LENDING CASE STUDY

Presenters :

Dhananjay Buva
Shriniket Pimparkar

OBJECTIVES

- To determine likelihood of default using exploratory data analysis techniques

- Data Clean up and standardization

- Data Understanding

- Univariate Analysis

- Segmented Univariate Analysis

- Bivariate Analysis

- Conclusions with Each Analysis

DATA BACKGROUND

- No of columns : 111 with data types

  float64(74), int64(13), object(24)

- No of rows : 39717

- The loan records range from 2007 to 2011 while credit information ranges from 2007 to 2016

DATA CLEAN UP

- Dropped empty columns (55 no's)

- Dropped columns with Unique Values (9 no's)

- Delete the rows with "Current" loan status

- Delete the columns with <30% not-null data (2 no's)

- Drop the unnecessary columns which are not useful for analysis

- Shape of data after cleanup : 29 columns and 38577 rows

- Action Items : earliest_cr_line has records till 2068, authenticity of this data to be seen.

DATA STANDARDIZATION

- Convert and standardizing data types of columns as applicable

- Removing special characters from columns (like %)

- Replace Nan with 0 as applicable

- Set floating numerals to the precision of 2

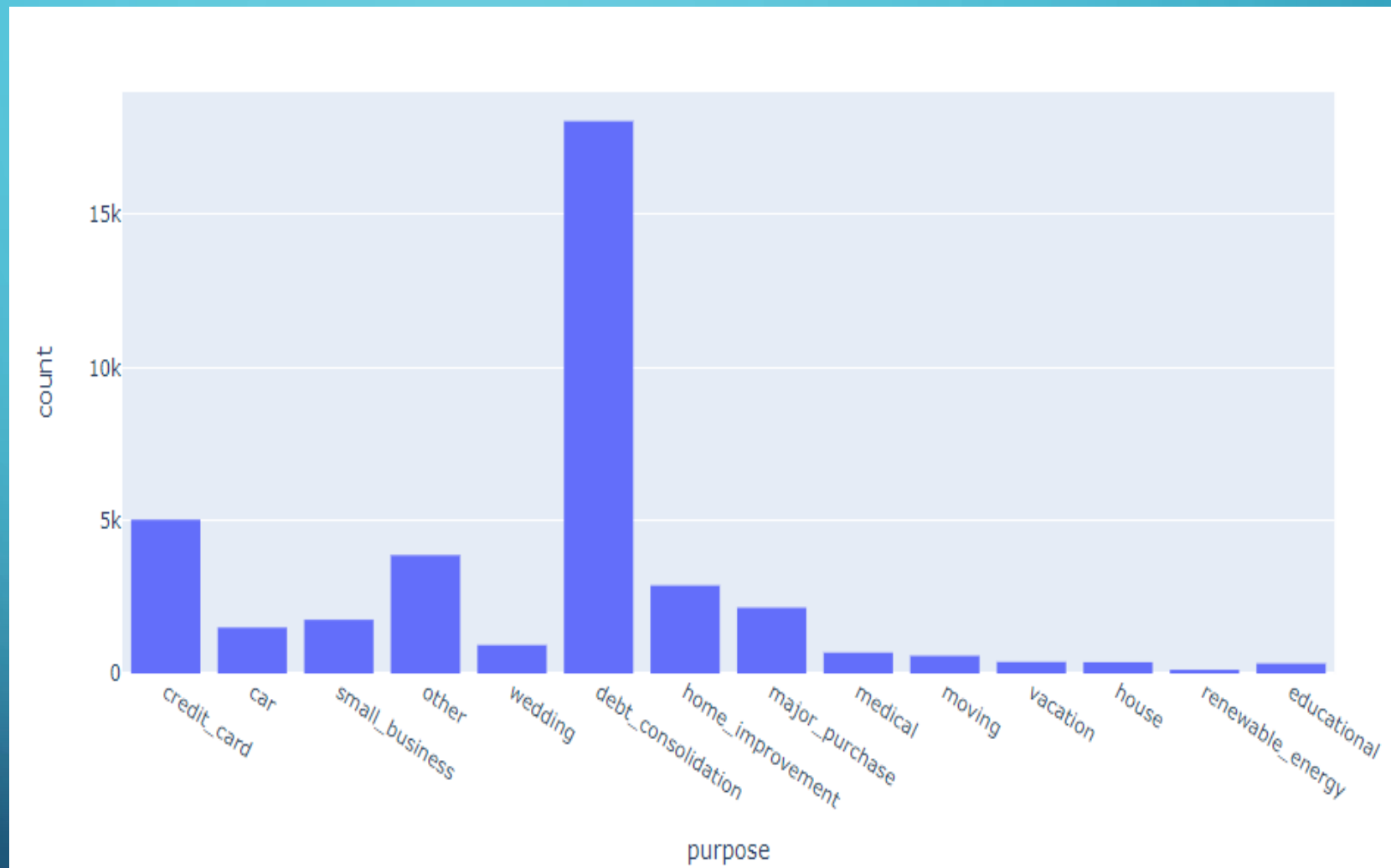- Outlier treatment

PURPOSE COLUMN ANALYSIS

Conclusion:

Top 3 reasons for taking loan are :
1. Debt consolidation
2. Credit card
3. Other
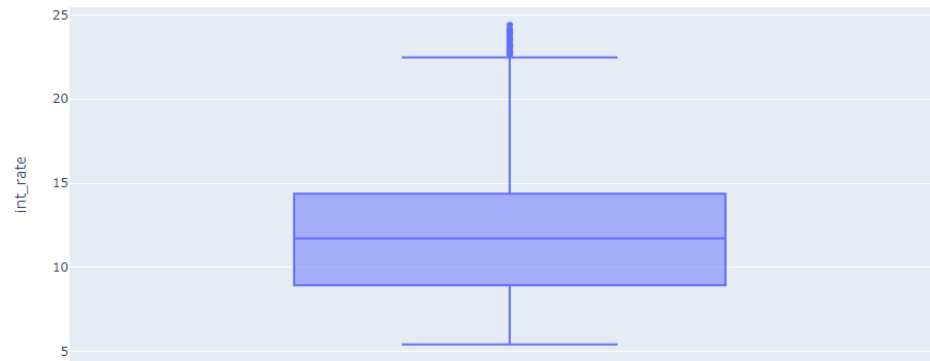And bottom 2 are :
1. Renewable energy
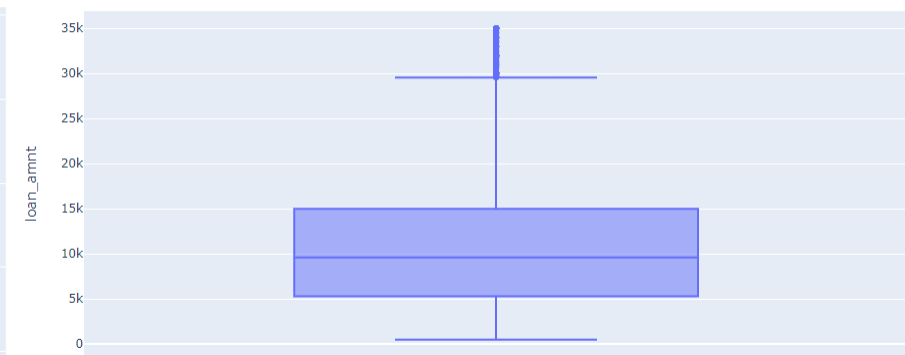2. Education

UNIVARIATE ANALYSIS : QUICK LOOK

CONCLUSIONS:
1. Out of total data, ~15% loans were charged off
2. Outliers adjusted for annual income
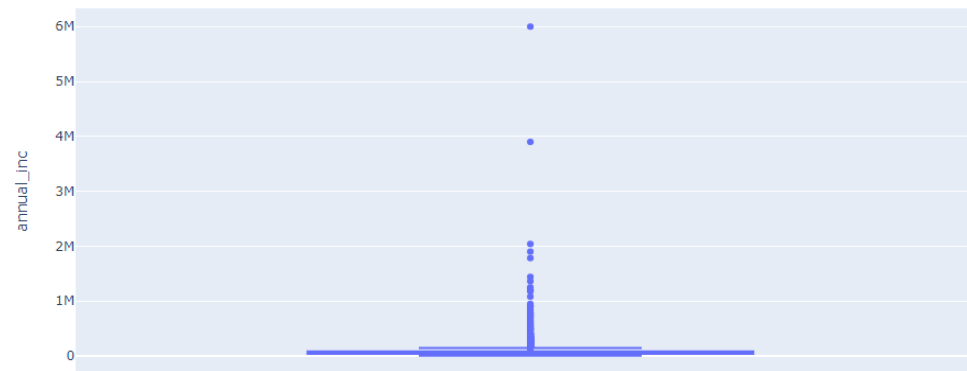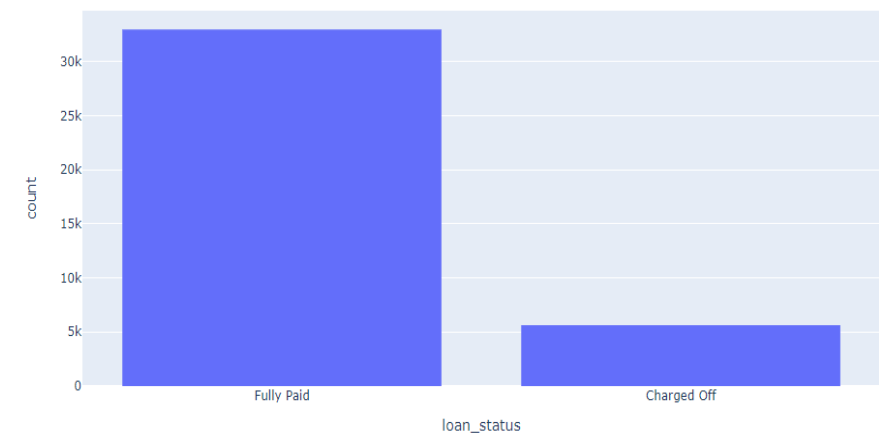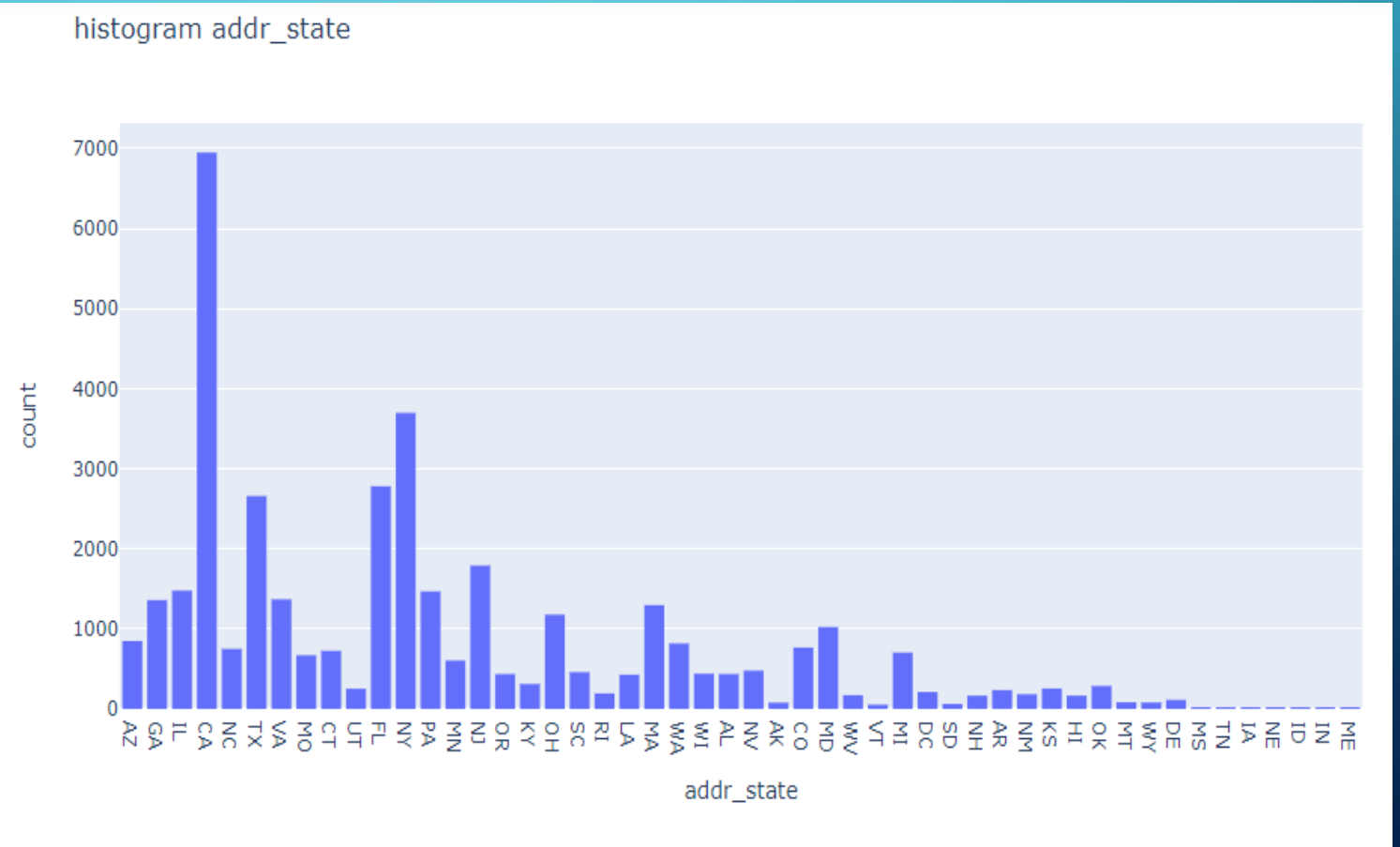3. The loan amount ranged till 35k, median being 10k

GEOGRAPHICAL DISTRIBUTION OF LOAN APPLICANTS

Conclusion:
States like CA , FL ,NY, NJ have higher loan applicants.

CORRELATIONS

|	loan_amnt	funded_amnt	int_rate	installment	annual_inc	open_acc	revol_bal	revol_util	total_rec_prncp	total_rec_int
loan_amnt	1.00	0.98	0.30	0.93	0.27	0.18	0.31	0.06	0.85	0.73
funded_amnt	0.98	1.00	0.30	0.96	0.26	0.18	0.31	0.07	0.86	0.74
int_rate	0.30	0.30	1.00	0.28	0.05	0.01	0.09	0.47	0.17	0.52
installment	0.93	0.96	0.28	1.00	0.27	0.17	0.31	0.09	0.85	0.64
annual_inc	0.27	0.26	0.05	0.27	1.00	0.16	0.28	0.02	0.26	0.19
open_acc	0.18	0.18	0.01	0.17	0.16	1.00	0.29	-0.09	0.16	0.12
revol_bal	0.31	0.31	0.09	0.31	0.28	0.29	1.00	0.30	0.28	0.24
revol_util	0.06	0.07	0.47	0.09	0.02	-0.09	0.30	1.00	0.02	0.19
total_rec_prncp	0.85	0.86	0.17	0.85	0.26	0.16	0.28	0.02	1.00	0.68
total_rec_int	0.73	0.74	0.52	0.64	0.19	0.12	0.24	0.19	0.68	1.00

Conclusion:

Loan amount, funded amount, installment ,total rec principal , total rec interest are correlated

Interest rate is positively correlated with revolving Util, loan amount.

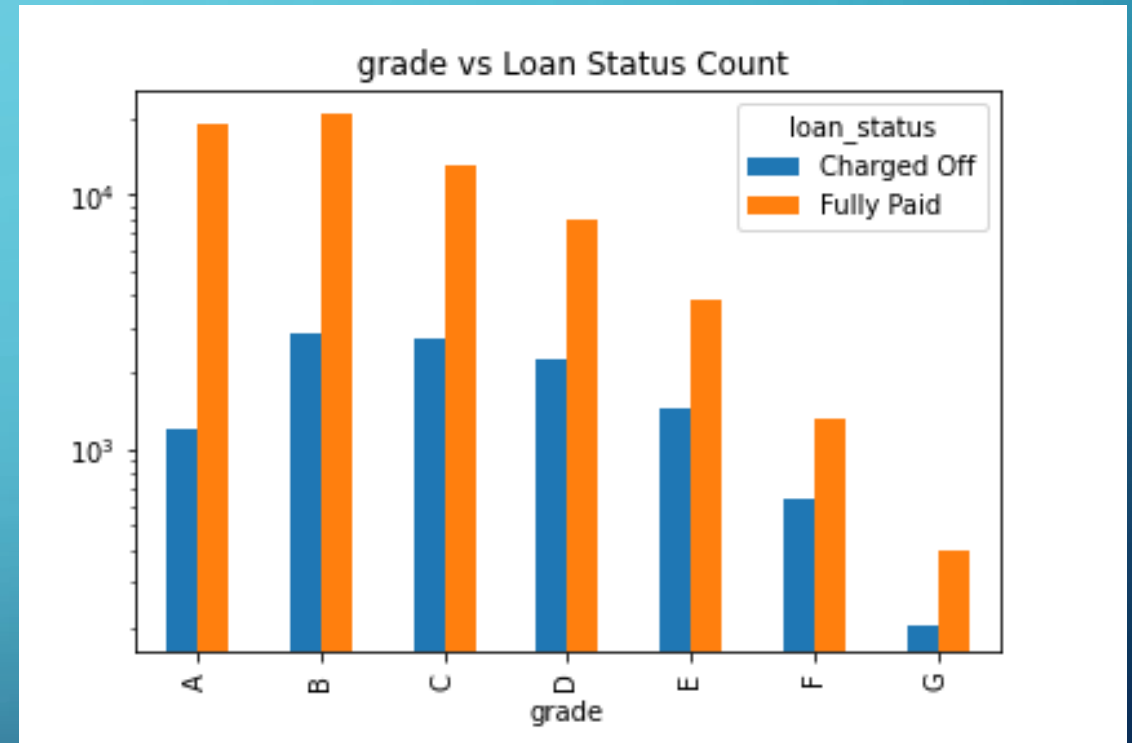# IMPACT OF PURPOSE ON LOAN STATUS

Conclusions:

1. Purpose impacts by ~17%.

2. Top 3 purposes(in order) leading to default are –

   1. Small business
   2. Renewable energy
   3. Educational
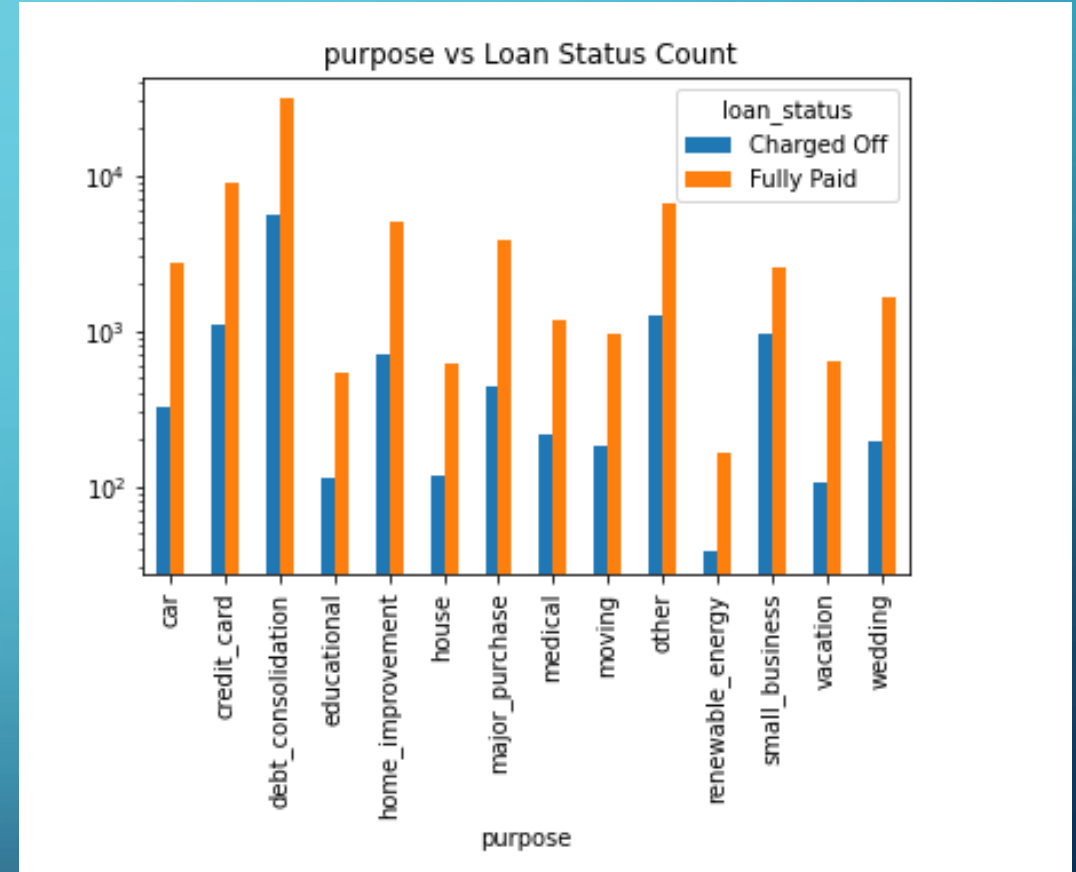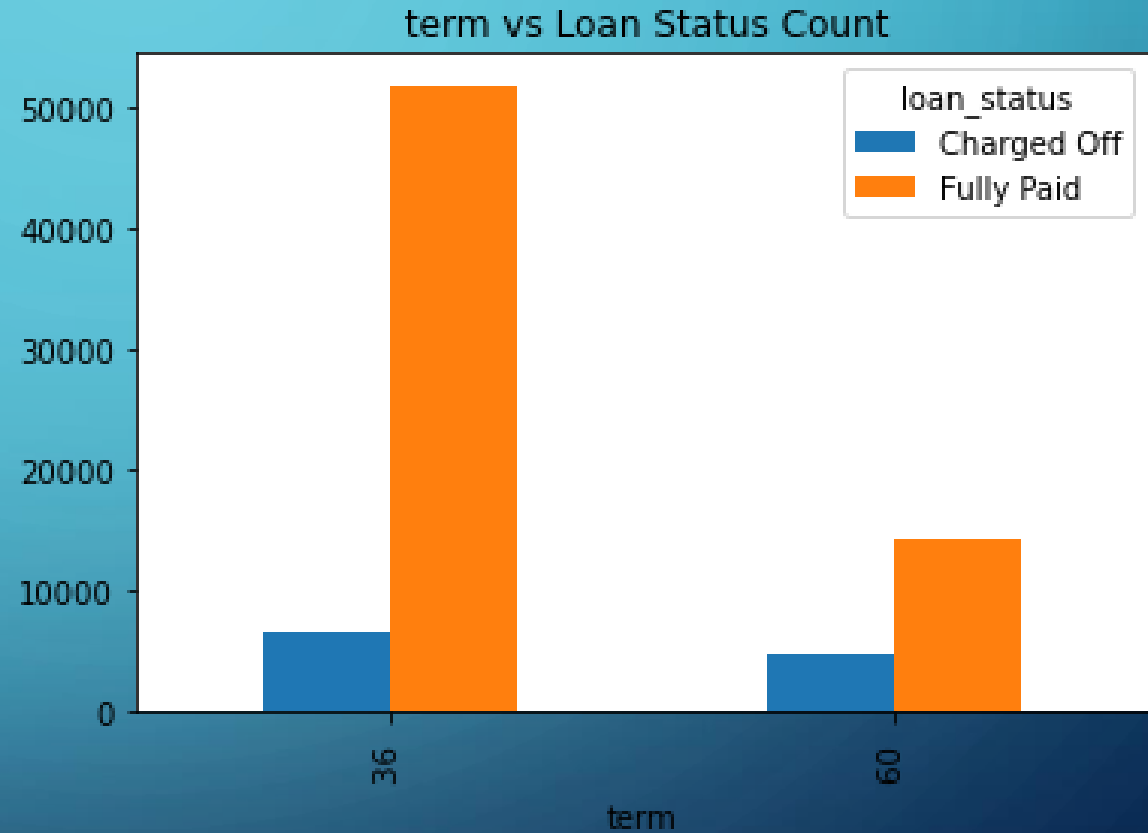
IMPACT OF PURPOSE ON LOAN STATUS

Conclusions:
1. Grade has around 26% impact
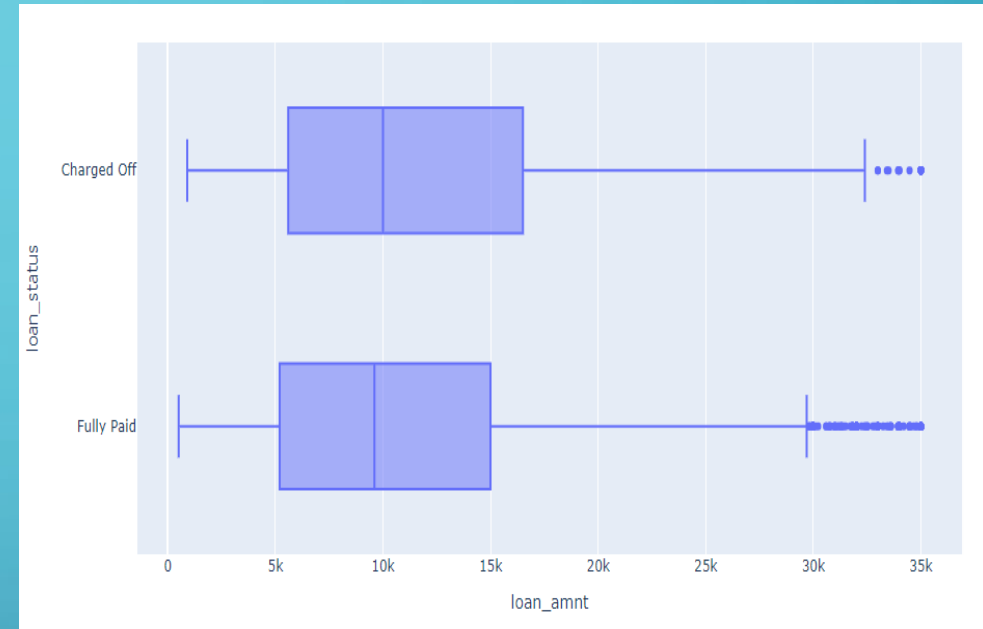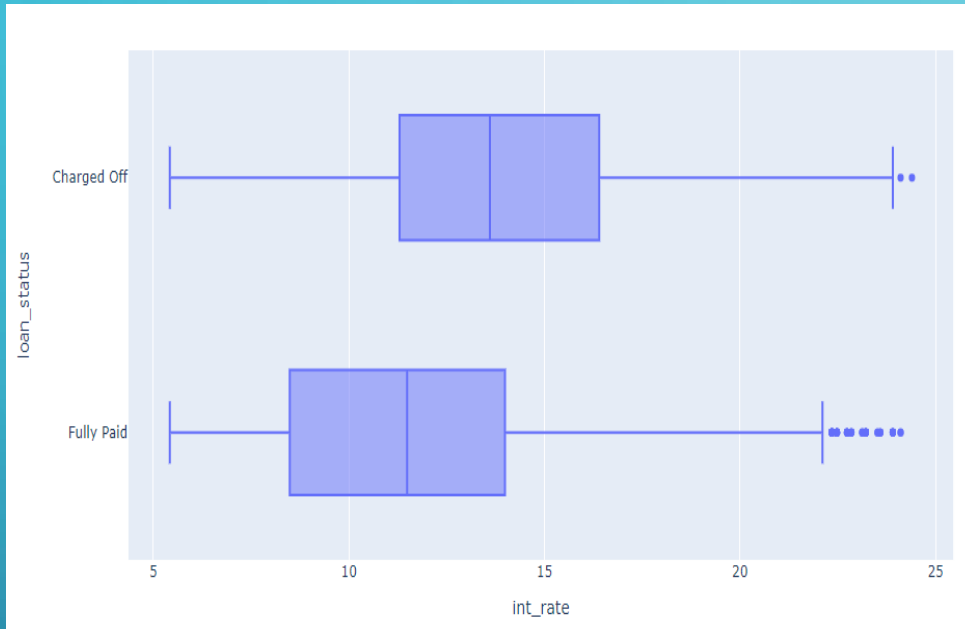2. As grade decreases from A to G , number of defaults increase.

IMPACT OF TERM ON LOAN STATUS

Conclusions:
1. Term has ~ 14% impact
2. Defaults increase as term increases

term vs Loan Status Count

LOAN AMOUNT & INTEREST RATE VS LOAN STATUS

Conclusions:
1. Since interest rate and loan amount are correlated , seeing impact together
2. Loan Amount has ~8% impact and interest rate has > 25% impact
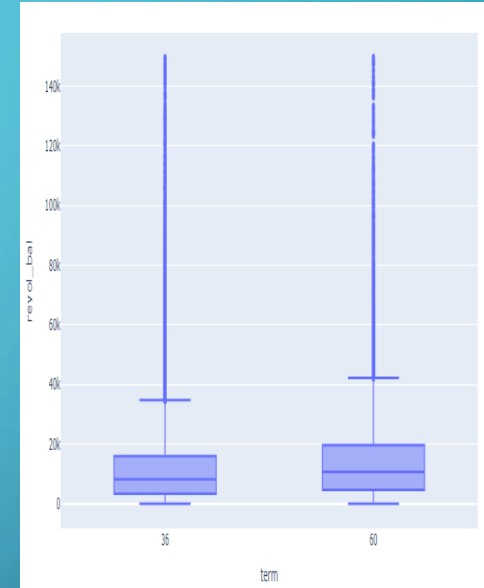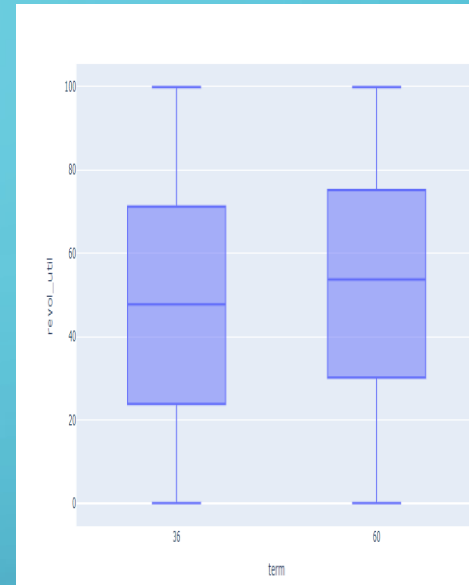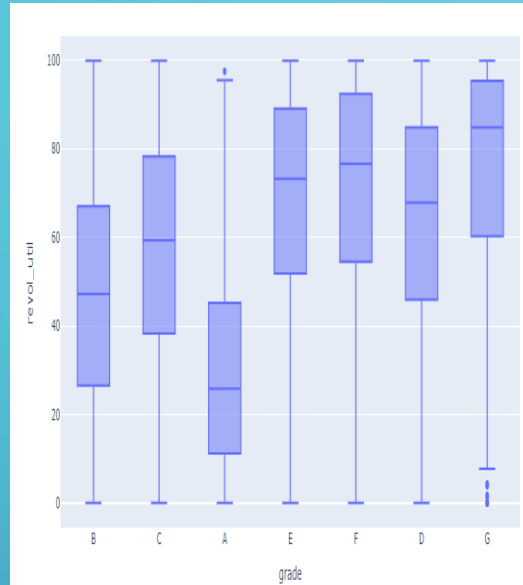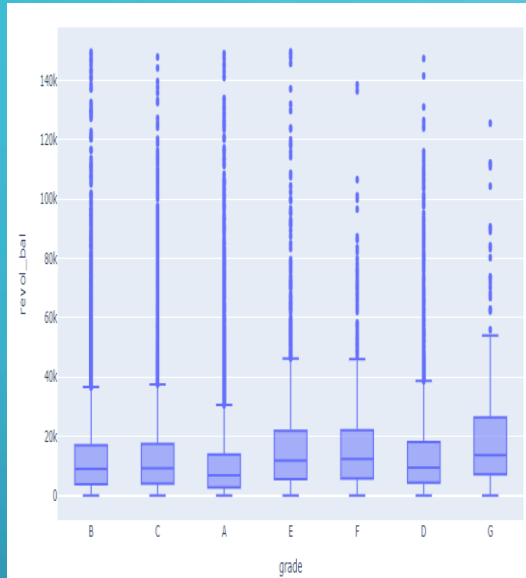3. As loan amount / interest rate increase, chances of default increase

IMPACT OF REVOLVING UTILIZATION RATE

Conclusions:
1. Revolving util has around 12% impact
2. In general, we can say that as revolving util increases, chance of default increase
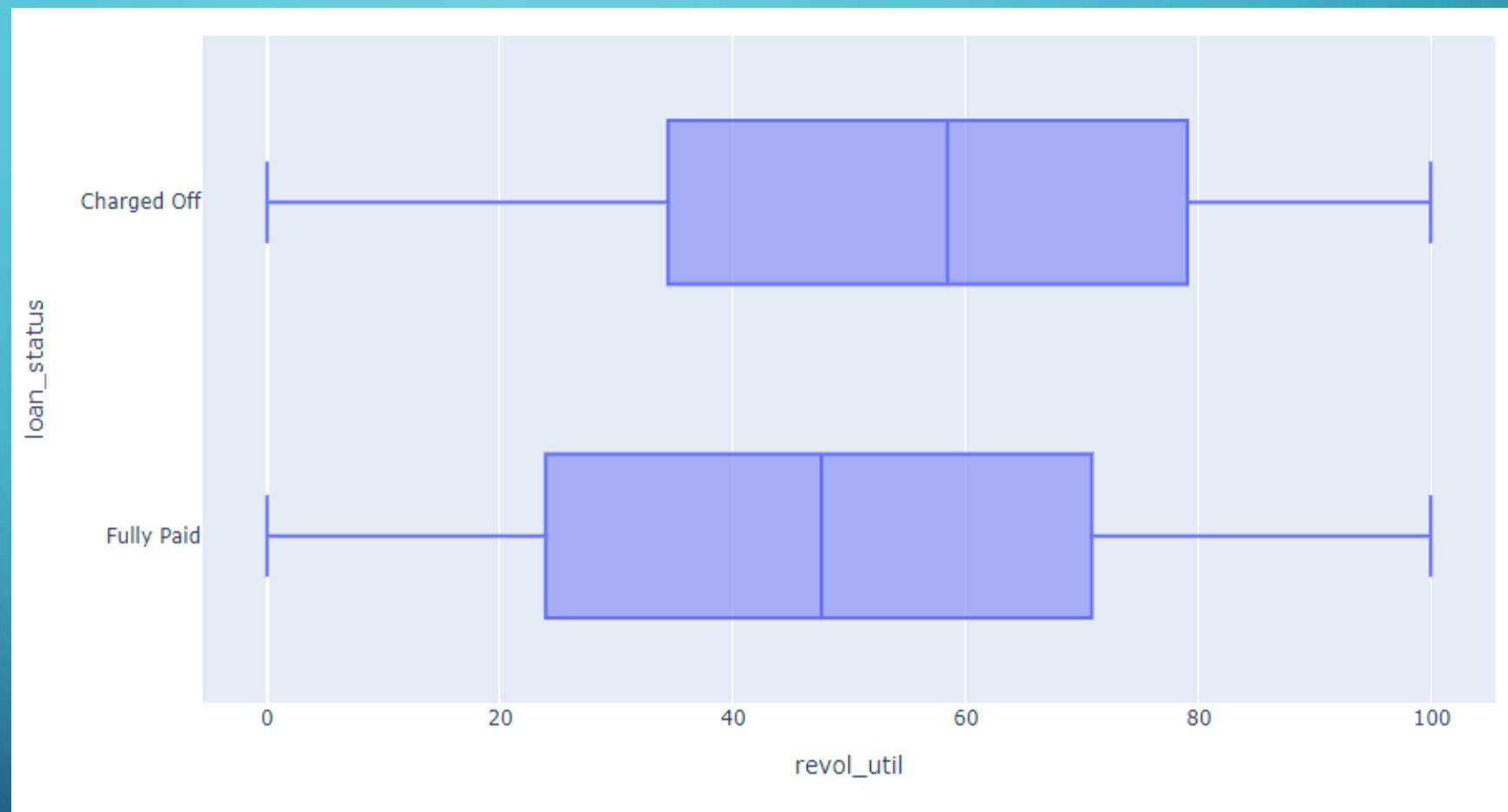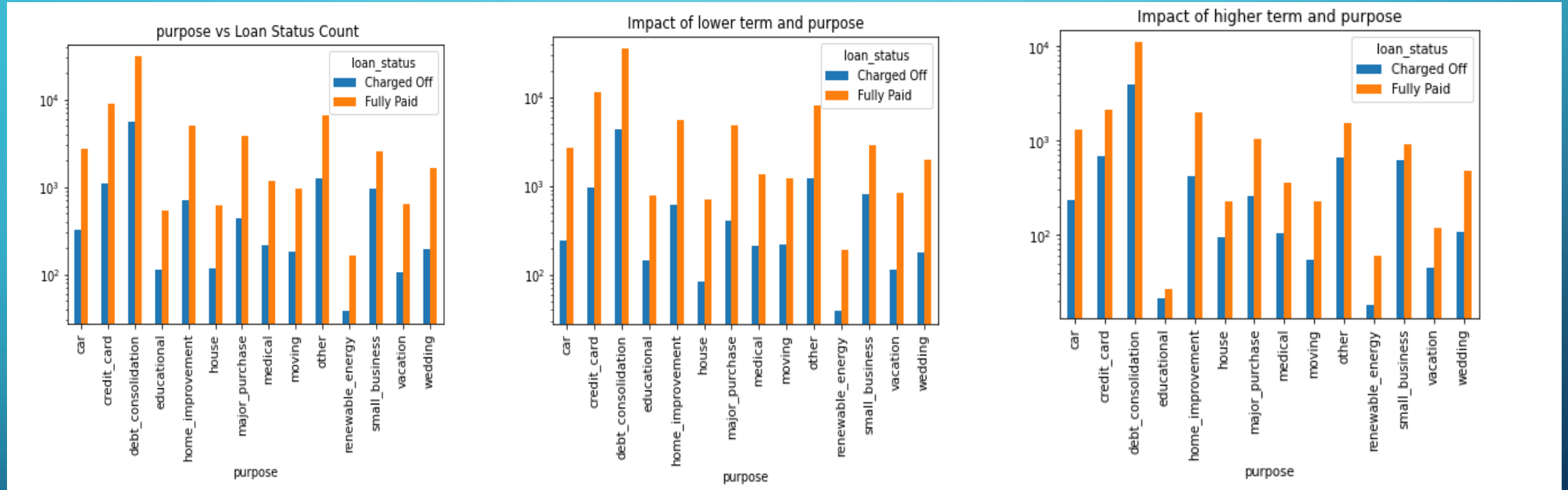
GRADE VS REVOLVING BALANCE AND UTILIZATION

Conclusions :
1. As grade decreases from A to G , revolving balance and its utilization increases
2. As term increases , revolving balance and its utilization increases

IMPACT OF TERM AND PURPOSE

Conclusion:

As term increases, impact of purpose is higher , the total magnitude being ~ 28 %.
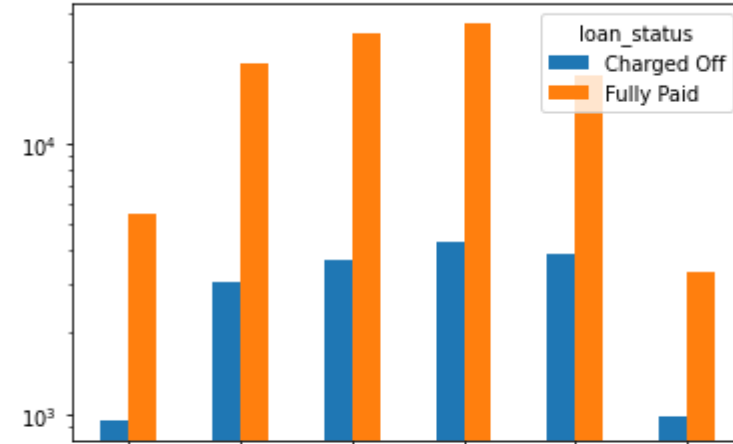
This can be seen for categories like education, debt consolidation.

IMPACT OF LOAN AMOUNT AND ANNUAL INCOME

Conclusion:

1. Combined impact of loan amount and annual income is ~ 9%.
2. When both income and loan amount are higher (loan amount > 15k and income > 80k) , chances of default are higher.
3. When ratio of loan amount by income is > 0.5 , chances of default are higher.
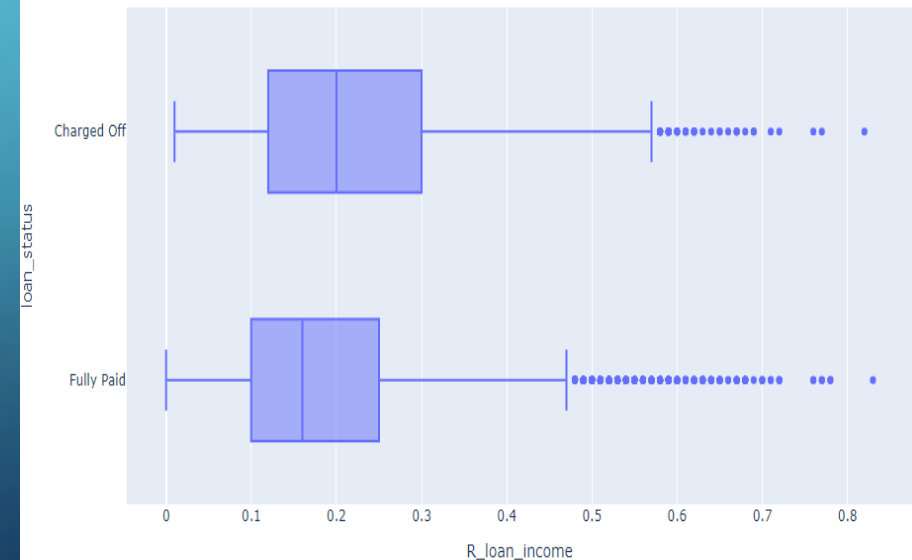
SUMMARY

- A quick look at data revealed good insights.

- By proper selection of parameters, likelihood of default can be guessed.

- There is possibility of further analysis which can provide lot more insights.