



**Kampus
Merdeka**
INDONESIA JAYA

MSIB
Meningkatkan Inovasi Berkeadilan

ibimbing

BUILDING DATA PLATFORM FOR ONLINE RETAIL COMPANY

FINAL PROJECT



TEAM 6



Elsa Rahma Hidayani



Gabriel Khoirul Huga



Yara Rivayanti



Gadis Syahrani Elhakim



Indah Monica Dewi



Latar Belakang

Topik ini melibatkan pengelolaan data secara efisien dalam suatu organisasi. Data pipeline memainkan peran kunci dalam mentransfer, mengolah, dan menyimpan data dari berbagai sumber. Entity-Relationship Diagrams (ERD) membantu memodelkan hubungan antara entitas dalam basis data.

Tools ETL (Extract, Transform, Load) digunakan untuk mengambil, mengubah, dan memuat data ke dalam sistem. Ini penting untuk memastikan konsistensi dan integritas data. Data modeling melibatkan perancangan struktur data untuk memahami dan mengorganisir informasi. Sementara itu, data visualization memanfaatkan grafik dan visualisasi untuk menyajikan data dengan cara yang mudah dimengerti.

Pengintegrasian dan penerapan efektif dari elemen-elemen ini mendukung pengambilan keputusan yang lebih baik dan membantu organisasi mengoptimalkan penggunaan data mereka.



Tujuan

Proyek ini dibentuk untuk merangkum, menyimpan, dan menganalisis data dari berbagai sumber guna memberikan wawasan mendalam tentang operasi bisnis. Dalam konteks ini, kami akan:

- Mengidentifikasi beragam segmen pelanggan berdasarkan perilaku pembelian mereka.
- Mengevaluasi kinerja produk dengan memperhatikan aspek penjualan dan popularitasnya.
- Memantau serta menganalisis aktivitas login pelanggan untuk pemahaman yang lebih baik.
- Memantau proses pemenuhan pesanan dan melacak status pesanan dengan cermat.
- Menilai efektivitas kampanye kupon untuk meningkatkan strategi pemasaran.
- Mengevaluasi kinerja pemasok dalam hal kualitas produk dan ketepatan pengiriman.

Dengan pendekatan ini, diharapkan proyek ini dapat dilaksanakan secara efektif dan efisien, memberikan kontribusi maksimal terhadap pemahaman mendalam tentang berbagai aspek operasional bisnis yang sedang berlangsung.



Kampus
Merdeka
INDONESIA JAYA

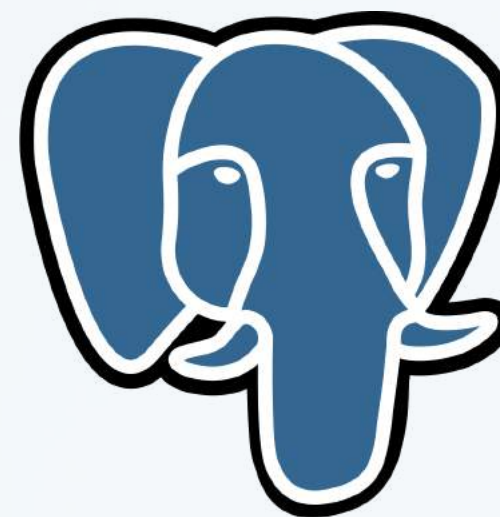
MSIB

ibimbing

TOOLS



Apache
Airflow



Metabase



Polaris

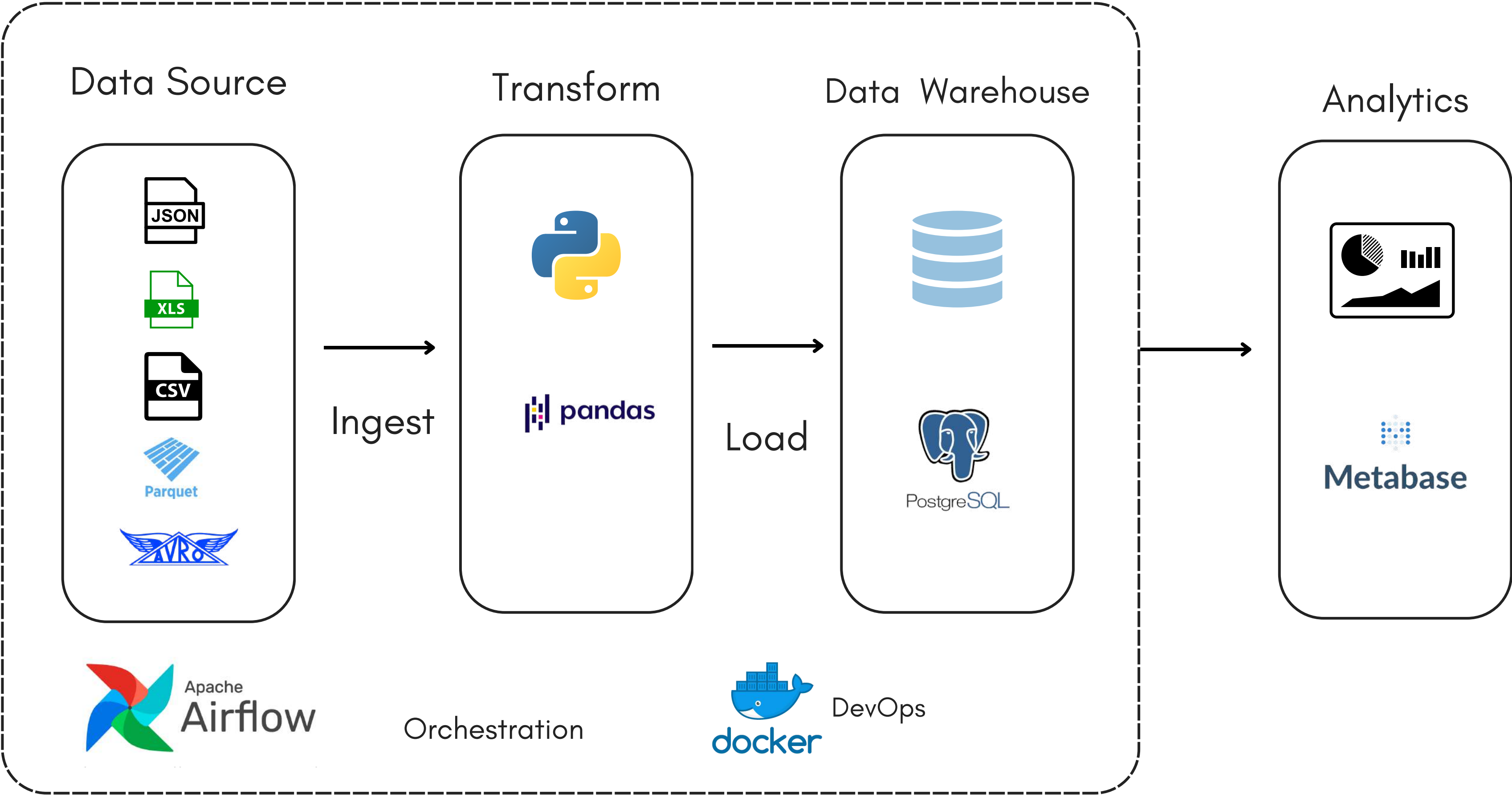


pandas

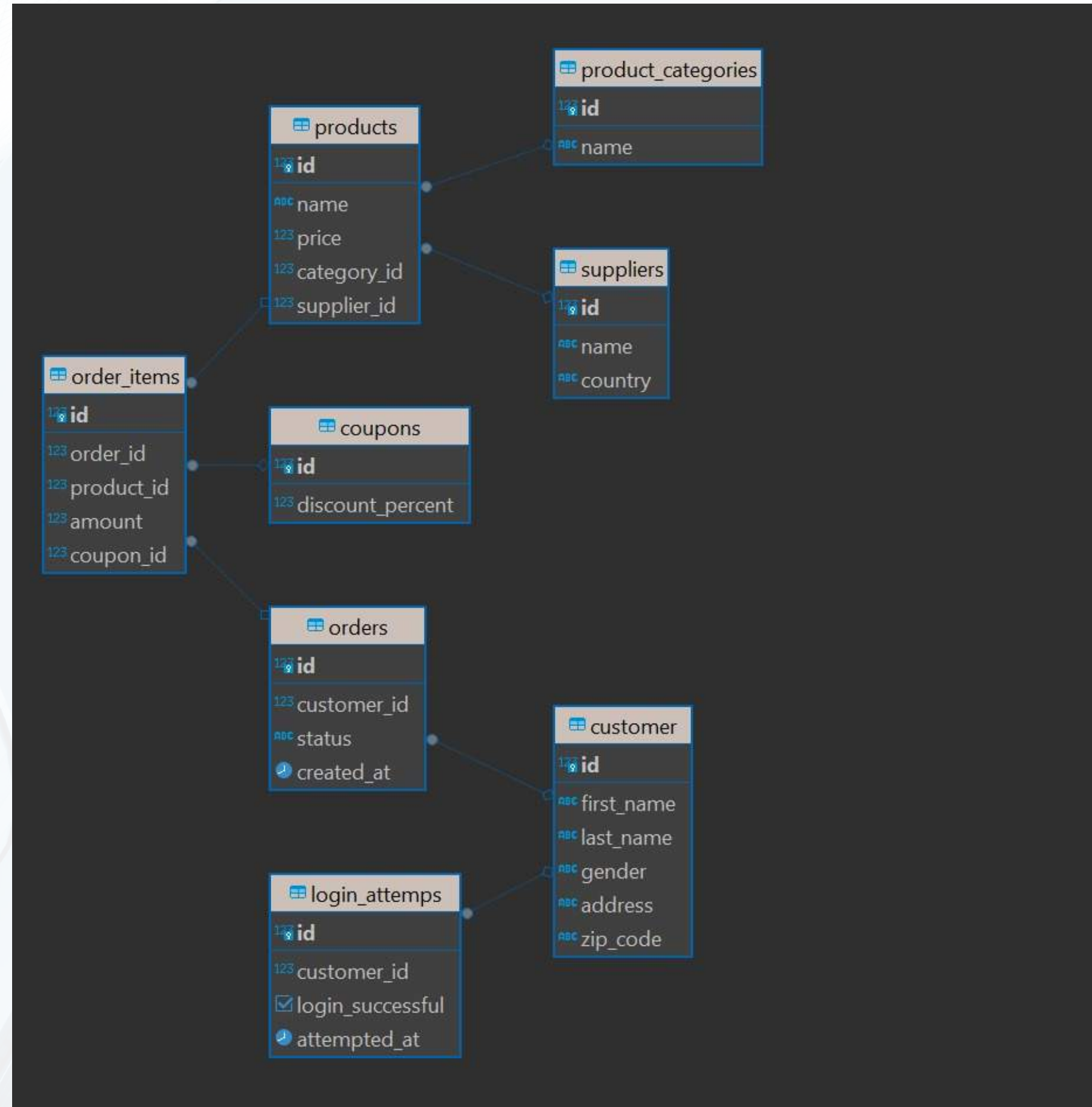


docker

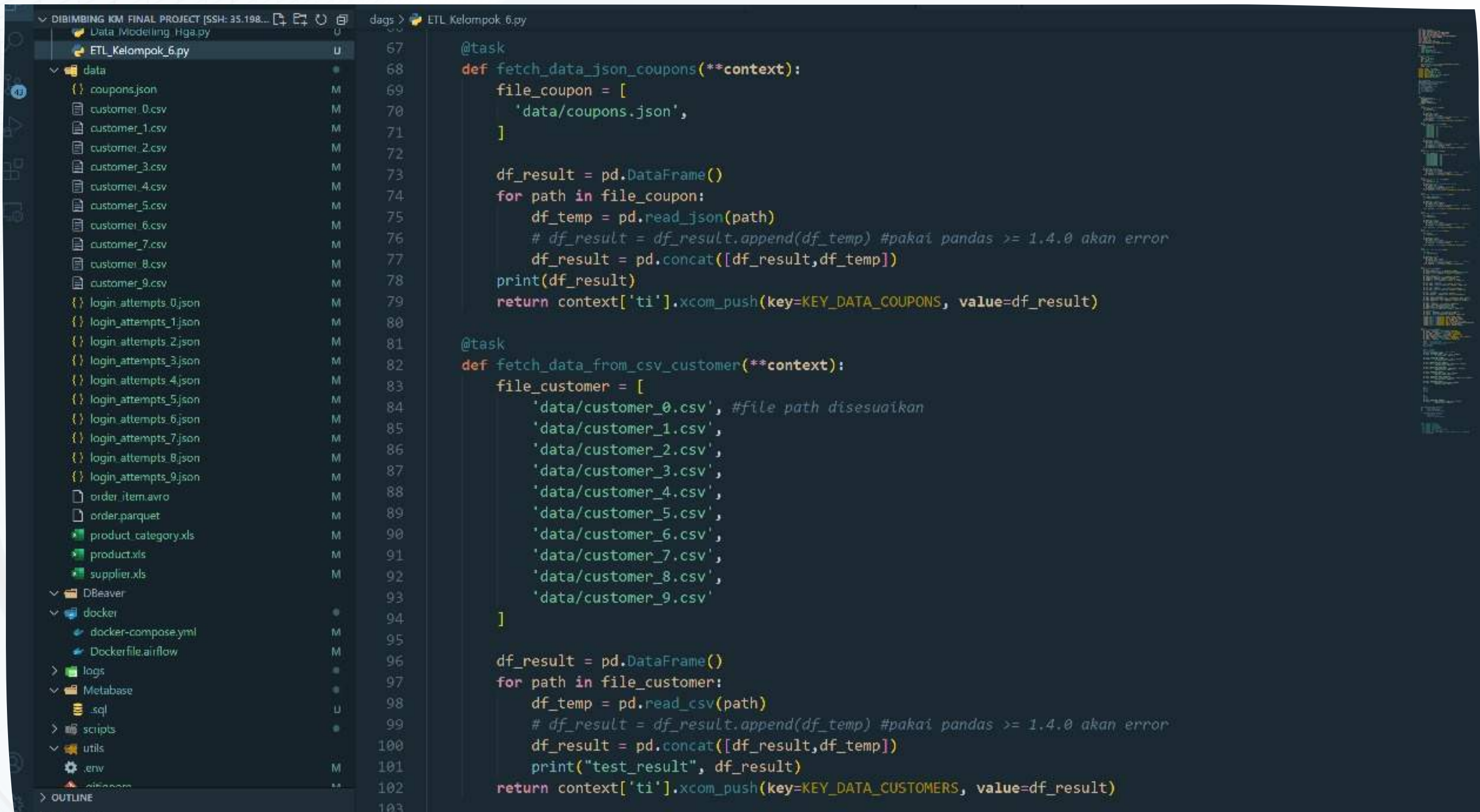
Data Platform Architecture



ERD



ETL



```

@task
def fetch_data_json_coupons(**context):
    file_coupon = [
        'data/coupons.json',
    ]

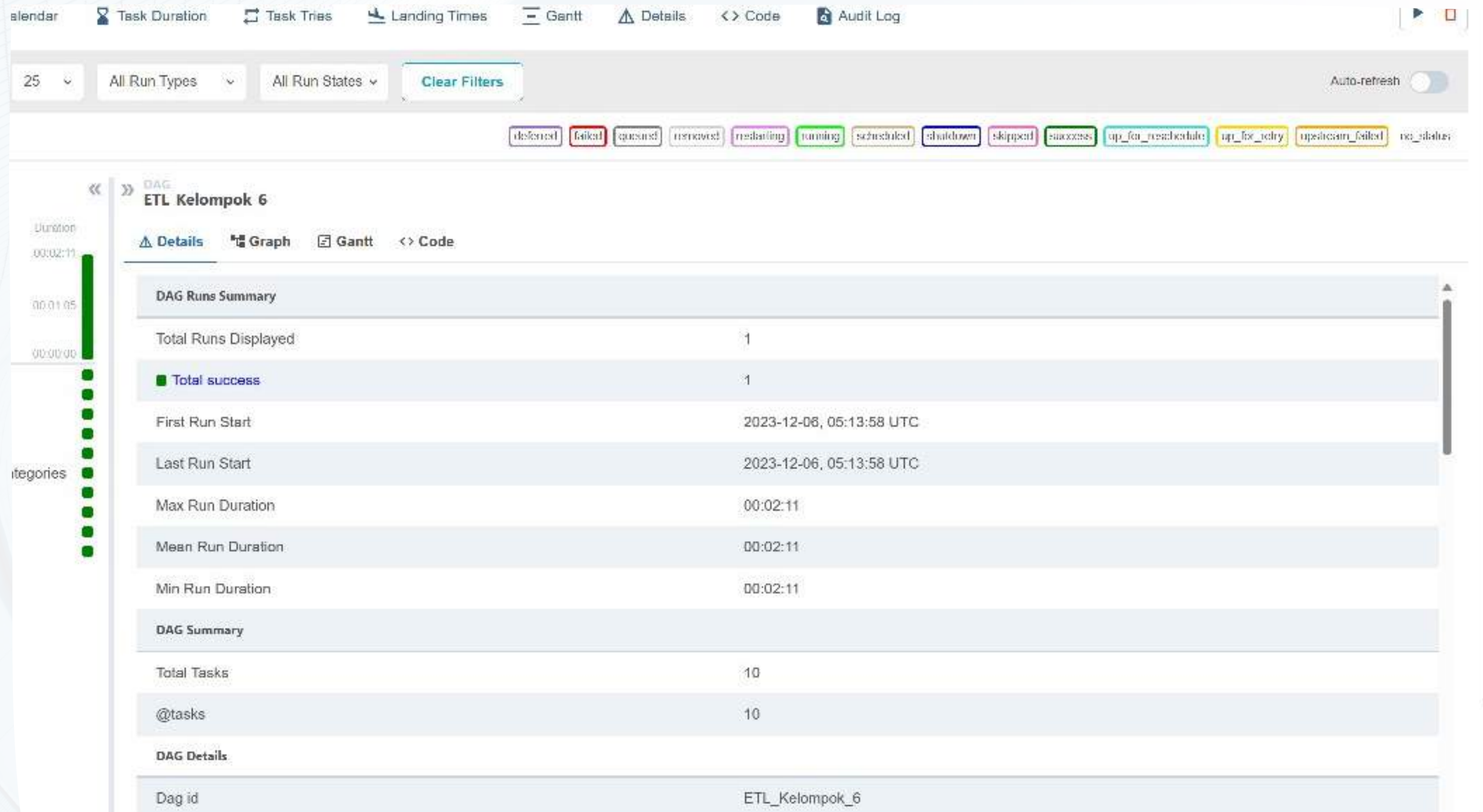
    df_result = pd.DataFrame()
    for path in file_coupon:
        df_temp = pd.read_json(path)
        # df_result = df_result.append(df_temp) #pakai pandas >= 1.4.0 akan error
        df_result = pd.concat([df_result, df_temp])
    print(df_result)
    return context['ti'].xcom_push(key=KEY_DATA_COUPONS, value=df_result)

@task
def fetch_data_from_csv_customer(**context):
    file_customer = [
        'data/customer_0.csv', #file path disesuaikan
        'data/customer_1.csv',
        'data/customer_2.csv',
        'data/customer_3.csv',
        'data/customer_4.csv',
        'data/customer_5.csv',
        'data/customer_6.csv',
        'data/customer_7.csv',
        'data/customer_8.csv',
        'data/customer_9.csv'
    ]

    df_result = pd.DataFrame()
    for path in file_customer:
        df_temp = pd.read_csv(path)
        # df_result = df_result.append(df_temp) #pakai pandas >= 1.4.0 akan error
        df_result = pd.concat([df_result, df_temp])
    print("test_result", df_result)
    return context['ti'].xcom_push(key=KEY_DATA_CUSTOMERS, value=df_result)

```


ETL











ETL

Diagram

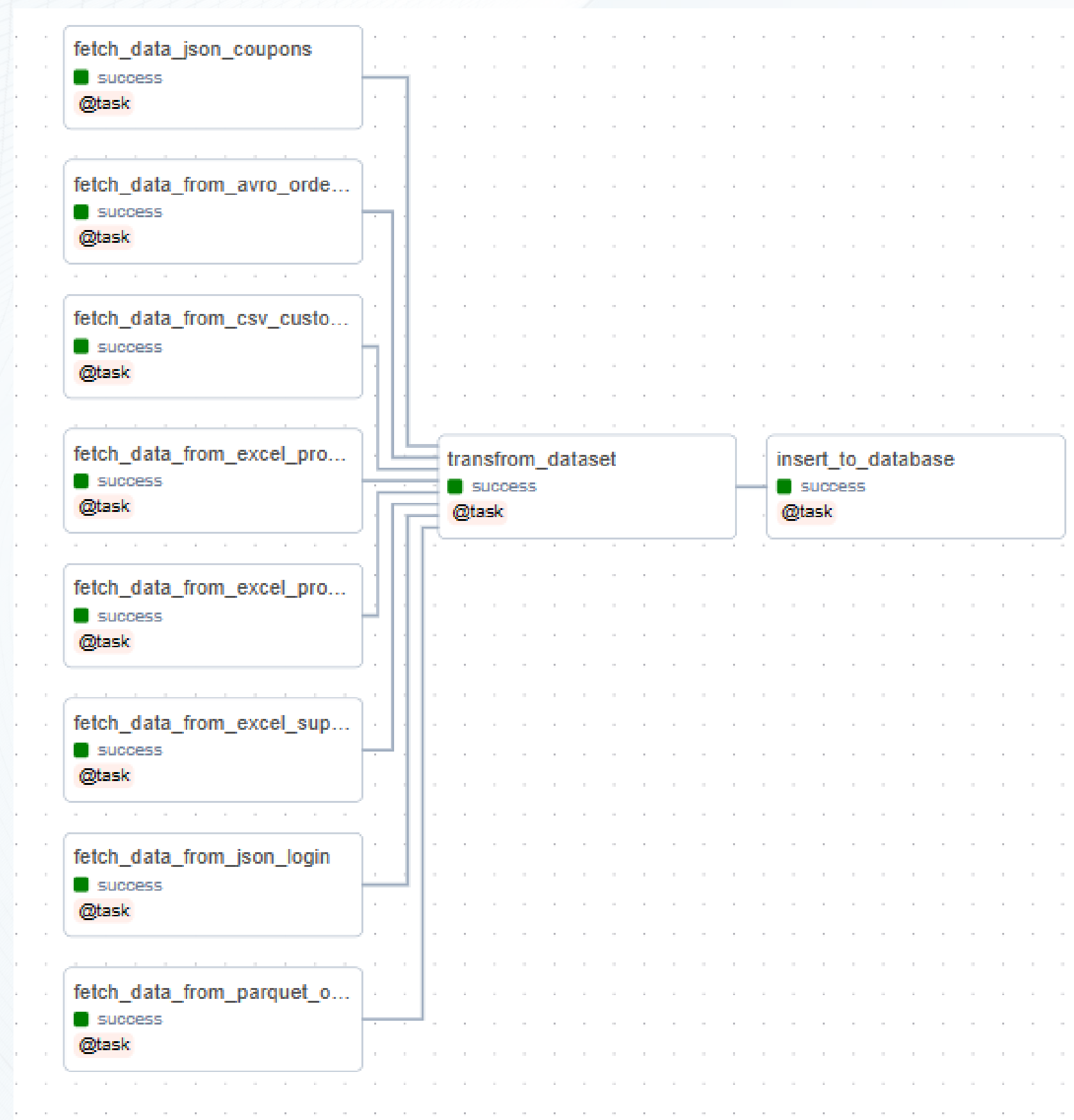
data_warehouse Databases data_warehouse Schemas public

Namespace ID: 2200

Owner: [pg_database_owner](#)

Table Name	Object ID	Owner	Tablespace	Row Count Estimate	Has Row-Level Security	Partitions	Partition by	Extra
 coupons	17,123	user	pg_default	-1	[]	[]		
 customer	17,126	user	pg_default	10,000	[]	[]		
 login_attempts	17,131	user	pg_default	1,000,000	[]	[]		
 order_items	17,134	user	pg_default	1,000,000	[]	[]		
 orders	17,141	user	pg_default	100,000	[]	[]		
 product_categories	17,146	user	pg_default	-1	[]	[]		
 products	17,151	user	pg_default	6,072	[]	[]		
 suppliers	17,156	user	pg_default	157	[]	[]		

Graph



Data Modelling

Dim Table

```
1
2
3 CREATE TABLE "Fact and Dim Tables".customer_dim (
4     customer_id integer PRIMARY KEY,
5     first_name varchar,
6     last_name varchar,
7     address varchar,
8     gender varchar,
9     zip_code varchar
10 );
11
12 -- Assuming "id" is actually "customer_id" in the customers table
13 -- Insert data into "Fact and Dim Tables".customer_dim from the customers table
14 INSERT INTO "Fact and Dim Tables".customer_dim (customer_id, first_name, last_name, address, gender, zip_code)
15 SELECT
16     id as customer_id,
17     first_name,
18     last_name,
19     address,
20     gender,
21     zip_code
22 FROM
23     "Data Modelling Metabase".customer;
24
25 CREATE TABLE "Fact and Dim Tables".product_dim (
26     product_id integer PRIMARY KEY,
27     name varchar, -- Corrected column name from "names" to "name"
28     price float,
29     category_id integer,
30     supplier_id integer
31 );
```


Data Modelling

Dim Fact

```
4 CREATE TABLE "Fact and Dim Tables".sales (  
5   id serial PRIMARY KEY,  
6   order_id integer,  
7   product_id integer,  
8   customer_id VARCHAR, -- Assuming customer_id is a string, adjust the type accordingly  
9   supplier_id integer,  
10  category_id integer,  
11  coupon_id varchar (255),  
12  amount integer,  
13  price float,  
14  status text,  
15  created_at date  
16 );  
17  
18 -- Insert data into the fact.sales table using a SELECT statement  
19 INSERT INTO "Fact and Dim Tables".sales (  
20     order_id,  
21     product_id,  
22     customer_id,  
23     supplier_id,  
24     category_id,  
25     coupon_id,  
26     amount,  
27     price,  
28     status,  
29     created_at  
30 )
```

```
31 SELECT  
32     o.id as order_id,  
33     oi.product_id,  
34     c.id as customer_id,  
35     p.supplier_id,  
36     p.category_id,  
37     oi.coupon_id,  
38     oi.amount,  
39     p.price,  
40     o.status,  
41     o.created_at as date  
42 FROM  
43     "Data Modelling Metabase".customer c  
44     JOIN orders o ON c.id = o.customer_id  
45     JOIN order_items oi ON o.id = oi.order_id  
46     JOIN products p ON oi.product_id = p.id  
47     JOIN suppliers s ON p.supplier_id = s.id  
48     JOIN login_attempts l ON c.id = l.customer_id;  
49
```


Data Visualization

Coupon Effectiveness Analysis:

Objective: Evaluate the effectiveness of coupon campaigns

Data Sources :

- Coupons for coupon information.
- Order_items for coupon usage in orders.

Analytics Queries :

- Calculate the impact of coupons on total revenue.
- Assess the popularity of different types of coupons.
- Measure the conversion rate of orders with and without coupons.
- Display coupon performance metrics on the dashboard

Data Visualization

Customer Segmentation

Objective : Identify different customer segments based on their purchasing behavior.

Data Sources :

- Customers for customer information.
- Orders for order history.

Analytics Queries :

- Aggregate total spending per customer.
- Categorize customers into segments (e.g., high spenders, frequent shoppers, one-time buyers).
- Visualize customer segments on the dashboard.

Data Visualization

Order Fulfillment Metrics

Objective : Monitor order fulfillment and track order status

Data Sources :

- Orders for order information

Analytics Queries :

- Track the status distribution of orders (returned).
- Analyze return rate for each supplier.
- Identify bottlenecks in the fulfillment process.
- Visualize order status trends on the dashboard.

Data Visualization

Product Performance Analysis

Objective : Evaluate the performance of products in terms of sales and popularity.

Data Sources :

- fact.products for product information.
- fact.order_items for order details.

Analytics Queries :

- Calculate total revenue generated by each product.
- Identify best-selling products.
- Analyze product sales trends over time.
- Display top-performing products on the dashboard.

Data Visualization

Supplier Performance Metrics

Objective : Assess the performance of suppliers in terms of product quality and delivery.

Data Sources :

- fact.suppliers for supplier information.
- fact.products for product-supplier relationships

Analytics Queries :

- Evaluate supplier reliability based on order fulfillment.
- Monitor product return rates by supplier.
- Identify top-performing suppliers.
- Provide supplier performance scores on the dashboard.

Data Visualization

Login Activity Monitoring

Objective : Monitor and analyze customer login activity.

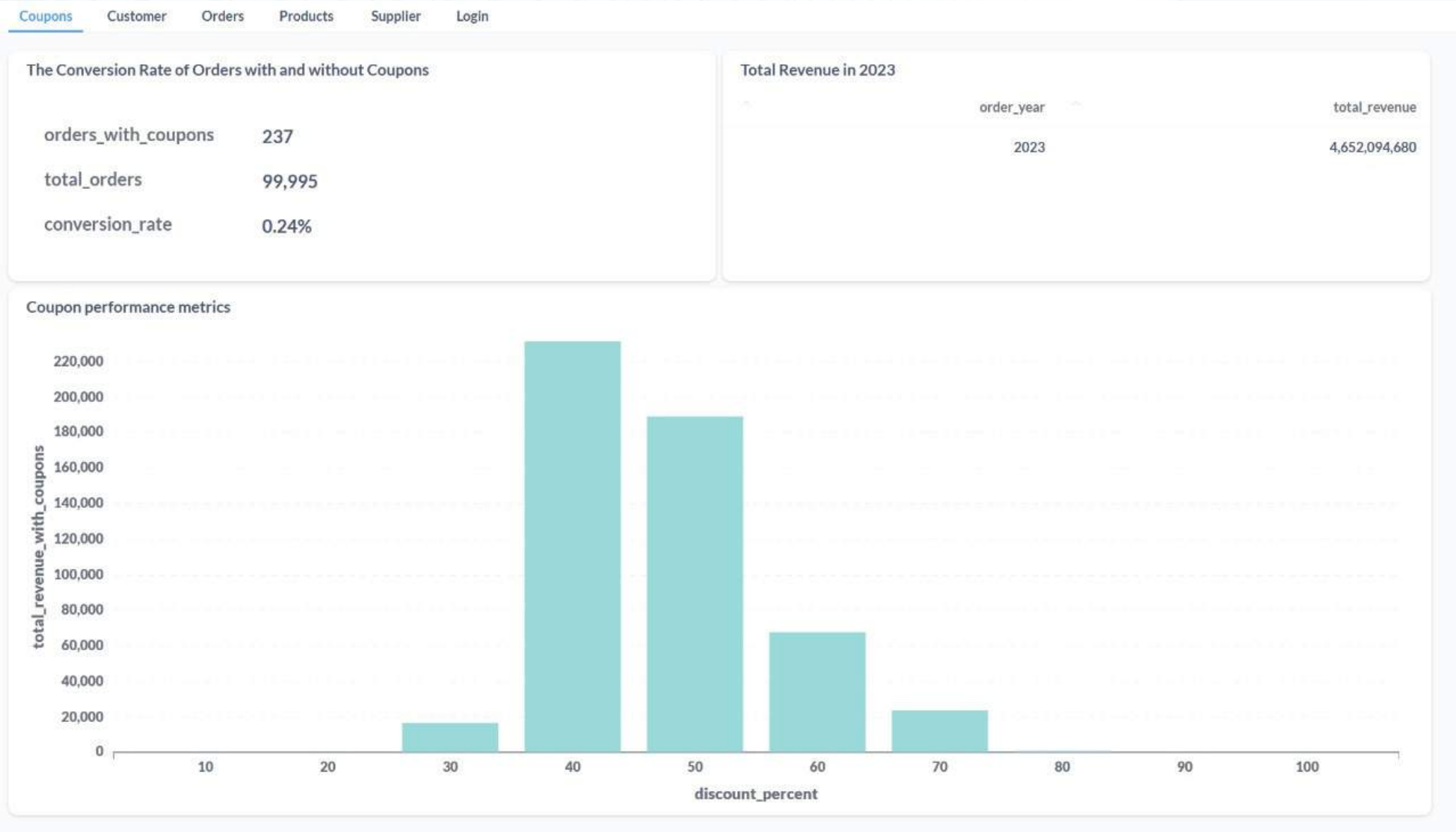
Data Sources :

- fact.login_attempt_history for login attempts.

Analytics Queries :

- Count successful and unsuccessful login attempts.
- Identify login patterns (e.g., peak login hours).
- Highlight any unusual login activity.
- Provide insights into account security on the dashboard

Data Visualization



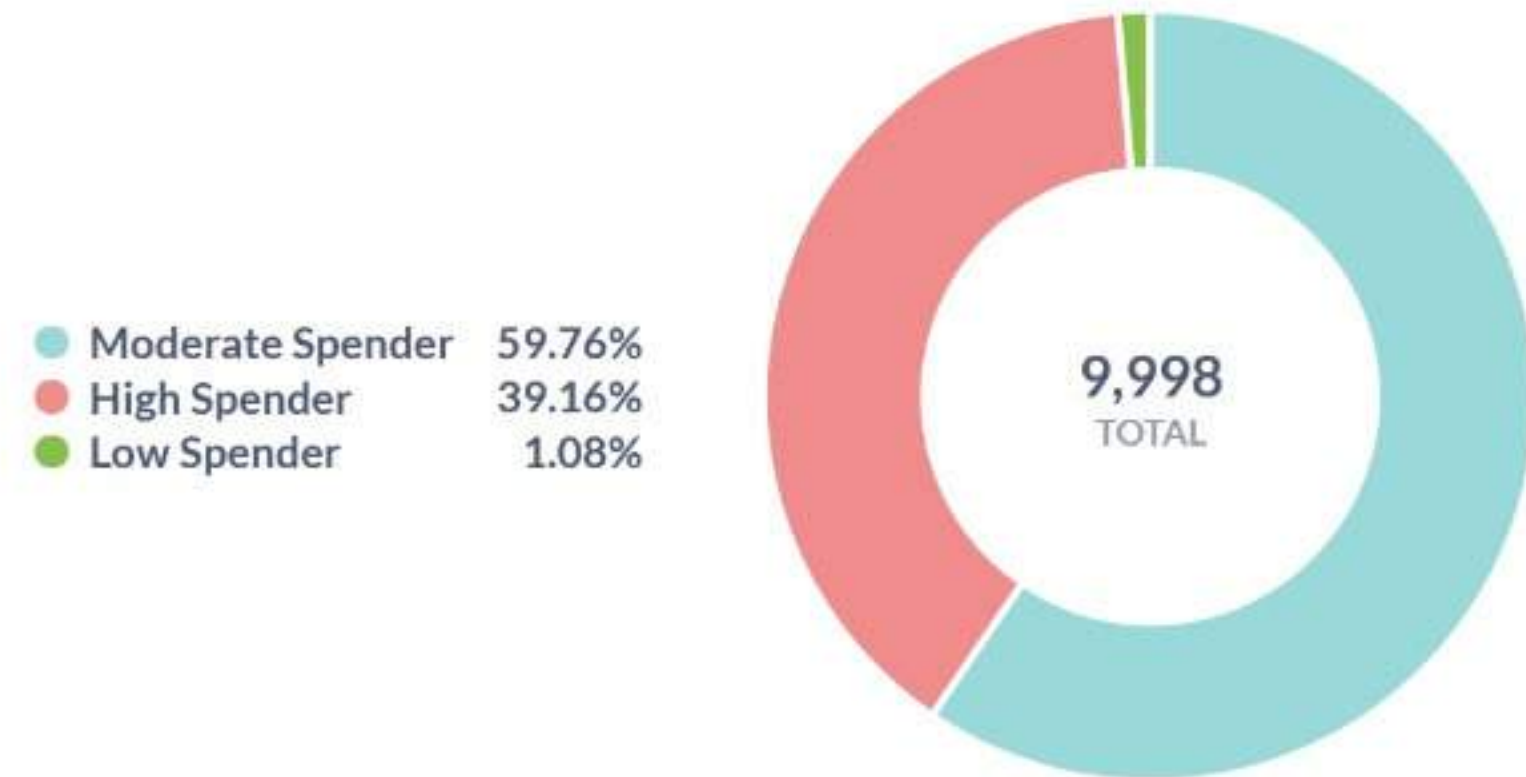
Data Visualization

Coupons Customer Orders Products Supplier Login

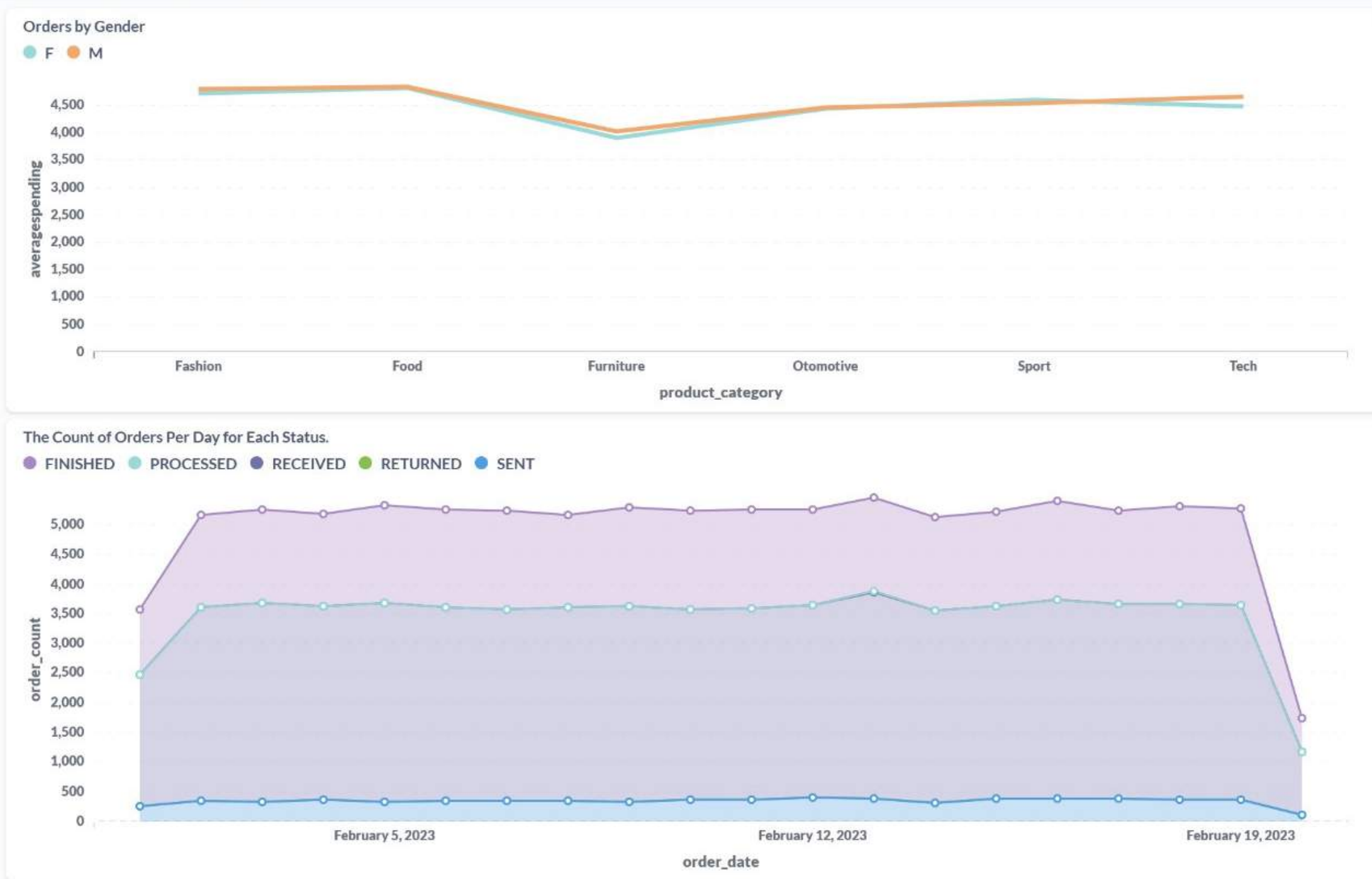
Total Customer by Gender



Categorize customers into segments



Data Visualization



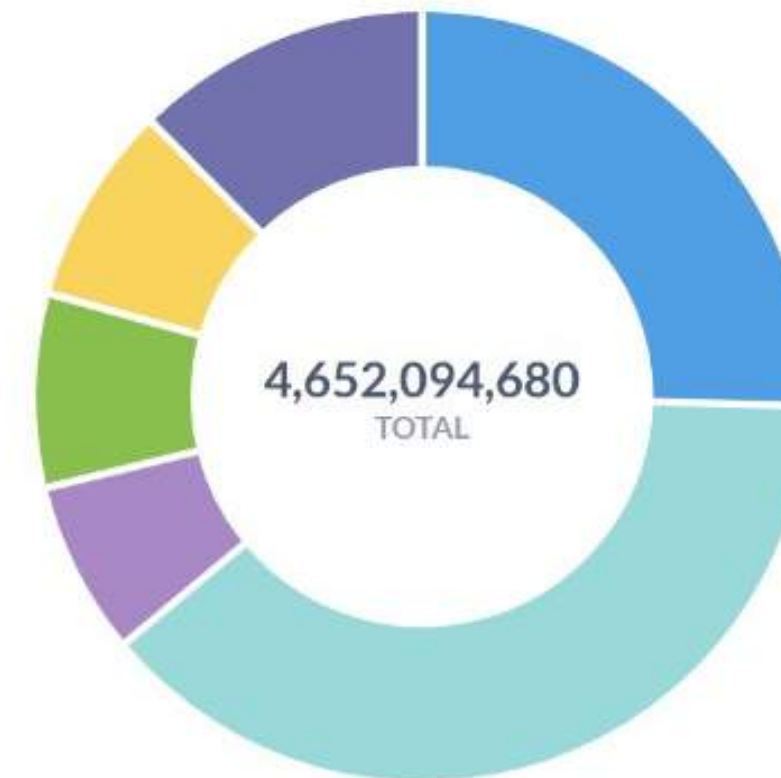
Data Visualization

Sales by Product Price

product_price	total_amount_sold
5	2,877,546
10	5,764,018
15	8,820,388
25	11,734,081
50	8,744,353
75	6,155,477
100	2,965,645
1,000	2,958,966

Sales by Product Category

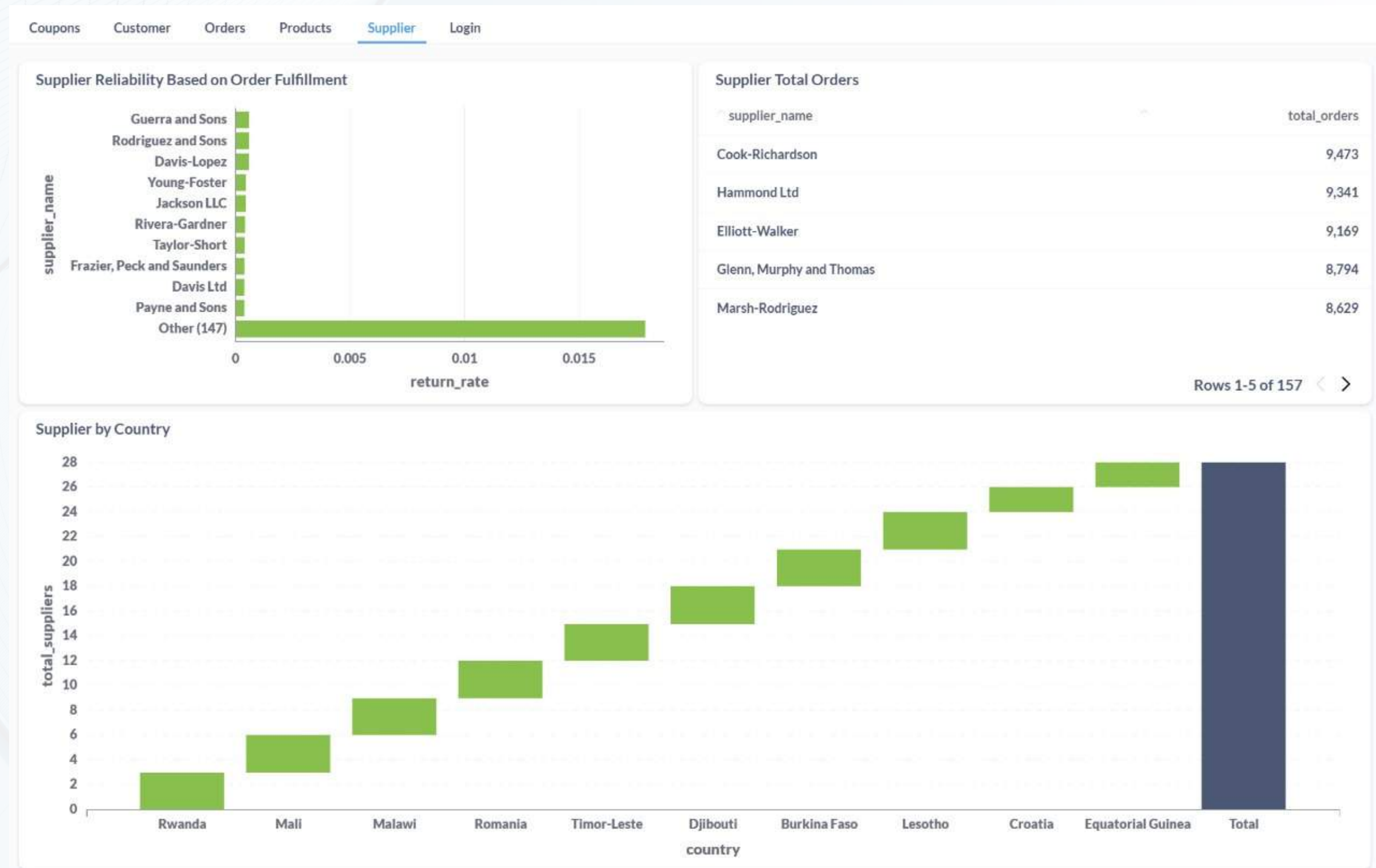
Fashion	25.52%
Food	38.95%
Furniture	7.09%
Otomotive	7.96%
Sport	8.21%
Tech	12.28%



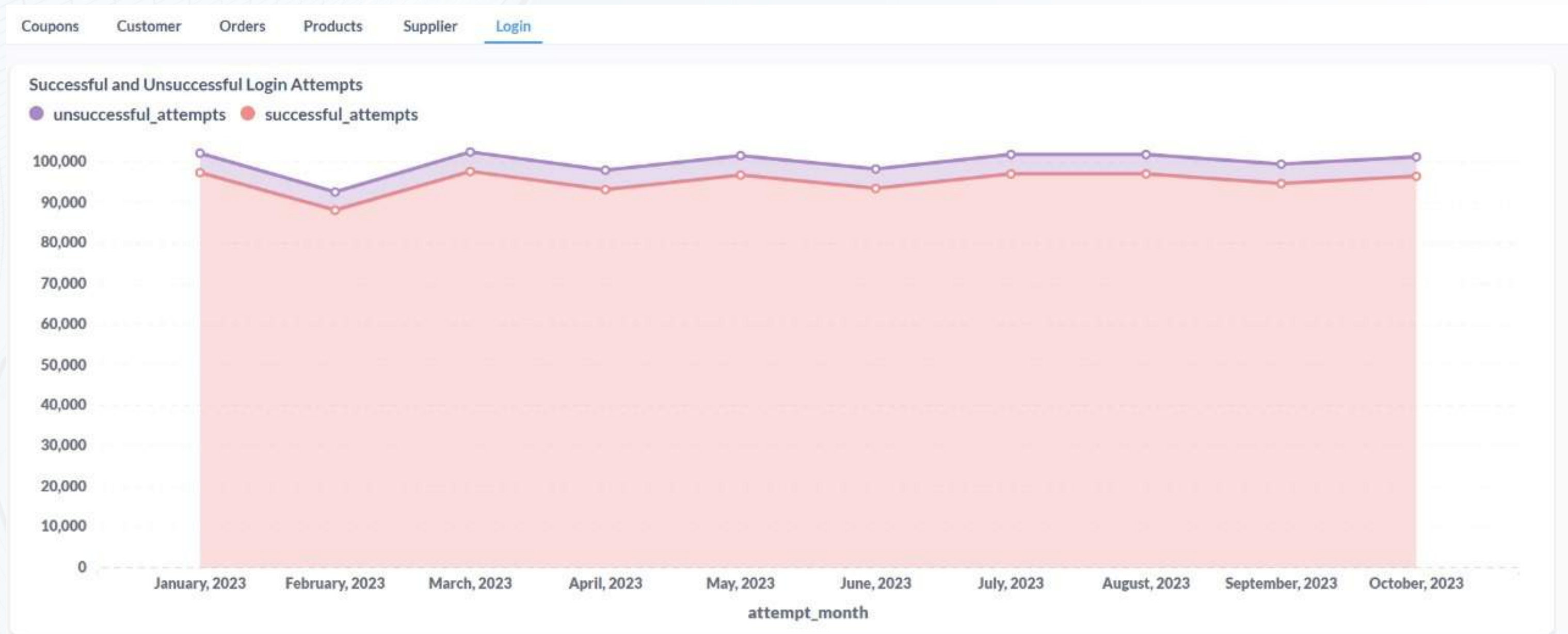
4,652,094,680

Total Revenue in the Last Year

Data Visualization



Data Visualization



Sesi Tanya Jawab

10 Menit



TERIMA KASIH

*Atas Waktu dan Kesediaan Anda Untuk
Mendengarkan Presentasi dari Kami.*

