**Assignment: E-commerce Sales Data Analysis for Business Insights**

**Objective:**  
Analyze e-commerce sales data to derive actionable insights for improving business decisions. Perform various analyses, including customer, product, time-based, and geographical analyses, and provide recommendations supported by data analysis.

**Part 1: Dataset and Tools**

1. Use an open-source or simulated e-commerce dataset (e.g., from Kaggle or a dataset generated by the instructor).
2. Tools: Python with libraries like pandas, numpy, matplotlib, seaborn, and scikit-learn.

**Part 2: Experiment/Assignment Breakdown**

**Experiment 1: Data Loading and Preprocessing**

**Objective:** Load the dataset, clean missing/erroneous values, and prepare it for analysis.  
**Code:**

import pandas as pd

import numpy as np

# Load the dataset

data = pd.read\_csv("ecommerce\_sales.csv") # Replace with the actual file name

# Display dataset information

print("Dataset Info:")

print(data.info())

# Handle missing values

data.fillna(0, inplace=True)

# Convert date column to datetime format

data['Amount'] = pd.to\_datetime(data['Amount'])

# Display the first few rows of the cleaned dataset

print("\nCleaned Data:")

print(data.head())

**Experiment 2: Customer Analysis**

**Objective:** Analyze customer behavior (e.g., purchase frequency, total spend, and segmentation).  
**Code:**

python

Copy code

# Total spend by each customer

customer\_analysis = data.groupby('Customer ID')['Sales'].sum().sort\_values(ascending=False)

print("Top 10 Customers by Total Spend:")

print(customer\_analysis.head(10))

# Purchase frequency

customer\_frequency = data['Customer ID'].value\_counts()

print("\nTop 10 Customers by Purchase Frequency:")

print(customer\_frequency.head(10))

**Experiment 3: Product Analysis**

**Objective:** Identify top-performing products and products with low sales.  
**Code:**

python

Copy code

# Total sales by product

product\_sales = data.groupby('Product Name')['Sales'].sum().sort\_values(ascending=False)

print("Top 10 Products by Sales:")

print(product\_sales.head(10))

# Visualizing product performance

import matplotlib.pyplot as plt

top\_products = product\_sales.head(10)

top\_products.plot(kind='bar', color='skyblue', figsize=(10, 6))

plt.title("Top 10 Products by Sales")

plt.xlabel("Product Name")

plt.ylabel("Total Sales")

plt.xticks(rotation=45)

plt.show()

**Experiment 4: Time-Based Analysis**

**Objective:** Analyze sales trends over time (e.g., monthly or daily sales trends).  
**Code:**

python

Copy code

# Monthly sales trend

data['Month'] = data['Order Date'].dt.to\_period('M')

monthly\_sales = data.groupby('Month')['Sales'].sum()

# Plotting monthly sales

monthly\_sales.plot(kind='line', figsize=(10, 6), marker='o')

plt.title("Monthly Sales Trend")

plt.xlabel("Month")

plt.ylabel("Total Sales")

plt.grid()

plt.show()

**Experiment 5: Geographical Analysis**

**Objective:** Analyze sales distribution across regions or countries.  
**Code:**

python

Copy code

# Total sales by region

region\_sales = data.groupby('Region')['Sales'].sum().sort\_values(ascending=False)

print("Sales by Region:")

print(region\_sales)

# Visualizing regional sales

region\_sales.plot(kind='pie', autopct='%1.1f%%', figsize=(8, 8), startangle=90, colors=['skyblue', 'orange', 'lightgreen'])

plt.title("Sales Distribution by Region")

plt.ylabel("")

plt.show()

**Experiment 6: Deriving Insights and Recommendations**

**Objective:** Summarize insights and propose actionable recommendations.

**Code:**

python

Copy code

# Insights

print("Insights:")

print("- Top customers contribute significantly to total revenue. Target them with loyalty programs.")

print("- High-performing products dominate sales. Consider increasing inventory for these products.")

print("- Seasonal trends indicate peak sales during certain months. Optimize marketing during these periods.")

print("- Regions with lower sales could benefit from targeted campaigns.")

**Deliverables**

1. Submit a Jupyter Notebook containing:
   * Code for all experiments.
   * Visualizations with proper titles and labels.
2. A PDF report summarizing:
   * Insights from each analysis.
   * Recommendations to improve marketing strategies, inventory management, and regional targeting.

**Expected Outcomes**

1. **Customer Analysis:** Identify top customers and their purchasing patterns.
2. **Product Analysis:** Recognize best-selling and underperforming products.
3. **Time-Based Analysis:** Determine sales trends to identify peak periods.
4. **Geographical Analysis:** Understand the sales distribution to enhance regional strategies.
5. **Actionable Insights:** Use the analyses to recommend data-driven business improvements.

This assignment ensures practical exposure to real-world business data analysis and decision-making.