

Projeto Capacitação em Ciência de Dados para o SEN e AP Investimento apoiado pelo Plano de Recuperação e Resiliência (PRR) e pelos Fundos Europeus Next Generation EU¹

Programação Metodológica

1. ORGANIZADOR

Instituto Nacional de Estatística

2. IDENTIFICAÇÃO DO/A FORMADOR/A

Nome	N.º Emp (aplicável a trabalhadores do INE)
Bartholomeus Johannes Schoenmakers	12432

N.º CCP	Data de emissão	Data de validade (se aplicável) ¹
F730743/2022	27/07/2022	Clique ou toque para introduzir uma data.

Nome	N.º Emp (aplicável a trabalhadores do INE)
Luis Osório do Carmo Ferreira	16667

N.º CCP	Data de emissão	Data de validade (se aplicável) ¹
F730358/2022	20/07/2022	Clique ou toque para introduzir uma data.

Nome	N.º Emp (aplicável a trabalhadores do INE)
Sónia Patrícia Folgado C. Borges Quaresma Gonçalves	14508

N.º CCP	Data de emissão	Data de validade (se aplicável) ¹
EDF 460645/2007 DC	14/11/2007	Clique ou toque para introduzir uma data.

3. IDENTIFICAÇÃO DA AÇÃO

Designação
Python para estatísticas oficiais - Basic Data Science

Formato	Local
Presencial	Delegação do Porto

Data de Início	Data de Fim	Horário	Duração
08/04/2024	12/04/2024	ver observações	30 horas

¹ Conforme indicado na Portaria n.º 994/2010, de 29 de setembro, os CAPs, emitidos ao abrigo do Decreto regulamentar n.º 66/94, de 18 de novembro, com as alterações introduzidas pelo Decreto Regulamentar n.º 26/97, de 18 de junho, incluindo aqueles que tenham sido renovados nos termos do disposto na Portaria n.º 1119/97, de 5 de novembro, consideram-se emitidos sem dependência de qualquer período de validade, não carecendo de ser objeto de renovação.

Observações sobre a calendarização, horário ou outra

Módulo 1

Formador Luís Ferreira

08/04/2024 - 9:30 -12:30/14:00 – 17:00 (6 horas)

09/04/2024 - 9:30 -12:30/14:00 – 17:00 (6 horas)

Módulo 2

Formadora Sónia Quaresma

10/04/2024 - 9:30 -12:30/14:00 – 17:00 (6 horas)

11/04/2024 - 9:30 -12:30/14:00 – 17:00 (6 horas)

Módulo 3

Formador Bartholomeus Schoenmakers

12/04/2024 - 9:30 -12:30/14:00 – 17:00 (6 horas)

Objetivos

No final do curso os formandos deverão ser capazes de:

- Reconhecer código Python
- Escrever programas simples em Python, ler input do utilizador e escrever no ecrã
- Escrever código que executa operações aritméticas
- Identificar em scripts em Python alguns constructos fundamentais: variáveis, ciclos, listas e funções básicas do Python
- Distinguir entre Estatística e Inferência
- Determinar medidas de tendência central
- Determinar medidas de dispersão
- Calcular o desvio padrão e variância
- Utilizando o package NumPy, trabalhar com ARRAYS e ARRAYS 2D
- Utilizando o package NumPy, calcular estatísticas básicas
- Saber explicar o que é um modelo realcional
- Escrever código Python para efectuar uma lição a base de dados
- Executar pesquisas em bases de dados (Queries)
- Conhecer a package Pandas, nomeadamente os conceitos de Series e DataFrames
- Escrever em Markdown num Jupyter notebook
- Importar dados públicos a partir de fontes externas
- Criar gráficos básicos utilizando pyplot do matplotlib e seaborn



Programa

Configuração da Formação - O tema da formação apresenta um grau de complexidade elevado, requerendo foco e capacidade de abstração. A formação terá uma configuração presencial, de forma a potenciar o envolvimento e motivação dos formandos.

Estrutura Modular e Respectiva Carga Horária - Os conteúdos programáticos do curso de Basic Data Science serão organizados em 3 módulos, segundo um desenvolvimento sequencial.

Módulo 1 - 12 horas - Módulo de introdução a conceitos básicos da linguagem de programação Python. Tópicos abordados no módulo:

- Execução de Código (interativo / scripts)
- Funções básicas (Escrever no ecrã)
- Aritmética (Python como calculadora)
- Definir e alterar variáveis
- Tipos de dados
- Integers, Floats, Booleans, Strings
- Conversão de tipos
- Organização do Código (em Funções e Packages)
- Controlo da execução (condicional, ciclos)
- Funções mais usadas
- Introdução a Listas
- Strings e Dicionários
- Packages (Importar package e importar seletivamente)
- Zen Of Python

Módulo 2 – 12 horas - O módulo é dedicado a apresentar as estatísticas básicas em Python. Conhecer as packages NumPy e Pandas do Python é essencial para os Data Scientists. O módulo permitirá também a familiarização com os aspectos fundamentais de bases de dados relacionais e o modelo relacional para gestão das mesmas. Serão abordados elementos como as conexões às bases de dados e as interações através de consultas básicas de SQL. As bases de dados são onnipresentes e incontornáveis no processo de Data Science.

- Estatística descritiva e inferência
- Medidas de tendência central
- Medidas de dispersão; quartis, quantis, etc
- Variância e desvio padrão em Python
- NumPy
- NumPy Arrays e 2D Arrays
- Subsetting 2D NumPy Arrays e aritmética 2D
- NumPy: Estatísticas Básicas
- Introdução às bases de dados e modelos relacionais
- Conexões às bases de dados, engines and connection strings
- Queries SQL e pesquisas em base de dados
- Exploração do Pandas: Series and DataFrames

Módulo 3 – 6 horas - Este módulo tem como objetivo apresentar a utilização de jupyter notebooks para documentação de scripts e visualização dos seus resultados. Mostra-se como dados de pandas dataframes podem ser visualizados em diferentes tipo de gráficos, utilizando o Pyplot do Matplotlib.

- Introdução a trabalhar com Jupyter Notebooks e Markdown
- Utilização de dados de Pandas para visualização
- Importar dados públicos a partir de fontes externas
- Diferentes tipos de gráficos e Técnicas de visualização
- Criar gráficos básicos utilizando pyplot do matplotlib e seaborn



Metodologia

Será usada uma mistura de métodos, recorrendo maioritariamente ao método demonstrativo como fio condutor da narrativa possibilitando o desenvolvimento sequencial da apresentação dos conteúdos.

Numa primeira fase, sincrética, o formador proporcionará uma visão global da operação que o formando deverá aprender, explicando globalmente o conteúdo, o processo e o tempo necessário para o realizar. Esta demonstração / explicação é feita

num ritmo normal de execução da tarefa em aprendizagem.

Este método será intercalado com o método activo, para levar os formandos a experimentar e manipular os materiais e permitir aferir a aquisição de alguns conceitos no decurso da formação.

Serão utilizadas ainda técnicas pedagógicas de tipo expositivo (por exemplo, em momentos de introdução de temas, sínteses parciais e finais).

Recursos Pedagógicos

A acção de formação terá em sala um computador portátil para o formador e um para cada formando, um videoprojector e um quadro branco com as respectivas canetas.

Recursos Materiais e Logísticos

Critérios de seleção dos espaços de formação:

- Computadores funcionais equipado com placa de som, microfone, colunas de som e com ligação à Internet e à intranet e pré-instalação de Oracle, Ambiente Anaconda com os pacotes de Python instalados NumPy, Pandas, Scikit-learn, Seaborn, Matplotlib, SciPy, PyPlot, Getpass, cx_Oracle, instalação de Visual Studio Code com Plugin de Python da Microsoft
- Sala com boa luminosidade, ventilação, temperatura e isolada de ruídos perturbadores ao bom funcionamento;
- Espaço equipado com todos os recursos didáticos necessários;
- Mobiliário que respeite as regras de ergonomia dos formadores e dos formandos;
- Espaço amplo o suficiente para permitir a concretização de dinâmicas de grupo e da disposição das mesas em "U", no sentido da facilitação da comunicação;
- Local de fácil acesso;
- Espaço cuja limpeza é assegurada diariamente.

Avaliação

Não haverá avaliação. Será emitido um certificado de participação no curso se o formando tiver registado uma assiduidade mínima de 95% sobre a duração global do curso.



4. DESTINATÁRIOS

Destinatários	N.º Participantes
Técnicos do INE e das EDCs	12 (n.º máximo)

Requisitos Prévios

¹ www.recuperar Portugal.gov.pt