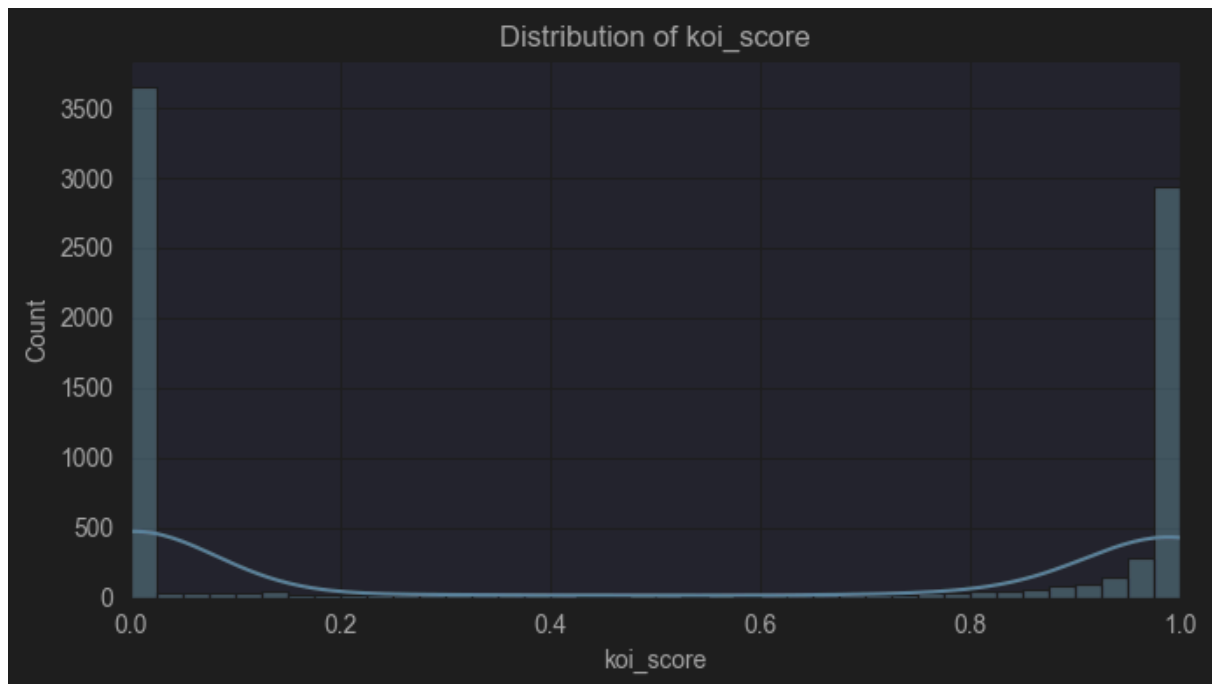# RESULTS

## koi_score Distribution



**Summary statistics:**

```
count    7994.000000
mean        0.483829
std         0.477009
min         0.000000
25%         0.000000
50%         0.371000
75%         0.999000
max         1.000000
```

# Exploratory Data Analysis

## Feature Transformations

**Log10 transformation:**

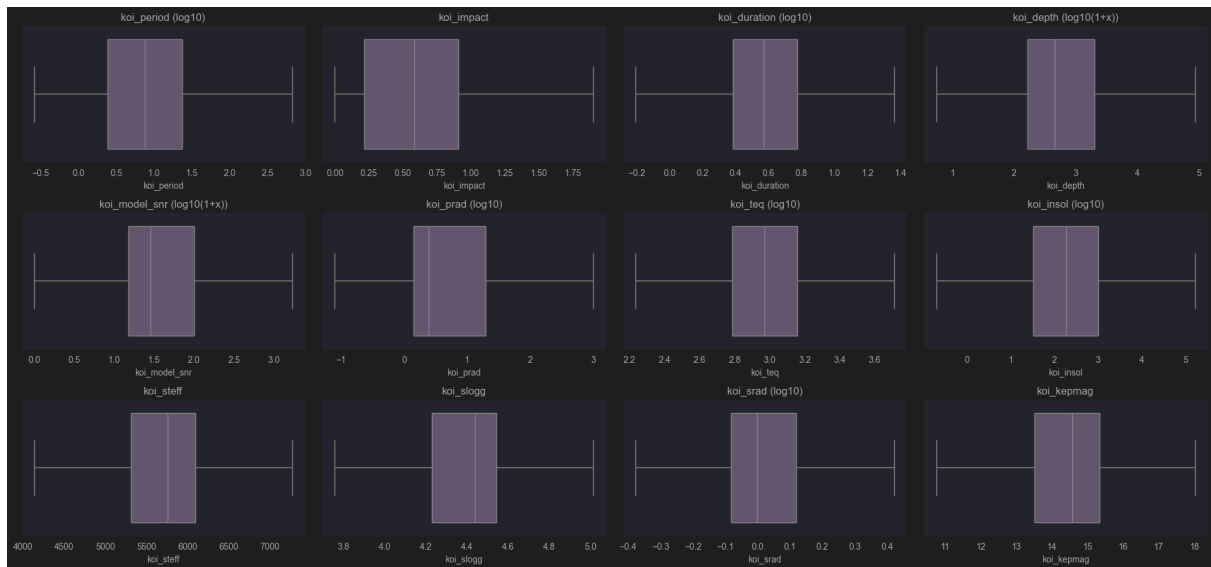- koi_period , koi_duration , koi_prad , koi_teq , koi_insol , koi_srad

**Log10(1+x) transformation:**

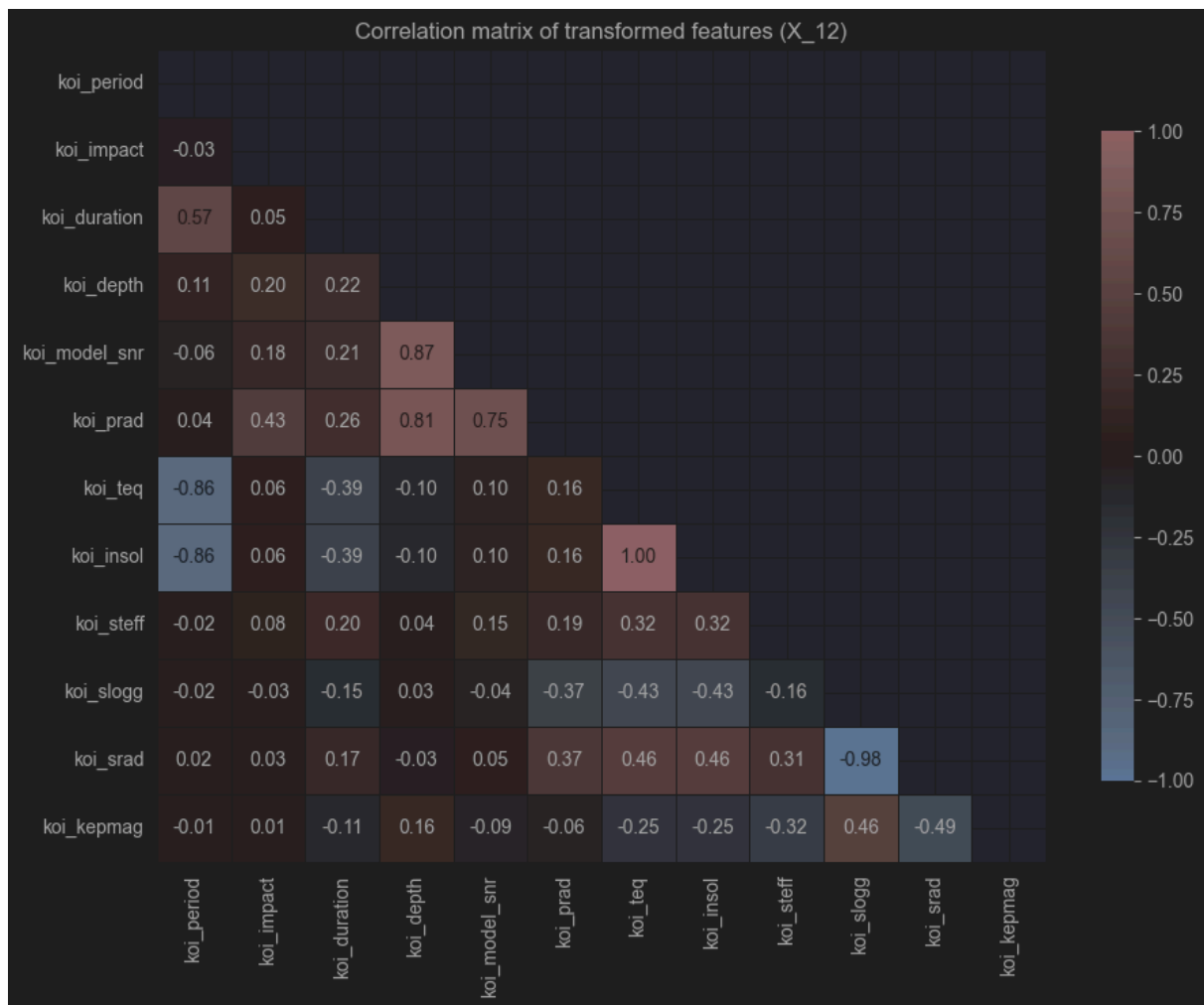- koi_depth , koi_model_snr

**Linear (no transformation):**

- koi_impact , koi_steff , koi_slogg , koi_kepmag

## Boxplots



# Correlation Analysis

## Pre-PCA Correlation Matrix (X_12)

Correlation matrix of transformed features (X_12)

# Principal Component Analysis

## Eigenvalue Decomposition

**Explained variance by component:**

|     | eigenvalue | explained_% | cumulative_% |
|-----|-----------|-------------|--------------|
| PC1 | 3.737318  | 31.14       | 31.14        |
| PC2 | 2.971272  | 24.76       | 55.90        |
| PC3 | 2.111366  | 17.59       | 73.49        |
| PC4 | 0.994332  | 8.29        | 81.78        |
| PC5 | 0.926532  | 7.72        | 89.50        |
| PC6 | 0.646095  | 5.38        | 94.88        |
| PC7 | 0.439769  | 3.66        | 98.54        |
| PC8 | 0.115413  | 0.96        | 99.51        |

| | | | |
|---|---|---|---|
| PC9 | 0.052262 | 0.44 | 99.94 |
| PC10 | 0.006205 | 0.05 | 99.99 |
| PC11 | 0.000935 | 0.01 | 100.00 |
| PC12 | 0.000001 | 0.00 | 100.00 |

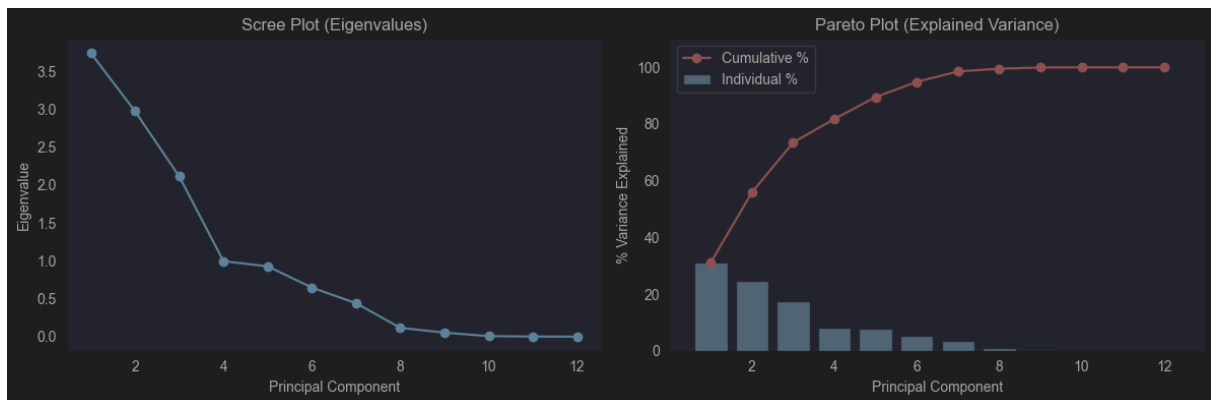## Loading Matrix

**Eigenvector matrix A (feature contributions to each PC):**

```
              PC1    PC2    PC3    PC4    PC5    PC6    PC7    PC8    PC9
PC10   PC11   PC12
koi_period   -0.3076 0.3040 0.3926 -0.0245 0.0357 0.0068 -0.3113 -0.11
37 -0.0458 0.0357 0.7374 0.0002
koi_impact    0.0844 0.1888 -0.1620 -0.0895 0.9193 -0.1488 0.0617 -0.20
20 0.1055 0.0048 -0.0012 0.0000
koi_duration -0.0884 0.3615 0.3099 0.2177 0.0183 0.3202 0.7762 0.08
13 0.0773 0.0070 -0.0028 -0.0000
koi_depth     0.0460 0.4749 -0.3316 -0.0401 -0.2271 -0.0109 -0.1465 0.07
52 0.7623 0.0295 0.0055 0.0000
koi_model_snr 0.1457 0.4288 -0.3093 0.1115 -0.2771 -0.2171 0.1108 -0.62
00 -0.4074 -0.0104 -0.0058 -0.0000
koi_prad      0.2270 0.4680 -0.1662 -0.1551 0.0555 0.0547 -0.1302 0.666
5 -0.4587 -0.0122 0.0018 -0.0000
koi_teq       0.4657 -0.2013 -0.1526 0.0500 -0.0182 0.1100 0.1436 0.0120
0.0520 0.0277 0.4246 -0.7070
koi_insol     0.4657 -0.2013 -0.1526 0.0500 -0.0182 0.1100 0.1436 0.0120
0.0520 0.0272 0.4243 0.7072
koi_steff     0.2240 0.0909 0.1336 0.7823 0.1262 0.2887 -0.4073 -0.038
6 0.0026 -0.1279 -0.1695 -0.0001
koi_slogg    -0.3607 -0.1033 -0.3821 0.3969 0.0476 -0.0989 0.0720 0.16
99 -0.0774 0.6985 0.1254 0.0003
koi_srad      0.3790 0.1088 0.3942 -0.2615 -0.0238 0.1555 -0.1464 -0.176
0 0.0259 0.7011 -0.2269 0.0002
koi_kepmag   -0.2408 -0.0316 -0.3529 -0.2510 0.0495 0.8268 -0.1110 -0.
2091 -0.1060 -0.0145 -0.0029 0.0000
```
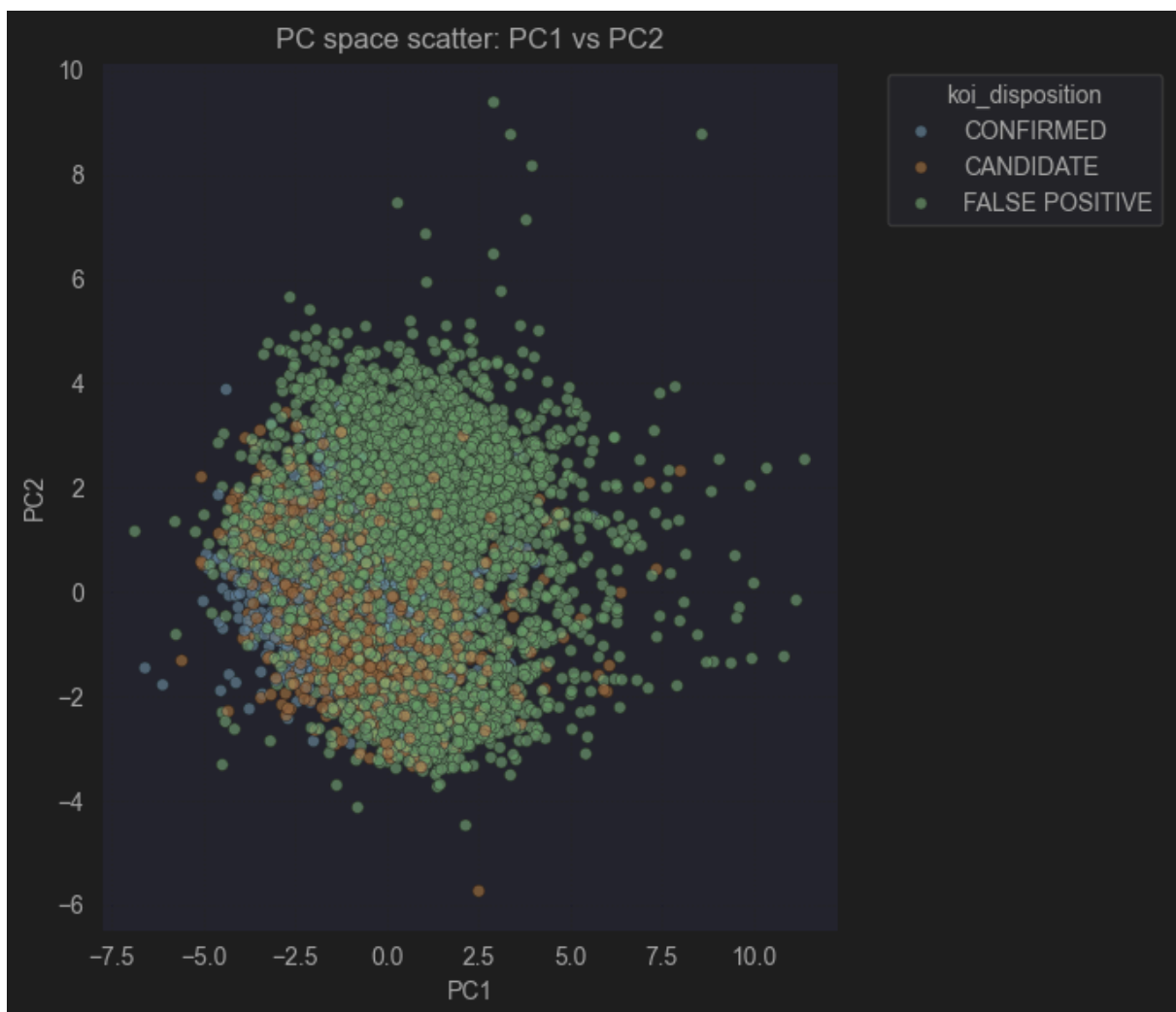
## Visualization

# Scree Plot



# Pareto Plot (Cumulative Variance)



# Loadings Heatmap

Loadings (A): weights of original features on PC1..PC5

| | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| koi_period | -0.31 | 0.30 | 0.39 | -0.02 | 0.04 |
| koi_impact | 0.08 | 0.19 | -0.16 | -0.09 | 0.92 |
| koi_duration | -0.09 | 0.36 | 0.31 | 0.22 | 0.02 |
| koi_depth | 0.05 | 0.47 | -0.33 | -0.04 | -0.23 |
| koi_model_snr | 0.15 | 0.43 | -0.31 | 0.11 | -0.28 |
| koi_prad | 0.23 | 0.47 | -0.17 | -0.16 | 0.06 |
| koi_teq | 0.47 | -0.20 | -0.15 | 0.05 | -0.02 |
| koi_insol | 0.47 | -0.20 | -0.15 | 0.05 | -0.02 |
| koi_steff | 0.22 | 0.09 | 0.13 | 0.78 | 0.13 |
| koi_slogg | -0.36 | -0.10 | -0.38 | 0.40 | 0.05 |
| koi_srad | 0.38 | 0.11 | 0.39 | -0.26 | -0.02 |
| koi_kepmag | -0.24 | -0.03 | -0.35 | -0.25 | 0.05 |

# Top Contributors by Component

**PC1** (31.14% variance): koi_teq , koi_insol , koi_srad , koi_slogg , koi_period

**PC2** (24.76% variance): koi_depth , koi_prad , koi_model_snr , koi_duration , koi_period

**PC3** (17.59% variance): koi_srad , koi_period , koi_slogg , koi_kepmag , koi_depth

# Feature Spaces Summary

- **X_std** (standardized features): 7,994 × 12

- **X_pca5** (first 5 PCs, 89.50% variance): 7,994 × 5