Lending Club Case Study

Naga Kartheek Kothapalli

Problem Statement

– Lending Club is the largest online loan marketplace.

– Facilitates personal loans, business loans, and financing of medical procedures.

– The data contains 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, through exploratory data analysis.
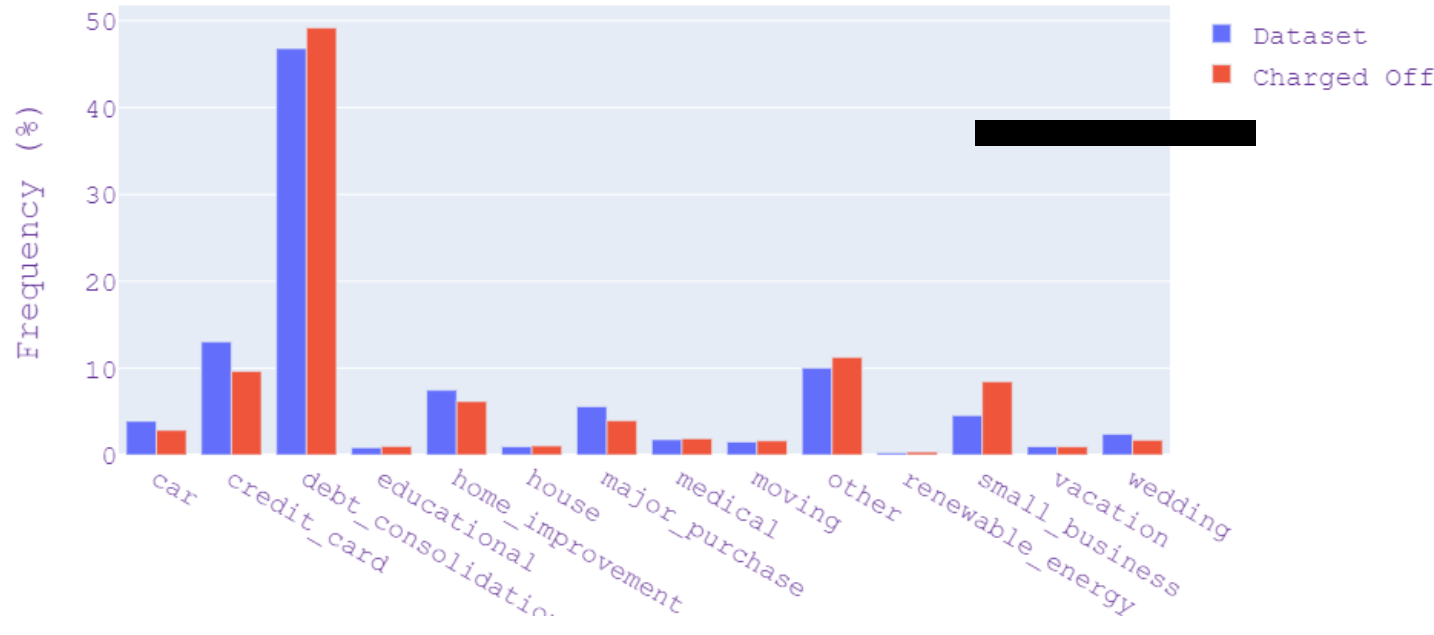
Analysis – Status Distribution

– Around 14.6% of the loans are charged off.

– Aim is to identify the factors likely to increase this risk of default.
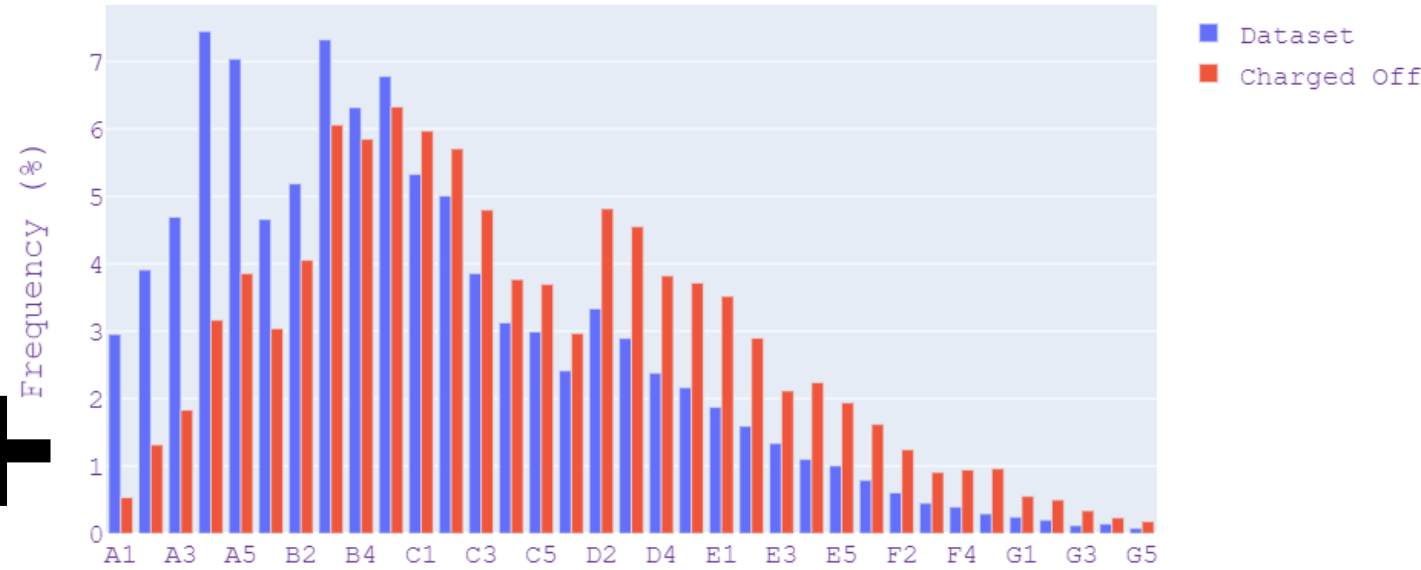

Distribution of loan_status
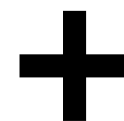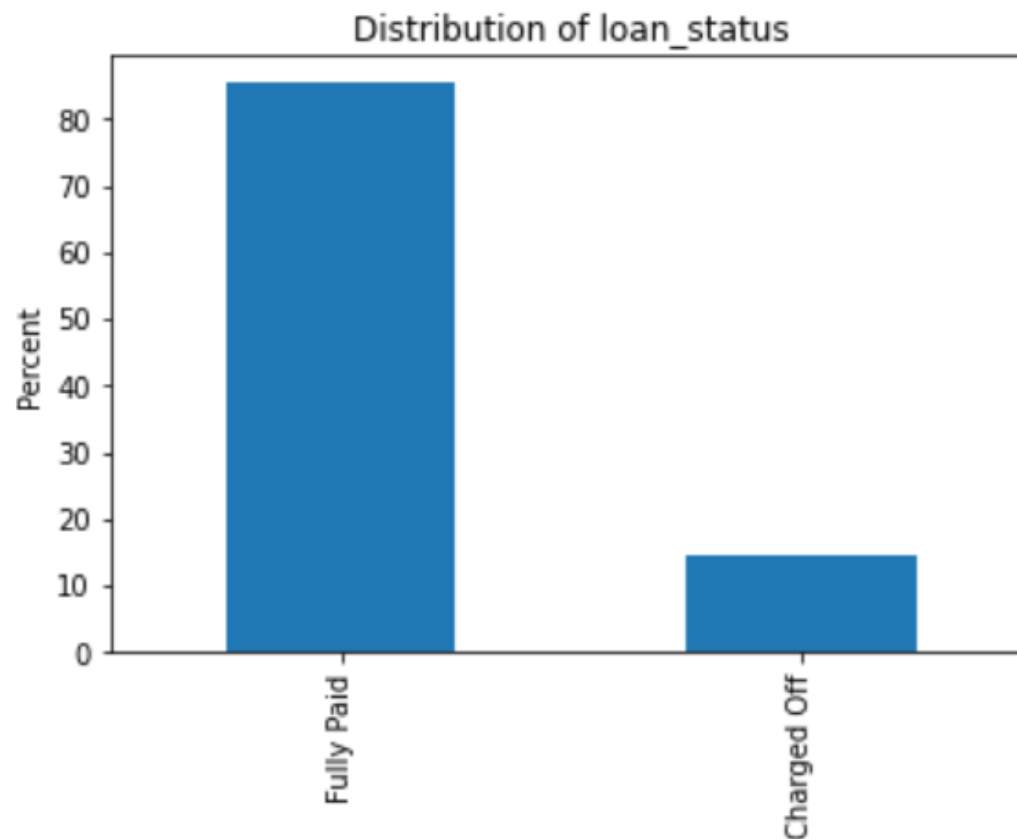
– Higher chance of default for lower grades.

– Most of the loans are sanctioned to California, Texas, Florida and New York. Florida has a slightly higher chance of charging off.
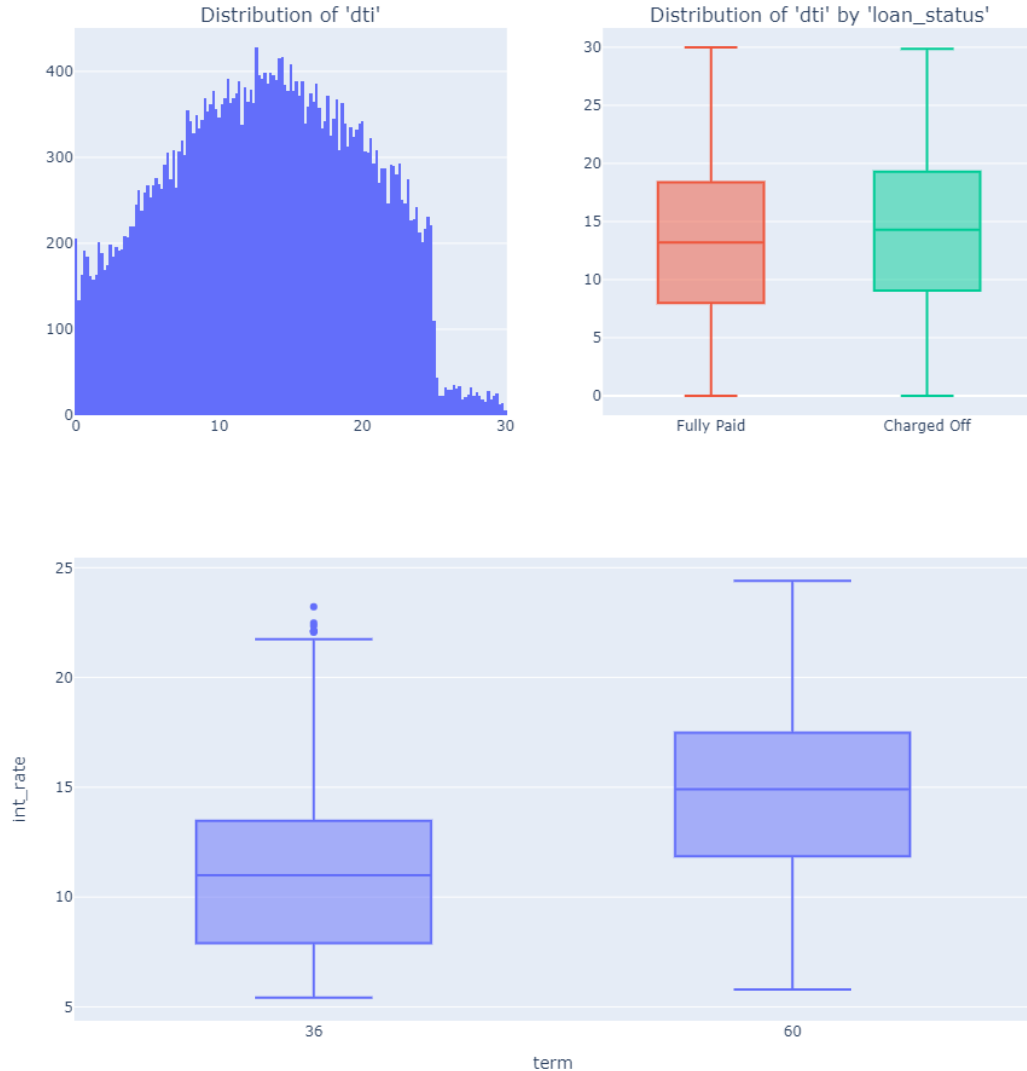
– Those with income status verified have higher probability of charge-off!

– Defaults are significantly higher in proportion for 60-month loan tenures

Distribution of 'dti'
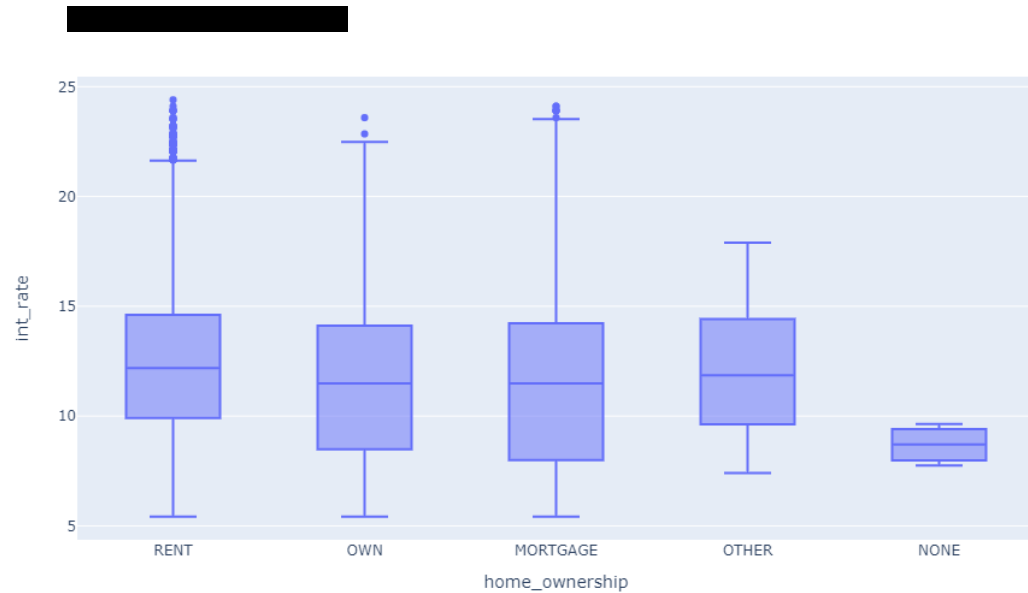
Distribution of 'dti' by 'loan_status'
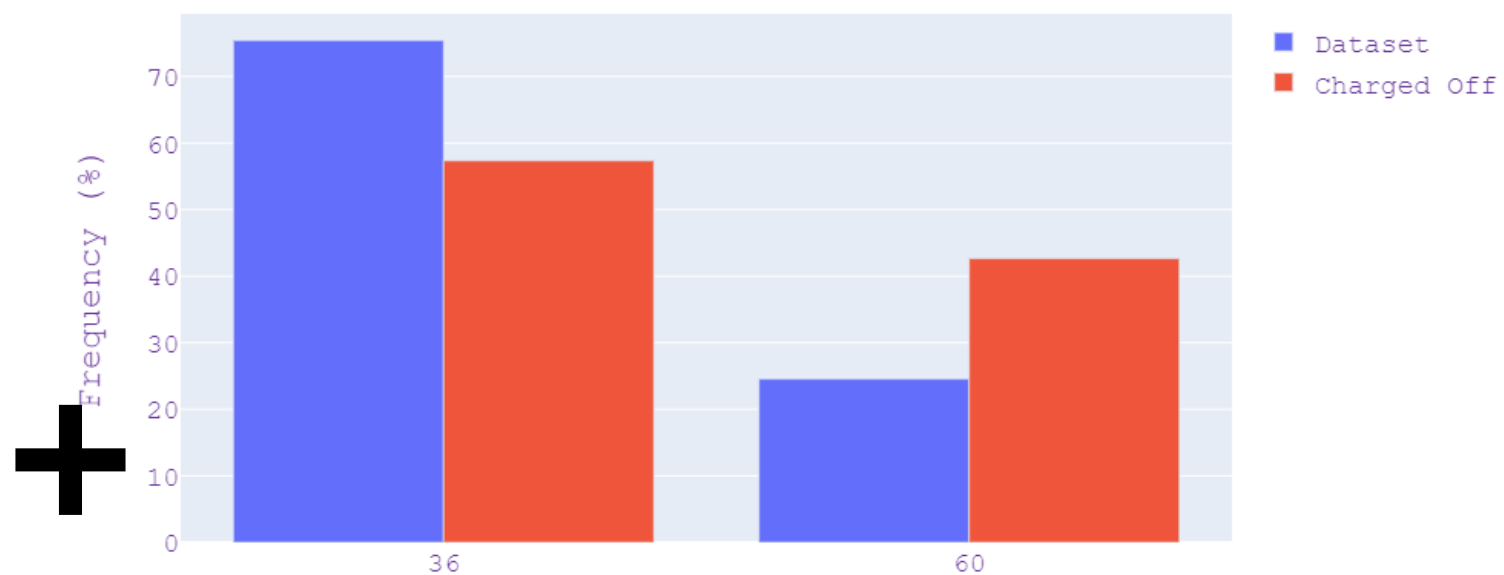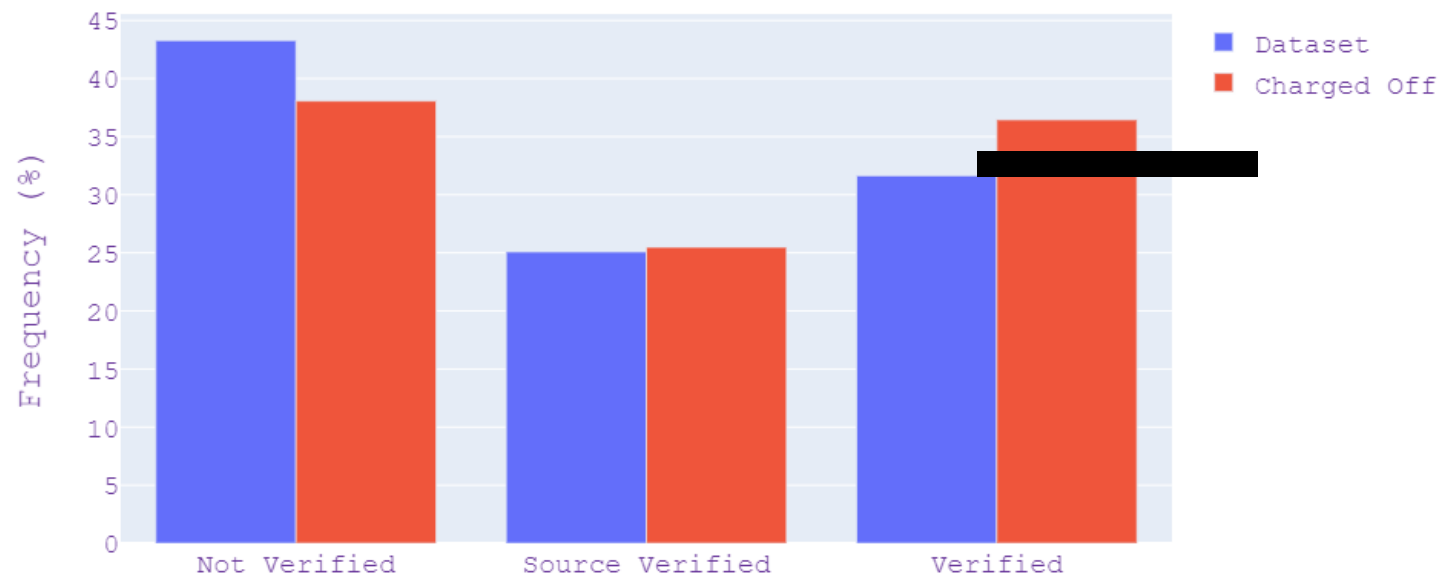
- Interestingly, there is a big drop in dti (debt-to-income) values after 25.

- As expected, dti of defaulters is a bit higher.

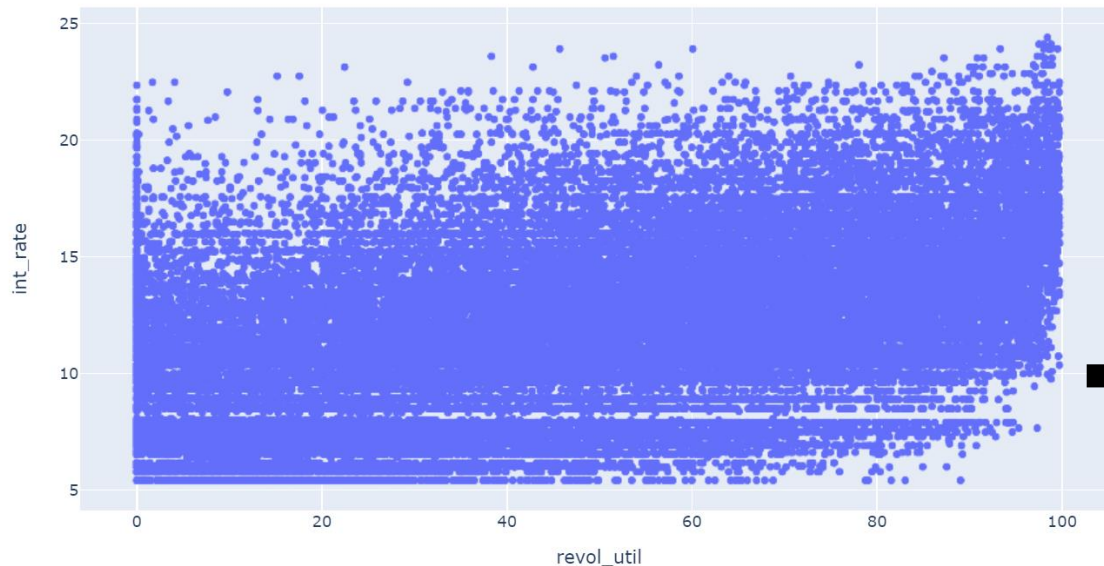- Interest rates are higher for loans with longer tenure!

+

– There is an increasing trend of higher interest rates for higher risk loans (based on sub-grade)
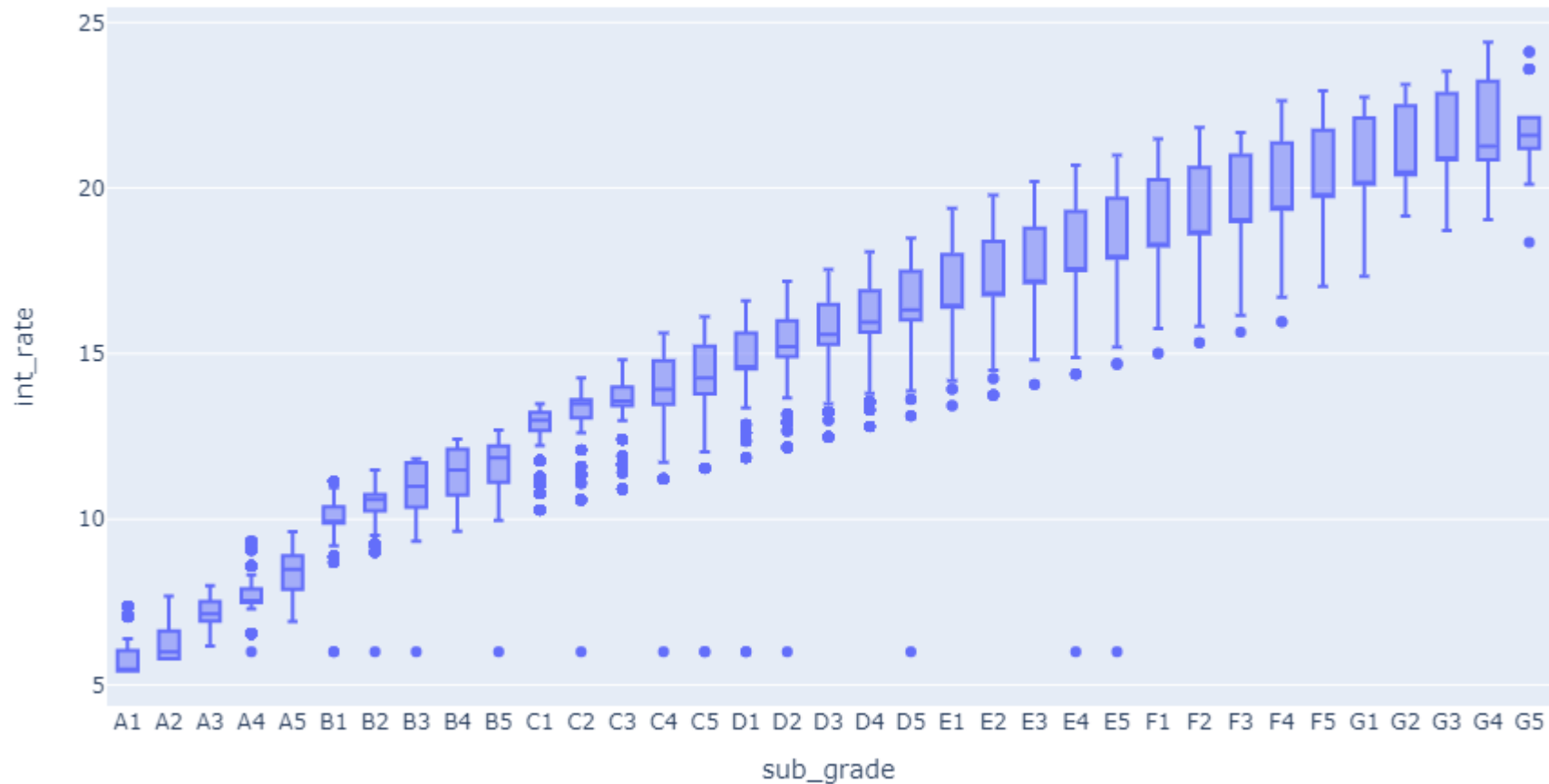
– Although, not much, those with own home tend NOT to have very high interest rates compared to mortgagers/tenants

– Interest rate is decently correlated with revol_util (revolving credit line utilization). This could be because higher the utilization, lower would be the credit score.

Distribution of 'last_pymnt_amnt'

Distribution of 'last_pymnt_amnt' by 'loan_status'
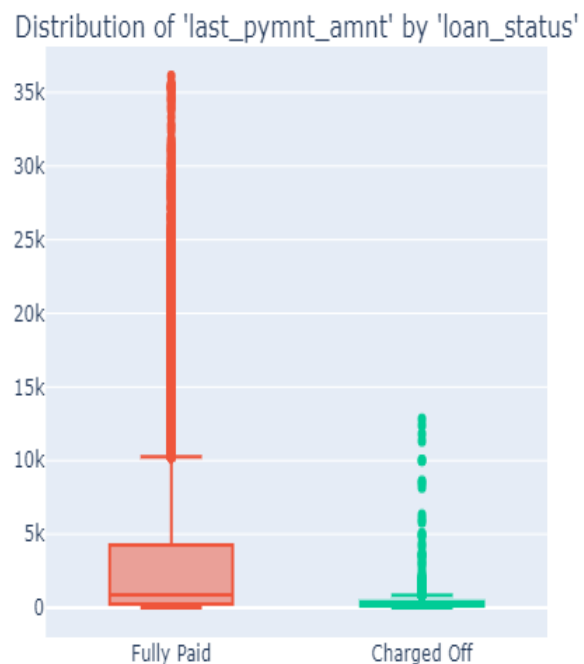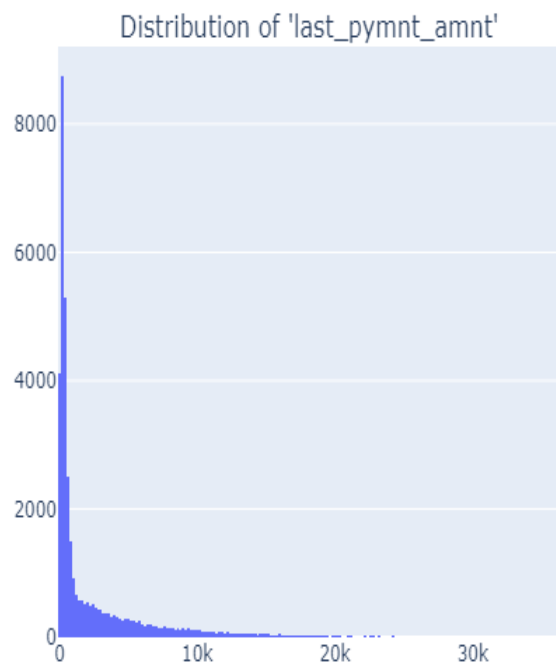
While those who pay fully paid a lot as the last payment, defaulters hardly paid anything.

Thank you!