

# Customer Shopping Behavior Analysis

## Ringkasan Proyek

Proyek ini menganalisis perilaku belanja pelanggan menggunakan data transaksi dari 3.900 pembelian di berbagai kategori produk. Tujuan utama adalah untuk mengidentifikasi pola pengeluaran, segmen pelanggan, preferensi produk, dan perilaku berlangganan guna mendukung pengambilan keputusan strategis bisnis.

## Ringkasan Dataset

- Rows: 3,900
- Columns: 18
- Fitur Utama:
  - Demografi pelanggan (Usia, Jenis Kelamin, Lokasi, Status Langganan)
  - Rincian pembelian (Produk yang Dibeli, Kategori, Jumlah Pembelian, Musim, Ukuran, Warna)
  - Perilaku berbelanja (Diskon yang Digunakan, Kode Promo yang Digunakan, Pembelian Sebelumnya, Frekuensi Pembelian, Peringkat Ulasan, Jenis Pengiriman)
- Data yang Hilang: 37 nilai di kolom Review Rating

## Eksplorasi Data Analysis Menggunakan Python

Kami memulai dengan persiapan dan pembersihan data di Python:

- **Data Loading:** Mengimpor dataset menggunakan pandas.
- **Eksplorasi Awal:** Menggunakan df.info() untuk memeriksa struktur dan .describe() untuk statistik ringkasan.

	Customer ID	Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating	Subscription Status	Shipping Type	Discount Applied
count	3900.000000	3900.000000	3900	3900	3900	3900.000000	3900	3900	3900	3900	3863.000000	3900	3900	39
unique	Nan	Nan	2	25	4	Nan	50	4	25	4	Nan	2	6	
top	Nan	Nan	Male	Blouse	Clothing	Nan	Montana	M	Olive	Spring	Nan	No	Free Shipping	
freq	Nan	Nan	2652	171	1737	Nan	96	1755	177	999	Nan	2847	675	22
mean	1950.500000	44.068462	Nan	Nan	Nan	59.764359	Nan	Nan	Nan	Nan	3.750065	Nan	Nan	Nan
std	1125.977353	15.207589	Nan	Nan	Nan	23.685392	Nan	Nan	Nan	Nan	0.716983	Nan	Nan	Nan
min	1.000000	18.000000	Nan	Nan	Nan	20.000000	Nan	Nan	Nan	Nan	2.500000	Nan	Nan	Nan
25%	975.750000	31.000000	Nan	Nan	Nan	39.000000	Nan	Nan	Nan	Nan	3.100000	Nan	Nan	Nan
50%	1950.500000	44.000000	Nan	Nan	Nan	60.000000	Nan	Nan	Nan	Nan	3.800000	Nan	Nan	Nan
75%	2925.250000	57.000000	Nan	Nan	Nan	81.000000	Nan	Nan	Nan	Nan	4.400000	Nan	Nan	Nan
max	3900.000000	70.000000	Nan	Nan	Nan	100.000000	Nan	Nan	Nan	Nan	5.000000	Nan	Nan	Nan
Discount Applied	Promo Code Used	Previous Purchases	Payment Method	Frequency of Purchases										
3900	3900	3900.000000	3900	3900										
2	2	Nan	6	7										
No	No	Nan	PayPal	Every 3 Months										
2223	2223	Nan	677	584										
Nan	Nan	25.351538	Nan	Nan										
Nan	Nan	14.447125	Nan	Nan										
Nan	Nan	1.000000	Nan	Nan										
Nan	Nan	13.000000	Nan	Nan										
Nan	Nan	25.000000	Nan	Nan										
Nan	Nan	38.000000	Nan	Nan										
Nan	Nan	50.000000	Nan	Nan										

- Penanganan Data yang Hilang:** Memeriksa nilai null dan mengisi nilai yang hilang pada kolom Review Rating menggunakan nilai median rating dari setiap kategori produk.
- Standarisasi Kolom:** Mengubah nama kolom menjadi snake case untuk meningkatkan keterbacaan dan dokumentasi.
- Feature Engineering:**
  - Membuat kolom age\_group dengan mengelompokkan usia pelanggan.
  - Membuat kolom purchase\_frequency\_days dari data pembelian.
- Pemeriksaan Konsistensi Data:** Memverifikasi apakah discount\_applied dan promo\_code\_used; menghapus promo\_code\_used.
- Integrasi Database:** Menghubungkan skrip Python ke PostgreSQL dan memuat DataFrame yang telah dibersihkan ke dalam database untuk analisis SQL.

## Analisis Data menggunakan SQL (Business Transactions)

Kami melakukan analisis terstruktur di PostgreSQL untuk menjawab pertanyaan bisnis:

- Pendapatan Berdasarkan Jenis Kelamin** – Membandingkan total pendapatan yang dihasilkan oleh pelanggan pria versus wanita.

	gender text	revenue numeric
1	Female	75191
2	Male	157890

2. **Pengguna Diskon dengan High-Spending** – Mengidentifikasi pelanggan yang menggunakan diskon tetapi tetap menghabiskan lebih dari rata-rata jumlah pembelian.

	customer_id bigint	purchase_amount bigint
1	2	64
2	3	73
3	4	90
4	7	85
5	9	97
6	12	68
7	13	72
8	16	81
9	20	90
10	22	62
11	24	89

Total rows: 839    Query complete 00:00:00

3. **5 Produk Teratas Berdasarkan Rating** – Menemukan produk dengan peringkat ulasan rata-rata tertinggi.

	item_purchased text	Average Product Rating numeric
1	Gloves	3.86
2	Sandals	3.84
3	Boots	3.82
4	Hat	3.80
5	Skirt	3.78

4. **Perbandingan Jenis Pengiriman** – Membandingkan rata-rata jumlah pembelian antara pengiriman Standar dan Ekspres.

	shipping_type text	round numeric
1	Standard	58.46
2	Express	60.48

5. **Pelanggan Berlangganan vs. Non-Berlangganan** – Membandingkan rata-rata pengeluaran dan total pendapatan berdasarkan status berlangganan.

	subscription_status text	total_customers bigint	avg_spend numeric	total_revenue numeric
1	Yes	1053	59.49	62645.00
2	No	2847	59.87	170436.00

6. **Discount-Dependent Products** – Mengidentifikasi 5 produk dengan persentase pembelian diskon tertinggi.

	item_purchased text	discount_rate numeric
1	Hat	50.00
2	Sneakers	49.66
3	Coat	49.07
4	Sweater	48.17
5	Pants	47.37

7. **Segmentasi Pelanggan** – Mengelompokkan pelanggan ke dalam New, Returning, and Loyal segments berdasarkan riwayat pembelian.

	customer_segment text	Number of Customers bigint
1	Loyal	3116
2	New	83
3	Returning	701

8. **3 Produk Teratas per Kategori** – Mencantumkan produk yang paling banyak dibeli dalam setiap kategori.

	item_rank bigint	category text	item_purchased text	total_orders bigint
1	1	Accessories	Jewelry	171
2	2	Accessories	Sunglasses	161
3	3	Accessories	Belt	161
4	1	Clothing	Blouse	171
5	2	Clothing	Pants	171
6	3	Clothing	Shirt	169
7	1	Footwear	Sandals	160
8	2	Footwear	Shoes	150
9	3	Footwear	Sneakers	145
10	1	Outerwear	Jacket	163
11	2	Outerwear	Coat	161

9. **Repeat Buyers & Subscriptions** – Memeriksa apakah pelanggan dengan lebih dari 5 pembelian lebih cenderung berlangganan.

	subscription_status text	repeat_buyers bigint
1	No	2518
2	Yes	958

10. **Revenue by Age Group** – Menghitung kontribusi pendapatan total dari setiap kelompok usia.

	age_group	total_revenue
	text	numeric
1	Young Adult	62143
2	Middle-aged	59197
3	Adult	55978
4	Senior	55763

## Dashboard di Power BI

Akhirnya, kami membangun dashboard interaktif di Power BI untuk menampilkan wawasan secara visual.



## Rekomendasi Bisnis

- **Boost Subscriptions** – Promosikan manfaat eksklusif untuk pelanggan berlangganan.

- **Customer Loyalty Programs** – Berikan hadiah kepada pembeli berulang untuk memindahkan mereka ke segmen “Loyal”.
- **Review Discount Policy** – Seimbangkan peningkatan penjualan dengan pengendalian margin.
- **Product Positioning** – Sorot produk terlaris dan berperingkat tinggi dalam campaigns.
- **Targeted Marketing** – Fokus upaya pada high-revenue age groups dan pengguna express-shipping.