**Alcohol Dataset:**

The columns in student\_alcohol.csv are school,sex,age,address,famsize,Pstatus,Medu,Fedu, Mjob,Fjob,reason,guardian,traveltime,studytime,failures,schoolsup,famsup,paid,activities,nursery,higher,internet,romantic,famrel,freetime,goout,Dalc,Walc,health,absences,G1,G2,G3

1.Remove the headers

2.Rename to '.txt'

3.Copy the local file in hdfs

#hadoop dfs -put 'file' /

**LOAD Dataset:**

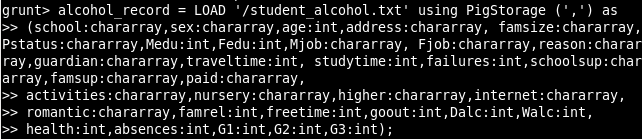
alcohol\_record = LOAD '/student\_alcohol.txt' using PigStorage (',') as

(school:chararray,sex:chararray,age:int,address:chararray, famsize:chararray,Pstatus:chararray,Medu:int,Fedu:int,Mjob:chararray, Fjob:chararray,reason:chararray,guardian:chararray,traveltime:int, studytime:int,failures:int,schoolsup:chararray,famsup:chararray,paid:chararray,

activities:chararray,nursery:chararray,higher:chararray,internet:chararray,

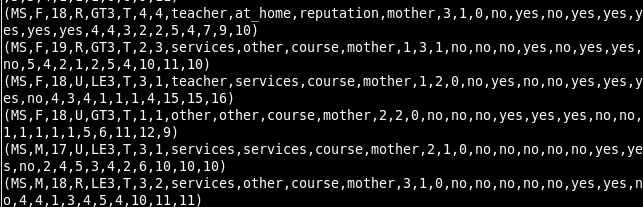
romantic:chararray,famrel:int,freetime:int,goout:int,Dalc:int,Walc:int,

health:int,absences:int,G1:int,G2:int,G3:int);



**View Dataset:**

dump alcohol\_record;



**Analysing Dataset:**

**1)Counts of different school categories**

GroupBySchool = GROUP alcohol\_record BY school;

dump GroupBySchool;

countbyschool = foreach GroupBySchool generate group , COUNT(alcohol\_record.school);

 dump countbyschool;

**2)Counts of Male and Female:**

GroupByGender = GROUP alcohol\_record BY sex;

dump GroupByGender;

countbygender = foreach GroupByGender generate group , COUNT(alcohol\_record.school);

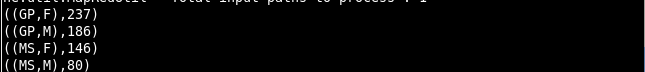
 dump countbygender;

**3)Count by School and Gender:**

GroupBySchoolGender = GROUP alcohol\_record by (school,sex);

dump GroupBySchoolGender;

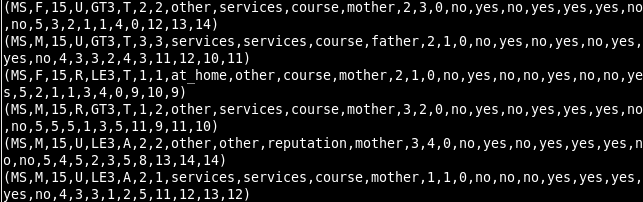
countbyschoolgender = foreach GroupBySchoolGender GENERATE group,COUNT(alcohol\_record.school);

dump countbyschoolgender;

**4) Students Below 15 consuming alcohol:**

filter\_data = FILTER alcohol\_record by age<16;

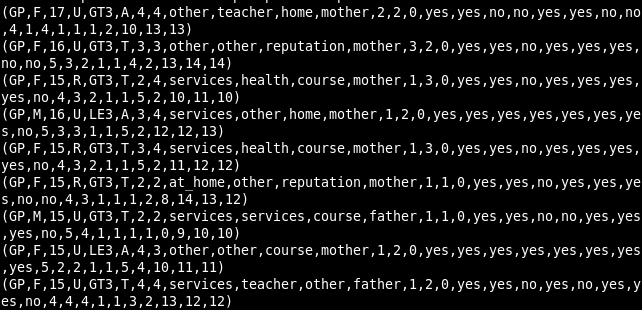
dump filter\_data;



**5) Students with SchoolSupport and FamilySupport:**

filter\_data = FILTER alcohol\_record by (schoolsup=='yes') AND (famsup=='yes') ;

dump filter\_data;



**6) Students Whose Mother or Father is uneducated:**

filter\_data = FILTER alcohol\_record by Medu==0 or Fedu==0;

dump filter\_data;

