Titanic Data Analysis

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

T = load '/user/hdusr/project1/titanic/' using PigStorage(',') as (passengerid:int,survived:chararray,Pclass:int,Name:chararray,Sex:chararray,Age:int,Sibsp:int,Parch:int,Ticket:chararray,Fare:float,Cabin:chararray,Embarked:Chararray);

cl = group T by Pclass;

av = foreach cl generate group, AVG(T.Fare)

DUMP av;

(1,84.15468752825701)

(2,20.66218318109927)

(3,13.675550210257411)

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

al = filter T by survived == '1' AND Embarked == 'S';

cl = group al by (Pclass,survived);

alive = foreach cl generate group.Pclass, COUNT(al.survived);

DUMP alive

(1,74)

(2,76)

(3,67)

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

de = filter T by survived == '0';

mf = group de by (Pclass,Sex);(output will be like(1,male),77)

dead = foreach mf generate group, COUNT(de.survived);

DUMP dead

((1,male),77)

((1,female),3)

((2,male),91)

((2,female),6)

((3,male),300)

((3,female),72)