

USO INTELIGENTE DOS DADOS



IAPMEI



Powered by CAPACITAR PARA EMPREENDER

ÍNDICE

CAPÍTULO 1 NOTA INTRODUTÓRIA Faz-se uma breve introdução através de um breve enquadramento do que é o uso inteligente dos dados na área da Ciência dos Dados e de todas as suas subáreas, bem como as suas vantagens ao aplicar corretamente a utilização inteligente dos dados. Também são abordados os temas da informação versus conhecimento e salienta-se os aspetos principais da importância de estudar os dados.	4
CAPÍTULO 2 CIÊNCIA DOS DADOS Explica-se o conceito de Ciência dos Dados com os seus domínios e as áreas e também a interdisciplinaridade nos setores económicos. Faz-se referência à importância da análise & visualização dos dados e aos processos de aprendizagem automática. Também é apresentada a necessidade da Ciência dos Dados para a tomada de decisão.	10
CAPÍTULO 3 OS DADOS COMO INFORMAÇÃO IMPORTANTE PARA AS ORGANIZAÇÕES Aborda-se a Ciência dos Dados, a sua análise e a visualização destes como um aspeto importante a considerar para o negócio das organizações, salientando as oportunidades de negócio das empresas. Também são dados exemplos de mercado que recorrem à utilização da Ciência dos Dados e explica-se o que é um cientista de dados.	22
CAPÍTULO 4 PLATAFORMAS TECNOLÓGICAS Apresentam-se as plataformas tecnológicas e tecnologias atualmente mais utilizadas na Ciência dos Dados e também as principais plataformas de apoio são referidas.	33
CAPÍTULO 5 OPORTUNIDADES, SOLUÇÕES E TENDÊNCIAS Descreve-se quais as oportunidades de carreira para um cientista de dados e que soluções existem nas várias áreas/setores. Também são descritas as tendências na Ciência dos Dados.	39
CAPÍTULO 6 O IMPACTO DO USO INTELIGENTE DOS DADOS Aborda-se a importância da exploração inteligente dos dados, a previsão como resultado da boa aplicação da exploração dos dados e a Ciência dos Dados e a garantia de privacidade e anonimato.	44
CAPÍTULO 7 CASOS DE ESTUDO Nos casos de estudo referem-se algumas empresas de Ciência de Dados que estão a emergir em Portugal e também empresas de Ciência de Dados internacionais que estão a emergir em 2022. Refere ainda algumas fontes de financiamento para <i>startups</i> de Ciência de Dados.	50

ÍNDICE	
GLOSSÁRIO	61
REFERÊNCIAS	64
ÍNDICE DE FOTOGRAFIAS, FIGURAS E TABELAS	67



1. NOTA INTRODUTÓRIA

Faz-se uma breve introdução através de um enquadramento do que é o uso inteligente dos dados na área da Ciência dos Dados e de todas as suas subáreas, bem como as suas vantagens ao aplicar corretamente a utilização inteligente dos dados. Também são abordados os temas da informação versus conhecimento e salienta-se os aspetos principais da importância de estudar os dados.

Fotografia de [Shubham Dhage](#) disponível em [Unsplash](#)

1.1. Enquadramento

Nos últimos anos, os termos "Ciência de Dados" e "*Big Data*" têm sido alvo de discussão entre as comunidades académicas e empresariais como sendo tópicos que poderão ajudar as empresas naquilo que é a sua adaptação estratégica às tendências destas novas metodologias. (Redman & Davenport, 2021)

A verdade é que estes temas, quando bem explorados e bem aplicados nas atuais (ou novas) soluções a nível empresarial, poderão fazer a diferença no que diz respeito ao posicionamento da empresa a nível estratégico e tecnológico, bem como a identificação de novas oportunidades, apoio na tomada de decisão, exploração de novos nichos de mercado, entre muitas outras aplicações. (Avantika Monnappa, 2022)

Neste capítulo, pretende-se fazer um enquadramento do que é o uso inteligente dos dados na área da Ciência dos Dados e de todas as suas subáreas, bem como as suas vantagens ao aplicar corretamente a utilização inteligente dos dados.

1.2. Informação vs. Conhecimento

Na História, existem Eras, Períodos e Épocas. Apesar de "ainda" não ter sido atribuído um nome para o período atual, as empresas e organizações consideram que, para fins estratégicos, que estamos na Era da Informação ou na Idade da Informação. E porquê? Porque desde o surgimento das redes e da *Internet* na década de 90, e de todas as Tecnologias de Informação (TI) aliadas a essa "nova" tecnologia, que houve uma grande quantidade e disponibilidade de informação a ser criada, digitalizada e transformada. (Tucci, 2016)

No entanto, sempre houve uma grande dificuldade entre saber a distinção entre estes dois grandes termos, muito utilizados hoje em dia: Informação e Conhecimento. O que significa ter informação e ter conhecimento? Ter informação é a mesma coisa que ter conhecimento? Não! No entanto, estes dois conceitos estão bastante interligados, podendo ser afirmado que "Não existe conhecimento sem informação!".

Um dos aspetos factuais é que tanto o conhecimento como a informação são construídos com base em dados. Os dados são a fonte de toda a informação e conhecimento que hoje em dia as empresas utilizam para todo e qualquer tipo de funções. No entanto, continuamos com o mesmo problema: O que realmente quer dizer ter "Informação" e o que realmente quer dizer ter "Conhecimento"?

Ora, a principal diferença reside apenas e simplesmente naquilo que são as definições das etapas num processo normalmente designado por "Extração de Conhecimento", sendo o "Conhecimento" a última etapa do processo e a "Informação" aquilo que é extraído depois do processamento dos dados.

A figura 1.1 apresenta um diagrama daquilo que pode exemplificar o processo de "Extração de Conhecimento".



Figura 1.1: Fases do Processo de Extração de Conhecimento.
Fonte: Adaptado de Guru99.com

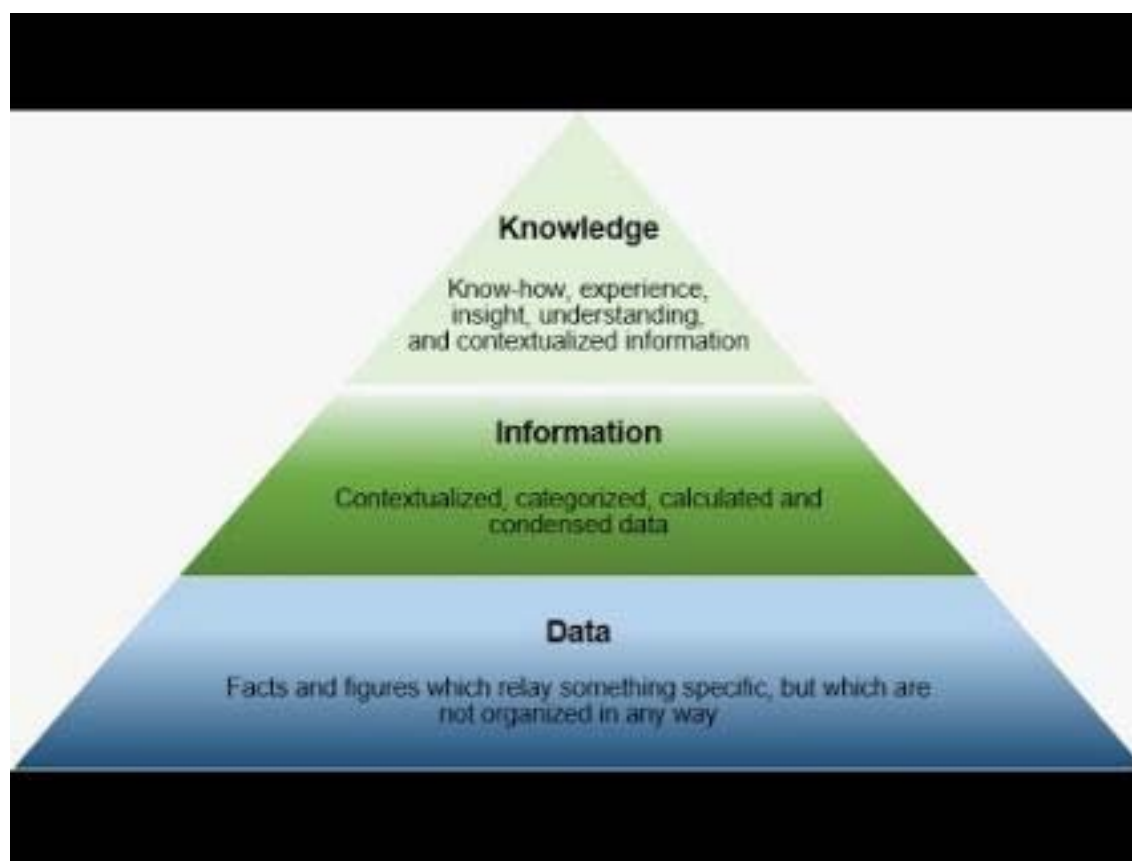
Este processo tem como objetivo, como o próprio nome indica, extrair conhecimento. Partindo dos dados, existe a necessidade de processar os mesmos para obter algum tipo de informação. Esses dados podem ser obtidos de várias formas, partindo do princípio de que são geralmente dados multimídia, como por exemplo textos, tabelas, gráficos, imagens, vídeos, que por sua vez podem ser obtidos / capturados através de documentos, bases de dados, câmaras, sensores, entre muitas outras possibilidades. Este processamento faz parte do processo de "Extração de Informação".

No entanto, esta informação necessita de ter um enquadramento, um contexto significado, um objetivo. E é através destes aspetos que se extrai o conhecimento. Podemos dizer que o conhecimento é a combinação do processamento de informação, experiência e introspeção acerca dos dados, e está diretamente relacionado com o *know-how* dos indivíduos ou da organização em si.

O Conhecimento está presente na empresa se o mesmo estiver presente nos seus colaboradores, sabendo que o mesmo provém da experiência e de toda a informação que foi processada.

A informação e conhecimento também têm uma particularidade no que toca à e partilha e transferência: A Informação é facilmente partilhável e transferível, enquanto o Conhecimento não. Para ter Conhecimento é necessário aprender, e é a grande aposta das novas soluções hoje em dia: saber que para serem melhoradas, inovadas, otimizadas e transformadas, é necessário ter Conhecimento. (Taylor, 2021)

Para completar interactivamente esta secção, sugerimos a visualização do seguinte vídeo:



Disponível em [Youtube](#)

1.3. A importância de estudar os dados

No mundo empresarial, é cada vez mais importante haver aquilo que é um entendimento e compreensão acerca dos dados que são adquiridos para os mais diversos fins. E devido à grande necessidade de recolher informações, existe uma grande quantidade de dados que são obtidos todos os dias nas empresas, e que são fruto da sua atividade diária. Mas, qual será a utilidade "extra" desses mesmos dados?

Anteriormente, muitas empresas não olhavam criticamente e estrategicamente para os dados. Apenas usavam os dados como meio de guardar informação acerca dos seus clientes, como por exemplo, as subscrições de um certo serviço. Mas hoje em dia, e devido à grande competitividade que existe no mundo empresarial, houve a necessidade de olhar para os dados de uma forma mais cuidada, e tentar perceber onde poderão melhorar, como por exemplo, responder de uma forma mais otimizada àquilo que são as necessidades dos clientes, como também estabelecer relações duradouras e de qualidade com os mesmos. (Wang, 2021)



Figura 1.2: Os conceitos base da análise dos dados.
Fonte: Adaptado de HealthMinds

Devido a esta necessidade, as empresas começaram a apostar estrategicamente em novas metodologias para estudar os seus clientes ou os seus produtos e/ou serviços. Como? Através da utilização da Ciência dos Dados e de todas as suas metodologias e tecnologias envolventes.

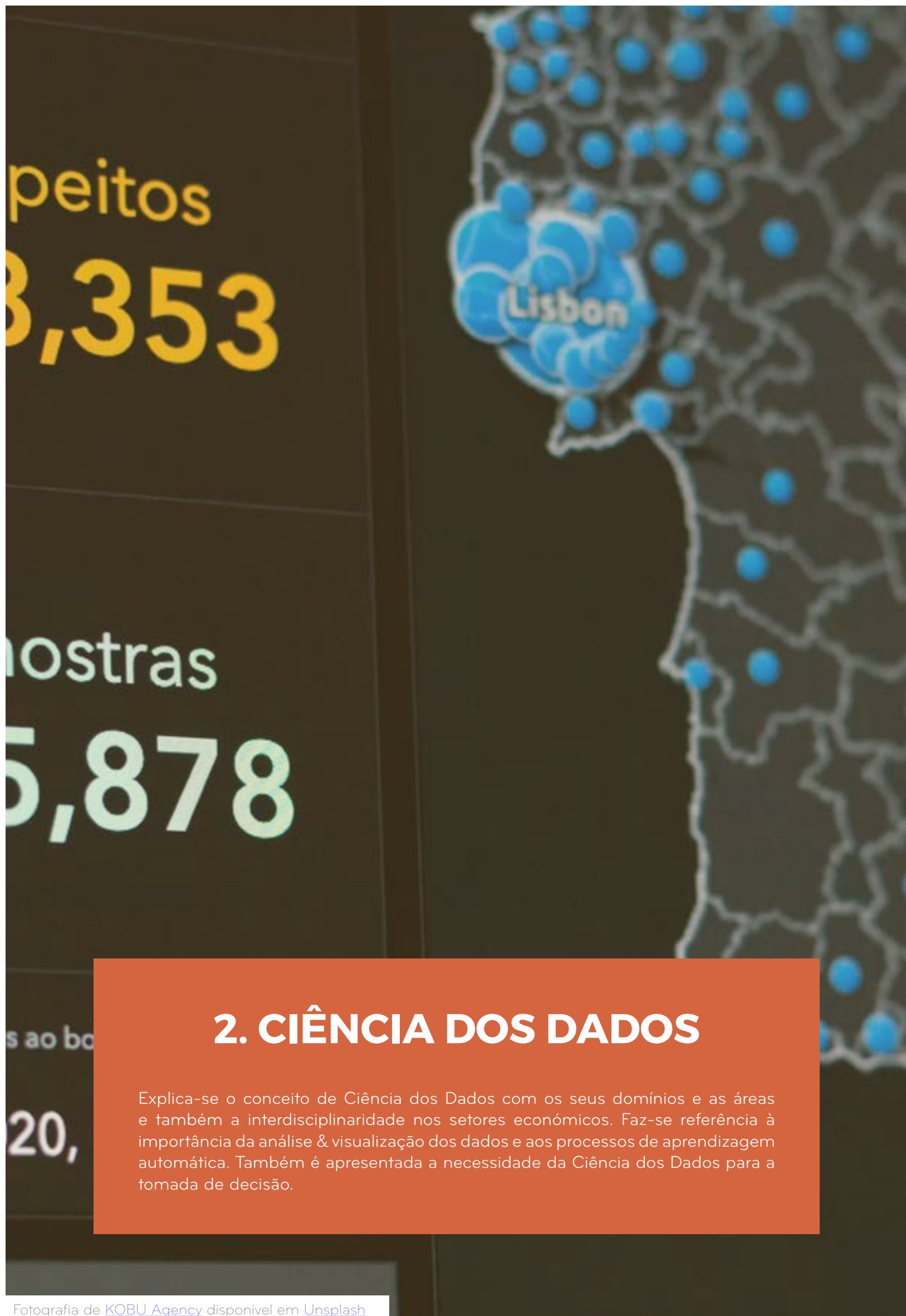
A importância de estudar os dados, ou por outras palavras, a importância da análise dos dados, assenta em diversos pilares de exploração, tais como: o desenvolvimento de novos produtos e/ou serviços; exploração de mercados, identificando aquilo que são as necessidades dos clientes e apontando soluções otimizadas para os mesmos (*targeted marketing*); otimização de processos internos e externos, que visam o melhoramento de técnicas atuais bem como atualização das mesmas.

E tudo isto tem como base os objetivos transversais a toda e qualquer empresa: mais e melhor, a custo competitivo, com qualidade, assertividade e em tempo útil. (Ohri, 2020)

Recomendamos o seguinte vídeo sobre análise de dados e a sua importância no estudo dos dados



Disponível em [Youtube](#)



2. CIÊNCIA DOS DADOS

Explica-se o conceito de Ciência dos Dados com os seus domínios e as áreas e também a interdisciplinaridade nos setores económicos. Faz-se referência à importância da análise & visualização dos dados e aos processos de aprendizagem automática. Também é apresentada a necessidade da Ciência dos Dados para a tomada de decisão.

Fotografia de [KOBU Agency](#) disponível em [Unsplash](#)

2.1. O que é a Ciência dos Dados

Sem ter conhecimento, muitos de nós já interagimos com a Ciência dos Dados de uma maneira ou outra. Hoje em dia, o simples facto de executar uma pesquisa na *Internet* ou, via *smartphone*, ver as promoções de refeições numa aplicação de pedidos *mobile* em restaurantes, é por si só uma interação com produtos que implementam a Ciência dos Dados nas suas funcionalidades.

Hoje em dia, pode-se considerar que a Ciência dos Dados está presente em muitas tarefas comuns do nosso quotidiano, e mesmo sem saber, somos alimentados constantemente com soluções que visam melhorar essas mesmas tarefas. (Igual & Seguí, 2020)

A Ciência dos Dados é muitas vezes considerada aquilo que é a combinação de várias áreas científicas tais como a matemática, a estatística, a ciência da computação, a engenharia, a programação e a inteligência artificial. Esta combinação, quando aplicada aos dados, dá origem à Ciência dos Dados. (Skiena, 2017)

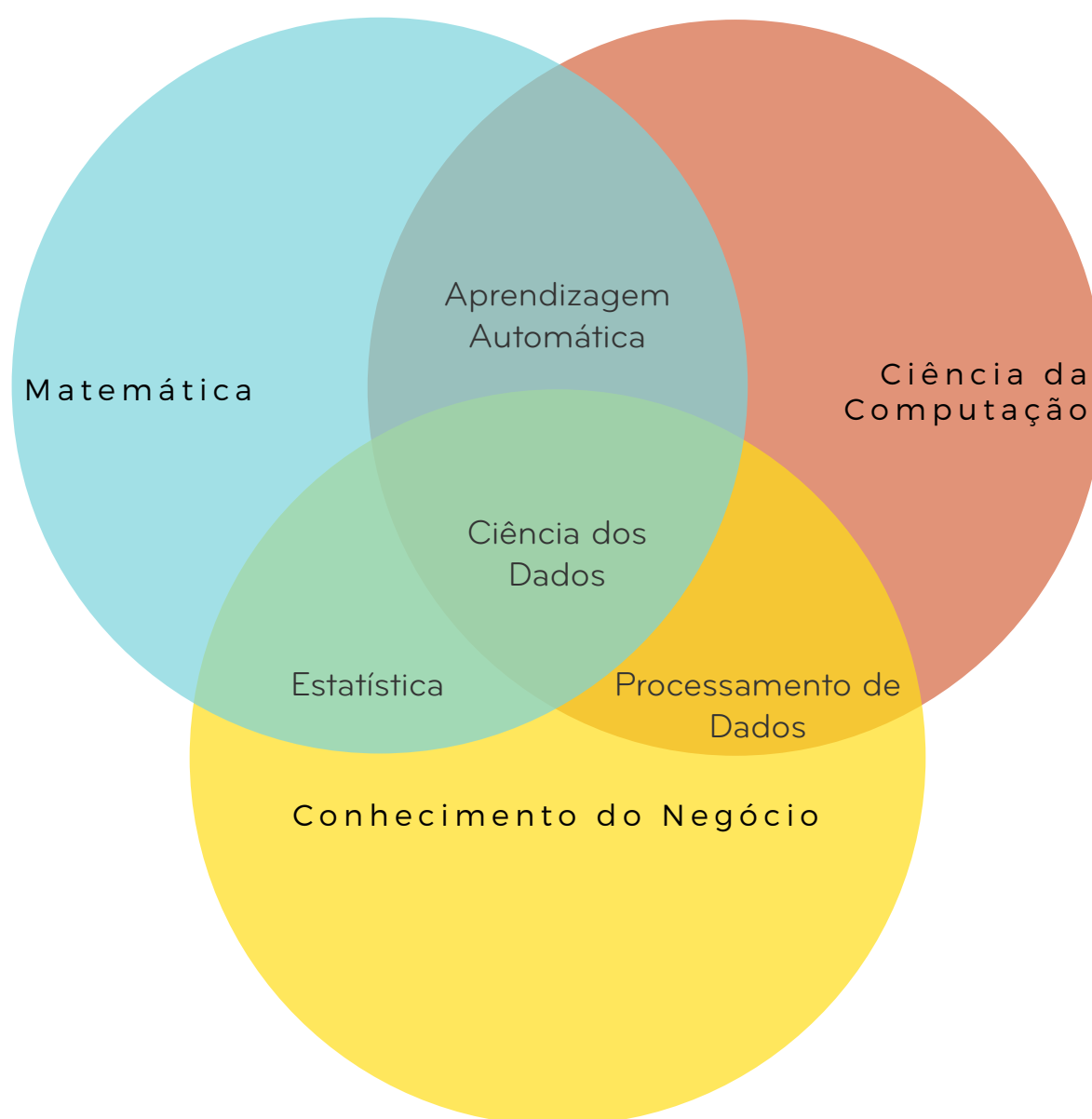


Figura 2.1: Os domínios / áreas de atuação que constituem a Ciência dos Dados.
Fonte: Adaptado de: Analance

Algumas destes domínios já são bem antigos, como é o caso da matemática e da estatística. As áreas da ciência da computação e do *Machine Learning* são, logicamente, as mais recentes. Mas é também devido a estas áreas que houve uma grande evolução a nível tecnológico nas outras áreas. No entanto, se estas áreas já existem há tanto tempo, qual é a razão de serem uma moda apenas nestes últimos anos?

Esta moda deve-se aos avanços tecnológicos provocados pelo drástico aumento do poder computacional e da quantidade de dados disponível (*Big Data*). Anteriormente, apenas as grandes empresas com poder económico tinham os meios para executar aquilo que era a exploração de dados com as ferramentas da altura, que utilizavam métodos de extração de informação tradicionais. Hoje em dia, devido à massificação da informação, e com o aparecimento de novas tecnologias de exploração dos dados, tornou-se bastante mais acessível e possível de extrair conhecimento com base nos dados.

Em suma, a Ciência dos Dados é a área que explora os dados utilizando metodologias de processamento, análise, visualização e compreensão dos dados para tirar conclusões que ajudam na tomada de decisão. (Das, 2016; Schutt & O'Neil, 2013)

Aconselhamos a visualização do seguinte vídeo e sobre aquilo que é a desmistificação da Ciência dos Dados.



Disponível em [Youtube](#)

2.3. Ciência dos Dados e a sua Natureza Interdisciplinar

A Ciência dos Dados, por si só, não possui um contexto específico. Não tem uma área de atuação específica. Não tem um domínio de aplicação específico. É errado considerar que apenas se pode implementar metodologias baseadas em Ciência dos Dados em áreas tais como, por exemplo, a medicina ou informática.

A Ciência dos Dados é uma área versátil, flexível, adaptável, que tanto pode ser aplicada às áreas ditas mais comuns como a ciência da computação e a medicina, como também às mais variadas áreas tais como as ciências sociais, engenharias, economia, ciências da saúde, naturais, entre outras. (Anand, 2021)

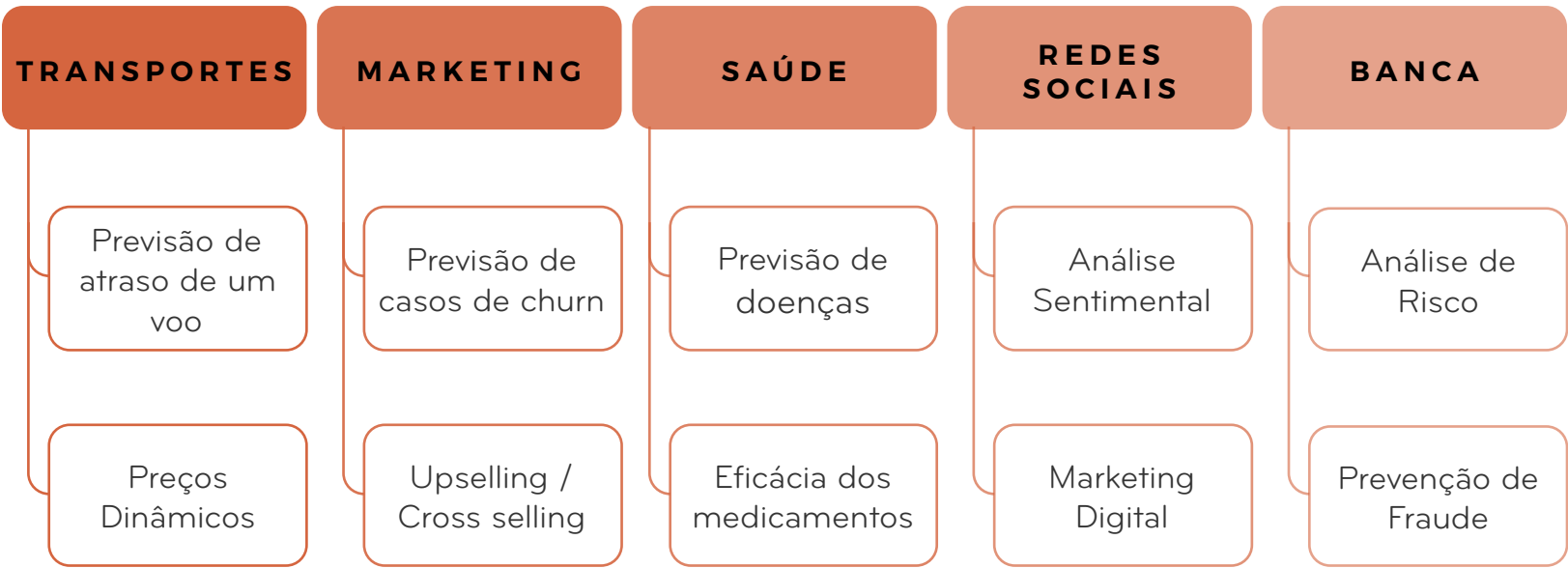


Figura 2.2: Exemplos de áreas dos setores económicos onde são aplicadas as metodologias suportadas por Ciência dos Dados.
Fonte: Adaptado de LucidInsights

Sempre que o âmbito seja explorar os dados com a finalidade de extrair conhecimento, de modo a otimizar processos, inovar soluções e melhorar relações com o cliente, então podemos aplicar Ciência dos Dados. Não existe uma fronteira entre as limitações da Ciência dos Dados e o seu uso real. Se houver dados, podemos aplicar estas metodologias.

Por exemplo, empresas como o **Facebook** utilizam a Ciência dos Dados para direcionar publicidade baseada naquilo que é o nosso perfil de interesses, capturando aquilo que são as nossas preferências, estando constantemente a analisar o nosso comportamento dentro da plataforma, bem como o tipo de conteúdo que visualizamos ou publicamos.

A **Amazon**, uma das maiores lojas *online*, utiliza metodologias suportadas em Ciência dos Dados para melhorar a experiência do utilizador. É através destas metodologias que implementam o seu sistema de recomendação de modo a realizar análises preditivas acerca dos seus produtos, fazendo recomendações com base no perfil consumidor do cliente, melhorando a satisfação do cliente.

A **Uber**, a empresa responsável pela aplicação *mobile* com o mesmo nome, implementa estas metodologias para otimizar as rotas de deslocação dos veículos tendo em conta as condições de trânsito atuais, bem como as condições meteorológicas, entre outros tipos de perturbações. Também é com base na Ciência dos Dados que oferece o melhor preço aos seus clientes, baseado naquilo que é a procura de veículos e oferta de condutores no momento.

A **Airbnb**, considerada como uma das "maiores empresas de alojamento, mas que não possui qualquer alojamento em sua posse", também implementa a Ciência dos Dados naquilo que é a sua estratégia de apresentação dos resultados de pesquisa por um alojamento. Com base nas preferências do cliente, apresenta os melhores resultados para esse cliente em primeiro lugar, de modo a aumentar as vendas.

Em 2014, a **Airbnb** descobriu que muitos utilizadores acabavam por abandonar a página por acederem aos resultados de alojamentos "nas redondezas de um certo destino". Também em 2014, a **Airbnb** lançou uma nova versão da plataforma que, ao invés de apresentar os alojamentos nas redondezas desse destino, apresenta os alojamentos com melhor reputação (*top travel destinations*), melhorando em 10% as vendas totais (Data Flair Tutorials, 2021)

De modo a apresentar mais exemplos da aplicação interdisciplinar da Ciência dos Dados, recomendamos a visualização do seguinte vídeo que refere aquilo que são as Top 20 aplicações de Ciência dos Dados no mundo empresarial.



2.3. A Análise & Visualização como motor para a retirada de conclusões

Quando se está perante grandes quantidades de dados, por vezes, não existe um método simples para saber se a informação está pronta a ser processada, e se a mesma foi ou não, efetivamente, bem processada. Qualquer cientista de dados necessita de verificar e validar todos os dados antes e depois de serem processados de modo a garantir que a informação que cria e/ou transmite representa exatamente aquilo que os dados são, de modo a evitar más interpretações e/ou conclusões acerca dos mesmos.

Sabendo que existe uma grande responsabilidade em garantir a fiabilidade e veracidade da representação dos nossos dados, que mecanismos existem para que realmente seja possível confirmar se a informação processada vai ao encontro do esperado? Que aspeto têm realmente os nossos dados, e que conclusões podemos retirar dos mesmos? (Skiena, 2017)

É através de ferramentas de visualização, das mais variadas formas, que se torna possível “olharmos” literalmente para a forma dos dados, de modo a realizar a Análise Exploratória de Dados, que é a análise mais comum e mais utilizada para visualizar informações sobre os nossos dados.

É nesta análise que se identificam as variáveis de interesse e se efetua estudos estatísticos sobre a mesma (médias, correlações, variâncias, relações entre variáveis, extração de características, etc.). É também nesta análise que se apresentam sob forma de gráficos (*plots*) os estudos anteriormente mencionados.

Uma das grandes vantagens de se recorrer a mecanismos de visualização de dados é a deteção de erros. Quando já se está perante a visualização daquilo que são os nossos dados, é também através da análise dos mesmos que se possíveis erros nos dados, que podem ter originado na própria fonte (*source* dos dados) ou do resultante da sua manipulação / transformação.

É extremamente importante garantir que os nossos dados, quando processados, não resultam em informações falaciosas ou erros que podem causar problemas nos processos seguintes, por exemplo, no processamento através de algoritmos suportados por Aprendizagem Automática (*Machine Learning*).

A visualização também se torna importante quando queremos transmitir os nossos dados, a informação e conhecimento extraído dos mesmos. O processo de "comunicação" dos dados, muito frequente em elaboração de resultados, conclusões e posteriores relatórios, quando recorre a mecanismos de visualização torna a partilha de informações para terceiros bastante mais intuitiva



Figura 2.3: Pontos de invasão terrestre na Ucrânia. Créditos Infografia: Jaime Figueiredo
Fonte: *Expresso.pt*, Data: 25/02/2022

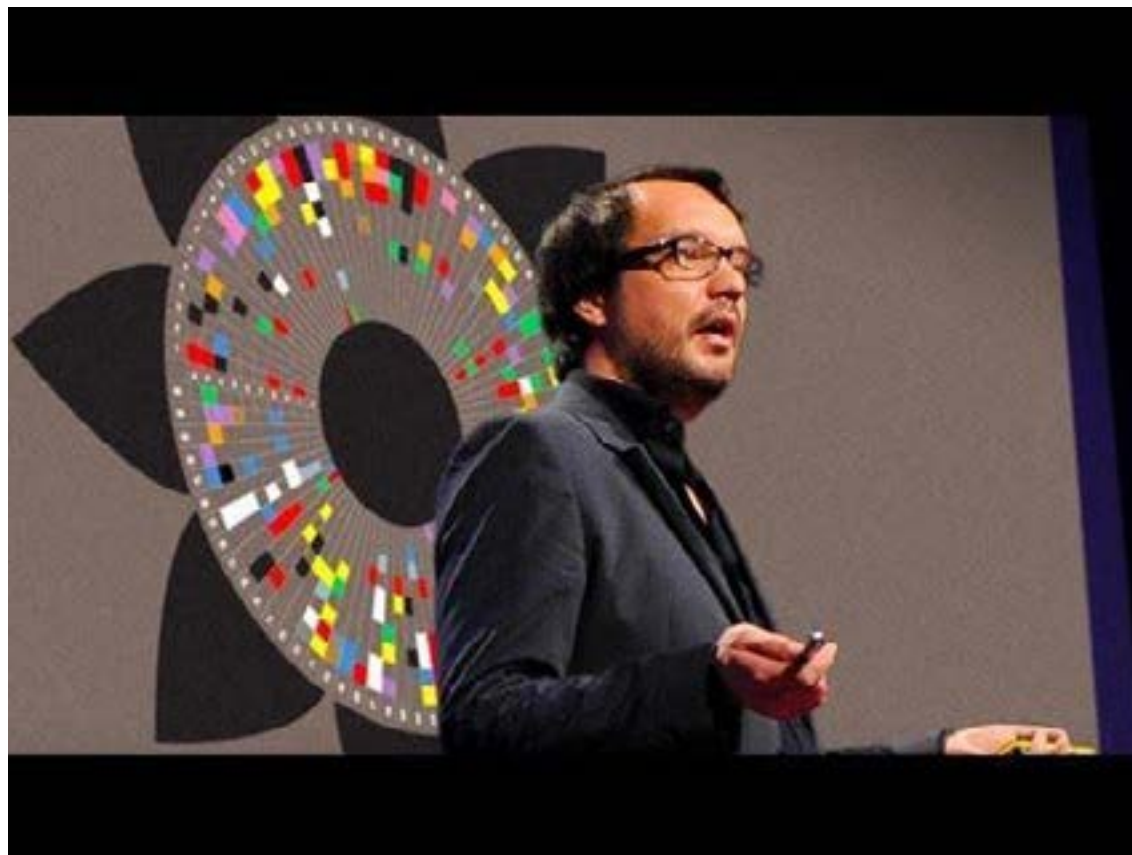
e compreensível. (Grus, 2019)

Tomamos como exemplo a Figura 5. No presente ano, a Rússia invadiu a Ucrânia via meios aéreos e terrestres. A Figura 5 representa o avanço das tropas russas sobre o leste da Ucrânia, através de meios terrestres. E, utilizando de um esquema de cores a rosa, a representação das regiões sob controlo russo, sendo que os círculos (também a rosa) representam os ataques da ofensiva russa.

Estas conclusões foram retiradas através da elaboração de um grafismo, do tipo região geográfica que, através de indicação através de cores e legendas, foi possível retirar rapidamente conclusões acerca do estado atual da invasão russa. Ora isto demonstra claramente o poder da visualização para a demonstração de resultados e retirada rápida de conclusões.

Existem muitos mais mecanismos que permitem demonstrar o poder da visualização. Tomando a máxima muito conhecida da "imagem vale mais que 1000 palavras", estes mecanismos continuam a ser das ferramentas mais importantes para a visualização de informação e extração rápida de conhecimento que permitem a retirada eficaz de conclusões acerca dos dados. (Analytiks, 2020)

Para concluir, recomenda-se a visualização do seguinte vídeo sobre aquilo que é a "beleza" da visualização de dados.



Disponível em [Youtube](#)

2.4. *Machine Learning* como ferramenta essencial na Ciência dos Dados

Hoje em dia, as empresas tendem estar a par daquilo que são as novas tendências e apostas no mercado. Uma delas é a utilização de algoritmos suportados por Aprendizagem Automática (ou, em inglês, *Machine Learning* - ML) nos seus produtos, serviços, processos internos, entre outros. Isto deve-se ao facto de, nos dias de hoje, os algoritmos de ML estarem cada vez mais otimizados, evoluídos e acessíveis a todos, tornando-os ferramentas de interesse.

No entanto, a sua aplicação depende bastante do caso de uso e dos resultados esperados, e quando não aplicado corretamente, aquilo que seria uma solução passa a ser um problema. Visto ser uma área que requer algum conhecimento técnico específico, os algoritmos suportados por ML necessitam de ser corretamente aplicados de modo a produzir resultados esperados. E aplicação correta depende de alguns fatores, sendo um deles o correto processamento dos dados de entrada (*input*).

A Ciência dos Dados, ao incorporar os domínios de atuação referidos na figura 3 do presente documento, possui a capacidade de explorar e garantir que os dados ganham a forma necessária para produzir os resultados esperados. E mesmo que não sejam esperados, produzir resultados que originem conclusões que são fulcrais na tomada de decisão.

O ML, um ramo da Inteligência Artificial, é uma área que permite, com a utilização de algoritmos específicos, produzir modelos de previsão. Estes modelos de previsão são depois utilizados para criar ferramentas de suporte à tomada de decisão. É com estas ferramentas que as empresas hoje em dia otimizam os seus produtos e/ou serviços, bem como criam novos processos e novas soluções, sempre com o objetivo de inovar, mantendo aquilo que é o acompanhamento das tendências de mercado, tornando-se mais competitivas. (Dönmez, 2013)

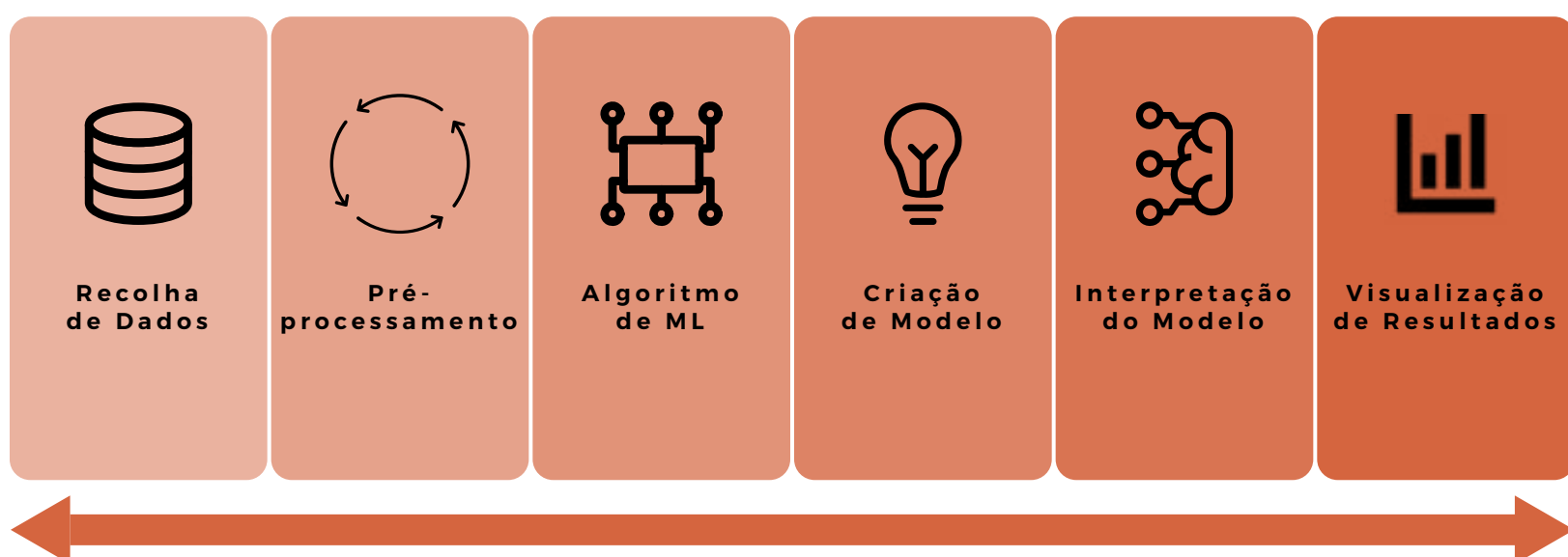


Figura 2.4: Processos da Aprendizagem Automática aplicada aos dados
Fonte: Adaptado de Medium

A Ciência dos Dados fornece a linguagem e as técnicas necessárias para a compreensão e manipulação dos dados. Envolve a concepção, recolha, análise e interpretação de dados numéricos, com o objetivo de extrair padrões e outras informações úteis. A Aprendizagem Automática, que está intimamente relacionada com a Ciência dos Dados, lida com a concepção de algoritmos e recursos informáticos que aprendem com os dados.

É através da utilização destes algoritmos que residem as vantagens da utilização destes mecanismos, que são essenciais para o processamento dos dados, permitindo a automatização de processos que serão o suporte para a tomada de decisão. (Sundaresan, 2011)

Para ajudar à compreensão do processo do ML integrado na Ciência dos Dados, recomendamos o seguinte vídeo:



Disponível em [Youtube](#)

2.5. Ciência dos Dados e Tomada de Decisão: O par perfeito

A tomada de decisão sempre foi um tema bastante sensível, pois uma decisão pode ser a diferença entre estar ou não um passo mais à frente, criando vantagem competitiva. Tomar uma decisão que impactue negativamente aquilo que é o *modus operandi* de uma empresa, poderá trazer consequências graves para o futuro das empresas no que diz respeito ao posicionamento estratégico e competitivo da empresa.

A Ciência dos Dados tem como objetivo final de melhorar a tomada de decisão, uma vez que é de extrema importância para o negócio, e usa uma metodologia chamada Tomada de Decisão baseada em dados (ou, em inglês DDDM – *Data-Driven Decision Making*). A DDDM refere-se à prática de tomada de decisão a partir dos dados baseando-se na análise de dados e não apenas por observação ou intuição.

Por exemplo, um *marketeer* poderia direcionar os seus anúncios com base apenas na sua longa experiência e o seu "olho" para decidir sobre a sua estratégia. Ou então, poderia basear a sua seleção na análise de dados relativos à forma como os consumidores reagem aos diferentes anúncios.

Big Data é um termo que se refere às enormes quantidades de dados - tanto estruturados como não estruturados - que inundam diariamente as empresas. Este tipo de dados pode ser difícil de tratar devido à sua elevada dimensão e volume. Mas o que as empresas fazem com os dados é o que realmente importa; não é simplesmente a variedade ou quantidade de dados que é significativa; é o que as mesmas fazem com isso. A análise de grandes quantidades de dados pode produzir percepções que ajudam a melhorar a tomada de decisões e dar confiança quando se realizam ações empresariais importantes. E quando aplicado à análise de dados, o **Big Data Analytics** é o processo de analisar esses mesmos grandes volumes de dados para descobrir padrões ocultos, correlações, e outras percepções anteriormente desconhecidos. Com a tecnologia hoje disponível, é possível analisar dados e extrair respostas dos mesmos quase instantaneamente. Este é um processo que seria mais longo e menos eficiente se fosse feito utilizando ferramentas mais tradicionais de inteligência empresarial.



Figura 2.5: Tomada de Decisão baseada em Dados (DDM).

Fonte: *IdealMarketing*

Um exemplo de uma tomada de decisão baseada em dados ocorreu num dos estabelecimentos **Walmart**, nos Estados Unidos da América, quando o furacão Frances estava a caminho, atravessando as Caraíbas e ameaçando um ataque directo à costa atlântica da Flórida. Em Bentonville, os diretores executivos das lojas Walmart decidiram que a situação oferecia uma grande oportunidade para ganharem vantagem competitiva utilizando uma das suas armas mais recentes: a tecnologia preditiva.

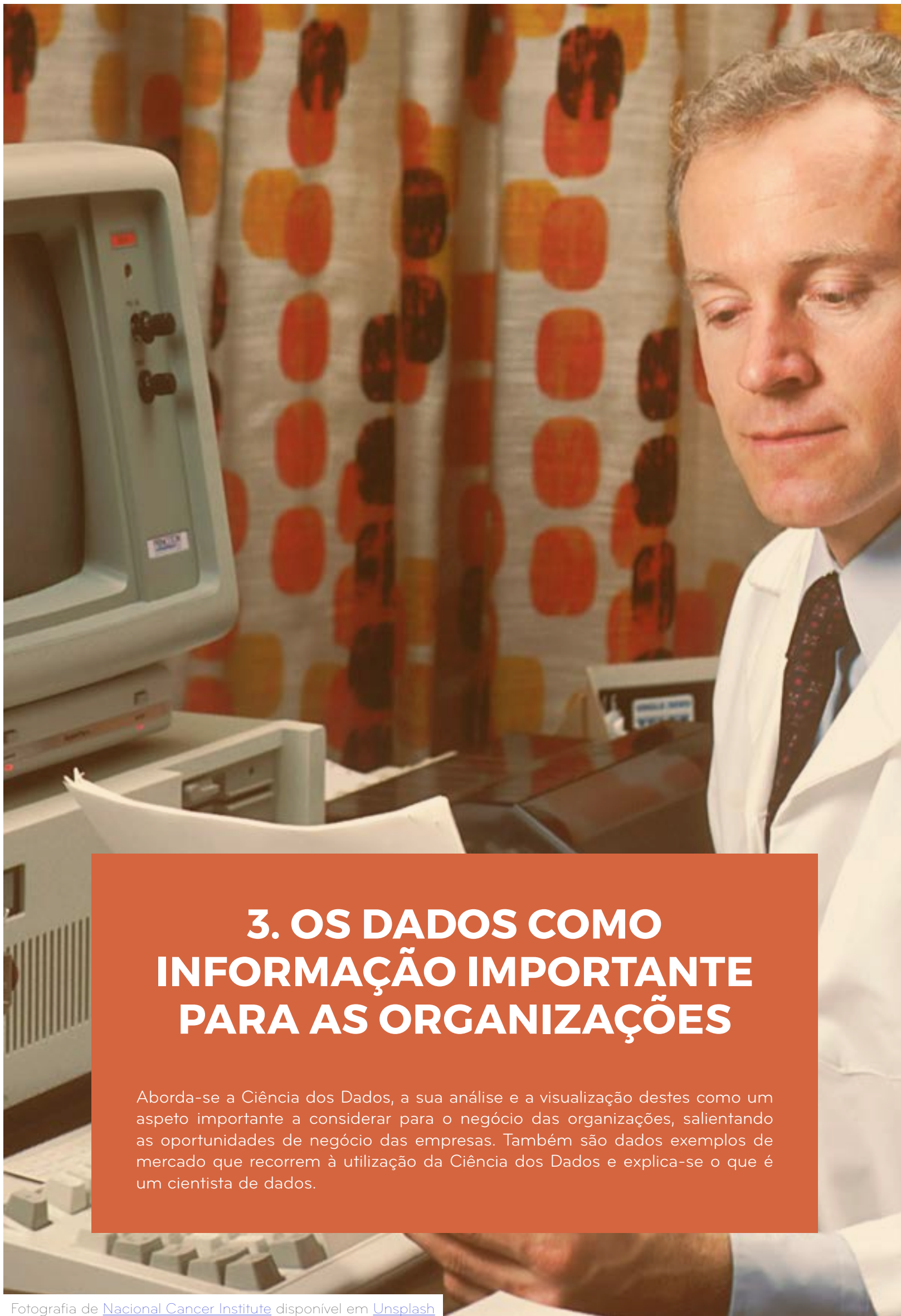
Uma semana antes da tempestade, Linda Dillman, chefe de informação do Walmart, pressionou o seu pessoal a apresentar previsões baseadas no que tinha acontecido quando o Furacão Charley atingiu várias semanas antes. Com o apoio dos triliões de *bytes* de histórico de compras que está armazenado nas bases de dados do Walmart, Dillman sentiu que a empresa poderia "*começar a prever o que vai acontecer, em vez de esperar que aconteça*" (Provost & Fawcett, 2013)

Estar "um passo a frente" é, hoje em dia, uma missão importante naquilo que é a visão das empresas e nas suas apostas de negócio. Saber que existem metodologias que, quando exploradas, permitem tomar melhores decisões, é também um sinal que existe um cuidado acrescido com aquilo que é o *fine-tuning* das decisões de negócio.

De modo a fazer uma ligação entre a tomada de decisão e o papel importante do ML, recomendamos a visualização do seguinte vídeo:



Disponível em [Youtube](#)



3. OS DADOS COMO INFORMAÇÃO IMPORTANTE PARA AS ORGANIZAÇÕES

Aborda-se a Ciência dos Dados, a sua análise e a visualização destes como um aspeto importante a considerar para o negócio das organizações, salientando as oportunidades de negócio das empresas. Também são dados exemplos de mercado que recorrem à utilização da Ciência dos Dados e explica-se o que é um cientista de dados.

Fotografia de [Nacional Cancer Institute](#) disponível em [Unsplash](#)

Com o recurso à **Ciência dos Dados** é possível transformar, acelerar e garantir a eficácia e eficiência de modelos de negócio existentes. Na base está a informação que se consegue captar, a capacidade de computação (local, distribuída e em grandes centros de dados) e a eficiência dos sistemas de comunicação, para uma análise cuidada e apoiada por algoritmos que dão suporte à otimização da decisão (figura 3.1).



Figura 3.1: Análise suportada pela Ciência dos Dados

As empresas que identificam e criam oportunidades suportadas em **Ciência dos Dados** conseguem apresentar vantagem competitiva e, conseqüentemente, transformar os seus modelos de negócios e desenvolver novos também, para um crescimento inovador. Através dos dados que conseguem (obtidos pelas várias fontes) e também com aspetos inovadores de **análise destes dados**, as empresas conseguem melhorar e otimizar a oferta de serviços e também o apoio e a relação com o cliente.

As empresas que utilizam a **Ciência dos Dados** conseguem descobrir novas oportunidades de receita, atendimento ao cliente, produtos e serviços.

À medida que a Ciência dos Dados continua a otimizar os **processos de negócio**, verifica-se uma grande mudança de foco para a descoberta do valor real com a utilização da Ciência dos Dados.

Atualmente, as empresas estão a acelerar o crescimento dos seus negócios recorrendo ao elevado **volume, valor, variedade, velocidade, e veracidade (5Vs)** dos dados (Figura 3.2).

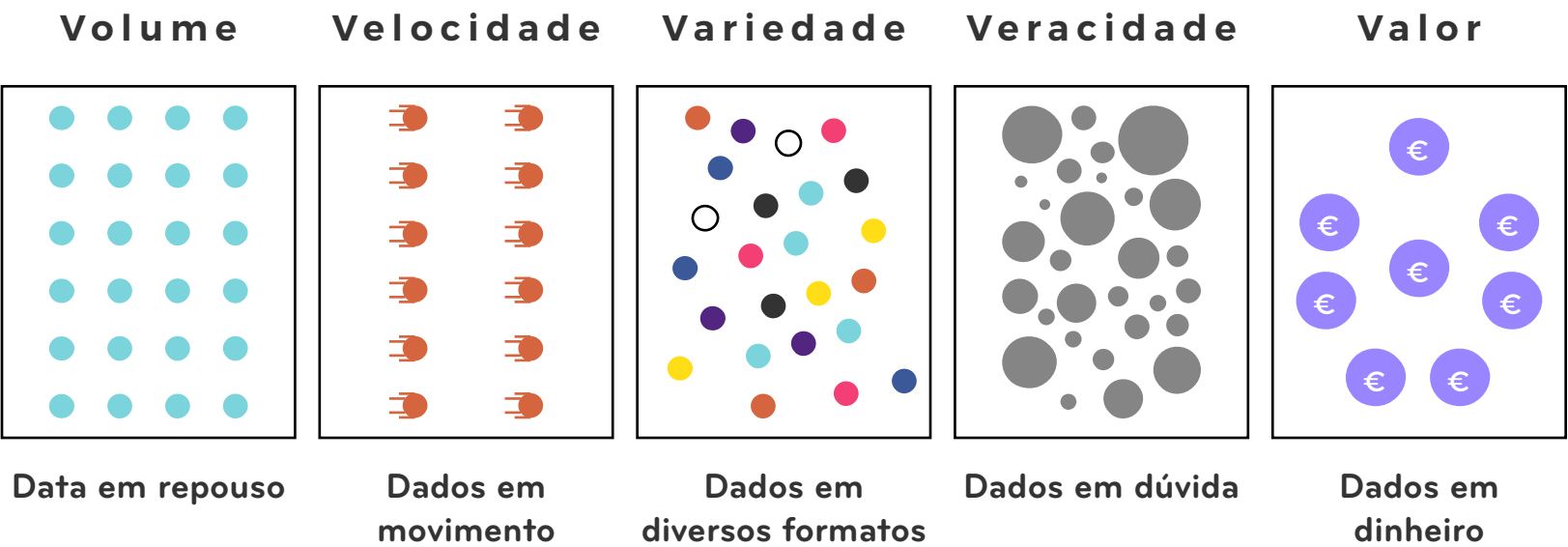


Figura 3.2: Os 5Vs do *Big Data* - (adaptado de Information Catalyst - informationcatalyst.com)

3.1. As oportunidades da ciência dos dados para o negócio das empresas

A Ciência dos Dados ainda não é totalmente conhecida pelas empresas e pelo potencial que representa. A Ciência dos Dados considera dados que são recebidos de várias fontes e, conseqüentemente em vários formatos. Os dados chegam principalmente desestruturados, sendo necessário a utilização de algoritmos para estruturar e analisar a sua informação útil. Atualmente existe poder computacional com poder de processamento para organizar os dados e melhorar a qualidade resultante da análise destes. Com o poder computacional existente (através dos avanços da eletrônica, - *hardware* e *software*) é possível obterem-se decisões que permitem ajudar a orientar no caminho e estratégia do negócio.

As Tecnologias de Computação em Nuvem, *Inteligência Artificial* e *Internet das Coisas* permite que a Ciência dos Dados possa evoluir como uma excelente ferramenta de suporte à inovação de negócios e na transformação dos atualmente existentes.

A Ciência de Dados é uma combinação de inferência de dados, algoritmos, estatística e tecnologias que combinados permitem dar um excelente suporte ao negócio, através de análise mais complexas com decisões mais eficazes. Com a Ciência dos Dados é possível analisar comportamentos, tendências e situações complexas para tomar decisões de negócios mais inteligentes. Por exemplo, os motores de busca da Google, Bing, Yahoo, etc. recorrem a algoritmos de Ciência de Dados para mostrar os principais resultados da pesquisa (quase em tempo real).

3.2. A importância da visualização de dados para o negócio das empresas

A **visualização de dados** é muito importante para o negócio. Através de uma visualização de dados altamente eficaz facilita a análise dos decisores das organizações, direções ou departamentos. Com os volumes de informação atual seria impossível uma análise para uma decisão adequada. A visualização permite às organizações públicas e privadas a terem a informação adequada resultante da complexidade utilizada para a apresentação destes. Através de análise avançadas e modelação visual, é possível obterem-se resultados com impacto na previsão do comportamento do consumidor e também de indicadores de desempenho chave – ou *Key Performance Indicators* (KPIs), que suportam a decisão dos negócios.

Alguns exemplos de ferramentas de referência especializadas de visualização de dados, disponíveis tanto na forma de código aberto quanto desenvolvidas de forma customizada, são apresentados na figura 3.3.

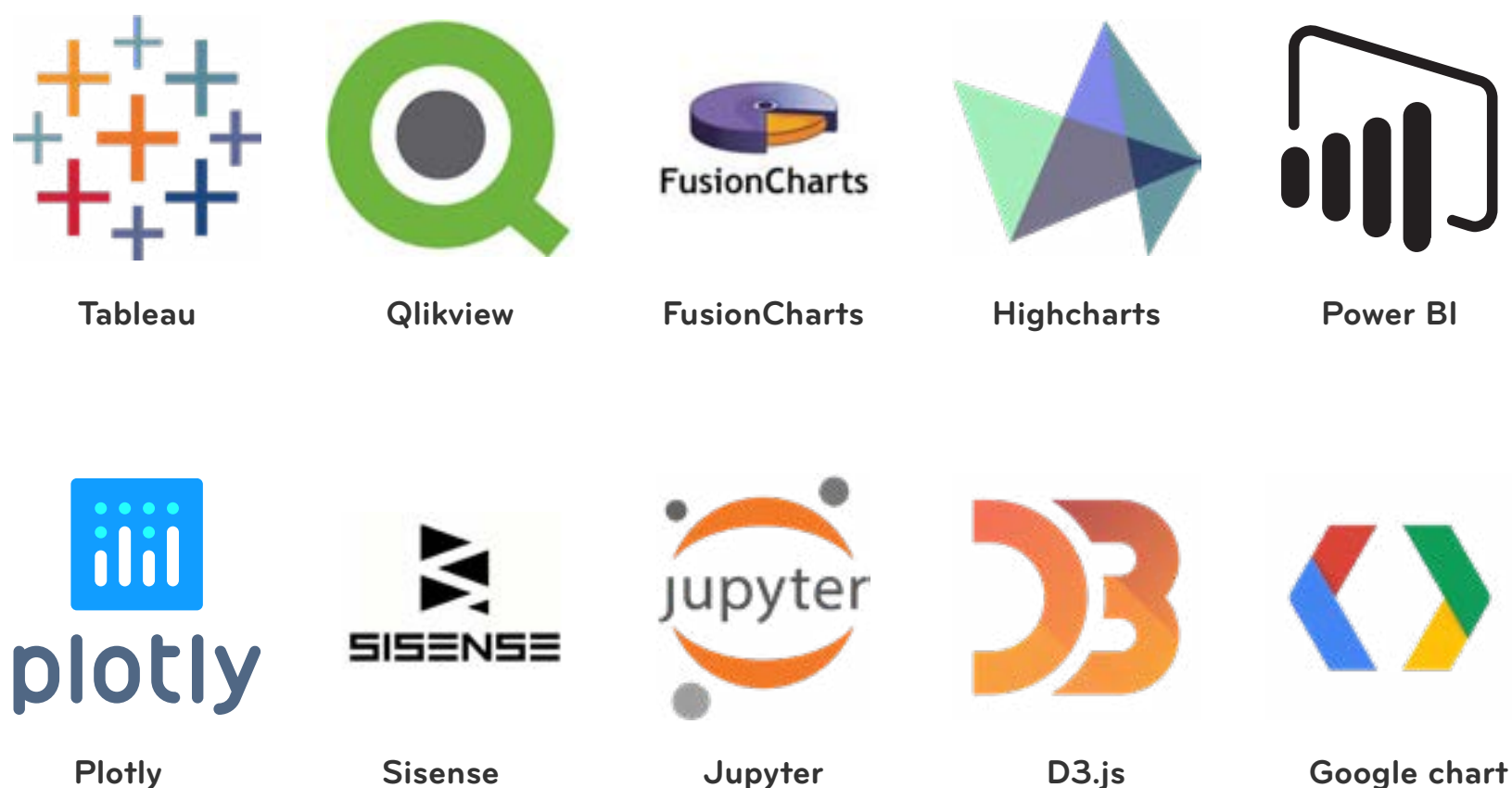


Figura 3.3: Software utilizado para a visualização de dados.

Um Painel de Visualização de Informação (em inglês, *Dashboard*) é uma interface visual que contém informações, métricas e indicadores de desempenho referentes a uma empresa ou organização. O objetivo deste painel é que estejam representados os números relevantes para a estratégia de negócio e para o alcance dos objetivos organizacionais.

Benefícios da Visualização de dados:

- i) Permite que os dados (depois de tratados) sejam apresentados em tabelas e gráficos e permite a visualização dos dados de perspectivas diferentes com um *dashboard* mais completo e com maior valor na informação apresentada;
- ii) Permite observar os dados em função da geografia (no espaço e no tempo) sob a forma de dados geolocalizados, com indicadores importantes para a decisão;
- iii) Ajuda as empresas a tomar decisões inteligentes, conseguindo "entrar" nos dados e desenvolver análises relevantes para a decisão. As empresas podem observar padrões, tendências, correlações que suportem, e.g., a diminuição de custos, o aumento da receita, as necessidades de recursos humanos, os clientes (tipo, a sua localização, etc..), falhas produtivas, ações comerciais e de crescimento e expansão do negócio;
- iv) Permite organizar os dados massivos de forma intuitiva para facilitar a análise pretendida e facilitar a decisão.

3.3. A importância da análise dos dados nos negócios

A análise de dados pode identificar o caminho para os vários ganhos de negócios, e.g., novos fluxos de receita, práticas de marketing eficientes, maior envolvimento do cliente, eficiência operacional avançada e vantagens competitivas. A análise de dados permite que os cientistas de dados possam analisar um volume de dados cada vez maior, de várias fontes, e com vários formatos que são explorados pela utilização de ferramentas convencionais de *Business Intelligence*.

Por outro lado, a análise de dados ajuda os órgãos das empresas a obterem valor com os dados que têm, de forma a tomarem melhores decisões. A análise de dados envolve aplicações complexas, e.g., modelos preditivos, algoritmos estatísticos e análise de hipóteses, com sistemas analíticos de elevado desempenho.

Os dados obtidos e a Ciência de Dados permitem descobrir novas oportunidades e a extração de informações dos dados permite melhorar produtos, operações mais eficientes, lucros mais elevados e maior satisfação por parte dos clientes. A análise de dados permite encontrar valor de várias formas, e.g., Figura 3.4.



Figura 3.4: Várias formas do valor dos dados.

3.4. Exemplos de mercados que recorrem e utilizam a Ciência de Dados

Diferentes mercados recorrem à Ciência dos Dados para melhorar a gestão e os procedimentos na otimização de processos, produtos e serviços, e.g.,:

TELECOMUNICAÇÕES

Os fornecedores de serviços de telecomunicações recorrem à análise de dados para analisar e processar milhares de documentos, imagens, manuais, dados de clientes e relatórios (em tempo real) para ajudar as suas equipas no suporte ao cliente, esclarecendo as dúvidas dos clientes com mais eficiência. Esta utilização tem impacto na economia de tempo e dinheiro, relativamente à capacidade de resolução de problemas dos clientes.

RETALHO

O atendimento ao cliente mudou muito nos últimos anos, e os compradores esperam que os retalhistas compreendam as necessidades dos clientes e no tempo adequado. Neste sentido, o retalho recorre à Ciência dos Dados para responder de forma eficaz e eficiente às necessidades dos clientes. Um exemplo disso é a utilização de análises preditivas para prever tendências, necessidades futuras e estratégias de negócios para incrementar o lucro dos retalhistas e dos seus clientes.

SAÚDE

O setor de saúde recorre à Ciência dos Dados para fazer análises em várias dimensões, e.g., os fornecedores de serviços de saúde estão a utilizar a análise de dados para extrair informações importantes do paciente. Os padrões e outras informações obtidas através dos dados ajudam os médicos a identificar pacientes com doenças crónicas e outras doenças, de forma a poder responder com tratamento em tempo oportuno e adequado para reduzir o risco.

FABRICAÇÃO

As empresas de fabrico utilizam análise de dados para ajustar e otimizar a produção, a fim de evitar excessos de *stock* e equilibrar a relação necessidade-procura. Através da análise preditiva, os fabricantes produzem produtos personalizados para responder às expectativas do cliente, reduzindo as despesas de *stock* e infraestruturas de armazenagem.

BANCA & SEGUROS

O setor da banca & seguros tem vários desafios, e.g., na forma de alerta de fraude de títulos, análise de *ticks*, detecção de fraude de cartão, arquivamento de auditoria, relatórios de risco de crédito, visibilidade comercial, etc. Com o uso da análise de dados, os bancos e outras instituições financeiras conseguem monitorizar o mercado financeiro de forma eficaz. Os principais bancos e instituições financeiras estão a desenvolver análise comercial para dar suporte à negociação, recorrendo à Ciência dos Dados..

Existem muitos outros setores, e.g., *mass media*, energia, tecnologias, organizações governamentais, moda, entretenimento, etc., que recorrem à análise de dados para otimizar os negócios, a satisfação dos clientes e, consequentemente, a sua fidelização.

3.5. Cientistas de Dados: A chave para o conhecimento nas empresas

Na última década, os cientistas de dados tornaram-se recursos humanos fundamentais e estão presentes em quase todas as organizações. Estes profissionais são indivíduos bem capacitados, com conhecimentos teóricos, técnicos e práticos de alto nível, capazes de construir algoritmos quantitativos complexos para organizar e sintetizar grandes quantidades de informação utilizada para responder a perguntas e conduzir estratégias na sua organização.

Os cientistas de dados precisam de ser curiosos e orientados para os resultados, com conhecimentos e capacidades de comunicação excepcionais, específicos da atividade onde desempenham a sua função, que lhes permitam explicar resultados altamente técnicos aos seus homólogos não técnicos. (Qian, 2021)



PROGRAMAÇÃO

As duas principais linguagens de programação usadas em Ciência dos Dados (*Data Science*) são *Python* e *R*.



MATEMÁTICA

Entender conceitos de otimização e modelagem matemática.



ESTATÍSTICA

Conseguir construir modelos estatísticos e realizar inferências (previsões).



CRIATIVIDADE

Conseguir pensar fora da caixa, propor soluções que saiam do senso comum.

Figura 3.5: Capacidades técnicas de um cientista de dados.
Fonte: *Sigmoidal.ai*

Possuem uma forte formação quantitativa em estatística, matemática e álgebra, bem como conhecimentos de programação com foco no armazenamento de dados, Prospeção de Dados (*Data Mining*) e modelação para construir e analisar algoritmos.

Os cientistas de dados experientes são responsáveis por desenvolver as melhores práticas de uma empresa, desde a limpeza ao processamento e armazenamento de dados. Trabalham de forma transversal e funcional com outras equipas em toda a sua organização, tais como *marketing*, operações, relação com o cliente, etc. É uma profissão bastante procurada e o salário reflete-se devido à grande responsabilidade das suas funções (Berkeley University, 2018).

Os cientistas de dados são aquilo que hoje em dia se considera serem a chave para o conhecimento em "primeira mão" nas empresas. São o veículo para um posicionamento estratégico de excelência, tornando as empresas mais evoluídas nos nichos de mercado que exploram.

Recomendamos a visualização do vídeo que se segue, sobre as *skills* mais importantes de um cientista de dados.



Disponível em [Youtube](#)



4. PLATAFORMAS TECNOLÓGICAS

Apresentam-se as plataformas tecnológicas e tecnologias atualmente mais utilizadas na Ciência dos Dados e também as principais plataformas de apoio.

Fotografia de [ThisisEngineering_RAEng](#) disponível em [Unsplash](#)

4.1. A utilização de plataformas tecnológicas

Um cientista de dados é considerado aquele que é responsável por extrair, manipular, pré-processar e gerar previsões a partir de dados. Para o fazer, necessita de várias ferramentas estatísticas e linguagens de programação. Um cientista de dados usa, geralmente, uma (ou várias) plataformas tecnológicas que permitem facilitar o seu trabalho ao longo dos processos.

Estas plataformas são bastante úteis pois permitem, de uma forma acessível, rápida e intuitiva, de analisar e processar os dados, sendo que algumas também permitem que, através da utilização de *dashboards*, apresentar aquilo que foi o resultado do estudo dos dados.

Tal como referido anteriormente, a Ciência dos Dados tem uma natureza interdisciplinar, podendo ser aplicada a qualquer área científica ou setor económico. No entanto, é na área tecnológica que residem as plataformas que permitem processar os dados de modo a aplicar todas as metodologias até à tomada de decisão. A próxima secção apresenta as aplicações tecnológicas que permitem implementar as metodologias inerentes à Ciência dos Dados.



Figura 4.1: Plataformas tecnológicas direccionadas para a Ciência dos Dados
Fonte: INSAID - International School of Artificial Intelligence & Data Science

4.2. Aplicações de apoio à aplicação da Ciência dos Dados

Na Ciência dos Dados, existem plataformas que recolhem, analisam, e que processam os dados. Existem também as que são um "tudo-em-um", que permitem realizar todos os processos na sua íntegra. Plataformas gratuitas, pagas e também de código aberto. Umas necessitam de experiência em linguagens de programação, outras nem tanto. A lista que se segue apresenta o top 8 de aplicações mais utilizadas para a Ciência dos Dados (DataFlair, 2021; Mujthaba et al., 2020):

[RapidMiner](#)

O RapidMiner é uma ferramenta que permite implementar ciclo de vida completo da modelação de previsão. Tem todas as funcionalidades para a preparação de dados, construção de modelos, validação e implementação.

Figura 4.2: Funcionalidades do RapidMiner

Fonte: *Edureka*

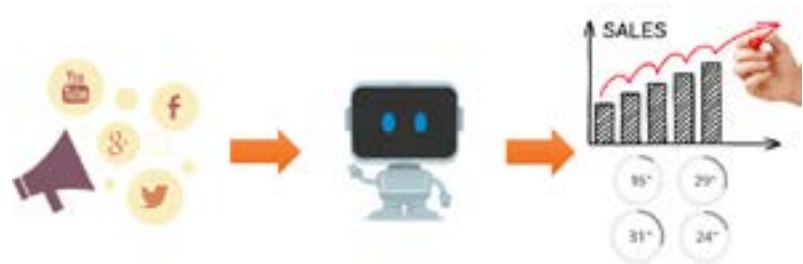


[Data Robot](#)

O Data Robot é uma plataforma que permite implementação de algoritmos baseados em aprendizagem automática. Indicado para cientistas de dados, diretores executivos, engenheiros de *software*, e profissionais de TI.

Figura 4.3: Recolha de dados de redes sociais e posterior análise utilizando a plataforma Data Robot.

Fonte: *Data Robot*



[Alteryx](#)

A Alteryx fornece uma plataforma para descobrir, preparar e analisar os dados. Ajuda também a encontrar conhecimentos mais profundos através da implementação (*deploy*) e partilha da análise em escala.

Figura 4.4: Manipulação de dados de diversas fontes no Alteryx.

Fonte: *TrueCue*



[SAS](#)

O SAS é uma ferramenta da Ciência dos Dados que é especificamente concebida para operações estatísticas. O SAS é um *software* proprietário de código fechado que é utilizado por grandes empresas para analisar dados. O SAS utiliza a linguagem de programação SAS de base que para a realização de modelação estatística.

Figura 4.5: Funcionalidades da plataforma SAS.
Fonte: Edureka



[Apache Spark](#)

Apache Spark ou simplesmente Spark é uma ferramenta analítica bastante poderosa de código aberto construída em torno da velocidade, facilidade de utilização, e análise de fluxo. É distribuído pela fundação Apache.

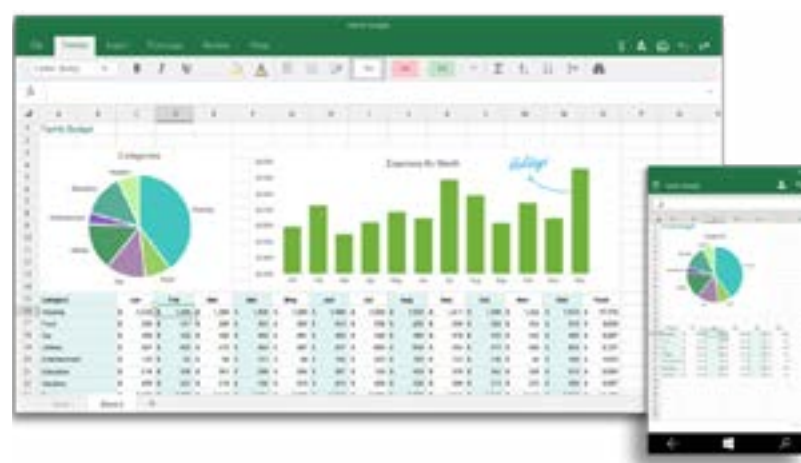
Figura 4.6: Ferramentas suportadas para integração com o Apache Spark.
Fonte: DataBricks



[Microsoft Excel](#)

O Excel é uma poderosa ferramenta de análise para a Ciência dos dados. Embora tenha sido a ferramenta tradicional para manipular folhas de cálculo, o Excel ainda mantém o seu prestígio e continua a ser a ferramenta de análise de dados mais amplamente utilizada.

Figura 4.7: Exemplo de um criado através do Microsoft Excel
Fonte: VHV.rs

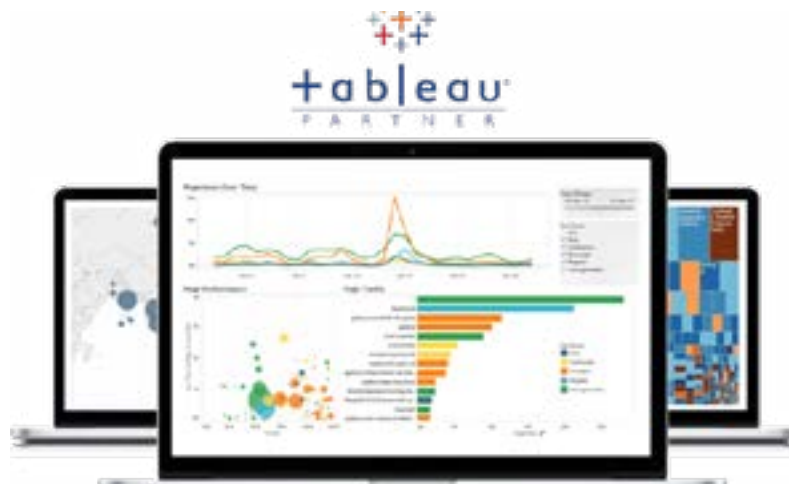


[Tableau](#)

O Tableau é um *software* de visualização de dados que possui a funcionalidade de elaborar gráficos poderosos para fazer visualizações interativas. Está centrado nas indústrias que trabalham na área da inteligência empresarial.

Figura 4.8: Exemplo de um dashboards criado através da plataforma Tableau.

Fonte: VHV.rs



[Jupyter](#)

O Jupyter é uma ferramenta amplamente popular que foi concebida para responder aos requisitos da Ciência dos Dados. É gratuita e 100% de código aberto. Ajuda os programadores a fazer *software* de código aberto e experimentar a programação interativa. O Jupyter suporta múltiplas linguagens como Julia, *Python* e R. É uma ferramenta utilizada para escrever código em tempo real e permite apresentar visualizações e conteúdo interativo.



Figura 4.9: Jupyter exemplificado na linguagem de programação *Python*

Fonte: DataCamp

[MATLAB](#)

MATLAB® é uma plataforma de programação que foi criada principalmente para engenheiros e cientistas desenvolverem e conceberem sistemas e produtos que transformam o mundo. A linguagem MATLAB está no centro do ambiente MATLAB. É uma linguagem baseada em matrizes que permite a expressão mais natural da matemática computacional.

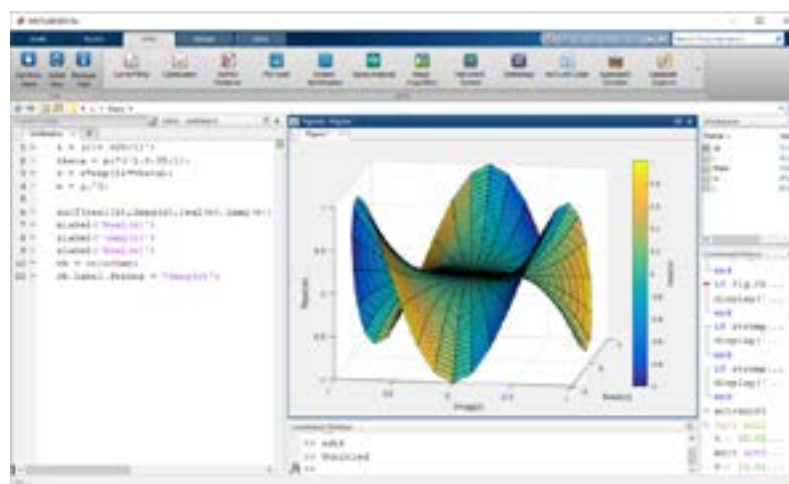


Figura 5: Fonte: Princeton Research Computing (researchcomputing.princeton.edu)

[RStudio \(Posit\)](#)

RStudio, agora denominado de Posit, é um ambiente de desenvolvimento integrado (em inglês, IDE) que suporta as linguagens de programação R e *Python*. Vem com um terminal de desenvolvimento, um editor com realce de sintaxe que permite a execução direta do código, bem como ferramentas de visualização, histórico, depuração de erros, e gestão do ambiente de trabalho. RStudio é uma aplicação *desktop* que está disponível tanto em versões *open-source* como em edições comerciais (Windows, Mac e Linux).

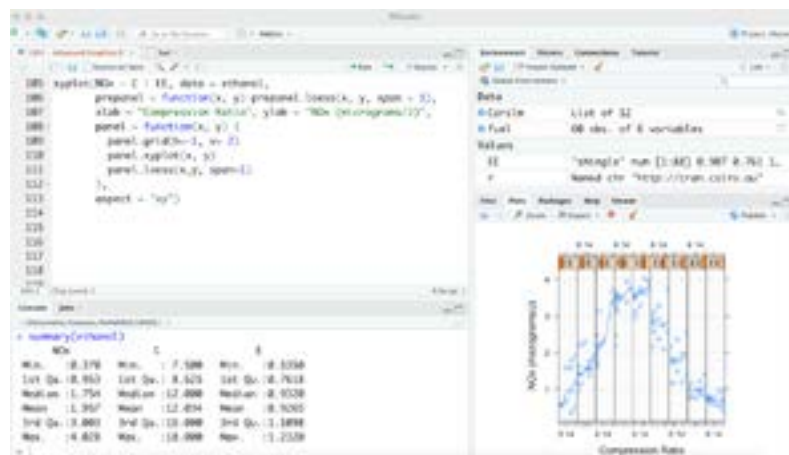


Figura 5.1: Fonte: *SolutionMetrics* (solutionmetrics.com.au)

[MySQL](#)

O MySQL é um sistema de gestão de bases de dados relacional (em inglês, RDBMS) baseado na linguagem de consulta estruturada (em inglês, SQL), criado pela Oracle. É uma tecnologia bem conhecida no contexto atual do *Big Data*. O MySQL é por vezes referido como a base de dados mais popular do mundo, e está atualmente em ampla e bem sucedida utilização em todas as indústrias. É evidente que todas as empresas envolvidas com dados comerciais ou TI em geral devem ter em conta uma compreensão fundamental do funcionamento do MySQL.



Figura 5.2: Fonte: *Simplilearn* (simplilearn.com)

[TensorFlow](#)

TensorFlow, criada pela Google, é uma ferramenta de código aberto que possibilita a implementação de algoritmos baseados em *Machine Learning*. Inclui um ecossistema amplo e flexível de ferramentas, bibliotecas, e recursos comunitários, que não só permite aos programadores conceber e implementar rapidamente aplicações impulsionadas pelo *Machine Learning*, mas também permite aos investigadores desenvolver e implementar novos algoritmos.



Figura 5.3: Fonte: *DataScientest* (datascientest.com)



5. OPORTUNIDADES, SOLUÇÕES E TENDÊNCIAS

Descreve-se quais as oportunidades de carreira para um cientista de dados e que soluções existem nas várias áreas/setores. Também são descritas as tendências na Ciência dos Dados.

Fotografia de [National Cancer Institute](#) disponível em [Unsplash](#)

5.1. Quais as principais oportunidades na Ciência dos Dados

Diariamente são produzidas elevadas quantidades de bytes de dados, pelo que se julga oportuno abordar o tema das oportunidades de carreira para a Ciência dos Dados, i.e., **Cientista de Dados**.

De acordo com uma revisão efetuada pela Harvard Business, Um Cientista de Dados "é um profissional de elevado nível com competências e capacidade para fazer análises no mundo do *Big Data*". Neste sentido, os Cientistas de Dados são profissionais altamente apetecíveis para o setor de *Big Data Analytics* e TI (edureka, 2022).

As oportunidades de carreira em Ciência de Dados são evidentes devido à escassez de profissionais qualificados que possam "trabalhar" os dados para ajudar na tomada de decisão, de forma mais eficaz e eficiente.

Verifica-se que existe uma necessidade deste tipo de profissionais para integrar os quadros de grandes organizações (privadas e públicas) mas também de *startups*, em que os seus negócios estão fortemente ligados à tomada de decisão com suporte na análise de dados.

As estatísticas apontam para um crescimento das carreiras relacionadas com a Ciência de Dados de 33,8% entre 2016 e 2026, de acordo com o *Bureau of Labor Statistics* (BLS), sendo descrito como o crescimento ocupacional mais rápido na área matemática da indústria.

Um Guia de Carreira em Ciência de Dados pode ser encontrada em *discoverdatascience* (2022), que apresenta diferentes tipos de carreiras no campo da Ciência de Dados e resume as qualidades de um Cientista de Dados.

5.2. Que soluções existem na Ciência dos Dados

Os cientistas de dados mudaram quase todos os setores. Na medicina, os seus algoritmos ajudam a prever os efeitos colaterais do paciente. No desporto, os seus modelos e métricas redefiniram o "potencial atlético". As aplicações de Ciência de Dados abordaram até o tráfego, com modelos de otimização de rotas que capturam as horas de pico típicas e as de baixo fluxo de fim de semana.

APLICAÇÕES E EXEMPLOS DE CIÊNCIA DE DADOS

- Identificação e prevenção de doenças
- Recomendações de saúde personalizadas
- Otimização de rotas de envio em tempo real
- Identificação de talentos dos próximos atletas de classe mundial
- Combate à fraude fiscal
- Posicionamento automatizado de anúncios digitais
- Algoritmos que ajudam a encontrar o amor
- Previsão de taxas de encarceramento

A **Ciência de Dados** não deve ser confundida com análise de dados. Ambos os campos são formas de entender *Big Data*, e ambos geralmente envolvem a análise de bases de dados massivos usando R e *Python*. Esses pontos de sobreposição significam que os campos geralmente são tratados como um campo, mas diferem de maneiras importantes. Por um lado, têm diferentes relações com o tempo. Os analistas de dados sintetizam *Big Data* para responder a perguntas concretas fundamentadas no passado, por exemplo, "Como nossa base de assinantes cresceu de 2016 a 2019?" Noutras palavras, extraem *Big Data* para obter percepções sobre o que já aconteceu. Enquanto isso, os cientistas de dados baseiam-se em *Big Data*, criando modelos que podem prever ou analisar o que vem a seguir. Claro que é impossível modelar perfeitamente todas as complexidades da vida real. Tal facto ficou evidenciado na famosa frase do estatístico George E.P. Box [1976]: "Todos os modelos estão errados, mas alguns são úteis". Ainda assim, a *Ciência de Dados* no seu melhor pode fazer recomendações informadas sobre as principais áreas de incerteza (Rice, 2019).

Alguns exemplos de soluções que envolvem a Ciência dos Dados são apresentados na tabela 5.1. Muitas outras soluções desenvolvidas por *startups* estão a surgir diariamente.

Tabela 5.1: Exemplos de soluções que envolvem a Ciência dos Dados.

ÁREAS/SOLUÇÃO			
1	2	3	4
Saúde			
Análise de imagens médicas	Fenotipagem de doenças	Deteção de tumores malignos	Monitorização do paciente
Viagem & Logística			
Marketing direcionado	Gestão e otimização de armazéns	Otimização de rotas	Previsão de preços de viagens e alojamento
Desporto			
Previsão de resultados em jogos	Monitorização da saúde e desempenho físico	Apoio à tomada de decisão no <i>scouting</i>	Visualização de mapas de calor dos jogadores
Governação			
Deteção de fraude e evasão a impostos	Monitorização de atividade criminal	Defesa contra ataques cibernéticos	Resposta rápida de serviços de emergência
Prevenção de Crime			
Videovigilância inteligente	Análise de perfil e historial	Processamento de Linguagem Natural	Policiaimento preditivo
Finanças			
Análise de risco	Análise de dados em tempo real	Algoritmia para <i>trading</i> em mercados financeiros	Previsão de cotações em mercados financeiros
Comércio eletrónico e Marketing			
Otimização de preço e de entregas	Sistemas de recomendação	Otimização da estratégia de produto / serviço	Previsão de cadeias de distribuição
Vida Social			
Otimização de recursos no apoio domiciliário	Algoritmia para aprendizagem adaptativa	Assistência remota e virtual	Distribuição otimizada de recursos de ajuda humanitária

5.3. Quais as tendências da Ciência dos Dados

No contexto da Ciência dos Dados as empresas pretendem adotar modelos orientados por dados para otimizar os processos de negócio e tomarem as melhores decisões com base em *insights* analíticos de dados.

Com a pandemia, as PMEs e grandes empresas tiveram de se adaptar às mudanças em menos tempo, implicando um aumento dos investimentos em Ciência de Dados.

À medida que as empresas confiam na análise de dados para superar vários desafios, verifica-se novas tendências que estão a surgir nos vários setores. As tendências de Inteligência Artificial para 2022 do Gartner são um exemplo de desenvolvimento. As tendências foram divididas em três grandes mudanças: aceleração da mudança, operacionalização do valor do negócio e distribuição de tudo (dados e *insights*).

Independentemente do setor, as empresas estão a valorizar muito a Ciência dos Dados como uma excelente ferramenta de suporte ao negócio. As tendências mais importantes para Ciência dos Dados são apresentadas na figura 5.1.



Figura 5.1: Principais tendências de Ciência dos Dados para 2022.

6. O IMPACTO DO USO INTELIGENTE DOS DADOS

Aborda-se a importância da exploração inteligente dos dados, a previsão como resultado da boa aplicação da exploração dos dados e a Ciência dos Dados e a garantia de privacidade e anonimato.

Fotografia de [Alexander Sinn](#) disponível em [Unsplash](#)

6.1. A importância da exploração inteligente dos dados

O objetivo da exploração de dados é aprender sobre as características e potenciais problemas de um conjunto de dados sem a necessidade de formular previamente suposições sobre os mesmos. Atualmente, a exploração de dados tornou-se como um requisito obrigatório em todos os projetos de Ciência dos Dados. Mas de que forma pode exploração inteligente dos dados pode trazer vantagens? (Piringer, 2020)

As técnicas de exploração manuais / tradicionais de dados podem ajudar-nos a identificar áreas específicas de interesse, o que é exequível, mas ficam aquém de uma investigação mais profunda. É aqui que o ML pode levar a análise de dados para o nível seguinte.



Figura 6.1: As etapas para uma exploração inteligente dos dados.

Fonte: Adaptado de Sisudata

Tanto a exploração de dados tradicional, como na exploração de dados suportada por algoritmos de ML, podem identificar padrões e ajudar a tirar conclusões a partir dos dados. Mas o ML permite aos cientistas de dados extrair informação em grandes bases de dados muito mais rapidamente, minimizando possíveis erros.

Com dados disponíveis, mais do que nunca, muitas empresas são confrontadas com uma abundância de dados. No entanto, as mesmas não dispõem de recursos suficientes para os analisar e processar. É aqui que entram os algoritmos de ML.

O *software* utilizado para implementar ML também podem tornar os dados muito mais fáceis de "digerir". Ao pegar em pontos específicos nos dados e exportá-los para visualizações de dados, tais como gráficos de barras ou gráficos de dispersão, as empresas podem extrair informação significativa muito rapidamente e sem gastar tempo a interpretar e a questionar os resultados. (Henn, 2021)

6.2. A previsão como resultado da boa aplicação da exploração dos dados

As previsões são também um tema sensível, pois uma boa previsão, p. ex. o estado do tempo para amanhã, irá permitir aos agricultores orientarem melhor os seus trabalhos em função do estado meteorológico. Este exemplo é outra prova que, se houver possibilidade de fazer previsões, e se as mesmas forem ao encontro da realidade, é certo que a prévia tomada de decisão irá ter resultados bastante melhores.

A Análise Preditiva enquadra-se na análise avançada, e é utilizada para fazer previsões sobre eventos desconhecidos que possam vir a desenrolar-se no futuro. A análise preditiva é a utilização dos dados em algoritmos e técnicas de ML para identificar a probabilidade de resultados futuros com base em dados históricos. O objetivo é estudar o que aconteceu, de modo a fornecer uma melhor avaliação do que irá acontecer no futuro. (SAS, 2016)



Figura 6.2: Comparação entre Aprendizagem Automática e Análise Preditiva.
Fonte: Adaptado de BMC Blogs

No entanto, um equívoco comum é dizer que a análise preditiva e a aprendizagem automática têm a mesma definição. A Aprendizagem Automática e a Análise Preditiva abordam o problema de forma diferente. Por vezes, a Análise Preditiva também se enquadra com diversas aplicações de ML. As técnicas suportadas por ML são mais adaptativas, mais recentes, e têm maiores graus de liberdade, pelo que pode ser mais flexível com a sua abordagem a um problema. A análise preditiva já existe há mais tempo e é mais processual na sua utilização. (Johnson, 2020)

As previsões são extremamente importantes para a tomada de decisão, pois beneficiam daquilo que é a experiência anterior para poder fazer correções e melhorias dos processos atuais.

Recomendamos a visualização do seguinte vídeo onde aborda os temas como a análise preditiva e as técnicas suportadas por ML.



Disponível em [Youtube](#)

6.3. A Ciência dos Dados e a garantia de privacidade e anonimato

A privacidade é também um dos temas que gera mais controvérsia na Ciência dos Dados, pois quando se lida com dados sensíveis, tem que existir uma garantia do lado das empresas sobre a segurança dos mesmos, de modo a evitar fugas de informação sensíveis, e ser potenciais alvos de crimes cibernéticos.

Existem conflitos na aplicação das metodologias de Ciência dos Dados, no que diz respeito à transparência e proteção dos dados. Sem a recolha de dados e a livre circulação da informação, a Ciência dos Dados não existiria, mas quanto mais informação se recolhe, mais complicada se torna a essa mesma proteção.

Os cientistas de dados precisam de encontrar um método para assegurar que os dados são simultaneamente seguros e acessíveis (para quando é necessária uma divulgação dos mesmos), ao mesmo tempo que precisam de tornar os dados partilháveis, e também de permitir que as pessoas manipulem a informação, se necessário.



Figura 6.3: GDPR é um regulamento da União Europeia que garante a privacidade da informação
Fonte: *Medium*

Há algumas formas de enfrentar estes desafios. A anonimização dos dados é uma delas: anonimizar os dados recolhidos e separar a informação das pessoas reais às quais se aplica. De facto, muitos regulamentos de privacidade de dados, incluindo o Regulamento Geral de Proteção de Dados (RGPD), tornaram a anonimização de dados, um requisito obrigatório. Do ponto de vista empresarial, isto não é a abordagem mais ideal, pois a anonimização de dados não é reversível. Se se tornar anónima toda a informação pessoal dos dados recolhidos de possíveis clientes, é bastante difícil depois utilizá-los no âmbito da Ciência dos Dados.

No entanto, existem métodos que garantem a integridade dos dados, ao mesmo tempo que garantem a privacidade, sendo que um deles é a segmentação de dados. Ao invés de guardar informação pessoal sobre os clientes, as empresas "aglomeram" os dados em categorias bastante amplas (grupos etários, áreas geográficas, etc.), assegurando ao mesmo tempo que os dados não podem ser convertidos novamente na sua forma granular, garantindo assim o anonimato dos dados e a preservação da identidade privada dos seus clientes. (Wieringa et al., 2021)

Para uma revisão mais detalhada sobre técnicas de garantida de privacidade e anonimato em Ciência dos Dados, aconselhamos a visualização do seguinte vídeo.



Disponível em [Youtube](#)

7. CASOS DE ESTUDO

Fotografia de [Walkator](#) disponível em [Unsplash](#)

7.1. Algumas empresas de Ciência de Dados que estão a emergir em Portugal

Alguns exemplos de *startups* e outras empresas portuguesas que já estão consolidadas e desenvolvem a sua atividade centrada na Ciência dos Dados são apresentadas como bons exemplos de Portugal com equipas dinâmicas e de valor para o mercado.

- ✎ A [Unbabel](#) é uma *startup* portuguesa que está a revolucionar o mundo da tradução *online*, entregando traduções de qualidade através de uma combinação de tecnologia de IA e reguladores de qualidade humana, de forma a fornecer serviços rápidos às empresas. Tem sede em Lisboa e São Francisco (Figura 7.1).



Figura 7.1: Unbabel na tradução *online* com a Ciência dos Dados.

- ✎ A [Uniplaces](#) tem uma plataforma *online* de aluguer de quartos, que rapidamente passou a dominar o mercado imobiliário, não só em Lisboa, mas também em várias outras capitais europeias. A simplicidade do seu site contribui muito para o sucesso da marca. A sua sede fica em Lisboa e abriu recentemente escritórios em Londres, Madrid e outras grandes cidades europeias. A figura 7.2 mostra o exemplo da plataforma.

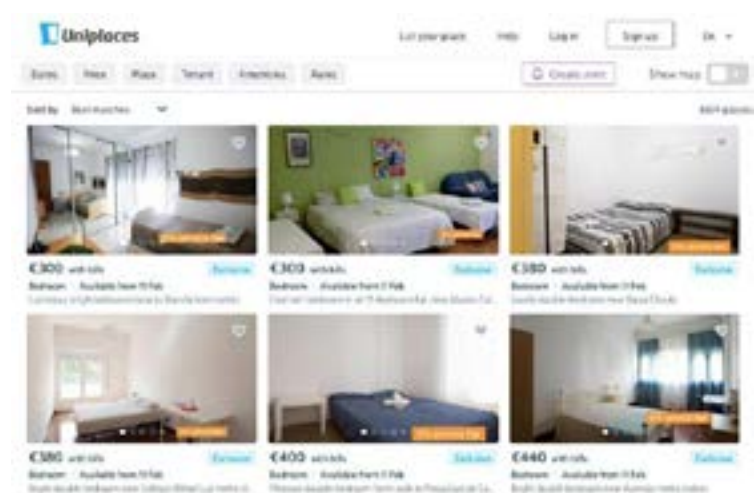


Figura 7.2: Plataforma da Uniplaces.

✍ A [Prodsmart](#), uma empresa portuguesa, obteve um financiamento de outras *startups* portuguesas que tinham necessidade de facilitar os processos produtivos das pequenas empresas. A aplicação fornece análises e gestão inteligente para a otimização da produção. Os relatórios de acompanhamento da produção consideram informações obtidas em tempo real (Figura 7.3) através de dispositivos móveis utilizados no chão de fábrica, permitindo o acompanhamento e análise (em tempo real) do fluxo de produção, e.g., pedidos, trabalhadores, materiais, máquinas, manutenção, desperdício, tempo, custos.



Figura 7.3: Análise em tempo real da Prodsmart.

✍ A [James](#), também conhecida no passado por CrowdProcess, desenvolveu uma plataforma de inteligência artificial para risco de crédito. O sistema foi projetado por e para especialistas em crédito com o objetivo de garantir que as equipas de risco tenham acesso contínuo a algoritmos utilizados na Ciência dos Dados e requisitos atualizados de validação e monitorização.

✍ A [Talkdesk](#) teve início em 2011, após os fundadores terem detetado um *gap* no mercado ao nível de *software de call center*. A empresa desenvolve *software* que permite que os utilizadores possam configurar os serviços de *call center* por meio de um processo simples e rápido. Os vários serviços são assentes na computação na nuvem.

✍ A [Defined.ai](#) considera modelos de elevado desempenho com dados de treino de algoritmos de IA a nível mundial. A Defined.ai fornece uma plataforma para treino de IA. Conta com mais de 40 clientes, e.g., Mastercard, BMW, EDP, Grupo Mello Saúde, Amazon e Sony, entre outros no EUA. Atuam ao nível da Banca & Seguros, Retalho e Telecomunicações. Atualmente, a empresa está avaliada em 35,2 milhões de euros e está presente em três continentes, com uma equipa de 140 pessoas.

✍ [Feedzai](#) teve um investimento de 74 milhões de euros e está avaliada em 521 milhões de euros. Tem como objetivo principal a segurança da atividade bancária recorrendo à inteligência artificial antifraude. É uma das *startups* de tecnologia financeira com maior potencial na Europa. Atualmente, conta com escritórios nas cidades de Lisboa, Porto, Coimbra, Silicon Valley, Nova Iorque e Reino Unido.

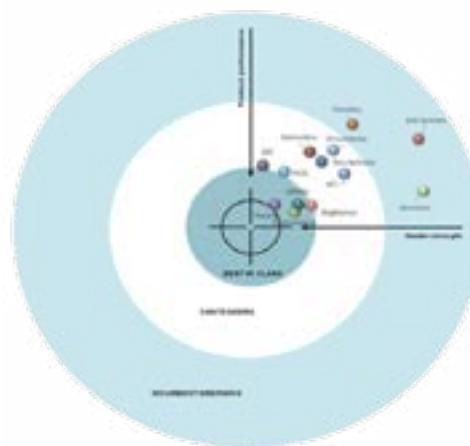


Figura 7.4: Feedzai na segurança da atividade bancária.

✎ A [Rely.ai](#) é uma *startup* no setor de inteligência artificial que disponibiliza uma plataforma para transformar a monitorização de dados em informação para ajudar a planear o negócio. O objetivo da Detech.ai é utilizar a engenharia da fiabilidade dos *sites* para melhorar a experiência do cliente. Esta *startup* foi fundada em 2021 com sede em Almada e teve um investimento de 400 mil euros distribuído por quatro investidores.

Figura 7.5: Rely.ai na melhoria da experiência do utilizador.



✎ A [Gorodata](#) é uma empresa que trabalha nas indústrias de inteligência artificial e de *software*. A empresa gere uma plataforma suportada na nuvem que envolve a inteligência artificial para classificar os dados, de forma a identificar informações relevantes para aumentar as vendas das empresas. A Gorodata oferece indicadores de compra personalizáveis para prestar um serviço às empresas para saber quando os seus serviços são necessários. A *startup* foi fundada em 2020 e tem a sede em Cascais, Portugal. Atualmente é composto por entre um e dez funcionários. Em abril de 2021, A empresa teve uma ronda de financiamento de 475 mil euros.



Figura 7.6: Gorodata plataforma de análise de vendas suportada pela Ciência dos Dados.

✎ Outras empresas como a [WeChangers](#), [Growhappy](#), [UNI2ALL](#), [Sensei](#), [Meight](#), [Tesselo](#), [Shimejito](#), [Virtuleap](#), [Visor.AI](#), [Criadam](#), [Probely](#), [UpHill](#), etc... são algumas *startups* portuguesas e com investimento que utilizam a Ciência de Dados nos seus modelos de negócio.

7.2. Empresas de Ciência de Dados internacionais que estão a emergir em 2022

As empresas globais estão na senda da utilização da Ciência de Dados. As empresas estão preparando, organizando, desenvolvendo e usando os dados com a ajuda de várias ferramentas de Ciência de Dados. As **startups emergentes de Ciência de Dados** estão a oferecer ferramentas inovadoras de última geração para a gestão e análise de grandes volumes de dados. Um conjunto de *startups* estão a adoptar várias abordagens para fornecerem soluções a empresas de vários setores. De acordo com a Analytics Insight, as seguintes *startups* são as 10 melhores que estão a emergir em 2022 (Sanyal, 2019):

- ✍ A [Cropln](#) é líder em termos do modelo *Software as a service* (SaaS) para o setor agrícola (Figura 7.7). No setor agrícola existem vários desafios e a Cropln começou a sua atividade a fornecer soluções que consideram a análise financeira, a previsão do tempo, a interpretação de dados, monitorização por satélite, *geo-tagging*, Inteligência Artificial e análise de dados. A missão da Cropln é poder rastrear o setor agrícola.



Figura 7.7: A Cropln aplica a Ciência de Dados num modelo SaaS ao setor agrícola.

✍ A [Firebolt](#) é uma *startup* que melhora as plataformas de armazenamento de dados em nuvem para análises eficazes e eficientes. Esta empresa tem uma solução SaaS que oferece um desempenho rápido de consulta, além de combinar simplicidade, elasticidade e baixo custo da nuvem, com a inovação ao nível da análise. Ele foi desenvolvido especialmente para a [Amazon AWS](#) e possui um mecanismo de consulta SQL eficiente que separa armazenamento e computação, permitindo que clientes criem vários recursos isolados na mesma base de dados (Figura 7.8).

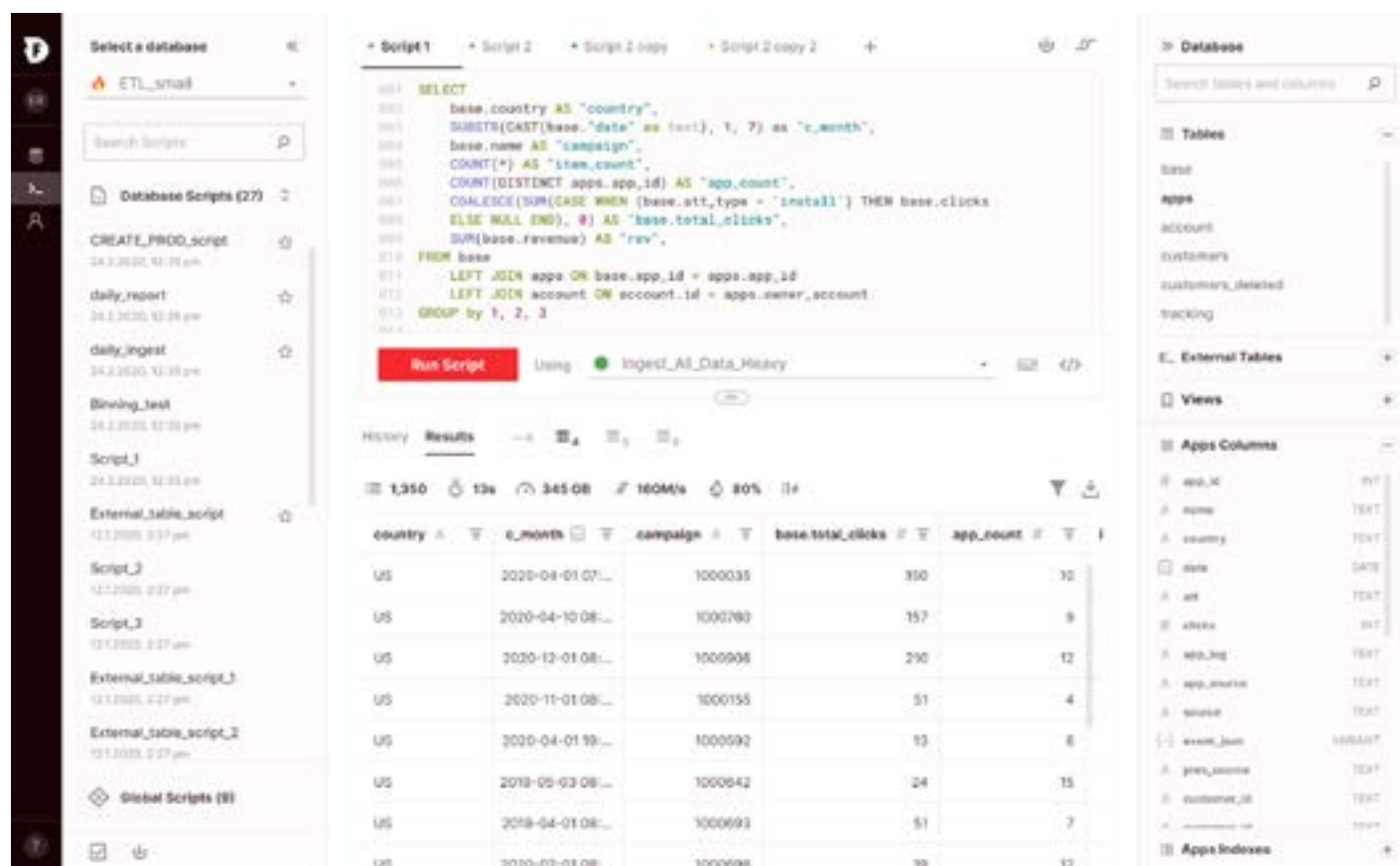


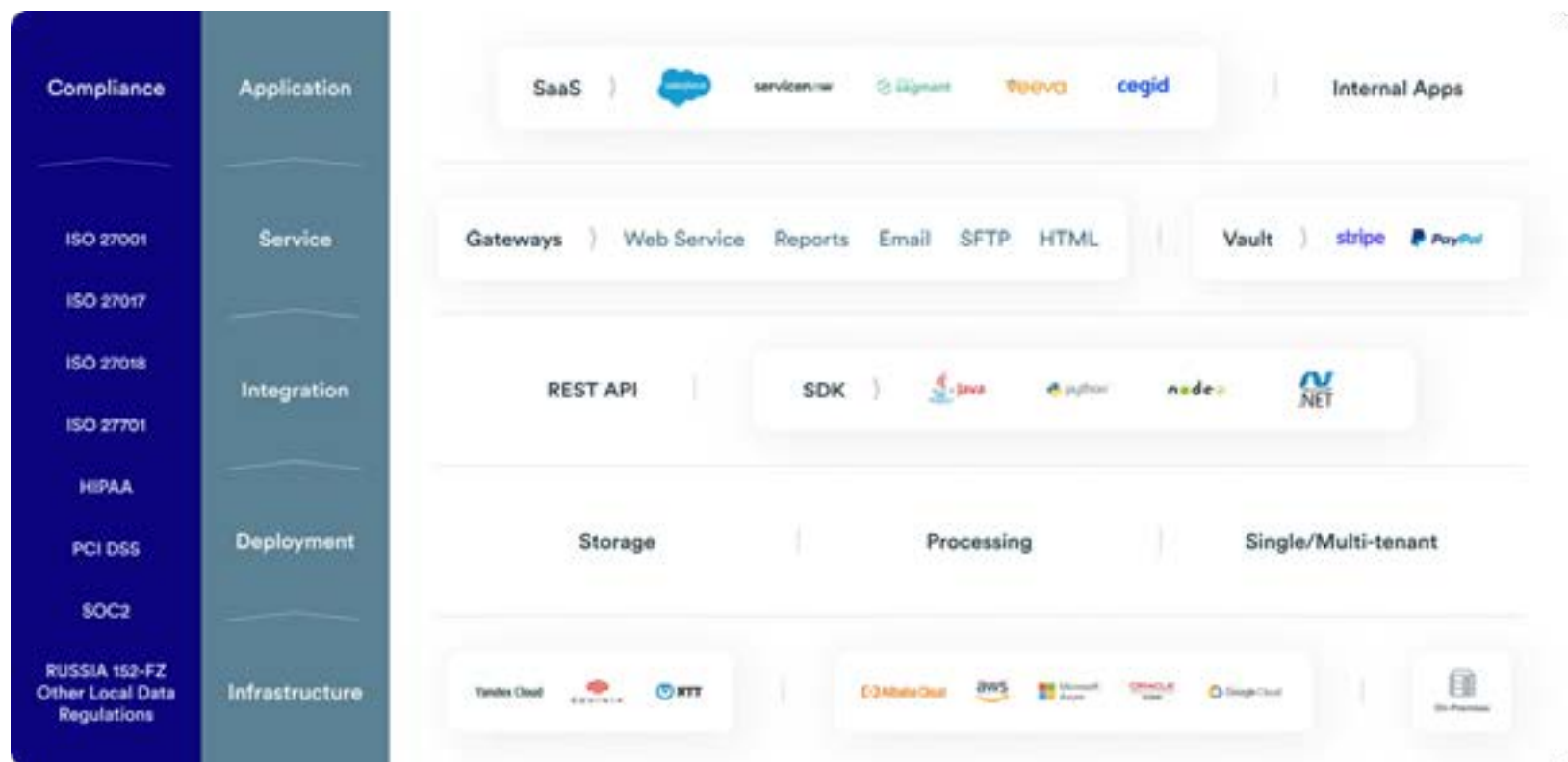
Figura 7.8: A Firebolt tem soluções SaaS de computação e armazenamento de dados.

✍ A [Lymbyc](#) é uma *startup* que tem uma equipa dedicada de cientistas e analistas de dados que desenvolveram ferramentas analíticas robustas suportadas por tecnologias avançadas. O principal objectivo desta solução avançada de ciência de dados é criar um sistema prático e otimizado de tomada de decisões para os líderes empresariais, de modo a que estes possam estruturar a previsão de negócios utilizando a inteligência empresarial através de entradas de dados utilizando o processamento de linguagem natural (PLN).



Figura 7.9: A Lymbyc tem uma solução SaaS de Análise de Dados para líderes de empresas.

🔗 A [InCountry](#) é uma plataforma de armazenamento, processamento e regulação de dados (como serviço) que permite a conformidade global de dados para negócios internacionais. A plataforma armazena, processa e regula com segurança os dados de negócios no país de origem, ajudando as multinacionais a escalar, proporcionando fácil integração em poucos minutos. A empresa foi fundada em 2019 com sede em São Francisco e opera em mais de 65 países.




 A [3LOQ Labs](#) permite que empresas e retalhistas identifiquem o comportamento de clientes individuais, com base em ações e contextos anteriores, recorrendo a algoritmos avançados de *machine learning*. A 3LOQ fortalece os sistemas de gestão financeira (dinheiro e ativos) em termos de serviços bancários digitais para que os clientes finais possam obter uma comunicação simplificada e personalizada (Figura 7.11).

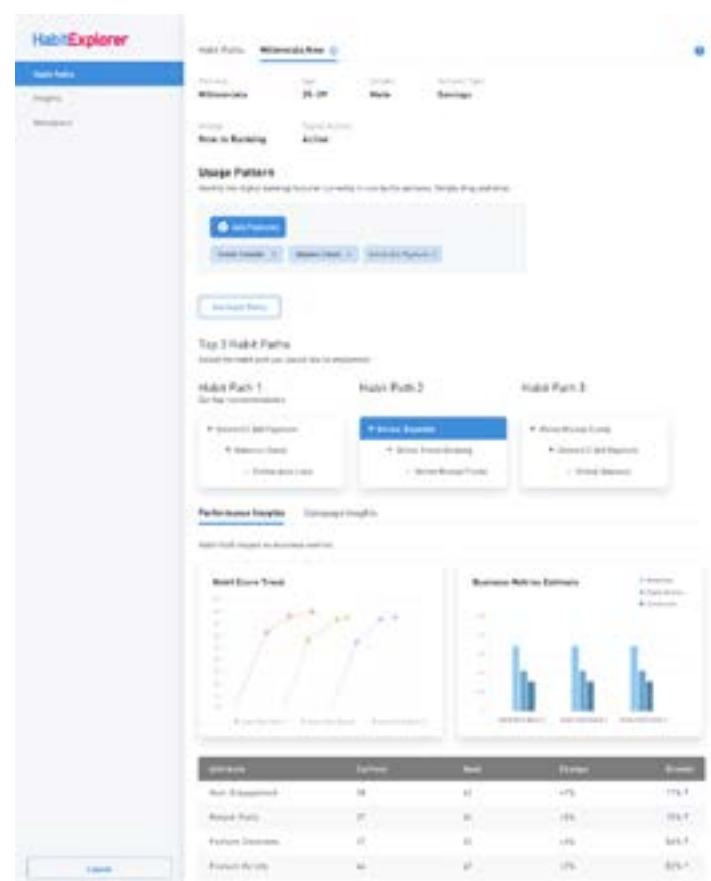


Figura 7.11: A 3LOQ Labs possui uma solução de gestão financeira inteligente.

- ✍ A [CloudNine AI](#) faz a gestão de dados orientado por IA (sem código) para empresas de saúde, removendo a carga de codificação para engenheiros que resolvem os problemas de interoperabilidade. A *startup* foi fundada em 2019 e é otimizada para análise preditiva e segurança na nuvem. A CloudNine AI tem um conjunto de serviços relacionados com a Ciência dos Dados.



Figura 7.12: A CloudNine AI possui uma solução de gestão financeira inteligente.

- ✍ A [Razorthink](#) foi estabelecida como uma empresa emergente que fornece soluções de IA. Esta empresa foi financiada por investidores individuais. A empresa visa a computação otimizada de IA, automação, modelos dinâmicos e ferramentas de Ciência de Dados. A Razorthink tem uma equipa de mais de 100 profissionais que apoiam pequenas e grandes empresas para otimizar serviços ao cliente baseados em Inteligência Artificial, gestão de risco, deteção de fraudes e outros serviços (figura 7.13).

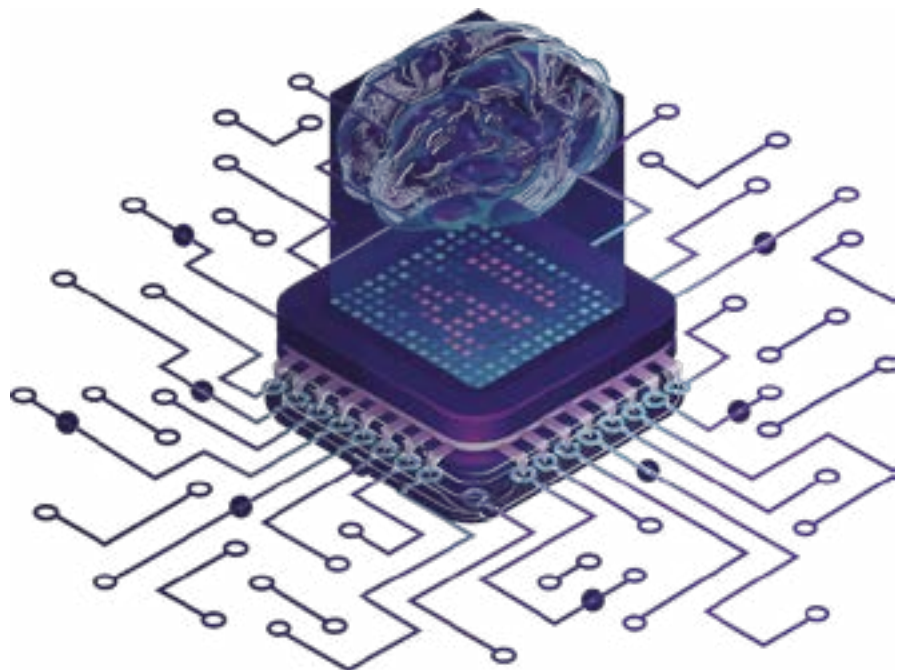


Figura 7.13: A Razorthink possui uma solução de previsão para aumentar vendas e lucros.

- ✍ A [Math Company](#) é uma *startup* que surgiu como um empreendimento colaborativo para especialistas em análise de dados. Atualmente, a empresa conta com mais de 50 funcionários e é emergente de análise de dados e do *machine learning*, visualização de dados, *design thinking* e conhecimento de domínio para dar suporte aos seus clientes.



Figura 7.15: Empresa MATH Company.

- ✎ A [Zendrive](#) utiliza os dados de sensores móveis para fornecer informações que melhoram a segurança de passageiros e motoristas em todo o mundo. A indústria global de transporte está evoluindo e a Zendrive está a liderar o caminho para a segurança de automação de transporte e disponibilização de serviço móvel.



Figura 7.16: A Zendrive tem uma Plataforma inteligente de risco de mobilidade com mitigação de risco.

- ✎ Um exemplo de outras *startups* promissoras, na área da Ciência dos Dados, são a [Avo](#), [Bigeye](#), [DataKitchen](#), [Deepnote](#), [Dhiva.ai](#) e [Hex Technologies](#) (Avidon, 2021).

7.3. Fontes de financiamento para *startups* de Ciência de Dados

✍ Para além do [IAPMEI](#), da [Portugal Ventures](#), [Caixa Capital](#), [RAIZE](#), [Previsional](#), entre outras, que investem em *startups* de diferentes formas e em diferentes estados de desenvolvimento, a capital de risco [Faber](#) lançou o Faber Tech II para investir em *startups* que têm o objetivo de acelerar a transformação digital de empresas que utilizam a Ciência de Dados nas suas várias vertentes. Este fundo tem a expectativa de obter pelo menos 30 Milhões de Euros em *fundraising* (Computerworld, 2020).

Este fundo direciona o investimento para empresas em fase inicial (pré-séries A) que operam e lideram áreas da Ciência dos Dados e com o objetivo principal da **transformação digital de empresas e indústrias a nível global**.

Em termos geográficos o fundo dedica os seus investimentos a *startups* que utilizem Portugal e Espanha como plataforma de lançamento para o mercado global, explorando o seu potencial como um dos centros tecnológicos em ascensão da Europa.

✍ A Faber já executou diversos investimentos, entre eles uma participação na ronda de financiamento Séries A de 15,6 milhões de euros feita pela [SWORD Health](#) (*startup* portuguesa que criou o primeiro sistema de fisioterapia digital e que junta equipas de Ciência de Dados com equipas clínicas), a liderança da ronda *pre-seed* da [YData](#) (uma plataforma que ajuda os cientistas de dados a melhorar a qualidade dos dados e a garantir a sua privacidade), e a participação na ronda *pre-seed* da [Emotai](#), que desenvolveu uma plataforma de análise e melhoria de desempenho para atletas de desportos virtuais, baseada em neurociência.

✍ O fundo continua a angariar investidores que pretendem apostar cada vez mais em ativos numa abordagem especializada. Além de vários investidores privados (particulares e institucionais), o Faber Tech II tem como investidores principais o [Fundo Europeu de Investimento](#) (FEI) – maior investidor institucional Europeu nesta classe de ativos – e a Instituição Financeira de Desenvolvimento (IFD) que, em conjunto, assumiram um compromisso de 15 milhões de euros ao abrigo do programa Portugal Tech.

Glossário

- **Algoritmo**

Qualquer programa (normalmente *software*) que é desenvolvido para ser implementado em máquinas para realizar processos finitos, para solucionar problemas, seguindo um conjunto definido de operações.

- **Análise Preditiva (*Predictive Analysis*)**

Consiste em fazer previsões sobre resultados futuros através do estudo das tendências dos dados atuais e passados. Pode utilizar algoritmos suportados por aprendizagem automática para extrair a informação necessária dos dados e padrões de comportamento futuros.

- **Análise de Dados (*Data Analysis / Analytics*)**

Envolve extração, limpeza, transformação, modelação e visualização de dados com o objetivo de extrair informação importante e que pode ser útil para tirar conclusões e tomar decisões.

- **Aprendizagem Automática (*Machine Learning*)**

Subárea da Inteligência Artificial que implementa algoritmos que executam tarefas e que, de uma forma autónoma, aprendem informações acerca dos dados. É utilizada para criar modelos que realizam previsões acerca dos dados.

- **Ciência dos Dados (*Data Science*)**

Área interdisciplinar que se concentra na extração de conhecimento a partir de grandes quantidades de dados. Abrange a análise e preparação de dados, como também a apresentação de resultados de modo a basear decisões de alto nível numa organização. Incorpora competências de informática, matemática, estatística e *domain expertise*.

- **Competitividade**

Característica ou capacidade de qualquer organização em cumprir a sua missão, com mais sucesso que outras organizações. Baseia-se na capacidade de satisfazer as necessidades e expectativas dos clientes ou cidadãos aos quais serve, no seu mercado objetivo, de acordo com a sua missão específica, para a qual foi criada.

- **Conhecimento**

Familiaridade ou consciência, de alguém ou algo, tais como factos, competências ou objetos que contribuem para a sua compreensão. Adquirido de diversas formas e fontes, tais como a perceção, razão, memória, investigação científica, educação e prática.

- **Dados**

Observações documentadas que podem ser resultantes de medições, recolhas, capturas e extrações. Geralmente, os dados necessitam de ser processados para extração de informação e conhecimento. Descrevem características de entidades sobre as quais podem operar algoritmos.

- **Prospecção de Dados (*Data Mining*)**

Processo de extração de informação, padrões e tendências a partir de dados em bruto. A prospecção de dados é um método que é utilizado para investigar, analisar, e demonstrar dados para encontrar informação útil. É uma subárea da Análise de Dados.

- **Estratégia**

Conjunto de movimentos e ações competitivas que uma empresa utiliza para atrair clientes, competir com sucesso, reforçar o desempenho, e alcançar objetivos organizacionais. Descreve como o negócio deve ser levado a cabo para atingir os fins desejados.

- **Exploração de Dados (*Data Exploration*)**

Processo inicial da Análise de Dados em que os analistas de dados utilizam a visualização de dados e técnicas estatísticas para descrever caracterizações de conjuntos de dados, tais como precisão, quantidade, e exatidão, de modo a melhor compreender a natureza dos dados.

- **Informação**

Dados processados, organizados e/ou classificados, que tem valores significativos para o quem analisa. A informação é o conjunto de dados processados, que integra o processo de Extração de Conhecimento, onde se baseiam a tomada de decisões e ações.

- **Interdisciplinaridade**

Método ou ideia que funde conceitos, e que consiste na implementação de uma integração de diferentes campos e/ou áreas científicas, que são tradicionalmente pensados como áreas singulares e/ou separadas.

- **Jupyter Notebook**

Ferramenta gratuita, de código aberto e interativa que permite implementar blocos de notas com funções computacionais. Possibilita combinar linguagens de programação, texto, recursos multimédia, visualização de dados, operações computacionais e algoritmos de ML num único documento.

- **Mercado**

Método que uma empresa utiliza para vender produtos ou serviços a um grupo específico de clientes. Tipicamente, os mercados comerciais facilitam as vendas de uma empresa para outra nos casos em que uma empresa planeia reutilizar ou revender os produtos ou serviços de outra empresa.

- **Nicho**

Segmento de um mercado maior que tem as suas próprias exigências e preferências. É definido como sendo uma agregação de esforços estratégicos para um grupo bem definido de clientes alvo.

- **Python**

Linguagem de programação de alto nível, interpretada, orientada a objetos e de alto nível com semântica dinâmica, utilizada para programação de aplicações em geral. É frequentemente utilizada para construir *websites*, *software*, automatizar tarefas, implementar algoritmos e efetuar análises de dados.

- **Processamento de Dados**

Transformação e manipulação de dados para produzir informação significativa e de relevo. É considerado um subprocesso da Extração de Conhecimento e permite converter os dados em bruto em conteúdo significativo e utilizável por algoritmos de aprendizagem automática.

- **Sistema**

Plataforma física e/ou virtual capaz de executar instruções armazenadas que interpretam dados armazenados e/ou lidos a partir de algum *input*, realizam operações, e produzem um *output*.

- **Tarefa**

Ação ou conjunto de ações a serem realizadas por uma máquina através de um ou vários algoritmos. É a execução das ações específicas que resultam da aplicação dos modelos gerados por algoritmos suportados por aprendizagem automática.

- **Tomada de Decisão (*Decision Making*)**

Processo de recolha de informação e avaliação eficaz entre duas ou mais alternativas, com a finalidade de alcançar um resultado esperado. É também um processo de raciocínio que pode ser baseado em dados, pressupostos, preferências e crenças do decisor.

- **Visualização de Dados (*Data Visualization*)**

Representação gráfica de informação e dados num formato pictórico ou gráfico (Exemplo: digramas, gráficos e mapas). As ferramentas de visualização de dados proporcionam uma forma acessível de ver e compreender tendências, padrões e *outliers* nos dados.

Referências

Analytiks. (2020). Why Data Visualization Is Important - Analytiks. Analytiks.Co. <https://analytiks.co/importance-of-data-visualization/>

Anand, A. (2021). 10 Data Science Applications in Real Life. AnalyticSteps. <https://www.analyticssteps.com/blogs/10-data-science-applications-real-life>

Avantika Monnappa. (2022). Why Data Science Matters And How It Powers Business Value? Simplilearn, 1–8. <https://www.simplilearn.com/why-and-how-data-science-matters-to-business-article>

Avidon, E. (2021). 8 analytics startups to watch over the next year. TechTarget. <https://www.techtarget.com/searchbusinessanalytics/feature/8-analytics-startups-to-watch-over-the-next-year>

Berkeley University. (2018). What is Data Science? | The Data Science Career Path. Berkeley School of Information. <https://ischoolonline.berkeley.edu/data-science/what-is-data-science/>

Computerworld. (2020). Faber investe 30 milhões de euros em *startups* ibéricas de "data science" | Computerworld. <https://www.computerworld.com.pt/2020/11/13/faber-investe-30-milhoes-de-euros-em-startups-ibericas-de-data-science/>

Das, S. R. (2016). Data Science: Theories, Models, Algorithms, and Analytics. In S. R. Das.

Data Flair Tutorials. (2021). Top 6 Data Science Use Cases that are Changing the World. Data Flair Tutorials. <https://data-flair.training/blogs/data-science-use-cases/>

DataFlair. (2021). 14 Most Used Data Science Tools for 2021 – Essential Data Science. <https://data-flair.training/blogs/data-science-tools/>

Dönmez, P. (2013). Introduction to Machine Learning. Natural Language Engineering, 19(2), 285–288. <https://doi.org/10.1017/s1351324912000290>

discoverdatascience, (2022) Data Science Career Guide Overview, <https://www.discoverdatascience.org/career-information/>

Edureka, (2022). Data Science Career Opportunities: Your Guide To Unlocking Top Data Scientist Jobs, <https://www.edureka.co/blog/data-science-career-opportunities-your-guide-to-top-data-scientist-jobs>

Grus, J. (Software engineer). (2019). Data Science from Scratch: First Principles with Python.

Henn, B. (2021). What is Data Exploration? Sisudata. <https://sisudata.com/blog/what-is-data-exploration-in-machine-learning>

Igual, L., & Seguí, S. (2020). Introduction to Data Science. <https://doi.org/10.4018/978-1-7998-3053-5.ch001>

Johnson, J. (2020). Predictive Analytics vs Machine Learning: What's The Difference? Bmc Blogs. <https://www.bmc.com/blogs/machine-learning-vs-predictive-analytics/>

Mujthaba, G. M., Ameen, A. Al, Kolhar, M., & Rahmath, M. (2020). Data Science Techniques, Tools and Predictions. International Journal of Recent Technology and Engineering, 8(6), 5661–5668. <https://doi.org/10.35940/ijrte.f9887.038620>

Ohri, A. (2020). Importance Of Data Analytics in 2021 | Jigsaw Academy. Importance Of Data Analytics in 2021. <https://www.jigsawacademy.com/blogs/business-analytics/importance-of-data-analytics/>

Piringer, H. (2020). Benefits of Data Exploration. Visplore.Com. <https://visplore.com/benefits-of-data-exploration/>

Provost, F., & Fawcett, T. (2013). Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. O'Reilly Media, Inc.

Qian, L. (2021). What is a Data Scientist? March. <https://www.mastersindatascience.org/careers/data-scientist/>

Redman, T. C., & Davenport, T. H. (2021). 4 Ways to Democratize Data Science in Your Organization. 1–7. <https://hbr.org/2021/03/4-ways-to-democratize-data-science-in-your-organization>

Rice, M. (2019). 17 Data Science Applications & Examples You Should Know. Built-In. <https://builtin.com/data-science/data-science-applications-examples>

Sanyal, S. (2019). Top 10 Emerging Data Science Startups To Keep an Eye on 2022. Analytics Insight. <https://www.analyticsinsight.net/top-10-emerging-data-science-startups-to-keep-an-eye-on-2022/>

SAS. (2016). Predictive Analytics: What it is and why it matters | SAS. Predictive Analytics Overview, 1–12. https://www.sas.com/pt_pt/insights/analytics/predictive-analytics.html

Schutt, R., & O'Neil, C. (2013). Doing Data Science: Straight Talk.

Skiena, S. S. (2017). The Data Science Design Manual. In Springer. <http://link.springer.com/10.1007/978-3-319-55444-0>

Sundaresan, N. (2011). Data science and machine learning at scale. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6911 LNAI(November), 10. https://doi.org/10.1007/978-3-642-23780-5_9

Taylor, D. (2021). Information vs Knowledge: Key Differences. Guru99. <https://www.guru99.com/information-vs-knowledge-difference.html>

Tucci, L. (2016). What is information age? In TechTarget. <https://www.techtarget.com/searchcio/definition/Information-Age>

Wang, J. (2021). The Importance of Data Analytics in Today's Business World. IMC. <https://www.imc.edu.au/news-archive/the-importance-of-data-analytics-in-today-s-business-world>

Wieringa, J., Kannan, P. K., Ma, X., Reutterer, T., Risselada, H., & Skiera, B. (2021). Data analytics in a privacy-concerned world. Journal of Business Research, 122(July 2018), 915–925. <https://doi.org/10.1016/j.jbusres.2019.05.005>

Fotografias

Fotografia de [Shubham Dhage](#) disponível em [Unsplash](#)
Fotografia de [KOBU Agency](#) disponível em [Unsplash](#)
Fotografia de [Nacional Cancer Institute](#) disponível em [Unsplash](#)
Fotografia de [ThisisEngineering RAEng](#) disponível em [Unsplash](#)
Fotografia de [National Cancer Institute](#) disponível em [Unsplash](#)
Fotografia de [Alexander Sinn](#) disponível em [Unsplash](#)
Fotografia de [Walkator](#) disponível em [Unsplash](#)

Figuras

[Figura 1.1: Fases do Processo de Extração de Conhecimento.](#)
[Figura 1.2: Os conceitos base da análise dos dados.](#)
[Figura 2.1: Os domínios / áreas de atuação que constituem a Ciência dos Dados.](#)
[Figura 2.2: Exemplos de áreas dos setores económicos onde são aplicadas as metodologias suportadas por Ciência dos Dados.](#)
[Figura 2.3: Pontos de invasão terrestre na Ucrânia. Créditos Infografia: Jaime Figueiredo.](#)
[Figura 2.4: Processos da Aprendizagem Automática aplicada aos dados.](#)
[Figura 2.5: Tomada de Decisão baseada em Dados \(DDM\).](#)
[Figura 3.1: Análise suportada pela Ciência dos Dados.](#)
[Figura 3.2: Os 3Vs dos dados que suporta o crescimento dos negócios.](#)
[Figura 3.3: Software utilizado para a visualização de dados.](#)
[Figura 3.4: Várias formas do valor dos dados.](#)
[Figura 3.5: Capacidades técnicas de um cientista de dados.](#)
[Figura 4.1: Plataformas tecnológicas direcionadas para a Ciência dos Dados](#)
[Figura 4.2: Funcionalidades do RapidMiner](#)
[Figura 4.3: Recolha de dados de redes sociais e posterior análise utilizando a plataforma Data Robot.](#)
[Figura 4.4: Manipulação de dados de diversas fontes no Alteryx.](#)
[Figura 4.5: Funcionalidades da plataforma SAS.](#)
[Figura 4.6: Ferramentas suportadas para integração com o Apache Spark.](#)
[Figura 4.7: Exemplo de um criado através do Microsoft Excel](#)
[Figura 4.8: Exemplo de um dashboards criado através da plataforma Tableau.](#)
[Figura 4.9: Jupyter exemplificado na linguagem de programação Python](#)
[Figura 5.1: Principais tendências de ciência dos dados para 2022.](#)
[Figura 6.1: As etapas para uma exploração inteligente dos dados.](#)
[Figura 6.2: Comparação entre Aprendizagem Automática e Análise Preditiva.](#)
[Figura 6.3: GDPR é um regulamento da União Europeia que garante a privacidade da informação](#)
[Figura 7.1: Unbabel na tradução online com a ciência dos dados.](#)
[Figura 7.2: Plataforma da Uniplaces.](#)
[Figura 7.3: Análise em tempo real da Prodsmart.](#)
[Figura 7.4: Feedzai na segurança da atividade bancária.](#)
[Figura 7.5: Detech.ai na melhoria da experiência do utilizador.](#)
[Figura 7.6: Gorodata plataforma de análise de vendas suportada pela ciência dos dados.](#)
[Figura 7.7: A CropIn aplica a ciência de dados num modelo SaaS ao setor agrícola.](#)
[Figura 7.8: Firebolt com soluções SaaS de computação e armazenamento de dados.](#)
[Figura 7.9: Lymbyc com solução SaaS de Análise de Dados para líderes de empresas.](#)
[Figura 7.10: InCountry tem uma plataforma para Análise de Dados para negócio como solução SaaS.](#)
[Figura 7.11: 3LOQ Labs com uma solução de gestão financeira inteligente.](#)
[Figura 7.12: CloudNine AI com uma solução de gestão financeira inteligente.](#)
[Figura 7.13: Razorthink com uma solução de previsão para aumentar vendas e lucros.](#)
[Figura 7.15: Empresa MATH Company.](#)
[Figura 7.16: Zendrive com Plataforma inteligente de risco de mobilidade com mitigação de risco.](#)

Tabelas

Tabela 5.1: Exemplos de soluções que envolvem a ciência dos dados.

AUTORES

António Raimundo e
Pedro Sebastião

**ENTIDADE
PROMOTORA**

IAPMEI, Agência para a Competitividade e Inovação, I.P.
Departamento de Valorização e Capacitação Empresarial
Departamento de Empreendedorismo e Financiamento

**COORDENAÇÃO &
REVISÃO**

AUDAX – Centro de Inovação e Empreendedorismo do ISCTE-IUL
Augusto Albuquerque

DESIGN GRÁFICO

I AM - The Creative House

DATA DE EDIÇÃO

Outubro de 2022

COPYRIGHT

2022, IAPMEI

CONCEÇÃO

audax _iscte



ISBN: 978-972-8191-61-0

cofinanciado por



UNIÃO EUROPEIA

Fundo Europeu
de Desenvolvimento Regional