openml_benchmark_results - f1_weighted per Task