Q1. Frequent Intemsets

a)

Pass1: Counters: a:5, b:6, c:2, d:3, e:2 Frequent Items: a,b,d
Pass 2: Frequent Items: a, b, d Counters: ab:4, ad:2, bd:3 Frequent 2-Items: ab, bd
Pass 3: Frequent 2-Items: ab, bd Counters: abd: 2 Frequent 3-Itemsets: None.
Frequent Itemsets: {a}, {b}, {d}, {ab}, {bd}

b)

The results from the previous computing is enough since conf{b->d} = $\frac{\sup(\{b,d\})}{\sup(\{b\})} = \frac{\frac{3}{7}}{\frac{6}{7}} = \frac{1}{2}$

Both {b} and {b,d} are counted before.

The support for {b->d} is $support\{b \rightarrow d\} = \frac{Times(\{b \rightarrow d\})}{Total\ numbers} = \frac{3}{7}$

c)

Hash Function: f(i, j) = (i+j)%3

Pass1: Counters for the itemsets: 1:2, 2:2, 3:2, 4:2, 5:1, 6:1, 7:2 f(1,3)=1; f(1,4)=2, f(3,4)=1, f(4,5)=0, f(2,7)=0, f(1,6)=1 Counters for the buckets: B0:3, B1:3, B2:1 Frequent Items: 1, 2, 3, 4, 7
Pass 2: Frequent Items: 1, 2, 3, 4, 7 Counters for the buckets: B0:3, B1:3, B2:1 Counters for the itemsets: (1,3):1, (2,7):2, (3,4):1 Frequent 2-Items: (2, 7)
Pass 3: Frequent 2-Items: (2, 7) No frequent 3-itemsets combination
Frequent Itemsets: {1}, {2}, {3}, {4}, {7}, {2,7}

Question 2

2.1

provided with a dataset ("clustering_data.csv") and a text document describing the task.

Question 4: Practical exercise on FI and AR, 30/100 points. You are provided with a dataset ("mammographic_masses.csv") and a text document describing the task.

Q2.1

Step 1

P_1: Blue ; P_4: Yellow.

P_2 $d(P_2, P_1) = 0.5$, $d(P_2, P_4) = \sqrt{1 + (\frac{1}{2})^2} = 1.118$, $\Rightarrow P_2$ is assigned to Blue clustering

P_3 $d(P_3, P_1) = 1.118$, $d(P_3, P_4) = 0.5 \Rightarrow P_3$ is assigned to Yellow Clustering

P_5 $d(P_5, P_1) = 4$, $d(P_5, P_4) = \sqrt{3^2+1} = 3.162 \Rightarrow P_5$ is assigned to Yellow Clustering

P_6 $d(P_6, P_1) = \sqrt{4^2+1} = 4.123$, $d(P_6, P_4) = 3 \Rightarrow P_6$ is assigned to Yellow Clustering

P_7 $d(P_7, P_1) = \sqrt{5^2+1} = 5.099$, $d(P_7, P_4) = 4 \Rightarrow P_7$ is assigned to Yellow Clustering

We computed the centroids $\bar{C}_1$, $\bar{C}_2$ as follows:

$$\bar{C}_1 = \frac{P_1 + P_2}{2} = (0, \frac{1}{4}), \quad \bar{C}_2 = \frac{P_3 + P_4 + P_5 + P_6 + P_7}{5} = (3, \frac{7}{10})$$

Step 2:

$\bar{C}_1$: Blue $\bar{C}_2$: Yellow

P_1 $d(P_1, \bar{C}_1) = 0.25$, $d(P_1, \bar{C}_2) = 3.081 \Rightarrow P_1$ gets blue

P_2 $d(P_2, \bar{C}_1) = 0.25$, $d(P_2, \bar{C}_2) = 3.007 \Rightarrow P_2$ remains in Blue clusters

P_3 $d(P_3, \bar{C}_1) = 1.031$, $d(P_3, \bar{C}_2) = 2.010 \Rightarrow P_3$ gets blue

P_4 $d(P_4, \bar{C}_1) = 1.25$ $d(P_4, \bar{C}_2) = 2.022 \Rightarrow P_4$ gets blue

P_5 $d(P_5, \bar{C}_1) = 4.008$, $d(P_5, \bar{C}_2) = 1.221 \Rightarrow P_5$ remains in Yellow Clusters

P_6 $d(P_6, \bar{C}_1) = 4.070$, $d(P_6, \bar{C}_2) = 1.044 \Rightarrow P_6$ remains in Yellow Clusters

P_7 $d(P_7, \bar{C}_1) = 5.056$ $d(P_7, \bar{C}_2) = 2.224 \Rightarrow P_7$ remains in Yellow Clusters

So, the $\hat{C}_1$ and $\hat{C}_2$ should be:

$$\hat{C}_1 = \frac{P_1 + P_2 + P_3}{3} = (\frac{1}{3}, \frac{1}{3}), \quad \hat{C}_2 = \frac{P_4 + P_5 + P_6 + P_7}{4} = (\frac{7}{2}, \frac{3}{4})$$

Step 3:

$\hat{C}_1$: Blue, $\hat{C}_2$: Yellow

P_1 $d(P_1, \hat{C}_1) = 0.471$, $d(P_1, \hat{C}_2) = 3.579 \Rightarrow P_1$ remains blue

P_2 $d(P_2, \hat{C}_1) = 0.373$, $d(P_2, \hat{C}_2) = 3.509 \Rightarrow P_2$ remains Blue

P_3 $d(P_3, \hat{C}_1) = 0.687$, $d(P_3, \hat{C}_2) = 2.512 \Rightarrow P_3$ remains Blue

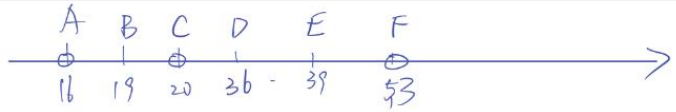
P_4 $d(P_4, \hat{C}_1) = 0.943$, $d(P_4, \hat{C}_2) = 2.512 \Rightarrow P_4$ remains Blue

P_5 $d(P_5, \hat{C}_1) = 3.682$, $d(P_5, \hat{C}_2) = 0.901 \Rightarrow P_5$ remains Yellow

P_6 $d(P_6, \hat{C}_1) = 3.727$, $d(P_6, \hat{C}_2) = 0.559 \Rightarrow P_6$ remains Yellow

P_7 $d(P_7, \hat{C}_1) = 4.714$, $d(P_7, \hat{C}_2) = 1.521 \Rightarrow P_7$ remains Yellow

Q 2.2



Take A, C, F as centroid.

Step1

Let A be the Red clustering
C be the Yellow clustering
and F be the Green clustering.

B $d(BA)=3$, $d(BC)=1$, $d(BF)=34$, B is assigned to Yellow
D $d(DA)=20$, $d(DC)=16$, $d(DF)=17$, D is assigned to Yellow
E $d(EA)=23$, $d(EC)=19$, $d(EF)=14$, E is assigned to Green

$$\bar{C}_1=A,\quad \bar{C}_2=\frac{B+C+D}{3},\quad \bar{C}_3=\frac{E+F}{2}$$

$$\Rightarrow \bar{C}_1=16,\quad \bar{C}_2=25,\quad \bar{C}_3=46$$

Step2.

B $d(B\bar{C}_1)=3$, $d(B\bar{C}_2)=6$, $d(B\bar{C}_3)=27$, B gets Red
C $d(C\bar{C}_1)=4$, $d(C\bar{C}_2)=5$, $d(C\bar{C}_3)=26$, C gets Red
D $d(D\bar{C}_1)=20$, $d(D\bar{C}_2)=11$, $d(D\bar{C}_3)=10$, D gets Green
E $d(E\bar{C}_1)=23$, $d(E\bar{C}_2)=14$, $d(E\bar{C}_3)=7$, E remains Green
F $d(F\bar{C}_1)=37$, $d(F\bar{C}_2)=25$, $d(F\bar{C}_3)=7$, F remains Green

$$\hat{C}_1=\frac{A+B+C}{3}=18.33,\quad \hat{C}_2=\text{null},\quad \hat{C}_3=\frac{E+F}{2}=46$$

Step3

A $d(A\hat{C}_1)=2.33$, $d(A\hat{C}_2)=\text{null}$, $d(A\hat{C}_3)=30$, A remains Red
B $d(B\hat{C}_1)=0.67$, $d(B\hat{C}_2)=\text{null}$, $d(B\hat{C}_3)=27$, B remains Red
C $d(C\hat{C}_1)=1.67$, $d(C\hat{C}_2)=\text{null}$, $d(C\hat{C}_3)=26$, C remains Red
D $d(D\hat{C}_1)=17.67$, $d(D\hat{C}_2)=\text{null}$, $d(D\hat{C}_3)=10$, D remains Green
E $d(E\hat{C}_1)=20.67$, $d(E\hat{C}_2)=\text{null}$, $d(E\hat{C}_3)=7$, E remains Green
F $d(F\hat{C}_1)=34.67$, $d(F\hat{C}_2)=\text{null}$, $d(F\hat{C}_3)=7$, F remains Green

The clustering doesn't change, the algorithm terminates.