

## Springboard Project 18 (PySpark)

The aim of this project is to gain some experience with large data sets and parsing them with PySpark. After installing PySpark, I performed a few basic exercises and created an RDD to get started. With the RDD in place, I then do various mapping exercises, to count and pluralize words.

Next, a brief mini-project is completed using the text of *Julius Shakespeare*. I count the number of words, the first 20 words, and then remove stop words before counting again. Finally, a histogram to assess word frequency is produced.