SQL Comprehensive Report on Retail Sales Data Analysis

INTRODUCTION: This project involves using SQL to explore, clean, and analyze retail sales data, also setting up a retail sales database, performing exploratory data analysis (EDA), and answering specific business questions through SQL queries.

OBJECTIVE: The main objectives of this project is to;

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.

PROCESS: These are the processes and steps taken;

1. Database Creation and Data Import I began the project by creating a new database in SSMS name Retail_Sales Database. After setting up the database, I imported the given dataset into it for cleaning, preparation, and analysis.

2. Initial Data Exploration The first step was to study the structure of the dataset. During exploration, I noticed a spelling error in one of the field names 'quantiy' instead of 'quantity' This was corrected to maintain clarity and consistency. I also checked for null values and duplicate records to ensure the integrity of the data. Additionally, I ran a query to remove any trailing spaces in text fields which could interfere with grouping or filtering operations.

3. Data Type Corrections Upon reviewing the dataset, I discovered that most fields had been stored as VARCHAR. This was not ideal since numerical and date operations would be inefficient or prone to error. To correct this:

• Converted total_sale, price_per_unit, and cogs from VARCHAR to DECIMAL for accurate arithmetic operations.

• Converted age, transaction_id, customer_id, and quantity to INT because they are numerical identifiers or whole numbers.

• Converted sales_date and sales_time to proper DATE and TIME data types respectively, allowing for effective date-time operations.

4. Data Cleaning Summary

• Corrected field name spelling errors.

• Removed unnecessary spaces.

• Checked for null and duplicate records.

• Assigned appropriate data types to each column. These steps ensured that the dataset was reliable and ready for analysis.

5. Business Questions and SQL Queries After cleaning the data, I proceeded to answer the given business questions using SQL queries. Some of the tasks I performed include:

• Total and Average Sales Analysis: Used SUM() and AVG() functions to calculate revenue and average sales across different dimensions.

• Top Performing Products/Months: Applied RANK() and ORDER BY to identify which months had the highest sales per year.

• Customer and Transaction Insights: Grouped sales data by customer and transaction IDs to derive meaningful patterns.

• Profitability Measures: Calculated total_sale - cogs where required to estimate gross profit. For example, to identify the best sales month in each year, I applied a window function (RANK() OVER (...)) that ranked months by total sales and returned the top one per year.

6. Insights and Conclusion Through this process, I was able to transform raw, inconsistent sales data into a structured, reliable dataset suitable for analysis. The queries provided insights such as:

• Which months had the highest sales performance each year.

• Trends in average sales values.

• Customer and product performance metrics. By systematically cleaning the dataset, converting datatypes, and applying SQL queries, I was able to provide meaningful answers to the business questions. This process highlights the importance of data quality in deriving accurate insights and supporting informed decision

TOOL USED: SSMS

CONCLUSION: With the approach used above , we are now sure that the data is thoroughly cleaned and business questions asked is answered accordingly.