**DATA SET DESCRIPTION**

Column1: District.IDI4N 1M1varchar

Column2: ,Distributer nameshellvarchar

Column3: Buy rate (million)$957.70varchar

Column4: Sell rate(million)$5779.92varchar

Column5: volumeIN(millioncubic litter)933,int

Column6: volume OUT(millioncubic litter)

**Problem Statements**

**1. What is the total amount of petrol in volume sold by every distributer.**

*R1 = Load 'Data' Using csvloader() as (District: Chararray, Distributor: Chararray, Buy\_rate: int, Sell\_rate: int, Volume\_IN: int, Volume\_out: int, Year:int);*

*R2 = group R1 by Distributor;*

*R3 = foreach R2 generate Distributor, sum(R1.Volume\_out);*

*Dump R3;*

**2. which are the top 10 distributersID's for selling petrol and also display the amount of petrol sold in volume.**

*R2 = group R1 by distributor;*

*R3 = foreach generate Distributor, sum(R1.Volume\_out) as Petrol\_Sold, sum(R1.Sell\_rate) as Sold\_Amount;*

*R4 = filter R3 10;*

*Dump R4;*

**3. list 10 year where consumption of petrol is more with the distributer id who sold it.**

**4. find distributer name who sold petrol in least amount.**

*R2 = group R1 by distributor;*

*R3 = foreach generate Distributor, sum(R1.Volume\_out) as Petrol\_Sold;*

*R4 = foreach R3 generate Distributor, MIN(Petrol\_sold);*

*Dump R4;*