

# Superstore 数据分析报告（SQL 项目）

## 1. 项目概述（Project Overview）

- 项目名称：Superstore 数据分析项目
- 项目工具：MySQL + DBeaver + Excel
- 数据来源：Kaggle（Superstore 数据集）
- 数据规模：约 9,000 行订单数据，包含客户、商品、地区、销售额、利润等字段
- 项目目标：
  - 熟练掌握 SQL 查询与分析能力
  - 提取有价值的业务洞察（销售、利润、客户行为）
  - 输出可复用的 SQL 模板与项目经验

## 2. 数据预处理（Data Preparation）

- 原始文件：CSV 文件，包含订单明细信息
- 处理步骤：
  - 使用 MySQL 创建数据库 superstore，导入 orders.csv 表
  - 字段命名规范化（如 Order Date → order\_date）
  - 类型转换（如将 order\_date 从 VARCHAR 转换为 DATE）

示例代码：

```
WITH orders_clean AS (  
    SELECT STR_TO_DATE(order_date, '%m/%d/%Y') AS order_dt,  
           sales  
    FROM orders  
)
```

## 3. 分析方向与问题设计（Business Questions）

本项目分为两个阶段，共设计了 40+ 道分析题目，覆盖以下方向：

类别	代表问题
销售趋势	各月份销售额趋势如何？
地域对比	哪些地区的利润率更高？
客户行为	哪些客户连续多年来有订单？
产品分析	哪些子类销售额高但利润低？
时间序列	环比增长、同比趋势如何？

## 4. 样例分析展示（示例结构）

### 02\_queries.Q1. 按月统计销售额和利润额

分析目标：识别年度中销售的旺季与淡季

- **SQL 简要逻辑：**
  - 使用 DATE\_FORMAT(order\_dt, '%Y-%m') 转换为年月；
  - 按年月聚合销售额及利润额，按时间排序；
- **关键 SQL 片段：**

```
WITH orders_clean AS (
    SELECT STR_TO_DATE(order_date, '%m/%d/%Y') AS order_dt,
           sales,
           profit
    FROM orders
)
SELECT DATE_FORMAT(order_dt, '%Y-%m') AS ym,
       SUM(sales) AS total_sales,
       SUM(profit) AS total_profit
FROM orders_clean
GROUP BY DATE_FORMAT(order_dt, '%Y-%m')
ORDER BY ym;
```

- **结果简述：**
  - 销售额在每年 11~12 月呈现明显上涨；
  - 季节性波动明显，建议提前做促销计划；
  - 销售额与利润额波动趋势并不完全一致，需分析具体品类的贡献

## 5. 高价值客户行为分析

### 02\_queries.Q12. 哪些客户多次复购?

- 逻辑概述：
  - 对每位客户统计 DISTINCT 的订单数；
  - 使用 HAVING 筛选 > 5 次；
- SQL 示例：

```
SELECT customer_id,  
       COUNT(DISTINCT order_id) AS order_count  
FROM orders  
GROUP BY customer_id  
HAVING order_count > 5  
ORDER BY order_count DESC, customer_id;
```

- 结果说明：
  - 共有 464 位客户下单超过 5 次；
  - 这些客户多次下单，忠诚度高于整体客户群，建议针对性营销。

---

## 6. 可视化图表展示

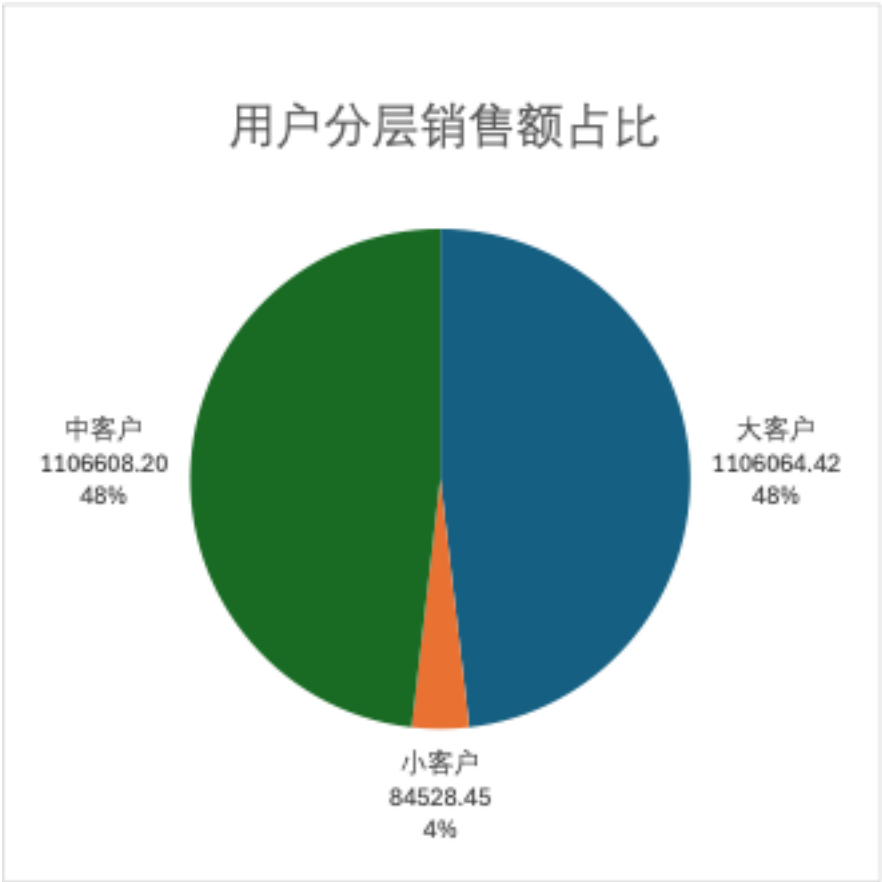
- 月份销售额及利润额趋势（折线图）



- 月份销售额及环比、同比增长率趋势（折线图）



- 用户分层销售额占比（饼图）



---

## 7. 结论总结 (Key Insights)

- 📌 每年 11 月、12 月为销售高峰;
- 📌 西部地区利润率相对较高, 建议加强投放;
- 📌 某些产品子类销售额高但利润负值, 存在成本结构问题;
- 📌 用户分层分析显示高价值客户(销售额前 20%的用户)贡献 48% 左右的销售额。

---

## 8. 项目亮点与反思 (Reflection)

- 熟练运用了 CTE、窗口函数、子查询等中高级 SQL 技术;
- 编写了两阶段共 40+ 道 SQL 题, 涵盖多种分析需求;
- 输出了可复用的 SQL 模板、命名规范、查询清单;
- 遗留改进方向: 可结合 BI 工具输出更完整可视化结果。

---

## 9. 附录 (Appendix)

- 所有 SQL 查询脚本: query1.sql, query2.sql
- 数据来源: Kaggle - Superstore
- GitHub 项目地址: [https://github.com/AureSong/sql-projects/tree/main/03\\_mock\\_projects/superstore\\_sql\\_parctice](https://github.com/AureSong/sql-projects/tree/main/03_mock_projects/superstore_sql_parctice)
- 项目用时: 7 天, 含数据导入、清洗、分析、文档撰写