

# What is there that Can be Learned?

**Erniel B. Barrios**

*Professor of Statistics*

*School of Statistics*

*University of the Philippines Diliman*

# Call and Data Records

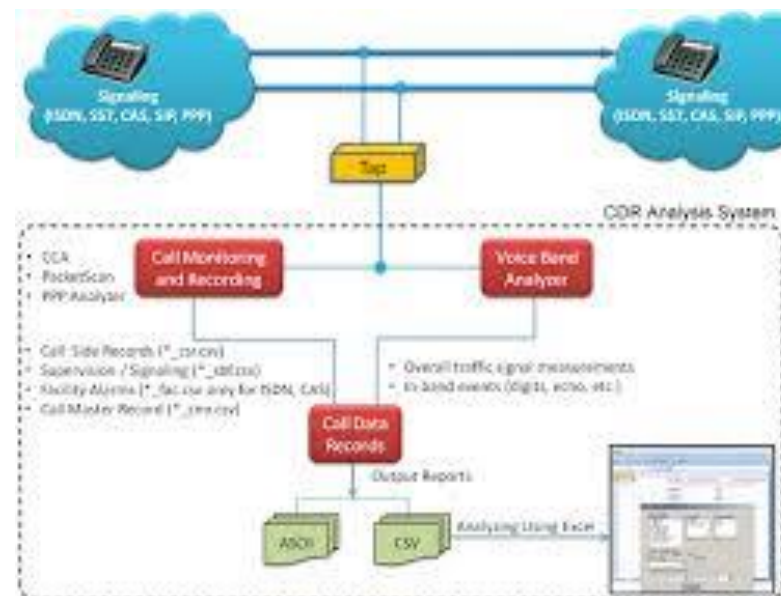
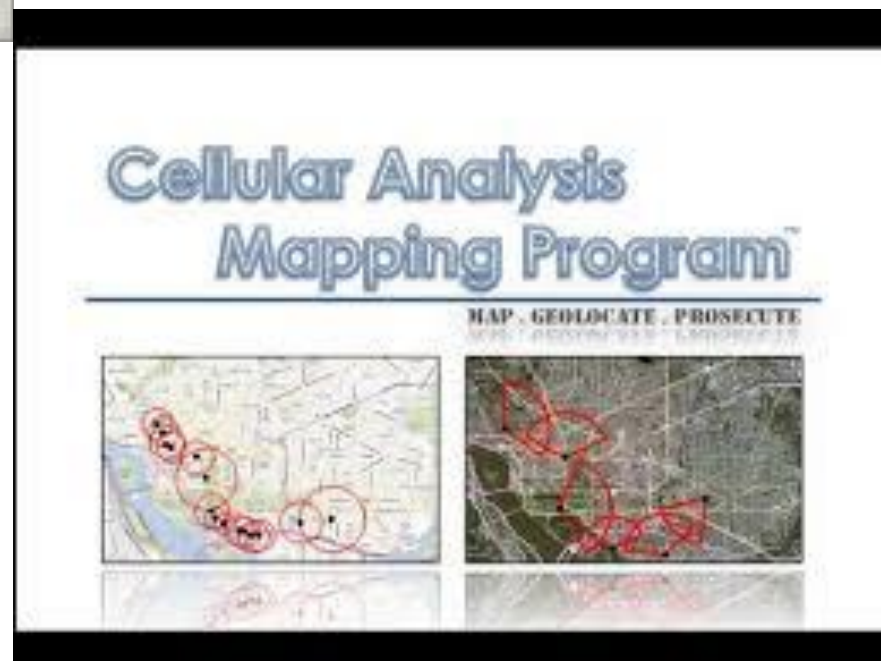
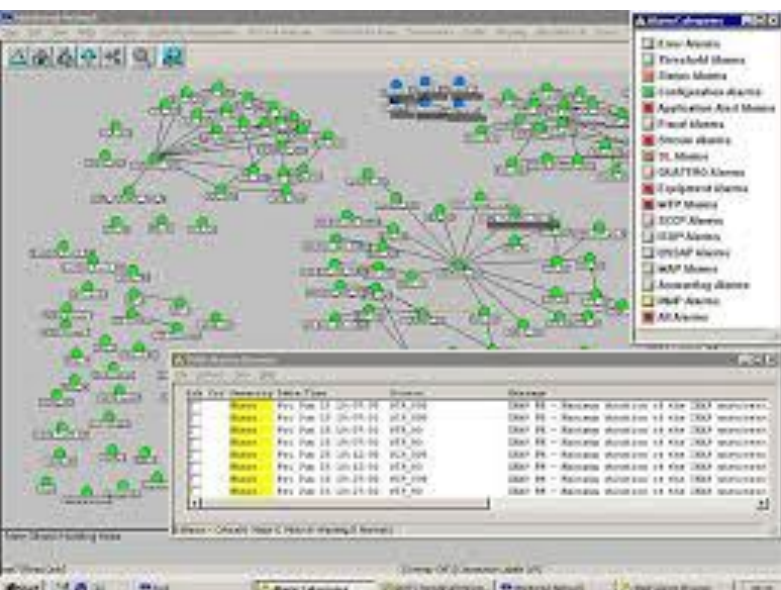


Table 1: CDR Data Sample

ProfileID	Call ID	Dir	Calling	Called	Start	Terminet	Duration	Stat Code
ATTGAP01	111160111500-1	New York(02-0)	3016241111	3016241111	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:34	Normal
ATTGAP01	111160111500-2	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:41	Normal
ATTGAP01	111160111500-3	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:42	Normal
ATTGAP01	111160111500-4	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:43	Normal
ATTGAP01	111160111500-5	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:44	Normal
ATTGAP01	111160111500-6	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:45	Normal
ATTGAP01	111160111500-7	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:46	Normal
ATTGAP01	111160111500-8	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:47	Normal
ATTGAP01	111160111500-9	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:48	Normal
ATTGAP01	111160111500-10	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:49	Normal
ATTGAP01	111160111500-11	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:50	Normal
ATTGAP01	111160111500-12	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:51	Normal
ATTGAP01	111160111500-13	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:52	Normal
ATTGAP01	111160111500-14	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:53	Normal
ATTGAP01	111160111500-15	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:54	Normal
ATTGAP01	111160111500-16	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:55	Normal
ATTGAP01	111160111500-17	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:56	Normal
ATTGAP01	111160111500-18	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:57	Normal
ATTGAP01	111160111500-19	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:58	Normal
ATTGAP01	111160111500-20	New York(02-0)	3016241200	3016241200	11/16/2011 11:15:24	11/16/2011 11:16:20	00:31:59	Normal



# Point of Sales/Payment System





# ATM



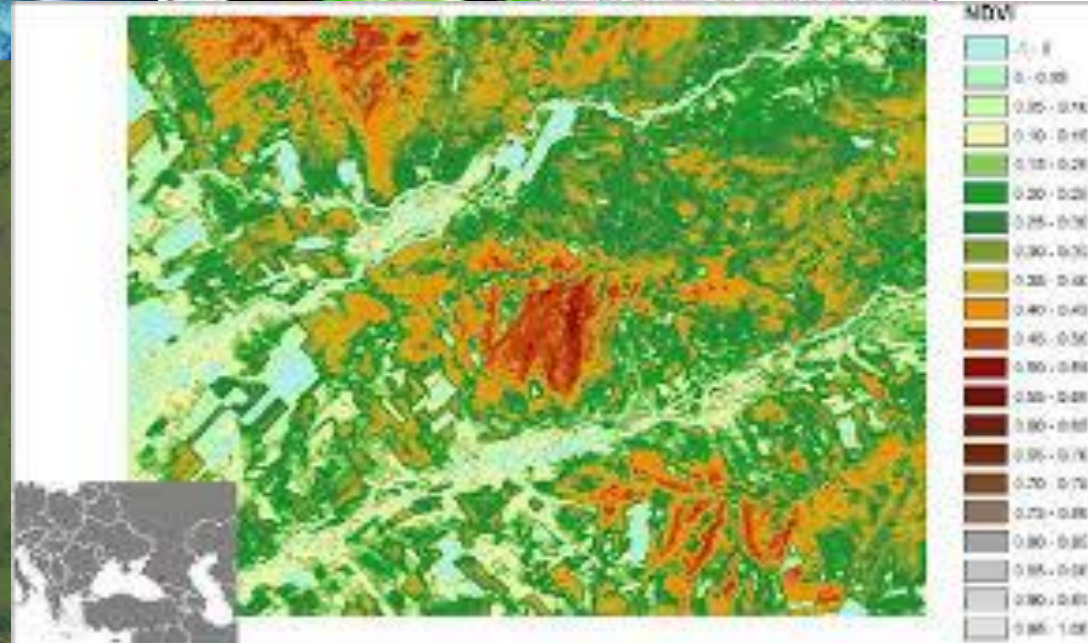
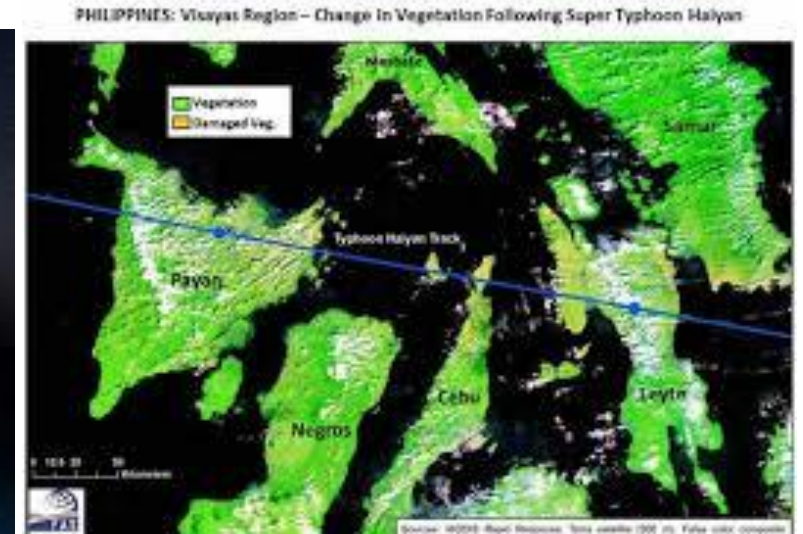


# Social Media



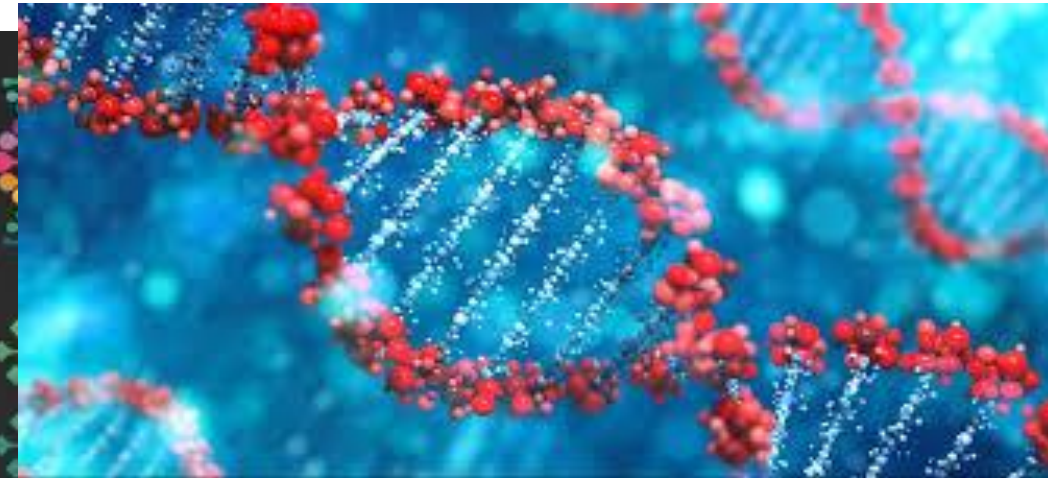


# Satellite Images



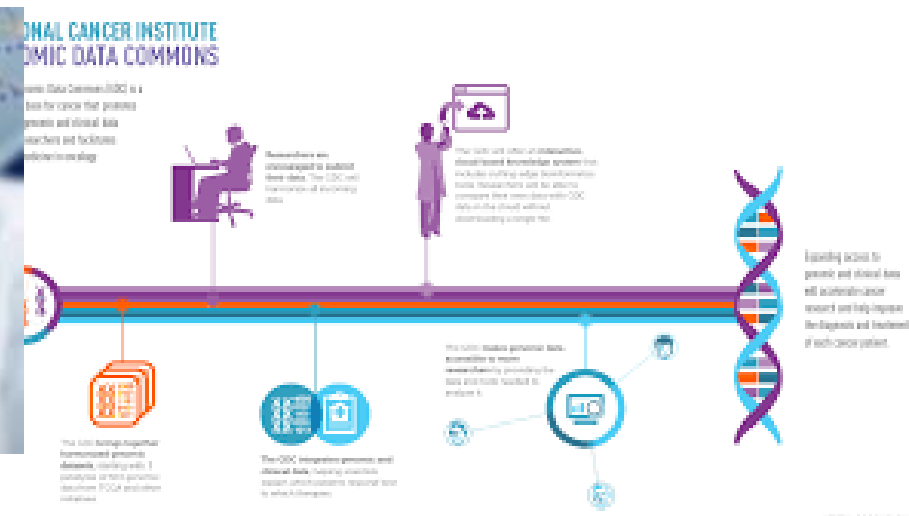


# DNA Sequences



170 180 190

ATCTCTTGGCTCCAGCATCGATGAAGAACGCG  
TCATTTAGAGGAAGTAAAAGTCGTAACAAGC  
GAACTGTCAAAACTTTTAACAACGGATCTCT  
TGTTGCTTCGGCGGGCGCCCGCAAGGGTGCC  
GGCCTGCCGTGGCAGATCCCCAACGCCGGG  
TCTCTTGGCTCCAGCATCGATGAAGAACGCG  
CAGCATCGATGAAGAACGCGAGCGAAACGCG  
CGATACCTTCTGAGTGTTCTTAGCGAACTGT  
CGGATCTCTTGGCTCCAGCATCGATGAAGA  
ACAACGGATCTCTTGGCTCCAGCATCGATGA  
CGGATCTCTTGGCTCCAGCATCGATGAAGA  
GATGAAGAACGCGAGCGAAACGCGATATGTAAT



# What is Big Data?

## Volume

The amount  
of data

## Variety

The types  
of data

## The 4 V's of Big Data

## Velocity

The frequency of  
data

## Veracity

The quality  
of data



# Volume: scale of data

- 90% of today's data has been created in just the last 2 years
- Every day we create 2.5 quintillion bytes of data or enough to fill 10 million Blu-ray discs
- 40 zettabytes (40 trillion gigabytes) of data will be created by 2020, an increase of 300 times from 2005, and the equivalent of 5,200 gigabytes of data for every man, woman and child on Earth
- Most companies in the US have over 100 terabytes (100,000 gigabytes) of data stored

# Variety: different forms of data





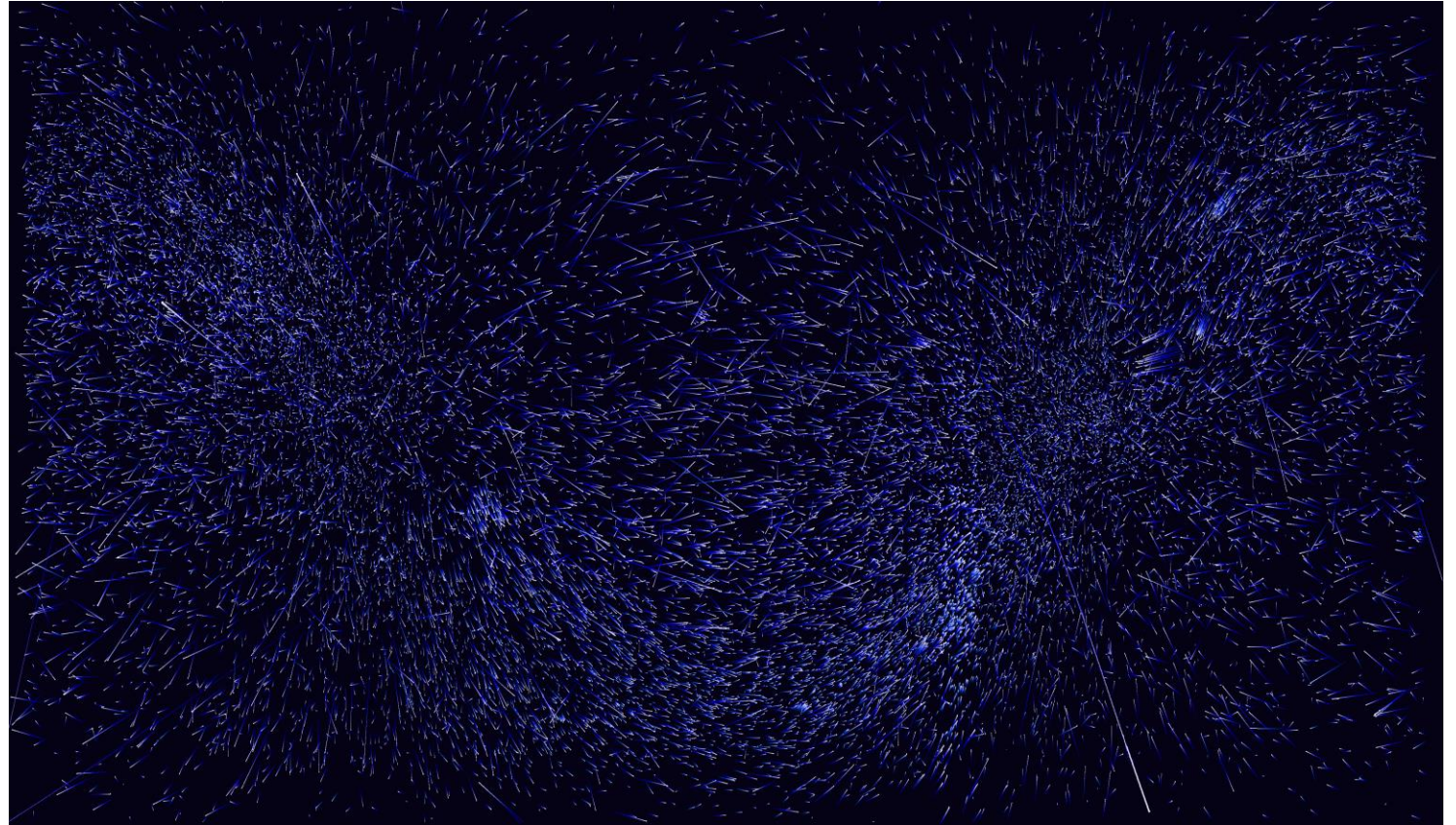
# Velocity: analysis of streaming data





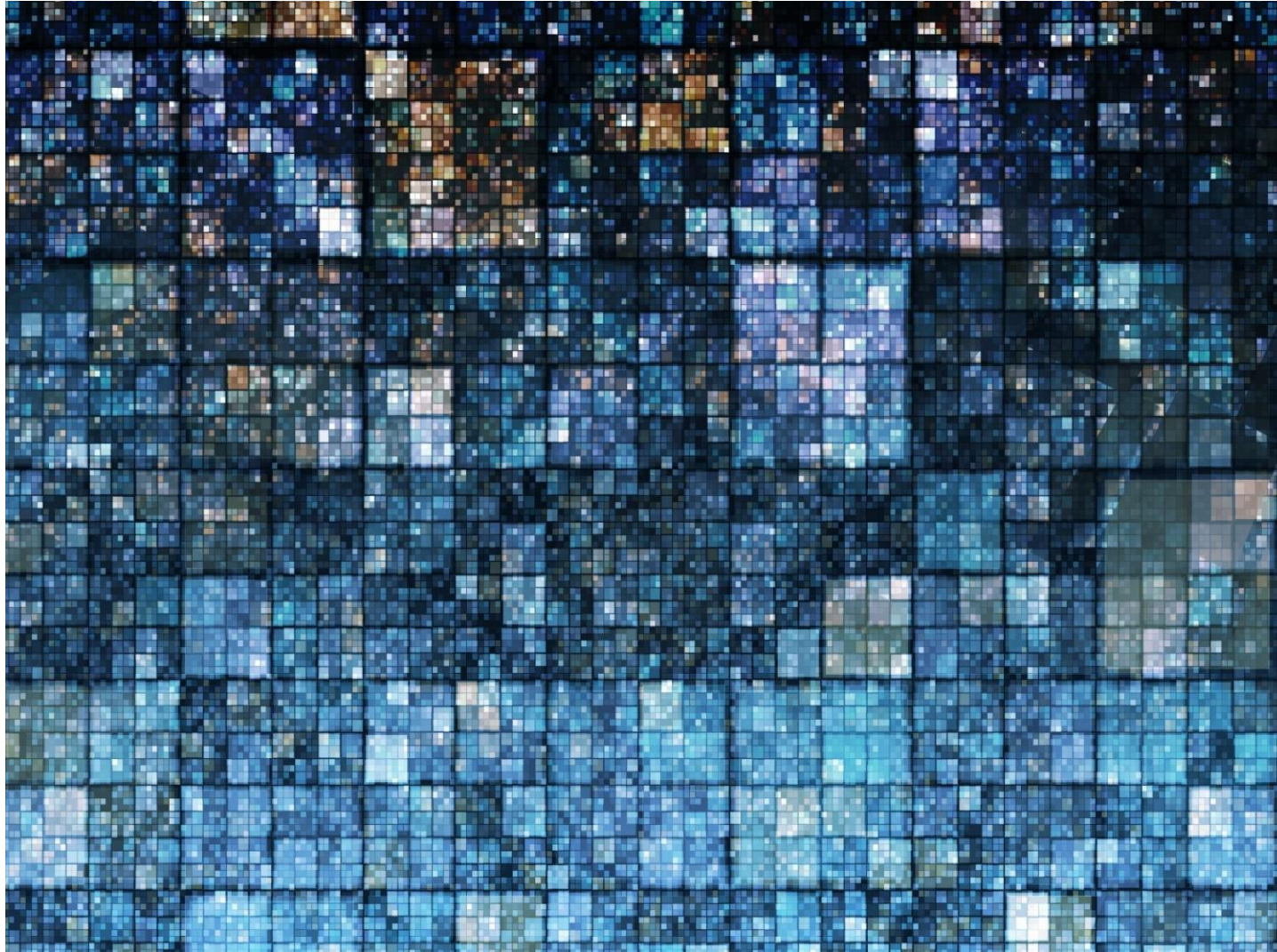
# Veracity: trustworthiness of data

- Origin
- Authenticity
- Trustworthiness
- Completeness
- Integrity





# What is in Big Data?

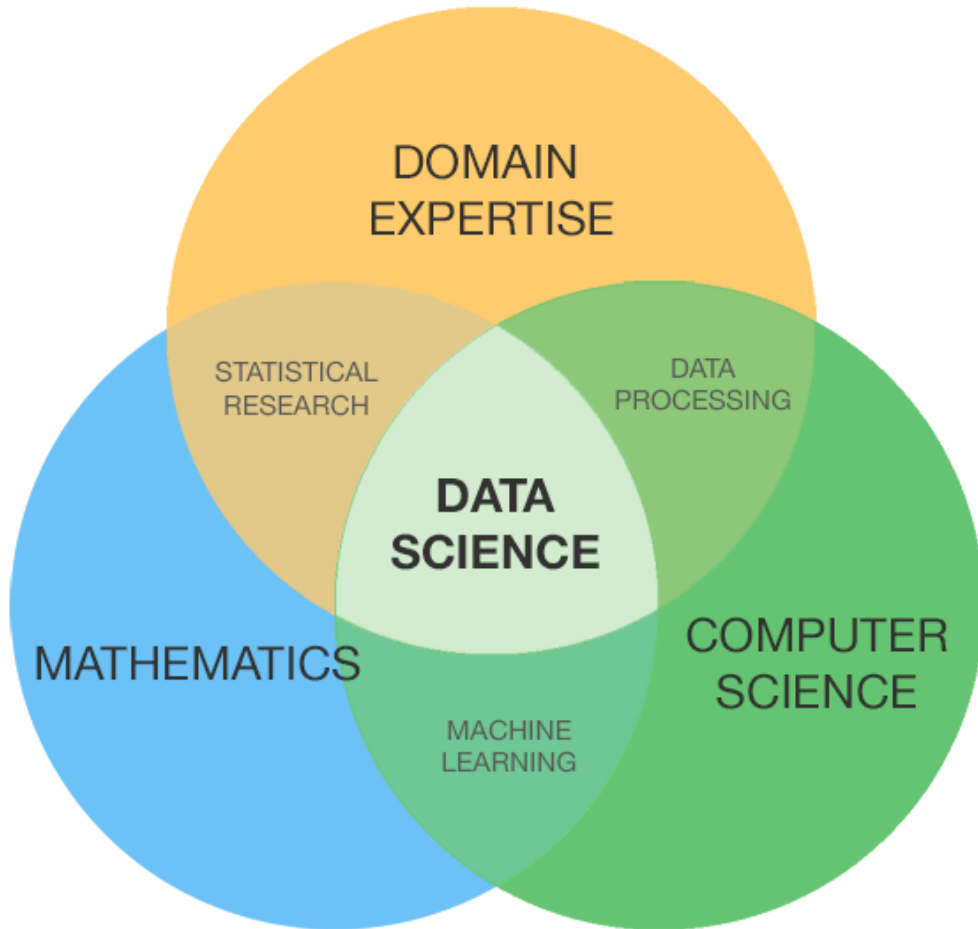


**What is there?**

**Is it worth it?**

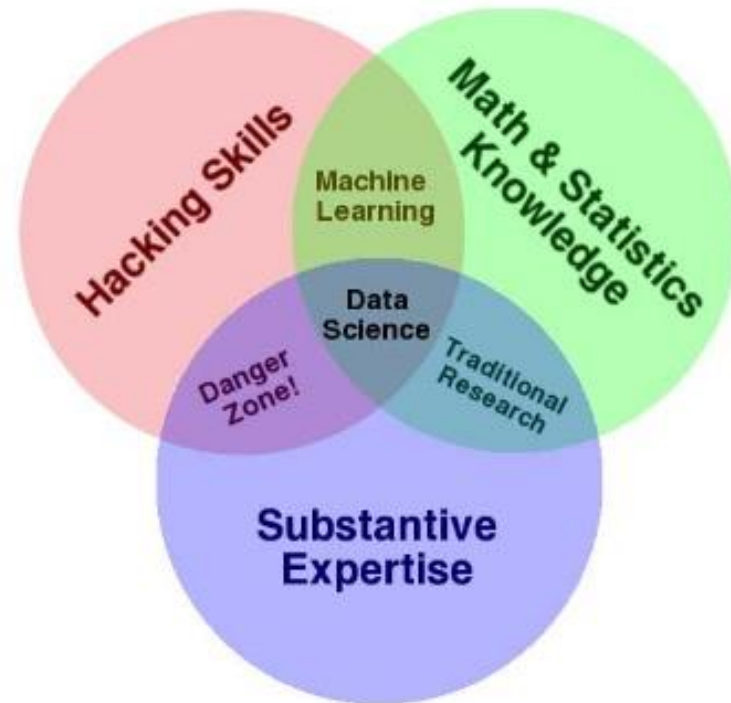
**How to go about it?**

# Data Science



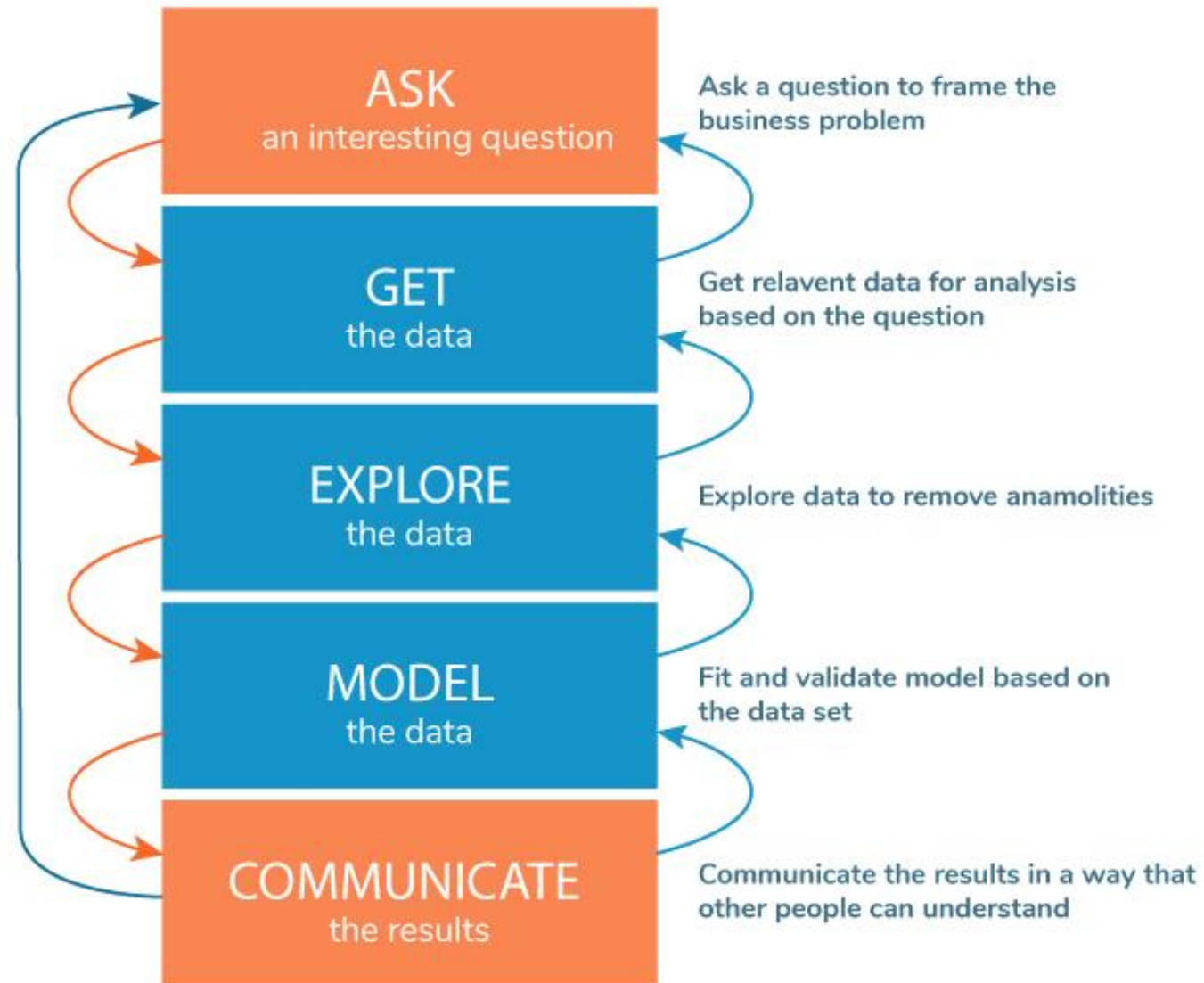
Source: Palmer, Shelly. *Data Science for the C-Suite*.  
New York: Digital Living Press, 2015. Print.

Drew Conway's Venn Diagram

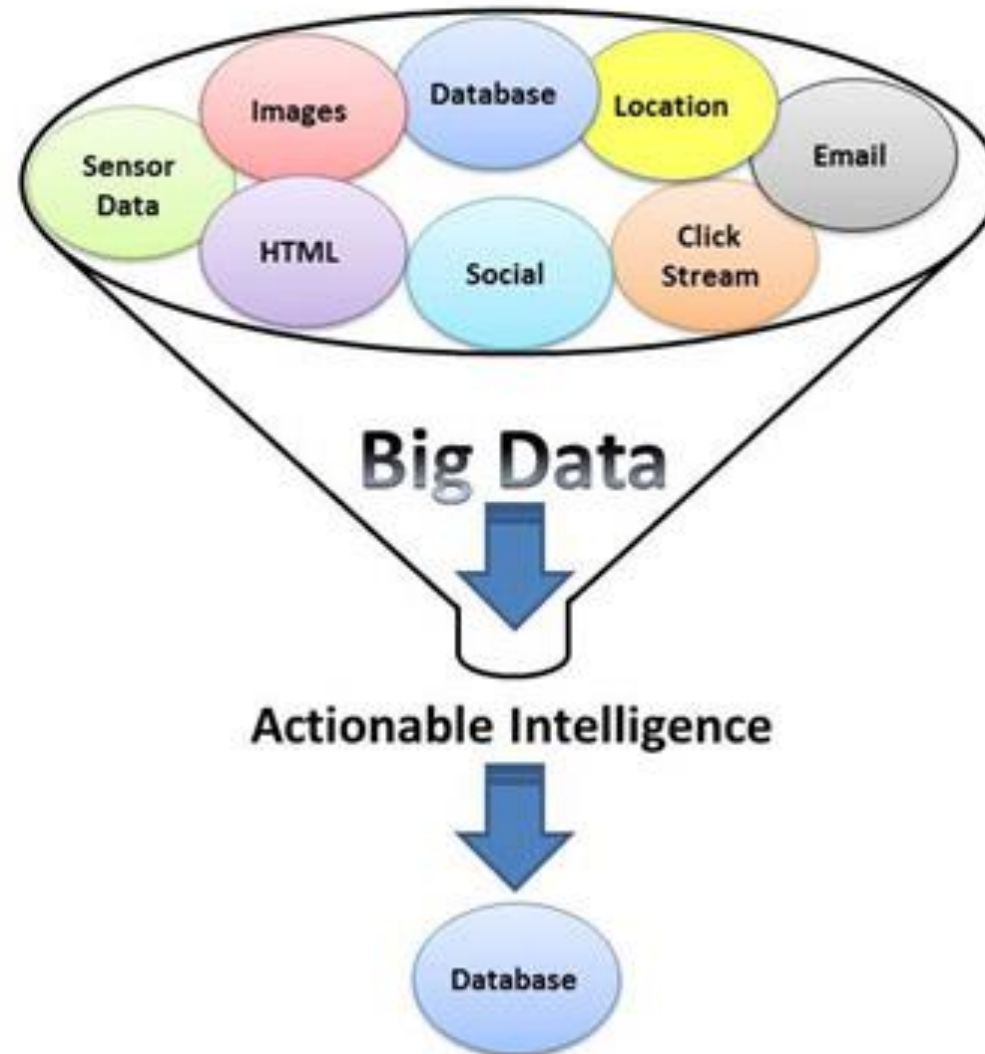




# THE DATA SCIENCE PROCESS

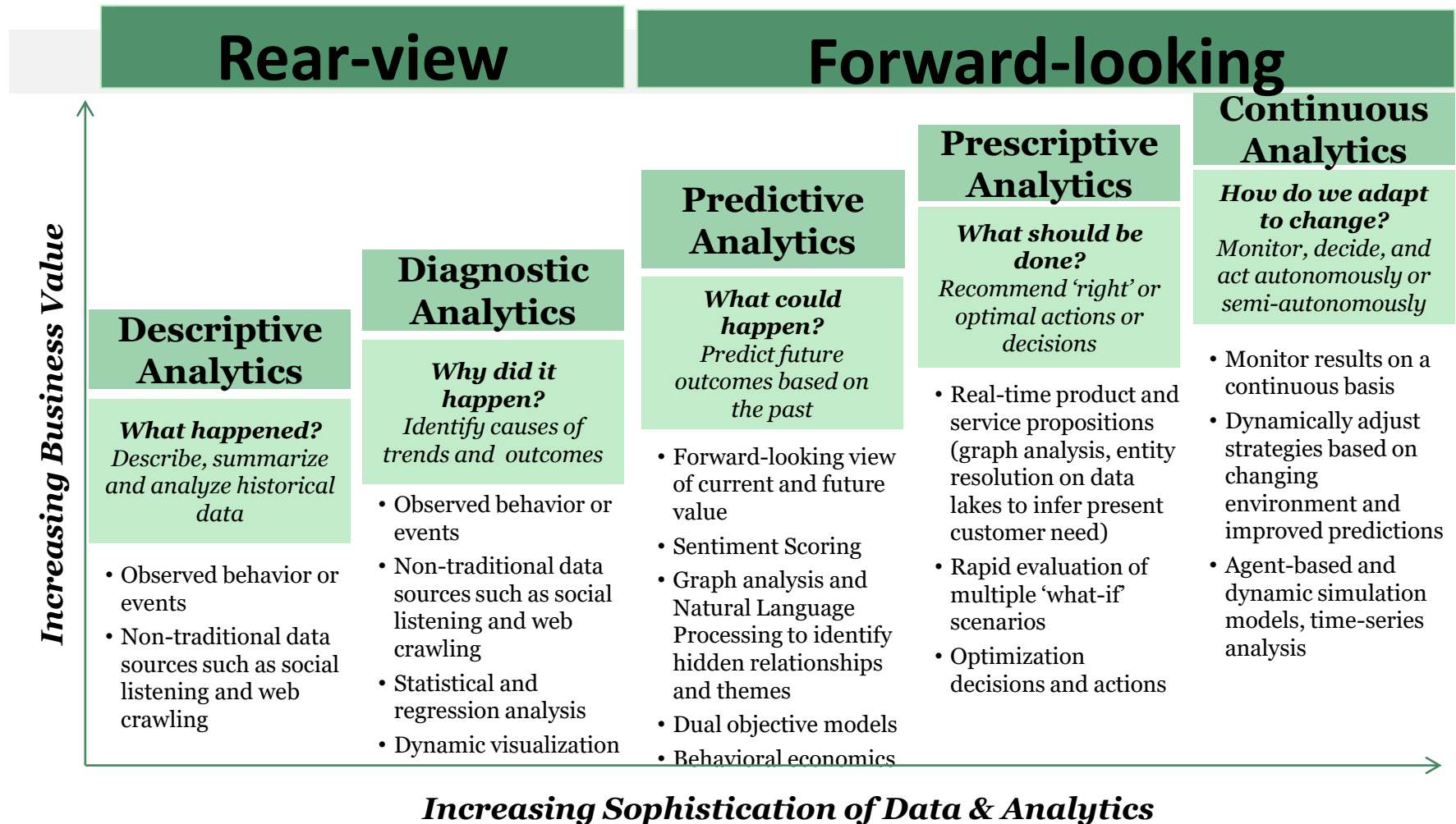


# Data Mining





# Analytics

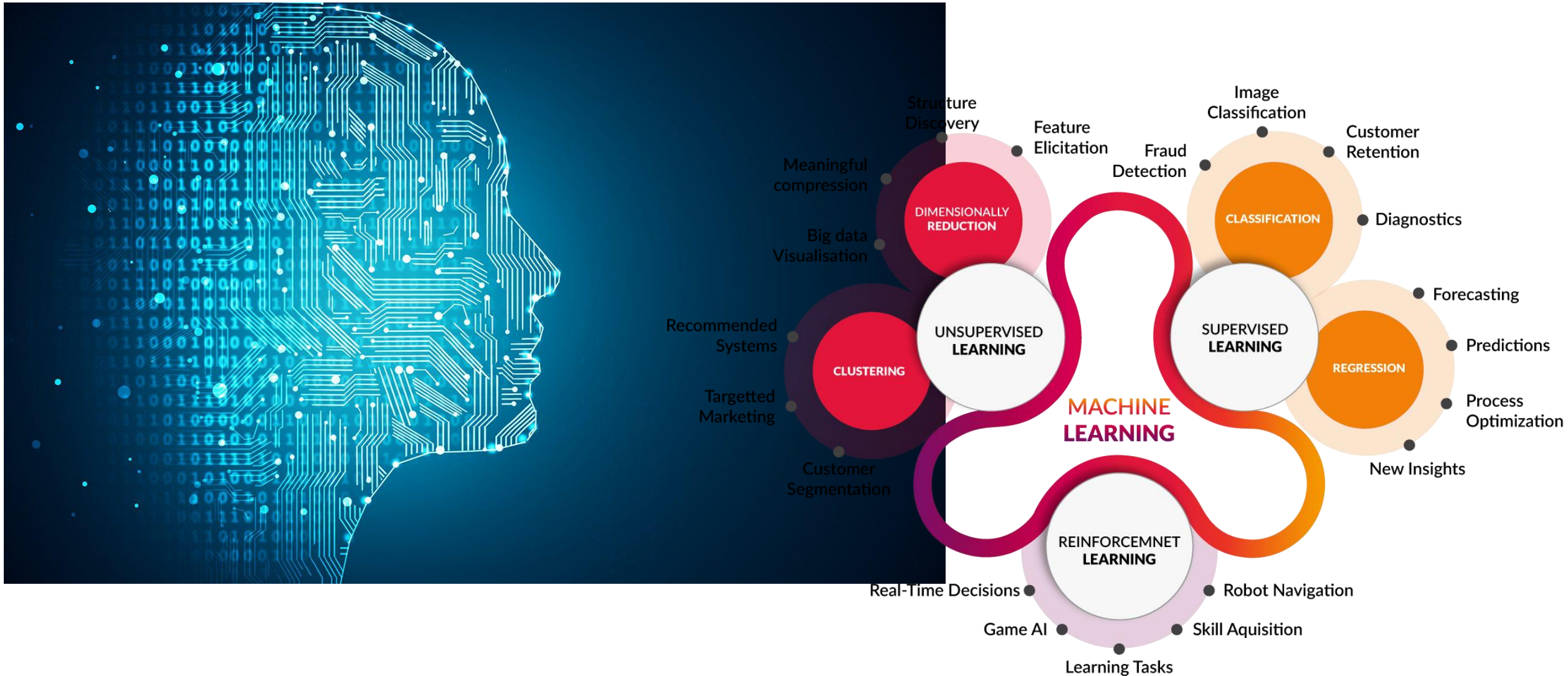


# Artificial Intelligence

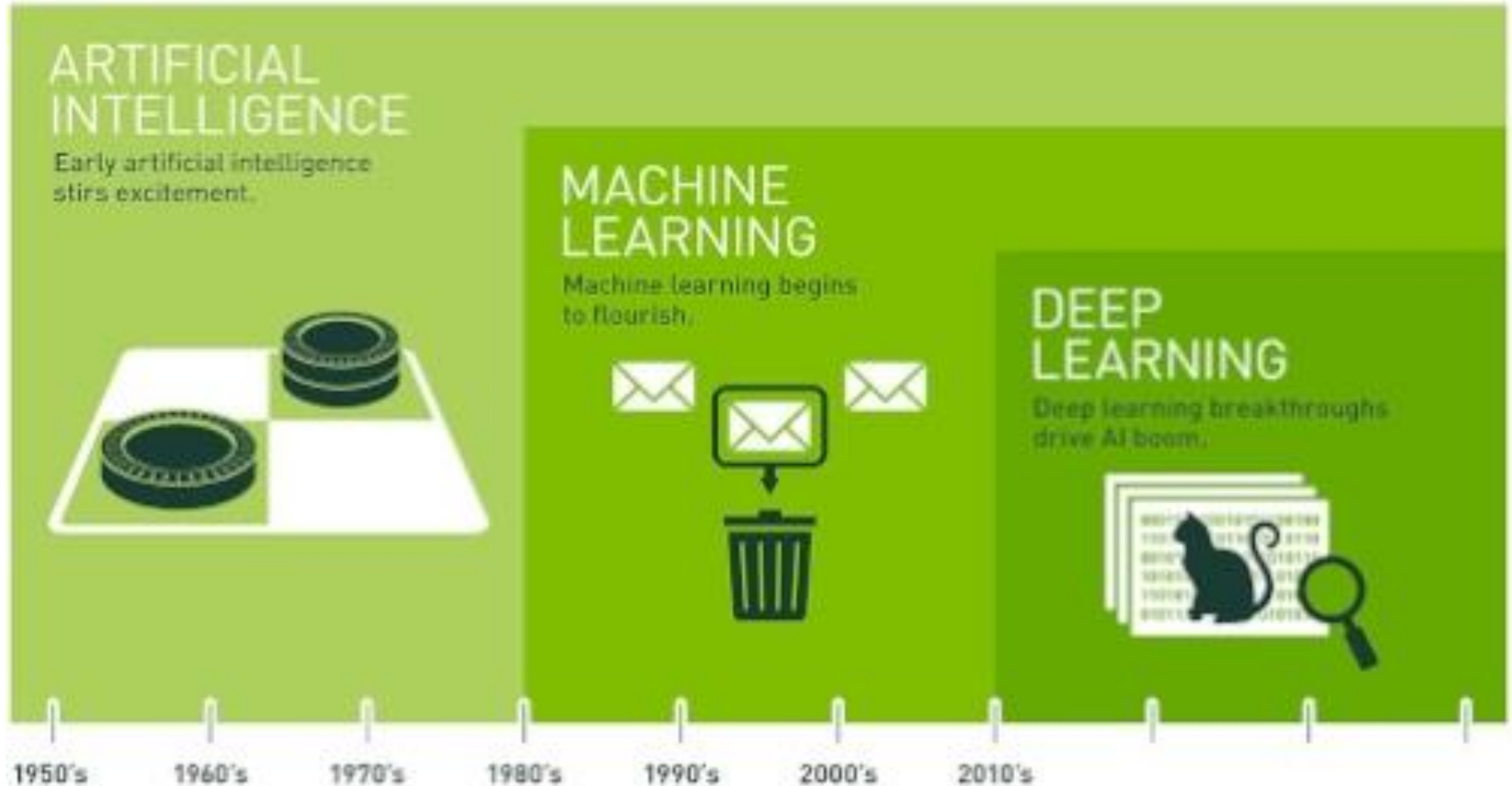




# Machine Learning



# Insights





# Insights



STRATEGY &amp; INNOVATION

Provide insight to guide business strategy, self-innovation opportunities and innovation priorities

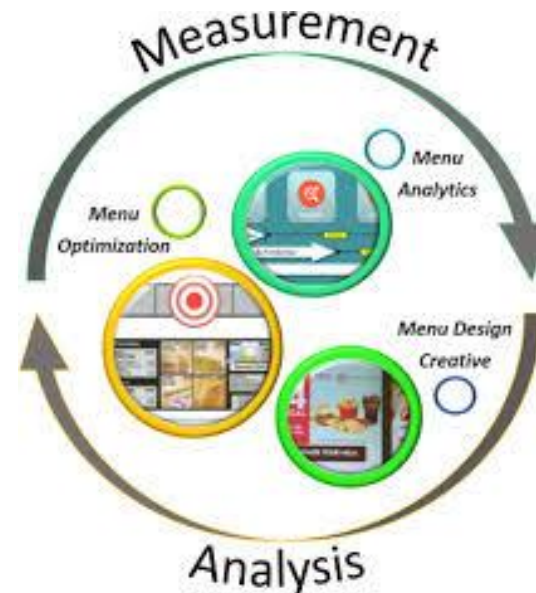


## MARKETING & GO-TO-MARKET

Identify the brand hurdles that must be addressed for each segment and how can we best reach the target consume

## PORTFOLIO PLANNING & PRODUCT DEVELOPMENT

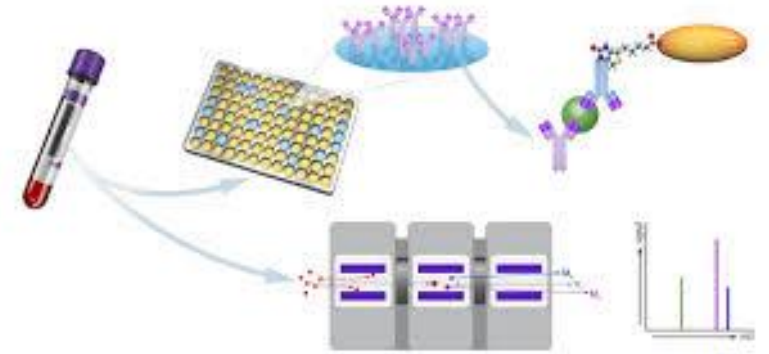
### Guide portfolio structure



# Biomarkers



## DISEASE RELATED BIOMARKERS



Development  
of new diagnostic  
biomarkers

Co-development  
of companion  
diagnostics

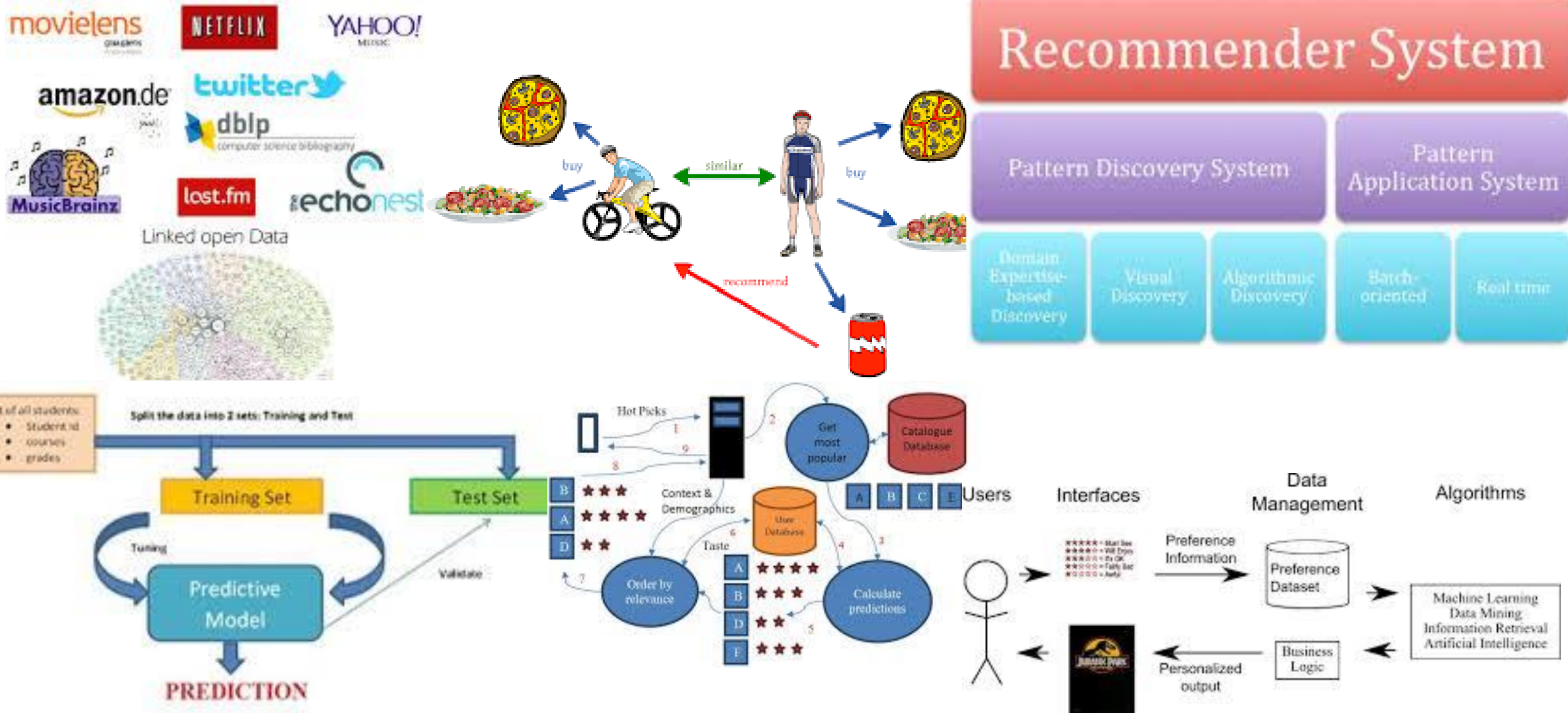
Tools to improve  
drug development

Identification and validation  
of new drug targets

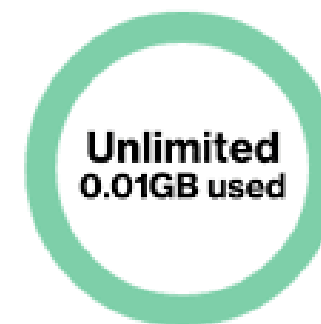




# Recommender System



# Loyalty Plan



LOYALTY PLAN  
\$15 / 30 MIN 1YR

[View plan details](#)

All usage is estimated.

[Manage plan >](#)

30 days to go

## How to Keep Your Customers Coming Back



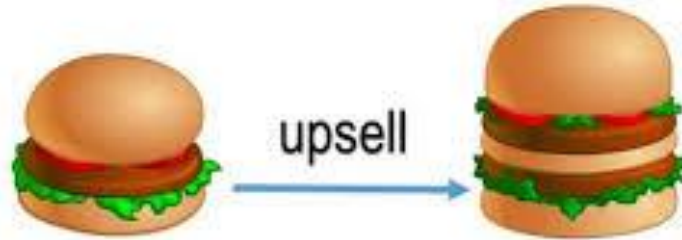
My Trial Pack				
Product	Package	Quantity	DP (Rs.)	MRP
1. Tooth Paste	100 gm	1	90/-	90/-
2. Tooth Brush	1 pc set	1	110/-	110/-
3. Bath Soap	1 pc set	2	120/-	120/-
4. Mouthwash	200ml	1	180/-	180/-
5. Conditioner	100ml	1	140/-	140/-
6. Shampoos 200 ml	200 ml	1	80/-	80/-
7. Tea	250 gm	1	150/-	150/-
8. Blue Bar Soap	100 gm	1	180/-	180/-
9. Star Lotion	100 ml	2	80/-	160/-
10. Beyond Blue	100 ml	2	110/-	220/-
11. One Day Cream	400 ml	2	120/-	240/-
12. One Day Scrub	400 ml	2	120/-	240/-
13. Nivea deodorant powder	100 gm	1	150/-	150/-
14. Nivea Pro-Gel deodorant	200ml	1	140/-	140/-
15. Nivea Sunscreen Lotion	300 ml	1	100/-	100/-
16. Nivea Sunscreen Cream	100 ml	1	100/-	100/-
17. Nivea Sunscreen Spray	100 ml	1	100/-	100/-
18. Nivea Sunscreen Stick	100 ml	1	100/-	100/-
19. Nivea Sunscreen Gel	100 ml	1	100/-	100/-
20. Nivea Sunscreen Cream	100 ml	1	100/-	100/-
Total Amount			5060/-	5060/-

Retail Profit = Rs. 887 [MRP - DP difference]  
 Offer Product = Rs. 400 on MRP (upto)  
 Cash Back offer = Rs. 105 min. on BV (points on product)  
 Total Benefits = Rs. 1,392 Minus 1% on MRP + 1% on 1% on MRP





# Upselling/Cross Selling



# Customer Relationship Management





# Possible Sources of Human Resources

- Statistics/Mathematics
- Computer Science/Information Technology
- Science and Engineering
- Business and Economics
- Social Science, Arts and Humanities
- Agriculture, Medicine, Fisheries

# Big Data and Analytics in the Philippines

- 2000-2006: FMCG – Data from different sources, varying frequencies
- 2003 – 2006: TELCO
- 2012-Present: Energy
- 2014-2015: IT Company
- 2007-2010: QSR
- UA&P: BSBA (Management, Sp. Bus Analytics)
- UA&P: Master in Applied Business Analytics
- AIM: MS Data Science
- UPSS: Prof. Master in Data Science (Analytics)



# SDG



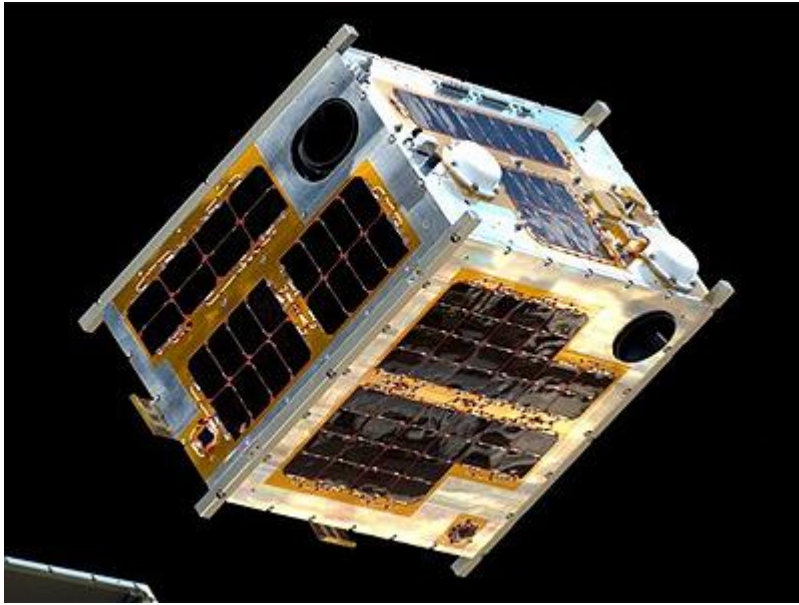


# Official Statistics





# Diwata 1



# Diwata – Data Receiver



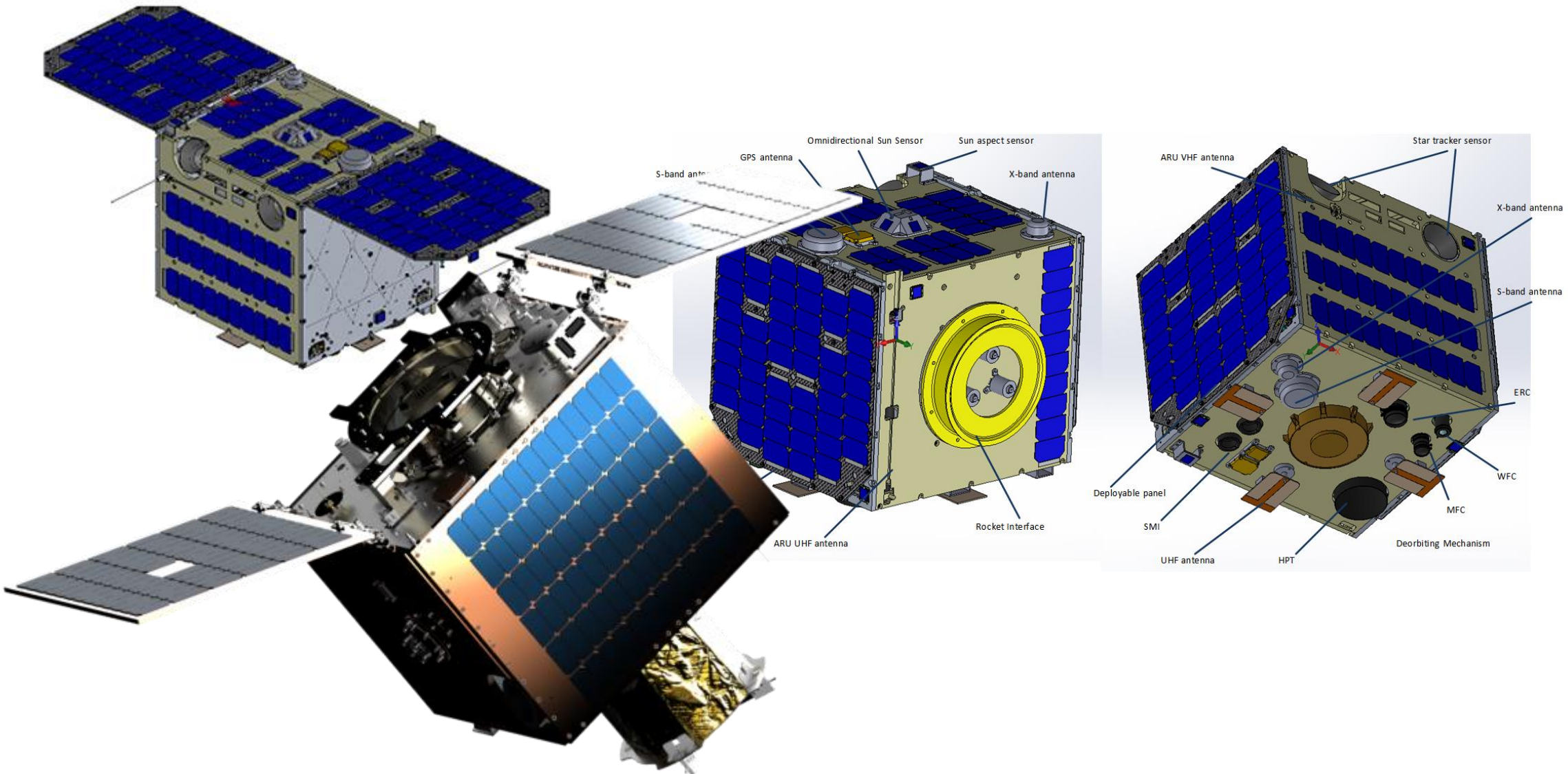
2.4m at Tohoku University, Sendai



1.8m at ASTI-DOST, Diliman



# Diawata 2



# National ID System

- **Coverage**

- Unique Philippine ID Number
- Full Name
- Blood Type
- Marital Status
- Date of Birth
- Place of Birth
- Address
- Front-Facing Photo

- **Registration**

- PSA
- Local Civil Registry
- GSIS
- SSS
- PhilHealth
- Home Dev Mutual Fund
- COMELEC
- Phil Postal Corp



# The Challenge....

- Dealing with data growth
- Generating insights in a timely manner
- Recruiting and retaining talents
- Integrating disparate data sources
- Validating data
- Securing big data

**Thank you.**