

🏠 (/) / Courses (/courses) / 6240 Parallel Data Processing with Map-Reduce (/courses/5) / 3 - In-Class (/courses/5) / 06 - Starting Spark

Assignment: 06 - Starting Spark

Bucket: 3 - In-Class

Due Date: 2016-02-19

Grading Hidden? no

Teams? no

Description:

Start from the in-class [Spark Samples](#)

(<https://d1b10bmlvqabco.cloudfront.net/attach/ij4yepvz8v3zf/i3hms81gzknwa/ikpumhgckszx/SparkSamples.tar.gz>)

1. Write a Spark program that finds the (player, year) that hit the most home runs per dollar of salary.
2. Find the (name, year) of the pitcher (appears in Pitching.csv) with the best batting average (Hits / At Bats).

Assignment Download: ()

Your Submissions

New Submission (/assignments/50/submissions/new)

Date	Status	Automatic	Teacher	Score	Link
2016-02-19 23:25:53 -0500	✓	0 / 100	100.0 / 100	100.0 / 100	View (/submissions/3738)

Course Page (<http://www.ccs.neu.edu/home/ntuck/courses/2016/01/cs6240/index.html>) | Piazza (<https://piazza.com/class/ij4yepvz8v3zf>)

Bottlenose copyright © 2012-2015 Nat Tuck. Licensed under the GNU Affero GPL (/agpl-3.0.txt) v3 or later. Source at github (<http://www.github.com/NatTuck/bottlenose>). The development team takes no responsibility for death or serious injury that may result from use of this program.

ajax-status: none