

Terra Store

Product Recommendation

Written by Revaldi

BUSINESS CONTEXT

Terra Store is looking to enhance its marketing strategy by predicting customer purchase behavior based on historical data. The company wants to build an AI-powered application that can provide insights into which products a customer is likely to purchase next.

GUIDELINE

- 1 **Data Pipeline**
- 2 **Data Exploration & Preprocessing**
- 3 **Model Development**
 - Random Forest
 - XGBoost
- 4 **Web Development**
- 5 **API**

DATA PIPELINE

Data syntetics with faker library

```
fake = Faker()
categories = product_detail['category'].unique()

# Generate purchase_history data
purchase_history_data = []
for _ in range(100):
    purchase_history_data.append({
        'customer_id': random.randint(1, 20),
        'product_id': random.randint(1, 105),
        'purchase_date': fake.date_between(start_date='-1y', end_date='today')
    })

purchase_history = pd.DataFrame(purchase_history_data)

# Generate product_detail data
product_detail_data = []
product_ids = range(1, 106) # generate 100 product_id unik from 1 - 100
for product_id in product_ids:
    product_detail_data.append({
        'product_id': product_id,
        'category': random.choice(categories),
        'price': random.randint(10, 1000),
        'ratings': round(random.uniform(1, 5), 1)
    })

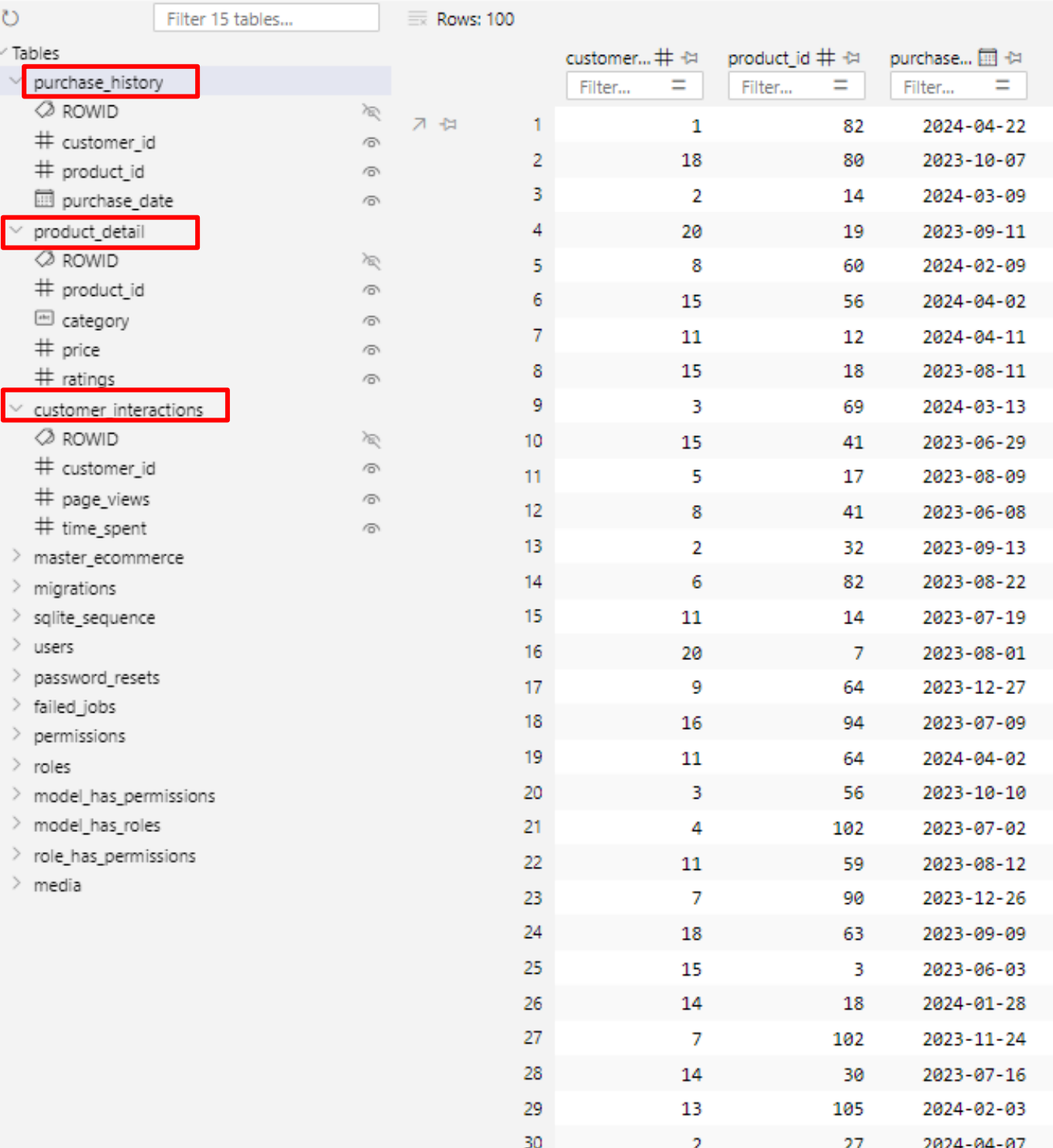
product_detail = pd.DataFrame(product_detail_data)

# Generate customer_interactions data
customer_interactions_data = []
for _ in range(100):
    customer_interactions_data.append({
        'customer_id': random.randint(1, 20),
        'page_views': random.randint(1, 50),
        'time_spent': random.randint(30, 300)
    })

customer_interactions = pd.DataFrame(customer_interactions_data)

return purchase_history, product_detail, customer_interactions
```

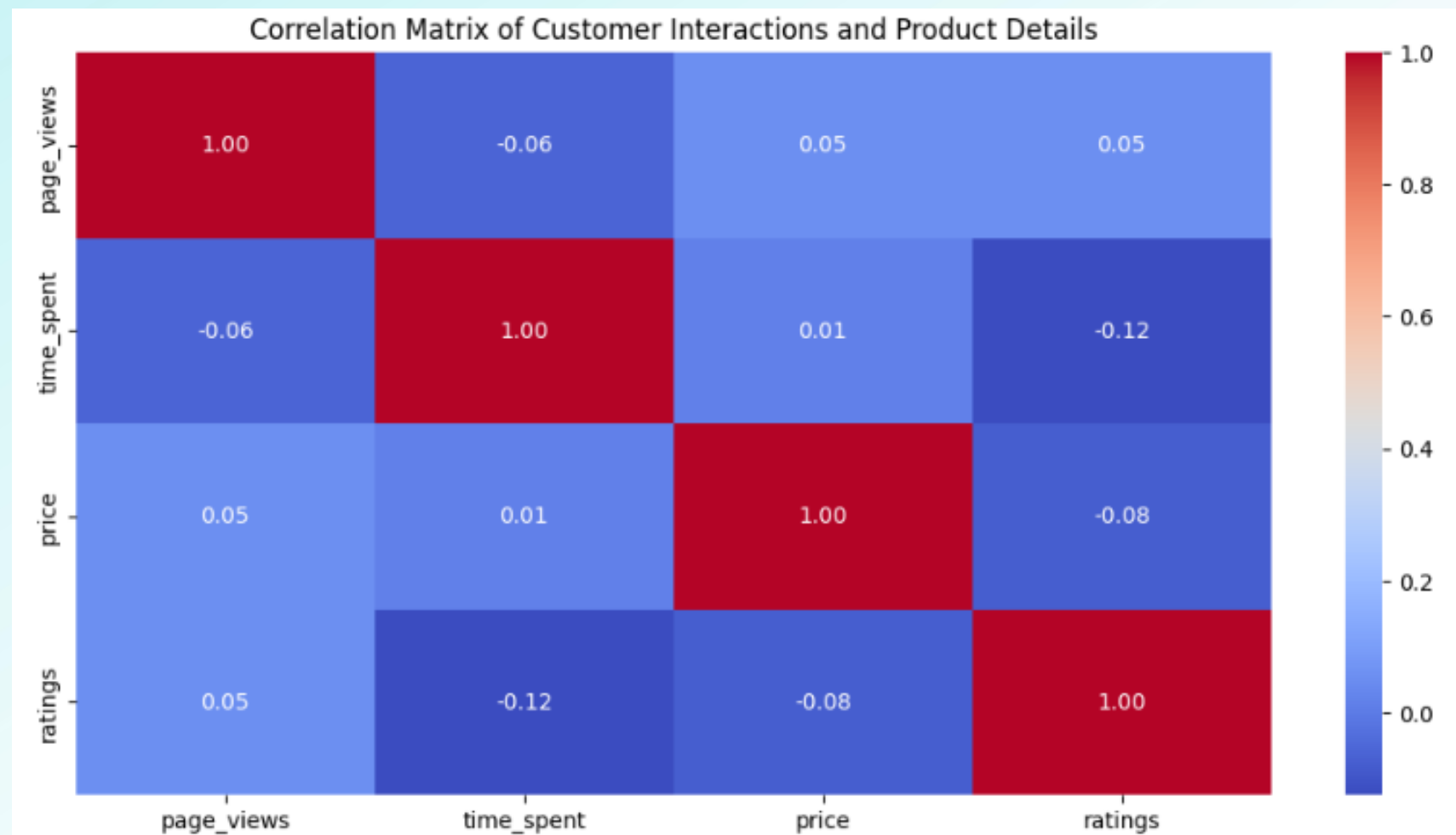
save to sqlite3



Filter 15 tables... Rows: 100

Tables	customer... #	product_id #	purchase... #
	Filter...	Filter...	Filter...
purchase_history			
ROWID			
customer_id	1	82	2024-04-22
product_id	18	80	2023-10-07
purchase_date	2	14	2024-03-09
product_detail			
ROWID			
product_id	20	19	2023-09-11
category	8	60	2024-02-09
price	15	56	2024-04-02
ratings	11	12	2024-04-11
customer_interactions			
ROWID			
customer_id	15	41	2023-06-29
page_views	5	17	2023-08-09
time_spent	8	41	2023-06-08
master_ecommerce			
migrations			
sqlite_sequence			
users			
password_resets			
failed_jobs			
permissions			
roles			
model_has_permissions			
model_has_roles			
role_has_permissions			
media			

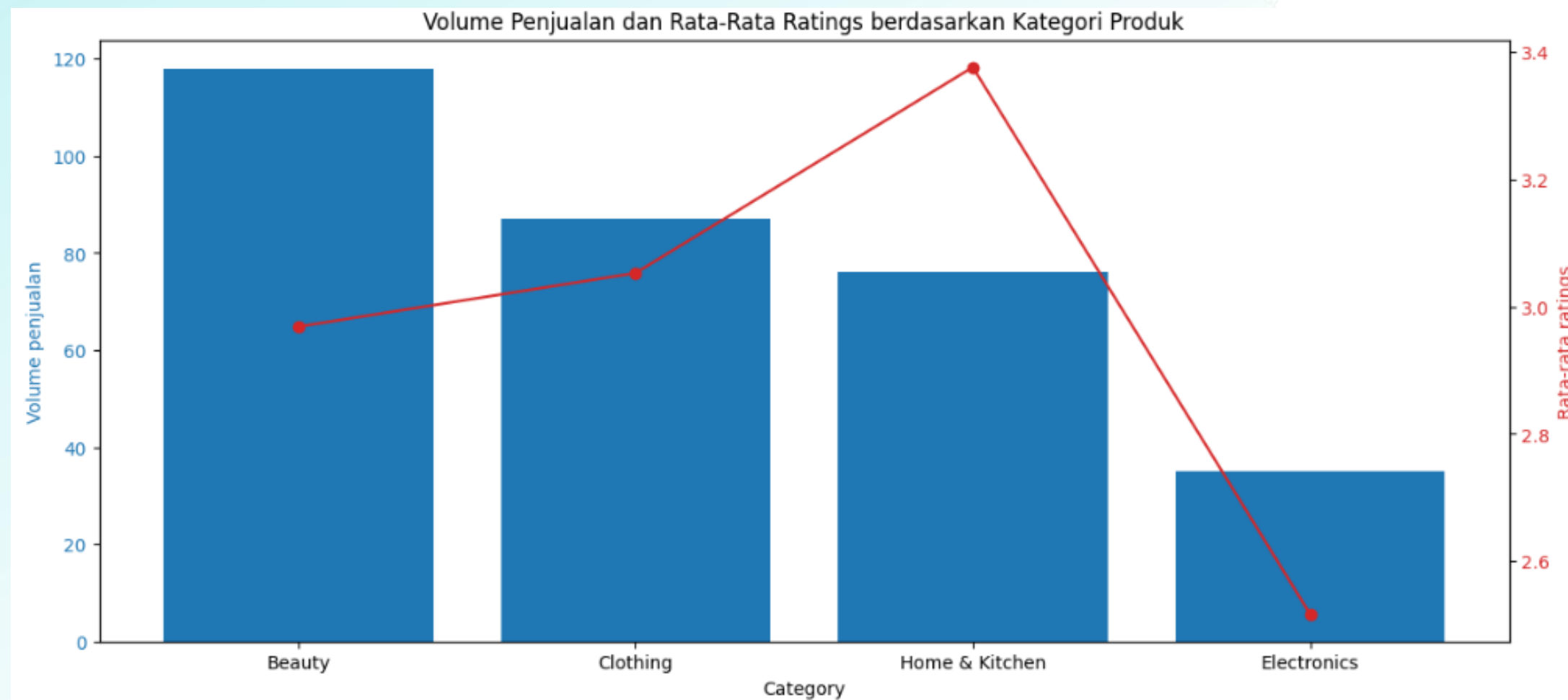
EDA & PREPROCESSING



- **page_views** dan **time_spent**: Ada korelasi positif, yang menunjukkan bahwa user yang melihat lebih banyak halaman cenderung menghabiskan lebih banyak waktu di situs. Hal ini menunjukkan bahwa konten yang menarik berpotensi meningkatkan waktu penelusuran.
- **page_views** dan **price**: Ada korelasi yang negatif, yang menunjukkan bahwa variabel tersebut sangat lemah atau bahkan tidak ada hubungan linear yang signifikan di antara keduanya. Peningkatan atau penurunan jumlah page_views tidak berdampak secara konsisten pada price, dan sebaliknya.
- **time_spent** dan **price**: time_spent yang dihabiskan di situs menunjukkan korelasi yang lemah dengan harga produk yang dibeli.
- **ratings**: page_views dan spent_time korelasinya sangat lemah dengan ratings, yang mungkin menunjukkan bahwa keputusan user untuk melihat halaman atau spent_time tidak terlalu dipengaruhi oleh peringkat produk.

note: korelasi mendekati 1 menunjukkan korelasi positif. Korelasi mendekati -1 menunjukkan tidak adanya hubungan linear antara dua variabel.

EDA & PREPROCESSING



Karna ini pakai data dummy jadi seperti ada anomali, kategori yang punya volume penjualan terbanyak malah memiliki rata2 rating paling kecil.

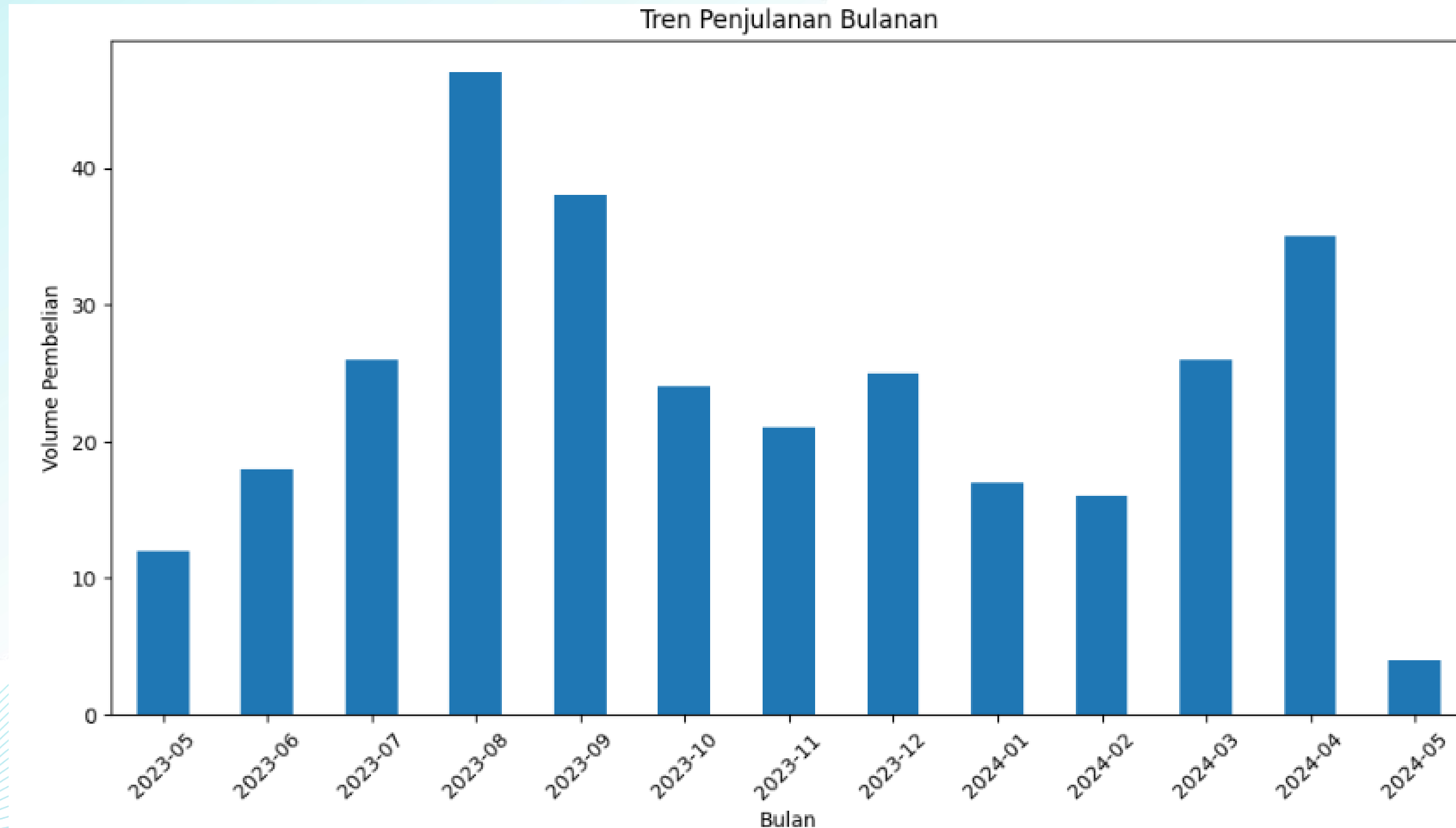
Pada real case nya rata2 rating mempengaruhi keputusan pembelian.

****Insight****

1. **Category performance:** Kategori dengan penjualan lebih tinggi dan rating yang lebih baik mungkin diprioritaskan untuk keputusan dilakukan promosi. Sebaliknya, kategori dengan penjualan dan peringkat lebih rendah mungkin memerlukan penyesuaian pemasaran atau peningkatan kualitas produk.

2. **Strategi pemasaran:** Memahami kategori mana yang punya kinerja baik dalam hal penjualan dan kepuasan pelanggan bisa membantu menyesuaikan strategi pemasaran.

EDA & PREPROCESSING



selama data cutoff 1 tahun, penjualan paling rendah ada di bulan april 2024, dan paling tinggi di agustus 2023.
tren ini bisa digunakan untuk analisis tren kedepannya saat pick session.

EDA & PREPROCESSING

	category	product_id	Sales_Volume	Average_Price	Average_Ratings
0	Home & Kitchen	41	14	315.0	4.7
1	Home & Kitchen	90	8	971.0	1.6
2	Home & Kitchen	39	7	941.0	2.4
3	Home & Kitchen	31	6	309.0	1.5
4	Home & Kitchen	53	5	174.0	1.7
5	Home & Kitchen	81	4	665.0	2.3
6	Home & Kitchen	32	3	203.0	1.4
7	Home & Kitchen	33	3	838.0	2.5
8	Home & Kitchen	47	3	357.0	3.4
9	Home & Kitchen	73	3	375.0	2.6

INSIGHT EDA

****INSIGHT****

1. Penjualan Tertinggi dan Rating Tertinggi: Home 102 dan Home & Kitchen 88 menunjukkan kinerja yang baik berdasarkan volume penjualan yang tinggi dan tingkat kepuasan pelanggan.
2. Sensitivitas Harga: Produk dengan harga lebih tinggi seperti Home & Kitchen 14 menunjukkan volume penjualan yang sangat rendah, menunjukkan sensitivitas harga dalam kategori ini. bisa di analisis lanjutan feedback dari pelanggan
3. Rating: Home & Kitchen 88, ini penjualannya paling tinggi dan rating yang tinggi juga, yang bisa menjadi area untuk memperbanyak pemasaran.

****REKOMENDASI****

1. Strategi Penetapan Harga: Untuk barang dengan harga tinggi dengan penjualan lebih rendah, bisa dipertimbangkan untuk menyesuaikan harga atau meningkatkan kualitas dan fitur produk untuk meningkatkan rating dan kepuasan pelanggan.
2. Meningkatkan Kualitas Produk atau Layanan kepada Pelanggan: Untuk produk dengan rating yang rendah, teliti keluhan atau feedback dari pelanggan untuk meningkatkan kualitas produk.
3. Area Fokus Promosi: Produk yang berkinerja baik bisa dijadikan kampanye pemasaran untuk meningkatkan penjualan yang lebih masif.

MODEL DEVELOPMENT

Metode Machine Learning yang digunakan pada Ecommerce Terra Store ini adalah
Random Forest & XGBoost

MODEL TESTING & EVALUATION:

Teknik pengukuran dalam model testing & evaluation:

- Training Accuracy dan Loss
- Validation Accuracy dan Loss
- Confusion Matrix

BEST MODEL

```
best_model = get_best_model()
```

```
=====
Best model      : XGBClassifier
Validation accuracy: 0.9550561797752809
Best model params : {'n_estimators': 25}
=====
```

Confusion Matrix:

```
[[26  0  0  0]
 [ 0 21  0  0]
 [ 1  0 25  0]
 [ 1  0  1 36]]
```

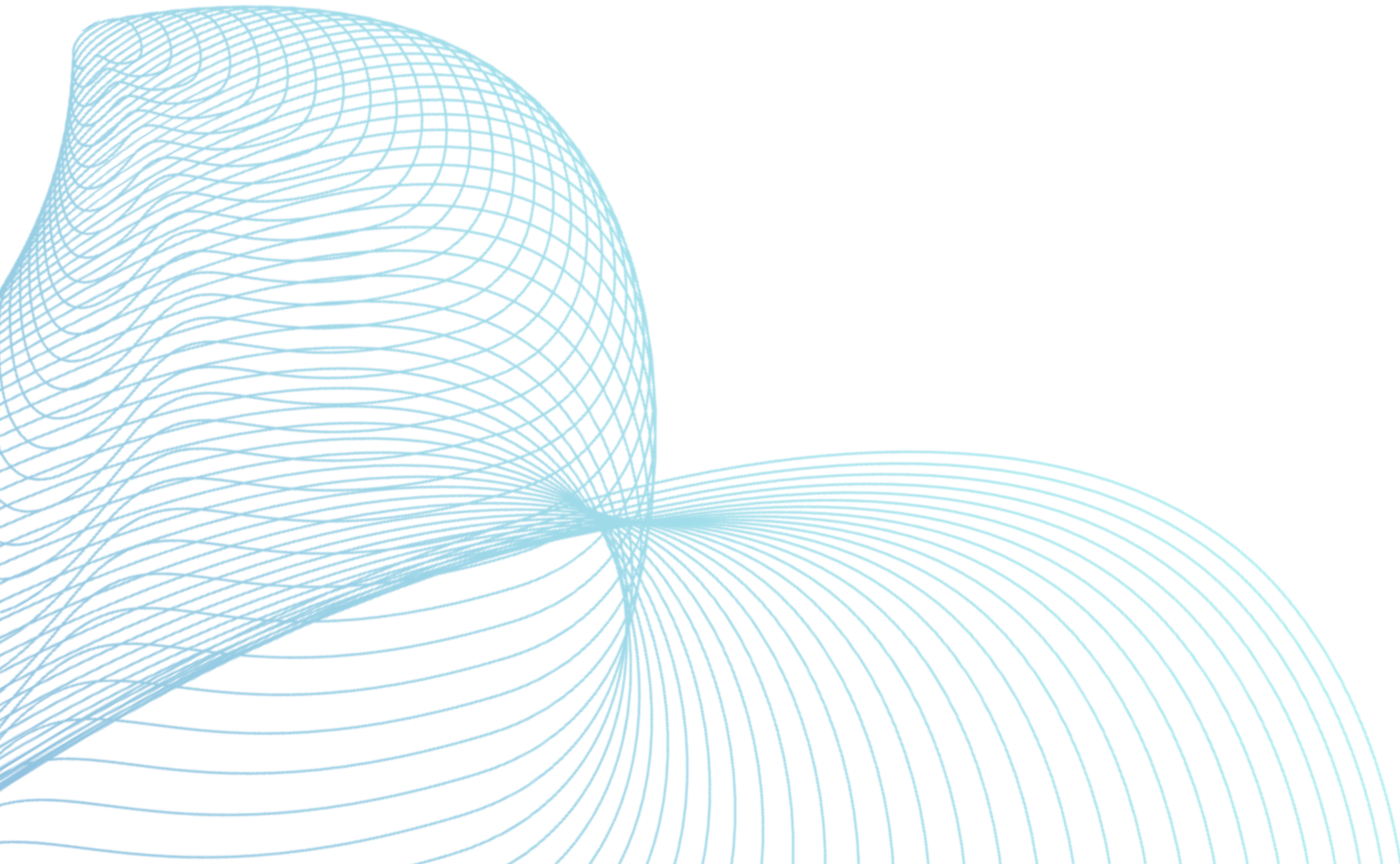
Accuracy: 0.972972972972973

	user_id	y_pred	y_test
231	11	0	0
374	6	3	3
55	16	1	1
381	15	2	3
70	13	3	3
...
11	19	0	3
281	9	3	3
22	7	2	2
375	4	3	3
477	19	2	2

111 rows × 3 columns

model Inference

API



WEB DEVELOPMENT



Dashboard

Pengaturan User

SuperAdmin

Dashboard

Predictive Analytics for Terra Store

Home > Dashboard

32

Stock Beauty

View statistics

124

Stock Electronics

View statistics

45

Stock Clothing

View statistics

45

Stock Home & Kitchen

View statistics

SALES TREND

This Week

Last Week

1400

1200

1000

800

600

400

200

0

MON

TUE

WED

THU

FRI

SAT

SUN

570

Total Orders

2.2%

\$500

Total Earnings

4.2%

264

9.3%

PRODUCT RECOMMENDATION

🔍
🧪 Filters ▾

Search by cusomer id..



CUSTOMER_ID	PRODUCTS
1	Clothing
3	Beauty, Electronics
4	Home & Kitchen
12	Home & Kitchen
45	Beauty
2	Cloths
6	Home & Kitchen, Cloths
20	Beauty
21	Home & Kitchen

Prev
 1
2
3
4
 Next

<https://github.com/revaldianggara/predictive-analytics-for-ecommerce>