



Forums / General Discussion

Help

## What's next on your data science "to do" list?

Subscribe for email updates.

HOME



Announcements

Videos

Syllabus

CAPSTONE PROJECT

**Task 0 Info:**  
 Understanding the problem

**Task 1 Info:** Data acquisition and cleaning

**Task 2 Info:** Exploratory analysis

**Task 3 Info:** Modeling

**Task 4 Info:** Prediction

**Task 5 Info:** Creative exploration

**Task 6 Info:** Data product

**Task 7:** Slide Deck

ASSIGNMENTS

Quizzes

Report Submission

Milestone Report Rubric

Final Project Rubric

COMMUNITY

Discussion Forums

Hangout with Swiftkey

Course Wiki

Help Articles

 No tags yet. + Add Tag

Sort replies by: Oldest first Newest first Most popular

Michael Haymes (Signature Track) · 23 days ago

This specialization is winding down. It's been a fun process. For those of you continuing to train in Data Science, what's next? Do you have specific areas you're looking to focus on improving your skills? Any good courses or textbooks you're thinking of tackling?

Here's my "to do list"...what's yours?

Math

- Stanford Courses on Intermediate Linear Algebra
- Advanced stats (???)

CS/Programming

- More R (haven't found any good online courses yet)
- More Octave/Matlab (??)
- Coursera Courses on algorithm design/coding linear algebra

Machine Learning

- Read *Elements of Statistical Learning*
- Sebastian Thrun, Tibshirani or Geoffrey Hinton course on machine learning, specifically Neural Networks and some methods I don't know like Naive Bayes
- Start a blog and dive deeper into Shiny to demonstrate the Machine Learning I've picked up.

Fun Stuff

- Arduino!

 11  · flag
Patrick Chamberlain (Signature Track) · 23 days ago

First thing on my list is to watch all of the Stanford NLP course from the beginning

 6  · flag

+ Comment

Maciej Szymkiewicz (Signature Track) · 23 days ago

- Advanced Data Analysis Software Development with R, Neural Networks, Web Scale Data Mining and Processing, Mining massive datasets on the educational platform of the ICS PAS - all in progress
- HarvardX Biomedical Data Science series on edX - starts January 19

 5  · flag
Hendrik D'Oosterlinck (Signature Track) · a day ago

Dear Maciej,

any chance we can get the course material somewhere? These look great!

 0  · flag

+ Comment

Alexandre GADEBSKI (Signature Track) · 23 days ago

I was thinking about the Coursera Data Mining specialization which will start next February  
<https://www.coursera.org/specialization/datamining/20/courses>

The Text retrieval and Text Mining courses from it might be a good continuation for NLP.

That being said, the courses look quite challenging (proficiency in Java or C++, artificial intelligence are in the list of requirements).

7  · flag

Patrick Chamberlain Signature Track · 5 days ago

Thanks for pointing that out @Alexandre - looks so interesting I've signed up.

It seems to be spread over a whole year even though it's only 5 one month courses + a capstone. I've done a bit of Java and C++ in the past, hopefully it'll come back to me ..

1  · flag

Michael Lyubinin Signature Track · 5 days ago

Yes, thank you. I signed up for the first course too. See you guys there :)

(I will be taking it in parallel with <https://www.coursera.org/course/compmethods>, which is a little scary with 10-12 hours/week estimate, so will see how it goes).

0  · flag

Mark Henry Butler Signature Track · 5 days ago

Its not that bad time wise. The lectures are around 2 - 3 hours per week. I like the professor - he is funny - but it is just videos of actual lecture (which I always find harder to take in) and very little assessment, no certification. He gives you a big chunk of reading like a book - a few people on my course were disciplined and had a reading group around that - I think they got more out of it. It's more signal processing (radar etc, with some PCA and SVD) than these courses. It could be a good course but they need to adapt it more for the MOOC format.

If you haven't done it, I really recommend this The Analytics Edge. It's the best data science one I have done to date: <https://www.edx.org/course/analytics-edge-mitx-15-071x-0#.VJLyUVxGjUY>

Good luck with it Michael!

2  · flag

David Hood COMMUNITY TA · 5 days ago

I audited compmethods- I knew I was going to have to many commitments but thought I'd follow along for a while. Even auditing it I would say I got a good general idea of processes for signal analysis.

1  · flag

+ Comment

Johannes Eric Victorine van Mulken Signature Track · 23 days ago

I did the first 8 weeks of the Information Theory course, Info Theory is imo a must for senior datascientists:

<https://www.coursera.org/course/informationtheory>

I hope it will rerun and i can do the 15 weeks,it is one of the toughest on coursera.

Furthermore the game theory courses of Stanford are recommended, part 2 starts soon:

<https://www.coursera.org/course/gametheory2>

I havent finished any of the analytic combinatorics classes but I sense these are for advanced algo builders:  
<https://www.coursera.org/course/ac>

There are so many courses I hope that will run again:

probabilistic graphic models by Daphne Koller

automata by Jeff Ullman

discrete optimization

So much to learn, we're blessed

0  · flag

+ Comment



Charles McGuinness

Signature Track

· 23 days ago



For me, there were three areas where I felt like I didn't get the depth of knowledge I wanted.

1. The first is in some basic Statistics and Inference. There's a lot of material that was crammed into that class (from p-values to distribution types etc.) that I'd like to re-study in depth.
2. The second is in building predictive models. I signed up for but didn't really participate in Andrew Ng's machine learning class, and I hope to run through that before the materials go offline.
3. The third is NLP as well, so that's probably on my list after I get through the first two.

I also note that there are some non-MOOC directions to consider -- UC Berkley has an online Data Sciences masters program (kind of pricey compared to Coursera), and Georgia Tech has a much less pricey online CS Masters that appears to have a data science-like track in it. That's still down the road, though.

Now, if Johns Hopkins wants to offer an online Data Science graduate degree, I'd definitely consider that. I've gotten fond of our trio of instructors...

0  · flag

David Hood

COMMUNITY TA

· 23 days ago



Charles, for number 1 you might want to look at the 10 week Data Analysis and Statistical Inference course that Duke does on Coursera. It is very well regarded as an introductory stats one (disclosure, I was a CTA on the last run). <https://www.coursera.org/course/statistics>

7  · flag

Charles McGuinness

Signature Track

· 23 days ago



You're like the CTA for everything!

4  · flag

Jason Wilkinson

Signature Track

· 22 days ago



Well Charles, hopefully this program, along with the linear algebra (see ulaff.net and Strang's book which I found very approachable) and python moocs (Coursera) I've taken has been good preparation for the Georgia Tech OMSCS as I'm starting next month!

0  · flag

Michael Lyubinin

Signature Track

· 8 days ago



I 2nd David's comment. Duke course is one of the best I took on Coursera so far.

0  · flag

Frank Evans

Signature Track

· 8 days ago



Put me on the list for that one. I did it while I was waiting for the Capstone to start up. I really learned a lot, and found it to be a really good class on statistical testing. I wish it had more R in it, but the application of the class is top notch.

0  · flag

[+ Comment](#)



Patrick Chamberlain

Signature Track

· 9 days ago



<http://www.kaggle.com/c/word2vec-nlp-tutorial> looks worth a look (via the Data Science group on LinkedIn)

2  · flag

[+ Comment](#)



Michael A. Wise

Signature Track

· 9 days ago



Well, I already have a job, but partially as a result of all this Data Science activity that I have been doing, I have received a new assignment to work on a Recommender System for an internal project. Very happy about this too. So I am reading up on Collaborative vs. Content vs. Knowledge Based systems, etc. As that is work I can devote quite a few hours a day to it, this certificate and even the experience gained on this project are all extremely relevant to it.

---

Other than that I have a goal to get a lot better at Bayesian Techniques, especially Bayesian Networks. This is going to require a lot of reading, I think I have figured out the next textbooks to absorb - the dog book seems to be the

level I need, followed by Bishop's ML and Pattern Recognition (which is more ESL level).

The upcoming course that most excites me (I have a couple including that Duke Statistical Inference course), is the upcoming one on AI Planning from the University of Edinburgh. Short, but interesting.

---

And I do intend to eventually absorb ESL. But I think it will be spring before I get to it, and a year or two (or more) to really grok it all.

1  · flag

 **Mark Henry Butler** [Signature Track] · 6 days ago

Yes it would be really wonderful if Coursera could organize a course with either [Chris Bishop](#), [David Mackay](#) or [Michael Jordan](#).

It turns out David Mackay does already have some videos available

1  · flag

[+ Comment](#)

 **Taylor Hemby** [Signature Track] · 8 days ago

There is also a National Data Science Bowl that kicks off today and runs through March. The goal is to create a imagery identification algorithm for plankton populations. The details can be found at:

<http://www.datasciencenationalbowl.com/>

2  · flag

 **Sepanda Pouryahya** [Signature Track] · 8 days ago

This looks like a pretty interesting project! Anyone thinking of putting together a Kaggle team? (I've never tried anything out on Kaggle myself, but I'm mighty tempted!).

0  · flag

 **Vamshideep Devershetty** [Signature Track] · 8 days ago

I am planning to put together a team, is anyone interested in joining me?

0  · flag

[+ Comment](#)

 **Eduardo Díez Báez** [Signature Track] · 8 days ago

On January 16th I'll try the final project —capstone— of the Duke specialization.

Each separate courses possess uniquely wonderful qualities and I hope the capstone follow the same line. Perhaps the best three courses I have taken so far in Coursera without thereby belittle any other course, especially this also amazing Johns Hopkins' specialization.

0  · flag

[+ Comment](#)

 **Dinesh Kumar SRB** [Signature Track] · 8 days ago

I may try this

<https://www.coursera.org/course/algo>

0  · flag

[+ Comment](#)

 **Ephraim Baron** [Signature Track] · 8 days ago

I'll be taking Linear Algebra - Foundations to Frontiers through edX starting in late January

<https://www.edx.org/course/linear-algebra-foundations-frontiers-utaustinx-ut-5-02x#.VI8xRnuEl48>

1  · flag

[+ Comment](#)

Maher Harb

Signature Track

· 8 days ago



Next on my to do list:

- Try out some Kaggle competitions. This one in particular seems interesting: <http://www.kaggle.com/c/helping-santas-helpers>
- Take a course that specifically addresses big data (something we haven't dealt with in this specialization).
- Learn python.
- Beef up my knowledge of statistical learning. This book has been a great companion during the specialization (free to download): <http://www-bcf.usc.edu/~gareth/ISL/>

1  · flag

[+ Comment](#)

Johannes Eric Victorine van Mulken

Signature Track

· 8 days ago



intro into statistical learning is rerunning within a couple of weeks

<http://online.stanford.edu/course/statistical-learning>

10  · flag



Kurt Kuppens

Signature Track

· 7 days ago



Thanks Johannes for sharing.

I was looking for a course to refresh and improve some statistical concepts using R.  
I just registered.

0  · flag



Kevin Markham

COMMUNITY TA

· 7 days ago



The Stanford course is excellent! Here are the videos from the last iteration of that course:  
<http://www.dataschool.io/15-hours-of-expert-machine-learning-videos/>

4  · flag



Richard Ambler

Signature Track

· 6 days ago



This is on my to-do list too. The course text is what Maher posted above. It looks like a really detailed practical tutorial on a broad range of topics.

0  · flag



Nitin Gupta

Signature Track

· 4 days ago



Thanks Johannes. I signed up too.

Those who are looking for a more rudimentary start, I wholeheartedly recommend Prof. Ani Adhikari's [Intro to Statistics](#) series over at edX. Start with Stats2.1x (Descriptive Statistics), then 2.2x (Probability) & 2.3x (Inference). Even though the course material is archived, just watch the videos. She's a brilliant instructor.

0  · flag



Anonymous

· a day ago



I did Stanford Course as well, real good one and a good free book, can recommend.

0  · flag



Knut Behrends

Signature Track

· a day ago



I have registered for this one, too. First impression is: Very approachable and thorough.

The videos are available right now, but the automatically graded exercises aren't.

0  · flag

+ Comment

Peggy Fan Signature Track · 8 days ago

Excellent list!

I want to do more kaggle competitions too- perhaps this is a good place to form kaggle teams?

If you are interested in doing competition for social good, here is <http://www.drivendata.org/competitions/>  
It just got started but it's another place to try out your skills.

0  · flag

+ Comment

Bradley Skaggs Signature Track · 8 days ago

edX's [Introduction to Big Data with Apache Spark](#) and [Scalable Machine Learning](#) classes

4  · flag

+ Comment

 Kevin Markham COMMUNITY TA · 7 days ago

Here is a list of resources I created for the students in my data science class:  
<https://github.com/justmarkham/DAT3/blob/master/resources.md>

Besides online classes, it lists some face-to-face educational programs, recommend blogs, aggregators, books, and more.

Hope that is helpful to some of you!

Kevin

6  · flag

Sonya Crofton Signature Track · 7 days ago

Definitely bookmarking this page. Thanks!

1  · flag

Michael Haymes Signature Track · a day ago

Kevin, thank you for posting these resources. This is an excellent list and exactly the kind of resource I was looking for. Much appreciated.

0  · flag

+ Comment

Sonya Crofton Signature Track · 7 days ago

So many great suggestions listed in this thread. I will miss the forums the most about this course.

My next project is the Challenges in Global Affairs Capstone (Jan 19th). I will attempt to write a data driven research paper in international relations. Too bad I didn't take courses from the Duke Reasoning, Data Analysis, and Writing Specialization.

Even though I am really tempted to take more classes, I think I need to first polish my portfolio and resume so I can do this stuff full-time :)

1  · flag

 Eduardo Díez Báez Signature Track · 6 days ago

You always have the chance to take those three Duke courses.  
Each one is a pearl.

0  · flag

[+ Comment](#)

Zecca Lehn

Signature Track

· 6 days ago [□](#)

Looking forward to integrating Data Science via Hadoop / Mahout / R -- bigger contingency tables!

I purchased Become a Certified Hadoop Developer | Training | Tutorial for \$18 on Black Friday (regularly \$198). <https://www.udemy.com/hadoop-tutorial/?dtcode=2jQjqr23feT#/>

Also getting started with reading Hadoop: The Definitive Guide, recommended by the course instructor (4th edition due in April 2015). <https://www.udemy.com/u/zeccalehn/>

[□ 0 □](#) · flag

[+ Comment](#)  
Molly J. Markey

Signature Track

· 5 days ago [□](#)

I am enrolled in the edX course "MIT 6.00.1x Introduction to Computer Science and Programming Using Python", which begins on January 7th: <https://www.edx.org/course/introduction-computer-science-mitx-6-00-1x-0#.VJLywyvFOI>

I'm looking forward to learning some Python, to supplement what I have learned about using R in this specialization, and improve my programming skills.

[□ 1 □](#) · flag

  
Alexandre GADEBSKI

Signature Track

· 5 days ago [□](#)

It is a very good introduction to Python. I found the second part of this course MIT [6.00.2x Introduction to Computational Thinking and Data Science](#) more challenging and time-consuming than the 1st part. These courses are also a part of Xseries (specialization) Foundations of Computer Science <https://www.edx.org/xseries>.

6.00.1 and 6.00.2 are the only courses that have been run so far, others are getting postponed. According to their site, the next course to be run is 6.004.1x Computation Structures: Digital Circuits. But I am more interested in two Java classes that were postponed to fall 2015 and spring 2016.

[□ 2 □](#) · flag

  
Molly J. Markey

Signature Track

· 5 days ago [□](#)

Thanks for this information, Alexandre! I'm glad to hear that this course will be a good one. I'll keep a lookout for the next scheduled run of the 2nd portion of the course, too (currently, there are no future offerings listed--but perhaps one will appear in Feb. or March?).

[□ 0 □](#) · flag

[+ Comment](#)

Maher Harb

Signature Track

· 5 days ago [□](#)

It's interesting to contrast how Insight prepares it's fellows to the data science program:  
[http://insightdatascience.com/blog/preparing\\_for\\_insight.html](http://insightdatascience.com/blog/preparing_for_insight.html)

I know couple of ex-colleagues who went through that program and landed jobs in silicon valley right after completion. I wonder how the Coursera specialization ranks against their 6 week training.

[□ 1 □](#) · flag

  
Anonymous· 5 days ago [□](#)

It doesn't, not even close. I think there are lots of brilliant people who took the JHU Cert (like yourself) who will do great but this program is not going to become highly reputed. I think someone who only does analytics edge+DASI+NG's class+reads ISL will be more knowledgeable than someone who only does this cert.

[□ 0 □](#) · flag



Michael A. Wise

Signature Track

· 5 days ago [□](#)

Looking at the insights program, it seems to be weak in R (non-existent?). Also, as I have pointed out before, although Andrew Ng's course is a brilliant introduction and I loved it, its neglect of ensemble methods (Bagging, Boosting and Random Forests) is a serious flaw. However judging it without touching it would be stupid, so who knows, maybe it is better. Frankly though, I don't see why one needs a course in Python. It is a pretty pedestrian language, but one with some great libraries.

0  · flag

 Michael A. Wise Signature Track · 5 days ago

Actually I take it back and agree with you... but I think one needs to point out that one would have invested over double the amount of time to do what you suggest, even without reading ISL.

0  · flag

 Mark Henry Butler Signature Track · 4 days ago

Pedestrian? At least it supports [list comprehensions](#) and [higher order functions](#). Of course [you can do this type of thing in most languages](#) - but [in some you have to work harder than others](#) :-)

I had a job interview for Spotify last year for a data engineer job. I do Java / Scala development normally but they insisted I did a programming assignment in Python. It took me twice as long as it should have because it was the first time I had used the language and resulted in them not taking my application further. I understand [other companies use Python heavily too](#).

So instead of getting cross I enrolled in several Python courses. I did the Udacity's [Introduction to Computer Science](#). Even though this is an introduction course I got something a lot out of it - you get to implement Page Rank at the end. Then I did the Udacity [Programming Languages course](#) - this isn't using programming languages - it's an introduction to how you build them - so it covers building lexers / parsers / interpreters and compilers. It is a very good course.

Then I did the two Python MIT courses on EdX - although I confess I did not watch the videos for the first two weeks of the first one - [Introduction to Computer Science and Programming using Python](#). Despite the title the second one [Introduction to Computational Thinking](#) is not so much about data science like this course, more about algorithms and artificial life models. These courses are very high quality as they come from MIT. They have a lot of practical assessment which I like as that forces me to practice the skills they are teaching.

Then I just finished the Coursera [Interactive Programming With Python](#). This is focused on writing games - so will be less relevant to people here.

Then in my professional work I do use Python quite a bit now. It's good to know a scripting language - the other candidate is Ruby - that has some nice features too and slightly better support for functional programming. It wins for web development but thanks to libraries like [Numpy](#) / [Scipy](#) / [Scikitlearn](#) and [Pandas](#) Python has the edge for data science. Plus it is supported on [Hadoop](#) and [Spark](#) which are two important platforms for large scale data science.

Mind you I have language prejudices too - don't ask me to do PHP or Perl - maybe I will get over them too some day ;-)

3  · flag

[+ Comment](#)

 Maciej Szymkiewicz Signature Track · 5 days ago

Here is quite interesting reading: <http://www.amstat.org/education/pdfs/guidelines2014-11-15.pdf>

0  · flag

[+ Comment](#)

 Chao Yang Signature Track · 5 days ago

A hands-on tutorial on deep learning: <http://danielnouri.org/notes/2014/12/17/using-convolutional-neural-nets-to-detect-facial-keypoints-t...>

1  · flag

[+ Comment](#)

 Dan Flath Signature Track · a day ago

I may take a look at geospatial intelligence  
<https://www.coursera.org/course/geoint>

Also - metaacademy has some interesting looking courses

[http://www.metacademy.org/course\\_guides/](http://www.metacademy.org/course_guides/)

For example - I would like to understand how Deep Belief Networks work.

In the Natural Language Processing line, I would like to create an app to do part-of-speech tagging.

1  · flag

[+ Comment](#)

 **Patrick Chamberlain** [Signature Track] · 10 hours ago

This Kaggle tutorial/competition looks interesting <https://www.kaggle.com/c/word2vec-nlp-tutorial> - NLP with word vectors in Python.

0  · flag

[+ Comment](#)

 **Dan Flath** [Signature Track] · 3 hours ago

Attending conferences keeps you connected to the field. Here's one that interests me this coming February.

<http://theinnovationenterprise.com/summits/predictive-analytics-innovation-summit-san-diego2015>

I went to a conference with the same name organized by the same people in Chicago last month (Nov 2014) and it was fantastic.

0  · flag

 **Michael Lyubinini** [Signature Track] · 2 hours ago

Here's another one: <http://strataconf.com/big-data-conference-ca-2015>

BTW, if anyone's going to Strata and is interested in getting together, drop me a note - <my first name>@<my last name>.com

0  · flag

[+ Comment](#)

 **Manish Sharma Kolachalam** [Signature Track] · 32 minutes ago

Maybe a little late but since this discussion forum has pointed out interesting courses i have enrolled for, I thought I should share this. I must say I am probably a bit of an oddball in this course with my interest in behavioural Sciences (which I think are pretty important)

There's a math heavy course that I don't know if I am up for but seems interesting by Antonio Rangel:

Principles of economics with calculus

<https://www.edx.org/course/principles-economics-calculus-caltechx-ec1101x#.VJmu5F4AKA>

And a course about network analysis from Cornell

<https://www.edx.org/course/networks-crowds-markets-cornellx-info2040x#.VJmt9F4AKA>

0  · flag

[+ Comment](#)

New post

To ensure a positive and productive discussion, please read our [forum posting policies](#) before posting.

B	I	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> Link	<code>	<input type="checkbox"/> Pic	Math	Edit: Rich <input type="checkbox"/>	Preview

Make this post anonymous to other students

Subscribe to this thread at the same time

Add post