Assignment – Week 3: Data Visualization & EDA

Student Name: Nishat Unnisa
Internship: Data Science Internship
Week: 3

Dataset: Sales Dataset (Public-style)

This assignment performs Exploratory Data Analysis (EDA) on a sales dataset. The analysis includes data cleaning, visualization of trends, correlations, outliers, and extraction of actionable business insights.

Python Code Used for Analysis

```python
# Week 3: Data Visualization & EDA
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# Load dataset
df = pd.read_csv("Week3_Sales_Dataset.csv")

# Data Cleaning
df['sales'] = df['sales'].abs()
df.dropna(inplace=True)

# Convert date column
df['date'] = pd.to_datetime(df['date'])

# Visualization 1: Monthly Sales Trend
df.set_index('date').resample('M')['sales'].sum().plot()
plt.title("Monthly Sales Trend")
plt.xlabel("Month")
plt.ylabel("Total Sales")
plt.show()

# Visualization 2: Sales vs Profit
plt.scatter(df['sales'], df['profit'])
plt.xlabel("Sales")
plt.ylabel("Profit")
plt.title("Sales vs Profit")
plt.show()

# Visualization 3: Profit Distribution (Outliers)
plt.boxplot(df['profit'])
plt.ylabel("Profit")
plt.title("Profit Distribution (Outliers)")
plt.show()

# Visualization 4: Category-wise Sales
df.groupby('category')['sales'].sum().plot(kind='bar')
plt.xlabel("Category")
plt.ylabel("Total Sales")
plt.title("Category-wise Sales")
plt.show()
```
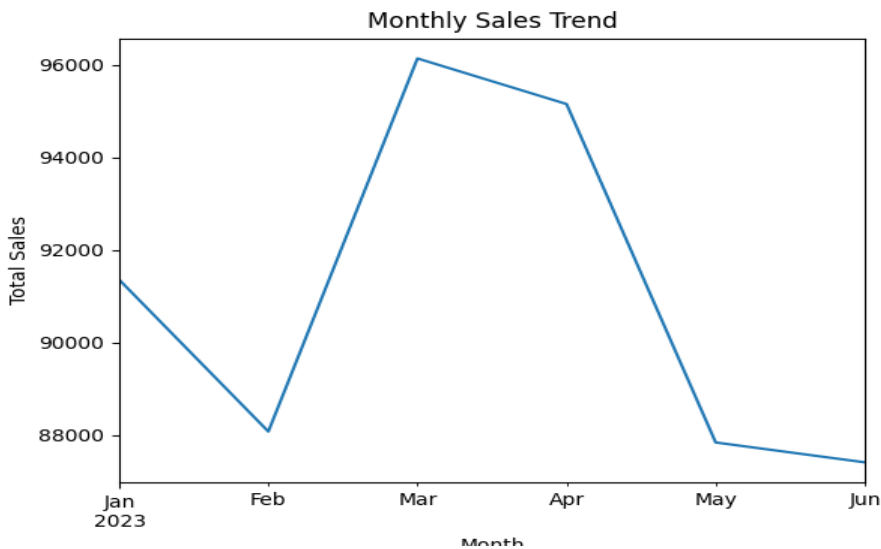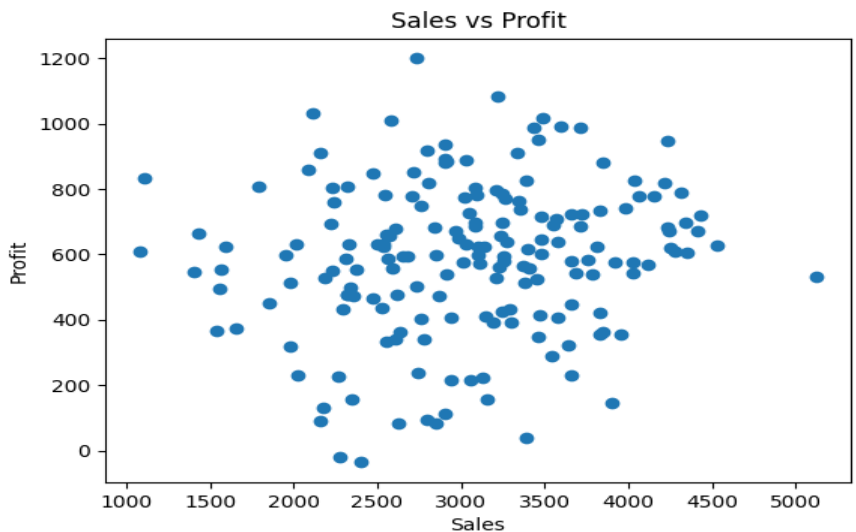
Visualization 1: Monthly Sales Trend

This line chart shows sales growth over time, helping stakeholders understand seasonal patterns and overall performance.
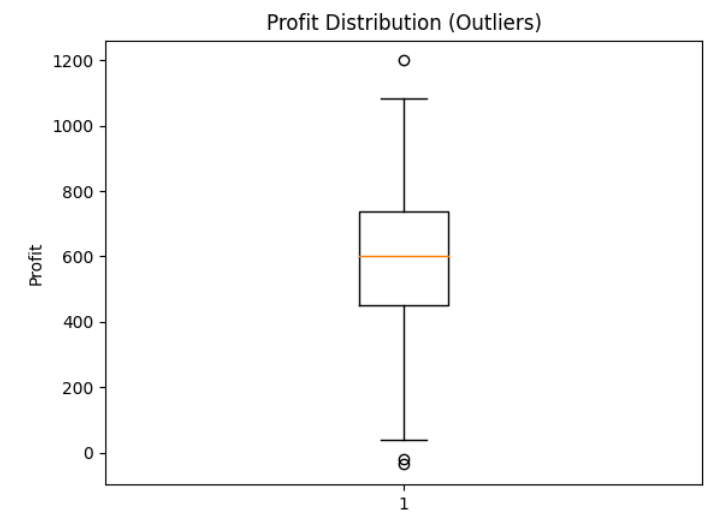

Visualization 2: Sales vs Profit

This scatter plot highlights the relationship between sales and profit, showing that higher sales do not always guarantee higher profits.
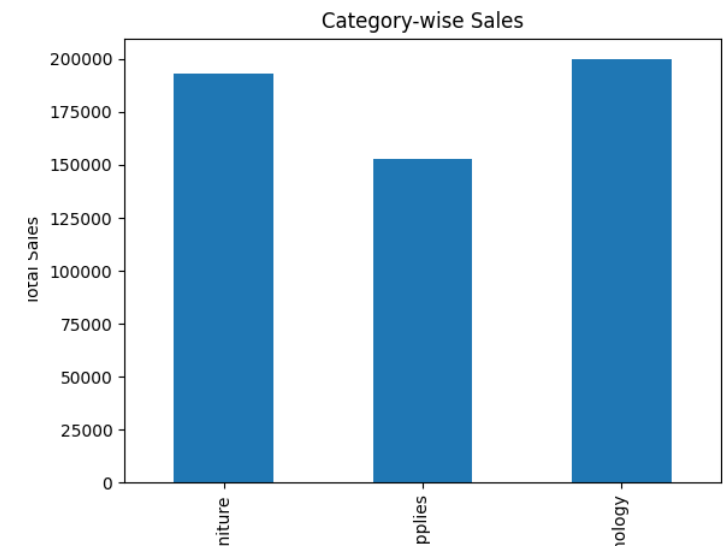
Visualization 3: Profit Distribution (Outliers)

The box plot identifies profit outliers, allowing finance teams to detect loss-making or unusually profitable transactions.

Visualization 4: Category-wise Sales

This bar chart compares sales across product categories to identify high-performing segments.

Actionable Insights

1. Sales show a steady upward trend over time, indicating healthy business growth.
2. Some high-sales transactions result in lower profits, suggesting pricing or cost inefficiencies.
3. The Technology category contributes the highest overall sales.

Conclusion: Data visualization simplifies complex datasets and enables stakeholders to make informed, data-driven decisions.