

E-Commerce Sales Performance Dashboard – Online Retail Dataset

Hello, saya Hanif Prasetyo, disini saya mencoba menganalisis E-Commerce Sales Performance Dashboard – Online Dataset.

1. Pendahuluan

Dataset *Online Retail* mencakup lebih dari 540 Ribu transaksi pada periode Desember 2010 – Desember 2011 dengan sekitar ± 4.027 produk unik. Dari semua produk itu kebanyakan dijual berupa *Giftware*, Dekorasi Rumah, Perlengkapan Sederhana hingga Mainan Anak-Anak dengan karakteristik harga unit yg relatif rendah namun volume penjualan tinggi.

Dengan variasi produk yang luas dan perilaku pelanggan yang beragam, perusahaan menghadapi sejumlah tantangan bisnis:

A. Terjadi kesulitan untuk mengidentifikasi kategori produk secara akurat karena mengandung banyak sekali kata yang tidak sesuai pada tempatnya seperti deskripsi produk yang tidak konsisten, ada data yang non-produk seperti biaya administrasi atau ongkos kirim serta transaksi return dan adjustment.

Dengan kondisi yang seperti ini kita harus membuat proses kategorisasi produk menjadi lebih kompleks dan sangat membutuhkan tahap data cleaning yang mendalam atau menyeluruh.

B. Program promosi yang relatif kecil (1-2% dari total transaksi) berupa potongan harga, komisi serta lain halnya. Kontribusi yang sangat rendah membuat efektivitas kegiatan promosi sulit untuk dievaluasi secara mencalam, hal ini menjadi indicator yang sangat penting untuk strategi awal pemasaran.

C. Return rate yang tinggi juga menambah tantangan bagi profitabilitas perusahaan. Khususnya seperti United Kingdom yang mendominasi return.

Project ini bertujuan untuk membangun sebuah *Dashboard* analisis penjualan yang menyajikan gambaran menyeluruh tentang performa produk, pola pembelian pelanggan, kontribusi promosi dan potensi risiko return.

Fokus utama analisis ini adalah **Mengoptimalkan Strategi Promosi Berdasarkan Pola Pembelian**, misalnya dengan mengidentifikasi waktu transaksi tertinggi, tren musiman, serta dampak promosi terhadap volume penjualan.

Dengan insight yang dihasilkan, manajemen dapat mengambil keputusan yang lebih tepat terkait strategi stok, perencanaan promosi, dan pengelolaan resiko return sehingga kinerja penjualan dapat ditingkatkan secara berkelanjutan.

2. Tujuan Analisis

Dari tujuan analisis ini dilakukan untuk mendukung pengambilan keputusan berbasis data, khususnya dalam **Mengoptimalkan Strategi Promosi Berdasarkan Pola Pembelian**. Berikut Rincian tujuannya:

A. Mengukur Kinerja Penjualan Secara Menyeluruh

Melalui KPI Utama: Total Sales, Total Quantity, Total Transaksi, Total Customer, Average Order Value (AOV), Return Rate, dan Promo.

- B. Menganalisis pola pembelian pelanggan berdasarkan dimensi waktu (harian, mingguan, bulanan, musiman, jam transaksi) serta dimensi geografis (country/region) untuk memahami kapan dan di mana transaksi paling tinggi terjadi.
- C. Mengevaluasi Dampak Promosi dan Strategi Harga
Melalui ini kita bisa dengan melihat distribusi harga, kontribusi promo terhadap total penjualan, dan hubungan promo dengan return.
- D. Mengidentifikasi Produk Berkinerja Tinggi & Return yang Tinggi,
Dengan mengidentifikasi ini kita bisa melihat seberapa sering diminati produk tersebut, termasuk produk terlaris, produk penyumbang revenue terbesar, serta produk dengan return tertinggi.
- E. Menyediakan Insight Visual Dalam Bentuk *Dashboard*
Dengan *Dashboard* yang terstruktur (KPI utama → tren & perilaku → harga & promo → return) sehingga mudah dipahami oleh manajemen dan dapat langsung digunakan sebagai dasar strategi promosi, pengelolaan stok, dan mitigasi risiko.

3. Data Understanding, Data Quality Issues, Data Cleaning

A. Data Understanding

Dataset Online Retail terdiri dari 541.909 baris transaksi dengan 8 kolom utama. Berikut penjelasan struktur data secara ringkas:

- 1) InvoiceNo
Nomor unik untuk setiap transaksi. Nilai dengan awalan “C” menandakan cancellation atau transaksi return.
- 2) StockCode
Kode produk unik. Sebagian besar merepresentasikan produk normal, namun terdapat sejumlah entri non-produk seperti ongkos kirim atau biaya administrasi.
- 3) Description
Nama produk. Berisi deskripsi teks, namun terdapat entri kosong maupun teks tidak valid yang memerlukan pembersihan.
- 4) Quantity
Jumlah unit produk yang dijual. Nilai negatif mengindikasikan pengembalian barang (return).
- 5) InvoiceDate
Tanggal dan waktu transaksi. Format datetime ini memungkinkan analisis berdasarkan hari, bulan, minggu, jam, dan pola musiman.
- 6) UnitPrice
Harga per unit produk. Sebagian besar bernilai normal (>1), tetapi terdapat harga nol, sangat rendah (<1), maupun negatif.
- 7) CustomerID
Identitas pelanggan. Tidak semua transaksi memiliki ID, sehingga sebagian data dianggap anonymous transaction.
- 8) Country
Negara asal pelanggan. Data mencakup lebih dari 30 negara, dengan konsentrasi terbesar di UK.

B. Data Quality Issues

Dari Dataset *Online Retail* mencakup lebih dari 540 Ribu transaksi menunjukkan bahwa hasil eksplorasi menemukan sejumlah data mengalami masalah kualitas data pada Dataset *Online Retail* antara lain:

- 1) InvoiceNo
Tidak semua berformat numerik. Sebagian diawali huruf “C”, yang menandakan cancellation atau transaksi return.
- 2) StockCode
 - i. Sebagian besar adalah kode produk reguler.
 - ii. Namun terdapat 18 entri anomali yang tidak merepresentasikan produk, misalnya: AMAZONFEE, BANK CHARGES, POST, gift_0001_10–50.
- 3) Description
Berisi teks nama produk, tetapi terdapat ±20 ribu entri anomali berupa nilai kosong, karakter tidak valid (“????”), atau teks yang tidak jelas.
- 4) Quantity
Mayoritas bernilai positif, namun ditemukan sejumlah nilai negatif, yang mengindikasikan barang dikembalikan (return) atau error pencatatan.
- 5) UnitPrice
 - i. Sebagian besar bernilai normal (>1).
 - ii. Namun terdapat harga nol (0.00), sangat rendah (<1), bahkan negatif.
- 6) CustomerID
 - i. Tidak semua transaksi memiliki ID pelanggan.
 - ii. Ditemukan banyak transaksi dengan nilai missing pada kolom ini, meski detail penjualan tetap tercatat.

C. Data Cleaning

- 1) Data Cleaning
Berdasarkan isu yang ditemukan pada tahap sebelumnya, dilakukan beberapa langkah pembersihan data untuk memastikan hasil analisis valid:
- 2) InvoiceNo
Transaksi dengan awalan “C” diidentifikasi sebagai cancellation atau return.

Langkah cleaning: ditandai dengan flag IsReturn = 1, sehingga bisa dianalisis terpisah dari transaksi normal.
- 3) StockCode
Ditemukan 18 entri anomali yang tidak merepresentasikan produk reguler (contoh: AMAZONFEE, BANK CHARGES, POST, gift_0001_xx).

Langkah cleaning: entri-entri ini dipisahkan dari dataset penjualan utama agar tidak memengaruhi perhitungan revenue.
- 4) Description
Terdapat ±20 ribu deskripsi yang kosong atau berisi teks tidak valid.

Langkah cleaning: dilakukan pencocokan berbasis StockCode untuk memperbaiki deskripsi. Jika tidak ditemukan padanan, baris tersebut ditandai sebagai Unknown Product.

5) Quantity

Nilai negatif dipertahankan karena merepresentasikan transaksi return.

Langkah cleaning: dibuat flag khusus agar transaksi return bisa dihitung secara terpisah dari transaksi normal.

6) UnitPrice

Nilai nol, sangat rendah (<1), atau negatif dianggap tidak wajar.

Langkah cleaning: entri tersebut ditandai sebagai AnomalyPrice, sehingga bisa dikecualikan dari analisis pendapatan.

7) CustomerID

Sejumlah besar transaksi tidak memiliki CustomerID.

Langkah cleaning: transaksi tanpa CustomerID tetap digunakan dalam analisis penjualan total, tetapi dikecualikan dari analisis perilaku pelanggan (misalnya customer segmentation atau retention).

D. Data Preparation

Setelah tahap cleaning, dilakukan proses *data preparation* untuk menambahkan variabel turunan dan transformasi data agar lebih siap dianalisis. Langkah-langkah utama meliputi:

1) **Perhitungan Nilai Penjualan (TotalSales)**

- Dibuat kolom baru:

$$TotalSales = Quantity \times UnitPrice$$

- Variabel ini menjadi metrik utama untuk analisis revenue per transaksi maupun agregasi penjualan.

2) **Identifikasi Transaksi Return (IsReturn)**

- Berdasarkan *InvoiceNo*:
 - Normal $\rightarrow 0$
 - Return (awalan "C") $\rightarrow 1$
- Flag ini memudahkan analisis tingkat return dan perbandingan dengan transaksi normal.

3) **Klasifikasi Harga (LabelPrice)**

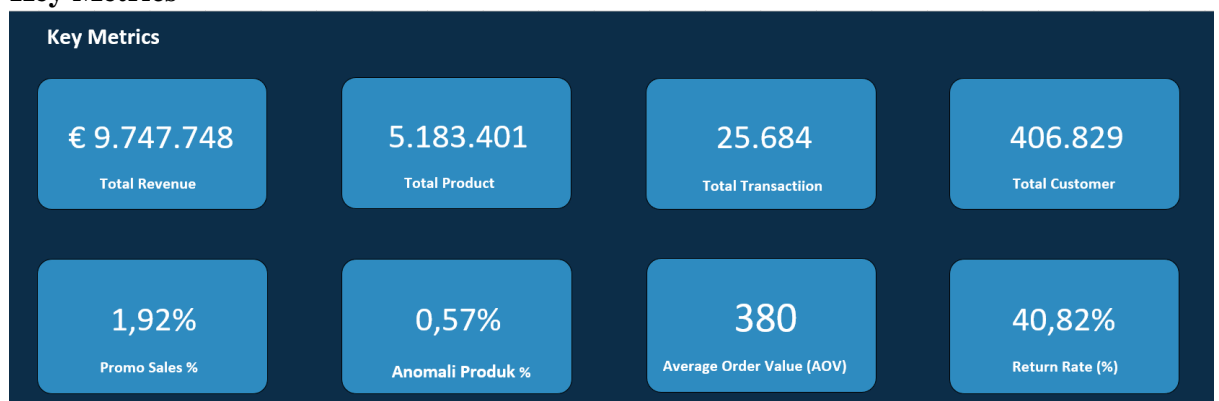
- UnitPrice dikelompokkan dalam kategori:
 - Free ($=0$)
 - Very Low (<0.05)
 - Low (<0.5)
 - Medium-Low (<1)

- Normal (>1)
- Kategori ini membantu analisis distribusi harga dan deteksi harga tidak wajar.
- 4) **Kategorisasi Jenis Transaksi (TransactionType)**
 - Berdasarkan *StockCode*:
 - Df_Clean Product: produk reguler
 - Df_Promo: entri diskon/voucher
 - Anomaly/Service: ongkos kirim, biaya bank, dll.
 - Dengan kategori ini, kontribusi promo maupun anomali bisa dipisahkan dari revenue murni.
- 5) **Ekstraksi Dimensi Waktu (Datetime Breakdown)**
 - Dari *InvoiceDate*, dibuat kolom tambahan:
 - Day
 - Month
 - Quarter
 - DayOfWeek (Senin–Minggu)
 - Hour (jam transaksi)
 - Transformasi ini memungkinkan analisis tren musiman dan pola pembelian pelanggan per jam.
- 6) **Customer Flag**
 - Transaksi dikelompokkan menjadi:
 - With CustomerID → pelanggan teridentifikasi.
 - Anonymous → pelanggan tanpa ID.
 - Hal ini penting untuk memisahkan analisis penjualan total dengan analisis perilaku pelanggan.

4. Dashboard Visual

Baris pertama dashboard menampilkan indikator utama performa penjualan, yang memberikan gambaran cepat sebelum masuk ke analisis detail.

A. Key Metrics



1) Total Revenue

- *Cara hitung*: $SUM(Quantity \times UnitPrice)$
- *Insight*: Menggambarkan nilai penjualan bersih dari seluruh transaksi.

2) Total Product

- *Cara hitung*: Jumlah unik **StockCode** (hanya produk reguler, tidak termasuk anomaly/promo).
- *Insight*: Memberikan gambaran variasi produk yang diperdagangkan.

3) Total Transaction

- *Cara hitung: Jumlah unik **InvoiceNo** (tidak termasuk invoice return).*
- *Insight: Metrik ini menunjukkan berapa banyak pesanan yang diproses.*

4) Total Customer

- *Cara hitung: Jumlah unik **CustomerID** (hanya baris dengan ID valid)..*
- *Insight: Menunjukkan seberapa luas basis pelanggan yang tercatat.*

5) Promo Sales %

- *Cara hitung: $(\text{Sales dari TransactionType} = \text{Promo} \div \text{Total Revenue}) \times 100\%$*
- *Insight: Promo hanya berkontribusi $\pm 2\%$ dari total penjualan, relatif rendah.*

6) Anomaly Product

- *Cara hitung: Jumlah baris dengan StockCode anomali (misalnya POST, BANK CHARGES, AMAZONFEE, dll..*
- *Insight: Menunjukkan jumlah transaksi non-produk yang perlu dipisahkan agar analisis penjualan akurat.*

7) Average Order Value (AOV)

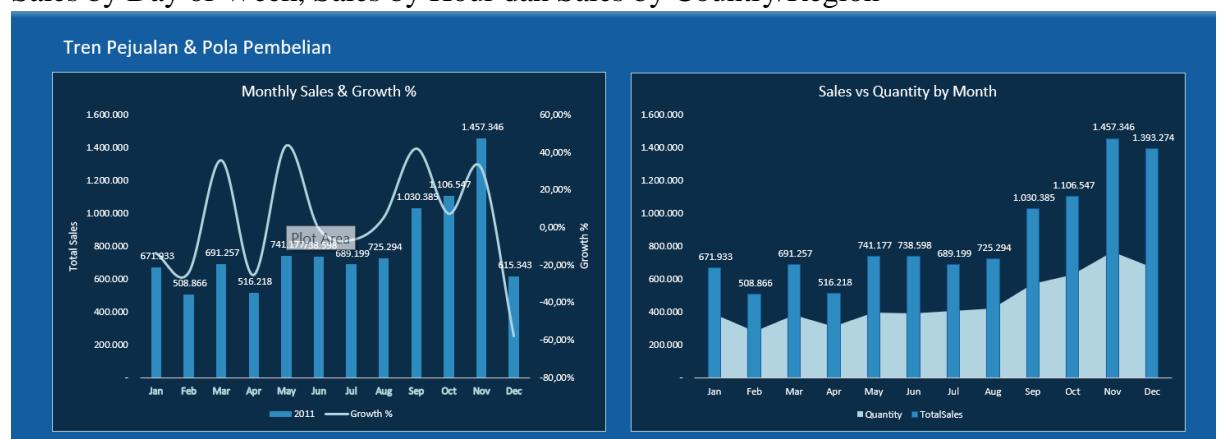
- *Cara hitung: $(\text{Total Quantity Return} \div \text{Total Quantity}) \times 100\%$*
- *Insight: Return rate relatif tinggi ($\pm 40\%$), menjadi risiko bagi profitabilitas*

8) Return Rate (%)

- *Cara hitung: Jumlah unik **StockCode** (hanya produk reguler, tidak termasuk anomaly/promo).*
- *Insight: Memberikan gambaran variasi produk yang diperdagangkan.*

B. Sales Trends and Buying Patterns

Baris ini berisi 5 grafik utama: Sales by Month/Year, Sales vs Quantity per Month , Sales by Day of Week, Sales by Hour dan Sales by Country/Region





1) Sales by Monty/Year

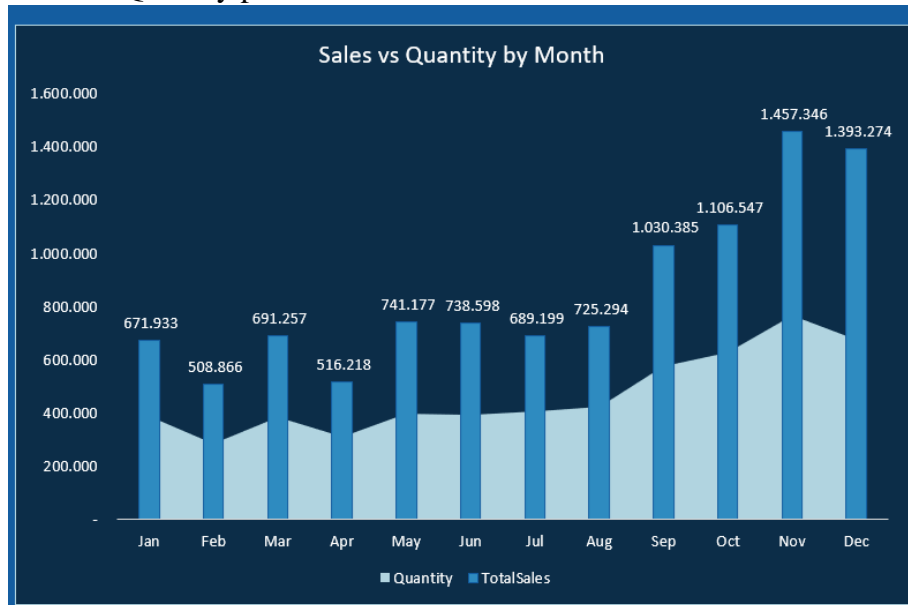


- *Cara hitung:* SUM(TotalSales) per Bulan untuk Total Bulanan dan untuk Growth (pertumbuhan) :

$$\frac{(\text{total penjualan bulan bulan ini} - \text{Total penjualan bulan sebelumnya})}{\text{Total Penjualan bulan ini}}$$

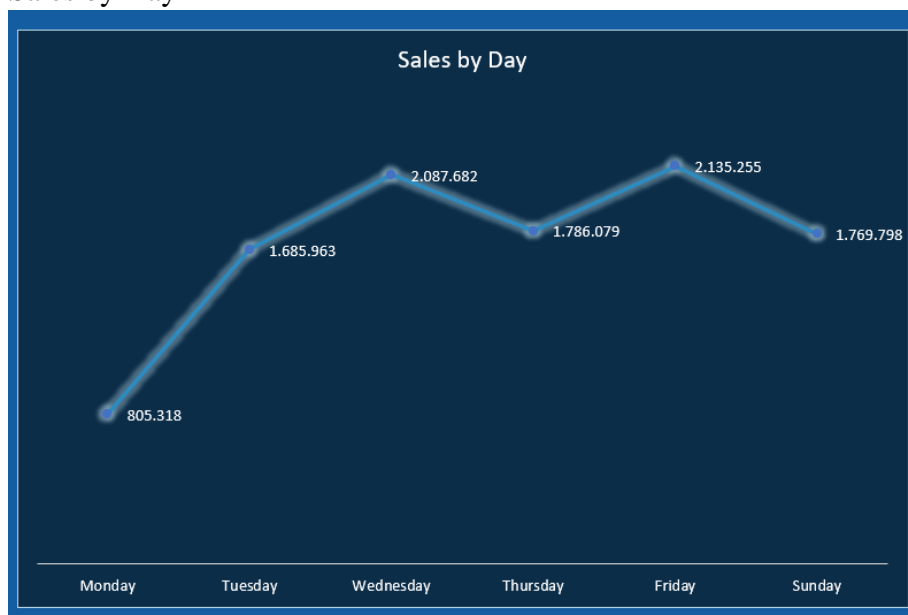
- *Insight:* Penjualan meningkat signifikan menjelang akhir tahun (Q4), mengindikasikan adanya pola musiman (*seasonality*).

2) Sales vs Quantity per Month



- *Cara hitung:* SUM(TotalSales) dan SUM(Quantity) per Bulan
- *Insight:* Ada bulan dengan *quantity* tinggi tapi *revenue* rendah → indikasi produk murah yang laku keras. Sebaliknya *quantity* rendah dengan *revenue* tinggi tapi indikasi produk *premium*.

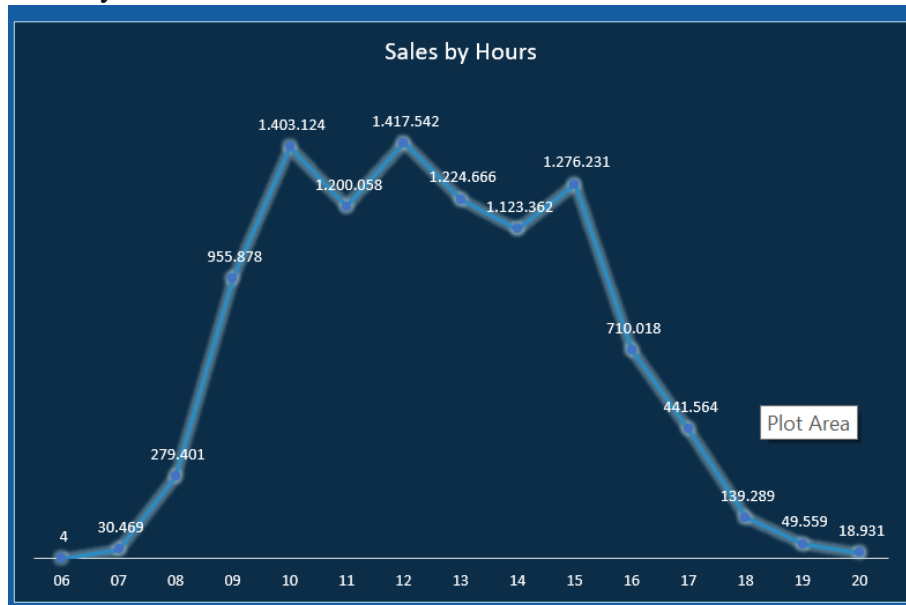
3) Sales by Day



Sales by Day of Week:

- *Cara hitung:* SUM(TotalSales) Dikelompokkan berdasarkan hari (Day of Week)
- *Insight:* Penjualan cenderung stabil sepanjang minggu dengan sedikit penurunan di akhir pekan.

4) Sales by Hours



- *Cara hitung:* SUM(TotalSales) Dikelompokkan berdasarkan Jam dari InvoiceDate (*Hours*)
- *Insight:* Aktivitas transaksi tertinggi terjadi jam 10–15 siang, membuka peluang promosi berbasis jam (*time-based campaign*).

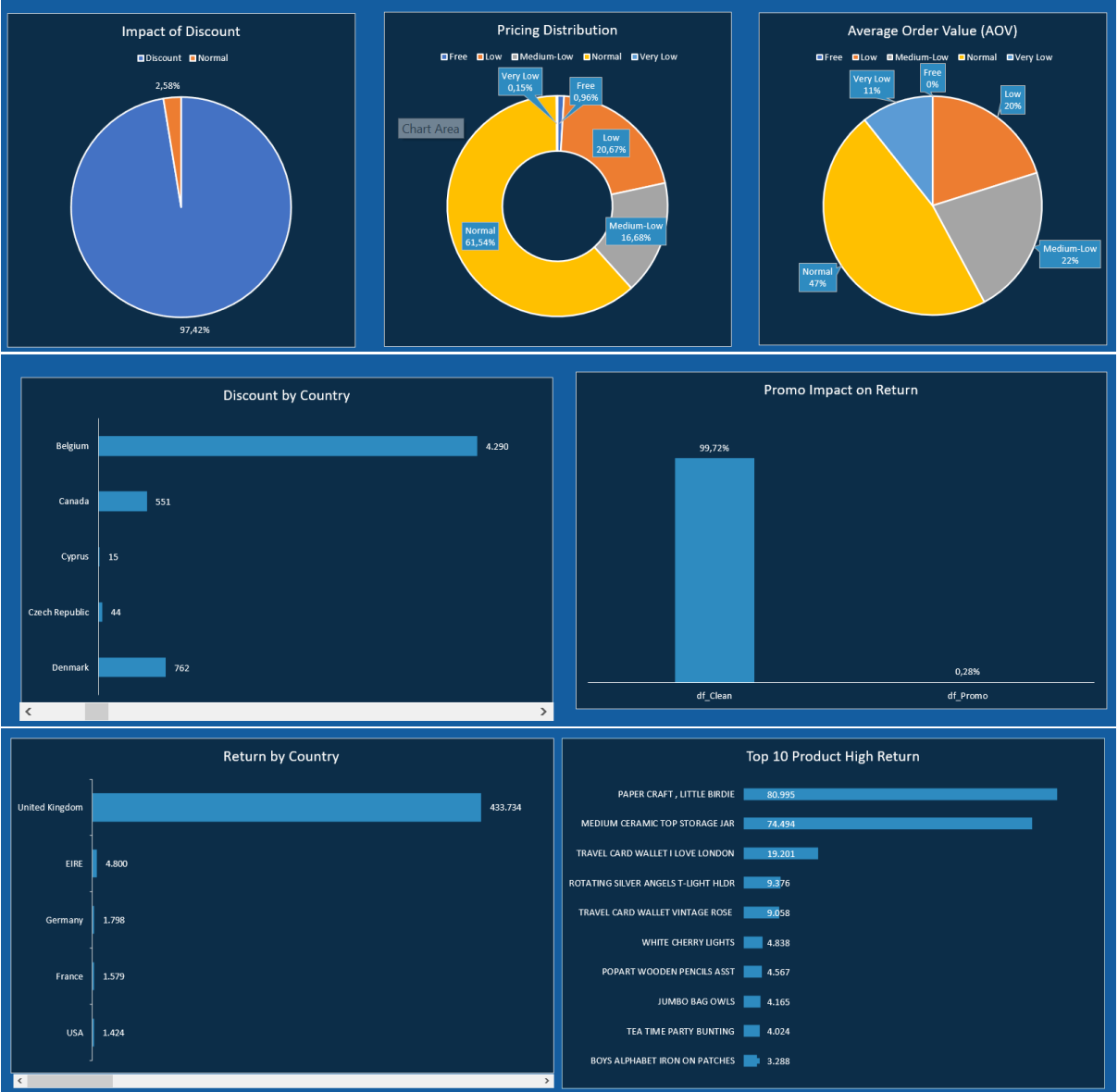
5) Sales by Country



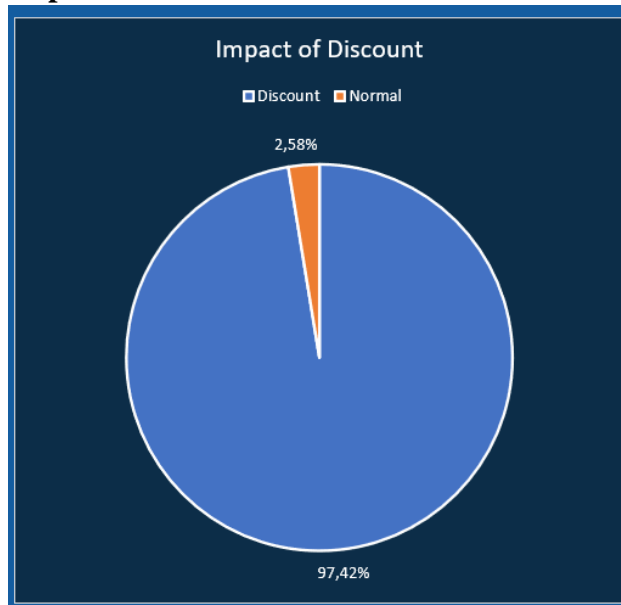
- *Cara hitung:* SUM(TotalSales) Per *Country*
- *Insight:* UK mendominasi, tapi lebih dari 30 negara lain juga berkontribusi, menunjukkan peluang ekspansi pasar internasional.

C. Discount, Return, Product

Berisi visualisasi distribusi harga, Kontribusi Promo, dan performa produk terlaris

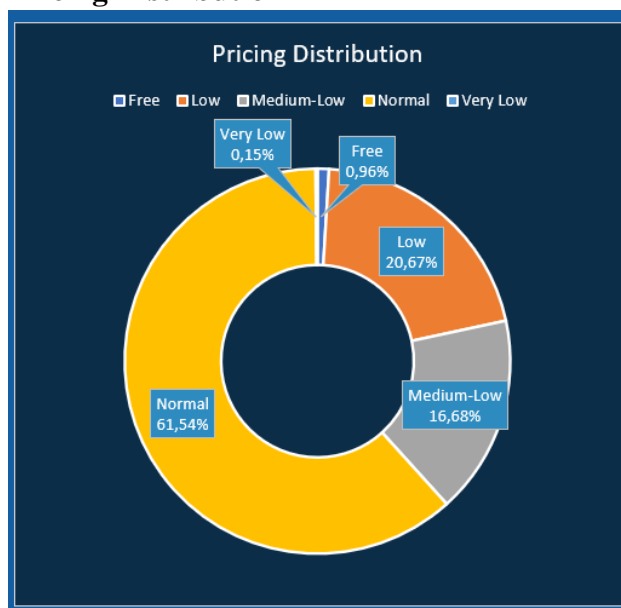


1) Impact of Discount



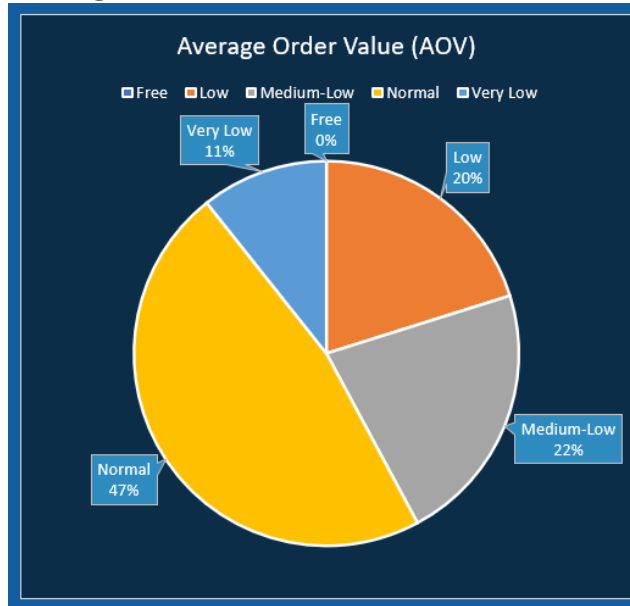
- *Cara hitung:* SUM(TotalSales) dibagi antara transaksi diskon (*Discount*) vs Non-Diskon (Normal)
- *Insight:* Diskon hanya berkontribusi $\pm 2-3\%$ dari total sales, menunjukkan promosi belum signifikan memengaruhi penjualan.

2) Pricing Distribution



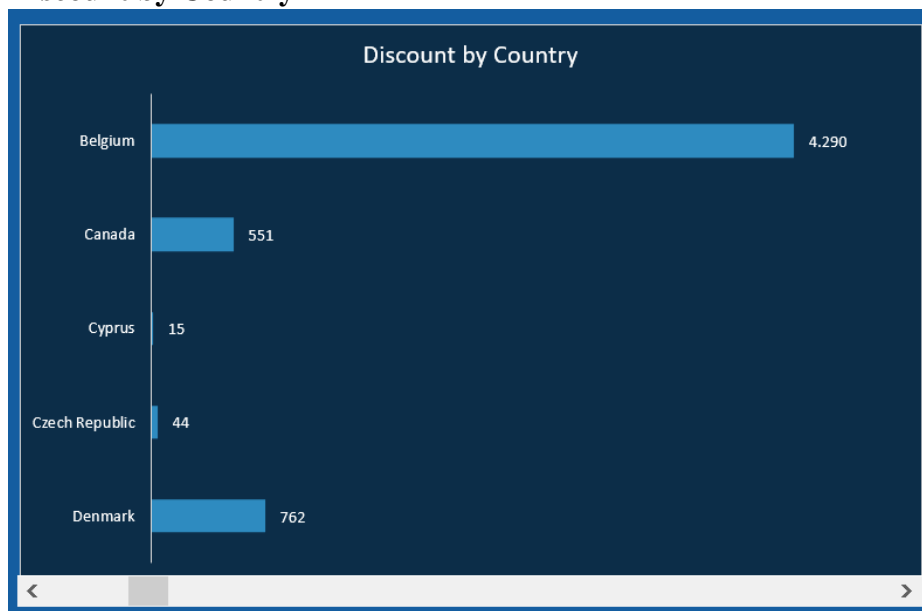
- *Cara hitung:* Isi distribusi UnitPrice (dikelompokkan dalam LabelPrice: Free, Low, Normal, dll.)
- *Insight:* Mayoritas produk dijual dengan harga normal (>1), sedangkan produk gratis/berharga sangat rendah hanya sebagian kecil.

3) Average Order Value (AOV)



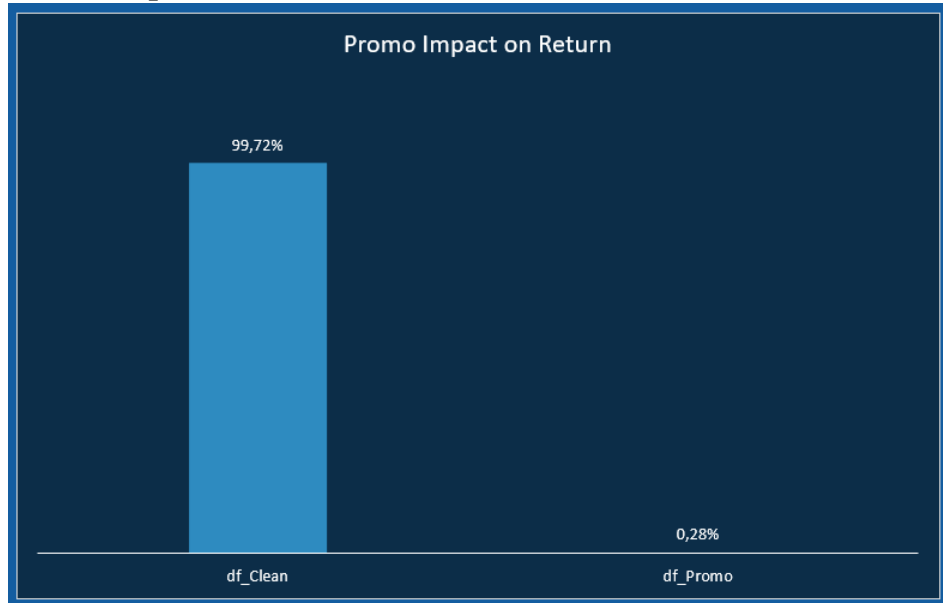
- *Cara hitung:* Isi distribusi $\frac{\text{TotalSales}}{\text{TotalTransaction}}$
- *Insight:* AOV relatif stabil di kisaran tertentu, dapat menjadi acuan strategi upselling/cross-selling.

4) Discount by Country



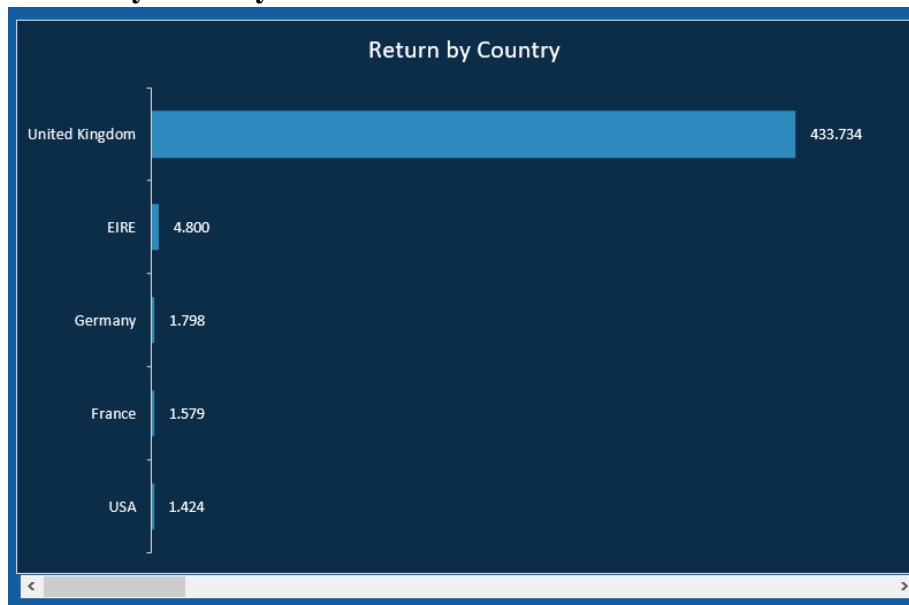
- *Cara hitung:* Isi distribusi Persentase promo Sales terhadap TotalSales per Negara (Country)
- *Insight:* Diskon paling banyak dimanfaatkan di pasar tertentu, tapi kontribusinya tetap kecil secara global.

5) Promo Impact on Return



- *Cara hitung:* Perbandingan return *rate* pada transaksi dengan Promo vs Non Promo (df_Promo vs df_Clean)
- *Insight:* Diskon paling banyak dimanfaatkan di pasar tertentu, tapi kontribusinya tetap kecil secara global.

6) Return by Country



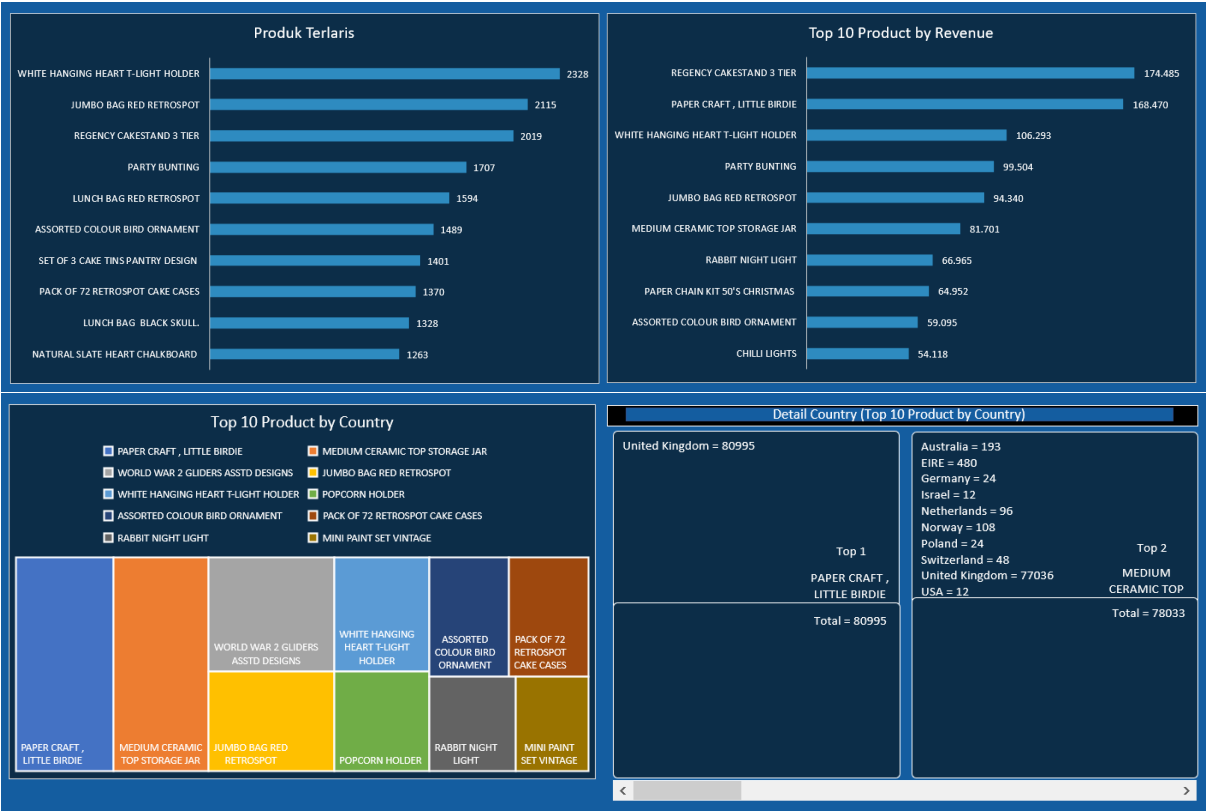
- *Cara hitung:* SUM(Quantity) dengan nilai negative per negara
- *Insight:* United Kingdom menyumbang return terbesar sekitar 433.734, menjadi fokus utama mitigasi risiko.

7) Top 10 Product High Return

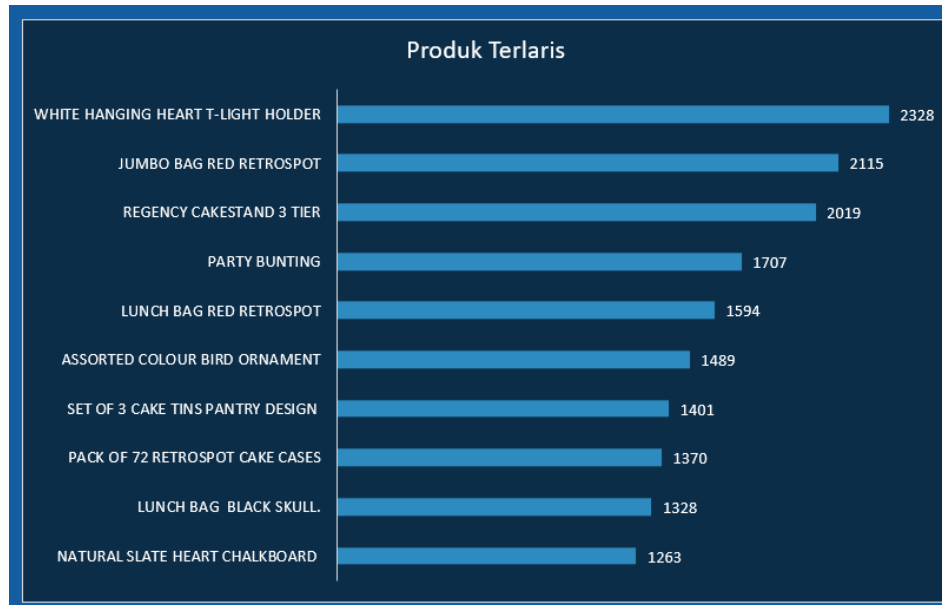


- Cara hitung: Produk dengan Total Return tertinggi berdasarkan Quantity
- *Insight:* Beberapa produk populer justru juga memiliki return tinggi, indikasi masalah kualitas atau mismatch ekspektasi.

D. Top Product

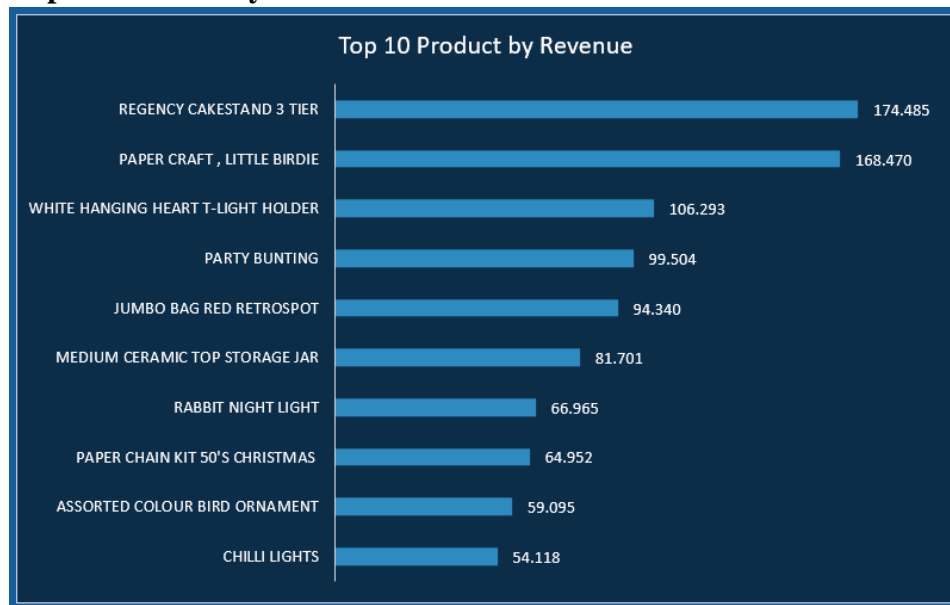


1) Produk Terlaris



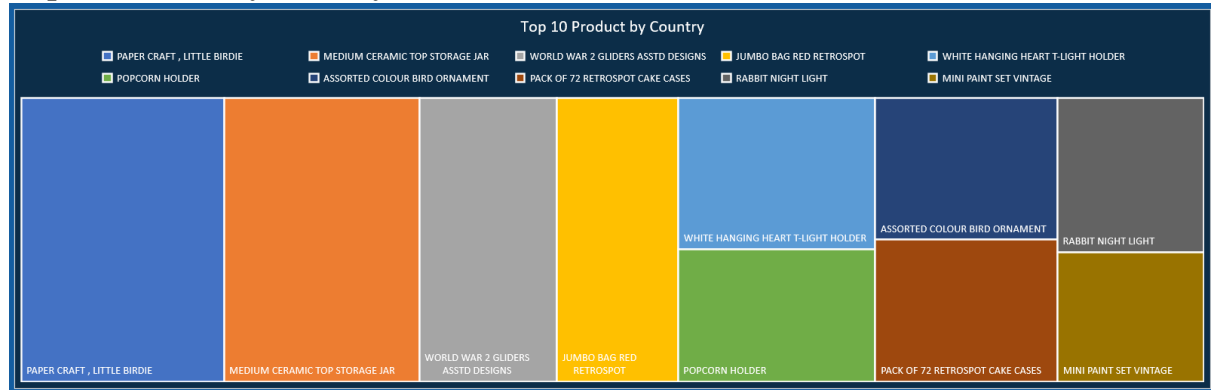
- Cara hitung: Produk dengan Quantity terjual terbanyak
- *Insight*: Produk murah dan mass-market mendominasi daftar ini.

2) Top 10 Product by Revenue



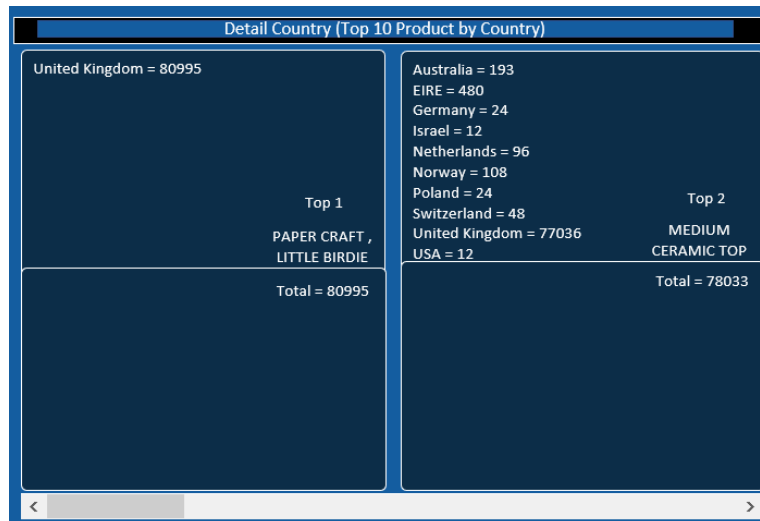
- Cara hitung: Produk dengan TotalSales Tertinggi
- *Insight*: Produk premium meski terjual lebih sedikit tetap berkontribusi besar ke revenue.

3) Top 10 Product by Country



- Cara hitung: Produk terlaris di tiap negara berdasarkan Sales/Quantity
- *Insight*: Preferensi produk berbeda-beda, misalnya dekorasi lebih populer di UK, sedangkan produk mainan/DIY lebih menonjol di negara lain.

4) Detail Country (Top 10 Product by Country)



Cara mencari = Produk terlaris di setiap negara dan detail dari negara apa aja yang penikmat produk tersebut

Insight = setiap produk berbeda-beda preferensi masing-masing negara, ada yang menyukai 1 produk seperti UK yang mendominasi banyaknya membeli 1 produk

5. Key Result & Insight

A. Sales Performance

- 1) Total penjualan mencapai sekitar €9,7 juta (jika di konversikan ke Dollars ialah 11.421.192,64) dari lebih dari 25 ribu transaksi dengan 5,1 juta produk terjual.
- 2) Pola penjualan menunjukkan bahwa produk berharga murah mendominasi dari sisi jumlah unit, sementara produk premium menjadi penyumbang terbesar revenue.
- 3) Average Order Value (AOV) sekitar €380, mengindikasikan rata-rata nilai keranjang belanja pelanggan relatif moderat.

B. Customer & Market Trends

- 1) Terdapat sekitar 407 ribu pelanggan unik, meskipun sebagian besar transaksi tidak memiliki CustomerID sehingga tercatat sebagai anonymous transactions.
- 2) Pasar terbesar berasal dari United Kingdom, namun lebih dari 30 negara lain turut berkontribusi terhadap penjualan.
- 3) Terdapat pola musiman, di mana penjualan mencapai puncak pada September–Desember).
- 4) Pola harian menunjukkan aktivitas belanja tertinggi pada pukul 10.00–15.00, sehingga berpotensi menjadi target promosi berbasis waktu.

C. Promotions & Pricing

- 1) Promosi hanya berkontribusi sekitar 1–3% dari total penjualan, sehingga efektivitasnya masih sangat terbatas.
- 2) Distribusi harga memperlihatkan sebagian besar produk dijual dengan harga normal ($> \text{€}1$), sementara produk gratis atau berharga sangat rendah hanya muncul dalam jumlah kecil.
- 3) Diskon bervariasi antar negara, namun kontribusinya secara keseluruhan tetap rendah.

D. Returns & Risks

- 1) Tingkat return sangat tinggi, Return Rate sekitar 40%, didominasi oleh UK dan EIRE.
- 2) United Kingdom merupakan negara dengan kontribusi return terbesar.
- 3) Beberapa produk terlaris juga masuk dalam kategori produk dengan return tinggi, mengindikasikan adanya isu kualitas atau ketidaksesuaian ekspektasi pelanggan.

E. Top Products

- 1) Produk dekorasi dan giftware berharga murah mendominasi dari sisi kuantitas penjualan.
- 2) Produk premium, meskipun lebih sedikit terjual, memberikan kontribusi signifikan terhadap revenue.
- 3) Preferensi produk berbeda di tiap negara, menegaskan perlunya strategi penjualan yang lebih tersegmentasi secara geografis.

6. Business Impact

A. High Return Rate (~40%)

Return tinggi (~40%) menggerus profit margin karena biaya logistik & pengelolaan stok meningkat. Jika dibiarkan, bisa menurunkan kepuasan pelanggan dan brand trust. Tingkat return yang sangat tinggi berpotensi menurunkan profitabilitas karena biaya logistik, pengelolaan stok, dan layanan pelanggan meningkat signifikan. Jika tidak ditangani, margin keuntungan akan terus tergerus meskipun volume penjualan besar.

B. Low Promotion Effectiveness

Promosi yang hanya berkontribusi $\pm 1\text{--}3\%$ terhadap total penjualan menunjukkan bahwa strategi pemasaran saat ini tidak optimal. Hal ini berarti peluang peningkatan penjualan melalui promosi masih belum dimanfaatkan sepenuhnya.

C. Time-based and Seasonal Patterns

Adanya pola transaksi tertinggi pada jam 10–15 (transaksi paling tertinggi di jam 10, 12, 15) dan lonjakan besar di Bulan September - Desember membuka peluang untuk

strategi promosi berbasis waktu dan musiman. Jika dimanfaatkan, hal ini dapat meningkatkan efektivitas promosi dan memperkuat perencanaan stok.

D. Revenue Concentration on Premium Products

Produk premium memberikan kontribusi besar terhadap revenue meskipun volume penjualannya lebih kecil. Dampaknya, perusahaan memiliki peluang untuk meningkatkan profitabilitas melalui strategi *upselling* atau *cross-selling* pada produk bernilai tinggi.

E. Geographic Variation

Preferensi produk berbeda di tiap negara, dengan UK sebagai pasar dominan dan UK juga sebagai sumber return terbesar dilanjut dengan EIRE untuk posisi kedua sumber return terbesar. Ini menandakan perlunya strategi lokal (*localized strategy*) untuk memaksimalkan potensi pasar internasional sekaligus mengurangi risiko return.

7. Conclusion & Recommendations

A. Conclusion

- 1) Analisis menunjukkan bahwa bisnis retail ini memiliki karakteristik volume tinggi namun margin tipis, dengan dominasi produk murah yang laku keras.
- 2) Produk premium meskipun jumlahnya lebih sedikit, berkontribusi besar terhadap revenue dan menjadi pendorong profitabilitas.
- 3) Tingkat return yang sangat tinggi ($\pm 40\%$) menjadi tantangan serius, terutama di pasar UK dan EIRE, yang berpotensi menekan margin keuntungan.
- 4) Aktivitas transaksi terbukti dipengaruhi oleh pola musiman (September – Desember) dan pola harian (jam 10–15), sehingga terdapat peluang promosi berbasis waktu.
- 5) Promo kontribusinya masih rendah ($\pm 2\%$), sehingga dampaknya terhadap peningkatan penjualan belum signifikan.

B. Recommendations

Berdasarkan hasil analisis, beberapa Langkah strategis yang direkomendasikan:

- 1) Return Management
Fokus untuk penurunan *return rate* di berbagai, terutama di pasar UK & EIRE dengan cara meningkatkan kontrol kualitas, memperjelas deskripsi produk, serta evaluasi kebijakan retur.
- 2) Promotion Strategy
 - Mengembangkan promosi berbasis waktu, khususnya pada jam sibuk (10–15 siang) dan periode akhir tahun (September - Desember) untuk memaksimalkan momentum penjualan.
 - Mengarahkan promosi ke produk dengan margin lebih tinggi untuk meningkatkan profitabilitas.
- 3) Product Portfolio Optimization
 - Mempertahankan produk murah sebagai volume driver, namun meningkatkan proporsi produk premium untuk menyeimbangkan revenue.

- Mengurangi ketergantungan pada produk dengan return tinggi meskipun penjualannya besar.
- 4) Market Expansion & Segmentation
 - Mengoptimalkan potensi pasar internasional dengan menyesuaikan strategi produk sesuai preferensi tiap negara.
 - Mengembangkan strategi lokal (localized strategy) untuk meningkatkan penetrasi pasar non-UK.
 - 5) Customer Data Improvement

Mendorong pencatatan CustomerID pada setiap transaksi untuk memperkuat analisis perilaku pelanggan, segmentasi, dan strategi retensi jangka panjang.

8. Sumber data & referensi

A. Data Source

Online Retail Dataset tersedia secara public melalui website Kaggle (Online-Retail). Dataset ini berisi Catatan transaksi Retail And Shopping di area UK pada Desember 2010 – Desember 2011, dengan ± 541.000 data dan ± 4.000 product.

Berikut Link Dataset Online Retail

<https://www.kaggle.com/datasets/tunguz/online-retail>

B. References

Dalam proses analisis, terdapat beberapa referensi yang digunakan sebagai acuan untuk memahami struktur dataset (misalnya arti kolom, jenis transaksi, dan karakteristik data). Referensi ini tidak sepenuhnya diikuti, melainkan digunakan sebagai bahan tambahan untuk memperkaya pemahaman.

Berikut referensi:

Holland, Sean Luke. Online Retail Data Analysis. Medium

<https://medium.com/@seanlukeholland/online-retail-data-analysis-75f3e443828c>