Data Analysis Report

# 1. Null Values Analysis

No null values found.

# 2. Normality Assessment

Non-normal features: 76

Top 10: 293, 325, 203, 20, 310, 184, 175, 250, 304, 9

Normal features: 12

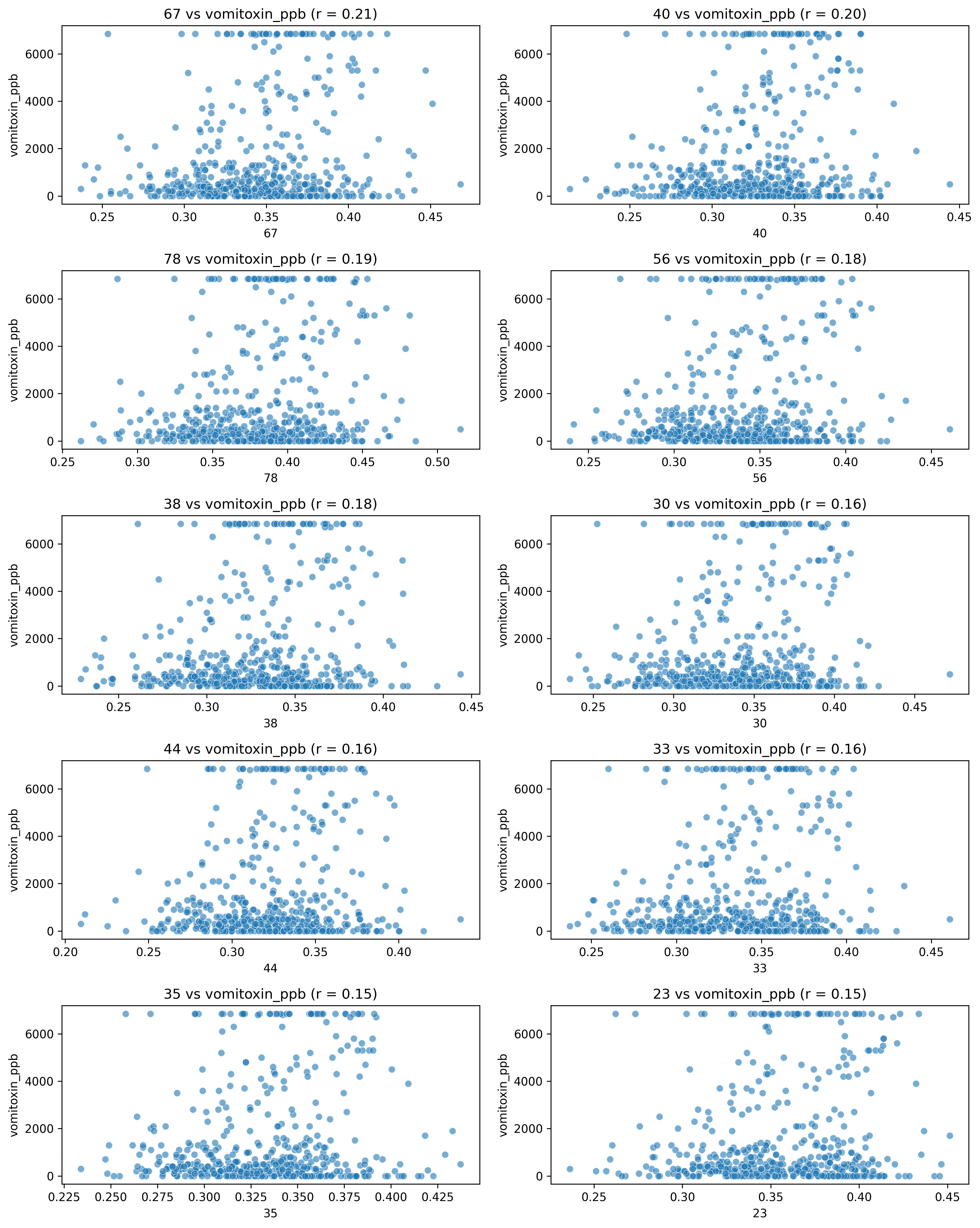
Top 10: 40, 30, 38, 67, 56, 33, 26, 96, 78, 44

# 3. Correlation Analysis

Top 10 Feature-Target Correlations:

|  |  |  |
| --- | --- | --- |
| Feature | Correlation | Distribution |
| 67 | 0.21 | Normal |
| 40 | 0.20 | Normal |
| 78 | 0.19 | Normal |
| 56 | 0.18 | Normal |
| 38 | 0.18 | Normal |
| 30 | 0.16 | Normal |
| 44 | 0.16 | Normal |
| 33 | 0.16 | Normal |
| 35 | 0.15 | Normal |
| 23 | 0.15 | Non-normal |

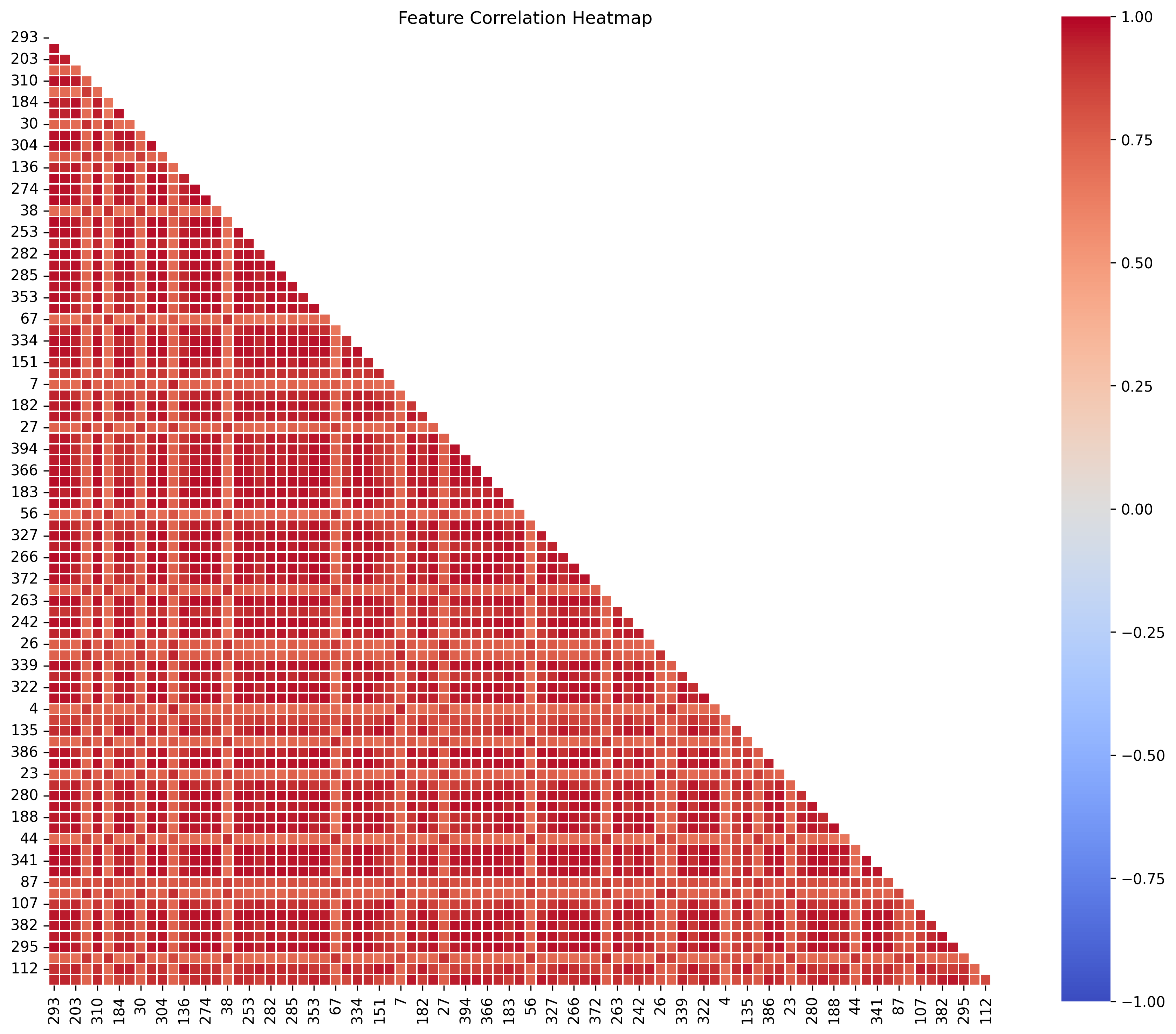
Top Features vs Target:



Top 10 Feature-Feature Correlations:

|  |  |  |
| --- | --- | --- |
| Feature 1 | Feature 2 | Correlation |
| 278 | 274 | 0.98 |
| 278 | 290 | 0.98 |
| 325 | 327 | 0.98 |
| 293 | 278 | 0.98 |
| 250 | 240 | 0.98 |
| 274 | 290 | 0.98 |
| 278 | 308 | 0.98 |
| 290 | 295 | 0.98 |
| 265 | 260 | 0.98 |
| 260 | 271 | 0.98 |

Feature Correlation Heatmap:



# 4. Skewness Analysis

Skewed features: 0

# 5. Outlier Analysis

|  |  |  |
| --- | --- | --- |
| Feature | Outliers | Outlier % |
| 135 | 15 | 3.00% |
| 111 | 14 | 2.80% |
| 161 | 14 | 2.80% |
| 147 | 14 | 2.80% |
| 107 | 14 | 2.80% |
| 184 | 13 | 2.60% |
| 136 | 13 | 2.60% |
| 386 | 12 | 2.40% |
| 412 | 12 | 2.40% |
| 138 | 12 | 2.40% |

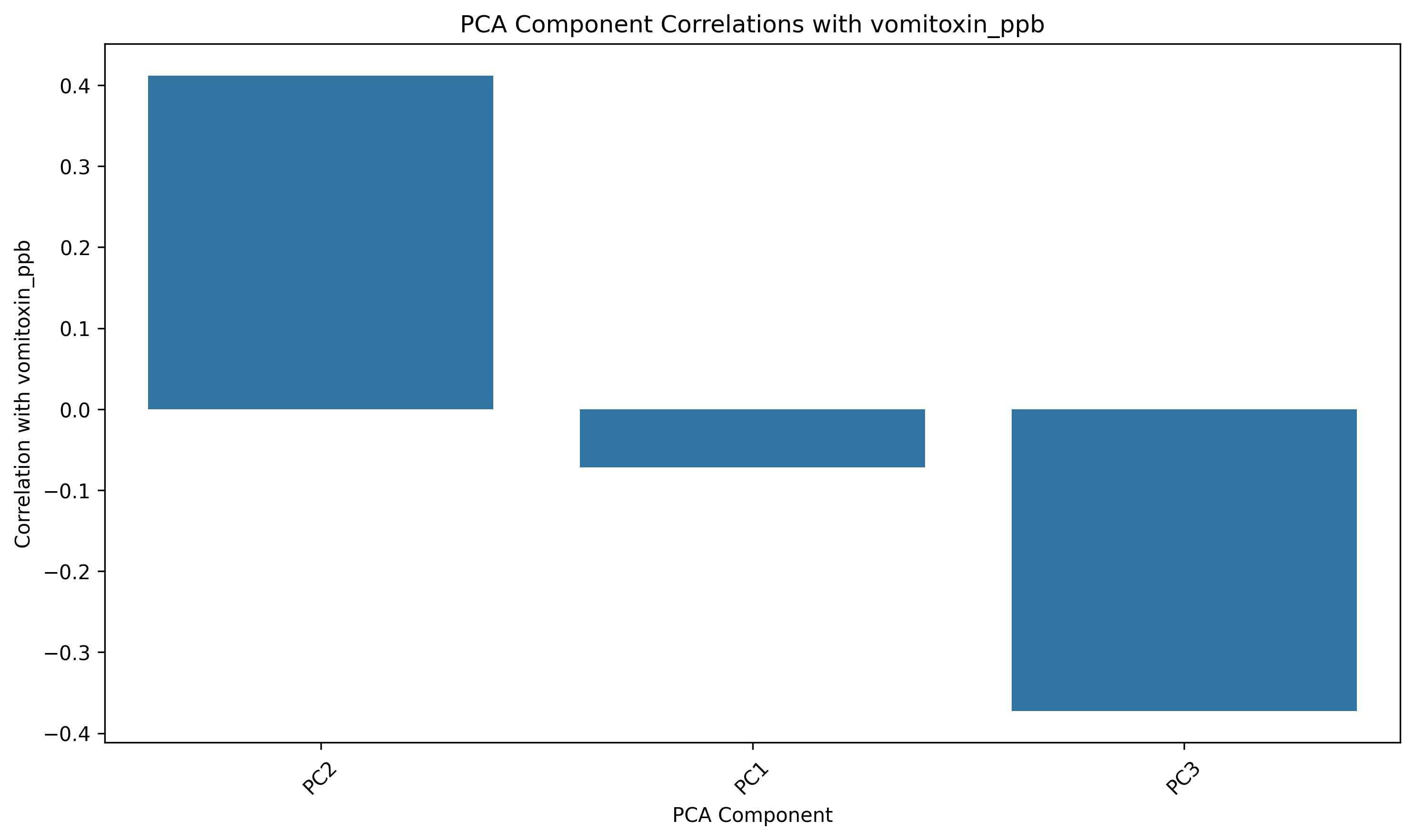
# 6. PCA Analysis

Components for 95.0% variance: 3

PCA Component Correlation with Target:

|  |  |
| --- | --- |
| Component | Correlation with Target |
| PC2 | 0.412 |
| PC1 | -0.072 |
| PC3 | -0.372 |

PCA Component Correlations with Target:



Component 1 Top Loadings:

|  |  |
| --- | --- |
| Feature | Loading |
| 278 | 0.112 |
| 295 | 0.112 |
| 271 | 0.112 |
| 308 | 0.112 |
| 274 | 0.112 |
| 310 | 0.112 |
| 290 | 0.112 |
| 262 | 0.112 |
| 280 | 0.112 |
| 293 | 0.112 |

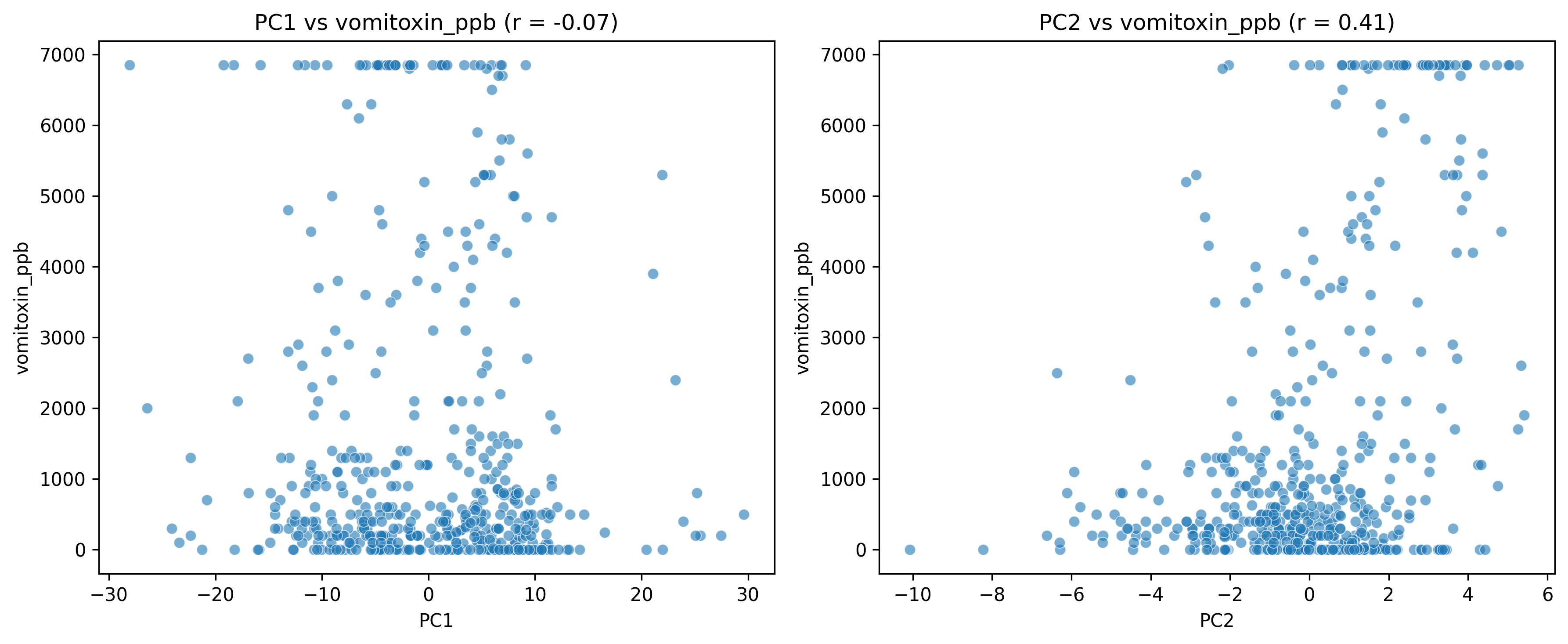
Component 2 Top Loadings:

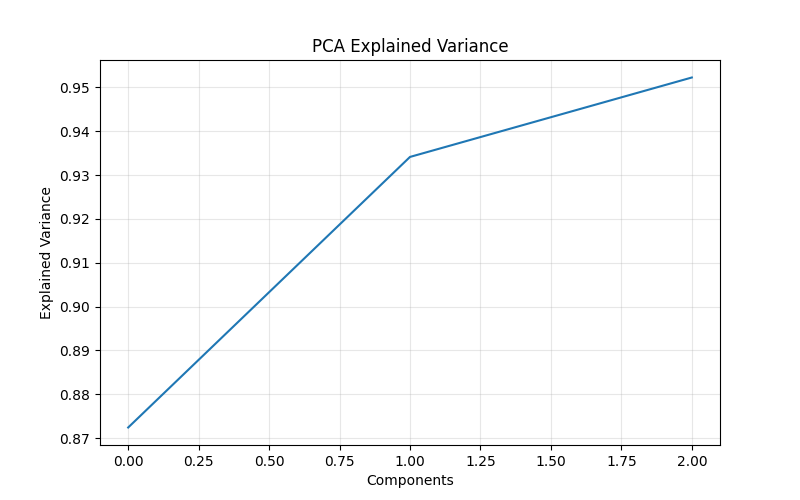
|  |  |
| --- | --- |
| Feature | Loading |
| 40 | 0.232 |
| 38 | 0.230 |
| 30 | 0.223 |
| 44 | 0.223 |
| 35 | 0.222 |
| 33 | 0.219 |
| 23 | 0.218 |
| 56 | 0.217 |
| 67 | 0.217 |
| 20 | 0.215 |

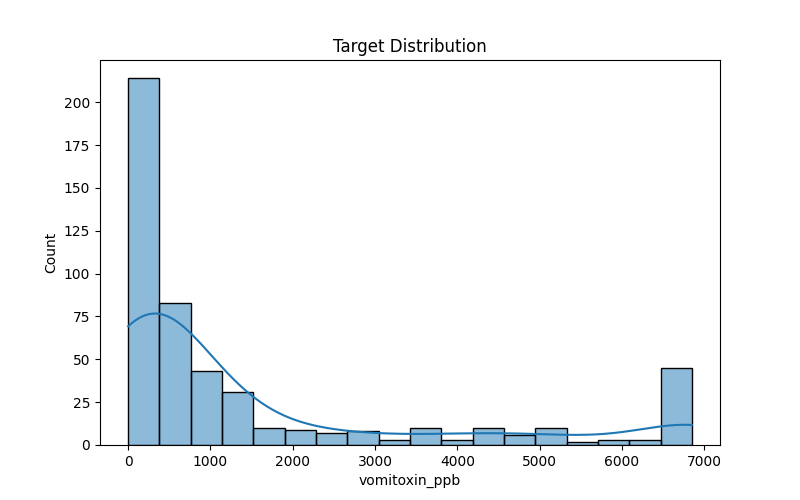
Component 3 Top Loadings:

|  |  |
| --- | --- |
| Feature | Loading |
| 107 | 0.218 |
| 111 | 0.213 |
| 103 | 0.213 |
| 112 | 0.211 |
| 117 | 0.210 |
| 96 | 0.207 |
| 135 | 0.193 |
| 436 | -0.190 |
| 147 | 0.180 |
| 136 | 0.176 |

Top PCA Components vs Target:







# 7. Target Variable Analysis

Target Variable Statistics:

|  |  |
| --- | --- |
| Statistic | Value |
| Count | 500 |
| Mean | 1520.4060 |
| Median | 500.0000 |
| Std | 2157.6483 |
| Min | 0.0000 |
| Max | 6850.0000 |

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: 1.5691

Kurtosis: 1.0395

