Aim of the Assignment cum Project——————————

Aim of the Assignment cum Project——————————

The aim of this project/assignment is to predict the flight delays based on the concept of classification based on the class priors as "Flight Delayed" and "Flights Not Delayed" We use the concept of supervised modelling in which based on the train data, we train our system so that we can predict the value of required dependent variable for any unseen test dataset.

IMPLEMENTATION :

This code is implemented using Map-reduce infrastructure for evaluating the linear regression for the average price of the Airlines w.r.t. to variables

- AirTimea
- Distance

For almost all the airlines, the root mean square error for the Air-Time vs Average-Ticket-Price graph is less than the root mean square error for the Distance-Traveled vs Average-Ticket-Price graph. This proves that Air-Time is a better predictor variable than Distance Traveled for the Criterion variable (Average Ticket Price).

Calculating the average ticket price from the slopes and the y-intercepts of the Air-Time regression graphs of various Airlines, we conclude F9 has the least operating cost. The sequence of Airlines (cheapest to dearest) is below : Order is as below:

*** STEPS TO BE EXECUTED in below SEQUENCE FOR AUTOMATED GENERATOON OF myReport.pdf

- make jar
- make permissiona
- make hstart
- make pseudo
- make hstop
- make clean
- make emr
- make report

GRAPHICAL REPRESENTATION

LINEAR REGRESSION -> Average Price with distance and Average Price with Air time:

[1] “2.96493832370479 is the slope for time graph for the Airline AA”
[2] “1.1693616655077 is the slope for time graph for the Airline AS”
[3] “2.95044671838792 is the slope for time graph for the Airline B6”
[4] “3.89522871839555 is the slope for time graph for the Airline DL”
[5] “2.90789562595067 is the slope for time graph for the Airline EV”
[6] “0.313153079599492 is the slope for time graph for the Airline F9”
[7] “3.03023400104918 is the slope for time graph for the Airline HA”
[8] “2.90724483836948 is the slope for time graph for the Airline MQ”
[9] “2.94429908188421 is the slope for time graph for the Airline OO”
[10] “4.76342437964737 is the slope for time graph for the Airline UA”
[11] “3.84003498039684 is the slope for time graph for the Airline US”
[12] “3.00279213400312 is the slope for time graph for the Airline VX”
[13] “1.15138071871333 is the slope for time graph for the Airline WN”
[1] “76.408709585957 is the y-intercept for time graph for the Airline AA”
[2] “19.7624471679235 is the y-intercept for time graph for the Airline AS”
[3] “74.1218013864907 is the y-intercept for time graph for the Airline B6”
[4] “106.763492896288 is the y-intercept for time graph for the Airline DL”
[5] “75.9929624074941 is the y-intercept for time graph for the Airline EV”
[6] “98.9248844866689 is the y-intercept for time graph for the Airline F9”
[7] “37.7067824121906 is the y-intercept for time graph for the Airline HA”
[8] “70.3228747134723 is the y-intercept for time graph for the Airline MQ”
[9] “69.4978213895832 is the y-intercept for time graph for the Airline OO”
[10] “144.256611084893 is the y-intercept for time graph for the Airline UA”
[11] “102.031253877006 is the y-intercept for time graph for the Airline US”
[12] “63.9101006807591 is the y-intercept for time graph for the Airline VX”
[13] “23.8987886278092 is the y-intercept for time graph for the Airline WN”