Scenario 1:

Flight Delay Analysis

An airline tracks flight delays (in minutes) for 20 flights. Analyze the flight delays to calculate percentiles, detect outliers, and evaluate the overall distribution.

ANSWER

```
import numpy as np
import pandas as pd

# Given flight delay data in minutes for 20 flights
delay_data = [5, 10, 8, 15, 20, 5, 12, 14, 10, 18, 7, 9, 21,
22, 16, 14, 17, 19, 23, 11]

# Create a DataFrame
df = pd.DataFrame({"Delay": delay_data})

# Calculate Q1 and Q3
q1 = np.percentile(df["Delay"], 25)
q3 = np.percentile(df["Delay"], 75)

# Calculate IQR
iqr = q3 - q1

# Calculate lower and upper bounds
lower_bound = q1 - 1.5 * iqr
upper_bound = q3 + 1.5 * iqr

# Detect outliers
outliers = df[(df["Delay"] < lower_bound) | (df["Delay"] >
upper_bound)]

print(f"Q1: {q1}")
print(f"Q3: {q3}")
print(f"IQR: {iqr}")
print(f"Lower Bound: {lower_bound}")
print(f"Upper Bound: {upper_bound}")
print("Outliers:")
print(outliers)

# Summary statistics
mean_delay = np.mean(df["Delay"])
median_delay = np.median(df["Delay"])
```

```
print(f"Mean Delay: {mean_delay}")
print(f"Median Delay: {median_delay}")
```

OUTPUT

```
Q1: 9.75
Q3: 18.25
IQR: 8.5
Lower Bound: -3.0
Upper Bound: 31.0
Outliers:
Empty DataFrame
Columns: [Delay]
Index: []
Mean Delay: 13.8
Median Delay: 14.0
```

Scenario 2:

Employee Salary Analysis

A company wants to analyze the salary distribution of its employees to understand the central tendency and determine whether the data is skewed.

ANSWER

```
#Approach
# Calculate the mean, median, and mode to understand central tendency.
# Calculate skewness to determine whether the data is symmetric or skewed.
# Skewness > 0 indicates right (positive) skew.
# Skewness < 0 indicates left (negative) skew.
# Skewness ≈ 0 indicates symmetric distribution.

import numpy as np
import pandas as pd
from scipy.stats import skew, mode

# Sample salary data
salaries = [45000, 48000, 47000, 52000, 49000, 51000, 90000,
47000, 48000, 50000]

mean_salary = np.mean(salaries)
median_salary = np.median(salaries)
mode_salary = mode(salaries).mode[0]
skewness = skew(salaries)

print(f"Mean Salary: {mean_salary}")
print(f"Median Salary: {median_salary}")
print(f"Mode Salary: {mode_salary}")
print(f"Skewness: {skewness}")
```

OUTPUT

```
Mean Salary: 52700.0
Median Salary: 48500.0
Mode Salary: 47000
Skewness: 2.547189312037542
```

Scenario 3:

Product Sales Analysis

A retail store records product sales over 15 days. Create a frequency distribution table and visualize the sales data using appropriate charts.

Answer

```
import pandas as pd
import matplotlib.pyplot as plt

# Sample sales data for 15 days
sales = [25, 30, 28, 45, 55, 60, 22, 80, 95, 120, 33, 29, 27,
35, 40]

# Create pandas Series from sales data
sales_series = pd.Series(sales)

# Calculate frequency distribution
freq_dist = sales_series.value_counts().sort_index()

print("Frequency Distribution Table:")
print(freq_dist)

# Plot bar chart for frequency distribution
plt.figure(figsize=(10, 4))
freq_dist.plot(kind='bar', color='skyblue')
plt.title('Frequency Distribution of Sales')
plt.xlabel('Sale Value')
plt.ylabel('Frequency')
plt.show()

# Plot histogram of sales data
```

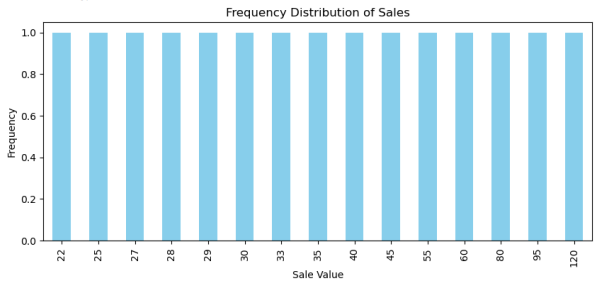
```
plt.figure(figsize=(10, 4))
plt.hist(sales, bins=10, color='lightgreen', edgecolor='black')
plt.title('Sales Histogram')
plt.xlabel('Sale Value')
plt.ylabel('Count')
plt.show()
```

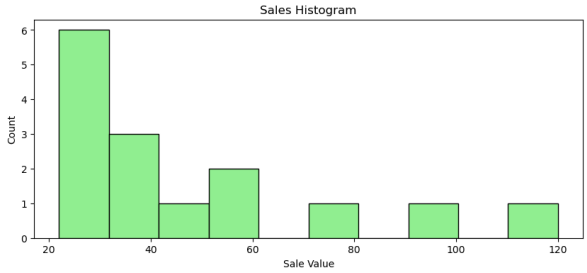
OUTPUT

Frequency Distribution Table:

22	1
25	1
27	1
28	1
29	1
30	1
33	1
35	1
40	1
45	1
55	1
60	1
80	1
95	1
120	1

Name: count, dtype: int64





Scenario 4:

Student Exam Performance Analysis

A school wants to analyze the exam performance of students across three subjects: Mathematics, Science, and English. How can Data Science concepts be applied to understand their performance?

Answer

Data Science can be applied to analyze student exam performance across Mathematics, Science, and English by:

1. Descriptive Statistics:
    - Calculate mean, median, mode, and standard deviation for each subject to understand central tendency and variability.
2. Visualization:
    - Plot histograms, boxplots, or violin plots per subject to visualize score distributions and detect outliers.
3. Correlation Analysis:
    - Compute correlation between subjects to see if performance in one relates to others.
4. Clustering/Segmentation:
    - Group students based on performance patterns using clustering algorithms to identify struggling or excelling groups.
5. Predictive Modeling:
    - Use regression or classification models to predict future performance or identify risk factors.
6. Trend Analysis:
    - Analyze score trends over time or across exams.

This approach provides holistic insights into student performance to support targeted interventions and improvements.

Scenario 5:

Clinical Trial for Diabetes Medication

A pharmaceutical company conducted a clinical trial with two groups: one receiving medication and the other a placebo. Perform a hypothesis test to determine the effectiveness of the medication.

Answer

Data Science concepts can be applied to understand student exam performance in Mathematics, Science, and English with the following steps:

1. Descriptive Statistics: Calculate averages (mean, median), spread (variance, standard deviation), and central tones for each subject to gauge performance levels.
2. Visual Analytics: Use histograms, boxplots, or violin plots to visualize each subject's score distribution and identify outliers or skewness.
3. Correlation Analysis: Measure relationships between subjects' scores to see if performing well in one affects others.
4. Clustering: Segment students into groups (high, medium, low performers) using clustering algorithms to tailor analysis or interventions.
5. Trend and Comparative Analysis: Track improvement or regression across exams or compare subject-wise performance across classes or demographics.
6. Predictive Modeling: Develop models forecasting student success or risk based on past scores.

Implementing these data-driven techniques provides actionable insights for educators to support students effectively.