Tyler White

# VBA Final Project: Automated Web Stats Collection for Adobe Industry Strategy Group

## Executive Summary

The Industry Strategy Group within the Omnture Business Unit at Adobe consults with the management teams of strategic clients to help them know how to make the most of their digital business. These engagements are referred to as Strategic Value Assessments and include a competitive analysis of the client's top competitors which involves gathering publicly available data about website traffic/conversion metrics to identify sub-category trends and benchmarks. The purpose of this tool is to automate process of gathering and analyzing those data. When I began doing this work this summer it normally took about two days to gather data and perform a complete analysis. By the end of the summer I could complete the task in two hours. This tool is able to do most of the work in two minutes.

## Implementation Documentation

The process of gathering the requisite data and preparing it for analysis can be broken into 5 parts; First, scrapping the data from the various public websites that measure web traffic/conversion metrics. Second, parsing the data and organizing into a usable form. Third, formatting the data into comparable units. Forth, building summary reports and charts to facilitate further analysis. And lastly, creating a form that allows the user to select companies/domains for comparison.

### Scrapping:

The first few traffic measurement websites I attempted to pull data from were fairly strait forward. Since they are open to the public free of any passwords I was able to use a standard web query to pull data for any real domain available in their data set. However not all of the data sources I had access through (on behalf of Adobe) allowed for this method so I rewrote the code to take advantage of the "agent" module provided in class. You'll notice that those data that I was able to gather using passwords provided by Adobe are obscured from the dataset and this version of the script will not attempt to collect them.

### Parsing:

After scrapping the contents of each webpage I needed pull out the relevant data and sort it appropriately. This is done using "find" and "offset" commands that rely on the usual structure of each

website, but this method fails due to various inconsistencies in the data. I used a handful of error catches to identify these inconsistencies and deal with them appropriately.

### Formatting:

Besides simple text formatting changes the form of the data needed to be changed on occasion. In some cases percentages were shown as whole numbers that would be off by a magnitude if simply converted to percentages and in some cases labels such as “1.9B”, “100k”, and “19M” needed to be converted to whole numbers which required tricky manipulation to preserve the correct number of digits.

### Data Summaries and Charts:

Part of the purpose of this tool is to aggregate all the data Adobe’s Industry Strategy team will use to later indentify overall trends, but when the user pulls data for a given website he/she is generally most interested in that website and its category. In order to facilitate category analysis the script leaves the user viewing a “Summary Data” page that only show the most up-to-date data for the domains he/she has chosen to analyze. This summary data is supplemented with distribution charts to help visualize important patterns in the data for the category in question.

### User Form:

I will be the predominant user of this tool, however I felt it was necessary to make the tool such that other could use it if necessary. I built as simple button attached to a user form to help the user know what to do and how to select companies/domains for analysis. The user form brings up a list of the domains for which I know the relevant data is available and allow the user to select the subset or category he/she would like to analyze.

## Learning and Conceptual difficulties

I was able to figure out the majority of this project by following similar patterns outlined in by the class assignments. My difficulties were in the finer details of execution.  Even after working through all the difficulties of finding and pulling out particular data points within a page of data I found over and over again that slight changes in the websites structure would throw off the script. After testing enough to feel confident that my error catching can account for most variations in website structure I can’t know for sure if I need to test more.

I felt that formatting the data would be a fairly simple step, but I found that the many different text forms were difficult to account for and required a sequence of modifications to change correctly in all cases. I originally recorded my own steps for formatting the data in order to write the script, but I later found that those specific steps only worked when more than one company/domain was being analyzed. When working with only a single domain my method cause excel to freeze up and quit which made it very difficult to identify what was causing the error and account for it.

The last major difficult I ran into resulting from repeated use of the script and calling for the “agent” module to pull website data. At times the “agent” module would not create a tab for the “web query” as directed. After much investigation I found that this and other problems would resolve themselves if I

simply restarted the script. I resorted to exiting the script with a message box directing the user to retry in these cases.