Sales Data Analysis Project Report

Introduction

This project focuses on analyzing sales data using SQL to extract actionable insights. The goal is to answer diverse business questions, from basic data retrieval to complex analytical challenges, showcasing a strong command of SQL for data analysis.

Dataset Description

The project is based on a relational database named **`sales`**, which consists of five interconnected tables:

- **customers**: This table contains information about customers, including a unique `customer_code`, `customer_name`, and `customer_type` (e.g., 'Brick & Mortar', 'E-Commerce').
- **date**: This table provides date-related details, which can be utilized for time-series analysis.
- **markets**: This table encompasses data pertaining to various markets, such as `markets_code`, `markets_name`, and their corresponding `zone`.
- **products**: This table stores information regarding products, including `product_code` and `product_type`.
- **transactions**: This central table records all sales transactions. It is linked to the other tables via foreign keys and includes metrics such as `sales_qty`, `sales_amount`, `profit_margin_percentage`, and `cost_price`.

Data Analysis and Insights

The data analysis is structured into three sections based on the complexity of the questions: Easy, Medium, and Hard. The entirety of our comprehensive set of questions and queries can be found within the accompanying SQL script file. This following section, however, is dedicated to showcasing a curated selection of examples that stand out due to their complexity, the interesting challenges they present, or the innovative solutions they embody.

Project Highlights: Key Questions & SQL Solutions -

Tracking Growth: Total Transactions Per Year (Level - “Medium”) (Query no. 5)

Question: How many transactions were there each year?

Business Insight: This query provides a high-level view of business activity over time. By tracking the total number of transactions each year, leadership can quickly assess growth, identify trends, and understand the overall trajectory of the business. A steady increase in transactions year-over-year is a strong indicator of a healthy, growing company. 📈

```
select year(order_date) as order_year , count(product_code) as
number_of_transactions
from transactions
group by order_year
order by order_year ;
```

Optimizing Regional Sales: Top Product by Market (Level - “Hard”) (Query no. 3)

Question: For each market, find the product with the highest sales amount.

Business Insight: This powerful query helps tailor business strategy to regional tastes. By identifying the best-selling product in each market, a company can optimize local inventory, create targeted marketing campaigns, and ensure that promotional efforts are focused on the products that resonate most with that specific audience. It answers the crucial question: "What is our hero product in each city?" 🗺️

SQL Solution (Using a Common Table Expression - CTE):

To solve this multi-step problem elegantly, this solution uses a **Common Table Expression (CTE)**. A CTE, defined using the `WITH` clause, creates a temporary, named result set (`ProductSalesByMarket`) that can be referenced in the main query. This makes the code more readable and easier to debug than a complex subquery.

1. **The CTE (`MarketProductSales`) :** First, the CTE calculates the total sales for each product within each market.
2. **2nd CTE (`RankedSales`) :** Then, it uses the `RANK()` window function to assign a rank to every product based on its sales *within that market*.
3. **The Final Query :** The final `SELECT` statement then queries this temporary `MarketProductSales` table and simply filters for the rows where the rank is 1, giving us the top-selling product for each market.

```sql
WITH MarketProductSales AS (
     select t.market_code , t.product_code , sum(t.sales_amount) as total_sales
     from transactions as t
     group by  t.market_code , t.product_code
) ,
RankedSales AS (
     select mps.market_code , mps.product_code , mps.total_sales ,
     rank() over(partition by mps.market_code order by mps.total_sales desc) as
sales_rank
    from MarketProductSales as mps )

select rs.market_code , m.markets_name , rs.product_code , rs.total_sales ,
m.zone
from RankedSales as rs
left join markets as m
on rs.market_code = m.markets_code
where sales_rank = 1
;
```

Maximizing Profitability: Top 3 Products (Level - “Hard”) (Query no. 5)

Question: Which are the top 3 most profitable products?

Business Insight: While revenue is important, **profit** is the ultimate measure of success. This query pinpoints the top 3 products that contribute the most to the bottom line. With this information, the business can double down on what works—allocating more ad spend, ensuring healthy stock levels, and training the sales team to prioritize these high-margin items. 💰

SQL Solution (Using a Subquery):

This solution uses a **subquery** (the inner `SELECT` statement) to first calculate the total profit for every product. The outer query then takes the results from this subquery, orders them, and selects the top 3. This is a clear and effective way to break down a complex problem into logical steps.

```sql
select *
from
(
select product_code , round(sum(profit_margin),2) as
total_profit_margin ,
rank () over (order by round(sum(profit_margin),2) desc ) as
profit_rank
from transactions
group by product_code
order by total_profit_margin desc
) as mm
where mm.profit_rank < 4
;
```

Tools and Technologies

- **Database:** MySQL
- **Query Language:** SQL

Conclusion & Key Takeaways

This project uses SQL to analyze sales data, providing actionable insights. It demonstrates comprehensive SQL understanding, from basic retrieval to advanced analytics. The findings can inform business decisions on CRM, product, and market strategy.

- **Identified Key Profit Drivers:** The analysis successfully pinpointed the top 3 most profitable products, providing clear direction for where to focus sales and marketing efforts to maximize profitability.
- **Revealed Regional Customer Preferences:** By identifying the top-selling product in each market, the business can now implement data-driven, localized strategies for inventory and marketing.
- **Established Performance Baselines:** The project established clear KPIs, such as annual transaction growth and monthly revenue, which can be used to track business performance over time.

Future Work & Next Steps

- Customer **Segmentation:** "The next step could be to perform a more detailed customer segmentation analysis (e.g., high-value vs. one-time buyers) to create even more targeted marketing campaigns."
- **Time-Series Forecasting:** "Using the monthly revenue data, a future project could involve building a time-series model to forecast sales for the upcoming quarters."
- **Basket Analysis:** "Analyze which products are frequently purchased together to identify cross-selling opportunities."

Appendix: Full Project Script

The complete SQL script containing all queries—from basic to advanced—is available for review. For the best viewing experience and to see the project in its entirety, please visit the project's GitHub repository.

[Link to your GitHub Repository Here]