Final Project

Kelompok 7

"Hiduplah seolah engkau mati besok. Belajarlah seolah engkau hidup selamanya."

Dibimbing X Kampus Merdeka

MEET OUR TEAM

Animni Fiddaroini

UNIVERSITAS INDRAPRASTA PGRI

Alya Nabila

UNIVERSITAS DIPONEGORO

Fathur Rahman S.

UNIVERSITAS NEGERI YOGYAKARTA

Gloria W.Z.

UNIVERSITAS BINA SARANA INFORMATIKA

Imas Siti M.

UNIVERSITAS NAHDLATUL ULAMA YOGYAKARTA

TABLE OF CONTENT

LATAR BELAKANG PROJECT

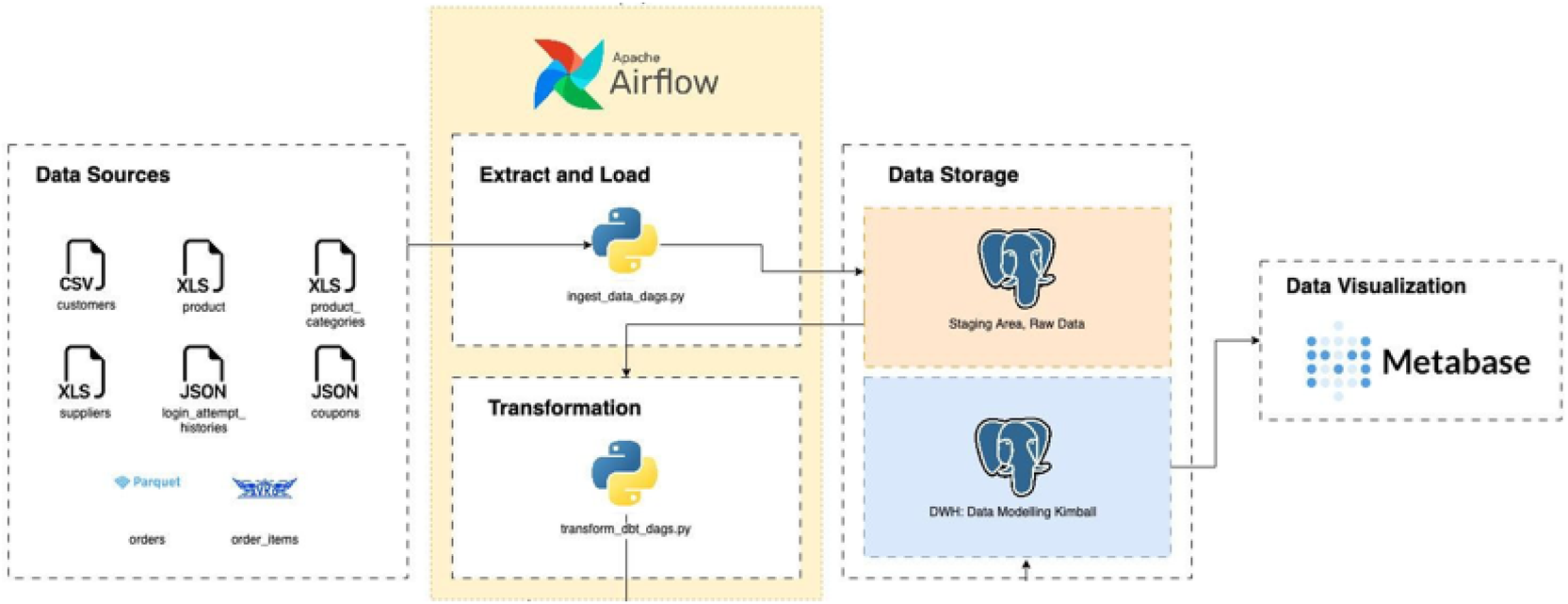
Data Engineer adalah orang yang ahli dalam bidang TI untuk mengolah hingga mengumpulkan berbagai macam data yang diperlukan oleh perusahaan dalam berbagai bidang. Data Engineer berperan besar dalam menyimpan, menganalisis dan mengolah data bahkan dalam jumlah besar di hampir seluruh bidang industri pekerjaan.

Sebagai seorang data enigeer, kita dipekerjakan di perusahaan ritel online AS yang menjual produk pelanggan umum langsung ke pelanggan dari berbagai pemasok di seluruh dunia. Kita membangun infrastruktur data menggunakan data yang dihasilkan yang dibuat untuk mencerminkan data dunia nyata dari perusahaan teknologi terkemuka.

Tasks

- ETL/ELT Job Creation using Airflow
- Data Modeling in Postgres
- Dashboard Creation with Data Visualization
- Craft a Presentation Based on Your Work

Data Platform Architecture

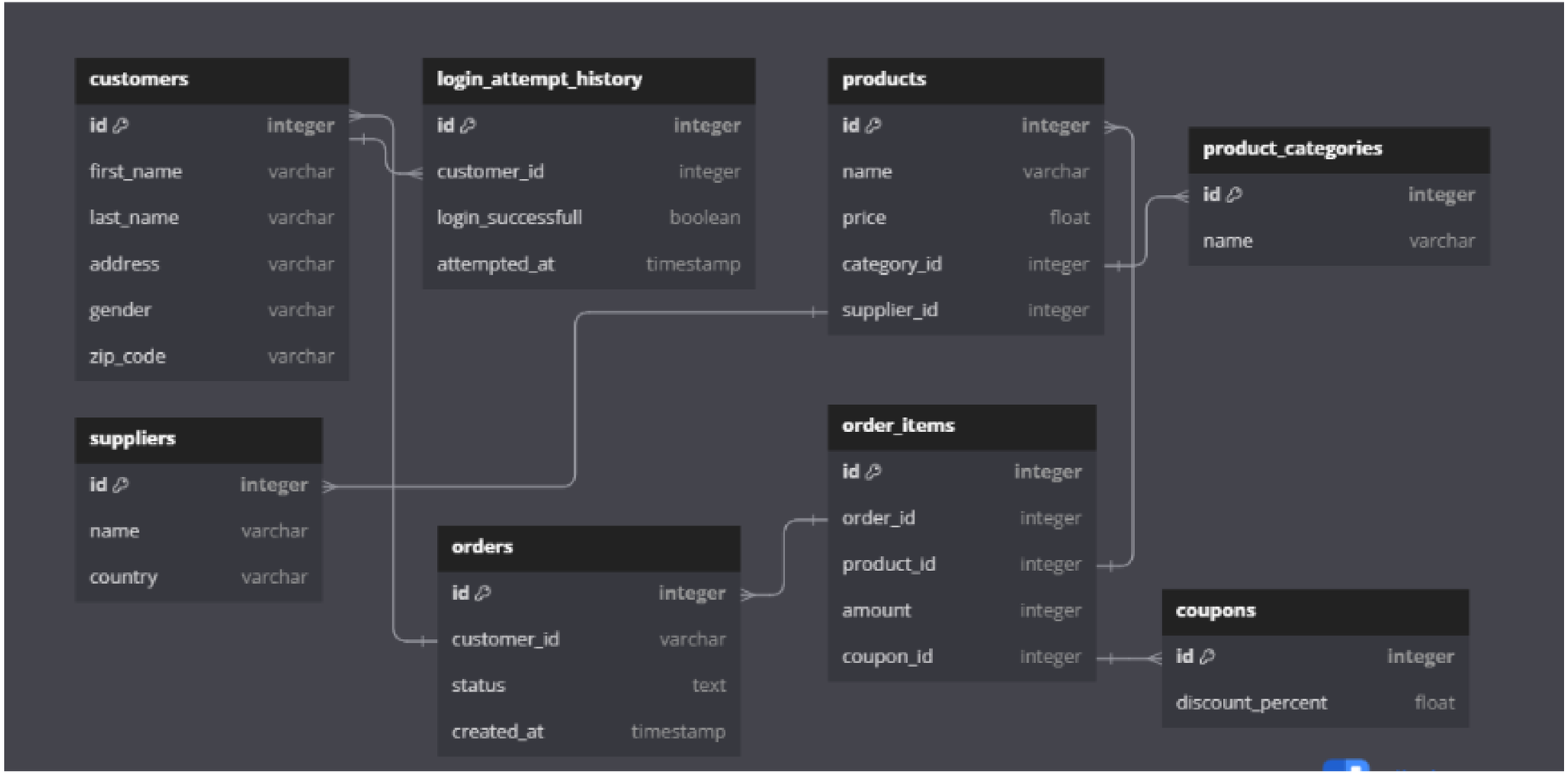


docker

Tech Stack

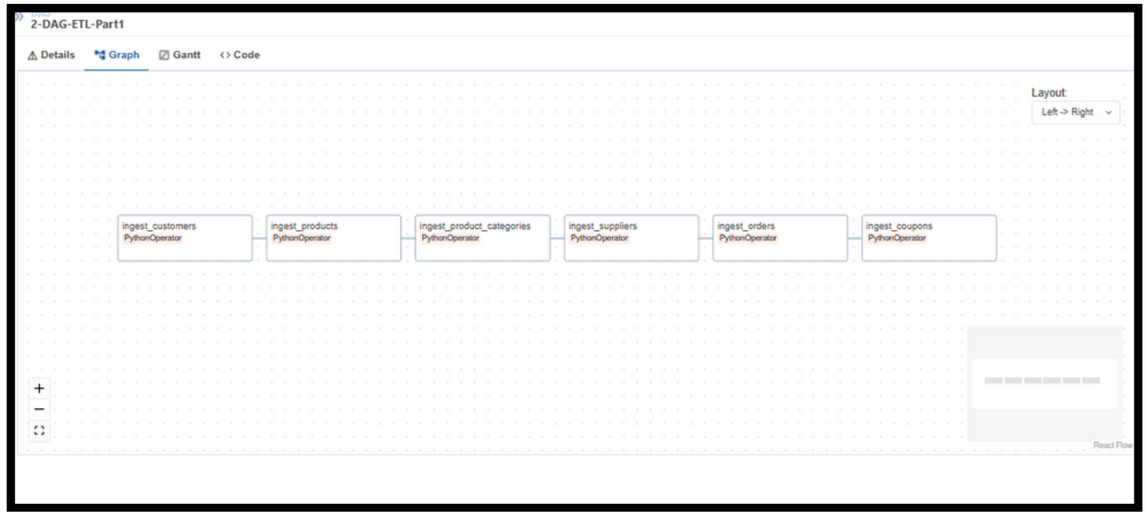
- Docker
- Python
- Airflow
- Postgresql
- Metabase

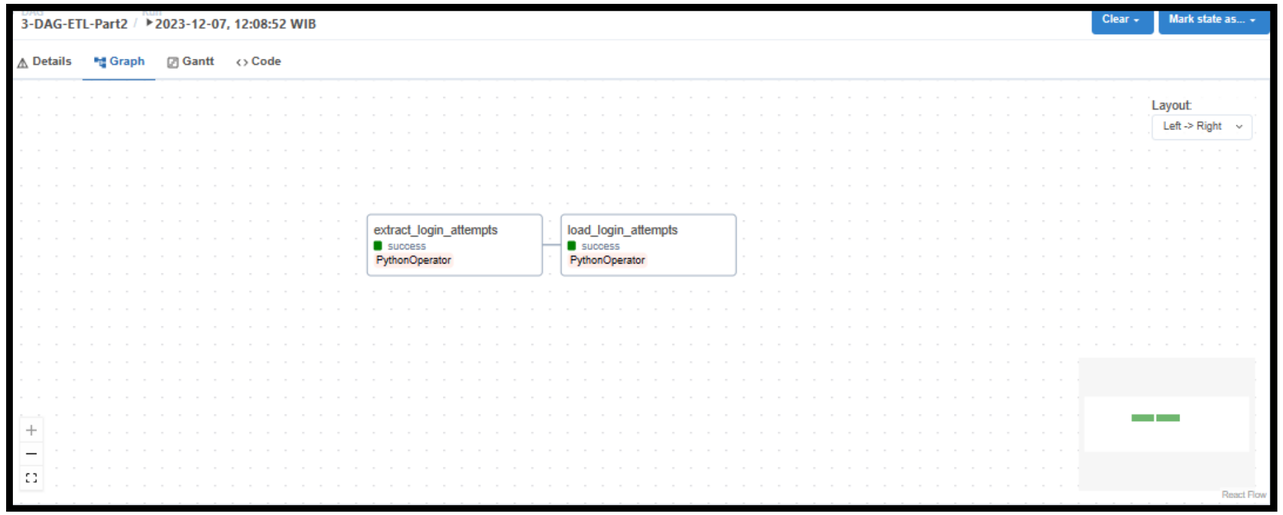
Data source

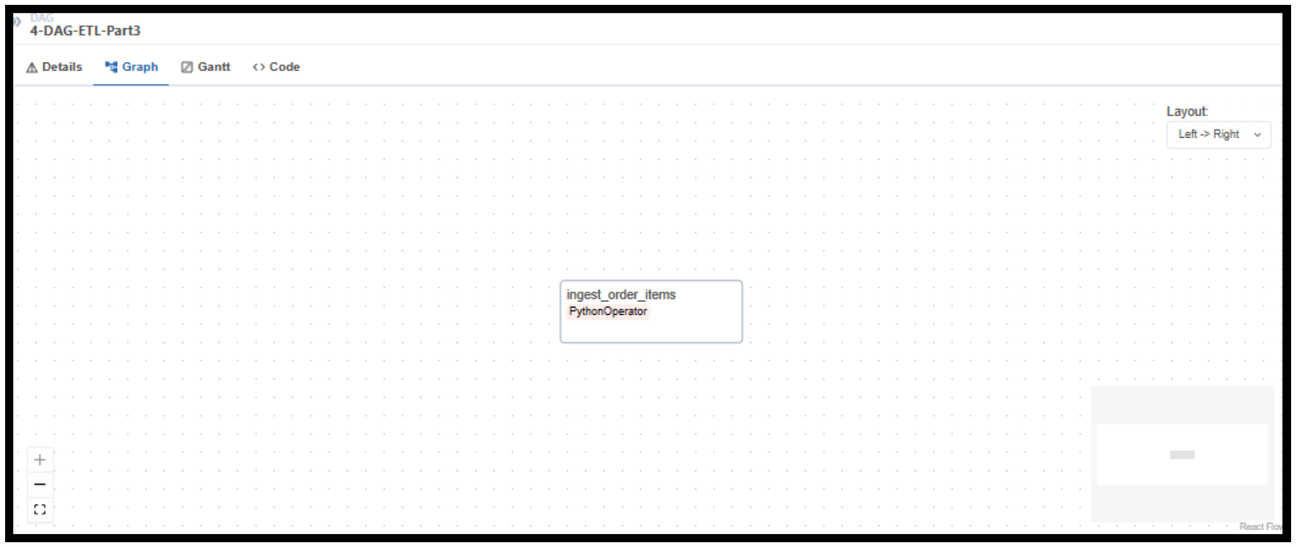

customers
id integer
first_name varchar
last_name varchar
address varchar
gender varchar
zip_code varchar
login_attempt_history
id integer
customer_id integer
login_successfull boolean
attempted_at timestamp
products
id integer
name varchar
price float
category_id integer
supplier_id integer
product_categories
id integer
name varchar
suppliers
id integer
name varchar
country varchar
order_items
id integer
order_id integer
product_id integer
amount integer
coupon_id integer
orders
id integer
customer_id varchar
status text
created_at timestamp
coupons
id integer
discount_percent float

DOKUMENTASI

ETL Job Creation using Airflow

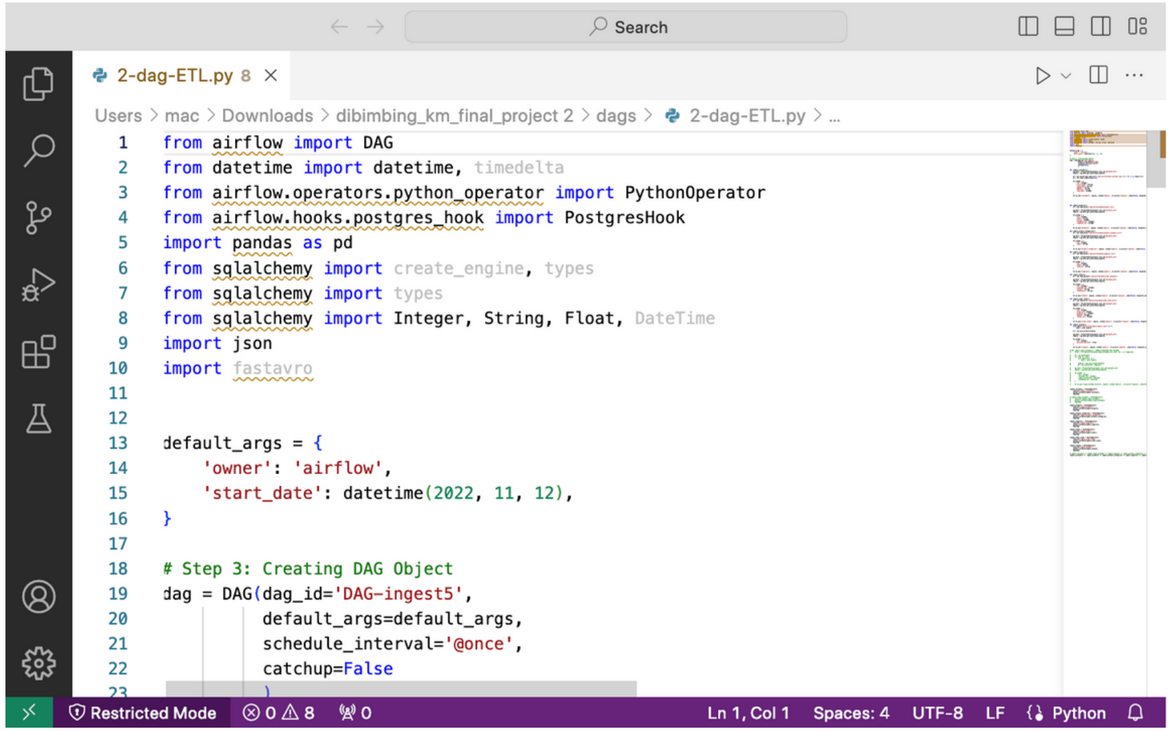


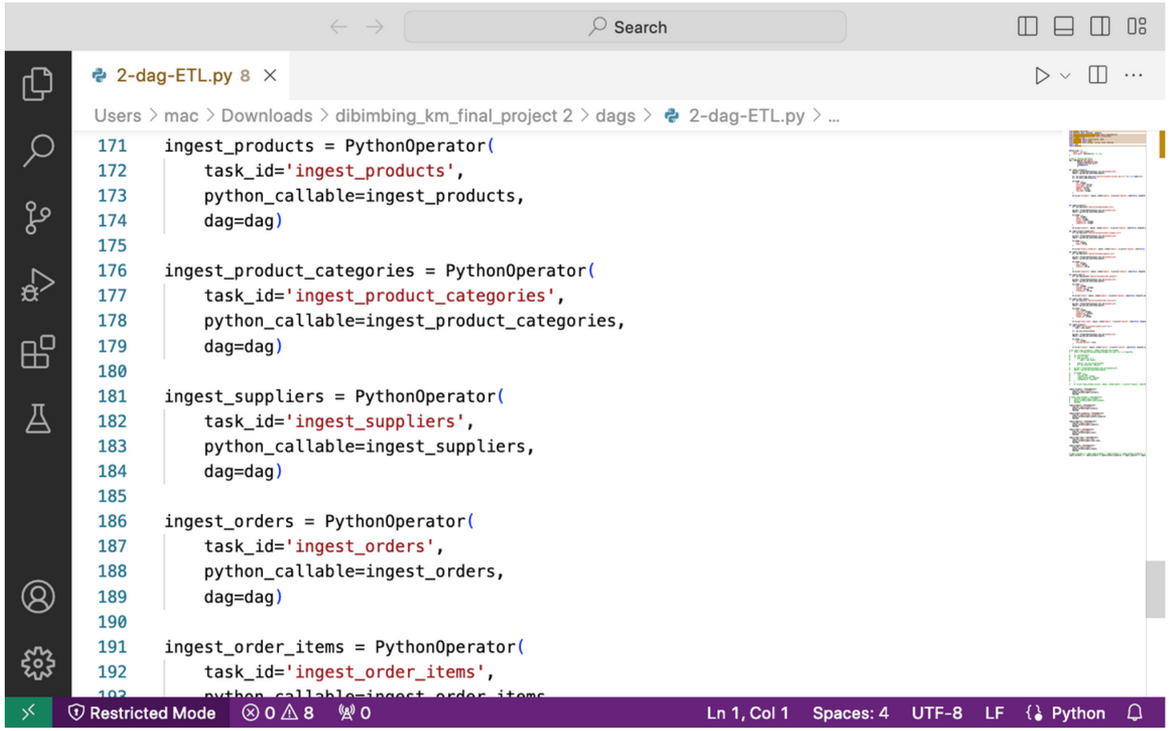




DOKUMENTASI

Ingest Data





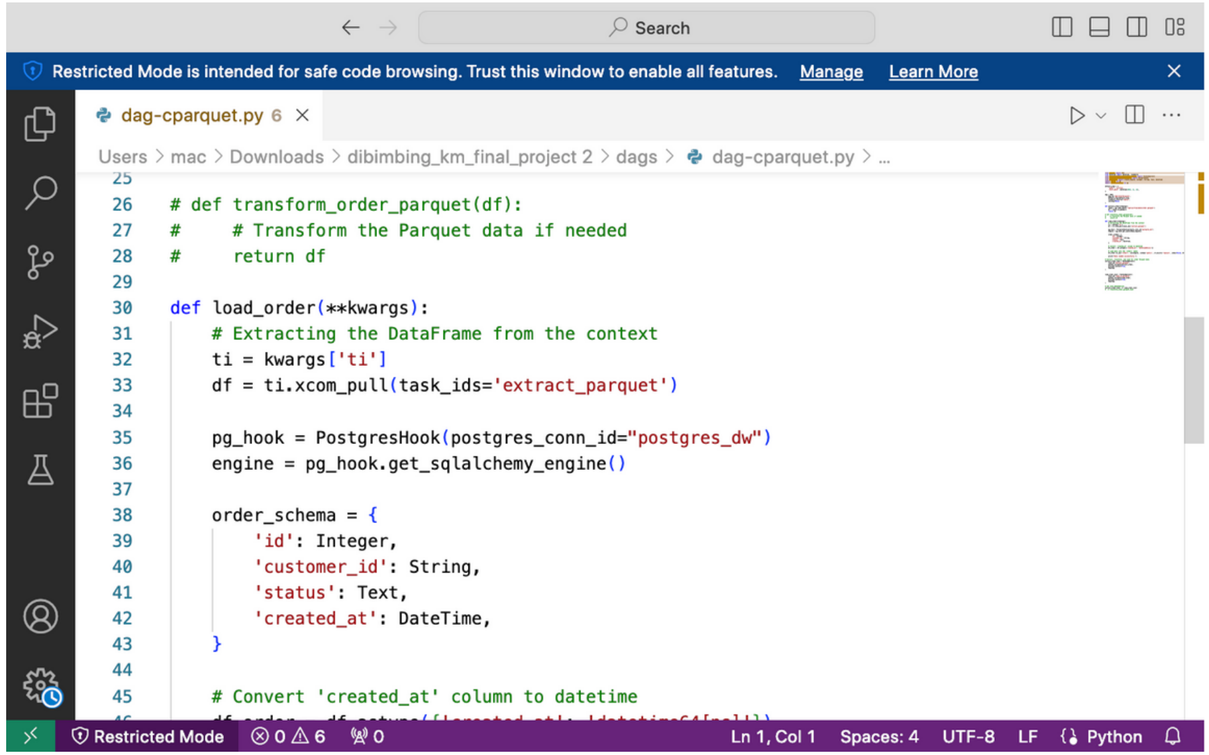
DOKUMENTASI

Transform

- Menghapus Kolom berlebih di tabel suppliers, customers, dan product_categoeris
- Menggabungkan file customers yang terpisah
- Menyesuaikan bentuk dari data json dan avro menjadi dataframe
- Mengubah tipe data sesuai dengan rancangan ERD

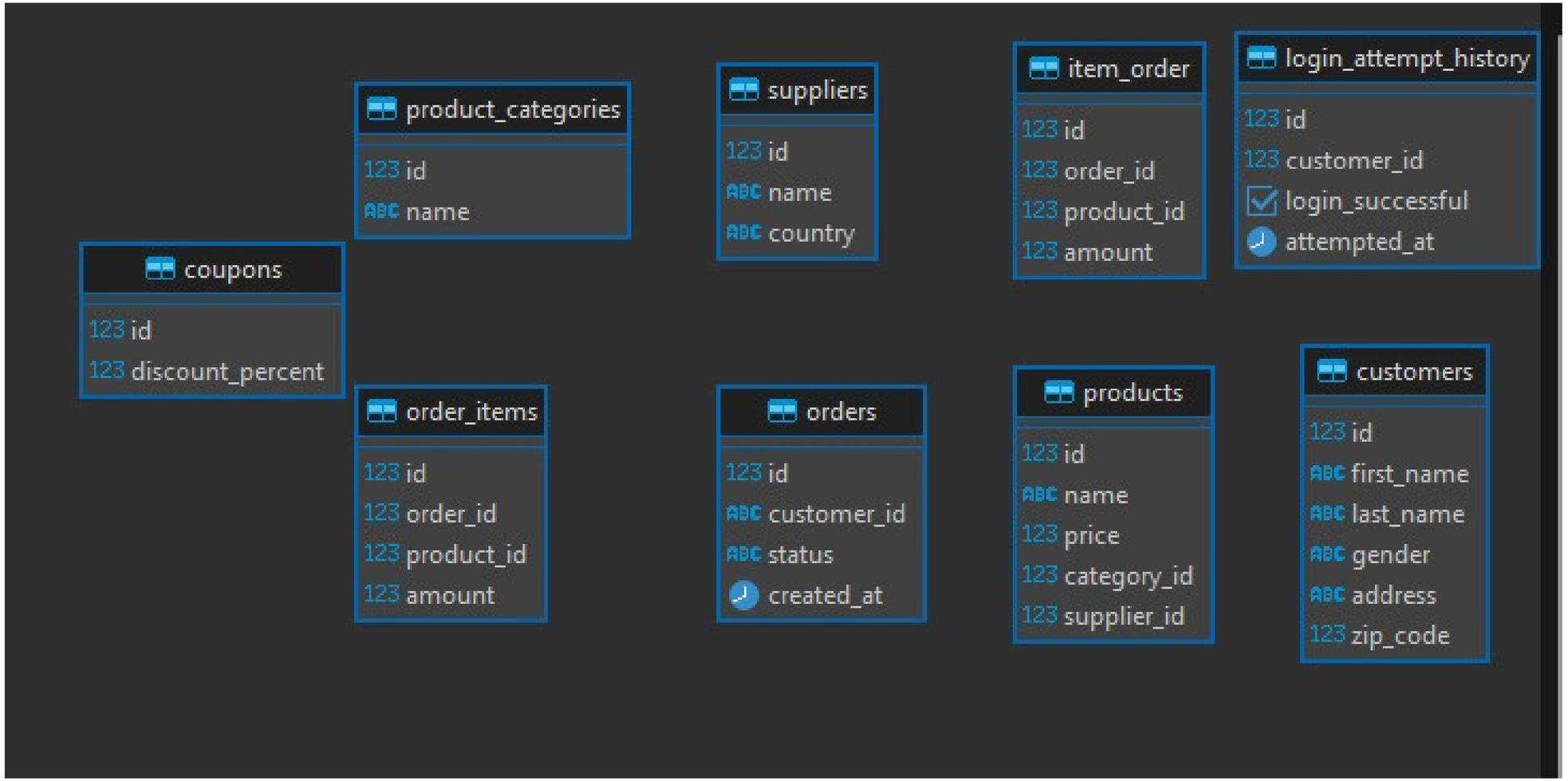
Load

- Memuat data yang diolah kedalam data warehouse



DOKUMENTASI

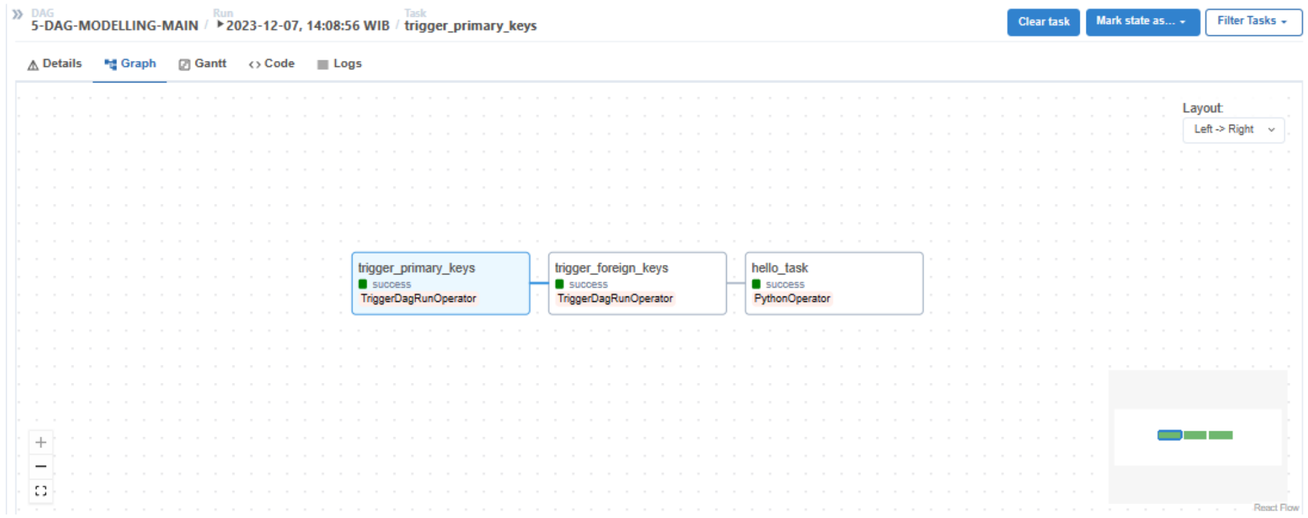
Default ETL Sebelum Modeling



DOKUMENTASI

Data Modelling

- Menentukan Primary keys dari masing-masing tabel
- Mengatur relasi tabel sesuai dengan ERD yang direncanakan



DOKUMENTASI

Data Modelling

Membuat dua tabel baru:

- sales_per_category yang didapat dengan menghitung total penjualan per kategori dengan cara: amount dikalikan price. Data diambil dari tabel order_items dan products.

- return_rate_per_category yang didapat dengan merata-rata total jumlah dari setiap kategori. Tabel ini dibuat dengan tujuan untuk mengetahui kategori mana yang memiliki pengembalian tinggi, dan mengetahui faktor apa yang menyebabkannya.

```
19  # Task untuk membuat tabel sales_per_category
20  create_sales_per_category_table = PostgresOperator(
21      task_id='create_sales_per_category_table',
22      sql="""
23      CREATE TABLE IF NOT EXISTS sales_per_category AS
24      SELECT
25        p.category_id,
26        c.name AS category,
27        SUM(oi.amount * p.price) AS total_sales
28      FROM order_items oi
29      INNER JOIN products p ON oi.product_id = p.id
30      INNER JOIN product_categories c ON c.id = p.category_id
31      GROUP BY p.category_id, c.name;
32      """,
33      postgres_conn_id='postgres_dw',
34      autocommit=True,
35      dag=dag,
36  )
```

DOKUMENTASI

Perubahan yang dilakukan

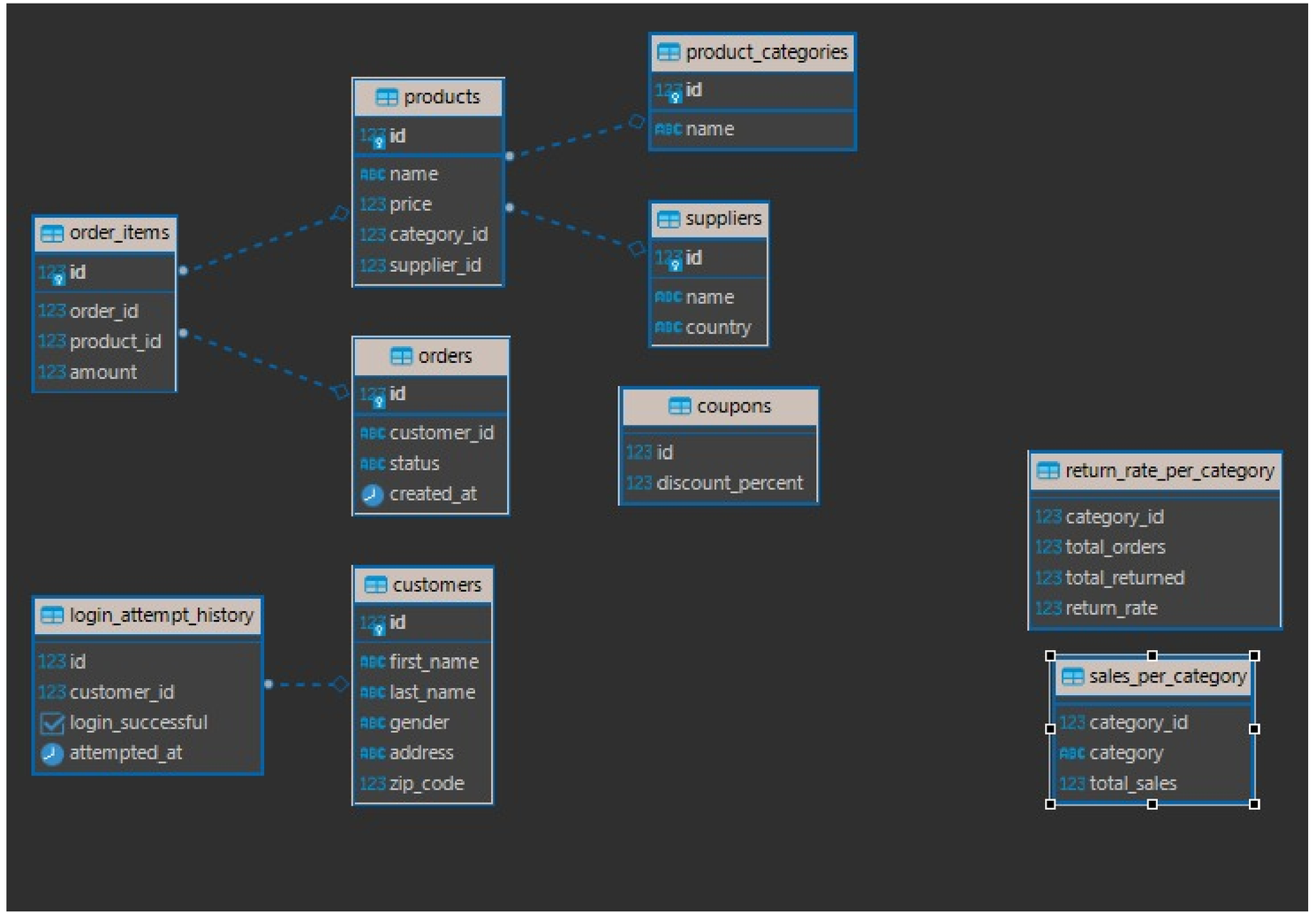
requirements.txt

```
pandas
fastavro
pyarrow
xlrd
cramjam
avro
python-snappy
copy
```

Menginstall library yang di pakai

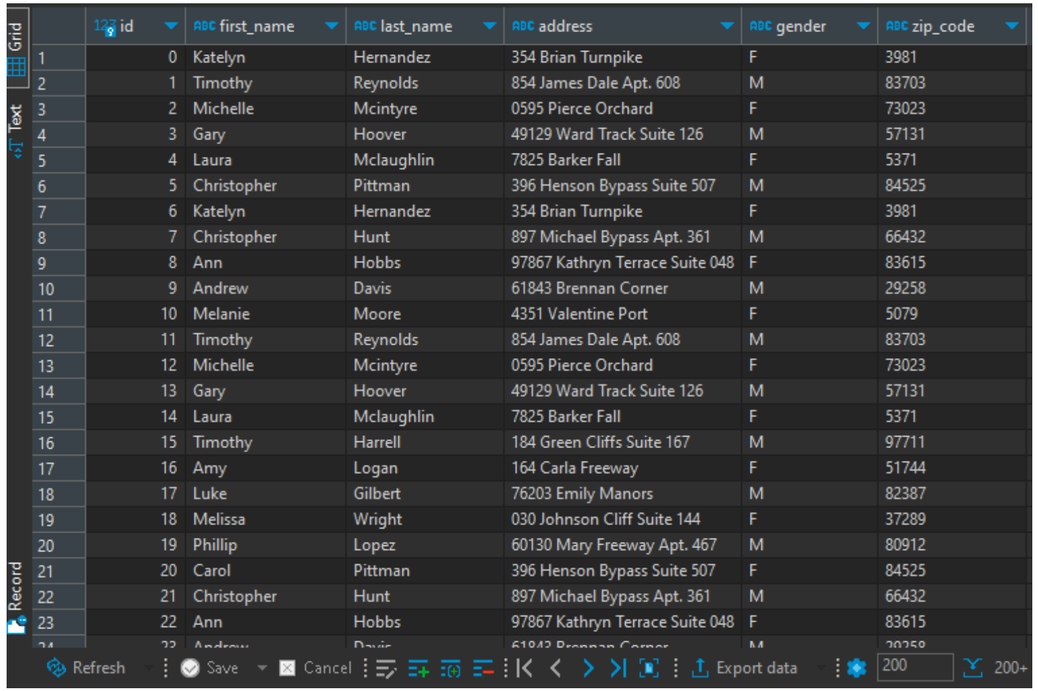
DOKUMENTASI

Data Modeling in Postgres



DOKUMENTASI

Data Preview at PostgreSQL

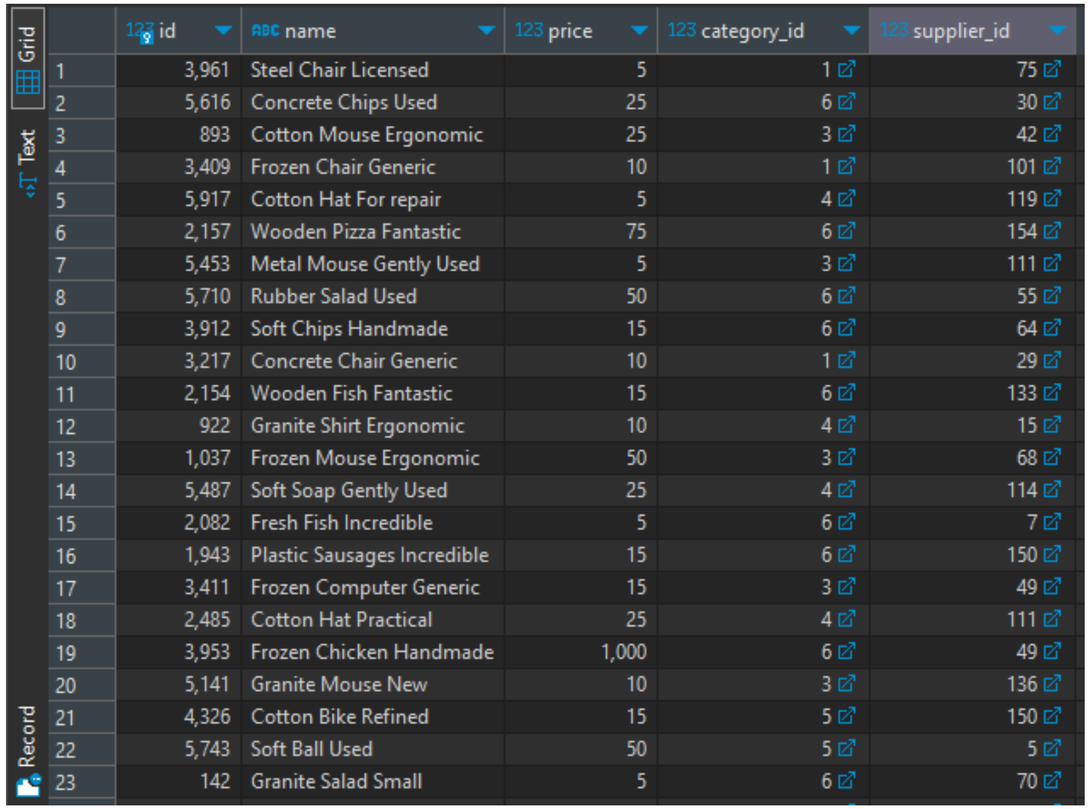


	id	first_name	last_name	address	gender	zip_code
1	0	Katelyn	Hernandez	354 Brian Turnpike	F	3981
2	1	Timothy	Reynolds	854 James Dale Apt. 608	M	83703
3	2	Michelle	Mcintyre	0595 Pierce Orchard	F	73023
4	3	Gary	Hoover	49129 Ward Track Suite 126	M	57131
5	4	Laura	Mclaughlin	7825 Barker Fall	F	5371
6	5	Christopher	Pittman	396 Henson Bypass Suite 507	M	84525
7	6	Katelyn	Hernandez	354 Brian Turnpike	F	3981
8	7	Christopher	Hunt	897 Michael Bypass Apt. 361	M	66432
9	8	Ann	Hobbs	97867 Kathryn Terrace Suite 048	F	83615
10	9	Andrew	Davis	61843 Brennan Corner	M	29258
11	10	Melanie	Moore	4351 Valentine Port	F	5079
12	11	Timothy	Reynolds	854 James Dale Apt. 608	M	83703
13	12	Michelle	Mcintyre	0595 Pierce Orchard	F	73023
14	13	Gary	Hoover	49129 Ward Track Suite 126	M	57131
15	14	Laura	Mclaughlin	7825 Barker Fall	F	5371
16	15	Timothy	Harrell	184 Green Cliffs Suite 167	M	97711
17	16	Amy	Logan	164 Carla Freeway	F	51744
18	17	Luke	Gilbert	76203 Emily Manors	M	82387
19	18	Melissa	Wright	030 Johnson Cliff Suite 144	F	37289
20	19	Phillip	Lopez	60130 Mary Freeway Apt. 467	M	80912
21	20	Carol	Pittman	396 Henson Bypass Suite 507	F	84525
22	21	Christopher	Hunt	897 Michael Bypass Apt. 361	M	66432
23	22	Ann	Hobbs	97867 Kathryn Terrace Suite 048	F	83615

Grid Text Record

Refresh Save Cancel Export data 200 200+

customers



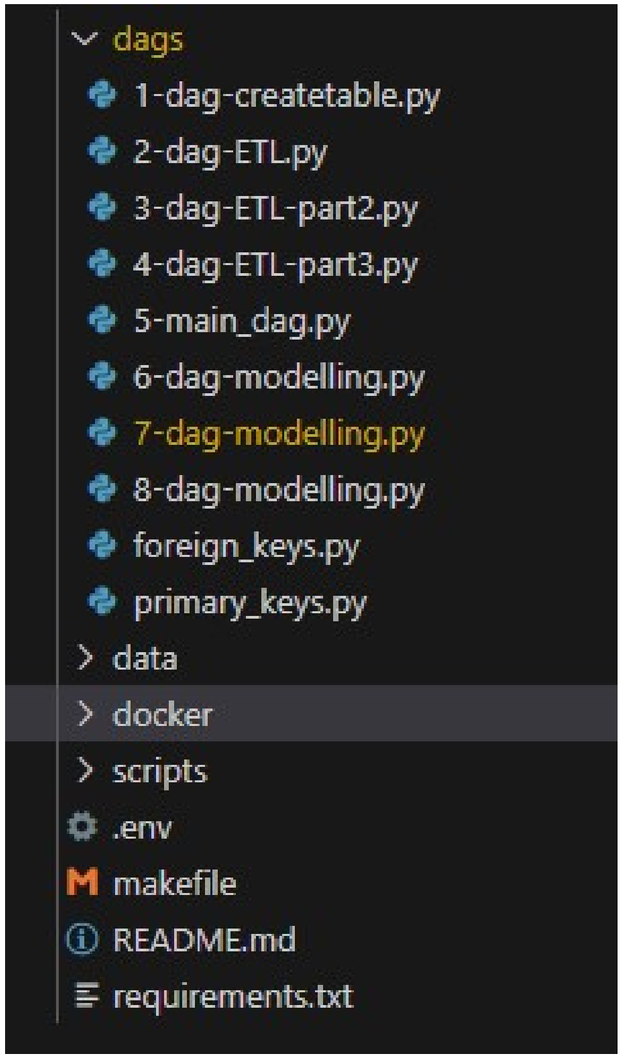
	id	name	price	category_id	supplier_id
1	3,961	Steel Chair Licensed	5	1	75
2	5,616	Concrete Chips Used	25	6	30
3	893	Cotton Mouse Ergonomic	25	3	42
4	3,409	Frozen Chair Generic	10	1	101
5	5,917	Cotton Hat For repair	5	4	119
6	2,157	Wooden Pizza Fantastic	75	6	154
7	5,453	Metal Mouse Gently Used	5	3	111
8	5,710	Rubber Salad Used	50	6	55
9	3,912	Soft Chips Handmade	15	6	64
10	3,217	Concrete Chair Generic	10	1	29
11	2,154	Wooden Fish Fantastic	15	6	133
12	922	Granite Shirt Ergonomic	10	4	15
13	1,037	Frozen Mouse Ergonomic	50	3	68
14	5,487	Soft Soap Gently Used	25	4	114
15	2,082	Fresh Fish Incredible	5	6	7
16	1,943	Plastic Sausages Incredible	15	6	150
17	3,411	Frozen Computer Generic	15	3	49
18	2,485	Cotton Hat Practical	25	4	111
19	3,953	Frozen Chicken Handmade	1,000	6	49
20	5,141	Granite Mouse New	10	3	136
21	4,326	Cotton Bike Refined	15	5	150
22	5,743	Soft Ball Used	50	5	5
23	142	Granite Salad Small	5	6	70

Grid Text Record

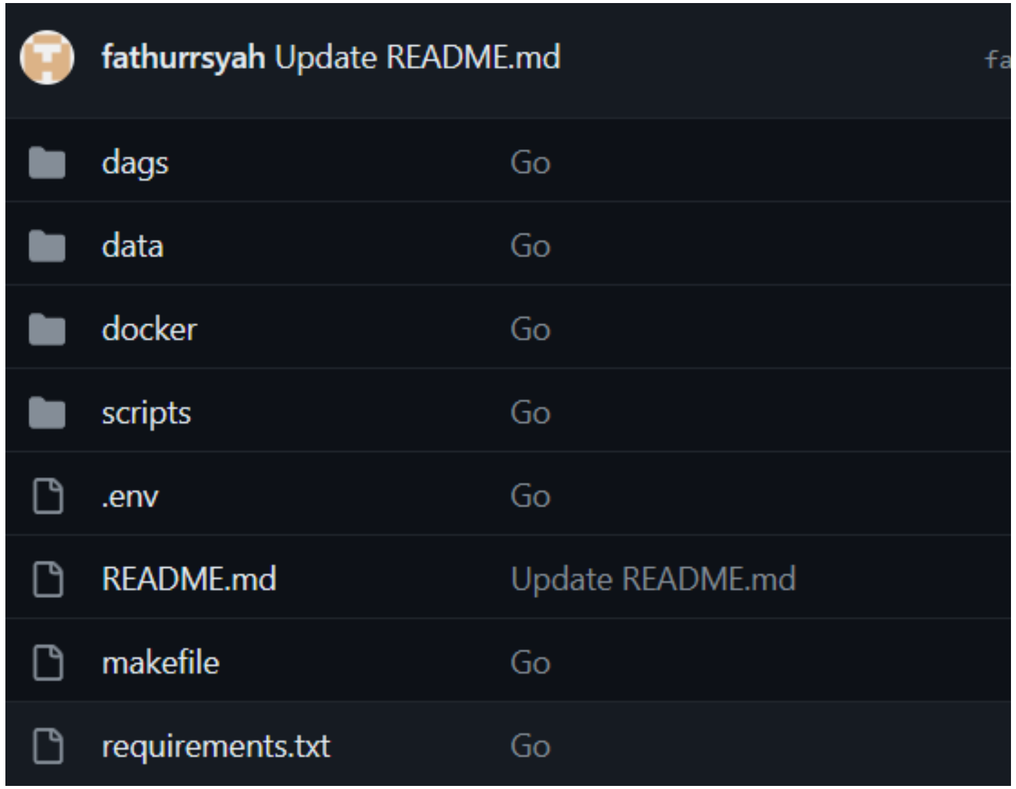
products

DOKUMENTASI

Struktur Folder

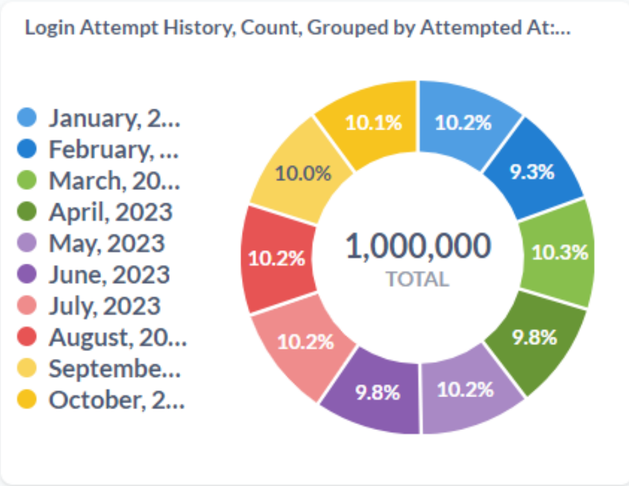


Dapat diakses di Github

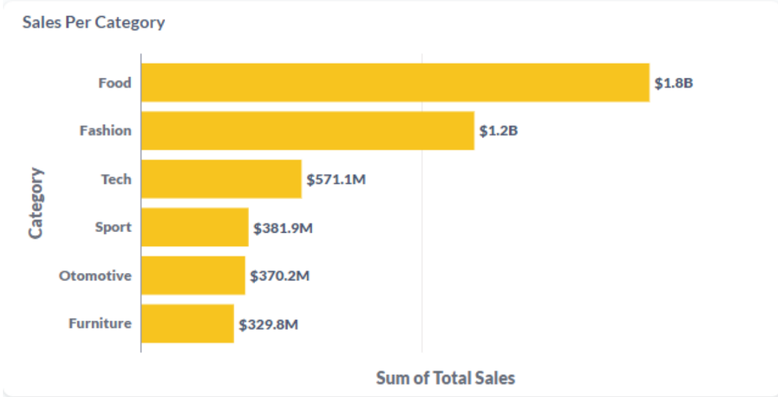


DOKUMENTASI

Dashboard Creation with Data Visualization



Menghitung jumlah Customer selama satu Tahun belakangan

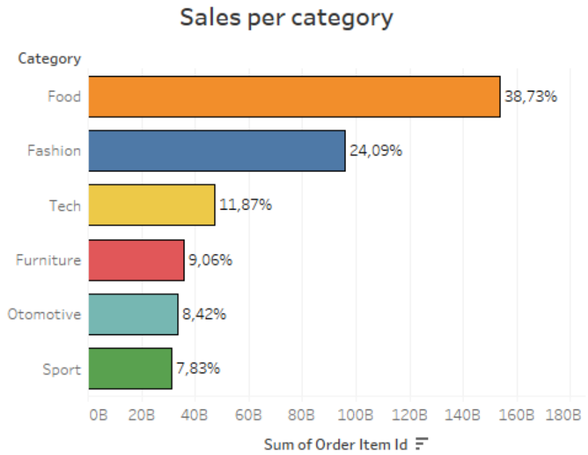


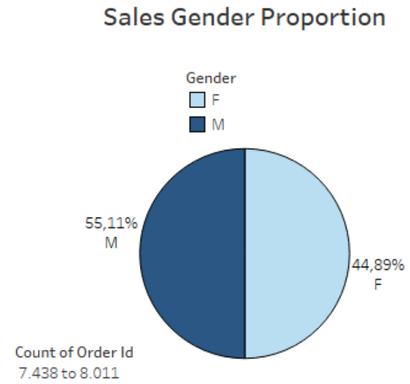
Penjualan Per Kategori dan Jumlah Total Penjualan, dikelompokkan serta diurutkan berdasarkan Jumlah Total Penjualan secara menurun

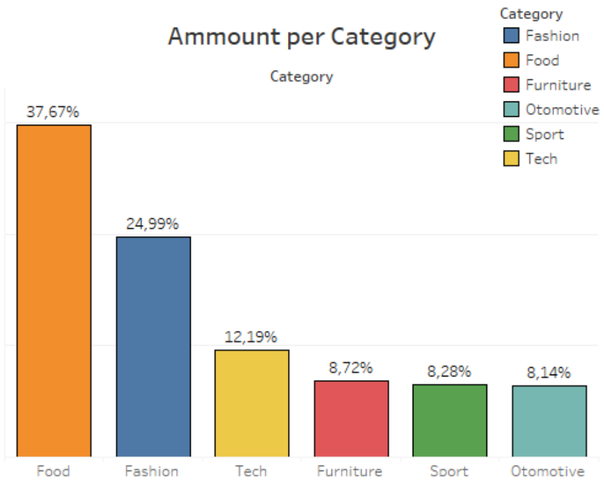
menampilkan jumlah total pelanggan yang memiliki jumlah 10, 12, 8, atau 6 di setiap kode pos yang ada dalam tabel "zipcode_customer"

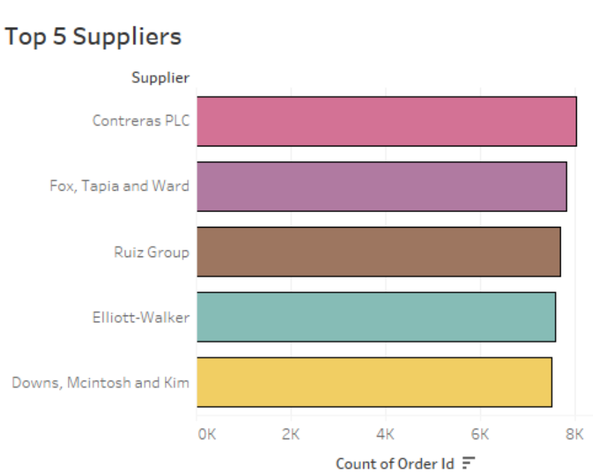
DOKUMENTASI

Pembuatan Dashboard menggunakan tableau



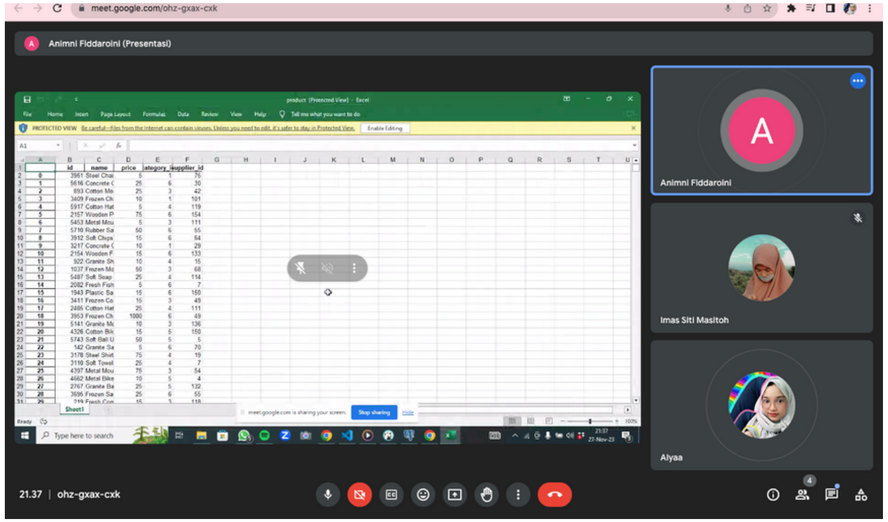


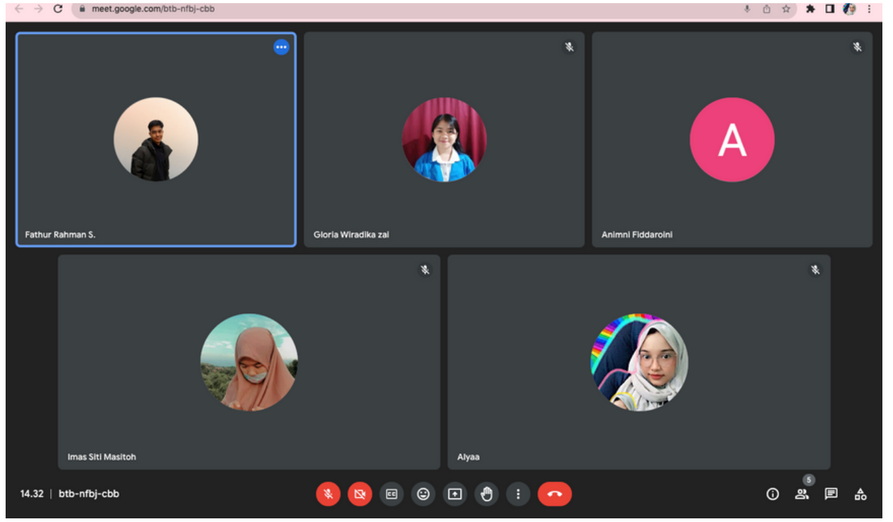


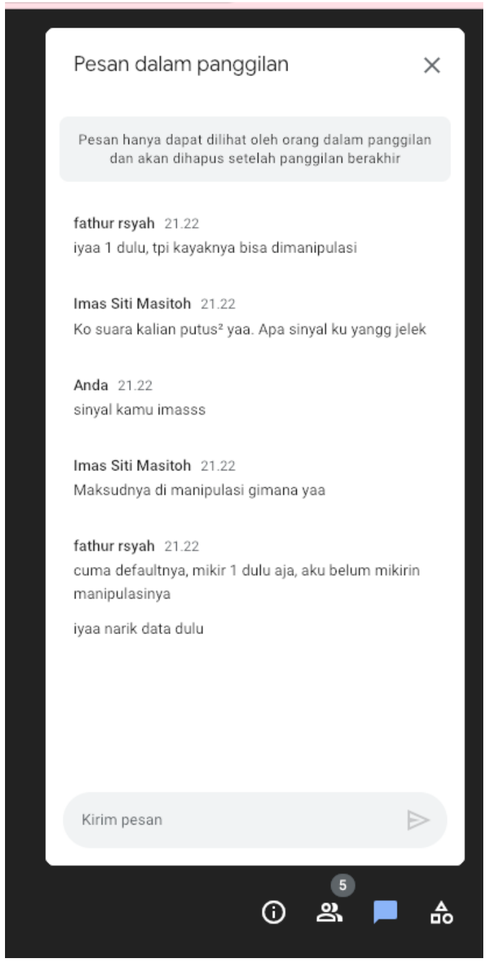


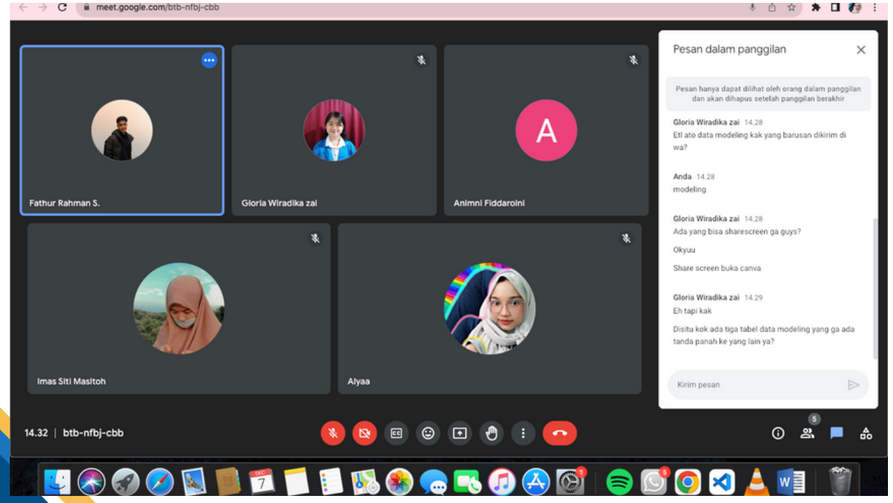
DOKUMENTASI

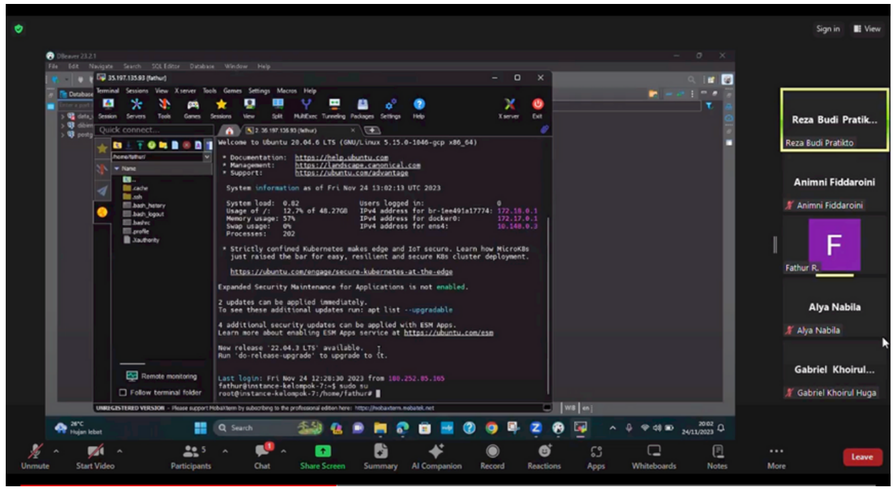
Aktivitas Diskusi Kelompok 7











THANK YOU

Dibimbing X Kampus Merdeka