

# Online Wine Price Harvesting Challenge

The company's on-trade markets and distribution teams are facing an intriguing dilemma: they need to track wine prices on online stores to guide their distribution and pricing strategies. The goal is to create a solution to harvest information from online sales platforms, addressing two crucial aspects – harvesting & treating the data and presenting the data for the markets and distribution teams to analyze.

---

## The Challenge

As a programming student, you have the opportunity to dive into this challenge. Here are the two dimensions you need to address:

### 1. Technical Dimension

#### Daily Information Collection - Web Scraping:

- Web Scraping 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 (listed below) 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.

## Conclusion

*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.

# More information

---

## Challenge Details:

### 1. Websites to Scrape data from:

- [Continente Online](https://www.continente.pt/) (https://www.continente.pt/)
- [Garrafeira Soares](https://www.garrafeira-soares.pt/) (https://www.garrafeira-soares.pt/)
- [Supermercado El Corte Ingles](https://www.elcorteingles.pt/supermercado/) (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; onwine.pt; granvine.com; vinha.pt; adegga.com; portugalvineyards.com)

---

### 2. 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).

# More information

				
	<b>Mateus Rosé</b>	<b>Mateus Sparkling</b>	<b>Trinca Bolotas</b>	<b>Papa Figos</b>
<b>Brand</b>	Mateus	Mateus	Herdade do Peso	Casa Ferreirinha
<b>Sub Brand</b>	N.A.	N.A.	Trinca Bolotas	Papa Figos
<b>Wine Name</b>	Mateus Rosé Original	Mateus Sparkling Brut Rosé	Trinca Bolotas Tinto	Papa Figos Branco
<b>Harvest year</b>	NA	NA	2021*	2022*
<b>Capacity</b>	750 ml	750 ml	750 ml	750 ml
<b>EAN</b>	5601012011500	5601012001310	5601012004427	5601012011920

\*Year may not always be referenced. When referenced, however, this data should be collected.

Find detailed product sheets in attached documents.

### 3. 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**

# More information

**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.

---

## 4. 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.

---

## Submission

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

Submission details:

- How to: **README** file explaining how your application works and the commands you need to run it
- Technical part directory: **webscraping/**
- Dashboard part directory: **hack\_dashboard/**

# More information

## 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)
- 

## Other useful information:

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

## 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”

**| Please remember that it is your responsibility to ensure that the web scraping you undertake is both ethical and lawful. |**