

Mestrado em Informática

Pós-Graduação em Data Science and Digital Transformation

(Ambientes de) Programação para Ciência de Dados

Projeto 1

OBJETIVOS

Pretende-se com este projeto utilizar o Python e as suas bibliotecas para explorar e interpretar os dados que constam nos ficheiros **incendios_sc.csv** e **estados.csv**

O ficheiros de dados foram obtido a partir de dados públicos disponíveis em <https://queimadas.dgi.inpe.br/>. O ficheiro **incendios_sc.csv** apresenta informação relativa a eventos de incêndios (queimadas) detetados por satélite em municípios brasileiros pertencentes ao bioma Amazónia. O ficheiro **estados.csv** relaciona os municípios com os respetivos estados federais.

Devem ser dadas às resposta às questões do enunciado com base na análise efetuada com Python e nas bibliotecas adequadas. O trabalho é entregue através do PAE em formato *jupyter notebook*. Se precisar de entregar mais do que um ficheiro, deve submeter no PAE um único arquivo comprimido. O trabalho é sujeito a discussão em data a combinar,

PRIMEIRA PARTE

A primeira parte do trabalho consiste na utilização de Python e suas bibliotecas para explorar e interpretar a informação que consta nos dois ficheiros.

Realize uma análise inicial aos dados constantes em ambos os ficheiros e interprete o seu conteúdo, respondendo a questões como:

1. **incendios_sc.csv**
 1. Qual é a dimensão do conjunto de dados?
 2. Qual o nome das colunas e que tipo de dados que lhe estão associados?
 3. Existem valores NaN? Se sim, em quantidade significativa? Em que colunas?
 4. Qual é o intervalo temporal a que o conjunto de dados se reporta?
 5. Analise os dados da coluna "Satelite" para dar resposta às questões:
 - a. Quais são os satélites dominantes no dataset?
 - b. Os satélites disponíveis foram sempre os mesmos ao longo dos anos, ou houve alterações? Se sim, quais
 - c. Qual é o satélite mais antigo registado no dataset? Encontra-se ainda em operação?
 - d. Existem satélites que deixaram de operar? Quais
2. **estados.csv**
 1. Qual é o tamanho total do dataset (número de linhas e colunas)?
 2. Quantos municípios e quantos estados estão representados no dataset?
 3. Existem nomes de municípios ou estados repetidos? Identifique-os.
 4. O dataset contém valores NaN?

SEGUNDA PARTE

A segunda parte do trabalho consiste em analisar os focos de incêndio temporal e espacialmente.

1. Comece por criar um conjunto de dados (a gravação para ficheiro .csv é opcional) com as seguintes colunas: Ano, Mês, Estado, Nº de ocorrências.

A partir deste novo conjunto de dados, analise os seguintes aspetos, apresentando evidências que sustentem as conclusões:

2. Evolução anual dos incêndios

- a. Como evoluiu o número total de incêndios ao longo dos anos?
 - b. Observa-se alguma tendência significativa?
 - c. Em que anos se registaram o menor e o maior número de ocorrências?
3. Distribuição mensal e análise por estado
 - a. Como se distribui o número de incêndios ao longo dos meses do ano (usar os valores médios no período representado)?
 - b. Para cada estado, determine: o número total de incêndios, o valor médio e o desvio-padrão das ocorrências.
 - c. Que estados apresentam o maior e o menor número de incêndios?
4. Análise temporal dos estados extremos: considerando os estados identificados na questão anterior, analise a evolução anual do número de incêndios em cada um deles.