**SQL Analysis – Retail Data Insights**

**1. Introduction**

Retail businesses rely on data-driven insights to improve sales, understand customer behavior, and optimize inventory. This project focuses on analyzing retail sales data using SQL to extract meaningful insights. The analysis includes customer purchases, top-selling products, revenue trends, and other key metrics. The insights are further visualized in Power BI for better interpretation.

**2. Dataset Overview**

The dataset used for this project contains the following columns:

* **Transaction ID** (Integer, Primary Key)
* **Date** (Date)
* **Customer ID** (String)
* **Gender** (String)
* **Age** (Integer)
* **Product Category** (String)
* **Quantity** (Integer)
* **Price per Unit** (Integer)
* **Total Amount** (Integer)

The dataset was imported into MySQL, and data cleaning techniques were applied to ensure accuracy and completeness.

**3. Data Cleaning using SQL**

**3.1 Handling Missing Values**

To identify missing values, we used:

SELECT COLUMN\_NAME, COUNT(\*) - COUNT(COLUMN\_NAME) AS missing\_values

FROM retail\_sales\_data

GROUP BY COLUMN\_NAME;

Missing values were handled appropriately based on the data type and significance of the column.

**3.2 Checking for Duplicates**

Duplicate records were identified and removed using:

SELECT \*, COUNT(\*)

FROM retail\_sales\_data

GROUP BY Transaction\_ID, Date, Customer\_ID, Product\_Category, Quantity

HAVING COUNT(\*) > 1;

**3.3 Basic Statistics**

To analyze the key statistics of sales data:

SELECT

COUNT(\*) AS total\_rows,

MIN(Total\_Amount) AS min\_sales,

MAX(Total\_Amount) AS max\_sales,

AVG(Total\_Amount) AS avg\_sales,

SUM(Total\_Amount) AS total\_sales

FROM retail\_sales\_data;

**4. Data Analysis & Insights**

**4.1 Best-Selling Products**

To determine the top-selling products:

SELECT Product\_Category, SUM(Total\_Amount) AS total\_sales

FROM retail\_sales\_data

GROUP BY Product\_Category

ORDER BY total\_sales DESC

LIMIT 5;

**4.2 Identifying Peak Sales Months**

SELECT MONTH(Date) AS Month, SUM(Total\_Amount) AS total\_sales

FROM retail\_sales\_data

GROUP BY Month

ORDER BY total\_sales DESC;

**4.3 Most Valuable Customers**

SELECT Customer\_ID, SUM(Total\_Amount) AS total\_spent

FROM retail\_sales\_data

GROUP BY Customer\_ID

ORDER BY total\_spent DESC

LIMIT 5;

**4.4 Sales by Customer Demographics**

**Sales by Gender:**

SELECT Gender, SUM(Total\_Amount) AS total\_sales

FROM retail\_sales\_data

GROUP BY Gender;

**Sales by Age Group:**

SELECT

CASE

WHEN Age BETWEEN 10 AND 20 THEN '10-20'

WHEN Age BETWEEN 21 AND 30 THEN '21-30'

WHEN Age BETWEEN 31 AND 40 THEN '31-40'

WHEN Age BETWEEN 41 AND 50 THEN '41-50'

WHEN Age BETWEEN 51 AND 60 THEN '51-60'

ELSE '60+'

END AS Age\_Group,

COUNT(\*) AS Customers\_Count

FROM retail\_sales\_data

GROUP BY Age\_Group

ORDER BY Age\_Group;

**5. Visualization Using Power BI**

After extracting insights using SQL, we visualized the data using Power BI. The following visualizations were created:

* **Total Revenue Trend**: Line chart displaying revenue over time.
* **Top-Selling Product Categories**: Bar chart showing product sales.
* **Customer Demographics Analysis**: Pie charts for gender and age group distribution.
* **Sales by Month**: Bar chart representing peak sales months.

**6. Conclusion**

This project demonstrated how SQL can be used to clean, explore, and analyze retail data. By leveraging SQL queries, we identified key sales trends, customer behaviors, and revenue insights. These findings were further visualized in Power BI for better decision-making. The approach helps businesses optimize inventory, improve marketing strategies, and enhance customer experiences.

**Project Completed by:** Bala **Tools Used:** MySQL, Power BI