Uncovering Coffee Trends with SQL

用 SQL 挖掘咖啡趋势

- ## Background(背景)

The coffee industry shows distinct trends in Southeast Asia and North America. Southeast Asia is experiencing rapid expansion, with local brands like Café Amazon, ZUS Coffee, and Kopi Kenangan scaling quickly through extensive outlet growth and high sales volume. In contrast, North America is a mature market where chains like Starbucks are adapting through store renovations and strategic shifts, while rising costs lead to price increases. These trends highlight the importance of tracking key metrics such as store counts, sales volume, and regional expansion.

咖啡行业在东南亚和北美呈现显著差异。东南亚市场快速扩张，本土品牌如 Café Amazon、ZUS Coffee 和 Kopi Kenangan 通过大量开店和惊人销量迅速崛起；北美市场则趋于成熟，星巴克等品牌通过门店改造和策略调整应对竞争，同时成本上涨导致零售价上调。这些趋势凸显了对门店数量、销量及区域扩张等关键指标进行数据分析的重要性。

- ## Data Sources(数据来源)

1、Baristacoffeesalestbl

Sourced from

https://www.kaggle.com/datasets/yashparab/barista-coffee-sales-data-for-eda-csv

2、caffeine_intake_tracker

Sourced from https://www.kaggle.com/datasets/prekshad2166/caffeine-intake-tracker-csv

3、coffeesales

Sourced from https://www.kaggle.com/datasets/visabelsarahsargunar/coffee-sales-dataset

4、consumerpreference

Sourced from https://datadryad.org/dataset/doi:10.25338/B8993H

5、list_coffee_shops_in_kota_bogor

Sourced from https://www.kaggle.com/datasets/mrbuitenzorg/list-coffee-shops-in-kota-bogor

6、.top-rated-coffee

Sourced from https://www.kaggle.com/datasets/asimmahmudov/top-rated-coffee

- **Task List(任务列表)**

1. Count the number of product categories and records per category (Table: baristacoffeesalesTBL)

统计产品类别数量及各类别记录数（表：baristacoffeesalesTBL）

2. Group by customer gender and loyalty member type, show record counts and repeat customer distribution (Table: baristacoffeesalesTBL)

按顾客性别和会员类型分组，统计记录数及复购用户分布（表：baristacoffeesalesTBL）

3. Sum total amount by product category and customer discovery source; compare two query versions and explain differences (Table: baristacoffeesalestbl)

按产品类别和客户发现来源汇总总金额，比较两种查询结果差异（表：baristacoffeesalesTBL）

4. For coffee consumption, show average focus level and sleep quality by time of day and gender (Table: caffeine_intake_tracker)

分时段和性别统计饮用咖啡后的平均专注度与睡眠质量（表：caffeine_intake_tracker）

5. Identify and list problematic data records (Table: list_coffee_shops_in_kota_bogor)

识别并列出数据中存在问题的记录（表：list_coffee_shops_in_kota_bogor）

6. Calculate spending before and after 12:00, note data issues and handling method (Table: coffeesales)

统计中午 12 点前与后的消费金额，指出数据问题并说明处理方法（表：coffeesales）

7. Group by pH ranges, compute average Liking, FlavorIntensity, Acidity, and Mouthfeel (Table: consumerpreference)

按 pH 值区间分组，计算喜好度、风味强度、酸度与口感的平均值（表：consumerpreference）

8. Join four tables, list top 3 store-shop combinations by total sales per month (Tables: coffeesales, list_coffee_shops_in_kota_bogor, top-rated-coffee, baristacoffeesalestbl)

联合四表，按月统计销售额最高的前 3 家店铺组合（表：coffeesales、list_coffee_shops_in_kota_bogor、top-rated-coffee、baristacoffeesalestbl）