Script	Companion should detect …	Spark-UI / symptom to verify
taxi_skewed_join_small.py	Skewed shuffle & disabled broadcast; single-file write	One long straggler task; shuffle read skew; 1 output file
taxi_collect_driver_small.py	Excessive collect() to driver	Driver GC / memory spike; executors idle
taxi_too_many_partitions_small.py	Excessive shuffle partitions & duplicate shuffle	Two 10 000-task stages; high scheduler delay
taxi_cache_no_unpersist_small.py	Large DF cached but never unpersisted	Storage tab shows DF cached; later jobs GC/spill
taxi_python_udf_small.py	Python UDF instead of built-in expression	Task runtimes dominated by Python exec; no Whole-stage codeg
taxi_cartesian_join_small.py	Cartesian (cross) join	Plan shows ‘CartesianProduct’; huge shuffle read
taxi_many_small_files_small.py	Writes thousands of tiny files	FileOutputCommitter tasks; 5 000 small files in GCS
taxi_no_compression_small.py	Output Parquet with compression disabled	Environment shows codec=none; output size unusually large
taxi_multi_cache_small.py	Multiple large caches not released	Storage tab: several cached DFs; executor memory near cap
taxi_rdd_conversion_small.py	Unnecessary DF → RDD → DF conversion	Plan loses Tungsten/columnar; extra serialization stage