

# Pandas DataFrames for F.A.S.T. binned analysis at CMS

*Tuesday, July 10, 2018 11:30 AM (15 minutes)*

Many analyses on CMS are based on the histogram, used throughout the workflow from data validation studies to fits for physics results. Binned data frames are a generalisation of multidimensional histograms, in a tabular representation where histogram bins are denoted by category labels. Pandas is an industry-standard tool, providing a data frame implementation that allows easy access to “big data” scientific libraries, including I/O, visualisation, and machine learning tools.

F.A.S.T. (Faster Analysis Software Taskforce) runs regular hackdays and tutorials in the UK. We present a new toolkit where Pandas DataFrames are used as the basis for binned physics analysis. We demonstrate how this engenders faster, more robust, and more flexible development, using fewer lines of code, as well as improving accessibility for newcomers.

The toolkit is presented in the context of a typical CMS search analysis, where we look for evidence of new physics signal in a multidimensional parameter space, but is not analysis specific. The code structure is simple and modular, with built-in bookkeeping and documentation. We envisage its adoption by other UK binned analyses, and plan associated tools, helping physicists to focus more on “what” their analysis should do rather than technically “how” it is done.

**Primary authors:** Dr KRIKLER, Benjamin (University of Bristol (GB)); DAVIGNON, Olivier (University of Bristol (GB)); Dr KRECZKO, Lukasz (University of Bristol (GB)); LINACRE, Jacob Thomas (Fermi National Accelerator Lab. (US)); OLAIYA, Emmanuel Olatunji (STFC-Rutherford Appleton Laboratory (GB)); SAKUMA, Tai (University of Bristol (GB))

**Presenter:** Dr KRIKLER, Benjamin (University of Bristol (GB))

**Session Classification:** T6 - Machine learning and physics analysis

**Track Classification:** Track 6 –Machine learning and physics analysis