

Data Intake Report

Name: Large Dataset Processing with Pandas, Dask, and Modin

Report date: 04 April 2024

Internship Batch: LISUM28

Version: 1.0

Data intake by: Minseok Kim

Data intake reviewer: Minseok Kim

Data storage location: <https://www.kaggle.com/datasets/skihikingkevin/pubg-match-deaths>

Tabular data details: kill_match_stats_final

Total number of observations	13440891
Total number of files	1
Total number of features	12
Base format of the file	.csv
Size of the data	2.01 GB

Note: Replicate same table with file name if you have more than one file.

Proposed Approach:

- Since it is large dataset, Dask would be the best tool.