**2311cs020140**

**Assignment 1: Sales and Effectiveness Analysis**

**Objective:**  
Explore the relationship between marketing spend, sales, and drug effectiveness across different regions and age groups. Create visualizations using matplotlib and seaborn.

**Instructions:**

1. Load the dataset.
2. Perform data cleaning (check for missing values, duplicates).
3. Create the following visualizations:
   * A **bar plot** showing total sales per region.
   * A **scatter plot** to visualize the relationship between Marketing\_Spend and Sales.
   * A **boxplot** comparing drug effectiveness across different age groups.
   * A **line plot** showing the sales trend for each product over different trial periods.
   * A **heatmap** of the correlation between Sales, Marketing\_Spend, and Effectiveness.
4. Based on the visualizations, summarize any patterns or trends you observe.

**Expected Outcome:**

* Insights on how marketing spend impacts sales.
* Analysis of which age groups have higher drug effectiveness.
* Regional sales distribution.

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# Load the dataset (adjust the path to your dataset)

# Check for missing values

print("Missing values in the dataset:")

print(df.isnull().sum())

# Check for duplicates

print("\nNumber of duplicate rows in the dataset:")

print(df.duplicated().sum())

# Drop rows with missing values or handle them as appropriate

df = df.dropna() # Example: dropping rows with missing values (alternative: df.fillna())

# Drop duplicates if any

df = df.drop\_duplicates()

# Task 1: Bar plot showing total sales per region

plt.figure(figsize=(8, 6))

sales\_per\_region = df.groupby('Region')['Sales'].sum()

sales\_per\_region.plot(kind='bar', color='skyblue')

plt.title('Total Sales per Region')

plt.xlabel('Region')

plt.ylabel('Total Sales')

plt.xticks(rotation=45)

plt.tight\_layout()

plt.show()

# Task 2: Scatter plot to visualize the relationship between Marketing\_Spend and Sales

plt.figure(figsize=(8, 6))

sns.scatterplot(x='Marketing\_Spend', y='Sales', data=df, color='green')

plt.title('Relationship between Marketing Spend and Sales')

plt.xlabel('Marketing Spend')

plt.ylabel('Sales')

plt.tight\_layout()

plt.show()

# Task 3: Boxplot comparing drug effectiveness across different age groups

plt.figure(figsize=(8, 6))

sns.boxplot(x='Age\_Group', y='Effectiveness', data=df, palette='Set2')

plt.title('Drug Effectiveness across Different Age Groups')

plt.xlabel('Age Group')

plt.ylabel('Effectiveness')

plt.tight\_layout()

plt.show()

# Task 4: Line plot showing the sales trend for each product over different trial periods

plt.figure(figsize=(10, 6))

sns.lineplot(x='Trial\_Period', y='Sales', hue='Product', data=df, marker='o')

plt.title('Sales Trend for Each Product over Trial Periods')

plt.xlabel('Trial Period')

plt.ylabel('Sales')

plt.legend(title='Product')

plt.tight\_layout()

plt.show()

# Task 5: Heatmap of the correlation between Sales, Marketing\_Spend, and Effectiveness

correlation\_matrix = df[['Sales', 'Marketing\_Spend', 'Effectiveness']].corr()

plt.figure(figsize=(8, 6))

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm', fmt='.2f', vmin=-1, vmax=1)

plt.title('Correlation between Sales, Marketing Spend, and Effectiveness')

plt.tight\_layout()

plt.show()