**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>passenger\_data = LOAD ' titanic\_data\_file.csv ' 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 passenger\_data BY Pclass;

>result = FOREACH group\_data GENERATE AVG(passenger\_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 passenger\_data BY Survived==0 AND Embarked=='S';

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

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

>DUMP Result;

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

> filtered\_data = FILTER passenger\_data BY Survived==1;

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

>Result1= FOREACH group\_data GENERATE group, COUNT(filtered\_data.Sex=="male") AS Male\_Count;

>DUMP Result1;

>Result2= FOREACH group\_data GENERATE group, COUNT(filtered\_data.Sex=="female") AS Female\_Count;

>DUMP Result2;