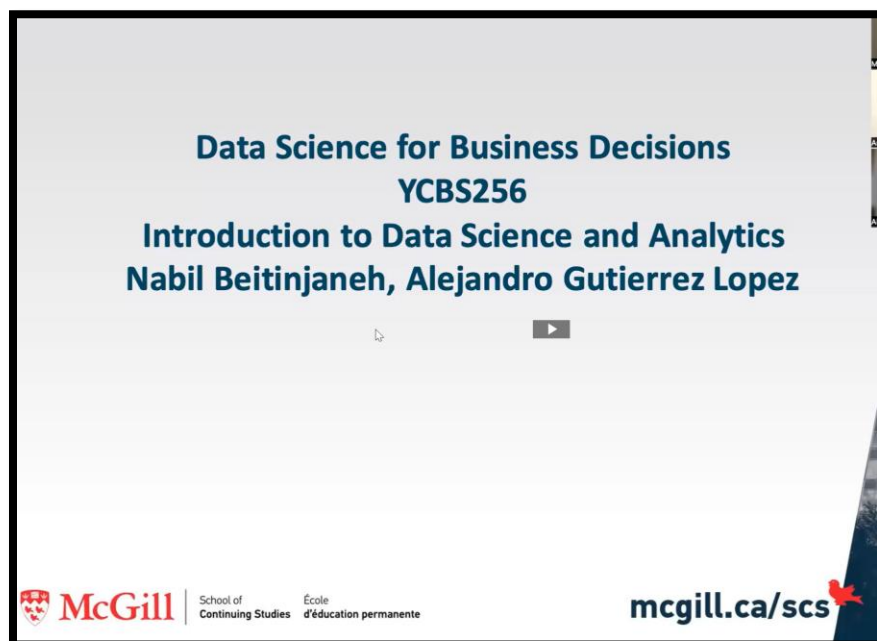
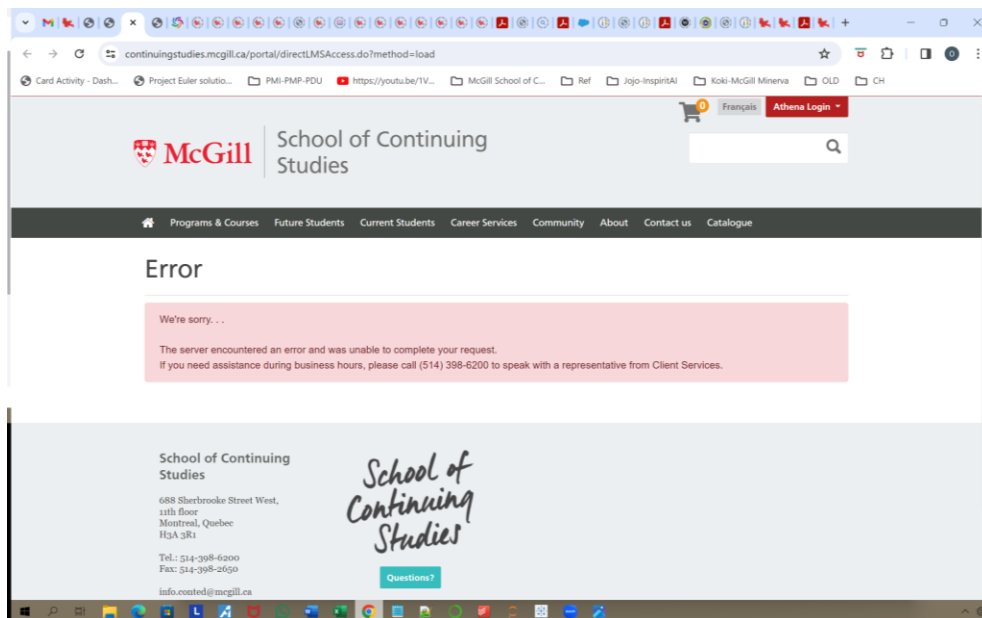


what is phone # and email address for IT

Michelangelo Chouinard 7:29 PM
514-398-6200

extension 1 > 4 > 2 if I remember

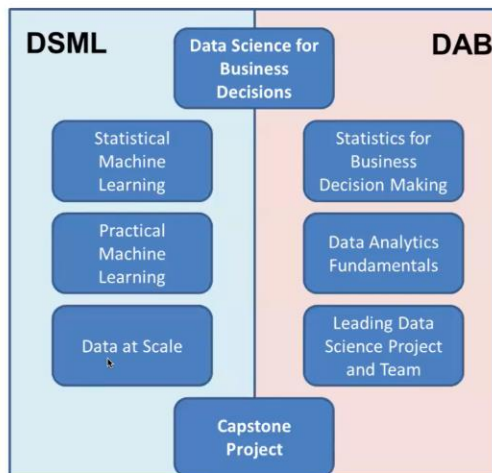
askd.scs@mcgill.ca



Session Overview

- Class Introductions
- Course fit within the overall program
- Course logistics (preparation, synchronous & asynchronous sessions,
 - myCourses
 - The Course Outline
 - Assessments, assignments, participation
 - Tools (Alteryx and Tableau)
 - DataCamp access
 - Computer Lab 1285
 - How to reach us
- Why Data Science and Analytics?

Recommended Path(s) for McGill Certificate




Skills

Please refer to the Study Plan document shared in the McGill One Drive

Study Plan

Data Science &
Machine Learning

Data Analytics
for Business



Grid of video call participants including Kayla Gauthier, Kevin D., Amir Kumar, Minelle El Haddad, Anas Khan, and Anayman Nascimbeno.


McGill | School of Continuing Studies | École d'éducation permanente

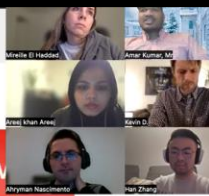
4

Advantages

Q

If you were a top long distance runner, how much are you willing to pay to buy a special pair of shoes that will shave off a couple of minutes of a marathon?





Grid of video call participants including Minelle El Haddad, Anas Khan, Kevin D., Anayman Nascimbeno, and Han Zhang.

McGill | School of Continuing Studies | École d'éducation permanente

5

Advantages

Q If you were a top long distance runner, how much are you willing to pay to buy a special pair of shoes that will shave off a couple of minutes of a marathon?

A Such a shoe exists...it's called the Vaporfly.

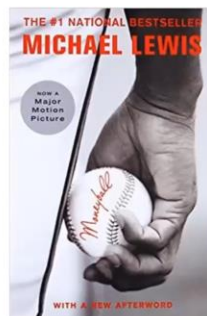
[Do Nike's New Shoes Give Runners an Unfair Advantage?](#)

[Do Nike's New Marathon Shoes Actually Make you Run Faster?](#)

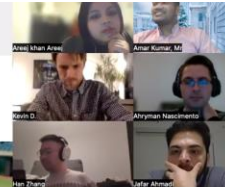


Why Data Science?

2003



2011



Zoom Meeting You are viewing Peter Komar, MSc's screen View Options

Recording

Frederick... Josh Gall... Han Zhang J Joshua Akue Michelange Cho...

GM Looked for Undervalued Players

- Believed that winning 90 games would help get the team into the finals
- Had a limited budget
- Used non-standard statistics to identify undervalued players
 - On-base percentage is important
 - Slugging percentage is important
 - Batting average is over-valued

Salaries and Wins, 1998-2001

Average yearly wins

Average yearly salary (in millions)

McGill School of Continuing Studies Ecole d'Études permanentes

8

Recording

After Publication of Book / Adoption of Methods

Still a strong team but difference is smaller!

Source : The Analytics Edge, Coursera

Salaries and Wins, 2002-2012

Average yearly wins

Average yearly salary (in millions)

McGill School of Continuing Studies Ecole d'Études permanentes

9

Zoom Meeting

You are viewing Amir Komar, MSc screen


View Options

Recording


Kajla Gauthier, Yusef Gulam, Frederic B., Han Zhang, Jayang Zhang, Joshua Akse

Data Science

"The two most important things in Data Science are the 'data' and the 'science'."
- Kirk D Borne (@kirkdborne)



"Data Science is the art of turning data into actions. It's all about the tradecraft. Tradecraft is the process, tools and technologies for humans and computers to work together to transform data into insights."



McGill School of Continuing Studies Enroll Education permanently

10

The book in the mcgill library I can access with mcgill ID

Zoom Meeting

Participants (11)



Find a participant

Han Zhang, Jafar Almadhi, Jayang Zhang, Joshua Akse, Jung Bae, Kajla Gauthier, Kevin D., Minsang Mangul, Mervyn Farhang, Michelangelo Chouard, Minville El Haddad, Nadia, Olga Sabatini, Rishi Jain, Paul Charlton, Ralph Bow-Fadel, Ronali, Samah Lurabi, Samuel Brink

Analytic Techniques (1/2)

List of techniques, algorithms, approaches, etc.

The Field Guide to Data Science, 2015

McGill School of Continuing Studies Enroll Education permanently

11

You can now send non-urgent feedback. Click "Give feedback" on the toolbar.

Give Feedback

Zoom Meeting

Osama Wanis Suparna Baral Joshua Akue PARTH Jain

Recording

Analytic Techniques (2/2)

Source: [Four Types of Analytics and Cognition](#)

McGill School of Continuing Studies École d'éducation permanente

12

Zoom Meeting

Osama Wanis Suparna Baral Joshua Akue PARTH Jain

Recording

Data to Insight (1/3)

McGill School of Continuing Studies École d'éducation permanente

13

Zoom Meeting

Osama Wanis | Amar Kumar, Mr | Kevin D. | Suparna Barai | Joshua Akue | PARTH Jain

Recording

Data to Insight (2/3)

2019 This Is What Happens In An Internet Minute

Source: [Visual Capitalist](#)

McGill School of Continuing Studies École d'éducation permanente

14

Zoom Meeting

You are viewing Amar Kumar, Mr's screen View Options

Osama Wanis | Amar Kumar, Mr | Kevin D. | Suparna Barai | Joshua Akue | PARTH Jain

Recording

Data to Insight (3/3)

2018 This Is What Happens In An Internet Minute

2019 This Is What Happens In An Internet Minute

Source: [Visual Capitalist](#)

McGill School of Continuing Studies École d'éducation permanente

15

Zoom Meeting

Participants (11)

Find a participant

Participants:

- Yan Zhang
- Yan Zhang
- Yan Zhang
- Yan Zhang
- Yan Zhang
- Yan Zhang
- Yan Zhang
- Yan Zhang
- Yan Zhang
- Yan Zhang
- Yan Zhang

The 4Vs of Big Data

Volume
SCALE OF DATA

Variety
DIFFERENT FORMS OF DATA

Velocity
ANALYSIS OF STREAMING DATA

Veracity
IMPORTANCE OF DATA

The FOUR V's of Big Data

Source: [the 4Vs of Big Data](#)

IBM

McGill School of Continuing Studies

16

Recording

The 5th V is Value

OUR CONSULTANT HAS BEEN MINING DATA ALL DAY. THE RESULTS ARE QUITE SHOCKING.

ACCORDING TO THE DATA, SALES ARE ALWAYS HIGHEST WHEN I DO THIS...

www.dilbert.com

Source: [The 5th V is Value](#)

Reading: [The Value of Real-Time Data Analytics](#) – Forbes

McGill School of Continuing Studies

17

Zoom Meeting

Osama Waniis Kevin D. Joshua Akue PARTH Jain

Recording

Is the Scientific Method Obsolete?

Are theories and models useless?
Should we focus on getting more data or building better algorithms?

"...faced with massive data, this approach to science – hypothesize, model test – is becoming obsolete....who knows why people do what they do? The point is they do it, and we can track and measure it..."

[The End of Theory](#) (Wired)
[More data or better models?](#)
[Could Big Data be the end of theory in science?](#)

Thanks to John Nikolopoulos for sharing!

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

'Petabytes allow us to say: "Correlation is enough" (...)

~~We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.'~~

Chris Anderson, WIRED.com, 2008
http://archive.wired.com/wired/archive/16.03/ai_theory

McGill School of Continuing Studies École d'éducation permanente

18

Check the post/sharing by John Nikolopoulos

Recording

Is it about Data Intensive Computing?

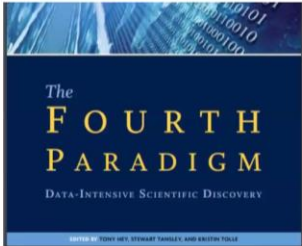
"This book is about a new, fourth paradigm for science based on data intensive computing.

Data-intensive science consists of three basic activities: capture, curation, and analysis.

Data analysis covers a whole range of activities throughout the workflow pipeline, including the use of databases, analysis and modeling, and then data visualization."

[Read the Review](#)
[More on Science and Data](#)

Thanks to John Nikolopoulos for sharing!



The
**FOURTH
PARADIGM**
DATA-INTENSIVE SCIENTIFIC DISCOVERY

EDITED BY TONY HEY, STEWART TANSLEY, AND EUSEBIO VALLE

McGill School of Continuing Studies École d'éducation permanente

19

Check also the above

Is there a Unifying Theory?

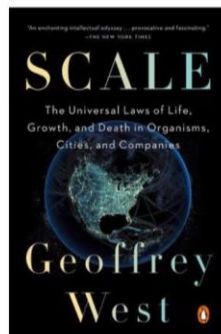
"The basic mathematical laws of physics governing growth in the physical world apply equally to biological, political and corporate organisms. On its face, his book's objective is to contribute to an overarching behavioral science of what it calls *highly complex systems*."

~ Jonathan A. Knee

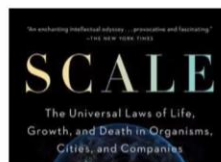
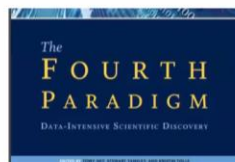
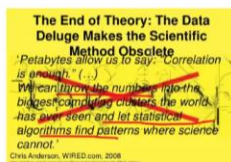
[From the New York Times](#)

[The hidden maths of organisms, cities and companies \(The Economist\)](#)

Thanks to John Nikolopoulos for sharing his reading list!



What 'Science' are we talking about?



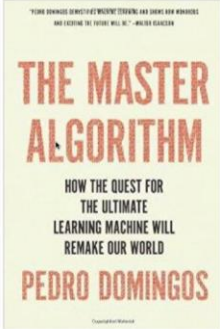
Where does Data Science fit on this spectrum?

Now for Machine Learning

The Master Algorithm *"is an introduction to machine learning for a non-technical audience. It describes the five major paradigms in machine learning - symbolism, connectionism, evolutionism, Bayesianism and analogism - including the main ideas, the history, the leading researchers, applications, strengths and weaknesses, etc. It's framed as a quest for the eponymous master algorithm - a general-purpose learner capable of learning anything from data. In that search, we visit each of the five schools of thought to see what we can learn from it, and then we try to put all the pieces together into one overarching algorithm that is not the master algorithm yet, but hopefully brings us a step closer."*

~ Pedro Domingos

[The Master Algorithm](#)



22

The Master Algorithm



Data Scientists

	Data science, due to its interdisciplinary nature, requires an intersection of abilities: hacking skills , math and statistics knowledge , and substantive expertise in a field of science.
	Hacking skills are necessary for working with massive amounts of electronic data that must be acquired, cleaned, and manipulated.
	Math and statistics knowledge allows a data scientist to choose appropriate methods and tools in order to extract insight from data.
	Substantive expertise in a scientific field is crucial for generating motivating questions and hypotheses and interpreting results.
	Traditional research lies at the intersection of knowledge of math and statistics with substantive expertise in a scientific field.
	Machine learning stems from combining hacking skills with math and statistics knowledge, but does not require scientific motivation.
	Danger zone! Hacking skills combined with substantive scientific expertise without rigorous methods can beget incorrect analyses.

Source: [How to Become a Data Scientist Before you Graduate](#)

McGill School of Continuing Studies Education permanente

23

Data Scientists

	Data science, due to its interdisciplinary nature, requires an intersection of abilities: hacking skills , math and statistics knowledge , and substantive expertise in a field of science.
	Hacking skills are necessary for working with massive amounts of electronic data that must be acquired, cleaned, and manipulated.
	Math and statistics knowledge allows a data scientist to choose appropriate methods and tools in order to extract insight from data.
	Substantive expertise in a scientific field is crucial for generating motivating questions and hypotheses and interpreting results.
	Traditional research lies at the intersection of knowledge of math and statistics with substantive expertise in a scientific field.
	Machine learning stems from combining hacking skills with math and statistics knowledge, but does not require scientific motivation.
	Danger zone! Hacking skills combined with substantive scientific expertise without rigorous methods can beget incorrect analyses.

Source: [How to Become a Data Scientist Before you Graduate](#)

McGill School of Continuing Studies Education permanente

23



Danger zone! Hacking skills combined with substantive scientific expertise without rigorous methods can beget incorrect analyses.

23

Recording Sign in

The Third Wave Data Scientist

Source: [The Third Wave Data Scientist](#)

McGill School of Continuing Studies Enroll Education permanently

24

Recording Sign in

No Ordinary Disruption: Class Discussion

The 4 Global Forces breaking All the Trends

McGill School of Continuing Studies Enroll Education permanently

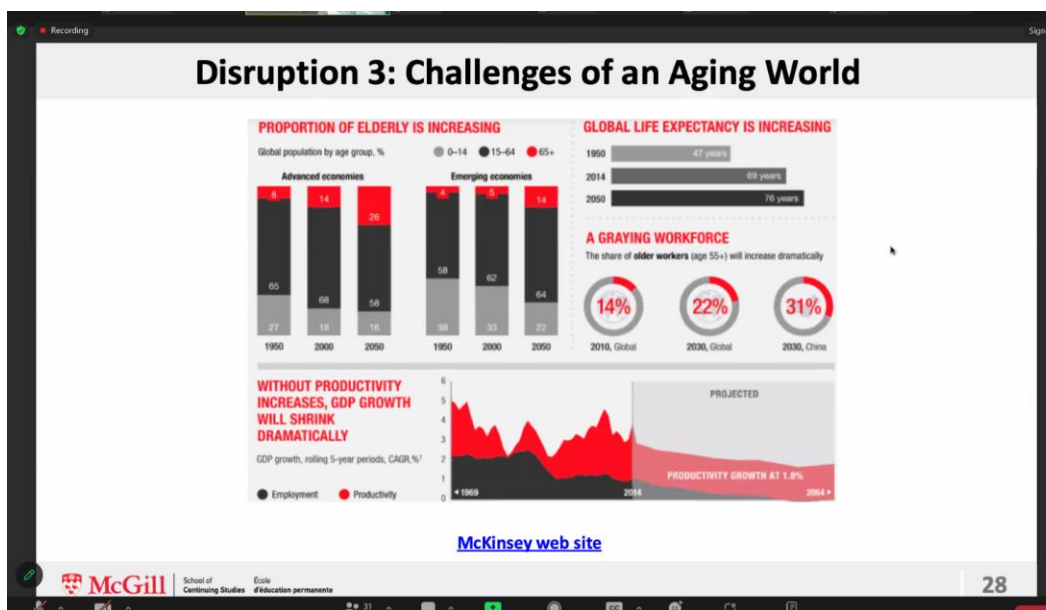
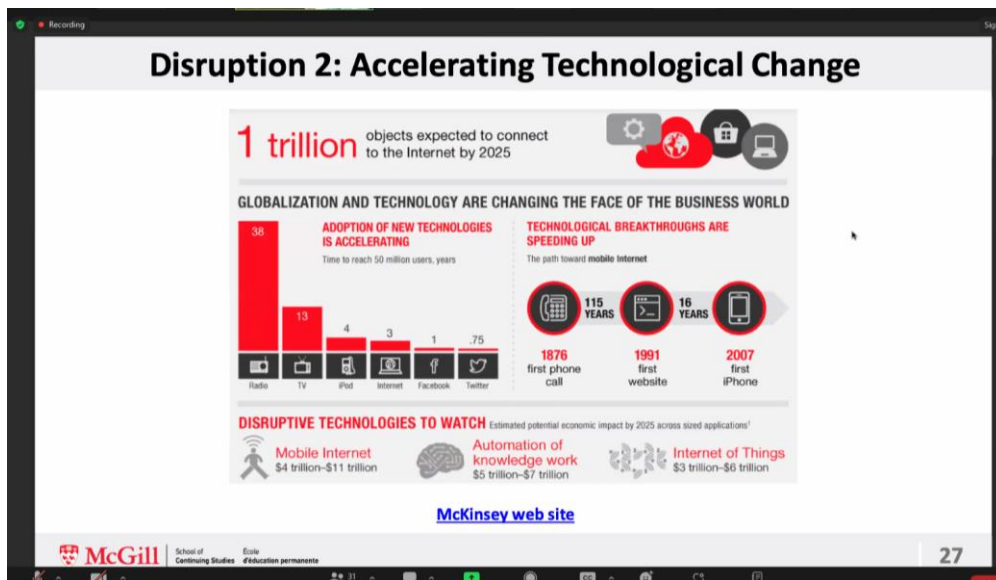
25

Disruption 1: The Age of Urbanization

McKinsey web site

McGill School of Continuing Studies Enroll Education permanently

26



Recording Page 30 of 68 Sign in

What Strategy Is: Discussion



[Michael Porter : What Strategy Is?](#)

McGill School of Continuing Studies École d'éducation permanente 30

Recording Sign in

Traditional vs. Big Bang Strategy



[Big Bang Disruption Topples Traditional Business Strategy](#)

McGill School of Continuing Studies École d'éducation permanente 31

Recording

Sign in

Data Strategy

The Data-Strategy Spectrum

A company's industry, competitive and regulatory environment, and overall strategy will inform its data strategy.

Hospitals operate in highly regulated environments where data quality and protection are paramount. They emphasize defense over offense.

Banks are heavily regulated and require strong data defense. But they operate in dynamic markets and so typically devote equal attention to data offense.

Retailers are less regulated, work with limited sensitive personal data, and must react rapidly to competition and market changes. They typically emphasize offense over defense.

FROM "WHAT'S YOUR DATA STRATEGY?," BY LEANDRO DALLEMULE AND THOMAS H. DAVENPORT, MAY-JUNE 2017 © HBR.ORG

Source: "What's your data strategy?"
by Leandro Dallemule and Thomas H. Davenport (2017)
Harvard Business Review, 95(3), 112-121

McGill

School of Continuing Studies

Esse
et education permanente

33

The Elements of Data Strategy

	DEFENSE	OFFENSE
KEY OBJECTIVES	Ensure data security, privacy, integrity, quality, regulatory compliance, and governance	Improve competitive position and profitability
CORE ACTIVITIES	Optimize data extraction, standardization, storage, and access	Optimize data analytics, modeling, visualization, transformation, and enrichment
DATA-MANAGEMENT ORIENTATION	Control	Flexibility
ENABLING ARCHITECTURE	SSOT (Single source of truth)	MVOTs (Multiple versions of the truth)

Source: "What's your data strategy?"
by Leandro Dallemule and Thomas H. Davenport (2017)
Harvard Business Review, 95(3), 112-121

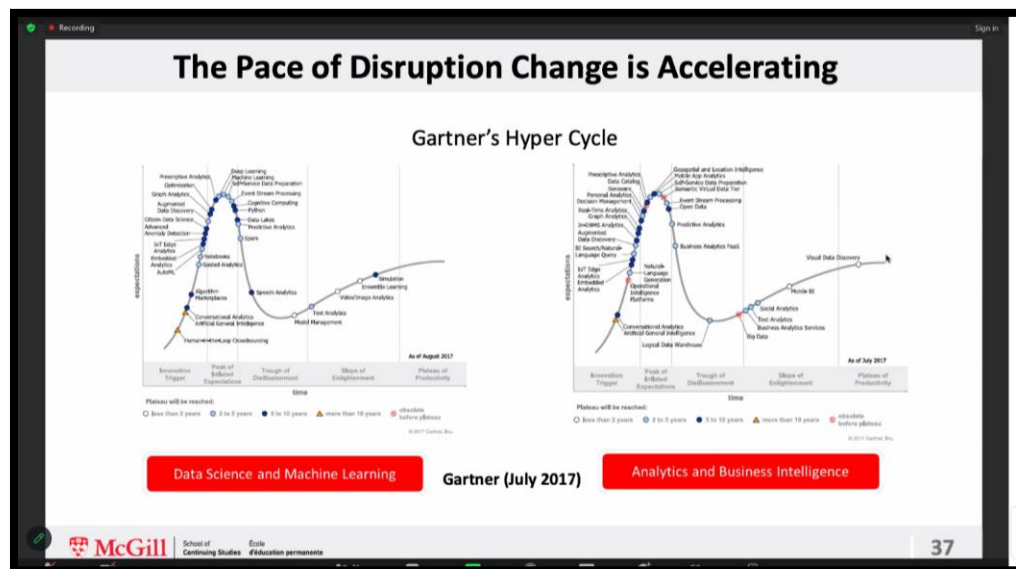
Various Models Used by Business

- **Freemium model** : Basic products/services are provided free but users are charged for advance features
- **Pay as you go or Subscription Model** : Pay only for services which are used
- **Customer experience model** : provide the customer experience never before
- **On-Demand model** : provide customer service on demand with speed
- **Marketplace model** : provide a platform for buyer and seller interact with each other directly
- **Free model** : provide the typical services to users free and sell their behavior data to different businesses
- **Crowd-sourcing model** : receive money for engaging crowd for common goal, innovation, problem solving
- **Bundling model** : selling similar products or services together
- **Gamification model** : use of game like feature to simplify the interaction

Source: [All You Need To Know About Business Models In Digital Transformation](#) – Sandeep Raut

McGill School of Continuing Studies École d'éducation permanente 35

Stack Video Endpoints Chat Share Screen Record Screen Captions Reaction Apps Notes



The Unicorn Club – Top 20 Private Tech Companies

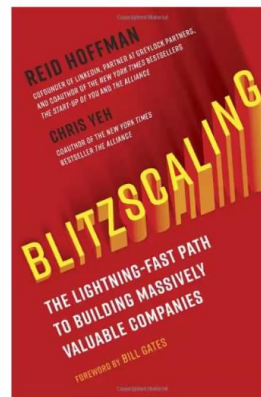
Company	Valuation (\$B)	Date Joined	Country	Industry
Toutiao (Bytedance)	\$75	04/07/2017	China	Digital Media/ AI
Didi Chuxing	\$56	12/31/2014	China	On-Demand
WeWork	\$47	02/03/2014	United States	Facilities
JUUL Labs	\$38	12/20/2017	United States	Consumer Electronics
Airbnb	\$29.30	7/26/2011	United States	eCommerce/Marketplace
Stripe	\$22.50	1/23/2014	United States	Fintech
SpaceX	\$18.50	12/01/2012	United States	Other Transportation
Epic Games	\$15	10/26/2018	United States	Gaming
GrabTaxi	\$14	12/04/2014	Singapore	On-Demand
Bitmain Technologies	\$12	07/06/2018	China	Blockchain
Samumed	\$12	08/06/2018	United States	Biotechnology
Palantir Technologies	\$11	05/05/2011	United States	Big Data
Global Switch	\$11.08	12/22/2016	United Kingdom	Computer Hardware & Services
Infor	\$10	11/16/2016	United States	Internet Software & Services
DJI Innovations	\$10	05/06/2015	China	Hardware
One97 Communications (operates Paytm)	\$10	05/12/2015	India	Fintech
Go-Jek	\$10	08/04/2016	Indonesia	On-Demand
Coupang	\$9	5/28/2014	South Korea	eCommerce/Marketplace
Guazi (Chehaoduo)	\$9	03/12/2016	China	eCommerce/Marketplace
Coinbase	\$8	08/10/2017	United States	Fintech

Note that companies such as Uber were recently floated via IPO

Source : [Research Unicorn Companies](#)

How to get there?

- Presents a set of processes and frameworks in order to scale
- Prioritize growth over efficiency
- Needed in industries with high demand-side network effects (winner takes all markets)

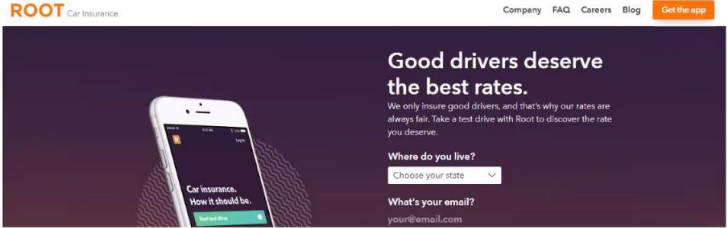


Recording

Osama Wanis Amar Kumar, Mr Frederic B. Joshua Akue Aissatou Balde PARTH Jain Sign in

Real-Time, Context-Based Insurance

- For Tesla vehicle owners: Reduces price of policy the longer the vehicle runs in autonomous mode
- Rationale: autonomous mode is safer than manual drivers (Risk Assessment and Management)
- Root accesses GPS, accelerometers and gyroscopes data on your smartphone to evaluate risk
- How much time the vehicle is in use, frequent destinations, whether drivers change lane excessively, their driving speeds, to what extent they respect traffic rules, or if they use their smartphone while driving
- 30% of applicants rejected
- Accessing vehicle data would provide additional information
- Tesla is working with industry players and is considering becoming an insurer



McGill School of Continuing Studies École d'éducation permanente


40

Recording

Sign in


Data Product for Farmers – John Deere (1/2)

John Deere uses sensors added to their equipment to help farmers manage their fleet and to decrease downtime of their tractors as well as to save on fuel. The information is combined with historical and real-time weather data, soil conditions, crop features and many other data sets.



Variable Rate Application

When you put down seed and chemical where you need it most, you'll save seed and chemical (and money).



Field and Water Management

Whether it's water or soil, your resources are precious. Make the most of them with automated water and field management. Whether it's reducing irrigation costs or being able to have one person create ditches and levees in less time, you'll save on labor and fuel costs, as well.

McGill School of Continuing Studies École d'éducation permanente

41

Leave

Before we leave...

'Would you tell me please, which way I ought to go from here,' (said Alice).

'That depends a good deal on where you want to get to,' said the Cat (Cheshire).

'I don't much care where,' said Alice.

'Then it doesn't matter which way you go,' said the Cat.

—Lewis Carroll, *Alice in Wonderland*

For Next Week

- Do the readings!
- Get your McGill e-mail
- Install Alteryx
- Install Tableau
- Login to DataCamp