Slice of Analysis: A Comprehensive SQL Journey Through Pizza Sales Data

Overview

This project report details my progression through three stages of SQL queries—basic, intermediate, and advanced—using a robust pizza sales dataset. The focus of the report is to discuss my findings, the skills I applied, and the strategies I adopted for handling large datasets and complex SQL functionalities. The aim is to demonstrate my proficiency in SQL for potential employers or educational purposes.

Project Setup and Data Import

The project initiated with establishing a MySQL database environment and importing a substantial pizza sales dataset containing over 21,300 records. The data was organized into four key tables: Orders, Pizza_Types, Pizzas, and Order_Details. My initial task involved efficiently importing this dataset using MySQL Workbench, which required configuring data types, defining primary keys, and ensuring the integrity of each field to prevent issues during the import process. This foundational step was crucial for facilitating the subsequent data manipulation and queries.

SQL Query Development

Basic Queries: At the basic level, I focused on crafting SQL queries to create and populate tables. This stage was essential for understanding table relationships and practicing simple SQL commands to retrieve data. I executed queries to display all records from each table, verifying the successful import and integrity of the data.

Intermediate Queries: Transitioning to intermediate queries, I began to manipulate the data more extensively. This involved employing JOINs to connect different tables and extract meaningful relationships, such as matching pizzas with their orders and calculating total sales per pizza type. This phase was pivotal in enhancing my ability to derive insights from multiple data sources and honed my skills in using aggregate functions like SUM and COUNT within GROUP BY clauses.

Advanced Queries: The most challenging and rewarding part of the project involved utilizing Common Table Expressions (CTEs) and Window Functions. I performed complex data analysis tasks such as calculating the percentage contribution of each pizza category to total revenue and analyzing sales trends over time. Additionally, I implemented ranking functions to determine the top-selling pizzas, involving nested subqueries and strategic use of PARTITION BY and ORDER BY clauses within the RANK() function.

Key Insights and Findings

Several key insights were uncovered throughout the project:

- **Sales Distribution:** Advanced queries revealed the sales distribution across different times of the day, identifying peak sales periods.
- **Revenue Analysis:** I calculated the cumulative revenue over time, providing a clear picture of growth and seasonal fluctuations.

- **Product Performance:** The ranking queries highlighted the best-performing pizza types, suggesting potential business strategies for focusing on popular products or promoting less popular ones.

Challenges and Solutions

Handling the large dataset presented significant challenges, particularly in maintaining performance and managing memory. I implemented batching techniques and used indexes to optimize query execution times. Learning to define efficient primary keys and using the right JOIN types (INNER JOIN, LEFT JOIN) were critical in managing data relationships without compromising query performance.

Conclusion

This project not only enhanced my SQL skills but also strengthened my problem-solving abilities. I learned to methodically approach large datasets, applying best practices in data management and query optimization to extract actionable insights. This experience has significantly prepared me for real-world data analysis tasks and has been a valuable addition to my professional portfolio.