

# Analisis Penjualan Supermarket X Melalui Pendekatan Berbasis Sains Data

Adinda Nur 'Azizah<sup>1</sup>, Sarah Lamtua Tobing<sup>2</sup>, Lintang Dzunuraini<sup>3</sup>, Nauraesa Denali<sup>4</sup>

Fakultas Ilmu Komputer

Universitas Pembangunan Nasional "Veteran" Jakarta

Jl. Rs. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450

2310501035@mahasiswa.upnvj.ac.id<sup>1</sup>, 2310501015@mahasiswa.upnvj.ac.id<sup>2</sup>,  
[2310501030@mahasiswa.upnvj.ac.id](mailto:2310501030@mahasiswa.upnvj.ac.id)<sup>3</sup>, 2310501033@mahasiswa.upnvj.ac.id<sup>4</sup>

## Abstrak

Studi ini bertujuan untuk menganalisis penjualan di Supermarket X menggunakan pendekatan berbasis sains data yang menggunakan visualisasi data untuk mendapatkan pemahaman yang lebih dalam tentang kebiasaan pembelian pelanggan. Pengaruh diskon terhadap penjualan per kategori, waktu puncak pembelian, dan pola pembelian selama musim liburan dipelajari melalui data penjualan. Untuk membuat grafik yang informatif, metode penelitian menggunakan pengolahan data menggunakan pustaka Python seperti Pandas, Matplotlib, dan Seaborn. Hasil analisis menunjukkan bahwa diskon yang diberikan terbesar berdampak pada kategori pakaian, dengan hari Minggu sebagai hari pembelian tertinggi. Selain itu, pola pembelian musim liburan telah meningkat, yang memiliki peluang strategis untuk promosi dan manajemen stok. Studi ini memberikan informasi yang dapat digunakan untuk membantu supermarket dalam membuat keputusan yang lebih baik dalam pengelolaan bisnis.

**Kata Kunci:** Supermarket, Ritel, Diskon, Penjualan, Sains Data, Python

## Abstract

*This study aims to analyze sales at Supermarket X using a data science-based approach that uses data visualization to gain a deeper understanding of customer purchasing habits. The effect of discounts on sales per category, peak buying times, and buying patterns during the holiday season were studied through sales data. To create informative graphs, the research method uses data processing using Python libraries such as Pandas, Matplotlib, and Seaborn. The results of the analysis show that the discounts given have the greatest impact on the clothing category, with Sunday being the highest purchase day. In addition, holiday season buying patterns have increased, which has strategic opportunities for promotions and stock management. This study provides information that can be used to assist supermarkets in making better decisions in business management.*

**Keywords:** Supermarket, Retail, Discount, Sales, Data Science, Python

## 1 Pendahuluan

### 1.1 Latar Belakang

Dalam era persaingan bisnis yang semakin ketat, supermarket besar berupaya untuk meningkatkan penjualan dengan memahami pola belanja pelanggan. Seiring dengan peningkatan ekspektasi pelanggan dan munculnya berbagai opsi belanja, seperti e-commerce, persaingan dalam industri ritel, khususnya supermarket, semakin ketat. Dalam situasi seperti ini, memahami perilaku belanja pelanggan sangat penting untuk mempertahankan daya saing dan meningkatkan pendapatan. Untuk menarik dan mempertahankan pelanggan, supermarket harus membuat strategi yang efektif untuk mengelola stok, promosi, dan pengalaman belanja. Supermarket dapat menemukan peluang peningkatan penjualan melalui optimalisasi strategi diskon, perencanaan stok, dan promosi musiman dengan menggunakan analisis data yang tepat.

Tujuan dari analisis ini adalah untuk mengetahui bagaimana berbagai variabel, seperti diskon, periode musiman, dan waktu pembelian, memengaruhi penjualan supermarket. Selain itu, analisis ini mengkaji perbedaan pola belanja antara hari kerja dan akhir pekan serta dampak musim liburan terhadap penjualan berbagai jenis barang. Analisis ini melihat perilaku belanja pelanggan melalui analisis data penjualan harian selama satu tahun. Informasi ini dapat

digunakan untuk membuat strategi promosi, pengelolaan stok, dan segmentasi pasar yang lebih efisien. Diharapkan penelitian ini dapat membantu mengembangkan rencana bisnis supermarket yang lebih peka terhadap permintaan pelanggan yang berbasis bukti.

## 1.2 Rumusan Masalah

1. Bagaimana diskon mempengaruhi jumlah penjualan untuk setiap kategori produk?
2. Bagaimana pola pembelian produk pada hari kerja dan akhir pekan?
3. Bagaimana pola pembelian selama musim liburan dan apa rekomendasi yang dapat diberikan untuk penjualan di masa mendatang?

## 2 Landasan Teori

### 2.1 Sains Data

Sains data adalah bidang yang menggunakan statistik, pemrograman, dan teknologi untuk mendapatkan pengetahuan dari sejumlah besar data. Melalui data, supermarket dapat memperoleh informasi berharga data konsumen. Ini sangat berguna dalam memahami pola perilaku yang sulit terprediksi. Sehingga supermarket dapat memahami pola perilaku belanja yang sulit diprediksi. Menurut Provost dan Fawcett (2013), sains data terdiri dari tiga elemen penting, yaitu eksplorasi data, pemodelan, dan komunikasi hasil analisis.

#### 2.1.1 Manfaat Pendekatan Sains Data untuk Supermarket

Metode seperti machine learning dan analitik prediktif memungkinkan perusahaan untuk menganalisis data secara otomatis dan membuat prediksi lebih akurat tentang tindakan konsumen. Data yang teranalisis bisa berasal dari berbagai sumber, termasuk media sosial, transaksi pembelian, dan interaksi pelanggan. Dengan begitu, perusahaan dapat mengidentifikasi preferensi konsumen, kebiasaan berbelanja, serta tren yang sedang berkembang.

Melalui data science, pola konsumen yang sebelumnya sulit terpahami dapat diungkap dengan lebih jelas. Hal ini memberikan perusahaan wawasan yang lebih dalam tentang cara untuk menyesuaikan produk atau layanan mereka agar lebih sesuai dengan kebutuhan pasar. Selain itu, data science dapat membantu perusahaan untuk meningkatkan efisiensi operasional dan mengurangi biaya. Dengan mengidentifikasi area yang membutuhkan perbaikan dan mengoptimalkan proses bisnis, perusahaan dapat mencapai kinerja yang lebih baik. Misalnya, analitik prediktif dapat digunakan untuk memprediksi permintaan produk dan mengelola persediaan dengan lebih efisien, sehingga mengurangi biaya penyimpanan dan risiko kehabisan stok. Tidak hanya itu, melalui penerapan machine learning, perusahaan juga dapat meningkatkan pengalaman pelanggan dengan menyediakan layanan yang lebih personal dan relevan. Algoritma machine learning dapat menganalisis data interaksi pelanggan untuk merekomendasikan produk atau layanan yang sesuai dengan preferensi dan kebutuhan individu, sehingga meningkatkan kepuasan dan loyalitas pelanggan.

Secara keseluruhan, dengan menggabungkan kekuatan machine learning dan analitik prediktif, perusahaan dapat memanfaatkan data untuk mengungkap wawasan berharga yang dapat meningkatkan kinerja, efisiensi, dan pengalaman pelanggan. Ini membuka peluang baru untuk inovasi dan pertumbuhan yang berkelanjutan.

#### 2.1.2 Pendekatan Sains Data Terhadap Penjualan

Untuk menghasilkan wawasan yang dapat ditindaklanjuti, data penjualan diolah melalui pendekatan berbasis sains data. Beberapa langkah diambil dalam proses ini, seperti:

1. Pengumpulan dan Pembersihan Data: Untuk memastikan kelengkapan dan akurasi, data inventaris, demografi pelanggan, dan transaksi dianalisis.
2. Eksplorasi Data: Menemukan pola yang tersembunyi dengan menggunakan teknik visualisasi.

Ada beberapa keterbatasan dengan teknik visualisasi yang digunakan dalam penelitian ini:

- Keterbatasan pada Interaktivitas: Matplotlib dan Seaborn menggunakan visualisasi statis, sehingga kurang interaktif dibandingkan dengan alat seperti Tableau atau Power BI.
  - Keterbatasan Pemrosesan Data Besar: Dalam kasus di mana Python dapat menangani dataset besar, analisis dengan Matplotlib dan Seaborn dapat menjadi lambat.
  - Keterbatasan dalam Interpretasi Data: Beberapa pola atau insight tidak dapat diamati hanya dengan visualisasi grafik. Akibatnya, analisis tambahan diperlukan menggunakan teknik statistik atau machine learning.
3. **Pemodelan Data:** Analisis perilaku pelanggan dengan menggunakan algoritma seperti regresi, klasifikasi, atau clustering.
  4. **Evaluasi dan Implementasi Hasil:** Informasi yang diperoleh dari model digunakan untuk membantu bisnis membuat keputusan.

## 2.2 Teori Perilaku Konsumen

Perilaku konsumen mencakup berbagai tindakan yang dilakukan oleh orang saat berinteraksi dengan barang, seperti mencari informasi dan melakukan evaluasi setelah pembelian. Menurut Schiffman dan Kanuk (2000), perilaku konsumen didefinisikan sebagai "perilaku yang ditampilkan pelanggan dalam mencari, membeli, menggunakan, mengevaluasi, dan membuang produk, layanan, dan ide yang mereka harapkan dapat memenuhi kebutuhan mereka".  
1. Definisi ini menunjukkan bahwa perilaku konsumen tidak hanya terbatas pada transaksi pembelian semata-mata. Ini juga menunjukkan bahwa seluruh proses yang menghasilkan keputusan pembelian juga termasuk dalam perilaku konsumen. Faktor-faktor seperti psikologis, sosial, dan budaya memengaruhi perilaku konsumen. Sangat penting untuk memahami perilaku ini untuk mengetahui kebutuhan dan preferensi pelanggan.

## 2.3 Analisis Data dan Visualisasi

Analisis eksploratif data (EDA) adalah langkah awal untuk menganalisis dataset. EDA membantu dalam mengidentifikasi pola, tren, dan anomali dalam data. Sedangkan, visualisasi data adalah representasi grafis dari informasi ataupun data yang mana merupakan komponen penting dari analisis berbasis sains data. Alat seperti Tableau, Power BI, dan Python (matplotlib, seaborn) membantu *stakeholders* memahami hasil analisis.

Dengan visualisasi data, pola-pola dalam data menjadi lebih mudah dipahami dan diakses (Madyatmadja, et al., 2022) dan dapat digunakan untuk membuat strategi penjualan yang efektif, misalnya data disajikan dengan grafik batang ataupun lainnya. Hal ini dapat memudahkan pemahaman informasi yang kompleks kepada masyarakat awam sehingga membantu mengurangi adanya misinformasi. Menurut Guntara (2023), memvisualisasikan data dapat membantu melihat hubungan dan insight yang mungkin tersembunyi dalam data mentah, sehingga proses pengambilan keputusan menjadi lebih efektif dan tepat sasaran. Visualisasi data juga dapat memudahkan pemahaman informasi yang kompleks kepada masyarakat awam sehingga menghindari adanya misinformasi.

### 2.3.1 Analisis Data Penjualan

Analisis data penjualan adalah proses mengevaluasi data transaksi untuk menemukan pola, kecenderungan, dan korelasi yang terkait dengan strategi bisnis. Tujuan dari analisis ini adalah:

- Mengidentifikasi pola pembelian konsumen, seperti pembelian berulang (*market basket analysis*).
- Menganalisis tren penjualan, seperti tren musiman atau permintaan naik/turun.
- Mengoptimalkan strategi pemasaran yang mencakup pengelolaan inventaris dan rekomendasi produk (Chaudhuri et al., 2011).

## 2.4 Teknik Statistik dan Machine Learning

Analisis penjualan berbasis data menggunakan banyak teknik statistik dan algoritma pengajaran mesin. Contohnya:

1. Regresi Linier dan Logistik: Untuk memprediksi jumlah penjualan atau kemungkinan pembelian
2. Clustering (K-Means, DBSCAN): Mengelompokkan pelanggan menurut pola pembelian mereka.
3. Algorithm of Apriori: untuk menganalisis hubungan antara produk yang sering dibeli bersama.
4. Analisa Serial Waktu: Ini digunakan untuk mempelajari tren musiman atau prediksi penjualan masa depan (Shmueli et al., 2017).

## 2.5 Studi Terkait

Penelitian telah menunjukkan bahwa teknik sains data dapat meningkatkan kinerja perusahaan. Sebuah penelitian yang dilakukan oleh Herawan et al. (2018) menemukan bahwa analisis data penjualan yang didasarkan pada *machine learning* dapat meningkatkan efisiensi inventaris hingga 25%. Studi lain yang dilakukan oleh Linoff dan Berry (2011) menemukan bahwa analisis pasar keranjang dapat meningkatkan penjualan hingga 15% melalui promosi yang ditargetkan.

## 3 Hasil Analisis

### 3.1 Dataset

Pengumpulan data dalam analisis ini menerapkan data sekunder. Data sekunder, dengan cara data yang dikumpulkan dari pihak lain atau tidak secara langsung yang diperoleh dari buku, website, diagram, grafik, dan lain-lain. Dataset yang digunakan dalam penelitian ini mencakup informasi penjualan harian dari sebuah supermarket selama satu tahun. Variabel data yang dianalisis merupakan jenis produk, kategori produk, harga, diskono, dan waktu pembelian.

### 3.2 Teknik Analisis Data

Analisis data pada pasar yang juga dikenal sebagai *market basket*, bertujuan untuk menentukan kombinasi item yang tepat untuk setiap transaksi penjualan. Tujuannya adalah untuk menggabungkan barang yang sering muncul atau dibeli oleh konsumen (Ramadhan dan Setiawan, 2021). Dalam analisis keranjang belanja (*market basket analysis*), diperlukan data minimal dua item produk pada setiap transaksi. Variabel yang digunakan dapat meliputi identifikasi transaksi dan nama item. Berdasarkan aturan asosiasi yang dihasilkan dari analisis pada umumnya, parameter probabilitas ditentukan berdasarkan nama item (Haristyarini dan Yustanti, 2021).

Analisis asosiasi adalah metode data mining yang bertujuan untuk menemukan kombinasi produk. Tahap utama dalam *association rules* adalah untuk mengidentifikasi kombinasi dengan pola frekuensi tinggi dalam basis data (Umar et al., 2022). Selain itu, algoritma apriori digunakan untuk mengungkap aturan asosiasi yang menghubungkan kombinasi item (Rifqy Alfian et al., 2019). Analisis asosiasi menjadi dasar teknik data mining untuk menemukan pola berfrekuensi tinggi. Prosesnya terdiri dari dua tahapan, yaitu analisis model frekuensi tinggi, dimana pada fase ini, dilakukan pengalihan gabungan barang yang melengkapi persyaratan minimal bobot dukungan.

$$\text{Support}(A) = \frac{\text{JML Transaksi } A}{\text{Total Transaksi}} \times 100\%$$

$$\text{Support}(A, B) = \frac{\text{JML Transaksi } A \text{ dan } B}{\text{Total Transaksi}} \times 100\%$$

Sedangkan tahapan lainnya, yaitu pembentukan aturan asosiasi, dimana setelah pola dengan frekuensi tinggi ditemukan, langkah selanjutnya adalah membentuk aturan asosiasi yang memenuhi batas minimum tingkat kepercayaan.

$$Confidence P(B|A) = \frac{JML\ Transaksi\ A\ dan\ B}{JML\ Transaksi\ A} \times 100\%$$

Agar dapat memilih aturan asosiasi yang akan digunakan, langkah pertama adalah melakukan pengurutan berdasarkan nilai dukungan (*support*) dan tingkat kepercayaan (*confidence*). Aturan-aturan tersebut kemudian dipilih sebanyak *n* atau dipilih sejumlah aturan dengan hasil terbesar (Rifqy Alfiyan et al., 2019).

### 3.3 Metode Penelitian

Data pada penelitian ini berasal dari dataset fiktif yang disusun berdasarkan metodologi yang dijelaskan dalam jurnal "Visualisasi Data Laporan Penjualan Toko Online Melalui Pendekatan Data Science Menggunakan Google Colab" (Guntara, 2023). Visualisasi dilakukan untuk mempelajari pengaruh diskon terhadap penjualan per kategori produk serta waktu puncak pembelian. Data diolah menggunakan Python dengan pustaka Pandas untuk manipulasi data dan Matplotlib serta Seaborn untuk visualisasi grafik.

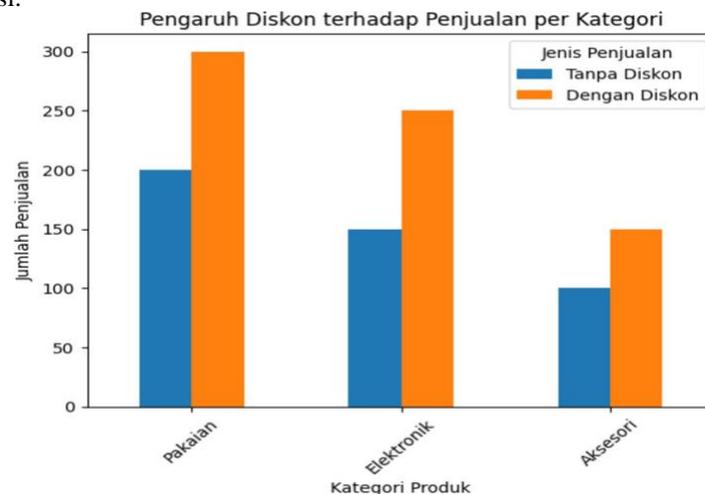
Pustaka ini dipilih karena keunggulannya dalam pengolahan dan visualisasi data dibandingkan dengan alat seperti Tableau.

- Pandas: Digunakan untuk mengubah dan menganalisis data yang disusun dalam tabel. Keunggulannya adalah fleksibilitas pengolahan data dan kemampuan untuk menangani dataset besar dengan efisien.
- Matplotlib: Meskipun agak rumit dibandingkan dengan alat visualisasi seperti Tableau, namun Matplotlib memungkinkan Anda membuat berbagai jenis grafik dengan tingkat kustomisasi yang tinggi.
- Seaborn: Membantu membuat grafik statistik dengan tampilan yang lebih menarik daripada Matplotlib. Selain itu, Seaborn dapat diintegrasikan dengan Pandas, yang memungkinkan analisis data yang lebih mudah dipahami.

Pustaka Python memungkinkan analisis yang lebih kompleks dan integrasi dengan algoritma pengajaran mesin, yang menjadikannya lebih fleksibel dan otomatis dibandingkan Tableau. Analisis ini bertujuan untuk memberikan wawasan tentang pola penjualan yang dapat membantu pengambilan keputusan strategis dalam bisnis e-commerce.

#### 3.3.1 Pengaruh Diskon Terhadap Penjualan

Untuk menganalisis pengaruh diskon terhadap jumlah penjualan barang pada setiap kategori produk, perlu dilakukan analisis regresi.

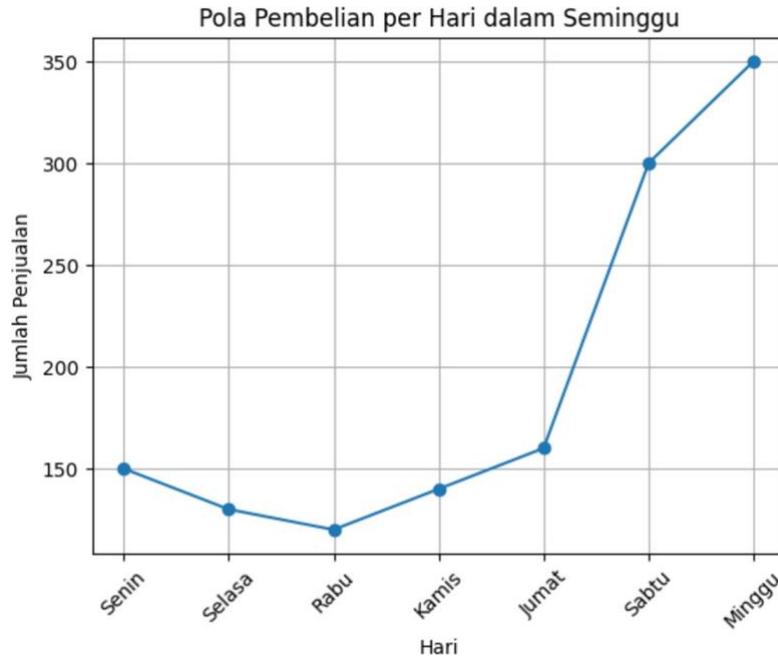


Gambar 1. Grafik Batang Penjualan Berdasarkan Kategori

Grafik di atas menunjukkan bahwa tren penjualan meningkat seiring bertambahnya diskon pada beberapa kategori produk.

### 3.3.2 Waktu Puncak Pembelian

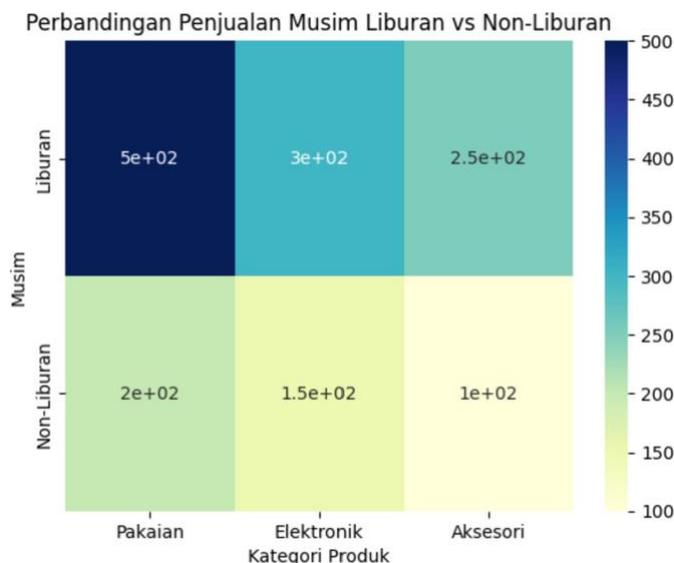
Analisis waktu pembelian menunjukkan bahwa waktu puncak terjadi peningkatan signifikan pada penjualan produk terjadi pada akhir pekan. Sedangkan, pola pembelian di hari kerja cenderung lebih stabil atau stagnan dan rendah.



Gambar 2. Stokastik Waktu Puncak Pembelian

### 3.3.3 Pola pembelian di Musim Liburan

Untuk menganalisis perbedaan pola pembelian di musim liburan, kita dapat membandingkan data penjualan selama musim liburan dengan hari kerja biasa. Berdasarkan analisis di sebuah supermarket, selama musim liburan terjadi lonjakan penjualan yang signifikan, terutama pada kategori produk pakaian dan makanan.



### 3.4 Rencana Penelitian Masa Depan

Beberapa pengembangan berikut dapat dilakukan untuk penelitian lebih lanjut:

1. Analisis Sentimen pada Ulasan Pelanggan: Mempelajari kepuasan dan preferensi pelanggan dengan menggunakan proses pengolahan bahasa natural (NLP).  
Sebagai contoh, supermarket dapat mengumpulkan ulasan pelanggan dari media sosial atau platform e-commerce. Dengan menggunakan model NLP seperti Sentiment Analysis berbasis Transformer (BERT), dapat dikategorikan apakah ulasan bersifat positif, negatif, atau netral. Misalnya, jika banyak pelanggan mengeluhkan stok produk tertentu yang sering habis, maka supermarket dapat meningkatkan persediaan produk tersebut untuk meningkatkan kepuasan pelanggan. (Liu et al., 2020).
2. Pembelajaran Mesin untuk Prediksi Penjualan: Memprediksi tren penjualan berdasarkan faktor musiman dan promosi dengan menggunakan model seperti regresi linier.  
Sebagai contoh, regresi linier dapat digunakan untuk memprediksi peningkatan penjualan selama liburan dengan menggunakan variabel seperti jumlah diskon yang diberikan, hari dalam seminggu, dan tren historis penjualan. Jika model memprediksi lonjakan penjualan tertentu, supermarket dapat menyesuaikan stok dan strategi promosi dengan lebih akurat. (Chaudhuri et al., 2011).
3. Optimisasi Harga dengan Algoritma Apriori: Menentukan strategi bundling dan harga diskon terbaik dengan menganalisis hubungan antara produk yang sering dibeli bersama.  
Sebagai contoh, jika analisis Apriori menemukan bahwa pelanggan yang membeli kopi cenderung juga membeli susu, supermarket dapat menawarkan diskon bundling untuk kedua produk ini guna meningkatkan penjualan sekaligus kepuasan pelanggan. (Agrawal & Srikant, 1994).

## 4 Kesimpulan

Beberapa poin penting yang dapat disimpulkan berdasarkan hasil analisis dengan menggunakan pendekatan sains data adalah sebagai berikut:

### 1. Pengaruh Diskon dan Musim Liburan

Hasil analisis menunjukkan bahwa diskon memiliki pengaruh positif yang signifikan terhadap penjualan setiap kategori produk, terutama pada kategori pakaian. Diskon yang diberikan mampu meningkatkan daya tarik konsumen, sehingga penjualan pada kategori ini mencapai nilai tertinggi dibandingkan kategori lainnya. Minggu adalah hari dengan penjualan tertinggi, jadi Anda dapat meningkatkan jumlah transaksi dengan mengadakan diskon.

### 2. Waktu puncak pembelian

Analisis data mengindikasikan bahwa waktu puncak pembelian terjadi pada hari Minggu. Hal ini menunjukkan bahwa konsumen lebih aktif berbelanja pada akhir pekan, kemungkinan karena mereka memiliki lebih banyak waktu luang untuk berbelanja. Peluang ini dapat dimanfaatkan supermarket untuk mengadakan promosi khusus pada hari tersebut.

### 3. Pola pembelian di musim liburan

Selama musim liburan, terjadi peningkatan signifikan dalam volume pembelian. Konsumen cenderung membeli dalam jumlah lebih besar, khususnya pada kategori kebutuhan elektronik dan makanan meningkatkan ketersediaan barang-barang yang mengalami peningkatan permintaan, seperti pakaian dan makanan. menggunakan analisis prediktif untuk menentukan jumlah stok optimal untuk mencegah kelebihan stok atau kehabisan barang. Hal ini menunjukkan perlunya pengelolaan stok dan promosi yang lebih intensif selama periode tersebut untuk memenuhi kebutuhan konsumen dan meningkatkan pendapatan.

## 5 Referensi

- Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An Overview of Business Intelligence Technology *Communications of the ACM*, 54(8), 88–98.
- Guntara, R. G. (2023). Visualisasi Data Laporan Penjualan Toko Online Melalui Pendekatan Data Science Menggunakan Google Colab. *Ulil Albab: Jurnal Ilmiah Multidisiplin*, Vol. 2, No. 6, 2091-2100.
- Harahap, M., Rozi, F., Yennimar, Y., & Siregar, S. D. (2021). Analisis Wawasan Penjualan Supermarket dengan Data Science. *Data Sciences Indonesia (DSI)*, 1(1), 1-7.
- Herawan, T., Deris, M. M., & Abawajy, J. H. (2018). *Advances on Data Mining and Applications in Business and Social Management*. Springer.
- Linoff, G. S., & Berry, M. J. A. (2011). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* (3rd ed.). Wiley.
- Madyatmadja, E. D., Ridho, M. N., Pratama, A. R., Fajri, M., & Novianto, L. (2022). Penerapan Visualisasi Data Terhadap Klasifikasi Tindak Kriminal di Indonesia. *Infotech: Journal of Technology Information*, Vol. 8, No. 1, 61-68.
- Ma'ruf, D. (2018). Sistem Informasi Manajemen: Konsep, Arsitektur, dan Implementasi. Andi.
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media.
- Rachmadhoni, A., & Tantowi, A. (2019). Analisis dan Perancangan Sistem Informasi Restoran menggunakan Metode PIECES dan Waterfall. *Jurnal Sistem Informasi*, 11(1), 35-40.
- Schiffman, L.G., & Kanuk, L.L. (2000). *Consumer Behavior*. Prentice Hall.
- Shmueli, G., Patel, N. R., & Bruce, P. C. (2017). *Data Mining for Business Analytics: Concepts, Techniques, and App*