

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 EU<sup>1</sup>

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) <sup>1</sup>
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) <sup>1</sup>
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) <sup>1</sup>
EDF 460645/2007 DC	14/11/2007	Clique ou toque para introduzir uma data.

## 3. IDENTIFICAÇÃO DA AÇÃO

Designação
Python para estatísticas oficiais - Intermediate Data Science

Formato	Local
Presencial	Delegação do Porto do INE

Data de Início	Data de Fim	Horário	Duração
15/04/2024	22/04/2024	ver observações	30 horas

<sup>1</sup> 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

15/04/2024 - 9:30 -12:30/14:00 – 17:00 (6 horas)

16/04/2024 - 9:30 -12:30/14:00 – 17:00 (6 horas)

### Módulo 2

Formadora Sónia Quaresma

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

18/04/2024 - 9:30 -12:30/14:00 – 17:00 (6 horas)

### Módulo 3

Formador Bartholomeus Schoenmakers

22/04/2024 - 9:30 -12:30/14:00 – 17:00 (6 horas)

## Objetivos

No final do curso os formandos deverão ser capazes de:

- Fazer a instalação do Anaconda e implementar environments específicos
- Instalar Packages Python
- Utilizar Python em Ambiente Integrado do Desenvolvimento (IDE) ou em Jupyter
- Utilizar classes e módulos
- Criar funções em Python sobre listas e dicionários
- Conhecer os princípios base de Python Comprehensions
- Conhecer os princípios base para escrever programas em Python
- Saber manipular DataFrames do Python
- Obter estatísticas a partir de um DataFrame
- Explorar as funcionalidades do Scikit e fazer pré-processamento de dados em Scikit
- Identificar os princípios base de Machine Learning
- Criar modelos básicos de classificação e regressão e fazer a sua avaliação
- Descrever os princípios base de dados geográficos e com estes podem ser utilizados em Python
- Fazer manipulações básicas de dados geográficos
- Fazer visualizações de dados geográficos utilizando os packages matplotlib, geopandas e folium
- Compreender como dados de OpenStreetmap e Google Maps podem ser utilizados num ambiente Python



## Programa

### CONFIGURAÇÃO DA FORMAÇÃO

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 Intermediate Data Science serão organizados em 3 módulos, segundo um desenvolvimento sequencial.

#### Módulo 1 – 12 horas

Módulo dedicado à configuração do ambiente de desenvolvimento e continuação da introdução de conceitos de programação em Python. Tópicos abordados no módulo:

- Instalar Anaconda
- Environments
- Instalar Packages (usando o pip)
- Jupyter Notebook
- IDE (VSCode)
- Usar Classes e Módulos
- Definir classes e Métodos
- Funções sobre Listas e Listas bidimensionais
- Comprehensions (List e Dictionary)
- Zen Of Python

#### Módulo 2 – 12 horas

O módulo é dedicado à familiarização com os conceitos de Data Science usando Pandas e Scikit. Pandas é uma das bibliotecas Python mais populares, sendo usada para tudo, desde manipulação de dados até à sua análise. Este módulo foca-se na manipulação de DataFrames para permitir extrair, filtrar e transformar conjuntos de dados para análise. O Scikit é a biblioteca Python primordial para a Data Science e permitirá para explorar todos os seus principais conceitos, introduzindo a aprendizagem supervisionada e os seus diferentes métodos de Regressão e Classificação.

- Exploração da package Pandas
- DataFrames
- Reorganizar a DataFrame; linhas e colunas (adição e cálculo)
- summary statistics na DataFrame; medidas de tendência central
- Exploração do scikit package
- Pré-processamento usando Scikit
- introduction ao Machine Learning: Modelos de Regressão e Classificação
- Modelos de Regressão usando Statsmodels package
- Modelos básicos de classificação usando árvores no scikit
- Avaliação de Modelos de Regressão e Classificação

#### Módulo 3 – 6 horas

Este módulo tem como objetivo fornecer uma introdução prática à visualização de resultados obtidos de análise de dados no anterior módulo e uma introdução a visualização de dados geográficos.

- Introdução a trabalhar com dados geográficos
- Formatos de dados geográficos
- Projeções de dados geográficos
- GeoPandas
- Visualização de dados geográficos utilizando as bibliotecas geopandas, matplotlib e folium
- Introdução a utilização dos dados de OpenStreetmap e Google Maps num ambiente Python



## 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, Leafmap, Folium, Matplotlib, SciPy, PyPlot, Geopandas, 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

Técnicos do INE e das EDCs

### N.º Participantes

12 (n.º máximo)

### Requisitos Prévios

Esta formação é continuação da formação Basic Data Science. Espera-se dos participantes de conhecer o conteúdo desta formação.

<sup>1</sup> [www.recuperar.pt](http://www.recuperar.pt)