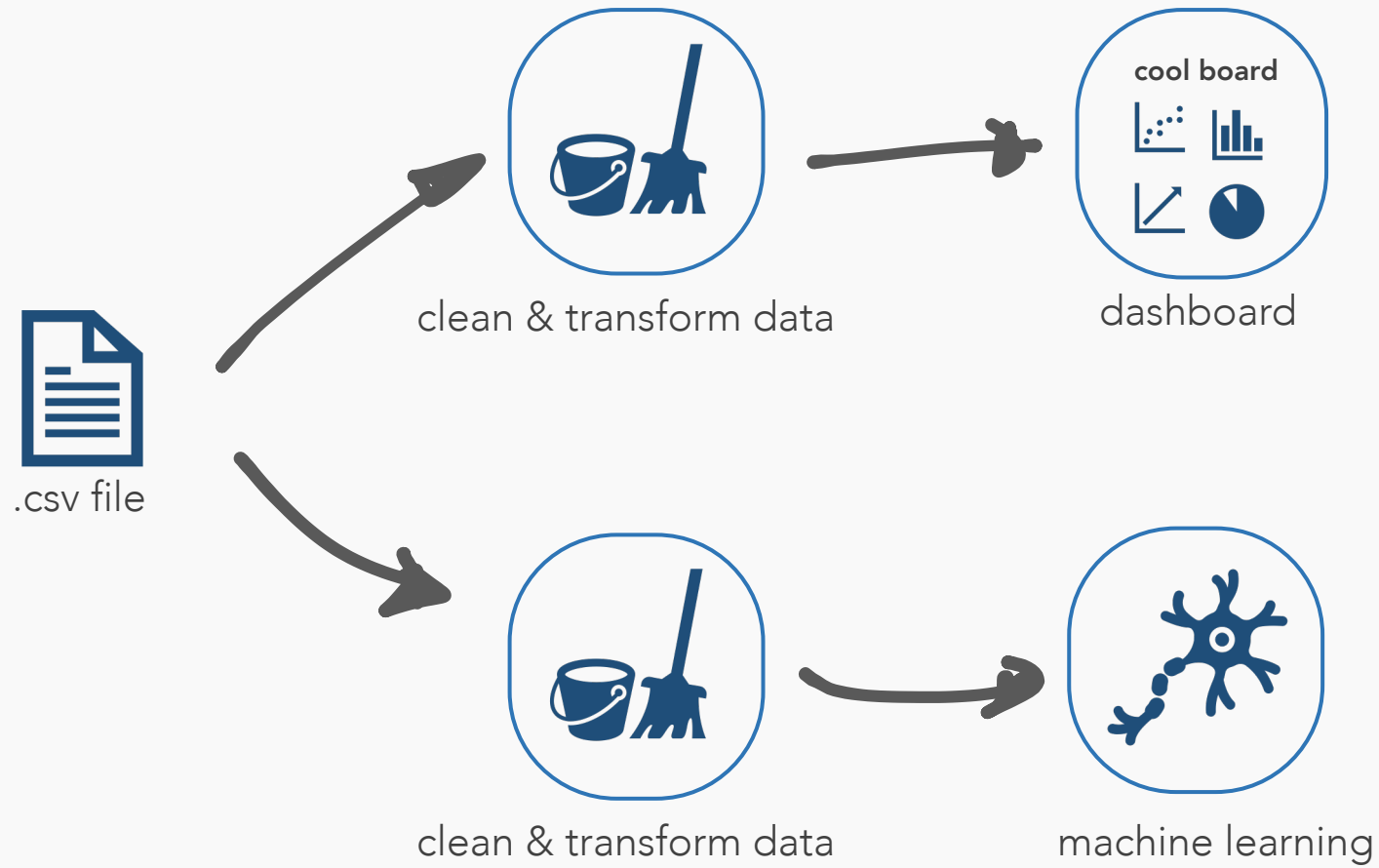


kokchun giang

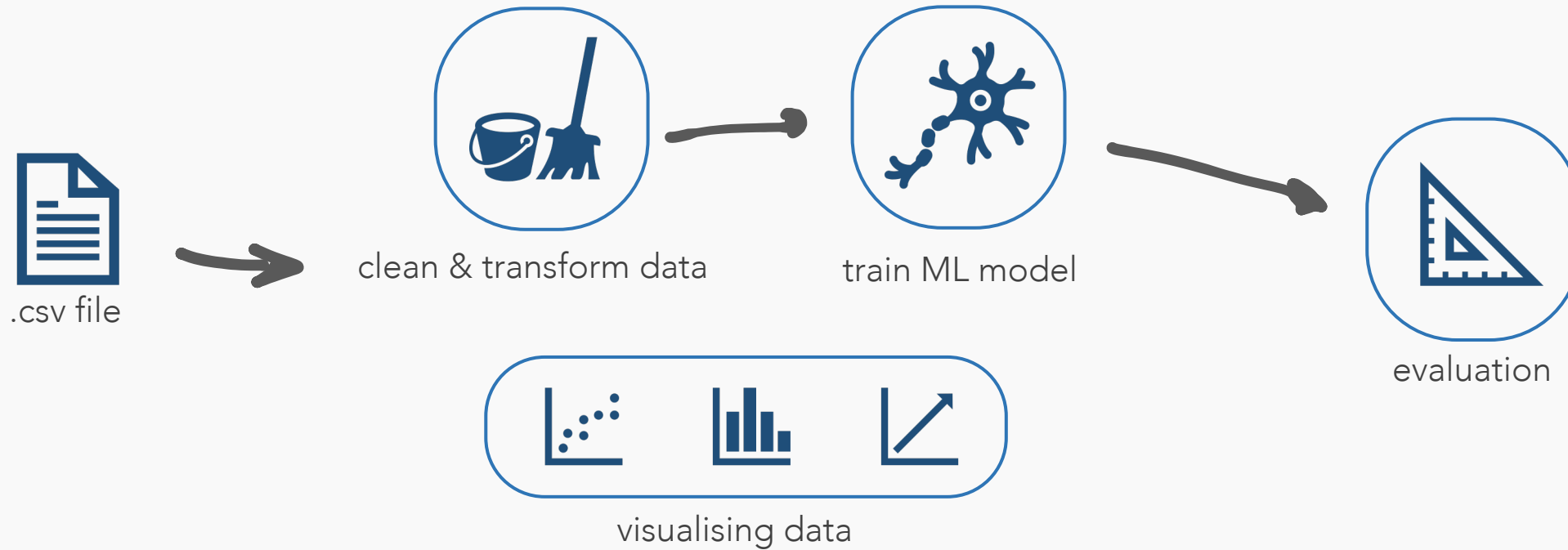
the need for a **data engineering** pipeline



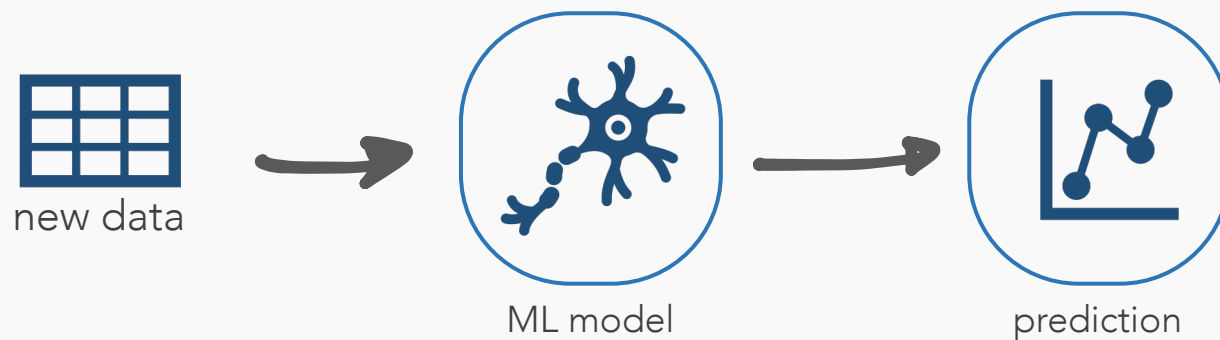
what have we **learnt** so far?

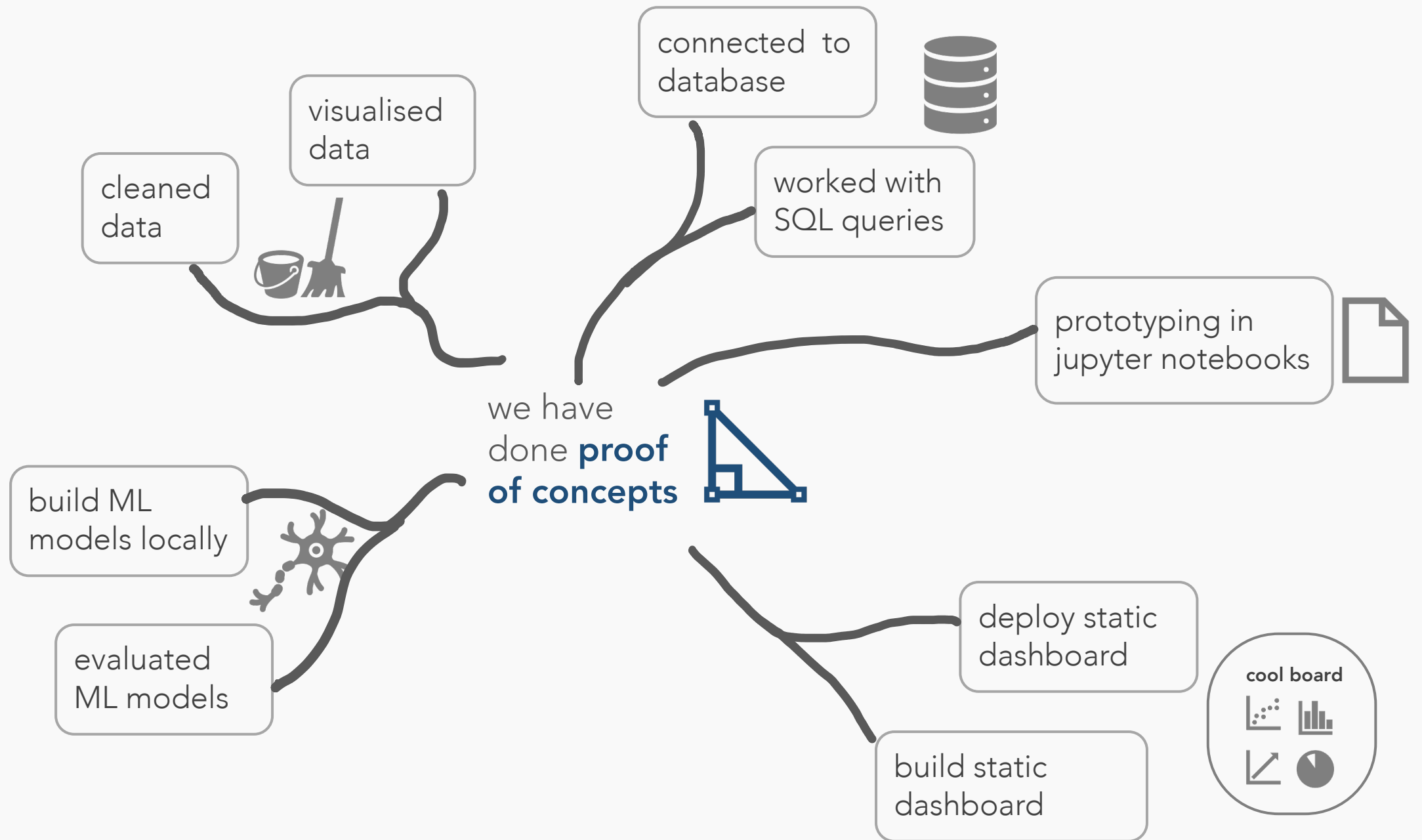


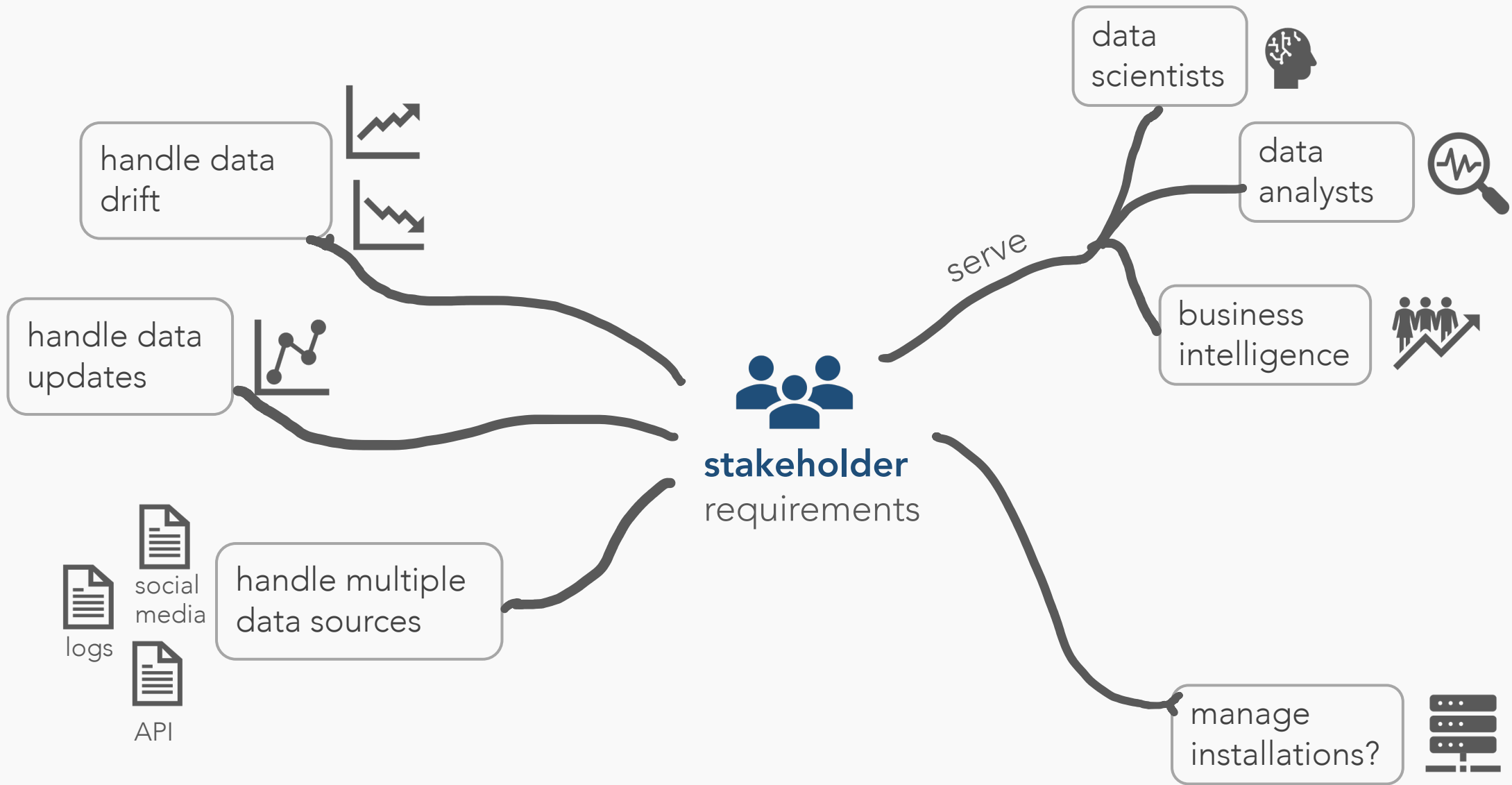
current **ML workflow** simplified – training part



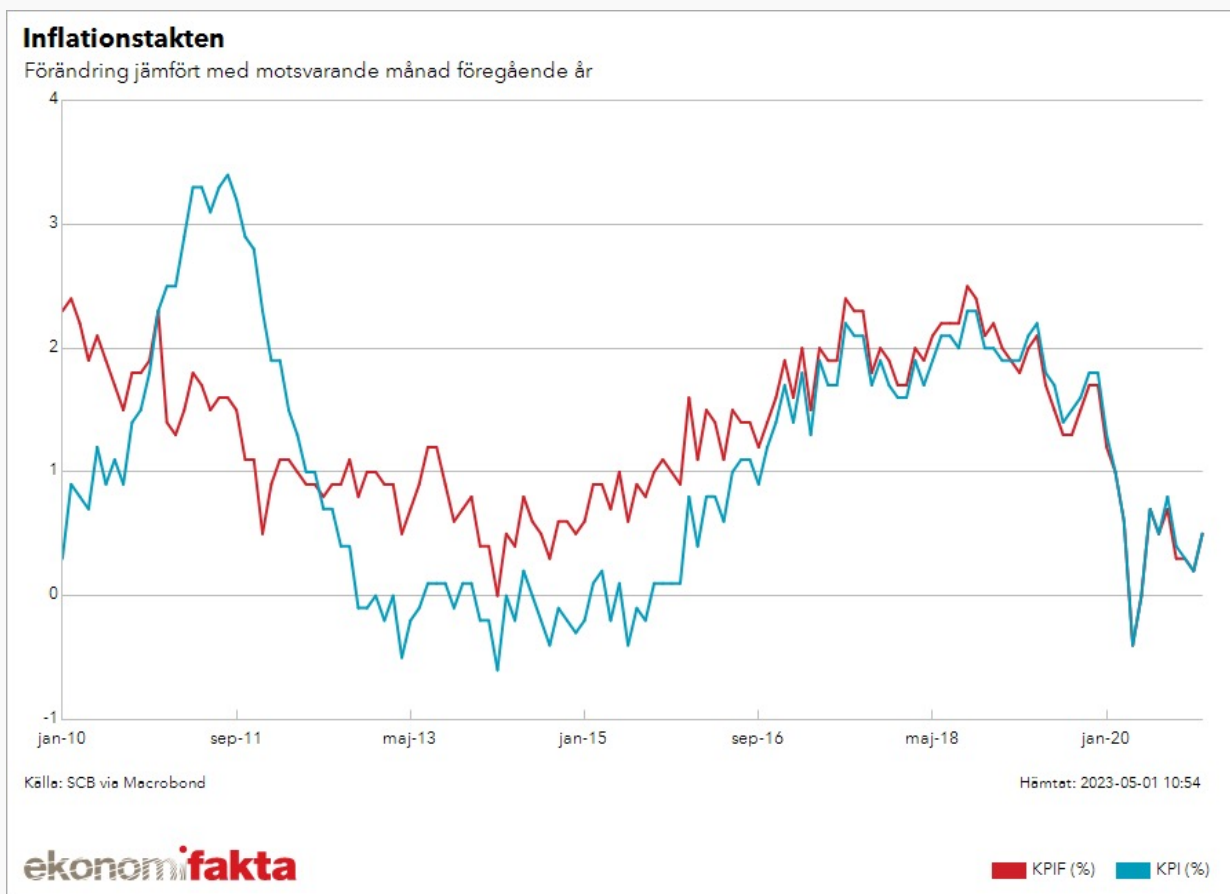
current **ML workflow** simplified – inference part





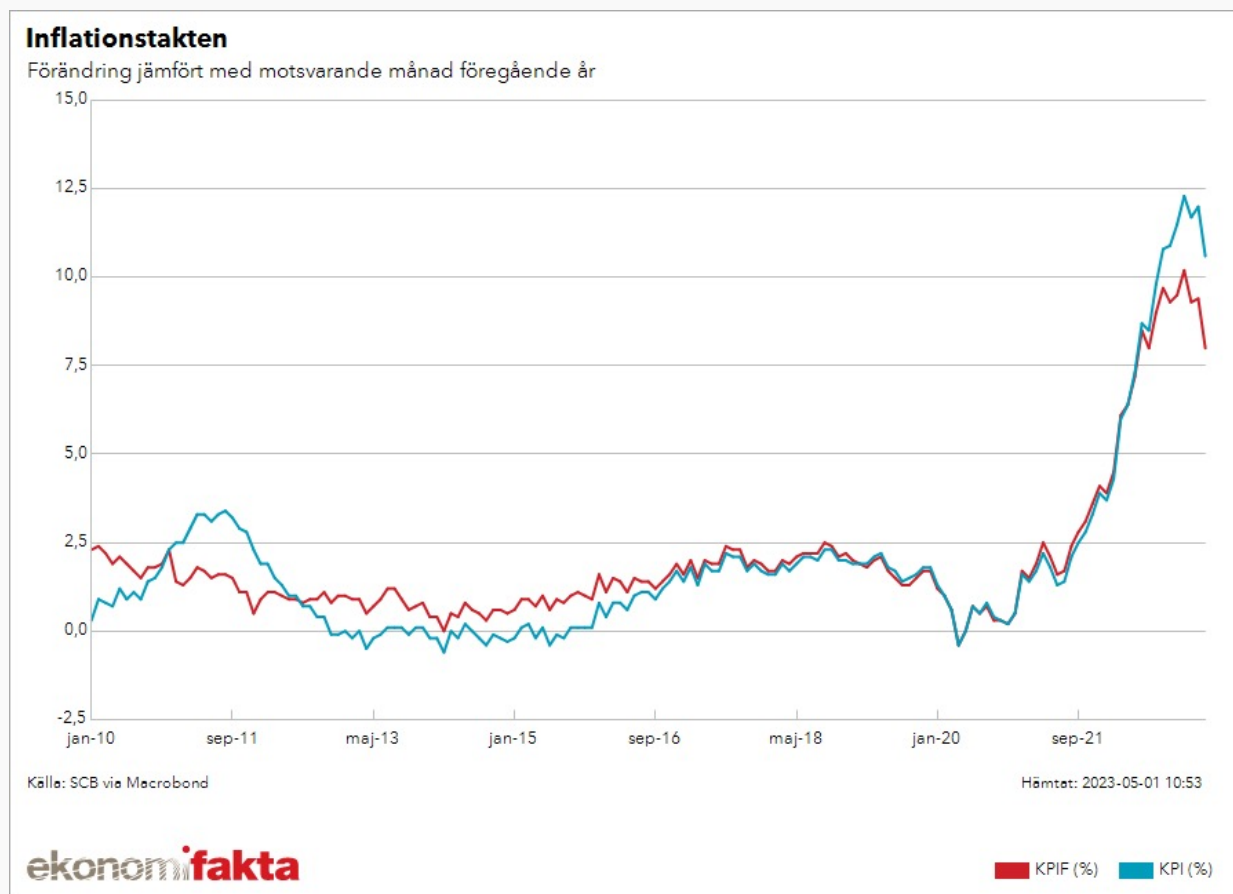


example of **data drift**



train a model on
Swedish inflation
data 2010-2020

example of **data drift**



data evolved to
invalidate the
previously
trained model

desired wishes after **data updates**

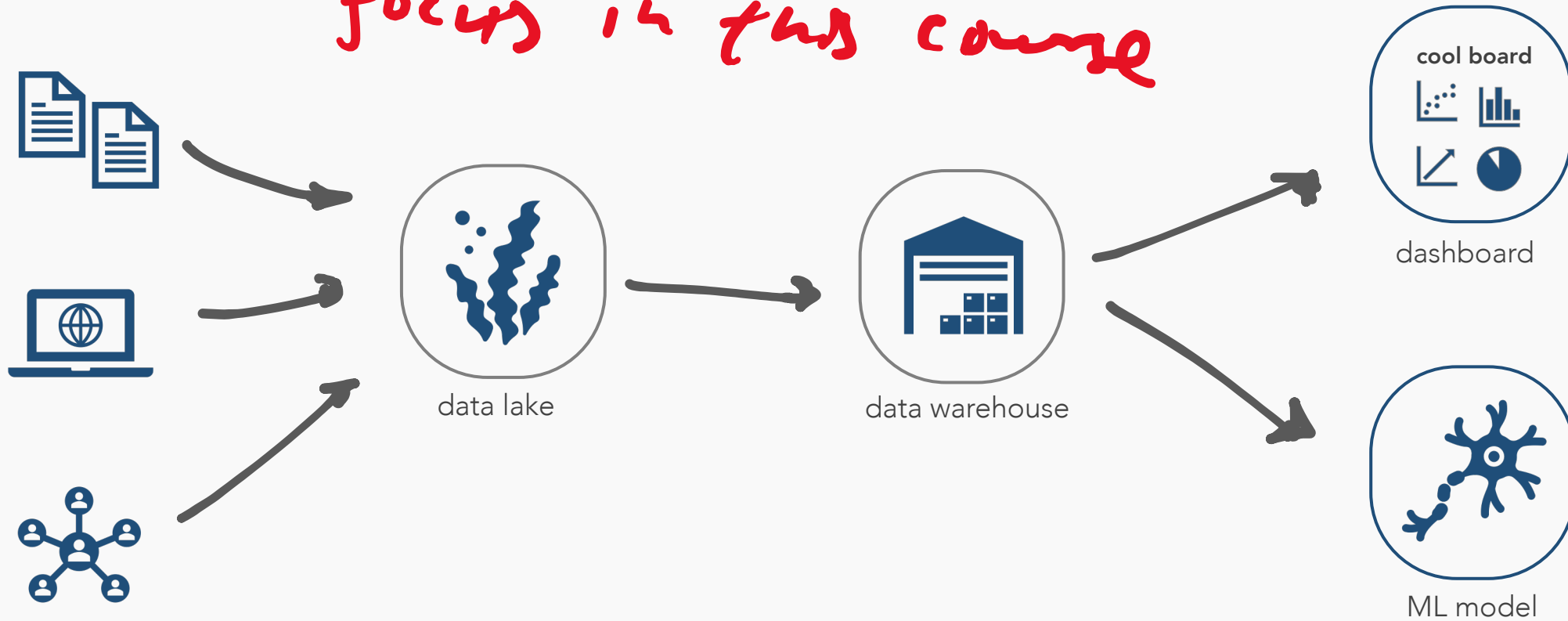
✓ dashboard visualise
updated data

✓ ML models have up
to date parameters

✗ manually run scripts
when data is updated

data pipeline for the rescue

focus in this course



different
data sources