Competitors

1.
[Free online tool for Data Cleaning (datablist.com)](datablist.com)

Here is a quick summary of the clean-up operations found in this article:

- [Convert text to Datetime, Number, Boolean](#)
- [Convert HTML to text (remove HTML tags)](#)
- [Remove extra spaces from texts](#)
- [Normalize your data](#)
- [Remove symbols from texts](#)
- [Split full name into First Name and Last Name](#)
- [Deduplicate items](#)
- [Extract email addresses, URLs, etc. from texts](#)
- [Write custom transformations with JavaScript code](#)
- [Validate Email Addresses](#)

Images

Data

Untitled Collection

Customize your properties

Step 2/3 - Scraped Organizations.csv

Select the file columns you want to import and define your column data types.

Name	Data Type	Import?
Index	Number	⬤
Organization Id	Text	⬤
Name	Text	⬤
Scraped on	Text	⬤
Website	Url	⬤
Country	Text	⬤
Description	Text	⬤
Founded	Number	⬤
Industry	Text	⬤
Number of employees	Number	⬤
To Contact	Text	⬤

Help and Contact

Organizations [⏸ Sync Paused] [🔍 Search] [▽] | Import Export Clean Enrich Edit [New +] ⋮

n	Founded	Industry	Number of er		ems		
d global system e...	1988	Dairy	3930	🔍 Duplicates Finder			
nted fault-tolerant...	2018	Library	7871	🔲 Find & Replace			
tualized bifurcate...	2003	Hospital / Health Care	3095	⇄ Text to Datetime, Number, Ch...			
gurable next gene...	1989	Supermarkets	5422	💬 Ask for more			
tangible methodol...	2013	Nanotechnology	3135	valid			

aused [🔍 Search] [▽]

Long Text   +

Lorem ipsum dolor sit amet...

Lorem ipsum dolor sit amet,...

Lorem ipsum dolor sit amet,...

Lorem ipsum dolor sit amet,...

Lorem ipsum dolor sit amet,...

Lorem ipsum dolor sit amet,...

Lorem ipsum dolor sit amet,...

Lorem ipsum dolor sit amet,...

Lorem ipsum dolor sit amet,...

Lorem ipsum dolor sit amet,...

Lorem ipsum dolor sit amet,...

Bulk Edit   ✕

— **Options**

Edit data on
Long Text   ⌄

Action
Change text case   ⌄

✓ Uppercase
Lowercase
Capitalize all words
Capitalize only first word

Cancel   **Preview edits**

---

ur-tete-2 ⏸ Sync Paused   🔍 Search   ▽

Title	Location
Chasseur de tête indépendant chez CH...	Boulogne-Bill
Chasseur de Têtes Recrutement Débus...	Paris
Recrutement, Chasseur de tête, Conseil...	Paris et périp
Chasseur de tête	Paris
Chasseuse de têtes Headhunter chez M...	Paris
Talent acquisition chasseur de tête che...	Paris
Consultante Senior / Chasseur de Tête -...	Bordeaux et p
chasseur de tete	Paris et périp
Recruteur/ Chasseur de Tête	Paris

Bulk Edit   ⦾   ✕

— **Options**

Edit data on
Title   ⌄

Action
Remove Symbols   ⌄

This action will remove non-letter and non-digit characters from your text.

☑ Keep punctuation

If checked, will remove all symbols except punctuation ('-', '.', ',', ':', '!', '?', '@', etc.).

2. Open refine

    a. Exploration -

You can check data types at any time by:

- clicking "edit" on a single cell (where you can also edit the type)
- creating a Custom Text Facet on a column, and inserting `type(value)` into the Expression field. This will generate the data type in the preview, and you can facet by data type if you press OK.

The data types supported are:

- string (one or more text characters)
- number (one or more characters of numbers only)
- boolean (values of "true" or "false")
- date (ISO-8601-compliant extended format with time in UTC: YYYY-MM-DDTHH:MM:SSZ)

OpenRefine recognises two further data types as a result of its own processes:

- error
- null

    b. Transforming -

This section of ways to improve data is organised by their appearance in the menu options in OpenRefine. You can:

- change the order of rows or columns
- edit cell contents within a particular column
- transform rows into columns, and columns into rows
- split or join columns
- add new columns based on existing data, with fetching new information, or through reconciliation
- convert your rows of data into multi-row records.
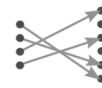
Main features

Faceting

Drill through large datasets using facets and apply operations on filtered views of your dataset.

Clustering

Fix inconsistencies by merging similar values thanks to powerful heuristics.

Reconciliation

Match your dataset to external databases via reconciliation services.

Infinite undo/redo

Rewind to any previous state of your dataset and replay your operation history on a new version of it.

Privacy

Your data is cleaned on your machine, not in some dubious data laundering cloud.

Wikibase

Contribute to Wikidata, the free knowledge base anyone can edit, and other Wikibase instances.

And much more to discover in our documentation.

OpenRefine Icelandic women artists   Permalink

Facet / Filter | Undo / Redo 29 / 29

12 rows

Show as: **rows** records   Show: 5 10 25 **50** rows

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?
Watch these screencasts

		All	Artist	Lifespan	Profession
☆	🖓	1.	Anna Jóelsdóttir	born 1947	contemporary artist
			☑☑ Hübschmannov'a, Anna (11)		
			☑☑ Macková, Anna (11)		
			☑☑ Neborová, Anna (11)		
			☑☑ Tweeddale, Anna (11)		
			☑☑ Anders, Anna (11)		
			☑☑ Beeck, Anna (11)		
			☑☑ Heindl, Anna (11)		
			☑☑ Jermolaewa, Anna (11)		
			☑☑ Waser, Anna (11)		
			☑☑ Matoušková, Anna (11)		
			☑☑ Create new item		
☆	🖓	2.	Fridriksdottir, Gabriela Choose new match	born 1971	painter, sculptor
☆	🖓	3.	Helgadóttir, GerÐur Choose new match	1928–1975	sculptor, stained-glass artist
☆	🖓	4.	Gunnfríður Jónsdóttir	1889–1968	sculptor
			☑☑ Jónsdóttir, Jóní (19)		
			☑☑ Create new item		
☆	🖓	5.	Sveinsdóttir, Júlíana Choose new match		
☆	🖓	6.	Sigurdardóttir, Katrín Choose new match		
☆	🖓	7.	Kristín Jónsdóttir		
			☑☑ Jónsdóttir, Jóní (19)		
			☑☑ Create new item		
☆	🖓	8.	Matthiasdottir, Louisa Choose new match		
☆	🖓	9.	Margret the Adroit		

[edit] Match this Cell | Match All Identical Cells | Cancel

Jónsdóttir, Jóní
(Icelandic artist, born 1972)

55 rows

Show as: **rows** records   Show: 5 10 25 **50** rows   Sort ▾

	Remove sort	
	Reorder rows permanently	
By Author	▶	Sort...
By Publication year	▶	Reverse
		Remove sort

			All	▾ Author	▾ Title			▾ Country
☆	👎	24.	A. S. Byatt	Possession			United Kingdom
☆	👎	39.	Alan Hollinghurst	The Line of Beauty			
☆	👎	18.	Anita Brookner	Hotel du Lac			
☆	👎	53.	Anna Burns	Milkman	2018	Novel	
☆	👎	42.	Anne Enright	The Gathering	2007	Novel	
☆	👎	43.	Aravind Adiga	The White Tiger	2008	Novel	India
☆	👎	32.	Arundhati Roy	The God of Small Things	1997	Novel	India
☆	👎	27.	Barry Unsworth	Sacred Hunger	1992	Historical novel	United Kingdom
☆	👎	25.	Ben Okri	The Famished Road	1991	Magic realism	Nigeria
☆	👎	55.	Bernardine Evaristo	Girl, Woman, Other	2019	Experimental novel	United Kingdom
☆	👎	2.	Bernice Rubens	The Elected Member	1970	Novel	United Kingdom
☆	👎	10.	David Storey	Saville	1976	Novel	United Kingdom
☆	👎	38.	DBC Pierre	Vernon God Little	2003	Black comedy	Australia
☆	👎	48.	Eleanor Catton	The Luminaries	2013	Historical novel	New Zealand
☆	👎	52.	George Saunders	Lincoln in the Bardo	2017	Historical/experimental novel	United States of America
☆	👎	31.	Graham Swift	Last Orders	1996	Novel	United Kingdom
☆	👎	44.	Hilary Mantel	Wolf Hall	2009	Historical novel	United Kingdom
☆	👎	47.	Hilary Mantel	Bring Up the Bodies	2012	Historical novel	United Kingdom
☆	👎	45.	Howard Jacobson	The Finkler Question	2010	Comic novel	United Kingdom
☆	👎	33.	Ian McEwan	Amsterdam	1998	Novel	United Kingdom
☆	👎	12.	Iris Murdoch	The Sea, the Sea	1978	Philosophical novel	Ireland
☆	👎	3.	J. G. Farrell	Troubles	1970	Novel	United Kingdom
☆	👎	6.	J. G. Farrell	The Siege of Krishnapur	1973	Novel	United Kingdom
☆	👎	17.	J. M. Coetzee	Life & Times of Michael K	1983	Novel	South Africa

alteryx

3. Alterix - Designer Cloud

No-code/Low-code

Useable by anyone regardless of technical skill level.

Enterprise Grade

Unlimited scalability, built-in governance, best-in-class security, and available on multiple clouds.

Faster Insights

Reduce time to value with an intuitive and collaborative data preparation solution.

Find Tool

DateTime

RegEx

Text To Columns

Transform

Arrange

Count Rows

Cross Tab

Summarize

Transpose

Weighted Average

My First Workflow

100%

Customers

SUMMARIZE (12)

Select Columns to Add Actions*   + Add Action

	Column Name	Type	
	CUSTOMERID	String	
	LOCATIONID	String	
	CUSTOMERSEGM...	String	
	CURRENTCUSTO...	String	
	FIRSTNAME	String	
	LASTNAME	String	

SAMPLE RESULTS ⓘ / Summarize (12) / Output   68 Rows   3 of 3 Columns ⌄

	Row ID	CUSTOMERSEGM...	CITY	Count
Messages	1	Corporate	DENVER	390
Input	2	Reccurring	DENVER	294
Output	3	New	AURORA	137
	4	New	DENVER	256
	5	Corporate	AURORA	222

Find Tool

DateTime

RegEx

Text To Columns

Transform

Arrange

Count Rows

Cross Tab

Summarize

Transpose

Weighted Average

My First Workflow

Customers

Count - Descending

SORT (13)

Sort by Column* ⓘ

Column Name	Order	
Count	Descending ⌄	
	Ascending ⌄	

Rename

Change Data Type

Remove

Filter

Sort

Clean Up Data

Filter Count   Clear

Operator

= (Equals)

Value

Enter value...

Apply

SAMPLE RESULTS ⓘ / Sort (13) / Output

	Row ID	CUSTOMERSEGM...	CI
Messages	1	Corporate	DENVE
Input	2	Reccurring	DENVE
Output	3	New	DENVE
	4	Rental	DENVE
	5	Corporate	AUROR

File ▾   View ▾

Find Tool

JOIN (10) ⓘ

In/Out
Input Data
Output Data
Text Input

Preparation
Data Cleansing
Filter
Formula

Sales Analysis between Different SKUs

Data Input   Data Prep and Blend   Data Output

M_CUSTO MERS – 12

M_CUSTO M_TRANSA CTIONS – 10

61%

Output Anchor Diagrams

L   J   R

Join by Specific Columns*

Left Columns   Right Columns

CUSTOMERID   CUSTOMER_ID

Select   Select

SAMPLE RESULTS ⓘ / Join (10) / Join   2,678 Rows   11 of 11 Columns ▾

Row ID	LOCATIONID	CUSTOMERSEGM...	CURRENTCUSTO...	CITY
	100	13%	Corporate 35%	No 84%	DENVER 44%
	105	13%	Reccurring 23%	Yes 16%	AURORA 23%
	106	12%	New 22%		ARVADA 10%
	101	12%	Rental 20%		BROOMFIELD 6%
1	100		Corporate	No	DENVER
2	100		Corporate	No	DENVER
3	100		Reccurring	No	DENVER
4	100		Reccurring	No	DENVER

Messages
Left
Right
Left
Join
Right

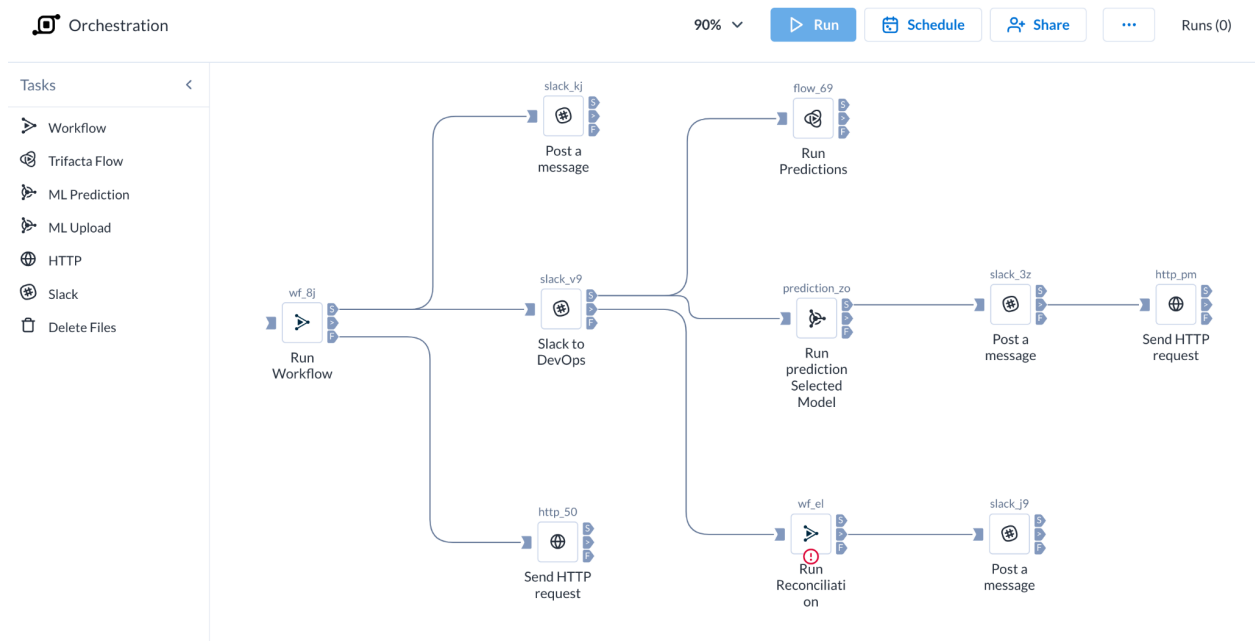SUGGESTIONS FOR [CUSTOMERSEGMENT]

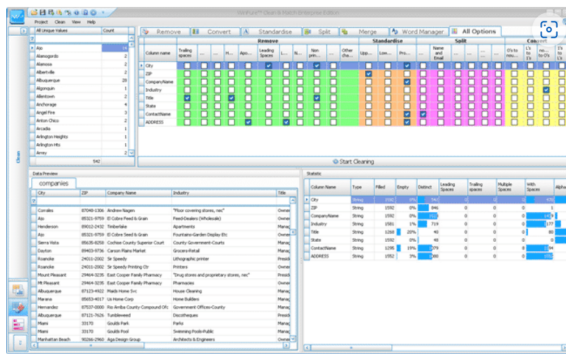SELECT
Rename [CUSTOMERSEGMENT] to [CUSTOMERSEGMENT_new]
Rename [CUSTOMERSEGMENT] to [customersegment]
Remove [CUSTOMERSEGMENT]

SORT

Overview   Library   Connections   Data   Schedules   Jobs

Orchestration   90% ▾   ▷ Run   Schedule   Share   ···   Runs (0)

Tasks
Workflow
Trifacta Flow
ML Prediction
ML Upload
HTTP
Slack
Delete Files

slack_kj
Post a message

flow_69
Run Predictions

wf_8j
Run Workflow

slack_v9
Slack to DevOps

prediction_zo
Run prediction Selected Model

slack_3z
Post a message

http_pm
Send HTTP request

http_50
Send HTTP request

wf_el
Run Reconciliati on

slack_j9
Post a message

WinPure™

4.

DATA CLEANSING MADE EASY

Featuring the revolutionary WinPure CleanMatrix™ to provide a simple yet sophisticated method of applying a whole host of data cleaning processes onto your data. Containing a full set of tools to help clean, complete, correct, standardize and transform your data, the CleanMatrix™ is divided into 7 sections, each section focusing on a specific data cleaning operation. All settings can then be saved and used on other similar data sets thus saving an incredible amount of time.

One-click Cleaning

Clean & Match features a one-click data cleaning mode which will process all the clean options across multiple columns at the same time.

Instantly Fixes Data Quality Issues

Included is a powerful data profiling tool which will scan each data list and provide over 30 different statistics ranging from % filled/empty cells to most common values & counts. It also features red & amber colouring to highlight potential data quality issues such as leading/trailing spaces, dots, hyphens, etc all of which can be fixed from a single click to help clean and correct your data, and to prepare it better for data matching.

We are on

Write a message

> Its comprehensive and powerful features really does set this aside from other data cleansing tools

WHY CHOOSE WINPURE™ CLEAN & MATCH

More Accurate Data

Make Better Decisions with Cleaner, more Accurate and Duplicate Free Data.

Easy To Use

Specially designed for anyone to use, without any special training required

Security

Locally installed software so all processing is done on premise and not online

5. CleanLab
    a. Github free data cleaning tool (Images)

cleanlab/cleanlab: The standard data-centric AI package for data quality and machine learning with messy, real-world data and labels. (github.com)

Top 7 list I found: The 7 Best Data Cleaning Tools for 2023 [Pros and Cons] (careerfoundry.com)