**Regional Sales Reporting of Fortune 500 Retail Chain: Analyzing Sales Data with SQL, Python, and Excel**

**Project Description:**

**This project involves using SQL to subset a dataset of sales records, Python for further data manipulation, and Excel to create various visualizations to analyze sales performance. The goal is to extract meaningful insights about sales trends, product performance, and regional sales distribution.**

**Dataset:**

**A sales dataset with the following structure:**

| **OrderID** | **OrderDate** | **Region** | **Product** | **Quantity** | **Price** | **TotalSales** |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | **2023-01-15** | **North** | **Product A** | **10** | **20** | **200** |
| **2** | **2023-02-20** | **South** | **Product B** | **5** | **50** | **250** |
| **3** | **2023-03-10** | **East** | **Product C** | **7** | **30** | **210** |
| **4** | **2023-04-05** | **West** | **Product A** | **8** | **20** | **160** |
| **5** | **2023-05-20** | **North** | **Product D** | **3** | **100** | **300** |
| **...** | **...** | **...** | **...** | **...** | **...** | **...** |

**Queries and Commands:**

**1. Subsetting Dataset using SQL**

**a. Extracting sales data for a specific region**

**SELECT \* FROM sales\_data**

**WHERE Region = 'North';**

**b. Extracting sales data for a specific product**

**SELECT \* FROM sales\_data**

**WHERE Product = 'Product A';**

**c. Extracting sales data within a specific date range**

**SELECT \* FROM sales\_data**

**WHERE OrderDate BETWEEN '2023-01-01' AND '2023-03-31';**

**d. Aggregating total sales by region**

**SELECT Region, SUM(TotalSales) as TotalSales**

**FROM sales\_data**

**GROUP BY Region;**

**e. Aggregating total sales by product**

**sql**

**SELECT Product, SUM(TotalSales) as TotalSales**

**FROM sales\_data**

**GROUP BY Product;**

**2. Data Manipulation using Python**

**After subsetting the data using SQL, we can use Python for additional data manipulation before exporting to Excel.**

**a. Required Libraries:**

**python**

**import pandas as pd**

**import sqlite3**

**b. Connecting to the Database and Executing Queries:**

**python**

**# Connect to SQLite database**

**conn = sqlite3.connect('sales.db')**

**# Query data**

**query = "SELECT \* FROM sales\_data WHERE Region = 'North';"**

**north\_sales\_data = pd.read\_sql\_query(query, conn)**

**query = "SELECT \* FROM sales\_data WHERE Product = 'Product A';"**

**product\_a\_data = pd.read\_sql\_query(query, conn)**

**query = "SELECT \* FROM sales\_data WHERE OrderDate BETWEEN '2023-01-01' AND '2023-03-31';"**

**date\_range\_data = pd.read\_sql\_query(query, conn)**

**query = "SELECT Region, SUM(TotalSales) as TotalSales FROM sales\_data GROUP BY Region;"**

**total\_sales\_by\_region = pd.read\_sql\_query(query, conn)**

**query = "SELECT Product, SUM(TotalSales) as TotalSales FROM sales\_data GROUP BY Product;"**

**total\_sales\_by\_product = pd.read\_sql\_query(query, conn)**

**# Close the connection**

**conn.close()**

**c. Exporting Data to Excel:**

**python**

**with pd.ExcelWriter('sales\_analysis.xlsx') as writer:**

**north\_sales\_data.to\_excel(writer, sheet\_name='North Sales', index=False)**

**product\_a\_data.to\_excel(writer, sheet\_name='Product A Sales', index=False)**

**date\_range\_data.to\_excel(writer, sheet\_name='Date Range Sales', index=False)**

**total\_sales\_by\_region.to\_excel(writer, sheet\_name='Sales by Region', index=False)**

**total\_sales\_by\_product.to\_excel(writer, sheet\_name='Sales by Product', index=False)**

**3. Creating Visuals using Excel**

**Some visualization created in excel -**

**a. Total Sales by Region**

**b. Sales Trend Over Time**

**c. Total Sales by Product**

**d. Sales Quantity by Product and Region**

**Example Dataset in Excel:**

**OrderID | OrderDate | Region | Product | Quantity | Price | TotalSales**

**--------|-----------|--------|---------|----------|-------|-----------**

**1 | 2023-01-15| North | A | 10 | 20 | 200**

**2 | 2023-02-20| South | B | 5 | 50 | 250**

**3 | 2023-03-10| East | C | 7 | 30 | 210**

**4 | 2023-04-05| West | A | 8 | 20 | 160**

**5 | 2023-05-20| North | D | 3 | 100 | 300**

**...**

**Conclusion:**

**Based on the analysis of the sales data, several key insights have been identified:**

**Regional Performance:**

**The North region has the highest total sales, contributing significantly to the company's overall revenue.**

**The South region shows potential for growth, but current sales are lower compared to other regions.**

**Product Performance:**

**Product A is the top-performing product across all regions, indicating strong market demand.**

**Product D, while having higher individual sales prices, has lower sales volumes, suggesting it may be a niche product.**

**Sales Trends Over Time:**

**Sales peak in the first quarter of the year, particularly in January and February, likely due to post-holiday season buying behavior.**

**There is a noticeable dip in sales during the summer months, which could be attributed to seasonal factors affecting customer purchasing habits.**

**Sales Strategy Insights:**

**Focusing marketing efforts in the South region could help boost its sales, leveraging the strong performance seen in the North.**

**Promoting Product A in regions where it is already performing well could further increase its sales volume.**

**Exploring strategies to increase sales of Product D, such as targeted promotions or bundling with other products, could help improve its market share.**

**These insights can guide strategic decisions to enhance sales performance, optimize product offerings, and tailor marketing efforts to regional preferences and seasonal trends.**