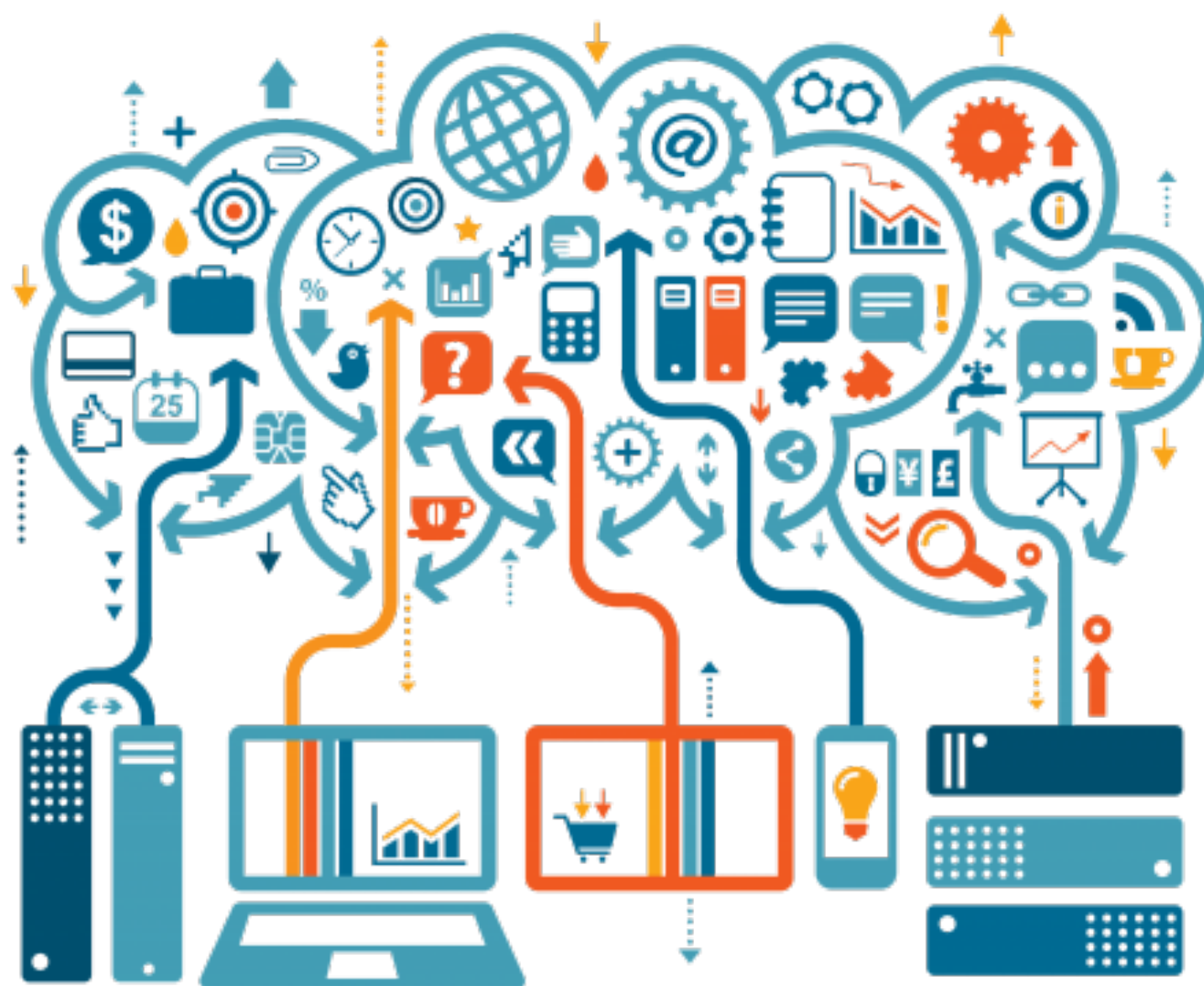# Introdução a Ciencia de Dados

Hitoshi Nagano, Ph.D.

POR QUE DATA SCIENCE?

- 3?

- 4?

- 10?

- https://www.psychologicalscience.org/pdf/ps/mind_variables.pdf

- Através de dados reais
- a máquina (computador) entende os padrões…
- …e ajuda a predizer um resultado (classe ou valor numérico)
- predição é usada para tomada de uma decisão de negócios

Forbes
INSIGHTS

BLAZING THE TRAIL FROM DATA TO INSIGHT TO ACTION

EXECUTIVE BRIEFING

The enterprises leading the way in data analytics are demonstrating an enormous capability to capture, process, scale and make available data to their organizations and across their networks of customers, partners and suppliers. But is this enough? The challenge is to develop this data into insights that can be applied to increase business value—moving from data to insight to action. Executives say there is a direct correlation between well-designed analytics programs and success in the marketplace. However, while many executives are embracing customer data analytics to guide their businesses, the infrastructures and processes they require to support and sustain such efforts still lag.

These are the findings of a new survey of 105 executives of large global organizations, conducted by Forbes Insights in partnership with SAS, which explores the depth of their embrace of data analytics. The survey finds that while half of large enterprises (with $500 million or more in annual revenue) have tightly integrated customer data analytics into their key processes, there are many areas that are still works in progress.

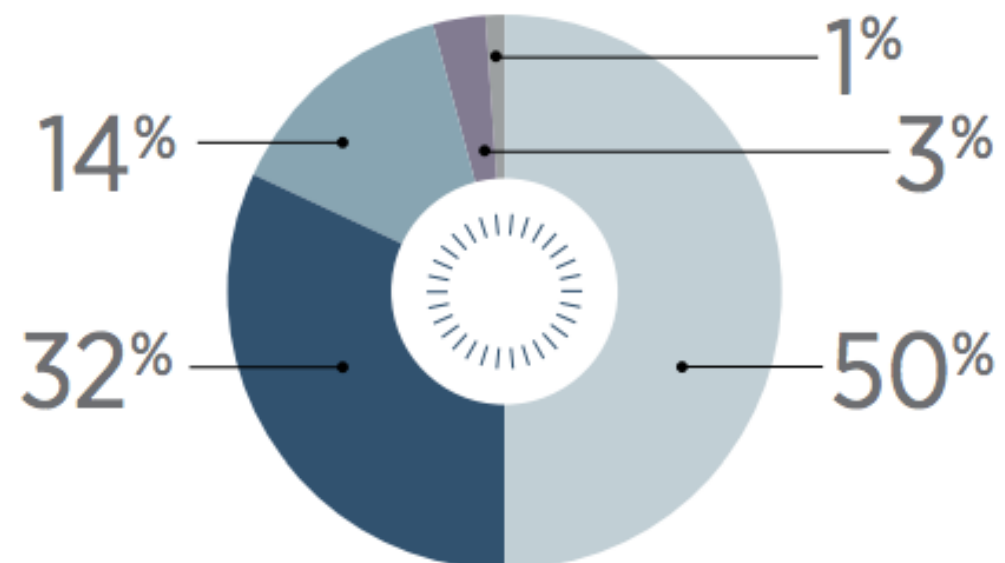**KEY FINDINGS FROM THIS SURVEY INCLUDE THE FOLLOWING:**

- Half of large enterprises are still in the learning stages of customer analytics, and are in the process of applying these analytics to enhance their customer experience. The results of such efforts, no

Fonte: "BLAZING THE TRAIL FROM DATA TO INSIGHT TO ACTION", Forbes

FIGURE 1

IN YOUR EFFORTS TO BECOME A MORE CUSTOMER-CENTERED BUSINESS, WHICH OF THE FOLLOWING STATEMENTS HOLDS TRUE FOR YOUR ORGANIZATION?
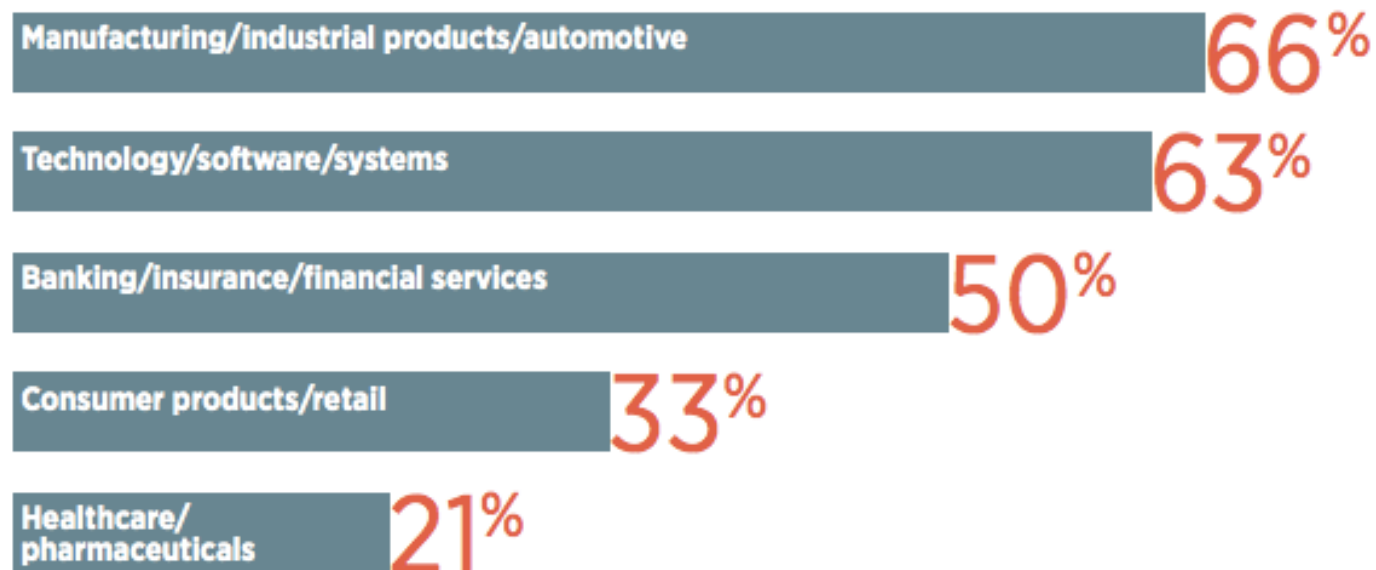
We provide a superior customer experience through a well-developed and designed enterprise data analytics effort

We are rapidly and systematically developing data analytics capabilities to improve our customer experience

We have data analytics capabilities in many parts of the organization, and are beginning to apply these analytics to improve the customer experience

We are not using analytics for customer experience

We are still working on developing data analytics and have difficulty providing consistent customer experience

1%

3%

14%

32%

50%

FIGURE 2

PROVIDE SUPERIOR CUSTOMER EXPERIENCE THROUGH ANALYTICS– BY INDUSTRY GROUP

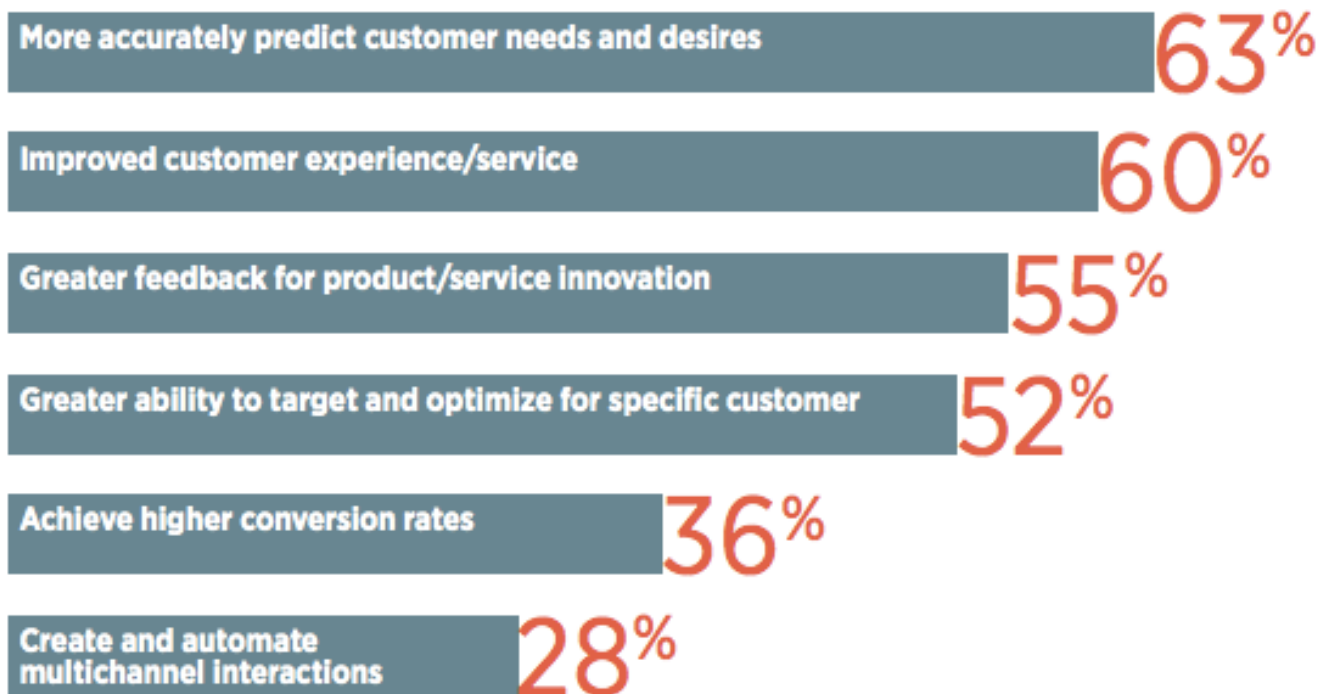| Industry Group	%
Manufacturing/industrial products/automotive	66%
Technology/software/systems	63%
Banking/insurance/financial services	50%
Consumer products/retail	33%
Healthcare/pharmaceuticals	21%

FIGURE 3

WHAT ROLE HAS DATA ANALYTICS PLAYED IN YOUR ORGANIZATION'S ABILITY TO DELIVER A SUPERIOR CUSTOMER EXPERIENCE?

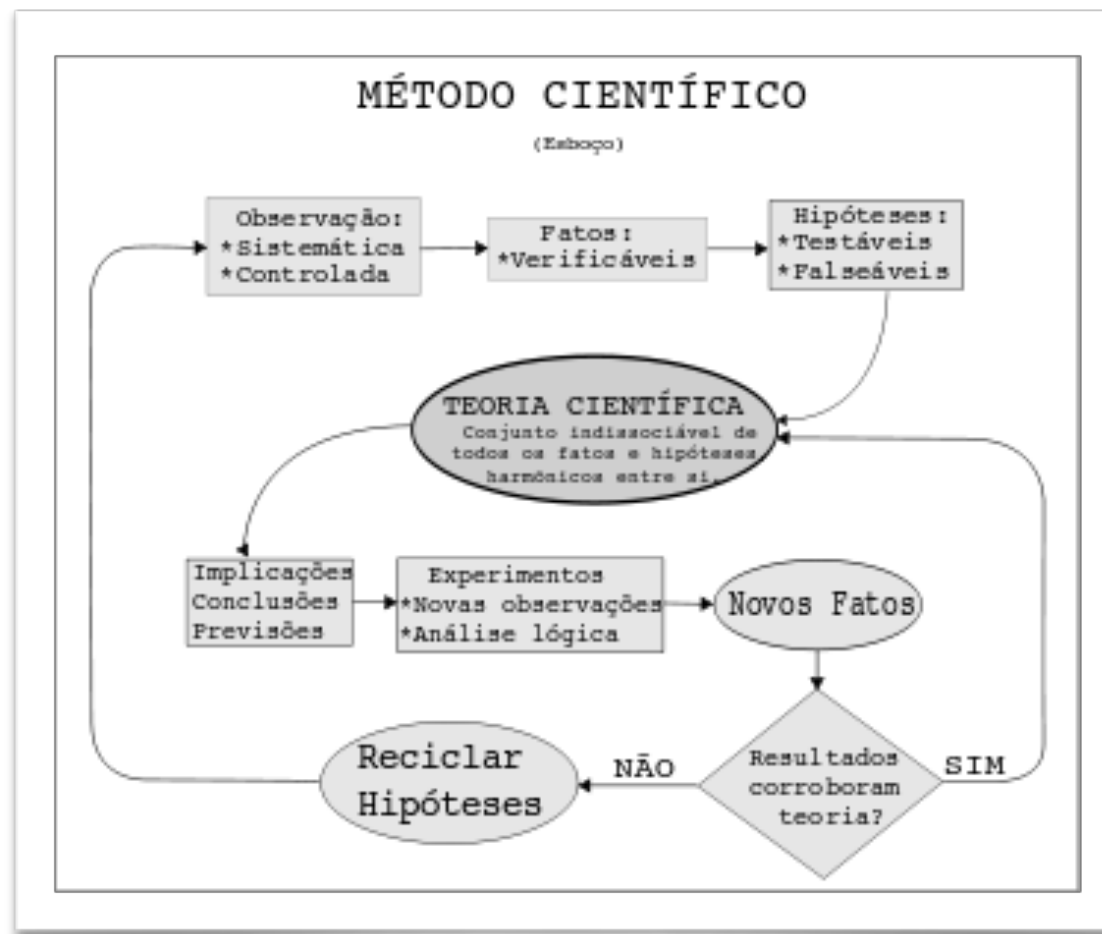Data analytics has had an extremely positive impact on our customer experience

Data analytics has pushed the needle in a positive direction

Data analytics has shown potential in areas where it has been applied

Data analytics has not played a significant role yet

8%

5%

32%

54%
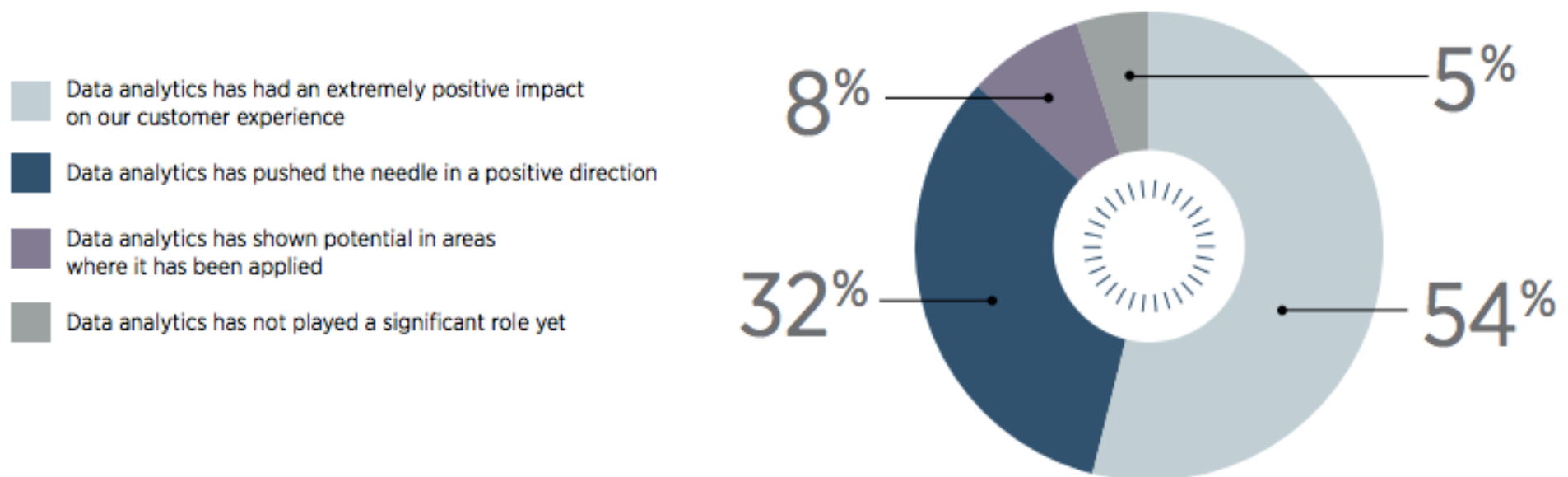
FIGURE 9 WHAT BENEFITS CAN BE GAINED FROM ACHIEVING A MORE COMPLETE OR UNIFIED VIEW OF THE CUSTOMER?

Benefit	%
More accurately predict customer needs and desires	63%
Improved customer experience/service	60%
Greater feedback for product/service innovation	55%
Greater ability to target and optimize for specific customer	52%
Achieve higher conversion rates	36%
Create and automate multichannel interactions	28%

- ciencia em "ciencia de dados"
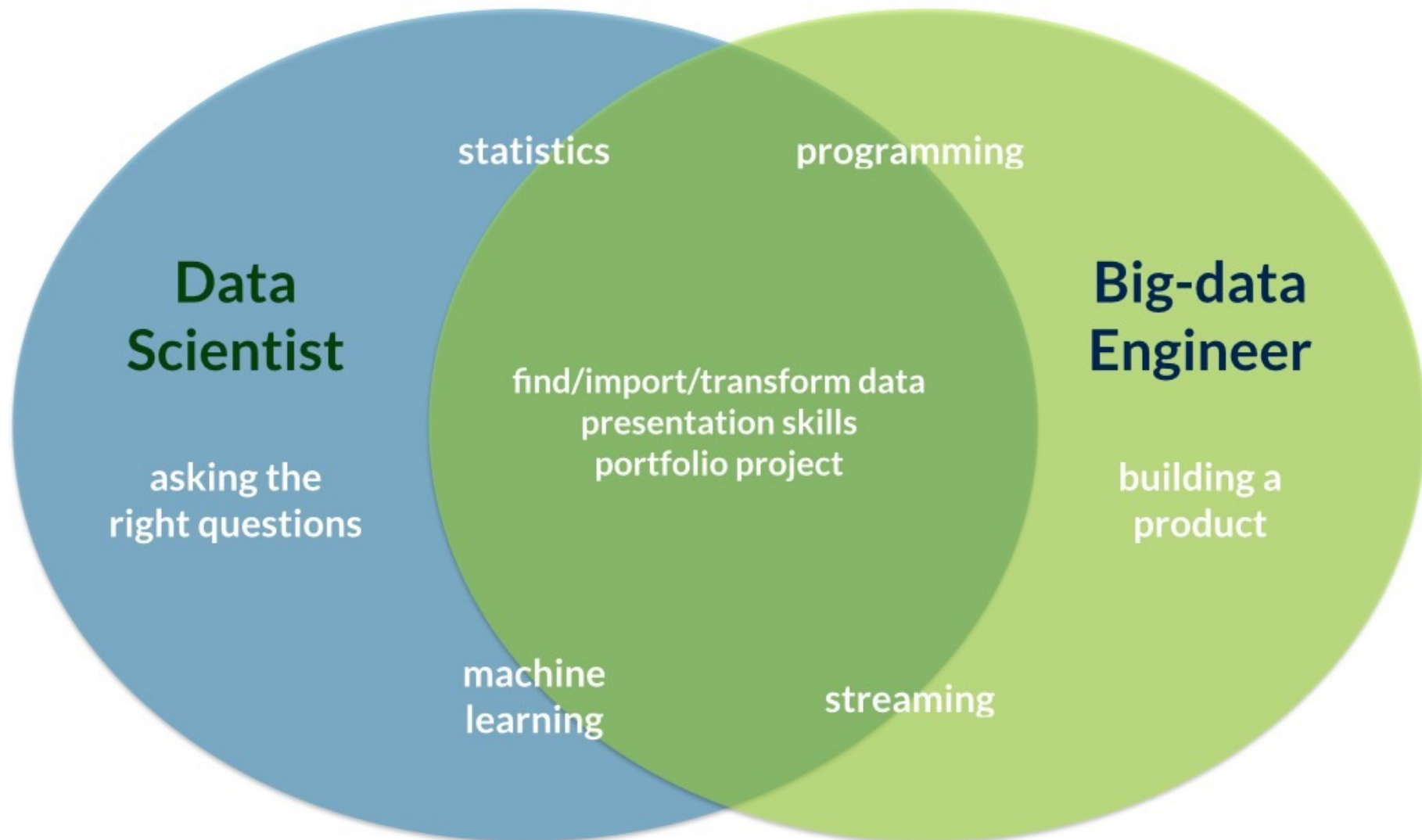- método científico

PERFIL DO CIENTISTA DE DADOS

– estatística + matemática

– computação + ferramentas

– perspicácia comercial + intuição

   (ex: domain knowledge)

– explorador + inquisitivo
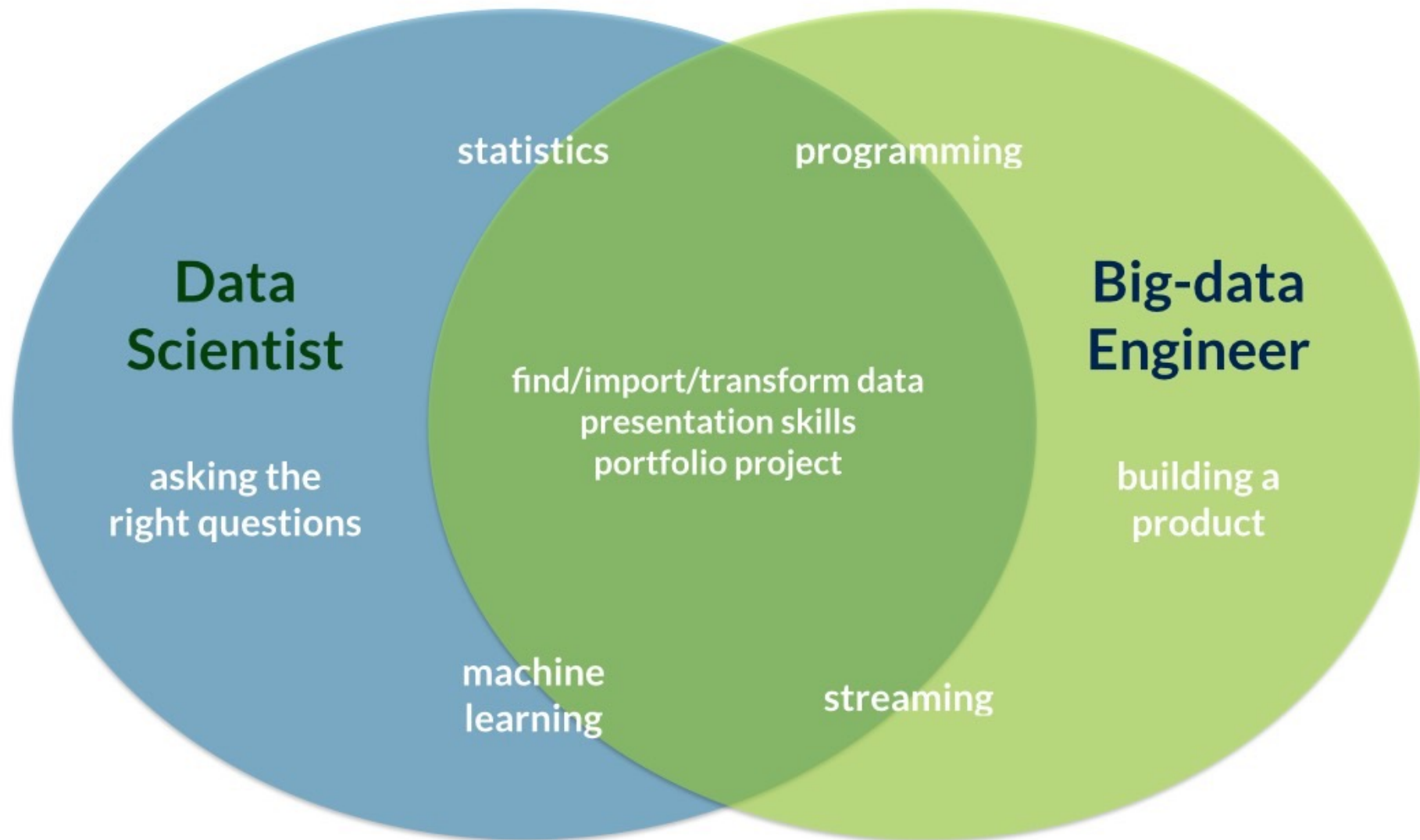
– arte e estética, story telling

```
def f(n):
  if n == 1:
    return 1
  else:
    return n*f(n-1)
```
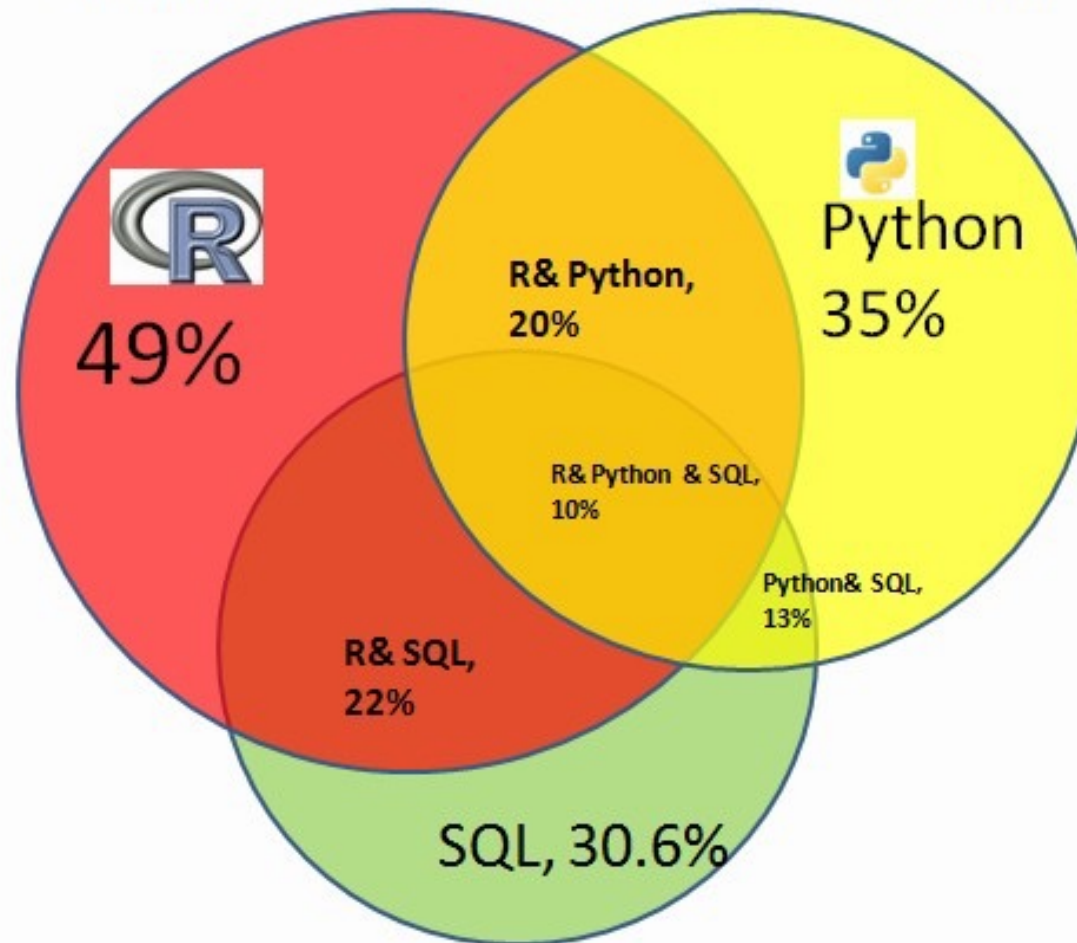
$$lim_{z \to -\infty} \frac{1}{1 + e^{-z}}$$

$$lim_{z \to \infty} \frac{1}{1 + e^{-z}}$$

churn, up-sell & cross-sell, lifetime? RFV? NBO? CR, CTR, CPA, CPC?

KDnuggets 2014 Poll: Languages used for Analytics/Data Mining

R 49%

Python 35%

R & Python, 20%

R & Python & SQL, 10%

Python & SQL, 13%

R & SQL, 22%

SQL, 30.6%

http://www.kdnuggets.com/2014/08/four-main-languages-analytics-data-mining-data-science.html

PROJETOS DE DATA SCIENCE

- existe um problema de negócios?

- a empresa precisa da análise?

- tenho dados?
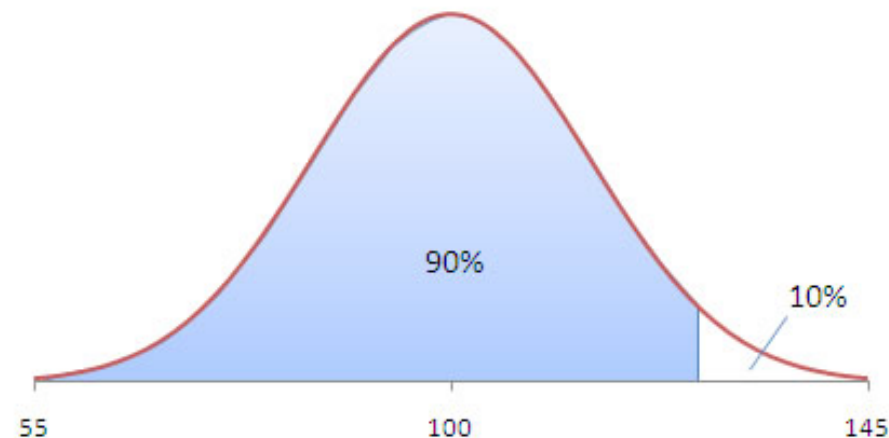
- consigo extrair boas features?

- existe um padrão?

- predição de risco de churn

- clusterização de clientes

- extração de tópicos

- recomendação de produtos

- score de clientes

- reordernação de inadimplentes na fila de cobrança

- predição de severidade de falhas em redes

- predição de convulsões epilépticas

- modelagem de funil de conversão

MERCADO DE TRABALHO
DATA SCIENCE

FGV MANAGEMENT

FGV MANAGEMENT

- start-ups

- consultorias

- telecom, varejo, e-commerce, bancos, governo, farmaceuticas, …

- http://corporate.canaltech.com.br/noticia/vivo/vivo-cria-nucleo-de-bi-e-big-data-em-departamento-com-120-profissionais-74597/

- Forbes 2014
  (https://www.forbes.com/sites/gilpress/2015/10/21/the-number-of-data-scientists-has-doubled-over-the-last-4-years/#5027125538d4)
  Quantidade: **11430 profissionais**

- Stitchdata 2015
  (https://www.stitchdata.com/resources/reports/the-state-of-data-science/)
  Quantidade: **11400 funcionários empregados em empresas**

- KDnuggets 2014
  (http://www.kdnuggets.com/2014/03/how-many-data-scientists-are-there.html)
  Quantidade: **150K a 250K** pessoas tecnicamente capacitadas (perfil LinkedIn)

OUTROS FATOS E PIADAS SOBRE DATA SCIENCE

- http://priceonomics.com/whats-the-difference-between-data-science-and/

- https://www.analyticsvidhya.com/blog/2015/10/job-comparison-data-scientist-data-engineer-statistician/

- http://blog.revolutionanalytics.com/2013/05/statistics-vs-data-science-vs-bi.html

Analysis Tool	Similar Superhero	Super Powers in Common
R	Batman	• Detective Work • Intelligence • Cunning • Usage of Tools • More Brain than Muscles
Python	Superman	• Muscle Power • Super Strength • Elegance • Wide Range • More Muscles than Brain

http://i2.wp.com/ucanalytics.com/blogs/wp-content/uploads/
2015/10/R.jpeg?resize=198%2C150

- …machine learning faz com que nossa ação seja mais mais assertiva.

- A gestão data-driven nos leva a decisões mais assertivas

PEQUENAS Empresas & GRANDES Negócios

BUSCA   OK

movimento empreenda
extreme makeover
empreendedor de sucesso
melhores franquias

EMPRESAS Negócios
COMO FATURAR COM O NOVO MERCADO DE ALIMENTAÇÃO

home   como começar   dia a dia   franquias   banco de ideias   startups   mei | REVISTA   ASSINE

DIA A DIA > GESTÃO

TAMANHO DO TEXTO   A- A+

Como tomar decisões mais assertivas na sua empresa

Quando surge aquele pepino, como conduzir a tomada de decisão? Veja algumas dicas para resolver

Da Endeavor Brasil - 02/10/2015

BUSCA DE FRANQUIAS

- montei um data lake, meu data warehouse,

  gastei uma 💰 💰 💰 , o que fazer agora?

- minha infra estrutura Hadoop para meus 100M de

  clientes

https://www.surveysystem.com/sscalc.htm

https://cloud.google.com/products/calculator/

```
join_us.py
> import Analytics_Center_of_Excellence as
ACE
> print ACE.vagas.title
[1] "Cientistas de Dados"
> print ACE.vagas.to_know
[1] "O cientista de dados do ACE irá atuar
em projetos atendendo diversas áreas do
banco (CRM, Crédito, Cobrança, Fraudes,
...) e terá como objetivo atacar problemas
desafiadores em ciência de dados.
As soluções envolvem profundo conhecimento
teórico e prático das principais técnicas
de machine learning e ferramentas do mundo
Big Data. Venha para o ACE!"
> print ACE.tecnologias.to_know
```

```
> help ACE.vagas
[1] "Mande e-mail para
    cienciadedados@itau-unibanco.com.br
com o assunto sendo o resultado de:

    sum_hex([x.Unicode\
        .replace("U+","") for\
        x in [🏦, 🎲, 💻, 📚, 🚀]])

Atenção: sintaxe aproximadamente correta,
use apenas como motivação."
```

Banco Itaú recrutando Cientistas de Dados…