Analysis of Train Set Results

The classification report for the train set indicates the performance of the logistic regression model during training:

1. Class 0 (Unsubscribed):
    - Precision (0.95): The model is highly precise in predicting class `0`. When the model predicts a customer is unsubscribed, it is correct 95% of the time.
    - Recall (1.00): The model perfectly identifies all unsubscribed customers in the training set.
    - F1-score (0.98): The harmonic mean of precision and recall indicates excellent performance for class `0`.
2. Class 1 (Subscribed):
    - Precision (0.99): When the model predicts a subscription, it is correct 99% of the tim
    - Recall (0.60): The model identifies only 60% of actual subscribed customers. This is quite low compared to class `0`, indicating potential underfitting for this class.
    - F1-score (0.75): Lower compared to class `0`, showing an imbalance in the model's effectiveness across classes.
3. Accuracy (0.96):
    - The overall accuracy is very high, but this could be influenced by the class imbalance (more unsubscribed records than subscribed).
4. Macro Avg:
    - Precision (0.97) and Recall (0.80) indicate the model performs well overall but struggles with class `1`.
5. Weighted Avg:
    - These metrics (precision = 0.96, recall = 0.96) are weighted by class frequencies, reflecting the model’s strong performance on the dominant class (`0`).

Analysis of Test Set Results

The test set classification report reveals how well the model generalizes:

1. Class 0 (Unsubscribed):
    - Precision (0.95): The model maintains its high precision when predicting unsubscribed customers.
    - Recall (0.97): Slightly reduced compared to the train set but still excellent.
    - F1-score (0.96): Shows that the model generalizes well for class `0`.
2. Class 1 (Subscribed):
    - Precision (0.72): Lower than in the train set, but still reasonable for class `1`.

    - Recall (0.65): A moderate drop compared to the train set recall (0.60). This indicates the model struggles to identify subscribed customers in unseen data.
    - F1-score (0.68): The model's performance for class 1 has declined further in generalization.
3. Accuracy (0.93):
    - A slight drop compared to the train set accuracy (0.96), which is expected, but the high value indicates good generalization.
4. Macro Avg:
    - Precision (0.83) and Recall (0.81) indicate a significant disparity between the two classes.
5. Weighted Avg:
    - Precision (0.92) and Recall (0.93) suggest the model performs well overall, heavily influenced by class 0.

Observations:

The dataset is imbalanced, with many more 0 (unsubscribed) records than 1 (subscribed) records.This imbalance likely causes the model to favor class 0, as evidenced by much higher precision, recall, and F1 scores for this class.The recall for class 1 is consistently lower in both train and test sets, showing the model struggles to correctly identify customers likely to subscribe.This may stem from the imbalance and limitations of logistic regression in handling such disparities.The model generalizes well for class 0 but shows a slight drop in performance for class 1 when applied to the test set.There is no significant overfitting, as the metrics are reasonably consistent across the train and test sets. However, the model's weakness in identifying class 1 remains evident.