8/29/19

**What is data science? Article**

Data science is not just statistics, you need computation, coding, and visualization skills. It is all encompassing to work with data. You could come from a physics, social sciences, computer science, or statistics background.

1. Potential project could be text analysis of data science job postings
2. Tweets about products and their sales? Or profit or something like that?
3. Use deep learning to analyze the amount of advertisements on popular podcasts

**Article: What is Deep Learning?** This was the title on Springboard, but the article didn’t attempt to answer that question and the title on the site was very different.

**How to Break into Data Science Article**

https://www.linkedin.com/pulse/one-technique-land-your-dream-data-science-job-matt-fornito/

**Data Science Capstone Project**

Project 1 should be simpler (no web scraping), overall, keep both projects simple and boring if you can so that you can keep it clean. Answer a question that is useful to someone. BE VALUABLE.

Is it a real problem?

Is there data available?

Is the data acquirable and clean?

Thoughts and Sites for Data

* I want to look at newer data (2017+) as it is more likely to answer a relevant question?
* [www.quandl.com](http://www.quandl.com) has financial data, could you do something with implied volatility vs actual volatility. I like the idea of options and trading, so this is at least mildly appealing
  + Insider Trades <https://www.quandl.com/databases/SF2/data>
  + ARIMA model for stock prices (what is this?)
* Government Data <https://www.data.gov/>
* Data is Plural email list archive <https://docs.google.com/spreadsheets/d/1wZhPLMCHKJvwOkP4juclhjFgqIY8fQFMemwKL2c64vk/edit#gid=0>
  + Economic mobility, not sure if this is suitable for a first capstone
* UCI ML repository <http://archive.ics.uci.edu/ml/index.php>
* Kaggle, public data competitions <https://www.kaggle.com/competitions>
* Mode Analytics <https://mode.com/blog/five-public-dataset>
* Google Dataset search, could be a useful resource
  + <https://toolbox.google.com/datasetsearch>
  + http://www.google.com/publicdata/directory
* I want to look

Write three potential projects with a brief description of the problem you are solving and where you will get the data.

https://quant.stackexchange.com/questions/16481/why-do-we-usually-model-returns-and-not-prices

<https://www.dataquest.io/blog/free-datasets-for-projects/>

<https://docs.google.com/document/d/1uUqwmw2qnHmdcw7aTiExf4yIhDczI-EbYkU0_Dluono/edit?usp=sharing>