In [1]:
```
#import libraries

import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from pandas import Series ,DataFrame
import seaborn as sns
import statsmodels.api as sm

%matplotlib inline
sns.set_style('whitegrid')
```

In [2]:
```
#import dataset
dataset = pd.read_csv('https://spark-public.s3.amazonaws.com/dataanalysis/loansDa
dataset.head()
```

Out[2]:

	Amount.Requested	Amount.Funded.By.Investors	Interest.Rate	Loan.Length	Loan.Purpose
81174	20000	20000.0	8.90%	36 months	debt_consolidatior
99592	19200	19200.0	12.12%	36 months	debt_consolidatior
80059	35000	35000.0	21.98%	60 months	debt_consolidatior
15825	10000	9975.0	9.99%	36 months	debt_consolidatior
33182	12000	12000.0	11.71%	36 months	credit_carc

In [3]:
```
#required data for this project are
```

In [4]:
```
dataset['FICO.Range'][0:10]
```

Out[4]:
```
81174    735-739
99592    715-719
80059    690-694
15825    695-699
33182    695-699
62403    670-674
48808    720-724
22090    705-709
76404    685-689
15867    715-719
Name: FICO.Range, dtype: object
```

In [5]:
```
dataset['Interest.Rate'][0:10]
```

Out[5]:
```
81174     8.90%
99592    12.12%
80059    21.98%
15825     9.99%
33182    11.71%
62403    15.31%
48808     7.90%
22090    17.14%
76404    14.33%
15867     6.91%
Name: Interest.Rate, dtype: object
```

In [6]:
```
dataset['Loan.Length'][0:10]
```

Out[6]:
```
81174    36 months
99592    36 months
80059    60 months
15825    36 months
33182    36 months
62403    36 months
48808    36 months
22090    60 months
76404    36 months
15867    36 months
Name: Loan.Length, dtype: object
```

Data Cleaning

We need to remove 'months' from Loan.Length , '%' from Interest.Rate & we need to parse the string from FICO.Range

In [7]:
```
# import new dataset after cleaning up and ready for direct use
# Now we import another file which is data cleaned

loans = pd.read_csv('C:\\Users\\bittu\\loan.csv')
loans.head()
```

Out[7]:

	Interest.Rate	FICO.Score	Loan.Length	Monthly.Income	Loan.Amount
6	15.31	670	36	4891.67	6000
11	19.72	670	36	3575.00	2000
12	14.27	665	36	4250.00	10625
13	21.67	670	60	14166.67	28000
21	21.98	665	36	6666.67	22000

In [8]:
```
int_rate = loans['Interest.Rate']
loan_amt = loans['Loan.Amount']
fico_scope = loans['FICO.Score']
```

In [9]:
```
y = np.matrix(int_rate).transpose()
```

In [10]:
```
x1= np.matrix(fico_scope).transpose()
x2 = np.matrix(loan_amt).transpose()
```

In [11]:
```
x = np.column_stack([x1,x2])
```

In [12]:
```
x3 = sm.add_constant(x)
```

OLS - Ordinary Least Squre

In [13]:
```
model = sm.OLS(y,x3)
```

In [14]:
```
model_fit = model.fit()
```

In [15]:
```
print('the P values are :',model_fit.pvalues)
print('the R - Squared value are : ',model_fit.rsquared)
```

```
the P values are : [0.00000000e+000 0.00000000e+000 5.96972978e-203]
the R - Squared value are :  0.6566326246493586
```

In [16]:
```
#P values should be less than 0.05 we got 0.0000 it's good
#R values are in -1 to 1 and we got 0.657 it's good
```

Summary

We have linear multivarialbe regression model for intrest rate and based of this our multivariable for intrest rate based of fico score and loan amount

The intrest rate is influenced by both fico score and loan amount