We will be using Berka Dataset from (<https://www.researchgate.net/deref/http%3A%2F%2Fsorry.vse.cz%2F~berka%2Fchallenge%2Fpkdd1999%2Fdata_berka.zip>)

The details for tables are as follows:

1. **ACCOUNTS**

Each record describes static characteristics of an account

Size: 4500 records

1. **CLIENTS**

Each record describes characteristics of a client

Size: 5369 records

1. **DISPOSITON (DISP)**

Each record relates together a client with an account i.e. this relation describes the rights of clients to operate accounts

Size: 5369 records

1. **PERMANENT ORDERS, Debits only (ORDER)**

Each record describes characteristics of a payment order

Size: 6471 records

1. **TRANSACTIONS (TRANS)**

Each record describes one transaction on an account

Size: 1056320 records

1. **LOANS**

Each record describes a loan granted for a given account

Size: 682 records

1. **CREDIT CARDS (CARD)**

Each record describes a credit card issued to an account

Size: 892 records

1. **DEMOGRAPHIC DATA (DISTRICT)**

Each record describes demographic characteristics of a district

Size: 77 records