3 BODY PROBLEM – OUR APPROACH

1. Test set does not give velocities – it gives positions x and y for the 3 bodies at t=0 for various simulations, and initial velocity is always 0. (??? - why are there various times that have the positions at t=0 in this data?)
2. Train set – as some regions of the table with probably “faulty” simulations, where there is a sequence of lines with all values (except ID) = 0; this is data that will skew our model, so we aim to strip the train set of it with code/function that does something of the sort:

* if line all 0 (except ID):
  + line.strip()
* else:
  + Skip
* (this is an important part of our data preprocessing)

1. To account for data outliers and because position, velocity and time are measured using different scales, it is probably reasonable to use some type of standardisation.
2. To the aims in 2. and 3., we started by printing out a table with descriptive statistics of our variables across all simulations – summary\_stats.xlsx