

Assignment 1 – Exploratory data analysis

Group Members:

- Francisco Gouveia 2023186817 - franciscoggouveia@gmail.com
- Miguel Pedroso 2019218176 - mmsspmpedroso@gmail.com
- Rui Santos 2020225542 - rpsantos@student.dei.uc.pt

Dataset(s):

- Nasa:
https://www.cs.huji.ac.il/labs/parallel/workload/l_nasa_ipsc/index.html

The dataset contains three 3 months of records. The workload on the 128-node iPSC/860 consists of interactive and batch jobs for computational aerospace applications. This machine is located in the NAS systems Division at NASA.

- Atlas:
https://www.cs.huji.ac.il/labs/parallel/workload/l_inl_atlas/index.html
Log containing several months of accounting data from a large linux cluster with 1152 clusters, each with 8 nodes, for a total of 9216 processors, used to run large parallel jobs that cannot be performed on less powerful machines.

EDA tools:

For this assignment we will use python (pandas, matplotlib, seaborn) and datatab.