Lending Club Case Study

SUBMISSION

Name: Prince Adlakha
Pratik Kumar

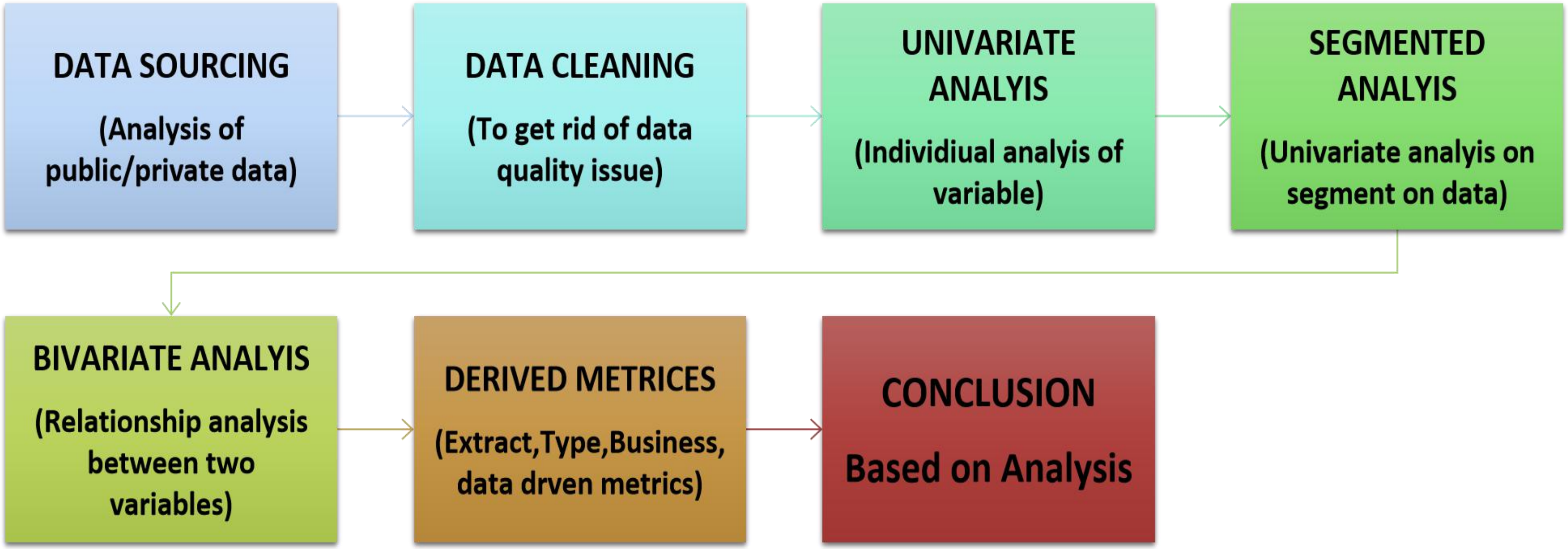
Background :

Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

Business Objective:

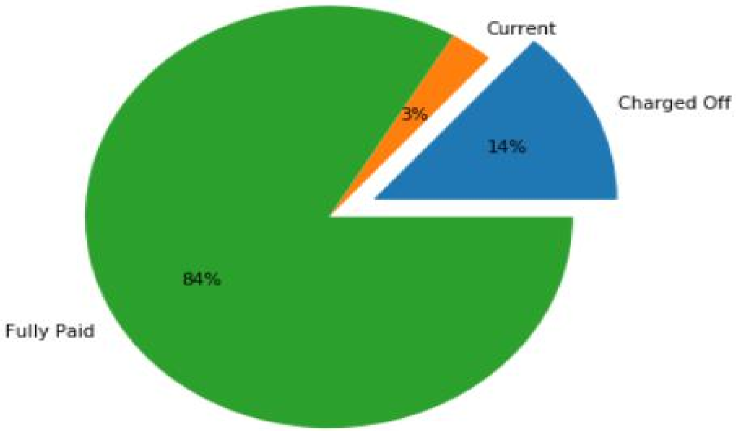
Company wants to understand the driving factors for loan default. Company can utilize this analysis for portfolio and risk assessment.

FlowChart :

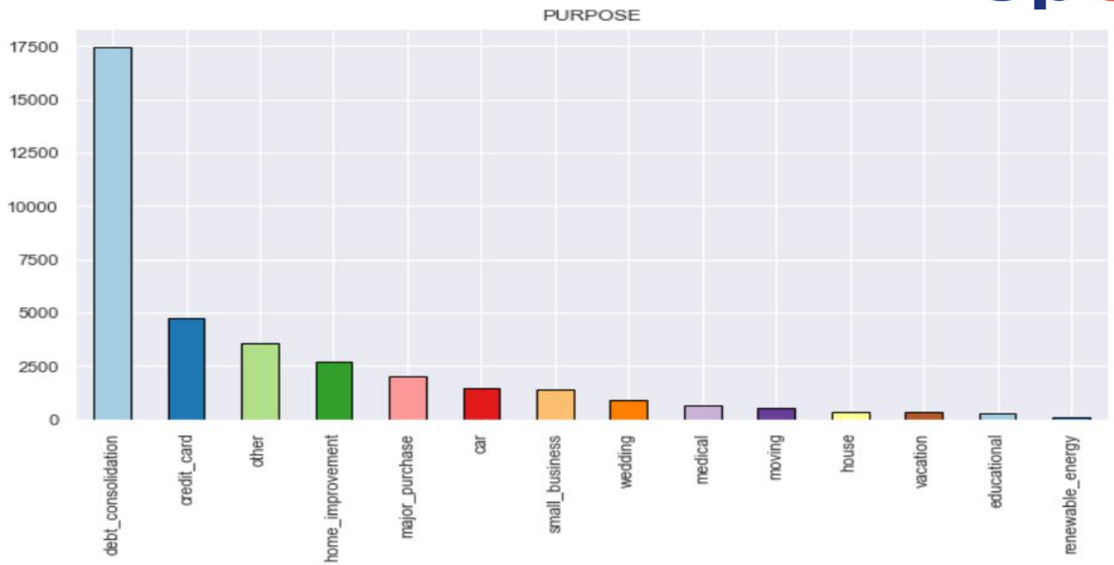

DATA SOURCING
(Analysis of public/private data)
DATA CLEANING
(To get rid of data quality issue)
UNIVARIATE ANALYIS
(Individiual analyis of variable)
SEGMENTED ANALYIS
(Univariate analyis on segment on data)
BIVARIATE ANALYIS
(Relationship analysis between two variables)
DERIVED METRICES
(Extract,Type,Business, data drven metrics)
CONCLUSION
Based on Analysis

Top Insights from Data

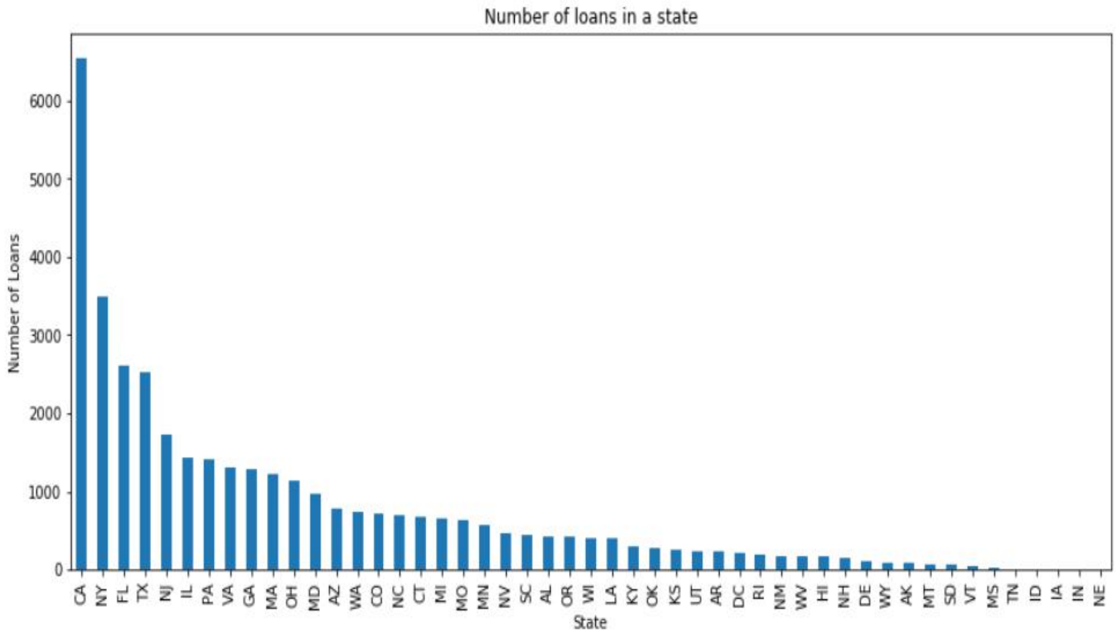
UpGrad



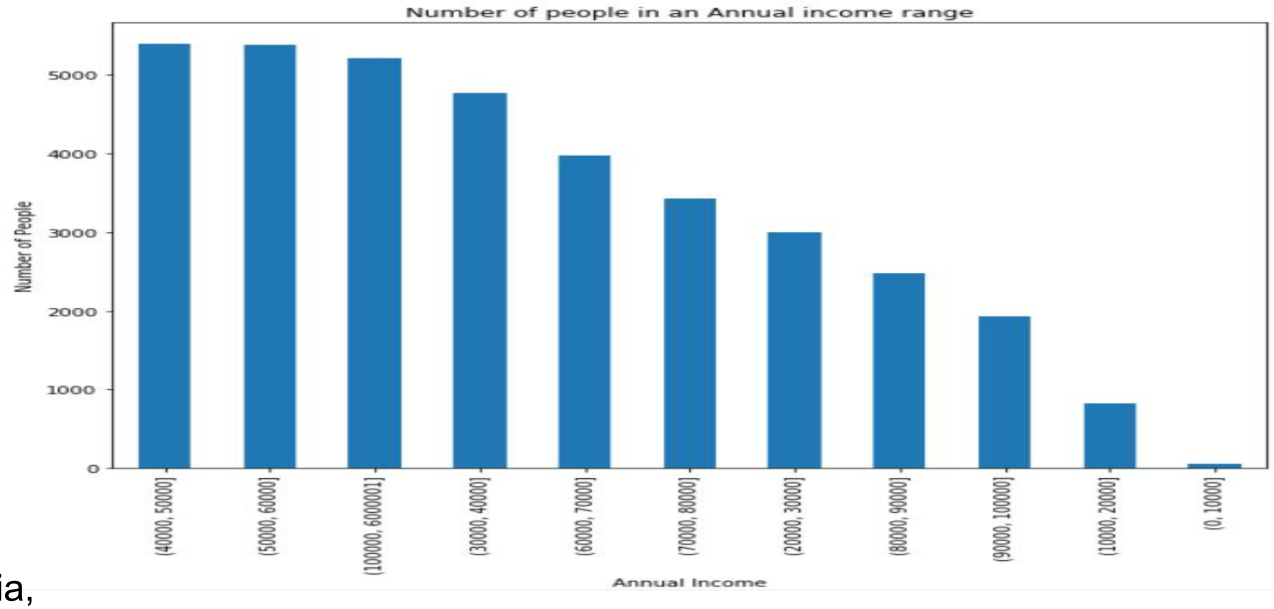
Overall Default rate is 14% in Lending Club



Most Popular loan products are debt_consolidation, credit_card, home_improvement

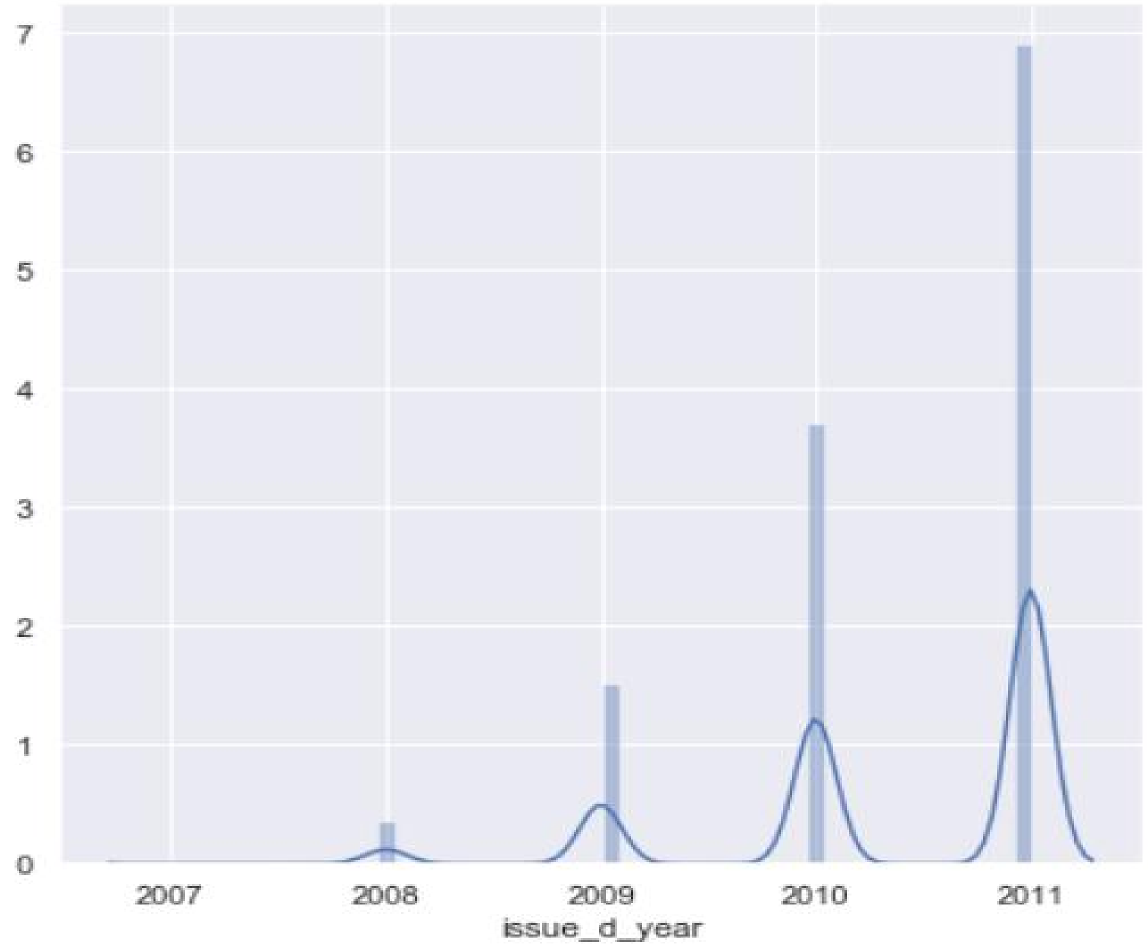


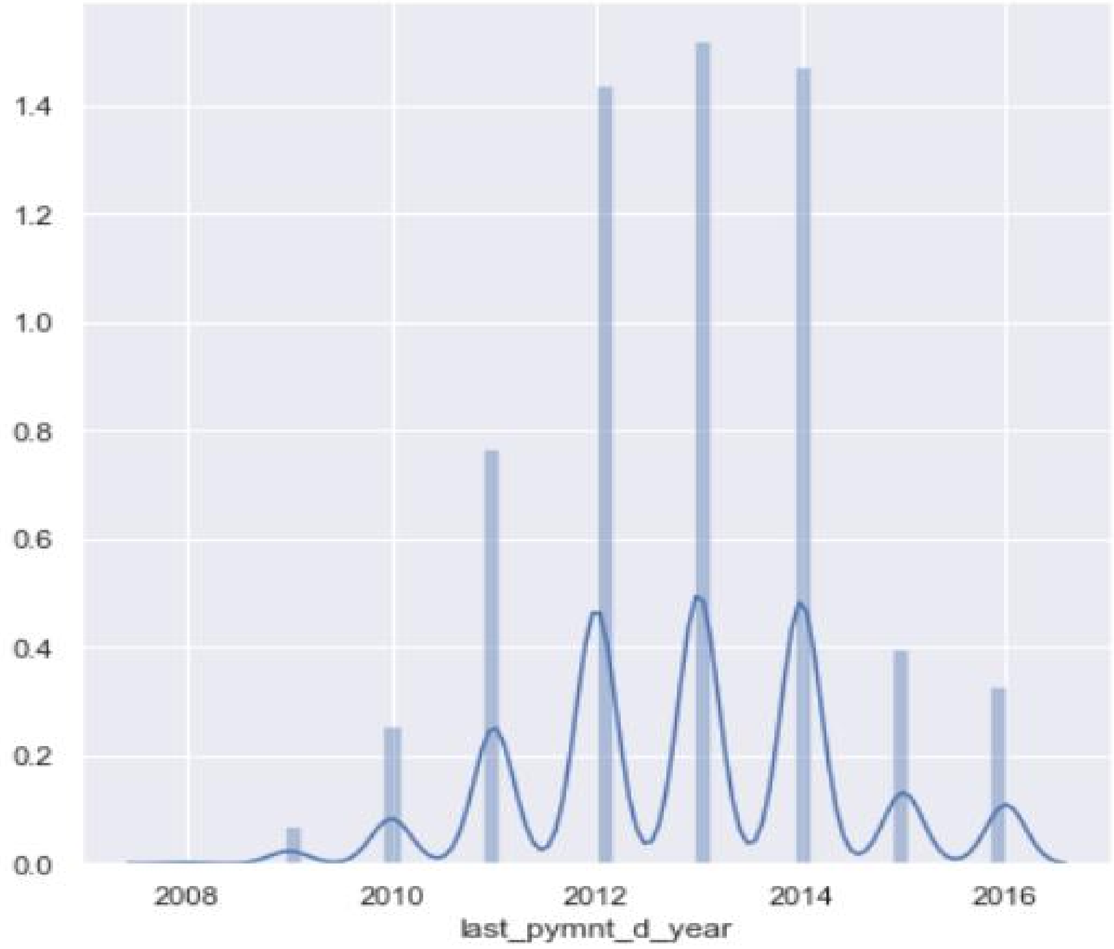
The most number of loans are distributed in big states like California, New York and Florida



The most number of loans are in range of 3000 to 15000 USD

Insights from Time frame of Data

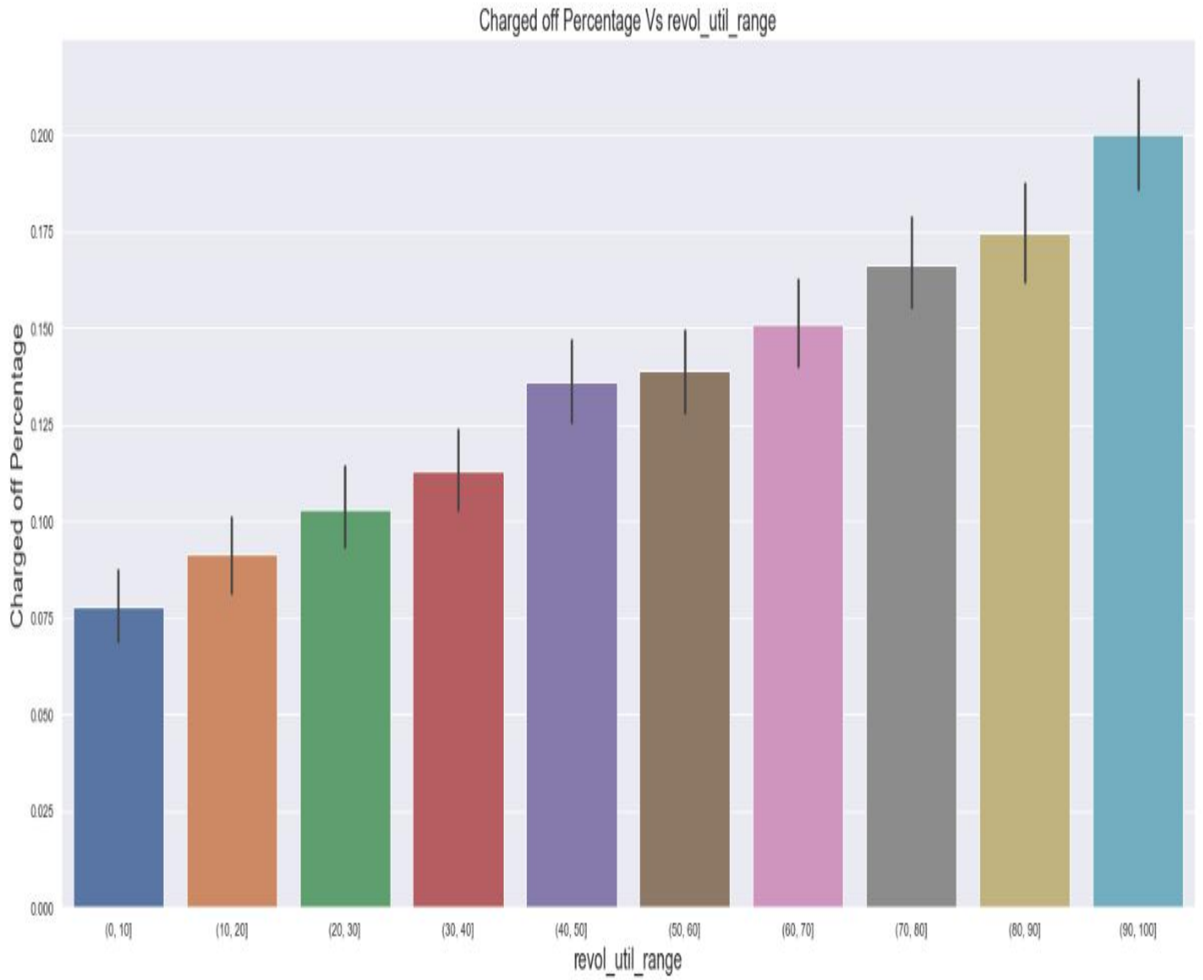




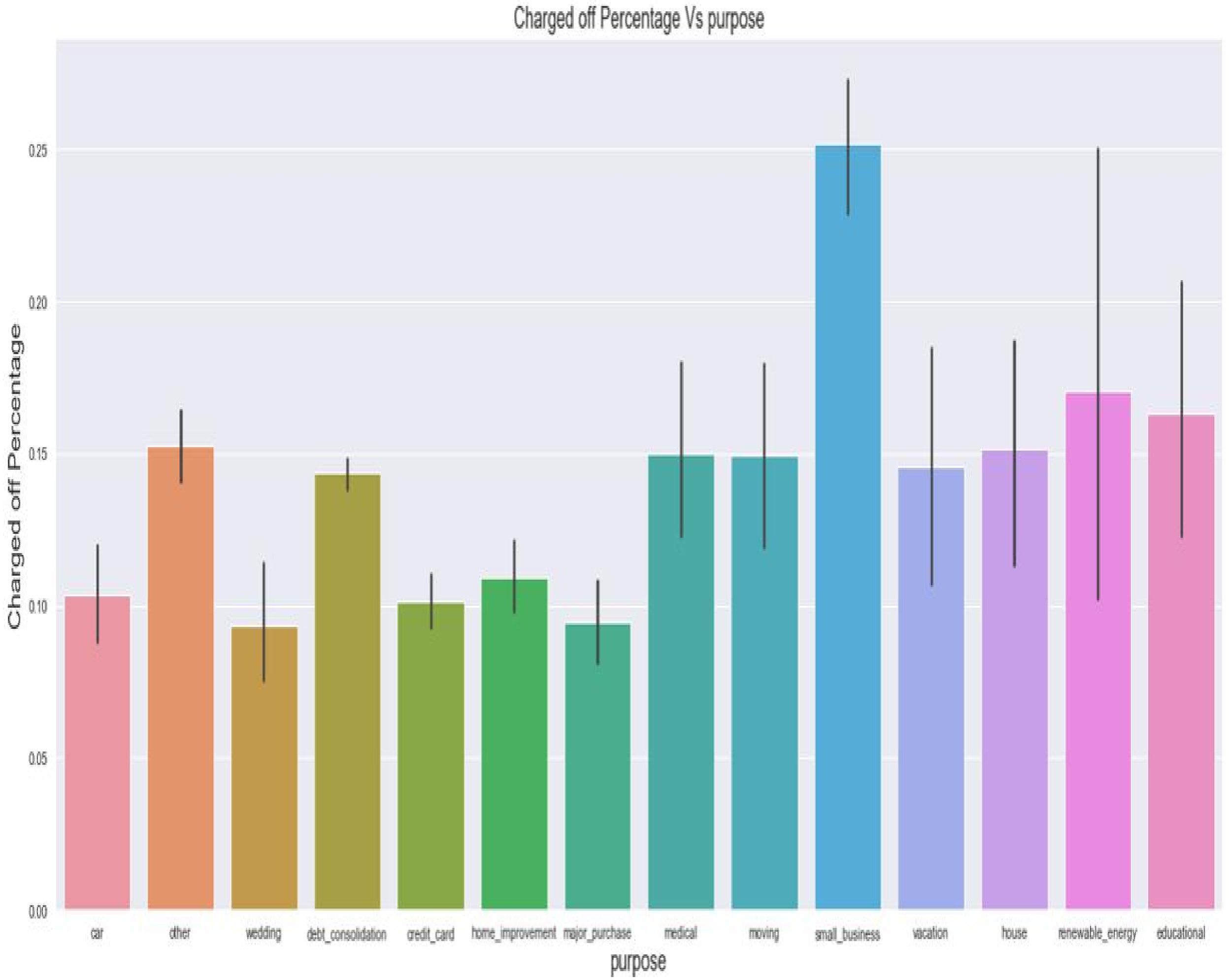
We see that there is a sudden spike in number of loans issued in year 2011, 2010 indicating recession

We also observe that there is sudden increase in loan defaulters in years 2010-2014 indicating an economic slowdown during these years

Loan Defaulters Study-1

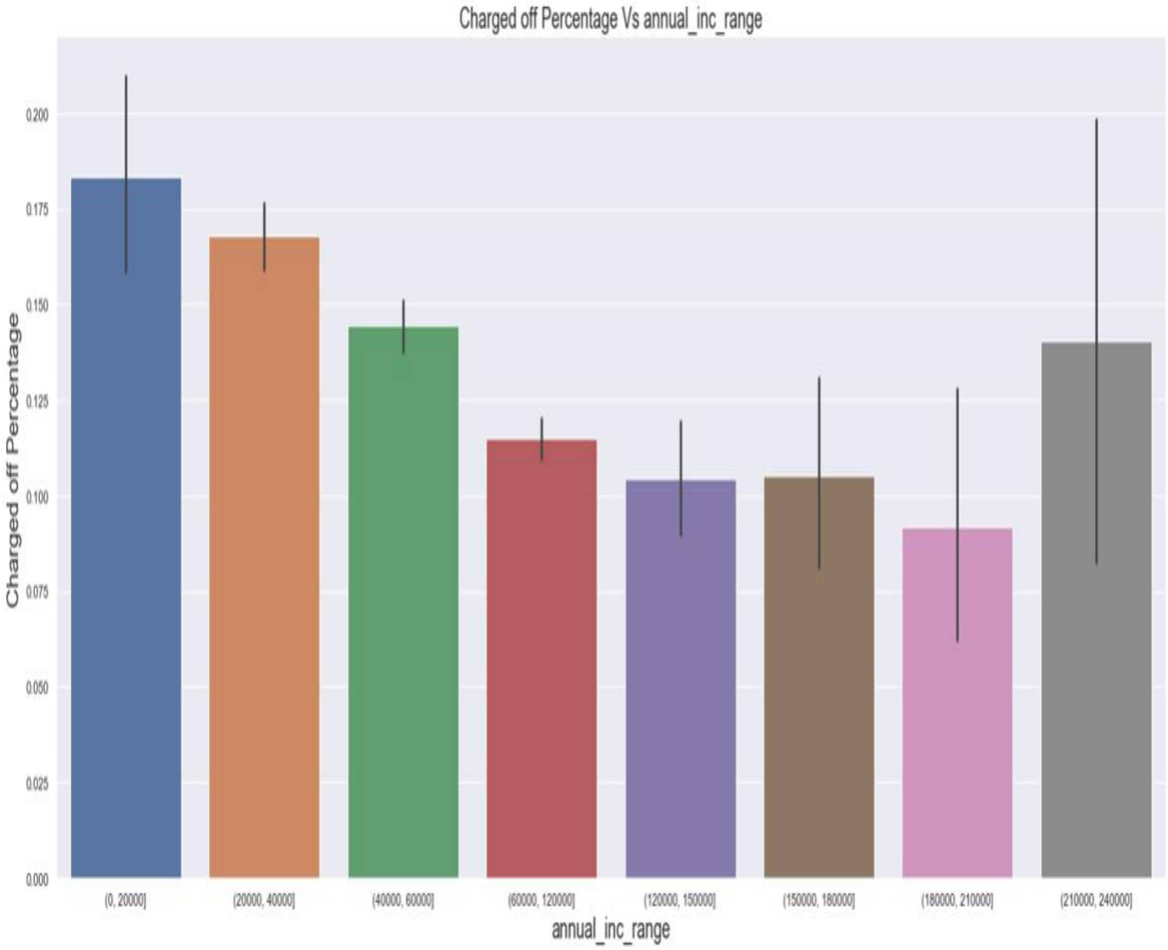


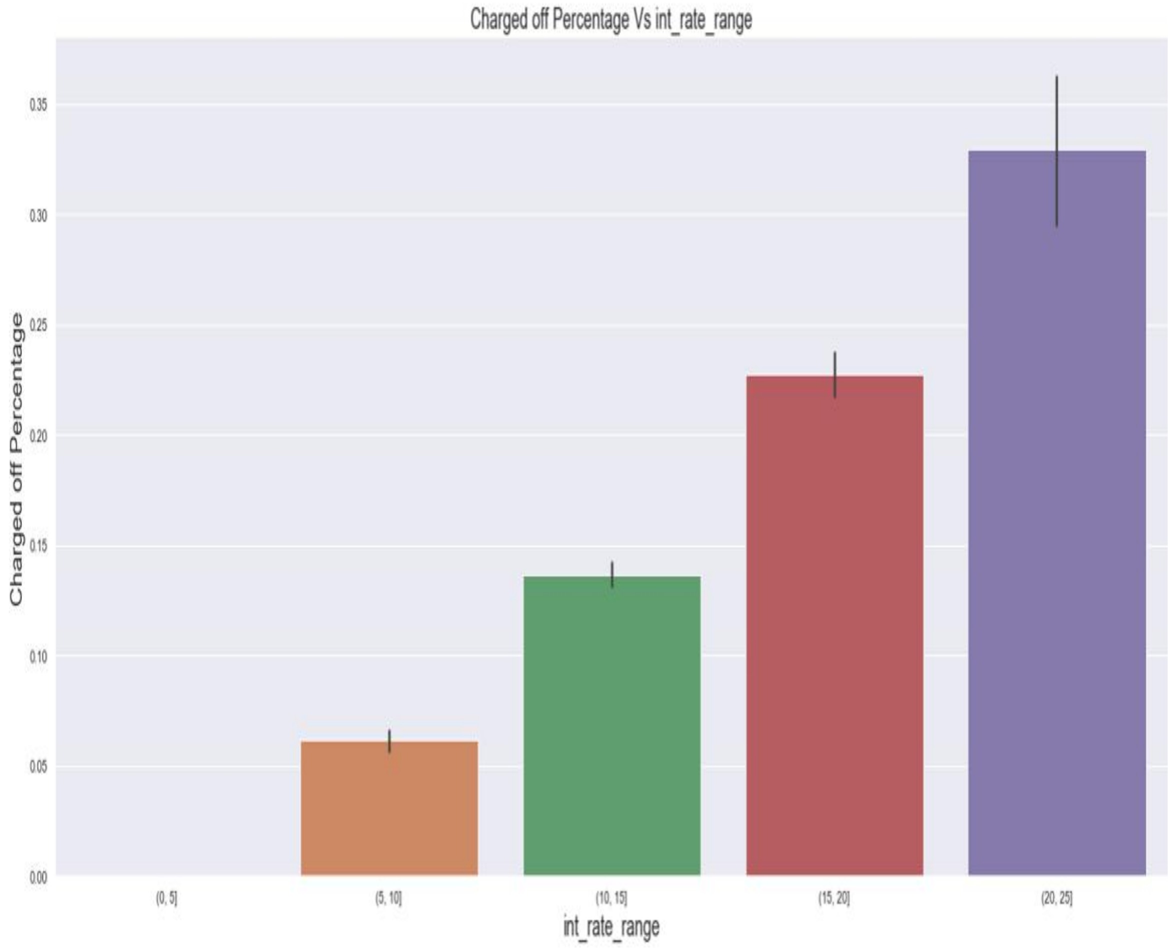
If a borrower has large revolving utilization, there are more chances that borrower will default on a loan



Small Business: 27%, Renewable energy: 17%,
House: 15%, Debt Consolidation: 14%, Credit card: 10%

Loan Defaulters Study-2

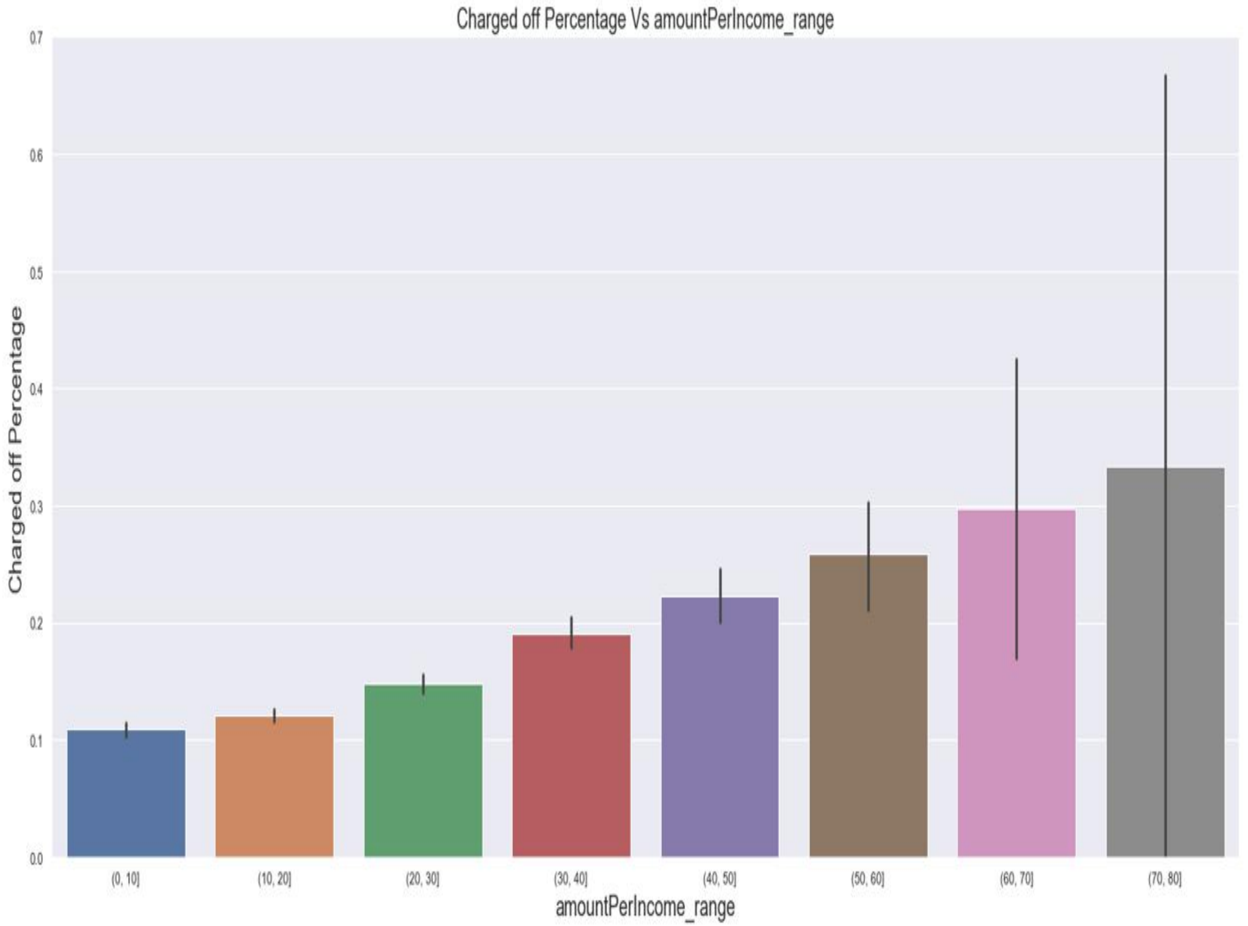


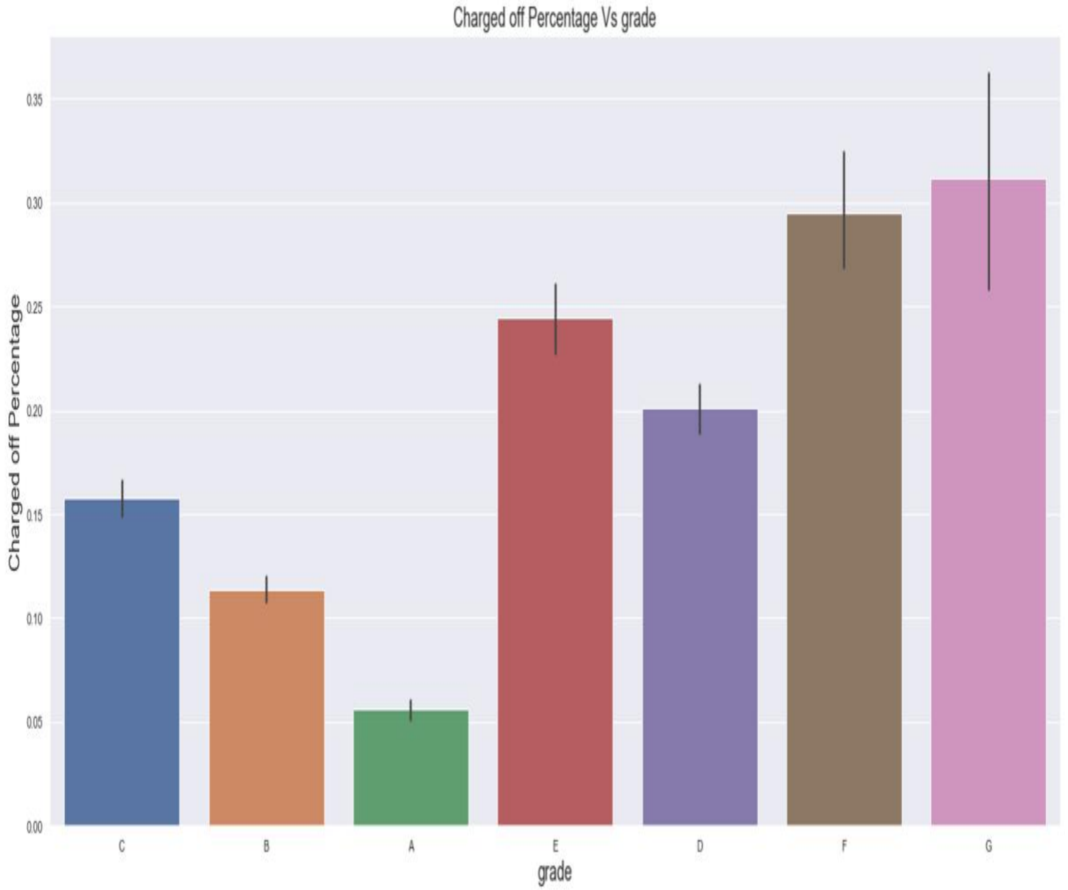


We observe higher default rate generally for lower range of salaries.

We observe higher default percentage for higher rate of interests.

Loan Defaulters Study-3





We observe if loan amount is more than 30% of the annual income of a borrower, the chances of borrower defaulting on loan is 20-30%

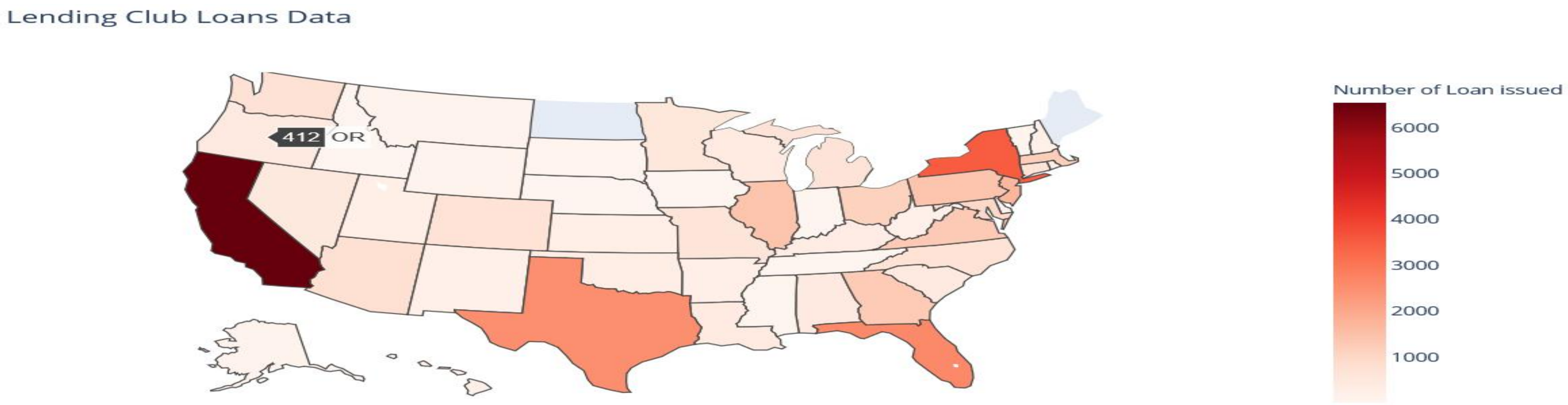
We observe higher default percentage in G,F,E,D respectively

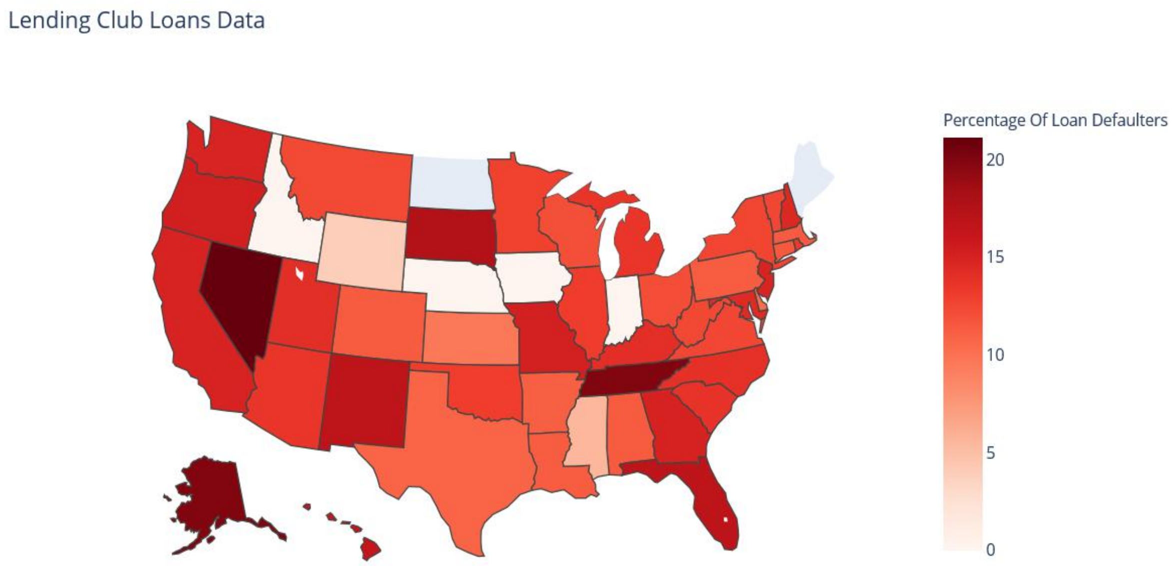
Loan Defaulters Study-4

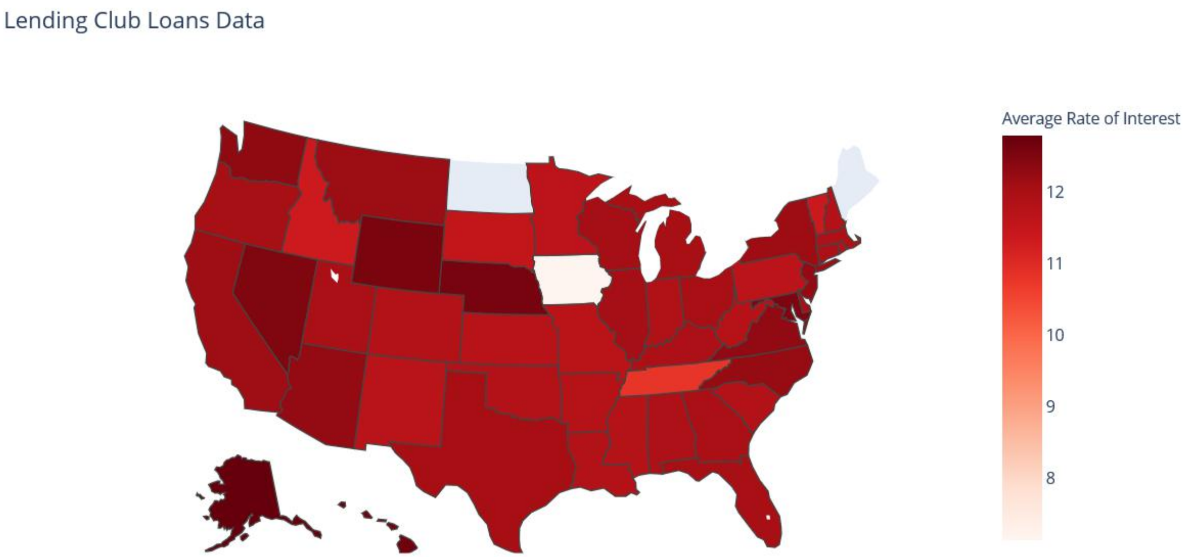
We observe that there is not much impact of verification on Loan Defaulters which raises some questions. Ideally there should be less number of defaults in cases where proper verification is done. Hence there can be 2 hypothesis:

1) Verification status is done correctly and all people could not replay the loan due to economic slowdown
2) Source of income was not verified properly and this may highlight the possibilities of corruption in Loan Department

USA-State wise Heat Map







iiit-b UpGrad

Conclusions

On the basis of the analysis, below are our conclusions for Lending Club Case Study:

Top Leading states in Number of Loan Applications	Top Leading states in Percentage of Defaults
California(6536)	Nevada(21.09%)
New York(3494)	Alaska(20%)
Florida(2616)	Tennessee(20)
Texas(2516)	South Dakota(17.54)

Top Loan Default Observations	Suggestions
Loan application for Small Business	Do not give large amount of loans for Small Business Loan Applications or charge less interest rate
More than 30% of Amount to Income Percentage	Do not Approve loan amounts which exceeds 30% of the annual income of borrower
More than 50% of Revolving Utilization	Charge higher interest rate on Loan accounts with higher utilization rate
Loan Application for Grade D,E,F,G	Higher Rate of Interest for these category of Loans

Recommendations:

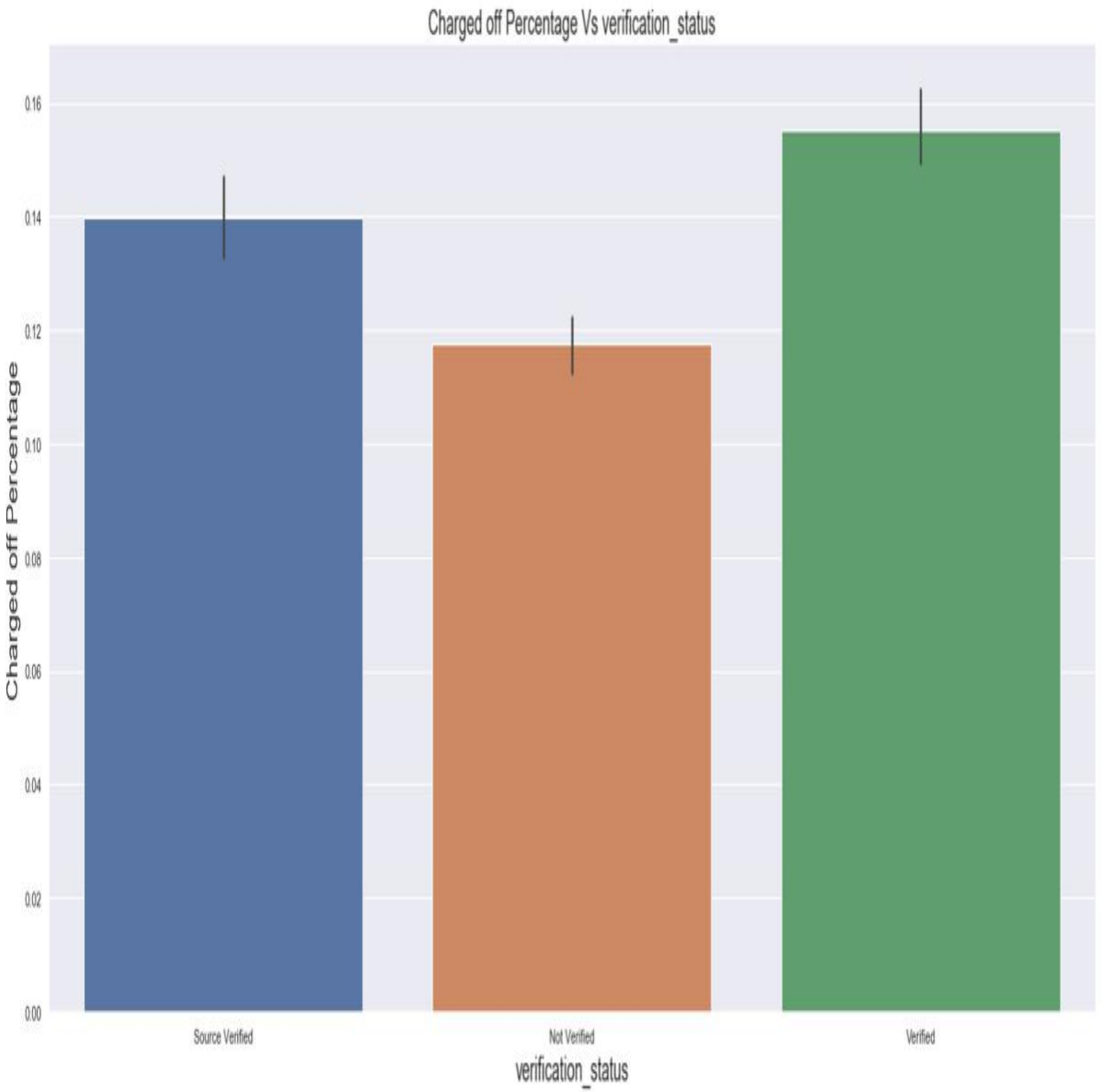
1) Stop approving loans where amount/income is higher than 30%
2) Reduce the number of approvals where purpose is small business
3) Stop approving loans if revolving utilization is greater than 75%
4) Stop approving loans to people with prior bad records
5) Do strict verifications in states leading in Default percentage score

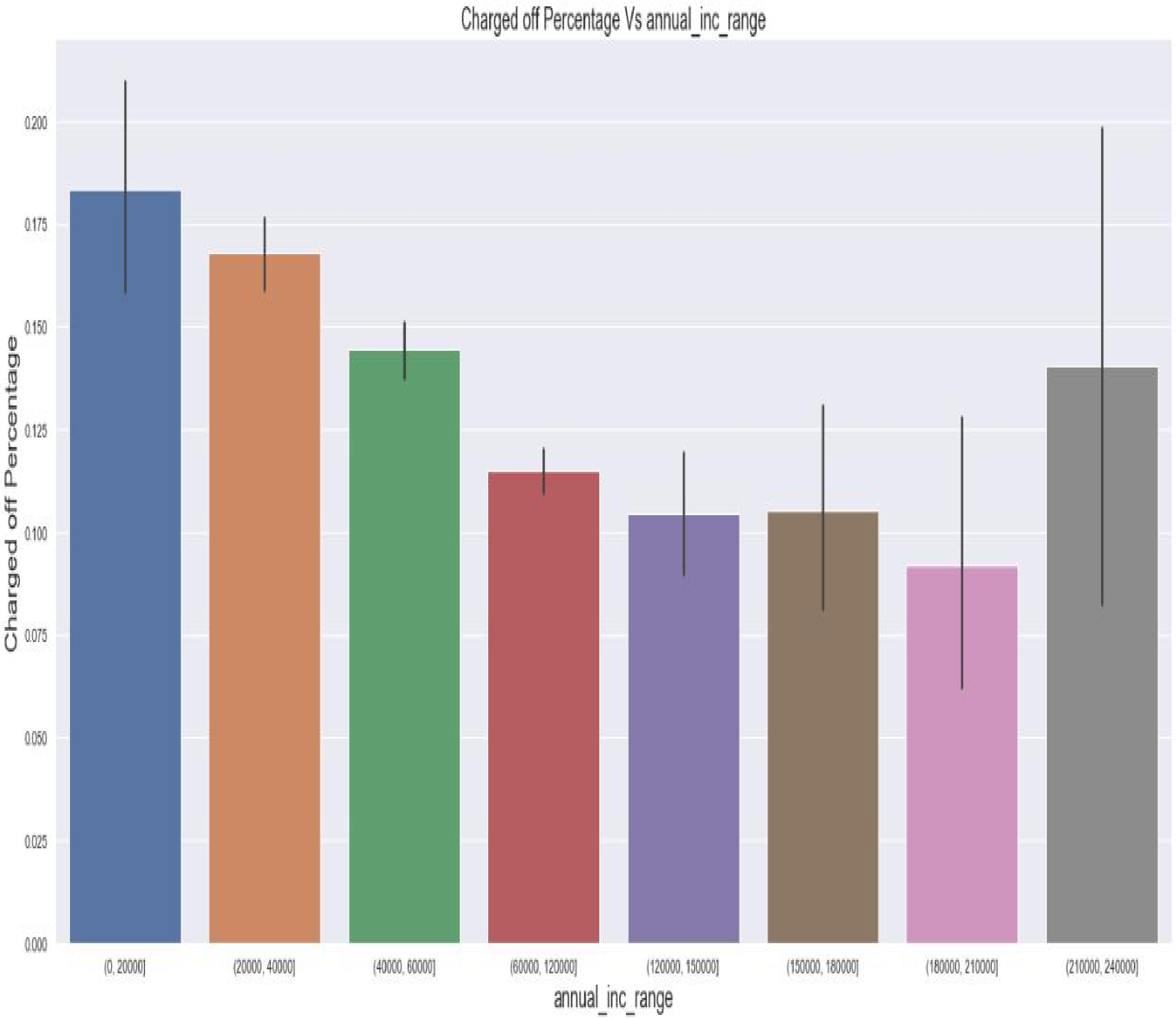
Open Questions:

1) Why People with higher annual income range also tend to default on loans?
2) Why there is no impact of verification on default rate? Is there a possibility of corruption?

Refer Appendix for plots

Appendix-1


Charged off Percentage Vs verification_status
Charged off Percentage
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
Source Verified
Not Verified
Verified
verification_status


Charged off Percentage Vs annual_inc_range
Charged off Percentage
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
(0, 20000]
(20000, 40000]
(40000, 60000]
(60000, 120000]
(120000, 150000]
(150000, 180000]
(180000, 210000]
(210000, 240000]
annual_inc_range