Contribution ID: **7**   Type: **not specified**

Taking Raw Data Towards Analysis

Tuesday, 24 February 2015 14:30 (1 hour)

Taking the understanding of multi-variate interactions from lecture one; this lecture aims to demonstrate the descriptive power of various clustering algorithms in Python with a brief introduction to large scale data preprocessing software such as Hadoop.

Targeted audience: Anyone wishing to learn more about how to look at data. Data scientists in this context can be either physicists or computer scientists or people involved in marketing or recruitment.

Benefits of attending the lecture: Demonstration of the importance of understanding the nature of the data that were working with before beginning to analyse it. Examples of various software for data processing and visualisation will also be given.

Prerequisites: This lecture will be self contained and easy to follow for anyone with some experience in basic computing and statistics.

Presenter: CROFT, Vincent Alexander (NIKHEF (NL))

Session Classification: Exploring EDA, Clustering and Data Preprocessing