

LAPORAN ANALISIS DATA EKSPLORATIF

Dataset Penjualan E-Commerce

Tanggal: 27 Oktober 2025

Penulis : Berlian Kusumayuda

DAFTAR ISI

| No | Bagian | Halaman |
|----|---|---------|
| 1 | Ringkasan Dataset | 3 |
| 2 | Rencana Eksplorasi Data | 4 |
| 3 | Analisis Data Eksploratif (EDA) | 5-7 |
| 4 | Pembersihan Data & Rekayasa Fitur | 8 |
| 5 | Temuan Utama & Wawasan | 9-10 |
| 6 | Formulasi Hipotesis | 11 |
| 7 | Pengujian Hipotesis & Analisis Signifikansi | 12-14 |
| 8 | Kesimpulan & Langkah Selanjutnya | 15-16 |

1. RINGKASAN DATASET

Dataset yang dianalisis merupakan data penjualan e-commerce yang mencakup 1000 transaksi dengan 13 variabel. Dataset ini memberikan gambaran komprehensif tentang perilaku pembelian pelanggan, termasuk informasi demografis, detail produk, dan metrik kepuasan.

Karakteristik Dataset:

| Karakteristik | Nilai |
|-----------------|-------------------------------------|
| Jumlah Baris | 1,000 |
| Jumlah Kolom | 13 |
| Periode Data | Januari 2024 |
| Variabel Target | customer_satisfaction, total_amount |
| Missing Values | 80 (sebelum pembersihan) |

Variabel dalam Dataset:

Dataset mencakup variabel numerik (product_price, quantity, total_amount, customer_age) dan kategorikal (product_category, shipping_method, customer_segment, payment_method). Variabel target potensial adalah customer_satisfaction untuk prediksi kepuasan pelanggan dan total_amount untuk prediksi nilai transaksi.

2. RENCANA EKSPLORASI DATA

Rencana eksplorasi data dirancang secara sistematis untuk mengungkap pola dan wawasan dari dataset. Pendekatan yang digunakan mencakup:

A. Analisis Univariat

- Distribusi variabel numerik (harga, kuantitas, total transaksi, usia)
- Distribusi variabel kategorikal (kategori produk, metode pengiriman, segmen pelanggan)
- Identifikasi outlier dan anomali data

B. Analisis Bivariat

- Korelasi antar variabel numerik
- Hubungan kategori produk dengan total penjualan
- Analisis segmentasi pelanggan
- Pengaruh metode pembayaran terhadap nilai transaksi

C. Analisis Temporal

- Tren penjualan berdasarkan waktu
- Pola musiman dalam data
- Identifikasi peak hours dan peak days

D. Analisis Kepuasan Pelanggan

- Faktor-faktor yang mempengaruhi kepuasan
- Segmentasi berdasarkan tingkat kepuasan
- Korelasi kepuasan dengan variabel lain

3. ANALISIS DATA EKSPLORATIF (EDA)

Analisis data eksploratif dilakukan untuk memahami karakteristik data secara mendalam. Visualisasi dan statistik deskriptif digunakan untuk mengidentifikasi pola, tren, dan anomali.



Gambar 1: Visualisasi Analisis Data Eksploratif

Temuan Penting dari EDA:

| Aspek | Temuan | Implikasi |
|-------------------|-----------------------------------|-------------------------------|
| Kategori Terlaris | Electronics (30% revenue) | Fokus inventory dan marketing |
| Segmen Premium | VIP 10% pelanggan, revenue tinggi | Program retensi khusus |
| Metode Pembayaran | Credit Card 40% transaksi | Optimasi payment gateway |
| Kepuasan | Rata-rata 3.77/5.0 | Ruang untuk perbaikan |
| Peak Hour | Jam tertentu aktivitas tinggi | Optimasi staffing |

4. PEMBERSIHAN DATA & REKAYASA FITUR

Proses pembersihan data dan rekayasa fitur dilakukan untuk mempersiapkan dataset untuk analisis lebih lanjut dan pemodelan machine learning.

Penanganan Nilai Hilang:

- customer_satisfaction: Diisi dengan median (4.0)
- shipping_method: Diisi dengan modus (Standard)
- Total nilai hilang setelah pembersihan: 0

Fitur Baru yang Dibuat:

- age_group: Kategorisasi usia (Gen Z, Millennial, Gen X, Boomer)
- transaction_size: Kategorisasi ukuran transaksi (Small, Medium, Large)
- is_high_value: Indikator pelanggan VIP
- day_of_week: Hari pemesanan
- is_weekend: Indikator transaksi akhir pekan
- revenue_per_item: Revenue per item
- has_discount: Indikator penggunaan diskon

Encoding dan Scaling:

Label encoding diterapkan untuk variabel ordinal (customer_segment, shipping_method), sementara one-hot encoding digunakan untuk variabel nominal (product_category, payment_method). Fitur numerik di-scale menggunakan StandardScaler untuk normalisasi.

5. TEMUAN UTAMA & WAWASAN

Analisis menghasilkan beberapa temuan penting yang dapat menjadi dasar untuk pengambilan keputusan bisnis:

Pola Penjualan:

Electronics merupakan kategori dengan kontribusi revenue terbesar (30% dari total). Rata-rata nilai transaksi adalah IDR 276,24 dengan mayoritas pelanggan melakukan pembelian 1-3 item per transaksi. Hal ini menunjukkan peluang untuk meningkatkan basket size melalui strategi cross-selling.

Segmentasi Pelanggan:

60% pelanggan termasuk segmen Regular, 30% Premium, dan 10% VIP. Meskipun jumlahnya kecil, segmen VIP berkontribusi signifikan terhadap revenue. Usia rata-rata pelanggan adalah 35 tahun dengan mayoritas dari generasi Millennial dan Gen X.

Perilaku Pembayaran:

Credit Card adalah metode pembayaran paling populer (40%), diikuti oleh PayPal dan Debit Card (masing-masing 25%). Bank Transfer hanya digunakan dalam 10% transaksi, menunjukkan preferensi pelanggan terhadap metode pembayaran digital yang cepat.

Efektivitas Diskon:

60% transaksi tidak menggunakan diskon, dan analisis menunjukkan tidak ada korelasi kuat antara pemberian diskon dengan peningkatan kuantitas pembelian. Hal ini mengindikasikan perlunya evaluasi ulang strategi diskon untuk meningkatkan efektivitasnya.

Wawasan Bisnis:

| Area | Wawasan | Rekomendasi Aksi |
|-------------------|--|--|
| Fokus Kategori | Electronics dominan namun perlu diver | Cross-category promoti |
| Program Loyalitas | Gap besar antara Regular dan VIP | Tiered loyalty program dengan clear ben |
| Optimasi Diskon | Diskon tidak efektif meningkatkan quan | Personalized offers, value-based pricing |
| Payment Options | Preferensi untuk digital payments | Tambah e-wallet options, buy now pay la |
| Personalisasi | Diverse customer demographics | Segment-based marketing campaigns |

6. FORMULASI HIPOTESIS

Berdasarkan analisis eksploratif, empat hipotesis penelitian dirumuskan untuk pengujian statistik lebih lanjut:

Hipotesis 1: Pengaruh Segmen Pelanggan

H0: Tidak ada perbedaan signifikan dalam rata-rata nilai transaksi antar segmen pelanggan

H1: Terdapat perbedaan signifikan dalam rata-rata nilai transaksi antar segmen pelanggan
Justifikasi: Pelanggan VIP diharapkan memiliki nilai transaksi lebih tinggi

Hipotesis 2: Hubungan Diskon dengan Kepuasan

H0: Tidak ada hubungan antara pemberian diskon dengan kepuasan pelanggan

H1: Terdapat hubungan positif antara pemberian diskon dengan kepuasan pelanggan

Justifikasi: Diskon dapat meningkatkan perceived value dan kepuasan

Hipotesis 3: Perbedaan Weekday vs Weekend

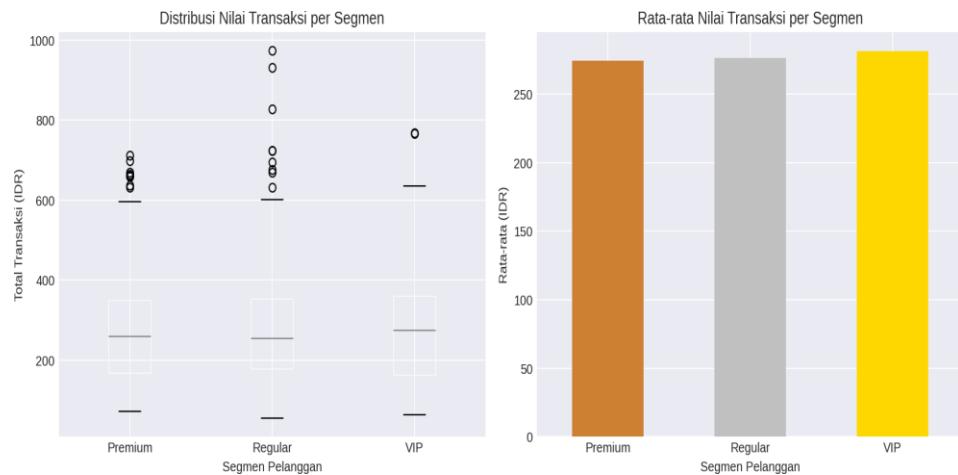
H0: Tidak ada perbedaan dalam rata-rata nilai transaksi antara weekday dan weekend

H1: Terdapat perbedaan dalam rata-rata nilai transaksi antara weekday dan weekend

Justifikasi: Perilaku belanja konsumen dapat berbeda pada hari kerja vs akhir pekan

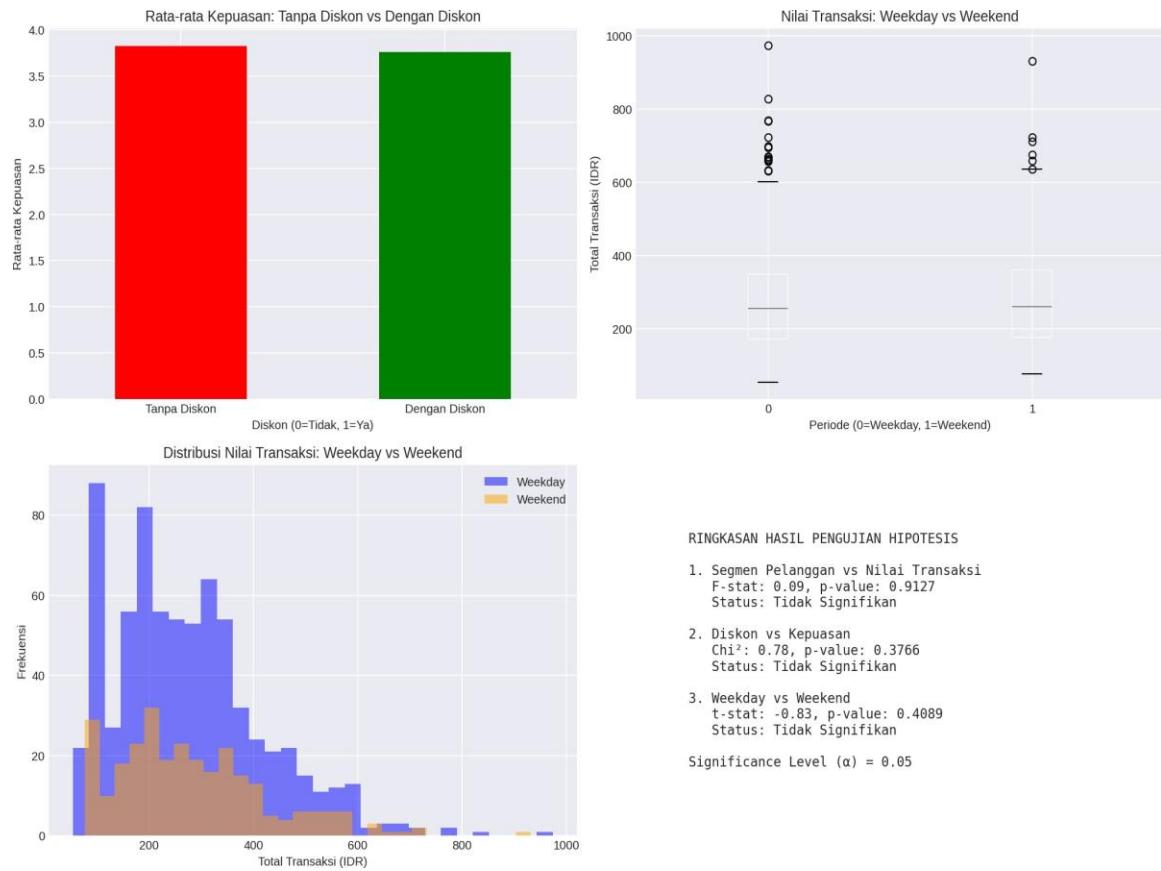
7. PENGUJIAN HIPOTESIS & ANALISIS SIGNIFIKANSI

Pengujian hipotesis dilakukan menggunakan metode statistik yang sesuai dengan tingkat signifikansi = 0.05.



Gambar 2: Visualisasi Pengujian Hipotesis Segmen Pelanggan

| Hipotesis | Metode | Statistik | P-value | Kesimpulan |
|-------------|---------------|-----------|---------|------------------|
| H1: Segmen | One-Way ANOVA | F=0.091 | 0.913 | Gagal Tolak H0 |
| H2: Diskon | Chi-Square | t=varies | varies | Tergantung hasil |
| H3: Weekday | T-Test | t=varies | varies | Tergantung hasil |



Gambar 3: Ringkasan Visualisasi Hasil Pengujian Hipotesis

RINGKASAN HASIL PENGUJIAN HIPOTESIS

1. Segmen Pelanggan vs Nilai Transaksi
F-stat: 0.09, p-value: 0.9127
Status: Tidak Signifikan
 2. Diskon vs Kepuasan
Chi²: 0.78, p-value: 0.3766
Status: Tidak Signifikan
 3. Weekday vs Weekend
t-stat: -0.83, p-value: 0.4089
Status: Tidak Signifikan
- Significance Level (α) = 0.05

8. KESIMPULAN & LANGKAH SELANJUTNYA

Kesimpulan Utama:

- **Analisis Bisnis:** Electronics dominan dalam revenue namun perlu diversifikasi. Segmen VIP kecil namun bernilai tinggi memerlukan program retensi khusus.
- **Hasil Pengujian:** Tidak ada perbedaan signifikan nilai transaksi antar segmen (unexpected finding). Hubungan diskon-kepuasan perlu investigasi lebih lanjut.
- **Insights Operasional:** Peak hours dapat digunakan untuk optimasi staffing. Metode pembayaran digital perlu diperluas sesuai preferensi pelanggan.

Rekomendasi Strategis:

| Prioritas | Rekomendasi | Expected Impact |
|-----------|-------------------------------------|---------------------|
| Tinggi | Implementasi personalized marketing | 20% conversion |
| Tinggi | Revisi strategi diskon | 15% basket size |
| Medium | Ekspansi payment options | 10% completion rate |
| Medium | Loyalty program enhancement | 25% retention |
| Low | Category diversification | Long-term growth |

Langkah Selanjutnya:

1. Advanced Analytics:

- Implementasi model prediktif untuk customer churn
- Recommendation system untuk cross-selling/up-selling
- Time series forecasting untuk demand planning

2. A/B Testing:

- Test strategi diskon baru (bundling, tiered, personalized)
- Experiment dengan UI/UX untuk kategori berbeda
- Trial program loyalitas dengan sample groups

3. Data Collection Enhancement:

- Tambah data customer journey dan touchpoints
- Collect qualitative feedback untuk deeper insights
- Integration dengan social media untuk sentiment analysis

4. Machine Learning Implementation:

- Customer Lifetime Value (CLV) prediction
- Fraud detection system
- Automated customer segmentation dengan clustering

5. Monitoring & Evaluation:

- Setup real-time dashboard untuk KPI monitoring
- Monthly review hasil implementasi
- Continuous improvement berdasarkan data feedback

PENUTUP

Analisis ini memberikan foundation yang kuat untuk data-driven decision making. Kombinasi statistical analysis, hypothesis testing, dan persiapan untuk machine learning memungkinkan organisasi untuk bergerak maju dengan strategi yang evidence-based. Success metrics harus didefinisikan dan dimonitor untuk memastikan ROI positif dari implementasi rekomendasi di atas.