

This program reports on an aspect of data quality in a logical table. Specifically, it looks for duplicate column values in a row. The logical table is represented by a text file. The duplicate checker reads rows of data, where each record of the file is a logical row and multiple column values are separated from each other as specified by the file format. One supported file format is CSV. The reporter then computes the required reports.

The program operates entirely in memory and executes these three phases:

1. Read in column data, first row are labels
2. Process data to cumulate counts for reports (uses a `DupChecker` implementation)
3. print reports as required from the cumulated data

The program is packaged as a runnable Jar and contains supporting libraries. Reports are triggered by program parameters and are optional. The report definitions and program (command line) options to trigger them are:

- **-f** Defines the input file path to check.
- **-l** Display a formatted listing of the input data.
- **-s** Display a Summary Report that prints the null, complete, incomplete, and duplicate row-column combinations.

You execute the program using:

```
%JAVA_HOME%\bin\java -jar DupCheck.jar {options above}
```

**Note these conceptual definitions:**

- A *complete* column has 2 or more characters, fewer than 2 are incomplete.
- Two non-null columns are *duplicates* if their values are the same in the same row.