L'AKADÉMIE DU DIGITAL · PROGRAMA DE FORMAÇÃO Certificação Qualiopi em preparação

Data Scientist

FOR-DEV-M5-A · Desenvolvimento, DevOps e Dados

Do dado ao modelo: analisar, modelizar e prever com a ciência de dados

DURAÇÃO
35 horas — 5 dias

PREÇO
2?450 € s/ IVA

PÚBLICO
Escolas e centros de formação

PRAZO DE ACESSO
Acesso em 2 dias úteis após validação da necessidade e dos pré-requisitos.

Preço estabelecido com base em 35 h a 70 € s/ IVA por hora. Percurso dedicado a escolas e centros de formação, estabelecido por orçamento; preço indicativo e negociável consoante o volume e o contexto.

O módulo alterna sequências teóricas e oficinas práticas ao longo de cada dia. Abrange todo o ciclo de um projeto de dados: manipulação e limpeza dos dados, análise exploratória, estatística, machine learning supervisionado e não supervisionado, avaliação e apresentação de resultados. O fio condutor é Python/scikit-learn, complementado por um panorama do ecossistema de dados.

Objetivos pedagógicos

- Compreender o papel do data scientist e o ciclo de vida de um projeto de dados
- Manipular, limpar e explorar dados com Python (pandas, NumPy)
- Conduzir uma análise exploratória e produzir visualizações pertinentes
- Compreender os fundamentos estatísticos úteis para a modelização
- Construir, treinar e avaliar modelos de machine learning (supervisionado e não supervisionado)
- Apresentar e interpretar os resultados, conhecendo os limites, os enviesamentos e o ecossistema de ferramentas

Público-alvo

- Estudantes e formandos das escolas e centros de formação parceiros
- Perfis técnicos ou analíticos que queiram orientar-se para a ciência de dados
- Programadores, analistas ou estatísticos em evolução de competências

Pré-requisitos

São exigidas bases de programação (de preferência Python — módulo «Fundamentos» recomendado), bem como à-vontade com a lógica e os números. Noções de matemática e de estatística facilitam a progressão. Estes elementos são validados por um formulário online e / ou por uma entrevista prévia.

- Dominar as bases da programação (de preferência Python).

- Sentir-se à vontade com a lógica, os números e o raciocínio analítico.
- Ter noções de matemática / estatística (uma mais-valia).
- Ter autonomia de trabalho e curiosidade pelos dados.

Programa

Dia 1 — A profissão e a manipulação dos dados

Profissão e ecossistema de dados	1h15	Papel do data scientist, ciclo de vida de um projeto; panorama das linguagens (Python, R, SQL) e dos ambientes (Jupyter, Colab, Databricks). Exposição interativa, panorama das ferramentas, exemplos de projetos.	Perguntas orais; questionário de escolha múltipla sobre o ecossistema.
Familiarização com o ambiente	2h15	Descobrir o ambiente e carregar conjuntos de dados. Trabalho prático guiado (Jupyter), importação de dados, primeiros passos.	Carregamento de dados validado.
Manipulação com pandas	1h00	Estruturas de dados, indexação, transformação. Exposição, demonstração, exemplos pandas/NumPy.	Questionário de escolha múltipla sobre pandas.
Transformar um conjunto de dados	2h30	Manipular e transformar um conjunto de dados real. Trabalho prático, conjunto de dados real, operações progressivas.	Manipulações validadas.

Dia 2 — Limpeza e análise exploratória (EDA)

Qualidade e preparação dos dados	1h15	Valores em falta, outliers, feature engineering. Exposição, exemplos de dados sujos, métodos de limpeza.	Perguntas orais; questionário de escolha múltipla.
Limpar um conjunto de dados	2h15	Tratar valores em falta, outliers e incoerências. Trabalho prático guiado, conjunto de dados com ruído, acompanhamento.	Conjunto de dados limpo validado.
EDA e visualização	1h00	Análise exploratória; matplotlib, seaborn, abertura ao Plotly e a ferramentas de BI (Power BI, Tableau). Exposição, exemplos de visualizações, panorama das ferramentas de visualização.	Questionário de escolha múltipla sobre a EDA e a visualização.
Conduzir uma EDA	2h30	Explorar e visualizar um conjunto de dados. Trabalho prático, produção de gráficos, interpretação.	EDA e visualizações validadas.

Dia 3 — Estatística e fundamentos do ML

Estatística para a ciência de dados	1h15	Estatística descritiva e inferencial úteis. Exposição, exemplos numéricos, interpretação.	Perguntas orais; questionário de escolha múltipla.
Análise estatística	2h00	Analisar estatisticamente um conjunto de dados. Trabalho prático guiado, cálculos e testes, interpretação.	Análise estatística validada.
Princípios do machine learning	1h15	Supervisionado/não supervisionado, train/test, sobreajuste. Exposição, esquemas, exemplos, panorama scikit-learn / XGBoost / LightGBM.	Questionário de escolha múltipla sobre os princípios de ML.
Primeiro modelo	2h00	Construir um primeiro modelo simples (regressão ou classificação). Trabalho prático guiado em scikit-learn, conjunto de dados preparado.	Primeiro modelo validado.

Dia 4 — Modelização supervisionada e avaliação

Algoritmos supervisionados	1h15	Regressão, árvores, k-NN…; pipeline de modelização. Exposição, comparação de algoritmos, esquema de pipeline.	Perguntas orais; questionário de escolha múltipla.
Treinar modelos	2h15	Treinar vários modelos supervisionados. Trabalho prático guiado em scikit-learn, comparação de modelos.	Modelos treinados validados.
Avaliação e otimização	1h00	Métricas, validação cruzada, hiperparâmetros. Exposição, matrizes de confusão, curvas, exemplos.	Questionário de escolha múltipla sobre a avaliação.
Avaliar e otimizar	2h30	Avaliar e otimizar um modelo. Trabalho prático, validação cruzada, ajuste dos hiperparâmetros.	Modelo avaliado e otimizado validado.

Dia 5 — Não supervisionado, interpretação e projeto

Não supervisionado e ética	1h00	Clustering, redução de dimensão; enviesamentos, limites, ética dos modelos. Exposição, exemplos de clustering, casos de enviesamento.	Questionário de escolha múltipla sobre o não supervisionado e a ética.
Clustering	2h00	Aplicar um clustering e interpretar os resultados. Trabalho prático guiado, conjunto de dados, interpretação.	Clustering interpretado validado.
Apresentação de resultados e data storytelling	1h00	Apresentar e comunicar os resultados aos decisores. Exposição, princípios de data storytelling, exemplos.	Questionário de escolha múltipla sobre a apresentação de resultados.
Miniprojeto de dados e apresentação de resultados	2h30	Conduzir um miniprojeto de dados de ponta a ponta e apresentá-lo. Trabalho prático de projeto, apresentações, revisão coletiva, questionário final de escolha múltipla, inquérito de satisfação.	Avaliação sumativa: projeto de dados + questionário final de escolha múltipla.

Modalidades e meios pedagógicos

Formação presencial nas nossas instalações ou nas suas, em grupo reduzido. Alternância de contributos teóricos, oficinas práticas e simulações.

Material necessário :

16 GB de RAM recomendados

- Os formandos devem dispor de um PC com Wi-Fi e 16 GB de RAM recomendados.
- Efetivo adaptado ao formato escola / centro de formação.
- Percurso dedicado a escolas e centros de formação, estabelecido por orçamento.

Modalidades de avaliação

Avaliação das aprendizagens durante e no final da formação; certificado de conclusão entregue a cada formando.

Acessibilidade

Formação acessível a pessoas com deficiência. Contacte-nos antes da inscrição para estudarmos as adaptações necessárias.

Financiamento

Orçamento ou convenção de formação a pedido. A certificação Qualiopi está em preparação: a elegibilidade para financiamentos OPCO será aberta após a sua obtenção.

Auditoria de certificação prevista para T1 2027. Sem elegibilidade OPCO antes da sua obtenção.