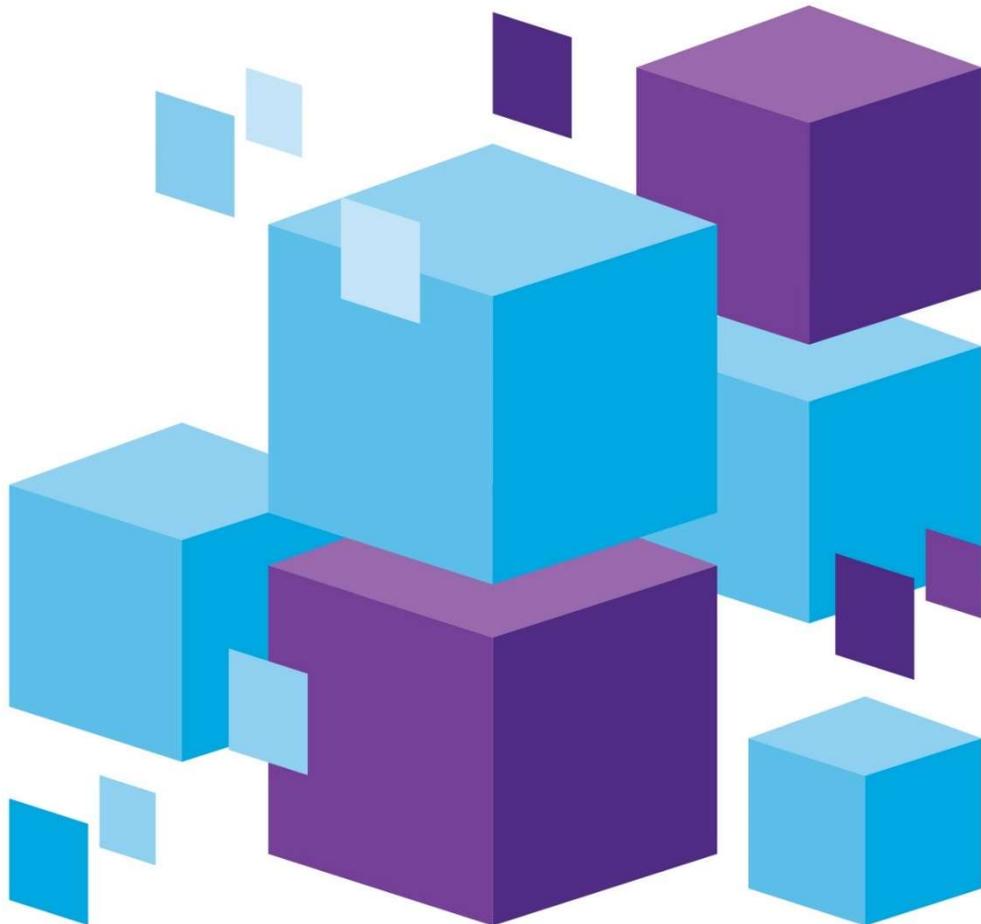


DEZEMBRO DE 2021



# MANUAL DA PROFISSÃO DE DATA SCIENCE

DATA SCIENCE PLAYBOOK

CARREIRAS, COMPETÊNCIAS E TÁTICAS  
*DATA SCIENCE PORTUGUESE ASSOCIATION*

## Tabela de conteúdos

Introdução .....	3
Como usar este Manual?.....	5
A Data Science Portuguese Association .....	5
Os Desafios para a “Democratização” de Data Science.....	5
O Playbook .....	6
A Quem Se Destina o Playbook .....	7
Como usar o Playbook .....	8
Porquê certificar?.....	8
Que Recursos vai encontrar no Playbook .....	10
<i>Benchmark</i> de Autodiagnóstico .....	10
O que é Data Science e AI?.....	11
Os principais domínios de Data Science e AI .....	11
Matemática .....	12
Ciências da Computação .....	12
Conhecimento de Negócio .....	13
Os principais perfis em Data Science e AI .....	13
Competências gerais.....	16
Identificação das competências de cada domínio de dados .....	16
Mapeamento de competências .....	19
Percursos e Níveis .....	21
Metodologia.....	21
<i>Foundational</i> .....	23
Data Analyst.....	24
Data Engineer.....	27
Machine Learning Engineer.....	30
Data Scientist .....	33
O Playbook de Táticas .....	37
Contexto .....	37
PME - Intermédia.....	39
Grande Empresa - Iniciada.....	40

Grande Empresa - Intermédia.....	41
Profissional a mudar de carreira – <i>Reskilling</i> .....	43
Profissional - a Expandir a sua Carreira .....	44
Estudante - Intermédio.....	45
Estudante - Finalista .....	46
Certificação Profissional.....	48
Níveis de certificação.....	48
<i>Journey</i> e opções de candidatura .....	52
Recursos Formativos.....	55
Apoios e Agradecimentos.....	58
Quem nos apoiou? .....	58
Agradecimentos .....	59
Onde nos inspiramos? .....	59
Autoria .....	60

# Introdução

## ***"Vivemos numa era de transformação a grande velocidade!"***

A integração digital que atravessamos tem, mesmo com as vantagens e riscos que implica, uma consequência inevitável no trilho que deixamos na nossa passagem e na utilização dos vários dispositivos e serviços que usufruímos. A capacidade de transmitir, armazenar e combinar dados cria uma oportunidade relevante para que se transforme a informação em conhecimento e, eventualmente, em inteligência.

No entanto, existe um sem número de desafios no acesso e na utilização desses mesmos dados, sejam eles de privacidade, ética, regulação, propriedade, ou sobre as capacidades necessárias para manusear e tirar proveito dessa informação.

Contudo, há uma consciência clara em relação a quem terá uma posição de vantagem neste contexto. Será, sem dúvida, aquele que melhor souber gerir o processo de transformação de informação em inteligência. Numa perspetiva pessoal, as assimetrias da informação trazem vantagens para quem tem mais, isto num debate entre duas partes. A nível de carreira, essas capacidades são altamente valorizadas e em sentido crescente neste segmento. Não menos importante, a nível profissional e económico, uma empresa pode transformar a maior inteligência, que provêm dos dados, em novos produtos, melhores campanhas de comunicação e marketing, políticas de preços ou novos mercados que poderão ser desenvolvidos.

Este processo de transformação da informação em inteligência é conhecido por **Data Science**. Esta é considerada ciência, com uns toques de arte, na extração e organização dos dados para descobrir padrões, prever comportamentos futuros ou tirar conclusões que permitam tomar melhores decisões.



Pretendemos com este **Playbook** ou **Manual da Profissão de Data Science** reforçar o potencial das técnicas nas empresas e despoletar o interesse dos profissionais, tentando explicar que, embora seja uma matéria complexa, é hoje possível e fácil de aceder. Os conhecimentos e processos estão acessíveis a qualquer empresa ou profissional, independentemente da sua área de atividade, estrutura humana ou económica ou do grau de especialização.



**A era da democratização do acesso ao Data Science está mais viva que nunca.** Esperamos que com este documento e a plataforma Data Science Playbook o possam ajudar no acesso imediato e no desenvolvimento de conhecimentos e técnicas de Data Science.

# Como usar este Manual?

## A Data Science Portuguese Association

A Data Science Portuguese Association (DSPA) foi criada em 2017 com o principal objetivo de contribuir ativamente para o desenvolvimento da profissão de Data Science, muitas vezes apelidada como a profissão mais sexy do mundo. Tentamos seguir um caminho diferente, face ao que têm feito outras associações, por acreditarmos ser possível juntar indivíduos, empresas e escolas para criar um veículo que acelere a utilização das técnicas de Data Science e não apenas para *networking*. Tudo isto se torna ainda mais importante pelo potencial de Data Science em resolver problemas da sociedade, ser fundamental para crescimento económico e proporcionar carreiras profissionais mais desafiantes e mais bem remuneradas.

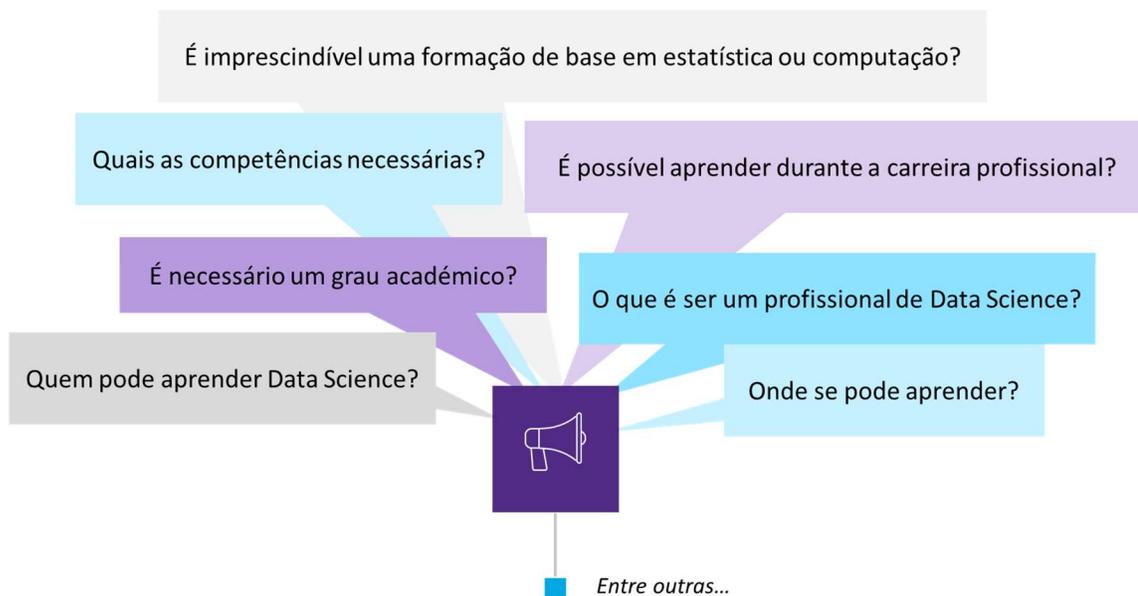


Os fundadores da DSPA sentiram a vontade de contribuir de forma concreta para algo de novo, principalmente num tecido empresarial carenciado como o português. Esta vontade surge por se sentir demasiadas barreiras à volta de Data Science. A dificuldade de acesso a estas tecnologias devido à complexidade, o défice de conhecimento das aplicações ao negócio e o acesso a profissionais preparados são algumas das razões válidas para querer impulsionar o crescimento da Data Science em Portugal.

## Os Desafios para a “Democratização” de Data Science

Ao longo do decorrer das atividades da DSPA, nomeadamente nos nossos fóruns de debate com empresas, profissionais e estudantes deparámo-nos com um conjunto

de perguntas elementares, de natureza recorrente e sem uma resposta óbvia, tais como:



Por este motivo, democratizar o acesso a estas tecnologias tem sido o foco da nossa atuação, explicando as suas aplicações, os benefícios que se podem obter, a execução e quais as competências necessárias para ser profissional nesta área.

## O Playbook

Com este Playbook pretendemos contribuir para a resposta do desafio:

**Como se tornar um profissional de Data Science e Inteligência Artificial (IA)?**

Centramo-nos nesta questão por dois motivos:

- **Mais quantidade:** o incremento do número de profissionais com conhecimento e formação irá acelerar a utilização das técnicas;
- **Menos complexidade:** atualmente o significado de ser um profissional de Data Science é pouco claro. Quem pode aprender? É necessário um grau académico? É possível aprender durante a carreira profissional? É imprescindível ter formação

de base em estatística ou computação? Onde se pode aprender as competências necessárias?

Às respostas destas questões daremos o nome de Data Science Playbook. Este será um material que funcionará como manual de introdução à profissão, como também um guia prático para se iniciar ou progredir na profissão.

No Playbook encontram-se os seguintes componentes:

- **O conteúdo:** as carreiras, as competências, os níveis, entre outros;
- **O processo:** passos e tempos para saber e progredir;
- **As táticas:** ações e atividades em função de uma situação e/ou objetivo particular.

A utilidade e impacto do Playbook serão medidos pelo número de empresas e profissionais que o possam vir a utilizar, sendo essa a verdadeira medida do sucesso e contribuição.

Acreditamos que mais clareza sobre a profissão e as ferramentas para incentivar a progressão dos profissionais podem impulsionar mais pessoas e mais empresas a utilizar técnicas de Data Science. Existem também outros elementos da nossa estratégia nas quais queremos e estamos focados, tais como a implementação de princípios de ética na profissão e o acesso a ferramentas de utilização prática, como modelos preditivos de negócio.

## A Quem Se Destina o Playbook



O Playbook destina-se a qualquer profissional ou empresa que pretenda aprofundar o seu conhecimento sobre as profissões de Data Science. Este será o documento onde podem encontrar informação estruturada sobre os percursos profissionais, as suas características e recomendações para a evolução na carreira.

Os principais destinatários são as Pequenas e Médias Empresas (PME's) e os seus profissionais, muito influenciado pelas características do tecido empresarial português. Contudo, a informação aqui presente poderá ser igualmente útil a qualquer cidadão ou empresa interessada no tema.

Os principais temas abordados são: (1) Carreiras e as variantes da profissão; (2) Conhecimentos mais relevantes para adquirir; (3) Níveis de progressão e competências necessárias; (4) Exemplos de cursos e formações, aplicações práticas dos diferentes domínios, artigos e outras informações interessantes sobre cada domínio.

Também aqui as PME's poderão encontrar ferramentas de autodiagnóstico relativas à sua maturidade em Data Science, como também da visão da maturidade nacional e internacional nesta matéria.

## Como usar o Playbook

Com este documento irá:

1. Conhecer a profissão de Data Science, as suas variantes, áreas de aplicação, níveis de maturidade e competências;
2. Avaliar a sua maturidade, em especial se for uma PME ou uma empresa de grande dimensão;
3. Entender a estrutura, progressão e metodologia proposta;
4. Escolher o percurso que pretende aprofundar, as competências que deverá adquirir e onde as poderá aprender;
5. Certificar as suas competências.

## Porquê certificar?

A certificação de profissionais é outro dos objetivos deste projeto e da DSPA, pois acreditamos no poder das certificações profissionais como uma ferramenta de comprovação de competências para empregadores e candidatos. As certificações permitem também orientar os profissionais numa carreira e no seu progresso, criando continuamente objetivos adicionais, de acordo com uma estrutura dentro da especialização escolhida.

A certificação é um comprovativo prestado, por uma entidade acreditada, sobre o nível de conhecimento e, por vezes, uma aplicação prática, de um indivíduo numa determinada profissão. São normalmente atribuídos por entidades creditadas como associações profissionais, escolas oficiais ou por agências privadas de concessão de certificados.

É normal que as certificações desta natureza tenham um prazo de validade temporal e que tenham de ser renovadas após verificação de condições pré-definidas. Esta renovação geralmente requer uma aprendizagem continuada para garantir atualização nos avanços da disciplina e, por vezes, um sistema de créditos ou de cursos de desenvolvimento profissional reconhecidos e aprovados.

Um dos principais objetivos de um programa de certificação profissional é o interesse em elevar os padrões de qualidade do setor. Os programas de certificação requerem metodologias e processos muito claros e definidos, auditoria e independência para garantir a credibilidade e reconhecimento pela indústria. As certificações são muito comuns em áreas como a saúde, aviação, construção, tecnologia ou finanças. O reconhecimento tem como objetivos principais a facilidade de contratação de profissionais, a candidatura de profissionais em processos competitivos e a garantia de atualização de conhecimentos por parte de profissionais já empregados.

Uma certificação profissional é distinta de um processo de credenciação profissional, pois estes são normalmente atribuídos por ordens profissionais ou agências estatais, fornecendo um licenciamento individual obrigatório por lei para o exercício de uma profissão.

Existem situações em que a acumulação de certificados profissionais pode ter reconhecimento parcial em créditos académicos, em condições bem definidas, mas normalmente estes dois sistemas constituem referenciais bem distintos, embora muitas vezes se complementem mutuamente.

A recente evolução muito rápida da economia, sociedade e, principalmente, da tecnologia veio criar grande disrupção em modelos, processos e conhecimentos necessários para exercer uma profissão de maneira competitiva. Isto tem impacto a nível de uma mais rápida obsolescência do conhecimento tradicional e obriga a uma maior velocidade de mudança e de aprendizagem em cada profissão.

Este fenómeno deu origem a pequenos cursos de aprendizagem e, por consequência, pequenas certificações. É uma tendência que está a tornar o desenvolvimento profissional mais granular e mais conveniente, permitindo uma maior flexibilidade e adequação na utilização do tempo dedicado à aprendizagem e atraindo mais profissionais para estes processos

## Que Recursos vai encontrar no Playbook

- Estrutura da profissão de Data Science;
- Características e conhecimentos gerais;
- Percursos de especialização;
- Competências específicas;
- Autodiagnóstico para empresas;
- Questionário de autodiagnóstico;
- Panorama de maturidade europeia e nacional;
- Que tipos de cursos existem disponíveis para aprender Data Science;
- Onde se aprende, exemplo e links;
- Exemplos de aplicação prática;
- Links selecionados de casos demonstrativos de aplicação prática;
- Conteúdos de interesse tais como artigos, vídeos, livros recomendados, entre outros;
- Processo de certificação;
- Metodologia de certificação, os seus requisitos e passos a dar;
- Interação com uma rede de profissionais (*em construção*);
- Informação sobre a rede de profissionais registados (*em construção*);
- Meios de interação entre as empresas e os membros da rede (*em construção*).

## Benchmark de Autodiagnóstico

Com o objetivo de analisar o estado da arte de Data Science e AI no mundo corporativo e de potenciar nas empresas uma visão comparativa do estado de maturidade, a DSPA disponibiliza o *benchmark* de autodiagnóstico.

Este método de reconhecimento corporativo será instanciado via um questionário de autodiagnóstico que captura as especificidades de utilização de Data Science e AI das empresas. A captura em escala desta informação potência a apresentação de um relatório (*report* de autodiagnóstico) que permitirá às empresas contextualizar e analisar o seu estado da arte quando comparado com o mercado nacional e internacional.

# O que é Data Science e AI?

## Os principais domínios de Data Science e AI

Durante os últimos anos, o domínio dos dados foi alvo de grandes evoluções. O incremento da capacidade de entrega de soluções de Inteligência Artificial, a democratização de acesso a soluções de Big Data e Cloud e os novos avanços tecnológicos criaram um vasto conjunto de funções que exigem diferentes tipos de expertise. De forma a descrever os impactos destas evoluções, é necessário analisar as áreas de conhecimento que influenciam o domínio de dados, no qual se destacam três principais: **Matemática**, **Ciências da Computação** e **Negócio**.

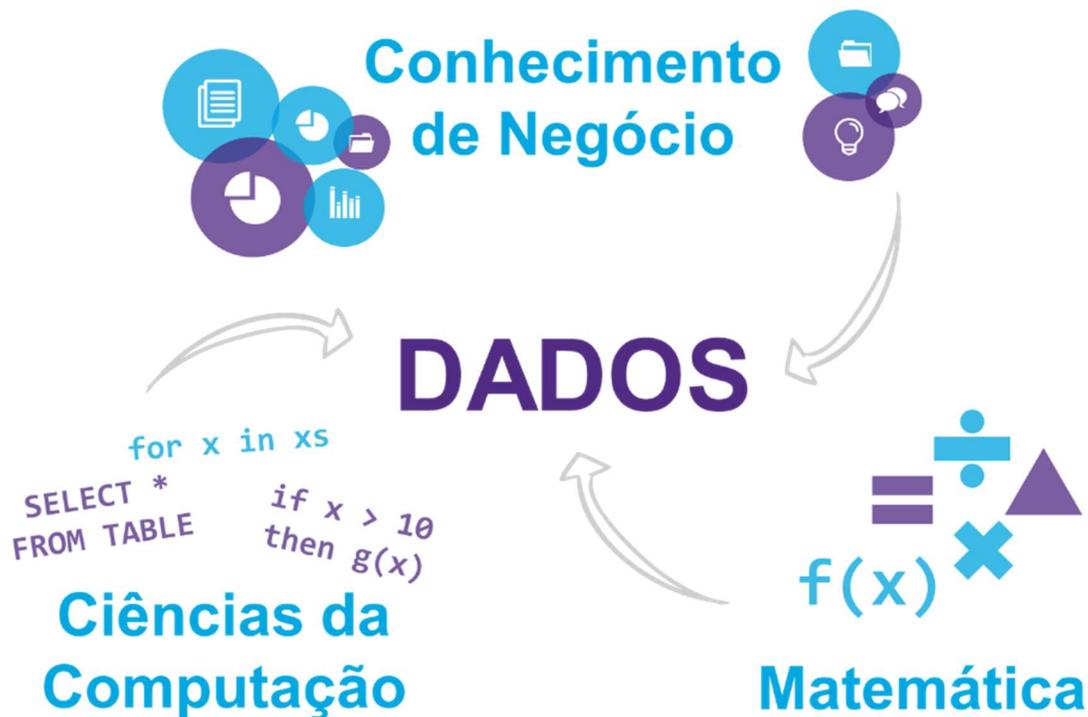


Figura 1 - Principais áreas de conhecimento do domínio dos dados

De acordo com o diagrama da figura 1, destacam-se três áreas de conhecimento principais:

## Matemática

O campo da Matemática é vasto e abrangente, pelo que é essencial compreender quais os tópicos fundamentais para trabalhar em Data Science e AI. Há três tópicos recorrentes que se destacam: **Cálculo**, **Álgebra linear** e **Estatística**.

- a. **Cálculo**: os princípios de cálculo são muito importantes, pois são utilizados para enunciar as funções utilizadas na formulação de algoritmos para atingir o seu objetivo. Por exemplo: compreender um algoritmo de gradiente e como as diferentes variáveis podem afetar o sistema modelado;
- b. **Álgebra linear**: a maioria dos cálculos requer a utilização de álgebra linear. Desde a redução das dimensões nos dados até ao cálculo de redes neurais, todas estas operações computacionais são baseadas em álgebra linear;
- c. **Estatística**: é utilizada para analisar e visualizar dados e, conseqüentemente, inferir percepções úteis. É o domínio com o qual os profissionais devem estar mais familiarizados e ter um nível de conhecimento aprofundado. Adicionalmente, existem outros domínios, tal como a matemática discreta, a teoria dos grafos ou a teoria da informação, que permitiram um alargamento do conhecimento.<sup>1</sup>

## Ciências da Computação

As ciências da computação são uma área de conhecimento multidisciplinar heterógeno, no qual é expectável que o profissional domine um subconjunto de tópicos. Desses tópicos, destacam-se os seguintes:

- Gestão de fluxos de dados;
- Desenho e desenvolvimento de *pipelines* de dados;
- Familiarização de diversas linguagens de programação;
- Utilização e gestão de pacotes e software adequados;
- Execução de *data wrangling*;
- Gestão de base de dados;

---

<sup>1</sup> <https://towardsdatascience.com/mathematics-for-data-science-e53939ee8306>

- Comunicação através de visualização de dados adequadas;
- Computação em Cloud e *Big Data*.

## Conhecimento de Negócio

Em Data Science e AI é essencial compreender e conhecer o negócio. Só então será possível extrair informações úteis e conteúdos relevantes a partir dos extensos dados que possam ser fornecidos.

*"Business knowledge is a business owner's extensive reservoir of understanding on customers' needs and preferences, business environments and their dynamics, staff skills, experiences and potentials, and the business' overall foreseeable direction."<sup>2</sup>*

No Conhecimento de Negócio destacam-se duas peças fundamentais: **Processos** e **Core Business**.

- a. **Processos**: são o conjunto de procedimentos internos relacionados com as operações desempenhadas na área, que influenciam e impactam a geração de dados e a relação entre os mesmos e entre diversas entidades;
- b. **Core Business**: é o aglomerado de conhecimento de todas as dimensões do negócio.

## Os principais perfis em Data Science e AI

A conjugação das várias áreas de competência permite derivar um conjunto de perfis correspondentes à heterogeneidade de tarefas necessárias para o domínio de dados: **Data Analyst, Data Engineer, Machine Learning Engineer** e **Data Scientist**.

---

<sup>2</sup> <https://www.mybusiness.com.au/management/4458-the-importance-of-business-knowledge-for-smes>>

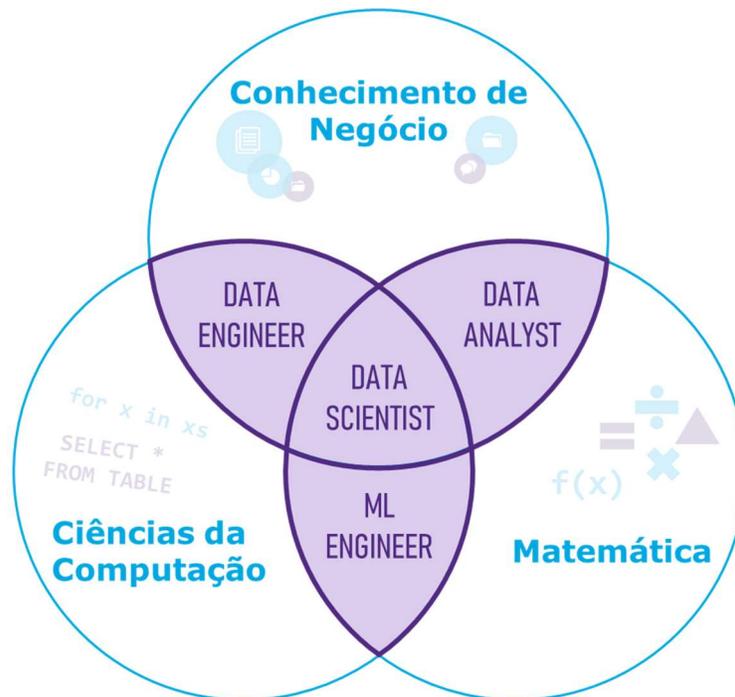


Figura 2 - Mapeamento entre perfis e áreas de conhecimento

Como tal, será apresentada uma caracterização inicial de alto nível dos vários perfis, os quais serão descritos com maior detalhe na secção de competências gerais.



Os **Data Analysts** conjugam o conhecimento teórico-prático para identificar e comunicar *insights*, permitindo à camada de gestão, *stakeholders* e outros executivos de uma organização uma tomada de decisão mais informada.<sup>3</sup>



Os **Data Engineers** são responsáveis pela criação e gestão de sistemas que ingerem e transformam os dados, garantindo que estes são armazenados e disponibilizados para consumo dos restantes perfis.<sup>4</sup>

<sup>3</sup> <https://hbr.org/2018/12/what-great-data-analysts-do-and-why-every-organization-needs-them>

<sup>4</sup> <https://www.oreilly.com/content/data-engineering-a-quick-and-simple-definition/>



O **Machine Learning Engineer** (ML Engineer) utiliza as “receitas” desenvolvidas pelos Data Scientists e industrializa-as em sistemas de produção que permitem executar algoritmos avançados em escala e com governo.<sup>5</sup>



Os **Data Scientists** têm a tarefa de encontrar soluções baseadas em dados para problemas de negócios com recurso a algoritmia avançada. Tendem a ser altamente especializados, nomeadamente na preparação de dados, análise estatística, visualização de dados e Machine Learning.<sup>6,7</sup>

---

<sup>5</sup> <https://www.oreilly.com/radar/what-are-machine-learning-engineers/>

<sup>6</sup> Stanton J. et al, Interdisciplinary Data Science Education, 10.1021/bk-2012-1110.ch006

<sup>7</sup> Walker M., The professionalisation of data science, 10.1504/IJDS.2015.069048

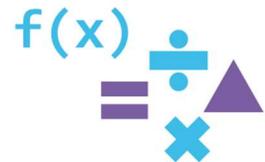
# Competências gerais

## Identificação das competências de cada domínio de dados

Com base nas áreas de conhecimento identificadas para os diferentes perfis, é necessário aprofundar e identificar um conjunto de competências que caracterizam as áreas de conhecimento. Desse modo, pode-se materializar quais os componentes chave e o conjunto de ações e tarefas que as caracterizam, com o objetivo de mapeá-las com cada um dos perfis.

Cada área de conhecimento será analisada num conjunto de competências-chave que melhor tipificam a área de atuação da Data Science e AI.

## Matemática



A **Estatística** é utilizada em Data Science e AI para analisar e visualizar dados e inferir percepções úteis, na qual é esperado um grau de familiarização e nível de conhecimento elevado, devido à sua transversalidade.



Os princípios de **Cálculo e Álgebra Linear** são importantes, pois são utilizados para enunciar as funções utilizadas na formulação de algoritmos, como por exemplo compreender um algoritmo de gradiente e como as diferentes variáveis podem afetar o sistema modelado.

*"Machine Learning is a subset of Artificial Intelligence that studies computer algorithms that improve automatically through experience.<sup>8</sup>"*



Aprendizagem automática para a Data Science e AI é essencial pois existe a necessidade de gerir e operar grandes quantidades de dados e porque o processo de tomada de decisão é centrado nos dados. Para o processamento e modelação dos dados, utilizam-se alguns dos mais conhecidos algoritmos, tais como o K-nearest neighbours, Random Forests, Naive Bayes, modelos de regressão, entre outros.

## Ciências da Computação

```
for x in xs
SELECT *
FROM TABLE
if x > 10
then g(x)
```

A grande maioria dos temas relacionados com Data Science e AI dependem intrinsecamente de **Ferramentas de Programação** ou de software.

*"Any software program or utility that aids software developers or programmers in creating, editing, debugging, maintaining and/or performing any programming or development-specific task.*



*Programming Skills for Data Science brings together all the fundamental skills needed to transform raw data into actionable insights.<sup>9</sup>"*

Um exemplo será a utilização de linguagens de programação, tal como Python e R, instanciadas através de Notebooks ou scripts, dentro de *pipelines* de *Big Data* utilizando soluções na Cloud ou *on-prem*.

<sup>8</sup> Mitchell, Tom (1997). Machine Learning. New York: McGraw Hill. ISBN 0-07-042807-7. OCLC 36417892.

<sup>9</sup> <https://towardsdatascience.com/top-10-skills-for-a-data-scientist-in-2020-2b8e6122a742>

A **Visualização e Comunicação de Dados** tem como objetivo contar de forma intuitiva e clara a história que os números representam. Pode ser sob a forma de infografias, gráficos, entre outros. Há três pilares que a visualização de dados deve seguir:



1. Suporta o público
2. Estabelece um enquadramento claro
3. Conta uma história

Os *designers* de visualização estão responsáveis pela análise dos dados a fim de os compreender e concretizarem com a visualização correta para o consumidor da comunicação. Os bons designers sabem não só como escolher o gráfico e a gama de dados certos, mas também como ser um contador de histórias convincente através da visualização.<sup>10</sup>



**Data Wrangling** refere-se ao processo de transformação e mapeamento de dados de um formato para outro com a intenção de o tornar mais apropriado e útil para as etapas que se seguem na sua utilização, tais como transformação ou análise. Este processo representa uma grande parte das tarefas realizadas em iniciativas de Extract Transform and Load (ETL), mas também em esforços de enriquecimento de dados para a aplicação de algoritmos de Machine Learning ou para a apresentação de dados em camadas de visualização.

*"Is the systematic application of engineering approaches to the development of software. <sup>11</sup>"*



De forma a garantir a escalabilidade e *governance* dos ativos tecnológicos, é imperativo a aplicação de conhecimento e boas práticas de **Engenharia de Software**. Estas permitem ter controlo sobre o código desenvolvido (e.g., GitOps, versionamento, CI/CD), sobre a qualidade através de baterias de teste (e.g., *Test Driven Development*, *Unit Testing*) e seguindo padrões que permitam a fácil integração e escalabilidade (i.e., aplicação de padrões de desenho standards).

---

<sup>10</sup> <https://hbr.org/2013/04/the-three-elements-of-successf>

<sup>11</sup> Abran, Alain; Moore, James W.; Bourque, Pierre; Dupuis, Robert; Tripp, Leonard L. (2004). Guide to the Software Engineering Body of Knowledge. IEEE. ISBN 978-0-7695-2330-9

# Conhecimento de Negócio



A **Intuição de Dados** é definida pela maneira como o indivíduo aborda os dados. Quando confrontado com grandes volumes de informação essa intuição auxilia a compreensão e contextualização.



A intuição pode ser útil quer no entendimento do contexto de negócio dos Dados, bem como auxiliar na rápida captação de possíveis erros e incongruência nos Dados, alavancando o *know-how* que um indivíduo possui sobre o *core business* e os processos associados à atividade em questão.

## Mapeamento de competências

Com base nos perfis identificados e no conjunto de áreas de competências, o seguinte mapeamento é efetuado com o objetivo de identificar quais as áreas de competência com maior relevância para cada um dos perfis de Data Science e AI:<sup>12</sup>

Tabela 1 - Mapeamento de competências por cada perfil

	Data Analyst	Data Engineer	ML Engineer	Data Scientist
Estatística	●	●	●●	●●
Cálculo e Álgebra Linear	●	●	●●	●
<i>Machine Learning</i>	●	●	●●	●●
Ferramentas de programação	●●	●●	●●	●●

<sup>12</sup> <https://blog.udacity.com/2014/11/data-science-job-skills.html>

Visualização e Comunicação de Dados				
<i>Data Wrangling</i>				
Engenharia de Software				
Intuição de Dados				



*Very Important:*

Qualificação ou competência essencial para o cargo



*Somewhat Important*

Qualificação ou competência importante para o cargo, mas não essencial



*Not that important*

Qualificação ou competência não essencial para o cargo

# Percursos e Níveis

## Metodologia

De forma a simplificar o enquadramento de cada perfil no ecossistema de dados, é estabelecido um conjunto de três graus de experiência expectáveis que, transversalmente, caracterizam a evolução dos profissionais ao longo dos vários perfis de Data Science e AI.

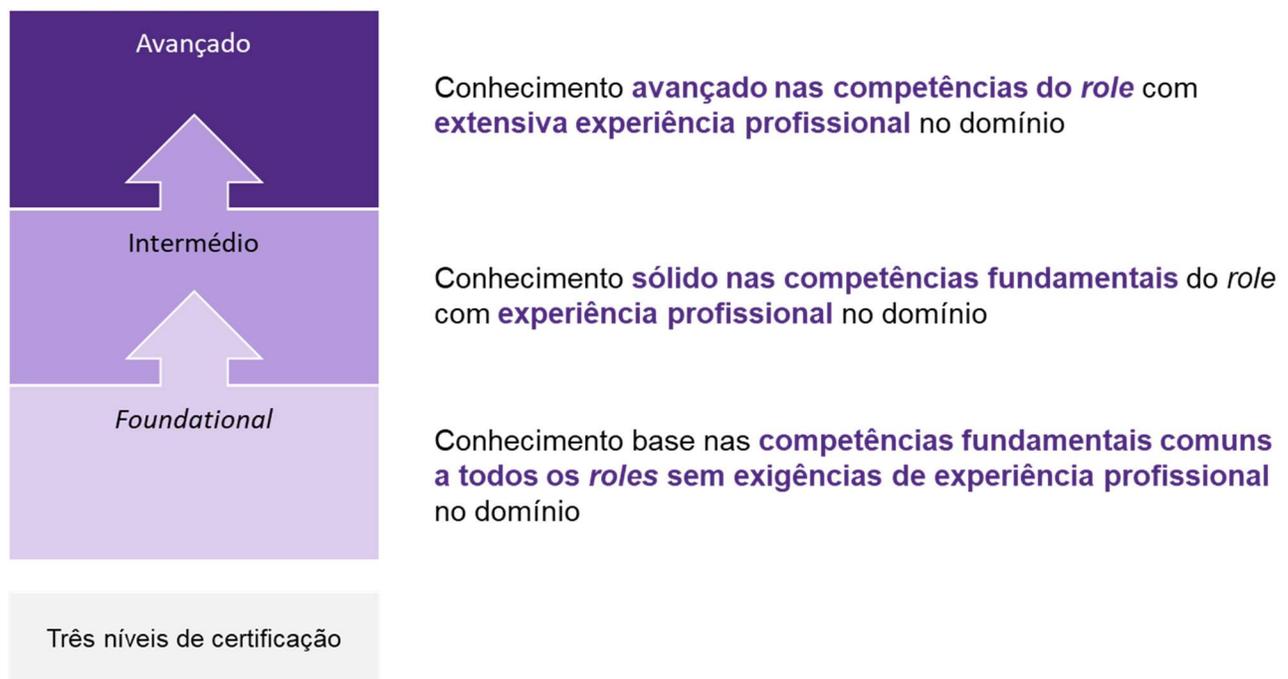


Figura 3 - Níveis de certificação

O primeiro nível – *Foundational* – representa um estado de maturidade do profissional que corresponde ao conjunto de competências fundamentais comuns a todos os perfis de Data Science e AI, servindo de base a partir do profissional que estará apto para iniciar a sua especialização em qualquer um dos perfis de Data Science e AI.

Como tal, o *landscape*, que mapeia os perfis e os níveis, desdobrar-se-á num nível transversal de *Foundational*, seguido de dois níveis, um de intermédio e outro de avançado, para cada um dos perfis.

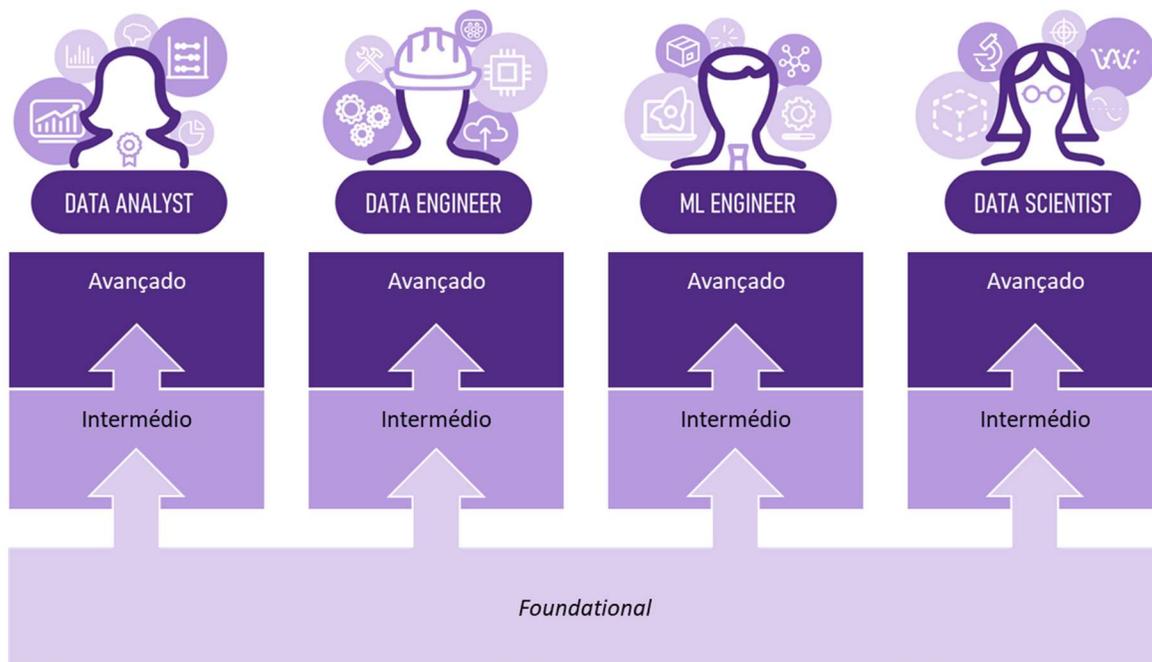


Figura 4 - Mapeamento de níveis com perfis

O nível intermédio e avançado representam o profissional que conseguiu especializar-se, de forma multidimensional nas temáticas específicas a cada um dos perfis de Data Science, i.e., evidenciando domínio a nível de tecnologia, ferramentas, processos, desenho e gestão associado à tipologia de trabalho do perfil. A diferenciação entre os três níveis será detalhada na secção de Certificação Profissional, onde será evidenciado os processos e metodologia da DSPA para a aferição das competências de profissionais em Data Science.

De seguida, serão detalhadas as competências *core* expectáveis para cada perfil de Data Science, incluindo o nível *Foundational*.

## Foundational

### Competências

Competência	Descrição
Ferramentas de Programação	<p>Um perfil <i>Foundational</i> deverá reunir os básicos para interpretar ou desenvolver pequenos blocos de código. Para isso, é esperado conhecimento acerca de variáveis em código, distinção dos vários tipos de dados usados em programação (<i>strings</i>, inteiros, booleanos, <i>doubles</i>, etc.) e conhecimento alto nível do que são listas e <i>arrays</i>.</p> <p>Alguém com este nível de competências deverá ser capaz de interpretar blocos de código com condições (<i>if</i>, <i>else</i>, <i>else if</i>), reconhecer diferentes sintaxes e ter um entendimento superficial de funções.</p>
Visualização e Comunicação de Dados	<p>A apresentação de resultados de forma gráfica deverá estar patente na sua formação, seja através de ferramentas como Microsoft Excel ou outras que sirvam o mesmo propósito. Espera-se que alguém com perfil <i>Foundational</i> saiba distinguir <i>pie charts</i>, gráficos de barras, gráficos de dispersão, entre outros, e que saiba adequar a escolha dos mesmos a cenários práticos.</p>
Intuição de Dados	<p>A identificação de padrões básicos e tendências nos dados trabalhados deve estar bem patente, conseguindo interpretar, ainda que de forma superficial, os cenários apresentados.</p>
Estatística	<p>Conceitos como população e amostra são cruciais para entendimento básico de Estatística. Identificação de distribuições normais, cálculo de médias e desvio padrão não devem ser procedimentos estranhos.</p>
<i>Data Wrangling</i>	<p>No nível <i>Foundational</i>, deve ser evidenciada a capacidade de olhar para um <i>dataset</i> e identificar incongruências a nível da formatação, através de, por exemplo, distinguir o que são valores <i>null</i> e valores <i>blank</i>. O contacto com conceitos como <i>substrings</i>, <i>trims</i>, <i>concat</i> ou diferentes formatações para datas é benéfico.</p>

Competência	Descrição
<i>Machine Learning</i>	Um perfil Foundational deve ser capaz de explicar sucintamente o que é <i>Machine Learning</i> , qual o seu objetivo e exemplificar com vários cenários.
Engenharia de Software	Os conhecimentos básicos de Engenharia de Software como protocolos de comunicação (HTTP, TCP, UDP, entre outros), memória (caching, RAM, ROM, HDD, entre outros), identificação de arquiteturas simples e modelos relacionais deverão ser evidenciados. O contacto com diferentes IDE's ( <i>Integrated Development Environment</i> ), como Eclipse ou IntelliJ, é também um dos requisitos.
Cálculo e Álgebra Linear	Depreende-se que os candidatos com este perfil tenham formação académica nas áreas de Cálculo e Álgebra Linear, assumindo-se que operações com derivadas, primitivas, matrizes e cálculo espacial (e.g., integrais triplos) façam parte do seu domínio.

## Data Analyst

### Caraterização



Binómio problema-solução. Se existe um problema, o Data Analyst descobri-lo-á e, infalivelmente, chegará a uma solução. Munidos de mentes curiosas e sentidos matemáticos apuradíssimos. Bons comunicadores e com uma componente técnica versátil para poder fazer frente às diferentes situações que aparecem

Um Data Analyst deve apresentar uma grande componente de pensamento crítico, boa capacidade de comunicação, aptidão para resolução de problemas e saber sustentar as suas análises em visualizações simples e interativas, pelo que, no domínio técnico, terá alguma experiência com ferramentas que envolvam dados (e.g., Excel, Bases de Dados, linguagens de programação) e subsequentes visualizações (e.g., PowerBI, Looker, Tableau, QlikSense, Microsoft Excel). Estes elementos devem ser conjugados com um forte raciocínio matemático/lógico, uma

excelente capacidade de interpretação, modelação e atenção ao detalhe, que são características que devem estar bem desenvolvidas.

### Exemplos de Tecnologias

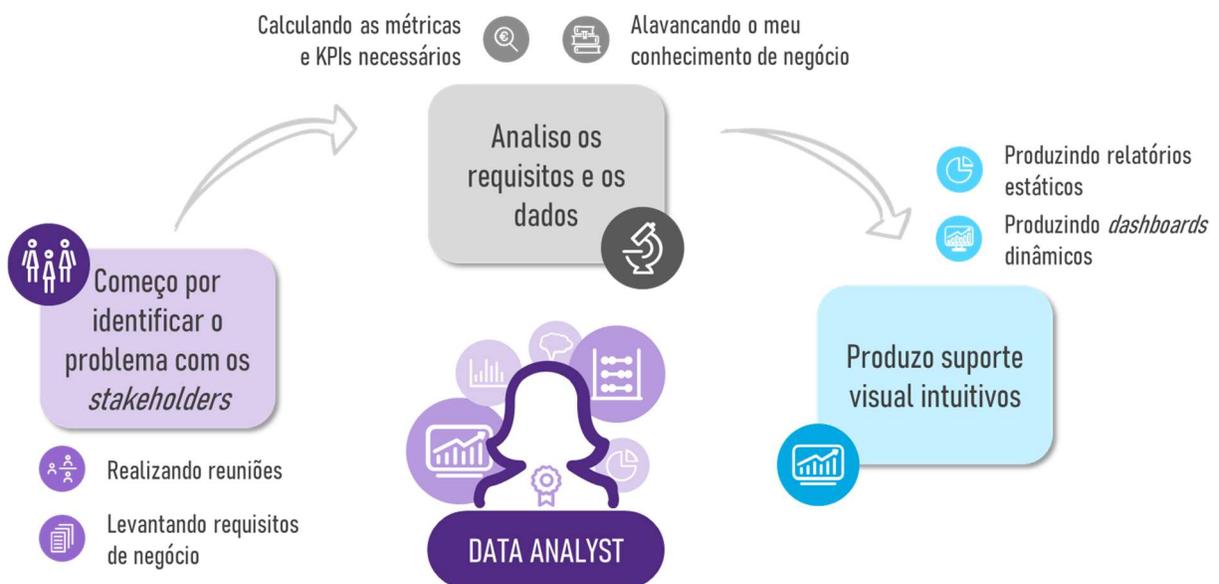


## Competências

Competência	Descrição
Ferramentas de Programação	<p>Para um Data Analyst, as capacidades de programação são menos exigentes que para os outros perfis. O Data Analyst tende a dominar linguagens como SQL e R, e software como PowerBI, Tableau, SAS, Excel, RapidMiner, QlikView.</p> <p>O Data Analyst deve estar focado na manipulação de informação, de forma a apresentar e retirar inferências de negócio.</p>
Visualização e Comunicação de Dados	<p>A visualização de informação é extremamente importante para o Data Analyst, uma vez que é a partir dela que são retiradas as conclusões e informações. O trabalho deste perfil é focado na preparação da informação para ser apresentada da maneira mais correta para ser comunicada e interpretada.</p> <p>Algumas ferramentas úteis são o PowerBI, Tableau, QlikSense, Microsoft Excel, entre outros.</p>
Intuição de Dados	<p>O Data Analyst usa a sua intuição em relação aos dados para a interpretação e análise dos mesmos, gerando inferências de negócio, representadas na forma de visualizações e relatórios adequados ao contexto</p>
Estatística	<p>Um Data Analyst pode recorrer a métodos estatísticos que permitam avaliar a qualidade de um conjunto de dados, das implicações estatísticas a analisar e como retirar conclusões importantes dos dados.</p>
<i>Data Wrangling</i>	<p>Competência não essencial para o cargo.</p>

Competência	Descrição
<i>Machine Learning</i>	Competência não essencial para o cargo.
Engenharia de Software	Competência não essencial para o cargo.
Cálculo e Álgebra Linear	Competência não essencial para o cargo.

### Descrição de caso prático



De forma a produzir inferências de negócio para os *stakeholders* relevantes, as tarefas típicas do Data Analyst centrar-se-ão na recolha de dados, como por exemplo a partir de um Data Warehouse, na análise e identificação de padrões, métricas e KPIs (*Key Performance Indicators*), recorrendo a ferramentas de análise e/ou programação, nomeadamente SQL, e na produção de elementos visuais, recorrendo a ferramentas de visualização, de forma a finalizar com um relatório com as conclusões retiradas (e.g., *dashboard*, relatórios estático).

# Data Engineer

## Caraterização



Um Data Engineer é o **todo-o-terreno** responsável pela estrutura que **ingere, transforma e carrega** os **dados** provenientes de vários sistemas diferentes numa estrutura com valor a partir da qual os Data Analysts, ML Engineers e Data Scientists podem desenvolver o seu trabalho

Ao Data Engineer é exigido o conhecimento detalhado de um leque expressivo de tecnologias com diferentes linguagens de programação associadas, pedindo-se um perfil analítico, versátil, curioso e com domínio de código. As tecnologias utilizadas em engenharia de dados evoluem a uma velocidade significativa, pelo que é preciso mostrar proatividade para uma aprendizagem contínua. A atenção ao detalhe e a capacidade de comunicação são de carácter obrigatório. Ter uma visão alargada do seu trabalho é bastante útil, já que um Data Engineer tem que acompanhar o fluxo de dados com origens diferentes e através de sistemas diferentes.

### Exemplos de Tecnologias



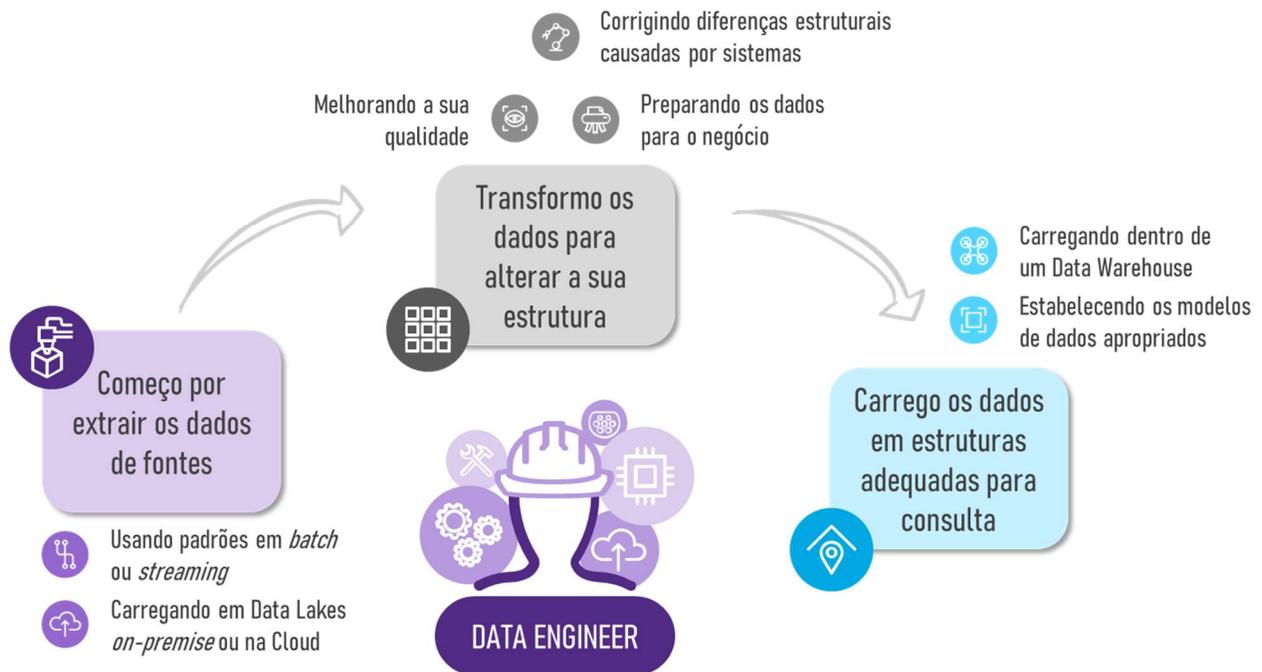
Azure



## Competências

Competência	Descrição
Ferramentas de Programação	Para o Data Engineer é importante dominar um ou mais domínios de atuação, nomeadamente linguagens de programação como Python, Java e Scala, <i>frameworks</i> de processamento, orquestração e ingestão (e.g., Apache Spark, Apache Beam Airflow, Kafka), ferramentas de ETL ( <i>Extract, Transform and Load</i> ) como Pentaho ou Talend, ambientes Cloud (AWS, GCP, Azure) e os seus serviços de ingestão, ETL e Data Warehousing, tal como tecnologias Big Data Hadoop.
Visualização e Comunicação de Dados	O Data Engineer pode usar algumas bibliotecas das linguagens de programação com que está familiarizado para comunicar informação sobre os dados que está a trabalhar.
Intuição de Dados	Para um Data Engineer, a intuição pode ser útil para facilmente se familiarizar com os dados.
Estatística	Um Data Engineer aplica métodos de estatística nos pipelines de dados em escala para apresentar a informação necessária a consumidores de dados.
<i>Data Wrangling</i>	O Data Engineer é responsável pela industrialização dos processos de transformação e manipulação para ambientes de produção.
<i>Machine Learning</i>	Competência não essencial para o cargo.
Engenharia de Software	O Data Engineer é responsável pelos processos mais complexos a desenvolver, requerendo capacidades de engenharia de software elevadas. O Data Engineer deverá ser capaz de desenhar, definir arquitetura, implementar e efetuar a manutenção dos processos de ingestão e tratamento de dados de ponta a ponta, seguindo as boas práticas de desenvolvimento de software (e.g., GitOps, CI/CD, TDD).
Cálculo e Álgebra Linear	Competência não essencial para o cargo.

## Descrição de caso prático



O caso prático de um *Data Engineer* focar-se-á nos elementos periféricos do *Data Lake*, onde o seu ónus é centrado. O *Data Engineer* ficará responsável por todas as camadas de ingestão e processamento de dados que disponibilizam os mesmos no *Data Lake* de forma a serem consumidos por outros utilizadores e aplicações. O *Data Engineer* estará focado no desenvolvimento das *pipelines* de tratamento de informação, utilizando um leque diversificado de ferramentas, tecnologias, *frameworks* e linguagens (e.g., SQL, Python, serviços Cloud, Spark), interagindo diretamente com infraestruturas, de forma a garantir robustez e escalabilidade aos processos, trabalhando maioritariamente com dados em formato *raw*, de grande volume e em tempo real.

# Machine Learning Engineer

## Caraterização



Catalisador de um projeto de Machine Learning (ML). Ao pensar no role de ML Engineer, chegamos de imediato a dois conceitos: modelação e escalabilidade. O objetivo do seu trabalho é garantir a execução de complexos modelos de ML, servindo-se, para o efeito, de um vasto repertório de linguagens de programação

Alguém com apetência para Machine Learning Engineer deverá sentir-se extremamente confortável num ambiente de desenvolvimento/programação. Pedese uma componente matemática e estatística forte, resiliência, já que existem muitos modelos que falham, uma excelente capacidade na identificação de padrões, para além de, claro, ter uma mente curiosa. O raciocínio iterativo é visto como uma vantagem, já que as soluções de Machine Learning seguem desenvolvimentos dessa tipologia (iterando de versão em versão até chegar ao modelo final). O raciocínio analítico e a compreensão/definição de métricas são requisitos primordiais. O Machine Learning Engineer deve comunicar de forma clara e concisa, já que, muitas vezes, é necessário explicar conceitos a outros profissionais não familiarizados com as matérias.

### Exemplos de Tecnologias

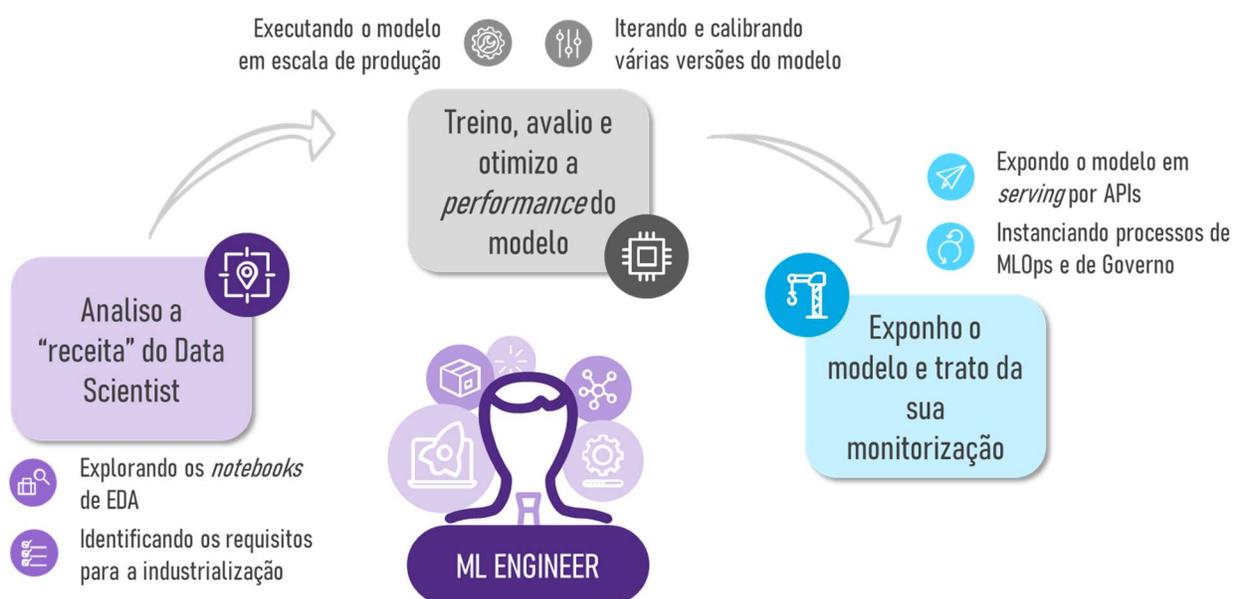


## Competências

Competência	Descrição
Ferramentas de Programação	Para o Machine Learning Engineer é importante ser fluente no cruzamento de matemática com a programação. É neste ponto que se inserem ferramentas mais complexas de programação para o desenvolvimento de modelos e algoritmos. O <i>Machine Learning Engineer</i> apresentará um domínio profundo de linguagens de programação, como Python, R, MatLab, Scala ou Java, de bibliotecas, como Tensorflow, Keras, SciKit, SparkML, e de <i>frameworks</i> de operacionalização de pipelines de <i>Machine Learning</i> , como kubeflow, MLFlow.
Visualização e Comunicação de Dados	Para um Machine Learning Engineer, a visualização, embora não essencial, é também importante. Um Machine Learning Engineer com conhecimentos profundos de programação pode usar algumas bibliotecas que providenciam visualização como, por exemplo, as bibliotecas Plotly ou Dash de Python.
Intuição de Dados	A intuição relativamente aos dados para um Machine Learning Engineer é extremamente importante. Com a experiência que advém do seu trabalho, o <i>Machine Learning Engineer</i> tem uma certa sensibilidade que lhe permite manipular os dados e conhecer que técnicas são apropriadas para cada exercício, evitando assim erros na interpretação e obtenção de resultados e conclusões.
Estatística	Estatística é extremamente importante para o Machine Learning Engineer, uma vez que recorre a métodos estatísticos que permitem avaliar a qualidade de um modelo e de um algoritmo, bem como retirar conclusões importantes dos dados.
<i>Data Wrangling</i>	Competência não essencial para o cargo.
<i>Machine Learning</i>	O Machine Learning Engineer é responsável pelas tarefas relacionadas com Machine Learning, pelo que é essencial que esteja familiarizado com os diferentes tipos de estruturação de dados, diferentes tipos de modelos e algoritmos.

Competência	Descrição
Engenharia de Software	O Machine Learning Engineer precisa de ter noções de engenharia de software de forma a executar tarefas relativas à implementação de algoritmos, operações computacionais e operações de industrialização/ <i>serving</i> .
Cálculo e Álgebra Linear	É através do Cálculo e da Álgebra que são construídos os modelos e algoritmos, bem como funções de custo e objetivo utilizadas em Machine Learning.

### Descrição de caso prático



O Machine Learning Engineer estará responsável por traduzir os modelos e inferências do Data Scientist para ambientes de produção com escalabilidade e capacidades de monitorização, de forma a sistematizar a criação de modelos em repositórios próprios, com capacidades de gestão de ciclo de vida e monitorização de hiper parâmetros e métricas de performance. Como tal, o Machine Learning Engineer trabalhará com tecnologias de transformação de dados para *feature engineering* (e.g., Python), treino de modelos em escala, disponibilização de modelos para padrões de consumo de inferência preditiva e na gestão da infraestrutur e *frameworks* de instanciação de pipelines de Machine Learning, incluindo temáticas de Governance.

# Data Scientist

## Caraterização



Metade hacker, metade analista. Se a verdade existe, o Data Scientist vai descobrir. Máquinas estatísticas com elevada diferenciação a nível de ferramentas de programação. Não há um único segredo que uma montanha de dados consiga esconder destes profissionais. Ilusionistas que transformam números em soluções

Os Data Scientists devem combinar uma forte componente matemática/estatística com competências de programação, versatilidade e grande expertise técnica, já que é necessário conhecimento de um vasto leque de ferramentas com recurso a diferentes linguagens de programação. A apetência para o detalhe, polivalência e curiosidade para acompanhar proactivamente o crescimento da área e pensamento crítico fazem parte da personalidade de alguém que assume este perfil. A criatividade e resiliência são características valorizadas, pois os desenvolvimentos de novas metodologias/aplicações nesta área atravessam muitas fases de baixa performance e a apresentação de resultados deve ser feita através de visualizações claras com a capacidade de passar inferências de forma competente.

### Exemplos de Tecnologias



## Competências

Competência	Descrição
Ferramentas de Programação	Por ser a carreira mais multidisciplinar, o Data Scientist deverá utilizar um conjunto de ferramentas focadas na produção de modelos e algoritmos, com fluência em Python, R ou MatLab, na utilização de bibliotecas de modelação e no uso de Notebooks.
Visualização e Comunicação de Dados	Tal como para o Data Analyst, as ferramentas de visualização também são bastante importantes para o Data Scientist. Contudo, a versatilidade deste perfil permite que ele escolha quais as ferramentas que pretende usar. Destacam-se algumas bibliotecas de Python, como o Plotly ou Dash, mas também o PowerBI, Tableau, QlikSense, Microsoft Excel, entre outras.
Intuição de Dados	Do mesmo modo que a intuição é importante para o Machine Learning Engineer, ela é também importante para o Data Scientist, uma vez que têm tarefas muito semelhantes. A sensibilidade e a experiência anterior trazem uma maturidade para a análise e manipulação de dados, que, posteriormente, evitam erros na interpretação e obtenção de resultados e conclusões.
Estatística	A estatística é fundamental para um Data Scientist, uma vez que é recorrendo a métodos estatísticos que a avaliação da qualidade de um modelo e de um algoritmo é efetuada, tal como as conclusões importantes a retirar dos dados.
<i>Data Wrangling</i>	O Data Scientist é responsável pela transformação e manipulação de dados durante os primeiros passos do ciclo de vida de exploração de dados e resolução de casos práticos.
<i>Machine Learning</i>	O Data Scientist lida também com tarefas relacionadas com Machine Learning, sendo essencial que esteja familiarizado com os diferentes tipos de estruturação de dados, de modelos e de algoritmos.

Competência	Descrição
Engenharia de Software	Embora importante, as competências de Engenharia de Software não são uma área de foco para o Data Scientist. No entanto, O Data Scientist deve estar atento e a par das boas práticas de desenvolvimento de software (e.g., GitOps, CI/CD, TDD).
Cálculo e Álgebra Linear	É através do Cálculo e da Álgebra que são construídos os modelos e algoritmos, bem como funções de custo e objetivo utilizadas em ML.

## Descrição de use case



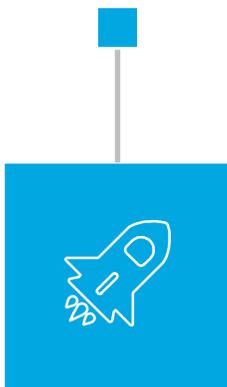
O caso prático de Data Science será o mais completo e complexo, uma vez que, no limite, poderá contemplar competências de todos os restantes perfis. O caso prático típico do Data Scientist começaria pelo processo de descoberta, onde o profissional terá de entender o problema e planear a sua solução. Depois, o processo de adquirir os dados, que, num nível mais elevado, poderia ter de recorrer à extração de uma base de dados, seguindo-se o processo de refinamento, armazenamento e, finalmente, analisar os dados que lhe serão úteis para a resolução do problema. Depois da exploração e dados, segue-se a fase de modelos e algoritmos, onde iria

ter de escolher quais os mais adequados ao seu problema e proceder a alguns testes. Com a sua escolha final de modelo, irá proceder à afinação e recolha de resultados. Por fim, poderia ainda utilizar algumas ferramentas de visualização para demonstrar as suas conclusões.

# O Playbook de Táticas

## Contexto

Fazendo uma analogia com o desporto, um Playbook é como um “livro de táticas”. Ou seja, é um manual com todas as peças e componentes que compõem a abordagem de uma pessoa ou empresa para realizar um determinado objetivo. Um Playbook em ambiente de negócio pode incluir “definições e conceitos, objetivos concretos e caminhos possíveis para lá chegar”, de forma a apoiar um profissional a desenhar uma resposta consistente a um desafio. Neste caso específico, o desafio seria iniciar e progredir as técnicas e conhecimentos de Data Science e AI.



Neste Playbook, para além da estrutura da profissão, carreiras, níveis e competências, apresentamos aqui algumas “táticas” possíveis, especialmente para uma organização ou profissional que se queira iniciar neste domínio.

As “táticas” que vos apresentamos são apenas exemplos, não exclusivos, nem exaustivos, mas que pretendem apoiar quem está a entrar neste mundo complexo, mas com vários caminhos a tomar, mais imediatos e acessíveis do que se pode pensar numa abordagem convencional.

O Playbook de táticas define quatro tipos de entidade – PME’s, Grande Empresa, Profissional e Estudante-, vários níveis de maturidade - Iniciado, Intermédio, Reskiling, Expansão e Finalista -, e quatro intervalos temporais - Já, Amanhã, Curto Prazo e Médio Prazo.

## Data Science Playbook

### Táticas Para Iniciar e Progredir Em Data Science



Entidade	Maturidade	Já (0 a 1 Mês)	Amanhã (1 a 2 Meses)	Curto Prazo (2 a 6 Meses)	Medio Prazo (6 a 12 Meses)
PME	Iniciada	1.1	1.2	1.3	1.4
	Intermédia	2.1	2.2	2.3	2.4
Grande Empresa	Iniciada	3.1	3.2	3.3	3.4
	Intermédia	4.1	4.2	4.3	4.4
Profissional	Reskilling	5.1	5.2	5.3	5.4
	Expansão	6.1	6.2	6.3	6.4
Estudante	Intermédio	7.1	7.2	7.3	7.4
	Finalista	8.1	8.2	8.3	8.4

*Figura 5 - Táticas para iniciar e progredir em Data Science*

## PME - Iniciada

### 1.1 Já (0-1 mês)

- Estudar a definição e as competências gerais;
- Realizar o autodiagnóstico e analisar os resultados;
- Interpretar os resultados e traçar um plano.

### 1.2 Amanhã (1-2 meses)

- Identificar três profissionais para especializar (e.g. diretor de IT, analista financeiro, analista comercial);
- Estudar o percurso de *Foundational*, competências, artigos, entre outros.

### 1.3 Curto Prazo (2-6 meses)

- Selecionar um curso de formação profissional (40-80h) sobre fundamentos de Data Science;
- Realizar o curso, o exame e os trabalhos práticos, se for o caso;

- Identificar um projeto de aplicação prática na empresa para cada profissional que realizou o curso.

#### 1.4 Médio Prazo (6-12 meses)

- Candidatar-se à certificação em Data Science *Foundational*;
- Marcar e realizar a preparação para o exame;
- Implementar o projeto interno individual asignado;
- Realizar o exame e obter a devida certificação.

## PME - Intermédia

#### 2.1 Já (0-1 mês)

- Realizar o autodiagnóstico e analisar o resultado;
- Interpretar o resultado e traçar um plano;
- Selecionar percursos e profissionais para desenvolver.

#### 2.2 Amanhã (1-2 meses)

- Mobilizar os quadros críticos em cada área para a realização do percurso de *Foundational*;
- Selecionar metade dos recursos mais experientes para especialização;
- Escolher os percursos para os recursos mais experientes.

#### 2.3 Curto Prazo (2-6 meses)

- Selecionar um curso de formação profissional (40-80h) sobre fundamentos de Data Science;
- Os quadros críticos devem realizar o curso, exame e trabalhos práticos, se for o caso;
- Identificar um projeto de aplicação prática na empresa para cada profissional que realizou o curso;
- Estes recursos devem então candidatar-se ao nível de certificação profissional *Foundational* e realizar o exame de certificação;
- Estudar os percursos selecionados pelos recursos mais experientes (*skills, use cases, artigos, livros, cursos disponíveis, entre outros*).

## 2.4 Médio Prazo (6-12 meses)

- Os quadros críticos devem realizar o seu projeto de aplicação prática na empresa (cada profissional que realizou o curso);
- Os profissionais mais experientes deverão selecionar individualmente o percurso de especialização pretendido e um curso de formação profissional (40-80h) sobre esse mesmo percurso;
- Os profissionais mais experientes deverão realizar o seu projeto de aplicação prática na empresa para consolidar conhecimentos;
- Os profissionais mais experientes deverão candidatar-se ao nível de certificação profissional adequado (*Foundational*, Intermédio ou Avançado), em função do seu percurso, conhecimentos e experiência prática.

## Grande Empresa - Iniciada

### 3.1 Já (0-1 mês)

- Estudar a definição e as competências gerais;
- Realizar o autodiagnóstico e analisar o resultado;
- Interpretar o resultado e traçar um plano de ação.

### 3.2 Amanhã (1-2 meses)

- Identificar um grupo multidisciplinar de profissionais (quatro a seis) para especializar (e.g. diretor de IT, analista financeiro, analista comercial, analista de produção, analista de marketing, gestor de produto, entre outros);
- Este grupo deve estudar o percurso de *Foundational*, competências, artigos, entre outros;
- Identificar 3/4 profissionais já com alguma maturidade em análise de dados e desafiá-los para uma especialização em percursos profissionais concretos.

### 3.3 Curto Prazo (2-6 meses)

- Deverá ser selecionado um curso de formação profissional (40-80h) sobre fundamentos de Data Science para este grupo inicial;
- O grupo deverá realizar o curso selecionado, exercícios, exame e trabalhos práticos, se for o caso;

- Identificar um projeto de aplicação prática na empresa para cada profissional que realizou o curso;
- Os profissionais mais experientes em Data Science deverão selecionar individualmente os percursos de especialização pretendido e um curso de formação profissional (40-80h) sobre esse mesmo percurso;
- Os profissionais mais experientes deverão realizar o seu projeto de aplicação prática na empresa para consolidar conhecimentos;
- Os profissionais mais experientes deverão candidatar-se ao nível de certificação profissional adequado (*Foundational*, Intermédio ou Avançado) em função do seu percurso, conhecimentos e experiência prática.

### 3.4 Médio Prazo (6-12 meses)

- A empresa deverá apoiar os projetos e os casos de aplicação iniciados para garantir que estes se transformam em resultados concretos para a empresa;
- A empresa deverá também aferir o grau de evolução dos colaboradores que aderiram a este processo de desenvolvimento e avaliar o seu melhor enquadramento e utilidade para a empresa;
- Em função desta análise, a empresa deverá estimular alguns destes profissionais a planearem a fase seguinte do seu desenvolvimento profissional em Data Science;
- A empresa deverá repetir o exercício, identificando um segundo grupo de profissionais iniciados e um segundo grupo de profissionais mais experientes em diferentes áreas, no sentido de iniciarem o seu processo de desenvolvimento, aprendizagem e certificação.

## Grande Empresa - Intermédia

### 4.1 Já (0-1 mês)

- Assumindo a existência de um grupo *core* já focado em Data Science, mobilizar um novo grupo multidisciplinar de profissionais a estudar as definições e competências gerais;
- Realizar o autodiagnóstico, incluindo a este novo grupo, e analisar o resultado;
- Interpretar o resultado em conjunto e atualizar e alargar o atual plano de ação.

#### 4.2 Amanhã (1-2 meses)

- Identificar neste novo grupo multidisciplinar de profissionais – quatro a seis indivíduos - para especializar (e.g. técnicos de IT, analista financeiro, analista comercial, analista de produção, analista de marketing, gestor de produto, entre outros.);
- Este grupo deve estudar o percurso de *Foundational*, competências, artigos, entre outros;
- Identificar alguns profissionais (3/4) do grupo *core*, já com alguma maturidade em analítica de dados, e desafiar a que estes se especializem em percursos profissionais concretos ou progridam na sua especialização;
- Os profissionais do grupo mais preparados devem fazer uma candidatura à certificação *Foundational* e efetuar a preparação individual adequada.

#### 4.3 Curto Prazo (2-6 meses)

- Deverá ser selecionado um curso de formação profissional (40-80h) sobre fundamentos de Data Science para o segundo grupo;
- O segundo grupo deverá realizar o curso selecionado, exercícios, exame e trabalhos práticos;
- Deverá ser identificado um projeto de aplicação prática na empresa para cada profissional do segundo grupo que realizou o curso;
- Os profissionais mais experientes em Data Science deverão selecionar individualmente os percursos de especialização pretendidos e um curso de formação profissional (40-80h) sobre esse mesmo percurso;
- Os profissionais mais experientes deverão realizar o seu projeto de aplicação prática na empresa para solidificar os conhecimentos;
- Os profissionais mais experientes deverão candidatar-se ao nível de certificação profissional adequado (Intermédio ou Avançado) em função do seu percurso, conhecimentos e experiência prática.

#### 4.4 Médio Prazo (6-12 meses)

- A empresa deverá apoiar os projetos e os casos de aplicação iniciados, e garantir que estes se transformam em resultados concretos para a empresa;
- A empresa deverá também aferir o grau de evolução dos colaboradores que aderiram a este processo de desenvolvimento e avaliar o seu melhor enquadramento e utilidade para a empresa;

- Em função desta análise, a empresa deverá estimular alguns destes profissionais a planearem a fase seguinte do seu desenvolvimento profissional em Data Science;
- A empresa deverá repetir o exercício, identificando terceiro grupo de profissionais iniciados e um terceiro grupo de profissionais mais experientes, em diferentes áreas, no sentido de iniciarem o seu processo de desenvolvimento, aprendizagem e certificação.

## Profissional a mudar de carreira – *Reskilling*

### 5.1 Já (0-1 mês)

- Estudar a descrição geral da profissão, e as suas aplicações, e avaliar o interesse em seguir uma nova carreira nesta área;
- Deverá estudar as definições e as competências gerais;
- Poderá realizar a autoavaliação e analisar o resultado;
- Interpretar o resultado e traçar um plano pessoal de desenvolvimento.

### 5.2 Amanhã (1-2 meses)

- Deverá estudar o percurso de *Foundational*, competências, artigos relacionados, entre outros.

### 5.3 Curto Prazo (2-6 meses)

- Selecionar um curso de formação profissional (40-80h) sobre fundamentos de Data Science;
- Realizar o curso, exame e trabalhos práticos, se for o caso;
- Identificar um projeto de aplicação prática na sua empresa ou num tema social de interesse público.

### 5.4 Médio Prazo (6 - 12 meses)

- Candidatar-se à certificação em *Data Science Foundational*;
- Marcar e realizar a preparação para o exame;
- Implementar o projeto interno individual asignado;
- Realizar o exame e obter o certificado;
- Procurar ativamente novas funções na sua empresa ou num empregador que necessite dos novos *skills* adquiridos.

# Profissional - a Expandir a sua Carreira

## 6.1 Já (0-1 mês)

- Estudar a descrição geral da profissão, as suas aplicações e avaliar as áreas de maior aplicabilidade à sua profissão atual (analista, *controller*, especialista, em funções financeiras, comerciais, operacionais, IT, entre outros);
- Deverá estudar as definições e competências gerais;
- Poderá realizar a autoavaliação e analisar o resultado;
- Interpretar o resultado, selecionar o percurso profissional em que pretende desenvolver-se e traçar um plano pessoal de desenvolvimento.

## 6.2 Amanhã (1-2 meses)

- Caso tenha adquirido os conhecimentos e alguma experiência anteriormente, deve avaliar se deve candidatar-se ao nível intermédio do percurso selecionado e fazer a sua candidatura;
- Caso contrário, deverá estudar o percurso de *Foundational*, competências, artigos relacionados, entre outros;
- Selecionar um curso de formação profissional (40-80h) sobre fundamentos de Data Science;
- Realizar o curso, exame e trabalhos práticos, se for o caso.

## 6.3 Curto Prazo (2-6 meses)

- Deverá identificar um projeto de aplicação prática na sua empresa e implementar o projeto interno individual assignado;
- Candidatar-se à certificação em Data Science *Foundational*, marcar e realizar a preparação para o exame;  
Realizar o exame e obter a certificação.

## 6.4 Médio Prazo (6-12 meses)

- Após a certificação, deverá identificar um projeto de aplicação prática na sua empresa e implementá-lo para sedimentar os conhecimentos adquiridos;
- Deverá selecionar um curso de formação profissional ou académica (60-120h) sobre o percurso selecionado (e.g. Data Analyst, Data Engineer, entre outros);
- Deverá realizar o curso, o exame e os trabalhos práticos, se for o caso;

- Deverá identificar e realizar os projetos de aplicação prática na sua empresa dos temas selecionados, adquirindo experiência prática;
- Candidatar-se à certificação de nível intermédia na especialidade escolhida, marcar e realizar a preparação para o exame;
- Realizar o exame e obter a devida certificação;
- Deverá procurar ativamente ajustar as suas funções da empresa às competências adicionais adquiridas.

## Estudante - Intermédio

### 7.1 Já (0-1 mês)

- Estudar a descrição geral da profissão e as suas aplicações. Avaliar também as áreas de maior aplicabilidade à sua formação e às suas preferências (Engenharia, Gestão, Economia, Matemática, Biologia, Enfermagem, entre outras);
- Deverá estudar as definições e competências gerais.

### 7.2 Amanhã (1 - 2 meses)

- Deverá estudar o percurso de *Foundational*, competências, artigos relacionados, entre outros;
- Selecionar um curso de formação profissional (40-80h) sobre fundamentos de Data Science.

### 7.3 Curto Prazo (2 - 6 meses)

- Realizar o curso, exame e trabalhos práticos, se for o caso;
- Candidatar-se à certificação em Data Science *Foundational*, marcar e realizar a preparação para o exame;
- Realizar o exame e obter a devida certificação;
- Deverá identificar um projeto de aplicação prática na sua universidade ou numa empresa que lance um desafio à escola, nos quais possa ter seguimento para um possível estágio.

### 7.4 Médio Prazo (6 - 12 meses)

- Após a certificação deverá candidatar-se a um estágio académico para uma função relacionada com Data Science numa empresa ou setor que o motive;

- Um critério de escolha deve ser a possibilidade de aplicação em projetos de conhecimentos adquiridos para continuar a sua progressão;
- Deverá estudar os diferentes percursos profissionais e selecionar aquele em que pretende desenvolver-se para traçar um plano pessoal de desenvolvimento;
- Deverá selecionar um curso de formação profissional ou académica (60-120h) sobre o percurso selecionado (e.g. Data Analyst, Data Engineer, entre outros);
- Deverá identificar e realizar projetos de aplicação prática na empresa onde realiza o estágio sobre os temas selecionados, adquirindo experiência prática;
- Deverá realizar o curso, exame e trabalhos práticos, se for o caso.

## Estudante - Finalista

### 8.1 Já (0-1 mês)

- Estudar a descrição geral da profissão e as suas aplicações, e avaliar as áreas de maior aplicabilidade à sua formação e às suas preferências (Engenharia, Gestão, Economia, Matemática, Biologia, Enfermagem, entre outras);
- Deverá estudar as definições e as competências gerais;
- Poderá realizar a autoavaliação e analisar o resultado;
- Interpretar o resultado, selecionar o percurso profissional em que pretende desenvolver-se e traçar um plano pessoal de desenvolvimento.

### 8.2 Amanhã (1 - 2 meses)

- Caso tenha já conhecimentos e alguma experiência (e.g. alguns ramos de Engenharia, Matemática, entre outros), deve avaliar se deve candidatar-se ao nível intermédio do percurso selecionado e fazer a sua candidatura;
- Caso contrário, deverá estudar o percurso de *Foundational*, competências, artigos relacionados, entre outros;
- Selecionar um curso de formação profissional (40-80h) sobre fundamentos de Data Science;
- Realizar o curso, exame e trabalhos práticos, se for o caso.

### 8.3 Curto Prazo (2 - 6 meses)

- Deverá identificar um projeto de aplicação prática numa empresa que o aceite para um estágio ou tenha um desafio concreto adequado ao seu perfil;
- Candidatar-se à certificação em Data Science *Foundational*, marcar e realizar a preparação para o exame, realizar o exame e obter o certificado.

### 8.4 Médio Prazo (6 - 12 meses)

- Após a certificação, deverá candidatar-se a uma evolução do seu estágio ou candidatar-se a uma posição de Data Science numa empresa ou setor que o motive;
- Um critério de escolha deve ser a possibilidade de aplicação em projetos de conhecimentos adquiridos para continuar a sua progressão;
- Deverá selecionar um curso de formação profissional ou académica (60-120h) sobre o percurso selecionado (e.g. Data Analyst, Data Engineer, entre outros);
- Deverá realizar o curso, exame e trabalhos práticos, se for o caso;
- Deverá identificar e realizar projetos de aplicação prática na sua empresa do percurso selecionado adquirindo experiência prática;
- Deverá estudar os diferentes percursos profissionais e deverá selecionar aquele em que pretende desenvolver-se, e traçar um plano pessoal de desenvolvimento;
- Candidatar-se à certificação de nível intermédio na especialidade escolhida, marcar e realizar a preparação para o exame, realizar o exame e obter a devida certificação;
- Deverá procurar ativamente ajustar as suas funções na empresa às competências adicionais adquiridas.

# Certificação Profissional

Face à crescente necessidade no mercado em guiar os profissionais e as empresas a responderem à contínua evolução associada ao paradigma de Data Science e AI, a DSPA apresenta um percurso de certificação que visa criar um *roadmap* de capacitação para os vários perfis e domínios associados à prática.

## Níveis de certificação

Para cada um dos quatro perfis de Data Science e AI, haverá um nível transversal comum a todos os perfis, o *Foundational*, e dois níveis de certificação possíveis para cada perfil, com um grau de exigência incremental para as competências do perfil e para o tempo de experiência profissional no domínio.

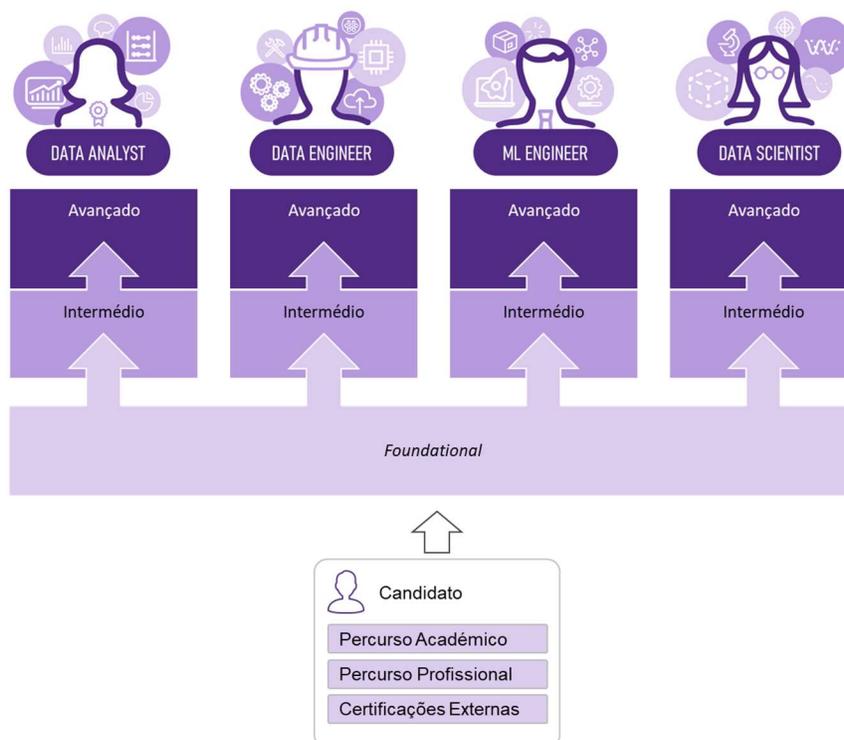


Figura 6 - Estrutura de certificação DSPA

- **Foundational:** Conhecimento base nas competências fundamentais comuns a todos os perfis, sem exigências de experiência profissional no domínio;
- **Intermédio:** Conhecimento sólido nas competências fundamentais do perfil com experiência profissional no domínio;
- **Avançado:** Conhecimento avançado nas competências do perfil com extensiva experiência profissional no domínio.

Ao candidatar-se para a certificação, serão requeridas informações declarativas referentes ao percurso académico, percurso profissional e certificações de especialidade.

### **Foundational**

O nível de certificação *Foundational* apresenta um método de avaliação específico, com um exame de escolha múltipla, onde o profissional será avaliado nas várias áreas de competência associadas ao paradigma de Data Science e AI.

Os critérios de avaliação do exame de escolha múltipla para o nível *Foundational* são:

<b>Critérios do exame de escolha múltipla</b>	<b>Valores</b>
Duração	90 minutos
Nota <i>target</i>	70% por área de competência
Nº de questões	80 questões
Conteúdo	10 questões por competência

A certificação terá uma validade de quatro anos e não está planeada a hipótese de renovação de certificação (e.g. para renovar após quatro anos, o profissional deverá realizar novamente o exame de certificação).

## **Intermédio**

O nível de certificação Intermédio apresenta três métodos de avaliação, um exame de escolha múltipla, nos quais o profissional será avaliado nas várias áreas de competência associadas ao perfil de Data Science e AI selecionado. O profissional terá de proceder à resolução de um problema e à respetiva defesa e apresentação do mesmo.

Os critérios de avaliação do exame de escolha múltipla para o nível Intermédio são:

<b>Critérios do exame de escolha múltipla</b>	<b>Valores</b>
Duração	120 minutos
Nota <i>target</i>	75% por área de competência
Nº de questões	60 questões
Conteúdo	Por competência relevante

Para o método de resolução de problema, o profissional terá que desenvolver uma solução para um problema referente ao perfil de Data Science e AI, no qual é expectável que sejam evidenciadas as competências técnicas do profissional. Para o problema, a DSPA fornecerá um enunciado, juntamente com um *dataset* base, onde serão especificados o desafio e o *output* pretendido.

Os critérios de avaliação da resolução do problema para o nível Intermédio são:

<b>Critérios de resolução do problema</b>	<b>Valores</b>
Tempo de resolução	7 dias
Tempo de resposta	15 dias
Nota <i>target</i>	75%
Critério de avaliação	1. Match entre resultado obtido e resultado esperado 2. Qualidade do trabalho desenvolvido

Para o método de defesa do problema, o profissional irá defender o problema resolvido, no qual a DSPA pretende avaliar a capacidade de articulação e aprofundar o conhecimento do profissional nas temáticas do perfil de Data Science.

Os critérios de avaliação da defesa do problema para o nível Intermédio são:

Critérios da defesa do problema	Valores
Duração	30 minutos a 60 minutos
Nota <i>target</i>	75%
Critério de avaliação	1. Metodologia / abordagem 2. Estrutura da solução 3. Domínio da solução

A avaliação para o nível intermédio será realizada da seguinte forma:

- Realização do exame de escolha múltipla (nota *target* de 75%);
- Caso o profissional atinja a nota mínima, realização do problema e defesa do problema (nota *target* de 75%).

A certificação terá uma validade de três anos com hipótese de renovação. Para o profissional renovar a certificação, será submetido a uma nova avaliação do exame de escolha múltipla de nível Intermédio para o perfil de Data Science e AI, estando dispensada a realização dos restantes métodos de avaliação.

## **Avançado**

O nível de certificação Avançado apresenta um método de avaliação, através da defesa de um *use case*. Para este método de avaliação, o objetivo da DSPA será avaliar a experiência em contexto real do profissional, onde será avaliada a sofisticação dos desafios encontrados e o domínio na matéria por parte do profissional.

Os critérios de avaliação da defesa do *use case* para o nível Avançado são:

Critérios da defesa do problema	Valores
Duração	30 minutos a 60 minutos
Nota <i>target</i>	75%
Critério de avaliação	<ol style="list-style-type: none"> <li>1. Metodologia / abordagem</li> <li>2. Estrutura da solução</li> <li>3. Domínio da solução</li> </ol>

A certificação terá uma validade de três anos com hipótese de renovação. Para o profissional renovar a certificação, será submetido a uma nova avaliação do exame de escolha múltipla de nível Intermédio para o perfil de Data Science e AI, estando dispensada a realização dos restantes métodos de avaliação.



## *Journey* e opções de candidatura

No percurso de certificação da DSPA, são disponibilizados dois métodos de candidatura: o *standard* e o *fast-track*.

## Standard

No método *standard*, a certificação *Foundational* serve como o nível de certificação base, que tem como objetivo certificar o profissional nas competências chave na temática de Data Science e AI. Após adquirida a certificação do nível *Foundational*, o profissional terá a opção de selecionar as quatro certificações do nível Intermédio para os perfis de Data Science e AI. Para a certificação de nível Avançado de cada perfil, existem três pré-requisitos necessários:

- Adquirir a certificação do nível intermédio do perfil;
- Apresentar mais de quatro anos de experiência relevante no desempenho do perfil;
- Submeter duas a três referências que comprovem a experiência realizada.

	Standard journey				
	Validade (anos)	Renovação	Pré requisito	Anos de experiência	Referenciação
Avançado	3	Exame intermédio	Intermédio do role	> 4	2 - 3
Intermédio	3	Exame intermédio	Foundational	N/A	N/A
Foundational	4	N/A	N/A	N/A	N/A

Figura 7 - Journey standard

## Fast-track

No método *fast-track*, a DSPA disponibiliza um método de acesso a certificações de nível Intermédio e Avançado para os perfis de Data Science e AI a profissionais que apresentem um conjunto de pré-requisitos adicionais que evidenciam experiência tácita nos domínios da certificação. A DSPA avaliará os pré-requisitos e, após confirmação da sua validade, será atribuído ao profissional a realização da certificação para o nível desejado. Os pré-requisitos são não acumulativos, pelo que

se o profissional desejar utilizar este método para diversos níveis e perfis de Data Science e AI, deverá apresentar novas evidências por cada pedido.

	Validade (anos)	Renovação	Fast-track		
			Pré requisito	Anos de experiência	Referenciação
Avançado	3	Exame intermédio	N/A	> 6	2 - 3
Intermédio	3	Exame intermédio	N/A	> 4	2 - 3
Foundational	4	N/A	Não é possível realizar <i>fast-track</i> para <i>Foundational</i>		

Figura 8 - Journey fast-track

# Recursos Formativos

## Oferta formativa por perfil e nível de competência

Na tabela, encontra-se uma lista de cursos recomendados, específicos para cada carreira, mas aconselhados para todos aqueles que pretendem alargar o domínio de conhecimento em Data Science e AI:

Perfil	Grau	Nome do Curso	Entidade	Descrição
ML Engineer	Foundational	Amazon Web Services/Google Cloud Platform/Azure	Udemy ( <a href="https://www.udemy.com/course/exam-microsoft-azure-4/">https://www.udemy.com/course/exam-microsoft-azure-4/</a> )	A Udemy disponibiliza uma série de cursos sobre o stack tecnológico AWS/Google Cloud Platform/Azure com vista à preparação para certificações de diferentes níveis nestas tecnologias. Considerando o stack tecnológico Microsoft, sugere-se a certificação de <i>Azure AI Fundamentals</i> para este nível.
ML Engineer	Intermédio	Amazon Web Services/Google Cloud Platform/Azure	Udemy ( <a href="https://www.udemy.com/course/ai-100-designing-and-implementing-an-azure-ai-solution/">https://www.udemy.com/course/ai-100-designing-and-implementing-an-azure-ai-solution/</a> )	A Udemy disponibiliza uma série de cursos sobre o stack tecnológico AWS/Google Cloud Platform/Azure com vista à preparação para certificações de diferentes níveis nestas tecnologias. Considerando o stack tecnológico Microsoft, sugere-se a certificação de <i>Designing and Implementing a Microsoft Azure AI Solution</i> para este nível.
ML Engineer	Avançado	Amazon Web Services/Google Cloud Platform/Azure	Udemy ( <a href="https://www.udemy.com/course/aws-machine-learning/">https://www.udemy.com/course/aws-machine-learning/</a> )	A Cloud Guru disponibiliza uma série de cursos sobre o stack tecnológico AWS/Google Cloud Platform/Azure com vista à preparação para certificações de diferentes níveis nestas tecnologias. Considerando o stack tecnológico AWS, sugere-se a certificação de <i>Machine Learning Specialty</i> para este nível.

Perfil	Grau	Nome do Curso	Entidade	Descrição
Data Analyst	Foundational	Amazon Web Services/Google Cloud Platform/Azure	Cloud Guru ( <a href="https://acloudguru.com/course/aws-cloud-practitioner-clf-c01-la">https://acloudguru.com/course/aws-cloud-practitioner-clf-c01-la</a> )	A Cloud Guru disponibiliza uma série de cursos sobre o stack tecnológico AWS/Google Cloud Platform/Azure com vista à preparação para certificações de diferentes níveis nestas tecnologias. Considerando o stack tecnológico AWS, sugere-se a certificação de <i>Cloud Practitioner</i> para este nível.
Data Analyst	Intermédio	Data Analyst in R	Dataquest ( <a href="https://www.dataquest.io/path/data-analyst-r/">https://www.dataquest.io/path/data-analyst-r/</a> )	Este curso aborda diversos tópicos na carreira de Data Analyst especializado em R, tais como: <ul style="list-style-type: none"> <li>• Conceitos de programação de nível iniciado e intermédio;</li> <li>• <i>Workflows</i> em R com recurso a bibliotecas como RStudio e tidyverse;</li> <li>• Conceitos de probabilidade e estatística;</li> <li>• Limpeza e visualização de dados;</li> <li>• Ferramentas colaborativas como Git ou bases de dados SQL.</li> </ul>
Data Analyst	Avançado	Amazon Web Services/Google Cloud Platform/Azure	Cloud Guru ( <a href="https://acloudguru.com/course/aws-certified-data-analytics-specialty">https://acloudguru.com/course/aws-certified-data-analytics-specialty</a> )	A Cloud Guru disponibiliza uma série de cursos sobre o stack tecnológico AWS/Google Cloud Platform/Azure com vista à preparação para certificações de diferentes níveis nestas tecnologias. Considerando o stack tecnológico AWS, sugere-se a certificação de <i>Data Analytics Specialty</i> para este nível.
Data Scientist	Foundational	Data Scientist in Python	Dataquest ( <a href="https://www.dataquest.io/path/data-scientist/">https://www.dataquest.io/path/data-scientist/</a> )	Este curso aborda fundamentos de Python, análise e visualização de dados, bases de dados SQL, conceitos de probabilidade e estatística e construção de modelos de aprendizagem automática. Não é exigida qualquer experiência prévia e é acompanhado de exemplos práticos e exercícios para o aluno.
Data Scientist	Intermédio	Amazon Web Services/Google Cloud Platform/Azure	Udemy ( <a href="https://www.udemy.com/course/dp100-azure/">https://www.udemy.com/course/dp100-azure/</a> )	A Udemy disponibiliza uma série de cursos sobre o stack tecnológico AWS/Google Cloud Platform/Azure com vista à preparação para certificações de diferentes níveis nestas tecnologias. Considerando o stack tecnológico Microsoft, sugere-se a certificação de <i>Azure Data Scientist Associate</i> para este nível.
Data Scientist	Avançado	SAS Certified Data Scientist	SAS Global Certification Program ( <a href="https://www.sas.com/en_us/training/academy-data-science.html">https://www.sas.com/en_us/training/academy-data-science.html</a> )	Criado para profissionais com capacidade para manipular e retirar valor de grandes quantidades de dados com uma grande variedade de ferramentas, capazes de fazer recomendações com base em complexos modelos de <i>Machine Learning</i> e implementar soluções escaláveis dentro do ambiente SAS. Este é um programa que oferece três certificações de nível profissional em <i>Data Curation</i> , <i>Advanced Analytics</i> e <i>AI &amp; Machine Learning</i> .

Perfil	Grau	Nome do Curso	Entidade	Descrição
Data Engineer	Foundational	Amazon Web Services/Google Cloud Platform/Azure	Udemy  ( <a href="https://www.udemy.com/course/azure-dp-900/">https://www.udemy.com/course/azure-dp-900/</a> )	A Udemy disponibiliza uma série de cursos sobre o stack tecnológico AWS/Google Cloud Platform/Azure com vista à preparação para certificações de diferentes níveis nestas tecnologias. Considerando o stack tecnológico Microsoft, sugere-se a certificação de <i>Azure Data Fundamentals</i> para este nível.
Data Engineer	Intermédio	Amazon Web Services/Google Cloud Platform/Azure	Udemy  ( <a href="https://www.udemy.com/course/google-cloud-certification-associate-cloud-engineer/">https://www.udemy.com/course/google-cloud-certification-associate-cloud-engineer/</a> )	A Udemy disponibiliza uma série de cursos sobre o stack tecnológico AWS/Google Cloud Platform/Azure com vista à preparação para certificações de diferentes níveis nestas tecnologias. Considerando o stack tecnológico Google, sugere-se a certificação de <i>Associate Cloud Engineer</i> para este nível.
Data Engineer	Avançado	Amazon Web Services/Google Cloud Platform/Azure	A Cloud Guru  ( <a href="https://acloudguru.com/course/google-cloud-certified-professional-data-engineer-la">https://acloudguru.com/course/google-cloud-certified-professional-data-engineer-la</a> )	A Cloud Guru disponibiliza uma série de cursos sobre o stack tecnológico AWS/Google Cloud Platform/Azure com vista à preparação para certificações de diferentes níveis nestas tecnologias. Considerando o stack tecnológico Google, sugere-se a certificação de <i>Professional Data Engineer</i> para este nível.

É importante ressaltar que a inclusão destes cursos deve ser encarada como referência ou *guideline*, não obedecendo qualquer um deles a um carácter obrigatório para alguém ser considerado para o perfil em questão, apesar de ser altamente recomendado que se siga o percurso aqui proposto. Acresce o facto de haver múltiplas plataformas diferentes para preparação de cada uma das certificações sugeridas, o que faz com que a coluna Entidade sirva de mero exemplo.

# Apoios e Agradecimentos

## Quem nos apoiou?

Este Projeto, cujo conceito, objetivos estratégicos e objetivos operacionais foram desenvolvidos pela DSPA – Data Science Portuguese Association, foi viabilizado e realizado graças ao apoio de diversas entidades:

- **Portugal 2020**, através do Sistema de Apoio a Ações Coletivas (eixo da Qualificação), no âmbito do Programa Operacional Competitividade e Internacionalização, com um financiamento parcial a fundo perdido enquadrado na prioridade de investimento na qualificação e inovação das PME através da concessão de apoio à criação e ao alargamento de capacidades avançadas de desenvolvimento de produtos e serviços das PME;
- **Câmara Municipal do Fundão**, entidade e município que desde cedo apoiou a DSPA no desenvolvimento e *empowerment* de uma visão estratégica e de abrangência alargada, regional e nacional, disponibilizando recursos para a operacionalização de projetos e convocando a DSPA para fóruns de debate com vista à implementação de metodologias, ferramentas e iniciativas potenciadoras da criação e alargamento do conhecimento e das capacidades dos ecossistemas de formação e ensino, social e económico;
- Ao longo da implementação das ações e iniciativas do Projeto, a DSPA tem vindo a contar com o apoio, participação e contributos de representantes de entidades de naturezas tão diversas e representativas como Câmaras Municipais das Regiões Centro e Norte, Comissões Coordenadoras do Desenvolvimento Regional do Norte e do Centro, Associações Empresariais, Laboratórios de Investigação, Laboratórios Colaborativos e Centros de Interface Tecnológico, Agências de Investimento Regionais, empresas de Grande, Média e Pequena dimensão, e Profissionais, a título individual.

## Agradecimentos

A equipa coordenadora do Projeto e a Direção da DSPA gostaria de registar o reconhecimento devido a todas as pessoas e entidades que, a par das anteriores, colaboraram em regime de prestação de serviços e em regime pro-bono no seu desenvolvimento, destacando-se entre os últimos alguns dos membros fundadores e dos órgãos sociais da DSPA.

Paralelamente, registamos também um agradecimento muito especial aos membros convidados do Comité de Supervisão e Aconselhamento Estratégico do Projeto que, conforme nota anterior, conferiram em regime pro-bono a qualidade e fiabilidade da utilização e da utilidade prática das ferramentas desenvolvidas no âmbito do Projeto: **António Vidigal**, ex-CEO da EDP Inovação e atual consultor nos setores da Energia e das TIC, **Bernardo Caldas**, Head of Data Science & AI do NOVO BANCO, **Gonçalo Marques Oliveira**, Executive Committee Member e CIO do Pestana Hotel Group, **Leid Zejnilovic**, Assistant Professor na Nova SBE e Chairman da DSSG Foundation, e **Ricardo Chaves**, Chief Commercial Officer da SIBS.

## Onde nos inspiramos?

Como em tantas matérias na atualidade, a riqueza da informação disponível para pesquisa e aprendizagem é enorme e, para este projeto, a DSPA inspirou-se em quatro grandes grupos de organizações e movimentos.

As ordens profissionais que acreditam oficialmente profissionais para o exercício das profissões e que, embora assimétricas no seu modelo operativo e utilidade, contribuem para uma sociedade que necessita de maior regulação para protegermos os consumidores e o desenvolvimento económico. Alguns exemplos:

- <https://www.occ.pt/pt/inscricao/exames-de-avaliacao-profissional/>
- <https://portal.oa.pt>
- <https://www.ordemenfermeiros.pt/faqs/certificacao-individual-de-competencias/>

As empresas de tecnologia que, por via da sua inovação em software e hardware, têm vindo a promover a aprendizagem e certificação de profissionais que possam operar os seus produtos. Alguns exemplos:

- <https://www.cisco.com/c/en/us/training-events/training-certifications/certifications.html>
- <https://aws.amazon.com/certification/>
- <https://docs.microsoft.com/en-us/learn/certifications/browse/>
- <https://trailhead.salesforce.com/credentials/administratoroverview/>

As empresas de formação em modelos MOOC (Massive Online Open Courses) que, por via do seu modelo inovador de conteúdos abertos e estruturados, vieram reinventar o mundo da micro aprendizagem e micro certificação, alargando a milhões de novos alunos um conhecimento antes restrito a quem tinha meios ou proximidade ao sistema de ensino. Alguns exemplos:

- <https://www.coursera.org/degrees>
- <https://www.edx.org/verified-certificate>
- <https://www.udacity.com/course/data-scientist-nanodegree--nd025>
- <https://business.udemy.com>

Outras associações profissionais, algumas precursoras numa era pré-tecnológica, em especialidades como a Logística ou Produção Industrial, ou mais recentemente entidades congéneres à DSPA, que em outros continentes vieram inspirar este trabalho. Referir também publicações de regulamentação governamental, como o do governo de UK que, para um fim específico, serviu também de forte inspiração para esta publicação. Alguns exemplos:

- <https://www.ascm.org/learning-development/certifications-credentials/>
- <https://www.iaop.org/Content/23/193/3037/>
- <https://www.gov.uk/guidance/data-scientist>
- <https://www.adasci.org>
- <https://www.dasca.org/about-data-science-council-of-america>
- <https://www.iccp.org/index.html>

## Autoria

Este documento foi elaborado pela Deloitte Portugal, em parceria com DSPA – Data Science Portuguese Association, tendo por base artigos e informação de entidades como Harvard Business Review, Gartner, Forrester, The MIT Press, Oxford University Press, entre outros. A partir da informação recolhida foi reunido neste documento um

conjunto de definições relativas ao âmbito de Data Science, bem como linhas gerais orientadoras das carreiras de Data Analyst, Machine Learning Engineer, Data Engineer e Data Scientist.