

INTRODUCTION TO DATA DRIVEN DECISION MAKING



Andreas Zaras, Data Scientist

CHARACTERISTICS OF TODAY'S BUSINESS ENVIRONMENT

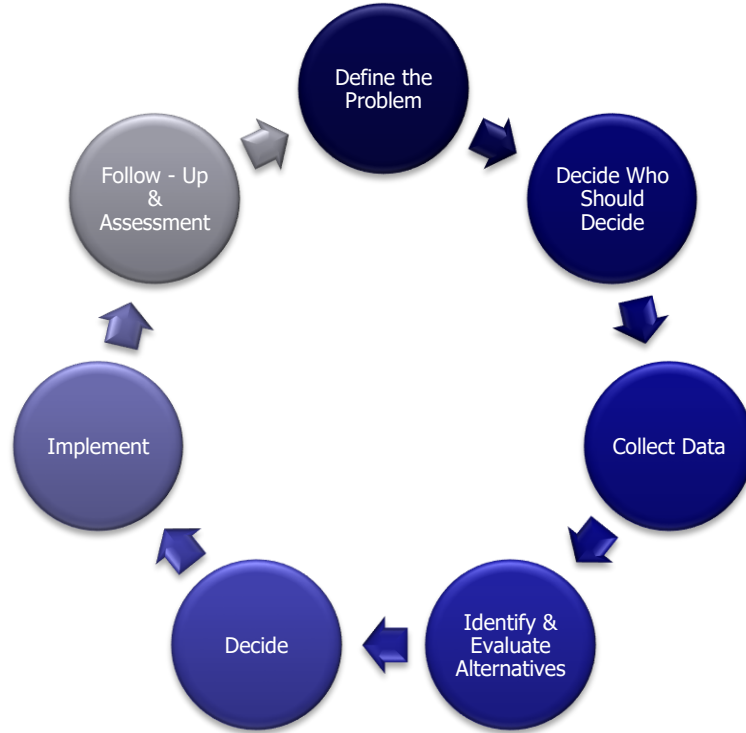
It is characterized by:



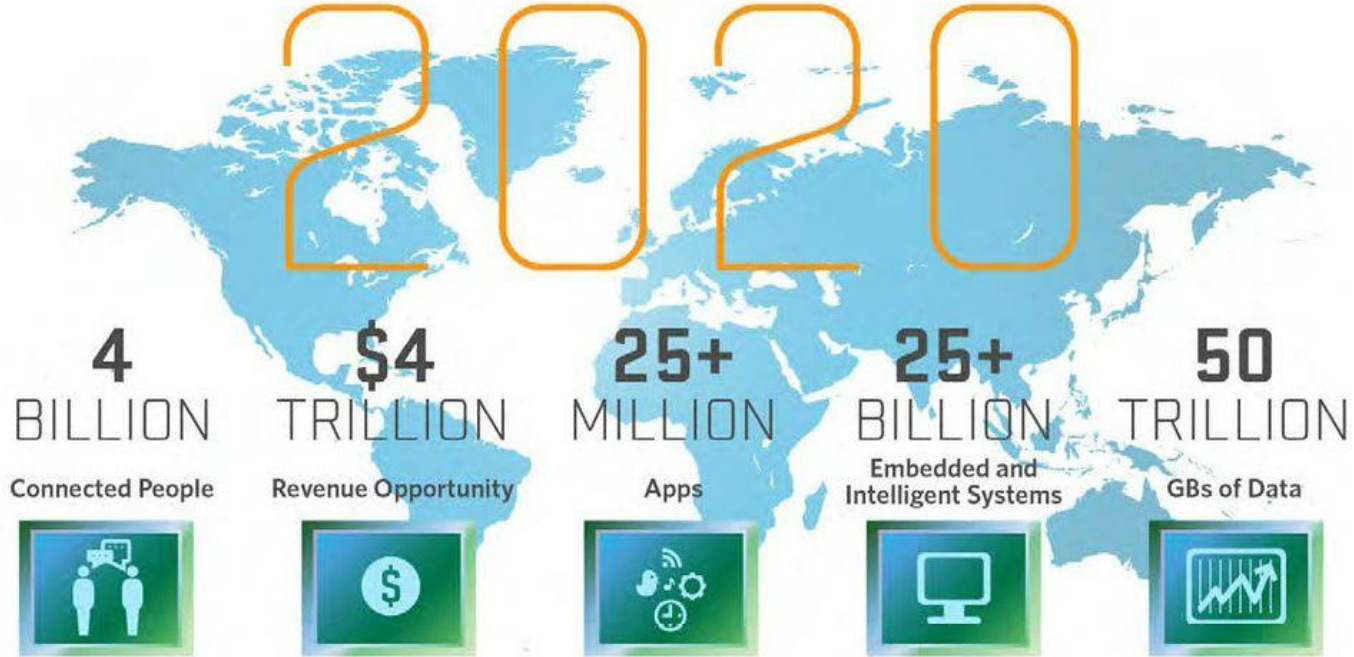
Organizations strive to survive by acquiring competitive advantage.



THE DECISION MAKING PROCESS

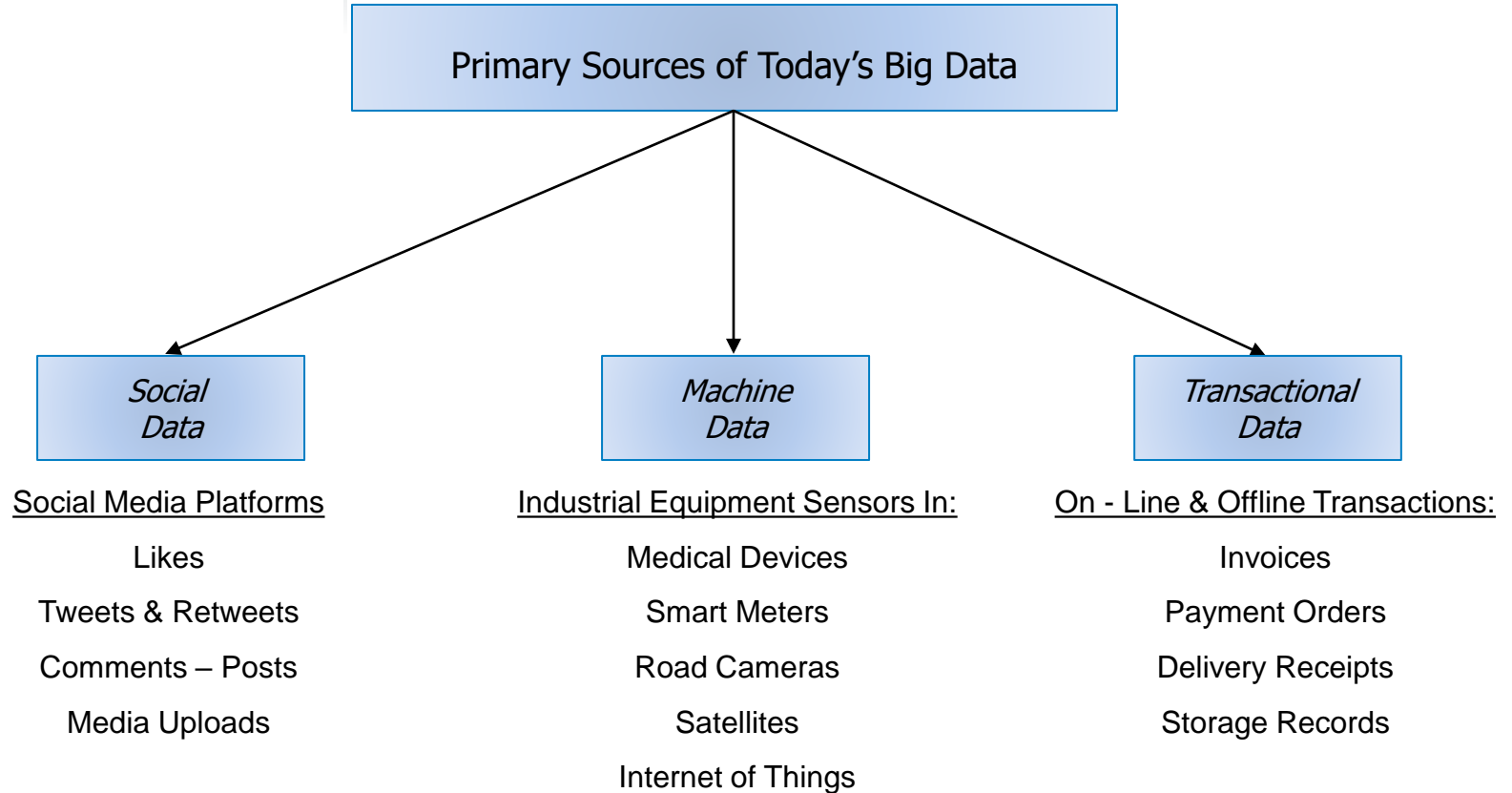


LIVING IN A DATA FLOODED WORLD! (1 / 6)



Source: Mario Morales, IDC

LIVING IN A DATA FLOODED WORLD! (2 / 6)



Some Big Data Statistics

90% of all data in the world has been created in the last two years
(Source: IBM).

Internet users generate about 2.5 quintillion bytes of data each day – equal to the total ants on the globe times 100 (Source: Data Never Sleeps).

Today it would take a person approximately 181 million years to download all the data from the internet
(Source: Physics.org).

In 2020 there will be 40 x more bytes of data than there are stars in the observable universe
(Source: Data Never Sleeps).

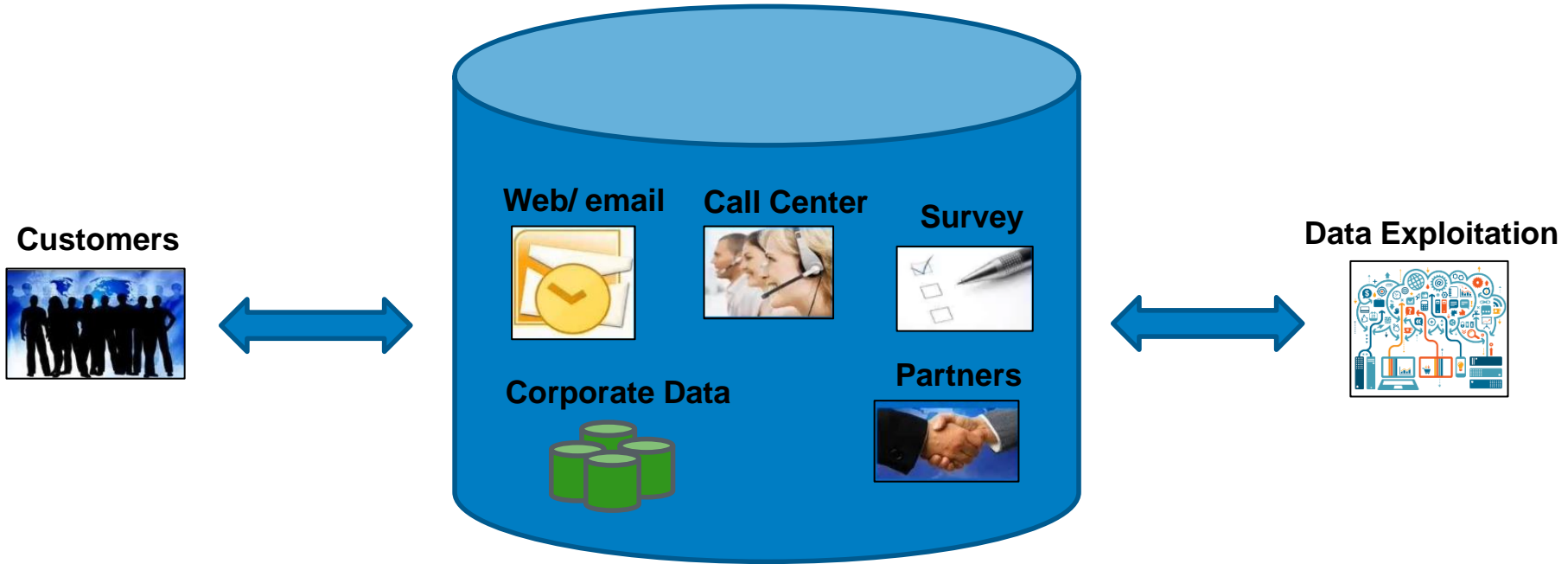
By 2020, every person will generate 1.7 megabytes in just a second.
(Source: Data Never Sleeps)

LIVING IN A DATA FLOODED WORLD! (4 / 6)



Source: Data Never Sleeps 9.0

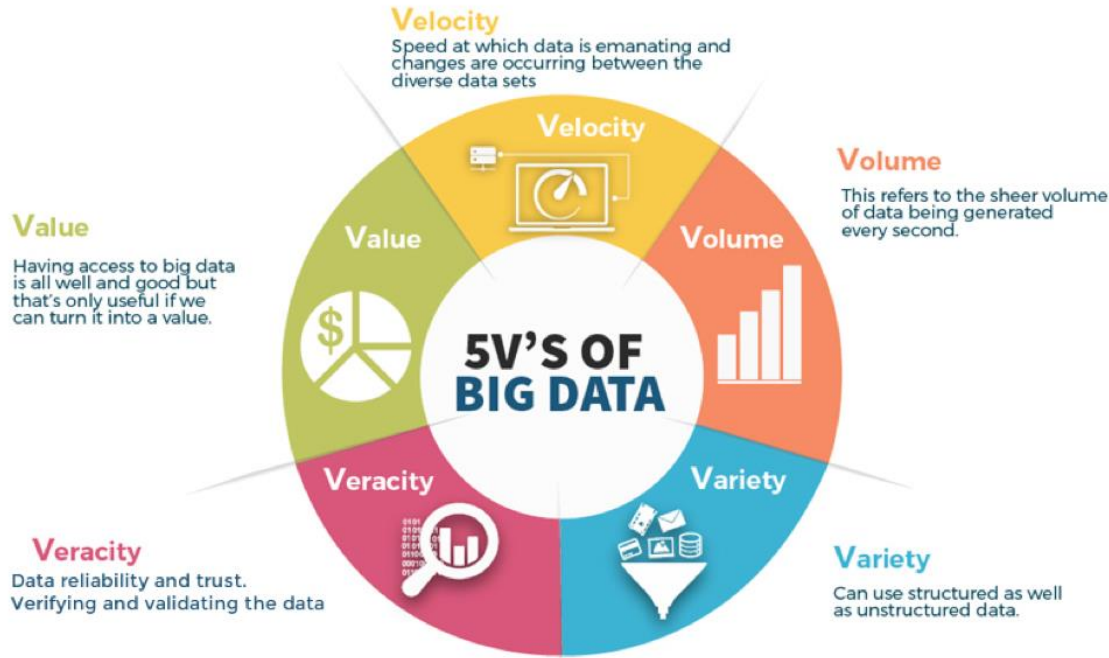
LIVING IN A DATA FLOODED WORLD! (5 / 6)



LIVING IN A DATA FLOODED WORLD! (6 / 6)

Big Data Definition

The point at which the volume, velocity, and variety of data exceed an organization's storage or computation capacity for accurate and timely decision making



A NEW PROFESSION IS BORN: THE DATA SCIENTIST!



Data
Access



Data
Processing



Business
Intelligence



Business
Analytics



Business
Knowledge

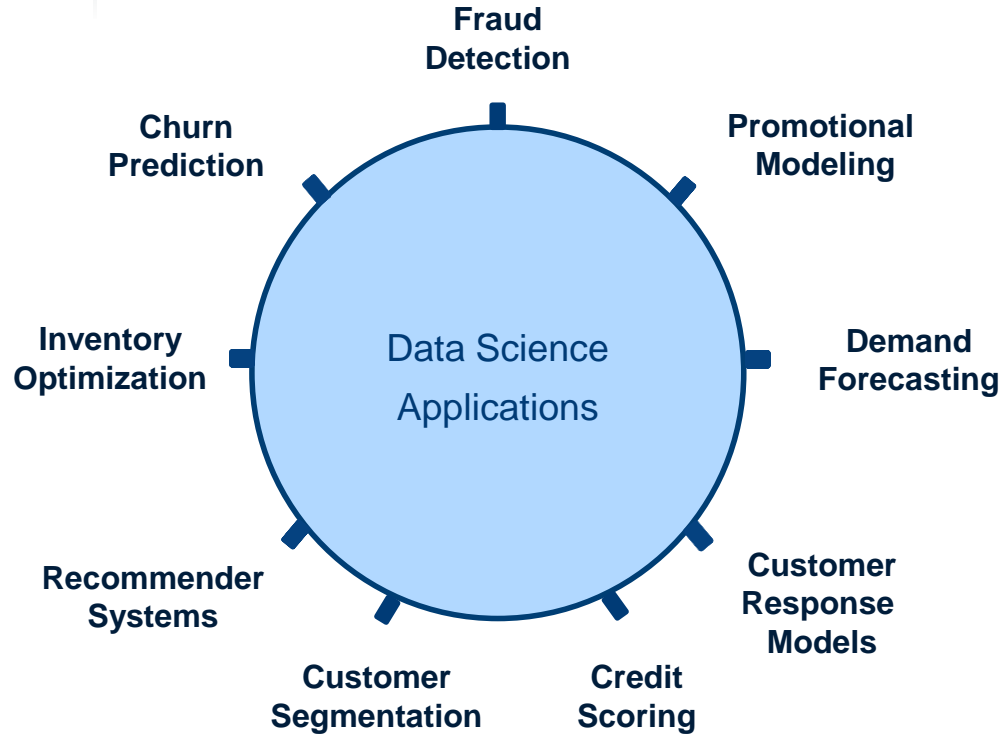


Presentation
of Results



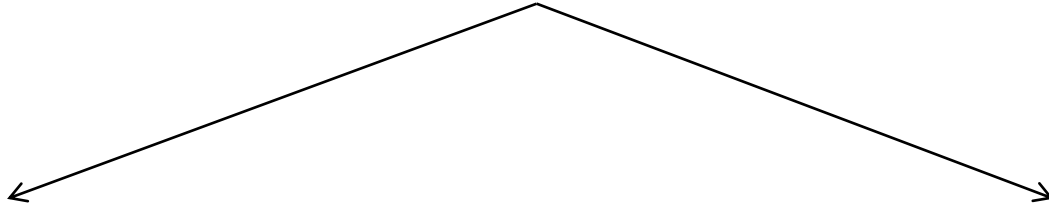
The Skills of a Data Scientist

APPLICATIONS OF DATA SCIENCE



Information Systems

(Hardware, Software, Data, People, Networks)



Data Capturing

Data Exploitation

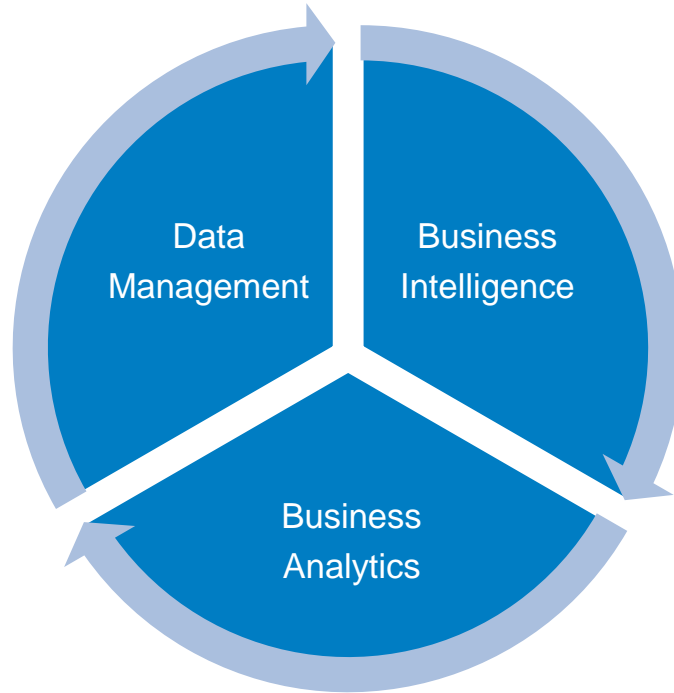
Transactional Systems

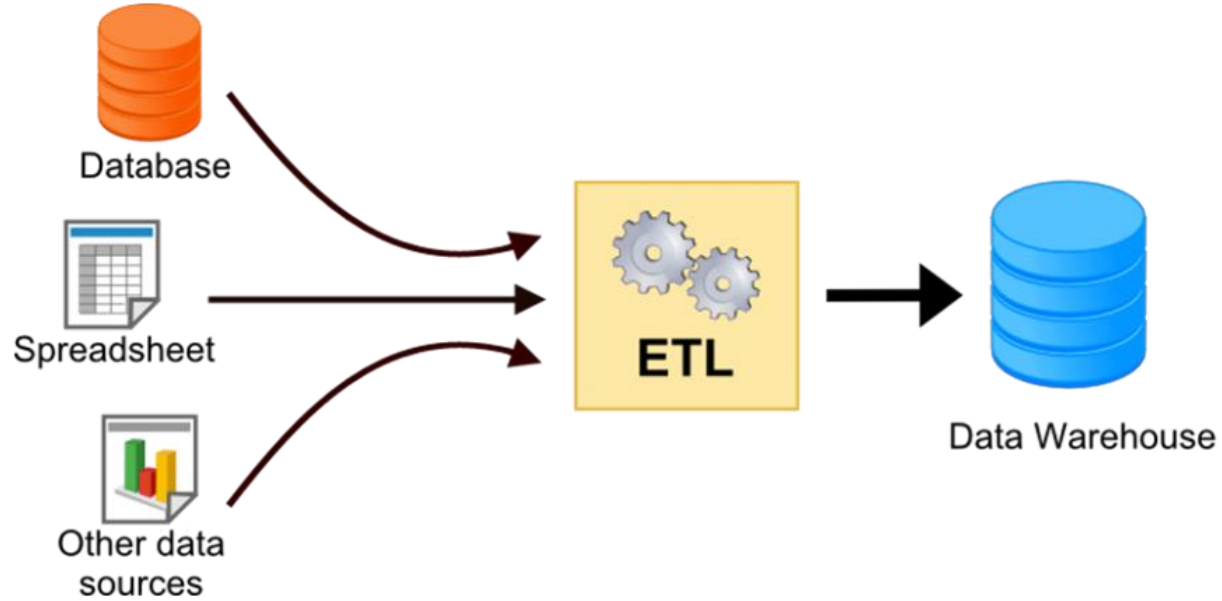
(Support Day to Day Operations)

Decision Support Systems

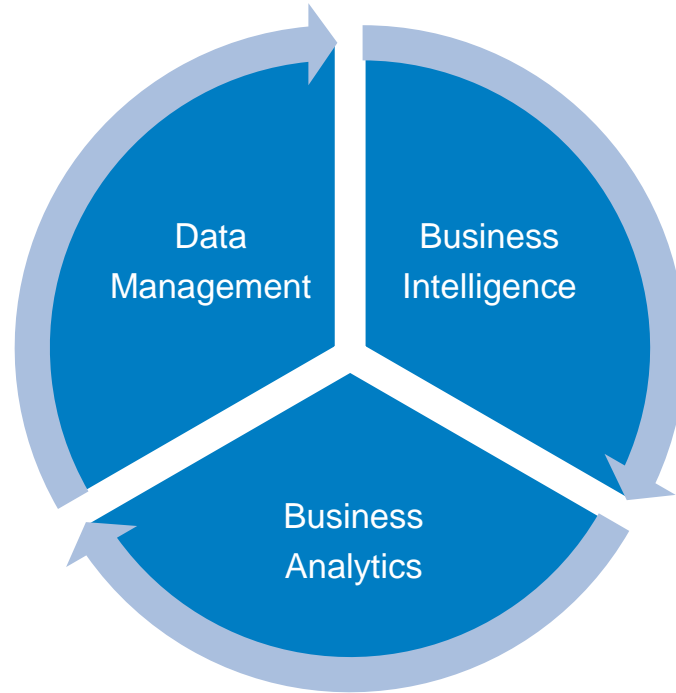
(Support Decision Making)

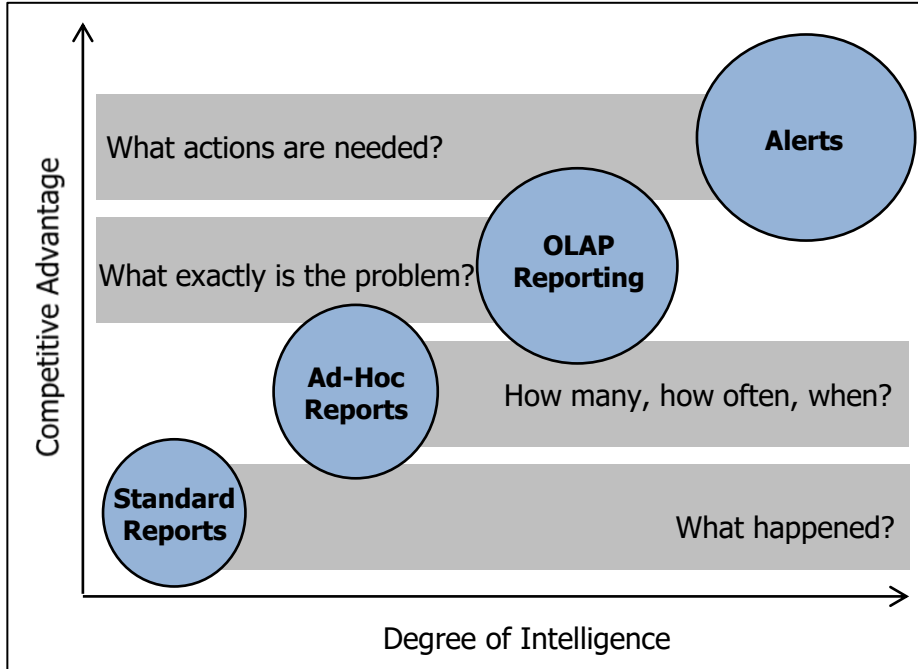
STAGES OF DATA EXPLOITATION



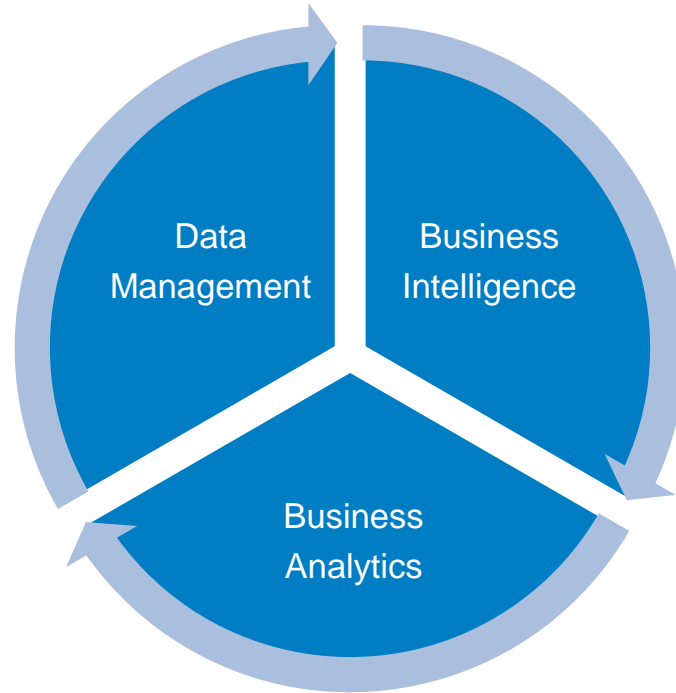


STAGES OF DATA EXPLOITATION

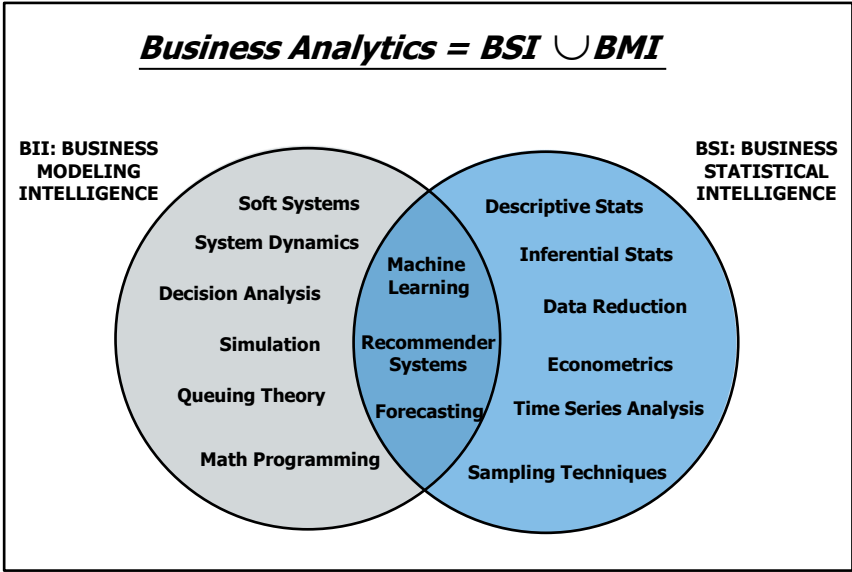


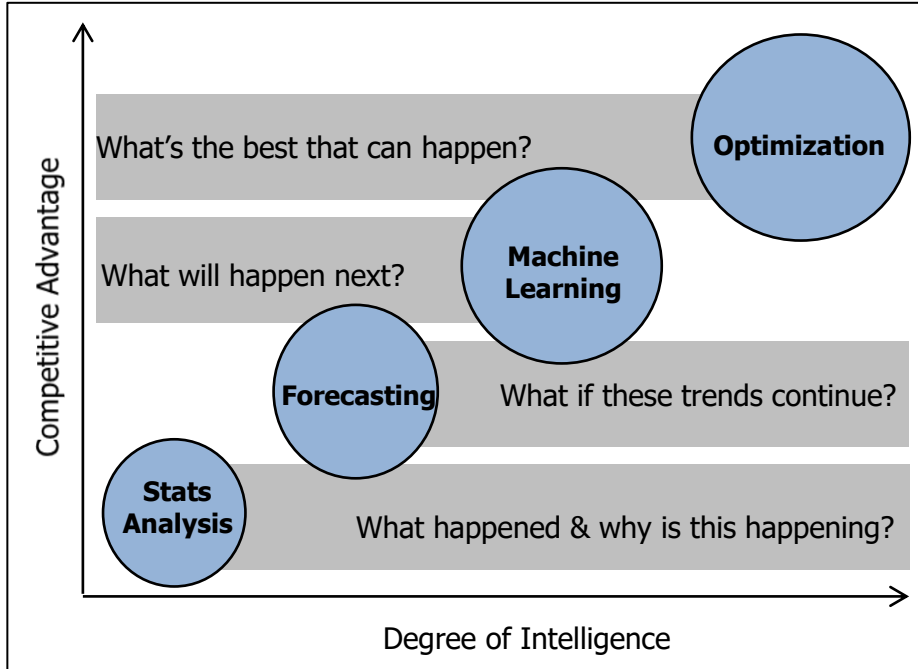


STAGES OF DATA EXPLOITATION

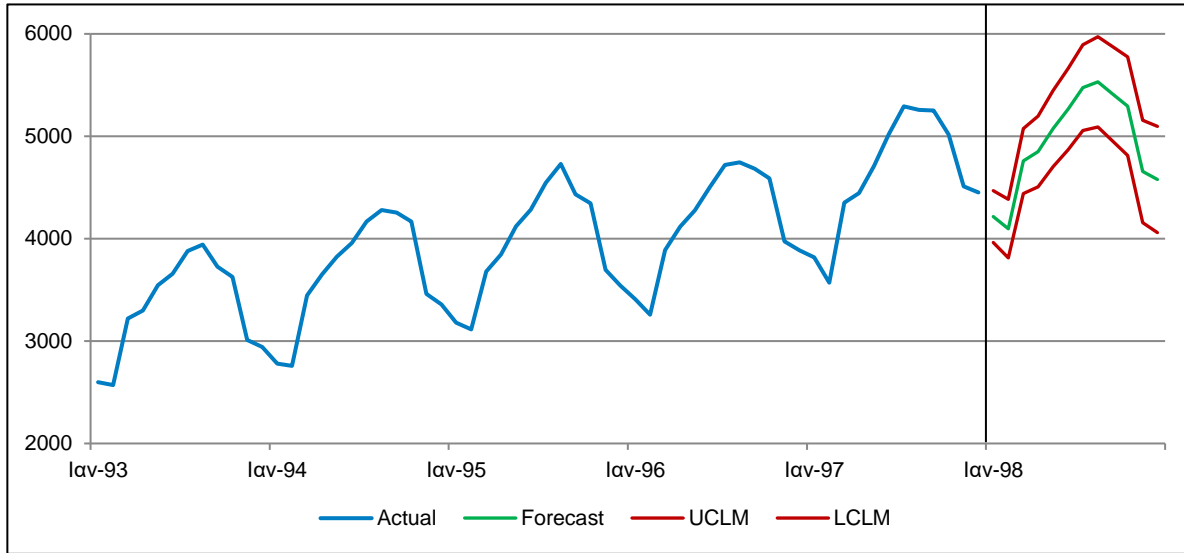


Business Analytics (BA), as defined by the International Institute of Analytics in 2010, is: "The broad use of data and quantitative analysis to support the decision making process...."





FORECASTING



Date	Demand (Units)
January 1993	2554
February 1993	2890
March 1993	3240
.....
October 1997	3390
November 1997	3212
December 1997	3019

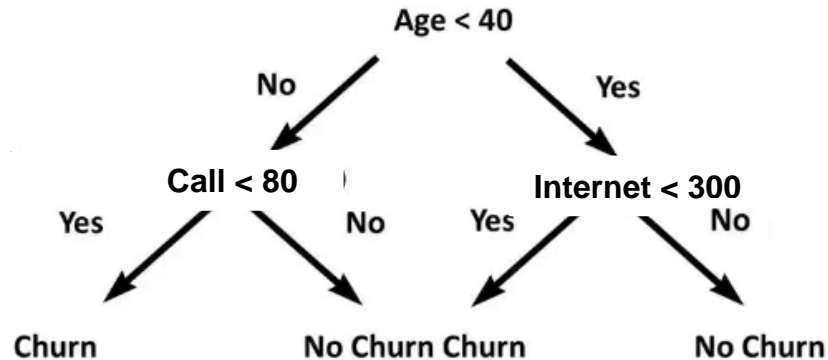


Forecast
(Produced in 31/12/1997)

January 1998	4480
February 1998	4670
.....
November 1998	4789
December 1998	4760

MACHINE LEARNING: THE CASE OF CUSTOMER CHURN

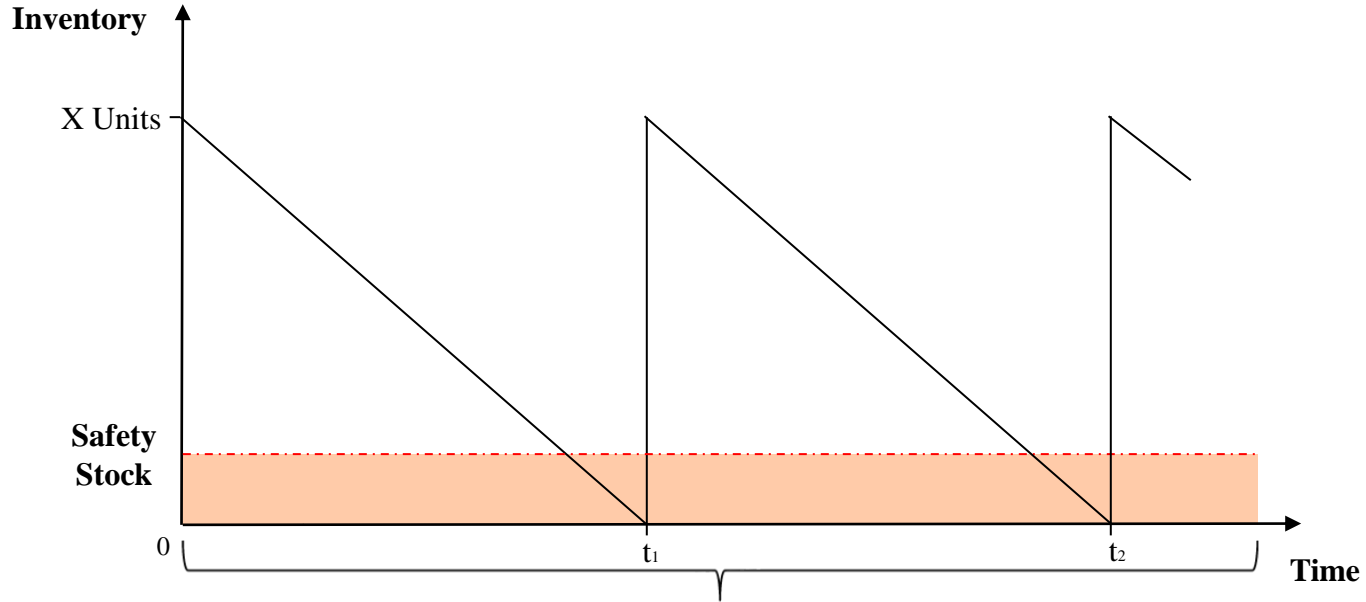
Customer	Age	SMS (#)	Call (Min)	Internet (MB)	Churn
John	35	100	30	500	Yes
Sophie	18	200	60	300	No
Victor	38	50	120	400	No
Laura	44	25	80	600	Yes



FORECASTING VS MACHINE LEARNING

Organization	Forecasting	Machine Learning
e - Commerce	Forecast the demand for each SKU for the next six months.	Predict which customers are likely to respond to a promotion.
Telco	Forecast how many new customers will sign contract the next quarter.	Predict which customers are likely to churn (change provider).

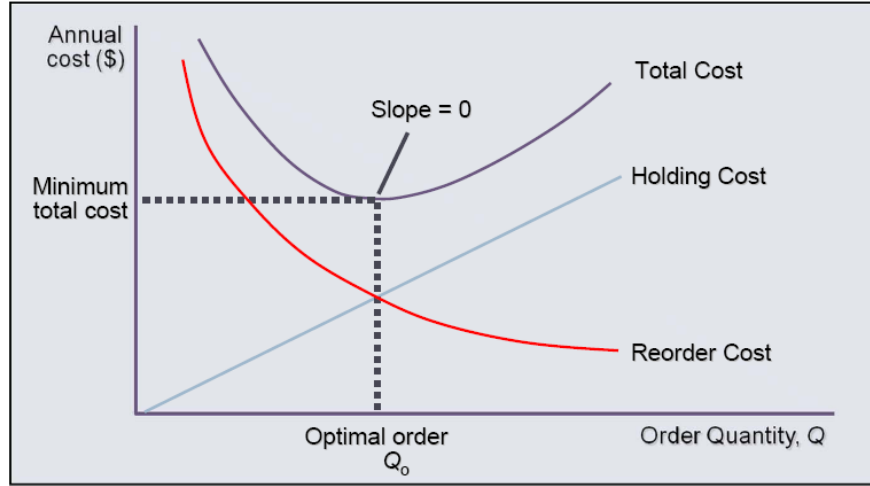
OPTIMIZATION EXAMPLE: THE CASE OF INVENTORY



When to order?

How much inventory should we order?

What should be the level of the safety stock?



Decision Variables

- When to order?
- How much should we order?
- How much should the safety stock be?

Objectives

- Minimize inventory cost
- Min customer dissatisfaction
- ⇒ Min stock outs ⇒ Min lost customers

Inventory Total Cost

Holding Cost e.g.. Insurance, Security, Obsolescence, Rent

Reorder Cost e.g. Transportation, Order, Inspection, Communication



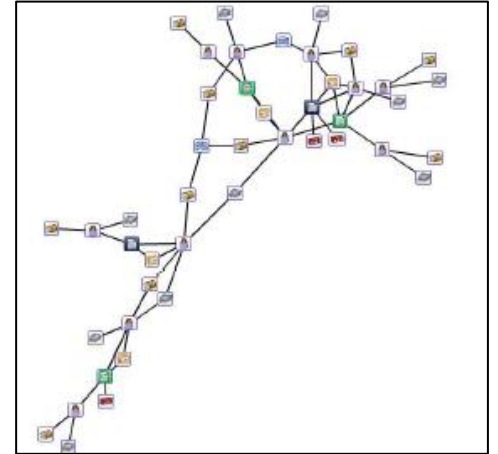
Social Network Definition

It can be any set of nodes connected by edges in a particular business setting.
Examples of social networks:

- Telephone calls between customers of a telco provider.
- E-mail traffic between people.
- Spread of illness between patients
- Research papers connected by citations

Social Network Analytics (SNA)

SNA comprise of a variety of mathematical and statistical metrics derived from the data of a social network and that can provide insights and unhide useful information.



DEFINITION OF RECOMMENDER SYSTEMS

In a general way, recommender systems are algorithms aimed at suggesting relevant items to users (items being movies to watch, text to read, products to buy or anything else depending on industries).



NETFLIX EXAMPLE

User \ Movie	Rambo II	Rocky IV	Harry Potter	Lord of the Rings	Game of Thrones
John	5	5	?	?	1
Sophie	4	1	1	1	?
Victor	5	4	1	1	?
Laura	?	?	4	4	4
Patrick	?	?	5	5	5
Clarisse	1	1	?	4	?

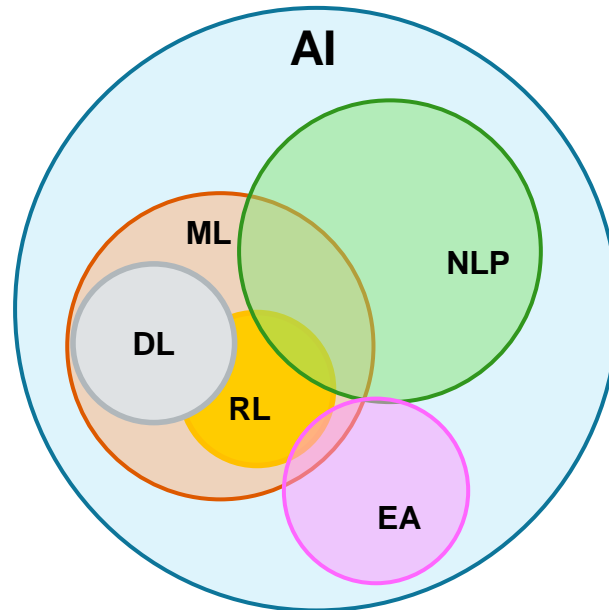
Computer Vision, often abbreviated as CV, is defined as a field of study that seeks to develop techniques to help computers “see” and understand the content of digital images such as photographs and videos.



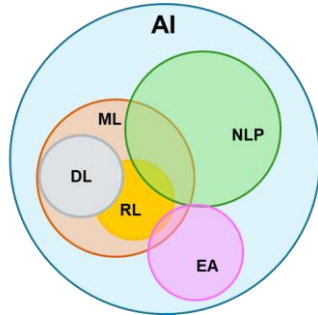
Applications of Computer Vision

- Face Recognition
- Number Plate Recognition
- Autonomous Driving
- Cancer Detection (X – Rays)
- Tumor Detection
- Mask Detection
- Theft Detection
- Social Distance
- Waiting Time Analytics
- Customer Tracking
- Cancer Cell Classification

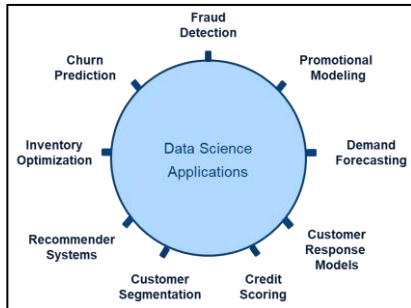
Artificial Intelligence is the field of science that is occupied with transforming machines in a way so they can think and make decisions as human beings, mainly by learning from huge amounts of past data. Analytics is in the heart of AI which comprises of Machine Learning Techniques, Natural Language Processing and Evolutionary Algorithms.



ARTIFICIAL INTELLIGENCE



Weak or Narrow
AI



Strong AI or
Artificial General Intelligence





61%

of organizations identified machine learning and AI as the most significant data initiative for next year

Source: Machine Learning and AI survey, O'Reilly Media and MemSQL, 2018

Tomorrow, AI will impact your industry

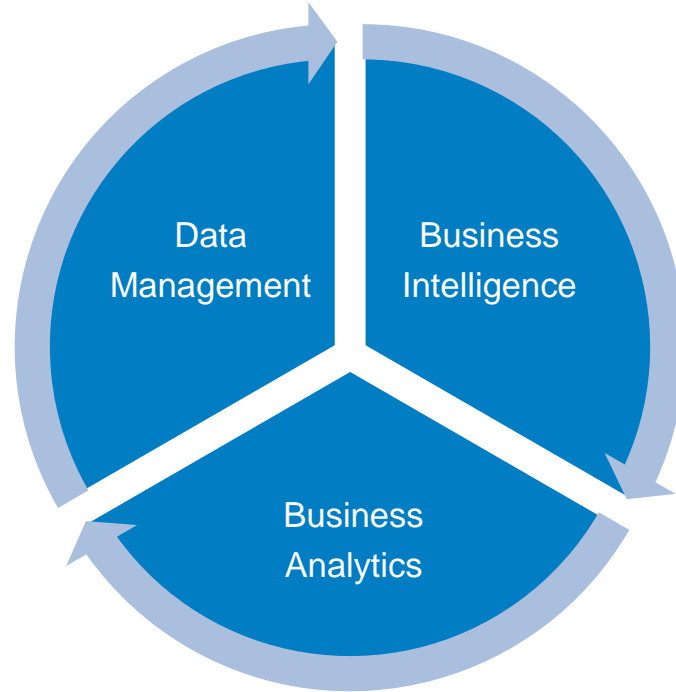


3x

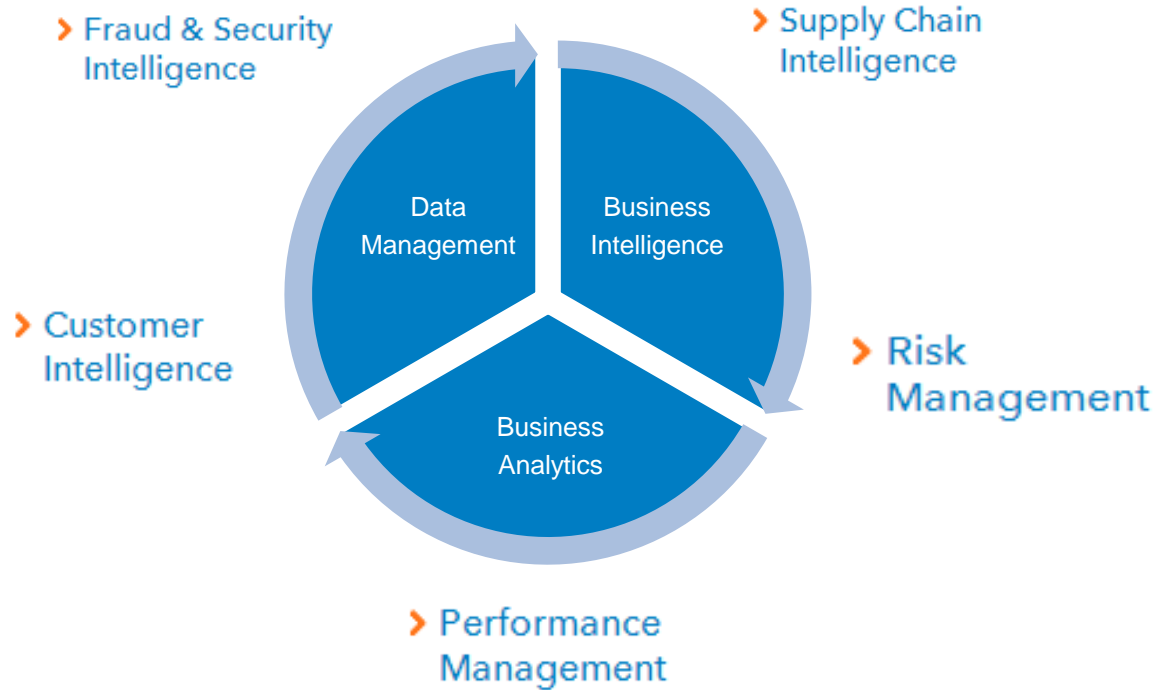
Increased investment in 2017 for AI technology compared to investment in 2016

Source: Business Tech Predictions: 10 Ways AI, Big Data, and Cloud evolved in 2017

STAGES OF DATA EXPLOITATION



IMPORTANT SAS BUSINESS SOLUTIONS



THE ANALYTICS PROCESS MODEL



Identify
Business
Problem



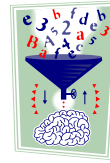
Identify
Data
Sources



Select
the
Data



Clean
the
Data



Transform
the
Data



Analyze
the
Data



Interpret,
Evaluate
and Deploy
The Model



Preprocessing

Analytics

Post-
Processing

THANK YOU!



Andreas Zaras, Data Scientist