



CONTEÚDOS PROGRAMÁTICOS DETALHADOS

Objetivo do Curso:

Continuar-se-á o curso anterior, usando os mesmo materiais dos exercícios, com o objectivo de montar um projecto de Data Prepping, que consistirá em reunir dados de múltiplas fontes de forma automatizada, tratá-los e uni-los numa fonte de dados única, preparada para análise posterior na óptica de Data Science. A abrangência do projecto permitirá contactar com uma grande quantidade de módulos de Python, oferecendo um overview bastante completo da oferta de soluções que estão disponíveis na linguagem.

Nota: não é requisito obrigatório ter frequentado previamente o curso de iniciação.

Conteúdos Programáticos:

- Boas práticas e recomendações de projectos em Python
- O que é um processo de ETL?
- Data Prepping = Data Clean + Data Munging + Data Enrichment
- Exercício: Ideação, definição e setup inicial do projecto
- ETL: “E”xtracting data:
 - Ficheiros: CSV e Excel
 - Bases de dados
 - Online: API’s, Scraping, FTP
- Exercício:
 - Identificação de fontes a usar
 - Definição de estratégia de extração
 - Automatização do processo de extração
- ETL: “T”ransforming data:
 - Visualizar dados
 - Interpretar dados
 - Limpar dados
 - Dar qualidade aos dados
- Exercício: trabalhar com os dados extraídos sobre os tópicos referidos acima
- ETL: “L”oading data
 - Definição de estratégia de modelo de dados único
 - Exercício: modelação e criação de repositório único de dados