Function References List

There are four libraries imported for this analysis: MatplotLib, numpy, pandas and root_numpy. Root is a framework created by CERN for large scale data analysis such as the one you are about to attempt (and hopefully complete!) and pandas also contains functions related to data manipulation and analysis. Numpy is python's "fundamental package for scientific computing" and MatplotLib allows graphs to be drawn (using the same functions as in MATLAB). The list below gives some functions that you will find useful, although you can use other functions from the libraries given, common arguments used in the functions are given. Note that you should type them as library.function(arguments).

numpy

sqrt(n) - Returns the square root of n.

mean(data) - Returns the mean of given data.

sum('logical expression') - Returns the number of occurences where the logical expression is true in the data. Like countIF in excel.

maximum(data1,data2) - Compares two datasets and returns an array containing the element-wise maxima.
http://docs.scipy.org/doc/numpy-1.10.0/reference/generated/numpy.maximum.html

minimum(data1,data2) - Compares two datasets and returns an array containing the element-wise minima.
http://docs.scipy.org/doc/numpy-1.10.1/reference/generated/numpy.minimum.html

MatplotLib

hist(data, bins = n, range = [x,y]) - Plots a histogram of given data in *n* equally spaced bins over the range *x* to y. e.g. hist(data.H1_PX, bins = 40, range = [-100000,100000]).

hist2d(data1,data2, bins = n,) - Plot a 2D histogram from two datasets, with n^2 bins.

scatter(x,y,c,alpha) - Plot a scatter plot of x vs y, with colour 'c' and alpha blending value (between 0-1). http://matplotlib.org/api/pyplot_api.html#matplotlib.pyplot.scatter

Pandas

Note: The rest of these functions apply to a panda data structure!!!

DataFrame.head(n) - When applied to a panda data structure produces a table of the first n rows of data in the strucutre.

DataFrame.eval(expression) - Evaluate an expression in the context of the DataFrame. E.G. data['M'] = data.eval('E^2 - P^2').

You can find more information here:

http://pandas.pydata.org/pandas-docs/stable/generated/pandas.eval.html

DataFrame.min() - Returns the minimum value for each axis in a dataframe (an axis is equivelant to a column)

DataFrame.max() - Returns the maximum value for each axis in a dataframe

DataFrame.query(expression) - Allows selection from a DataFrame variable using an expression of logical rules. If a DataFrame has a column called "Mass" then for a DataFrame of particle data...

FilteredDataFrame = DataFrame.query("Mass < 200") would return a new DataFrame containing all the data from the original, but only if the particles mass was less than 200.