

# Programa Take-Home MLE Challenge

## Welcome

Thank you for your interest in joining the Data & AI team at Programa! We're excited to see what you can build.

This take-home challenge is designed to give you an opportunity to demonstrate your technical skills in a realistic scenario that reflects the kind of problems we solve at Programa. There are no trick questions—we genuinely want to see how you approach problem-solving, structure your code, and handle real-world data challenges.

---

### Timeline & Expectations

- **Duration:** 5 calendar days from when you receive this challenge
- **Estimated effort:** 8–12 hours (you don't need to spend all 5 days on this!)

## Interior Schedule Extraction API

### Objective

Interior designers often create Excel-based product schedules containing product names, materials, finishes, prices, dimensions—and sometimes images.

Your task is to **build a Python-based REST API** that accepts `.xlsx` file uploads and converts them into structured JSON format so that we can import the schedules into programa platform.

NEW: Save hours of time on each project by adding products directly to your Schedule from any ORL [Learn more](#)

Studio Alexandra

Vincent Li

Summary

Financial

Procurement

Approvals

Navigate to Section

Search

Filter

Sort

Projects

TOOLS

Time Tracking

To-Do List

Studio

Trade Portal Discover

Procurement Hub

Purchase Orders

Invoices

LIBRARIES

Address Book

Image Library

Product Library

Schedule Guides

Seating 41

Product Description

PRODUCT DETAILS

Enter Doc Code

KINESIS PERSONAL - VISI...

PRODUCT NAME

Technogym

BRAND

1700.0

WIDTH (MM)

0.0

LENGTH (MM)

2100.0

HEIGHT (MM)

450.0

DEPTH (MM)

1

QTY

2-4 wks

LEAD TIME

Technogym Australia

GIULIO GASPERINI

Draft

Couch for living room

PRODUCT DETAILS

Enter Doc Code

Orange Bubble Couch

PRODUCT NAME

Lapis

BRAND

0.0

WIDTH (MM)

0.0

LENGTH (MM)

760.0

HEIGHT (MM)

1000.0

DEPTH (MM)

1

QTY

6-8 wks

LEAD TIME

LAPIS

MARCELLO VALENTE

Quoting

Armchair

PRODUCT DETAILS

CH02

Throne Chair

PRODUCT NAME

New Tendancy

BRAND

620

WIDTH (MM)

-

LENGTH (MM)

750

HEIGHT (MM)

470

DEPTH (MM)

1

QTY

16-18 wks

LEAD TIME

Supplier

Click to add supplier

Internal Review

Accent Chair - Op 1

LIBRARY

CH02

Triangolo Chair

PRODUCT NAME

Frama

BRAND

480

WIDTH (MM)

-

LENGTH (MM)

680

HEIGHT (MM)

400

DEPTH (MM)

3

QTY

4-6 wks

LEAD TIME

Frama

LISETTE BIRCH

Hidden

Accent Chair - Op 2

HALLWAY

CH03

Pudica

PRODUCT NAME

Matter Made

BRAND

365

WIDTH (MM)

-

LENGTH (MM)

965

HEIGHT (MM)

450

DEPTH (MM)

3124

QTY

In Stock

LEAD TIME

Leibal

SIM PETERS

Approved

accent chair

PRODUCT DETAILS

Enter Doc Code

-

PRODUCT NAME

-

BRAND

-

WIDTH (MM)

-

LENGTH (MM)

-

HEIGHT (MM)

-

DEPTH (MM)

3.33

QTY

-

LEAD TIME

Supplier

Click to add supplier

Draft

+ New

The primary challenge lies in the **variability of schedule formats** from different designers—including inconsistent column names, merged cells, multiple sheets, and diverse layouts. The goal is to **accurately extract and standardize** product data to support seamless schedule imports into our platform.

## Requirements

### 1. Build an API

- Use **FastAPI** or **Flask**
- Implement the endpoint:
  - **POST /parse**
  - Accepts an Excel **.xlsx** file upload (you will be provided with a sample of schedules)
  - Returns structured JSON in the format specified below
- Ensure the implementation is **modular, clean, and production-ready**

### 2. Extract Data

Expected fields to extract:

Field	Description	Type
<code>doc_code</code>	Drawing or reference code	string
<code>product_name</code>	Product display name	string
<code>brand</code>	Product brand	string
<code>colour</code>	Colour name	string
<code>finish</code>	Surface finish	string
<code>material</code>	Main material	string
<code>width</code>	Width in mm	integer
<code>length</code>	Length in mm	integer
<code>height</code>	Height in mm	integer
<code>qty</code>	Quantity	integer
<code>rrp</code>	Recommended retail price	decimal
<code>feature_image</code>	Image filename (optional)	string
<code>product_description</code>	Short description	string
<code>product_details</code>	Additional specifications	string

- Input data may span **multiple sheets**
- Expect **merged cells**, inconsistent headers, and missing values

## Expected JSON Output Format

```
{
  "schedule_name": "Lighting Schedule",
  "products": [
    {
      "doc_code": "L1",
      "product_name": "Minimalist Pendant Light",
      "brand": "Lighting Co",
      "colour": "Black",
      "finish": "Brushed Brass",
      "material": "Metal",
      "width": 300,
```

```
"length": 300,  
"height": 600,  
"qty": 1,  
"rrp": 489,  
"feature_image": "blackpendantlight.jpg",  
"product_description": "A simple pendant light ideal for modern interiors.",  
"product_details": "Install at 2.4m height; supplied with dimmable bulb."  
}  
]  
}
```

## Suggested Libraries

You may use any Python libraries you see fit. For Excel parsing, you might find libraries like `openpyxl` or `pandas` useful.

## Bonus Points

These features are **not required**, but may earn you bonus credit:

- Extract embedded images or image URLs (if available)
- Handle merged cells gracefully
- Handle inconsistent headers and edge cases robustly
- Include a `Dockerfile` for containerization
- Include basic test cases
- Include a **system architecture diagram** (optional but encouraged) showing how you'd deploy this API in a production environment

## Project Structure

- Place any sample `.xlsx` files under `/data`
- Keep code modular (e.g., `/app/parser.py`, `/app/models.py`)
- Include a `README.md` with:

- Setup instructions
  - Example request/response
  - Optional architecture diagram
- 

## Evaluation Criteria

- **Code quality:** Clean, modular, well-documented
  - **Robustness:** Handles diverse real-world input formats
  - **Accuracy:** JSON output structure and correctness
  - **API design:** RESTful and resilient
  - **Bonus:** Docker, tests, image support, architecture diagram
- 

## Submission

- Upload your solution to GitHub
  - Include a detailed system architecture diagram showing your proposed production deployment setup
  - Invite: [joshiain](#) , [foozieakbar](#) and [vinceliorg](#) as collaborators
  - Share the repository link with us via email
- 

✨ **Note:** Image support is a nice-to-have, not a requirement. Focus on accuracy, robustness, and thoughtful code design. Good luck, and have fun!