POWERED BY liveguide

# Columbia Engineering: My Story

**Francisco Javier Arceo**

CU SCHOOL OF ENGINEERING

CLASS OF 2016


## Today's Agenda

- Meet Francisco
- Campus Tour
- Academic Highlights
- Stories Outside the Classroom

POWERED BY liveguide

CLASS OF 2016

# Meet Francisco

## Background

Chicago, Illinois
**HOMETOWN**

Illinois State University
**ALMA MATER**

Clemson University
**GRADUATE SCHOOL**

Data Scientist – Commonwealth Bank
**CURRENT ROLE**

Economics, BS
**UNDERGRADUATE MAJOR**

Economics & Statistics, MS
**GRADUATE MAJOR**

POWERED BY liveguide

CLASS OF 2016

# Meet Francisco

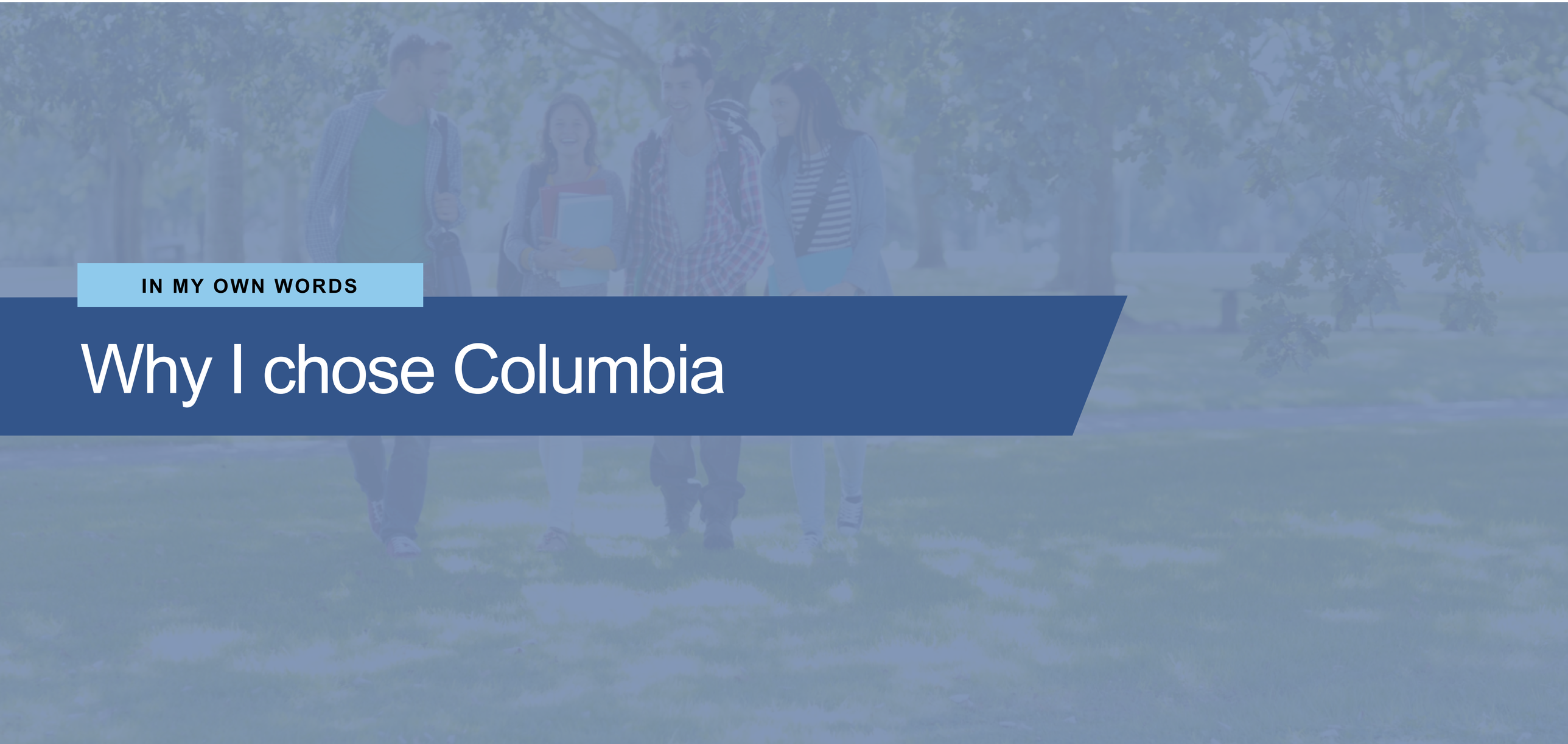
## Before Columbia

- [Clemson University]
  GRADUATE STUDENT

- [American International Group]
  DATA SCIENTIST

- [Commonwealth Bank]
  DATA SCIENTIST

POWERED BY liveguide



IN MY OWN WORDS

# Why I chose Columbia

POWERED BY liveguide

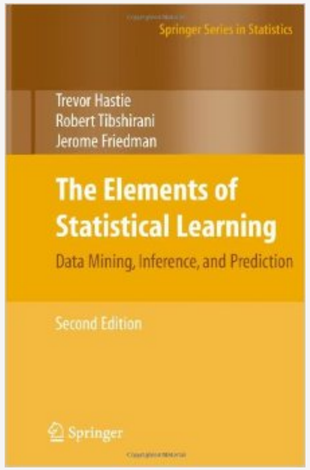
CLASS OF 2016

# Meet Francisco

## Columbia Favorites

[Machine Learning for Data Science]

COMS4721



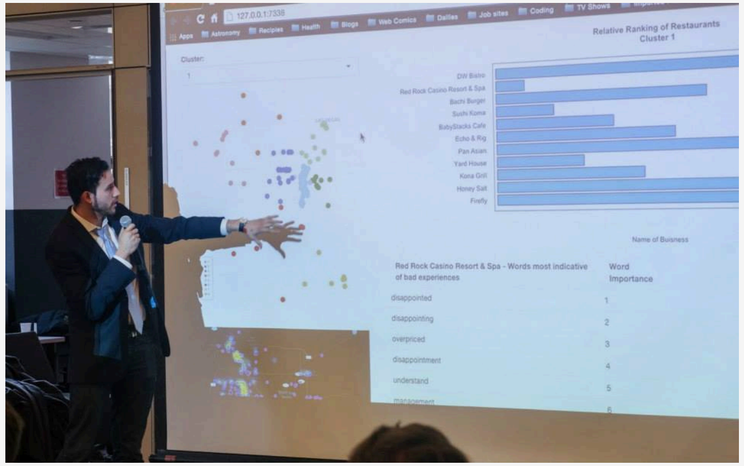

[Columbia Data Science Society]

STUDENT CLUBS

Columbia Data Science Society

[ICornellTech Hackathon with DSI Classmates]

MOMENT



POWERED BY liveguide

CLASS OF 2016

# Meet Francisco

## Where I am now?

[Commonwealth Bank]

**DATA SCIENTIST**

[Columbia University in the City of New York]

**SECOND YEAR STUDENT**

POWERED BY liveguide

CLASS OF 2016

# Meet Francisco

## 2 Truths and a Lie…

- I used to salsa dance

- I used to deliver pizza

- I used to compete as a boxer

# Top Questions Asked

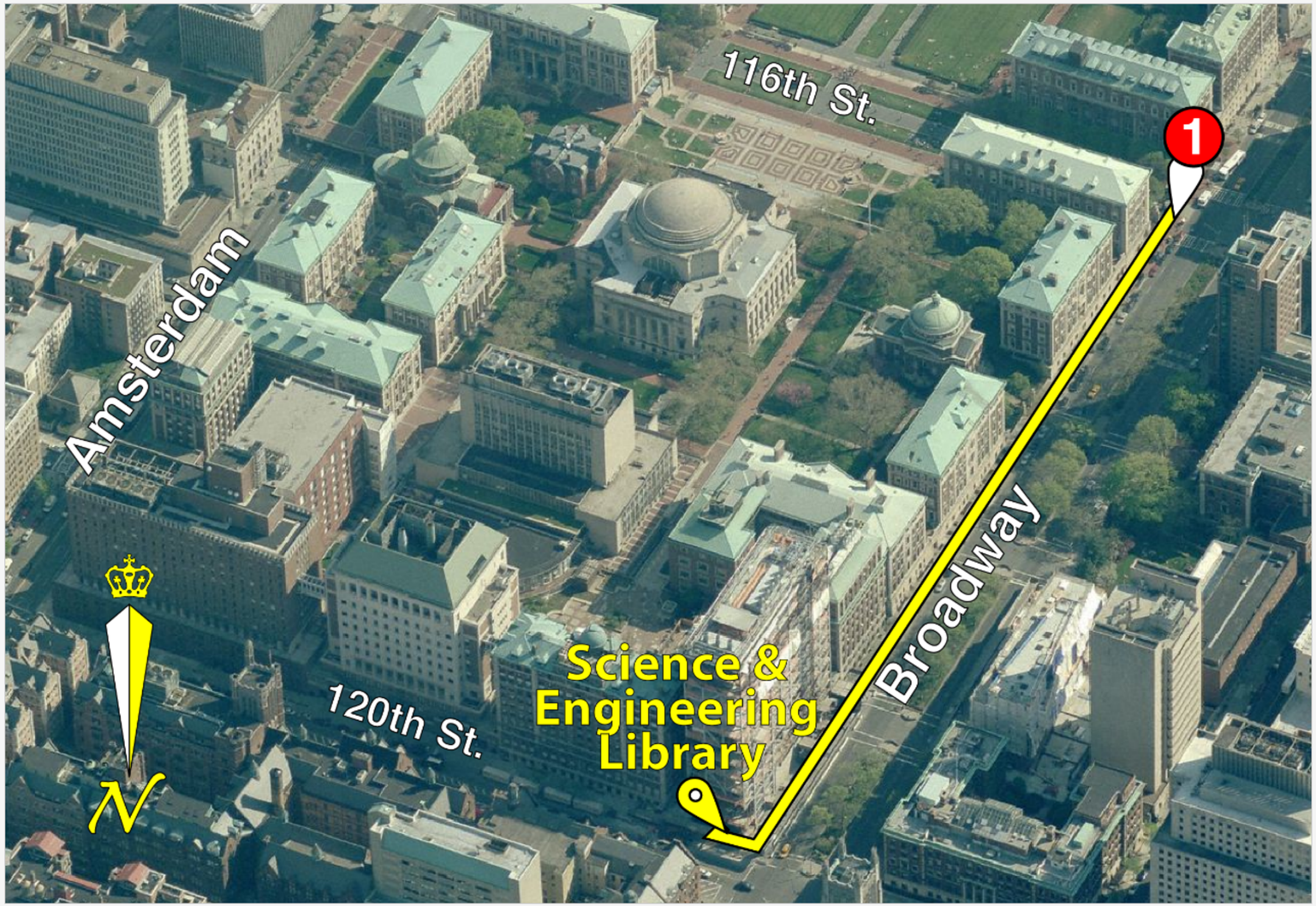
1. Do you think there was a benefit in going through the certificate program?
2. What skillsets has the program helped you hone and how have these skills impacted your work?
3. What are the career opportunities for data scientists?

# Agenda

- A Short Tour of Campus
- Academic Highlights
- Stories Outside the Classroom

POWERED BY liveguide

# A Short Tour of Campus



# Engineering Campus Highlights

POWERED BY liveguide

POWERED BY liveguide

# Engineering Campus Highlights

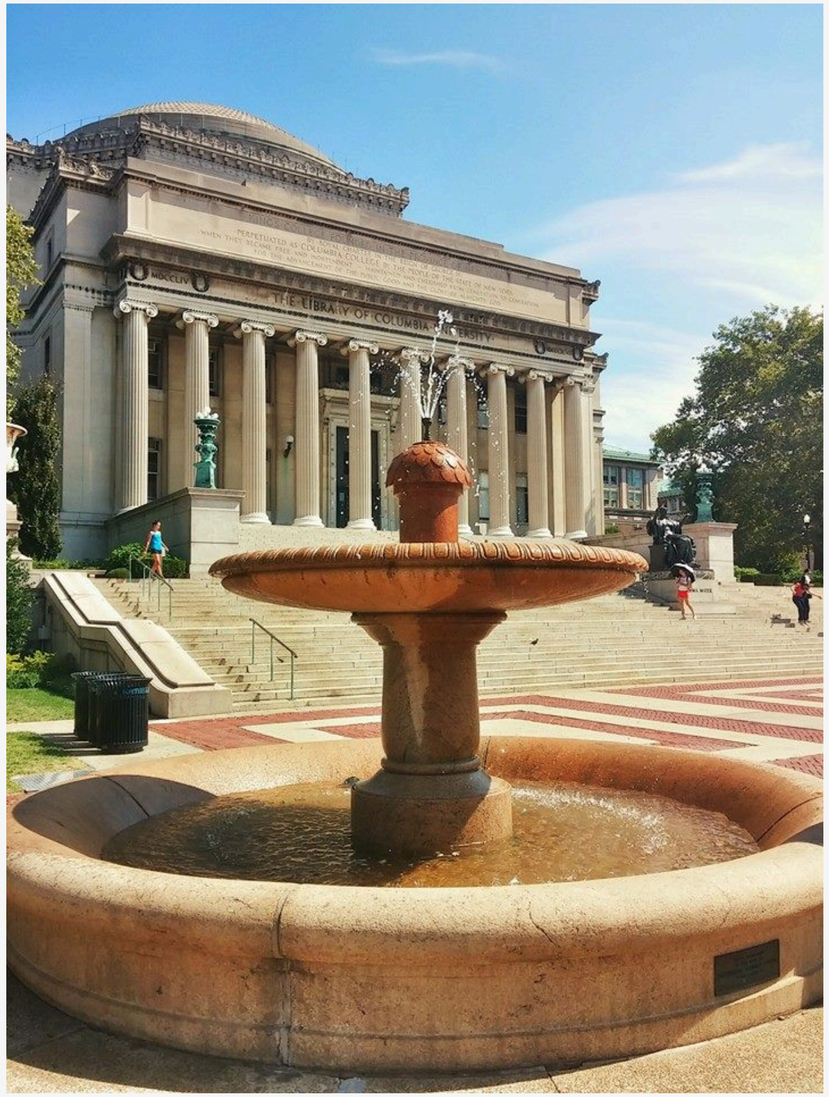
POWERED BY liveguide

POWERED BY liveguide

# Columbia University Campus

POWERED BY liveguide

POWERED BY liveguide



# Columbia University Campus

POWERED BY liveguide

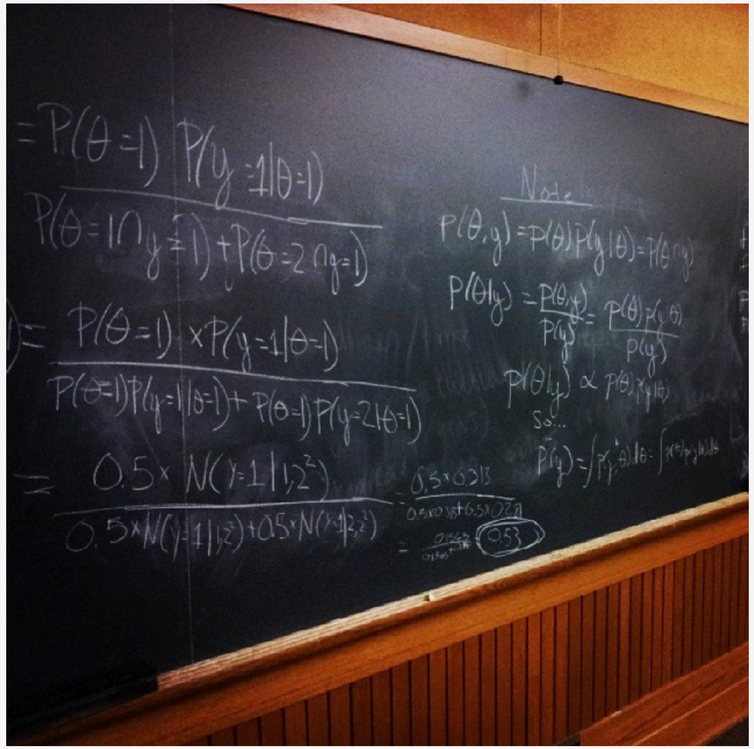
# My First Time on Campus

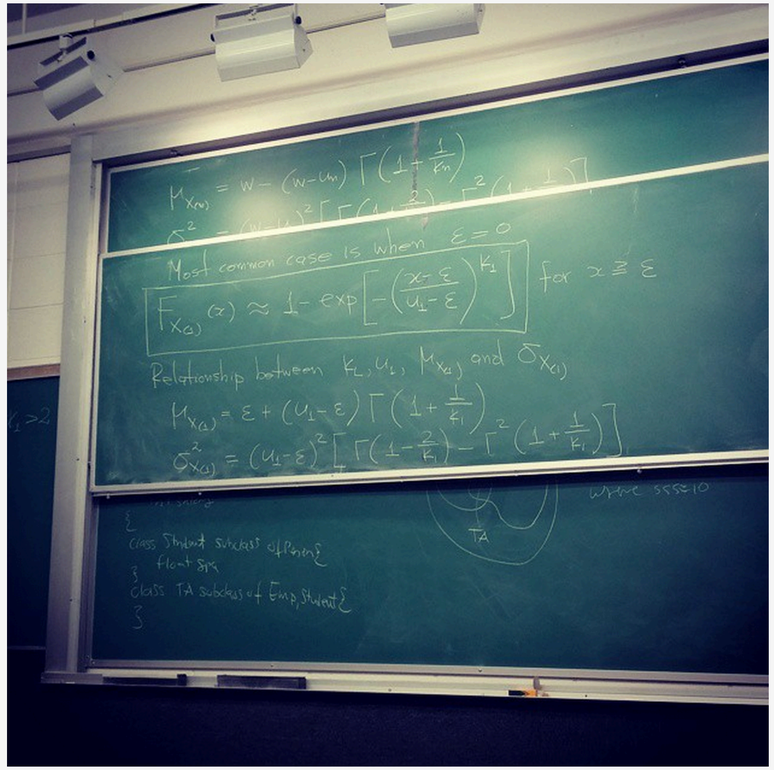
POWERED BY liveguide

POWERED BY liveguide

# The Part of Campus I Call Home

POWERED BY liveguide





Avery, Mudd, and Uris

POWERED BY liveguide

POWERED BY liveguide

# Questions

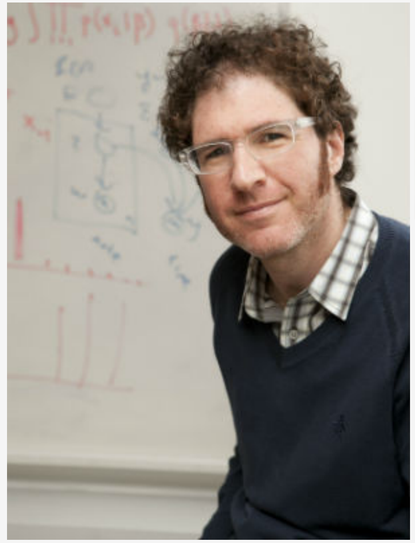
POWERED BY liveguide

# Agenda

- A Short Tour of Campus ✔
- Academic Highlights
- Stories Outside the Classroom

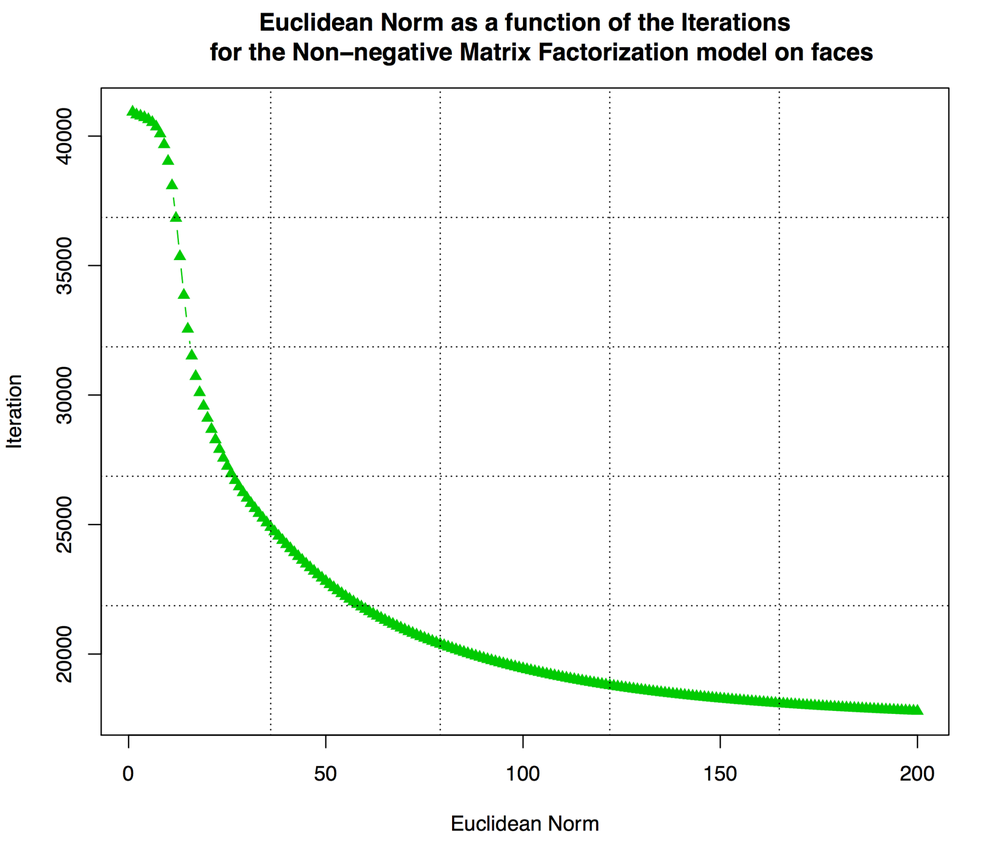
# Academic Highlights

# The People



Wonderful faculty and staff

# Machine Learning



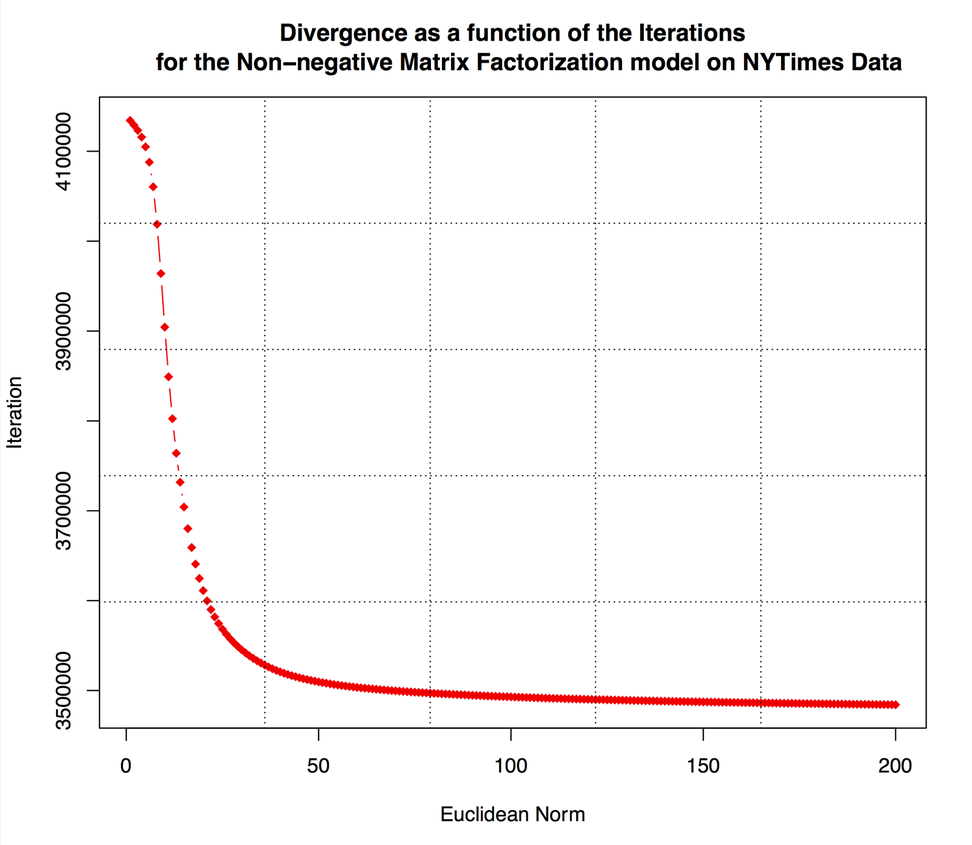

For 3 images in the data set, below is a plot of the original image and the column of W for which the corresponding weight in H is the largest.

Figure 4: Original Image and the corresponding max(W)

# Non-Negative Matrix Factorization & Image Processing

POWERED BY liveguide

# Machine Learning

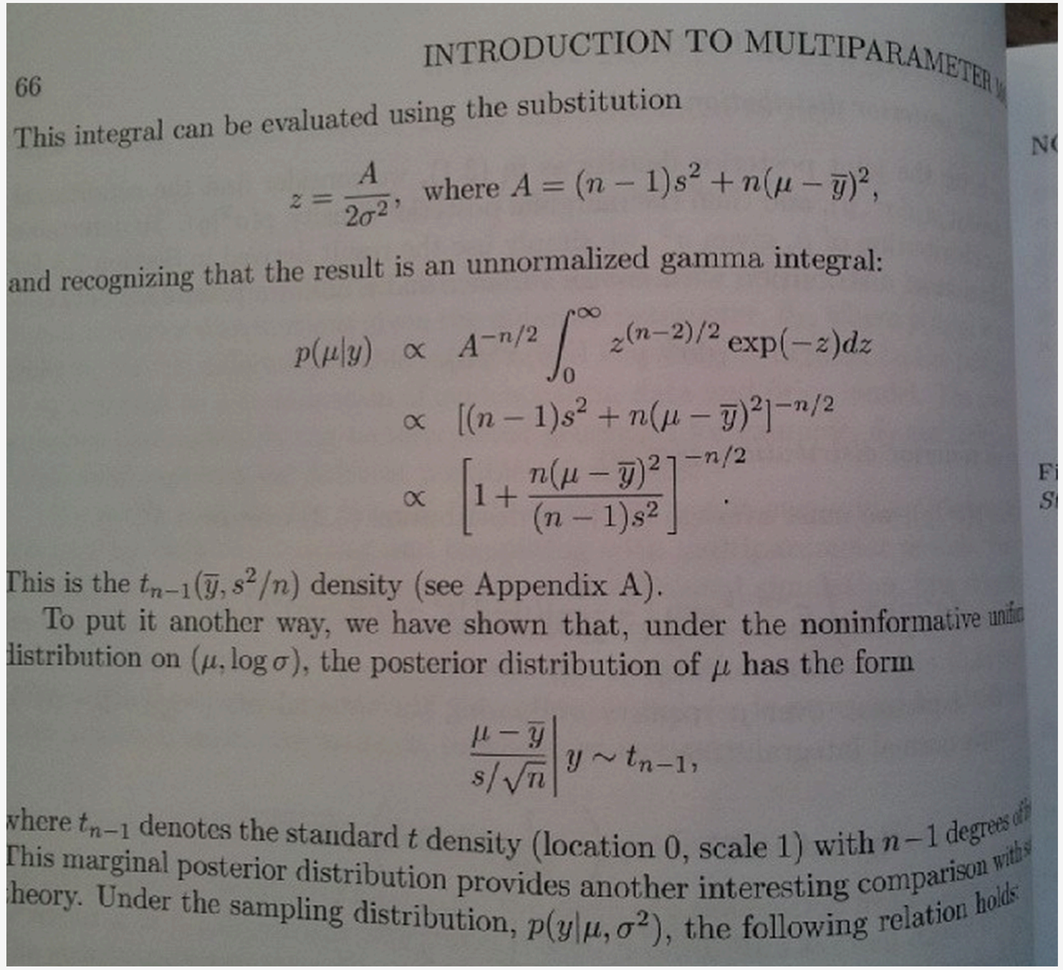



| Column 4 | Weight_4 | Column 8 | Weight_8 | Column 13 | Weight_13 | Column 22 | Weight_22 |
|---|---|---|---|---|---|---|---|
| school | 0.04494903 | wear | 0.0138742 | fact | 0.00757925 | father | 0.06229501 |
| student | 0.03212563 | man | 0.01361706 | piece | 0.00687064 | son | 0.05431684 |
| class | 0.01572991 | woman | 0.01308201 | page | 0.00673745 | mrs | 0.05260641 |
| college | 0.01476465 | stand | 0.00966269 | material | 0.00624996 | mother | 0.0406026 |
| program | 0.01415204 | white | 0.00929137 | case | 0.00571273 | graduate | 0.03266972 |
| child | 0.01401803 | black | 0.00843294 | different | 0.0055802 | daughter | 0.03244185 |
| education | 0.01366386 | dress | 0.00772902 | keep | 0.00547874 | marry | 0.02915262 |
| teacher | 0.01211609 | eye | 0.00717126 | small | 0.00501173 | president | 0.02329818 |
| community | 0.01131118 | young | 0.00675153 | little | 0.00497038 | retire | 0.02181934 |
| group | 0.01037284 | blue | 0.00625887 | surface | 0.00486741 | receive | 0.02173939 |

*Non-Negative Matrix Factorization*

Natural Language Processing

# Statistical Inference and Algorithms



66 INTRODUCTION TO MULTIPARAMETER

This integral can be evaluated using the substitution

$$z = \frac{A}{2\sigma^2}, \quad \text{where } A = (n-1)s^2 + n(\mu - \overline{y})^2,$$

and recognizing that the result is an unnormalized gamma integral:

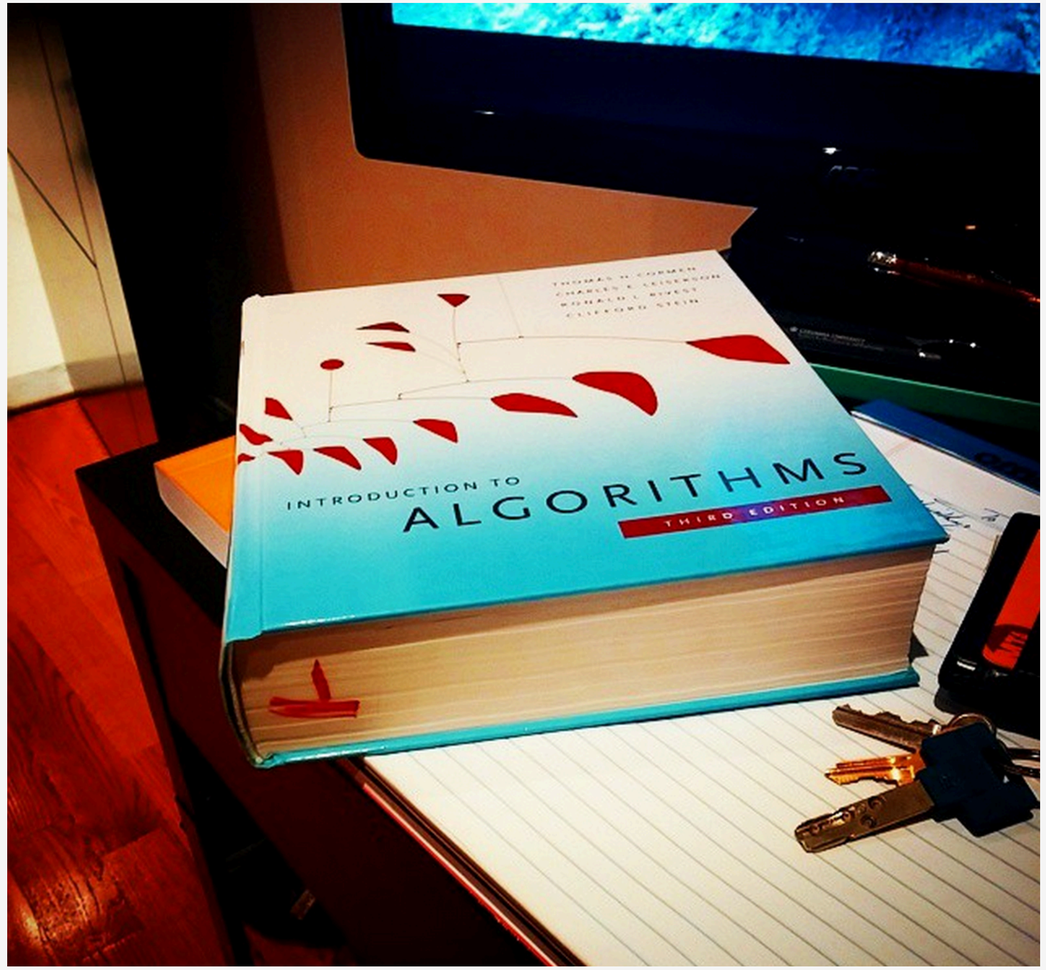
$$\begin{aligned} p(\mu|y) &\propto A^{-n/2} \int_0^\infty z^{(n-2)/2} \exp(-z) dz \\ &\propto [(n-1)s^2 + n(\mu - \overline{y})^2]^{-n/2} \\ &\propto \left[1 + \frac{n(\mu - \overline{y})^2}{(n-1)s^2}\right]^{-n/2}. \end{aligned}$$

This is the $t_{n-1}(\overline{y}, s^2/n)$ density (see Appendix A).

To put it another way, we have shown that, under the noninformative unifo distribution on $(\mu, \log \sigma)$, the posterior distribution of $\mu$ has the form

$$\left.\frac{\mu - \overline{y}}{s/\sqrt{n}}\right| y \sim t_{n-1},$$

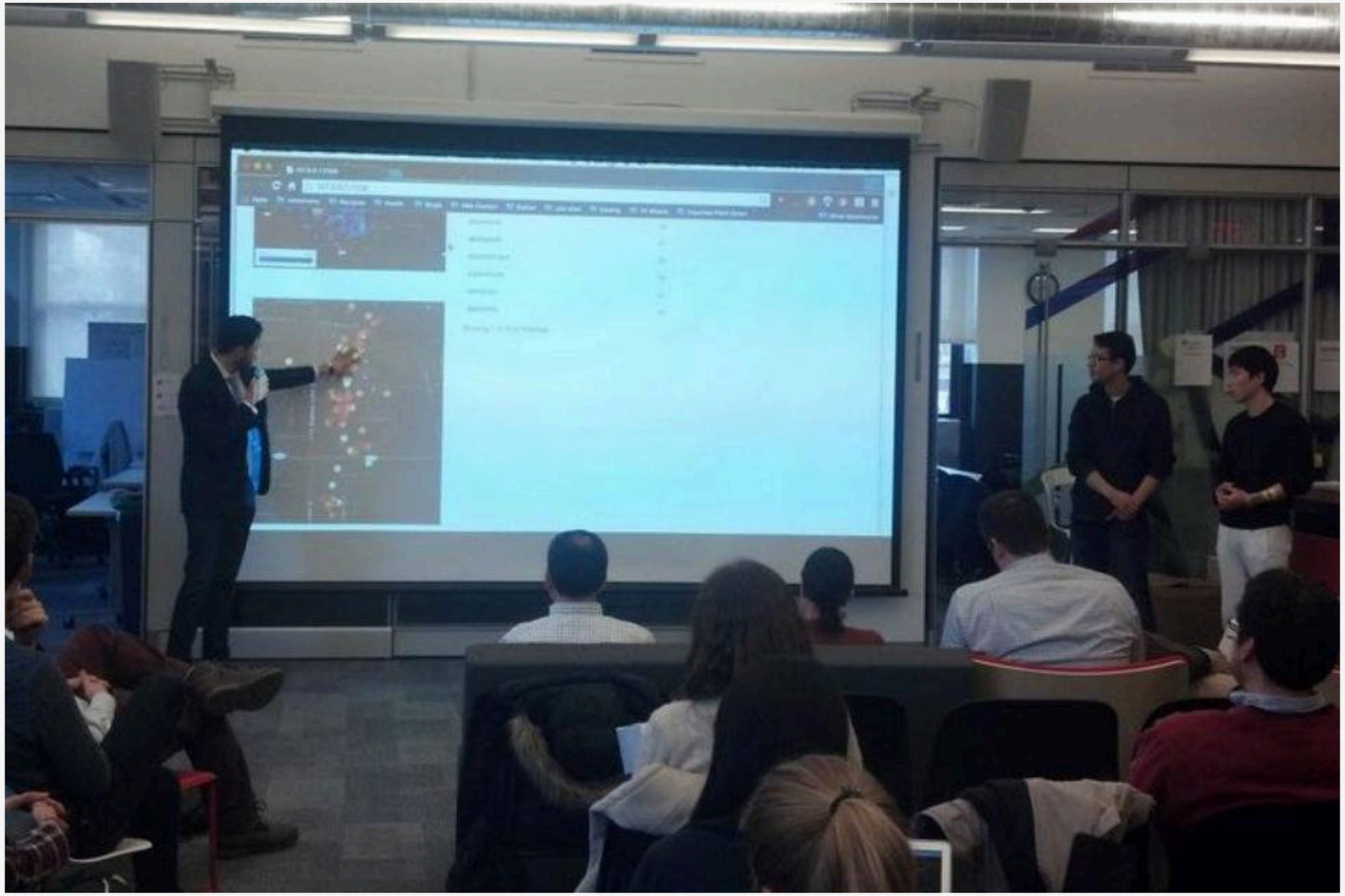
where $t_{n-1}$ denotes the standard $t$ density (location 0, scale 1) with $n-1$ degrees of fr This marginal posterior distribution provides another interesting comparison with s heory. Under the sampling distribution, $p(y|\mu, \sigma^2)$, the following relation holds:




Awesome books and material

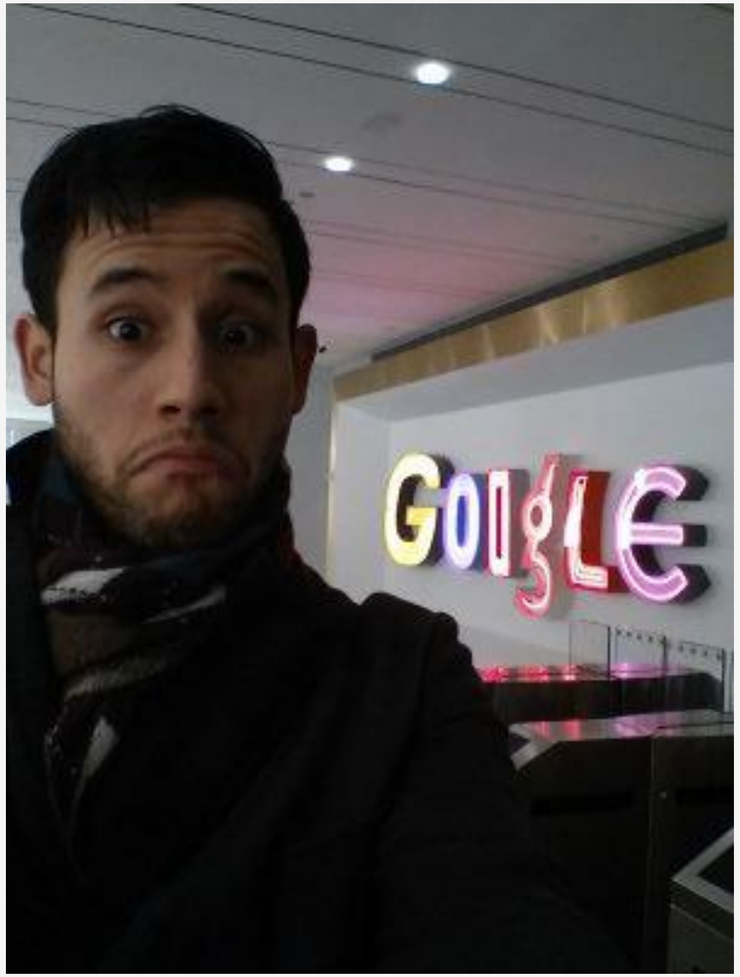
POWERED BY liveguide

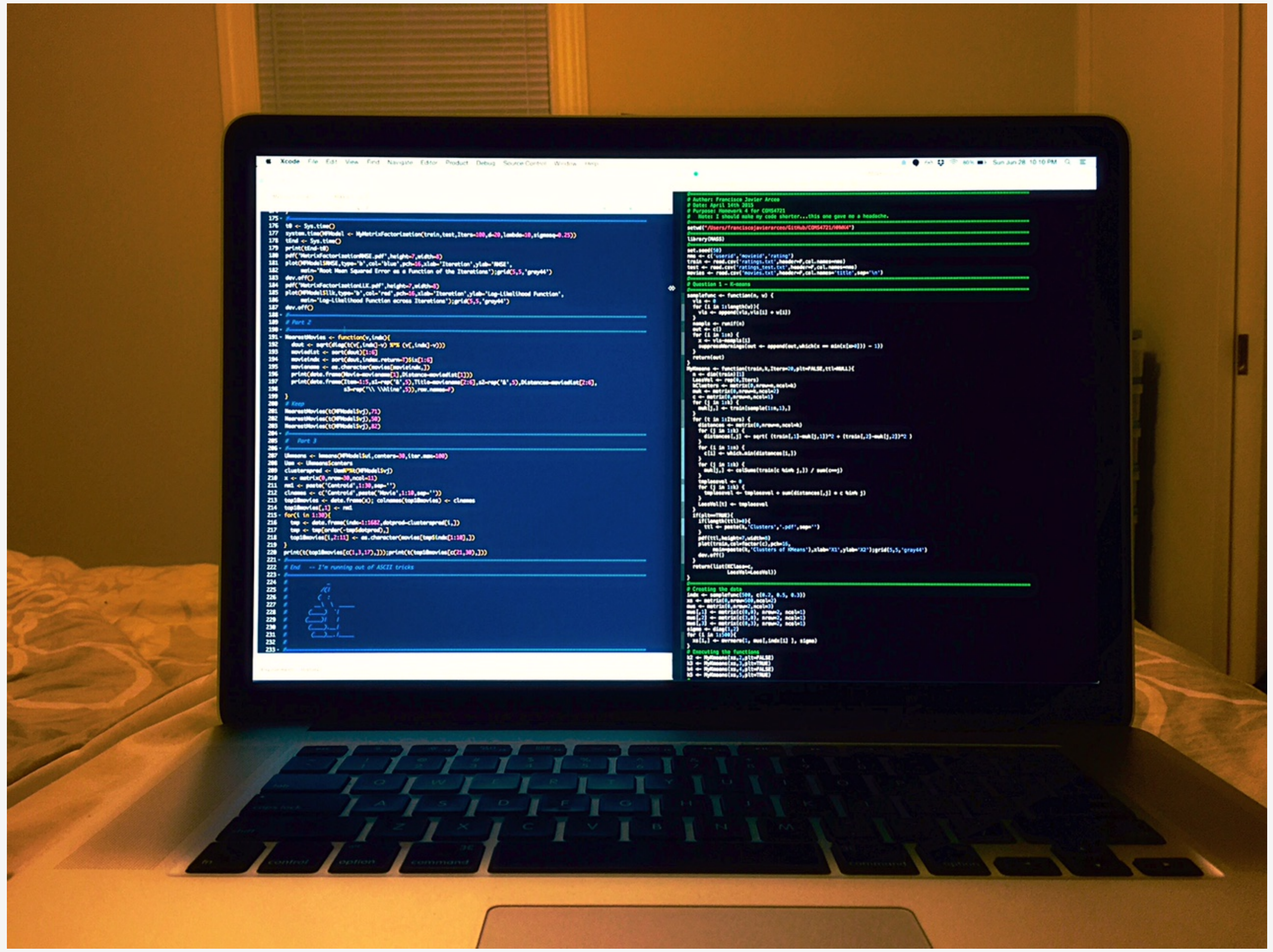
# Data Science Hackathon



DataLions!

# Data Science Hackathon



Data Science Hackathon

POWERED BY liveguide

POWERED BY liveguide

# Questions

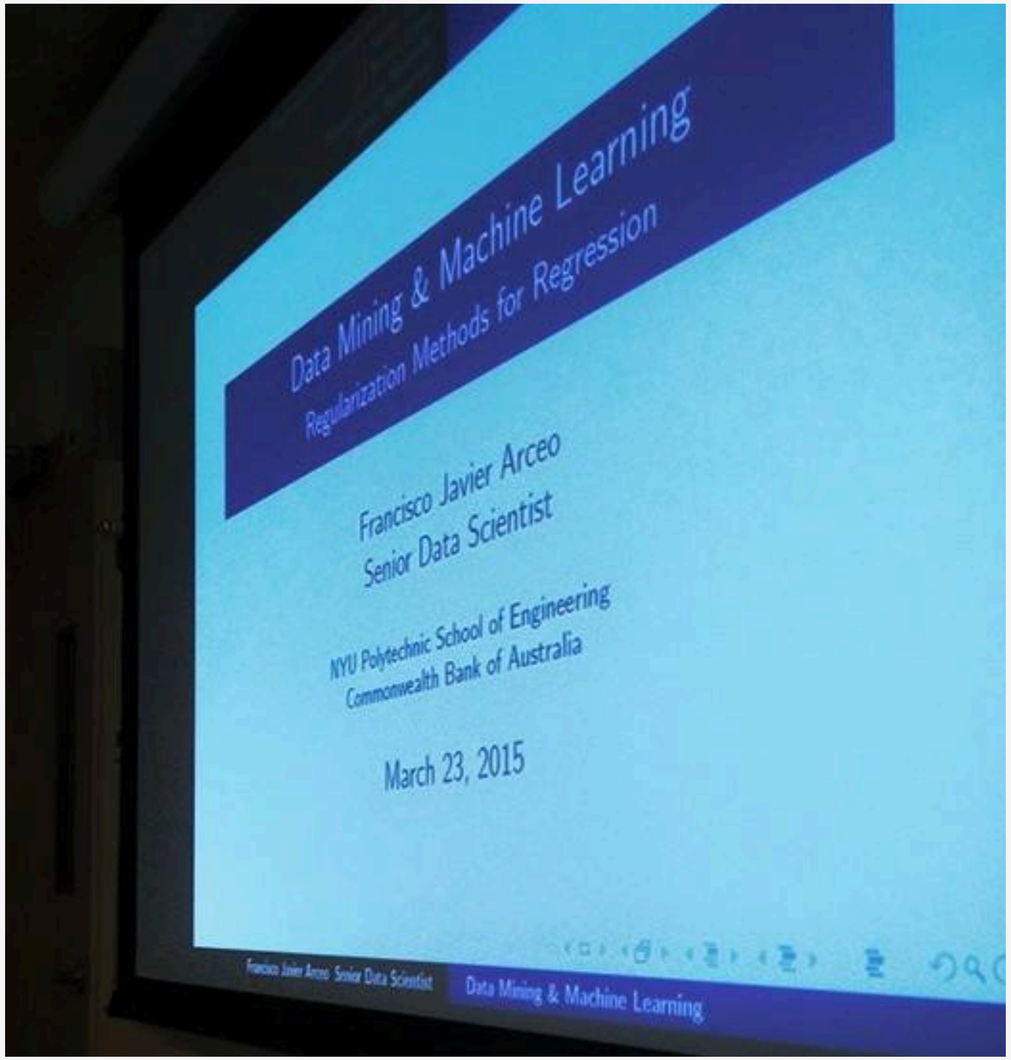
POWERED BY liveguide

# Agenda

- A Short Tour of Campus ✔
- Academic Highlights ✔
- Stories Outside the Classroom

POWERED BY liveguide

# Stories Outside the Classroom




# Invited Talks

POWERED BY liveguide

## LASSO Regression

$$\hat{\boldsymbol{\beta}}_{lasso}(\lambda) = S(\hat{\beta}, \lambda) \equiv sign(\hat{\beta})(|\hat{\beta}| - \lambda)_+ \tag{11}$$

$$S(\hat{\beta}, \lambda) = \begin{cases} \hat{\beta} - \lambda, & if\, \hat{\beta} > 0 \text{ and } \lambda < |\hat{\beta}| \\ \hat{\beta} + \lambda, & if\, \hat{\beta} < 0 \text{ and } \lambda < |\hat{\beta}| \\ 0, & if\, \lambda > |\hat{\beta}|. \end{cases}$$

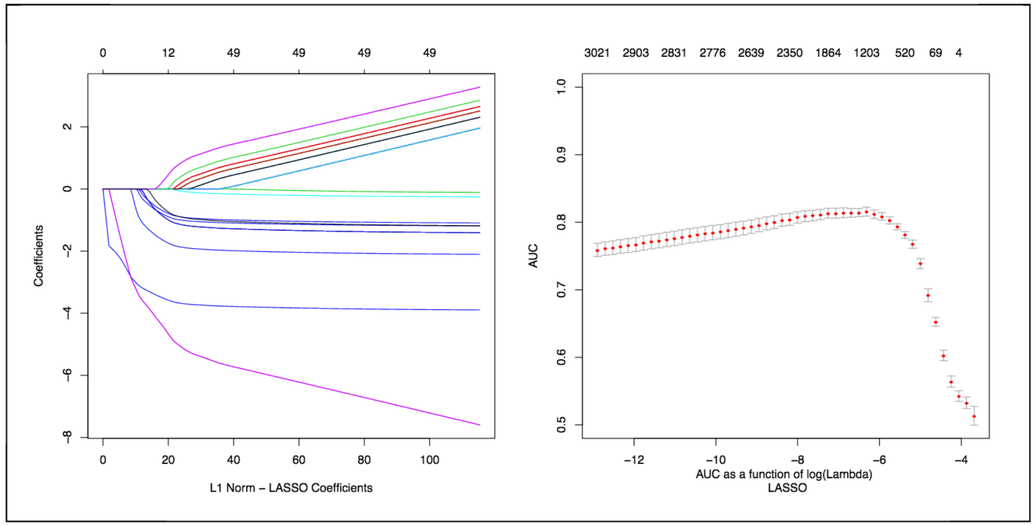
$$\hat{\beta}^t_{j,lasso} \leftarrow S(\hat{\beta}^{t-1}_{j,lasso} + \sum_{i=1}^{n} x_{ij}(y_i - \hat{y}_i{}^{t-1}), \lambda) \tag{12}$$

- Equation (11) is called the soft-threshold operator
- Apply soft-threshold operator as we cycle through our features at the $t^{th}$ iteration (12) updating the residual along the way
- LASSO does variable selection by setting select variables to 0!





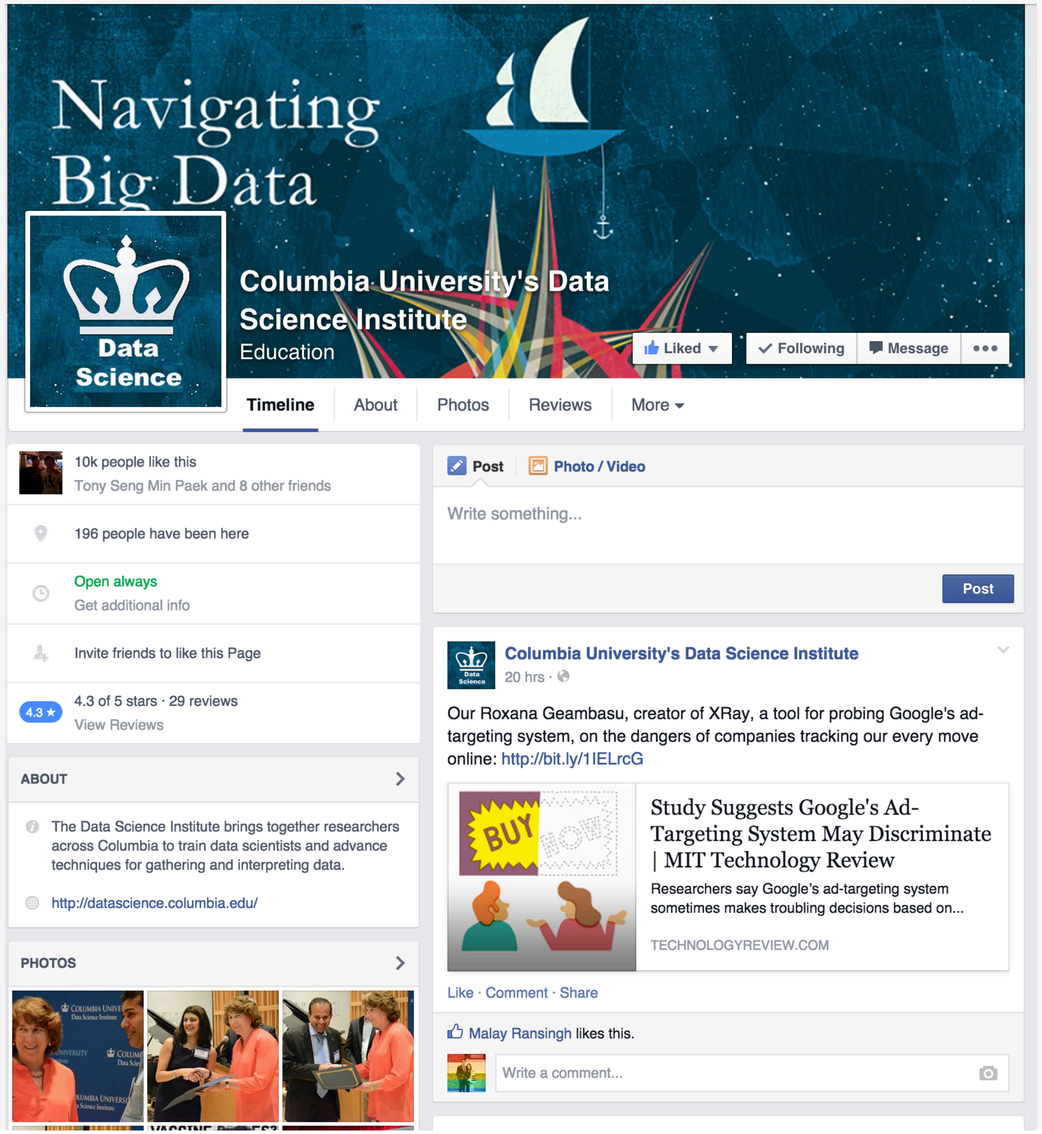
## LASSO Model Performance

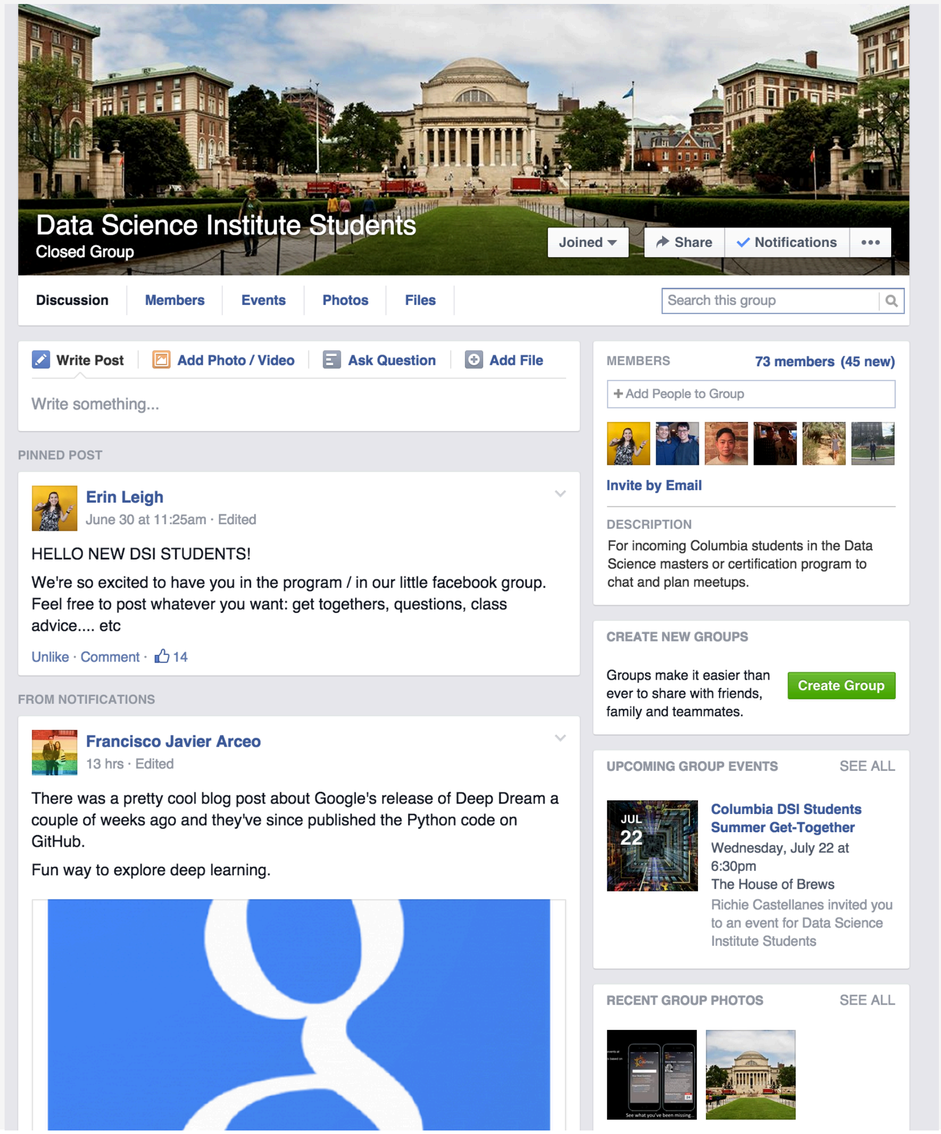





# Presentation at NYU Poly

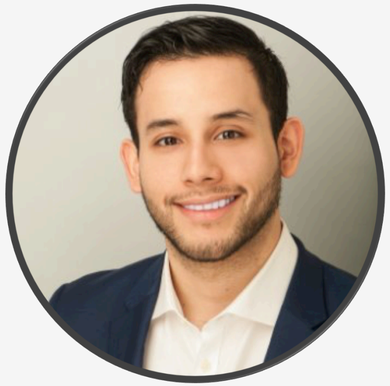
POWERED BY liveguide




# Great Online Community

POWERED BY liveguide



# A Great Network

POWERED BY liveguide

# Questions

POWERED BY liveguide

POWERED BY liveguide

CLASS OF 2016

# Francisco

## 2 Truths and a Lie…

- I used to salsa dance

- I used to deliver pizza

- I used to compete as a boxer

POWERED BY liveguide

# Follow Up with Me

- Admissions Website: http://datascience.columbia.edu/master-of-science-in-data-science
- My Email: fja2114@columbia.edu
- More on Francisco: www.franciscojavierarceo.github.io
- More on Liveguide: www.onliveguide.com