Assignment 02

Total Marks: 20

Highlights of the Dataset

This is an Airline Dataset having 227,496 flight information of an US Airport in the year 2011.

There is total 21 columns in the dataset. Below is the brief of the columns.

Column Name	Description
Year	the year of departure
Month	the month of departure
DayofMonth	the day of the month of departure
DayOfWeek	the day of week of departure
DepTime	departure time in local time
ArrTime	arrival time in local time
UniqueCarrier	unique abbreviation for a carrier
FlightNum	flight number
TailNum	airplane tail number
ActualElapsedTime	elapsed time of flight, in minutes
AirTime	flight time, in minutes
ArrDelay	arrival delays in minutes
DepDelay	departure delays in minutes
Origin	origin airport code
Dest	destination airport code
Distance	distance of flight, in miles
TaxiIn	taxi in time in minutes
TaxiOut	taxi out time in minutes
Cancelled	cancelled indicator: 1 = Yes, 0 = No
CancellationCode	A = carrier, B = weather, C = national air system, D = security
Diverted	diverted indicator: 1 = Yes, 0 = No

Requirements

i. Answer the questions given in next page and create a PDF report appropriately explaining the results. Add all the R visualizations in the relevant questions to use in your explanation.

ii. Submit both PDF & R script.

Question 1: Consider the months as Winter (January, Feburuary, March) and Summer (June, July, August) seasons. Do planes depart delayed more in winter than in summer? If yes, how significant the delay pattern is? *(Marks: 2)*

Hints: Calculate Mean, Median & SD of both groups. To test significance, use T-test.

Question 2: Show the actual elapsed time of flights going to Atlanta Airport (ATL) in a histogram. Please add additional reference lines showing the mean and median. *(Marks: 2)*

Hints: First step is to create the histogram and then add mean & median lines using "abline".

Question 3: Identify relationship between distance of a flight and below mentioned variables, if there's any. If there's relation, then please identify the significance of it. *(Marks: 3)*

i. Examine whether average distance of flights differs according to the day of the week.
ii. Examine whether distance of a flight is somehow correlated with the taxi out time. Use appropriate chart to visualize the outcome.

Hints:

For (i), dependent variable 'Distance' is numeric & independent variable 'Day of the week' is categorical variable. You can use boxplot to show the relationship.

For (ii), both variables are numeric. Use correlation to find if there's any relationship and use correlation test to evaluate the significance. Draw a plot to show the correlation.

Question 4: What is the average delay time for flights from Houston to Atlanta in comparison to other destinations? *(Marks: 3)*

Hints:

i. Use aggregate function to assess the mean general delay time for every destination.
ii. Use indexing to get information of Atlanta.
iii. Use summary function to get the maximum and minimum for all the average delay times.

Question 5: What is the worst destination to go to from Houston? *(Marks: 3)*

Hints:

i. Compare the result of Atlanta with other destinations in terms of mean & median value.
ii. Find the index of the Airport with most average delay time for flights from Houston to that airport.

Question 6: How many flights are flying from Houston Airport (HOU) and how many from George Bush Intercontinal airport (IAH) per month. Create a loop to calculate the number of flights in each of the 12 months. *(Marks: 4)*

Hints:

i. You can use 'for' loop and repeat it for every month.
ii. Use 'print' function to see a monthly table.

Question 7: Create scatterplot with two groups in different colours: Compare flights starting at HOU (red dots) and flights starting at IAH Airport (green squares) in relation to distance of a flight and Taxi in time. *(Marks: 3)*

Hints:

i. Create two subsets of data – one for HOU & the other for IAH.
ii. Now create a blank plot.
iii. Then add red dots (pch = 16) for HOU and green squares (pch = 22) for IAH. You can use 'points' command to add the points.

Good Luck!