

Marketing Campaign Analysis – Insight Summary

1. Project Objective

This project analyzes customer behavior and purchasing patterns to identify which customer segments are most likely to accept marketing campaigns. The focus is on data cleaning, exploratory data analysis (EDA), and translating findings into actionable business insights.

2. Dataset Overview

- Each row represents one customer.
- Target variable: Response (1 = accepted the campaign, 0 = did not accept).
- Key feature groups:
 - Demographic: Education, Marital_Status, Country, Year_Birth.
 - Behavioral: Recency, NumWebVisitsMonth, NumWebPurchases, NumStorePurchases, NumCatalogPurchases.
 - Monetary: Income, MntWines, MntFruits, MntMeatProducts, MntFishProducts, MntSweetProducts, MntGoldProds.
 - Campaign history: AcceptedCmp1–5, previous campaign acceptance.

3. Data Quality & Preprocessing

- Menghapus spasi pada nama kolom dan mengubah kolom tanggal (Dt_Customer) menjadi tipe datetime.
- Kolom Income berisi simbol dolar (\$) dan tanda koma, sehingga dibersihkan lalu dikonversi ke tipe float.
- Missing value pada Income diisi menggunakan median (lebih tahan terhadap outlier dibanding mean).
- Outlier ekstrem pada Year_Birth (< 1900) dihapus karena tidak realistis.
- Dibuat beberapa fitur turunan:
 - Customer_Age = Dt_Customer.year – Year_Birth
 - Num_Dependents = Kidhome + Teenhome
 - Dt_Customer_Year dan Dt_Customer_Month
 - TotalAmount_Spent = total pengeluaran dari seluruh kategori produk
 - Totalpurchases_feature = total frekuensi pembelian dari semua channel

4. Exploratory Data Analysis – Key Insights

- Education:
 - Customer dengan pendidikan tinggi (terutama PhD) memiliki proporsi penerimaan campaign yang lebih besar.
- Marital Status:
 - Customer yang dikelompokkan sebagai Single (termasuk kategori YOLO, Alone, Absurd) cenderung lebih responsif dibanding yang Married.
- Num_Dependents:
 - Customer tanpa tanggungan (Num_Dependents = 0) lebih banyak menerima campaign dibanding yang memiliki anak.
- Customer Tenure:
 - Customer yang lebih lama bergabung (Dt_Customer_Year lebih kecil) menunjukkan kecenderungan lebih tinggi menerima campaign.
- Recency:
 - Nilai Recency rendah (baru saja berbelanja) berkaitan dengan probabilitas lebih tinggi menerima campaign baru.
- Spending:
 - Customer dengan TotalAmount_Spent sekitar 1000–2000 dolar merupakan segmen yang paling

sering menerima campaign.

5. Model & SHAP (High-Level)

- Model yang digunakan: RandomForestClassifier sebagai model utama.
- Metrik evaluasi utama: ROC-AUC pada test set.
- Hasil SHAP menunjukkan fitur paling berpengaruh terhadap penerimaan campaign adalah:
 - Recency (pembelian terakhir yang lebih baru meningkatkan peluang respons).
 - Dt_Customer_Year (customer yang join lebih baru cenderung responsif di awal).
 - Riwayat penerimaan campaign sebelumnya (AcceptedCmp1, AcceptedCmp3, AcceptedCmp5).
 - TotalAmount_Spent dan Income (customer bernilai tinggi lebih mudah dikonversi).
 - Pembelian pada produk premium seperti MntMeatProducts dan MntGoldProds.

6. Business Questions & Answers

Q1. Produk apa yang paling sukses (revenue tertinggi)?

A1. Produk Wine memiliki total nilai pembelian tertinggi, diikuti oleh produk daging dan produk premium lainnya. Wine dapat diprioritaskan sebagai produk utama dalam campaign.

Q2. Channel mana yang underperform?

A2. CatalogPurchases merupakan channel dengan performa terendah dibanding web dan store. Channel ini dapat dievaluasi kembali efektivitasnya atau dialihkan budget-nya ke channel yang lebih efektif.

Q3. Channel mana yang paling baik performanya?

A3. StorePurchases memiliki rata-rata pembelian tertinggi di antara customer yang menerima campaign. Interaksi langsung di toko berpotensi besar untuk meningkatkan conversion.

Q4. Negara mana yang memiliki tingkat keberhasilan campaign tertinggi?

A4. Mexico dan Spain memiliki conversion rate tertinggi. Negara-negara ini dapat menjadi fokus utama untuk campaign berikutnya karena menunjukkan respon yang paling positif.

7. Recommendations

- Fokuskan campaign pada segmen berikut:
 - Customer dengan Recency rendah (baru berbelanja).
 - Customer dengan TotalAmount_Spent tinggi dan income menengah–tinggi.
 - Customer yang sebelumnya pernah menerima campaign (AcceptedCmp* = 1).
- Optimalkan channel toko (store) sebagai channel utama, sambil mengevaluasi efektivitas channel katalog.
- Prioritaskan negara dengan conversion rate tinggi seperti Mexico dan Spain, kemudian uji strategi serupa di negara lain.
- Gunakan insight demografis (Education dan Marital_Status) untuk menyusun pesan campaign yang lebih personal dan relevan.