Data Mining; Assignemt 1

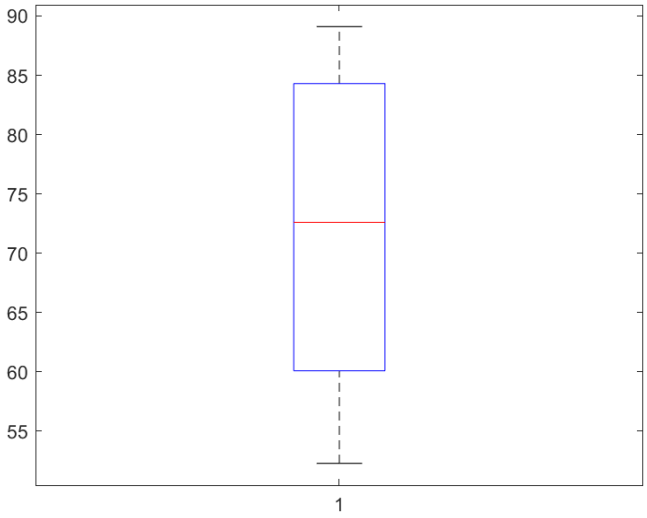
Name: Jesse Annan | ID: 002708111

September 2022

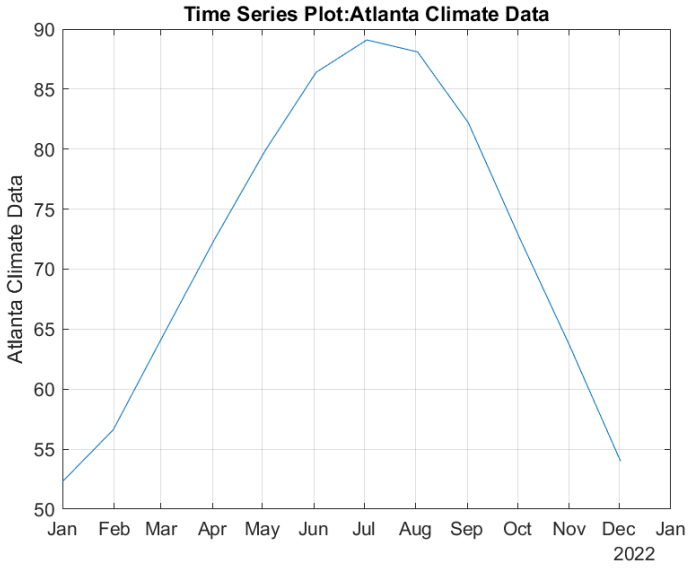
1. 1.1) (a) Mean = 31.20
   (b) Median = 27
   (c) Mode = 27

2. 2.1) (a) MIN = 52.30
   (b) Q1 = 60.10
   (c) Q2 (Median) = 72.60
   (d) Q3 = 84.30
   (e) MAX = 89.10

   2.2) No outliers, this is because per the interquartile range (IQR) rule, no value(s) was 1.5 × IQR away. The interquartile range is the difference between the Q3 and Q1.



   2.3) From the plot we can see that it will be cold during the first and last 3 months of the year, and about room temperature during April and October. The plot also points out that it gets warm from May through September but very warm during July.



3. 3.1) All are nominal attributes
   3.2) similarity between "David" and "Susan" $S_{David,Susan} = \frac{2}{3}$
   3.3) similarity between "Susan" and "Lisa" $S_{Susan,Lisa} = \frac{1}{3}$
4. 4.1) All are binary attributes
   4.2) similarity between "Tom" and "Mat" $S_{Tom,Mat} = \frac{4}{4+3+1} = .5$
   4.3) similarity between "Mat" and "Lucy" $S_{Mat,Lucy} = \frac{3}{3+2+2} = \frac{3}{7}$
5. 5.1) All are numeric attributes
   5.2) Pearson's Correlation Coefficient
   5.3) similarity between "A" and "B" $S_{A,B} = 0.7866$
   5.4) similarity between "B" and "C" $S_{B,C} = 0.8331$

6. 6.1) All are ordinal attributes

   6.2) similarity between "Kevin" and "John"

Kevin	**John**
1	0.75
$\frac{2}{3}$	1
1	0.5

   6.3) similarity between "John" and "Daniel"

John	**Daniel**
0.75	0.5
1	$\frac{1}{3}$
0.5	0

7. Normalized Data (min-max normalization)

Patient	**Height**
Tom	0.4615
Mat	0.8462
Lucy	0
Brain	1