

\$ jupyter notebook

etl\_properties.py

X +

```
{"ts": "2026-02-21T08:09:26.347Z", "level": "INFO", "msg": "Job 31 finished: count at NativeMethodAccessorImpl.java:0, took 24.108126 ms", "context": {"call_site_short_form": "count at NativeMethodAccessorImpl.java:0", "job_id": "31", "time": "24.108126"}, "logger": "DAGScheduler"}
```

== COMPLETENESS BY COLUMN ==

area_raw	27/	29	93.10%
id	29/	29	100.00%
url	29/	29	100.00%
scraped_at	29/	29	100.00%
source_page	29/	29	100.00%
title	29/	29	100.00%
location	29/	29	100.00%
price_raw	29/	29	100.00%
object_id	29/	29	100.00%
description	29/	29	100.00%
features	29/	29	100.00%

OVERALL AVG COMPLETENESS: 99.37%

```
{"ts": "2026-02-21T08:09:27.097Z", "level": "INFO", "msg": "Removed broadcast_24_piece0 on c0c9286d8932:42021 in memory (size: 9.2 KiB, free: 434.3 MiB)", "context": {"block_id": "broadcast_24_piece0", "free_memory_size": "434.3 MiB", "host_port": "c0c9286d8932:42021", "original_memory_size": "9.2 KiB"}, "logger": "BlockManagerInfo"}
```

```
{"ts": "2026-02-21T08:09:27.105Z", "level": "INFO", "msg": "Removed broadcast_31_piece0 on c0c9286d8932:42021 in memory (size: 6.5 KiB, free: 434.3 MiB)", "context": {"block_id": "broadcast_31_piece0", "free_memory_size": "434.3 MiB", "host_port": "c0c9286d8932:42021", "original_memory_size": "6.5 KiB"}, "logger": "BlockManagerInfo"}
```

```
{"ts": "2026-02-21T08:09:27.123Z", "level": "INFO", "msg": "Removed broadcast_27_piece0 on c0c9286d8932:42021 in memory (size: 6.5 KiB, free: 434.4 MiB)", "context": {"block_id": "broadcast_27_piece0", "free_memory_size": "434.4 MiB", "host_port": "c0c9286d8932:42021", "original_memory_size": "6.5 KiB"}, "logger": "BlockManagerInfo"}
```

do\_f, logger: DAGSchedulerImpl

```
{"ts": "2026-02-21T08:09:27.740Z", "level": "INFO", "msg": "Job 34 finished: count at NativeMethodAccessorImpl.java:0, took 40.141847 ms", "context": {"call_site_short_form": "count at NativeMethodAccessorImpl.java:0", "job_id": "34", "time": "40.141847"}, "logger": "DAGScheduler"}
```

CLEAN COUNT: 27

```
{"ts": "2026-02-21T08:09:27.998Z", "level": "INFO", "msg": "Code generated in 42.551657 ms", "context": {"total_time": "42.551657"}, "logger": "CodeGenerator"}
```

```
{"ts": "2026-02-21T08:09:28.003Z", "level": "INFO", "msg": "Registering RDD 112 ($anonfun$withThreadLocalCaptured$1 at FutureTask.java:264) as input to shuffle 19", "context": {"creation_site": "$anonfun$withThreadLocalCaptured$1 at FutureTask.java:264", "rdd_id": "112", "shuffle_id": "19"}, "logger": "DAGScheduler"}
```

```
{"ts": "2026-02-21T08:09:28.356Z", "level": "INFO", "msg": "Job 36 finished: jdbc at NativeMethodAccessorImpl.java:0, took 168.36626 ms", "context": {"call_site_short_form": "jdbc at NativeMethodAccessorImpl.java:0", "job_id": "36", "time": "168.36626"}, "logger": "DAGScheduler"}
```

ETL FINISHED SUCCESSFULLY

```
{"ts": "2026-02-21T08:09:28.606Z", "level": "INFO", "msg": "Invoking stop() from shutdown hook", "logger": "SparkContext"}
```

```
{"ts": "2026-02-21T08:09:28.608Z", "level": "INFO", "msg": "SparkContext is stopping with exitCode 0 from run at Executors.java:539.", "context": {"exit_code": "0", "stop_site_short_form": "run at Executors.java:539"}, "logger": "SparkContext"}
```

```
"$ jupyter notebook
```