

LAPORAN TUGAS MANDIRI

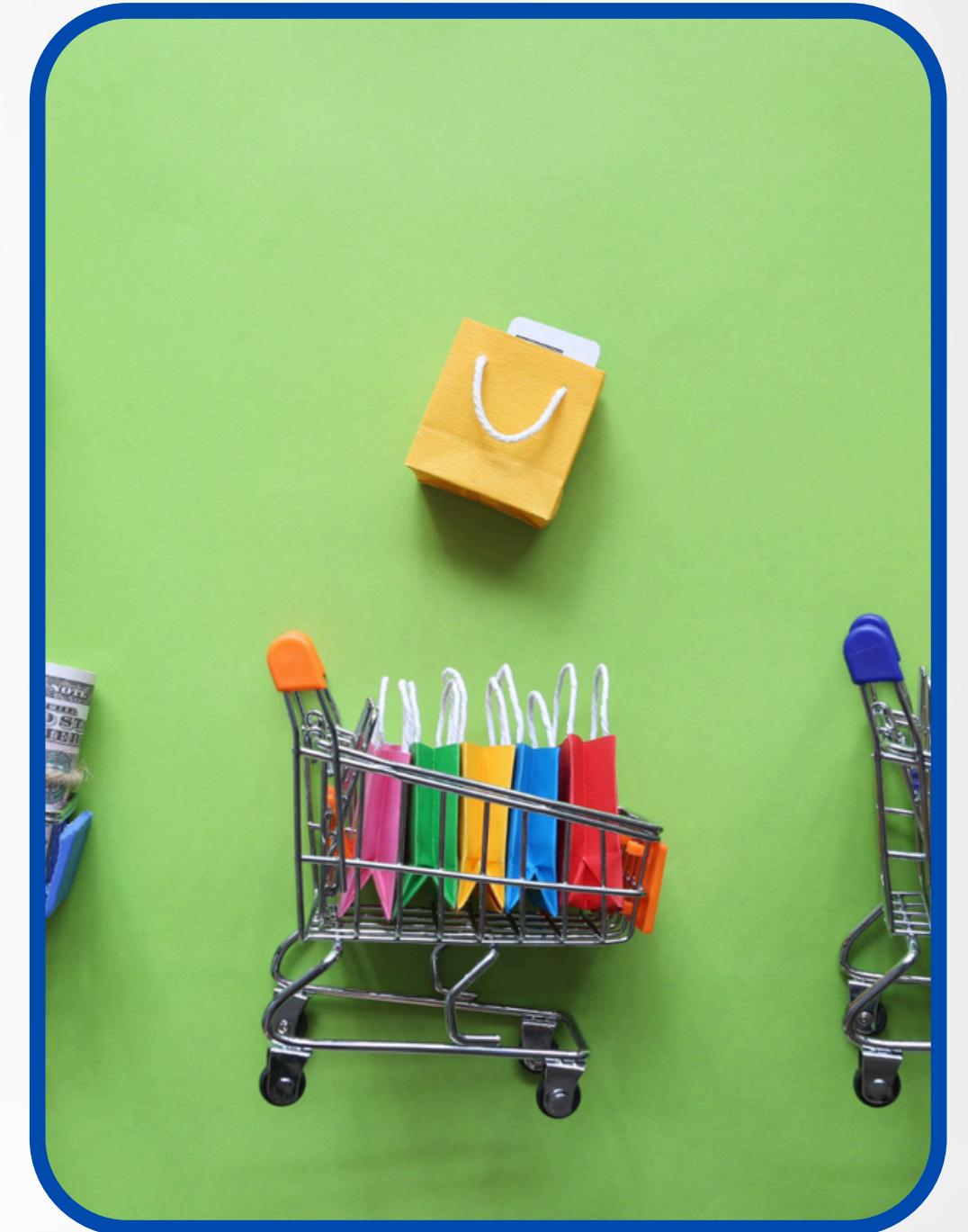
ANALISIS DATA E-COMMERCE UNTUK PERTUMBUHAN BISNIS

Muhammad Farhan Al Hafizh

OVERVIEW DATA

Data yang diberikan merupakan data penjualan di e-commerce RetailKita.

Data terbagi menjadi 2 tabel master yakni tabel customers dan products, serta 1 tabel transaksi yakni tabel transactions



APA TUJUAN ANALISIS?

Tujuan dari analisis data ini adalah untuk mengidentifikasi produk terlaris, pelanggan teratas, dan trend penjualan yang terjadi di e-commerce RetailKita.

Serta untuk meningkatkan retensi pelanggan, hasil dari analisis ini nantinya akan digunakan untuk mengidentifikasi potensi churn pelanggan

ANALISIS YANG DILAKUKAN

1

Top 5 produk terlaris

2

Top 10 pelanggan dengan pembelian tertinggi

3

Trend penjualan

PEMBERSIHAN DATA CUSTOMERS DAN TRANSACTIONS

Hasil dari profiling kualitas seluruh data pada tabel ditemukan bahwa seluruh tabel bersih dari nilai duplikat, tetapi tabel customers (customer_location) dan transactions (quantity) memiliki nilai null

Tindakan yang diambil:

1. Mengisi nilai null pada tabel customer dengan nilai modus.
2. Mengisi nilai null pada tabel transactions dengan nilai mean, agar tidak mengubah sebaran datanya.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 3 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   customer_id      100 non-null    object  
 1   join_date        100 non-null    object  
 2   customer_location 95 non-null    object  
dtypes: object(3)
memory usage: 2.5+ KB
0
{'customer_id': 0, 'join_date': 0, 'customer_location': 5}
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 498 entries, 0 to 497
Data columns (total 5 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   transaction_id  498 non-null    object  
 1   customer_id     498 non-null    object  
 2   product_id      498 non-null    object  
 3   timestamp       498 non-null    object  
 4   quantity        496 non-null    float64 
dtypes: float64(1), object(4)
memory usage: 19.6+ KB
0
{'transaction_id': 0, 'customer_id': 0, 'product_id': 0, 'timestamp': 0, 'quantity': 2}
```

CATATAN PENTING PEMBERSIHAN DATA CUSTOMERS DAN TRANSACTION

- Pengisian nilai NULL pada tabel customer menggunakan nilai modus dilakukan daripada men-drop nilai NULL tersebut walaupun hanya berjumlah (5%), dilakukan karena menghindari munculnya nilai NULL lainnya pada saat penggabungan dataset, dikarenakan customer ID yang hilang saat didrop.
- Pengisian nilai Null pada tabel transaction dengan nilai mean dilakukan agar persebaran datanya tidak berubah drastis.

PENGGABUNGGAN DAN PROSES ETL SEDERHANA

CUSTOMERS, PRODUCTS, DAN TRANSACTIONS

Setelah data diberisihkan, data digabungkan agar menjadi satu tabel analitik.

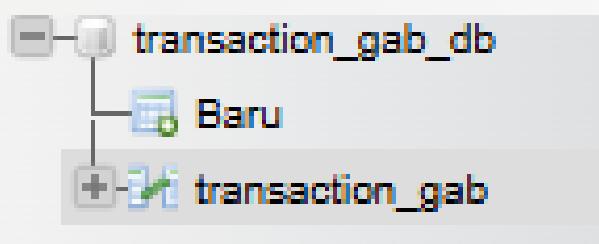
Selain pembersihan data, beberapa proses ETL sederhana juga dilakukan untuk menunjang analisis ini. beberapa proses yang dilakukan adalah sebagai berikut:

- Membuat kolom total_price (price*quantity)
- Menyamakan format tanggal pada kolom timestamp dan join_date menjadi dd/mm/YYYY
- Menambah kolom month berisi tahun-bulan suatu transaksi terjadi (mengantisipasi error pada saat analisis menggunakan kolom timestamp yang diperbarui)

df_merged.head(5)												
	transaction_id	customer_id	product_id	quantity	timestamp_fmt	customer_location	join_date_fmt	product_name	product_category	price	total_price	month
0	TRX0001	CUST001	PROD005	2.0	2023-01-20 10:05:00	Jakarta	2023-01-15	T-Shirt Katun	Pakaian	150000	300000.0	2023-01
1	TRX0002	CUST002	PROD002	1.0	2023-01-22 14:30:00	Surabaya	2023-01-20	Smartphone X	Elektronik	8500000	8500000.0	2023-01
2	TRX0003	CUST001	PROD003	1.0	2023-02-01 09:15:00	Jakarta	2023-01-15	Wireless Mouse	Elektronik	250000	250000.0	2023-02
3	TRX0004	CUST003	PROD011	1.0	2023-02-08 18:45:00	Bandung	2023-02-05	Novel Fiksi "Senja"	Buku	95000	95000.0	2023-02
4	TRX0005	CUST005	PROD001	1.0	2023-03-05 11:20:00	Jakarta	2023-03-01	Laptop Pro 15	Elektronik	15000000	15000000.0	2023-03

LOAD DATASET KE DATABASE LOKAL

Setelah data siap digunakan, saya mengekspor dataset dalam bentuk CSV dan mengimportnya ke database lokal menggunakan MySQL via XAMPP yang saya berinama transaction_gab_db



transaction_id	customer_id	product_id	quantity	timestamp_fmt	customer_location	join_date_fmt	product_name	product_category	price	total_price	month
TRX0031	CUST024	PROD009	1.0	2023-12-22 13:45:00	Jakarta	2023-10-20	Matras Yoga	Olahraga	300000	300000.0	2023-12
TRX0033	CUST025	PROD001	1.0	2024-01-12 09:10:00	Jakarta	2023-11-08	Laptop Pro 15	Elektronik	15000000	15000000.0	2024-01
TRX0029	CUST022	PROD008	1.0	2023-11-20 20:30:00	Surabaya	2023-09-14	Sepatu Lari	Olahraga	750000	750000.0	2023-11
TRX0017	CUST013	PROD012	1.0	2023-07-18 11:10:00	Yogyakarta	2023-05-03	Buku Motivasi "Bangkit"	Buku	120000	120000.0	2023-07
TRX0018	CUST015	PROD002	1.0	2023-08-05 15:30:00	Surabaya	2023-06-01	Smartphone X	Elektronik	8500000	8500000.0	2023-08
TRX0019	CUST002	PROD018	1.0	2023-08-15 18:00:00	Surabaya	2023-01-20	Jaket Hoodie	Pakaian	400000	400000.0	2023-08
TRX0027	CUST005	PROD002	1.0	2023-11-01 11:50:00	Jakarta	2023-03-01	Smartphone X	Elektronik	8500000	8500000.0	2023-11
TRX0021	CUST017	PROD016	1.0	2023-09-10 13:20:00	Semarang	2023-07-04	Sprei Katun Jepang	Rumah Tangga	250000	250000.0	2023-09
TRX0022	CUST016	PROD028	1.0	2023-09-15 16:00:00	Bandung	2023-06-11	Topi Baseball	Pakaian	120000	120000.0	2023-09
TRX0023	CUST001	PROD011	1.0	2023-09-25 19:45:00	Jakarta	2023-01-15	Novel Fiksi "Senja"	Buku	95000	95000.0	2023-09
TRX0498	CUST048	PROD008	1.0	2025-03-01 12:00:00	Surabaya	2024-10-15	Sepatu Lari	Olahraga	750000	750000.0	2025-03
TRX0025	CUST018	PROD004	1.0	2023-10-18 12:30:00	Jakarta	2023-07-21	Mechanical Keyboard	Elektronik	800000	800000.0	2023-10
TRX0014	CUST012	PROD007	1.0	2023-06-22 17:40:00	Medan	2023-04-19	Celana Jeans	Pakaian	450000	450000.0	2023-06
TRX0013	CUST009	PROD014	1.0	2023-06-10 14:05:00	Semarang	2023-03-09	Panci Set Anti Lengket	Rumah Tangga	550000	550000.0	2023-06
TRX0002	CUST002	PROD002	1.0	2023-01-22 14:30:00	Surabaya	2023-01-20	Smartphone X	Elektronik	8500000	8500000.0	2023-01
TRX0003	CUST001	PROD003	1.0	2023-02-01 09:15:00	Jakarta	2023-01-15	Wireless Mouse	Elektronik	250000	250000.0	2023-02
TRX0004	CUST003	PROD011	1.0	2023-02-08 18:45:00	Bandung	2023-02-05	Novel Fiksi "Senja"	Buku	95000	95000.0	2023-02
TRX0005	CUST005	PROD001	1.0	2023-03-05 11:20:00	Jakarta	2023-03-01	Laptop Pro 15	Elektronik	15000000	15000000.0	2023-03
TRX0006	CUST004	PROD008	1.0	2023-03-10 20:00:00	Medan	2023-02-10	Sepatu Lari	Olahraga	750000	750000.0	2023-03
TRX0007	CUST006	PROD009	1.0	2023-03-15 13:10:00	Yogyakarta	2023-03-12	Matras Yoga	Olahraga	300000	300000.0	2023-03
TRX0028	CUST021	PROD027	1.0	2023-11-08 17:10:00	Jakarta	2023-09-02	Webcam HD	Elektronik	450000	450000.0	2023-11
TRX0009	CUST008	PROD006	1.0	2023-04-25 19:22:00	Jakarta	2023-02-19	Kemeja Flanel	Pakaian	350000	350000.0	2023-04
TRX0010	CUST007	PROD015	1.0	2023-05-01 12:00:00	Surabaya	2023-01-28	Blender Multifungsi	Rumah Tangga	400000	400000.0	2023-05
TRX0011	CUST010	PROD017	1.0	2023-05-28 08:30:00	Jakarta	2023-03-25	Headphone Bluetooth	Elektronik	600000	600000.0	2023-05
TRX0012	CUST001	PROD022	1.0	2023-06-02 10:11:00	Jakarta	2023-01-15	Power Bank 10000mAh	Elektronik	300000	300000.0	2023-06

ANALISIS PRODUK TERLARIS

```
SELECT product_name, SUM(total_price) AS total_sales  
  -> FROM transaction_gab  
  -> GROUP BY product_name  
  -> ORDER BY total_sales DESC  
  -> LIMIT 5;
```

product_name	total_sales
Laptop Pro 15	390000000
Smartphone X	238000000
Sepatu Lari	18750000
Mechanical Keyboard	14400000
Tablet 10 inch	13500000

Output yang dihasilkan menunjukkan Top 5 produk terlaris di e-commerce RetailKita. Hasil analisis menunjukkan bahwa Laptop Pro 15 dan Smartphone X memiliki penjualan yang jauh lebih tinggi dibanding tiga produk teratas lainnya.

ANALISIS PELANGGAN PALING BERHARGA

```
SELECT customer_id, SUM(total_price) AS total_purchase
    -> FROM transaction_gab
    -> GROUP BY customer_id
    -> ORDER BY total_purchase DESC
    -> LIMIT 10;
```

Output yang dihasilkan menunjukkan Top 10 pelanggan dengan total pembelian terbanyak.

customer_id	total_purchase
CUST005	111600000
CUST033	90600000
CUST021	62500000
CUST001	56590000
CUST029	52200000
CUST041	51800000
CUST018	50200000
CUST037	31500000
CUST025	31300000
CUST071	30000000

month	total_revenue
2023-01	24668145.161290325
2023-02	33380000
2023-03	53080000
2023-04	4930000
2023-05	7630000
2023-06	10620000
2023-07	14760000
2023-08	39200000
2023-09	19660000
2023-10	22470000
2023-11	43208145.16129032
2023-12	79920000
2024-01	48950000
2024-02	5180000
2024-03	5680000
2024-04	35600000
2024-05	6160000
2024-06	52390000
2024-07	58045000
2024-08	56830000
2024-09	35125000
2024-10	57010000
2024-11	16990000
2024-12	23750000
2025-01	47610000
2025-02	6660000
2025-03	2770000
2025-04	400000
2025-05	120000

29 rows in set (0.047 sec)

ANALISIS PELANGGAN PENJUALAN PER BULAN

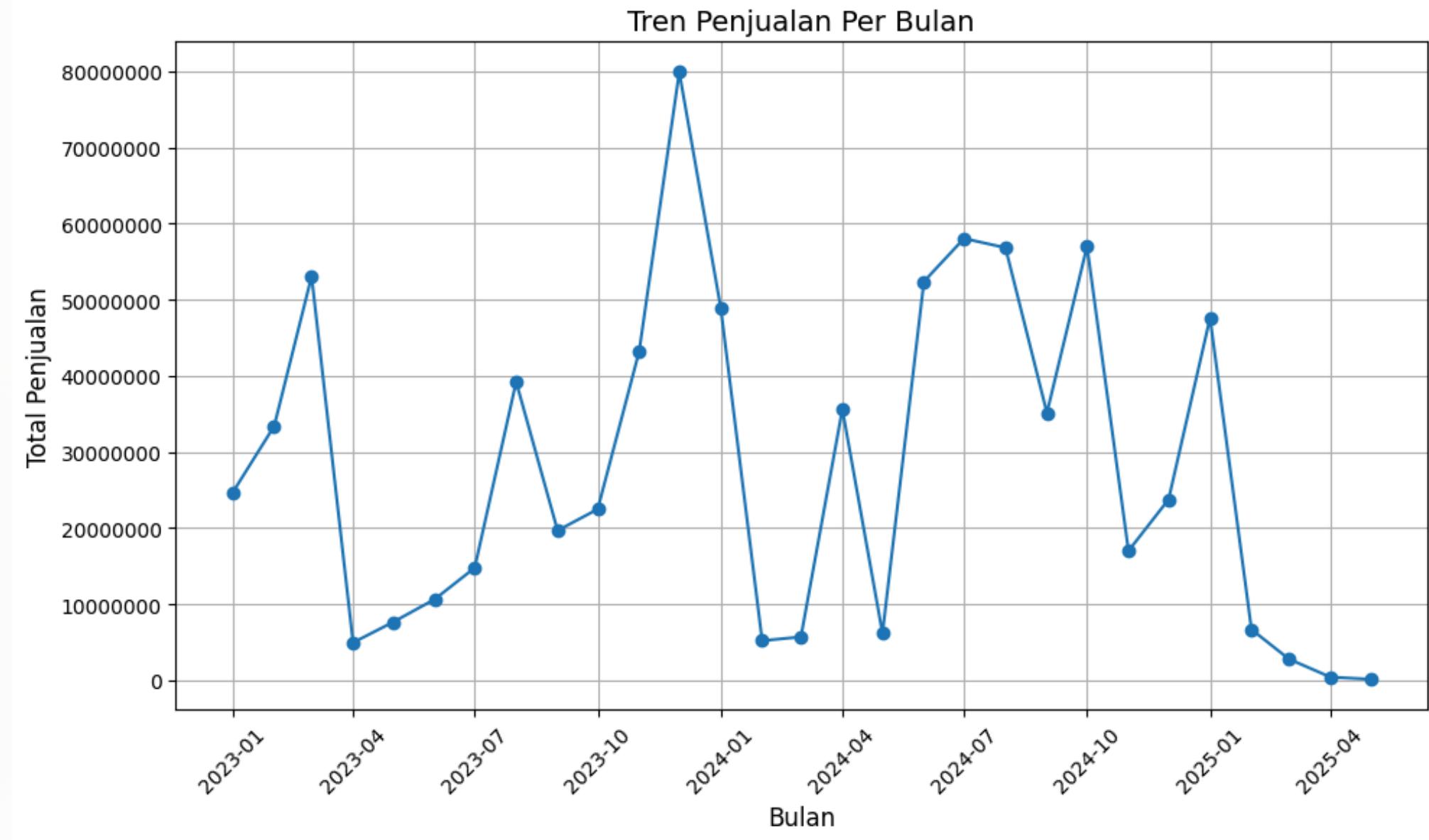
```
SELECT month,
       -> SUM(total_price) AS total_revenue
       -> FROM transaction_gab
       -> GROUP BY month
       -> ORDER BY month ASC;
```

Output yang dihasilkan menunjukkan total revenue yang didapatkan oleh e-commerce RetailKita setiap bulannya.



VISUALISAI TREND PENJUALAN PER BULAN

Visualisasi menunjukkan bahwa secara keseluruhan trend penjualan mengalami fluktuasi. Visualisasi ini juga menunjukkan bahwa di awal Q2 penjualan konsisten menurun dan cenderung rendah. walaupun demikian, puncak tertinggi penjualan terjadi pada bulan desember 2023, dan secara konsisten data tersebut menunjukkan bahwa setiap akhir Quartal penjualan mengalami kenaikan.



IDENTIFIKASI PELANGGAN BERPOTENSI CHURN

Saya mendefinisikan churn sebagai aktivitas berhenti berlangganan, dimana seorang customer tidak melakukan pembelian dalam kurun waktu 3 buan (asumsi) atau memiliki rasio pembelian yang kecil. lalu beberapa fitur yang saya ambil untuk melakukan feature engineering dan analisis ini di antaranya ada customer_id, timestamp_fmt, join_date_fmt, dan total_price. Terakhir sebelum melakukan identifikasi, seluruh data dikelompokkan berdasarkan ID customer.

	customer_id	total_purchase	last_transaction	join_date	active_days	purchase_ratio	cluster	days_since_last_purchase	is_churn_candidate
0	CUST001	5.659000e+07	2025-01-20 19:00:00	2023-01-15	736	76888.586957	1	100	0
1	CUST002	2.955815e+07	2025-01-05 10:00:00	2023-01-20	716	41282.325644	1	116	0
2	CUST003	1.175000e+06	2024-07-15 13:00:00	2023-02-05	526	2233.840304	0	290	1
3	CUST004	4.330000e+06	2024-07-20 15:00:00	2023-02-10	526	8231.939163	0	285	1
4	CUST005	1.116000e+08	2025-01-25 09:00:00	2023-03-01	696	160344.827586	2	96	0



MENGHADIRKAN FITUR BARU

Dari kolom-kolom yang telah dipilih sebelumnya, saya membuat fitur-fitur baru untuk mengidentifikasi potensi churn pelanggan. berikut merupakan fitur yang ditambahkan:

- **total_purchase** : total pembelian yang telah dilakukan pelanggan (total_price.sum())
- **last_transaction** : transaksi terakhir (timestamp_fmt.max())
- **active_days** : lama seorang pelanggan aktif berlangganan (join_date - last_transaction)
- **purchase_ratio** : rasio pembelian yang dilakukan pelanggan per harinya (total_purchase / active_days)
- **days_since_last_purchase** : jumlah hari sejak pembelian terakhir pelanggan (timestamp_fmt.max()-last_transaction)

MODEL ML YANG DIGUNAKAN DECISION TREE

Penggunaan model dilakukan karena Fitur yang menggambarkan potensi seorang pelanggan melakukan churn (is_churn_candidate) dikalkulasi dengan sedikit rumit. jadi walaupun data nya sedikit, tetap diperlukan model klasifikasi yang robust

KMEANS

Penggunaan model dilakukan untuk membuat sebuah fitur baru yang mengelompokkan customer berdasarkan 'purchase_ratio', 'total_purchase', 'active_days' secara unsupervised.

CATATAN PENTING IDENTIFIKASI POTENSI CHURN PELANGGAN

- Pembuatan purchase_ratio dilakukan untuk melihat aktivitas belanja seorang pelanggan selama berlangganan. karena pelanggan yang terindikasi churn akan memiliki nilai rasio yang rendah.
- Pelanggan yang berpotensi churn, diasumsikan tidak aktif melakukan pembelian selama lebih dari 3 bulan (180 hari).
- Pembuatan fitur days_since_last_purchase dilakukan untuk melihat apakah pelanggan masih aktif berbelanja dalam kurun waktu 180 hari dari hari ini (transaksi terakhir pada kolom timestamp_fmt)

```
new_agg_df['is_churn_candidate'] = (
    (new_agg_df['purchase_ratio'] < new_agg_df['purchase_ratio'].quantile(0.25)) |
    (new_agg_df['days_since_last_purchase'] > 180) # misal lebih dari 3 bulan gak beli
).astype(int)
```

THANK YOU

VISIT US



+123-456-7890



www.reallygreatsite.com