

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,nurse
ry,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);
```

```
grunt> 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:charar  
ray,guardian:chararray,traveltime:int, studytime:int,failures:int,schoolsup:char  
array,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;
```

```
(MS,F,18,R,GT3,T,4,4,teacher,at home,reputation,mother,3,1,0,no,yes,no,yes,yes,yes,yes,yes,4,4,3,2,2,5,4,7,9,10)
(MS,F,19,R,GT3,T,2,3,services,other,course,mother,1,3,1,no,no,no,yes,no,yes,yes,no,5,4,2,1,2,5,4,10,11,10)
(MS,F,18,U,LE3,T,3,1,teacher,services,course,mother,1,2,0,no,yes,no,no,yes,yes,yes,no,4,3,4,1,1,1,4,15,15,16)
(MS,F,18,U,GT3,T,1,1,other,other,course,mother,2,2,0,no,no,no,yes,yes,yes,no,no,1,1,1,1,1,5,6,11,12,9)
(MS,M,17,U,LE3,T,3,1,services,services,course,mother,2,1,0,no,no,no,no,no,yes,yes,no,2,4,5,3,4,2,6,10,10,10)
(MS,M,18,R,LE3,T,3,2,services,other,course,mother,3,1,0,no,no,no,no,no,yes,yes,no,4,4,1,3,4,5,4,10,11,11)
```

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;
```

```
(GP,423)
(MS,226)
```

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;
```

```
(F,383)
(M,266)
```

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;
```

```
((GP,F),237)  
((GP,M),186)  
((MS,F),146)  
((MS,M),80)
```

4) Students Below 15 consuming alcohol:

```
filter_data = FILTER alcohol_record by age<16;  
dump filter_data;
```

```
(MS,F,15,U,GT3,T,2,2,other,services,course,mother,2,3,0,no,yes,no,yes,yes,yes,no  
,no,5,3,2,1,1,4,0,12,13,14)  
(MS,M,15,U,GT3,T,3,3,services,services,course,father,2,1,0,no,yes,no,yes,no,yes,  
yes,no,4,3,3,2,4,3,11,12,10,11)  
(MS,F,15,R,LE3,T,1,1,at_home,other,course,mother,2,1,0,no,yes,no,no,yes,no,no,ye  
s,5,2,1,1,3,4,0,9,10,9)  
(MS,M,15,R,GT3,T,1,2,other,services,course,mother,3,2,0,no,yes,no,yes,yes,yes,no  
,no,5,5,5,1,3,5,11,9,11,10)  
(MS,M,15,U,LE3,A,2,2,other,other,reputation,mother,3,4,0,no,yes,no,yes,yes,yes,n  
o,no,5,4,5,2,3,5,8,13,14,14)  
(MS,M,15,U,LE3,A,2,1,services,services,course,mother,1,1,0,no,no,no,yes,yes,yes,  
yes,no,4,3,3,1,2,5,11,12,13,12)
```

5) Students with SchoolSupport and FamilySupport:

```
filter_data = FILTER alcohol_record by (schoolsup=='yes') AND (famsup=='yes') ;  
dump filter_data;
```

```
(GP,F,17,U,GT3,A,4,4,other,teacher,home,mother,2,2,0,yes,yes,no,no,yes,yes,no,no  
,4,1,4,1,1,1,2,10,13,13)  
(GP,F,16,U,GT3,T,3,3,other,other,reputation,mother,3,2,0,yes,yes,no,yes,yes,yes,  
no,no,5,3,2,1,1,4,2,13,14,14)  
(GP,F,15,R,GT3,T,2,4,services,health,course,mother,1,3,0,yes,yes,no,yes,yes,yes,  
yes,no,4,3,2,1,1,5,2,10,11,10)  
(GP,M,16,U,LE3,A,3,4,services,other,home,mother,1,2,0,yes,yes,yes,yes,yes,yes,ye  
s,no,5,3,3,1,1,5,2,12,12,13)  
(GP,F,15,R,GT3,T,3,4,services,health,course,mother,1,3,0,yes,yes,no,yes,yes,yes,  
yes,no,4,3,2,1,1,5,2,11,12,12)  
(GP,F,15,R,GT3,T,2,2,at_home,other,reputation,mother,1,1,0,yes,yes,no,yes,yes,ye  
s,no,no,4,3,1,1,1,2,8,14,13,12)  
(GP,M,15,U,GT3,T,2,2,services,services,course,father,1,1,0,yes,yes,no,no,yes,yes  
,yes,no,5,4,1,1,1,1,0,9,10,10)  
(GP,F,15,U,LE3,A,4,3,other,other,course,mother,1,2,0,yes,yes,yes,yes,yes,yes,ye  
,yes,5,2,2,1,1,5,4,10,11,11)  
(GP,F,15,U,GT3,T,4,4,services,teacher,other,father,1,2,0,yes,yes,no,yes,no,yes,y  
es,no,4,4,4,1,1,3,2,13,12,12)
```

6) Students Whose Mother or Father is uneducated:

```
filter_data = FILTER alcohol_record by Medu==0 or Fedu==0;  
dump filter_data;
```

```
(GP,M,15,U,GT3,T,4,0,teacher,other,course,mother,2,4,0,no,no,no,yes,yes,yes,yes,  
no,3,4,3,1,1,1,0,12,11,11)  
(GP,F,19,U,GT3,T,0,1,at_home,other,course,other,1,2,2,no,yes,no,no,no,no,no,3  
,4,2,1,1,5,0,9,10,11)  
(GP,M,16,U,GT3,T,1,0,other,other,reputation,mother,2,2,0,no,yes,no,yes,yes,yes,y  
es,yes,4,3,2,1,1,3,0,16,17,18)  
(GP,M,16,U,GT3,T,0,2,other,other,other,mother,1,1,0,no,no,no,no,no,yes,yes,no,4,  
3,2,2,4,5,0,11,12,11)  
(GP,F,17,U,LE3,T,0,2,at_home,at_home,home,father,2,3,0,no,no,no,no,yes,yes,yes,n  
o,3,3,3,2,3,2,0,14,14,15)  
(GP,F,20,U,GT3,T,1,0,other,other,reputation,mother,2,1,1,yes,no,no,no,yes,yes,ye  
s,yes,5,3,1,1,1,5,5,8,10,10)  
(MS,F,16,R,GT3,T,0,2,other,other,other,mother,2,1,0,no,yes,no,yes,yes,yes,no,no,  
3,2,3,1,2,2,0,12,11,12)
```