

Project Document: SQL-Based Analysis of Retail Sales Performance

1. Introduction

In the retail industry, data-driven decision-making is essential for optimizing sales strategies, improving operational efficiency, and enhancing customer experience. This project focuses on leveraging SQL to analyze a retail sales transactions dataset. By querying and analyzing structured sales data, we aim to extract valuable insights that will aid in business strategy formulation and operational improvements.

2. Project Objective

The primary goal of this project is to utilize SQL to analyze sales transactions data and extract meaningful insights in the following key areas:

- **Market Analysis:** Understanding customer purchasing behavior, product preferences, and sales trends.
 - **Operational Efficiency:** Identifying patterns in sales performance to optimize inventory and supply chain management.
 - **Sales Performance:** Evaluating high- and low-performing products and store branches.
 - **Customer Satisfaction:** Analyzing customer feedback and ratings to improve service quality.
-

3. Dataset Overview

The dataset contains transactional sales data with the following key attributes:

- **Invoice ID:** Unique identifier for each transaction.
- **Branch:** Store branch where the transaction took place (A, C).
- **City:** Location of the store (Yangon, Naypyitaw).
- **Customer Type:** Member or Normal customer.
- **Gender:** Male or Female.
- **Product Line:** Category of the product purchased (e.g., Health and Beauty, Electronic Accessories).
- **Unit Price:** Price per unit of the product.
- **Quantity:** Number of units purchased.
- **Tax 5%:** Tax applied to the total price.
- **Total:** Total amount paid, including tax.

- **Date & Time:** Timestamp of the transaction.
 - **Payment Method:** Mode of payment (Ewallet, Cash, Credit Card).
 - **COGS:** Cost of Goods Sold.
 - **Gross Margin Percentage:** Percentage of revenue exceeding COGS.
 - **Gross Income:** Earnings from the transaction before deductions.
 - **Rating:** Customer satisfaction rating.
-

4. SQL Analytical Approach

The SQL-based analysis will be structured into the following stages:

4.1 Data Preprocessing

- Checking for **duplicate entries** and ensuring data consistency.
- Handling **missing values**, if any.
- Converting **date and time** fields into appropriate SQL formats.

4.2 Exploratory Data Analysis (EDA) Using SQL Queries

1. **Customer Segmentation:**
 - Query to count transactions by customer type.
 - Analyze average spending per customer type.
2. **Sales Trend Analysis:**
 - SQL query to determine sales performance over time.
 - Identify peak sales days and time slots.
3. **Product Line Performance:**
 - Query to rank product lines by total revenue.
 - Compute average quantity sold per product category.
4. **Payment Method Insights:**
 - Query to determine the most preferred payment methods.
 - Evaluate the correlation between payment methods and customer satisfaction.

4.3 Performance Analysis Using SQL

1. **Branch and City-Wise Sales Performance:**
 - Query to compare sales revenue across branches and cities.
2. **Customer Type Revenue Contribution:**
 - Query to determine whether members or normal customers contribute more to revenue.
3. **Product Line Profitability:**
 - SQL query to compute the highest profit margins by product category.
4. **Gross Income & Margins Analysis:**
 - Calculate total gross income and gross margin percentages using SQL.

4.4 Customer Satisfaction Analysis

- Analyze customer ratings by product line and store branch.
 - Identify factors influencing higher customer satisfaction scores.
-

5. Tools and Techniques

For this project, the following tools and techniques will be utilized:

- **SQL (MySQL/PostgreSQL/SQL Server)** for querying, data manipulation, and aggregation.
- **SQL Functions:** GROUP BY, ORDER BY, JOIN, WHERE, HAVING, CASE statements.