Data Science & Python Resources

WHAT TO READ:

1. Short reads:
    a. KDNuggets – the ultimate source on all things data science (www.kdnuggets.com)
    b. Analytics Vidya – has great tutorials on common data science techniques (www.analyticsvidhya.com/blog/)
    c. Towards Data Science – there are a few helpful gems. Read the top articles. (www.towardsdatascience.com)
2. Long reads:
    a. An Introduction to Statistical Learning with Applications in R
    b. Automate the Boring Stuff with Python
    c. Learn Python the Hard Way
    d. Storytelling with Data

TWITTER RESOURCES

1. Fei Fei Li - chief of Machine Learning at Google; she makes Google happen (@drfeifei)
2. Hilary Mason (@hilarymason)
3. Kirk Borne (@kirkborne) (my boss at Booz Allen Hamilton & top 3 data science influencer)
4. Dr. GP Pulipaka (@gp_pulipaka)
5. Data for Black Lives (@data4blacklives)
6. Gil Press (@GilPress)
7. Ben Lorica (@bigdata)

DATA SCIENCE KNOWLEDGE BASE:

1. **Expertise** in Probability & Statistics
2. Proficiency in linear algebra, multivariate calculus
3. Proficiency in data visualization
4. Proficiency in Python (or any programmatic language for statistical calculations)
5. Proficiency in SQL
6. Daily practice, practice, practice with data science problems

1. **STATISTICS**[1]**:**
    a. Read "An Introduction to Statistical Learning with Applications in R" aka "the ISLR" (http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf)
        i. Find the Python version here (https://github.com/JWarmenhoven/ISLR-python)
2. **LINEAR ALGEBRA:**
    a. Khan Academy (www.khanacademy.org/math/linear-algebra)
    b. Read The Matrix Cookbook: (http://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf)
3. **MULTIVARIATE CALCULUS:**
    a. Khan Academy (www.khanacademy.org/math/multivariable-calculus)
4. **DATA VISUALIZATION:**

[1] Seriously, master statistics.

   a. Read Storytelling with Data: A Data Visualization Guide for Business Professionals (https://www.amazon.com/Storytelling-Data-Visualization-Business-Professionals/dp/1119002257)
5. **PYTHON:**
   a. Read Learn Python the Hard Way (https://learnpythonthehardway.org/) to become proficient
   b. Read Automate the Boring Stuff with Python (www.automatetheboringstuff.com) to make your life easier
   c. Datacamp/Dataquest/Udacity/Udemy[2] or any other learning platform.
6. **SQL:**
   a. Learn Structured Query Language to manipulate databases (https://www.khanacademy.org/computing/computer-programming/)
7. **DAILY PRACTICE[3]:**
   a. Again, work through the ISLR: https://github.com/JWarmenhoven/ISLR-python
   b. **Datacamp** - Explore the **Python for Data Science** track. Cost = $25/mo. Explore the student discount. Has daily practice mode. (www.datacamp.com)
   c. **Kaggle**** – an online competition site. Try your hand at old projects with have solution code (www.kaggle.com)
   d. **Codewars** – practice your skills in a competitive format

[2] It doesn't matter which platform you use. Pick one that you're able to finish.

[3] I don't really endorse Datacamp because it isn't free. I endorse daily practice. And Datacamp makes it easier than most platforms.