TASK2P - BTC / USD Daily Closing Price In 2023

Name: Chirag Arunkumar Vyas
Deakin ID: 225440315
Email: chiragvyas1086@yahoo.com

Unit: SIG731 – Data Wrangling
TASK: TASK2P

INDEX

Introduction

In this report, we analyse the daily closing prices of Bitcoin (BTC) quoted in United States Dollars (USD) over the course of the year 2023. The objective is to load, preprocess, and investigate the behaviour of these prices through a sequence of exploratory and statistical tasks. Particular focus is placed on the third quarter of 2023, during which we compute standard descriptive measures, examine the daily evolution of prices, and study the corresponding price differences. The analysis proceeds without the use of the pandas library or explicit for-loops, thereby highlighting the strengths of NumPy and SciPy for vectorised numerical computation. All computations and visualisations are documented and explained in a structured and readable manner, with interpretations offered for the numerical summaries and plots.

1.1 Data Acquisition Downloading BTC USD Daily Prices for 2023

- In this section, we download the raw daily BTC-to-USD price data from Yahoo Finance, covering the period from 1 January 2023 to 31 December 2023. At this stage no preprocessing is performed.
- The data is stored locally as a CSV file, which will be inspected and cleaned in subsequent steps.

In [1]:
```
import yfinance as yf
import pandas as pd

# Define the ticker and date range
ticker_symbol = "BTC-USD"  # For Bitcoin in US Dollars
# ticker_symbol = "BTC-INR" # For Bitcoin in Indian Rupees, as seen in [6, 12]
start_date = "2023-01-01"
end_date = "2024-1-1"

# Download the data
# Use interval="1d" for daily data, or choose others like "1h", "1wk" [8]
btc_data = yf.download(ticker_symbol, start=start_date, end=end_date, interval="1d")

# Display the first few rows
print("Bitcoin Data:")
print(btc_data.head())

# Save the data to a CSV file
```

```
btc_data.to_csv("bitcoin_historical_data.csv")
print("\nData saved to bitcoin_historical_data.csv")
```

```
C:\Users\Chirag Pc\AppData\Local\Temp\ipykernel_18136\700505939.py:12: FutureWarning: YF.download() has changed argument auto_a
djust default to True
  btc_data = yf.download(ticker_symbol, start=start_date, end=end_date, interval="1d")
[*********************100%***********************]  1 of 1 completed
Bitcoin Data:
Price              Close          High           Low          Open  \
Ticker           BTC-USD       BTC-USD       BTC-USD       BTC-USD
Date
2023-01-01  16625.080078  16630.439453  16521.234375  16547.914062
2023-01-02  16688.470703  16759.343750  16572.228516  16625.509766
2023-01-03  16679.857422  16760.447266  16622.371094  16688.847656
2023-01-04  16863.238281  16964.585938  16667.763672  16680.205078
2023-01-05  16836.736328  16884.021484  16790.283203  16863.472656

Price            Volume
Ticker          BTC-USD
Date
2023-01-01   9244361700
2023-01-02  12097775227
2023-01-03  13903079207
2023-01-04  18421743322
2023-01-05  13692758566

Data saved to bitcoin_historical_data.csv
```

2.1 Data Preparation and Import

- The raw BTC-USD data obtained from Yahoo Finance contained multiple metadata rows as well as several columns (Date, Close, High, Low, Open, and Volume). Prior to analysis, the dataset was manually pre-processed using a spreadsheet application. In particular, the first three non-data rows were removed, all columns except the daily closing prices were deleted, and the resulting single column of observations was exported in CSV format.
- Each line of the exported file contains one numerical closing price, with no header row or supplementary information.
- In this section, the cleaned dataset is imported into the environment as a NumPy vector named rates. This vector forms the basis for all subsequent computations and visualisations.

In [2]:
```
import numpy as np

rates = np.loadtxt(
    r"E:\2. MDS DEAKIN UNIVERSITY\7. SIG 731 - Data wragling submissions\TASK 2 P\BTC_USD_data_cleaned.csv",
    encoding="utf-8-sig"
)

rates.shape
```

Out[2]:
```
(365,)
```

- The import yields a one-dimensional NumPy array of length 365, corresponding to the 365 daily closing BTC-USD prices observed throughout the year 2023.
- These observations enable quarter-specific slicing, computation of descriptive statistics, and graphical analysis in the subsequent sections of the report.

3.1 Quarter 3 Descriptive Analysis

- Quarter 3 of 2023 corresponds to days 182–273 inclusive within the 365-day observation period. Using zero-based indexing in NumPy, this range is obtained via indices 181–272.
- In this section, the Q3 observations are extracted and standard descriptive statistics are computed to summarise the central tendency and dispersion of the BTC-USD daily closing prices during this period

In [3]:
```
# Extract Q3 (days 182–273 inclusive, using 0-based indexing)
q3 = rates[181:273]
```

- We now determine the arithmetic mean, minimum, first quartile, median, third quartile, maximum, standard deviation, and interquartile range using vectorised NumPy operations.

In [4]:
```
import numpy as np

def summary_statistics(x):
    """Return mean, min, Q1, median, Q3, max, std, and IQR for a 1D NumPy array x."""
```

```
    mean = np.mean(x)
    x_min = np.min(x)
    q1 = np.percentile(x, 25)
    median = np.median(x)
    q3 = np.percentile(x, 75)
    x_max = np.max(x)
    std = np.std(x, ddof=0)
    iqr = q3 - q1
    return mean, x_min, q1, median, q3, x_max, std, iqr
```

In [5]:
```
q3 = rates[181:273]
q3_mean, q3_min, q3_q1, q3_med, q3_q3, q3_max, q3_std, q3_iqr = summary_statistics(q3)
```

- The resulting quantities are displayed in a structured and readable form.

In [6]:
```
print("╔═══════════════════════════════════════════════╗")
print("║   Q3 2023 BTC-USD — Summary Statistics        ║")
print("╠═══════════════════════════════════════════════╣")
print(f"║ Arithmetic mean:     {q3_mean:>12.2f}             ║")
print(f"║ Minimum:             {q3_min:>12.2f}             ║")
print(f"║ Q1:                  {q3_q1:>12.2f}             ║")
print(f"║ Median:              {q3_med:>12.2f}             ║")
print(f"║ Q3:                  {q3_q3:>12.2f}             ║")
print(f"║ Maximum:             {q3_max:>12.2f}             ║")
print(f"║ Std deviation:       {q3_std:>12.2f}             ║")
print(f"║ IQR:                 {q3_iqr:>12.2f}             ║")
print("╚═══════════════════════════════════════════════╝")
```

```
╔═══════════════════════════════════════════╗
║   Q3 2023 BTC-USD — Summary Statistics    ║
╠═══════════════════════════════════════════╣
║ Arithmetic mean:         28091.33         ║
║ Minimum:                 25162.65         ║
║ Q1:                      26225.56         ║
║ Median:                  28871.82         ║
║ Q3:                      29767.07         ║
║ Maximum:                 31476.05         ║
║ Std deviation:            1827.04         ║
║ IQR:                      3541.51         ║
╚═══════════════════════════════════════════╝
```

3.2 Discussion of Q3 Statistical Results

These values summarise the behaviour of BTC-USD closing prices over the third quarter of 2023. The arithmetic mean reflects the overall level during the period, whereas the median and quartiles indicate the location and spread of the central portion of the distribution. The interquartile range measures the dispersion of the middle 50% of the observations, and the standard deviation quantifies overall variability relative to the mean. The minimum and maximum provide the observed price extremes within the quarter.

4.1 Q3 Price Plot

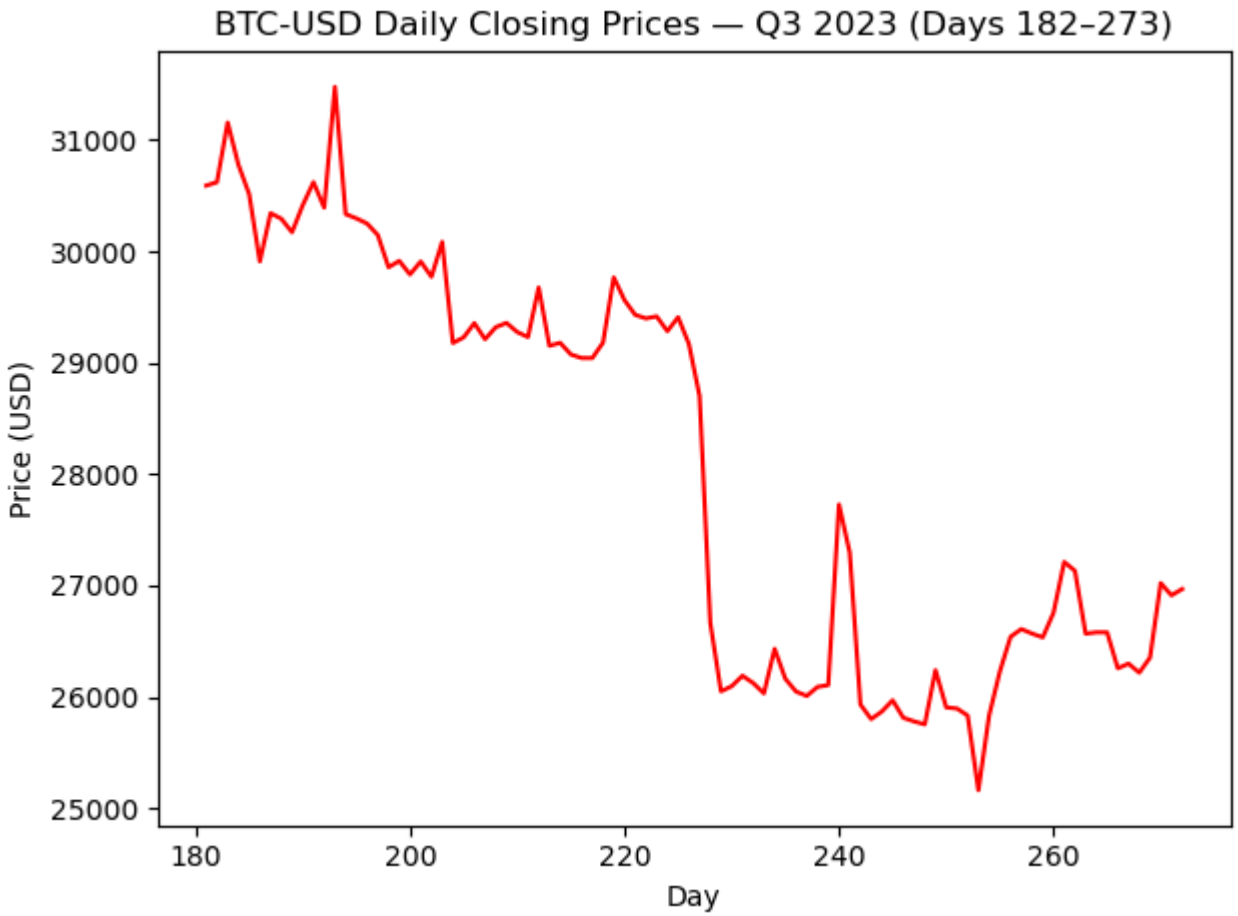
- Having examined the numerical summary of Q3 2023, we now visualise the BTC-USD closing prices over the same period. - - Days are numbered according to the assignment convention, with day 182 corresponding to 1 July and day 273 to 30 September.
- The resulting plot illustrates the day-to-day variation in cryptocurrency prices during the third quarter of 2023

In [7]:
```
import matplotlib.pyplot as plt
%matplotlib inline
```

In [8]:
```
import matplotlib.pyplot as plt
import numpy as np

def plot_q3_prices(rates):
    """Plot the Q3 BTC-USD daily closing prices using the assignment convention."""
```

```
    days = np.arange(181, 273)  # day numbering for display
    q3 = rates[181:273]         # 0-based indexing for underlying data

    plt.plot(days, q3, color='red')
    plt.title("BTC-USD Daily Closing Prices — Q3 2023 (Days 182–273)")
    plt.xlabel("Day")
    plt.ylabel("Price (USD)")
    plt.tight_layout()
    plt.show()
```

In [9]:
```
plot_q3_prices(rates)
```



- The plot reveals several notable features of price behaviour during Q3 2023.
- The quarter begins at comparatively elevated price levels, followed by a gradual downward trajectory interspersed with short-term fluctuations. In the latter half of the quarter, the series exhibits sharper declines and occasional abrupt movements, characteristic of the high volatility associated with cryptocurrency assets.
- Toward the end of the period, the prices appear to stabilise somewhat, oscillating within a narrower band without a clear directional trend.
- Overall, the visual evidence confirms that BTC-USD prices during Q3 2023 were highly variable and lacked sustained upward or downward momentum.

4.2 Extremal Day Positions in Q3

- In addition to the descriptive statistics and visual analysis, the assignment requires the identification of the specific days within Q3 2023 on which the lowest and highest BTC-USD closing prices were observed.
- Following the convention that day 182 corresponds to 1 July, the daily indices for Q3 are extracted and mapped to the corresponding day numbers.

In [10]:
```python
import numpy as np

def q3_extremal_days(rates):
    """
    Return the assignment-style day numbers and prices for the minimum and maximum
    BTC-USD closing prices in Q3 2023. Day 182 denotes 1 July.
    """
    q3 = rates[181:273]         # extract Q3 period using 0-based indexing
    min_idx = np.argmin(q3)
    max_idx = np.argmax(q3)
    min_day = min_idx + 182     # convert to assignment day numbering
    max_day = max_idx + 182
    min_price = q3[min_idx]
    max_price = q3[max_idx]
    return min_day, min_price, max_day, max_price
```

In [11]:
```python
def print_q3_extremal_days(min_day, min_price, max_day, max_price):
    """
```

```
    Display the extremal Q3 prices and day numbers in a clean, readable format.
    """
    print("┌──────────────────────────────────────────────────┐")
    print("│     Q3 2023 BTC-USD — Extremal Day Results       │")
    print("├──────────────────────────────────────────────────┤")
    print(f"│ {'Lowest price on day':<22} {min_day:>6}  ({min_price:>10.2f})  │")
    print(f"│ {'Highest price on day':<22} {max_day:>6}  ({max_price:>10.2f})  │")
    print("└──────────────────────────────────────────────────┘")
```

In [12]:
```
min_day, min_price, max_day, max_price = q3_extremal_days(rates)
print_q3_extremal_days(min_day, min_price, max_day, max_price)
```

```
┌──────────────────────────────────────────────────┐
│     Q3 2023 BTC-USD — Extremal Day Results       │
├──────────────────────────────────────────────────┤
│ Lowest price on day       254  (  25162.65)  │
│ Highest price on day      194  (  31476.05)  │
└──────────────────────────────────────────────────┘
```

- The highest closing price in Q3 2023 occurred early in the quarter, whereas the lowest appeared toward the end.
- This temporal pattern aligns with the downward trend visible in the Q3 plot and corroborates the descriptive statistics, which indicated a decline in central tendency across the period.

4.3 Report Analysis of Highest and Lowest Q3 Closing Prices

- The highest observed closing price in Q3 2023 occurred on day 194, while the lowest closing price occurred on day 254.
- According to the assignment's day numbering convention, day 182 corresponds to 1 July; thus, the maximum price was attained relatively early in the quarter and the minimum near the end.
- This temporal positioning corroborates the general downward movement visible in the Q3 price plot and further reflects the volatile dynamics of the cryptocurrency market over the examined period.

5.1 Analysis of Daily Price Changes via Boxplot

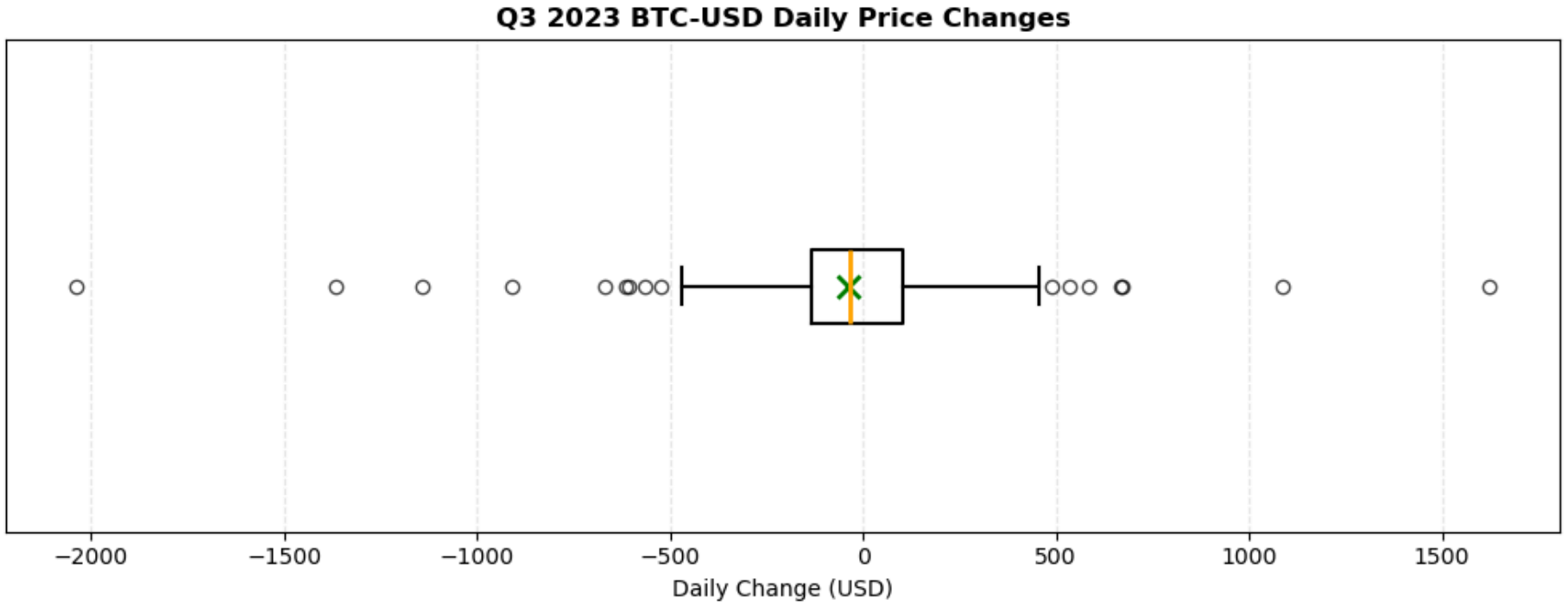
- Daily movements in the BTC-USD closing price provide insight into short-term volatility over the course of Q3 2023. To examine this behaviour, we compute daily differences using the vectorised numpy.
- diff operator applied to the Q3 segment of the data. These differences quantify the day-to-day increases and decreases in the asset's value.
- A horizontal box-and-whisker plot is then employed to visualise the distribution of these daily price changes.
- This representation highlights the central location, spread, and skewness of the distribution, while also illustrating the presence of extreme movements through whiskers and potential outliers.
- The arithmetic mean of the daily changes is plotted using a green "x" marker to provide a point of comparison against the distribution's median.

In [13]:
```python
import numpy as np

def q3_daily_changes(rates):
    """Return daily price changes for Q3 2023 based on numpy.diff."""
    q3 = rates[181:273]
    return np.diff(q3)
```

In [14]:
```python
def plot_q3_changes_boxplot(changes):
    """Draw an enhanced horizontal boxplot of Q3 daily price changes with mean marker."""

    mean_val = np.mean(changes)

    plt.figure(figsize=(10, 4))  # wider aspect for readability

    # prettier boxplot settings
    plt.boxplot(
        changes,
        vert=False,
        showmeans=False,
        showfliers=True,
        boxprops=dict(linewidth=1.5),
        whiskerprops=dict(linewidth=1.5),
        capprops=dict(linewidth=1.5),
        flierprops=dict(marker='o', markersize=6, alpha=0.7),
        medianprops=dict(color='orange', linewidth=2)
    )
```

```
    # mean marker
    plt.plot(mean_val, 1, marker='x', color='green', markersize=10, markeredgewidth=2)

    # formatting
    plt.title("Q3 2023 BTC-USD Daily Price Changes", fontsize=12, fontweight='bold')
    plt.xlabel("Daily Change (USD)", fontsize=10)

    # remove y-axis clutter
    plt.yticks([])

    # grid for easier reading
    plt.grid(axis='x', alpha=0.3, linestyle='--')

    plt.tight_layout()
    plt.show()
```

In [15]:
```
changes = q3_daily_changes(rates)
plot_q3_changes_boxplot(changes)
```



5.2 Interpretation of the Boxplot

- The horizontal box-and-whisker plot reveals several notable characteristics of the Q3 2023 BTC-USD daily price changes.
- The central box is relatively narrow, indicating that the interquartile range of daily movements was modest.
- This suggests that, on most days, price changes were moderate and clustered around the median.
- The arithmetic mean, highlighted with a green “x”, lies slightly to the right of the median, implying a mild right-skew in the distribution of daily changes.
- In other words, unusually large upward movements occurred more frequently than unusually large downward movements relative to the typical daily fluctuations.
- A substantial number of points lie outside the whiskers on both sides of the distribution, confirming the presence of outliers.
- These extreme movements, both positive and negative, are consistent with the high volatility commonly observed in cryptocurrency markets.
- The negative outliers appear more numerous and more extreme in magnitude than the positive ones, which suggests that sharp downward corrections were more prevalent than large upward spikes during this quarter.

6.1 Outlier Identification and Interpretation

- To further analyse the variability of daily price movements, we identify outliers within the Q3 2023 daily price changes using the classical interquartile range (IQR) criterio
- . Under this rule, an observation (x) is declared an outlier if

$$x < Q_1 - 1.5 \cdot \mathrm{IQR} \quad \text{or} \quad x > Q_3 + 1.5 \cdot \mathrm{IQR}$$

where

$$\mathrm{IQR} = Q_3 - -Q_1.$$

This computation is performed in a fully vectorised manner using NumPy's relational operators, thereby avoiding explicit loops while enabling efficient identification of unusually large positive or negative deviations in the dail

- y price changes. The resulting outlier count provides a quantitative indication of extreme short-term price movements relative to typical daily behaviour. daily behaviour.

In [16]:
```python
def count_outliers_IQR(changes):
    """
    Count outliers in the daily price changes using the IQR rule (no loops).
    Outlier: < Q1 - 1.5*IQR or > Q3 + 1.5*IQR.
    """
    q1 = np.percentile(changes, 25)
    q3 = np.percentile(changes, 75)
    iqr = q3 - q1
    lower = q1 - 1.5 * iqr
    upper = q3 + 1.5 * iqr

    outliers = (changes < lower) | (changes > upper)
    return np.sum(outliers), lower, upper
```

In [17]:
```python
def print_outlier_summary(count, lower, upper):
    """Display outlier information in a compact and readable form."""
    print("┌──────────────────────────────────────────────┐")
    print("│  Q3 2023 BTC-USD — Outlier Diagnostics       │")
    print("├──────────────────────────────────────────────┤")
    print(f"│ {'Lower bound:':<18} {lower:>12.2f}           │")
    print(f"│ {'Upper bound:':<18} {upper:>12.2f}           │")
    print(f"│ {'Outlier count:':<18} {count:>12d}           │")
    print("└──────────────────────────────────────────────┘")
```

In [18]:
```python
count, lower, upper = count_outliers_IQR(changes)
print_outlier_summary(count, lower, upper)
```

```
┌──────────────────────────────────────┐
│  Q3 2023 BTC-USD — Outlier Diagnostics  │
├──────────────────────────────────────┤
│ Lower bound:            -494.48      │
│ Upper bound:             464.15      │
│ Outlier count:               16      │
└──────────────────────────────────────┘
```

6.2 Interpretation of Outlier Results

The IQR-based diagnostic identifies **16 outliers** in the Q3 2023 daily price changes. Given that outliers are defined as changes more negative than

−494.48 USD or more positive than **+464.15 USD**, these correspond to days exhibiting unusually large downward or upward movements relative to typical daily fluctuations. The magnitude of the lower threshold being slightly more extreme than the upper threshold suggests that sharp negative corrections occurred more forcefully than positive surges during this period.

In the context of cryptocurrency markets, such outliers are not unexpected; they are consistent with the high volatility and episodic price shocks that characterise digital asset behaviour. These extreme movements may be attributable to macroeconomic announcements, shifts in market sentiment, or reactions to news events that provoke rapid changes in trading activity. The relatively large number of outliers reinforces the conclusion that BTC-USD prices in Q3 2023 experienced a considerable degree of short-term unpredictability.

7.1 Final Conclusion

- This report examined the daily BTC-USD closing prices for the year 2023 with a particular focus on the third quarter (days 182–273).
- After manual data preparation and import, Q3 was subjected to a series of numerical and graphical analyses.
- The descriptive statistics indicated a clear downward shift in central tendency over the period, and the time series plot confirmed the presence of substantial short-term fluctuations without a sustained upward or downward trend.
- The highest and lowest observed prices were found to occur early and late in the quarter, respectively, consistent with the overall declining trajectory.
- Daily price changes were subsequently analysed to quantify short-term volatility.

- The boxplot representation illustrated a moderate interquartile range alongside numerous unusually large movements. The IQR-based diagnostic identified sixteen outliers, underscoring the presence of episodic price spikes and corrections characteristic of cryptocurrency markets.
- These findings collectively highlight the volatile and reactive nature of BTC-USD prices during Q3 2023, reflecting both standard day-to-day variability and periodic extreme deviations from typical behaviour.

========== END OF TASK2P ==========

In [ ]: