Meaning of important columns:

* kepid: Target identification number, as listed in the [Kepler Input Catalog (KIC)](https://ui.adsabs.harvard.edu/abs/2011AJ....142..112B/abstract). The KIC was derived from a ground-based imaging survey of the Kepler field conducted prior to launch. The survey's purpose was to identify stars for the Kepler exoplanet survey by magnitude and color. The full catalog of 13 million sources can be searched at the [MAST archive](http://archive.stsci.edu/kepler/kic10/search.php). The subset of 4 million targets found upon the Kepler CCDs can be searched via the [Kepler Target Search form](http://archive.stsci.edu/kepler/kepler_fov/search.php). The Kepler ID is unique to a target and there is only one Kepler ID per target.
* kepoi\_name: A KOI is a target identified by the Kepler Project that displays at least one transit-like sequence within Kepler time-series photometry that appears to be of astrophysical origin and initially consistent with a planetary transit hypothesis

* kepler\_name: [These names] are intended to clearly indicate a class of objects that have been confirmed or validated as planets—a step up from the planet candidate designation.
* koi\_disposition: The disposition in the literature towards this exoplanet candidate. One of CANDIDATE, FALSE POSITIVE, NOT DISPOSITIONED or CONFIRMED.
* koi\_pdisposition: The disposition Kepler data analysis has towards this exoplanet candidate. One of FALSE POSITIVE, NOT DISPOSITIONED, and CANDIDATE.
* koi\_score: A value between 0 and 1 that indicates the confidence in the KOI disposition. For CANDIDATEs, a higher value indicates more confidence in its disposition, while for FALSE POSITIVEs, a higher value indicates less confidence in that disposition.

Columns which are dropped:

* **koi\_pdisposition**.
  + This column describes whether a KOI [Kepler Object of Interest] is considered a candidate for an exoplanet, or a false positive, or a confirmed exoplanet.)
  + **koi\_disposition** is the same, but also incorporates newer, peer-reviewed/published information from the Exoplanet Archive, so keep that and get rid of **koi\_pdisposition,** which is from the Kepler pipeline.
* **koi\_tce\_plnt\_num** and **koi\_tce\_delivname.**
  + **koi\_tce\_plnt\_num** is the TCE (Threhold-Crossing Event) Planet Number federated to the KOI.
  + **koi\_tce\_delivname** is TCE delivery name corresponding to the TCE data federated to the KOI.
* all the error columns (**koi\_period\_err1**, **koi\_impact\_err2**, etc...)
  + These columns contain uncertainty values for their corresponding columns, e.g. **koi\_period\_err1** and **koi\_period\_err2** are the positive and negative uncertanties for koi\_period.
* **kepler\_name.**
  + Kepler number name in the form "Kepler-N," plus a lower-case letter, identifying the planet.
  + Only the confirmed planets have these, everything else has a NaN value here. when we try to drop rows with NaN values, everything besides the confirmed exoplanets will get dropped... which we don't want. We can always cross-reference from the cleaned-out dataset back to this one if we really want the object's Kepler name.

Drop the data.