

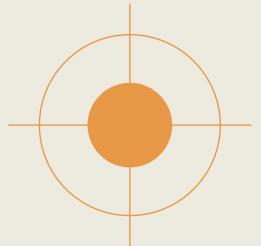


ANALISIS
PENJUALAN
RITEL &
Segmentasi
PELANGGAN





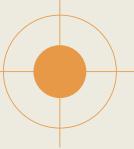
Retail Sales Dataset merupakan data penjualan ritel dan demografi penjualan. Dataset ini mencakup detail penting seperti ID Transaksi, Tanggal, ID Pelanggan, Jenis Kelamin, Usia, Kategori Produk, Kuantitas, Harga per Unit, dan Jumlah Total. Project ini melakukan Exploratory Data Analysis (EDA) dan Analisis RFM (Recency, Frequency, Monetary) untuk memperoleh wawasan, meningkatkan segmentasi pelanggan, dan menjawab pertanyaan bisnis yang telah diperoleh dengan proses analisis.



PENJELASAN DATASET

[Ritel_sales_dataset](#)





ATTRIBUT



- Transaction ID: Kode unik untuk setiap transaksi, digunakan untuk pelacakan dan referensi
- Date: Tanggal transaksi, memberikan wawasan dari waktu ke waktu
- Customer ID: Kode unik untuk setiap pelanggan
- Gender: Jenis kelamin pelanggan (Male/Female), memberikan wawasan tentang pola pembelian berdasarkan Jenis kelamin
- Age: Usia pelanggan
- Product Category: Kategori produk yang dibeli pelanggan, untuk membantu memahami preferensi produk
- Quantity: Jumlah unit produk yang dibeli pelanggan
- Price per Unit: Harga per satuan unit produk
- Total Amount: Nilai moneter total dari transaksi, yang menunjukkan dampak keuangan dari setiap pembelian



EDA

(Exploratory Data Analysis)

01

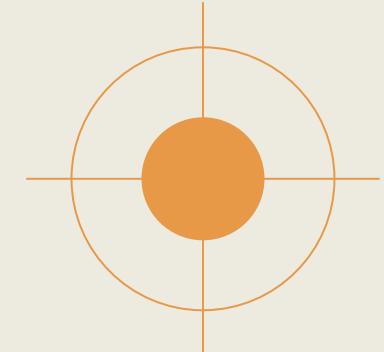
Mengeksplorasi kinerja penjualan produk dari waktu ke waktu (misalnya, tren pendapatan bulanan, efek musiman)

02

Menganalisis demografi pelanggan (misalnya, distribusi jenis kelamin, pola berdasarkan usia).

03

Mengidentifikasi kategori produk dengan penjualan terbanyak dan dampaknya terhadap pendapatan.





RFM



(Recency, Frequency, Monetary)



01

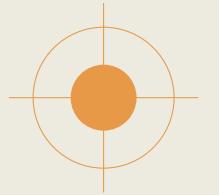
Penilaian RFM digunakan untuk mengelompokkan pelanggan ke dalam segmen yang dapat ditindaklanjuti (Champions, Loyal Customers, At-Risk, dan Hibernating).

02

Memahami karakteristik pelanggan yang menunjukkan penurunan keterlibatan dan bagaimana strategi yang dapat melibatkan mereka kembali.

03

Memberikan rekomendasi untuk upaya pemasaran yang dipersonalisasi berdasarkan segmentasi pelanggan.



LIBRARY



```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
from IPython.display import display
```





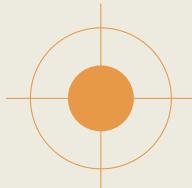
DETAIL DATASET



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 9 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Transaction ID  1000 non-null    int64  
 1   Date              1000 non-null    object  
 2   Customer ID      1000 non-null    object  
 3   Gender            1000 non-null    object  
 4   Age               1000 non-null    int64  
 5   Product Category 1000 non-null    object  
 6   Quantity          1000 non-null    int64  
 7   Price per Unit   1000 non-null    int64  
 8   Total Amount      1000 non-null    int64  
dtypes: int64(5), object(4)
memory usage: 70.4+ KB
```

	Transaction ID	Age	Quantity	Price per Unit	Total Amount
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	500.500000	41.39200	2.514000	179.890000	456.000000
std	288.819436	13.68143	1.132734	189.681356	559.997632
min	1.000000	18.00000	1.000000	25.000000	25.000000
25%	250.750000	29.00000	1.000000	30.000000	60.000000
50%	500.500000	42.00000	3.000000	50.000000	135.000000
75%	750.250000	53.00000	4.000000	300.000000	900.000000
max	1000.000000	64.00000	4.000000	500.000000	2000.000000

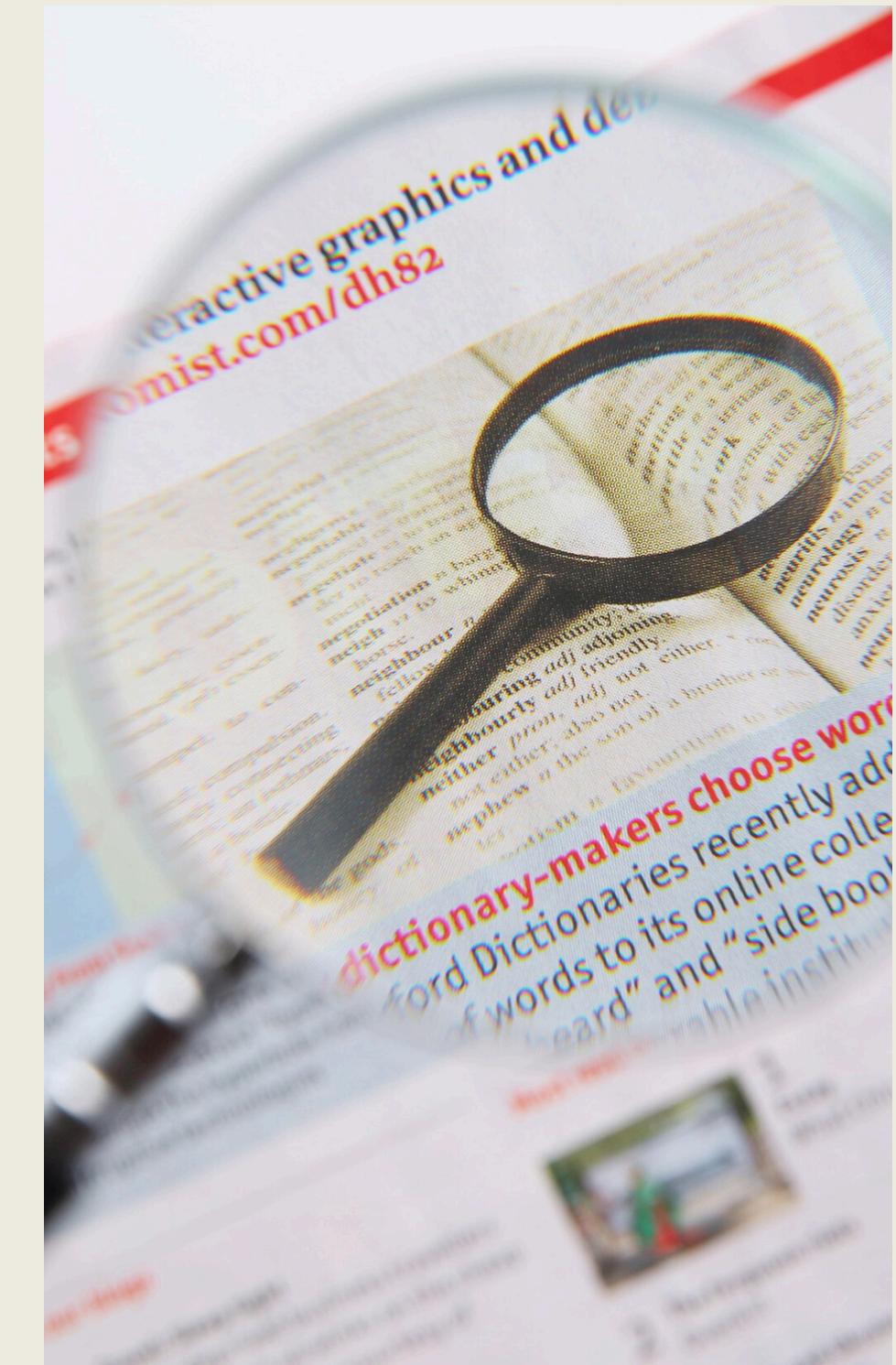
Penjelasan



Dataset terdiri dari 1000 baris dan 9 kolom, dengan kolom bertipe numerik yaitu Transaction ID, Age, Quantity, Price per Unit, dan Total Amount. Sedangkan kolom bertipe kategorik yaitu Date, Customer ID, Gender, dan Product Category.

Berdasarkan deskripsi data:

- Usia rata-rata pelanggan adalah 41,3 tahun.
 - Rata-rata jumlah produk yang dibeli per transaksi adalah sekitar 2–3 item.
 - Usia pelanggan berkisar antara 18 hingga 64 tahun.
 - Jumlah produk yang dibeli per transaksi berkisar antara 1 hingga 4 item.
 - Harga per unit berkisar antara 25 hingga 500, dengan harga rata-rata sebesar 179,89.
 - Total nilai transaksi berkisar antara 25 hingga 2000, dengan rata-rata 456.

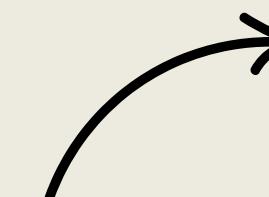




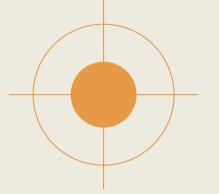
MENAMBAHKAN BEBERAPA KOLOM

```
# Mengubah kolom `Date` menjadi tipe Datetime
df['Date'] = pd.to_datetime(df['Date'])
# Memisahkan Kolom Tanggal dan menambah kolom nama hari dan nama bulan
df['Days'] = df['Date'].dt.day_name()
df['MonthDate'] = df['Date'].dt.day
df['Month'] = df['Date'].dt.month
df['Month'] = df['Date'].dt.month_name()
df['Year'] = df['Date'].dt.year
```

```
# Mengelompokkan usia berdasarkan kategori usia (muda, tua, senior)
df['Age Group'] = pd.cut(df['Age'], bins=[0, 30, 50, 100], labels=['Muda', 'Dewasa', 'Lansia'])
# Menghitung keuntungan penjualan per unit produk
df['Revenue per Item'] = df['Total Amount'] / df['Quantity']
df.head()
```



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 15 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Transaction ID  1000 non-null   int64  
 1   Date             1000 non-null   datetime64[ns]
 2   Customer ID    1000 non-null   object  
 3   Gender           1000 non-null   object  
 4   Age              1000 non-null   int64  
 5   Product Category 1000 non-null   object  
 6   Quantity         1000 non-null   int64  
 7   Price per Unit  1000 non-null   int64  
 8   Total Amount    1000 non-null   int64  
 9   Days             1000 non-null   object  
 10  MonthDate       1000 non-null   int32  
 11  Month            1000 non-null   object  
 12  Year             1000 non-null   int32  
 13  Age Group        1000 non-null   category
 14  Revenue per Item 1000 non-null   float64 
dtypes: category(1), datetime64[ns](1), float64(1), int32(2), int64(5), object(5)
memory usage: 102.8+ KB
```



DATA PREPOCESSING

1

Jumlah Nilai yang hilang	
Transaction ID	0
Date	0
Customer ID	0
Gender	0
Age	0
Product Category	0
Quantity	0
Price per Unit	0
Total Amount	0
Days	0
MonthDate	0
Month	0
Year	0
Age Group	0
Revenue per Item	0

2

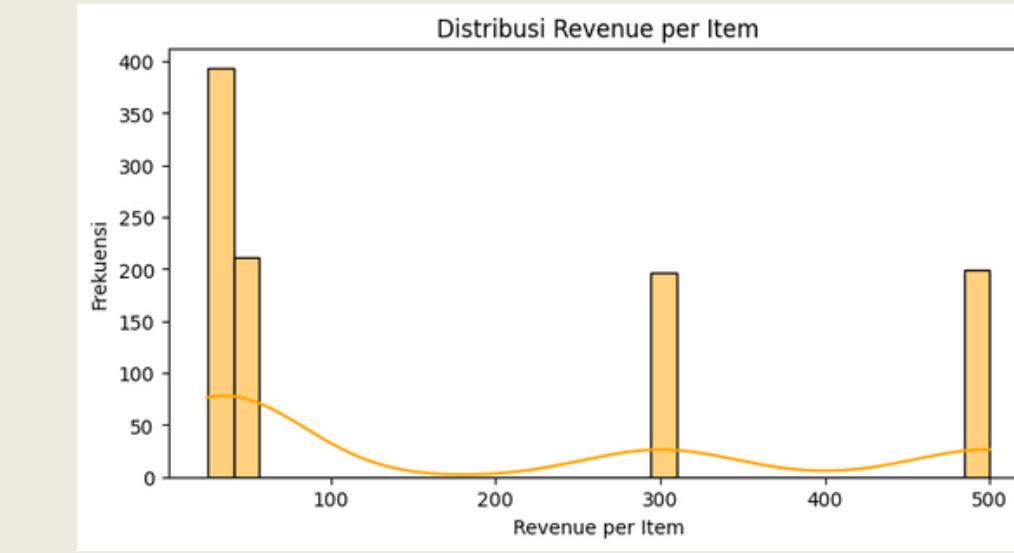
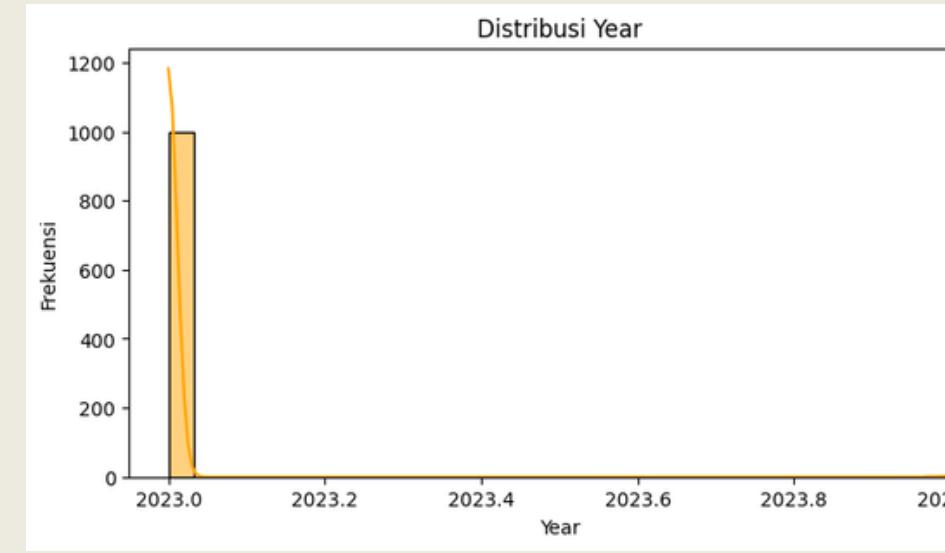
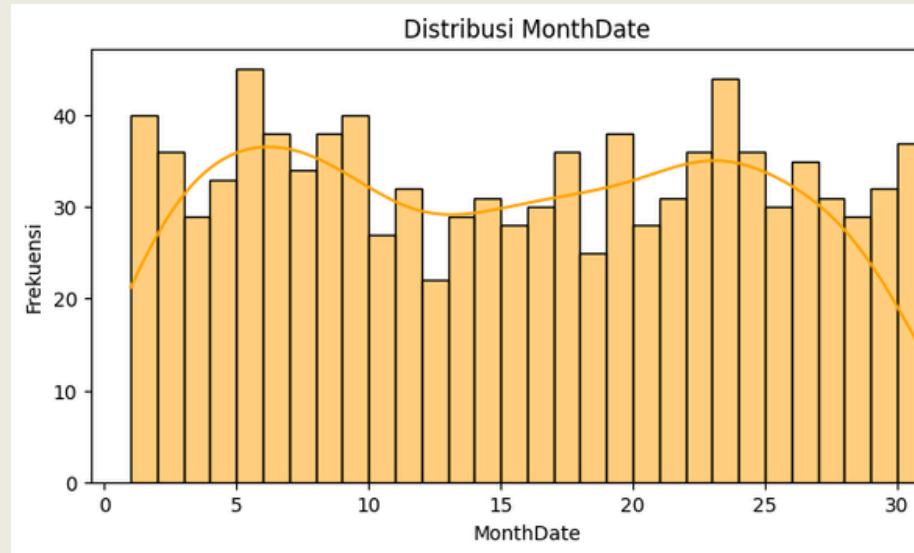
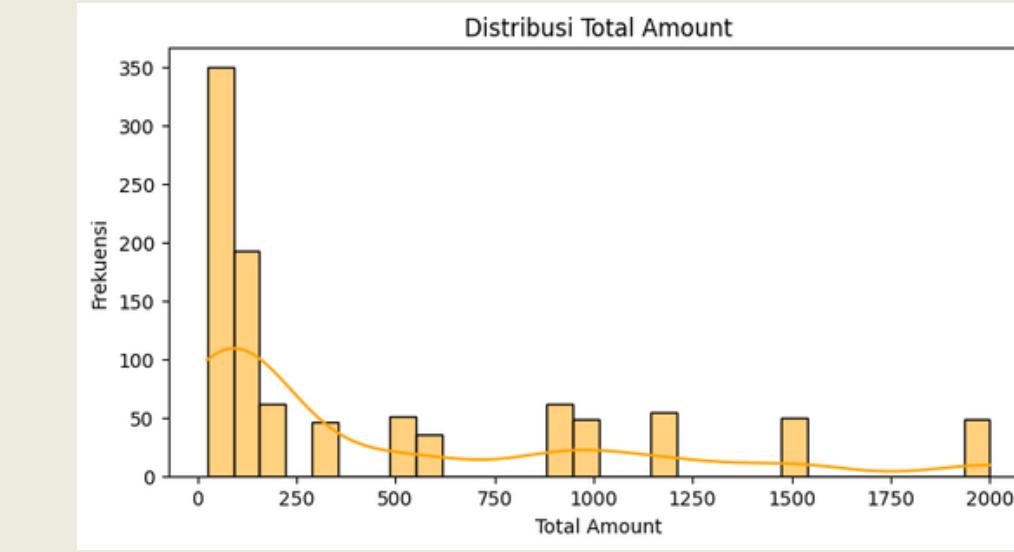
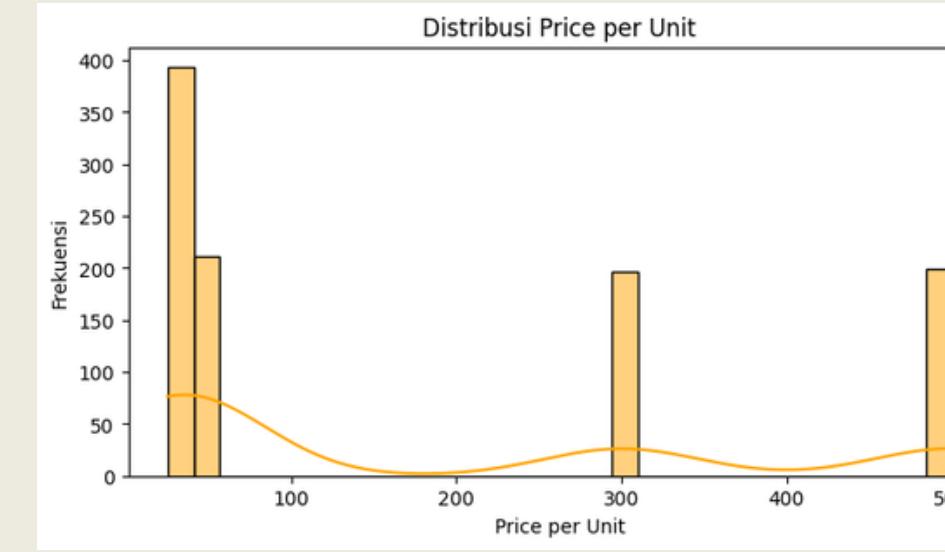
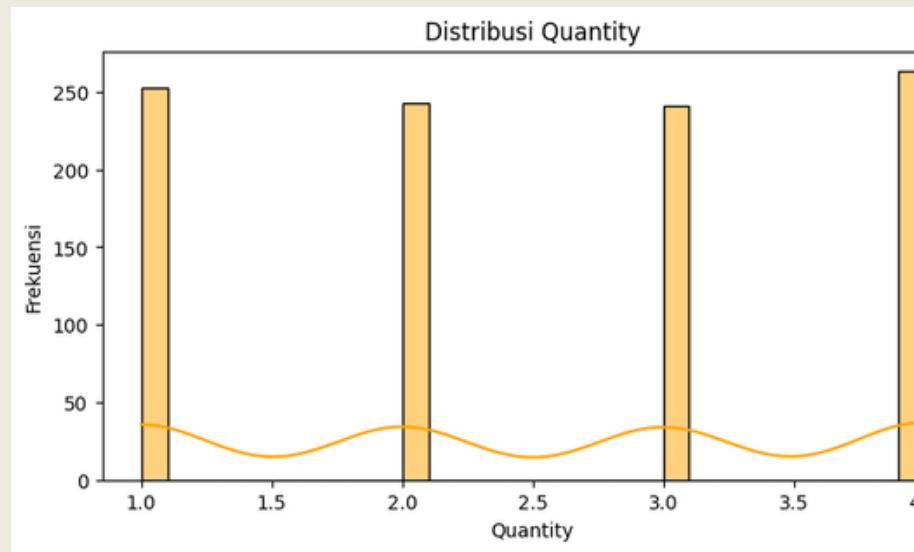
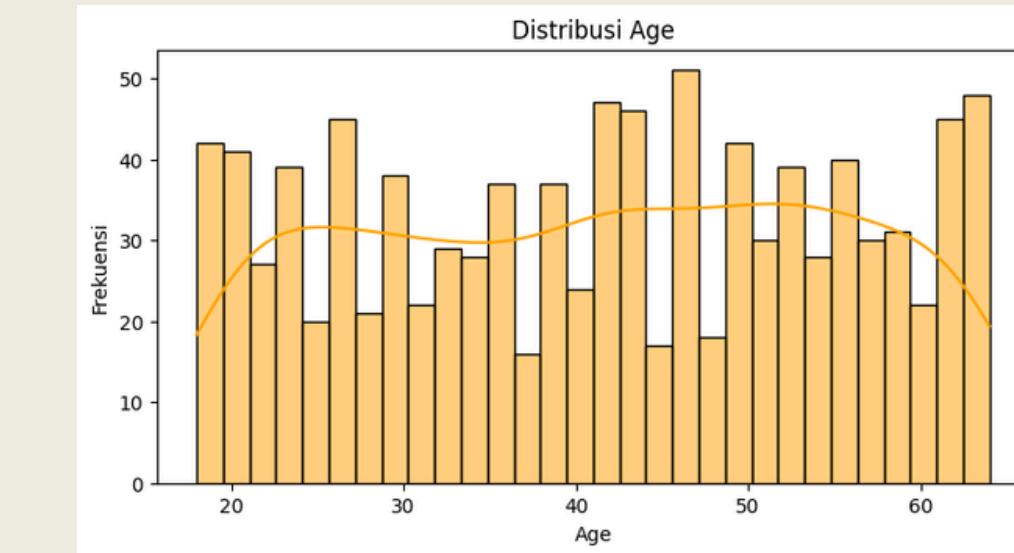
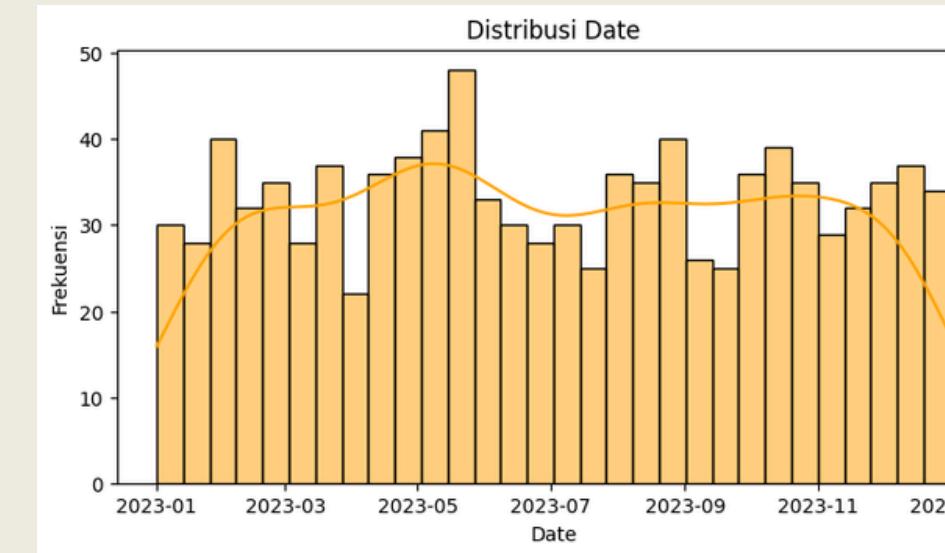
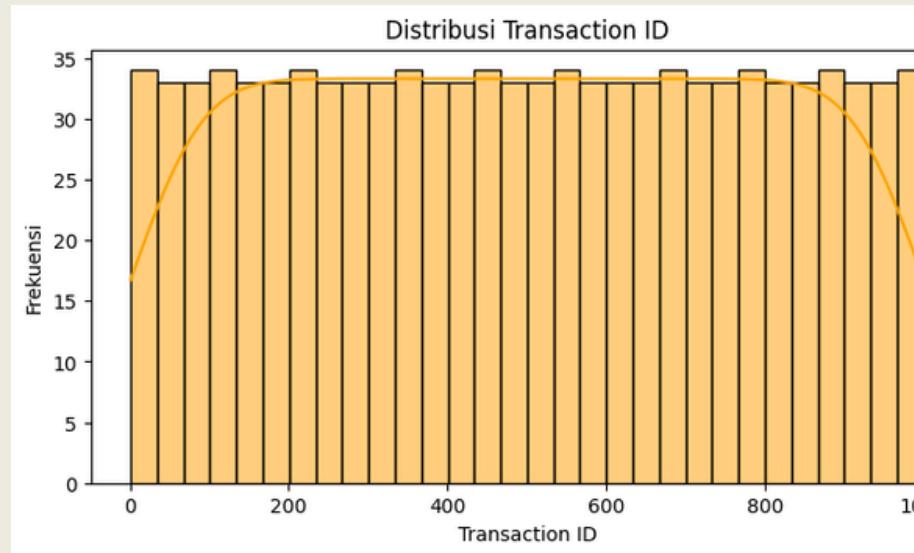
Jumlah Data terduplicat: 0

Outlier pada Data

- ▶ Outlier terdeteksi di Transaction ID: 0
- ▶ Outlier terdeteksi di Age: 0
- ▶ Outlier terdeteksi di Quantity: 0
- ▶ Outlier terdeteksi di Price per Unit: 0
- ▶ Outlier terdeteksi di Total Amount: 0
- ▶ Outlier terdeteksi di Revenue per Item: 0

3

EDA untuk Data Numerik



Penjelasan



Pada Plot Distribusi Transaksi ID diperoleh:

- Setiap rentang Transaction ID memiliki jumlah frekuensi yang serupa, sekitar 33–35 transaksi per rentang ID tertentu.
- Ini menunjukkan bahwa data transaksi terdistribusi secara acak dan seragam, tanpa adanya konsentrasi atau cluster pada ID tertentu.

Pada Plot Distribusi Date:

- Distribusi transaksi cukup merata sepanjang tahun, namun terjadi peningkatan signifikan pada bulan Mei 2023 dan terdapat penurunan transaksi yang terlihat pada Juli dan Agustus 2023.
- Pola ini bisa mengindikasikan adanya pengaruh oleh waktu maupun musim.
- Distribusi tidak simetris sempurna, tetapi cukup seimbang tanpa outlier ekstrem.

Pada Plot Distribusi Age:

Distribusi cukup merata pada rentang usia, dengan sedikit variasi pada beberapa kelompok umur. Meskipun distribusi tidak simetris sempurna tetapi cukup seimbang tanpa outlier.

Pada Plot Distribusi Quantity:

Kuantitas produk yang dibeli pelanggan terbatas hanya pada angka 1 hingga 4 unit per transaksi. Distribusinya cukup seragam, dengan frekuensi pembelian yang hampir seimbang di setiap jumlah kuantitas. Hal ini mengindikasikan bahwa pelanggan memiliki kebiasaan belanja yang serupa dalam hal jumlah produk yang dibeli.

Pada Plot Distribusi Price per Unit:

Harga produk per unit hanya terbatas pada nilai tertentu seperti 25, 50, 300, dan 500. Sebagian besar transaksi terjadi pada produk dengan harga lebih rendah (25 dan 50), menunjukkan bahwa barang dengan harga ekonomis lebih sering dibeli oleh pelanggan. Distribusi ini bersifat diskret, bukan menyebar merata, yang mengindikasikan bahwa jenis produk dalam dataset ini memiliki rentang harga tetap atau kategori harga tertentu.

Penjelasan

Pada Plot Distribusi Total Amount:

Distribusi Total Amount menunjukkan bahwa mayoritas pelanggan melakukan pembelian dengan total nilai transaksi yang relatif kecil. Hanya sebagian kecil pelanggan yang melakukan pembelian dalam jumlah besar. Pola ini umum dalam penjualan ritel, di mana sebagian besar transaksi bernilai rendah hingga sedang, sedangkan pembelian besar lebih jarang terjadi. Distribusi ini bersifat skewed (miring ke kanan), mencerminkan ketimpangan nilai transaksi.

Pada Plot Distribusi MonthDate:

Distribusi transaksi cukup merata dari hari ke hari dalam sebulan, Distribusi tidak simetris sempurna, tetapi cukup seimbang tanpa outlier ekstrem.

Pada Plot Distribusi Year:

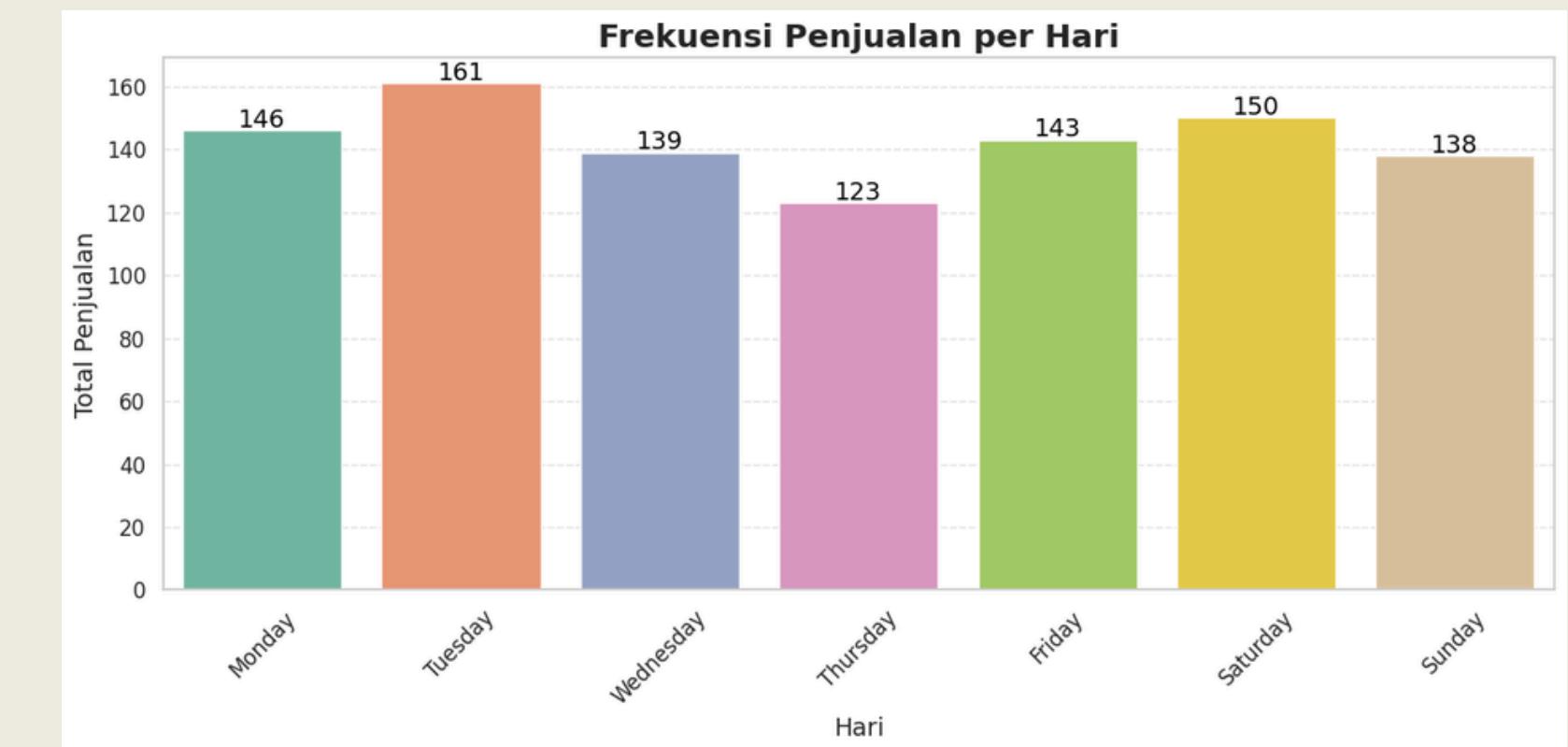
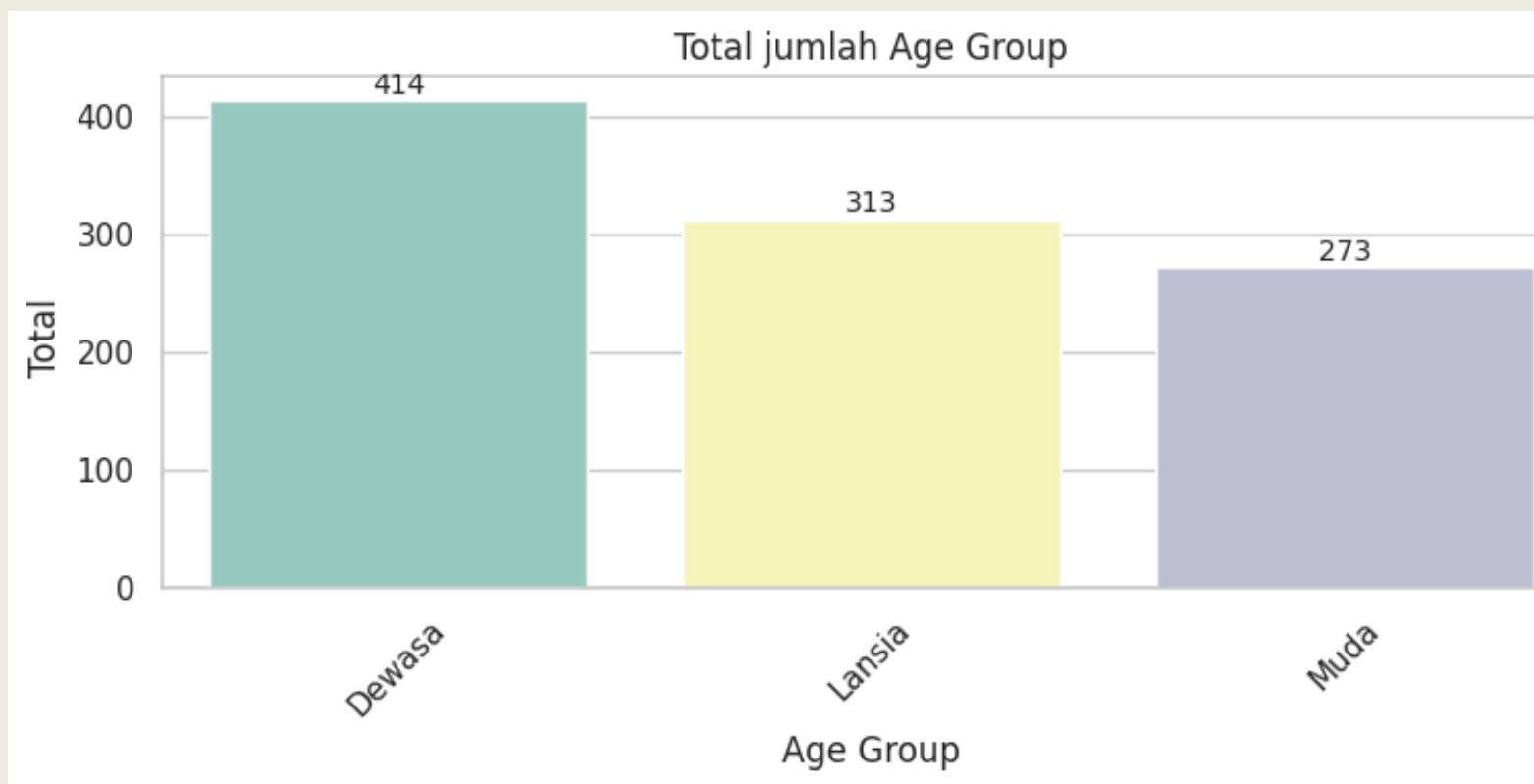
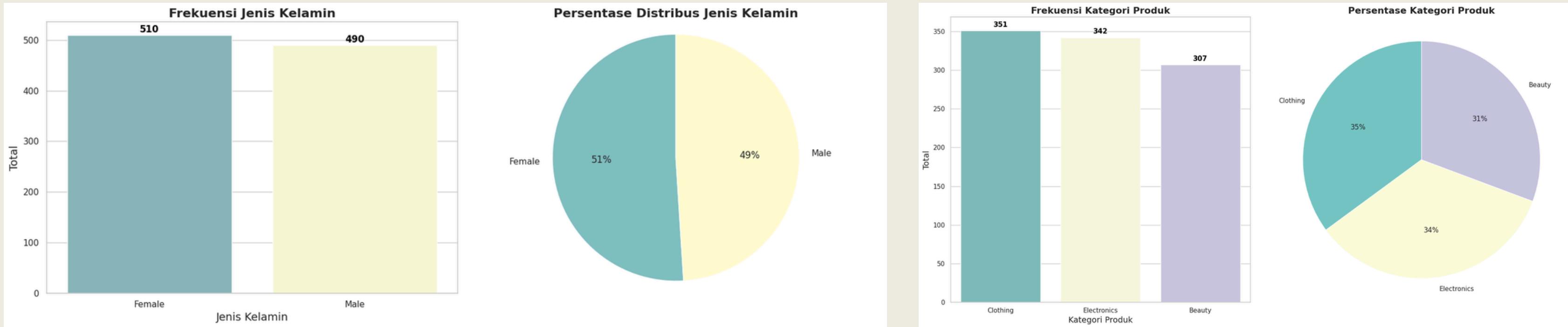
Sebagian besar data transaksi terkonsentrasi pada tahun 2023. Hal ini menunjukkan bahwa data penjualan yang tersedia mayoritas berasal dari periode tersebut. Sementara itu, data untuk tahun 2024 masih sangat terbatas atau belum terkumpul secara signifikan. Oleh karena itu, analisis tren tahunan sebaiknya difokuskan hanya pada tahun 2023, karena tahun 2024 belum mencerminkan pola yang cukup kuat untuk dianalisis.

Pada Plot Distribusi Revenue per item:

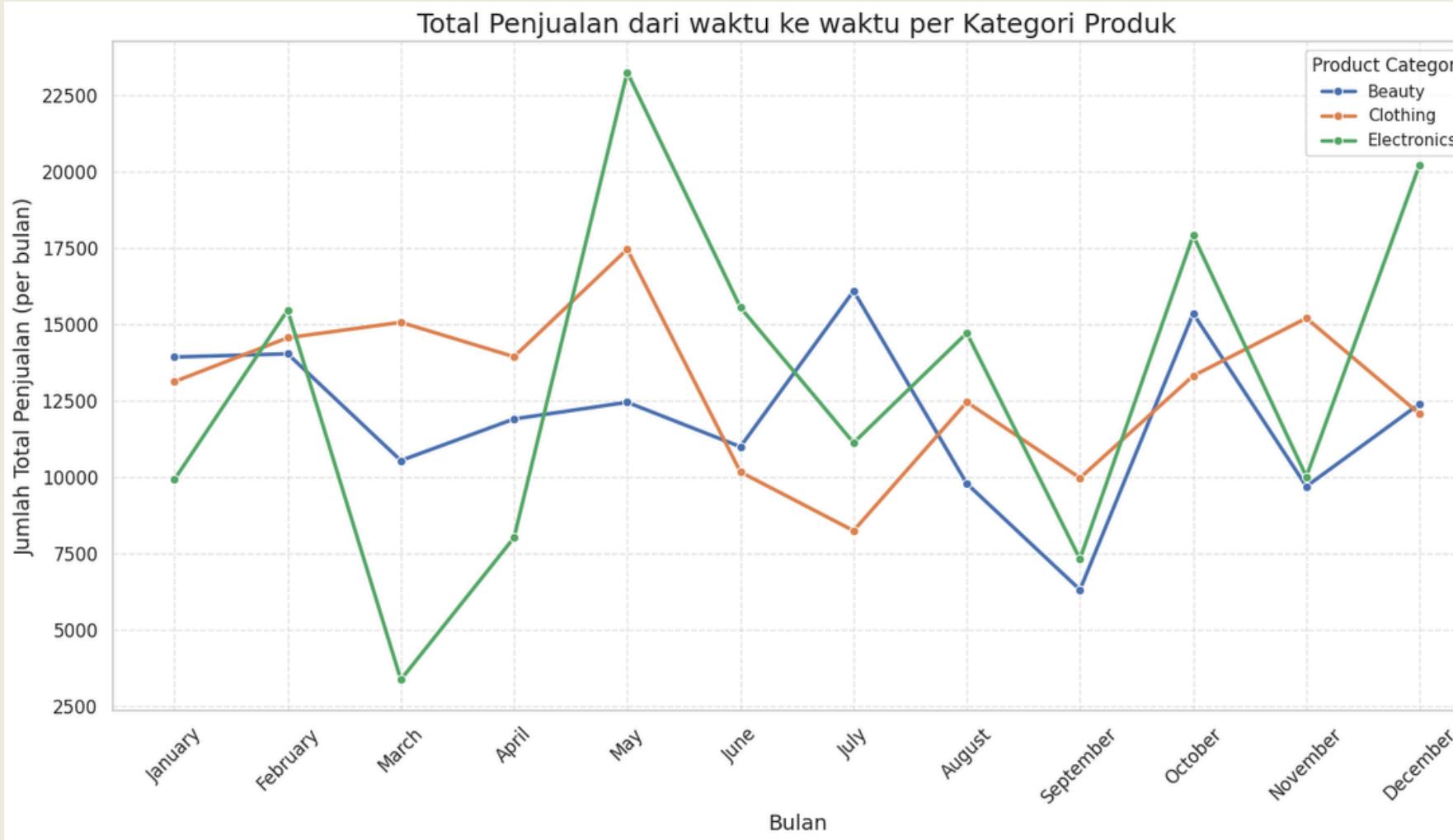
Distribusi pada plot Revenue selaras dengan distribusi harga per item, menunjukkan bahwa banyaknya penjualan pada harga yang lebih murah memberikan kontribusi cukup baik untuk keuntungan penjualan.



EDA untuk Data Kategorik

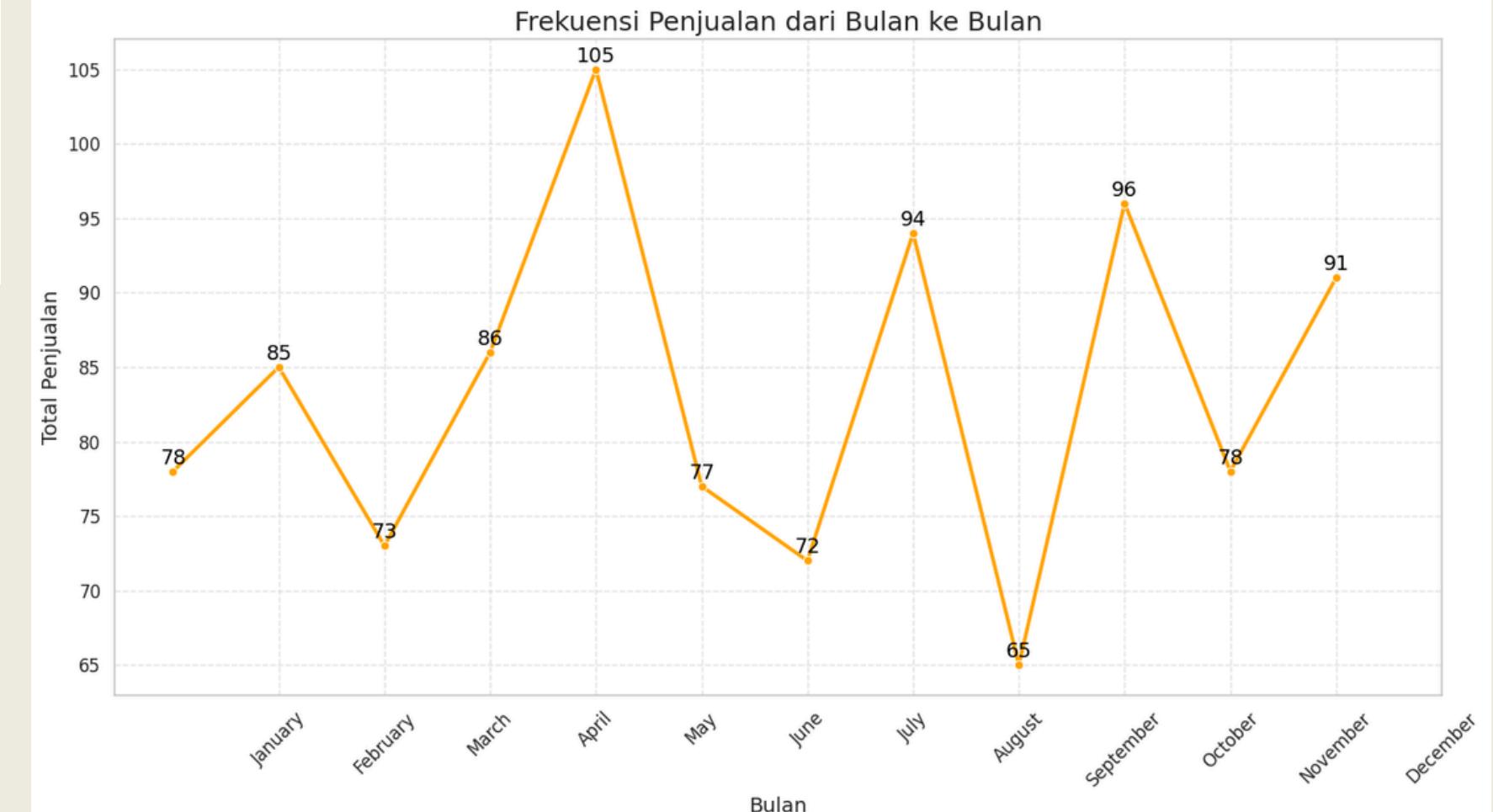


EDA untuk Data Kategorik



Per Kategori Produk

Seluruh Penjualan Produk





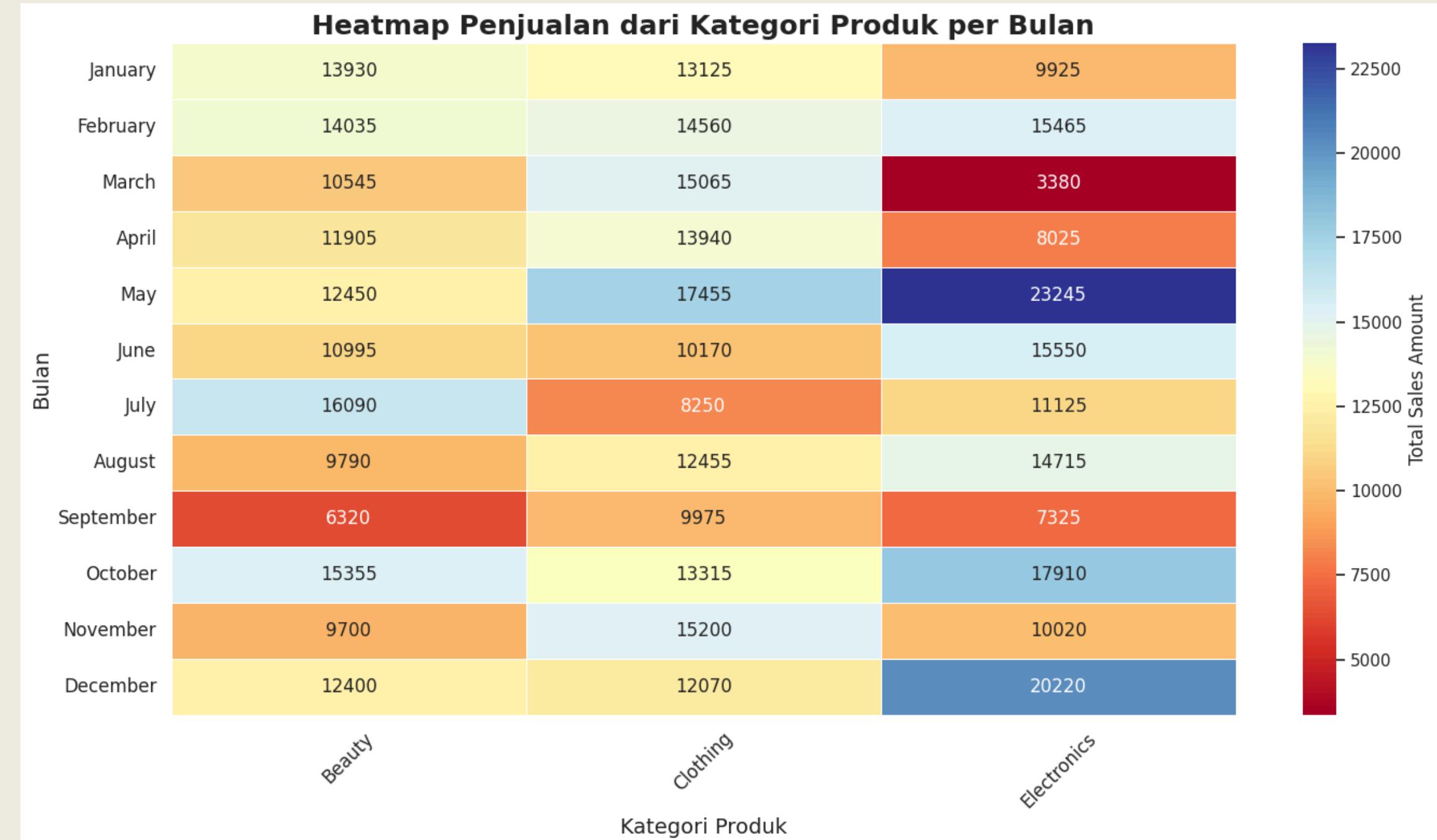
Penjelasan

- Tercatat jumlah pelanggan wanita sebanyak 510 dan pria 490, dengan porsi yang cukup seimbang. Sedikit lebih tingginya jumlah pembeli wanita kemungkinan mencerminkan kecenderungan perilaku belanja yang lebih aktif, namun hal ini perlu dikaji lebih lanjut agar tidak menimbulkan asumsi bias.
- Kategori produk yang paling banyak terjual adalah pakaian, diikuti oleh elektronik, dan terakhir produk kecantikan. Penjualan produk kecantikan yang lebih rendah kemungkinan disebabkan oleh segmen pasar yang lebih sempit, yaitu mayoritas wanita.
- Total pembelian berdasarkan kelompok usia dengan pembelian tertinggi adalah dewasa, diikuti lansia, dan terakhir usia muda. Hal ini mungkin dipengaruhi banyaknya kebutuhan dan daya beli kelompok usia dewasa.
- Penjualan tertinggi terjadi pada bulan Mei 2023, bertepatan dengan musim kemarau. Kenaikan ini mungkin dipicu oleh kebutuhan pakaian tertentu seperti pakaian ringan, atau bisa juga karena adanya promosi musiman.
- Hari dengan penjualan tertinggi adalah Selasa, sedangkan yang terendah adalah Kamis. Hal ini bisa menjadi acuan untuk penjadwalan promosi mingguan, misalnya dengan meningkatkan kampanye atau diskon khusus di hari Kamis untuk meratakan performa.

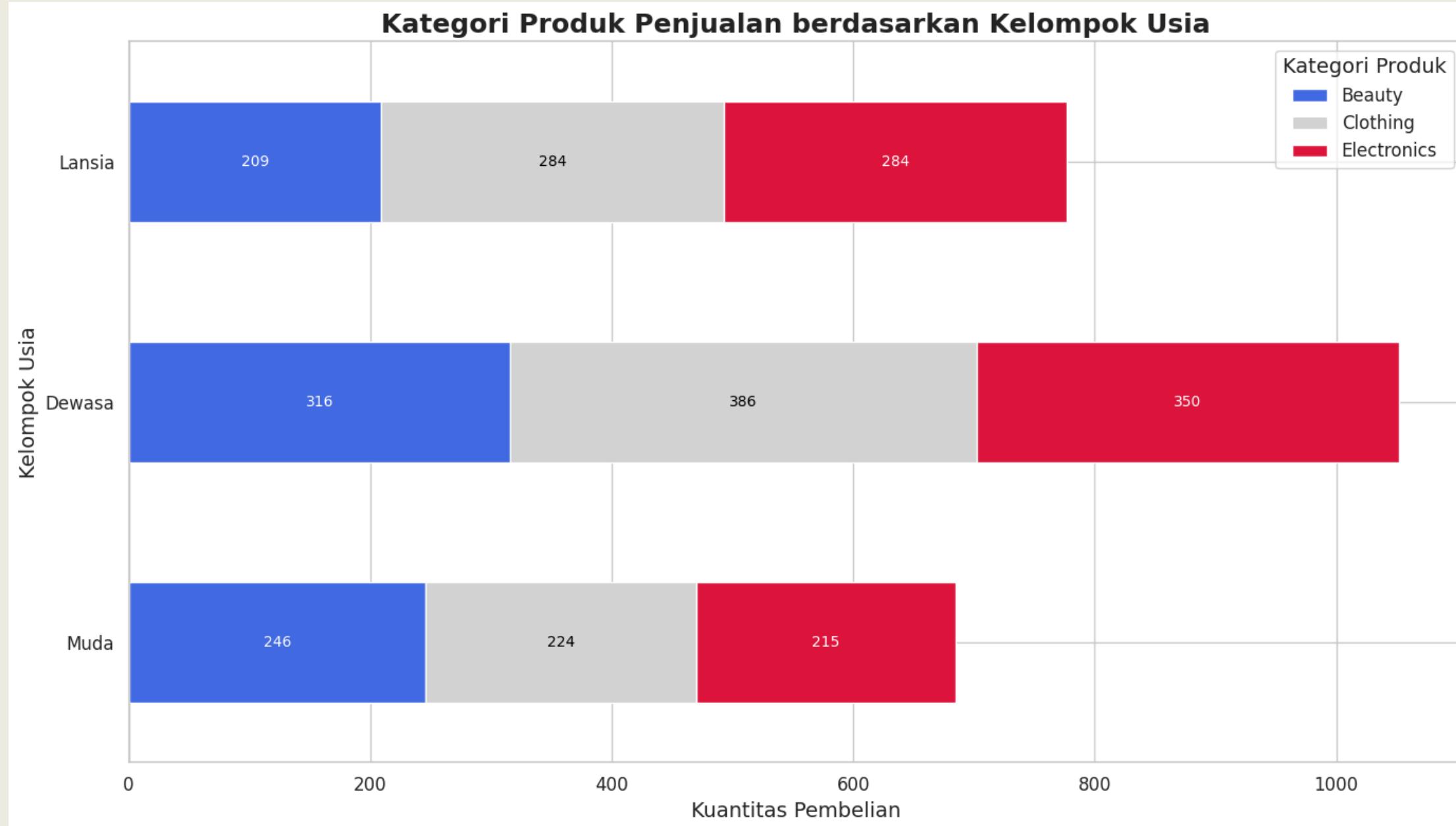
HeatMap Penjualan Produk

Berikut adalah korelasi antara kategori produk yang terjual dengan bulan. Berdasarkan kategori produk, bulan-bulan yang mempunyai kemungkinan penjualan terbaik yaitu,

Kategori	Bulan Terbaik	Nilai Penjualan
Beauty	July	16090
Clothing	May	17455
Electronics	May	23245

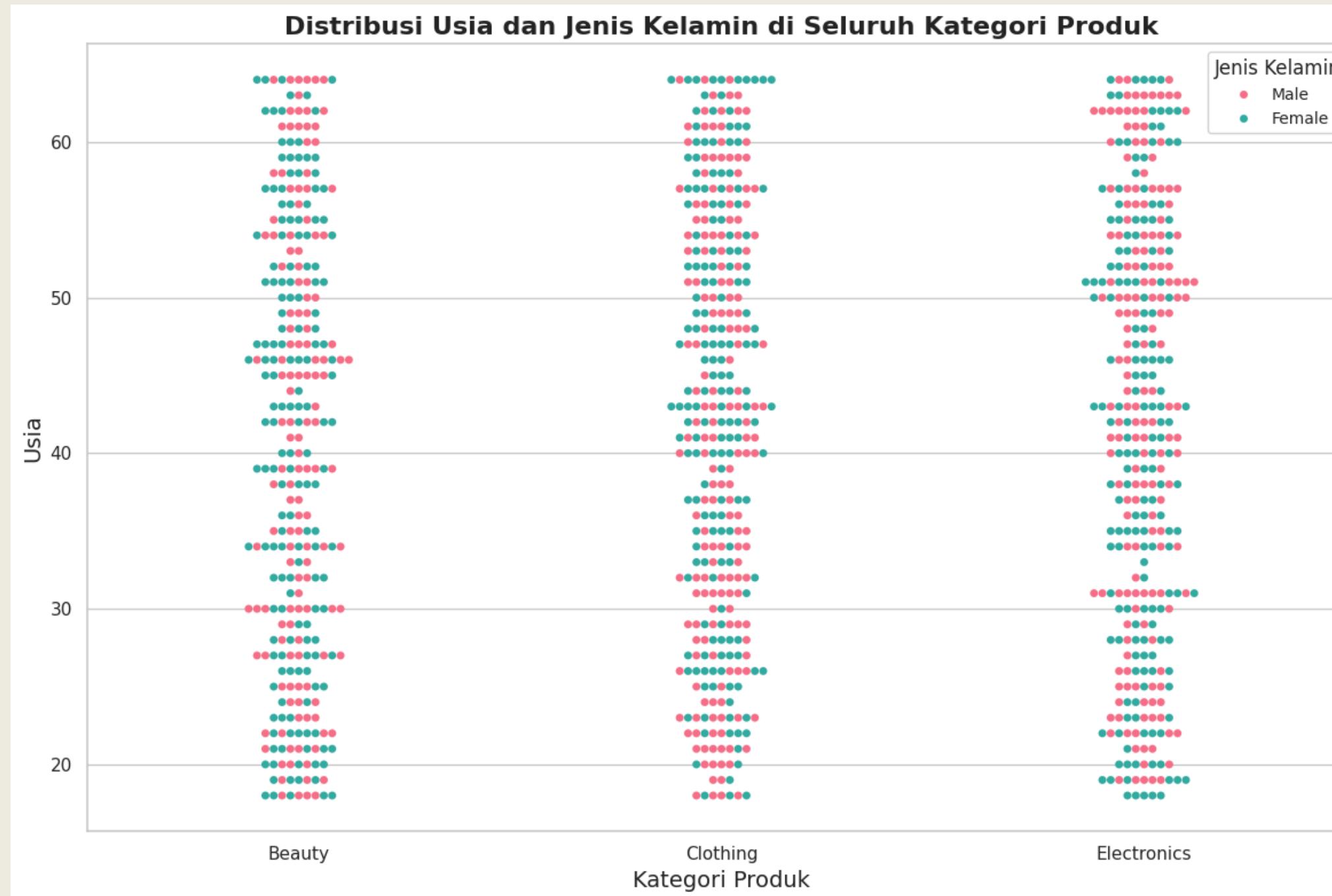


Penjualan Produk berdasarkan Kelompok Usia

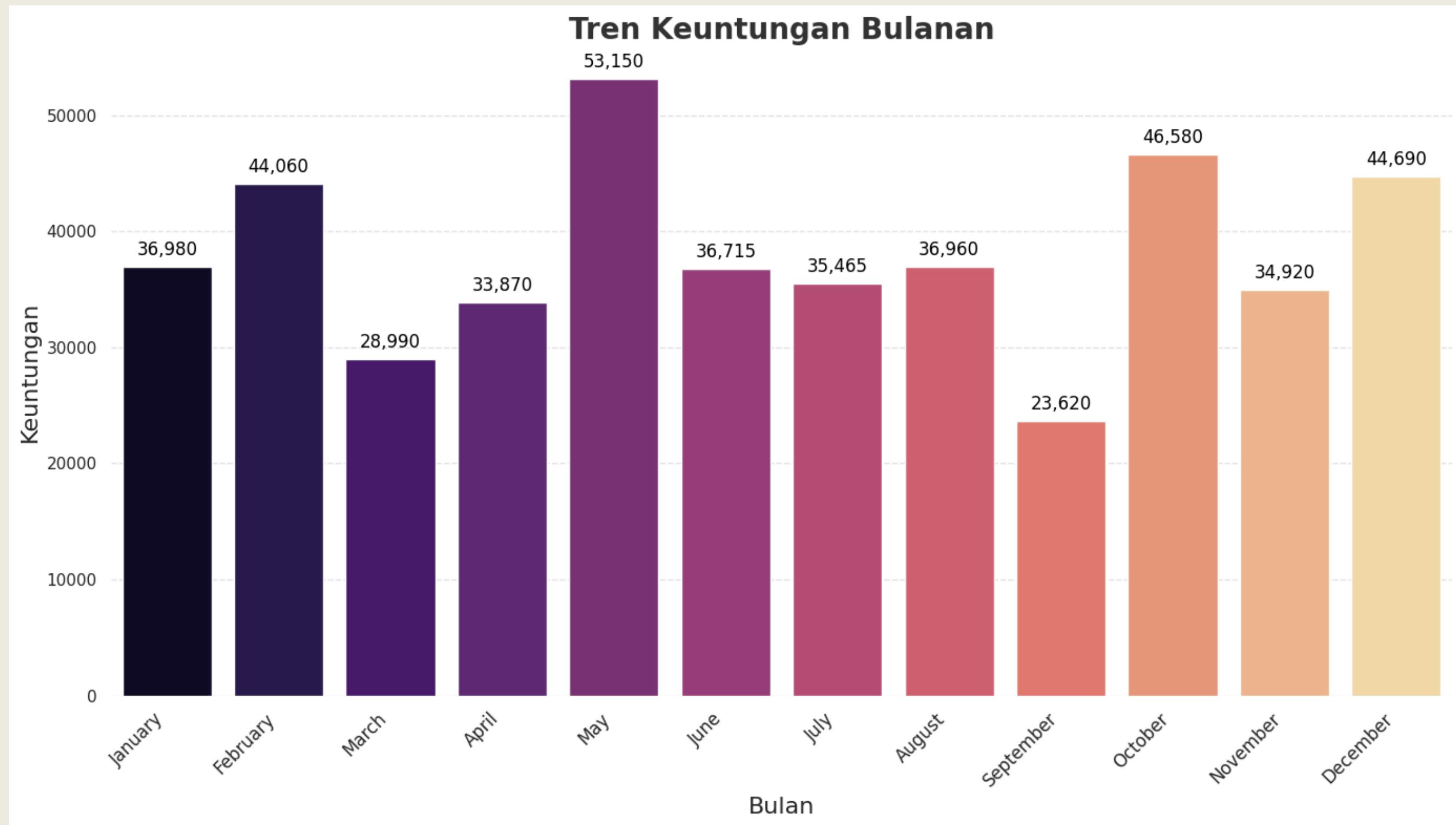


- Kelompok Dewasa merupakan penyumbang pembelian tertinggi secara keseluruhan, terutama pada kategori Clothing.
- Kelompok Lansia paling banyak membeli produk Electronics dibanding kategori lain.
- Kelompok Muda cenderung merata dalam pembelian, dengan sedikit dominasi pada kategori Beauty.

Penjualan Produk berdasarkan Sebaran Usia dan Jenis Kelamin

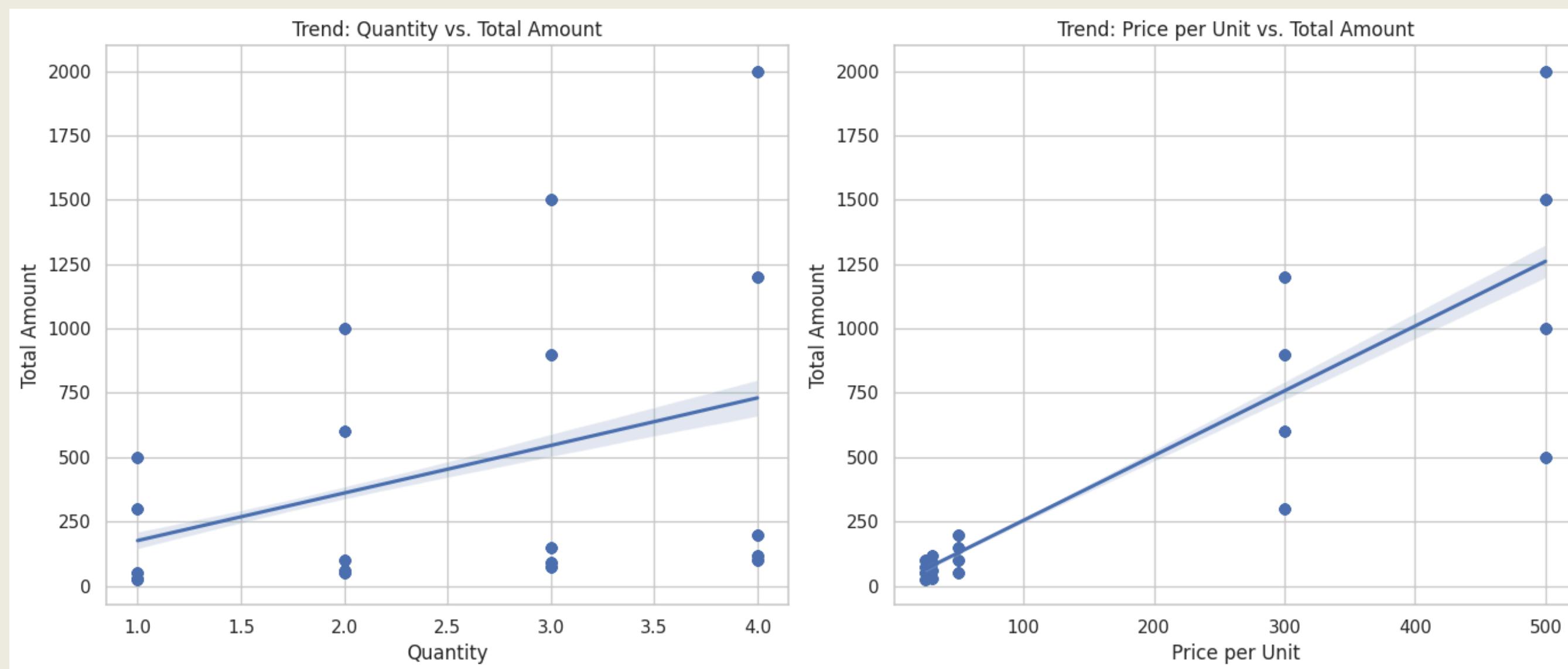


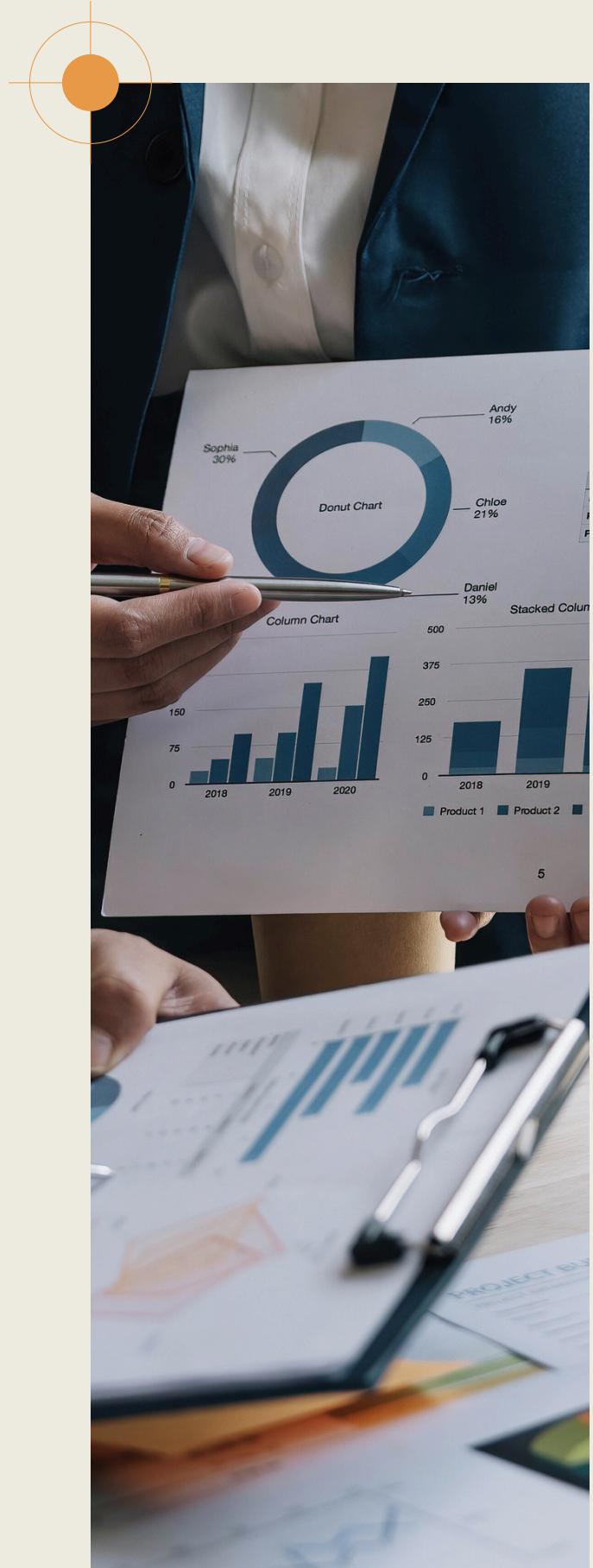
- Ketiga kategori produk (Beauty, Clothing, dan Electronics) memiliki sebaran usia konsumen yang relatif merata antara usia 18 hingga 65 tahun.
- Tidak ada dominasi yang mencolok antara pria dan wanita dalam masing-masing kategori — distribusi gender tampak seimbang.
- Produk Clothing dan Electronics diminati oleh rentang usia yang lebih luas dibandingkan produk Beauty.
- Hal ini menunjukkan bahwa semua kelompok usia dan gender berkontribusi terhadap pembelian di tiap kategori produk, meskipun dengan sedikit variasi preferensi.



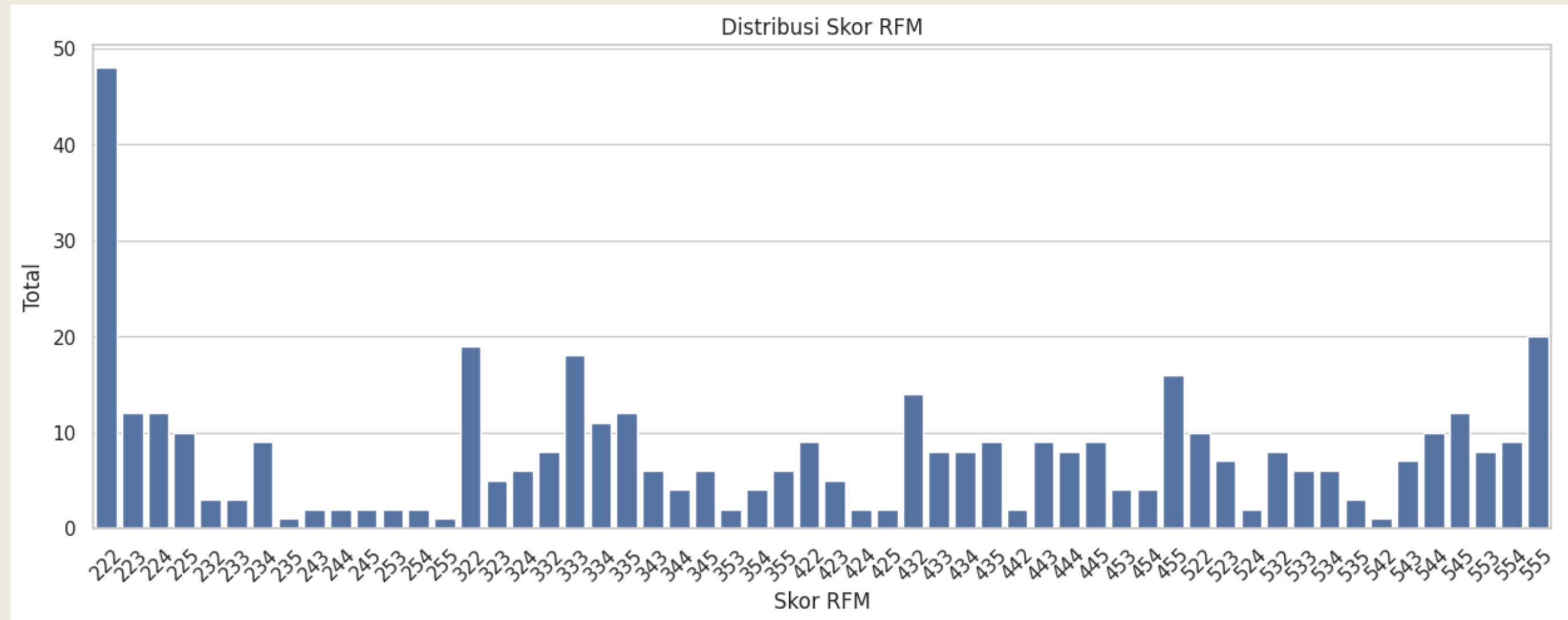
Berdasarkan grafik keuntungan, keuntungan terbesar ada pada bulan Mei (53.15%), hal ini selaras dengan tingginya penjualan pada bulan Mei.

- Quantity vs. Total Amount: Terdapat hubungan positif antara kuantitas barang yang dibeli dengan total jumlah pembayaran, tetapi hubungan ini tidak terlalu kuat. Ini mengindikasikan bahwa jumlah pembelian saja belum cukup untuk menjelaskan besar kecilnya total pembayaran.
- Price per Unit vs. Total Amount: Terdapat korelasi yang lebih kuat antara harga satuan dan total pembayaran. Artinya, semakin tinggi harga per unit produk, semakin besar total pembayaran, menunjukkan bahwa harga satuan lebih berpengaruh terhadap total transaksi dibandingkan kuantitas.



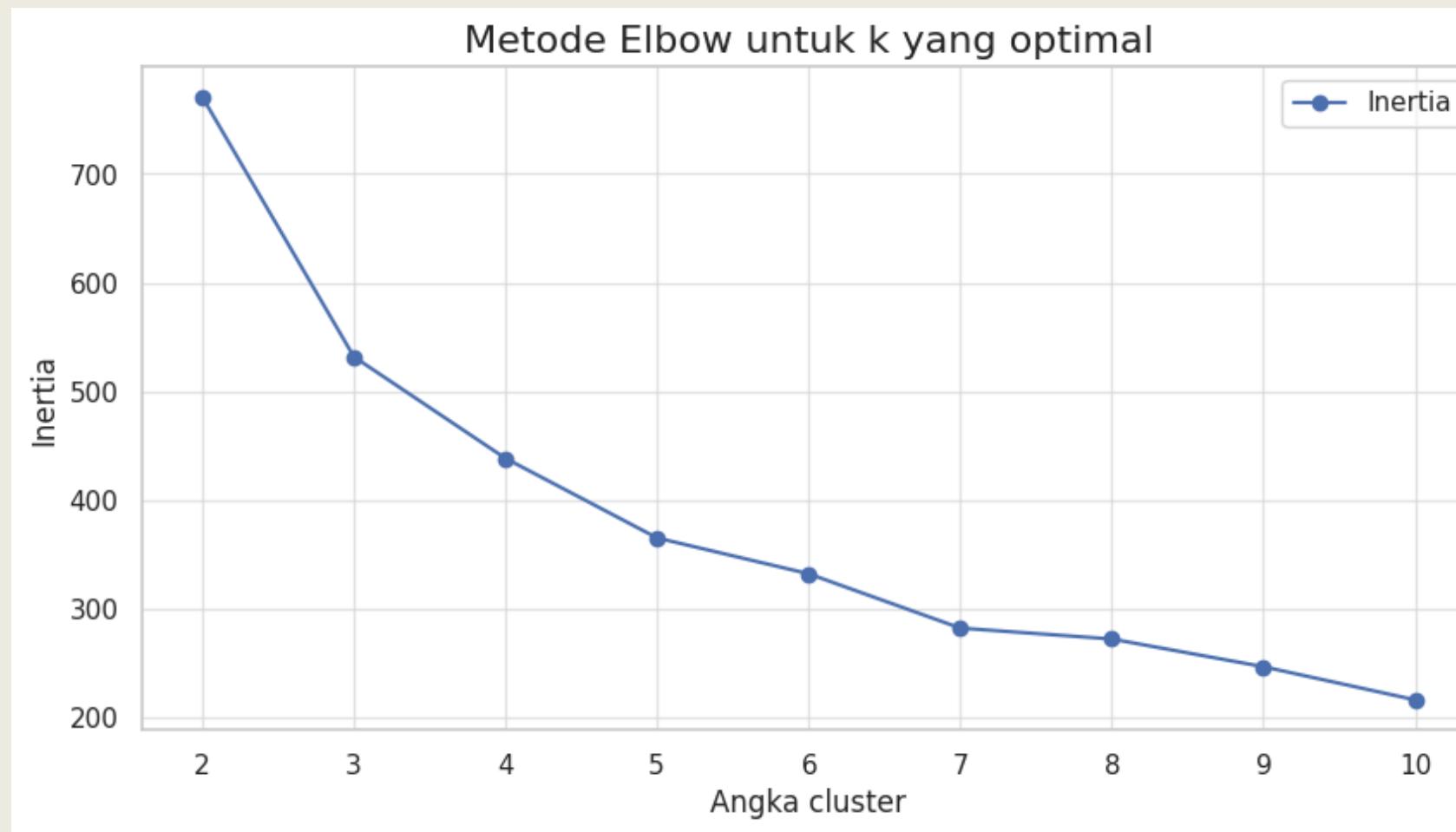


RFM



Dari plot dapat dilihat bahwa skor RFM paling tinggi berada pada skor 222, ini menunjukkan mayoritas pelanggan berada di segmen pasif atau tidak aktif — mereka jarang bertransaksi dan tidak menghasilkan banyak nilai bagi bisnis.

K-Means untuk Clustering RFM



Menentukan jumlah kluster yang optimal (diperoleh 4)

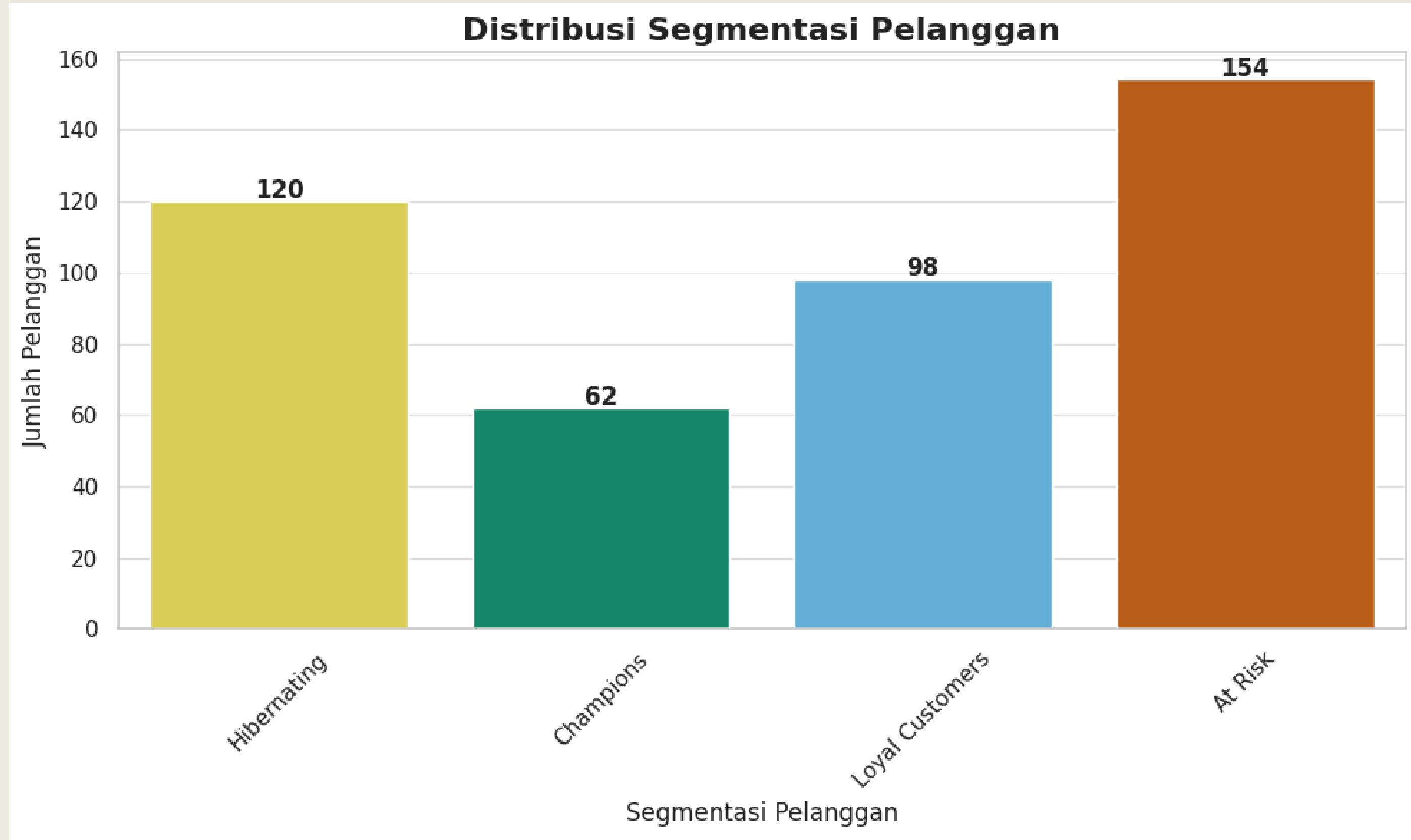
Tabel karakteristik rata-rata setiap kluster berdasarkan dimensi RFM

Ringkasan Cluster (KMeans):				
Cluster_KMeans	Recency	Frequency	Monetary	
0	259.858333	1.291667	664.041667	
1	59.548387	4.145161	2852.580645	
2	72.112245	3.469388	1027.040816	
3	81.850649	1.610390	641.590909	

Segmentasi pelanggan ditentukan berdasarkan hasil klaster

Segment_KMeans	Count
0	At Risk
1	Hibernating
2	Loyal Customers
3	Champions

SEGMENTASI PELANGGAN



Rekomendasi



- **Kelompok Pelanggan Champion (62 pelanggan)**

1. Memberikan penghargaan atas loyalitas yang dilakukan baik bentuk voucher, fasilitas VIP, diskon khusus, maupun penawaran eksklusif.
2. Pertahankan keterlibatan pelanggan melalui email ucapan terimakasih maupun ucapan khusus dalam setiap perayaan mereka.

- **Kelompok Pelanggan Loyal Customer (98 pelanggan)**

1. Perkuat loyalitas pelanggan dengan adanya promo khusus, bonus belanja, maupun program berhadiah.
2. Jaga keterlibatan dengan membagikan informasi terkini tentang produk baru, diskon, maupun acara khusus anggota.

- **Kelompok Pelanggan At Risk (154 pelanggan)**

1. Atasi penurunan dengan adanya kampanye yang dipersonalisasi melalui email maupun sosial media pelanggan.
2. Berikan informasi terkait dengan produk baru serta diskon agar pelanggan dapat tertarik kembali untuk membeli produk.

- **Kelompok Pelanggan Hibernating (120 pelanggan)**

1. Bangkitkan kembali minat mereka dengan adanya promo besar-besaran, kampanye "Kami Merindukanmu", ataupun menawarkan paket musiman.
2. Lakukan survei untuk mengetahui alasan utama pemutusan hubungan dan gunakan umpan balik untuk meningkatkan layanan.



THANK
YOU

arfinadhifa3@gmail.com



[More detail!!](#)

