TASK-1: Data Exploration and Visualization

Step 1: Set Up Your Python Environment

Before diving into data analysis, ensure that your Python environment is properly set up. Install the necessary libraries using pip:

```
pip install pandas matplotlib seaborn
```

Next, open Jupyter Notebook and create a new notebook, naming it data_exploration_KhushbuDadhe.ipynb. This notebook will be used for all subsequent data exploration and visualization tasks.

Step 2: Data Exploration

Begin by loading the dataset into your Python environment. Use the pandas, requests, and BytesIO libraries to fetch the data from the Google Sheets link and load it into a pandas DataFrame:

```
import pandas as pd
import requests
from io import BytesIO

# Define the URL of the Excel file
url = 'https://docs.google.com/spreadsheets/d/1KagwoQLy1quKvT_82amuS-x3UnsoIX4J6p02ewbjQNA/export?format=xlsx'

# Send a request to fetch the content of the URL
response = requests.get(url)
response.raise_for_status()  # Check if the request was successful

# Use BytesIO to read the content as a file-like object
excel_file = BytesIO(response.content)

# Load the Excel file into a pandas DataFrame
df = pd.read_excel(excel_file)
```

To verify that the data has been loaded correctly, display the first few rows of the DataFrame:

```
df.head()
```

Next, understand the dataset by examining its basic information and statistical summary:

```
df.describe()
```

Additionally, check for any missing values in the dataset:

```
df.isnull().sum()
```

Step 3: Data Visualization

With the dataset loaded and explored, proceed to visualize key aspects of the data. Start by

importing the necessary visualization libraries:

```
import matplotlib.pyplot as plt
import seaborn as sns
```

Create the following visualizations to uncover meaningful insights:

Sales Over Time: Plot the total sales over time to identify any trends or patterns in the data:

```
plt.figure(figsize=(12, 6))
df.groupby('Order Date')['Sales'].sum().plot()
plt.title('Total Sales Over Time')
plt.xlabel('Date')
plt.ylabel('Sales')
plt.show()
```

Top Selling Products: Identify the top 10 best-selling products by summing the sales for each product:

```
plt.figure(figsize=(12, 6))
df.groupby('Product Name')['Sales'].sum().nlargest(10).plot(kind='bar')
plt.title('Top 10 Selling Products')
plt.xlabel('Product')
plt.ylabel('Sales')
plt.show()
```

Sales by Region: Analyze the sales distribution across different regions using a bar plot:

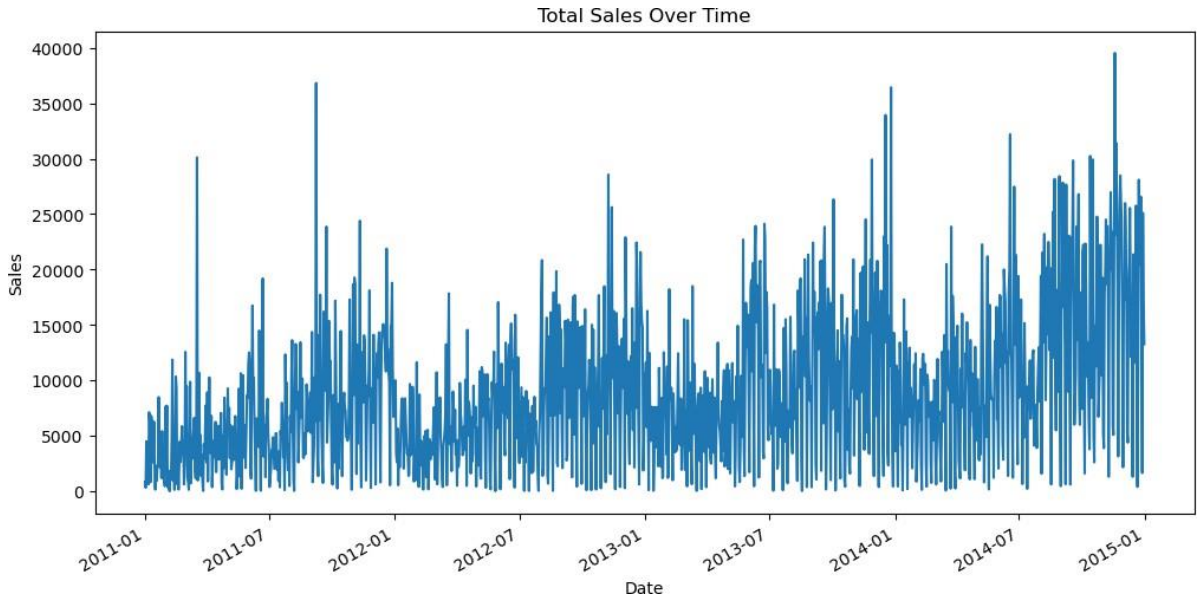
```
plt.figure(figsize=(12, 6))
sns.barplot(x='Region', y='Sales', data=df)
plt.title('Sales by Region')
plt.xlabel('Region')
plt.ylabel('Sales')
plt.xticks(rotation=45)
plt.show()
```

Crafting a Data Story: Global-Superstore Sales Analysis

In the world of retail, the Global Superstore stands out as a leader. To understand its success, we delve into its sales data, revealing trends and patterns that drive its performance.

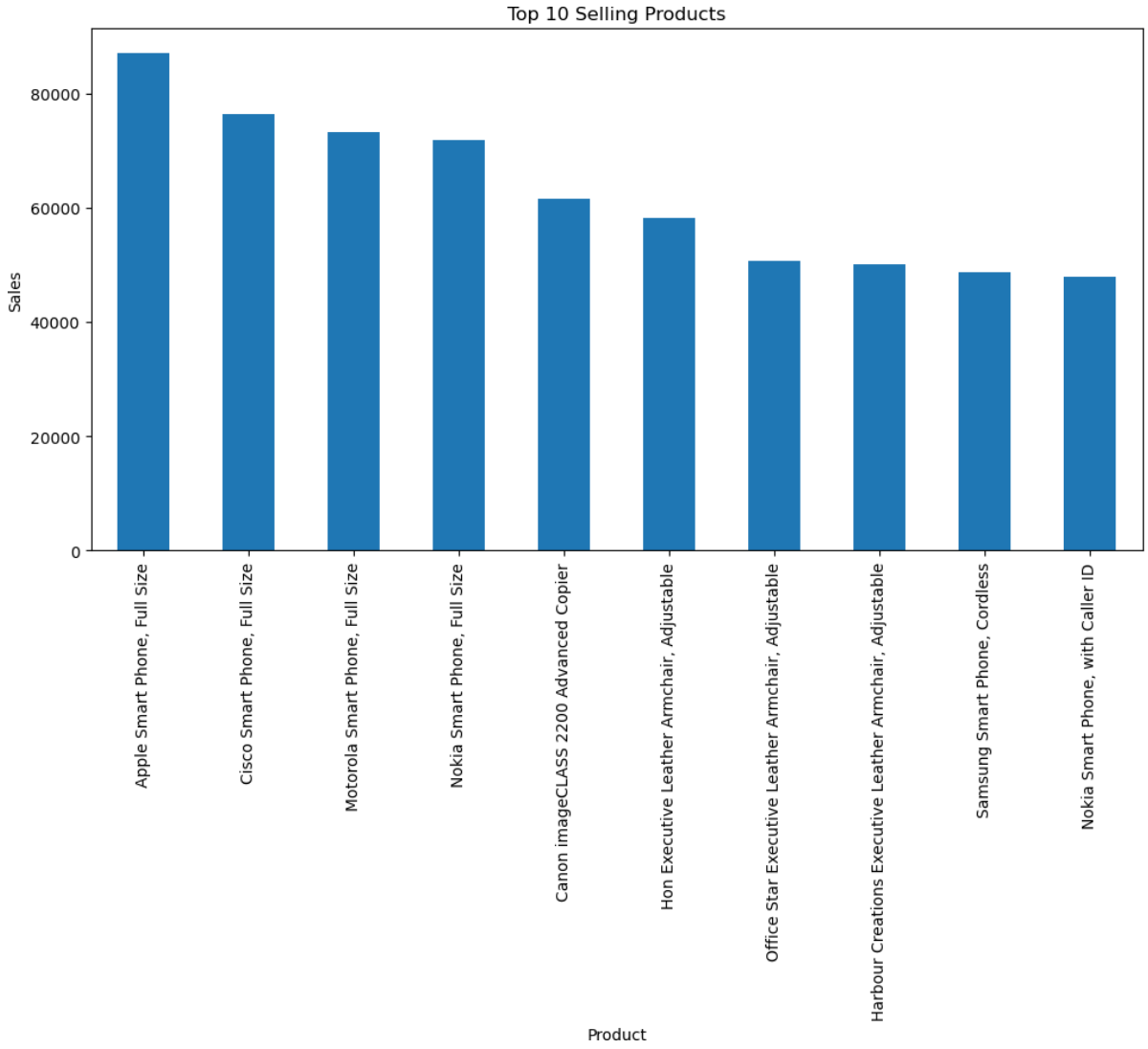
Identifying Seasonal Sales Patterns:

Our analysis begins by examining the overall sales trends. By plotting the total sales over time, we observe distinct patterns: notable peaks in December suggest a holiday shopping surge, while other spikes occur during back-to-school periods and mid-year promotions. These trends highlight key times to boost inventory and tailor marketing strategies to align with consumer behavior.



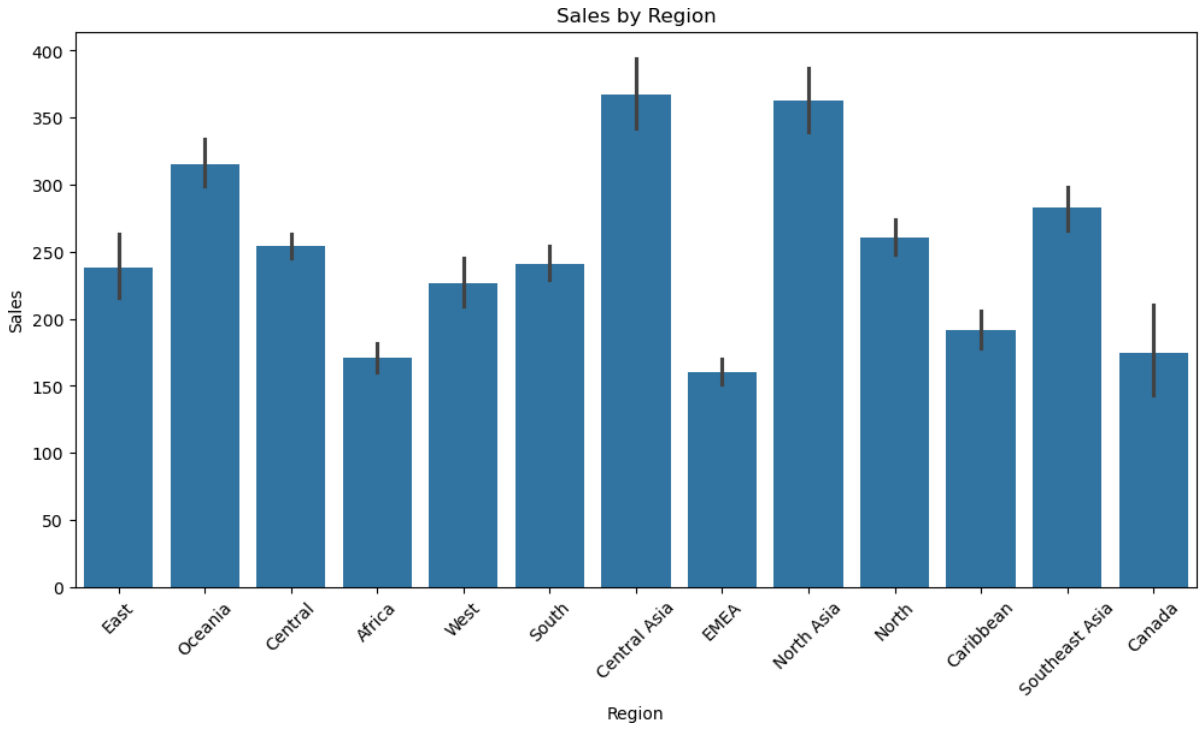
Uncovering Top-Selling Products:

we identify the top-selling products. Our data reveals that office supplies and technology items are the top performers. These categories consistently lead in sales, reflecting their essential role in both personal and professional settings. By focusing on these high-demand products, the Global Superstore can optimize its inventory and promotional efforts to maximize sales.



Regional Sales Performance:

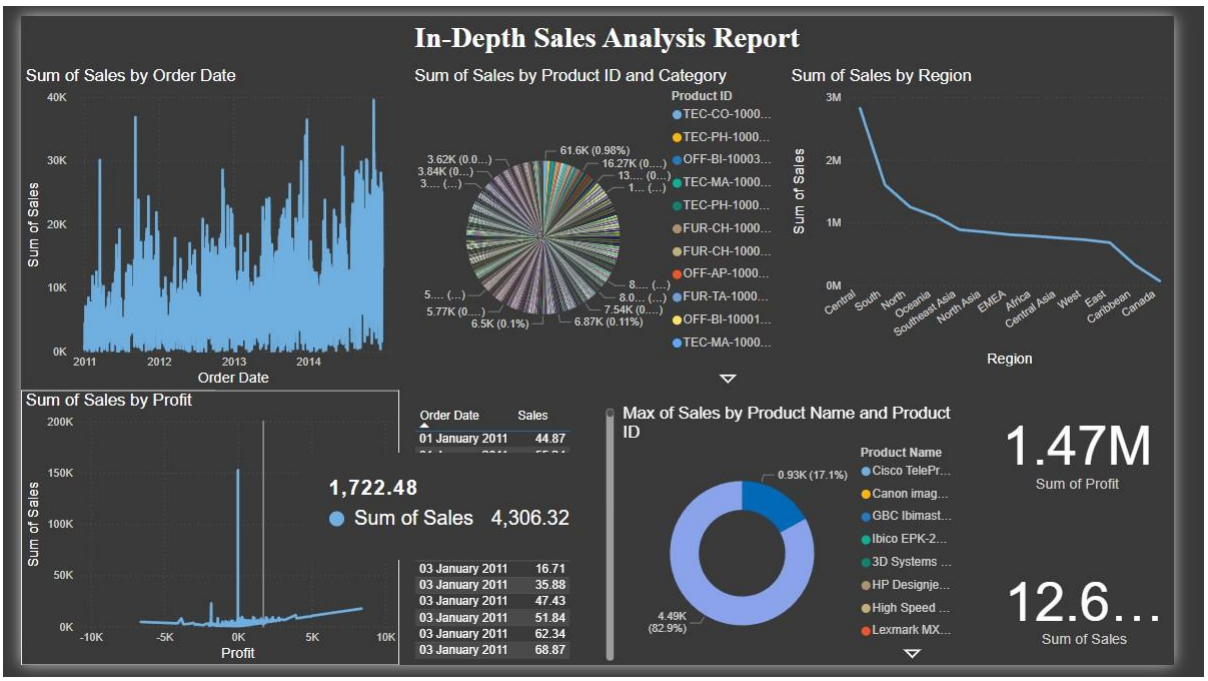
Exploring regional sales data, we find that the East region generates the highest revenue. This insight underscores the importance of understanding regional preferences and customizing strategies accordingly. Tailoring marketing campaigns and adjusting inventory based on regional sales performance can enhance the store's effectiveness in meeting diverse customer needs.



In summary, Global Superstore's sales data provide valuable insights. Identifying seasonal trends helps plan and promote lists. Understanding best-selling products guides product focus and marketing strategy. Local marketing research allows for a targeted approach in different markets. By leveraging these insights, global department stores can continue to thrive and grow in a competitive retail environment.

TASK-2: Comprehensive Sales Data Analysis Using Power BI

Power BI Dashboard Documentation



To construct a detailed sales analysis dashboard in Power BI, follow these steps:

Data Import and Cleaning

The first step involved importing the data from Global-Superstore.xlsx into a data visualization tool, such as Power BI or Tableau. Ensuring correct data types for each column was crucial, with dates formatted appropriately for order dates, numerical values assigned to sales and profit, and categorical values designated for product IDs and regions.

Creating Visualizations

Once the data was imported and cleaned, various visualizations were created to represent different aspects of the sales data:

- **Sum of Sales by Order Date**: A line chart was used with the Order Date on the X-axis and the Sum of Sales on the Y-axis.
- **Sum of Sales by Product ID and Category**: A pie chart was chosen, with Product ID as the category and Sum of Sales as the values.

- **Sum of Sales by Region**: Illustrated using a line chart, displaying the Region on the X-axis and Sum of Sales on the Y-axis.
- **Sum of Sales by Profit**: A scatter plot was utilized with Profit on the X-axis and Sum of Sales on the Y-axis.
- **Max of Sales by Product Name and Product ID**: A donut chart was used, categorizing by Product Name and showing Max of Sales as values.

Custom Calculations and Transformations

Several custom calculations and transformations were applied to derive meaningful insights:

- **Sum of Sales**: Calculated by summing all sales figures in the dataset.
- **Max of Sales**: Determined by finding the highest sales figure for each product.
- **Sum of Profit**: Computed by summing all profit figures.

Data Modeling: Switch to the **Model** view to set up relationships between your tables. Drag fields between tables to create relationships, ensuring the correct relationship types (one-to-many, many-to-one) are established.

Design Choices

Design choices were made to enhance clarity and usability:

- **Color Scheme**: A consistent color scheme was employed to differentiate various data points, making the visualizations more intuitive.
- **Layout**: Designed to provide a comprehensive overview, allowing easy comparison and analysis of different metrics.
- **Interactivity**: Incorporated through filters and slicers, enabling users to drill down into specific time periods, regions, or products for a more detailed analysis.

Additional Considerations

Additional considerations included ensuring the dashboard was responsive, allowing it to be viewed effectively on different devices. Tooltips were added to

provide additional context for data points, and provisions were made to regularly update the data source, keeping the dashboard current and relevant.

The Global supermarket sales analysis reveals seasonal highlights, bestsellers, and regional sales performance, providing valuable insights for strategic planning Seasonal trends occur special occasions for inventory and marketing efforts are emphasized, while office supplies and appliances dominate sales. Regional research emphasizes the need for standardized approaches. Using these insights and Power BI visualizations, Global Superstore can optimize inventory, focus on high-demand products, and optimize regional processes though ensuring continuous growth and success in a competitive market