

Análise e organização dos dados

- Analisar os dados com o profiling:
 - `import pandas_profiling`
 - `pandas_profiling.ProfileReport(df2017.drop(columns='coords'))` – não dar ao profiling a coluna das coordenadas para não o entupir...
- Analisar fora do Python
 - Exportar os dados para ficheiros Excel
 - Abrir no Excel
 - Transformar os dados numa tabela (seleccionar todos os dados e fazer “Insert Table”)
 - Observar cada uma das colunas do Excel e tirar conclusões
 - Voltar ao Python e melhorar os dados
 - Repetir o processo até ter os dados devidamente organizados

Algumas operações de limpeza dos dados

- Os fogos de 2010 e 2011 não têm data – talvez seja melhor descartar
- 2012 e 2013 têm muita informação em falta – talvez seja melhor descartar para já
- 2014 e 2015 têm uma coluna “ANO” que poderá ser apagada (desde que se crie uma coluna “Ano” com o ano de cada exemplo quando se lêem os dados)
- 2014 tem fogos com GHInicio vazio e AREA_HA=0 que podem ser apagados
- 2015 tem fogos em que o “Dia” é zero que podem ser apagados
- 2016 tem fogos em que o “Dia” é zero ou vazio e outros em que DHInicio está vazio e que podem ser apagados
- 2017 tem fogos em que o “Dia” é zero ou vazio que podem ser apagados
- 2018 tem fogos em que “DHInicio” é zero ou vazio que podem ser apagados
- 2015 tem 'DHfim', 'Dhinicio' e 'AreaTotal' que correspondem a 'DHFim', 'DHInicio', 'AREA_HA' respetivamente nos outros anos
- 2016 tem 'Dhinicio' e não 'DHInicio' (renomear)
- 2018 tem 'AreaTotal' e não 'AREA_HA' (renomear)
- As colunas 'Tipo' e 'Concelho' têm strings com acentos que aparecem “desformatados” – traduzir utilizando dicionários

Objetivo imediato desta reorganização dos dados

- Todas as linhas devem ter area, data, tipo e local (será necessário verificar este facto)
- As colunas Tipo e Concelho devem ter valores únicos para a mesma coisa
- ...