**Retail Order Data Analysis**

This project focuses on the analysis and visualization of retail order data, where we work with a dataset that contains various details about retail transactions. The goal is to provide insights into sales performance, profit margins, discounts, order trends, and other key metrics related to retail business performance. Below is a detailed description of each component of the project:

**1. Dataset Overview:**

The project uses a dataset called orders.csv, which contains detailed information about retail orders. The data includes various attributes such as:

* **Order details**: Order ID, Order Date, Ship Mode, Postal Code, City, State, Country, Segment, and Region.
* **Product details**: Product ID, Category, Sub-category, List Price, Cost Price, Quantity, and Discount Percent.
* **Revenue and Profit details**: Discount, Sales price and Profit were calculated.

**2. Preprocessing the Data:**

Before analysis, the dataset is preprocessed:

* **Missing Values Handling**: Missing values are dealt with by filling them with zeros or removing rows with critical missing values like ‘List Price’.
* **Renaming Columns**: The column names are standardized and renamed to lowercase for consistency.
* **Handling Text Data**: Text columns (e.g., City, State) are stripped of any leading or trailing spaces to ensure clean data.
* **Creating Additional Columns**: The project calculates some additional columns such as:
  + discount: Discount amount calculated using the list\_price and discount\_percent.
  + sale\_price: The final sale price after applying the discount.
  + profit: Profit earned on each order, calculated as (sale\_price - cost\_price) \* quantity.

**3. Data Transformation:**

The data is split into two tables for easier analysis:

* **Table 1 (Order Information)**: Contains order-related information (e.g., Order ID, Date, Ship Mode, Segment, etc.).
* **Table 2 (Product Information)**: Contains product-related information (e.g., Category, Sub-category, Product ID, Cost Price, List Price, etc.).
* These tables are stored in a PostgreSQL database called retailorder for querying and further analysis.

**4. Database Connection & SQL Queries:**

The project uses PostgreSQL as the backend database to store and query data. Through Python and psycopg2, the project establishes a connection to the database and executes SQL queries to retrieve the required insights.

**5. Data Visualization with Streamlit and Plotly:**

A **Streamlit app** is created to display the results of various SQL queries in an interactive manner. Users can select a specific query from a dropdown list, and the corresponding results are displayed in tables and charts.

* **Plotly** is used for data visualization, which allows interactive bar charts, pie charts, and line charts to represent various metrics, such as:
  + **Top Revenue-Generating Products**
  + **Cities with the Highest Profit Margins**
  + **Total Discount Given for Each Category**
  + **Average Sale Price per Product Category**
  + **Year-over-Year Revenue Growth**
  + **Monthly Revenue Trends**

These visualizations provide clear, actionable insights that help retail businesses understand their performance.

**6. Key SQL Queries Implemented:**

Several key business questions are addressed using SQL queries, such as:

* **Top 10 Highest Revenue-Generating Products**: A query identifies the top 10 products based on total revenue.
* **Top 5 Cities with the Highest Profit Margins**: A query finds the cities with the highest profit margins.
* **Total Discount Given for Each Category**: A query calculates the total discount given for each product category.
* **Average Sale Price per Product Category**: A query calculates the average sale price for products within each category.
* **Profit Margin of Subcategories**: A query calculates the profit margins of subcategories within different product categories (Furniture, Technology, etc.).
* **Year-over-Year Growth in Total Revenue**: A query calculates the percentage growth in total revenue compared to the previous year.

**7. Key Metrics Analyzed:**

The project aims to provide several metrics related to retail performance:

* **Revenue**: Total revenue generated per product, region, category, year, month, etc.
* **Profit Margins**: Profit margins by city, region, or subcategory.
* **Discounts**: Total discount given and its impact on sales.
* **Quantity of Orders**: Analysis of the top segments with the highest order quantities.
* **Order Trends**: Revenue trends over time (monthly, quarterly, yearly).

**8. Streamlit Dashboard:**

The interactive **Streamlit app** serves as the user interface, where users can:

* Select the query of interest.
* View tabular results for the selected query.
* Interactively explore the visualizations (e.g., bar charts, pie charts, line charts).
* Download the visualized data if required.

**9. Deployment:**

Once the app is developed, it can be deployed to a web server for access by business analysts or other users who need to explore the retail data and derive insights.

**10. Summary of Insights Provided:**

This project can help retail businesses by providing insights into:

* The most profitable products and categories.
* The impact of discounts on revenue.
* The cities or regions with the highest profit margins.
* Monthly and yearly revenue trends.
* Profit margins by product subcategory.

**Technologies Used:**

* **Python**: For data processing and analysis.
* **Pandas**: For data manipulation and handling.
* **PostgreSQL**: For data storage and querying.
* **Streamlit**: For building an interactive web application.
* **Plotly**: For data visualization (bar charts, pie charts, line charts).

This project is a robust solution for retail data analysis and can be extended or customized to fit specific business needs. It combines SQL querying, data visualization, and an interactive web interface to provide actionable insights into retail business performance.