**2311cs020140**

**Assignment 2: Drug Effectiveness and Side Effects Comparison**

**Objective:**  
Compare drug effectiveness and side effects by different products, regions, and trial periods. Visualize these comparisons using matplotlib and seaborn.

**Instructions:**

1. Load the dataset.
2. Perform any necessary data cleaning.
3. Create the following visualizations:
   * A **bar plot** comparing the average Effectiveness for each drug across different regions.
   * A **violin plot** to show the distribution of Effectiveness and Side\_Effects for each product.
   * A **pairplot** to explore relationships between Effectiveness, Side\_Effects, and Marketing\_Spend.
   * A **boxplot** comparing Effectiveness for different trial periods.
   * A **regression plot** to analyze how Marketing\_Spend affects drug Effectiveness.
4. Based on the visualizations, provide an analysis of:
   * Which product has the best overall effectiveness.
   * The correlation between effectiveness and side effects.

**Expected Outcome:**

* Insights into how the effectiveness of each drug varies by region and trial period.
* Understanding the trade-off between drug effectiveness and side effects.

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

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

df = pd.read\_csv('drug\_effectiveness\_data.csv')

# 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())

# Handle missing values (e.g., dropping or filling)

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

# Drop duplicates if any

df = df.drop\_duplicates()

# Task 1: Bar plot comparing the average Effectiveness for each drug across different regions

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

avg\_effectiveness\_per\_region = df.groupby(['Region', 'Product'])['Effectiveness'].mean().unstack()

avg\_effectiveness\_per\_region.plot(kind='bar', figsize=(10, 6))

plt.title('Average Effectiveness for Each Drug across Different Regions')

plt.xlabel('Region')

plt.ylabel('Average Effectiveness')

plt.xticks(rotation=45)

plt.tight\_layout()

plt.show()

# Task 2: Violin plot to show the distribution of Effectiveness and Side\_Effects for each product

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

sns.violinplot(x='Product', y='Effectiveness', data=df, inner='quart', palette='muted')

plt.title('Distribution of Effectiveness for Each Product')

plt.xlabel('Product')

plt.ylabel('Effectiveness')

plt.tight\_layout()

plt.show()

# Violin plot for Side\_Effects

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

sns.violinplot(x='Product', y='Side\_Effects', data=df, inner='quart', palette='muted')

plt.title('Distribution of Side Effects for Each Product')

plt.xlabel('Product')

plt.ylabel('Side Effects')

plt.tight\_layout()

plt.show()

# Task 3: Pairplot to explore relationships between Effectiveness, Side\_Effects, and Marketing\_Spend

sns.pairplot(df[['Effectiveness', 'Side\_Effects', 'Marketing\_Spend']], hue='Product', diag\_kind='kde', palette='Set2')

plt.suptitle('Pairplot of Effectiveness, Side Effects, and Marketing Spend', y=1.02)

plt.tight\_layout()

plt.show()

# Task 4: Boxplot comparing Effectiveness for different trial periods

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

sns.boxplot(x='Trial\_Period', y='Effectiveness', data=df, palette='Set3')

plt.title('Effectiveness Comparison across Different Trial Periods')

plt.xlabel('Trial Period')

plt.ylabel('Effectiveness')

plt.tight\_layout()

plt.show()

# Task 5: Regression plot to analyze how Marketing\_Spend affects drug Effectiveness

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

sns.regplot(x='Marketing\_Spend', y='Effectiveness', data=df, scatter\_kws={'s': 20}, line\_kws={'color': 'red'})

plt.title('Relationship between Marketing Spend and Drug Effectiveness')

plt.xlabel('Marketing Spend')

plt.ylabel('Effectiveness')

plt.tight\_layout()

plt.show()