



# THE DEVELOPER'S CONFERENCE

**{Data Science} e  
{Machine Learning} na prática**

**Guilherme de Cleve Farto**

Framework/Engenharia – TOTVS Supply Agro

# Agenda

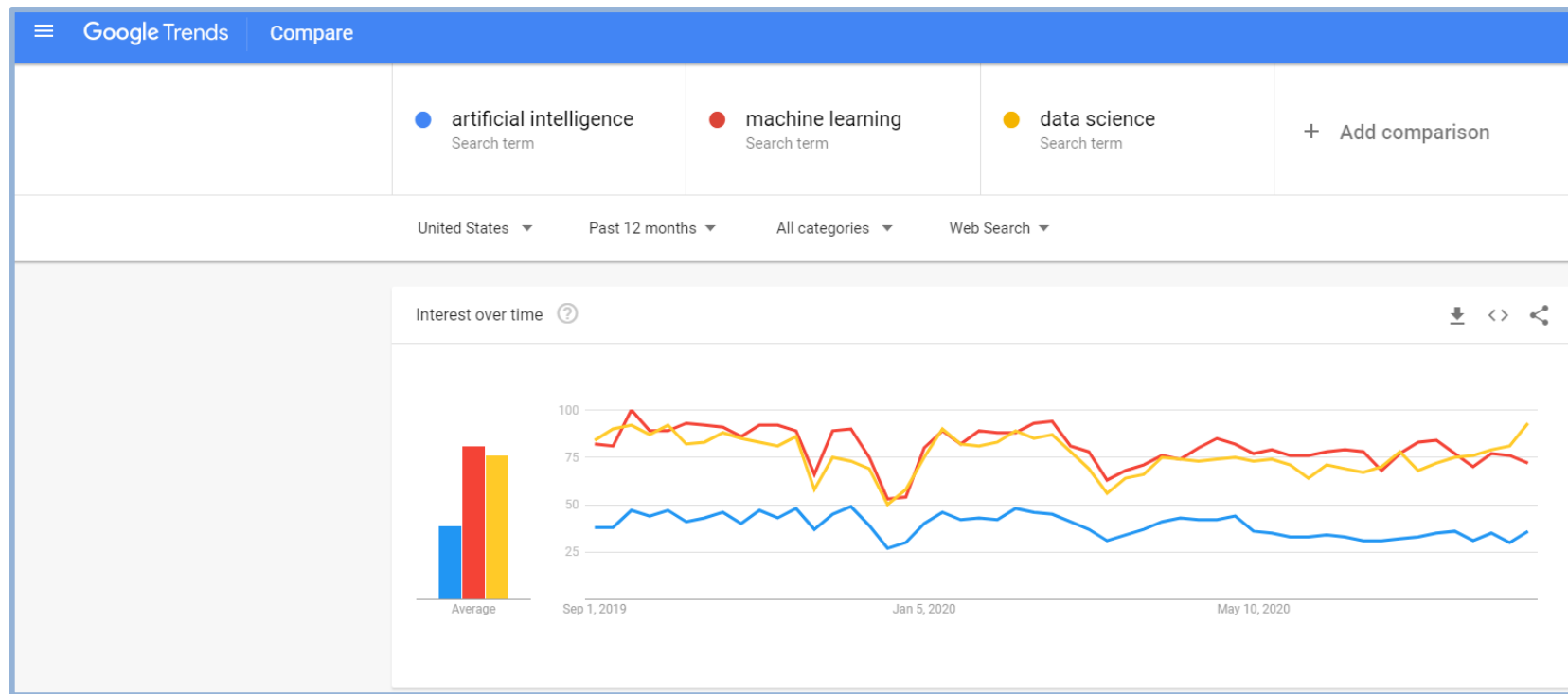


- *Trends*
- *IA e evolução*
- *Áreas relacionadas*
- *Data Science*
- *Plataformas*
- *Demonstração*
- *TOTVS e Carol*
- *Livros*
- *Próximos passos*

# Trends



THE  
DEVELOPER'S  
CONFERENCE

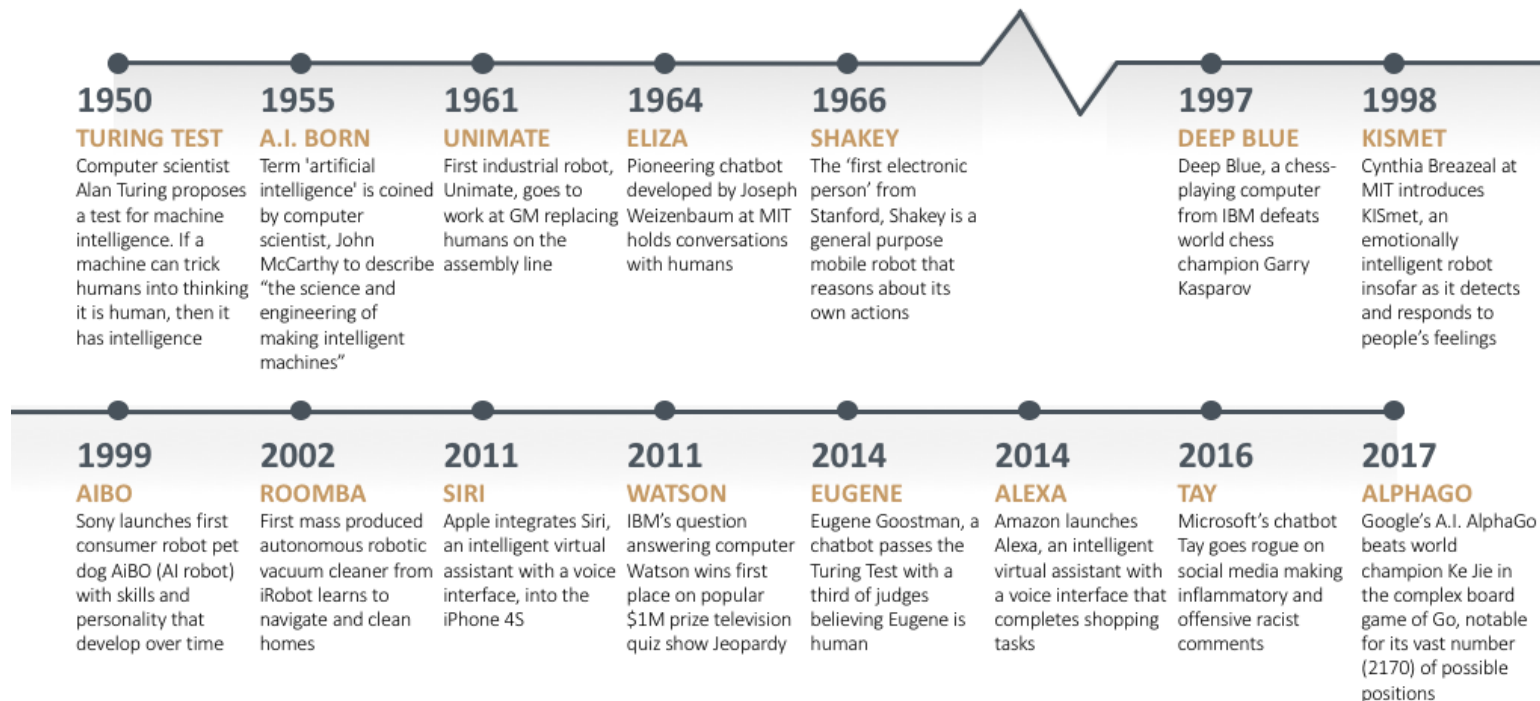


Fonte: Google Trends

# IA e evolução



THE  
DEVELOPER'S  
CONFERENCE



# Áreas relacionadas



## ***Artificial Intelligence***

Área que estuda como habilitar (ou imitar/simular) o comportamento humano em máquinas

## ***Machine Learning***

Subárea de IA que adota métodos e modelos estatísticos para habilitar ou melhorar a experiência de aprendizado de máquinas

## ***Deep Learning***

Subárea de IA e *Machine Learning* que explora técnicas avançadas de aprendizado de máquina por meio de redes multicamadas

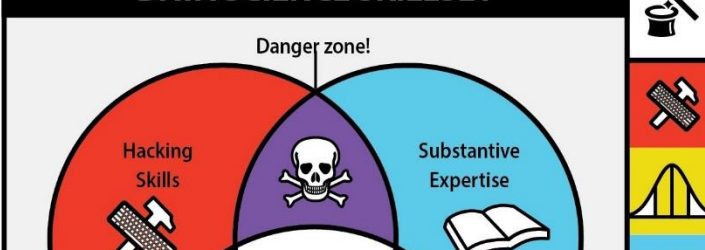
# Áreas relacionadas

## Data Science

Área interdisciplinar que

- Matemática e Estatística
- Ciência da Computação
- Domínio de Negócio

### DATA SCIENCE SKILLSET



Data science, due to its interdisciplinary nature, requires an intersection of abilities: **hacking skills**, **math and statistics knowledge**, and **substantive expertise** in a field of science.

**Hacking skills** are necessary for working with massive amounts of electronic data that must be acquired, cleaned, and manipulated.

**Math and statistics knowledge** allows a data scientist to choose appropriate methods and tools in order to extract insight from data.

**Substantive expertise** in a scientific field is crucial for generating motivating questions and hypotheses and interpreting results.

**Traditional research** lies at the intersection of knowledge of math and statistics with substantive expertise in a scientific field.

**Machine learning** stems from combining hacking skills with math and statistics knowledge, but does not require scientific motivation.

**Danger zone!** Hacking skills combined with substantive scientific expertise without rigorous methods can beget incorrect analyses.

Harvard  
Business  
Review



October 2012 Issue

DATA

## Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

**W**hen Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

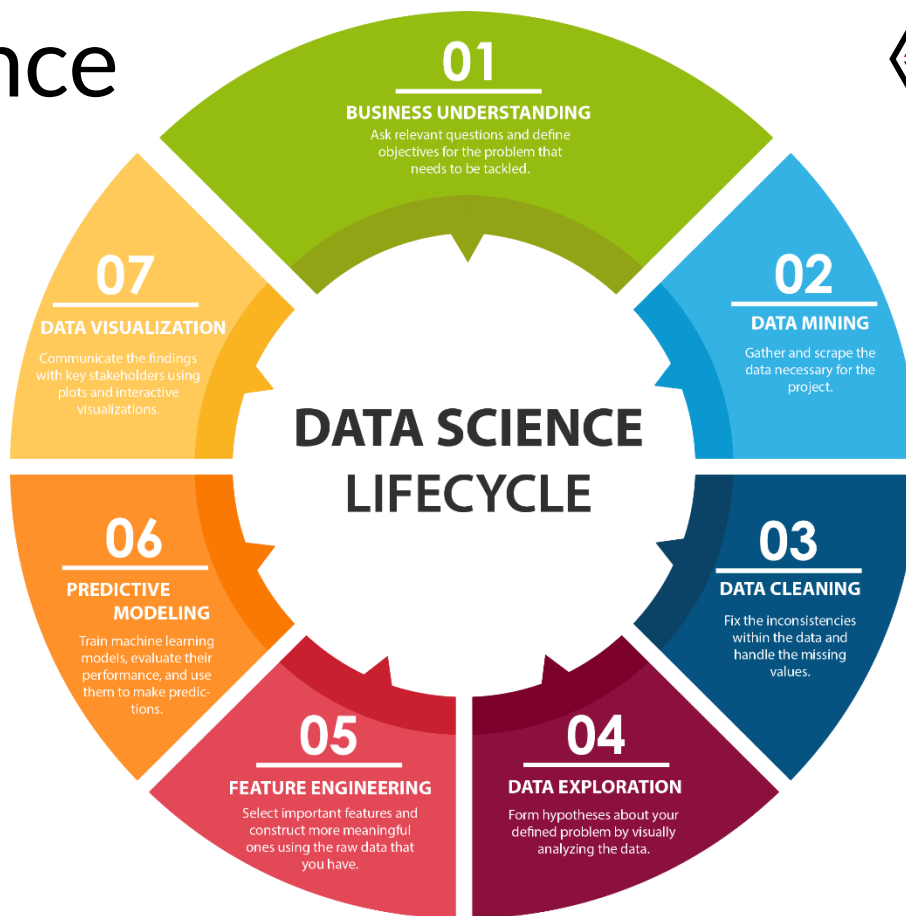
Fonte: <https://www.quora.com/e>  
<https://towardsdatascience.com/>

# Data Science

## *Ciclo de vida*



THE  
DEVELOPER'S  
CONFERENCE



Fonte: <https://addiai.com/>

# Plataformas



- **Python**
- Análise de dados
  - **Pandas**
- Matemática e estatística
  - **NumPy, SciPy e statsmodels**
- Aprendizado de máquina e algoritmos
  - **scikit-learn**
- Visualização de dados
  - **Matplotlib, Seaborn e Plotly**



# Plataformas

## TIOBE Index for August 2020

August Headline: Holiday season in the programming language world

Aug 2020	Aug 2019	Change	Programming Language	Ratings	Change
1	2	▲	C	16.98%	+1.83%
2	1	▼	Java	14.43%	-1.60%
3	3		Python	9.69%	-0.33%
4	4		C++	6.84%	+0.78%
5	5		C#	4.68%	+0.83%
6	6		Visual Basic	4.66%	+0.97%
7	7		JavaScript	2.87%	+0.62%
8	20	▲▲	R	2.79%	+1.97%
9	8	▼	PHP	2.24%	+0.17%
10	10		SQL	1.46%	-0.17%

# Plataformas



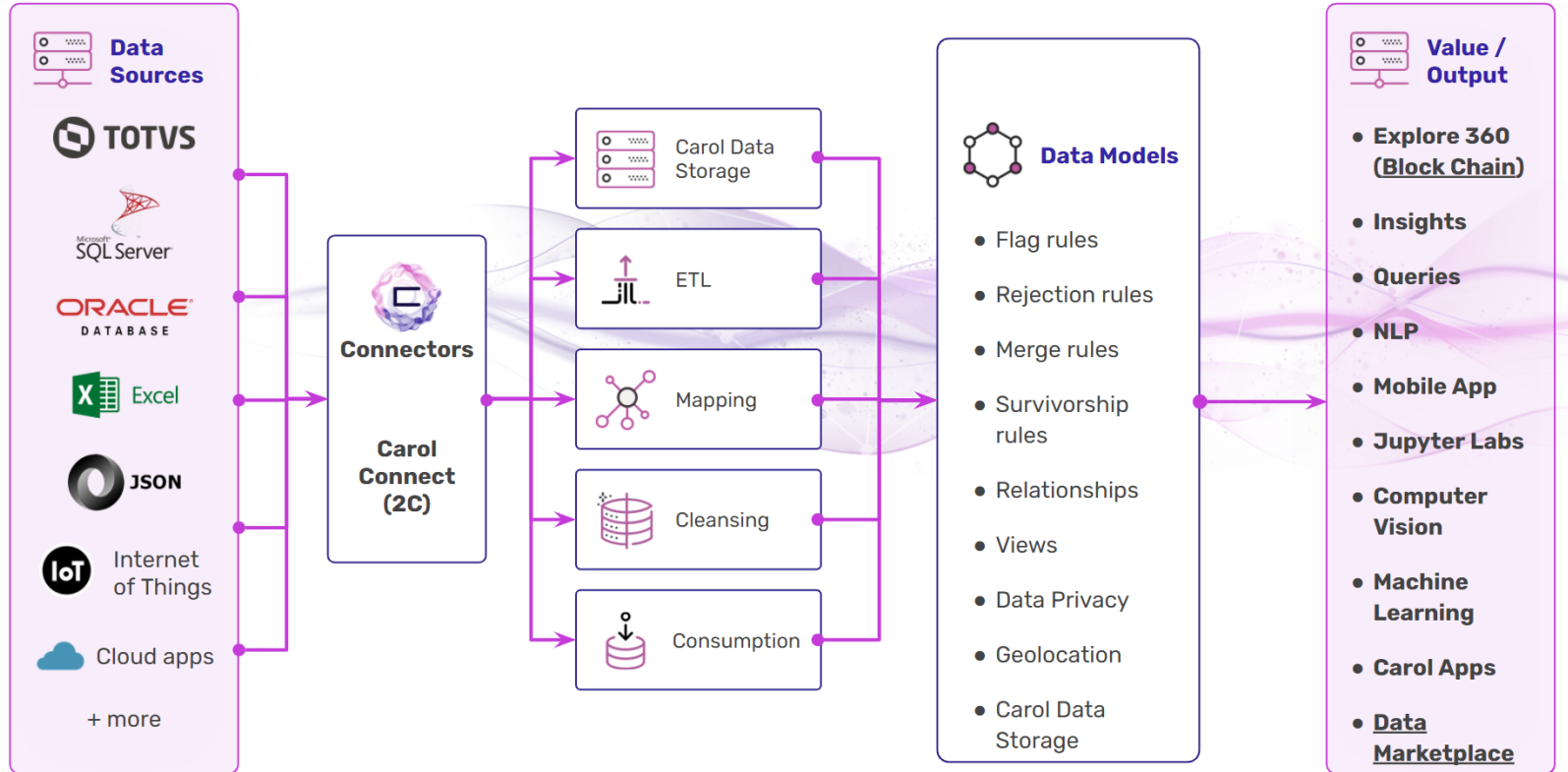
## ➤ Anaconda 3

- Versionamento Python
- Módulos e pacotes (dependências)
- Ferramentas/IDEs: JupyterLab e Notebook
- Desenvolvimento de plugins e aceleradores

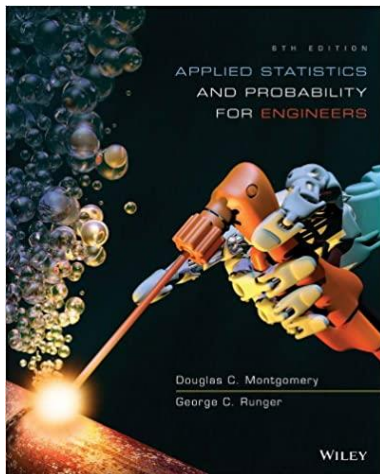
#2

# Demonstração

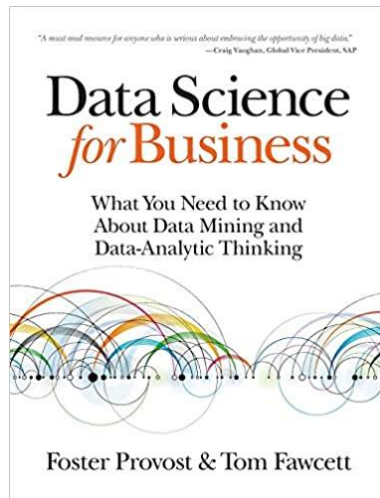
# TOTVS e Carol



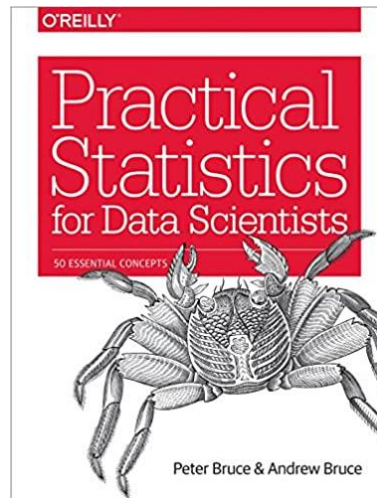
# Livros



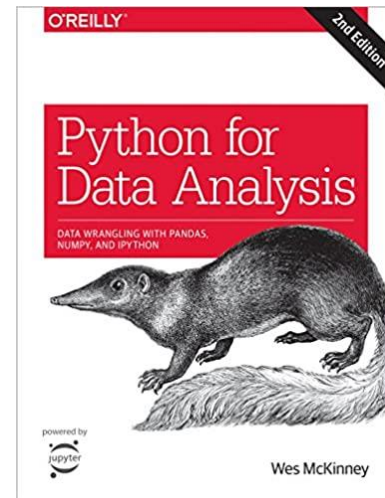
**Applied Statistics and Probability for Engineers**  
Douglas C. Montgomery  
George C. Runger



**Data Science for Business**  
Foster Provost  
Tom Fawcett



**Practical Statistics for Data Scientists: 50 Essential Concepts**  
Peter Bruce  
Andrew Bruce



**Python for Data Analysis**  
Wes McKinney

# Próximos passos



- Kaggle
- Google Dataset Search
- Automated Machine Learning (AutoML)
- Data + Dev + Ops = Data DevOps
- Apache Airflow: workflows de tarefas baseados em grafos;
- Luigi: construção de pipelines complexos em batch;

# Referências



- DataCamp: <https://www.datacamp.com/>
- TOTVS e IA: <https://www.totvs.com/inteligencia-artificial/>
- TOTVS e Carol: <https://docs.carol.ai/>
- TOTVS Developers (Podcast): <https://developers.totvs.com/podcast/>
  - *TOTVS Developers #2 - Inteligência Artificial*
  - *TOTVS Developers #16 - TOTVS CAROL*

# Obrigado!



## Guilherme de Cleva Farto



[guilherme.farto@gmail.com](mailto:guilherme.farto@gmail.com)



<https://www.linkedin.com/in/guilherme-farto/>



<https://github.com/guilhermefarto>



<https://twitter.com/gclevea>





THE  
DEVELOPER'S  
CONFERENCE



# THE DEVELOPER'S CONFERENCE