

# **Project Analisis Dan Visualisasi Data Superstore**



Nama penyusun : Muhamad Azril Rafsanzani

Tanggal Penyusunan : 07/11/2025

## DAFTAR ISI

<b>Bab I Pendahuluan .....</b>	<b>3</b>
1.1 Latar Belakang.....	3
1.2 Rumusan Masalah.....	3
1.3 Tujuan Analisis .....	3
1.4 Manfaat Analisis.....	3
<b>Bab II Metodologi Analisis Data .....</b>	<b>4</b>
2.1 Sumber dan Jenis Data .....	4
2.2 Alat Analisis .....	4
2.3 Tahapan Analisis Data .....	4
<b>Bab III Hasil dan Pembahasan .....</b>	<b>5</b>
3.1 Melakukan Data Cleaning .....	5
3.2 Melakukan Data Processing.....	6
3.3 Visualisasi Data, Grafik data Sales.....	7
3.4 Grafik data Profit .....	8
3.5 Frekuensi produk yang paling banyak terjual .....	9
<b>Bab IV Kesimpulan dan Saran .....</b>	<b>10</b>
4.1 Kesimpulan .....	10
4.2 Saran .....	10
<b>Lampiran .....</b>	<b>11</b>
A. Potongan Kode Program (Python).....	11

## **I. Pendahuluan**

### **1.1 Latar Belakang :**

Analisis dan visualisasi data dilakukan untuk mengetahui tren tahunan penjualan, profit, serta produk yang paling sering dibeli oleh pelanggan selama periode 2014–2017. Melalui analisis ini, dapat diketahui bagaimana perkembangan penjualan dari tahun ke tahun, seberapa besar keuntungan yang diperoleh, dan produk apa saja yang paling diminati.

Kegiatan ini penting karena hasil analisis dapat membantu perusahaan memahami pola pembelian pelanggan dan kinerja penjualan secara keseluruhan. Dengan visualisasi data, informasi yang diperoleh menjadi lebih mudah dipahami dan dapat digunakan sebagai dasar dalam pengambilan keputusan serta perencanaan strategi bisnis di masa depan.

### **1.2 Rumusan masalah :**

1. Bagaimana tren penjualan (*sales*) dari tahun 2014 hingga 2017?
2. Bagaimana perkembangan profit perusahaan dari tahun 2014 hingga 2017?
  
3. Produk apa saja yang paling sering dibeli oleh pelanggan selama tahun 2014–2017?

### **1.3 Tujuan Analisis :**

1. Mengetahui tren penjualan (*sales*) perusahaan dari tahun 2014 hingga 2017.
2. Menganalisis perkembangan profit perusahaan selama periode 2014 – 2017.
  
3. Mengetahui produk yang paling sering dibeli oleh pelanggan.

### **1.4 Manfaat Analisis :**

1. Sebagai bahan evaluasi kinerja perusahaan, khususnya dalam melihat perkembangan penjualan dan profit dari tahun ke tahun.

2. Membantu manajemen dalam pengambilan keputusan, terutama dalam menentukan strategi pemasaran, promosi, dan pengelolaan produk.
3. Mengetahui produk yang paling diminati pelanggan, sehingga perusahaan dapat memprioritaskan stok, pengadaan, dan distribusi barang yang memiliki permintaan tinggi.
4. Mengidentifikasi tren bisnis dari waktu ke waktu, yang dapat digunakan untuk memprediksi potensi penjualan di periode selanjutnya.
5. Mendukung perencanaan strategi peningkatan profit.

## II. Metodologi Analisis

### 2.1 Sumber Data :

Data yang digunakan dalam analisis ini diperoleh dari dataset yang tersedia pada platform *Data Courses* di Coursera. Dataset tersebut berisi informasi penjualan, profit, dan produk yang dibeli pelanggan selama periode tahun 2014 hingga 2017.

sample\_-\_superstore  
XLS File

### 2.2 Jenis dan Bentuk Data :

Data yang digunakan bersifat kuantitatif karena memuat nilai numerik seperti jumlah penjualan dan profit, serta kualitatif untuk menggambarkan kategori produk yang dibeli pelanggan. Selain itu, data ini juga berbentuk time series karena mencakup periode tahunan dari 2014 sampai 2017.

### 2.3 Alat Analisis:

Proses analisis dilakukan menggunakan Microsoft Excel untuk menampilkan dataset mentah maupun clean, serta Python (melalui library seperti *Pandas* dan *Matplotlib*) untuk proses analisis yang lebih mendalam dan visualisasi tren data.

## **2.4 Langkah-Langkah Analisis.**

Tahapan analisis dilakukan melalui beberapa langkah berikut:

- **Collect Data :** Memperoleh informasi yang akan digunakan dalam proses analisis.
- **Cleaning Data:** memeriksa dan menghapus data yang tidak valid, kosong, atau duplikat agar hasil analisis lebih akurat.
- **Pengolahan Data:** melakukan pengelompokan data berdasarkan tahun dan produk untuk mengetahui tren penjualan dan profit tahunan.
- **Visualisasi Data:** menyajikan hasil analisis dalam bentuk grafik dan diagram untuk memperjelas tren penjualan, profit, serta produk yang paling sering dibeli pelanggan.

## **III. Hasil dan Pembahasan**

Analisis data dilakukan dengan menggunakan bahasa pemrograman Python melalui platform Google Colab sebagai lingkungan pemrograman berbasis cloud. Penggunaan Python dipilih karena memiliki berbagai library yang mendukung proses analisis data, seperti *Pandas* untuk pengolahan data, *Matplotlib* dan *Seaborn* untuk visualisasi, serta *NumPy* untuk perhitungan numerik.

Seluruh proses analisis dimulai dari tahap pembersihan data (*data cleaning*), pengolahan data (*data processing*), hingga pembuatan grafik dan visualisasi untuk menampilkan hasil secara lebih informatif. Melalui tahapan tersebut, diperoleh informasi mengenai tren penjualan dan profit dari tahun 2014 hingga 2017, serta produk-produk yang paling sering dibeli oleh pelanggan.

Setelah semua tahap analisis dilakukan, berikut disajikan hasil dan pembahasan yang menggambarkan pola penjualan, perkembangan profit, serta preferensi pelanggan terhadap produk dalam kurun waktu empat tahun tersebut.

### **3.1 Melakukan Data cleaning :**

```

contains_8888 = (super == 8888).sum().sum()
contains_9999 = (super == 9999).sum().sum()
contains_strip = (super == '-').sum().sum()
print("Data contains 8888: ", contains_8888)
print("Data contains 9999: ", contains_9999)
print("Data contains strip: ", contains_strip)

Data contains 8888: 1
Data contains 9999: 0
Data contains strip: 0

super.replace(8888, np.nan, inplace=True)

contains_8888 = (super == 8888).sum().sum()
contains_9999 = (super == 9999).sum().sum()
contains_strip = (super == '-').sum().sum()
print("Data contains 8888: ", contains_8888)
print("Data contains 9999: ", contains_9999)
print("Data contains strip: ", contains_strip)

.. Data contains 8888: 0
Data contains 9999: 0
Data contains strip: 0

```

Pada tahap ini, kode program dijalankan untuk mendeteksi adanya nilai yang tidak sesuai atau tidak valid di dalam dataset, salah satunya berupa nilai anomali seperti ‘8888’ yang tidak memiliki makna sebenarnya dalam konteks data penjualan maupun profit.

Dari hasil pemeriksaan, ditemukan masih terdapat satu entri data yang mengandung nilai tidak valid tersebut. Nilai tersebut kemudian diganti dengan nilai yang sesuai agar tidak mengganggu hasil analisis. Setelah proses penggantian dilakukan, data kembali diperiksa menggunakan kode yang sama untuk memastikan bahwa seluruh nilai tidak valid telah dihapus dan diperbaiki.

### 3.2 Melakukan Data Processing :

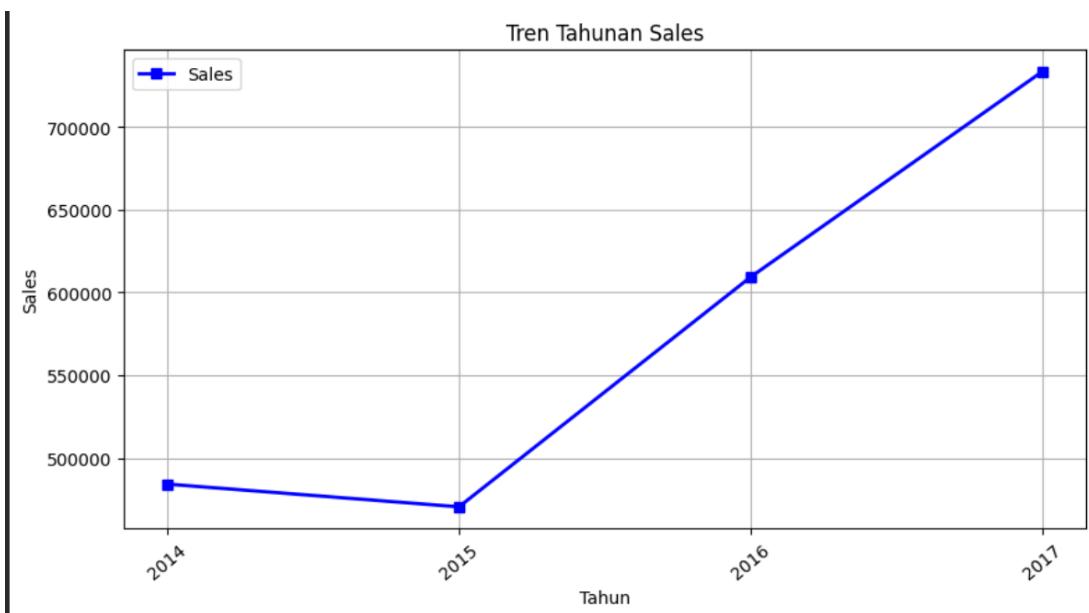
#Untuk menampilkan deskripsi statistik (nilai average, standar deviasi, nilai minimum, nilai maksimum dan kuartil)								
D = super.describe()								
	Row ID	Order Date	Ship Date	Postal Code	Sales	Quantity	Discount	Profit
count	9994.000000	9994	9994	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000
mean	4997.500000	2016-04-30 00:07:12.259355648	2016-05-03 23:06:58.571142912	55190.379428	229.858001	3.789574	0.156203	28.656896
min	1.000000	2014-01-03 00:00:00	2014-01-07 00:00:00	1040.000000	0.444000	1.000000	0.000000	6599.978000
25%	2499.250000	2015-05-23 00:00:00	2015-05-27 00:00:00	23223.000000	17.280000	2.000000	0.000000	1.728750
50%	4997.500000	2016-06-26 00:00:00	2016-09-29 00:00:00	56430.500000	54.490000	3.000000	0.200000	8.666500
75%	7495.750000	2017-05-14 00:00:00	2017-05-18 00:00:00	90008.000000	209.940000	5.000000	0.200000	29.364000
max	9994.000000	2017-12-30 00:00:00	2018-01-05 00:00:00	99301.000000	22638.480000	14.000000	0.800000	8399.976000
std	2885.163629	Nan	Nan	32063.693350	623.245101	2.225110	0.206452	234.260108

Setelah data berhasil dibersihkan dari nilai yang tidak valid, proses dilanjutkan ke tahap pengolahan data (data processing). Pada tahap ini dilakukan perhitungan

statistik deskriptif menggunakan kode Python untuk mengetahui gambaran umum dari data, seperti nilai rata-rata (mean), kuartil, nilai maksimum, dan minimum.

Tahap ini bertujuan untuk memahami karakteristik dasar dari data penjualan dan profit sebelum dilakukan analisis lebih lanjut. Dengan mengetahui hasil deskriptif statistik, kita dapat melihat pola dan persebaran data secara keseluruhan sehingga menjadi dasar untuk proses visualisasi dan interpretasi hasil berikutnya.

### 3.3 Visualisasi data, Grafik data Sales :



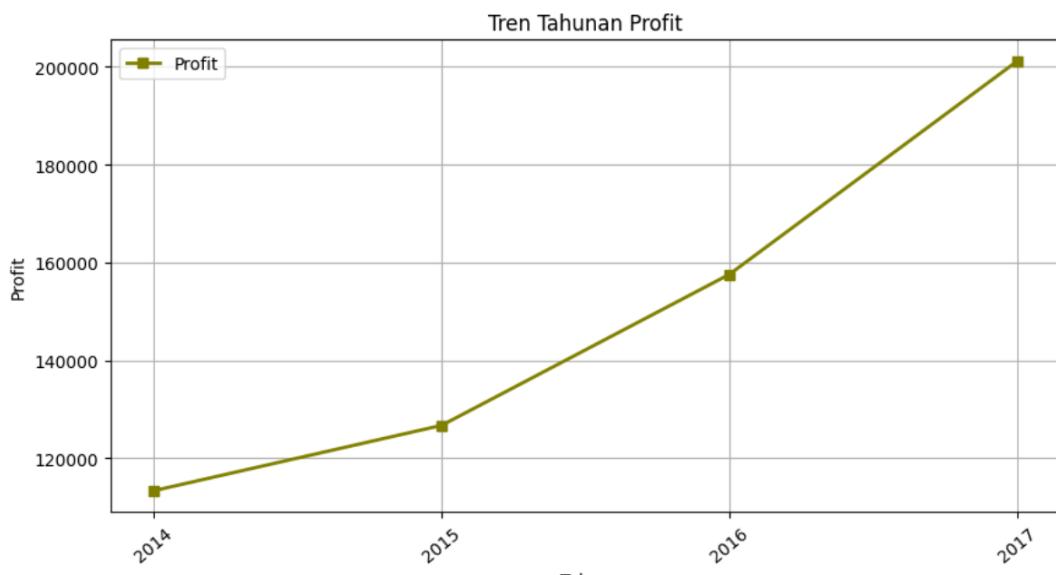
Berdasarkan hasil analisis dan visualisasi data, grafik tren tahunan penjualan (Sales) menunjukkan perubahan yang cukup dinamis dari tahun 2014 hingga 2017. Meskipun sempat menurun di awal periode, secara keseluruhan penjualan mengalami peningkatan yang signifikan hingga akhir periode analisis.

- **2014 ke 2015:** Penjualan mengalami sedikit penurunan dari sekitar **485.000** menjadi **470.000**. Periode ini merupakan satu-satunya tahun di mana terjadi penurunan penjualan selama masa analisis.
- **2015 ke 2016:** Penjualan meningkat tajam dari sekitar **470.000** menjadi **610.000**, yang menunjukkan adanya perbaikan strategi bisnis serta peningkatan permintaan dari pelanggan.

- **2016 ke 2017:** Penjualan terus mengalami kenaikan signifikan hingga mencapai sekitar **730.000**, menjadi titik tertinggi selama empat tahun terakhir dan mencerminkan kinerja penjualan yang sangat baik.

Secara keseluruhan, hasil analisis menunjukkan bahwa meskipun sempat turun di awal, **penjualan perusahaan terus tumbuh dengan baik** dan mencerminkan kinerja yang semakin positif dari tahun ke tahun.

### 3.4 Grafik data Profit :

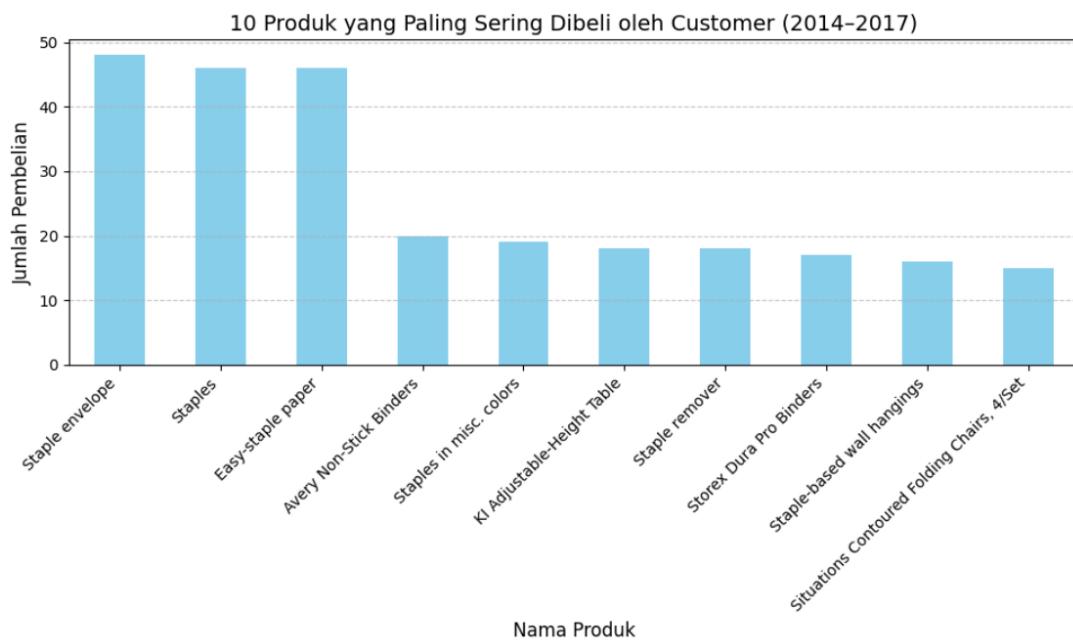


Berdasarkan hasil analisis dan visualisasi data, grafik tren tahunan menunjukkan bahwa profit perusahaan mengalami peningkatan yang konsisten dan signifikan setiap tahunnya selama periode 2014 hingga 2017.

- **Tahun 2014 ke 2015:** Profit meningkat dari sekitar **113.000** menjadi **126.000**. Kenaikan ini tergolong moderat, menandakan awal pertumbuhan yang stabil.
- **Tahun 2015 ke 2016:** Profit terus mengalami peningkatan dari sekitar **126.000** menjadi **158.000**. Pertumbuhan ini lebih tajam dibandingkan tahun sebelumnya, menunjukkan adanya peningkatan performa bisnis yang cukup signifikan.
- **Tahun 2016 ke 2017:** Profit melonjak tajam hingga mencapai titik tertinggi sekitar **201.000**. Lonjakan ini mencerminkan adanya akselerasi pertumbuhan yang kuat dan efisiensi yang lebih baik pada tahun terakhir periode analisis.

Secara keseluruhan, tren ini menggambarkan bahwa kinerja perusahaan terus membaik dari tahun ke tahun. Peningkatan profit yang konsisten mengindikasikan adanya strategi bisnis dan pengelolaan operasional yang efektif, serta respon pasar yang positif terhadap produk yang ditawarkan.

### 3.5 Frekuensi produk yang paling banyak terjual :



Hasil analisis frekuensi pembelian menunjukkan 10 produk yang paling sering dibeli oleh pelanggan. Produk yang menempati posisi teratas adalah **Staple envelope** dengan jumlah penjualan sebanyak **48 kali**.

Dua produk berada di posisi berikutnya dengan jumlah penjualan yang hampir setara, yaitu **Staples (Isi Staples)** dan **Easy-staple paper (Kertas Staples)**, yang masing-masing terjual sebanyak **46 kali**. Selanjutnya, terdapat penurunan frekuensi penjualan yang signifikan: **Avery Non-Stick Binders** terjual **20 kali**, diikuti oleh **Staples in misc. colors** sebanyak **19 kali**.

Di posisi menengah hingga bawah, terdapat **KI Adjustable-Height Table** dan **Staple remover**, yang masing-masing terjual sebanyak **18 kali**. Kemudian, **Storex Dura Pro Binders** terjual **17 kali**, **Staple-based wall hangings** sebanyak **16 kali**, dan yang paling sedikit dalam daftar 10 teratas adalah **Situations Contoured Folding Chairs, 4/Set**, dengan penjualan sebanyak **15 kali**.

## **IV. Kesimpulan dan saran**

### **4.1 Kesimpulan**

Berdasarkan hasil analisis data penjualan dan profit perusahaan dari tahun 2014 hingga 2017, dapat disimpulkan bahwa kinerja bisnis menunjukkan tren pertumbuhan yang positif. Penjualan sempat menurun sedikit pada tahun 2015, namun berhasil meningkat tajam pada tahun-tahun berikutnya hingga mencapai puncaknya pada tahun 2017. Penjualan yang awalnya berada di kisaran 485.000 pada tahun 2014 sempat turun menjadi 470.000 pada tahun 2015, namun kemudian melonjak tajam hingga mencapai 730.000 pada tahun 2017. Hal ini menunjukkan bahwa perusahaan mampu beradaptasi dan memperbaiki strategi penjualannya dengan baik.

Profit perusahaan juga mengalami peningkatan yang stabil setiap tahun, mencerminkan efisiensi operasional dan pengelolaan yang efektif. Dari sekitar 113.000 pada tahun 2014, profit meningkat secara konsisten hingga mencapai 201.000 di tahun 2017. Kenaikan ini menunjukkan bahwa perusahaan tidak hanya berhasil meningkatkan penjualan, tetapi juga mampu mengelola biaya dan sumber daya secara efisien, sehingga menghasilkan keuntungan yang lebih tinggi setiap tahunnya.

Selain itu, hasil analisis menunjukkan bahwa produk yang paling sering dibeli pelanggan adalah kebutuhan kantor dasar seperti *Staple envelope*, *Staples*, dan *Easy-staple paper*. Produk-produk tersebut menjadi kontributor utama dalam volume transaksi dan berperan penting dalam menjaga stabilitas penjualan perusahaan.

Secara keseluruhan, perusahaan berhasil menunjukkan performa yang kuat dan konsisten, baik dari segi penjualan maupun profitabilitas, yang menandakan arah pertumbuhan bisnis yang berkelanjutan.

### **4.2 Saran**

Berdasarkan hasil analisis yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan oleh perusahaan untuk mendukung peningkatan kinerja di masa mendatang, yaitu:

1. Mempertahankan dan mengembangkan strategi penjualan yang efektif. Perusahaan perlu melanjutkan strategi yang terbukti berhasil meningkatkan penjualan setelah tahun 2015, seperti peningkatan promosi, penguatan hubungan dengan pelanggan, serta diversifikasi saluran distribusi agar jangkauan pasar semakin luas.
2. Meningkatkan efisiensi dan inovasi operasional. Mengingat profit mengalami peningkatan stabil setiap tahun, perusahaan dapat terus

memperkuat efisiensi biaya produksi dan distribusi. Selain itu, penerapan teknologi digital dalam proses bisnis dapat membantu meningkatkan produktivitas serta pengambilan keputusan berbasis data.

3. Fokus pada produk unggulan dan pengembangan produk baru. Produk seperti *Staple envelope*, *Staples*, dan *Easy-staple paper* terbukti menjadi penopang utama penjualan. Perusahaan sebaiknya mempertahankan kualitas dan ketersediaan produk tersebut, serta melakukan inovasi pada lini produk serupa untuk menarik pelanggan baru dan menjaga loyalitas pelanggan lama.
4. Melakukan evaluasi berkala terhadap tren pasar. Analisis data penjualan dan profit sebaiknya dilakukan secara rutin untuk memantau perubahan tren dan kebutuhan pelanggan. Dengan demikian, perusahaan dapat mengambil langkah strategis lebih cepat dalam merespons perubahan pasar.

## V. Lampiran

### A. Potongan Kode Program (Python)

```
#Untuk menampilkan ukuran data
#Jumlah baris
print('jumlah sampel =', super.shape[0])
#Jumlah kolom
print('Jumlah variabel (kolom)= ', super.shape[1])

Jumlah sampel = 9994
Jumlah variabel (kolom)=  21

#Untuk melihat nama kolom
nama_kolom = super.columns
nama_kolom

Index(['Row ID', 'Order ID', 'Order Date', 'Ship Date', 'Ship Mode',
       'Customer ID', 'Customer Name', 'Segment', 'Country', 'City', 'State',
       'Postal Code', 'Region', 'Product ID', 'Category', 'Sub-Category',
       'Product Name', 'Sales', 'Quantity', 'Discount', 'Profit'],
      dtype='object')
```

Kode di atas untuk mengetahui informasi simple dari dataset atau data yang akan di analisis, seperti ingin mengetahui berapa total jumlah sampel dan variabel.

```
# Plot

var = 'Profit'
plt.figure(figsize=(10,5))
yearly.plot(marker='s', color='olive', linewidth=2, legend=True)

plt.xlabel('Tahun')
plt.ylabel(var)
plt.title(f'Tren Tahunan {var}')

plt.xticks(yearly.index, rotation=40)
plt.grid(True)

plt.show()
```

Kode di atas adalah untuk membuat visualisasi data dengan menggunakan grafik pada variabel Profit.

```
# Plot

plt.figure(figsize=(10,5))
yearly_sales.plot(marker='s', color='blue', linewidth=2, legend=True)

plt.xlabel('Tahun')
plt.ylabel(var)
plt.title(f'Tren Tahunan {var}')

plt.xticks(yearly_sales.index, rotation=40)
plt.grid(True)

plt.show()
```

Kode di atas adalah untuk membuat visualisasi data dengan menggunakan grafik pada variabel Sales.

```
# Membuat grafik batang untuk 10 produk teratas
plt.figure(figsize=(10,6))
top10_products.plot(kind='bar', color='skyblue')

# Memberikan judul dan label sumbu
plt.title('10 Produk yang Paling Sering Dibeli oleh Customer (2014–2017)', fontsize=14)
plt.xlabel('Nama Produk', fontsize=12)
plt.ylabel('Jumlah Pembelian', fontsize=12)
plt.xticks(rotation=45, ha='right') # Memiringkan label agar tidak tumpang tindih
plt.grid(axis='y', linestyle='--', alpha=0.7)

# Menampilkan grafik
plt.tight_layout()
plt.show()
```

Kode di atas adalah untuk membuat visualisasi data grafik pada variabel Product Name dan menampilkan 10 produk yang paling banyak dibeli.