

## ANALYSIS OF DATA SETS FOR A RETAIL STORE

---

### Data set description:

The data set that we will be using is a Ta - Feng Grocery data set which can be downloaded from the following dropbox link:

<https://www.dropbox.com/s/tsd5zd8a7afmzs7/D11-02.ZIP?dl=0>

The data set contains the data of the following 4 months:

- D11: Transaction data collected in November, 2000
- D12: Transaction data collected in December, 2000
- D01: Transaction data collected in January, 2001
- D02: Transaction data collected in February, 2001

### Format of Transaction Data:

- First line: Column definition in Traditional Chinese
- Second line and the rest: data columns separated by “,”

### Column definition:

- Transaction date and time (time invalid and useless)
- Customer I.D
- Age: 10 possible values
  - A <25,B 25-29,C 30-34,D 35-39,E 40-44,F 45-49,G 50-54,H 55-59,I 60-64,J >65
- Residence Area: 8 possible values, A-F: zip code area: 105,106,110,114,115,221,G:others, H: Unknown Distance to store, from the closest: 115,221,114,105,106,110
- Product subclass
- Product ID
- Amount
- Asset
- Sales price

**Problem Statement:**

- Find out the customer I.D for the customer who has spent the maximum amount in a month and a year. Let's do this for year on year basis as well.
- Find out the top 4 or top 10 product being sold.
- Find out the top grossing product and the product subclass for the age group A, B, C etc.....