

SPARK JOY IN YOUR DATA

A Marie Kondo–Inspired Guide to Data Cleaning

Dr. Jessica Breen | 4.1.25 | American University

MARIE KONDO'S 6 RULES OF TIDYING

- Commit Yourself to Tidying Up
- Imagine Your Ideal Lifestyle
- Finish Discarding First
- Tidy by Category, Not by Location
- Everything Has a Home
- Ask Yourself if It Sparks Joy

0. BACKUP YOUR DATA

Before you spark joy, spark caution!

1. COMMIT YOURSELF TO TIDYING UP

Embrace a positive mindset.

2. IMAGINE YOUR IDEAL LIFESTYLE

Visualize the end state of your data clean, organized, and ready for analysis.

3. FINISH DISCARDING FIRST

VISUAL INSPECTION:

IDENTIFY ‘CLUTTERED CORNERS’

Just like scanning a cluttered room, do a quick scan for obvious issues.

CATEGORIZE:

DETERMINE WHAT ‘SPARKS JOY’

Decide which parts of the data are necessary.

THANK AND RELEASE:

REMOVING DUPLICATES

Eliminate duplicates and irrelevant entries.

4. TIDY BY CATEGORY, NOT BY LOCATION

FOLDING NEATLY:

HANDLING MISSING DATA AND INCONSISTENCIES

Make entries uniform and tidy, akin to folding clothes neatly.

LOCATE & REPLACE MISSING DATA

Go to Home > Find & Select > Replace

EXTRA SPACES & HIDDEN CHARACTERS

```
=TRIM(text) removes extra spaces  
=CLEAN(text) removes non-printable characters  
=TRIM(CLEAN(text)) combines the two!
```

DATA TYPES

```
=VALUE(text) converts text to number  
=TEXT(value, format_text) converts numbers to  
formatted text
```

TRANSFORMING TEXT

```
=LOWER(text) transforms text to all lowercase  
=UPPER(text) transforms text to all uppercase  
=PROPER(text) capitalizes 1st letter of each  
word
```

ROUNDING NUMERIC DATA

```
=ROUND(number, num_digits) rounds to nearest  
=ROUNDUP(number, num_digits) rounds up  
=ROUNDDOWN(number, num_digits) rounds down
```


KONMARI YOUR COLUMNS:

SPLITTING & COMBINING

Give each column a proper place.

MERGING & SPLITTING DATA

To split data, go to Data → Text to Columns.

To merge using Concatenation:

```
=CONCAT(A2, " ", B2)  
=A2 & " " & B2
```

ADVANCED DATA SPLITTING

```
=LEFT(text, [num_chars])  
=RIGHT(text, [num_chars])  
=MID(text, start_num, num_chars)
```

REALLY ADVANCED DATA SPLITTING

```
=FIND(find_text, within_text, [start_num])  
=SEARCH(find_text, within_text, [start_num])  
=LEN(text)
```

TIDY CATEGORIES:

ENSURING CONSISTENT AND CORRECT VALUES

Eliminate confusion with standardized labels and validated entries.

USING A REFERENCE TABLE

```
=VLOOKUP(lookup_value, table_array,  
col_index_num, [range_lookup])  
=XLOOKUP(lookup_value, lookup_array,  
return_array, [if_not_found], [match_mode],  
[search_mode])
```

5. EVERYTHING HAS A HOME

Keep “like items” together, properly labeled.

6. ASK YOURSELF IF IT SPARKS JOY

STEP BACK AND APPRECIATE:

SPOT CHECKING WITH PIVOT TABLES AND CHARTS

Confirm the newly organized dataset truly “sparks joy.”

STORE WITH GRATITUDE:

ORGANIZING AND DOCUMENTING

Give your final dataset the place it deserves, just like a neatly organized drawer.

PUTTING IT INTO PRACTICE

Let's apply the concepts we've learned to a real dataset.

GREET OUR DATASET

ENVISION OUR IDEAL DATASET

What do we want this dataset to look like?

FINISH DISCARDING FIRST

QUICK VISUAL INSPECTION

Just like scanning a cluttered room, spot big issues first.

HIDING/REMOVING UNNECESSARY COLUMNS

Decide which columns “spark joy” and which don’t.

REMOVING DUPLICATES

“Thank and release” duplicate entries.

TIDY BY CATEGORY

HANDLING MISSING DATA AND INCONSISTENCIES

Achieve uniform, tidy entries.

SPLITTING & COMBINING COLUMNS

Give every column a proper place, just like folding items neatly in a drawer.

TIDY CATEGORIES WITH DATA VALIDATION

Prevent future clutter and keep data consistent.

CHECKING OUR PROGRESS

Verify that the dataset is approaching our “ideal state.”

WRAP UP

Maintaining Joy and Good Data Habits

Q&A