

Analisis *Outlier* Transaksi Retail Rumah Tangga Menggunakan Segmentasi Multidimensi dan Uji *Kruskal-Wallis*

Kiki Mustaqim¹⁾, Dini Septiana²⁾, Selma Nabila Razani³⁾, Woro Isti Rahayu⁴⁾

^{1,2,3,4)} Universitas Logistik dan Bisnis Internasional (ULBI)

Received: 27 July 2025

Accepted: 01 February 2026

Published: 1 June 2026



*diniseptiana239@gmail.com

Kata Kunci: *Outlier*, Retail, Segmentasi Multidimensi, *Interquartile Range*, *Kruskal-Wallis*.

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak : *Outlier* pada data transaksi penjualan sering kali diabaikan atau dihapus tanpa analisis lebih lanjut, padahal keberadaannya berpotensi menyimpan informasi penting terkait perilaku konsumen dan pola penjualan ekstrem. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang memengaruhi kemunculan transaksi *Outlier* pada data penjualan retail peralatan rumah tangga menggunakan pendekatan eksploratif dan segmentasi multidimensi. Identifikasi *Outlier* dilakukan pada variabel *Revenue* menggunakan metode *Interquartile Range* (IQR) ekstrem dengan konstanta $k = 7,5$, yang dirancang untuk memfokuskan analisis pada transaksi dengan nilai yang sangat ekstrem. Data kemudian disegmentasi berdasarkan perilaku konsumen (*Frequency* dan *Monetary*), kategori harga produk, wilayah pembelian berdasarkan benua, serta dimensi waktu berdasarkan hari dan bulan. Untuk menguji perbedaan distribusi *Revenue* antar segmen, digunakan uji *Kruskal-Wallis* sebagai metode non-parametrik yang sesuai untuk data yang tidak berdistribusi normal dan mengandung nilai ekstrem. Hasil penelitian menunjukkan bahwa terdapat perbedaan signifikan dalam distribusi *Revenue Outlier* antar segmentasi customer behavior, kategori harga produk, dan wilayah geografis. Sementara itu, dimensi waktu berdasarkan hari tidak menunjukkan perbedaan signifikan, namun segmentasi berdasarkan bulan menunjukkan perbedaan signifikan dalam distribusi *Revenue Outlier*. Temuan ini memberikan wawasan penting bagi pengelola retail dalam memahami karakteristik transaksi ekstrem serta menyusun strategi pemasaran dan manajemen risiko yang lebih efektif.

PENDAHULUAN

Seiring dengan pertumbuhan populasi, perubahan gaya hidup, dan perkembangan teknologi, permintaan terhadap produk-produk rumah tangga terus meningkat, mendorong keberagaman produk yang ditawarkan di pasar. Dalam kondisi tingginya volume transaksi yang terjadi setiap hari, data penjualan menjadi aset yang sangat bernilai [1]. Data ini dapat dimanfaatkan untuk memahami perilaku konsumen, memprediksi tren penjualan, dan merumuskan strategi pemasaran yang lebih tepat sasaran.

Namun, dalam pengolahan data transaksi sering kali ditemukan nilai-nilai yang berada jauh dari distribusi data normal, yang dikenal sebagai *Outlier*. Keberadaan *Outlier* ini tidak hanya dapat menyebabkan distribusi data menjadi tidak normal, tetapi juga berpotensi memengaruhi akurasi dalam pengambilan kesimpulan atau keputusan pada suatu penelitian [2]. Di sisi lain, nilai-nilai tersebut dapat menyimpan informasi penting mengenai pola perilaku konsumen ekstrem yang sering terlewatkan dalam analisis konvensional. Sebagai contoh, satu transaksi pembelian peralatan rumah tangga dalam jumlah besar oleh satu pelanggan, meskipun jarang terjadi, dapat menghasilkan nilai *Revenue* yang jauh lebih tinggi dibandingkan transaksi lainnya dan mencerminkan adanya segmen pelanggan bernilai tinggi yang strategis bagi perusahaan retail.

Pentingnya pemahaman mengenai *Outlier* mendorong perlunya penelitian lebih lanjut terhadap transaksi-transaksi yang mengandung nilai ekstrem tersebut. Salah satu metode yang efektif untuk mendeteksi *Outlier* adalah *Interquartile Range* (IQR) [3]. Metode ini sederhana namun dapat menangkap nilai-nilai ekstrem dalam distribusi data, terutama dalam kasus *Outlier* univariate, yang sangat relevan dalam konteks analisis data retail.

Selanjutnya, analisis data dapat dilakukan melalui segmentasi multidimensi. Segmentasi multidimensi dilakukan berdasarkan Frequency dan Monetary untuk memahami karakteristik konsumen yang berpotensi melakukan transaksi ekstrem [4]. Faktor-faktor lain seperti kategori harga produk (murah, sedang, mahal), wilayah pembelian (berdasarkan benua), serta dimensi waktu (hari dan bulan) juga dianalisis untuk mengungkap pola-pola terkait kemunculan transaksi *Outlier*. Dimensi waktu digunakan untuk mengidentifikasi pola musiman transaksi ekstrem.

Untuk menganalisis perbedaan antar kelompok, Uji Kruskal-Wallis dipilih sebagai metode non-parametrik yang efektif dalam membandingkan dua kelompok atau lebih tanpa mengharuskan data berdistribusi normal [5]. Pemilihan metode ini didasarkan pada fakta bahwa distribusi data yang tidak normal dan mengandung nilai ekstrem membuat uji non-parametrik lebih sesuai. Berdasarkan prinsip robust statistics, uji Kruskal-Wallis lebih tahan terhadap skewness dan *Outlier*, karena metode ini berbasis pada peringkat, bukan nilai rata-rata, sehingga menghasilkan estimasi yang lebih akurat terhadap populasi data yang heterogen.

Selain itu, analisis eksplorasi data (Exploratory Data Analysis/EDA) dilakukan untuk memvisualisasikan distribusi dan pola dalam setiap segmen. Teknik visualisasi seperti boxplot, histogram, dan scatter plot digunakan untuk menggali insight awal dan mendukung interpretasi hasil uji statistik. Melalui kombinasi EDA dan uji Kruskal-Wallis pada setiap kelompok, penelitian ini diharapkan dapat memetakan faktor-faktor yang memengaruhi transaksi *Outlier* dan memberikan kontribusi praktis bagi pengelola retail dalam merumuskan strategi bisnis yang lebih efektif. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam memahami karakteristik transaksi *Outlier* serta mendukung perumusan strategi bisnis retail yang lebih efektif.

TINJAUAN LITERATUR

Tinjauan literatur pada penelitian ini menunjukkan bahwa analisis *Outlier* dalam data transaksi retail tidak hanya berfokus pada aspek teknis deteksi anomali, tetapi juga pada pemahaman konteks bisnis yang melatarbelakanginya.

1. *Outlier* dalam Data Transaksi

Outlier atau nilai ekstrim dalam data transaksi sering kali menjadi perhatian utama dalam analisis data retail. *Outlier* yang tidak terdeteksi dapat menyebabkan distorsi distribusi data, yang pada gilirannya dapat memengaruhi akurasi keputusan bisnis atau mencerminkan perilaku konsumen ekstrem yang bernilai strategis. Beberapa penelitian menunjukkan bahwa deteksi *Outlier* dengan *Interquartile Range* (IQR) tetap menjadi metode yang banyak digunakan karena kemampuannya menangani data dengan distribusi yang tidak normal [6]. Namun, pendekatan berbasis machine learning mulai diterapkan untuk mendeteksi *Outlier* yang lebih kompleks dan non-linear pada data retail besar.

2. Segmentasi Konsumen Berdasarkan Frequency dan Monetary

Segmentasi konsumen dengan pendekatan RFM (Recency, Frequency, Monetary) merupakan teknik yang populer dalam memahami pola perilaku pembelian konsumen [7]. Beberapa penelitian terbaru menunjukkan bahwa segmentasi yang lebih spesifik berdasarkan frekuensi transaksi (Frequency) dan nilai transaksi (Monetary) mampu mengidentifikasi pelanggan dengan potensi transaksi ekstrem yang tinggi, seperti *Big Spender* dan *High-Value Outliers*. Segmentasi ini juga memungkinkan perusahaan untuk mengarahkan upaya pemasaran yang lebih terfokus pada kelompok konsumen dengan nilai transaksi tertinggi.

3. Segmentasi Harga Produk

Dalam analisis retail, harga produk sering digunakan untuk mengelompokkan produk berdasarkan

harga, dengan kategori seperti Murah, Sedang, dan Mahal. Penelitian yang dilakukan oleh Memmen et al., 2025 menunjukkan bahwa produk dengan harga lebih tinggi sering kali menghasilkan transaksi *Outlier* yang lebih besar, baik dalam jumlah transaksi maupun nilai total pembelian [8]. Segmentasi harga ini tidak hanya membantu dalam pemetaan pasar, tetapi juga berfungsi untuk menentukan strategi penetapan harga yang tepat.

4. Wilayah Pembelian dan Analisis Geospasial

Pola pembelian sering kali dipengaruhi oleh faktor geografis, di mana konsumen di wilayah tertentu menunjukkan perilaku pembelian yang berbeda. Geografis menjadi faktor penting dalam menganalisis perbedaan perilaku konsumen, karena berbagai wilayah menunjukkan preferensi produk yang berbeda [9]. Di Asia dan Eropa, pola transaksi dengan produk mahal lebih sering ditemukan pada konsumen dengan daya beli tinggi, sedangkan wilayah lain cenderung lebih sensitif terhadap harga. Analisis berbasis wilayah ini memberikan insight tambahan untuk merumuskan strategi distribusi produk.

5. Penggunaan Uji Kruskal-Wallis dalam Analisis *Outlier*

Uji Kruskal-Wallis adalah metode statistik non-parametrik yang digunakan untuk membandingkan dua atau lebih grup sampel yang independen [10]. Dalam konteks penelitian ini, uji ini digunakan untuk menganalisis perbedaan antar segmentasi variabel konsumen, kategori harga produk, wilayah pembelian, dan dimensi waktu terhadap terjadinya transaksi *Outlier*. Metode ini dipilih karena data transaksi retail tidak selalu mengikuti distribusi normal dan mengandung nilai ekstrem yang berpotensi melanggar asumsi uji parametrik.

Dibandingkan dengan uji ANOVA, yang mensyaratkan normalitas dan homogenitas varians, uji Kruskal-Wallis lebih unggul dalam menangani data yang bersifat skewed dan mengandung *Outlier*, karena pengujian dilakukan berdasarkan peringkat (rank-based), bukan nilai rata-rata. Dengan demikian, pengaruh nilai ekstrem terhadap hasil pengujian dapat diminimalkan. Oleh karena itu, penggunaan uji Kruskal-Wallis dalam penelitian ini dinilai lebih sesuai dan memberikan hasil yang lebih robust dalam mengidentifikasi perbedaan antar kelompok pada data transaksi retail yang heterogen [11]. Uji ini lebih robust terhadap skewness dan *Outlier* dibandingkan uji parametrik, sehingga cocok untuk digunakan dalam penelitian ini.

6. Analisis Eksplorasi Data (Exploratory Data Analysis/EDA)

Analisis Eksplorasi Data (EDA) adalah pendekatan penting dalam memvisualisasikan dan menganalisis pola data sebelum melakukan analisis lebih lanjut [12]. EDA membantu peneliti memahami struktur data, distribusi data, dan hubungan antar variabel [13]. Dalam penelitian ini, teknik visualisasi seperti boxplot, histogram, dan scatter plot digunakan untuk mengeksplorasi distribusi transaksi *Outlier* dan untuk memberikan gambaran yang lebih jelas mengenai pola-pola dalam data. Visualisasi ini juga membantu mendukung hasil uji statistik dan memberikan insight awal mengenai faktor-faktor yang memengaruhi transaksi *Outlier*.

METODE PENELITIAN

Penelitian menggunakan pendekatan kuantitatif, dengan fokus pada analisis statistik data numerik. Data transaksi penjualan dari retail peralatan rumah tangga dianalisis untuk mengidentifikasi *Outlier* dan faktor-faktor yang memengaruhinya.

Tahapan Penelitian

Tahapan penelitian disusun secara sistematis untuk memastikan proses analisis data dilakukan secara terstruktur dan dapat direplikasi. Mekanisme tahapan penelitian sebagai berikut:

1. Identifikasi Masalah

Penelitian dimulai dengan mengidentifikasi permasalahan pada data transaksi retail, yaitu adanya *Outlier* yang mempengaruhi distribusi data dan keputusan bisnis. Data transaksi dengan nilai ekstrem sering dihapus, padahal nilai tersebut dapat memberikan insight penting tentang perilaku konsumen.

2. Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari dataset publik "An Online Shop Business – Sales Transaction Dataset v4a" yang tersedia pada platform Kaggle. Dataset ini merekam transaksi penjualan retail peralatan rumah tangga yang terjadi selama periode Januari hingga Desember 2019. Dataset awal terdiri dari 536.350 baris data dan 8 atribut, yang mencakup informasi utama seperti *TransactionNo*, *Date*, *ProductNo*, *ProductName*, *Price*, *Quantity*, *CustomerNo*, dan *Country*. Dataset ini dipilih karena bersifat terbuka, terdokumentasi dengan baik, serta banyak digunakan dalam penelitian analisis transaksi dan perilaku konsumen, sehingga memungkinkan replikasi penelitian secara independen.

3. Pengolahan Data

Pengolahan data dilakukan menggunakan bahasa pemrograman Python dengan memanfaatkan pustaka Pandas dan NumPy untuk manipulasi data, serta Plotly untuk visualisasi eksploratif. Pada tahap ini, atribut tanggal transaksi dikonversi ke dalam format *datetime* untuk memudahkan ekstraksi dimensi waktu. Nilai *Revenue* dihitung untuk setiap transaksi sebagai hasil perkalian antara *price* dan *quantity*. Data yang telah diolah kemudian dianalisis secara eksploratif untuk mengidentifikasi pola distribusi transaksi serta indikasi awal keberadaan nilai ekstrem.

4. Pembersihan Data

Tahapan pembersihan data dilakukan untuk memastikan kualitas dan konsistensi data sebelum analisis lebih lanjut. Proses ini mencakup identifikasi dan penghapusan data duplikat, yang berjumlah 4.671 baris, guna menghindari bias akibat pencatatan transaksi berulang [14]. Selain itu, baris data dengan nilai hilang (*missing values*) pada atribut *CustomerNo* dihapus karena tidak dapat merepresentasikan perilaku pelanggan secara valid. Setelah seluruh proses pembersihan dilakukan, jumlah data berkurang dari 536.350 transaksi menjadi 489.533 transaksi, yang selanjutnya digunakan sebagai dasar analisis.

5. Identifikasi *Outlier*

Identifikasi transaksi *Outlier* pada variabel *Revenue* dilakukan menggunakan metode *Interquartile Range* (IQR) dengan pendekatan ekstrem. Metode ini dipilih karena bersifat non-parametrik dan efektif dalam menangani data dengan distribusi tidak normal serta mengandung nilai ekstrem [15]. Kuartil pertama dan ketiga ditentukan menggunakan persentil ke-5 (Q_1) dan ke-95 (Q_3) untuk menyesuaikan dengan karakteristik data yang sangat skewed. Transaksi dikategorikan sebagai *Outlier* ekstrem apabila nilai *Revenue* berada di luar batas ekstrem yang ditentukan dengan konstanta $k = 7,5$. Rumus yang digunakan:

$$IQR = Q3 - Q1$$

Dengan:

$IQR = \text{Interquartile Range}$

$Q1 = \text{kuartil pertama (25\%)}$

$Q3 = \text{kuartil ketiga (75\%)}$

Transaksi dikategorikan sebagai *Outlier* ekstrem apabila nilai *Revenue* memenuhi kondisi batas bawah dan batas atas:

$$Revenue > Q1 - k \times IQR$$

$$Revenue < Q3 + k \times IQR$$

Dengan:

$Revenue = \text{nilai pendapatan (price} \times \text{quantity)}$

$k = \text{nilai konstanta (7.5)}$

Nilai $k = 1,5$ umumnya digunakan untuk mendeteksi mild *Outlier*, sedangkan nilai k yang lebih besar digunakan untuk mengidentifikasi *Outlier* yang lebih ekstrem. Dalam penelitian ini, digunakan nilai konstanta $k = 7,5$ untuk mendeteksi *Outlier* ekstrem. Dalam penelitian ini digunakan nilai konstanta $k = 7,5$ untuk mendeteksi *Outlier* ekstrem, dengan tujuan memfokuskan analisis pada transaksi yang benar-benar berada jauh dari distribusi mayoritas data. Pemilihan nilai ini didasarkan pada karakteristik data transaksi retail yang highly skewed, sehingga penggunaan nilai k yang lebih besar dapat menghindari klasifikasi transaksi bernilai tinggi yang masih wajar sebagai *Outlier*,

sehingga analisis benar-benar difokuskan pada transaksi yang merepresentasikan perilaku pembelian ekstrem.

Untuk mengevaluasi pengaruh pemilihan nilai konstanta k , dilakukan uji sensitivitas dengan membandingkan nilai $k=3,0$, $k=5,0$, dan $k=7,5$. Hasil uji menunjukkan bahwa pada $k=3,0$ jumlah *Outlier* relatif tinggi namun pola signifikansi antar segmentasi belum konsisten. Pada $k=5,0$ jumlah *Outlier* berkurang dan perbedaan antar segmen mulai terlihat, sedangkan pada $k=7,5$ diperoleh jumlah *Outlier* yang lebih sedikit dengan hasil uji Kruskal-Wallis yang signifikan dan lebih stabil pada segmentasi utama. Temuan ini menunjukkan bahwa nilai k yang lebih besar mampu menyaring transaksi bernilai tinggi yang masih wajar dan memfokuskan analisis pada transaksi ekstrem.

Tabel 1. Uji Sensitivitas Deteksi *Outlier* terhadap Variasi Nilai Konstanta k

Nilai k	Jumlah <i>Outlier</i>	Pola Signifikansi Utama
3.0	Tinggi	Kurang signifikan pada customer, harga, dan wilayah
5.0	Sedang	Cukup signifikan pada customer, harga, dan wilayah
7.5	Rendah	Signifikan pada customer, harga, dan wilayah

Mengidentifikasi transaksi dengan nilai *Revenue* yang benar-benar ekstrem dan menghasilkan temuan statistik yang lebih stabil. Berdasarkan kriteria tersebut, dari total 489.533 transaksi setelah pembersihan data, sebanyak 1.312 transaksi teridentifikasi sebagai transaksi *Outlier* ekstrem dan menjadi fokus utama dalam analisis selanjutnya.

6. Segmentasi Data

Data yang telah diidentifikasi sebagai *Outlier* kemudian dikelompokkan dalam beberapa segmen: Segmentasi customer behavior dilakukan berdasarkan dua dimensi utama, yaitu *Frequency* dan *Monetary*. Dimensi *Frequency* merepresentasikan jumlah transaksi yang dilakukan oleh setiap pelanggan selama periode pengamatan, sedangkan dimensi *Monetary* merepresentasikan total nilai *Revenue* yang dihasilkan oleh pelanggan tersebut.

Aturan segmentasi customer behavior ditetapkan sebagai berikut:

- High-Value Outlier*: pelanggan yang terlibat dalam transaksi *Outlier* ekstrem dan memiliki nilai *Monetary* pada kuantil tertinggi distribusi.
- Big Spender*: pelanggan dengan nilai *Monetary* tinggi berdasarkan distribusi kuantil, namun tidak termasuk dalam transaksi *Outlier* ekstrem.
- Low Priority*: pelanggan dengan nilai *Frequency* dan *Monetary* relatif rendah dibandingkan kelompok lainnya.

Selain segmentasi customer behavior, transaksi *Outlier* juga dianalisis berdasarkan beberapa dimensi tambahan, yaitu:

- Segmentasi Harga: Produk dikelompokkan ke dalam kategori murah, sedang, dan mahal berdasarkan kuartil distribusi harga.
- Segmentasi Wilayah: Berdasarkan atribut *Country*, transaksi dikelompokkan ke dalam wilayah geografis utama, yaitu Eropa, Asia, Australia, dan wilayah lainnya.
- Segmentasi Waktu: Segmentasi waktu dilakukan berdasarkan hari dalam satu minggu serta periode bulan untuk mengidentifikasi pola musiman transaksi ekstrem.

7. Pengujian Metode

Uji Kruskal-Wallis digunakan untuk membandingkan perbedaan antar segmen. Uji ini dipilih karena data tidak berdistribusi normal dan mengandung *Outlier*. Nilai H dihitung dan dibandingkan dengan p -value untuk menentukan signifikansi. Untuk menguji perbedaan distribusi *Revenue* antar segmen, digunakan uji Kruskal-Wallis dengan rumus sebagai berikut:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

dengan:

H = statistik Kruskal-Wallis

- k = jumlah kelompok
 N = total observasi
 n_i = jumlah observasi pada kelompok ke- i
 R_i = jumlah ranking pada kelompok ke- i

Nilai statistik H mengikuti distribusi Chi-Square dengan derajat bebas $df = k - 1$. Keputusan pengujian didasarkan pada nilai p-value dengan tingkat signifikansi 5%. Sebagai pelengkap uji signifikansi, ukuran efek dihitung menggunakan epsilon squared (ϵ^2) untuk uji Kruskal-Wallis guna menilai kekuatan perbedaan antar kelompok. Sebagaimana rumus yang digunakan:

$$\epsilon^2 = \frac{H - k + 1}{N - k}$$

8. Interpretasi Statistik Kruskal-Wallis

Nilai statistik H yang dihasilkan dibandingkan dengan nilai kritis distribusi Chi-Square pada derajat bebas tertentu. Apabila nilai p-value lebih kecil dari tingkat signifikansi ($p < 0,05$), maka hipotesis nol ditolak, yang mengindikasikan adanya perbedaan signifikan antar kelompok [16]. Sebaliknya, apabila p-value lebih besar dari 0,05, maka tidak terdapat bukti statistik yang cukup untuk menyatakan adanya perbedaan distribusi *Revenue* antar segmen.

Ukuran efek ϵ^2 digunakan untuk menginterpretasikan besarnya perbedaan antar kelompok secara praktis. Nilai ϵ^2 yang semakin besar menunjukkan perbedaan distribusi yang semakin kuat antar kelompok, sedangkan nilai yang mendekati nol mengindikasikan perbedaan yang lemah secara praktis.

9. Analisis Visualisasi Eksploratif

Visualisasi eksploratif menggunakan boxplot, histogram, dan scatter plot untuk menganalisis distribusi dan hubungan antar variabel, serta untuk mendeteksi adanya *Outlier* yang mempengaruhi pola transaksi.

10. Penarikan Kesimpulan dan Saran

Berdasarkan hasil pengujian dan analisis, kesimpulan ditarik tentang faktor-faktor yang memengaruhi transaksi *Outlier*. Saran diberikan untuk meningkatkan strategi pemasaran dan pengelolaan transaksi retail.

HASIL PENELITIAN

Analisis dilakukan melalui segmentasi multidimensi yang mencakup customer behavior, harga produk, wilayah, dan waktu untuk mengidentifikasi pola transaksi *Outlier*, dimana data yang diambil di luar dari batas *Revenue* -247.5 pound (negatif pada barang yang di batalkan) dan 2820.5 pound. Hasil dari analisis uji Kruskal-Wallis yang digunakan pada tiap segmen menunjukkan temuan berikut:

- Segmentasi Customer: Tiga segmen utama, yaitu *High-Value Outlier*, *Big Spender*, dan *Low Priority* menunjukkan perbedaan signifikan dalam distribusi *Revenue*. Uji Kruskal-Wallis menunjukkan perbedaan signifikan dalam distribusi *Revenue Outlier* antar segmen customer ($H(2) = 153,922$, $p < 0,001$, $\epsilon^2 = 0,727$), dengan jumlah observasi masing-masing segmen yaitu *High-Value Outlier* ($N = 2$), *Big Spender* ($N = 125$), dan *Low Priority* ($N = 85$).
- Segmentasi Harga: Tiga segmen harga (Murah, Sedang, dan Mahal) diperbandingkan. Uji Kruskal-Wallis menunjukkan nilai $H = 9,289$ dengan p-value 0,0096, yang mengindikasikan perbedaan signifikan dalam distribusi *Revenue* antar segmen harga.
- Segmentasi Wilayah: Pengelompokan berdasarkan wilayah geografis (Eropa, Asia, Australia, dan Lainnya) menunjukkan bahwa wilayah Australia memiliki kontribusi dominan terhadap transaksi *Outlier*, dengan hasil Uji Kruskal-Wallis menunjukkan perbedaan signifikan dalam distribusi *Revenue Outlier* di wilayah Australia berbeda secara signifikan dibandingkan wilayah lain ($H(3) = 15,2163$, $p = 0,0016$, $\epsilon^2 = 0,0093$), dengan jumlah observasi masing-masing wilayah yaitu Eropa ($N = 123$), Asia ($N = 21$), Australia ($N = 56$), dan Lainnya ($N = 1112$).
- Segmentasi Waktu: Pada segmentasi waktu segmentasi hari, uji Kruskal-Wallis tidak menunjukkan perbedaan signifikan dalam distribusi *Revenue Outlier* ($H(2) = 1,40206$, $p = 0,4961$, $\epsilon^2 \approx 0$), dengan

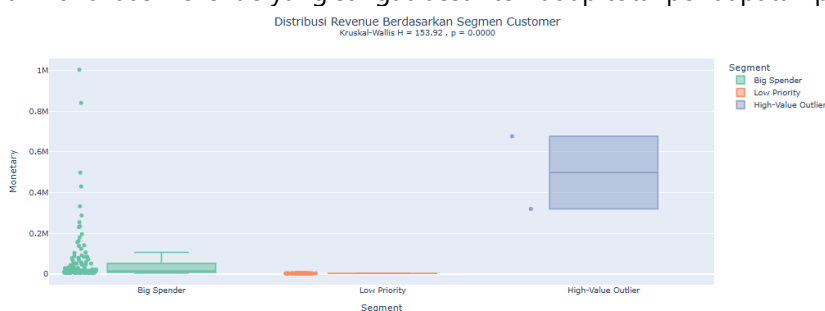
jumlah observasi *Mon–Wed* (N = 231), *Thurs–Fri* (N = 472), dan *Sat–Sun* (N = 609).” sedangkan pada segmentasi bulan, uji Kruskal-Wallis menunjukkan perbedaan signifikan dalam distribusi *Revenue Outlier* antar periode ($H(3) = 10,3029$, $p = 0,0162$, $\epsilon^2 = 0,0056$), dengan jumlah observasi Jan–Feb–Dec (N = 162), Mar–Apr–May (N = 256), Jun–Jul–Aug (N = 396), dan Sep–Oct–Nov (N = 498).

Ukuran efek (ϵ^2) dilaporkan untuk melengkapi hasil signifikansi statistik, mengingat ukuran sampel yang besar dapat menghasilkan nilai p yang sangat kecil meskipun kekuatan perbedaan antar kelompok relatif kecil. Perlu dicatat bahwa beberapa uji Kruskal-Wallis dilakukan pada berbagai segmentasi, sehingga potensi kesalahan akibat multiple testing tidak sepenuhnya dapat dihindari dan menjadi keterbatasan penelitian ini.

PEMBAHASAN

Pembahasan difokuskan pada interpretasi hasil analisis statistik dan visualisasi eksploratif untuk memahami pola transaksi *Outlier* dalam konteks bisnis retail.

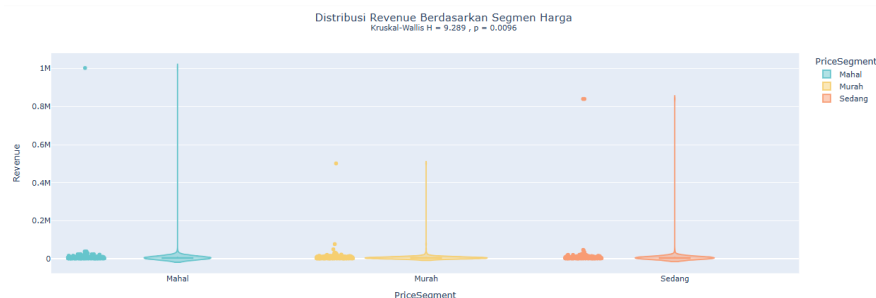
- Hasil visualisasi distribusi *Revenue* berdasarkan segmentasi customer behavior ditunjukkan pada Gambar 1. Segmentasi ini membagi pelanggan ke dalam tiga kelompok utama, yaitu *High-Value Outlier*, *Big Spender*, dan *Low Priority*. Terlihat bahwa segmen *High-Value Outlier* memiliki distribusi *Revenue* yang paling tinggi dengan sebaran nilai ekstrem yang signifikan, meskipun jumlah observasinya relatif sedikit. Hal ini menunjukkan bahwa sebagian kecil pelanggan mampu memberikan kontribusi *Revenue* yang sangat besar terhadap total pendapatan perusahaan.



Gambar 1 Analisis Visualisasi berdasarkan Segmen Customer

Segmen *Big Spender* berkontribusi besar dengan sebaran yang lebih luas, sementara segmen *Low Priority* memiliki kontribusi *Revenue* yang relatif kecil. Segmentasi ini mendukung konsep Pareto Principle temuan ini diperkuat oleh proporsi kontribusi *Revenue* antar segmen pelanggan. Berdasarkan hasil analisis, segmen *High-Value Outlier* yang hanya mencakup sekitar 0,94% dari total pelanggan menyumbang sekitar 10,91% dari total *Revenue Outlier*. Sementara itu, segmen *Big Spender* yang mencakup sekitar 58,96% pelanggan memberikan kontribusi dominan sebesar 85,94% dari total *Revenue*, sedangkan segmen *Low Priority* mencakup sekitar 40,10% pelanggan namun hanya berkontribusi sekitar 3,16% terhadap total *Revenue*. Pola ini secara kuantitatif mengonfirmasi penerapan Pareto Principle dalam konteks transaksi *Outlier* retail, di mana sebagian kecil pelanggan memberikan kontribusi yang tidak proporsional terhadap total pendapatan.

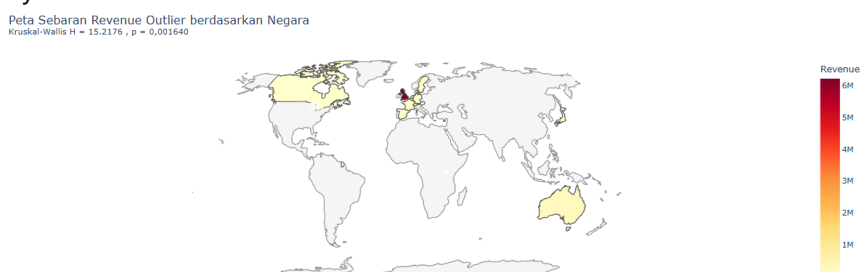
- Pada segmentasi harga, segmen Mahal menunjukkan distribusi *Revenue* yang lebih lebar, dengan *Outlier* bernilai lebih tinggi. Temuan ini mendukung pemikiran bahwa produk dengan harga lebih tinggi sering kali menghasilkan transaksi ekstrem yang lebih besar. Sebaliknya, segmen Murah cenderung memiliki *Revenue* yang lebih terpusat pada kisaran yang lebih rendah, yang memberi indikasi bahwa produk murah lebih cenderung dibeli dalam jumlah besar meski dengan nilai transaksi yang lebih rendah.



Gambar 2. Analisis Visualisasi berdasarkan Segmen Harga

Sebaliknya, kategori harga murah cenderung menghasilkan transaksi dengan nilai *Revenue* yang lebih rendah namun memiliki frekuensi pembelian yang lebih tinggi. Pola ini menunjukkan adanya trade-off antara volume transaksi dan nilai transaksi, yang perlu dipertimbangkan dalam strategi penetapan harga dan promosi produk.

- c. Analisis distribusi transaksi *Outlier* berdasarkan wilayah geografis ditampilkan pada Gambar 3. Perbedaan ini juga didukung oleh hasil ujif Kruskal-Wallis yang menunjukkan perbedaan signifikan antar wilayah.

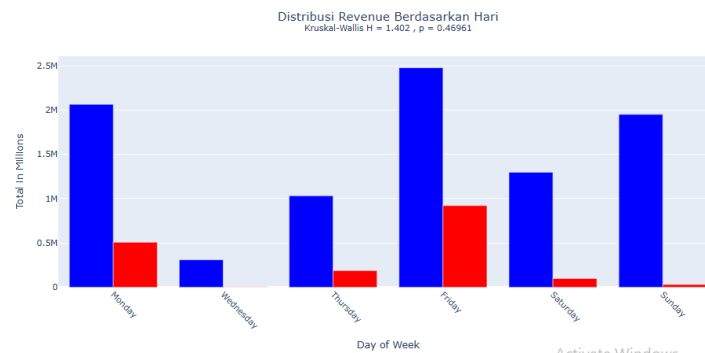


Gambar 3. Analisis Visualisasi berdasarkan Segmen Wilayah

Hasil ini menunjukkan bahwa pola transaksi ekstrem di Australia berbeda dengan kawasan Eropa dan Asia, yang dominannya transaksi *Outlier* di Australia dapat dijelaskan melalui beberapa kemungkinan faktor. Perbedaan ini kemungkinan dipengaruhi oleh faktor daya beli konsumen, pola pembelian dalam jumlah besar, serta perbedaan harga dan biaya distribusi antar wilayah.

Meskipun dataset yang digunakan tidak secara eksplisit memuat informasi kampanye pemasaran regional, promosi khusus, atau kebijakan distribusi yang lebih agresif di wilayah Australia pada periode penelitian (tahun 2019) dapat berkontribusi terhadap peningkatan transaksi ekstrem. Meskipun dataset yang digunakan tidak secara eksplisit memuat informasi kampanye pemasaran, pola transaksi yang muncul mengindikasikan bahwa faktor kontekstual wilayah berperan penting dalam pembentukan transaksi *Outlier*. Insight ini dapat dimanfaatkan oleh pengelola retail untuk merancang strategi pemasaran dan distribusi berbasis wilayah yang lebih terarah dan adaptif.

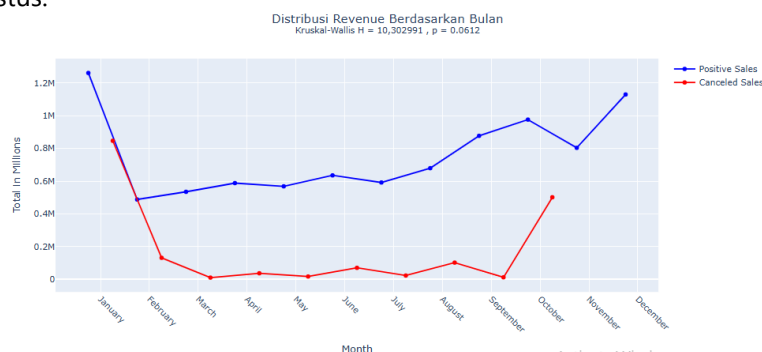
- d. Distribusi *Revenue Outlier* berdasarkan hari dalam satu minggu ditunjukkan pada Gambar 4. Meskipun terdapat variasi nilai *Revenue* antar hari, hasil uji Kruskal-Wallis menunjukkan bahwa perbedaan tersebut tidak signifikan secara statistik.



Gambar 4. Analisis Visualisasi berdasarkan Segmen Hari

Hal ini menunjukkan bahwa transaksi ekstrem tidak dipengaruhi secara kuat oleh faktor hari operasional. Temuan ini mengindikasikan bahwa transaksi ekstrem tidak terikat secara kuat pada hari tertentu, melainkan lebih dipengaruhi oleh konteks musiman dan strategi bisnis jangka menengah seperti promosi bulanan atau kuartalan. Oleh karena itu, dimensi waktu berbasis bulan dinilai lebih relevan dalam menjelaskan pola transaksi *Outlier* dibandingkan dengan segmentasi berbasis hari.

- e. Distribusi *Revenue* berdasarkan segmentasi bulan ditampilkan pada Gambar 5. Hasil uji Kruskal-Wallis menunjukkan adanya perbedaan signifikan antar periode bulan, khususnya pada periode Juni hingga Agustus.



Gambar 5. Analisis Visualisasi berdasarkan Segmen Bulan

Pada segmentasi waktu, hasil uji Kruskal-Wallis menunjukkan bahwa distribusi *Revenue Outlier* pada periode Juni–Agustus berbeda secara signifikan dibandingkan dengan periode awal dan akhir tahun. Perbedaan signifikan pada periode Juni–Agustus dapat dikaitkan dengan faktor musiman dalam konteks bisnis retail. Pada periode tersebut, terutama di belahan bumi utara, terjadi musim liburan musim panas yang sering diikuti dengan peningkatan aktivitas belanja rumah tangga, baik untuk kebutuhan pribadi maupun persiapan liburan. Selain itu, banyak perusahaan retail memanfaatkan pertengahan tahun sebagai momentum pelaksanaan promosi dan diskon musiman, yang berpotensi mendorong terjadinya transaksi bernilai tinggi dalam waktu singkat.

KESIMPULAN

Terdapat perbedaan signifikan dalam distribusi *Revenue Outlier* pada retail peralatan rumah tangga berdasarkan segmentasi customer behavior, kategori harga produk, wilayah geografis, dan dimensi waktu tertentu. Segmentasi yang didasarkan pada Frequency dan Monetary menunjukkan bahwa pelanggan dalam kategori *High-Value Outlier* memiliki kontribusi yang sangat besar terhadap total *Revenue*, meskipun jumlah pelanggan dalam segmen ini relatif sedikit. Temuan ini menegaskan pentingnya pendekatan pemasaran berbasis data yang berfokus pada identifikasi dan retensi pelanggan bernilai tinggi.

Dari perspektif manajerial, hasil penelitian ini mengindikasikan bahwa manajer retail perlu memberikan perhatian khusus pada segmen *High-Value Outlier* melalui strategi yang lebih terpersonalisasi, seperti program loyalitas eksklusif, penawaran khusus untuk pembelian bernilai tinggi, atau layanan pelanggan prioritas. Pendekatan ini diharapkan dapat meningkatkan retensi pelanggan strategis serta

memaksimalkan kontribusi *Revenue* jangka panjang. Selain itu, segmentasi berdasarkan harga menunjukkan bahwa produk dengan kategori harga tinggi memiliki potensi lebih besar dalam menghasilkan transaksi ekstrem, sehingga pengelolaan portofolio produk premium perlu dioptimalkan melalui strategi penetapan harga dan promosi yang tepat.

Segmentasi wilayah menunjukkan bahwa Australia memiliki karakteristik transaksi yang berbeda secara signifikan dibandingkan wilayah lainnya. Implikasi manajerial dari temuan ini adalah perlunya penyesuaian strategi distribusi dan pemasaran berbasis wilayah, misalnya dengan mengintensifkan kampanye pemasaran atau pengelolaan stok produk bernilai tinggi di wilayah dengan potensi transaksi ekstrem yang besar. Pada segmentasi waktu, meskipun perbedaan signifikan ditemukan pada periode bulanan tertentu, segmentasi berdasarkan hari tidak menunjukkan pengaruh yang cukup kuat. Oleh karena itu, perencanaan promosi dan penataan stok sebaiknya lebih difokuskan pada pola musiman bulanan dibandingkan pada variasi harian.

Secara keseluruhan, penelitian ini memberikan kontribusi dalam memahami peran transaksi *Outlier* sebagai sumber informasi strategis dalam industri retail. Hasil penelitian dapat dimanfaatkan oleh manajer retail untuk merumuskan strategi bisnis yang lebih efektif dan berbasis data, khususnya dalam pengelolaan pelanggan bernilai tinggi, optimasi produk premium, serta penyesuaian strategi pemasaran berdasarkan wilayah dan waktu.

REFERENCES

- [1] T. Prasetya, J. E. Yanti, A. I. Purnamasari, and A. R. Dikananda, "Analisis Data Transaksi Terhadap Pola Pembelian Konsumen Menggunakan Metode Algoritma Apriori," *INFORMATICS Educ. Prof.*, vol. 6, no. 1, pp. 43–52, 2021.
- [2] A. E. Fadlilah, Y. H. Chrisnanto, and A. K. Ningsih, "IDENTIFIKASI ANOMALI DATA AKADEMIK MENGGUNAKAN DBSCAN OUTLIER DETECTION," *Pros. Semin. Nas. Sains dan Teknol.*, vol. 12, no. 1, pp. 336–342, 2022, doi: 10.36499/psnst.v12i1.7012.
- [3] D. Dallah, H. Sulieman, and A. Al Zaatreh, "Empirical Evaluation of the Relative Range for Detecting Outliers," *ENTROPY*, pp. 1–24, 2025, doi: doi.org/10.3390/e27070731.
- [4] F. A. Maresti, "PENERAPAN EXPLORATORY DATA ANALYSIS (EDA) DAN ANALISIS RECENCY , FREQUENCY , AND MONETARY (RFM) UNTUK SEGMENTASI PELANGGAN E-," *COMPETITIVE*, vol. 19, no. 1, pp. 14–25, 2024.
- [5] J. Jamco and A. M. Balami, "Analisis Kruskal-Wallis Untuk Mengetahui Konsentrasi Belajar Mahasiswa Berdasarkan Bidang Minat Program Studi Statistika Fmipa Unpatti," *Param. J. Mat. Stat. dan Ter.*, vol. 1, no. 1, pp. 29–34, 2022, doi: 10.30598/parameter.v1i1pp29-34.
- [6] L. Azhari, N. Wulandari, F. Adiningrat, and A. D. Alexander, "Model Prediksi Penyakit Jantung dengan Penanganan Outlier Menggunakan Interquartile Range dan Extreme Gradient Boosting," *J. Inf. Syst. Res.*, vol. 6, no. 2, pp. 1398–1406, 2025, doi: 10.47065/josh.v6i2.6390.
- [7] E. F. L. Awalina and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail," *J. Teknol. dan Inf.*, vol. 13, no. 2, pp. 122–137, 2023, doi: 10.34010/jati.v13i2.10090.
- [8] M. Memmen, L. Ipsen, and J. Schulz, "Sellers ' inflation , price dispersion and substitutability: Schumpeter meets Lerner," *Bamb. Econ. Res. Gr.*, no. 209, 2025.
- [9] I. Oktariantara, "ANALISIS PERILAKU KONSUMEN TERHADAP PENGAMBILAN KEPUTUSAN PEMBELIAN MERK PRODUK," *JUMPER*, vol. 3, no. 1, pp. 338–344, 2024, doi: https://doi.org/10.51771/jumper.v3i1.1088.
- [10] S. S. Asih, A. Tripena, and A. Wardayani, "Studi Perbandingan Uji Nonparametrik K-Sampel Independen," *INFORMATICS Educ. Prof.*, vol. 2, no. 8, pp. 3503–3514, 2023.
- [11] F. F. Ritonga, Baldan Zulzamzami;Muliani, "ANALISIS TARAF OPTIMAL TERHADAP PENURUNAN HASIL PRODUKSI PADI DI KABUPATEN LANGKAT," *Gamma-Pi J. Mat. dan Terap.*, vol. 5, pp. 28–38, 2023.
- [12] N. T. M. Sagala and F. Y. Aryatama, "Exploratory Data Analysis (EDA): A Study of Olympic Medallist," *Sist. J. Sist. Inf.*, vol. 11, no. September, pp. 578–587, 2022, [Online]. Available:

<https://sistemasi.ftik.unisi.ac.id/index.php/stmsi>

- [13] T. Ramadhani;Ramadhanu;Hidayat, "Exploratory Data Analysis (EDA) untuk Mengetahui Distribusi Data Kualitas Susu Sapi," vol. 23, pp. 68–76, 2024.
- [14] N. Hanifah, M. Irwan, and P. Nasution, "Manajemen Data Yang Efektif: Solusi Untuk Mencegah dan Mengatasi Duplikasi Data Dalam Perusahaan," *J. Manag.*, vol. 3, no. 1, 2025, doi: [://doi.org/](https://doi.org/).
- [15] F. Fadillah, Annisa.k., Humairo, Sakinah Zainuri2; Siti, "PERAN UJI HIPOTESIS PENELITIAN PERBANDINGAN MENGGUNAKAN STATISTIKA NON PARAMETRIK DALAM PENELITIAN," *jurnal.asrypersadaquality*, vol. 3, no. 1, 2024, [Online]. Available: <https://jurnal.asrypersadaquality.com/index.php/alittihadu/article/view/131/182>
- [16] D. Herinanto, B. H. S. Utami, D. Arif, and M. Gumanti, "Analisis Chi Square Zona Wilayah Marketing Terhadap Penjualan Produk Ekonomi Kreatif," *eCo-Buss*, vol. 6, no. 09, pp. 1626–1637, 2024.