

# CSV File Splitter

A scalable Node.js tool to split large CSV files into smaller chunks, with advanced email and telephone processing. Supports both API and CLI usage.

## Features

- Splits huge CSVs into multiple files with a configurable number of records per file
- Handles semicolon-separated emails and telephones, creating separate rows/fields as needed
- Skips compliance notice lines and empty email records
- Outputs a detailed summary in `summary.txt`

## CLI Usage

### 1. Install dependencies:

```
npm install
```

### 2. Run the splitter on a CSV file:

- Place your CSV file in `src/uploads/`.

#### Normal Split (by batch size):

```
node csv-split.js <filename> [--batch=SIZE]
```

- `--batch=SIZE` (optional): Number of records per output file. If not provided, uses `CSV_SPLIT_SIZE` from environment or defaults to 100000.

#### Range Split (by row range and batch size):

```
node csv-split.js <filename> --range=START:END [--batch=SIZE]
```

- `--range=START:END` (required for range mode): Only rows from START to END (inclusive) will be processed (1-based, header not counted).
- `--batch=SIZE` (optional): Number of records per output file within the range. If not provided, all selected rows go into a single file.
- The output files and `summary.txt` will be created in a timestamped folder under `uploads/splits/` by default.

## API Usage

### 1. Start the server:

```
npm run dev
```

### 2. Call the split endpoint:

- POST to `/api/split-company-contacts`
- The input file is taken from `src/uploads/input.csv` by default, or set `CSV_INPUT_PATH` in your `.env`.
- Example with `curl`:

```
curl -X POST http://localhost:4000/api/split-company-contacts
```

- The response will include summary details and the output directory path.

## Example Summary Output

Below is an example of the summary table generated after splitting a CSV file as `summary.txt` :

```
=====
CSV Split Summary
=====
Input:          D:\CSV_file_splitter\src\uploads\input-test.csv
Output Directory:  D:\CSV_file_splitter\src\uploads\splits\input-test_20250829095306
Files Created:    11
Total Records:   1086548
Records Without Email: 1058
Records Processed: 1085490
Chunk Size:      100000
=====
```

## Environment Variables

- `CSV_INPUT_PATH` - Path to the input CSV file (default: `src/uploads/main_file.csv`)
- `CSV_SPLIT_SIZE` - Number of records per output file (default: 100000)

## Output

- Split CSV files with headers in a `uploads/split/`
- `summary.txt` in the output directory with a formatted summary table

## Example

### CLI Examples:

```
# Normal split (default batch size or from ENV)
node csv-split.js input.csv

# Normal split with custom batch size
node csv-split.js input.csv --batch=50000

# Range split (rows 1 to 1500, batch size 50)
node csv-split.js input.csv --range=1:1500 --batch=50

# Range split (rows 1001 to 2000, all in one file)
node csv-split.js input.csv --range=1001:2000
```

### API:

```
curl -X POST http://localhost:4000/api/split-company-contacts
```