Presented by
Group 3

Retail
Data Warehouse

This dataset represents a realistic retail business

Dataset

Dataset คือชุดข้อมูลที่นำมาใช้ในการทำโครงงาน โดย Dataset นี้เป็นข้อมูลเกี่ยวกับ การขายสินค้าและการดำเนินงานของร้านค้าปลีก (Retail) ซึ่งประกอบด้วยข้อมูลลูกค้า สินค้า ร้านค้า คำสั่งซื้อ การชำระเงิน การจัดส่ง และการคืนสินค้า

รายละเอียดตารางข้อมูล

ลำดับ	ตาราง	จำนวน Records	จำนวน Columns	รายละเอียด
1	employees	1,000	3	ข้อมูลพนักงาน
2	returns	30,000	3	ข้อมูลการคืนสินค้า
3	products	10,000	4	ข้อมูลสินค้า
4	suppliers	200	2	ข้อมูลผู้จัดจำหน่าย
5	categories	30	2	ข้อมูลประเภทสินค้า
6	promotions	50	2	ข้อมูลโปรโมชั่น
7	stores	100	2	ข้อมูลสาขา
8	customers	50,000	3	ข้อมูลลูกค้า
9	payments	300,000	3	ข้อมูลการชำระเงิน
10	orders	300,000	5	ข้อมูลคำสั่งซื้อ
11	order_items	600,000	5	รายละเอียดสินค้าในคำสั่งซื้อ
12	shipments	300,000	3	ข้อมูลการจัดส่ง
Dataset หลักทั้ง 12 ตารางมีข้อมูลรวมทั้งหมด 1,631,380 Records และ 39 Columns				

3. OLTP (Online Transaction Processing)

OLTP คือ ระบบฐานข้อมูลที่ใช้จัดเก็บและประมวลผลธุรกรรมที่เกิดขึ้นจากการดำเนินงานประจำวันขององค์กร เพื่อบันทึก แก้ไข และเรียกใช้ข้อมูลได้อย่างรวดเร็ว ถูกต้อง รองรับธุรกรรมจำนวนมากและผู้ใช้งานพร้อมกัน

Dataset: ระบบธุรกิจค้าปลีก (Retail Business) | 12 ตาราง | 1,631,380 Records | 39 Columns

3.1 ข้อมูลหลัก (Master Data)

ตาราง	รายละเอียด	ตัวอย่างคอลัมน์	Records
customers	ข้อมูลลูกค้า	customer_id, city, signup_date	50,000
products	ข้อมูลสินค้า	product_id, category_id, supplier_id, price	10,000
categories	ประเภทสินค้า	category_id, category_name	30
suppliers	ผู้จัดจำหน่าย	supplier_id, supplier_name	200
stores	ข้อมูลสาขา	store_id, store_name	100
employees	ข้อมูลพนักงาน	employee_id, employee_name, store_id	1,000
promotions	ข้อมูลโปรโมชัน	promotion_id, promotion_name	50

3.2 ข้อมูลธุรกรรม (Transaction Data)

ตาราง	รายละเอียด	ตัวอย่างคอลัมน์	Records
orders	คำสั่งซื้อ	order_id, customer_id, store_id, order_date, promotion_id	300,000
order_items	รายการสินค้าในคำสั่งซื้อ	order_item_id, order_id, product_id, qty, price	600,000
payments	การชำระเงิน	payment_id, order_id, amount	300,000
shipments	การจัดส่ง	shipment_id, order_id, status	300,000
returns	การคืนสินค้า	return_id, order_item_id, refund	30,000

รวมข้อมูลทั้งหมด 1,631,380 Records | 12 ตาราง | 39 Columns

3.3 กระบวนการทำงานของ OLTP

1. **ลูกค้า (Customer)**
   ลูกค้าเก็บในตาราง customers โดยใช้ customer_id ระบุลูกค้าแต่ละราย
2. **การสั่งซื้อสินค้า (Order)**
   สร้างคำสั่งซื้อในตาราง orders โดยใช้ order_id และเชื่อมโยงกับ customer_id
3. **รายการสินค้าในคำสั่งซื้อ (Order Item)**
   บันทึกรายละเอียดสินค้าลงใน order_items โดยใช้ order_id เชื่อมโยงกับคำสั่งซื้อ และ product_id เชื่อมโยงกับสินค้า
4. **การชำระเงิน (Payment)**
   บันทึกการชำระเงินใน payments โดยใช้ order_id ระบุว่าการชำระเงินเกี่ยวข้องกับคำสั่งซื้อใด
5. **การจัดส่ง (Shipment)**
   สร้างข้อมูลการจัดส่งใน shipments และใช้ status ระบุสถานะการจัดส่ง
6. **การคืนสินค้า (Return)**
   บันทึกการคืนสินค้าใน returns โดยใช้ order_item_id ระบุสินค้าที่คืน

Customer → Order → Order Item → Payment → Shipment → Return

3.4 ความสัมพันธ์ของข้อมูลที่สนับสนุน OLTP

- customer_id → เชื่อมโยงลูกค้ากับคำสั่งซื้อ
- order_id → เชื่อมโยงคำสั่งซื้อกับรายการสินค้า, การชำระเงิน และการจัดส่ง
- product_id → เชื่อมโยงรายการสินค้าเข้ากับข้อมูลสินค้า
- category_id → เชื่อมโยงสินค้าเข้ากับประเภทสินค้า
- supplier_id → เชื่อมโยงสินค้าเข้ากับผู้จัดจำหน่าย
- store_id → เชื่อมโยงคำสั่งซื้อและพนักงานกับสาขา
- promotion_id → เชื่อมโยงคำสั่งซื้อกับโปรโมชั่น
- order_item_id → เชื่อมโยงรายการสินค้าเข้ากับการคืนสินค้า

3.5 เหตุผลที่ Dataset มีลักษณะเป็น OLTP

1. มีตารางที่บันทึกธุรกรรมโดยตรง เช่น orders, order_items, payments, shipments และ returns
2. มีตารางข้อมูลหลัก เช่น customers, products, stores, categories, suppliers, employees และ promotions
3. ข้อมูลมีรายละเอียดในระดับรายการธุรกรรม (เช่น order_items ระบุสินค้า จำนวน ราคา ในแต่ละคำสั่งซื้อได้)
4. มีข้อมูลจำนวนมาก รวม 1,631,380 Records โดยส่วนใหญ่เป็นข้อมูลธุรกรรม
5. ตารางต่าง ๆ เชื่อมโยงด้วยรหัสสำคัญ เช่น order_id, customer_id, product_id
6. โครงสร้างสะท้อนกระบวนการธุรกิจค้าปลีก ตั้งแต่สั่งซื้อจนถึงการคืนสินค้า

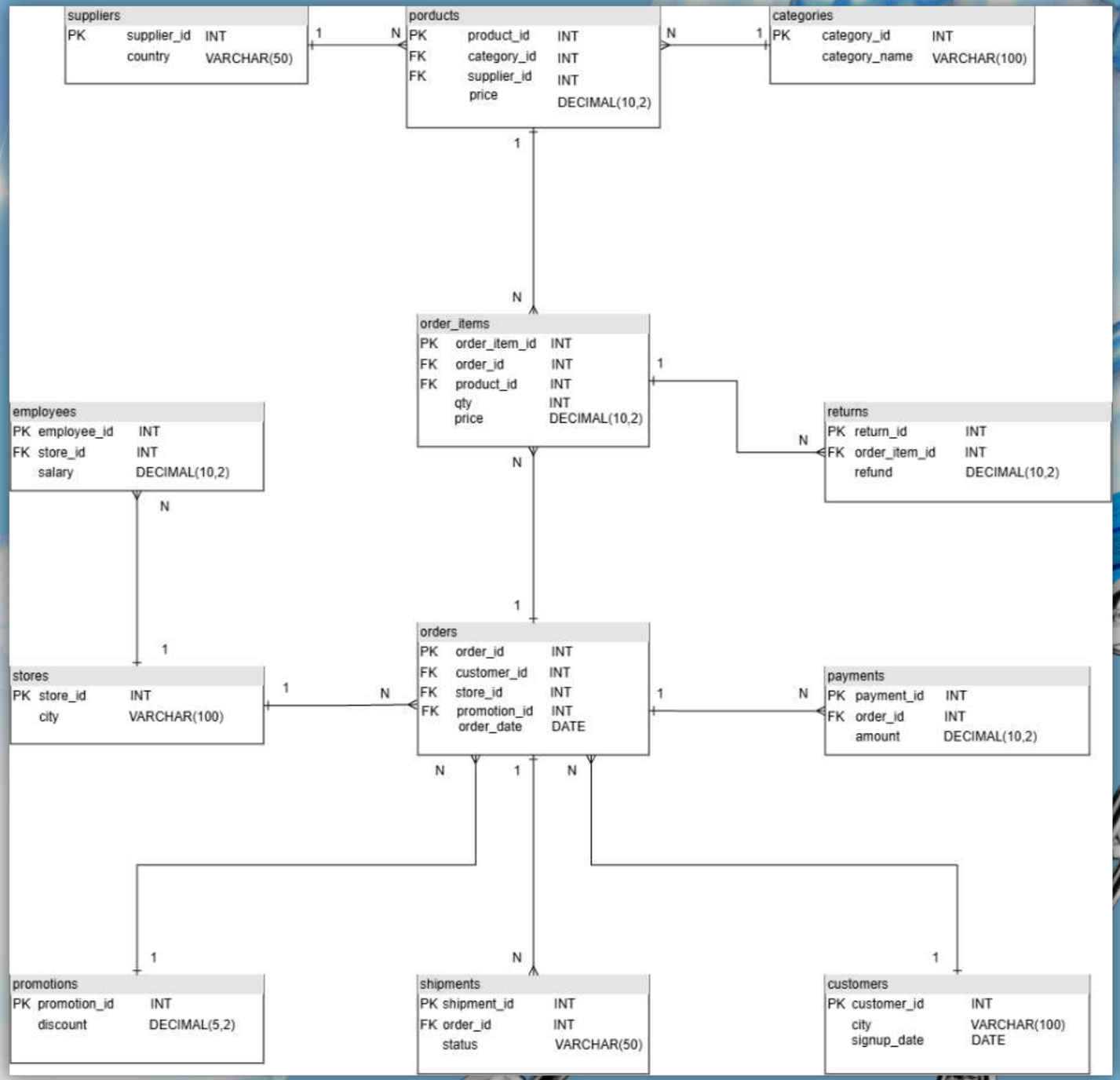
3.6 สรุปการวิเคราะห์ OLTP

Dataset นี้สอดคล้องกับ OLTP เนื่องจากเป็นข้อมูลที่ใช้สนับสนุนกระบวนการดำเนินงานของธุรกิจค้าปลีก มีทั้งข้อมูลหลักและข้อมูลธุรกรรมที่เชื่อมโยงกันอย่างเป็นระบบ โดยมุ่งเน้นการจัดเก็บข้อมูลธุรกรรมในระดับรายละเอียดของ Transaction เพื่อรองรับการทำงานประจำวันอย่างมีประสิทธิภาพ

ER

Database Relationships

Table	Relationship	Table
categories	1:N	products
suppliers	1:N	products
products	1:N	order_items
orders	1:N	order_items
order_items	1:N	returns
customers	1:N	orders
stores	1:N	orders
promotions	1:N	orders
orders	1:N	payments
orders	1:N	shipments
stores	1:N	employees



Business Questions + KPI แนวทางการวิเคราะห์

1	ยอดขายรวมของบริษัทในแต่ละเดือนเป็นอย่างไร?	**-Total Sales / -Revenue**	รวมยอดขายตามเดือน แล้วดูแนวโน้มขึ้นลงด้วย Line Chart
2	เดือนใดมียอดขายสูงสุดและต่ำสุด?	**-Monthly Revenue**	รวม Revenue รายเดือนและจัดอันดับสูงสุดไปต่ำสุด
3	สินค้าใดสร้างรายได้สูงที่สุด?	**-Product Revenue**	รวมยอดขายแยกตามสินค้า แล้วเรียงจากมากไปน้อย
4	สินค้าใดขายได้จำนวนชิ้นมากที่สุด?	**-Units Sold**	รวมจำนวนสินค้าที่ขายได้ (Quantity) ตาม Product
5	หมวดสินค้าใดทำรายได้ดีที่สุด?	**-Category Revenue**	Join Products กับ Categories แล้วเปรียบเทียบ Revenue แต่ละ Category
6	ร้านค้าใดมียอดขายดีที่สุด?	**-Store Revenue**	รวมยอดขายตาม Store แล้วจัดอันดับร้าน
7	ร้านค้าใดมียอดขายเฉลี่ยต่อคำสั่งซื้อสูงที่สุด	**-Revenue per Store / -Average Order Value**	คำนวณ Revenue จำนวน Order ของแต่ละ Store แล้วเปรียบเทียบ
8	ลูกค้ากลุ่มใดสร้างรายได้ให้บริษัทมากที่สุด?	**-Customer Revenue**	รวมยอดซื้อของลูกค้าแต่ละคน แล้วแบ่งกลุ่ม Top/Medium/Low

9	ลูกค้ากลับมาซื้อซ้ำมากน้อยแค่ไหน?	**-Repeat Purchase Rate**	คำนวณสัดส่วนลูกค้าที่มี Order มากกว่า 1 ครั้ง แล้วคำนวณเป็น % ของลูกค้าทั้งหมด
10	ลูกค้าโดยเฉลี่ยใช้จ่ายต่อคำสั่งซื้อเท่าไร?	**-Average Order Value (AOV)**	Total Revenue ÷ Number of Orders
11	โปรโมชั่นช่วยเพิ่มยอดขายหรือไม่?	**-Promotion Lift / -Revenue with Promotion**	เปรียบเทียบยอดขายช่วงมี Promotion กับไม่มี Promotion
12	โปรโมชั่นใดสร้างรายได้จากการขายมากสุด?	**-Promotion Revenue / -ROI**	รวมยอดขายของแต่ละ Promotion และ แยกตาม Promotion แล้วจัดอันดับ
13	สินค้าหรือหมวดหมู่ใดมีมูลค่าการคืนสินค้าสูงที่สุด	**-Total Refund**	Returned Quantity ÷ Sold Quantity × 100 แล้วจัดอันดับสินค้า
14	บริษัทมีการจัดส่งสินค้าตรงเวลาหรือไม่?	**-On-Time Delivery Rate**	เปรียบเทียบวันที่คาดว่าจะส่งกับวันที่ส่งจริง แล้วคำนวณ % ส่งตรงเวลา
15	Supplier รายใดสร้างรายได้จากสินค้าสูงที่สุด?	**-Supplier Revenue**	เชื่อม Supplier ⟶ Product ⟶ Order Items แล้วรวม Revenue
16	ยอดขายมีแนวโน้มเติบโตหรือหดตัวเมื่อเทียบกับช่วงก่อนหน้าหรือไม่?	**-Sales Growth Rate**	คำนวณ ((ยอดขายปัจจุบัน - ยอดขายก่อนหน้า) ÷ ยอดขายก่อนหน้า) × 100

KPI ที่สำคัญ

1.Sales

- Total Revenue
- Total Orders
- Units Sold
- Average Order Value (AOV)
- Sales Growth Rate

2.Customer

- Total Customers
- Active Customers
- Repeat Purchase Rate
- Customer Revenue

3.Product

- Top-Selling Product
- Top Revenue Product
- Category Revenue

4.Returns

- Return Rate

5.Store

- Store Revenue
- Average Order Value By Store
- Top Performing Store

6.Promotion

- Promotion Revenue
- Promotion Lift

7.Logistics

- On-Time Delivery Rate
- Average Delivery Time

8.Employee

- Sales per Employee

9.Supplier

- Supplier Revenue

10.Payment

- Payment Method Usage

Multidimensional Data Model Design

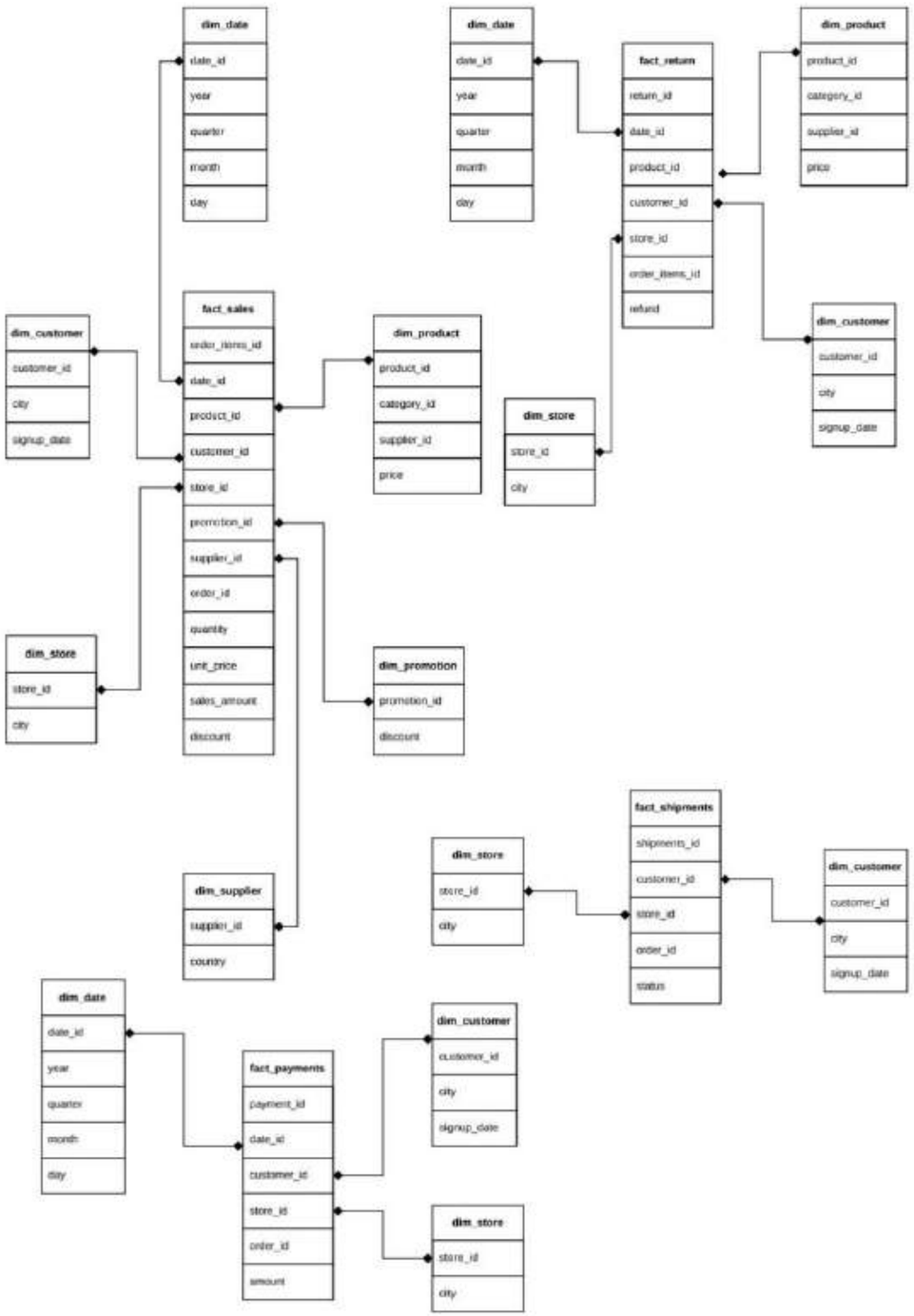
Fact Tables

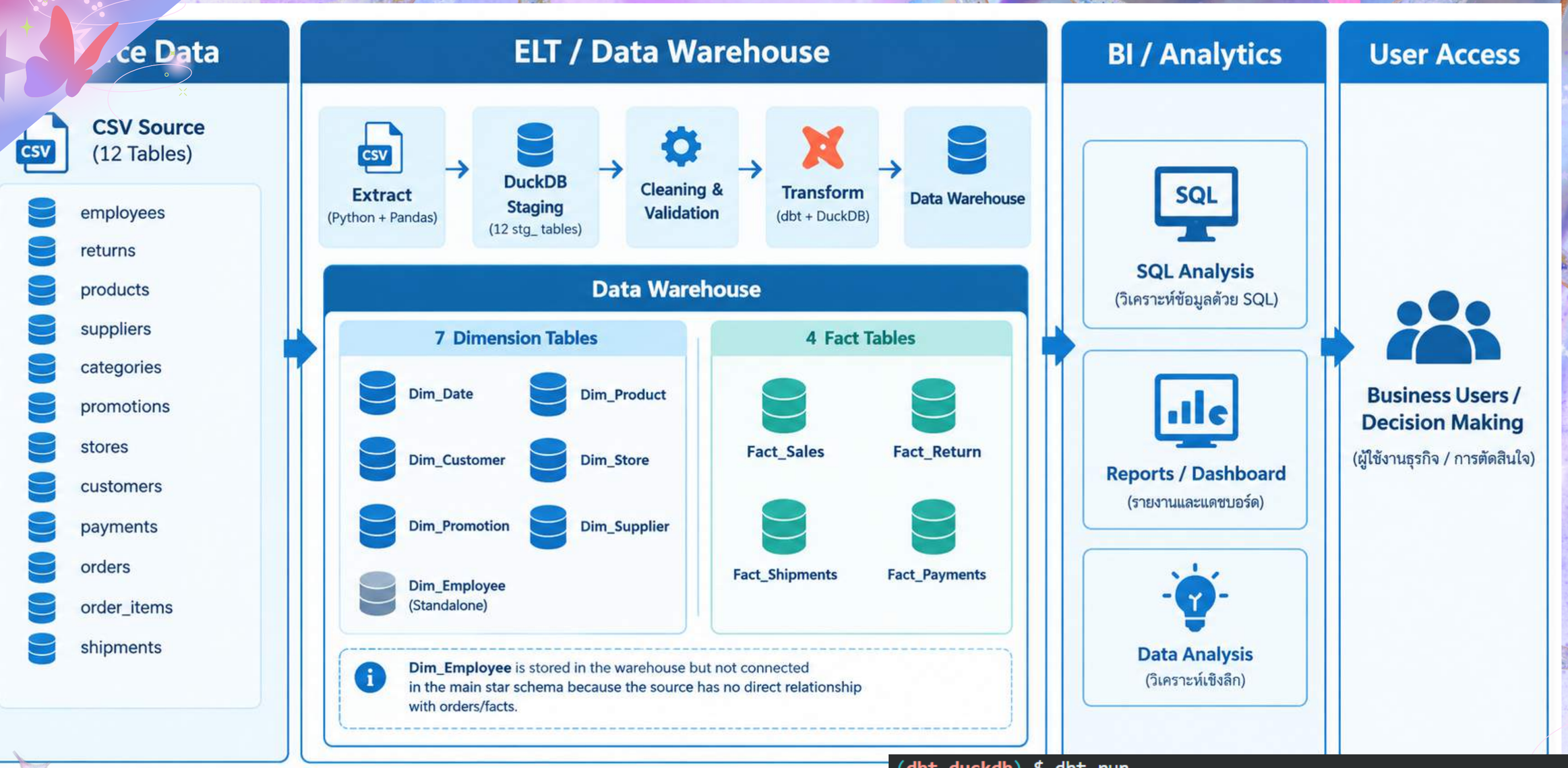
Table	Grain / Purpose	Base Measures	Measure Type	Calculated Measures
`Fact_Sales`	1 Order Line Item	`quantity`, `sales_amount`, `discount`, `unit_price`	Additive: `quantity`, `sales_amount`, `discount` / Non-Additive: `unit_price`	`Average Selling Price`, `AOV`, `Sales Growth Rate`, `Promotion Revenue`, `Supplier Revenue`
`Fact_Returns`	1 Return Transaction	`refund`	Additive	`Total Refund`, `Return Rate*`
`Fact_Shipments`	1 Shipment	`shipment_count`	Additive	`Shipment Status Rate`
`Fact_Payments`	1 Payment Transaction	`amount`	Additive	`Total Payment Amount`, `Payment Transaction Count`

Dimension Tables

Table	Type	Primary Key	Main Attributes	Purpose
`Dim_Date`	Dimension	`date_id`	`day`, `month`, `quarter`, `year`	วิเคราะห์ข้อมูลตามช่วงเวลา
`Dim_Product`	Dimension	`product_id`	`category_id`, `category_name`, `supplier_id`, `price`	วิเคราะห์สินค้าและหมวดหมู่
`Dim_Customer`	Dimension	`customer_id`	`city`, `signup_date`	วิเคราะห์ลูกค้าและพฤติกรรมการซื้อ
`Dim_Store`	Dimension	`store_id`	`city`	วิเคราะห์ยอดขายตามสาขา
`Dim_Promotion`	Dimension	`promotion_id`	`discount`	วิเคราะห์ประสิทธิภาพของ Promotion
`Dim_Supplier`	Dimension	`supplier_id`	`country`	วิเคราะห์ Supplier และรายได้จากสินค้า
`Dim_Employee`	Dimension	`employee_id`	`store_id`, `salary`	วิเคราะห์ข้อมูลพนักงานและสาขา

Data Model Diagram






```
(dbt_duckdb) $ dbt run
20:05:58  Completed successfully
20:05:58
20:05:58  Done. PASS=25 WARN=0 ERROR=0 SKIP=0 NO-OP=0 REUSED=0 TOTAL=25
```

DW_DB

สร้าง DW_DB ด้วย DuckDB สำหรับจัดเก็บข้อมูลธุรกิจค้าปลีกจำนวน 12 Core Tables และพัฒนา Web Interface ด้วย **Streamlit** เพื่อเชื่อมต่อตรวจสอบโครงสร้าง และแสดงข้อมูลจากฐานข้อมูล

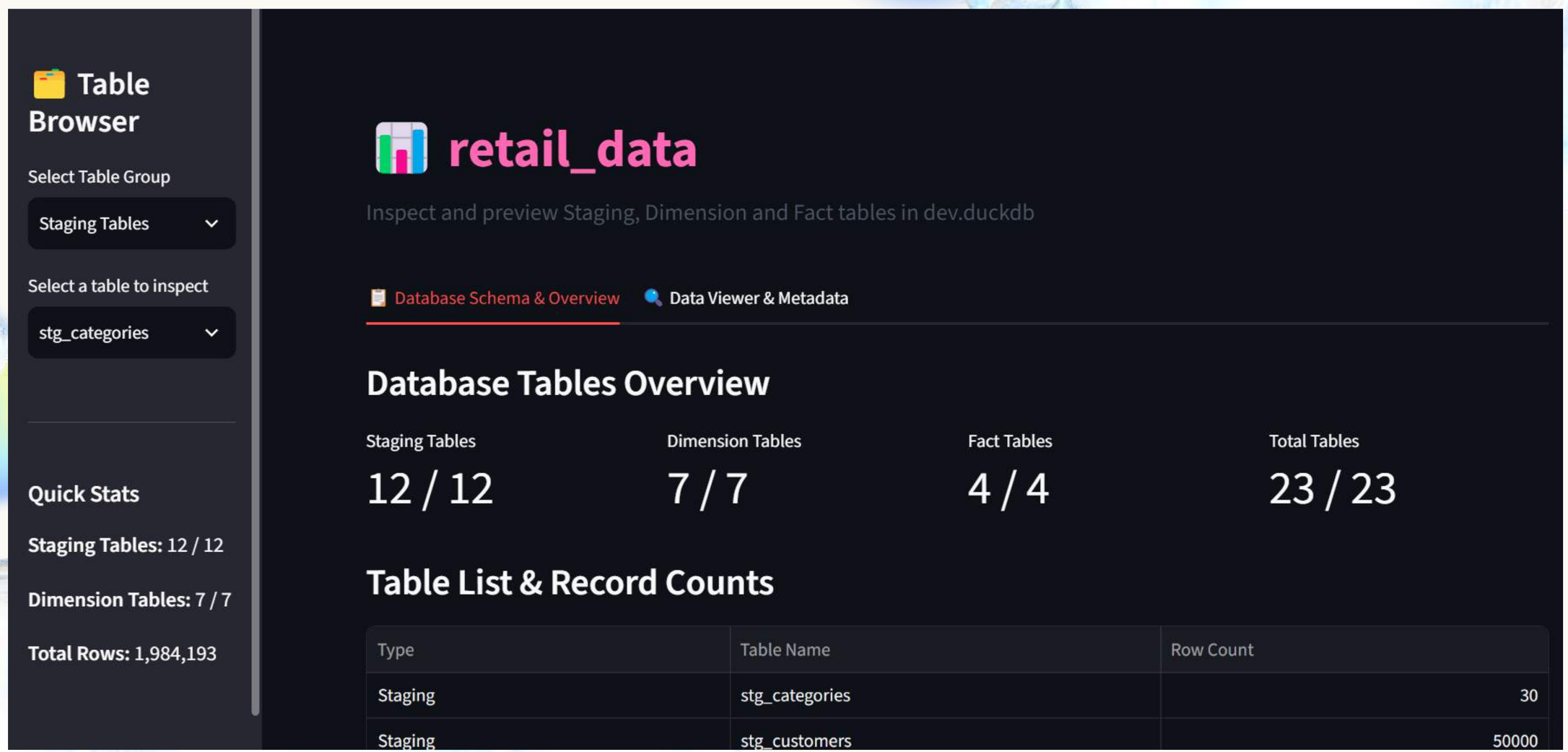
```python
STAGING_TABLES = [
    "stg_categories",
    "stg_customers",
    "stg_employees",
    "stg_order_items",
    "stg_orders",
    "stg_payments",
    "stg_products",
    "stg_promotions",
    "stg_returns",
    "stg_shipments",
    "stg_stores",
    "stg_suppliers",
]
```

```python
# ---------------------
# Fact Tables
# ---------------------

FACT_TABLES = [
    "fact_payments",
    "fact_returns",
    "fact_sales",
    "fact_shipments",
]
```

```python
DIMENSION_TABLES = [
    "dim_customer",
    "dim_date",
    "dim_employee",
    "dim_product",
    "dim_promotion",
    "dim_store",
    "dim_supplier",
]
```

DW_DB

ตรวจสอบโครงสร้างและปริมาณข้อมูลใน DW_DB
ผ่าน Web Interface โดยพบข้อมูล**ครบทั้ง 12 Core Tables**
รวม 1,591,380 Records



Dashboard_app.py

Mini-Project [Codespaces: fluffy fiesta]

[Preview] README.md | schema.yaml | dashboard_app.py 9+ | dbt_project.yml | {} manifest.json

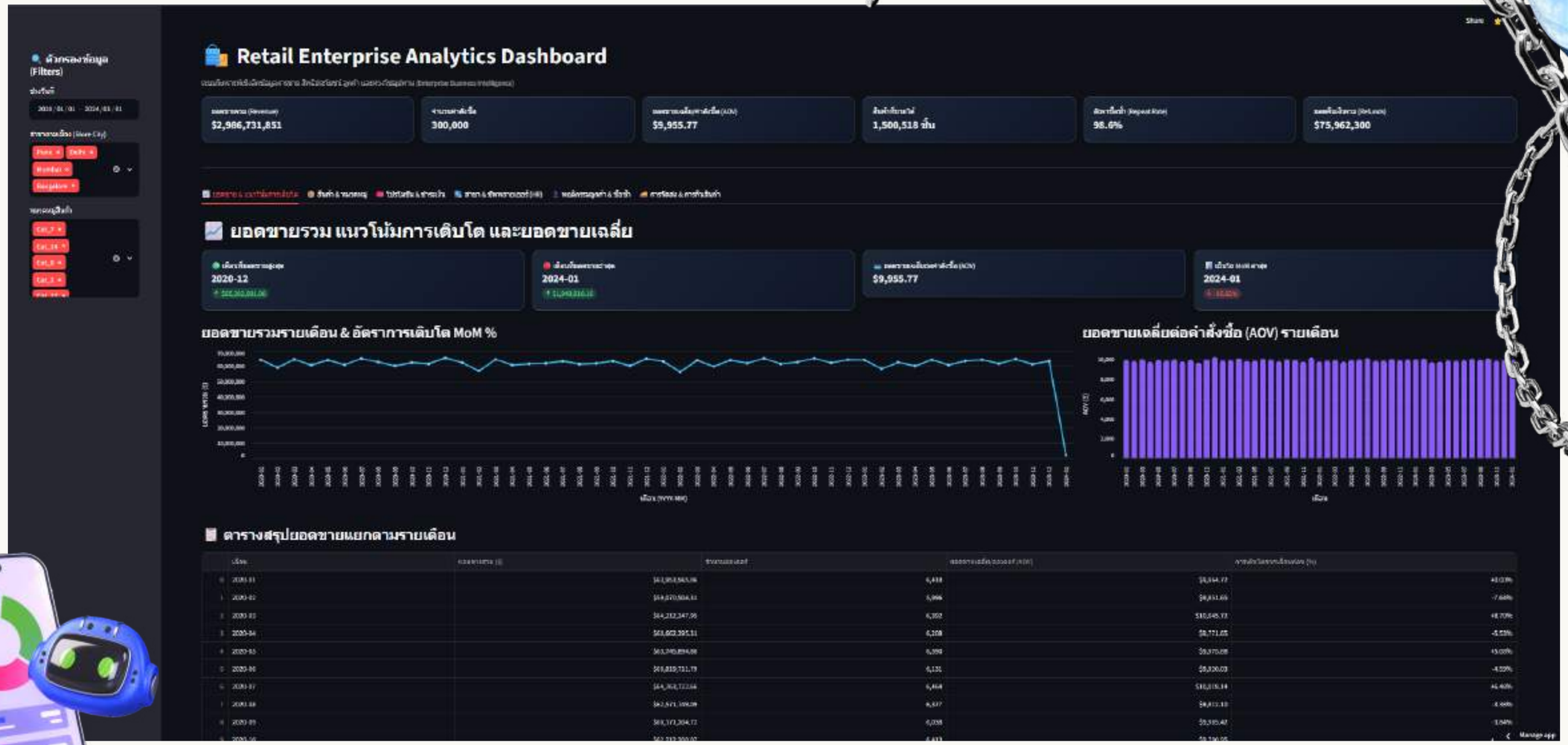
retail_data > dashboard_app.py

```python
import os
from pathlib import Path
import duckdb
import pandas as pd
import streamlit as st
import altair as alt

# --- Page Config & Modern Styling ---
st.set_page_config(
    page_title="Retail Enterprise Dashboard",
    page_icon="📦",
    layout="wide",
    initial_sidebar_state="expanded",
)

st.markdown("""
<style>
    div[data-testid="stMetric"] {
        background: linear-gradient(135deg, rgba(30, 41, 59, 0.9), rgba(15, 23, 42, 0.95));
        border: 1px solid rgba(255, 255, 255, 0.12);
        border-radius: 12px;
        padding: 16px 18px;
        box-shadow: 0 4px 12px rgba(0, 0, 0, 0.25);
        backdrop-filter: blur(8px);
    }
    div[data-testid="stMetricLabel"] > label {
        font-size: 0.85rem !important;
        font-weight: 600 !important;
        color: #94a3b8 !important;
        letter-spacing: 0.5px;
    }
    div[data-testid="stMetricValue"] > div {
```

0 ⚠ 20 | 2 | Pull Request #5 | Apichai Aimwong (1 week ago) | Ln 38, Col 33 | Spaces: 4 | UTF-8

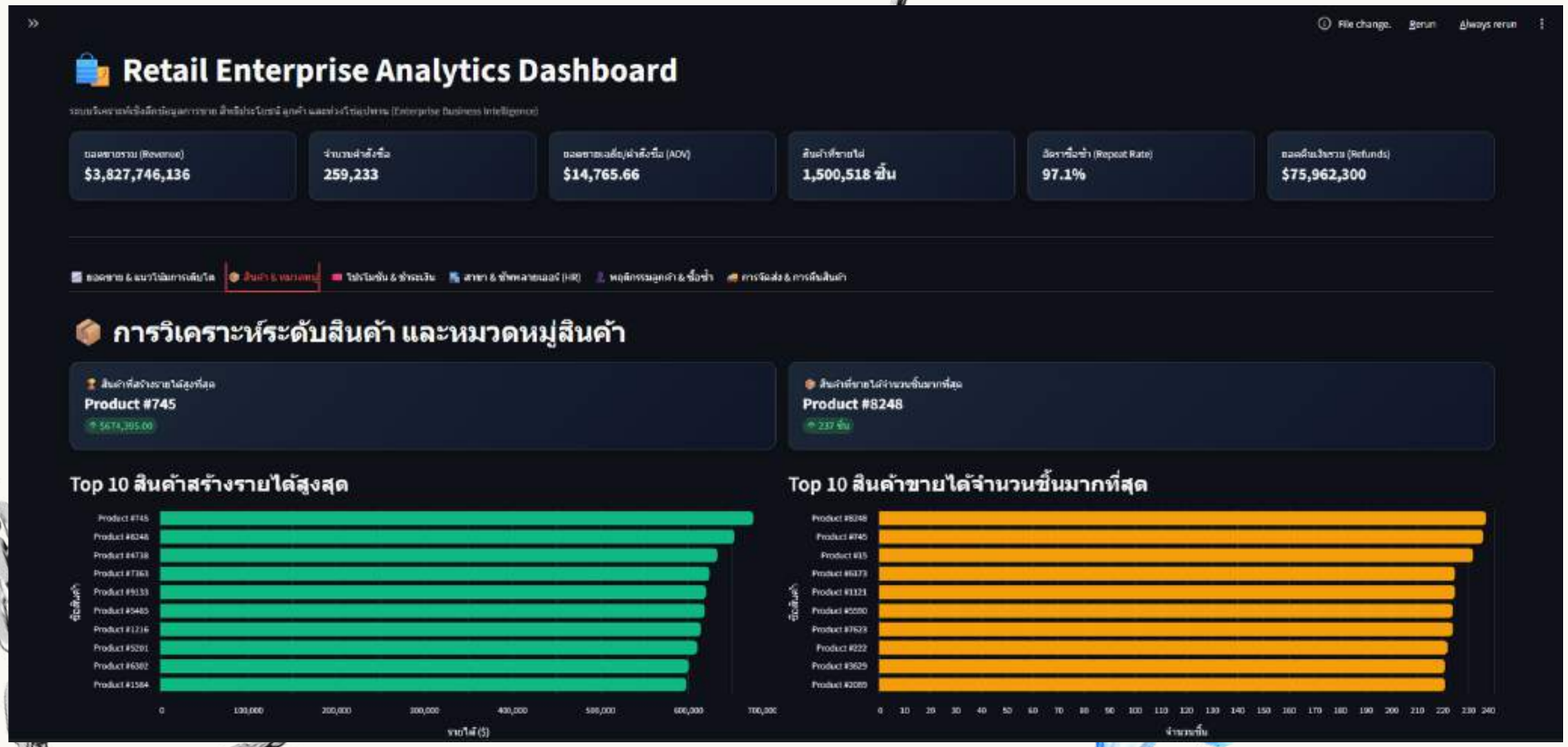
ได้ทำการสร้าง Dashboard_app.py ใน code spaces ในการสร้างได้ดูถึงข้อมูลต่างๆในแดชบอร์ดซึ่งวิเคราะห์ข้อมูลธุรกิจค้าปลีกแบบครบวงจร (OLAP + HR Analytics) เชื่อมโยงข้อมูลการขาย ขนส่ง โปรโมชัน และทรัพยากรบุคคล

Dashboard



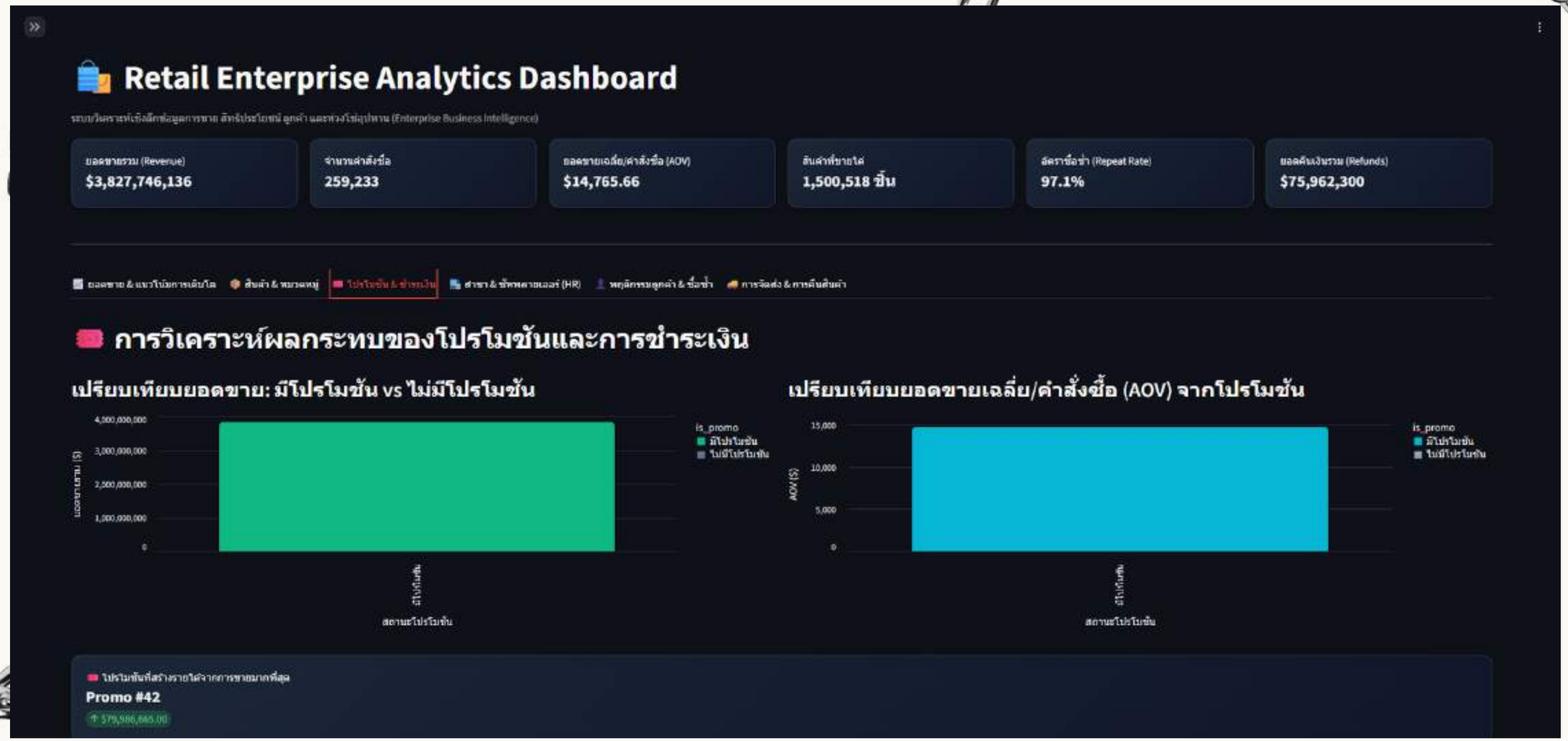
ยอดขาย & เวลา: ติดตามเทรนด์รายวัน/เดือน/ไตรมาส และหมวดหมู่สินค้าที่ทำรายได้สูงสุด

Dashboard_app.py



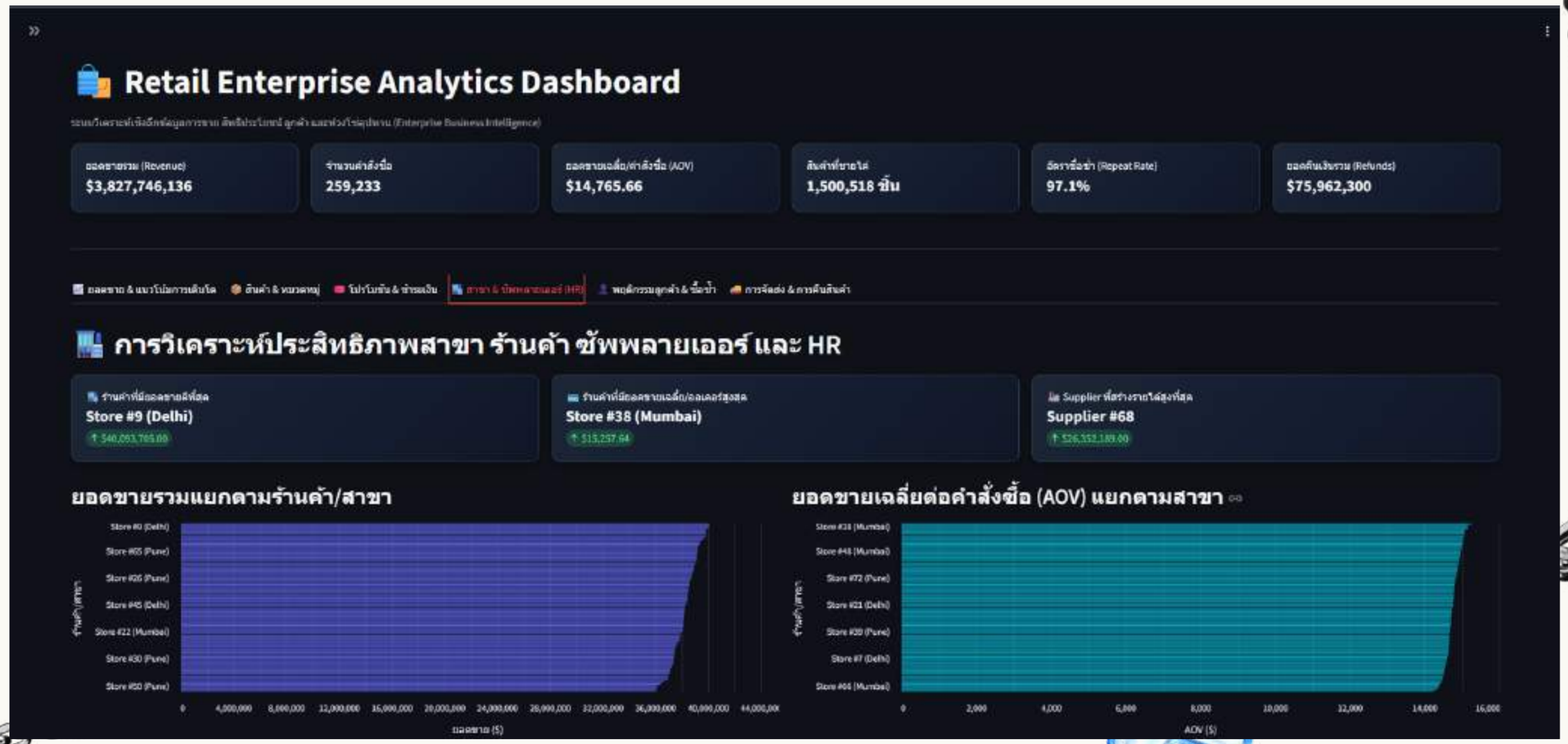
มอนิเตอร์มูลค่าเงินคืน (Refunds) แยกตามหมวดหมู่ และสัดส่วนสถานะการจัดส่ง

Dashboard_app.py



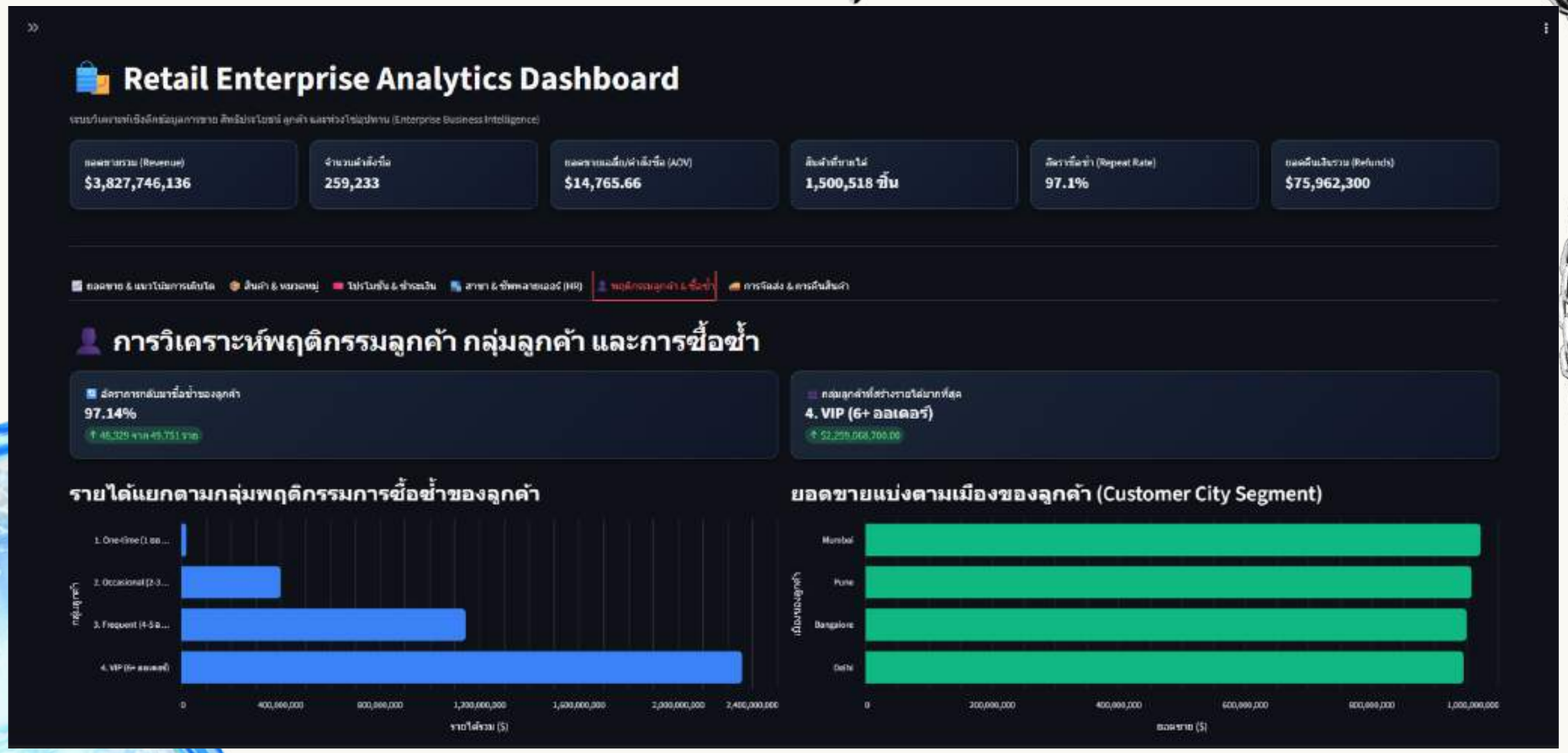
โปรโมชัน & ชำระเงิน: ประเมินผลกระทบของ % ส่วนลดต่อยอดขาย และเปรียบเทียบยอดขายกับเงินที่รับจริง

Dashboard_app.py



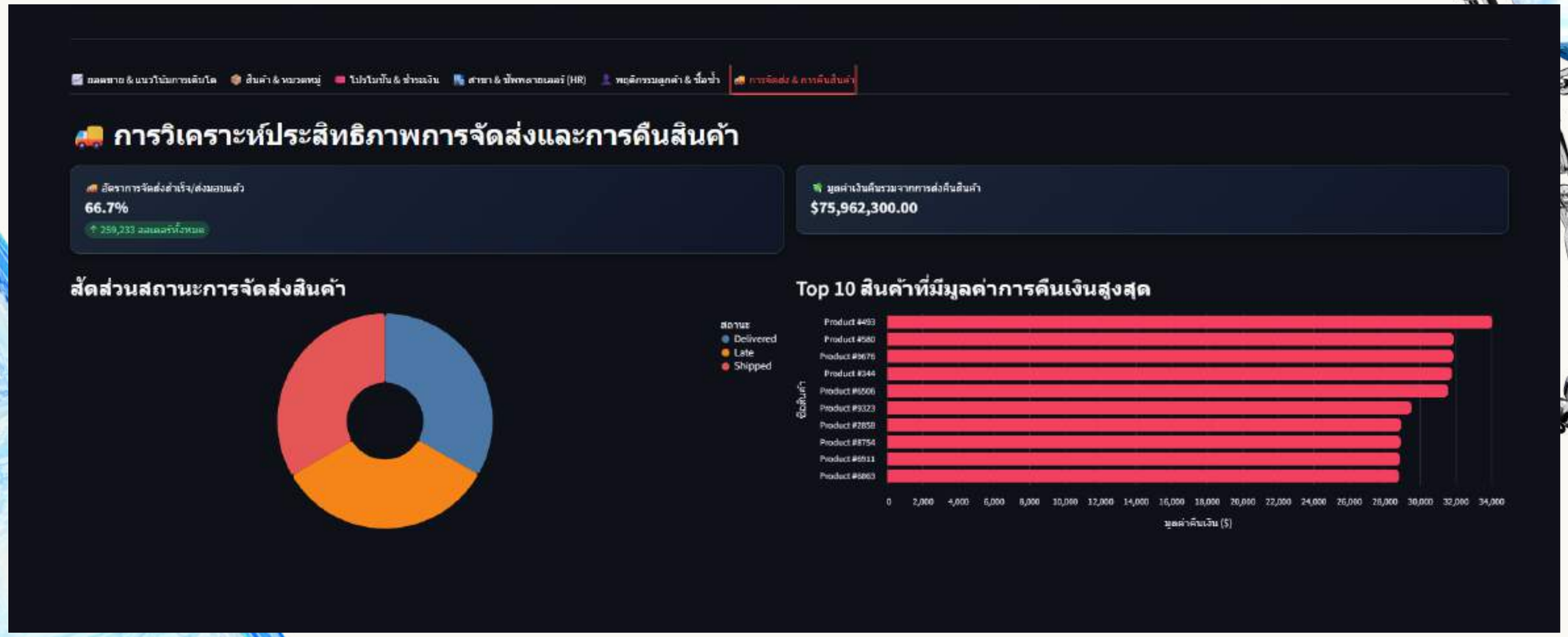
สาขา & HR: วิเคราะห์รายได้แยกตามเมือง/ประเทศซัพพลายเออร์ พร้อมวัดประสิทธิภาพพนักงานและสัดส่วนค่าจ้างต่อรายได้ (Payroll to Revenue %)

Dashboard_app.py



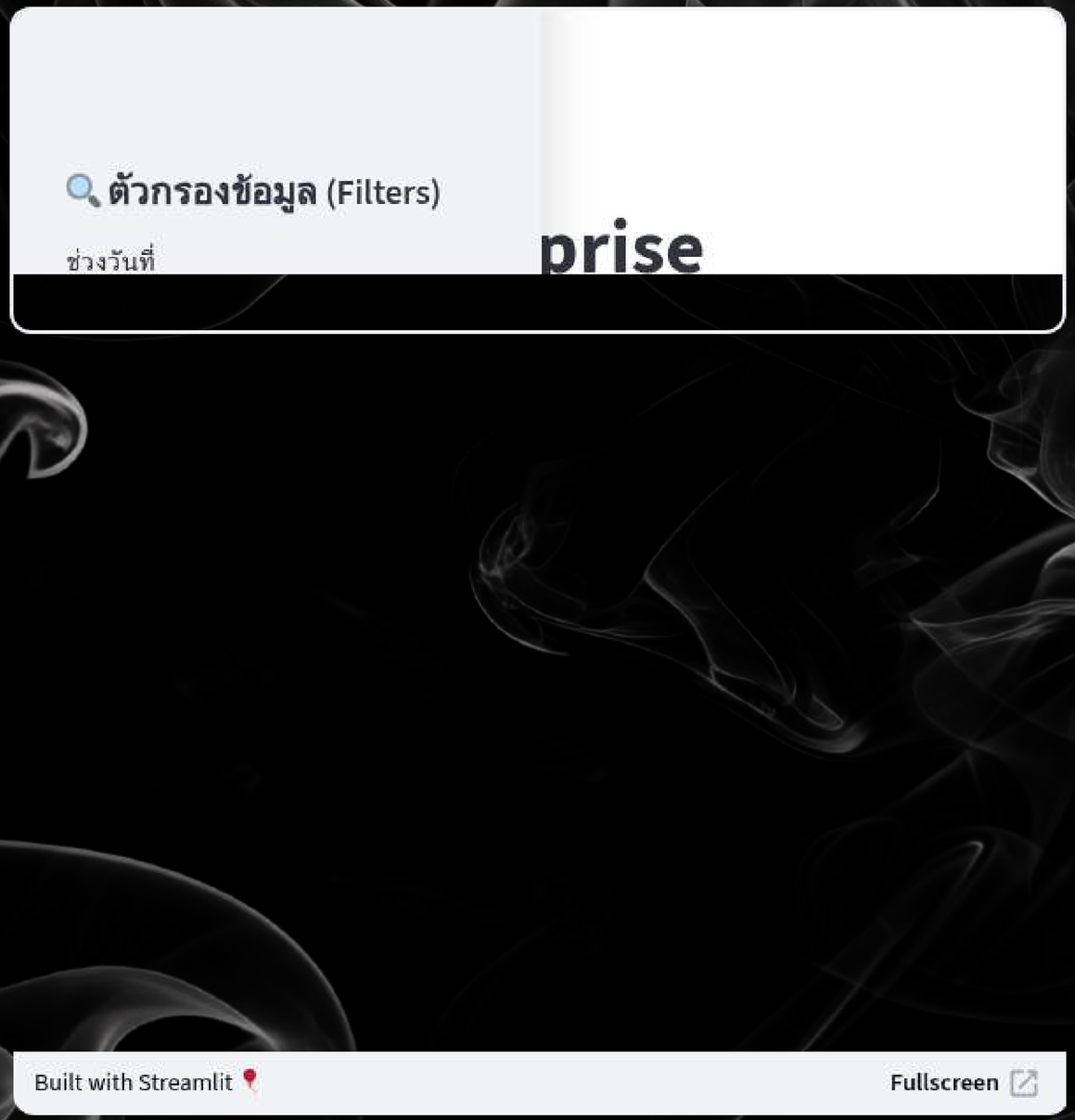
พฤติกรรมลูกค้า: จัดอันดับ Top Spenders และวิเคราะห์ ยอดซื้อตามปีที่สมัครสมาชิก (Cohort Analysis)

Dashboard_app.py



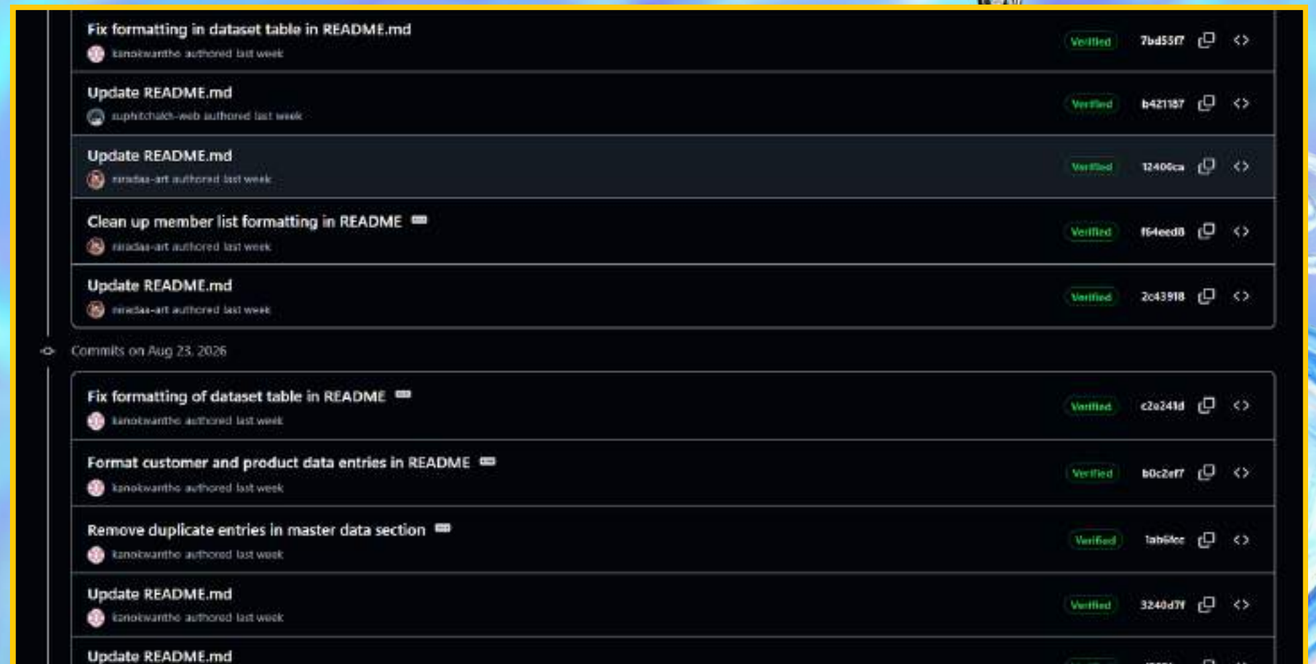
การวิเคราะห์ประสิทธิภาพการจัดส่งและการคืนสินค้า : บอกสัดส่วนสถานะการจัดส่งสินค้า

Retail Enterprise Analytics Dashboard

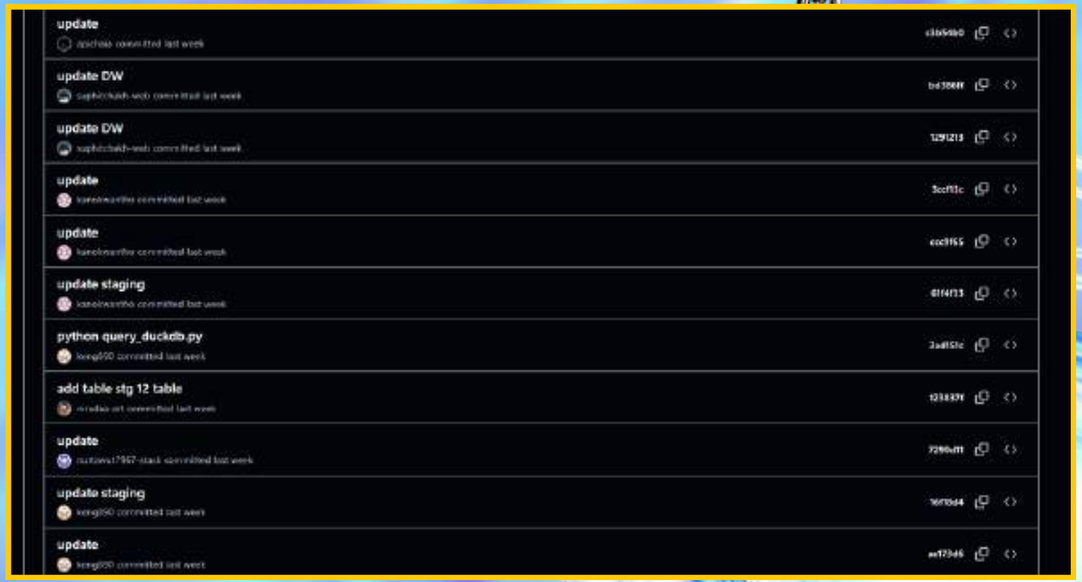


Retail Enterprise Analytics Dashboard

แสดงประวัติการทำงานใน git



แสดงประวัติการทำงานใน git



แสดงประวัติการทำงานใน git

SUMMARIZE

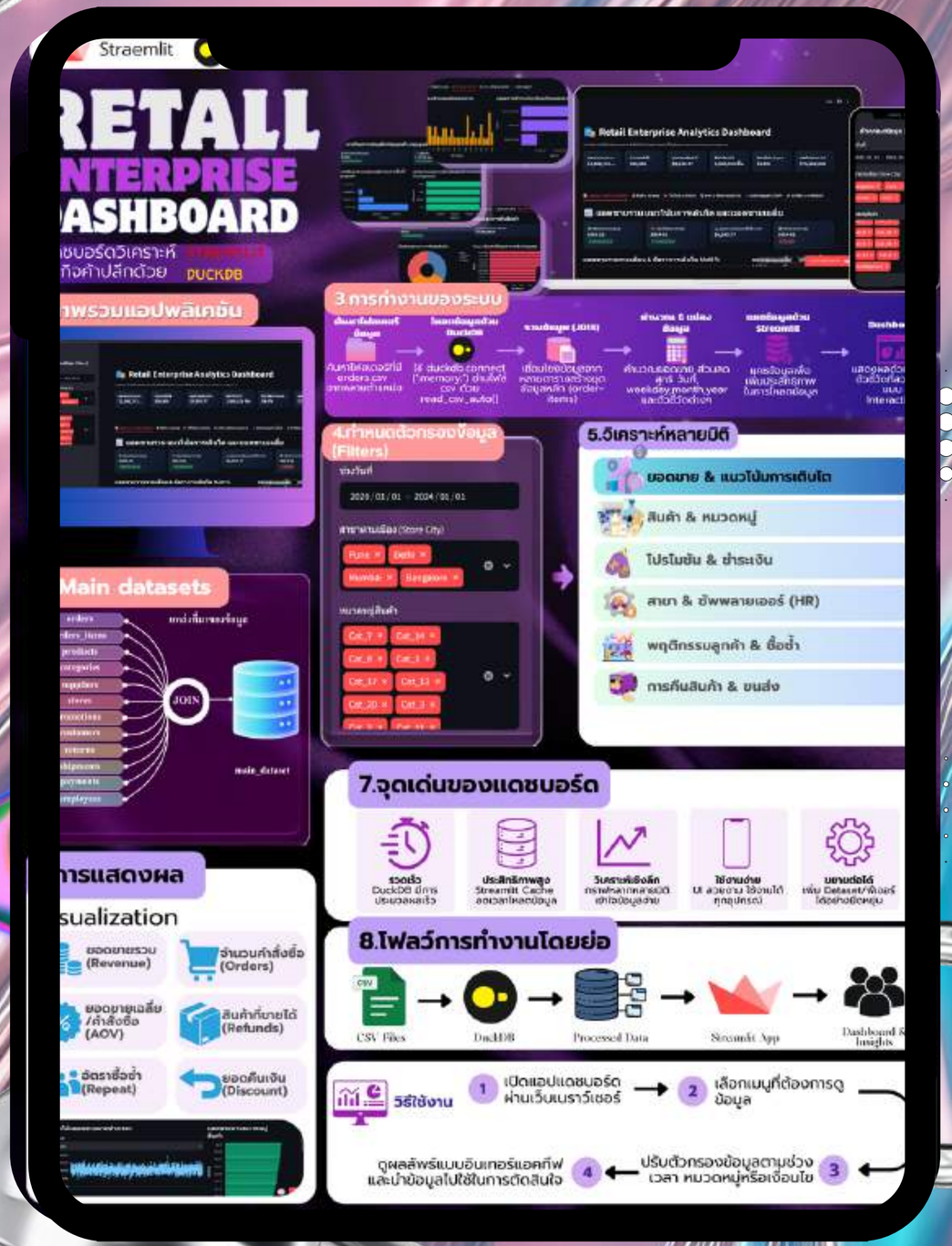
Operational Database

Business Questions

Multidimensional Data Model

ELT & dbt

Interactive Dashboard



THANK you

1.นางสาวกนกวรรณ ทองเทพ รหัสนักศึกษา 673020243-5

2.นายณัฐวุฒิ กำจัดภัย รหัสนักศึกษา 673020251-6

3.นางสาวณิรดา อนุนิวัฒน์ รหัสนักศึกษา 673020252-4

4.นายภูธิป ต้นโลห์ รหัสนักศึกษา 673020261-3

5.นางสาวสุพิชชา คำสิงห์ รหัสนักศึกษา 673020265-5

6.นายสุวิชชา ผาสุข รหัสนักศึกษา 673020267-1

7.นายอพิชัย อิ่มวงค์ รหัสนักศึกษา 673020269-7