

# Queens College

## CSCI 381 / 780

### Data Analytics

Alex Pang  
Spring 2021

What do you think you are taking?

Data Science, Data Mining

Decision Science, Business Intelligence

Artificial Intelligence, Machine Learning

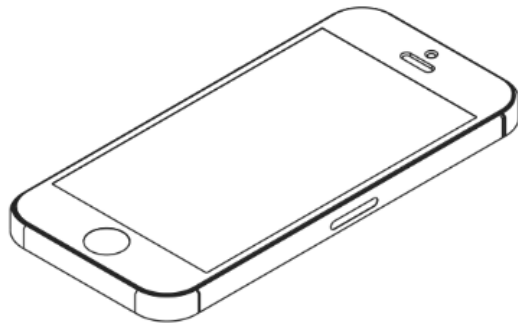
Statistical Learning, Data Analytics

Why are you taking this class?

# What are you taking this class

---

Go to **www.menti.com** and use the code 26 92 92 9



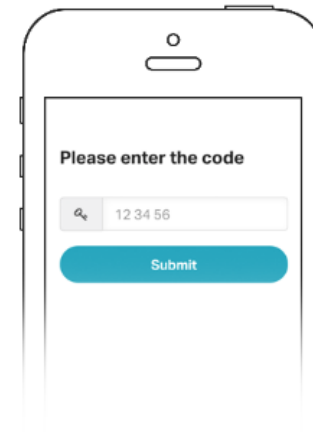
1

Grab your phone

www.menti.com

2

Go to [www.menti.com](http://www.menti.com)



3

Enter the code 91 44 43 and vote!

## Results

Go to [www.menti.com](https://www.menti.com) and use the code 26 92 92 9

# Why do you take this class?



## Data Science as a career?

---

California nurse meets baby she helped save 28 years ago – and he's a doctor



In 2012, Harvard Business Review said Data Scientist will be the Sexiest Job of the 21<sup>st</sup> Century



DATA

## Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE



## Data Science as a career (2019 survey)

---

### The 25 highest-paying entry-level jobs for college grads



1. Data scientists make \$95,000 a year.



2. Software engineers make \$90,000 a year.



4. Investment banking analysts make \$85,000 a year.





# High Demand – Who is hiring?

Answer: EVERYONE Old and New companies



**Senior Data Scientist, Analytics - Trust**  
Airbnb  
San Francisco, CA  
via AngelList

🕒 2 days ago 🧑‍💼 Full-time



**Data Analyst Intern**  
PayPal  
New York, NY  
via Glassdoor

🕒 5 days ago 🧑‍💼 Internship



**Customer Engineer, Data & Analytics**  
Microsoft  
New York, NY  
via Microsoft Jobs

🕒 23 days ago 🧑‍💼 Full-time



**Data Scientist**  
Microsoft  
New York, NY  
via LinkedIn



**Data Scientist, Analytics**  
Facebook  
New York, NY  
via Facebook

🕒 Over 1 month ago 🧑‍💼 Full-time



**Data Scientist, Product Analytics**  
Facebook  
New York, NY  
via Facebook

**Data Scientist, Product Analytics**

FB Facebook  
New York, NY

[Apply on Facebook](#)

[Apply on FOX8 Jobs](#)

[Apply on JobServe](#)

[Apply on Indeed](#)



IBM Watson™



C3.ai

**Data Scientist/ Predictive Analytics Actuary job**

🏢 [GENERAL MOTORS](#) 📍 WARREN, MI



**Data Scientist**

Redwood City, California; New York, New York

As a Data Scientist, you will participate in the definition of new analytics capabilities able to provide our customers with the



[https://www.youtube.com/watch?v=jgj6FYghPh8&ab\\_channel=BeyondourMoney](https://www.youtube.com/watch?v=jgj6FYghPh8&ab_channel=BeyondourMoney)

CSCI 381/780  
Data Analytics  
Your instructor

---



# Reality vs Expectation

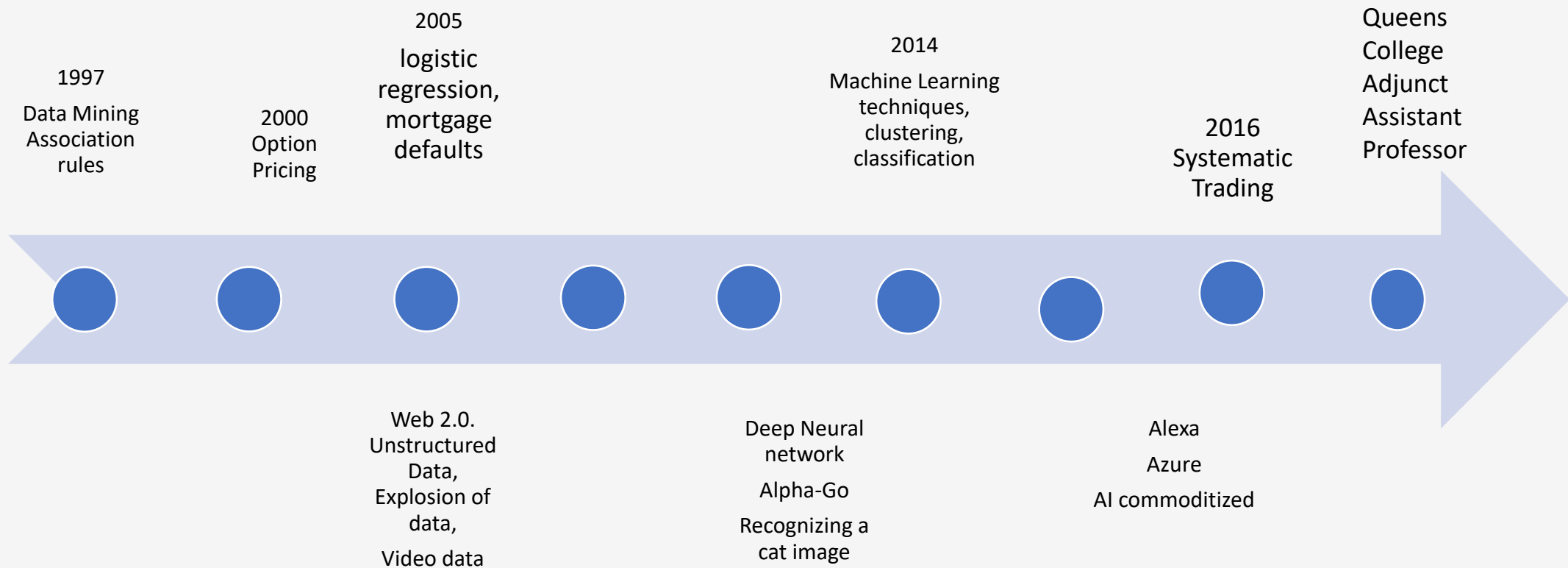
---



# CSCI 381/780 Data Analytics – Instructor Bio

---

## My Life As a Quant, Statistician and Data Scientist





# Data Mining – where it all started



## Association Rule Mining

- Given a set of transactions, find rules that will predict the occurrence of an item based on the occurrences of other items in the transaction

### Market-Basket transactions

TID	Items
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

### Example of Association Rules

$\{\text{Diaper}\} \rightarrow \{\text{Beer}\},$   
 $\{\text{Milk, Bread}\} \rightarrow \{\text{Eggs, Coke}\},$   
 $\{\text{Beer, Bread}\} \rightarrow \{\text{Milk}\},$

Implication means co-occurrence,  
not causality!

Remember Target Marketing these days?

# CSCI 381/780 Data Analytics – Instructor Bio





---

## Exploratory Mining and Pruning Optimizations of Constrained Associations Rules

Article

Full-text available

Jan 2002 · ACM SIGMOD Record

 Raymond T. Ng ·  Laks Lakshmanan ·  Alex Pang ·  Jiawei Han

From the standpoint of supporting human-centered discovery of knowledge, the present-day model of mining association rules suffers from the following serious shortcomings: (i) lack of user exploration and control, (ii) lack of focus, and (iii) rigid notion of relationships. In effect, this model functions as a...

[Show more](#)







## Optimization of Constrained Frequent Set Queries with 2-variable Constraints

Article

Full-text available

Nov 1999 · ACM SIGMOD Record

 Laks Lakshmanan ·  Raymond T. Ng ·  Jiawei Han · [...] ·  Alex Pang

Currently, there is tremendous interest in providing ad-hoc mining capabilities in database management systems. As a first step towards this goal, in [15] we proposed an architecture for supporting constraint-based, human-centered, exploratory mining of various kinds of rules including associations, introduced the notion of constrained frequent se...

[Show more](#)



# CSCI 381/780 Data Analytics – Instructor Bio

---

## Quantitative Analyst, Technologist and Educator

- Lehman Brothers, Merrill Lynch, Mortgage originator and investor and REIT
- Multiple multi-billion hedge funds
- Financial software start-up
- ...
- Friends For Impact (non-profit organization) advisors
- Complete Foundation Creative Technology Center
- Queens College Tech Incubator volunteer/mentors
- Two times outstanding teaching assistants awards (Physics and Computer Science)

Me:

Love to share what I know  
Love to help in whatever ways possible

You:

Learn something incredibly useful for your  
career and have fun at the same time