Problem Statement:

Analyse the customer lending data set provided by the bank and identify the risks of loan defaults at all categories.

the company wants to understand the driving factors (or driver variables) **behind loan default, i.e. the variables which are strong indicators of default.**

Approach taken to identify the risks

1. Loaded the data into the python pandas object
2. Removed the columns that are null, same values across all columns, irrelevant or un-useful columns for the analysis
3. Removed the data of rows that are duplicates, nulls and loan status not equal to current
4. Cleansed the data to more readable format by removing the special characters
5. Replaced the null values with mode
6. Identified the outliers and removed them from our calculations
7. Now, plotted the data with the counts of loan status = 'Charged off' against the below fields to analyse the data closely and to draw a pattern
   a. Grade
   b. Sub grade
   c. Home ownership
   d. Purpose
   e. Utilization groups
   f. Segmented Interest rates
   g. Annual income groups
   h. Total credit lines (Total_acc_groups)
   i. Employment length
   j. Verification status

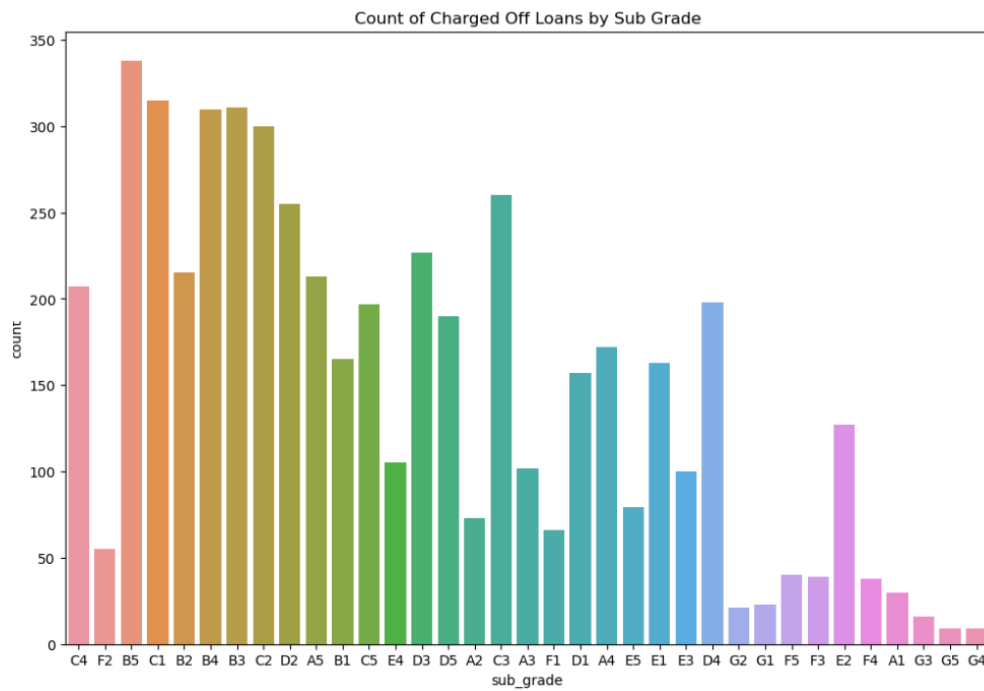Lastly, analysed all the plots and concluded the risk elements while granting a loan to the customer.
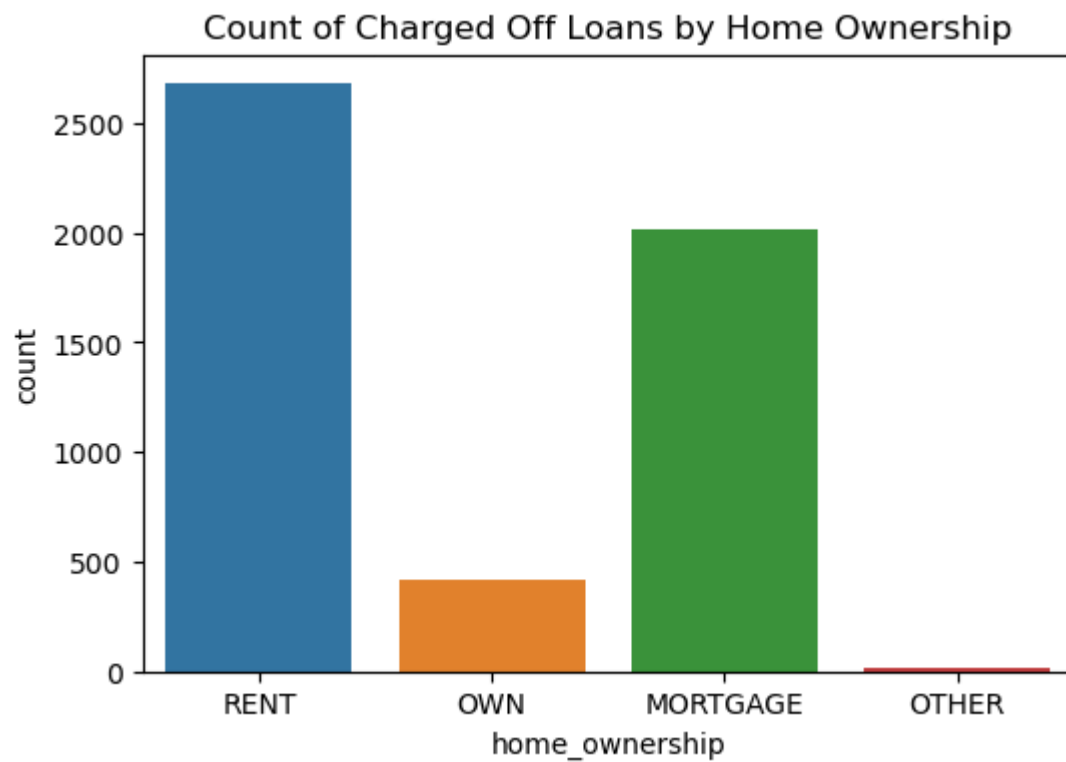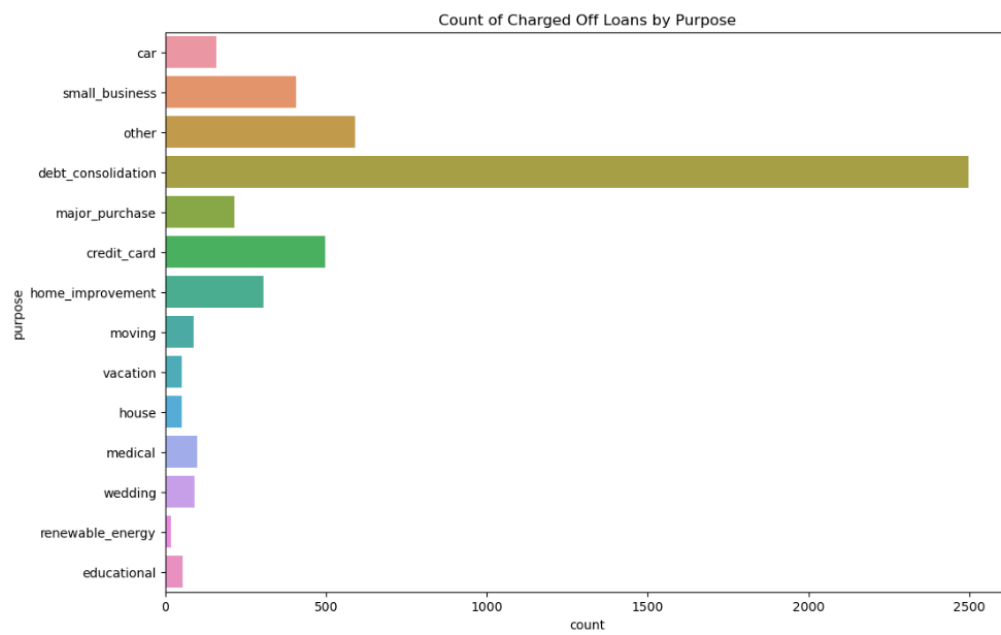
Plots identified

1. **Grade Vs Charge off**

<Axes: xlabel='grade', ylabel='count'>
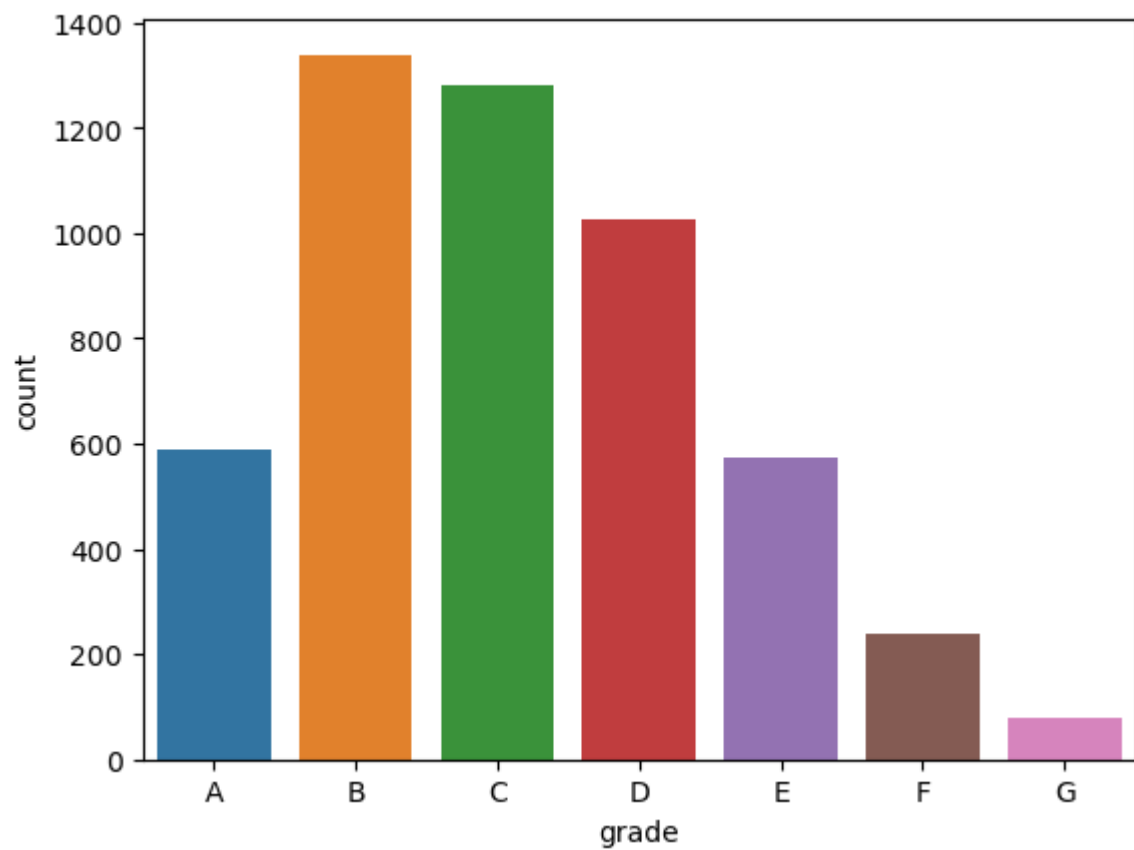
2. Sub-Grade Charge off stats

Count of Charged Off Loans by Sub Grade

3. Home Ownership Vs Charged off stats

Count of Charged Off Loans by Home Ownership
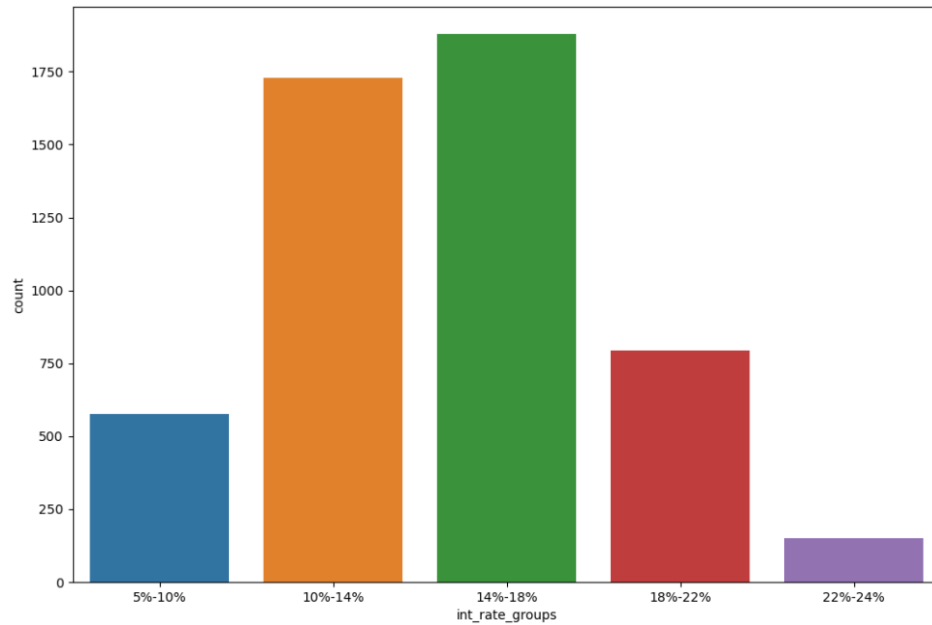
4. Charged off by Purpose of the loan
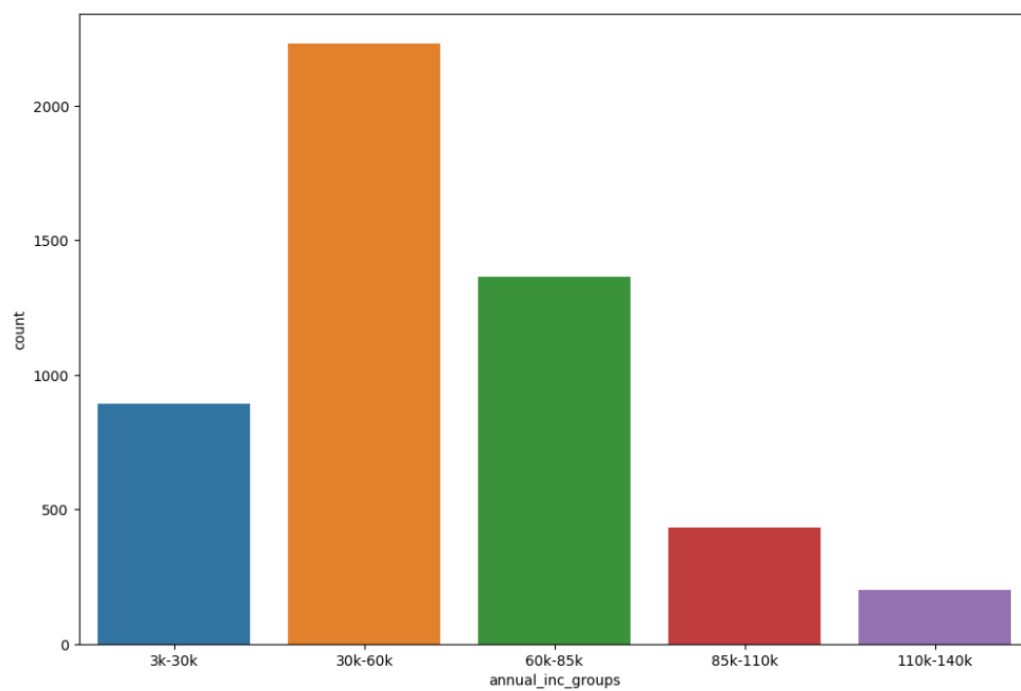

Count of Charged Off Loans by Purpose

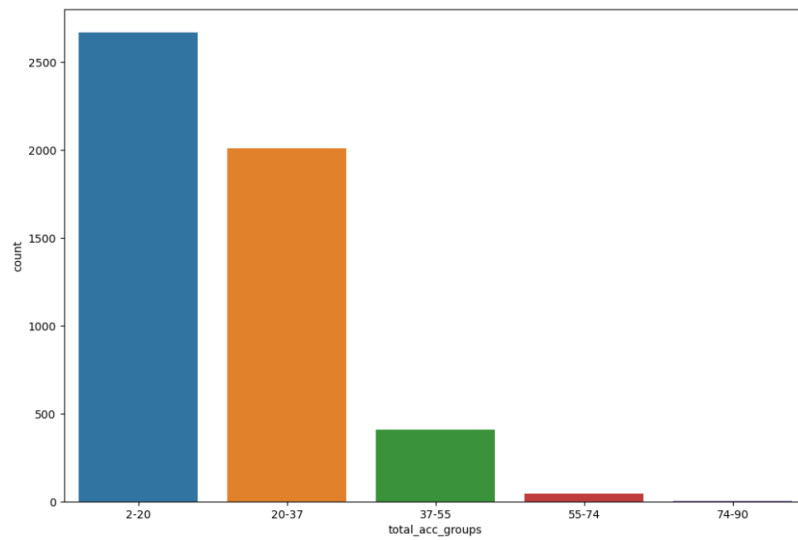## 5. Utilization groups vs charge off stats
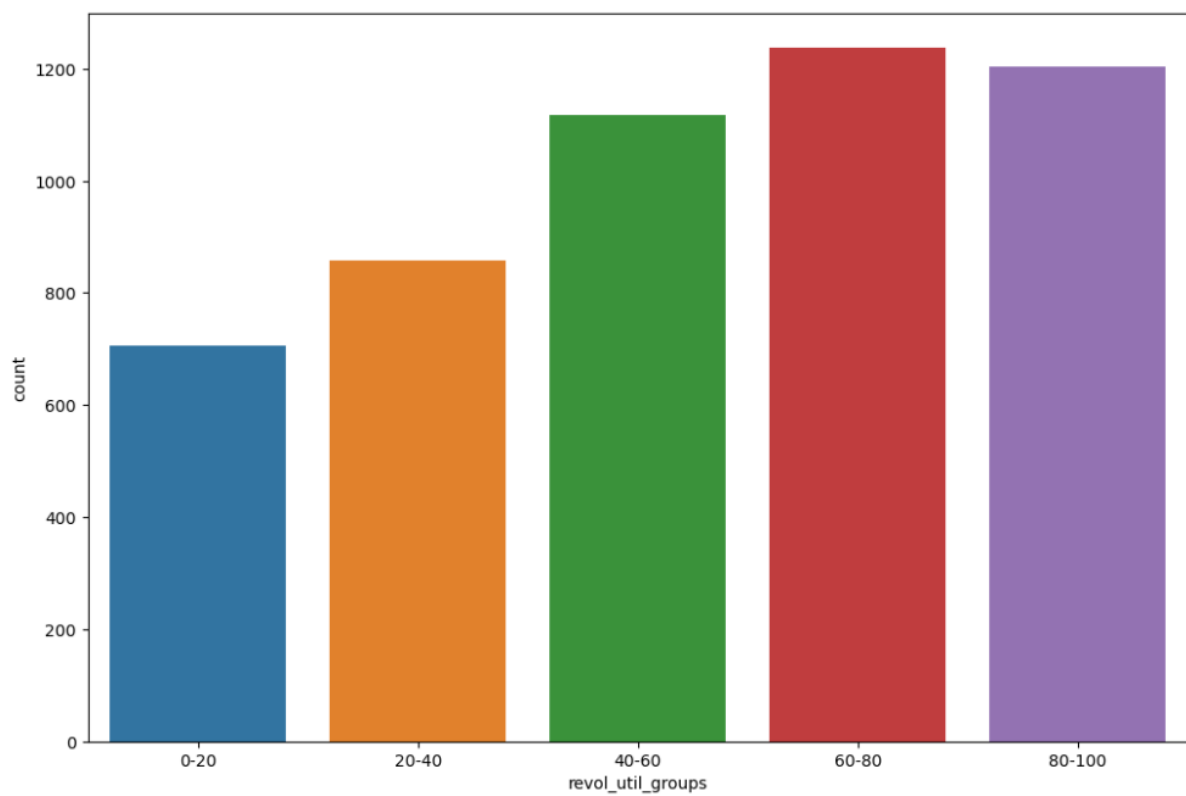
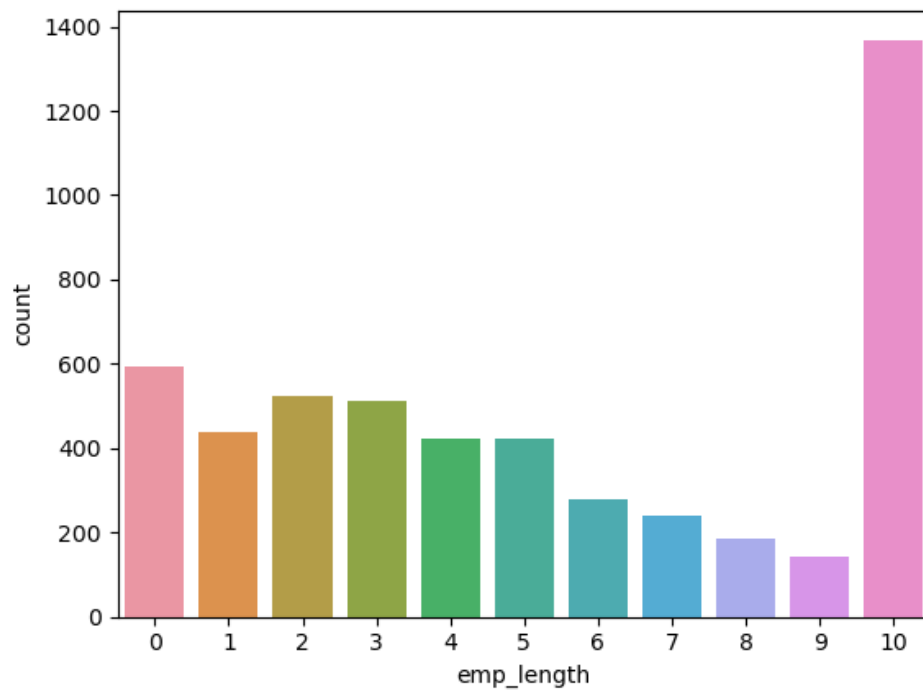## 6. Interest rate groups vs charged off plot

7. Annual income groups vs Charged off Plot
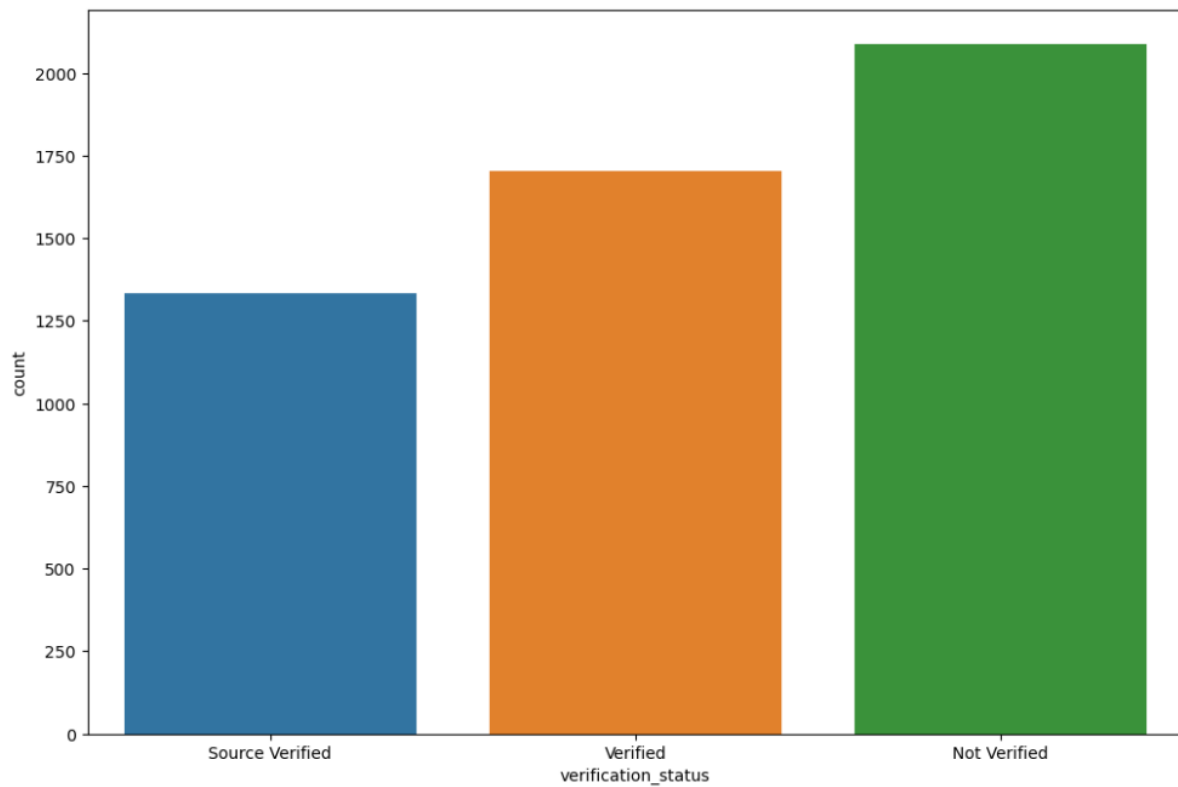
8. Total Credit lines of customer Vs charged off plot

9. Employment length Vs charged off Plot

Verification Vs charged off Plot

After thoroughly analysing all the plots above, below are points to be noted:

* Customers with verification status Not Verified tend to default more

* Customers who took the loan with purpose as Debt consolidation tend to default more

* Customers whose grade is B tend to default more

* Customers whose sub grade is B5, C1, B4, B3, C2, D2 tend to default more

* Customers whose employment length is more than 10 years tend to default more

* Customers whose current credit line is in between 2-20 tend to default more

* Customers whose interest rate is in between 14%-18% tend to default more

* Customers whose annual income is in between 30k-60k tend to default more

* Customers whose home ownership is RENT tend to default more