Factor Analysis model notes:

* Both PCA and FA to be run on explanatory variables
* Synthetic values. Variables called latent variables
* Some assumptions: features are metric, numeric, corr (‘r’) > 0.3, obs>100, sample is homogenous
* Factors ~ 1 are good
* Factor 1 has the biggest impact. Factors that have the highest loading or none at all should be retained.
* Those that have a weak loading should be removed.

Principal Component Analysis:

* Reduces variables down to synthetic components called principal components
* Factors and components are what is left of a dataset after redundancy and noise has been stripped out
* Variance ratio: how much variance is explained by the components found
* Explained variance: how much information is compressed into the first few components
* Keep at least 70% of dataset’s information
* Some info is tied up with noise, and outliers. So don’t want 100%
* Here: 1st component: 92%, 1st+2nd component: 98%
* How to use once you’ve isolated them:
  + Components can be used as input variables for ML algorithms, such as a classifier