Analytics and Engineering, and What It Means to You

AIAA Southern California Aerospace Systems and Technology (SoCal ASAT) Conference

Saturday, November 9, 2019

Doubletree Club Hotel, Orange County Airport
7 Hutton Centre Drive, Santa Ana, CA 92707

AIAA OC *Section Website:* *https://engage.aiaa.org/orangecounty/home*

Joseph Eugene (Gene) Justin, MS, PhD
AIAA Associate Fellow
University of Redlands, Adjunct
9 November 2019

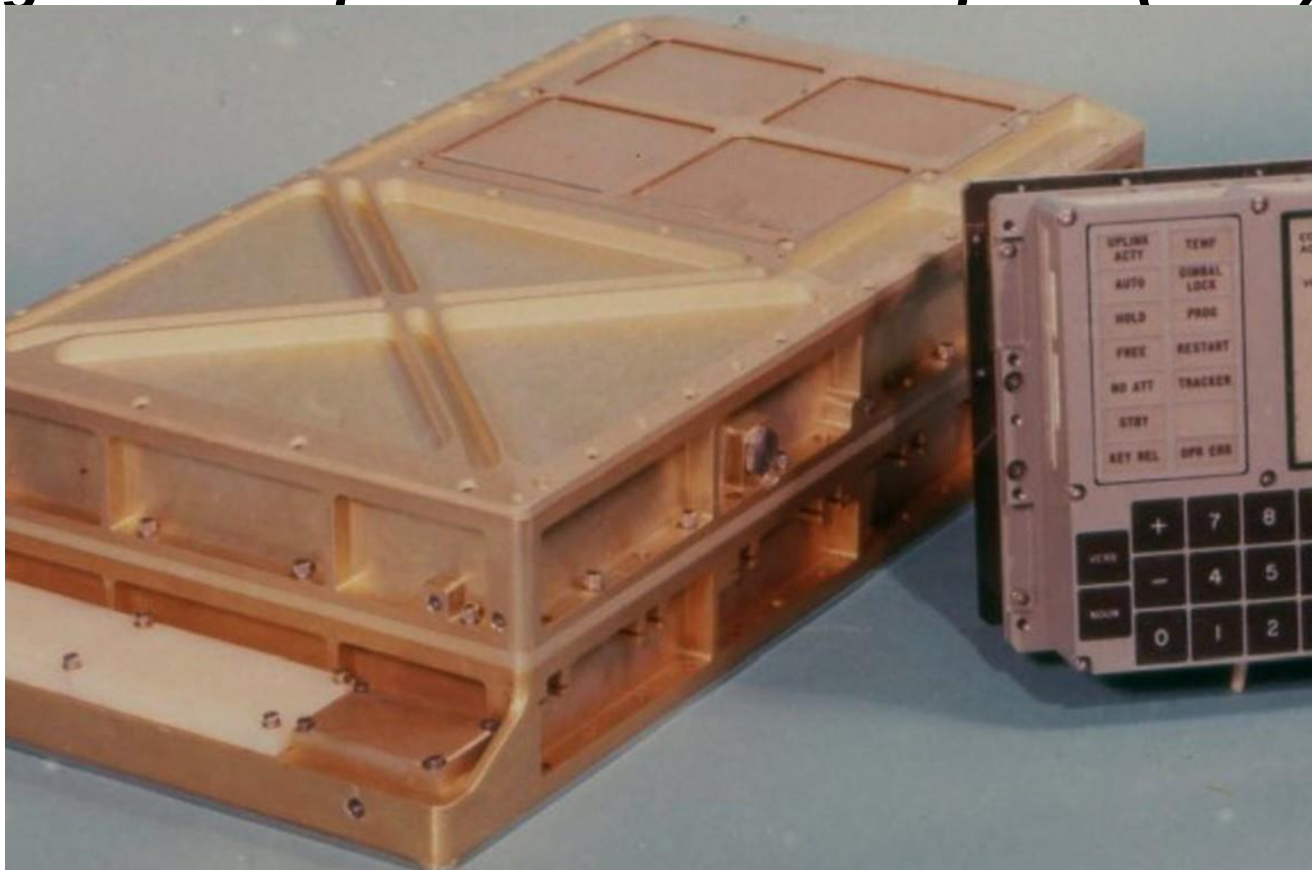
- ***BS Engineering USAFA, MS Astro-Aero Ohio State, MA USC, EMBA UCLA, PhD Mgmt Drucker***
- ***AF Assignments:***
- ***Wright-Patterson AFB, F.E. Warren AFB, Norton AFB, USAF Academy, Rand Corp, Pentagon***
- ***AIAA Life Associate Fellow, IEEE Member***

Summary

- **An Academic survey here on the Rise of Modern Analytics including Big Data, Data Mining, and Data Based Decisions.**
  - The raise of the Modern Big Data and Number Crunchers.
  - The impact on Engineering.
  - See Next Charts
- **Factors discussion: The impact of the Internet, Massive Databases, the massive spread of computers worldwide, and the faster and faster speed of PCs devices (Moore's Law*)**
  - The wide public acceptance of the conveniences of Cell Phones, Online shopping, Social Media versus concerns about privacy and data misuse.
  - Engineering is not isolated from Analytics and Number Crunching.
  - In fact, Analytics is changing how we are thinking about many things, from Baseball to Deep Learning.
  - It is also challenging us about the role of Judgment, Intuition, and Experience versus the Modern Big Data World.
- **Will discuss the future, room for both Experts and Crunchers?**

***Gordon E. Moore,1965, Co-founder of Intel.** The number of transistors that can be packed into a given unit of space (a chip) will double about every two years**.** In general, speed/power of a computer will double every 2 years. A new generation every few years.

Looking Back -- Apollo Guidance Computer (AGC) -- 1969



- **See NewAlas.com, RealClearScience.com (1 November 2019)**
- **Raytheon, 24 × 12.5 × 6.5 ins, and 70 lbs.**
- **Hardwired for particular tasks. Not a general purpose computer.**
- **72KB of Read Only Memory (ROM), 32,768 bits of RAM memory. equivalent to 589,824 bits. 12-microsecond clock speed.**
- **Much, much less powerful than iPhone 11 (64 GBs)**
- **Gigabyte = one billion (10^9) or, strictly, 2^{30} bytes.**

Apple II e

- Apple II e ~ June 1977, US $1,298 with 4K RAM, US $2,638 with 48K RAM
- Floppy: 143K
- OS: Basic
- Anything missing from this phone?

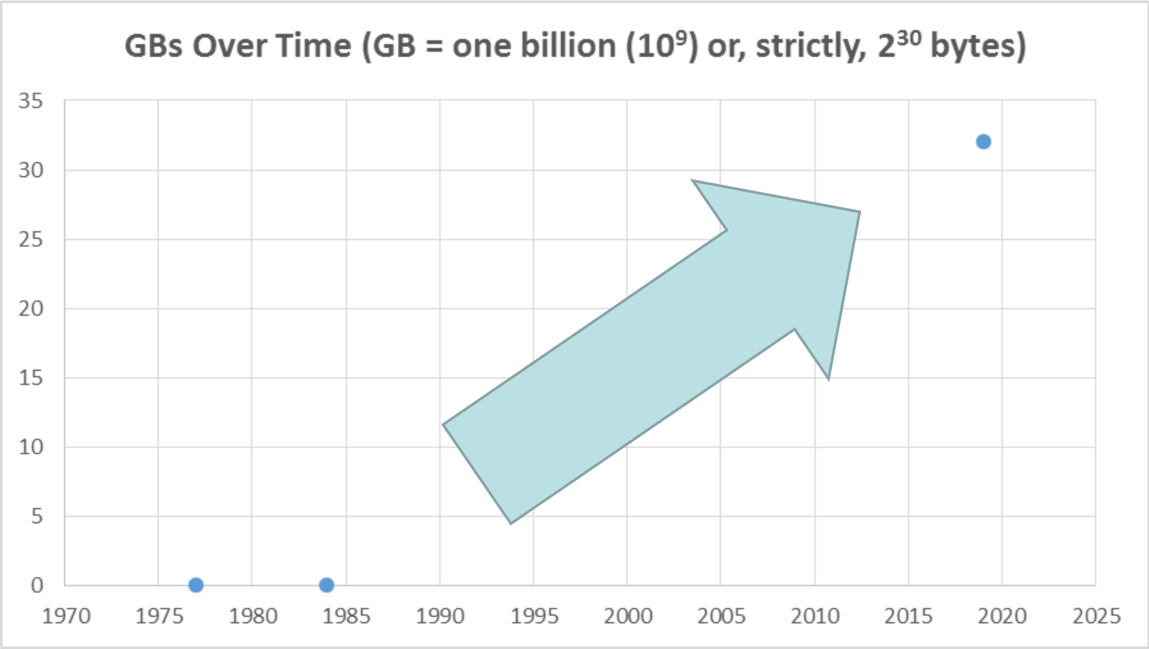
Apple First Mac (Macintosh 128K)

- **Apple First Mac (Macintosh 128K), 24 January 1984**
- **Nine-inch screen, diskettes, and $2,500 price tag**
- **“A PC Computer for Everyone. That anyone can afford…”**
- **Apples’ Game Changer for the Early Mass Produced “Modern” PCs Market, Mouse/Graphs (rivals Commodore, Atari, IBM, etc.)**

MacBook Pro (MBP)

- **MacBook Pro, info as of 2019, $1,499.00: Fast Intel Core i5 and i7 Processors, 13 Inch and 15 inch models**
- **Upgrades for 16 GB, 32 GB**

Memory (GBs) Over Time



"Moore's Law": In general, speed/power of a computer have increased over the years, as demand for faster more powerful personal computers increase. A new generation every few years.

*Look Back Milestones**

- **Apollo 11 Landing on the Moon – 24 July 1969**
- **Woodstock – 15-18 August 1969**
- **Microprocessors ~ 1969** (single integrated circuit chip solid-state electronics, computer basic central processing unit building block)

- **Walmart – 2 July 1962 (Mkt Cap now >$300B, >$36M/hr in Sales)**
- **First PC -- ~1971**
- **Barcode Scanners – Summer 1974**
- **Microsoft Started – 4 April 1975 (Mkt Cap now >$1T)**
- **Apple – 1 April 1976 (Mkt Cap >$1T). 1st Apple 2e – April 1977**
- **Dell Computers – 1 February 1984 (Mkt Cap >$30B)**
- **Worldwide Web (www) Available to Public – 6 August 1991**
- **Smart Phones – 1992**

- **Google, Various Web Searches, 2019 (Ballpark numbers).**
- **Mkt Cap = Number of shares outstanding x Current market price of one share.**

Amazon Fulfillment Center

- **Amazon started 5 July 1994. Today, one of their Fulfillment Center**

Stater Bros Distribution Center

- **Stater Bros Distribution Center, San Bernardino, CA**

*Look Back Milestones, Continued**

- **Amazon -- 5 July 1994 (Mkt Cap >$800B, >300 transactions/sec)**
- **UPS around since 28 August 1907.**
- **Google – 4 September 1998 (Revenue >$136B/yr, Alphabet Mkt Cap >$800B)**
- **9/11 – 11 September 2001**
- **Walmart 1St RFIDs (Radio Freq Id Tags) – June 2003**
- **Facebook – February 2004 (Mkt Cap >$500B, > 4,000 photos/sec)**
- **“Money Ball” Baseball Movie – 19 September 2011**
- **5G Phones – 2019-2020???**
- **Amazon Drone Deliveries ~ ??? (For Real 2019-2020?)**

*** Various Web Searches, 2019 (Ballpark Numbers)**

*Terms**

- **Analytics – Analysis looking for trends in data and statistics**
- **Main types of Analytics (Statistics)**
  - Descriptive – Describes historic and real time data distribution, trends, and patterns of interest – e.g. A Batter's hits toward First or Third
  - Inferential/Predictive **–** Drawing "observations" / "conclusions" from data. Subjected to tests, random variation (e.g., observational errors, sampling variation).
    - **Test Analytics – Difference tests**
    - **Predictive Forecasts – "Likelihood" of a "future outcome". Accuracy of predictions is not 100%, as it is based on probabilities – e.g. Wine yield and eventual rating based on weather, soil, picking conditions, timing**
      - Not causality, But Association
      - Theory Based versus Data Based
      - Beware spurious variables and linkages, human bias and motivation
    - **Regression, ANOVA (Analysis of Variances), Forecasts, Monte Carlo Sim**
- **When did Analytics start –** Some date it start back to WWII in 1940s. Actually, Statistics is older than that. Modern Statistics in early 1900s. First College Major in 2014? at Denison, Ohio. First online tools date back to 1990s**.**

*Various Web/Video Searches

Terms, Continued*

- **Random Tests, Very Large Samples, Real Time Feedback**
  - Random, independent inputs, equal chance of occuring
  - Public Data, and Internet – allows for Randomization Options
  - Very Large Sample Sizes – allows for Significant Tests
  - Allows for simple models, listening to what the data says
  - Less need to expert models, preconceived models
  - e.g. – Change web site design – some see one site, others other site

*Various Web/Video Searches

Analytics Example 1

- **Wine Model -- Orley Ashenfelter's circa 1990s (Ayres, Page 2):**
- **Wine Quality =**

  12.14500

  + 0.00117 Winter Rainfall

  + 0.06140 Average Growing Season Temperature

  – 0.00386 Harvest Rainfall
- **Countered expert opinions**
- **Model was reliable in forecasting quality**
- **The model is based upon data that showed that low levels of harvest rain and higher average summer temperatures is linked to greater wine qualities**

Example 2 – A Website

- **Option A**

Create Job Inventory	Post a Job	Search Resumes
• Words • Learn More	• Words • Learn More	• Words • Learn More

- **Option B**

Buy Job Postings	Post a Job	Search and Buy Resumes
• Words • Learn More	• Words • Learn More	• Words • Learn More

- **Example based upon (Ayres, Page 57) Monster.com / Offermatica.com results**
- **Tested with Random Assignments to Different Website Combinations**
- **Measured Site View Time for Combinations, sales, etc.**
- **PS: Make sure to check with Firms, Ethics, etc. as needed**

*Example 3 – Two Sample Test**

- **Example – Data on Two Competing Procedures**
- **Standard Stat Tools – Simple Software Tools available**
- **Lowers the need for "Expert" Help**
- **Large Samples, Randomization makes Controls and Tests easier**
- **Example here:**
  - Small Sample t Test, But Steps Applies to large Samples.
  - Two Sample Test; T Score -0.6.
  - Result: For this example, the Different was not significant, and Alpha level of .10

	Xw = The W Method (Minutes)*	Xa = The A Method (Minutes)
	2	3
	4	7
	9	5
	3	8
	2	4
		3
Sum	20	30
n	5	6
=SumX/n	4	5

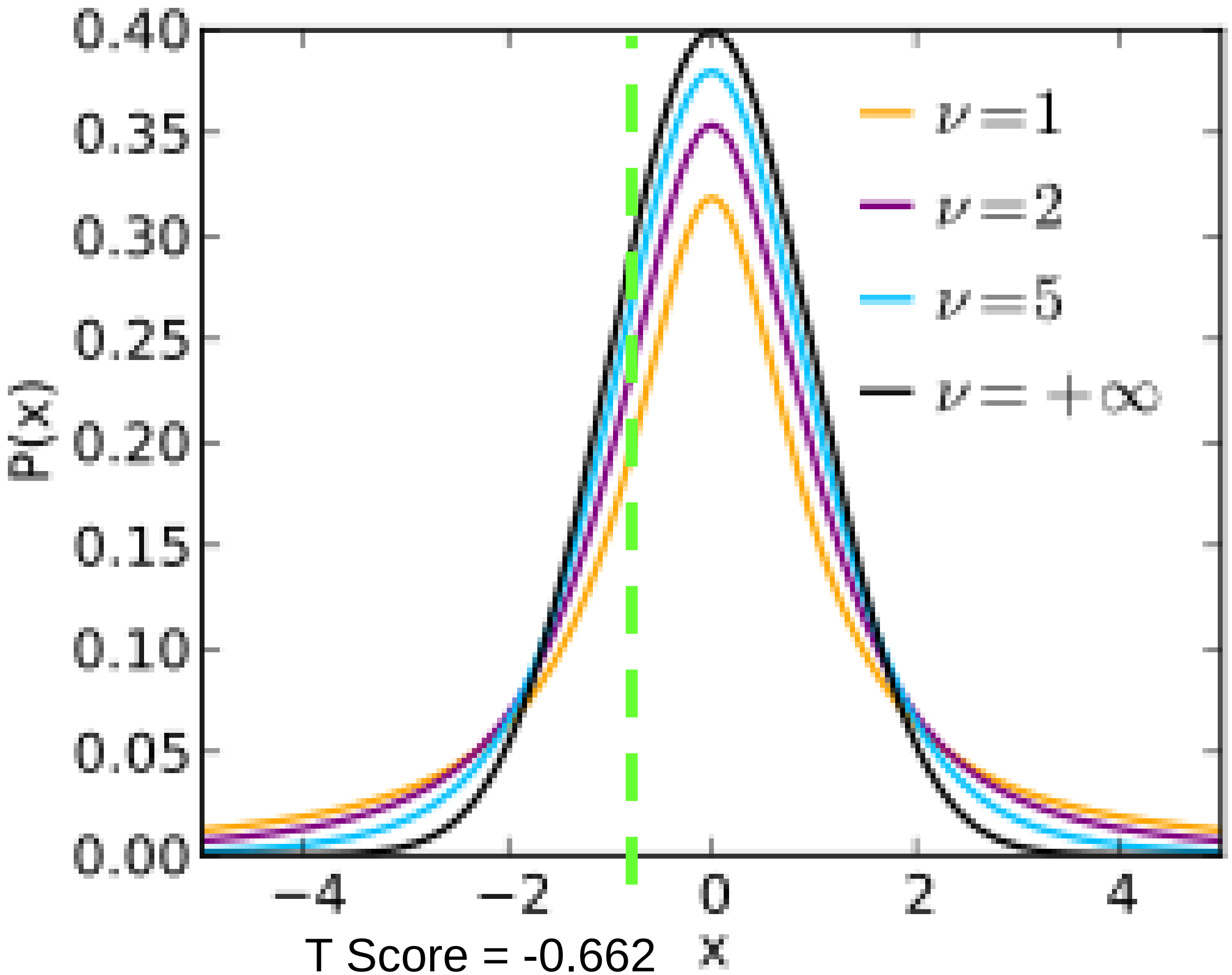
Two Different Assembly Methods

*Lind, Marchal, Wathen (2018), Statistical Techniques in Business & Economics, Pages 361-364.

Example 3 – Two Sample Test, Continued

	Xw = The W Method (Minutes)*	Xa = The A Method (Minutes)		(Xw-XwMean)	( )**2	(Xa-XaMean)	( )**2	
	2	3		-2	4	-2	4	
	4	7		0	0	2	4	
	9	5		5	25	0	0	
	3	8		-1	1	3	9	
	2	4		-2	4	-1	1	
		3				-2	4	
Sum	20	30		Sum	34	Sum	22	
n	5	6						
=SumX/n	4	5						
*Lind, Marchal, Wathen (2018), Statistical Techniques in Business & Economics, *Pages 361-364.								
Assume Two Sample Test, with Unknown Population Std Dev, i.e. Mean Xw = Mean Xa								
Samples Independent								
Two Populations Normal, Equal Std Dev: Can Pool Std Dev								
Sw = SQRT([(Xw-XwMean)**2]/nw-1) = SQRT(34/(5-1)) = 2.915						8.5	2.915476	
Sa = SQRT([(Xa-XaMean)**2]/na-1) = SQRT(22/(6-1)) = 2.0976						4.4	2.097618	
Sample Pooled Variance								
Sp**2 = [(nw-1) Sw**2 + (na-1) Sa**2)]/(nw + na -2) =						6.222222222		
t Test								
t = (Mean Xw - Mean Xa/SQRT(Sp**2*(1/nw + 1/na) =						=(4-5)/1.51045	-0.662	
t Score = Not Significant					dof = nw + na -2 = 9			
p Value > .20								
Difference in Means is Not Significant								

Example 3 – Two Sample Test, Continued



T Score = -0.662

Deg of Freedom = v = nw + np - 2 = 9

P valve = > 20% (Not Significant Difference in Xw and Xa Test

Example 3 – Excel Data Analysis Add On: Output

Xw	Xa	t-Test: Two-Sample Assuming Equal Variances		
2	3			
4	7		*Variable 1*	*Variable 2*
9	5	Mean	4	5
3	8	Variance	8.5	4.4
2	4	Observations	5	6
	3	Pooled Variance	6.222222	
		Hypothesized Mean Difference	0	
		df	9	
		t Stat	-0.66205	
		P(T<=t) one-tail	0.262263	
		t Critical one-tail	1.833113	
		P(T<=t) two-tail	0.524526	
		t Critical two-tail	2.262157	

T Score = -0.66205

Deg of Freedom = df = nw + np - 2 = 9

P valve (Two-Tail) 0.524526 (Not Significant Difference in Xw and Xa Test

Terms, Continued*

- **Big Data --** Extremely large data sets. Viewed and tested to find patterns, trends, and associations. Often company internal, or on web. Often people related, patterns, trends, behavior, transactions, and interactions. e.g. Social Media, E-Commerce Data.
  - e.g. The most common RAM size laptop PCs now ~ 8GB (Gigabytes)
  - Gigabyte = one billion (10^9) or, strictly, 2^{30} bytes.
  - Higher-end ~ 12GB and 16GB.
  - Gaming laptops ~ 24GB or 32GB
  - Library of Congress collection is > 10 terabytes of memory.
  - One terabyte is ~1,000 gigabytes.
  - World's Printed collection – one copy ~ > a petabyte — 2^{50} bytes
  - Google, Amazon, Microsoft and Facebook store > 1,200 petabytes (i.e.,1.2 million terabytes) between them

*Google, ScienceFocus, Various Web/Video Searches

Terms*, Continued

- **Data Mining –** Looking at large databases. Looking for trends, patterns, correlated data.
- **Data/Evidence Based Decisions** – Decisions grounded at looking at data before making a decision.
- **The Gold Standard in Data Based Decisions --** Random, Large Samples.

*Various Web/Video Searches

Terms, Continued*

- **Modern Number Crunchers --** Person, computer or software. That runs fast calculations with large data sets.
- **Equations -- Assertion that two expressions are equivalent.**
  - Left side of the equal sign equals/is equivalent to the right hand side.
- **Algorithms –** Step by step procedures to solve a problem, given these inputs here is the desired output to increase the chance of the desire result.
  - e.g., A Recipe, step by step procedures
  - e.g. Wine yield/grade given the weather, grapes, picking time, etc.
  - e.g. Baseball infielders shift for an opposing batter to provide coverage for where the batter probably will hit most often.
  - Algorithm versus equation:

*Various Web/Video Searches

Analytics Flow Diagram

- **Data Driven based upon Scientific Method, Statistic Methods.**
- **May be more to it than this Diagram but Analytics challenges us to simplify and to look at what the data is telling us.**
- **Collect, and Analyze the Data**
- **Challenges preconceived models**
- **Challenges established Expert Models**
- **Almost Real Time Feedback, and Simple Analytics Tools**

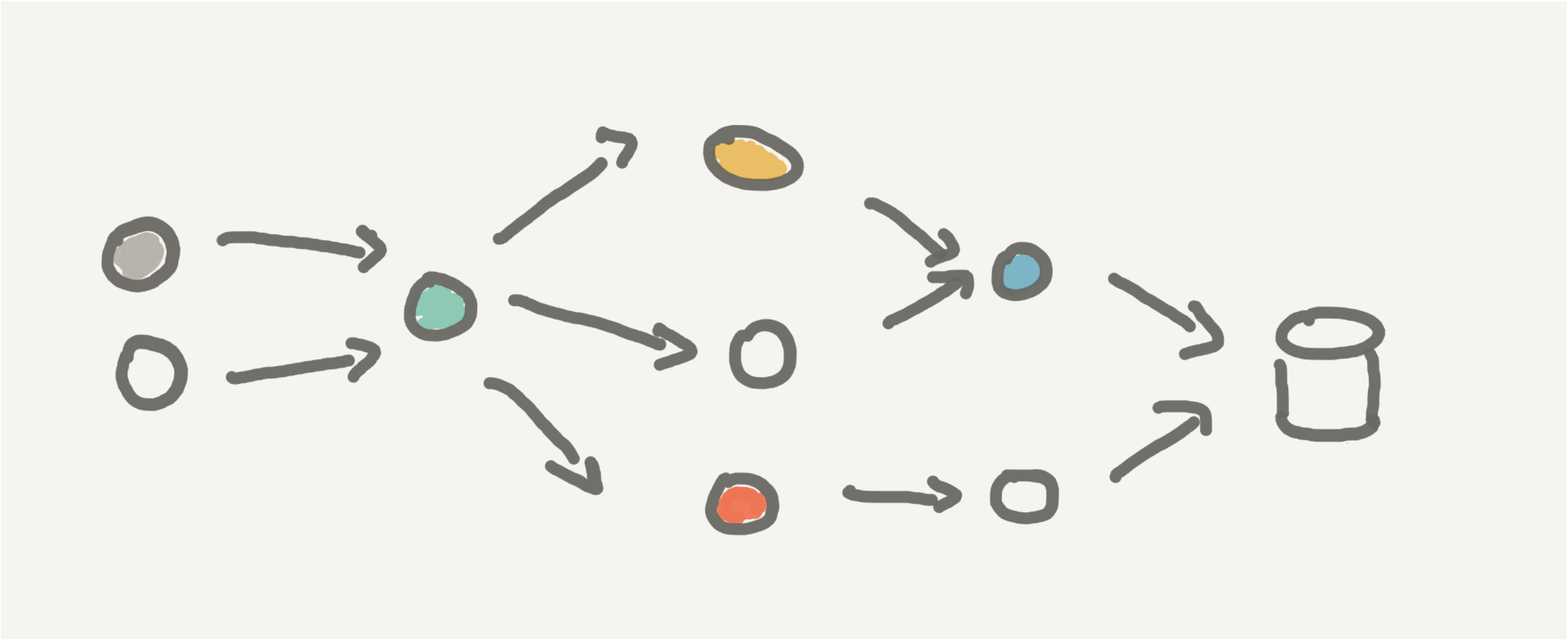
Data-Driven Decision Making*

***Northeastern University**

https://www.northeastern.edu/graduate/blog/data-driven-decision-making/

Engineering Data

Is this Your Engineering Data?



*Button Blog, 2019

Engineering

- **Engineering Data –** State of the art data, Innovation data, or new Knowledge data:
  - Data that eventually becomes the body of engineering knowledge
  - Data that leads to an engineering design and a final product for the customer and the community.
  - Engineering data is captured in Technical Performance Measurement (TPMs), Requirements, Drawings, Specifications, and Standards.
  - Also, Performance Data, Schedules, Agile, Scrums, Storyboards
- **Engineering typically wants:**
  - Repeatable means and small variance (3 Sigma, 6 Sigma) in data, hardware, standards
  - Tight quality standards (3 Sigma, 6 Sigma) and controls
  - Reliable engineering models, including software, and stochastic inputs for Kalman Filters, and Model Based Engineering models.

Engineering and Analytics

- **Analytics** -- Challenges us to look at patterns and trends in Data.
- **Challenges** -- Preconceived Expert Models, Complex Sampling
- **Analytics** -- Challenges us with very low cost, fast feedback, large sample, simple randomization, simple tests, and simple models
- **So far, Analytics uses** -- E-Commerce and Public Large Data sets.
- **Analytics has changed some industries** -- Retail, E-Commerce, Service, Information.
- **Not clear what is next, or what the next Hi-Tech shift will impact.**
- **Engineering** -- Already is changing and challenging preconceived models (websites, customer transactions, and customer expectations, needs, and wants).
- **What is next** -- Still to be determined. But, stay tune.
- **And, yes need both Experts and Data Crunchers** -- Best if in same people

Summary

- **Analytics** -- Uses E-Commerce and Public Large Data sets.
  - Has changed industries: Retail, E-Commerce, Service, Information.
  - Challenges us to look at patterns and trends in Data.
  - Challenges Preconceived Expert Models, and Complex Samples.
  - Uses very low cost, fast feedback, large sample, simple randomization, simple tests, and simple models
- **Engineering** -- Is changing and challenging preconceived models
  - Websites, customer transactions, and customer expectations, needs, and wants.
- **Need** -- Both Experts and Data Crunchers, best if in same people
- **What is next** -- Still to be determined. Create your own Future.
- **So, it will be interesting times**

- **Back Up**

Market Cap Data

Company*	Mkt Cap* $ USD
Amazon	$878.90B
Apple	$1.09T
Dell	$35.97B
Facebook	$534.90B
Google/Alphabet Class A	$874.27B
Microsoft	$1.10T
Wal-Mart	$333.52B
*Market Cap, Google.com: October 30, 2019, ~0800 PST Hours	
Mkt Cap* = The number of shares outstanding x the current market price of one share. (*Google.Com)	

References

- Ayres, I. (2008). Super Crunchers: Why thinking-by-numbers is the new way to be smart. New York, NY: Bantam Books. ISBN-10: 0553384732; ISBN-13: 978-0553384734
- Google, ScienceFocus, and Other Various Web Searches, Last Searches, 30 October 2019
- Lind, Marchal, Wathen (2018), Statistical Techniques in Business & Economics, Pages 361-364.
- Monster.com Scores Millions, http://www.offermatics.com/stories-1.7.html
- NewAlas.com and Realclearscience.com, 1 November 2019.

  https://www.realclearscience.com/articles/2019/07/02/your_mobile_phone_vs_apollo_11s_guidance_computer_111026.html
- Northeastern University, 1 November 2019: (https://www.northeastern.edu/graduate/blog/data-driven-decision-making/)

- **Simple tools are available**
  - Excel (with “data analysis pack” add-in)
  - Minitab