

Live #11

# Os principais erros de Metodologia e Data Storytelling

Parte 2

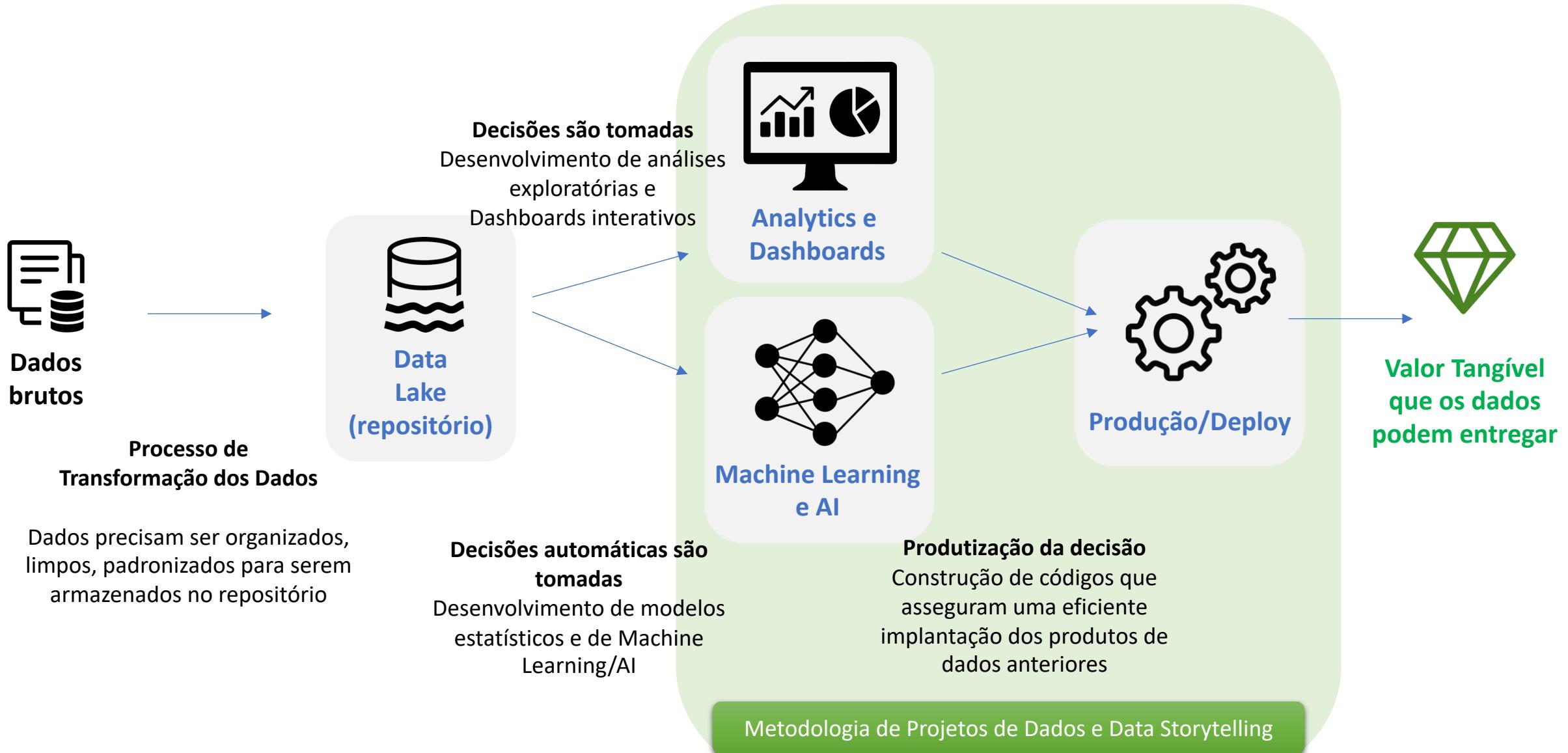
07/03/2023



Preditiva.ai

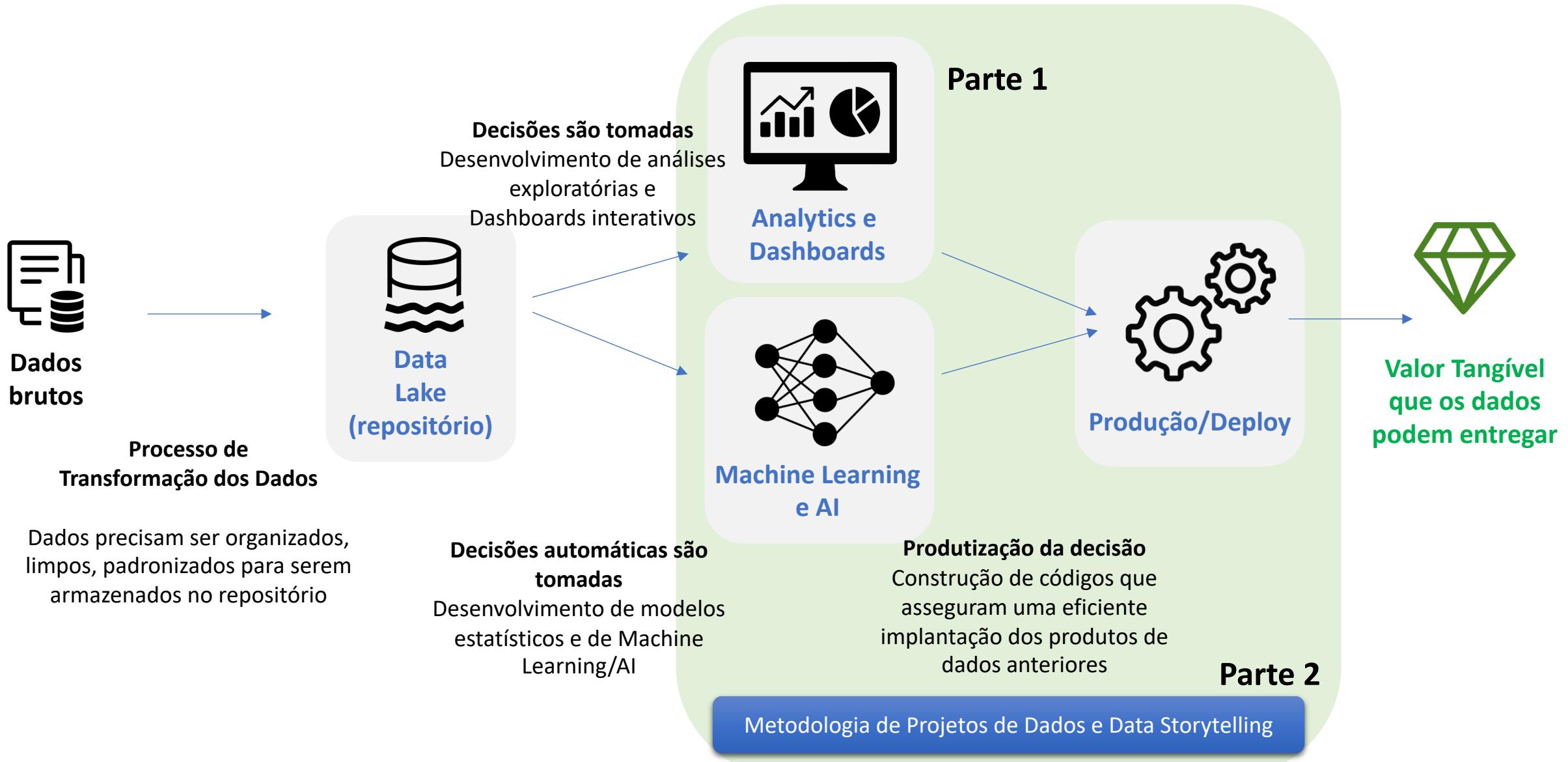
# Jornada dos dados

Existem muitas chances de erros na jornada dos dados



# Jornada dos dados

Existem muitas chances de erros na jornada dos dados



**Mas antes de continuar... uma dúvida?**

**Já deu like e compartilhou  
com seus amigos?**

# Erros de Metodologia

## Overview



**Análise de Dados** é uma máquina de geração de valor. Mas para que ela funcione, precisamos de pelo menos três “engrenagens”. São elas:



- ✓ Matemática fundamental
- ✓ Técnicas de Otimização
- ✓ Estatística Descritiva
- ✓ Probabilidades
- ✓ Inferência (ex: Testes de Hipótese...)
- ✓ Modelagem Estatística
- ✓ *Machine Learning*
- ✓ Entre outros...

- ✓ Planilhas e Tipos de Arquivos
- ✓ Linguagens de programação
- ✓ Bancos de dados
- ✓ Visualização de Dados
- ✓ Infra estrutura (DWs, Data Lakes etc)
- ✓ ETL / Pipelines de Dados
- ✓ Entre outros...

- ✓ Entendimento dos processos internos
- ✓ Alinhamento com os objetivos da empresa
- ✓ Entrosamento com o time e principais *stakeholders*
- ✓ Análise da concorrência
- ✓ Entre outros...

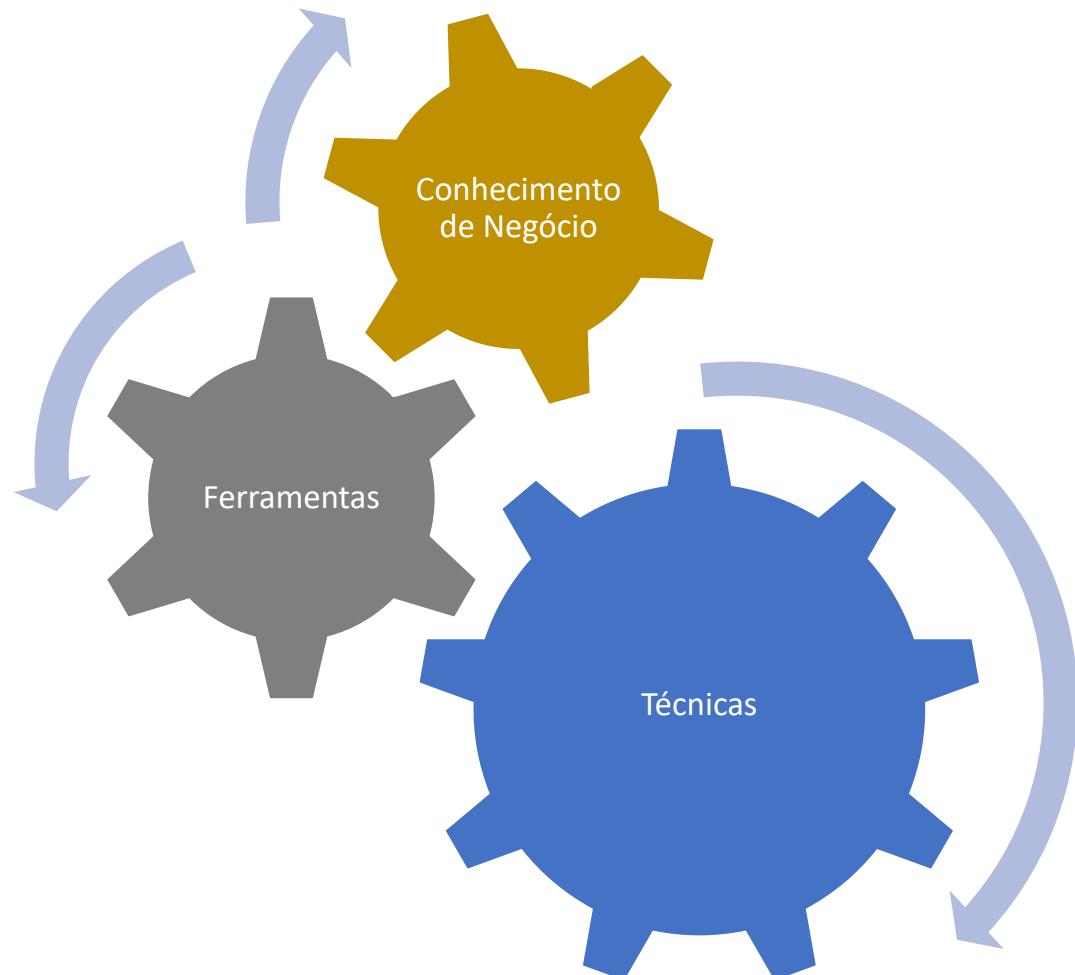
# Erros de Metodologia

## Overview

No entanto, essas “engrenagens” funcionando sozinhas **não** entregam o real valor para as áreas das empresas como deveriam.

Felizmente, existe uma forma eficiente para que as engrenagens funcionem em sinergia:

Aplicando-se uma  
**Metodologia de Projetos de Dados.**



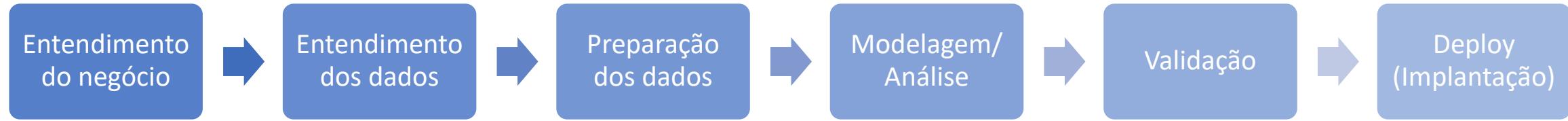
# Erros de Metodologia

## Overview do CRISP-DM



Preditiva.ai

Recapitulando os passos da metodologia. Utilize essas etapas como um “check-list” do seu projeto:



<input type="checkbox"/> Objetivo	<input type="checkbox"/> Descrição dos dados	<input type="checkbox"/> Seleção de variáveis	<input type="checkbox"/> Escolha da técnica estatística que responde o problema	<input type="checkbox"/> Verificação dos critérios de sucesso	<input type="checkbox"/> Plano de implantação
<input type="checkbox"/> Premissas	<input type="checkbox"/> Coleta	<input type="checkbox"/> Limpeza dos dados	<input type="checkbox"/> Desenvolvimento do estudo analítico ou modelo	<input type="checkbox"/> Validação do estudo/ modelo	<input type="checkbox"/> Monitoramento do modelo/KPI's
<input type="checkbox"/> Riscos envolvidos	<input type="checkbox"/> Análise Exploratória	<input type="checkbox"/> Formatação dos dados	<input type="checkbox"/> Finalização	<input type="checkbox"/> Apresentação/ Data Storytelling	<input type="checkbox"/> Finalização da documentação
<input type="checkbox"/> Custos x Benefícios	<input type="checkbox"/> Verificação da qualidade dos dados ( <i>Sanity Check</i> )			<input type="checkbox"/> Atualização do Roadmap	
<input type="checkbox"/> Critérios de sucesso					
<input type="checkbox"/> Planejamento do projeto e início da documentação					

# Erros de Metodologia

## Erro do tipo 3

### Erro do tipo 1

Falso **Positivo**



### Erro do tipo 2

Falso **Negativo**



### Erro do “tipo 3”

**Desenvolver algum projeto de dados apenas  
porque é possível desenvolvê-lo**

- Analistas estão desenvolvendo **Dashboards** apenas porque hoje é mais fácil (Power BI)
- Cientistas de Dados estão desenvolvendo modelos de **Machine Learning** complexos demorando meses para melhorar 1% do resultado
- Gestores pedem estudos e não sabem o que vão fazer com eles (nem os analistas)

### O erro:

Começar um projeto sem um objetivo claro e que gere valor para a empresa

### Como evitá-lo?

Muitos analistas somente entendem o objetivo geral e já saem construindo *queries* e criando *dashboards*. Grande erro! Nesta etapa, sempre tenha as respostas para seguintes perguntas:

- Qual o objetivo deste trabalho?
- O que queremos conhecer? O que queremos mudar na área com esse projeto?
- Já existe algo realizado ou em andamento feito por alguém? Quais os resultados?
- Supondo que consigamos atingir o objetivo, o que vem depois? Como este trabalho será usado?

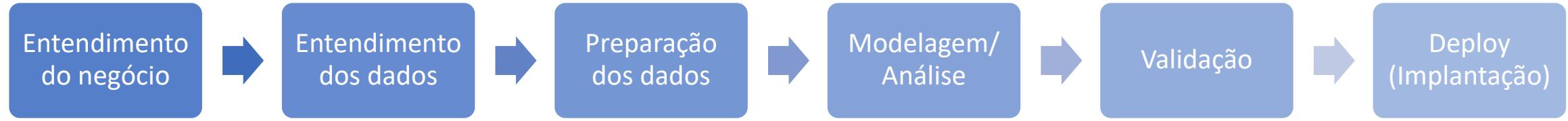
# Erros de Metodologia

## Overview do CRISP-DM



Preditiva.ai

Recapitulando os passos da metodologia. Utilize essas etapas como um “check-list” do seu projeto:



<input type="checkbox"/> Objetivo	<input type="checkbox"/> Descrição dos dados	<input type="checkbox"/> Seleção de variáveis	<input type="checkbox"/> Escolha da técnica estatística que responde o problema	<input type="checkbox"/> Verificação dos critérios de sucesso	<input type="checkbox"/> Plano de implantação
<input type="checkbox"/> Premissas	<input checked="" type="checkbox"/> <b>Coleta</b>	<input type="checkbox"/> Limpeza dos dados	<input type="checkbox"/> Desenvolvimento do estudo analítico ou modelo	<input type="checkbox"/> Validação do estudo/ modelo	<input type="checkbox"/> Monitoramento do modelo/KPI's
<input type="checkbox"/> Riscos envolvidos	<input type="checkbox"/> Análise Exploratória	<input type="checkbox"/> Formatação dos dados	<input type="checkbox"/> Finalização	<input type="checkbox"/> Aprovação pelo cliente do trabalho	<input type="checkbox"/> Finalização da documentação
<input type="checkbox"/> Custos x Benefícios	<input type="checkbox"/> Verificação da qualidade dos dados ( <i>Sanity Check</i> )			<input type="checkbox"/> Atualização do Roadmap	
<input type="checkbox"/> Critérios de sucesso					
<input type="checkbox"/> Planejamento do projeto e início da documentação					

# Erros de Metodologia

## Dados ruins



Invista um tempo suficiente para coletar e preparar bons dados, pois caso contrário seu estudo/modelo pode não servir para nada.

***"Garbage In, Garbage Out"***

# Erros de Metodologia

## Exemplo

Base “bruta” extraída do repositório de dados

Nome (A)	Data de Nascimento (B)	Renda (C)	Estado (D)
Sr Marcelo	20/01/1985	2000 mil	São Paulo
Sra Luana	14/06/1995	3000 mil	sao paulo
Sr João	11/01/1991	1200 mil	RJ
Sra Valquiria	23/07/1981	-	Rio de Janeiro

Base “tratada” após tarefas de tratamento

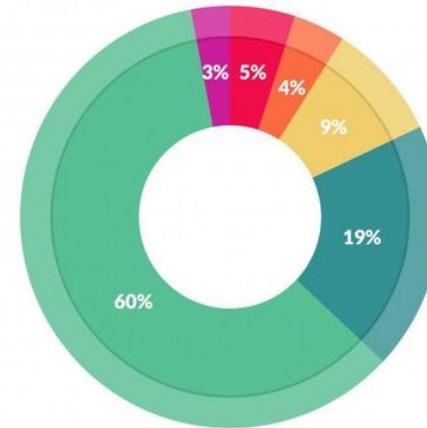
Variável criada a partir de (A)	Variável criada a partir de (E)	Variável criada a partir de (B)	Variável criada a partir de (F)	Variável criada a partir de (C)	Variável criada a partir de (D)
Variável criada a partir de (A)	Variável criada a partir de (E)	Variável criada a partir de (B)	Variável criada a partir de (F)	Variável criada a partir de (C)	Variável criada a partir de (D)
Pronome (E)	Sexo	Data de Nascimento	Idade (F)	Faixa Etária	Renda
Sr	Masculino	20/01/1985	36	33-38	R\$ 2.000
Sra	Feminino	14/06/1995	26	23-28	R\$ 3.000
Sr	Masculino	11/01/1991	30	28-33	R\$ 1.200
Sra	Feminino	23/07/1981	40	38-43	R\$ 2.067

Feature Engineering

Feature Engineering + Categorização

Ajuste de tipo de dados + imputação de missing

Padronização + Ajuste de Case Sensitive



What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets: 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

Fonte: <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says>

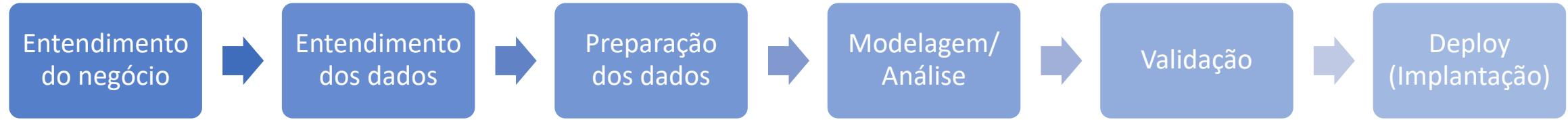
# Erros de Metodologia

## Overview do CRISP-DM



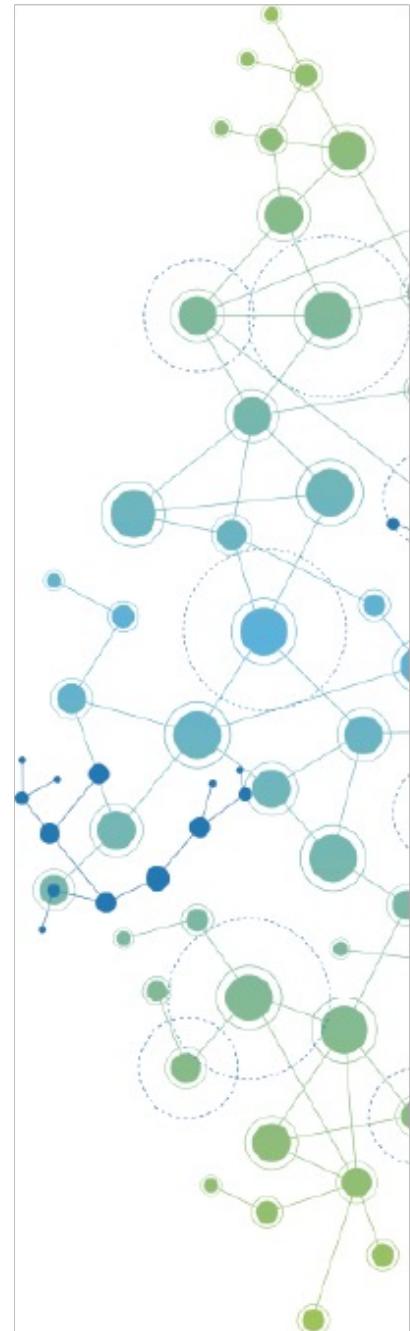
Preditiva.ai

Recapitulando os passos da metodologia. Utilize essas etapas como um “check-list” do seu projeto:



<input type="checkbox"/> Objetivo	<input type="checkbox"/> Descrição dos dados	<input type="checkbox"/> Seleção de variáveis	<input type="checkbox"/> Escolha da técnica estatística que responde o problema	<input type="checkbox"/> Verificação dos critérios de sucesso	<input type="checkbox"/> Plano de implantação
<input type="checkbox"/> Premissas	<input type="checkbox"/> Coleta	<input type="checkbox"/> Limpeza dos dados	<input type="checkbox"/> Desenvolvimento do estudo analítico ou modelo	<input type="checkbox"/> Validação do estudo/ modelo	<input type="checkbox"/> Monitoramento do modelo/KPI's
<input type="checkbox"/> Riscos envolvidos	<input type="checkbox"/> Análise Exploratória	<input type="checkbox"/> Formatação dos dados	<input type="checkbox"/> Finalização	<b><input checked="" type="checkbox"/> Apresentação /Data Storytelling</b>	<input type="checkbox"/> Finalização da documentação
<input type="checkbox"/> Custos x Benefícios	<input type="checkbox"/> Verificação da qualidade dos dados ( <i>Sanity Check</i> )			<input type="checkbox"/> Atualização do Roadmap	
<input type="checkbox"/> Critérios de sucesso					
<input type="checkbox"/> Planejamento do projeto e início da documentação					

# Elementos visuais ruins



# Data Storytelling

## Gráficos



Qual o problema com essa matriz de correlação de preços de ação?

Lembrando que queremos correlações pequenas para diversificar o risco da carteira

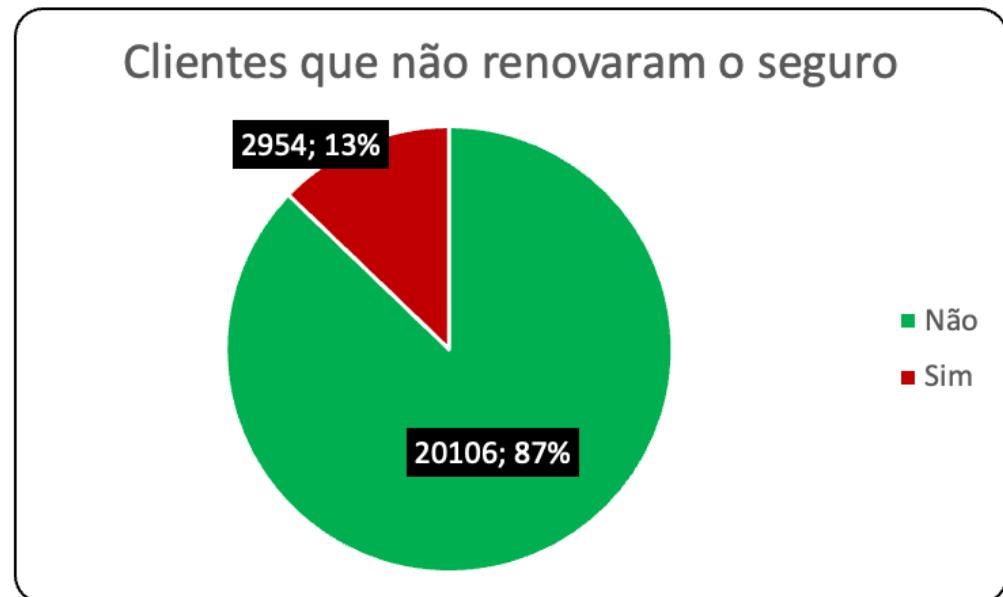
MATRIZ HISTÓRICA DE CORRELAÇÃO										
	<i>Arezzo&amp;Co</i>	<i>B3</i>	<i>Klabin</i>	<i>Localiza</i>	<i>Lojas Americanas</i>	<i>Multiplan</i>	<i>Petrobras</i>	<i>Rede Atacadão</i>	<i>Sul América</i>	<i>Vale</i>
<i>Arezzo&amp;Co</i>										
<i>B3</i>		0,52								
<i>Klabin</i>	0,27		0,28							
<i>Localiza</i>	0,61	0,61		0,24						
<i>Lojas Americanas</i>	0,42	0,40	0,27		0,44					
<i>Multiplan</i>	0,57	0,59	0,30	0,66		0,43				
<i>Petrobras</i>	0,46	0,59	0,35	0,56	0,35		0,57			
<i>Rede Atacadão</i>	0,27	0,44	0,30	0,39	0,33		0,39	0,36		
<i>SulAmerica</i>	0,59	0,59	0,33	0,63	0,46		0,63	0,57	0,37	
<i>Vale</i>	0,37	0,50	0,37	0,42	0,30		0,40	0,63	0,30	0,44

# Data Storytelling

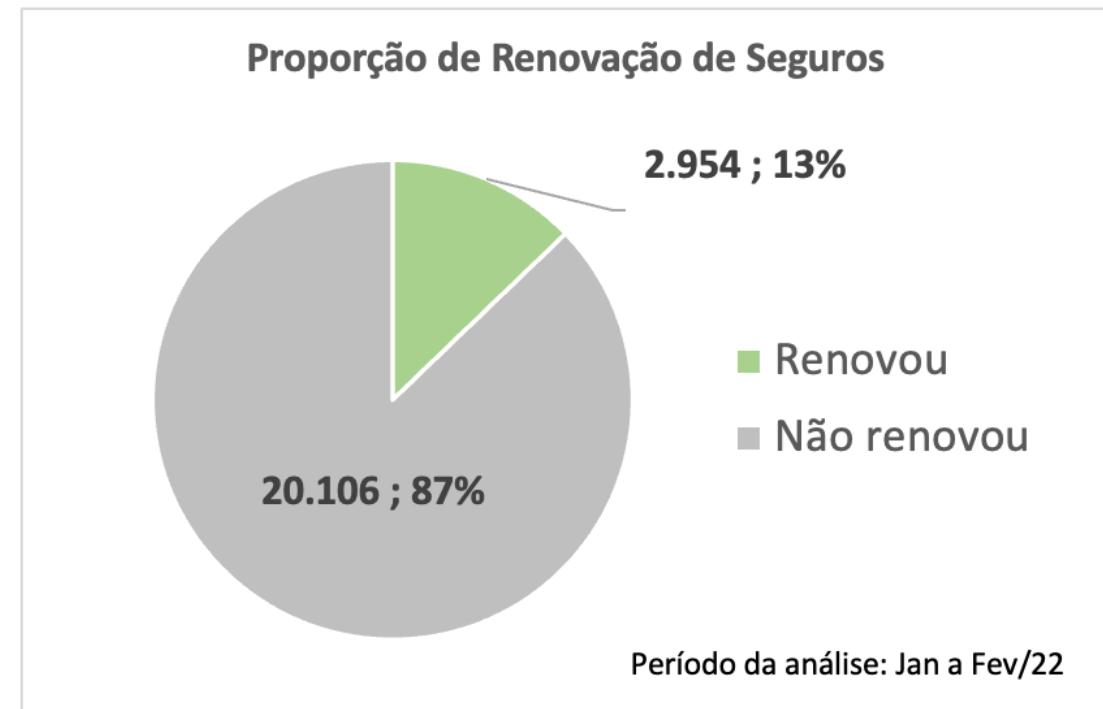
## Gráficos



**Objetivo do trabalho:** Encontrar formas dos clientes renovarem a garantia extendida.



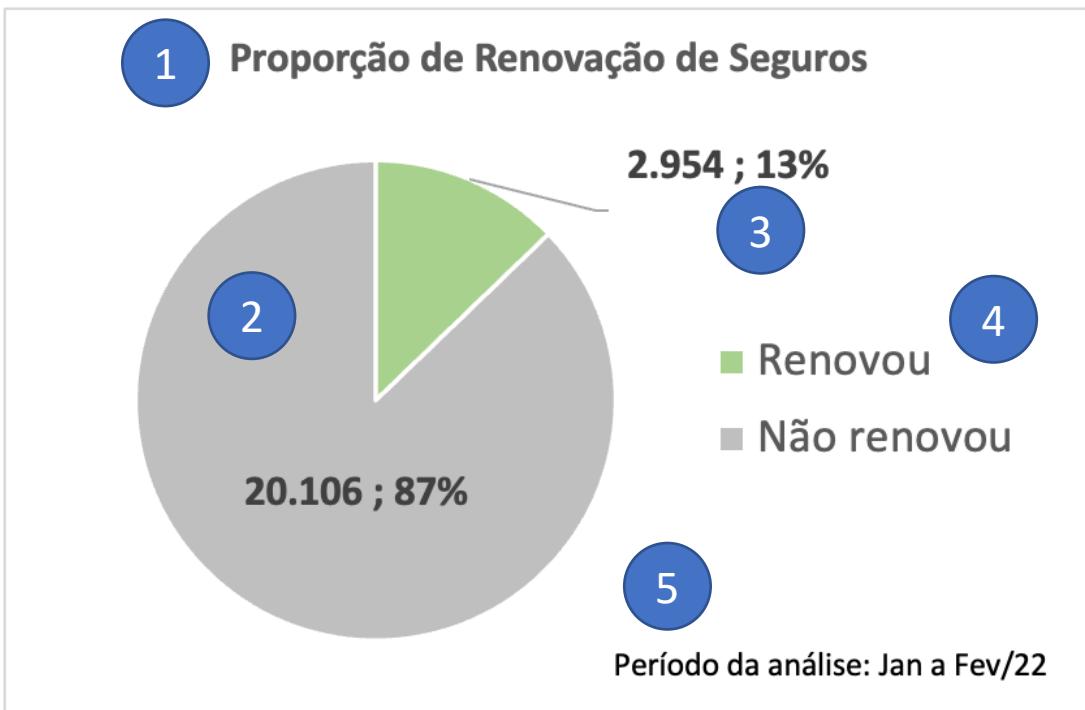
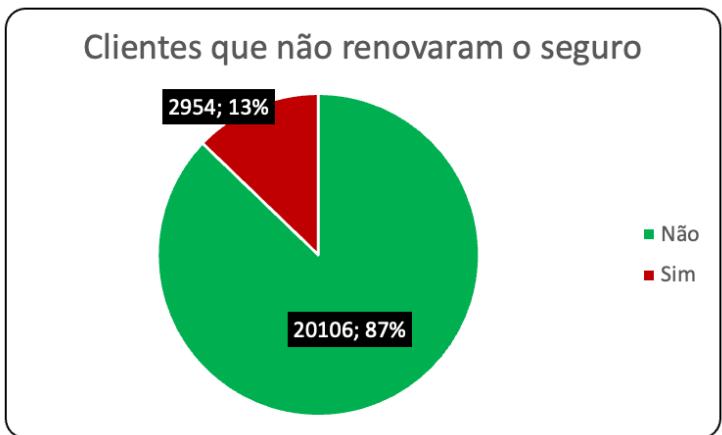
**Qual o problema com  
esse gráfico?**



**Opção melhor**

# Data Storytelling

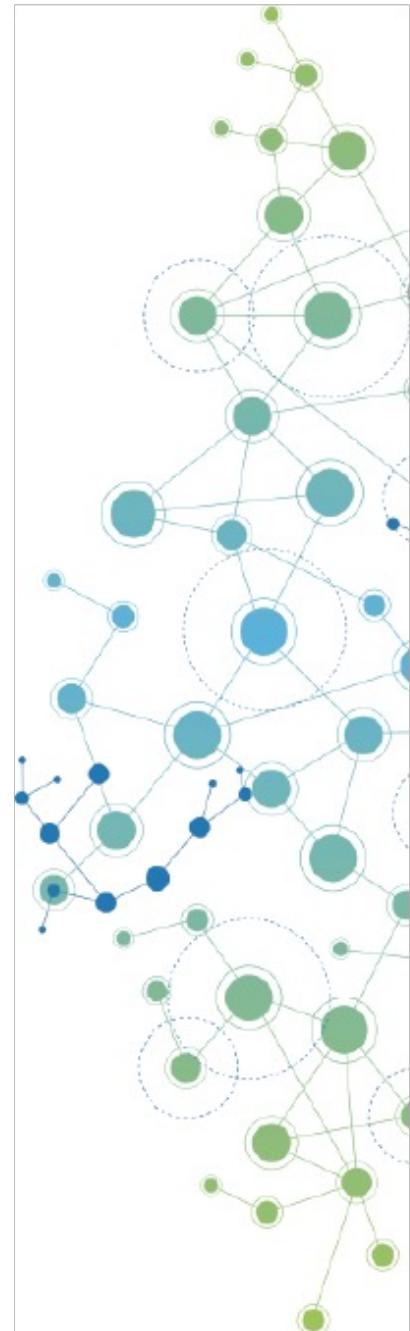
## Gráficos

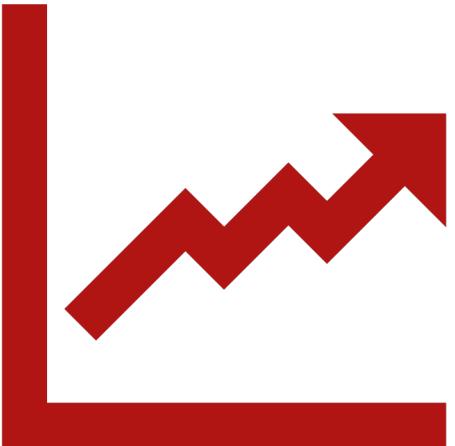


## Por que é melhor?

- 1 Títulos mais intuitivos
- 2 Cores com a psicologia certa e com tons com melhor saturação (não cansa os olhos)
- 3 Formatação numérica adequada e tamanho de fonte visível
- 4 Legendas auto explicativas
- 5 Informações adicionais relevantes

# Falta de Storytelling





# Reduzindo o TurnOver

ESTRATÉGIAS PARA EVITAR O ESTÍMULO A  
TROCAR DE EMPRESA COM BASE NOS DADOS  
DOS COLABORADORES

# Database

- ▶ Foi avaliado o histórico de **1470** funcionários;
- ▶ Entre as informações analisadas, destacam-se as seguintes:
  1. **Quantidade de Horas Extras realizadas**
  2. **Salário médio – considerando a formação**

# Horas Extras

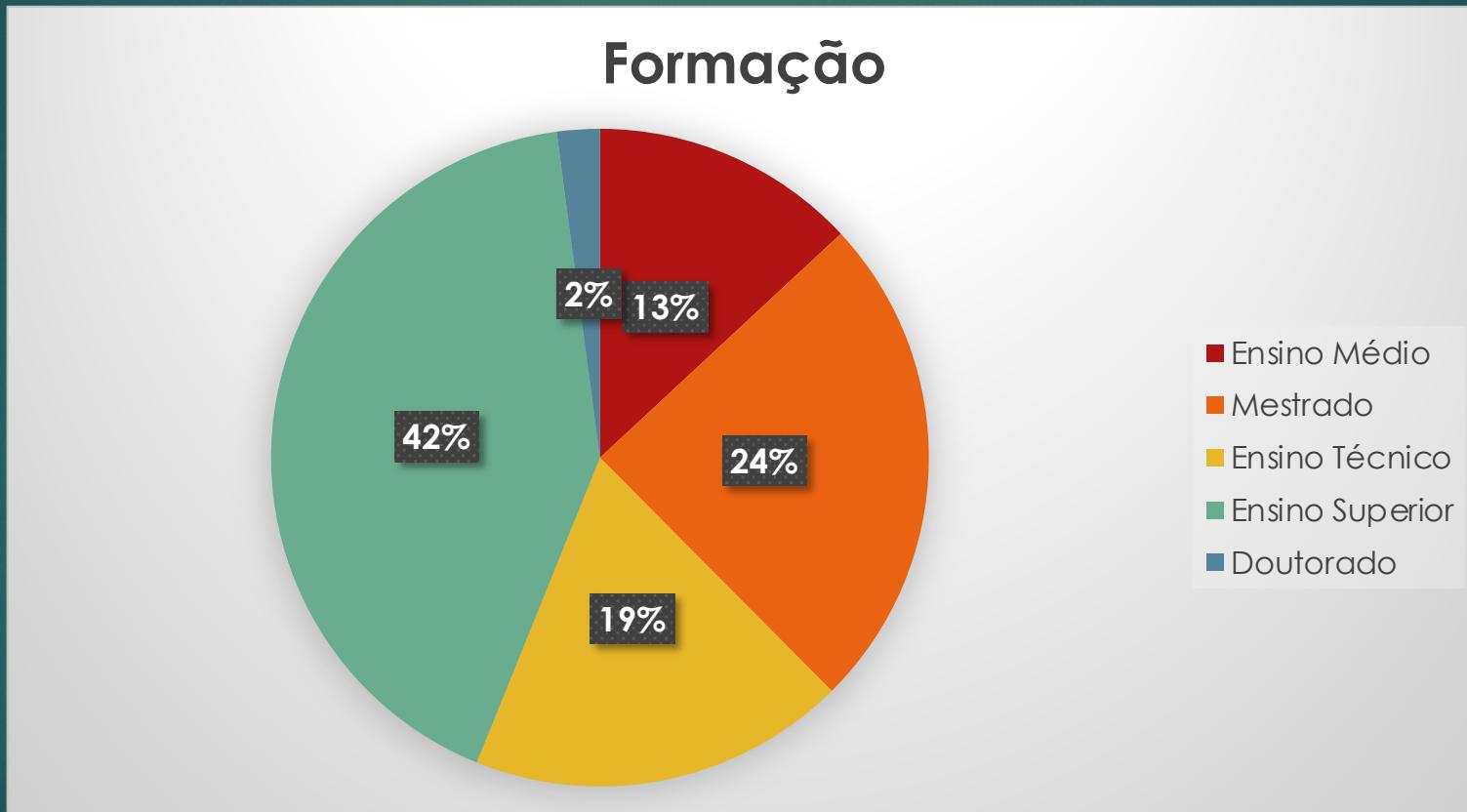
- ▶ Dos 237 funcionários que deixaram a empresa no período;
  - ❖ 127 realizaram horas extras, correspondendo a 54% do total
  - ❖ 110 não realizaram horas extras, correspondendo a 46% do total
- ▶ No mesmo período, dos 1233 funcionários que permanecem;
  - ❖ 289 realizaram horas extras, correspondendo a 23% do total
  - ❖ 944 não realizaram horas extras, correspondendo a 77% do total

# Horas Extras

- ▶ Considerando apenas estas informações, sugere-se que a realização de horas extras é um fator que possui correlação com o maior TurnOver de funcionários
- ▶ Cabe ao setor de RH validar com os gestores a política de utilização de horas extras, considerando este cenário

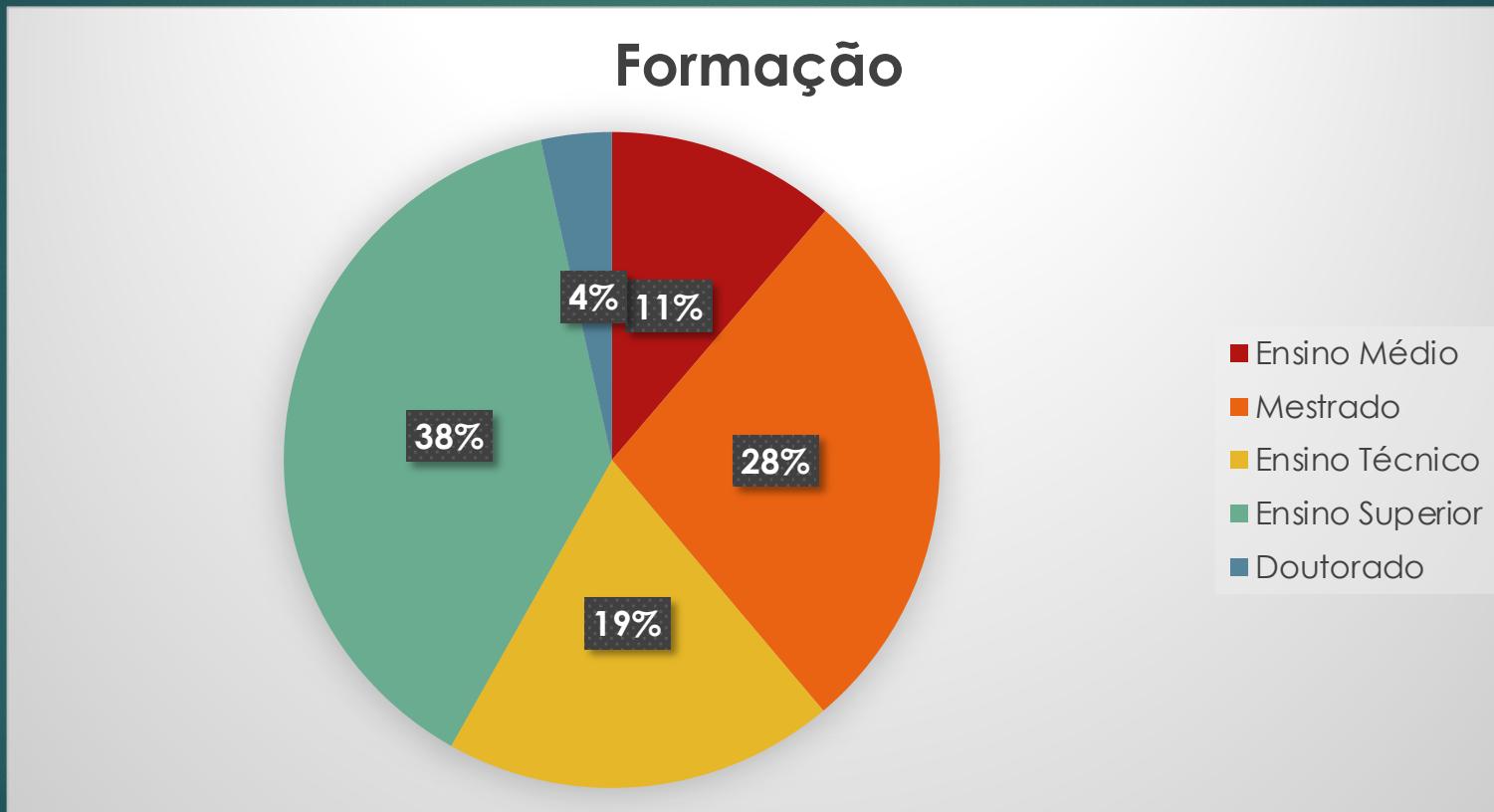
# Salário Médio vs. Formação

- ▶ Entre os funcionários que **sairam** da empresa:



# Salário Médio vs. Formação

- ▶ Entre os funcionários que **permaneceram** da empresa:



# Salário Médio vs. Formação

- ▶ Percebe-se que mesmo com percentuais similares em relação à formação, a média salarial entre quem **deixou** a empresa e quem **permaneceu** possui grande discrepância:

CARGO	SALÁRIO MÉDIO		% DIFERENÇA
	SAIU	FICOU	
Ensino Médio	4.360,16	5.926,13	<b>-26%</b>
Mestrado	5.335,16	7.087,81	<b>-25%</b>
Ensino Técnico	4.282,55	6.586,06	<b>-35%</b>
Ensino Superior	4.770,24	6.882,92	<b>-31%</b>
Doutorado	5.850,20	8.559,91	<b>-32%</b>

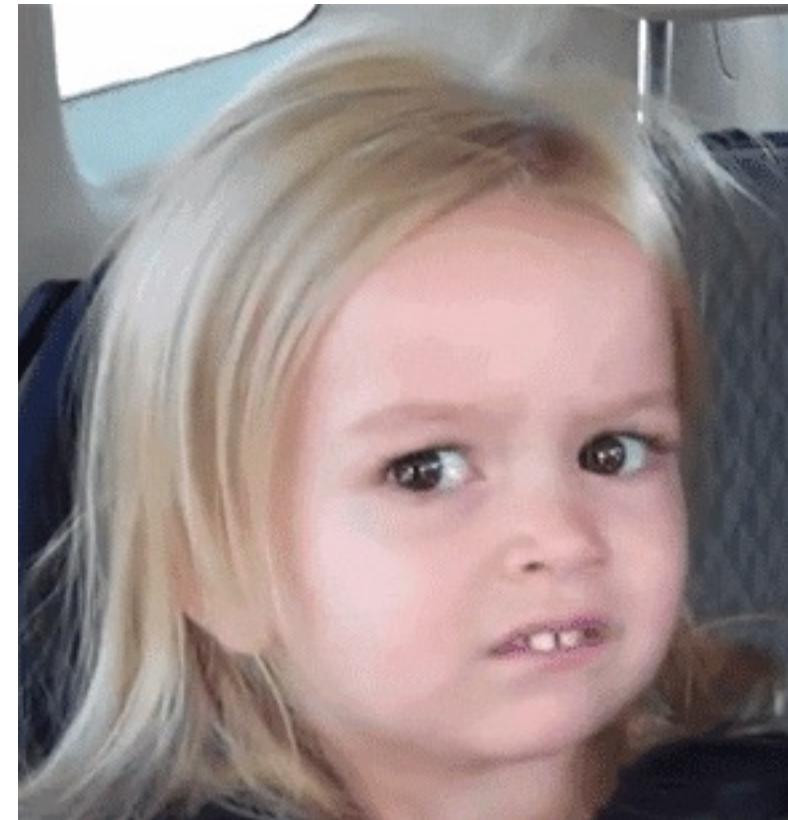
# Salário Médio vs. Formação

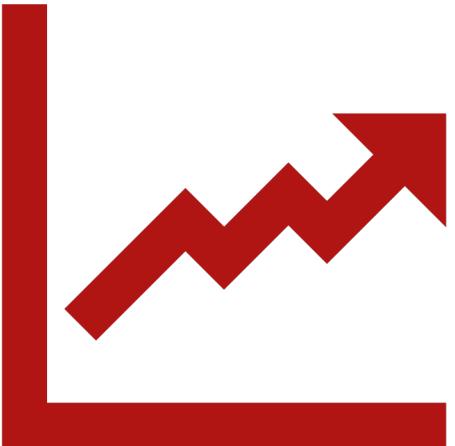
- ▶ Conclui-se que discrepância salarial entre funcionários com formações similares é um motivo que pode ser correlacionado com o alto TurnOver, especialmente quando a diferença é grande
- ▶ Cabe ao setor de RH, conjuntamente com os gestores, determinar de maneira mais clara a política de compensação salarial, atrelada às especificidades de cada formação para evitar perder funcionários para o mercado



► Obrigado!

# Ficou claro?





# Reduzindo o TurnOver

ESTRATÉGIAS PARA EVITAR O ESTÍMULO A  
TROCAR DE EMPRESA COM BASE NOS DADOS  
DOS COLABORADORES

# Database

- ▶ Foi avaliado o histórico de **1470** funcionários;
- ▶ Entre as informações analisadas, destacam-se as seguintes:
  1. **Quantidade de Horas Extras realizadas**
  2. **Salário médio – considerando a formação**

Por que foi analisado?

Qual o período analisado?

“destacam-se” em que sentido?

# Horas Extras

- ▶ Dos 237 funcionários que deixaram a empresa no período;
  - ❖ 127 realizaram horas extras, correspondendo a 54% do total
  - ❖ 110 não realizaram horas extras, correspondendo a 46% do total
  
- ▶ No mesmo período, dos 1233 funcionários que permanecem;
  - ❖ 289 realizaram horas extras, correspondendo a 23% do total
  - ❖ 944 não realizaram horas extras, correspondendo a 77% do total

Técnica não adequada  
(Não se deve comparar os percentuais pois o risco de interpretação errada é grande)

A informação mostrada carece de visuais e organização. Uma tabela simples bem formada seria muito mais fácil de ser interpretado.

# Horas Extras

- ▶ Considerando apenas estas informações, sugere-se que a realização de horas extras é um fator que possui correlação com o maior *TurnOver* de funcionários
- ▶ Cabe ao setor de RH validar com os gestores a política de utilização de horas extras, considerando este cenário

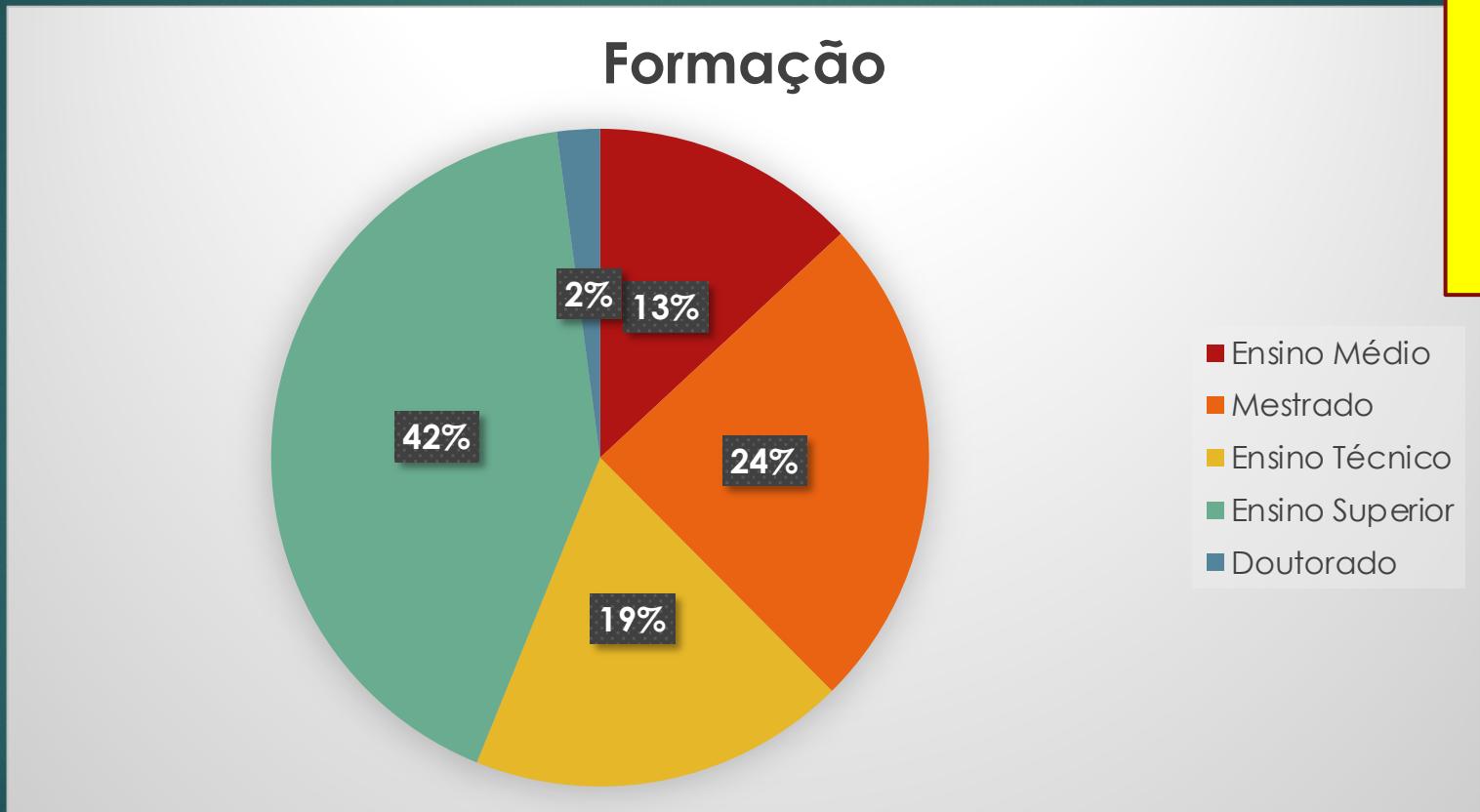
“Considerando apenas estas informações” ?

Como assim, deveria considerar outras?

Sugestão fraca pois a explicação não convence.

# Salário Médio vs. Formação

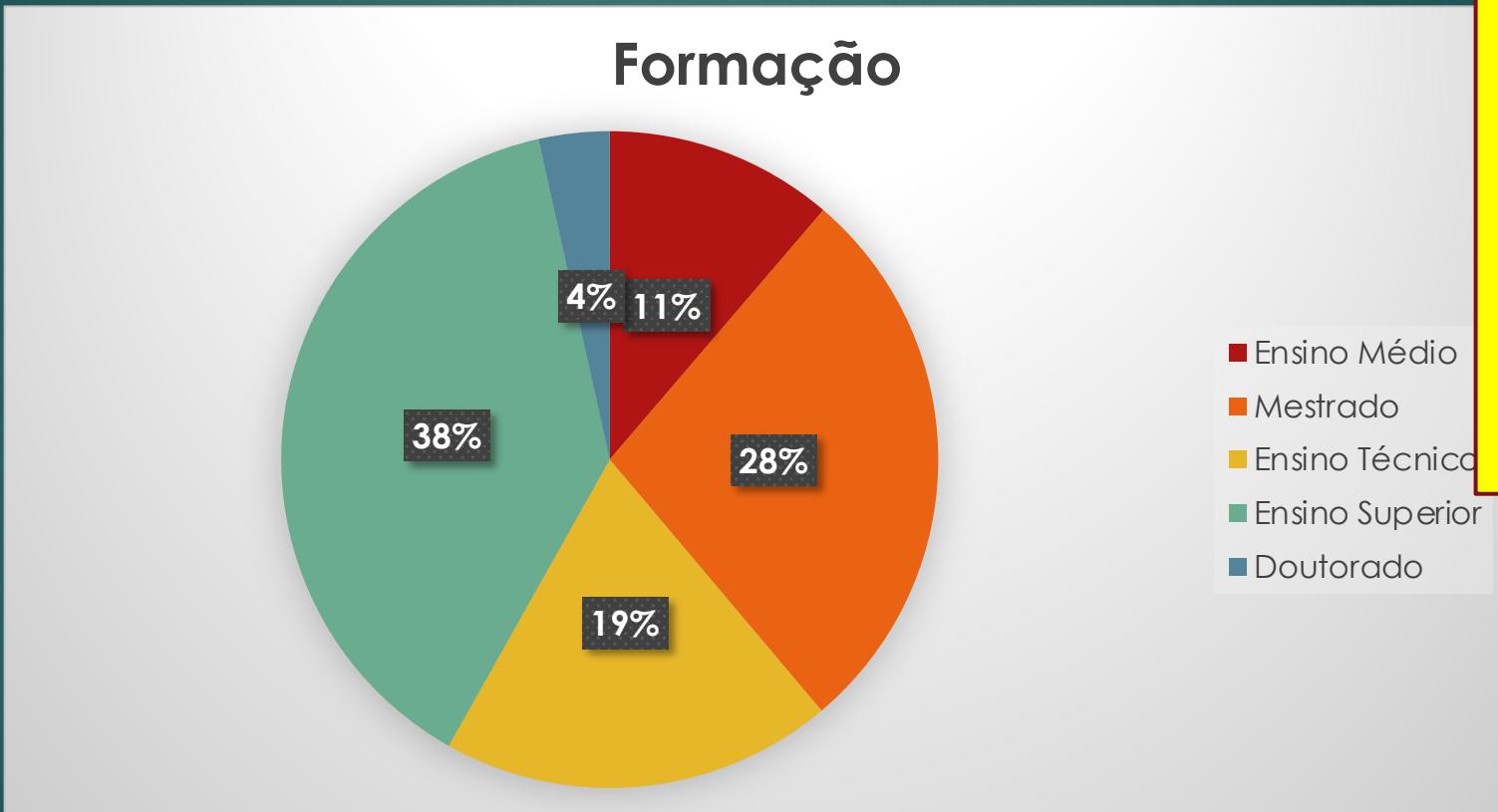
- ▶ Entre os funcionários que **sairam** da empresa:



Salário Médio? Cadê?

# Salário Médio vs. Formação

- Entre os funcionários que **permaneceram** da empresa:



Como as pizzas estão em slides diferentes, as pessoas não conseguem comparar as grandezas.

Além disso, o visual escolhido não é adequado, pois aumentase o esforço mental para interpretação.

# Salário Médio vs. Formação

- Percebe-se que mesmo com percentuais similares em relação à formação, a média salarial entre quem **deixou** a empresa e quem **permaneceu** possui grande discrepância:

CARGO	SALÁRIO MÉDIO		% DIFERENÇA
	SAIU	FICOU	
Ensino Médio	4.360,16	5.926,13	-26%
Mestrado	5.335,16	7.087,81	-25%
Ensino Técnico	4.282,55	6.586,06	-35%
Ensino Superior	4.770,24	6.882,92	-31%
Doutorado	5.850,20	8.559,91	-32%

Risco de conclusão errada. Não se pode comparar médias como medida absoluta da diferença entre grupos.

A média é medida sensível a outliers e na maioria das vezes não resume a realidade do comportamento analisado.

# Salário Médio vs. Formação

- ▶ Conclui-se que discrepância salarial entre funcionários com formações similares é um motivo que pode ser correlacionado com o alto TurnOver, especialmente quando a diferença é grande
- ▶ Cabe ao setor de RH, conjuntamente com os gestores, determinar de maneira mais clara a política de compensação salarial, atrelada às especificidades de cada formação para evitar perder funcionários para o mercado

E as outras variáveis?  
Tanta informação que  
não foi analisada?

Plano de ação por  
prioridade?



Preditiva.ai

# Projeto de Minimização do TurnOver - RH

Versão 1

Mar/23

# Projeto: Minimização de Turnover

## Contexto de negócio e objetivo do trabalho

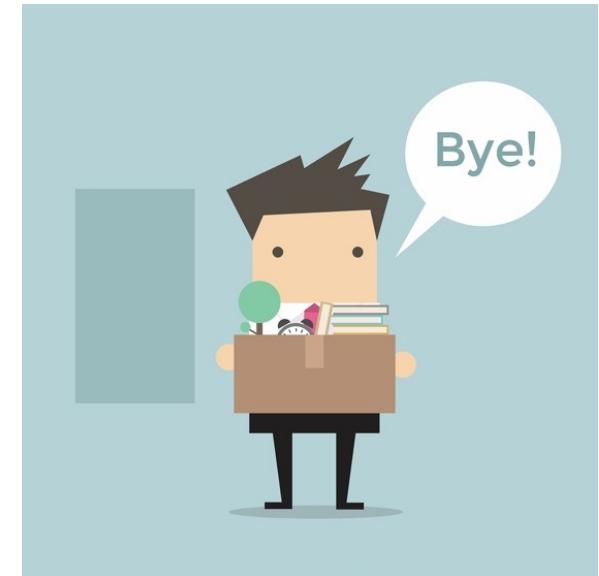


O **Turnover** (rotatividade de funcionários) é um grande problema para as empresas. Sempre que um funcionário deixa um determinado trabalho, a empresa perde dinheiro e tempo com novas entrevistas e treinamentos do novo funcionário. Isso sem falar da perda de produtividade do setor afetado por esse *turnover*. São muitas as questões que fazem um funcionário deixar a empresa, entre eles: Melhores oportunidades, clima organizacional ruim, chefes ruins, baixo equilíbrio entre vida pessoal e profissional, entre outros.

Para tentar entender quais as características que fazem um funcionário ficar ou deixar uma empresa de Tecnologia, o time de **RH nos forneceu** uma base com **1470 funcionários** que deixaram ou permaneceram na companhia no último ano.

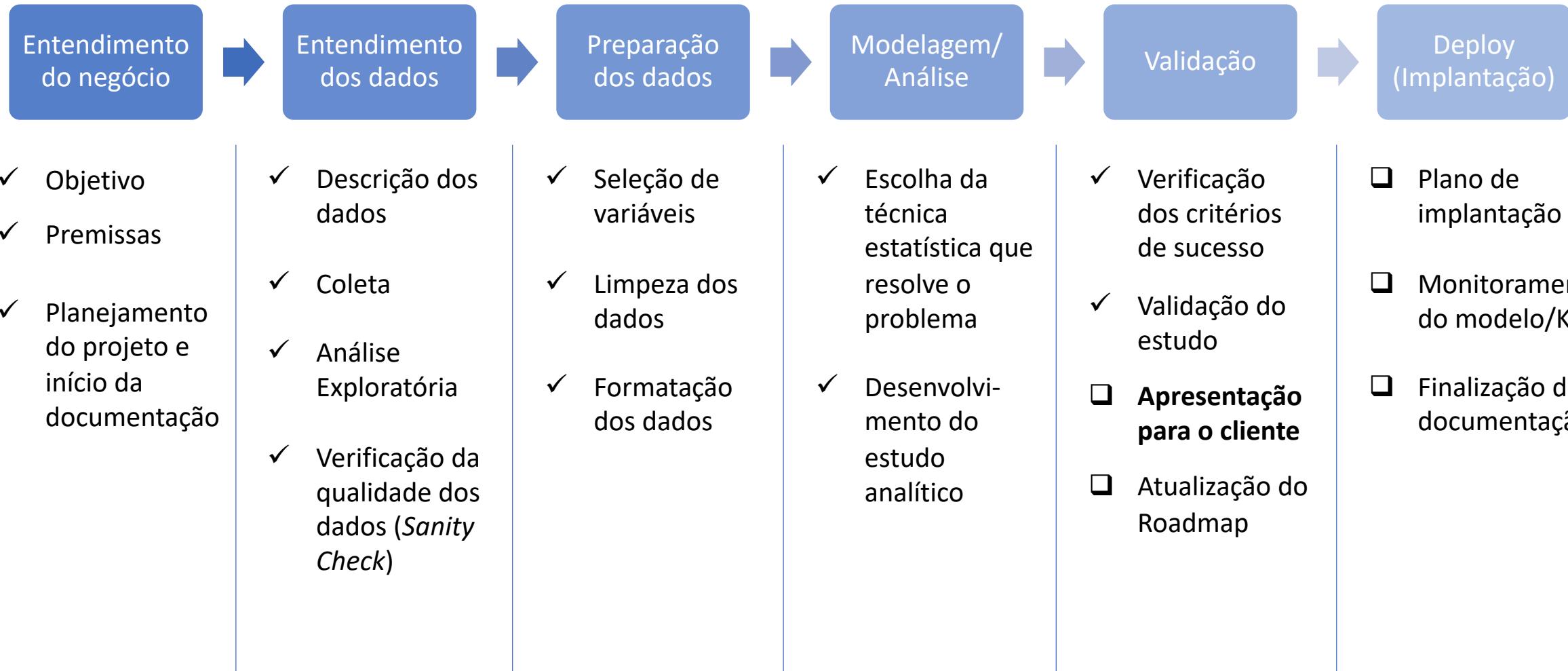
O resultado desse levantamento gerou **19 possíveis fatores** que explicam o comportamento do turnover.

Portanto, o objetivo deste trabalho é: **Buscar quais políticas/fatores da empresa deveriam mudar de forma a minimizar o turnover? Ou seja, quais planos de ação podemos sugerir para o RH da empresa?**



# Projeto: Minimização de Turnover

## Etapas do projeto



# Projeto: Minimização de Turnover

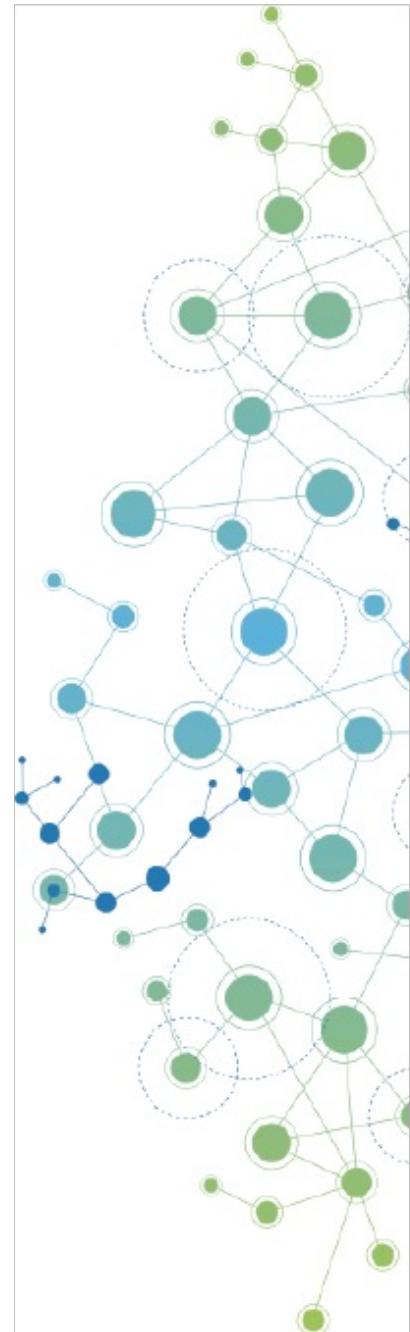
## Premissas e limitações



Seguem abaixo informações importantes sobre o trabalho:

- O **estudo considera apenas um ano de comportamentos** (2021 a 2022). Em versões futuras deste trabalho, podemos pegar mais tempo e verificar se as conclusões mudam.
- A **base foi preparada pelo time de TI**. Estamos supondo que todas as queries de extração foram executadas corretamente e não temos maiores problemas oriundos de risco operacional.
- O turnover estudado aqui refere-se ao chamado **Turnover Voluntário**, ou seja, rotatividade realizada quando o colaborador pede demissão. Desta forma, a empresa tem menos custos de recisão.
- O **Turnover mensurado no último ano foi de 16%**. Como o RH não nos passou uma expectativa de redução de Turnover, vamos considerar como “**critério de sucesso**” para este trabalho uma redução de 6 pp. Portanto, nossa meta para Turnover será de no **máximo 10%**.

# 1) Análise exploratória



# Projeto: Minimização de Turnover

## Descrição das variáveis



Variável	Descrição
ID	Matrícula do funcionário
<b>Funcionário_deixou_a_empresa</b>	<b>Marcação sem funcionário deixou a empresa recentemente (Target)</b>
Idade	Idade do funcionário
Frequência de Viagens	Frequência de viagens a trabalho do funcionário
Distância_do_trabalho	Distância em Km até o trabalho
Formação	Nível de formação
E-Sat	Satisfação com o clima organizacional
Gênero	Gênero do funcionário
Estado_Civil	Estado civil do funcionário
Salário	Salário mensal
Qte_Empresas_Trabalhadas	Quantidade de empresas que o funcionário já trabalhou
Faz_hora_extras?	Se funcionário costuma fazer hora extra
Perc_de_aumento	Percentual de aumento de salário de 2018 a 2019
Qte_ações_da_empresa	Qte de lotes de ações da empresa que o funcionário possui
Tempo_de_carreira	Tempo em anos que o funcionário tem de carreira
Horas_de_treinamento	Qte de horas de treinamento que o funcionário teve no ano passado
Equilibrio_de_Vida	Nota que o funcionário deu para seu equilíbrio entre vida pessoal e profissional
Tempo_de_empresa	Tempo em anos que o funcionário trabalha na empresa
Anos_no_mesmo_cargo	Qte de tempo em anos que o funcionário atua no mesmo cargo
Anos_desde_a_ultima_promocao	Qte de tempo em anos que o funcionário teve a última promoção
Anos_com_o_mesmo_chefe	Qte de tempo em anos que o funcionário responde para o mesmo chefe

# Projeto: Minimização de Turnover

## Análise exploratória da base



Para verificar quais fatores estão mais relacionados com o Turnover, vamos utilizar a técnica do **Information Value (IV)**. Essa medida é responsável por mensurar o “**poder de separação/discriminação**” que uma variável possui sobre nossa variável target (no caso, o turnover).

Se uma variável tem poder forte, isso significa que uma ou mais categorias da variável tem um alto ou baixo nível de turnover, sendo útil estuda-la com mais profundidade.

Nas páginas a seguir, vamos trazer tabelas com os principais *insights* das variáveis com bom poder de discriminação.

Variável	IV	Grau de Discriminação do Turnover
Salário	0,42	Forte
Faz_hora_extras?	0,40	Forte
Tempo_de_empresa	0,34	Forte
Tempo_de_carreira	0,34	Forte
Idade	0,31	Forte
Anos_no_mesmo_cargo	0,27	Médio
Anos_com_o_mesmo_chefe	0,26	Médio
Estado_Civil	0,22	Médio
Frequência de Viagens	0,13	Médio
Qte_Empresas_Trabalhadas	0,10	Médio
E-Sat	0,10	Médio
Distância_do_trabalho	0,08	Fraco
Equilibrio_de_Vida	0,06	Fraco
Qte_ações_da_empresa	0,06	Fraco
Horas_de_treinamento	0,05	Fraco
Perc_de_aumento	0,05	Fraco
Anos_desde_a_ultima_promocao	0,02	Muito Fraco
Formação	0,01	Muito Fraco
Gênero	0,00	Muito Fraco

# Projeto: Minimização de Turnover

## Análise exploratória da base



Salário	Turnover			Probabilidade de Turnover		IV
	Sim	Não	Total Geral	% Sim	% Não	
1009-3008	113	282	395	48%	23%	0,18
3009-5008	50	307	357	21%	25%	0,01
5009-7008	27	257	284	11%	21%	0,06
7009-9008	13	88	101	5%	7%	0,00
9009-11008	22	96	118	9%	8%	0,00
11009-13008	3	27	30	1%	2%	0,01
13009-15008	4	48	52	2%	4%	0,02
15009-17008		37	37	0%	3%	0%
17009-19008	1	45	46	0%	4%	0,07
19009-21008	4	46	50	2%	4%	0,02
<b>Total Geral</b>	<b>237</b>	<b>1233</b>	<b>1470</b>		<b>16%</b>	<b>0,36</b>

Colaboradores que ganham até 3000 reais têm probabilidade de 29% de turnover, 13 pp acima da média geral.

Faz Hora Extras	Turnover			Probabilidade de Turnover		IV
	Sim	Não	Total Geral	% Sim	% Não	
Não	110	944	1054	46%	77%	0,15
Sim	127	289	416	54%	23%	0,25
<b>Total Geral</b>	<b>237</b>	<b>1233</b>	<b>1470</b>		<b>16%</b>	<b>0,40</b>

Colaboradores que fazem hora extra têm probabilidade de 31% de turnover, 15 pp acima da média geral. Hipótese: Fazer hora extra constantemente pode degradar a qualidade de vida.

Tempo de empresa	Turnover			Probabilidade de Turnover		IV
	Sim	Não	Total Geral	% Sim	% Não	
0-2	102	240	342	43%	19%	0,19
3-5	60	374	434	25%	30%	0,01
6-8	29	217	246	12%	18%	0,02
9-11	28	206	234	12%	17%	0,02
12-14	4	52	56	2%	4%	0,02
15-18	4	50	54	2%	4%	0,02
>18	10	94	104	4%	8%	0,02
<b>Total Geral</b>	<b>237</b>	<b>1233</b>	<b>1470</b>		<b>16%</b>	<b>0,30</b>

Colaboradores com até 2 anos de empresa têm probabilidade de 30% de turnover, 14 pp acima da média geral. Hipóteses: Talvez a cultura da empresa não tem agradado ou mesmo a integração com o time e gestor.

Idade	Turnover			Probabilidade de Turnover		IV
	Sim	Não	Total Geral	% Sim	% Não	
18-22		27	30	57	11%	0,14
23-27		32	121	153	14%	0,01
28-32		70	236	306	30%	0,05
33-37		43	289	332	18%	0,01
38-42		21	222	243	9%	0,06
43-47		17	146	163	7%	0,02
48-52		9	64	73	4%	0,00
>53		18	125	143	8%	0,01
<b>Total Geral</b>	<b>237</b>	<b>1233</b>	<b>1470</b>		<b>16%</b>	<b>0,31</b>

Colaboradores com idade até 22 anos têm probabilidade de 47% de turnover, 31 pp acima da média geral. Embora a taxa seja alta, a quantidade e colaboradores nesta faixa é pequena, não trazendo tanto impacto quanto outras variáveis.

# Projeto: Minimização de Turnover

## Análise exploratória da base



Viagens	Turnover					Probabilidade de Turnover	IV
	Sim	Não	Total Geral	% Sim	% Não		
Não viaja	12	138	150	5%	11%	8%	0,05
Viaja raramente	156	887	1043	66%	72%	15%	0,01
Viaja frequentemente	69	208	277	29%	17%	25%	0,07
<b>Total Geral</b>	<b>237</b>	<b>1233</b>	<b>1470</b>			<b>16%</b>	<b>0,12</b>

Colaboradores que viajam frequentemente têm probabilidade de 25% de turnover, 9 pp acima da média geral. Hipóteses: Viagens constantes podem diminuir a qualidade de vida.

E-Sat	Turnover					Probabilidade de Turnover	IV
	Sim	Não	Total Geral	% Sim	% Não		
Baixo	72	212	284	30%	17%	25%	0,08
Médio	43	244	287	18%	20%	15%	0,00
Alto	62	391	453	26%	32%	14%	0,01
Muito Alto	60	386	446	25%	31%	13%	0,01
<b>Total Geral</b>	<b>237</b>	<b>1233</b>	<b>1470</b>			<b>16%</b>	<b>0,10</b>

Colaboradores com baixa satisfação com o trabalho têm probabilidade de 25% de turnover, 9 pp acima da média geral. Hipóteses: Variável alto explicativa. É preciso entender as causas da insatisfação.

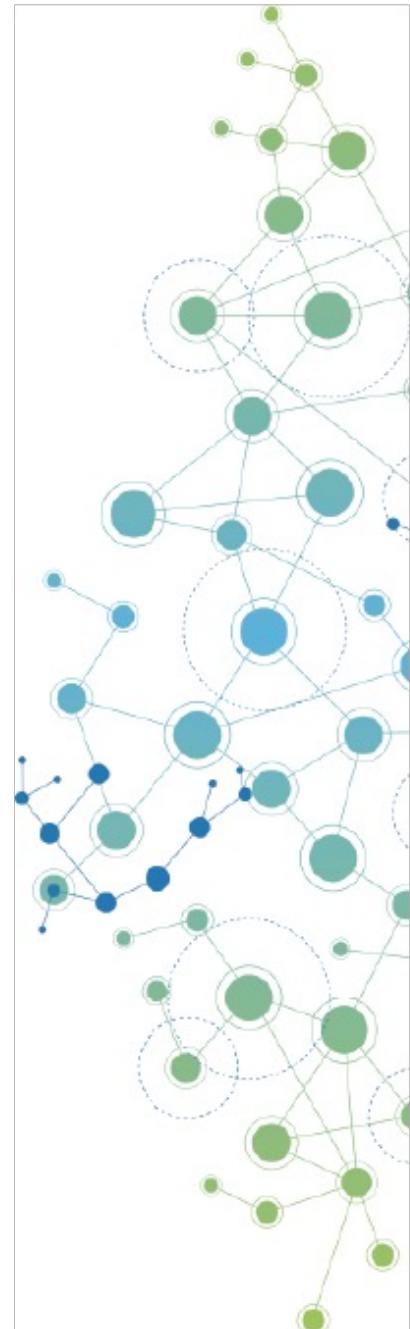
Estado Civil	Turnover					Probabilidade de Turnover	IV
	Sim	Não	Total Geral	% Sim	% Não		
Solteiro	120	350	470	51%	28%	26%	0,13
Casado	84	589	673	35%	48%	12%	0,04
Divorciado	33	294	327	14%	24%	10%	0,05
<b>Total Geral</b>	<b>237</b>	<b>1233</b>	<b>1470</b>			<b>16%</b>	<b>0,22</b>

Colaboradores solteiros têm probabilidade de 26% de turnover, 10 pp acima da média geral. Hipóteses: Solteiros podem estar relacionados com baixa idade, que já foi analisada no slide anterior. O problema aqui é que não podemos fazer nada em relação ao estado civil devido à questão ética envolvida.

Equilíbrio de Vida	Turnover					Probabilidade de Turnover	IV
	Sim	Não	Total Geral	% Sim	% Não		
Ruim	25	55	80	11%	4%	31%	0,05
Bom	58	286	344	24%	23%	17%	0,00
Muito bom	127	766	893	54%	62%	14%	0,01
Excelente	27	126	153	11%	10%	18%	0,00
<b>Total Geral</b>	<b>237</b>	<b>1233</b>	<b>1470</b>			<b>16%</b>	<b>0,07</b>

Colaboradores com baixo equilíbrio de vida têm probabilidade de 31% de turnover, 15 pp acima da média geral. Hipóteses: Da mesma forma que o E-Sat, é importante entender as causas do baixo equilíbrio. Altas horas extras? Frequência de Viagens maior?

## 2) Planos de ação sugeridos e próximos passos



# Projeto: Minimização de Turnover

## Planos de ação



Com base nos *insights* obtidos anteriormente, sugerimos os seguintes planos de ação:

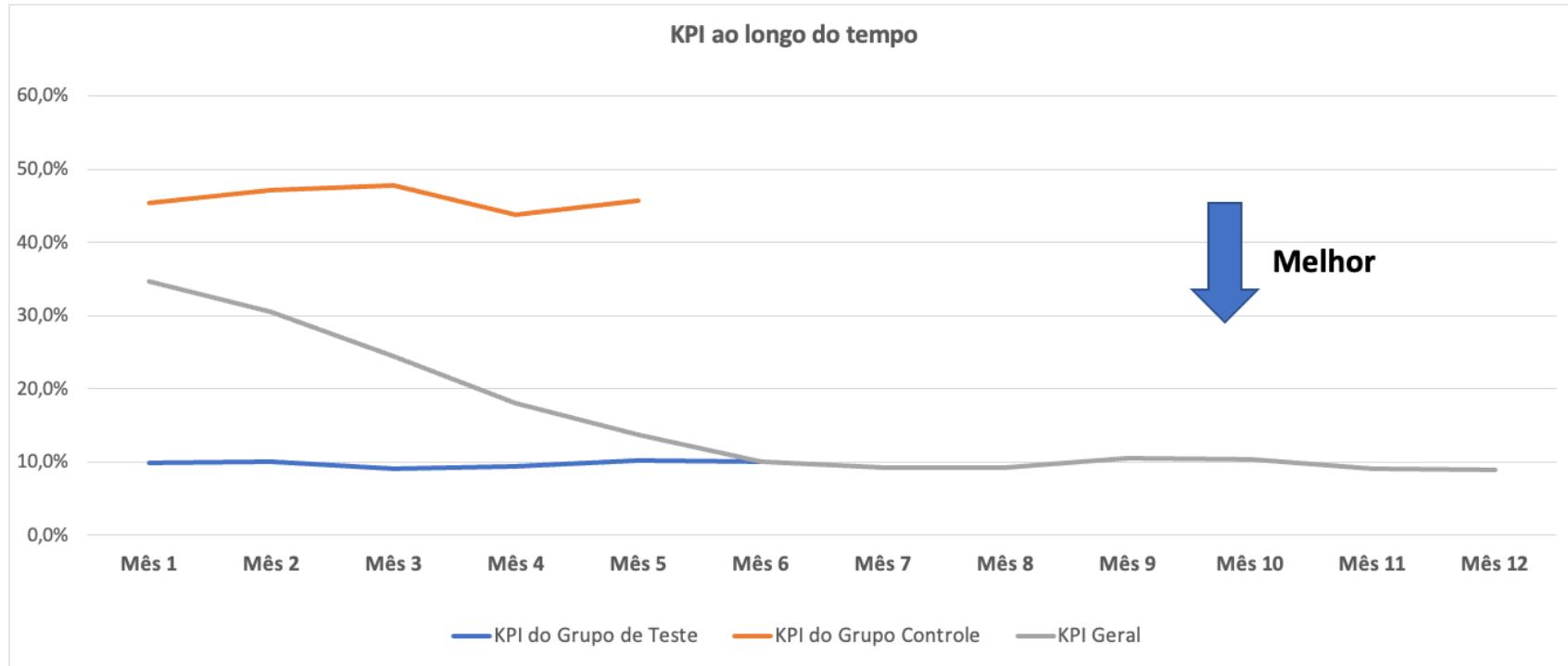
Prioridade	O que fazer?	Por que?	Quem?	Como fazer?	Qual o custo?
1	Revisão da política de Horas Extras	Colaboradores que fazem hora extra têm probabilidade de 31% de turnover.	Time de RH	Criar um grupo controle e teste de pessoas que podem fazer hora extra e não podem. Assim podemos mensurar a real diminuição de turnover ao longo do tempo com KPI's e Dashboards de acompanhamento.	Diminuição da produtividade de algumas áreas que precisam que os colaboradores trabalhem mais horas.
2	Revisão salarial	Colaboradores que ganham até 3000 reais têm probabilidade de 29% de turnover.	Time de Remuneração	Verificar se os salários da empresa estão atualizados de acordo com o mercado. Caso não estejam, estudar a possibilidade de adequação salarial.	Aumento da folha de pagamento.
3	Revisão da política de Viagens Corporativas	Colaboradores que viajam frequentemente têm probabilidade de 25% de turnover.	Time de RH	Criar um grupo controle e teste de pessoas que podem viajar e não podem. Assim podemos mensurar a real diminuição de turnover ao longo do tempo com KPI's e Dashboards de acompanhamento.  Além disso, intensificar ferramentas de trabalho remoto na empresa, mudando processos e a cultura.	Diminuição da produtividade de algumas áreas que precisam que os colaboradores trabalhem localmente. No entanto, pode diminuir o custo com alternativas de trabalho remoto.
4	Revisão da Integração e Execução da Cultura	Colaboradores com até 2 anos de empresa têm probabilidade de 30% de turnover.	Time de Cultura e Onboarding	Verificar se o onboarding dos colaboradores está adequado. Isso envolve acessos, integração com o time, liderança e feedback.  Também faz constantes mensurações de como a cultura organizacional está sendo seguida e executada pelos gestores.	Apenas horas de projeto dos times envolvidos.
5	Pesquisa de Satisfação mais detalhada	Colaboradores com baixa satisfação com o trabalho têm probabilidade de 25% de turnover.	Time de Cultura	Como não temos dados das causas da baixa satisfação, é importante conduzir uma pesquisa mais abrangente com perguntas qualitativas e abertas dos motivos para baixa satisfação.	Fornecedor externo com a ferramenta de pesquisa.

# Projeto: Minimização de Turnover

Exemplo de acompanhamento do estudo após implantação



Plano de Ação	Revisão da política de Horas Extras
Por que?	Colaboradores que fazem hora extra têm probabilidade de 31% de turnover.
O que foi feito?	Criação de grupo controle e teste de pessoas que podem fazer hora extra e não podem.
KPI em acompanhamento	Percentual de Turnover (% de colaboradores que deixaram a empresa)



	Mês 1	Mês 2	Mês 3	Mês 4	Mês 5	Mês 6	Mês 7	Mês 8	Mês 9	Mês 10	Mês 11	Mês 12
--	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------

% do Grupo Teste (grupo com aplicação da mudança)	30%	45%	60%	75%	90%	100%	100%	100%	100%	100%	100%	100%
% do Grupo Controle	70%	55%	40%	25%	10%	0%	0%	0%	0%	0%	0%	0%

KPI do Grupo de Teste	9,8%	10,1%	9,0%	9,4%	10,2%	10,0%	9,2%	9,3%	10,6%	10,3%	9,1%	9,1%
KPI do Grupo Controle	45,4%	47,2%	47,8%	43,7%	45,7%	0%	0%	0%	0%	0%	0%	0%
KPI Geral	34,7%	30,5%	24,5%	18,0%	13,8%	10,0%	9,2%	9,3%	10,6%	10,3%	9,1%	9,0%

# Projeto: Minimização de Turnover

## Próximos passos



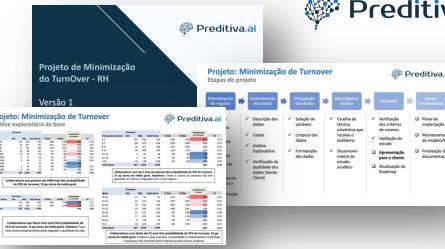
O que vem em seguida?

- Alinhamento com os times de RH sobre os planos de ação** que serão executados. Criação de cronograma para execução desses planos.
- Criação de um **dashboard** de acompanhamento do Turnover com os grupos controle e teste de cada plano de ação sugerido.
- Extração automática dos dados** para acompanhamento dos planos de ação.
- Inclusão de novas variáveis** e um período maior de tempo para análise.
- Considerar outros tipos de turnover**, como o demissional. Assim podemos entender os colaboradores mais propensos a serem desligados.
- Realizar o cálculo do impacto que o turnover traz** para a empresa. Assim podemos ter uma estimativa financeira do problema.
- Criação de um modelo de propensão de turnover**. Assim os times de RH podem saber os colaboradores com maior risco de turnover e tomar ações preventivas.
- Criação de um **modelo preditivo** de turnover com base em informações capturadas no momento da contratação. A ideia é ter uma boa noção se um candidato tem chances de pedir demissão nos próximos x meses após a contratação.

Viu a diferença?

# Comparação entre os dois trabalhos

Vamos entender



Preditiva.ai

<b>Objetivo claro</b>	Sem objetivo claro e contextualizado	Objetivo e contexto fornecido logo no início
<b>Premissas (datas de referência, filtros etc)</b>	Sem premissas, datas nem nada	Premissas alinhadas logo de início
<b>Esforço Mental e Visual</b>	Grande	Pequeno - Foco no que importa!
<b>Psicologia das cores</b>	Errada - Cor Vermelha em Uma formação	Correta - Vermelho nas categorias de maior turnover
<b>Saturação e Contraste</b>	Cores fortes e pesadas (verde escuro)	Cores mais brandas com saturação adequada
<b>Rótulos e Legendas</b>	Tabelas mal formatadas	Tabelas formatadas com destaque para o mais importante com formatação condicional
<b>Tamanho de Textos</b>	Tamanho adequado	Tamanho adequado
<b>Metodologia usada</b>	Não apresentada. Trabalho sem ordem lógica.	Apresentada - CRISP-DM
<b>Técnicas adequadas</b>	Não. Comparações incorretas com médias e proporções em gráficos de pizza	Sim, com técnica de associação IV trazendo as variáveis que mais estão associadas com o Turn Over
<b>Riqueza e aplicabilidade dos planos de ação</b>	Não se avaliou todas as possibilidades de variáveis. As conclusões ficaram fragilizadas devido às técnicas incorretas para o problema	Todas as variáveis foram analisadas e os planos de ação foram gerados por ordem de prioridade de impacto no turnover
<b>Próximos passos e chamada para ação</b>	Não mostrados e sem chamada para ação. Risco alto de nada ser feito após o trabalho	Vários passos alinhados e sugeridos, acima como uma chamada para ação forte, gerando engajamento

**Legal Vini, maaaaas, como aprendo tudo isso e muito mais?**

**Já conhece a Preditiva?**



# Preditiva.ai

Somos a única escola **100% focada** no ensino  
de Ciência de Dados e Analytics

A **Preditiva** nasceu do sonho de ajudar os brasileiros a ter excelentes oportunidades no mercado de trabalho através da **Alfabetização de Dados** e do desenvolvimento de uma Mentalidade Analítica.

Para empresas, ajudamos a preparar seus times, facilitando a transição para uma **cultura Data-Driven**, com decisões de negócio totalmente orientadas por Dados.

# Formação Gerando Valor com Dados

Aprenda as técnicas, ferramentas e metodologias que estão revolucionando os negócios



## Conteúdo

São mais de 160h (e aumentando), considerando:

- Aulas gravadas
- Bônus de Carreira
- Mentorias
- Exercícios e Aplicações
- Estudos de Caso dirigidos
- Projetos de Dados



## Suporte ao Aluno

Tire suas dúvidas utilizando o espaço de perguntas em cada aula. Nossos professores estão sempre à disposição para te ajudar



## Mentorias ao Vivo (Semanais)

Nesses encontros, aproveite para:

- Tirar dúvidas sobre conteúdo: técnicas e ferramentas
- Fazer networking com professores Preditiva a outros alunos
- Falar sobre desafios profissionais, carreira e outros



## Acesso imediato e disponível por 2 anos

Nesse período aproveite ao máximo assistindo aos conteúdos, participando das mentorias ao vivo e enviando perguntas aos professores quantas vezes quiser, tudo no seu ritmo de aprendizado

**Inscreva-se agora**

**15 dias de garantia incondicional**



# Preditiva.ai

Escola de Dados

[www.preditiva.ai](http://www.preditiva.ai)  
[comercial@preditiva.ai](mailto:comercial@preditiva.ai)

## Venha embarcar conosco nessa jornada!

Siga-nos nas redes sociais



[@preditiva.analytics](https://www.instagram.com/preditiva.analytics)



[www.linkedin.com/school/preditiva-analytics](https://www.linkedin.com/school/preditiva-analytics)





Preditiva.ai