Math 4522

Open-Ended Lab

Date: 10.11.22

Marks: 20

Time: 40 min

Answer the following questions here for the given dataset. Submit your live script file in google classroom.

Import the ‘flight’’ dataset into your workspace. The time-dependent variables in the dataset are in minutes.

1. MATLAB live script documentation 2

2. Which of the attributes do you think should not be considered for outlier analysis? Justify your answer. 2
   Ans:

3. Use any suitable plotting mechanism to demonstrate the presence of outliers for ‘Departure_delay’ variable. (Show it in your live script file) 2
   Which of the following data visualization techniques do you think best represents the outliers? Justify.
   - Line Plot
   - Boxplot
   - Histogram
   - Stem plot
   - Bar plot

Ans:

4. Identify the upper-limit and lower-limit of outliers as well as the number of outliers detected using the following methods. 4
    - Mean
    - Median
    - Quartile
    - Percentile (5, 95)

    Ans:

5. Which of the above mechanisms do you think is the most appropriate to remove/modify outliers from the data for that feature variable? Justify your answer. 3

    Ans:

6. How does the asymmetry in the distribution of the data around zero affect the outlier detection mechanisms? What is the best approach to deal with such extremely skewed data? 3

Ans:

7. Can you determine the canceled flights with the help of departure_delay variable using a polynomial or non-linear regression model? Justify your answer. 2

Ans:

8. With respect to the feature variable ‘Schedule_elapsed_time’, which of the other variables in the dataset can be used for linear regression analysis and prediction? For instance, Schedule_elapsed_time vs actual_elapsed_time - you need to determine whether you can draw a regression line for this task. 2

Ans: