Data Exploration and Visualization Exercise

Overview

In this exercise, we'll explore a dataset replicated from a PostgreSQL database into Google BigQuery. We'll be creating wide tables to facilitate easier reporting and visualization using Google Data Studio. The aim is to derive business insights from the data with a focus on sales analysis, customer behavior, and supply chain management.

Objectives

1. **Sales Analysis:** Understand sales trends, best-selling products, and sales performance.
2. **Customer Analysis:** Identify high-value customers and analyze purchasing patterns.
3. **Supply Chain Analysis:** Evaluate supplier performance and product availability.

Task 1: Creating Wide Tables

1.1 Sales Wide Table

The Sales Wide Table will facilitate analysis and reporting on sales data by combining relevant fields from the Products, Order Details, Orders, Customers, and Employees tables.

SQL Query:

```
CREATE OR REPLACE VIEW
  `your_project_id.your_dataset_id.SalesWide`
AS
  (WITH
    OrderWithCustomer AS (
    SELECT
      o.*,
      c.CompanyName, c.ContactName, c.Country, c.Region
    FROM
      `your_project_id.your_dataset_id.Orders` AS o
    JOIN
      `your_project_id.your_dataset_id.Customers` AS c
    ON
      o.CustomerID = c.CustomerID ),
    OrderDetailWithProduct AS (
```

```sql
    SELECT
      od.*,
      p.ProductName, p.CategoryID
    FROM
      `your_project_id.your_dataset_id.OrderDetails` AS od
    JOIN
      `your_project_id.your_dataset_id.Products` AS p
    ON
      od.ProductID = p.ProductID )
  SELECT
    p.ProductID, p.ProductName, p.CategoryID,
    od.OrderID, od.UnitPrice, od.Quantity, od.Discount,
    o.OrderDate, o.RequiredDate, o.ShippedDate, o.ShipVia,
    o.CompanyName, o.ContactName, o.Country, o.Region,
    e.EmployeeID, e.LastName, e.FirstName, e.Title
  FROM
    OrderDetailWithProduct AS od
  JOIN
    OrderWithCustomer AS o
  ON
    od.OrderID = o.OrderID
  JOIN
    `your_project_id.your_dataset_id.Employees` AS e
  ON
    o.EmployeeID = e.EmployeeID);
```

1.2 Supplier Wide Table

The Supplier Wide Table will assist in analyzing and reporting on supply chain data by combining fields from the Suppliers, Products, and Categories tables.

SQL Query:

```sql
CREATE OR REPLACE VIEW
  `your_project_id.your_dataset_id.SupplierWide`
AS
  (WITH
    ProductWithCategory AS (
    SELECT
      p.*,
      c.CategoryName, c.Description
    FROM
```

```
    `your_project_id.your_dataset_id.Products` AS p
  JOIN
    `your_project_id.your_dataset_id.Categories` AS c
  ON
    p.CategoryID = c.CategoryID )
SELECT
  s.SupplierID, s.CompanyName, s.ContactName, s.Country,
  p.ProductID, p.ProductName, p.UnitPrice,
  p.CategoryID, p.CategoryName, p.Description
FROM
  `your_project_id.your_dataset_id.Suppliers` AS s
JOIN
  ProductWithCategory AS p
ON
  s.SupplierID = p.SupplierID);
```

Note: Ensure to replace `your_project_id` and `your_dataset_id` with your actual BigQuery project and dataset ID.

Task 2: Exploration and Visualization

Using Google Data Studio, explore the wide tables created and derive insights related to:

- **Product Sales:** Identify the top-selling products.
- **Customer Behavior:** Analyze purchasing patterns of high-value customers.
- **Employee Performance:** Evaluate the sales performance of employees.
- **Supplier Analysis:** Assess the performance and contribution of suppliers.

Task 3: Report Generation

Create a report in Google Data Studio and share it with your peers. Ensure your report contains:

- **Visualizations:** Various charts, graphs, and tables to represent the data.
- **Insights:** Interpretations or observations derived from the visualizations.

Submission

Submit the SQL queries used, the Google Data Studio report link, and a brief document containing insights and observations derived from the data.

Evaluation Criteria

- **SQL Queries:** Accuracy and efficiency of SQL code.
- **Visualization:** Clarity and relevance of visualizations in the report.
- **Insights:** Depth and relevance of derived insights and observations.