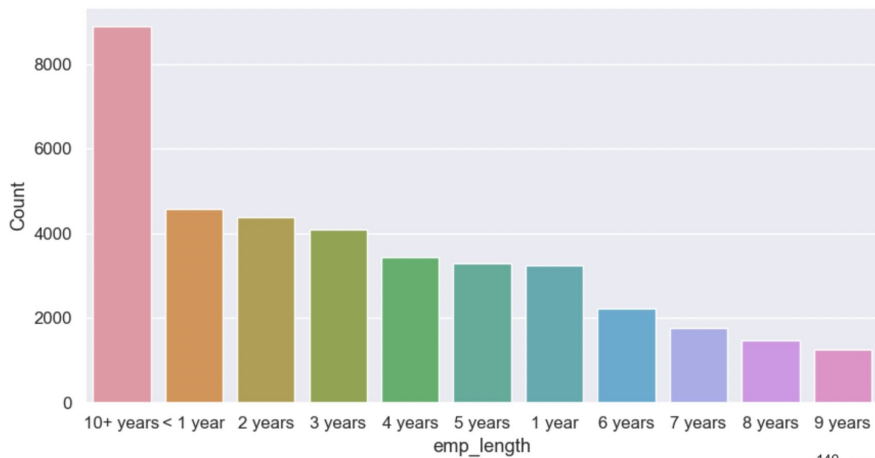
Current is ongoing loan, **not considered** during analysis

Borrower's state
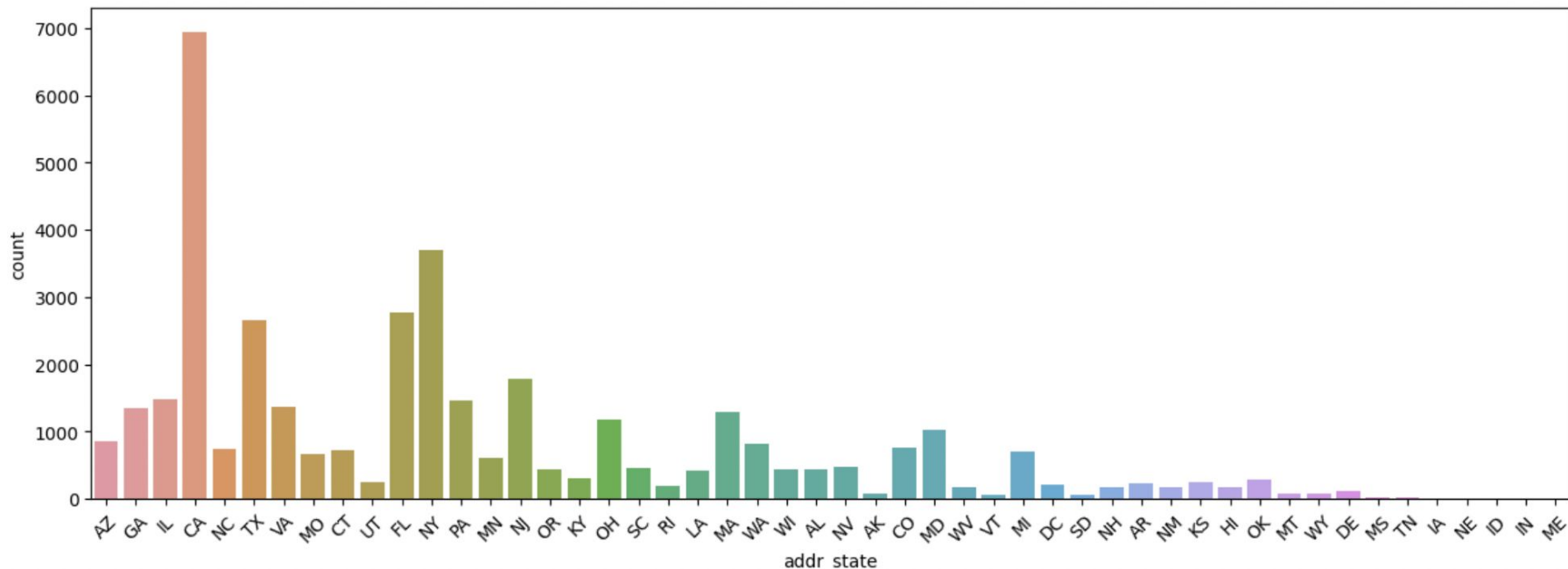
Observation: Most of the loan application for LC comes from **CA, NY, FL & TX**

Grade and sub_grade

Observation: Most of the loan application are from people of grade **A B & C.**
Not considering **subgrade**

Employee Length and Employee Title

Observation: Most loan application are either from someone starting their career or someone having **10+ years of experience.**

Observation: **US-ARMY & BOA** employee has applied for loan more than any other company.
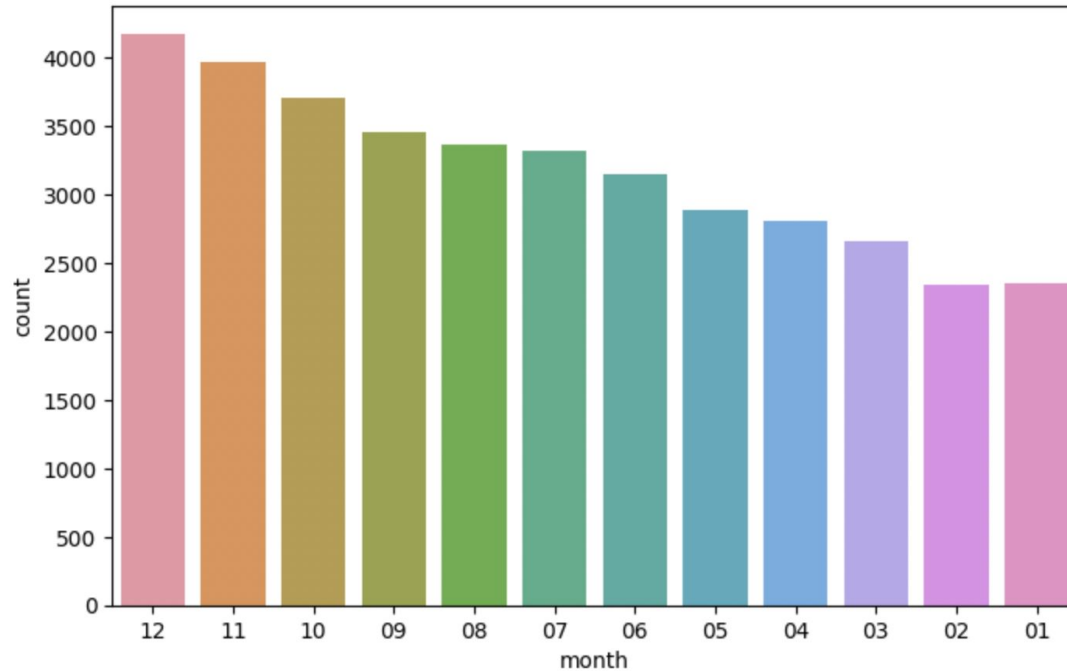
Term and Home ownership

Observation: Most of the member opt for short term loan option i.e **36 months.**
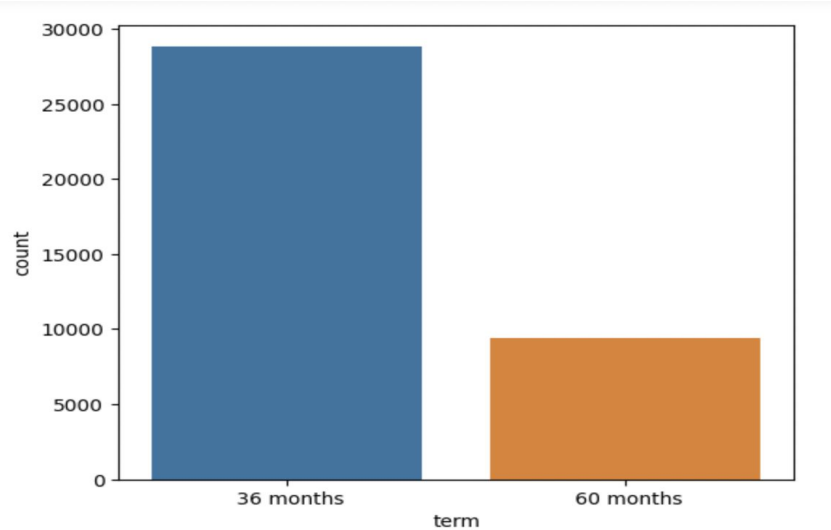
Observation: As we can see in the above plot people who **own house** are **less likely to take loan**
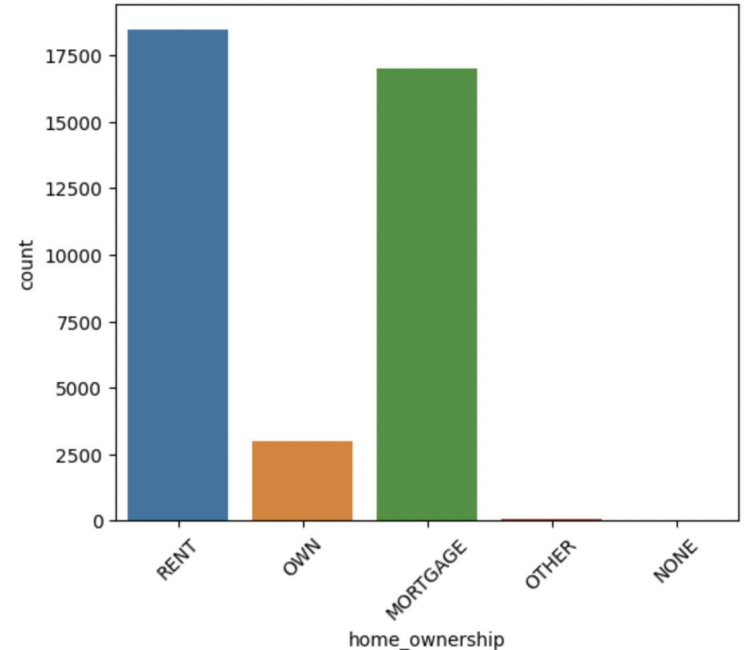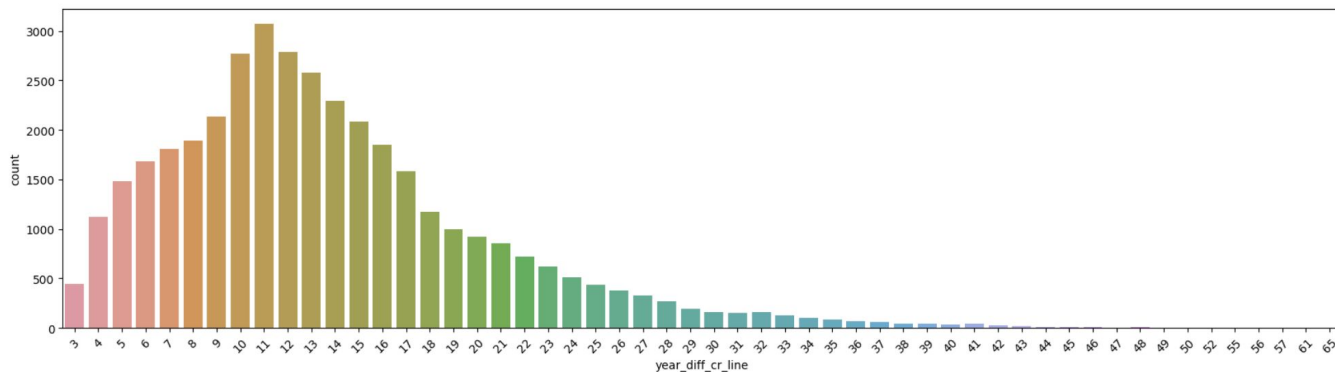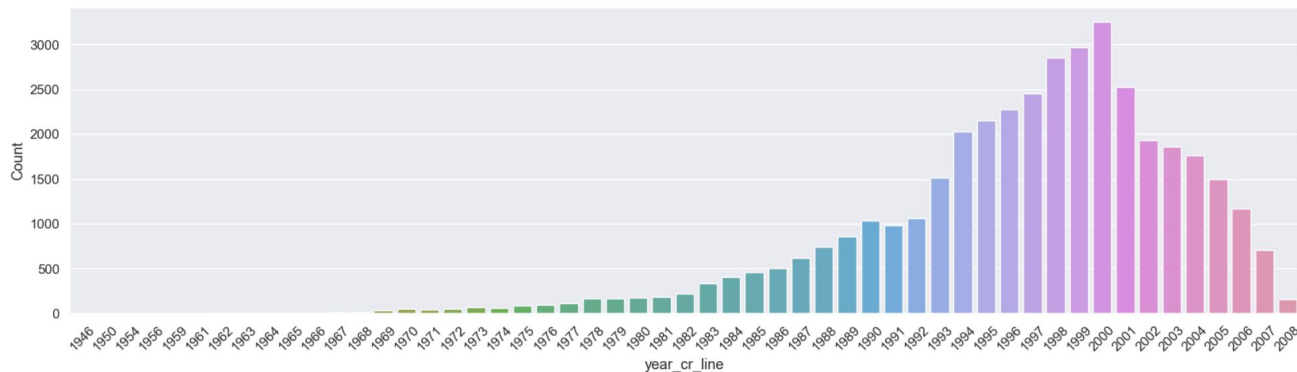
Loan issued date
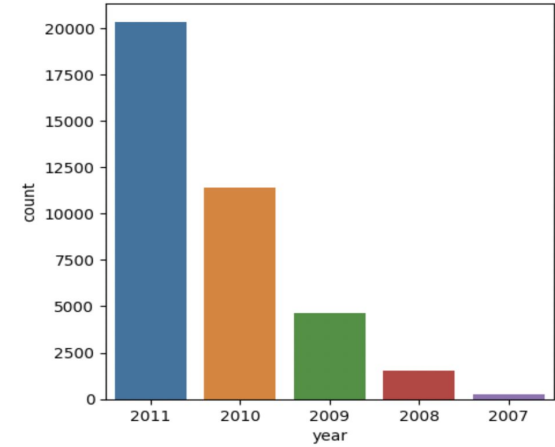
Observation:
From 2007 to 2011 there is **2x increase** in borrowers.
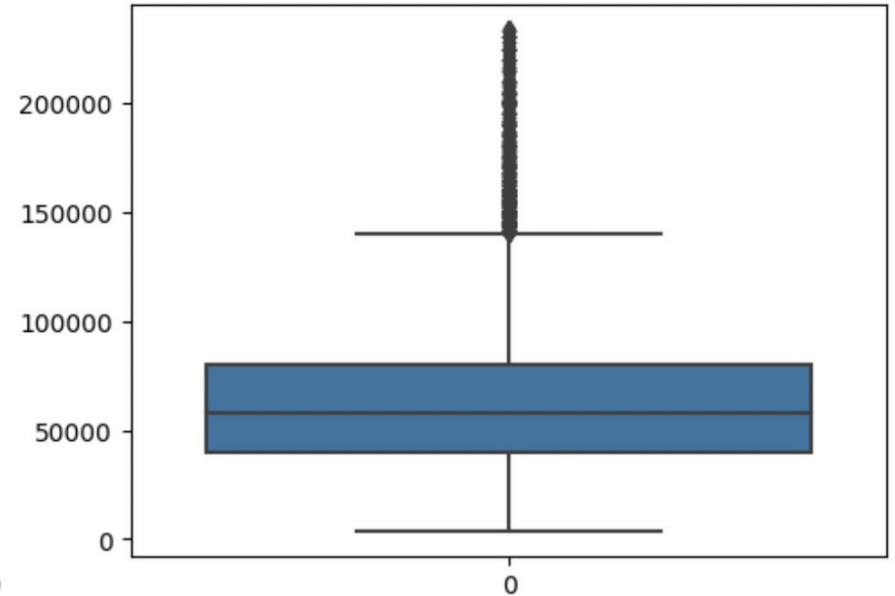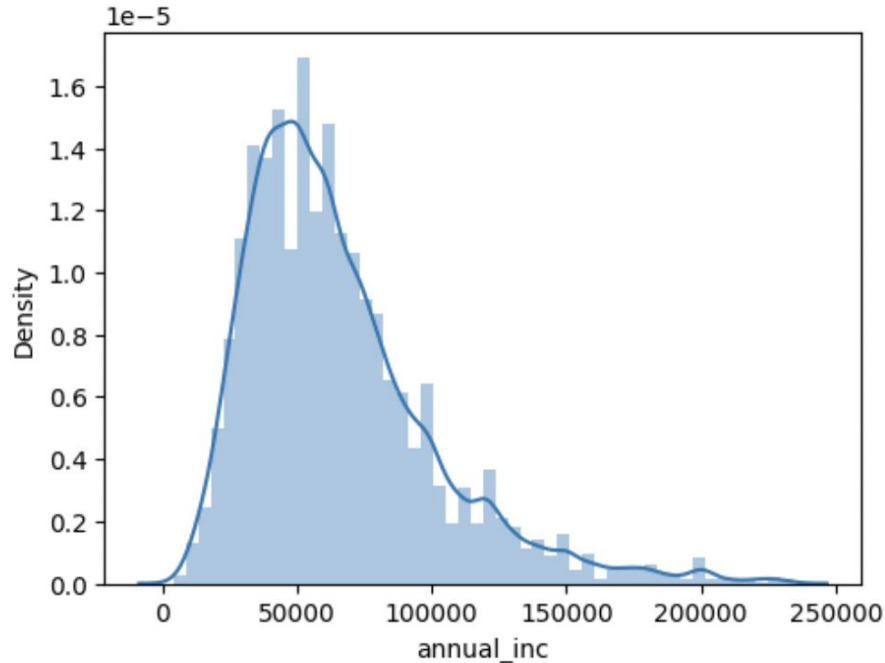Most people take loan towards **year end (Dec, Nov, Oct)**
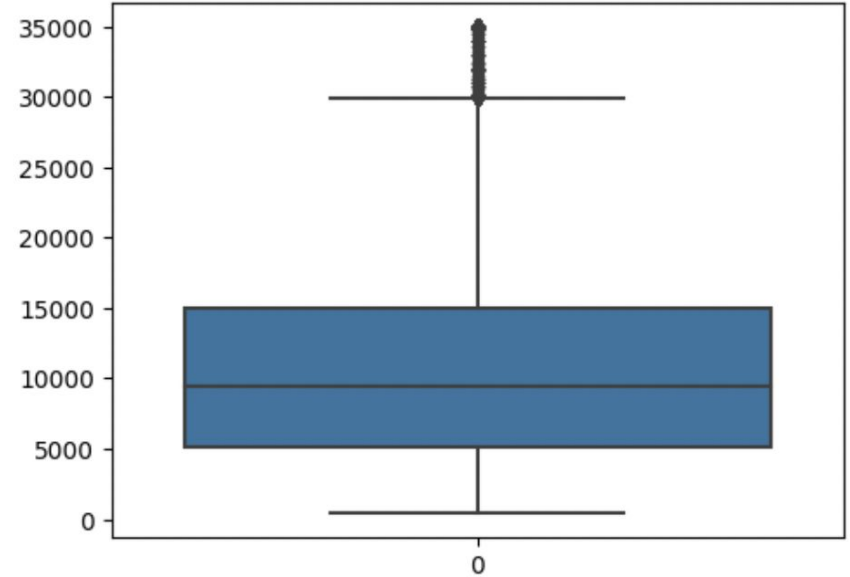
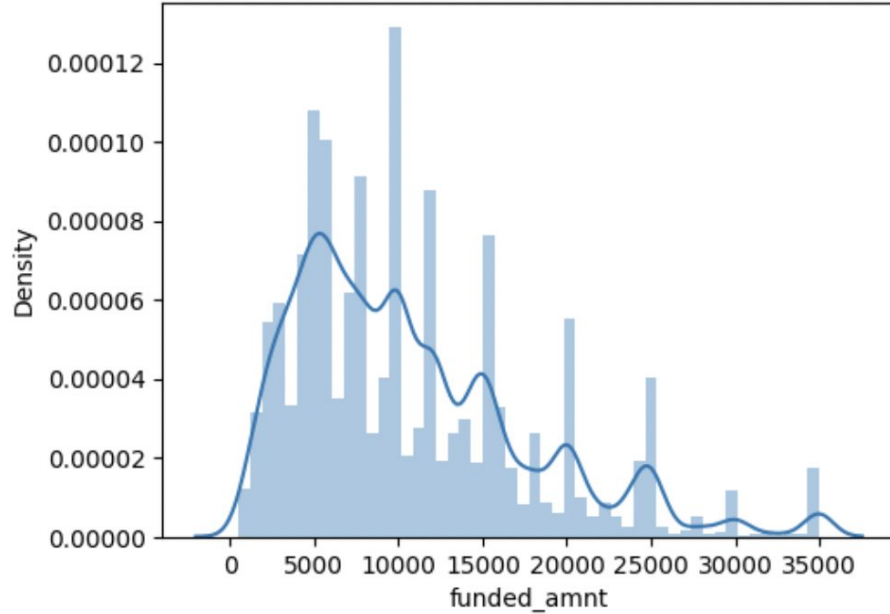# Earliest credit line (first loan date)

Observation:
1. Most of the customers have started taking loan from **late 80s.**
2. People take loan on avg after **13 years.**
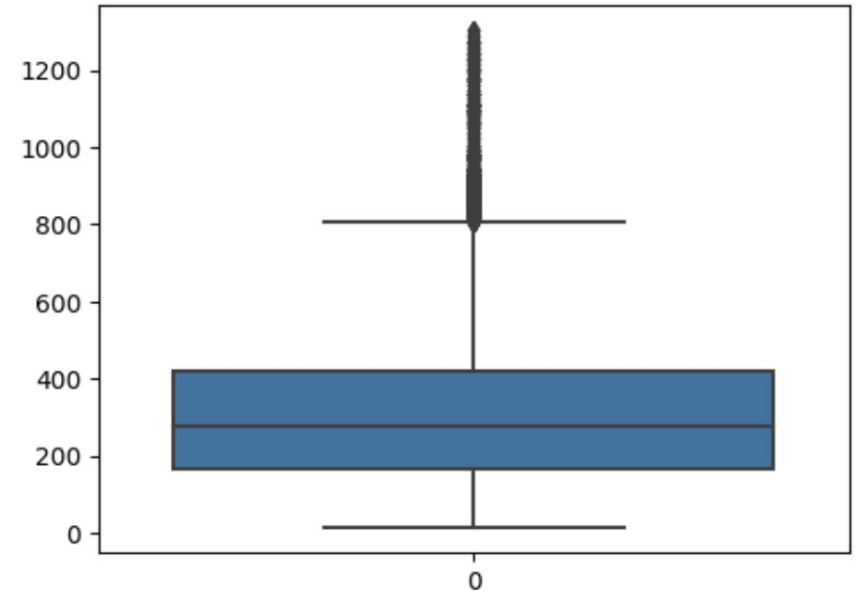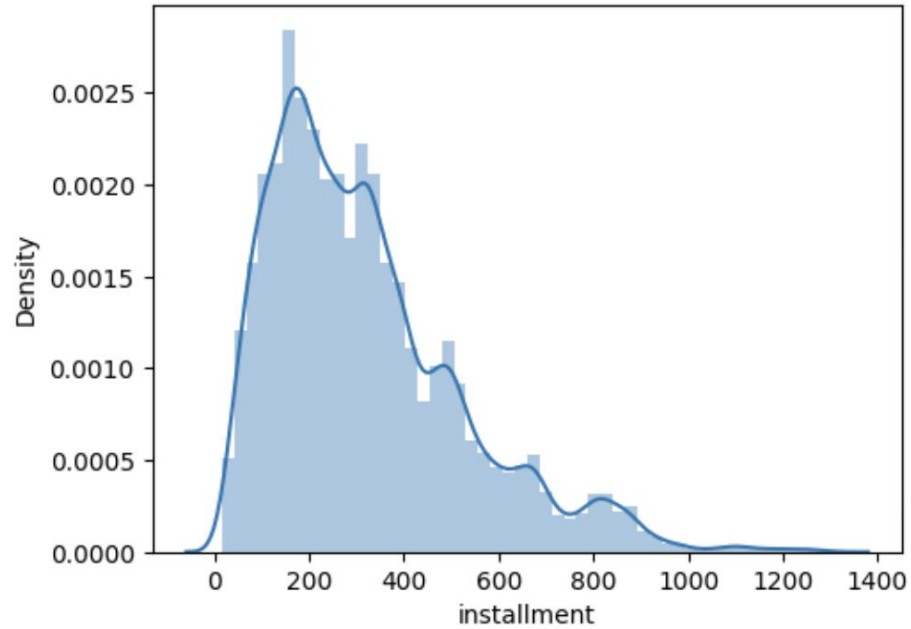
Annual income

Observation: Annual income is **right skewed**, which is true cause there are few people with higher salary range, Also there are f**ew outliers**. **99%** of people are having incomes **below 23k**.
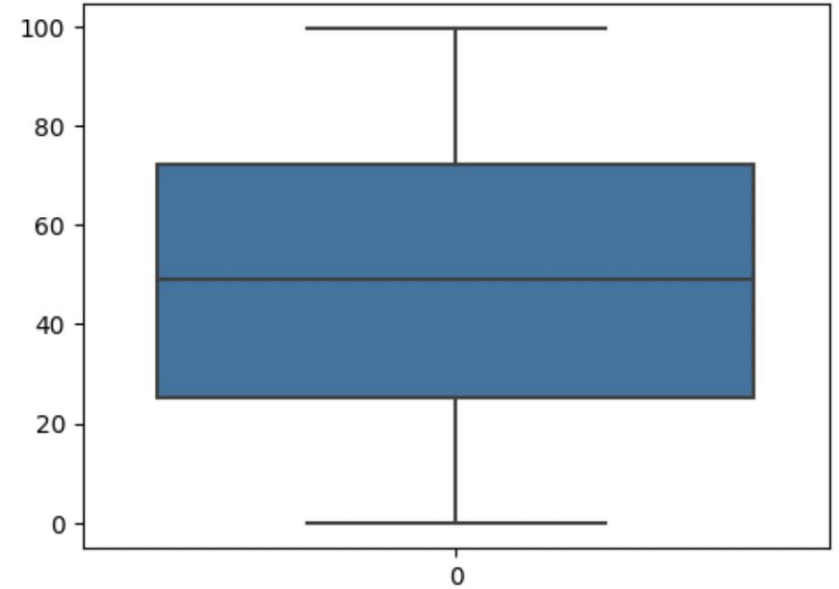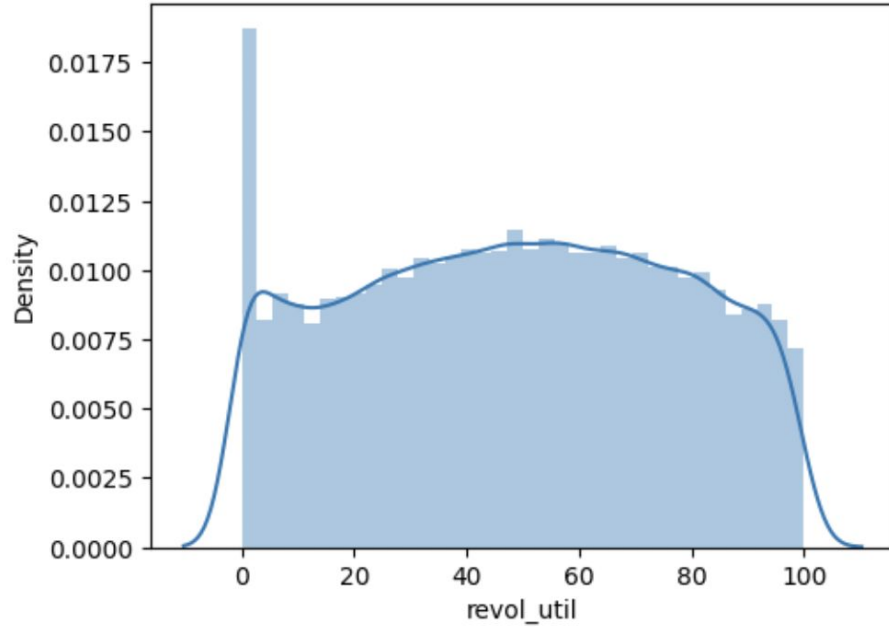
Funded Amount

Observation: Most common range of funded amount is between **5k-15k**
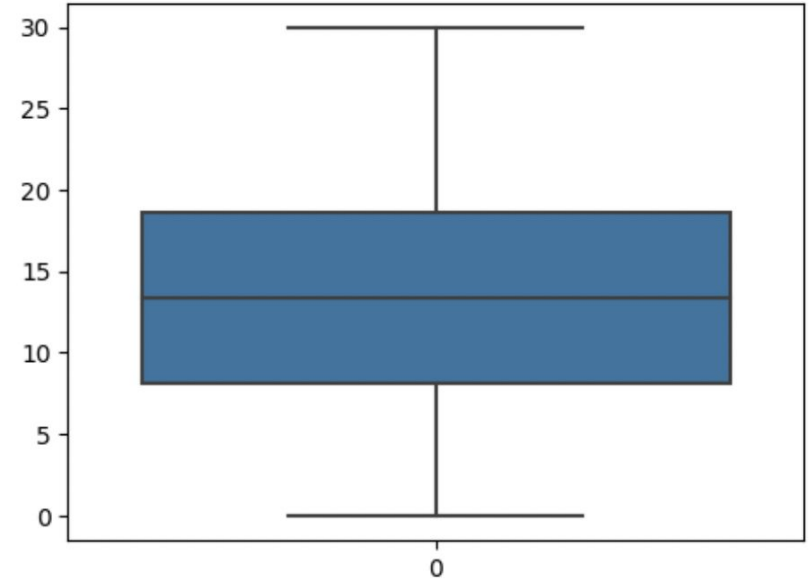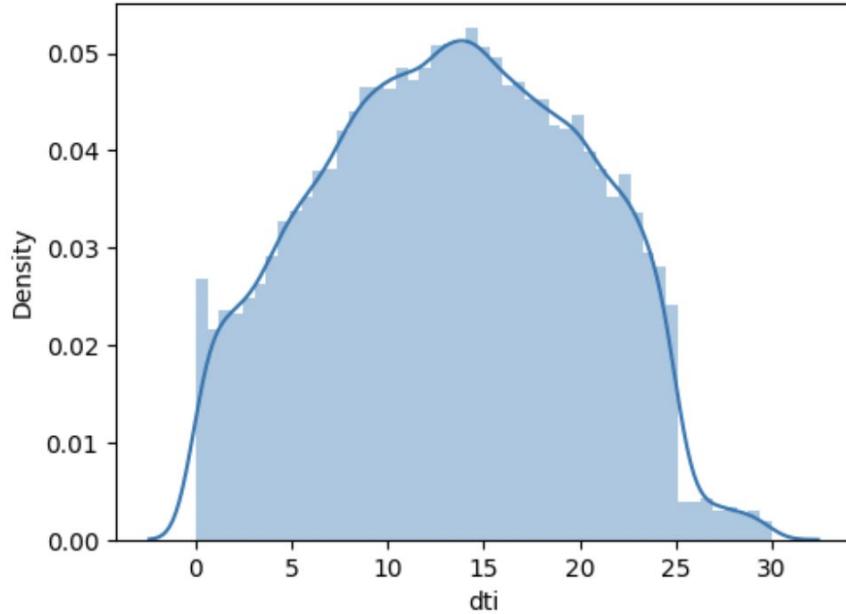
Installment

Observation: Most loan application has installment amount in the range of **200-400.**
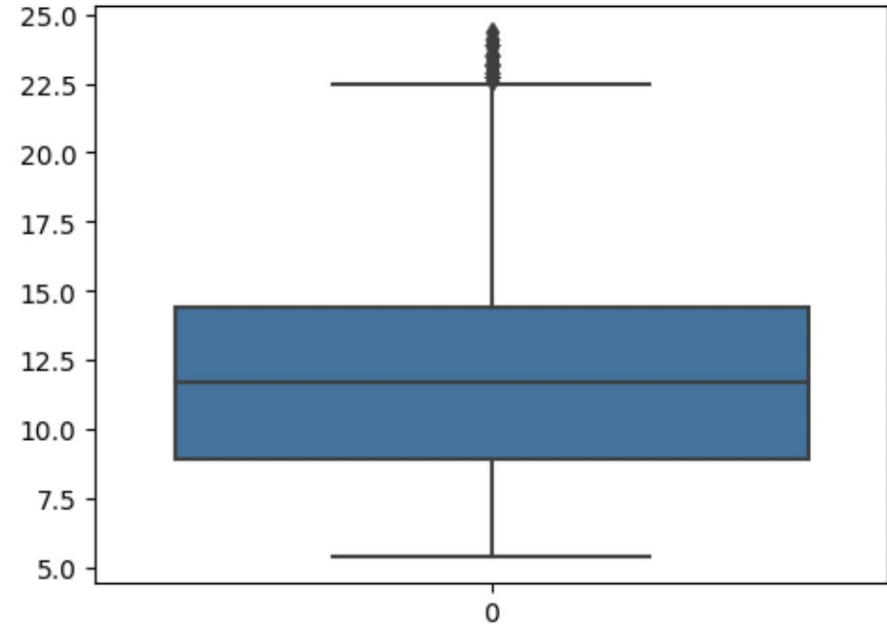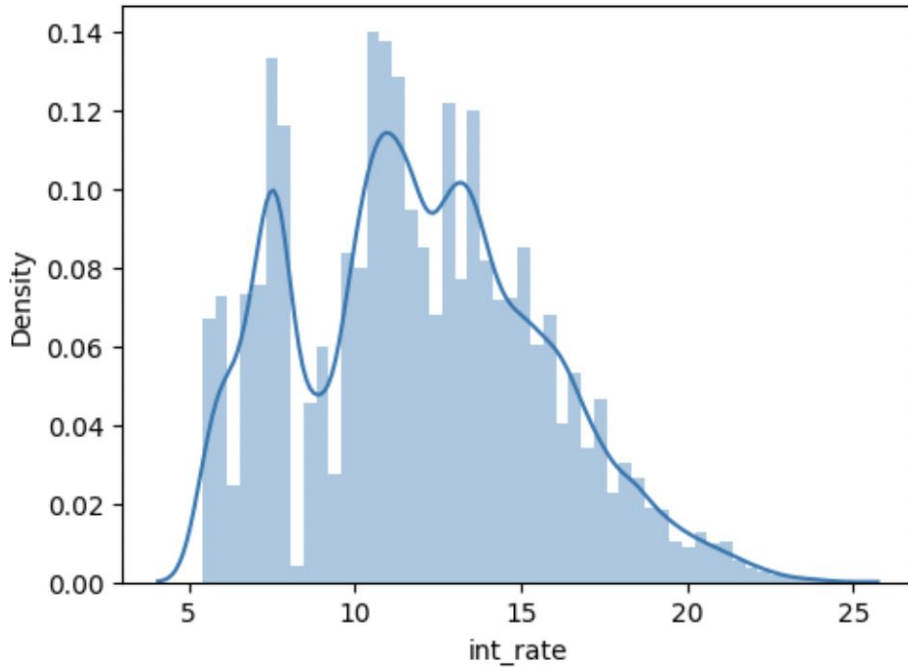
Revol util

Observation: Most of the member have **40-60%** of credit limit consumed.

Dti -> ratio of loan/monthly income

Observation: All the applicants have **dti < 30%** which is good thing. Addition to that there is **no outliers** in dti.
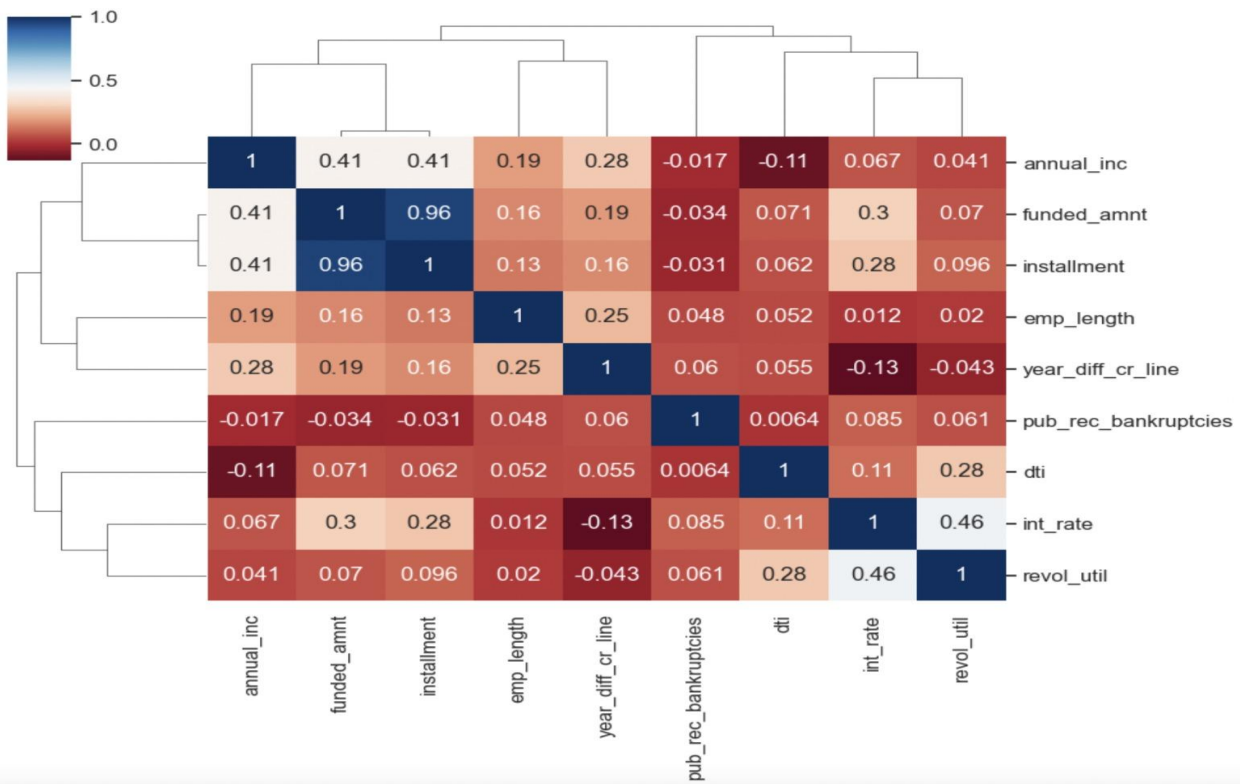
Interest Rate

Observation: Most applicant had to pay interest rate of **9-14%**.
Frequency distribution of interest rate is **right skewed**, few people took loan at higher rates of interest i.e., **17-22.5%**

Bivariate Data Analysis

Correlation between numerical features.
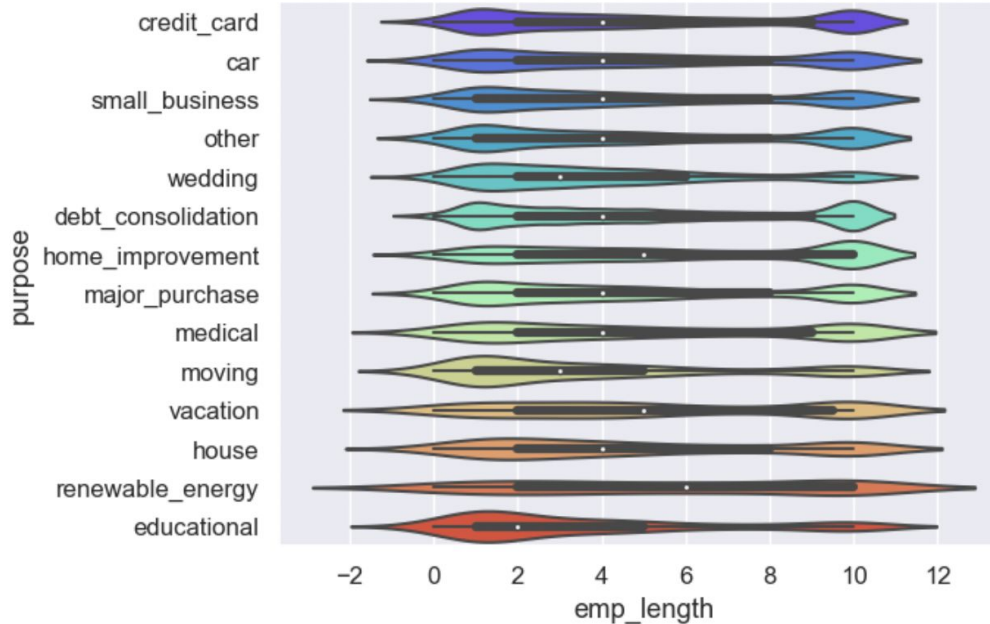

Observations:

Positive Correlation

1. Funded amount & Installment are highly correlated as higher the funded amount , installments to be paid will be higher too.
2. Funded amount is also moderately correlated to annual income & interest rate.

Negative Correlation

3. dti is - monthly debt paid/monthly income, hence when lesser the income high is the dti ratio. when income increases, dti will decrease.
4. interest rate is decreases for some who has been taking loan from long time, as int_rate seem to be negatively correlated to years passed from earliest credit line of a user.

Purpose vs Employee length

Observation: As observed in Univariate analysis - most loan application are either from beginner of 10+ year experienced. Here we can see that beginners generally take loan for educational, moving and wedding purpose, but experienced members generally take loan for medical & home improvement.
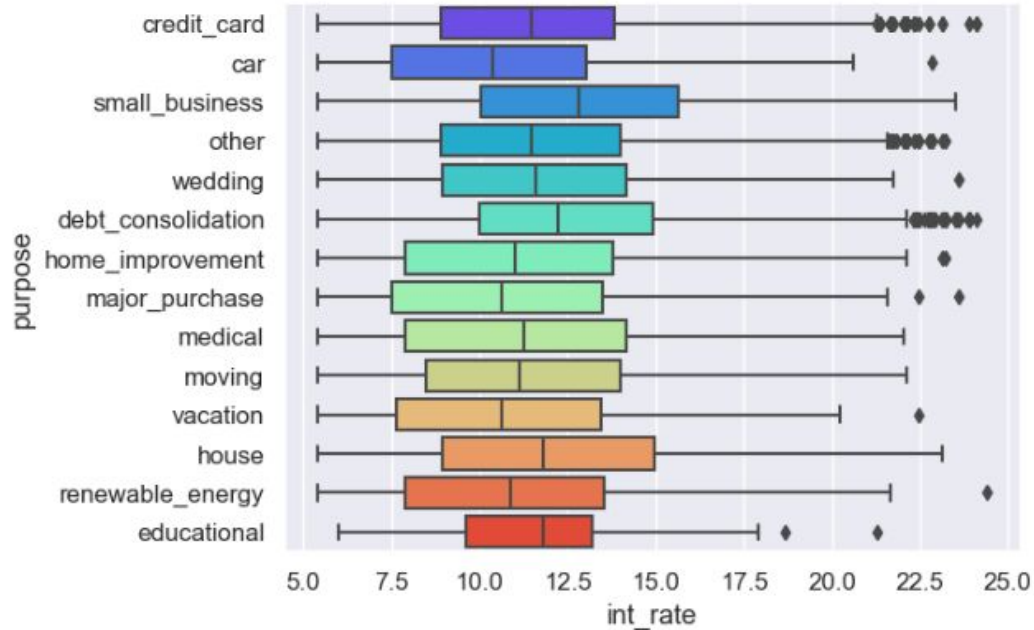
Purpose vs Funded Amount
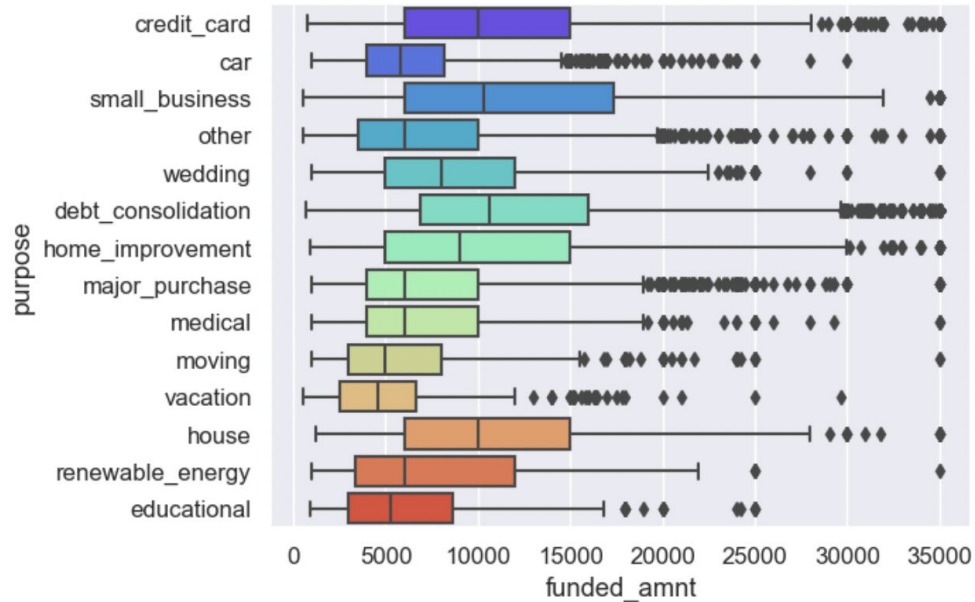
Observation:
1. High loan amount is applied for small businesses, house & debt consolidation.
2. Loan amount for car, education & vacation and comparatively smaller than other reasons.
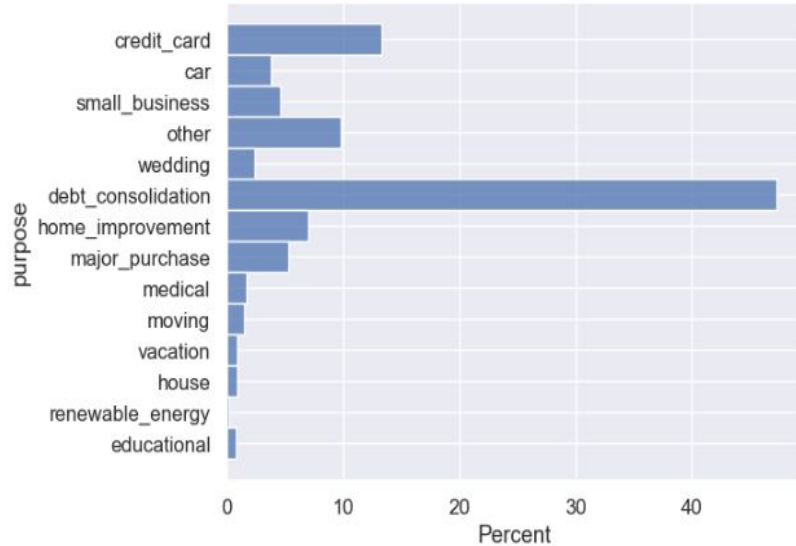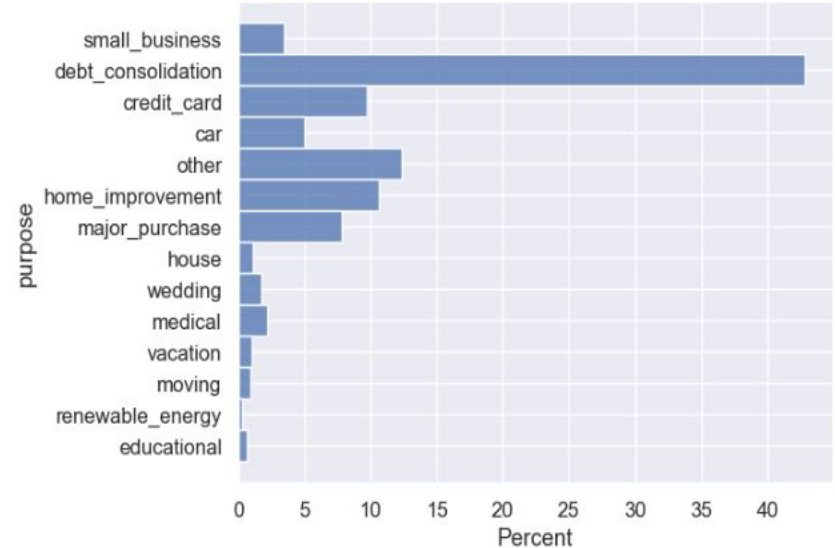
Purpose vs Interest Rate

Observation:
Interest rates for small businesses are generally higher as risk involved in loan repayment is higher.
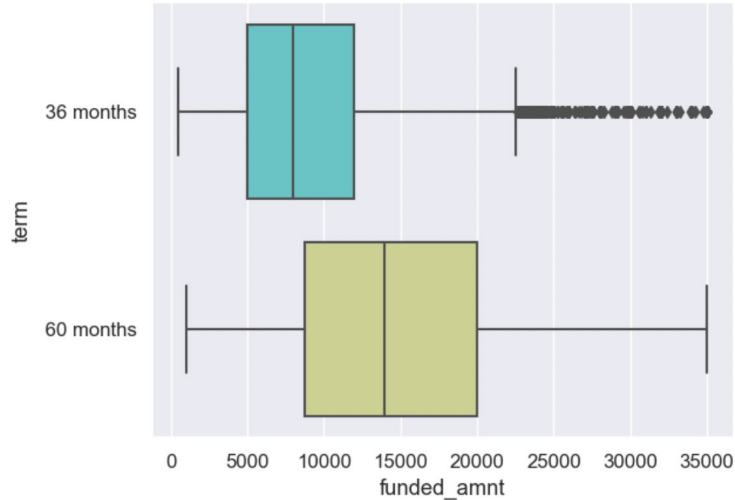
Purpose vs Home Ownership

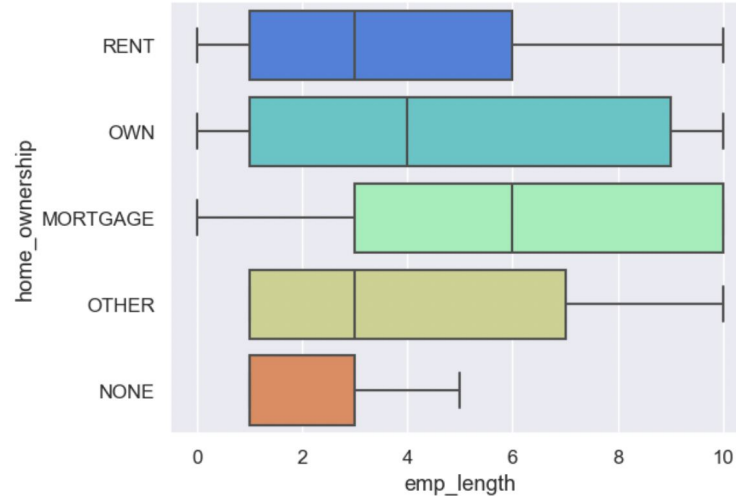Owns home

Does not owns home

Observation: Although most of the **purpose** for both category of **homeowners** seems same but we can see More members who has their own house apply for loan for '**major purchase**'.

Term vs Funded Amount

Home ownership vs emp_length

Observation: Funded amount is higher for long term loan. generally if a person applies for big loan they opt for long term payment options.

Observation: Most of the applicant lives in rent till 6 years of experience. Applicants who owns their own house, we can see its graph is left skewed, which means their is high chance someone owns house only after significant year of experience.
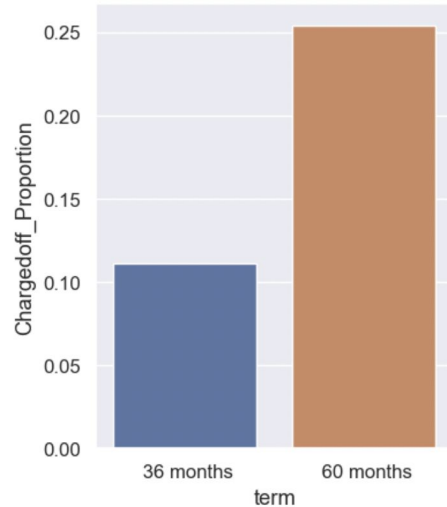
grade vs int_rate
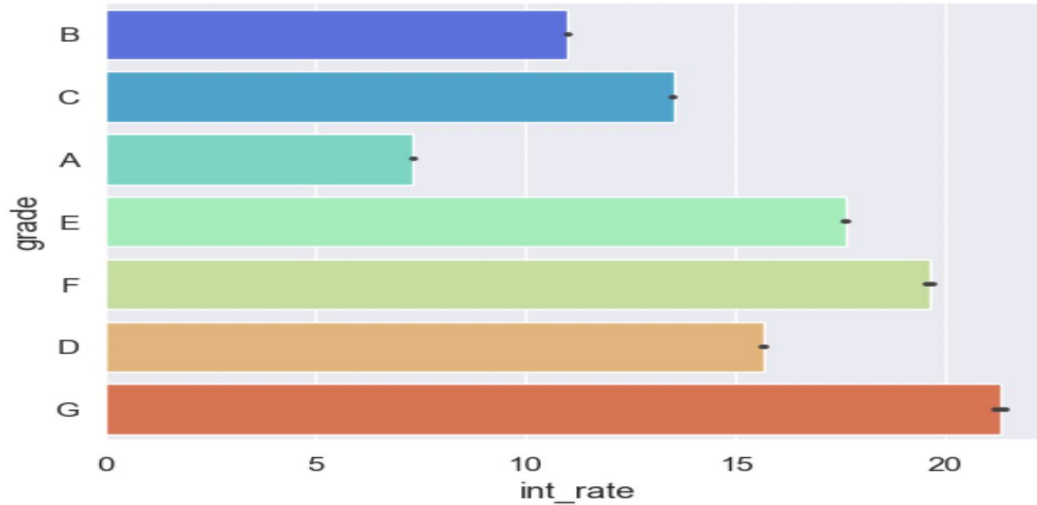
Observation: Interest Rates are higher for grade **D,E,F,G** as compared to grade **A,B,C.** We also observed more loan applicants are from **A, B & C** Grades.
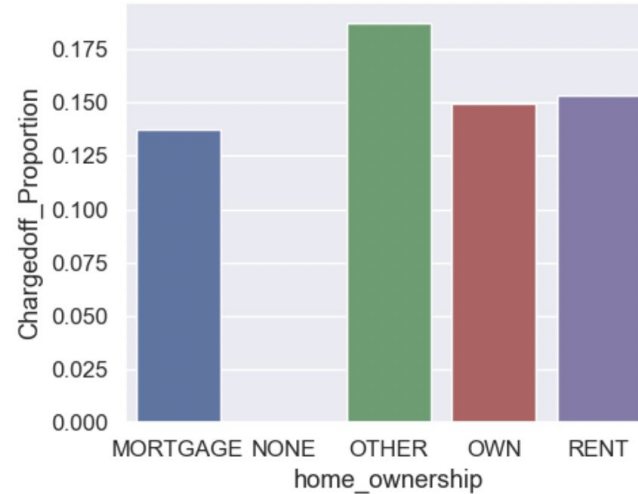
All features vs Label
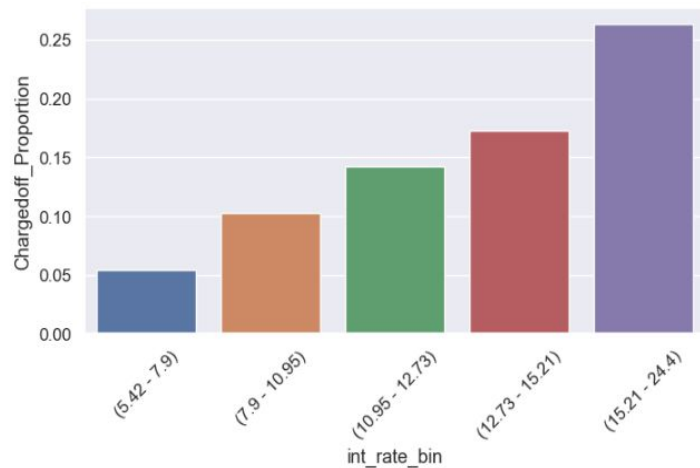
Charged off vs term

Observation: Long Term loan has higher chances of getting charged off. We have also observed that funded amount is significantly higher for long term loans as well.
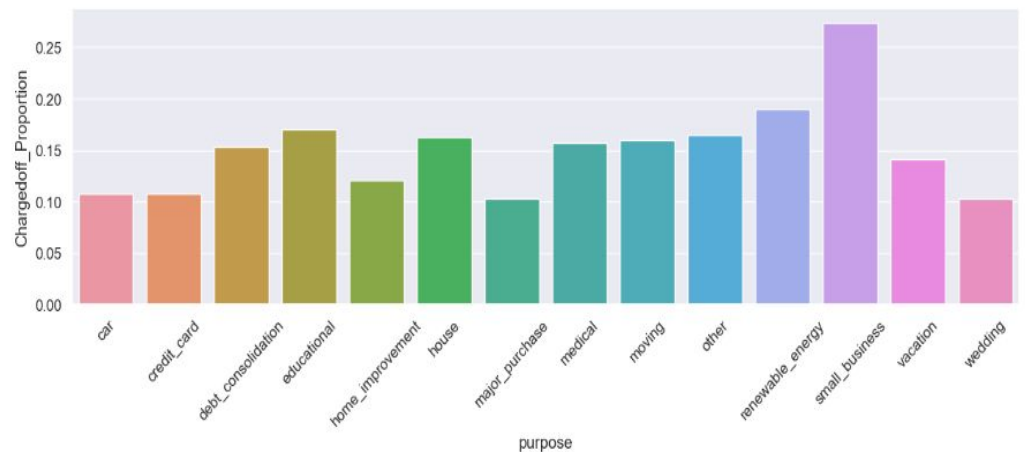
Charged off vs Home ownership

Observation: Observation - People who don't share their home_ownership status are more likely to charged off
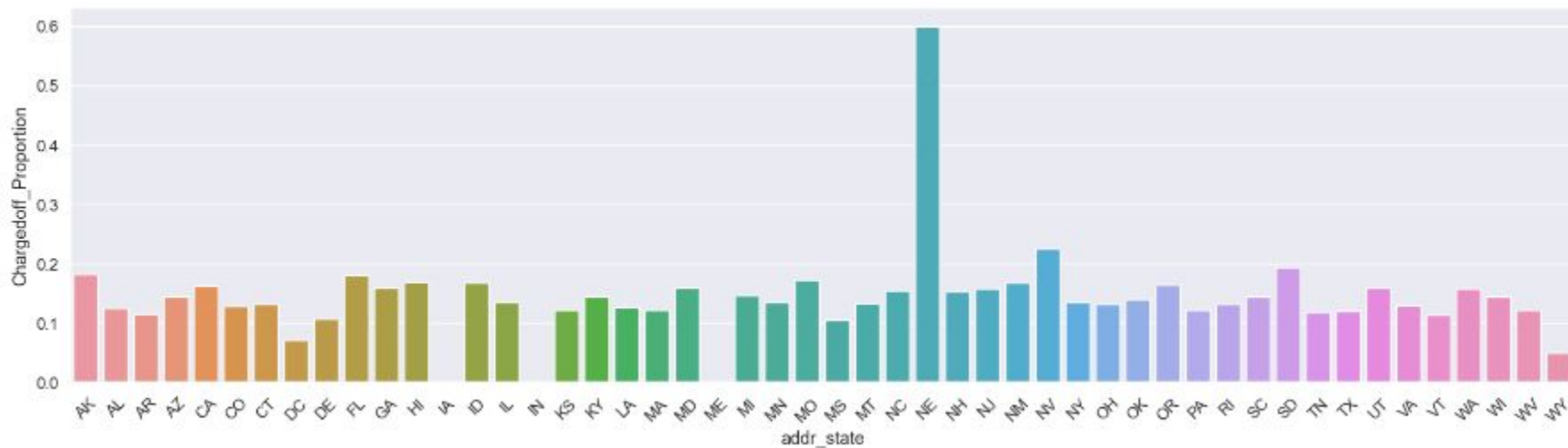
Charged off vs interest_rate

Charged off vs purpose

Observation: Charged Off rate shows positive 2x relation with interest rate.

Observation: Small business being the riskiest purpose, has **highest rate(>25%)** of Charged Off.
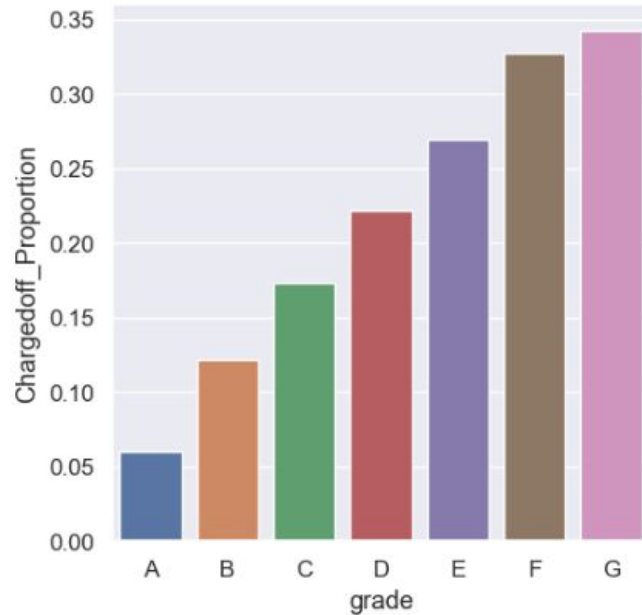
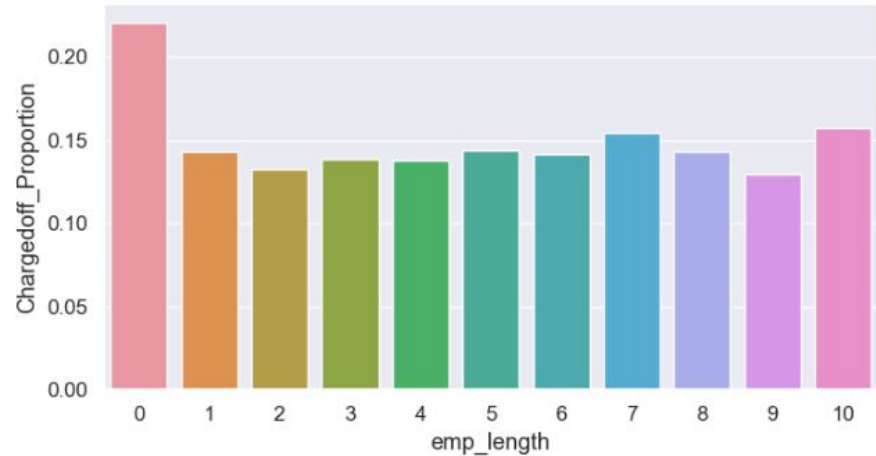# Charged Off Rate vs State

Observation:
Applicants from NE has **6%** Charged Off rate which is higher than applicants from any other state, and has low Charged Off rate below **2%**

Charged off vs grade

Observation: Higher chance of Charged Off from applicant of Grade **E, F & G**, as we noticed earlier, their interest rate are also higher when compared to other Grades.

Charged off vs emp_length

Observation: Applicant just starting their career has higher rate of Charged Off cases. i.e **0-1 yrs**
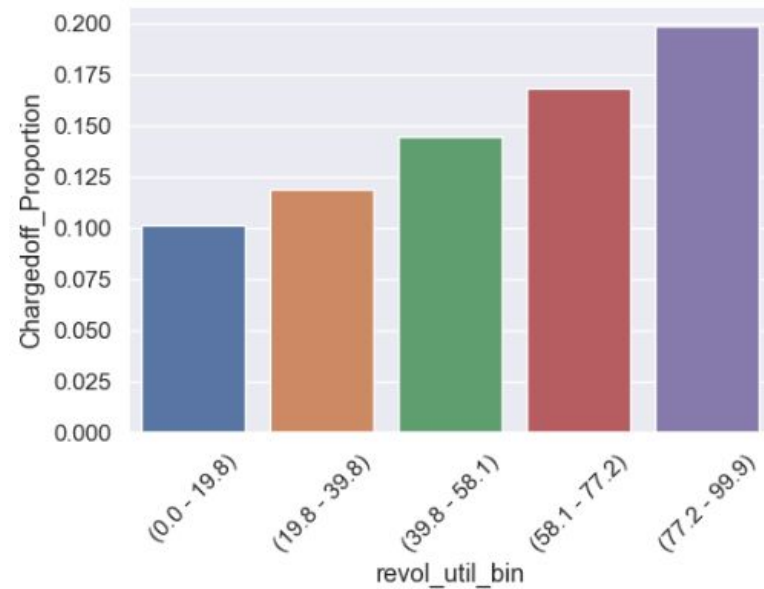
Annual income vs Charged Off Rate

revo_util vs Charged Off Rate
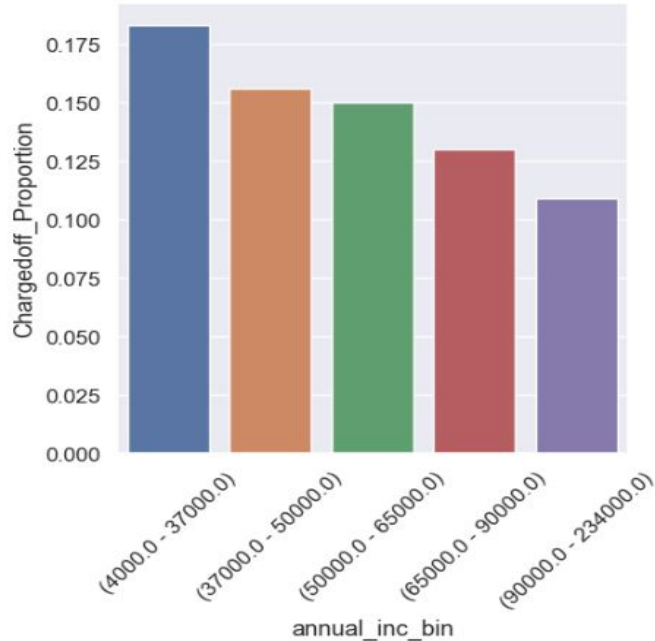
Observation:
Charged Off rate is higher for applicant having low annual income and shows linear decline with increase in income.

Observation:
Applicant who has consumed 70-100% of their credit limit are more likely to charge off as credit limits are set according to applicant's potential to loan back.

Deriving Factors

1. **Lower grade(A, B, C)** has **low chance** of charged off
2. **Low interes**t rate has **low** change of charged off
3. Employee with **less experience** are **more likely** to be charged off.
4. Basis on **state (Nebraska)** has **high** chance of charged off

THANK YOU