

ANALISIS PENJUALAN SEBUAH TOKO ROTI DI PRANCIS SELAMA 12 BULAN

Ahmad Muwafiqul Adli¹, Ariela Safmi Ramdhani², Najlia Intani³, Reva Anwar⁴

Universitas Koperasi Indonesia
Kawasan Pendidikan Tinggi JL. Raya Jatinangor No.KM.20.
ahmadmuwafiquladli@gmail.com¹, arielsafmi@gmail.com²,
najliaintani@gmail.com³, hi.revaanwar@gmail.com⁴

ABSTRAK

Penelitian ini dilatarbelakangi oleh meningkatnya permintaan produk bakery yang dipengaruhi faktor tren musim. Untuk merespons hal itu, maka peneliti melakukan penelitian yang bertujuan untuk mengidentifikasi tren permintaan dengan menghadirkan analisis dan visualisasi data penjualan toko roti. Tujuan yang hendak dicapai pada penelitian ini antara lain: (1) Mengembangkan analisis penjualan toko roti menggunakan Python untuk memvisualisasikan pola dan tren data, (2) Menyediakan informasi yang mudah dipahami untuk mendukung pengambilan keputusan manajemen produksi dan strategi penjualan. Adapun hasil penelitian ini adalah: (1) Program berbasis Python yang dapat menganalisis penjualan toko roti dengan cepat, meskipun data berjumlah besar, (2) Visualisasi interaktif yang dihasilkan memudahkan memahami pola penjualan, mendeteksi produk terlaris, dan merancang strategi penjualan yang tepat.

Kata kunci: Analisis Data, Python, Penjualan, Toko Roti.

1. PENDAHULUAN

Industri bakery berkembang pesat seiring dengan meningkatnya permintaan produk roti, kue, dan pastry. Keberhasilan operasional bakery bergantung tidak hanya pada kualitas produk, tetapi juga pada efisiensi perencanaan produksi, pengelolaan stok, dan perkembangan produk. Tantangan utama adalah memahami pola permintaan yang dipengaruhi oleh faktor tren musim. Dengan kemajuan data, perkembangan produk dapat dianalisis lebih mendalam untuk mengidentifikasi tren permintaan. Penelitian ini bertujuan untuk mengidentifikasi tren penjualan, pola permintaan dan pengembangan produk yang tepat, sehingga dapat membantu bakery merencanakan produksi lebih efisien dan meningkatkan daya saing industri.

2. TINJAUAN PUSTAKA

2.1 Konsep Dasar Python

Python adalah bahasa pemrograman tingkat tinggi yang serbaguna, mudah dipelajari, dan banyak digunakan untuk berbagai keperluan. Program Python biasanya terdiri dari serangkaian perintah yang ditulis dalam file teks dengan ekstensi (.py). Program ini sering digunakan untuk membangun aplikasi web, analisis data, kecerdasan buatan, otomatisasi tugas, dan pengembangan perangkat lunak.

2.2 Library

Dalam pemrograman, library adalah kumpulan kode yang telah ditulis sebelumnya (berisi penction, class, modul) dan dapat digunakan kembali untuk menyelesaikan tugas-tugas tertentu. Tujuan utama dari library yaitu menyediakan solusi siap pakai. Macam-macam library yang dipakai dalam program ini adalah IPython, Kagglehub, Matplotlib, NumPy, OS, Pandas, Plotly, Seaborn, Sklearn, Statsmodels, dan Warnings.

2.3 Array

Array adalah struktur data yang menyimpan sekumpulan nilai dengan tipe data yang sama dalam satu variabel. Array berguna untuk operasi matematika yang cepat, seperti perhitungan statistik atau

transformasi data yang besar. Dalam program, jika menggunakan array, akan menggunakan numpy untuk operasi numerik pada data yang besar, terutama untuk kolom numerik seperti penjualan.

2.4 List

List di Python adalah struktur data dasar yang dapat menyimpan beberapa nilai. Mungkin dalam file ini, list digunakan untuk menyimpan beberapa nilai hasil analisis atau data kategori. Dalam program ini, list digunakan untuk menyimpan nama hari atau bulan, dan menyimpan hasil analisis.

2.5 Matriks

Matriks dalam konteks ini dapat digunakan untuk menyimpan data dalam bentuk dua dimensi, misalnya, data penjualan per produk dan waktu tertentu. Menggunakan numpy untuk matriks sering digunakan untuk mengolah data dalam bentuk grid.

2.6 Grafik

Untuk visualisasi grafik, seperti tren penjualan atau perbandingan penjualan berdasarkan hari atau produk, matplotlib dan seaborn adalah pustaka yang sering digunakan. Berikut bentuk-bentuk grafik yang digunakan untuk visualisasi data program ini diantaranya grafik batang dan grafik garis.

2.8 Statistik Deskriptif

Statistik deskriptif adalah proses untuk merangkum dan menggambarkan karakteristik utama dalam dataset, seperti rata-rata, median, variansi, dan lainnya. Tujuan statistik deskriptif adalah untuk memahami pola dalam data, menentukan sebaran nilai, dan memperoleh informasi awal tentang dataset sebelum melakukan analisis lebih lanjut.

a. Ukuran Pemusatan (Central Tendency)

Ukuran ini digunakan untuk menggambarkan nilai pusat atau "titik tengah" dalam dataset. Tiga ukuran pemusatan yang umum adalah:

- Mean (Rata-rata): Jumlah semua nilai dibagi dengan jumlah data. Digunakan untuk menggambarkan nilai rata-rata dalam dataset.
- Median: Nilai tengah dari dataset ketika data diurutkan. Digunakan ketika data memiliki distribusi yang miring (skewed).
- Mode: Nilai yang paling sering muncul dalam dataset. Berguna untuk data kategorikal atau nominal.

b. Ukuran Penyebaran (Dispersion)

Ukuran ini menggambarkan sejauh mana data tersebar dari nilai pusat. Beberapa ukuran penyebaran yang umum adalah:

- Range: Selisih antara nilai maksimum dan nilai minimum dalam dataset.
- Variance (Varian): Rata-rata kuadrat deviasi dari mean. Menunjukkan seberapa besar variasi data dari rata-rata.
- Standard Deviation (Simpangan Baku): Akar kuadrat dari variance. Memberikan informasi tentang seberapa tersebar data relatif terhadap rata-rata.
- Interquartile Range (IQR): Rentang antara kuartil pertama (Q1) dan kuartil ketiga (Q3), yang digunakan untuk mengukur penyebaran data tanpa dipengaruhi oleh outlier.

c. Visualisasi Data

Statistik deskriptif sering kali dipadukan dengan visualisasi data untuk membantu menggambarkan informasi secara lebih mudah dipahami. Beberapa jenis visualisasi yang digunakan adalah:

- Histogram: Menunjukkan distribusi frekuensi data dalam bentuk batang.
- Boxplot: Menampilkan sebaran data, median, dan outlier.
- Pie Chart: Memperlihatkan proporsi masing-masing kategori dalam data.
- Bar Chart: Menunjukkan perbandingan antar kategori.

d. Skewness dan Kurtosis

- Skewness (Kemiringan): Mengukur seberapa miring distribusi data. Jika skewness positif, data cenderung berada di sisi kiri dengan ekor panjang di kanan, dan sebaliknya jika skewness negatif.
- Kurtosis: Mengukur ketinggian puncak distribusi data. Kurtosis yang tinggi menunjukkan distribusi data yang sangat tajam (peaked), sementara kurtosis rendah menunjukkan distribusi yang lebih datar.

2.9 Agregasi

Agregasi adalah proses menggabungkan data berdasarkan kriteria tertentu dan menghitung nilai agregat seperti total, rata-rata, atau jumlah elemen dalam kelompok tersebut. Agregasi digunakan untuk menganalisis data dalam kategori atau kelompok tertentu, misalnya, total penjualan per hari, rata-rata penjualan per bulan, atau penjualan tertinggi berdasarkan produk.

2.10 Tipe Data

Tipe data adalah kategori atau jenis data yang digunakan dalam program untuk menentukan bagaimana suatu nilai atau variabel akan disimpan, diolah, dan digunakan. Tipe data memastikan bahwa operasi yang dilakukan pada data sesuai dengan format atau jenis datanya. Berikut tipe data yang digunakan: Numerik (Int, Float), Boolean, String, List, Tuple, Set, Dictionary dan Date/Time.

2.11 Plot

Plot adalah representasi grafis data yang lebih sering digunakan dalam analisis statistik atau sains data untuk menggambarkan hubungan antar data atau distribusi. Plot yang digunakan diantaranya, Scatter plot, Box plot, Histogram, Line plot, dan Sunburst.

2.12 Google Colab

Google Colab adalah platform berbasis cloud yang mendukung eksekusi file Python berformat .ipynb, ideal untuk analisis data besar seperti data penjualan toko roti. Dengan akses ke GPU/TPU dan pustaka Python seperti pandas atau matplotlib, Colab mempermudah pemrosesan data, visualisasi, dan kolaborasi secara real-time tanpa memerlukan instalasi software tambahan.

2.13 Sreadsheet

Spreadsheet, seperti Google Sheets atau Excel, adalah alat yang berguna untuk menyimpan, mengelola, dan menganalisis data tabular. Spreadsheet memudahkan inspeksi awal, seperti mendeteksi anomali, membersihkan data, dan melakukan perhitungan otomatis. Data dalam format .csv atau .xlsx dapat diimpor ke Python untuk analisis lebih lanjut menggunakan pustaka seperti pandas.

2.14 Korelasi Statistik

Korelasi adalah ukuran statistik yang menggambarkan hubungan antara dua variabel, menunjukkan sejauh mana satu variabel berhubungan dengan variabel lainnya. Dalam analisis data, korelasi dapat memberikan wawasan penting mengenai apakah perubahan pada satu variabel memengaruhi variabel lainnya, dan jika ya, bagaimana arah serta kekuatan hubungan tersebut. Korelasi umumnya dihitung menggunakan koefisien korelasi Pearson, yang berkisar antara -1 hingga 1. Nilai 1 menunjukkan korelasi positif sempurna, -1 menunjukkan korelasi negatif sempurna, dan 0 menunjukkan tidak ada korelasi.

Dalam Python, korelasi dapat dihitung menggunakan pustaka Pandas dengan fungsi `corr()`, yang secara otomatis menghitung korelasi Pearson antar kolom dalam DataFrame. Untuk korelasi yang lebih kompleks atau visualisasi hubungan antar variabel, kita bisa menggunakan Seaborn untuk membuat heatmap yang menggambarkan kekuatan dan arah korelasi.

3. METODE PENELITIAN

Metode penelitian yang digunakan dalam analisis data ini adalah Studi Literatur dan Analisis Data, yang dilakukan secara deskriptif menggunakan pendekatan kuantitatif. Data sekunder diperoleh dari dataset penjualan toko roti harian yang tersedia di platform Kaggle, serta literatur pendukung dari jurnal dan sumber lainnya. Studi literatur dilakukan untuk meninjau referensi terkait data, teknik analisis deskriptif, dan penerapan Python dalam pengolahan dataset besar. Analisis data dilakukan dengan statistik dan visualisasi hasilnya menggunakan bahasa pemrograman Python di Google Colab. Python dipilih untuk menangani dataset besar yang berjumlah ratusan ribu baris, yang tidak efisien jika diolah menggunakan Microsoft Excel. Data yang telah diolah kemudian disimpan dalam bentuk file CSV, yang menjadi basis untuk analisis dan pembuatan visualisasi.

Tatalaksana metode penelitian dapat diuraikan sebagai berikut:

1. Analisis dan Perencanaan
Perencanaan dimulai dengan memahami langkah-langkah yang dibutuhkan dalam normalisasi data, analisis deskriptif, dan pembuatan visualisasi untuk mencapai tujuan penelitian.
2. Desain Struktur Sistem
 - a. Menentukan metode normalisasi yang akan digunakan.
 - b. Menyiapkan struktur program Python untuk memuat data, melakukan normalisasi, dan menyimpan hasilnya.
 - c. Menentukan visualisasi yang sesuai, seperti grafik garis, histogram, dan diagram batang.
3. Pembuatan Program
Program dibuat menggunakan Python di Google Colab dengan memanfaatkan pustaka seperti pandas, numpy, dan matplotlib untuk pengolahan data dan visualisasi. Program ini dirancang untuk menghasilkan file CSV yang berisi data yang sudah dinormalisasi dan mendukung pembuatan plot visual.
 - a. Pembersihan Data
Menangani data yang tidak lengkap, duplikat, atau tidak valid.
 - b. Statistik Deskriptif
Menganalisis tren dan pola kuantitatif, seperti rata-rata penjualan, distribusi, dan korelasi antar variabel.
 - c. Visualisasi Data
Membuat berbagai grafik dan plot untuk memahami pola penjualan menggunakan pustaka Python seperti matplotlib dan seaborn.
4. Pengujian Awal
Pengujian awal dilakukan untuk memastikan langkah normalisasi dan proses penyimpanan data berjalan dengan benar. Hasil pengujian diverifikasi dengan membandingkan data asli dan data yang telah dinormalisasi.
5. Evaluasi
Evaluasi dilakukan untuk menilai efektivitas normalisasi data dan visualisasi yang dihasilkan. Proses ini bertujuan mengidentifikasi potensi error, memastikan kualitas visualisasi, dan memastikan data sesuai dengan tujuan analisis.
6. Pengujian Akhir
Setelah evaluasi dan perbaikan, dilakukan pengujian akhir untuk memastikan bahwa semua fungsi berjalan dengan baik dan visualisasi memberikan insight yang sesuai dengan kebutuhan analisis. Hasil akhir berupa data yang telah diolah dan visualisasi siap digunakan untuk pengambilan keputusan.

4. HASIL PENELITIAN DAN PEMBAHASAN

4.1 Deskripsi dan Karakteristik Data

Dataset ini berisi informasi penjualan harian toko roti, mencakup jumlah produk terjual, jenis produk, tanggal transaksi, pendapatan, dan harga per item. Data ini digunakan untuk menganalisis pola penjualan, produk terlaris, prediksi penjualan, serta korelasi antara produk dan pendapatan, sehingga membantu manajer toko membuat keputusan terkait stok, promosi, dan pengembangan produk.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 234005 entries, 0 to 234004
Data columns (total 7 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Unnamed: 0      234005 non-null  int64
1   date            234005 non-null  object
2   time            234005 non-null  object
3   ticket_number   234005 non-null  float64
4   article         234005 non-null  object
5   Quantity        234005 non-null  float64
6   unit_price      234005 non-null  object
dtypes: float64(2), int64(1), object(4)
memory usage: 12.5+ MB
```

Gambar 1 Informasi Data

Disimpan dalam format CSV, dataset ini berasal dari Kaggle berisi data dari 2 Januari 2021 sampai 30 September 2022 bersifat time series dan mencakup 234.005 baris data dari indeks 0 sampai 234004 dan 7 kolom transaksi selama lebih dari 600 hari, dengan ukuran data sebesar lebih dari 12,5 MB.

Dataset ini terdiri dari tujuh atribut utama:

1. Unnamed: 0: Kolom ini berisi indeks otomatis, kemungkinan hasil pembacaan dari file CSV. Tidak memiliki informasi khusus terkait transaksi. Jenis datanya Kuantitatif (Numerik, Diskret). Karakteristik nilai berisi 0-234004, dan 0% missing values.
2. date: Berisi tanggal transaksi dalam format string (teks), mencatat kapan transaksi terjadi. Jenis datanya Kategorikal (Nominal, Representasi Waktu). Karakteristik nilai berisi tanggal selama periode waktu tertentu, 0% missing values, formatnya YYYY-MM-DD.
3. time: Waktu transaksi terjadi pada hari tertentu. Tipe datanya kategorikal (Nominal, Representasi Waktu). Karakteristik nilai format waktu HH:MM, dan 0% missing values.
4. ticket_number: Nomor unik untuk setiap transaksi (identifikasi transaksi). Jenis datanya kuantitatif (Numerik, Kontinu). Karakteristik nilai berupa angka desimal, dan 0% missing values.
5. article: Jenis atau nama produk yang dijual dalam transaksi. Jenis datanya kategorikal (Nominal). Karakteristik nilai Berisi nama-nama produk, seperti "Roti", "Donat", "Croissant", dll, dan 0% missing values.
6. Quantity: Jumlah unit produk yang dijual dalam transaksi. Jenis datanya Kuantitatif (Numerik, Kontinu). Karakteristik nilai berupa angka desimal, 0% missing values, dan Rentang nilai perlu dicek lebih lanjut, tetapi biasanya dimulai dari 0 ke atas
7. Unit Price: Harga per unit produk dalam transaksi. Jenis datanya kategorikal (Nominal, tetapi seharusnya kuantitatif jika dalam format numerik). Karakteristik nilai formatnya teks (object) mengandung simbol mata uang, dan 0% missing values.

Semua kolom memiliki 234.005 nilai non-null, sehingga tidak ada data kosong (0% missing values) dalam dataset ini. Sehingga datanya lengkap untuk dilakukan analisis. Atribut-atribut ini memberikan informasi lengkap untuk mengidentifikasi pola penjualan, mengevaluasi performa produk, dan menganalisis pendapatan toko berdasarkan transaksi harian.

Perubahan format atribut data dilakukan untuk membuat dataset lebih terstruktur dan siap dianalisis. Nama

kolom disesuaikan, seperti 'date' menjadi 'Tanggal' dan 'article' menjadi 'Menu', agar lebih mudah dipahami. Di kolom 'Kuantitas', tanda titik ('.') diganti dengan spasi (' ') untuk menghindari kesalahan konversi data ke tipe numerik. Simbol mata uang '€' dihapus dan koma (',') diganti dengan titik ('.') pada kolom 'Harga/Item' agar harga dikenali sebagai angka desimal.

Perubahan tipe data juga dilakukan agar setiap kolom memiliki format yang tepat, seperti mengubah 'Tanggal' menjadi 'datetime64[ns]' dan 'Kuantitas' serta 'No_Transaksi' menjadi integer. Kolom 'Pendapatan' ditambahkan dengan mengalikan 'Kuantitas' dan 'Harga/Item'. Selain itu, kolom 'Bulan' dan 'Hari' ditambahkan untuk analisis pola penjualan. Semua perubahan ini bertujuan untuk menyederhanakan pengolahan data dan meningkatkan nilai informatif dataset untuk proses analisis data.

```
<class 'pandas.core.frame.DataFrame'>
Index: 232679 entries, 0 to 234004
Data columns (total 10 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Unnamed: 0           232679 non-null  int64
1   Tanggal              232679 non-null  datetime64[ns]
2   Jam                  232679 non-null  object
3   No_Transaksi         232679 non-null  int64
4   Menu                 232679 non-null  object
5   Kuantitas            232679 non-null  int64
6   Harga/Item           232679 non-null  float64
7   Bulan                232679 non-null  object
8   Hari                 232679 non-null  object
9   Pendapatan           232679 non-null  float64
dtypes: datetime64[ns](1), float64(2), int64(3), object(4)
memory usage: 19.5+ MB
```

Gambar 2 Informasi Data Setelah Perubahan Format

Dataset menjadi terdiri dari 232.679 entri dengan 10 kolom, yaitu 'Unnamed: 0', 'Tanggal', 'Jam', 'No_Transaksi', 'Menu', 'Kuantitas', 'Harga/Item', 'Bulan', 'Hari', dan 'Pendapatan'. Semua kolom memiliki data yang lengkap, tanpa nilai yang hilang (non-null). Tipe data pada kolom 'Tanggal' adalah datetime, sementara kolom 'Harga/Item' dan 'Pendapatan' menggunakan tipe float64. Kolom lainnya, seperti 'No_Transaksi', 'Kuantitas', 'Bulan', dan 'Hari', memiliki tipe data integer dan object sesuai dengan fungsinya. Dataset ini memiliki total ukuran data sebesar lebih dari 19,5 MB.

4.2 Pembersihan Data

Pembersihan data adalah tahap awal yang penting dalam proses analisis data, untuk memperbaiki dan memvalidasi data agar sesuai dengan kebutuhan analisis. Proses ini meliputi identifikasi dan perbaikan kesalahan seperti format yang tidak konsisten, nilai yang hilang, duplikasi, serta penggantian atau penyesuaian nilai yang tidak relevan. Selain itu, pembersihan data juga memastikan bahwa setiap kolom memiliki tipe data yang tepat, sehingga data dapat diproses dengan benar. Dengan langkah-langkah ini, dataset akan lebih terstruktur dan siap digunakan untuk analisis yang lebih akurat dan efisien.

	Unnamed: 0	Tanggal	Jam	No_Transaksi	Menu	Kuantitas	Harga/Item	Bulan	Hari	Pendapatan
0	0	2021-01-02	08:38	150040	BAGUETTE	1	0.90	Jan	Sat	0.90
1	1	2021-01-02	08:38	150040	PAIN AU CHOCOLAT	3	1.20	Jan	Sat	3.60
2	4	2021-01-02	09:14	150041	PAIN AU CHOCOLAT	2	1.20	Jan	Sat	2.40
3	5	2021-01-02	09:14	150041	PAIN	1	1.15	Jan	Sat	1.15
4	8	2021-01-02	09:25	150042	TRADITIONAL BAGUETTE	5	1.20	Jan	Sat	6.00

Gambar 3 Tampilan 5 Baris Pertama Data Awal Sebelum Data Cleansing

4.2.1 Segmentasi Data

Data awal berjumlah 232.679 entri dari tahun 2021-12-02 sampai 2022-09-30, dipotong sampai akhir tahun 2021 (31 Desember 2021) karena karena fokus analisis diperlukan pada periode tersebut untuk tujuan spesifik, seperti memahami tren atau pola penjualan pada akhir tahun. Selain itu, pemotongan data sampai Desember 2021 dapat mengurangi kompleksitas analisis dan memastikan konsistensi dalam pengolahan data jika periode setelahnya tidak relevan atau belum sesuai dengan kriteria yang diinginkan dalam penelitian atau analisis.

<

Gambar 4 Output Mengecek Baris Terakhir Setelah Dipotong

Kolom 'Unnamed: 0' dihapus karena tidak mengandung informasi yang relevan untuk analisis, dan kemungkinan besar merupakan indeks atau kolom tambahan yang tidak memberikan nilai tambah pada dataset. Sementara itu, kolom 'Jam' dihapus karena informasi waktu yang lebih detail sudah tercakup dalam kolom 'Tanggal', sehingga kolom ini dianggap redundan dan tidak diperlukan untuk analisis lebih lanjut. Penghapusan kedua kolom ini membantu menyederhanakan dataset dan fokus pada data yang lebih penting.

4.2.2 Duplikat Data



758

Gambar 5 Output Jumlah Data Duplikat

Setelah melakukan pengecekan dengan perhitungan, jumlah data duplikat yang ditemukan sebanyak 758 entri. Berikut adalah 5 baris pertama dari data yang terduplikat:

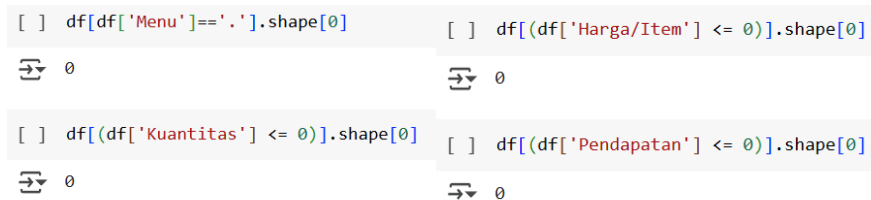


	Tanggal	No_Transaksi	Menu	Kuantitas	Harga/Item	Bulan	Hari	Pendapatan
34	2021-01-02	150061	COUPE	1	0.15	Jan	Sat	0.15
36	2021-01-02	150061	COUPE	1	0.15	Jan	Sat	0.15
40	2021-01-02	150063	COUPE	1	0.15	Jan	Sat	0.15
43	2021-01-02	150063	COUPE	1	0.15	Jan	Sat	0.15
218	2021-01-02	150154	TRADITIONAL BAGUETTE	1	1.20	Jan	Sat	1.20

Gambar 6 Tampilan 5 Baris Pertama Data Duplikat

Setelah memverifikasi keberadaan duplikat, data duplikat ini kemudian dihapus untuk memastikan bahwa dataset hanya berisi entri yang unik, sehingga analisis yang dilakukan lebih akurat dan tidak terpengaruh oleh data yang terulang.

4.2.3 Validasi Data



```
[ ] df[df['Menu']==''].shape[0]
0

[ ] df[(df['Harga/Item'] <= 0)].shape[0]
0

[ ] df[(df['Kuantitas'] <= 0)].shape[0]
0

[ ] df[(df['Pendapatan'] <= 0)].shape[0]
0
```

Gambar 7 Cek Validasi Data Kolom

Hasil validasi data menunjukkan nilai output 0 untuk keempat pengecekan, ini berarti tidak ada entri yang memiliki nilai tidak valid atau salah pada kolom.. Artinya, semua data dalam kolom tersebut sudah bersih dan valid, sehingga tidak ada perbaikan atau penghapusan yang diperlukan pada tahap ini. Dataset siap digunakan untuk analisis lebih lanjut.


4.2.4 Data Hilang



	Total	Persentase dari Missing Value
Tanggal	0	0.0
No_Transaksi	0	0.0
Menu	0	0.0
Kuantitas	0	0.0
Harga/Item	0	0.0
Bulan	0	0.0
Hari	0	0.0
Pendapatan	0	0.0

Gambar 8 Total dan Persentase Data Hilang

Output menunjukkan tidak memiliki data yang hilang dan persentase data hilang untuk setiap atributnya adalah 0%, ini berarti bahwa semua kolom dalam dataset telah terisi lengkap tanpa ada nilai yang kosong atau hilang. Setiap entri di dalam dataset memiliki informasi yang valid di setiap kolom, yang sangat penting untuk memastikan analisis yang akurat dan hasil yang dapat diandalkan. Tidak adanya data yang hilang mengindikasikan bahwa proses pengumpulan dan pemrosesan data telah dilakukan dengan baik, dan tidak ada perlu tindakan pembersihan atau pengisian nilai yang hilang. Dengan kondisi seperti ini, dataset siap untuk dianalisis lebih lanjut tanpa kekhawatiran terhadap kualitas data.



```
<class 'pandas.core.frame.DataFrame'>
Index: 130108 entries, 0 to 131419
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Tanggal          130108 non-null  datetime64[ns]
1   No_Transaksi     130108 non-null  int64
2   Menu             130108 non-null  object
3   Kuantitas        130108 non-null  int64
4   Harga/Item       130108 non-null  float64
5   Bulan            130108 non-null  object
6   Hari             130108 non-null  object
7   Pendapatan       130108 non-null  float64
dtypes: datetime64[ns](1), float64(2), int64(2), object(3)
memory usage: 8.9+ MB
```

Gambar 9 Informasi Data Terbaru Setelah Dibersihkan

Dataset ini merupakan data terbaru yang telah melalui proses pembersihan dan validasi, memastikan bahwa setiap entri dalam dataset ini lengkap dan tidak ada nilai yang hilang (non-null). Dengan 130.108 entri, data ini mencakup informasi yang relevan dan terstruktur dengan baik mengenai transaksi penjualan bakery, yang meliputi tanggal transaksi, menu yang terjual, kuantitas, harga per item, bulan, hari, dan pendapatan yang dihasilkan. Proses pembersihan data ini memastikan bahwa semua data yang digunakan dalam analisis akurat, konsisten, dan siap untuk digunakan dalam berbagai aplikasi analitik, seperti pemodelan prediksi atau analisis tren penjualan dan analisis deskriptif.



	Tanggal	No_Transaksi	Menu	Kuantitas	Harga/Item	Bulan	Hari	Pendapatan
0	2021-01-02	150040	BAGUETTE	1	0.90	Jan	Sat	0.90
1	2021-01-02	150040	PAIN AU CHOCOLAT	3	1.20	Jan	Sat	3.60
2	2021-01-02	150041	PAIN AU CHOCOLAT	2	1.20	Jan	Sat	2.40
3	2021-01-02	150041	PAIN	1	1.15	Jan	Sat	1.15
4	2021-01-02	150042	TRADITIONAL BAGUETTE	5	1.20	Jan	Sat	6.00

Gambar 10 Tampilan 5 Baris Pertama Data Setelah Dibersihkan

4.3 Dimensi Data

```
[ ] df['Menu'].nunique()
```

 125

Gambar 11 Total Menu Item

Ini menunjukkan total jumlah jenis menu yang tersedia di bakery, baik itu roti, kue, atau pastry dimana memberikan gambaran tentang variasi produk yang dijual oleh bakery, yang penting untuk memahami keanekaragaman pilihan produk yang tersedia untuk pelanggan.



```
array(['BAGUETTE', 'PAIN AU CHOCOLAT', 'PAIN', 'TRADITIONAL BAGUETTE',
      'CROISSANT', 'BANETTE', 'BANETTINE', 'SPECIAL BREAD', 'COUPE',
      'SAND JB EMMENTAL', 'KOUIGN AMANN', 'BOULE 200G', 'BOULE 400G',
      'GAL FRANGIPANE 6P', 'CAMPAGNE', 'MOISSON', 'CAFE OU EAU',
      'BRIOCHE', 'CEREAL BAGUETTE', 'SEIGLE', 'COMPLET',
      'GAL FRANGIPANE 4P', 'COOKIE', 'FICELLE', 'PAIN AUX RAISINS',
      'GAL POMME 6P', 'GAL POMME 4P', 'FINANCIER X5', 'VIK BREAD',
      'DIVERS VIENNOISERIE', 'GACHE', 'SANDWICH COMPLET', 'PAIN BANETTE',
      'GRAND FAR BRETON', 'QUIM BREAD', 'SPECIAL BREAD KG',
      'GD KOUIGN AMANN', 'BOULE POLKA', 'DEMI BAGUETTE',
      'CHAUSSON AUX POMMES', 'BAGUETTE GRAINE', 'DIVERS CONFISERIE',
      'SUCETTE', 'DIVERS BOULANGERIE', 'BOISSON 33CL', 'PATES',
      'FORMULE SANDWICH', 'DIVERS SANDWICH', 'CROISSANT AMANDES',
      'PAIN CHOCO AMANDES', 'SACHET VIENNOISERIE', 'DIVERS PATISserie',
      'NANTAIS', 'CHOCOLAT', 'PAIN S/SEL', 'FONDANT CHOCOLAT',
      'GAL POIRE CHOCO 6P', 'GAL POIRE CHOCO 4P', 'GALETTE 8 PERS',
      'SAND JB', 'SACHET DE CROUTON', 'GRANDE SUCETTE', 'DEMI PAIN',
      'TARTELETTE', 'FLAN', 'PARIS BREST', 'SAVARIN', 'FLAN ABRICOT',
      'BAGUETTE APERO', 'MILLES FEUILLES', 'CHOU CHANTILLY', 'ECLAIR',
      'ROYAL 4P', 'TARTE FRUITS 6P', 'TARTE FRUITS 4P', 'NOIX JAPONAISE',
      'THE', 'BRIOCLETTE', 'ROYAL 6P', 'ECLAIR FRAISE PISTACHE',
      'TRIANGLES', 'TROPEZIENNE', 'TROPEZIENNE FRAMBOISE', 'ROYAL',
      'TARTE FRAISE 6P', 'TARTELETTE FRAISE', 'TARTE FRAISE 4PER',
      'FRAISIER', 'NID DE POULE', 'TARTELETTE CHOC', 'PAIN DE MIE',
      'CRUMBLE', 'FINANCIER', 'DIVERS BOISSONS', 'CAKE', 'VIENNOISE',
      'TRAITEUR', 'PAIN GRAINES', 'PLATPREPARE6,50', 'PLATPREPARE5,50',
      'PLATPREPARE7,00', 'FORMULE PLAT PREPARE', 'ST HONORE', 'BROWNIES',
      'RELIGIEUSE', 'PLATPREPARE6,00', 'DELICETROPICAL',
      'CRUMBLECAMEL OU PISTAE', 'PT NANTAIS', 'GD NANTAIS',
      'DOUCEUR D HIVER', 'TROIS CHOCOLAT', 'GD FAR BRETON', 'TARTE FINE',
      'ENTREMETS', 'BRIOCHE DE NOEL', 'FRAMBOISIER', 'BUCHE 4PERS',
      'BUCHE 6PERS', 'GD PLATEAU SALE', 'BUCHE 8PERS', 'PT PLATEAU SALE',
      'REDUCTION SUCREES 12', 'PAIN NOIR', 'REDUCTION SUCREES 24'],
      dtype=object)
```

Gambar 12 List Item Menu

Dengan informasi ini, kita dapat mengetahui produk apa saja yang dijual dan seberapa sering setiap item dipilih oleh pelanggan. Ini juga membantu dalam menganalisis popularitas setiap produk dan membuat keputusan terkait pengelolaan stok atau pengembangan produk baru.

```
[ ] int(df['Kuantitas'].sum())
```

↗ 203538

Gambar 13 Total Produk Terjual

Informasi ini penting untuk mengevaluasi kinerja penjualan bakery secara keseluruhan dan dapat digunakan untuk merencanakan strategi pemasaran atau produksi, seperti memperkirakan permintaan di masa depan.

```
[ ] df['No_Transaksi'].nunique()
```

↗ 75758

Gambar 14 Total Transaksi Penjualan

Ini mencerminkan tingkat aktivitas atau jumlah pelanggan yang melakukan pembelian. Menganalisis total transaksi membantu memahami volume kunjungan pelanggan dan efektivitas strategi pemasaran yang diterapkan oleh bakery.

```
[ ] int(np.ceil(df['Pendapatan'].sum()))
```

↗ 305499

Gambar 15 Total Pendapatan (dalam satuan Euro)

Pendapatan dihitung berdasarkan harga per item dan kuantitas produk yang terjual. Informasi ini memberikan gambaran tentang seberapa sukses bisnis bakery dalam menghasilkan keuntungan, serta membantu dalam perencanaan keuangan dan evaluasi profitabilitas usaha.

Secara keseluruhan, informasi ini memberikan wawasan yang komprehensif tentang kinerja operasional bakery, mulai dari keanekaragaman menu, volume penjualan, hingga analisis pendapatan, yang dapat membantu pengelola bakery dalam mengambil keputusan yang lebih baik terkait produksi, stok, strategi bisnis, pengembangan produk.

4.4 Statistik Deskriptif

Statistik deskriptif adalah metode untuk meringkas, menyajikan, dan menggambarkan data agar lebih mudah dipahami tanpa membuat generalisasi terhadap populasi. Dalam analisis ini, ukuran pemusatan seperti mean (rata-rata), median (nilai tengah), dan modus (nilai yang paling sering muncul) digunakan untuk memberikan gambaran tentang titik pusat data. Selain itu, ukuran penyebaran seperti rentang, varians, dan simpangan baku membantu memahami sejauh mana data tersebar dari nilai rata-rata. Ukuran posisi, seperti persentil ke-10, 25 (kuartil pertama), 50 (median), 75 (kuartil ketiga), dan 90, digunakan untuk membagi data ke dalam bagian tertentu sehingga distribusinya dapat dipahami dengan lebih jelas.

Analisis statistik deskriptif juga mencakup pemeriksaan distribusi data, seperti nilai minimum dan maksimum. Tabel ringkasan dan visualisasi grafis, seperti histogram, boxplot, dan scatter plot, digunakan untuk menyajikan data secara informatif dan mempermudah identifikasi pola, tren, atau anomali. Dengan statistik deskriptif, data yang kompleks dapat diubah menjadi informasi terstruktur yang membantu dalam pengambilan keputusan atau analisis lebih lanjut.

4.4.1 Sortir Data



	Tanggal	No_Transaksi	Menu	Kuantitas	Harga/Item	Bulan	Hari	Pendapatan
50552	2021-06-12	179931	CAFE OU EAU	200	1.0	Jun	Sat	200.0
74999	2021-07-31	194199	GD NANTAIS	11	11.0	Jul	Sat	121.0
25378	2021-04-04	165104	ROYAL 6P	4	18.0	Apr	Sun	72.0
75000	2021-07-31	194199	GD KOUIGN AMANN	9	7.5	Jul	Sat	67.5
115245	2021-10-31	217821	TRADITIONAL BAGUETTE	55	1.2	Oct	Sun	66.0

Mengurutkan dari pendapatan Terendah:

	Tanggal	No_Transaksi	Menu	Kuantitas	Harga/Item	Bulan	Hari	Pendapatan
2971	2021-01-15	151885	DIVERS SANDWICHS	1	0.07	Jan	Fri	0.07
71346	2021-07-25	192047	COUPE	1	0.15	Jul	Sun	0.15
71341	2021-07-25	192044	COUPE	1	0.15	Jul	Sun	0.15
71335	2021-07-25	192041	COUPE	1	0.15	Jul	Sun	0.15
71315	2021-07-25	192035	COUPE	1	0.15	Jul	Sun	0.15

Gambar 16 Output 5 Baris Pertama Urutan Data Terbesar dan Data Terkecil

Ini memungkinkan analisis dengan cepat mengidentifikasi transaksi yang menghasilkan pendapatan paling banyak dan paling sedikit, yang dapat berguna untuk analisis lebih lanjut.

4.4.2 Statistik Deskriptif Tanggal



Rata-Rata Tanggal: 2021-07-06 20:24:20.718787840
Standar Deviasi Tanggal: 92 days 09:40:13.062989266

Persentil:

10%: 2021-02-27T00:00:00.000000000
25%: 2021-04-25T00:00:00.000000000
50%: 2021-07-16T00:00:00.000000000
75%: 2021-09-08T00:00:00.000000000
90%: 2021-11-07T00:00:00.000000000

Nilai Minimum Tanggal: 2021-01-02 00:00:00
Nilai Maksimum Tanggal: 2021-12-31 00:00:00

Gambar 17 Output Statistik Deskriptif Tanggal

Interpretasi:

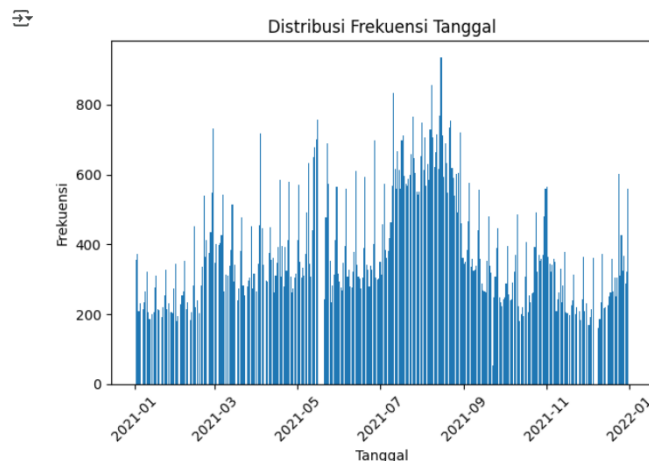
1. Rata-rata tanggal transaksi penjualan di toko roti adalah sekitar tanggal 6 Juli 2021 pukul 20:24. Ini menunjukkan bahwa sebagian besar transaksi terjadi di sekitar pertengahan tahun 2021.
2. Standar deviasi tanggal adalah sekitar 92 hari. Ini menunjukkan bahwa data tanggal tersebar relatif luas di sekitar rata-rata, dengan rentang variasi sekitar 92 hari.
3. Persentil:
 - 10%: 10% dari transaksi terjadi sebelum tanggal 27 Februari 2021.
 - 25%: 25% dari transaksi terjadi sebelum tanggal 25 April 2021.
 - 50% (Median): 50% dari transaksi terjadi sebelum tanggal 16 Juli 2021.
 - 75%: 75% dari transaksi terjadi sebelum tanggal 8 September 2021.
 - 90%: 90% dari transaksi terjadi sebelum tanggal 7 November 2021.

Persentil memberikan gambaran tentang distribusi data tanggal dan menunjukkan sebaran data pada titik-titik persentase tertentu.

4. Nilai Minimum Tanggal: Tanggal transaksi paling awal dalam dataset adalah 2 Januari 2021.
5. Nilai Maksimum Tanggal: Tanggal transaksi paling akhir dalam dataset adalah 31 Desember 2021.

Kesimpulan:

Data transaksi penjualan toko roti mencakup periode satu tahun penuh, dari 2 Januari 2021 hingga 31 Desember 2021. Rata-rata transaksi terjadi di sekitar pertengahan tahun, dan data tanggal tersebar relatif luas di sekitar rata-rata. Persentil memberikan informasi lebih rinci tentang distribusi data tanggal.



Gambar 18 Grafik Distribusi Frekuensi Tanggal

Interpretasi :

Grafik menunjukkan distribusi frekuensi data berdasarkan tanggal. Sumbu X menampilkan tanggal, mulai dari awal 2021 hingga awal 2022. Sumbu Y menunjukkan frekuensi atau jumlah data untuk setiap tanggal. Grafik memiliki puncak dan lembah, yang menunjukkan fluktuasi jumlah data pada tanggal yang berbeda.

Grafik ini menunjukkan bahwa toko roti cenderung lebih ramai pada akhir pekan (Sabtu dan Minggu), dengan frekuensi transaksi yang lebih tinggi dibandingkan dengan hari-hari lainnya dalam seminggu. Hal ini ditunjukkan oleh batang yang lebih tinggi pada hari Sabtu dan Minggu pada grafik. Sebaliknya, hari kerja (Senin-Jumat) menunjukkan frekuensi transaksi yang lebih rendah, mengindikasikan bahwa toko roti mungkin lebih sepi pada hari-hari tersebut.

Variasi frekuensi transaksi cukup signifikan, menunjukkan bahwa penjualan toko roti tidak konsisten dari hari ke hari.

Terdapat pola mingguan yang jelas, dengan peningkatan frekuensi transaksi menjelang akhir pekan dan penurunan pada awal pekan.

Grafik ini tidak menunjukkan adanya data outlier yang signifikan, artinya tidak ada tanggal dengan frekuensi transaksi yang sangat berbeda (jauh lebih tinggi atau lebih rendah) daripada tanggal-tanggal lainnya.

Kesimpulan:

Toko roti mengalami peningkatan penjualan pada akhir pekan, sementara penjualan pada hari kerja cenderung lebih rendah. Variasi frekuensi transaksi menunjukkan bahwa penjualan tidak konsisten dari hari ke hari, namun terdapat pola mingguan yang jelas dengan peningkatan penjualan menjelang akhir pekan.

4.4.3 Statistik Deskriptif No Transaksi

```
Rata-Rata No_Transaksi: 188726.30739846895
Standar Deviasi No_Transaksi: 22194.666977343422

Nilai Minimum No Transaksi: 150040
Nilai Maksimum No Transaksi: 227279

Persentil No_Transaksi:
10%: 157945.7
25%: 169638.75
50%: 188847.0
75%: 207645.0
90%: 219537.3
```

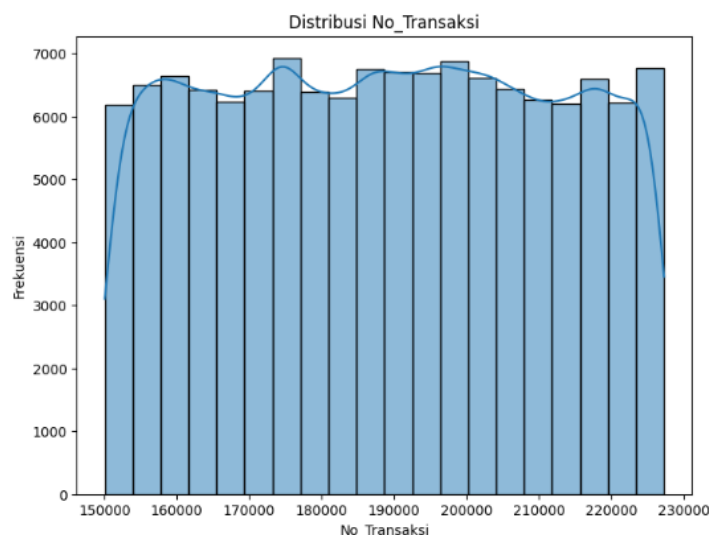
Gambar 19 Output Statistik Deskriptif No_Transaksi

Interpretasi :

Gambar ini menunjukkan statistik deskriptif dari data "No_Transaksi". Informasi yang ditampilkan meliputi rata-rata, standar deviasi, nilai minimum dan maksimum, serta persentil dari data transaksi. Rata-rata nomor transaksi memberikan gambaran umum tentang nilai tengah dari seluruh data transaksi yang ada. Standar Deviasi menunjukkan adanya variasi yang cukup signifikan di antara nomor-nomor transaksi. Nilai transaksi terendah adalah 150040, sedangkan nilai tertinggi adalah 227279. Hal ini menunjukkan rentang nomor transaksi yang cukup luas. Persentil ini membantu kita memahami distribusi data dengan lebih baik.

Kesimpulannya :

Data ini dapat digunakan untuk analisis lebih lanjut, seperti mengidentifikasi pola transaksi, menentukan anomali, atau mengevaluasi kebijakan yang mungkin berdampak pada distribusi nomor transaksi. Pemahaman yang mendalam tentang distribusi data ini penting untuk pengambilan keputusan yang lebih baik dalam konteks manajemen data dan analisis transaksi.



Gambar 20 Grafik Distribusi No_Transaksi

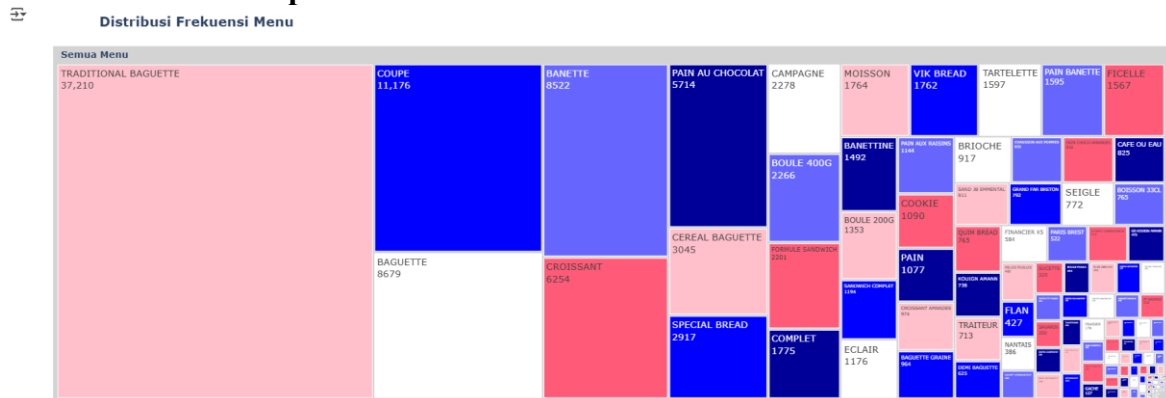
Interpretasi:

Gambar ini menunjukkan histogram distribusi dari "No_Transaksi" dengan sumbu x mewakili "No_Transaksi" dan sumbu y mewakili "Frekuensi". Histogram ini memiliki batang-batang yang menunjukkan frekuensi dari nomor transaksi dalam rentang tertentu, dengan rentang nomor transaksi dari 150000 hingga 230000. Histogram ini menunjukkan bahwa frekuensi nomor transaksi relatif merata di seluruh rentang, dengan beberapa fluktuasi kecil di beberapa titik.

Kesimpulan:

Kesimpulan dari histogram distribusi No Transaksi adalah bahwa frekuensi nomor transaksi relatif merata di seluruh rentang, dengan beberapa fluktuasi kecil di beberapa titik. Puncak frekuensi terlihat berada di sekitar nomor transaksi 160000, 170000, dan 220000, sementara frekuensi sedikit menurun di sekitar nomor transaksi 150000 dan 230000. Ini menunjukkan bahwa ada variasi dalam jumlah transaksi, namun distribusinya cenderung merata secara keseluruhan. Informasi ini dapat digunakan untuk analisis lebih lanjut dalam mengidentifikasi pola transaksi atau anomali yang mungkin ada.

4.4.4 Statistik Deskriptif Menu



Gambar 21 Distribusi Frekuensi Menu

Interpretasi :

Gambar tersebut menunjukkan distribusi frekuensi menu berdasarkan jumlah penjualan. Menu dengan penjualan tertinggi adalah Traditional Baguette dengan jumlah penjualan mencapai 37.210. Diikuti oleh Coupe dengan 11.176 penjualan dan Baguette dengan 8.679 penjualan. Menu dengan penjualan paling rendah yaitu Gache dengan 127 penjualan.

Kesimpulan:

Ada perbedaan yang signifikan dalam popularitas setiap menu. Traditional Baguette menjadi menu yang paling diminati dengan jumlah penjualan mencapai 37.210, jauh lebih tinggi dibandingkan menu lainnya. Menu Coupe dan Baguette juga menunjukkan penjualan yang relatif baik, masing-masing dengan 11.176 dan 8.679 penjualan. Sementara itu, Gache memiliki penjualan yang sangat rendah, hanya 127 unit, yang mengindikasikan bahwa menu ini kurang diminati oleh pelanggan. Temuan ini dapat membantu bakery dalam mengoptimalkan produksi dan promosi, dengan fokus pada menu yang lebih laris dan mempertimbangkan strategi untuk meningkatkan penjualan menu yang kurang populer.

4.4.5 Statistik Deskriptif Kuantitas

➡ Rata-Rata Kuantitas: 1.5643772865619332
Standar Deviasi Kuantitas: 1.2608046442163279

Nilai Minimum Kuantitas: 1
Nilai Maksimum Kuantitas: 200

Persentil Kuantitas:

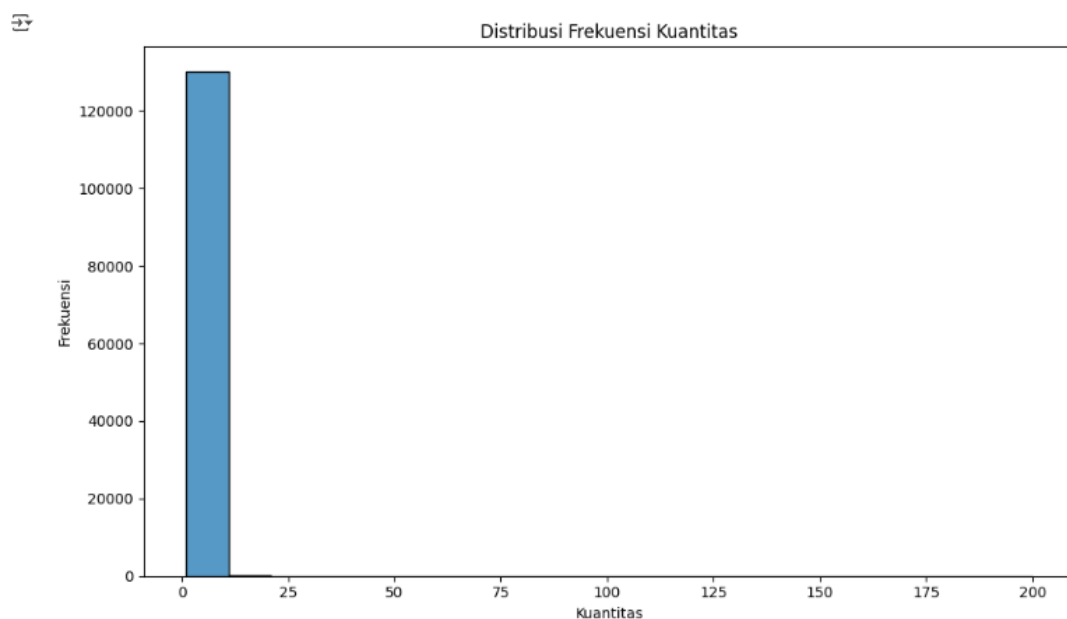
10%	1.0
25%	1.0
50%	1.0
75%	2.0
90%	3.0

Gambar 22 Output Statistik Deskriptif Kuantitas

Interpretasi : Interpretasi : Deskriptif Statistik Kuantitatif menunjukkan bahwa sebagian besar data kuantitas berkumpul di nilai yang relatif rendah, dengan 50% data memiliki kuantitas 1 atau kurang. Namun, ada juga beberapa data dengan kuantitas yang jauh lebih tinggi, hingga 200, yang mengakibatkan standar deviasi yang cukup besar. Hal ini menggambarkan bahwa meskipun sebagian besar data berada di nilai rendah, ada variasi yang signifikan dalam data kuantitas tersebut.

Kesimpulan :

Kesimpulan ini memberikan gambaran yang jelas tentang distribusi dan variasi dalam data kuantitas, serta adanya outlier yang signifikan. Pemahaman ini penting untuk analisis lebih lanjut dalam konteks manajemen data dan interpretasi statistik yang relevan.



Gambar 23 Distribusi Frekuensi Kuantitas

Interpretasi:

Gambar di atas menunjukkan sebuah histogram yang menggambarkan distribusi frekuensi pendapatan, dengan nilai maksimum sekitar 120.000. Histogram ini memperlihatkan bahwa sebagian besar pendapatan berada di kisaran 0 hingga 25, dengan jumlah frekuensi yang sangat tinggi di kisaran tersebut. Hal ini menunjukkan bahwa mayoritas populasi memiliki pendapatan yang relatif rendah.

Kesimpulannya:

Kesimpulan dari histogram distribusi frekuensi pendapatan adalah bahwa sebagian besar pendapatan berada di kisaran 0 hingga 25, dengan jumlah frekuensi yang sangat tinggi di kisaran tersebut. Hal ini menunjukkan bahwa mayoritas populasi memiliki pendapatan yang relatif rendah. Distribusi ini dapat menggambarkan ketimpangan pendapatan yang signifikan dalam populasi. Data ini penting untuk analisis lebih lanjut mengenai kebijakan ekonomi dan sosial yang dapat membantu mengurangi ketimpangan pendapatan dan meningkatkan kesejahteraan masyarakat.

4.4.6 Statistik Deskriptif Harga/Item

```
Rata-Rata Harga/Item: 1.6037385864051403
Standar Deviasi Harga/Item: 1.6335578188633602

Nilai Minimum Harga/Item: 0.07
Nilai Maksimum Harga/Item: 44.0

Persentil Harga/Item:
10%: 0.6
25%: 1.05
50%: 1.2
75%: 1.5
90%: 2.5
```

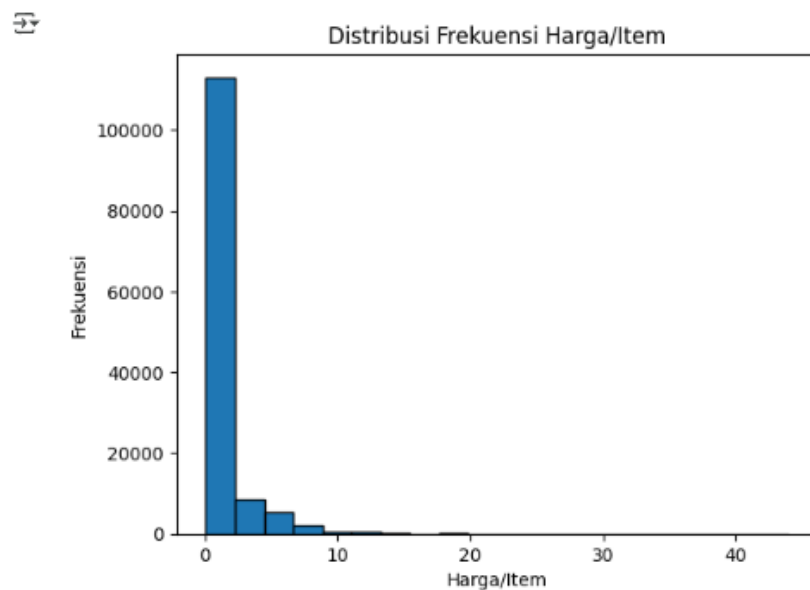
Gambar 24 Output Statistik Deskriptif Harga/Item

Interpretasi :

Kesimpulan dari statistik deskriptif harga per item dalam dataset ini menunjukkan adanya variasi harga yang signifikan, dengan rata-rata harga sebesar 1.60 dan standar deviasi 1.63. Harga item berkisar dari 0.07 hingga 44.0, menunjukkan rentang yang cukup besar. Dari data persentil, terlihat bahwa sebagian besar item memiliki harga yang relatif rendah, dengan 90% item memiliki harga di bawah 2.5. Namun, ada beberapa item dengan harga yang jauh lebih tinggi, menunjukkan ketimpangan harga yang cukup signifikan dalam dataset. Distribusi ini mencerminkan mayoritas item dijual dengan harga rendah, sementara hanya sebagian kecil yang memiliki harga tinggi. Data ini penting untuk analisis lebih lanjut mengenai penetapan harga dan strategi pemasaran.

Kesimpulan :

Statistik Deskriptif menunjukkan harga per item dalam dataset tertentu. Distribusi harga ini mencerminkan ketimpangan harga yang cukup signifikan dalam dataset tersebut.



Gambar 25 Distribusi Frekuensi Harga/Item

Interpretasi:

Gambar tersebut menampilkan histogram yang menggambarkan distribusi frekuensi harga per item. Dapat dilihat bahwa mayoritas item memiliki harga yang rendah, dengan frekuensi terbanyak berada di kisaran 0 hingga 1 unit harga. Frekuensi kemudian menurun drastis seiring dengan meningkatnya harga per item, menandakan bahwa semakin mahal harga suatu item, semakin sedikit jumlah item yang memiliki harga tersebut.

Kesimpulan:

Gambar tersebut merupakan histogram yang menunjukkan distribusi frekuensi harga per item. Pada gambar, dapat dilihat bahwa sebagian besar item memiliki harga yang rendah (sekitar 0-5), dengan frekuensi terbanyak berada di kisaran 0-1. Semakin tinggi harga item, semakin sedikit jumlah item dengan harga tersebut. Hal ini menunjukkan bahwa data tersebut memiliki distribusi yang miring ke kanan atau right skewed.

4.4.7 Statistik Deskriptif Pendapatan

```
➡ Rata-Rata Pendapatan: 2.348034479048175
   Standar Deviasi Pendapatan: 2.5071319995227688

   Nilai Minimum Pendapatan: 0.07
   Nilai Maksimum Pendapatan: 200.0

   Persentil Pendapatan:
   10%: 0.6
   25%: 1.2
   50%: 1.5
   75%: 2.5
   90%: 4.8
```

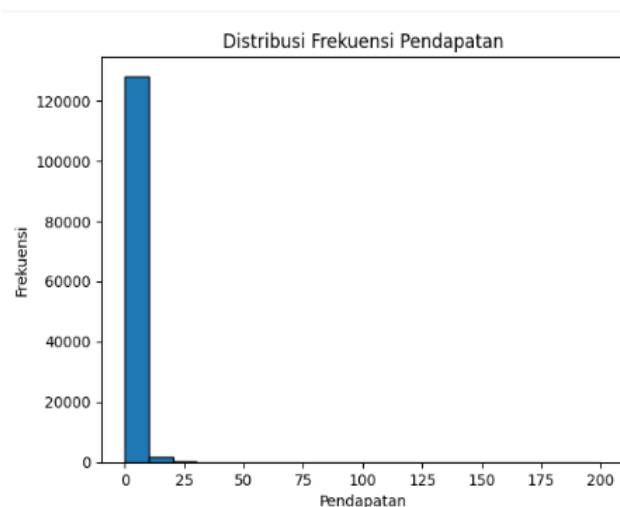
Gambar 26 Output Statistik Deskriptif Pendapatan

Interpretasi:

Gambar di atas menunjukkan statistik pendapatan yang mencakup rata-rata, distribusi pendapatan yang tidak merata dalam populasi, dengan mayoritas berada pada pendapatan yang lebih rendah namun beberapa individu memiliki pendapatan yang sangat tinggi. Ini bisa menjadi indikator penting untuk analisis ketimpangan ekonomi dan kebutuhan untuk intervensi kebijakan.

Kesimpulan:

Statistik pendapatan yang ditampilkan dalam gambar adalah adanya ketimpangan yang signifikan dalam distribusi pendapatan. Rata-rata pendapatan berada di angka 2.35, namun dengan standar deviasi yang cukup tinggi yaitu 2.51, menunjukkan adanya variasi yang luas di antara pendapatan individu. Rentang pendapatan yang sangat besar dari 0.07 hingga 200 juga mengindikasikan ketidakmerataan yang mencolok. Persentil pendapatan dari 10% hingga 90% menunjukkan lonjakan signifikan, dari 0.6 hingga 4.8, menggambarkan bahwa sebagian besar populasi memiliki pendapatan yang relatif rendah, sementara sebagian kecil memiliki pendapatan yang jauh lebih tinggi. Data ini memberikan gambaran penting untuk analisis lebih lanjut tentang ketimpangan ekonomi dan kebutuhan akan intervensi kebijakan yang tepat.



Gambar 27 Distribusi Frekuensi Pendapatan

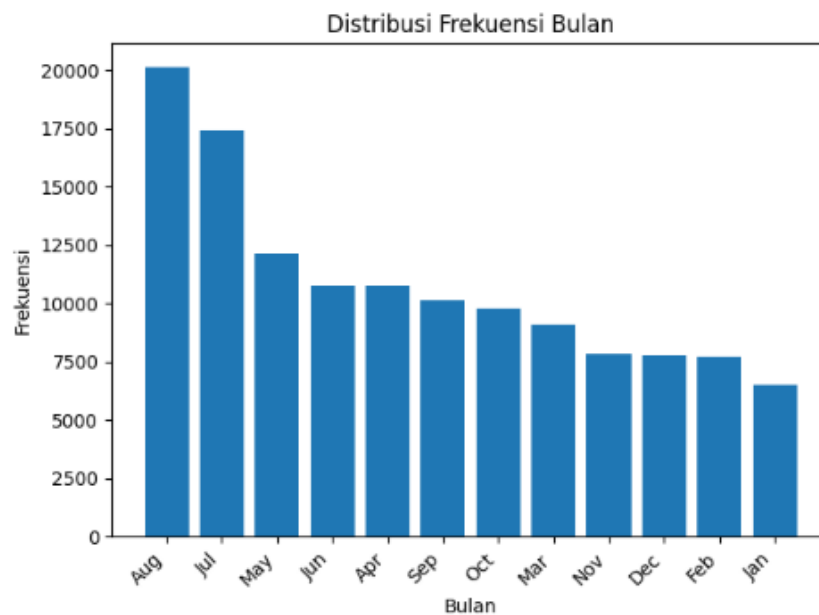
Interpretasi :

Gambar di atas menunjukkan sebuah histogram yang menggambarkan distribusi frekuensi pendapatan. dengan nilai maksimum sekitar 120.000. Histogram ini memperlihatkan bahwa sebagian besar pendapatan berada di kisaran 0 hingga 25, dengan jumlah frekuensi yang sangat tinggi di kisaran tersebut. Hal ini menunjukkan bahwa mayoritas populasi memiliki pendapatan yang relatif rendah.

Kesimpulan :

Kesimpulan dari histogram distribusi frekuensi pendapatan adalah bahwa sebagian besar pendapatan berada di kisaran 0 hingga 25, dengan jumlah frekuensi yang sangat tinggi di kisaran tersebut. Hal ini menunjukkan bahwa mayoritas populasi memiliki pendapatan yang relatif rendah. Distribusi ini dapat menggambarkan ketimpangan pendapatan yang signifikan dalam populasi. Data ini penting untuk analisis lebih lanjut mengenai kebijakan ekonomi dan sosial yang dapat membantu mengurangi ketimpangan pendapatan dan meningkatkan kesejahteraan masyarakat.

4.4.8 Statistik Deskriptif Bulan



Gambar 28 Distibusi Frekuensi Bulan

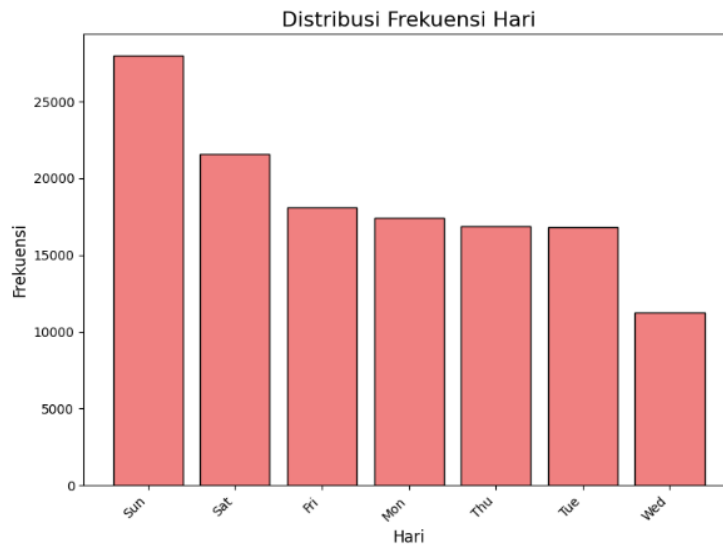
Interpretasi:

Gambar di atas menunjukkan sebuah diagram batang yang menggambarkan distribusi frekuensi bulanan. Dari grafik ini, terlihat bahwa bulan Agustus (Aug) memiliki frekuensi tertinggi, diikuti oleh bulan Juli (Jul) dan Mei (May). Frekuensi menurun secara bertahap dari bulan Agustus hingga Januari, dengan bulan Januari memiliki frekuensi terendah. Grafik ini menarik karena menunjukkan pola distribusi frekuensi yang menurun dari pertengahan tahun hingga awal tahun berikutnya, yang bisa relevan untuk analisis tren musiman atau perencanaan kegiatan berdasarkan frekuensi bulanan.

Kesimpulan:

Kesimpulan dari diagram batang distribusi frekuensi bulanan tersebut adalah bahwa bulan Agustus memiliki frekuensi tertinggi, diikuti oleh bulan Juli dan Mei. Frekuensi kemudian menurun secara bertahap dari bulan Agustus hingga Januari, dengan bulan Januari memiliki frekuensi terendah. Pola ini menunjukkan adanya puncak aktivitas atau kejadian tertentu pada pertengahan tahun, yang menurun menjelang akhir dan awal tahun berikutnya. Data ini dapat memberikan wawasan berharga untuk analisis tren musiman atau perencanaan kegiatan berdasarkan frekuensi bulanan.

4.4.9 Statistik Deskriptif Hari



Gambar 29 Distribusi Frekuensi Hari

Interpretasi:

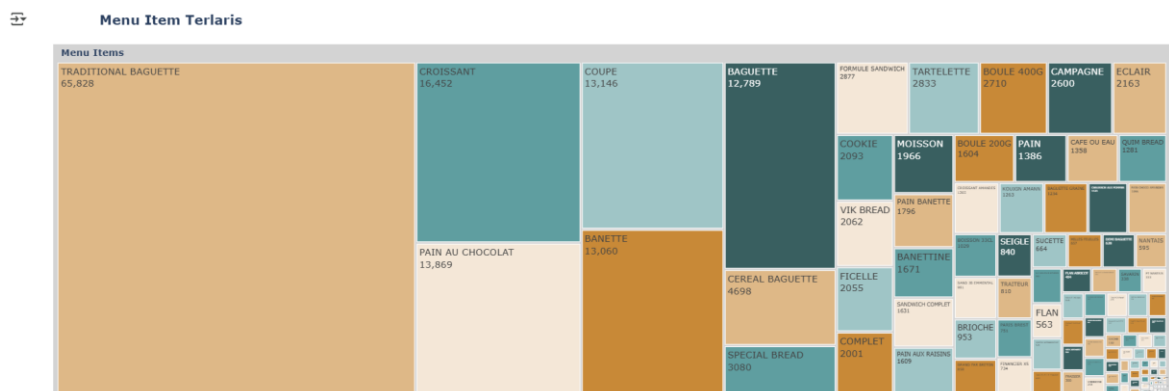
Gambar ini menunjukkan sebuah diagram batang yang menggambarkan distribusi frekuensi hari dalam seminggu. Judul diagram ini adalah "Distribusi Frekuensi Hari". Sumbu vertikal (y) diberi label "Frekuensi" dan menunjukkan jumlah frekuensi yang berkisar dari 0 hingga 30.000. Sumbu horizontal (x) diberi label "Hari" dan menunjukkan hari-hari dalam seminggu, mulai dari Minggu (Sun), Sabtu (Sat), Jumat (Fri), Senin (Mon), Kamis (Thu), Selasa (Tue), hingga Rabu (Wed).

Kesimpulan:

Kesimpulan dari diagram batang distribusi frekuensi hari dalam seminggu adalah bahwa hari Minggu memiliki frekuensi tertinggi, diikuti oleh Sabtu, Jumat, Senin, Kamis, dan Selasa. Hari Rabu memiliki frekuensi terendah. Hal ini menunjukkan bahwa aktivitas atau kejadian tertentu paling sering terjadi pada hari Minggu, sementara paling jarang terjadi pada hari Rabu. Pola ini bisa memberikan wawasan tentang kebiasaan, aktivitas, atau perilaku pada hari-hari tertentu dalam seminggu. Analisis lebih lanjut dapat mengungkap alasan di balik distribusi ini.

4.5 Visualisasi Data

4.5.1 Visualisasi Data Menu Item Terlaris



Gambar 30 Distribusi Penjualan Item Menu

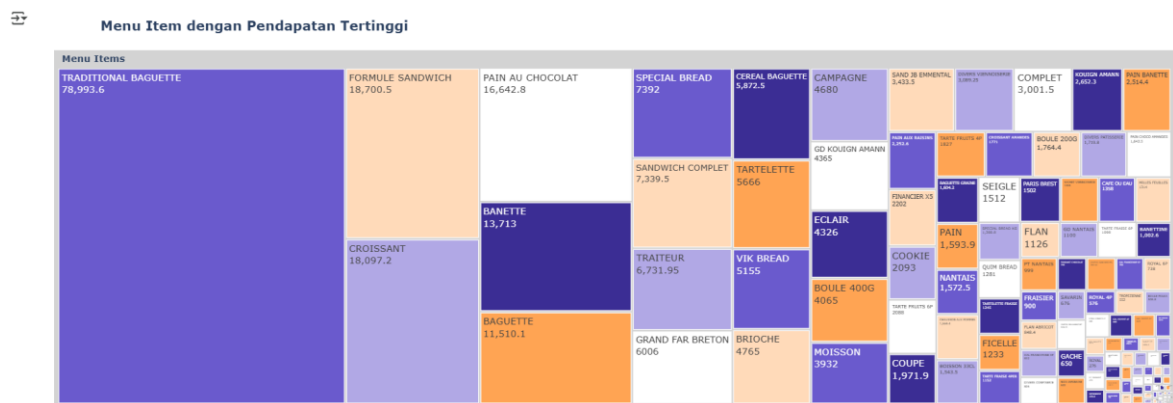
Interpretasi:

Gambar treemap di atas menunjukkan distribusi penjualan item menu berdasarkan popularitasnya. Traditional Baguette mendominasi penjualan dengan angka yang sangat tinggi, yaitu sebanyak 65.828 penjualan, menjadikannya produk paling laris dan favorit pelanggan. Di posisi berikutnya, Croissant dan Pain au Chocolat masing-masing mencatatkan penjualan sebesar 16.452 dan 13.869, menunjukkan bahwa produk roti dan pastry ini juga sangat diminati.

Kesimpulan:

Data ini mencerminkan pentingnya Traditional Baguette sebagai produk utama yang perlu difokuskan dalam produksi, pemasaran, dan pengembangannya.. Selain itu, produk seperti Croissant dan Pain au Chocolat juga memiliki potensi untuk dikembangkan lebih lanjut. Untuk item dengan penjualan rendah, bisnis dapat mengevaluasi apakah produk tersebut tetap dipertahankan, dikurangi, atau mungkin dipromosikan lebih efektif guna meningkatkan daya tariknya kepada pelanggan. Interpretasi ini memberikan wawasan strategis bagi toko roti untuk mengoptimalkan portofolio produknya berdasarkan permintaan pelanggan.

4.5.2 Visualisasi Data Menu Item dengan Pendapatan Tertinggi



Gambar 31 Distribusi Menu Item dengan Pendapatan Tertinggi

Interpretasi:

Produk dengan pendapatan tertinggi dalam dataset adalah "Traditional Baguette" dengan total pendapatan sebesar 79.993,6, menunjukkan bahwa produk ini merupakan item unggulan dan paling diminati konsumen. Item lain seperti "Formule Sandwich" (18.700,5), "Pain au Chocolat" (16.542,8), dan "Croissant" (18.057,2) juga memberikan kontribusi besar, mengindikasikan popularitas tinggi di kalangan pelanggan. Sementara itu, produk seperti "Baguette" (11.510,1) dan "Banette" (13.713) berada pada kategori pendapatan menengah, tetap memainkan peran penting meskipun kontribusinya tidak sebesar produk utama. Beberapa item dengan pendapatan rendah, seperti "Ficelle" (1.233), "Pain" (1.593,9), dan "Nantais" (1.572,5), kemungkinan merupakan produk niche atau kurang populer. Secara umum, dominasi "Traditional Baguette" menunjukkan ketergantungan yang signifikan terhadap produk ini, sementara diversifikasi menu dengan berbagai item lainnya membantu memenuhi preferensi konsumen yang beragam.

Kesimpulan:

Toko Roti bisa memperkuat strategi promosi atau produksi pada item seperti "Traditional Baguette" dan "Formule Sandwich" karena mereka memberikan kontribusi besar. Evaluasi Produk Niche: Item dengan pendapatan rendah dapat dievaluasi apakah perlu peningkatan pemasaran, pengurangan, atau bahkan inovasi untuk meningkatkan daya tariknya.

4.5.3 Visualisasi Data Pendapatan Per-Bulan



Gambar 32 Grafik Pendapatan Per-bulan

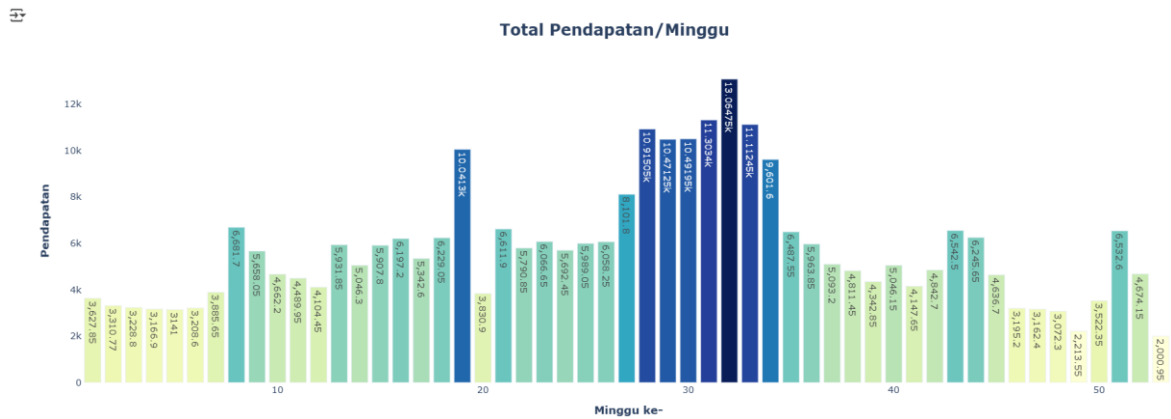
Interpretasi:

Berdasarkan analisis pendapatan bulanan, Agustus mencatat pendapatan tertinggi sebesar 48.713,2k, diikuti oleh Juli dengan 42.150,25k, kemungkinan dipengaruhi oleh faktor musiman, promosi besar, atau puncak aktivitas tertentu. Sebaliknya, pendapatan terendah terjadi pada Januari (15.335,27k) dan Februari (16.916,95k), menunjukkan awal tahun sebagai periode dengan aktivitas penjualan yang lebih rendah. Tren pendapatan secara umum meningkat pada pertengahan tahun, khususnya dari Mei hingga Agustus, sebelum menurun secara bertahap menjelang akhir tahun, mengindikasikan adanya pola musiman seperti aktivitas musim panas atau liburan tertentu. Beberapa bulan, seperti Juni (24.848,7k) dan April (24.838,5k), menunjukkan pendapatan yang stabil, mencerminkan konsistensi penjualan pada periode tersebut.

Kesimpulan:

Pendapatan bulanan menunjukkan adanya pola musiman yang signifikan, dengan puncak pendapatan terjadi pada Agustus, kemungkinan dipengaruhi oleh aktivitas musiman, promosi, atau peningkatan permintaan selama musim panas. Sebaliknya, awal tahun, khususnya Januari dan Februari, menjadi periode dengan pendapatan terendah, mencerminkan aktivitas penjualan yang lebih rendah pada waktu tersebut. Tren ini menunjukkan pentingnya strategi yang fokus pada memaksimalkan pendapatan selama puncak musim dan meningkatkan penjualan di bulan-bulan dengan aktivitas rendah. Stabilitas pendapatan pada bulan-bulan tertentu, seperti April dan Juni, menunjukkan potensi untuk mempertahankan performa penjualan yang konsisten di luar musim puncak.

4.5.4 Visualisasi Data Total Pendapatan/Minggu



Gambar 33 Distribusi Pendapatan Per-minggu

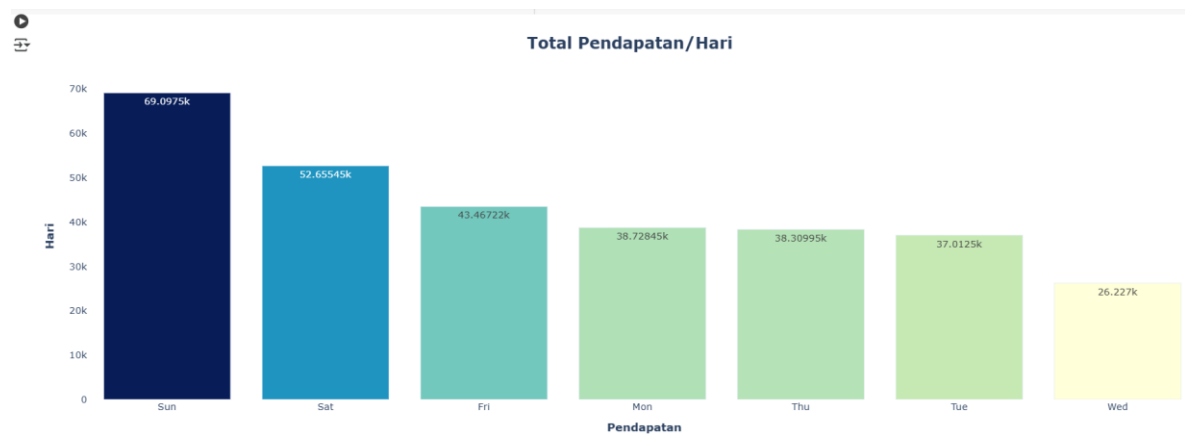
Interpretasi:

Grafik menunjukkan fluktuasi pendapatan mingguan sepanjang periode yang diamati, dengan sumbu X merepresentasikan minggu keberapa dan sumbu Y menunjukkan jumlah pendapatan (dalam ribuan atau jutaan). Puncak pendapatan tertinggi tercatat pada minggu ke-39 dengan nilai sekitar 13.047k, sementara minggu-minggu lainnya, seperti minggu ke-20 dan ke-38, juga menunjukkan pendapatan yang menonjol. Sebagian besar minggu lainnya memiliki pendapatan yang lebih stabil dengan rata-rata di kisaran 4.000–7.000. Tren peningkatan signifikan pada minggu-minggu tertentu kemungkinan disebabkan oleh faktor-faktor khusus seperti promosi, musim liburan, atau aktivitas bisnis tertentu. Sebaliknya, minggu-minggu dengan pendapatan lebih rendah, seperti minggu ke-1, ke-41, dan ke-52, mungkin mencerminkan dampak akhir periode bisnis, faktor musiman, atau penurunan aktivitas secara keseluruhan.

Kesimpulan:

Pendapatan mingguan menunjukkan fluktuasi yang signifikan sepanjang periode yang diamati, dengan beberapa minggu mengalami puncak pendapatan, terutama pada minggu ke-39, yang bisa dipengaruhi oleh faktor musiman, promosi, atau aktivitas khusus lainnya. Minggu-minggu dengan pendapatan lebih rendah, seperti minggu pertama dan terakhir tahun, kemungkinan mencerminkan penurunan aktivitas atau faktor musiman, yang mempengaruhi performa penjualan. Secara keseluruhan, meskipun ada variasi dalam pendapatan setiap minggunya, pola-pola ini memberikan wawasan tentang pengaruh faktor eksternal terhadap kinerja pendapatan, yang bisa digunakan untuk merencanakan strategi penjualan lebih efektif di masa mendatang.

4.5.5 Visualisasi Data Pendapatan/Hari



Gambar 34 Distribusi Pendapatan Per-hari

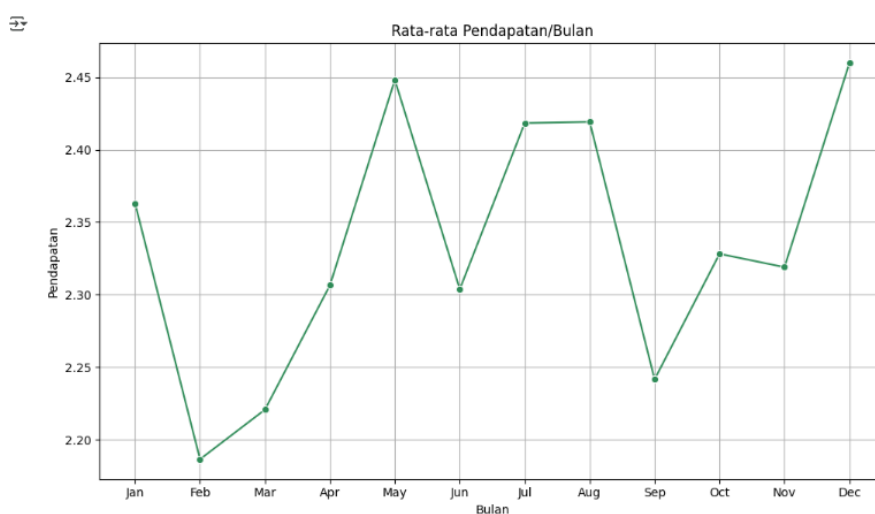
Interpretasi:

Diagram batang horizontal tersebut menunjukkan total pendapatan berdasarkan hari dalam seminggu. Dari visualisasi, terlihat bahwa hari Minggu (Sun) mencatat pendapatan tertinggi sebesar 69.097k, diikuti oleh Sabtu (Sat) dengan 52.655k, mengindikasikan bahwa akhir pekan adalah waktu paling produktif dalam menghasilkan pendapatan. Sebaliknya, hari Rabu (Wed) memiliki pendapatan terendah sebesar 26.227k, dengan hari-hari kerja lain seperti Selasa (Tue) dan Kamis (Thu) juga menunjukkan pendapatan yang relatif rendah dibandingkan akhir pekan.

Kesimpulan:

Dapat disimpulkan bahwa akhir pekan adalah waktu yang ideal untuk memaksimalkan pendapatan. Strategi bisnis seperti promosi, diskon, atau peluncuran produk baru dapat dilakukan di akhir pekan untuk memanfaatkan tingginya aktivitas ekonomi. Untuk hari kerja dengan pendapatan rendah, perusahaan dapat mempertimbangkan promosi khusus untuk menarik lebih banyak pelanggan dan meningkatkan pendapatan.

4.5.6 Visualisasi Data Rata-Rata Pendapatan/Bulan



Gambar 35 Grafik Rata-rata Pendapatan Per-bulan

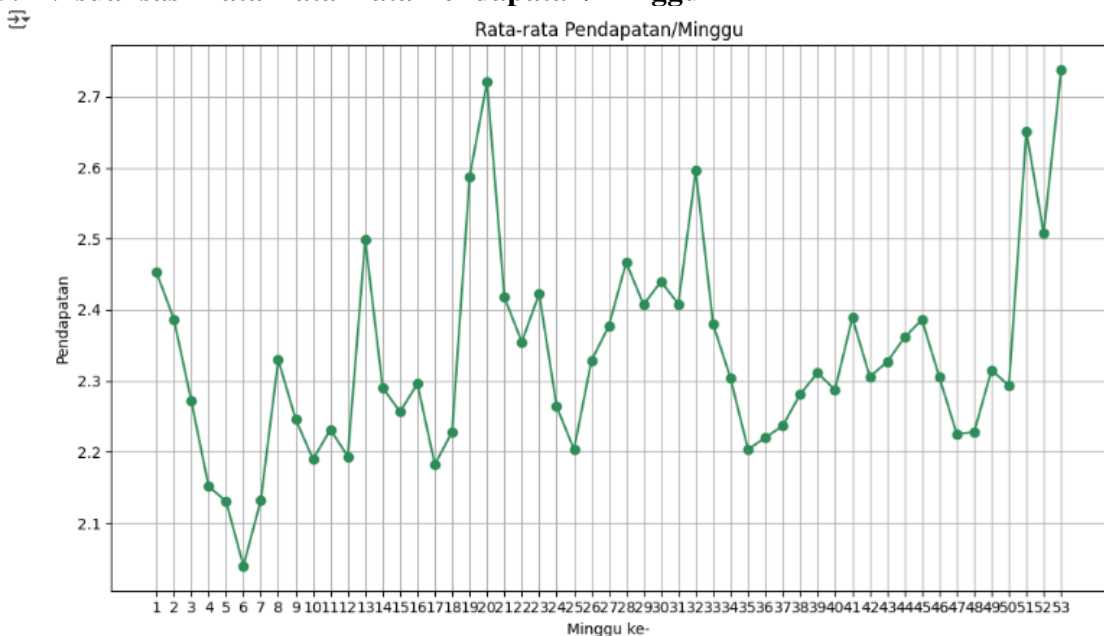
Interpretasi:

Grafik ini menunjukkan rata-rata pendapatan per bulan selama satu tahun. Pendapatan bergerak naik dan turun selama tahun berjalan dengan beberapa puncak dan lembah. Ada penurunan tajam dari Januari ke Februari, kemudian peningkatan signifikan dari Maret hingga Mei. Pendapatan tertinggi terjadi pada bulan Mei dan Desember, mendekati atau melebihi angka 2,45. Pendapatan terendah tercatat pada Februari dan September, berada di sekitar 2,20. Setelah penurunan di September, pendapatan menunjukkan tren meningkat secara konsisten hingga Desember.

Kesimpulan:

Ada kenaikan tajam di awal tahun (Februari ke Mei). Penurunan terjadi pada pertengahan tahun (Mei hingga September). Pendapatan cenderung meningkat kembali menjelang akhir tahun (Oktober hingga Desember). Naiknya pendapatan pada bulan Mei dan Desember bisa jadi terkait dengan momen tertentu, seperti hari raya, liburan, atau musim tertentu yang meningkatkan aktivitas ekonomi.

4.5.7 Visualisasi Data Rata-Rata Pendapatan/Minggu



Gambar 36 Grafik Rata-rata Pendapatan Per-minggu

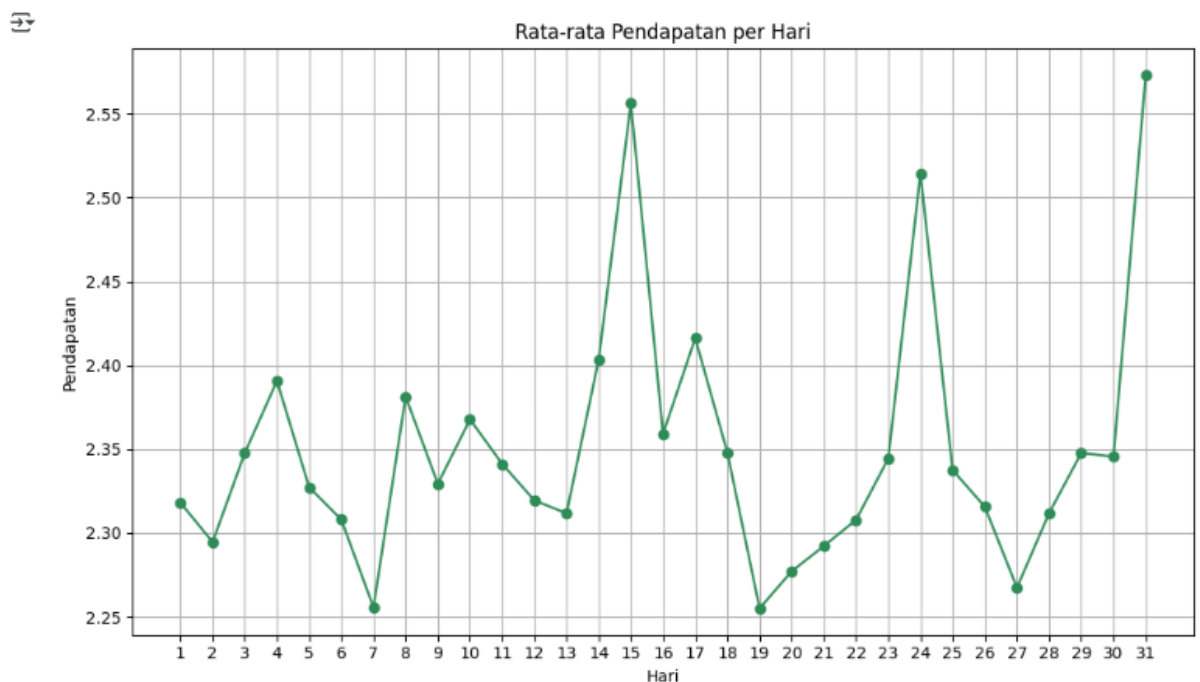
Interpretasi:

Grafik yang ditampilkan menunjukkan rata-rata pendapatan per minggu sepanjang 53 minggu (1 tahun). Grafik memperlihatkan pola yang fluktuatif dengan naik-turunnya pendapatan secara signifikan di beberapa minggu. Beberapa minggu memiliki penurunan tajam, seperti sekitar minggu ke-4 dan ke-12. Pendapatan tertinggi terjadi pada minggu ke-19 dan minggu ke-53, mendekati angka 2,7. Pendapatan terendah terlihat di minggu ke-4, berada di bawah angka 2,1. Pendapatan cenderung meningkat mulai dari minggu ke-47 hingga minggu ke-53.

Kesimpulan:

Meskipun fluktuasi ada, ada kecenderungan peningkatan pendapatan yang signifikan pada minggu-minggu terakhir (menuju akhir periode). Beberapa bagian grafik menunjukkan stabilitas pendapatan, misalnya di tengah periode (minggu ke-25 hingga 40), dengan sedikit variasi.

4.5.8 Visualisasi Data Rata-Rata Pendapatan/Hari



Gambar 37 Grafik Rata-rata Pendapatan Per-hari

Interpretasi:

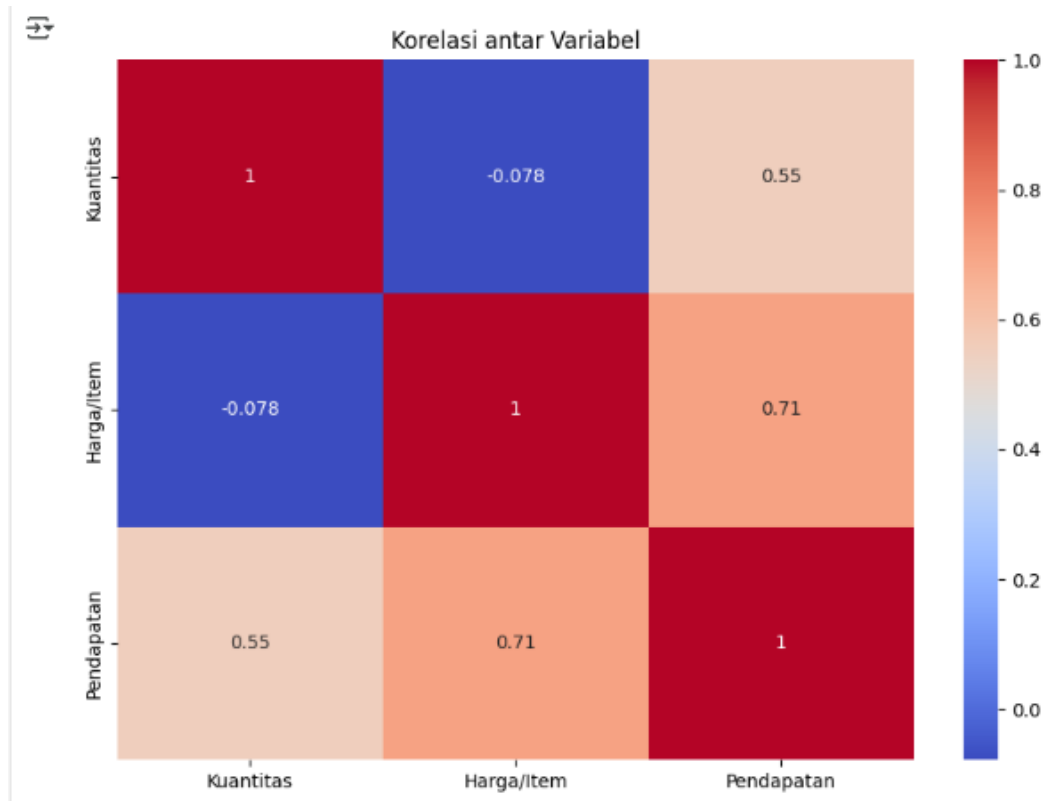
Grafik ini menunjukkan rata-rata pendapatan per hari dalam satu bulan (berisi hingga 31 hari). Pendapatan mengalami fluktuasi signifikan dari hari ke hari. Ada beberapa puncak tinggi di hari-hari tertentu, seperti hari ke-15, 24, dan 31, dengan pendapatan mencapai angka sekitar 2,55 atau lebih. Pendapatan terendah terjadi pada hari ke-6, 13, dan 27, dengan angka mendekati atau sedikit di atas 2,25. Ada peningkatan pendapatan yang jelas mulai dari hari ke-28 hingga hari ke-31. Hal ini sering dikaitkan dengan periode akhir bulan, di mana aktivitas belanja cenderung meningkat.

Kesimpulan:

Ada lonjakan pendapatan yang signifikan di beberapa titik waktu, seperti di pertengahan bulan (hari ke-15) dan akhir bulan (hari ke-30). Setelah lonjakan, sering diikuti penurunan tajam, menunjukkan adanya variasi atau pola yang mungkin terkait dengan faktor tertentu seperti hari tertentu atau promosi. Lonjakan di akhir bulan (hari ke-30) dapat dikaitkan dengan penutupan laporan keuangan atau promosi akhir bulan. Fluktuasi signifikan mungkin dipengaruhi oleh pola aktivitas pelanggan tertentu atau peristiwa spesifik di hari-hari tertentu.

4.6 Korelasi Data

4.6.1 Korelasi Antar Variabel



Gambar 38 Korelasi Variabel

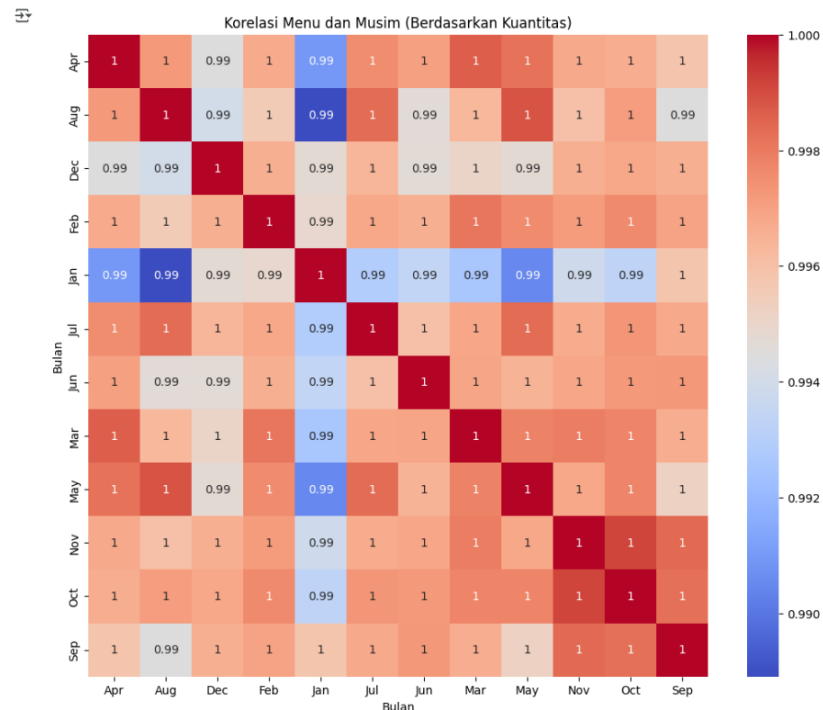
Interpretasi:

- Hubungan Kuantitas dan Harga/Item: Nilai korelasi sebesar -0.078 menunjukkan hubungan yang sangat lemah dan negatif. Artinya, kenaikan atau penurunan kuantitas tidak memiliki hubungan yang berarti dengan harga per item.
- Hubungan Kuantitas dan Pendapatan: Nilai korelasi sebesar 0.55 menunjukkan hubungan positif sedang. Artinya, jika kuantitas meningkat, pendapatan cenderung meningkat pula dengan korelasi sedang.
- Hubungan Harga/Item dan Pendapatan: Nilai korelasi sebesar 0.71 menunjukkan hubungan positif yang cukup kuat. Artinya, jika harga per item meningkat, pendapatan cenderung meningkat pula.

Kesimpulan:

Kuantitas memiliki hubungan yang lebih erat dengan pendapatan dibandingkan dengan harga per item, meskipun korelasi antara harga per item dan pendapatan lebih kuat. Hal ini menunjukkan bahwa peningkatan jumlah produk yang terjual cenderung memiliki dampak lebih besar terhadap pendapatan, meskipun kenaikan harga per item juga dapat meningkatkan pendapatan. Di sisi lain, hubungan antara harga per item dan kuantitas hampir tidak ada, yang berarti kedua variabel ini tidak saling memengaruhi secara langsung. Korelasi berkisar antara -1 (hubungan negatif sempurna) hingga 1 (hubungan positif sempurna), dengan nilai 0 menunjukkan tidak ada hubungan. Oleh karena itu, korelasi yang lebih tinggi, seperti 0.71, menunjukkan hubungan yang lebih signifikan dibandingkan dengan nilai yang mendekati 0, seperti -0.078.

4.6.2 Korelasi Menu dan Musim (Berdasarkan Kuantitas)



Gambar 39 Korelasi Menu dan Musim (Berdasarkan Kuantitas)

Interpretasi :

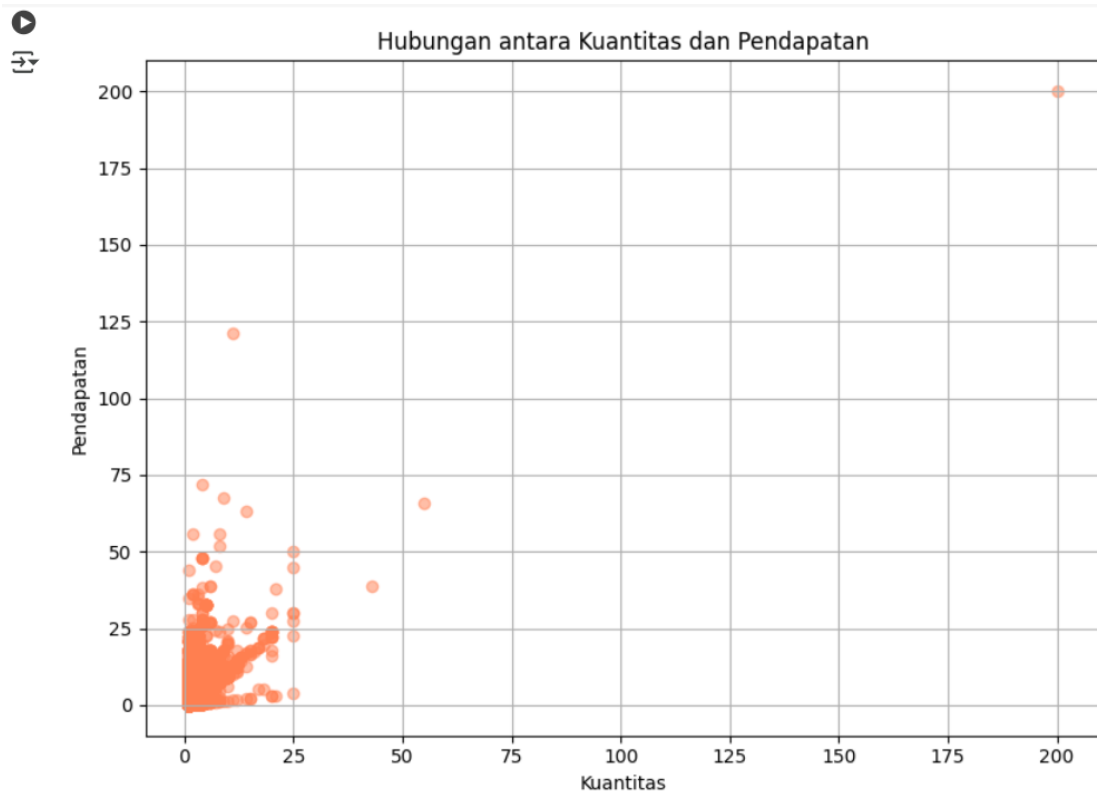
Heatmap memvisualisasikan matriks korelasi antara penjualan berbagai menu di setiap bulan. Berikut poin-poin interpretasinya: Merah: Menunjukkan korelasi positif. Artinya, jika penjualan satu menu meningkat di suatu bulan, penjualan menu lain juga cenderung meningkat di bulan yang sama. Semakin merah warnanya, semakin kuat korelasinya. Biru: Menunjukkan korelasi negatif. Artinya, jika penjualan satu menu meningkat di suatu bulan, penjualan menu lain cenderung menurun di bulan yang sama. Semakin biru warnanya, semakin kuat korelasinya. Putih: Menunjukkan korelasi mendekati nol. Artinya, tidak ada hubungan linear yang kuat antara penjualan dua menu di bulan tersebut.

Angka yang ditampilkan di setiap sel heatmap menunjukkan nilai korelasi antara dua menu di bulan tersebut. Nilai berkisar dari -1 hingga +1. Nilai +1 menunjukkan korelasi positif sempurna, -1 menunjukkan korelasi negatif sempurna, dan 0 menunjukkan tidak ada korelasi.

Kesimpulan :

Heatmap menunjukkan bahwa penjualan beberapa menu dipengaruhi oleh musim (bulan). Menu-menu tertentu mungkin lebih populer di bulan-bulan tertentu, sementara yang lain kurang populer. Heatmap juga menunjukkan hubungan antara penjualan berbagai menu. Beberapa menu mungkin memiliki korelasi positif yang kuat, menunjukkan bahwa mereka sering dibeli bersamaan. Ada juga menu yang memiliki korelasi negatif, menunjukkan bahwa mereka cenderung dibeli secara terpisah. Informasi dari heatmap dapat digunakan untuk membuat strategi bisnis toko roti yang lebih baik.

4.6.3 Hubungan Antara Kuantitas dan Pendapatan



Gambar 40 Hubungan Antara Kuantitas dan Pendapatan

Interpretasi:

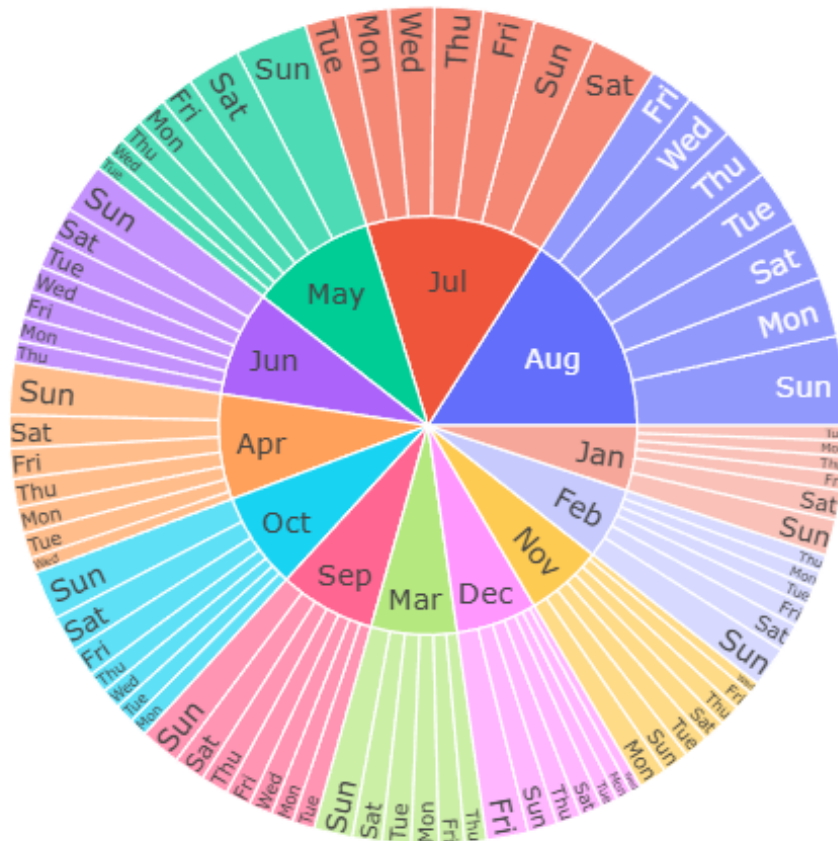
Scatter Plot menampilkan setiap data sebagai titik pada grafik. Sumbu X mewakili kuantitas barang yang terjual, dan sumbu Y mewakili pendapatan yang dihasilkan. Warna titik adalah coral dan memiliki transparansi ($\alpha=0.5$) agar titik yang tumpang tindih tetap terlihat.

Scatter Plot ini menunjukkan hubungan positif antara kuantitas dan pendapatan. Semakin banyak kuantitas barang yang terjual, semakin tinggi pendapatan yang dihasilkan. Hal ini terlihat dari titik-titik yang cenderung membentuk pola menaik dari kiri bawah ke kanan atas.

Kesimpulan:

Terdapat korelasi positif antara kuantitas dan pendapatan. Artinya, ketika kuantitas penjualan meningkat, pendapatan juga cenderung meningkat. Pengaruh Kuantitas terhadap Pendapatan: Kuantitas penjualan memiliki pengaruh yang signifikan terhadap pendapatan. Semakin banyak barang yang terjual, semakin besar pendapatan yang dihasilkan. Dengan mengetahui kuantitas penjualan, kita dapat memperkirakan pendapatan yang akan dihasilkan. Semakin tinggi kuantitas, semakin tinggi pula prediksi pendapatannya.

4.6.4 Hubungan Pendapatan Per-Hari dan Per-Bulan



Gambar 41 Hubungan Pendapatan Per-hari Dan Per-bulan

Penjelasan :

Program ini sangat berguna untuk analisis data hierarkis dan mendukung pengambilan keputusan berbasis data. Diagram sunburst memberikan wawasan visual yang cepat tentang pola pendapatan per bulan dan hari. Pusat Diagram (menunjukkan nama bulan). Bulan yang memiliki sektor lebih besar berarti menghasilkan pendapatan yang lebih tinggi. Dari diagram ini kita dapat melihat bulan mana yang memberikan pendapatan tertinggi. Mengidentifikasi hari-hari tertentu yang lebih produktif untuk setiap bulan. Menemukan pola yang berguna, seperti apakah akhir pekan lebih menguntungkan dibandingkan hari kerja. Sektor tiap hari menunjukkan kontribusi hari tersebut terhadap pendapatan di bulan terkait. Warna digunakan untuk membedakan bulan dan memberikan tampilan yang jelas.

Kesimpulan :

Dari *chart*, kita dapat melihat bahwa bulan-bulan tertentu (misalnya, bulan dengan potongan terbesar) cenderung memiliki pendapatan yang lebih tinggi dibandingkan bulan lainnya. Demikian juga, hari-hari tertentu dalam seminggu (misalnya, hari dengan potongan terbesar dalam lingkaran bulan) cenderung memiliki pendapatan yang lebih tinggi dibandingkan hari lainnya. Sebaliknya, bulan dan hari dengan potongan terkecil menunjukkan pendapatan yang lebih rendah. *Sunburst chart* juga dapat membantu mengidentifikasi pola pendapatan. Misalnya, kita mungkin melihat bahwa pendapatan cenderung lebih tinggi pada akhir pekan atau selama bulan-bulan liburan tertentu.

5. KESIMPULAN

Penelitian ini menunjukkan bahwa analisis data penjualan menggunakan Python memberikan wawasan penting mengenai pola permintaan produk bakery dan tren musiman yang memengaruhi penjualan. Dengan teknik statistik deskriptif dan visualisasi data, manajemen dapat merencanakan strategi pemasaran dan produksi yang lebih efektif. Pembersihan dan normalisasi data sebelum analisis sangat penting untuk akurasi hasil. Integrasi teknologi seperti Python dan Google Colab memungkinkan pengelolaan dataset besar secara efisien, meningkatkan kepuasan pelanggan dan memperkuat posisi pasar.

Untuk meningkatkan efisiensi operasional, disarankan agar toko roti memanfaatkan teknologi analisis data lebih lanjut, diversifikasi produk, serta melatih staf. Monitoring dan evaluasi berkala, kolaborasi dengan pihak ketiga, dan pengembangan sistem manajemen stok berbasis data akan membantu dalam pengelolaan inventaris dan meningkatkan daya saing toko roti.

6. DAFTAR PUSTAKA

- [1]. Chopra, S., Meindl, P., & Kalra, A. (2021). *Supply chain management: Strategy, planning, and operation* (7th ed.). Pearson.
- [2]. Few, S. (2009). *Now you see it: Simple visualization techniques for quantitative analysis*. Analytics Press.
- [3].Gimbert, M. (2021). *French Bakery Daily Sales*. Kaggle, tersedia pada <https://www.kaggle.com/datasets/matthieugimbert/french-bakery-daily-sales>. Diakses pada 01 Januari 2025.
- [4]. Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
- [5]. Kumar, A., & Sharma, M. (2016). Statistical methods for forecasting retail sales in bakery business. *Journal of Operations Research*, 8(3), 201–213.
- [6]. Lee, J., Choi, Y., & Han, S. (2018). Data-driven decision making in retail: Predicting customer behavior in bakeries. *Journal of Retail Analytics*, 5(2), 113–127.
- [7]. Peng, Y., & Zhang, D. (2017). Machine learning for time series forecasting: A comparative study of algorithms. *Journal of Applied Data Science*, 3(4), 89–105.
- [8]. Sharma, A., Bhatnagar, R., & Gupta, P. (2020). Customer behavior analysis and forecasting in the retail sector. *International Journal of Business Intelligence and Data Mining*, 14(1), 49–68.