# Exercise description

Kindly write a command line tool to analyze the content of log files. The tool should accept arguments as input and return an output after having performed operations on the given input.

## Requirements:

- Python >= 3.7
- Please provide a Dockerfile in your solution with the application as its entry point

## Program arguments

(We expect the below to appear as command line options in the tool you are writing)

- Input data:
  - Path to one or more plain text files, or a directory
- Operations:
  - Most frequent IP
  - Least frequent IP
  - Events per second
  - Total amount of bytes exchanged
- Output
  - Path to a file to save output in plain text JSON format.

## Sample data

As sample input file please use Squid Proxy access logs from the following URL:

https://www.secrepo.com/squid/access.log.gz

*Data from SecRepo website https://www.secrepo.com/#about published under a Creative Commons Attribution 4.0 International License*

The data is in CSV format with 10 fields. After the second field they are separated by a space:

1157689324.156 1372 10.105.21.199 TCP_MISS/200 399 GET http://www[.]google-analytics[.]com/__utm.gif? badeyek DIRECT/66.102.9.147 image/gif

which, after proper parsing should be split as follow:

Field 1: 1157689324.156 *[Timestamp in seconds since the epoch]*
Field 2: 1372 *[Response header size in bytes]*
Field 3: 10.105.21.199 *[Client IP address]*
Field 4: TCP_MISS/200 *[HTTP response code]*
Field 5: 399 *[Response size in bytes]*
Field 6: GET *[HTTP request method]*
Field 7: http://www.google-analytics.com/__utm.gif? *[URL]*
Field 8: badeyek *[Username]*
Field 9: DIRECT/66.102.9.147 *[Type of access/destination IP address]*
Field 10: image/gif *[Response type]*

## Considerations

- Different input log formats might be supported in the future
- Operations might be extended with new ones in future
- Output might be saved in different formats in the future
- Our alerting might be based on this tool so its operation must be fault tolerant and its output must be correct

Since this is an open exercise, the above considerations are not hard requirements of the solution.
However, they can serve as guidelines for the initial application structure to make the extension of the code effortless in the future.
If questions arise during development, please make assumptions and document it in the code by using comments.

## Assignment delivery

Please send us a link to a public repository of your choice containing a link to your tool. The deadline has been communicated to you together with this exercise description. We will review your work and follow up with you within a week of your submission.