Project Report

Retal Analysis

Introduction :

In today's data-driven retail environment, understanding customer demographics and purchase behavior is essential for driving sales, improving marketing strategies, and enhancing product offerings. This project leverages transactional, product, and customer datasets to explore key sales patterns and customer insights. Using SQL for data preparation and Power BI for visualization, the objective is to uncover actionable insights from retail data.

Abstract:

This project involves building a full-stack data analysis and visualization pipeline using SQL and Power BI. The dataset consists of three tables: **Customers**, **Transactions**, and **Product Category Info**, each offering vital information about customer profiles, product classifications, and sales records. SQL was used to clean, merge, and extract insights, while Power BI was used to create interactive dashboards that present:

- Sales trends over time
- Customer age group analysis
- Product category performance
- Store-type wise sales insights
- Return vs. purchase patterns

Tools:

- **SQL (MySQL / PostgreSQL / SQL Server)**:
  For data exploration, joins, filtering, aggregations, and computed columns (e.g., age calculation).
- **Power BI Desktop**:
  For data modeling, DAX measures, and creating dashboards with advanced visualizations and slicers.
- **Power Query Editor** (within Power BI):
  For transforming and cleaning data (changing data types, removing nulls, etc.).

Steps Involved:

1. Data Understanding and Cleaning (in SQL & Power BI)

- Imported the datasets into Power BI.
- Inspected data types and cleaned invalid values.
- Converted DOB to date type, removed null or future dates.
- Filtered negative values to analyze returns separately.

2. Relationship Modeling

- Created relationships in Power BI Model View:
    - Customers[Customer_ID] → Transactions[cust_id]
    - Prod_cat_info[prod_cat_code] → Transactions[prod_cat_code]
    - Prod_cat_info[prod_subcat_code] → Transactions[prod_subcat_code]

3. SQL Queries for Pre-Analysis

- Identified top-selling product categories.
- Calculated total returns using negative quantity and amount.
- Derived customer age from DOB.
- Segmented customers into age groups using SQL CASE statements.

4. DAX Measures Created in Power BI

- TotalSales = SUM(Transactions[total_amt])

- TotalQuantity = SUM(Transactions[Qty])

- AvgTaxPerTransaction = AVERAGE(Transactions[Tax])

- ReturnAmount = CALCULATE(SUM(Transactions[total_amt]), Transactions[Qty] < 0)

- CustomerAge = DATEDIFF(Customers[DOB], TODAY(), YEAR)

- AgeGroup = SWITCH(TRUE(), …) *(e.g., "18–25", "26–40", etc.)*

◆ **5. Dashboard Design in Power BI**

Key visuals added:

- **Column Chart**: Total Sales by Age Group

- **Stacked Bar Chart**: Product Category vs. Quantity Sold

- **Donut Chart**: Store Type Contribution to Total Sales

- **Line Chart**: Monthly Sales Trend

- **Card KPIs**: Total Revenue, Total Returns, Avg. Rate

- **Slicers**: Gender, City, Store Type

Conclusion:

This project successfully demonstrates how to transform raw customer and transactional data into meaningful business insights using SQL and Power BI. Key findings include:

- Customers aged 26–40 drive the highest sales volume.

- Certain product categories contribute disproportionately to revenue.

- Physical stores outperform other types in average transaction value.

- Return rates are higher in specific subcategories, indicating quality or satisfaction issues.

By integrating SQL for data shaping and Power BI for visual storytelling, the project showcases practical data analytics capabilities that support business decision-making.