7/17/2025

Use simulations to illustrate careful data synthesis

* We want to achieve data minimization
* What is the minimum amount of variables that we need to include for a business-relevant purpose?
* Compute PCs on original data
* Option One Synthesis steps
  + Compute PCs from original data
  + Synthesize those principal components
  + Reverse the transformation from synthetic principal components back into original features
  + Compute PMSE ratio using original and synthetic features
* Option Two Synthesis Steps
  + Compute PCs from original data
  + Synthesize those principal components
  + Compute pMSE ratio using original and synthetic principal components
  + Reverse the transformation after optimization process
* Double check logistic regression code