SOLUTIONS

1. A, D, F
2. 1) C, 2) A, 3) B, 4) E
3. D
4. A (but D is also acceptable as months are not ordered properly)
5. Recency, frequency, and monetary value
6. B
7. C

8. You can consider that (3,4) and (5,2) are two temporary data points acting as initiators.

Solution: There are two seed nodes that is (3,4) and (5,2) given in the question. Thus, you need to calculate the Euclidean distance of each of the seed points to all the three data points.

Please find below attached the screenshot where you can see the calculations with respect to seed points.

Since for all the three data points, it is (3,4) data point which is the nearest to all the data points. But the question is K=2. So it means you need to form 2 clusters.

Ideally you should put all the three data points in (3,4) but it is asked to create 2 clusters (k=2). In this situation, since the min distance to (3,4) is 178 (2,4), so it goes in (3,4) cluster. Next, we need to calculate the centroid of (3,4) and (2,4), which is (2.5, 4). From this centroid calculate the distance 163 (5, 7) and 112 (1,1). Out of two data points, the distance between centroid and 112 (1,1) is minimum. So you can put it in that cluster. So one iteration is enough.
9. Use recall and precision formula.

<table>
<thead>
<tr>
<th>Recall</th>
<th>TP/TP+FN = 50/(50 + 25) = 50/75 = 0.666</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>TP/TP+FP = 50/(50 + 30) = 50/80 = 0.625</td>
</tr>
</tbody>
</table>

10. A
11. B, C, D
12. Follow the equation in the Question 11. Now put X1, X2, X3, X4 with the values provided in Question 12.
13. C