Lending Club Analysis

The dataset has been obtained from https://www.kaggle.com/wordsforthewise/lending-club.

In [1]:

```python
import pandas as pd
```

In [2]:

```python
import warnings
warnings.filterwarnings("ignore")
```

In [3]:

```python
# zf = zipfile.ZipFile('/content/drive/MyDrive/test/archive.zip')
# zf.extractall('temp')
```

In [4]:

```python
accepted = pd.read_csv('temp/accepted_2007_to_2018q4.csv/accepted_2007_to_2018Q4.csv')
```

In [5]:

```python
from eda import EDA
eda = EDA(accepted)
```

Data Cleaning

In [6]:

```python
data=eda.cleanData() #cleaning dataset
```

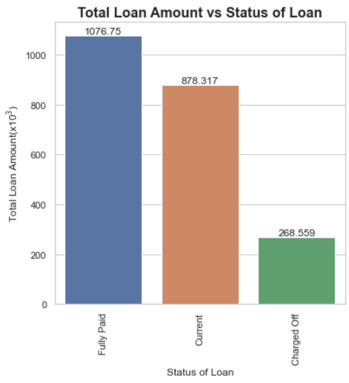
Exploratory Data Analysis

In this section there are 10 subparts with visualizations accompanied by its description.

1) Total Loan Amount vs Status of Loan

In [7]:

```python
eda.plot1(data)
```



```
Percentage of loans that are fully paid= 47.63 %
Percentage of loans that are current= 38.85 %
Percentage of loans that are charged off= 11.88 %
```

Insights:

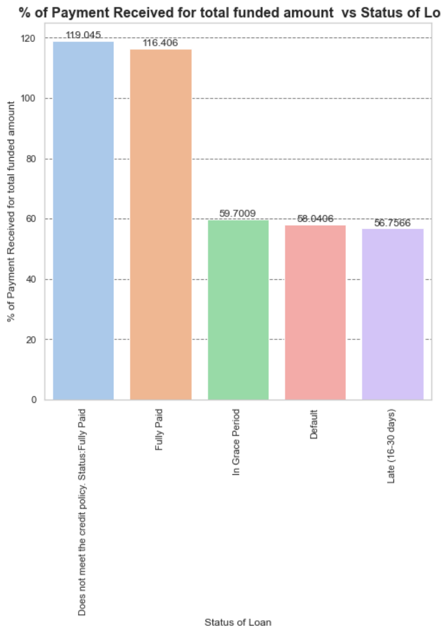
From the plot of Total Loan Amount vs Status of Loan for Lending Club, we can see that around 47.63% of loans are paid off loans and 11.88% are charged off loans from the time period of 2007 to 2018.

The charged off loans can be considered to be a loss for the Lending Club as it would no longer receive further payments for the same. The Lending Club can try to sell these charged off loans to a third party. The ongoing loans are approximately 38.85%.

2) % of Payment Received for total funded amount vs Status of Loan

In [8]:

```python
eda.plot2(data)
```



Insights:

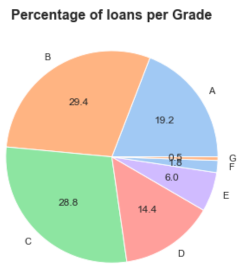
The above plot represents the payments received till date for total amount funded with respect to the type of loan status. The value above 100 for fully paid loans indicates that the payment was commpleted and the Lending Club made around 19% profit on the same. The Lending Club is able to recover approximately 58% of loans that are defaulted.

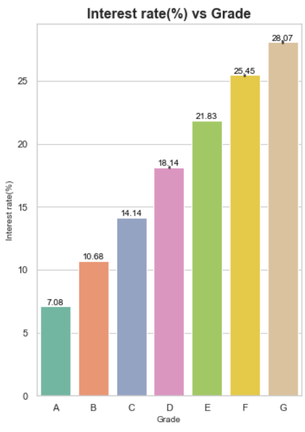
Also, Lending Club is able to recover only 60% and 58% of in grace period and late loans respectively which makes it clear that such loans result in a loss for Lending Club.

3) Types of loans & interest rates per Grade

In [9]:

```python
eda.plot3(data)
```





Insights:

Lending Club assigns a grade to the loan type ranging from A to G.

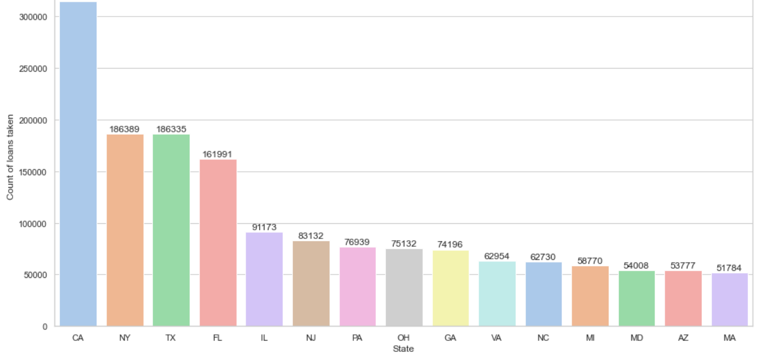
From the left plot, we can see that B and C grade category have the maximum loans granted. The best grade A, is offered approximately 19.2% of the total loans. As the grades get lower, the loans offered to them are also less i.e. grades like E, F and G have 6%, 1.8% and 0.5% respectively of the total loans offered.

From the right plot it is observed that, A grade has the lowest interest rates and G grade has the highest interest rates.

4) States which took maximum number of loans

In [10]:

```python
eda.plot4(data)
```



Insights:

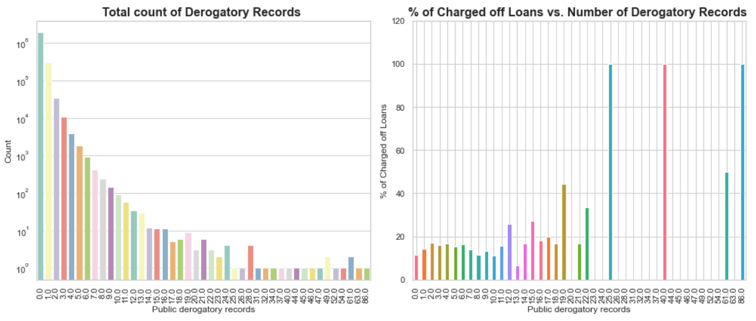
The above plot shows the top 15 states of USA that have taken the most number of loans from Lending Club from the year 2007-2018. The state of California has approximately double the count of loans taken in comparison to New York.

We can conclude that the top 5 states are California, New York, Texas, Florida and Illinois. Looking at the plot it makes sense that these states are in top 5 because of them being a hub for businesses and also because they are highly populated states.

5) Public derogatory records

In [11]:

```
eda.plot5(data)
```



```
Percentage of customers with no derogatory records= 84.17 %
Percentage of customers with one derogatory record= 13.51 %
Percentage of customers with two derogatory records= 1.51 %
```

Insights:

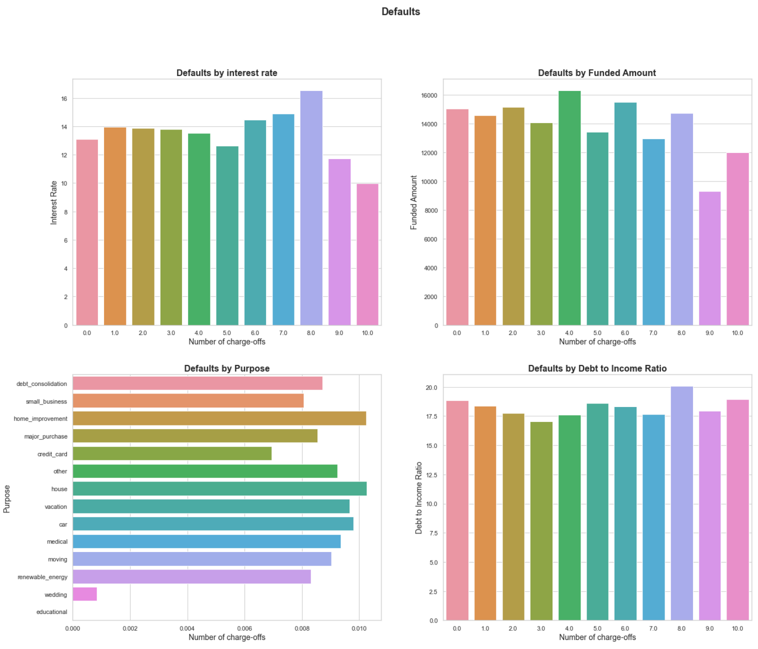
The chart on the left side shows the total count of the derogatory records with a majority of customers having no public derogatory record.

The chart on the right shows the percentage of charged off loans against the number of derogatory records. It shows that having more than 1 derogatory records increases the possibility of charged off loan.

6) Defaults by Interest Rate, Funded Amount, Purpose,Debt to Income Ratio(DTI)

In [12]:

```
eda.plot6(data)
```



Insights:

The top left plot is a plot of Interest rates vs number of charge offs. We can see that the interest rate is low in case of no chaarge offs. However, when the number of charge offs increase the interest rate increases as well. The interest rate for number of charges=5 is lower than the interest rate in the case of no charge offs. The Lending Club should take a look at this and investigate further.

The top right plot is a plot of funded amount vs number of charge offs. We can see that the funding amount is more in the case of number of charge offs=4, Lending Club needs to investigate this issue as well. Also, the Lending Club needs to take care further as there is a lot of money funded in number of charge offs greater than 3.

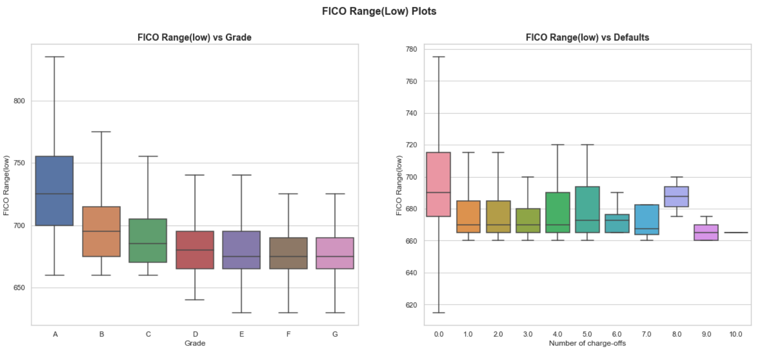
The bottom left plot is a plot of purpose vs number of charge offs. We can see that the most number of charge offs occur in the case of home loans and home improvement. Also, it is observed that in cases like education there are no charge offs occuring making them safe for investment.

The bottom right plot is a plot of debt to income(DTI) ratio vs number of charge offs. We can see that the DTI ratio is unable to give us a lot of information about the number of charge offs as for all the cases it seems to be around the value of 25%.

7) Lower Range of FICO Score by Grade, Defaults

In [13]:

```
eda.plot7(data)
```



Insights:

A FICO score ranging from 650 to 750 is considered as a good score. A customer having good FICO score indicates that he/she is more likely to get loan at low interest rates.

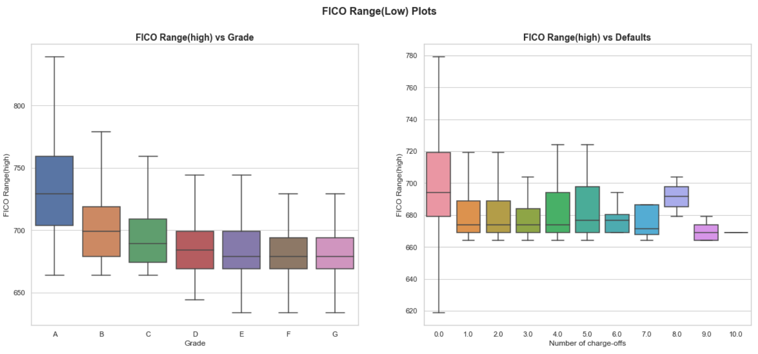
From the left plot, we can see that for grade A, the median of FICO score (low) is also a lot more than the rest of the grades. The medians for FICO score(low) is almost same in the case of the worst grades i.e. E,F and G.

From the right plot, we can see that for customers with no defaults have their median FICO score(low) higher than the rest.

8) Higher Range of FICO Score by Grade, Defaults

In [14]:

```
eda.plot8(data)
```



Insights:

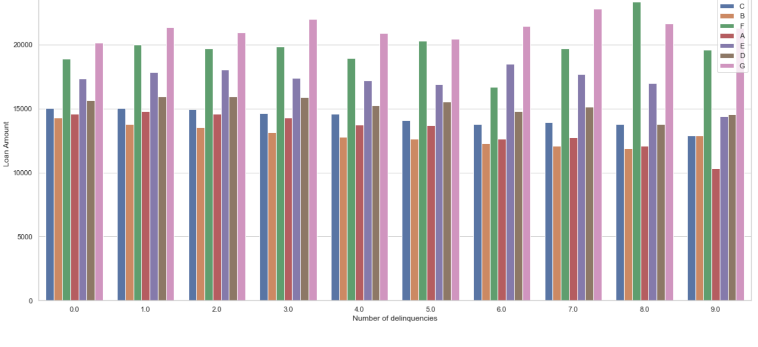
From the left plot, we can see that for grade A, the median of FICO score (high) is quite more than the rest. The medians for FICO score high is almost same in the case of the worst grades i.e. E,F and G.

From the right plot, we can see that for customers with no defaults have their FICO score (high) as compared to the other cases.

9) Loan Amount vs Number od delinquencies for each Grade

In [15]:

```
eda.plot9(data)
```



Insights:

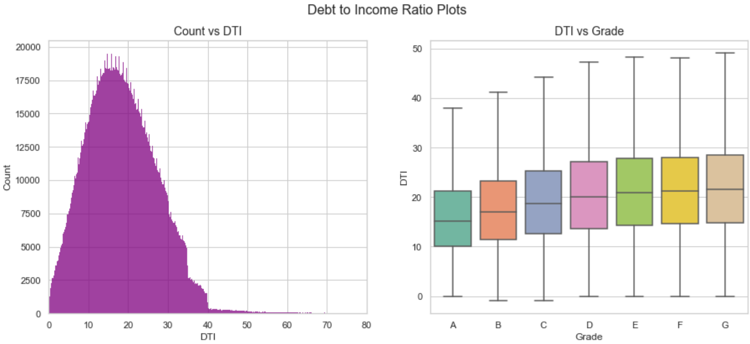
In the above plot we have delinquencies on the x axis and the loan amount on the y axis. Here, delinquencies means the number of times the borrower has been 30+ days past due on a payment of his/her loan.

From the above plot we can see that there is an investment of amount in cases where the number of delinquencies is greater than 3. Lending Club needs to take into account that it should restrict itself when the number of delinquencies is greater than 3 as these are not a safe investment.

10) Distribution of DTI as per Grades

In [16]:

```
eda.plot10(data)
```



Insights:

The debt-to-income (DTI) ratio helps to measure the income a person with respect to the debt. A low DTI ratio shows a balance between debt and income.

From the left plot it is clearly visible that the DTI ratio of approximately 20% is most occuring in the data. As the DTI ratio increases, it indicates that the customer has a lot of debt (perhaps more than the income earned).

The plot on the right side confirms our findings from the left plot, i.e. the best grade i.e. A and B have a low DTI ratio as compared to the poor grades which have a high DTI ratio indicating a lot of debt.

CONCLUSION

The Exploratory Data Analysis helped look at the Lending Club data in depth. The following conclusions can be drawn from EDA:

1) There are arouand 47.63% of loans which are paid off loans and 11.88% are charged off loans from the time period of 2007 to 2018.

2) The Lending Club makes around 19% profit on fully paid loans. Also, it is able to only recover approximately 58% of loans that are defaulted.

3) The interest rates are in accordance with the Lending Club policy i.e. it is correctly imposing higher interest rate for lower grades starting from D to G and has less interest rates for grades like A, B and C.

4) Lending Club should focus on states like California, New York, Texas, Florida and Illinois as these are the states from where most number of loans are taken.

5) A majority of customers have no public derogatory record. Having more than 1 derogatory record increases the possibility of charged off loan.

6) Investigate the case where the interest rate for number of charges equal to 5 is lower than the interest rate as compared to the case of no charge offs. Investigate further in the case of funding amount being more for number of charge offs equal to 4.

7) Lending Club should invest in loans with purpose like Education, so as to increase profits as it has no defaults till date.

8) The lower range of FICO Score(low) for grade A is a lot more than other grades indicating it is a good FICO score. Customers with no defaults have their FICO score(high) higher as compared to the other grades.

9) Lending Club can take steps to restrict itself by reducing its investments in cases where the loan amount is greater than 10,000 and number of delinquencies is greater than 3 as they are not safe.

10) A and B have a low DTI ratio as compared to the poor grades like E,F and G which have a high DTI ratio indicating a lot of debt.A low DTI ratio shows a balance between debt and income indicating A and B grades are good.