

Análise Exploratória de Dados

- **Formato do curso:** Presencial e Live Training
- **Preço:** 920€
- **Duração:** 18 horas

A Análise Exploratória de Dados é uma das principais fases da estrutura de um projecto de Data Science. Contribui para uma diminuição das dimensões do conjunto de dados, tornando o problema menos complexo, e ainda a levar menos tempo a efectuar o treino dos modelos de machine learning, num passo subsequente. Além disso, a Análise Exploratória de Dados é uma actividade fundamental para termos uma primeira noção do tipo de dados que temos à nossa disposição.

Assim, os principais objectivos desta análise são:

- Maximizar os insights no conjunto de dados
- Relevar estruturas e padrões características do conjunto de dados
- Detectar anomalias e outliers
- Extrair e seleccionar variáveis mais importantes
- Avaliar premissas existentes (e.g., intuições de negócio actuais)
- Preparar todo o conjunto de dados para um modelo mais simples e com capacidade preditiva e explicativa elevada

Destinatários

- Data Scientists / Data Analysts / Business Analysts
- Managers
- IT

Pré-requisitos

- Programação Básica em Python (com conhecimentos de pandas e numpy)
- Noções básicas de Estatística

Objectivos

No final da ação de formação os participantes deverão estar aptos a:

- Perceber o que é a Análise Exploratória de Dados e como se encaixa no workflow de Data Science
 - Criar visualizações objetivas e eficazes de dados que resultem em ações concretas
 - Criar sumários estatísticos de dados
 - Seleccionar e justificar variáveis mais relevantes para utilizar num modelo de machine learning
 - Reduzir a dimensão do conjunto de dados através de seleção e extração de variáveis
 - Construir uma *pipeline* composto por diferentes módulos de pré-processamento de dados
-

Metodologia

Aulas expositivas devidamente complementadas com trabalhos práticos.

Existirão casos de estudo com debate entre os alunos e orientados pela docente. Estes casos consistem na apresentação da situação de uma organização (real ou ficcionada) onde são descritos os desafios que se lhe colocam. Os alunos deverão propor uma solução para o problema de acordo com a matéria exposta e o seu estudo.

Programa

- Visualização de dados para Análise Exploratória de Dados
- Sumário estatístico dos dados
- Seleção de variáveis preditivas
- Redução de dimensão/número de variáveis
- Data Science Pipelines