Columns

🏷️ IDs & Labels (don’t train on IDs)

- **kepid / EPIC / TIC** — Object ID (Kepler/K2/TESS).

  Ex: `TIC 12345678` → use for grouping/splitting, not as a feature.

- **koi_name / kepoi_name / toi** — Candidate name.

  Ex: `KOI-1234.01` / `TOI-700 b`.

- **disposition** — Ground truth label.

  KOI: `koi_disposition` (Confirmed / Candidate / False Positive)

  TOI: `tfopwg_disposition` (PC / FP / KP / APC)

  K2: `archive_disposition` (CANDIDATE / CONFIRMED / FP)

🌑 Transit Shape & Timing

- **period** (`koi_period`, `toi_period`) — Days between dips.

  Ex: `3.52 d` → short-period “hot” planet candidate.

- **duration** (`koi_duration`, `toi_duration`) — Hours the dip lasts.

  Ex: `2.8 h` → helps judge geometry & sanity.

- **depth_ppm** (`koi_depth`, `toi_depth`) — Dip depth in parts-per-million.

  Ex: `10,000 ppm` (=1%) → big planet signal.

- **epoch / t0** (`koi_time0bk`, `toi_transit_epoch`) — When first transit happens.

  Ex: `2457000.123 (BJD)` → for plotting/folding.

- **impact_parameter** (`koi_impact`) — How central the transit is (0 center, ~1 grazing).

  Ex: `b = 0.2` U-shaped; `b = 0.9` V-shaped → grazing, be cautious.

📣 Signal Quality / Vetting Hints

- **SNR / MES** (`koi_snr`, `koi_model_snr`, `mes`) — Signal strength.

  Ex: `MES = 12` → decent detectability.

- **num_transits** (`koi_transit_count`, sometimes derived) — How many dips observed.

Ex: `N=5` → stronger confidence than single-transit.

- **odd_even_test** (various flags/columns) — Odd vs even depth difference.

  Ex: "Odd≠Even" → eclipsing binary warning.

- **secondary_depth** / **eclipse_flag** — Dip at phase ~0.5.

  Ex: `clear secondary` → likely binary, not planet.

- **vshape / shape_flag** — U vs V shape indicator (if present).

  Ex: `V-shaped` → grazing binary risk.

🌟 Star (host) Properties

- **st_teff / koi_steff** — Star temperature (K).

  Ex: `5777 K` Sun-like → sets expected sizes/brightness.

- **st_logg / koi_slogg** — Surface gravity (cgs).

  Ex: `4.4` → main-sequence star (good). Low values can mean giant star issues.

- **st_rad / koi_srad** — Star radius (R⊙).

  Ex: `1.0 R⊙` → depth translates cleanly to planet size.

- **st_mass / koi_smass** — Star mass (M⊙).

  Ex: `0.9 M⊙` → context for orbits.

- **[Fe/H] / metallicity** (`st_metfe`) — Star metal content.

  Nice to have for trends; not critical for baseline.

- **mag** (`kepmag`, `Tmag`) — Brightness in mission band.

  Ex: `Tmag=10.5` → brighter = better follow-up.

🪐 Planet (if provided/derived)

- **planet_radius_re** (`koi_prad`) — Planet radius (Earth radii).

  Ex: `11.2 R⊕` ≈ Jupiter-size → depth should be large.

- **insolation_flux / teq** (`koi_insol`, `koi_teq`) — Star energy / equilibrium temp.

  Ex: `teq=1200 K` → hot world; sanity check with period.

🦠 Contamination & Quality

- **crowding / contamination** — Nearby starlight diluting transit.

  Ex: `crowding=0.7` → 30% of light is from neighbors; depth underestimates size.

- **centroid_offset / motion** — Image center shifts during transit.

  Ex: `offset > 3σ` → background eclipsing binary likely.

- **quality flags** (`data_quality`, `vetting_flags`, `disposition_score`) — Pipeline/QA hints.

  Ex: low score or bad flags → treat prediction cautiously.

✅ Minimal Feature Set to Start (tabular MVP)

1. `period`
2. `duration`
3. `depth_ppm`
4. `impact_parameter`
5. `SNR/MES`

6. `num_transits` (or derive)
7. `st_teff`
8. `st_logg`
9. `st_rad`
10. `kepmag/Tmag`
11. `planet_radius_re` (if present)
12. `crowding/contamination`
13. `centroid_offset` (if present)
14. `odd_even_test`
15. `secondary_depth / eclipse_flag`

> Keep IDs for grouping/splits; don't feed them into the model. Use disposition as your label.

🧩 Tiny Examples (one-liners)

- **depth_ppm=800** & **st_rad=1.0 R⊙** → shallow dip → likely small planet.
- **period=0.8 d**, **V-shape**, **odd_even mismatch** → eclipsing binary suspect.
- **SNR high**, **num_transits≥3**, **U-shape**, **no secondary**, **good centroid** → planet-like.