

Spreadsheet (Dataset 1 & 2)

Applications	Issues & Actions	References
1. Conditional Formatting & Filtering	a. Locate empty cells	2942/09/12/2022 Yue Wo Villa 2 Rooms 2 Living Rm/Flat E Lower Floor Block 09 Tuen Mun Ferry \$8,800 431 sq.ft. (SFA) 518 sq.ft. (GFA)@817
	- found a page’s records are missing (due to some sort of loading error during web scraping)	2943/09/12/2022 St Martin Phase 1 1 Room 2 Living Rm/Flat A1 Middle Floor Tower 5 Pak Shek Kok \$13,500 379 sq.ft. (SFA) N/A
		2944/09/12/2022 The Regent 3 Rooms 2 Living Room/Flat / Higher Floor Tower 18 Tai Po Mid-Levels \$26,000 819 sq.ft. (SFA) N/A
		2945/09/12/2022 Tai Po Centre Phase 2 2 Rooms 1 Liv/Flat E Higher Floor Block 04 Tai Po Town Centre \$10,800 299 sq.ft. (SFA) 423 sq.ft. (GFA)@826
		2946
		2947
		2948
		2949
		2950
		2951
	2952	
	2953	
	2954	
	2955	
	2956	
	2957	
	2958	
	2959	
	2960	
	2961	
	2962	
	2963	
	2964	
	2965	
	2966	
	2967	
	2968	
	2969/05/12/2022 Vibe Centro 2 Rooms 2 Living Room/Flat B Lower Floor Tower 1A Kai Tak \$20,000 538 sq.ft. (SFA) N/A	
	2970/05/12/2022 Manhattan Hill Flat F Lower Floor Block 3 Mei Foo Station \$21,800 559 sq.ft. (SFA) N/A	
	2971/05/12/2022 Lei Szeious 2 Rooms 1 Living Room/Flat A Higher Floor Block 3 (Lantau/Sai Wai Ho \$23,000 556 sq.ft. (SFA) 744 sq.ft. (GFA)@831	
	2972/05/12/2022 Double Cove Phase 1 2 Rooms 2 Liv/Flat A Middle Floor Block 3 Wo Kai Sha \$17,000 497 sq.ft. (SFA) N/A	

## 2. Conditional Functions (“IF”, “IFERROR”, “FIND”, “MID”...etc)

### a. Split the information in “Premises” column into separate cells

- spot the text pattern in the “Premises” column
- use “IF”, “FIND”, “MID” to locate the information and extract the corresponding text to new columns
- new columns: “Building”, “Floor”

### b. Locate ERROR cells by filter

- fix the error

### c. Format columns for consistency

- “Date” column: **date (YYYY-MM-DD)**
- “Transaction Rent”, “SFA”, “GFA”: **numeric**

### d. Categorize the unit size (SFA/GFA) in range for better visualization later

- “Less than 400 sq.ft.” // “400 - 800 sq.ft.” // “801 - 1200 sq.ft.” // “1201 - 1500 sq.ft.” // “1501 - 2000 sq.ft.” // “More than 2000 sq.ft.”

### e. Pick few random samples to verify the cleaned data

=IF(MID(J3,FIND("Room",J3)-5,4)<>"ving",MID(J3,1,FIND("Room",J3)-3),MID(J3,1,FIND("Living Room",J3)-3))			
J	K	L	M
Premises	Floor	Building	Floor
Yue Fai Court	Flat 7 Middle Floor Block B	#VALUE!	Middle
The Regent 3 Rooms 2 Living Rooms (Garden)	Flat A Tower 7	The Regent	N/A
Wing Cheung Court 3 Rooms 2 Living Rooms	Flat C Middle Floor	Wing Cheung Court	Middle
Academic Terrace 2 Rooms 2 Living Rooms	Flat H Middle Floor Block 1	Academic Terrace	Middle
Green Lodge 3 Rooms 2 Living Rooms (Carpark, Garden)	23 Ma Fung Ling Road (Apartment)	Green Lodge	N/A
The Avenue Phase 2 1 Living Room (Balcony, Utility Platform)	Flat R Lower Floor Tower 2	The Avenue Phase 2	Lower
Nan Fung Plaza 3 Rooms 2 Living Rooms	Flat C Middle Floor Block 6	Nan Fung Plaza	Middle
The Pavila Farm The Pavila Farm 1 3 Rooms 2 Living Rooms (Balcony)	Flat B Lower Floor Tower 2A	The Pavila Farm The Pavila Farm 1	Lower
Park Central Phase 2 3 Rooms 2 Living Rooms	Flat G Higher Floor Block 01	Park Central Phase 2	Higher

## Sample spreadsheet functions:

### a. Extract the building name

- =IFERROR(MID(B2,1,FIND("-",B2)-1),B2)

### b. Categorize the floor information

- =IF(ISERROR(FIND("Lower Floor",B2))=FALSE, "Lower", IF(ISERROR(FIND("Middle Floor",B2))=FALSE, "Middle", IF(ISERROR(FIND("Upper Floor",B2))=FALSE, "Upper", "N/A")))

### c. Extract numeric parts from the size information

- =IFERROR(MID(H2,1,FIND("ft",H2)-1),"N/A")

<b>3. Manual adjustment</b>	<p><b>a. Adjust namings under “District” column</b></p> <ul style="list-style-type: none"><li>- some naming discrepancies between source data are spotted by filtering</li><li>- i.e. “Shatin” vs “Sha Tin”</li><li>- manually align the naming</li></ul> <p><b>b. Create new columns: “Unit Rate (SFA/GFA)”, “Source”</b></p> <p><b>c. Combine the two datasets into one single spreadsheet</b></p>	
-----------------------------	--	--