INTRODUCTION

This benchmarking software "**Benchmark.py**" is a powerful tool designed to measure and compare the performance of different scripts on specific data sets. It's built with Python and uses a variety of scripts including Python, Polars, and Peaks for benchmarking.

The software works by running each script on a specified data file for a certain number of batches. The elapsed time for each run is recorded and the average time is calculated. The results are then written to a CSV file and displayed on the screen.

The software also includes a feature to organize the output files into batch folders and benchmark folders following a specific naming convention. This makes it easy to keep track of different runs and compare results.

Before starting the benchmark process, the software checks the Outbox directory to ensure it only contains Benchmark folders that match the current case's naming convention or Batch folders. If there are other files or folders in the Outbox that do not follow this pattern, the process will not start.

This benchmarking software is an excellent tool for anyone looking to measure and improve the performance of their scripts. It provides detailed, accurate results and organizes them in a way that's easy to understand and compare. Whether you're a developer looking to optimize your code or a data scientist trying to choose the most efficient script for your data, this benchmarking software can help you make informed decisions.

MOTIVATION OF BUILDING THIS SOFTWARE

The author developed this benchmarking software to provide an effective and efficient means of comparing the performance of Polars and Peaks dataframes across different use cases and implementation environments.

The software supports a variety of implementations, including Polars running in Rust, Python and Peaks in Go, Python, C#, and Rust. This wide range of support allows users to test and compare the performance of these dataframes in the environment that is most relevant to their use case.

By automating the benchmarking process and providing detailed, easy-to-understand results, this software saves users time and effort. It eliminates the need for manual testing and comparison, which can be time-consuming and prone to error.

Furthermore, by organizing the output files into batch folders and benchmark folders following a specific naming convention, it makes it easy to keep track of different runs and compare results. This organization also ensures that the Outbox directory remains clean and manageable, even after multiple benchmark runs.

In summary, this benchmarking software is a valuable tool for anyone working with Polars and other dataframe software. It offers a streamlined, automated solution for performance testing and comparison, helping users to make informed decisions about their dataframe implementation.

WARNING: IMPROPER USE OF THIS SOFTWARE

This benchmarking software is designed for testing purposes only. It should be run on a dedicated testing machine with test data. Here are some important guidelines to follow:

1. **Do not use this software on production machines or with sensitive data.** The software performs intensive operations that can significantly impact system performance. It also does not include any features for securing or anonymizing data.
2. **Always back up your data before running the software.** The software moves and renames files in the Outbox directory. If an error occurs, it could result in data loss.
3. **Ensure the Outbox directory is correctly configured before starting a benchmark.** The software checks the Outbox directory before starting and will not run if it finds files or folders that do not match the expected pattern.
4. **Do not interrupt the software while it's running a benchmark.** Interrupting the process could lead to incomplete results and could leave your Outbox directory in an inconsistent state.
5. **Use the software responsibly.** Running benchmarks can be resource-intensive. Be mindful of other processes and users that might be affected by a benchmark run.

Failure to adhere to these guidelines could result in data loss, system slowdowns, or inaccurate benchmark results. Use this software responsibly and at your own risk.