EDA Case Study

By Abhishek and Robin

Problem Statement

Solving this case study will give you an idea about how real business problems are solved using EDA. In this case study, apart from applying the techniques you have learnt in EDA, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.

Data Set of Problem

There are many Columns in this dataset, and we divided them in three categories mentioned below:-

- Customer Demographic :-

id,member_id,emp_title,emp_length,home_ownership,annual_inc,zip_code,addr_state,

- Loan Variable :-

loan_amnt,funded_amnt,funded_amnt_inv,term,int_rate,installment,grade,sub_grade,verification_status,issue_d,loan_status,pymnt_plan,initial_list_status,total_acc,revol_bal,revol_util,total_pymnt,total_pymnt_inv,total_rec_prncp,total_rec_int,total_rec_late_fee,recoveries,collection_recovery_fee,last_pymnt_d,last_pymnt_amnt,policy_code

- Customer Behaviour :-
  url,desc,purpose,title,delinq_2yrs,earliest_cr_line,inq_last_6mths,mths_since_last_delinq,mths_since_last_record,pub_rec,mths_since_last_major_derog,

Solutions

First, we have to read data from the CSV file after that we have cleaned and filled the data by removing unwanted rows and columns, We also remove outliers from this data set

Removing the Outliers

Here we describe the 'loan_amnt' and we got to know that there are outliers present in our data and we remove those outliers after that we got to know that most of the loan_amnt is from Rs 5000 to 14000. We can also perform this operation on 'installment' and 'annual_inc' column and remove outliers from them also. With these steps we clean our data further.

This is the box plot before removal of outlier
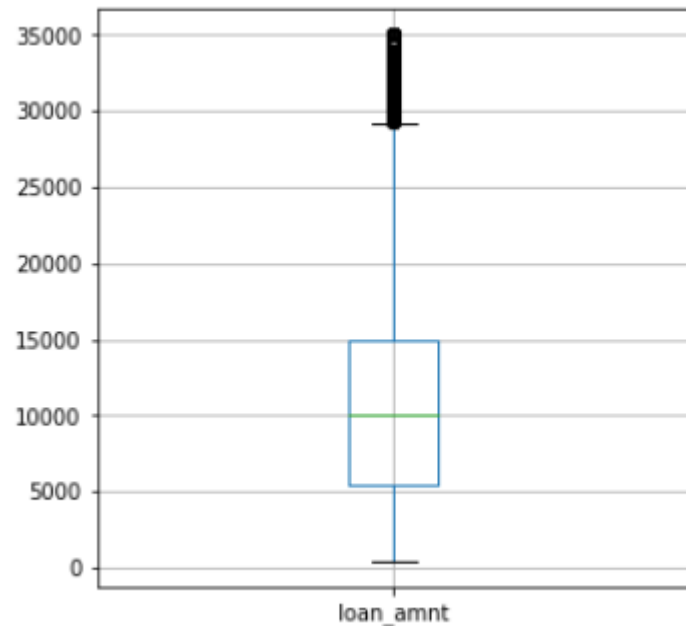
This is after the removal of outlier
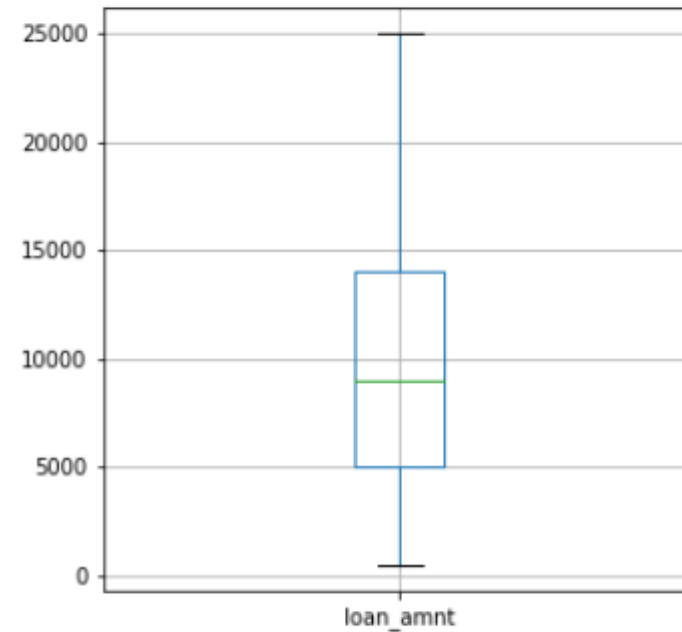
Loan_Status

Fig. 1

• Here we can see that total 14.2% people who defaulted their loan and 83.3% people who cleared all their dues and 2.5% people are currently paying their loan

Fig. 2

• In this picture we have only taken those who have Fully Paid or Charged off their loan because we can not decide on the basis of Current that who can Charged off their loan in future so we can continue our insights with only two scenarios of loan_status that is Fully paid and Charged Off
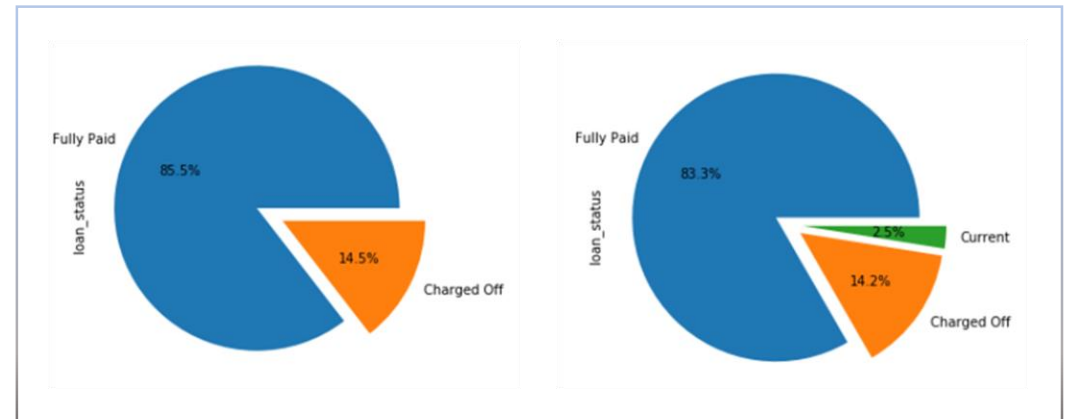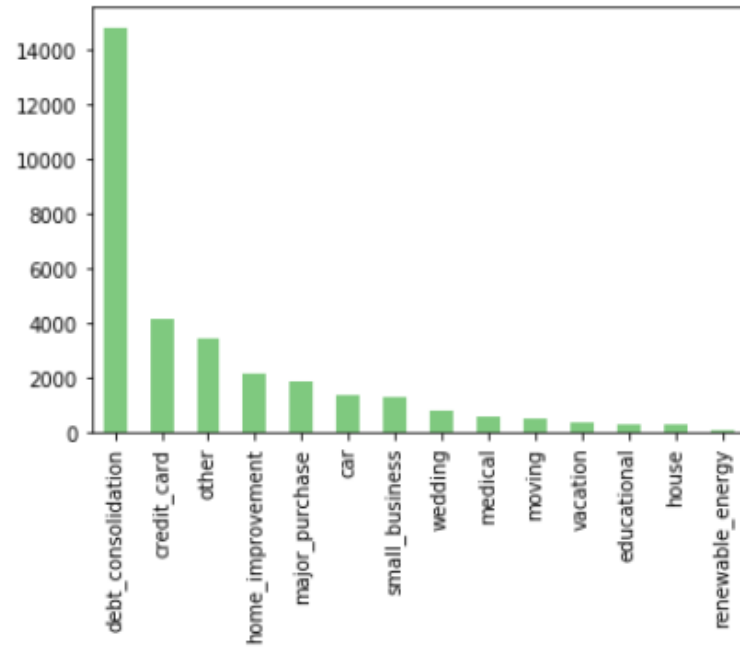
Number of loan fund Month-Year wise

This picture shows the statistics that how many loan funded at a period.
We can clearly see that the graph is increasing from Jun-2007 to Nov-2011 that means the company is in profit and doing good business by providing loan to the people
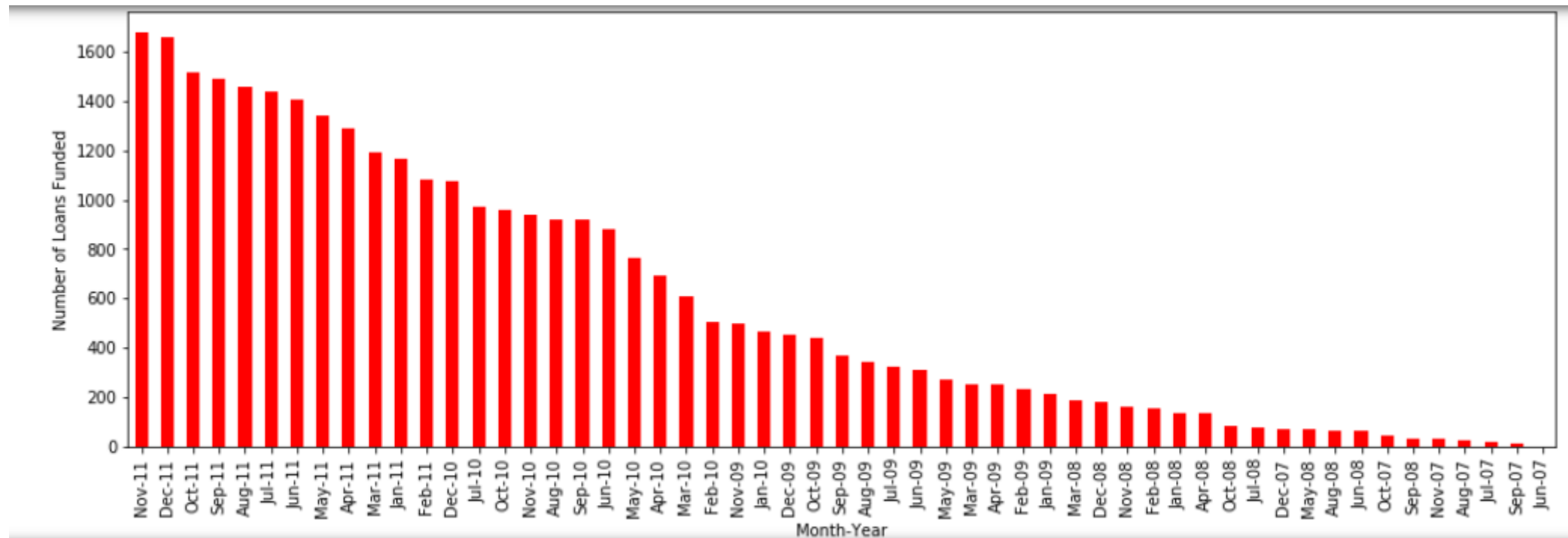
Purpose of Loan

- Here is very interesting insight that most of the people more than 60% taking loan to repay the other debt_consolidation

Loan_status Vs Purpose

From this insight most of the people who took loan has Fully paid their loan and around 10 to 20 % who failed to pay.

Interesting one is that approximately 27% people who took loan for small_business, are failed to pay the loan this give us insight that loan given for the purpose for small_business is more risky than others
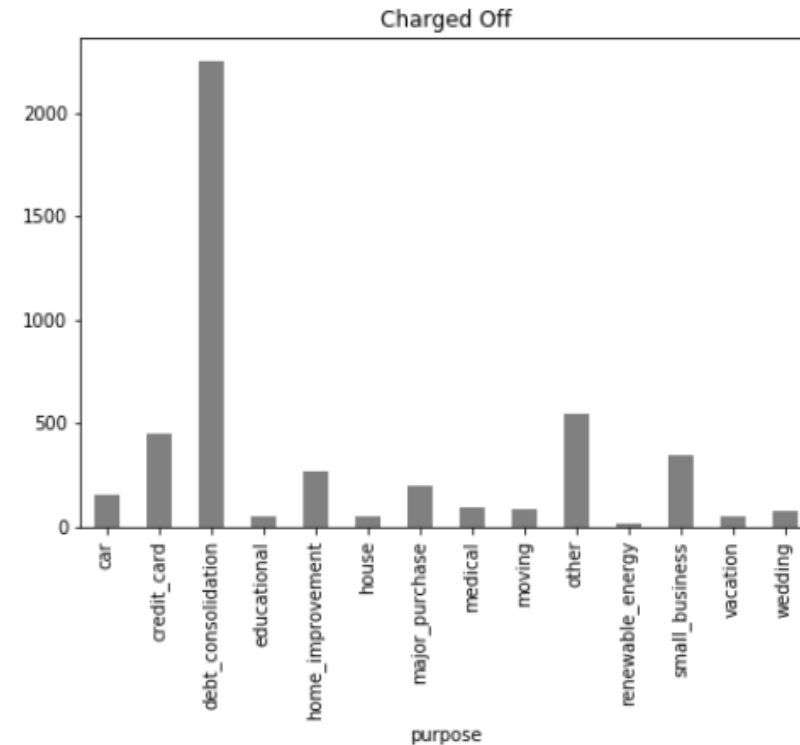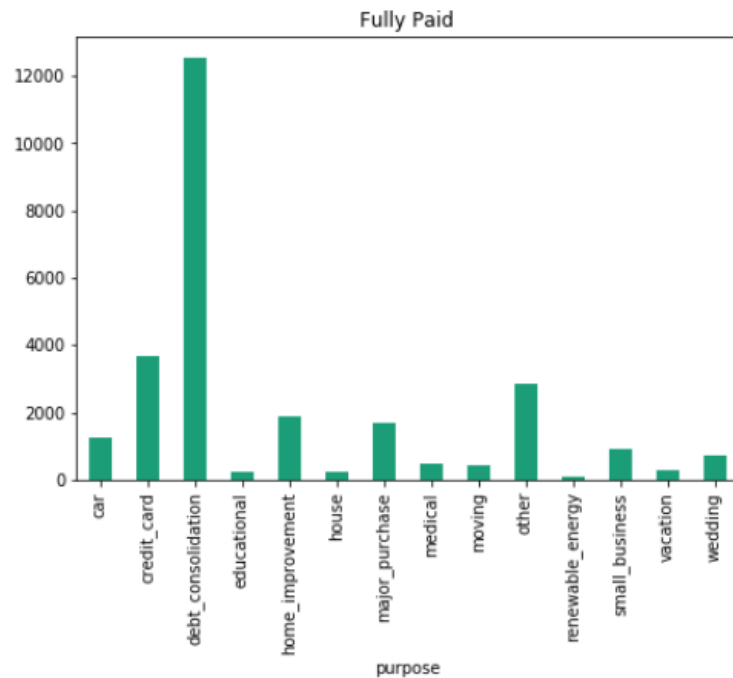
Term Vs loan_status

In this we can see that most of the people took loan for the period of 36 months and also who have taken loan for 36 months are more likely to Charged off and fully paid their loans
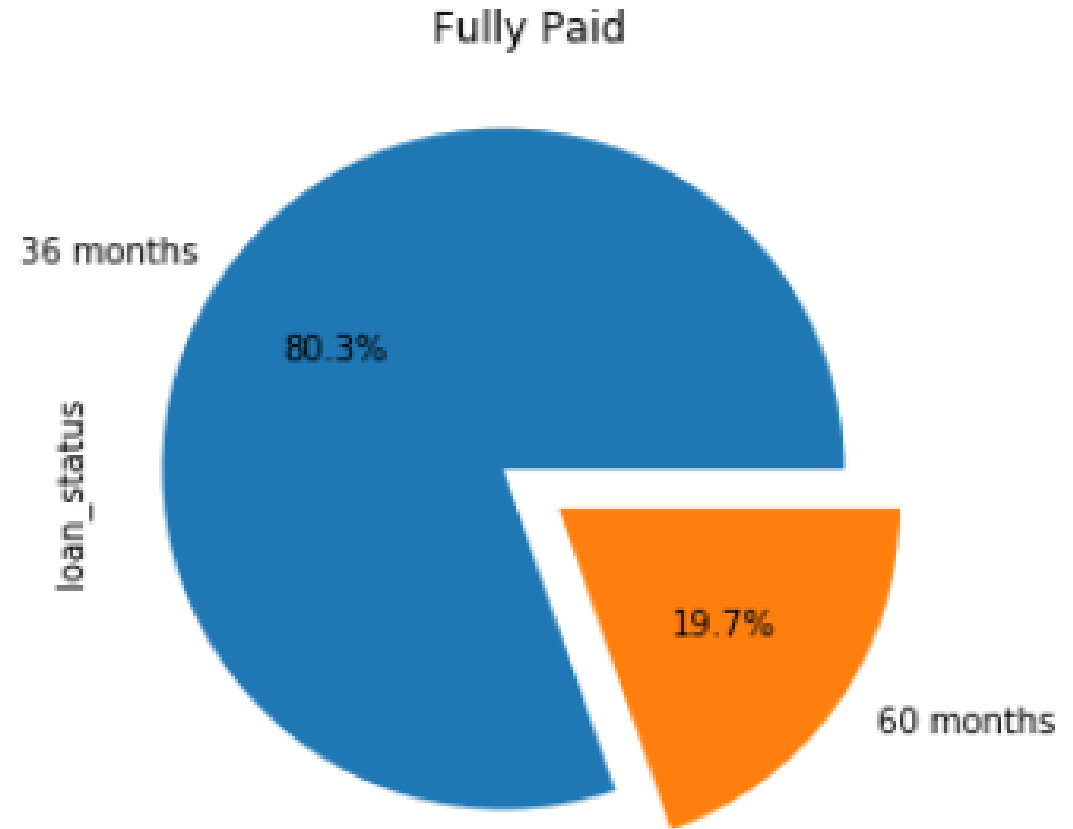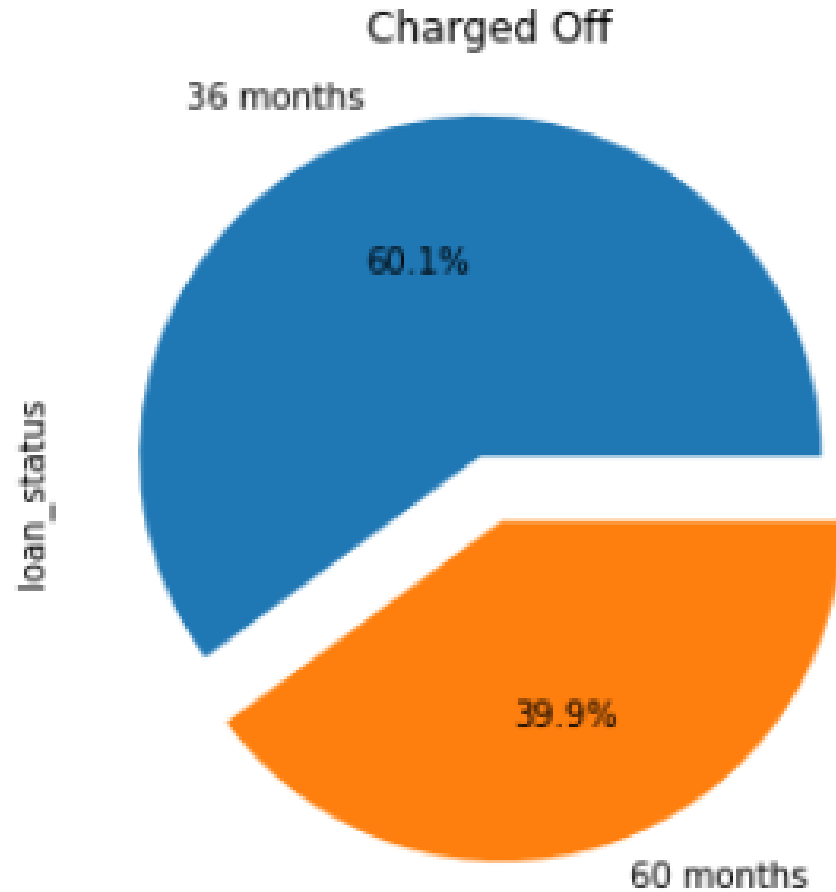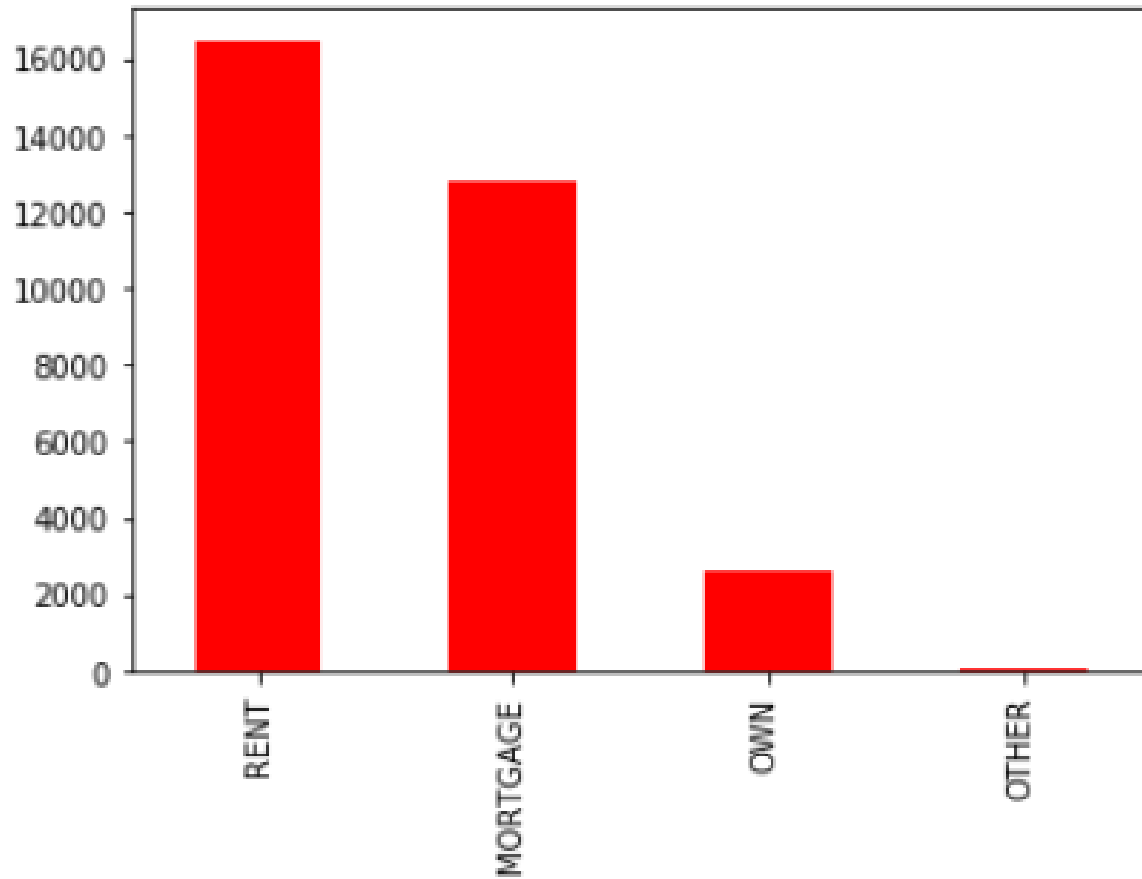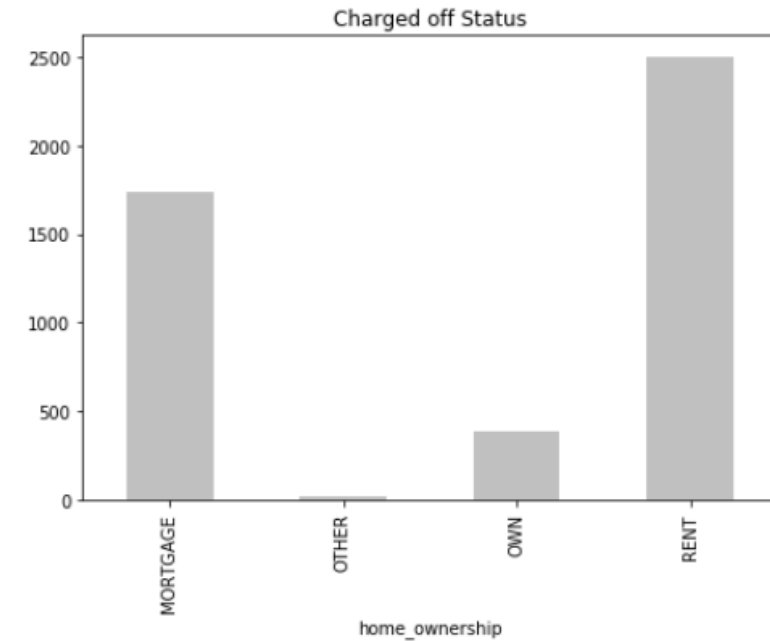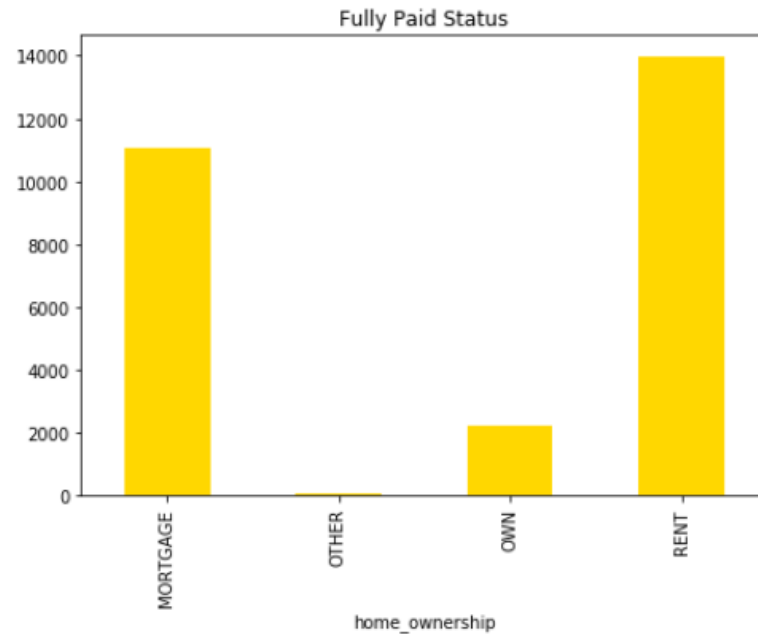
House_ownership

Most of the people who lives in a rented place are taking more loans than other

Loan_status vs home_ownership

Here the highest percentage is approximately 19% people who Charged off their loan having home_ownership status 'Other'. Around 10 to 15% with different home_ownership status are also Charged off their loan
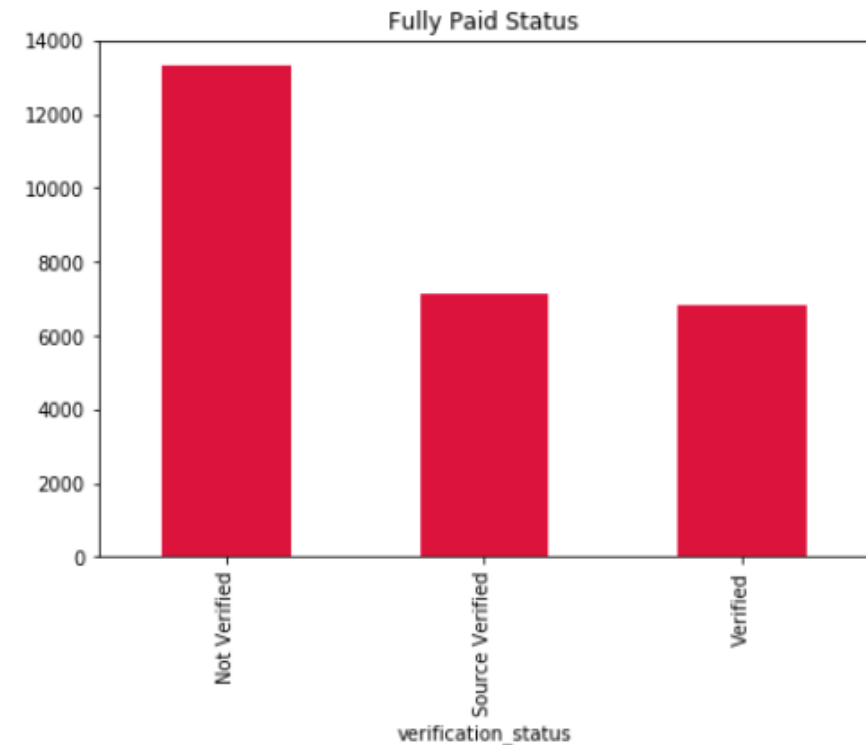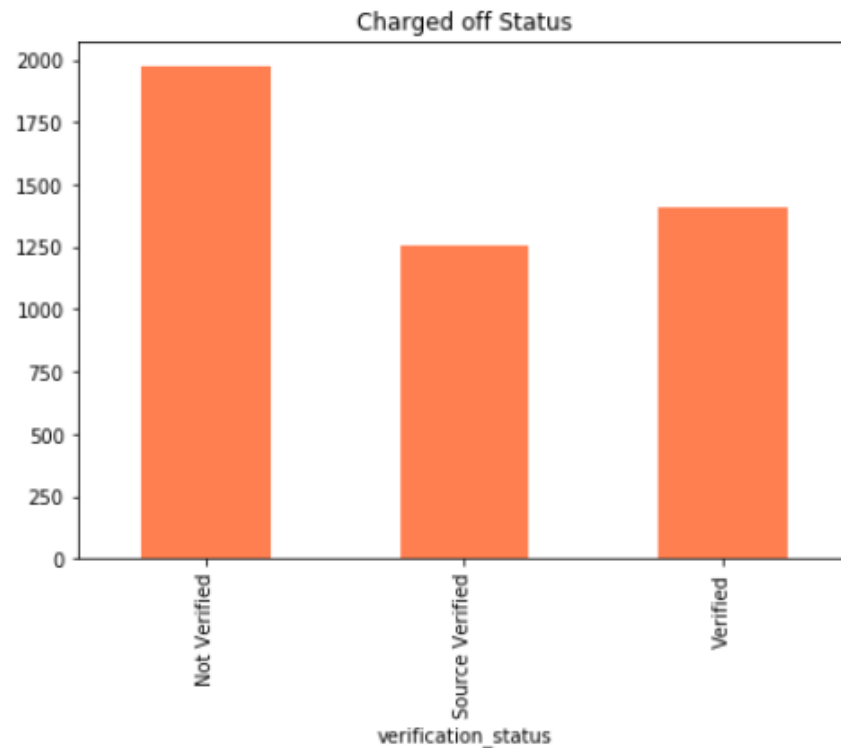
Loan_status Vs verification_status
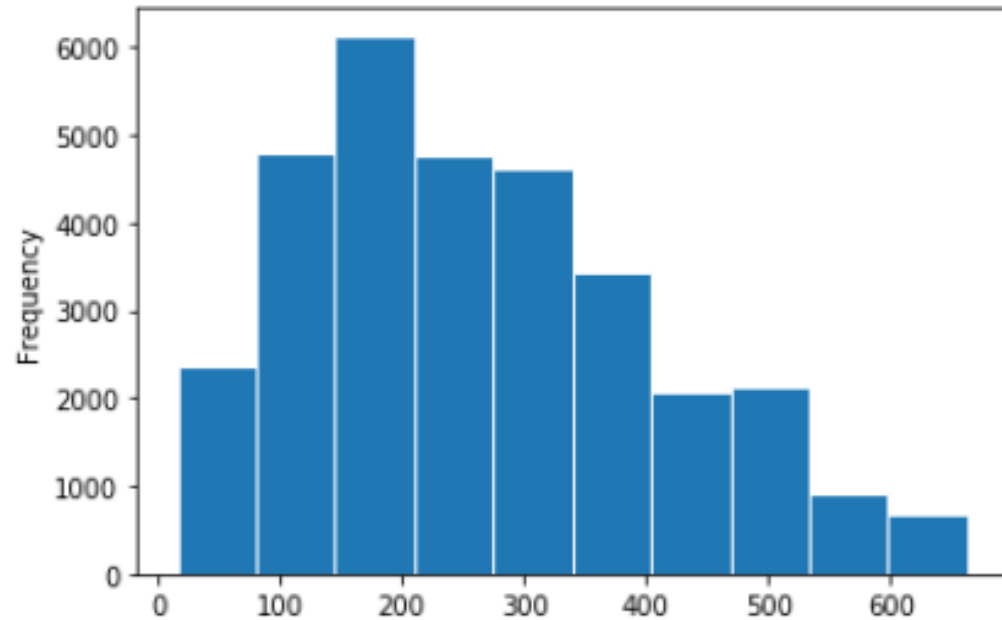
Here the insight is that :-

The people whose income is Verified are more likely to Charged off their loan (around 17%)

The people whose income is Source verified are Charged off their loan (around 15%)

The people whose income is Not verified are also Charged off their loan (around 13%)
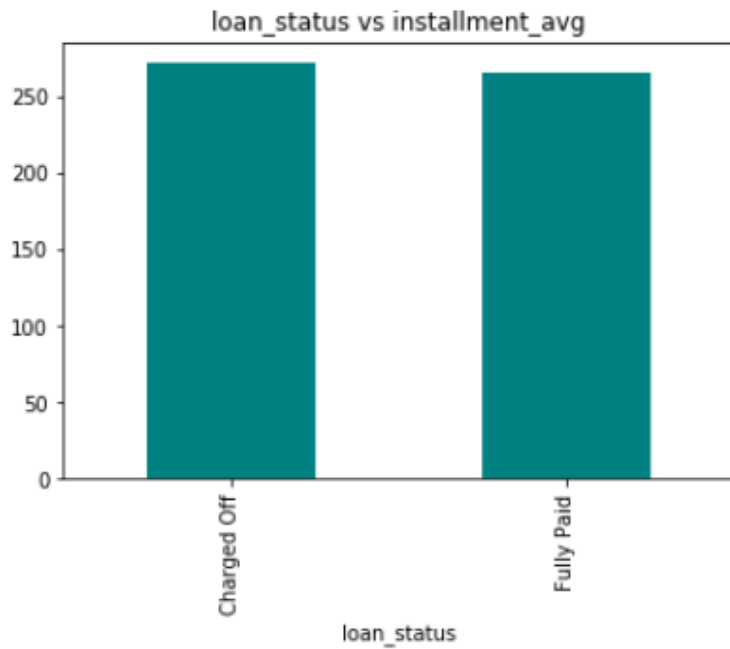
Histogram of installment

- Here the average installment is around Rs 270 per month and highest is 664 Rs. Per month and lowest is around 17 Rs approx.
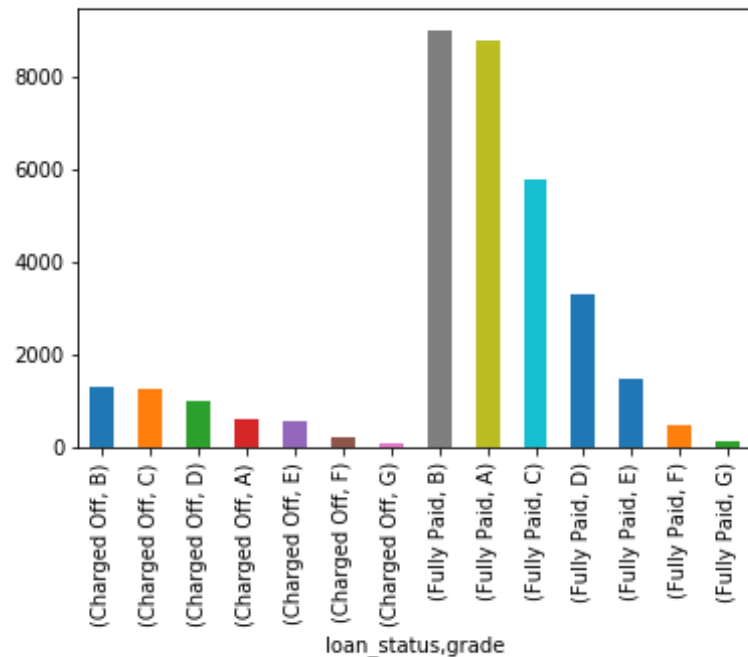
Loan_status Vs Installment

loan_status vs installment_avg

- In this the people who Charged off having average installments around 336 Rs. And people who have Fully Paid their loan having average installments around 320 Rs. From this we conclude that we have to make installments of less range.
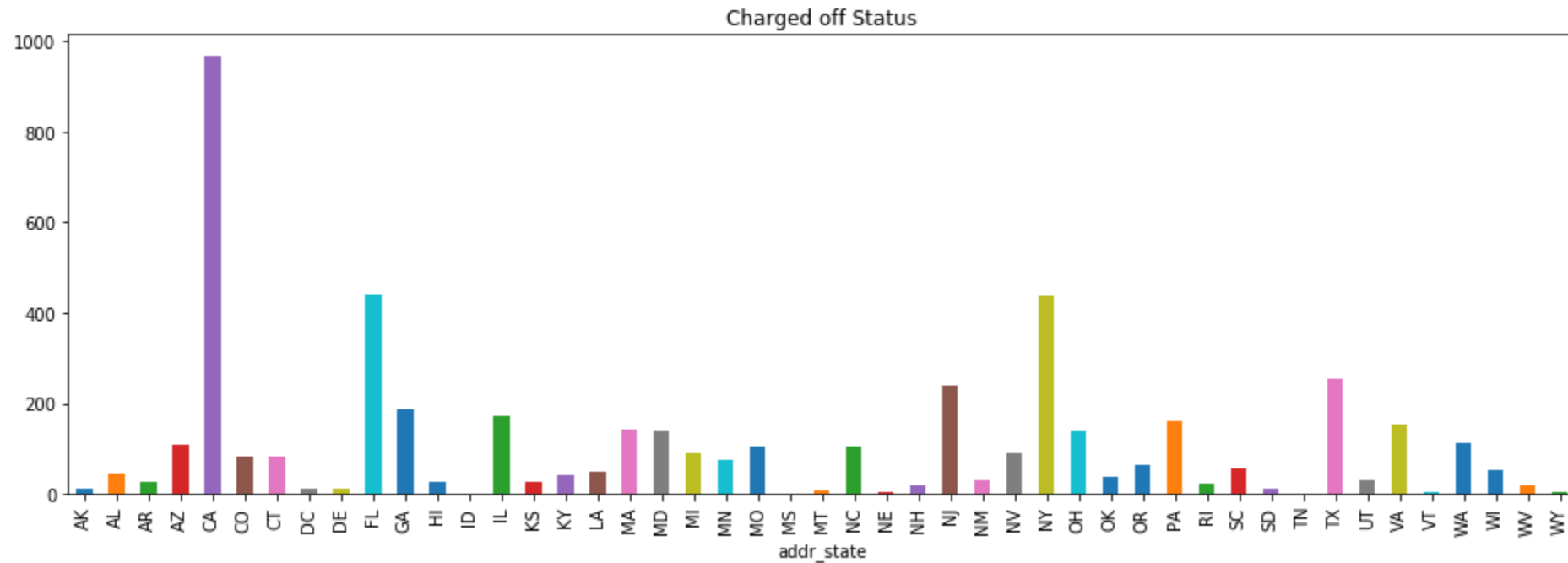
Loan_status Vs grade

In this the grade from A to G represents risk from low to high, as shown in figure we can insight that till grade C we have less risk of charged off loan that is less than 20% after grade C we have high risk of Charged off loans as it is more than 20% and grade E have highest percentage approx. 33% who Charged off their loans

addr_state Vs loan_status

State named as 'CA' having more cases of Charged Off loans approximately 30%

Thank You