EDA CASE STUDY(Consumer Finance Company)

SUBMISSION

Group Name:

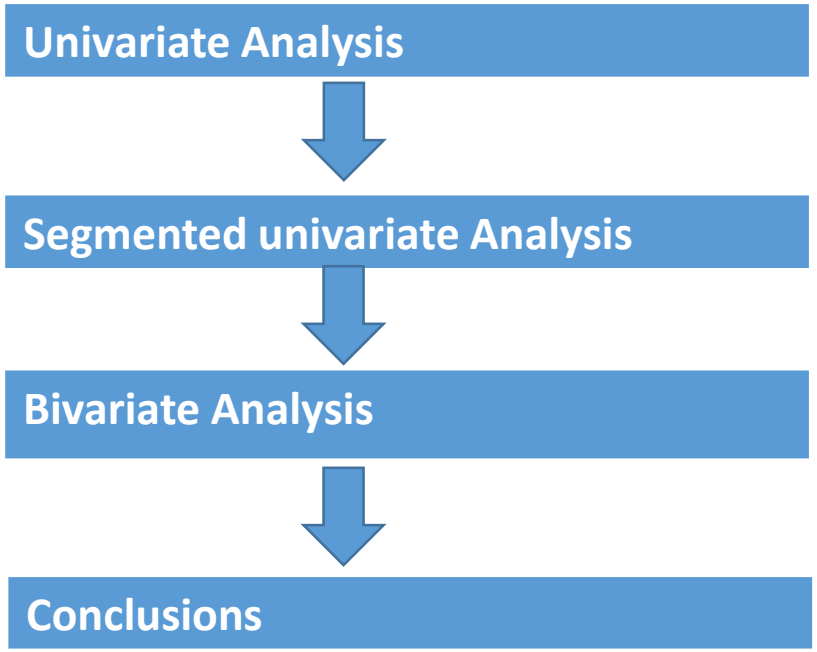
1. Sudheer Kumar Payyavula

Abstract:

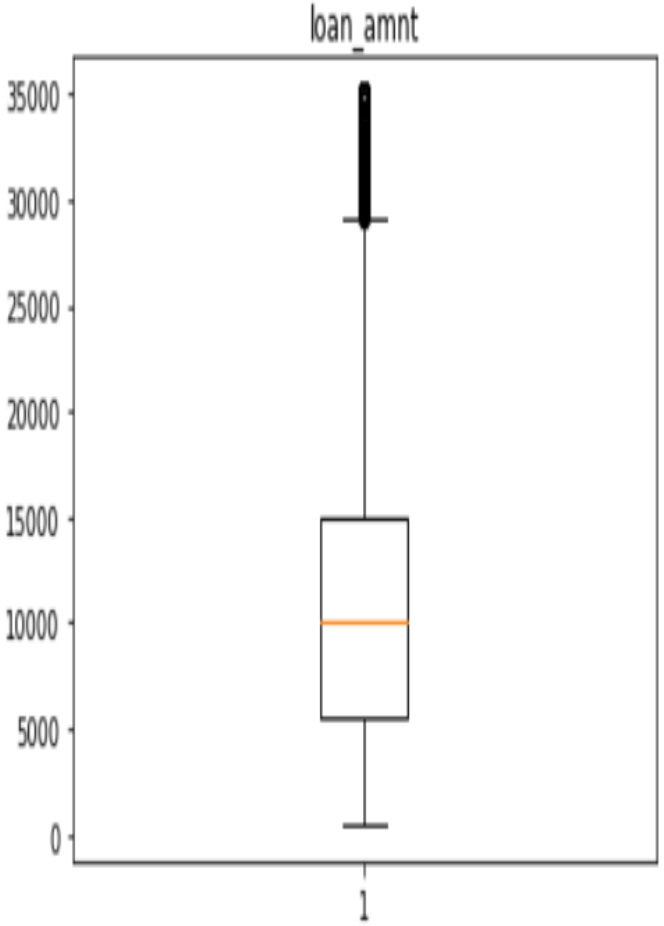
- This presentation contains the analysis of a consumer finance companies dataset.
- Our main intention of this analysis is to determine the driving factors for the loan charged off's (Loan defaults)
- Let's look into the data and get the conclusions

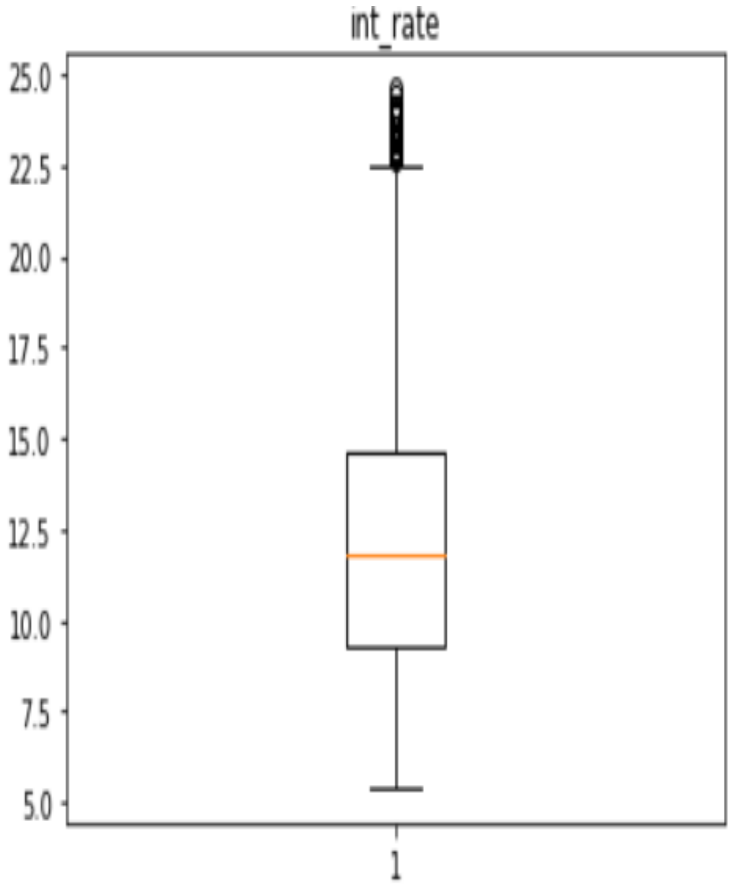
Problem solving methodology:

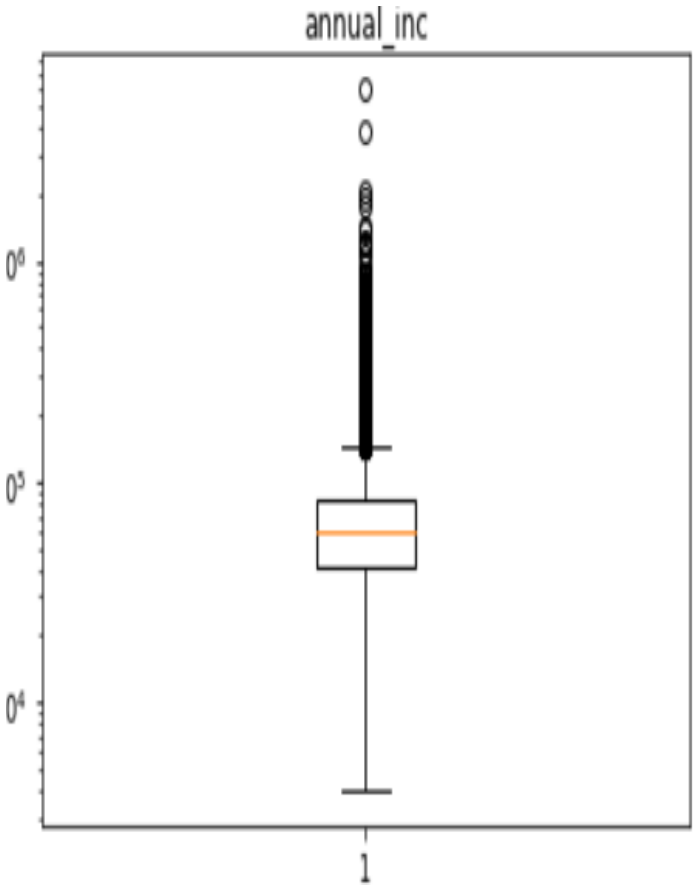
- Followed below approach for the analysis.



Univariate Analysis- Box plots:







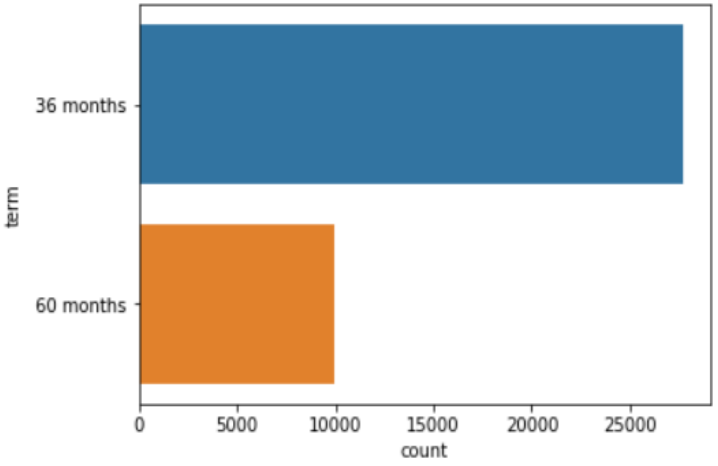
Univariate Analysis:

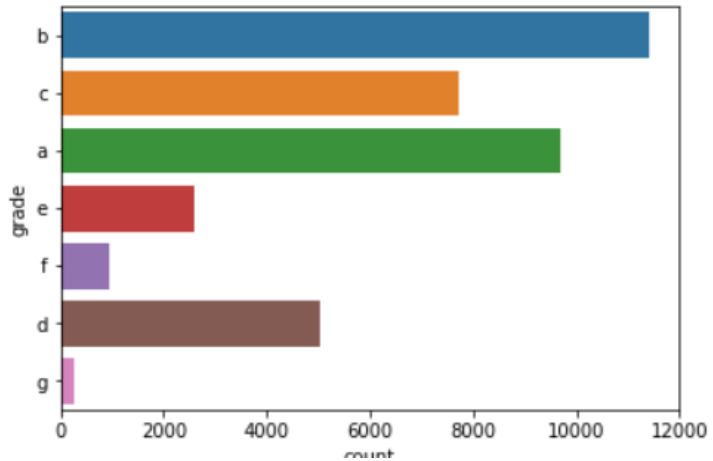
- For this we have plotted Boxplots for all the quantitative variables to check for the outliers.
- We have plotted Bar graphs to understand the categorical variables.
- Calculated summary metrics for all the quantitative variables.

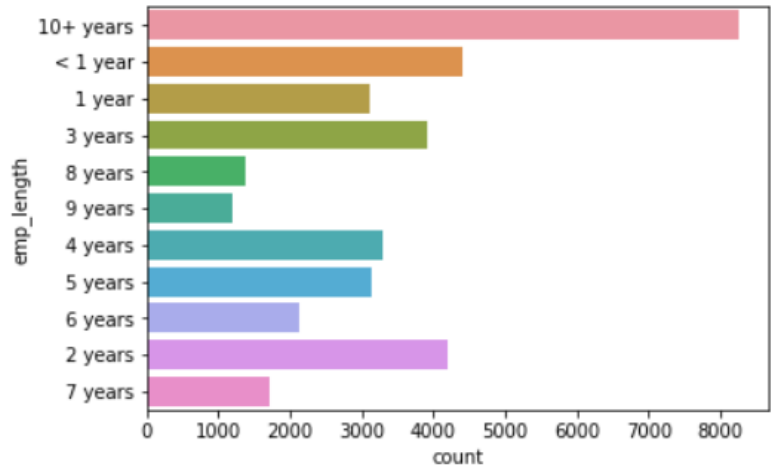
Univariate Analysis- Box Plots-Conclusions:

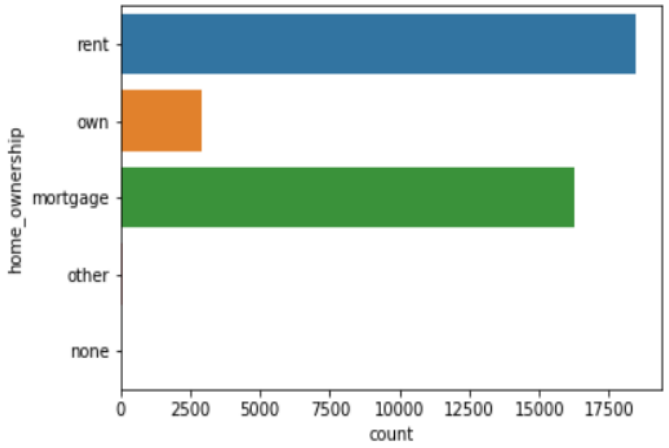
- Loan_amnt is widely spread we are seeing few no.of outliers.These may not impact the analysis so we kept them as it is.
- Funded_amnt and funded_amnt_inv are widely spread we are seeing few no.of outliers.These may not impact the analysis so we kept them as it is.
- int_rate is widely spread we are seeing few no.of outliers.These may not impact the analysis so we kept them as it is.
- dti has no outliers.
- Most values of the fields delinq_2yrs and pub_rec are '0' that's why all the values are concentrated at '0' in boxplot.
- Most values of inq-last_6mths are '0' and 1 that's why all the values are concentrated at '0' and 1 in boxplot.
- total_paymnt and total_paymnt_inv are widely spread and have good no.of outliers this is because loan_amount's wide distribution.
- total_rec_late_fee and collection_recovery_fee fields contain more no.of '0' values that's why it is concetrated at 0 in boxplot so we can ignore the outliers here.
- We have annual_inc field which have more no.of outliers which may affect our analysis.So we have removed the outliers for this.

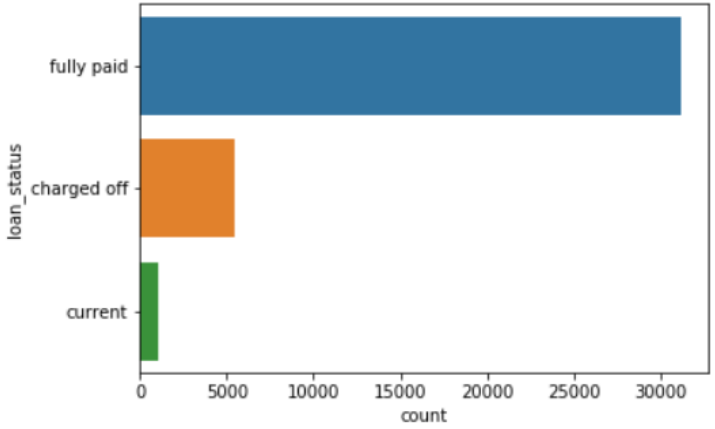
Univariate Analysis- Bar plots:

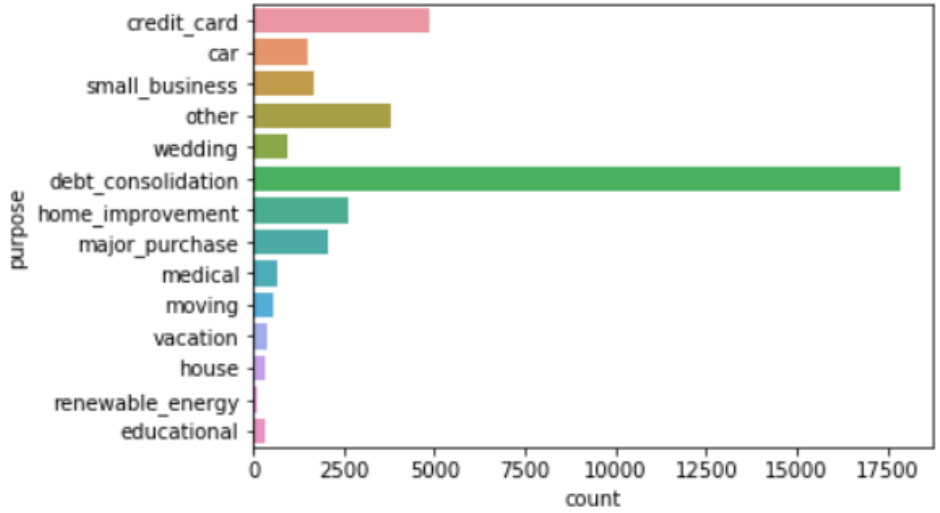












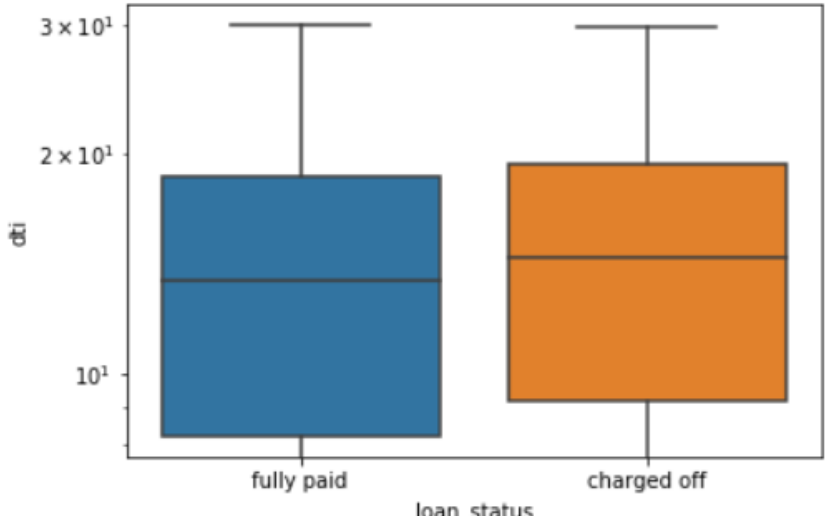
Univariate Analysis- Bar Plots-Conclusions:

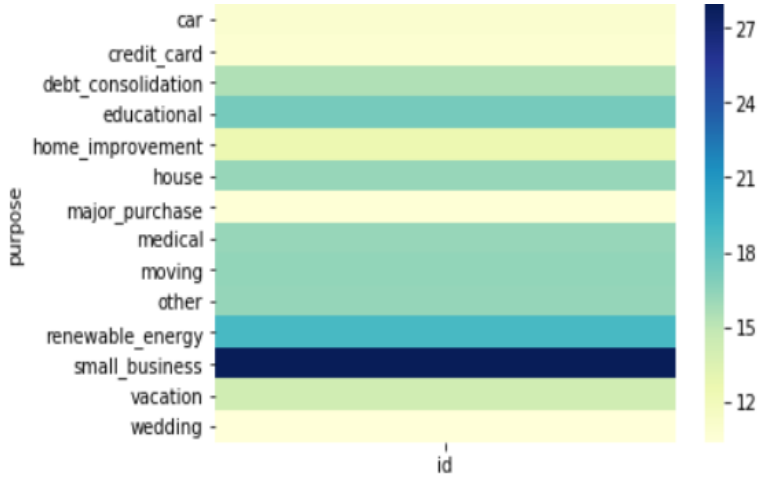
- 'loan_terms' are either 36 or 60 months out of which more loans are for 36 months term.
- Most of the loans fall under the grades 'a','b' and 'c'.
- Most of the loans of the dataset are for 10+ years experienced people.
- Very less borrower's have own house.
- Considerable no.of charged-off's are there in the dataset.
- Most of the loans taken for the purpose of 'Debt consolidation'.
- All the applications of the dataset are for 'Individual' type.

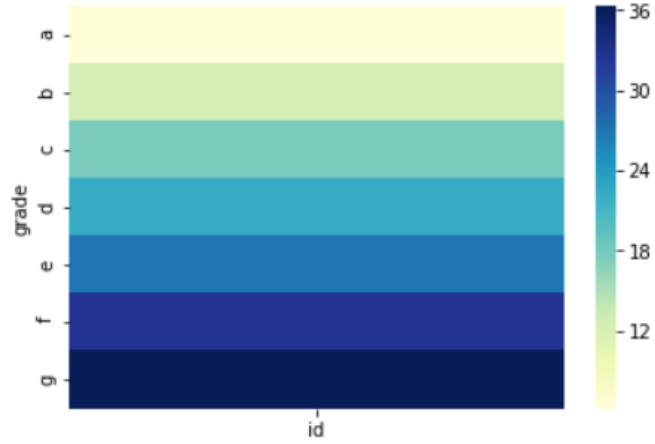
Segmented Univariate Analysis:

- For this we have plotted Boxplots for all the quantitative variables by grouping of loan status.
- Concluded the driving factors of loan default from the plots.

Segmented Univariate Analysis- Box plots:







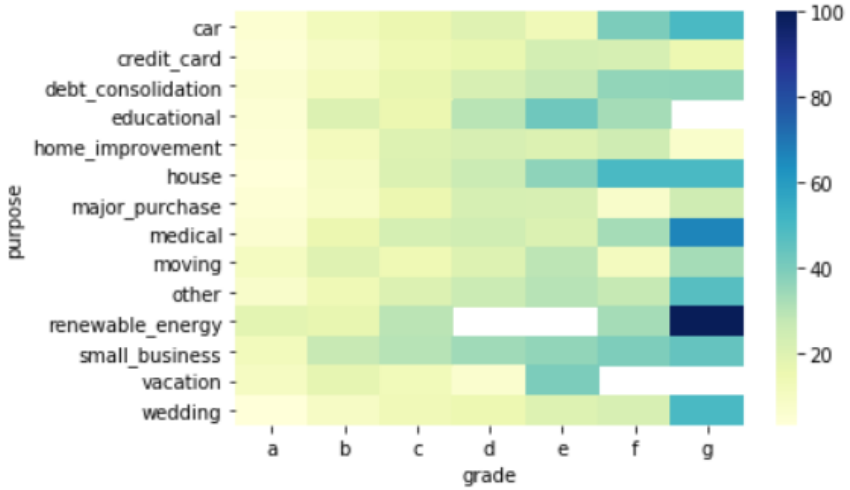
Segmented univariate Analysis- Box Plots-Conclusions:

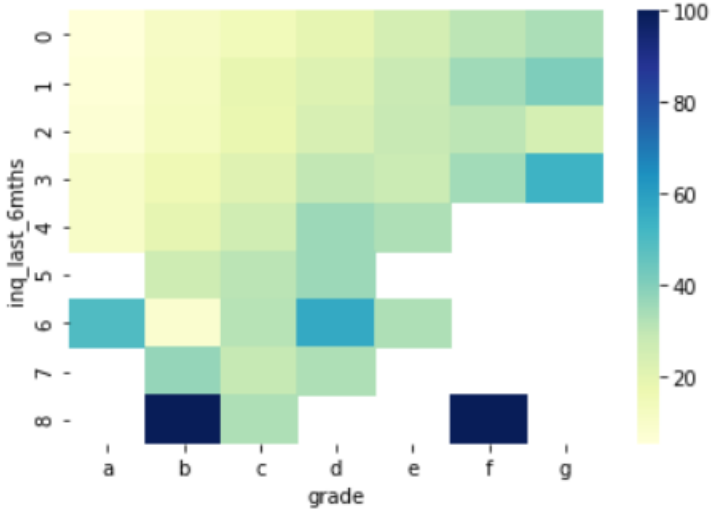
- From boxplot we can observe that Loan_amount and Funded_amount are little higher in the case of defaulters
- Debt-to-income ratio's median is slight higher for defaulter's so DTI would be a factor i.e more the DTI then more the chance of defaulting the loan.
- Interesr rate is clearly looking to be higher for charged off's. So more the interest rates then more the chance of a loan to be getting defaulted.
- Annual_income is also looks to be lower for the defaulter's as per the boxplot. But it is very less variation.
- Revolving_utility is higher for charged off's.
- 'Inquiry in last 6 months' is also differing for 'fully paid' and 'charged off's'.Customers who are having more than 1 inquiry in past 6 months are defaulting more.
- Purpose of the loan is also seems to be important factor. Loans taken for the small business are defaulted more.
- Grade of the customer based on credit score is also a factor as per the heatmap. Lower the grade then more the chance of defaulting the loan.

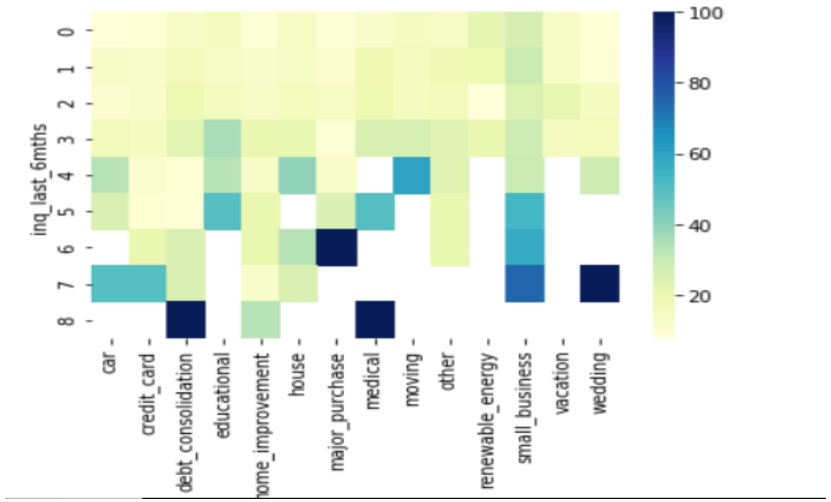
Bivariate Analysis:

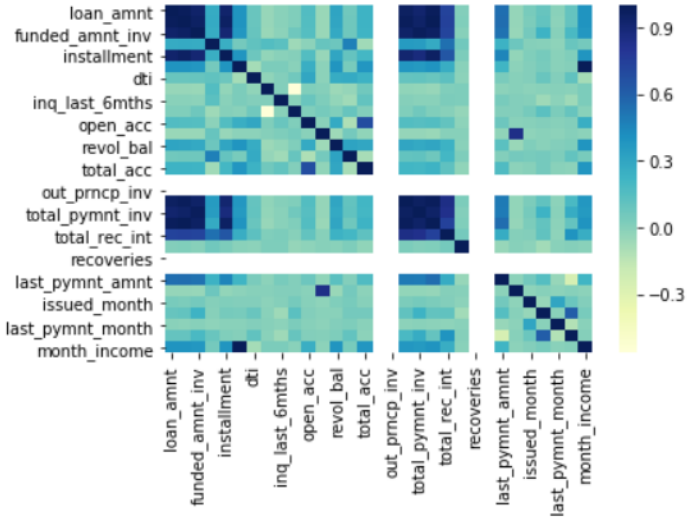
- For this we have found correlation among all the numeric variables.
- Found the combination of driver variables which impact the loan defaults.

Bivariate Analysis- Heat maps:









Bivariate Analysis-Conclusions:

- we can conclude that combinations of 'purpose', 'grade' and 'inquiry in last 6 months' are major factors for the loan default.
- From above we can say that the fields 'loan_amnt', 'loan_amnt_inv', 'funded_amnt', 'installment' are positively correlated.
- ‘Loan amount' and 'oustanding principle' are positively corelated.

Conclusion-Driving factors of the loan default:

Major factors:

1. *Debt to Income ratio* – More the DTI ratio then more the chance of a loan getting defaulted.
2. *Interest rate* - More the interest ratio then more the chance of a loan getting defaulted.
3. *Revolving utility* - More the revolving utility then more the chance of a loan getting defaulted.
4. *Grade given to customer based on credit score* – if the grade of the customer lowers then chance of loan default increases.
5. *Purpose of the loan* – If purpose of the loan is for ‘small business’, ’renewable energy’ and ‘education’ then there is more chance for loan default.

Minor factors:

6. *Annual income* – Lesser the annual income then higher the chance of loan default.
7. *Inquiry in last 6 months* – if higher the no .of inquiries made in past 6 months then higher the chance of loan default.
8. *Public recorded bankruptcies* -if higher the no .of public recorded bankruptcies then higher the chance of loan default.

Combinations of factors:

1. Combinations of 'purpose', 'grade' and 'inquiry in last 6 months' are factors for the loan default.