LENDING CLUB CASE STUDY

Exploratory data Analysis on the loan data from Lending Club

Sunil Kumar C

Business problem:

- The data given contains the information about past loan applicants and whether they ‘defaulted’ or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.
- If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study
- In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Steps followed in EDA:

- Data Understanding
- Data Cleaning
- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
- Observations & Recommendations

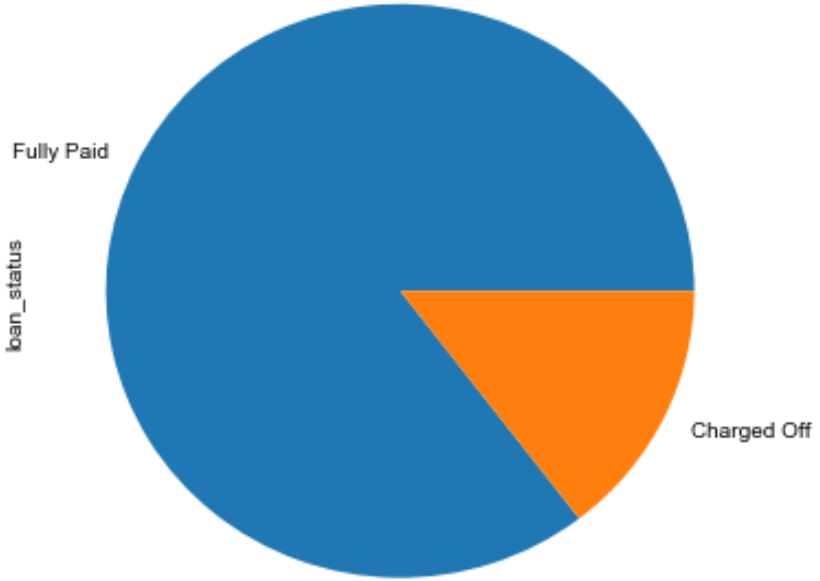
- **Data Understanding :** The total dataset consisted of 111 variables that describe the various aspects of an applicant such as Demographic information, Loan details as well as post loan payment data.

- **Data Cleaning :** There are many variables that does not help in analysis as per the business requirement of the Company which were removed.

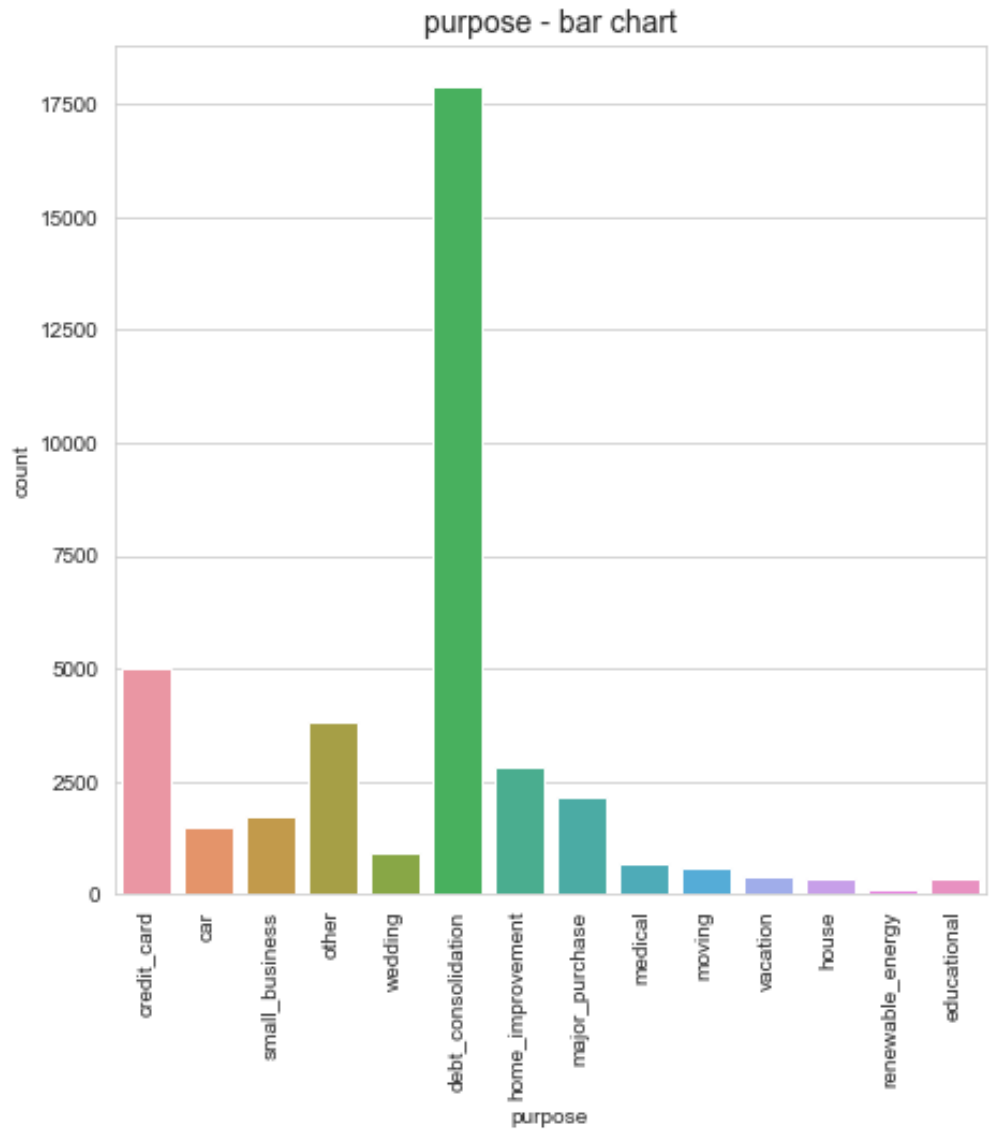
- **Data Analysis :** Once the data is cleaned and important variables are identified, Univariate, Bivariate and Multivariate analysis is performed

Univariate Analysis on Important variables:

- Out of total loans, 14% are charged off.
- Median loan amount is 10000.
- Interest rates range from 5% - 24% and the median is 11.7%.
- Majority of loans are graded A,B,C.
- Median Annual income is 58000.
- Majority of the loans are from CA, followed by NY, FL, TX



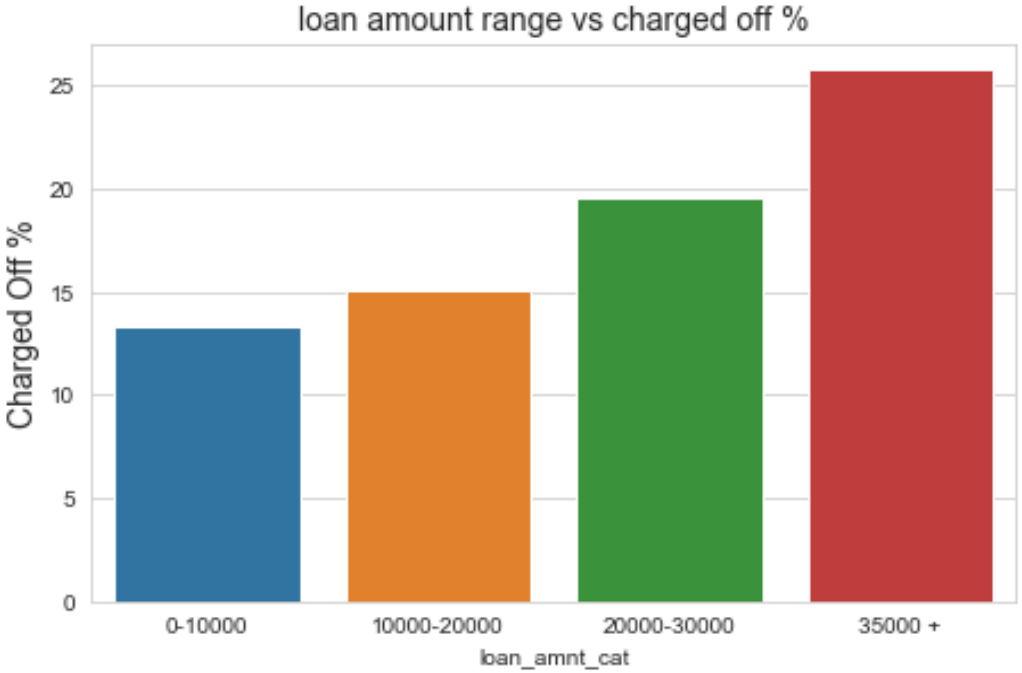
- The purpose for majority of the loans is Debt Consolidation followed by Credit card and others.
- The number of loans almost doubled every year from 2007 to 2011.
- The average instalment amount is 322 and the median is 277

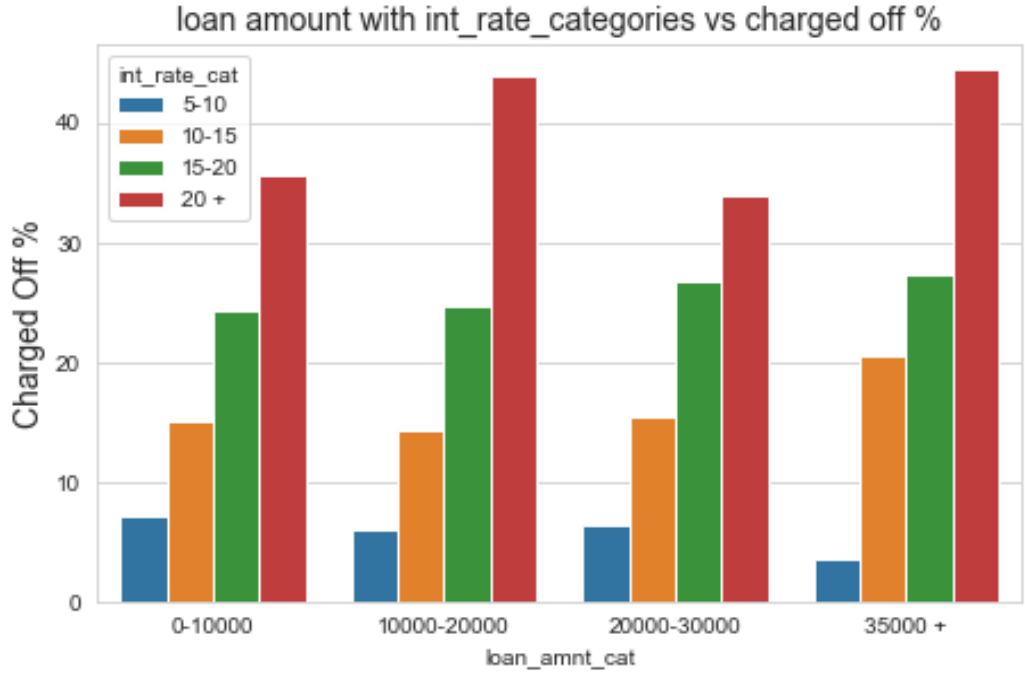


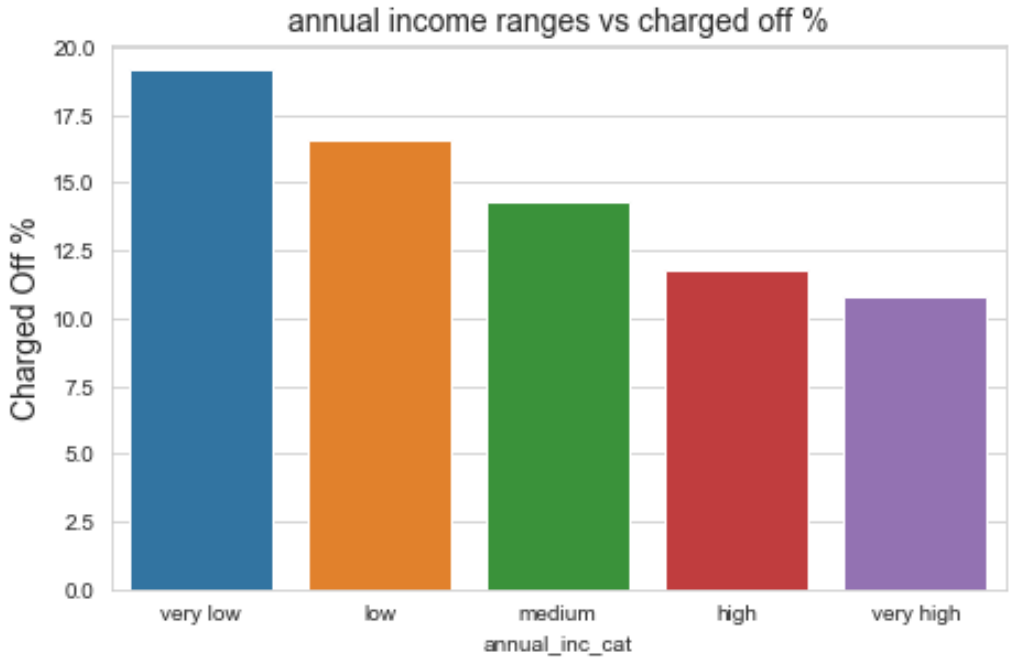
Bivariate & Multivariate Analysis:

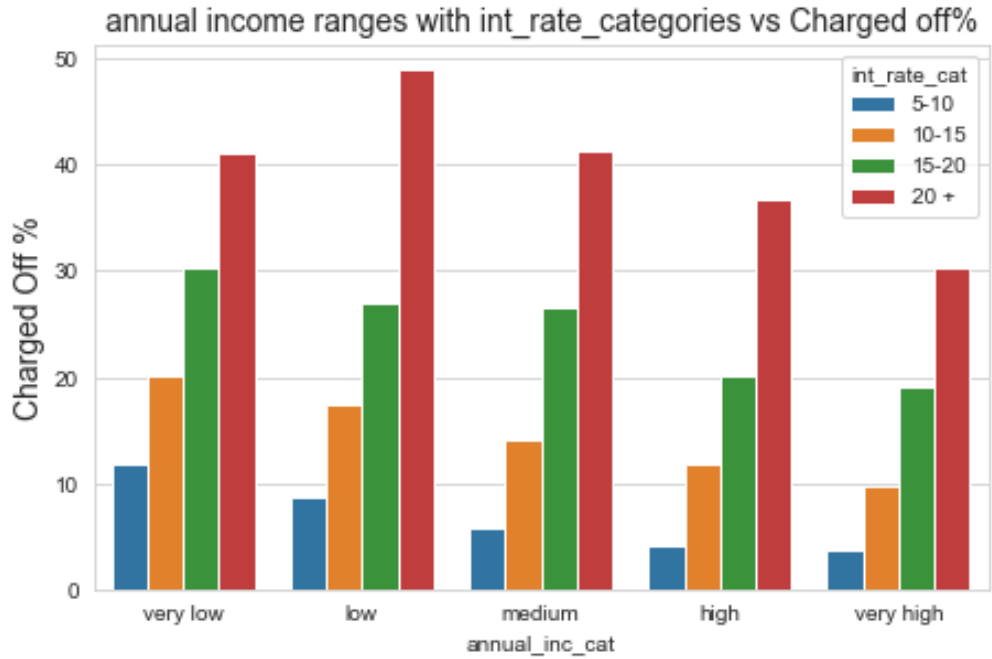
- Very low and low income groups are more prone to default than those with higher annual incomes.
- Interest rates are positively correlated to default rates, As the interest rates increases the rate of default also increases.
- Loan Amount is also a factor for default, higher the loan amount, higher is the default rate.
- DTI (Debt to Income) ratio is also positively correlated to Default rate. A high DTI has more propensity to default

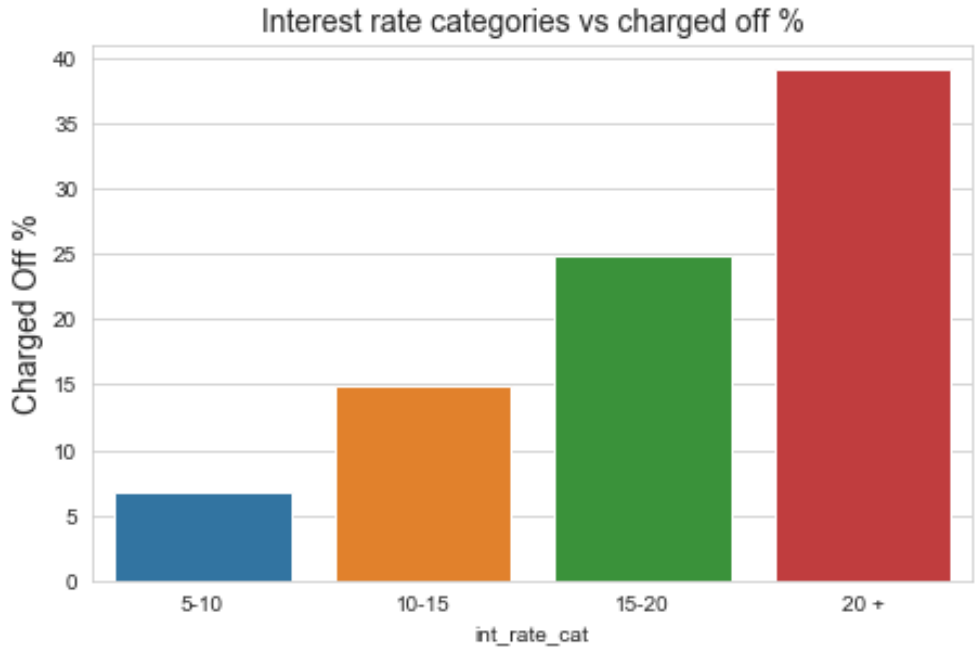
- For employees with different work experiences the rate of default ranges between 12% - 16%.
- Though the number of loans doubled every year, the default rates have been consistent at 12.5 % - 16%.
- Though Debt Consolidation is the largest purpose for borrowers, the default rate in this category is equal to average default rate of 15%.
- Small business category has 25+ % default rate, but the number of loans are very small to make an impact.
- As we move ahead from grades A to G, the rate of default increases from 5% to 35%.
- Repayment term of 60 months has a default rate of 25% whereas 36 months term has a default rate of 11%

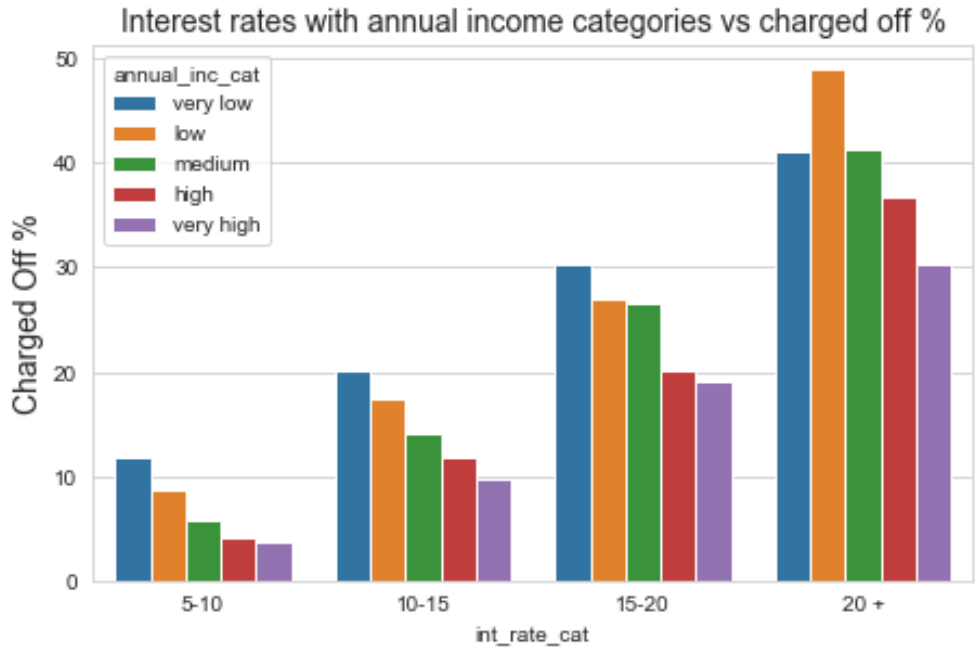

loan amount range vs charged off %
Charged Off %
0
5
10
15
20
25
0-10000
10000-20000
20000-30000
35000 +
loan_amnt_cat

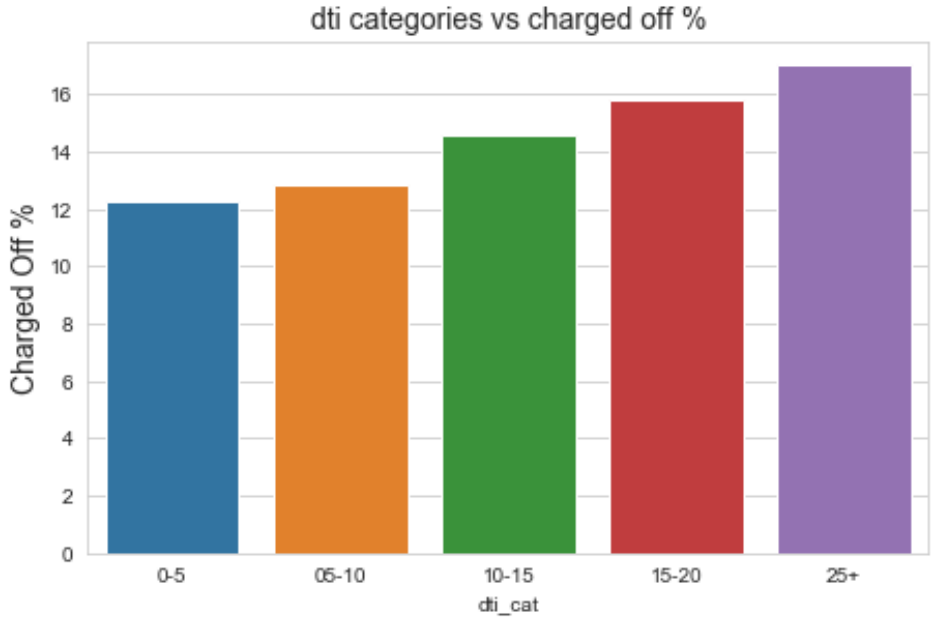

loan amount with int_rate_categories vs charged off %
int_rate_cat
5-10
10-15
15-20
20 +
Charged Off %
0
10
20
30
40
0-10000
10000-20000
20000-30000
35000 +
loan_amnt_cat

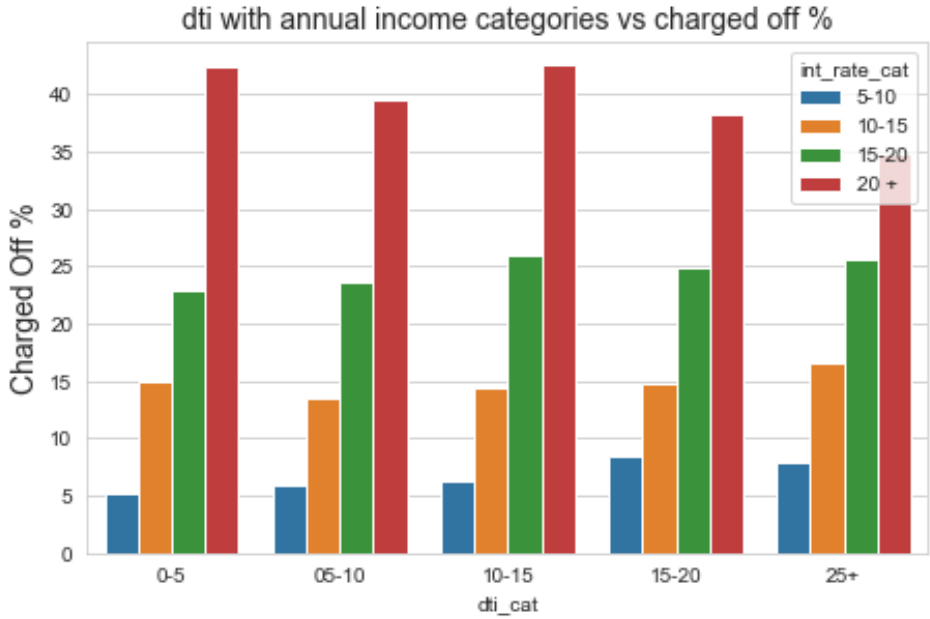

annual income ranges vs charged off %
Charged Off %
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
very low
low
medium
high
very high
annual_inc_cat


annual income ranges with int_rate_categories vs Charged off%
int_rate_cat
5-10
10-15
15-20
20 +
Charged Off %
0
10
20
30
40
50
very low
low
medium
high
very high
annual_inc_cat


Interest rate categories vs charged off %
Charged Off %
0
5
10
15
20
25
30
35
40
5-10
10-15
15-20
20 +
int_rate_cat

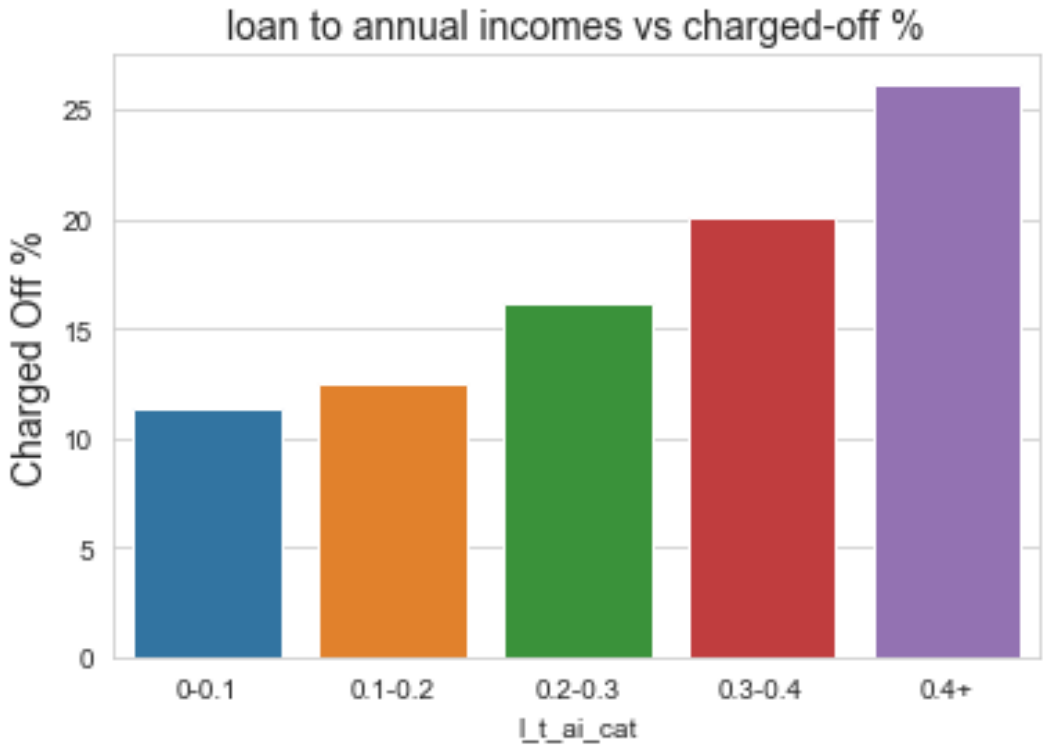

Interest rates with annual income categories vs charged off %
annual_inc_cat
very low
low
medium
high
very high
Charged Off %
0
10
20
30
40
50
5-10
10-15
15-20
20 +
int_rate_cat


dti categories vs charged off %
Charged Off %
0
2
4
6
8
10
12
14
16
0-5
05-10
10-15
15-20
25+
dti_cat


dti with annual income categories vs charged off %
int_rate_cat
5-10
10-15
15-20
20 +
Charged Off %
0
5
10
15
20
25
30
35
40
0-5
05-10
10-15
15-20
25+
dti_cat

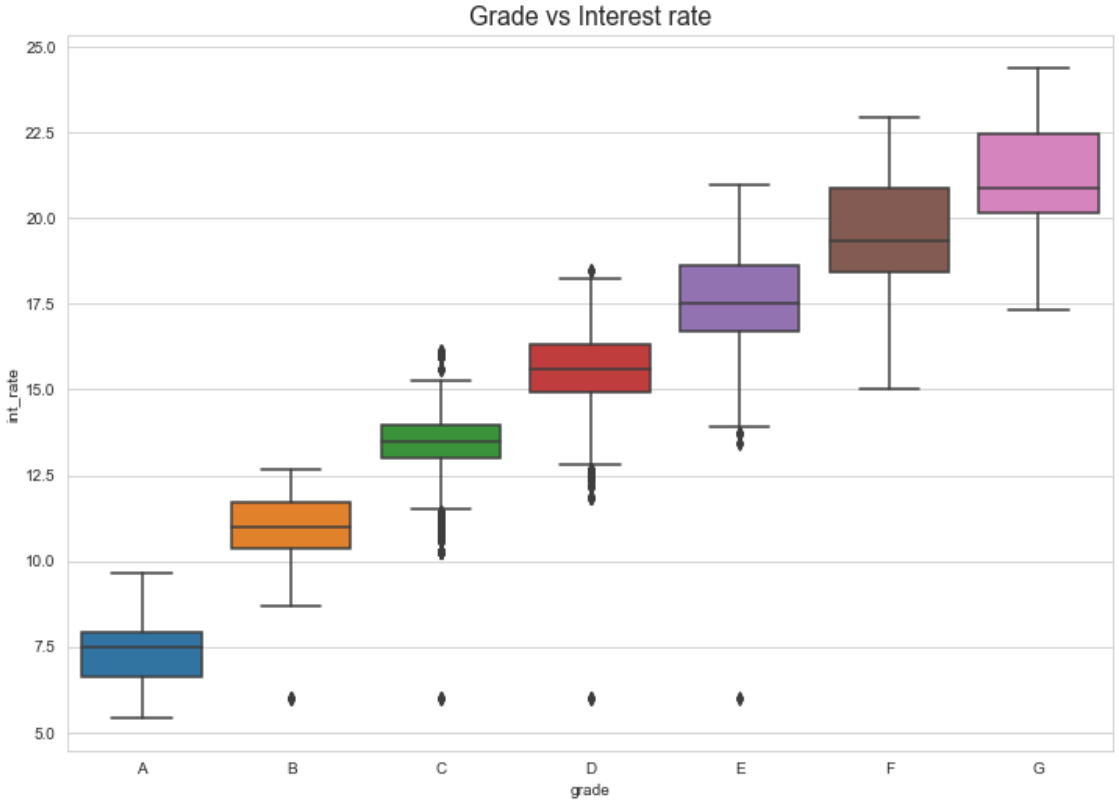
Derived Variable – Loan to annual income(LTAI)

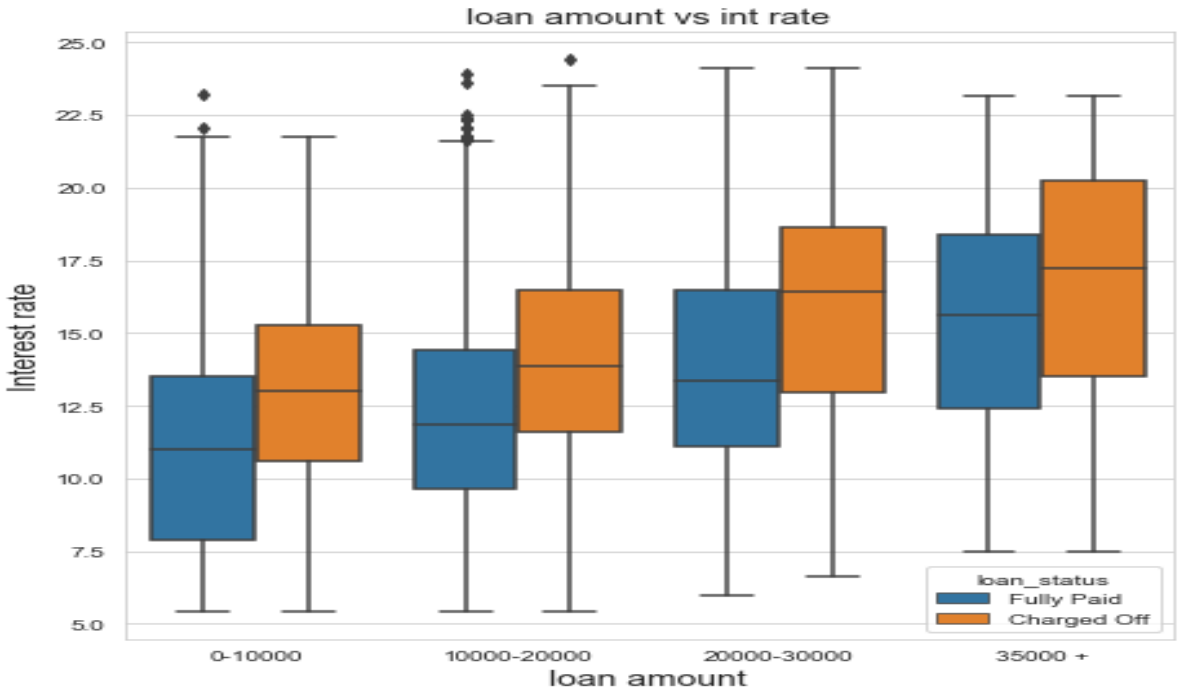
- The DTI variable alone is not enough for proper analysis so we created a new variable Loan to annual income ratio. This helps to determine the characteristics of a probable default.
- The default rate increases as the LTAI increases

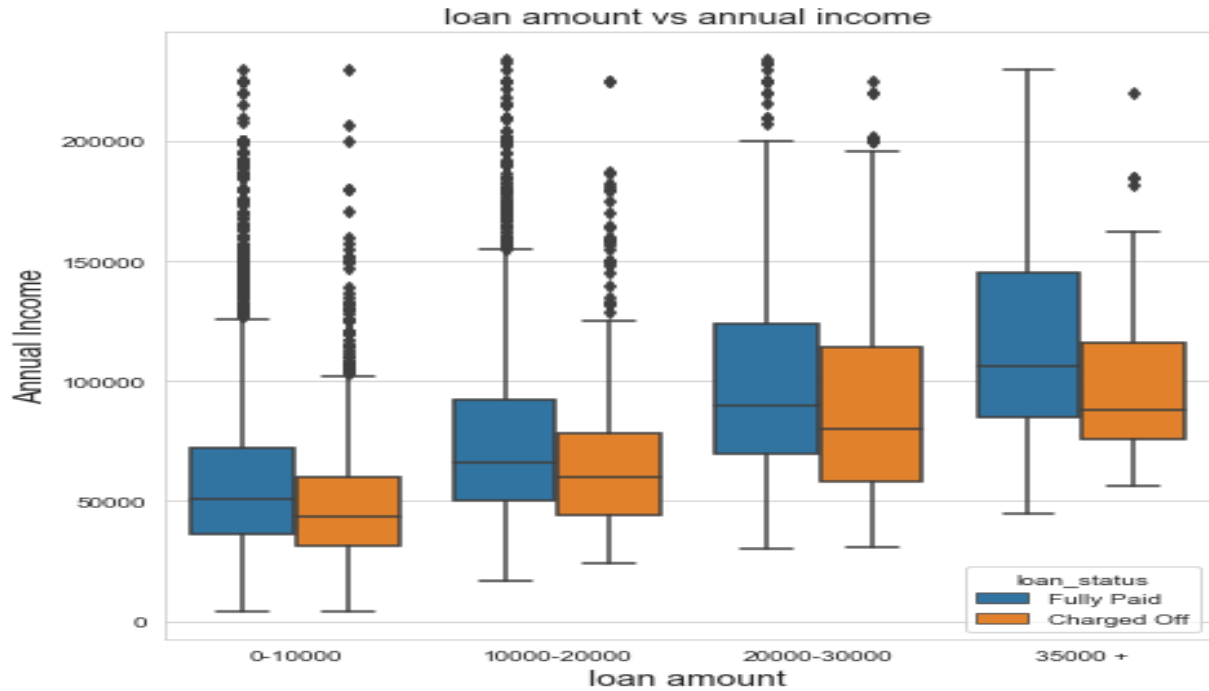


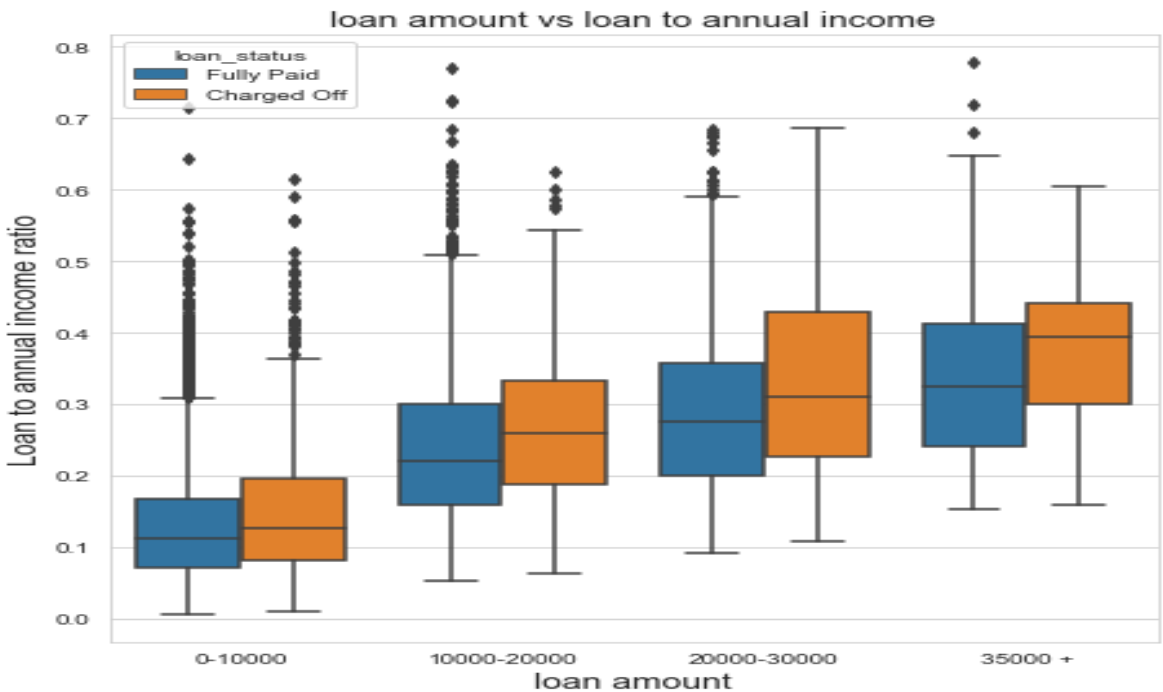
How interest rates are charged ?

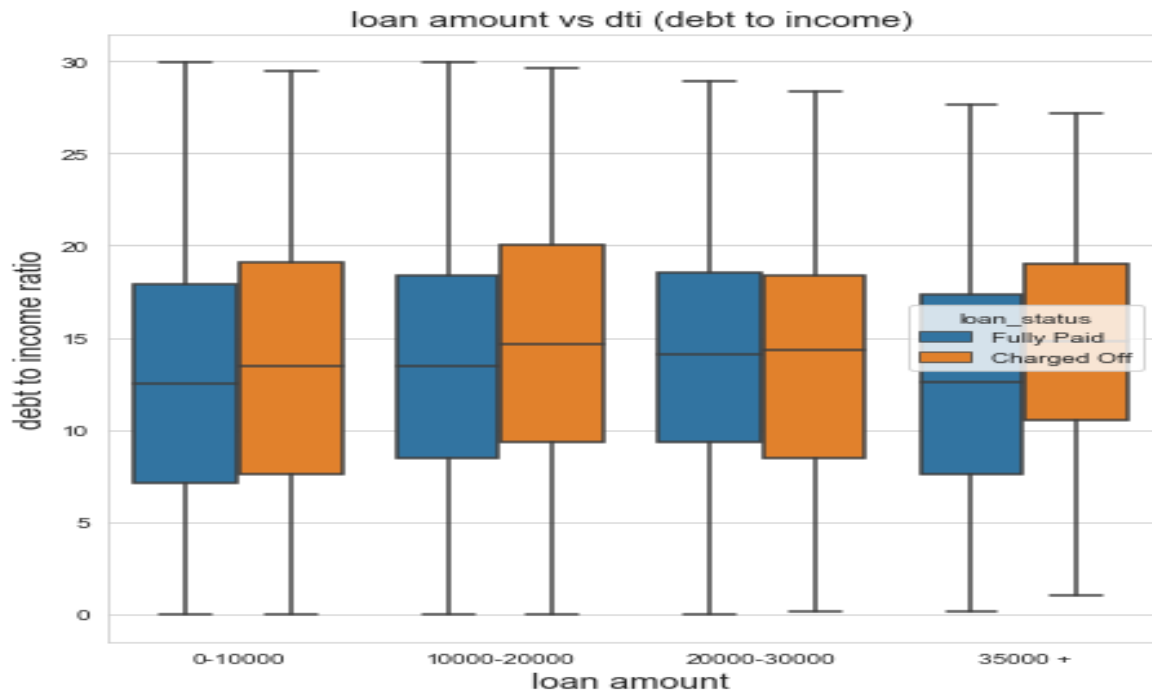
- The interest rate is decided by the grade of the loan.
- Loans with bad grades are charged more interest rates.
- From the plot we can see that for A grade the median interest rate is 7.5% and it goes upto 24% for G Grade
- In the next visualization we see the difference between fully paid and charged off loans across various key factors




loan amount vs int rate
Interest rate
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
loan_status
Fully Paid
Charged Off
0-10000
10000-20000
20000-30000
35000 +
loan amount


loan amount vs annual income
Annual Income
200000
150000
100000
50000
0
loan_status
Fully Paid
Charged Off
0-10000
10000-20000
20000-30000
35000 +
loan amount


loan amount vs loan to annual income
Loan to annual income ratio
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
loan_status
Fully Paid
Charged Off
0-10000
10000-20000
20000-30000
35000 +
loan amount


loan amount vs dti (debt to income)
debt to income ratio
30
25
20
15
10
5
0
loan_status
Fully Paid
Charged Off
0-10000
10000-20000
20000-30000
35000 +
loan amount

Observations:

- The table shows the difference between Charged off and Fully paid loans with respect to their annual Incomes.
- For a same annual income range, charged off loans have higher median loan amount, interest rate, DTI, LTAI than fully paid loans.
- Instalment and Annual income are lower for charged off than that off fully paid loans.
- This can be attributed as the reason behind defaults

annual_inc_cat	loan_status	loan_amnt	int_rate	installment	annual_inc	dti	l_t_ai
very low	Charged Off	4500	12.870	136.33	20800.0	11.920	0.253456
	Fully Paid	4400	11.140	137.99	21000.0	11.725	0.225000
low	Charged Off	8000	13.060	238.96	39996.0	14.580	0.211111
	Fully Paid	7500	11.490	225.58	40000.0	14.160	0.193182
medium	Charged Off	12000	13.990	339.06	62400.0	14.880	0.192308
	Fully Paid	10000	11.480	304.36	64000.0	13.400	0.153846
high	Charged Off	15000	14.110	456.12	90000.0	11.880	0.168675
	Fully Paid	12000	11.580	348.29	90000.0	12.550	0.130435
very high	Charged Off	20000	14.615	495.93	125000.0	12.980	0.135710
	Fully Paid	15000	11.860	417.34	125000.0	10.600	0.109234

- In general, defaults are caused by applicants taking higher loan amount with higher interest rates and small instalments than their peers in same income group.
- Relatively higher DTI and LTAI are driving factors for defaults.

l_t_ai_cat	loan_status	int_rate	annual_inc	dti	l_t_ai
0-0.1	Charged Off	12.730	62000.0	12.930	0.070000
	Fully Paid	10.805	72700.0	11.405	0.068355
0.1-0.2	Charged Off	13.430	56000.0	13.560	0.150000
	Fully Paid	11.460	61200.0	12.560	0.147222
0.2-0.3	Charged Off	14.220	54000.0	14.370	0.250000
	Fully Paid	12.420	53000.0	13.940	0.241523
0.3-0.4	Charged Off	14.650	48000.0	15.110	0.338983
	Fully Paid	12.260	48000.0	15.380	0.340940
0.4+	Charged Off	14.655	42500.0	15.670	0.447751
	Fully Paid	12.420	40000.0	14.740	0.441667

annual_inc_cat	loan_status	loan_amnt	int_rate	installment	annual_inc	dti	l_t_ai
very low	Charged Off	4500	12.870	136.33	20800.0	11.920	0.253456
	Fully Paid	4400	11.140	137.99	21000.0	11.725	0.225000
low	Charged Off	8000	13.060	238.96	39996.0	14.580	0.211111
	Fully Paid	7500	11.490	225.58	40000.0	14.160	0.193182
medium	Charged Off	12000	13.990	339.06	62400.0	14.880	0.192308
	Fully Paid	10000	11.480	304.36	64000.0	13.400	0.153846
high	Charged Off	15000	14.110	456.12	90000.0	11.880	0.168675
	Fully Paid	12000	11.580	348.29	90000.0	12.550	0.130435
very high	Charged Off	20000	14.615	495.93	125000.0	12.980	0.135710
	Fully Paid	15000	11.860	417.34	125000.0	10.600	0.109234

Recommendations:

- Exercise caution over loans given to very low and low income groups.
- Reduction of Exposure to poorer graded loans (Grades D,E,F,G).
- Compare the relative LTAI ratio values from the tables provided above to identify a probable default loan application.
- Instead of providing a loan with a term of 60 months with lower instalments, providing a lesser loan amount with a term of 36 months that matches their LTAI and instalment capacity may reduce default.
- It is better to provide a higher loan amount to high annual income applicant having a lower LTAI at lower interest rate than providing a lower loan amount to low income applicant having a higher LTAI at a high interest rate