Links and Lists Extraction Checklist

A printable checklist for collecting links, list items and dropdown options from web pages, with a short CSS selector cheat sheet. Pairs with the CSV templates in this pack.

Links

- [ ] Decide which links you need: all links, only links in one section, or only links to files (PDF, CSV).
- [ ] Keep both the link text and the full URL (absolute, starting with https://).
- [ ] Remove duplicates and tracking parameters you do not need.
- [ ] Mark internal versus external links in a separate column.
- [ ] Spot check 5 links by opening them.

Lists and cards

- [ ] Pick one repeating item (a card or list row) and list the fields inside it.
- [ ] Check that every item has the same fields; leave blanks instead of shifting columns.
- [ ] Scroll or load more until the item count matches the page.
- [ ] Keep the item position (1, 2, 3) so you can restore page order.

Dropdowns and selects

- [ ] Capture both the visible label and the option value.
- [ ] Note which options are disabled or placeholders like "Select one".
- [ ] If options depend on another dropdown, record the parent choice in its own column.

Stay on the right side of the rules

- [ ] Only collect what you are allowed to use under the site terms.
- [ ] Email addresses are personal data in many places. Do not collect them for cold email.
- [ ] Go slowly and avoid repeated bulk downloads from the same site.

CSS selector cheat sheet

Selector	Matches	Example
a	every link	a
.class	elements with a class	.product-card
#id	one element by id	#results
parent child	child anywhere inside parent	#results a
parent > child	direct child only	ul > li
[attr$=".pdf"]	attribute ends with text	a[href$=".pdf"]
[attr*="text"]	attribute contains text	a[href*="/product/"]
:nth-child(n)	item by position	tr:nth-child(2)
select option	all dropdown options	#country option