Notes from 14 December 2016

Spark

* Feature rich but lacks features that enterprises want
* Spark allows people to curate data at a much larger scale
  + More efficient than Hadoop (Map reduce)
  + Also more efficient
* Spark is less verbose than Java MR
* Servian bot is built on IBM Watson, coded with PAS (sort of like the customer service box that you interact with on a website) – the more data you have and you can use to train, the more automated you can make the outcome
* They run the spark meetup in Sydney

Presentation

* Webscraping packages: requests page and beautiful soup
* Datamaps: cool map choices with javascript
* D3 is easier if you want to embed into website (compared to python)

Presentation comments

* Changes in Kiva’s policy over time and map to rate of expiry
* Looking at number of lenders that jump on during the lifecycle of the project
* What did you learn about the models from this
  + Decision trees and random forests are good for predicting but not so good with showing feature important