

LAPORAN HASIL DATA ANALYST

Nama Perusahaan : SuperStore
Nama Data Analyst : Tazkiyatul Zahra
NPM : 1121436
Jenjang/Program Studi : S1/Sistem Informasi
Tanggal Laporan : 14 Juni 2025

Laporan Hasil Data Analyst ini Dibuat

DAFTAR ISI

DAFTAR ISI.....	2
1. PENDAHULUAN	3
1.1 Latar Belakang	3
1.2 Tujuan Analisis.....	4
2. IDENTIFIKASI KEBUTUHAN BISNIS	4
3. PENGUMPULAN & PERSIAPAN DATA.....	4
3.1 Sumber Data.....	4
3.2 Proses Praproses Data	5
3.2.1 Pemeriksaan dan penghapusan data kosong	6
3.2.2 Membuat Copy Dataset.....	6
3.2.1 Cek Outlier.....	6
3.2.2 Deteksi dan penghapusan data duplikat	8
4. ANALISIS DATA	8
4.1 Eksplorasi Variabel Utama.....	8
4.2 Visualisasi Data	9
4.2.1 Persentase Persentase Profit berdasarkan Order_Date(Year)	9
4.2.2 Jumlah Sales Berdasarkan Region	10
4.2.3 Profit berdasarkan Category dan Segmen	12
4.2.4 Perbandingan Sales dan Profit berdasarkan Kategori	13
4.2.5 Top 5 Produk yang Salesnya Terbanyak.....	14
5. TEMUAN & <i>INSIGHT</i>	16
6. PELUANG BISNIS	16
7. KESIMPULAN DAN REKOMENDASI	17

1. PENDAHULUAN

1.1 Latar Belakang

Superstore adalah toko buatan di AS yang menjual perabot rumah, bahan kerja, dan *gadget*. Dalam era digital dan globalisasi saat ini, industri ritel menghadapi tantangan yang semakin kompleks, mulai dari perubahan perilaku konsumen yang cepat, persaingan yang semakin ketat, hingga kebutuhan untuk mengelola rantai pasok dan inventaris secara efisien.

Untuk tetap kompetitif dan relevan di pasar, Superstore harus mampu mengadopsi strategi bisnis yang berbasis data dan teknologi modern. Salah satu pendekatan yang sangat krusial adalah penerapan *Business Intelligence* (BI), yang memungkinkan perusahaan untuk mengumpulkan, mengintegrasikan, dan menganalisis data dari berbagai sumber secara komprehensif dan *real-time*. Dengan *Business Intelligence*, Superstore dapat memperoleh wawasan mendalam mengenai performa penjualan, tren pasar, preferensi pelanggan, serta efektivitas berbagai program pemasaran dan distribusi.

Di pasar ritel kompetitif yang berkembang, membuat keputusan berdasarkan data sangat penting. Oleh karena itu, janji superstore untuk menggunakan pemeriksaan data untuk memahami pola penjualan sesuai dengan bidang dan jenis produk, memastikan pendekatan iklan yang sesuai, mengawasi inventaris lebih baik, dan meningkatkan kepuasan dan kesetiaan klien. Dengan metode ini, perusahaan dalam mengoptimalkan manajemen persediaan dengan memprediksi permintaan

secara lebih akurat, sehingga dapat menghindari masalah overstocking maupun kekurangan stok yang dapat berdampak negatif pada kepuasan pelanggan dan biaya operasional.

1.2 Tujuan Analisis

Tujuan analisis dataset Superstore adalah:

- Mengetahui persentase profit superstore pertahun
- Mengetahui total sales berdasarkan region
- Mengetahui profit superstore berdasarkan segmen dan kategori
- Mengetahui profit vs sale berdasarkan kategori
- Mengetahui 5 produk yang laku terjual

2. IDENTIFIKASI KEBUTUHAN BISNIS

Identifikasi kebutuhan bisnis berfokus menentukan apa yang dibutuhkan oleh bisnis. Mengidentifikasi kebutuhan bisnis bertujuan untuk menjabarkan kebutuhan sistem, informasi, dan solusi untuk menyelesaikan masalah bisnis. Identifikasi kebutuhan bisnis Superstore dapat diuraikan:

1. Jumlah profit penjualan berdasarkan tahun dan produk
2. Distribusi penjualan berdasarkan wilayah (Region)

3. PENGUMPULAN & PERSIAPAN DATA

3.1 Sumber Data

Sumber *dataset* Superstore diperoleh dari hasil unduhan *link* <https://github.com/WuCandice/Superstore-Sales-Analysis/blob/main/dataset/Superstore%20Dataset.csv> dengan format file csv. *Tools* untuk menganalisis dan mempersiapkan data menggunakan Bahasa pemrograman Python (pandas, seaborn) di Google Colaboratory.

3.2 Proses Praproses Data

Preprocessing data merupakan tahapan awal pada analisis data yang bertujuan membersihkan, mengubah, dan mempersiapkan data mentah agar dapat dianalisis lebih lanjut. Tahapan *preprocessing* dapat meliputi penghapusan data kosong, membuat *copy dataset*, cek *outlier*, standarisasi format kolom, deeksi dan penghapusan data duplikat, dan format tanggal dan konversi tipe data.

Hal pertama yang dilakukan adalah menginstall dan mengimport library yang dibutuhkan pada preprocessing data.

```
pip install numpy
```

Gambar 3.2.1 *Install Library Numpy*

```
pip install pandas
```

Gambar 3.2.2 *Install Library Pandas*

```
pip install seaborn
```

Gambar 3.2.3 *Install Library Seaborn*

Setelah dilakukan *install library*, selanjutnya mengimpor *library* yang telah di install dan membaca *dataset* yang akan digunakan untuk analisis.

```
%%time
import numpy as np
import pandas as pd
import seaborn as sns
from matplotlib import pyplot as plt
sns.set()
%matplotlib inline
```

Gambar 3.2.4 *Impor Library*

```
fordGB = pd.read_csv("201902-fordgobike-tripdata.csv")
fordGB.head()
```

Gambar 3.2.5 *Membaca Dataset*

Dataset yang dibaca merupakan dataset yang telah di *upload* ke Google Colaboratory dari File Explorer tempat dimana menyimpan *dataset* di penyimpanan lokal. Dataset Superstore Dataset.csv terdapat 20 kolom dan 9994 baris. Kolom

dataset mencakup 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, dan Profit.

3.2.1 Pemeriksaan dan penghapusan data kosong

Langkah yang pertama pada preprocessing data adalah mengidentifikasi apakah terdapat nilai kosong pada dataset. Data kosong atau missing value dapat mempengaruhi hasil analisis karena tidak ditangani dengan baik.

```
#jumlah missing value  
store.isnull().sum()
```

Gambar 3.2.1.1 Mengidentifikasi *Missing Value*

Pada gambar 3.2.1.1 pemeriksaan missing value dilakukan menggunakan fungsi `store.isnull().sum()`. Setelah dilakukan pemeriksaan *missing value* tidak terdapat *output* yang menunjukkan banyaknya *missing value* pada setiap kolom.

Selanjutnya dilakukan visualisasi apakah data masih terdapat *missing value* menggunakan fungsi `sns.heatmap(store.isnull())` yang menampilkan *heatmap missing value*.

3.2.2 Membuat Copy Dataset

Sebelum melakukan transformasi pada data, dilakukan *copy dataset* untuk menjaga data mentah agar tidak rusak atau berubah karena manipulasi data selama proses *preprocessing*. Dengan membuat salinan data dapat kembali ke versi awal apabila terjadi kesalahan pada tahapan selanjutnya. Untuk membuat *copy dataset* menggunakan fungsi `store _copy = store.copy()`.

3.2.1 Cek Outlier

Outlier merupakan nilai data yang jauh berbeda dari Sebagian nilai data lainnya. *outlier* dapat mempengaruhi analisis statistik, seperti *mean* dan regresi data yang dapat menyebabkan bias pada hasil analisis. Untuk mengidentifikasi pertama

dapat dilihat dari nilai min dan max setiap kolom, jika jauh dari nilai mean maka data dapat disebut outlier.

```
store.describe(include="all")
```

Gambar 3.2.3.1 Membandingkan Nilai Min dan Max

Pada gambar 3.2.3.1 merupakan fungsi yang di run untuk melihat perbandingan nilai max dan min dengan nilai mean. Terdapat 2 kolom yang memiliki outlier, yaitu Sales dan Profit. Setelah dilakukan kepastian bahwa kolom yang dianggap outlier memiliki outlier menggunakan boxplot dengan fungsi `plt.figure(figsize = (10,10))` dan `plt.boxplot(store_copy["duration_sec"])`.

Untuk menghapus outlier dapat menggunakan metode statistik *Interquartile Range* (IQR) dan boxplot dengan fungsi persamaan. Jika ditemukan outlier lebih tinggi dari Q3 akan dihapus menggunakan 2 persamaan.

```
# persamaan 1
Q1 = store_copy["Sales"].quantile(0.25)
Q3 = store_copy["Sales"].quantile(0.75)
IQR = Q3 - Q1

# persamaan 2
outlierQ3 = Q3 + 1.5 * IQR

# membuang data outlier
store_copy = store_copy[(store_copy["Sales"] < outlierQ3)]
store_copy["Sales"].describe()
```

Gambar 3.2.3.2 Persamaan IQR

Pada gambar 3.2.3.2 nilai kuartil pertama (Q1) dan kuartil ketiga (Q3) dihitung dan nilai IQR diperoleh dengan mengurangi Q1 dari Q3, yang memberikan rentang data tengah (middle 50%) yang dianggap sebagai data normal. Setelah proses penghapusan selesai, dilakukan pemeriksaan statistik deskriptif terhadap kolom `duration_sec` menggunakan fungsi `describe()`.

3.2.2 Deteksi dan penghapusan data duplikat

Duplikat data dapat terjadi karena input ganda, sehingga untuk mengidentifikasi menggunakan fungsi duplicated(). Untuk mengetahui bahwa data duplikat adalah data yang bernilai True dan perlu dihapus. Untuk menghapus data duplikat dapat dilakukan menggunakan fungsi drop_duplicates().

```
print(f"jumlah kolom data copy : {store_copy.shape[1]}")  
print(f"jumlah Baris data copy : {store_copy.shape[0]}")
```

Gambar 3.2.4.1 Menghapus Duplikat

Pada gambar 3.2.4.1 setelah dilakukan penghapusan duplikat, dilakukan pengecekan kembali jumlah baris dan kolom untuk melihat jumlah baris apakah berkurang atau tidak.

4. ANALISIS DATA

4.1 Eksplorasi Variabel Utama

Eksplorasi variabel utama merupakan proses awal pada analisis data yang bertujuan untuk memahami karakteristik dasar dari variabel – variabel penting dalam dataset. Tujuannya untuk mengenali jenis data, struktur data, dan nilai yang menonjol atau outlier. Variable utama pada dataset Superstore adalah:

1. Waktu (Order Date, Ship Date)
2. Penjualan dan Keuntungan (Sales, Profit, Discount, Quantity)
3. Pelanggan & Segmentasi (Customer ID, Segment)
4. Produk (Category, Sub-Category, Product Name)
5. Geografis (Region, State, City)

4.2 Visualisasi Data

4.2.1 Persentase Profit berdasarkan Order_Date (Year)

Visualisasi persentase profit berdasarkan tahun (Order Date (Year)) dibuat untuk memahami sejauh mana efisiensi keuntungan yang diperoleh Superstore dari total penjualannya setiap tahun.

```
#mengubah typedata order_date menjadi datetime
store_copy['Order Date'] = pd.to_datetime(store_copy['Order Date'])

#ekstrak tahun dari order_date
store_copy['Order Year'] = store_copy['Order Date'].dt.year
```

Gambar 4.2.1.1 Persentase Penyewa

Pada gambar 4.2.1 dilakukan mengubah format data pada kolom Order Date menjadi datetime

```
#menghitung total profit pertahun
profit_per_year = store_copy.groupby('Order Year')['Profit'].sum().reset_index()

#profit keseluruhan
total_profit = profit_per_year['Profit'].sum()

#persentase profit pertahun
profit_per_year['Profit %'] = (profit_per_year['Profit'] / total_profit) * 100
```

Gambar 4.2.1.2 Menghitung Total Profit

Gambar 4.2.2 menghitung total profit superstore pertahun secara keseluruhan. Setelah itu dibuat persentase untuk memudahkan membaca data.

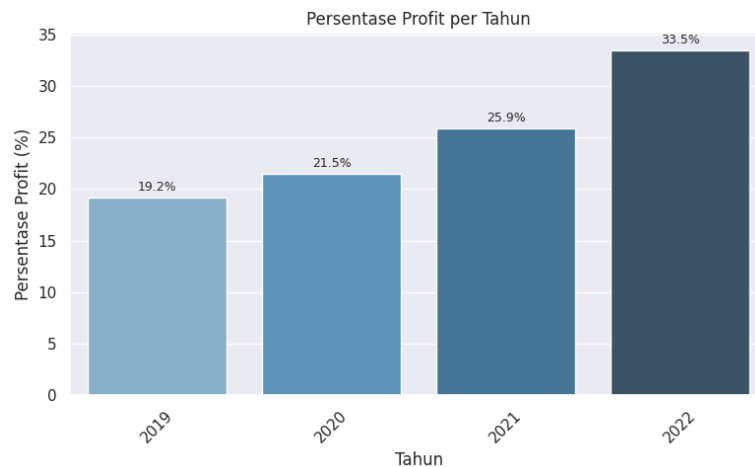
```
#Visualisasi
plt.figure(figsize=(8, 5))
ax = sns.barplot(data=profit_per_year, x='Order Year', y='Profit %', palette='Blues_d')

for container in ax.containers:
    ax.bar_label(container,
                  labels=[f'{v:.1f}%' for v in container.datavalues],
                  label_type='edge',
                  fontsize=9,
                  padding=3)

plt.title('Persentase Profit per Tahun')
plt.ylabel('Persentase Profit (%)')
plt.xlabel('Tahun')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

Gambar 4.2.1.3 Membuat Visualisasi Data

Gambar 4.2.1.3 membuat visualisasi data dari persentase profit dari tahun 2019 – 2022 menggunakan bar chart.



Gambar 4.2.1.4 Visualisasi Persentase Profit per Tahun

Gambar 4.2.1.4 menunjukkan tren peningkatan persentase profit Superstore dari tahun 2019 hingga 2022. Kenaikan ini mencerminkan efisiensi bisnis yang semakin baik dalam mengelola penjualan dan biaya, serta kemungkinan penerapan strategi yang lebih efektif, seperti pengurangan diskon yang tidak perlu atau fokus pada produk dengan margin keuntungan lebih tinggi.

4.2.2 Jumlah Sales Berdasarkan Region

Visualisasi Sales berdasarkan Region dibuat untuk mengetahui seberapa besar kontribusi penjualan dari tiap wilayah. Hal ini penting untuk mengidentifikasi wilayah dengan performa terbaik maupun terendah, sehingga perusahaan dapat mengatur strategi pemasaran, distribusi, dan alokasi sumber daya secara lebih tepat sasaran.

```
#menghitung total sales per region
region_sales = store_copy.groupby('Region')['Sales'].sum().reset_index()

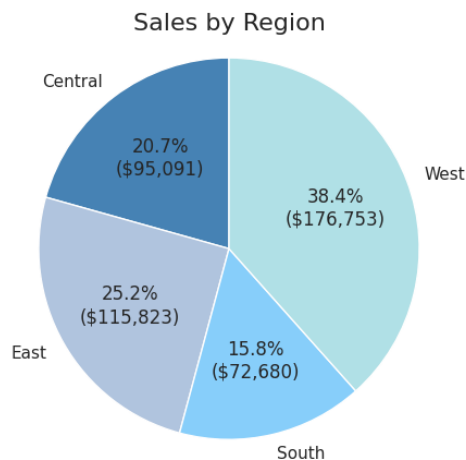
# Pie chart
plt.figure(figsize=(8,5))

# Pie plot
plt.pie(
    region_sales['Sales'],
    labels=region_sales['Region'],
    autopct=lambda p: f'{p:.1f}%\n({p * region_sales["Sales"].sum() / 100:,.0f})',
    startangle=90,
    colors=['steelblue', 'lightsteelblue', 'lightskyblue', 'powderblue']
)

plt.title('Sales by Region', fontsize=16)
plt.axis('equal') # Agar pie berbentuk lingkaran
plt.show()
```

Gambar 4.2.2.1 Menghitung Jumlah Sales dan Memvisualisasikan

Pada gambar 4.2.2.1 data dikelompokkan berdasarkan Region dan dihitung total penjualannya. *Pie chart* ditampilkan dengan proporsi persentase dan nilai dolar penjualan masing-masing wilayah. Warna yang digunakan bervariasi untuk membedakan tiap region, dan `startangle=90` digunakan agar tampilan pie dimulai dari posisi atas.



Gambar 4.2.2.2 Visualisasi *Sales by Region*

Gambar 4.2.2.2 menunjukkan bahwa wilayah West memberikan kontribusi penjualan terbesar sebesar 38,4% atau sekitar \$176.753, diikuti oleh East dengan 25,2%. Wilayah Central menyumbang 20,7%, sedangkan South memiliki kontribusi terkecil yaitu 15,8% atau sekitar \$72.680. Visualisasi ini memberikan gambaran yang jelas mengenai performa penjualan antar wilayah.

4.2.3 Profit berdasarkan Category dan Segmen

Visualisasi profit berdasarkan Category dan Segment dibuat untuk memahami kontribusi keuntungan dari setiap kategori produk (seperti Furniture, Technology, dan Office Supplies) di masing-masing segmen pelanggan (Consumer, Corporate, dan Home Office).

```
# Group data
grouped = store_copy.groupby(['Segment', 'Category'])['Profit'].sum().reset_index()

# Colors
colors = sns.color_palette('Blues', n_colors=3)

# Pivot data agar mudah diplot
pivot_store = grouped.pivot(index='Segment', columns='Category', values='Profit')

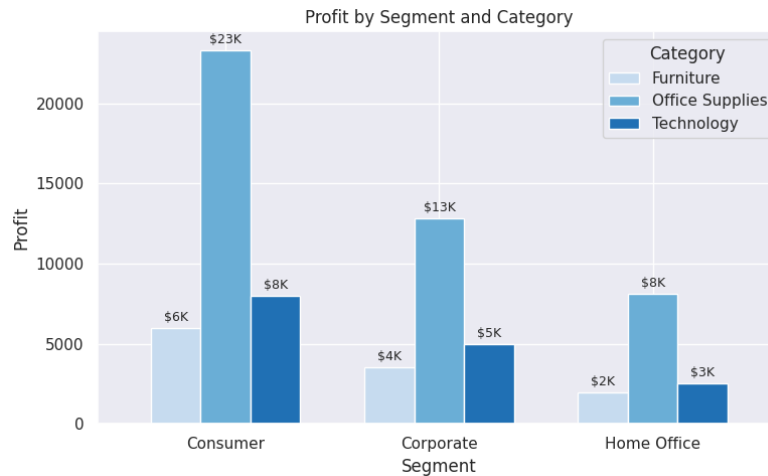
# Plot grouped bar chart
ax = pivot_store.plot(kind='bar', figsize=(8,5), width=0.7, color=colors)

# Tambahkan angka di atas bar
for container in ax.containers:
    ax.bar_label(container,
                  labels=[f'{v/1e3:.0f}K' for v in container.datavalues], # Format ribuan (K)
                  label_type='edge',
                  fontsize=9,
                  padding=3)

plt.title('Profit by Segment and Category')
plt.ylabel('Profit')
plt.xlabel('Segment')
plt.xticks(rotation=0)
plt.legend(title='Category')
plt.tight_layout()
plt.show()
```

Gambar 4.2.3.1 Menghitung Total Profit

Gambar 4.2.3.1 data dikelompokkan berdasarkan Segment dan Category, lalu dihitung total profit-nya. Data kemudian diubah menjadi format pivot agar memudahkan plotting. Bar chart yang dihasilkan memperlihatkan perbandingan profit tiap kategori produk dalam masing-masing segmen pelanggan.



Gambar 4.2.3.2 Visualisasi Profit Berdasarkan Segment dan Category

Gambar 4.2.3.2 menunjukkan bahwa segmen Consumer menyumbang profit tertinggi, khususnya dari kategori Office Supplies sebesar \$23K. Sementara itu, segmen Corporate dan Home Office menunjukkan profit yang lebih rendah secara keseluruhan, dengan kontribusi utama tetap berasal dari kategori Office Supplies. Kategori Furniture mencatat profit paling kecil di semua segmen.

4.2.4 Perbandingan Sales dan Profit berdasarkan Kategori

Visualisasi perbandingan Sales dan Profit berdasarkan Kategori dibuat untuk mengetahui sejauh mana penjualan (sales) pada setiap kategori produk mampu menghasilkan keuntungan (profit). Meskipun suatu kategori memiliki penjualan tinggi, belum tentu memberikan profit yang besar.

```
#Grouping data
data = store_copy.groupby('Category')[['Sales', 'Profit']].sum()

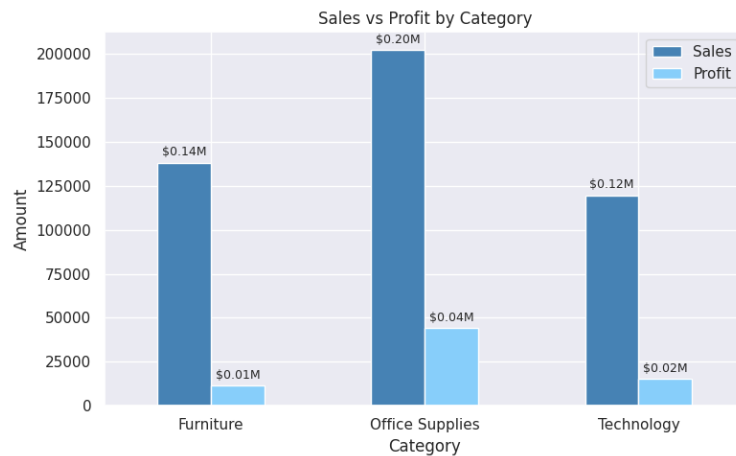
#Plot
ax = data.plot(kind='bar', figsize=(8,5), color=['steelblue', 'lightskyblue'])
plt.title('Sales vs Profit by Category')
plt.ylabel('Amount')
plt.xticks(rotation=0)
plt.tight_layout()

#Tambah label angka di atas bar
for container in ax.containers:
    ax.bar_label(container,
        labels=[f'${v/1e6:.2f}M' for v in container.datavalues],
        label_type='edge', fontsize=9, padding=3)

plt.show()
```

Gambar 4.2.4.1 Menghitung Jumlah Sales dan Profit

Gambar 4.2.4.1 data dikelompokkan berdasarkan Category, lalu dijumlahkan untuk kolom Sales dan Profit. Hasilnya diplot sebagai bar chart dengan dua batang untuk setiap kategori—satu untuk sales dan satu untuk profit—menggunakan warna berbeda (steelblue untuk Sales dan lightskyblue untuk Profit).



Gambar 4.2.4.2 Visualisasi Sales vs Profit Berdasarkan Category

Gambar 4.2.4.2 menunjukkan bahwa kategori Office Supplies memiliki nilai penjualan tertinggi, diikuti oleh Furniture dan Technology. Namun, profit Office Supplies juga memberikan keuntungan terbesar, sementara Furniture meskipun memiliki penjualan tinggi, hanya menghasilkan profit yang sangat kecil. Hal ini mengindikasikan bahwa tidak semua kategori yang memiliki penjualan tinggi otomatis memberikan profit yang tinggi.

4.2.5 Top 5 Produk yang Salesnya Terbanyak

Visualisasi Top 5 Produk dengan Sales Terbanyak dibuat untuk mengidentifikasi produk-produk yang memberikan kontribusi penjualan tertinggi terhadap total pendapatan perusahaan.

```

top5_quantity = store_copy.groupby('Product Name')['Quantity'].sum().sort_values(ascending=False).head(5)

# Plot
plt.figure(figsize=(8, 5))
bars = plt.barh(top5_quantity.index, top5_quantity.values, color='steelblue')

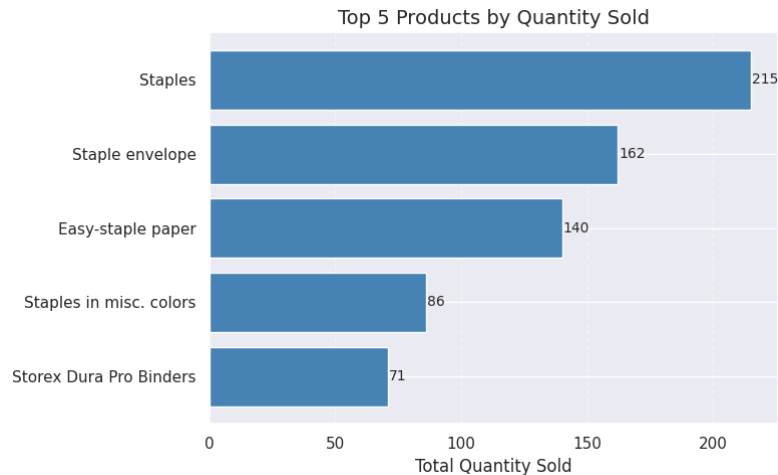
# Tambahkan label jumlah di ujung bar
for bar in bars:
    plt.text(bar.get_width() + 0.5,          # posisi X (kanan bar)
             bar.get_y() + bar.get_height()/2, # posisi Y (tengah bar)
             f"{int(bar.get_width())}",      # angka quantity
             va='center', fontsize=10)

# Format chart
plt.xlabel('Total Quantity Sold', fontsize=12)
plt.title('Top 5 Products by Quantity Sold', fontsize=14)
plt.gca().invert_yaxis() # supaya urutan dari paling besar ke kecil
plt.grid(axis='x', linestyle='--', alpha=0.5)
plt.tight_layout()
plt.show()

```

Gambar 4.2.5.1 Menghitung Top 5 Produk

Gambar 4.2.5.1 mengelompokkan data berdasarkan nama produk dan menghitung total jumlah (quantity) terjual untuk masing-masing produk, lalu mengurutkannya secara menurun dan mengambil 5 produk teratas. Setelah itu, dibuat grafik batang horizontal menggunakan matplotlib untuk menampilkan lima produk tersebut beserta jumlah penjualan masing-masing.



Gambar 4.2.5.2 Visualisasi Top 5 Produk

Gambar 4.2.5.2 Visualisasi "Top 5 Products by Quantity Sold" di atas menampilkan lima produk dengan jumlah penjualan terbanyak berdasarkan kuantitas dari keseluruhan data penjualan. Grafik batang horizontal menunjukkan bahwa produk Staples memiliki jumlah penjualan tertinggi (215 unit), diikuti oleh

Staple envelope, Easy-staple paper, Staples in misc. colors, dan Storex Dura Pro Binders.

5. TEMUAN & *INSIGHT*

Dari hasil visualisasi terlihat bahwa keuntungan perusahaan setiap tahunnya bervariasi, yang menunjukkan adanya ketidakstabilan profit yang bisa disebabkan oleh kondisi pasar atau faktor internal perusahaan. Dilihat dari wilayah penjualan, wilayah West menjadi penyumbang penjualan tertinggi, disusul oleh East, Central, dan South. Artinya, wilayah West memiliki potensi pasar yang kuat, sedangkan South masih memiliki ruang untuk ditingkatkan. Selain itu, produk dalam kategori Technology tampak memberikan kontribusi profit yang cukup besar meskipun tidak dijual sebanyak Office Supplies atau Furniture. Dari sisi pelanggan, penjualan terbagi dalam segmen Consumer, Corporate, dan Home Office, dengan peluang untuk mengembangkan strategi penjualan yang disesuaikan dengan kebutuhan tiap segmen.

6. PELUANG BISNIS

Dari hasil analisis, terdapat beberapa peluang bisnis yang bisa dimanfaatkan. Wilayah South yang penjualannya masih rendah bisa menjadi target ekspansi dengan strategi promosi yang lebih gencar. Produk Technology juga memiliki peluang besar karena keuntungannya tinggi, sehingga bisa dikembangkan lebih lanjut atau dipaketkan dengan produk lain. Selain itu, segmen Home Office meskipun kecil, punya potensi pasar yang bagus karena tren kerja dari rumah yang terus meningkat.

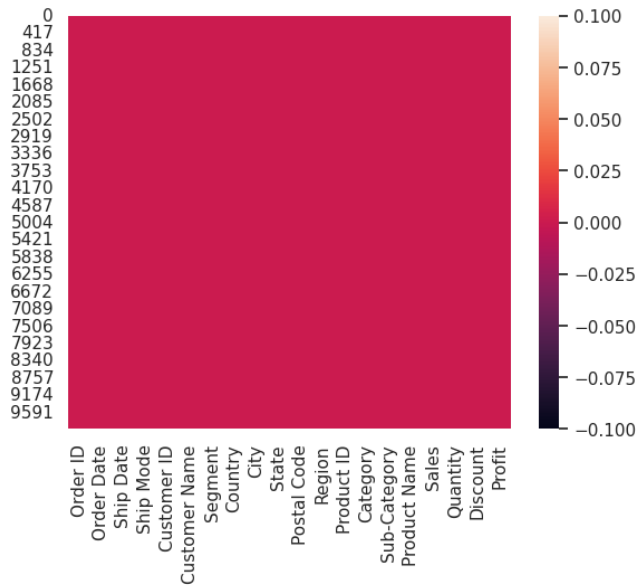
7. KESIMPULAN DAN REKOMENDASI

Data Superstore menunjukkan pentingnya pembersihan data untuk mendapatkan insight yang akurat. Dari hasil analisis, perusahaan disarankan untuk:

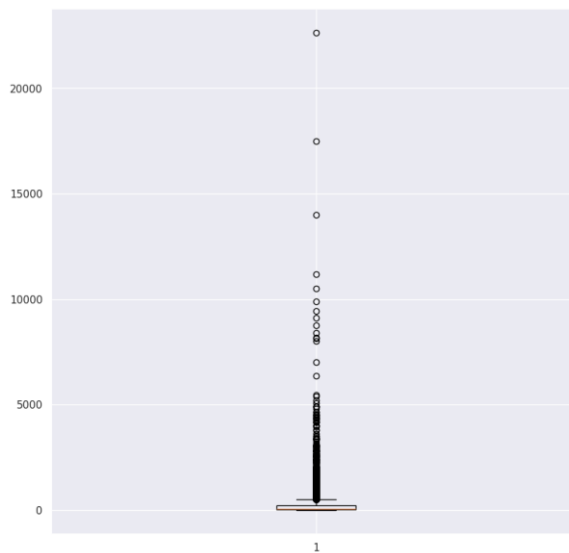
1. Fokus pada pengembangan produk dan promosi di wilayah yang belum optimal seperti South.
2. Meningkatkan lini produk yang menguntungkan, khususnya pada kategori Technology yang memberikan margin profit tinggi.
3. Melakukan kampanye yang disesuaikan dengan segmen pelanggan, termasuk memberi perhatian lebih pada segmen Home Office yang berpotensi tumbuh.

LAMPIRAN-LAMPIRAN

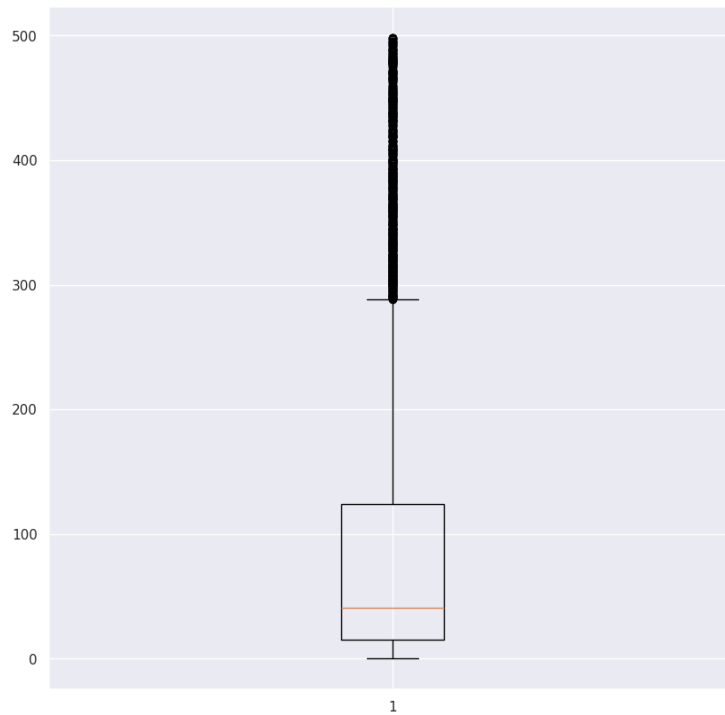
Heatmap Missing Value



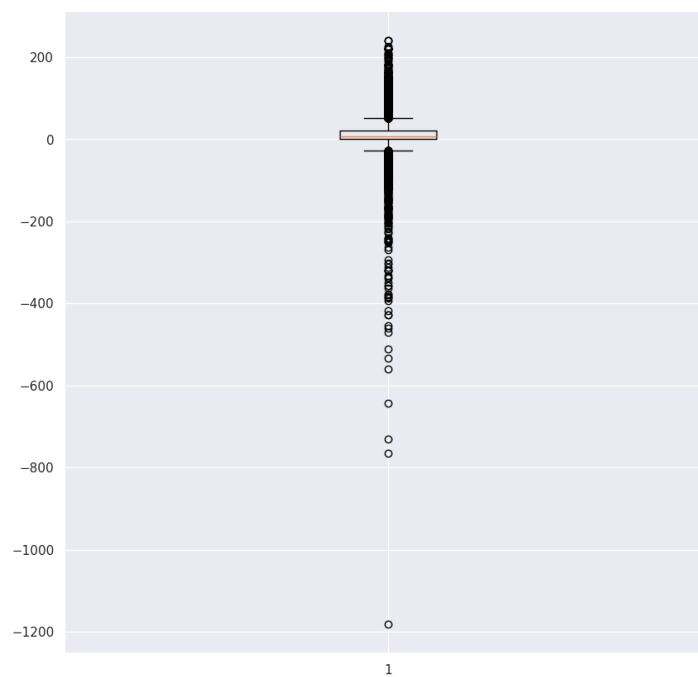
Boxplot (Sales) Melihat *Outlier*



BoxPlot (Sales) Setelah *Outlier* Dihapus



Boxplot (Profit) Melihat *Outlier*



BoxPlot (Profit) Setelah *Outlier* Dihapus

