



USO DE ANÁLISE DE CORRESPONDÊNCIA SIMPLES PARA EXAMINAR AS PRÁTICAS DE GESTÃO DE PROJETOS EM DATA SCIENCE

THE USE OF SIMPLE CORRESPONDENCE ANALYSIS TO EXAMINE PROJECT MANAGEMENT PRACTICES IN DATA SCIENCE.

Recebido em 11.10.2024 Aprovado em 21.11.2024

Avaliado pelo sistema double blind review

DOI: <https://doi.org/10.32888/cge.v12i3.65048>

Gabriel Henrique Pereira Nascimento

gabrielhenrique.hpn@gmail.com

Especialista em Data Science e Analytics pela USP/ESALQ – Piracicaba/São Paulo, Brasil

<https://orcid.org/0000-0001-9409-7947>

Murilo Henrique Tank Fortunato

mtank@live.com

Professor orientador do curso de Data Science, Pecege – Piracicaba/São Paulo/ Brasil

<https://orcid.org/0000-0002-1695-8120>

Resumo

Muitos projetos em ciência de dados enfrentam dificuldades, frequentemente devido à falta de atenção nas fases decisórias ou à ausência de uma metodologia estruturada para sua execução. A adoção de metodologias de gestão de projetos pode ser uma solução para os desafios que as organizações encontram na administração de iniciativas nesse setor. Este estudo teve como objetivo aplicar a Análise de Correspondência Simples, utilizando um algoritmo desenvolvido em R, para explorar a relação entre práticas de gestão de projetos em ciência de dados e seus principais benefícios. Como resultado, foram identificadas associações entre benefícios e práticas de gestão, alinhadas com a literatura existente, como: o cumprimento de prazos e a distribuição de tarefas nas equipes associados à metodologia Kanban; a satisfação do cliente relacionada às Metodologias Ágeis e Lean Manufacturing; e o engajamento da equipe e a obtenção de resultados eficazes vinculados ao SCRUM. Os achados deste estudo contribuem para fornecer diretrizes no processo de seleção e implementação de práticas de gestão de projetos pelas empresas, ao evidenciar e relacionar os principais benefícios dessas ferramentas no contexto da ciência de dados.

Palavras-chave: ANACOR. Benefícios Percebidos. SCRUM.

Abstract

Many data science projects face challenges, often due to a lack of attention during decision-making phases or the absence of a structured methodology for execution. Adopting project management methodologies can be a solution to the challenges organizations encounter in managing initiatives in this field. This study aimed to apply Simple Correspondence Analysis, using an algorithm developed in R, to explore the relationship between project management practices in data science and their key benefits. As a result, associations were identified between benefits and management practices, aligned with existing literature, such as: meeting deadlines and task distribution within teams associated with the Kanban methodology; customer satisfaction related to Agile and Lean Manufacturing methodologies; and team engagement and effective outcomes linked to SCRUM. The findings of this study contribute to providing guidelines for the selection and implementation of project management practices by companies, highlighting and relating the main benefits of these tools in the context of data science.

Keywords: ANACOR. Perceived Benefits. SCRUM.

Introdução

A ciéncia de dados (DS) é um campo emergente que se concentra na análise de dados para gerar insights que auxiliem na tomada de decisões (Escudero e Fortunato, 2024; Fernandes e Fortunato, 2024). Esse campo permite a identificação de correlações, classificações, previsões de eventos, reconhecimento de padrões e inferências (Das et al., 2015; Saltz et al., 2018). O crescimento acelerado de estudos em DS dentro das organizações tem levado a investimentos significativos em talentos e projetos na área (Saltz et al., 2018), principalmente devido à necessidade de gerenciar grandes volumes de dados para decisões eficazes. Apesar da crescente relevância do DS, a sua emergéncia traz diversos desafios relatados na literatura.

Um dos principais desafios enfrentados na ciéncia de dados está ligado à complexidade e diversidade dos dados, que exigem das empresas a capacidade de lidar com sua disponibilidade e análise, definir metas claras, engajar e coordenar as equipes envolvidas e estabelecer cronogramas realistas (Saltz, 2015). Além disso, muitos projetos em DS são iniciados apenas com base nas habilidades de analistas e equipes, sem um foco claro nos resultados esperados (Saltz, 2015). Isso resulta em uma alta taxa de falhas, que pode ser atribuída à negligéncia nas fases de tomada de decisão ou à falta de uma metodologia estruturada para a execução dos projetos (Grander et al., 2022; Saltz e Shamshurin, 2016; Saltz e Krasteva, 2022).

Nesse cenário, a adoção de metodologias estruturadas de gestão de projetos surge como uma solução viável para os desafios enfrentados pelas organizações no gerenciamento de projetos em DS e Big Data (Saltz et al., 2017b). A implementação de uma metodologia de projeto é crucial para reduzir falhas e aumentar as chances de sucesso em projetos de DS. Os benefícios são variados, permitindo que todos os participantes compreendam o processo como um todo, escolham as ferramentas de análise adequadas, concentrem-se no que realmente é importante e melhorem a fluidez do processo, resultando em resultados mais eficazes (Rybicki, 2019; Saltz, 2015).

A literatura revela um aumento no número de estudos sobre boas práticas de gestão de projetos em DS. Algumas metodologias se destacam nesse contexto, como: Metodologia Ágil, SCRUM, Kanban, Lean Manufacturing, CRISP-DM, KDDM, TDSP e Cookiecutter DS (Rybicki, 2019; Saltz e Crowston, 2017; Saltz et al., 2018; Saltz, 2021; Saltz e Krasteva, 2022; Saltz et al., 2022).

Embora o número de estudos sobre gestão de projetos em DS esteja crescendo, a literatura ainda carece de abordagens sistemáticas para organizar as principais boas práticas (Rybicki, 2019; Saltz e Crowston, 2017; Saltz et al., 2018) e de uma metodologia padronizada, representando uma lacuna significativa (Grander et al., 2022; Saltz e Krasteva, 2022). É essencial, portanto, identificar os principais benefícios que as ferramentas de gestão de projetos podem proporcionar no contexto específico de DS, a fim de auxiliar na seleção, implementação e execução de projetos mais eficientes e bem-sucedidos. A metodologia deve ser adaptada ao contexto de DS para garantir resultados eficazes (Saltz e Hotz, 2021).

Diante disso, o objetivo deste trabalho de conclusão de curso é aplicar a Análise de Correspondência Simples para examinar a relação entre as principais boas práticas de gestão de projetos em DS e os benefícios percebidos

de sua aplicação nas empresas. Para isso, um algoritmo foi desenvolvido em R para conduzir a análise, e um levantamento do tipo survey foi realizado para a coleta de dados, dada a sua viabilidade em termos de custo e agilidade (Miguel e Ho, 2010).

Este estudo visa contribuir para o entendimento do tema e fornecer insumos tanto para a comunidade acadêmica quanto para as organizações, no que diz respeito à seleção das práticas de projetos de DS mais adequadas a diferentes contextos, conforme seus principais benefícios. Além disso, compreender as tendências de boas práticas e seus benefícios pode ajudar a direcionar melhor os investimentos em projetos de DS, promovendo a obtenção de resultados positivos nas organizações (Saltz et al., 2018).

Gestão de projetos em ciência de dados

Neste trabalho, diversos embasamentos teóricos foram fundamentais para a construção do conhecimento sobre a gestão de projetos em ciência de dados. Primeiramente, Escudero e Fortunato (2024); Fernandes e Fortunato (2024), apresentam uma análise contemporânea da relevância da análise de dados no processo de tomada de decisões, destacando sua importância para a eficácia e precisão das escolhas organizacionais. A obra de Das et al. (2015) contribui significativamente para a compreensão das técnicas de correlação e previsão, que são essenciais para a modelagem e interpretação de grandes volumes de dados, proporcionando uma base sólida para a análise preditiva no contexto de ciência de dados.

Saltz e seus co-autores (2015, 2018) discutem metodologias amplamente aplicadas em projetos de ciência de dados, além dos principais desafios enfrentados pelos profissionais da área. Eles ressaltam as dificuldades inerentes à complexidade e à dinâmica desses projetos, sugerindo abordagens para superar obstáculos e alcançar os objetivos desejados. Grander et al. (2022) e Saltz e Shamshurin (2016) enfatizam a importância de adotar uma abordagem estruturada na gestão de projetos, propondo frameworks que garantem a organização e o alinhamento das equipes com os objetivos do projeto, minimizando os riscos e maximizando os resultados.

Rybicki (2019) e Saltz e Crowston (2017) oferecem uma visão detalhada sobre boas práticas na gestão de projetos em ciência de dados, com ênfase em metodologias ágeis como SCRUM e Kanban, que promovem flexibilidade, colaboração contínua e adaptação às mudanças rápidas que caracterizam este campo. Essas metodologias são fundamentais para o sucesso dos projetos, pois garantem uma gestão eficiente dos recursos e tempo.

Por fim, as contribuições de Miguel e Ho (2010) e Saltz e Hotz (2021) foram cruciais para a definição da abordagem metodológica adotada neste trabalho. Suas pesquisas ajudaram a construir uma análise robusta das relações entre práticas de gestão e os benefícios percebidos pelos profissionais de ciência de dados, proporcionando uma compreensão mais profunda sobre como as estratégias de gestão impactam diretamente os resultados e a eficiência dos projetos.

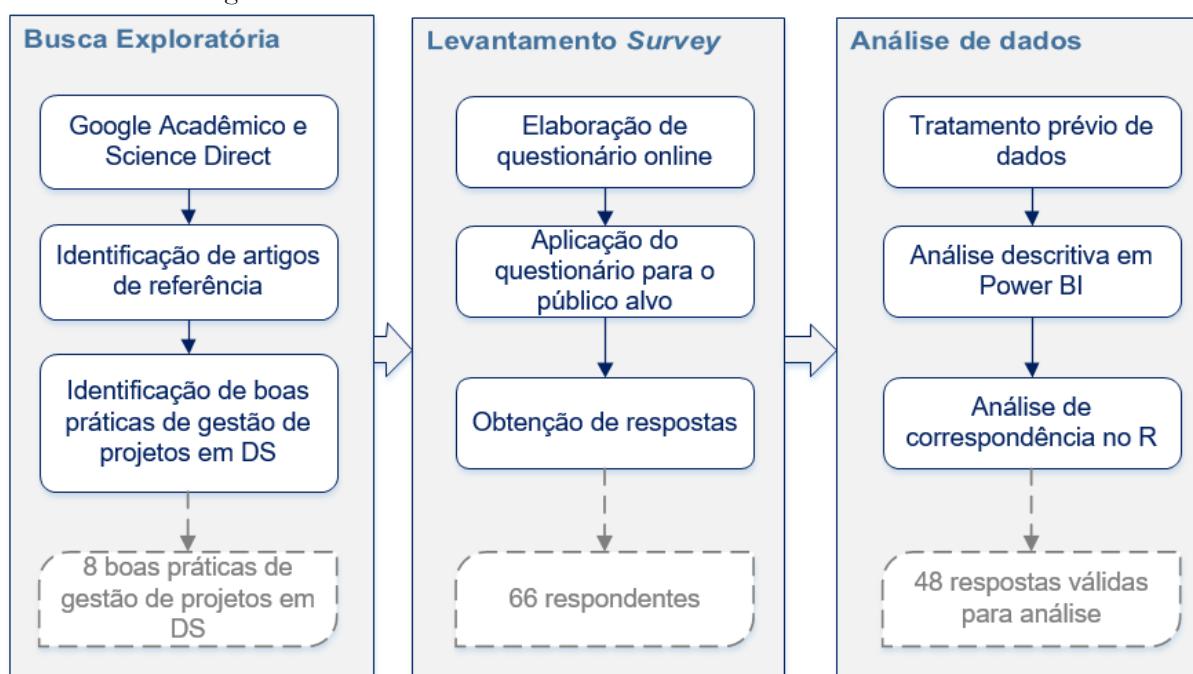
Essa base teórica não só sustenta a análise dos achados da pesquisa, mas também orienta as práticas de gestão de projetos em ciência de dados, promovendo um entendimento mais completo e integrado dos fatores que influenciam o sucesso desses projetos.

Procedimentos metodológicos

Este trabalho de conclusão de curso consiste em uma pesquisa exploratória e descritiva, com o objetivo de identificar as relações entre os benefícios percebidos na prática e as boas práticas de gestão de projetos em ciência de dados, conforme descrito na literatura. Adicionalmente, adota uma abordagem quantitativa (Martins, 2010), utilizando um método estatístico para analisar os dados e revelar associações entre as duas variáveis mencionadas.

A Figura 1 esquematiza os procedimentos metodológicos adotados neste trabalho de conclusão de curso.

Figura 1.
Procedimentos metodológicos



Fonte: Dados originais da pesquisa

A primeira etapa da pesquisa envolveu uma revisão da literatura para identificar artigos fundamentais sobre gestão de projetos em ciência de dados (DS). Foram realizadas buscas em plataformas como Google Acadêmico e Science Direct, utilizando termos relacionados a projetos e DS. Os estudos de Saltz e Shamshurin (2016), Saltz e Crowston (2017), Saltz et al. (2018) e Saltz e Krasteva (2022) se destacaram nesse campo e foram utilizados como referências para a extração das boas práticas de gestão de projetos em DS, uma vez que esses autores são frequentemente citados na área.

Identificaram-se um total de oito boas práticas amplamente empregadas no contexto de DS: Metodologia Ágil, SCRUM, Kanban, Lean Manufacturing, CRISP-DM, KDDM, TDSP e Cookiecutter DS (Rybicki, 2019; Saltz e Crowston, 2017; Saltz et al., 2018; Saltz, 2021; Saltz e Krasteva, 2022; Saltz et al., 2022).

Essas práticas foram utilizadas como base para o desenvolvimento do levantamento do tipo survey, com o objetivo de identificar os principais benefícios percebidos e os desafios na aplicação das ferramentas de gestão em DS nas organizações. O uso do survey é vantajoso, pois permite ao pesquisador avaliar uma amostra de maneira probabilística ou não, extraíndo conclusões a partir dos dados coletados (Miguel e Ho, 2010). Neste caso, foi adotada uma abordagem não probabilística, pois o foco está em identificar tendências de uso de práticas de projetos em DS, sem a necessidade de generalizar para toda a população (Miguel e Ho, 2010).

Para a coleta de dados, foi utilizado um questionário online desenvolvido na ferramenta Forms da Microsoft. O questionário foi elaborado e distribuído de forma não estruturada, por meio de redes sociais e e-mail, direcionado a profissionais que atuam diretamente com DS. O questionário aplicado pode ser encontrado no Apêndice.

Após 30 dias de coleta, foram obtidas 66 respostas. A partir dos dados coletados, realizou-se um tratamento inicial para eliminar respostas duplicadas e excluir participantes que não se enquadram no público-alvo. Os dados foram, então, utilizados para criar um dashboard no Power BI, permitindo uma análise descritiva, incluindo gráficos que segmentam o público respondente em termos demográficos, atividades realizadas e principais práticas, benefícios e desafios em gestão de DS.

Na sequência, aplicou-se a Análise de Correspondência Simples para avaliar a relação entre as principais boas práticas de gestão de projetos e os benefícios percebidos nas empresas. Este método é adequado para identificar associações entre as variáveis analisadas (Greenacre, 2017; Miguel e Ho, 2013).

Os dados foram preparados no Excel, onde foram organizados e tratados para garantir sua qualidade. Após o tratamento, foram registradas 222 observações oriundas das respostas de 48 participantes (amostra não probabilística). Cada observação relacionou um benefício percebido a uma boa prática, permitindo que um único respondente indicasse mais de um benefício para a mesma prática. A tabela resultante continha duas colunas: 'Métodos' e 'Benefícios', e foram excluídas ferramentas que não foram mencionadas.

Finalmente, os dados foram importados para o R-Studio. A seção seguinte apresentará a descrição do algoritmo utilizado em R para realizar a análise de correspondência simples. O arquivo importado é denominado 'Corresp_Novo_Correto'.

Instalação de pacotes e visualização dos dados no R-Studio

```
# Carregar os pacotes necessários
```

```
install.packages("readxl") # Instala o pacote readxl, utilizado para ler arquivos Excel
```

```
library(readxl) # Carrega o pacote readxl  
  
install.packages('ca') # Instala o pacote ca, utilizado para análise de correspondência  
  
library(ca) # Carrega o pacote ca  
  
# Importar os dados do Excel  
  
Corresp_Novo_Correto <- read_excel("caminho do arquivo") # Lê os dados de um arquivo  
  
Excel especificado e armazena na variável Corresp_Novo_Correto  
  
# Gerar a tabela de contingência  
  
tabela_contingencia <-  
  
table(Corresp_Novo_Correto$Métodos, Corresp_Novo_Correto$Desafios) #Cria uma tabela de contingência  
com base nos dados importados, onde as linhas representam os Métodos e as colunas representam os Desafios  
  
print("Tabela de Contingência:") #Imprime uma mensagem indicando o início da tabela de contingência  
  
print(tabela_contingencia) #Imprime a tabela de contingência gerada
```

Teste qui-quadrado, resíduos e mapa perceptual no R-Studio

```
# Realizar o teste Qui-quadrado  
  
qui2 <- chisq.test(tabela_contingencia) #Calcula o teste de Qui-quadrado com base na tabela de contingência  
  
# Exibir o resultado do teste Qui-quadrado  
  
print("Teste Qui-quadrado:") # Imprime uma mensagem indicando o início dos resultados do teste  
  
print(qui2) #Imprime os resultados do teste Qui-quadrado  
  
# Exibir as frequências observadas  
  
print("Frequências Observadas:") #Imprime uma mensagem indicando o início das frequências observadas  
  
print(qui2$observed) #Imprime as frequências observadas  
  
# Exibir as frequências esperadas  
  
print("Frequências Esperadas:") # Imprime uma mensagem indicando o início das frequências esperadas  
  
print(qui2$expected) #Imprime as frequências esperadas calculadas  
  
# Calcular os resíduos padronizados
```

```
print('Resíduos Padronizados:') #Imprime uma mensagem indicando o início dos resíduos padronizados  
print(qui2$residuals) #Imprime os resíduos padronizados calculados  
  
# Calcular os resíduos padronizados ajustados  
  
print('Resíduos Padronizados Ajustados:') #Imprime uma mensagem indicando o início dos resíduos padronizados ajustados  
  
print(qui2$stdres) #Imprime os resíduos padronizados ajustados calculados  
  
# Realizar a análise de correspondência simples  
  
resultado_acs <- ca(tabela_contingencia, graph = TRUE) #Realiza a análise de correspondência simples com base na tabela de contingência  
  
# Exibir o resultado da análise  
  
print(resultado_acs) #Imprime os resultados da análise de correspondência simples  
  
# Plotar o mapa perceptual  
  
plot(resultado_acs) #Plota o mapa perceptual gerado a partir da análise de correspondência simples
```

Apresentação e análise dos resultados

Análise Descritiva e Discussão dos Resultados

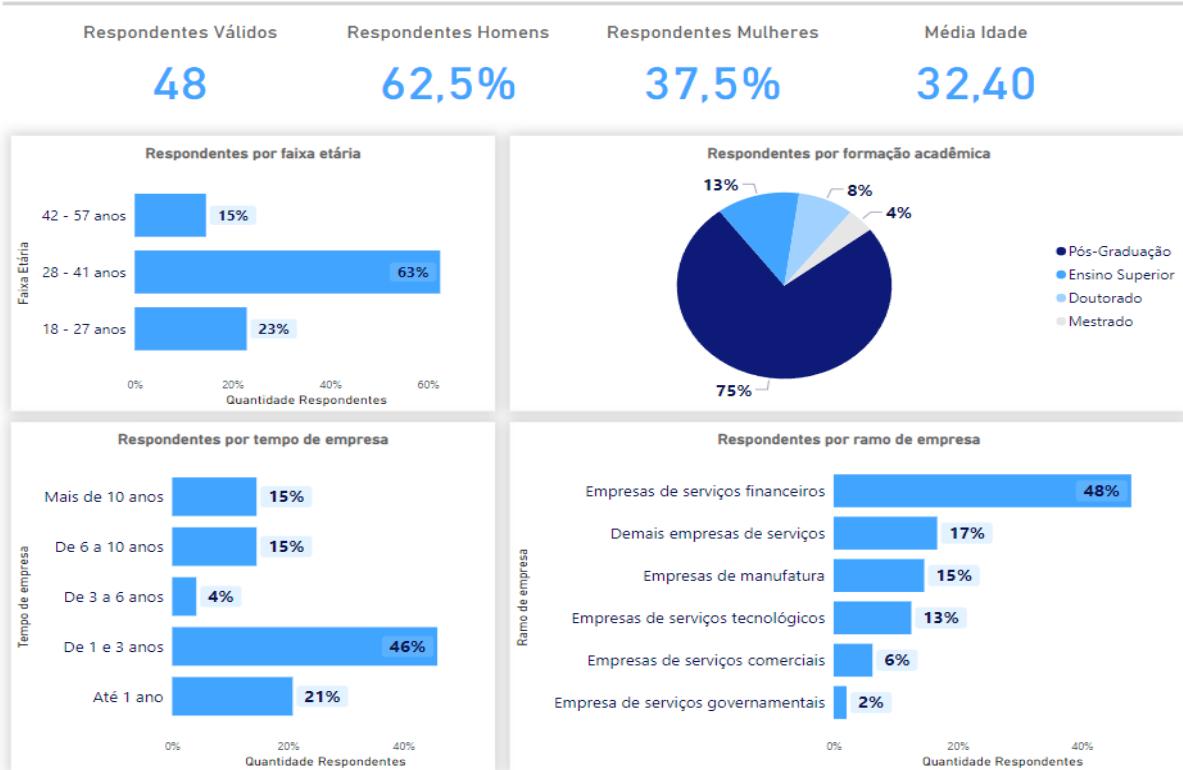
Nesta seção, são apresentados os resultados obtidos a partir do levantamento de tipo survey, conforme os procedimentos metodológicos descritos anteriormente. Após a coleta dos dados, foi desenvolvido um dashboard que organizou as informações em três categorias principais: perfil demográfico, perfil de atividades e boas práticas de gestão de projetos em ciência de dados (DS).

Perfil Demográfico

A primeira parte do dashboard, ilustrada na Figura 2, exibe o perfil demográfico dos participantes do survey. Foram validadas 48 respostas, e, entre os respondentes, 62% são do sexo masculino e 37% do sexo feminino, com uma idade média de 32 anos. Este dado reforça, conforme apontado por Smith et al. (2020), que a representatividade de gênero na área de ciência de dados continua a ser um desafio, com a sub-representação das mulheres nesse campo.

Figura 2.
Parte I ‘dashboard’ (perfil demográfico)

Perfil Demográfico



Fonte: Dados originais da pesquisa.

Em relação à formação acadêmica, 75% dos participantes têm pelo menos um diploma de graduação e pós-graduação stricto sensu. Quanto à área de atuação, 48% dos participantes estão no setor de serviços financeiros, 16% em serviços gerais e 12% em manufatura, evidenciando a presença de cientistas de dados em setores críticos para o processamento e análise de grandes volumes de dados.

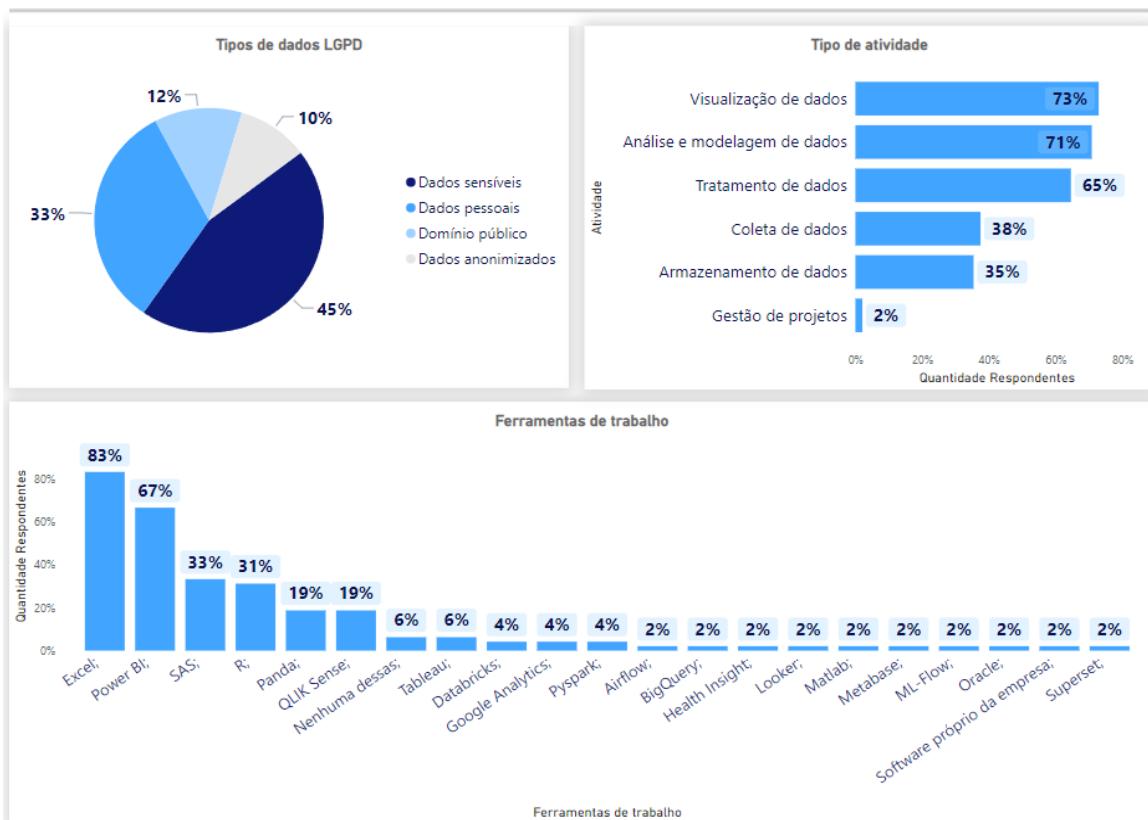
Além disso, 66% dos participantes possuem, pelo menos, três anos de experiência nas suas empresas, o que reflete uma sólida formação técnica e acadêmica, já que 54% desse grupo também possui pós-graduação.

Perfil de Atividades

A segunda parte do dashboard, mostrada na Figura 3, descreve o perfil de atividades dos profissionais participantes. Aqui, destaca-se que 44% dos participantes lidam diretamente com dados sensíveis, enquanto 32% trabalham com dados pessoais e 22% com dados públicos ou anonimizados. Isso reflete a diversidade de tipos de dados com os quais os cientistas de dados interagem, conforme a definição de ciência de dados de Saltz e Stanton (2017), que inclui coleta, preparação, análise, visualização e gerenciamento de dados.

Figura 3.
Parte II ‘dashboard’ (perfil de atividade)

Perfil de atividades



Fonte: Dados originais da pesquisa.

Além disso, 73% dos respondentes trabalham com visualização de dados, 70% com análise e modelagem de dados, e 64% com tratamento de dados. Essas atividades são essenciais para o ciclo de vida da ciência de dados, onde a visualização desempenha um papel crucial na comunicação eficaz dos resultados (Saltz e Stanton, 2017).

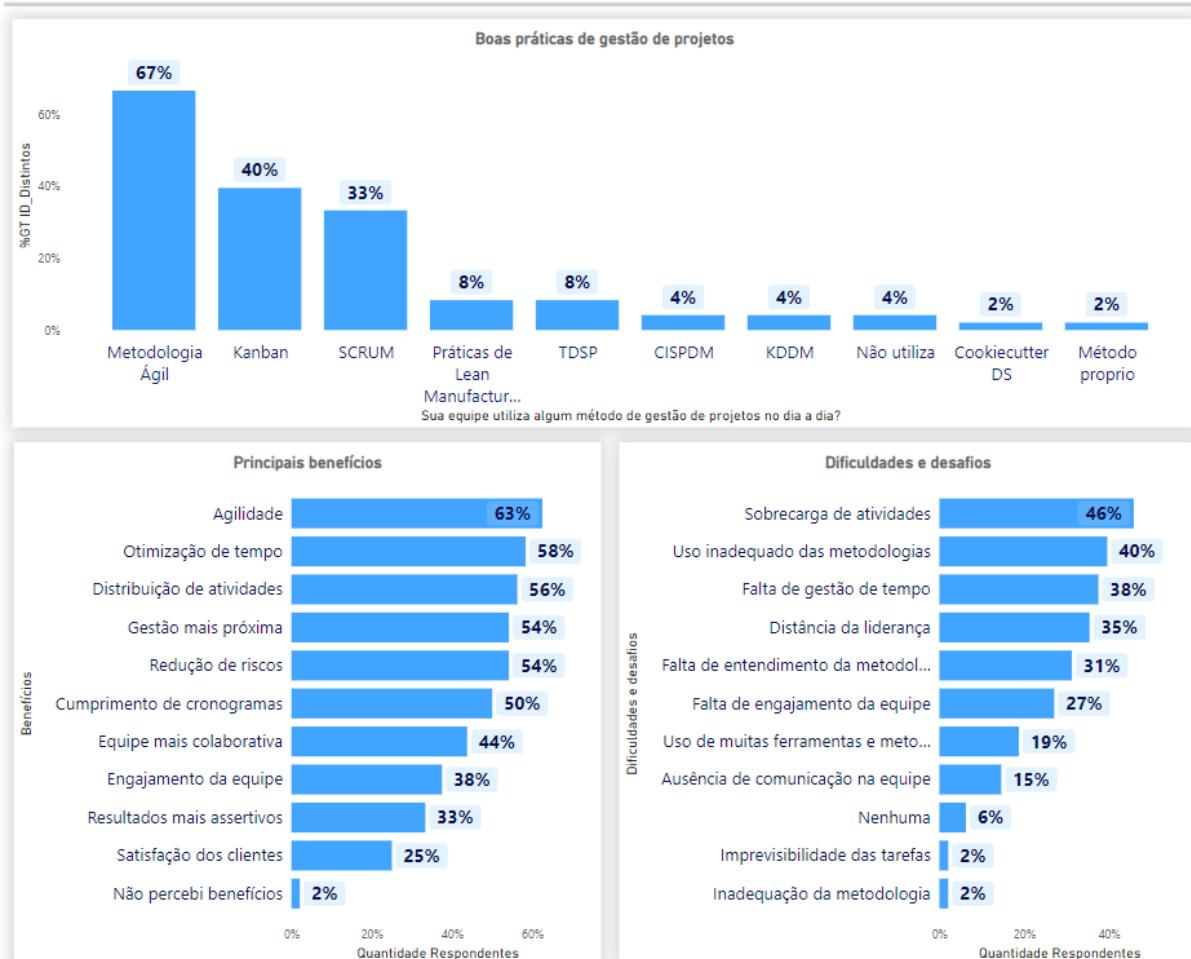
Em relação às ferramentas utilizadas, 83% dos participantes utilizam o Excel, 66% o Power BI e 33% o SAS. Essas ferramentas são vitais para diferentes etapas do processo de ciência de dados: Excel para o tratamento inicial de dados, Power BI para visualização e SAS para análise/modelagem. Ferramentas como R (31%) e Pandas (18%) também são usadas para tratamento e modelagem de dados.

Boas Práticas de Gestão de Projetos em Ciência de Dados

A terceira parte do dashboard, apresentada na **Figura 4**, destaca as boas práticas de gestão de projetos de DS, além dos principais benefícios e desafios observados pelos profissionais da área. A Metodologia Ágil se destaca como a prática mais adotada (66%), seguida pelo Kanban e Lean Manufacturing, que são utilizados por 47% dos participantes. Essas práticas refletem a tendência de utilização de abordagens flexíveis e eficientes, amplamente discutidas na literatura (Saltz e Crowston, 2017; Saltz et al., 2018).

Figura 4.
Parte III ‘dashboard’ (boas práticas, benefícios e desafios)

Métodos, benefícios e desafios



Fonte: Dados originais da pesquisa.

A metodologia SCRUM, embora amplamente reconhecida (Smith et al., 2020), foi adotada por 33% dos respondentes, enquanto outras metodologias como TDSP, CRISP-DM e KDD apresentaram taxas de adoção mais baixas (18%).

Quanto aos benefícios percebidos, os participantes destacaram principalmente a agilidade, a otimização de tempo e a melhor distribuição de tarefas como os principais resultados das práticas de gestão adotadas, com 62%, 58% e 56% de concordância, respectivamente. Isso está em consonância com a literatura que sugere que práticas de gestão ágeis ajudam a reduzir o tempo de resposta a mudanças e otimizar processos (Saltz e Crowston, 2017; Saltz et al., 2022).

Desafios Identificados

A pesquisa também revelou desafios significativos enfrentados pelos profissionais da área. Aproximadamente 46% dos participantes mencionaram a sobrecarga de atividades como um dos principais obstáculos para a adoção de metodologias de gestão. Outros desafios importantes incluem o uso inadequado de ferramentas (40%) e a falta de uma boa gestão de tempo (38%).

Análise de Correspondência

Com base nos benefícios reconhecidos para as práticas de gestão de projetos de DS, a análise de correspondência foi realizada utilizando o software R. As tabelas 1 e 2 mostram as frequências observadas e esperadas no levantamento, e os resultados do teste Qui-quadrado (Tabela 3) indicaram uma relação significativa entre as variáveis, com um valor de Qui-quadrado de 74,26, significando que há uma associação relevante entre as boas práticas de gestão e os benefícios percebidos, já que o p-valor foi inferior a 0,05.

Tabela 1.
Frequências observadas

Práticas em DS	AG	CC	DA	EC	EE	GP	OT	RA	RR	SB	SC
KB	4	6	7	4	3	5	4	1	3	0	1
LM	2	1	1	1	1	1	1	1	1	0	1
MA	16	13	15	14	11	15	16	9	17	0	8
MP	1	0	0	0	0	1	0	0	1	0	0
SCRUM	4	4	4	1	3	3	5	4	3	0	1
TDSP	1	0	0	0	0	0	1	0	1	1	0

Fonte: Dados originais da pesquisa.

Nota1: Kanban [KB]; Lean Manufacturing [LM]; Metodologia Ágil [MA]; Método Próprio [MP]; Team Data Science Process [TDSP]

Tabela 2.
Frequências esperadas

Práticas em DS	AG	CC	DA	EC	EE	GP	OT	RA	RR	SB	SC
KB	4,7	4,1	4,6	3,4	3,0	4,2	4,6	2,5	4,4	0,1	1,8
LM	1,3	1,1	1,3	0,9	0,8	1,2	1,3	0,7	1,2	0,0	0,5
MA	16,9	14,4	16,3	12,0	10,8	15,0	16,3	9,0	15,6	0,6	6,6
MP	0,3	0,3	0,3	0,2	0,2	0,3	0,3	0,2	0,3	0,0	0,1
SCRUM	4,0	3,4	3,8	2,8	2,5	3,6	3,8	2,1	3,7	0,1	1,5
TDSP	0,5	0,4	0,4	0,3	0,3	0,4	0,4	0,2	0,4	0,0	0,2

Fonte: Dados originais da pesquisa.

Nota1: Kanban [KB]; Lean Manufacturing [LM]; Metodologia Ágil [MA]; Método Próprio [MP]; Team Data Science Process [TDSP]

Tabela 3.

Resultado do Teste Qui-Quadrado

Teste Qui-Quadrado	Resultados
Qui-Valor	74,2
Grau de liberdade	50,0
p-valor	0,01

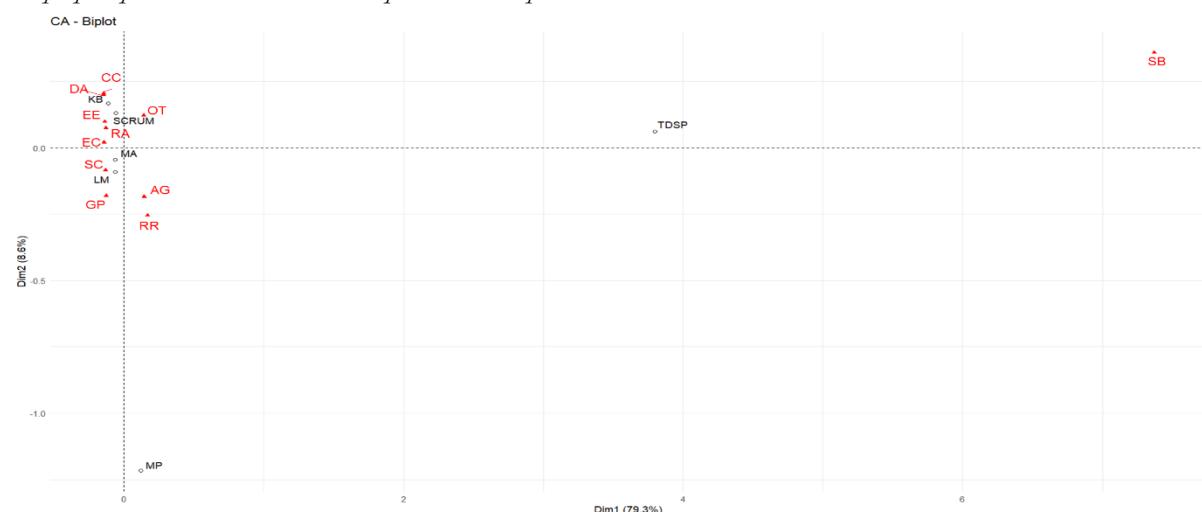
Fonte: Dados originais da pesquisa.

Considerações finais

A partir dessa premissa, um mapa perceptual foi gerado, permitindo observar as associações entre as práticas de gestão e os benefícios percebidos. O Mapa Perceptual (Figura 5 e Figura 6) ilustra as relações encontradas.

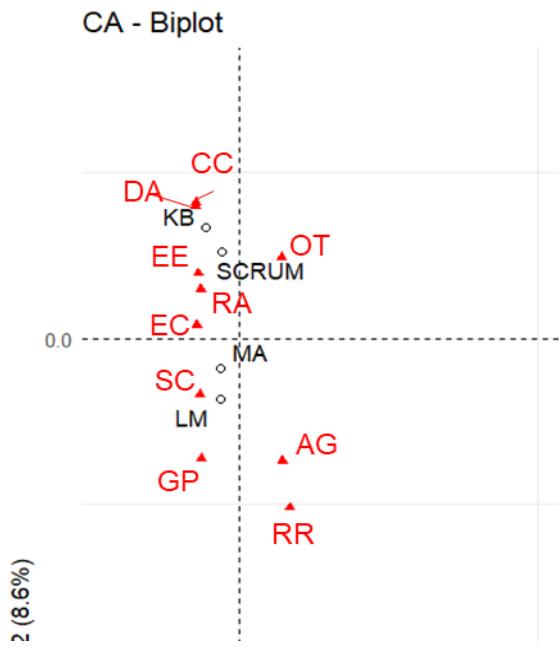
Figura 5.

Mapa perceptual da Análise de Correspondência Simples



Fonte: Dados originais da pesquisa.

Figura 6.
Recorte do mapa perceptual da análise de correspondência simples



Fonte: Dados originais da pesquisa.

Discussão

A análise de correspondência revelou associações significativas entre algumas práticas de gestão de projetos em ciência de dados (DS) e os benefícios percebidos pelos profissionais. Dentre as práticas mais adotadas, o Kanban demonstrou uma associação clara com os benefícios de cumprimento de prazos e melhor distribuição de atividades, corroborando com a literatura que destaca o Kanban como uma metodologia eficiente para a priorização e organização de tarefas dentro de projetos de DS (Saltz e Crowston, 2017). A associação entre o Kanban e o benefício da distribuição de atividades reflete sua aplicação na visualização e gerenciamento contínuo de fluxos de trabalho, o que facilita a gestão de tarefas e promove uma maior produtividade e agilidade nas entregas.

A Metodologia Ágil foi identificada como fortemente ligada à satisfação do cliente, o que reforça a ideia de que a flexibilidade e a capacidade de adaptação das metodologias ágeis são componentes chave para a entrega de valor ao cliente de forma rápida e eficiente (Schwaber e Sutherland, 2020). A associação entre Ágil e a satisfação do cliente também pode ser vista como um reflexo do foco da metodologia na entrega incremental e na comunicação contínua com os stakeholders, permitindo ajustes durante o processo de desenvolvimento do projeto, o que aumenta a percepção de valor entregue ao cliente final.

O SCRUM, uma das abordagens ágeis mais populares, foi associado a engajamento da equipe e resultados mais assertivos, seguindo a literatura que destaca o papel central da colaboração e da comunicação eficiente entre membros da equipe como parte fundamental da sua estrutura. O foco do SCRUM em ciclos iterativos de trabalho, chamadas de sprints, permite que as equipes de DS se ajustem rapidamente às mudanças nos requisitos

ou nos resultados das análises, o que contribui para maior assertividade nos resultados e aumenta o engajamento dos profissionais ao longo do projeto.

O Lean Manufacturing foi vinculado à satisfação do cliente, um benefício que está diretamente relacionado à sua filosofia de minimizar desperdícios e otimizar processos, proporcionando entregas mais rápidas e com melhor qualidade. Isso se alinha com o objetivo do Lean de aumentar a eficiência, reduzir custos e melhorar os resultados de um projeto. Essa associação sugere que, no contexto de DS, a aplicação do Lean pode contribuir para uma gestão mais eficaz dos recursos e prazos, levando a uma maior satisfação dos clientes que esperam entregas rápidas e eficientes.

Por outro lado, algumas práticas não demonstraram as associações esperadas. TDSP e Método Próprio, por exemplo, não apresentaram uma conexão clara com os benefícios perceptíveis pelos participantes do estudo. Isso pode indicar que, embora essas metodologias sejam adotadas em algumas organizações, sua implementação ou resultados ainda não foram suficientemente avaliados ou documentados para gerar uma associação clara. A ausência de associações significativas pode também refletir uma falta de maturidade na aplicação dessas metodologias, o que torna difícil observar os benefícios que elas poderiam trazer para os projetos de DS.

Considerações finais

Este trabalho que teve como objetivo aplicar a Análise de Correspondência Simples para investigar a relação entre as principais boas práticas de gestão de projetos em Data Science (DS) e os benefícios percebidos em seu uso pelas empresas. A partir dessa análise, foi possível identificar algumas associações relevantes, o que contribui para a compreensão das práticas de gestão mais eficazes no contexto de projetos de DS. Do ponto de vista metodológico, a aplicação da Análise de Correspondência Simples se mostrou uma ferramenta válida e eficiente para identificar padrões entre variáveis categóricas, oferecendo uma nova perspectiva para estudos na área de gestão de projetos.

Embora este estudo tenha proporcionado contribuições significativas, também apresenta limitações que devem ser consideradas na interpretação dos resultados. O caráter exploratório do levantamento de dados e a amostra não probabilística constituem restrições que comprometem a generalização dos achados para toda a população de profissionais de ciência de dados. Além disso, embora a análise de correspondência e os mapas perceptuais sejam úteis para visualizar as associações, as relações entre as variáveis podem ser mais complexas do que o que foi identificado neste estudo. A ausência de associações significativas em algumas práticas de gestão sugere que existem fatores adicionais que precisam ser considerados em análises futuras.

Com base nos resultados obtidos, é possível destacar a necessidade de pesquisas adicionais que explorem mais profundamente como diferentes metodologias de gestão são implementadas em projetos de DS e quais são seus impactos reais nos resultados alcançados. Estudos que investiguem comparativamente práticas ágeis, como SCRUM e Kanban, com metodologias mais estruturadas, como CRISP-DM ou KDD, podem trazer insights

valiosos sobre as sinergias e limitações dessas abordagens no contexto de DS. Além disso, a realização de pesquisas com amostras mais amplas e representativas permitirá testar a robustez das conclusões e verificar a aplicabilidade das práticas de gestão em diferentes ambientes organizacionais e culturais.

Ademais, a incorporação de métodos qualitativos, como entrevistas e estudos de caso, pode fornecer uma compreensão mais profunda dos desafios e benefícios na implementação das práticas de gestão em projetos de DS, permitindo identificar as condições que favorecem o sucesso dessas metodologias. Essas investigações ajudarão a preencher as lacunas existentes, fornecendo uma base sólida para a formulação de recomendações práticas e eficazes para profissionais da área.

Em síntese, este estudo contribui para o entendimento das práticas de gestão de projetos em Data Science e dos benefícios percebidos com sua adoção, mas também revela a complexidade desse campo emergente. A contínua exploração de novas metodologias, aliada à validação e aprofundamento dos achados, é essencial para aprimorar as práticas de gestão e garantir resultados mais eficientes e impactantes nos projetos de ciência de dados. A crescente importância da DS nas organizações exige que pesquisadores e profissionais se unam na busca por soluções cada vez mais eficazes para enfrentar os desafios do futuro.

Referências

- DAS, M., Cui, R., Campbell, D. R., & Agrawal, G. (2015). Towards methods for systematic research on big data. In *International Conference on Big Data*, 2015. Anais (pp. 2072-2081).
- David, A., Corbett, A. M., Dutra, J. L., & Straw, A. G. (2021). Eight practices for data management to enable team data science. *Journal of Clinical and Translational Science*, 5, e4.
- Escudero, T. A., & Fortunato, M. H. T. (2024). Construção de “Key Risk Indicators” para mitigação de riscos operacionais no setor de pagamentos. *Cadernos de Gestão e Empreendedorismo*, 12(1), 34-48.
- Fernandes, G., & Fortunato, M. H. T. (2024). Análise de correspondência múltipla: Avaliando a influência de BI e dataviz em consultoria de sistemas. *Gestão Contemporânea*, 14(2), 25-40.
- Grander, G., Silva, L. F., Gonzalez, E. D. R. S., & Penha, R. (2022). Framework for structuring big data projects. *Electronics*, 11, 3540.
- Greenacre, M. (2017). *Correspondence analysis in practice* (3rd ed.). Nova Iorque: Chapman and Hall CRC.
- Martins, R. (2010). Abordagens quantitativa e qualitativa. In P. A. C. Miguel (Ed.), *Metodologia de pesquisa em engenharia de produção e gestão de operações* (pp. 47-63). Rio de Janeiro: Elsevier.
- Rybicki, J. (2019). Best practices in structuring data science projects. In *Information Systems Architecture and Technology: Proceedings of the 39th International Conference on Information Systems Architecture and Technology* (pp. 348-357).

- Saltz, J., Hotz, N., Wild, D., & Stirling, K. (2018). Exploring project management methodologies used within data science teams. In *America Conference on Information Systems*.
- Saltz, J. S. (2021). CRISP-DM for data science: strengths, weaknesses, and potential next steps. In *IEEE International Conference on Big Data* (pp. 2337-2344).
- Saltz, J. S. (2015). The need for new processes, methodologies, and tools to support big data teams and improve big data project effectiveness. In *IEEE International Conference on Big Data* (pp. 2066-2071).
- Saltz, J. S., & Krasteva, I. (2022). Current approaches for executing big data science projects: A systematic literature review. *Journal of Computer Science*, 8, e862.
- Saltz, J. S., & Shamshurin, I. (2016). Big data team process methodologies: A literature review and the identification of key factors for a project's success. In *International Conference on Big Data* (pp. 2872-2879).
- Saltz, J., & Crowston, K. (2017). Comparing data science project management methodologies via a controlled experiment. In *50th Hawaii International Conference on System Sciences* (pp. 1013-1022).
- Saltz, J., Hotz, N., & Sutherland, J. (2022). Achieving lean data science agility via data-driven SCRUM. In *55th Hawaii International Conference on System Sciences* (pp. 7287-7296).
- Saltz, J., Shamshurin, I., & Connors, C. (2016). A framework for describing big data projects. In *Business Information Systems Workshops: International Workshops* (pp. 183-195).
- Saltz, J., & Stanton, J. (2017). *An introduction to data science*. Los Angeles: SAGE.
- Sutherland, J., & Schwaber, K. (2020). *The Scrum guides: The definitive guide to scrum: The rules of the game*. Retrieved from <https://scrumguides.org/docs/scrumguide/v2020/2020-Scrum-Guide-US.pdf>
- .