Regional Sales Data Analysis Project

A comprehensive analysis using Python, SQL, and Power BI
Following the EDA approach

Table of contents

01

Introduction

Objective:

The main goal of this project was to analyze store sales data to gain insights into sales performance, customer behavior, and product profitability. By leveraging tools like Python for data cleaning, SQL for querying the database, and Power BI for creating visualizations, the project aimed to provide actionable recommendations to improve business outcomes.

رواد مصر الرقمية

Scope:

The data analysis focused on several key areas:

- **Sales Trends:** Understanding overall sales performance across different channels, time periods, and locations.
- **Customer Insights:** Identifying key customer demographics and purchasing patterns to improve targeted marketing.
- **Product Profitability:** Assessing which products generate the highest profit and identifying underperforming products.
- **Regional Performance:** Analyzing sales across various regions and stores to optimize distribution and warehouse strategies.

Tools Used:

1. **Python:** Used for cleaning and preparing the dataset, handling missing values, standardizing data formats, and performing basic exploratory data analysis.
2. **SQL:** Employed to answer key business questions related to sales, customers, and products, including filtering and aggregating data.
3. **Power BI:** Used for building interactive dashboards to visualize sales trends, customer demographics, and other insights.

Key Insights:

1. **Sales Performance:** Identified the best-performing sales channels and time periods, showing seasonal spikes.
2. **Customer Analysis:** Discovered that higher-income households tend to spend more, especially in specific product categories.
3. **Product Profitability:** Determined the most profitable products and those that required improvement or discounting.
4. **Regional Analysis:** Revealed regional disparities in sales, with certain warehouses and states performing significantly better than others.

Deliverables:

1. Cleaned and structured data tables for efficient analysis.

2. SQL queries to answer specific business-related questions.

3. Power BI dashboards with visual insights for sales,

   customer, and product data.

02

Data Overview

Data Discription

The dataset used in this analysis consists of multiple tables, each representing different aspects of the store's sales operations.

The data is structured to provide insights into sales transactions, customer information, store locations, and products, with key metrics such as order quantities, discounts, prices, and profits.

```
# # Display sheet names and a preview of the first sheet
sheet_names = xls.sheet_names
sheet_names
```

```
['Sales Orders Sheet',
 'Customers Sheet',
 'Store Locations Sheet',
 'Products Sheet',
 'Regions Sheet',
 'Sales Team Sheet']
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7991 entries, 0 to 7990
Data columns (total 16 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   OrderNumber     7991 non-null    object
 1   Sales Channel   7991 non-null    object
 2   WarehouseCode   7991 non-null    object
 3   ProcuredDate    7991 non-null    datetime64[ns]
 4   OrderDate       7991 non-null    datetime64[ns]
 5   ShipDate        7991 non-null    datetime64[ns]
 6   DeliveryDate    7991 non-null    datetime64[ns]
 7   CurrencyCode    7991 non-null    object
 8   _SalesTeamID    7991 non-null    int64
 9   _CustomerID     7991 non-null    int64
 10  _StoreID        7991 non-null    int64
 11  _ProductID      7991 non-null    int64
 12  Order Quantity  7991 non-null    int64
 13  Discount Applied 7991 non-null   float64
 14  Unit Price      7991 non-null    float64
 15  Unit Cost       7991 non-null    float64
dtypes: datetime64[ns](4), float64(3), int64(5), object(4)
memory usage: 999.0+ KB
```

رواد مصر الرقمية

03

Data Cleaning (Python)

Confirming no missing values & no duplicates
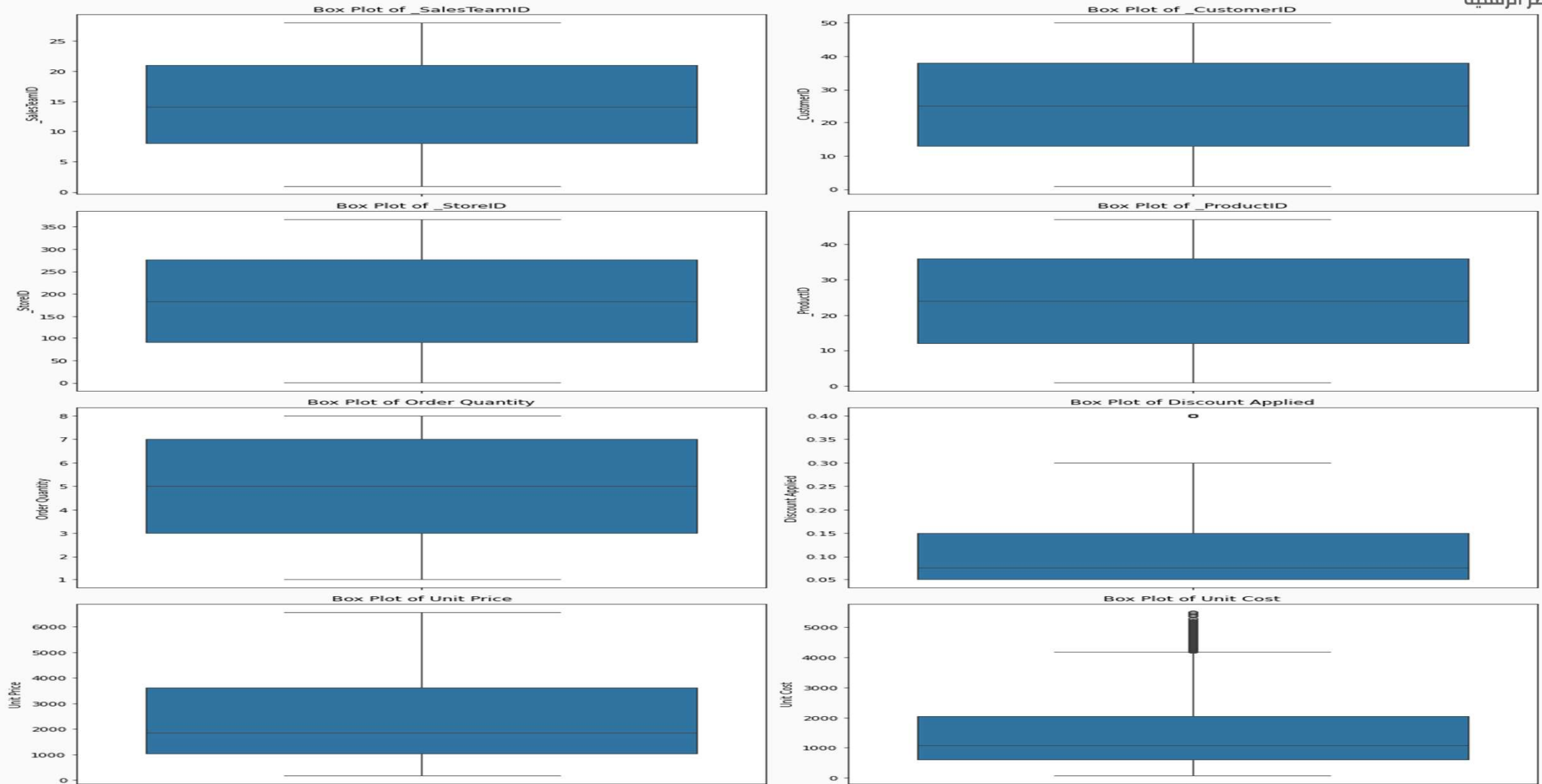
```
df.isna().sum()
```

```
OrderNumber        0
Sales Channel      0
WarehouseCode      0
ProcuredDate       0
OrderDate          0
ShipDate           0
DeliveryDate       0
CurrencyCode       0
_SalesTeamID       0
_CustomerID        0
_StoreID           0
_ProductID         0
Order Quantity     0
Discount Applied   0
Unit Price         0
Unit Cost          0
dtype: int64
```
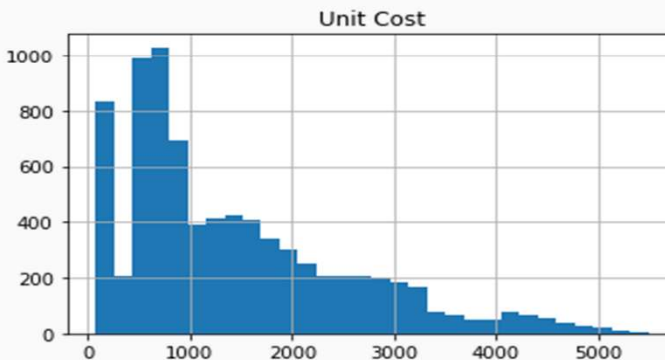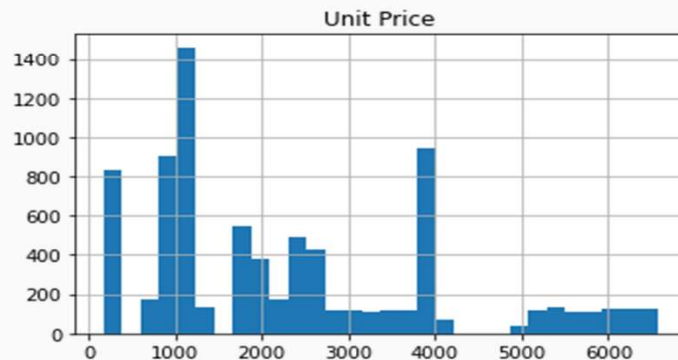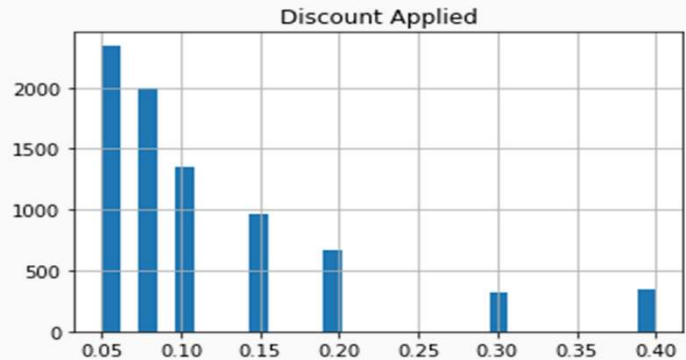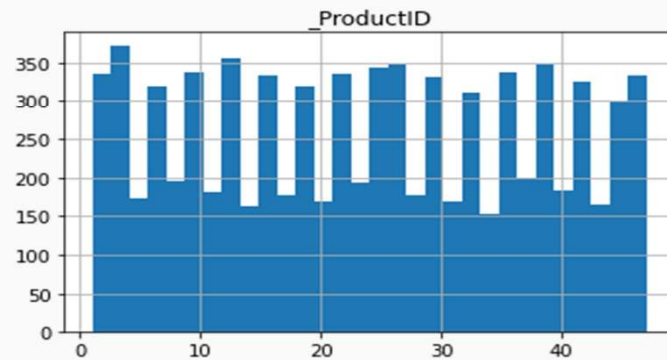
```
df.duplicated().sum()
```

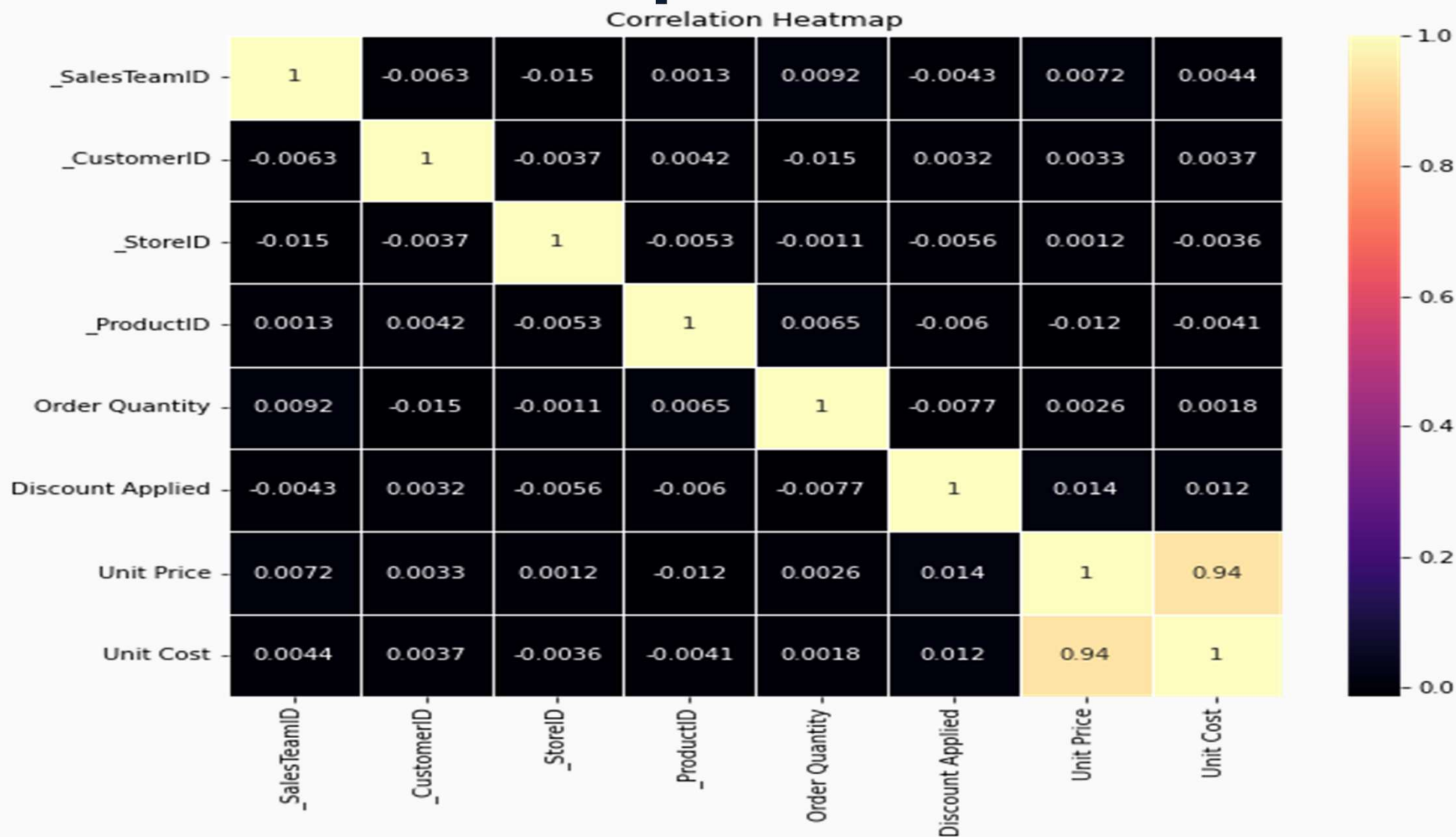```
np.int64(0)
```

رواد مصر الرقمية

# Box plots for outlier visualization

# histograms

# Correlation heatmap



Correlation Heatmap

Sales Channel Performance

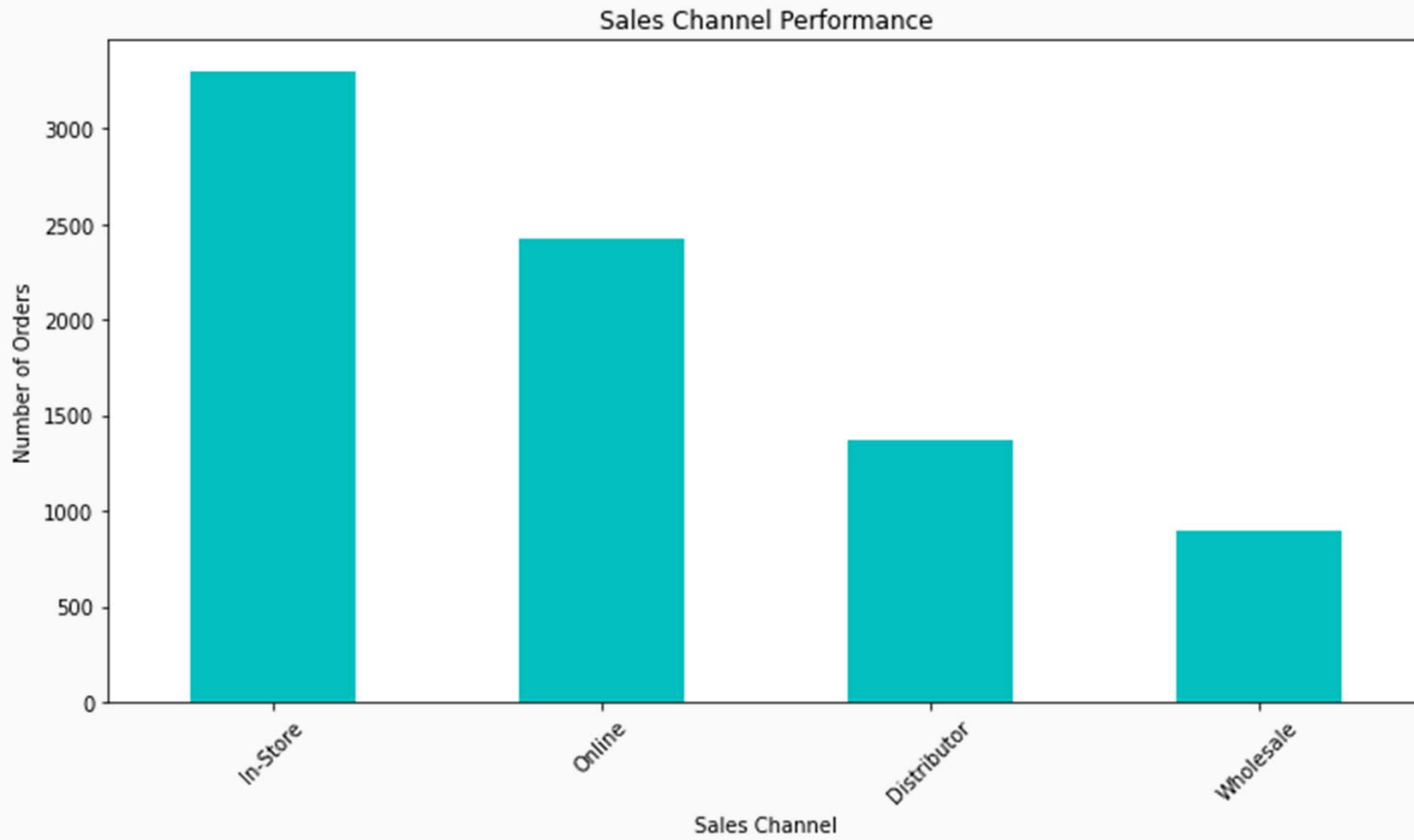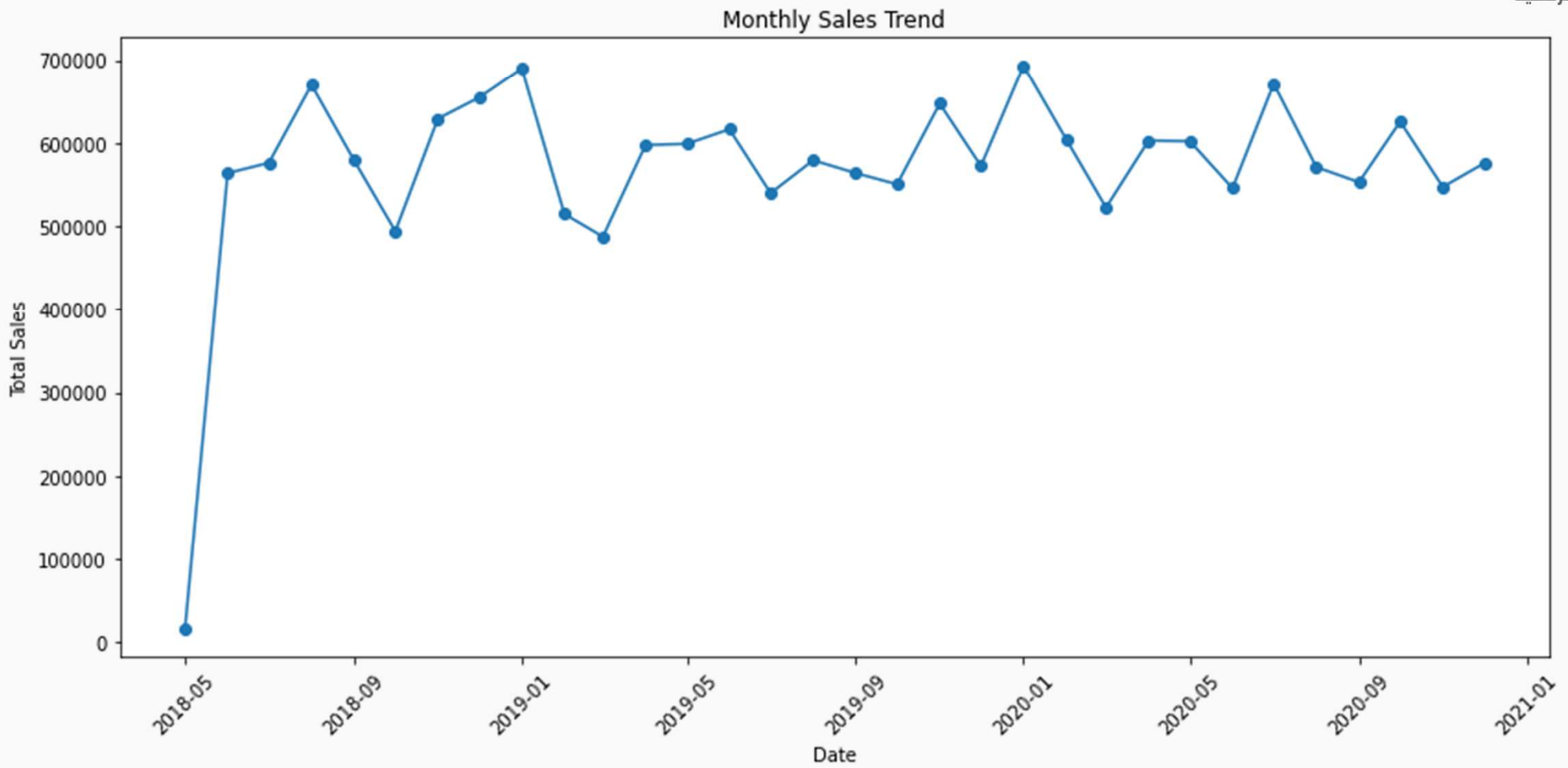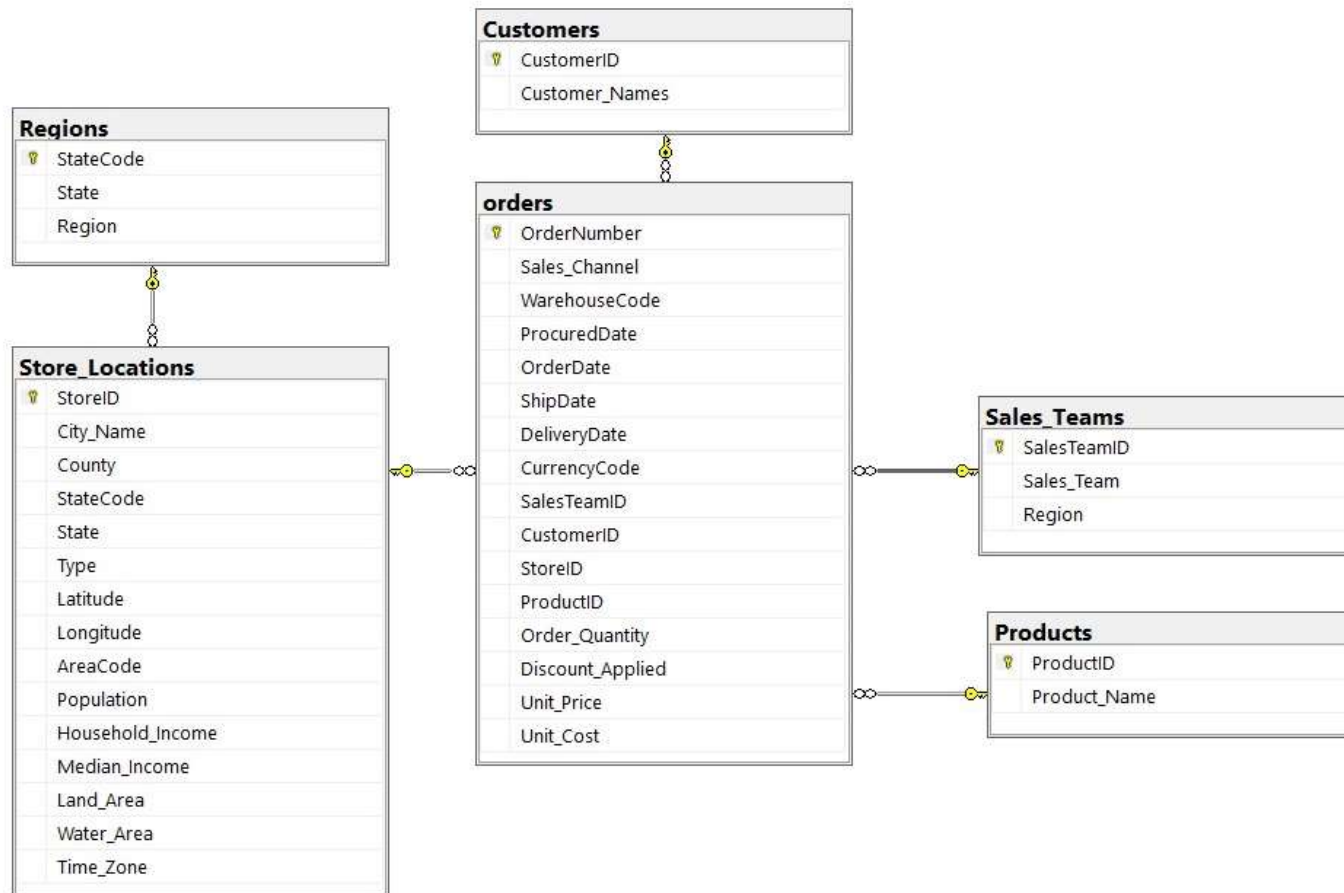Monthly Sales Trend
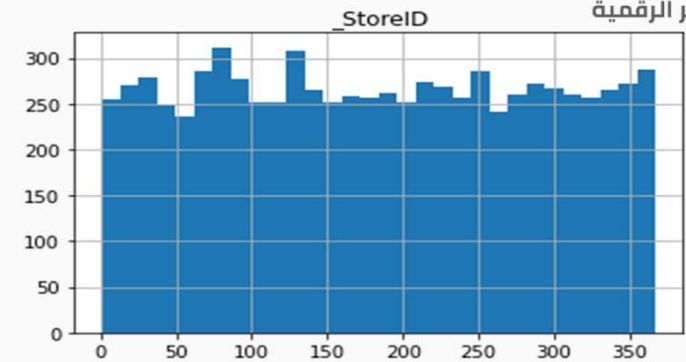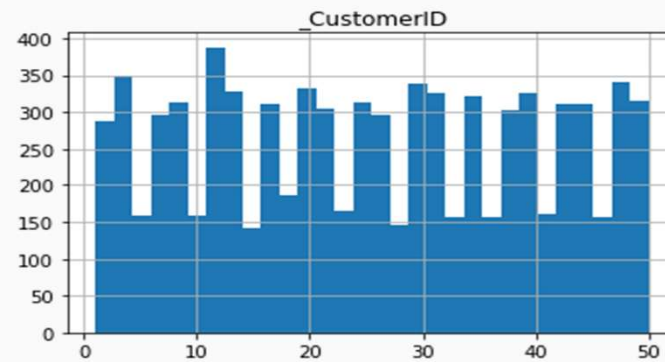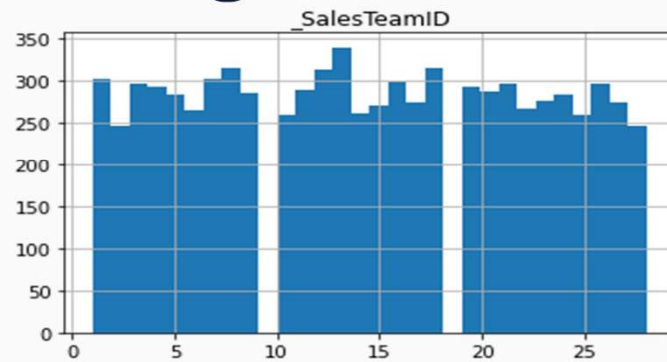
**04**

# Questions & Insights (SQL Queries)

# Data Modeling

# All Questions

## Sales Performance Analysis
- What are the total sales over time?
- How do sales vary across different sales channels (e.g., online vs. offline)?
- What is the average order value (AOV) across different stores or regions?
- Which products generate the most revenue or profit?
- What is the profit margin across various products and regions?
- What is the impact of discounts on sales and profitability?
- What is the trend of sales over specific time periods (daily, monthly, quarterly, annually)?
- How do shipping and delivery dates affect the profitability or customer satisfaction?

## Product Analysis
- Which products have the highest demand (based on order quantity)?
- Which products offer the highest profit margins?
- What is the average price and cost for each product category?
- Are there specific products that are consistently discounted, and what is their impact on profit?
- How does product performance vary across different regions or sales channels?

## Customer Analysis
- Which customers contribute the most to sales and profitability?
- What is the repeat customer rate?
- What is the average order quantity or value per customer?
- Are there any trends in customer behavior based on geographic location (city, county, or state)?
- How does household income, median income, or population affect customer purchasing behavior?
- How do time zones influence customer buying patterns?
- Which store locations generate the highest sales or profit?

## Regional and Store Analysis
- How do sales vary by city, county, or state?
- How does region (e.g., West, East, etc.) affect sales performance?
- How does population density or household income affect sales at a store level?
- What is the relationship between store location (latitude/longitude) and sales volume?
- How does the time between order date and delivery date affect sales performance?

## Time and Delivery Analysis
- Are there patterns or trends in sales based on the delivery date (e.g., holiday seasons)?
- How efficient are warehouses in fulfilling orders, based on ship and delivery times?
- How do procurement lead times (ProcuredDate to OrderDate) affect overall sales?
- Which sales team generates the most revenue?

## Sales Team Performance
- How does sales team performance vary across regions?
- What is the relationship between sales team performance and product or customer type?
- How do discounts applied by specific sales teams affect profit margins?
- How do sales and profitability vary by geographic factors such as county, state, and region?

## Geographic Analysis
- What impact does the store's proximity to customers (latitude/longitude) have on sales?
- Do areas with higher household income or population result in higher sales volumes?
- How does the time zone affect store sales and shipping patterns?
- How does unit cost vary across different products and stores?

## Cost and Profit Analysis
- What is the overall profit margin across different sales channels, regions, or time periods?
- How does the application of discounts affect overall profitability?
- Which products or stores offer the highest profit margin after considering costs and discounts?

رواد مصر الرقمية

# Answers Samples

```sql
--(3) Average sales per each region and store (What is the average order value (AOV) across different stores or regions?)
WITH Store_Revenue AS (
    SELECT
        sl.StoreID,
        r.Region,
        SUM(o.Order_Quantity * o.Unit_Price * (1 - o.Discount_Applied)) AS Total_Revenue,
        COUNT(DISTINCT o.OrderNumber) AS Total_Orders
    FROM orders o
    JOIN [Store_Locations] sl ON sl.StoreID = o.StoreID
    JOIN Regions r ON r.StateCode = sl.StateCode
    GROUP BY sl.StoreID, r.Region),
Region_Revenue AS (
    SELECT
        r.Region,
        SUM(o.Order_Quantity * o.Unit_Price * (1 - o.Discount_Applied)) AS Total_Revenue,
        COUNT(DISTINCT o.OrderNumber) AS Total_Orders
    FROM orders o
    JOIN [Store_Locations] sl ON sl.StoreID = o.StoreID
    JOIN Regions r ON r.StateCode = sl.StateCode
    GROUP BY r.Region)
SELECT
    'Store' AS Level,
    sl.StoreID AS Store_ID,
    sl.City_Name AS City_store,
    r.Region,
    ROUND(sr.Total_Revenue / sr.Total_Orders, 2) AS AOV
FROM Store_Revenue sr
JOIN [Store_Locations] sl ON sl.StoreID = sr.StoreID
JOIN Regions r ON r.StateCode = sl.StateCode

UNION ALL

SELECT
    'Region' AS Level,
    NULL AS Level_ID,     -- No StoreID for region-level aggregation
    r.Region AS Level_Name,
    r.Region,
    ROUND(rr.Total_Revenue / rr.Total_Orders, 2) AS AOV
FROM Region_Revenue rr
JOIN Regions r ON r.Region = rr.Region;
```

```sql
--(7) (What is the trend of sales over specific time periods (daily, monthly, quarterly, annually)?)
--Daily Sales Trend:
SELECT
    CONVERT(DATE, OrderDate) AS Sales_Date,
    SUM(Order_Quantity * Unit_Price) AS Total_Sales
FROM orders
GROUP BY
    CONVERT(DATE, OrderDate)
ORDER BY
    Sales_Date;
-- Monthly Sales Trend:
SELECT
    YEAR(OrderDate) AS Year,
    MONTH(OrderDate) AS Month,
    SUM(Order_Quantity * Unit_Price) AS Total_Sales
FROM orders
GROUP BY
    YEAR(OrderDate), MONTH(OrderDate)
ORDER BY
    Year, Month;
--Quarterly Sales Trend:
SELECT
    YEAR(OrderDate) AS Year,
    CEILING(MONTH(OrderDate) / 3.0) AS Quarter,
    SUM(Order_Quantity * Unit_Price) AS Total_Sales
FROM orders
GROUP BY
    YEAR(OrderDate), CEILING(MONTH(OrderDate) / 3.0)
ORDER BY
    Year, Quarter;
--Annual Sales Trend:
SELECT
    YEAR(OrderDate) AS Year,
    SUM(Order_Quantity * Unit_Price) AS Total_Sales
FROM orders
GROUP BY
    YEAR(OrderDate)
ORDER BY
    Year;
```

```sql
--(38) (How does the application of discounts affect overall profitability?)
-- Total Revenue With and Without Discounts
SELECT
    SUM(Order_Quantity * Unit_Price * (1 - Discount_Applied)) AS Sales_With_Discounts,
    SUM(Order_Quantity * Unit_Price) AS Sales_Without_Discounts
FROM orders;
-- Total Profit With and Without Discounts
SELECT
    SUM(Order_Quantity * Unit_Price * (1 - Discount_Applied) - Order_Quantity * Unit_Cost) AS Profit_With_Discounts,
    SUM(Order_Quantity * Unit_Price - Order_Quantity * Unit_Cost) AS Profit_Without_Discounts
FROM orders;
-- Profit Margin With and Without Discounts
SELECT
    ROUND(
        (SUM(Order_Quantity * Unit_Price * (1 - Discount_Applied)) - SUM(Order_Quantity * Unit_Cost))
        / NULLIF(SUM(Order_Quantity * Unit_Price * (1 - Discount_Applied)), 0), 2) AS Profit_Margin_With_Discounts,
    ROUND((SUM(Order_Quantity * Unit_Price) - SUM(Order_Quantity * Unit_Cost))
    / NULLIF(SUM(Order_Quantity * Unit_Price), 0), 2) AS Profit_Margin_Without_Discounts
FROM orders;
-- Impact of Discounts by Product
-- CTE to calculate revenue and profit margins
WITH Discounted_Sales AS (
    SELECT
        p.Product_Name,
        SUM(o.Order_Quantity * o.Unit_Price * (1 - o.Discount_Applied)) AS Sales_With_Discount,
        SUM(o.Order_Quantity * o.Unit_Price) AS Sales_Without_Discount,
        SUM(o.Order_Quantity * o.Unit_Cost) AS Total_Cost
    FROM orders o
    JOIN Products p ON o.ProductID = p.ProductID
    GROUP BY p.Product_Name)
SELECT
    Product_Name,
    Sales_With_Discount,
    Sales_Without_Discount,
    Total_Cost,
    ROUND((Sales_With_Discount - Total_Cost) / NULLIF(Sales_With_Discount, 0), 2) AS Profit_Margin_With_Discount,
    ROUND((Sales_Without_Discount - Total_Cost) / NULLIF(Sales_Without_Discount, 0), 2) AS Profit_Margin_Without_Discount
FROM Discounted_Sales;
```

05

Data Modeling & Visualization (Power BI)
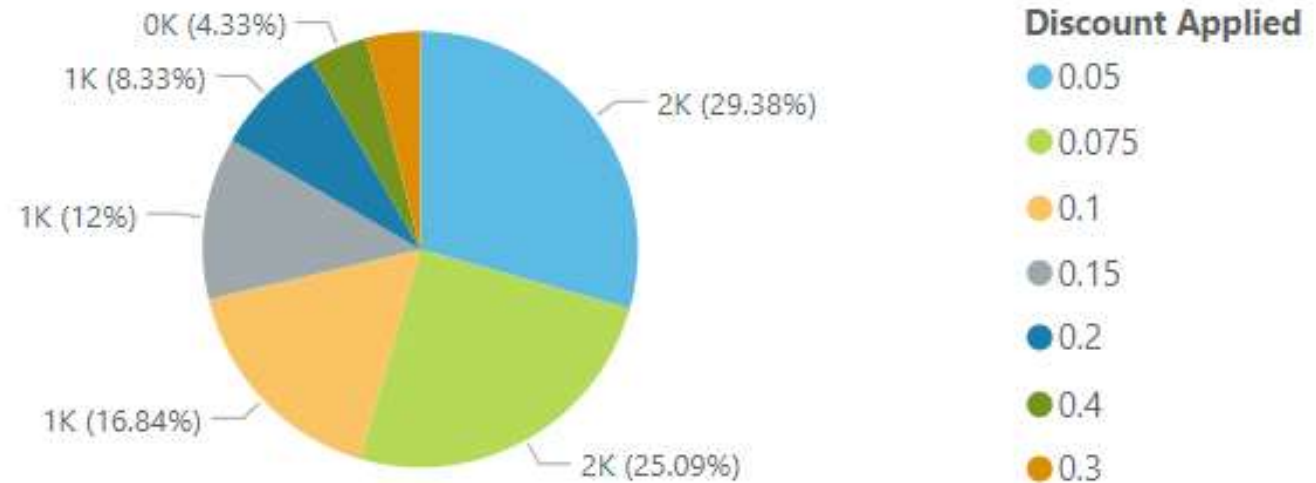
Dashboard

Results & Key Findings:

- **Sales Channels:** Online sales are growing significantly faster than in-store sales.
- **Discount Strategy:** While discounts drive sales, they negatively impact profit margins.
- **Top Products:** Cookware is the best-performing product by revenue, but photo frames yield the highest profit margins.
- **Regional Performance:** The South region dominates in terms of sales volume, while the West region excels in profitability.
- **Customer Insights:** Repeat customers, especially corporate clients, are a crucial driver of profitability.
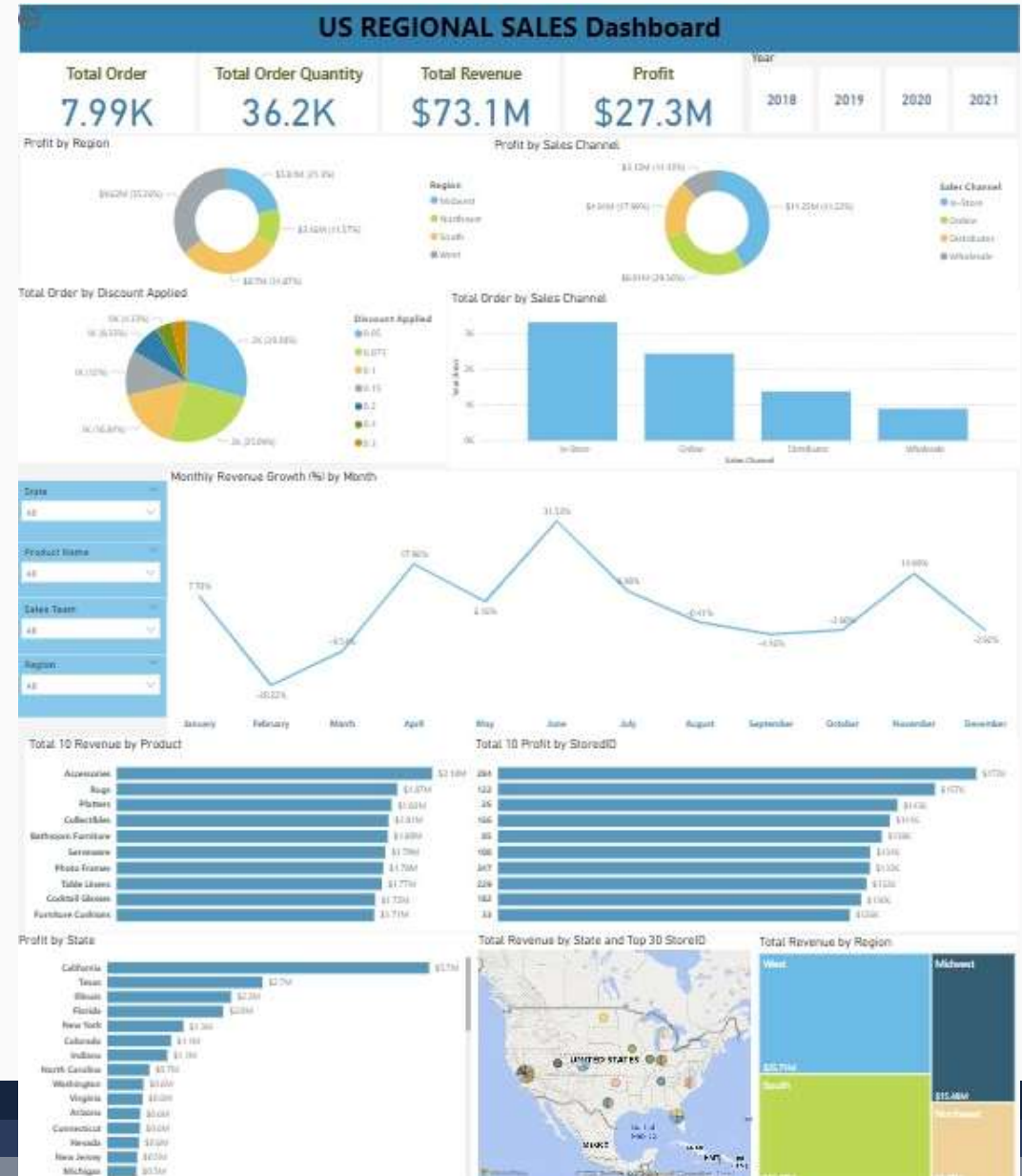
Question: What is the impact of discounts on sales and profitability?

Optimize Discount Strategies: Focus on strategic discounting for high-margin products to boost sales without significantly impacting profitability.
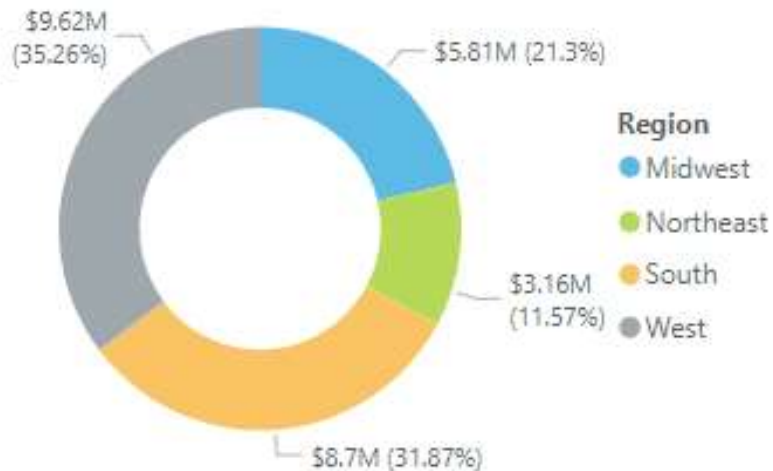

Total Order by Discount Applied

Discount Applied
- 0.05
- 0.075
- 0.1
- 0.15
- 0.2
- 0.4
- 0.3

- 0K (4.33%)
- 1K (8.33%)
- 2K (29.38%)
- 1K (12%)
- 1K (16.84%)
- 2K (25.09%)

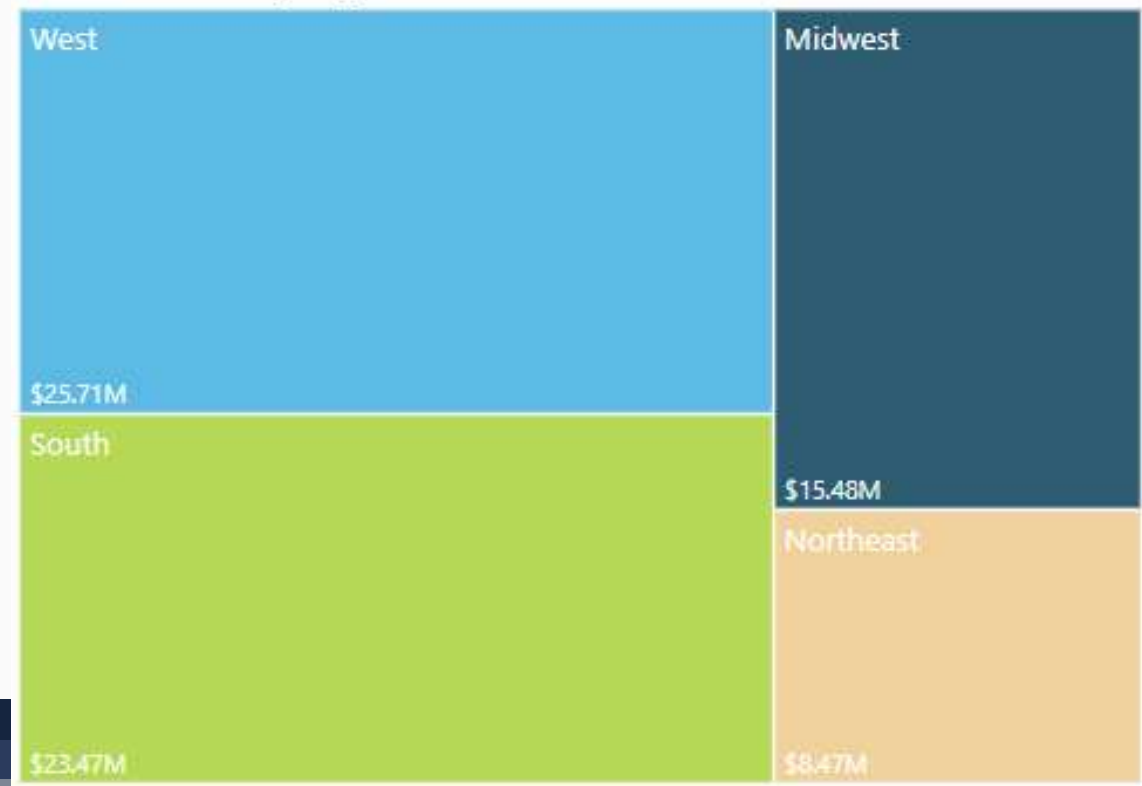# Question: What is the average order value (AOV) across different stores or regions?

Focus on the West Region: Expand operations in the West, where higher AOV and profitability are observed.

Profit by Region

- $9.62M (35.26%)
- $5.81M (21.3%)
- $3.16M (11.57%)
- $8.7M (31.87%)

Region
- ● Midwest
- ● Northeast
- ● South
- ● West

Total Revenue by Region

- West $25.71M
- Midwest $15.48M
- South $23.47M
- Northeast $8.47M

Question: How do sales vary across different sales channels (e.g., online vs. offline)?

Enhance Online Sales Channels: Invest in online marketing and logistics to capitalize on the growing online sales trend.

Profit by Sales Channel

$3.12M (11.43%)

$4.91M (17.99%)

$11.25M (41.22%)

$8.01M (29.36%)

Sales Channel
- In-Store
- Online
- Distributor
- Wholesale

06

Conclusion

Conclusion

This project successfully analyzed store sales data by utilizing a combination of Python, SQL, and Power BI. The analysis provided valuable insights into sales performance, customer demographics, product profitability, and regional disparities. By standardizing and cleaning the data, we ensured that the information was consistent and reliable for decision-making.

The regional sales analysis provided critical insights into how sales performance varies across channels, products, and regions. By optimizing discount strategies, focusing on profitable regions, and enhancing online channels, the company can achieve sustainable growth.

Key insights include:

1. **Sales Performance Trends:** Identified best-selling periods and most effective sales channels.
2. **Customer Insights:** High-income households were found to be key drivers of sales, indicating opportunities for targeted marketing.
3. **Product Profitability:** Certain products contributed disproportionately to profits, while others showed opportunities for improvement.
4. **Regional Disparities:** Specific regions outperformed others, suggesting the need for a re-evaluation of distribution strategies.

By visualizing these insights using Power BI, the findings were made accessible and actionable, guiding future business strategies aimed at maximizing sales efficiency and profitability.

Thanks !

Do you have any questions for us?

رواد مصر الرقمية

Our Mentor

Mennatullah Selim

Presenters

➢Ahmed Mokhtar Mohamed
➢Abdulrahman Wheed
➢Ali Abdulhalim
➢Maher Ahmed
➢Mahmoud Mounir

Resources

All sources and files for the project can be found here:

https://drive.google.com/drive/folders/1-
4awDQFdIegQ6lt8vTOiehB-_PF3XRDf?usp=sharing

رواد مصر الرقمية