Project - Retail Store Data

----------------------------------

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:**

data columns separated by “;”

fields terminated by '\;'

**Column definition:**

1)Transaction date and time (Time is invalid)

2)Customer I.D

3)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

4)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

5)Product subclass (category)

6)Product ID

7)Qty or Number of units

8)Total Cost

9)Total Sales

1) Count of unique customers and total sales for each age group and for a given month = Jan

2) Count of customers and total sales for one age group(A) for all products

3) Area wise sales

4) top 10 viable products

5) all loss making products

Total number records : 817741 records