

O Cientista de Dados

Agente Educacional

Sérgio M. Dias

Agenda

Big Data

Ciência de Dados

Competências do cientista de dados

O que os cientistas de dados fazem?



Big Data

A cada segundo:

100.000 *tweets* circulam

547 *websites* são criados

mais de 2 milhões *de pesquisas* (Google)

48h *de vídeos* são baixadas no YouTube

684.478 *itens* são compartilhados no Facebook...

Fonte:
<https://www.webpagefx.com/internet-real-time/>



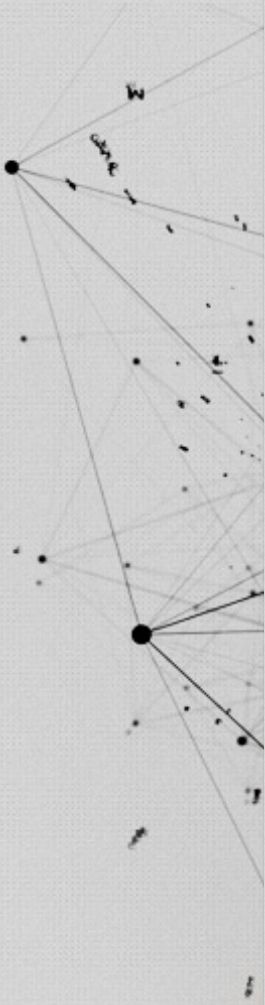
Big Data

Em governo (Brasil):

Mais de 7 milhões *de notas fiscais eletrônicas (NFe) por dia*

Mais de 16 bilhões *de NFe autorizadas...*

Fonte:
<http://www.nfe.fazenda.gov.br/portal/principal.aspx>



Como lidar com este "dilúvio" de dados?

A palavra mais importante no termo “ciência de dados” não é “dados”, mas “ciência”.



Fonte:
<http://www.economist.com/node/15579717>

Ciência de Dados

1 - A partir da necessidade de análise desse emaranhado de dados surgiu uma “nova” área da ciência, a chamada **ciência de dados**

- O “**quarto paradigma**” da ciência
- A profissão mais “**sexy**” do século 21
- Uma nova **buzzy word**!

...

Fontes:

Tansley, S.; Tolle, K.M. *The Fourth Paradigm: Data-intensive Scientific Discovery*. Microsoft Research.

<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>

www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/

...

2 - As atividades executadas pelo “cientista de dados”, em menor escala em relação ao volume de dados, são bastante antigas

Fontes:


Tansley, S.; Tolle, K.M. *The Fourth Paradigm: Data-intensive Scientific Discovery*. Microsoft Research.

<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>

www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/

O Cientista de Dados


Por que Ciência de Dados?


 **WIRED**


Tell Your Kids to Be Data Scientists, Not Doctors


BUSINESS CULTURE DESIGN GEAR SCIENCE

SHARE

 SHARE

 TWEET

 COMMENT

 EMAIL

PARTNER CONTENT LINDA BURTCH. BURTCH WORKS

TELL YOUR KIDS TO BE DATA SCIENTISTS, NOT DOCTORS




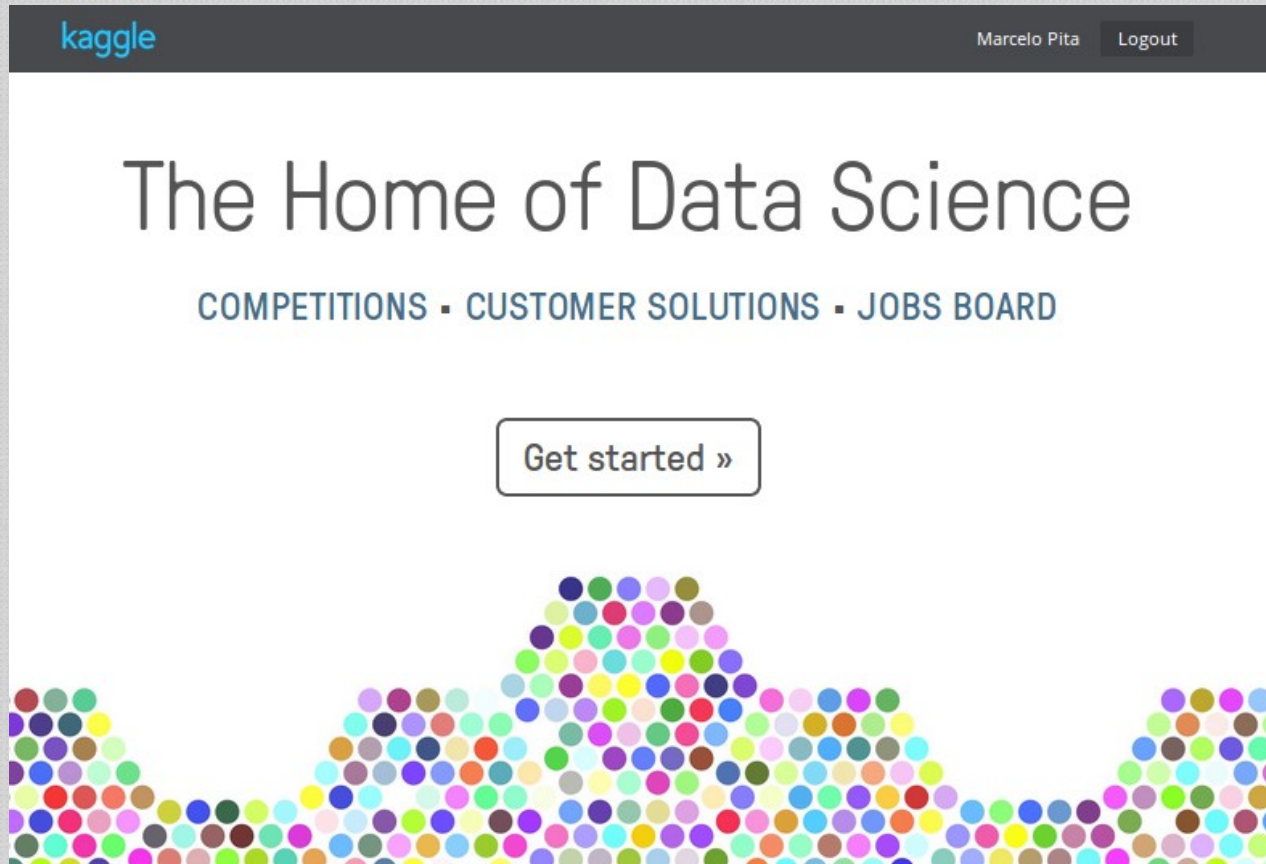
Image: ifindkarma/Flickr

Fonte:

<https://www.wired.com/insights/2014/06/tell-kids-data-scientists-doctors/>

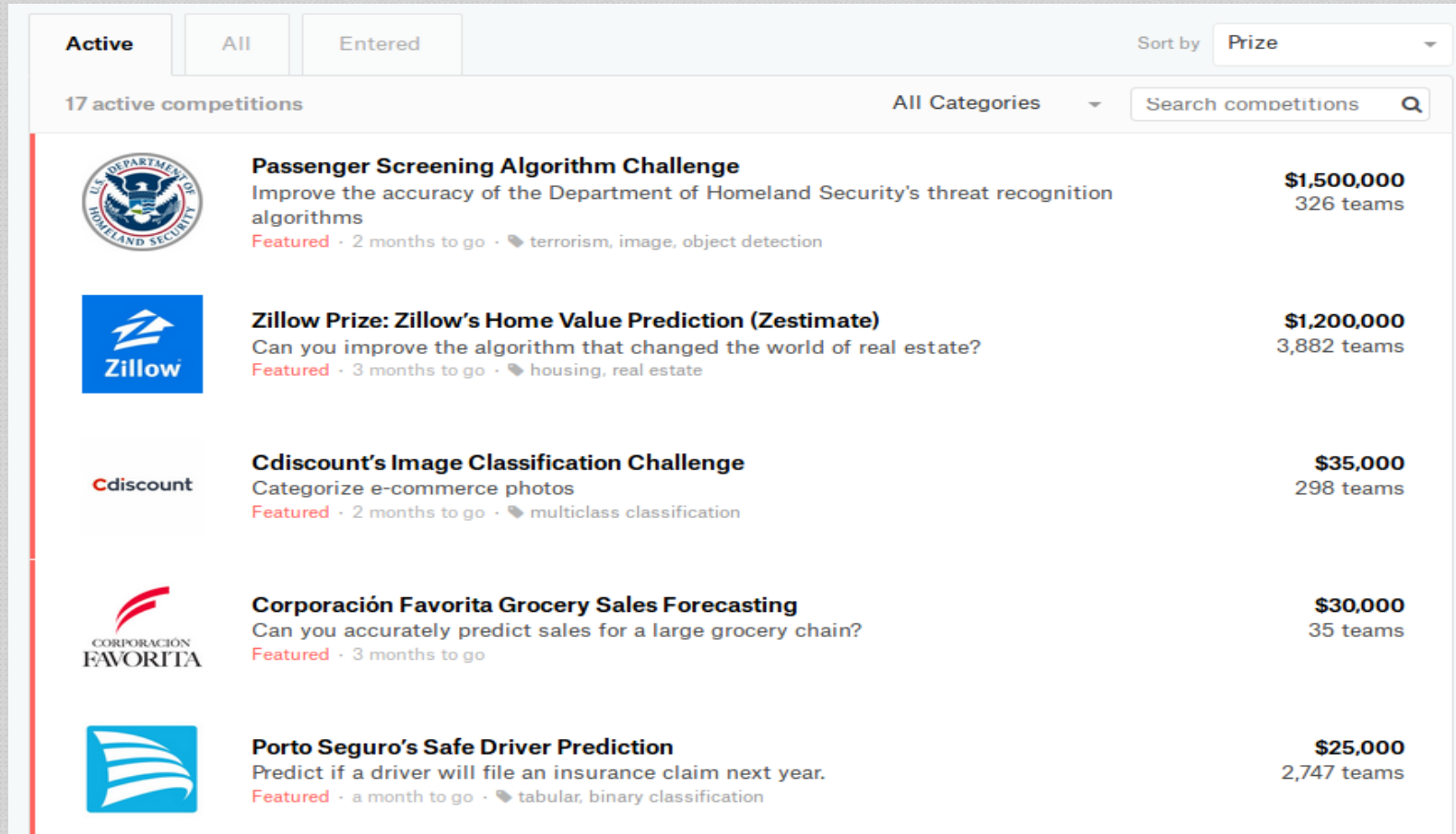
O Cientista de Dados

Por que Ciência de Dados?








O Cientista de Dados

Por que Ciência de Dados?



The screenshot shows the Kaggle competitions page with filters for 'Active' competitions, sorted by 'Prize'. It lists five competitions with their respective logos, titles, descriptions, prize amounts, and team counts.

Competition Logo	Competition Title	Description	Prize	Teams
	Passenger Screening Algorithm Challenge	Improve the accuracy of the Department of Homeland Security's threat recognition algorithms <i>Featured</i> · 2 months to go · terrorism, image, object detection	\$1,500,000	326 teams
	Zillow Prize: Zillow's Home Value Prediction (Zestimate)	Can you improve the algorithm that changed the world of real estate? <i>Featured</i> · 3 months to go · housing, real estate	\$1,200,000	3,882 teams
	Cdiscount's Image Classification Challenge	Categorize e-commerce photos <i>Featured</i> · 2 months to go · multiclass classification	\$35,000	298 teams
	Corporación Favorita Grocery Sales Forecasting	Can you accurately predict sales for a large grocery chain? <i>Featured</i> · 3 months to go	\$30,000	35 teams
	Porto Seguro's Safe Driver Prediction	Predict if a driver will file an insurance claim next year. <i>Featured</i> · a month to go · tabular, binary classification	\$25,000	2,747 teams

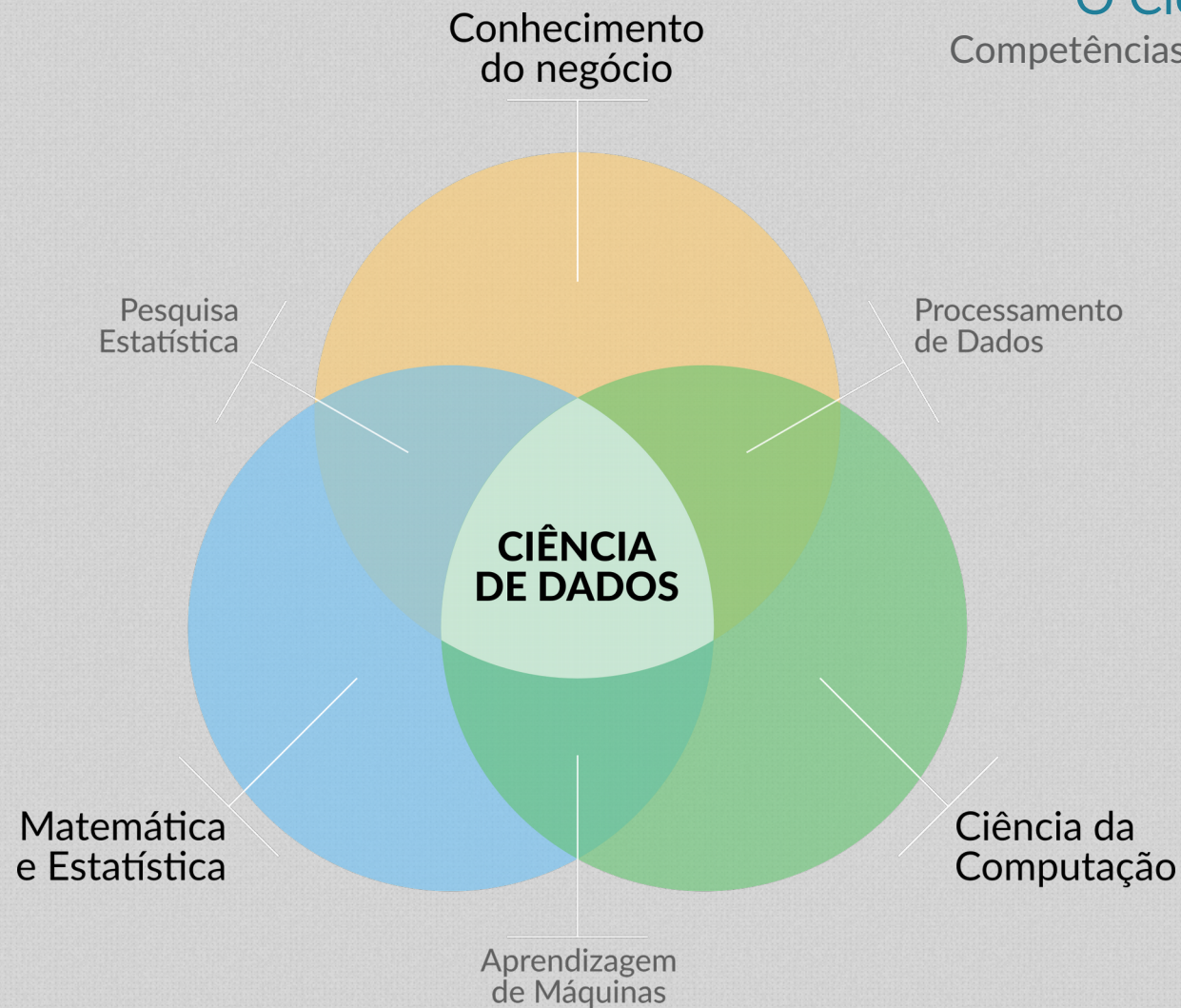


Competências do cientista de dados



O Cientista de Dados

Competências do cientista de dados





O que os cientistas de dados fazem?

“Considerando: Processos | Ambientes | Projetos”

O Cientista de Dados

O que os cientistas de dados fazem?

- **Definem** hipóteses e perguntas
- **Definem** os conjuntos de dados ideais
- **Determinam** que dados podem ser acessados
- **Adquirem** os dados
- **Pré-processam** os dados
- **Realizam** análise de dados exploratória
- ...

O Cientista de Dados

O que os cientistas de dados fazem?

...

- Realizam modelagem estatística dos dados
- Interpretam resultados de análises
- Escrevem relatórios sobre os resultados
- Criam modelos/componentes/códigos reusáveis
- Compartilham modelos e resultados com outras pessoas



Obrigado!

Agente Educacional

Sérgio M. Dias

sergio.dias@serpro.gov.br | #31 6539

Demais agentes educacionais sobre o assunto

Marcelo Pita | marcelo.pita@serpro.gov.br | #81 8794

Gustavo Torres | gustavo.gamatorres@serpro.gov.br | #31 6950