




資料庫整理介面

1. 連接到資料庫

選擇您的資料庫類型並提供連接詳細資訊。

CSV 

選擇文件 SQLite, csv :   SuperMarket Analysis.csv

[連接](#)

2. 訓練模型(三個方法組合任意組合選取 至少選取一個方法)

✓ a. Schema分析陳述

```
"Date" TEXT,  
"Time" TEXT,  
"Payment" TEXT,  
"cogs" REAL,  
"gross margin percentage" REAL,  
"gross income" REAL,  
"Rating" REAL  
)
```

✓ b. 資料的知識背景文件

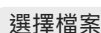

用於添加業務術語定義或其他關於資料的知識背景。

"Supermarket Sales Data: This dataset records individual transactions from a supermarket chain in Myanmar, with detailed attributes for each sale. The data includes the following columns:

Invoice ID: Unique identifier for each transaction.

Branch: The branch location of the supermarket (e.g., Yangon, Naypyitaw, Mandalay).

City: The city in which the supermarket branch is located.

✓ c. SQL 問答配對

用於提供自然語言問題與對應 SQL 查詢的配對範例。

選擇檔案 尚未選取檔案

從 SQL 生成問答

| SQL說明 | SQL語法 | 操作 |
|-------------------------------|--|----|
| <div>每個城市的總銷售額是多少？</div> | <div>-- 1. 查詢每個城市的總銷售額</div> <div>SELECT City, SUM("Sales") AS total_sales</div> <div>FROM SuperMarket_Analysis</div> <div>GROUP BY City</div> | 儲存 |
| <div>各分店的平均評分是多少？</div> | <div>-- 2. 計算每個分店的平均評分 (Rating)</div> <div>SELECT Branch, AVG(Rating) AS avg_rating</div> <div>FROM SuperMarket_Analysis</div> <div>GROUP BY Branch</div> | 儲存 |
| <div>哪十張發票的銷售額最高？</div> | <div>-- 3. 找出消費金額最高的前 10 張發票</div> <div>SELECT "Invoice ID", "Sales"</div> <div>FROM SuperMarket_Analysis</div> <div>ORDER BY "Sales" DESC</div> <div>LIMIT 10</div> | 儲存 |
| <div>男女顧客平均購買數量有何差異？</div> | <div>-- 4. 比較男女顧客的平均購買數量</div> <div>SELECT Gender, AVG(Quantity) AS avg_qty</div> <div>FROM SuperMarket_Analysis</div> <div>GROUP BY Gender</div> | 儲存 |
| <div>各城市不同產品線的總銷售額是多少？</div> | <div>-- 5. 計算不同產品線在各城市的總銷售額</div> <div>SELECT City, "Product line", SUM("Sales") AS total_sales</div> <div>FROM SuperMarket_Analysis</div> <div>GROUP BY City, "Product line"</div> | 儲存 |
| <div>每個分店毛利總額最高的產品線是什麼？</div> | <div>-- 6. 找出每個分店毛利率最高的產品線</div> <div>WITH branch_profit AS (</div> <div> SELECT Branch, "Product line", SUM("gross income") AS total_income</div> <div> FROM SuperMarket_Analysis</div> <div> GROUP BY Branch, "Product line"</div> <div>)</div> <div>SELECT *</div> <div>FROM branch_profit bp</div> <div>WHERE total_income = (</div> <div> SELECT MAX(total_income)</div> <div> FROM branch_profit b2</div> <div> WHERE b2.Branch = bp.Branch</div> <div>)</div> | 儲存 |
| | | |

| | | |
|---|---|----|
| 請問每個城市中，按銷售額排列後各發票的排名是多少？ <div>✓</div> | -- 7. 對每個城市的發票按銷售額排名 SELECT "Invoice ID", City, "Sales", RANK() OVER (PARTITION BY City ORDER BY "Sales" DESC) AS rank_in_city FROM SuperMarket_Analysis | 儲存 |
| 哪些產品線的平均評分超過 7.5？ <div>✓</div> | -- 8. 查找平均評分大於 7.5 的產品線 SELECT "Product line", AVG(Rating) AS avg_rating FROM SuperMarket_Analysis GROUP BY "Product line" HAVING AVG(Rating) > 7.5 | 儲存 |
| 2025-01 每日的總銷售額是多少？ <div>✓</div> | -- 9. 查出 2025-01 每日的總銷售額 SELECT DATE("Date") AS sales_day, SUM("Sales") AS total_sales FROM SuperMarket_Analysis WHERE "Date" LIKE '2025-01%' GROUP BY DATE("Date") | 儲存 |
| 各顧客類型中，消費金額≥500元的顧客與<500元的顧客各有多少人？ <div>✓</div> | -- 10. 分析顧客類型的高消費 vs 低消費比例 SELECT "Customer type", SUM(CASE WHEN "Sales" >= 500 THEN 1 ELSE 0 END) AS high_spenders, SUM(CASE WHEN "Sales" < 500 THEN 1 ELSE 0 END) AS low_spenders FROM SuperMarket_Analysis GROUP BY "Customer type" | 儲存 |

訓練模型

3. 提出問題

描述這個資料集

提問

思考過程：

****SuperMarket_Analysis 資料集概覽****

| 主要欄位 | 資料型別 | 商業意義 | 典型值範例 |
|-------------------------|------|-----------------------|---------------|
| ----- ----- ----- ----- | | | |
| `Invoice ID` | TEXT | 每筆交易的唯一識別碼（保留前導零與連字元） | `INV-0012345` |
| `Branch` | TEXT | 店鋪編號（用於地理區段分析） | `BR01` |
| `City` | TEXT | 所在城市 | `Yangon` |

```
| `Customer type` | TEXT | 顧客分類 (Member / Normal) | `Member` |
| `Gender` | TEXT | 顧客性別 | `Male` |
| `Product line` | TEXT | 商品大類別 | `Health & Beauty` |
| `Unit price` | REAL | 單位售價 (不含稅) | `4.99` |
| `Quantity` | INTEGER | 購買數量 | `3` |
| `Tax 5%` | REAL | 5% 稅金 | `0.75` |
| `Sales` | REAL | 總銷售額 (含稅) | `15.24` |
| `Date` | TEXT | 交易日期 (YYYY-MM-DD) | `2025-10-01` |
| `Time` | TEXT | 交易時間 (HH:MM:SS) | `14:30:05` |
| `Payment` | TEXT | 付款方式 | `Cash` |
| `cogs` | REAL | 成本 (Cost of Goods Sold) | `8.50` |
| `gross margin percentage` | REAL | 毛利率 ((Sales-cogs)/Sales*100) | `44.44` |
| `gross income` | REAL | 毛利 (Sales-cogs) | `6.74` |
| `Rating` | REAL | 顧客滿意度評分 (0-5) | `4.2` |
```

****商業用途****

- ****營收與毛利分析****：計算不同分店、城市、產品線、支付方式的總銷售與毛利。
- ****顧客分群****：區分會員與非會員，分析交易行為與滿意度。
- ****時間趨勢****：根據 `Date` 或 `Time` 欄位觀察每日、每週或每小時銷售波動。
- ****產品線績效****：評估各類別的平均毛利率與銷售貢獻。
- ****支付方式效益****：比較各付款方式的毛利差異，判斷最利潤高的通道。

****數據完整性規則****

- `Invoice ID` 必須唯一（可設為主鍵）。
- `Tax 5%` 應等於 $\text{Unit price} * \text{Quantity} * 0.05$ 。
- `gross margin percentage` 應等於 $((\text{Sales} - \text{cogs}) / \text{Sales}) * 100$ 。
- `Rating` 限 0-5。
- 所有金額欄位應為非負數。

****維護與安全****

- 每日 ETL 進行資料導入與衍生欄位計算。
- 建議對 `Date`、`Branch`、`Product line` 等欄位建立索引以加速查詢。
- 資料庫層級加密、傳輸加密與最小權限原則保護敏感資訊。

這份資料集是零售分析、財務報表與 BI 報表的核心來源，可直接用於 Power BI、Tableau 等可視化工具，支援決策制定與績效監控。

生成的 SQL：

結果：