



Exploratory Data Analysis of Credit Risk Data

Pramisha Thapaliya
Md Mominul Islam
Sakib Faisal
Tahmid Alam

Dataset and Introduction

Field Name	Usage	Description
Debt to Income Ratio	Input	Total monthly debt payments divided by monthly income
Is Borrower Homeowner	Input	Is the Borrower a Homeowner?
Amount Borrowed	Information Only	Loan Amount
Current Delinquencies, Delinquencies last 7 years	Input	Number of accounts delinquent at time of loan application
Revolving Credit Balance	Input	Revolving credit is credit card debt.
Bank Card Utilization	Input	Total credit card balance on all cards divided by total credit line on all cards
Employment Status	Input	
Income	Input	
Bad	Target	1= Bad and 0=Good

- Initial data dimension: 18,987 observations with 30 variables
- Final data dimension: 14,289 observations with 15 variables, including 3 categorical, 11 numeric and one target.

Data Reformatting

Variable	Irregular value(s)	Number of missing values	Interpretation	Decision
Debt to Income Ratio	NA	659	Data wasn't recorded or found	Remove rows
Current Delinquencies	NA	61	Maybe not investigated	Remove rows
Borrower Occupation	Blank	-	Not mentioned in the dataset	Ignore the whole column
Revolving Credit Balance	NA	4105	Data wasn't recorded or found	Remove rows

- Initial data dimension: 18,987 observations with 30 variables
- Final data dimension: 14,289 observations with 15 variables, including 3 categorical, 11 numeric and one target.
- We ignored the missing values.

Summary Statistics

Table 1. Summary Statistics of numeric variables

Variables	Min.	1 st Quantile	Median	Mean	3 rd Quantile	Max.
Debt to Income Ratio	0.00	0.14	0.22	0.35	0.32	10.01
Amount Borrowed	1000	3000	5000	6742	8850	25000
Current Delinquencies	0.00	0.00	0.00	1.14	1.00	50.00
Delinquencies Last 7 years	0.00	0.00	0.00	5.70	6.00	99.00
Public Records Last 10 Years	0.00	0.00	0.00	0.42	1.00	30.00
Inquiries Last 6 Months	0.00	0.00	2.00	2.75	4.00	46.00
Revolving Credit Balance	0	1169	5000	15478	14821	1435667
Bank Card Utilization	0.00	0.24	0.64	0.57	0.89	5.95
Length Status Months	0.00	0.00	0.00	29.82	31.00	554.00
Principal Balance	0	1629	2865	4104	5234	25000

Table 2. Summary Statistics based on employment status

Employment Status	Debt to Income Ratio	Amount Borrowed	Current Delinquencies	Delinquencies Last 7 Years	Public Records Last 10 Years	Inquiries Last 6 Months
Full-time	0.29	6729.0	1.17	5.81	0.42	2.79
Not employed	3.17	3879.0	1.56	6.04	0.33	2.05
Part-time	0.58	4679.0	0.68	3.84	0.21	1.72
Retired	0.50	4914.0	1.69	7.51	0.63	2.23
Self-employed	0.87	9185.0	0.86	4.61	0.49	3.04

Correlation Matrix for Numeric Variables



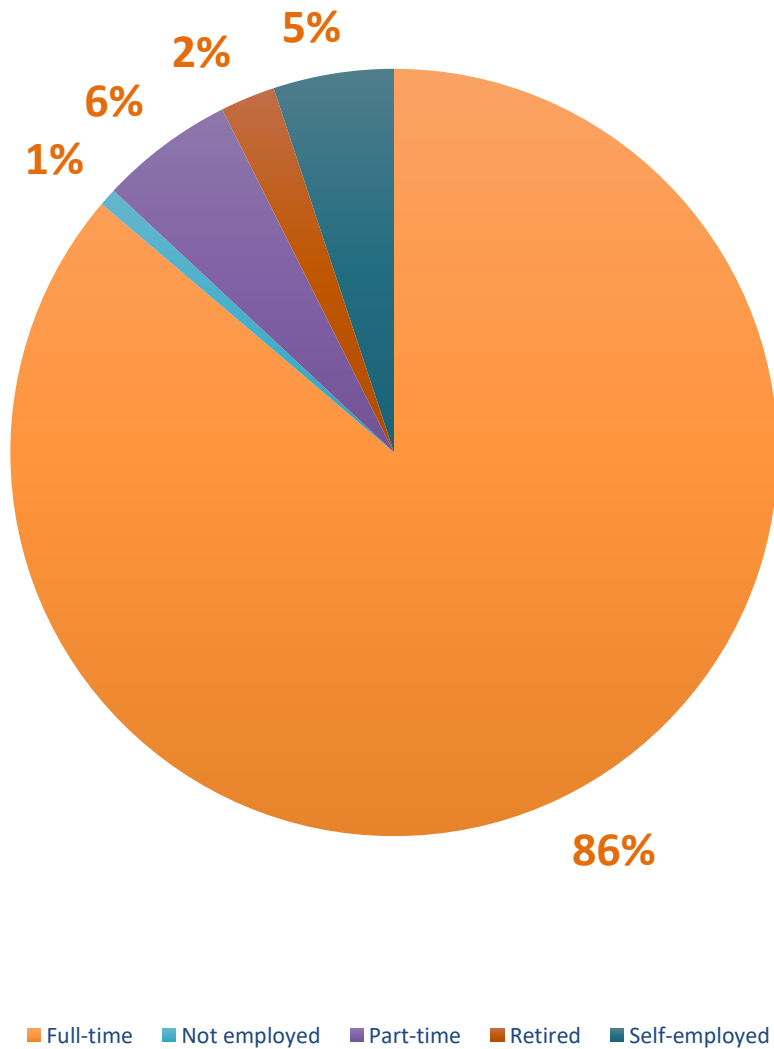


Figure1.Relationship between employment status and home ownership (not having any home)

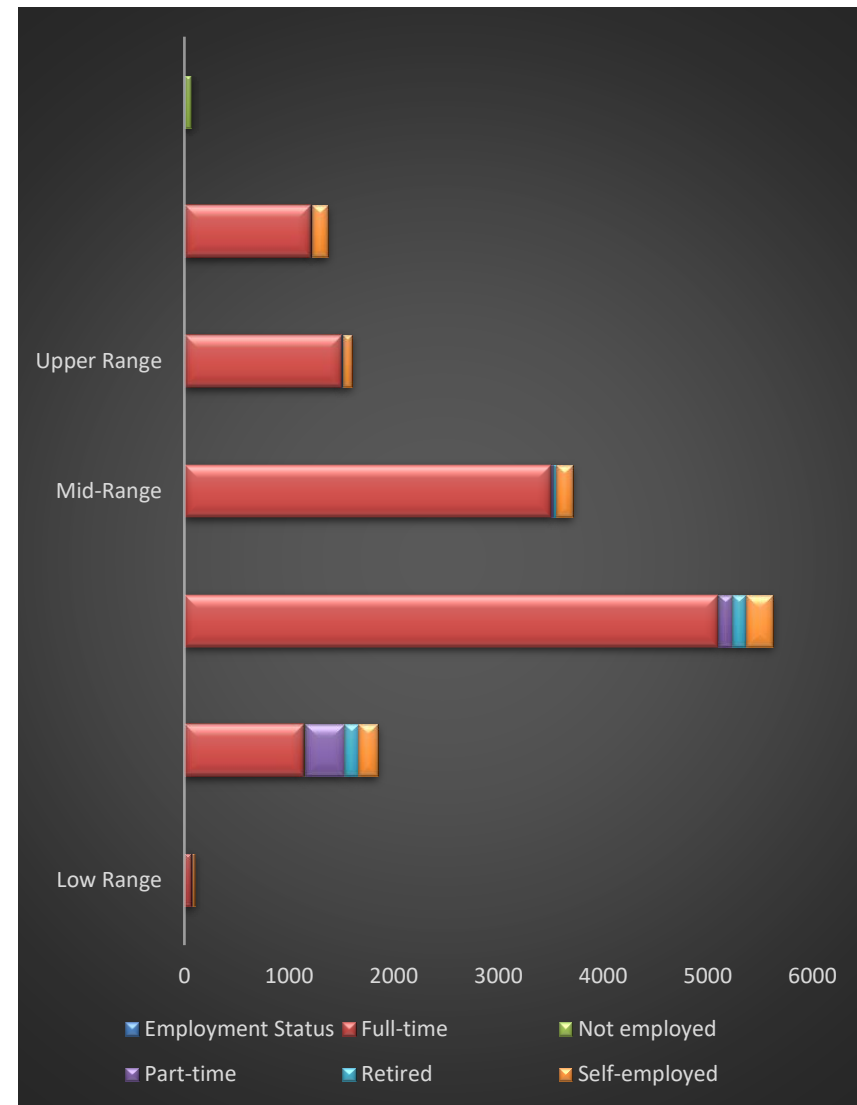


Figure2. Employment Status vs. Income

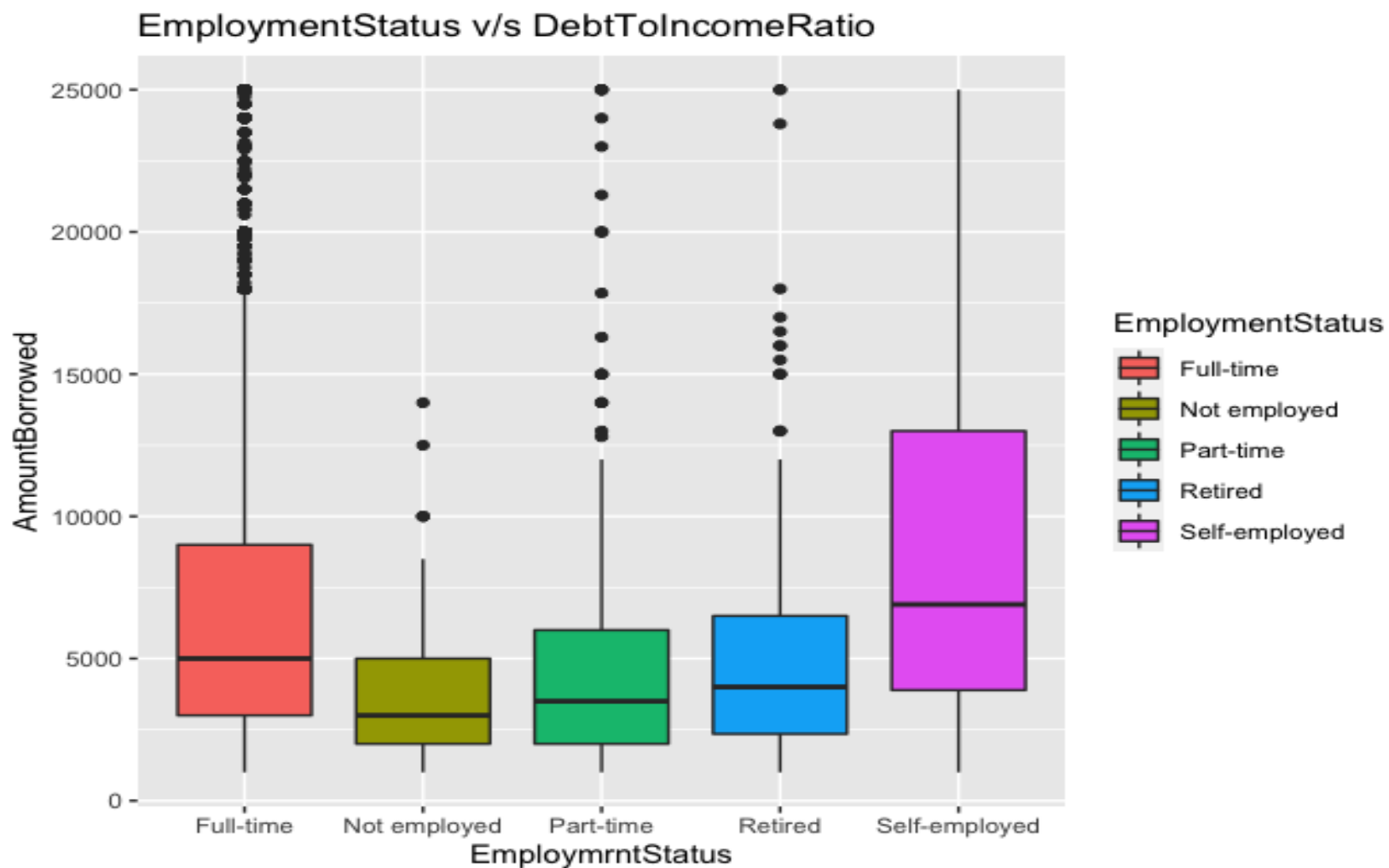


Figure 3. Boxplot of Employment Status vs. Debt to Income Ratio

Barplot of Categorical Variables vs Target (Bad or Good)

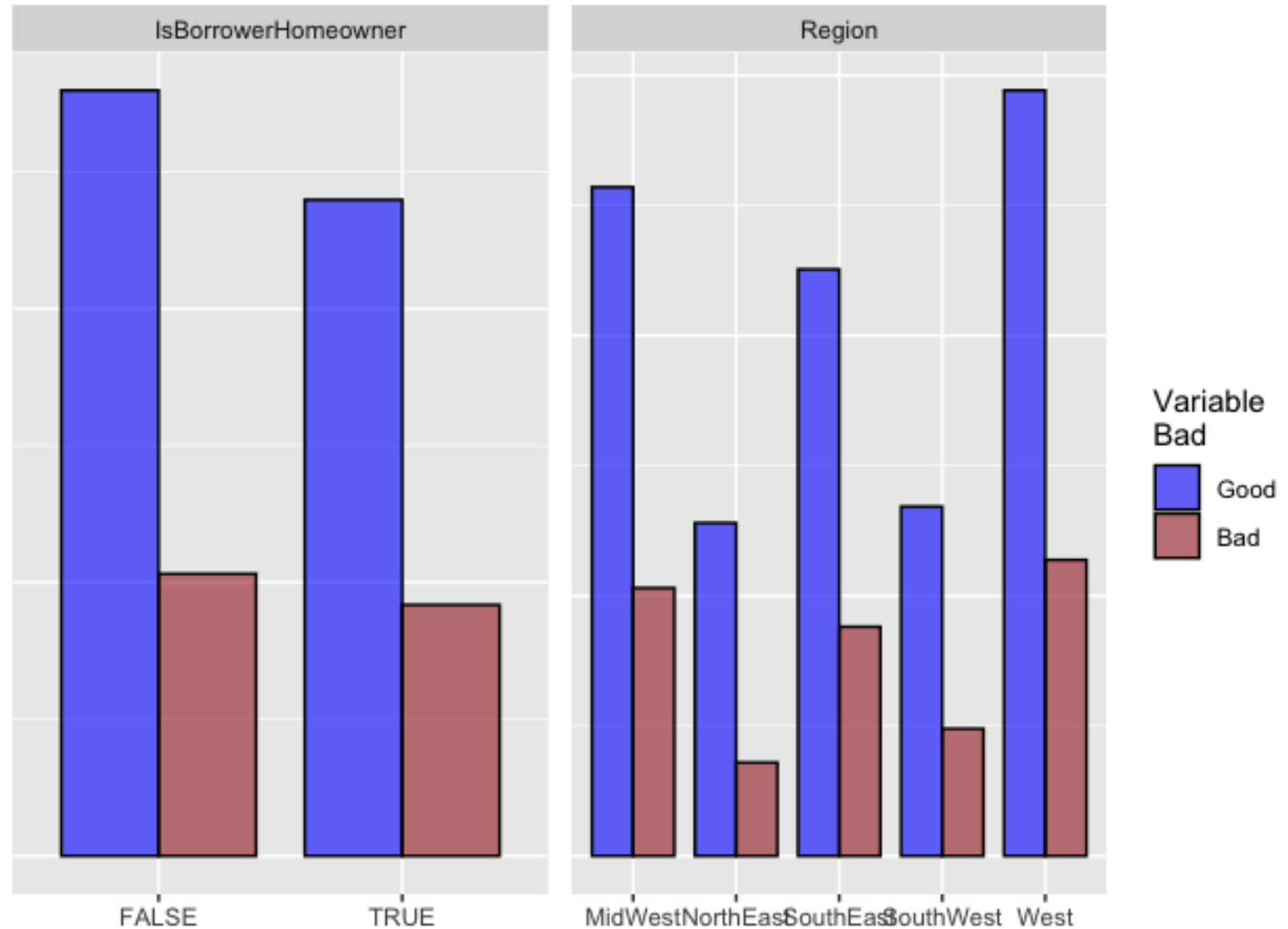


Figure 4. Boxplot of Employment Status vs. Debt to Income Ratio



Conclusion

- Did the exploratory data analysis of credit risk data
- Helped to understand the data and scenario better
- Will proceed to modeling next time