Inspire…Educate…Transform.

Certificate Program in Engineering Excellence

Orientation – Batch 28

25th March, 2017

Scholarship

Day 1

Entrance Test results and awards

Day 45

Project Presentation, Viva and awards

Day 22

Scholarship Test – Viva and awards

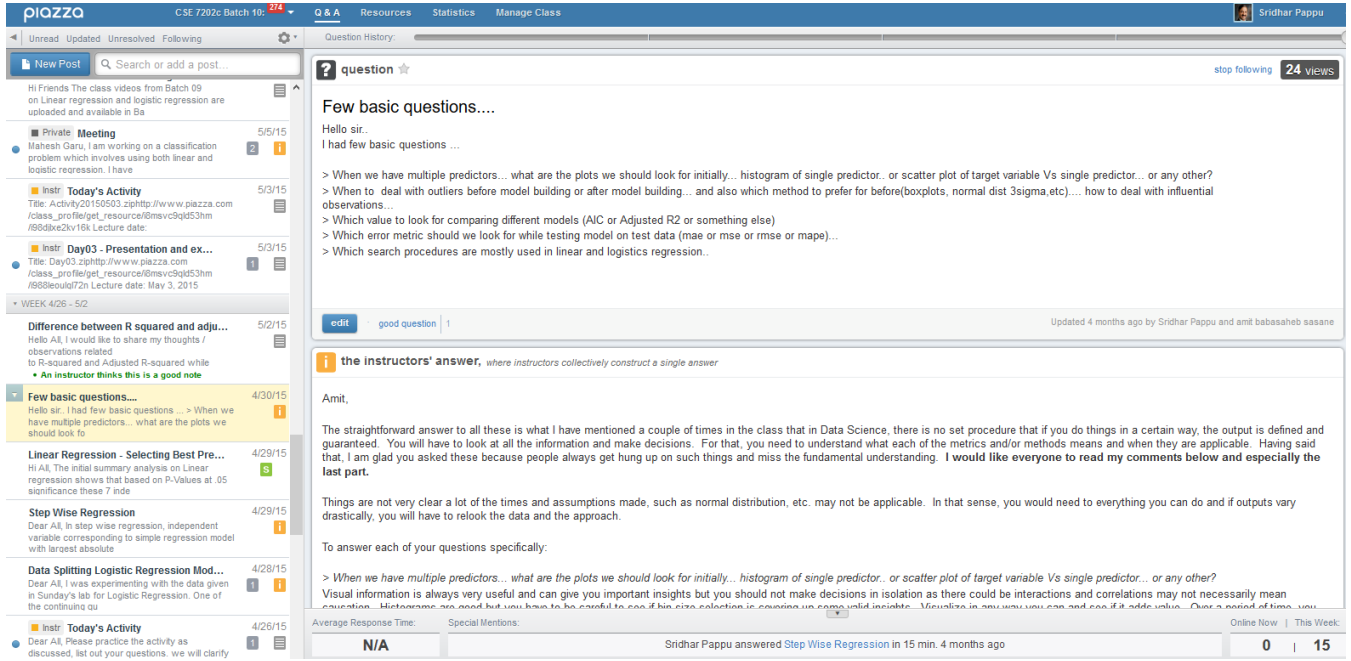
- 10% of your batch revenue is allocated as the total scholarship amount.
- Number of scholarships and the amount disbursed will depend on the performance at each phase. The academic team in-charge will conduct the evaluations and it is up to their discretion to decide the amount and number of scholarships.
- A student can claim either partial or full fee amount in this process.
- The maximum a student can gain is INR 3,00,000
- Internship announcements as required by INSOFE

Class Structure

- 4 hours lecture…focus on explaining concepts.
- 4 hours lab…focus on hands-on.
- Hands-on activities using R & Python. Do not underestimate its importance. Will have focus in exams.
- Welcome to come to office during weekdays and clarify questions. But please make an appointment as Data Scientists are also involved in CPR activities.

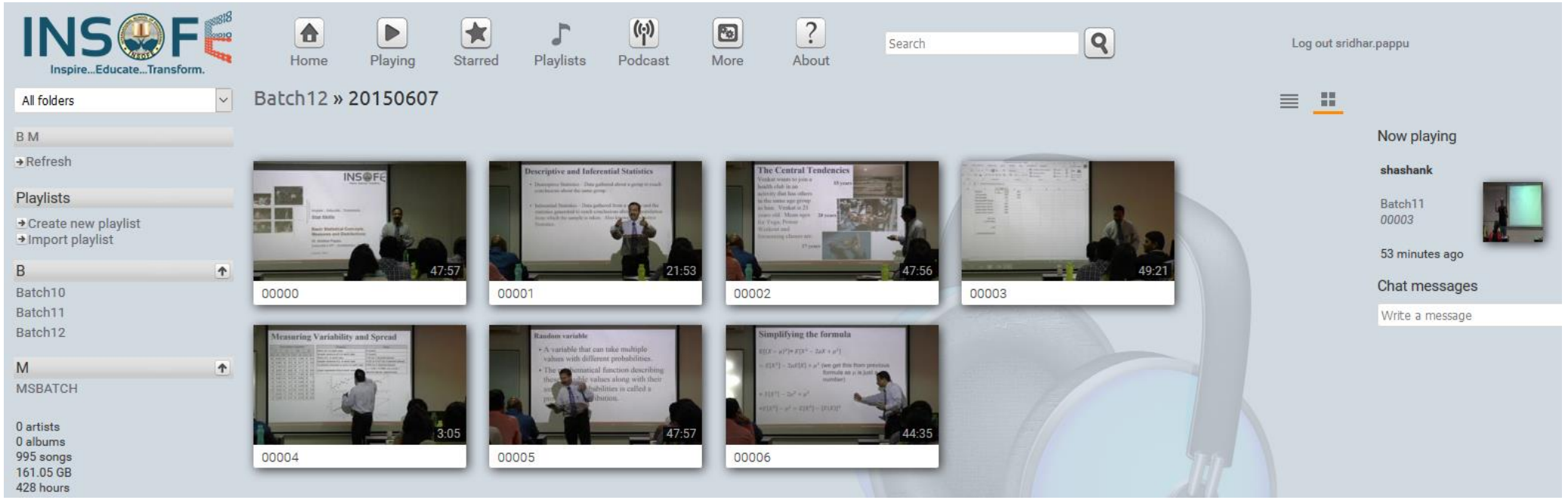
Course Material and Discussion

Piazza



- Ask a lot of questions
- Post questions, assignments, etc. in the correct module; else, they will be lost.

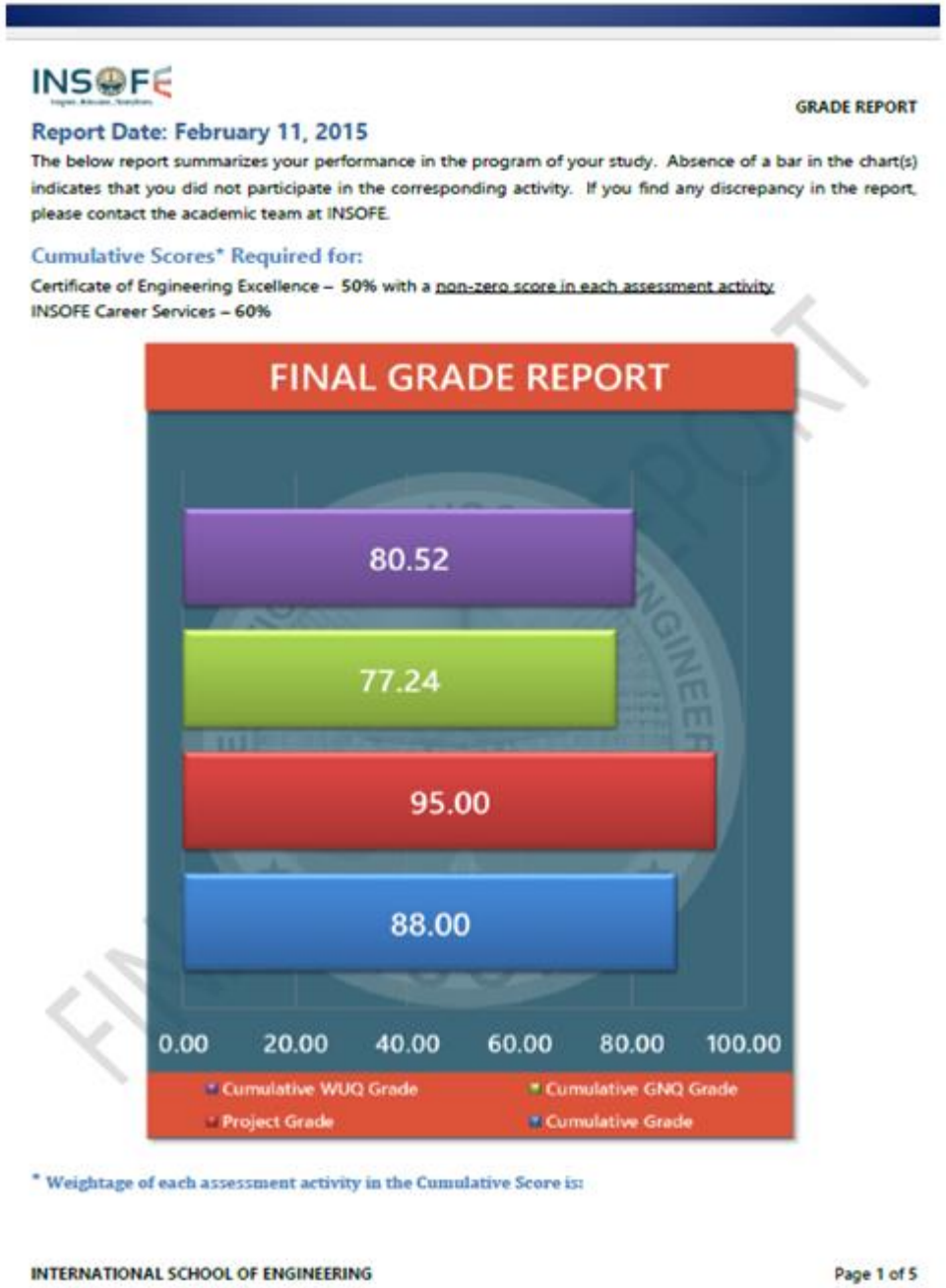
Video Recordings



- Do not miss classes. Video ≠ Classroom
- Will be available for 3 months after the program.
- Contact if you need extension.

Assessment

- ~30 WUQs, 6 GNQs, Project
- Weightage
  - Project: 40%
  - GNQ: 30%
  - WUQ: 10%
  - Feedback: 5%
  - Attendance (85%+): 5%
  - Lab exercises: 10%
- Non-zero cumulative score in each a must



INSOFE

GRADE REPORT

Report Date: February 11, 2015

The below report summarizes your performance in the program of your study. Absence of a bar in the chart(s) indicates that you did not participate in the corresponding activity. If you find any discrepancy in the report, please contact the academic team at INSOFE.

Cumulative Scores* Required for:

Certificate of Engineering Excellence – 50% with a non-zero score in each assessment activity

INSOFE Career Services – 60%

* Weightage of each assessment activity in the Cumulative Score is:

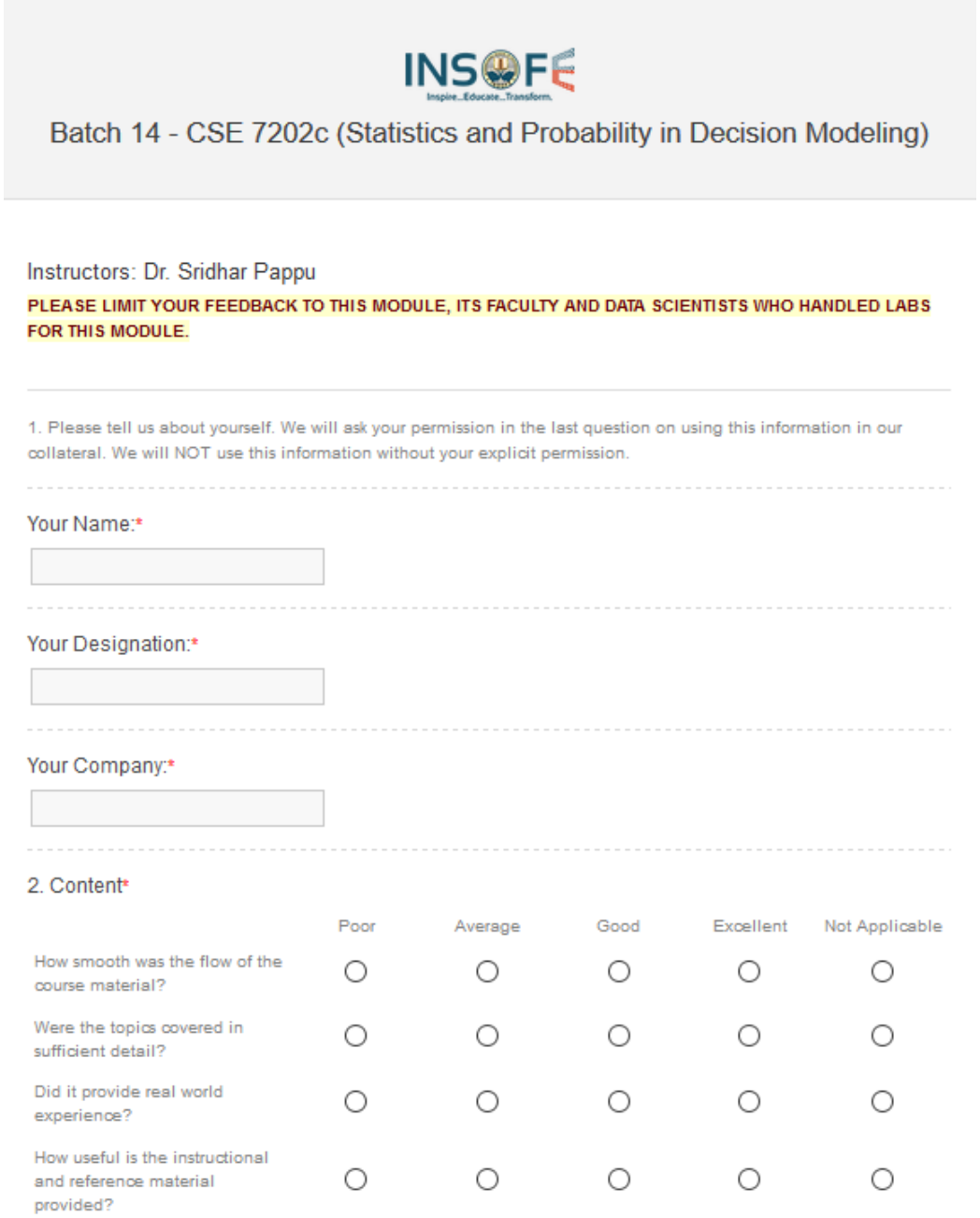
INTERNATIONAL SCHOOL OF ENGINEERING

Page 1 of 5

Feedback

Zoho Survey

- 5% grade as incentive.
- One per module.
- Provide as soon as you receive the link.



INSOFE
Inspire...Educate...Transform.

Batch 14 - CSE 7202c (Statistics and Probability in Decision Modeling)

Instructors: Dr. Sridhar Pappu

PLEASE LIMIT YOUR FEEDBACK TO THIS MODULE, ITS FACULTY AND DATA SCIENTISTS WHO HANDLED LABS FOR THIS MODULE.

1. Please tell us about yourself. We will ask your permission in the last question on using this information in our collateral. We will NOT use this information without your explicit permission.

Your Name:*

Your Designation:*

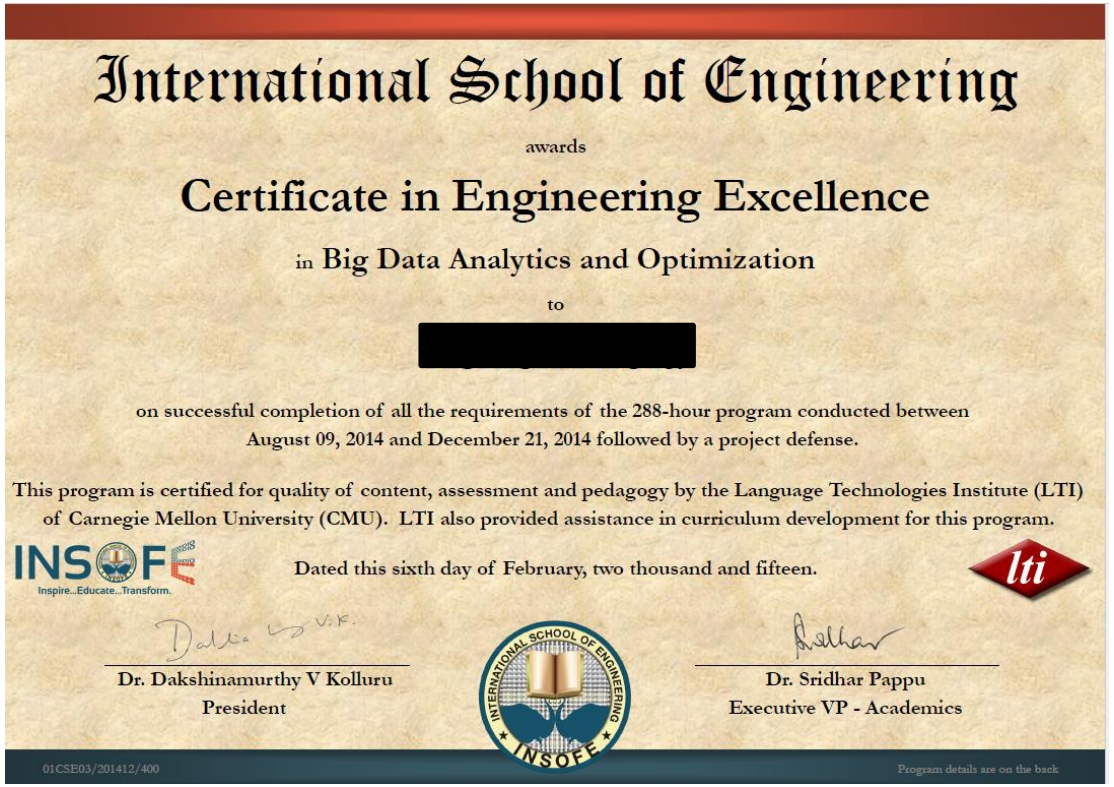
Your Company:*

2. Content*

	Poor	Average	Good	Excellent	Not Applicable
How smooth was the flow of the course material?	○	○	○	○	○
Were the topics covered in sufficient detail?	○	○	○	○	○
Did it provide real world experience?	○	○	○	○	○
How useful is the instructional and reference material provided?	○	○	○	○	○

Certification (or not)

- No certificate if attendance is below **70%** and cumulative grade **< 50%.**
- You have to work really hard to be (in)eligible for this.



International School of Engineering

awards

Certificate in Engineering Excellence

in Big Data Analytics and Optimization

to

[redacted]

on successful completion of all the requirements of the 288-hour program conducted between August 09, 2014 and December 21, 2014 followed by a project defense.

This program is certified for quality of content, assessment and pedagogy by the Language Technologies Institute (LTI) of Carnegie Mellon University (CMU). LTI also provided assistance in curriculum development for this program.

INSOFE Inspire...Educate...Transform.

Dated this sixth day of February, two thousand and fifteen.

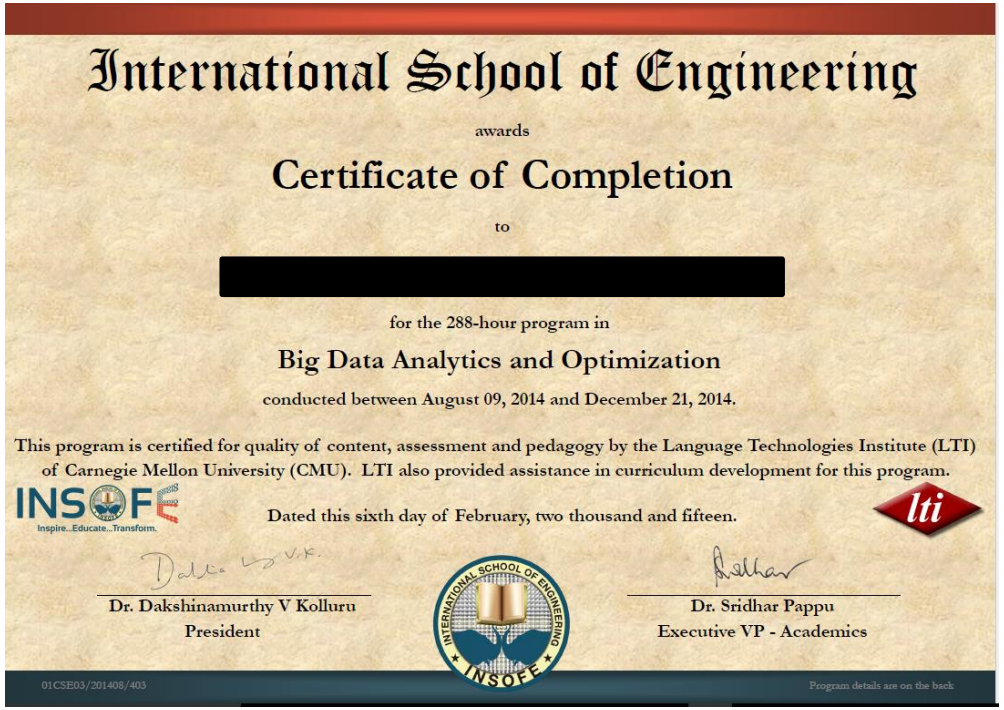
lti

Dr. Dakshinamurthy V Kolluru
President

Dr. Sridhar Pappu
Executive VP - Academics

01CSE03/201412/400

Program details are on the back



International School of Engineering

awards

Certificate of Completion

to

[redacted]

for the 288-hour program in

Big Data Analytics and Optimization

conducted between August 09, 2014 and December 21, 2014.

This program is certified for quality of content, assessment and pedagogy by the Language Technologies Institute (LTI) of Carnegie Mellon University (CMU). LTI also provided assistance in curriculum development for this program.

INSOFE Inspire...Educate...Transform.

Dated this sixth day of February, two thousand and fifteen.

lti

Dr. Dakshinamurthy V Kolluru
President

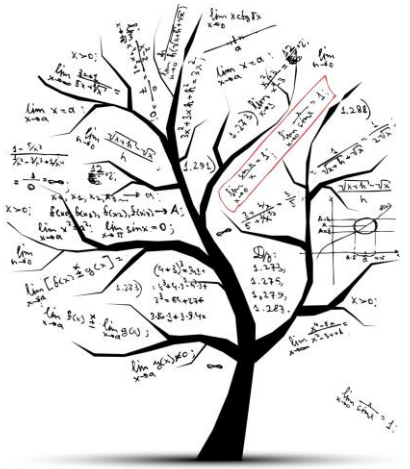
Dr. Sridhar Pappu
Executive VP - Academics

01CSE03/201408/403

Program details are on the back

Seeking Help

- Contact relevant person(s) only.
- All contact details provided in the Guidelines document.



For any subject-matter help

Maheshkumar Duvvarapu| Pavan Srungaram| Vivek Bakaraju |Shilpa Kadam

All payments and loan related

Ravindra Karanam

For any in-class issues

Yugandhar Reddy
Chakradhar Reddy

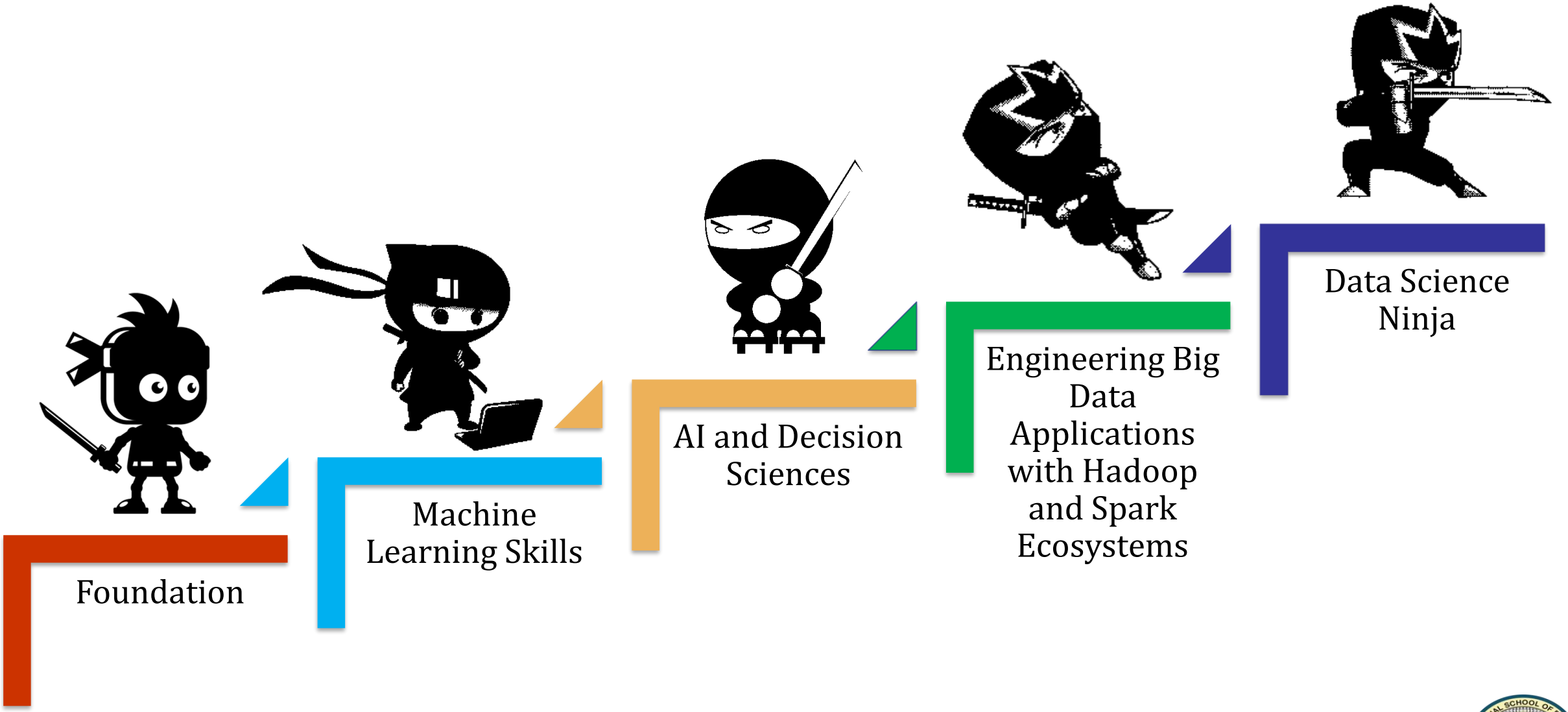
Some Good Practices

- Classes start at 9 AM. Please respect others' time.
- Please switch off phones or keep them in silent mode. Please do not distract others.
- Please do not browse internet unless instructed to do so. We want you to learn.
- We take attendance in the last hour of the day. We do want you to learn.
- Please do not ask questions on topics not yet covered. No ransom will be paid for hijacking class discussions.
- Refer to the Guidelines document in your welcome kit.
- Write all your exams. In case you are likely to miss any exam, please post a note in piazza in advance. Before Wednesday of the following week, take the exam.

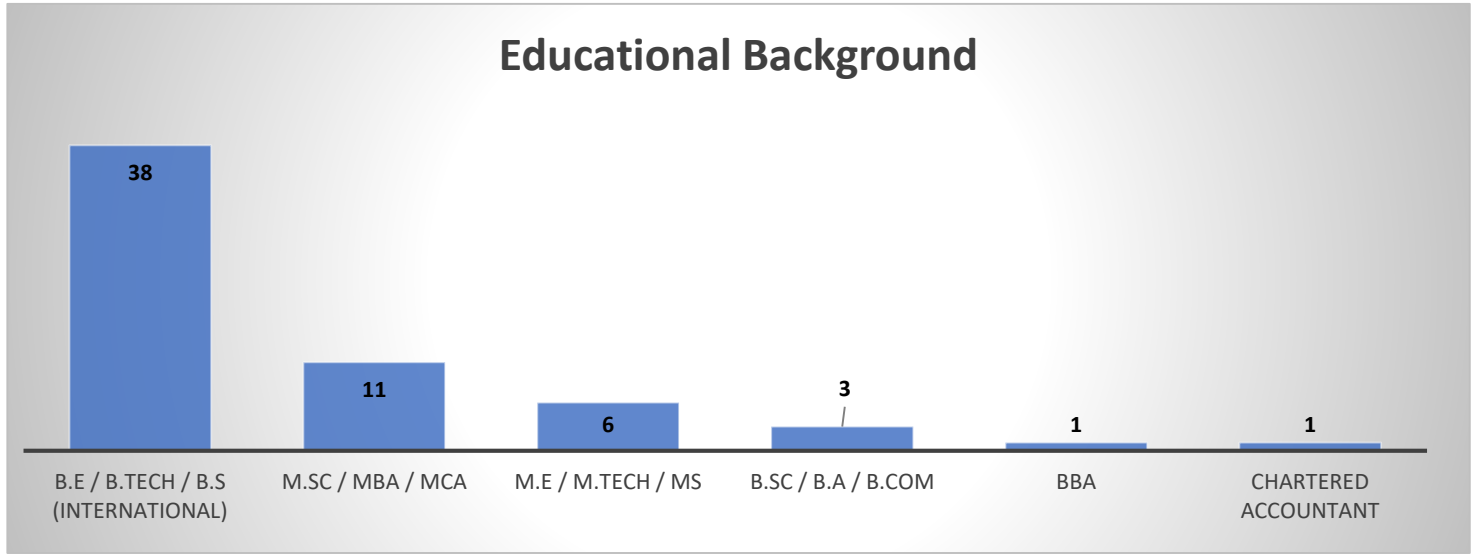

UP

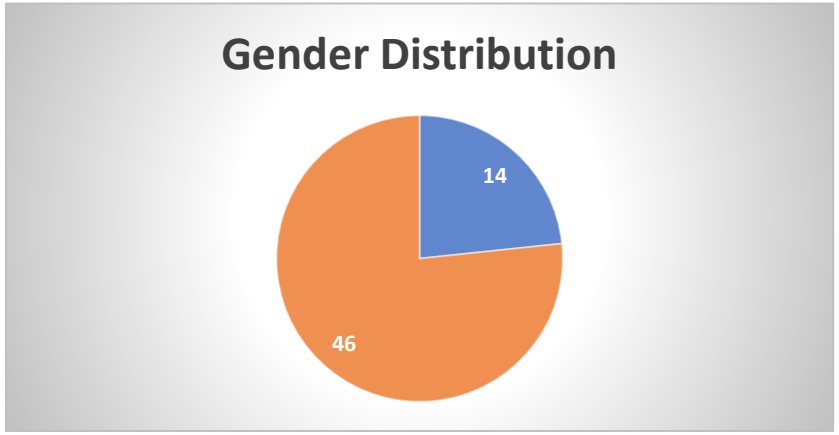

UP

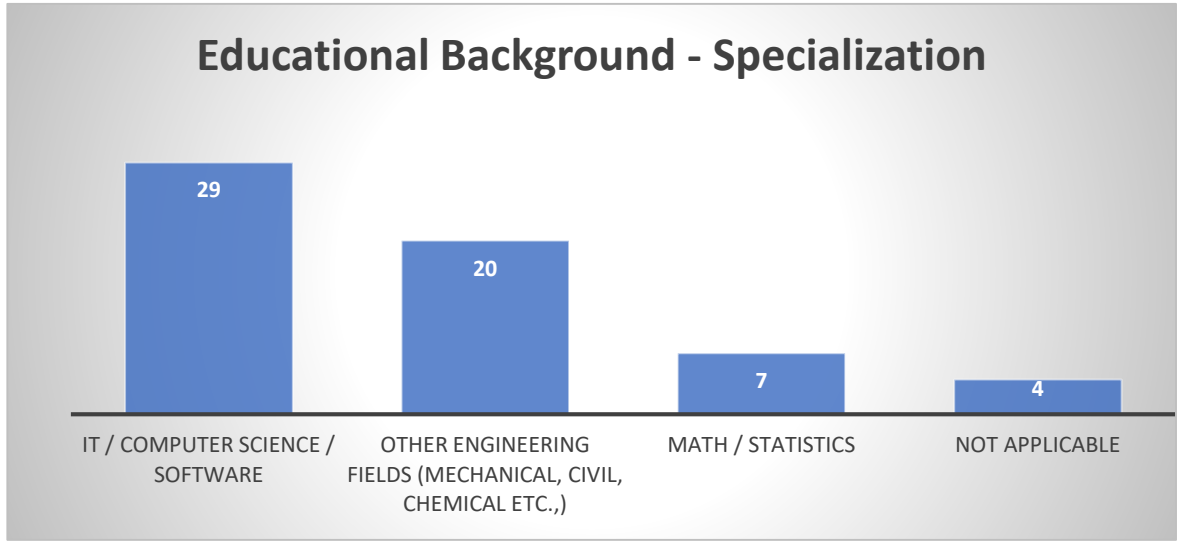
Curriculum structure – how are we heading?

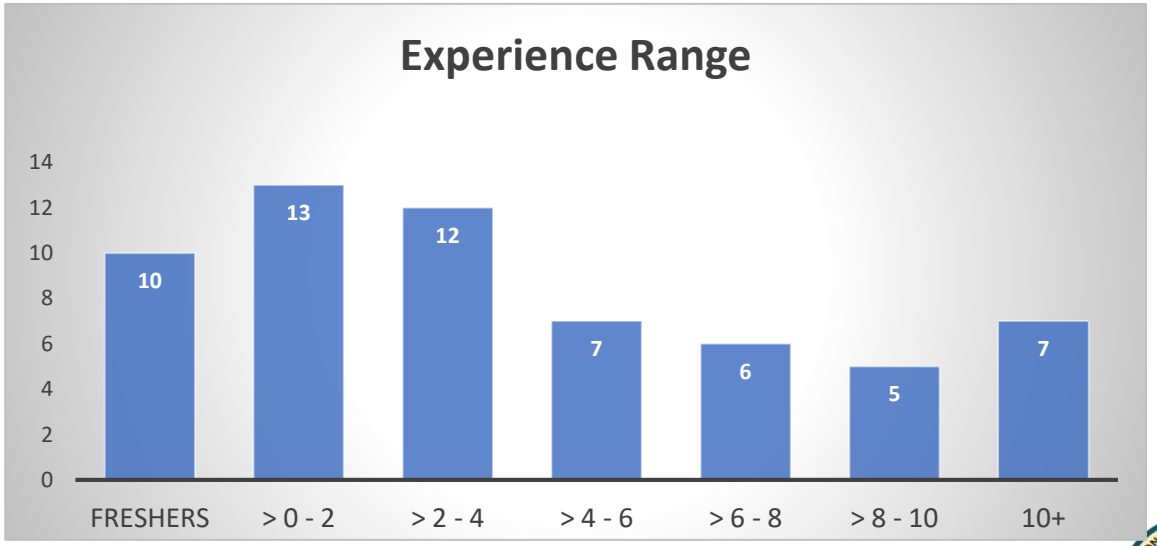


Background


Educational Background
38
11
6
3
1
1
B.E / B.TECH / B.S (INTERNATIONAL)
M.SC / MBA / MCA
M.E / M.TECH / MS
B.SC / B.A / B.COM
BBA
CHARTERED ACCOUNTANT


Gender Distribution
14
46


Educational Background - Specialization
29
20
7
4
IT / COMPUTER SCIENCE / SOFTWARE
OTHER ENGINEERING FIELDS (MECHANICAL, CIVIL, CHEMICAL ETC.,)
MATH / STATISTICS
NOT APPLICABLE


Experience Range
14
12
10
8
6
4
2
0
10
13
12
7
6
5
7
FRESHERS
> 0 - 2
> 2 - 4
> 4 - 6
> 6 - 8
> 8 - 10
10+

Break

The General Tasks of a Data Scientist

20,000-FOOT VIEW OF DATA SCIENCE

Data Science

*"Goal is in **extracting meaning from data** and **creating data products** and seeks to **use all available and relevant data** to effectively **tell a story** that can be easily **understood by non-practitioners.**"*

- From Wikipedia (not on latest version of the page)

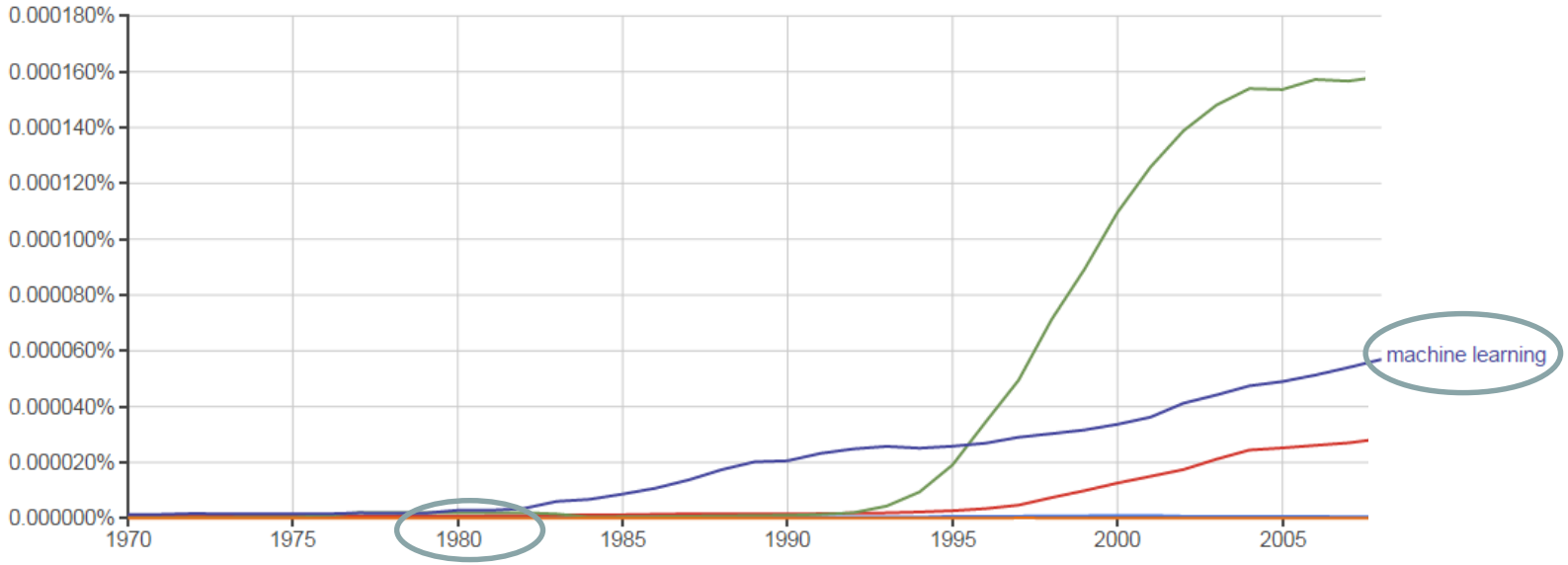
Clarifying the Jargon

- Machine Learning
- Predictive Analytics/Data Mining
- Big Data Analytics
- Data Science

ALL THESE TERMS ARE PART OF THE SAME FIELD USED AT DIFFERENT TIMES BY PEOPLE WITH DIFFERENT BACKGROUNDS

Clarifying jargon: Chronology

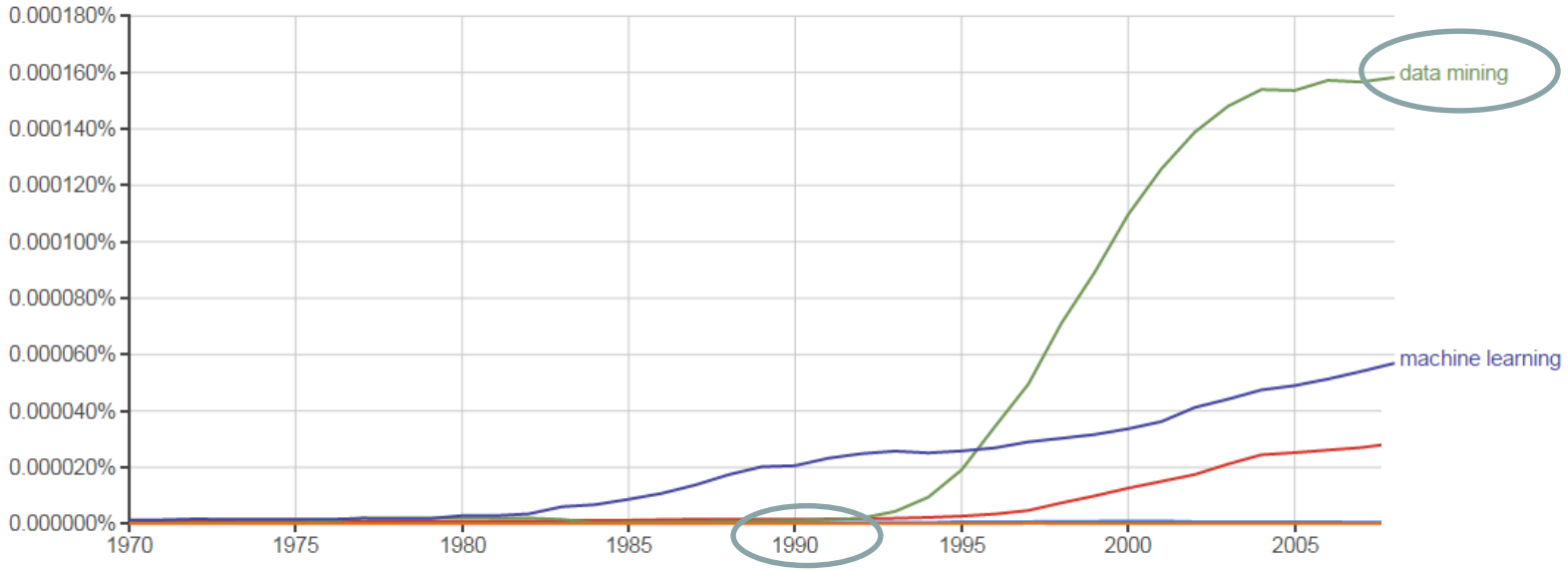
- Machine learning – 1980's
  - Computing Departments called it
  - Focus was on algorithm and the amount of data is limited



Source: *Google Books chart comparing the frequency of occurrence of "big data", "business intelligence", "data mining", "data science" y "machine learning" in the historical records of this service.*
www.mikelnino.com

Clarifying jargon: Chronology

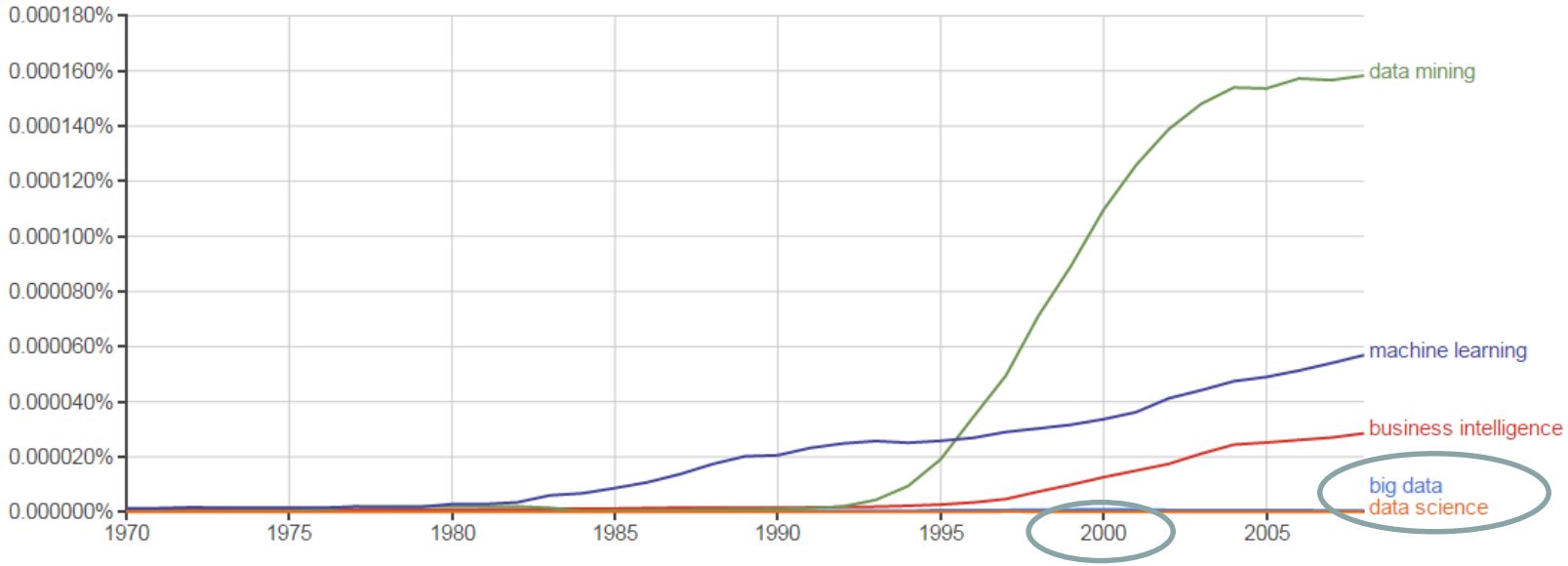
- Predictive analytics | Data Mining – 1990's
  - Business world started adopting Data Analytics
  - Used algorithms that are developed and applied on large amount of data



Source: *Google Books chart comparing the frequency of occurrence of "big data", "business intelligence", "data mining", "data science" y "machine learning" in the historical records of this service.*
www.mikelnino.com

Clarifying jargon: Chronology

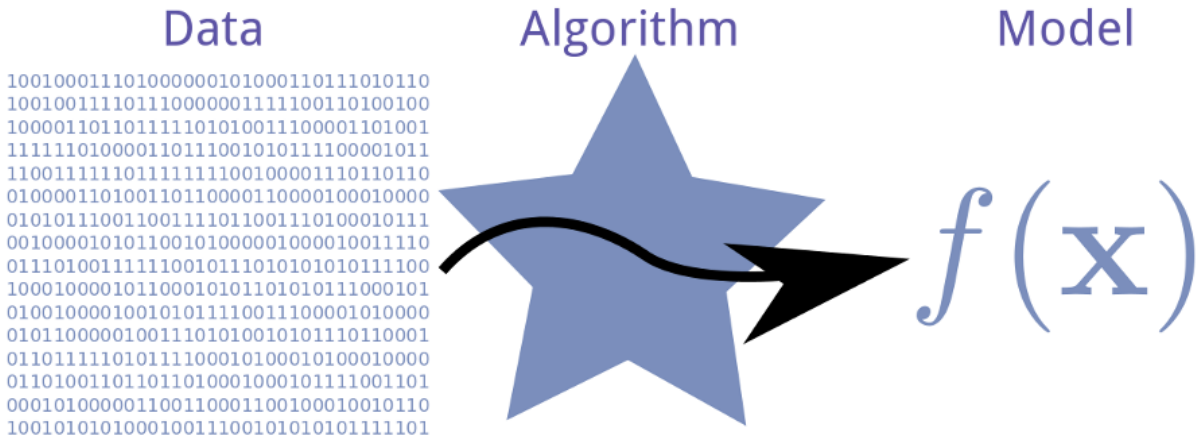
- Big Data Analytics – 2000's
  - Focus was on computing on big volume of data
  - Google, Yahoo, Facebook, Twitter etc. used to apply algorithms on much large amount of data



Source: *Google Books chart comparing the frequency of occurrence of "big data", "business intelligence", "data mining", "data science" y "machine learning" in the historical records of this service.*
www.mikelnino.com

Clarifying jargon: Chronology

- Data Science – over the last decade
  - Term is used to indicate a field where complex algorithms work on large volumes of data to solve important business problems to non-practitioners
  - A lot of emphasis on Visualization and story telling



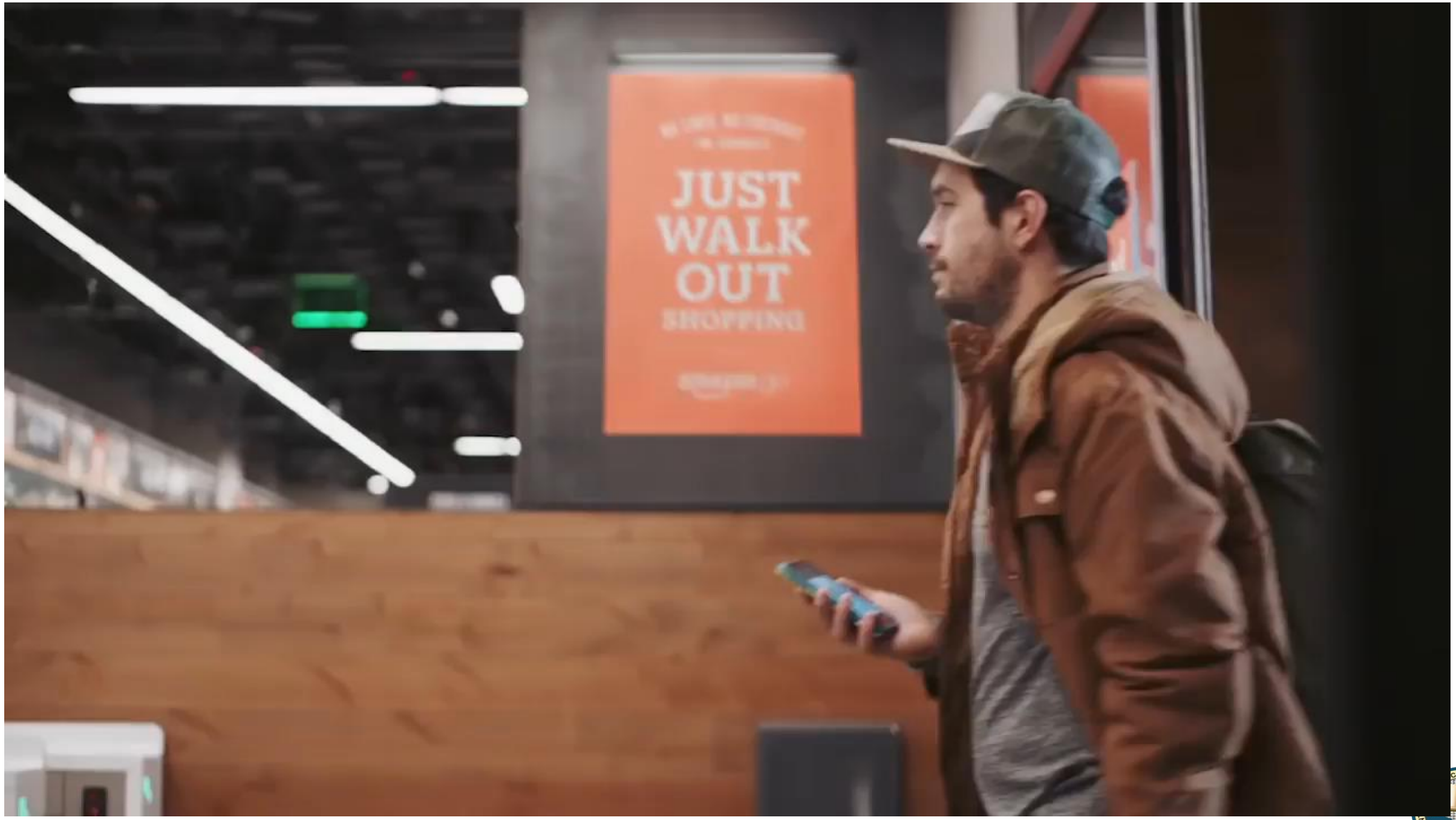
Source: *Google Books chart comparing the frequency of occurrence of "big data", "business intelligence", "data mining", "data science" y "machine learning" in the historical records of this service.* www.mikelnino.com

What are various sources of data?

Conventional data sources

Evolution with internet of things

https://www.youtube.com/watch?v=Q3ur8wzzhBU

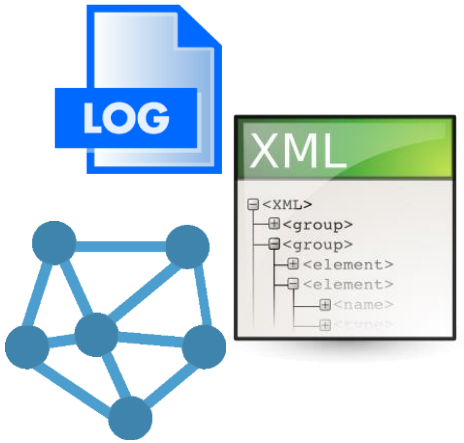

JUST
WALK
OUT
SHOPPING

What are various Types of data?

Examples

Structured Data

Semi-structured Data



Unstructured Data

Volume

Variety

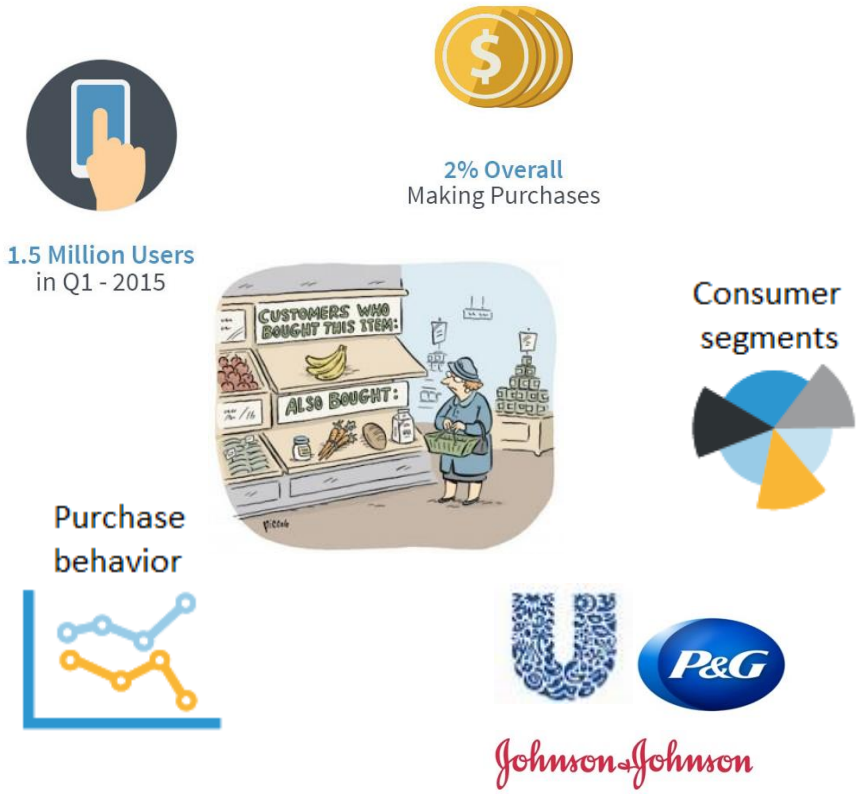
Velocity

Big versus Nano data

- Extremely small data sets (often found in healthcare, national calamities) are challenging
  - Most techniques fail to catch subtle trends
  - Deep learning, ensembles are able to tackle these problems better than before

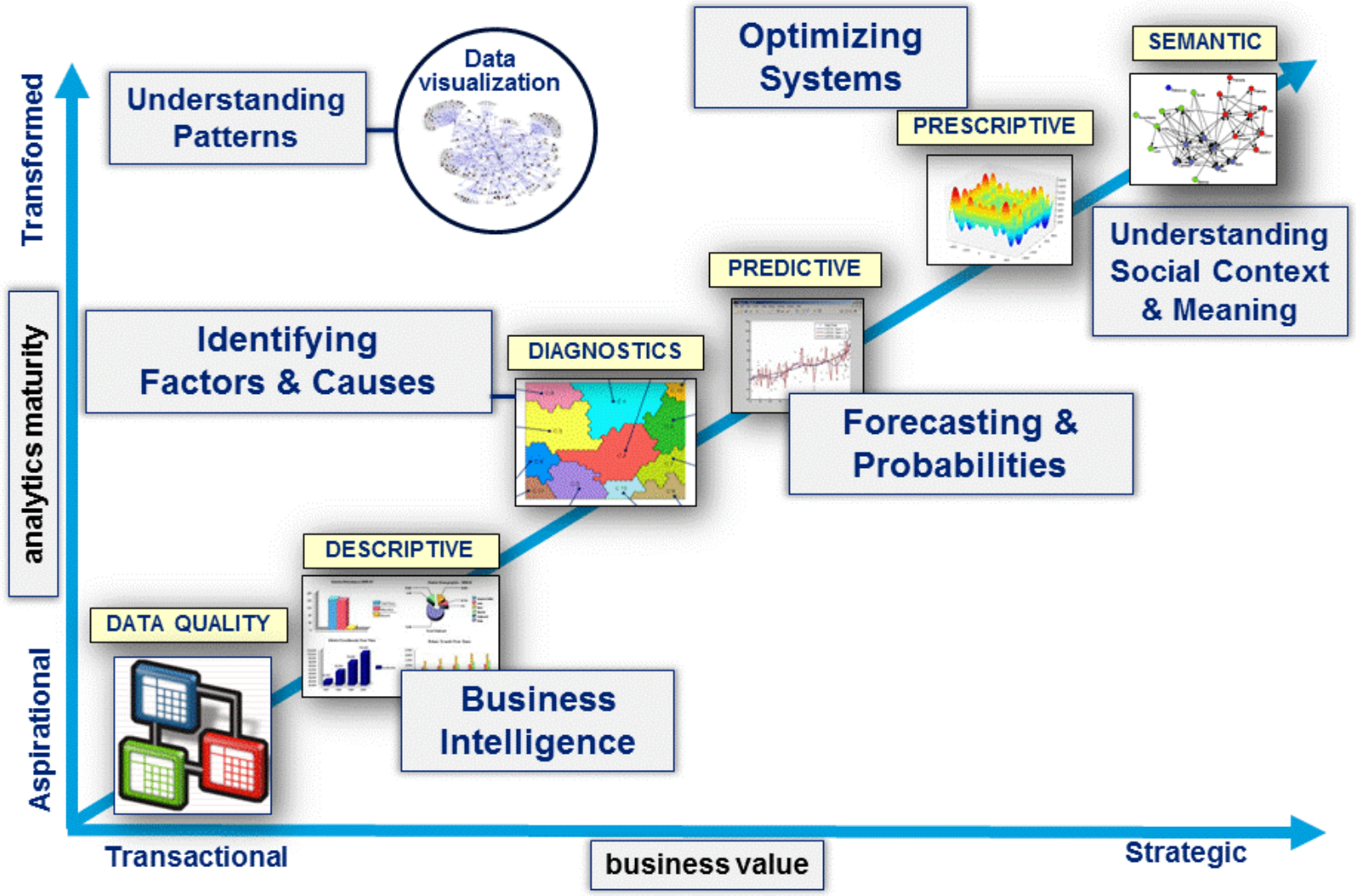
What does the business need?

Quickly leverage data using data analytical methods in strategic decision making



- FMCG is a consumer driven business and enterprises continuously look out for ways to:
  - Innovate new products and services
  - Get visibility on trade promotions
  - Get Quick insights on consumer behaviour
  - Counter fierce competition, etc.

BI to Predictive Analytics & Beyond



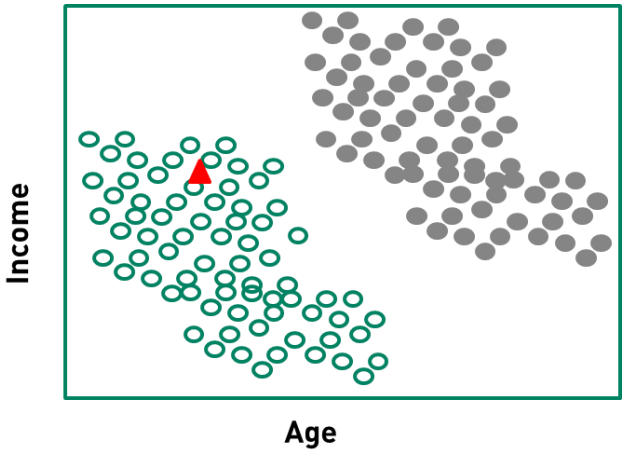
Types of problems solved in Data Science

Machine Learning Tasks

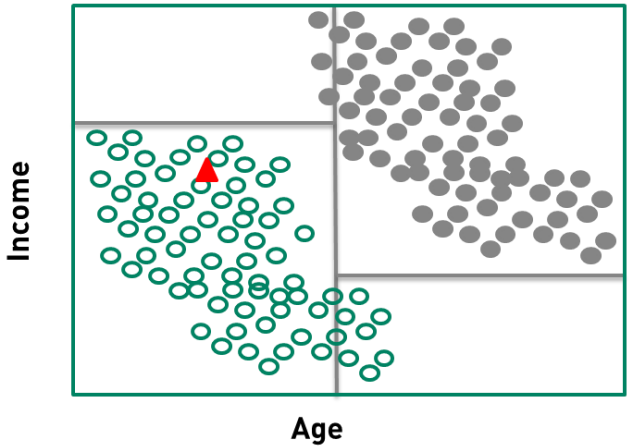
- Classification
- Regression
- Clustering
- Optimization

SOME CLASSIFICATION METHODS

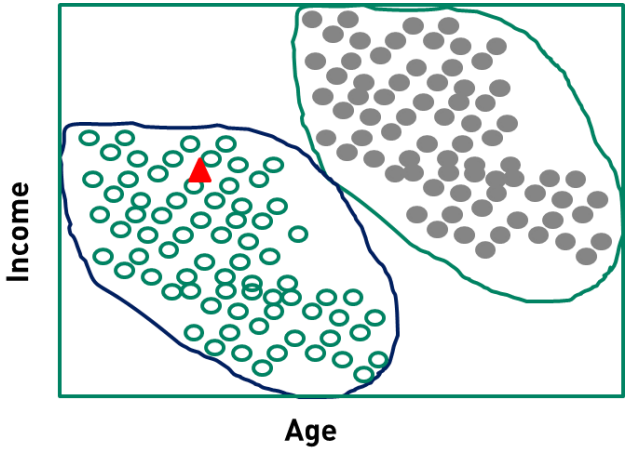
Binary Classification



Decision Trees

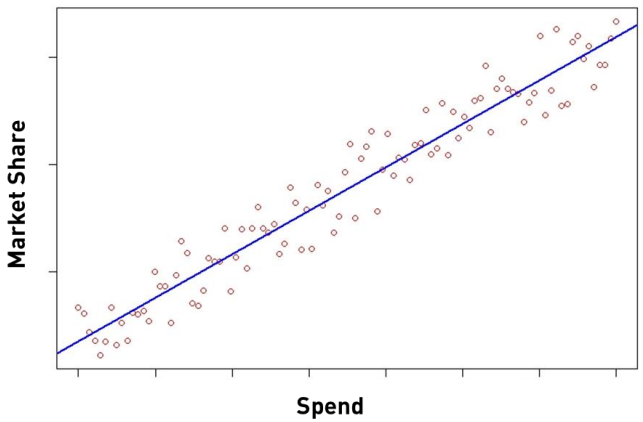


Naïve Bayes

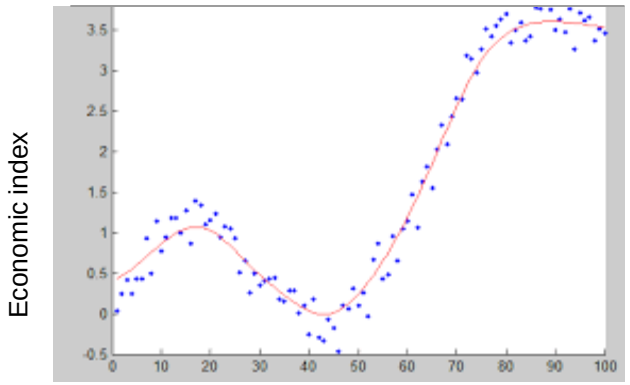


SOME PREDICTION METHODS

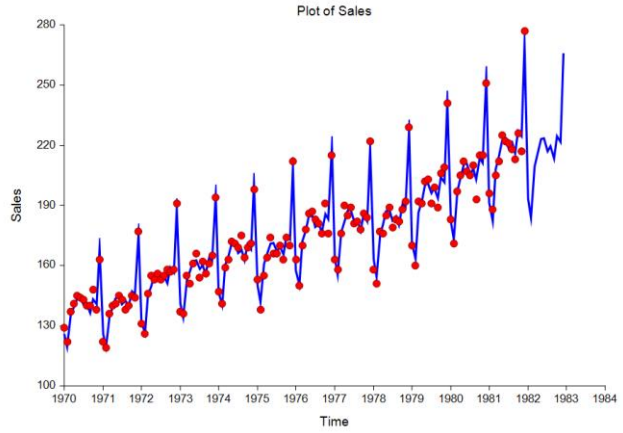
Linear Regression



Polynomial Regression



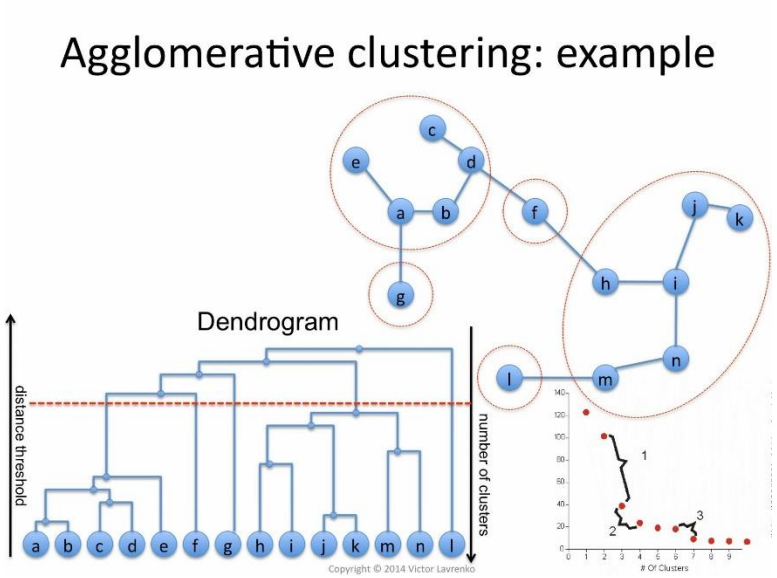
Time Series Forecasting

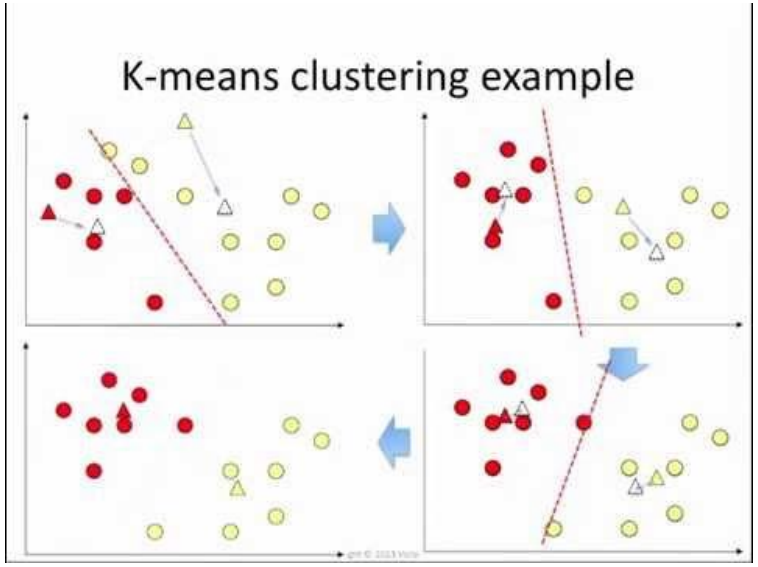


Learning Method

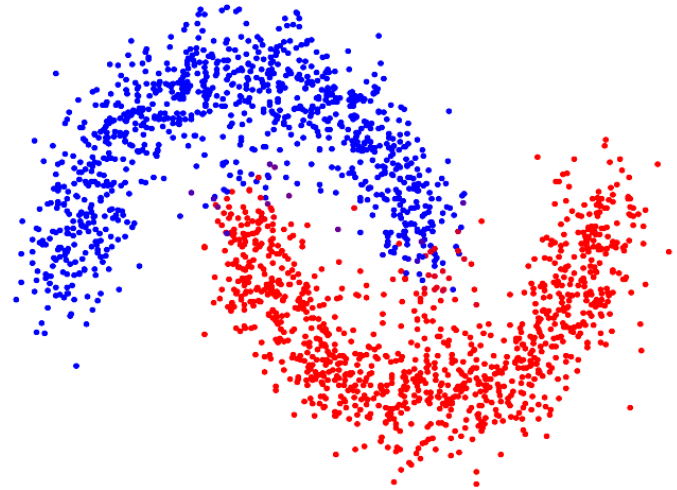
Supervised Learning

Some Clustering methods





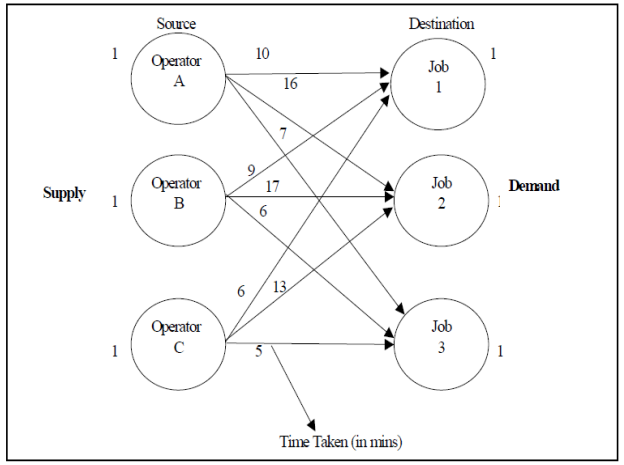
Spectral clustering example



Unsupervised Learning

SOME OPTIMIZATION METHODS

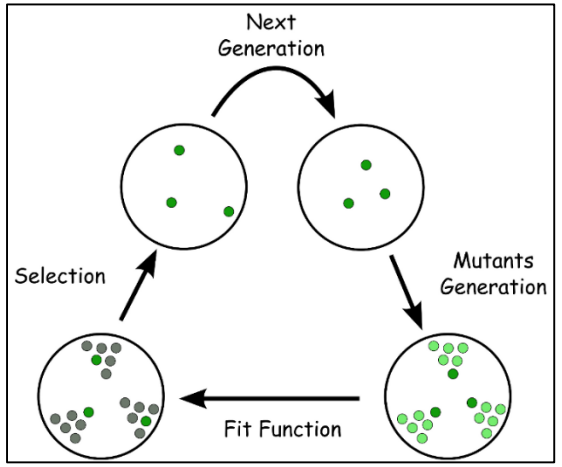
Assignment Problem



Transportation Problem

Source	Destination D_1	D_2	D_3	D_4	Supply
S_1	19	30	50	10	7
S_2	70	30	40	60	9
S_3	40	8	70	20	18
Demand	5	8	7	14	

Genetic Algorithm



Key terms: Minimize / Maximize

A Catalog of methods

Descriptive statistics

- Mean, Median, Mode
- Correlations
- Sampling & Distributions
- T Test, F Test
- Normal distribution
- Poisson distribution. Etc.

Classification (Unsupervised)

- Clustering
  - K-means
  - Hierarchical
- Association Rules, etc.

Prediction (Supervised)

- Multiple linear regression
- Neural nets
- Support vector machines
- Gradient Boosting
- Etc.

Classification (Supervised)

- Logistic regression
- Decision trees
- Bayesian analysis
- Random Forest
- Etc.

Optimization

- Operations Research
- Linear Programming
- Genetic Algorithm, etc.

WHY SO MANY METHODS?

There are many different methods for prediction and classification.

Each method has its advantages and disadvantages.

The usefulness of a particular method can depend on factors such as –

- the particular goal of the analysis
- size of the data set
- the types of patterns that exist in the data
- whether the data meet some underlying assumptions of the method

Different methods can lead to different results, and their performance can vary.

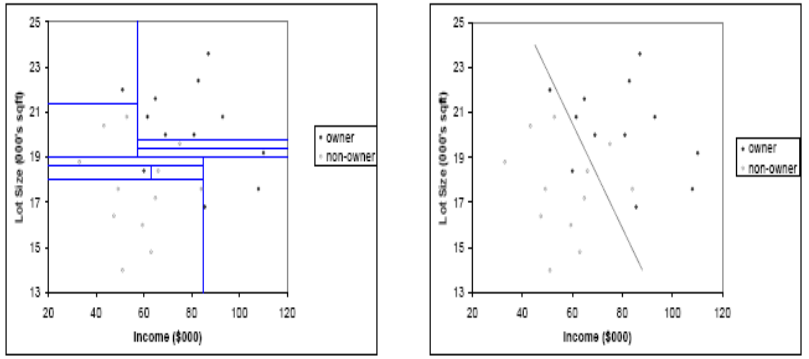
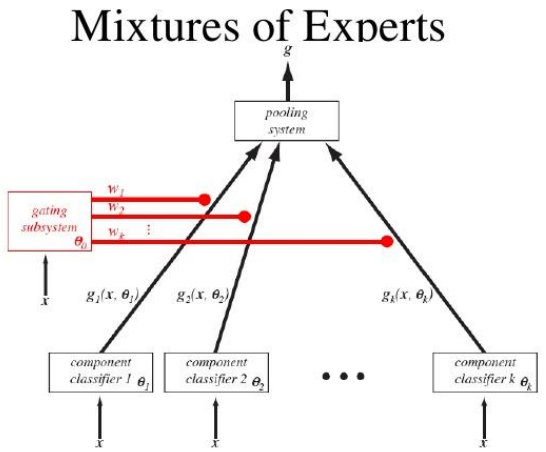


Figure 1.1: Two different methods for separating buyers from non-buyers



The General Tasks of a Data Scientist

1. Get a little domain understanding
2. Define the problem statement well

Understanding the Problem

The Approach

Results - The Streak!

The General Tasks of a Data Scientist

3. Pre-process data to fix data issues like duplicates, missing values, etc.
4. Visualize data to the extent possible for better understanding and to see basic patterns
5. Identify what kind of a problem it is (Prediction/Forecasting, Classification, Optimization and/or Managing Big Data)
6. Identify appropriate modeling techniques and build models
7. Analyze results and iterate, as needed; DO NOT trust software outputs blindly – Remember: Garbage In, Garbage Out
8. Visualize outputs and Communicate

High-level steps

Insight : frequency of shoppers across marks

Broadly, here are the steps

Store data

Retrieve data

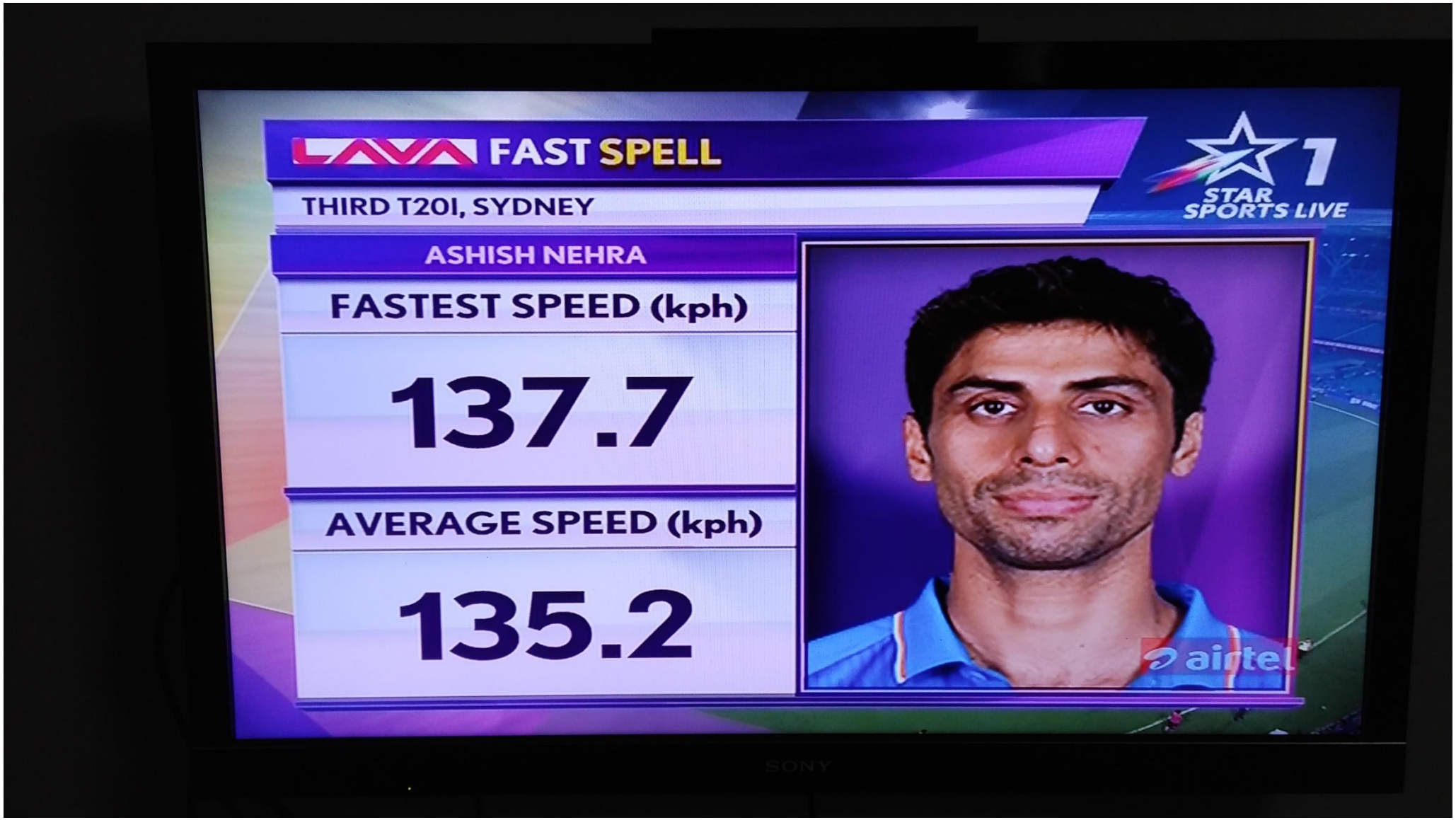
Process data

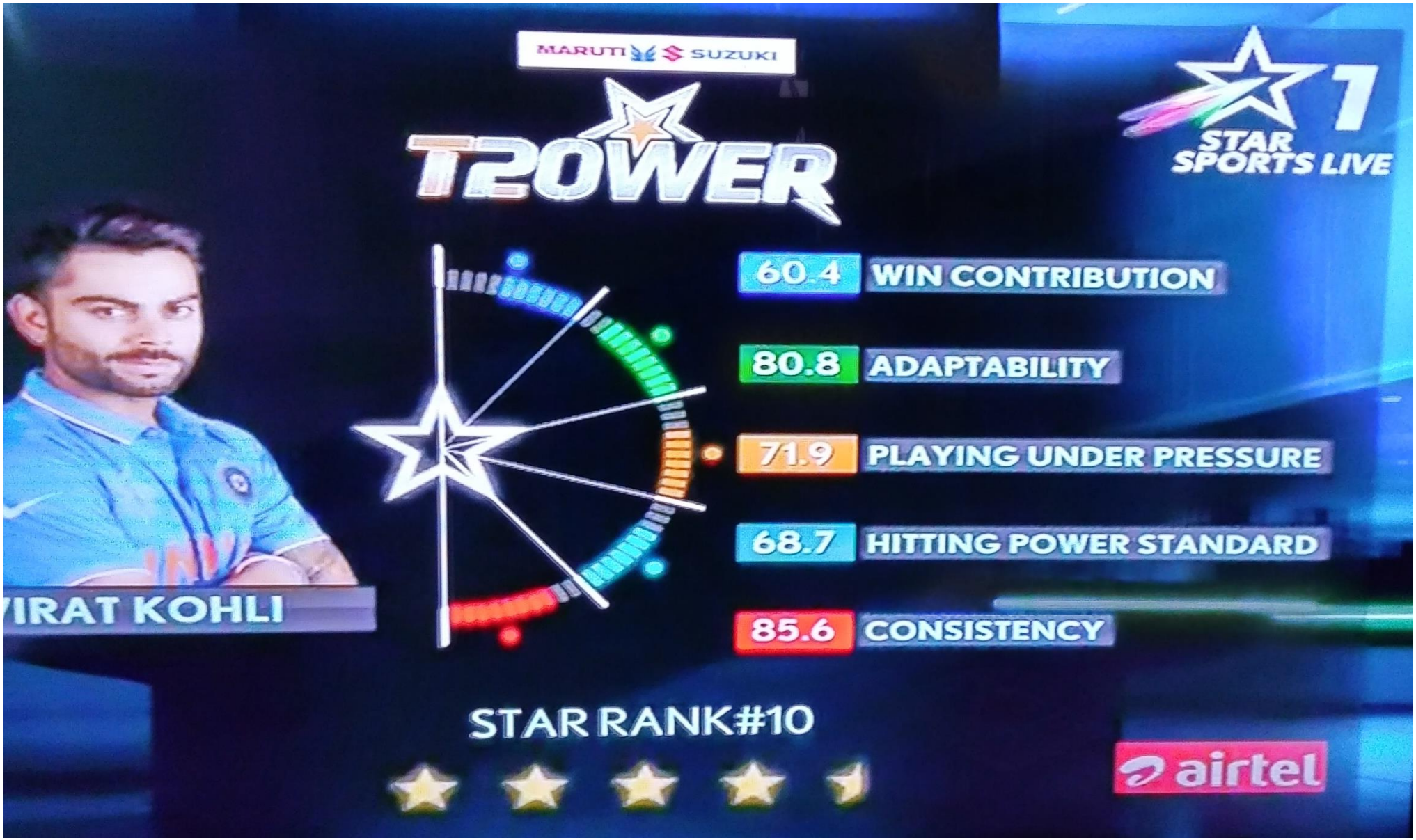
Analyze/Visualize Data

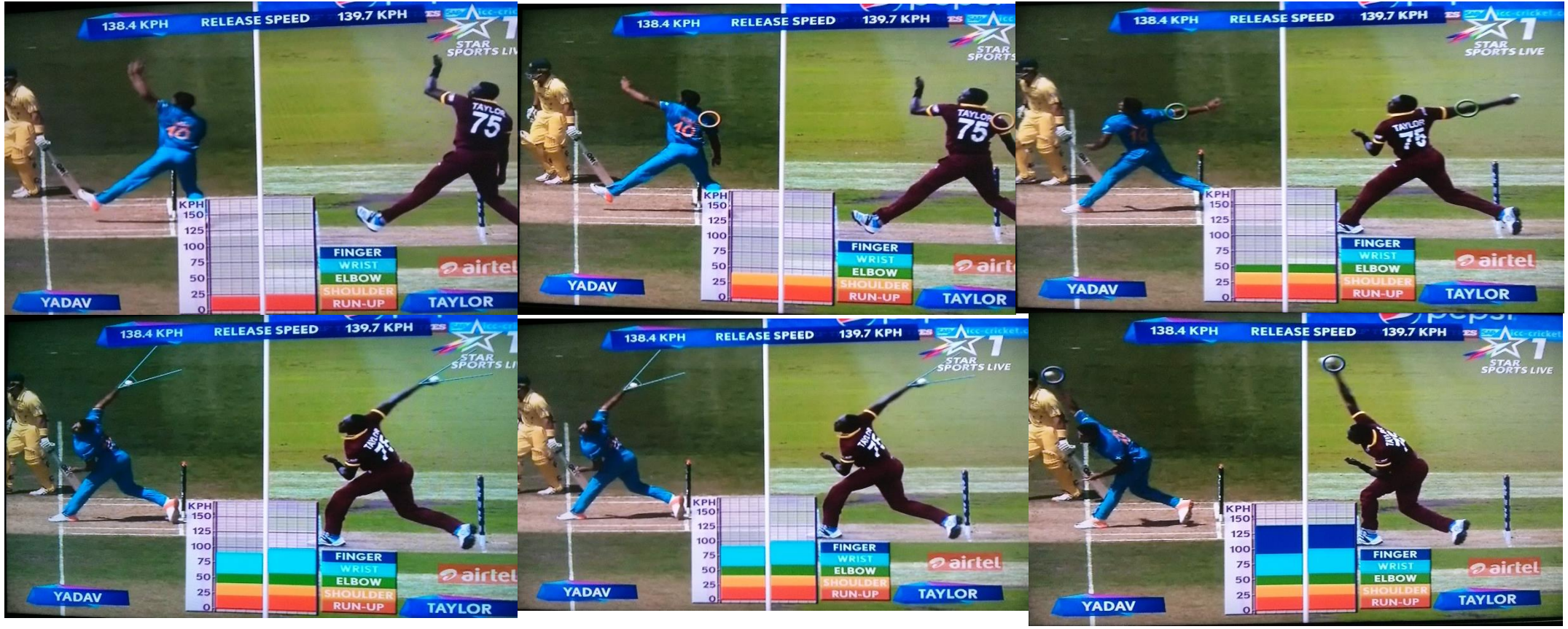
Interpreting Results and Story Telling

Capture data

Applications and Significance of Data Science


LAVA FAST SPELL
THIRD T20I, SYDNEY
STAR SPORTS 1 LIVE
ASHISH NEHRA
FASTEST SPEED (kph)
137.7
AVERAGE SPEED (kph)
135.2
airtel
SONY


MARUTI SUZUKI
T20WER
STAR SPORTS LIVE 1
60.4 WIN CONTRIBUTION
80.8 ADAPTABILITY
71.9 PLAYING UNDER PRESSURE
68.7 HITTING POWER STANDARD
85.6 CONSISTENCY
IRAT KOHLI
STAR RANK#10
airtel


138.4 KPH
RELEASE SPEED
139.7 KPH
STAR SPORTS LIVE
KPH
150
125
100
75
50
25
0
FINGER
WRIST
ELBOW
SHOULDER
RUN-UP
YADAV
TAYLOR
airtel


REDEVILS
0868851

Descriptive Statistics

World T20 Finals 2014:

Kohli scored 77;

Spinners picked 3 wickets

Stock Market Trends

Google searches can predict stock market crashes: Study

London, July 29: A rise in Google searches for terms relating to business and politics can predict a future stock market crash, researchers have claimed.

A team of researchers from Warwick Business School in the UK and Boston University in the US has developed a method to automatically identify topics that people search for on Google before subsequent stock market falls. Applied to data between 2004 and 2012, the method shows that increases in searches for business and politics preceded falls in the stock market.

The researchers suggest that this method could be applied to help identify warning signs in search data before a range of real world events. "Search engines, such as Google, record almost everything we search for," said Chester Curme, research fellow at Warwick Business School and lead author of the study. "Records of these search queries allow us to learn about how people gather information online before making decisions in the real world. So there's potential to use these search data to anticipate what large groups of people may do," Mr Curme said.

In previous studies, Mr Curme and his colleagues, Tobias Preis and Suzy Moat of WBS, and H Eugene Stanley of BU, have demonstrated that usage data from Google and Wikipedia may contain early warning signs of stock market moves. — *PTI*

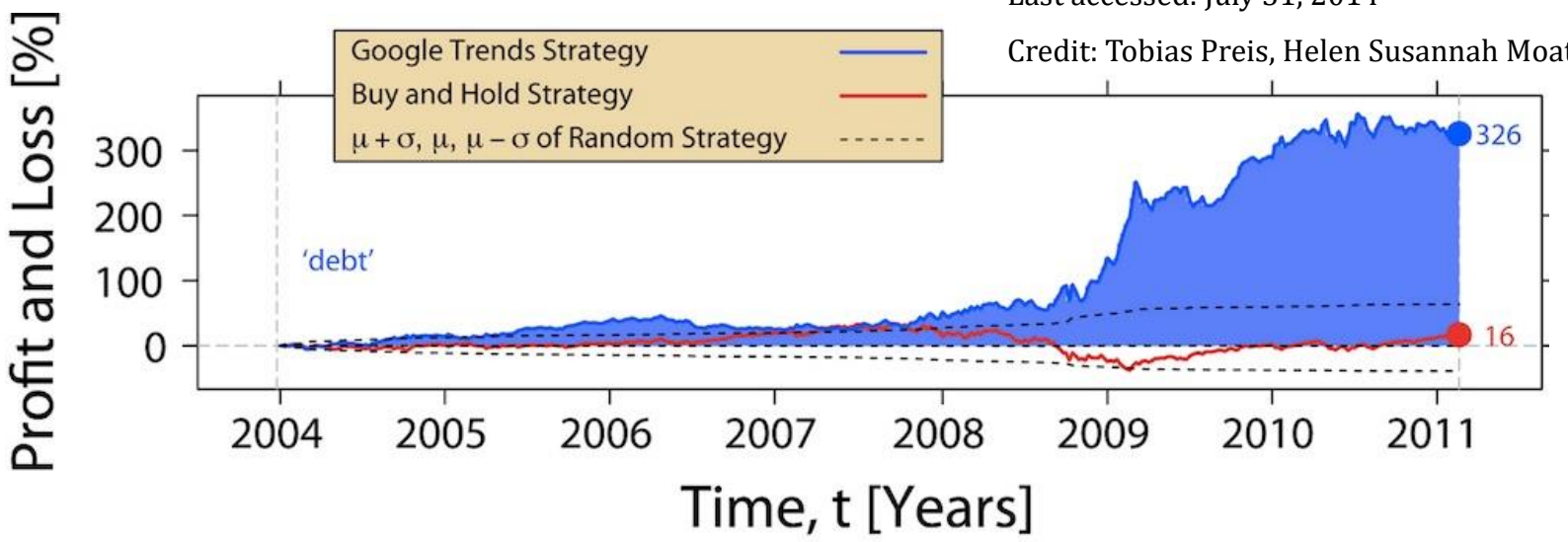
Source: http://epaper.deccanchronicle.com/articledetailpage.aspx?id=732781

Last accessed: July 31, 2014

Source: http://www.livescience.com/29016-google-predicts-stock-market.html

Last accessed: July 31, 2014

Credit: Tobias Preis, Helen Susannah Moat, H. Eugene Stanley



Medical non-adherence is likened to be an invisible epidemic

Non-adherence = # of doses not taken or taken incorrectly that jeopardizes patient's therapeutic outcome

Note: Lack of adherence to non-pharmacologic treatments, such as recommended lifestyle changes, is not included

Stakeholders	Medical non-adherence impact	Est. $$$
• Insurers • Employers • Patients	• Increases heathcare costs due to disease-related complications	• $ 290 B Source: New England Healthcare Institute
• Pharmaceutical companies • Pharmacies • Pharmacy benefit managers	• Erodes profits due to prescriptions never filled and medications not taken often enough	• $ 188 B Source: Capgemini

INDUSTRY ACTIONS

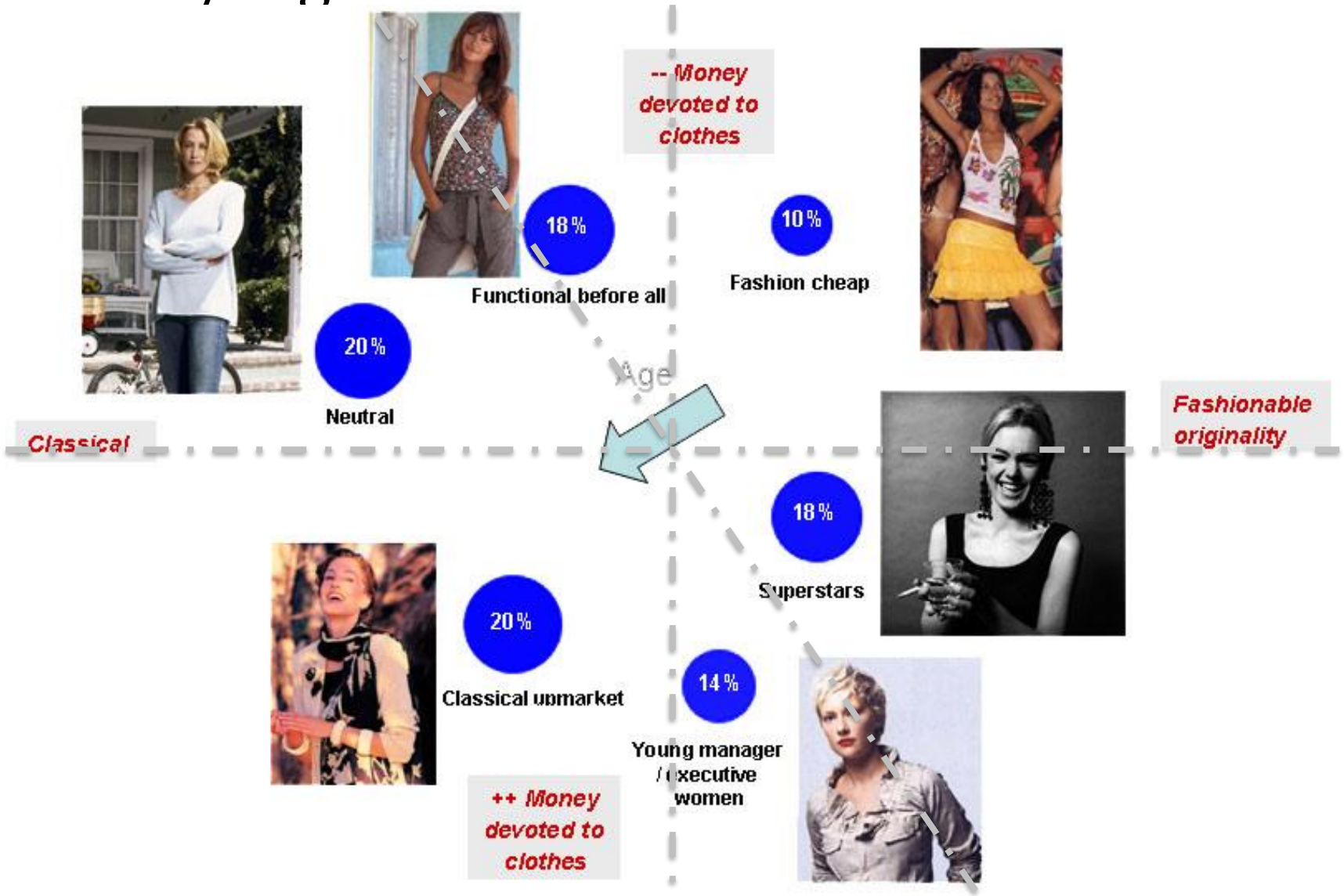
- Free drugs (*Netherlands*)
- Lower co-pays (*Aetna*)
- Reminders
- Reward points (*HealthPrize*)
- More targeted reminders (*RxAnte* uses algorithms)

The entire medical industry is scrambling to address this issue !!!

Advanced analytics may provide the insights to help solve this puzzle

Source: http://pharmaceuticalcommerce.com/business_finance?articleid=26718

Classification/Segmentation



Source: http://www.bayesia.com/en/applications/marketing.php

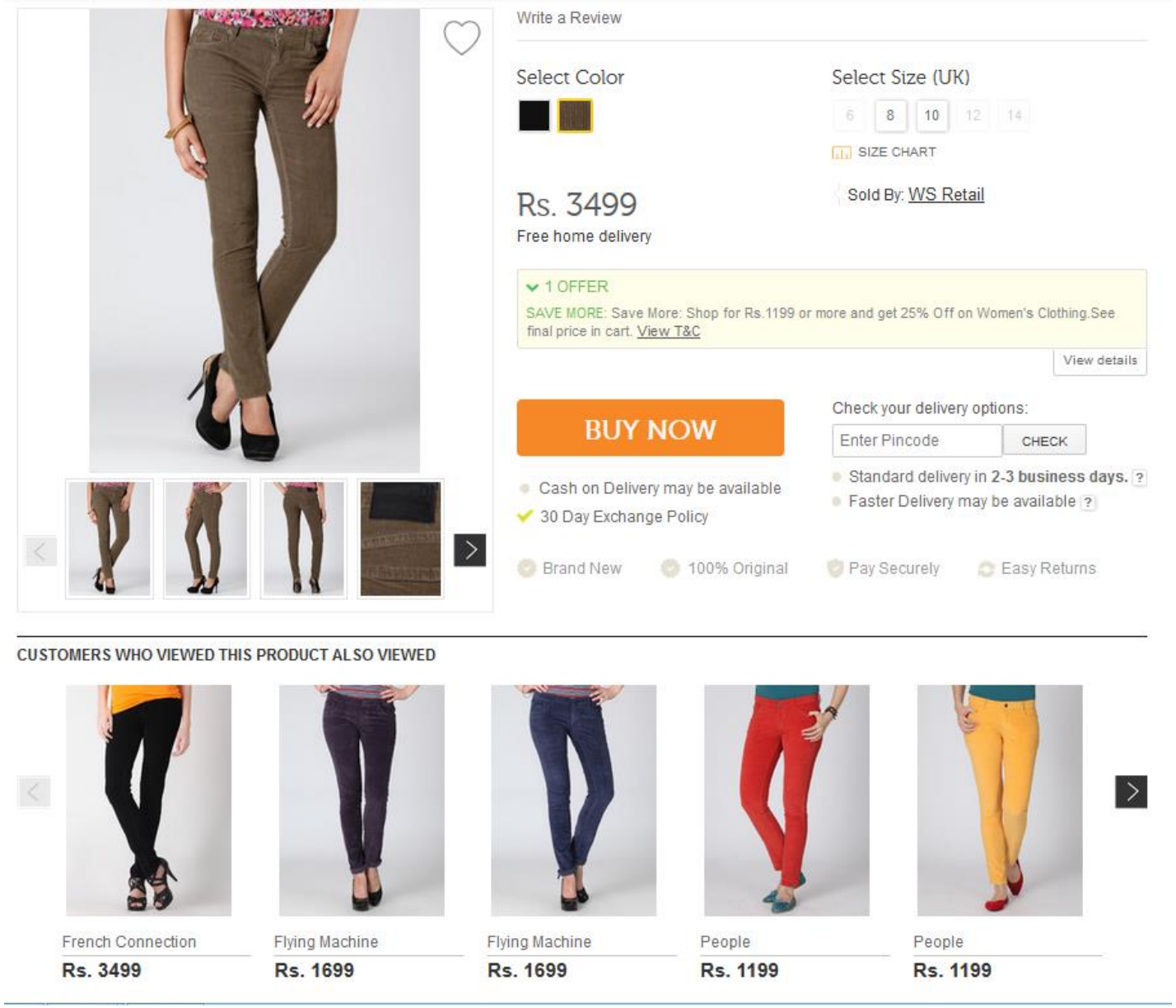
Last accessed: August 4, 2014

Churn can be heartbreaking

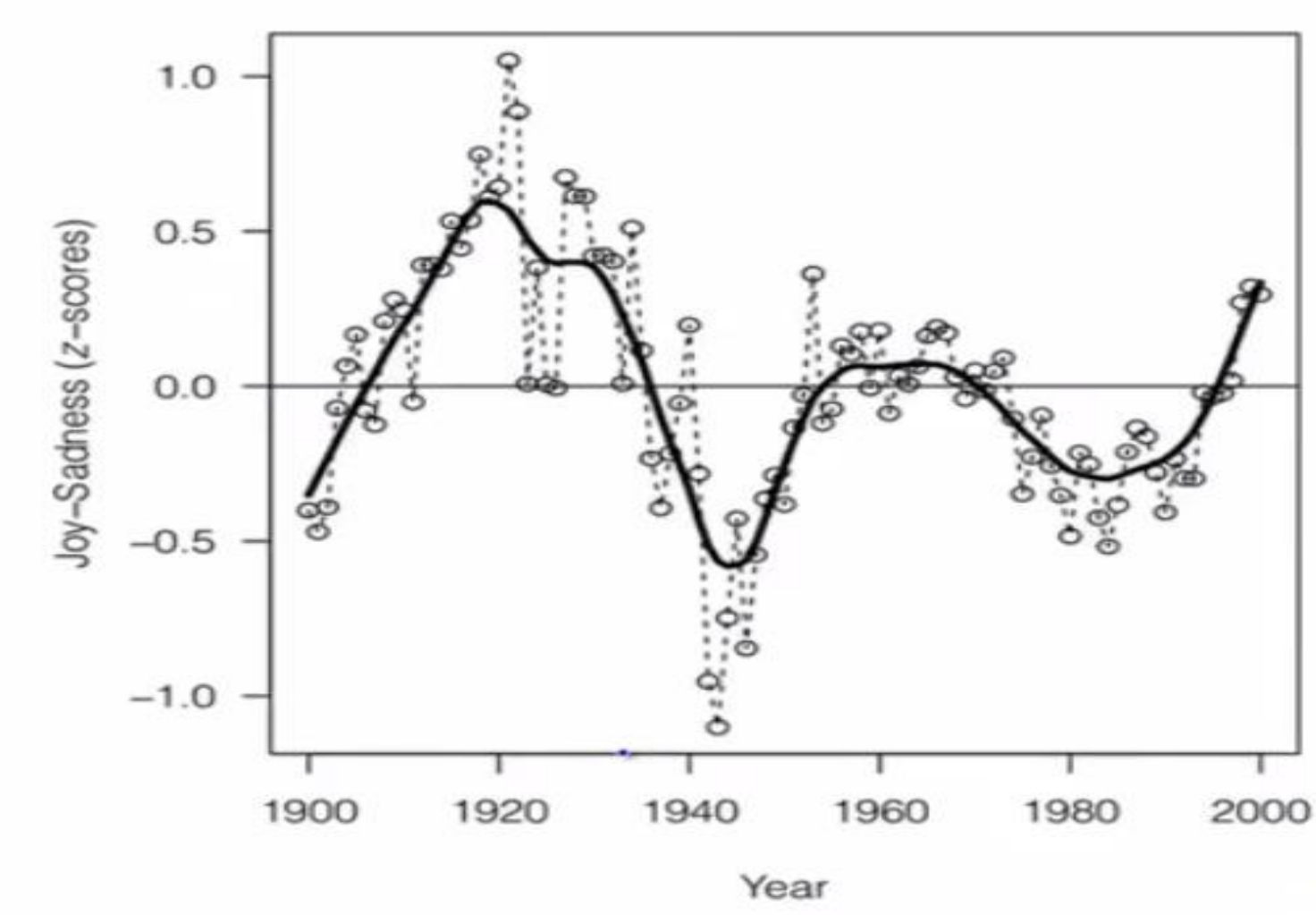
Source: http://princesswithapen.hubpages.com/hub/Is-my-girlfriend-cheating-on-me-Signs-of-a-cheating-girlfriend

Last accessed: August 04, 2014

Recommendation Engine



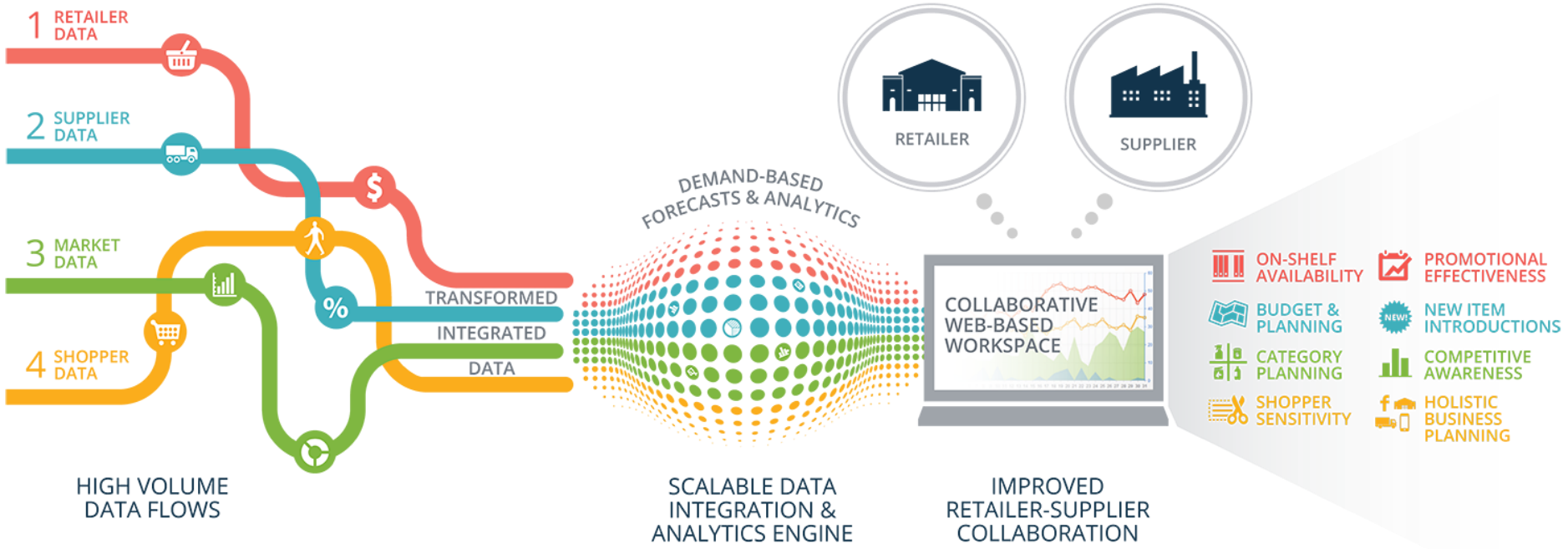
Text Mining: Are Americans happy or unhappy?



Customer Analytics

Retail and Telecommunications

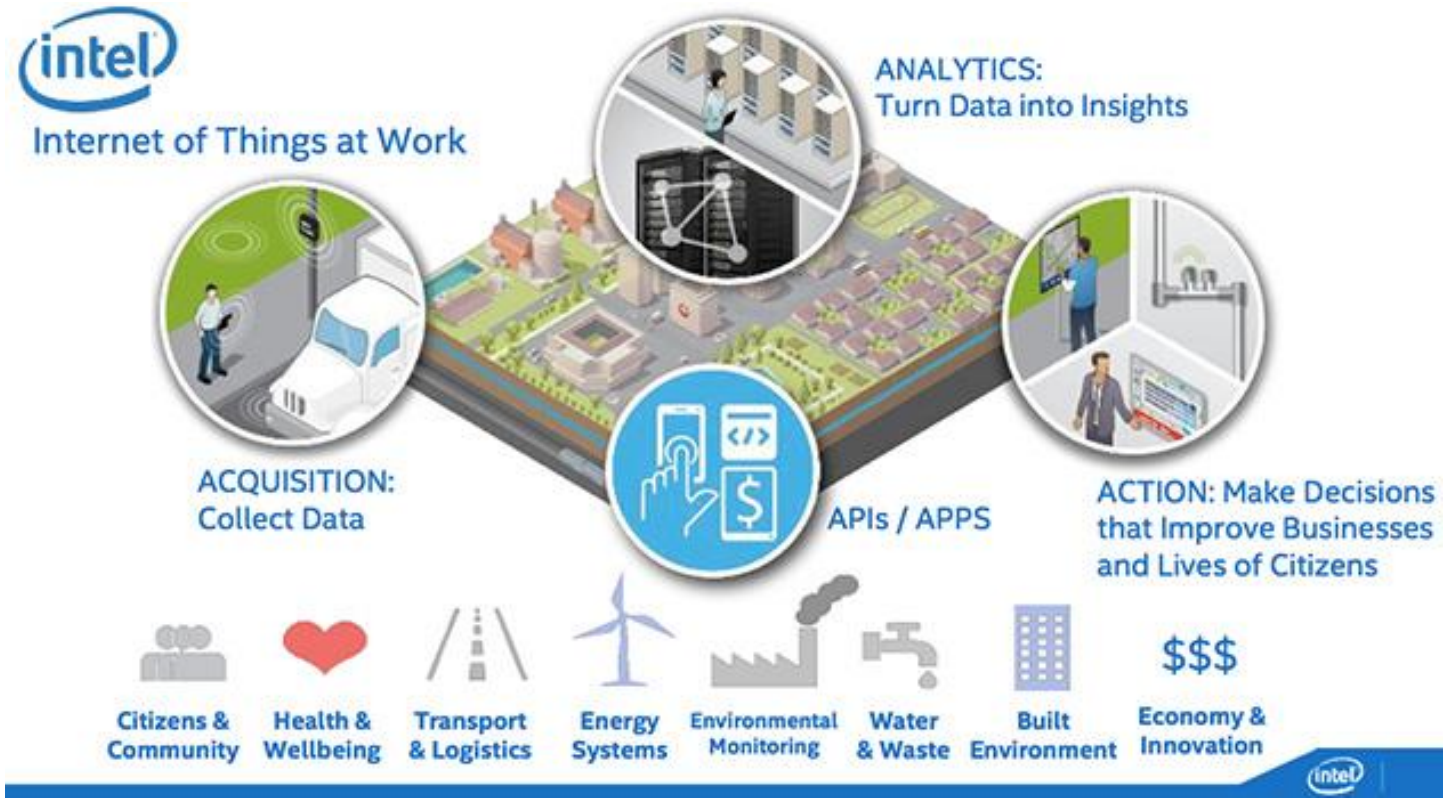
- Customer retention; Enhancing supply chain efficiency; Improving customer service quality; Planning store locations; Cross-selling and upselling; Recommendation systems; Sales forecasting



Other important applications

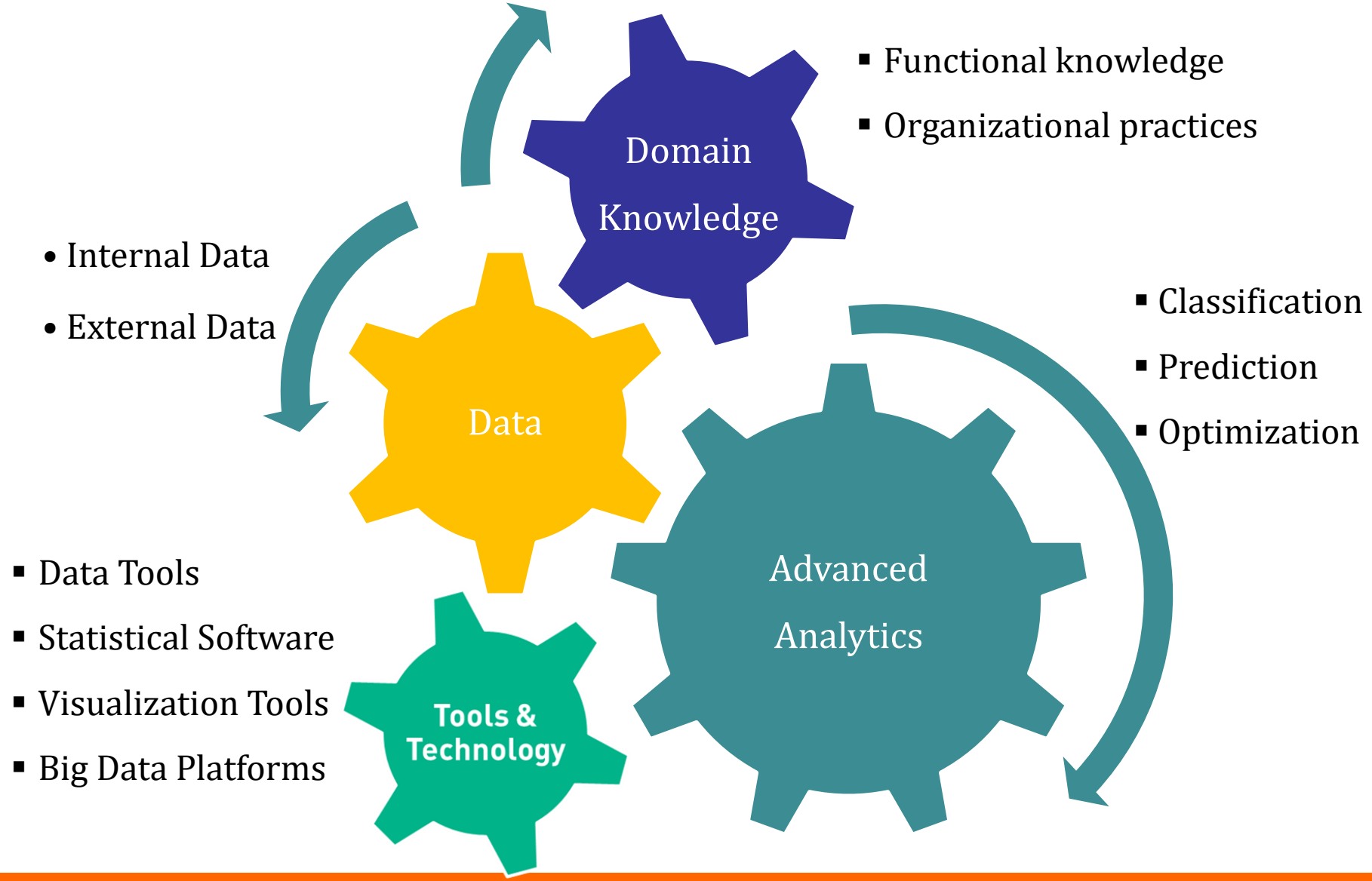
Government

- Policy planning; Effective use of resources; Security against terrorist attacks; Effective policing by understanding crime patterns; Weather predictions; Calamity predictions

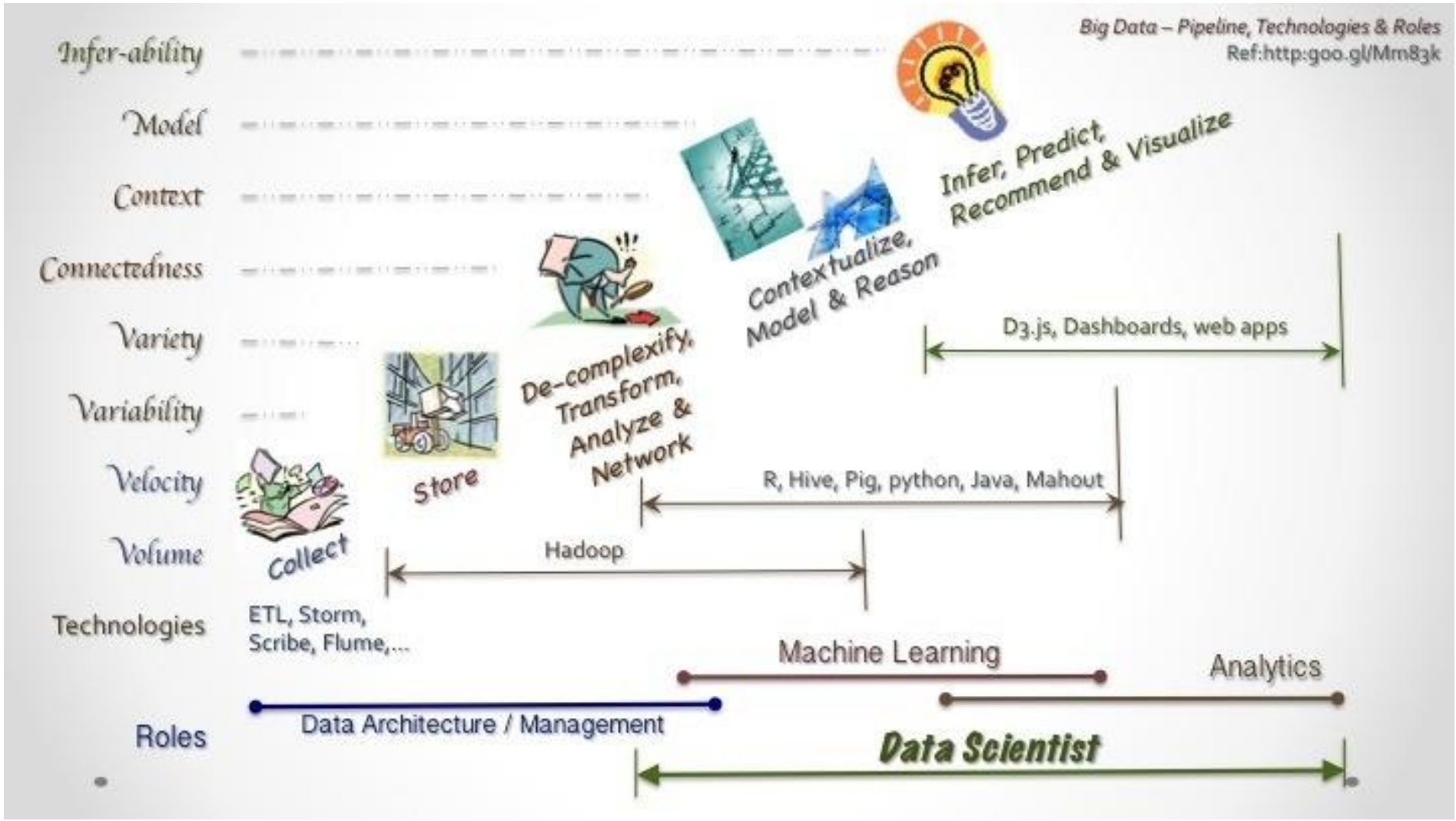



Mobile Data
Messages
Social Media
Email
Audio
Data Everywhere!
You can’t escape it!

You Can Make Better Decisions Using Data Science / Big Data Analytics

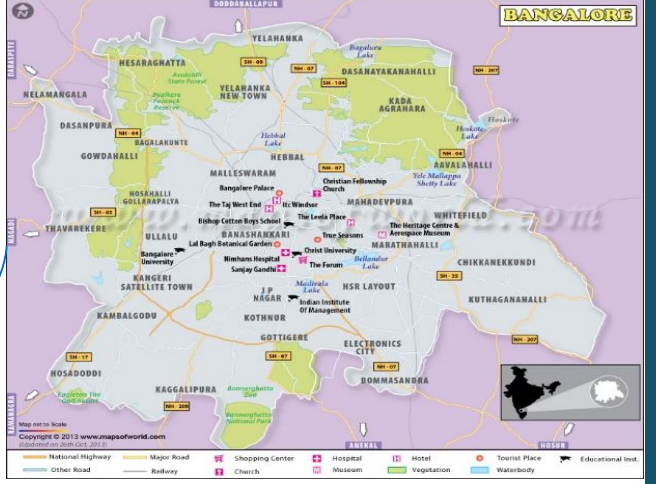


Roles



https://www.quora.com/What-are-the-most-valuable-skills-to-learn-for-a-data-scientist-now

INSOFE
Inspire...Educate...Transform.



HYDERABAD

Office and Classrooms

Plot 63/A, Floors 1&2, Road # 13, Film Nagar, Jubilee Hills, Hyderabad - 500 033
+91-9701685511 (Individuals)
+91-9618483483 (Corporates)

BENGALURU

Office

Incubex, #728, Grace Platina, 4th Floor, CMH Road, Indira Nagar, 1st Stage, Bengaluru – 560038
+91-9502334561 (Individuals)
+91-9502799088 (Corporates)

Classroom

KnowledgeHut Solutions Pvt. Ltd., Reliable Plaza, Jakkasandra Main Road, Teacher's Colony, 14th Main Road, Sector – 5, HSR Layout, Bengaluru - 560102

Social Media

Web:	http://www.insofe.edu.in
Facebook:	https://www.facebook.com/insofe
Twitter:	https://twitter.com/Insofeedu
YouTube:	http://www.youtube.com/InsofeVideos
SlideShare:	http://www.slideshare.net/INSOFE
LinkedIn:	http://www.linkedin.com/company/international-school-of-engineering

This presentation may contain references to findings of various reports available in the public domain. INSOFE makes no representation as to their accuracy or that the organization subscribes to those findings.