***COURSERA: Introduction to Data Science in Python***

**WEEK 1- Python Fundamentals**

* Python = used in 8 out of 10 of the US's top CS programs
* Python programs tend to have minimal templating + more natural constructs for typical tasks you might need to accomplish
* Python is also full-featured + a very general programming language w/ a lot of built-in libraries + it excels at manipulating data, network programming, + databases.
* It's mature + there's plenty of resources available from books to online courses.
* Python has a significant set of DS libraries one can use, the base of which is the **SciPy Ecosystem**
* Interest in DS is at an all-time high + really has exploded in popularity in the last couple of years.
* The history of data science goes back prior to 2004 + the popularity of interest in the area comes from the network + data-driven society we find ourselves living in.
* When people think of the term *data scientist*, they tend to think of Google or Amazon or Facebook, places with big AI research teams
* Certainly these are some amazing companies who are doing great things with DS, but data scientists aren’t just limited to careers w/ tech companies.
* Almost every company is turning to DS to better understand how to build products, serve customers + leverage new opportunities.
* Companies aren't the only one (Universities + academia)
* The need for data driven intelligence + skills is growing in companies + organizations throughout the world
* The techniques + methodologies of DS stem from the fields of CS+ stats.
* Ex: Drew Conway’s Venn diagram 🡪 suggests DS = intersection of hacking skills, math + stats knowledge, + substantial expertise (a bit of an oversimplification)
* DS is definitely one of those areas where you ask 10 people + get 10 different answers.
* 1 thing missing from this diagram is the underlining need for the scientific inquiry.
* You don't necessarily get this for having good hacking skills or math + stats knowledge.
* A good data scientist bring skepticism, experimentation, simulation, + replication to bear on understanding a given phenomena.
* Whether it be trying to predict sales at a coffee shop, cluster online store products, determine who's likely to win in election, or mine health data from physiological sensors like Fitbits.
* Also need some communication skills in the form of charting, graphing, + related visualizations.
* A good data scientist is one who can communicate findings clearly to others 3:59
* David Donoho, Professor of Statistics at Stanford provides excellent commentary of views of the field in a paper entitled “50 years of Data Science”
* He describes the field broadly as being made up of 6 activities
* **Data Exploration + Preparation** = involves cleaning data + manipulating it for further analysis.
* **Data Representation + Transformation** = Tabular structures, text data, + graph-based data, etc.
* **Computing w/ Data** = R + Python are fundamental here
* **Pipelining** = Data Scientists need to be able to work w/ different languages for different parts of an analysis project
* Unlike enterprise software projects, where you might use 1 language for implementing all functionality you need, modern DS projects can span many different languages + computing paradigms
* Knowing when to use the right tool for the job is an important attribute.
* **Data Modeling** = a big space, but usually concerned w/ about **predictive modeling** + **generative modeling**.
* Increased interest in predictive modeling is helping to fuel the current DS push.
* The modern world, w/ massive data streams + significant computational power, gives us opportunity to rapidly experiment w/ making predictions, allowing us to innovate + evaluate new DS techniques.
* **Data Visualization + Presentation** = charting, graphing, 3D visualization, interactive environments
* Information Visualizationis a whole field of its own.
* A meta-activity = **The Science About DS =** understanding what works + what doesn't in DS, + building ways to leverage these discoveries.
* See examples of this as new tools + paradigms of computing are invented to change the way DS works.
* DS is becoming a fundamental way of understanding the world around us.
* While many institutions are starting to offer DS masters degrees + certificate, it's important to think of DS as a sort of **epistemology**, or a *way of knowing*.
* DS thinking can be useful in a variety of disciplines + careers as a way of approaching problems + critical thinking skills