DATATTACK

2.0

CHALLENGE INFORMATION

NEB-FEUP/ICBAS
núcleo de estudantes de bioengenharia

IEEE
University of Porto - Student Branch

ni
aefeup

PROBLEM

Can you use data to optimize the performance of the National Emergency and Civil Protection Authority (Proteção Civil)?

GOAL

The goal of this challenge is to develop a data-driven framework to optimize the work of the National Emergency and Civil Protection Authority.

DATASET

The National Emergency and Civil Protection Authority has made the history of events (occurrences) publicly available - dates, locations and number of involved professionals (terrestrial and aerial). Occurrences are updated live on the Authority's website. An example of this data can be consulted here.

DATA DICTIONARY

Occurrences

- Numero: Occurrence Identifier
- DataOcorrencia: Occurrence Date
- DataFechoOperacional: Occurrence Closing Date
- Natureza: Objective of the occurrence (e.g. Protecção e Assistência a Pessoas e Bens / Assistência em Saúde / Trauma)
- EstadoOcorrencia: Occurrence State (Closed, False Alarm, ...)
- Distrito: The District (i.e., "Distrito") of the place of the occurrence
- Concelho: The Municipality (i.e., "Concelho") of the place of the occurrence
- Freguesia: The Parish (i.e., "Freguesia") of the place of the occurrence
- Localidade: The local of the place of the occurrence
- Latitude: The latitude of the place of the occurrence

- Longitude: The longitude of the place of the occurrence
- NumeroMeiosTerrestresEnvolvidos: Number of involved vehicles (in land)
- NumeroOperacionaisTerrestresEnvolvidos: Number of people involved in land operations
- NumeroMeiosAereosEnvolvidos: Number of involved helicopters
- NumeroOperacionaisAereosEnvolvidos: Number of people involved in aerial operations

For simplification purposes, we'll use the data from 2016, containing 120 000 occurrences. If you have time, you can include data from other years.

You can use extra data sources, such as:

- Geographical Location of the Parishes (i.e., "Freguesias"), available at https://dados.gov.pt/pt/datasets/freguesias-de-portugal/ (Note: requires GeoPandas)
- Geographical Location of the Municipalities (i.e., "Concelhos"), available at https://dados.gov.pt/pt/datasets/concelhos-de-portugal/ (Note: requires GeoPandas)
- Populational Density (https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&contecto=pi&indOcorrCod=0008337&selTab=tab0&xlang=pt)

SUGGESTIONS

Some examples of things you can do are:

- Predict the number of future occurrences and the number of resources that you will need to allocate, per location
- Visualize the occurrences over time and try to find relevant patterns (e.g. correlate with population density, occurrence nature with development level and topology of the area ...)

- Understand which type of occurrence is most common per location, and how that creates a bottleneck in Authority's services
- Measure data quality and give recommendations for creating better datasets

DELIVERABLES

Participants are expected to share a GoogleDrive link to a folder containing the following documents:

- Code used to develop their solution(s) (e.g., Python scripts, Jupyter Notebooks, Google Collaboratory Notebooks)
- The presentation that will be used for the final pitch session (e.g., PDF, PPT)

EVALUATION

At the end, you will create a presentation of your work. Your solution will be evaluated according to:

Solution

- Exploratory Data Analysis: how well you explored and summarized the data
- Technical Rigor: how rigorous were you in modeling the problem
- Impact: how impactful is your solution for being used by Proteção Civil. Think about making your solution actionable!
  - How many resources are you saving?
  - What bottlenecks have you identified and how much do they cost?
  - How much can they be improved?

Presentation

- Clarity
- Technical and Non-technical communication skills

USEFUL LINKS

- **Plotting**
  - Folium
  - Matplotlib
  - Plotly
- **Modeling**
  - scikit-learn
- **Data Processing**
  - GeoPandas
  - pandas