Lending Club Case Study

Group Members:

Abhinav Kumar Dubey(Group Facilitator)

Gautam Kumar

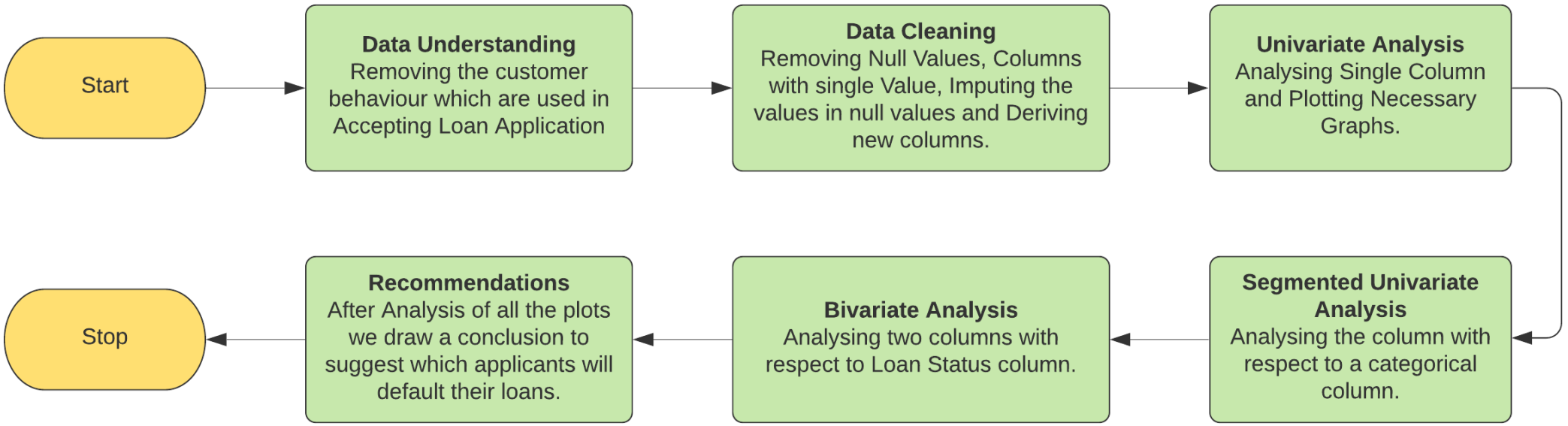
Problem Statement

- You work for a **consumer finance company** which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two **types of risks** are associated with the bank's decision:
  - If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
  - If the applicant is **not likely to repay the loan,** i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

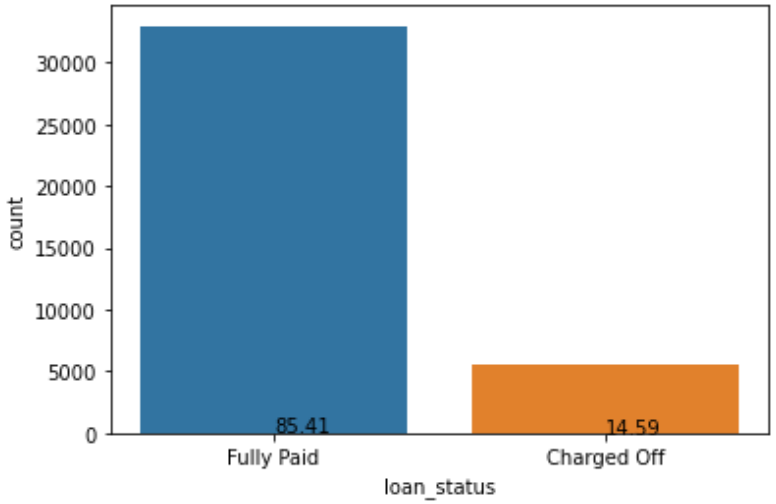
Business Objectives

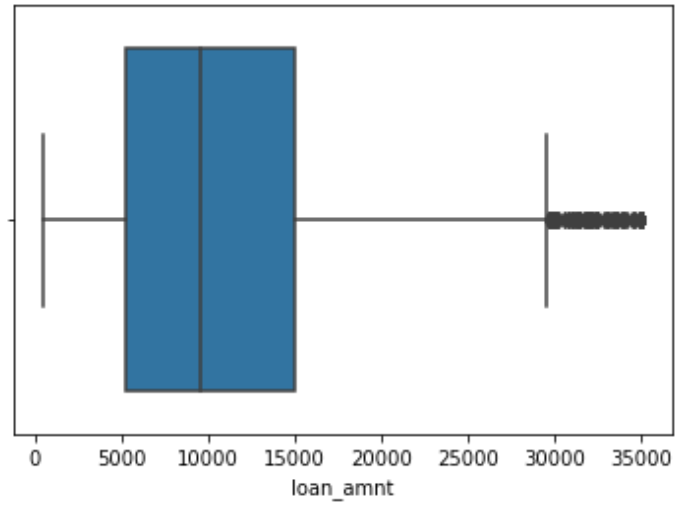
- lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who **default** cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters’.
- 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.

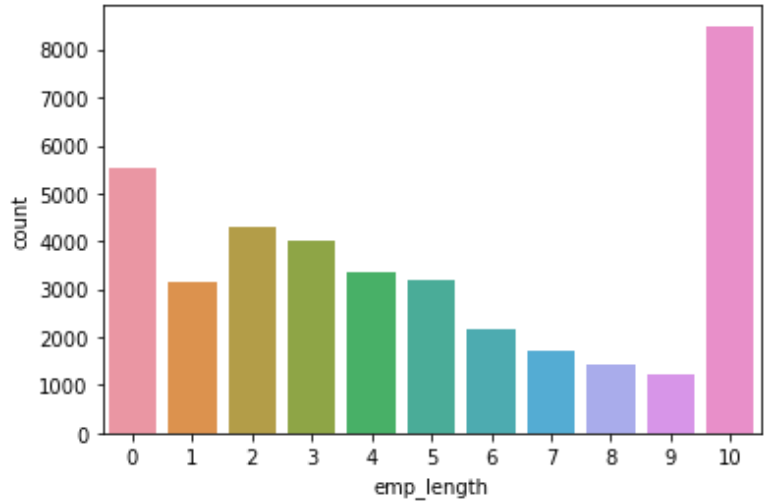
Methodology



Univariate Analysis

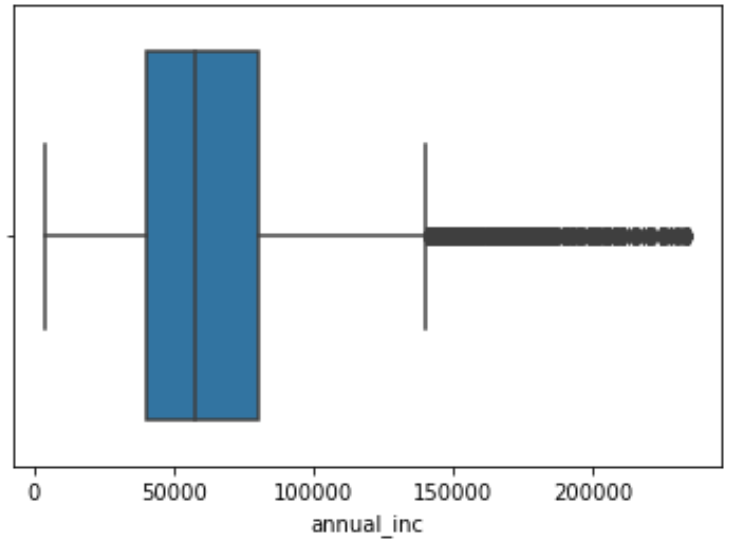


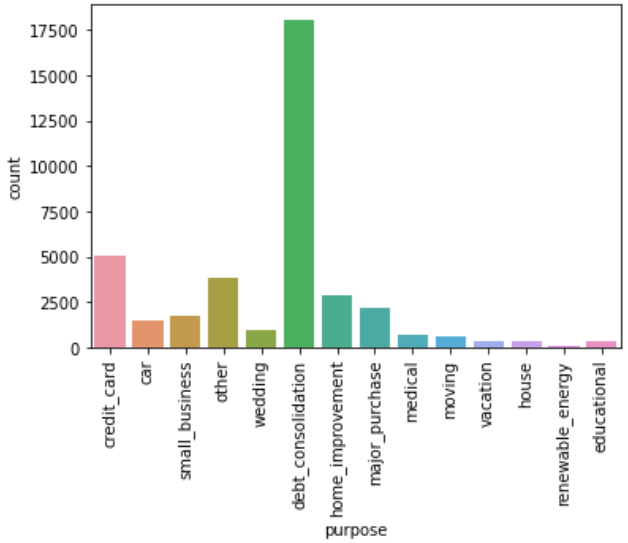


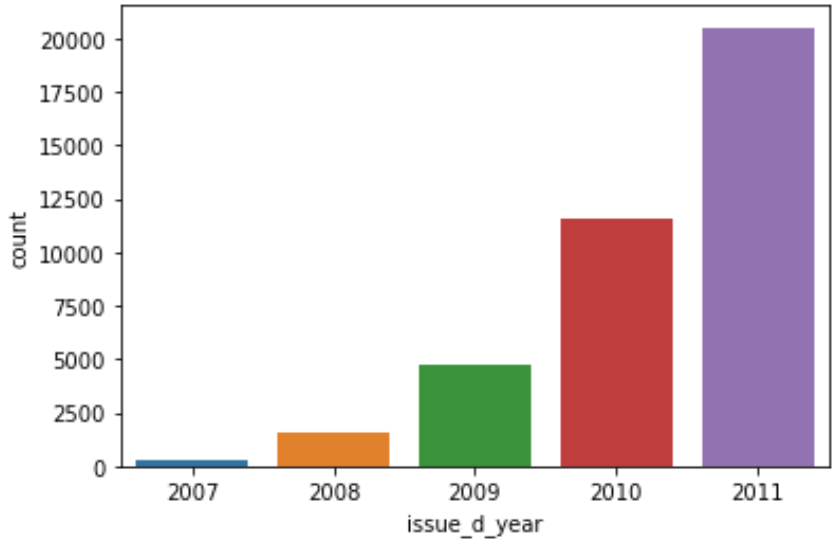


- **85.41%** of the Applicants paid off the full loan where as **14.59%** of applicants are defaulters.
- In the boxplot we can see that the most number of data lies between 5,000 and 15,000, and the maximum loan taken is 35,000 but after 30,000 mark, the loan borrowed above that mark is very low.
- Around 8000+ applicants have work experience of 10 years or more.

Univariate Analysis(Contd..)

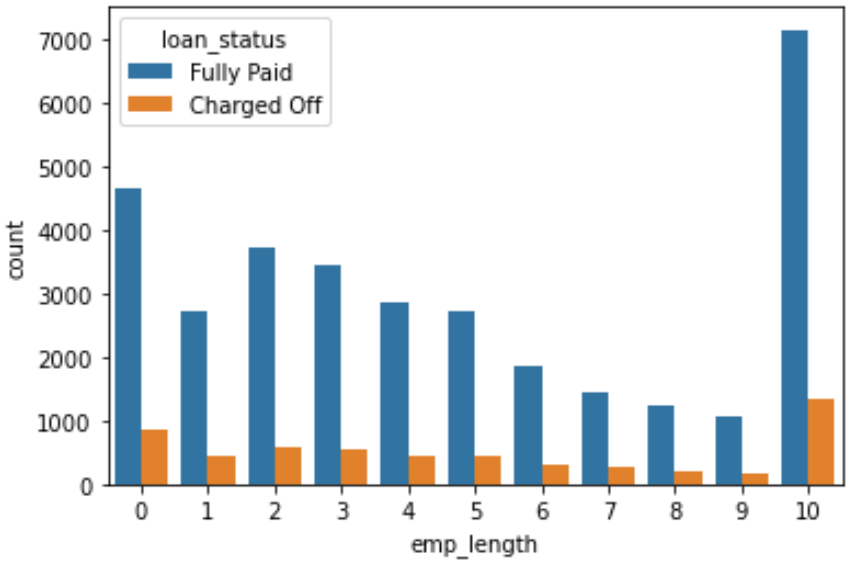


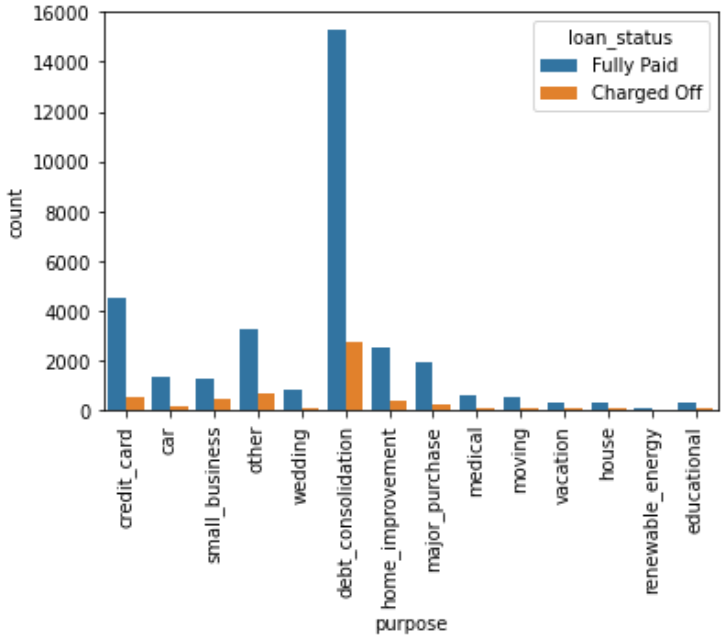


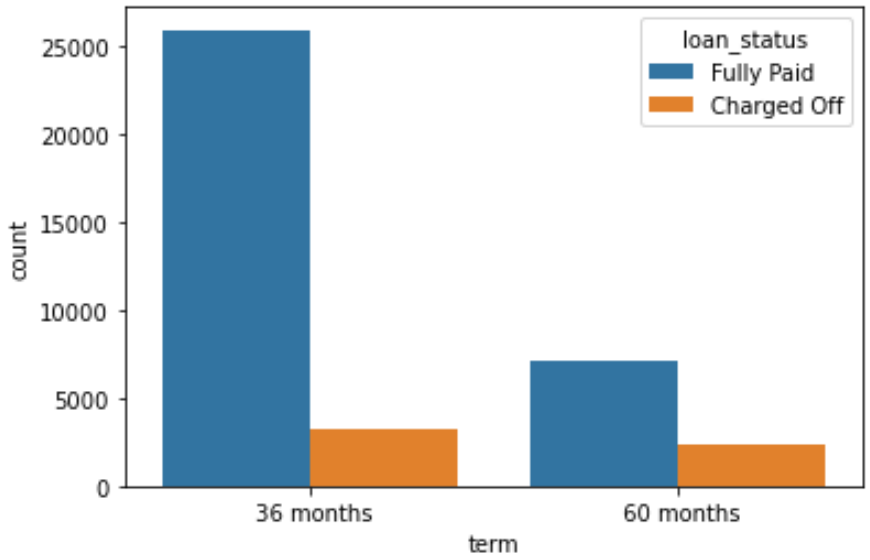


- After removal of outlier we can see that in annual income most data lies between 2,000$ and 9,000$.
- We can see that most people have applied loan for **Debt Consolidation** and very few have applied **Renewable Energy.**
- We can inference from the above plot that around 53% of loan was issue in the year of 2011.

Segmented Univariate Analysis

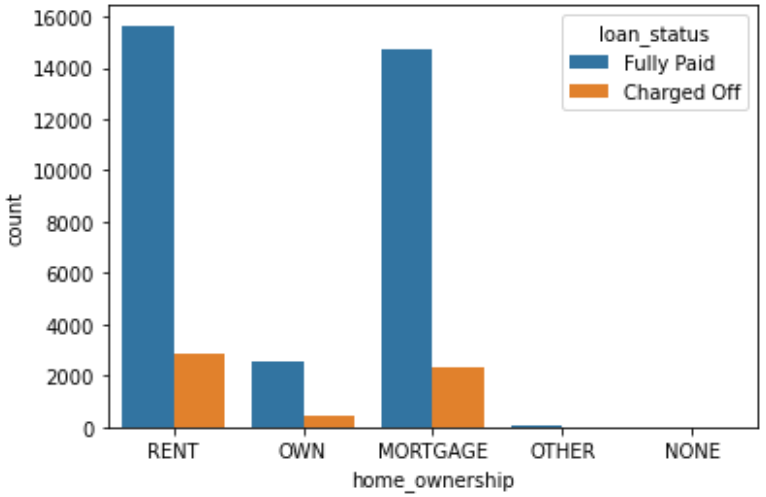


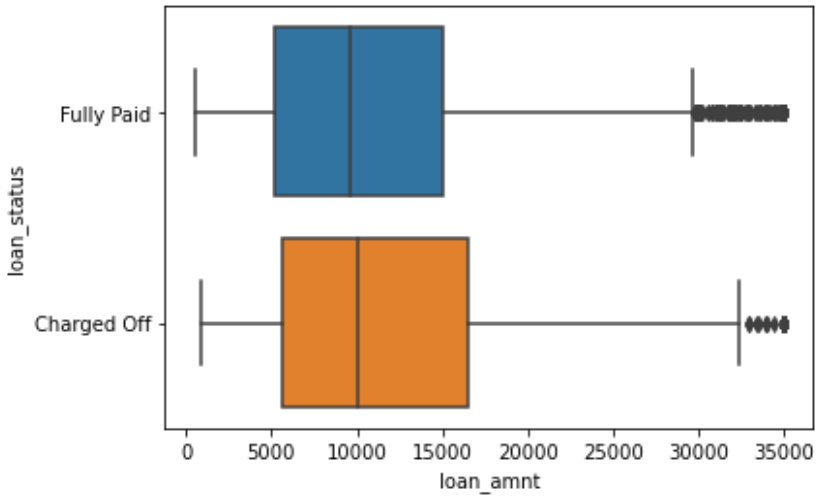


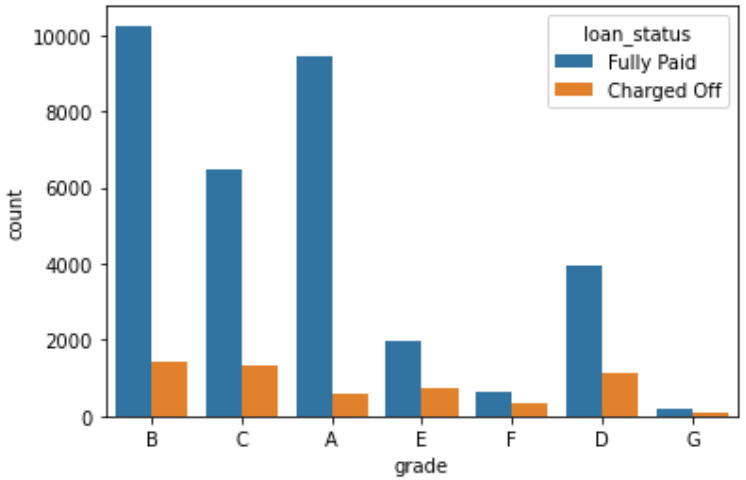


- From the above visualization we can confirm that most Employees who have work length of 10 years or more have returned the loans. But most of the applicants of work length of 10 years have also defaulted their loans.
- We can infer that applicants who took loan for debt consolidation are the ones who have almost paid all of their loans on the other hand, the highest number of defaulters are also from the same category.
- From the above plot we can draw an inference that the applicants who have taken loan with the term on 36 months are most likely to default their loans than the ones who have taken loan with term of 60 months.

Segmented Univariate Analysis(Contd..)

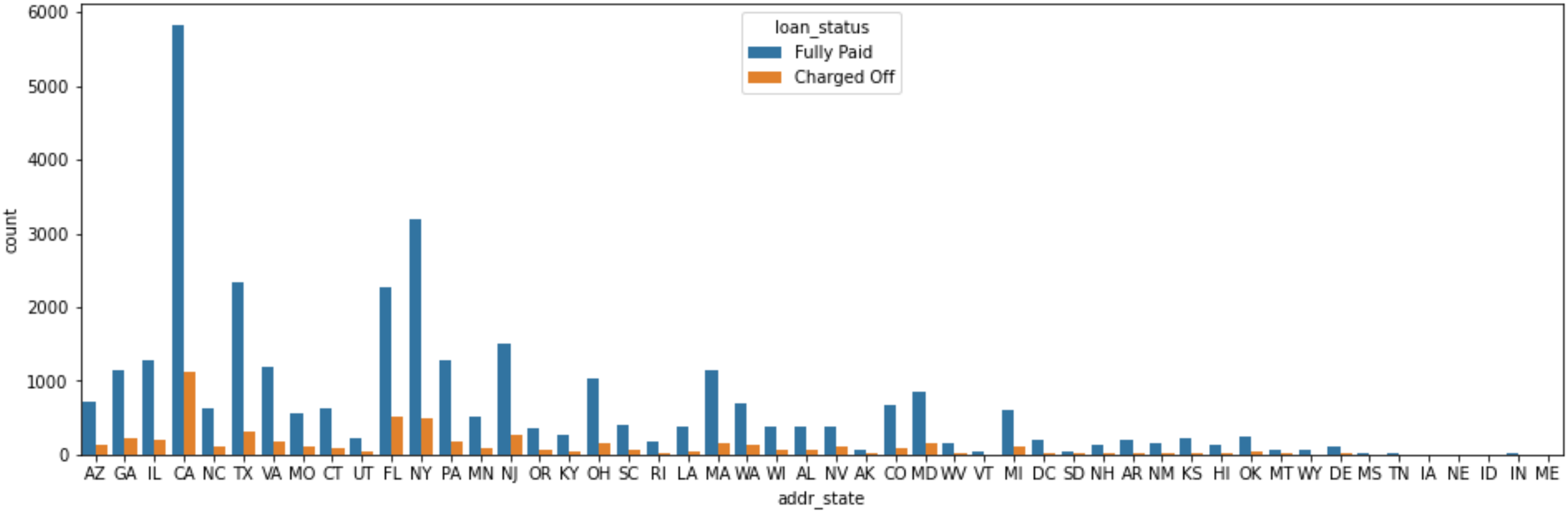






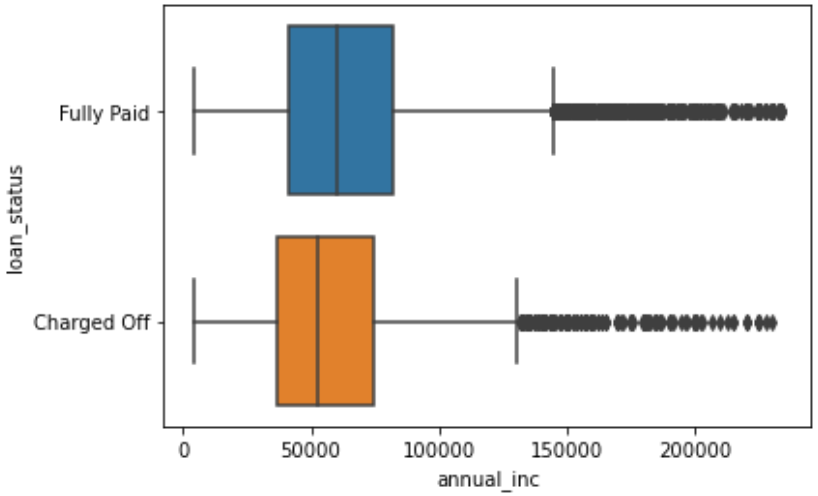
- People who live in rent and mortgage are more likely to default their loans then the other category in home ownership, but the Applicants living in rent are more likely to be defaulters than the Applicants of mortgage.
- The 25 percentile and median are closely related to each other. But as we can see the loan amounts which are getting defaulted are more than 15,000$. So basically that means that more the loan amount higher the chances that the loan will get defaulted.
- As we can see that the higher grades have more applicants who have paid loans but lower grades like E,F,G have comparatively less, so we can say that lower grades will default their loans more than higher grades.

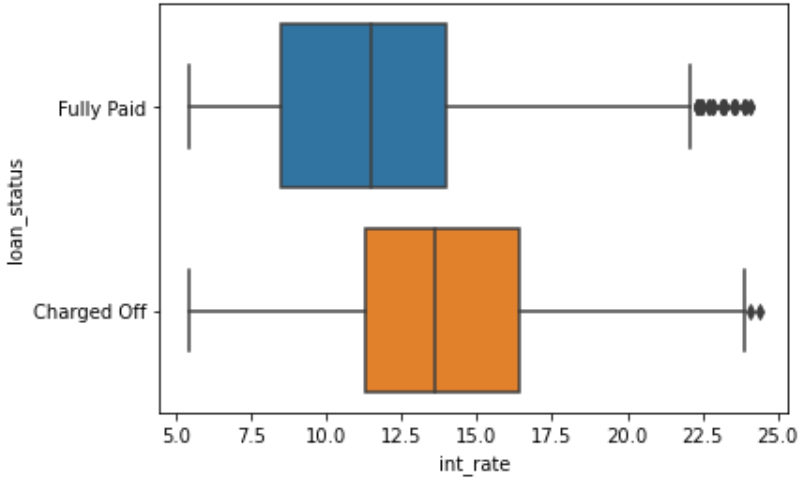
Segmented Univariate Analysis(Contd..)

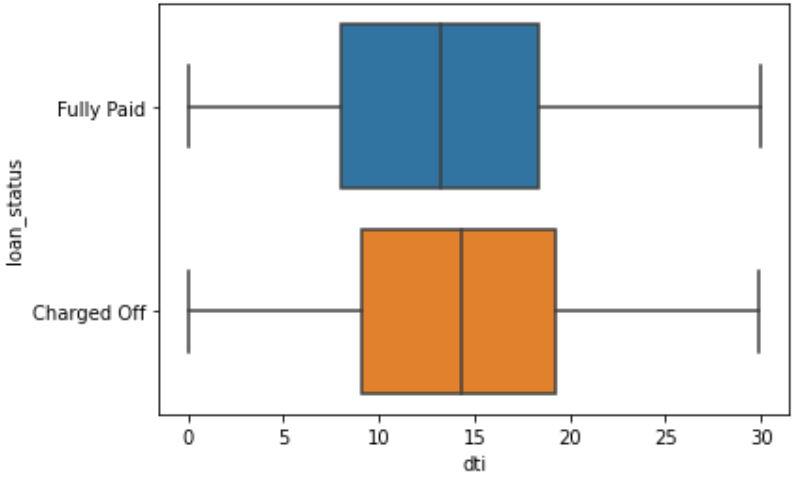


Applicants of CA, NY, FL and TX are more likely to default the loans.

Segmented Univariate Analysis(Contd..)

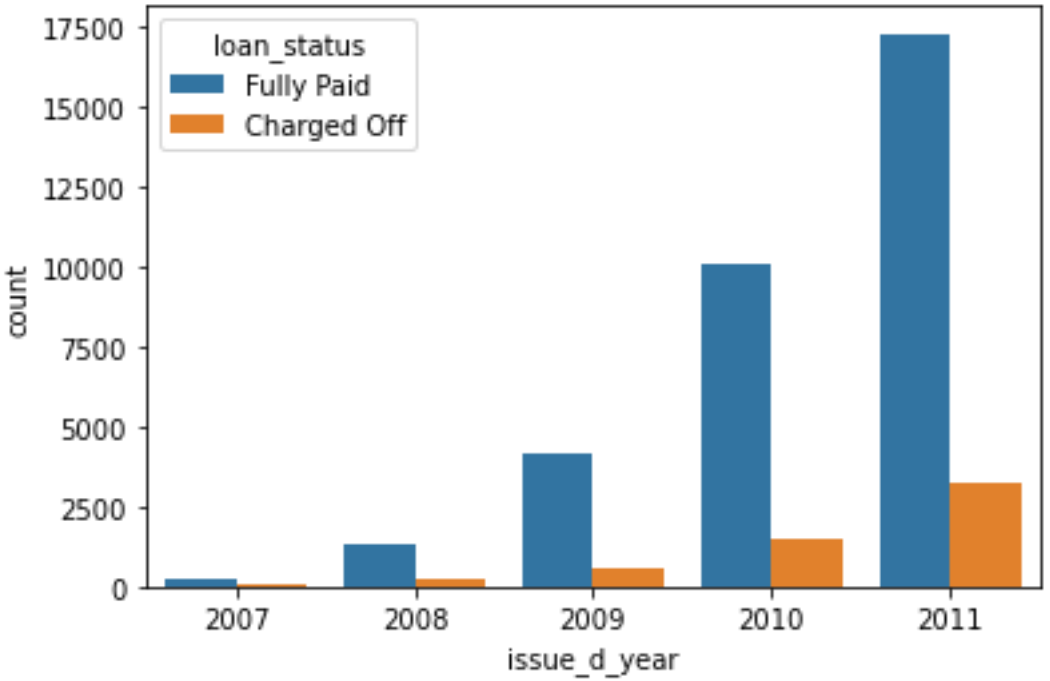


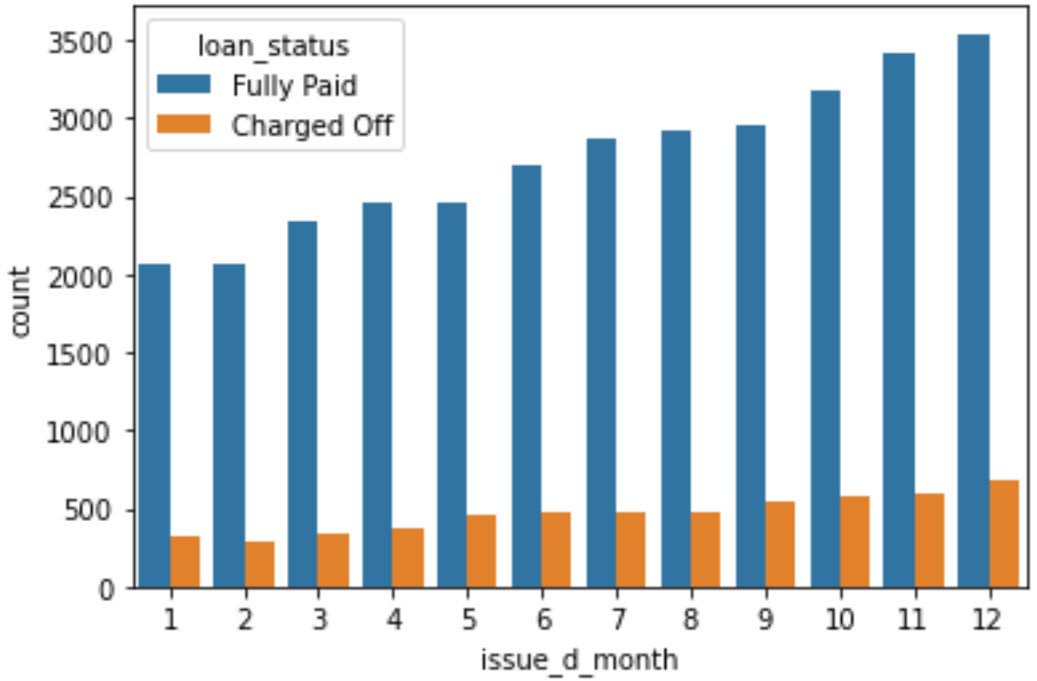




- After not considering the outlier, The 25th, median, 75th and the maximum varies a lot in the above plot. It clearly shows that the ones whose income is less have more chances of defaulting their loans than the ones who have more income.
- There is a high variation in data between the charged off interest rate and fully paid interest rate, which clearly indicates that more the interest rate more it is likely the applicants will be defaulters.
- From the above plot we can infer that the charged off applicants DTI variation is comparitively more than that of the fully paid.

Segmented Univariate Analysis(Contd..)

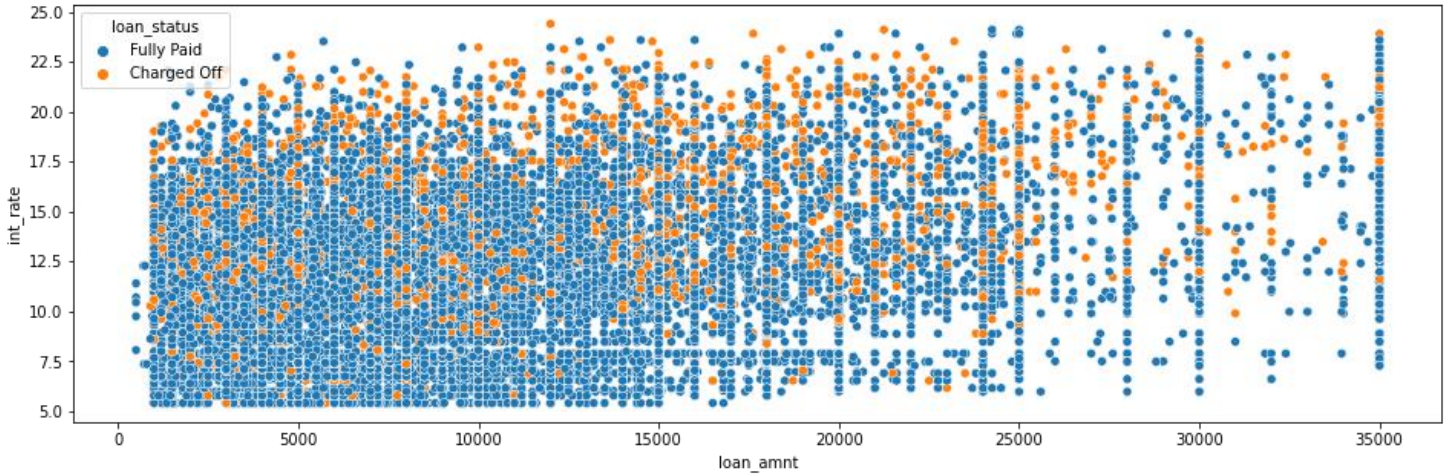




- The loan issued in 2011 and from the month of October - December are most likely to get defaulted.

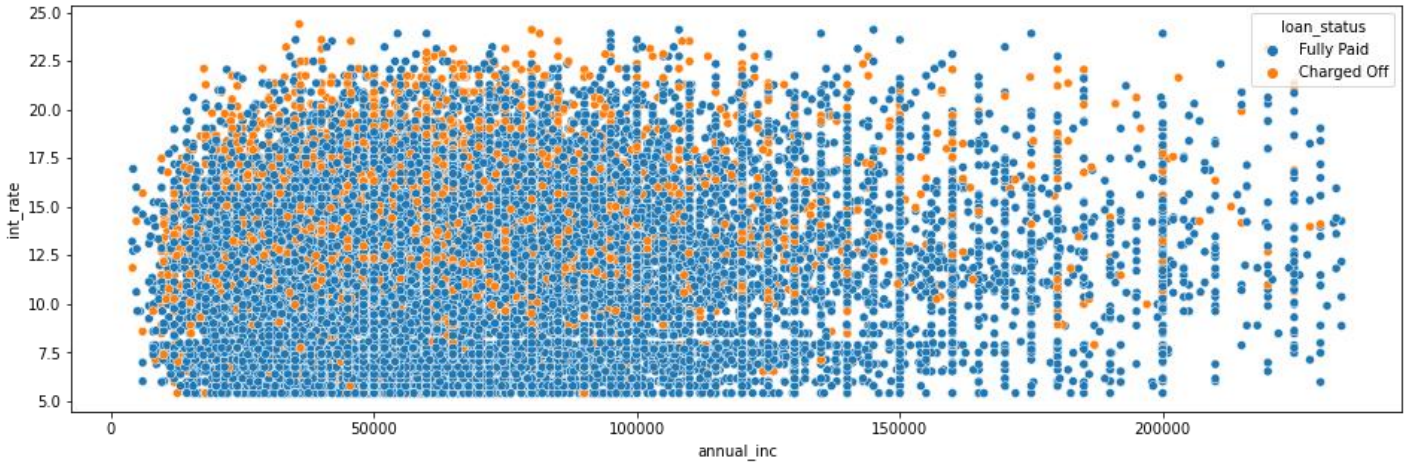
Bivariate Analysis

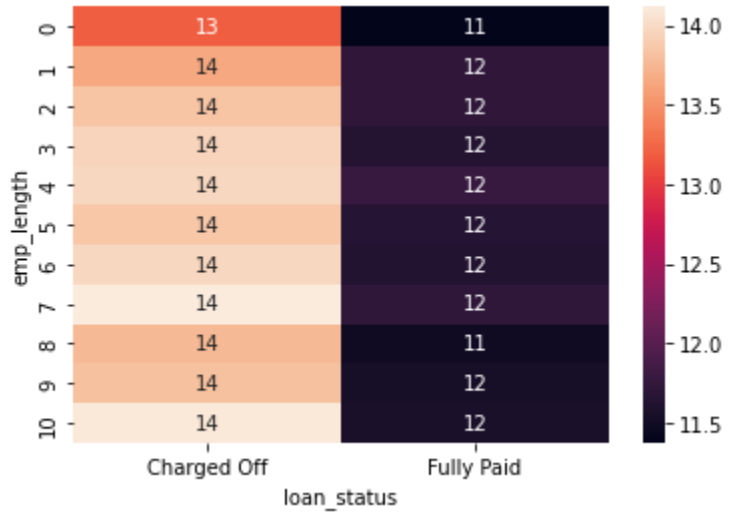
	loan_amnt	funded_amnt	funded_amnt_inv	int_rate	emp_length	annual_inc	dti
loan_amnt	1.000000	0.981790	0.937922	0.301265	0.156493	0.268999	0.062436
funded_amnt	0.981790	1.000000	0.956172	0.304930	0.156229	0.264798	0.062194
funded_amnt_inv	0.937922	0.956172	1.000000	0.297473	0.165112	0.251981	0.070663
int_rate	0.301265	0.304930	0.297473	1.000000	0.011633	0.048899	0.110913
emp_length	0.156493	0.156229	0.165112	0.011633	1.000000	0.120469	0.050561
annual_inc	0.268999	0.264798	0.251981	0.048899	0.120469	1.000000	-0.121530
dti	0.062436	0.062194	0.070663	0.110913	0.050561	-0.121530	1.000000

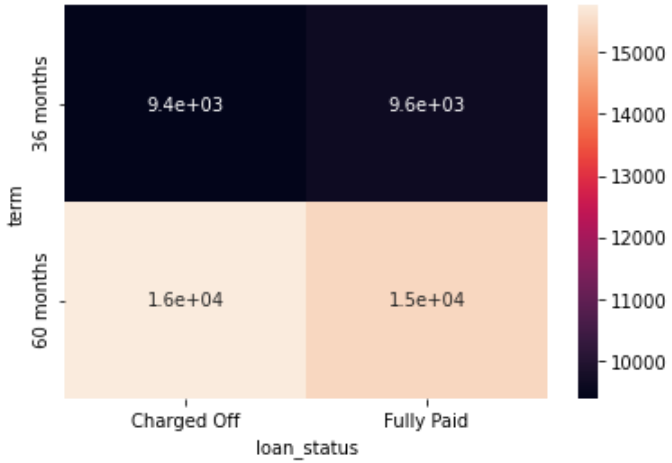


- From the correlation table we can conclude that loan_amnt, funded_amnt and funded_amnt_inv are highly correlated to each other since most of them have got the loans they asked for. And where as others columns are not correlated to each other in any way.
- As we can see that higher the loan_amnt, more the interest and hence most of loans are getting defaulted when there is high interest rate.

Bivariate Analysis(Contd..)







- From the figure we can inference that the loan has high chances to get defaulted if the applicant has their annual income less than 1500,00$.
- From the heatmap plot we can conclude that more the employee length, more the interest rate, hence more the chances of the loan to get defaulted.
- The amounts paid are highest in 36 months term and less in 60 months term. Hence we can conclude that more the terms, more likely that the loan is going to get defaulted.

Final Conclusions

- Loan amount of 15,000$ and more have high chances of getting defaulted.
- More the grade goes lower, less chances that the applicants will repay the loan.
- People will annual income less than 75,000 have more chances of defaulting the loan.
- The loans of term 60 months have more chances of getting defaulted.
- The states CA, DL, NY and TX have more defaulters, so the applicants of these states will default their loans more.
- Lending club should consider to check the DTI as more the DTI higher the chances that the loan will default.
- If the loan amount is high then the interest rate is also high that would make the applicants default their loans.
- The loan issued in 2011 and in the month of December are most likely to get defaulted.
- More the employee service length, more the interest rate, hence more the chances of the loan to get defaulted.