
Advanced Analytics Canvas

Data Storytelling

Luiz Mendes
3 de Maio 2018

Olá!

Meu nome é Luiz Felipe Mendes



Luiz Felipe Mendes
Data Scientist e
Co-fundador da Hekima

CONTATOS

luiz.mendes@hekima.com

REDES SOCIAIS

twitter.com/lfmendes

linkedin.com/in/lfomendes

instagram.com/bruce__frenchie

FORMAÇÃO

Mestrando em Ciência da Computação

ESPECIALIDADES

Aprendizado de Máquina

Sistemas de Recomendação

Data Storytelling

Cinema

Olá!

Meu nome é Luiz Felipe Mendes



Luiz Felipe Mendes
Data Scientist e
Co-fundador da Hekima

CONTATOS

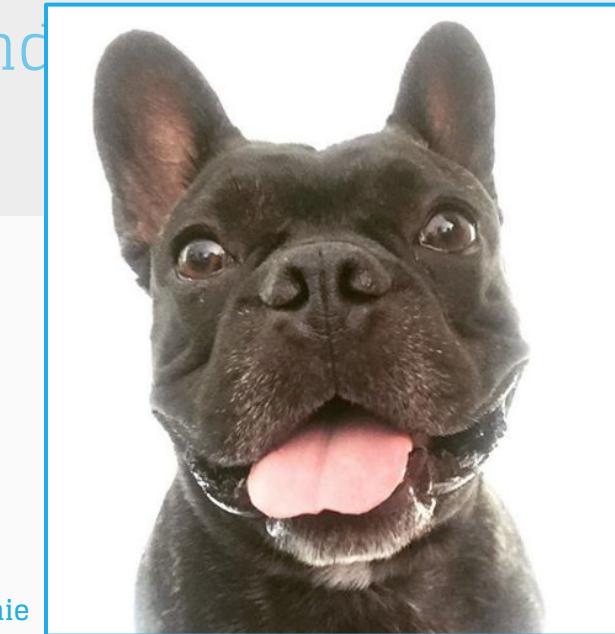
luiz.mendes@hekima.com

REDES SOCIAIS

twitter.com/lfmendes

linkedin.com/in/lfomendes

instagram.com/bruce_frenchie



Cinema

Sobre a Hekima

HEKIMA

Somos uma empresa de Big Data e Inteligência Artificial.

Aceleramos a transformação data-driven de organizações por meio da utilização estratégica e inteligente de dados.

**in
data
we
trust**



PIONEIRISMO

Uma das primeiras empresas brasileiras a construir produtos com Inteligência Artificial na nuvem.

LIDERANÇAS

Ivan M. Campos / COB



Rachel Horta / President CA



Marcia Asano / CEO



Luiz Temponi / President CFO



Thiago Cardoso / CTO



Carolina Bigonha / COO

NOVE ANOS DE EXPERIÊNCIA

desde 2008, aplicando tecnologias de *Big Data* e *Machine Learning*.

PHD'S, MESTRES E ESPECIALISTAS

em Ciência de Dados, Inteligência Artificial, Engenharia de Dados, Mineração de dados, Redes Complexas, Qualidade de Dados, Governança, Estatística, Matemática e outros.

Introdução

Objetivo

Workshop em forma de desafio para que consigam identificar demandas e criar projetos de soluções baseadas em Big Data Analytics e Aprendizado de Máquina.

Negócios + Desenvolvimento

Agenda

Apresentação do **Advanced Analytics Canvas**

Caso de uso do Canvas

Data **Storytelling**

Construção do **Pitch**

CANVAS

Advanced Analytics Canvas

Desafio

Pessoas

Dados

Análise

Ação

Ética de
dados

Desafio

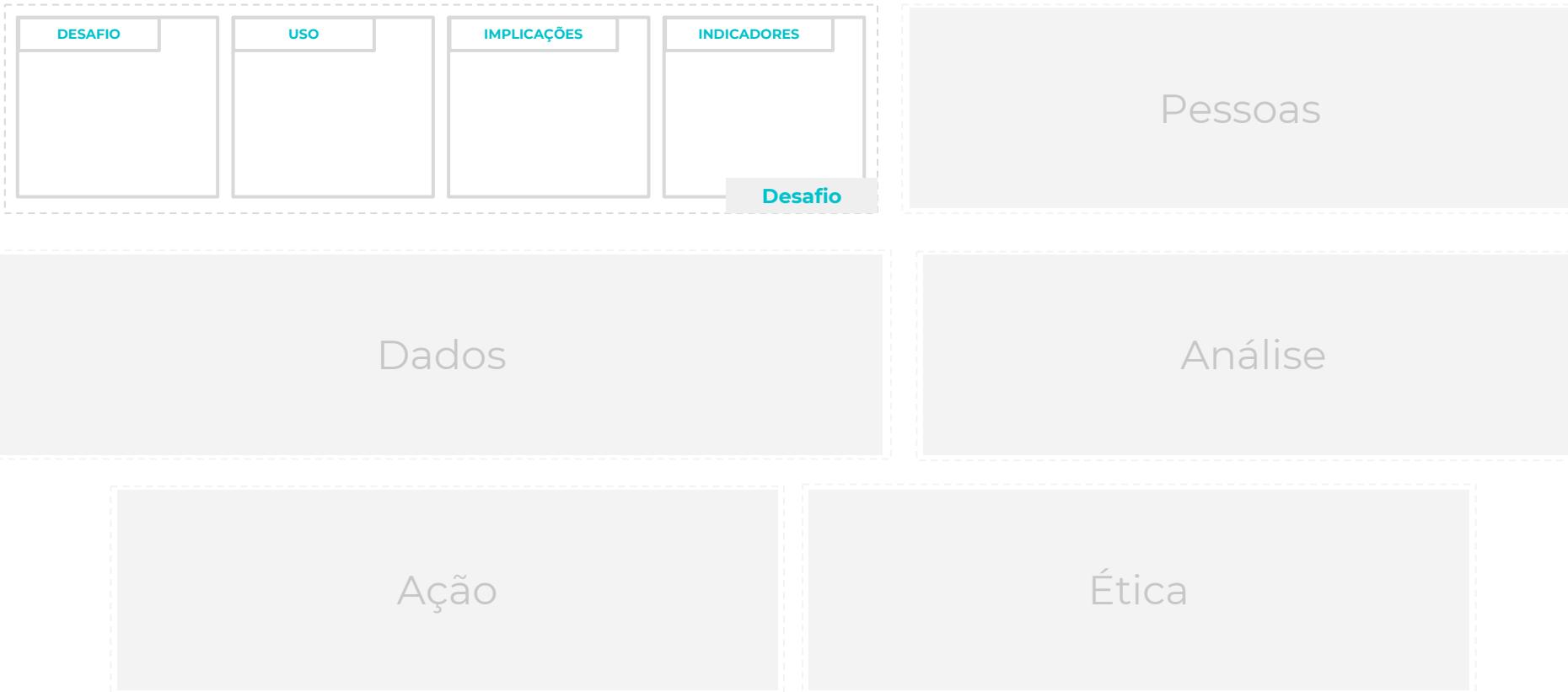
Pessoas

Dados

Análise

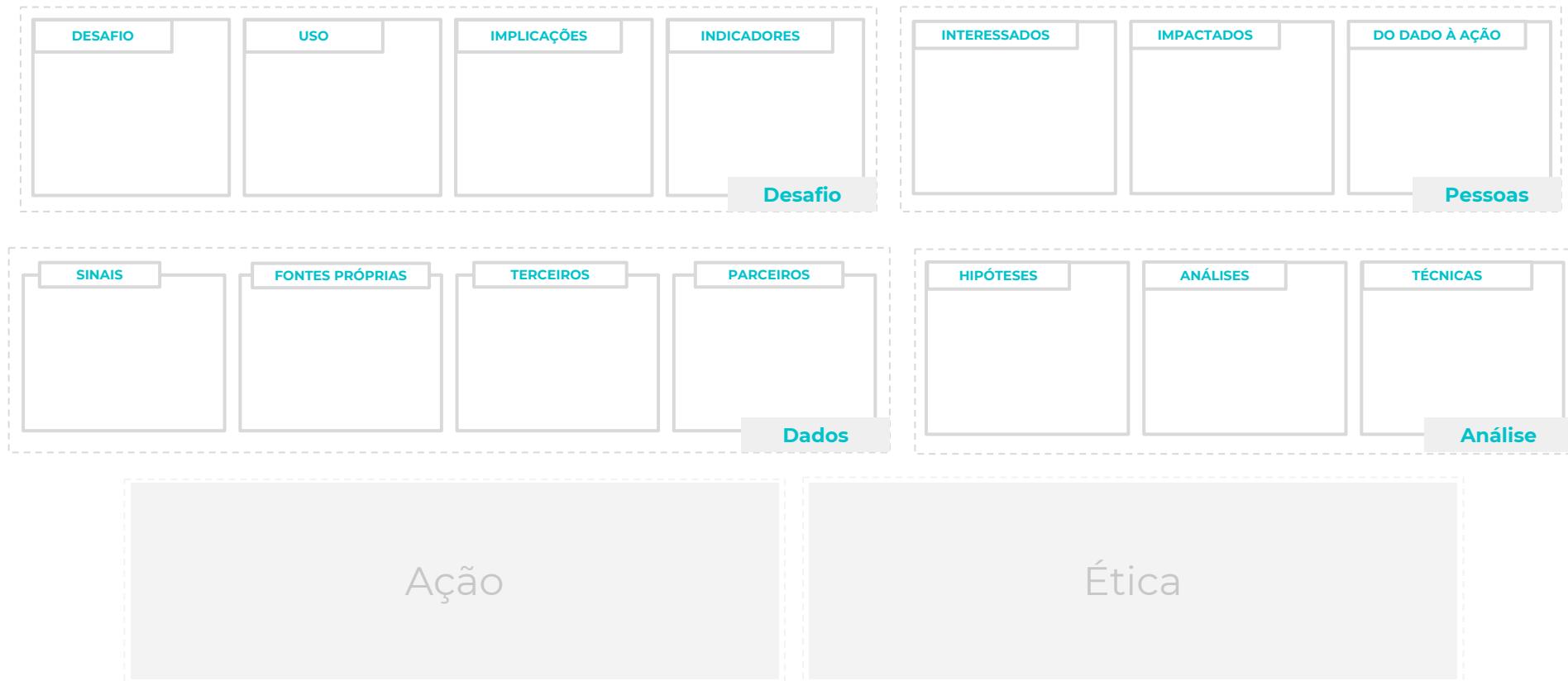
Ação

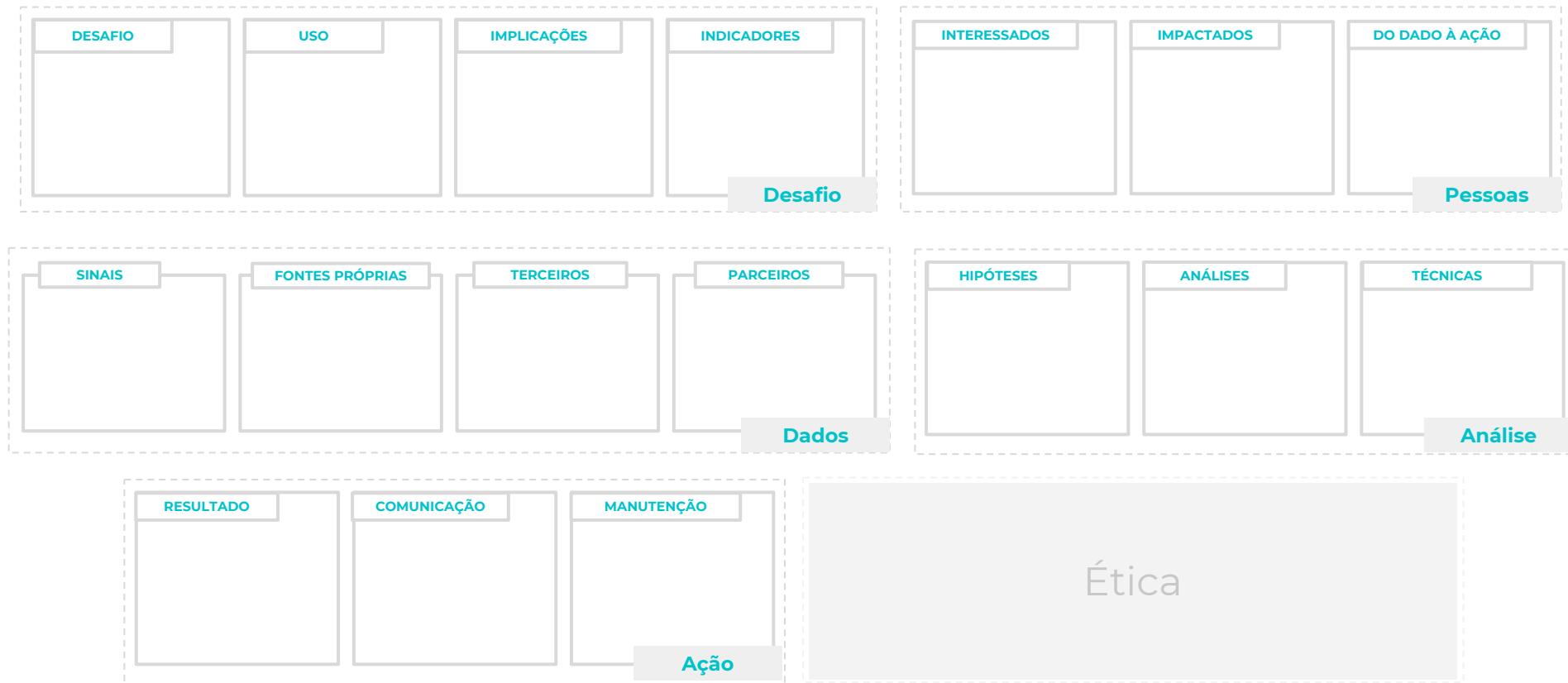
Ética

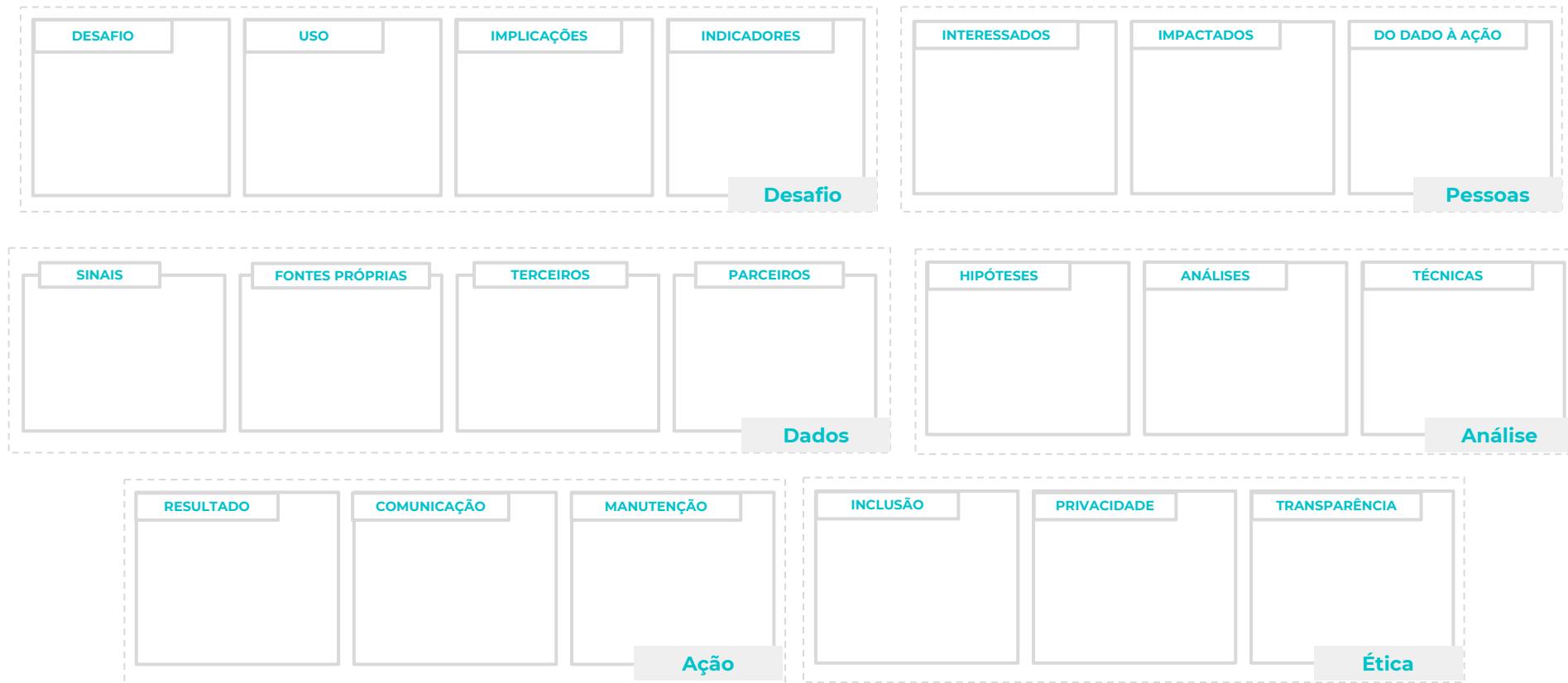












Análise e prevenção de
CHURN

Análise e Prevenção de Churn

- É comum que o custo de aquisição de novos clientes seja muito maior do que o custo de mantê-los
- Para manter ou aumentar a participação de mercado, o ideal é assegurar que clientes atuais permaneçam na carteira.
- Identificação e entendimento dos clientes + comportamento
- Trabalho de retenção:
 - Diversificar produtos e serviços
 - Melhorar segmentação e ações
 - Melhorar atendimento
 - Identificar o melhor mix de marketing
 - ...
- **Como analisar e prevenir evasão de clientes?**

Análise e Prevenção de Churn

Tipos de churn

Ativo - Tipo de evasão em que o cliente escolhe cancelar o vínculo com o serviço

Cancelamento - O aluno/cliente sai no meio do ano/contrato

Renovação - O cliente/aluno não renova o contrato

Reativo - Tipo de evasão em que o cliente não se lembra ou se importa em atualizar os dados cadastrais. Dessa forma, sua conta é cancelada.

Happy Churn - O cliente ou aluno termina o ciclo de uso do produto ou serviço (o aluno se forma, por exemplo)

Falso Churn - O cliente faz uso de um período de teste mas escolhe não continuar com o serviço

Análise e Prevenção de Churn

- Empresa: **Grupo de educação com várias faculdades**
- Análise e prevenção de **churn** na **educação privada de ensino superior**

Desafio

Pessoas

Dados

Análise

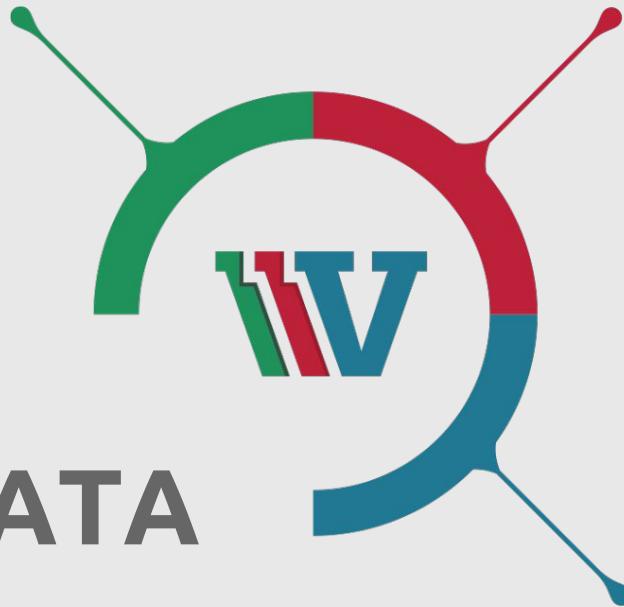
Ação

Ética de
dados

VOLUME

Bigger volume
of data

BIG DATA



VELOCITY

Need for faster
data analysis

VARIETY

Different types of
data, structured or
unstructured

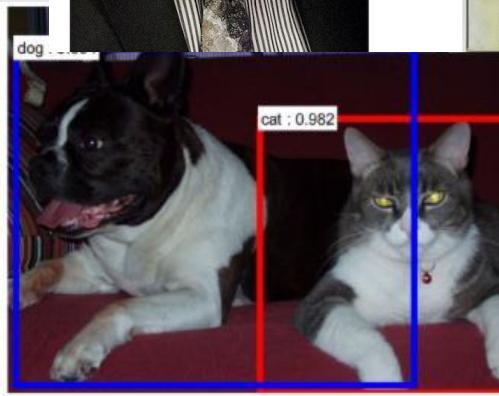
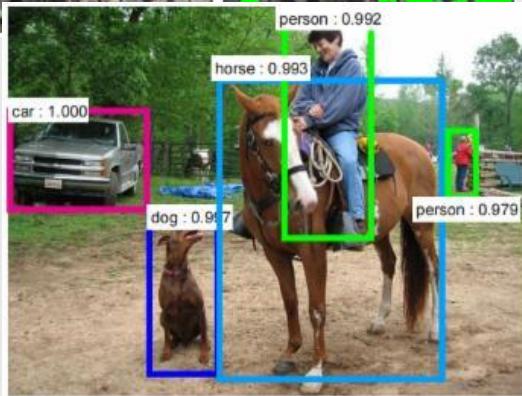


“ Big Data is
not actually
about the
data.



Gary King, professor and
director of the Institute
for Quantitative Social
Science at Harvard.

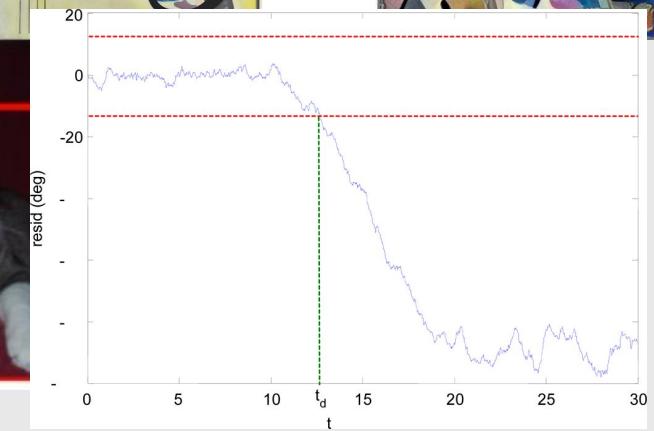
Todos sabem de algo que deve ser automatizado...



+



=



WHY ARE WE DOING
A/B TESTING?



WHY ARE WE PREDICTING
FUTURE TRENDS?



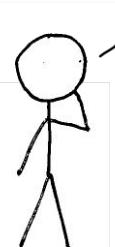
WHY?
WHY?
WHY!



WHY ARE WE ANALYZING
SENTIMENT?



WHY ARE WE CREATING
A RECOMMENDER SYSTEM?



Se você tivesse um modelo / solução com
100% de acurácia, o que faria a seguir?

Desafio

DESAFIO

Qual dor você está tentando resolver?
Como é o fluxo atual?
Como dados podem te ajudar?

USO

Como irão usar a informação/solução?
Como se encaixa no fluxo atual?

IMPLICAÇÕES

Em caso de sucesso:
Ganhos?

Em caso de falha?
Perdas?
Riscos?

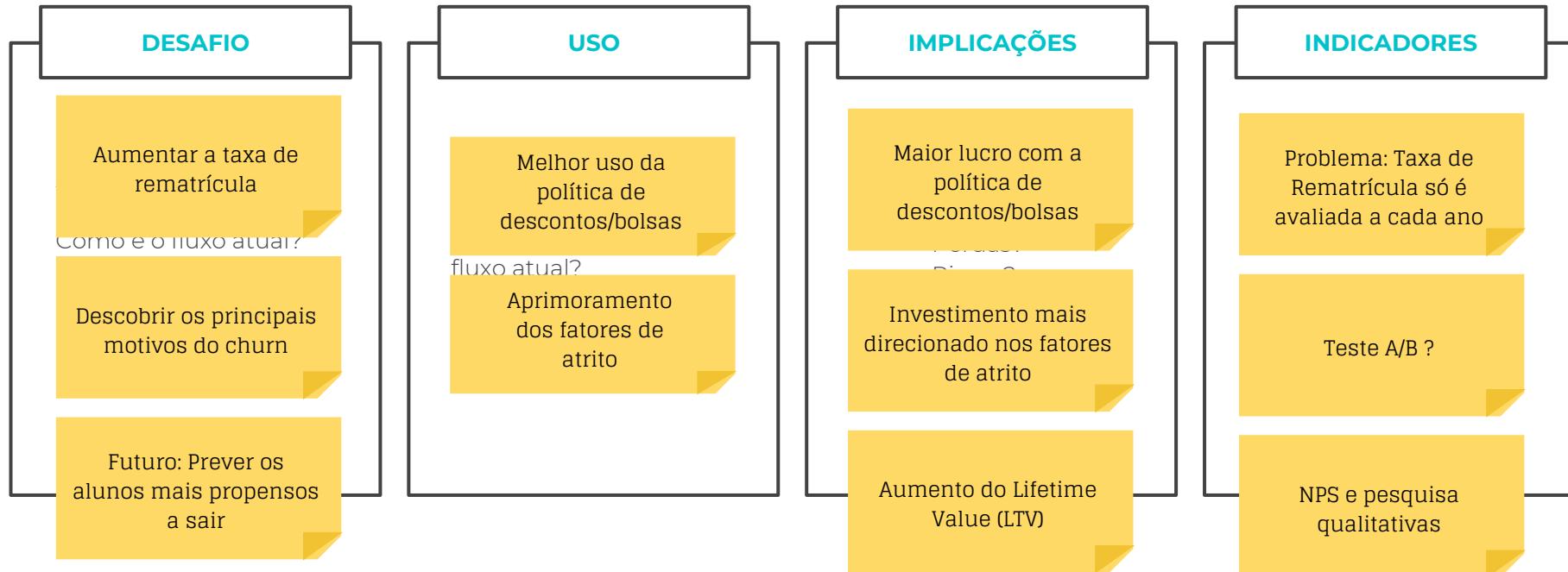
INDICADORES

Como medir o sucesso ou fracasso do projeto?



Desafio

Tempo para preenchimento: **10 minutos**



Desafio

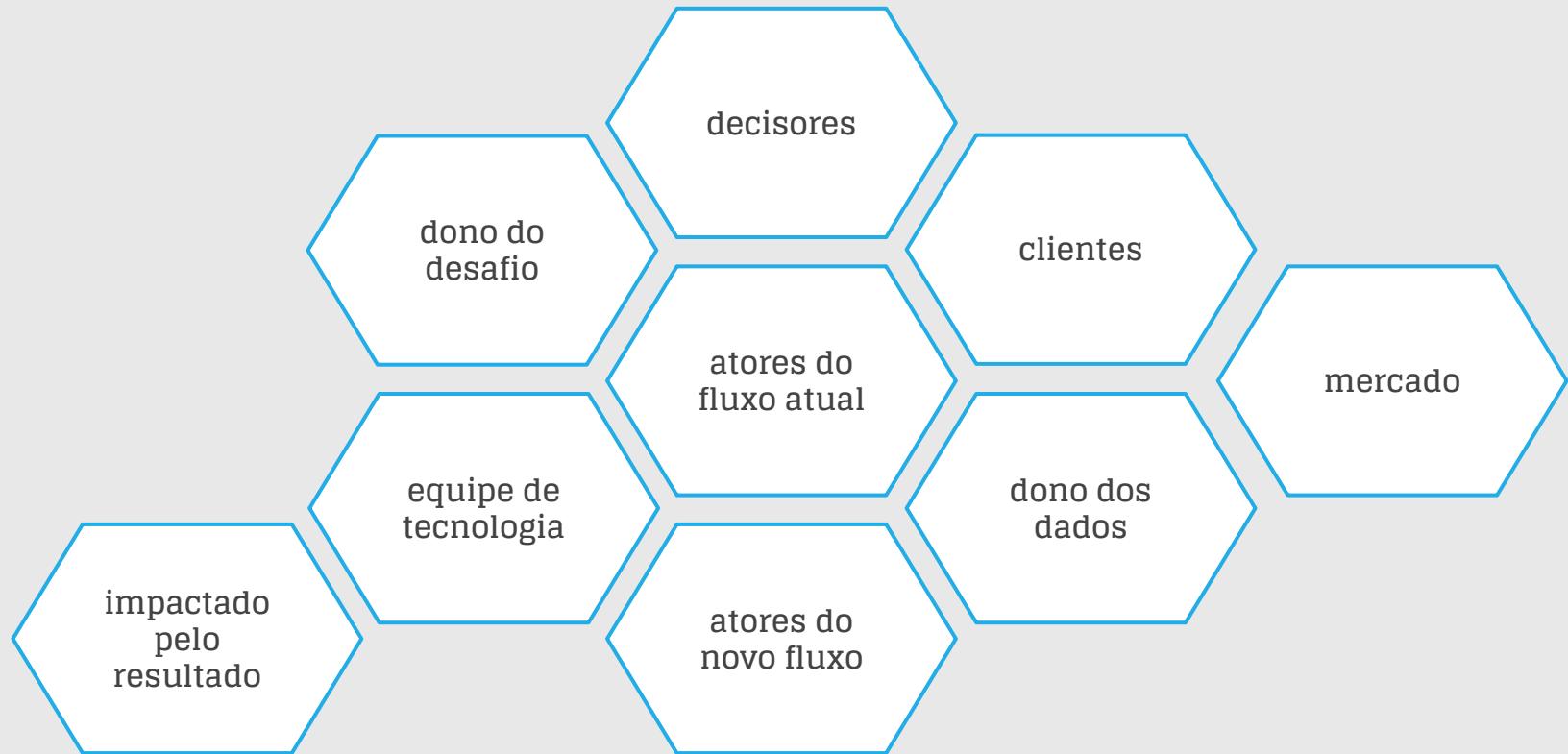
Pessoas

Dados

Análise

Ação

Ética de
dados



Pessoas

INTERESSADOS

Quem são os principais stakeholders interessados?

IMPACTADOS

Quem mais será impactado pela solução?

DO DADO À AÇÃO

Quem vai colocar o resultado em prática?

Qual a cadeia de valor associada ao projeto?



Pessoas

Tempo para preenchimento: **10 minutos**



cadeia de valor data-driven



Desafio

Pessoas

Dados

Análise

Ação

Ética de
dados

Creating a Data-Driven Organization - Carl Anderson



Desafio

Pessoas

Dados

Análise

Ação

Ética de
dados

**Dados empresariais**

CRM

Vendas

Logística

**Dados públicos**

Governo

Mercado

Indicadores

**Ambientes online**

Redes sociais

Blogs

Notícias

**Acervos**

Livros

Fotografia

DADOS INTERNOS

Accessible: available for querying

Coherent: correctly combined

Complete: no missing data

Trustworthy: data is correct

Consistent: between datasets

Defined: unambiguous meaning

Precise: data represents entity state

Relevant: to ongoing analysis

Recent: small time window between fetching and availability

Análise e Prevenção de Churn

- **Variáveis internas:**
 - Dados dos clientes
 - Comportamento em relação aos produtos
 - Ligações & atendimento
 - Meios de pagamento, atraso, inadimplência
- **Variáveis externas**
 - IBGE
 - Contexto político / sócio-econômico
 - Concorrentes

Dados

SINAIS

Quais são os sinais?

FONTES PRÓPRIAS

Quais fontes de dados você já possui?

TERCEIROS

Quais fontes de dados abertos ou fornecidas que poderiam ser utilizadas?

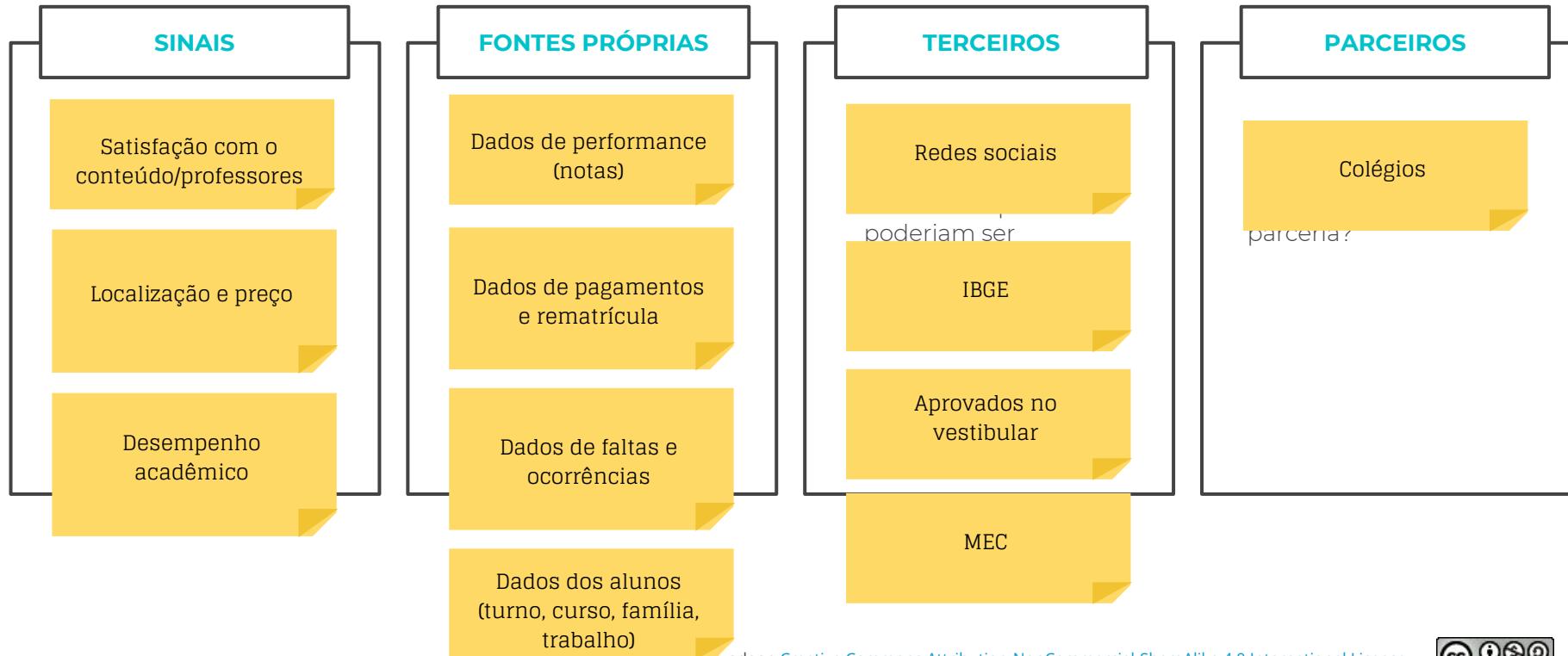
PARCEIROS

Quais fontes de dados estariam disponíveis mediante parceria?



Dados

Tempo para preenchimento: **10 minutos**



Desafio

Pessoas

Dados

Análise

Ação

Ética de
dados

ANÁLISE

- Análise descritiva
- Análise preditiva
- Análise prescritiva

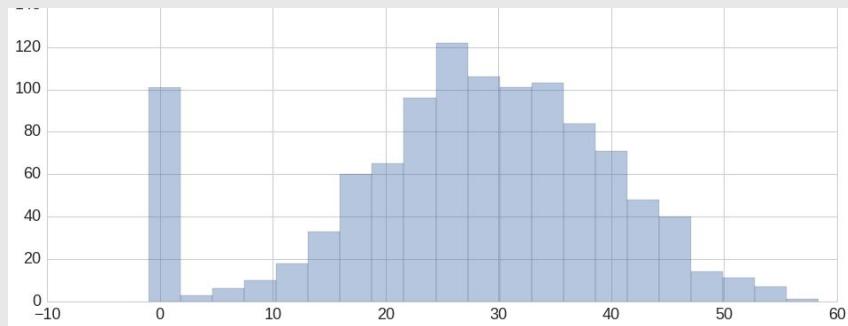
ANÁLISES DESCRIPTIVAS

- Análise exploratória
- Sumários e alertas
- Entendimento do passado e tendências
- Monitoramento de indicadores
- Normalmente feitas em BI tradicional

ANÁLISE EXPLORATÓRIA



- Box Plots
- Histograms
- Scatterplots
- Correlações
- Dicionários de dados



Histograma de idade (dados reais)

ANÁLISE EXPLORATÓRIA

Anscombe Quartet

4 datasets with two variables: x e y

Sample size: 11

Mean of x: 9

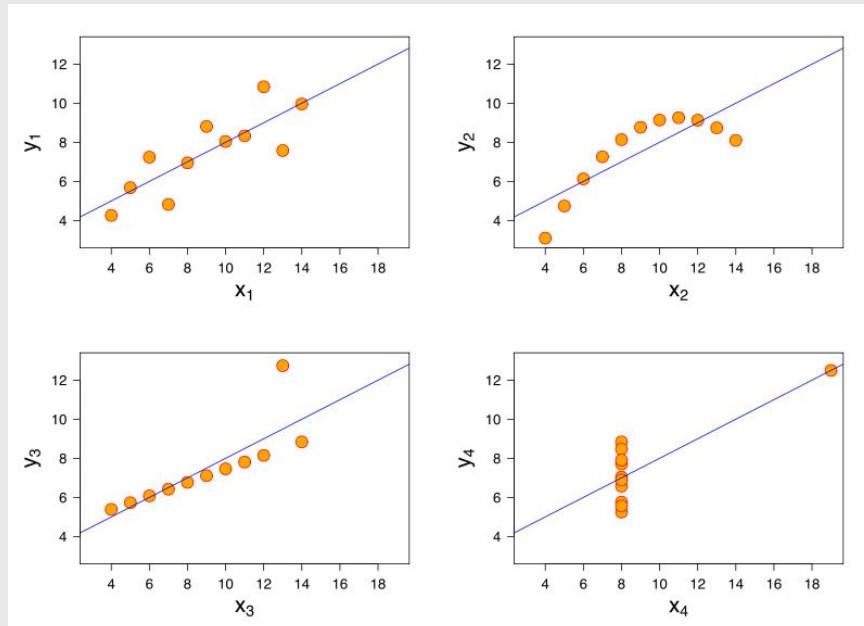
Variance of x: 11

Mean of y: 7.5

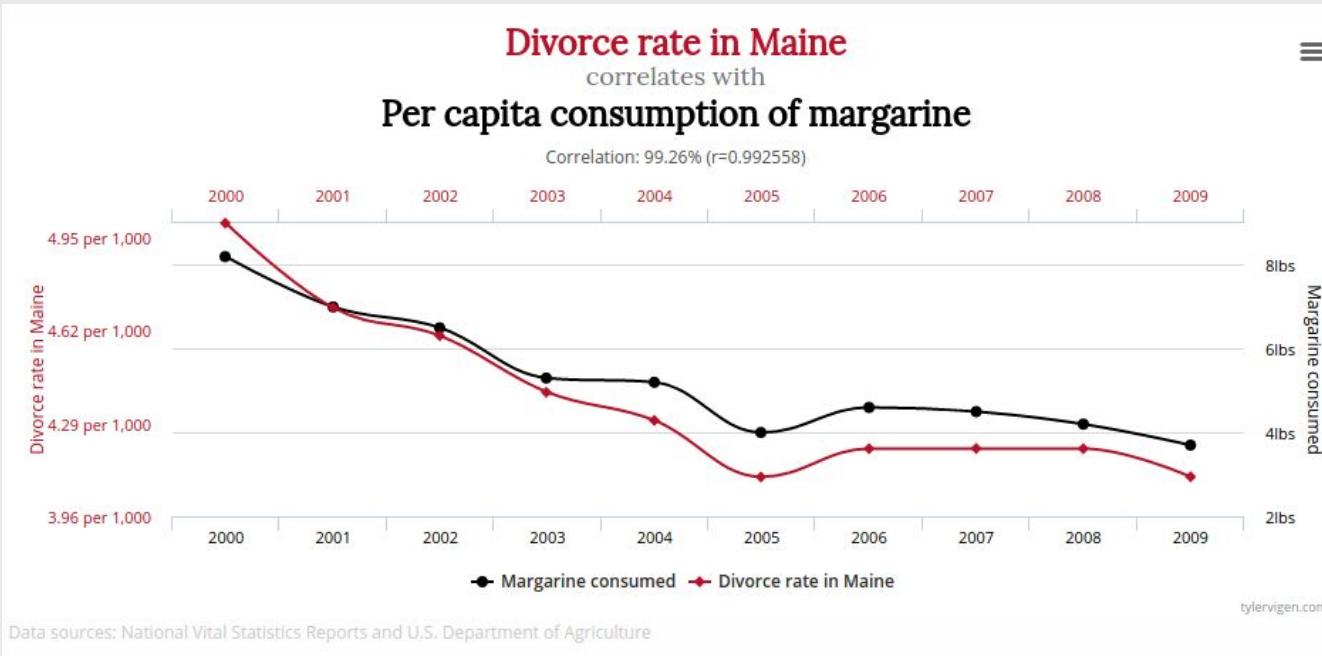
Variance of y: 4.12

Correlation between x e y: 0.816

Linear regression: $y = 4.00 + 0.50 x$



ANÁLISE EXPLORATÓRIA



ANÁLISE DESCRIPTIVA

Relatórios

Descriptivo

O quê?

Olhar para o passado

Gera perguntas

Dados -> Informação

Relatórios, dashboards, alertas

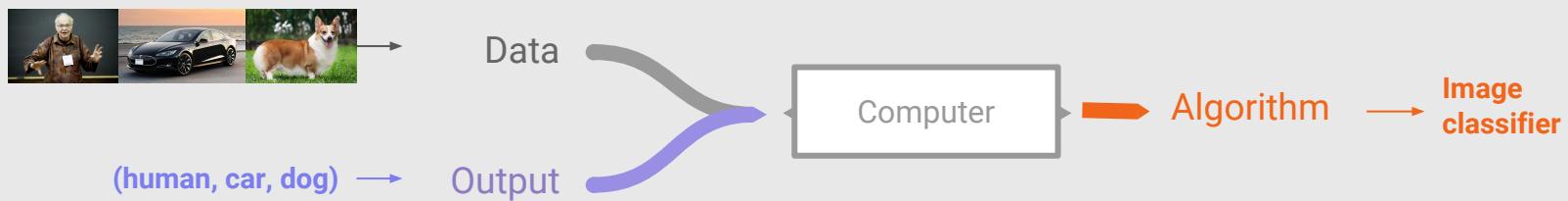
Sem contexto

ANÁLISES PREDITIVA & PRESCRITIVA

Relatórios	Análises
Descritivo	Prescritivo
O quê?	Por quê?
Olhar para o passado	Olhar para o futuro
Gera perguntas	Responde perguntas
Dados -> Informação	Dados + Informação -> Insights
Relatórios, dashboards, alertas	Descobertas, Recomendações
Sem contexto	Contexto + storytelling

MACHINE LEARNING

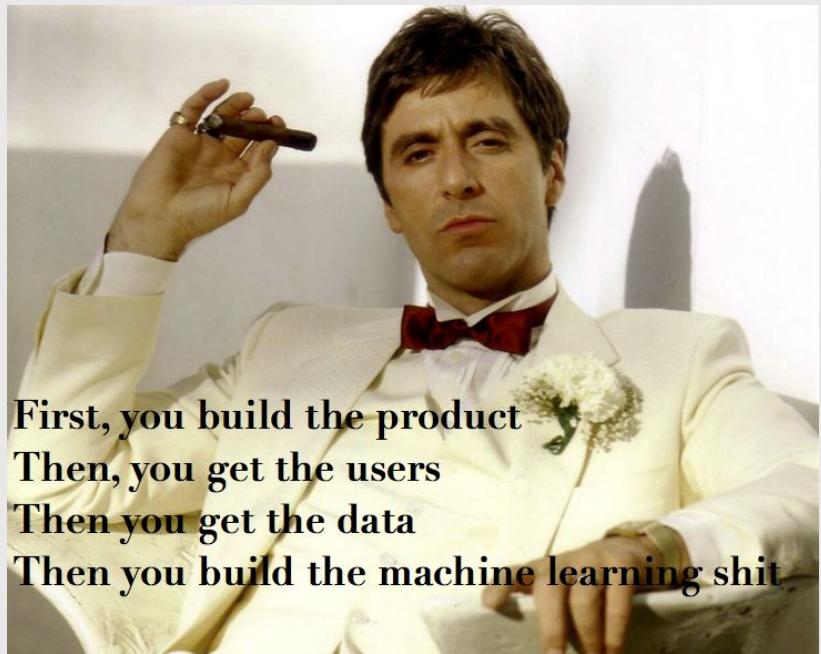
Machine Learning



MACHINE LEARNING

1. Não tenha medo de lançar um produto sem machine learning
2. Use Machine Learning em vez de uma heurística complexa

Tirado do "Rules of Machine Learning: Best Practices for ML Engineering" -- Google



Análise e Prevenção de Churn

- **Descriptivo**

- Quais são os atributos associados ao churn?
- Quais os grupos de churn (segmentação dos clientes)?
- Há efeitos sazonais e de tendência?
- Qual o custo/benefício das ações de prevenção?

- **Preditivo**

- Qual a propensão dos meus clientes ao churn?
- Para cada momento, quem tem maior risco?
- Dada uma ação, qual a chance de reverter?

- **Prescritivo**

- Qual a ação mais recomendada para cada grupo, cliente, motivo?
- Alertas para probabilidade de churn?

Análise

HIPÓTESES

Quais são as principais hipóteses de análise?

Qual será a abordagem?

Qual o primeiro palpite?

ANÁLISES

Para cada hipótese, quais análises serão necessárias?

Análise descritiva

Análise preditiva

Análise prescritiva

TÉCNICAS / MÉTODOS

Quais serão as técnicas empregadas para cada análise?



Análise

Tempo para preenchimento: **10 minutos**

HIPÓTESES

Quais fatores mais influenciam a não rematrícula e a saída?

Alunos que têm um desempenho ruim no curso tendem a sair?

Os alunos tendem a sair mais no início do curso?

ANÁLISES

Análise comparativa para descobrir os fatores

Análise previsiva

Modelo para predição com análise de importância das features

TÉCNICAS / MÉTODOS

Análise estatística e segmentação

Métodos baseados em Árvore de Decisão

Desafio

Pessoas

Dados

Análise

Ação

Ética de
dados

“Analytics is about **impact**...

In our company [Zynga], if you have brilliant
insight and you did great research and **no**
one changes, you get **zero credit**.”

Ken Rudin - Head of Analytics @ Facebook

Análise e Prevenção de Churn

Ações

- **Ações**

- Definição de ações de curto e longo prazo e custo de execução
- Mapeamento de ações de prevenção para cada tipo de churn (churner, potential churner, no churner)
- Priorização das ações para cada cliente por custo e risco

RESULTADO & MANUTENÇÃO

Relatório com lista de
clientes mais
propensos a evadir

Dashboard para
acompanhamento do
comportamento dos
alunos

Dashboard de
acompanhamento de
ações de prevenção

Aplicação para
disparo e gestão de
ações

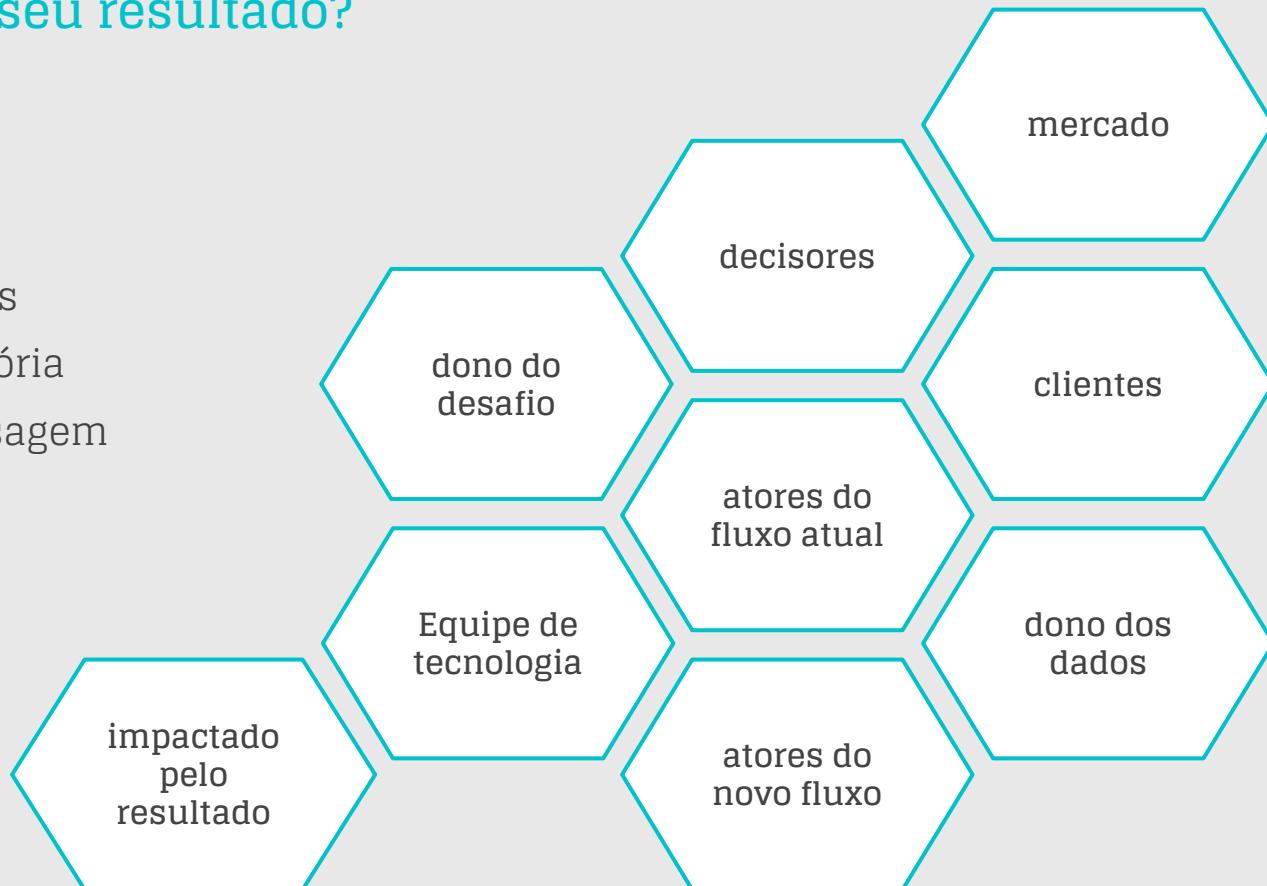
RESULTADO & MANUTENÇÃO

Relatório com lista de clientes mais propensos a evasão	Dashboard para acompanhamento do comportamento do aluno	<ul style="list-style-type: none">• Equipe• Manutenção do modelo• Infraestrutura cloud• Dashboard• Grupo controle
Dashboard de acompanhamento de ações de prevenção	Aplicação para disparo e gestão de ações	

Como comunicar seu resultado?

Ao final do processo

- Mostrar resultados
- Embasar sua história
- Passar uma mensagem



Ação

RESULTADO

Qual o principal resultado?

COMUNICAÇÃO

Como e para quem comunicar o resultado?

RECURSOS

Qual a infraestrutura e manutenção necessárias?
Equipe / skills?
Recursos financeiros?

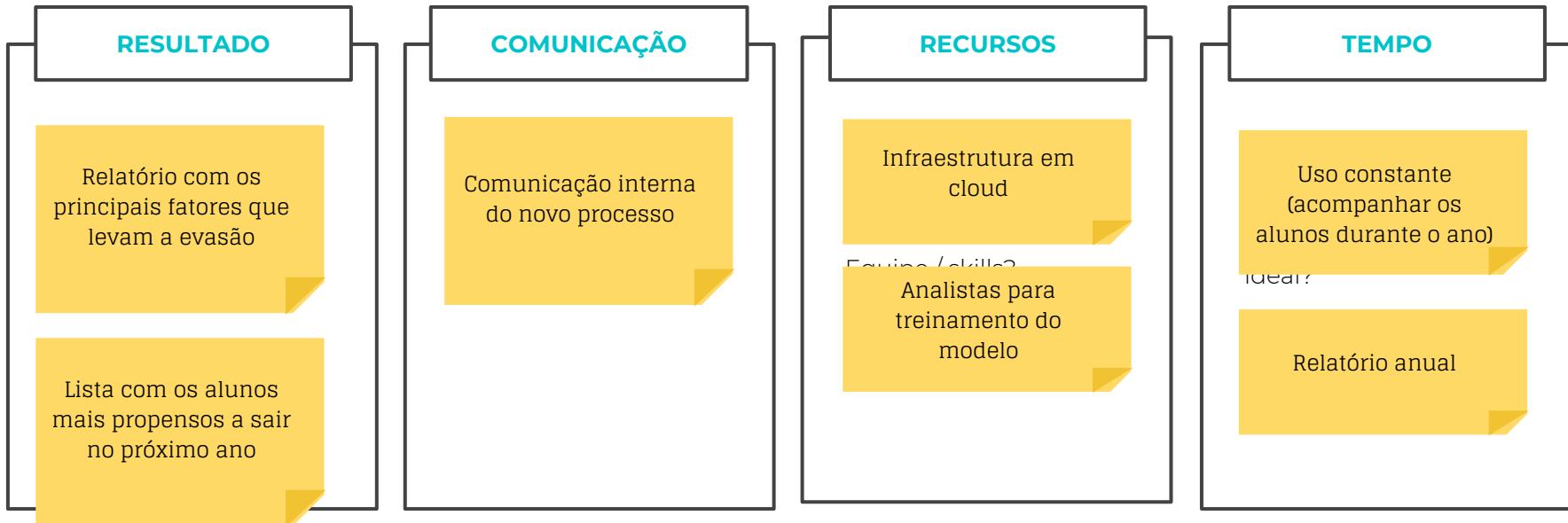
TEMPO

Qual o investimento de tempo?
Quando será o timing ideal?
Qual frequência?



Ação

Tempo para preenchimento: **10 minutos**



Desafio

Pessoas

Dados

Análise

Ação

Ética de
dados



The Menlo Report

Ethical Principles Guiding Information and
Communication Technology Research

August 2012



Homeland
Security

Science and Technology

Conjunto de **princípios básicos**
que guiam a identificação
e resolução de **desafios éticos**
em pesquisa de TIC.

1

RESPEITO
PELAS PESSOAS

2

BENEFICÊNCIA

3

JUSTIÇA

4

RESPEITO POR
LEIS E POLÍTICAS

1

RESPEITO PELAS PESSOAS

- Agentes **autônomos**
- **Proteção** a indivíduos com autonomia reduzida
- Participação **voluntária** e mediante **consentimento**.

2

BENEFICÊNCIA

- **Não faça o mal.**
- **Maximize** potenciais **benefícios** e **minimize** potenciais **riscos**.
- **Analizar sistematicamente** tanto os riscos quanto os benefícios.

3

JUSTIÇA

- Cada pessoa merece **consideração igual.**
- **Benefícios** igualmente distribuídos.
- **Fardos** igualmente distribuídos.
- Seleção **justa** de sujeitos.

RESPEITO POR
LEIS E POLÍTICAS

4

- **Auditoria**
- **Transparência** de resultados e métodos
- **Responsabilidade** por ações

Análise e Prevenção de Churn

- **Preparação dos dados**
 - Não levar em consideração características potencialmente discriminatórias
 - Há consentimento para uso das informações / há limites de privacidade sendo ultrapassados?
- **Definição de sucesso & análise dos resultado do modelo**
 - Meu resultado privilegia algum grupo?
 - Custo x benefício de ações

Ética

INCLUSÃO

Riscos relacionados a representatividade e participação

PRIVACIDADE

Riscos relacionados à privacidade de dados
Riscos relacionados ao consentimento de uso dos dados

TRANSPARÊNCIA

Riscos relacionados à transparência do projeto



Ética

Tempo para preenchimento: **10 minutos**

INCLUSÃO

Na preparação de dados, remover aspectos que indiquem raça e gênero

No resultado, identificar se há algum tipo de viés.

PRIVACIDADE

Permitir que os alunos
se for usar redes sociais,
deixar claro para o aluno
que é necessário o
consentimento de uso dos
dados

Não disponibilizar os dados
dos alunos para pessoas
de fora do projeto

TRANSPARÊNCIA

Pode ser feita uma
pergunta: Existe alguma forma de
deixar claro que o desconto
está atrelado a algum
motivo?

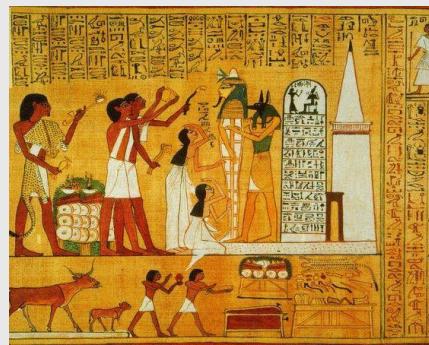
Storytelling

STORYTELLING

“

Transmitir uma história ou
mensagem de forma memorável

STORYTELLING



Storytelling acompanha a humanidade por quase toda sua existência.

STORYTELLING

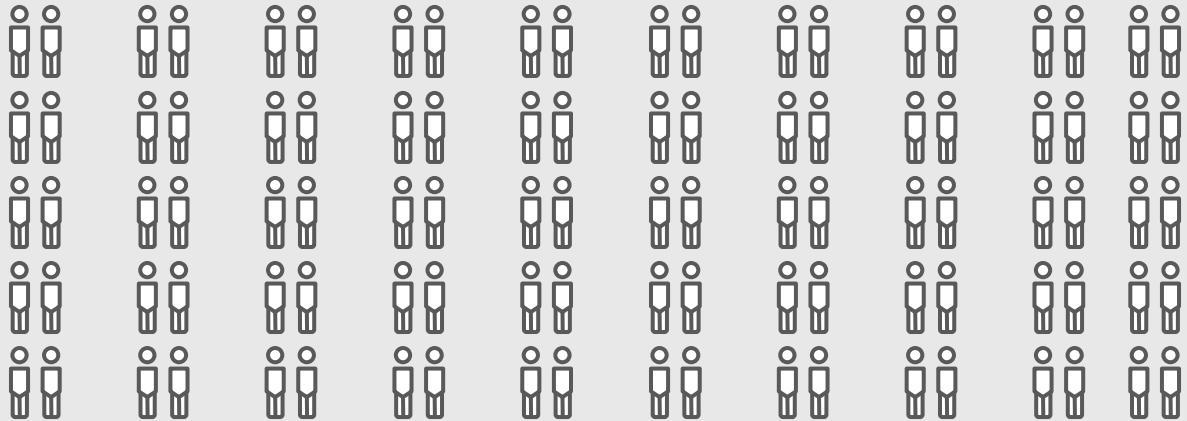


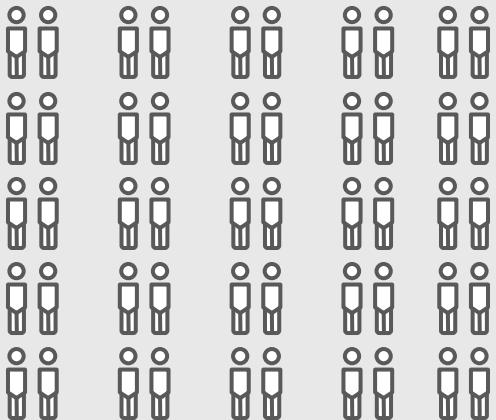
Storytelling acompanha a humanidade por quase toda sua existência.

STORYTELLING



CHIP HEATH

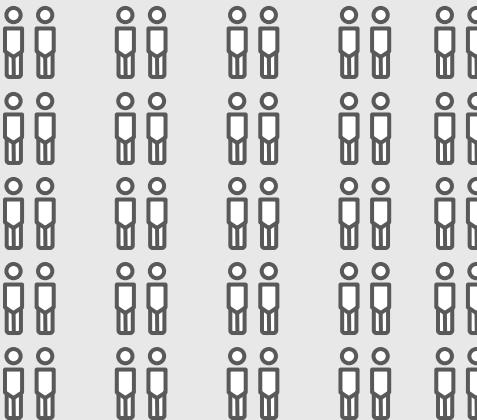




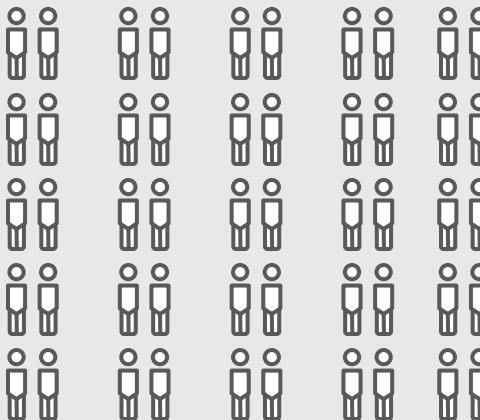
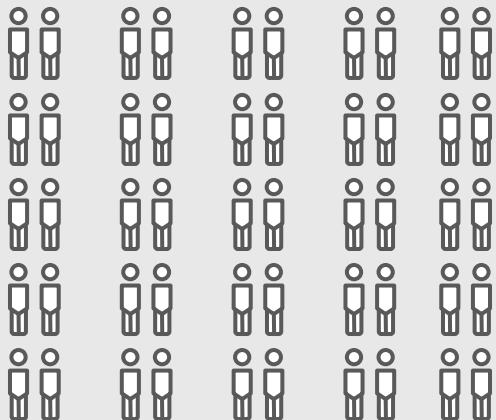
favoráveis



contrários



mesmos dados

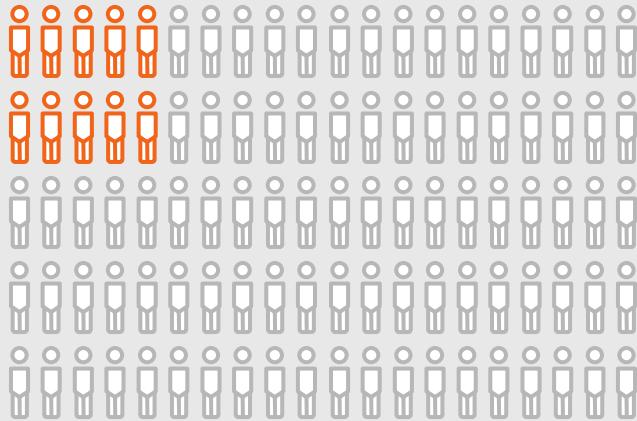


mesmo tempo, 1'



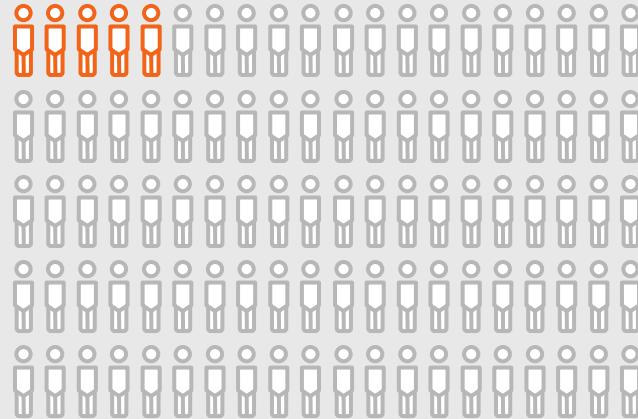
Cada grupo mostrou, em média,
2 a 3 estatísticas

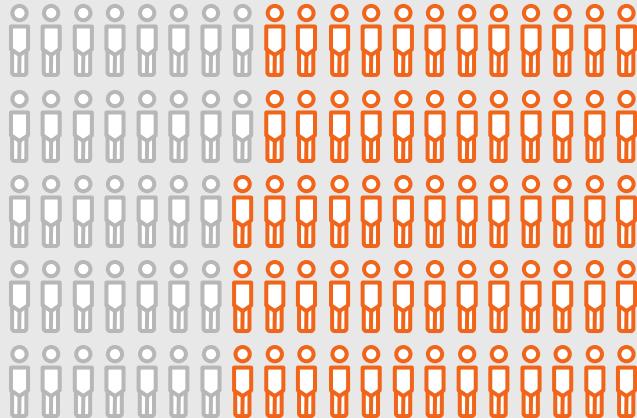
mas apenas 10%
contou histórias





apenas 5%
lembavam das
estatísticas envolvidas,
enquanto...





63%
lembaram
das histórias

STORYTELLING

As três partes de uma história

Primeiro Ato

The set up



O mundo do herói antes
de a aventura começar

Primeiro Ato

The set up



Segundo Ato

The conflict



O mundo do herói antes
de a aventura começar

O mundo do herói é
virado de ponta cabeça

Primeiro Ato

The set up



Segundo Ato

The conflict



Terceiro Ato

The resolution



O mundo do herói antes
de a aventura começar

O mundo do herói é
virado de ponta cabeça

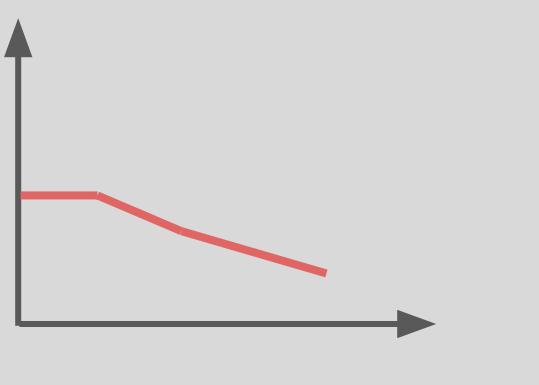
O mundo do herói
é transformado

STORYTELLING

As três partes de uma história com **Data Science**

Primeiro Ato

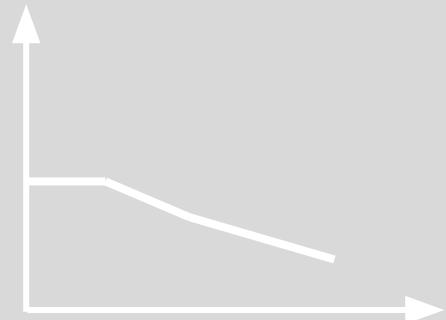
The set up



Estamos perdendo
clientes todo mês

Primeiro Ato

The set up

**Segundo Ato**

The conflict

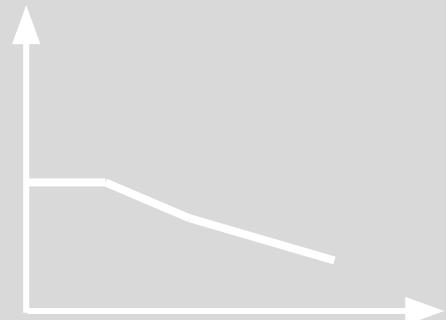
$$\begin{aligned} 2 &> -3 \\ 0.999\dots &= 1 \\ \infty &\approx 3.14 \\ \sqrt{2} &\approx 1.414 \\ 2^1 + 2 \cdot 3 &= 8 \\ 5^2 &= 25 \\ (1 - 2) + 3 &= 0 \\ 101_2 &= 5_{10} \end{aligned}$$

Estamos perdendo
clientes todo mês

Fizemos experimentos e
criamos um algoritmo

Primeiro Ato

The set up



Estamos perdendo
clientes todo mês

Segundo Ato

The conflict

$$\begin{aligned} 2 &> -3 \\ 0.999\dots &= 1 \\ \pi &\approx 3.14 \\ \sqrt{2} &\approx 1.414 \\ 1 + 2 \cdot 3 &= 7 \\ 5^{2^2} &= 25 \\ (1 - 2) + 3 &= 0 \\ 101_2 &= 5_{10} \end{aligned}$$

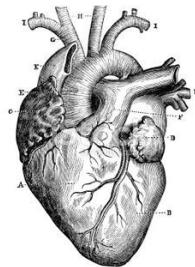
Terceiro Ato

The resolution

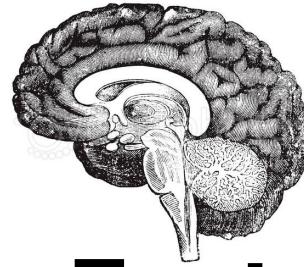
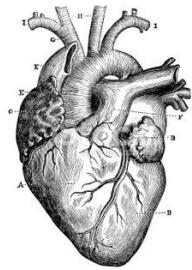


Fizemos experimentos e
criamos um algoritmo

**Se usarmos o algoritmo,
melhoramos a retenção**



Feelings +

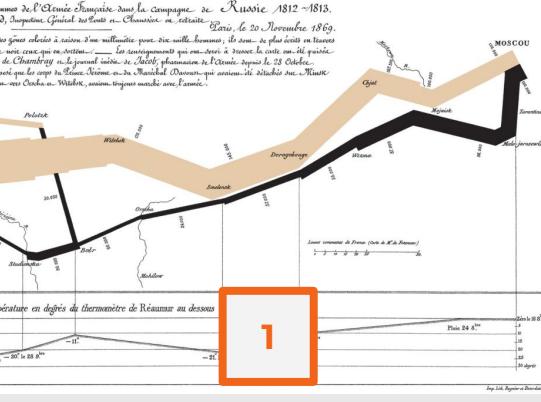


Feelings + Facts

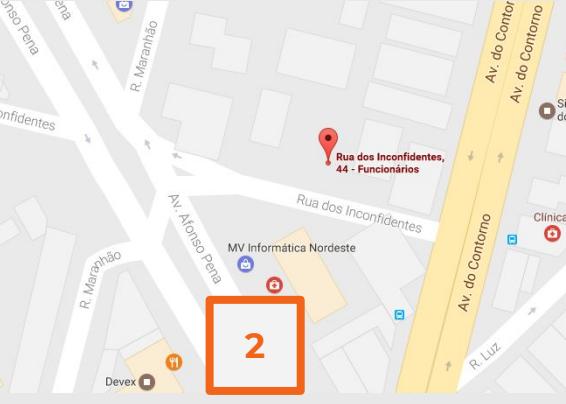
=

Feelings + Facts
= **Data Storytelling**

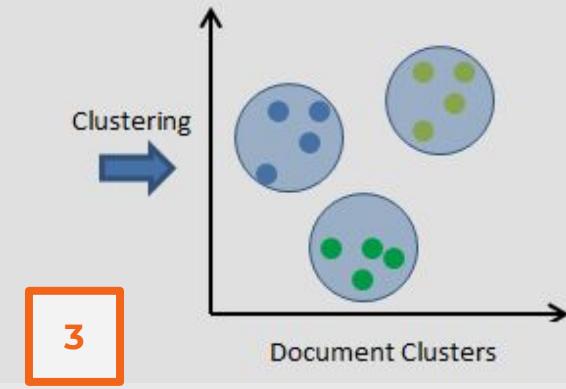




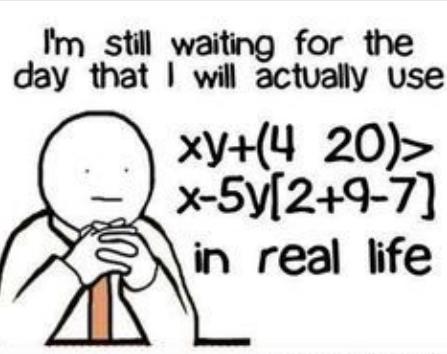
Um método de construir narrativas baseadas em dados.



Uma forma de “humanizar” os dados.

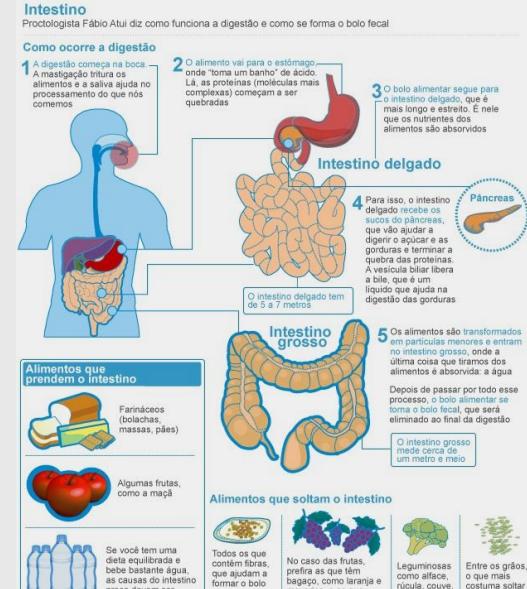


Um instrumento para transformar números, regressões e análises em entendimento.



A grande maioria das pessoas acha que matemática e dados são chatos.

Utilizar de uma boa história para mostrar os resultados democratiza a informação, dando capacidade de todos opinarem e tirarem conclusões.



STORYTELLING

pergunte-se

pergunte aos dados

descubra histórias

apresente

STORYTELLING

pergunte-se

pergunte aos dados

descubra histórias

apresente

quem é seu público?
qual o contexto/meio de comunicação?
qual é a técnica empregada?
qual o objetivo?

STORYTELLING

pergunte-se

pergunte aos dados

descubra histórias

apresente

o que eles oferecem?

qual a relação deles com outros dados?

STORYTELLING

pergunte-se

pergunte aos dados

descubra histórias

apresente

**fatos interessantes
mudanças
conexões
experiências pessoais**

STORYTELLING

pergunte-se

pergunte aos dados

descubra histórias

apresente

**conte uma história
simplifique
foque no que importa**

PITCH



Storytelling

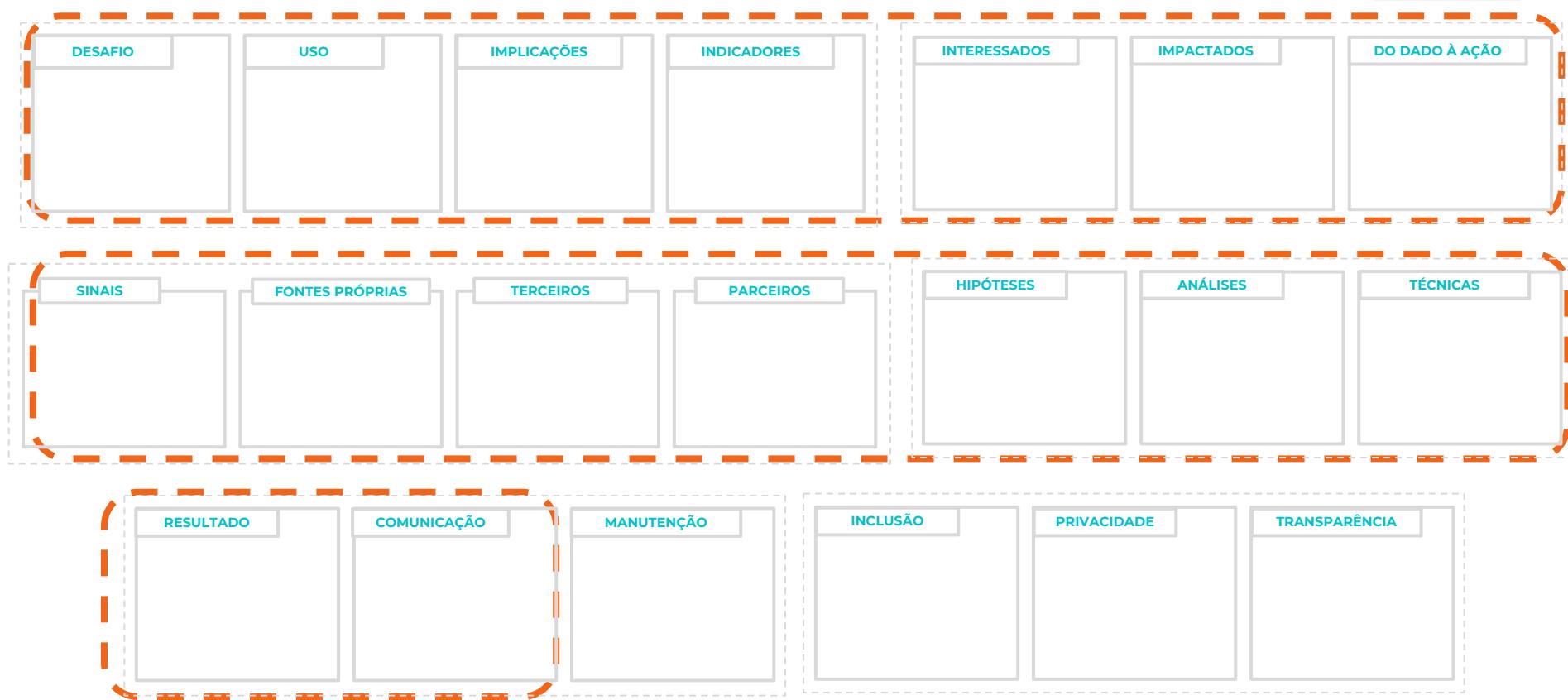
- Contexto
- Conflito Incitante
- Conflitos e Aproximação
- Proposta de Solução
- Desfecho
- Retorno

Pitch

- Contexto
- Problema
- Proposta de Solução
- Oportunidade
- Pedido!
- Problema
- Proposta de Solução
- Produto
- Mercado
- Time
- Financeiro

CANVAS Advanced Analytics Canvas

HEKIMA



CANVAS Pitch





Pitch

Tempo para criação: **5 minutos**

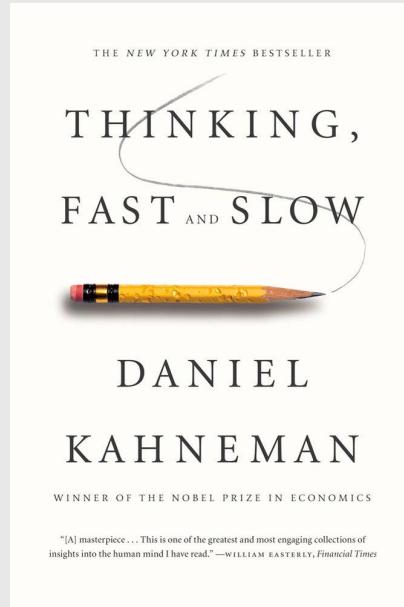
Recomendações

DATA STORYTELLING - Hans Rosling



TED Talks

Livro - Rápido e Devagar



HEKIMA

Nossos escritórios:

SÃO PAULO

Rua Casa do Ator, 919
1o. Andar. Vila Olímpia.

BELO HORIZONTE

Avenida Álvares Cabral,
1315, Sala 7 Lourdes

www.hekima.com

contato@hekima.com
+55 31 2516-4812