Login: ..... **Name: .....**

Question 1: Supervised vs. Unsupervised Learning (4 points)

For each of the following scenarios, select whether it is an example of supervised or unsupervised learning:

1. **A financial institution uses labeled data to predict whether a customer will default on a loan based on their credit history and income. (1p)**
   a) Supervised
   b) Unsupervised

2. **A self-driving car is trained to recognize pedestrians by processing large amounts of labeled images and video data. (1p)**
   a) Supervised
   b) Unsupervised

3. **A retailer uses a model to group customers into clusters based on purchasing patterns, without predefined labels, to personalize marketing strategies. (1p)**
   a) Supervised
   b) Unsupervised

4. **A healthcare provider builds a system that predicts disease progression by discovering hidden patterns in patient data without using labeled outcomes. (1p)**
   a) Supervised
   b) Unsupervised

Login: ..... **Name: .....**

Question 2: Linear Regression and Model Evaluation (4 points)

You are given the following dataset about advertising spend and product sales, and you need to perform linear regression:

Advertising Spend ($k)	Sales (units)
10	200
20	450
30	700
40	850
50	1000

1. **What is the equation of the best fit line (linear regression) for this data? (2p)**
   - **a) y = 20x + 100**
   - b) y = 25x + 150
   - c) y = 15x + 150

   Your solution:

2. **Based on the regression equation, predict the number of units sold if $35k is spent on advertising. (1p)**
   a) 725
   b) 780
   c) 800
   d) 900

   Your solution:

Login: ..... **Name: .....**

3. **Which method could you use to evaluate the linear regression model's performance in predicting new data? (1p)**
   a) Testing set error
   b) Training set error
   c) Euclidean distance
   d) Classification accuracy

Question 3: K-Means Clustering (4 points)
You are given the following data points, each with two features:

Point	Feature 1	Feature 2
A	3	4
B	4	5
C	6	8
D	7	7
E	2	2
F	8	8
G	7	6
H	3	3

Assume the initial cluster centroids are Point B and Point F.

1. **After the first iteration of K-Means, which points are assigned to Cluster 1 (centroid B)? (2p)**
   a) A, B, H
   b) A, B, E
   c) A, B, H, E
   d) B, C, D

Your solution:

Login: ..... **Name: .....**

2. **After the first iteration, what are the new centroids? (1p)**
   a) (3.0, 3.5) and (7, 7.25)
   b) (3.5, 4.0) and (8.0, 8.0)
   c) (2.5, 3.5) and (7.0, 7.0)
   d) (4.0, 4.0) and (7.5, 7.5)

Your solution:

3. **Which of the following is commonly used to determine the optimal number of clusters? (1p)**
   a) Gradient descent
   b**) Silhouette score**
   c) Cross-validation
   d) Mean Absolute Error

Login: ..... **Name: .....**

Question 4: Clustering analysis (4 points)

You used K-Means clustering on a retail customer dataset based on two features: income and spending habits. The algorithm produced the following clusters:

- **Cluster 1**: Low income, low spending.
- **Cluster 2**: High income, high spending.
- **Cluster 3**: Medium income, medium spending.

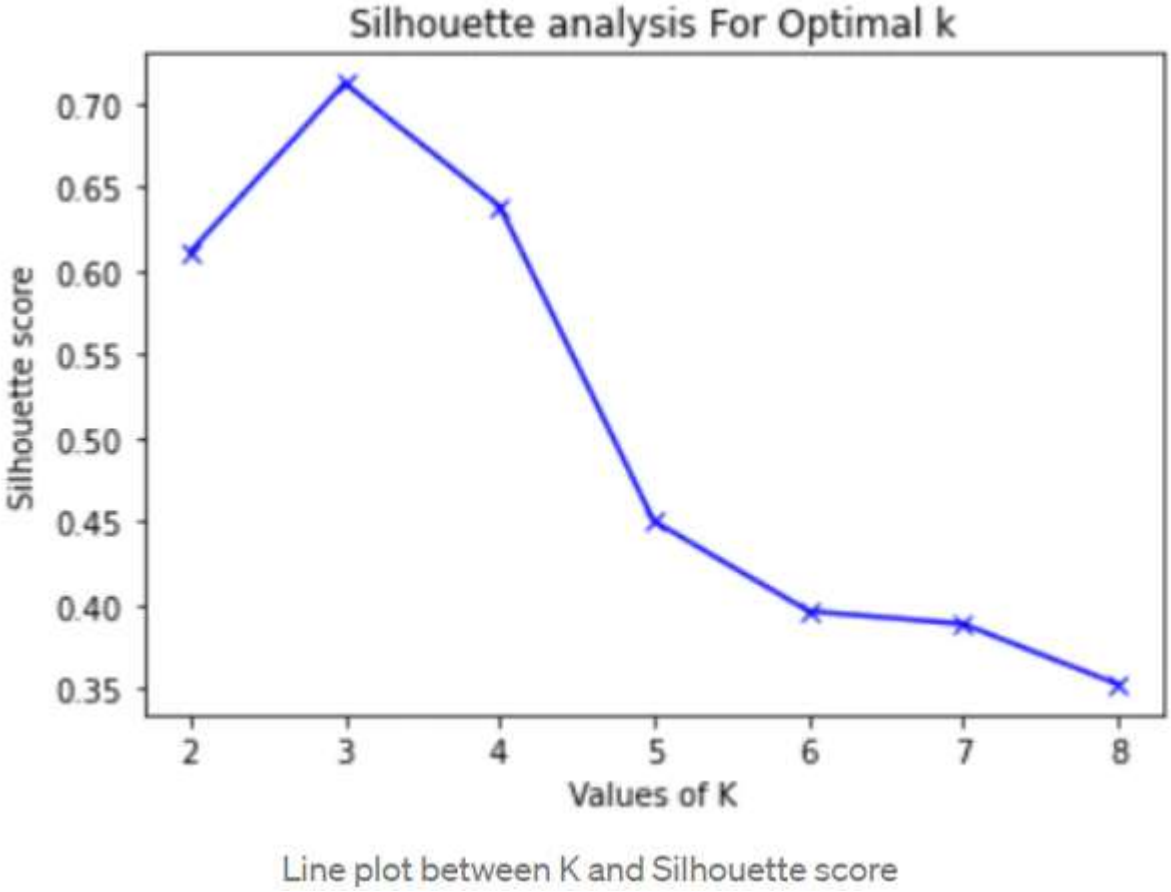
Answer the following questions:

1. **Which cluster should be targeted for luxury product promotions? (1p)**
   a) Cluster 1
   b) Cluster 2
   c) Cluster 3
   d) None of the above

   Why?

Login: ..... **Name: .....**

2. You have performed K-Means clustering on the dataset for various cluster numbers and calculated the Silhouette score.



Line plot between K and Silhouette score

Your goal is to estimate the optimal number of clusters. How many clusters would you pick? (2p)

a) 2
b) 3
c) 5
d) 8

Why?

3. **If the clusters are well-separated, what does this suggest about the customer groups? (1p)**
   a) They have clear and distinct purchasing behaviors
   b) They are very similar in spending habits
   c) The clusters are overlapping
   d) The clustering needs to be improved

Login: ..... **Name: .....**

Question 5: K-Nearest Neighbors (KNN) Classification (4 points)
You are given the following data to classify whether a person is "Healthy" (Yes) or "Not Healthy" (No) based on their hours of exercise per week and calorie intake.

Person	Hours of Exercise	Calorie Intake	Healthy?
1	1	3500	No
2	5	2500	Yes
3	4	2000	Yes
4	2	4000	No

A new person has 3 hours of exercise and a calorie intake of 2800.

1. **Using KNN with k=3, what is the predicted classification for the new person? (2p)**
   a) Healthy
   b) Not Healthy

Your solution:

2. **If k=1, what is the predicted classification? (1p)**
   a) Healthy
   b) Not Healthy

Your solution:

3. **How does increasing the value of k affect the model's bias and variance? (1p)**
   a) Increases bias, decreases variance
   b) Decreases bias, increases variance
   c) Increases both bias and variance
   d) Decreases both bias and variance

Login: ..... **Name: .....**