

DIGITAL SKILL FAIR 41.0

Presentasi Oleh: Sheila Novitasari

**Faculty of Data: Data Science
Analyst**

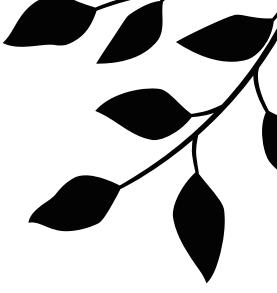


ABOUT ME & THE EVENT

ABOUT ME

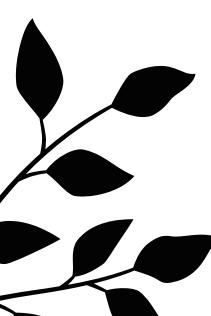
Seorang mahasiswa Statistika dengan antusiasme tinggi untuk belajar dan berkarir di bidang data. Mengikuti Digital Skill Fair 41.0 untuk mendalami fundamental Data Science dan membangun portofolio saya.





DIBIMBING.ID DIGITAL SKILL FAIR 41.0: FACULTY OF DATA

- 1** Program intensif 5 hari untuk upgrade skill di bidang data.
- 2** Fokus pada track Data Science Analyst.
- 3** Materi mencakup: Python Fundamentals, Exploratory Data Analysis (EDA), dan Career Preparation.



MY 5-DAY LEARNING JOURNEY

Day 1 & 2: Building the Foundation

- Mendapat gambaran umum industri data dan pengenalan dasar Python.
- Mengerjakan tugas pertama: program sederhana input data mahasiswa menggunakan logika if-elif-else.

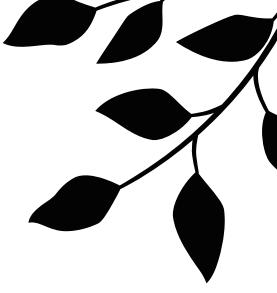
Day 3: Diving into The Core Skill – EDA

- Mempelajari proses Exploratory Data Analysis (EDA) dari data mentah hingga kesimpulan.
- Fokus pada teknik Data Cleansing: menangani data kosong (missing values) dan duplikat.

Day 4 & 5: Career Preparation & Showcase

- Mendapat wawasan tentang optimasi profil LinkedIn & GitHub.
- Belajar cara menyusun CV yang efektif dan mendapat kesempatan review portofolio.

PROJECT SHOWCASE: EXPLORATORY DATA ANALYSIS ON E- COMMERCE SALES DATA



BUSINESS PROBLEM (MASALAH BISNIS)

1

Membersihkan
mempersiapkan
transaksi
analisis.

2

Melakukan
deskriptif
menemukan pola dan
wawasan.

3

Menyajikan
kunci
visualisasi data.

Sebuah toko ritel online ingin memahami perilaku pelanggan dan performa penjualannya. Pertanyaan utama yang ingin dijawab adalah: Produk apa yang paling laris? Negara mana yang menjadi pasar terbesar? Dataset yang digunakan yaitu dataset Online Retail II dari Kaggle.



THE PROCESS: DATA CLEANSING & PREPARATION

LANGKAH 1: HANDLING MISSING VALUES

Problem

Ditemukan sejumlah besar data kosong pada kolom CustomerID.

Solution

Baris data tanpa CustomerID dihapus karena transaksi tidak dapat diatribusikan ke pelanggan spesifik, sehingga tidak relevan untuk analisis perilaku pelanggan.

Visual

Invoice	0	Invoice	0
StockCode	0	StockCode	0
Description	4382	Description	0
Quantity	0	Quantity	0
InvoiceDate	0	InvoiceDate	0
Price	0	Price	0
Customer ID	243007	Customer ID	0
Country	0	Country	0
			dtype: int64

Before

After

LANGKAH 2: HANDLING DUPLICATES

Problem

Terdeteksi adanya baris data transaksi yang terduplicasi.

Solution

Menghapus semua data duplikat untuk memastikan setiap transaksi dihitung satu kali dan menjaga akurasi analisis.

Visual

The screenshot shows a Jupyter Notebook interface with three code cells:

- Cell [13]:

```
# Mengecek apakah ada duplicate di seluruh kolom  
check_duplicate = df.duplicated().sum()  
print(f"Jumlah data yang duplikat= {check_duplicate}")
```

Output: 1.4s
... Jumlah data yang duplikat= 34335
- Cell [14]:

```
# Handling duplicates  
df = df.drop_duplicates()
```

Output: 1.2s
- Cell [15]:

```
# Mengecek duplicate setelah dihandle  
handle_duplicate = df.duplicated().sum()  
print(f"Jumlah data yang duplikat= {handle_duplicate}")
```

Output: 1.0s
... Jumlah data yang duplikat= 0

Red boxes highlight the output of the first cell (34335) and the last cell (0), indicating the removal of duplicates.

THE PROCESS: FEATURE ENGINEERING

1

Membuat kolom baru bernama TotalPrice untuk menghitung total nilai per transaksi.

2

Rumus: TotalPrice = Quantity * Price

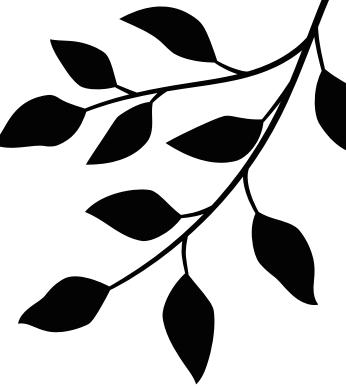
3

Tujuan: Kolom tersebut sangat penting untuk mengukur performa penjualan secara moneter, bukan hanya dari jumlah barang yang terjual.

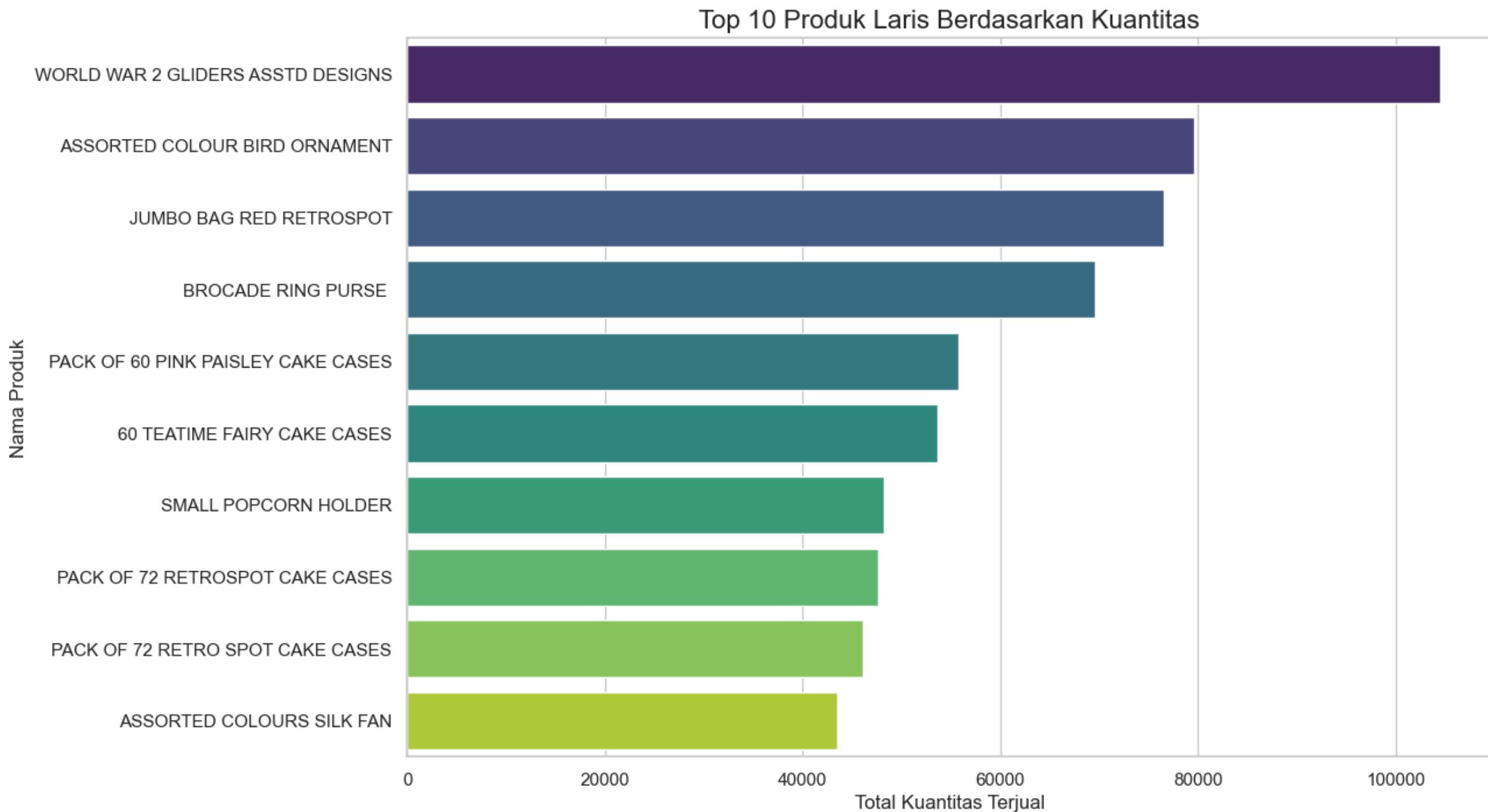
Description	Quantity	InvoiceDate	Price	Customer ID	Country	TotalPrice
15CM CHRISTMAS GLASS BALL	12	2009-12-01 07:45:00	6.95	13085.0	United Kingdom	83.40
20 LIGHTS PINK CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom	81.00
WHITE CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom	81.00
RECORD FRAME 7" SINGLE SIZE	48	2009-12-01 07:45:00	2.10	13085.0	United Kingdom	100.80
TRAWBERRY CERAMIC RINKET BOX	24	2009-12-01 07:45:00	1.25	13085.0	United Kingdom	30.00

KEY FINDINGS & BUSINESS INSIGHTS

TOP 10 PRODUK PALING LARIS

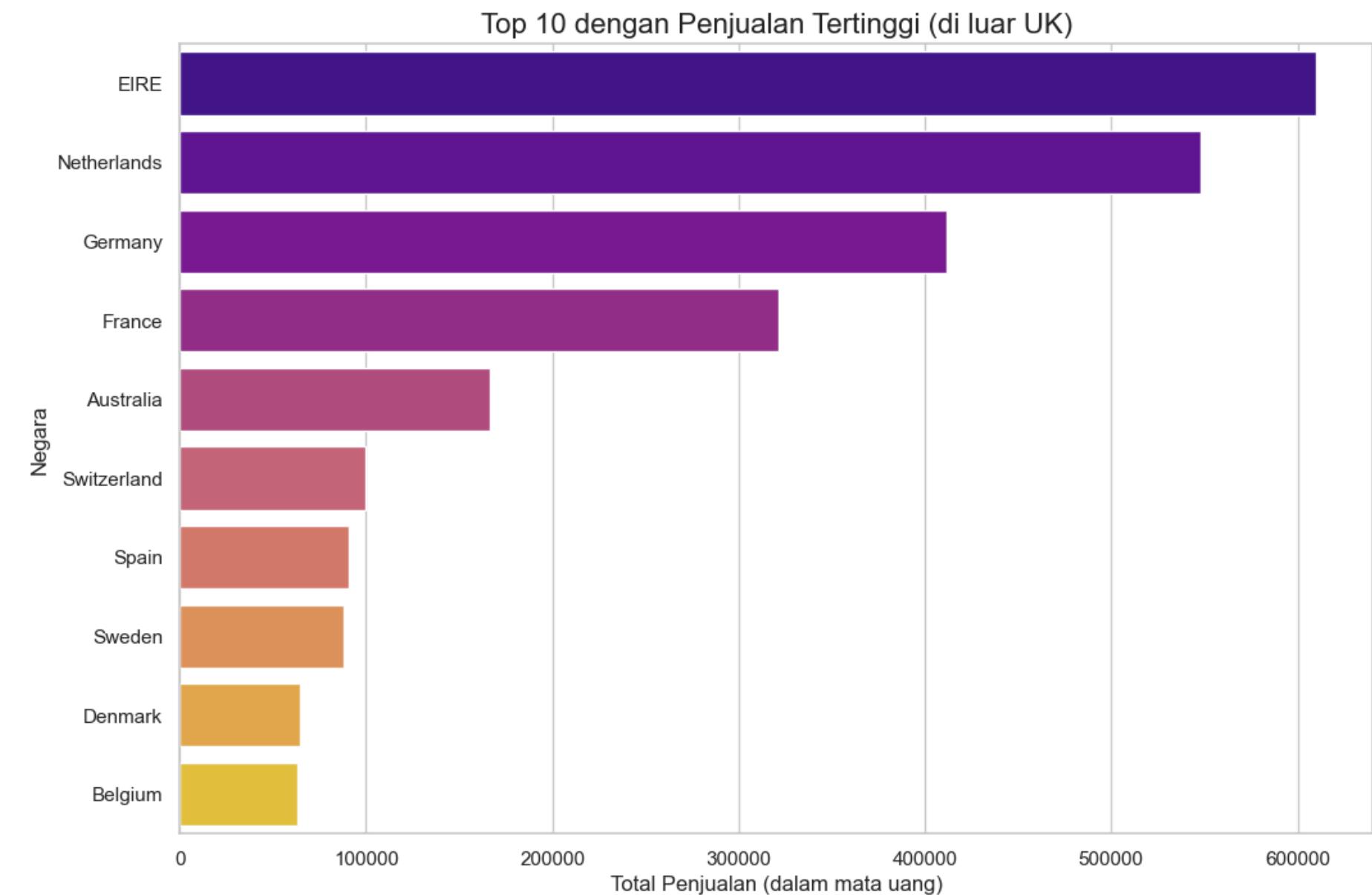


Produk 'World War 2 Gliders Asstd Design' dan 'Assorted Colour Bird Ornament' adalah yang paling diminati. Ini bisa menjadi dasar untuk manajemen stok dan strategi promosi.



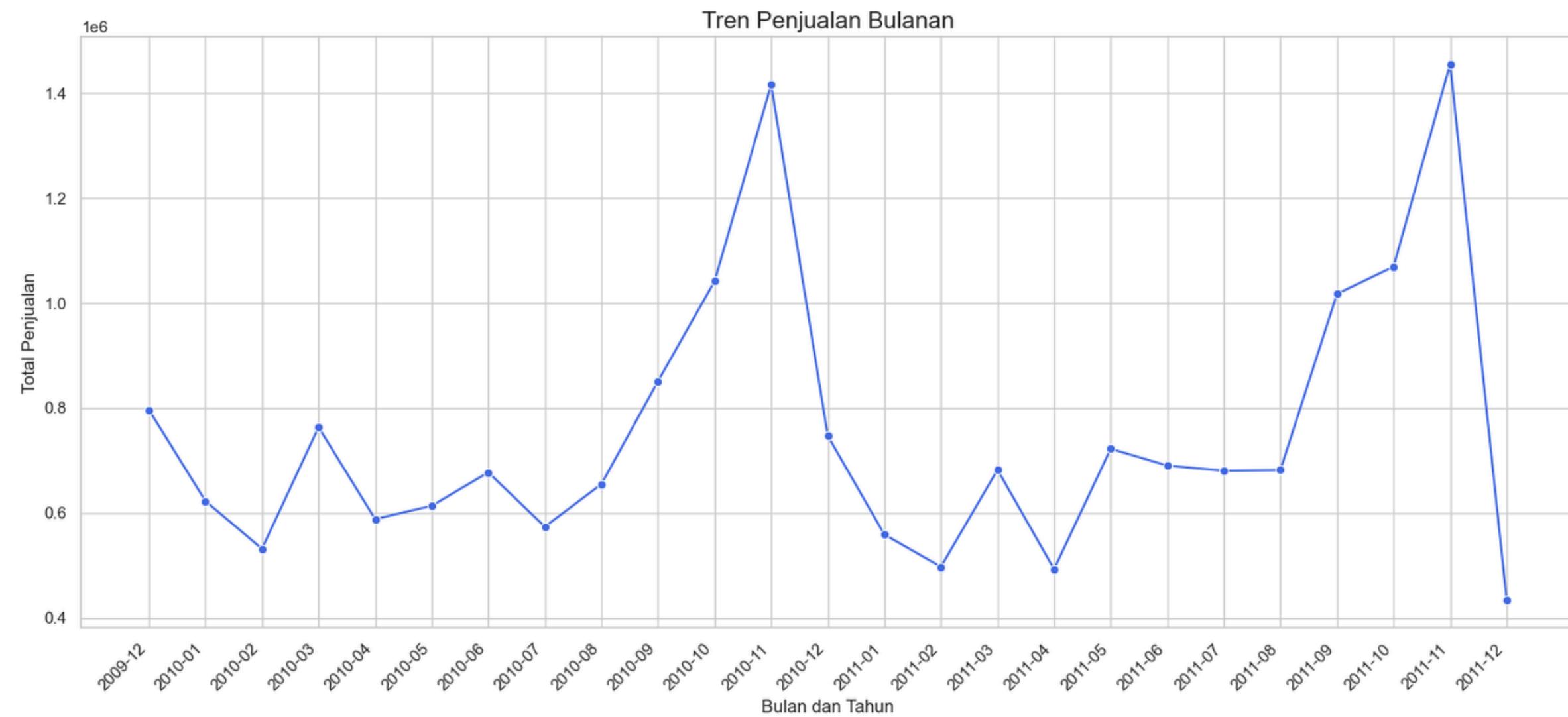
PENJUALAN BERDASARKAN NEGARA

Jika kita mengecualikan United Kingdom, 2 negara dengan penjualan tertinggi adalah EIRE (Irlandia) dan Netherlands (Belanda).



ANALISIS TREND PENJUALAN BULANAN

Grafik menunjukkan pola penjualan musiman yang kuat, dengan puncak penjualan terjadi pada bulan November setiap tahunnya, yang didorong oleh momentum belanja liburan.



CONCLUSION & KEY TAKEAWAYS

Melalui EDA, berhasil diidentifikasi produk dan pasar utama dari data penjualan. Proses pembersihan data sangat krusial untuk menghasilkan analisis yang akurat dan dapat diandalkan.

Pembelajaran Pribadi (Personal Learnings):

- "Memahami alur kerja seorang Data Analyst dari data mentah hingga insight."
- "Meningkatkan kemahiran teknis dalam menggunakan Python dan Pandas untuk manipulasi data."
- "Belajar pentingnya menceritakan data (data storytelling) melalui visualisasi."

TERIMA KASIH

**Saya sangat terbuka
untuk diskusi, feedback,
dan peluang kolaborasi.
Mari terhubung!**

- LinkedIn: <https://www.linkedin.com/in/shlnvt>
- GitHub: <https://github.com/shlnvt>
- Repository untuk kode lengkapnya:
<https://github.com/shlnvt/dibimbing-portofolio/tree/main>