www.kiet.edu
Delhi-NCR, Ghaziabad

KIET GROUP OF INSTITUTIONS
Connecting Life with Learning

SIMPLE SALES DATA VISUALIZATION

Submitted By -

PRIYANSHI

202401100300185

CSEAI(Sec-C)

Introduction-

This dataset represents **sales transactions** over a period of time. It includes details such as:

- **Date** – The transaction date.
- **Product** – The name/type of the product sold.
- **Units Sold** – The number of units sold per transaction.
- **Revenue** – The total revenue generated from the sale.

Purpose of Analysis

- **Identify sales trends over time** (e.g., peak sales periods).
- **Analyze product performance** (which products generate the most revenue).
- **Help in decision-making** (pricing, stock management, marketing strategies).

Methodology-

To analyze the sales data, we follow a structured approach:

1. Data Collection

- The dataset (sales_data.csv) is imported into Python using **pandas**.
- It contains key fields like **Date, Product, UnitsSold, and Revenue**.

2. Data Preprocessing

- **Handling Missing Values**: Checking for and handling missing or incorrect data.
- **Data Type Conversion**: Converting the "Date" column into a proper datetime format.
- **Column Name Validation**: Ensuring correct column names for analysis.

3. Exploratory Data Analysis (EDA)

- **Basic Statistics**: Mean, median, max, min of sales and revenue.
- **Trend Analysis**: Identifying sales growth patterns over time.
- **Product Performance**: Checking which products contribute the most revenue.

4. Data Visualization

- **Sales Trends**: Line charts showing revenue trends over time.
- **Product Sales Distribution**: Bar charts for best-selling products.
- **Seasonality Analysis**: Identifying high-sales periods.

5. Insights & Decision Making

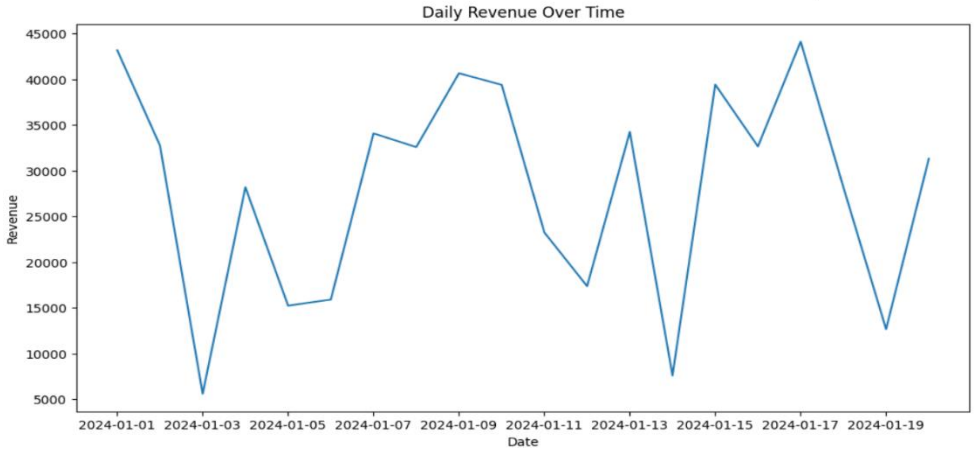
- Finding **best-performing products**.
- Identifying **peak sales periods** to optimize inventory and marketing.
- Understanding **revenue fluctuations** and trends.

Code-

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
# Load the dataset
data = pd.read_csv('sales_data.csv')

# Display the first few rows of the dataset
print(data.head())

# Convert the 'Date' column to datetime format
data['Date'] = pd.to_datetime(data['Date'])

# Plot the total revenue per product
plt.figure(figsize=(10, 6))
sns.barplot(x='Product', y='Revenue', data=data, estimator=sum)
plt.title('Total Revenue per Product')
plt.xlabel('Product')
plt.ylabel('Total Revenue')
plt.show()

# Plot the total units sold per product
plt.figure(figsize=(10, 6))
sns.barplot(x='Product', y='UnitsSold', data=data, estimator=sum)
plt.title('Total Units Sold per Product')
```

```
plt.xlabel('Product')
plt.ylabel('Total Units Sold')
plt.show()

# Plot the daily revenue over time
plt.figure(figsize=(12, 6))
sns.lineplot(x='Date', y='Revenue', data=data)
plt.title('Daily Revenue Over Time')
plt.xlabel('Date')
plt.ylabel('Revenue')
plt.show()
```

Output-



Credits-

- Image 1 taken form-

https://www.scottsdirectories.com/wp-content/uploads/2018/01/Sales-Data-You-Dont-Know-Youre-Missing-for-Your-Local-Toronto-Business.png

Image 2 (Kiet Logo) taken from-

https://kietalumni.com/joinkaa.php

- Image 3 Screenshot of code taken from-

https://colab.research.google.com/drive/1ZGRtTpvY7iU138_v8AvypzA82CfFXTeX