What is Tech? Data Science

Tatjana Predojev

tatjanapredojev@gmail.com

Women Who Code, February 2017

What is data science?

"Data Scientist (n.): someone who is better at statistics than any software engineer and better at software engineering than any statistician."

Josh Wills on twitter, 2012

No consensus about the correct definition!

- Tech end
  - machine learning (python, R), Big Data (Spark, h20), very technical (computer scientists)
- Business end
  - Excel, SAS, SPSS, lots of visualisation (Tableau, QlickView), less technical

Good news

- New field for everyone
  - goes beyond statistics (a lot more programming-oriented)

- Funny example: Spark initially released in 2012 vs. job ads requirements (at least five years experience in Spark)

Where to start?

- Love maths :)
- Free online course *Machine Learning* by Andrew Ng
- Another free course *Statistical Learning* by Hastie and Tibshirani
  - download free books both Intro and Elements
- Cloudera's free *Intro to Hadoop & MapReduce* hosted by Udacity
- *Algorithms* by Robert Sedgewick hosted by coursera

Get some experience!

- My story: management consulting and tech startup
- On a regular workday: read scientific papers, understand algorithms, "get your hands dirty" i.e. code, improve method accuracy, improve performance, understand and interpret results
- Normally, machine learning algorithms are not implemented directly, but applied (lots of pre- and post-processing)

Practice!

- Practice with open data
  - example: Renfe train schedules and delays

- Kaggle: form a team
  - protip: install Anaconda for package version hell

What is Tech?
Data Science

Tatjana
Predojev

What is Data
Science?

Where to
start?

Get experience

Conclusion

Sexiest job of the 21st century?

- Choose what you genuinely like doing, irrespective of the buzz

Thank you!