

CSVCompare 1.0

27 November, 2022

- a. **Overview:** CSVCompare is a small Windows-compatible utility application to compare multiple .CSV file contents and then generate a text file which documents instances of the same value(s).
- b. **Limitation:** CSVCompare can compare the contents of between 2 and 10 different .CSV files.
- c. **License:** CSVCompare is released as 'DonorWare' – if you find this application useful, then "Software by Daughtry" (aka: Author) would appreciate a small donation for saving you some time/effort. All rights to CSVCompare are reserved by Author. Author is not responsible for the loss of stored electronic information via the use of CSVCompare. This application is not crippled or limited in any way.
- d. **CSV File Format:** Most commonly associated with Microsoft Excel, a Comma Separated Values file (aka: a CSV) file can be structured two different ways: as a columnar list of data or a single line of values which are separated with a comma character. A CSV file uses a file extension of .CSV; typically, Windows will associate the file extension with whatever spreadsheet application exists on the computer.
- e. **Creating a CSV File:** There are numerous ways of creating a .CSV file. The two easiest ways:
 1. Text Editor/Windows Notepad: Create a new text file (also known as an ASCII file). Enter values which are either separated by a comma character or one value per line.
 2. Spreadsheet: Populate column A with values (one value per cell). When saving the file in Microsoft Excel, ensure the file is saved using the 'CSV (MS-DOS) (*.csv)' file format.
- f. **Configuration File:** CSVCompare uses a .Ini file to define what .CSV files will be compared against each other. A .Ini file is a small text file, typically created with Notepad or a text editor. The CSVcompare.Ini file is structured as such:

[Files]

CSV01=<filename.csv>

CSV02=<filename.csv>

CSV03=<filename.csv>

CSV04=<filename.csv>

CSV05=<filename.csv>

CSV06=<filename.csv>

CSV07=<filename.csv>

CSV08=<filename.csv>

CSV09=<filename.csv>

CSV10=<filename.csv>

Replace the <filename.csv> information with the name of the .CSV file. At least two .CSV files must be configured within the CSVCompare.Ini file. Delete lines within the CSVCompare.Ini that are unneeded; for example, if only three .CSV files will be compared (a.csv, b.csv, ccc.csv), the CSVCompare.Ini file would resemble:

[Files]

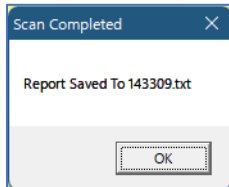
CSV01=a.csv

CSV02=b.csv

CSV03=ccc.csv

- g. **How It Works:** When the CSVCompare.Exe file is started, the CSVCompare.Ini file is opened and read; each of the defined .Csv files are opened and their contents are read into a unique array (i.e. memory). A uniquely-named text file is created/opened for writing output. Each array's contents are then compared against each other, in succession, until the last array is compared. If an array's element is found within another array, then a line is appended to the output text file. After the last array elements are compared, the output text file is amended with a reporter footer and then closed. A message box is then displayed onscreen to show the unique

filename that the output was generated to:



h. **Sample Scenario:** The CSVCompare.INI file looks like this:

```
[Files]
CSV01=A1.CSV
CSV02=B2.CSV
CSV03=C3.CSV
CSV04=D4.CSV
CSV05=E5.CSV
CSV06=F6.CSV
CSV07=G7.CSV
CSV08=H8.CSV
CSV09=I9.CSV
CSV10=J10.CSV
```

The contents of the ten .CSV files are:

A1	B2	C3	D4	E5
Volvo Cobra AC Nagant GM Puma Bugatti Bridglin Studebaker Geely Alfa Romeo DeLorean	Indian Jaguar Kia Crosley Land Rover Lincoln Lamborghini Matra Nissan	Chrysler Datsun Daihatsu Daimler Geo Honda Invicta Indian	Ford Dodge Plymouth Jaguar Ferrari Aston Marton Audi Bentley Chrysler	Yugo McClaren
F6	G7	H8	I9	J10
Opel Oldsmobile Packard Plymouth R.E.O. Reliant Scimitar Scion Toyota	McClaren Peugeot Maserati Auburn Checker Dort Motor Company Du Pont Motors Fisker Hudson Hupp Imperial MG Kaiser-Frazer	Toyota Volkswagon Porsche Ford Yugo Sterling Fiat BMW Acura Alfa Romeo	DeLorean Hummer AMC Crosley Duesenberg Excalibur Nash LaSalle Peerless Pontiac Stutz Studebaker Willys White	Daihatsu LaSalle

When the CSVCompare.Exe is executed, the resultant output file contains the following information:

```
File #1 is: D:\A\A1.CSV containing 12 items
File #2 is: D:\A\B2.CSV containing 9 items
File #3 is: D:\A\C3.CSV containing 8 items
File #4 is: D:\A\D4.CSV containing 9 items
File #5 is: D:\A\E5.CSV containing 2 items
File #6 is: D:\A\F6.CSV containing 9 items
File #7 is: D:\A\G7.CSV containing 13 items
File #8 is: D:\A\H8.CSV containing 10 items
File #9 is: D:\A\I9.CSV containing 14 items
File #10 is: D:\A\J10.CSV containing 2 items

Studebaker was found in D:\A\A1.CSV and D:\A\I9.CSV
Alfa Romeo was found in D:\A\A1.CSV and D:\A\H8.CSV
DeLorean was found in D:\A\A1.CSV and D:\A\I9.CSV
Indian was found in D:\A\B2.CSV and D:\A\C3.CSV
Jaguar was found in D:\A\B2.CSV and D:\A\D4.CSV
Crosley was found in D:\A\B2.CSV and D:\A\I9.CSV
Chrysler was found in D:\A\C3.CSV and D:\A\D4.CSV
Daihatsu was found in D:\A\C3.CSV and D:\A\J10.CSV
Ford was found in D:\A\D4.CSV and D:\A\H8.CSV
Plymouth was found in D:\A\D4.CSV and D:\A\F6.CSV
Yugo was found in D:\A\E5.CSV and D:\A\H8.CSV
McClaren was found in D:\A\E5.CSV and D:\A\G7.CSV
Toyota was found in D:\A\F6.CSV and D:\A\H8.CSV
LaSalle was found in D:\A\I9.CSV and D:\A\J10.CSV

=====
CSVCOMPARE UTILITY APPLICATION
Build Number: 11/27/2022
Company: Software by Daughtry
Support: http://www.sdaughtry.com
Report Generated On: Sun 11/27/2022 3:33:50 PM
Number of Seconds the Comparison Required: 1
```

i. Notes:

1. The file comparison operation is CaSe SeNsItIvE!! For example, a match would be found if two files contain the text 'LaSalle', but a match is NOT found if file (1) contains 'LaSalle' and file (2) contains 'LASALLE'.
2. Trailing spaces within .CSV file contents are problematic. This is also a problem with Microsoft Excel (and possibly other spreadsheet applications). If file (1) entries do NOT contain a trailing space, but some of file (2) entries do contain a trailing space, then a match will NOT be found. Microsoft Excel (and possibly other spreadsheet applications) will provide unreliable results within PivotTables and calculated data if some of the data (manually inputted or imported) sporadically contains trailing spaces. This problem may be mitigated by Excel add-on products; the use of Excel formula(s) to remove trailing spaces or conducting a search/replace operation against the contents of a .CSV file via a text editor prior to using CSVCompare.
3. Although not required, it is recommended the two files which comprise CSVCompare (i.e. CSVCompare.Exe, WBDVD64I.Dll), its configuration file (CSVCompare.Ini) and CSV files that will be compared reside in the same folder. WRITE (or MODIFY) permissions are required of this folder to ensure the resultant CSVCompare output report can be created and then written to).

- j. Scenarios:** A common usage scenario is comparing data archives against one another to find matches. For example, files 2/3/4/5 contain an archive of monthly data. File #1 should compare the new data (e.g. this month's data) that needs to be compared against the four archive .CSV files. Although this is not a requirement (because CSVCompare does a cross comparison of every .CSV files' content against the other in every possible direction), the resultant output report makes more 'sense' if the reader knows that the first .CSV file contained the 'new' information.

The sample files used to create this scant Readme file were purposely minimalistic. The author built this application to augment cyber analyses which needed to compare archived lists of TCP/IP addresses (and flag duplicate entries). This real world application compares four .CSV files

file #1 – this month's list of unique IP addresses

file #2 – company #1 unique IP address list which contained 23,274 entries

file #3 – company #2 unique IP address list which contained 43,749 entries

file #4 – company #2 unique IP address list which contained 11,193 entries

CSVCompare required 535 seconds (i.e. roughly 9 minutes) to generate the report, identifying 424 matches throughout the .CSV files.

- k. File Alteration:** CSVCompare does not purposely alter/modify the contents of a .CSV file during its operation. CSVCompare does not modify the Windows Registry database, nor does it require elevated privileges to operate. No file associations are made by CSVCompare. The output text file (aka: the report) unique filename is generated from the computer's current date and time
- l. Support:** Limited support is provided via the author's web site: <http://www.sdaughtry.com>.
- m. Application Download:** While it is possible this application may be posted on alternate web sites (e.g. Shareware or Freeware repositories), the Author does not recommend downloading this application from any web site other than the Software by Daughtry web site to ensure its not been tampered with.