ARF Young Pros Bootcamp 2017

YOUNG PROS SPONSORED BY:

- Viacom Media Networks
- Univision Communications Inc.
- Turner

HOST SPONSOR:

Google
Data Science 101: SQL, and Python, and SAS, OH MY!

Terry Sullivan  
Executive Director of Analytic Solutions

Samantha Reeves  
Research Manager, Analytic Solutions
Data Science 101 Workshop

What is Data Science? Why does it matter?

5 December, 2017
Agenda
1. Introduction
   - Vocabulary / handouts
   - What is data science?
   - Why does it matter?

2. Sample problem
   - Question: ad effectiveness, pre-launch
   - Approach
   - Tools + Shiny
Vocabulary List

If we use terms that aren’t on there, we will pay the first person $1 for mentioning it
What does data science mean to you? What words come to mind?
So...what is data science?

A new, interdisciplinary field of study and of work focused primarily on large datasets, with the goal of extracting meaning.
Did you study or take a class on any of these disciplines in school?

- Statistics: 35%
- Machine Learning: 12%
- Programming: 23%
- Artificial Intelligence: 11%
- Business Analytics: 8%
- Data Science: 7%
- None of these: 7%
Data science allows us to **ask** and **answer** new questions

- Data availability and storage capacity is rapidly advancing
  - More data available → answers to more questions
- Open-source technologies → more people participating
- Easy-to-host solutions
**Proprietary software is losing share of preference to open source**

**Database Engine Licenses Popularity Trend**

Source: Wired Magazine
Proprietary software is losing share of preference to open source

SAS, R, Python Preference Over Time

Source: Burtch Works Executive Recruiting Survey
Data science is becoming a degree and a career

Data Science Jobs (United States)

Source: Google Trends
Data science is becoming a degree and a career

GROWTH OF ANALYTICS AND DATA SCIENCE DEGREE PROGRAMS

Cumulative Number of Programs

Year

Source: Institute for Advanced Analytics
How this might impact your career
How this might impact your career

1. Advertising: get ready for addressable media

2. Segmentation and targeting:
   - Opt-in apps create a treasure trove of options
   - Geo-fencing
   - Media usage data

3. Generally:
   - Ignoring big data is risky
   - New options – methods and careers
Other potential impacts

- **Retail:**
  Online shopping changes purchase patterns, even seasonality.

- **Pricing:**
  Dynamic pricing is happening, models will have to adapt.

- **Category Forecasting:**
  Training data is getting alarmingly big, possibility for detail in predictors
Being a Data Scientist
You might like to be a data scientist if you like to wear many hats
We are always:

1. Keeping up-to-date on new theories/techniques
2. Cleaning data
3. Developing data hygiene processes
4. Writing code
5. Debugging code
6. Completely rewriting code
7. Sharing and presenting insights
General Process
Our goal is to generate a repeatable process to do the below data steps

- **Assess data**

- **Extract & clean data**
  - Options depend on the above. Spark, SQL, or SAS

- **Choose analytical tools**
  - Analyses can be done in Excel, SAS, R, Python, or C# with Environments

- **Write, test & run reusable analysis code**
  - SQL, PySpark, Python, R, SAS, C#

- **Identify and build the client’s story**
  - PowerPoint, Shiny, Flask, Excel, JavaScript, ...
Example Problem
Situation

- Brand with limited advertising budget
- Goal = maximize effectiveness of ads to drive sales
- But not sure if a new Internet advertising campaign will drive in-store sales
Complication

- **Limited advertising budget**: must be relatively sure new ad campaign will deliver desired results before rolling out nationwide.

- Further, we need **solid statistics** to prove the new direction will lead to success.
Question

Should the manufacturer pay to launch the ad nationwide?
Before we answer, we consider the data

Several types and of data from different sources may be helpful

The data that we use will guide what type of analysis is possible
What data would you most like to have if you had to answer this question?
Different solutions will arise depending on the data and the client

- Possible Answer 1: If the product and ad will be online, and click-stream data is available, do **online A/B**

- Possible Answer 2: If neither online nor offline data are available, there are always **surveys**

- Possible Answer 3: Depending on the data available and the needs of the client, numerous other tests might be possible
Possible Answer: Using city level scanner data, we can try A/B testing

- Launch the ad in selected cities
- Collect data
- Show how a larger audience would change their buying behavior if they see the ad
A/B Testing Shiny Tool

- Shiny Apps
Question & Answer
Thank you

TAKE THAT FOR DATA