

Data Analytics Pathway Assessment

Report By: Selorm Etse-Forfoe

A. Objective

The goal of this project is to predict whether a client will subscribe to a term deposit using historical marketing and client data. This will be extremely helpful for marketing teams, to target the right clients more effectively, minimizing waste and increasing conversion rates.

B. Key Findings from EDA

1. **Call Duration:** This appeared to be the strongest predictor. Longer calls are associated with higher subscription rates.
2. **Job Type:** Retired and student clients are more likely to subscribe; blue-collar and services clients are less likely to subscribe.
3. **Month:** Subscription success rates vary by month. May had many contacts but low success; March, June, and December performed better.
4. **Previous Contact:** Clients contacted in earlier campaigns (`contacted_before_campaign = 1`) are more likely to subscribe.
5. **Pdays:** A majority of clients had never been previously contacted (`pdays = -1`). A useful feature when engineered into a binary variable.

C. Modeling Results (Logistic Regression)

1. **Accuracy:** 83%

Out of all predictions the model made, 83% were correct (both for subscribers and non-subscribers). However, since the data is imbalanced (only 8.5% of clients subscribed), a high accuracy can be misleading. If the model predicted "no" for everyone, it would still score over 90% accuracy without being very useful.

2. **Recall for "Yes" (subscribed):** 78%

Out of all clients who actually subscribed, the model correctly identified 78% of them. This is one of the most important metrics for your case. In marketing, you don't want to miss potential subscribers.

Before, recall was only 23%, meaning the model missed most of the actual subscribers, because of class imbalance. Now after solving the problem of imbalance, it only misses about 22%, which is a big improvement.

3. **Precision for "Yes": 30%**

Out of all clients predicted as “Yes” by the model, only 30% actually subscribed.

This means there are false positives (the model sometimes predicts someone will subscribe when they won't). That's not ideal, but in marketing, it's often okay to reach out to a few uninterested clients, as long as you're catching the interested ones (confirmation from high recall).

4. **F1 Score for "Yes": 0.44**

The model is now very effective at identifying potential subscribers, even with class imbalance, thanks to the use of `class_weight='balanced'`.

D. Actionable Insights and Recommendations for the Business

1. **Longer Conversations Matter:** Agents should be trained to keep prospects engaged for at least 200–300 seconds, as call duration strongly correlates with subscriptions.
2. **Target Specific Profiles:**
 - Focus on **retired, students, and administrative** job types.
 - Re-engage **previously contacted** clients, they are more likely to convert.
3. **Avoid Over-Contacting:** Clients contacted too frequently in a single campaign (campaign > 3) showed lower interest.
4. **Time Your Campaigns:** Launching campaigns in **March, June, or December** may yield better results than May.