



Smart Extract

Learn how to use Smart Extract to import complex files without manual preparation



What is Smart Extract?

Smart Extract helps you extract structured, usable data from files that don't follow a standard tabular format.

Smart Extract is designed to work with more complex files that contain:

<i>Global or local metadata</i>	<i>Multiple header rows</i>	<i>Multiple tables within a sheet</i>
<i>Summary or total rows</i>	<i>Shifted data tables</i>	<i>Hierarchical data</i>

You'll be prompted to use Smart Extract when Flatfile detects these characteristics during the initial data upload and extraction process.

Warehouse: Miami DC				Warehouse: Seattle DC				Warehouse: New York DC						
Shipment ID	Date	Destination	Carrier	Status	Shipment ID	Date	Destination	Carrier	Status	Shipment ID	Date	Destination	Carrier	Status
WA_041	2025-06-13	Orlando	XPO Logistics	In Transit	WS_061	2025-06-22	Seattle	XPO Logistics	Delivered	WE_061	2025-06-03	Chicago	FedEx	In Transit
WA_042	2025-06-13	Miami	XPO Logistics	Pending	WS_062	2025-06-23	Atlanta	DHL	Delayed	WE_062	2025-06-03	Los Angeles	UPS	Delivered
WA_043	2025-06-14	Phoenix	UPS	In Transit	WS_063	2025-06-24	Dallas	FedEx	Delayed	WE_063	2025-06-04	Los Angeles	UPS	In Transit
WA_044	2025-06-15	Los Angeles	XPO Logistics	Delayed	WS_064	2025-06-25	Los Angeles	R-I-L Carrier	Delayed	WE_064	2025-06-05	New York	DHL	Pending
WA_045	2025-06-16	Los Angeles	XPO Logistics	Delayed	WS_065	2025-06-26	New York	FedEx	Delayed	WE_065	2025-06-06	Orlando	UPS	Delivered
WA_046	2025-06-17	Atlanta	DHL	In Transit	WS_066	2025-06-27	Miami	DHL	Pending	WE_066	2025-06-07	New York	R-I-L Carrier	Delivered

Customer: Aerovents				Customer: Acme				Customer: Veltrix			
Location ID	Last Delivered Address	Delivery C#	Delivered?	Location ID	Last Delivered Address	Delivery C#	Delivered?	Location ID	Last Delivered Address	Delivery C#	Delivered?
LC_001	2025-06-22 New York	DHL	Delivered	LC_001	2025-06-22 New York	DHL	Delivered	LC_013	2025-06-22 New York	DHL	Delivered
LC_002	2025-06-23 Atlanta	R-I-L Carrier	Delivered	LC_008	2025-06-23 Atlanta	R-I-L Carrier	Delivered	LC_014	2025-06-23 Atlanta	R-I-L Carrier	Delivered
LC_003	2025-06-24 Phoenix	XPO Logistics	Delivered	LC_009	2025-06-24 Phoenix	XPO Logistics	Delivered	LC_015	2025-06-24 Phoenix	XPO Logistics	Delivered
LC_004	2025-06-25 Dallas	XPO Logistics	Delivered	LC_010	2025-06-25 Dallas	XPO Logistics	Delivered	LC_016	2025-06-25 Dallas	XPO Logistics	Delivered
LC_005	2025-06-26 Los Angeles	UPS	Delivered	LC_011	2025-06-26 Los Angeles	UPS	Delivered	LC_017	2025-06-26 Los Angeles	UPS	Delivered
LC_006	2025-06-27 Seattle	FedEx	In Transit	LC_012	2025-06-27 Seattle	FedEx	In Transit	LC_018	2025-06-27 Seattle	FedEx	In Transit

Customer: Thornwell				Customer: Lunara One				Customer: Quanton			
Location ID	Last Delivered Address	Delivery C#	Delivered?	Location ID	Last Delivered Address	Delivery C#	Delivered?	Location ID	Last Delivered Address	Delivery C#	Delivered?
LC_025	2025-06-22 New York	DHL	Delivered	LC_031	2025-06-22 New York	DHL	Delivered	LC_037	2025-06-22 New York	DHL	Delivered
LC_026	2025-06-23 Atlanta	R-I-L Carrier	Delivered	LC_032	2025-06-23 Atlanta	R-I-L Carrier	Delivered	LC_038	2025-06-23 Atlanta	R-I-L Carrier	Delivered
LC_027	2025-06-24 Phoenix	XPO Logistics	Delivered	LC_033	2025-06-24 Phoenix	XPO Logistics	Delivered	LC_039	2025-06-24 Phoenix	XPO Logistics	Delivered
LC_028	2025-06-25 Dallas	XPO Logistics	Delivered	LC_034	2025-06-25 Dallas	XPO Logistics	Delivered	LC_040	2025-06-25 Dallas	XPO Logistics	Delivered
LC_029	2025-06-26 Los Angeles	UPS	Delivered	LC_035	2025-06-26 Los Angeles	UPS	Delivered	LC_041	2025-06-26 Los Angeles	UPS	Delivered
LC_030	2025-06-27 Seattle	FedEx	In Transit	LC_036	2025-06-27 Seattle	FedEx	In Transit	LC_042	2025-06-27 Seattle	FedEx	In Transit

Master Sh#	HIER0005	Weight (kg)	Temperature	Fragile	Dimensions
31	Container I Contents	8		Yes	56x25x37 cm
32	HIER0005 - Electronics	350		No	36x6x47 cm
33	HIER0005 - Hardware	15.3		No	

Shipment ID	Date	Deposit	Destination	Weight (kg)	Volume (m ³)	Cost (USD)	Carrier	Mode	Tracking #	Status	Driver	Truck ID	Dispatch TI	Arrival Time	Temp	Cont#	Haz
SUM10001	2025-06-01	Dallas DC	Chicago	1818	11.96	997	R-I-L Carrier	Truck	TRK000001	Delivered	Driver:399	TRUCK:687	2025-06-01	2025-06-01	No	No	No
SUM10002	2025-06-02	Dallas DC	Chicago	808	10.45	904	DHL	Air	TRK000002	Pending	Driver:610	TRUCK:654	2025-06-02	2025-06-02	No	No	No
SUM10003	2025-06-03	Dallas DC	Miami	1547	11.69	1336	FedEx	Ocean	TRK000003	Pending	Driver:186	TRUCK:369	2025-06-03	2025-06-03	No	No	Yes
SUM10004	2025-06-04	Seattle DC	Chicago	1495	7.35	885	DHL	Truck	TRK000004	Delivered	Driver:276	TRUCK:312	2025-06-04	2025-06-04	No	No	No

How does it work?

When you upload a file, Smart Extract will:

- 1: Begin an initial extraction pass**

Flatfile first runs a standard extractor and analyzes the file shape. If the file does not contain any complexities, you can proceed directly to mapping. This allows you to import simple files quickly and leverage Smart Extract only when necessary.
- 2: Assess the file shape**

If the file contains complexities—like metadata or multi header sheets—Flatfile will automatically recommend using Smart Extract.
- 3: Enable a preview**

Once Smart Extract analysis completes, you can review the structured output before officially importing your data.



Common use cases by industry

Complex files aren't exclusive to any one industry. Smart Extract accelerates data import for enterprises across verticals. Here are some examples:

Healthcare IT

Smart Extract helps healthcare IT teams structure and import complex clinical files.

- **EHR system exports:** Patient records or claims data pulled from systems like Epic or Cerner often include multiple tables, extra headers, or embedded metadata that prevent easy imports.
- **Clinical trial and care quality data:** These datasets frequently include summary rows, grouped sections, or non-standard formatting required for CMS or FDA reporting.
- **Lab results and code sets:** Lab exports may list tests, values, and codes in nested or repeated formats.

Fintech

Smart Extract helps fintech teams structure and import complex financial records.

- **KYC and transaction exports:** Files from banking platforms or compliance tools often contain multiple headers, embedded metadata, or summary rows that disrupt standard imports.
- **Loan and portfolio reports:** These files frequently include subtotal rows and hierarchical groupings by account or borrower.
- **Multi-currency or cross-border data:** International files may mix formats, currencies, or date conventions across rows or sections. Smart Extract structures the data into usable tables while preserving currency indicators or exchange notes.

Field Services

Smart Extract helps Field Services platforms structure and import inconsistent files from customers.

- **Customer-generated exports:** Implementation teams often receive job logs, service records, or dispatch summaries exported from legacy systems or spreadsheets. These files frequently include global metadata (like export notes or technician summaries), repeated header rows, and shifted tables.
- **Maintenance logs and inspection histories:** These datasets often include subtotals, summary rows, or nested task structures grouped by asset or location.
- **Technician compliance reports:** Certification records or audit logs frequently feature hierarchical data, irregular headers, or metadata mixed with form entries.



Where you'll find Smart Extract in Flatfile

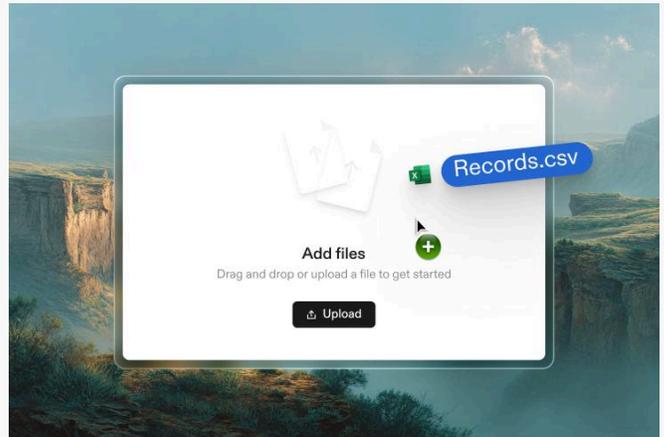
Once you're logged in to your Flatfile account, you'll find Smart Extract within the Files section within a Flatfile Space.

How to use Smart Extract:

1. Upload your file to Flatfile

Go to the Files section of your Space.

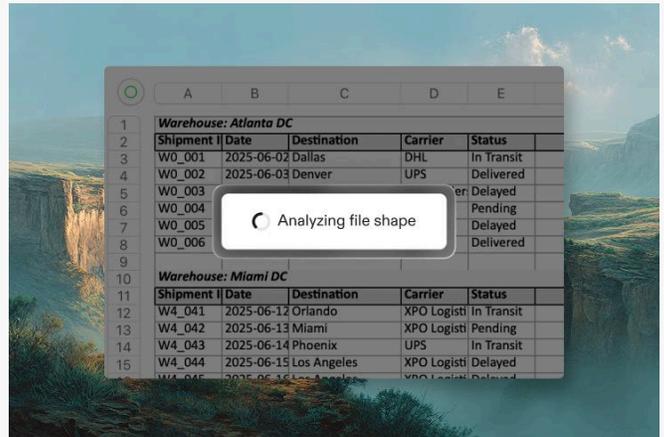
Drag and drop your CSV or Excel (XLSX) file or click to upload it.



2. Let Flatfile analyze the file structure

Flatfile runs an automatic scan of your file after upload.

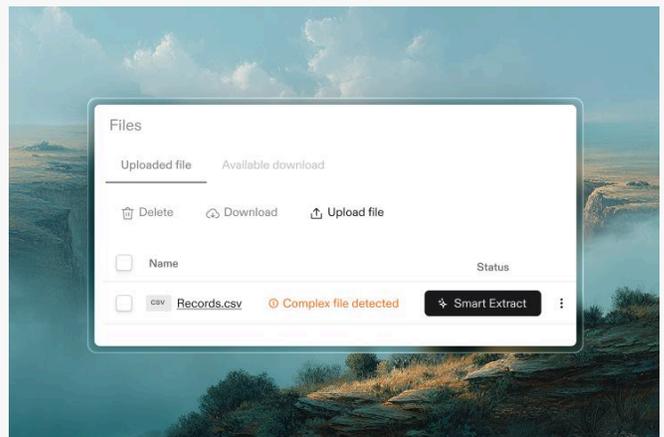
If it detects a complex structure—like sheets with multiple headers, global or local metadata, or hierarchical data—it will recommend using Smart Extract.



3. Look for the Smart Extract button

If Smart Extract is recommended, a button will appear next to your file as the primary action.

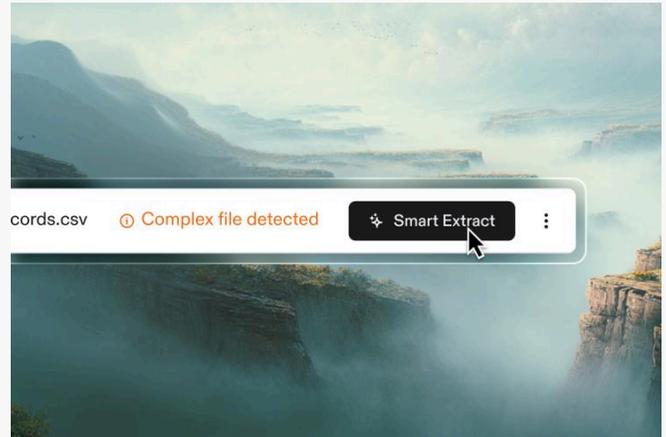
If Smart Extract is not triggered automatically, you can opt to run it manually using the three-dot menu to the right of the file name.





4. Run Smart Extract

Click “Smart Extract” to begin AI analysis.
Flatfile will build a custom extractor based on your file's structure.
This usually takes about 1–2 minutes, depending on file size.



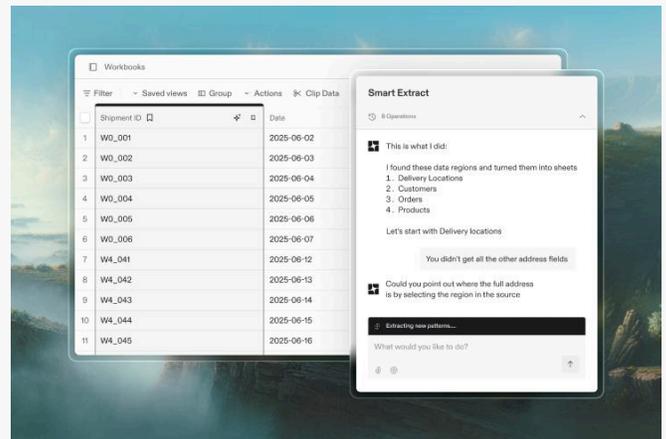
5. Review and adjust the extracted data

When Smart Extract finishes, click the blue “Review Extraction” button next to your file name to open the preview window.

You'll see:

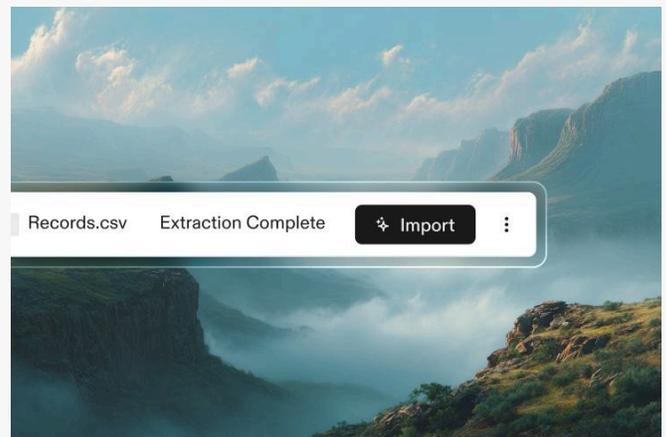
- The AI-generated preview of your file
- A chat interface where you can give instructions if the preview doesn't show exactly what you need the output to be

Use the chat assistant to refine the structure as needed.



6. Finalize the extraction

When everything looks correct, click “Import.”
The structured data will move into the next step of your workflow—such as mapping or transforming—based on your project setup.





Prompting best practices

Smart Extract is incredibly capable at recognizing data structure and formatting it accurately within Flatfile. However, occasionally you'll need to revise what you see in the preview with the chat assistant.

When you prompt the chat assistant to perform actions on your data, keep the following best practices in mind:

1. Be specific

A vague prompt like “there is missing data” could be interpreted many different ways. Telling the Smart Extract chat assistant exactly what you want changed removes any guesswork and ensures you get the exact result you want.

✓ Good prompt: “The column Address is missing information”

✗ Weak prompt: “There is missing information” (too vague)

2. Prompt structural adjustments, not data transformation

You'll be able to normalize values and perform other manipulations to values in your dataset at a later stage. When working with the Smart Extract assistant, focus on the structure that was imported and ensuring that all data is appropriately represented.

✓ Good prompt: “Only include the main table—ignore the summary above it.”

✗ Weak prompt: “Fix the formatting” (too vague, not structural)

3. Call out header rows or grouping structures explicitly

If the file has multi-row headers or grouped sections, tell Smart Extract where the real header is or how to flatten nested data.

✓ Good prompt: “The header row begins with ‘First Name’ but has a second header in the row above that needs to be combined.”

✗ Weak prompt: “The headers are messed up” (not enough detail)

4. Help the assistant understand how the data is grouped

When files use indentation, spacing, or repeated labels, describe how the data is organized to get to the desired output.

✓ Good prompt: “Status in row 1 is local metadata and is repeated every 100 or so rows. Status updates to active then to inactive later in the file.”

✗ Weak prompt: “Status: ‘Invalid’ isn't correct”

5. If unsure, ask for a structural summary first

You can use the assistant to describe how it interpreted the file before you give a command. This helps avoid misunderstandings.

✓ Good prompt: “What tables or sections did you detect in this sheet?”



Limitations

Smart Extract is specifically designed to handle the six types of complexities outlined early in the guide. There are a few limitations to keep in mind. Smart Extract is unable to:

Extract data from file types other than CSV or XLSX

Interpret visual-only formatting (e.g., colors, shading)

Handle files with entirely unstructured text blocks

Guess context with no repeatable structure or signal

Privacy and security

Data protection

All data is encrypted in transit and at rest, with isolated databases for each tenant or data upload to ensure strong logical segregation. We host our infrastructure on Amazon Web Services (AWS), using multiple availability zones for redundancy and uptime. Routine backups are fully encrypted and regularly tested to confirm restoration reliability.

AI access

Flatfile's AI operates entirely within its own infrastructure and does not send customer-uploaded data to third-party AI vendors. We do not use customer data for training external models, and our AI systems do not retain raw customer data for generalized model learning.

Access control

Each guest sees only the data they're allowed to access, based on assigned permissions. Admins can manage user access settings and view a complete audit log of every change made.

Data history

Flatfile automatically keeps a version history of your data, so you can see how it has changed over time. Admins can restore previous versions using saved snapshots, and all edits are recorded in the audit log for full traceability.



FAQs

Why don't I see Smart Extract in my account?

To have Smart Extract enabled in your account, reach out to the Success team (success@flatfile.com).

What file types are currently supported?

Smart Extract currently supports CSV and Excel (XLSX) files. Other file types, like PDF, aren't yet supported.

Does Smart Extract use my existing schema, validation rules, or mappings?

No, each file is analyzed independently. Smart Extract doesn't use past mappings, validation rules, or destination schemas from your project setup.

What if Smart Extract doesn't work for my file?

If Smart Extract isn't a good fit for your use case, you can use a custom extractor plugin. Custom extractors let you write your own logic for handling complex or unsupported file types, and then run them inside your Flatfile environment. You can find a library of prebuilt plugins [here](#).

Why isn't Smart Extract triggering when I upload the file?

Files with standard row-column layouts won't trigger Smart Extract. If your file does not contain any of the complexities that Smart Extract analyzes for, it will go through the normal import process.

Can I manually trigger Smart Extract for a file if it doesn't automatically trigger?

Yes. Under Files, navigate to the three dot icon to the right of your file name and click on Smart Extract from the drop down menu to trigger it manually.

The Smart Extract output isn't accurate; how can I fix this?

If the structure you see in the preview isn't correct, you can provide context and refine via the chat assistant.

Does Smart Extract modify my source files?

Never. All processing happens on a secure copy on our platform.

What if I need help?

You can find more resources for using Smart Extract in our [Knowledge Base](#). If you're still stuck or have questions, reach out to our team at support@flatfile.com.
