import kagglehub

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

from scipy import stats

# Download the dataset (you have already written this part)

path = kagglehub.dataset\_download("kukuroo3/body-performance-data/versions/12")

# Assume the dataset is in a CSV format, read it into a DataFrame

# Modify the path if necessary based on the actual file location

data = pd.read\_csv(path + "/body-performance.csv")

# Display the first few rows to understand the data structure

print(data.head())

# 1.1 Measure of Central Tendency: Mean, Geometric mean, Harmonic mean, Mode, Median

# Mean

mean = data.mean()

# Geometric Mean (using scipy)

geometric\_mean = stats.gmean(data)

# Harmonic Mean (using scipy)

harmonic\_mean = stats.hmean(data)

# Mode

mode = data.mode().iloc[0]

# Median

median = data.median()

# Print measures of central tendency

print("\nCentral Tendency Measures:")

print(f"Mean:\n{mean}")

print(f"\nGeometric Mean:\n{geometric\_mean}")

print(f"\nHarmonic Mean:\n{harmonic\_mean}")

print(f"\nMode:\n{mode}")

print(f"\nMedian:\n{median}")

# 1.2 Measure of Dispersion: Variance, Standard deviation, Shape of Data (Symmetric, Skewness), IQR/percentiles, Range, Mean Absolute Deviation(MAD)

# Variance

variance = data.var()

# Standard deviation

std\_dev = data.std()

# Skewness (indicates asymmetry of data)

skewness = data.skew()

# Interquartile Range (IQR)

Q1 = data.quantile(0.25)

Q3 = data.quantile(0.75)

IQR = Q3 - Q1

# Range (difference between max and min values)

data\_range = data.max() - data.min()

# Mean Absolute Deviation (MAD)

mad = data.mad()

# Print measures of dispersion

print("\nDispersion Measures:")

print(f"Variance:\n{variance}")

print(f"\nStandard Deviation:\n{std\_dev}")

print(f"\nSkewness:\n{skewness}")

print(f"\nInterquartile Range (IQR):\n{IQR}")

print(f"\nRange:\n{data\_range}")

print(f"\nMean Absolute Deviation (MAD):\n{mad}")

# 1.3 Correlation Between Features

# Calculate correlation matrix

correlation = data.corr()

# Print the correlation matrix

print("\nCorrelation Between Features:")

print(correlation)

# 1.4 Visualizing Data Distribution: Boxplot, Histograms, Density plots, Scatterplot, Bar chart

# Set up the matplotlib figure

plt.figure(figsize=(15, 10))

# Boxplot (for visualizing outliers and distribution)

plt.subplot(2, 3, 1)

sns.boxplot(data=data)

plt.title("Boxplot of Features")

# Histogram (for distribution of individual features)

plt.subplot(2, 3, 2)

data.hist(bins=20, figsize=(10, 6))

plt.suptitle("Histograms of Features")

# Density Plot (Kernel Density Estimate)

plt.subplot(2, 3, 3)

sns.kdeplot(data=data, shade=True)

plt.title("Density Plot of Features")

# Scatterplot (for relationships between two features)

plt.subplot(2, 3, 4)

sns.scatterplot(x=data.iloc[:, 0], y=data.iloc[:, 1])

plt.title("Scatterplot Between First Two Features")

# Bar Chart (assuming categorical data exists or summarizing counts of a column)

plt.subplot(2, 3, 5)

sns.countplot(x=data.columns[0], data=data)

plt.title("Bar Chart for First Feature")

# Adjust layout for better display

plt.tight\_layout()

plt.show()