

Dados e significados: Uma ilustração de aplicação *analytics*

Pedro Jácome de Moura Jr*

1. INTRODUÇÃO

Administradores vêm enfrentando o desafio de migrar o modelo decisório de suas organizações de uma fundamentação em intuição e heurísticas para uma fundamentação em dados e suas relações. Organizações orientadas por dados (ou *data-driven organizations*) esforçam-se para coletar, tratar, armazenar, organizar e utilizar dados relevantes para os seus processos operacionais e decisórios (McAfee *et al.*, 2012), numa clara evolução de conceitos prévios relacionados à gestão do conhecimento (p. ex: Davenport & Prusak, 1998) e do conhecimento como recurso capaz de produzir e manter vantagens competitivas (Grant, 1996).

Esse desafio compreende mudanças relacionadas à liderança, gestão de talentos (ou competências e habilidades), tecnologia, decisão e cultura organizacional (McAfee *et al.*, 2012). O aspecto central da mudança, no entanto, é a atenção aos dados, num espectro que contempla desde a criação/identificação/captura do dado até o seu emprego. Depreende-se, portanto, que esse espectro demande significativa multidisciplinaridade de competências. No âmbito da Administração, mais especificamente no âmbito da tomada de decisão como aspecto central da gestão (Simon, 1950), o foco desloca-se para um dos extremos do espectro citado, concentrando-se no emprego/uso do dado. Emprego/uso do dado é uma atividade tipicamente analítica (*business analytics* ou *data analytics*) e compreende análises descritivas, diagnósticas, preditivas e prescritivas (Porter & Heppelmann, 2015).

Entende-se que o objetivo central de *analytics* é extrair significados que não são evidenciados diretamente pelos dados originais (Agarwal & Dhar, 2014). Assim, depreende-se que as análises descritivas apresentam o comportamento dos dados tal como se apresenta após as etapas iniciais de coleta, tratamento, armazenamento e organização em bancos de dados; as análises diagnósticas apresentam inferências explicativas para o comportamento descrito; as análises preditivas apresentam antecipações de comportamentos futuros, lastreados no comportamento passado e em suas explicações; e as análises prescritivas apresentam recomendações de ações para que o comportamento desejado ocorra. A iteração e interação dessas “modalidades” *analytics* encontra respaldo no desenvolvimento e validação de teorias (vide Agarwal & Dhar, 2014; Gregor, 2006) e no desenvolvimento e validação de teorias fundamentadas em dados (*abductive theory/grounded theory*; Haig, 2005).

Com o objetivo de ilustrar *analytics* na prática administrativa, este documento sintetiza o processo de desenvolvimento de um aplicativo de software e o seu emprego em análises descritiva, diagnóstica, preditiva e prescritiva. O aplicativo foi desenvolvido e apresentado como parte do tópico TI03-Gestão e Ciência dos Dados, do curso Lideranças em Tecnologia da Informação, promovido pela Escola Nacional de Administração Pública (Enap) no período de 09 a 15 de março de 2018 e ministrado por este autor.

2. METODOLOGIA

Optou-se por adotar a modelagem dimensional e processamento analítico em tempo real (*online analytical processing*; OLAP) para desenvolvimento do aplicativo (Kimball & Ross, 2011). Adotou-se a plataforma Qlik Sense Cloud para desenvolvimento, em versão livre de licenciamento. O aplicativo OLAP (app, daqui em diante) foi desenvolvido para contemplar variadas perspectivas de análise, embora não exaustivamente. Os dados analisados foram selecionados arbitrariamente, considerando o interesse deste autor em ilustrar eventuais subtemas correlatos.

Os dados analisados são fornecidos pelo Tribunal de Contas da União (TCU) em sua página web (<http://portal.tcu.gov.br/comunidades/fiscalizacao-de-tecnologia-da-informacao/atuacao/perfil-de-governanca-de-ti/>) e não possuem identificação das instituições respondentes. Os dados apresentados, independentemente do nível de agregação, bem como as interpretações fornecidas, representam aqui interesse meramente ilustrativo e didático.

O formato inicial dos dados é planilha eletrônica (extensão .xlsx), contendo 157 variáveis medidas em escala de 1 a 5 com os seguintes significados: 1-Não se aplica; 2-Não adota; 3-Iniciou plano

* **Vinculação profissional:** Professor da Universidade Federal da Paraíba e Editor adjunto da MEPAD – Métodos e pesquisa em Administração.
Email: oitentabarradois@hotmail.com

para adotar; 4-Adota parcialmente; e 5-Adota integralmente. As variáveis estão agrupadas em 25 dimensões que tratam governança corporativa e governança de TI e os dados finais divulgados (após processamento pelo TCU) incluem o índice iGovTI2016, calculado segundo metodologia do TCU.

A primeira “pasta” do app contém artefatos de análise (gráficos, tabelas, indicadores etc) que procuram relacionar dados sobre governança corporativa e governança de TI; a segunda mantém essa relação, mas propõe a identificação de correlações estatísticas relevantes entre os dois níveis de governança; a terceira pasta relaciona dados dos planos de desenvolvimento institucionais (PDI) com dados dos planos diretores de TI (PDTI); a quarta pasta oferece artefatos para análise de governança (nos dois níveis já citados) e planos de dados abertos (PDA); a quinta pasta avança nessas relações ao relacionar PDA com dados de transparência e catálogo de serviços; e, por fim, a sexta pasta apresenta incipiente implementação de análise prescritiva baseada em aprendizagem de máquina, por meio de regressão linear simples.

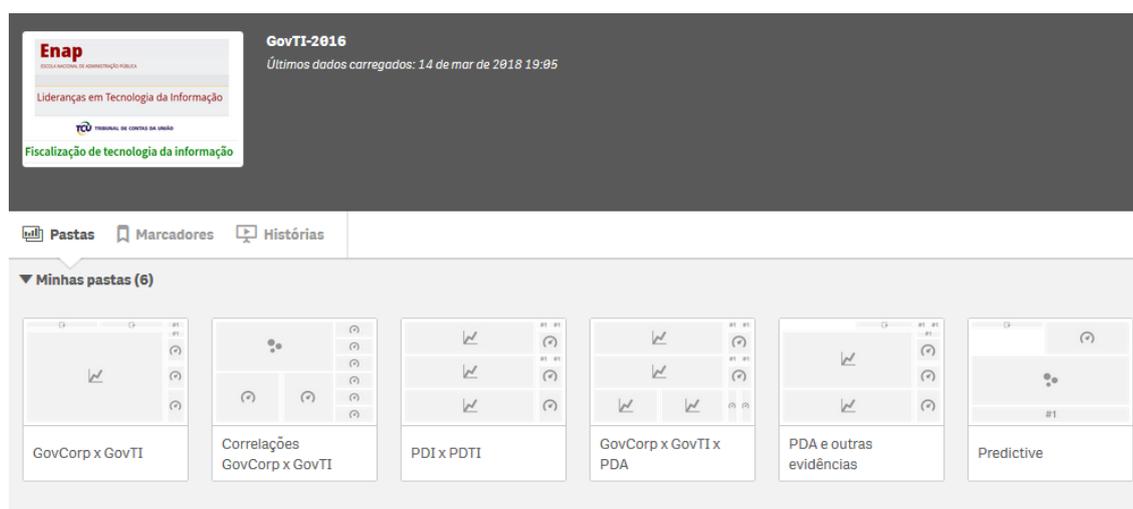


Figura 1 - Pastas de trabalho do aplicativo

3. RESULTADOS

Análises descritivas

As análises descritivas propostas são suportadas visualmente no app por meio de gráficos como o da Figura 2 (linhas). Observam-se comportamentos distintos das curvas de desempenho das notas médias de governança corporativa (linha azul) e governança de TI (linha vermelha). Como proposto, o artefato central (linhas) *descreve* uma determinada configuração dos dados. Há ainda alguns coeficientes de correlação, apresentados por meio de artefatos do tipo “mostrador com ponteiros”. Por se tratar de artefato voltado também para análises diagnósticas e preditivas, serão abordados oportunamente.

GovCorp x GovTI

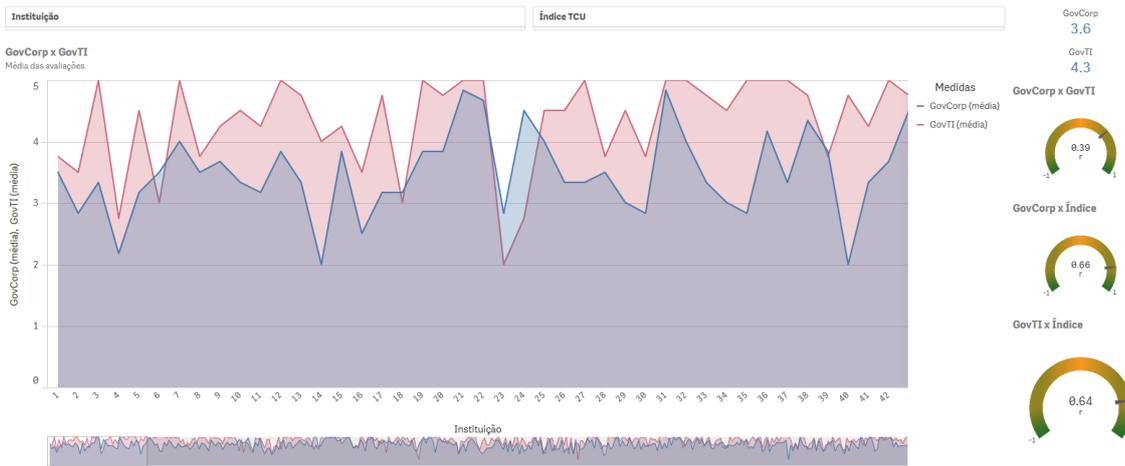


Figura 2 - Governança corporativa e governança de TI

PDI x PDTI

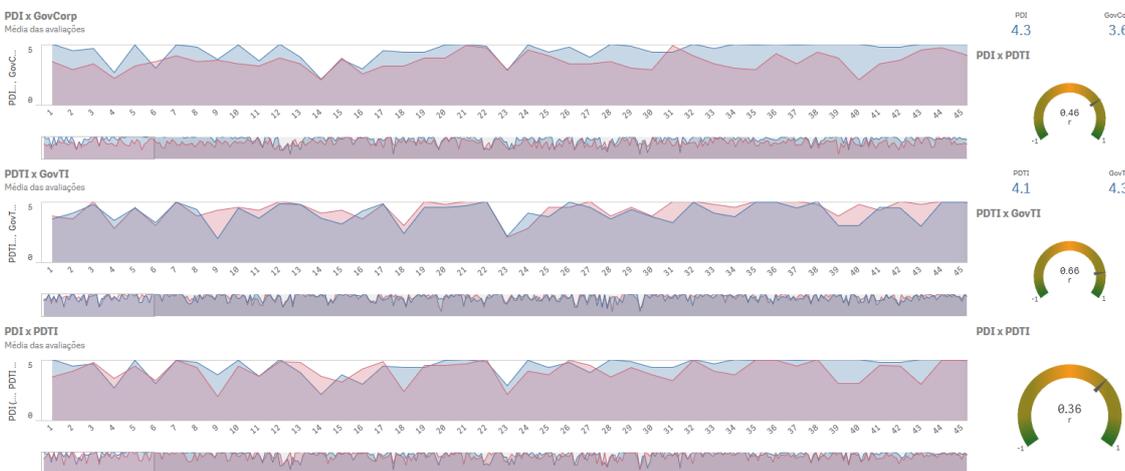


Figura 3 - Governança corporativa, PDI, PDTI e governança de TI

GovCorp x GovTI x PDA

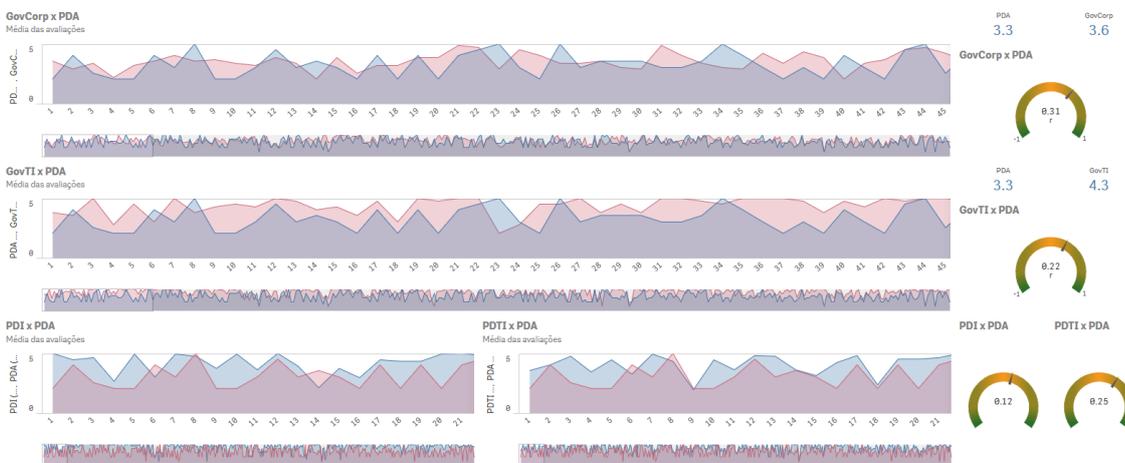


Figura 4 - Governança corporativa, PDA, PDI e governança de TI

PDA e outras evidências

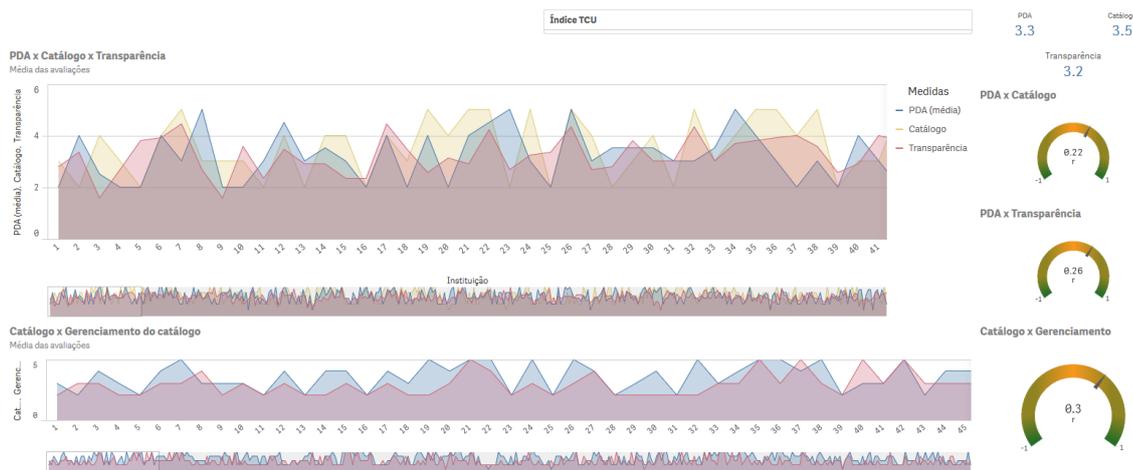


Figura 5 - PDA, catálogo de serviços e transparência

Possíveis análises diagnósticas

As análises descritivas sugerem níveis distintos entre os construtos analisados (governança, PDI, PDA etc.). O que provocaria as diferenças entre os níveis de governança corporativa e de TI, por exemplo? Uma possível explicação para os maiores níveis de governança de TI é a maior facilidade da área em lidar com padrões, protocolos e normativas.

Outra possível explicação seria falha no instrumento de coleta de dados, que pode dificultar o entendimento dos conceitos e, por conseguinte, a elaboração das respostas. Ainda, a eventual dificuldade do respondente (geralmente o CIO) em obter dados sobre governança corporativa, o que poderia produzir distorções nas respostas.

Ou seja, para cada situação identificada ou exposta pelos/nos dados, o analista atíça a sua curiosidade, busca novas relações e faz inferências como as aqui sugeridas. Assim, a título de exercício, quais outras situações os dados expõem e merecem ilações? As Figuras 2 a 5 podem oferecer subsídios para essas análises diagnósticas.

Análises preditivas

As análises preditivas estão baseadas e restritas aqui a correlações e regressões. Sabe-se que variáveis com histórica correlação forte ($\geq |.7|$) podem ser empregadas em estimativas de valores futuros, caso o valor de uma passe a ser desconhecido. Para inferências causais, no entanto, correlações não são suficientes, o que requer o emprego de outras técnicas. Este app implementa uma das mais comuns, em sua forma mais simples: regressão linear.

A análise das correlações (usando apenas por questões estéticas mostradores que simulam odômetros) sugere poucas variações conjuntas relevantes. A Figura 1 indica, por exemplo, que a governança corporativa e o índice iGovTI calculado pelo TCU possuem $r = 0.66$ (coeficiente de correlação). Isso sugere que quando a nota média das variáveis que medem governança corporativa varia, o iGovTI varia na mesma direção (p.ex: quando o valor de uma aumenta, o da outra aumenta também, linearmente).

Assim, é fácil observar que poucas correlações apresentarem r forte, o que indica que boa parte das variáveis do banco de dados utilizado não estão correlacionadas de modo a subsidiar análises preditivas. Tal inferência se confirma com a implementação da regressão para verificar a capacidade de determinação de governança corporativa sobre o iGovTI. Observa-se que a governança corporativa explica em média 44% do valor iGovTI (r^2), o que sugere a influência de outras variáveis na determinação de mais de 50% do índice.

Uma possibilidade de aprimoramento (e continuidade) dessa análise seria a implementação de regressão multivariada, buscando outras variáveis que possam aumentar o poder preditivo sobre o resultado iGovTI. Não obstante, as indicações prévias de baixas correlações não estimulam essa veia investigativa.

Correlações GovCorp x GovTI

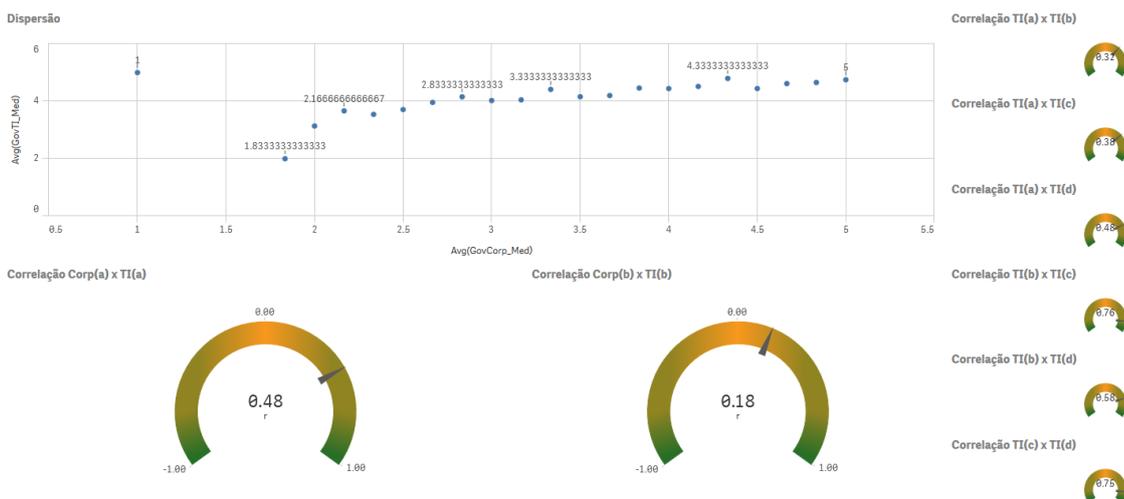
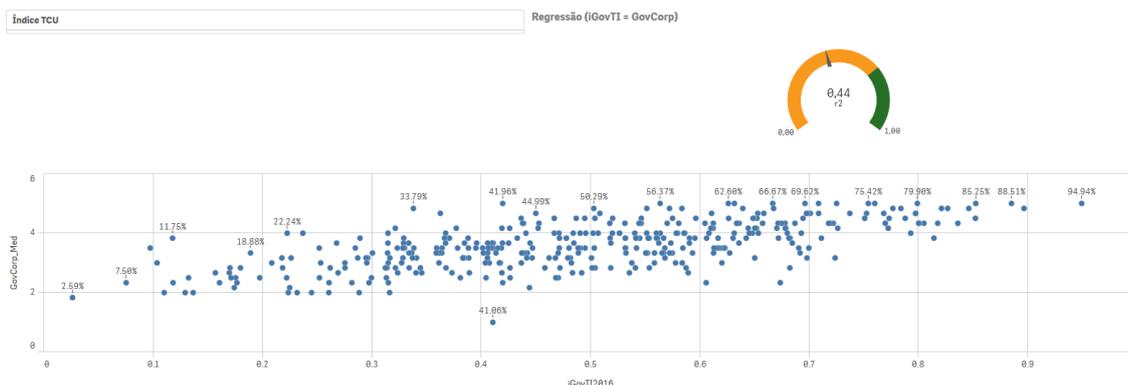


Figura 6 - Correlações entre variáveis de governança corporativa e de TI

Predictive



Equação da regressão
 $y = -0.05914162 + 0.15283152x$

Figura 7 - Regressão com iGovTI como variável dependente e governança corporativa como independente

Análise prescritiva

O percurso analítico aqui apresentado poderia ser resumido em: (1) aquisição de conhecimento sobre o comportamento dos dados, inicialmente por meio de análises descritivas; (2) inferências diagnósticas a partir das descrições, buscam explicar o comportamento; (3) conhecendo-se o comportamento e o que o provoca, busca-se estimar comportamentos futuros; e (4) pode-se então recomendar procedimentos que produzam os efeitos esperados.

Essa sequência tradicional de análise pode ter sido empregada pelo TCU, quando propõe a indução da implantação da governança corporativa por meio da implantação da governança de TI. Este parece-nos um bom exemplo de prescrição sendo posta em prática.

REFERÊNCIAS

Agarwal, R., & Dhar, V. (2014). Big data, data science, and analytics: The opportunity and challenge for IS research. *Information Systems Research*, 25(3), 443-448.
 Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Harvard Business Press.
 Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, 17(S2), 109-122.
 Gregor, S. (2006). The nature of theory in information systems. *MIS quarterly*, 611-642.

- Haig, B. D. (2005). An abductive theory of scientific method. *Psychological methods*, 10(4), 371.
- Kimball, R., & Ross, M. (2011). *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons.
- McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, 90(10), 60-68.
- Porter, M. E., & Heppelmann, J. E. (2015). How smart, connected products are transforming companies. *Harvard Business Review*, 93(10), 96-114.
- Simon, H. A. (1950). *Administrative behavior*. New York: Macmillan.