DATA SCIENCE FOR SOCIAL GOOD
PORTUGAL

# Data Science for Good

The author was not able to come up with a decent sub-title for this talk. He hopes to make it up to you with nice slides and exhaustive use of emoji.

DS@Work | DSPT Day 2019

Porto, 25th October 2019

# 🔴 Expectation management

- DSSG PT does **not conduct groundbreaking research**.
- DSSG PT is **not** a **company**.
- DSSG PT is **often recruiting**, but for **unpaid positions**.
- DSSG PT, from a purely **technical standpoint**, is a rather **boring endeavour**.

# 🟢 However...

- DSSG PT always works on **real-life problems**.
- DSSG PT delivers not papers, not prototypes, but usable **Data Science products**.
- DSSG PT only works on one type of problems: **those that matter**.

# Actually, think of us as a big pharma

An open **community** of **data scientists**, **data lovers** and **data enthusiasts** that wants to **tackle problems that really matter**. We believe in the **power of data to transform our society** for the better, for everyone.

Beneficiaries

Volunteers

Projects

# The formula

**Beneficiaries**

Charities
NGOs
Public Administration

+

**Volunteers**

Data Scientists
Data Enthusiasts
Developers
Designers
Domain experts

=

**Projects**

1-3 months; **Deliverable** (small, simple yet able to bring value - report, interactive visualization, predictive model, etc)

# Rotaract

A global organization

that empowers **students** and

**young professionals** to

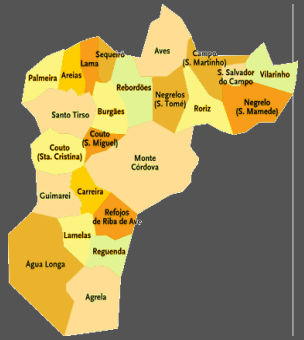
create **positive change**

around the world.

# Rotaract

## Santo Tirso

# Fundraising *Liga Portuguesa Contra o Cancro*



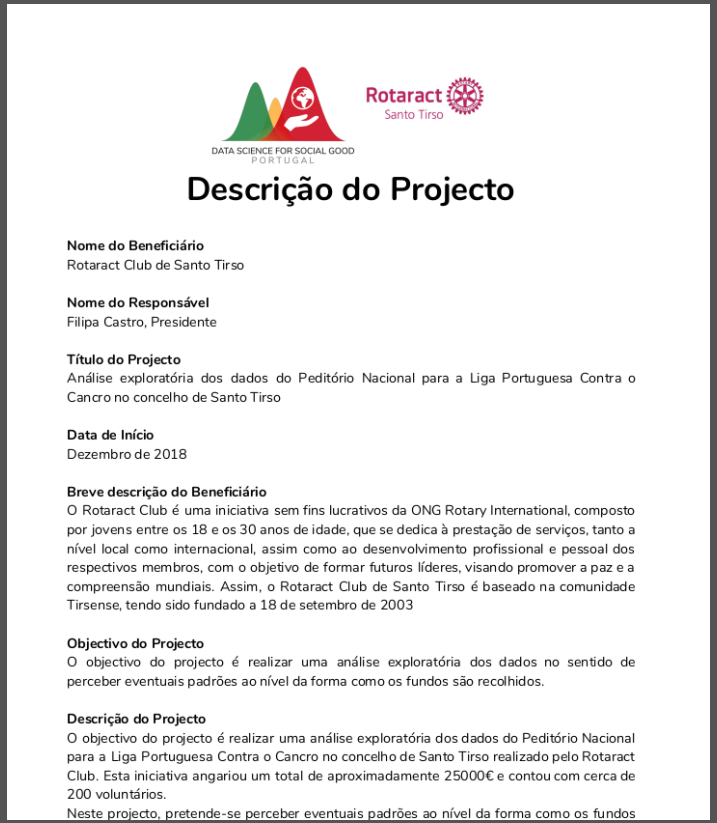

Meet with each council and authorities to raise awareness and collaborations

Manage volunteers (~100 across 3 days)

Collect and analyze Contributions (20-25k)

# Project's goals

- **Main goal:** Exploratory analysis of the fundraising data to find eventual patterns or anomalies.
  - Which parishes donate more? Can we correlate it with anything? Is it stable over the years?
  - What is the expected performance of a volunteer? Does it vary by year?
  - What makes one volunteer perform better than another?
  - Where is the fundraising more effective?
- **Deliverable:** Report that can be easily navigated and shown to relevant stakeholders



DATA SCIENCE FOR SOCIAL GOOD PORTUGAL

Rotaract Santo Tirso

## Descrição do Projecto

**Nome do Beneficiário**
Rotaract Club de Santo Tirso

**Nome do Responsável**
Filipa Castro, Presidente

**Título do Projecto**
Análise exploratória dos dados do Peditório Nacional para a Liga Portuguesa Contra o Cancro no concelho de Santo Tirso

**Data de Início**
Dezembro de 2018

**Breve descrição do Beneficiário**
O Rotaract Club é uma iniciativa sem fins lucrativos da ONG Rotary International, composto por jovens entre os 18 e os 30 anos de idade, que se dedica à prestação de serviços, tanto a nível local como internacional, assim como ao desenvolvimento profissional e pessoal dos respectivos membros, com o objetivo de formar futuros líderes, visando promover a paz e a compreensão mundiais. Assim, o Rotaract Club de Santo Tirso é baseado na comunidade Tirsense, tendo sido fundado a 18 de setembro de 2003

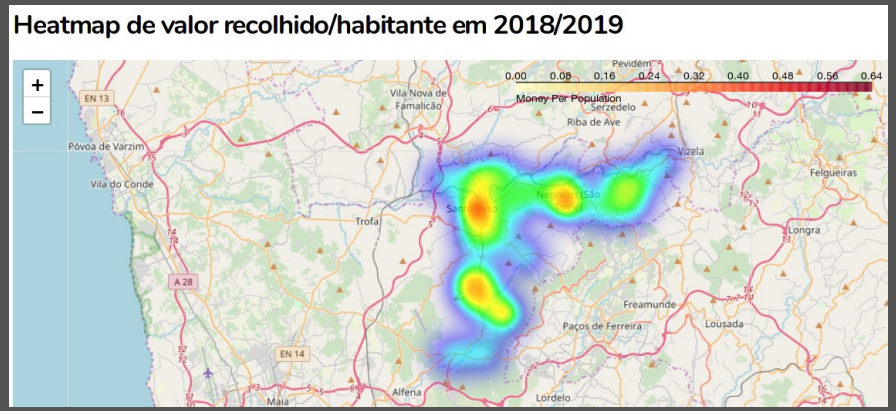
**Objectivo do Projecto**
O objectivo do projecto é realizar uma análise exploratória dos dados no sentido de perceber eventuais padrões ao nível da forma como os fundos são recolhidos.

**Descrição do Projecto**
O objectivo do projecto é realizar uma análise exploratória dos dados do Peditório Nacional para a Liga Portuguesa Contra o Cancro no concelho de Santo Tirso realizado pelo Rotaract Club. Esta iniciativa angariou um total de aproximadamente 25000€ e contou com cerca de 200 voluntários.
Neste projecto, pretende-se perceber eventuais padrões ao nível da forma como os fundos

# 🎯 Deliverables shape Projects

- Started with the traditional `Jupyter Notebook` + `pandas` + `Matplotlib` analyses (several plot types).

- Iterating with Beneficiary led to interactive heatmaps, implemented with the `folium` package.



EXPLORATION

PRODUCTION

WASTED TIME

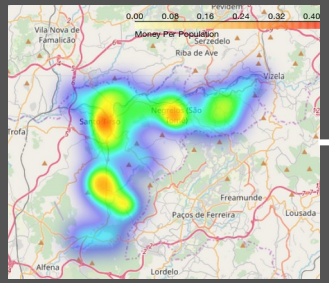
# The ecosystem comes to the rescue

plotly

Powerful Javascript visualization library with a Python API. Syntax is very **verbose** and **un-Pythonic**.

```python
import plotly.express as px
iris = px.data.iris()
fig = px.scatter(iris, x="sepal_width", y="sepal_length")
fig.show()
```

Recently launched high level Python API with a very `seaborn`-**like**, `pandas`-**friendly** interface



Dash

Build interactive webapps from Python/R (**no JS** and **minimal HTML/CSS**) with `plotly.express` **plots as first-class citizens**. Also, easy HTML embeddings!

# 🚀 Quick launch to production

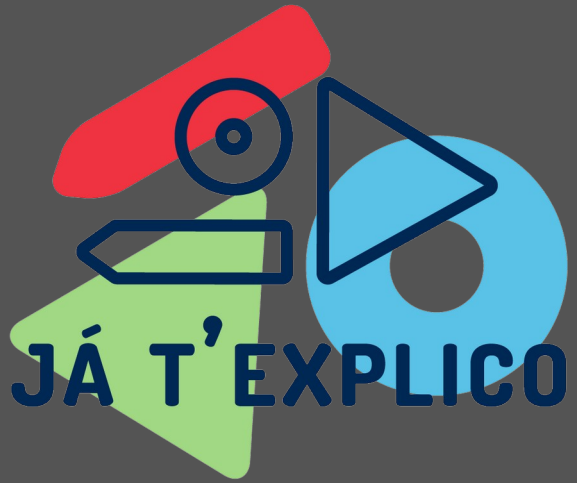
From a standard Python project tree and with heroku configured:

```
# Create a Procfile, a 1-liner specifying
entrypoint. Then:

$ heroku create a-dash-app
$ git push heroku master
$ heroku ps:scale web=1

# App available at a-dash-app.herokuapp.com
```

# Take-home messages

- Fast growing ecosystem of seamlessly integrated tools (`voilà`, `panel`, `steamlit`) that accelerate **exploration -> production**.
- **Production is where the fun's at**. DS has **the most value** not in GitHub repos, but in **people's hands**.
- Lots of organizations doing good things would **benefit from even the simplest DS tools and algorithms**. Enter **DSSG PT**.

# ❓ Current status




Ongoing

Soon!

# 🕐 One more thing...

**Partners & Supporters:** organizations who want to **contribute** directly/indirectly to our cause and projects and are **aligned with our mission and values**.

Yours? 👀

# Tag us on...

dssg.pt

dssg_pt

@dssgpt

/DSSGPortugal

...or grab your sticker @

**Community Corner**

DATA SCIENCE FOR SOCIAL GOOD

PORTUGAL