### 1. What is Logistic Regression?

**Logistic Regression is a statistical method for binary classification. It predicts the probability that an input belongs to a certain class using the logistic (sigmoid) function.**

---

### 2. What kind of problem does this code solve?

**This code solves a binary classification problem: predicting whether a user purchases a product or not based on age and estimated salary.**

---

### 3. What is the sigmoid function?

**The sigmoid function maps input values to a range between 0 and 1. It helps interpret model outputs as probabilities in logistic regression.**

---

### 4. Why did we use StandardScaler?

**We use `StandardScaler` to normalize the feature values so they have a mean of 0 and a standard deviation of 1. This improves model convergence and accuracy.**

---

### 5. What is a confusion matrix?

**A confusion matrix is a 2x2 table for binary classification problems showing true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).**

---

### 6. What do TP, FP, TN, FN mean?

- **TP: Correctly predicted positives**
- **FP: Incorrectly predicted positives**

- **TN: Correctly predicted negatives**
- **FN: Incorrectly predicted negatives**

---

### 7. What is accuracy?

**Accuracy is the ratio of correctly predicted samples (TP + TN) to total predictions. It shows how often the classifier is correct.**

---

### 8. How do you calculate precision?

**Precision = TP / (TP + FP)**
**It indicates how many of the predicted positives are actually positive.**

---

### 9. How do you calculate recall (sensitivity)?

**Recall = TP / (TP + FN)**
**It measures how many actual positives were correctly identified.**

---

### 10. What is the error rate?

**Error Rate = (FP + FN) / Total predictions**
**It represents the proportion of incorrect predictions.**

---

### 11. What is specificity?

**Specificity = TN / (TN + FP)**
**It measures how well the model identifies negatives.**

---

### 12. Why is logistic regression suitable here?

**It is suitable because the target variable is binary — purchase (1) or not (0) — which is ideal for logistic regression.**

---

### 13. What are the independent and dependent variables?

**Independent: Age and Estimated Salary**
**Dependent: Purchased (0 or 1)**

---

### 14. What does the `.fit()` function do?

**`.fit()` trains the logistic regression model using the training data.**

---

### 15. Why did we use `train_test_split()`?

**To divide the dataset into training and test sets, ensuring the model is trained and evaluated on different data for unbiased performance.**

---

### 16. What is the role of `random_state=42`?

**It ensures reproducibility of the train-test split by setting a seed for the random number generator.**

---

### 17. What evaluation metrics did we use?

**Accuracy, Precision, Recall, Error Rate, Specificity, F1-Score, and Confusion Matrix were used for evaluation.**

---

### 18. What does `.predict()` do?

**`.predict()` uses the trained model to predict the class labels for the test data.**

---

### 19. What is F1-score?

**The F1-score is the harmonic mean of precision and recall. It balances both metrics in one score.**

---

### 20. What does `.transform()` do in StandardScaler?

**It applies the learned scaling (mean and std) to the dataset, transforming it to a standardized format.**

---

### 21. Why is scaling important in logistic regression?

**Because logistic regression is sensitive to the scale of features. Unscaled features can dominate others and affect predictions.**

---

### 22. What solver did we use in Logistic Regression?

**We used `lbfgs`, a robust solver suitable for small to medium datasets.**

---

### 23. How does Logistic Regression handle non-linearity?

**Logistic regression assumes a linear relationship between features and the log-odds. It doesn't handle non-linearity well without feature engineering.**

---

### 24. What is the output of a logistic regression model?

**The model outputs a probability score between 0 and 1, which is then thresholded (usually at 0.5) to classify into 0 or 1.**

---

### 25. Why do we visualize the confusion matrix?

**To quickly see how many predictions were correct vs. incorrect in each category, helping us assess model performance clearly.**