

# Demo session

A demo session to display the housing data scraped

```
In [29]: import os
import pandas
import datetime
import pytz

import warnings
warnings.filterwarnings("ignore")
```

```
In [30]: now = datetime.datetime.now(tz=pytz.timezone("Asia/Kuala_Lumpur"))
json_file = f"../../output/data/{now:%Y-%m-%d}.json"
```

```
In [31]: data = pandas.read_json(json_file)
# remove duplicates
cleaned_data = data[~data.duplicated(subset="id")]
cleaned_data.head(5)
```

```
Out[31]:
```

	type	id	attributes	links
0	ads	100550027	{'date': '2024-02-11 23:59:17', 'modified_ts':...	{'image_baseurl': 'https://img.rnudah.com/imag...
1	ads	101268085	{'date': '2024-02-11 23:58:49', 'modified_ts':...	{'image_baseurl': 'https://img.rnudah.com/imag...
2	ads	100795490	{'date': '2024-02-11 23:57:46', 'modified_ts':...	{'image_baseurl': 'https://img.rnudah.com/imag...
3	ads	100549243	{'date': '2024-02-11 23:57:01', 'modified_ts':...	{'image_baseurl': 'https://img.rnudah.com/imag...
4	ads	105600121	{'date': '2024-02-11 23:56:52', 'modified_ts':...	{'image_baseurl': 'https://img.rnudah.com/imag...

```
In [32]: pandas.json_normalize(cleaned_data["attributes"])
```

Out[32]:

	date	modified_ts	building_id	type_name	category_name	lis
0	2024-02-11 23:59:17	1707667157	4d2836b7-847d-49fc-a0d9-5b4a53ea0815	sell	Apartment / Condominium	100550
1	2024-02-11 23:58:49	1707667129	NaN	sell	House	101268
2	2024-02-11 23:57:46	1707667066	NaN	sell	House	100795
3	2024-02-11 23:57:01	1707667021	NaN	sell	House	100549
4	2024-02-11 23:56:52	1707667012	NaN	sell	House	105600
...	...	...	...	...	...	...
3089	2024-02-12 20:43:23	1707741803	NaN	sell	House	104974
3090	2024-02-12 20:42:56	1707741776	55a5a803-59d5-404a-89e4-5d9523f89bc4	sell	Apartment / Condominium	101655
3091	2024-02-12 20:42:51	1707741771	NaN	let	Commercial Property	105357
3092	2024-02-12 20:42:24	1707741744	NaN	let	Commercial Property	105356
3093	2024-02-12 20:41:29	1707741689	NaN	sell	House	105628



Out[33]:

	date	modified_ts	building_id	type_name	category_name	lis
<b>0</b>	2024-02-11 23:59:17	1707667157	4d2836b7-847d-49fc-a0d9-5b4a53ea0815	sell	Apartment / Condominium	100550
<b>1</b>	2024-02-11 23:58:49	1707667129	NaN	sell	House	101268
<b>2</b>	2024-02-11 23:57:46	1707667066	NaN	sell	House	100795
<b>3</b>	2024-02-11 23:57:01	1707667021	NaN	sell	House	100549
<b>4</b>	2024-02-11 23:56:52	1707667012	NaN	sell	House	105600
...	...	...	...	...	...	...
<b>3089</b>	2024-02-12 20:43:23	1707741803	NaN	sell	House	104974
<b>3090</b>	2024-02-12 20:42:56	1707741776	55a5a803-59d5-404a-89e4-5d9523f89bc4	sell	Apartment / Condominium	101655
<b>3091</b>	2024-02-12 20:42:51	1707741771	NaN	let	Commercial Property	105357
<b>3092</b>	2024-02-12 20:42:24	1707741744	NaN	let	Commercial Property	105356
<b>3093</b>	2024-02-12 20:41:29	1707741689	NaN	sell	House	105628

3094 rows × 56 columns

```
In [34]: merged_data[merged_data.duplicated(subset="list_id")]
```

```
Out[34]:
```

	date	modified_ts	building_id	type_name	category_name	list_id	subject	
--	------	-------------	-------------	-----------	---------------	---------	---------	--

0 rows × 56 columns