■■ Retail Sales Analysis — SQL Project

Beginner Level | Database: p1_retail_db

■ Project Overview

This project demonstrates SQL skills and techniques typically used by data analysts to explore, clean, and analyze retail sales data. It involves setting up a retail sales database, performing exploratory data analysis (EDA), and answering specific business questions through SQL queries. It is ideal for beginners looking to build a solid foundation in SQL.

■ Objectives

1. Set up a retail sales database: Create and populate a retail sales database with the provided sales data.
2. Data Cleaning: Identify and remove any records with missing or null values.
3. Exploratory Data Analysis (EDA): Perform basic exploratory data analysis to understand the dataset.
4. Business Analysis: Use SQL to answer specific business questions and derive insights from the sales data.

■■ Project Structure

1. Database Setup

```
CREATE DATABASE p1_retail_db;

CREATE TABLE retail_sales
(
    transactions_id INT PRIMARY KEY,
    sale_date DATE,
    sale_time TIME,
    customer_id INT,
    gender VARCHAR(10),
    age INT,
    category VARCHAR(35),
    quantity INT,
    price_per_unit FLOAT,
    cogs FLOAT,
    total_sale FLOAT
);
```

2. Data Exploration & Cleaning

```
SELECT COUNT(*) FROM retail_sales;
SELECT COUNT(DISTINCT customer_id) FROM retail_sales;
SELECT DISTINCT category FROM retail_sales;

DELETE FROM retail_sales
WHERE sale_date IS NULL OR sale_time IS NULL OR customer_id IS NULL OR
      gender IS NULL OR age IS NULL OR category IS NULL OR
      quantity IS NULL OR price_per_unit IS NULL OR cogs IS NULL;
```

3. Data Analysis & Business Insights

```sql
-- Retrieve all columns for sales made on '2022-11-05'
SELECT * FROM retail_sales WHERE sale_date = '2022-11-05';

-- Find transactions for Clothing with quantity > 4 in Nov 2022
SELECT * FROM retail_sales
WHERE category = 'Clothing' AND TO_CHAR(sale_date, 'YYYY-MM') = '2022-11' AND quantity >=

-- Total sales per category
SELECT category, SUM(total_sale) AS net_sale, COUNT(*) AS total_orders FROM retail_sales

-- Average age of Beauty category buyers
SELECT ROUND(AVG(age), 2) AS avg_age FROM retail_sales WHERE category = 'Beauty';

-- Transactions where total_sale > 1000
SELECT * FROM retail_sales WHERE total_sale > 1000;

-- Transactions by gender in each category
SELECT category, gender, COUNT(*) AS total_trans
FROM retail_sales
GROUP BY category, gender
ORDER BY category;

-- Category with highest revenue
SELECT category, SUM(total_sale) AS total_revenue
FROM retail_sales
GROUP BY category
ORDER BY total_revenue DESC
LIMIT 1;

-- Highest selling month based on sales amount
SELECT EXTRACT(MONTH FROM sale_date) AS month, SUM(total_sale) AS total_revenue
FROM retail_sales
GROUP BY month
ORDER BY total_revenue DESC;

-- Best customer based on total sales amount
SELECT customer_id, SUM(total_sale) AS total_revenue
FROM retail_sales
GROUP BY customer_id
ORDER BY total_revenue DESC
LIMIT 1;
```

■ Key Learnings

• Strengthened SQL querying skills including filtering, grouping, and aggregation.
• Learned data cleaning techniques and data validation using SQL.
• Derived actionable insights from raw retail sales data.
• Gained hands-on experience with database creation and analysis workflow.
• Improved ability to interpret and visualize business trends from query results.

■ Conclusion

This project helped solidify my understanding of SQL and its application in real-world business data scenarios. Through systematic exploration and query-based problem-solving, I developed the ability to extract and communicate meaningful insights from data.