**DOCUMENTATION**

**MANDIRI INHEALTH DATA ENGINEER TECHNICAL TEST**

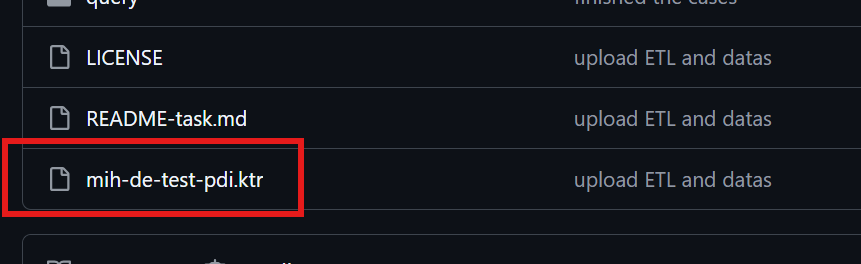
Mochammad Aditya Putra Suhendar

Data Engineer 1.3yoe at Akuntplus

[LinkedIn](https://www.linkedin.com/in/darhensu) [GitHub](https://www.github.com/darhensu)  [This task REPO](https://github.com/darhensu/MIH-TEST)  [Email](mailto:darhensu@gmail.com) [WhatsApp](wa.me/6285156595451)

1. **ETL**

Tools: Pentaho, MariaDB, DBeaver



Disini saya menggunakan PDI (Pentaho Data Integration), filenya dapat dilihat dengan format di atas di GitHub Repo saya dengan nama file ‘mih-de-test-pdi.ktr’. For additional notes, saya menggunakan PDI versi 9.3.

**EXTRACT**

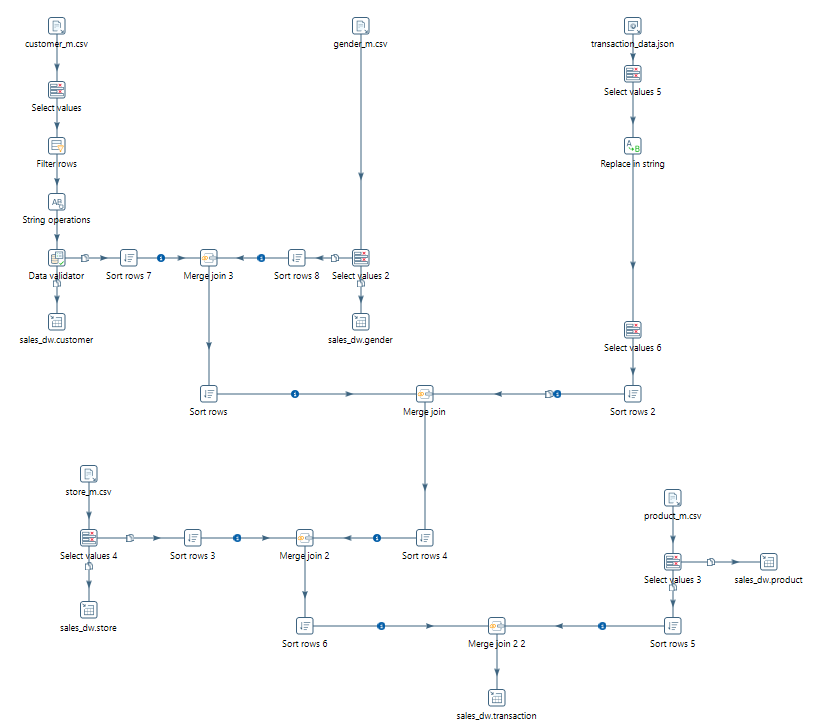
Saya melakukan ekstaksi data yang sudah disediakan dengan 4 CSV dan 1 JSON dengan box peruntukannya masing-masing.

**TRANSFORM**

Pada proses ini saya banyak melibatkan pada perubahan nama column agar tidak pusing dan dapat membedakan mana table biasa dan transaction, pada table transaction saya mengubah nominal yang varchar karena ada symbol mata uang ‘$’ dengan cara menghapusnya dan dijadikan number agar dapat menjawab case jika ditanyakan total pengeluaran kedepannya.

**LOAD**

Kemuadian step terakhir data di load ke MariaDB system localhost dengan MySQL ke setiap-setiap table masing-masing. Kecuali table transaction harus dijoinkan terlebih dahulu.

Berikut adalah gambaran kasar ETL yang sudah saya bangun:  
  


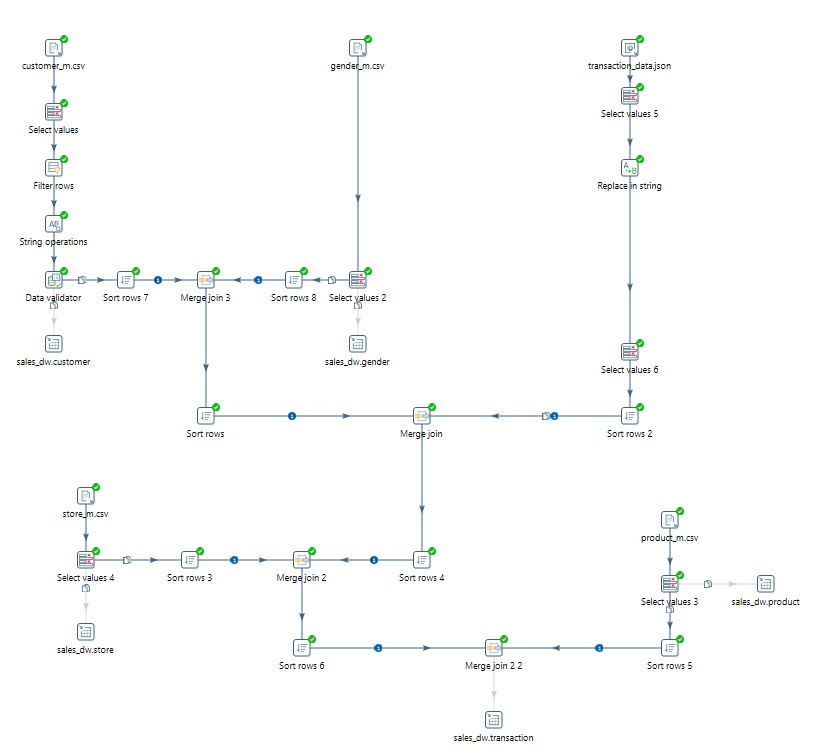
Sebelum dijalankan saya membuat databasenya terlebih dahulu dan juga table-tablenya agar sinkron antara keinginan, petunjuk dan juga default data type yang dibaca oleh Pentaho.

create database sales\_dw;

use sales\_dw;

|  |
| --- |
| Membuat table gender:  create table gender (  GenderID int(11) not null primary key,  Gender varchar(10),  ); |
| Membuat table customer:  create table customer (  CustomerID int(11) not null primary key,  GenderID int(11),  FirstName varchar(20),  LastName varchar(20),  Email varchar(28),  foreign key (GenderID) references gender (GenderID)  ); |
| Membuat table store:  create table store (  StoreID int(11) not null primary key,  StoreName varchar(100),  StoreLocation varchar(20)  ); |
| Membuat table product:  create table product (  ProductID int(11) not null primary key,  ProductName varchar(50),  ); |
| Membuat table transaction:  create table transaction (  trx\_id uuid not null primary key,  customer\_id int(11),  product\_id int(11),  store\_id int(11),  item\_quantity int(11),  item\_price decimal(10,2),  trx\_date date,  foreign key (customer\_id) references customer (CustomerID),  foreign key (product\_id) references product (ProductID),  foreign key (store\_id) references store (StoreID)  ); |

Baru setelah itu saya jalankan ETL di Pentaho yang hasilkan akan seperti dibawah ini:  
  
**DISCLAIMER: Mohon untuk diabaikan bagian untuk Loadnya tidak saya nyalakan karena sudah saya jalankan pada saat awal namun tidak terdokumentasikan, tapi yang pasti berhasil jika ingin dikosongkan terlebih dahulu table-tablenya.**

****

**CASE 1-6**

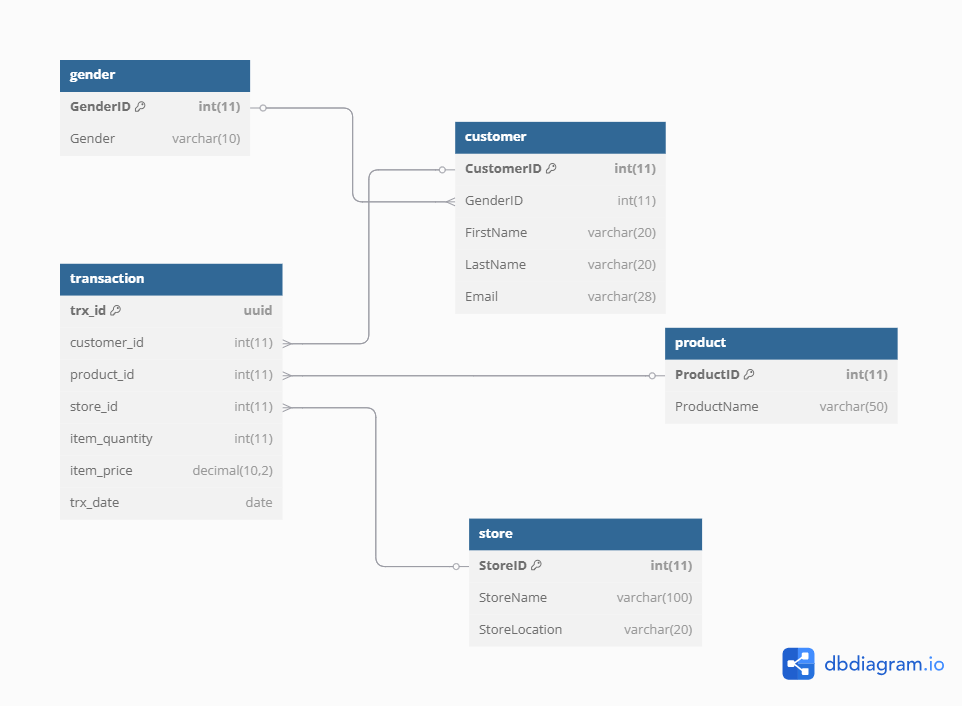
Tools: DBeaver

|  |
| --- |
| **# CASE 1**  **# Menampilkan 5 nama lengkap customer dengan pengeluaran paling banyak, urutan dari terbesar ke terkecil** |
|  |
| **# CASE 2**  **# Menampilkan 5 kota dengan total item terjual paling banyak, urutan dari terbesar ke terkecil** |
|  |
| **# CASE 3**  **# Menampilkan 5 product yang paling banyak dibeli oleh laki laki, urutan dari terkecil ke terbesar** |
|  |
| **# CASE 4**  **# Menampilkan 5 toko yang menjual product paling banyak pada bulan januari (akumulasi penjualan produk yang terjual hanya di bulan januari di semua tahun) , urutan dari terbesar ke terkecil** |
|  |
| **# CASE 5**  **# Menampilkan 5 email customer yang berasal dari pemerintahan (.gov) yang memiliki total pembelian terbesar**  **# di kota yang paling banyak menjual product (total item quantity) paling banyak kepada pembeli perempuan.**  **# HINT : cari terlebih dahulu kota apa yang paling banyak menjual barang untuk perempuan.**  **# kemudian di kota tersebut cari 5 customer dari pemerintahan yang melakukan pembelian terbesar** |
|  |
| **# CASE 6**  **# Menampilkan 5 nama lengkap customer dengan pengeluaran paling banyak,**  **# disertai dengan informasi produk apa saja yang pernah dibeli (unique),**  **# dan informasi toko mana saja tempat produk tersebut dijual (unique)** |
|  |

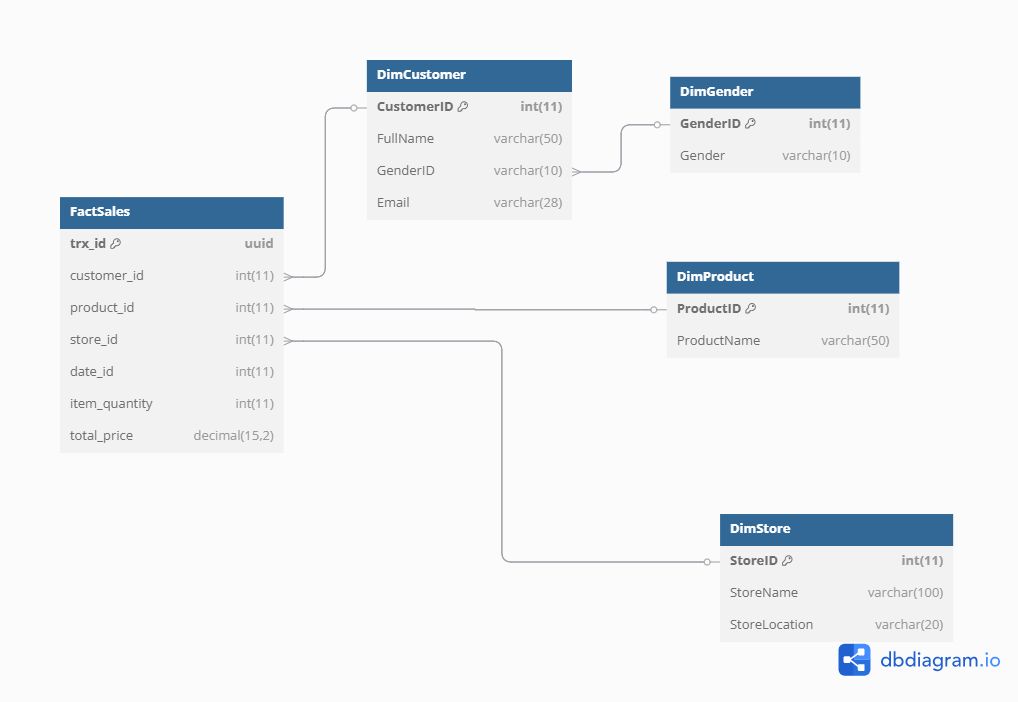
**ERD OLTP & OLAP**

Tools: dbdiagram.io

Saya menggunakan dbdiagram.io untuk membuat ERD OLTP dan OLAP model, berikut adalah ERD OLTP:

****

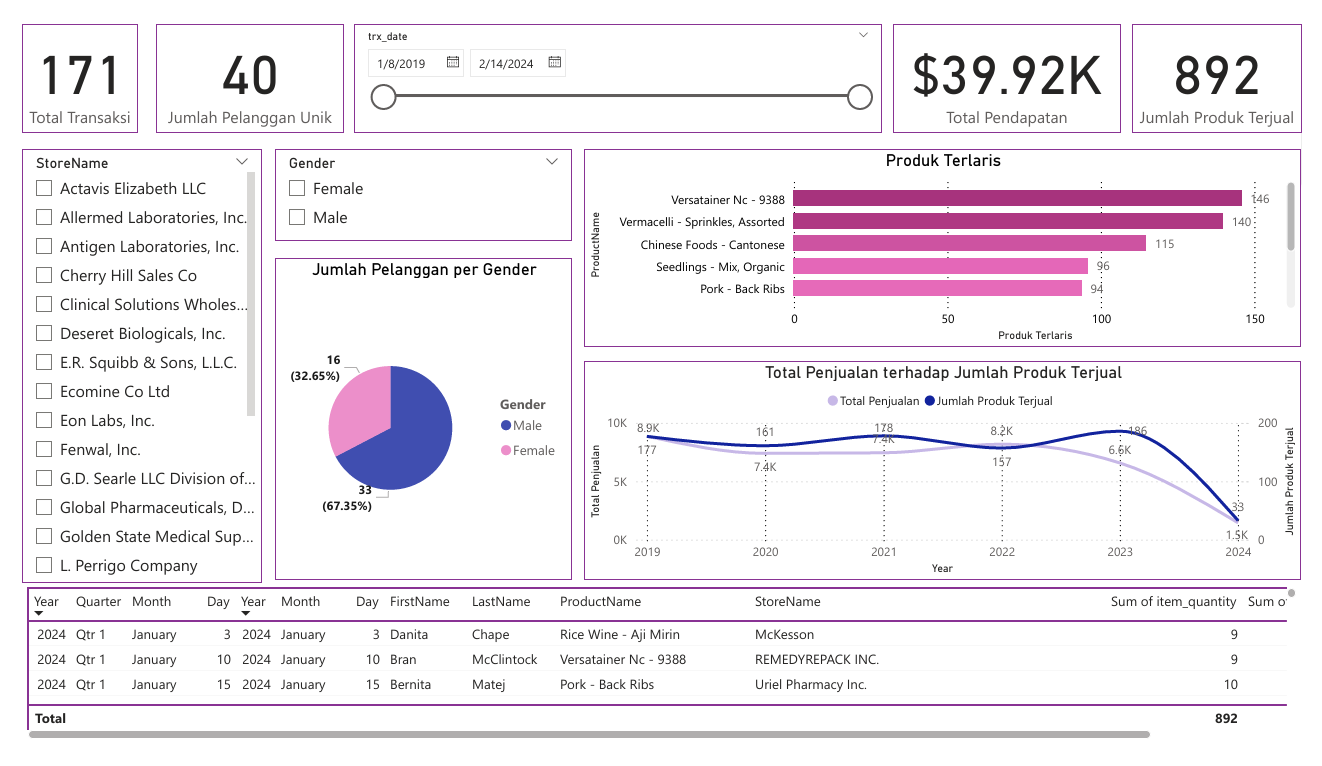
Berikut adalah model OLAP berbentuk star schema:



**VISUALIZATION**

Tools: PowerBI

Saya menggunakan PowerBI untuk melakukan visualization dengan beberapa charts, cards, dan juga filter slicers

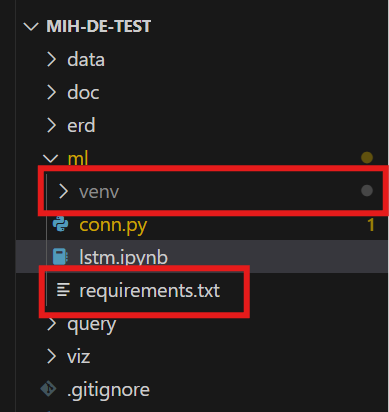


**SALES PREDICTION USING LSTM**

Tools: VSC

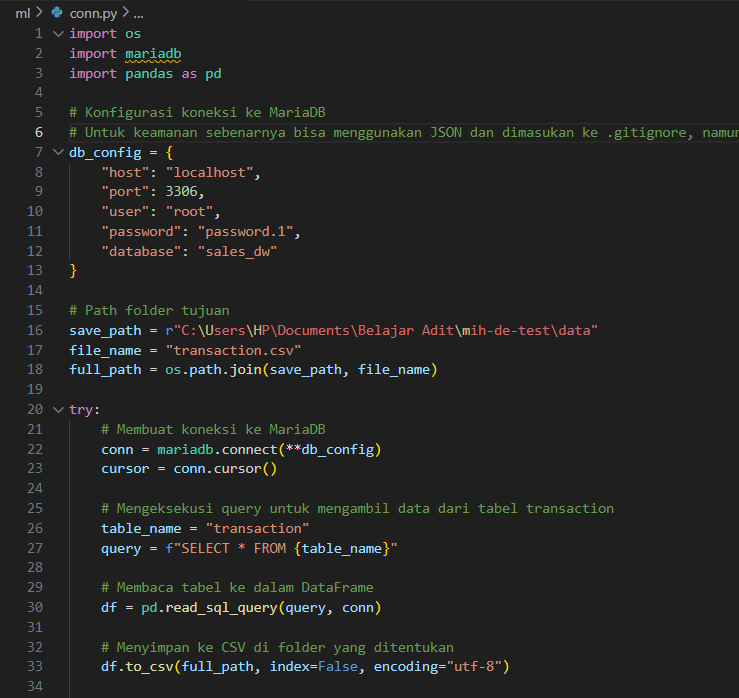
Pada tahap ini saya mencoba untuk membangun model LSTM dengan tujuan untuk memprediksi penjualan dari data transaction selama 30 hari kedepan, semua progress dan perkembangan dapat dilihat pada [GitHub REPO](https://github.com/darhensu/MIH-TEST).

1. **Pembuatan VENV, .gitignore dan Install requirements.txt**



Pembuatan cirtual environtment (venv) ditujukan untuk membuat suatu lingkungan dengan library dependencies untuk suatu project. Kemudian venv dimasukan kedalam .gitignore agar tidak ke push ke repo karena size yang besar. Requirements.txt berisi library yang perlu diinstall dan nantinya akan masuk kedalam venv yang sudah dibuat.

1. **Pembuatan conn.py**

****

Pembuatan file ini dimaksudkan untuk mendapatkan connection dari dan ke MariaDB untuk merubah table dan datanya menjadi csv yang kemudian diimport ke local. Disini saya tidak menggunaka JSON untuk credentialsnya karena menggunakan localhost, jika mengunakan database server diwajibkan untuk menggunakan JSON dan dimasukan kedalam .gitignore agar tidak dapat diakses sembarang orang.

1. **LSTM.ipynb**

**Karena keterbatasan waktu dokumen ini terlampir di** [GitHub REPO](https://github.com/darhensu/MIH-TEST). **Masih banyak yang perlu diperbaiki dan on progress, akan tetap saya lanjutkan untuk progress ini sementara dokumen ini saya kumpul terlebih dahulu sebelum 26/03/2025 16.00 WIB.**