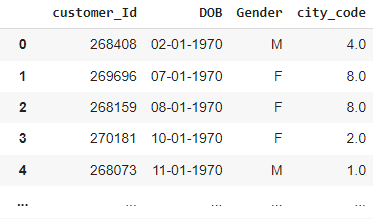
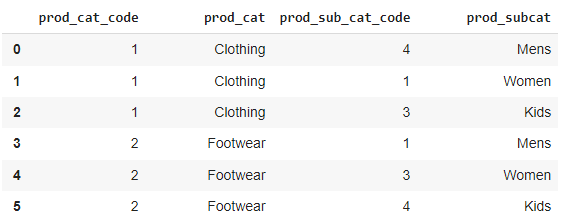
**Data Statistics Using Pandas & Python**

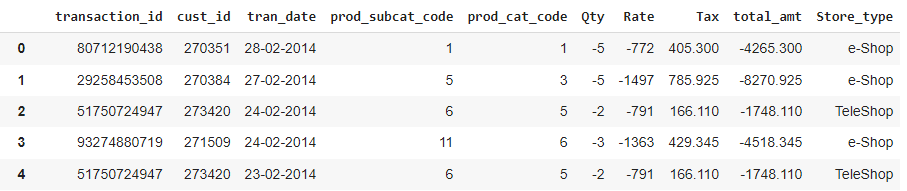
Sebuah perusahaan **retail** ingin melakukan analisa data dari data-data yang mereka kumpulkan di masa lalu. Hal ini digunakan untuk bisa menentukan kebijakan di masa mendatang.

**Data**

* Customers.csv
* Transactions.csv
* Prod\_cat\_info.csv







Bantuan Program

transaction\_customer = pd.merge(Transactions,Customer,left\_on='cust\_id',right\_on='customer\_Id')

transaction\_customer\_prod = pd.merge(transaction\_customer,Prod,left\_on=['prod\_subcat\_code','prod\_cat\_code'] ,right\_on=['prod\_cat\_code', 'prod\_cat\_code'])

Soal

1. Berapakah jumlah data yang tersedia
2. Deskripsikan statistika data sederhana (mean, max, min,..) menggunakan fungsi pandas
3. Bantulah perusahaan untuk bisa menentukan berapa banyak pelanggan berdasarkan jenis kelamin (**men vs women)**
4. Bantulah perusahaan untuk bisa menentukan product kategori (prod\_cat) mana saja yang paling banyak diminati laki-laki
5. Bantulah perusahaan untuk bisa menentukan product kategori (prod\_cat) mana saja yang paling banyak diminati perempuan.
6. Mana y store type yang paling banyak diakses oleh pelanggan untuk mencari produk di perusahaan retail ini.
7. Mana y bulan dengan penjualan tertinggi dan terendah selama satu tahun.

\*bantuan program

transaction\_customer\_prod['month'] = pd.DatetimeIndex(transaction\_customer\_prod['tran\_date']).month