Homework #5

Team 2:

Julian Carvajal Rico

James Platt Standard

Roberto Enriquez Vargas

Table 1. summarizes the performance of different file formats for data handling across five matrices (A, B, C, D and, E). For matrices A, B, and C, NumPy is the fastest format for both writing (generation time) and reading (loading time), indicating high efficiency in both creating and accessing data with NumPy. While for matrices D and E the time for writing and reading them is less using the HDF5 format.

However, when reading the file with the HDF5 explorer and finding the size of each matrix inside the dataset, for all the matrix the size was less than the CSV and NumPy formats. (Detailed values in the excel file included in the zip folder)

Table 1 - Matrix Analysis

Matrix	Fastest Write Format	Fastest Read Format	Smallest File Size Format
A	NumPy Format	NumPy Format	HDF5 Size
B	NumPy Format	NumPy Format	HDF5 Size
C	NumPy Format	NumPy Format	HDF5 Size
D	HDF5 Format	HDF5 Format	HDF5 Size
E	HDF5 Format	HDF5 Format	HDF5 Size

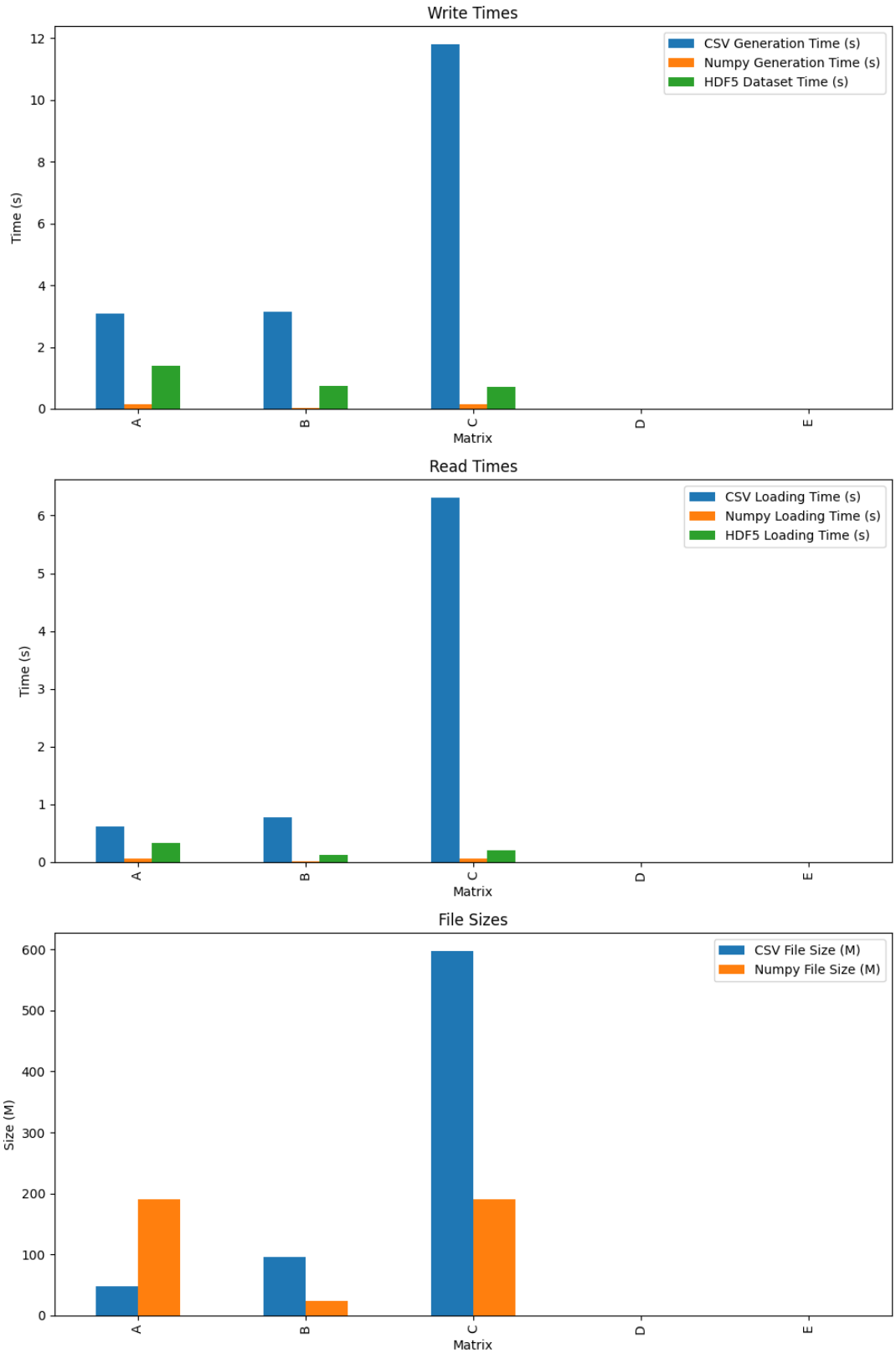

Figure 1 - Bar Plots Write Time, Read Times, and File Sizes