

## Temuan Utama (Bagian 2)

### 1. Produk Terlaris

- Produk dengan pendapatan tertinggi adalah **Laptop Pro 15 (Rp 390M)** dan **Smartphone X (Rp 238M)**.
- Produk kategori elektronik mendominasi kontribusi revenue.

### 2. Pelanggan Paling Berharga

- Pelanggan **CUST005 (Jakarta)** menyumbang pengeluaran tertinggi: **Rp 111,6M**.
- Dari 10 pelanggan teratas, **8 berasal dari Jakarta**, menandakan Jakarta sebagai pasar utama.

### 3. Tren Penjualan Bulanan

- Terlihat **fluktuasi signifikan** setiap bulan.
- Pendapatan tertinggi terjadi pada **Desember 2023 (Rp 79,9M)** → indikasi kuat adanya efek musiman (seasonal) seperti liburan akhir tahun.
- Pendapatan cenderung **meningkat kembali pertengahan 2024**, menunjukkan potensi pemulihan pasar setelah bulan-bulan rendah di awal tahun.

## Pendekatan Data Science untuk Prediksi Churn (Bagian 3)

### 1. Definisi Churn

- Dalam konteks ini, churn didefinisikan sebagai **pelanggan yang tidak melakukan transaksi lagi dalam 3 bulan terakhir**, meskipun sebelumnya aktif bertransaksi.

### 2. Jenis Model yang Digunakan

- Model yang tepat adalah **klasifikasi biner (Churn / Tidak Churn)**.
- Titik awal: gunakan **Logistic Regression** (mudah diinterpretasikan untuk bisnis) atau **Random Forest** (lebih kuat menangkap interaksi antar fitur).

## Fitur untuk Prediksi Churn dari Kolom yang Ada

### 1. Recency & Aktivitas Waktu

- timestamp → dipakai untuk hitung jarak hari sejak transaksi terakhir.
- year, month, day, weekday, week\_number → pola pembelian per periode (seasonality).

### 2. Frequency & Basket Behavior

- quantity → jumlah item per transaksi.
- basket\_size → ukuran keranjang belanja (berapa banyak barang dibeli).

### 3. Monetary (Nilai Transaksi)

- total\_price → nilai transaksi per order.
- revenue\_band → kategorisasi customer (Low / Medium / High spender).
- is\_high\_value\_product → indikasi apakah sering beli produk mahal.

#### 4. Profil Customer

- customer\_id → ID unik (tidak langsung dipakai sebagai fitur, tapi untuk agregasi).
- customer\_location → lokasi customer (Jakarta, Surabaya, dll).
- join\_date → dipakai untuk menghitung masa bergabung.
- customer\_tenure\_days → lama customer sejak join (umur pelanggan).
- is\_new\_customer → indikator customer baru ( $\leq 30$  hari).

#### 5. Produk

- product\_category → kategori produk favorit (Elektronik, Pakaian, Buku, dll).
- product\_name → bisa diolah ke dalam kategori level atas (atau one-hot encoding).
- price → harga produk (berkorelasi dengan spending power).

#### Feature Engineering yang akan saya lakukan

- **Recency**: MAX(timestamp) per customer → hitung berapa hari sejak transaksi terakhir.
- **Frequency**: jumlah transaction\_id per customer.
- **Monetary**: SUM(total\_price) per customer.
- **Loyalty**: customer\_tenure\_days.
- **Product Preference**: proporsi transaksi berdasarkan product\_category.
- **Spending Behavior**: rata-rata basket\_size, proporsi transaksi dengan is\_high\_value\_product=1.