Data description:

Daily median price data

Goal:

Obtain useful representation of price trends over time

Problems:

Outliers caused by combination of low trading volume and overpay for rare item characteristics, such as expensive stickers or extraordinary float values (most notably for FN items).

Data cleaning:

Outlier removal:

* Mean-based methods fail (outliers too extreme and frequent)
* Locally normalized methods perform poorly when the volatility changes drastically over time (overly sensitive when stable, and vice versa)
* Absolute methods are too crude and cant adapt to changing volatility

Solution – median-based hybrid method:

Modified z-score with offset:

* Still does not reliably identify outliers in very sparse data situations:
  + Solution: remove all data points with no neighbours within the rolling window range