My Journey From Advancement Data Analyst To Data Scientist

Renée M. P. Teate

APRA Data Analytics Symposium

August 8, 2018

Good morning!

- About Me
- What is Data Science?
- What can Data Science do for the fundraising industry?
- Should *you* become a data scientist?
- How to Learn Data Science / Hire a Data Scientist

#APRADAS2018
@becomingdatasci

Goals: Get everyone here up to speed on data science, and help those of you who want to grow into data science roles figure out what you need to learn in order to do so.

Write down your questions and thoughts… we'll have a discussion at the end of this presentation!

#APRADAS2018
@becomingdatasci

My Background

- James Madison University Integrated Science and Technology (BS, 2004)
- Started business out of undergrad
- Data Analytics & Reporting Department, JMU Division of Advancement
- Data Analyst role at Rosetta Stone
- Masters in Systems Engineering at University of Virginia

My Background

- JMU Advancement in full-time Data Analyst & Report Developer role
- Finished Masters Degree, started pursuing Data Science
- First Data Scientist job at Higher Ed Startup HelioCampus

My Start in Advancement:
Data Analyst at James Madison University

- Learned the “language” of university fundraising (domain knowledge)
- Started by building reusable queries for Ellucian Advance reports, later helped design data warehouse “reporting layer”
- JMU purchased IBM Cognos and we re-built reports to be interactive, self-service
- Data warehouse summary tables allowed those who didn’t know SQL to query/filter
- I had the opportunity to work with people across the Division of Advancement, and absorbed the business of fundraising while building data products. Partnered with Annual Giving to develop advanced segmentation tool.
- Attended my first APRA Data Analytics Symposium in 2015 in New Orleans

What I Do Now: Data Scientist at HelioCampus

- Data Engineering (SQL, AWS Redshift)
- Interactive Report & Dashboard Development in Tableau
- Predictive Modeling / Machine Learning
  - Application & Enrollment Forecasting
  - Financial Aid Offer Optimization
  - Retention Modeling
  - Donor Scoring
- Beyond Advancement
  - Full Student-Alumni Lifecycle

What is Data Science?

“What is Data Science?”

A collection of techniques that use mathematical computer programs to **find patterns in data**, often in order to predict something. Presenting those results in a usable form.

- Finding donors who might give enough to join a gift club next year, because they “look like” donors who joined a gift club this year
- Identifying geographic clusters of alumni in order to determine where your next alumni chapter should be formed
- Predicting how much a donor might give next year, based on their giving history (and others’ giving patterns), to create a custom ask amount
- Forecasting funds raised, based on seasonal and long-term giving trends
- Determining which donors might respond to an appeal to give to a certain fund, based on other funds they have given to

“What is the difference between a Data Analyst and a Data Scientist?”

Both need to be able to do this

Business Question

Data Question

Data Answer

Business Answer

Data Analysts might do some of the following to get a “Data Answer”...

- Query Data
- Build Summary Reports
- Slice & Dice data - filtering, sorting, customizing lists
- Create Data Visualizations
- Dig into Details
- Build “Data Analysis Products” (like “Self-Service” Interactive Reports)
- Find new data sources to integrate
- Find trends & forecast

Data Scientists do all of that as well, with additional skills and techniques for the “Data Answer” Step

- Machine Learning
  - Classification
  - Clustering
  - Regression
  - Time-Series Forecasting
  - Recommendation systems
- Often building repeatable analyses with code and data pipelines

- **Classification** (“supervised”)
  - Train the computer to split data into groups, by showing it examples of records that fall into each desired group so it can recognize patterns that define each group
    - *Finding donors who might give enough to join a gift club next year, because they “look like” donors who joined a gift club this year*

- **Clustering** (“unsupervised”)
  - Ask the computer to identify which records are “similar” to one another by calculating the “distance” between data points and breaking them into a specified number of groups
    - *Identifying geographic clusters of alumni in order to determine where your next alumni chapter should be formed*

- **Regression**
  - Show the computer a bunch of “input” values as well as the “output” values that resulted. It will learn the relationship between the inputs and outputs, so if you feed it a new set of inputs, it can predict an output value.
  - *Predicting how much a donor might give next year, based on their giving history (and others’ giving patterns), to create a custom ask*

- **Time-Series Forecasting**
  - Inputting values summarized over fixed time intervals, so the computer can identify patterns in the value over time (seasonal fluctuations, long-term trends), and forecast the value at a future time
  - *Forecasting Funds Raised*

- **Recommendation systems** (“collaborative filtering”)
  - Feed in a bunch of individuals’ ratings for (or interactions with) a set of items, and the computer will learn to predict which other items that person might like (or interact with) based on what other people who gave similar ratings for other items also liked.
  - *Think Netflix or Amazon: “People who liked ____ also liked ____”*
  - *Determining which donors might respond to an appeal asking them to give to a certain fund, based on other funds they have already given to*

Example Predictive Modeling Process

- Query Data from Source
- Feature Engineering
- Exploratory Data Analysis
- Develop Predictive Model
  - Algorithm Selection
  - Training and Testing
  - Evaluate Test Results
  - *Iterate*
- Optimize Model (Parameter Tuning)
- Get Feedback from Stakeholders
- *Iterate*
- Use Model to Forecast/Predict/Cluster Unseen Values
- Write Results to a Database
- Monitor model performance, evaluate results over time
- *Iterate*/Update Model

“What can data science do for our industry?”

Understand Your Constituents Better

Understand Your Fundraising Trends Better

Predict Outcomes of Outreach Actions

Spend Money and Time more Efficiently

“Should I become a data scientist?”

Does this stuff I’ve been talking about interest you?

Do you want to expand your data literacy, skill set,
and value to your organization?

Are you willing to put in some hard work,
and overcome your own doubts and ignore those of nay-sayers
and gatekeepers along the way?

Then, YES!

“Should I become a data scientist?”

I truly believe that ***anyone*** can become a data scientist.

Plus, every industry needs a whole bunch
of data-literate people right now.
And the need is growing.

Types of Data Scientist

Data Engineering Type

Data Analyst Type

Machine Learning Developer/Researcher

Types of Data Scientist

Data Engineering Type

Data Analyst Type

Machine Learning Developer/Researcher

“How can I become a data scientist?”

First, determine where you are,
and where you want to go.

The specific learning path you take
will be custom to you.

Skills to Learn

Mathematics/Statistics

Minimum

- Undergraduate-Level Statistics
- Data Visualization Techniques

Possibly also...

- Linear Algebra
- More Advanced Statistics
- Calculus

Computer Science

Minimum

- Coding for Data Manipulation & Summarization in 1 language (like python)
- Basic “Packaged” Machine Learning Techniques (like scikit-learn)

Then add on...

- Understanding databases & SQL
- Understanding how Machine Learning Algorithms work in detail “behind the scenes”
- Additional languages (like R) or tools/packages
- Software Engineering - Data Science Pipelines

Domain Knowledge

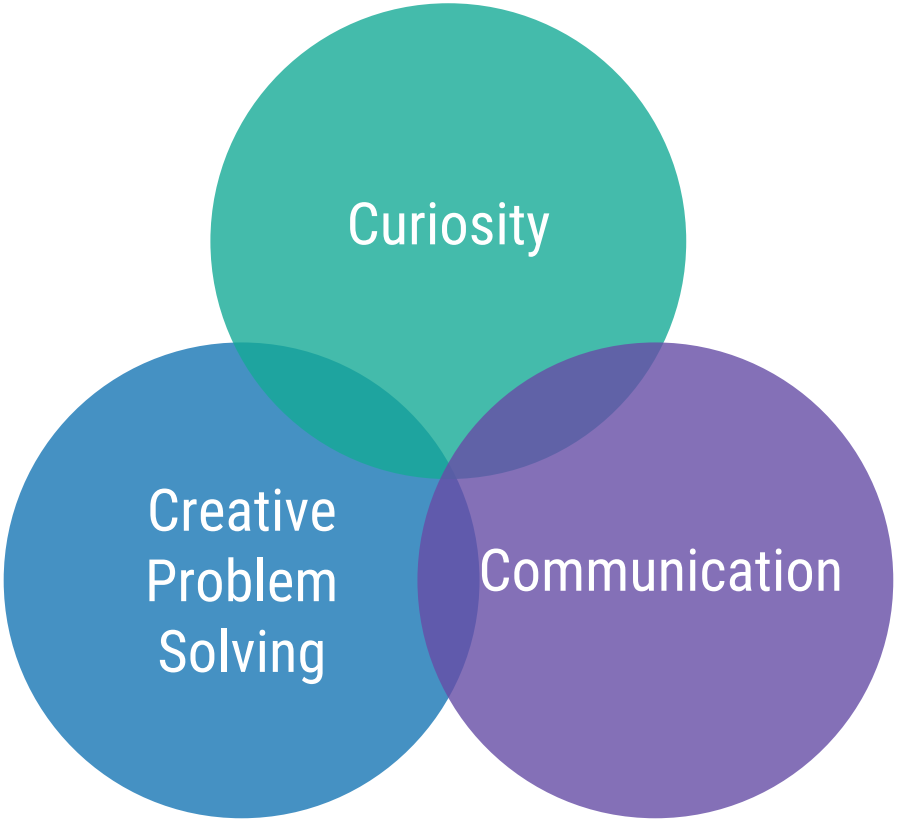
Minimum

- Understanding business enough to gather report requirements, and add value
- Communicating results to stakeholders

Next

- Communicate/translate between “technical” and “non-technical” staff
- Seek out work done by others in domain, determine what would be useful to your institution, and adapt for your own use

“Non-Technical” Skills for Good Data Science Practice



Project-Based Learning

1. Start by finding a dataset related to something you're familiar with and care about
2. Ask a question that could be answered using that dataset
3. Learn the skills as you go to complete the project (or morph it into something better as you learn new possibilities)

Learning Resources

Books - overviews, textbooks, topical guides, conceptual, tutorial, technology-specific

Data Science for Fundraising by Ashutosh Nandeshwar and Rodger Devine

Online courses and video tutorials - Khan Academy, DataCamp, DataQuest, Codecademy, Kaggle Docs, EdX, Coursera, and many many more (some of the paid ones are interactive)

Podcasts

Reviews on DataSciGuide.com (please add yours!)

Talk to me about your specific learning needs, and I'm happy to help recommend resources

Learning Communities

Email: Prospect-dmm Listserv

Twitter: #rstats, #r4ds, #py4ds, #pydata, #SoDS18, #pyladies...

Other Social Media: Facebook & LinkedIN

Slack Groups: Data Science Learning Club, DataEDU, Prospect-DMM?

Product-specific forums like Tableau community

In-Person User Groups and Meetups

Industry Conferences (and find people online via hashtags like #APRADAS18)

Hiring Data Scientists

- Determine your “analytics maturity” and needs first
  - Are you even ready for a Data Scientist?
- Avoid “wishlist” job postings
- Be prepared to hire entry-level, or pay high salaries
  - Industry is paying an average of $115K in the US
- Your best bet is probably to….

Grow your Data Analysts into Data Scientists!

It will be more efficient for you to grow a strong analyst from within your organization than to hire an outsider with machine learning skills and train them on your databases, your internal lingo and definitions, and the business of fundraising.

Chances are your analysts are learning data science in their free time, anyway!
Get the value from them while you can, and pay them as much as you can, because they will be heavily recruited.

Project Ideas

My First “Real-World” Data Science Project

First-Time Donors Predictive Model

Dataset: Non-Donors, plus first-time donors for the past 3 years

NOTE: Imbalanced dataset!

Example fields (would differ based on the type of school, data available, other factors):

Years Since Graduation

Location/Region

Job Status

Contactability

Past Appeals

College/Department

Other Project Ideas

Partner with an end-user to design an interactive report

Determine what would be good triggers for sending an appeal to upgrade to the next gift club level

Determine importance of loyalty vs gift size vs engagement to predicting future gifts above a certain threshold

Use clustering on giving and engagement history to customize outreach

Forecast campaign results

Identify best method for contacting people of different demographics (who responds best to which appeals)

Find the common features of your best appeals

Identify where more data needs to be collected or purchased in order to answer a high-value question

Identify patterns of giving among your constituents

Other Considerations

#protip

Keep in mind that the identified patterns often reflect business practices, not innate properties of the donor

Bias in Machine Learning

How good is your data?

Already building predictive models?
Level up: Data pipeline engineering, reproducible analysis, model monitoring/maintenance

This Week

Your Tasks for this Symposium

- Find someone who does what you hope to be able to do by next year, and ask them questions
- Find someone who is in a similar position with similar goals at another institution, stay in touch
- Take a lot of notes, because you can’t absorb it all in 2 days, but you’ll want to record terminology and tools to look up later, and remember the project ideas you have

How to Contact Me

Renée Teate
HelioCampus

renee.teate@heliocampus.com

@becomingdatasci
on Twitter

BecomingADataScientist.com
blog & podcast

I'll be around for the conference if you're curious to learn more!

Questions & Discussion

#APRADAS2018
@becomingdatasci