zSIMPLE SALES DATA VISUALISATION

|  |  |  |
| --- | --- | --- |
| **Name** | **Roll No.** | **Branch** |
| SURYANK BATHAM | 202401100300257  (2428CSEAI2055) | CSE(AI) |

# KIET GROUP OF INSTITUTIONS, GHAZIABAD, UTTAR PRADESH

INTRODUCTION

In today's competitive business environment, data-driven decision-making is crucial for optimizing sales strategies. Analyzing sales data helps businesses identify trends, monitor performance, and make informed decisions to boost revenue. This report focuses on **sales data visualization**, where we use Python to process and analyze sales records from a CSV file.

The primary objective is to **extract meaningful insights from sales data** by aggregating sales per product and presenting them visually through bar charts. By leveraging **Pandas for data manipulation and Matplotlib for visualization**, this analysis provides a clear understanding of which products perform well and which may need further attention.

This study will not only help in identifying best-selling products but also lay the foundation for deeper insights such as **seasonal trends, sales forecasting, and customer preferences**, which can be explored in future analyses.

UNDERSTANDING THE PROBLEM

Sales data plays a vital role in business growth, helping organizations track revenue, identify high-performing products, and optimize sales strategies. However, raw sales data stored in CSV files can be overwhelming and difficult to interpret without proper analysis and visualization.

The key challenge in this study is to **extract useful insights from raw sales data** by:

1. **Reading and processing the dataset** to understand the structure and contents.
2. **Aggregating sales figures** to identify the performance of different products.
3. **Visualizing sales trends** using charts to make data-driven conclusions.

By addressing this problem, businesses can make informed decisions, such as:

* Identifying **top-selling and low-performing products**
* Recognizing **patterns and trends** in sales
* Supporting **inventory management and marketing strategies**

This project aims to **simplify complex sales data into an easy-to-understand format**, making it more actionable for business stakeholders.

Top of Form

Bottom of Form

CHALLENGES IN THE PROBLEM

While analyzing and visualizing sales data provides valuable insights, there are several challenges that need to be addressed:

1. **Data Quality Issues**
   * Missing or incomplete sales records can lead to inaccurate analysis.
   * Inconsistent data formats (e.g., incorrect date formats, spelling errors in product names).
2. **Handling Large Datasets**
   * If the dataset is too large, processing and visualization may become slow.
   * Requires efficient data handling techniques like filtering and aggregation.
3. **Data Cleaning & Preprocessing**
   * Identifying and removing duplicate records.
   * Standardizing product names and categories for better grouping.
4. **Choosing the Right Visualization**
   * Selecting the correct type of chart (bar charts, line graphs, etc.) for better data representation.
   * Avoiding misleading visuals that may misinterpret trends.
5. **Dynamic & Real-Time Analysis**
   * Static CSV files do not update automatically, making real-time insights difficult.
   * Integrating with live databases or APIs would improve analysis.
6. **Interpreting Results Accurately**
   * Understanding trends correctly to make informed business decisions.
   * Avoiding false correlations that may lead to incorrect conclusions.

METHADOLOGY

To analyze and visualize the sales data, we used a combination of **data processing, aggregation, and visualization techniques**. The following methods were applied:

**1. Data Handling Methods**

* **File Uploading:** The files.upload() method was used to upload the CSV file to Google Colab.
* **Data Loading:** The pd.read\_csv() function from Pandas was used to read the CSV file into a DataFrame.
* **Data Cleaning:** Checked for missing values, duplicates, and incorrect data formats.

STEPS

** Upload the CSV File – Use Google Colab to upload the sales\_data.csv file.**

** Read and Load the Data – Load the CSV file using Pandas and check its structure.**

** Clean and Preprocess the Data – Handle missing values, remove duplicates, and standardize column names.**

** Process and Analyze the Data – Group sales by product and calculate total sales.**

** Visualize the Data – Create a bar chart to display product-wise sales.**

** Interpret Results – Identify best-selling and low-performing products based on sales trends.**

CODE

from google.colab import files

import pandas as pd

import io

# Upload the file

uploaded = files.upload()

# Print the uploaded file names

print("Uploaded files:", uploaded.keys())

# Use the correct filename dynamically

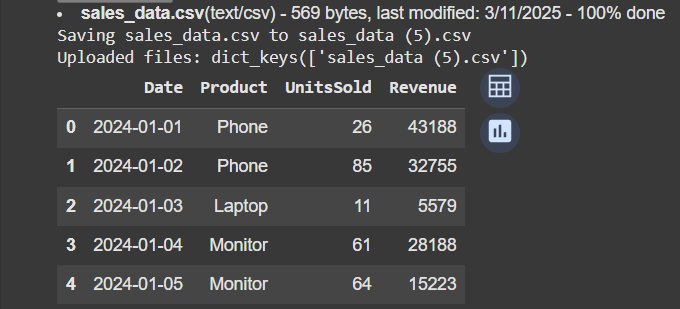
filename = list(uploaded.keys())[0]  # Get the actual file name

df = pd.read\_csv(io.BytesIO(uploaded[filename]))

# Display first few rows

df.head()

OUTPUT



CREDITS

ALGORITHM & INSPIRATION:

🡪Official Python Website: <https://www.python.org/>

TOOLS AND TECHNOLOGIES USED:

🡪 **Programming Language:** Python

🡪 **Development Environment:** GOOGLE Colab notebook