As demonstrated in class, there are some interesting inconsistencies in the Google Analytics data for the One-Stop-Shop that should cause us to question or at least be cautious in using this data. As such, it would be a valuable exercise to compare the Google Analytics Data for the One-Stop-Shop (both desktop and mobile versions) to access logs from the Apache server housing the One-Stop-Shop. That is what you are going to do in this project.

Ultimately, your task boils down to answering the following question and providing evidence to support your conclusion. **Is the Google Analytics Data consistent with the data found in the web logs over an extended period of time?** (Approximately one year or longer.)

In the files section of the class in Canvas, I have provided you with the following data for your analysis:

- A copy of the Apache access log for the One-Stop-Shop: *access.log*.
- Copies of the Google Analytics daily session data for both the desktop and mobile versions of the One-Stop-Shop.
- Copies of the Google Analytics referral counts for sites referring/linking to the desktop and mobile versions of the One-Stop-Shop.

You are to create and use original Python scripts that you develop to process and analyze the *access.log* file. The scripts can be extensions of those we developed in class. **You may not use third-party scripts, code, applications, products, etc. etc. that analyze web logs or similar.** Again, **you are to use original Python scripts that you develop for the purpose of this assignment.**

Here is some additional information that may be helpful:

- A (user) session is loosely defined as a continuous block of time in which an individual user is using the web application. I.e., there is a start and an end.
- Sometimes a cutoff/gap is defined to separate sessions. For instance, 20 minutes could be used to indicate that a gap in activity of 20 minutes or more results in the ending of a session.
- The *access.log* file does not directly indicate when users leave pages or close their browsers.
- User requests for both the desktop and mobile versions of the One-Stop-Shop are included in *access.log*. It should be possible to separate the desktop and mobile requests using data in the file for at least some of the data.
- There may have been several times in which this particular server did not serve as the production server for the One-Stop-Shop. At such times, there will be Google Analytics data but there will be no or minimal data in the access log file. (There were network issues with our host, which caused us to revert to a copy on local servers.)
- As mentioned in class, more information about Apache log files can be found here: https://httpd.apache.org/docs/2.4/logs.html (Links to an external site.)Links to an external site.. We use this log format: "%h %l %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{User-agent}i\""
- The desktop and mobile versions of the One-Stop-Shop use different mechanisms and frequencies to automatically refresh data.

- There are different ways, including some that are parameterized, to link to the desktop and mobile versions of the One-Stop-Shop. All of these different ways may be used to initiate a session. But, such links can be invoked in the midst of a session too.
- As we have seen, IP addresses alone might not uniquely identify individual users. In fact, it might be challenging to identify all individual users via the log file.

Your submittal should be a zip archive including the following:

- A two-page MS-Word or PDF document showing your conclusions and supporting evidence. You may include charts generated with third-party tools. (Excel, for example.) However, you may only include analysis (including chart data) that comes from the files provided, the script(s) you write, and the associated output.
- A single page summary of your approach. How did you process the data? What definitions and assumptions did you make? You can treat and present this single page as an appendix to your two page document and refer to items within it accordingly in the two-page document.
- Source code (Python files).