UPGRAD ASSIGNMENT

Our aim was to find out the reasons or the type of borrower who do not pay the installment and were declared as defaulter:-

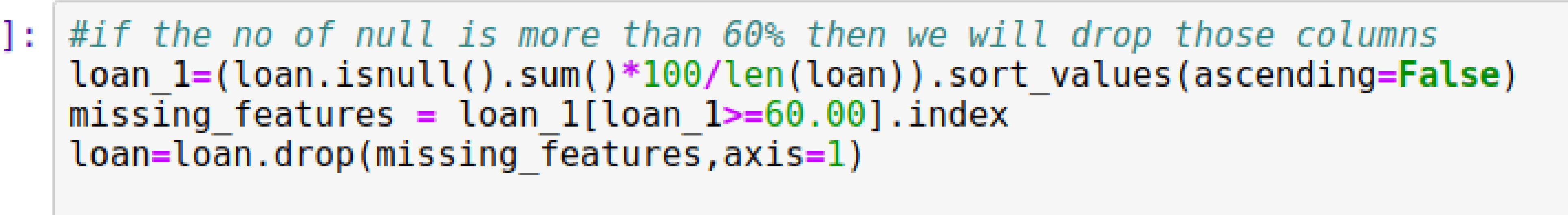
The columns which we have used are:-

- Annual income
- Interest rate
- Loan_status
- Pub_rec_bankrupcties rec
- Loan amount
- Verification status
- Grade(on the bais on record of borrower)
- Employment length

IN COLUMN LOAN INTEREST THE % SIGH HAD BEEN REMOVED TO MAKE THE COLUMN FLOAT AND MAKE IT SUITABLE TO PLOT GRAPHS

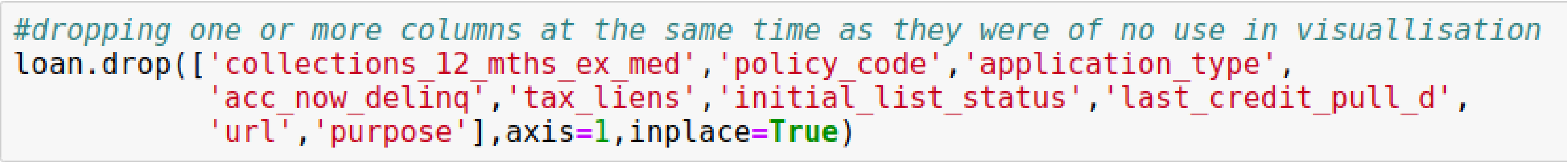
Data cleaning and Manipulation:-

- The columns which were having more than 60% of null values were removed . So from 111 columns only 45 columns were left after dropping such columns



```
]: #if the no of null is more than 60% then we will drop those columns
   loan_1=(loan.isnull().sum()*100/len(loan)).sort_values(ascending=False)
   missing_features = loan_1[loan_1>=60.00].index
   loan=loan.drop(missing_features,axis=1)
```

- Also some useless columns were dropped which seems to be of no use

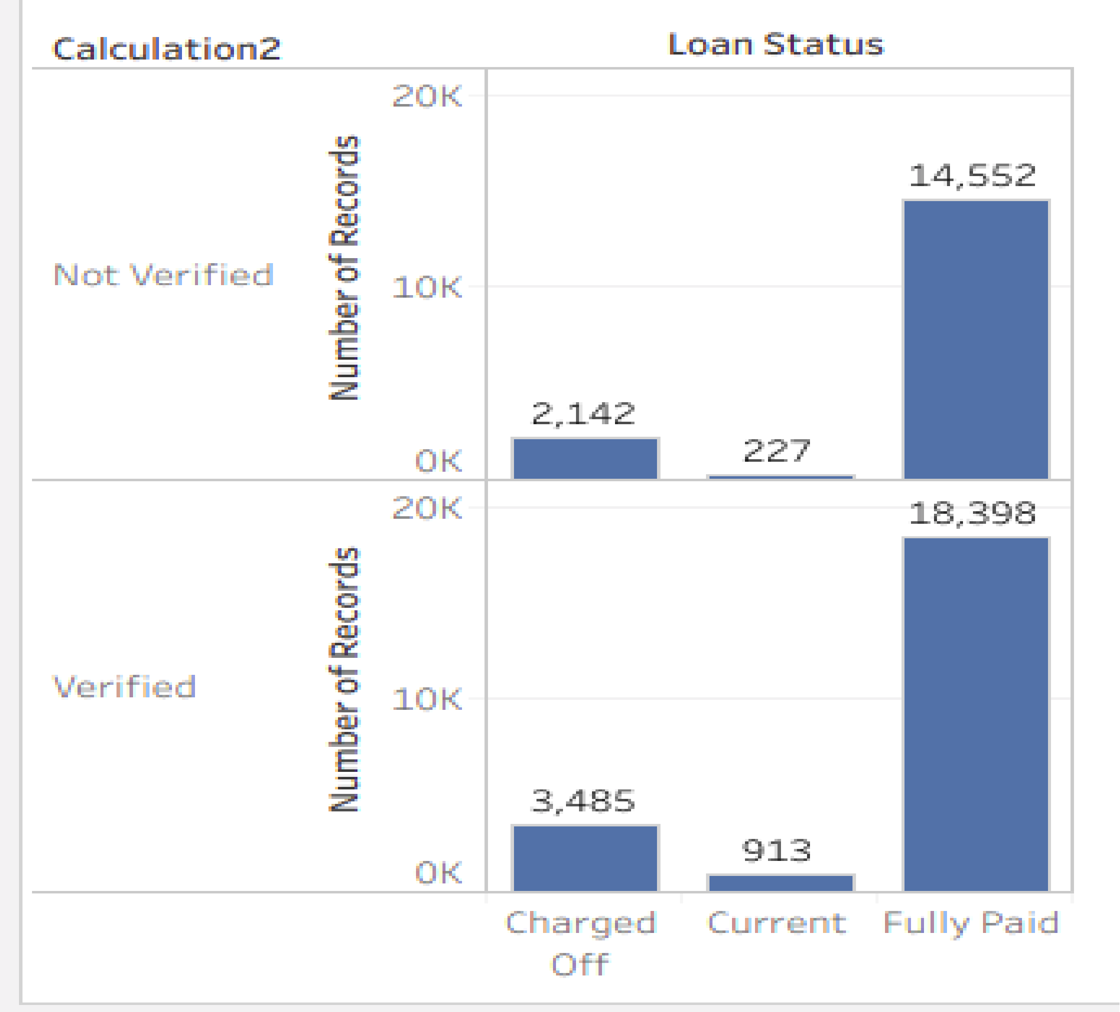


```
#dropping one or more columns at the same time as they were of no use in visualisation
loan.drop(['collections_12_mths_ex_med','policy_code','application_type',
           'acc_now_delinq','tax_liens','initial_list_status','last_credit_pull_d',
           'url','purpose'],axis=1,inplace=True)
```

- There was no duplicate values in loan id column.
- In column verification_status there were two values "source verified " and "verified" so we replaced all "source verified" to "verified".
- A new data set "newloan" had been made which consists record only for loan_status="charged off"

Analysing on the basis of verification status :-

- If we see the "charged off" bar of not verified ,



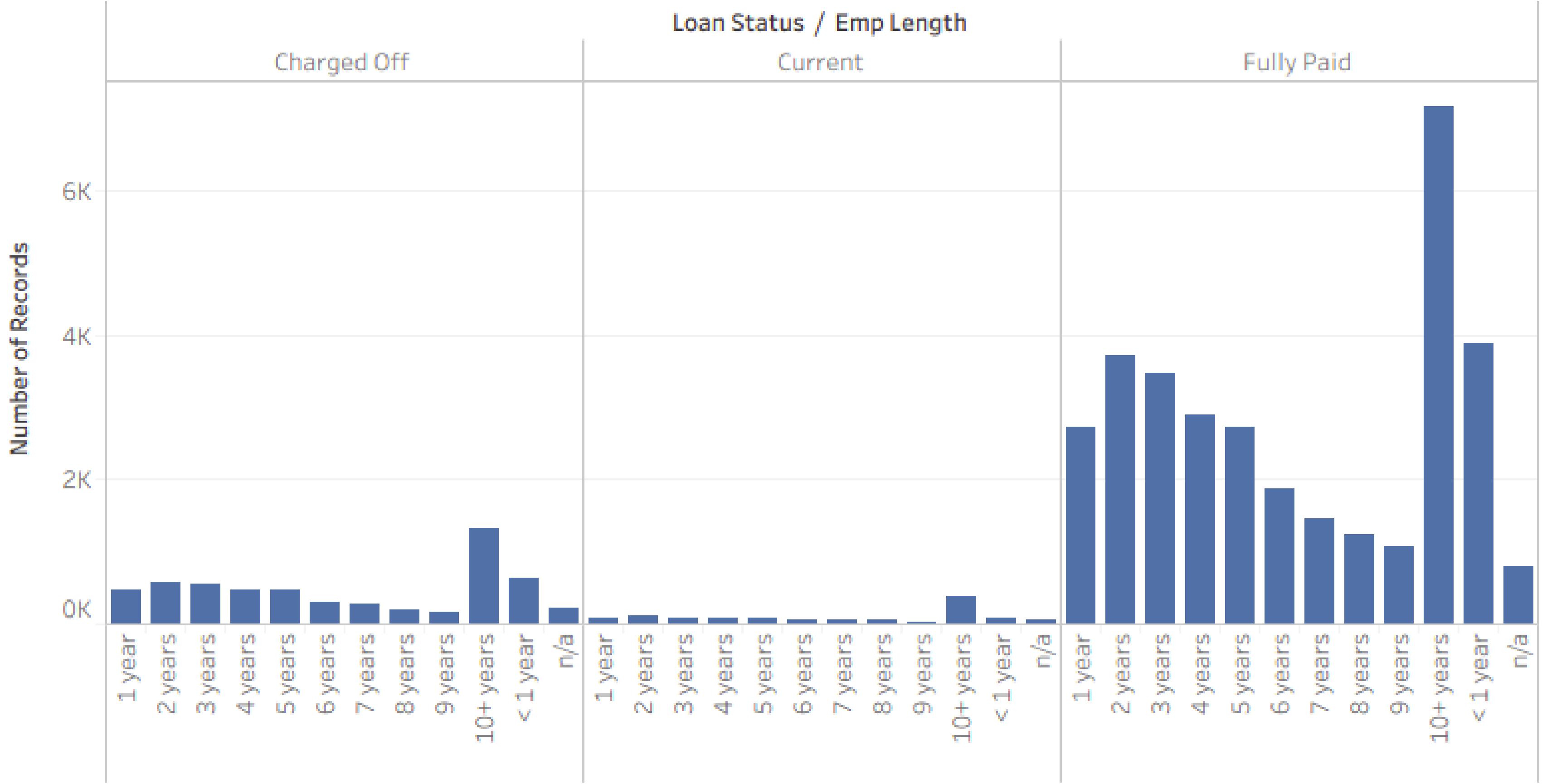
"Charged off" ratio in both the cases:--

Notverified=2142÷(2142+227+14552)=1.26
Verified=3845÷(3845+913+18398)=0.156

- There is not much difference in both the values as we are observing total 39717 no of records

- So the company should work more in verification area and provide loan to that people who clears all the verification

Analysing on the basis of emp length



- If we observe the ratio of records of loan stats "charged off" and "fully paid" so we cant judge on the basis of emp length .But we can say that most of the records had been seen from the people who had 10+ employment length
- The bank is benefitted with people with 10+ years emp length as the record of charged off is too small as compared to "Fully Paid"

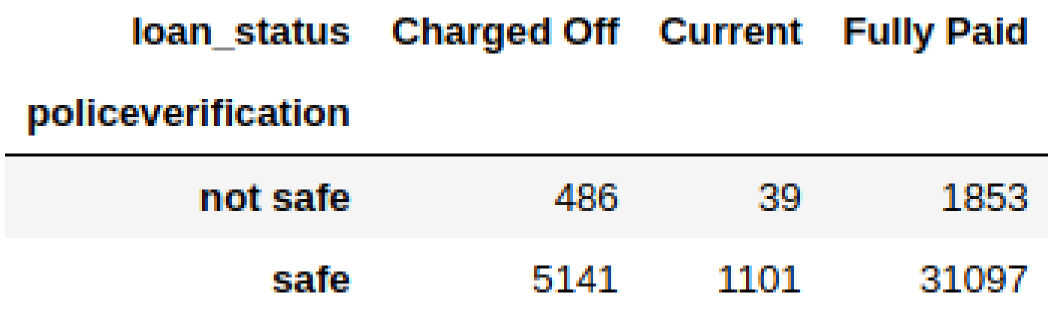
Some categories in which data is divided:-

LOAN AMOUNT	CATEGORY
LESS THAN 6000	LOW_LOAN
6000 TO 20000	MODERATE_LOAN
20000 TO 50000	HIGH_LOAN
MORE THAN 50000	VERY_HIGH

INTEREST RATE	CATEGORY
LESS THAN 10.0	LOW_INT
10.0 TO 15.0	MODERATE_INT
15.0 TO 20.0	HIGH_INT
MORE THAN 20.0	VERYHIGH_INT

PUB_REC_BANKRUPTIES	CATEGORY
0	SAFE
1	NOT SAFE

ANALYSIS ON THE PUB_REC_BANKRUPTICIES

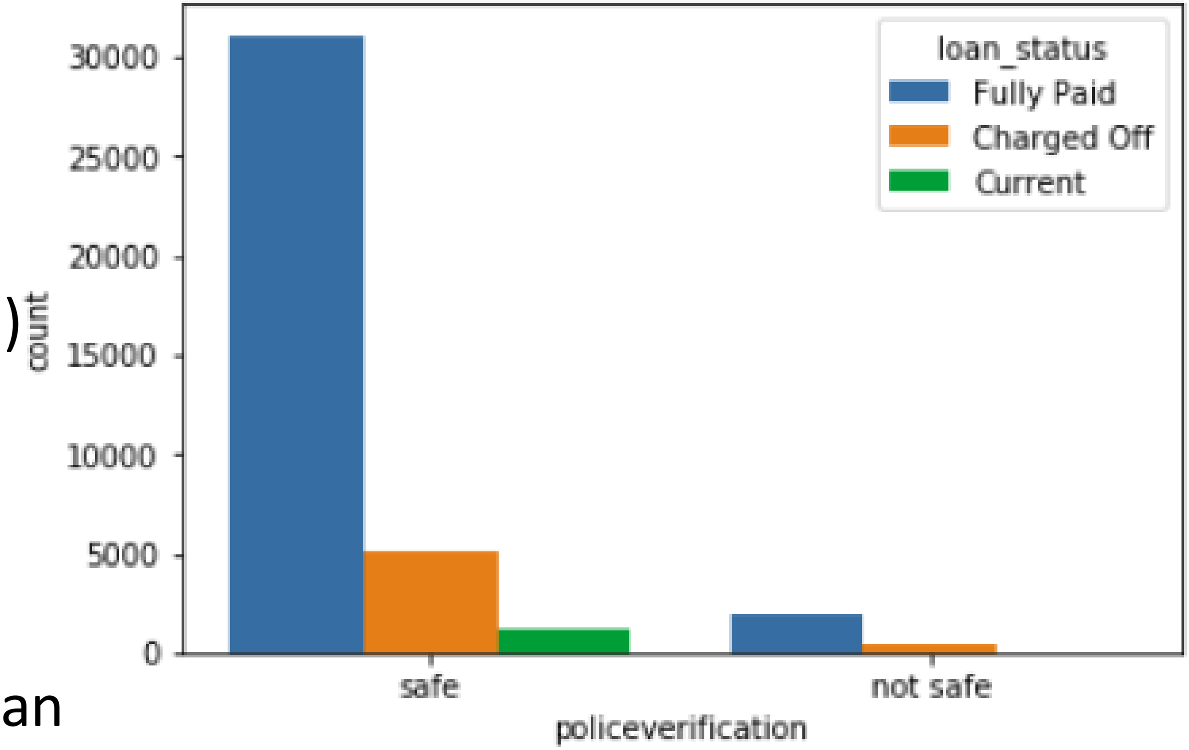


loan_status policeverification	Charged Off	Current	Fully Paid
not safe	486	39	1853
safe	5141	1101	31097

If we compare the ratio of total no of record to the loan staus "charged off" we get (from the cross tab)

Safe=5141/(5141+1101+31097)=0.137
Not safe=486/(486+39+1853)=0.2044

- So the bank is bearing a huge loss by providing loan to the category of not safe as the ratio of not safe "charged_off" is almost double than the safe one.



Analysis on the loan amount:-

- For the loan datasetMost of the loan is provided to moderate_income type and it has less no of Charged off record if we compare ratio with others

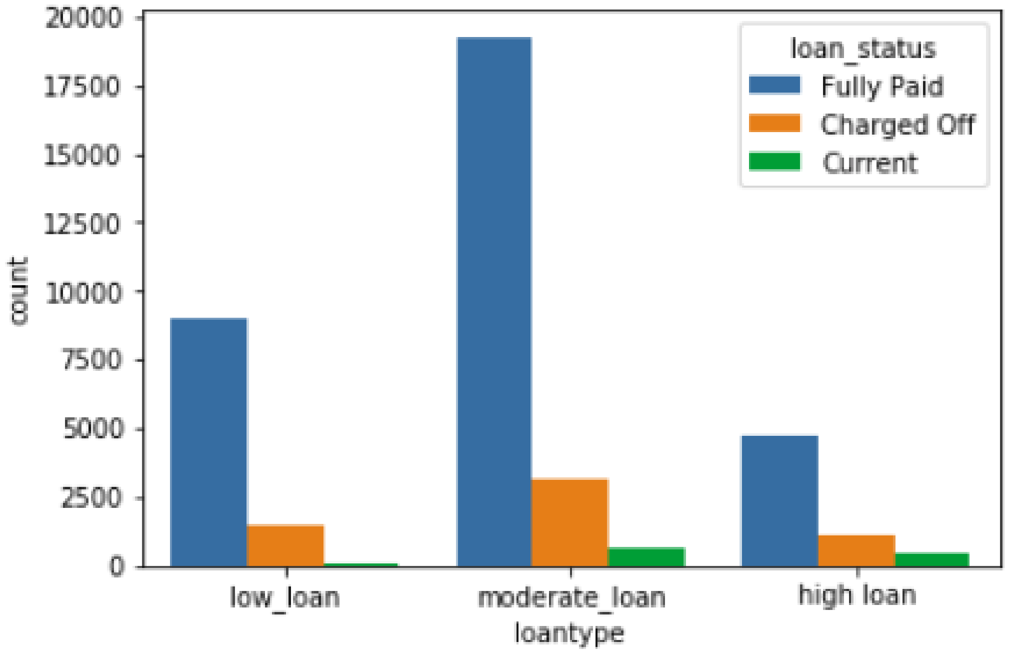
……………………………………………………………………………………

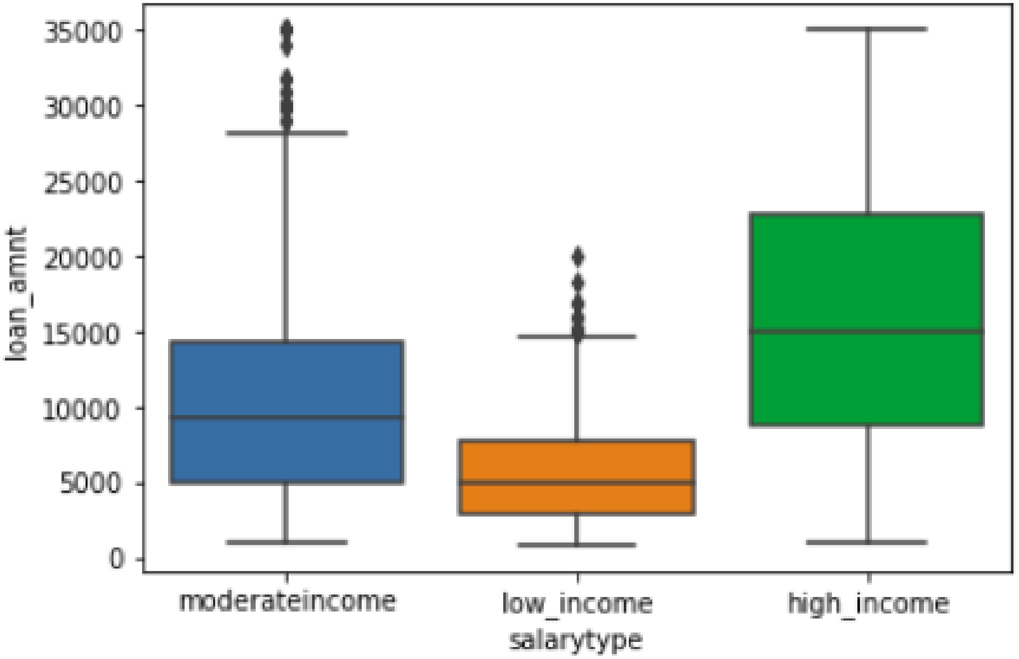
For the newloan datset(which consists of dataOnly for 'chargedoff' type loan_status)

- By box plot we analyse the loan_amnt with the Annua.. income of borrowers:-
- So we got maximum no of cheaters on the loan amount For moderate_income=approx 1000 to 15000

Low_income=approx 5000 to 7000

High_income=15000 to 24000

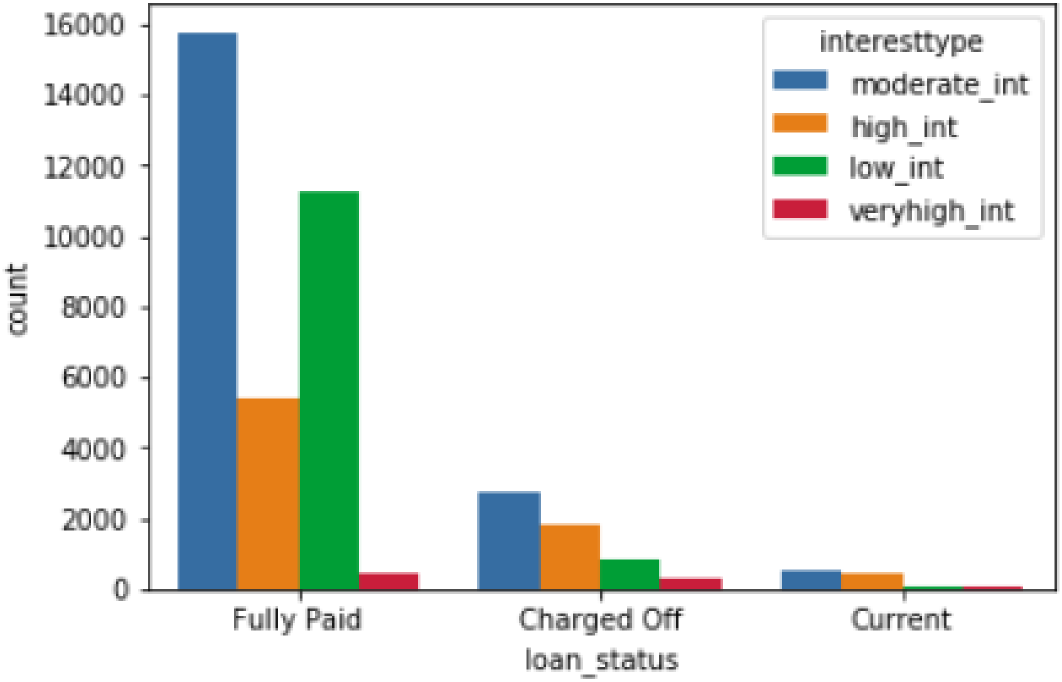




Analysis on loan interest:-

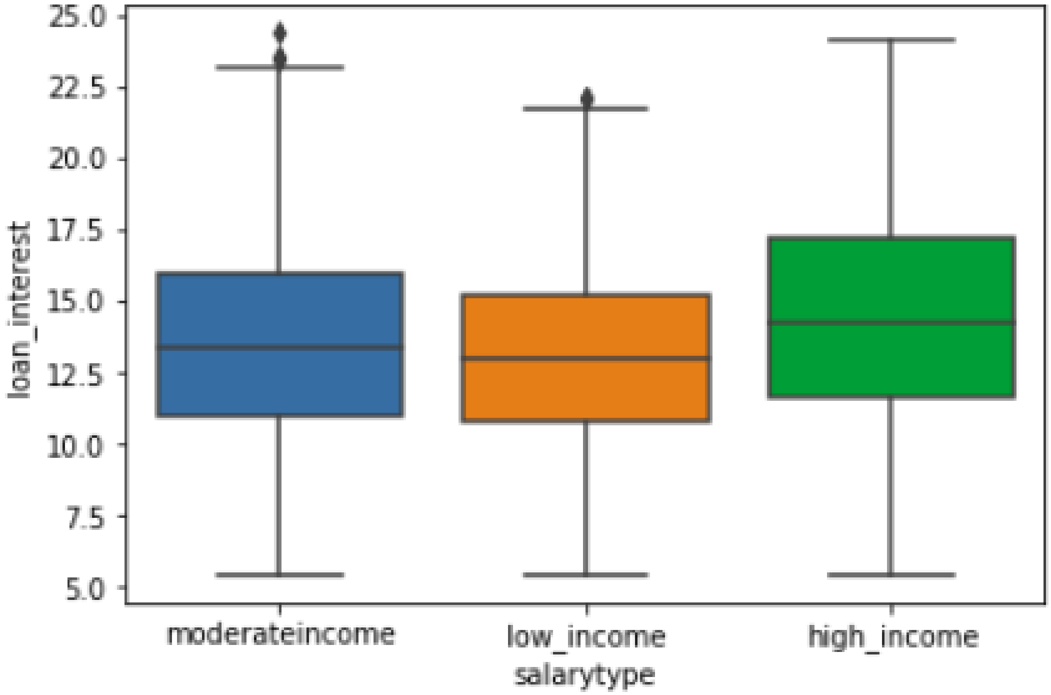
From loan dataset :

- observing loan status we can see that most of the record were for moderate interest
- High_interest comes under the most unpaid loan if we compare the no of records in fully paid and charged off loan status



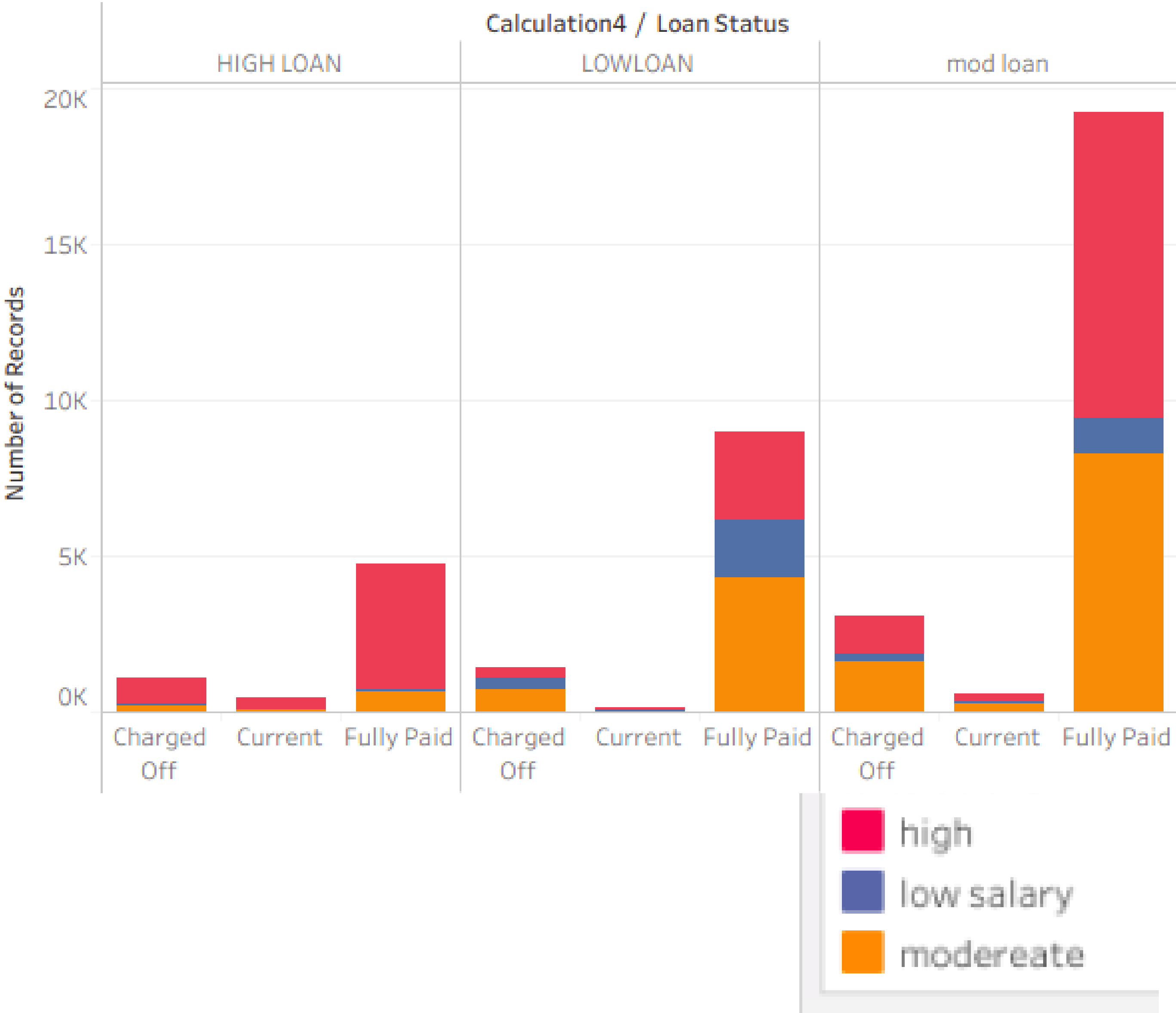
- From newloan dataset(consisting data only for value=charged off)
- Most of the people with these annual income fail to pay back the loan amount at these interest rate:-
- Moderate income=13 to 16.5
- Low_income =12.6 to 15.0
- High_income=14.5 to 17.3

(all the loan interest comes under moderate interest)



Dependency of loan amount and annual income and loan status

- We can see that most of the loans were borrowed of mod loan type from each type of annual income
- High loan were taken by almost by high salary people
- From this plot we can observe that very less people with low salary had taken loan from the bank
- Also the people with defaulter status are from moderate annual income but we can ignore this as ratio of total record of moderate salary and "charged off " status is low



Recommendations:-

- The bank should launch new policies so that the person with less employment length can opt for loan as most of the loan were taken by the people who had emp length 10+ years
- The bank should made a reform in the verification process as both the verified and not verified people are caught traitor if we compare the ratio
- The bank should take an affidavit by the people with pub_rec_bankruptcies record as not safe because most of the people who did not pay the loan is under this category
- Bank should make some new policies for low annual income people such as :-

1. providing loan with very low interest.
2. The loan duration should be long to make the full payment

- Almost all the loan amount had been fully paid with had low amount and low interest so the bank should not reject any loan id with low loan amount and keep low interest
- High interest should be only applied for those borrowers who had high or very high salary
- Or if provided high interest to moderate annual income then the loan duration should be enough months or years