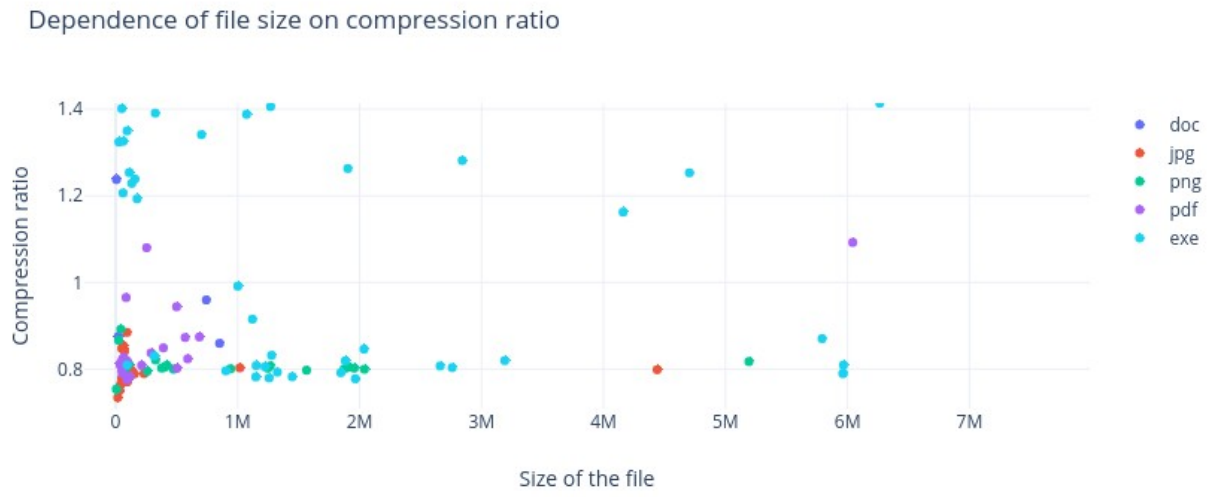
IT assignment 2

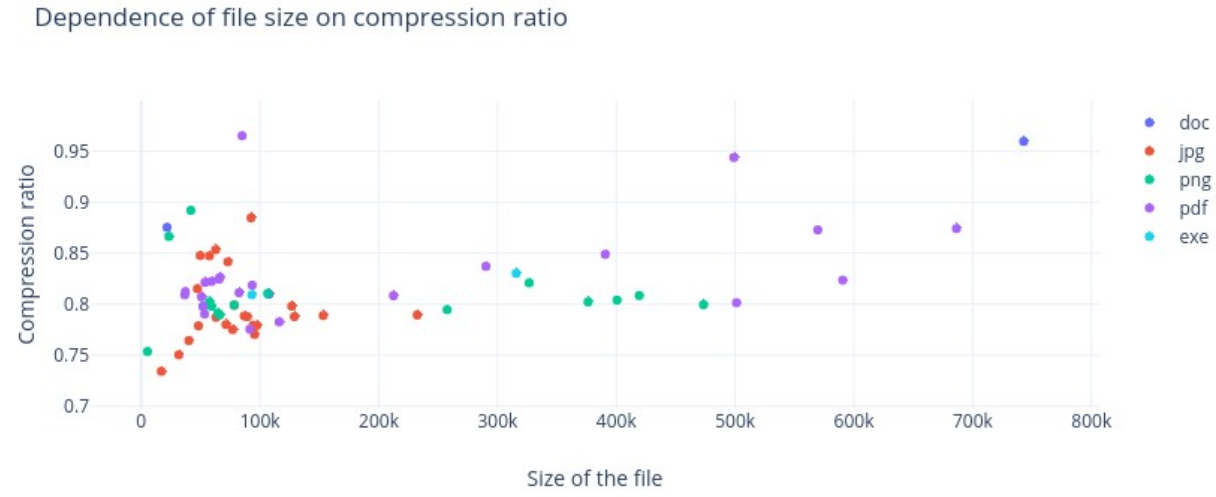
Author: Ruslan Sabirov

Group: B17-DS-02

Compression ratio table				
File type	**min**	**average ↓**	**max**	**std_dev**
doc	0.810	1.788	2.781	0.563
exe	0.778	1.143	3.171	0.468
pdf	0.756	0.885	1.579	0.170
png	0.754	0.806	0.892	0.025
jpg	0.734	0.797	0.885	0.034

** Table was sorted by descending average ratio. Minimum values for columns were colored by green, maximums – by red.*





$$\text{Compression Ratio} = \frac{\text{Uncompressed Size}}{\text{Compressed Size}}$$

The higher compression ratio is the better results were achieved by algorithm. As we can see from the table it is not useful to compress some files by LZ78 algorithm because their size would increase.

1. Does the compression ratio depends on the file type? If yes, then how?
Yes, compression ration depends on file type.
There is no need to apply LZ78 on already compressed file types (pdf, jpg, png), because it would just increase their size.
The best compression ratio was achieved on some of the exe-files, the worst – on some of the jpg-files. On the average LZ78 performs better on doc-files and worse on jpg-files (ratio of doc-files is 2,24 times higher than ratio of jpg-file).

2. Does the compression ratio depends on file size? (i.e. large text file vs small one, just text *.doc vs *.doc with with additional formatting).
As we can see from scatter graph compression ration does not depend on file size (we cannot imagine any function that would represent dependency).
Files that has additional formatting or inserted pictures tend to has lower compression rate.