EXPLORATORY DATA ANALYSIS

SUBMISSION

EMAIL- ID

GROUP FACILITATOR ⇒ GAURAV SINGH   gauravsingh13091990@gmail.com

GROUP MEMBERS ⇒   1)VIPIN KUMAR   vipinkumar535@gmail.com

2) SUDARSHAN RANGANATH SURVEPALLI   ssurvepalli@gmail.com

Objectives

Company provide online loans to customer at low interest rate. It provides different kinds of loan but lending loan to risky applicant may result in to huge financial loss to the company.The objective of the company is to understand various driving factors behind loan, i.e the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.
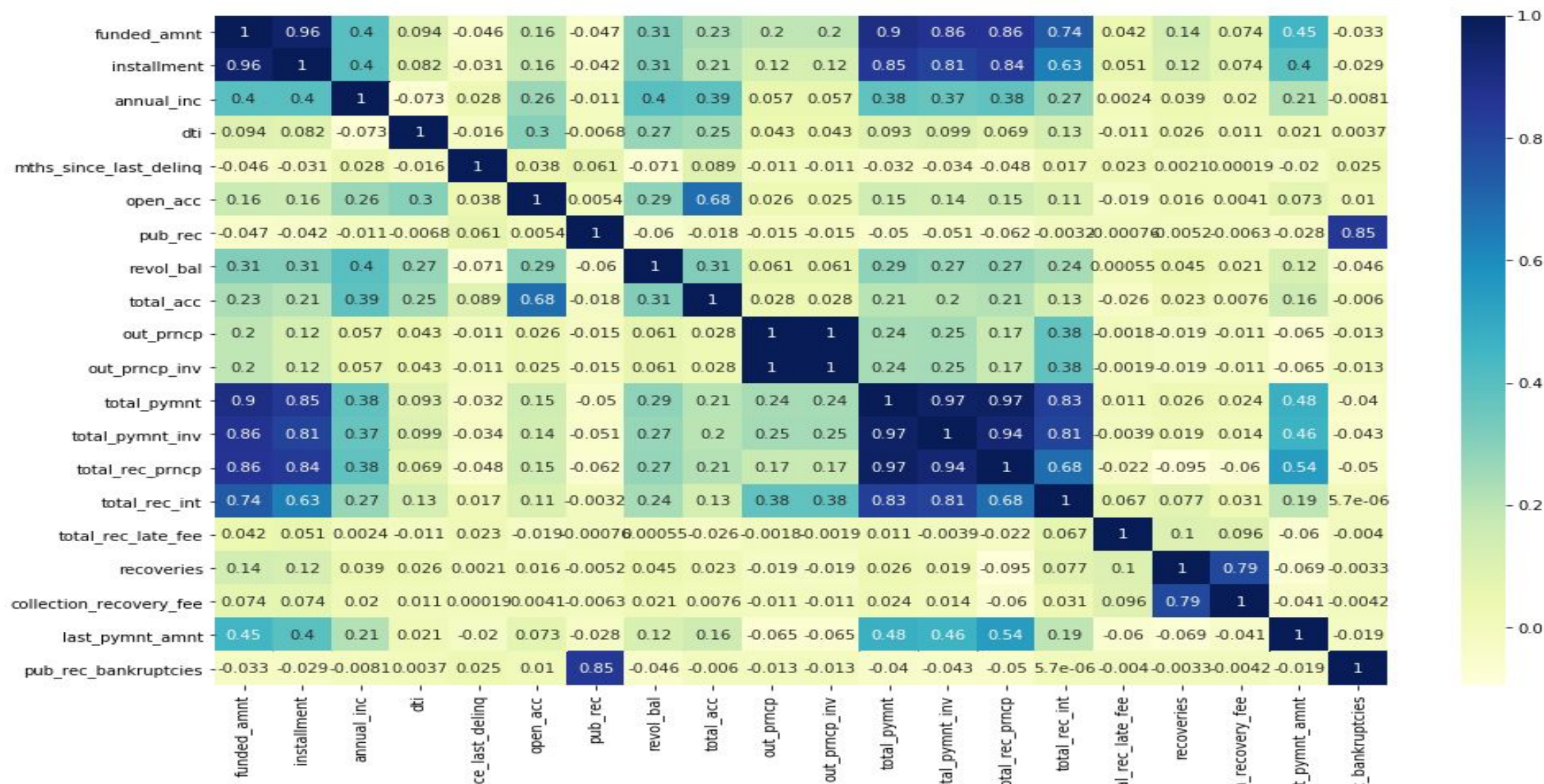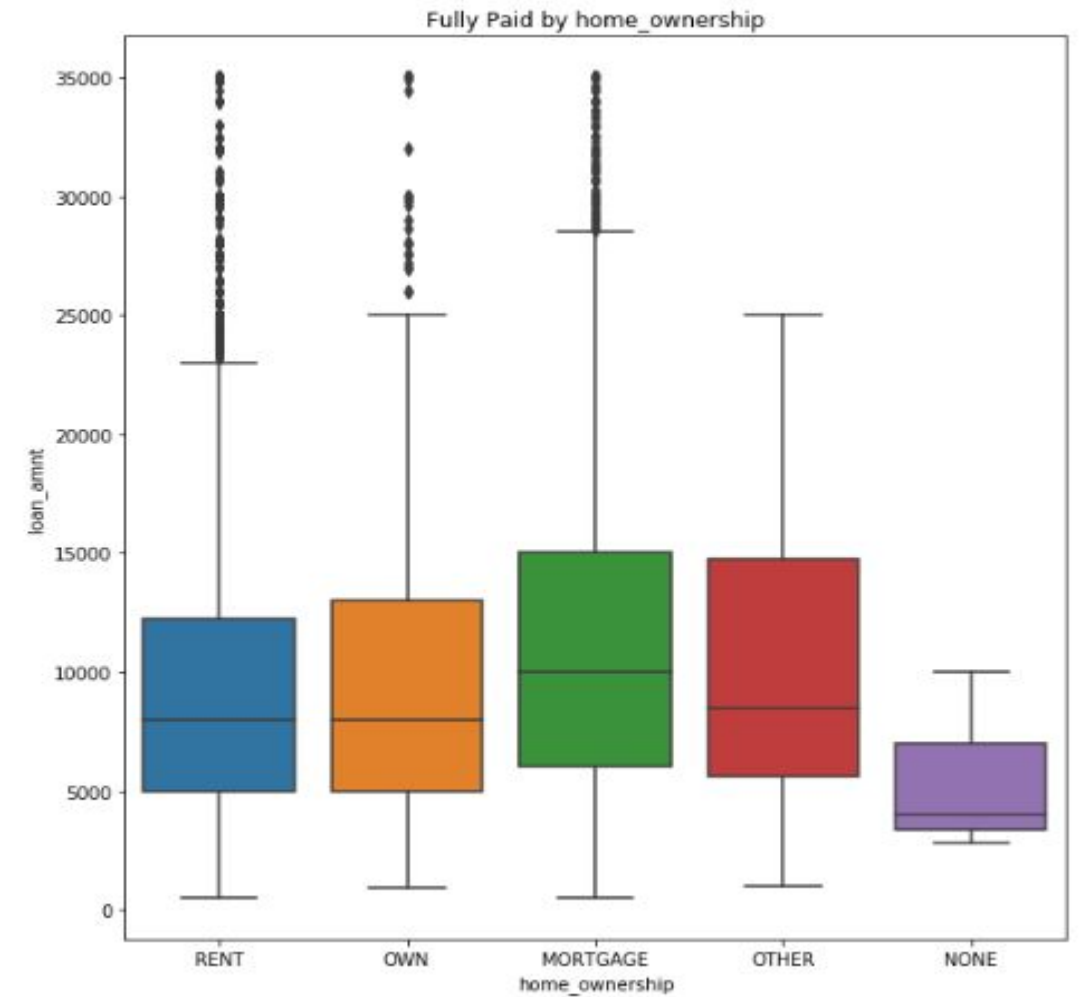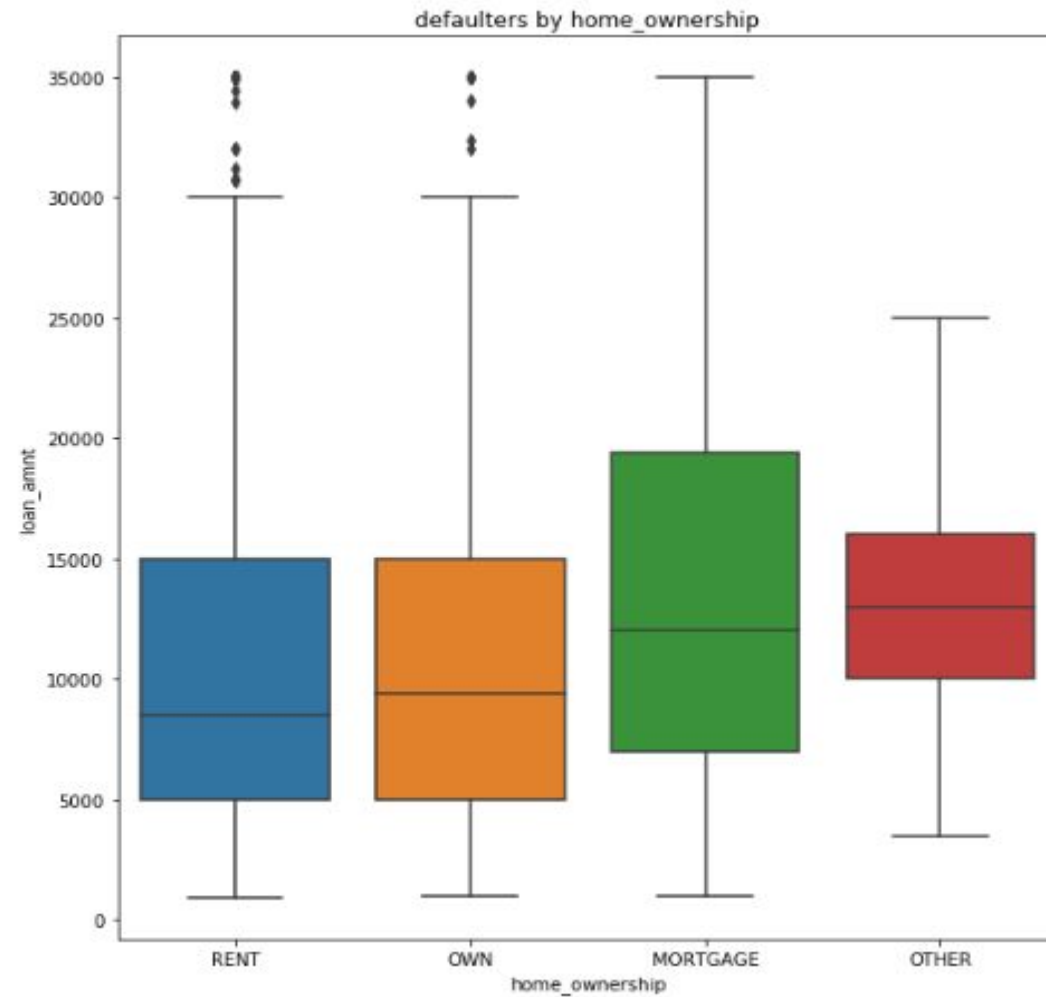
Methodology
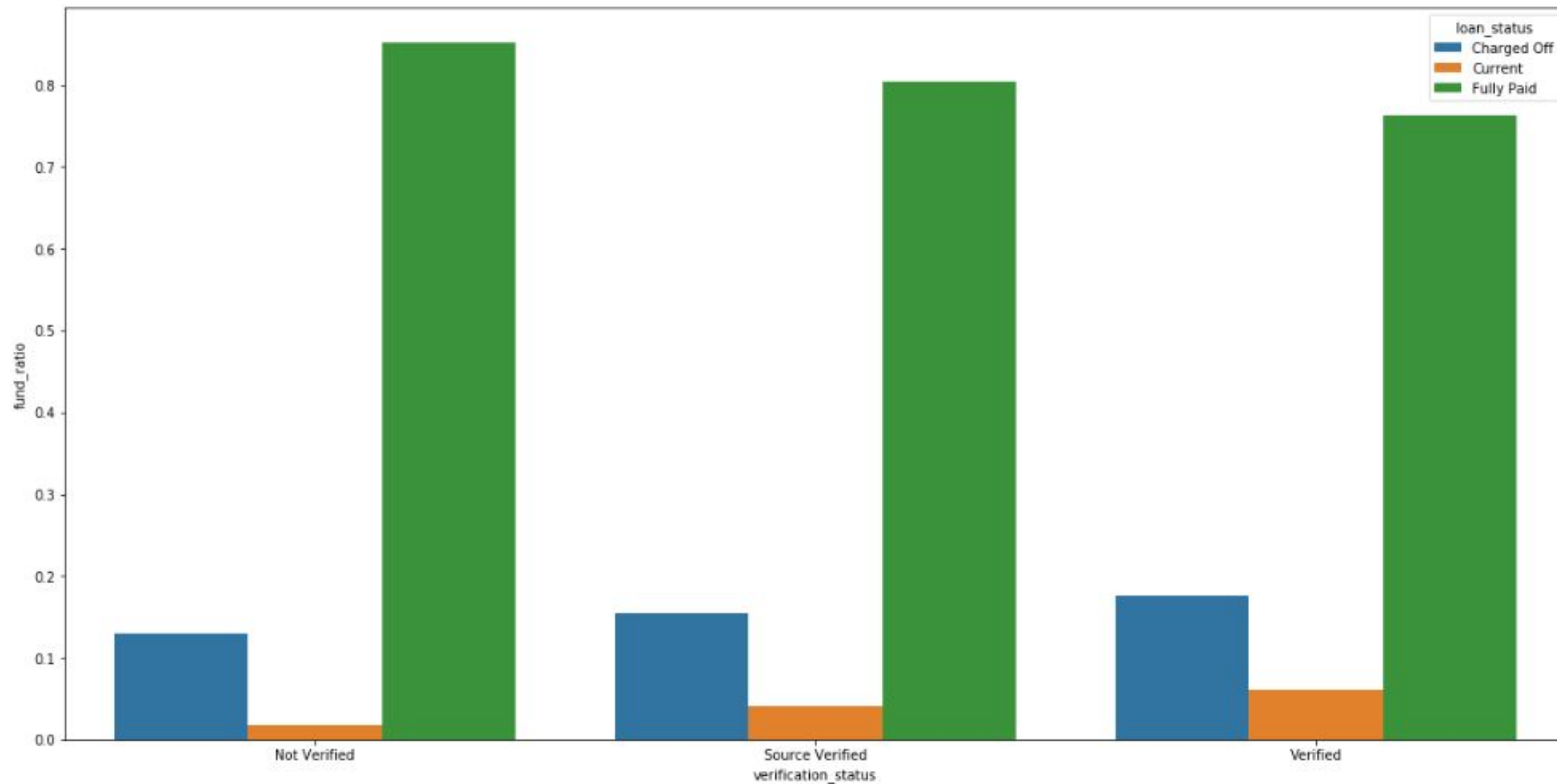
Use flow chart
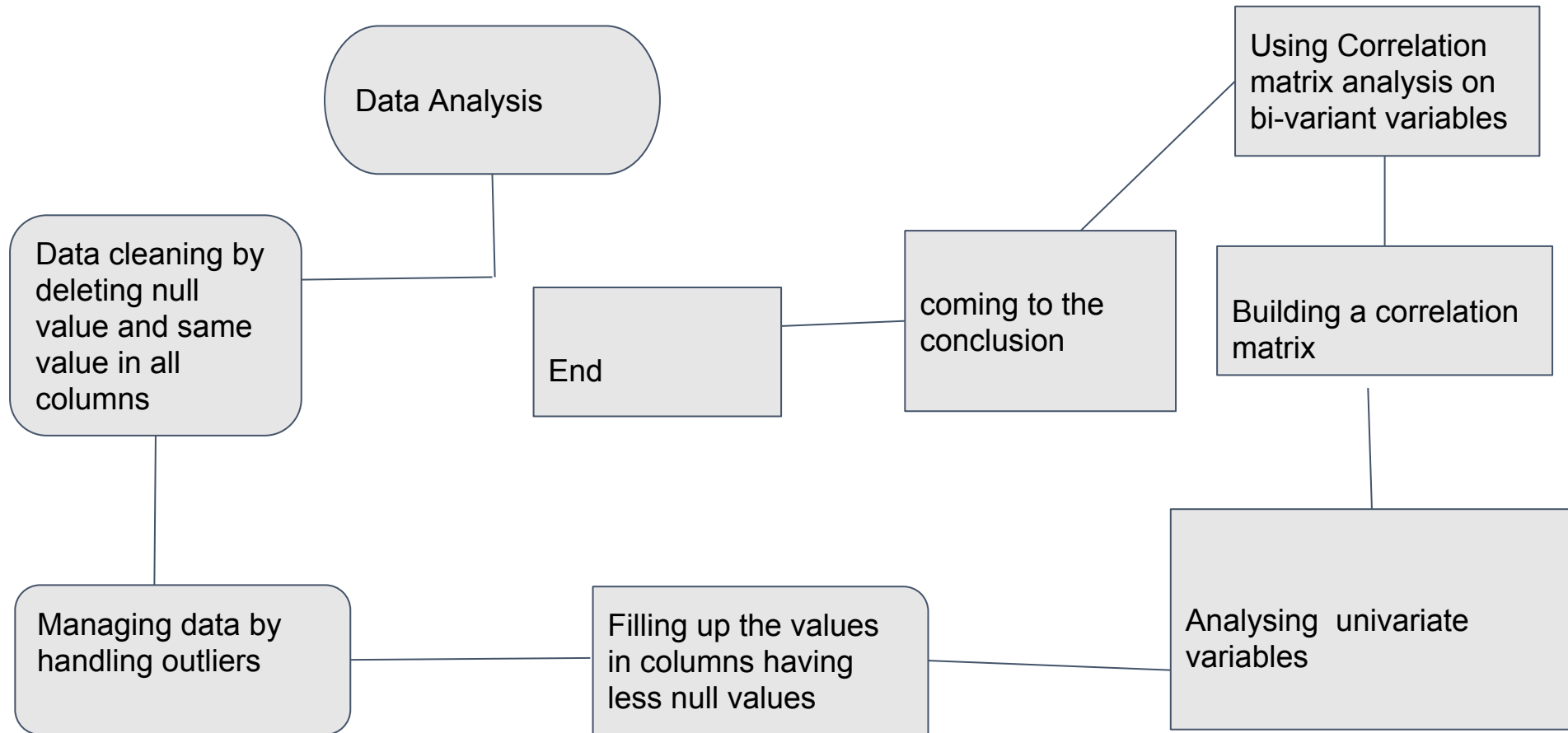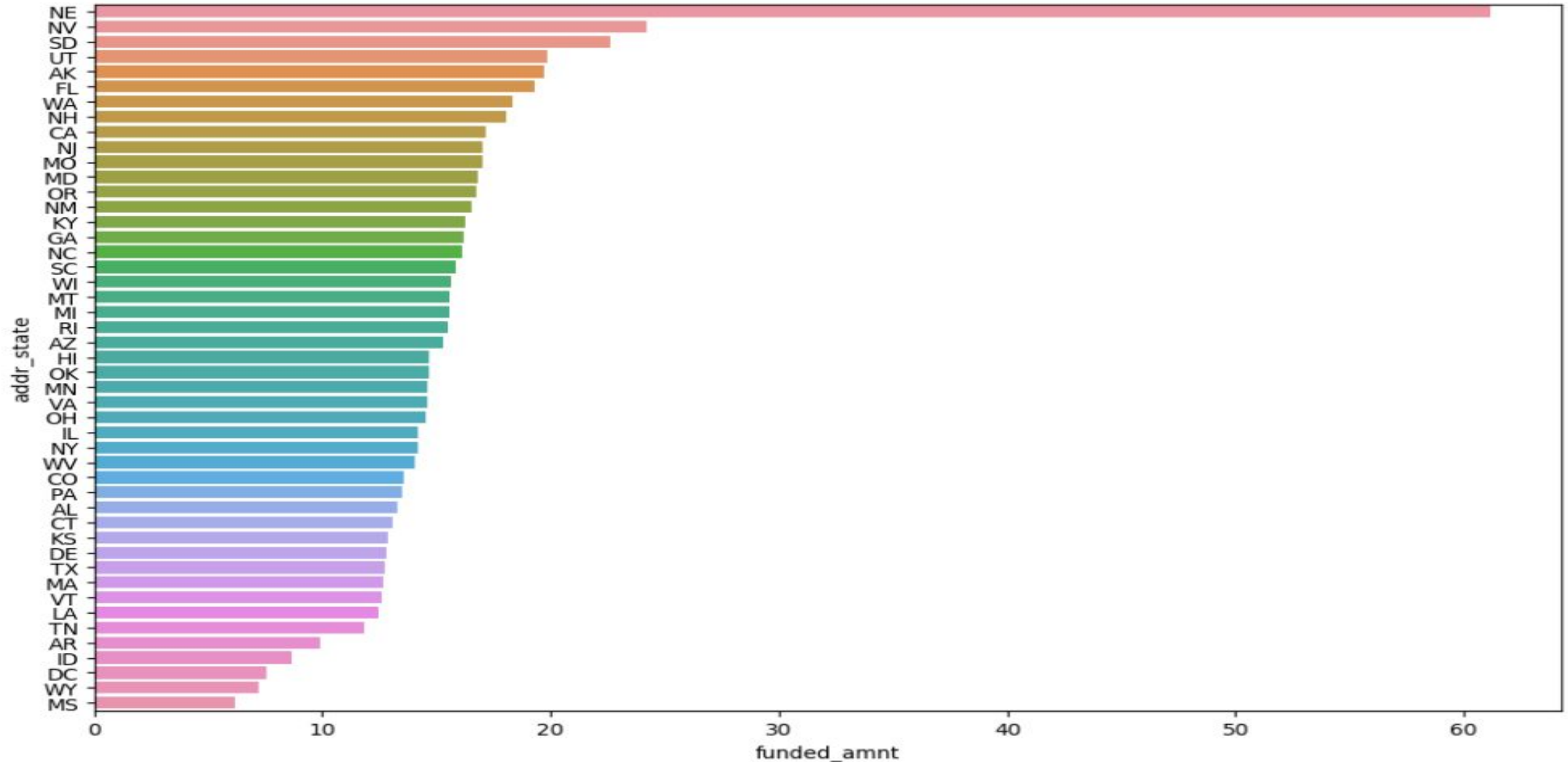
Bi-Variant Pair plot
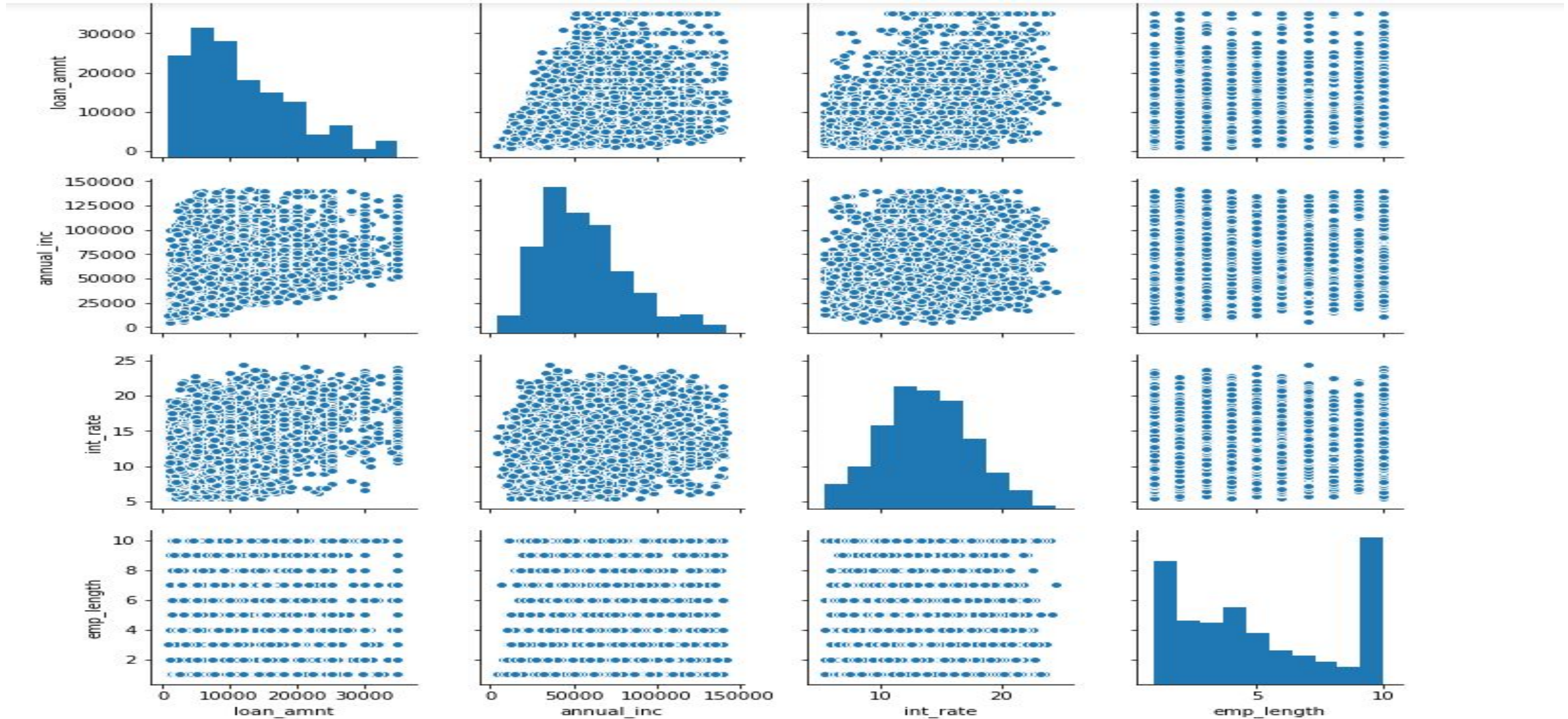
Correlation Heat-Map

Categorical Analysis

Verfied status analysis

State wise analysis

Co-relation metrice

	loan_amnt	annual_inc	int_rate	emp_length	installment	dti	term	pub_rec_bankruptcies
loan_amnt	1	0.46	0.34	0.19	0.93	0.078	0.39	-0.02
annual_inc	0.46	1	0.15	0.21	0.46	-0.029	0.13	0.03
int_rate	0.34	0.15	1	0.041	0.32	0.042	0.46	0.08
emp_length	0.19	0.21	0.041	1	0.17	0.062	0.14	0.051
installment	0.93	0.46	0.32	0.17	1	0.054	0.14	-0.018
dti	0.078	-0.029	0.042	0.062	0.054	1	0.066	0.018
term	0.39	0.13	0.46	0.14	0.14	0.066	1	0.02
pub_rec_bankruptcies	-0.02	0.03	0.08	0.051	-0.018	0.018	0.02	1

Conclusion

Applicants who never defaulted ⇒ 'None' category candidates under home ownership never defaulted

Most influencing factors for default ⇒

1) Verification status as verified have highest number of defaults.
2) State code NE have highest probability of loan default
3) Small business with high loan amount have are more likely to default
4) In case of derogatory public record available there is high probability of loan default
5) Applicants in F grade with high loan amount are more likely to default
6) 25% of people contribute to 50% of losses due to Charge-offs
7) Customers with High DTI rate have higher probability of default