**TITLE: SALES DATA ANALYSIS USING POWERBI and Python**

# **INTRODUCTION**

This report presents the step-wise approach to analyze the given sales data using PowerBI and Python. The goal is to extract meaningful insights through data visualization and statistical analysis. This report includes data loading, data cleaning, transformation, visualizations, and analysis to extract the valuable business insights.

# **OBJECTIVES**

* To transform and load the given data.
* To clean the data.
* To extract the business insights in dashboard.
* **Using PowerBI**

# **STEPS FOLLOWED**

1. **Loading Data into PowerBI**

* Open **Power BI Desktop**.
* Click **Home > Get Data > CSV** (or any other data source).
* Select the **sales\_data.csv** file and click **Load**.

1. **Data Transformation**

Before analysis, we clean and transform the data using Power Query Editor.

* 1. **Changing Data Types**
* Ensure **dates** are in the correct format (Date/Time).
* Ensure all numerical columns are Whole Number or Decimal.

**2.2. Removing Duplicates**

* Use **"Remove Duplicates"** in Power Query for unique records.

After transformations, **click “Close & Apply”** to save changes.

**2.3. Creating New Columns (DAX Calculations)**

* Click **Modeling>New Measure**

**Add DAX:**

* **Total Sales** = SUM(sales\_dataset[Sales])
* **Total Transactions** = COUNT(sales\_dataset[Order\_ID])
* **Yearly Sales** = CALCULATE(SUM(sales\_dataset[Sales]), YEAR(sales\_dataset[Order\_Date]))
* **Top Product Sales** =

MAXX(

TOPN(

1,

SUMMARIZE(sales\_dataset, sales\_dataset[Product\_Name], "Product Sales", SUM(sales\_dataset[Region])),

[Product Sales],

DESC

),

[Product Sales]

)

* **Top Product Name** =

MAXX(

TOPN(

1,

SUMMARIZE(sales\_dataset, sales\_dataset[Product\_Name], "Product Sales", SUM(sales\_dataset[Sales])),

[Product Sales],

DESC

),

sales\_dataset[Product\_Name] )

1. **Creating Visualizations in Power BI**

## **3.1. Key Performance Indicators (KPIs)**

* **Total Sales**: **Card Visual**
* **Best-Selling Product**: **Card Visual**
* **Total Transactions**: **Card Visual**

## **3.2. Sales Trends Over Time**

* **Visualization:** **Line Chart**
* **X-axis:** Month
* **Y-axis:** Total Sales
* **Insight:** Identifies peak sales months as **August ($0.30M)** and low sales months as **March ($0.12M)**
* **Recommendation: Offer seasonal discounts in March and October to boost sales.**

## **3.3. Regional Sales Performance**

* **Visualization:** **Bar Chart & Pie Chart**
* **Axis:** Region vs. Total Sales
* **Insight:** Central region dominates sales (**56% share**), while the North region has the lowest sales (**21.97%**).
* **Recommendation**: Focus on **marketing strategy** and **promotional campaign** in the North region to balance sales.

## **3.4. Product Performance**

* **Visualization:** **Bar Chart**
* **Axis:** Product Name vs. Sales
* **Insight:** **Nokia Smart Phone, Full size** is the top-selling product.
* **Recommendation:** Increase stocks and provide necessary accessories in bundle for the best-selling product to maximize revenue.

### **3.5. Customer Segmentation**

* **Visualization:** **Pie Chart**
* **Segments:** Very High, High, Medium, Low Spending Customers
* **Insight:** **39.17%** of customers are high spenders.
* **Recommendation:** Provide reward points and special offers for loyal customers to encourage repeat purchases.

# **4. Dashboard Interactivity (Filters & Slicers)**

To enhance usability, **filters and slicers** are added:

* **Date Slicer:** Select sales data for a specific year/month or All.
* **Region Slicer:** View sales by Central, North, South or All.
* **Product Category Slicer:** Analyze performance by Office Supplies, Technology, Furniture or All.

## **5. Country-Wise Sales Breakdown (Treemap)**

* **Visualization: Treemap**
* **Insight:** France, United Kingdom, and Germany are top-performing countries, while Denmark, Portugal, and Ireland have lower sales.
* **Recommendation:** Conduct market research to determine barriers to sales in underperforming countries and implement localized promotions.

# **6. Key Findings and Insights**

## **6.1. Overall Sales Performance**

* **Total Sales:** $2.35M (Significant increase)
* **Total Transactions:** 8K
* **Best-Selling Product:** Nokia Smart Phone, Full Size
* **Insight:** The increase in total sales suggests strong product demand and effective sales strategies.
* **Recommendation:** Strengthen supply chain management to maintain product availability, enhance marketing efforts for high-performing products, and explore expansion opportunities (online marketplace or retail partnerships) to sustain sales growth.

## **6.2. Monthly Sales Trends**

* **Peak Sales Month:** **August (~$0.30M)**
* **Lowest Sales Month:** **March (~$0.12M)**
* **Insight:** Sales peak in **August and September** but drop in **March and October**.
* **Recommendation:** Introduce limited-time offers to maintain sales momentum during off-peak months.

## **6.3. Regional Sales Breakdown**

* **Top-Performing Region:** **Central (56% of total sales)**
* **North and South regions contribute only ~44% combined.**
* **Insight:** Central dominates sales, while North and South regions need improvement.
* **Recommendation:** Focus marketing efforts and distribution improvements in the North and South region to balance sales distribution.

## **6.4. Country-Wise Sales Insights**

* **Top 3 Countries: France, Italy, Germany**
* **Low-Sales Countries:** **Denmark, Portugal, and Ireland**
* **Insight:** Sales are concentrated in a few major countries, indicating **market penetration gaps elsewhere**.
* **Recommendation:** Conduct market research to identify barriers to sales in under-performing countries and implement localized promotions.

## **6.5. Product Performance**

* **Top-Selling Product: Nokia Smart Phone, Full Size**
* **Top Product Categories:**
  + **Technology ($886K, 38%)**
  + **Office Supplies ($823K, 35%)**
  + **Furniture ($638K, 27%)**
* **Insight: Technology and Office Supplies drive the highest revenue while furniture is the lowest-performing category.**
* **Recommendation:** Increase stock levels and explore bundling options for the best-selling product to maximize revenue.

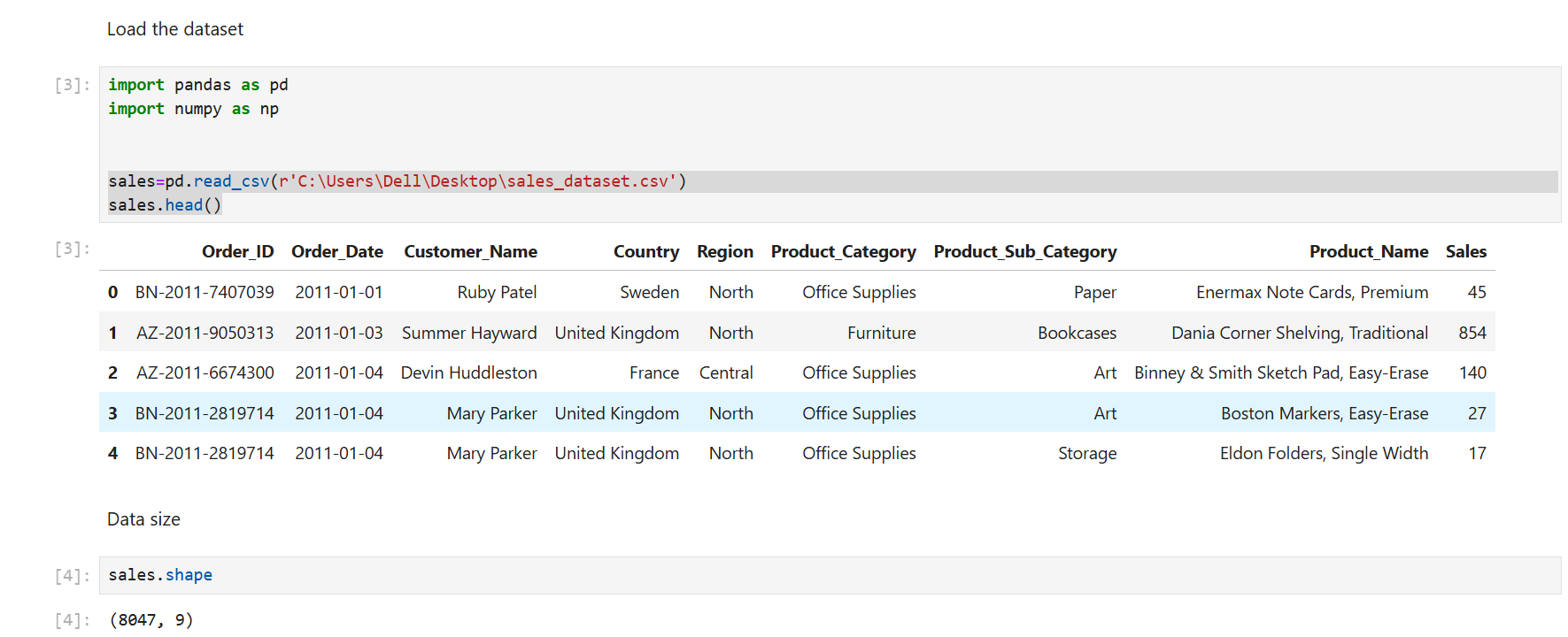
## **6.6. Customer Segmentation Analysis**

* **High-Spending Customers (Very High + High): ~70% of total sales**
* **Medium & Low-Spending Customers: Only ~30%**
* **Insight:** The majority of revenue comes from high-spending customers while medium and low-spending customers represent **a growth opportunity**.
* **Recommendation:** Introduce premium membership and special offers to encourage high-value purchases.
* **Python**

**Steps Followed:**

**1. Load the dataset**

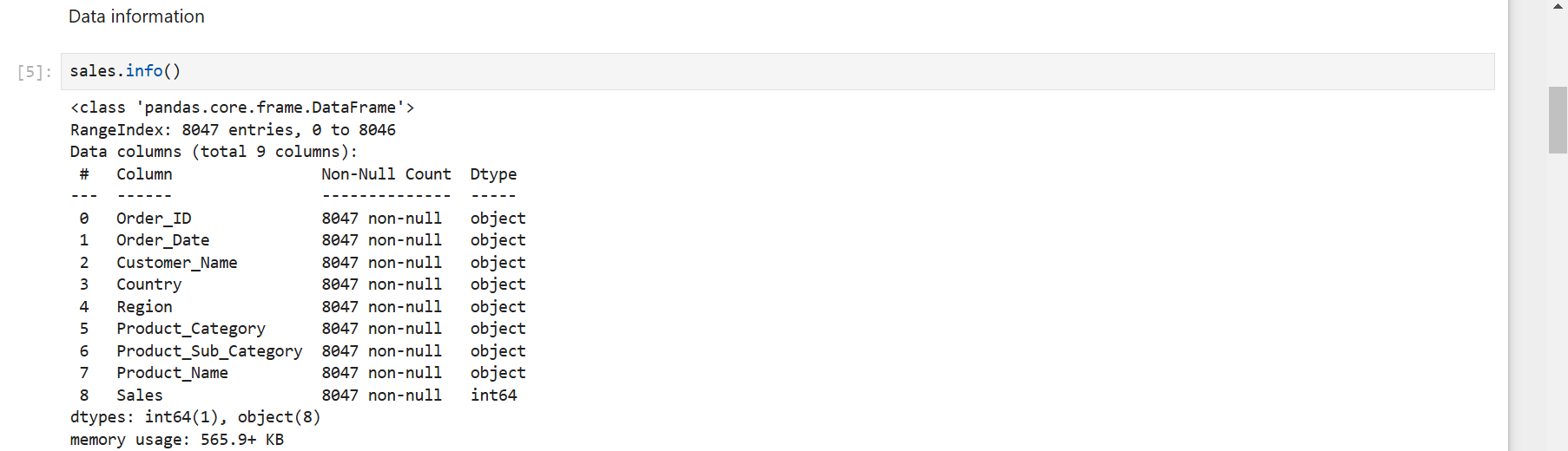
* **Objective**: Import the dataset into the working environment.
* **Action**:
  + Used **pandas** library to read the dataset from a CSV file.
  + File loaded using: **sales=pd.read\_csv(r'C:\Users\Dell\Desktop\sales\_dataset.csv')**
  + And **sales.shape** display the size of data



**Result**: Dataset successfully loaded into a pandas DataFrame and displayed the data size.

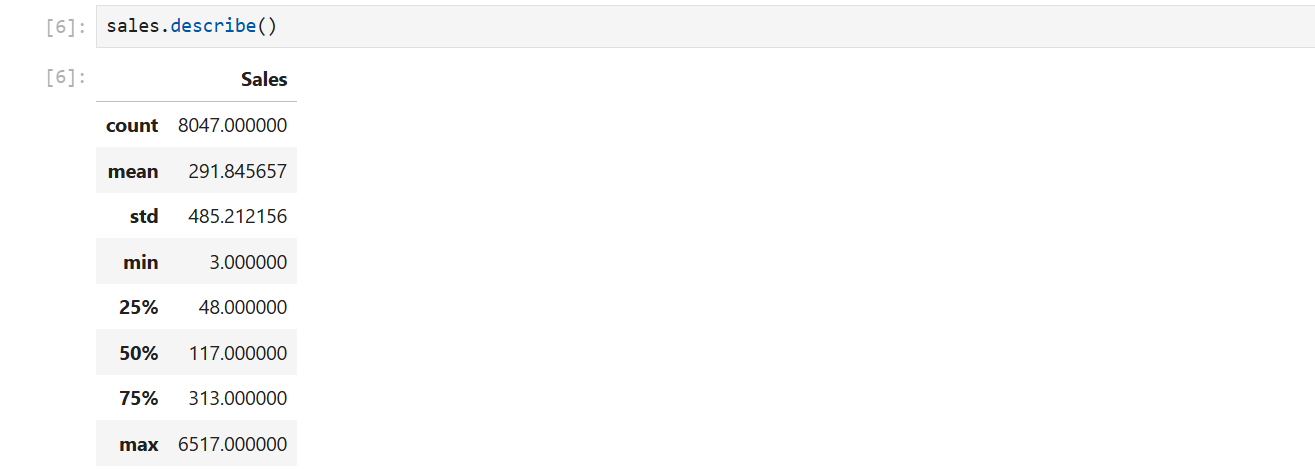
**2. Displayed the Data Information**

* **sales.info()** display the information of data.



**3. Displayed the sales description**

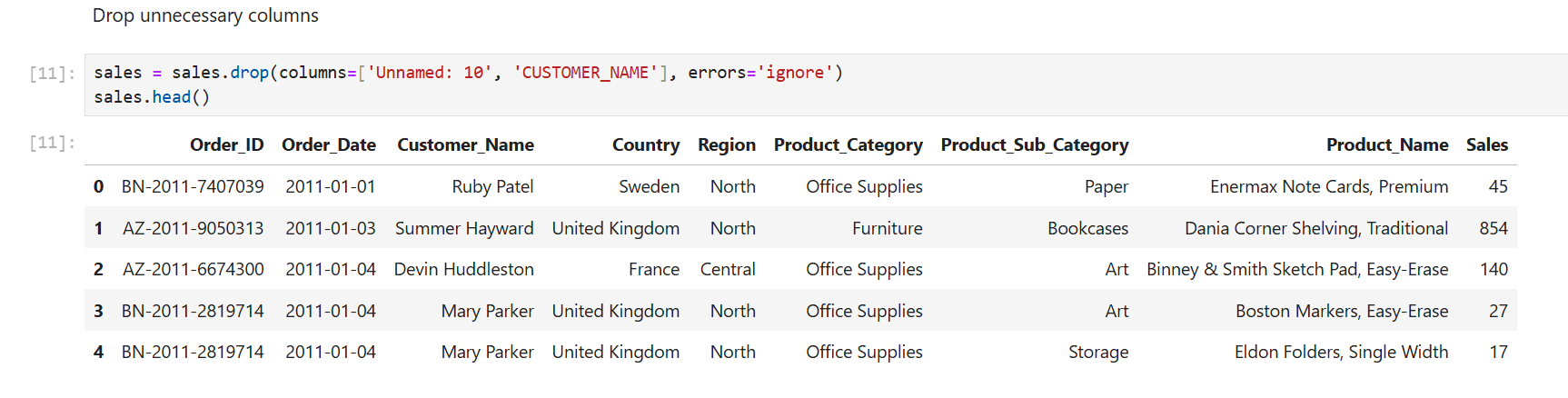
* **sales.describe()** displays the description of sales data**.**



**Result:** Displayed the description of data (count, mean, standard deviation, min, median, max value)

**4. Remove Duplicates and Irrelevant Data**

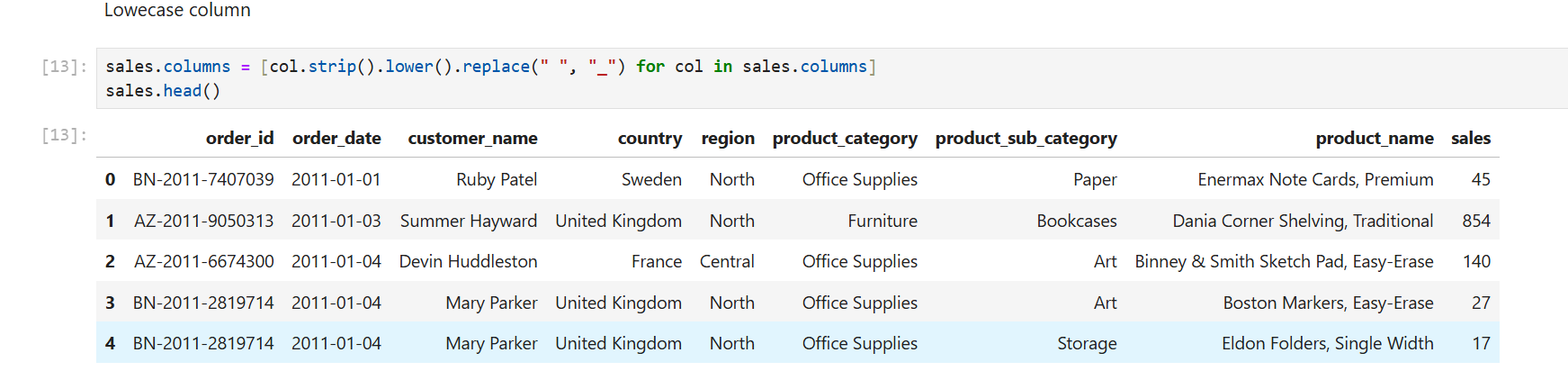
* **Objective**: Eliminate redundant entries and unnecessary columns.
* **Action**:
  + Checked for duplicate rows using duplicated().sum().
  + Removed duplicate rows using drop\_duplicates().
  + Dropped irrelevant columns.

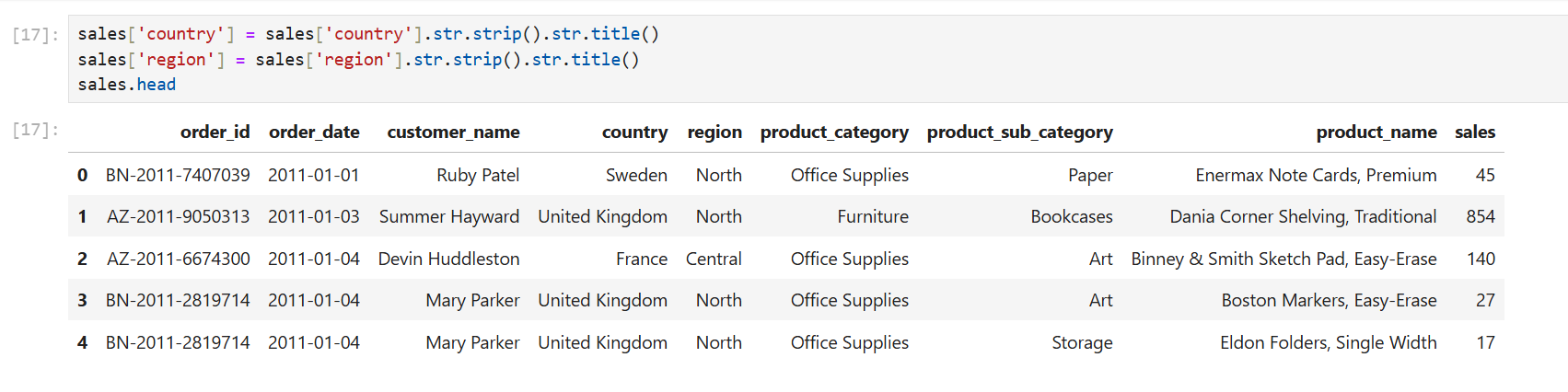


**Result**: Dataset size reduced by removing duplicates and irrelevant columns.

**4.1. To Standardize the Dataset**

* **Objective**: Ensure consistency in formatting and data types across the dataset.
* **Action**:
  + Standardized column names by converting to lowercase and replacing spaces with underscores.
  + Cleaned string columns by removing leading/trailing spaces and converting to title case.
* **sales.columns = [col.strip().lower().replace(" ", "\_") for col in sales.columns]** lowercase the column.



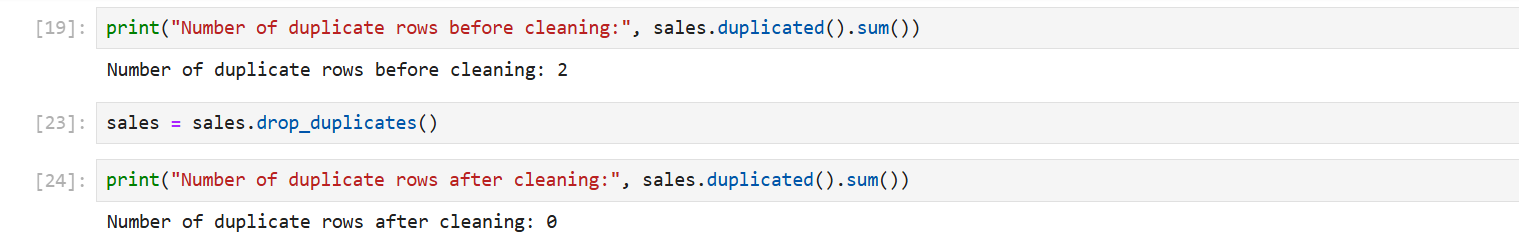


**Result**: Dataset is now uniform, with consistent formatting and types.

**4.2. To Handle the Missing Values and cleaning duplicate data**

* **Objective**: Address incomplete data to avoid errors during analysis.
* **Action**:
* Identified columns with missing values using **isnull().sum().**
* Applied the following strategies:
  + Numerical columns: Filled missing values with the column median.
  + Categorical columns: Filled missing values with the column mode.
  + **sales = sales.drop\_duplicates()** removes the duplicate data from the dataset.





* **Result**: Missing values and duplicate values were appropriately handled, ensuring no null and duplicatevalues remain.

**4.3. To Prepare the Dataset for Analysis**

* **Objective**: Make the dataset analysis-ready by ensuring it is clean, structured, and reliable.
* **Action**:
  + Validated the dataset by confirming no duplicates, missing values, or inconsistencies remain.
* **Result**: The dataset is fully prepared for exploratory data analysis and modeling.

## **4.4.** **Validation**

After cleaning, the dataset was validated:

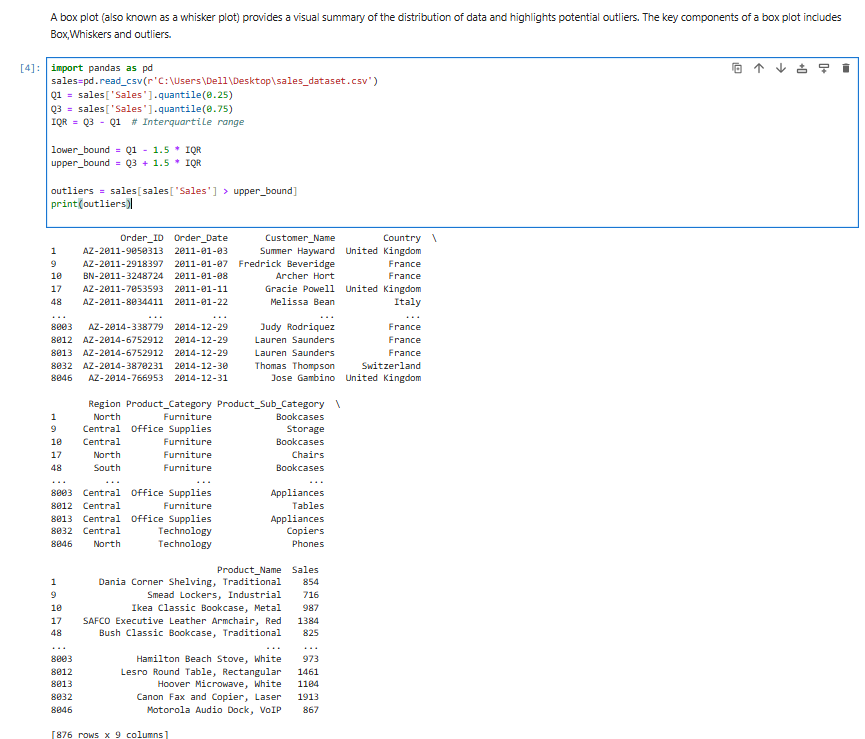
* Checked for duplicates: None found.
* Verified missing values: All missing values addressed.
* Confirmed consistent formats and types.

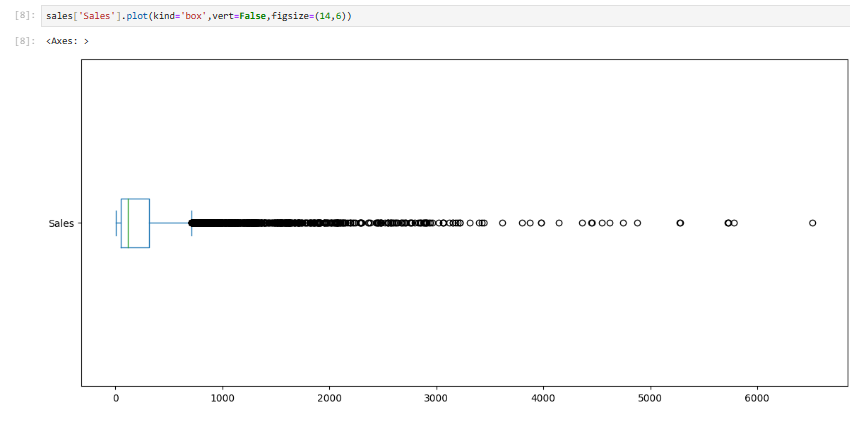
**4.5. Data Visualization**

Data Visualization is a graphical representation of data using charts, graphs, maps, and dashboards. It helps users to analyze trends, patterns, and insights by converting raw data into meaningful visual formats.

1. **Displaying Outliers**

A box plot (also known as a whisker plot) provides a visual summary of the distribution of data and highlights potential outliers. The key components of a box plot includes Box,Whiskers and outliers.

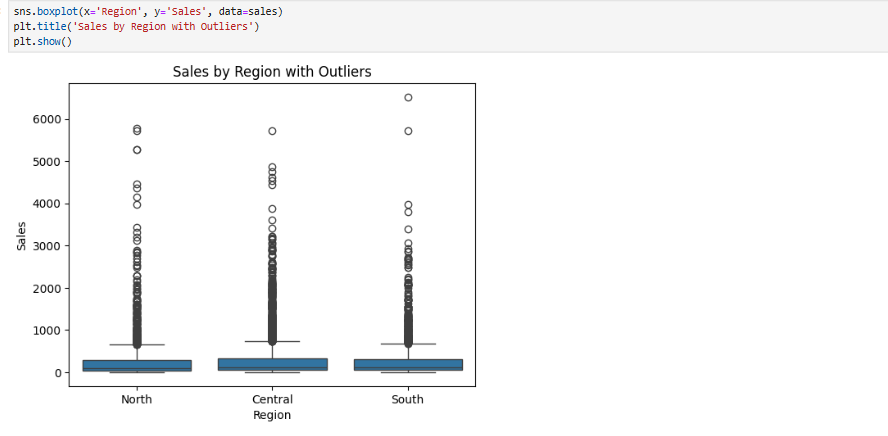


****

**Result:**

From above Visulization we get:

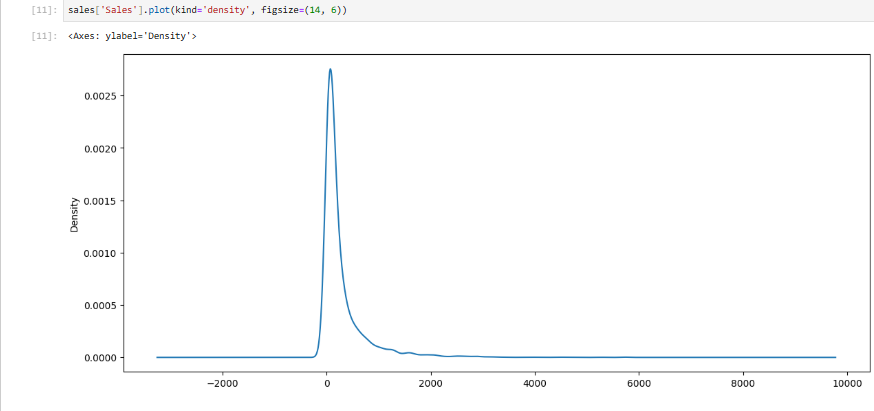
* Most sales values are concentrated at the lower range, as indicated by the narrow box and dense clustering near the median.
* A significant number of outliers exist, representing high sales values far from the main data distribution (to the right). This suggests occasional large transactions or orders.
* The distribution of sales is positively skewed, as the whisker extends much further to the right, showing a long tail of higher sales values.
* The compact interquartile range (IQR) indicates consistent sales for most transactions, except for a few large-value outliers.



**Result:** This boxplot displays the distribution of sales across different regions, highlighting the spread, central tendency, and potential outliers for each region. The x-axis represents the regions, while the y-axis shows the corresponding sales values.

1. **Density Plot**

A density plot is a graphical representation that shows the distribution of a continuous variable. It estimates the probability density function of the variable, providing insights into its underlying distribution.

****

**Result:**

1. Most sales are concentrated around low values, indicating frequent smaller purchases.

2. A long right tail suggests rare but significant high-value transactions.

3. Negative values might represent refunds, returns, or data errors requiring further investigation.

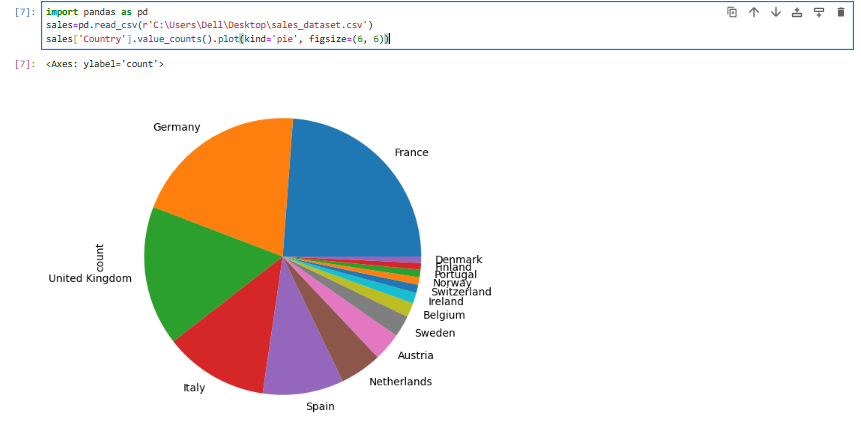
4. Positively skewed data indicates most sales occur at lower price points.

5. Small transactions dominate in frequency, but high-value sales contribute significantly to revenue.

6. Potential to introduce mid-tier products or adjust pricing strategies to capture more sales.

1. **Pie-chart**

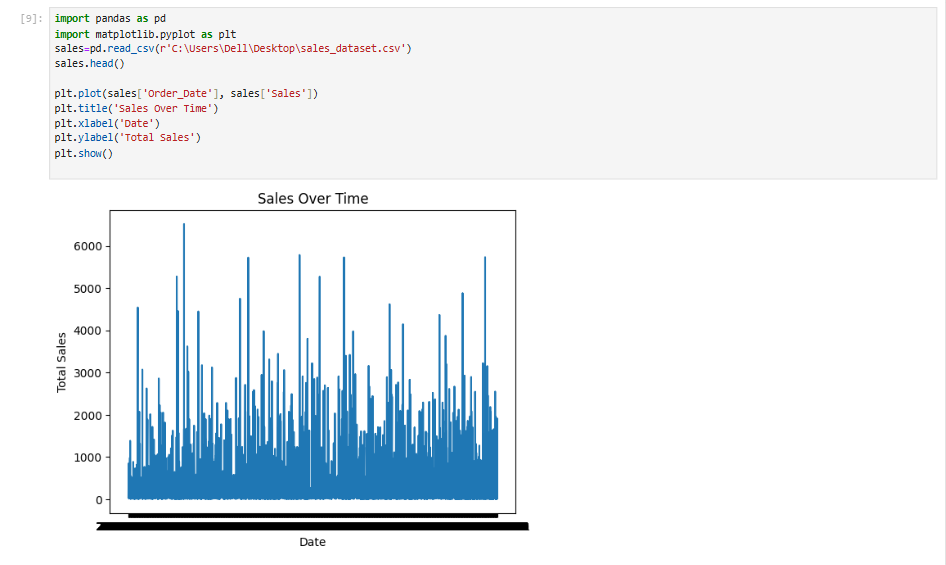
Pie-Chart is used to show the proportion of categories in a dataset. It helps visualize the distribution of a single variable as slices of a circle.



**Result:** The chart above visualizes the distribution of sales across different countries. It uses the `value\_counts()` function to count occurrences of each unique country in the `Country` column of the dataset. The `plot(kind='pie')` function then creates a pie chart, with the percentage share of each country labeled.

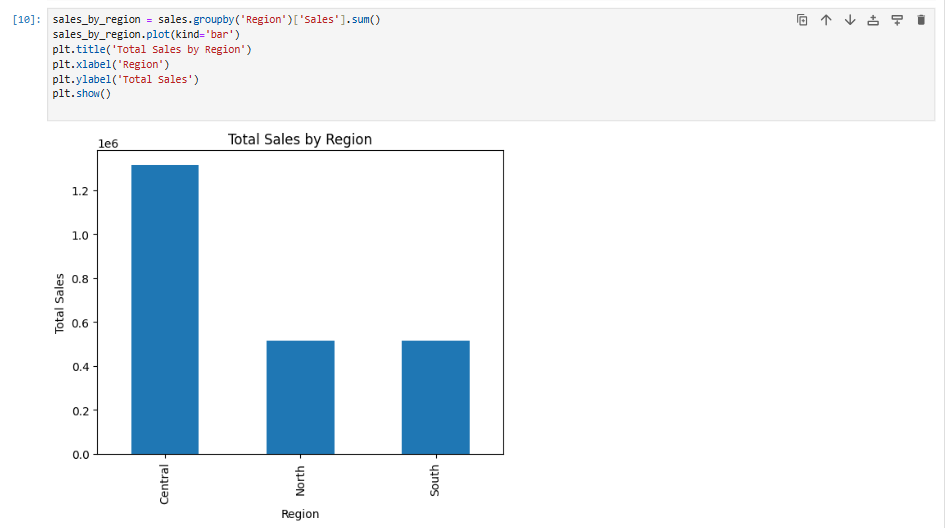
1. **Line Plot**

A **line plot** (or **time series plot**) is a type of data visualization that displays information as a series of points connected by a continuous line. It is commonly used to track changes over time.

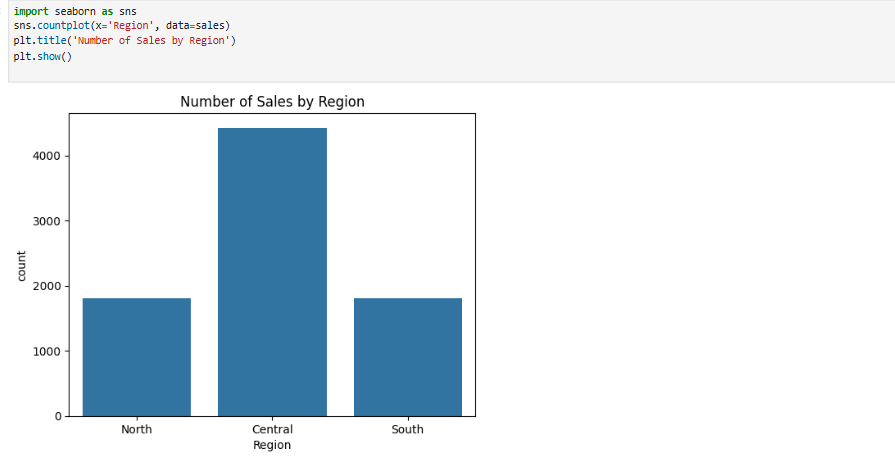
****

**Result:** This line chart illustrates the trend of sales over time. The x-axis represents the order dates, and the y-axis shows the corresponding sales amounts. The chart provides insights into sales patterns and variations over the specified period**.**

1. **Bar-chart**

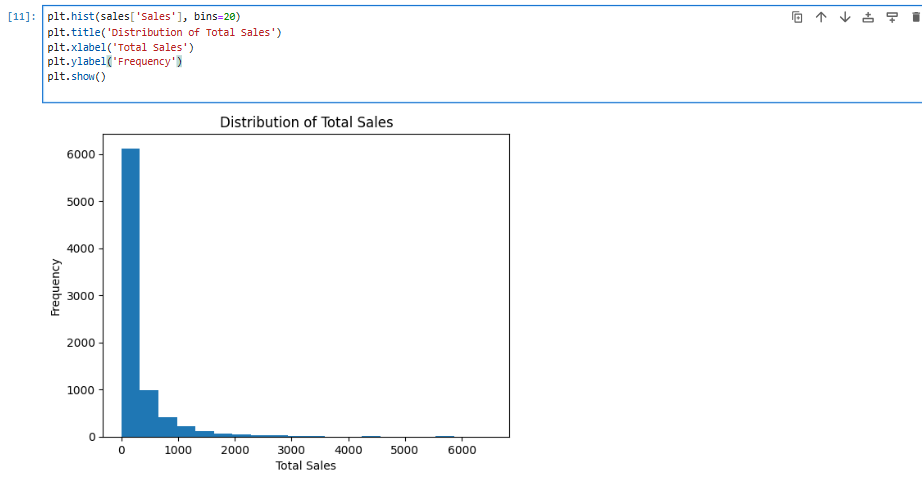
****

**Result:** This bar chart shows the total sales for each region. The data is grouped by the `Region` column, and the sales values are summed for each region. The x-axis represents the regions, and the y-axis shows the total sales.



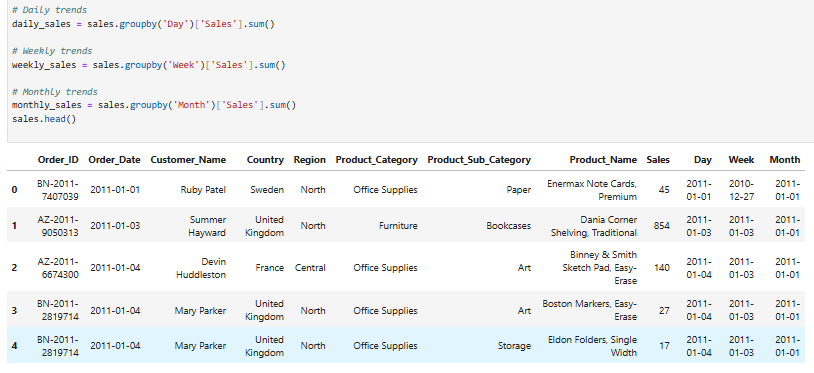
Result: This count plot shows the number of sales occurrences in each region. The x-axis represents the regions, and the y-axis indicates the count of sales. It provides insights into the frequency of sales across different regions.

1. **Histogram**

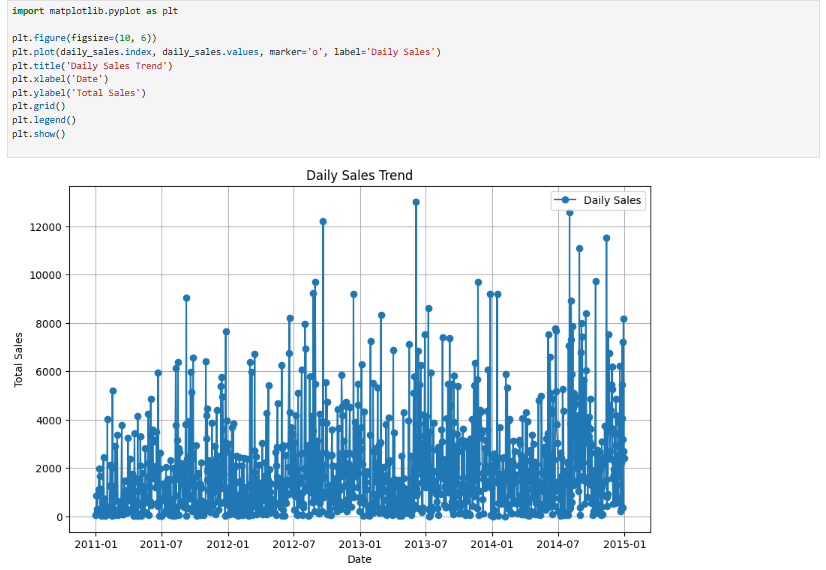
****

**Result:** This histogram visualizes the distribution of total sales values. The data is divided into 20 bins, representing ranges of sales amounts. The x-axis shows the total sales ranges, while the y-axis indicates the frequency of sales within each range.

* 1. **Analyzing Sales Trend**
* **Extract daily, weekly, or monthly periods from the Order\_Date column.**



* **Daily Sales Trends**



**Result:** This line chart visualizes the trend of total daily sales over time. Each point on the chart represents the total sales for a specific day.

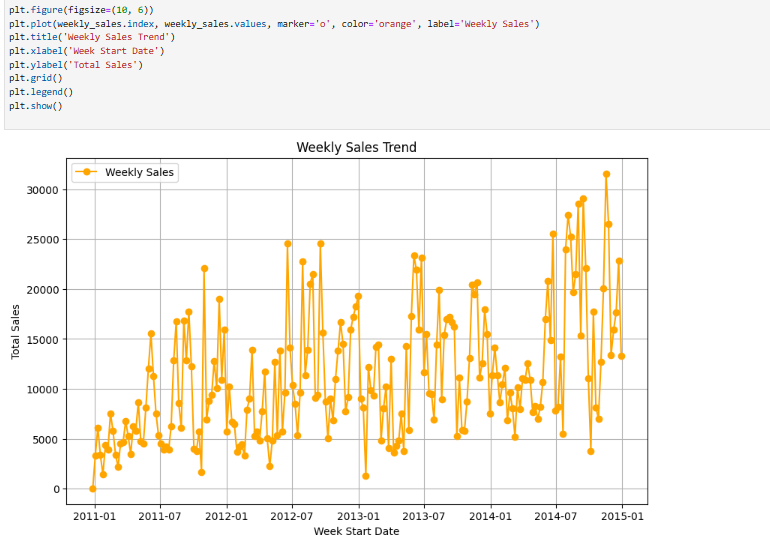
- The x-axis displays the dates.

- The y-axis shows the total sales for each day.

- The chart includes markers for each data point, a grid for better readability, and a legend for clarity.

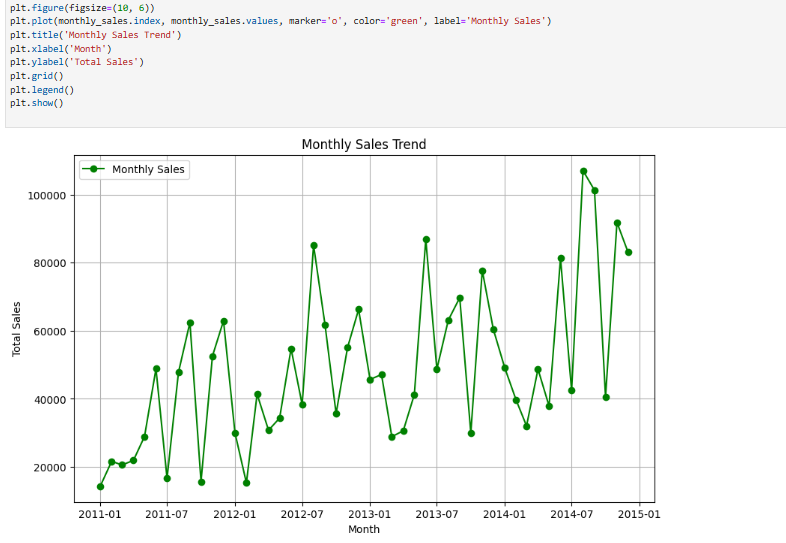
* **Weekly sales trends**

****



**Result:** This line chart illustrates the trend of total weekly sales over time. Each point represents the total sales for a specific week.

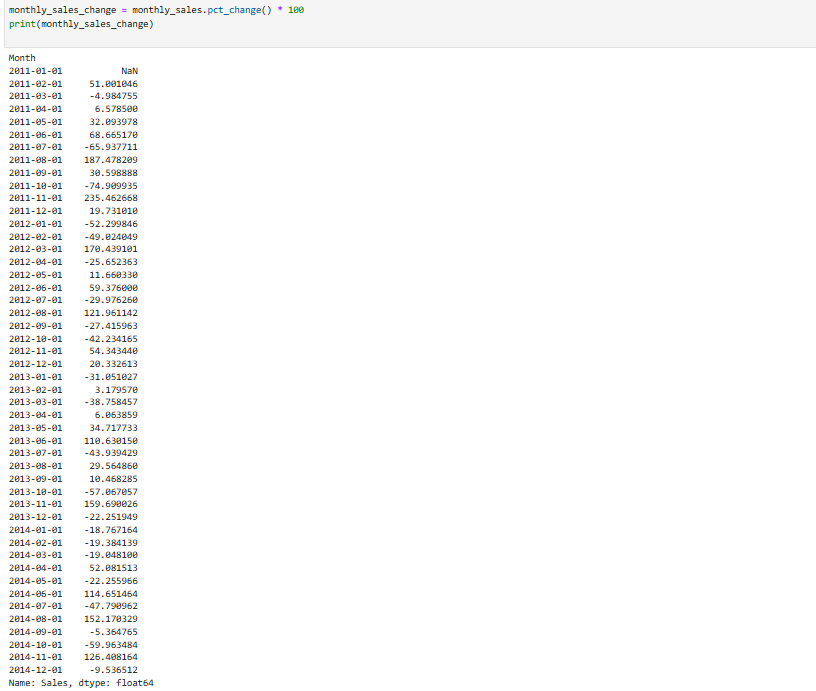
* The x-axis shows the start date of each week.
* The y-axis displays the total sales for the corresponding week.
* The chart uses an orange line with markers for better visualization, includes a grid for readability, and a legend for clarity.
* **Monthly Sales trends**



**Result:** This line chart visualizes the trend of total monthly sales over time. Each point represents the total sales for a specific month.

* The x-axis represents the months.
* The y-axis displays the total sales for each month.
* The chart uses a green line with markers for better visualization, includes a grid for clarity, and provides a legend for reference.

**Changing monthly sales to percentage**



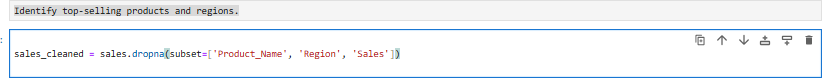
Result: This code calculates the month-over-month percentage change in total sales. The calculation helps identify the growth or decline rate in sales between consecutive months.

* **pct\_change():** Computes the percentage change between the current and previous elements in the **monthly\_sales** series.
* **\* 100**: Converts the ratio to a percentage.

The result is printed as a series where:

* The index represents the months.
* The values represent the percentage change in sales compared to the previous month.

**Identify top-selling products and regions.**



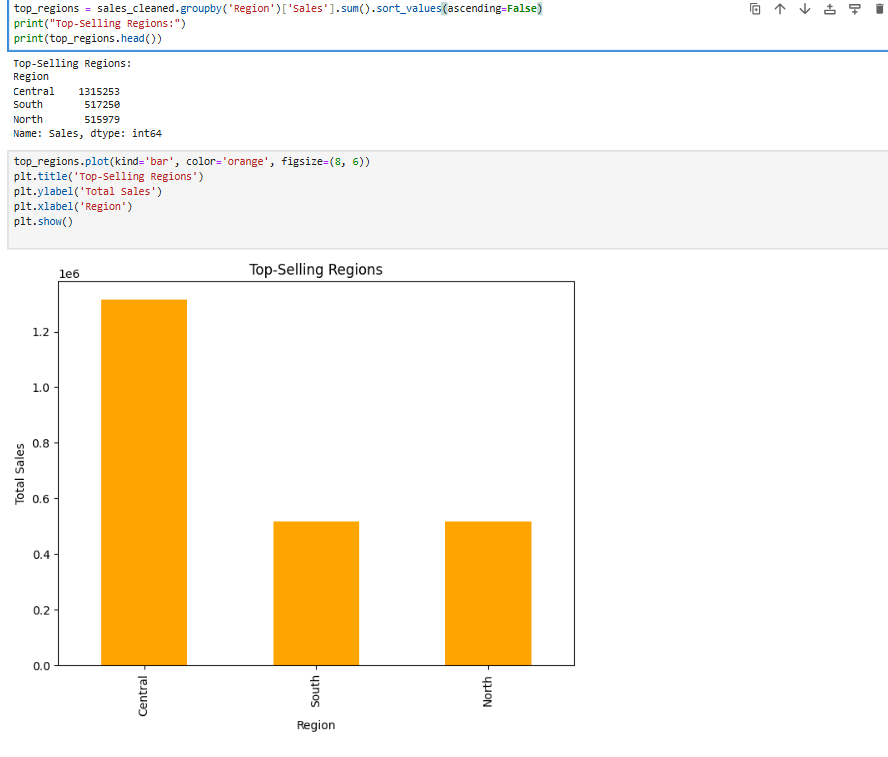
* **Display Top-selling products (Top-5)**



**Result:** This bar chart highlights the top 5 products with the highest total sales.

* The x-axis represents the product names.
* The y-axis displays the total sales for each product.
* The bars are displayed in a sky-blue color, with a clear title and labeled axes.

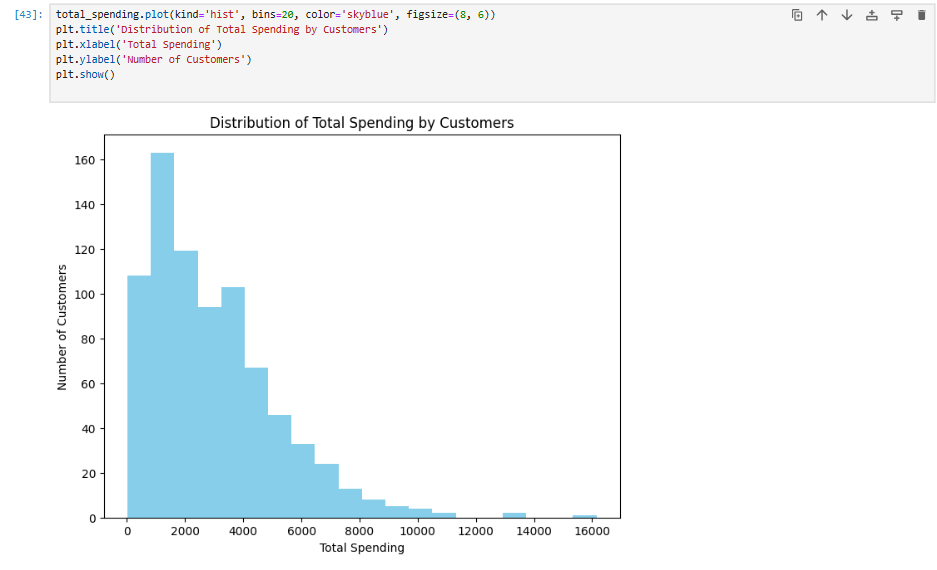
**Display Top-selling regions**



**Result:** This bar chart shows the regions with the highest total sales.

* The x-axis represents the regions.
* The y-axis displays the total sales for each region.
* The bars are displayed in an orange color, with a clear title and labeled axes for better understanding.

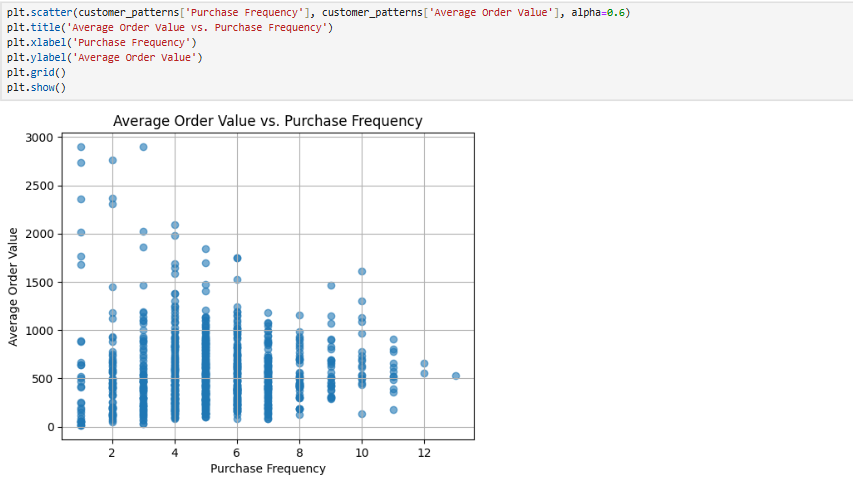
**Total Spending distribution**

****

**Result:** This histogram visualizes the distribution of total spending among customers.

* The x-axis represents the total spending ranges.
* The y-axis shows the number of customers within each spending range.
* The data is divided into 20 bins for better granularity, and the bars are displayed in a sky-blue color.

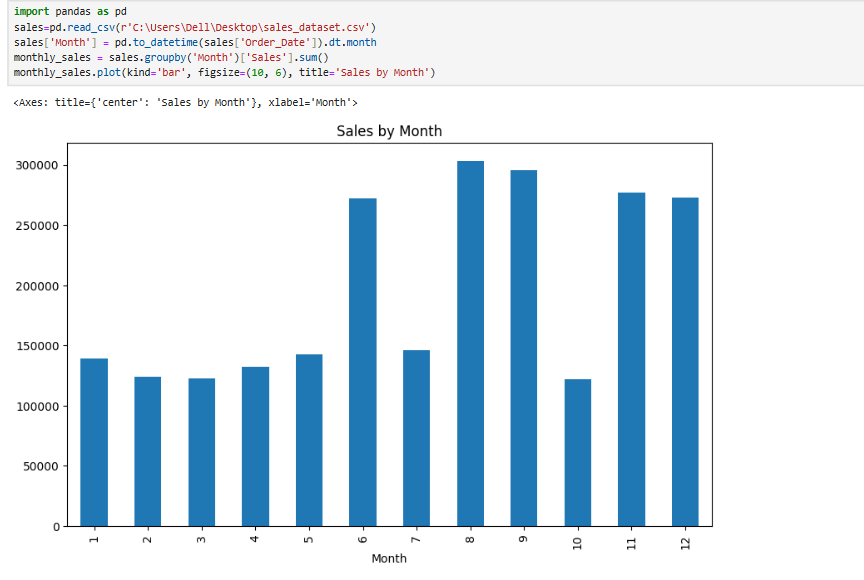
**Average Order Value vs. Purchase Frequency**

****

**Result:** This scatter plot examines the relationship between purchase frequency and average order value for customers.

* The x-axis represents the **purchase frequency** of customers.
* The y-axis shows the **average order value**.
* Each point represents a customer, with transparency (`alpha=0.6`) applied for overlapping points.
* A grid is included for better readability of the plot.

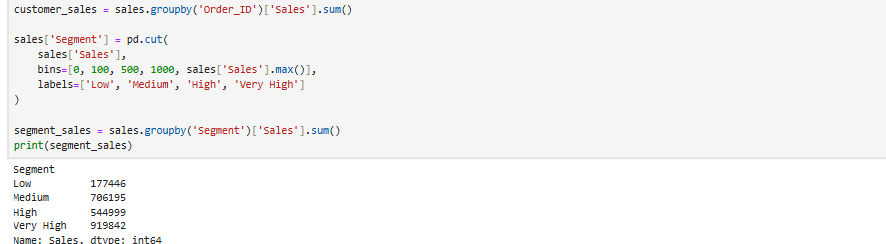
**What time of the year sales are highest?**



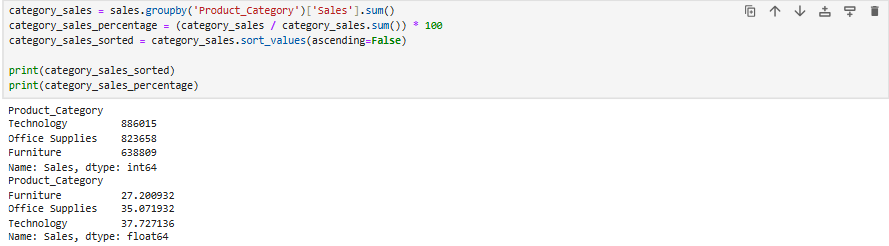
**Result:** This bar chart represents the sales by month to determine the time which has highest sales of the year.

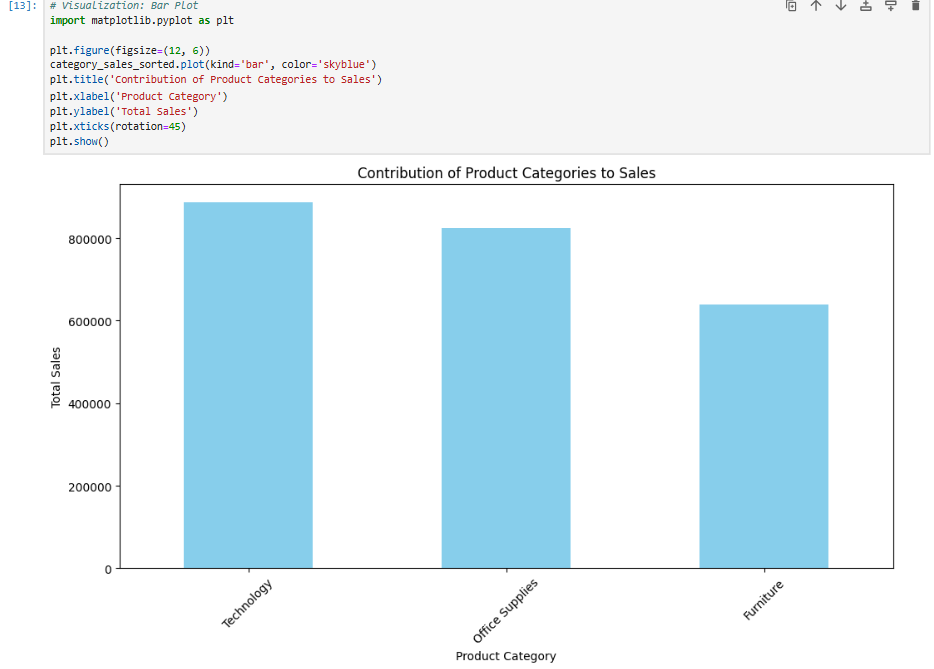
* The x-axis represents the months.
* The y-axis shows the sales value.

**Which customer segment drives the most revenue?**

****

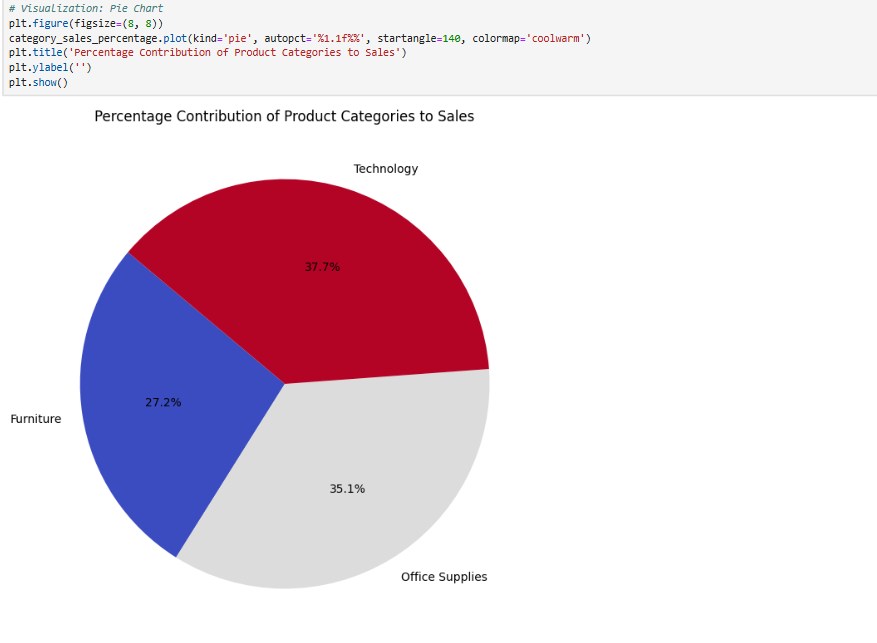
**How do product categories contribute to sales?**

****

****

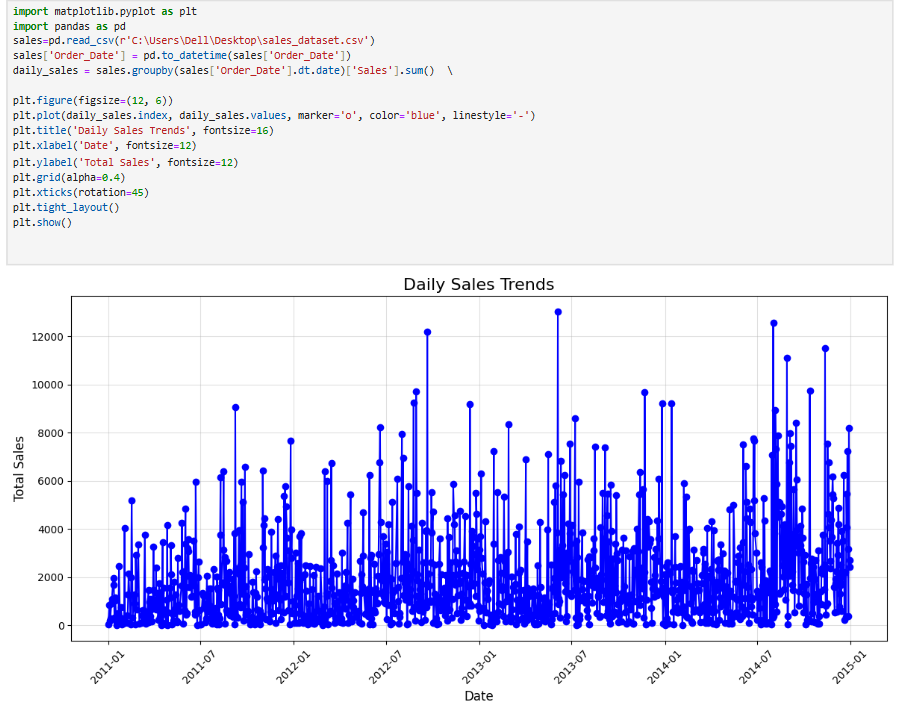
**Result:** The Bar chart above represents the contribution of Product Categories to the sales.

* The x-axis represents the Product Category.
* The y-axis represents the Total Sales.



**Result**: The chart above visualizes the contribution of Product Categories to Sales. It uses the **category\_sales\_percentage** function to convert the sales value to percentage. The **plot(kind='pie')** function then creates a pie chart, with the percentage share of each product categories.

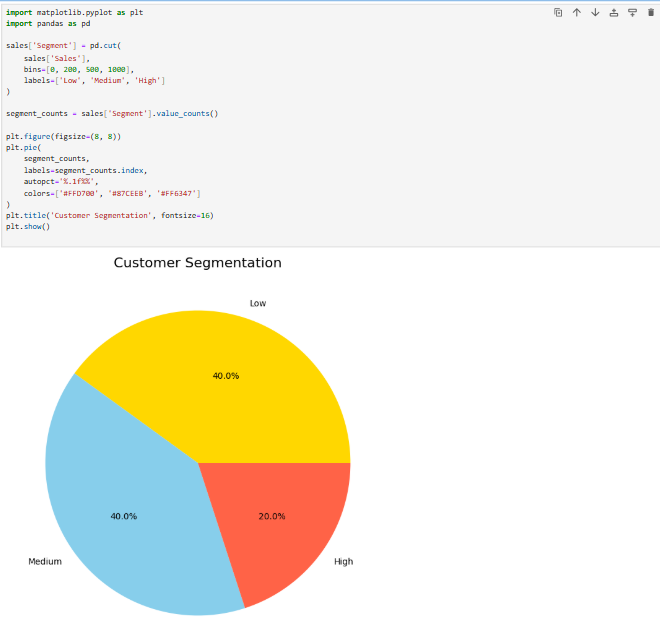
**Visualizations of Line charts for sales trends**



This line chart visualizes the daily sales trends over time, showing the total sales for each date.

* **Data Preparation**
* The **Order\_Date** column is converted to a datetime format using **pd.to\_datetime()**.
* Daily sales are calculated by grouping data by the order date and summing the sales values.
* **Chart Details**:
* **X-axis:** Displays the dates (rotated for readability).
* **Y-axis**: Represents the total sales for each day.
* A blue line with markers and gridlines highlights the trend.

**Pie-chart for customer segmentation**

****

**Result:** This pie chart visualizes the segmentation of customers based on their total sales amounts. Customers are categorized into three segments:

* **Low**: Sales below $200
* **Medium**: Sales between $200 and $500
* **High**: Sales above $500
* **pd.cut():** Categorizes sales values into the defined segments.
* **value\_counts()**: Counts the number of customers in each segment.

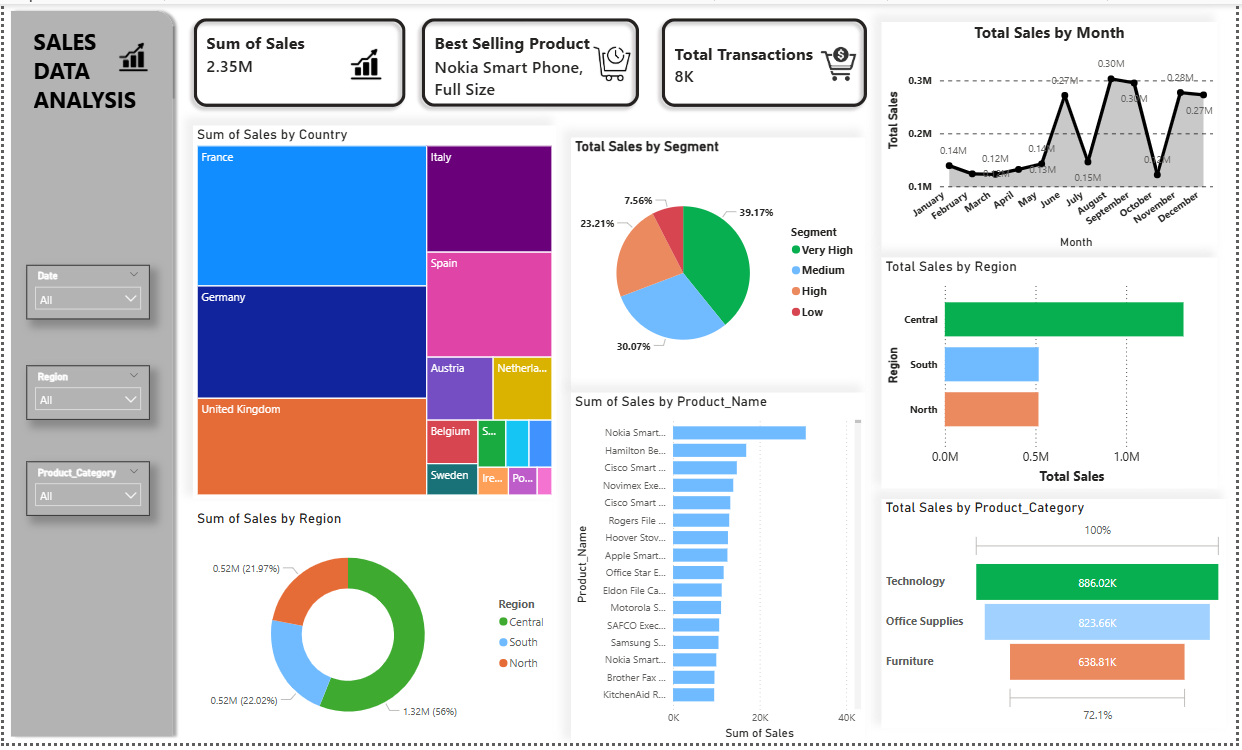
Note: The chart uses three distinct colors to represent the segments, with percentages displayed for each.

# **Conclusion**

Thus, **Power BI dashboard** and Python visualization provides valuable insights into **sales performance, customer behavior, and product trends**. The **interactive visualizations and KPIs** allow stakeholders to make informed business decisions.

**Appendices**

* 1. **PowerBI Dashboard**

****