

Seminário Criação de Diagnósticos e Planejamento rumo à Agenda 2030

Ciência de Dados para a Criação de Diagnósticos

Fabio Correa Xavier

Diretor do Departamento de Tecnologia da Informação – TCESP

Auditório nobre 'Professor José Luiz de Anhaia Mello'

08/11/2018

Agenda

- ▶ **Introdução**
- ▶ **Os dados**
- ▶ ***Data Science* - principais ferramentas e técnicas**
- ▶ **Trabalhando com um objetivo**
- ▶ **Um modelo simples**
- ▶ **Aperfeiçoando o modelo**
- ▶ **Considerações finais**

... e se fosse possível prever um surto de uma doença, *antes* que ele ocorresse?



Ferramenta usa dados do Google para prever surtos de gripe



PNAS Proceedings of the
National Academy of Sciences
of the United States of America

Keyword, Author, o

Home Articles Front Matter News Podcasts Authors

NEW RESEARCH IN

Physical Sciences

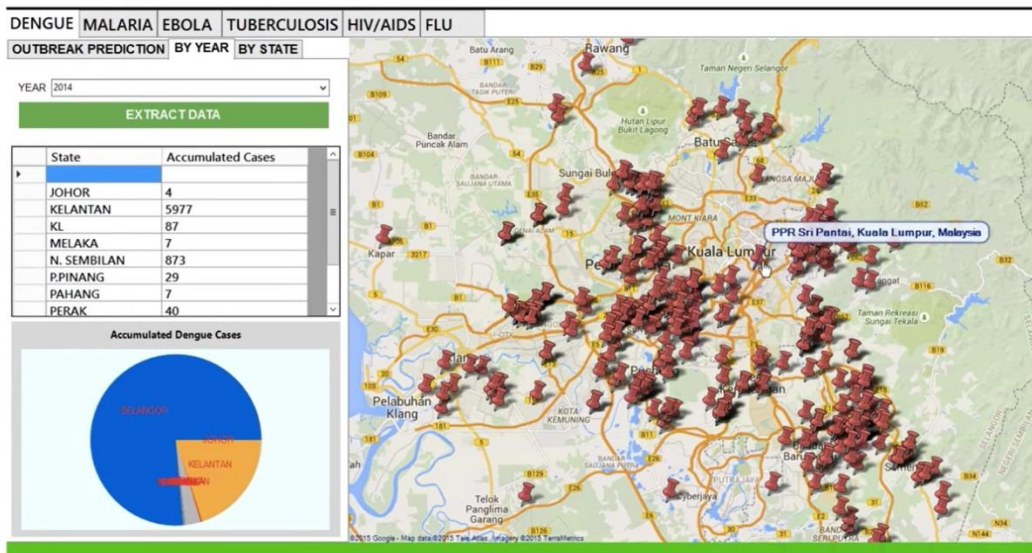
Social Sciences

**Accurate estimation of influenza epidemics
using Google search data via ARGO**



Início » Ciência » Startup usa inteligência artificial para prever surtos de dengue no Brasil

Startup usa inteligência artificial para prever surtos de dengue no Brasil



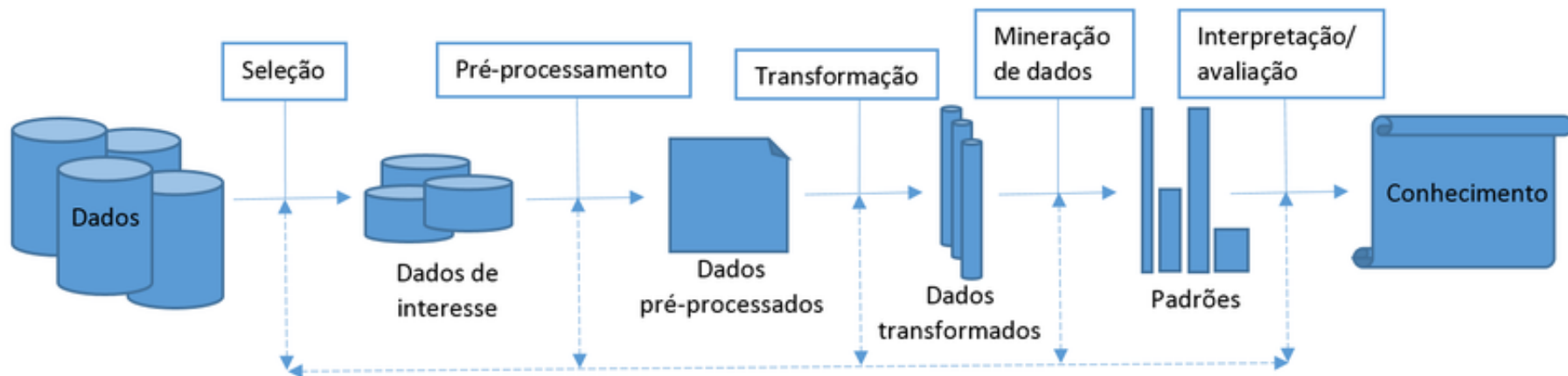
***“Data Science* consiste em **INTERROGAR** os dados até que eles confessem.”**

► Interrogar os dados? Quais deles???

- Acessibilidade
- Complexidade
- Diversidade
- Volume REALMENTE GIGANTESCO!



- ▶ **KDD – Extração de Conhecimento**
 - Por meio de relacionamentos de interesse



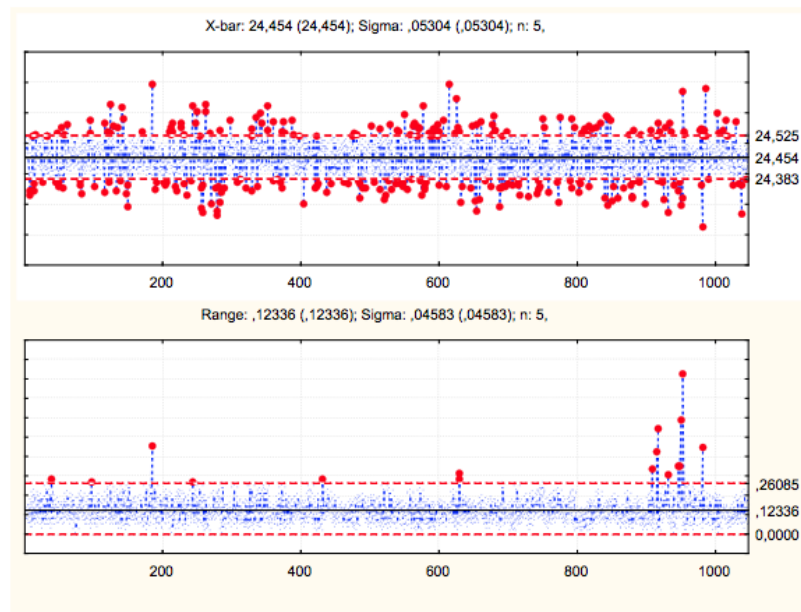
- ▶ **Análise de dados – definição de metodologia**

- ▶ ***Data Mining* – Mineração de Dados**
 - Exploração de grande quantidade de dados
 - Procura de padrões consistentes
 - Regras de associação
 - Sequências temporais
 - Detecção de relacionamentos entre variáveis

- ▶ ***Text Mining* - Mineração de Textos**

- ▶ **Clusterização**

- ▶ **Lógica *Fuzzy* (Análise de Palavras com *Fuzzy Matching*)**
- ▶ **Estatística Descritiva**
- ▶ **Regressão Linear**
- ▶ **Detecção de *Outliers*** ➔
- ▶ ***Web Crawler***



Ficou complicado? Comece pelo simples!

- ▶ **No início, simplicidade é tudo!**
 - Análises simples, mas precisas, que reflitam a realidade

- ▶ **Abordagem iterativa e incremental**
 - Reutilize o conhecimento adquirido nas iterações anteriores
 - Aumente a complexidade do modelo a cada iteração

- ▶ **Reanálise periódica dos dados**
 - Use o mesmo modelo, critérios de extração, exploração de grande quantidade de dados
 - Entenda as mudanças ocorridas ao longo do tempo

- ▶ **Importante: Aprenda com os dados!**

Comece analisando o objetivo a ser alcançado



4.2 Até 2030, garantir que todos as meninas e meninos tenham acesso a um desenvolvimento de qualidade na primeira infância, cuidados e educação pré-escolar, de modo que eles estejam prontos para o ensino primário

Como prever quantas vagas de creche o meu Município precisa criar até 2030?

X

► Demanda por vagas (em 2017)



- Vagas existentes: **5.636**
- Demanda por vagas: **7.944**
- Demanda não atendida: **2.308** vagas (**40,95%**)

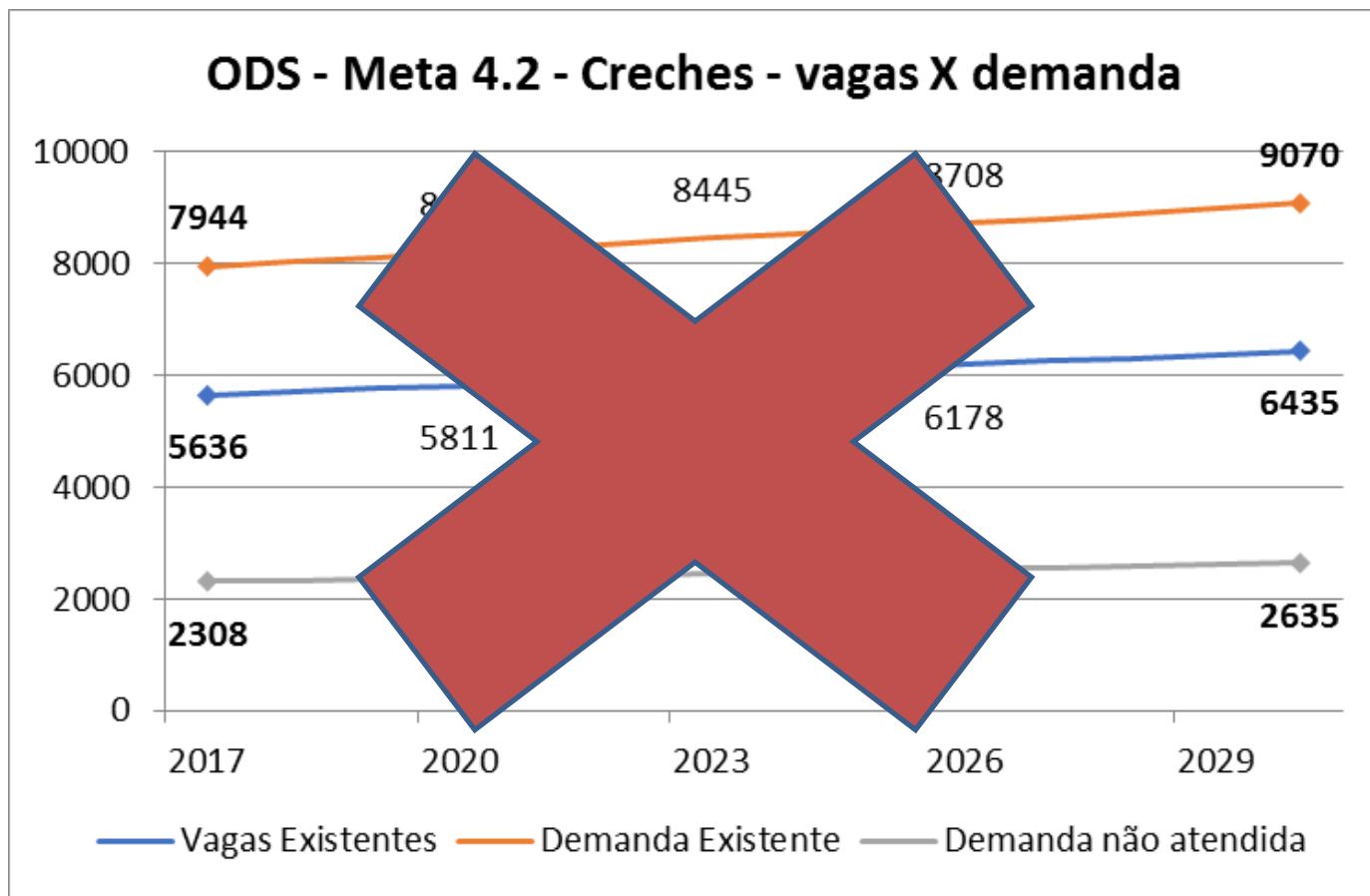
► Cálculo da taxa de crescimento populacional



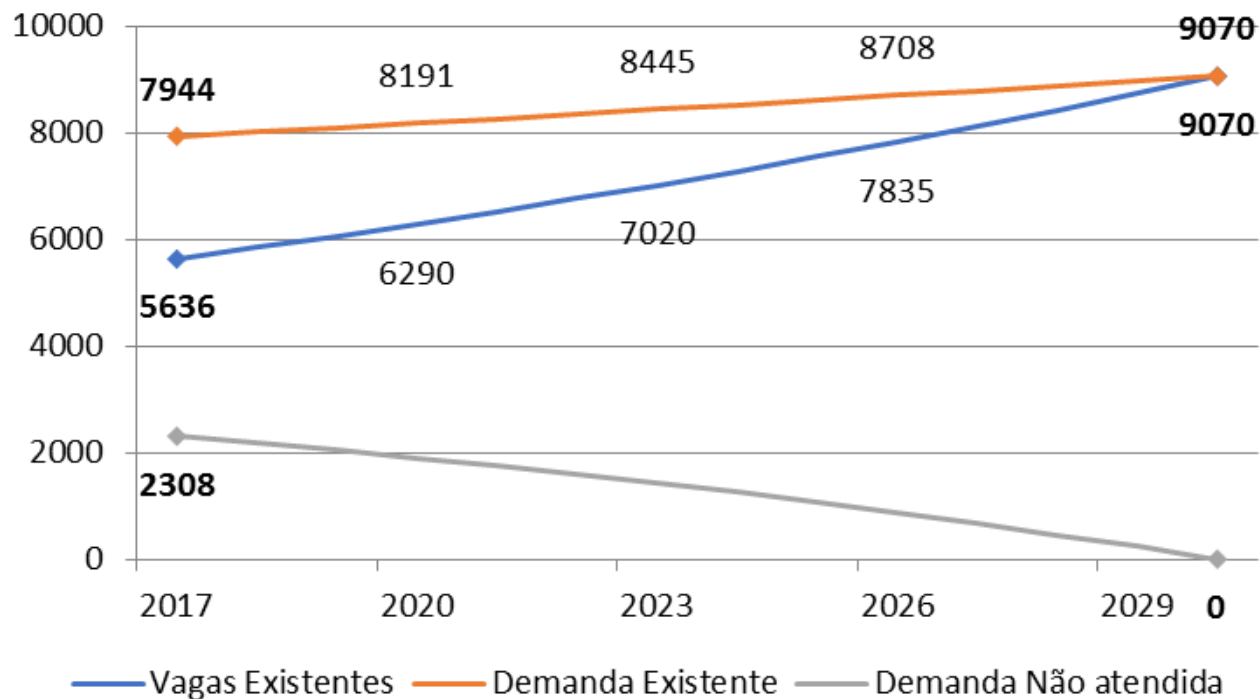
Períodos ▾	<u>População</u>
2014	273.854
2015	276.852
2016	279.626
2017	282.428
2018	285.257

- Taxa de Crescimento Populacional: **1,03% a.a**
- Fórmula:

$$\left[\left(\frac{f}{s} \right)^{\frac{1}{y}} - 1 \right] \times 100$$



ODS - Meta 4.2 - Creches - vagas X demanda



Tx. Crescimento da demanda: **1,03% a.a**

Tx. Crescimento das vagas: **3,73% a.a**

SEADE

Fundação Sistema Estadual
de Análise de Dados

Períodos ▾	População
2014	273.854
2015	276.852
2016	279.626
2017	282.428
2018	285.257

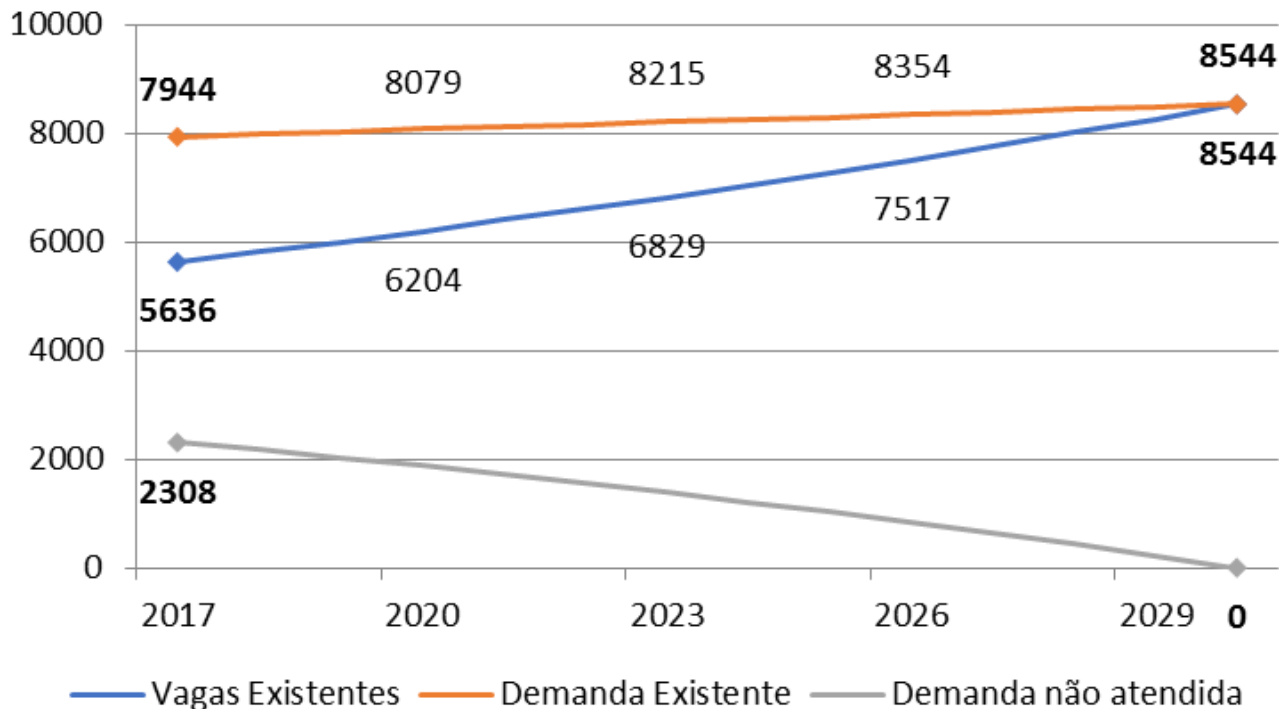
Períodos ▾	População de 0 a 4 Anos
2014	20.358
2015	20.750
2016	20.783
2017	20.807
2018	20.819

Ou

- Crescimento da demanda: **1,03% a.a**
- Crescimento de vagas: **3,73% a.a**
- Demanda total em 2030: **9.070**

- Crescimento da demanda: **0,56% a.a**
- Crescimento de vagas: **3,25% a.a**
- Demanda total em 2030: **8.544**

ODS - Meta 4.2 - Creches - vagas X demanda



Tx. Crescimento da
demanda: **0,56% a.a**

Tx. Crescimento
das vagas: **3,25% a.a**

Economia: **526 vagas**

▶ Zona Leste

- Vagas Existentes: **1.244 (22,08%)**
- Total de vagas necessárias em 2030: **1.886**
- Total de novas vagas criadas: **642**

▶ Zona Oeste

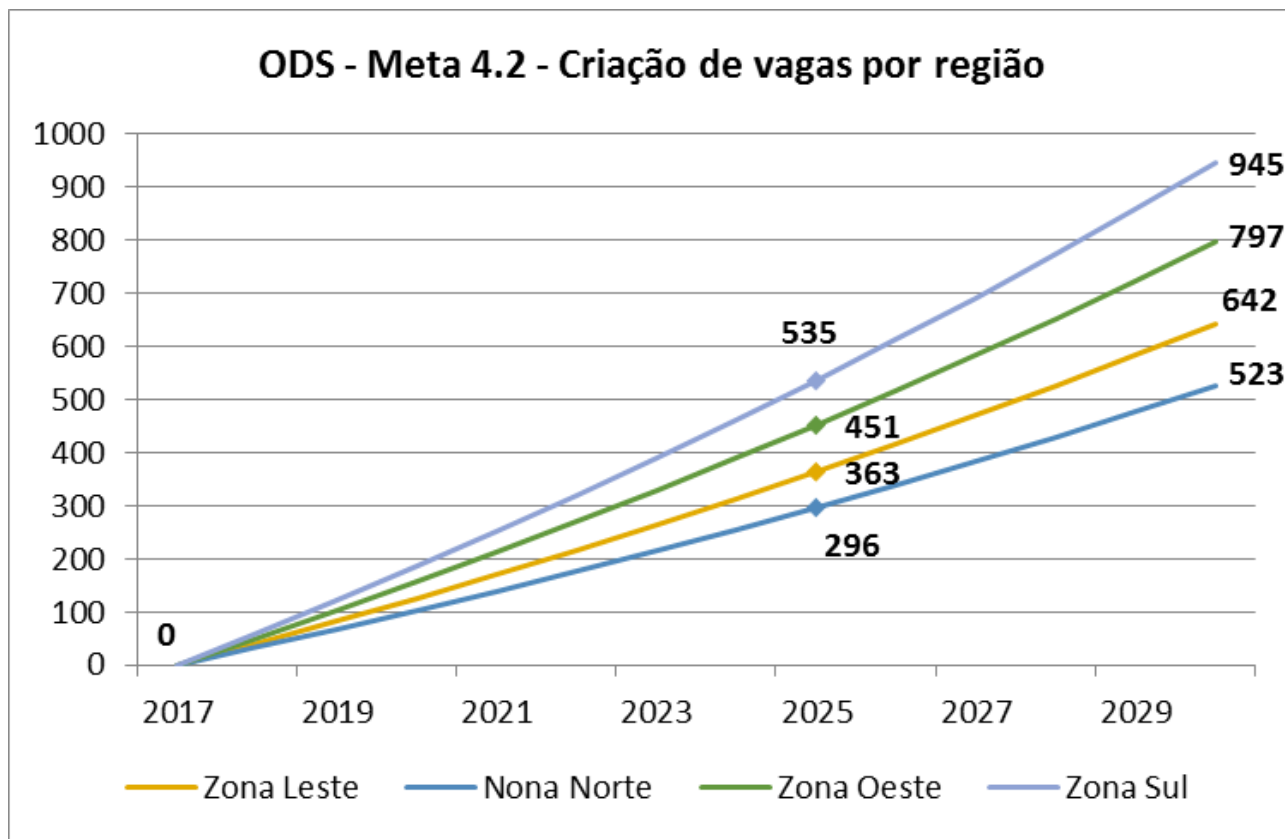
- Vagas Existentes: **1.545 (27,41%)**
- Total de vagas necessárias em 2030: **2.341**
- Total de vagas criadas: **797**

▶ Zona Norte

- Vagas Existentes: **1.014 (18,00%)**
- Total de vagas necessárias em 2030: **1.537**
- Total de vagas criadas: **523**

▶ Zona Sul

- Vagas Existentes: **1.832 (32,51%)**
- Total de vagas necessárias em 2030: **2.777**
- Total de vagas criadas: **945**





Educação

Pesquisa do IBGE revela que ocupação de creches tem impacto na ocupação das mães

O nível de ocupação das mulheres tem relação direta com a frequência de seus filhos a creches no país. Segundo estudo do Instituto Brasileiro de Geografia e Estatística (IBGE), com base em dados do Censo 2010, das mulheres que tinham filhos de até 3 anos em creche, 64% tinham emprego. Entre as mulheres com filhos que não frequentavam creche, o percentual era 41,2%.

▶ Zona Leste

- Renda per capita: **R\$ 693,12**
- Proporção inversa: **29,25**
- Total de novas vagas criadas: **850**
- Total de novas vagas antes: ~~642~~

▶ Zona Norte

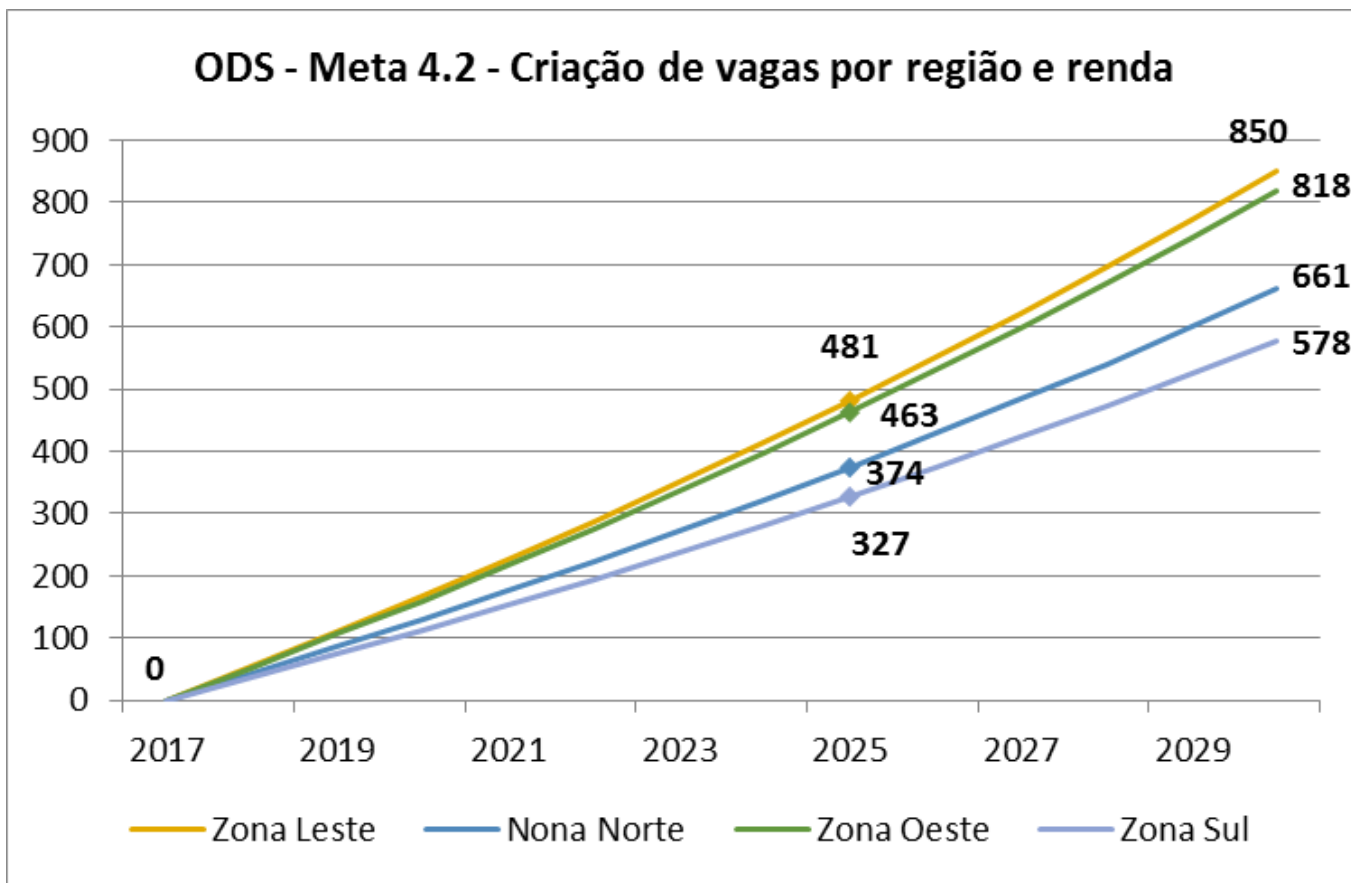
- Renda per capita: **R\$ 891,63**
- Proporção inversa: **22,74**
- Total de novas vagas criadas: **661**
- Total de novas vagas antes: ~~523~~

▶ Zona Oeste

- Renda per capita: **R\$ 720,78**
- Proporção inversa: **28,12**
- Total de novas vagas criadas: **818**
- Total de novas vagas antes: ~~797~~

▶ Zona Sul

- Renda per capita: **R\$ 1.019,42**
- Proporção inversa: **19,89**
- Total de novas vagas criadas: **578**
- Total de novas vagas antes: ~~945~~



- ▶ **Revisite seus dados e suas análises periodicamente**
 - Procure erros, incoerências, dados incorretos ou imprecisos
 - Verifique se o seu modelo espelha a realidade. Verifique os resultados!

- ▶ **Aprenda com os dados dos outros**
 - Se alguém já gastou \$\$\$ para desenvolver um estudo que se aplica à sua realidade, não gaste tudo de novo. Use-o!

- ▶ **Nenhum dado estatístico substitui o trabalho de campo**
 - São complementares. Teste os seus modelos *in loco*!
 - Colha os dados da sua localidade em campo. Maior chance de precisão

Seminário Criação de Diagnósticos e Planejamento rumo à Agenda 2030

Ciência de dados para criação de diagnósticos

Fabio Correa Xavier

OBRIGADO!

