

Lending Club Case Study

Data Summary

1. Given is complete loan data for all loans issued through the time period 2007 to 2011.
2. It has a total of 39,717 rows and 111.

Data Cleaning

1. There are many columns having only null value which can be dropped.
2. Dropped other columns which don't contain meaningful information.

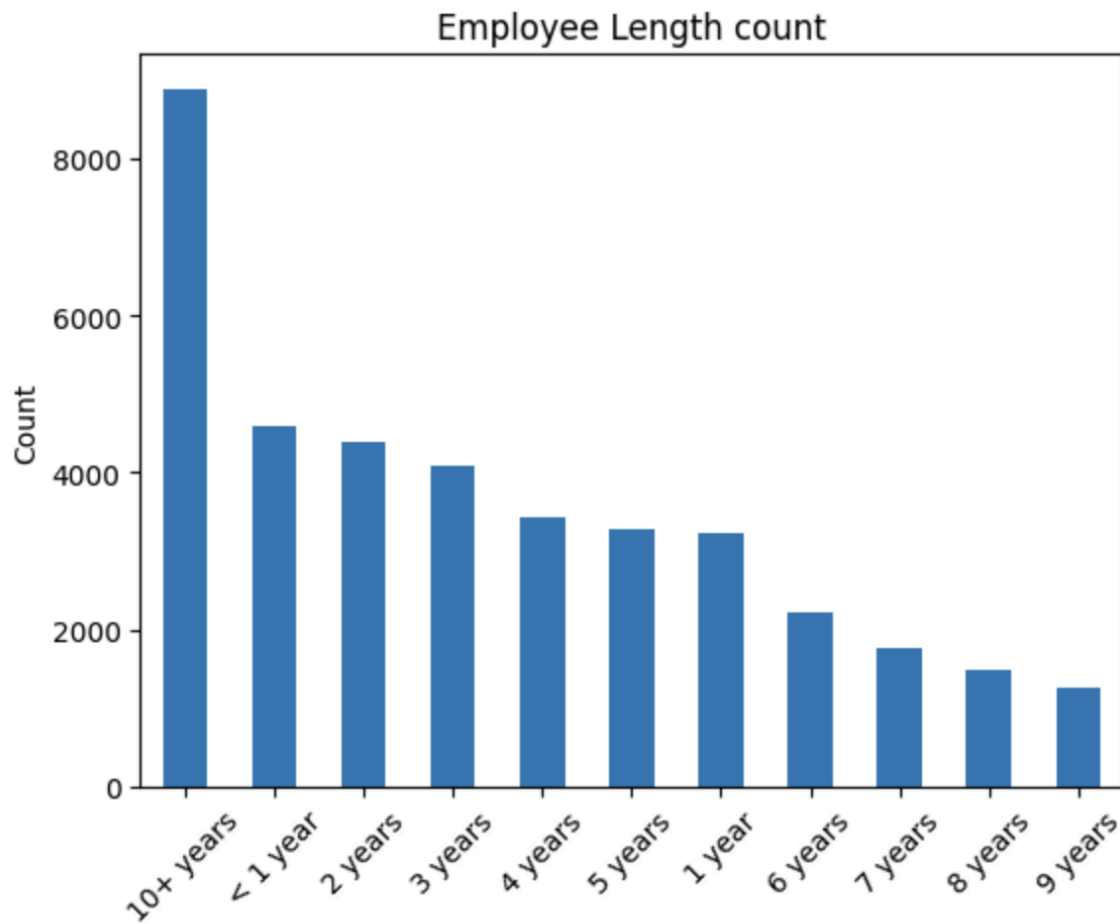
next_pymnt_d, mths_since_last_record, mths_since_last_delinq, desc, collections_12_mths_ex_med, delinq_amnt, chargeoff_within_12_mths, acc_now_delinq, pymnt_plan

3. Removed % from int_rate & revol_util columns
4. Updated the data type of various columns as int, float or string.

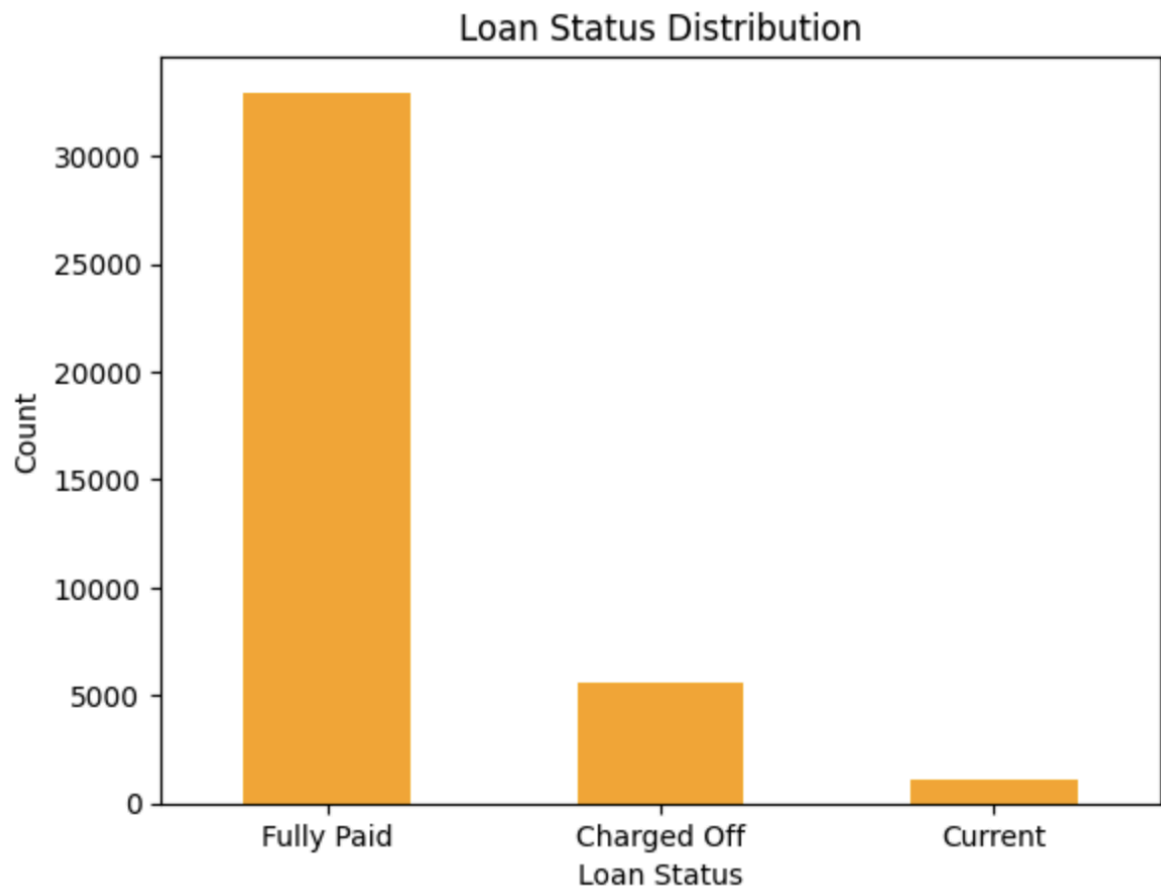
Derived Columns

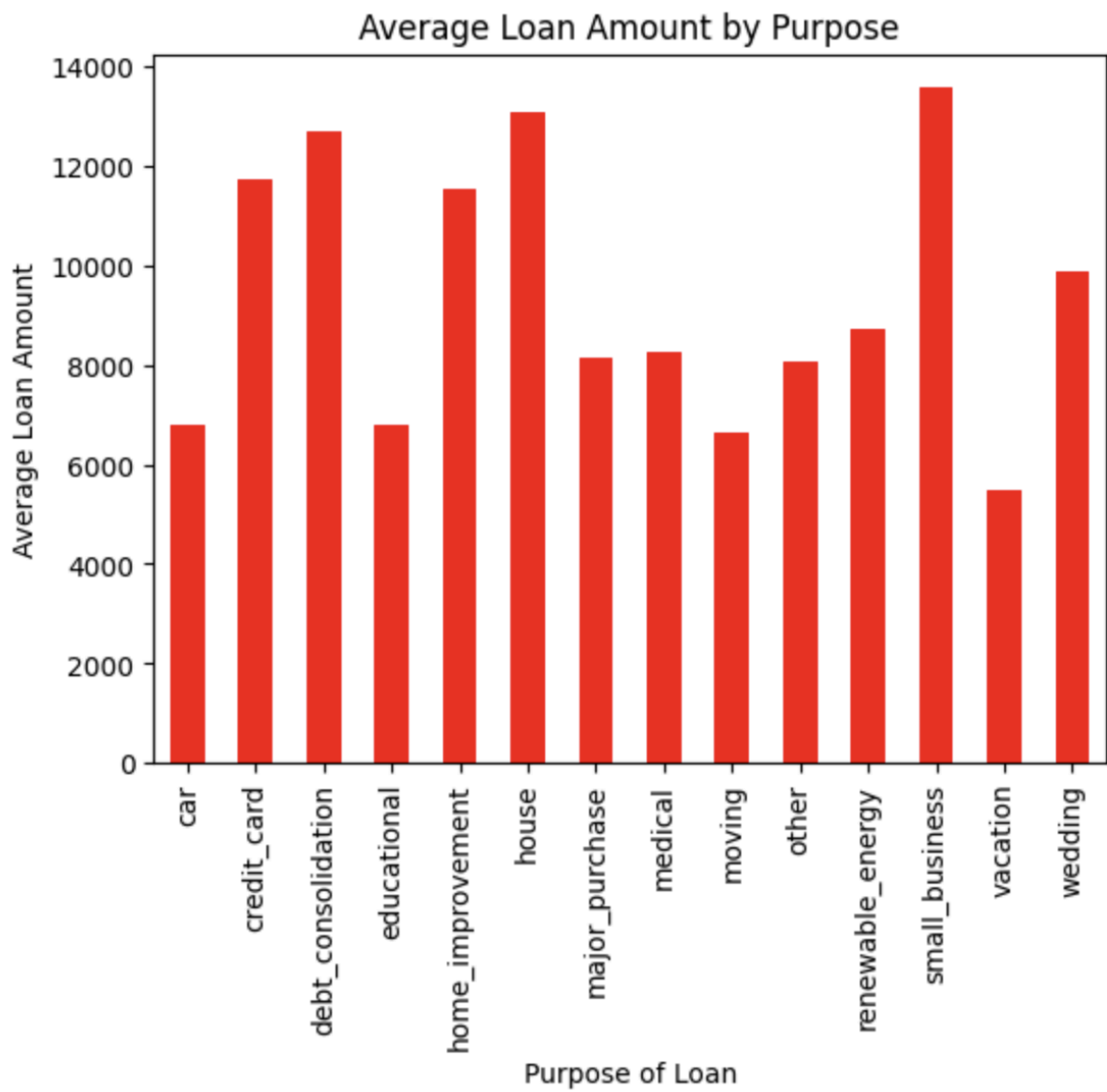
1. new column - **loan_status_deriv** - for loan status with numeric value
2. New column - **paying_capacity** - which is loan_amnt / annual_inc

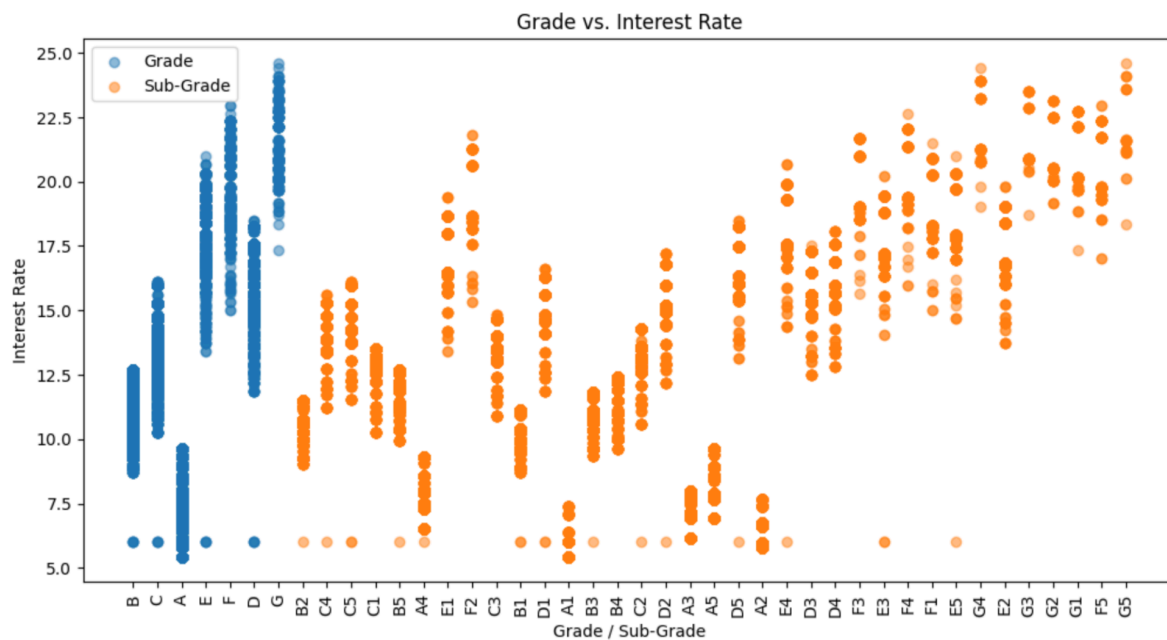
Univariate and Segmented Analysis



Employee Length





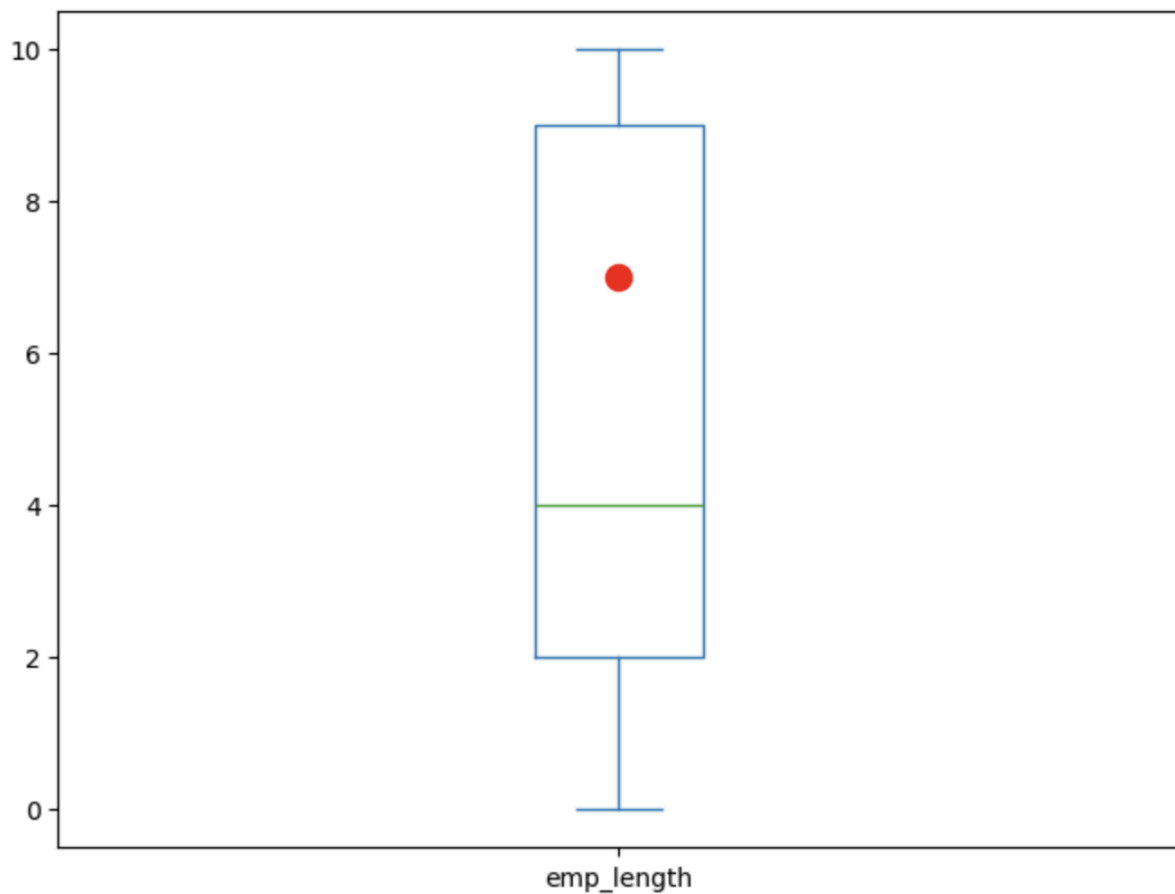


Driving Factors

Let's take a loan candidate. Let's say we are evaluating the application of a candidate with id 1069102.

DrivingFactor-1: Employment Length

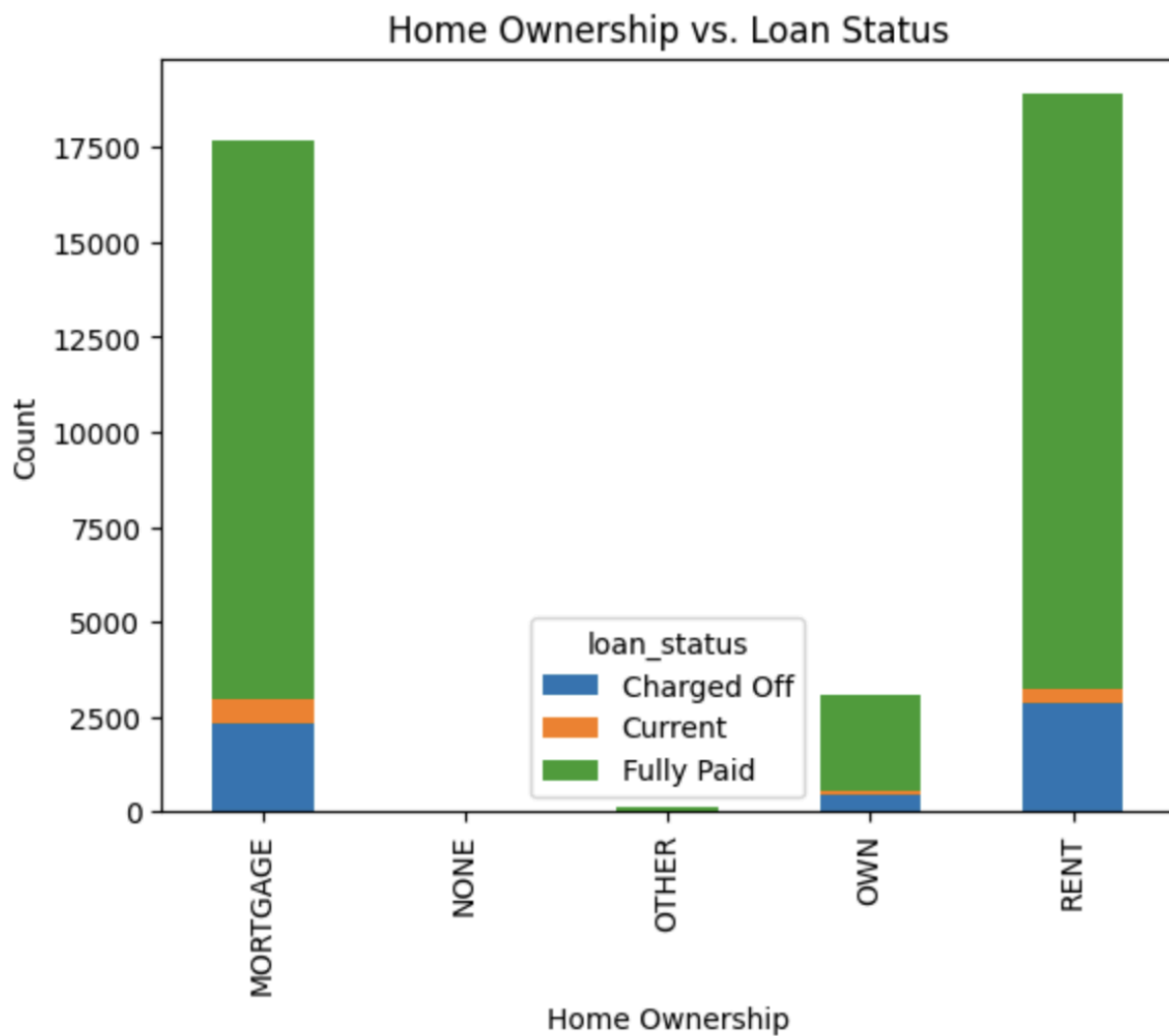
Below box plot shows that, amongst all the applicants, the given loan applicant has good employment length. Therefore, he is likely to have a stable income source and has lower credit risk.



DrivingFactor-2: Home Ownership

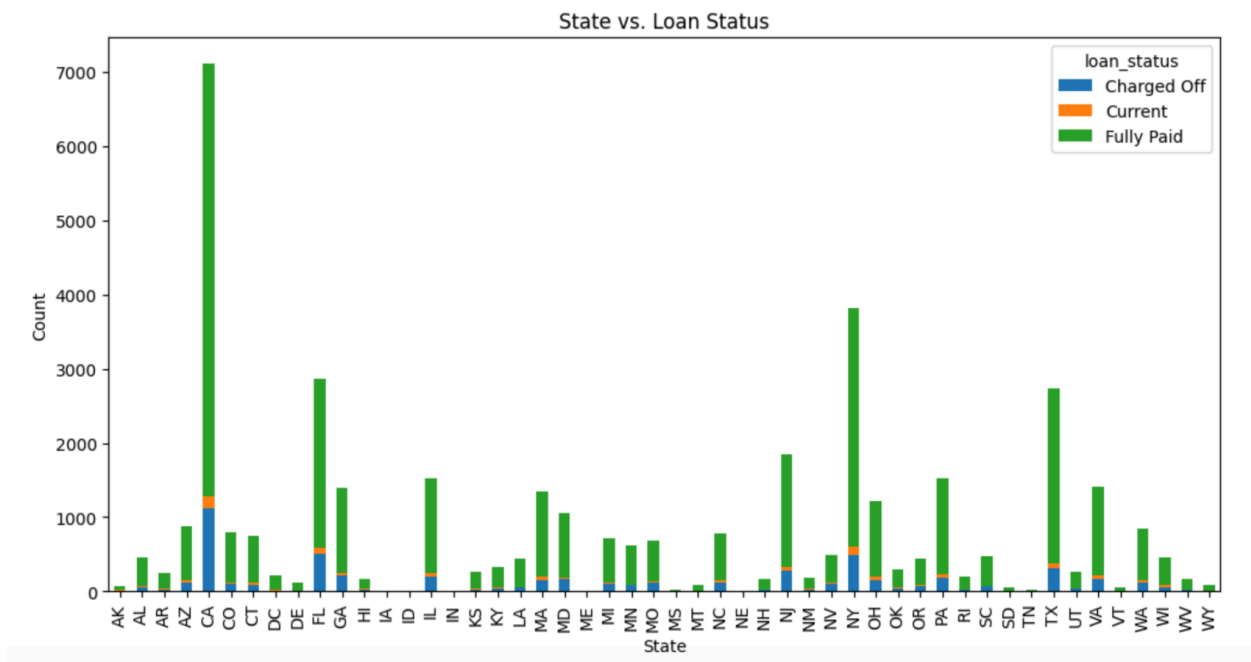
Below plot shows home ownership vs loan status. We can see that applicants having own home are less likely to get charged off

Given candidate's home ownership is MORTGAGE, hence he is doesn't have low credit risk from this parameter.



DrivingFactor-3: State Wise Analysis

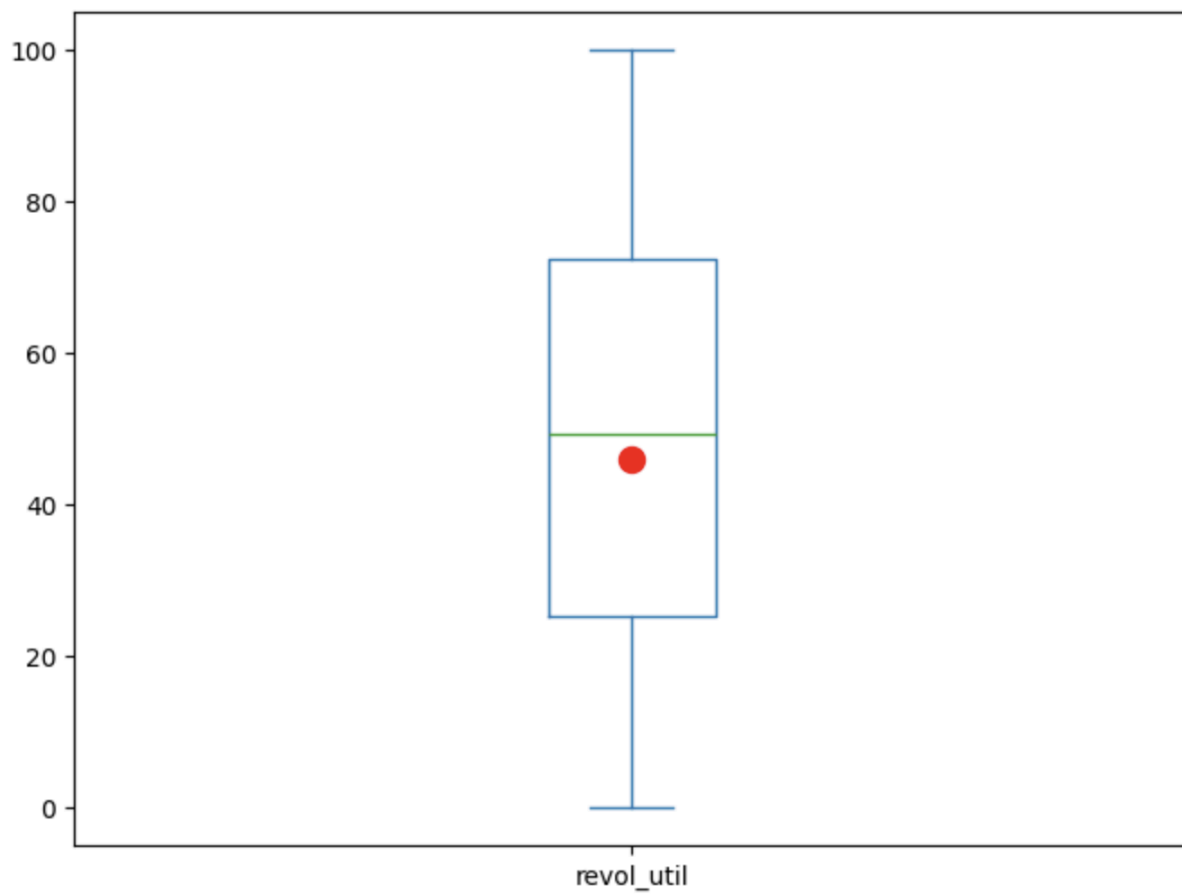
Given candidate is from state OH



DrivingFactor-4: Revolving Line Utilization

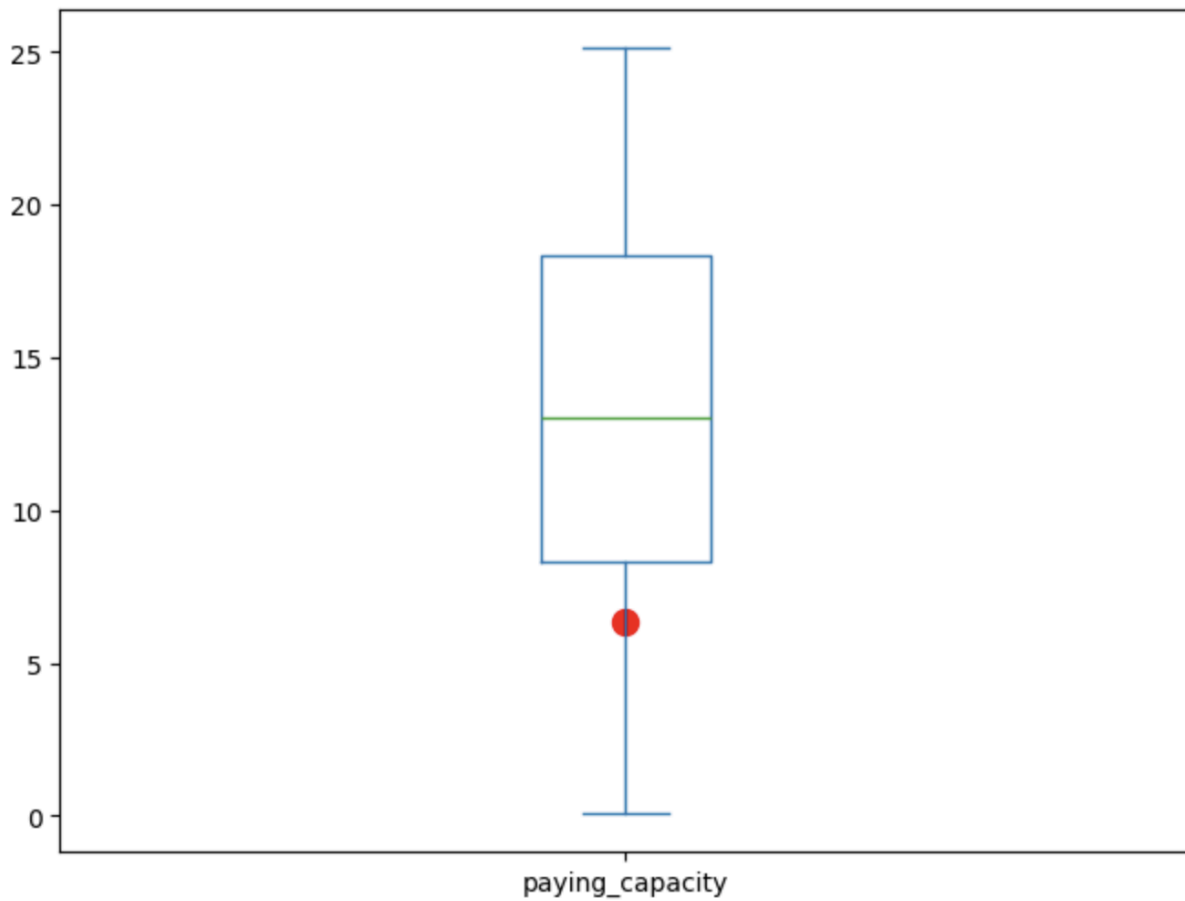
Lower revolving line utilization indicates lower credit risk of the candidate.

As seen in the below plot, candidate has lower credit risk then median wrt revolving line utilization



DrivingFactor-5: Derived Column: Paying Capacity

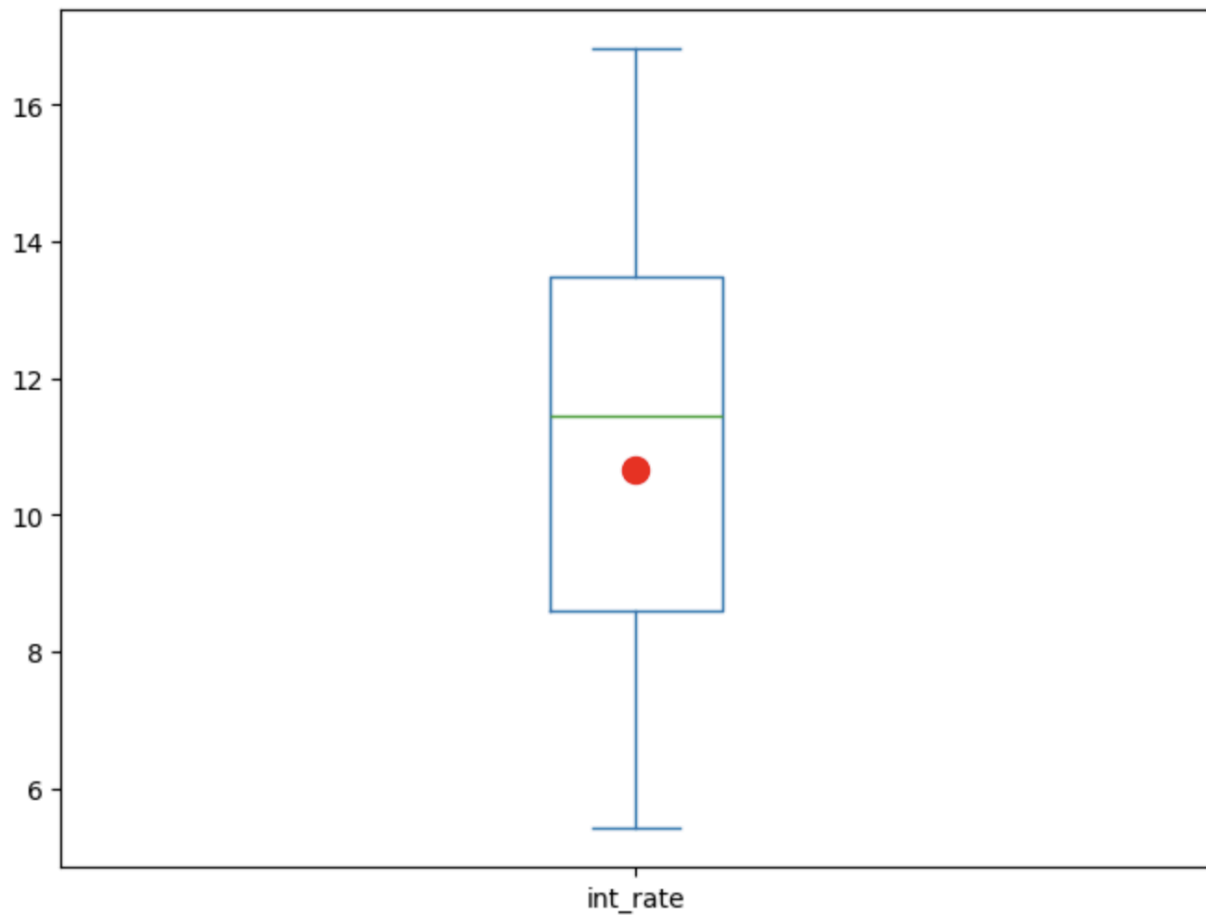
As shown in the below plot, candidate has good paying capacity for the loan given his/her annual income



DrivingFactor-6: Interest Rate

Candidates already having loans with higher interest rates have higher credit default risk.

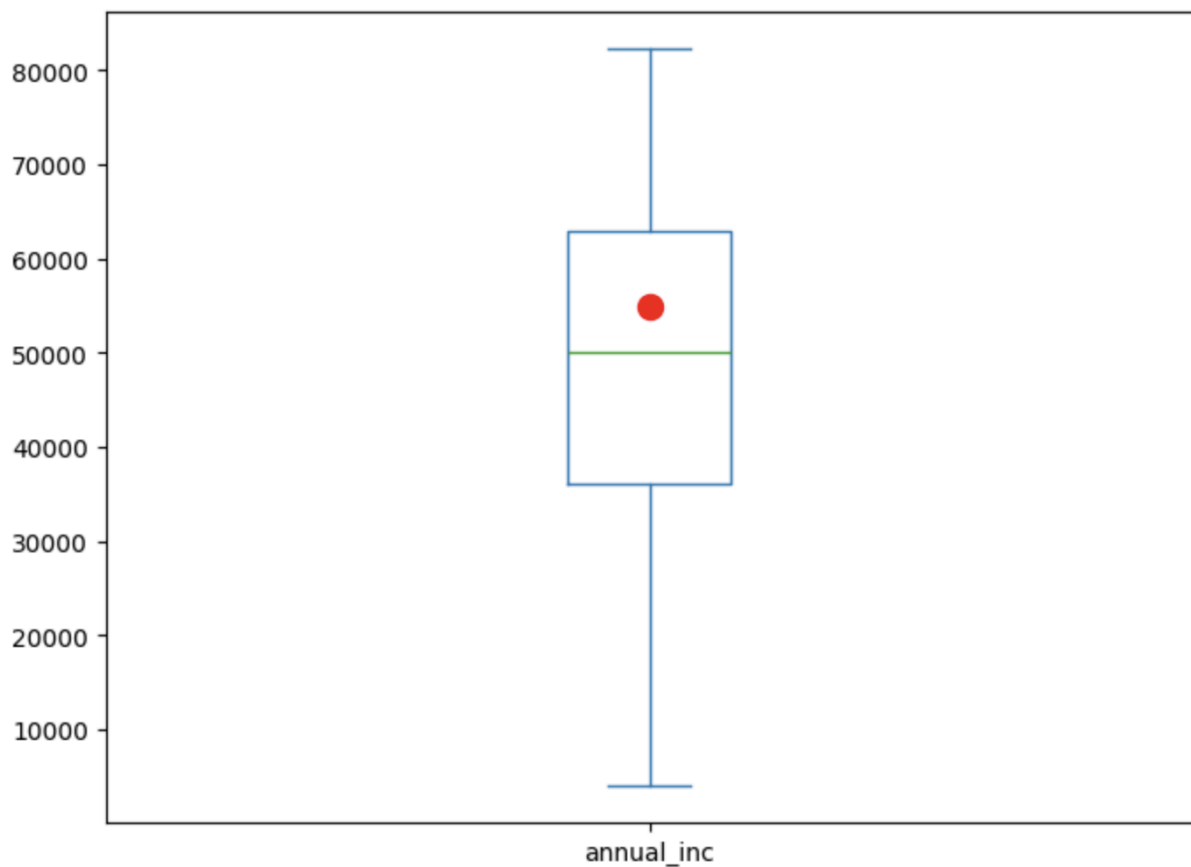
As shown in below plot, candidate has ongoing loans with interest rates lower than 50% compared to others.



DrivingFactor-7: Annual Income

Candidates having higher annual incomes have lower credit default risk. Also, if income source is verified, then it is a stronger indicator of lower credit risk of the candidate.

As shown in below plot, candidate has higher income than 50% of other candidates, which indicates lower credit risk.



Bivariate Analysis

Candidates having a higher public record of bankruptcy and higher public derogatory records have higher credit default risk.

There is a negative correlation of loan write-off for public bankruptcy records -0.04477231147893261

There is a negative correlation of loan write-off for public derogatory records -0.04821184032185699

Given candidate has total public bankruptcy records: 0

Given candidate has total public derogatory records: 0

Conclusion

From the given driving factors, a given loan applicant doesn't have high credit risk.