EXPLORATORY DATA ANALYSIS REPORT

Dataset: Duration

Generated: June 19, 2025 at 09:30 PM

Prepared by: Mini EDA Tool

Data Science Project

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY

This report presents a comprehensive exploratory data analysis (EDA) of the uploaded dataset. The analysis covers data quality assessment, statistical summaries, distribution analysis, and correlation studies to provide actionable insights for further data processing and modeling.

KEY METRICS:

- Total Records: 169
- Total Features: 4
- Missing Values: 5

2. DATASET OVERVIEW

2.1 Dataset Information

Dataset Shape: 169 rows × 4 columns

Memory Usage: 6.6 KB

2.2 Feature List

1. Duration
2. Pulse
3. Maxpulse
4. Calories

3. DATA QUALITY ASSESSMENT

3.1 Data Types

• int64: 3 columns

• float64: 1 columns

3.2 Missing Values Analysis

Columns with missing values:

• Calories: 5 (3.0%)

4. STATISTICAL ANALYSIS

4.1 Summary Statistics

Duration:

Mean: 63.85 | Std: 42.30 | Min: 15.00 | Max: 300.00
Q1: 45.00 | Median: 60.00 | Q3: 60.00

Pulse:

Mean: 107.46 | Std: 14.51 | Min: 80.00 | Max: 159.00
Q1: 100.00 | Median: 105.00 | Q3: 111.00

Maxpulse:

Mean: 134.05 | Std: 16.45 | Min: 100.00 | Max: 184.00
Q1: 124.00 | Median: 131.00 | Q3: 141.00

Calories:

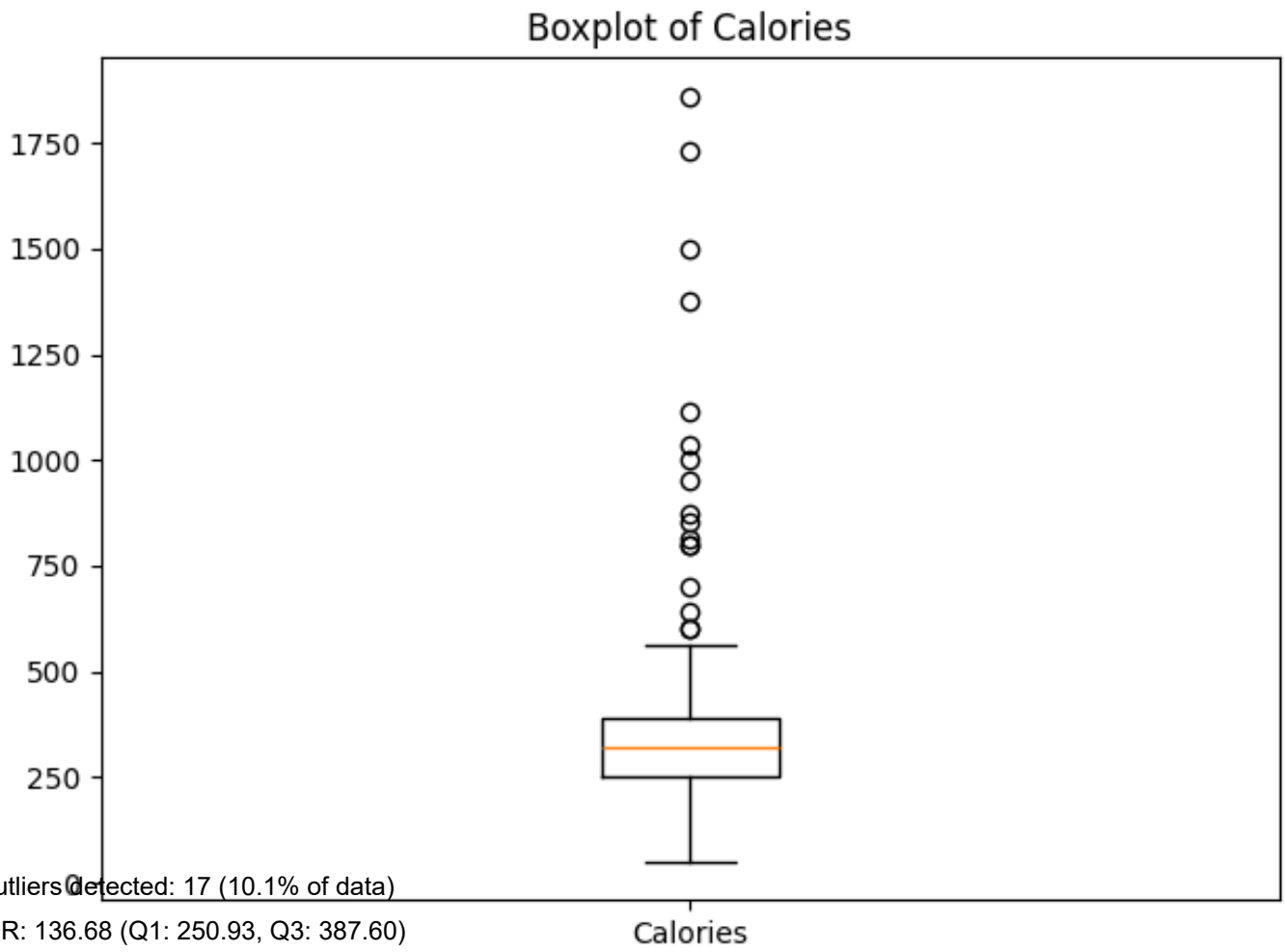
Mean: 375.79 | Std: 266.38 | Min: 50.30 | Max: 1860.40
Q1: 250.93 | Median: 318.60 | Q3: 387.60

5. DATA DISTRIBUTION ANALYSIS

The following visualizations provide insights into the distribution patterns of individual variables and their relationships. This analysis helps identify outliers, skewness, and potential data quality issues.

5. DATA VISUALIZATIONS

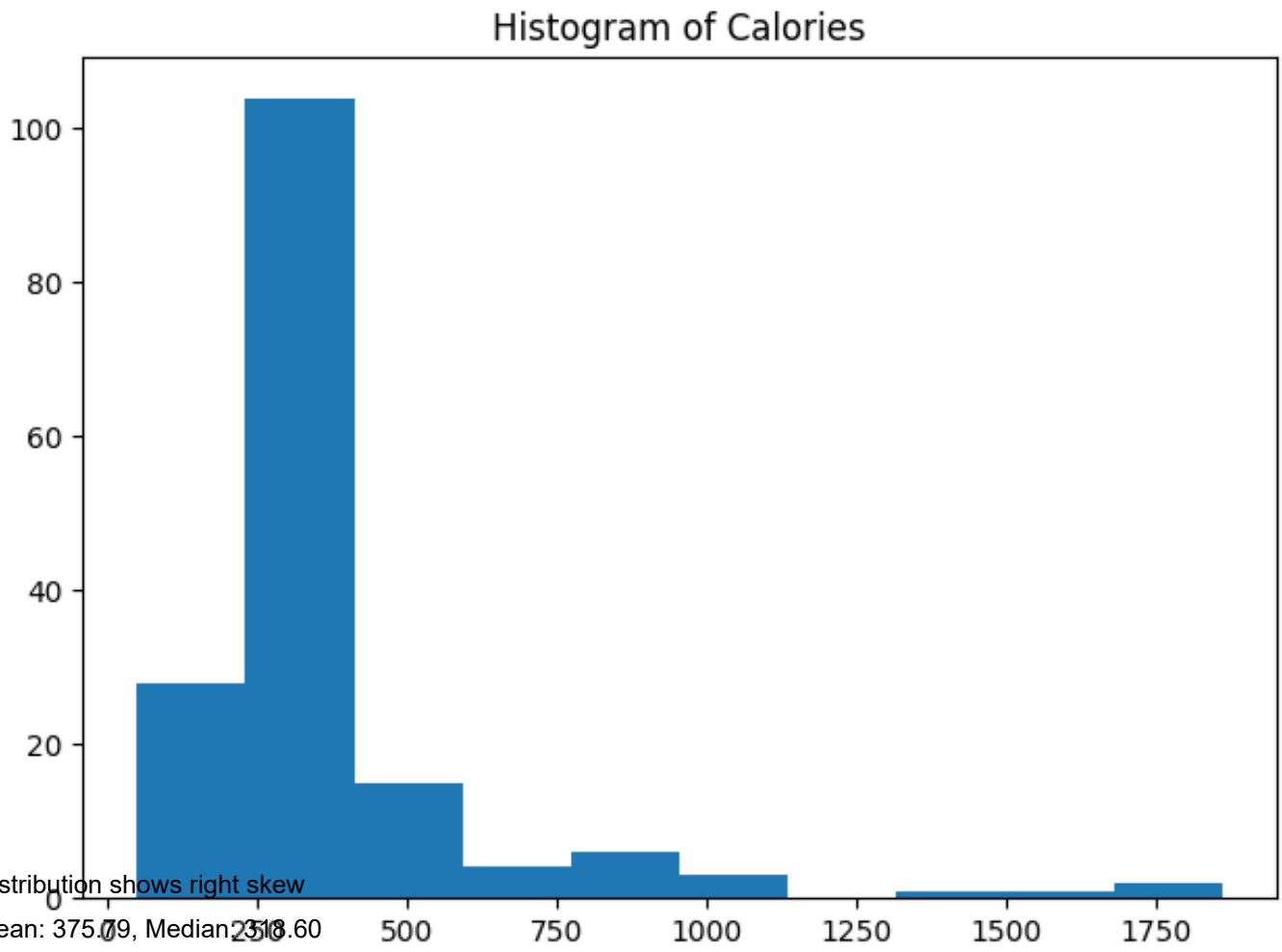
Calories Box



• Outliers detected: 17 (10.1% of data)

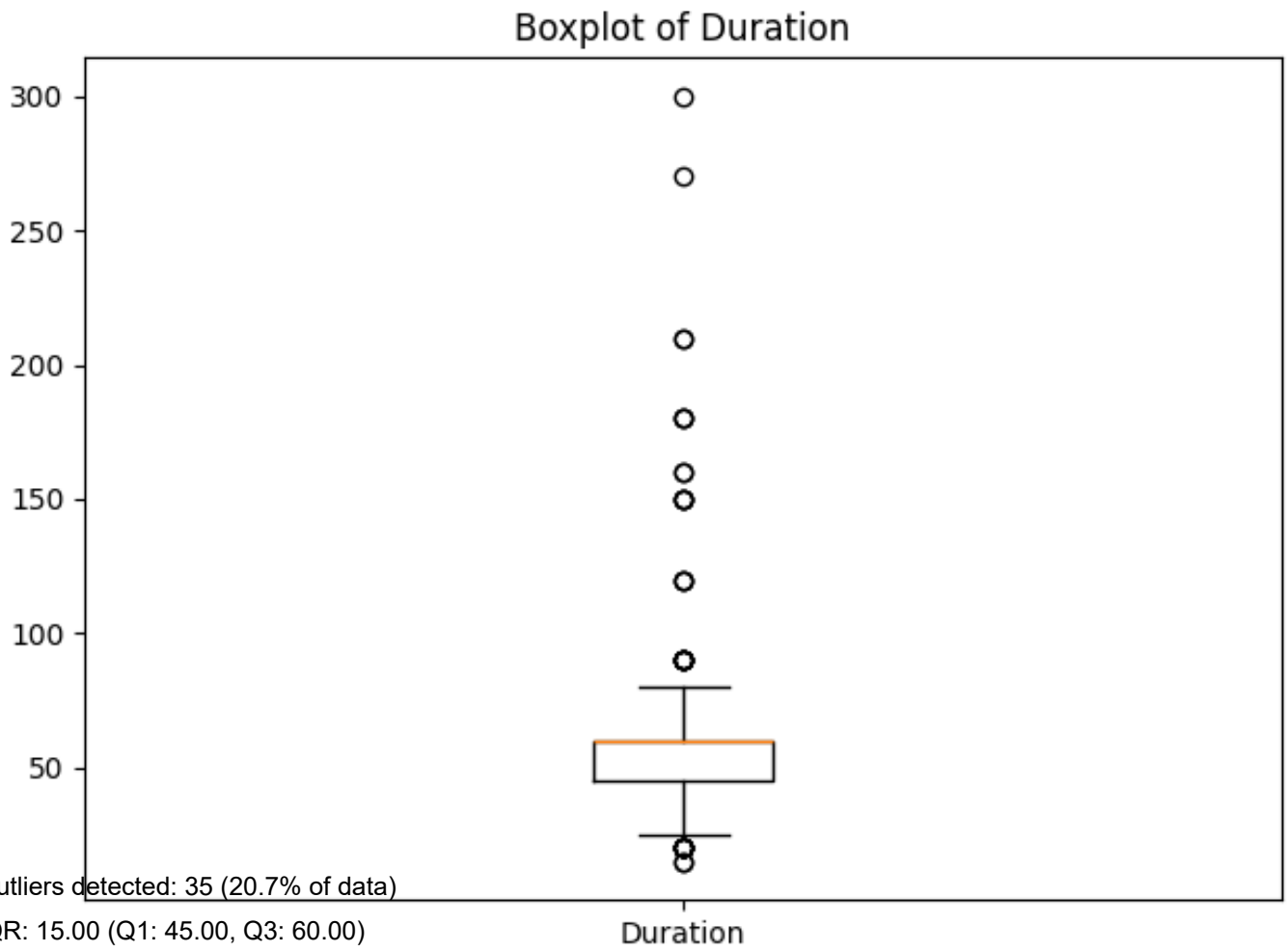
• IQR: 136.68 (Q1: 250.93, Q3: 387.60)

Calories Hist



- Distribution shows right skew
- Mean: 375.79, Median: 318.60

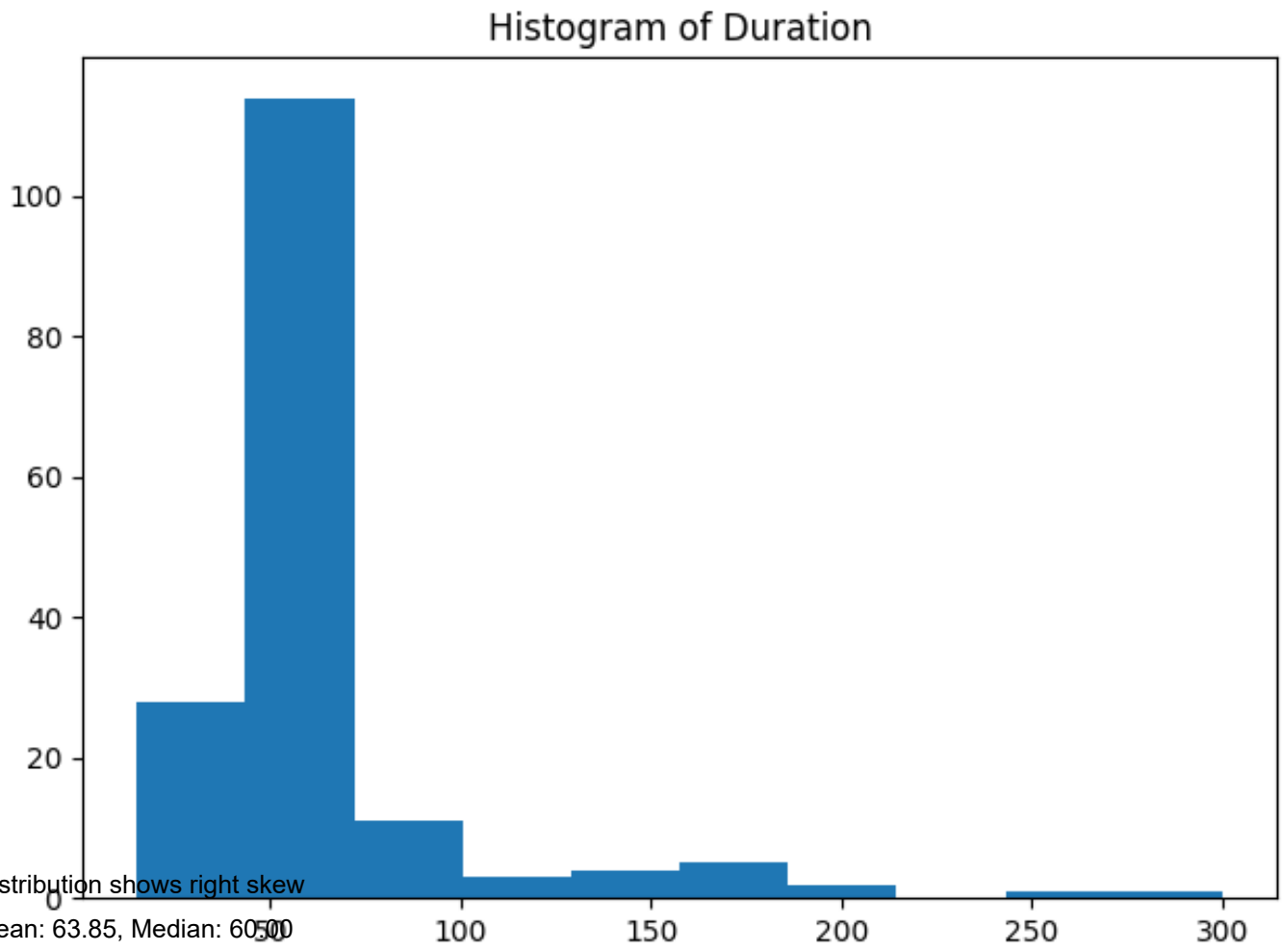
Duration Box



• Outliers detected: 35 (20.7% of data)

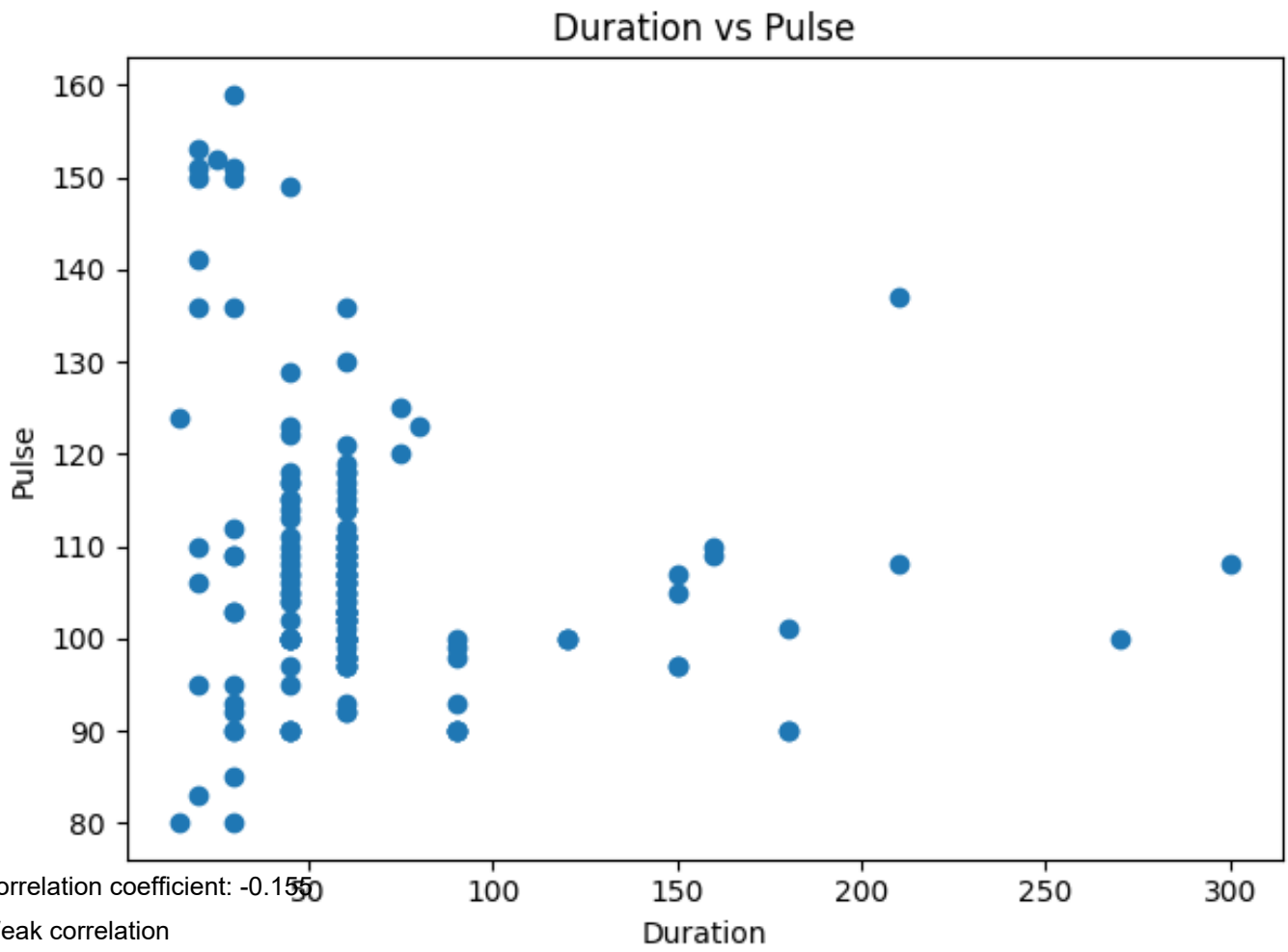
• IQR: 15.00 (Q1: 45.00, Q3: 60.00)

Duration Hist



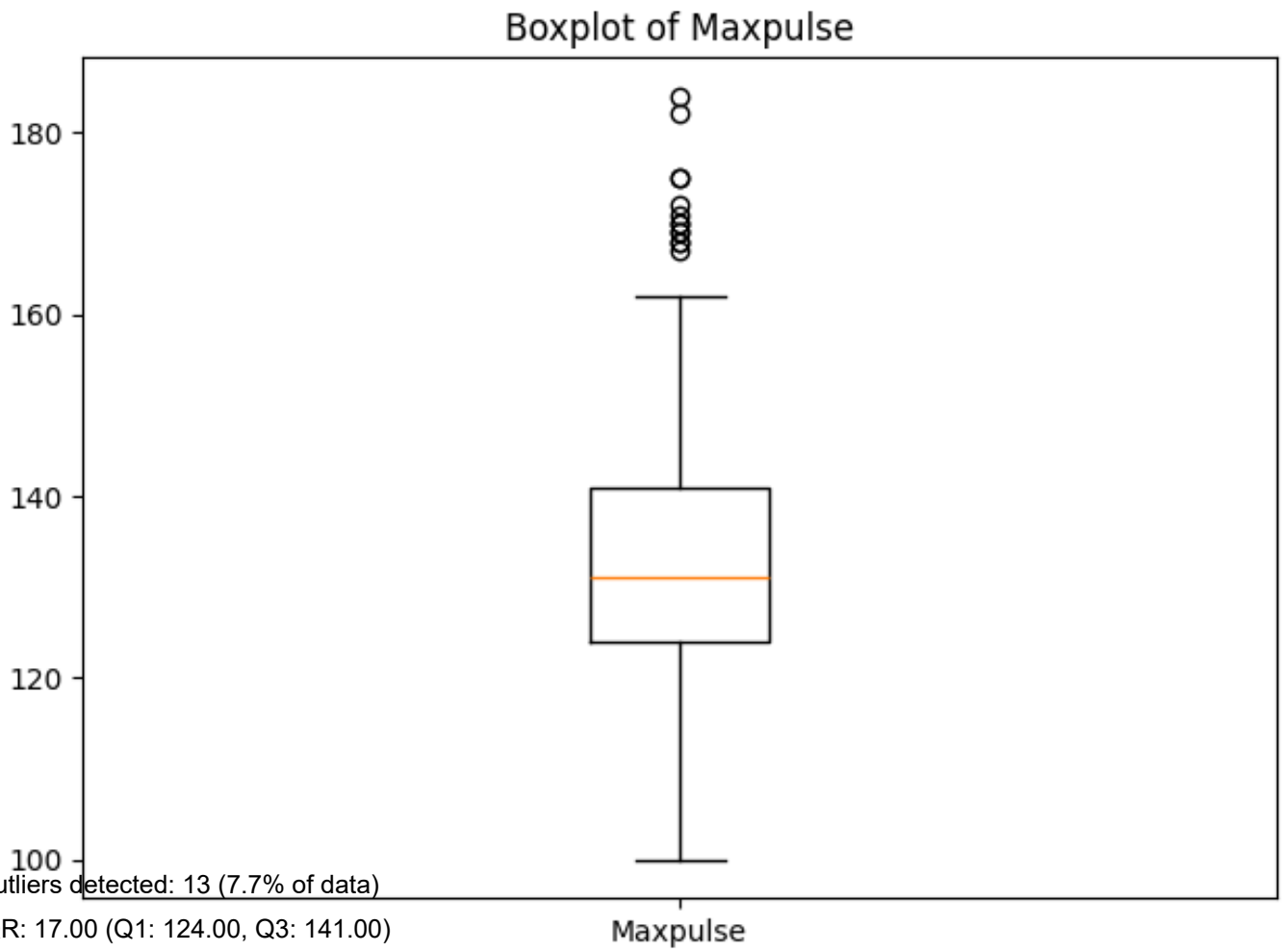
• Distribution shows right skew
• Mean: 63.85, Median: 60.00

Duration Vs Pulse



- Correlation coefficient: -0.155
- Weak correlation

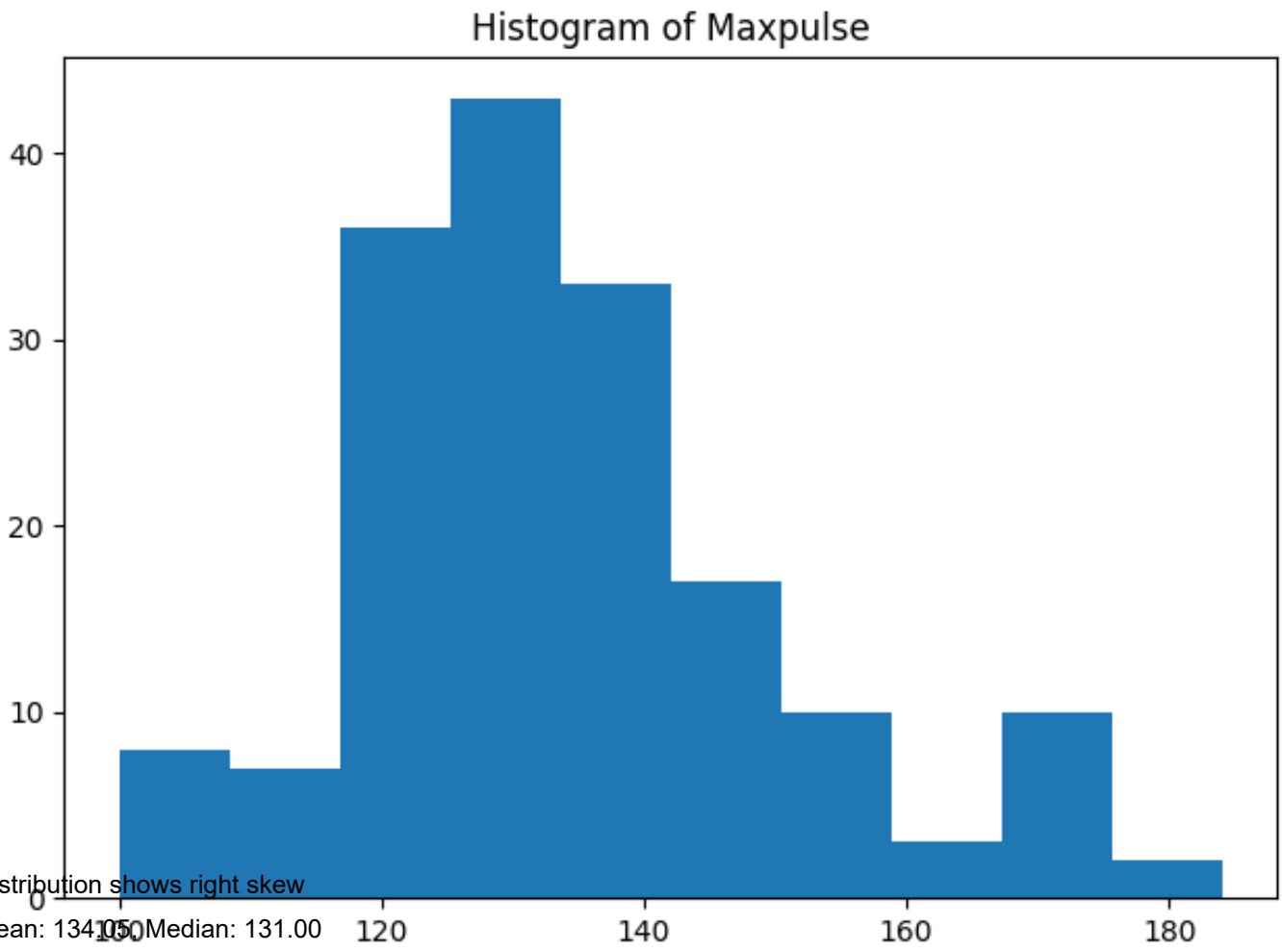
Maxpulse Box



• Outliers detected: 13 (7.7% of data)

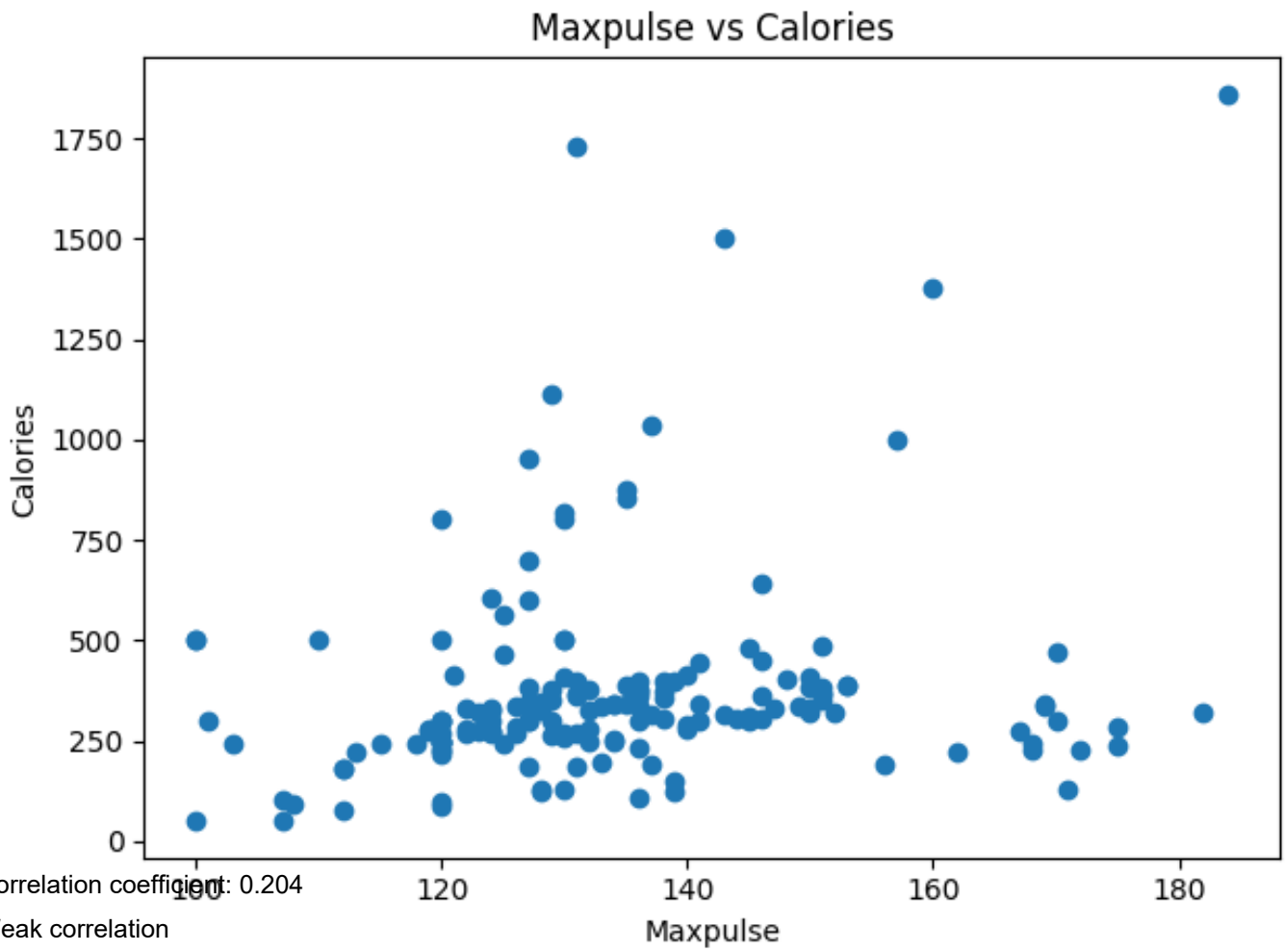
• IQR: 17.00 (Q1: 124.00, Q3: 141.00)

Maxpulse Hist



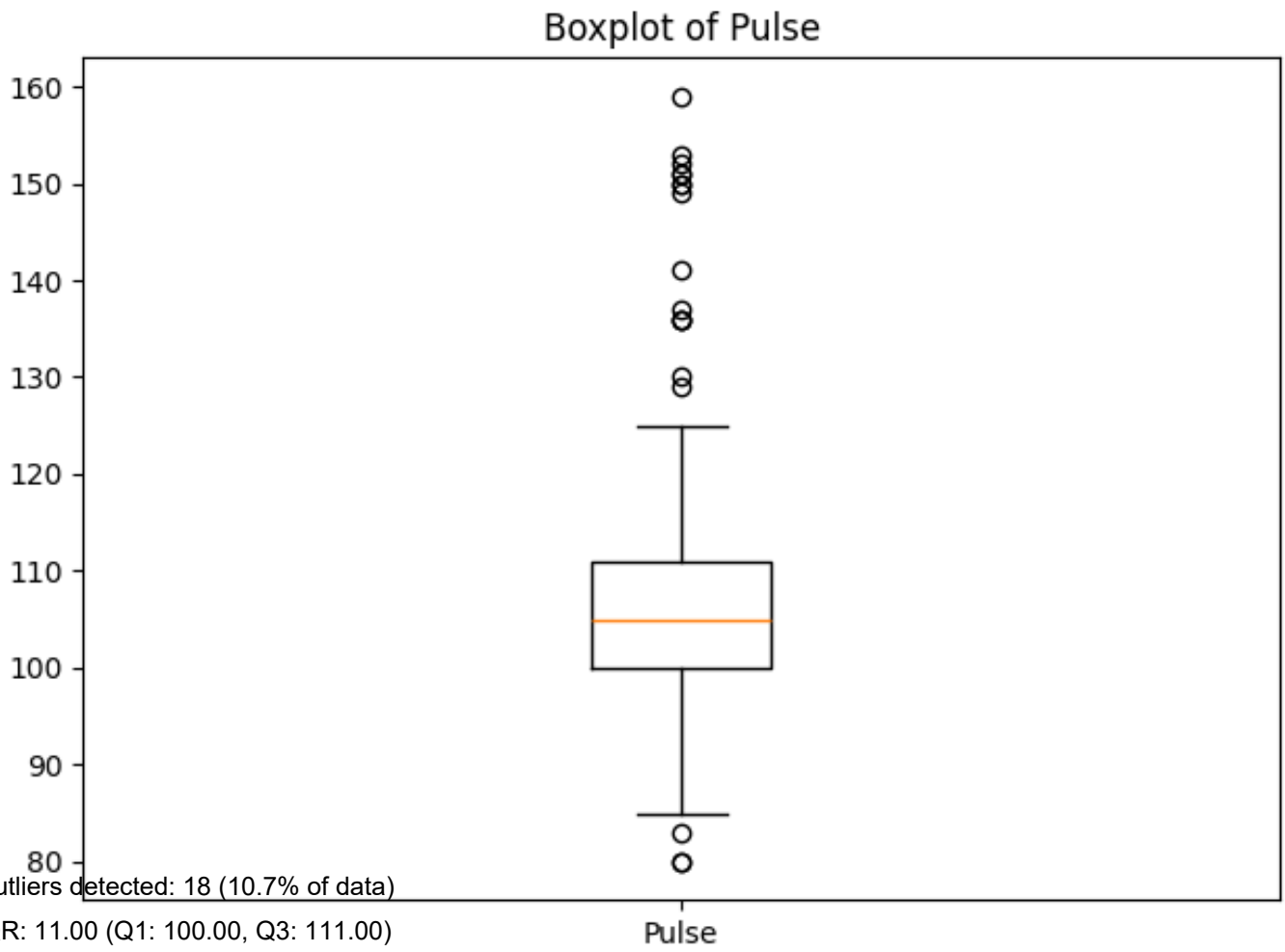
• Distribution shows right skew
• Mean: 134.05 Median: 131.00

Maxpulse Vs Calories



- Correlation coefficient: 0.204
- Weak correlation

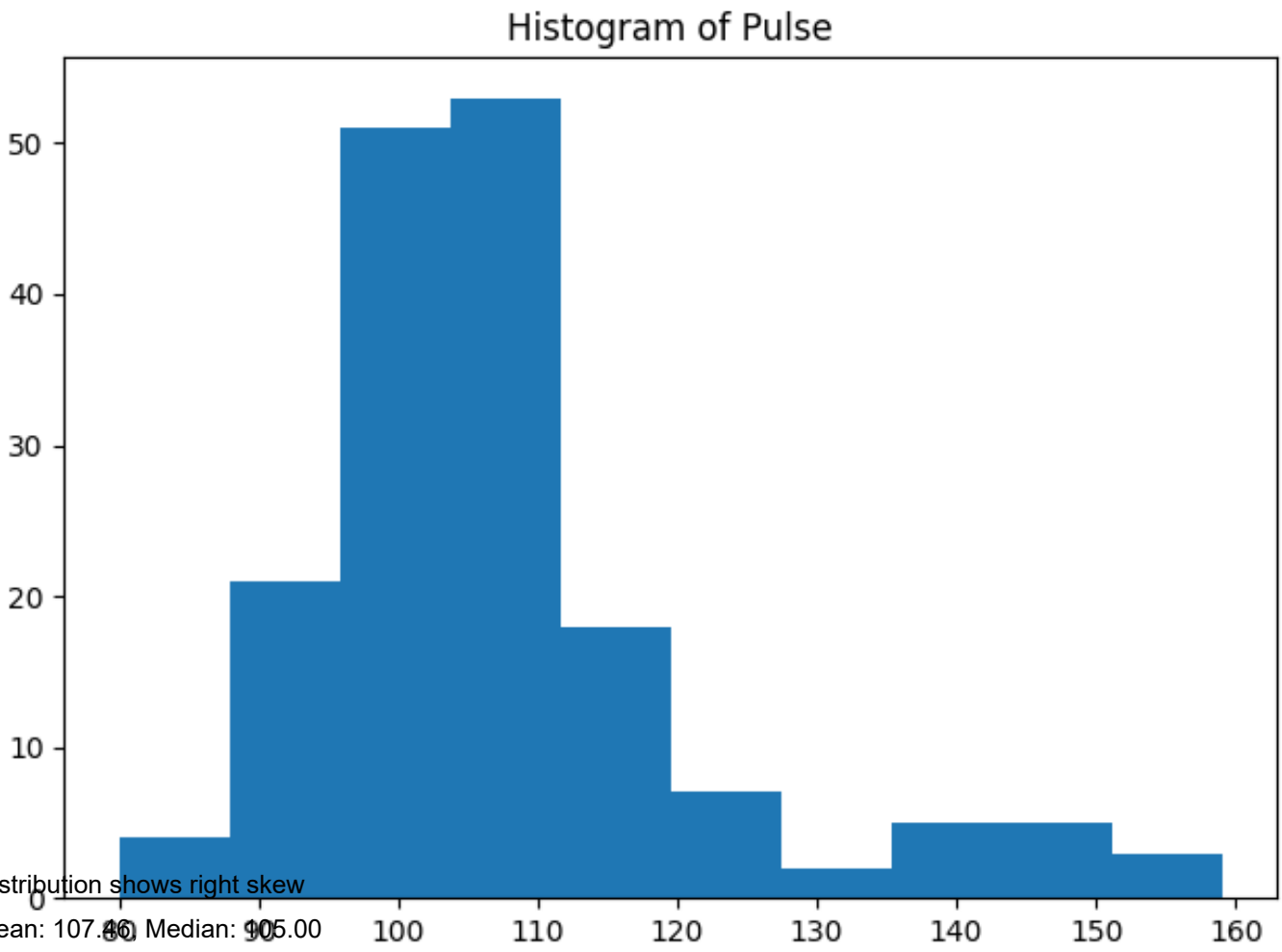
Pulse Box



• Outliers detected: 18 (10.7% of data)

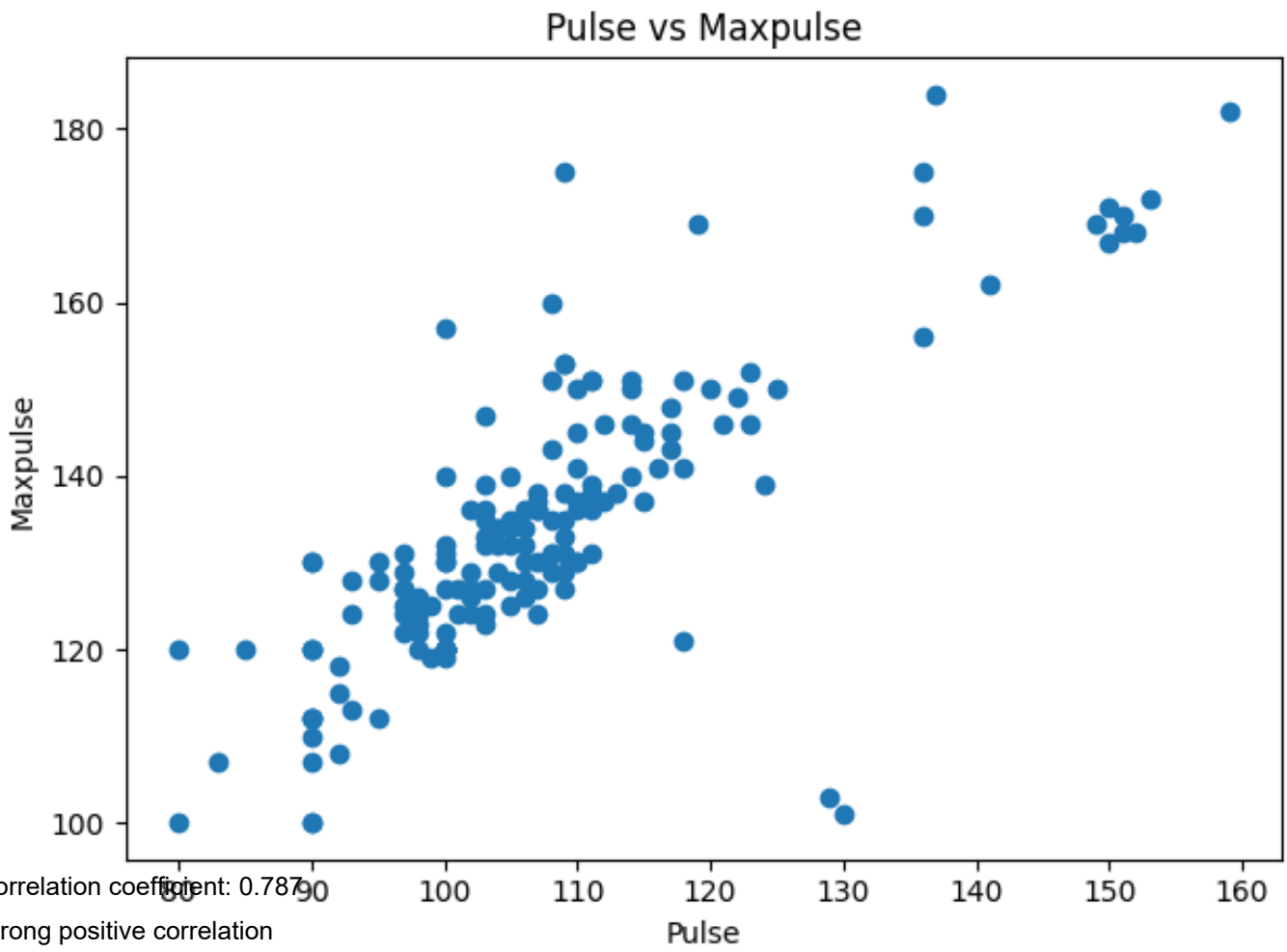
• IQR: 11.00 (Q1: 100.00, Q3: 111.00)

Pulse Hist



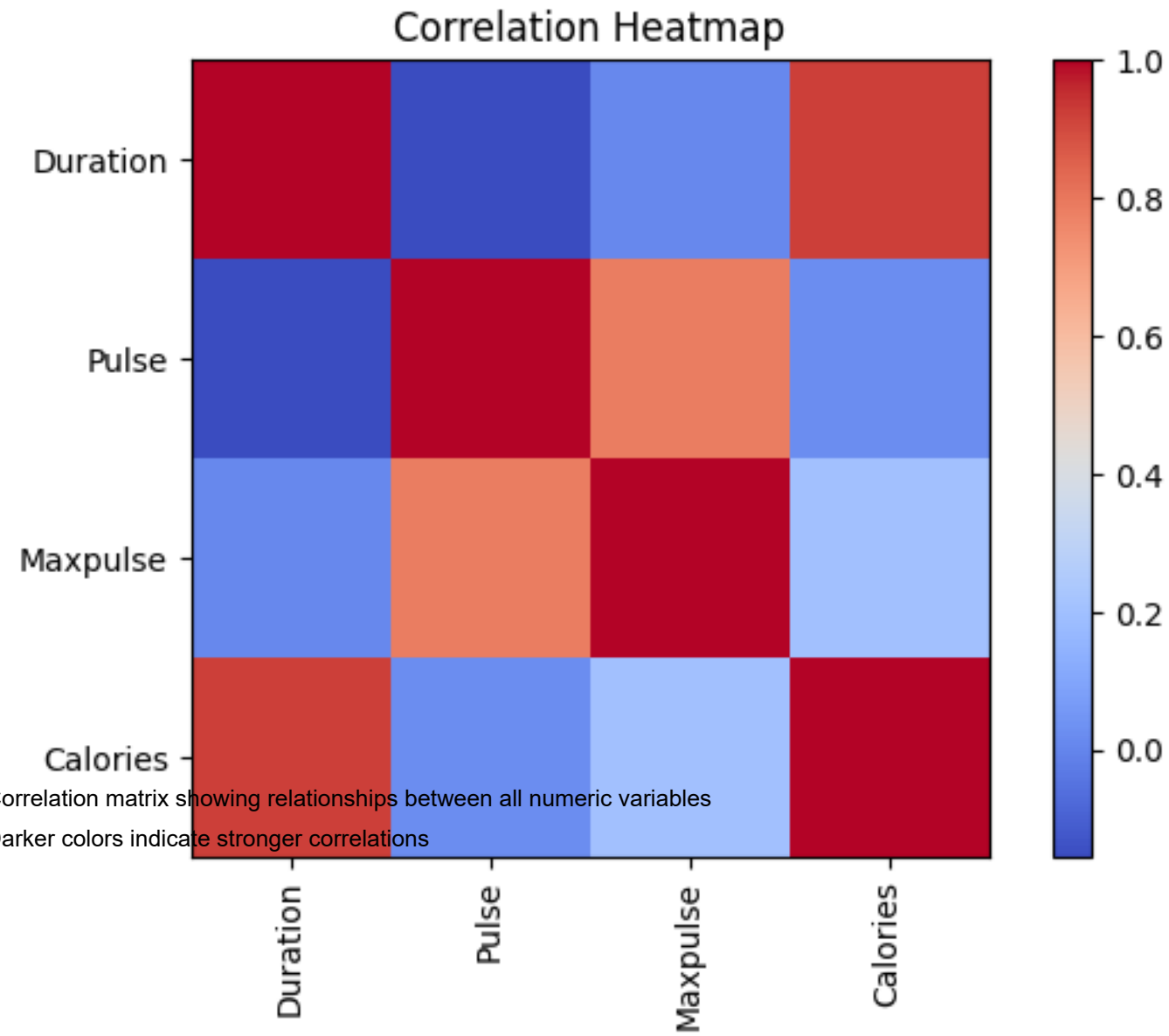
• Distribution shows right skew
• Mean: 107.46, Median: 105.00

Pulse Vs Maxpulse



- Correlation coefficient: 0.787
- Strong positive correlation

Heatmap



• Correlation matrix showing relationships between all numeric variables

• Darker colors indicate stronger correlations

6. CORRELATION ANALYSIS

Correlation analysis reveals relationships between numeric variables. Full visualizations are available in your analysis dashboard.

6.1 Strong Correlations (|r| > 0.7):

• Duration ↔ Calories: 0.923

• Pulse ↔ Maxpulse: 0.787

6.2 Moderate Correlations (0.3 < |r| < 0.7):

• No moderate correlations found

7. KEY INSIGHTS & RECOMMENDATIONS

7.1 Data Quality Issues

• Duration: 35 potential outliers
• Pulse: 18 potential outliers
• Maxpulse: 13 potential outliers
• Calories: 5 missing values
• Calories: 17 potential outliers

7.2 Recommendations

• Handle missing values through imputation or removal based on business context
• Investigate and treat outliers appropriately
• Consider feature engineering for highly correlated variables
• Apply appropriate scaling/normalization for skewed variables
• Validate data quality with domain experts
• Consider sampling strategies if dealing with imbalanced classes

8. APPENDICES

8.1 Detailed Statistics

Duration:

count: 169.0

mean: 63.846

std: 42.3

min: 15.0

25%: 45.0

50%: 60.0

75%: 60.0

max: 300.0

Pulse:

count: 169.0

mean: 107.462

std: 14.51

min: 80.0

25%: 100.0

50%: 105.0

75%: 111.0

max: 159.0

Maxpulse:

count: 169.0

mean: 134.047

std: 16.45

min: 100.0

25%: 124.0

50%: 131.0

75%: 141.0

max: 184.0

Calories:

count: 164.0

mean: 375.79

std: 266.38

min: 50.3

25%: 250.925

50%: 318.6

75%: 387.6

max: 1860.4