

Análise de Dados Prática

Texto Preliminar

Geraldo Xexéo

22 de julho de 2021 00:09

DRAFT

Copyright © 2020 Geraldo Xexéo.

Todos os direitos reservados.

Esta é um pré-publicação com o texto parcial em trabalho.

Contato com o autor pelo e-mail xexeo@ufrj.br.

DRAFT

Sumário

Lista de Figuras	v
Lista de Tabelas	vii
Lista de Programas	ix
I. Conceitos Iniciais	1
1. Introdução a Análise de Dados	3
1.1. O Ciclo de Vida dos Dados	6
1.2. A Prática da Análise de Dados	6
1.3. Tipos de Análise de Dados	7
1.4. Cenários de Análise de Dados	9
1.4.1. Temperatura dos Dados	9
1.5. Processos de Análise de Dados	10
1.6. Conclusão	11
1.7. Conceituação Genérica de Valor	12
1.8. Valor na Economia	13
1.8.1. Utilidade	14
1.8.2. Microeconomia Moderna e o Custo de Oportunidade	14
1.9. Valor em Marketing	15
1.9.1. Os Elementos do Valor - B2C	16
1.9.2. Os Elementos do Valor - B2B	17
1.9.3. Outras Técnicas	20
1.10. Priorização Baseada em Valor com o Cliente	21
1.10.1. Análise de Kano	21
1.11. Valor Financeiro de Projetos	22
1.11.1. Custo Total de Propriedade	24

1.11.2.	Retorno do Investimento	26
1.11.3.	Valor Presente Líquido	26
1.11.4.	Taxa Interna de Retorno	27
1.12.	Valor em Software	27
1.12.1.	Triângulo do Valor	28
1.12.2.	Valor em Métodos Ágeis	29
1.13.	Conclusão	29
1.14.	Exercícios	30
2.	Dados e Informação	31
2.1.	Usando Significados Diferentes	32
2.2.	Conclusão	34
3.	Tipos de Dados Convencionais	35
3.1.	Introdução às medidas	35
3.2.	Dados Nominais	36
3.3.	Dados Ordinais	41
3.3.1.	Escala de Likert	41
3.3.2.	Escalas Ordinais e Números	41
3.4.	IDEIA	43
II.	Introdução as Ferramentas	45
III.	Exploração Inicial dos Dados	47
IV.	Mineração de Dados	49
V.	Dados Textuais	51
VI.	Dados Georeferenciados	53
VII.	Grafos	55
VIII.	Entendendo o Negócio	57
4.	Modelagem de Processos de Negócio	59
4.1.	Processos de Negócio	60

4.2.	Exemplos de Processos	61
4.3.	Representação de Processos de Negócio	62
4.4.	Por que Entender os Processos de Negócio	62
4.5.	Exercícios	63
5.	BPMN	65
5.1.	Usos de BPMN	66
5.2.	Participantes	67
5.3.	Símbolos básicos	67
5.4.	Atividades	67
5.5.	Eventos	70
5.5.1.	Eventos Iniciais	71
5.5.2.	Eventos Finais	71
5.5.3.	Eventos Intermediários	72
5.6.	<i>Gateways</i>	73
5.7.	Fluxos	75
5.8.	Alguns Exemplos	76
5.9.	Eventos Anexados a Atividades	76
5.10.	Exercícios	76

DRAFT

Lista de Figuras

1.1.	Modelo Conceitual da Análise de Dados	5
1.2.	Ciclo de Vida dos Dados proposto por Wing (2019).	6
1.3.	Tipos de Análise de Dados	8
1.4.	A metodologia CRISP-DM (adaptado de (Chapman et al., 2000)) . .	11
1.5.	Elementos do Valor B2C. Fonte:Almqvist, Senior e Bloch, 2016	17
1.6.	Hierarquia das Necessidades de Maslow. Fonte: Wikipedia Commons por Felipe Sanchez (CC-BY-SA 3.0) e J. Finkelstein (GFDL)	18
1.7.	Elementos do Valor B2B. Fonte: Almqvist, Cleghorn e Sherer, 2018 .	19
1.8.	Cálculo do valor futuro de um investimento de R\$10.000,00 por mês com uma taxa de juros de 1%.	24
1.9.	Gráfico do TCO de impressoras em função de da quantidade de páginas impressas. Percebe-se que após aproximadamente 3500 páginas é melhor comprar uma impressora a laser mais cara, caso esse fosse o único fator de análise.	25
1.10.	O triângulo do valor para a maioria dos projetos.	29
2.1.	Pirâmide da hierarquia do DIKW	33
3.1.	Histograma da distribuição do estado civil das pessoas em uma amostra.	38
3.2.	Histograma do estado civil de pessoas em uma amostra, ordenado pela frequência.	38
3.3.	Gráfico de pizza de percentuais do estado civil de pessoas em uma amostra.	39
3.4.	Gráfico (histograma) um pouco mais sofisticado, mostrando várias informações que são baseadas na contagem de dados nominais simul- taneamente.	39

Lista de Figuras

3.5.	Gráfico construído com o software Tableau mostrando a quantidade e proporção de homens e mulheres de acordo com cor dos cabelos e dos olhos.	40
3.6.	Gráfico feito com o software Tableau, mostrando as proporções em uma amostra da cor do cabelo, como texto, e da cor dos olhos, como cor, permitindo comparar a distribuição entre homens e mulheres. . .	40
3.7.	Questionário com escala de Likert, produzido com o Google Forms. .	42
5.1.	Um exemplo de processo descrito em BPMN. Fonte: imagem do autor	66
5.2.	Exemplo de uso de um evento paralelo em BPMN	75
5.3.	Exemplo do atendimento da ordem de serviço em BPMN	76

DRAFT

Lista de Tabelas

1.1. Características típicas para as temperaturas dos dados	9
1.2. Resolvendo a classe de Kano por meio do cruzamento de duas perguntas (Moorman, 2012)	22
1.3. Valores usados para calcular o TCO de uma impressora.	25
1.4. Tabela de cálculo do Valor Presente Líquido, considerando saldos diferentes em cada período. Foi usada uma taxa de 5% a cada período. A fórmula para cada Valor Presente é $Caixa_i / (1 + Taxa)^{Período}$	27
3.1. Contagem e distribuição do estado civil de pessoas em uma amostra. . .	37
3.2. Resumo das características de cada tipo de dado.	42
5.1. Símbolos para os eventos que podem aparecer no início de um processo. .	69
5.2. Símbolos para os eventos que podem aparecer no início de um processo. .	71
5.3. Símbolos para os eventos que podem aparecer no fim de um fluxo de processo.	72
5.4. Eventos que podem aparecer no meio de um processo.	73
5.5. Símbolos para os gateway.	74

DRAFT

Lista de Programas

DRAFT

DRAFT

Parte I.

Conceitos Iniciais

DRAFT

DRAFT

Introdução a Análise de Dados

The value of an idea lies in the
using of it

(Thomas Edison)

Conteúdo

1.1.	O Ciclo de Vida dos Dados	6
1.2.	A Prática da Análise de Dados	6
1.3.	Tipos de Análise de Dados	7
1.4.	Cenários de Análise de Dados	9
1.5.	Processos de Análise de Dados	10
1.6.	Conclusão	11
1.7.	Conceituação Genérica de Valor	12
1.8.	Valor na Economia	13
1.9.	Valor em Marketing	15
1.10.	Priorização Baseada em Valor com o Cliente	21
1.11.	Valor Financeiro de Projetos	22
1.12.	Valor em Software	27
1.13.	Conclusão	29

O tópico deste livro é a Análise de Dados, minha tradução para *Data Analytics*¹. Este nome está muito em voga quando escrevo este texto, o que chamamos de *hype*, seguindo uma linha na história recente da computação aplicada que inclui outros termos, ou ainda ainda na moda, ou que começaram também como modismos e acabaram se estabelecendo. No Brasil, é comum usar os termos no original em inglês:

- Data Analytics, ou Análise de Dados;
- Data Science, ou Ciência de Dados;
- Data Mining, ou Mineração de Dados;
- Machine Learning, ou Aprendizado de Máquina;
- Business Intelligence, ou Inteligência de Negócios;
- Big Data;
- Text Analytics, e
- Data Warehousing.

Todos esses termos são na verdade *buzzwords* que giram em torno de um problema comum: obter valor para uma organização, empreendimento ou pesquisa a partir de dados, sejam eles em pequena ou grande quantidade, estruturados ou não.

Em uma tentativa de explicar a relação entre esses termos, apresento o modelo conceitual da figura 1.1. Nesse modelo mostramos que **dados** são guardados em **Bancos de Dados**, ou melhor, em **Sistemas Gerenciadores de Bancos de Dados**. Eles também podem ser guardados diretamente em **Data Warehouses**, ou o Data Warehouse pode usar um SGDB para sua implementação. Ambos, SGDB e Data Warehouse estão sujeitos a um Modelo de Dados que conceitualiza os dados, que representam **Informação**. Os dados podem se comportar como **Big Data**. A **Análise de Dados** se baseia na **Ciência de Dados**, usando de forma específica a **Mineração de Dados**, para analisar os dados, criando **visualizações** que permitem descobrir **Conhecimento**, e também permitindo realizar a **Inteligência de Negócios**, que também alcança o Conhecimento. Mais tarde, vamos tratar da questão de formas não estruturadas de dados, como texto e grafos.

Em especial, neste texto, não vou diferenciar Análise de Dados e Ciência de Dados, considerando que a primeira é a execução da segunda. Também vou considerar que *Data Mining*, a Mineração de Dados, é usada, na maior parte das vezes, em tarefas de Análise de Dados. Outros autores podem discordar dessa visão, relegando a Análise de Dados um trabalho menos nobre que o do Cientista de Dados, porém não é assim que vejo. Acredito que o Analista de Dados e o Cientista de Dados acabam se fundindo quando analisado o objetivo real de seu trabalho.

Doe (1666) fornece uma definição interessante para **Análise de Dados**: “O conjunto

¹Alguns autores separam, artificialmente, os termos em inglês *data analysis* e *data analytics*, afirmando que o primeiro trata dos dados do passado e o segundo das ações futuras. Como podemos ver, as definições mais modernas de *data analytics* cobrem passado e futuro.

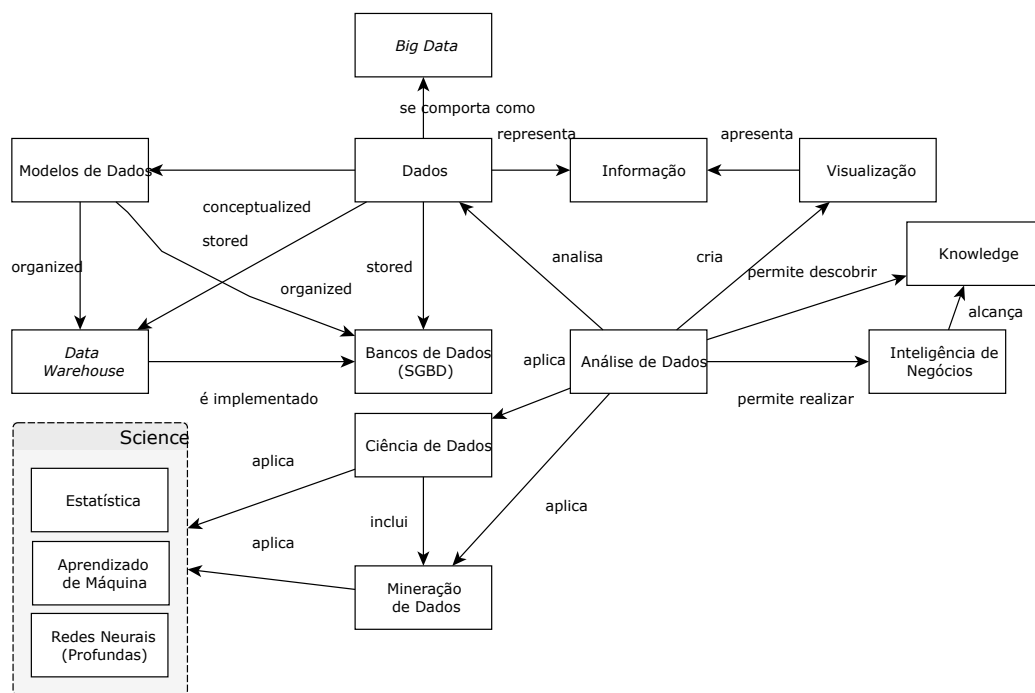


Figura 1.1.: Modelo Conceitual da Análise de Dados (do autor)

de abordagens quantitativas e qualitativas para buscar valor a partir de dados”.

Atualmente é comum que as empresas estejam afogadas em dados, isto é, elas possuem, ou podem adquirir facilmente, uma quantidade enorme de dados, porém não têm capacidade de usá-los, ou de entender como eles podem ser úteis. Imagine um supermercado, por exemplo, que pode atualmente associar todas as suas compras ao seu CPF. No Rio de Janeiro, muitos fazem isso, eu especialmente estou cadastrado em dois. Porém, o quanto eles usam essa informação? Quantas ações são feitas, baseadas no seu perfil de compras, específicas para você? Por causa da pandemia eu compro frequentemente no mesmo supermercado on-line, e ele é incapaz de sugerir meus itens comuns de compra, ou outros itens de compra, como oferecer feijão se eu compro arroz. Mesmo as promoções ditas personalizadas não fazem muito sentido: sempre recebo sugestão de comprar fraldas infantis, coisa que não faço há 15 anos! Ao contrário, empresas como a Amazon dedicam grande parte do seu esforço de marketing a tentar prever que produto pode ser interessante para seus clientes, frequentemente com sucesso.

A **Análise de Dados** pode, então, também ser definida como “a prática de criar soluções para o negócio a partir do estudo e manipulação dos dados, tanto específicas para uma questão imediata, que geram relatórios *ad-hoc*, como soluções que serão aplicadas de forma recorrente e precisarão ser implantadas em um sistema de produção”.

Momentaneamente, vou passar por cima da definição de valor. O próximo capítulo é dedicado a explicar algumas formas como valor pode ser entendido.



Figura 1.2.: Ciclo de Vida dos Dados proposto por Wing (2019).

1.1. O Ciclo de Vida dos Dados

A Análise de Dados é um dos passos do **Ciclo de Vida dos Dados**. Muitos modelos de Ciclo de Vida foram propostos para dados, sendo que um relatório do *Committee on Earth Observation Satellites* cita 55 modelos distintos CEOS WGISS, 2012.

Uma abordagem interessante, descrita rapidamente por CEOS WGISS (2012) a partir de uma apresentação de Govoni e Gunther (2006), descreve dois ciclos de vida: um para o produtor de dados e outro para o consumidor de dados. O primeiro é composto de Coleta, Análise, Publicação e Preservação, já o segundo é composto de Achar, Pegar, Entender e Usar.

Wing (2019) também tem uma abordagem interessante para o Ciclo de Vida dos Dados, que passa pelos passos geração, coleta, processamento, armazenagem, gestão, análise, visualização, interpretação e termina em humanos, como apresentado na figura 1.2.

Analizando vários ciclos, algumas preocupações importante parecem vir a tona apenas em alguns. Por exemplo, a capacidade de geração de dados pode ser muito maior que a capacidade de coletar (Wing, 2019), exigindo uma filtragem rápida inicial ou mesmo a aceitação de que somos incapazes de capturar todos os dados gerados, implicando em uma taxa de amostragem. Isso pode ser rapidamente associado a observações de objetos em movimento, por exemplo, onde temos que escolher essa taxa de amostragem, ou na captura de dados em experimentos de Física de Altas Energias, onde, devido a grande quantidade de dados gerados nos detectores, é necessário um filtro (*trigger*) online que seleciona apenas os eventos mais interessantes.

Outra preocupação interessante é a diferença dos ciclos de vida que exigem a preservação como passo final, para outros que falam da eliminação dos dados. No momento, por exemplo, a Lei Geral de Proteção de Dados (Brasil, 2019) no Brasil, e leis equivalentes no mundo, exigem que certos dados sejam apagados após um prazo específico.

Outros ciclos propostos também falam de: planejamento, curadoria, administração, gestão, distribuição, preservação em diferentes intervalos de tempo, migração, descoberta, experimentação etc (CEOS WGISS, 2012).

1.2. A Prática da Análise de Dados

A Análise de Dados é caracterizada por várias práticas:

- uso de software;
- criação de software, na forma de programas, scripts ou *workflows* em ferramentas especializadas;
- orientação ao negócio;
- orientação aos dados;
- orientação a hipóteses;
- ser baseada em iterações.

Dependendo do método, do contexto, dos profissionais envolvidos ou mesmo do autor de um livro ou artigo, o trabalho de análise de dados pode começar tanto com os dados, quanto com os negócios, mas certamente é orientado por ambos. Na verdade, começar por um ou por outro depende da oportunidade. Algumas vezes temos oportunidades de negócio, mas outras vezes temos oportunidades de dados. Certamente apenas como entendimento de ambos uma proposta de valor para o negócio será encontrada.

Os ciclos e iterações da Análise de Dados decorrem do fato que o entendimento dos dados é gradativo. Quanto mais os entendemos mais somos capazes de fazer hipóteses sobre o que é possível ou não. Devido a grande quantidade de dados, esse trabalho só é possível com o apoio de software, seja ele de prateleira, construído sob demanda ou adaptado.

1.3. Tipos de Análise de Dados

Existem quatro tipos de Análise de Dados, representadas na figura 1.3:

- **descritiva**, que se propõe a descobrir o que está acontecendo, em função de dados compreensivos, acurados e atualizados;
- **diagnóstica**, que busca identificar por que algo está acontecendo, por exemplo em busca de uma causa raiz;
- **preditiva**, que procura descobrir o que vai acontecer no futuro, e
- **prescritiva**, que procura uma indicação do que deve ser feito.

A figura 1.3 tenta indicar duas coisas: primeiro que há uma relação direta entre a complexidade de cada um desses tipos de análise de dados com o valor que ela pode trazer para a organização, depois que para uma forma mais complexa possa ser atingida, é necessário que a forma menos complexa tenha sido atingida antes. Isso porém não é sempre verdade, porque alguns modelos preditivos, como os gerados por redes neurais, podem não trazer uma explicação, isto é, diagnóstico ou descrição. Porém, teoricamente, essa descrição está embutida de alguma forma no modelo, nós é que não somos capazes, hoje, de interpretá-la.

Devemos levar em conta também que, pelo menos até agora, quase todo modelo de análise de dados é baseado no passado, isto é, criado em função das experiências existentes da organização, do mercado ou do mundo. É difícil imaginar um modelo de predição ou prescrição que não leve em conta algum conhecimento anterior, seja por aprendizado de

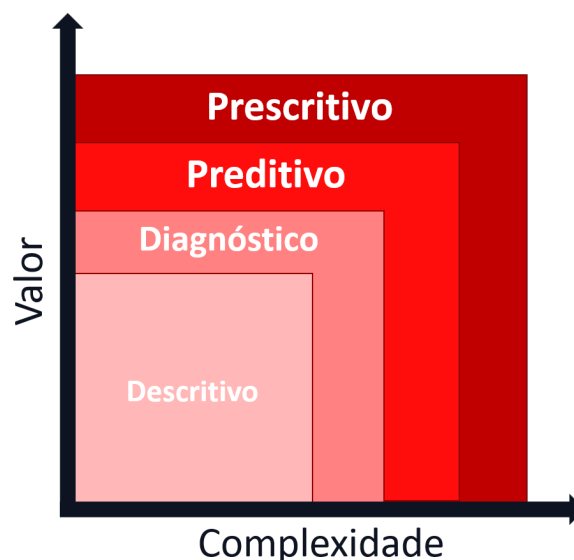


Figura 1.3.: Tipos de Análise de Dados

máquina, ou por conhecimento colocado explicitamente em modelos, por exemplo em tarefas de simulação.

A realização destas tarefas depende muito da disponibilidade de dados, por meio de tecnologias como Banco de Dados, relacionais ou não, Data Warehouses e Data Lakes. Os problemas mais difíceis envolvem Big Data, ou seja, dados que são de grande volume, variados e velozes, conhecidos como os três Vs do Big Data.

Para realizar a descrição de um negócio são necessários tantos relatórios e gráficos tradicionais, como o de faturamento mensal, como visões mais modernas, como visões georefenciadas do destino das entregas da empresa, ou dados demográficos de cliente e parceiros.

Já quando tratamos do diagnóstico, precisamos da habilidade de identificar a causa raiz, podendo usar instrumentos de controle de qualidade, como o gráfico de espinha de peixe, ou uma Análise de Pareto, ou ainda buscar ferramentas de business intelligence, como uma ferramenta OLAP, que nos permite ver a informação por vários ângulos.

Ambos os casos, que representam a análise do passado, podem ser acompanhados de abordagem mais sofisticadas, como Mineração de Dados, em busca de explicações, como regras de negócio, ou erros, como a busca de *outliers*. Um exemplo interessante é a mineração de processos, que pode mostrar se a empresa cumpre os processos que padronizou.

Os próximos dois casos se referem ao futuro, e dependem de ferramentas estatísticas ou de mineração de dados, tanto com modelos que podem ser completamente compreendidos, como no caso de árvores de decisão, quanto com modelos com alto grau de sofisticação, porém difíceis de entender, como redes de aprendizado profundas (*Deep*

Learning Networks), ou ainda de mecanismos de simulação e otimização.

1.4. Cenários de Análise de Dados

Podemos discutir também cenários onde a Análise de Dados acontece. São três os cenários principais:

- **sem dados**, onde começamos em uma situação onde, por motivos de negócio, os dados não estão disponíveis. Nesse caso é preciso criar dados, por exemplo com simulação ou ainda dados sintéticos a partir de modelos baseados em conhecimento prévio. Talvez seja surpreendente para alguns que exista um cenário de análise de dados onde iniciamos sem os dados, mas eles provêm de situações como lançamento de um novo produto ou serviço, criação de produtos, projeto ou processo onde os dados, artificiais, são criados de forma a permitir a análise de cenários;
- **com dados estruturados**, onde começamos a partir de dados já incorporados aos sistemas OLTP e OLAP da empresa, em sua grande parte já entendidos e organizados, de forma que já temos uma ideia de sua disponibilidade, qualidade e capacidade de servir para os quatro tipos de análise de dados, e
- **big data**, onde temos uma quantidade de dados com características de *big data*, principalmente variedade e volume, ainda não organizados pela empresa e muitas vezes com sua potencialidade ainda desconhecida.

1.4.1. Temperatura dos Dados

A **temperatura dos dados** é uma forma de tratar a frequência em que o dado é acessado, atualizado e mantido, durabilidade e a necessidade de manter um histórico sobre eles (Abbasi, 2021). O dado então pode variar de uma temperatura quente (*hot*) a fria (*cold*), com características exemplificadas na tabela 1.1.

Tabela 1.1.: Características típicas para as temperaturas dos dados

	Quente	Morno	Frio
Volume	MG-GB	GB-TB	PB-EB
Tamanho do item	B-KB	KB-MB	KB-TB
Latência	ms	ms, s	minutos, horas
Durabilidade	curta a longa	longa	muito longa
Velocidade dos Pedidos	muito alta	alta	baixa
Custo/GB	\$\$-\$	\$-€€	€

A ideia básica é que analisar dados on-line e que chegam rapidamente é um tipo de análise, e que avaliar dados guardados é outro. Porém os limites são fluídos. Existem

casos onde “dados frios” são poucos, como muitas vezes na Medicina, e outros onde os “dados quentes” tem um volume enorme.

1.5. Processos de Análise de Dados

Na prática, os processos de Análise de Dados se confundem com os processos de Data Science, Data Mining ou mesmo Business Intelligence. Existem várias descrições de processos para essas áreas, que são muito imbricadas, um mais e outros menos aceitos. Em geral, podemos avaliar que alguns processos começam com os dados, enquanto outros começam com o entendimento do negócio.

Três processos bastante conhecidos, adotados, mas talvez já antigos, são o CRISP-DM, o SEMMA e o KDD(Azevedo e Santos, 2008). A IBM propôs o ASUM-DM como evolução do CRISP-DM.

A metodologia **CRISP-DM**, ou **Cross Industry Standard Process for Data Mining**, é a mais divulgada. Ela é dividida em seis fases(Azevedo e Santos, 2008; Chapman et al., 2000), que se sub-dividem em tarefas genéricas, que por sua vez se dividem em tarefas especializadas, que também se dividem em instâncias de processo. As seis fases são:

- entendimento do negócio;
- entendimento dos dados;
- preparação dos dados;
- modelagem;
- avaliação, e
- implantação.

A ASUM-DM(IBM Corp., 2015) propõe cinco fases:

- analisar;
- projetar;
- configurar e construir;
- implantar e
- operar e otimizar.

A Microsoft propõe o **Team Data Science Process Lifecycle** (TDSP), dividido em quatro etapas que podem ocorrer em diferentes ordens(Microsoft, 2020):

- entendimento do negócio, sempre a inicial;
- aquisição de dados e compreensão;
- implantação
- modelagem

Apesar de ser importante seguir uma metodologia, a Análise de Dados é uma prática exploratória, e por isso iterativa e incremental. Muitas empresas a fazem de forma *ad-hoc* ou com processos próprios ou adaptados. A experiência também mostra que muitas

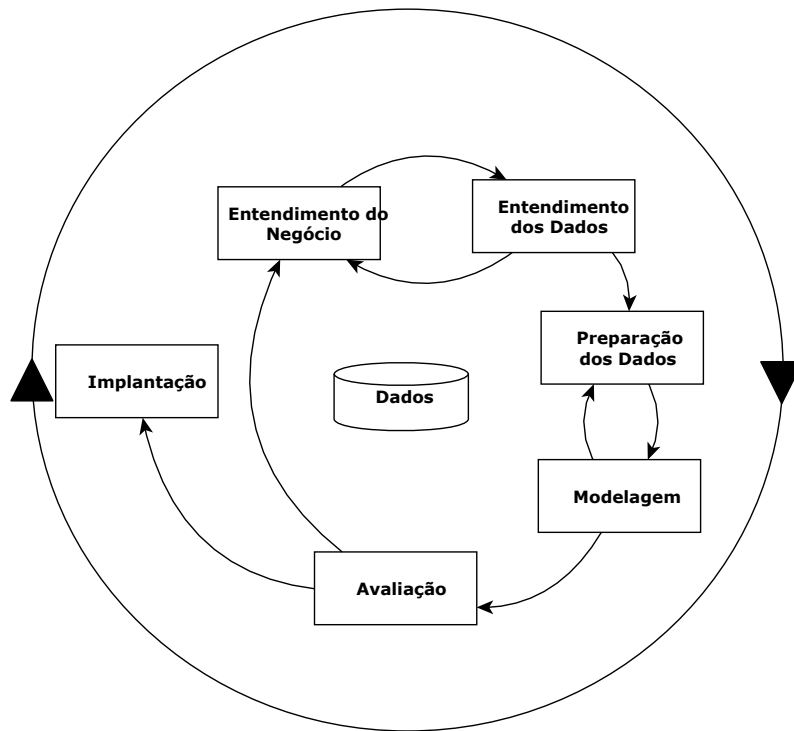


Figura 1.4.: A metodologia CRISP-DM (adaptado de (Chapman et al., 2000))

empresas afirmam seguir um processo, mas na prática apenas o definiram, não sendo plena sua implementação.

1.6. Conclusão

Este capítulo fez uma apresentação geral de vários conceitos de Análise de Dados.

O objetivo de qualquer projeto é **entregar valor às partes interessadas**. O conceito de valor, porém, tem várias acepções. Mesmo que muitos métodos e autores se baseiem neste objetivo, raramente valor é definido. Assim, fornecedores e clientes se perguntam: o que é valor? Como posso medir valor?

A resposta pode depender da escolha de uma visão comum do significado de valor que depende de todas as partes interessadas do sistema, logo, é influenciada também por todo o contexto onde o projeto está inserido, incluindo o *background* das partes interessadas.

Este capítulo busca mostrar as várias facetas da ideia de valor. Começa com uma abordagem genérica e depois mostra alguns entendimentos da palavra valor na economia, no marketing e na área de software. Obviamente há alguma similaridade entre as definições, mas os diferentes interesses entre as áreas também fornecem um conhecimento mais amplo do que é valor e de como as pessoas entendem o que é valor.

1.7. Conceituação Genérica de Valor

A palavra **valor** pode ser usada em vários sentidos (Cairncross, 1951, pg. 140):

- o **valor moral**, por exemplo, consideramos um *valor* a liberdade, a honestidade;
- o **valor estético**, por exemplo, damos *valor* a obra de Camões;
- o **valor em uso**, ou seja, ligado a utilidade, a como uma coisa atende as necessidades de alguém, como o valor da água, que é alto para a existência de vida, mesmo ela sendo barata;
- o **valor de troca**, por exemplo, o preço de uma casa a ser vendida, e
- o **valor ideal de troca**, o preço que um comprador imaginário pagaria pela casa.

Procuramos entender a acepção de valor com relação aos assuntos ligados aos negócios, em busca de entender **o que é valor para uma organização que está adquirindo ou desenvolvendo um software**. Isto não significa, porém, que a presença de atributos éticos e estéticos em um software não agregue valor, justamente o contrário. É possível, inclusive, que uma organização procure um produto que faça as mesmas funções que o produto que ela já possui, mas que seja mais agradável de usar, valor estético, ou garanta algumas propriedades ligadas ao valor ético, como privacidade.

Em todo caso, mesmo que a organização busque originalmente outro tipo de valor, no momento do projeto é mais provável que a interpretação dada ao conceito seja, na prática, ligada a fatores econômicos, financeiros ou legais, de acordo com a situação.

Da forma mais geral, é possível entender valor como **uma qualidade atribuída a algo ou alguém**, mas essa definição é muito ampla. Beleza, por exemplo, também é uma qualidade atribuída. Neste quadro em especial, valor é uma qualidade relacionada com a **utilidade que um produto ou serviço tem para os que o consomem ou usufruem**.

A princípio, valor econômico é “a importância que um indivíduo dá a determinado bem

ou serviço, seja para uso pessoal ou organizacional, seja para a troca” (Holanda Ferreira, 1986).

Essa definição do Dicionário Aurélio é interessante para projeto, porque já diz que o valor é ligado a importância dada ao bem ou serviço que está sendo recebido para o uso, ou sendo produzido para a troca.

Nas próximas seções veremos definições de valor nas seguintes áreas:

- Economia;
- Marketing;
- Projetos, e
- Software.

1.8. Valor na Economia

As tentativas de definir algo que pode ser chamado de valor na Economia tem, historicamente, também relação com o entendimento do preço dos objetos e serviços.

Na Economia Clássica, Smith, em 1776, primeiro reconhece que valor pode ter dois significados diferentes, um ligado a utilidade de um objeto particular, e outro ao poder de comprar outros bens. O primeiro seria o “valor de uso”, o segundo o “valor de troca” (Smith, 2003, Livro I, Cap. 4, para. 13)

Smith ainda discutiu valor, no sentido de preço, ou o preço relativo entre bens e serviços (Cairncross, 1951), com três abordagens (King e McLure, 2014), sendo que a terceira antecipou a teoria subjetiva de valor que apareceria mais tarde:

- o trabalho incorporado, que seria adequado às sociedades primitivas;
- a soma dos custos de produção, incluindo terra, capital e trabalho, mais adequado ao capitalismo, e
- a quantidade de trabalho e o incômodo de adquirir, ou que é poupado e pode ser imposto a outro, que é seu valor de troca (Strathern, 2003).

Podemos dizer que muitas empresas usam abordagens como essa para calcular o preço de seu software: calcular todos os custos de produção e somar um lucro, ou **markup**, sobre esse preço.

Também as organizações ou pessoas que estão pagando por um produto de software podem ser associadas a definição mais evoluída de Adam Smith: elas pagam para evitar algum incômodo.

Já em 1821, Ricardo (1996) propôs que para se saber o valor é necessário saber a utilidade, e que existiria um preço primário e natural, dado pela quantidade comparativa de trabalho necessário para a produção (King e McLure, 2014).

Novamente, a interpretação clássica de valor tem relação com a prática de desenvolvimento de software: utilidade para o usuário, custo de produção para o desenvolvedor.

1.8.1. Utilidade

Utilidade é o grau de satisfação que obtemos do consumo de bens e serviços (Krugman e Wells, 2013). Não é possível medir utilidade de forma prática, ou seja, dizer que algo tem 100 ou 2354 de utilidade para alguém. Porém, teoricamente, se discute a **função utilidade**, que é individual. Isso significa que cada pessoa tem uma satisfação diferente consumindo um produto, e não é possível comparar utilidade entre pessoas.

De acordo com o **princípio da diminuição da utilidade marginal**, cada unidade adicional de bens ou serviços adiciona menos a utilidade total do que a unidade prévia. Esse é um princípio genérico. A verdade é que existem produtos que tem utilidade marginal que aumenta, por exemplo, se você precisa de uma quantidade X para atingir um objetivo mínimo, a utilidade marginal não vai diminuir até você atingir X unidades, e pode até aumentar se for uma solução pior ter $X - 1$ unidades (Krugman e Wells, 2013). Um exemplo seria construir um muro de tamanho pré-determinado com tijolos, a utilidade dos tijolos não vai diminuir até atingir a quantidade necessária para o muro. Porém se você quer um muro que garanta privacidade, sem especificar a altura, a utilidade dos tijolos vai começar a diminuir quando certo tamanho é atingido, referente a uma privacidade necessária, até que a privacidade máxima seja alcançada e a utilidade marginal seja zero.

Software é um desses casos. Para um software funcionar, é necessária uma funcionalidade mínima que permita que ele comece a funcionar, porém após essa funcionalidade mínima, cada funcionalidade a mais vai seguir esse princípio, até o ponto que o cliente não mais vai querer pagar por uma funcionalidade, porque não vai valer a pena, já que o investimento vai ser maior que o retorno.

Utilidade é uma coisa importante no valor de produtos de software. Basicamente, em todas metodologias que usam o conceito de valor, o que se pode deduzir das explicações dadas é que **o mais útil em um certo contexto tem mais valor e deve ser priorizado**. Em software, utilidade pode ser um sinônimo de valor em alguns projetos.

1.8.2. Microeconomia Moderna e o Custo de Oportunidade

O **custo de oportunidade** indica “do que alguém deve desistir para obter o que deseja”(Greenlaw, Shapiro e Taylor, 2017), ou seja, indica a oportunidade perdida de se consumir ou usufruir outra coisa quando se consome ou usufrui de algo. É o “custo da próxima melhor alternativa” ao que está sendo consumido (Greenlaw, Shapiro e Taylor, 2017).

Segundo Krugman e Wells (2013), “o verdadeiro custo de algo é o seu custo de oportunidade”. Se alguém deseja um software, qual o verdadeiro custo disso? Isso envolve então o custo de oportunidade de uma segunda opção, por exemplo, não desenvolver o software e arcar com as consequências e investir o dinheiro.

Assim, novamente em software, o custo de oportunidade é um bom significado de valor.

Perguntas que podem ser feitas, em caso de discussões de valor, estão ligadas alternativa de não fazer o produto, ou não corrigir um erro. E, no cálculo de um valor financeiro, o preço das outras ofertas que vão ser recebidas pelo cliente.

O conceito de valor na Economia, como tratado aqui, nos permite trabalhar com algumas ideias onde o homem é visto como um ser racional, que faz cálculos e toma a decisão de forma acertada baseada nesses cálculos. Porém o homem também é um animal emocional, e sua visão de valor envolve necessidades e desejos, como veremos na abordagem que o Marketing dá ao conceito de valor.

1.9. Valor em Marketing

O estudo de valor em marketing se inicia com o entendimento do que são **necessidades**, ou seja, “os requisitos básicos dos seres humanos”: água, comida, recreação, educação, etc. (Kotler e Keller, 2012). Necessidades são estados onde se sente uma privação (Kotler, Armstrong et al., 2017). Em uma empresa, comprar um software que emite nota fiscal segundo a legislação vigente é uma necessidade.

Necessidades se transformam em **desejos** “quando são direcionadas para objetivos específicos que podem satisfazê-las”(Kotler e Keller, 2012). Desejos “são a forma que as necessidades tomam quando moldadas pela cultura e personalidade individual”(Kotler, Armstrong et al., 2017).

Por exemplo, a fome gera uma necessidade de se alimentar, que pode se configurar como o desejo de um hambúrguer para um americano, ou um prato feito de arroz, feijão, bife e batata frita para um brasileiro.

Da mesma forma, dada uma necessidade de um tipo de software, uma organização pode desejar características específicas adicionais, como ser feita de código aberto.

As necessidades podem ser (Kotler e Keller, 2012):

- **declaradas**, que são ditas pelo cliente, mas não são necessariamente verdade;
- **reais**, que são o que o cliente realmente precisa;
- **não declaradas**, que são o que o cliente espera receber;
- **de algo mais** (*delight*), que são o que o cliente gostaria, e
- **secretas**, que são como o cliente deseja ser visto pela sociedade.

Kotler e Keller (2012) ainda lembram que responder apenas a necessidade declarada pode não ser o bastante, e isso é verdade especialmente **em projetos de software, onde muitas vezes o cliente nem mesmo sabe suas necessidades reais**.

Uma **demanda** é um “desejo por um determinado produto, suportado por uma capacidade de pagar”(Kotler e Keller, 2012). Muitos querem lagosta no almoço, mas só podem pagar por algo mais barato.

Em software, a questão financeira é importante. É possível cotar um projeto que descrito em um parágrafo como algo entre 30 mil reais e quatro milhões de reais, dependendo da

1. Introdução a Análise de Dados

extensão do produto e da capacidade de operação necessária, como servidores e previsão de milhões de usuários.

Uma demanda pode ter oito estados (Kotler e Keller, 2012) em relação a um, ou mais, produtos ou serviços:

1. **demanda negativa**, quando os consumidores não gostam do produto e podem até pagar para evitá-lo;
2. **demanda não existente**, quando os consumidores não sabem que o produto existe ou não tem interesse no mesmo;
3. **demanda latente**, quando os consumidores tem um necessidade grande que não é satisfeita por nenhum produto existente;
4. **demanda declinante**, quando os consumidores passam a comprar menos ou deixam de comprar um produto;
5. **demanda irregular**, quando existe uma variação na compra do produto, em pequena escala de tempo (horas) ou grandes (estações do ano, meses);
6. **demanda plena**, onde os consumidores estão comprando todos os produtos postos no mercado;
7. **demanda excessiva**, onde os consumidores gostariam de comprar mais produtos do que podem ser fornecidos, e
8. **demanda indesejada**, onde os consumidores são atraídos por produtos com consequências sociais indesejadas.

Compreendendo essa sequência que começa com a necessidade, passa para o desejo e chega a demanda, é possível então falar de valor.

Para o marketing, o valor é a “a somatória dos benefícios tangíveis e intangíveis proporcionados pelo produto subtraída da somatória dos custos financeiros e emocionais envolvidos na aquisição desse produto”(Kotler e Keller, 2013). O valor é então uma “combinação de qualidade, serviço e preço”(Kotler e Keller, 2013). As percepções de valor “aumentam com a qualidade e o serviço, mas diminuem com o preço”(Kotler e Keller, 2013).

Segundo **Drucker:1974** o valor, para um cliente, é a satisfação de um desejo. Um dos objetivos do marketing é aumentar esse valor, gerando satisfação no cliente. A **satisfação** é então um “julgamento comparativo de uma pessoa sobre o desempenho percebido de um produto em relação as expectativas”(Kotler e Keller, 2013).

1.9.1. Os Elementos do Valor - B2C

Almquist, Senior e Bloch (2016) identificaram 30 elementos que fornecem valor para **consumidores**, i.e., pessoas, divididos em uma pirâmide² de quatro camadas³. Essa pirâmide é apresentada a seguir em forma de lista.

mais alta • **Impacto social**: auto-transcendência (ajudar outros ou a sociedade de forma mais

²A pirâmide é uma construção usada para passar a ideia que os níveis inferiores são essenciais para os níveis superiores poderem ser alcançados

³Uma versão interativa pode ser acessada em <https://media.bain.com/elements-of-value/>, link válido em 9/2/2020

ampla).

- **Mudança de vida:** fornecer esperança, auto-atualização, motivação, herança para as próximas gerações, afiliação/pertencimento.
- **Emocional:** reduzir ansiedade, recompensa pessoal, nostalgia, design/estética, representação de status ou aspirações, bem estar, valor terapêutico, diversão/entretenimento, aumentar a atração pessoal, fornecer acesso a outros itens de valor.
- **Funcional:** poupar tempo, simplificar, fazer dinheiro, reduzir risco, organizar, integrar aspectos diferentes da vida, conectar com outras pessoas, reduzir esforço, evitar problemas, reduzir custo, qualidade, variedade de escolha, apelo sensorial, informar.

mais baixa



Figura 1.5.: Elementos do Valor B2C. Fonte:Almquist, Senior e Bloch, 2016

Essa pirâmide pode ser comparada com a pirâmide de Maslow, apresentada na Figura 1.6, que mostra também as necessidades das pessoas em camadas cada vez mais abstratas. Sua aplicação é apropriada ao analisar valor em software que vai ser vendido para pessoas.

Os elementos podem ter importância diferente por pessoas e por indústria. Por exemplo, os elementos mais importantes para consumidores de *smartphones* são, em ordem decrescente (Almquist, Senior e Bloch, 2016): qualidade, reduzir esforço, variedade de escolha, organizar e conectar com outras pessoas⁴.

1.9.2. Os Elementos do Valor - B2B

Mais tarde, Almquist, Cleghorn e Sherer (2018) desenvolveram modelo semelhante para negócios B2B, obtendo uma pirâmide um pouco mais complicada, com 40 elementos,

⁴Apesar de ser a missão original do *smartphone*, seu valor como conexão é só o quinto mais valorizado.

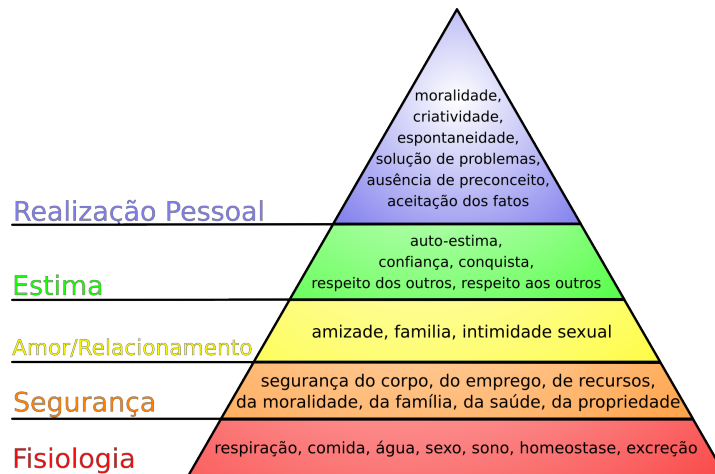


Figura 1.6.: Hierarquia das Necessidades de Maslow. Fonte: Wikipedia Commons por Felipe Sanchez (CC-BY-SA 3.0) e J. Finkelstein (GFDL)

onde há sub-níveis, representada na Figura 1.7 e na lista a seguir⁵. A importância de cada elemento novamente varia com a indústria. Essa segunda pirâmide é aplicável a software que vai ser vendido a organizações, porém leva em conta que as vendas são feitas para pessoas que participam das organizações. Os compromissos básicos são exigências para o vendedor.

mais alta • **Valor inspiracional**

- objetivo:
 - ◊ visão, ajuda o cliente a antecipar a direção do mercado;
 - ◊ esperança, dá aos clientes e usuários esperanças no futuro da empresa, e
 - ◊ responsabilidade social, ajuda o cliente a ser mais responsável socialmente.

• **Valor individual:**

- carreira:
 - ◊ expansão da rede (*network*), ajuda os usuários e colegas a expandir a rede profissional;
 - ◊ *marketability*, faz clientes e colegas mais *marketables* em seu campo, e
 - ◊ garantia de reputação, não atrapalha e pode aumentar a reputação do cliente no trabalho.
- pessoal:
 - ◊ design e estética, fornece produtos e serviços que são esteticamente agradáveis;
 - ◊ crescimento e desenvolvimento, ajuda usuários e colegas a crescer pessoalmente;
 - ◊ redução da ansiedade, ajuda clientes e outros na organização a se sentir mais seguros, e
 - ◊ diversão e vantagens, é agradável de interagir com ou dá recompensa de

⁵Uma versão interativa pode ser acessada em <https://media.bain.com/b2b-eov/index.html>, link válido em 9/2/2020.

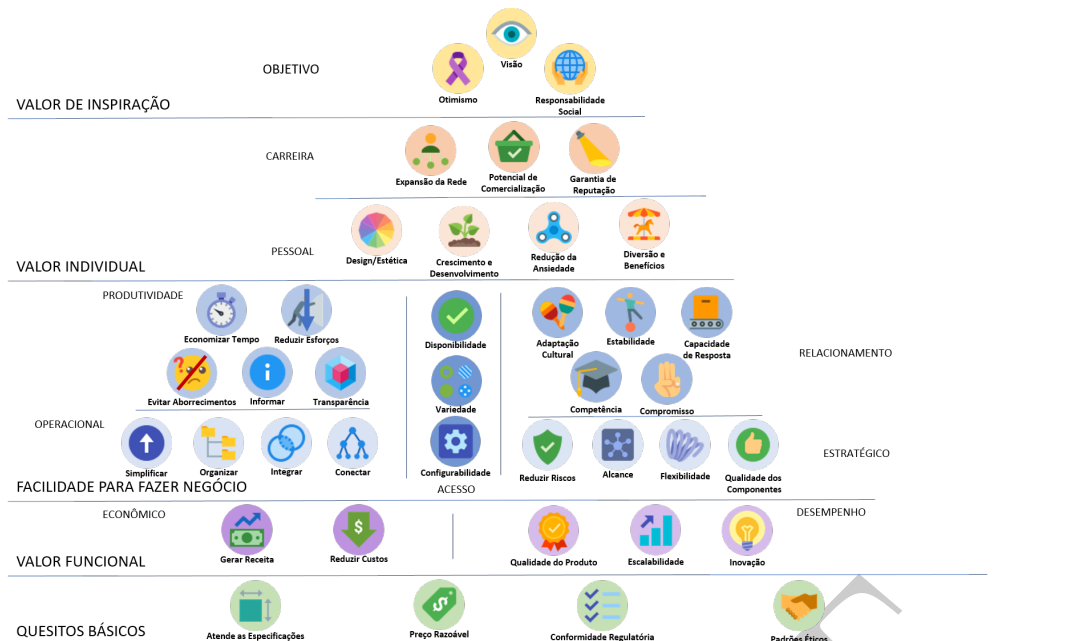


Figura 1.7.: Elementos do Valor B2B. Fonte: Almquist, Cleghorn e Sherer, 2018

alguma forma.

- **Valor da facilidade de fazer negócios:**

- produtividade:
 - ◇ poupar tempo;
 - ◇ reduzir esforço;
 - ◇ diminuir problemas;
 - ◇ informação, e
 - ◇ transparência, fornece uma visão clara da organização do cliente.
- operacional:
 - ◇ organização;
 - ◇ simplificação, reduz complexidade;
 - ◇ conexão, conecta organização e usuários com outros interna e externamente, e
 - ◇ integração, ajuda o cliente a integrar diversas facetas do negócio.
- acesso:
 - ◇ disponibilidade, garante que o bem ou serviço está disponível quando e onde necessário;
 - ◇ variedade, fornece uma variedade para escolha, e
 - ◇ configurabilidade, permite configurar o bem ou serviço de acordo com as necessidades do cliente.
- relacionamento:
 - ◇ responsividade, responde rápida e profissionalmente as necessidades da organização;
 - ◇ perícia, fornece *know-how* para o mercado ou indústria relevante;

1. Introdução a Análise de Dados

- ◇ compromisso, mostra que está compromissado com o sucesso do cliente;
- ◇ estabilidade, é uma empresa(bem ou produto) estável no futuro previsível, e
- ◇ ajuste cultural, se encaixa na cultura do cliente.
- estratégico:
 - ◇ redução de risco, protegendo o cliente;
 - ◇ alcance, permite o cliente operar em mais locais ou segmentos do mercado;
 - ◇ flexibilidade, vai além dos bens ou padrões comuns para permitir customização, e
 - ◇ qualidade de componentes, melhora a qualidade percebida dos produtos e serviços do cliente.

- **Valor funcional:**

- econômico:
 - ◇ redução de custos, e
 - ◇ aumento de receitas.
- desempenho:
 - ◇ qualidade do produto;
 - ◇ escalabilidade, e
 - ◇ inovação.

mais baixa ● **Compromissos básicos** (*table stakes*⁶)

- satisfazer especificações;
- preço aceitável;
- conformidade regulatória, e
- padrões éticos.

Ambas as pirâmides tentam ser exaustivas, mas foram criadas a partir de pesquisas de campo, o que quer dizer que podem existir outros elementos não detectados, porém isso não é esperado em geral.

Além disso, elas devem ser usadas em função dos elementos mais desejados em uma indústria ou cliente específico. Por exemplo, investigando a contribuição proporcional dos elementos para os clientes de infraestrutura de TI, com uma soma de 100%, (Almquist, Cleghorn e Sherer, 2018) encontraram que os três principais elementos são: qualidade do produto (7,8%), perícia (6,1%) e responsividade (5,5%). Poupar tempo, por exemplo, era décimo elemento na ordem e com proporção de apenas 3,0%. Na prática, a importância de cada elemento pode ser levantada com cada cliente.

1.9.3. Outras Técnicas

Tendo em vista que a finalidade do Marketing é engajar e gerenciar relações lucrativas com os clientes, é razoável que parte do trabalho de Marketing seja identificar o que é valor para o cliente, de forma que a organização possa atendê-lo com qualidade. Assim,

⁶No linguajar de negócios em inglês, *table stakes* é o mínimo que você deve fazer para um mercado ou negócio específico.

muitas técnicas de Marketing se dedicam a isso e está além deste livro cobrir todas.

Outras técnicas interessantes provenientes do Marketing são o QFD (Franceschini, 2016), adequada tanto a software sob encomenda quanto a produtos a serem lançados no mercado, e a Estratégia do Oceano Azul (Kim e Mauborgne, 2005), adequada apenas a novos produtos a serem lançados no mercado.

1.10. Priorização Baseada em Valor com o Cliente

A priorização é um processo para “determinar a importância relativa de informações”(IIBA, 2011). Ela permite várias abordagens:

- agrupamento, normalmente em poucas classes pré-determinadas;
- *ranking* ou ordenação;
- orçamento ou *time-boxing*, por meio da alocação de recursos finitos e
- negociação, em busca de um consenso.

Algumas das técnicas que podem ser usadas para criar uma estimativa relativa de valor com clientes e outras partes interessadas são (Satpathy et al., 2016):

- esquemas simples de pontuação, onde cada item do projeto analisado recebe um número entre uma variação pequena, como de 1 a 3 ou 1 a 5, ou uma rótulo como “Alto”, “Médio” ou “Baixo”, criando grupos ordenados;
- priorização **MoSCoW**, onde cada item é associado a um conceito do grupo “*Must Have*”, precisa ter, “*Should Have*”, devia ter, “*Could Have*”, podia ter e “*Won'tt Have*”, não vai ter (IIBA, 2011), criando grupos com significados específicos;
- **dinheiro de brinquedo** ou **100-pontos**, onde é dada uma quantidade de dinheiro de brincadeira⁷, como 100 moedas, ou pontos, para o cliente associar aos itens do projeto, algumas vezes submetido a algumas regras, como não ter X itens com o mesmo valor (Leffingwell e Widrig, 1999), criando uma ordem parcial baseada na alocação de recursos;
- **Análise de Kano** (Moorman, 2012), tratada na subseção a seguir;
- ordenação, usualmente feita com cartões com um item cada (J. Robertson e S. Robertson, 1998), mas também realizada em planilhas eletrônicas, possivelmente incluindo um processo em grupo, ou de negociação.

1.10.1. Análise de Kano

O modelo de Kano é um método para avaliar a reação emocional de clientes a características individuais de um produto. Kano detectou cinco respostas emocionais a essas características (Moorman, 2012):

- encantadores ou atrativos (*delighters*), “eu gosto disso”, que trazem satisfação se

⁷Facilmente obtida em uma caixa de Banco Imobiliário

1. Introdução a Análise de Dados

presentes, mas não trazem insatisfação se não estiverem presentes, normalmente porque são inesperadas e endereçam necessidades não reconhecidas;

- unidimensionais (*satisfiers*), que causam satisfação se estiverem presentes e insatisfação se não estiverem presentes;
- obrigatórios ou esperados (*dissatisfiers*), que causam insatisfação se não estiverem presentes, mas não causam satisfação se estiverem presentes;
- indiferentes (indifferent), ou
- indesejados.

Para classificar em uma dessas cinco respostas emocionais, são feitas duas perguntas ao cliente (Moorman, 2012):

- Como você vai se sentir se a funcionalidade estiver presente?
- Como você vai se sentir se a funcionalidade não estiver presente?

As respostas devem ser em uma escala ordinal:

- eu gosto disso;
- eu espero isso;
- eu sou neutro em relação a isso;
- eu posso tolerar isso, ou
- eu não gosto disso.

Finalmente, a classificação é feita cruzando as respostas na Tabela 1.2.

Tabela 1.2.: Resolvendo a classe de Kano por meio do cruzamento de duas perguntas (Moorman, 2012)

		Questão Negativa				
Questão Positiva	Gosto	Gosto	Espero	Neutro	Tolero	Não Gosto
	Espero	Conflito!	Atrativo	Atrativo	Atrativo	Unidimensional
	Neutro	Indesejado	Indiferente	Indiferente	Indiferente	Obrigatório
	Tolero	Indesejado	Indiferente	Indiferente	Indiferente	Obrigatório
	Não Gosto	Indesejado	Indesejado	Indesejado	Indesejado	Conflito!

As categorias de Kano não são permanentes, e as características de um produto, ou tipo de produto, vão migrando de atrativas, para unidimensionais, e para obrigatórias, e podem até mesmo se tornar indesejadas. O tamanho do celular, por exemplo, sofreu essa mudança com o tempo.

1.11. Valor Financeiro de Projetos

Apesar de não ser o objetivo principal desse capítulo, é interessante observar as formas de calcular o valor financeiro de um projeto, ou subprojeto.

Antes de tratar das estimativas financeiras usadas em projetos, é importante conhecer os seguintes conceitos, como definidos por **dalzoto:2015**:

- **principal** (P) é o capital inicial de um empréstimo ou uma aplicação financeira, também conhecido como **valor presente** (VP), **valor atual** (VA), **valor descontado**, ou por seu nome em inglês, **present value** (PV);
- **juro** (J) é a **remuneração** do capital emprestado, da parte de quem paga é uma despesa ou custo financeiro, da parte de quem recebe é um rendimento ou renda financeira;
- **montante** (S) é o **saldo**, ou **valor futuro** (VF), ou **future value** (FV), de um empréstimo ou de uma aplicação financeira. Pode ser também chamado de **valor de resgate**, ver equação 1.1.
- **prazo** (n) é o período de tempo que dura o empréstimo ou a aplicação financeira, podendo ser medido em dias, meses, anos...
- **prestação** ou **pagamento** (P,PGTO,PMT) se refere ao valor de pagamentos quando feitos em um número maior do que 1.
- **taxa de juros** (i) , ou simplesmente **taxa**, é o quociente entre juros e o principal, ver equação 1.2;

$$S = P + J \quad (1.1)$$

$$i = \frac{J}{P} \quad (1.2)$$

Analisando essas definições podemos detectar uma propriedade importante do dinheiro e que é a base da Matemática Financeira: o dinheiro muda de valor ao longo do tempo. Normalmente, como os cenários de inflação são muito mais comuns que os de deflação, R\$100,00 hoje compram mais que comprarão daqui a um ano e menos do que compravam um ano atrás. Por meio de cálculos como o do Valor Presente, como veremos no caso do Valor Presente Líquido, podemos comparar opções de investimentos ou empréstimos.

Calculadoras financeiras, como a HP-12C e o Excel permitem calcular facilmente o Valor Presente e o Valor Futuro com pagamentos, ou retiradas, fixas, em um período determinado, ao com uma taxa de juros fixas.

Por exemplo, se você tiver R\$10.000,00 por mês e puder aplicar em um projeto, deve procurar uma boa alternativa de investimento para poder saber se, financeiramente, vale a pena usar o dinheiro no projeto, ou se é melhor usar esse investimento. Com a fórmula do Valor Futuro (VF em português ou FV em inglês) esse cálculo pode ser facilmente feito no Excel, como mostrado na Figura 1.8.

Existem várias práticas que são utilizadas para calcular o valor financeiro de um projeto, entre elas (Satpathy et al., 2016):

- **Custo Total de Propriedade (TCO - Total Cost of Ownership)**, que calcula todo custo de um serviço ou produto, incluindo a realização e a o uso do mesmo.

Investimento Alternativo	
Taxa de Juros	1,00%
Período	12
Pagamento	R\$ 10.000,00
Valor Presente	R\$ 0,00
Valor Futuro	R\$ 126.825,03

(a) Valores e formatação

Investimento Alternativo	
Taxa de Juros	0,01
Períodos	12
Pagamento	10000
Valor Presente	0
Valor Futuro	=FV(Taxa_de_Juros;Períodos;Pagamento;Valor_Presente;0)*-1

(b) Fórmulas

Figura 1.8.: Cálculo do valor futuro de um investimento de R\$10.000,00 por mês com uma taxa de juros de 1%.

- Retorno do Investimento (ROI - *Return on Investment*), o método mais usado que é basicamente o lucro sobre o custo;
- Valor Presente Líquido - VPL (NPV - *Net Present Value*), onde se calcula a diferença entre a receita do projeto e seus custo ao longo do tempo, corrigido para o tempo presente, e
- Taxa Interna de Retorno - TIR (IRR - *Internal Rate of Return*) é a taxa de desconto de um investimento que torna o NPV zero para todos os fluxos de caixa de um projeto, não podendo ser calculado analiticamente, quanto maior for, mais desejável é o projeto (Hayes, 2019).

Devemos ter cuidado com projetos com resultados subjetivos, que não podem ser medidos diretamente pelo valor financeiro. Essa questão não será discutida nesse livro.

1.11.1. Custo Total de Propriedade

Quando se analisa o custo de um sistema é normal falar de Custo Total de Propriedade, conhecido pela sigla em inglês TCO (Total Cost of Ownership). O TCO é o valor presente de todos os custos a serem feitos durante a vida de um sistema, produto, serviço ou equipamento(anklesaria:2008).

Por exemplo, se decidirmos trocar todo o parque computacional de uma empresa que usa Windows para Linux, mesmo que o custo do Linux seja zero, o TCO é bem alto, pois envolve o processo de troca, novos profissionais, treinamento, etc... Outro exemplo comum é o da compra de uma impressora. Seu TCO não envolve apenas o custo da impressora, mas também o custo do material consumível, quando uma certa produção é prevista. Por

Tabela 1.3.: Valores usados para calcular o TCO de uma impressora.

Impressora	Laser A	Ink A	Laser B
Preço	R\$ 880,00	R\$ 400,00	R\$ 2.100,00
Preço Toner	R\$ 480,00	R\$ 105,00	R\$ 539,00
Páginas Toner	1000	480	12.000

isso é que grandes empresas compram menos impressoras, porém impressoras maiores e mais caras, para baixar o TCO, já que o custo por impressão delas acaba saindo menor.

Para o software desenvolvido vale o mesmo conceito. Qual seu TCO? Envolve o preço pago pelo software mais tudo que vai ser pago para possibilitar a implantação e utilização do produto (instalação, cursos de treinamento, manutenção mensal, etc...). Espera-se, em uma decisão racional, que o TCO seja menor que o benefício trazido pelo sistema.

Um exemplo típico é a análise a ser feita na compra de uma impressora. Impressoras mais caras normalmente tem o preço por página impressa mais barato, porém só a partir de um ponto de uma quantidade grande de impressões.

A Tabela 1.3 e a Figura 1.9 mostram como isso acontece. Nelas vemos qual a melhor alternativa, considerando apenas o preço da impressora e do tipo de impressão, qual a impressão mais barata. É possível, por exemplo, somar o gasto de energia a essa conta.

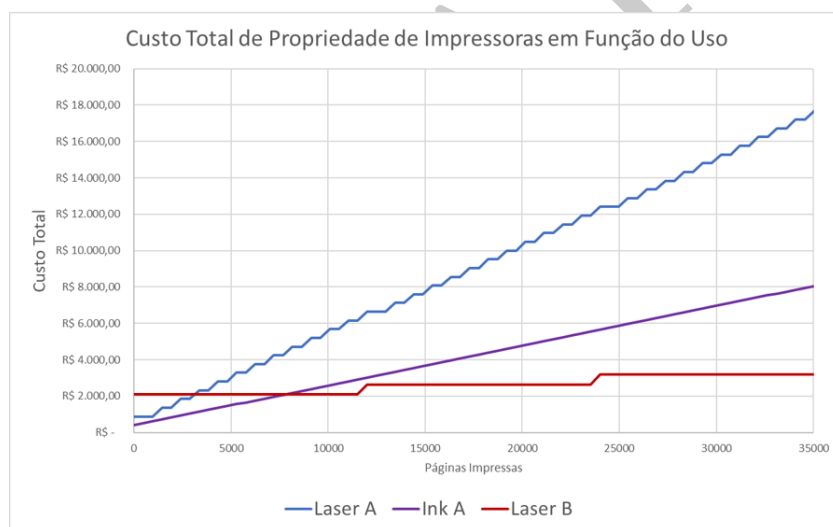


Figura 1.9.: Gráfico do TCO de impressoras em função de da quantidade de páginas impressas. Percebe-se que após aproximadamente 3500 páginas é melhor comprar uma impressora a laser mais cara, caso esse fosse o único fator de análise.

anklesaria:2008 divide os custos em possíveis classes:

- preço de compra, e
- custos internos, que se dividem em

1. Introdução a Análise de Dados

- **custos de aquisição**, como transporte, inspeção, armazenagem;
- **custos de uso**, como manutenção, garantias, e
- **custos do fim de vida**, como custos para desmontar um projeto, etc.

1.11.2. Retorno do Investimento

O cálculo do ROI depende de haver uma previsão para a receita corrente R e para o custo C do projeto. Sua fórmula é:

$$ROI = \frac{R - C}{C} \quad (1.3)$$

O cálculo do ROI de um projeto que está previsto custar R\$ 500.000,00 e pretende obter como benefício um valor de R\$ 700.000,00 é:

$$ROI = \frac{700 - 500}{500} = \frac{200}{500} = 40\% \quad (1.4)$$

O ROI pode ser usado para comparações com taxas que o mercado financeiro paga por um investimento do mesmo valor. Assim, so valeria a pena fazer um projeto, financeiramente falando, se ele resultasse em um retorno mais alto do que poderia ser obtido com o investimento em algum título do mercado.

Um cuidado ao analisar o ROI de um projeto é o fato de ser um valor percentual. Nesse caso, um projeto de R\$ 10.000,00 com um ROI de 150% pode parecer mais interessante que um projeto de R\$ 1.000.000,00 com um ROI de 1%, porém o segundo gera muito mais dinheiro que o primeiro.

1.11.3. Valor Presente Líquido

O Valor Presente Líquido é o valor atualizado para o presente de todos os gastos e ganhos de um projeto. Essa atualização é feita com uma taxa de desconto fixa. Isso significa que se você ganhar R\$ 100,00 daqui a um ano, e a inflação for de 5% ao ano, hoje isso tem o valor equivalente ao valor presente que, quando somado de 5%, resultasse em R\$100,00. A conta é $1/1,05$, que é aproximadamente R\$95,24.

A fórmula do VPL é(Kenton, 2019):

$$VPL = \sum_{t=1}^T \frac{C_t}{(1+i)^t} \quad (1.5)$$

onde C_t é o caixa líquido, receitas menos despesas, em um período t e i a taxa de desconto em um investimento alternativo, sendo T o número de períodos.

Esse cálculo é exemplificado na Tabela 1.4, onde usamos o método de calcular o saldo de cada período e atualizá-lo para o presente, com uma taxa fixa de 5%.

Período	Custo	Receita	Caixa	Valor Presente
0	R\$ 50.000,00	R\$ 0,00	-R\$ 50.000,00	-R\$ 50.000,00
1	R\$ 40.000,00	R\$ 52.500,00	R\$ 12.500,00	R\$ 11.904,76
2	R\$ 40.000,00	R\$ 52.000,00	R\$ 12.000,00	R\$ 10.884,35
3	R\$ 50.000,00	R\$ 61.500,00	R\$ 11.500,00	R\$ 9.934,13
4	R\$ 20.000,00	R\$ 31.000,00	R\$ 11.000,00	R\$ 9.049,73
5	R\$ 20.000,00	R\$ 30.500,00	R\$ 10.500,00	R\$ 8.227,02
6	R\$ 20.000,00	R\$ 30.000,00	R\$ 10.000,00	R\$ 7.462,15
7	R\$ 20.000,00	R\$ 29.500,00	R\$ 9.500,00	R\$ 6.751,47
Total				R\$14.213,63

Tabela 1.4.: Tabela de cálculo do Valor Presente Líquido, considerando saldos diferentes em cada período. Foi usada uma taxa de 5% a cada período. A fórmula para cada Valor Presente é $Caixa_i / (1 + Taxa)^{Período}$

1.11.4. Taxa Interna de Retorno

A Taxa Interna de Retorno é um cálculo um pouco mais complicado, exigindo um processamento de tentativa e erro. Ela indica de quanto deveria ser a taxa de desconto para o VPL ser zero, isto é, o resultado do projeto não ser nem positivo, nem negativo. O processo de calculá-la é numérico, mas o Excel e calculadoras financeiras fornecem funções para isso.

A fórmula que relaciona o VPL com o TIR é (Hayes, 2019):

$$0 = \text{VPL} = \sum_{t=1}^T \frac{C_t}{(1 + \text{TIR})^t} - C_0 \quad (1.6)$$

onde C_0 é o investimento inicial total.

Quanto mais alto for o TIR, melhor será o resultado do projeto, sempre lembrando que como o cálculo é percentual, isso depende do valor do projeto. Uma maneira de entender isso é que você precisaria de uma aplicação melhor que a taxa para o projeto não valer a pena financeiramente.

1.12. Valor em Software

Conhecendo algumas definições na Economia e no Marketing, podemos compreender melhor outras definições que encontramos na Engenharia de Software.

Segundo Erdogmus, Favaro e Halling (2006), **valor** é a “diferença entre benefícios e custos de um bem, ajustados ao risco, em um certo momento de tempo”.

Ele é dirigido por valores individuais ou coletivos. As partes interessadas esperam

1. Introdução a Análise de Dados

obter algum **benefício**, seja ele tangível ou intangível, econômico ou social, monetário ou utilitário, ou ainda estético ou ético (Biffi et al., 2006).

Valor significa esse benefício final, que existe geralmente nos olhos do observados e admite múltiplas caracterizações (Biffi et al., 2006).

O risco, que ainda não tínhamos encontrado em outras definições, é importante, e é parte das técnicas de gestão de projeto modernas considerá-lo. O risco pode ser incluído inclusive indiretamente no custo total previsto de um projeto multiplicando sua probabilidade pelo seu impacto. Isso faz pouco sentido para um risco único, mas já faz sentido para o conjunto total de riscos de um projeto.

Assim, o valor de algo pode ser visto como uma fórmula matemática, de uma forma básica como:

$$\text{benefício} - \text{custos} \quad (1.7)$$

Ou, considerando os riscos:

$$(\text{benefício} - \text{custos}) \times \text{correção por risco} \quad (1.8)$$

A partir dessa interpretação, busca valor é buscar os maiores benefícios a um custo que valha a pena, com risco baixo. Esses benefícios devem ser identificados no início do projeto, e estar alinhados com o planejamento estratégico da empresa.

É importante frisar que o valor deve ser em relação a organização, não ao software propriamente dito. Assim, algumas funcionalidades que poderiam tornar o software muito mais poderoso podem, na prática, não ser úteis, não trazer benefícios, custar caro demais ou ter risco muito alto. E isso depende de cada projeto, ou cada aplicação do software.

Gane e Sarson (1979) introduziram um o acrônimo **IRACIS**, que identifica três formas de agregar valor (Gane e Sarson, 1979; Ruble, 1997):

- Aumentar Faturamento (*Improve Revenue*);
- Evitar Custos (*Avoid Costs*), e
- Melhorar Serviços (*Improve Services*).

1.12.1. Triângulo do Valor

É comum se falar, em projetos de software, que existem três parâmetros, que são qualidade (ou escopo), tempo e custo, formando um triângulo, mas o cliente só pode definir dois. Apesar de ser tratado como um brincadeira, é realmente impossível arbitrar esses três valores de forma independente. Ou seja, a afirmação “eu quero um software com a funcionalidade X em tempo Y e custo Z” não pode ser feita sem conhecimento de causa (BeSeen, 2015; Kerzner, 2017).

A princípio, isso pode ser entendido a partir do fato que Boehm et al. (2000) mostraram que há uma relação entre a funcionalidade e qualidade desejada, e o esforço necessário

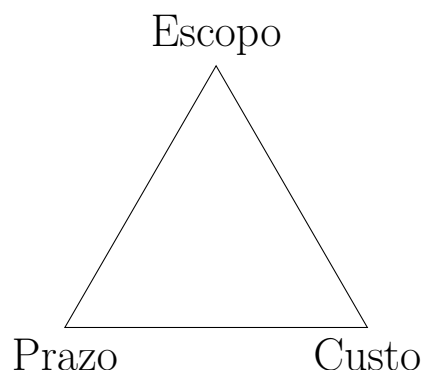


Figura 1.10.: O triângulo do valor para a maioria dos projetos.

para desenvolver o sistema, o que é óbvio. O que não era tão óbvio é que, além disso, dado um esforço, o tempo necessário para desenvolver o sistema também é determinado. Podem ser feitas certas modificações na previsão estatística, porém se forem feitas no sentido de diminuir o prazo necessário para a entrega do produto, elas implicam em risco maior. O assunto já foi tratado há muito anos por Brooks (1978) em seu livro adequadamente chamado *The Mythical Man-Month*.

Kerzner (2017) trata escopo, tempo e custo como restrições primárias que competem entre si em um projeto tradicional, e ainda apresenta restrições secundárias: reputação e imagem, risco, qualidade e valor. Já em projetos complexos, ainda segundo Kerzner, as restrições primárias passam a ser reputação e imagem, valor e qualidade, enquanto escopo, risco, custo e tempo passam a ser restrições secundárias. Podem aparecer também outros fatores, como segurança e estética, que aparece em parques de diversão, por exemplo.

1.12.2. Valor em Métodos Ágeis

É comum que métodos ágeis, como Scrum (Satpathy et al., 2016), proponham priorizar, de forma contínua, em ciclos curtos, os requisitos baseados no valor de negócio que é entregue aos clientes e usuários. O foco do Scrum, por exemplo, é tornar entregar valor, na forma de software executável, rapidamente, por meio de entregas incrementais a cada ciclo. Não há, porém, uma definição específica do significado da palavra valor, o que leva a necessidade de um entendimento maior do termo. Mesmo assim, cada história do usuário representa um valor a ser entregue ao cliente.

1.13. Conclusão

Não existe uma ideia unificada do que é valor, principalmente quando o conceito é comparado entre áreas distintas.

Como os clientes do analista de sistemas, partes interessadas nos projetos, estão em

1. Introdução a Análise de Dados

várias áreas diferentes, mesmo em um só projeto, então podem interpretar valor de formas diferentes também.

Mesmo assim, a principal função do desenvolvimento de software é produzir um software que tenha valor ao usuário, seja esse valor calculado por uma fórmula

Para isso é importante definir o que é valor no início de qualquer projeto, de maneira consensual entre as partes interessadas. Isso é uma prática ágil que deve sempre ser adotada.

A técnica MoSCoW também é muito prática e pode ser associada a qualquer outra. A partir dela pode ser mais fácil fazer uma ordenação de uma grande lista de requisitos, por exemplo.

Já a técnica de Kano é mais difícil de ser usada para projetos sob encomenda, mas é apropriada para o *design* de novos produtos.

O uso dos Elementos de Valor (Almquist, Cleghorn e Sherer, 2018; Almquist, Senior e Bloch, 2016) pode ser uma solução apropriada para o processo de discussão com as partes interessadas em busca de definições mais precisas e completas do verdadeiro valor.

1.14. Exercícios

Exercício 1.14-1: Procure outras definições de valor na Economia ou investigue mais sobre as definições dadas. Discuta a diferença entre essas definições e como elas mostram a evolução do entendimento do que é valor.

Exercício 1.14-2: Busque outras definições de valor, principalmente de outras áreas que não foram citadas neste capítulo.

Exercício 1.14-3: Vamos supor que Alice queira abrir um negócio. Para isso ela precisa de um software que permita registrar suas vendas. Sem esse software, por motivos legais, ela não pode vender. No mercado ela pode encontrar um software básico, que não atende todas as suas necessidades mas atende o requisito legal, por R\$ 100.000,00 por ano. Nesse caso, ela perderia a capacidade de mudar seus preços dinamicamente, o que faria com que ela deixasse de ganhar R\$ 240.000,00 por ano. Porém, a taxa de juros do mercado, se ela investir o dinheiro que usaria para o software novo é de 5% ao ano. Quanto ela pode pagar de aluguel, por ano, pelo software que atende completamente suas necessidades?

Exercício 1.14-4: Vá para o site <http://jogodeanalisedesistemas.xexeo.net/> e visite a Livraria Resolve. Identifique qual pode ser o conceito de valor para o projeto que essa organização deseja.

Dados e Informação

The goal is to turn data into information, and information into insight.

(Carly Fiorina)

Conteúdo

2.1.	Usando Significados Diferentes	32
2.2.	Conclusão	34

O tópico deste livro é Análise de Dados, onde buscamos transformar dados em informação, conhecimento e sabedoria, de forma a obter valor para uma organização. Mas o que significam essas quatro palavras?

Este capítulo trata de quatro palavras que no dia a dia parecem ser sinônimas, mas têm significados diferentes quando usadas na Análise de Dados: dado, informação, conhecimento e sabedoria.

Vamos recorrer a dicionários para ter uma definição inicial. Segundo o American Heritage, **informação**, no contexto de computadores, “é o dado quando processado, guardado ou transmitido”(American Heritage, 2019).

Já no dicionário Aurélio, informação, entre outros significados, pode ser “Conhecimento amplo e bem fundamentado, resultante da análise e combinação de vários informes”, “Coleção de fatos ou de outros dados fornecidos à máquina, a fim de se objetivar um processamento” ou ainda “Segundo a teoria da informação, medida da redução da incerteza, sobre um determinado estado de coisas, por intermédio de uma mensagem”.

Apesar de não estarmos diretamente envolvidos com a teoria da informação, não podemos deixar de notar a importância da definição que diz que a **informação reduz a incerteza por meio de uma mensagem**.

2.1. Usando Significados Diferentes

Dados, informação, conhecimento e sabedoria¹ são quatro palavras que na língua corrente possuem uma interseção de significados, porém são consideradas diferentes em áreas específicas, como Computação e Engenharia de Sistemas e Sistemas de Informação.

A distinção entre essas palavras e a existência de uma hierarquia de entendimento entre elas é uma discussão corrente, construída em cima de uma pirâmide que registra a complexidade do que esses termos significam, conhecida como hierarquia DIKW (Rowley, 2007).

Diversos autores propõe definições diferentes. Zins (2007) chega a apresentar 44 definições diferentes para dados, informação e conhecimento, fornecidas por especialistas no assunto reunidos em um painel, usando a metodologia *Critical Delphi* (Zins, 2012).

Seguindo a classificação de Zins (2007), preferimos uma conceituação baseada em um modelo onde dados e informação são fenômenos externos, enquanto conhecimento, e consequentemente sabedoria, são fenômenos internos. Isso significa que dados e informação são do domínio universal, externos a mente, enquanto conhecimento e sabedoria são do domínio subjetivo, internos a mente (Zins, 2007).

Dados são apenas os símbolos que usamos para representar a informação, o registro de diferentes aspectos de um fato ou fenômeno, como Os números que guardamos em um banco de dados. Dados não são interpretados, eles existem, são adquiridos de alguma forma, via coleta, pesquisa ou criação, guardados de outra forma e, possivelmente, apresentados em uma terceira. O computador é uma máquina programável que manipula dados.

Por outro lado, informação é o dado com significado, normalmente processado de forma a ser útil. Uma informação deve permitir responder perguntas como “quando”, “quanto”, “quem”, etc.

$$\text{Informação} = \text{Dado} + \text{Significado} \quad (2.1)$$

É necessário fazer um mapeamento entre dados e informação. Esse mapeamento pode ser simples ou complexo, dependendo de várias variáveis envolvidas, que vão desde decisões arbitrárias tomadas pelo desenvolvedor até padrões internacionais. Por exemplo, em muitos sistemas é preciso ter a informação do sexo de uma pessoa (masculino ou feminino). Para isso, são guardados um número (1 ou 0) ou uma letra (M ou F), que é o dado que faz o registro da informação.

¹Em inglês, *wisdom*

Existem outras definições para dados e informação, mas esta diferenciação entre símbolo e significado será suficiente neste livro. Em especial, na Teoria da Informação tem uma abordagem onde o significado não é importante, apenas a quantidade de mensagens possíveis(Claude E Shannon, 1963).

Definir conhecimento e sabedoria é muito mais complexo, sendo conceitos em que há mais incerteza ou mesmo desacordo sobre o que realmente significam formalmente, além do significado ambíguo dessas palavras na língua portuguesa ou inglesa. Porém, vamos tomar algumas decisões sobre esses significados, sempre no contexto deste livro.

Sumarizando vários autores, Rowley (2007) define conhecimento como “uma mistura de informação, entendimento, capacidades, experiências, habilidades e valores.” Além disso, o mesmo autor lembra a diferença entre conhecimento explícito, que está codificado em documentos, bancos de dados e outros registros, e tácito, que é está na mente das pessoas, ou nos processos que elas realizam. Essa diferença foi bastante explorada por Nonaka e Takeuchi (1995), que apontam também a dificuldade de processar computacionalmente conhecimento tácito.

Finalmente, Rowley (2007) afirma que vários autores tratando de dados, informação e conhecimento não falam de sabedoria, e aponta uma publicação anterior sua, que define sabedoria como “a capacidade de por em ação o comportamento mais apropriado, levando em conta o que é conhecido (conhecimento) e o que traz o bem maior (considerações sociais e éticas).”(Rowley, 2006).

Rowley (2007) também nos lembra que enquanto dados são muito fáceis de processar e contém pouco significado, e ficam na base da hierarquia, sabedoria é muito difícil de processar e fica no topo da mesma. Isso resulta em uma pirâmide como na Figura 2.1.

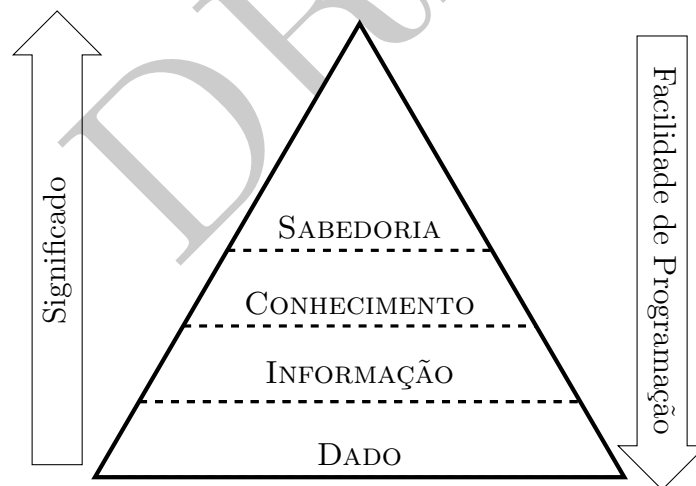


Figura 2.1.: Pirâmide da hierarquia do DIKW. Fonte:(Rowley, 2007)

2.2. Conclusão

Apesar do assunto deste livro ser conhecido como Análise de Dados, está claro que, para agregar valor a um empreendimento ou organização, o que queremos é gerar conhecimento e sabedoria a partir dos dados.

A tecnologia computacional trata do processamento de dados. Esses dados normalmente são apresentados de forma contextualizada que os transforma em informação.

Por exemplo, um sistema pode ter conter milhares de registros representando vendas, com números guardados em um formato binário que nada significa para o usuário, mas o usuário vai poder ver em um relatório o valor total vendido no mês, informação que é relevante para ele.

O conhecimento normalmente está mapeado nos algoritmos do sistema, nas regras de negócio implementadas. Alguns sistemas, que por exemplo usam aprendizado de máquina, podem inferir novos conhecimentos. Um exemplo disso seria o sistema de vendas descobrir que um tipo de item é mais vendido no início do mês, ou sempre junto de outro item, o que é chamado de análise de cesta de compras.

Já a sabedoria parece, por enquanto, estar ainda nas mãos dos seres humanos.

Tipos de Dados Convencionais

3.1. Introdução às medidas

A Análise de Dados trabalha basicamente com medidas do mundo real, possivelmente classificadas de alguma forma em suas dimensões, como podemos ver no modelo dimensional.

Medidas são um mapeamento do mundo real para o mundo formalDoe, 1666, normalmente na forma de um número ou símbolo, o valor, que caracteriza uma propriedade de um objeto no mundo real. Nós estamos acostumados com medidas como sexo de uma pessoa, distância de um caminho, peso de um produto. Palavras como quilograma, centímetro, são comuns para a gente, tanto que parecem naturais. As medidas, porém, são convenções. Os americanos, por exemplo, até hoje usam o sistema imperial, com pés, jardas e milhas para medir distância(Fenton e Pfleeger, 1998).

As medidas são criadas em alguns passos(Fenton e Pfleeger, 1998):

- identificar um atributo de uma entidade do mundo real;
 - altura de pessoas
 - temperatura do forno
- identificar relações empíricas para o atributo;
 - mais alta
 - mais quente
- identificar relações numéricas correspondentes a cada relação empírica;
- definir o mapeamento das entidades do mundo real com os números, e
- verificar que as relações numéricas preservam e são preservadas pelas relações empíricas.

Um ótimo exemplo da criação de medidas é a temperatura. No início não havia uma medida, e apenas eram feitas comparações, se algo era mais ou menos que outra coisa. Só em 1600 apareceu o primeiro instrumento que media se algo era mais quente que outra coisa. Com o tempo evoluíram o conceito para graus Fahrenheit (1720) ou

3. Tipos de Dados Convencionais

Celsius(1742), porém ainda com limitações, até que uma medida absoluta de temperatura foi encontrada, os graus Kelvin, em 1854. Mesmo assim, as escalas são arbitrárias. Por exemplo, na construção da escala Celsius se escolheu 0° como o ponto de congelamento da água e 100° como o ponto de fervura. Já na escala Kelvin, se escolheu 0° como o limite inferior de temperatura da natureza, onde não há movimento das partículas, o que faz sentido.

Medidas podem ser diretas ou indiretas. Uma medida direta não envolve outro atributo além do medido, como peso e largura. Já medidas indiretas envolvem outros atributos, como densidade e velocidade média.

As medidas tem algumas propriedades, que podem ou não possuir(Fenton e Pfleeger, 1998):

- Identidade, cada valor tem um único significado;
- Magnitude, valores possuem uma relação de ordem;
- Intervalos Iguais, diferenças entre os pontos da medida tem significado, e
- Valor de Zero Real, onde o zero não é arbitrário.

Levando em conta essas propriedades, as medidas podem ser divididas em(Fenton e Pfleeger, 1998):

- Categóricas ou Qualitativas, que se dividem em
 - Nominais, e
 - Ordinais
- Métricas ou Quantitativas, que se dividem em
 - Intervalares, e
 - De razão

3.2. Dados Nominais

Nos dados nominais os valores são apenas rótulos arbitrários que damos aos objetos, dividindo-os em categorias. Dizemos que números são pares ou ímpares, por exemplo.

Dados nominais não possuem ordem, e também não permitem operações aritméticas. Não podemos pegar 10 números que são pares e 10 que são ímpares e dizer que na média o conjunto é parímpar, por exemplo.

Na prática, qualquer representação simbólica ou numérica é válida para dados nominais. O sexo de uma pessoa, por exemplo, pode ser representado por palavras, como homem e mulher, por desenhos, por números como 0 e 1, por letras como M e F. Qualquer escolha é arbitrária e está apenas indicando os objetos no mundo real que estão em uma ou outra categoria.

Mesmo que sejam usados números, não há noção de ordem ou grandeza, assim não podemos dizer que alguém da categoria homem é maior ou menor que alguém da categoria mulher a partir dessa medida. Não podemos dar valores ou associar quantidades às

Tabela 3.1.: Contagem e distribuição do estado civil de pessoas em uma amostra.

Estado Civil	Frequência	Percentual (%)
Solteiro	125	25
Casado	175	35
Divorciado	50	10
Viúvo	75	15
Outro	75	15

classes.

Veremos mais tarde que as descrições possíveis são limitadas a contagens, percentuais e a uma medida estatística conhecida como moda, que é a classe com maior frequência em uma amostra.

Existe um tipo especial de dado nominal, quando só há duas categorias: dados dicotômicos. Um tipo comum de dado dicotômico é o resultado de uma pergunta do tipo que só aceita as respostas sim ou não, sendo que uma delas é obrigatória. Em contrapartida, dados como a cor de algo são chamados pelo nome, raramente usados, de politômicos.

Exemplos de medidas nominais são: cores, de olhos, cabelos, roupas etc; estado civil; país de origem, espécie de animais, etc.

A visualização de dados nominais é também restrita a gráficos que apresentam contagens e frequências, como histogramas, e gráficos que apresentam percentuais, como gráficos de pizza. A ordenação pode ser feita por frequência. Exemplos simples de visualização de dados nominais podem ser vistos na tabela 3.1 e nas figuras 3.1, 3.1 e 3.3, construídas com o Excel.

[hbt]

A figura 3.4 e mostra um histograma onde 3 variáveis nominais são apresentadas em um mesmo histograma, construído com o Excel (Pivot Graphic).

As figuras 3.5 e 3.6 mostram gráficos ainda mais sofisticados que permitem comparações específicas, construídos com o software Tableau.

3. Tipos de Dados Convencionais

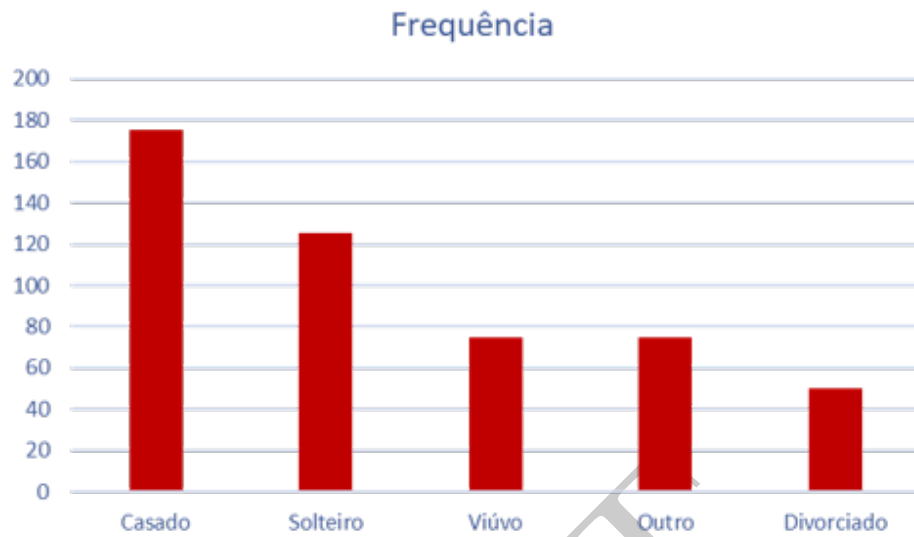


Figura 3.1.: Histograma da distribuição do estado civil das pessoas em uma amostra.

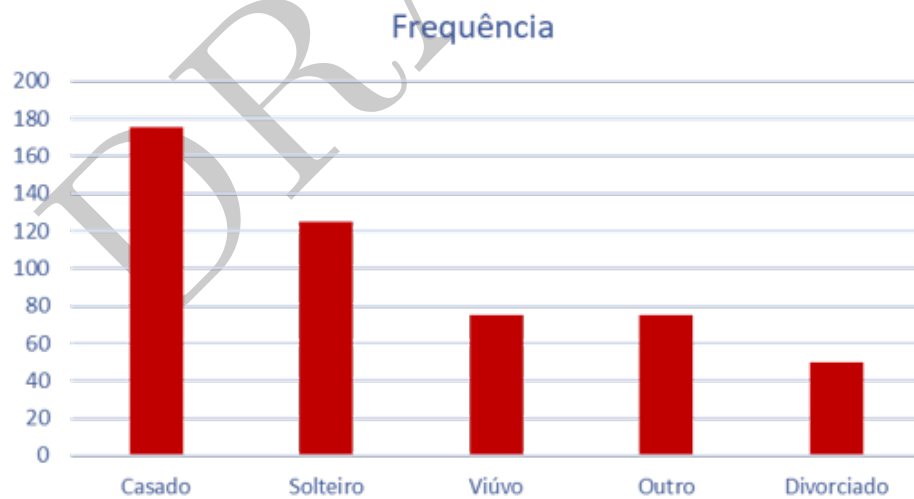


Figura 3.2.: Histograma do estado civil de pessoas em uma amostra, ordenado pela frequência.

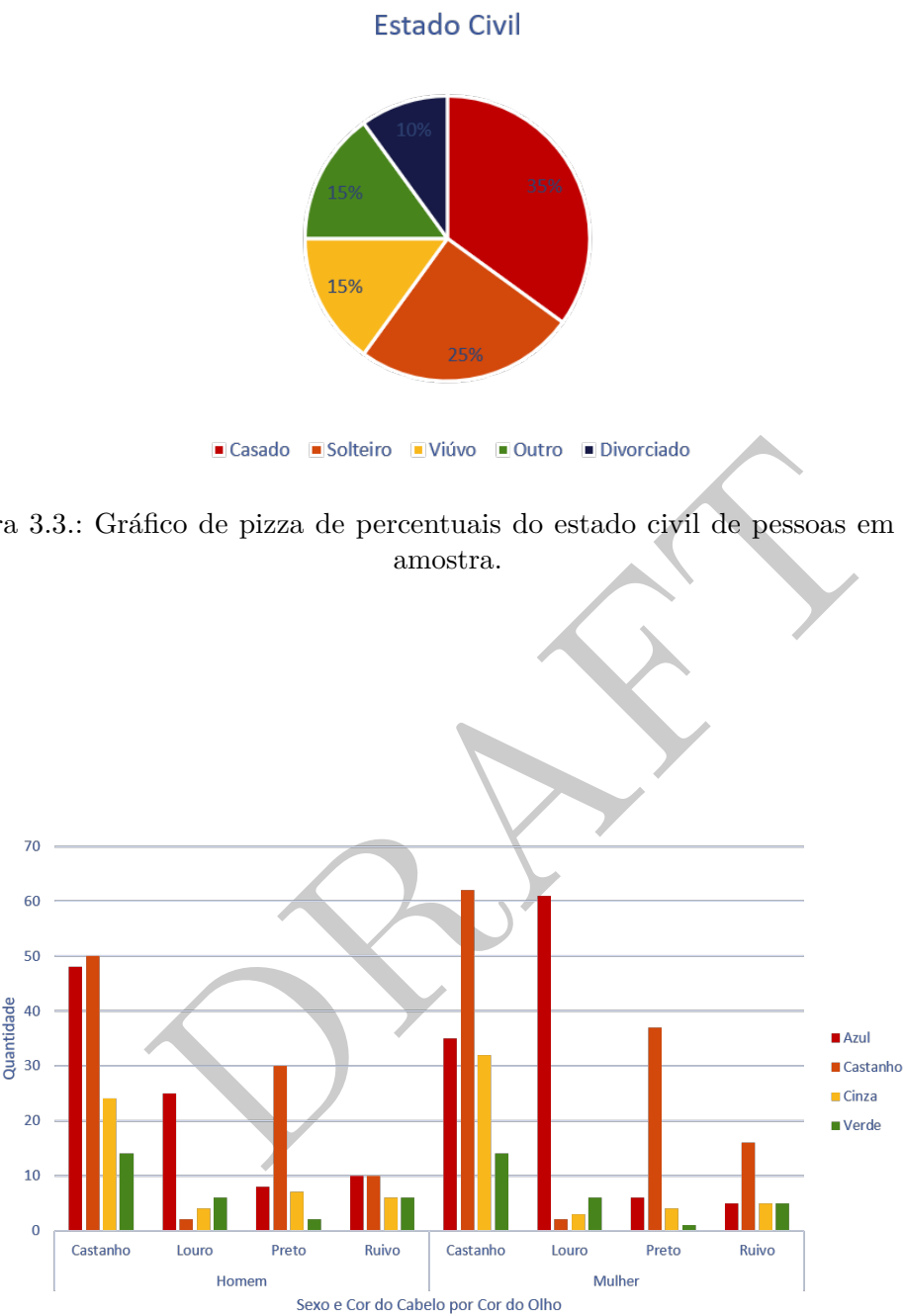


Figura 3.4.: Gráfico (histograma) um pouco mais sofisticado, mostrando várias informações que são baseadas na contagem de dados nominais simultaneamente.

3. Tipos de Dados Convencionais

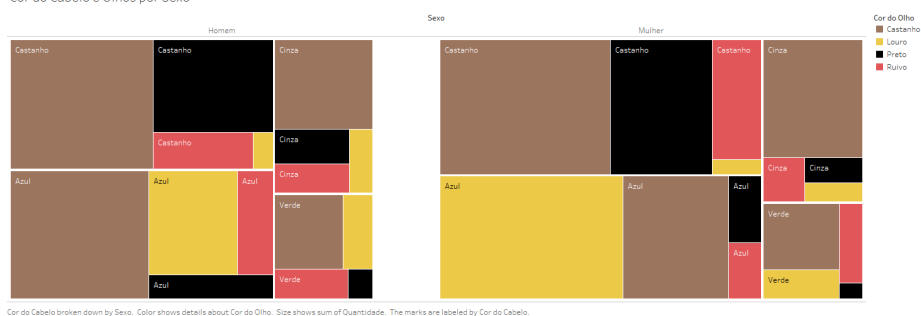
Proporção e quantidade de homens e mulheres de acordo com cor de olhos e cabelo



Sexo (color) and sum of Quantidade (size) broken down by Cor do Olho vs. Cor do Cabelo.

Figura 3.5.: Gráfico construído com o software Tableau mostrando a quantidade e proporção de homens e mulheres de acordo com cor dos cabelos e dos olhos.

Cor do Cabelo e Olhos por Sexo



Cor do Cabelo broken down by Sexo. Color shows details about Cor do Olho. Size shows sum of Quantidade. The marks are labeled by Cor do Cabelo.

Figura 3.6.: Gráfico feito com o software Tableau, mostrando as proporções em uma amostra da cor do cabelo, como texto, e da cor dos olhos, como cor, permitindo comparar a distribuição entre homens e mulheres.

3.3. Dados Ordinais

Dados ordinais são rótulos ordenados. É possível dizer que um é maior ou melhor que outro, mas não podemos quantificar quanto. Um exemplo de dados ordenados são: patentes do Exército (como tenente, capitão, major).

Dados ordinais muitas vezes são provenientes de questionários, onde há um número finito de opções. Por exemplo, uma escala de frequência (nunca, raramente, algumas vezes, frequentemente, sempre) indica quantas vezes alguém faz algo durante um período. Cada uma dessas classes claramente indica que faz mais vezes, porém não é possível determinar quanto mais. Outra escala de frequência é: horária, diária, semanal, mensal, trimestral, semestral e anual. Da mesma forma, como não há uma exatidão, não podem ser comparadas em sua magnitude.

É comum também que dados que estão em um domínio contínuo, como salário, sejam organizados em uma escala ordinal que fala de intervalos, como: até R\$1200,00, de R\$1201 a R\$5000 etc.

3.3.1. Escala de Likert

Um tipo muito frequente de dados ordinais são aqueles obtidos em um questionário que segue uma escala de Likert. Normalmente esta escala é composta de 5 graus de concordância com uma afirmação: discordo totalmente, discordo, neutro, concordo, concordo totalmente. A figura 3.7 é um exemplo pequeno de um questionário usado escala de Likert.

As Escalas de Likert são amplamente usadas em pesquisas de vários tipos, sendo originadas na Psicologia. Devem ter um número ímpar de itens, sempre tendo um item médio, e são aconselhados uma quantidade de 3, 5 ou 7 níveis, sendo 5 a quantidade mais comum.

3.3.2. Escalas Ordinais e Números

Muitas vezes escalas ordinais são associadas a números, porém, teoricamente, esses números não podem ser usados para fazer contas. Por exemplo, não é possível calcular a média. Na literatura científica esse princípio é muitas vezes questionado, a partir de argumentos razoáveis, porém muitos pesquisadores e analistas de dados realizam a média da opinião sem levar em consideração as questões.

Dados em escalas ordinais devem ser tratados com cuidado. Só devem ser calculadas as frequências, as modas e as medianas, mas não as médias.

- Quanto a X
 - Qualitativo
 - Quantitativo

3. Tipos de Dados Convencionais

Considerando as afirmativas abaixo, escolha seu grau de concordância com elas.

	Discorda totalmente	Discorda parcialmente	Nem concorda nem discorda	Concorda parcialmente	Concorda totalmente
Este livro será útil em sua profissão.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Este livro é mais fácil do que você pensava.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figura 3.7.: Questionário com escala de Likert, produzido com o Google Forms.

Tabela 3.2.: Resumo das características de cada tipo de dado.

Característica	Nominal	Ordinal	Intervalar	De Razão
ordem conhecida		✓	✓	✓
podem ser feitas contagens	✓	✓	✓	✓
admite uso da moda	✓	✓	✓	✓
admite uso da mediana		✓	✓	✓
admite uso da média			✓	✓
pode quantificar a diferença entre valores			✓	✓
pode somar ou subtrair valores			✓	✓
pode multiplicar e dividir valores				✓
possui um zero real				✓

- Quanto a Estrutura
 - Estruturado
 - Semi-Estruturado
 - Não-Estruturado
- Quanto a Y
 - Z
 - ◊ Classes
 - ◊ Ordinais
 - Numéricos
 - ◊ Intervalar
 - ◊ De razão
 - ◊ Apenas Contáveis

3.4. IDEIA

Os dados nominais e ordinais também podem ser divididos em Kinds, Fases e Papéis. Marcas de Carros são Kinds, Postos no Exército são fases e ser funcionário aluno ou professor são papéis.

DRAFT

Parte II.

Introdução as Ferramentas

DRAFT

DRAFT

Parte III.

Exploração Inicial dos Dados

DRAFT

DRAFT

Parte IV.

Mineração de Dados

DRAFT

DRAFT

Parte V.

Dados Textuais

DRAFT

DRAFT

Parte VI.

Dados Georeferenciados

DRAFT

DRAFT

Parte VII.

Grafos

DRAFT

DRAFT

Parte VIII.
Entendendo o Negócio

DRAFT

DRAFT

Modelagem de Processos de Negócio

The problem is that at a lot of big companies, process becomes a substitute for thinking. You're encouraged to behave like a little gear in a complex machine. Frankly, it allows you to keep people who aren't that smart, who aren't that creative.

(Elon Musk)

Conteúdo

4.1.	Processos de Negócio	60
4.2.	Exemplos de Processos	61
4.3.	Representação de Processos de Negócio	62
4.4.	Por que Entender os Processos de Negócio	62
4.5.	Exercícios	63

Por que processos de negócio?

Os sistemas de informação executam dentro das organizações para atender processos de negócio, logo é preciso entendê-los para melhor atender as partes interessadas.

Cada vez mais a Modelagem de Processos de Negócio vem sendo usada como uma ferramenta principal ou de auxílio do processo de desenvolvimento de software, incluindo

data warehouses e aplicações relacionadas, visando o levantamento completo dos requisitos do sistema.

4.1. Processos de Negócio

Uma das informações importante para o conhecimento de uma organização é entender o que ela faz e como ela faz para atingir seus objetivos de negócio.

Normalmente, isso é obtido por meio da descrição de seus processos de negócio.

Um **processo** “é uma coleção de tarefas inter relacionadas que são iniciadas em resposta a um evento, e que tem como objetivo alcançar um resultado específico para o consumidor do processo”(Scheel et al., 2015).

Processos aparecem em todos os lugares, e podem ser documentados de maneira mais ou menos formal. Um processo simples que usamos no dia a dia é o de fazer um café expresso em uma máquina de cápsulas. Os passos desse processo podem ser descritos como:

1. ligar a máquina de expresso;
2. escolher a cápsula de café;
3. colocar água na máquina;
4. colocar a cápsula na máquina
5. colocar a xícara no local indicado
6. apertar o botão de fazer café
7. remover a xícara

É possível notar que mesmo um processo simples que é feito no dia a dia apresenta dificuldades de descrição. Na lista anterior, por exemplo, não se discute o que fazer se a máquina não estiver na tomada, ou se uma cápsula já está na máquina, ou se já tem água na máquina. Isso significa que uma descrição de um processo simples de forma completa pode exigir um cuidado muito grande com detalhes, se tornando, no final, muito complexa.

Todas as organizações realizam processos. Esses processos podem ter nomes como “Atender Pedido do Cliente” ou “Fabricar Peça”.

A **modelagem de processos** demonstra como funciona a organização, passo a passo, no seu dia a dia. A partir dela pode ser possível levantar os pontos a serem automatizados de um processo e como os processos realmente realizados diferem dos processos normatizados da organização.

Um **processo de negócio** “uma coleção de tarefas e atividades (operações e ações de negócio) consistindo de empregados, materiais, máquinas, sistemas e métodos, que são estruturadas de forma a projetar, criar e entregas um produto ou serviço ao consumidor”(Scheel et al., 2015).

Processos de negócio são sempre descritos a partir de um ponto de vista que indica

a profundidade com que são tratados. Um processo de “Calcular o Preço Final”, por exemplo, tanto pode ser visto apenas como uma atividade única, como por várias etapas que passam por partes diferentes da organização, como ainda por um algoritmo detalhado.

Entre as características principais de processos de negócio é possível citar que eles sempre são sequências de passos, sempre respondem a uma necessidade ou demanda que define seu início e sempre chegam a um resultados específico e mensurável, seja ele o sucesso ou o fracasso do atendimento a necessidade ou demanda original.

A **Modelagem de Processos de Negócio** tem como objetivo entender os processos que ocorrem na organização, listá-los, descrevê-los por meio de uma representação adequado, indicando o passo a passo em um nível de abstração adequado ao propósito e, também, verificar se as representações correspondem a realidade.

Linguagens de descrição de processo de negócio, como ARIS e BPMN, se preocupam em fornecer uma abstração razoável para a descrição de processos. Normalmente essa abstração permite descrições tanto de um ponto de vista muito amplo, com a supressão da maioria dos detalhes, que visa mais entender o que é feito, como um ponto de vista razoavelmente detalhado, que permite ao leitor entender realmente como o processo é executado na organização, ou seja, como algo é feito. Cabe ao modelador do processo escolher e manter esse nível de abstração ao longo do trabalho.

4.2. Exemplos de Processos

É interessante ver alguns exemplos de processos sendo descrito na forma de uma narrativa simples que não inclui alternativas.

O processo de entrega de uma pizza pode ser descrito como:

1. O atendente recebe o pedido de pizza e bebidas
2. O pedido da pizza passa para o pizzaiolo, que prepara as pizzas e coloca na fila para assar
3. O pizzaiolo coloca as pizzas do pedido para assar
4. Quando as pizzas estão prontas, o pizzaiolo retira e passa as pizzas para seu ajudante colocar em caixas
5. Se foram pedidas bebidas, o atendente junta as bebidas ao pedido e o separa
6. O entregar pega o pedido completo e passa no caixa para pegar o troco
7. O entregador entrega a pizza
8. O entregador cobra a pizza do cliente
9. O entregador traz o dinheiro para o caixa

Já uma loja de construção pode ter o seguinte processo para atender um pedido:

1. O vendedor recebe o pedido
2. O vendedor levanta o preço de cada produto
3. O vendedor confirma com o cliente os produtos que vão ser comprados

4. Modelagem de Processos de Negócio

4. O vendedor envia a lista de produtos vendidos para o caixa
5. Enquanto o caixa faz a cobrança, O vendedor separa os produtos desejados
6. Enquanto o vendedor separa o pedido, o caixa cobra o pedido do cliente
7. Enquanto o vendedor separa o pedido, O caixa emite a nota fiscal
8. O vendedor junta a nota fiscal ao pedido
9. O vendedor entrega o pedido ao cliente

O que essas duas descrições permitem entender sobre processos de negócio, e mesmo processos em geral?

Primeiro, elas descrevem um passo a passo do que é feito. Todo processo é um passo a passo para realizar algo. Segundo, elas começam com um evento. Todo processo tem que ter um motivo para começar e deve atender uma necessidade definida por esse motivo.

Outra característica importante: os passos de um processo levam a organização a chegar mais perto do resultado desejado.

Além disso, no processo da pizzaria existe um passo opcional, enquanto no processo da loja de ferragens existem passos paralelos. Tanto passos opcionais como paralelos são comuns em processos de negócio.

Finalmente, os dois são descrições parciais do funcionamento do negócio.

4.3. Representação de Processos de Negócio

Várias são as formas de representar processos de negócio. É possível dividir essas formas em dois grupos básicos: formas textuais e formas gráficas.

Entre as formas textuais encontramos os Casos de Uso, as narrativas formais ou informais, linguagens declarativas e outras.

Entre as formas gráficas, podemos encontrar formas comuns mas hoje consideradas ultrapassadas, como fluxogramas e diagramas IDEF0 e IDEF3, formas usadas na academia mas com pouco uso na área de negócios, como Redes de Petri, e formas atuais em uso no mercado, como ARIS, Diagramas de Atividade e a principal ferramenta da atualidade: os diagramas BPMN.

4.4. Por que Entender os Processos de Negócio

Normalmente sistemas de informação estão inseridos e automatizam processos de negócio. Eles atuam garantindo que aconteçam, ajudando na sua gerência, monitoração e controle nos vários níveis da organização.

Durante os processos de negócio dados primitivos são criados e armazenados. São

os sistemas de informação que tratam esses dados, e possibilitam gerar informações adicionais, derivadas a partir deles.

Assim, entender os processos de negócio facilita a entender o ambiente dos sistemas de informação e seus requisitos.

4.5. Exercícios

Exercício 4.5-1: Descreva o processo de fazer o café expresso em uma máquina de cápsulas da forma mais detalhada possível, considerando as seguintes condições iniciais:

- a máquina de café está em cima do balcão, desligada da tomada e sem água;
- as xícaras estão no armário;
- as cápsulas de café ainda estão fechadas em sua caixa com 10 unidades, na despensa;
- as portas do armário e da despensa estão fechadas;
- é possível pegar água no filtro.

DRAFT

DRAFT

Conteúdo

5.1.	Usos de BPMN	66
5.2.	Participantes	67
5.3.	Símbolos básicos	67
5.4.	Atividades	67
5.5.	Eventos	70
5.6.	<i>Gateways</i>	73
5.7.	Fluxos	75
5.8.	Alguns Exemplos	76
5.9.	Eventos Anexados a Atividades	76
5.10.	Exercícios	76

Por que BPMN?

BPMN é o padrão *de facto* da modelagem de processos, ferramenta essencial para qualquer projeto de sistema de informação para uma organização

BPMN, sigla para *Business Process Model and Notation*, é uma linguagem gráfica para a modelagem de processos de negócio padronizada pela OMG e que está se tornando o padrão *de facto* da indústria.

A Figura 5.1 mostra um exemplo razoavelmente simples de um processo descrito em BPMN. Nesse diagrama podemos ver um processo que se inicia na tarefa “Receber

Pedido de Crédito” e continua na tarefa “Aprovar Pedido de Crédito”. Nesse momento encontramos um *gateway*, que nesse caso é uma decisão onde apenas um caminho pode ser seguido (o losango sem símbolo), e caso o pedido não tenha sido aprovado, seguimos para a tarefa “Avisar Resultado”, ou, caso o pedido tenha sido aprovado, seguimos para a tarefa “Efetuar Transação de Crédito” e então também para a tarefa “Avisar Resultado”, que sempre leva ao fim do processo.

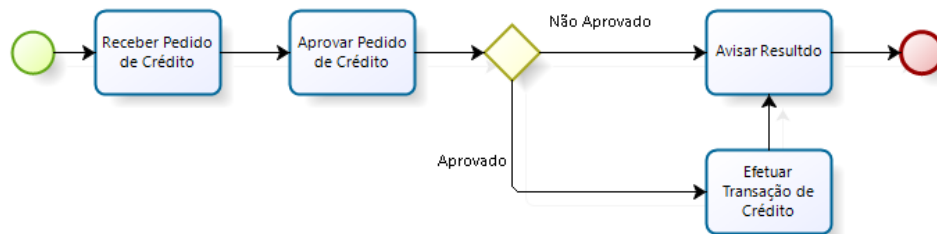


Figura 5.1.: Um exemplo de processo descrito em BPMN. Fonte: imagem do autor

5.1. Usos de BPMN

BPMN foi criado para unificar a representação de processos de negócio, tendo em vista a grande quantidade de linguagens existentes. Além disso, ele mira permitir não só a descrição de processos, mas também sua simulação e execução.

BPMN pode ser usado para criar **diagramas de processo**, **diagramas de colaboração** e **diagramas de coreografia** (OMG, 2013)(OMG, 2013).

Processos, ou orquestrações, podem ser privados ou públicos. Processos privados ainda podem ser executáveis ou não. Processos internos também são chamados de *workflows*, ou *orquestrações* no caso de serem suportados por *web services*(OMG, 2013), e descrevem processos dentro de organizações.

Um processo executável é modelado para ser executado, enquanto processos não executáveis são criados para documentação do comportamento de um processo de negócio da organização (OMG, 2013).

Um processo é **público** quando mostra a interação entre um processo privado e outro processo, ou participante, e sua representação só inclui as atividades usadas para se comunicar com outros participantes (OMG, 2013).

Uma **colaboração** “mostra as interações entre duas ou mais entidades de negócio”(OMG, 2013).

Uma **coreografia** é a “definição de um comportamento esperado, basicamente um contrato procedural, entre participantes interagindo”(OMG, 2013).

Este terá foco maior na descrição de processos privados e não tratará coreografias. Ele também não tem preocupação com a capacidade de executar os processos que apresenta.

5.2. Participantes

Participantes representam organizações ou papéis que participam em uma colaboração. São normalmente responsáveis pela execução de um processo dentro de uma pool (OMG, 2013).

5.3. Símbolos básicos

BPMN representa processos como uma sequência de tarefas, por meio de um fluxo de controle herdeiro do conceito de fluxograma, controlado por *gateways* e eventos.

Para isso são usados quatro 4 símbolos básicos:

1. **atividades**, representadas por caixas de bordas arredondadas;
2. **eventos**, representados por círculos;
3. **gateways**, representados por losangos, e
4. **fluxos**, representados por setas.

Além disso possui alguns símbolos adicionais para representar objetos de dados, armazenamento de dados, e mensagens, anotações e grupos. Também é possível criar novos símbolos representando artefatos, contanto que não sejam similares aos já existentes. Finalmente, processos podem ser organizados em *pools* e *raias*.

5.4. Atividades

As **atividades** são o trabalho que as organizações fazem. Se incluídas em um processo, passam a ser chamadas de **tarefas** ou **subprocessos**. Um processo é composto por um fluxo de atividades que podem ser genéricas ou possuir significado específico. Elas são representadas por caixas de cantos arredondadas e possuem variações descritas pela espessura da caixa, por símbolos no canto superior esquerdo e por símbolos na sua parte de baixo.

Atividades podem ser na verdade chamadas a outros processos ou tarefas (*call activities*) e representam um ponto onde processos ou tarefas globais, como sub-processos, são usados dentro de um processo. Elas são identificados por ter a borda mais grossa.

Atividades são **sempre** nomeadas no formato:

<verbo no infinitivo><objeto>

como em “coletar resíduos” ou “aprovar pedido”.

Tarefas podem ser mais detalhadamente especificadas, por uma marca no canto superior esquerdo, como (OMG, 2013):

- tarefa de envio, onde é enviada uma mensagem a um participante externo;

5. BPMN

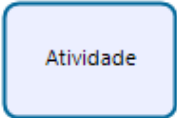
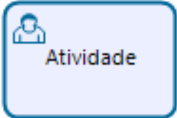
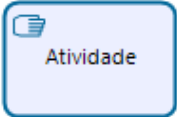
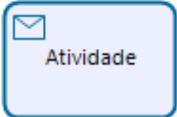
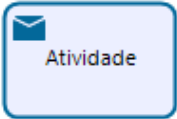
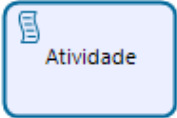
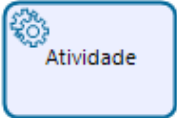
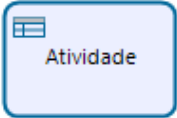
- tarefa de recepção, que espera a chegada de uma mensagem de um participante externo, muitas vezes usada para iniciar um processo;
- tarefa de *script*, que é executado por um *engine* de processo de negócio;
- tarefa de regra de negócio, que indica o uso de um *engine* que executa regras de negócio;
- tarefa de serviço, que usam um tipo de serviço, como um *web service* ou uma aplicação automatizadas;
- tarefa manual, feitas de forma totalmente manual, ou
- tarefa de usuário, feitas por uma pessoa com auxílio de uma aplicação.

Além disso, marcadores na base de tarefas e processos podem indicar (OMG, 2013):

- um subprocesso colapsado, por meio de uma caixa com um símbolo de soma;
- um *loop*, por meio de uma seta circular;
- múltiplas instâncias em sequência, por meio de 3 barras horizontais;
- múltiplas instâncias em paralelo, por meio de 3 barras verticais;
- uma compensação, por meio de dois triângulo ligados voltados para a esquerda, e
- *ad-hoc*, por meio de um til, apenas para subprocessos;

Também existem símbolos para subprocessos.

Tabela 5.1.: Símbolos para os eventos que podem aparecer no início de um processo.

Significado	Símbolo	Explicação
Simple		atividade simples
Usuário		atividade feita por um usuário
Manual		atividade feita manualmente
Recebimento		atividade que recebe
Envia		atividade que envia
Script		atividade feita por um script
Serviço		atividade feita por um serviço
Regra de Negócio		atividade que segue regra de negócio

5.5. Eventos

Eventos aparecem em três momentos:

1. **iniciais**, que indicam que um processo vai iniciar, e tem a borda simples;
2. **intermediários**, que afetam o fluxo de processo, e tem uma borda dupla, e
3. **fim**, que determinam onde um processo termina, que tem uma borda grossa.

Os eventos podem estar no fluxo do processo ou ligados diretamente a uma atividade. Nesse caso ele podem interromper ou não a atividade.

O círculo do evento possui duas possíveis bordas e um marcador interno, que pode ser preenchido ou não preenchido. As bordas dão indicação da localização do evento no processo. No caso de um evento ligado a uma tarefa, as bordas normais indicam que o evento interrompe a atividade, e bordas tracejadas indicam que o evento não interrompe.

Eventos podem ser sem tipo, ou especificados detalhadamente em subtipos. Existem várias subespecificações de eventos, mas nem todas podem ser aplicadas a todos os momentos em que o evento pode aparecer (OMG, 2013):

- **cancelamento**, reagindo a cancelamentos ou ativando cancelamentos;
- **compensação**, ativando ou tratando compensações;
- **condição**, reagindo a mudanças em condições do negócio ou relacionadas com as regras de negócio;
- **erro**, capturando ou lançando erros com nomes, semelhantes a exceções em linguagens de programação;
- **escalamento**, indicando a passagem do processo para um nível maior de responsabilidade;
- **ligação**, simplesmente ligando dois pontos como conectores, por exemplo em páginas diferentes;
- **mensagem**, indicando o recebimento ou envio;
- **múltiplo**, capturando um de vários eventos;
- **múltiplo e paralelo**, capturando vários eventos;
- **sinal**, enviados e recebidos através de vários processos e possivelmente tratados múltiplas vezes;
- **temporização**, indicando eventos cíclicos, momentos específicos, prazos e períodos ou faixas de tempo, e
- **terminação**, ativando a terminação final e imediata do processo.

É importante diferenciar um sinal de uma mensagem. Mensagens sempre tem claramente definidas quem as enviou e quem as recebe. Um sinal pode ser recebido por mais de um processo.








O subtipo é especificado por um desenho no interior do círculo que representa o evento. Se o desenho for escuro ou preenchido, significa que o evento está lançando seu sinal, seu erro, ou outra coisa. Se o símbolo for vazado, desenhado apenas com linhas, significa que

está capturando.

5.5.1. Eventos Iniciais

Os eventos iniciais possíveis são apenas o simples e mais seis tipos, como listado na tabela 5.2. Esses eventos nunca emitem nada, só capturam mensagens, sinais e múltiplos eventos de início, ou seja, respondem a gatilhos (OMG, 2013).










Tabela 5.2.: Símbolos para os eventos que podem aparecer no início de um processo.

Significado	Início	Explicação
Simple		evento simples de início
Mensagem		processo se inicia quando recebe uma mensagem
Temporizador		processo se inicia em um certo momento, ou uma data específica, ou outra questão relacionada a tempo, como um prazo
Condicional		processo se inicia quando uma condição ou regra de negócio se torna verdade
Sinal		processo se inicia com um sinal recebido
Múltiplo		processo se inicia quando um de vários eventos acontecem, bastando um acontecer
Múltiplo Paralelo		processo se inicia quando vários eventos acontecem

5.5.2. Eventos Finais

Os eventos finais podem ser simples ou de outros 8 subtipos. Eventos finais nunca recebem ou capturam, sempre enviam algo, ou seja, eles criam resultados (OMG, 2013). A Tabela 5.3 mostra os tipos de eventos possíveis no final de um fluxo de processo. É importante notar que os eventos normalmente só terminam o seu fluxo de processo, a não ser o evento de terminação que termina todos os fluxos.













Tabela 5.3.: Símbolos para os eventos que podem aparecer no fim de um fluxo de processo.

Significado	Início	Explicação
Simple		evento simples de fim
Mensagem		processo termina enviando uma mensagem
Sinal		processo termina enviando um sinal
Múltiplo		processo termina enviando vários eventos
Compensação		processo termina e lança compensação
Escalamento		processo termina e escala para nível superior
Erro		processo termina e avisa de erro
Cancelamento		processo termina e cancela
Terminação		processo termina completamente

5.5.3. Eventos Intermediários

Os eventos intermediários são o simples e mais 13, sendo que 7 recebem mensagens ou esperam algo acontecer e 6 enviam alguma comunicação para outros eventos. Em geral eles podem responder a gatilhos ou gerar resultados (OMG, 2013). A Tabela 5.4 mostra todos os eventos intermediários possíveis.

Tabela 5.4.: Eventos que podem aparecer no meio de um processo.

Significado	Símbolo de Recepção	Símbolo de Envio	Explicação
Simples			evento simples intermediário
Tempo			evento ativado por questão temporal
Mensagem			processo recebe ou envia uma mensagem
Sinal			processo recebe ou envia um sinal
Condição			condição acontece
Múltiplo			processo recebe ou envia vários eventos
Múltiplo			processo recebe vários eventos paralelos
Compensação			processo envia compensação
Escalamento			processo escala para nível superior

5.6. Gateways








Gateways servem para indicar o controle do fluxo do processo e são indicadas por um losango. No diagrama, elas podem dividir um fluxo em vários ou juntar vários fluxos em um único (OMG, 2013).

Um *gateway*, como mostra também a Tabela 5.5 pode ser de 6 tipos (OMG, 2013):

1. exclusivo;
2. inclusivo;
3. paralelo;
4. baseado em evento;
5. exclusivo baseado em eventos
6. paralelo baseado em eventos, e
7. complexo.

Um **gateway exclusivo** representa uma decisão de seguir um entre vários caminhos, ou de chegar a um lugar por um caminho entre muitos. É desenhada como um losango vazio, ou com um x dentro dele, conhecido como *gateway* exclusivo. O fluxo de saída pode ser marcado com uma barra diagonal para dizer que é o caminho *default*. Podemos associar o significado do *gateway* com uma instrução de *if* ou *case* de uma linguagem

Tabela 5.5.: Símbolos para os gateway.

Significado	Símbolo	Explicação
Exclusivo		escolhe um entre os caminhos
Inclusivo		escolhe alguns entre os caminhos
Paralelo		inicia vários fluxos simultaneamente
Baseado em eventos		inicia o fluxo do primeiro evento que acontece
Exclusivo Baseado em eventos		cada ocorrência de um evento subsequente ativa uma nova instância
Paralelo Baseado em eventos		inicia quando todos os eventos subsequentes ocorrem
Complexo		indica uma regra complexa de decisão

de programação. A Figura 5.1 mostra um gateway exclusivo sendo usado para indicar um caminho caso o pedido de crédito seja aprovado e outro caso não seja aprovado. Ela também demonstra o comportamento *default* do BPMN, que é considerar que se vários fluxos que entram em uma tarefa, apenas um precisa chegar a ela para ela iniciar¹.

Já um **gateway paralelo**, desenhado com um losango com um símbolo de uma cruz dentro dele, indica que um fluxo deve ser convertido em vários, que deverão ser seguidos em paralelo, ou vários fluxos devem ser combinados em um, sendo necessário que todos os caminhos cheguem ao *gateway* para poder continuar o processo. Eles são equivalentes as operações de *fork* ou *join*. A Figura 5.2 mostra um exemplo de uso desse *gateway*.

Temos um **gateway inclusivo** se houver um círculo dentro do losango, temos um caso onde um ou mais caminhos podem ser seguidos e todos devem ser avaliados.

O **gateway baseado em eventos**, o **gateway paralelo baseado em eventos**, e **gateway exclusivo baseado em eventos** são *gateways* bem interessantes. Todos eles depende da existência de eventos nos fluxos subsequentes.

¹Esse comportamento é o contrário do ARIS-EPC, onde todos os fluxos devem chegar a ela.

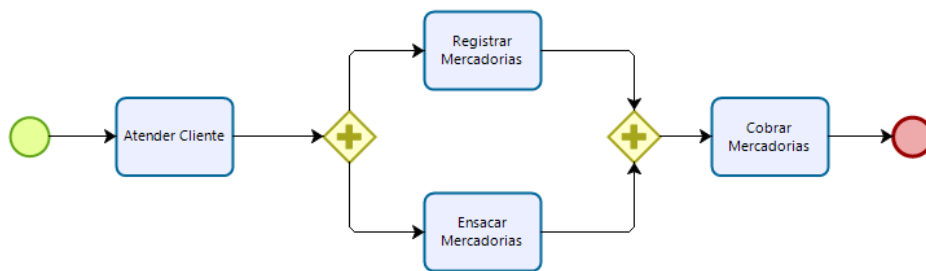


Figura 5.2.: Exemplo de uso de um evento paralelo em BPMN. Após receber o cliente em um caixa o registro de mercadorias e o ensacamento é feito em paralelo, só depois que ambos acabam se faz a cobrança ao cliente. Fonte: do autor.

Por exemplo, com o baseado em eventos, depois dele podemos ter dois fluxos, um com um evento que espera uma mensagem e outro que espera passar um prazo, indicando que a mensagem tem um prazo para chegar. Já com o paralelo baseado em eventos podemos seguir vários fluxos, cada fluxo tendo seu início quando o evento ocorre.

Finalmente o **gateway complexo**, que possui um asterisco dentro de si, significa uma comportamento complexo de condições ou sincronização.

5.7. Fluxos

Existem três tipos de fluxos (OMG, 2013):

1. **fluxo de sequência normal**, que indicam a ordem das atividades, e não iniciam em um evento intermediário, são linhas contínuas com uma seta indicando a direção;
2. **fluxo não controlado**, é um fluxo que não é afetado por condições e não passa em um *gateway*,
3. **fluxo condicional**, que pode ter uma expressão de condição que determina se vai ser seguido. Se sai de uma atividade, deve ter um pequeno losango no seu início. Caso saia de um *gateway* é uma linha comum.
4. **fluxo default**, que é usado em *gateways* exclusivos ou inclusivos e mostram o caminho *default* a ser seguido se todas as outras condições nos outros fluxos saindo desse *gateway* forem falsas.
5. **fluxo de mensagem**, que indica o envio de uma mensagem entre dois participantes, indicados por uma linha tracejada que se inicia em um círculo e acaba em uma seta aberta;
6. **fluxo de exceção**, que se inicia em um evento intermediário anexado a um processo, e
7. associações, que ligam elementos do processo a informações, representados por linhas pontilhadas, e não são fluxos de controle.

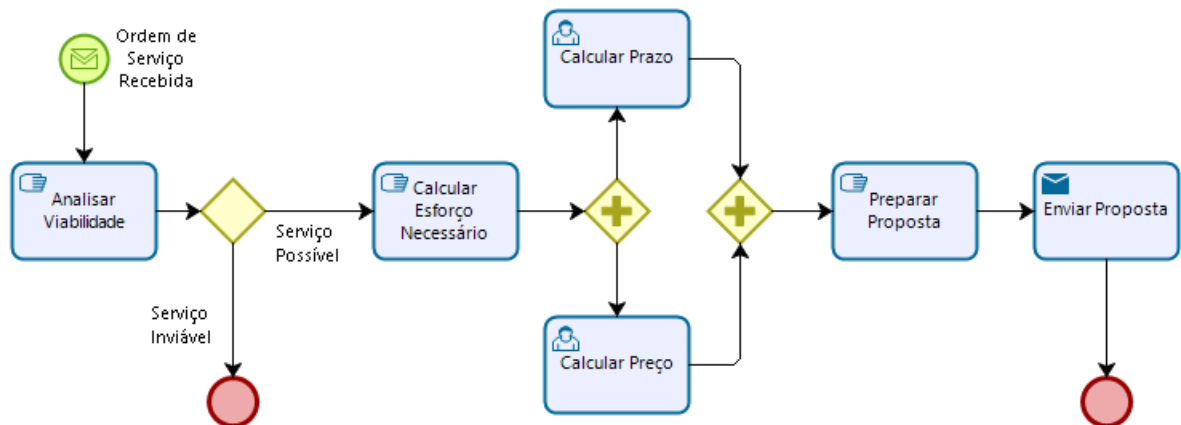


Figura 5.3.: Exemplo do atendimento da ordem de serviço em BPMN, também apresentado na Figura para Aris EPC. Fonte: do autor

5.8. Alguns Exemplos

A Figura 5.3 descreve, em BPMN, o atendimento a uma ordem de serviço. Ela representa o mesmo processo descrito em ARIS EPC na Figura ??.

5.9. Eventos Anexados a Atividades

Eventos iniciais e intermediários podem ser associados diretamente a uma atividade. Isso é feito colocando o evento sobre a borda da atividade.

Esses tipos de eventos podem interromper as atividades ou não. Caso não interrompam, seu círculo devem ser tracejados. Compensações e erros sempre interrompem a atividade.

5.10. Exercícios

Exercício 5.10-1: Vá para o site <http://jogodeanalisedesistemas.xexeo.net/> e visite a Livraria Resolve. A partir da sua visita desenhe todos os processos que você identificar, usando para isso o método ARIS-EPCe.

Exercício 5.10-2: Mostre as diferenças do que você conseguiu ou não representar com o EPCe e o BPMN.

Índice Remissivo

future value, 23
present value, 23
100-pontos, 21

Análise de Dados, 4, 5
Análise de Kano, 21

Bancos de Dados, 4

benefício, 28

Big Data, 4

BPMN, 65

atividade, 67

Gateway, 73

gateway

baseado em eventos, 74

complexo, 75

exclusivo, 73

inclusivo, 74

paralelo, 74

paralelos baseado em eventos, 74

participante, 67

processo

público, 66

processos

privados, 66

subprocesso, 67

tarefa, 67

Ciclo de Vida dos Dados, 6

Ciência de Dados, 4

colaboração, 66

Conhecimento, 4

coreografia, 66

CRISP-DM, 10

Cross Industry Standard Process for

Data Mining, 10

custo de oportunidade, 14

dados, 4, 32

Data Warehouses, 4

demanda, 15

estados, 16

descritiva, 7

desejo, 15

diagnóstica, 7

dinheiro de brinquedo, 21

fluxo condicional, 75

fluxo de exceção, 75

fluxo de mensagem, 75

fluxo de sequência normal, 75

fluxo default, 75

fluxo não controlado, 75

função utilidade, 14

Informação, 4

informação, 31

Inteligência de Negócios, 4

ÍNDICE REMISSIVO

IRACIS, 28

juro, 23

Kano, 21

markup, 13

Mineração de Dados, 4

Modelagem de negócio, 59

modelagem de processos, 60

Modelagem de Processos de Negócio, 61

montante, 23

MoSCoW, 21

necessidade, 15

pagamento, 23

prazo, 23

preditiva, 7

prescritiva, 7

prestação, 23

principal, 23

princípio da diminuição da utilidade
 marginal, 14

 processo, 60

 processo de negócio, 60

remuneração, 23

saldo, 23

satisfação, 16

Sistemas Gerenciadores de Bancos de
 Dados, 4

taxa, 23

taxa de juros, 23

Team Data Science Process Lifecycle, 10

temperatura dos dados, 9

utilidade, 14

valor, 12, 16, 27, 28

valor atual, 23

valor de resgate, 23

valor descontado, 23

valor futuro, 23

valor presente, 23

visualizações, 4

Bibliografia

- Abbasi, Asif (2021). *AWS Certified Data Analytics. Study Guide Specialty (DAS-C01) Exam*. Indianapolis, Indiana: SYBEX - John Wiley & Sons.
- Almquist, Eric, Jamie Cleghorn e Lori Sherer (2018). “The B2B Elements of Value”. *Harvard Business Review*, pp. 72–81. URL: <https://hbr.org/2018/03/the-b2b-elements-of-value> (acesso em 09/02/2020).
- Almquist, Eric, John Senior e Nicolas Bloch (1 de set. de 2016). “The Elements of Value”. *Harvard Business Review*, pp. 46–53. URL: <https://hbr.org/2016/09/the-elements-of-value> (acesso em 09/02/2020).
- American Heritage (2019). *The American Heritage Dictionary of the English Language*. URL: <https://www.ahdictionary.com/> (acesso em 25/12/2019).
- Azevedo, Ana e Manuel Filipe Santos (24 de jul. de 2008). “KDD, semma and CRISP-DM: A parallel overview”. Em: *Proceedings of the IADIS European Conference on Data Mining 2008*. Amsterdam, The Netherlands.
- BeSeen (17 de ago. de 2015). *The value triangle managing customers expectations*. BeSeen. URL: <https://www.beseen-marketing.co.uk/marketing-blog/the-value-triangle> (acesso em 09/02/2020).
- Biffl, Stefan et al., ed. (2006). *Value-Based Software Engineering*. Berlin, Heidelberg: Springer.
- Boehm, Barry W. et al. (2000). *Software Cost Estimation with Cocomo II with Cdrom*. 1ª ed. USA: Prentice Hall PTR. ISBN: 0130266922.
- Brasil (14 de ago. de 2019). *Lei Geral de Proteção de Dados Pessoais (LGPD)*.
- Brooks, Frederick P. (1978). *The Mythical Man-Month. Essays on Software Engineering*. 1ª ed. USA: Addison-Wesley Longman Publishing Co., Inc. ISBN: 0201006502.
- Cairncross, A. (1951). *Introduction to Economics*. Butterworth.
- CEOS WGISS (abr. de 2012). *Data life cycle models and concepts*. CEOS.WGISS.DSIG.TN01. Versão 13. Working group on Information Systems e Services - Committee on Earth Observation Satellites (CEOS).

- Chapman, Pete et al. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS e CRISP-DM Consortium.
- Claude E Shannon, Warren Weaver (1963). *The Mathematical Theory of Communication*. University of Illinois Press.
- Doe, Jonh (6 de jun. de 1666). *Sem Título - Precisa descobrir*.
- Erdogmus, Hakan, John Favaro e Michael Halling (2006). “Valuation of Software Initiatives Under Uncertainty: Concepts, Issues, and Techniques”. Em: *Value-Based Software Engineering*. Ed. por Stefan Biffl et al. Berlin, Heidelberg: Springer.
- Fenton, Norman E. e Shari Lawrence Pfleeger (1998). *Software Metrics: A Rigorous and Practical Approach*. 2nd. USA: PWS Publishing Co. ISBN: 0534954251.
- Franceschini, Fiorenzo (2016). *Advanced Quality Function Deployment*. CRC Press, p. 208. ISBN: 9781420025439.
- Gane, Chris P. e Trish Sarson (1979). *Structured Systems Analysis: Tools and Techniques*. 1ª ed. Prentice Hall Professional Technical Reference. ISBN: 0138545472.
- Govoni, D. L. e T. M. Gunther (2006). “Scientific Information Management at the U.S. Geological Survey: Issues, Challenges, and a Collaborative Approach to Identifying and Applying Solutions”. Em: *Geoinformatics 2006 Abstracts*. Scientific Investigations Report 2006-5201. U.S. Geological Survey. Reston, Virginia, pp. 19–20.
- Greenlaw, Steven A., David Shapiro e Timothy Taylor (2017). *Principles of Microeconomics for AP Courses*. 2ª ed. OpenStax, Rice University.
- Hayes, Adam (25 de jun. de 2019). *Internal Rate of Return IRR*. Investopedia. URL: <https://www.investopedia.com/terms/i/irr.asp> (acesso em 08/02/2020).
- Holanda Ferreira, Aurélio Buarque de (1986). *Novo Dicionário da Língua Portuguesa*. 2ª ed. Nova Fronteira.
- IBM Corp. (2015). *IBM Analytics Solution Unified Method (ASUM) - Delivery Process ASUM-DM*. URL: http://gforge.icesi.edu.co/ASUM-DM_External/index.htm#cognos.external.asum-DM_Teaser/deliveryprocesses/ASUM-DM_8A5C87D5.html (acesso em 30/01/2021).
- IIBA (2011). *Um guia para o Corpo de Conhecimento de Análise de Negócios (Guia BABOK) Version 2.0*. Ontário, Canadá: International Institute of Business Analysis.
- Kenton, Will (25 de jun. de 2019). *Net Present Value (NPV)*. Investopedia. URL: <https://www.investopedia.com/terms/n/npv.asp> (acesso em 08/02/2020).
- Kerzner, Harold (2017). *Project Management. A system approach to planning, scheduling and controlling*. 12ª ed. Hoboken, New Jersey: John Wiley & Sons.
- Kim, W. C. e R. Mauborgne (2005). *A Estratégia Do Oceano Azul*. Elsevier. ISBN: 9788535215243.
- King, J. E. e Michael McLure (2014). *History of the Concept of Value*. Discussion Paper 14.06. The University of Western Australia, Department of Economics. URL: <https://ideas.repec.org/p/uwa/wpaper/14-06.html>.
- Kotler, Philip, Gary Armstrong et al. (2017). *Principles of Marketing*. 7th European Edition. Pearsom.
- Kotler, Philip e Kevin Lana Keller (2012). *Marketing Management*. 14ª ed. Boston: Prentice Hall.
- (2013). *Administração de Marketing*. 14ª ed. São Paulo: Pearson.

- Krugman, Paul e Robin Wells (2013). *Microeconomics*. 3^a ed. New York: Worth Publishers.
- Leffingwell, Dean e Don Widrig (1999). *Managing Software Requirements: A Unified Approach*. USA: Addison-Wesley Longman Publishing Co., Inc. ISBN: 0201615932.
- Microsoft (17 de nov. de 2020). *What is the Team Data Science Process?* URL: <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview> (acesso em 30/01/2021).
- Moorman, Jan (9 de out. de 2012). *Leveraging the Kano Model for Optimal Results*. Article No :882. UX Magazine. URL: <https://uxmag.com/articles/leveraging-the-kano-model-for-optimal-results> (acesso em 08/02/2020).
- Nonaka, Ikujiro e Hirotaka Takeuchi (1995). *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. New York: Oxford University Press, p. 284.
- OMG (jan. de 2013). *Business Process Model and Notation (BPMN) Version 2.0.2*. Standard. Object Management Group. URL: <http://www.omg.org/spec/BPMN/2.0>.
- Ricardo, David (1996). *Princípios de Economia Política e Tributação*. Os Economistas. texto original de 1821. São Paulo: Editora Nova Cultural Ltda.
- Robertson, James e Suzzane Robertson (1998). *Complete Systems Analysis*. New York: Dorser House.
- Rowley, Jennifer (2006). “Where is the wisdom that we have lost in knowledge?” *Journal of Documentation* 62 (2), pp. 251–270.
- (2007). “The wisdom hierarchy: representations of the DIKW hierarchy.” *Journal of Information Science* 33.2, pp. 163–180.
- Ruble, David A. (1997). *Practical Analysis & Design for Client/Server and GUI Systems*. Upper Saddle River: Yourdon Press.
- Satpathy, Tridibesh et al. (2016). *A Guide to the SCRUM BODY OF KNOWLEDGE (SBOK Guide). A Comprehensive Guide to Deliver Projects using Scrum*. 2016^a ed. SCRUMStudy, VMEdU, Inc.
- Scheel, Henrik von et al. (2015). “Phase 1: Process Concept Evolution”. Em: *The Complete Business Process Handbook*. Elsevier, pp. 1–9. DOI: 10.1016/b978-0-12-799959-3.00001-x.
- Smith, Adam (1996). *A Riqueza das Nações. Investigação sobre sua natureza e suas causas*. Os Economistas. texto original de 1776. São Paulo: Editora Nova Cultural Ltda.
- (2003). *The Wealth of Nations*. texto original de 1854. Bantam Classics. ISBN: 0553585975.
- Strathern, Paul (2003). *Uma Breve História da Economia*. Zahar.
- Wing, Jeannette M. (1 de jul. de 2019). “The Data Life Cycle”. *Harvard Data Science Review* 1.1. DOI: 10.1162/99608f92.e26845b4. URL: <https://hdsr.mitpress.mit.edu/pub/577rq08d>.
- Zins, Chaim (fev. de 2007). “Conceptual Approaches for Defining Data, Information, and Knowledge: Research Articles”. *Journal American Society Information Science Technology* 58.4, pp. 479–493. ISSN: 1532-2882.
- (2012). *Critical Delphi Research Methodology*. URL: <http://www.success.co.il/critical-delphi.html> (acesso em 31/12/2019).

DRAFT