

**NAVEGANDO PELO
LABIRINTO DOS**

CARGOS DE DADOS

+extras

Kevin Takano

*Bacharel e Mestre em Ciência da Computação (UFAM)
Data Scientist da Hotmart.*



Quem sou eu?

Kevin Takano



Marido da
Anne

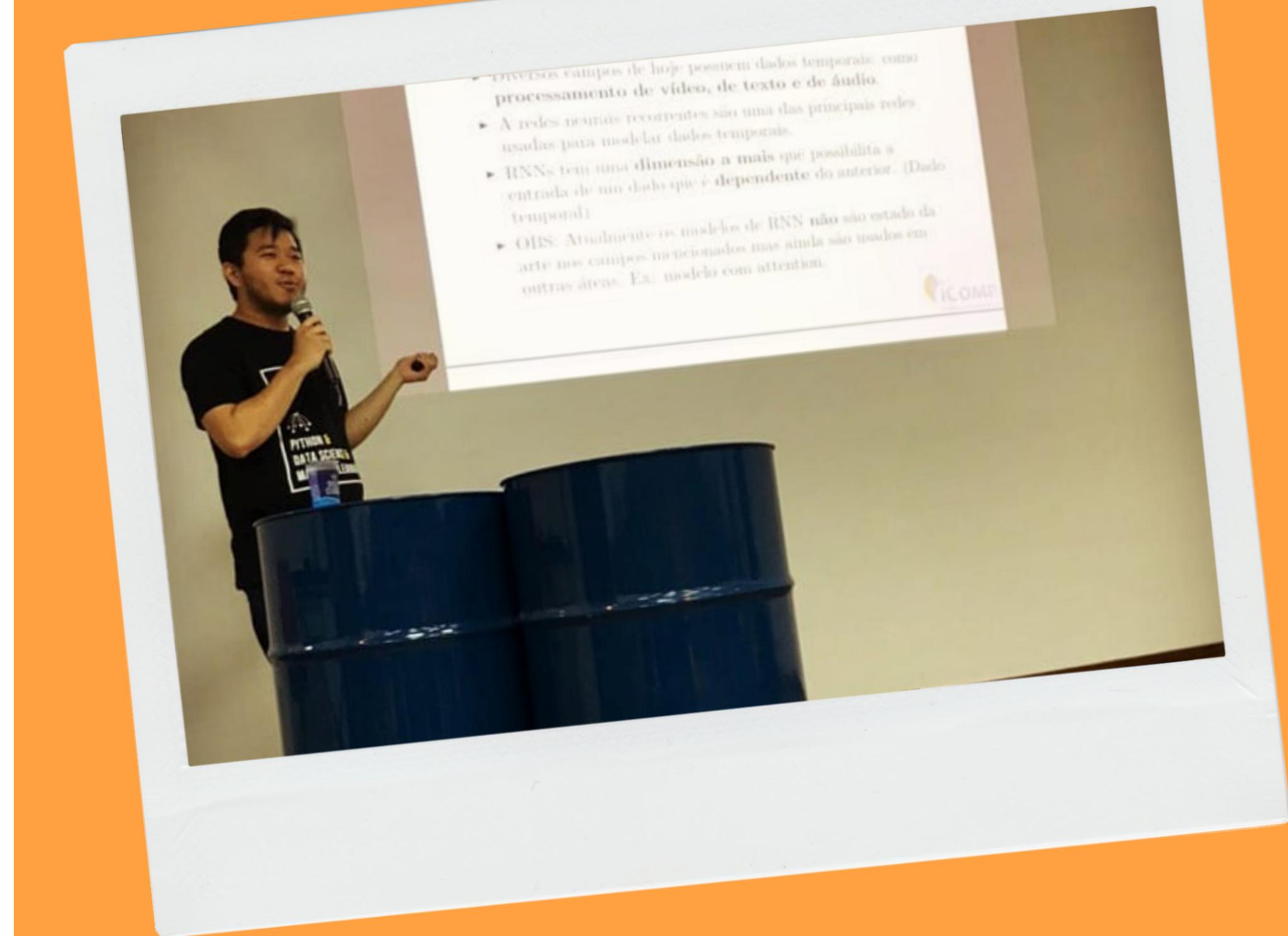


Pai do
Jiji e do Dio

Quem sou eu?

Kevin Takano

PyData 2019



OBJETIVO:

Desmistificar (ou tentar deixar menos confuso 😊) papéis na área de dados: data scientist, data engineer, ml engineer, etc.

SOBRE A PALESTRA



XP. de **Carreira** +
3943 **entrevistas** +
busca em **artigos**,
podcasts.

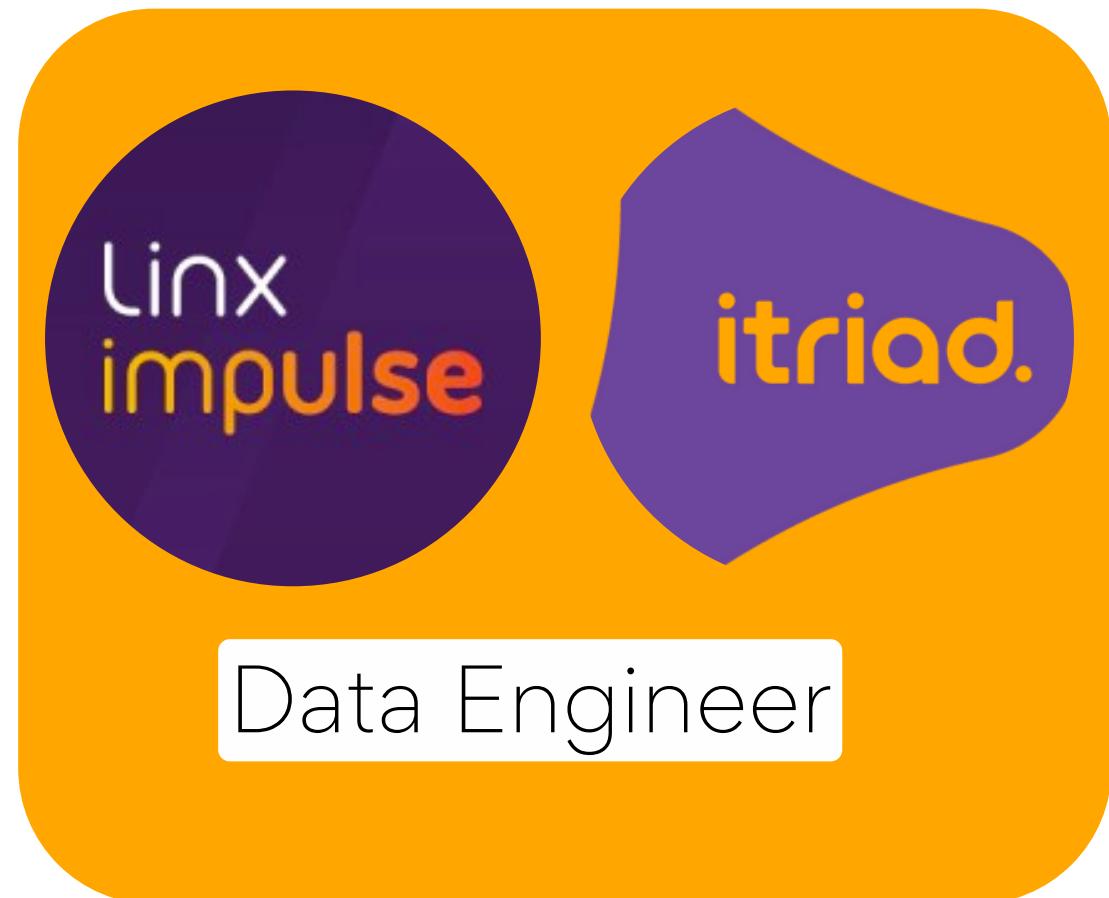
CARREIRA



CARREIRA



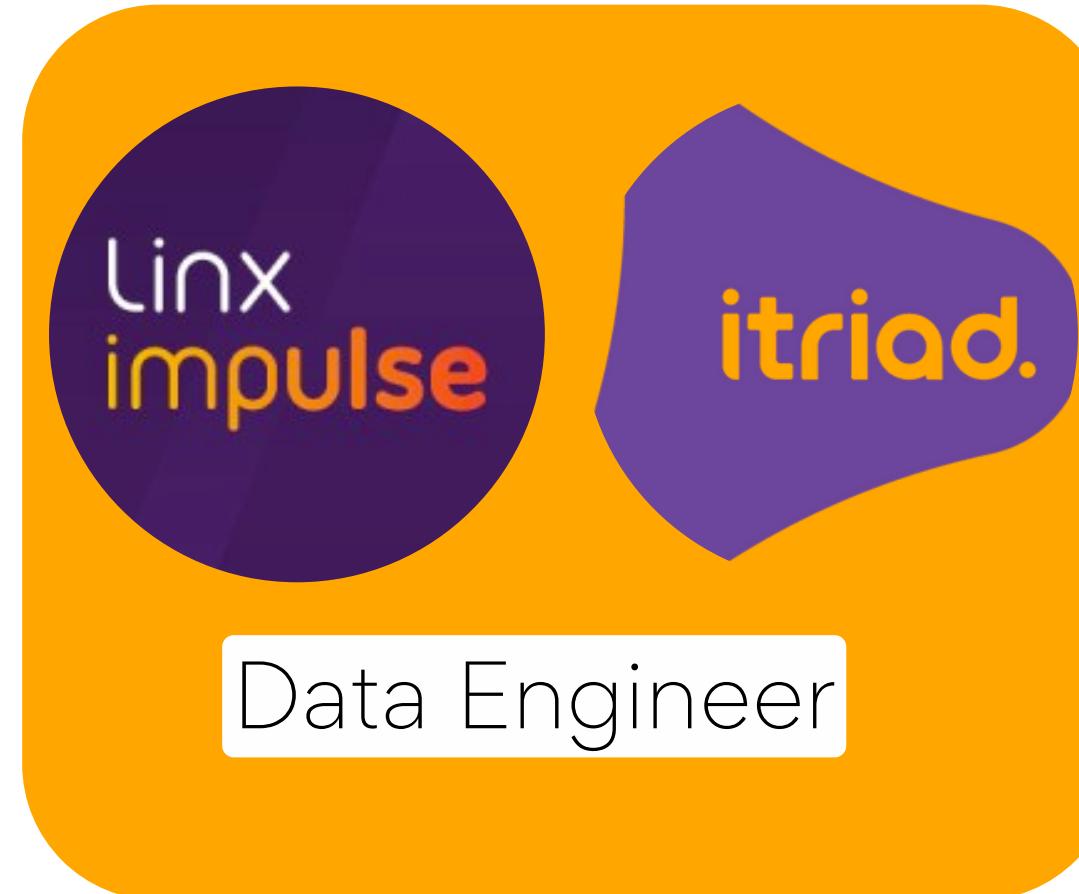
Sw Engineer



Data Engineer



CARREIRA



Why I gave up data science. Read this to understand.



Kevin Takano

5 min read · Aug 6, 2022

368 3

W+ ▶ ⬆ ...

CARREIRA



Sw Engineer



Data Engineer



Data Scientist



Machine Learning
Engineer/MLOPs
Engineer (primeiro
ano)

Data Scientist
(hoje)

Why I gave up data science. Read this to understand.



Kevin Takano

5 min read · Aug 6, 2022

368 3

...
W+ ▶ ⌂ ⌄

CONTEXTO

IMAGINE EM UM UNIVERSO PARALELO DA TERRA 2



TERRA 1

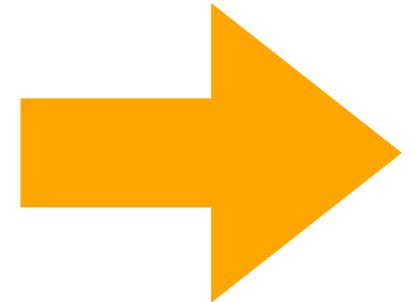


TERRA 2

IMAGINE EM UM UNIVERSO PARALELO DA TERRA 2



TERRA 2

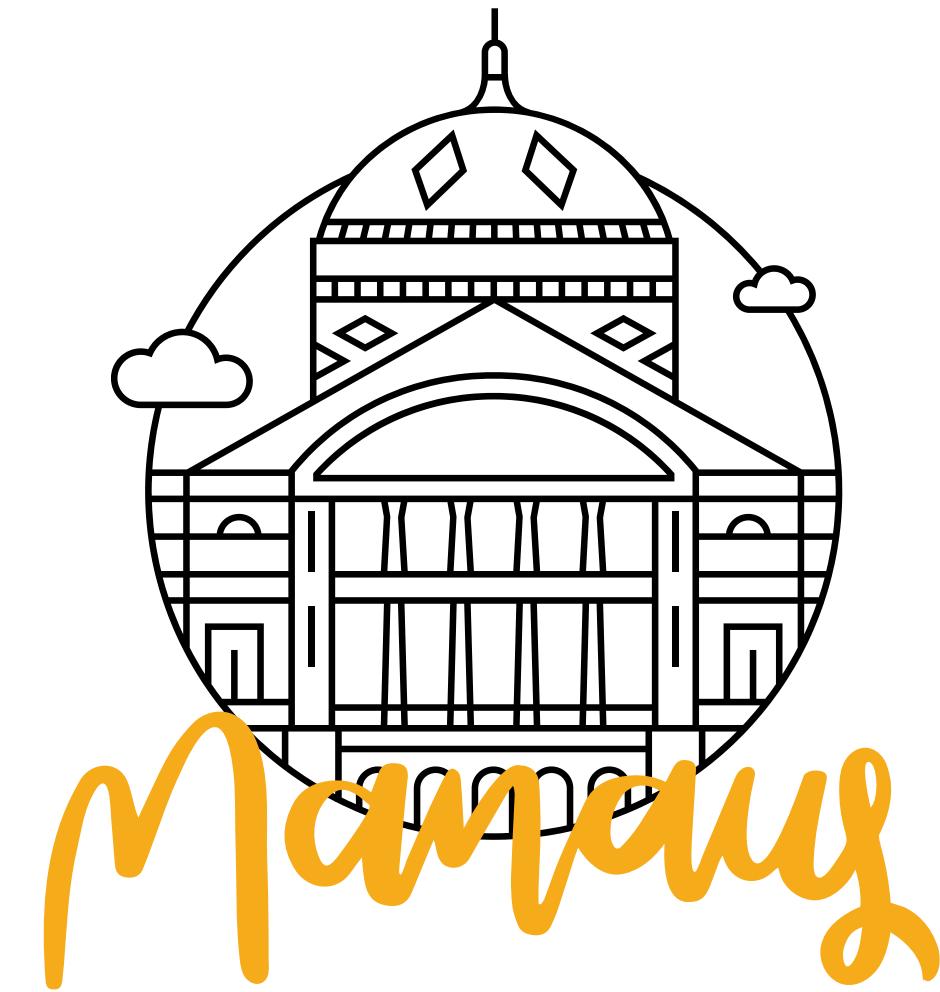
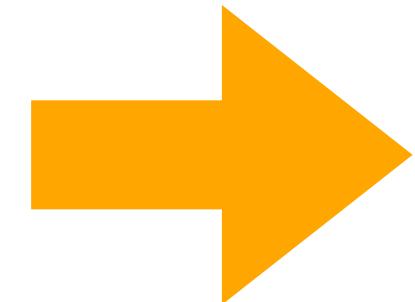


BRASILLL 2

IMAGINE EM UM UNIVERSO PARALELO DA TERRA 2

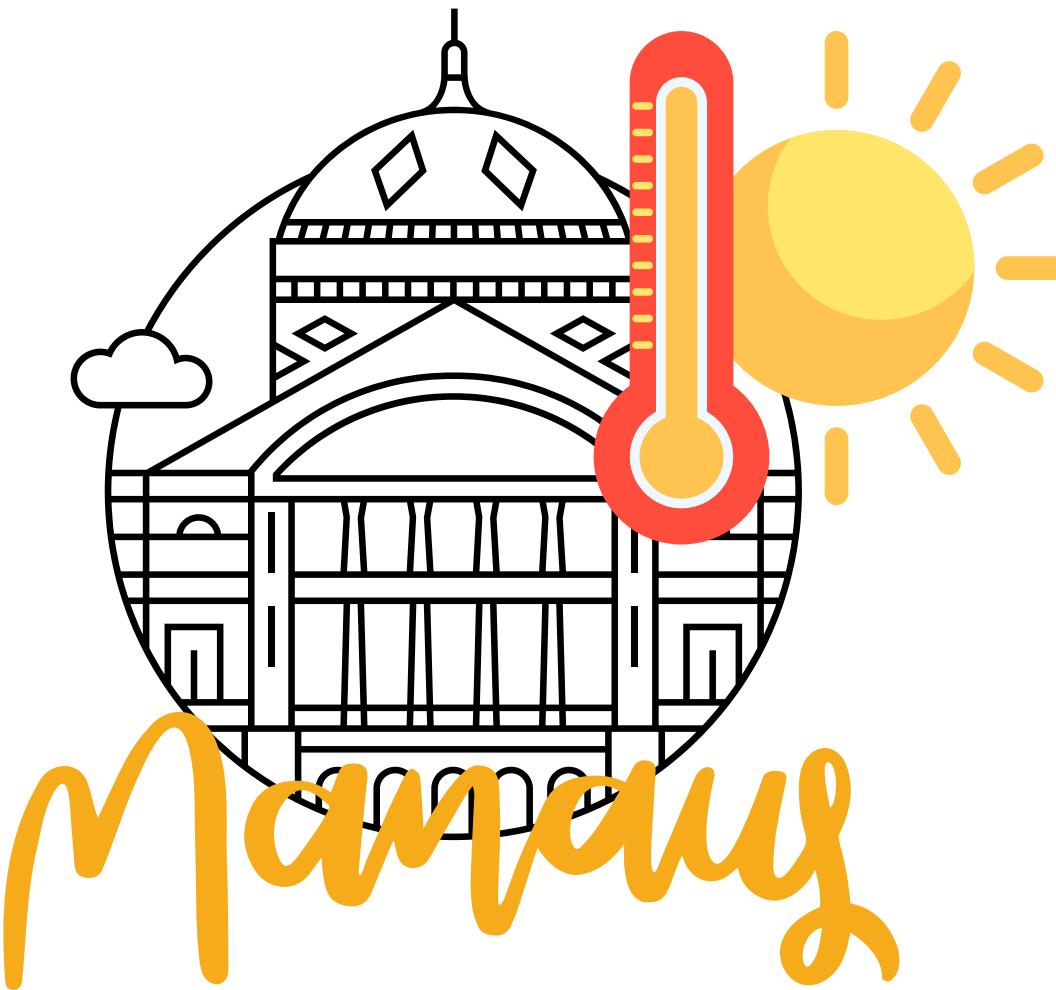


BRASIL 2



MANAUS 2

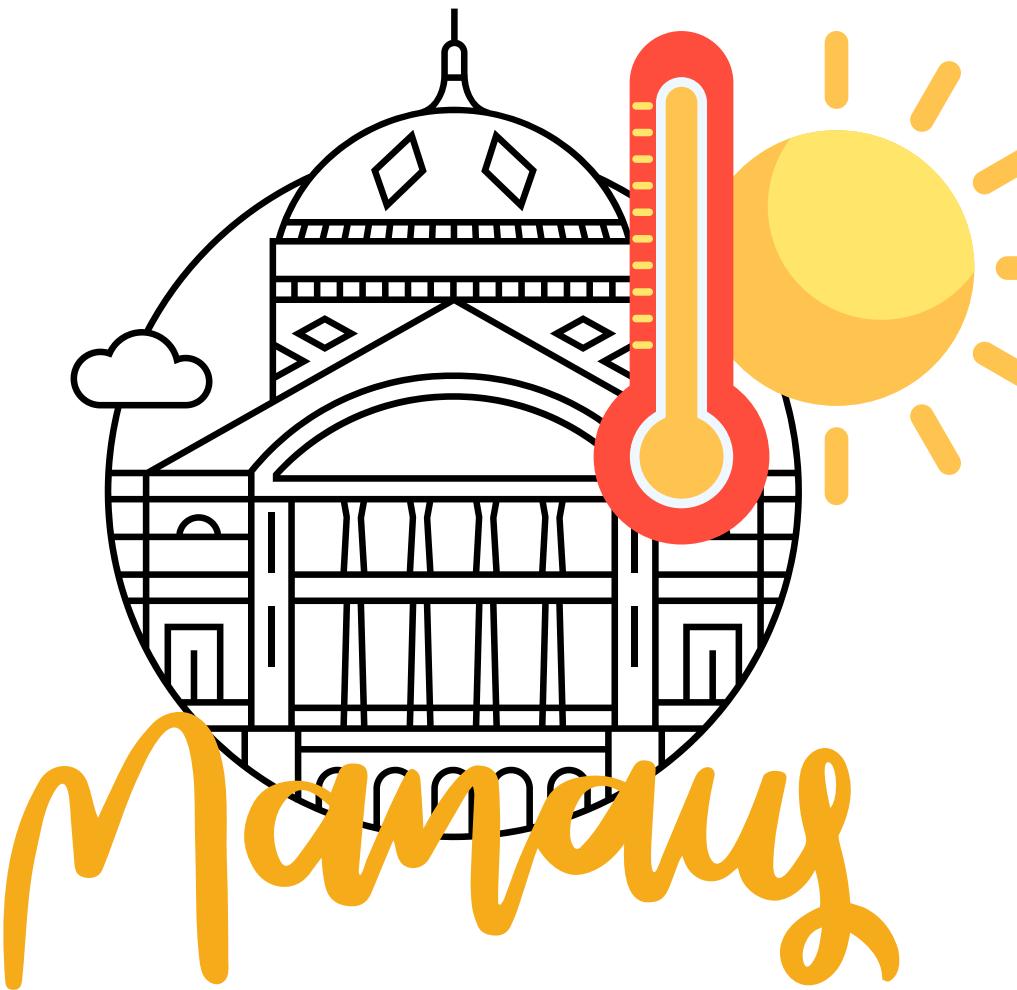
IMAGINE EM UM UNIVERSO PARALELO DA TERRA 2



Manaus

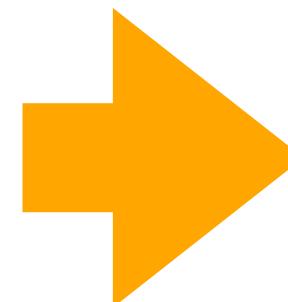
MANAUS 2

IMAGINE EM UM UNIVERSO PARALELO DA TERRA 2



MANAUS 2

Empresa de E-commerce
do Amazonas



Site da Mebol: Venda de
eletrônicos, móveis, etc.

IMAGINE EM UM UNIVERSO PARALELO DA TERRA 2

RECOMENDAMOS PRA VOCÊ

Eletroportáteis

Eletrônicos

Ar e Ventilação

Smartphones



★★★★★ (0)

Ar-Condicionado Janela Springer-
Midea 7500 BTus 127V...

220V



★★★★★ (34)

Split Springer Midea 9.000 BTUS
Inverter Xtreme WI-FI 220V...



★★★★★ (166)

Ventilador 2 em 1 Arno 40cm Turbo
Force 2 Em 1 Preto - VF42

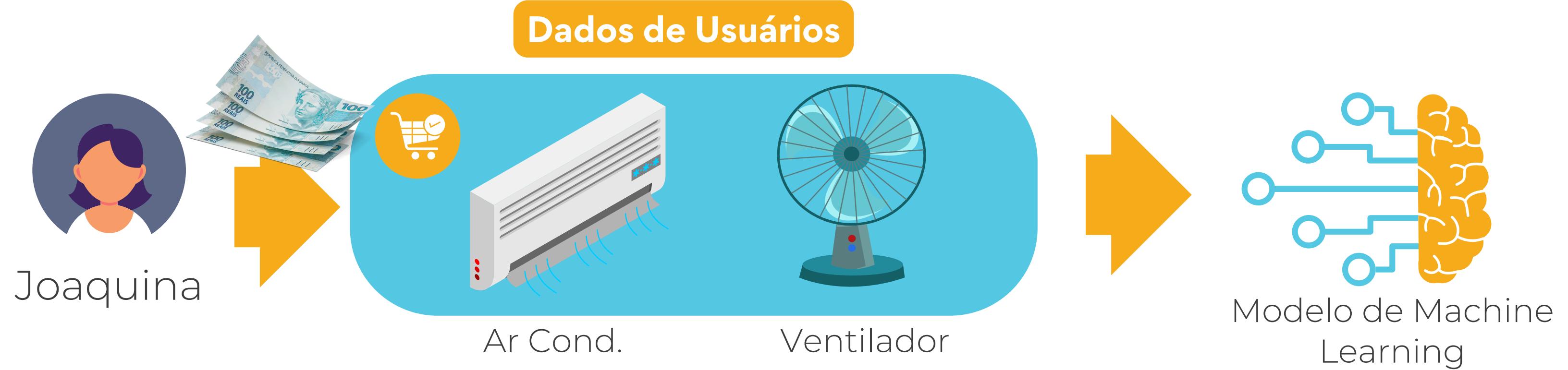


★★★★★ (80)

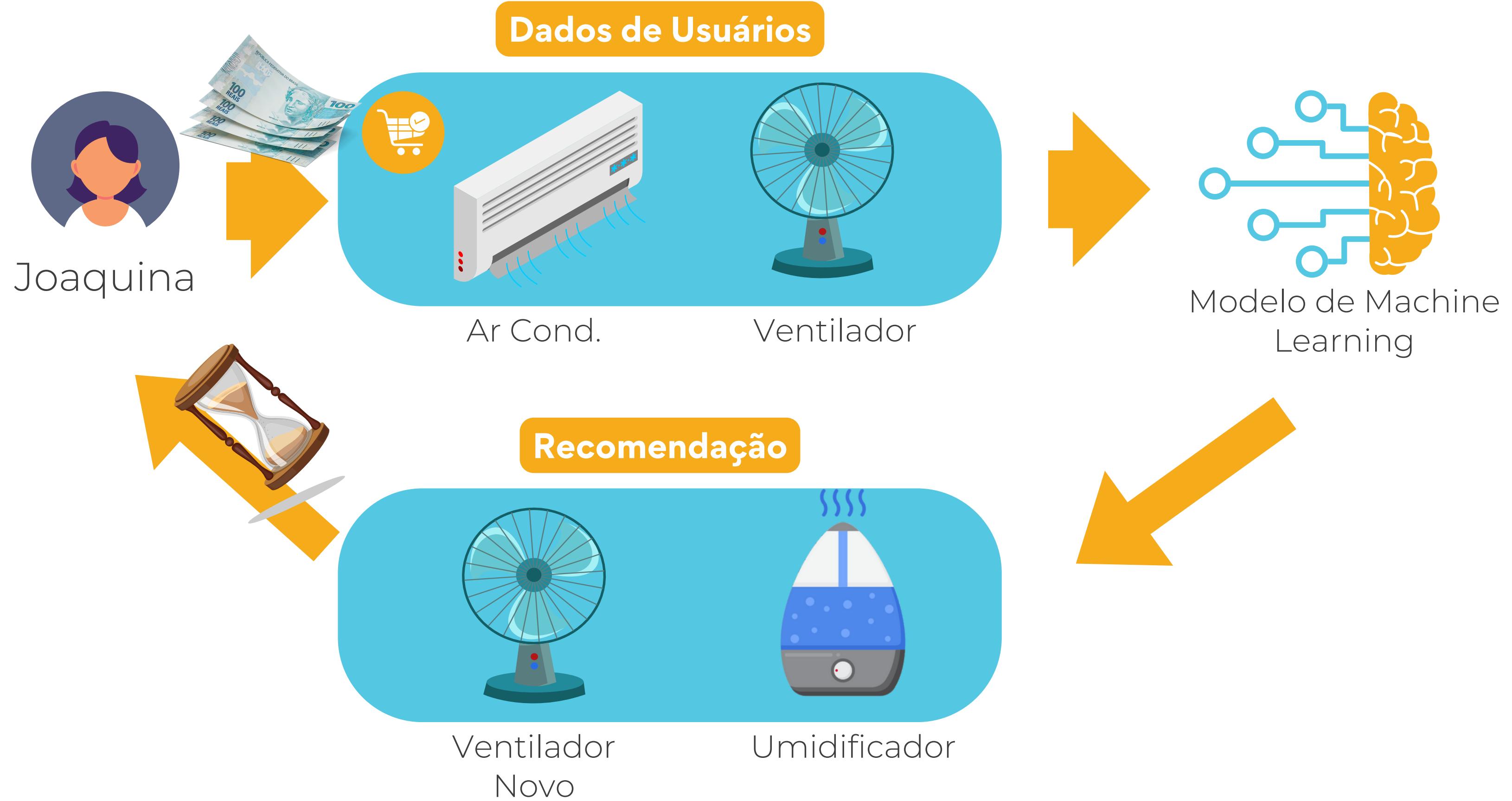
Ventilador de Parede Arno 40cm
Xtreme Force Breeze Preto VB4P

Querem criar um **sistema de
recomendação**

MEBOL QUER UM SISTEMA DE RECOMENDAÇÃO



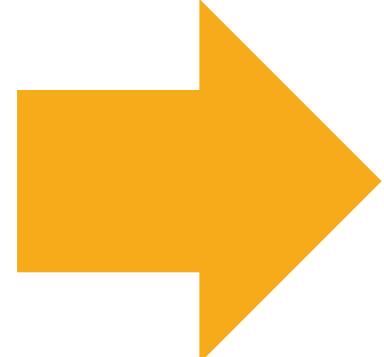
MEBOL QUER UM SISTEMA DE RECOMENDAÇÃO



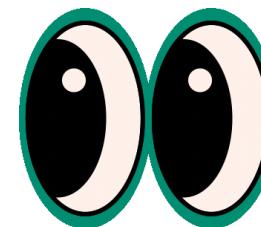
MEBOL PRECISA DE UM TIME



Divulga **vagas** para seu time de dados para criar o sistema de recomendação.



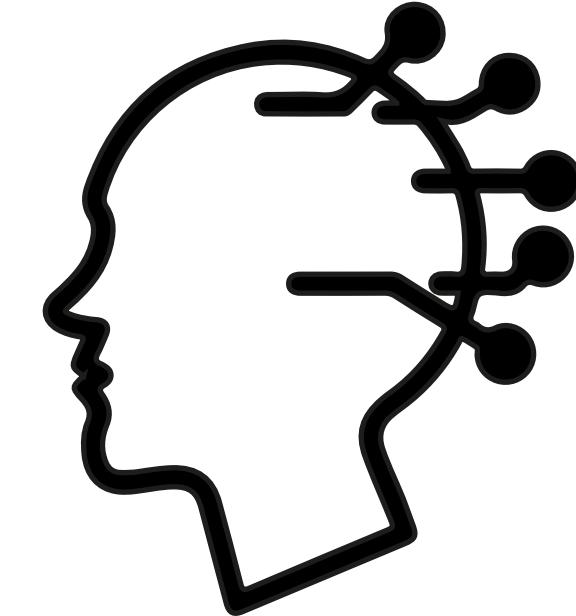
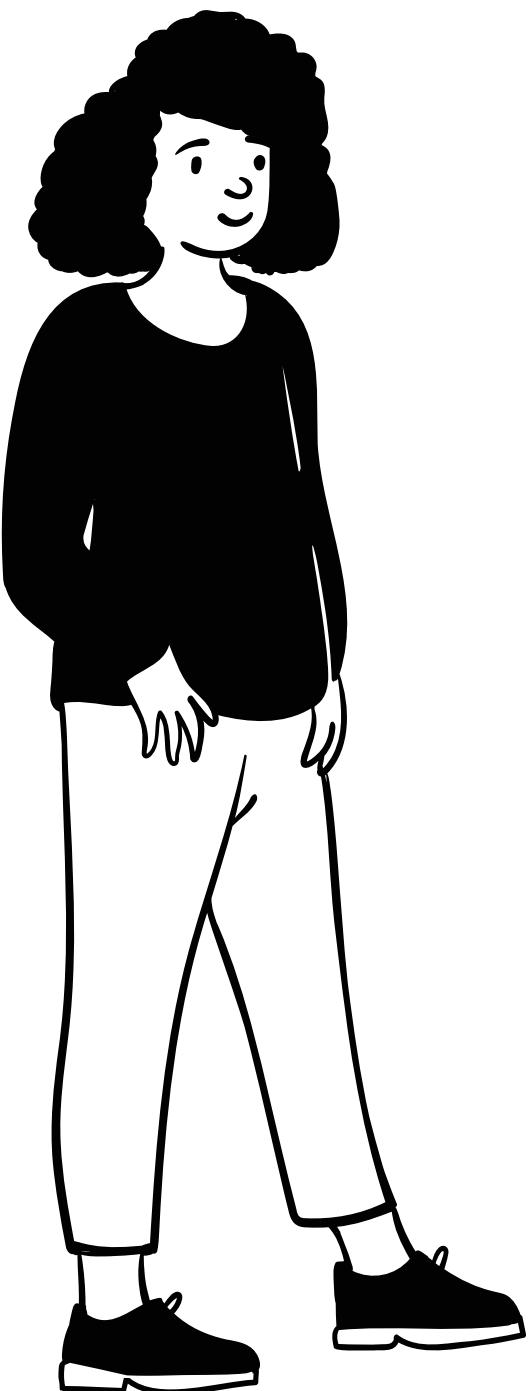
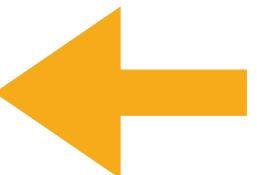
Alissa, uma jovem formanda em C.C olha as vagas da Mebol.



ALISSA



Formanda em C.C
na UFAM.



Fez pesquisa em iniciação
científica em redes neurais.



Está super empolgada para criar
muitos modelos como **cientista**
de dados

NO PRIMEIRO ANO

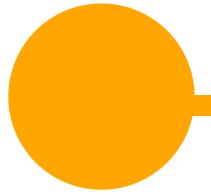


No decorrer do
ano..



Alissa consegue uma vaga de
estágio de ciência de dados
na MEBOL.

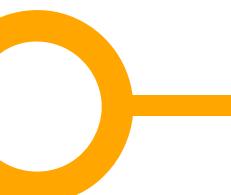
NO PRIMEIRO ANO



No decorrer do
ano..



Alissa consegue uma vaga de
estágio de ciência de dados
na MEBOL.



Alissa então quebra a cara a
área de ciência de dados.

ALISSA SE FRUSTRA

1 - Ela percebe que **passa 50%** do tempo apenas se **comunicando** com o time de negócio criando **apresentações**.



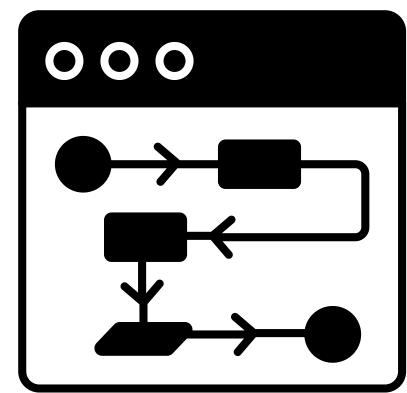
ALISSA SE FRUSTRA



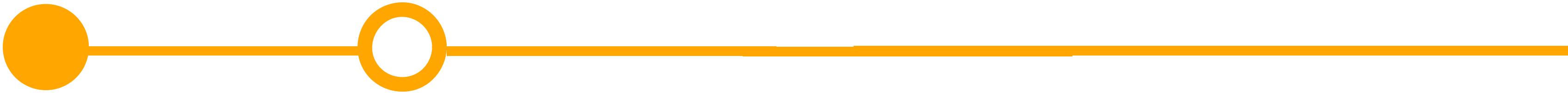
1 - Ela percebe que **passa 50%** do tempo apenas se **comunicando** com o time de negócio criando **apresentações**.



2 - Ela não aplica tanto seu conhecimento de **algoritmos, estrutura de dados, paradigmas** de programação na graduação.



NO PRIMEIRO ANO



No decorrer do
ano

Alissa, começa a
perceber que
talvez data science
não seja sua área e
pensa em **trocar**.

NO PRIMEIRO ANO

Nem os **recruiters**

sabem..



No decorrer do
ano

Alissa, começa a
perceber que
talvez data science
não seja sua área e
pensa em **trocar**.

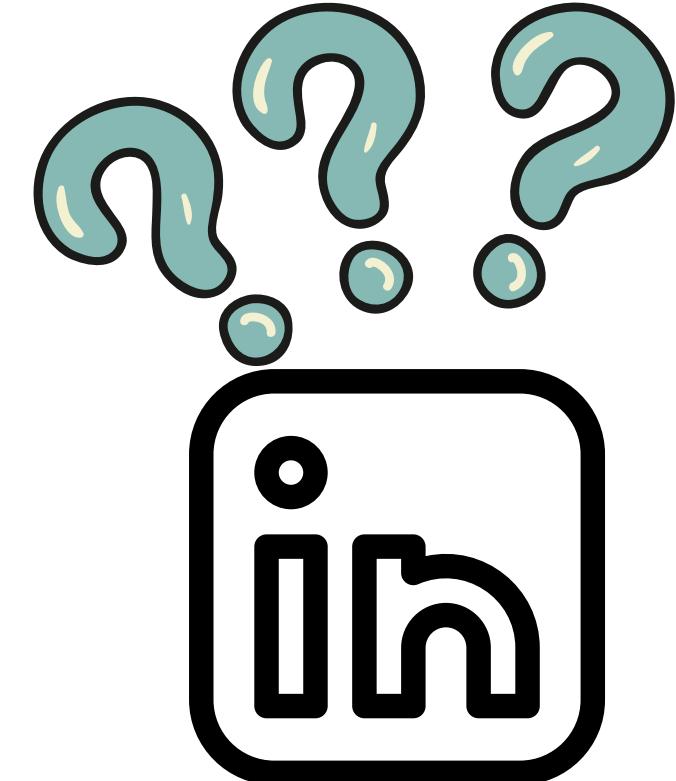
Alissa olha as vagas no
LinkedIn, e vê que a
área de dados é um
pouco bagunçada. E
por isso não seria tão
fácil escolher.

NO PRIMEIRO ANO



No decorrer do
ano

Alissa, começa a
perceber que
talvez data science
não seja sua área e
pensa em **trocar**.



Nem os **recruiters**
sabem..



Alissa olha as vagas no
LinkedIn, e vê que a
área de dados é um
pouco bagunçada. E
por isso não seria tão
fácil escolher.

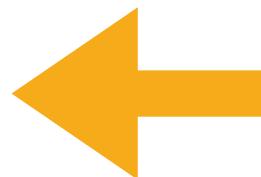


Logo, pede ajuda ao
seu professor amigo
Valter Branco.

VALTER BRANCO



Doutor em ciência
da computação na
área de machine
learning aplicado a
Química.



Gerente de dados no **Grupo**
Botocário.



Uma empresa de **perfumes e**
cosméticos da **Amazonia com uma**
cultura muito madura em dados.

ALISSA PEDE AJUDA AO PROF. VALTER BRANCO

Me ajuda a me
encontrar, tô
perdida..



Para entender os
cargos de dados, é
preciso entender a
jornada dos dados!

A JORNADA DOS
DADOS

A JORNADA DOS DADOS



A JORNADA DOS DADOS

1. COLETA



Objetivo: Coletar e extrair os dados para futuros produtos.

2. GERENCIAMENTO E ARMAZENAMENTO

Joaquina



5. APRENDIZAGEM E

7 MLOPs

4. DESENVOLVIMENTO E ANÁLISE

App da mebol



6. DE SERVIÇOS

Dados de usuários



A JORNADA DOS DADOS



A JORNADA DOS DADOS

1. COLETA



2. GERENCIAMENTO E ARMAZENAMENTO



3. TRA

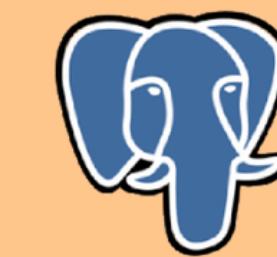
Objetivo:

1. Entender o dado e **definir qual a infra** para os dados: **postgres, redshift, athena.** etc.

5. APRENDIZAGEM E OTIMIZAÇÃO



Amazon
Redshift



7.M



Amazon Athena

PostgreSQ

4. EXPLORAÇÃO E ANÁLISE



6. DEPLOY E SERVING



SKILLS: BANCO DE DADOS, INFRA, BIG DATA

A JORNADA DOS DADOS

1. COLETA



2. GERENCIAMENTO
ARMAZENAMENTO

SKILLS: BANCO DE DADOS, INFRA, BIG DATA

3. TRA

Objetivo:

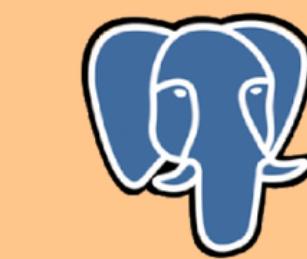
1. Entender o dado e **definir qual a infra** para os dados: **postgres, redshift, athena.** etc.

2. Desenvolver **datalake com ETLs.**

5. APRENDIZAGEM E
OTIMIZAÇÃO



Amazon
Redshift



7.M

Amazon Athena



PostgreSQ



A JORNADA DOS DADOS



A JORNADA DOS DADOS

Objetivo:

1. COLE



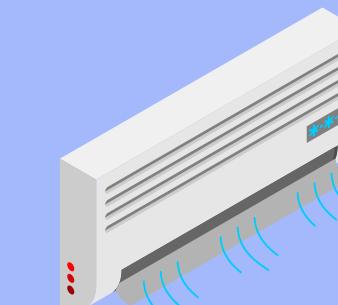
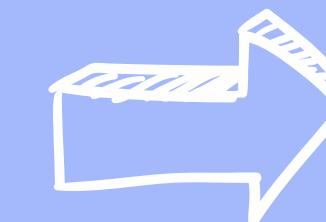
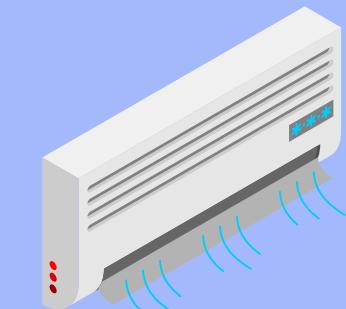
Limpeza e transformação de dados.

Remoção de ruído. Ex: preços negativos.



2. GERENCIAMENTO E ARMAZENAMENTO

Exemplo: Normalização de preço.



1000,00 R\$

1000.00 R\$



4. EXPLORAÇÃO E ANÁLISE



6. DEPLOY E SERVING

SKILLS: PIPELINES, BIG DATA, ALGORITMOS, TRANSFORMAÇÃO DE DADOS

A JORNADA DOS DADOS



4. EXPLORAÇÃO E ANÁLISE

A JORNADA DOS DADOS

Objetivo:

Análise de padrões dos dados, análise estatística e segmentação.

Ajudar a tomar decisões de negócio.

Exemplo de perguntas para análise:

1. Quais usuários têm mais probabilidade de aceitarem nosso cartão Mebol?

2. GERENCIAMENTO E ARMAZENAMENTO

4. EXPLORAÇÃO E ANÁLISE

6. DEPLOY E SERVING

SKILLS: ANÁLISE DE DADOS, COMUNICAÇÃO,
VISUALIZAÇÃO DE DADOS

4. EXPLORAÇÃO E ANÁLISE

A JORNADA DOS DADOS

Objetivo:

Com o análise de padrões,
pode-se gerar **Marketing Personalizado**.



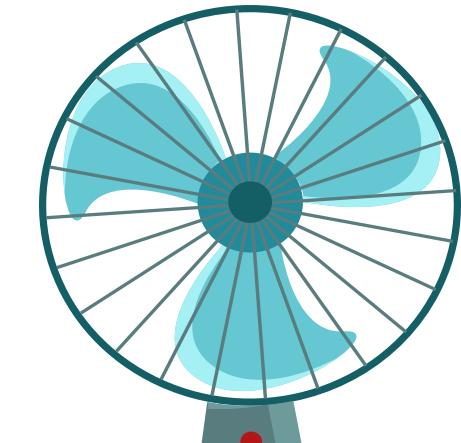
2. GERENCIAMENTO E ARMAZENAMENTO

Exemplo:

☀️ Refresque-se em Manaus: Oferta Especial de Ar Condicionado!

Olá **Valentina**,

Sabemos que Manaus **está fervendo este mês**! Por isso, temos uma oferta exclusiva na MEBOL só para você...



A JORNADA DOS DADOS



A JORNADA DOS DADOS

Objetivo:

Criação de modelos inteligentes para previsão de comportamentos para impulsionar o negócio.

Esta etapa envolve pesquisa e otimização de parâmetros.

SKILLS: MACHINE LEARNING, ESTATÍSTICA,
COMUNICAÇÃO

Exemplo de modelos:

1. Modelo de **recomendação** no site da MEBOL..



A JORNADA DOS DADOS



A JORNADA DOS DADOS

Objetivo:

Basicamente, uma vez que o modelo está pronto, precisa-se implantar o modelo em uma aplicação ou pipeline.



2. GERENCIAMENTO E ARMAZENAMENTO

SKILLS: ENG. DE SW, MACHINE LEARNING

1. COLETA

3. TRANSFORMAÇÃO

RECOMENDAMOS PRA VOCÊ

Eletroportáteis

Eletrônicos

Ar e Ventilação

Smartp



★★★★★ (0)

Ar-Condicionado Janela Springer-Midea 7500 BTus 127V...

220V



★★★★★ (34)

Split Springer Midea 9.000 BTUS Inverter Xtreme WI-FI 220V...



★★★★★ (166)

Ventilador 2 em 1 Arno 40cm Force 2 Em 1 Preto - VF42

4. EXPLORAÇÃO E ANÁLISE

6. DEPLOY E SERVING

A JORNADA DOS DADOS



OS CARGOS DOS
DADOS

OS CARGOS DE DADOS

RESPONSÁVEL: SOFTWARE ENGINEER

1. COLETA



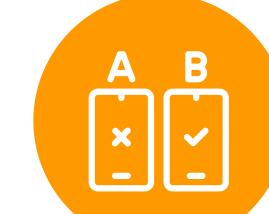
3. TRANSFORMAÇÃO



**2. GERENCIAMENTO E
ARMAZENAMENTO**

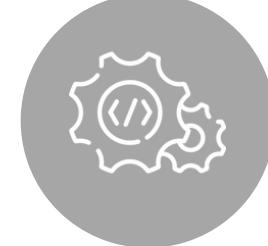


**5. APRENDIZAGEM E
OTIMIZAÇÃO**

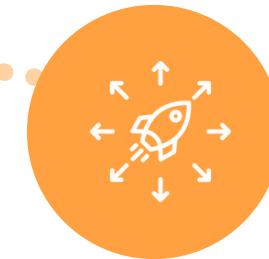


4. EXPLORAÇÃO E ANÁLISE

7. MLOPs



**6. DEPLOY E
SERVING**



SKILLS: BOAS PRÁTICAS DE PROGRAMAÇÃO,
BACK-END, FRONT-END, BANCO DE DADOS.

OS CARGOS DE DADOS

RESPONSÁVEL: DATA ENGINEER

1. COLETA



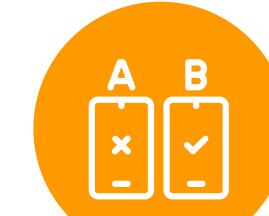
3. TRANSFORMAÇÃO



**2. GERENCIAMENTO E
ARMAZENAMENTO**

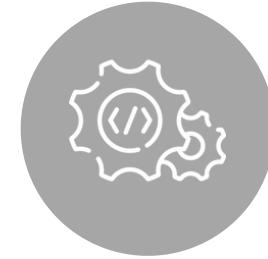


**5. APRENDIZAGEM E
OTIMIZAÇÃO**

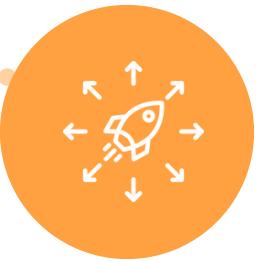


**4. EXPLORAÇÃO
E ANÁLISE**

7. MLOPs



**6. DEPLOY E
SERVING**



**SKILLS: PIPELINE, ETL, ORQUESTRAÇÃO DE
DADOS, BIG DATA**

OS CARGOS DE DADOS

RESPONSÁVEL: DATA ANALYST/ BUSINESS ANALYST



SKILLS: CAPACIDADE ANALÍTICA, VISUALIZAÇÃO DE DADOS, TRANSFORMAÇÃO DE DADOS, COMUNICAÇÃO.

OS CARGOS DE DADOS

RESPONSÁVEL: DATA SCIENTIST

1. COLETA



3. TRANSFORMAÇÃO



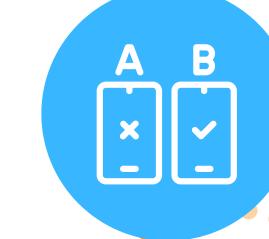
**2. GERENCIAMENTO E
ARMAZENAMENTO**



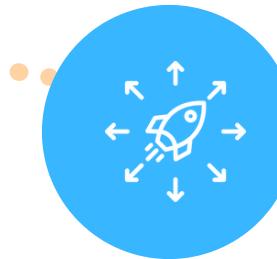
**4. EXPLORAÇÃO
E ANÁLISE**



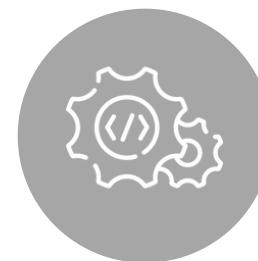
**5. APRENDIZAGEM E
OTIMIZAÇÃO**



**6. DEPLOY E
SERVING**



7. MLOPs



SKILLS: COMUNICAÇÃO, MACHINE LEARNING,
ESTATÍSTICA, ANÁLISE DE DADOS

OS CARGOS DE DADOS

RESPONSÁVEL: DATA SCIENTIST

1. COLETA



3. TRANSFORMAÇÃO

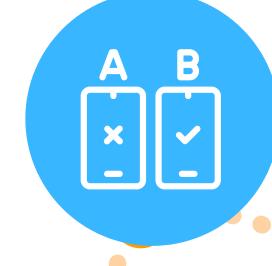


Em algumas empresas, o Cientista não faz o deploy. Isto vai para outro papel..

2. CREDENCIAMENTO E ARMAZENAMENTO



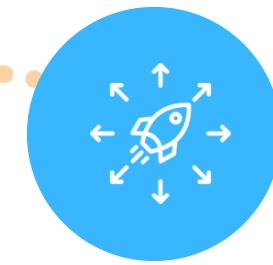
5. APRENDIZAGEM E OTIMIZAÇÃO



4. EXPLORAÇÃO E ANÁLISE



6. DEPLOY E SERVING



7. MLOPs



OS CARGOS DE DADOS

SKILLS: MACHINE LEARNING, ENG. DE SW, INFRA E ARQUITETURA DE SW.

1. COLETA



3. TRANSFORMAÇÃO



Em algumas empresas, o Cientista não faz o deploy. Isto vai para outro papel..

2. CREDENCIAMENTO E ARMAZENAMENTO



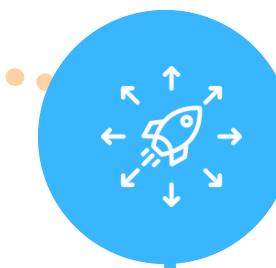
5. APRENDIZAGEM E OTIMIZAÇÃO



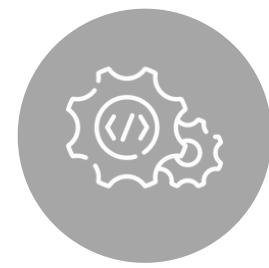
4. EXPLORAÇÃO E ANÁLISE



6. DEPLOY E SERVING

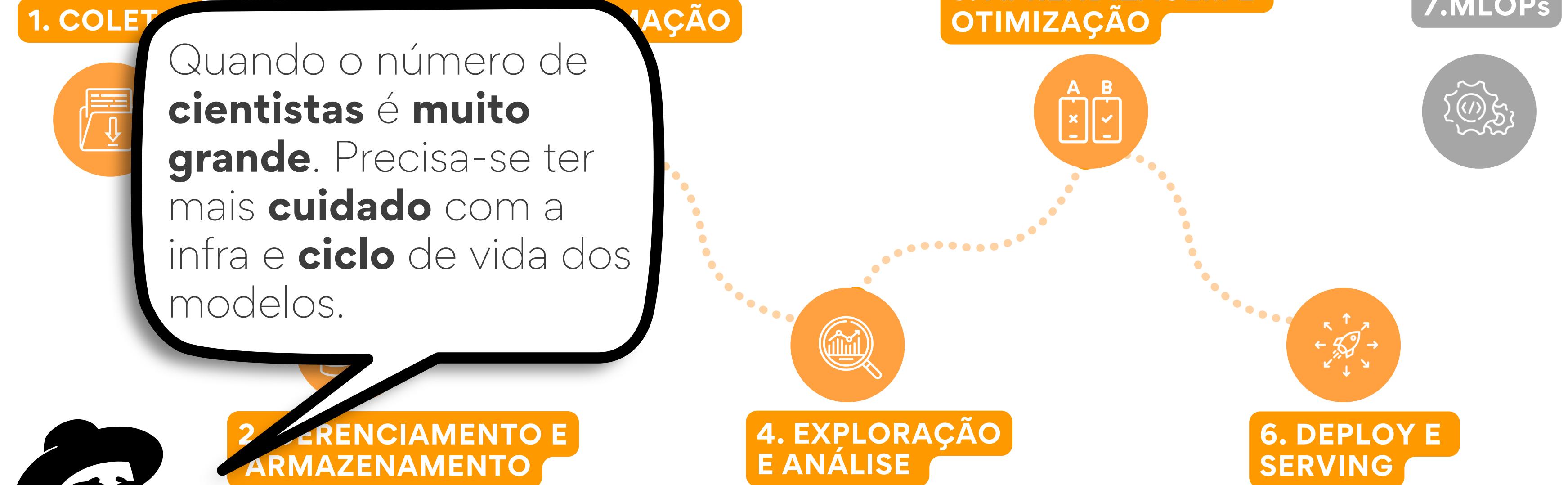


7. MLOPs



RESPONSÁVEL: MACHINE LEARNING ENGINEER

OS CARGOS DE DADOS



OS CARGOS DE DADOS

RESPONSÁVEL: MLOPS ENGINEER

1. COLET



2. TECN



Quando o número de **cientistas** é **muito grande**. Precisa-se ter mais **cuidado** com a infra e **ciclo** de vida dos modelos.

2. REFERENCIAMENTO E ARMAZENAMENTO

3. INTEGRAÇÃO

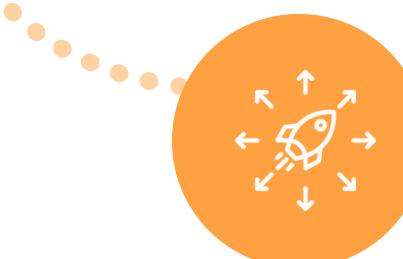


4. EXPLORAÇÃO E ANÁLISE

5. APRENDIZAGEM E OTIMIZAÇÃO



6. DEPLOY E SERVING



7. MLOPs



SKILLS: DEV OPS, ML, ENG. DE SW, INFRA E ARQUITETURA DE SW.

OS CARGOS DE DADOS

RESPONSÁVEL: ANALYTICS ENGINEER

1. COLETA



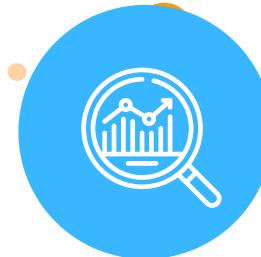
3. TRANSFORMAÇÃO



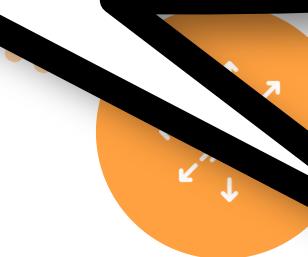
**2. GERENCIAMENTO E
ARMAZENAMENTO**



4. EXPLORAÇÃO E ANÁLISE



**6. DEPLOY E
SERVING**



SKILLS: COMUNICAÇÃO, PIPELINE, BANCO DE DADOS,
ETL, BIG DATA.

Quando a empresa **é maior**,
precisa-se ter um maior
cuidado nas **consultas** de
banco de dados e por isso
empresas **criaram um**
intermediário entre data
engineer e data **analyst**.



OS CARGOS DE DADOS

RESPONSÁVEL: AI RESEARCHER

1. COLETA

3. TRANSFORMAÇÃO

5. APRENDIZAGEM E
OTIMIZAÇÃO

7. MLOPs

Quando a modelagem é muito difícil e próximo da ciência..

2. DIFERENCIAMENTO E
ARMAZENAMENTO

4. EXPLORAÇÃO
E ANÁLISE

6. DEPLOY E
SERVING



SKILLS: MACHINE LEARNING, ESTATÍSTICA,
MATEMÁTICA, PESQUISA, COMUNICAÇÃO E ESCRITA.

OS CARGOS DE DADOS



RESPONSÁVEL: ML ENGINEER/ FULL STACK DATA SCIENTIST

1. COLETA



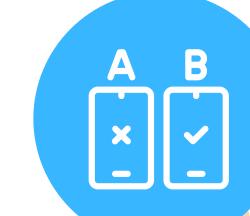
3. TRANSFORMAÇÃO



Em empresas
pequenas..

**GERENCIAMENTO E
ARMAZENAMENTO**

**5. APRENDIZAGEM E
OTIMIZAÇÃO**



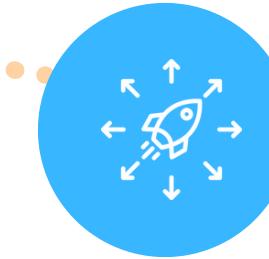
**4. EXPLORAÇÃO
E ANÁLISE**



7. MLOPs



**6. DEPLOY E
SERVING**

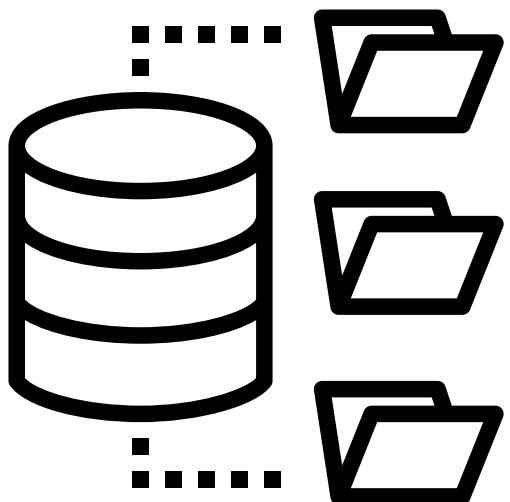


SKILLS: MACHINE LEARNING, ESTATÍSTICA, MATEMÁTICA, PESQUISA, COMUNICAÇÃO, ENG DE SW, ANÁLISE DE DADOS.

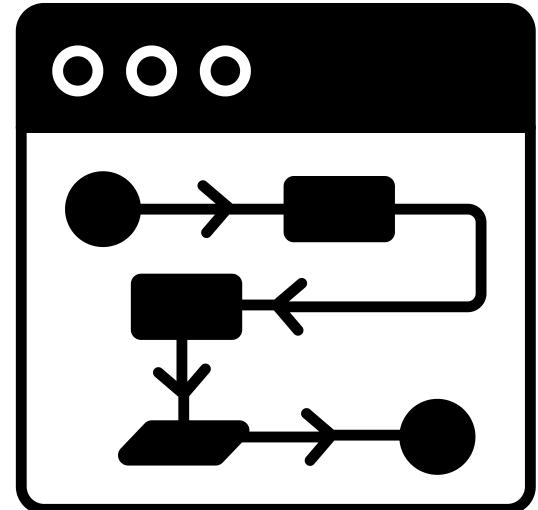


LOGO.

SE VC GOSTA DE..



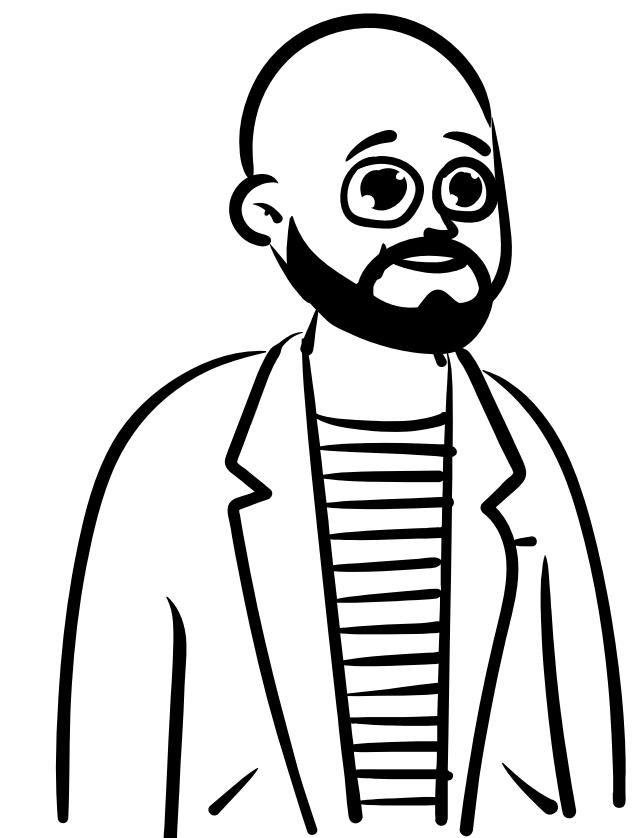
Banco de
Dados



Algoritmos



Python/Scala/Java

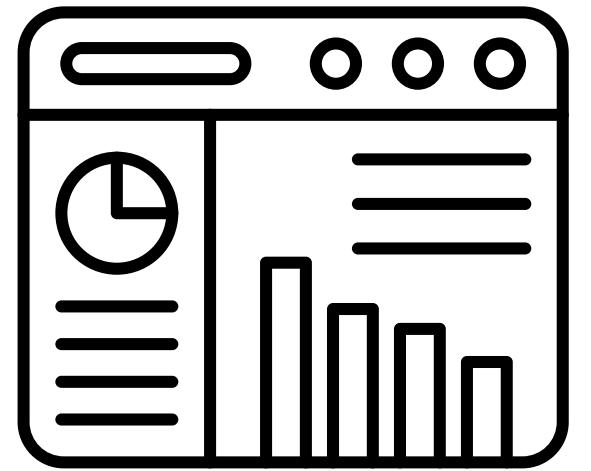


DATA ENGINEER

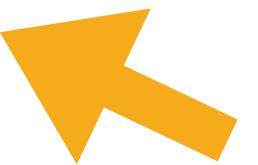


Big Data

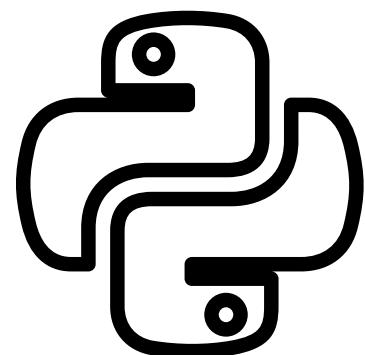
SE VC GOSTA DE..



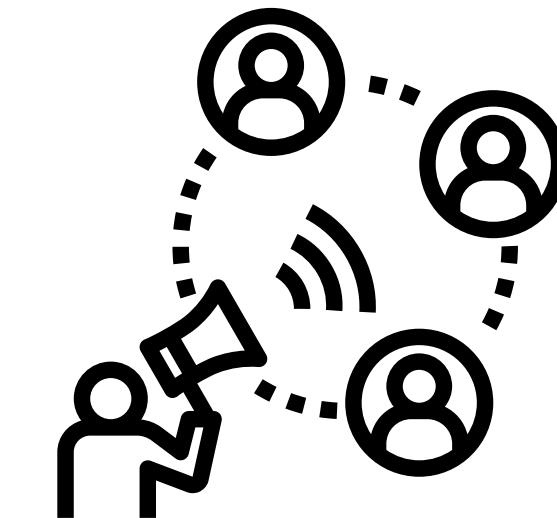
Dashboard



Investigar dados



Python/R



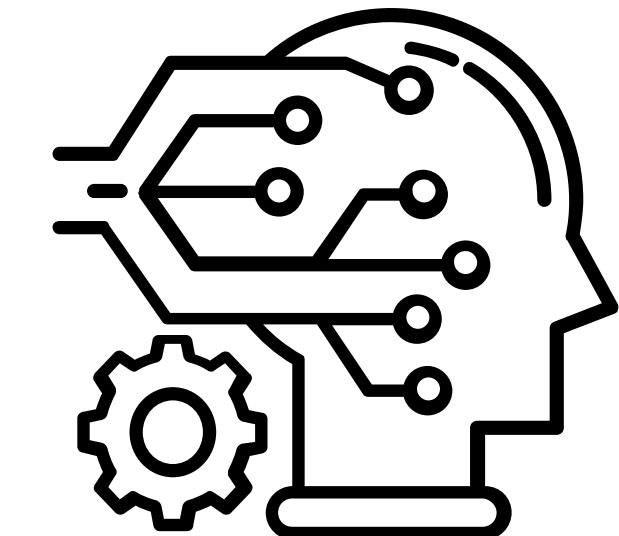
Comunicação/Negócio

DATA ANALYST

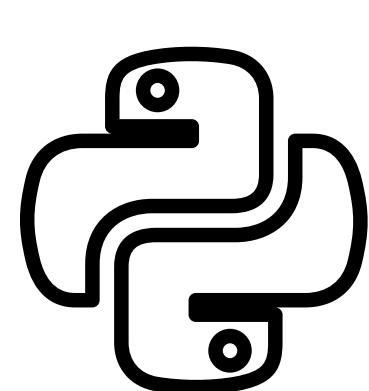
SE VC GOSTA DE..



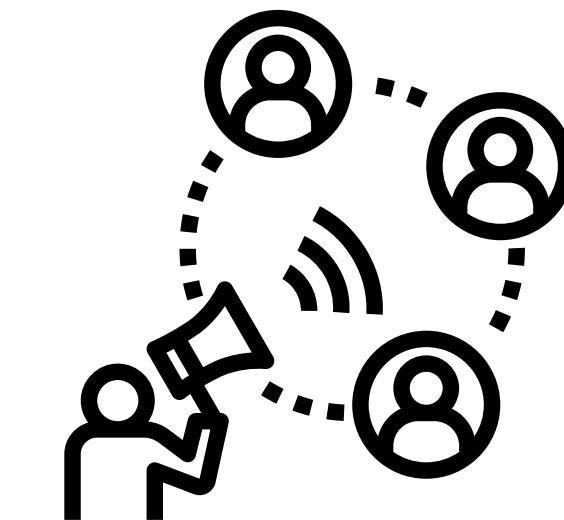
Estatística



Machine Learning



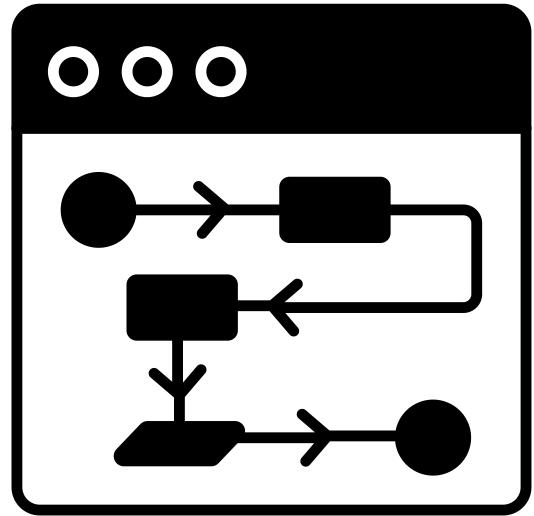
Python/R



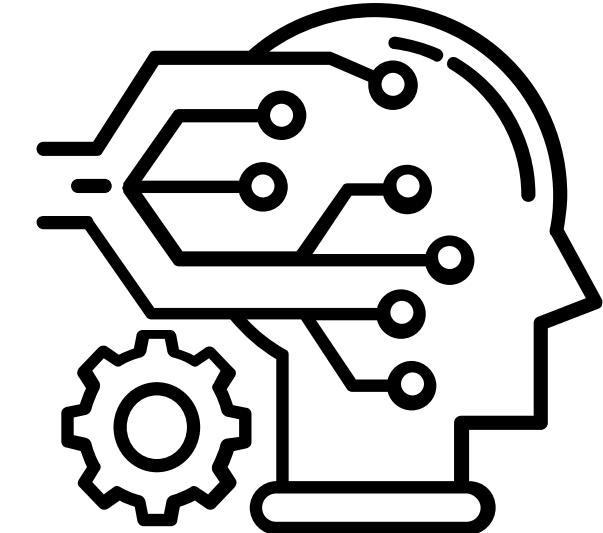
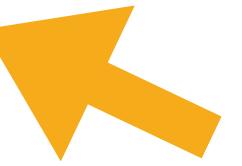
Comunicação/Negócio

DATA SCIENTIST

SE VC GOSTA DE..



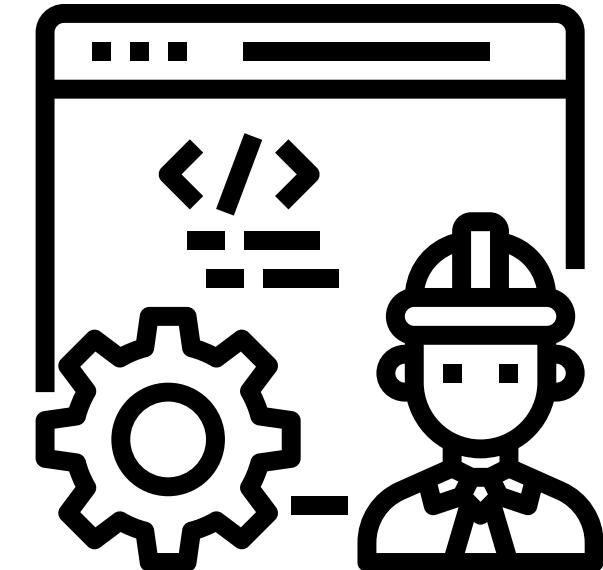
Algoritmos



Machine Learning



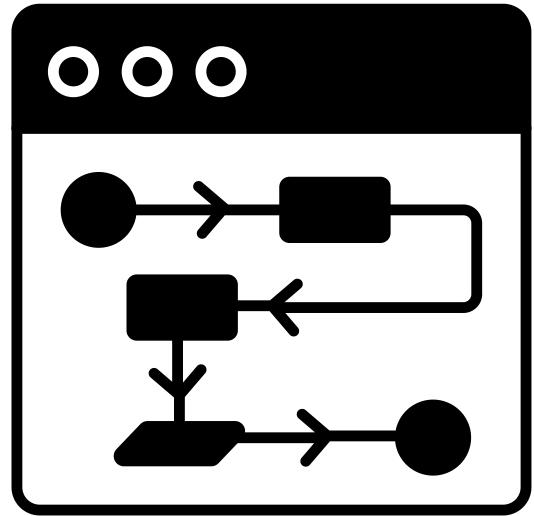
Python/Java/C++



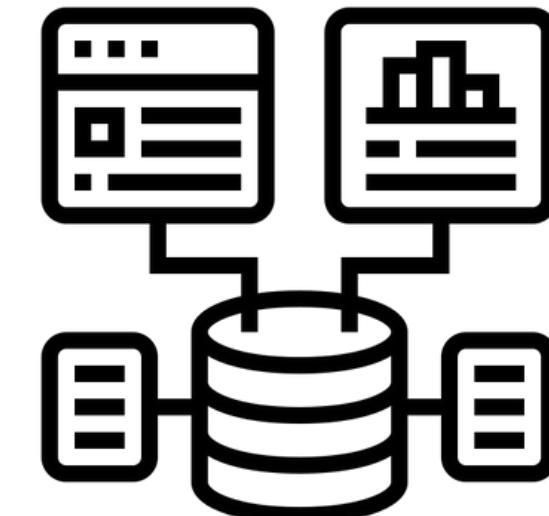
Engenharia de Software

MACHINE LEARNING ENGINEER

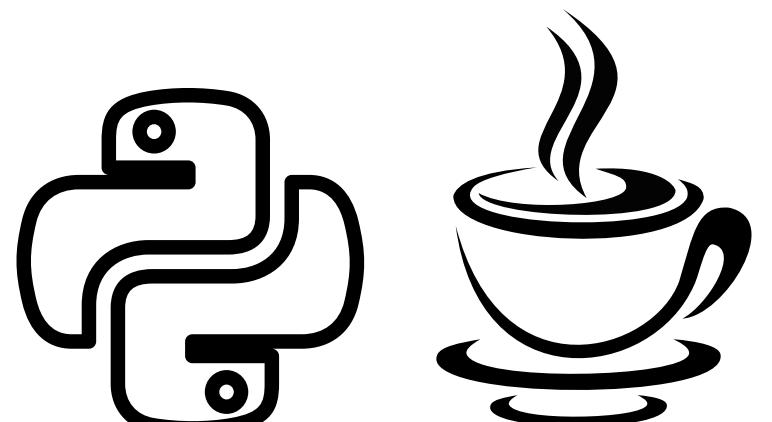
SE VC GOSTA DE..



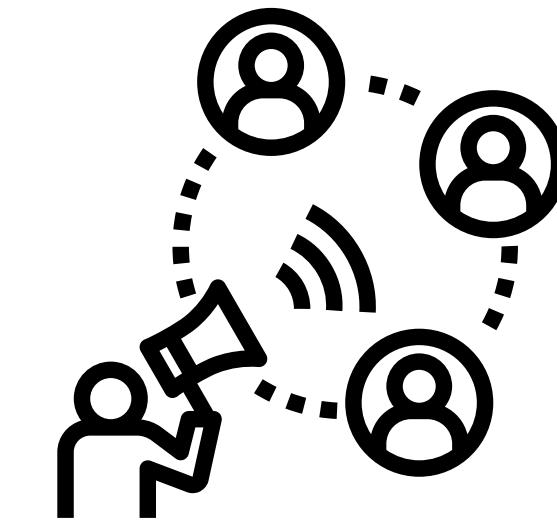
Algoritmos



Big Data



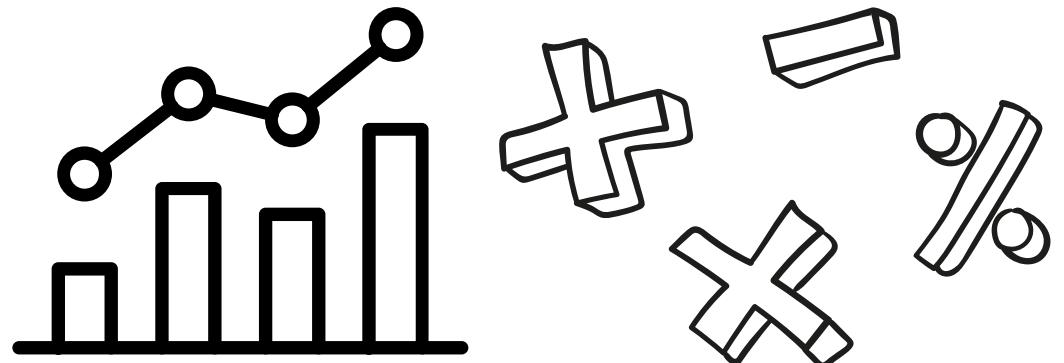
Python/Java



Comunicação/Negócio

ANALYTICS ENGINEER

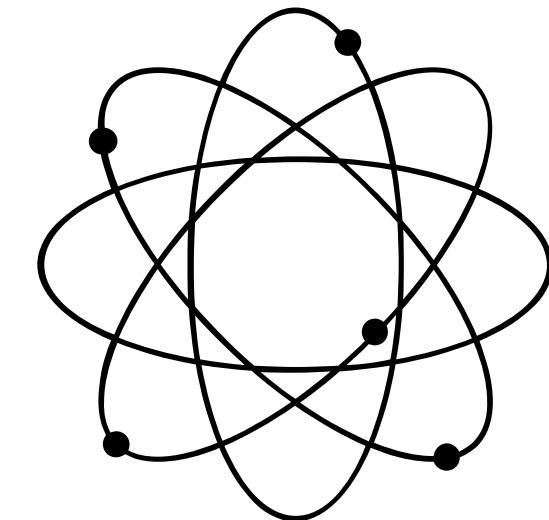
SE VC GOSTA DE..



Estatística/Matemática



doutoranda sofrida



Ciência/ML



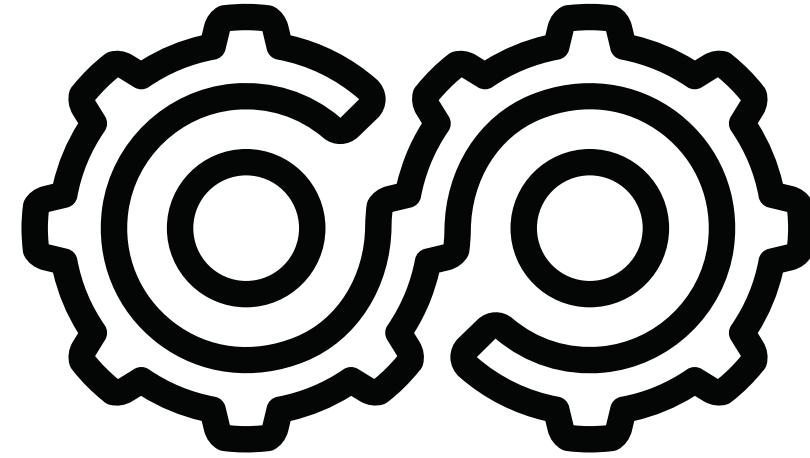
Python/R



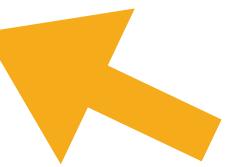
Escrever Artigos

AI RESEARCHER

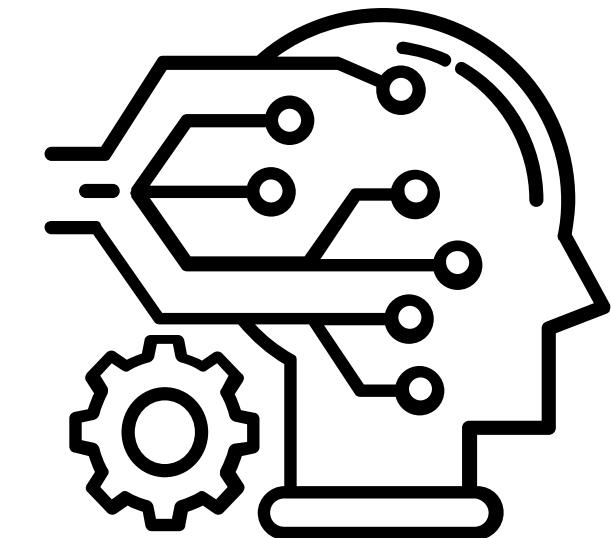
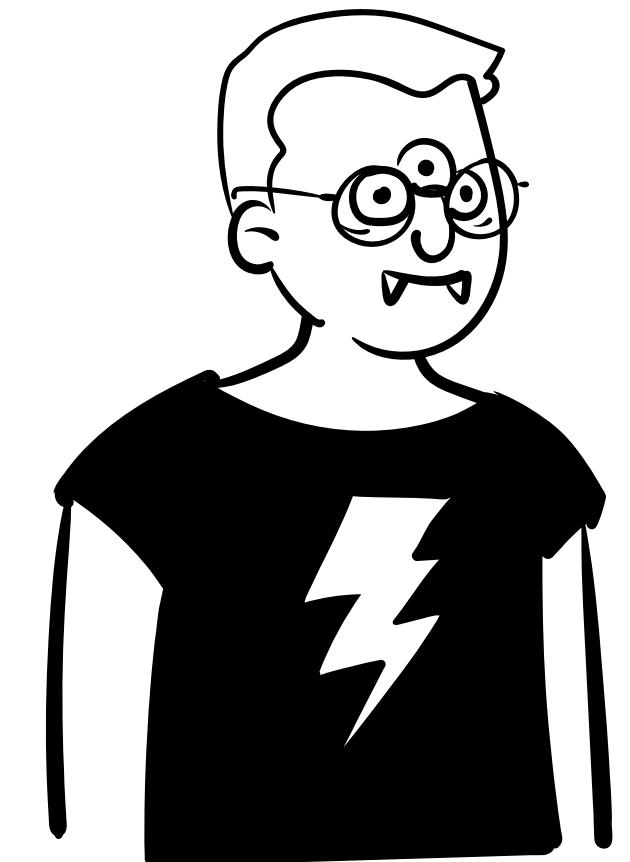
SE VC GOSTA DE..



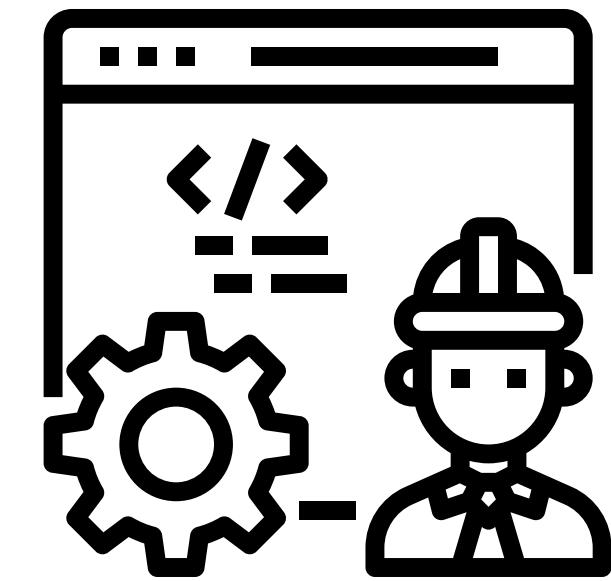
Devops



Infra (Aws, GCP,
Azure, etc)



Machine Learning



Engenharia de Software

MLOPS ENGINEER

EXTRA: MITOS NA CARREIRA DE DADOS

**1. VOU APLICAR REDES NEURAIS EM TODOS
OS PROBLEMAS.**



2. É NECESSÁRIO TER PHD PARA SER DATA SCIENTIST.

**3. DATA SCIENTIST É A VAGA QUE MAIS
PAGA BEM NA ÁREA DE DADOS.**

SPOILER: MLE (USA)

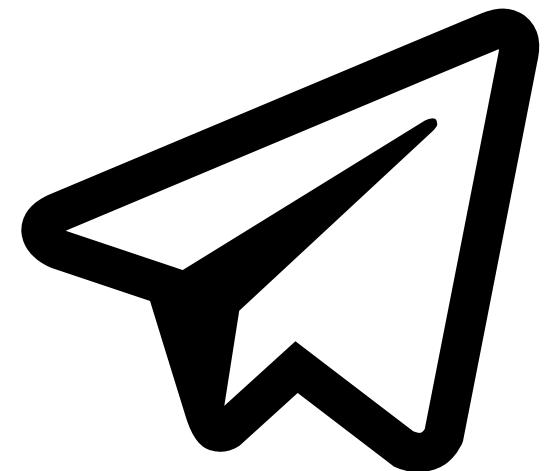
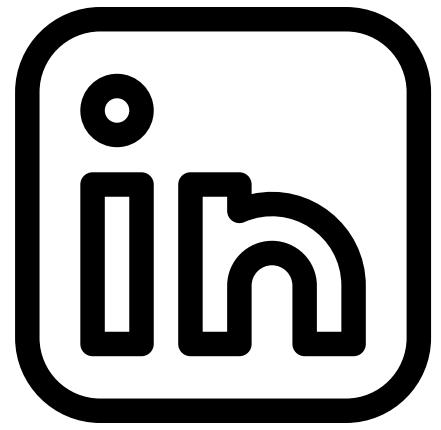
**4. DATA SCIENTISTS SÃO ENGENHEIROS DE
SOFTWARES.**

5. PYTHON SEMPRE É OBRIGATÓRIO PARA CARREIRA DE DADOS.

NÃO ESQUEÇA DO JAVINHA, SCALA, C++, R

6. DATA SCIENTIST EM 2023 É SÓ XETGPT.

7. FAZER DEPLOY DE MODELO É FÁCIL.



@KTAKANOPY

FIM

