Retail Orders Data Analysis

Downloading Data

Data Analysis using SQL

Data Cleaning and Processing

Loading Data

-By Deepanshu

Table of Contents

1. Introduction

2. Objectives and Scope

3. Workflow

4. Data Collection and Loading

5. Data Cleaning

6. Data Analysis

7. Insights and Results

Introduction

This project focuses on a comprehensive analysis of sales data to uncover key performance insights. Utilizing Python for data cleaning ensured the accuracy and reliability of the results. The analysis covers revenue generation, sales performance, and profit growth, providing valuable information to guide strategic decisions.

Objectives

- Identifying top revenue-generating products.
- Assessing regional sales performance.
- Comparing year-over-year sales growth.
- Highlighting profit growth trends.

Scope

- Data Cleaning: Used Python for thorough data cleaning to ensure accuracy.
- Performance Analysis: Assessed revenue generation, regional sales, and profit growth trends.

Workflow

Data Collection and Loading

Data Cleaning and Transformation

Data Storage and Management

Data Analysis and Insights

Data Collection and Loading

Data Sources

- Orders dataset containing information about order id, order date, product id, cost price, listed price, discount etc.

Data Loading Process

- Data was initially loaded into Python for preprocessing and then imported into the raw data layer of SQL.

DATA CLEANING

Null Handling

```
import numpy as np
df["Ship Mode"]=df["Ship
Mode"].replace({'unknown':np.nan,"Not
Available":np.nan})
```

To ensure data consistency and accuracy, we implemented null value handling by replacing occurrences of "unknown" and "Not Available" with NaN (Not a Number) in the "Ship Mode" column. This approach streamlines data preprocessing, enabling clearer insights during our analysis phase.

Standardizing Column Names

```
df.columns=df.columns.str.lower().str.replace(" ","_")
```

To ensure uniformity and ease of access, we standardized column names in our dataset. The transformation involved converting column names to lowercase and replacing spaces with underscores for consistency.

Product Id → **product_id**

Addition of New Columns

```
df['discount_per_unit']=df.list_price*df.discount_percent*0.01
df["sale_price"]=df.list_price-df.discount_per_unit
df["profit"]=df.sale_price - df.cost_price
```

To further enhance our dataset for comprehensive analysis, we introduced the following new columns:

- Discount per Unit: Calculated as a percentage of the list price.
- Sale Price: Derived by subtracting the discount per unit from the list price.
- Profit: Determined by subtracting the cost price from the sale price.

Conversion of Order Date to Datetime

```
df.order_date=pd.to_datetime(df.order_date)
```

To enhance analysis and ensure consistency in date handling, we converted the "order_date" column to datetime format . This transformation enables accurate date-based calculations and insights within our dataset.

DATA ANALYSIS

1. Find top 10 highest revenue generating products.

```sql
SELECT product_id,
         round(sum(sale_price*quantity),
        2) AS revenue
FROM orders
GROUP BY  product_id
ORDER BY  revenue DESC limit 10;
```

1.Top 10 highest revenue generating products.

Product ID	Revenue
TEC-CO-10004722	245,056
OFF-BI-10000545	163,777.7
TEC-MA-10002412	130,406.4
FUR-CH-10002024	120,090.7
TEC-PH-10001459	113,041.9
TEC-CO-10001449	107,388
OFF-BI-10003527	97,082.9
TEC-MA-10000822	89,622.3
FUR-BO-10002213	84,014.8
TEC-MA-10001047	81,549

2.Find top 5 highest selling products in each region.

```sql
with cte AS
    (SELECT region ,
          product_id ,
          sum(quantity) AS quantity,
          dense_rank( ) over(partition by region
    ORDER BY  sum(quantity) desc) AS r
    FROM orders
    GROUP BY  region,product_id)
SELECT region ,
         product_id,
         quantity,
         r
FROM cte
WHERE r<=5;
```

2.Top 5 highest selling products in each region.

Region	Product ID	Quantity	Ranking
Central	OFF-BI-10000301	34	1
Central	OFF-BI-10000756	33	2
Central	OFF-BI-10000546	29	3
Central	OFF-BI-10001249	29	3
Central	FUR-CH-10002304	27	4
Central	OFF-AP-10001947	27	4
Central	FUR-CH-10002880	26	5
East	OFF-PA-10001970	33	1
East	OFF-BI-10003656	32	2
East	OFF-FA-10000621	31	3
East	FUR-FU-10004848	31	3
East	OFF-FA-10002780	29	4
East	OFF-ST-10002615	29	4
East	OFF-BI-10001524	28	5
South	OFF-ST-10003716	26	1
South	FUR-CH-10000513	24	2
South	OFF-BI-10004728	24	2
South	OFF-BI-10000014	23	3
South	FUR-FU-10001731	21	4
South	OFF-BI-10000069	21	4
South	OFF-BI-10000977	20	5
South	OFF-BI-10001191	20	5
South	TEC-AC-10000023	20	5
South	TEC-PH-10001459	20	5
South	FUR-TA-10000198	20	5
West	TEC-AC-10003832	45	1
West	OFF-BI-10000174	32	2
West	OFF-BI-10001036	31	3
West	OFF-BI-10001670	29	4
West	OFF-ST-10002486	29	4
West	FUR-FU-10001979	28	5
West	TEC-AC-10002006	28	5

3.Find month over month growth comparison for 2022 and 2023 sales.

```sql
SELECT month(order_date) AS month,
          round(sum(case
    WHEN year(order_date)=2022 THEN
    sale_price*quantity
    ELSE 0 end),2) AS '2022', round(sum(case
    WHEN year(order_date)=2023 THEN
    sale_price
    ELSE 0 end),2) AS '2023'
FROM orders
GROUP BY  month(order_date)
ORDER BY  month(order_date);
```

3.Find month over month growth comparison for 2022 and 2023 sales.

Month	2022	2023
1	437,431.3	88,632.6
2	444,011.1	128,124.2
3	394,105.2	82,512.3
4	476,400.9	111,568.6
5	413,625.5	86,447.9
6	465,300.3	68,976.5
7	375,278.4	90,563.8
8	534,562.4	87,733.6
9	433,887	76,658.6
10	601,707.8	121,061.5
11	451,809.6	75,432.8
12	447,421.8	102,556.1

4.For each category which month had highest sales.

```sql
with cte AS
    (SELECT category,
        CONCAT(CAST(EXTRACT(MONTH
    FROM order_date) AS CHAR), '-', CAST(EXTRACT(YEAR
    FROM order_date) AS CHAR)) AS month ,
round(sum(sale_price*quantity),2) AS sales, dense_rank()
over(partition by category
    ORDER BY  sum(sale_price*quantity) desc) AS r
    FROM orders
    GROUP BY  category, CONCAT(CAST(EXTRACT(MONTH
    FROM order_date) AS CHAR), '-', CAST(EXTRACT(YEAR
    FROM order_date) AS CHAR)) )
SELECT category,
        month,
        sales
FROM cte
WHERE r=1;
```

4.For each category which month had highest sales.

Category	Month	Sales
Furniture	8-2023	230,523.5
Office Supplies	2-2023	287,244.6
Technology	10-2023	295,586.5

5.Which sub category had highest growth by profit in 2023 compare to 2022.

```sql
SELECT sub_category,
          sum(case
    WHEN year(order_date)=2022 THEN
    profit
    ELSE 0 end) AS "2022", sum(case
    WHEN year(order_date)=2023 THEN
    profit*quantity
    ELSE 0 end) AS "2023", (sum(case
    WHEN year(order_date)=2023 THEN
    profit*quantity
    ELSE 0 end) - sum(case
    WHEN year(order_date)=2022 THEN
    profit*quantity
    ELSE 0 end))*100.0/sum(case
    WHEN year(order_date)=2022 THEN
    profit*quantity
    ELSE 0 end) AS growth_pct
FROM orders
GROUP BY  sub_category
ORDER BY  growth_pct DESC ;
```

5.Sub category had highest growth by profit in 2023 compare to 2022.

Sub-Category	2022	2023	Growth (%)
Supplies	1,500.7	9,241.5	79.06
Machines	7,243.2	56,939.8	64.54
Binders	8,685.5	57,990.3	42.09
Envelopes	607.2	3,502.4	34.33
Storage	8,907.4	53,177.5	23.88
Phones	13,024.7	76,438.2	21.21
Labels	349.6	2,181.6	8.69
Accessories	7,387.2	40,613.7	4.68
Bookcases	5,459.5	25,776.1	1.10
Paper	3,058.9	16,211.9	0.38
Chairs	14,725.3	75,233.2	-4.37
Art	924.1	5,079.2	-4.76
Furnishings	4,236.2	16,428.4	-27.49
Tables	10,315.9	39,899.5	-31.94
Copiers	8,780.3	26,561.8	-37.80
Appliances	6,374.4	17,493.9	-52.83
Fasteners	40.1	87.9	-76.29

Insights and Results

Key Findings

Technology Leads Growth:

- Categories like Machines and Phones saw significant growth in 2023, with increases of 64.54% and 21.21% respectively.

Declines in Traditional Categories:

- Furniture categories such as Chairs (-4.37%) and Tables (-31.94%) faced challenges in sales growth.

Rising Demand in Office Essentials:

- Office Supplies, including Supplies and Binders, showed strong growth, indicating increased demand driven by evolving work environments.

Business Impact

Focus on High-Growth Technology:

- Prioritize products like Machines and Phones to capitalize on substantial growth opportunities.

Strategic Pivot in Furniture Categories:

- Address declines in traditional segments like Chairs and Tables through targeted adjustments and resource reallocation.

Expand in Office Supplies:

- Seize opportunities in growing categories such as Office Supplies (Supplies, Binders) to enhance market presence and meet evolving consumer demands.

Adapt to Remote Work Trends:

- Tailor product offerings to support remote work environments, ensuring alignment with changing consumer behaviors and needs.

Thank You!

Your engagement and attention are greatly appreciated.

Have a great day!