1. A real estate company wants to develop a system that predicts house prices based on square footage, number of bedrooms, and location.
   Q: Identify the problem type and outline the step-by-step logic to solve it.

   Problem Type: Supervised Learning – **Regression**
   Step-by-Step Logic:
   1. Collect data: square footage, bedrooms, location, and price.
   2. Preprocess data: handle missing values, encode categorical variables (e.g., location).
   3. Split into training and test sets.
   4. Train a regression model (e.g., Linear Regression, Random Forest).
   5. Evaluate model (e.g., RMSE, MAE).
   6. Use the model to predict prices for new properties.

2. A bank wants to build a model to detect fraudulent transactions by analyzing customer spending behavior and transaction history.
   Q: Identify the problem type and outline the step-by-step logic to solve it.

   Problem Type: Supervised Learning – **Classification**
   Step-by-Step Logic:
   1. Collect transaction and customer behavior data with labels (fraudulent or not).
   2. Preprocess data: clean, normalize, handle imbalanced classes (e.g., SMOTE).
   3. Split dataset into train/test sets.
   4. Train a classification model (e.g., Logistic Regression, XGBoost).
   5. Evaluate with metrics like accuracy, precision, recall, F1-score.
   6. Deploy model to detect fraud in real-time transactions.

3. A supermarket wants to segment its customers based on their shopping patterns to provide personalized promotions.
   Q: Identify the problem type and outline the step-by-step logic to solve it.

   Problem Type: Unsupervised Learning – **Clustering**
   Step-by-Step Logic:
   1. Collect data on customer purchases, frequency, basket size, etc.
   2. Preprocess data: normalize numeric features.
   3. Apply clustering algorithms (e.g., K-Means).
   4. Determine optimal number of clusters (e.g., Elbow Method).
   5. Analyze segments and label them based on shopping behavior.
   6. Use segments for targeted promotions.

4.A company wants to estimate an employee's salary based on their years of experience, job title, and education level.

Q: Identify the problem type and outline the step-by-step logic to solve it.

Problem Type: Supervised Learning – **Regression**

Step-by-Step Logic:

1. Gather data: experience, job title, education, and salary.
2. Encode categorical features (e.g., job title, education).
3. Normalize or scale numerical data.
4. Split data into training/testing sets.
5. Train regression model.
6. Evaluate and deploy model to predict salaries.

5.An email provider wants to automatically classify incoming emails as spam or not spam based on their content and sender details.

Q: Identify the problem type and outline the step-by-step logic to solve it.

Problem Type: Supervised Learning – **Classification**

Step-by-Step Logic:

1. Collect labeled email data (spam/ham).
2. Preprocess text data (cleaning, tokenization, vectorization like TF-IDF).
3. Train/test split.
4. Train classifier (e.g., Naive Bayes, SVM).
5. Evaluate performance (accuracy, F1-score).
6. Use model to classify incoming emails.

6.A business wants to analyze customer reviews of its products and determine whether the sentiment is positive or negative.

Q: Identify the problem type and outline the step-by-step logic to solve it.

Problem Type: Supervised Learning – **Classification (Text)**

Step-by-Step Logic:

1. Collect customer reviews labeled with sentiment (positive/negative).
2. Preprocess text (remove stopwords, stemming, vectorize).
3. Train/test split.
4. Train text classification model (e.g., Logistic Regression, LSTM).
5. Evaluate with precision/recall/F1.
6. Predict sentiment of new reviews.

7.An insurance company wants to predict whether a customer is likely to file a claim in the next year based on their driving history and demographics.

Q: Identify the problem type and outline the step-by-step logic to solve it.

Problem Type: Supervised Learning – **Classification**

Step-by-Step Logic:

1. Collect data on customers' driving history and demographics with claim status.
2. Encode categorical features, handle missing values.
3. Split into training and test sets.
4. Train a classification model (e.g., Decision Tree, Random Forest).
5. Evaluate model (AUC-ROC, accuracy).
6. Use it to predict likelihood of future claims.

8.A streaming platform wants to recommend movies to users by grouping them based on their viewing preferences and watch history.

Q: Identify the problem type and outline the step-by-step logic to solve it.

Problem Type: Unsupervised Learning – **Clustering / Collaborative Filtering (Recommender System)**

Step-by-Step Logic:

1. Collect user viewing history and preferences.
2. Use collaborative or content-based filtering.
3. Optionally, apply clustering (e.g., K-Means) to group similar users.
4. Recommend movies from similar groups or based on user-item similarity.
5. Continuously update recommendations as new data comes in.

9.A hospital wants to predict the recovery time of patients after surgery based on their age, medical history, and lifestyle habits.

Q: Identify the problem type and outline the step-by-step logic to solve it.

Problem Type: Supervised Learning – **Regression**

Step-by-Step Logic:

1. Gather data on patient demographics, health history, lifestyle, and recovery time.
2. Preprocess data: handle missing entries, encode categorical variables.
3. Train/test split.
4. Train regression model.
5. Evaluate using metrics like RMSE or MAE.
6. Predict recovery times for new patients.

10.A university wants to predict a student's final exam score based on study hours, attendance, and past academic performance.

Q: Identify the problem type and outline the step-by-step logic to solve it.

Problem Type: Supervised Learning – **Regression**

Step-by-Step Logic:

1. Collect data: study hours, attendance, past grades, and final exam scores.
2. Preprocess and clean the data.
3. Split dataset into training and testing sets.
4. Train regression model (e.g., Linear Regression).
5. Evaluate model using R^2, MAE, RMSE.
6. Predict exam scores based on new student input.