Task-2:

You've been given a dataset with information about customers of an online retail store. The dataset contains features such as age, gender, location, purchase history, and the amount spent on previous purchases. The goal is to build a machine learning model that can predict whether a new customer will make a purchase within the next week.

How would you approach this problem?

What steps would you take to build an effective predictive model?

To approach this problem, I'd start by diving into the dataset to understand its structure, identify any missing values, and explore trends or patterns in customer behavior. Once I have a clear picture, I'd preprocess the data by handling missing values, converting categorical variables into numerical formats, scaling numerical features, and creating new features that might be useful, such as the total amount spent by a customer or the time since their last purchase. If the dataset is imbalanced, I'd address this by using techniques like oversampling the minority class, undersampling the majority class, or adjusting class weights during model training.

After preprocessing, I'd split the data into training and testing sets. Then, I'd experiment with different machine learning algorithms, such as logistic regression, decision trees, random forests, or gradient boosting models like XGBoost or LightGBM. I'd fine-tune the hyperparameters of these models to optimize their performance. To evaluate the models, I'd use metrics like accuracy, precision, recall, F1-score.