Analysis Report

A summary of the calculated statistics(Before and After) removal of outliers (level by level) :

Stage	Mean	Median	Std Dev	Variance	Q1	Q2 (Median)	Q3	IQR
Original	106.258187	101.06	35.8010487	1281.71509	87.284	101.06	115.74	28.4559999
After Removing Moderate	106.294122	101.08	35.5504761	1263.83635	87.392	101.08	115.67	28.27800
After Removing High	103.105983	100.37	30.1132907	906.810282	86.9665	100.37	114.26	27.293500
After Removing Extreme	100.1117773	99.937	19.5786019	383.321653	86.6895	99.937	113.44	26.750500

- Q1 = 1st quartile, Q2 = median, Q3 = 3rd quartile, IQR = Q3 − Q1.
  Removing outliers progressively reduces mean and standard deviation while stabilizing the distribution

Itemized list of identified outliers with their scores and classifications :

Value	Outlier Score	Level
159.76	1.54694	Moderate
229.172	3.986	High
245.429	4.873	High
292.805	6.222	Extreme

- **Value:** The data point identified as an outlier.
- **Outlier Score:** Calculated using:

Outlier Score= data point - Q3 / IQR → if it is above upper bound

Outlier Score= Q1 - data point / IQR → if it is below lower bound

- **Classification:**
  - **Moderate:** Score between 1.5 and 3
  - **High:** Score between 3 and 5
  - **Extreme:** Score greater than 5

Discussion on Impact of Outliers :

Effect on Mean and Median

- **Mean:**
  Highly sensitive to extreme values. With outliers, the average changes a lot. After removing them, the average is closer to the middle
- **Median:**
  Remains stable through most removal levels. This shows that most of the data stays the same

Effect on Standard Deviation and Variance

- Both decrease after removing High and Extreme outliers
- This indicates that the dataset becomes more compact and less spread out

Possible Real-World Reasons for Outliers

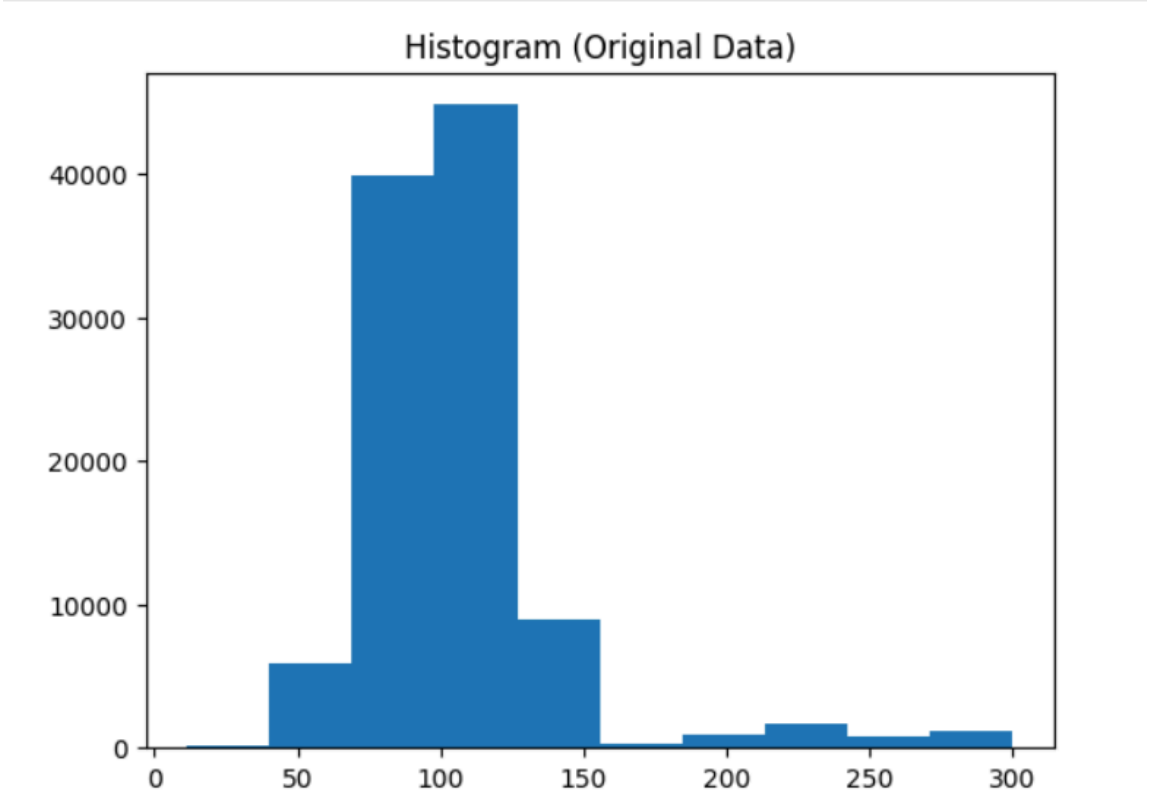
Outliers may appear due to:

- **Unique but Real Events**
  Unusual spikes from natural or human causes
- **Measurement or Sensor Errors**
  Broken or wrong equipment giving wrong values
- **Sampling Issues**
  Data coming from a different group
- **Natural Variability**
  Some things just naturally have extreme values
- **Data Entry Mistakes**
  Typing errors or wrong units (like 500 instead of 50)

Visualizations and Captions :

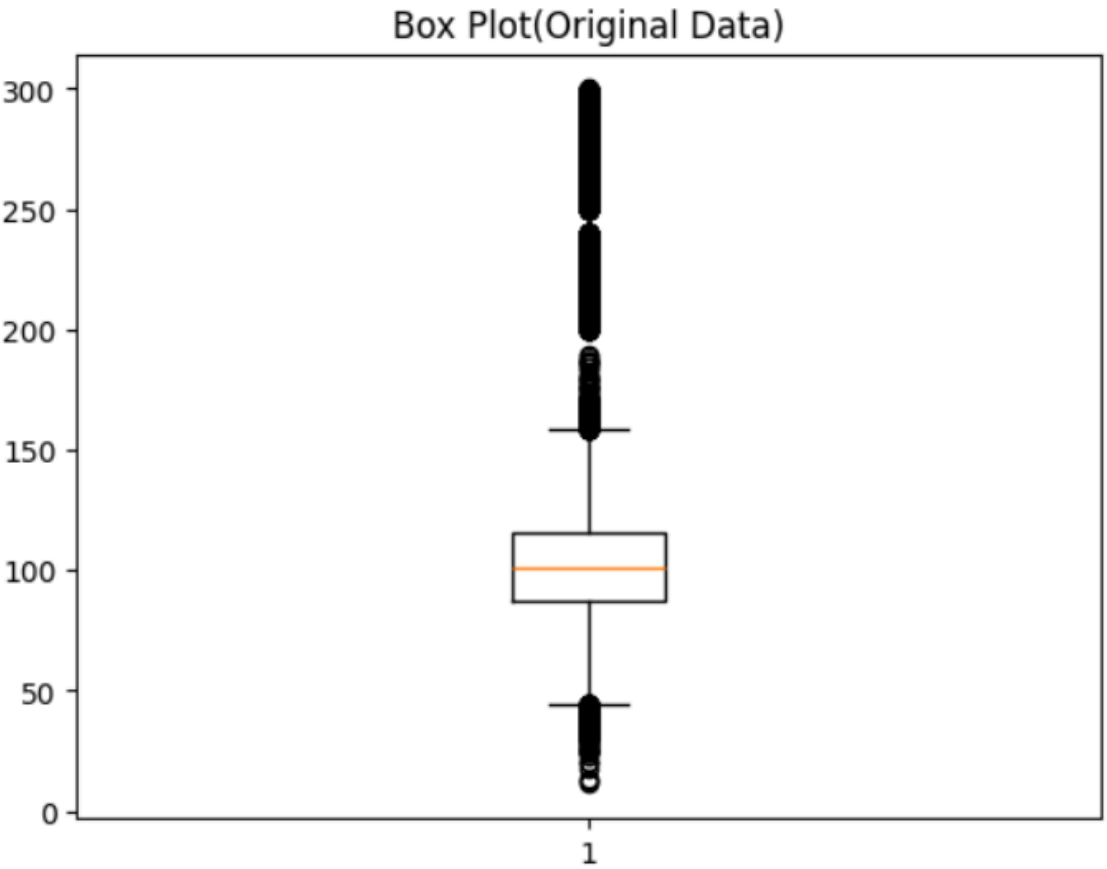
Histogram of Original Data :

The histogram shows the overall shape of the data distribution. Outliers appear as isolated bars far from the main cluster.



Boxplot of Original Data :

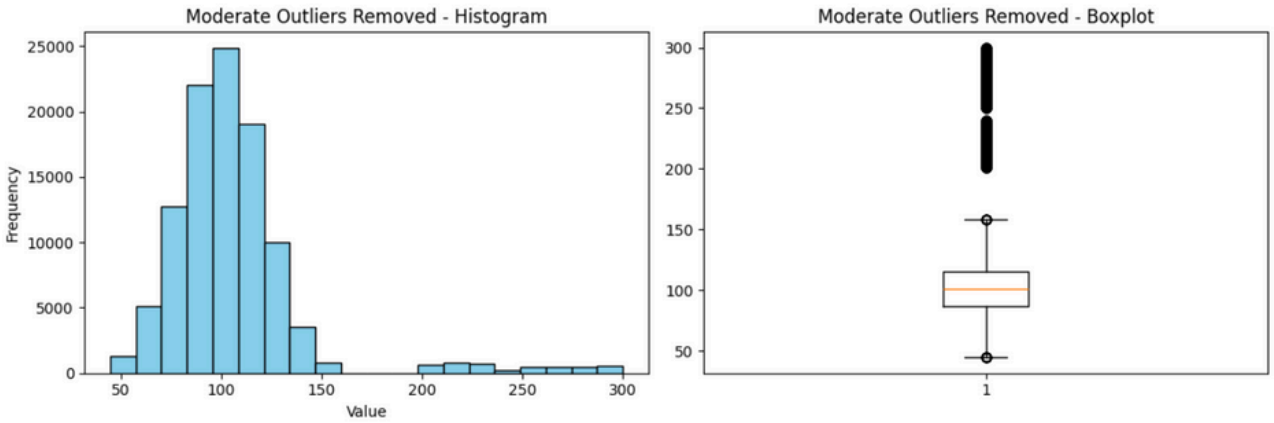
The box plot highlights outliers as points beyond the whiskers. Removing outliers tightens the distribution and shortens whiskers



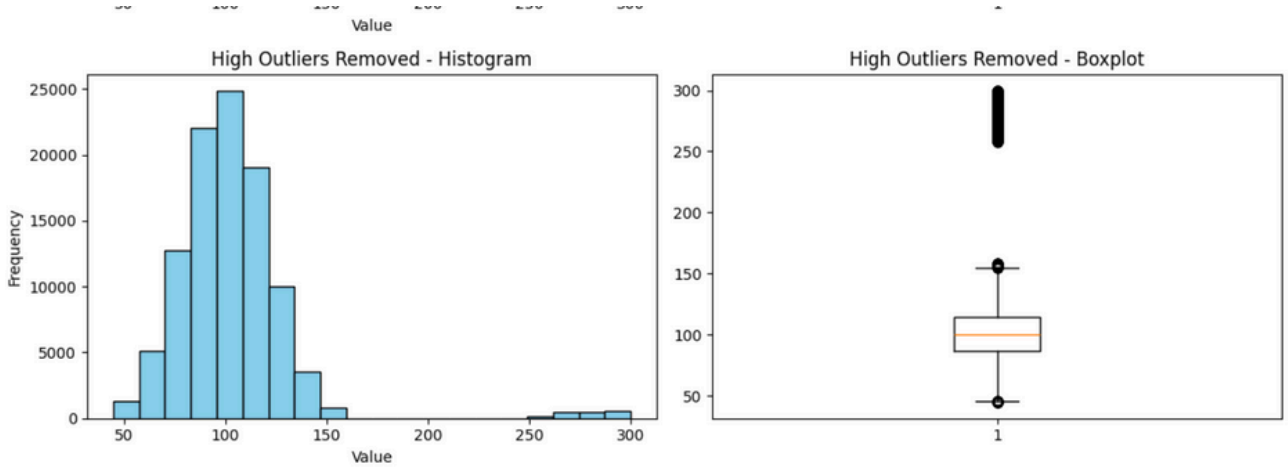
Histograms and Boxplots After Each Outlier Removal Level :

Demonstrates how data distribution becomes tighter when Moderate, High, and Extreme outliers are removed progressively

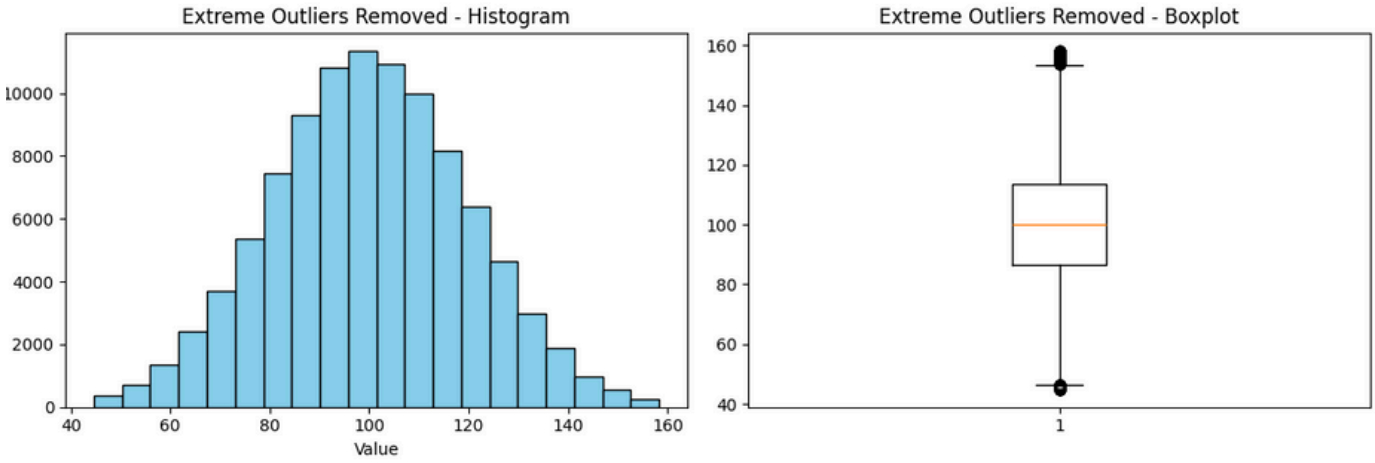
Moderate outlier Removed :



High outlier Removed :



Extreme outlier Removed :



Line Plot of Mean, Median, and Standard Deviation :

This plot shows how progressively removing Moderate → High → Extreme outliers reduces the effect of extreme values on mean and standard deviation.

The mean and standard deviation decrease after removing High and Extreme outliers, indicating that extreme values had a significant impact on overall dispersion. The median remains relatively stable, showing robustness to outliers.