Data introducton

Our dataset:

- 2200 rows
- 14 columns
- https://www.kaggle.com/prashanthsri12/bank-loan-interest-rate-dataset

Target

- Interest Rates

Numerical

- ID
- Amount Requested
- Amount Funded By Investors
- Loan Length
- Debt To Income Ratio
- Monthly Income
- Inquiries in the Last 6 Months
- FICO
- Open CREDIT Lines
- Revolving CREDIT Balance

Categorical

- Loan Purpose
- State
- Home Ownership

Pre-processing

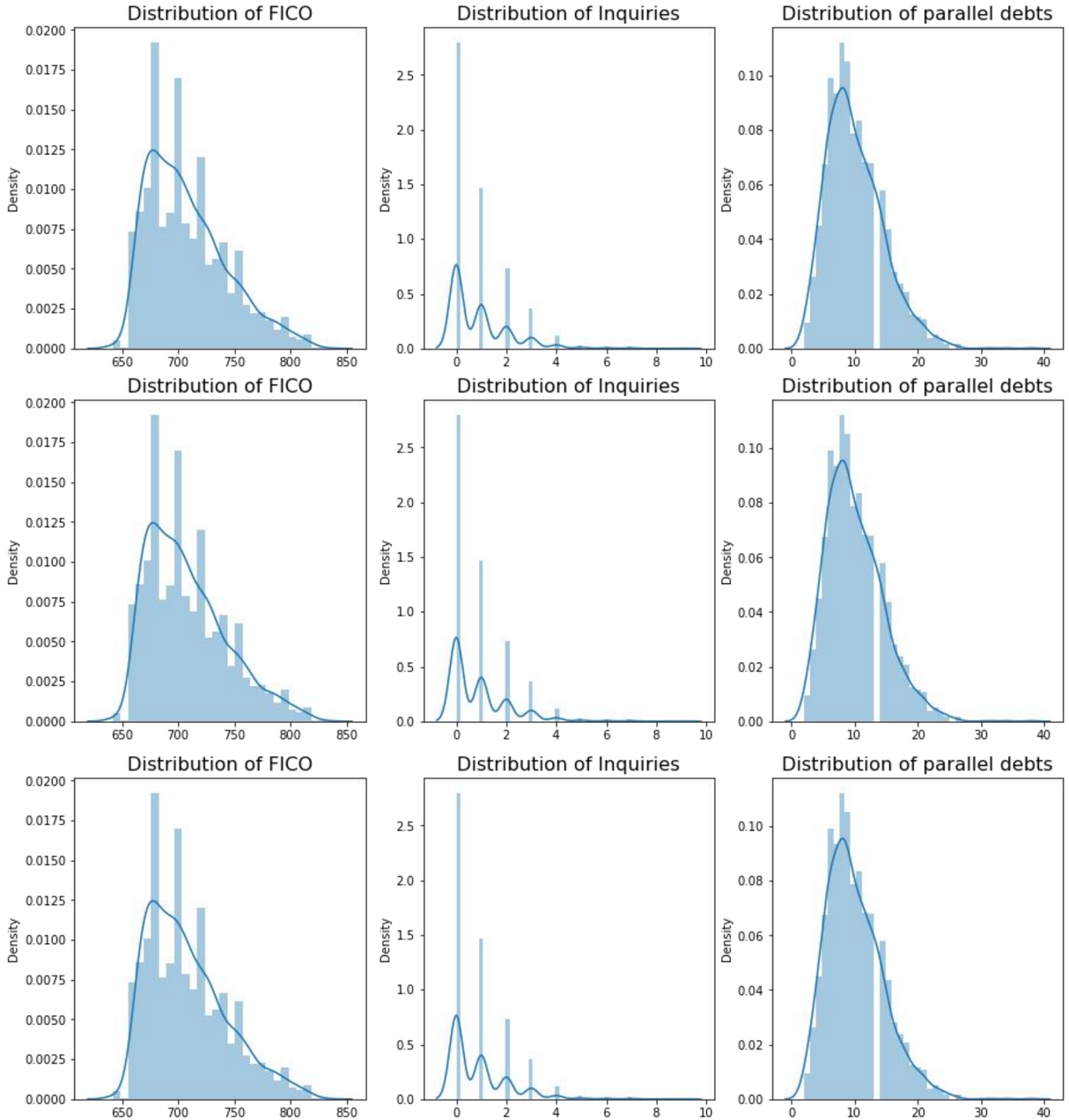
- **Omit N/A values** （14 rows）

- **Convert data formats**

  Convert characters of the data frame to numeric.

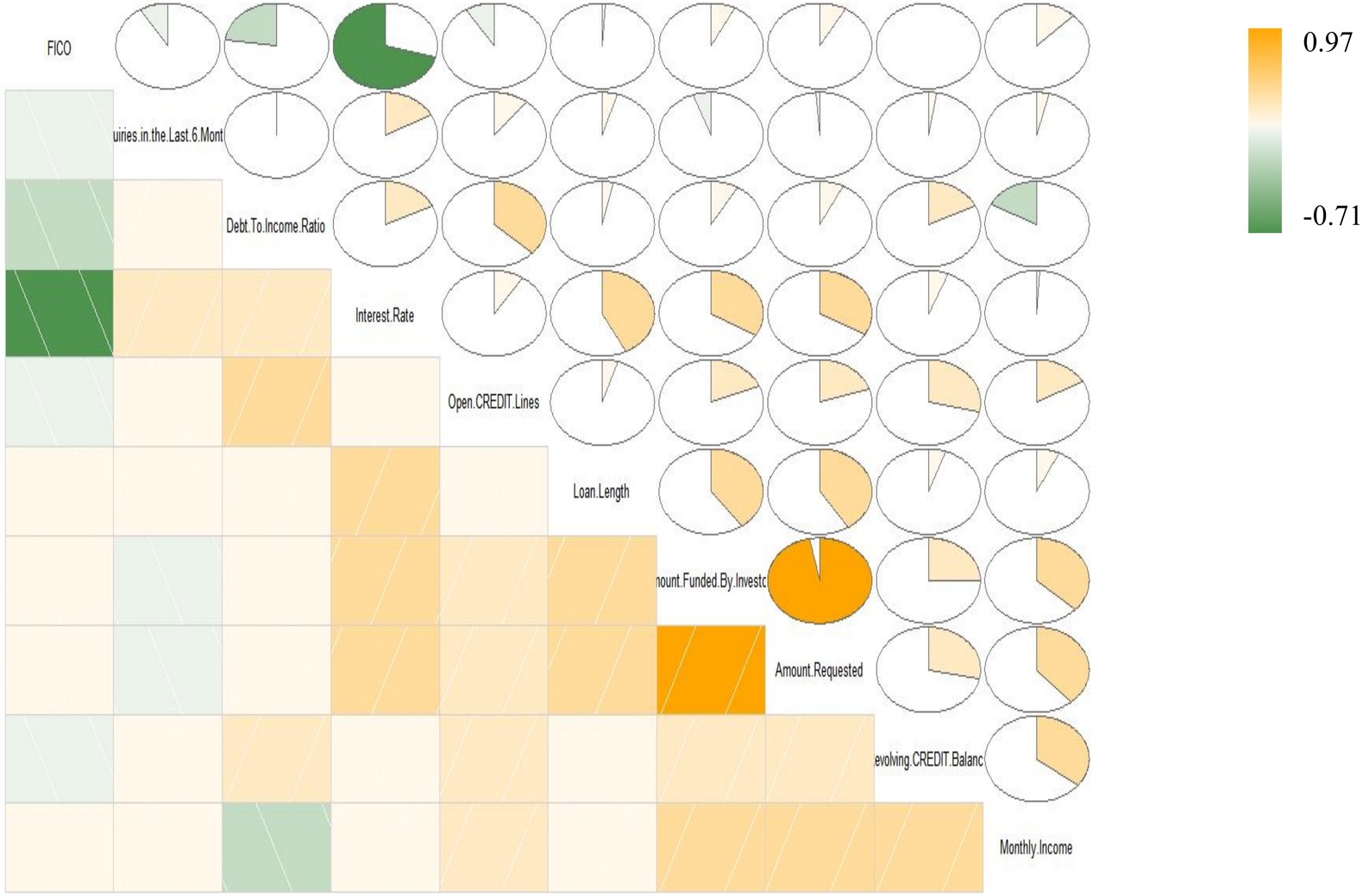
  Covert category variables into dummy variables

- **Standardization**

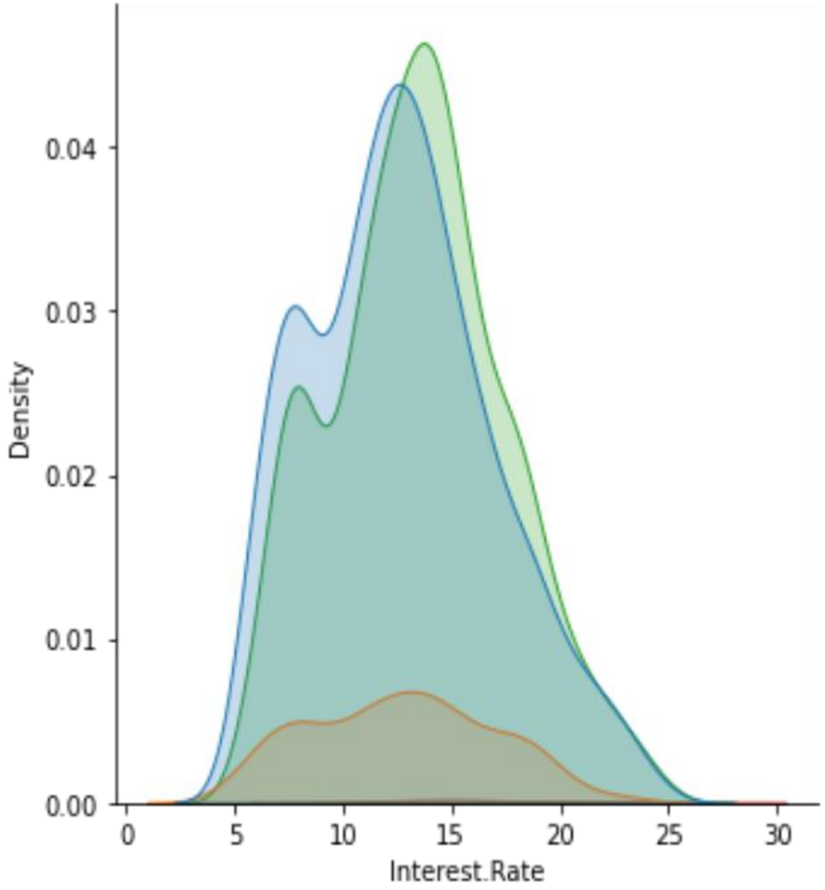


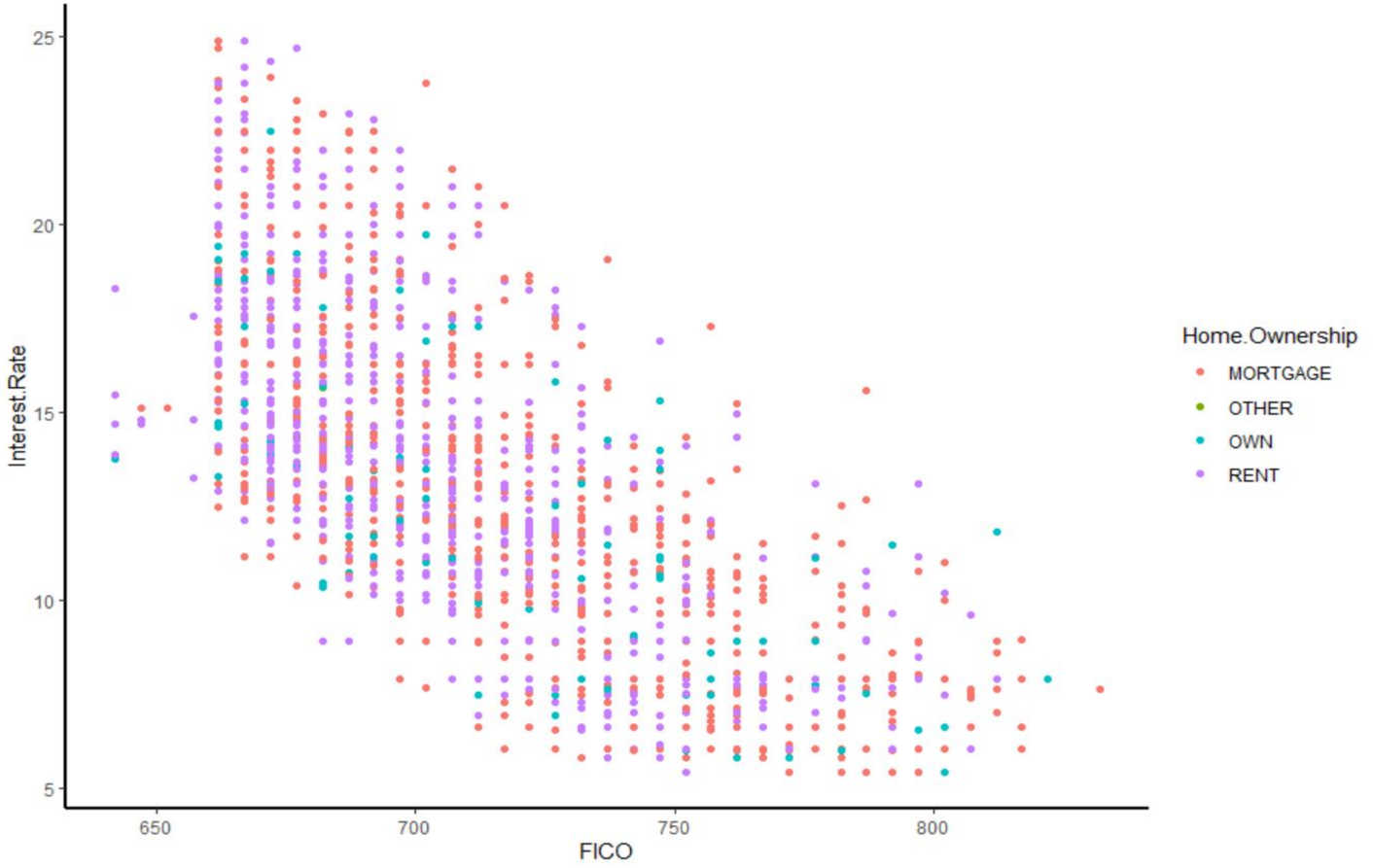
EDA

Correlation Matrix for all numeric variables



EDA

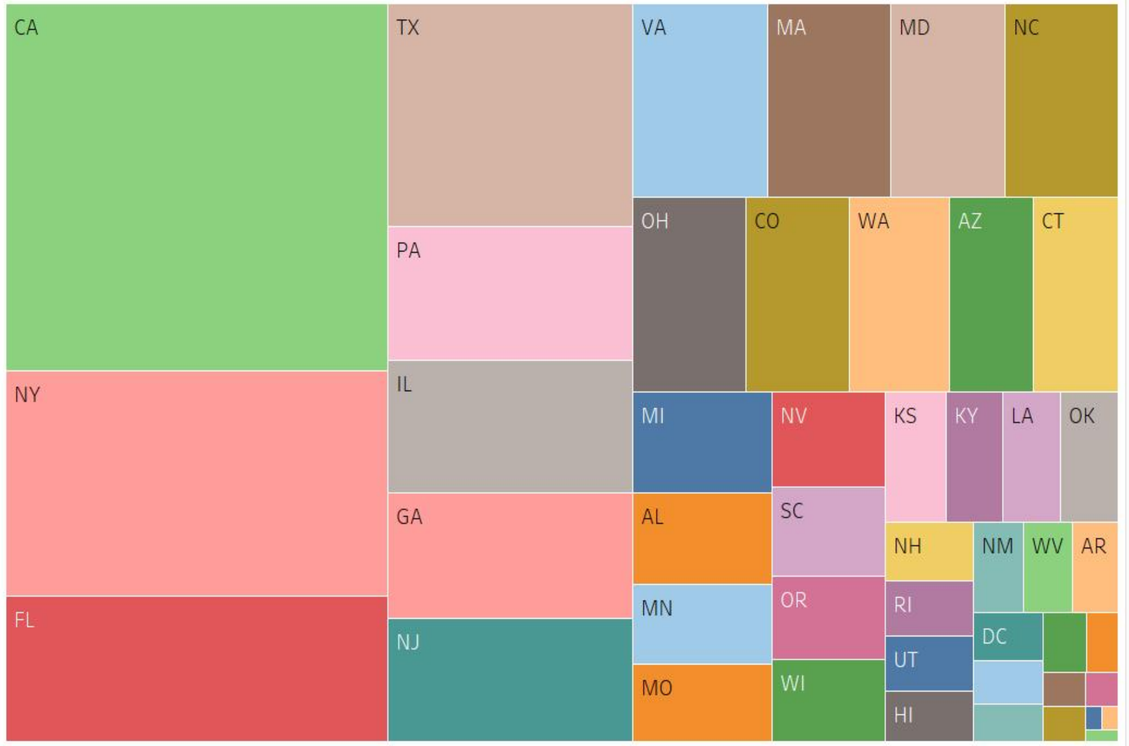

0.04
0.03
0.02
0.01
0.00
Density
0
5
10
15
20
25
30
Interest.Rate
Home Ownership
MORTGAGE
OWN
RENT
OTHER


25
20
15
10
5
Interest.Rate
650
700
750
800
FICO
Home.Ownership
MORTGAGE
OTHER
OWN
RENT

EDA

Difference =
Amount Requested -
Amount Funded by
Investors

EDA



Average interest rate per state

Average monthly income per state