SPCP's Excel Datasheet Cleaning Process

Step 1:

The first recommended step is to ensure that the columns and rows are formatted so that all the text is visible.

Before:

Visible minori: South Asian													
	Newfoundland	Prince Edv	Nova Scot	New Brun	Quebec 2(	Ontario 2(	Manitoba	Saskatche	Alberta 2(	British Col	Yukon 001	Northwes	Nunavut 00101 (8.7%)
Total - Age groups of the population in private households - 25% sample data	2645	920	7905	2535	90330	1150410	42060	29960	230930	365705	495	615	115
0 to 14 years	480	210	1440	535	19685	246445	9660	7920	55935	70985	70	130	10
0 to 4 years	140	70	545	185	6510	77375	3955	3020	19690	22460	0	20	10
5 to 9 years	165	60	460	200	6925	86195	3120	2570	19910	24355	35	70	0
10 to 14 years	170	70	435	145	6250	82870	2585	2325	16340	24165	30	45	10
15 to 64 years	2005	660	5750	1720	62950	792190	29655	20840	156650	248840	410	450	95
15 to 19 years	175	75	490	145	6110	80530	2625	1740	13550	25170	55	30	10
20 to 24 years	265	85	700	160	7665	85735	3360	2130	13815	27195	75	50	0
25 to 29 years	430	85	1300	315	8330	89315	4770	4500	21365	28155	105	90	10
30 to 34 years	285	115	930	230	7700	94745	5085	3805	24985	30205	35	50	15
35 to 39 years	190	90	535	180	7330	92640	4195	2455	22865	30220	30	45	15
40 to 44 years	175	50	490	205	6140	87605	3135	2135	18465	27250	35	60	0
45 to 49 years	155	60	415	170	5775	80740	2345	1805	13750	23995	35	45	10
50 to 54 years	120	20	375	140	5350	70565	1600	1105	10745	20375	20	25	10
55 to 59 years	95	50	310	75	4635	59305	1370	610	8970	18315	15	15	10
60 to 64 years	120	30	210	100	3915	51010	1175	560	8140	17970	10	30	10
65 years and over	155	55	715	280	7700	111775	2745	1200	18340	45875	20	30	0
65 to 69 years	80	10	195	105	2800	42155	1035	495	7150	16255	10	10	0
70 to 74 years	25	25	190	95	2170	29475	740	320	5095	12415	0	10	0
75 to 79 years	25	20	160	60	1345	20600	485	190	3230	8540	0	10	0
80 to 84 years	20	0	95	15	870	12105	315	120	1640	5245	0	0	0
85 years and over	0	0	70	0	515	7445	175	80	1225	3420	0	0	0
85 to 89 years	10	0	45	10	395	5455	115	50	870	2320	0	10	0
90 to 94 years	0	0	20	0	105	1585	35	25	270	800	0	0	0
95 to 99 years	0	0	10	0	15	335	0	0	60	235	0	0	0
100 years and over	0	0	0	0	10	70	20	0	20	75	0	0	0

After:

	A	B	C	D	E	F
1	Visible minority: South Asian	Newfoundland and Labrador	Prince Edward Island	Nova Scotia	New Brunswick	Quebec
2	Population By Age Range	2645	920	7905	2535	90330
3	0 to 14 years	480	210	1440	535	19685
4	0 to 4 years	140	70	545	185	6510
5	5 to 9 years	165	60	460	200	6925
6	10 to 14 years	170	70	435	145	6250
7	15 to 64 years	2005	660	5750	1720	62950
8	15 to 19 years	175	75	490	145	6110
9	20 to 24 years	265	85	700	160	7665
10	25 to 29 years	430	85	1300	315	8330
11	30 to 34 years	285	115	930	230	7700
12	35 to 39 years	190	90	535	180	7330
13	40 to 44 years	175	50	490	205	6140
14	45 to 49 years	155	60	415	170	5775
15	50 to 54 years	120	20	375	140	5350
16	55 to 59 years	95	50	310	75	4635
17	60 to 64 years	120	30	210	100	3915
18	65 years and over	155	55	715	280	7700
19	65 to 69 years	80	10	195	105	2800
20	70 to 74 years	25	25	190	95	2170
21	75 to 79 years	25	20	160	60	1345
22	80 to 84 years	20	0	95	15	870
23	85 years and over	0	0	70	0	515
24	85 to 89 years	10	0	45	10	395
25	90 to 94 years	0	0	20	0	105
26	95 to 99 years	0	0	10	0	15
27	100 years and over	0	0	0	0	10

Step 2:

Usually, there are many spelling mistakes or formatting errors. For example, there are random percentages in the headers of each table. Remove those as well as try to fix any minor spelling mistakes and shorten the names of the indicators.

Before:

72 *Total - Marital status for the population aged 15 years and over in private households - 25% sample data*

After:

76 Martial Status In Percentage

Created By: Yahya Saqib (Social Researcher)

SPCP's Excel Datasheet Cleaning Process

Step 3:

There could potentially be many leading or trailing empty character spaces. To make the future programmer's life easier, delete all of those and double check to see if there are any left.

Before:

89	Separated
90	Divorced
91	Widowed

After:

82	Separated
83	Divorced
84	Widowed

Step 4:

For each table's indicator, bold all the headers of the table. This is **crucial,** as the code for the website needs bolded text for it to register properly.

Before:

29	AGE GROUP	Newfoundland and Labrador 00000 (6.8%)

After:

29	**Age Group**	**Newfoundland and Labrador**

Step 5:

Try your best to fix the capitalization of each header for consistency purposes as well. The example above is a great example of a potential issue.

Step 6:

For each new table, please leave a one-row gap. This is **crucial**, as the code registers a new table being made within the code.

Before:

29	AGE GROUP	Newfoundland and Labrador 00000 (6.8%)	Prince Edward Islan
30	Total - Age groups of the population in private households - 25% sample data		
31	0 to 4 years	140	
32	5 to 9 years	165	
33	10 to 14 years	170	
34	15 to 24 years	440	
35	25 to 44 years	1080	
36	45 to 64 years	490	
37	65 years and over	155	
38	total -	2640	
39	Total - Age groups of the population in private households - 25% sample data		
40	0 to 4 years	5.3%	
41	5 to 9 years	6.3%	
42	10 to 14 years	6.4%	
43	15 to 24 years	16.7%	
44	25 to 44 years	40.9%	
45	45 to 64 years	18.6%	
46	65 years and over	5.9%	

Created By: Yahya Saqib (Social Researcher)

SPCP's Excel Datasheet Cleaning Process

After:

	Age Group	Newfoundland and Labrador	Prince Edward Island	Nova Scotia	New Brunswick	Quebec
30	0 to 4 years	140	70	545	185	6510
31	5 to 9 years	165	60	460	200	6925
32	10 to 14 years	170	70	435	145	6250
33	15 to 24 years	440	160	1190	305	13775
34	25 to 44 years	1080	340	3255	930	29500
35	45 to 64 years	490	160	1310	485	19675
36	65 years and over	155	55	715	280	7700
37	Total	2640	915	7910	2530	90335
38						
39	**Age Group In Percentage**	**Newfoundland and Labrador**	**Prince Edward Island**	**Nova Scotia**	**New Brunswick**	**Quebec**
40	0 to 4 years	5.3%	7.7%	6.9%	7.3%	7.2%
41	5 to 9 years	6.3%	6.6%	5.8%	7.9%	7.7%
42	10 to 14 years	6.4%	7.7%	5.5%	5.7%	6.9%
43	15 to 24 years	16.7%	17.5%	15.0%	12.1%	15.2%
44	25 to 44 years	40.9%	37.2%	41.2%	36.8%	32.7%
45	45 to 64 years	18.6%	17.5%	16.6%	19.2%	21.8%
46	65 years and over	5.9%	6.0%	9.0%	11.1%	8.5%
47	Total	100.0%	100.0%	100.0%	100.0%	100.0%

These should be all the steps that are important for the new spreadsheets to work and be implemented properly in the code.

If you have any other questions, feel free to contact me (Yahya Saqib) at ysaqib@spcpeel.com for any further questions.

This document also provides links to the original spreadsheet created by Dr. Mohanty and the reformatted spreadsheet created by me. This should be the best way to understand the changes that need to be made for future spreadsheets to work.

Google Drive Link **(You must have a SPC Email to gain access)**

Created By: Yahya Saqib (Social Researcher)