Escola Politécnica de Pernambuco

Especialização em Ciência de Dados e Analytics

Introdução a Ciência de Dados
Aula 1

Prof. Dr. Alexandre Maciel

alexandre.maciel@upe.br

Plano de Aula

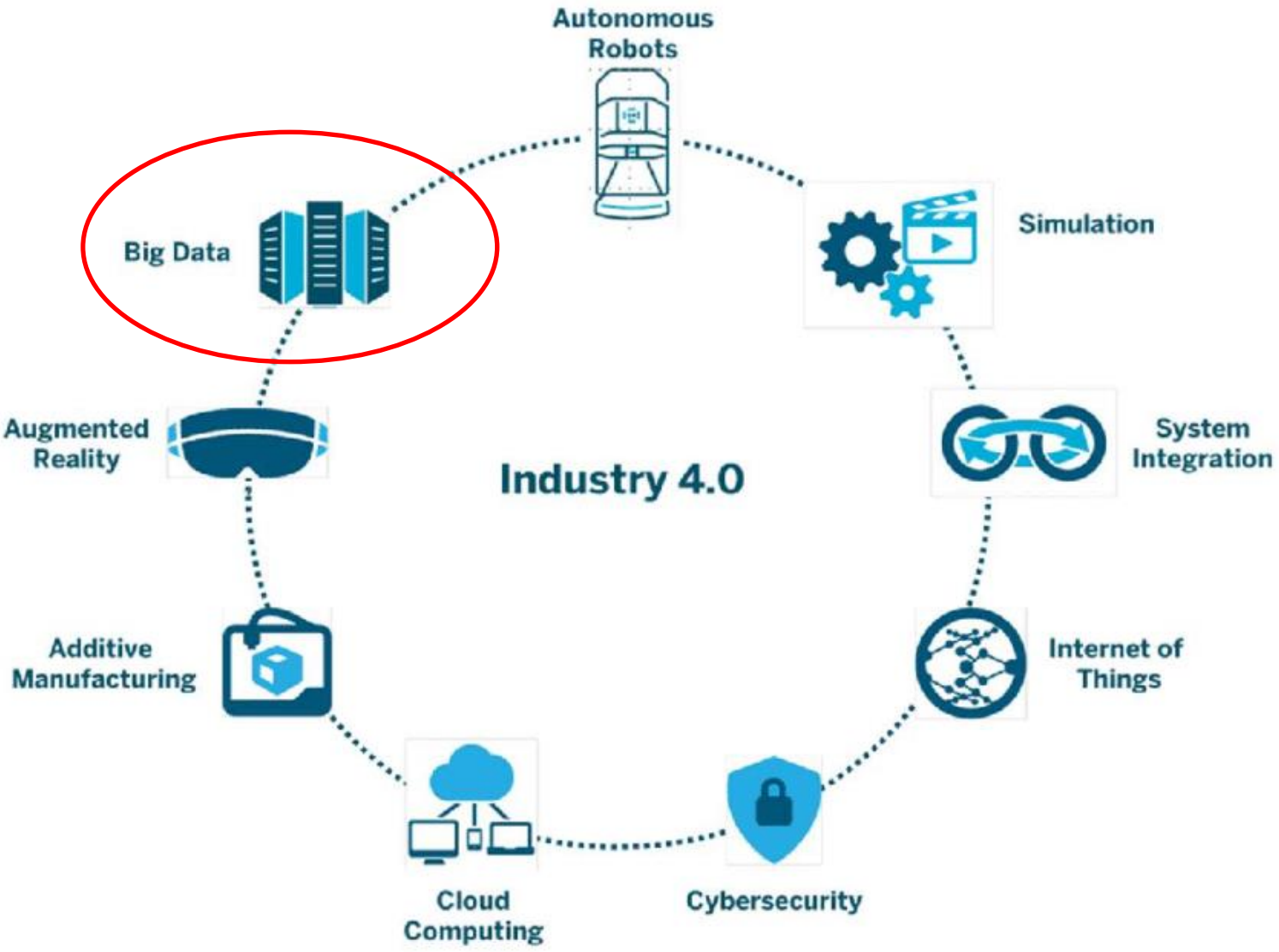
- **Objetivo**:
  - Apresentar aos alunos uma visão geral da área de ciência de dados.

- **Conteúdo**:
  - Motivação, desafios e definições.

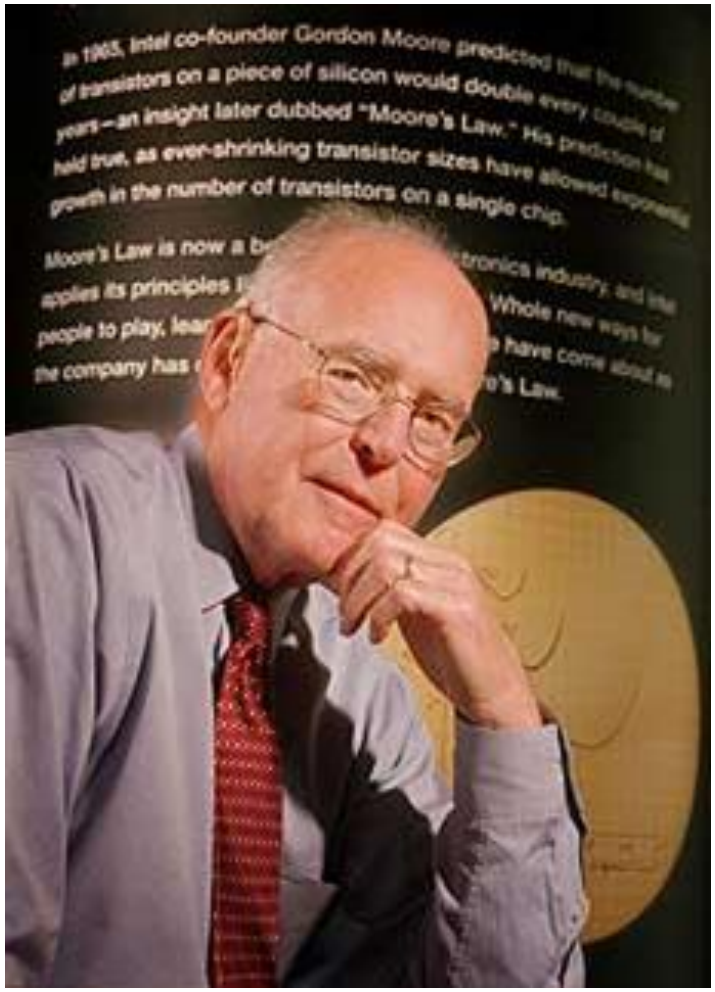
- **Referências**:
  - Amaral cap. 1, Castro, cap. 1, Provost cap. 1.

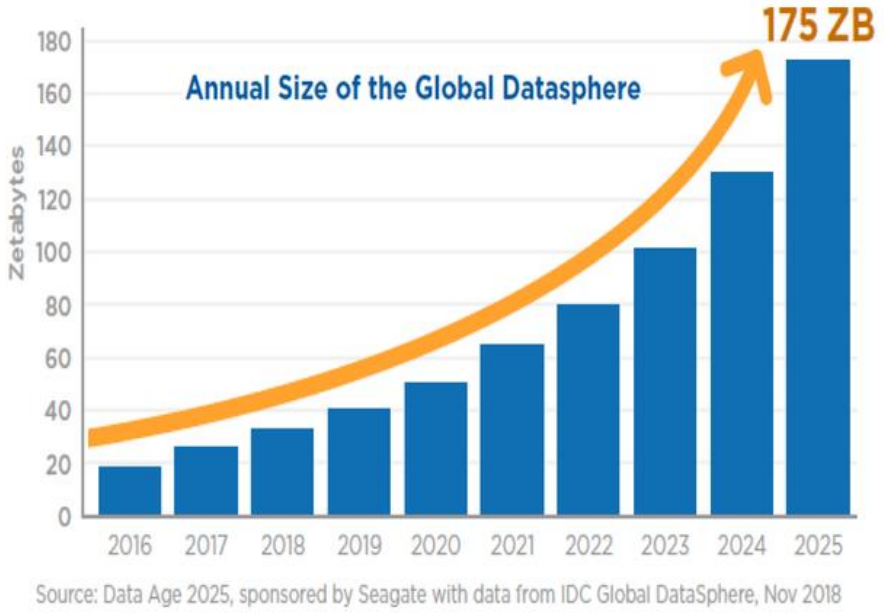
MOTIVAÇÃO

TECNOLOGIAS HABILITADORAS



LEI DE MOORE

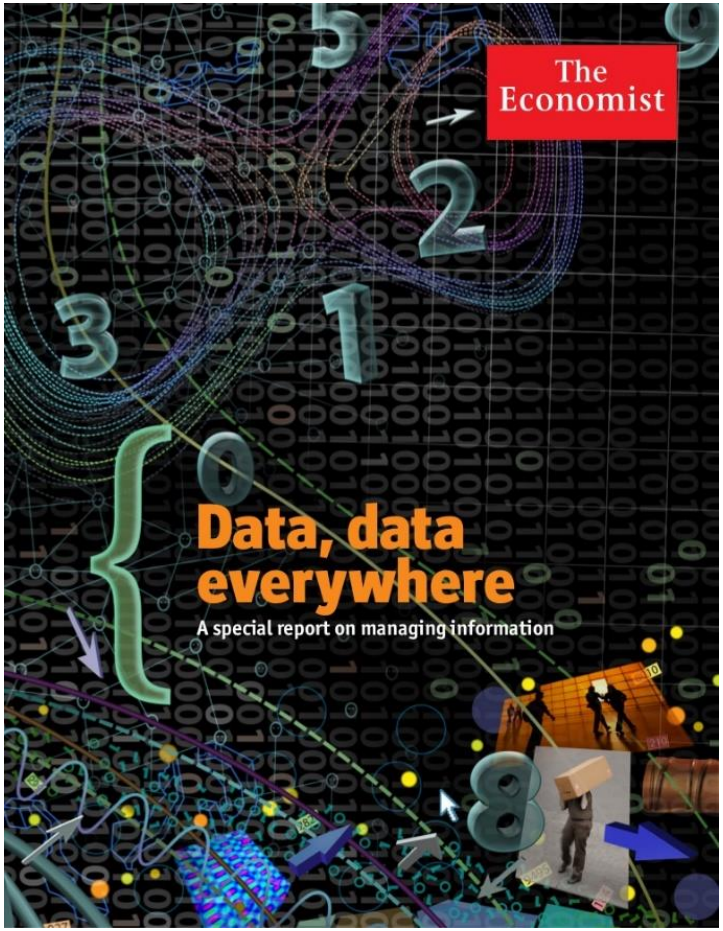




https://www.forbes.com/sites/tomcoughlin/2018/11/27/175-zettabytes-by-2025/#2a1356645459

https://newsroom.intel.com/press-kits/celebrating-the-50th-anniversary-of-moores-law/

NOVA ECONOMIA



FEVEREIRO 2010

Fuel of the future

Data is giving rise to a new economy

How is it shaping up?

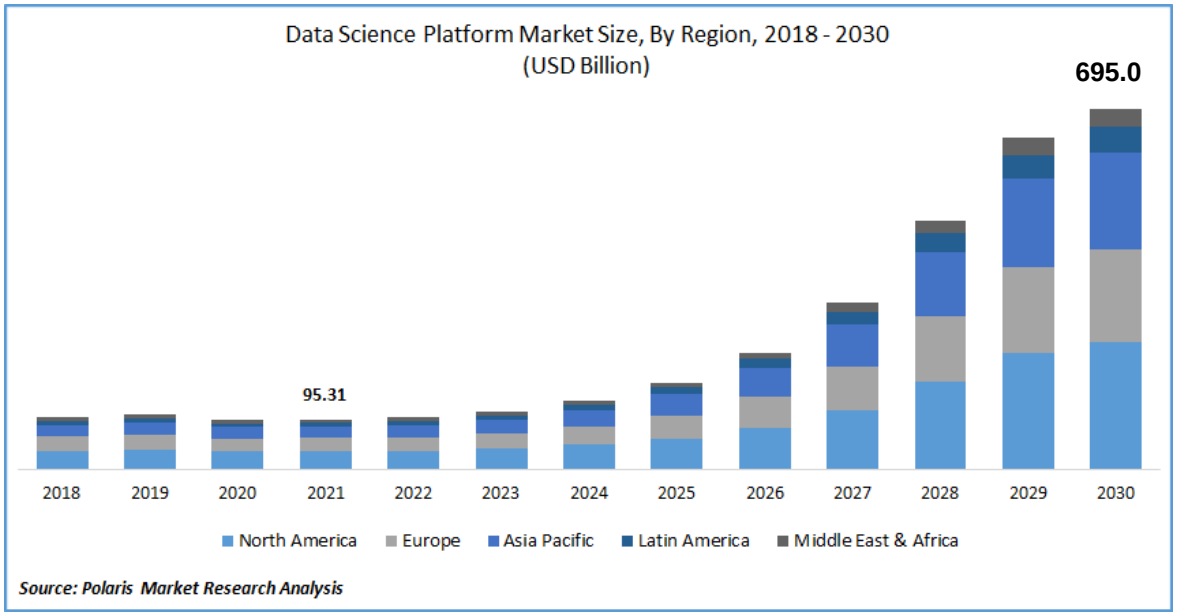
Regulating the internet giants

The world's most valuable resource is no longer oil, but data

The data economy demands a new approach to antitrust rules

MARÇO 2017

TAMANHO DO MERCADO

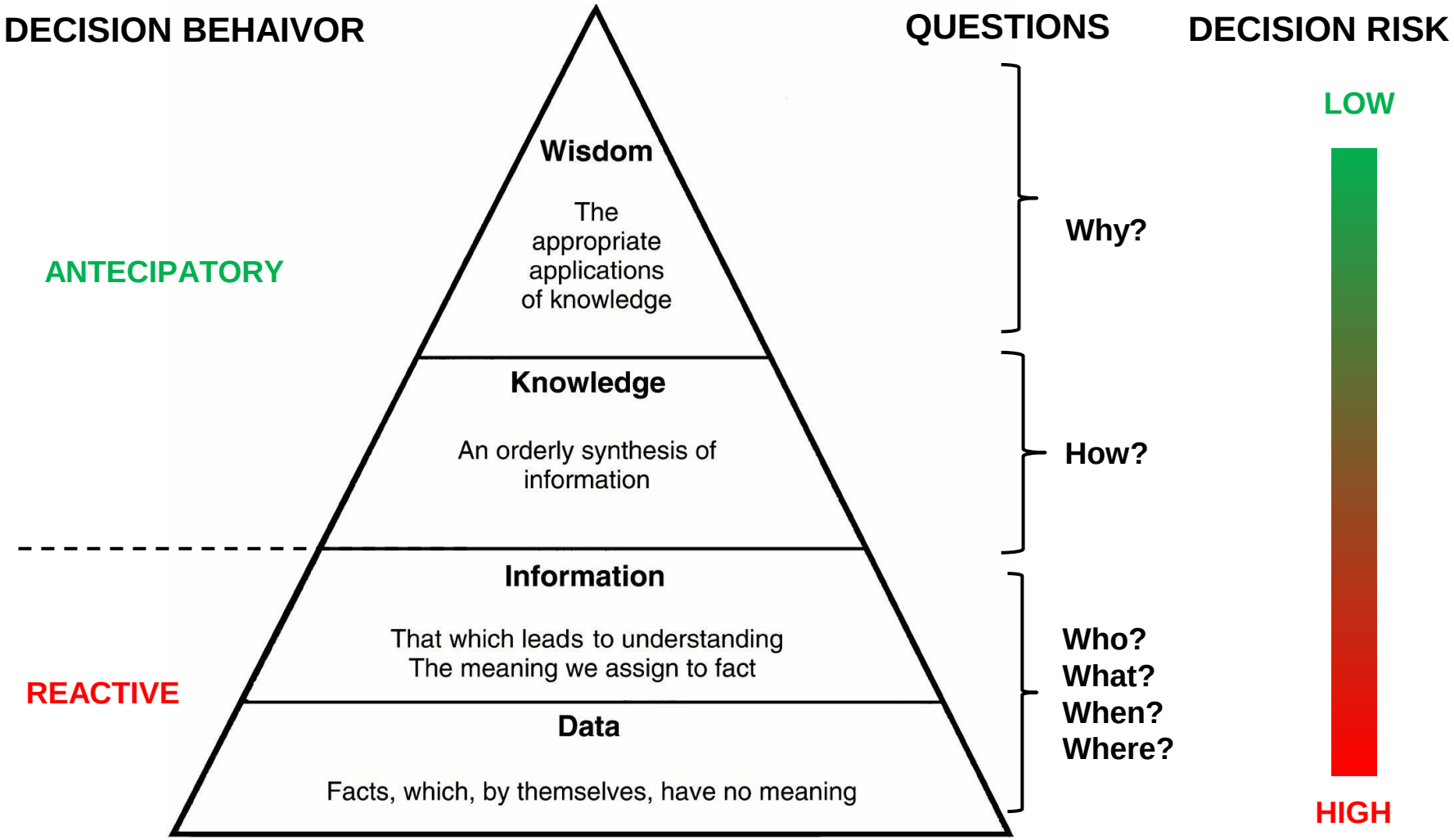


https://www.polarismarketresearch.com/industry-analysis/data-science-platform-market

DESAFIO

“DATA RICH BUT INFORMATION POOR SITUATION”

DIKW PYRAMID



DATA SCIENCE COMO ATIVO ESTRATÉGICO

Dados são potencialmente o maior ativo de uma empresa - se ela puder monetizá-los.
Brian Householder CEO Hitachi Vantara.

VANTAGEM COMPETITIVA

...quando uma empresa sustenta os lucros que excedem a média...

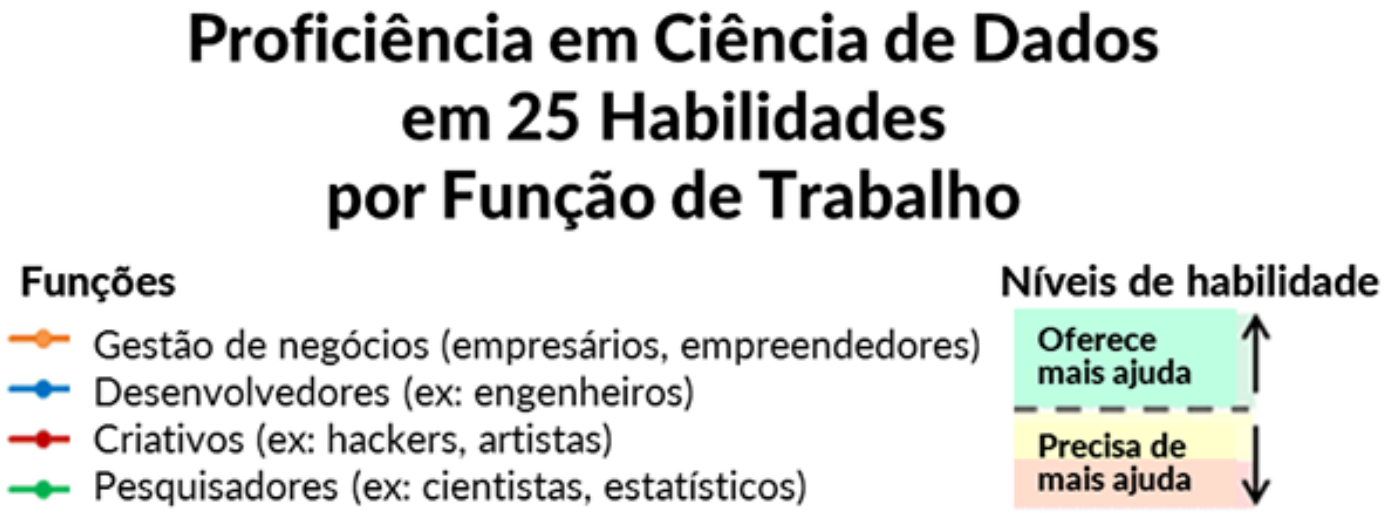
PENSAMENTO ANALÍTICO DE DADOS

- Como os dados podem ajudar na resolução de problemas de negócio?

- Análise de dados tornou-se crucial para a estratégia de negócios.

- Empresas *data-driven* em busca de obter a vantagem competitiva.

SKILLS

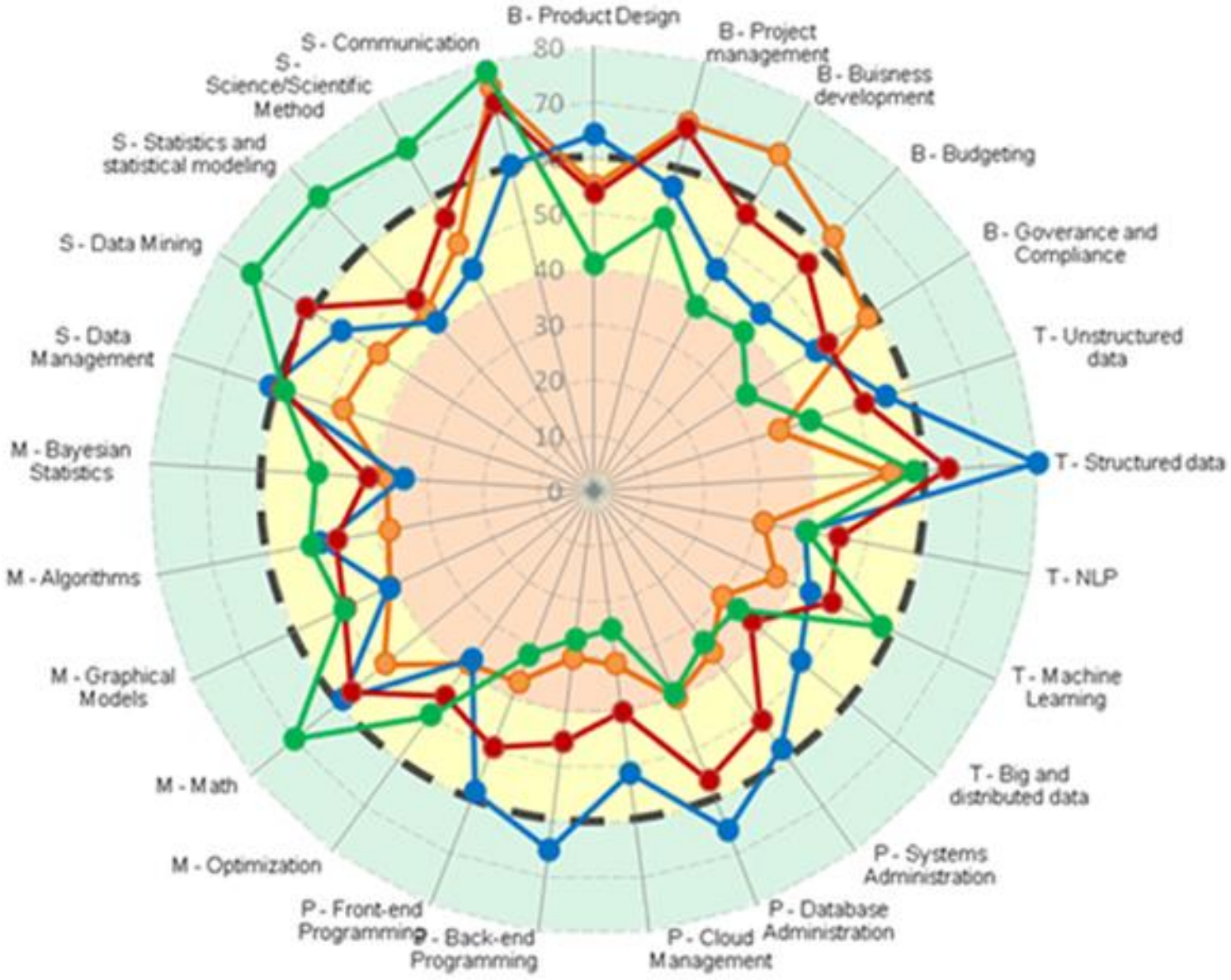


NOTA: Os dados são baseados em respostas de 490 profissionais de dados. Estes profissionais foram convidados a avaliar a sua proficiência em 25 habilidades usando uma escala de 0 (não sei) a 100 (expert). O gráfico é baseado em entrevistados que escolheram apenas uma função primária de trabalho. Negócios (n = 65); Desenvolvedor (n = 47); Criativo (n = 25); Pesquisador (n = 101)

AnalyticsWeek

BUSINESS BROADWAY

https://businessoverbroadway.com/2015/09/23/investigating-data-scientists-their-skills-and-team-makeup/



https://businessoverbroadway.com/2015/09/23/investigating-data-scientists-their-skills-and-team-makeup/

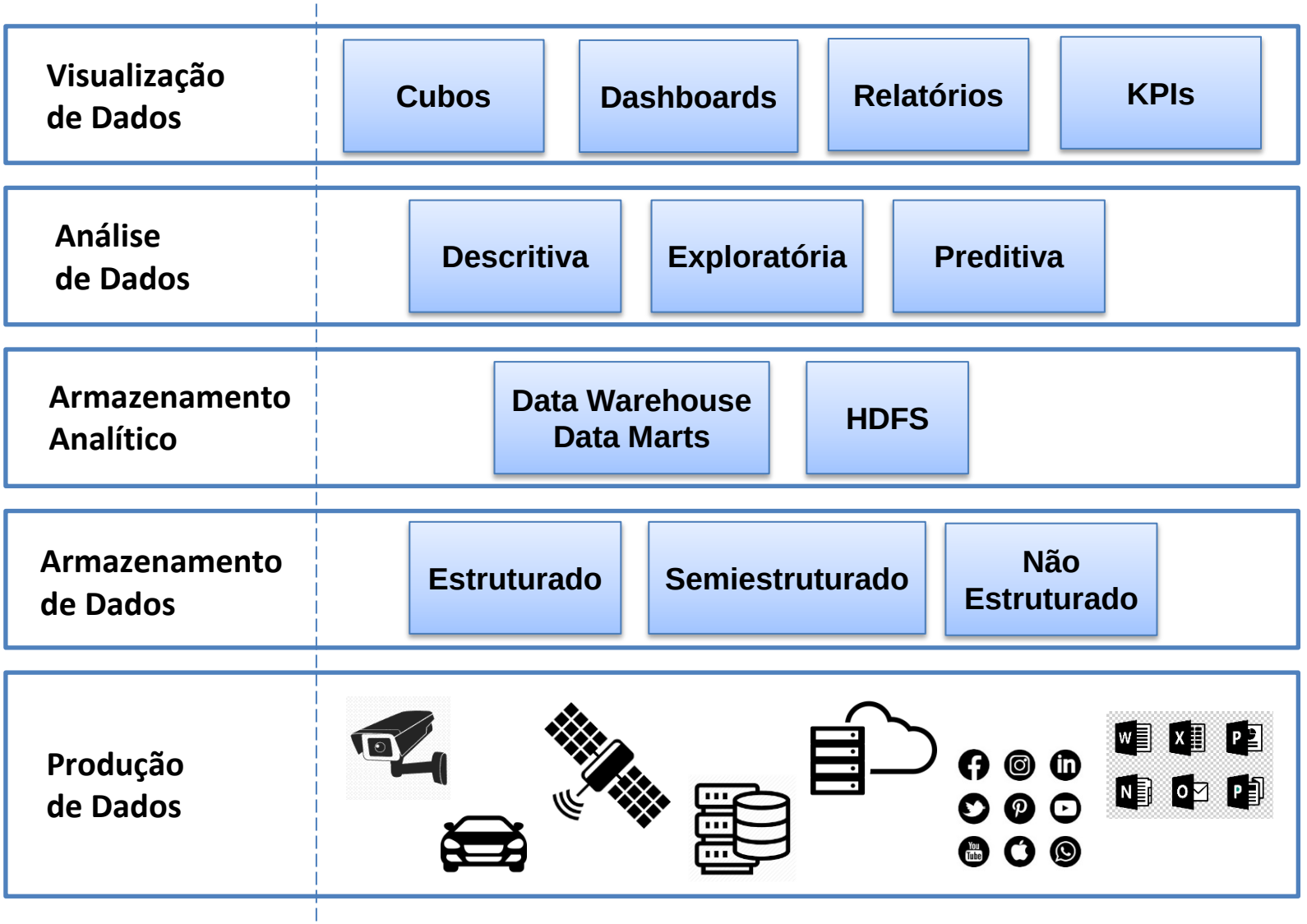
DEFINIÇÕES

Ciência de Dados

envolve **princípios**, **processos** e **técnicas** para compreender **fenômenos** por meio de análise de dados **automatizada**.

(Provost & Fawcett).

PANORAMA EM CIÊNCIA DE DADOS



DEFINIÇÕES

Inteligência Artificial

é a ciência e engenharia de fazer **máquinas inteligentes**, com objetivo de fazer com que elas **realizem tarefas** que, se **feitas por pessoas**, exigiriam inteligência.

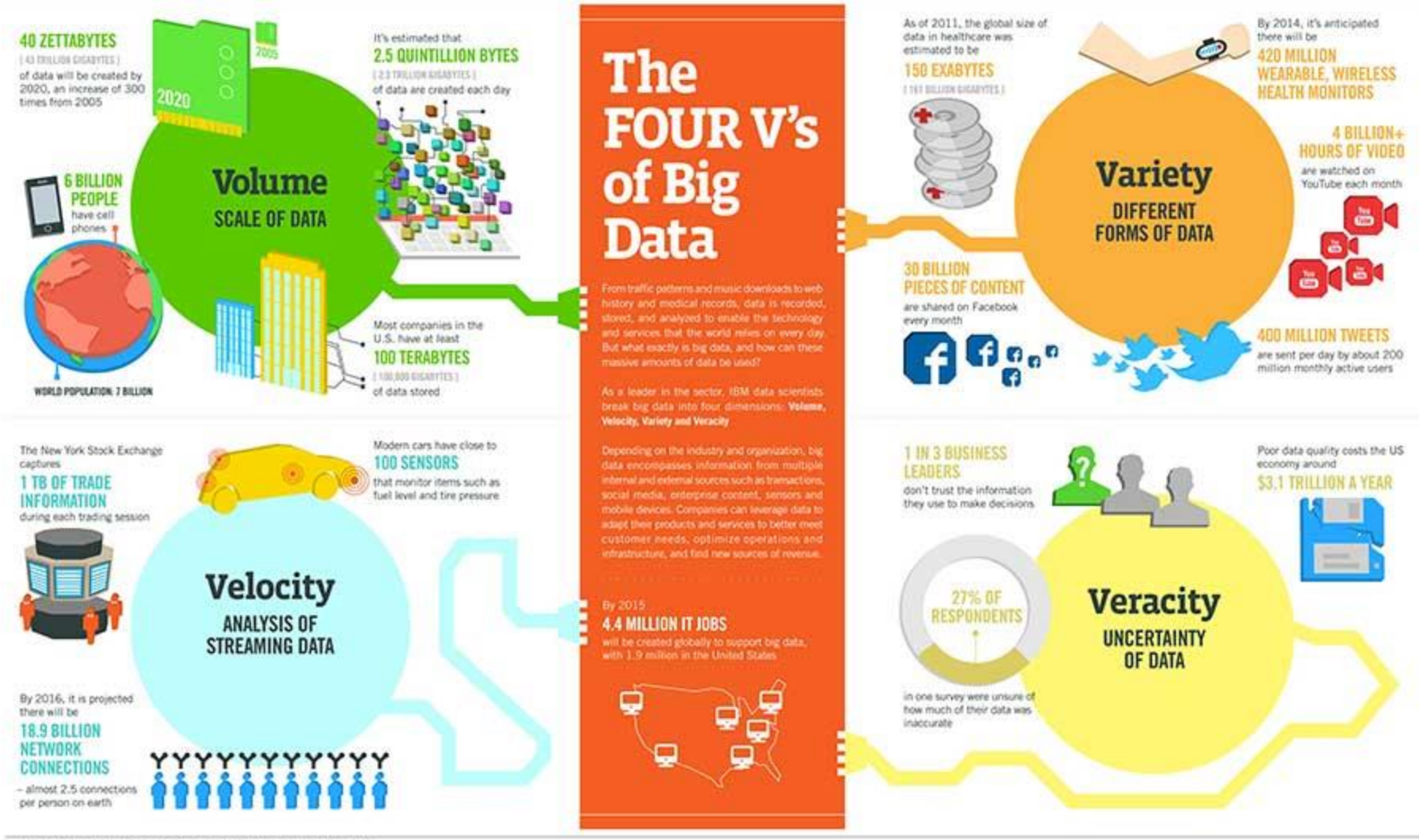
(John McCarthy).

DEFINIÇÕES

Big Data

significa conjuntos de dados que são **grandes demais** para os sistemas tradicionais de processamento e, portanto, exigem **novas tecnologias** para processá-los.

(Provost e Fawcett).


40 ZETTABYTES
[43 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005
2005
2020
It's estimated that
2.5 QUINTILLION BYTES
[2.3 TRILLION GIGABYTES]
of data are created each day
6 BILLION PEOPLE
have cell phones
WORLD POPULATION: 7 BILLION
Volume
SCALE OF DATA
Most companies in the U.S. have at least
100 TERABYTES
[100,000 GIGABYTES]
of data stored
The New York Stock Exchange captures
1 TB OF TRADE INFORMATION
during each trading session
Modern cars have close to
100 SENSORS
that monitor items such as fuel level and tire pressure
Velocity
ANALYSIS OF STREAMING DATA
By 2016, it is projected there will be
18.9 BILLION NETWORK CONNECTIONS
– almost 2.5 connections per person on earth
The FOUR V's of Big Data
From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?
As a leader in the sector, IBM data scientists break big data into four dimensions: Volume, Velocity, Variety and Veracity
Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.
By 2015
4.4 MILLION IT JOBS
will be created globally to support big data, with 1.9 million in the United States
As of 2011, the global size of data in healthcare was estimated to be
150 EXABYTES
[161 BILLION GIGABYTES]
By 2014, it's anticipated there will be
420 MILLION WEARABLE, WIRELESS HEALTH MONITORS
4 BILLION+ HOURS OF VIDEO
are watched on YouTube each month
Variety
DIFFERENT FORMS OF DATA
30 BILLION PIECES OF CONTENT
are shared on Facebook every month
400 MILLION TWEETS
are sent per day by about 200 million monthly active users
1 IN 3 BUSINESS LEADERS
don't trust the information they use to make decisions
Poor data quality costs the US economy around
$3.1 TRILLION A YEAR
27% OF RESPONDENTS
in one survey were unsure of how much of their data was inaccurate
Veracity
UNCERTAINTY OF DATA
Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS

DEFINIÇÕES

Analytics

uso amplo de **dados**, de análise estatística e quantitativa, de **modelos** descritivos e preditivos para orientar **decisões** e agregar **valor**.

(Davenport & Kim).

MATURIDADE EM ANALYTICS

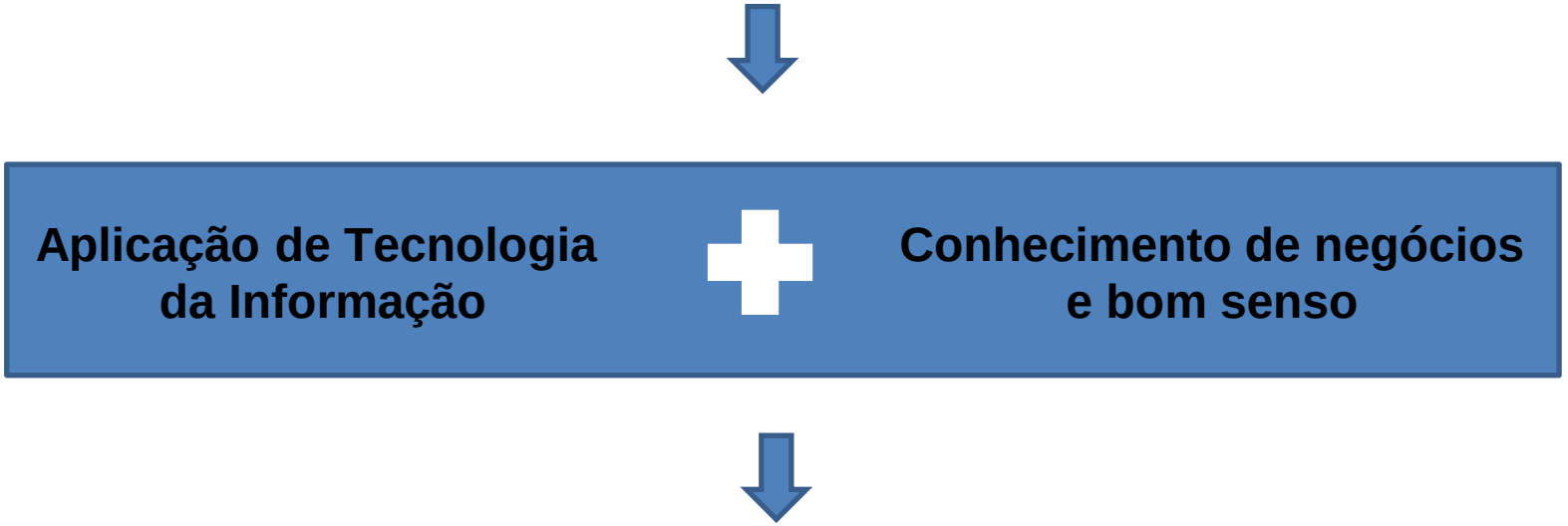
DEFINIÇÕES

Mineração de Dados

é o processo de **extração de conhecimento** em grandes quantidades de dados.
(Han e Kamber).

METODOLOGIAS

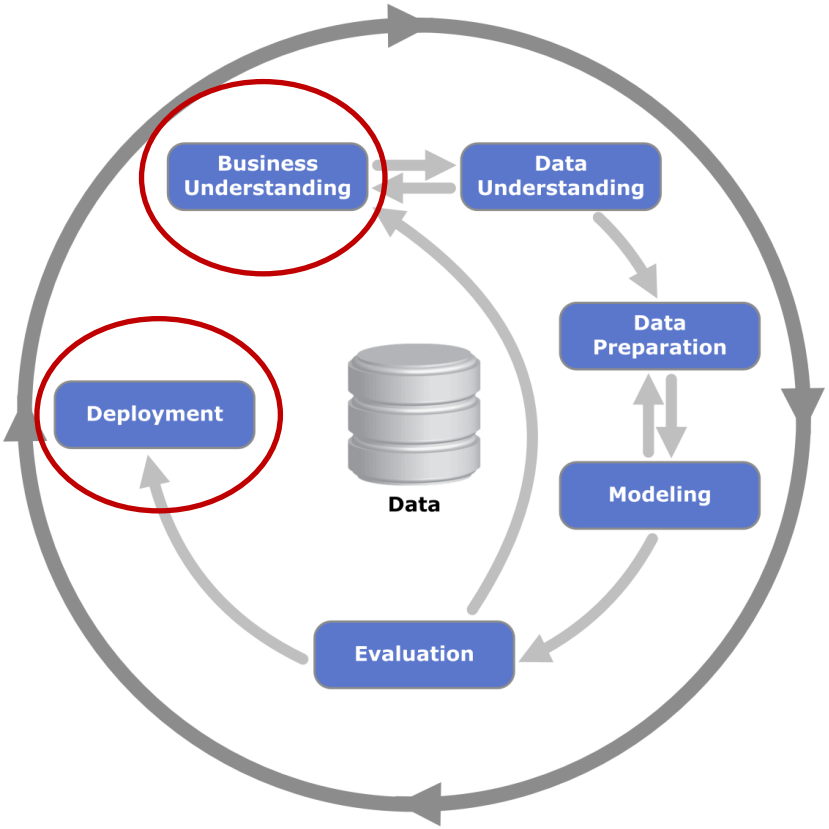
Ciência de Dados permite processo com estágios muito bem definidos



Projetos sistemáticos, sem esforços heroicos conduzidos ao acaso

METODOLOGIAS

CRISP-DM



Chapman *et al*, 2000

OSEMN

Mason and Wiggins, 2010

ATIVIDADE 1

1. Montem seus grupos (3-4 alunos).
2. Escolha um dos problemas disponibilizados para ser o seu projeto e descreva-o.
3. Entregue o primeiro slide de sua apresentação.
4. Faça o upload no Google Classroom.