**BUSINESS INTELLIGENCE AND ANALYTICS**

**ASSIGNMENT WEEK 10:**

**Total marks = 1 Marks**
**15 Qns * 1 marks = 15 marks**

1. What does RFM stand for in customer segmentation strategy? (1 Mark)
   A) Recency, Frequency, Management
   B) Recency, Frequency, Monetary value
   C) Revenue, Frequency, Management
   D) Recency, Firmness, Money

**Answer: B) Recency, Frequency, Monetary value**

2. Which of the following is NOT true about RFM analysis? (1 Mark)
   A) It requires detailed demographic data.
   B) It helps in predictive modelling techniques like churn modelling.
   C) It is useful for targeting mails.
   D) Customers with high scores in all three categories are likely to be the most loyal and profitable.

**Answer: A) It requires detailed demographic data.**

3. How are Butterflies categorized in RFM segmentation? (1 Mark)
   A) Highly loyal but not very profitable
   B) Profitable but disloyal customers
   C) Highly loyal and profitable
   D) Customers who generate neither loyalty nor profits

**Answer: B) Profitable but disloyal customers**

4. Which customer group is likely to need incentives to increase their spending and engagement? (1 Mark)
   A) True Friends
   B) Butterflies
   C) Barnacles
   D) Strangers

**Answer: C) Barnacles**

5. What does CLV stand for in RFM analyses? (1 Mark)
A) Customer Lifetime Value
B) Customer Loyalty Value
C) Customer Longevity Value
D) Customer Lifetime Volume

**Answer: A) Customer Lifetime Value**

6. How are R, F, and M typically combined to create composite scores in some methods? (1 Mark)
   A) Adding R, F, and M directly

B) Multiplying R by 5, F by 2, and M by 1
C) Dividing R, F, and M by a constant
D) Subtracting R from F and then adding M

**Answer: B) Multiplying R by 5, F by 2, and M by 1**

7. What SQL function is used for RFM analysis to scale RFM into a predefined range? (1 Mark)

   A) GROUP BY

   B) AVG()

   C) NTILE()

   D) MAX()

   **Answer: C) NTILE()**

8. In RFM analysis what does "Recency" refer to? (1 Mark)

   A) The total amount a customer spends

   B) The number of purchases made by a customer

   C) The time elapsed since a customer's last purchase

   D) The predicted net profit from a customer over a future horizon

**Answer: C) The time elapsed since a customer's last purchase**

9. What is the purpose of the `fit_predict` method in scikit-learn's K-means implementation? (1 Mark)
   A) Training the model
   B) Predicting cluster labels
   C) Evaluating model performance
   D) Visualizing data distribution

   **Answer: B) Predicting cluster labels**

10. Which Python package provides functionality for visualizing K-means clustering results using 2D and 3D plots?
    A) seaborn
    B) b. matplotlib
    C) c. pandas
    D) d. scikit-learn

    **Answer: B) `matplotlib`**

11. How is Recency (R) scaled after grouping Days since last order into 10 deciles? (1 Mark)
    a) It is scaled from 1-5 for better representation
    b) It is reversed, with the most recent customer receiving the highest R value
    c) It is not scaled as it represents the number of days directly

d) It is scaled logarithmically for better clustering

**Answer: B) It is reversed, with the most recent customer receiving the highest R value**

12. Which clustering algorithm assigns data points to the nearest cluster centroid? (1 Mark)
    a) K-Means
    b) DBSCAN
    c) Agglomerative
    d) Mean-Shift

**Answer: a. K-Means**

13. A retail company wants to segment its customers for targeted marketing campaigns. They have data on customer demographics (age, gender, income), purchase history (amount, frequency, categories), and online behaviour (website visits, clicks). Which features are most suitable for k-means clustering in this scenario? (1 Mark)

a) Demographics only (age, gender, income)
b) Purchase history only (amount, frequency, categories)
c) Online behaviour only (website visits, clicks)
d) A combination of all features

**Ans: d) A combination of all features**

14.True or False: In K-means clustering, each cluster is represented by its center (centroid) which corresponds to the median of points assigned to the cluster.

Ans: False

15. Out of the reasons elicited below, what would be a major reason for you not to choose K- means for clustering analysis? (1 Mark)

a) It is sensitive to noise and outlier data points and also sensitive to the initial placement of its cluster centers.
b) It always leads to complex cluster formation due to unequal size of cluster formed
c) Inter cluster distance is high for K-Means clustering
d) Accuracy of the model is comparatively low compared to other modes of clustering

**Ans: A) It is sensitive to noise and outlier data points and also sensitive to the initial placement of its cluster centers (centroids).**