Bayesian Imputation

While missing values are usually handled prior to the modeling phase of a data science project, it is worth noting an exception where missing values can be handled automatically as part of the modeling process. This is is the case when a model is treated in a *fully Bayesian* way, that is priors are used to govern parameters of the model. Then Expectation-Maximization, Markov Chain Monte Carlo (MCMC) or another method of inference can be use to infer both the parameters, hyper-parameters and missing values.

See the following resources to learn more.

PyMC3 - package for probabilistic programming in Python.

TensorFlow Probability - another package for Python that enables the Bayesian treatment of models

PyMC3 Getting Started - continue on to Case Study 2 to see how missing values are automatically imputed during inference