



CIO Connect Data Science Master Class
From Data Strategy to Implementation
Hong Kong, September 2018

Ikhlaq Sidhu
Chief Scientist & Founding Director, Sutardja Center
IEOR Emerging Area Professor Award
UC Berkeley

Opening Comments and Introductions

Day 1: Defining AI Opportunities

Day 2: Balanced: Scoping + Technical

***CIO Connect Data Science Master Class
From Data Strategy to Implementation***

Hong Kong, September 2018



Ikhlaq Sidhu, content author

CIO Connect Data Science Master Class - From Data Strategy to Implementation

Hong Kong, September 2018

Schedule	Day 1	Day 2
9:00	Welcome and Course Overview. Set Purpose AI, Machine Learning, and Data Science, Part I <ul style="list-style-type: none"> • Brief History • What is Machine Learning, Data Science, and AI • Today's technology in Industry 	Chatbot and Google APIs (IS/ECH) <ul style="list-style-type: none"> • Using Cloud based Cognitive Services • Google Cloud Introduction • Q/A on how to build/implement
10:30	AI, Machine Learning, and Data Science Part II: Data Science Tools: Where Does Data Come From, How is Data Formatted, Using Open Source Frameworks and APIs, Applying Machine Learning	Chatbot and Google APIs (IS/ECH) (continues) AI and Job Loss - as time permits
12:00	Lunch Break	Lunch Break
1:30	Data Science Algorithms and Code: Web Scrape, Machine Learning Algorithm Comparison with Resource Links <ul style="list-style-type: none"> • Open Sources Tools of Data Science, Data Science in Python Jupyter Notebooks • Machine Learning Algorithm Overview. Examples from Numpy & array computation Pandas 	Subgroup Coaching Sessions and Master Class Cases (4-5 strategy discussions or reviews of Project Low tech Demo) Topics: <ul style="list-style-type: none"> • Map from Business to Technology Strategy • Architectural Implementation, Systems Level vs Algorithm Level • Organizational Factors
3:30	Getting Value from Data & Business Models. What is your data strategy? EQ Considerations. Data Strategy Exercise <ul style="list-style-type: none"> • What are the challenges? What is the business strategy? • How can technology help? How can customer centric activity help (EQ)? • Group Exercise: Defining your competitive strategy to avoid disruption. • Project Work Time and Project Definition • Two In Class Data Strategy Mapping Examples is time permits 	Advanced Topic Overview <ul style="list-style-type: none"> • Resource Guide, Neural Networks, TensorFlow • Trade-off considerations with Freemium APIs X at Berkeley Labs Wrap up with Reflection on Concepts Learned. <ul style="list-style-type: none"> • Project Discussion Continued • Ways to Stay Connected



**HELLO
MY NAME IS**



Ikhlaq Sidhu

Founding Director & Chief Scientist
Sutardja Center for Entrepreneurship & Technology
IEOR Emerging Area Professor
UC Berkeley

About Me: Ikhlaq Sidhu

- All degrees: Electrical Engineering
- HW and SW Design: HP Laserjet Printer
- Networking & Communications: 75 patents including Skype Technology, Inventor of Year at 3Com Corp. IEEE Major Innovation Award 2018
- Executive Roles: USR, 3Com, Cambia (new venture): 3 products lines
- Professor & Founder/Director, Sutardja Center, and many spin outs ..., almost – 14 years



Elias Castro Hernandez,
Researcher, Data-X Lab at UC Berkeley's
Sutardja Center for Entrepreneurship &
Technology

Elias Castro Hernandez

- Elias Castro Hernandez, researcher in the Data-X Lab at UC Berkeley's Sutardja Center
- Data-driven product manager and lead developer of the Berkeley Innovation Index
- His major focus is applied mathematics with an emphasis in Operations Research and Management Science and as Data Science
- His interests include business processes automation, and all things FinTech

SUTARDJA CENTER AT BERKELEY

METRICS AT A GLANCE

Undergraduate:

- 12-14 Courses, 1500+ Undergraduates

X at Berkeley: Graduate, Labs and Professional

- 80+ Grad students
- 100+ Executives
- Labs: Data-X, Blockchain, Sustainable Food

Ecosystem:

- 14+ Global Partners
- 500+ Executives
- 50+ Investors



IEOR 135/290 Applied Data Science with Venture Applications

Supported by the Data-X Lab

33036) and

Instructor = Ikhlaq
Offered = Spring
Units = 3 (Lecture and

This course provides an understanding at the intersection of mathematical concepts and current computer applications of real world problems. The course will be presented in application organization of Collect, Combine, and Visualize. Applications of this course include industry sectors such as finance, health, engineering, transportation, energy, and many others.

COMPUTER SCIENCE TOOLS =

- PY pandas
- PY NumPy
- PY SciPy
- SQL
- Apache Spark

MATHEMATICAL CONCEPTS =

- Linear Algebra
- Probability
- Statistics
- Machine Learning
- Data Structures
- Transforms

Please complete this form if you're interested in receiving additional information about this course.

Ikhlaq H. Sabir, Author

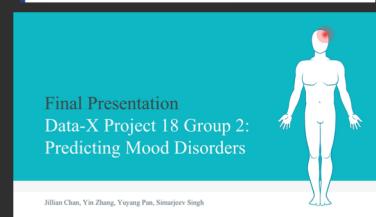
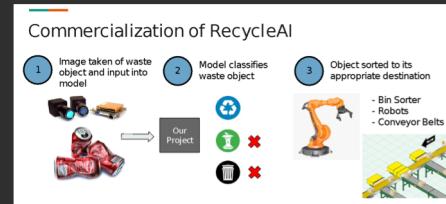
[http://data-x.org/ieor135](#)

IEOR 135/290 Applied Data Science with Venture Applications

Sample Data-X Projects

- Detection of fake news
- Prediction of long-term energy prices to solve Wall Street problem
- Prediction applications for stock market, sports betting, and more
- AI for crime detection, traffic guidance, medical diagnostics, etc.
- A version of Zillow that is recalculated with the effects of AirBnB income

and many many more...



Data-X: IEOR 135/290 Applied Data Science with Venture Applications by the Numbers

- Developed over two years
- 340+ alumni students
- 100+ industry experts in network
- 80+ great projects completed
- 6+ published research papers
- 10+ students got hired because of Data-X
- Amazing testimonials:

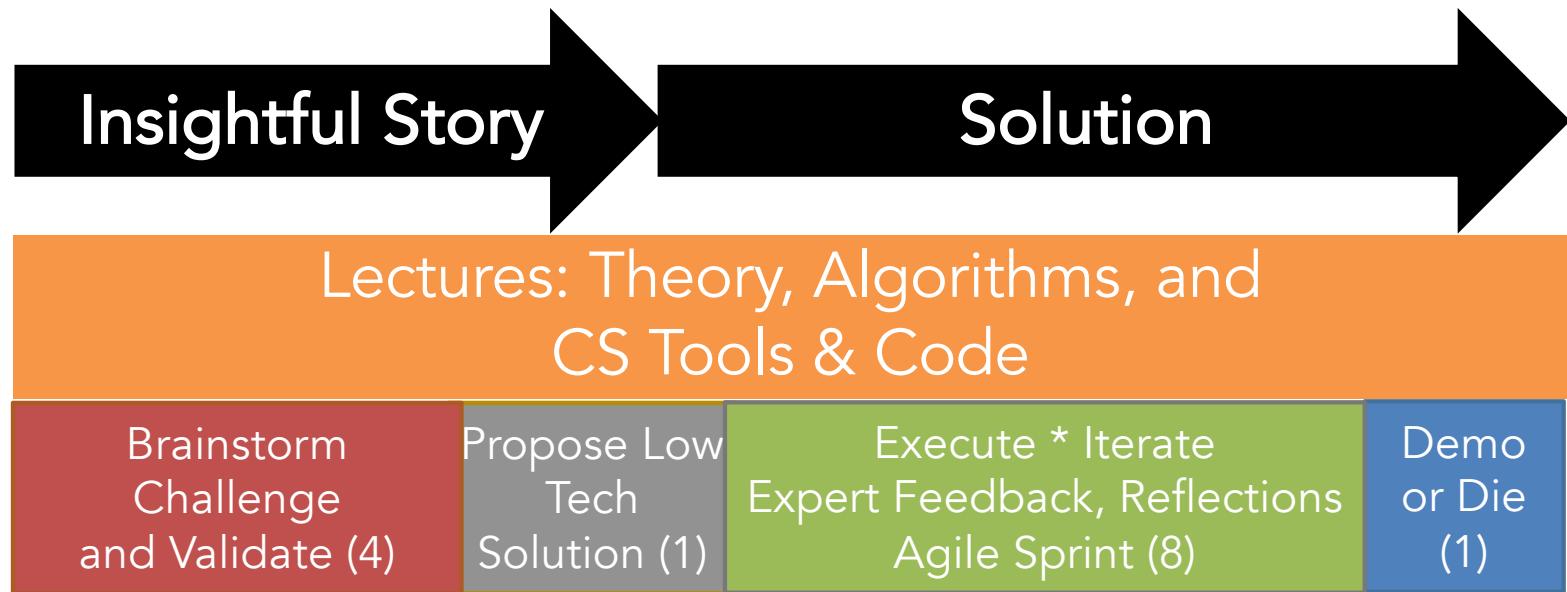
I think this class is so awesome because it teaches the tools and concepts that are most commonly used in workplace teams that are involved with data science and applied machine learning.



What is actually in this class?

- The ML stack most commonly used in creating ML/AI/Data applications
- Application and **systems** viewpoint of data and ML
- Implementation, architecture, and relevant processes to build real systems
- Statistical, rule based, and **hybrid decision systems**
- Connection with relevant **mathematical foundations** (optimization, entropy, correlation, LTI, prediction, classification)
- Practical insight into **advanced techniques** and tools: (eg. CNNs, NLP, scraping, recurrent networks, etc.)
- System **modeling** for data applications
- Application talks: Recommender systems, Blockchain, Spark etc.

How the Data-X Course Works:



Open-ended, real-world project: Typically 5 students, with available advisor network

Our Agenda and Purpose

1. Understand Data+AI and Business Application
 1. What, Why, and How for AI
 2. Learn the ML tools/stack
 3. Code samples / basic theory
 4. Starting point to go further
2. Project:
 1. Scope what you would like to learn or do next within your organization
3. Additional Goals:
 1. Teams – meet each other
 2. Diversity in perspective
 3. Pay it forward
 4. Global resources
 5. Staying connected

Why:
Relevance to
Organization

What:
Technology

How:
Implementation
with Innovation
Behaviors

An Overview of Data and AI Applications

Ikhlaq Sidhu, content author

Basic Concept of Working with Data

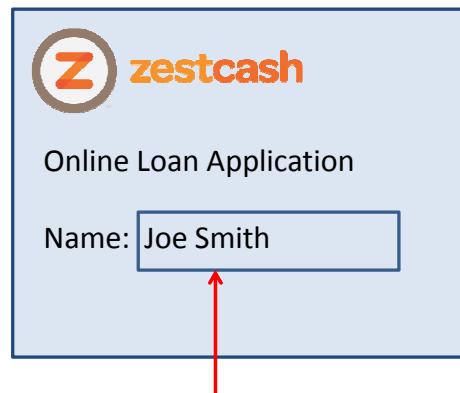


- Data Wrangling
- In Production

Example: Data and Information is a competitive advantage

Real-life Example: ZestCash

- “All data is credit data”



The data says: greater credit risk!

The data says: lesser credit risk!

Reference: Shomit Ghose

- Service provider of Gambling and Casinos
- Entry Card
- Pain points
- Intervention



Harrah's Casino: Knowing your customer

PLAY & WIN ▶

Reference: Supercrunchers

Top 8 Business Models Using Data

1. Knowing your customer, **better targeting and relationship.**
E.g. Target, Disney, Netflix
2. Improving **physical** product or service with **complimentary information:**
E.g. UPS, FedEx
3. Data-driven **reliability** or security
E.g. GE, BMW, Siemens
4. Information **Brokers, Arbitrage, and Trading Opportunities:**
E.g. Investment funds.
5. Improving the **customer journey/experience..**
E.g. Harrah's

6. Functional Applications: **HR/Hiring**, Operations etc..
Eg Walmart, Baseball, Sports
7. Efficiency or better performance **per dollar cost.**
E.G. General IT, SAP, etc
8. Risk Management, regulation, and **compliance**
Eg. Compliance 360



Why: More Simply

Customer
Insight/
Engagement

Operations:
Reliable &
Predictable

Security &
Fraud

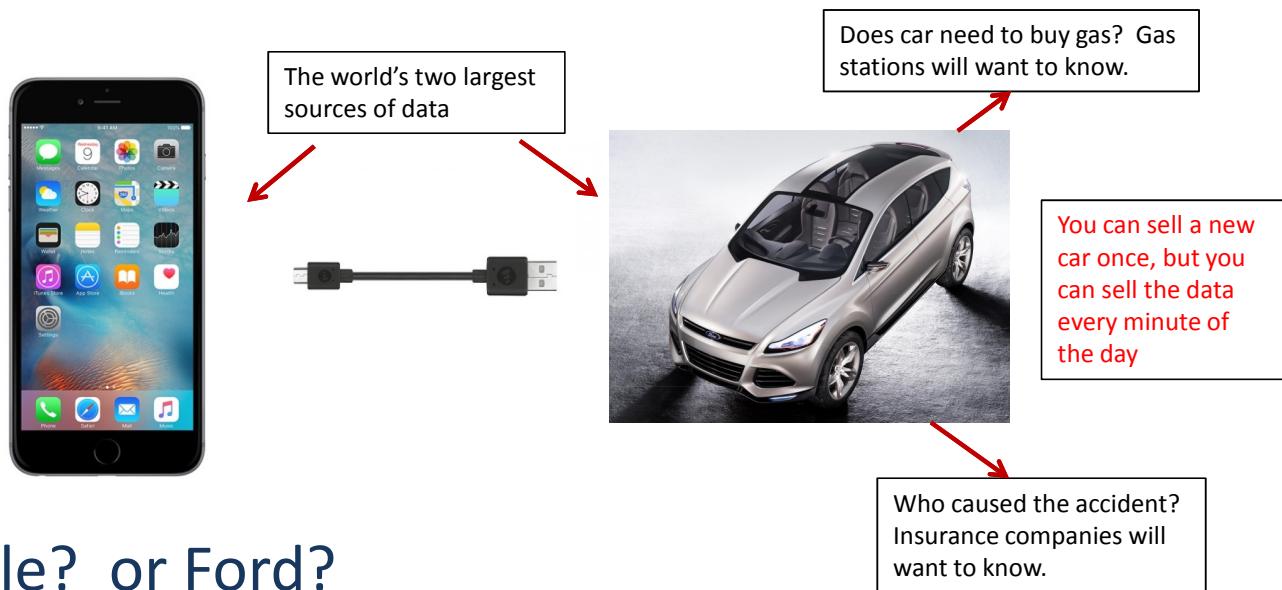


Compliance 360°

Financial Firms

Network Security

Who Will Control the Automobile?



- Google? or Ford?
 - Whoever has the better software and data science team
 - Winner will get the vast (and incredibly valuable) streams of auto data

The Core of Every Business in Future is Data and AI

The two key components of a business are resources (assets) and information (data)



=



resources

+



Information and
automated decisions

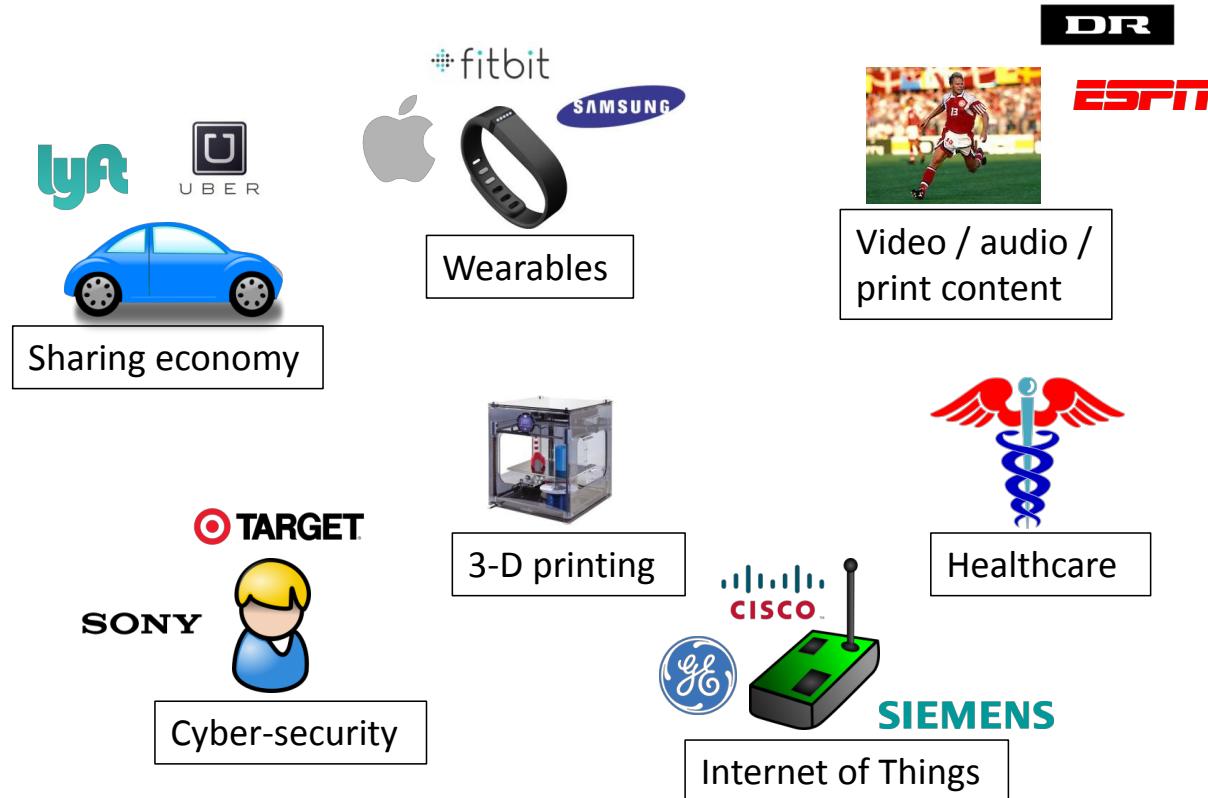
Less value
over time

More value
Over time

If you buy data, then everyone else has it also.

Ikhlaq Sidhu, content author

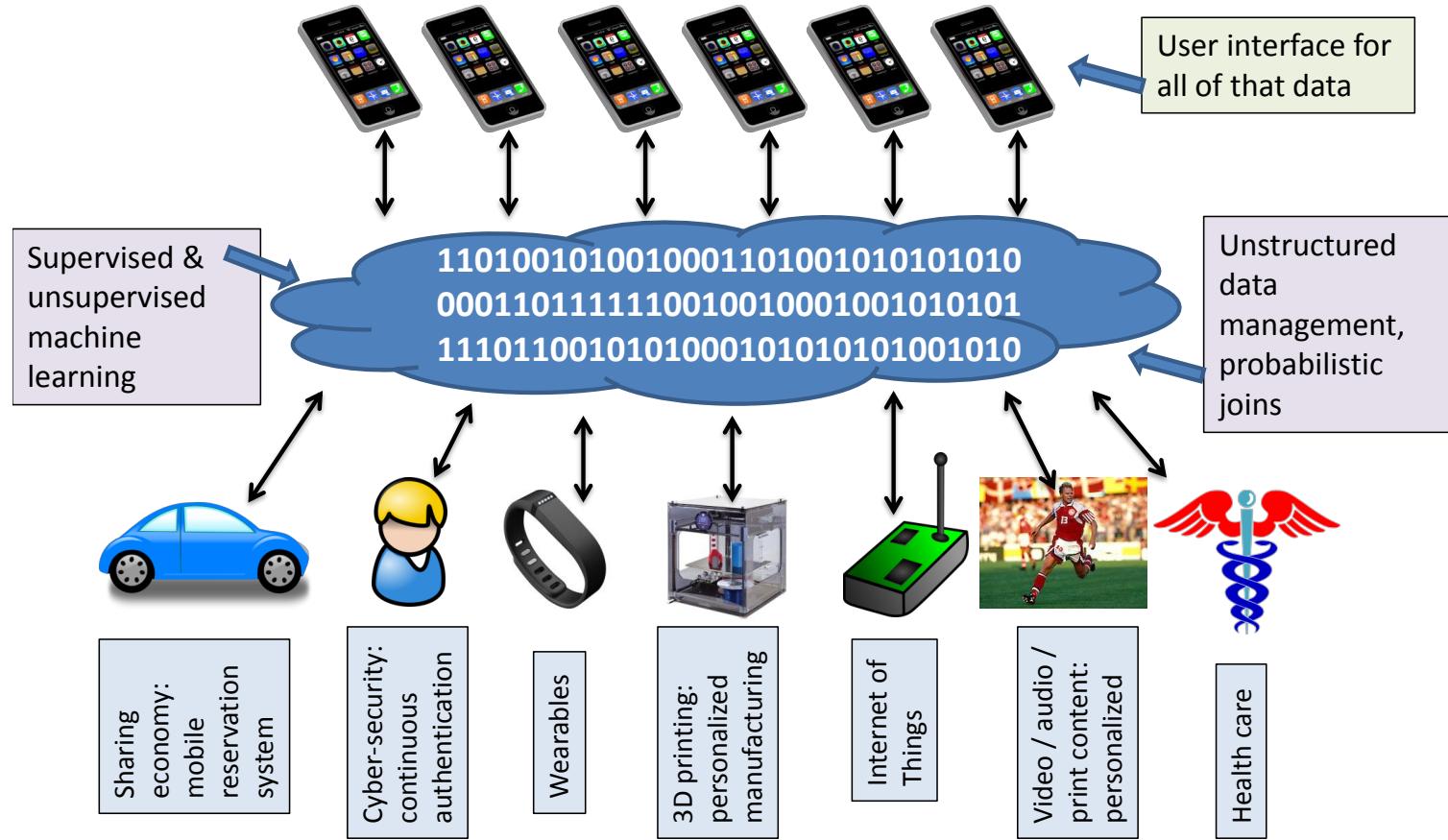
What One Thing Do the Following Have in Common?



Ref: Shomit Ghose

Ikhlaq Sidhu, content author

Everything is Data-Driven



Ref: Shomit Ghose

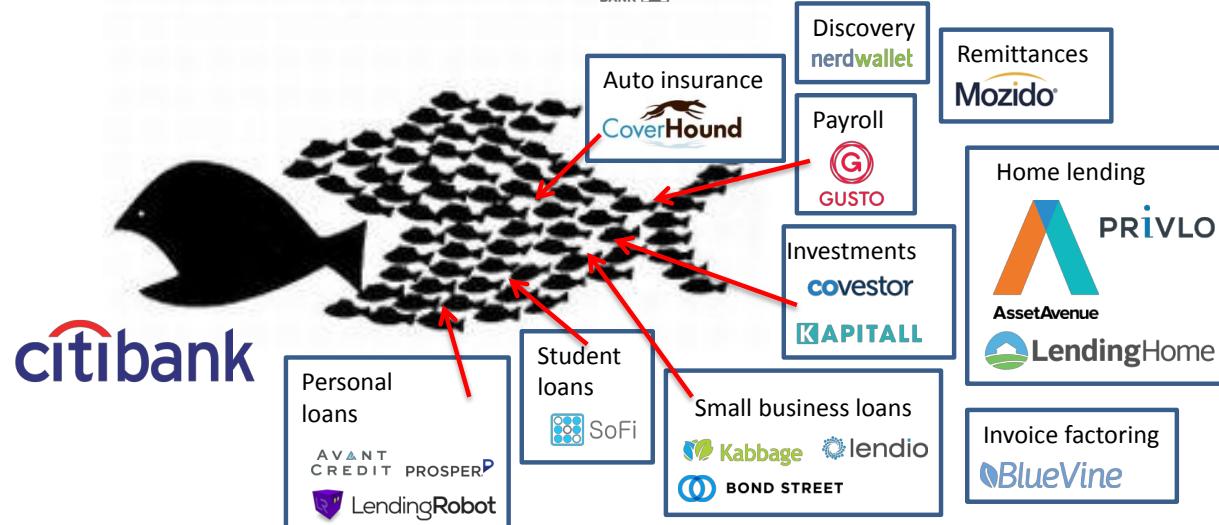
Ikhaiq Sidhu, content author

The Banking Industry: Disruption Through Data

Then:



Now:



Ref: Shomit Ghose

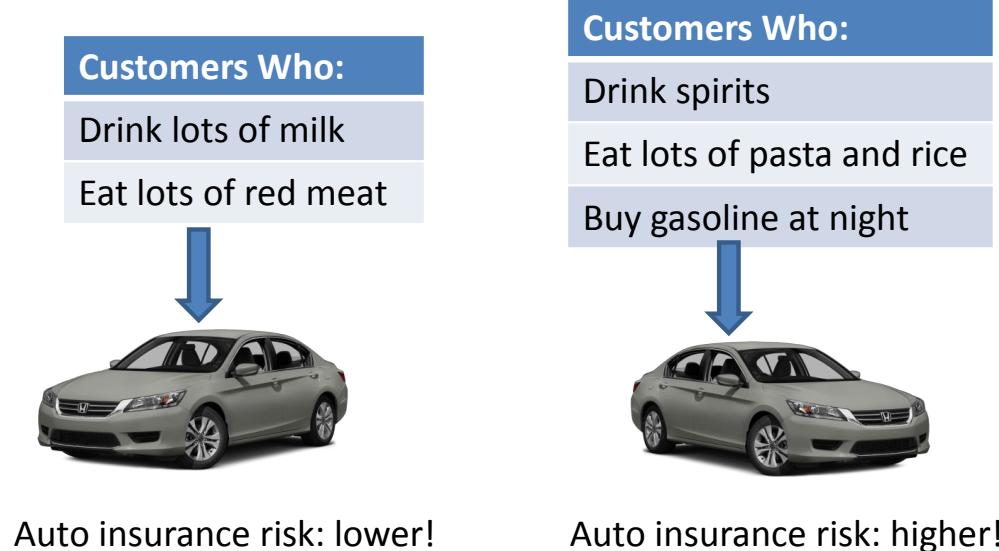
Ikhsaq Sidhu, content author

Data is deconstructing your legacy business

Ref: Shomit Ghose

Ikhlaq Sidhu, content author

Another Real Life Example



Reference: Shomit Ghose

Ikhlaq Sidhu, content author

But I'm in an Old, Commodity Business!

World's biggest brewer
Owner of 400 beer brands worldwide



App.



What kind of beer are people looking for?
Liking?
Where and when?
Anticipate customer needs: getting the right beer in front of the right person at the right time



Reference: Shomit Ghose

Beer business: dates to 5000 B.C.; definitely a commodity product

...Shomit Ghose, content author

Who Else Does This??



Reference: Shomit Ghose

Ikrinaq Sidhu, content author

The Data-driven Business Model

	Google	facebook.	amazon	Square
Create a data stream	  		 	
Make money by understanding data stream				

Reference: Shomit Ghose

Ikhlaq Sidhu, content author

“Devices” or “Data”?



A cleaning appliance or a data appliance??

The Washington Post

The Switch

Dust isn't the only thing your Roomba is sucking up. It's also gathering maps of your house.

By Hamza Shaban July 25

PRIVACY AND SECURITY

Roomba CEO Swears That He Will Never Sell Maps of Users' Homes, So Help Him God

Rhett Jones
7/28/11 5:45pm • Filed to: PRIVACY

62.5K 49 4



GIZMODO

Shomit Ghose, content author

Reference: Shomit Ghose

Data-Driven Animal Care



Hey, Old
McDonald! I'm
sick!!

The cow
can't talk,
but the
data can!

- Which cow is sick?
- Which cow is not growing?
- Do you medicate them all???

Can You Sell This as Data?

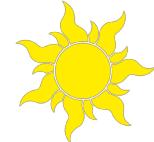
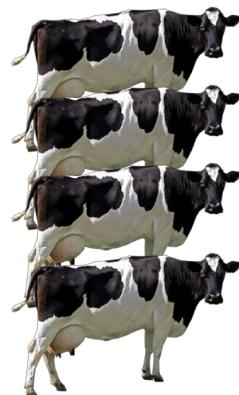


•

Reference: Shomit Ghose

Ikniaq Sidhu, content author

Selling the Data in Milk



All of the cows grew up on the same field eating the same grass

These cows did not incur an additional expense of pharmaceuticals: the data trail tells us!



Premium, anti-biotic-free milk that was **cheaper to produce than** the regular milk
Thank you, data!



These cows incurred an **additional expense of pharmaceuticals**



Reference: Shomit Ghose

Ikhlaq Sidhu, content author

The Source of Value: What Wins?

Mechanics	Company	Semantics
Social network		Data: individuals / businesses
Search Maps Email Self-driving cars		Data: individuals / businesses
Professional networking		Data: individuals / businesses
Work-out tracking		Data: individuals
Shopping lists		Data: individuals
Microblogging		???
???	Your start-up	Data: individuals / businesses

Reference: Shomit Ghose

Ikhlaq Sidhu, content author

Long Term, Who Wins in Mobile?

Device manufacturers...



htc



... or data companies?

Google

facebook



amazon

Long Term, Who Wins in Autos?

“Device” manufacturers...



... or data companies?

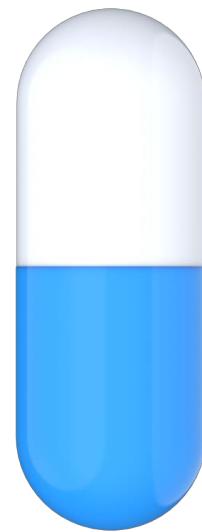


Long Term, Who Wins in Pharma?

“Device” manufacturers...



Johnson & Johnson



... or data companies?

Google

facebook

amazon

 **Microsoft**

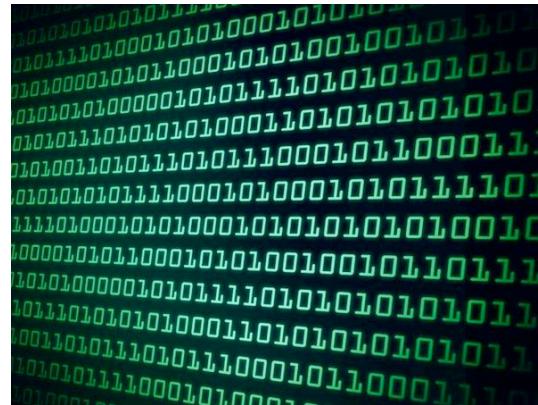
Anticipating the Next Industrial Revolution

Industrial Revolution 1.0



- Winner was whoever made something most cheaply
- Leveraged scale

Industrial Revolution 2.0



- Winner will be whoever makes best sense of the data
- Leveraging scale

Reference: Shomit Ghose

Ikhlaq Sidhu, content author

Contact:

Ikhlaq Sidhu

Founding Faculty Director, Center for Entrepreneurship & Technology

IEOR Emerging Area Professor, UC Berkeley

[sidhu @berkeley.edu](mailto:sidhu@berkeley.edu), scet.berkeley.edu