



JÁ VAI COMEÇAR MENTORIA DE EMPREGABILIDADE: A VISÃO DO RECRUTADOR

ENQUANTO ISSO...

- **Escolha um lugar confortável** para você sentar e se acomodar
- **Que tal pegar um snack** para matar a fome, **uma água**, um chá
- **Abra o chat, envie um “olá”** e #sentimento de como chega
- Que tal pegar **caderno e caneta para anotações**

T

**QUE BOM
QUE VOCÊ
VEIO!**



T



CHECK-IN & WARMUP



Allan Dieguez
AI Researcher / DS Manager

Head of Data Science

luiiza labs

E-Mail: allandieguez@gmail.com
Slack: [@allandieguez](#)
LinkedIn: <https://www.linkedin.com/in/allan-dieguez-a05a741b>

T

B.S.E Electronics and Computational
Engineering



M.Sc. Computer Science / Numerical
Optimization



Computer
Vision

Natural
Language
Processing

Geolocation &
Navigation

Customer
Value

Speech
Recognition

Dynamic
Pricing

AGENDA

- **Bloco 1: Processo Seletivo Na Visão do Recrutador**
 - + Intervalo - 5 min
- **Bloco 2: Construindo Seu Storytelling**
 - + Intervalo - 5 min
- **Bloco 3: Escolhendo a Empresa Mais Adequada**

O QUE VOCÊ ESPERA DA
AULA DE HOJE?



Processo Seletivo Na Visão do Recrutador



O PROCESSO SELETIVO



Missão: contratar o candidato com o **melhor fit com a vaga** e no **melhor momento de ambos** (a empresa e a pessoa).

Sucesso X Insucesso em um Processo Seletivo

- ✓ Alinhamento de **expectativas**
- ✓ **Senioridade** compatível com o **desafio da empresa**
- ✓ **Salário** compatível com **valor esperado** de entregas
- ✓ Mapeamento inicial da **evolução de carreira**
- ✗ Expectativas **ocultas**
- ✗ **Senioridade incompatível** com o **desafio da empresa**
- ✗ **Salário incompatível** com **valor esperado** de entregas
- ✗ Evolução da carreira **mal mapeada**

Etapas comuns de um Processo Seletivo

TRIAGEM DE CANDIDATOS

100%

Separar candidatos com as **skills certas**

ENTREVISTAS DE FIT CULTURAL

~25%

Separar candidatos de perfil **compatível com o da equipe**

ENTREVISTAS TÉCNICAS

~10%

Entender o quanto o candidato sabe **teoricamente** e observar **como ele se comunica** em um contexto de alta pressão

DESAFIO TÉCNICO

~5%

Verificar a **capacidade de solução** de um problema real, a **qualidade da documentação e codificação** e as **soft skills** apresentando a solução

T

PRIMEIRO CONTATO

A Triagem dos Candidatos



Fulano de Tal - **Cientista de Dados**

Sobre mim:

O que eu **busco**, o que eu **faz**,
qual a minha meta de vida
pessoal/**profissional**

Experiência:

Cientista de Dados Pleno em XYZ

Cientista de Dados Jr em ABC

Educação:

Graduação em <área de origem>

Mestrado/PhD em <área de especialização>

Skills:

Python

C/C++

Banco de Dados

Machine Learning

Estatística

SQL

Análise de Dados

Java

Projetos:

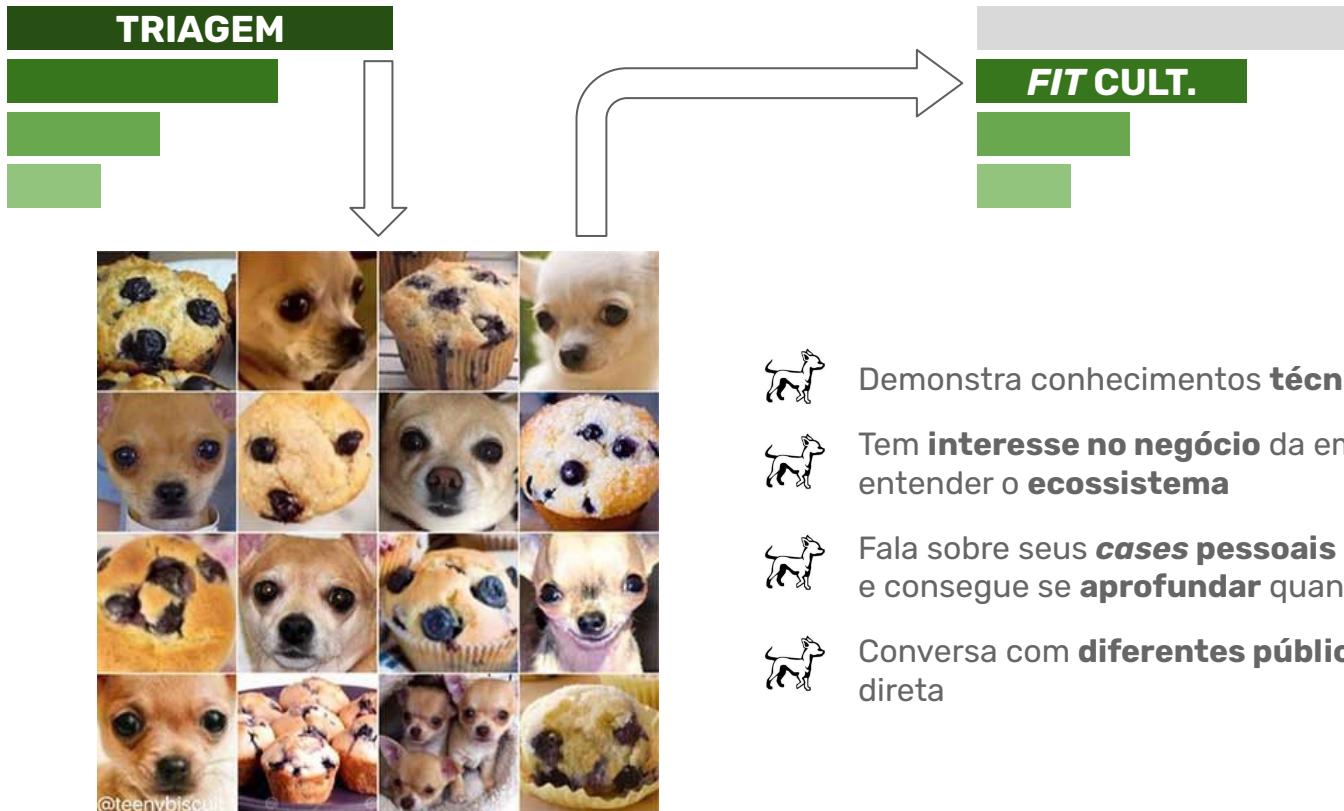
Site super complexo ponta a ponta

Análise de dados do tráfego em SP

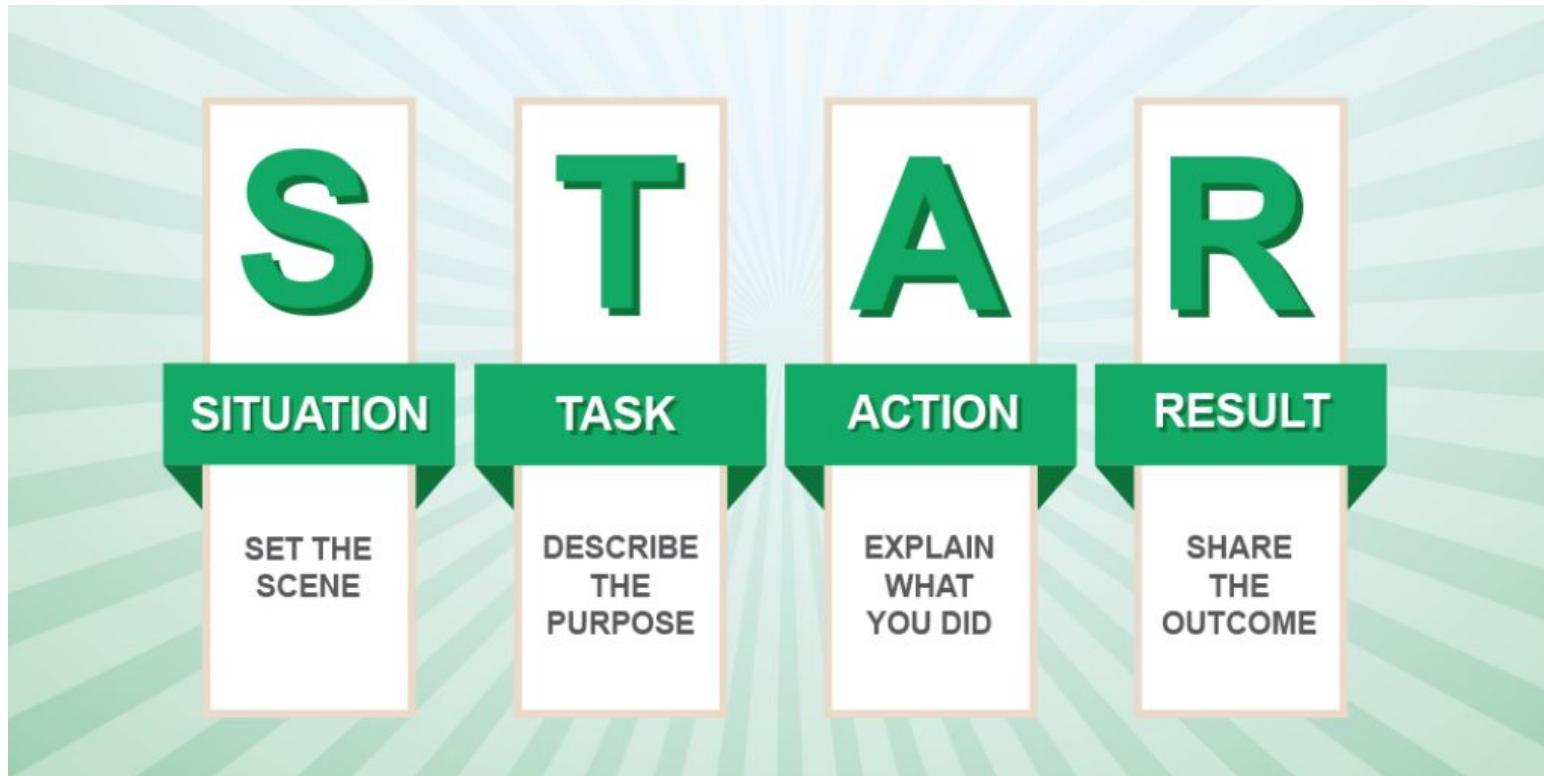
App mega útil para iphone

Sistema de Previsão de IBOVESPA

Primeiro contato: a **entrevista de fit cultural**



Método STAR



Explique seus **projetos** e **cases pessoais**

- Escolha um projeto com **aprendizados**
- Explique de forma sucinta, caprichando no **storytelling**
- O **método STAR** ajuda bastante aqui
- Fique de **olho no tempo**
- Atenção nas **perguntas do time**

PREVISÃO DE DEMANDA

Situação (Situation):

Devido a sazonalidade as demandas de vendas nunca eram acuradas.

Tarefa (Task):

Objetivo como cientista de dados é reduzir o erro das previsões em 30%.

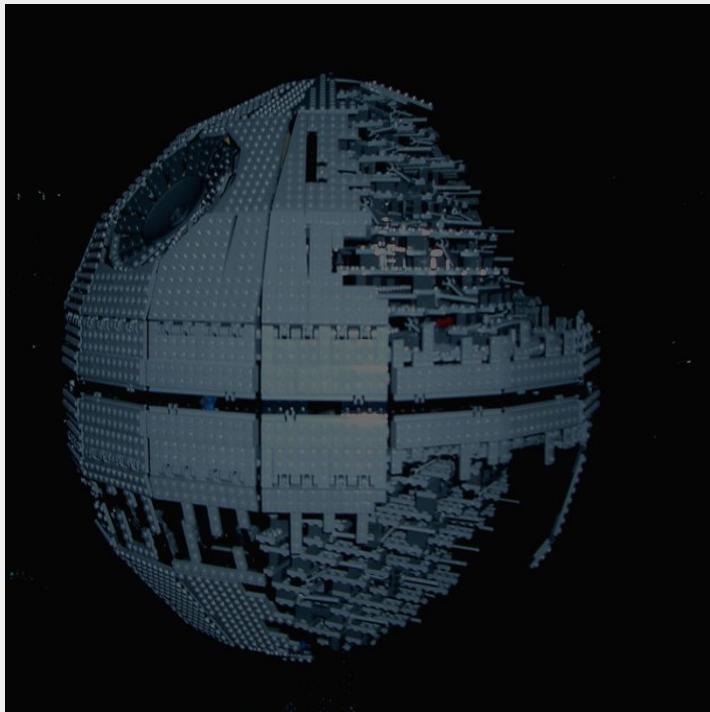
Ação (Action):

Desenvolvimento de um modelo de machine learning com séries temporais para prever a demanda.

Resultado (Result):

Redução do erro da previsão de demanda em 60% em comparação com o modelo anterior.

Mão na Massa: Método **S.T.A.R.**



- Dividir em grupos (~**5 pessoas**)
- Tempo de trabalho: **10 minutos**
- Escolher um *case* **por grupo** e descrever usando o **método STAR**
- Caso não tenham um *case* pessoal que queiram dividir, podem usar algum conhecido.
- Sugestão de fonte de cases de ML/DS: [Blog de Competições do Kaggle](#)

T

CONHECIMENTO TÉCNICO

T

Entrevistas Técnicas ou *Whiteboard*



Solução de **cases** e perguntas teóricas

- Tente **entender a pergunta** antes de responder
- Pergunte detalhes** até estar satisfeito
- Raciocine em voz alta**, deixe seu processo ser percebido
- Escreva, desenhe, rabisque
- Mais de uma solução é bom, **se pelo menos uma funciona**

ENTREVISTA MICROSOFT

- Qual a definição de p-valor? Como explicar o p-valor para clientes?
- Como identificar se uma nova observação é um *outlier*?
- Qual a diferença entre um *boxplot* e um histograma?
- Qual a diferença entre regressão *lasso* e *ridge*?
- O que é dilema viés-variância?
- Como definir o número de *clusters* adequado em uma análise?

Solução de problemas no *Live Coding*

- ❑ Mesmo racional de antes, comunique-se **o tempo inteiro** e pergunte **o quanto for necessário**
- ❑ **Protótipo antes** de escrever o código, corrija depois o que precisar
- ❑ Teste o raciocínio de tempos em tempos, faça **debug ao vivo**

PROGRAME EM PYTHON

Escreva um programa que some o vetor (1, 4, 2) com o vetor (4, 5, 8):

```
import numpy as np  
  
vetor1 = np.array([1, 4, 2])  
vetor2 = np.array([4, 5, 8])  
vetor3 = vetor1 + vetor2  
  
print('vetor3')
```

Desafio (ou *case*) técnico: **metodologia**

```
LICENSE           <- Makefile with commands like 'make data' or 'make train'  
Makefile          <- The top-level README for developers using this project.  
README.md  
data  
  external       <- Data from third party sources.  
  interim        <- Intermediate data that has been transformed.  
  processed      <- The final, canonical data sets for modeling.  
  raw            <- The original, immutable data dump.  
  
docs             <- A default Sphinx project; see sphinx-doc.org for details  
  
models           <- Trained and serialized models, model predictions, or model summaries  
  
notebooks         <- Jupyter notebooks. Naming convention is a number (for ordering),  
                   the creator's initials, and a short '-' delimited description, e.g.  
                   '1.0-jqp-initial-data-exploration'.  
  
references        <- Data dictionaries, manuals, and all other explanatory materials.  
  
reports           <- Generated analysis as HTML, PDF, LaTeX, etc.  
  figures         <- Generated graphics and figures to be used in reporting  
  
requirements.txt  <- The requirements file for reproducing the analysis environment, e.g.  
                   generated with 'pip freeze > requirements.txt'  
  
setup.py          <- Make this project pip installable with `pip install -e`  
src  
  __init__.py     <- Source code for use in this project.  
                  <- Makes src a Python module  
  
  data            <- Scripts to download or generate data  
    make_dataset.py  
  
  features         <- Scripts to turn raw data into features for modeling  
    build_features.py  
  
  models           <- Scripts to train models and then use trained models to make  
    |               predictions  
    |   predict_model.py  
    |   train_model.py  
  
  visualization  <- Scripts to create exploratory and results oriented visualizations  
    visualize.py  
  
tox.ini          <- tox file with settings for running tox; see tox.readthedocs.io
```

- Crie **hipóteses iniciais** sobre o problema para validar durante a solução
- Estruture a solução em etapas** para facilitar a leitura. Sugestão de etapas:
 - EDA
 - Modelagem
 - Avaliação da Solução
- Encontre um **benchmark** ou crie um **modelo baseline** inicialmente
- Versione o projeto em uma estrutura como a do **cookiecutter de data science**
- Registre seus *insights* e descobertas, assim como a sua conclusão em um **relatório técnico** ou um **diário de bordo**

Desafio (ou *case*) técnico: **apresentação**

- Estruture a apresentação de forma a **caber no tempo limite**
- Inicie com o enunciado e sua interpretação, trazendo **todos para a mesma página**
- Use ao máximo o **storytelling** para **descrever a solução** e suas descobertas
- Dê **prioridade aos resultados e descobertas**, só entre nos detalhes técnicos quando precisar
- Ouça com atenção** às dúvidas e *feedbacks* e **mantenha o aprendizado de todos na sala** como o foco principal



INTERVALO DE 5 MINUTOS



Construindo Seu *Storytelling*



**SE PREPARANDO PARA
O PROCESSO SELETIVO**

Antes de se candidatar



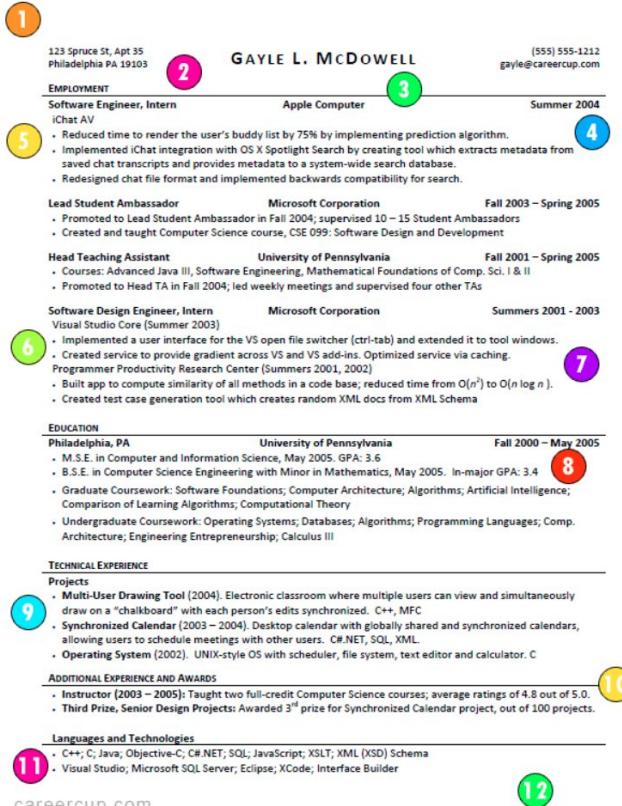
1. Entenda seus **pontos fortes e fracos** e sua **senioridade em relação à vaga**
2. Entenda os **tipos de problemas** que você gosta de resolver
3. Entenda mais **sobre a empresa** e seu nível de **maturidade *data-driven***
4. Prepare seu **currículo** e seu **portfólio**

Quem pode ser um Cientista de Dados?



- Engenheiros
- Cientistas da Computação
- Estatísticos
- Físicos
- Químicos
- Biólogos
- Advogados
- Jornalistas
- ...

Qual a impressão inicial que você quer passar?



Employment

Data Scientist **Loggi** **July 2016 - Present**

Driver Analytics

Improvements to the company's main platform regarding drivers' behavior.

- Designed and implemented a system for detecting delivery frauds using geolocation tracking.
- Designed and implemented a behavior classifier to detect when a driver deviates from the route.
- Designed and implemented an ETA predictor for an unassigned route and a set of candidate drivers.

Data Scientist **B2W** **November 2015 – July 2016**

Product Matching Platform for Marketplace

Port of Sieve's automatic product database matching platform to B2W, the company's current owner.

- Refactored product matching platform to handle massive datasets (~3M products); it currently processes up to 100 times more data consuming the same RAM and runs 20 times faster.
- Designed and implemented a product search platform prototype that learns automatically from user behavior, improving product matching accuracy on the run.
- Designed and implemented a product market structure estimator prototype that automatically improves its performance by processing data from semi structured databases.

Data Scientist **Big Data** **March 2015 – October 2015**

Consulting

Development and consulting projects for retail and marketing companies.

- Researched and developed a sales prediction system for a major retail client.
- Contributed on the research and development of a price optimization platform for retail clients.
- Analyzed the effect of specific TV advertisement on different groups of consumers.

Development

Development of support systems for extracting and processing external data.

- Developed several clusters based data extractors using AWS, StarCluster, Python and Bash scripts.
- Developed an event based system for automatic detection and correction of cluster node failure.
- Refactored spoken letters and numbers classifier pipeline and raised its F1 Score from 75% to 87%.

O 12 pontos do currículo orientado a entregas

1

One Page Resume: Recruiters do not read your resume; they do a 15 - 30 second "spot check" of your resume. When your resume is too long, it just takes your best stuff - the stuff that would have made the "one page cut" - and dilutes it with more mediocre content. Lengthy resumes do not make you more impressive, and [there are many other reasons to keep your resume short too](#). A good rule of thumb is to keep your resume to one page if you have less than 10 years of experience or *at most* two pages if you have more than 10 years of experience. And if you think you can't get your resume to just one page, trust me, [you can!](#) You just need to think about what is really important for a recruiter to see.

2

No Objectives: All an objective does is state, in a wordy way, what position you're interested in. The company already knows that because you applied for a particular position. At best, it'll just waste space. At worst, it'll limit you since it'll exclude other positions that might have been interesting to you.

3

Use a Resume Template with Columns: Unless you're great with design, you probably shouldn't be creating your own resume template. It'll most likely look sloppy. Use a template, and make sure it has multiple columns. Using three columns, for example, will allow you to put the company name, position, and date all on one line. This makes it easier to read *and* saves space.

4

Use Tables: If you're using Microsoft Word to create your resume (which you probably should), use Microsoft Word's "tables." Just make sure to hide the borders afterwards.

5

Short Bullets: Because resume screeners only spend 15 - 30 seconds on your resume, length bullets - anything that feels like a paragraph - just won't get read. Keep your bullets to one to two lines (with one line being better than two).

6

Accomplishment Oriented: Your bullets should focus on your *accomplishments* - that is, the impact you had - rather than your responsibilities. What did you build, create, design, optimize, lead, etc?

7

Quantify: Whenever possible, you should quantify your accomplishments. If you optimized something, by how much? If you won an award, out of how many people?

8

Resume: The general rule of thumb is to list your GPA if it's at least 3.0 or higher, but there are two important rules to know here: (1) You may choose to list your in-major GPA if it's higher than your overall GPA, but you need to specify that it's your in-major GPA. (2) If your school uses a different scale (such as a 10-point scale), you may want to convert your GPA to a 4.0-scale which will be more widely understood.

9

Projects: Most candidates should pick their top 3 - 5 projects to list on their resume. These can be academic required project or independent projects. They *do not* need to be completed or launched either. As long as you've done a "meaty" amount of work on them, that's good enough!

10

Additional Experience: You can put additional experience, like leadership activities or awards, in a section like this (changing the name of the section depending on what you list). Be careful here to focus on what really matters. If you're applying for a coding role, your role as an eagle scout in high school is probably not very important!

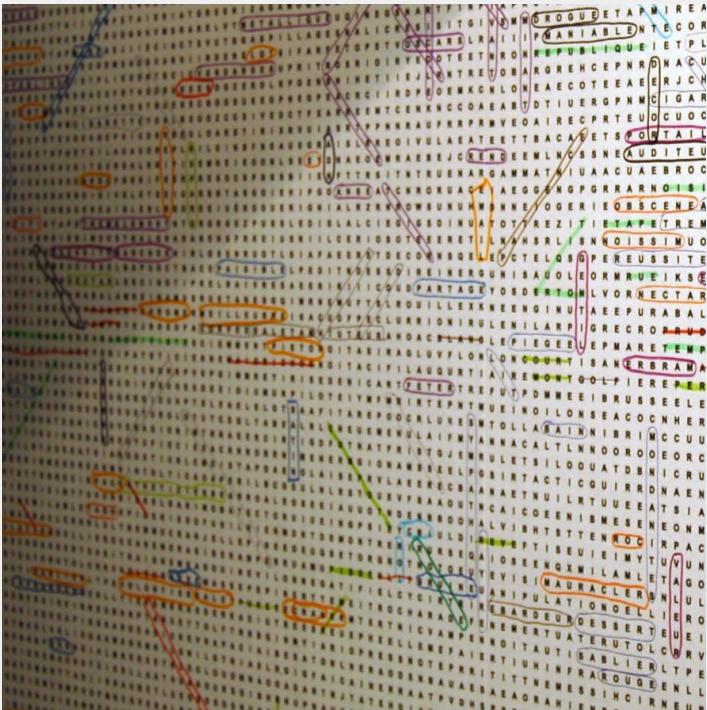
11

Languages and Technologies: It's a good idea to list your languages and technologies, but remember that anything you list here is "fair game" for the interviewer to test. If you want to list a language but you happen to be a bit rusty in it, consider listing it as something like: "C++ (Proficient), C# (Prior Experience), ..."

12

What did you NOT include?: Is there anything impressive / interesting that you've done that you left out? About 50% of candidates leave out an important project or other component of their experience because it wasn't finished / "official" / etc. If you've done it, and it's impressive enough to "make the cut" (you shouldn't just list *everything* you've ever done), then it belongs on your resume!

Mão na Massa: Construção de **Currículo**



- ❑ Dividir em grupos (**-5 pessoas**)
 - ❑ Tempo de trabalho: **10 minutos**
 - ❑ Escolher um currículo **por grupo** e reescrever no formato mostrado em sala.
 - ❑ Caso não tenham um currículo pronto, podem usar algum conhecido.
 - ❑ Sugestão de fonte de currículos: [Busca direcionada no Linkedin](#)

O QUE VOCÊ
REALMENTE SABE?

T

O que o seu **portfólio** diz sobre você?



**PROJETOS QUE VOCÊ
TRABALHOU OU QUE
ESTÁ TRABALHANDO**



**SUAS HABILIDADES DE
ENTENDIMENTO E
RESOLUÇÃO DE
PROBLEMAS**



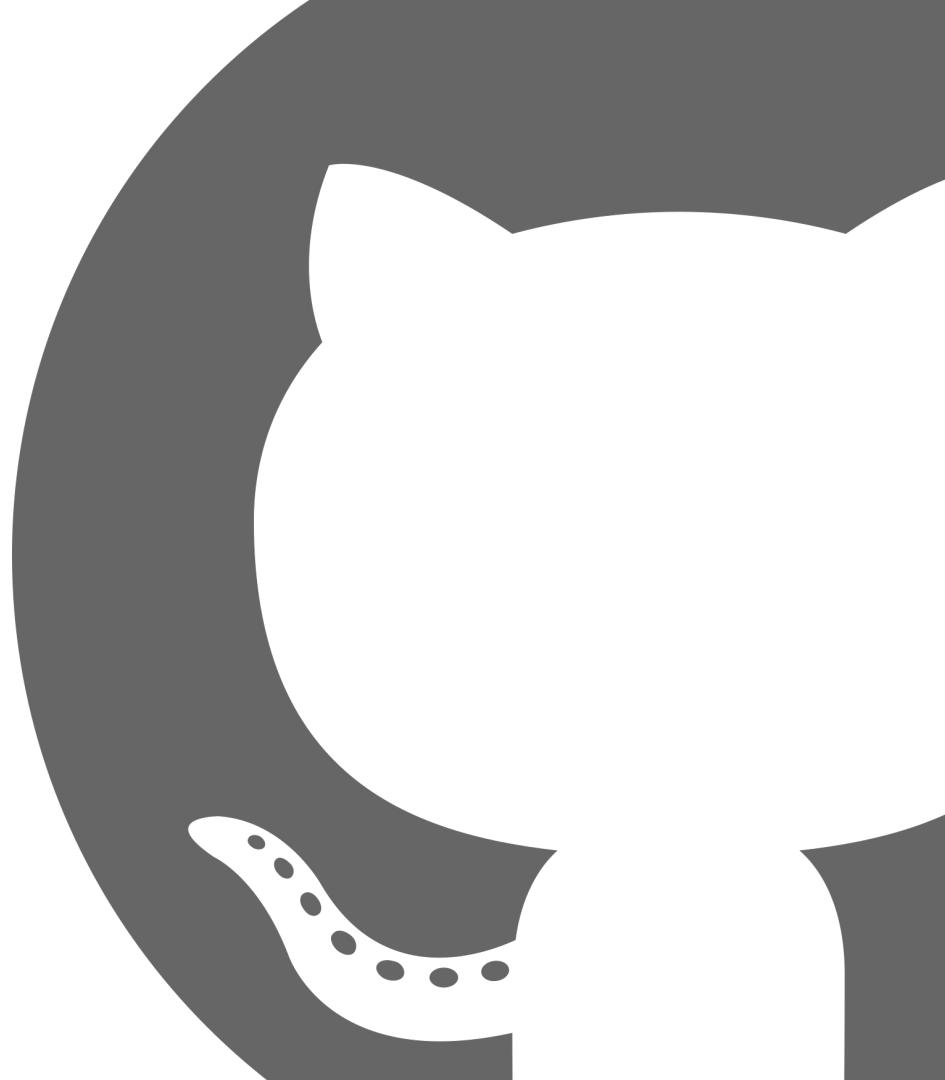
**TECNOLOGIAS E
LINGUAGENS QUE
VOCÊ TEM
CONHECIMENTO**



**SUAS CONTRIBUIÇÕES
PARA A COMUNIDADE DE
CIÊNCIA DE DADOS**

GITHUB

O quanto você sabe sobre
desenvolvimento?



T

GITHUB

**Siraj Raval**

II Sourcell

[Unfollow](#)

subscribe to my youtube channel!
www.youtube.com/c/sirajraval

San Francisco, CA

www.youtube.com/c/sirajraval

[Block or report user](#)

[Overview](#)[Repositories 430](#)[Projects 0](#)[Stars 20](#)[Followers 36.3k](#)[Following 4](#)

Popular repositories

[Learn_Machine_Learning_in_3_Months](#)

This is the code for "Learn Machine Learning in 3 Months" by Siraj Raval on Youtube

7.3k 2.4k

[learn_math_fast](#)

This is the Curriculum for "How to Learn Mathematics Fast" By Siraj Raval on Youtube

Python 2.9k 589

[Learn_Deep_Learning_in_6_Weeks](#)

This is the Curriculum for "Learn Deep Learning in 6 Weeks" by Siraj Raval on Youtube

2.6k 890

[Learn_Data_Science_in_3_Months](#)

This is the Curriculum for "Learn Data Science in 3 Months" By Siraj Raval on Youtube

2.4k 999

[Learn_Computer_Science_in_5_Months](#)

This is the Curriculum for "Learn Computer Science in 5 Months" By Siraj Raval on Youtube

1.7k 560

[YOLO_Object_Detection](#)

This is the code for "YOLO Object Detection" by Siraj Raval on Youtube

Python 1.5k 762

T

KAGGLE

O quanto você é bom em
solucionar problemas?



T

KAGGLE

**Mario Filho**

São Paulo, State of São Paulo, Brazil

Joined 7 years ago · last seen 3 days ago

<http://www.mariofilho.com/>Competitions
Grandmaster

Followers 657

Following 10

[Home](#)[Competitions \(32\)](#)[Notebooks \(3\)](#)[Discussion \(24\)](#)[Followers \(657\)](#)[Contact User](#)[Follow User](#)Competitions
Grandmaster

Current Rank

658
of 137,604

Highest Rank

12



7



11



6

Telstra Network ...

4 years ago
Top 1%1st

of 973

Caterpillar Tube ...

5 years ago
Top 1%1st

of 1323

Avito Duplicate ...

4 years ago
Top 1%3rd

of 548

Datasets
Contributor

Unranked



0



0



0

No dataset results

Notebooks
Contributor

Unranked



0



0



1

Discussion
Contributor

Unranked



1



0



16

live26 - https://...

20 days ago

16

votes

Top 100 Users b...

5 years ago

4

votes

Date Exploratio...

4 years ago

2

votes

Progression Sys...

4 years ago

15

votes

What was your ...

5 years ago

7

votes

Are your top sol...

6 years ago

6

votes

LINKEDIN

O que você **andou fazendo**
e qual o seu ***network***?



T

LINKEDIN



[Enviar mensagem](#)

[Mais...](#)

Jose Borbolla Neto · 1º

Data Science & Analytics Coordinator

São Paulo, São Paulo, Brasil · + de 500 conexões ·

[Informações de contato](#)



Tera

T

MEDIUM

O quanto você **comunica bem** idéias, conceitos e estudos?





A maior comunidade de Data Science do Brasil

[ÚLTIMOS POSTS](#)[ENGENHARIA](#)[CIÊNCIA](#)[ML](#)[PODCAST](#)[ESCREVA NO DH](#)[CONHEÇA O DATA HACKERS!](#)[Follow](#)

O dia a dia de um Machine Learning Engineer—Data Hackers Podcast 24

Saiba o que faz um Engenheiro de Machine Learning, seus desafios, e o que é preciso para começar na área



Paulo Vasconcellos
May 22 · 2 min read

T

arXiv.org

Qual a **sua profundidade** nos
assuntos que estudou?



Cornell University

We gratefully acknowledge support from the Simons Foundation and member institutions.

arXiv.org > cs > arXiv:2102.08941

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 10 Jan 2021]

Automatic Face Understanding: Recognizing Families in Photos

Joseph P Robinson

We built the largest database for kinship recognition. The data were labeled using a novel clustering algorithm that used label proposals as side information to guide more accurate clusters. Great savings in time and human input was had. Statistically, FiW shows enormous gains over its predecessors. We have several benchmarks in kinship verification, family classification, tri-subject verification, and large-scale search and retrieval. We also trained CNNs on FiW and deployed the model on the renowned KinWild I and II to gain SOTA. Most recently, we further augmented FiW with MM. Now, video dynamics, audio, and text captions can be used in the decision making of kinship recognition systems. We expect FiW will significantly impact research and reality. Additionally, we tackled the classic problem of facial landmark localization. A majority of these networks have objectives based on L1 or L2 norms, which inherit several disadvantages. The locations of landmarks are determined from generated heatmaps from which predicted landmark locations get penalized without accounting for the spread: a high scatter corresponds to low confidence and vice-versa. To address this, we introduced an objective that penalizes for low confidence. Another issue is a dependency on labeled data, which is expensive to collect and susceptible to error. We addressed both issues by proposing an adversarial training framework that leverages unlabeled data to improve model performance. Our method claims SOTA on renowned benchmarks. Furthermore, our model is robust with a reduced size: 1/8 the number of channels is comparable to SOTA in real-time on a CPU. Finally, we built BFW to serve as a proxy to measure bias across ethnicity and gender subgroups, allowing us to characterize FR performances per subgroup. We show performances are non-optimal when a single threshold is used to determine whether sample pairs are genuine.

Comments: PhD Thesis
Subjects: Computer Vision and Pattern Recognition (cs.CV)
Cite as: arXiv:2102.08941 [cs.CV]
(or arXiv:2102.08941v1 [cs.CV] for this version)

Submission history
From: Joseph Robinson [view email]
[v1] Sun, 10 Jan 2021 22:37:25 UTC (120,887 KB)

Bibliographic Tools Code Related Papers About arXivLabs

Bibliographic and Citation Tools

Bibliographic Explorer (What is the Explorer?)

Search... All fields Search
Help | Advanced Search

Download:

- PDF
- Other formats

Current browse context:
cs.CV
< prev | next >
new | recent | 2102
Change to browse by:
cs

References & Citations

- NASA ADS
- Google Scholar
- Semantic Scholar

Export BibTeX Citation

Bookmark



**PASSEI NO PROCESSO!
E AGORA?**

Depois da conquista da vaga, **o que fazer?**



- Nunca pare de estudar**, a indústria evolui todo dia
- DS é Vocaçao** e não Emprego, sua carreira te acompanha nas empresas
- Trabalhe sempre em projetos pessoais ou **colaborando em outros projetos**

Envolva-se com o **Ecossistema...** e desenvolva-o!

- ❑ Participe de *meetups* e eventos da área, **faça networking** e descubra o que o Mercado anda fazendo
- ❑ Palestre, **se exponha**, dê aulas, **esteja disponível** para mentorar novos cientistas
- ❑ **Fale com estranhos**, conecte-se a algum DS sênior no LinkedIn e peça um *feedback* honesto



INTERVALO DE 5 MINUTOS



Escolhendo a Empresa Mais Adequada



O MERCADO E A CIÊNCIA DE DADOS

Top 12 Data Science Related Roles (KD Nuggets, 2016)

1. DATA SCIENTIST

These people use their analytical and technical capabilities to extract meaningful insights from data.

SALARY: \$65,000 - \$110,000

2. DATA ENGINEER

They ensure uninterrupted flow of data between servers and applications and are also responsible for data architecture.

SALARY: \$60,0945 - \$124,635

3. BIG DATA ENGINEER

Big Data Engineers build the designs created by solutions architects. They develop, maintain, test and evaluate big data solutions within organizations.

SALARY: \$100,000 - \$165,000

4. MACHINE LEARNING SCIENTIST

They work in the research and development of algorithms that are used in adaptive systems. They build methods for predicting product suggestions and demand forecasting, and explore Big Data to automatically extract patterns.

SALARY: \$78,857 - \$124,597

5. BUSINESS ANALYTICS SPECIALIST

A business analytics specialist supports various development initiatives, assists in testing activities and in the development of test scripts, performing research in order to understand business issues, and developing practical cost-effective solutions to problems.

SALARY: \$50,861 - \$94,209

6. DATA VISUALIZATION DEVELOPER

They design, develop and provide production support of interactive data visualizations used across the enterprise. They possess an artistic mind that conceptualizes, design, and develop reusable graphic/data visualizations and uses strong technical knowledge for implementing these visualizations using the latest technologies.

SALARY: \$108,000 - \$130,000

7. BI ENGINEER

They have data analysis expertise and the experience of setting up reporting tools, querying and maintaining data warehouses. They are hands-on with big data and take a data driven approach to solving complex problems.

SALARY: \$96,710 - \$138,591

8. BI SOLUTION ARCHITECT

They come up with solutions quickly to help businesses in making time sensitive decisions, have strong communication & analytical skills, passion for data visualization, and a drive for excellence and self-motivation.

SALARY: \$107,000 - \$162,000

9. BI SPECIALIST

They are responsible for supporting an enterprise wide business intelligence framework. This position requires critical thinking, attention to detail, and effective communication skills.

SALARY: \$77,969 - \$128,337

10. ANALYTICS MANAGER

An analytics manager is responsible for configuration, design, implementation, and support of data analysis solution or BI tool. They are specifically required to analyze huge quantities of information gathered through transactional activity.

SALARY: \$83,910 - \$134,943

11. MACHINE LEARNING ENGINEER

Machine Learning engineer's final "output" is the working software, and their "audience" for this output consists of other software components that run autonomously with minimal human supervision. The decisions are made by machines and they affect how a product or service behaves.

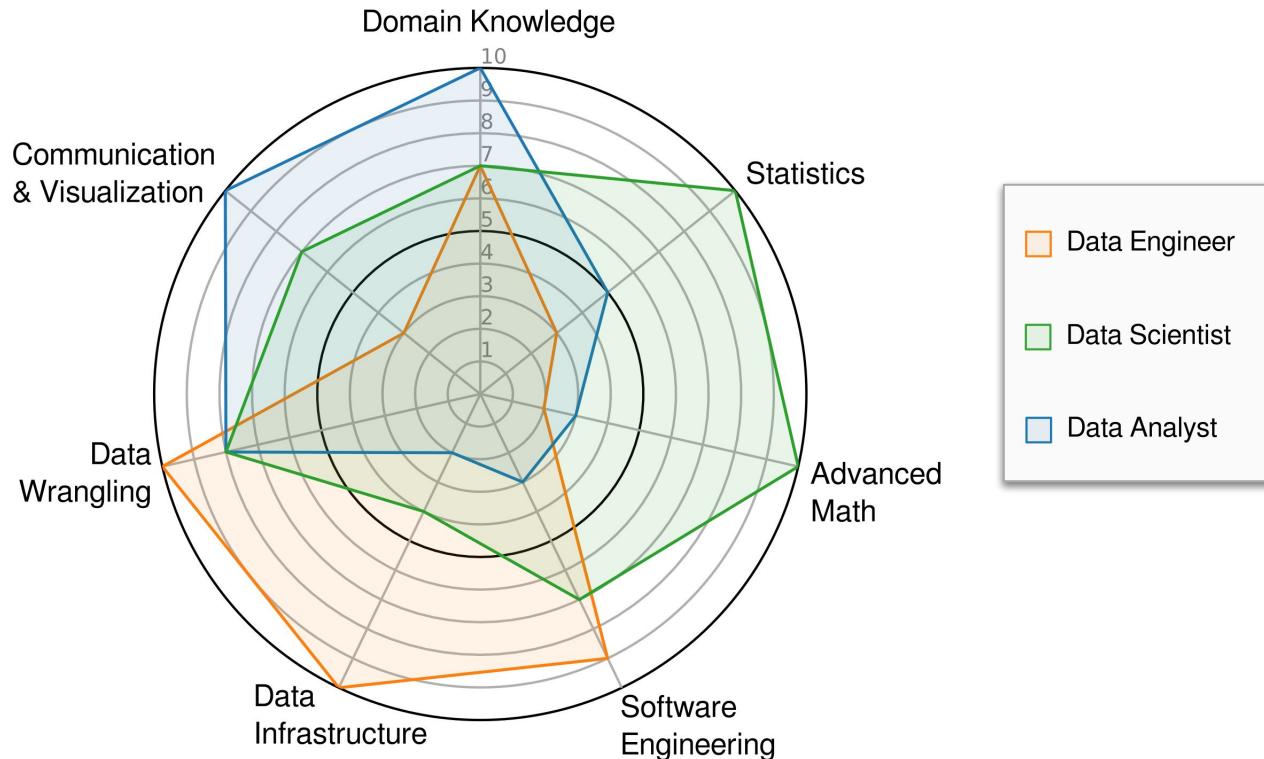
SALARY: \$96,710 - \$138,591

12. STATISTICIAN

They gather numerical data and then display it, and help companies to make sense of quantitative data and to spot trends and make predictions.

SALARY: \$57,000 - \$80,110

Comparação de *skillsets* dos papéis mais conhecidos

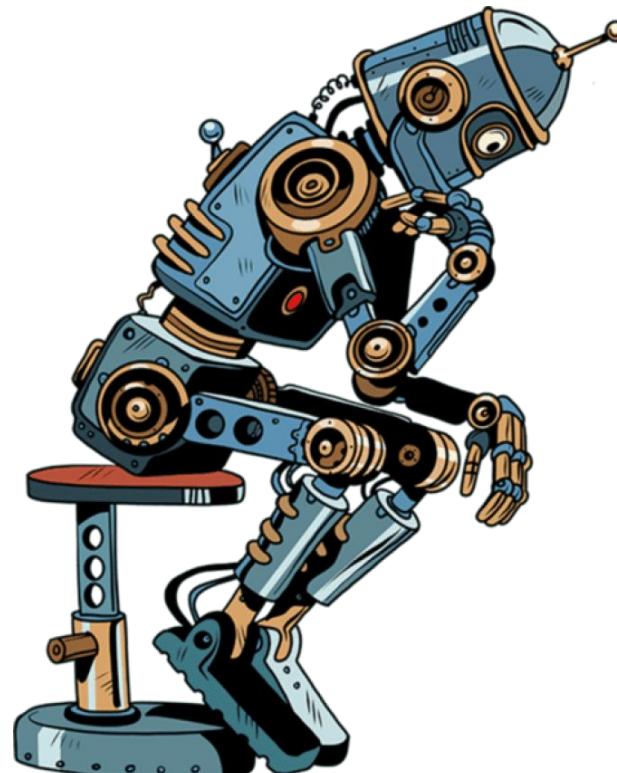




Ciência de Dados é **mais sobre ciência** do que sobre dados.

E se os dados **não forem o suficiente?**

- ? Não há **informação** nos dados
- ? Não há dados o **suficiente**
- ? O problema está **mal definido**



-  **Espírito Empreendedor**
-  **Formular Hipóteses**
-  **Criar Experimentos**
-  **Adicionar Sensores**

Exemplo:

Se o produto não varia o preço, **não há como prever** o efeito de preços diferentes na venda.

Como resolver o problema de **precificação de produtos** nesse contexto?



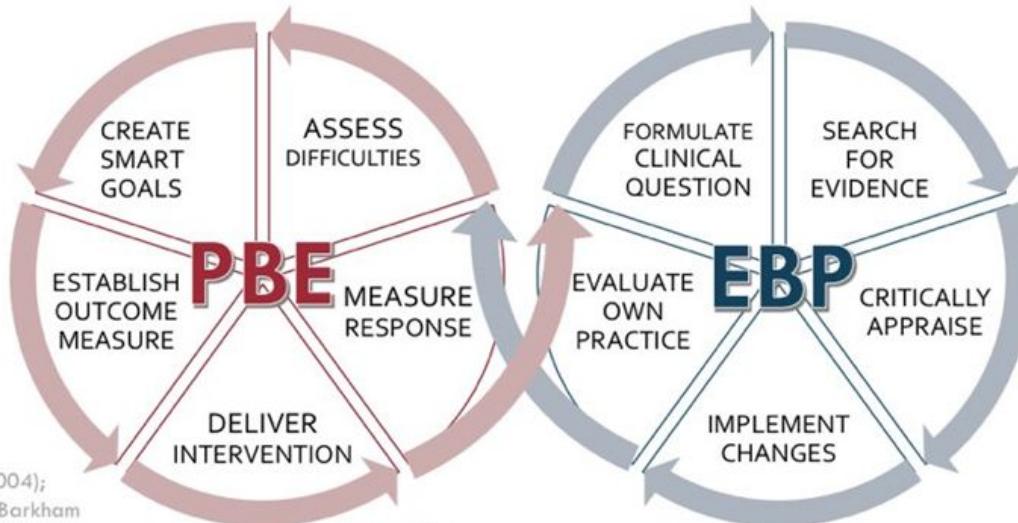
O trabalho do Cientista de Dados
é **validar hipóteses** a partir de
experimentações.

Hypothesis-Driven Development



Mentalidade **Data-Driven**

EVIDENCE-BASED PRACTICE <→ PRACTICE-BASED EVIDENCE
VIRTUOUS CYCLES



See Morgan (2004);
Green (2009); Barkham
& Mellor-Clark (2003)

- Coletar **Evidências**
- Quebrar **Silos**
- Desafiar **Tabus**
- Desafiar o **Status Quo**



#Swain_etalUK

NATHANIEL SWAIN 2016

THE UNIVERSITY OF MELBOURNE

73



MEDINDO A MATURIDADE DAS EMPRESAS



Como identificar que uma
empresa **não sabe** o que é
um **cientista de dados**?

CIENTISTA DE DADOS JR

Descrição:

Procuramos um profissional de dados para nos apoiar com análises preditivas para tomadas de decisão de nossa diretoria.

Requisitos:

- Prática em linguagem Python, R, Scala ou Java;
- Experiência com ferramentas de processamento de dados distribuídos (Hadoop, Hive, Spark, Flink, Beam, Storm, Samza);
- Experiência com SQL (Postgres);
- Experiência com Modelagem Estatística;
- Conhecimento em Machine Learning (Tensorflow, Keras, Xgboost, Matplotlib, Seaborn);
- Prática com ferramentas de visualização de dados (Tableau, Power Bi, QlikView);
- Domínio em Excel e Powerpoint.

Atividades:

- Construção e otimização de ETL;
- Gerenciamento dos Data Lakes da Companhia;
- Criação de Modelos de Machine Learning;
- Desenvolvimento de análises preditivas e prescritivas;
- Desenvolvimento de análises descritivas;
- Criação de Dashboards interativos;
- Criação de Apresentações Gerenciais e apresentação de resultados e análises para a diretoria;

Diferenciais:

- Mestrado ou Doutorado;
- Conhecimentos em Front-End;
- Prêmios em competições de Machine Learning;
- Publicações Científicas.

ENGENHEIRO DE DADOS

São responsáveis por **garantir que os dados estarão disponíveis** para análise de forma segura. Normalmente são os profissionais que constroem o ambiente de **big data**.



CIENTISTA DE DADOS

Utilizando habilidades em matemática, estatística e programação são capazes de **limpar, transformar e extrair valor dos dados**. Além disso, são responsáveis por criar modelos de machine learning e **realizar análises preditivas**.



ANALISTA DE DADOS

São os principais responsáveis por criar e **traduzir análises de dados para níveis gerenciais**. Normalmente tem amplo conhecimento na criação de **dashboards e apresentações de dados**.

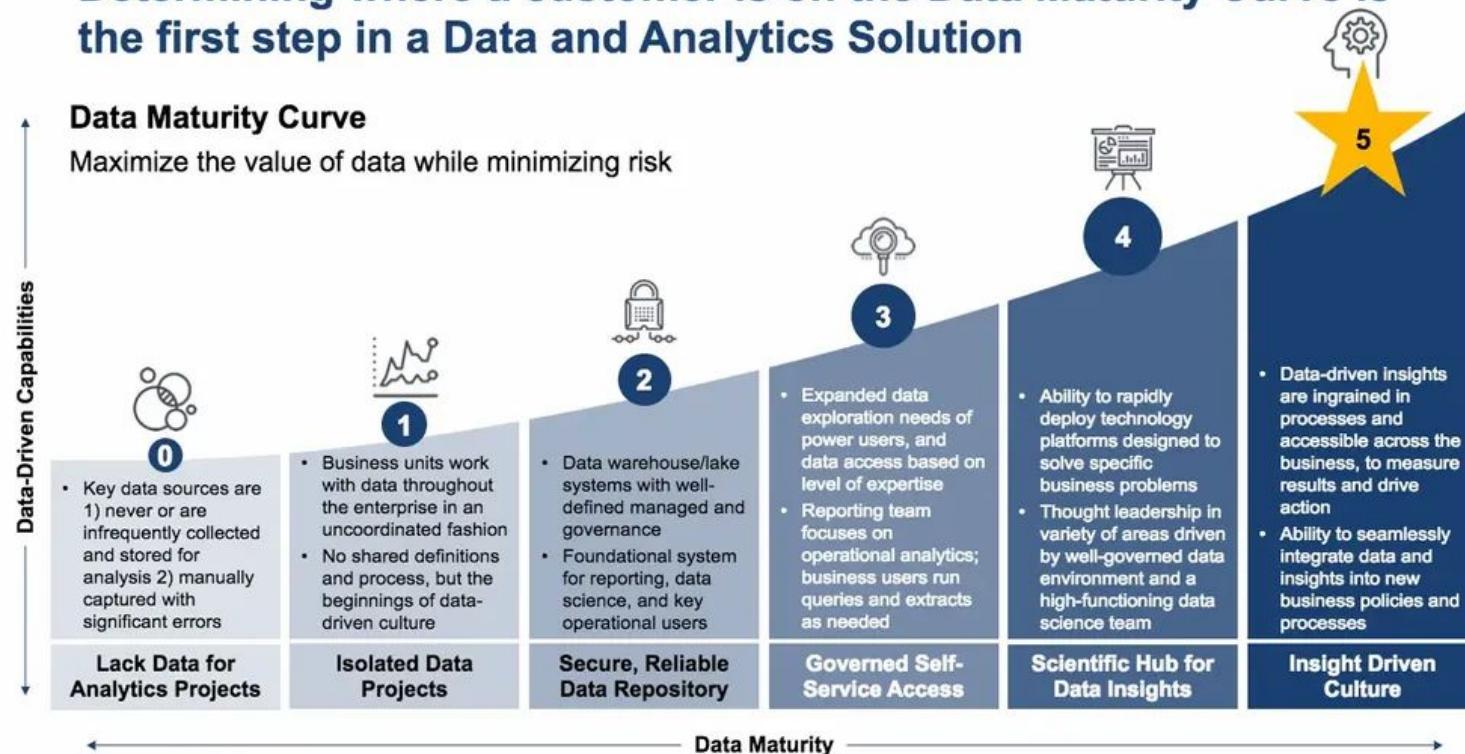




Empresas estão em
**diferentes estágios de
maturidade em dados.**

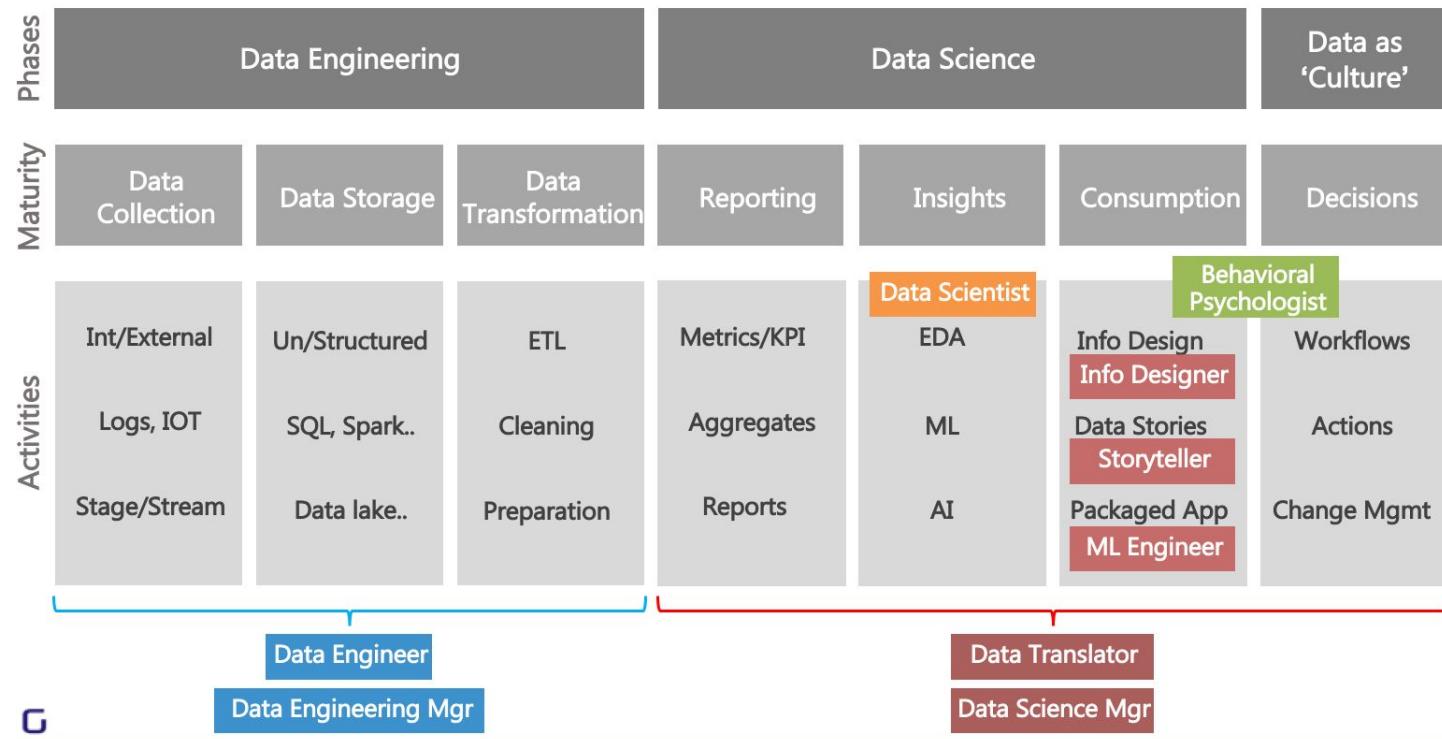
Diferentes estágios de maturidade das empresas

Determining where a customer is on the Data Maturity Curve is the first step in a Data and Analytics Solution



Os papéis nas etapas de maturidade *data-driven*

DATA SCIENCE ROLES ACROSS THE MATURITY LEVELS



Ferramenta para avaliação da **maturidade data-driven**

The 5x10 Data-Driven Maturity Model



Desafio Pós Aula: Maturidade *Data Driven*



- Escolher **uma empresa de interesse** e aplicar o Avaliação de Maturidade *Data Driven* usando como base as informações
- Analisar os resultados da avaliação e entender a **maturidade atual** da empresa.
- Avaliar se o momento da empresa é o certo para um cientista de dados começar ou se há ainda **débitos técnicos de dados**.



COMO FOI? 3QS

