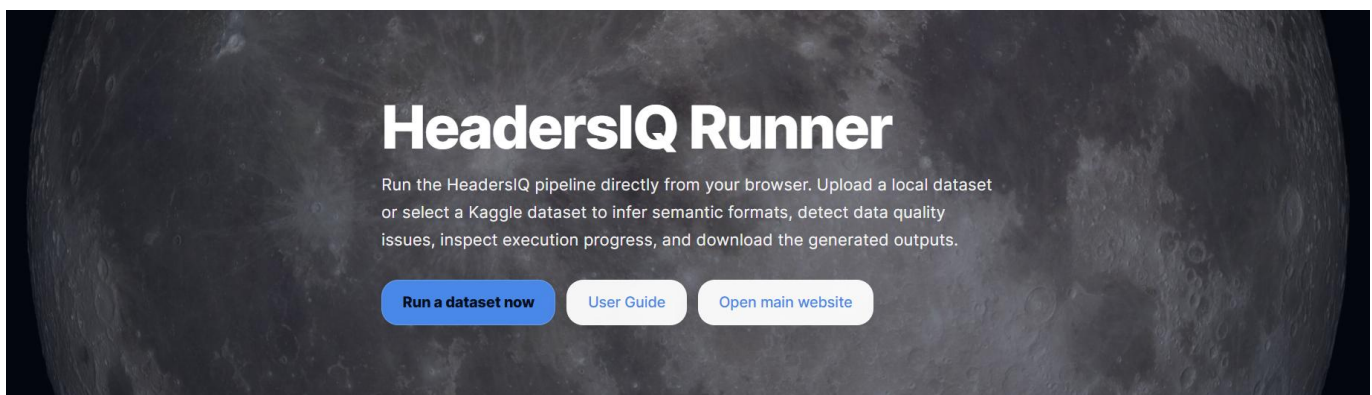


Headers IQ Code Runner

User Guide

Run local and Kaggle tabular datasets through the **HeadersIQ** semantic interpretation and Data Quality Assessment pipeline.



HeadersIQ Runner user-facing interface.

[Launch HeadersIQ Runner](#) | [Online User Guide](#)

Version 1.2 | 18 July 2026

Contents

1. Before you begin
2. Open the Runner
3. Choose a data source
4. Run the pipeline
5. Follow execution progress
6. Review and download the results
7. Supported tabular formats
8. Common problems and solutions
9. Privacy, security, and appropriate use
10. Quick user checklist

This guide is written for users of the public Code Runner. It does not describe the research implementation or source code.

1. Before you begin

HeadersIQ Runner is a browser-based interface for the current **HeadersIQ** semantic interpretation and Data Quality Assessment pipeline. No local installation is required.

- **Use a tabular dataset with column headers, in other words, the columns need clear names on them. The first row must contain the column names, not ordinary data values.**
- Use a supported local file or a tabular Kaggle dataset.
- Keep the progress page open until the result page appears.
- Allow additional time for large datasets.

Automatic header check. Before semantic interpretation begins, the **HeadersIQ** code checks whether the first suitable row appears to contain column names. If it appears to contain data, the run stops and the result page explains the detected reason.

Do not upload confidential, personal, sensitive, or proprietary data. The public service is a research prototype. HTTPS protects data while it is transmitted, but uploaded files and generated outputs may remain on the server during and after processing. Do not assume immediate deletion.

2. Open the Code Runner

1. Open <https://headersiq-runner.australiaeast.cloudapp.azure.com>
2. Select **Run a dataset now** to move directly to the input form.
3. Click **User Guide** to return to this manual at any time.

HeadersIQ Runner

Run the HeadersIQ pipeline directly from your browser. Upload a local dataset or select a Kaggle dataset to infer semantic formats, detect data quality issues, inspect execution progress, and download the generated outputs.

Run a dataset now User Guide Open main website

Run HeadersIQ on a local file or Kaggle dataset

Choose a local tabular file or a Kaggle dataset. The Runner executes the current HeadersIQ workflow and presents progress, summary metrics, and downloadable outputs in the browser.

Source mode

Upload my dataset

Dataset file

Drag and drop your dataset here

or click to browse. Supported formats: CSV, TXT, Excel (.xls, .xlsx), TSV, DATA, JSON, JSONL and Parquet. Files must be loadable as tabular data.

No file selected yet.

Dataset title

My dataset

Dataset URL

Optional source URL

Run pipeline Back to main site

Built for practical, reproducible runs

The public website presents the research, publications, and supporting artefacts. This Runner provides the execution layer for uploading tabular data, running the current HeadersIQ pipeline, following its progress, and downloading the generated outputs.

LARGE RUN
1.4M rows
Demonstrated in a large-scale dataset assessment.

EXECUTION
374 sec
An example of processing a substantial data volume.

DETECTED SIGNAL
7.7M cells
Violating cells identified in one large assessment.

- Upload a supported local tabular file or provide a Kaggle dataset.
- Follow the pipeline phases while the analysis runs.
- Inspect the HeadersIQ summary and download generated artefacts.
- This public service is a research prototype for demonstration and evaluation.

Figure 1. Current Runner interface. The form on the left is used to submit a local file or Kaggle dataset.

3. Choose a data source

The **Source mode** menu provides two options: **Upload my dataset** and **Kaggle dataset**.

3.1 Upload a local dataset

1. Keep **Upload my dataset** selected.
2. Drag the file into the upload area under **Dataset file** or click the area to browse your device.
3. Optionally enter a **Dataset title**. When left blank, the file name is used as a fallback.
4. Optionally enter the original **Dataset URL** for provenance.

New runner interface

Run HeadersIQ on a local file or Kaggle dataset

Choose a local tabular file or a Kaggle dataset. The Runner executes the current HeadersIQ workflow and presents progress, summary metrics, and downloadable outputs in the browser.

Source mode

Upload my dataset

Dataset file



Drag and drop your dataset here

or click to browse. Supported formats: CSV, TXT, Excel (.xls, .xlsx), TSV, DATA, JSON, JSONL and Parquet. Files must be loadable as tabular data.

No file selected yet.

Dataset title

My dataset

Dataset URL

Optional source URL

Run pipeline

Back to main site

Figure 2. Local upload mode. A dataset file is required; the title and source URL are optional.

3.2 Use a Kaggle dataset

1. Select **Kaggle dataset** from the **Source mode** menu.
2. Enter either owner/dataset-name or the full Kaggle dataset URL in the **Kaggle dataset input**.
3. Review the Kaggle area label. The default value is General.
4. Confirm that the Kaggle page contains a tabular file.

New runner interface

Run HeadersIQ on a local file or Kaggle dataset

Choose a local tabular file or a Kaggle dataset. The Runner executes the current HeadersIQ workflow and presents progress, summary metrics, and downloadable outputs in the browser.

Source mode

Kaggle dataset

Kaggle dataset input

owner/dataset-name or full Kaggle URL

This runner is intended for tabular Kaggle datasets. After download, HeadersIQ looks for a compatible file that can be loaded as tabular data. Some Kaggle pages contain images, audio, PDFs, NLP corpora, README files, nested folders, or other non-tabular files. In those cases, the run may fail because no suitable tabular file is found. If this happens, download the dataset manually and upload a specific tabular file instead.

Accepted tabular formats: CSV, CSV.GZ, TSV, TXT, DATA, Excel (.xls, .xlsx), JSON, JSONL and Parquet.

Kaggle area label

General

Run pipeline

Back to main site

Figure 3. Kaggle mode. The Runner downloads the dataset and searches for a compatible tabular file.

Some Kaggle pages contain images, audio, PDFs, README files, NLP corpora, archives, or nested folders instead of a directly usable table. When no compatible tabular file is found, download the required table manually and select Upload my dataset.

4. Run the pipeline

1. Review the selected file or Kaggle identifier.
2. Select **Run pipeline**.
3. Wait while the Runner validates the submission and starts the **HeadersIQ** workflow.

Execution time depends on file size, number of rows, number of columns, file format, and server workload. Avoid submitting the same run repeatedly while the first run is still active.

5. Follow execution progress

After the run starts, the progress page reports the current phase, state, last message, update time, and elapsed time. It redirects automatically when the pipeline finishes. If the header check fails, the page redirects to a failed-run result with a clear explanation.

Figures 4a-c show an actual execution of the Hotel bookings demand Kaggle dataset through the three pipeline phases.

Pipeline running

<https://www.kaggle.com/datasets/qucwang/hotel-bookings-analysis-dataset>

The pipeline is running in the background. This page updates the current phase automatically.

Current phase updated at 2026-07-18 15:18:40		
Elapsed at phase update [waiting]		
Current phase 1/3	State running	Job ID 9794e0e8-b34e-47b4-93fd-0021c506062a
Last message Harvesting Kaggle metadata and preparing input files		Next action This page will redirect to the result when the run finishes.

Figure 4a. Actual Phase 1/3 execution screen. HeadersIQ harvests Kaggle metadata and prepares the input files.

- **Current phase:** indicates how far the workflow has progressed.
- **State:** normally shows queued, running, finished, or failed.
- **Last message:** identifies the most recent pipeline activity.
- **Next action:** confirms that the page will redirect to the result.

Pipeline running

<https://www.kaggle.com/datasets/qucwang/hotel-bookings-analysis-dataset>

The pipeline is running in the background. This page updates the current phase automatically.

Current phase updated at
2026-07-18 15:19:15

Elapsed at phase update
15.76

Current phase
2/3

State
running

Job ID
9794e0e8-b34e-47b4-93fd-0021c506062a

Last message
Running semantic type detection

Next action
This page will redirect to the result when the run finishes.

Figure 4b. Actual Phase 2/3 execution screen. HeadersIQ runs semantic type detection.

Pipeline running

<https://www.kaggle.com/datasets/qucwang/hotel-bookings-analysis-dataset>

The pipeline is running in the background. This page updates the current phase automatically.

Current phase updated at
2026-07-18 15:19:18

Elapsed at phase update
19.3

Current phase
3/3

State
running

Job ID
9794e0e8-b34e-47b4-93fd-0021c506062a

Last message
Running data quality assessment and writing outputs

Next action
This page will redirect to the result when the run finishes.

Figure 4c. Actual Phase 3/3 execution screen. HeadersIQ performs the Data Quality Assessment and writes the outputs.

6. Review and download the results

A successful run displays the dataset name, summary metrics, status, pipeline timing, and available download buttons.

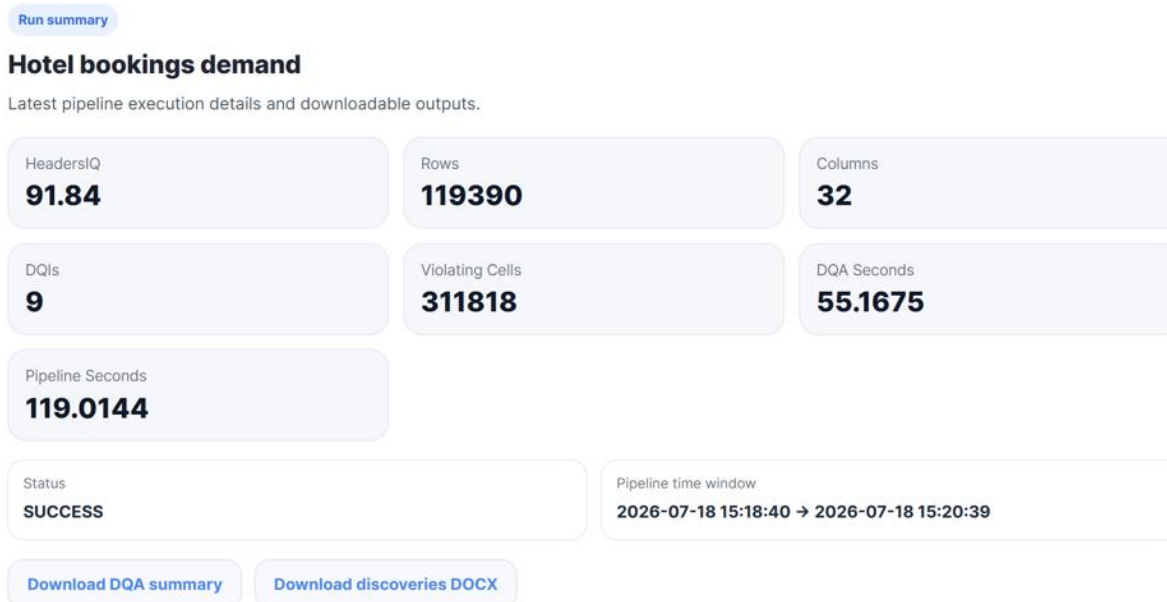


Figure 5. Actual successful result for the Hotel bookings demand Kaggle dataset.

Metric	Meaning
HeadersIQ	The overall <i>HeadersIQ</i> value returned by the current pipeline. Interpret it together with the detailed issues and outputs.
Rows	The number of dataset rows processed or reported by the pipeline.
Columns	The number of analysed columns.
DQIs	The total number of Data Quality Issues identified by the current assessment.
Violating Cells	The number of distinct cells associated with one or more detected issues.
DQA Seconds	Time spent in the Data Quality Assessment part of the workflow, when available.
Pipeline Seconds	Total elapsed time for the complete browser-triggered pipeline.
Status	SUCCESS indicates normal completion. FAILED indicates that the run stopped before completion; when available, the result page displays a clear explanation of the cause.

6.1 Available downloads

- **Download DQA summary:** downloads the generated Data Quality Assessment summary when available.
- **Download Discoveries DOCX:** downloads the generated discoveries document when available.
- A download button appears only when the corresponding file was generated and remains available on the server.

6.2 Failed runs

When the pipeline stops before completing the analysis, the result page explains why the run failed. Summary metric cards that were not produced are omitted, while the elapsed pipeline time and FAILED status remain visible.

Run summary

wine.data

The pipeline stopped before the analysis was completed.

Why the run failed

Header row not detected. Numeric-looking first-row values not found in the Formats Dictionary were detected: 1, 14.23, 1.71, 2.43, 15.6, 127, 2.8, 3.06, 0.28, 2.29, 5.64, 1.04, 3.92, 1065.

HeadersIQ requires a tabular dataset with column headers. Please add or correct the column names and run the dataset again.

Pipeline Seconds

11.6130

Status

FAILED

Pipeline time window

2026-07-18 22:07:44 → 2026-07-18 22:07:56

Run another dataset

Figure 6. Actual failed-run result for a dataset whose header row was removed. The Runner explains that the first row appears to contain data and reports the repeated values that supported the decision.

- **Why the run failed:** states the detected problem and the evidence found in the first row.
- **Pipeline Seconds:** shows how long the pipeline ran before it stopped.
- **Run another dataset:** returns to the input form after the dataset has been corrected.

7. Supported tabular formats

Source	Formats	Important condition
Local upload	CSV, Excel (.xls, .xlsx), TXT, DATA, TSV, JSON, JSONL, Parquet	The file must be loadable as tabular data and expose column names. All files must contain a header row.
Kaggle dataset	CSV, Excel (.xls, .xlsx), TXT, DATA, JSON, JSONL, Parquet, CSV.GZ, TSV	The Kaggle dataset must contain at least one compatible tabular file that the Runner can locate and load, with usable column names or schema fields.

8. Common problems and solutions

Message or symptom	Likely cause	What to do
Header row not detected.	The first suitable row appears to contain data rather than column names, for example, repeated values or numeric-looking values that are not defined as valid labels.	Add or correct the column names in the first row, save the dataset, and run it again.
Please choose a dataset file to upload.	No local file was selected.	Select or drag a supported file, then run the pipeline again.
No suitable tabular file is found in a dataset.	The page contains non-tabular content, nested files, or an unsupported structure.	Download the required table manually and upload that specific file.
Result not found or expired.	The application restarted or the in-memory job reference is no longer available.	Return to the form and start the run again.
The run remains active for a long time.	The dataset is large, complex, or the server is busy.	Keep the progress page open and allow additional time. Avoid repeated submissions.
A download button is missing.	The run failed, or that artefact was not generated or is no longer present.	Review the status and rerun the dataset if appropriate.
The Code Runner cannot open.	Temporary service, network, DNS, or certificate problem.	Refresh once, verify the HTTPS address, and try again later.

9. Privacy, security, and appropriate use

- The public address uses HTTPS to encrypt data while it travels between the browser and the Runner.
- HTTPS does not make a public research prototype suitable for confidential information.
- Uploaded files and generated outputs are processed and stored on the Runner server.
- The current interface does not promise immediate deletion or a fixed retention period.
- Use public, synthetic, anonymised, or otherwise non-sensitive data.
- Do not upload passwords, credentials, personal identifiers, protected records, commercial secrets, or restricted datasets.

10. Quick user checklist

- ☐ My data is tabular, and its first row contains column headers rather than data values.
- ☐ My file format is supported.
- ☐ My data is not confidential or sensitive.
- ☐ I selected the correct source mode.
- ☐ I kept the progress page open.
- ☐ I reviewed the status and summary metrics.
- ☐ I downloaded the available artefacts.