

Advanced Analytics Intern Case Study

Case Study Title: Audience Scoring for Direct Mail Marketing

Objective

Your task is to predict whether a customer is likely to be influenced by a Direct Mail (DM) campaign to purchase an incremental amount (in dollars). You'll use transaction-level data to generate customer-level features and build a model to score customer likelihood of influence. You will deliver both your code and a concise summary presentation.

Dataset

You are provided with two CSV files:

 **Sales_Data.csv (transaction-level data)**

Column Name	Description
Gold_Cust_ID	Unique customer ID
Post_Cd	ZIP/postal code
Tran_Dt	Transaction date
Dept ID	Department ID
Dept Name	Department name
Division	Product division
Channel	Sales channel (e.g., Online, Retail)
Online Store	Online store ID or name
Itm_Id	Item ID
Class Code	Item class code
Class Name	Item class name
Brand Name	Product brand
Price Paid	Price paid (per unit)
Units	Quantity purchased
Shipping Cost	Cost of shipping
DMCouponInd	Indicator if a DM coupon was used
Return_Ind	Indicator if the item was returned

DM_Dates.csv (mailing data)

Column Name	Description
Gold_Cust_ID	Unique customer ID
DM Mail Date	Date the DM was sent (nullable if no DM)

Part 1: Modeling Task

- Explore and clean the data.
- Aggregate transaction-level data to customer-level features.
- Join the datasets using Gold_Cust_ID.
- Define your own target variable based on transaction behavior and the DM Mail Date. Clearly explain your assumptions about how influence is measured.
- Train a classification model to predict the likelihood of a customer being influenced by DM.
- Evaluate the model using appropriate performance metrics (e.g., AUC, F1-score, precision/recall).
- Score each customer with a probability of influence.

Part 2: Presentation (5–10 slides)

Include the following:

1. Business problem and goal interpretation
2. Key insights from data exploration
3. Approach to feature engineering and modeling
4. How you defined and interpreted the target
5. Model performance summary
6. Strategic recommendation for marketing activation
7. Limitations and potential improvements

Deliverables

- Python Notebook (.ipynb)
- Slide Deck (.pptx or .pdf)
- Optional: Supporting visualizations or summary reports

Evaluation Criteria

- Clarity and logic of approach
- Quality of feature engineering and modeling
- Thoughtful definition of target variable
- Communication of insights and decisions
- Relevance of recommendations to marketing context

Thank you, we look forward to seeing how you think!