Data Analysis Report

# 1. Null Values Analysis

No null values found.

# 2. Normality Assessment

Non-normal features: 385

Top 10: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9

Normal features: 63

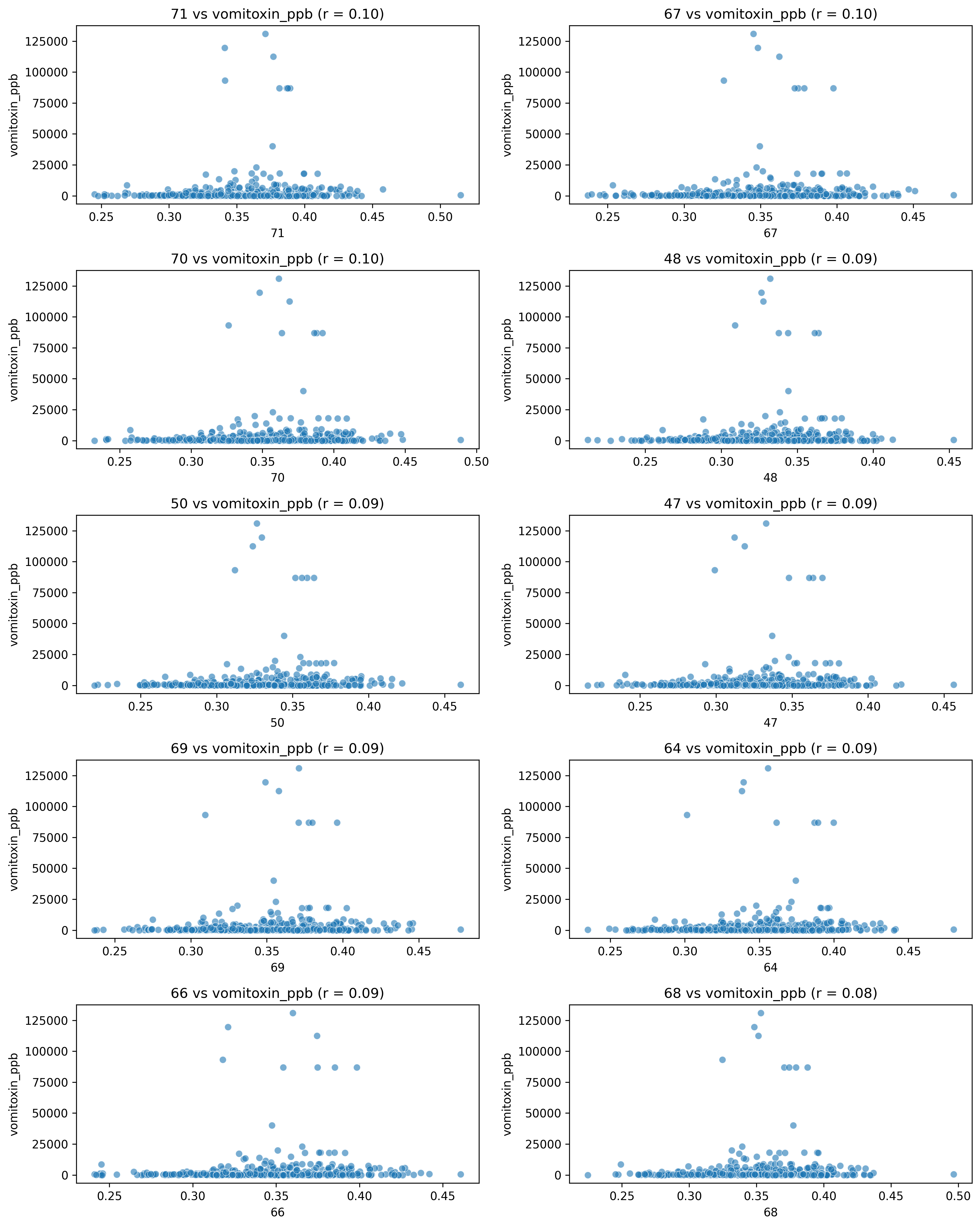
Top 10: 26, 28, 29, 30, 31, 32, 33, 34, 35, 36

# 3. Correlation Analysis

Top 10 Feature-Target Correlations:

|  |  |  |
| --- | --- | --- |
| Feature | Correlation | Distribution |
| 71 | 0.10 | Normal |
| 67 | 0.10 | Normal |
| 70 | 0.10 | Normal |
| 48 | 0.09 | Normal |
| 50 | 0.09 | Normal |
| 47 | 0.09 | Normal |
| 69 | 0.09 | Normal |
| 64 | 0.09 | Normal |
| 66 | 0.09 | Normal |
| 68 | 0.08 | Normal |

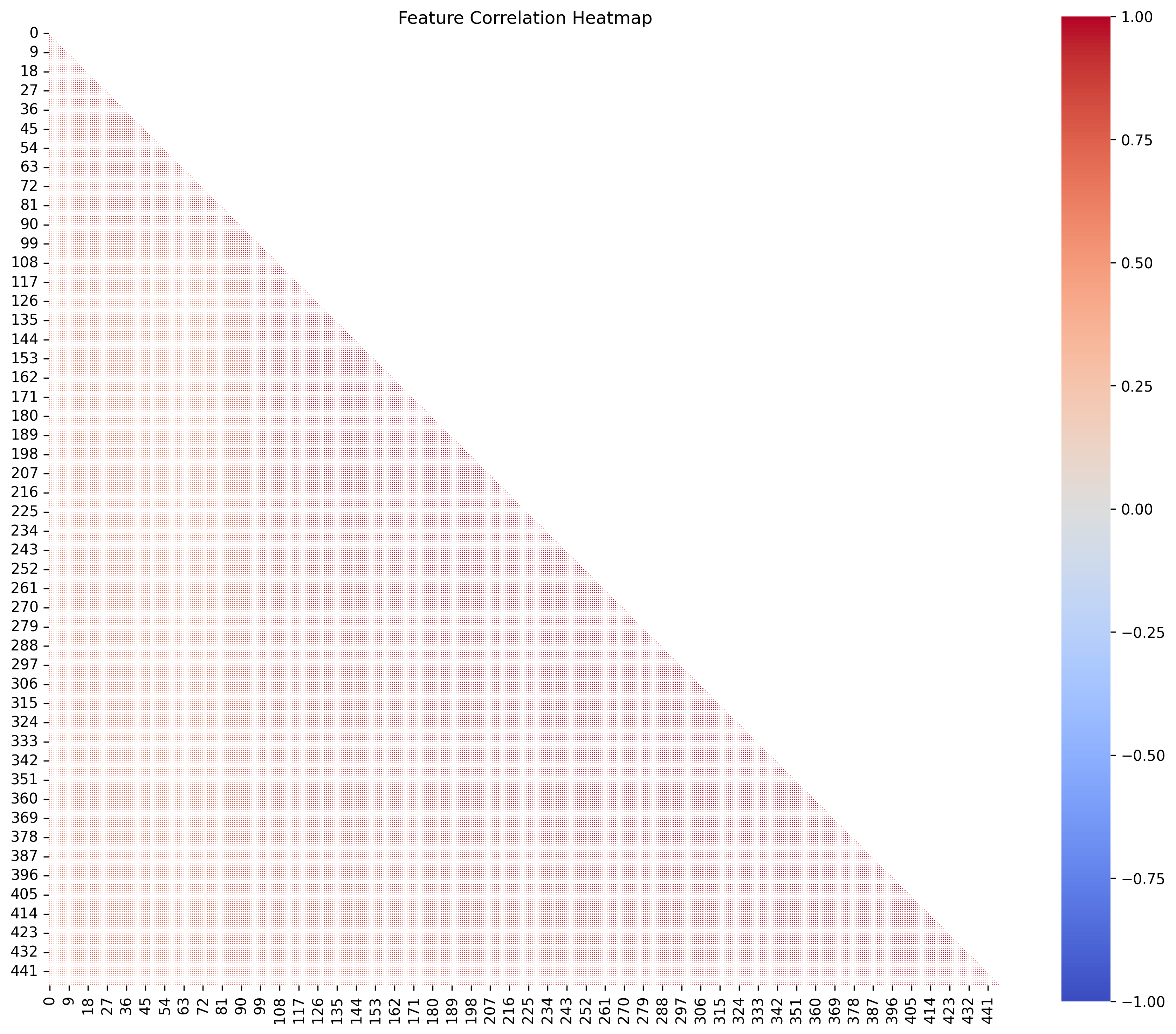
Top Features vs Target:



Top 10 Feature-Feature Correlations:

|  |  |  |
| --- | --- | --- |
| Feature 1 | Feature 2 | Correlation |
| 259 | 263 | 0.98 |
| 294 | 311 | 0.98 |
| 254 | 255 | 0.98 |
| 291 | 304 | 0.98 |
| 243 | 254 | 0.98 |
| 244 | 257 | 0.98 |
| 271 | 283 | 0.98 |
| 298 | 308 | 0.98 |
| 224 | 226 | 0.98 |
| 274 | 278 | 0.98 |

Feature Correlation Heatmap:



# 4. Skewness Analysis

Skewed features: 0

# 5. Outlier Analysis

|  |  |  |
| --- | --- | --- |
| Feature | Outliers | Outlier % |
| 100 | 17 | 3.40% |
| 97 | 16 | 3.20% |
| 144 | 16 | 3.20% |
| 419 | 15 | 3.00% |
| 135 | 15 | 3.00% |
| 143 | 15 | 3.00% |
| 422 | 14 | 2.80% |
| 407 | 14 | 2.80% |
| 111 | 14 | 2.80% |
| 147 | 14 | 2.80% |

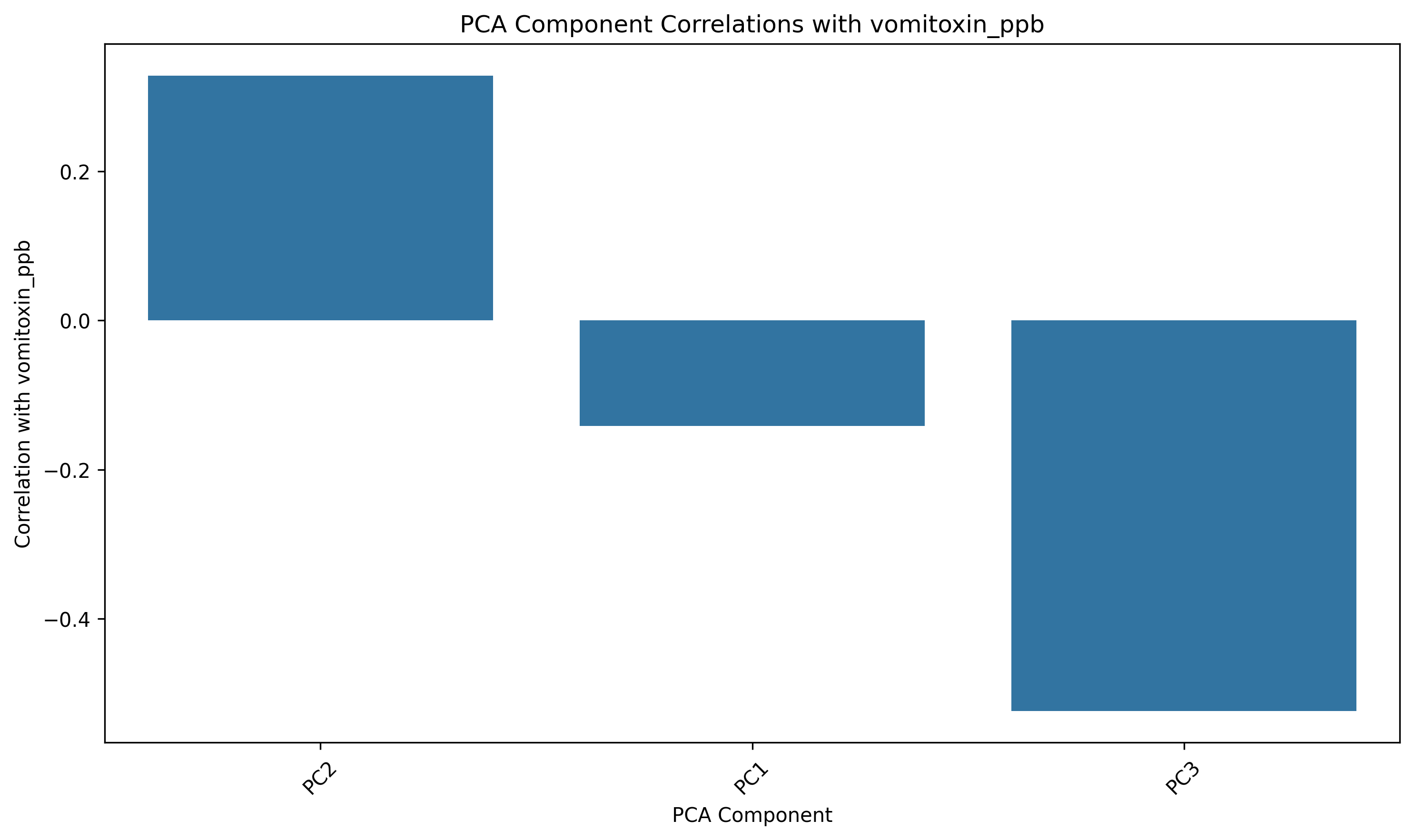
# 6. PCA Analysis

Components for 95.0% variance: 3

PCA Component Correlation with Target:

|  |  |
| --- | --- |
| Component | Correlation with Target |
| PC2 | 0.328 |
| PC1 | -0.142 |
| PC3 | -0.523 |

PCA Component Correlations with Target:



Component 1 Top Loadings:

|  |  |
| --- | --- |
| Feature | Loading |
| 278 | 0.050 |
| 295 | 0.050 |
| 271 | 0.050 |
| 309 | 0.050 |
| 286 | 0.050 |
| 269 | 0.050 |
| 294 | 0.050 |
| 314 | 0.050 |
| 291 | 0.050 |
| 277 | 0.050 |

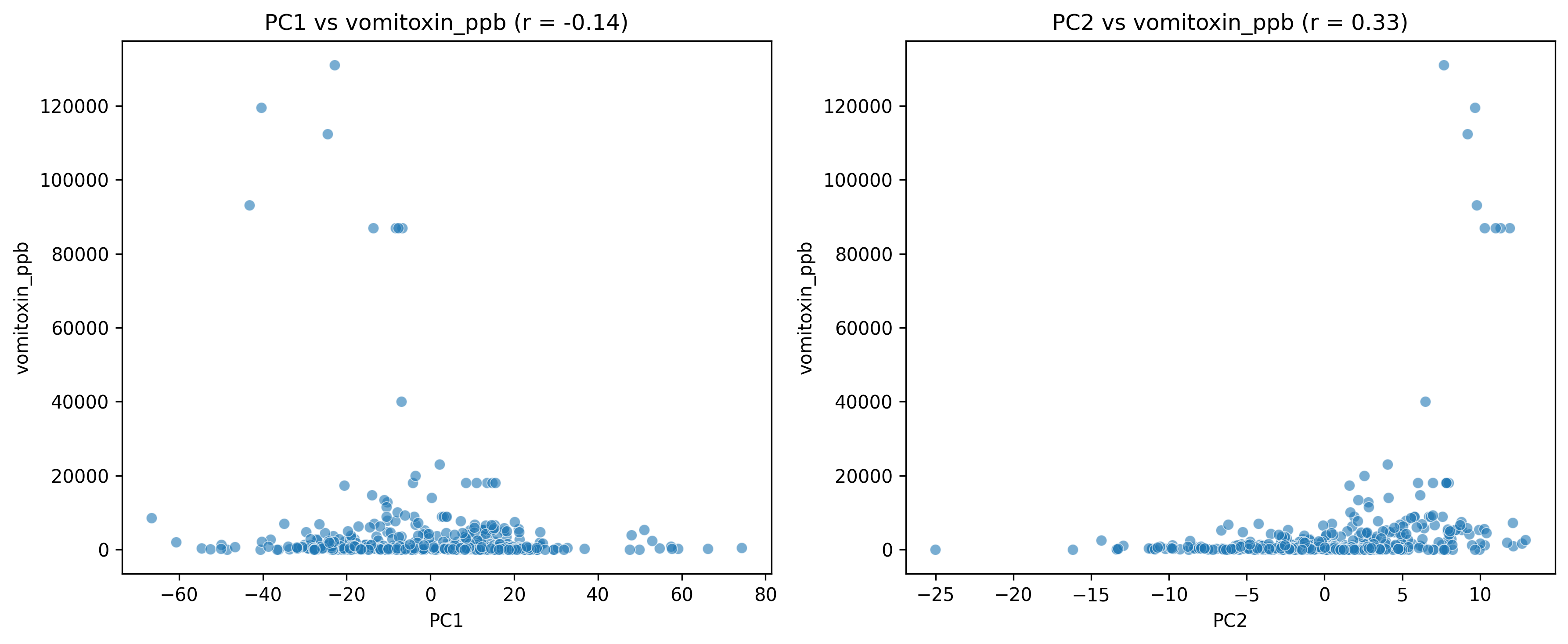
Component 2 Top Loadings:

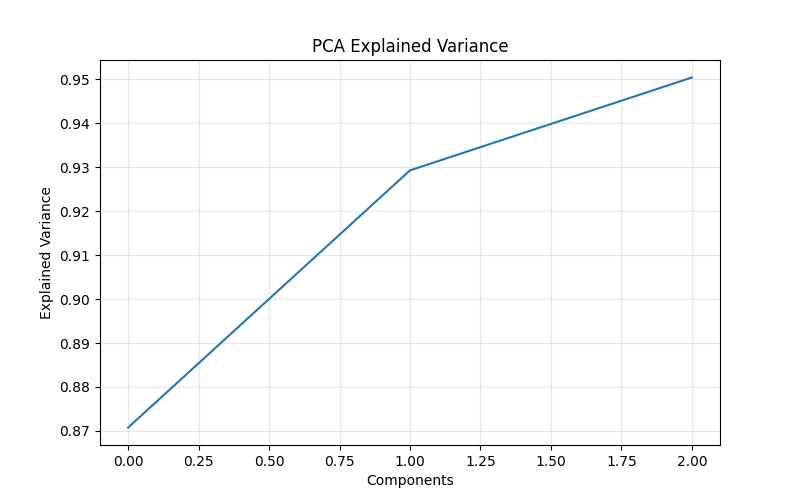
|  |  |
| --- | --- |
| Feature | Loading |
| 52 | 0.112 |
| 47 | 0.111 |
| 41 | 0.111 |
| 48 | 0.110 |
| 50 | 0.109 |
| 43 | 0.109 |
| 49 | 0.109 |
| 46 | 0.109 |
| 54 | 0.108 |
| 42 | 0.108 |

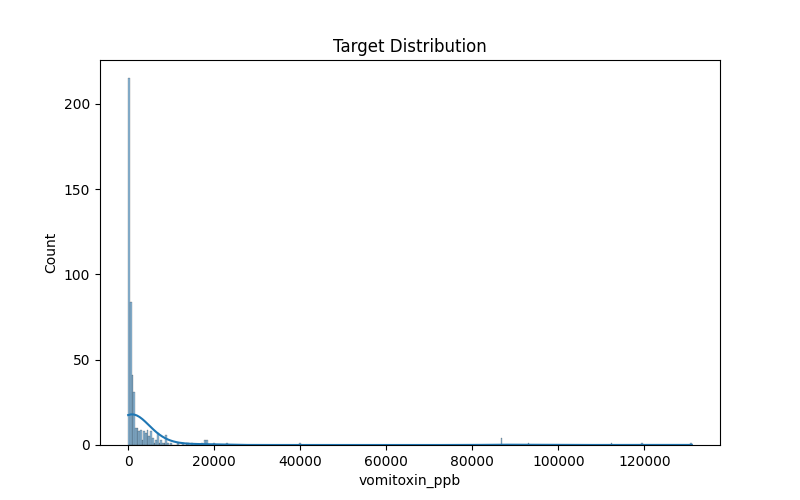
Component 3 Top Loadings:

|  |  |
| --- | --- |
| Feature | Loading |
| 104 | 0.089 |
| 108 | 0.087 |
| 110 | 0.087 |
| 107 | 0.086 |
| 111 | 0.085 |
| 106 | 0.085 |
| 112 | 0.085 |
| 114 | 0.085 |
| 120 | 0.084 |
| 109 | 0.084 |

Top PCA Components vs Target:







# 7. Target Variable Analysis

Target Variable Statistics:

|  |  |
| --- | --- |
| Statistic | Value |
| Count | 500 |
| Mean | 3410.0060 |
| Median | 500.0000 |
| Std | 13095.8035 |
| Min | 0.0000 |
| Max | 131000.0000 |

Target Variable Outliers:

Outlier count: 80

Outlier percentage: 16.00%

Lower bound: -2206.2500

Upper bound: 4043.7500

Target Variable Normality:

Shapiro-Wilk p-value: 0.0000

Distribution: Non-normal

Skewness: 7.2252

Kurtosis: 55.5447

