CS 4265 Homework #2

Write a Scala program in Spark to load the block_1 dataset, accessible from the Software Repository of the D2L course site, and do the following:

1. Parse all records (CSV)
2. Convert all fields except the last one to doubles. Questions marks should be NaN. The last field should be converted to a Boolean.
3. Output an array of statistics for each field.

What to submit?

Take a screenshot of your program output and turn it in with your Scala program in D2L.