xlmatch

Introduction

xlmatch is a tool for populating Excel documents by contents. You can think of it as a replacement for built-in INDEX/MATCH formulas. It has the following advantages over INDEX/MATCH:

1. Scriptability. This means that you can use it in script files (e.g. batch files on Windows or shell scripts on Unix-like systems).
2. Speed. It really is much faster than INDEX/MATCH.

Usage

```
xlmatch [-h] [-o [OUTPUT]]
[--dest-match [DEST_MATCH]] [--source-match [SOURCE_MATCH]]
[--dest-column [DEST_COLUMN]] [--source-column [SOURCE_COLUMN]]
[--dest-min-row [DEST_MIN_ROW]] [--source-min-row [SOURCE_MIN_ROW]]
[--dest-max-row [DEST_MAX_ROW]] [--source-max-row [SOURCE_MAX_ROW]]
[-n] [-i] [-c [COLOR_HIGHLIGHT]]
dest source
```

positional arguments:

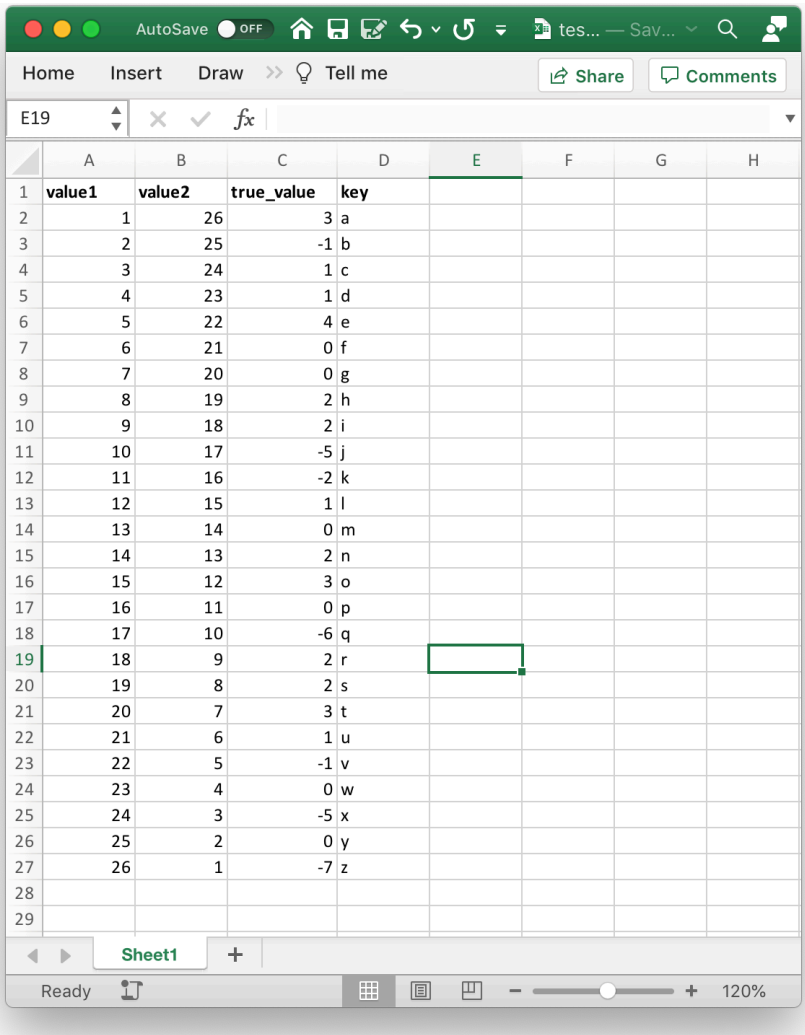
dest	destination document
source	source document

optional arguments:

-h, --help	show help message and exit
-o [OUTPUT], --output [OUTPUT]	output document (default: replace destination file)
--dest-match [DEST_MATCH]	column in the destination document used to match the content (default: B)
--source-match [SOURCE_MATCH]	column in the source document used to match the content (default: W)
--dest-column [DEST_COLUMN]	column in the destination document which will be populated (default: G)
--source-column [SOURCE_COLUMN]	column in the source document used as the source of data (default: AE)
--dest-min-row [DEST_MIN_ROW]	min row in the destination document (default: 2)
--source-min-row [SOURCE_MIN_ROW]	min row in the source document (default: 2)
--dest-max-row [DEST_MAX_ROW]	max row in the destination document (default: actual max row)
--source-max-row [SOURCE_MAX_ROW]	max row in the source document (default: actual max row)
-n, --no-backup	do not backup the destination document
-i, --ignore-case	ignore case and trailing/preceding spaces when matching
-c [COLOR_HIGHLIGHT], --color-highlight [COLOR_HIGHLIGHT]	set the background color of changed cells to the specified color (default: FFFF00)

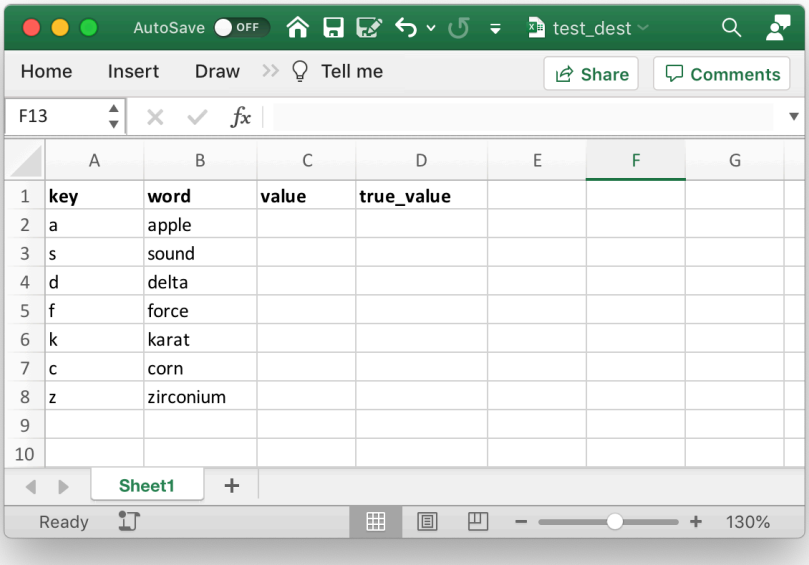
Example

Assume we have an Excel file named **`test_source.xlsx`** that contains Very Important Data. It is our **source file**.



	A	B	C	D
1	value1	value2	true_value	key
2	1	26	3	a
3	2	25	-1	b
4	3	24	1	c
5	4	23	1	d
6	5	22	4	e
7	6	21	0	f
8	7	20	0	g
9	8	19	2	h
10	9	18	2	i
11	10	17	-5	j
12	11	16	-2	k
13	12	15	1	l
14	13	14	0	m
15	14	13	2	n
16	15	12	3	o
17	16	11	0	p
18	17	10	-6	q
19	18	9	2	r
20	19	8	2	s
21	20	7	3	t
22	21	6	1	u
23	22	5	-1	v
24	23	4	0	w
25	24	3	-5	x
26	25	2	0	y
27	26	1	-7	z

Our goal is to take this data and copy it into the following Excel document named **`test_dest.xlsx`**, which is our **destination file**, based on **key**:

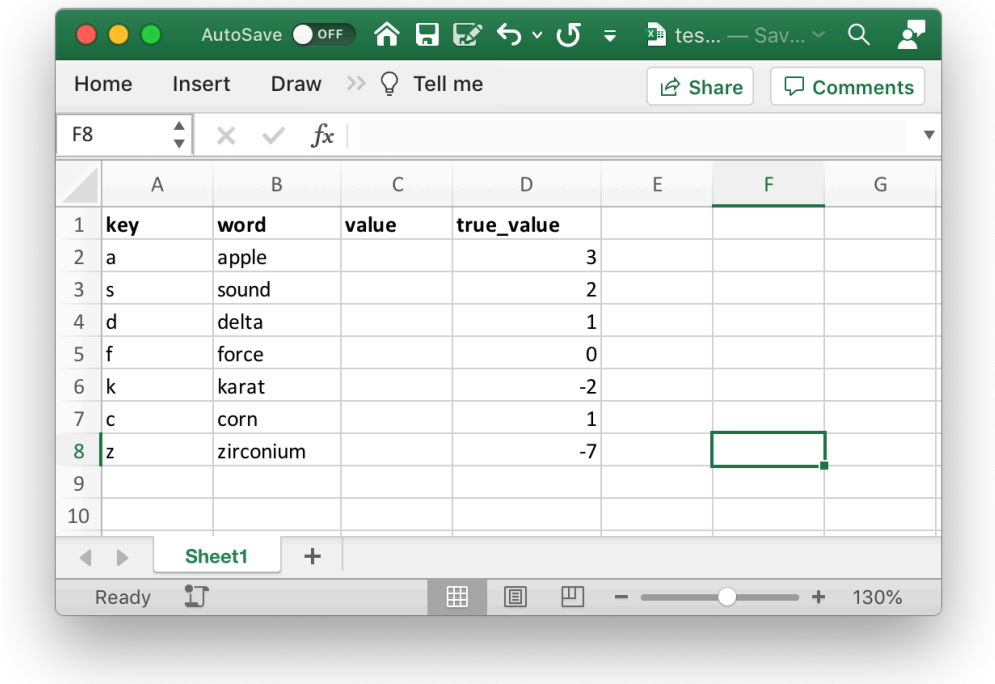


	A	B	C	D
1	key	word	value	true_value
2	a	apple		
3	s	sound		
4	d	delta		
5	f	force		
6	k	karat		
7	c	corn		
8	z	zirconium		
9				
10				

As you can see, **key** is located in column A in the destination file and in column D in the source file. To copy data from column C in the source file to column D in the destination file, we can use the following command:

```
xlmatch test_dest.xlsx test_source.xlsx --source-match A --dest-match D \
--source-column C --dest-column D --source-min-row 2 --dest-min-row 2
```

This gives us the following result:

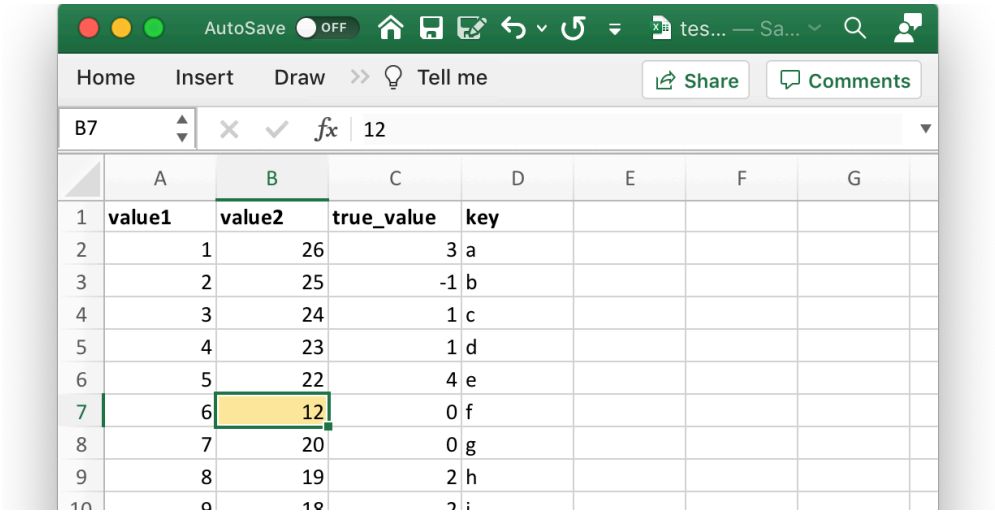


	A	B	C	D	E	F	G
1	key	word	value	true_value			
2	a	apple		3			
3	s	sound		2			
4	d	delta		1			
5	f	force		0			
6	k	karat		-2			
7	c	corn		1			
8	z	zirconium		-7			
9							
10							

Similarly, to copy data from column B in the source file to column C in the destination file, use the following command:

```
xlmatch test_dest.xlsx test_source.xlsx --source-match A --dest-match D \
--source-column B --dest-column C --source-min-row 2 --dest-min-row 2
```

Now let's assume that the values in column B associated with keys in the source file have changed:

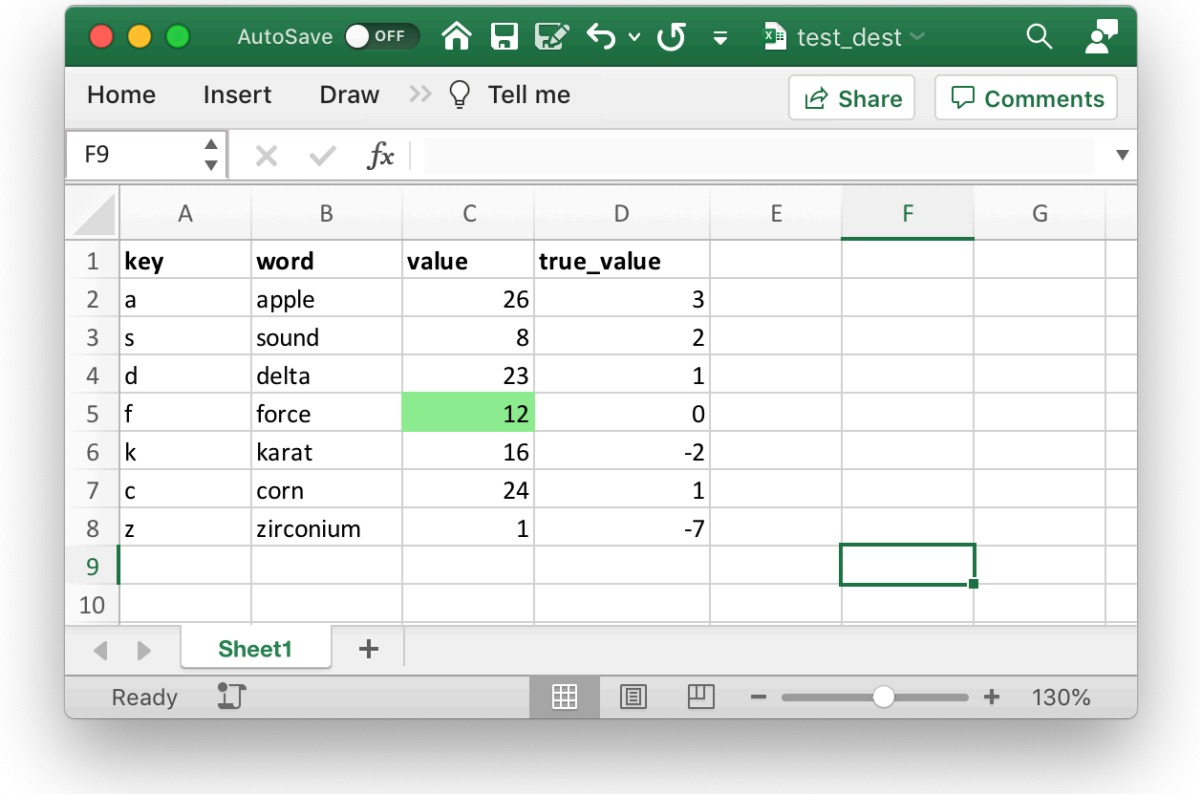


	A	B	C	D	E	F	G
1	value1	value2	true_value	key			
2	1	26	3	a			
3	2	25	-1	b			
4	3	24	1	c			
5	4	23	1	d			
6	5	22	4	e			
7	6	12	0	f			
8	7	20	0	g			
9	8	19	2	h			

We want to update our destination file but we also what to know exactly what has changed. In this case we can simply rerun the previous command adding the **`-c`** parameter to it:

```
xlmatch test_dest.xlsx test_source.xlsx --source-match A --dest-match D \
--source-column B --dest-column C --source-min-row 2 --dest-min-row 2 -c
```

The program will automatically change only those values that do not already match values in the destination document and highlight them:



	A	B	C	D	E	F	G
1	**key**	**word**	**value**	**true_value**			
2	a	apple	26	3			
3	s	sound	8	2			
4	d	delta	23	1			
5	f	force	12	0			
6	k	karat	16	-2			
7	c	corn	24	1			
8	z	zirconium	1	-7			
9							
10							

The default highlight color is yellow (**`FFFF00`**) but you can change it with the **`-c`** parameter, e.g. **`-c 00ffff`** will produce cyan highlights.

Important considerations

The tool works only with Excel 2003 and later files (**.xslx**), don't use it with .**xls** files.

Don't mix the destination and the source files, it can lead to data loss.

If the **`-n`** parameter is not specified, the program will save the destination file as **`<dest_file_name>_old.xlsx`**.

fuzzymatch

Introduction

fuzzymatch is similar to xlmatch, but uses fuzzy string matching instead of literal matching. It works much faster and produces better results than fuzzy matching add-ons available for Excel.

Usage

```
fuzzymatch [-h] [-o [OUTPUT]]
[--dest-match [DEST_MATCH]] [--source-match [SOURCE_MATCH]]
[--dest-column [DEST_COLUMN]] [--source-column [SOURCE_COLUMN]]
[--dest-min-row [DEST_MIN_ROW]] [--source-min-row [SOURCE_MIN_ROW]]
[--dest-max-row [DEST_MAX_ROW]] [--source-max-row [SOURCE_MAX_ROW]]
[-n] [-t [THRESHOLD]] [-w]
dest source
```

positional arguments:

dest	destination document
source	source document

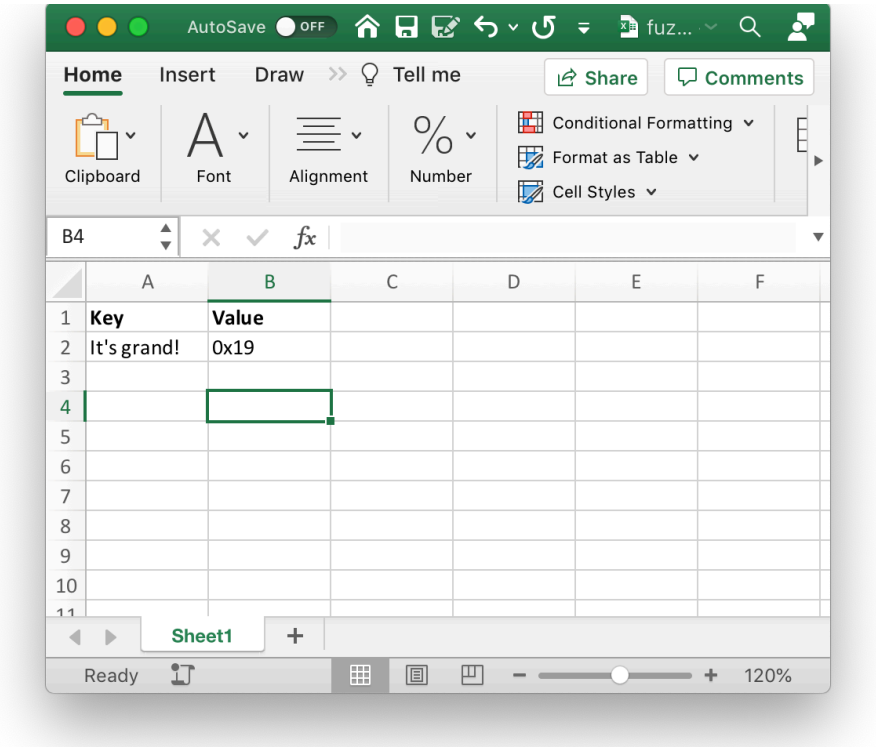
optional arguments:

-h, --help	show this help message and exit
-o [OUTPUT], --output [OUTPUT]	output document (default: replace destination file)
--dest-match [DEST_MATCH]	column in the destination document used to match the content (default: B)
--source-match [SOURCE_MATCH]	column in the source document used to match the content (default: W)
--dest-column [DEST_COLUMN]	column in the destination document which will be populated (default: G)
--source-column [SOURCE_COLUMN]	column in the source document used as the source of data (default: AE)
--dest-min-row [DEST_MIN_ROW]	min row in the destination document (default: 2)
--source-min-row [SOURCE_MIN_ROW]	min row in the source document (default: 2)
--dest-max-row [DEST_MAX_ROW]	max row in the destination document (default: actual max row)
--source-max-row [SOURCE_MAX_ROW]	max row in the source document (default: actual max row)
-n, --no-backup	do not backup the destination document
-t [THRESHOLD], --threshold [THRESHOLD]	minimum score that will be considered a match (default: 90)
-w, --weighted	use weighted ratio instead of simple ratio for calculating scores

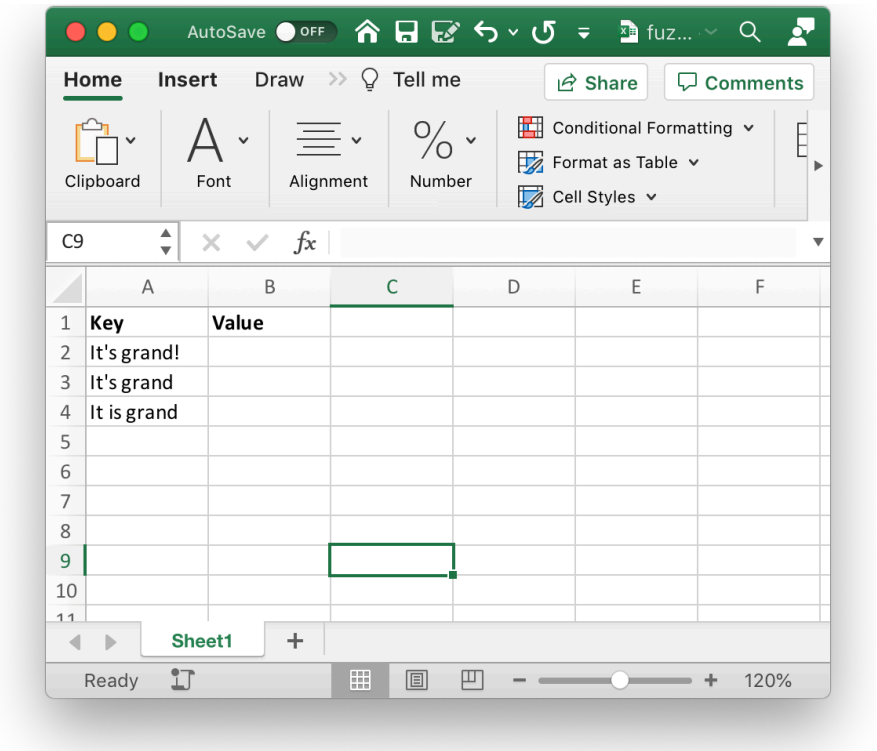
The **-w** parameter slows down the program significantly and its use is not recommended except for certain edge cases.

Example

Assume we have an Excel file named **`fuzzy_source.xlsx`** that contains Very Important Data. It is our **source file**.



Our goal is to take this data and copy it into the following Excel document named **`fuzzy_dest.xlsx`**, which is our **destination file**, based on **key**:

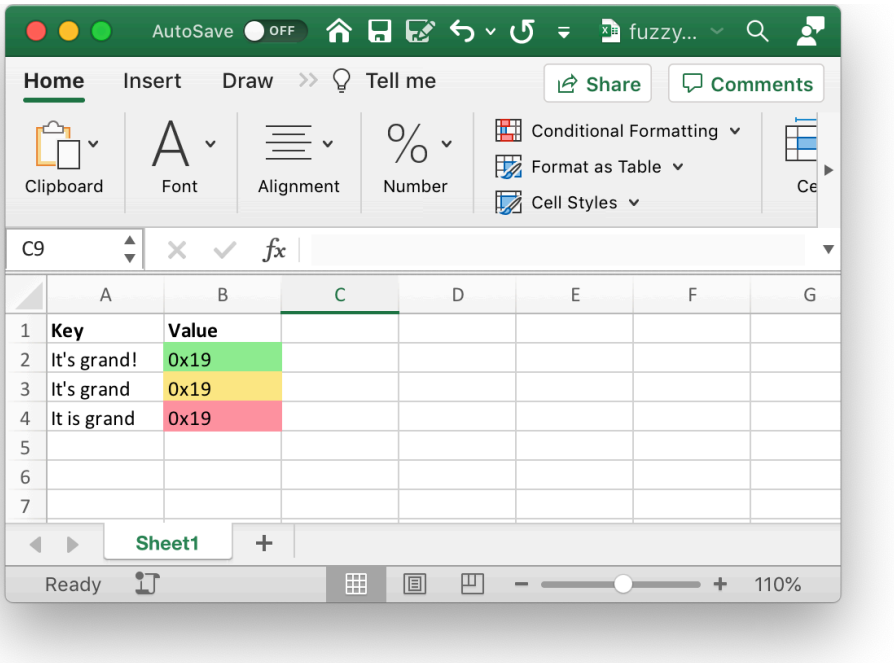


As you can see, some keys in the destination file do not match the key in the source file exactly, but we still want to copy the data if keys are similar enough.

To achieve this goal, we can use the following command:

```
fuzzymatch fuzzy_dest.xlsx fuzzy_source.xlsx \
--dest-match A --source-match A \
--dest-column B --source-column B --dest-min-row 2 --source-min-row 2
```

This gives the following result:



If the key is a literal match, the populated cell will be highlighted in green. If the key is a 100% fuzzy match, the populated cell will be highlighted in yellow. If the key is a fuzzy match with the score of less than 100, the populated cell will be highlighted in red.