1 Real Data set with no direct score

1.1 German Credit Dataset

The German Credit dataset hosted on (UCIs Machine Learning Repository) contains approximately 1000 observations, with 20 variables. The dependent variable that in all cases we will be trying to predict is whether or not an individual has good credit record. The selected attributes used in generating a ranking function are listed below:

Table 1: German Credit Dataset Selected Attributes

Attribute	Type	Order	Range/Value
Duration in month	Numerical	higher	4-72
Credit amount	Numerical	lower	250-18424
Employment length	Categorical	higher	1-5
Installment rate	Categorical	lower	1-4
Personal status and sex	Categorical	NA	1-5
Residence length	Categorical	higher	1-4
Age	Numerical	NA	19-75
Existing credit	Categorical	higher	1-4
Job	Categorical	higher	1-4

1.2 Adult Income Dataset

The Adult dataset hosted on (UCIs Machine Learning Repository) contains approximately 32000 observations, with 15 variables. The dependent variable that in all cases we will be trying to predict is whether or not an individual has an income greater than $50,000 a year.

Table 2: Adult Income Dataset Selected Attributes

Attribute	Type	Order	Range/Value
Age	Numerical	NA	17-90
Education	Numerical	higher	1-16
Race	Categorical	NA	White, Asian-Pac-Islander, AmerIndian-Eskimo, Other, Black
Sex	Binary	NA	Female,Male
Capital-gain	Numerical	higher	0-99999
Capital-loss	Numerical	lower	0-4356
Hours-per-week	Numerical	NA	0-99

2 Real Data set with ranking score

2.1 Lending Club Credit Dataset

The LendingClubCreditData dataset hosted on Lending Club contains approximately 115 attributes. This file contain complete loan data for all loans issued through the time period stated, including the current loan status (Current, Late, Fully Paid, etc.) and latest payment information. The file containing loan data through the "present" contains complete loan data for all loans issued through the previous completed calendar quarter. It includes approximately 42,538 observations, with 115 attributes.

Table 3: Loan Dataset Selected Attributes

Attribute	Type	Order	Range/Value
funded_amnt	Numerical	higher	500-35000
int_rate	Numerical	lower	5.42%-24.59%
annual_inc	Numerical	higher	1896-6000000
dti	Numerical	higher	0-29.9
revol_bal	Numerical	higher	0-1207359
emp_length	Numerical	higher	0-11

Table 4: Reject Dataset Selected Attributes

Attribute	Type	Order	Range/Value
Amount_Requested	Numerical	lower	0-1400000
Risk_Score	Numerical	higher	0-850
Debt-To-Income_Ratio	Numerical	lower	-1%-99999%
State	Categorical	NA	50 states
Employment_Length	Numerical	higher	0-11

2.2 Give Me Some Credit Kaggle Dataset

The GiveMeSomeCredit dataset hosted on Kaggle contains approximately 150000 observations, with 11 variables. There is also a test set which includes 101503 instances. The goal is to determine whether or not a loan should be granted.

All the attributes in the data set are listed below: **Serious Dlqin2yrs**: (binary) Person experienced 90 days past due delinquency or worse. Values: 0,1.

Revolving Utilization Of Unsecured Lines: (numerical) Total balance on credit cards and personal lines of credit except real estate and no installment debt like car loans divided by the sum of credit limits. Range from: 0.0-50708.0.

Age: (numerical) Age of borrower in years Range from: 0-109.

Table 5: Give Me Some Credit Dataset Selected Attributes

Attribute	Type	Order	Range/Value
Age	Numerical	NA	0-109
Number Of Time 30-59 Days Past Due Not Worse	Numerical	lower	0-98
DebtRatio	Numerical	lower	0-329664
MonthlyIncome	Numerical	higher	0-3008750

Number Of Time 30-59 Days Past Due Not Worse: (numerical) Number of times borrower has been 30-59 days past due but no worse in the last 2 years. Range from: 0-98.

DebtRatio: (numerical) Monthly debt payments, alimony,living costs divided by monthy gross income. Range from: 0.0-329664.0.

MonthlyIncome: (numerical) Monthly income. Range from: 0.0-3008750.0.

NumberOfOpenCreditLinesAndLoans: (numerical) Number of Open loans (installment like car loan or mortgage) and Lines of credit (e.g. credit cards) Range from: 0-58.

NumberOfTimes90DaysLate: (numerical) Number of times borrower has been 90 days or more past due. Range from: 0-98.

NumberRealEstateLoansOrLines: (numerical) Number of mortgage and real estate loans including home equity lines of credit. Range from: 0-54.

NumberOfTime60-89DaysPastDueNotWorse: (numerical) Number of times borrower has been 60-89 days past due but no worse in the last 2 years. Range from: 0-98.

NumberOfDependents: (numerical) Number of dependents in family excluding themselves (spouse, children etc.). Range from: 0-20.