

# HACKATHON

## Online Wine Price Harvesting Challenge



**Let us start by introducing ourselves:**



Sogrape is a family-owned wine company with a strong international presence and unique diversity.

Born from a visionary spark of innovation, Sogrape embraces an elegant journey of expertise, crafting iconic solutions that redefine the very essence of innovation.

From Portugal to the world.

Today, it produces wine in the country's main wine-growing regions and beyond its borders in Spain, Argentina, Chile and New Zealand.





# The Creator

## Fernando Van Zeller Guedes

Founder

Author of Portugal's only global wine brand

The founder of the company, back in 1942, demonstrated a gift for anticipating the future. His dream was to prove the quality of Portuguese wines to the world, and in a tremendous act of **creativity** he brought to life a light and refreshing rose wine – **Mateus!**

## MATEUS ROSÉ

Initially inspired by the canteens used by soldiers in World War I, the **canteen-shaped bottle** also had a strategic purpose – unusual, iconic and disruptive, it always ensured a **tactical position on store shelves**, being placed in front of taller wine bottles and with the label forefront.

This iconic bottle contained something completely different:

**A new style of wine, pink in color, refreshingly light and slightly sweet**, easy to drink and suited to the palates of the younger generations.



## DREAM

To be admired as the  
most successful  
family-owned wine  
company in the world

## SPIRIT

Sograpiness — Friendship  
& Happiness

Our purpose is to bring  
Friendship and  
Happiness to everyone  
we touch through our  
wonderful wines.

This is how it all begins.

## SOGRAPE'S INNOVATION GOAL

To leverage an **Ecosystem** that attracts and orchestrates **talent, knowledge and technology** to **build the next legacy of Sogrape.**

## 23 October

- 14:00 Check-in teams & networking
- 14:30 Welcome Note & Challenge Details
- 15:00 Teams Start Working
- 16:30 #1 Mentors Checkpoint

## 24 October

- 12:00 – 14:00 #2 Mentors Checkpoint

## 25 October

- 09:30 – 10:30 #3 Mentors Checkpoint
- 14:00 – 15:00 #4 Mentors Checkpoint
- 16:00 – Teams Stop Working
- 16:00 – 18:30 Pitches presentation - Jury pre-selects the finalist teams

## 26 October

- 09:00 – 10:00 Final Pitch
- 10:00 – 11:00 Coffee-Break
- 11:00 – 12:00 Winners announcement & closing remarks

# Mentoring and Evaluation Process





## RELY ON MENTORS

Please bear in mind that:

- The mentors are **NOT PART OF THE TEAM**, but rather ensure that the goal is always present, that all possibilities are explored and that the team is not trapped in any specificity that may compromise the completion of the challenge
- The Team's Mentors will be **AVAILABLE DURING CHECKPOINTS**
- During the pre-selection for the final presentation, **MENTORS DO NOT EVALUATE THE TEAMS**

## MEET THE MENTORS



António Pereira  
Custom Development  
Manager @ Sogrape



Vítor Branco  
IT Director @ 42  
Lisboa & 42 Porto



Inês Carola  
Innovation Specialist @  
Sogrape



Matthew Hogg  
Innovation Specialist @  
Sogrape



Maria Azevedo  
Office Manager @  
42 Porto



Mariana Martins  
Program Operations Manager  
@ 42 Porto



Vasco Gonçalves  
International Area  
Manager @ Sogrape



Rita Dias Costa  
Digital Communication  
Specialist @ Sogrape



Joana Martins  
Developer @ Sogrape

25th October - 16:00

Submit your work to the Github repository that will be made available to you.

Submission details:

- Technical part directory: `webscraping/`
- Dashboard part directory: `hack_dashboard/`

**Deliverables:**

- Executive Summary
- PPT with presentation
- Web Scraping demo
- Dashboard demo

25th October - 16:00

**Each Team presents their work** to the Intermediate Jury

- 1 Live pitch (5 minutes)

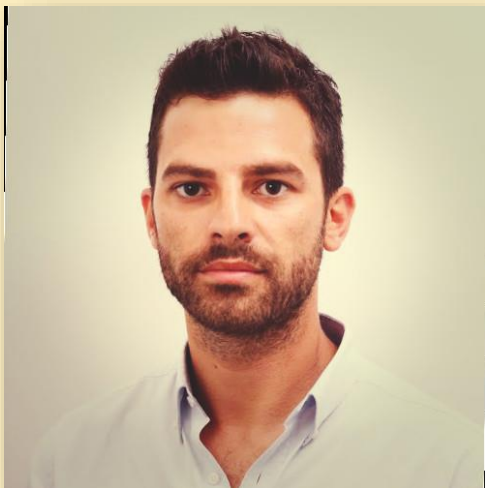
26th October - 09:00

**Finalists' announcement**

26th October - 11:00

**Winners' announcement**

## INTERMEDIATE EVALUATION – FINALISTS SELECTION



**Jorge Machado**  
Digital Transformation  
Manager @ Sogrape



**Tiago Costa**  
Data and Technology  
Specialist @ Sogrape



**Artur Salazar**  
Digital & Webmaster  
Specialist @ Sogrape

## MEET THE JURY



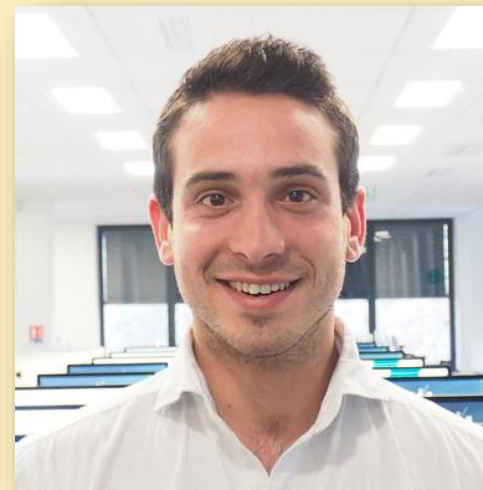
**Carlos Alves**  
Head of IT @ Sogrape



**Sara Mendes**  
Innovation Manager @  
Sogrape



**António Pereira**  
Custom Development  
Manager @ Sogrape



**Vítor Branco**  
IT Director @ 42 Lisboa &  
42 Porto

## 1ST PRIZE

**700€ (for the team) &  
follow-up implementation**

## 2ND PRIZE

**Lunch at Quinta do Seixo in  
Douro Valley (for the team)**

## 3RD PRIZE

**Visit to the Sandeman Wine  
Cellars in V.N. Gaia (for the team)**



# THE CHALLENGE

## Online Wine Price Harvesting Challenge



This challenge offers you the opportunity to immerse yourself in a realistic situation faced by industry professionals.

By designing a comprehensive solution that addresses the two dimensions - information collection and the creation of a user-friendly dashboard - you will be developing valuable skills in web scraping, data management, and interface design. The solution you create could have a significant impact on the operations of Sogrape and showcase your ability to tackle real-world programming challenges.



## 1. Technical Dimension

### Daily Information Collection - Web Scraping:

- Web Scrapping and Database Design
- Automated Data Retrieval and Structuring
- Web Content Aggregation and Database Architecture

You will be tasked with designing and implementing a web scraping system capable of collecting wine prices from national online stores on a daily basis. This involves creating automated scripts that navigate store pages, extract information about selected wine prices, and update the data regularly. You will need to deal with different website structures and potential security measures against scraping.

You will need to collect and store information about prices, dates, and the source of online stores, etc.

The model's design should not be customized for only one of the online stores, in order to make it scalable and adaptable to any other store.

## 2. Visualization and Interaction

### Display & user interaction with the data:

The second part of the challenge is to develop an interactive and user-friendly interface to present the collected data.

You will be responsible for creating a dashboard that allows users to filter data by countries, dates, products, etc.

The goal is to provide clear visualizations, such as graphs and tables, to help the team make informed strategic decisions.

## Challenge Details:

Websites to Scrape data from:

- Continente Online (<https://www.continente.pt/>)
- Garrafeira Soares (<https://www.garrafeirasoares.pt/>)
- Supermercado El Corte Ingles (<https://www.elcorteingles.pt/supermercado/>)

**EXTRA** – If you have the chance to scrape data from more websites as well as the ones listed in this challenge, extra points may be attributed. (For reference, other national websites that are available: [auchan.pt](http://auchan.pt); [onwine.pt](http://onwine.pt); [granvine.com](http://granvine.com); [vinha.pt](http://vinha.pt); [adegga.com](http://adegga.com); [portugalvineyards.com](http://portugalvineyards.com))

## Products to consider:

- Mateus Rosé Original (No harvest Year)
- Mateus Sparkling Rosé (No harvest year)
- Herdade do Peso Trinca Bolotas Tinto (May refer to harvest year)
- Casa Ferreirinha Papa Figos Branco (May refer to harvest year)

**EXTRA** – If you have the chance to scrape data for more products, in order to compare price fluctuations / discounts and such with competitors, extra points may be attributed. (For reference: Examples of comparative products respective to the products listed above are: Lancers Vinho Rosé; Casal Garcia Sparkling Espumante Rosé; Defesa do Esporão Tinto; Diálogo DOC Douro Vinho Branco).

# The Challenge



**Data Points to gather through web scraping:** The challenge involves scraping specific data points related to the selected wines with this information:

1. **Store Name**
2. **Wine Name**
3. **Harvest year (when possible)**
4. **Capacity**
5. **Price**
6. **Discounts** (Signal if product has a discounted price)
7. **Currency**
8. **Date of scraping**
9. **Location**

**EXTRA** – If you have the chance to scrape more data for the products such as: stock status (in stock/out of stock/low stock), product image, product description; pairing suggestions, extra points may be attributed. If you decide to do this, only consider the additional data for the 4 products listed in point 2.

## Data Visualization and Interaction Dimension

### Dashboard Requirements:

The dashboard that you, as a programming student, will develop should include the following features:

- **Filters:** Users can filter data by wine, market, and date.
- **Price Trends:** Visualizations depicting price trends over time for different wines in different markets.

**EXTRA** – If you have scraped prices for other products as mentioned in point 2, it should be possible to compare the prices / price trends in the dashboards.

## Other useful information:

Keep in mind that the name can differ from store to store! Similarly, prices may differ according to store locations (zip code).

### Product Names in Online Stores – Examples:

#### Continente Online:

- “Papa Figos DOC Douro Vinho Branco”
- “Mateus Sparkling Espumante Rosé Bruto”
- “Mateus Vinho Rosé”
- “Trinca Bolotas Regional Alentejano Vinho Tinto”

Note: Keep in mind that the product name can differ from store to store! There also might be differentiated prices for online stores on a store-by-store basis, based on the zip code entered during login.



25th October - 16:00



Submit your work to the Github repository that will be made available to you.

Submission details:

- Technical part directory: `webscraping/`
- Dashboard part directory: `hack_dashboard/`

25th October - 16:00



**Each Team presents their work** to the Intermediate Jury

- 1 Live pitch (5 minutes)

26th October - 09:00



**Finalists' announcement**

26th October - 11:00



**Winners' announcement**

## Deliverables

1. Executive Summary (suggestion: Approach to the Challenge; Main Recommendations and Main Conclusions).
2. PPT with presentation (for the pitch)
3. Web Scraping demo (video or interactive)
4. Dashboard demo (video or interactive)

## YOUR TEAM'S WORK WILL BE EVALUATED ON FOUR MAIN AREAS

### **Eliminatory requirements**

- Proposes a solution for the challenge presented

### **Impact, applicability and creativity**

- Solution clearly tackles the challenge
- Solution is general and applicable

### **Quality of the deliverable**

- Solution is clearly and concisely documented and presented

### **Extra points**

- The solution considers extra websites beyond the 4 listed in the standard challenge
- The solution considers extra products (competitors) beyond the 4 listed in the standard challenge
- The solution scrapes additional information for the products (stock status, product image, product description, pairing suggestions)

### **Technical Merit**

- Solution performs exploratory Data Analysis (EDA) & cleans the data
- Solution explores model interpretability and fairness
- Solution tests different modelling approaches
- Solution is scalable
- Solution explores configuration methods
- Solution performs results visualization in an Interactive Dashboard

# Evaluation Criteria

## 20% Impact & applicability

10%	Solution clearly tackles the challenge
10%	Solution is general and applicable

0 Strongly disagree	1 Somewhat disagree	2 Somewhat agree	3 Strongly agree
The solution does not tackle the challenge	The solution is remotely connected with the challenge	The solution partially addresses the challenge	The solution is a perfect fit with the challenge
The solution cannot be implemented	The implementation of the solution is questionable	The solution is general and its implementation is possible, but challenging	The solution is general and easily implementable in practice

## 60% Technical merit

10%	Performs Exploratory Data Analysis (EDA) & cleans the data
5%	Explores model interpretability and fairness
5%	Tests different modelling approaches
15%	Solution is scalable
10%	Explore solution configuration methods
15%	Performs results visualization in an Interactive Dashboard

Performs no EDA nor data cleaning	Performs superficial EDA or excludes messy data	Performs detailed EDA, identifying potential issues with data	Performs EDA to generate relevant insights applied in modelling, implementing strategies for cleaning and data imputation
Does not address model interpretability or fairness	Explores model interpretability or fairness	Explores model interpretability and fairness, without proposing a mitigation strategy	Explores interpretability and fairness in detail, proposing a mitigation strategy (or not needing it)
Tests a single model	Tests a single model, optimizing its formulation and parameters	Tests multiple models	Tests multiple models, optimizing its formulations and parameters, objectively selecting the best option
The solution cannot be scalable	The scalability of the solution is questionable	The solution scalability is possible, but challenging	The solution is scalable
Static with no possibility for execute changes and configurations	Configurations are possible, but require technical expertise	Built a simple interface where some configurations are allowed	Build UI interface for user to manage the solution with no technical
Performs no results visualization	Presents the results tabularly	Creates visually appealing graphics to present the results	Performs detailed multi-dimensional graphical analysis

## 20% Quality of the deliverable

20%	The solution is clearly and concisely documented and presented
-----	--

The solution cannot be easily perceived	The documentation and presentation is clear but not concise	The documentation and presentation is clear and concise but not appealing	The documentation and presentation is clear, well thought, concise and appealing
---	---	---	--

## 10% Extra Points

2,5%	The solution considers extra websites beyond the 4 listed in the standard challenge
5%	The solution considers extra products (competitors) beyond the 4 listed in the standard challenge
2,5%	The solution scrapes additional information for the products (stock status, product image, product description, pairing suggestions)

The solution does not consider additional websites	The solution considers 1 extra website	The solution considers 2 extra websites	The solution considers 3 or more extra websites
The solution does not consider additional products	The solution considers 1 extra product	The solution considers 2 extra products	The solution considers 3 or more extra products
The solution does not consider additional information	The solution considers 1 of the extra listed data points	The solution considers 2 of the extra listed data points	The solution considers 3 or more of the extra listed data points

# HACKATHON

## THE TEAMS



**Team 1**

gperfeit  
amenses-  
agpereir  
hicunha-  
luguimar  
passunca

**Team 2**

dcarrilh  
macastan  
plereno-  
llopes-f  
fduque-a  
dmeirele

**Team 3**

arocho-b  
tibarbos  
maricard  
ecarvalh  
gumoreir  
rpaiva-l  
dflores-

**Team 4**

tcosta-d  
fjourdan  
bsilva-c  
cacarval  
mborges-  
Joseferr

**Team 5**

jhogonca  
heda-sil  
psousa  
vbritto-  
tornelas  
Jburlama

**Team 6**

asousa-n  
malves-b  
dcaetano  
bjorge-m  
Damachad  
gamado-x

**Team 7**

ccosta-c  
gucamelo  
brumarti  
aconceic  
faneves  
hescoval

**Team 8**

jmarinho  
numartin  
ocaio-re  
ismirand  
Pabernar  
Aamaral-

**Team 9**

brolivei  
bsousa-d  
Ide-sous  
Luis-ffe  
pemirand  
rumachad  
dzinchen

**Team 10**

ride-sou  
brpereir  
joaoped2  
vivaccar  
bbento-e  
tchow-so

# HACKATHON

## Online Wine Price Harvesting Challenge

