

Smart UPC Lookup

This project provides a Python script (smart.py) to:

1. Read UPC codes from barcodes.csv.
2. Query the Go-UPC API (with retry/backoff) to retrieve product data.
3. Download product images (converting WebP to PNG when needed).
4. Write an Excel file (product_data.xlsx) containing UPC, Product Name, Brand Name, and Image URL.
5. Record any UPCs that failed Go-UPC lookup in failed_upcs.xlsx.
6. Query the Open Food Facts (OFF) API for any failed or partial entries to fill missing fields.
7. Download OFF images and save them to images/.
8. Write OFF results to off_products.xlsx.

Prerequisites

- **Python 3.7+**
- Internet access for API calls

Environment Setup

It's recommended to run the script inside a Python virtual environment to avoid polluting system packages:

Create a new venv (only once)

python3 -m venv venv

Activate it

source venv/bin/activate # macOS/Linux

venv\Scripts\activate.bat # Windows

Required Python Packages

Install dependencies via pip:

pip install requests openpyxl pillow

In order to run the script

python3 smart.py

Project Structure

```
— barcodes.csv      # Input CSV with a `barcode` column
— smart.py          # Main script
— images/           # Directory for downloaded images
— product_data.xlsx # Excel output from Go-UPC
— failed_upcs.xlsx  # Excel file listing UPCs Go-UPC could not find
— off_products.xlsx # Excel output from OFF API for failed UPCs
--- README.md       # This file
```

Usage

1. **Populate** barcodes.csv with one UPC per row, under the header barcode.
2. **Run** the script:

python3 smart.py

3. **Observe** console logs for progress and any fallback messages:
 - Please Wait... querying Go-UPC for {UPC}
 - 429 rate limit... retry messages
 - Fallback logs: field missing in Go-UPC, fetching from OFF API
 - Summary at end: counts of succeeded/failed for Go-UPC and OFF.
4. **Inspect** outputs:
 - **product_data.xlsx**: main results (UPC, Name, Brand, Image URL)
 - **failed_upcs.xlsx**: UPCs that Go-UPC could not find
 - **off_products.xlsx**: OFF API results for failed UPCs

- **images/**: downloaded and converted images, named {Brand}_{ProductName}.ext

Error Handling & Edge Cases

- **CSV errors**: missing file or missing barcode column will exit with an error.
- **Network/API errors**: timeouts, HTTP errors, rate limits are logged and/or retried.
- **Partial data**: missing name/brand/image from Go-UPC triggers an OFF API call.
- **Image download**: unsupported formats or I/O errors are caught and logged.
- **Excel save**: file write errors will abort with a message.

Customization

- **API key**: update API_KEY in smart.py if your key changes.
- **Retry settings**: adjust MAX_RETRIES, INITIAL_WAIT, and REQUEST_DELAY as needed.
- **Filename rules**: modify make_image_filename(...) for custom naming conventions.