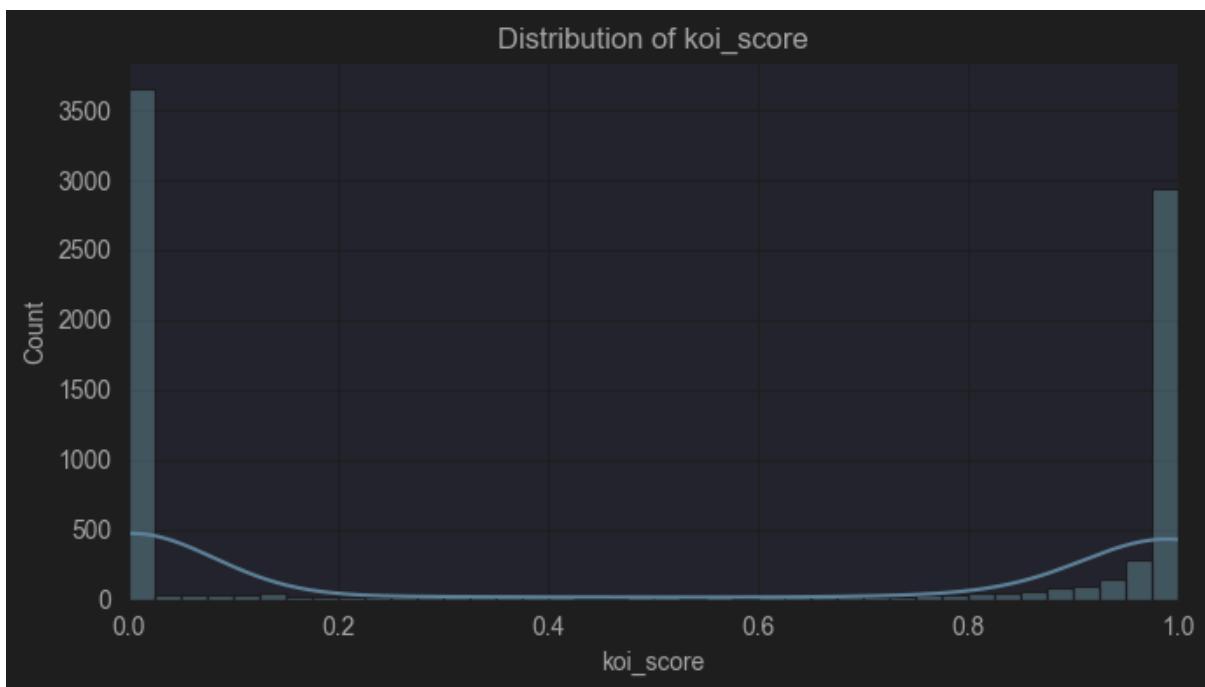


RESULTS

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`

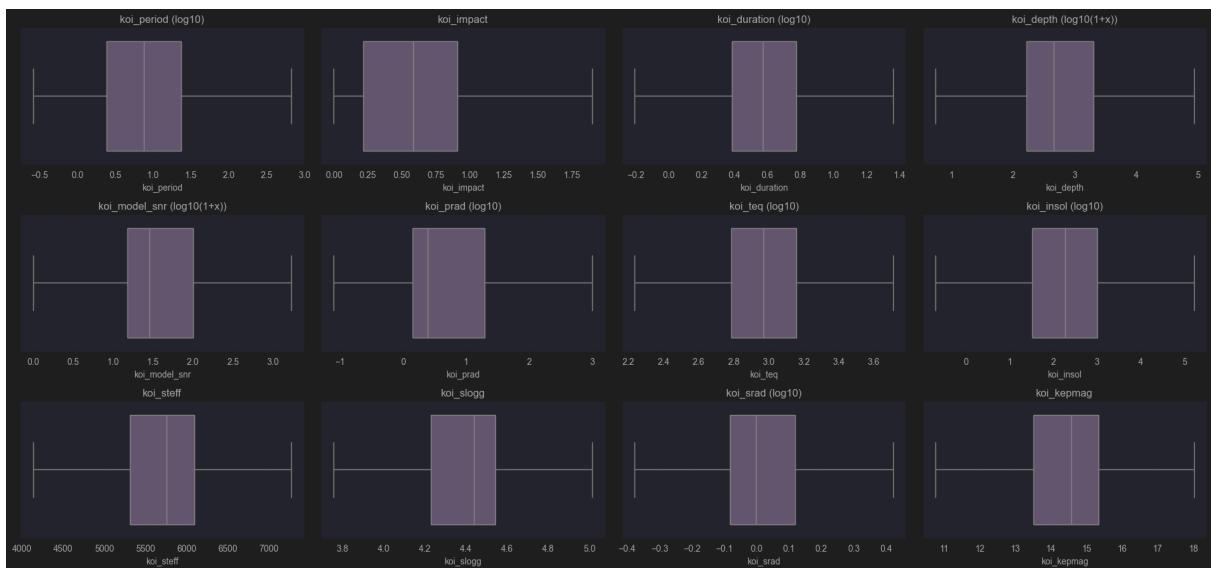
Log10(1+x) transformation:

- `koi_depth`, `koi_model_snr`

Linear (no transformation):

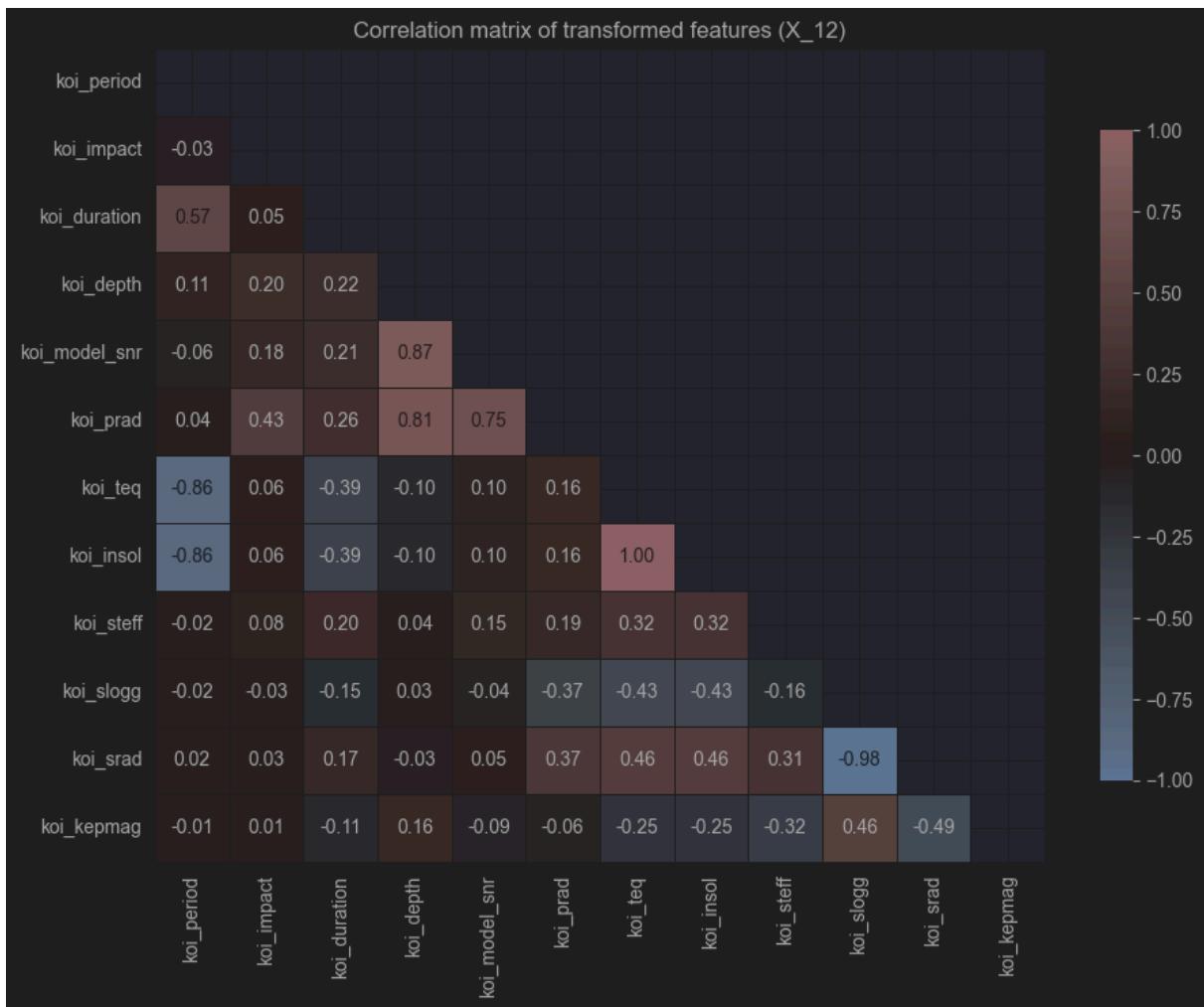
- `koi_impact`, `koi_steff`, `koi_slogg`, `koi_kepmag`

Boxplots



Correlation Analysis

Pre-PCA Correlation Matrix (X_12)



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

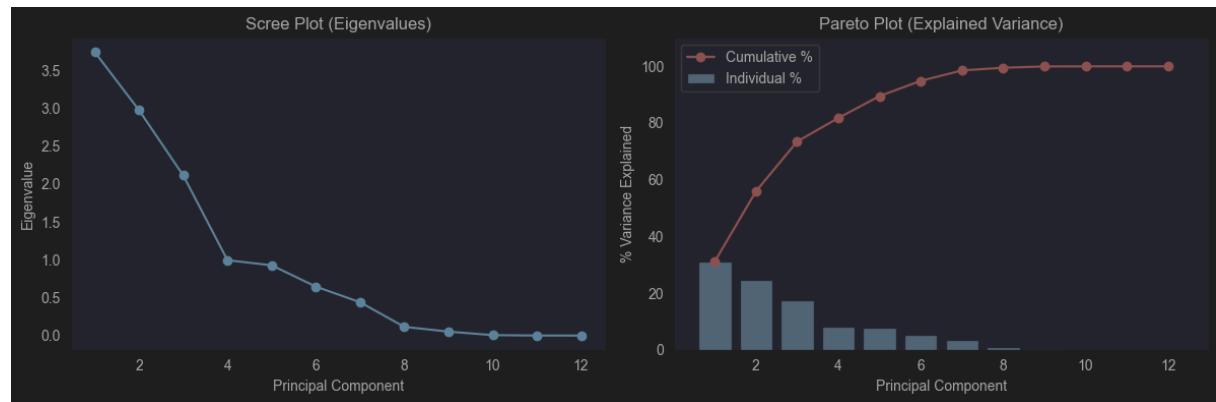
Eigenvector matrix A (feature contributions to each PC):

$$\mathbf{A} = \begin{bmatrix} -0.31 & 0.30 & 0.39 & -0.02 & 0.04 & 0.01 & -0.31 & -0.11 & -0.05 & 0.04 & 0.74 & 0.00 \\ 0.08 & 0.19 & -0.16 & -0.09 & 0.92 & -0.15 & 0.06 & -0.20 & 0.11 & 0.00 & 0.00 & 0.00 \\ -0.09 & 0.36 & 0.31 & 0.22 & 0.02 & 0.32 & 0.78 & 0.08 & 0.08 & 0.01 & 0.00 & 0.00 \\ 0.05 & 0.47 & -0.33 & -0.04 & -0.23 & -0.01 & -0.15 & 0.08 & 0.76 & 0.03 & 0.01 & 0.00 \\ 0.15 & 0.43 & -0.31 & 0.11 & -0.28 & -0.22 & 0.11 & -0.62 & -0.41 & -0.01 & -0.01 & 0.00 \\ 0.23 & 0.47 & -0.17 & -0.16 & 0.06 & 0.05 & -0.13 & 0.67 & -0.46 & -0.01 & 0.00 & 0.00 \\ 0.47 & -0.20 & -0.15 & 0.05 & -0.02 & 0.11 & 0.14 & 0.01 & 0.05 & 0.03 & 0.42 & -0.71 \\ 0.47 & -0.20 & -0.15 & 0.05 & -0.02 & 0.11 & 0.14 & 0.01 & 0.05 & 0.03 & 0.42 & 0.71 \\ 0.22 & 0.09 & 0.13 & 0.78 & 0.13 & 0.29 & -0.41 & -0.04 & 0.00 & -0.13 & -0.17 & 0.00 \\ -0.36 & -0.10 & -0.38 & 0.40 & 0.05 & -0.10 & 0.07 & 0.17 & -0.08 & 0.70 & 0.13 & 0.00 \\ 0.38 & 0.11 & 0.39 & -0.26 & -0.02 & 0.16 & -0.15 & -0.18 & 0.03 & 0.70 & -0.23 & 0.00 \\ -0.24 & -0.03 & -0.35 & -0.25 & 0.05 & 0.83 & -0.11 & -0.21 & -0.11 & -0.01 & 0.00 & 0.00 \end{bmatrix}_{12 \times 12}$$

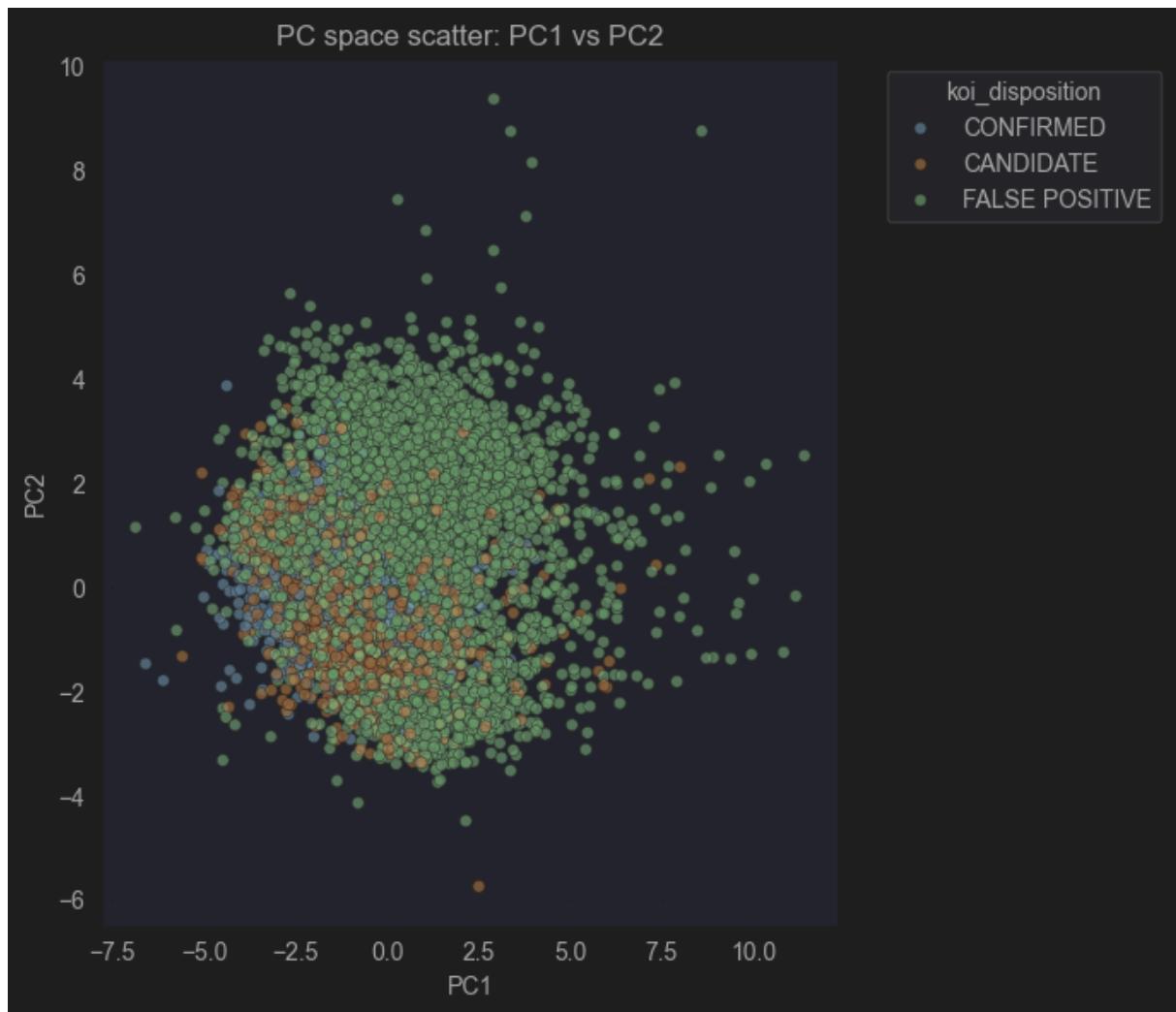
where rows represent features (koi_period, koi_impact, koi_duration, koi_depth, koi_model_snr, koi_prad, koi_teq, koi_insol, koi_steff, koi_slogg, koi_srad, koi_kepmag) and columns represent principal components (PC1 through PC12).

Visualization

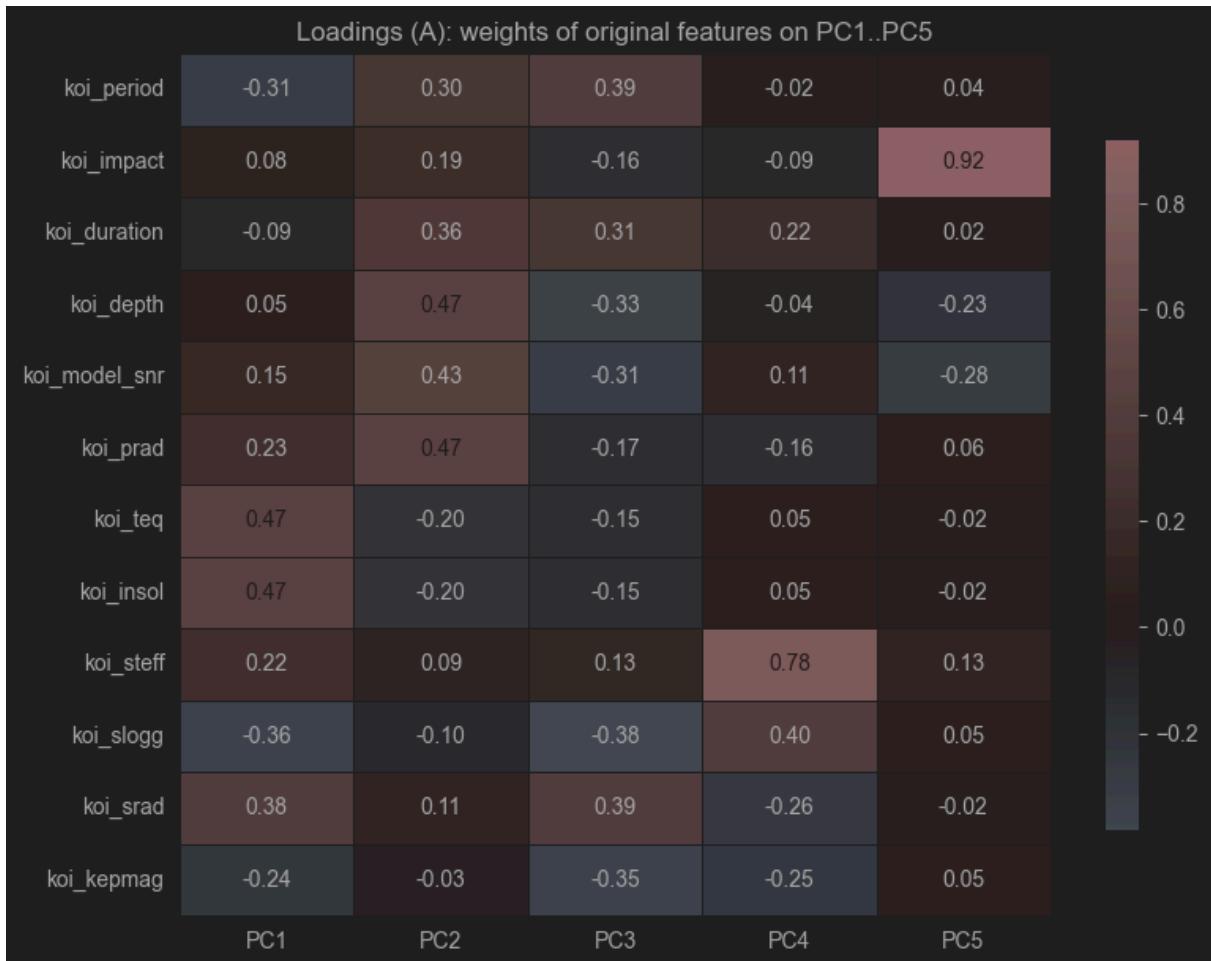
Scree Plot and Pareto Plot (Cumulative Variance)



Biplot



Loadings Heatmap



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): $7,994 \times 12$
- **X_pca5** (first 5 PCs, 89.50% variance): $7,994 \times 5$