

## PDF-to-Excel Kanban Analyzer

### Purpose

This script automatically extracts part information from a PDF file containing logistics labels and generates a summary Excel file. It is designed for analyzing Kanban-based shipments, allowing you to quickly determine:

- a list of unique items (Part Number);
- the Kanban code (Kanban);
- quantity per label (Lot);
- the number of times each item appears (Labels Count);
- the total quantity (Total Qty).

### How It Works

The script performs the following steps:

1. Loads and reads all text content from the PDF;
2. Uses regular expressions to extract:
3. Part Number (e.g., 040-K774-A000);
4. Kanban (e.g., K774);
5. Lot quantity (e.g., 4800 PC);
6. Builds a table with all matches;
7. Groups rows by Part Number + Kanban + Lot;
8. Counts how many labels were found for each part;
9. Calculates the total quantity (Lot × Labels Count);
10. Saves the result to an Excel .xlsx file.

### How to Use

1. Make sure you have a PDF file with standardized logistics labels;
2. Edit the script and replace the value of pdf\_path with the full path to your file;
3. Open a terminal and install the required Python libraries (only once) (pip install PyPDF2 pandas openpyxl)
4. Run the script;
5. An Excel file will be saved in the same folder with the extracted and grouped data.

### Important

- The PDF file must follow a consistent label structure with sections like (P) Part Number, (K) Kanban, Description, and Lot ... PC;
- For each Part Number + Lot, the quantity is assumed to be per label — the script does not sum mixed lots;
- The PDF must contain machine-readable text (not scanned images). For scanned PDFs, OCR is required.

### Requirements

Python

PDF file with extractable text content

Excel (for reading the output .xlsx file)