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:**

1)Transaction date and time (time invalid and useless)

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

6)Product ID

7)Qty or Number of units

8)Total Cost

9)Total Sales

**Problem Statement:**

Find out the customer I.D for the customer who has spent the maximum amount in a month and in all the 4 months.

Find out the top 4 or top 10 product being sold in the monthly basis and in all the 4 months..

Find out the top 5 grossing product and the top 5 product subclass for the age group A, B, C etc.....

TA = LOAD '/inputs/demo.txt' using PigStorage(' ') as (no:int, name:chararray);

PARTITIONED = GROUP DATA by name PARTITION BY org.apache.pig.test.utils.SimpleCustomPartitioner pa