**PROJECT I.2**

\*\*We have data file: titanic\_data\_file(PassengerId, Survived (survived=0 & died=1), Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked)

>pig -x local

grunt>titanic\_data = LOAD ' titanic\_data\_file.txt' USING PigStorage(',') AS (PassengerId:int, Survived:int, Pclass:int, Name:chararray, Sex:chararray, Age:int, SibSp:int, Parch:int, Ticket:chararray, Fare:double, Cabin:chararray, Embarked:char);

**1.In this problem statement we will find the average fare of each class.**

>group\_data = GROUP titanic\_data BY Pclass;

>result = FOREACH group\_data GENERATE group, AVG(titanic\_data.Fare);

>DUMP result;

**2.In this problem statement we will find the number of people alive in each class and are embarked in Southampton.**

>filtered\_data = FILTER titanic\_data BY Survived==0 AND Embarked=='S';

>group\_data = GROUP filtered\_data BY Pclass;

>Result = FOREACH group\_data GENERATE group, COUNT(filtered\_data.PassengerId);

>DUMP Result;

**3.In this problem statement we will find out number of male and female people died in each class.**

> died\_data = FILTER titanic\_data BY Survived==1;

>female\_data = FILTER died\_data BY sex=='female';

>male\_data = FILTER died\_data BY sex=='male';

>female\_group = GROUP female\_data BY pclass;

>male\_group = GROUP male\_data BY pclass;

>female\_count = FOREACH female\_group GENERATE group, COUNT(female\_data);

>DUMP female\_count;

>male\_count = FOREACH male\_group GENERATE group, COUNT(male\_data);

>DUMP male\_count;