

CSV 중복 정리 실습 1.0

같은 고객의 주문 두 건을 지우지 않도록, 중복 기준을 먼저 정하고 제외할 행을 검토하는 브라우저 앱입니다. 데이터가 맞는지 업무 관점에서 판단하는 일은 사용자가 합니다.

- 실행: <https://seoin.dev/samples/csv-cleaner/>
- 제작 글: <https://seoin.dev/vibe-coding-csv-cleaner/>
- 소스: <https://seoin.dev/samples/csv-cleaner/csv-cleaner-project.zip>

시작하기

ZIP을 풀고 [index.html](#)을 브라우저로 여세요. 설치·API 키·회원가입은 필요 없습니다. [core.js](#), [app.js](#), [style.css](#)를 같은 폴더에 두세요. 인터넷 연결 없이 로컬 파일로 실행할 수 있습니다. 글·출처 링크를 여는 데는 인터넷이 필요합니다.

1. 가상 주문 8행으로 시작을 누릅니다. 주문번호만 중복 기준으로 선택됩니다.
2. 읽은 데이터 8행, 남을 행 6행, 제외할 행 2행, 기준이 비어 유지한 행 2행인지 봅니다.
3. 원본 4행을 펼칩니다. 2행과 주문번호는 같지만 메모가 다릅니다. 이 행도 유지를 누르면 정리본은 7행입니다.
4. 기준 열과 제외 목록을 확인한 뒤 확인란을 체크하고 정리본을 내려받습니다.
5. [examples/orders.csv](#)를 직접 불러오면 처음에는 모든 열을 기준으로 선택합니다. 그대로 검사하면 완전히 같은 5행만 제외해 7행을 남깁니다. 예제 버튼의 초기 기준과 구별하세요.

CSV 정리만 한 번 필요하다면 엑셀의 기존 기능도 먼저 확인하세요. 이 실습은 별도 검토 화면과 업무 규칙을 직접 만드는 연습입니다. [Microsoft 중복 제거 안내](#)

파일과 개인정보

앱은 선택한 파일을 메모리에서 읽습니다. 원본 파일을 덮어쓰지 않습니다. CSV 내용·파일명·입력값을 서버로 전송하거나 브라우저 저장소에 기록하는 코드는 없습니다. 외부 API·분석 스크립트도 사용하지 않으며 CSP의 연결 정책을 `connect-src 'none'`으로 둡니다. 공개 주소를 여는 일반 웹 요청 자체는 웹 서버 로그에 남을 수 있습니다. 내려받은 파일의 보관과 삭제는 사용자가 관리합니다.

UTF-8 텍스트, 2MiB(화면에서는 2MB로 표기), 데이터 10,000행, 50열까지입니다. 쉼표·탭·세미콜론을 지원하며 첫 비어 있지 않은 행을 열 제목으로 읽습니다. 열 제목이 비거나 중복되면 멈춥니다. 빈 행은 건너뛰고 개수를 표시합니다. 기준 열에 빈 값이 하나라도 있으면 해당 행은 중복으로 합치지 않습니다.

행 번호는 열 제목을 1행으로 세는 CSV 레코드 번호입니다. 셀 안 줄바꿈은 별도의 행이 아닙니다. 열 제목 앞의 빈 행도 원본 번호에 포함합니다. 정리본의 순서는 원본 순서입니다. 비교 화면은 한 페이지 20행, 정리본 미리보기는 처음 20행이며 내보내기는 전체 결과입니다.

저장할 때의 범위

정리본 CSV는 UTF-8 BOM·침표 구분·큰따옴표 인용으로 저장합니다. 원본의 구분 문자·인용 모양·빈 행까지 그대로 복사하는 파일이 아니라 **남긴 셀 값을 보존하는 새 파일**입니다. 숫자로 변환하지 않아 **00123** 같은 문자열을 유지하지만, 엑셀이 CSV를 열 때 숫자로 해석하면 표시가 달라질 수 있습니다. 가져오기에서 해당 열을 텍스트로 지정하세요. [Microsoft 앞자리 0 안내](#)

=, +, -, @ 등으로 시작해 스프레드시트에서 수식으로 해석될 수 있는 값이 남아 있으면 CSV 저장을 차단합니다. 공백·전각 기호·일부 제어문자도 검사하므로 음수와 국제 전화번호도 걸릴 수 있습니다. 자동으로 값을 고치거나 작은 따옴표를 붙이지 않습니다. 원문을 확인하거나 문자열을 보존하는 JSON으로 보관하세요. **모든 스프레드시트 프로그램에서 안전함을 보증하는 변환기는 아닙니다.** [OWASP CSV Injection](#)

정리본 JSON은 **headers**와 문자열 배열 **rows**입니다. 검토 기록 JSON은 기준 열·순서·옵션·제외한 원본 행 번호·수동으로 살린 행 번호를 담습니다. 원본 셀 값은 담지 않지만 열 제목은 포함합니다. **.xlsx**, 셀 서식, 수식 계산, 여러 파일 합치기, 날짜의 의미 판단, 유사한 이름 추정 은 지원하지 않습니다.

직접 바꾸기

- 이 PDF의 요구사항: 구현 전에 정한 합격 조건
- 이 PDF의 요청 예시: AI에게 요청할 때 쓸 문장과 변경 연습
- [learning/naive.cjs](#): 일부러 단순하게 만든 실패 예제. 실제 AI 응답 기록이 아닙니다.
- [learning/failure-output.json](#): 같은 입력으로 단순 구현과 완성본을 실행한 결과
- 이 PDF의 제작·검증 기록: 제작·검증 범위와 남은 확인
- [verify.cjs](#): 처리 규칙 37개 검사

Node.js가 설치되어 있다면 이 폴더에서 다음을 실행하세요. 앱 자체 실행에는 Node가 필요하지 않습니다.

```
node verify.cjs
node learning/naive.cjs
```

구현 근거

CSV 인용 규칙은 [RFC 4180](#)을 참고합니다. 다양한 CSV가 존재하므로 모든 형식을 지원한다고 주장하지 않습니다. 파일 읽기는 [Blob.arrayBuffer](#)와 [TextDecoder.decode](#)를 사용합니다.

교육용 코드와 예제는 MIT License로 제공합니다. 실제 고객 데이터를 넣기 전에 가상 파일로 기대 결과를 확인하세요. 2026-09-07 · seoin.dev.

먼저 정한 요구사항

만들 목적

가상 주문 CSV에서 중복 후보를 찾아 사용자가 제외 여부를 결정하고, 원본과 별개인 정리를 받는다. 사람이 업무 규칙을 정해야 하는 지점을 화면에 드러낸다.

합격 조건

항목	조건
입력	CSV UTF-8, 2MiB 이하, 데이터 10,000행 이하, 50열 이하
초기 선택	사용자 파일은 모든 열. 가상 주문 버튼은 주문번호만 선택했다고 명시
같은 기준	선택한 열의 값이 모두 같아야 같은 묶음
기본 보존	먼저 나온 행. 사용자가 나중 행으로 바꿀 수 있음
공백·대소문자	기본 구별. 옵션은 비교에만 적용하고 원본 값은 수정하지 않음
빈 기준값	선택한 열 중 하나라도 비면 그 행은 유지
제외 검토	원본 행 번호와 남길 행 번호, 다른 값 표시, 후보별 추가 유지
입력·기준 변경	이전 결과와 검토 완료 상태를 무효화
잘못된 파일	조용히 일부만 처리하지 않고 오류를 알려 중단
출력	원본 순서로 CSV/JSON, 검토 기록은 별도 JSON
수식처럼 보이는 값	CSV 내보내기 중단, 값 자동 수정 없음, JSON은 문자열로 저장
개인정보	파일 내용 외부 전송·영구 저장·분석 이벤트 없음

예제의 정답

[orders.csv](#) 데이터는 8행이다. 열 제목이 1행이다.

- 주문번호 + 먼저 나온 행: 4·5행 제외, 2·3·6·7·8·9행 유지 → 6행.
- 주문번호 + 나중에 나온 행: 2·3행 제외 → 6행.
- 모든 열: 완전히 같은 5행만 제외 → 7행.
- 첫 설정에서 4행을 수동 유지: 5행만 제외 → 7행.
- 주문번호가 없는 8·9행은 서로 합치지 않는다.
- 같은 고객이어도 A002와 A003은 다른 주문이다.

범위 밖

엑셀 파일 직접 편집, 사람 이름의 유사도, 전화번호 자동 변환, 여러 파일 병합, 서버 저장, 로그인, AI API 호출, 금전·시간 절감 성과 추정.

AI에게 요청하고 직접 확인할 문장

다음은 독자가 재사용하도록 작성한 요청 예시입니다. 제작 당시 AI 대화의 원문이나 한 번에 성공한 기록이 아닙니다. 사용하는 AI·설정에 따라 결과가 달라집니다. 요구사항과 검증 파일은 함께 건네고 실제 업무 파일 대신 가상 데이터를 사용하세요.

처음 만들기

이 PDF의 '먼저 정한 요구사항'을 읽고 브라우저에서 실행하는 CSV 중복 정리 앱을 만들어 줘.
먼저 구현 범위와 확인이 필요한 결정을 짧게 적어 줘.
index.html, style.css, core.js, app.js로 나눠 줘.
외부 API·라이브러리·서버·로그인 없이 실행해야 해.
파일을 읽는 함수와 중복 판단 함수를 화면 코드에서 분리해 줘.
기준 열을 고르고, 제외할 원본 행을 확인하고, 필요하면 살린 뒤 저장하게 해 줘.
화면이 뜨는 것과 데이터가 맞는 것은 별개야.
orders.csv의 정답을 각각 비교하고 quotes.csv도 읽어 줘.

심표가 있는 메모에서 깨질 때

quotes.csv의 '심표, 보존'은 하나의 셀이야.
그런데 단순 split(',') 구현은 심표에서 셀을 둘로 나눠.
셀 안 줄바꿈과 큰따옴표 두 개로 표현한 따옴표도 보존해야 해.
RFC 4180의 인용 규칙을 참고해 읽기 함수를 수정해 줘.
원본 셀 배열과 CSV 저장 후 다시 읽은 셀 배열이 같음을 검사해 줘.
중복 판단과 내보내기 조건은 유지해 줘.

잘못 지워진 주문이 있을 때

같은 이메일의 A002와 A003은 서로 다른 주문이야.
고객 이메일을 기준으로 쓰면 다른 주문까지 합칠 수 있어.
주문번호만 골랐을 때 orders.csv의 6행이 남는지 확인해 줘.
빈 주문번호 8행과 9행도 남겨야 해.
앱이 기준을 몰래 바꾸지 않도록 현재 기준을 결과 위에 보여 줘.

자기 업무로 바꾸는 연습

아래 중 한 가지를 골라 기준과 정답부터 바꾸세요. 값이 실제로 같은 뜻인지는 본인이 판단해야 합니다.

- **출석 명단**: 이메일만 고르면 다른 날짜의 출석도 합쳐질 수 있습니다. **이메일 + 행사 날짜**가 같은 경우의 가상 파일부터 만드세요.
- **상품 목록**: 상품명만 고르면 색상·옵션이 다른 상품을 지울 수 있습니다. **상품번호 + 옵션**으로 기대 결과를 먼저 적으세요.
- **상담 기록**: 같은 고객의 후속 상담을 남겨야 한다면 고객번호 하나로 합치지 마세요. 중복 제거가 적합한 작업인지부터 다시 판단하세요.

이번에는 [업무]의 [열 이름들]이 모두 같은 행을 중복 후보로 보려 해.
가상 데이터 [행 수]행과 기대 결과 [남을 행 번호]를 준비했어.
어떤 기준 검사에 영향을 주는지 먼저 설명해 줘.
한 가지 변경만 적용하고 기존 verify.cjs와 새 가상 파일을 함께 검사해 줘.
기존 결과가 달라졌다면 이유를 보여 줘.

제작과 검증 기록 · 2026-09-07

이 앱은 seoin.dev의 바이브코딩 실전 프로젝트로 AI의 도움을 받아 제작했습니다. 모든 입력 예제는 가상 데이터입니다. 실제 회사의 업무 시간 절감이나 독자의 완주 성과를 측정한 사례가 아닙니다.

실제 작업에서 확인한 것

1. 중복 기준·빈 값·원본 보존·내보내기 조건을 먼저 적었습니다.
2. CSV 처리와 화면을 분리했습니다. Node.js v22.22.0에서 처리 규칙 37개가 통과했습니다.
3. Orca 브라우저에서 화면 상태와 파일 입력·내보내기를 검사했습니다. 다른 파일을 읽어도 이전의 '나중 행 유지' 설정이 남는 동작을 발견했습니다. 새 파일 내용을 읽으면 보존 순서와 공백·대소문자 옵션을 기본값으로 되돌리도록 수정했습니다.
4. 수정 후 브라우저 통합 검사 32개가 통과했습니다. 10,000행의 비교 페이지 이동, 기준 변경 후 이전 결과 잠금, 수동 보존, 입력 HTML의 문자열 표시, 느린 파일 읽기와 새 입력의 충돌을 포함합니다.
5. CSV·JSON 내보내기는 생성한 Blob을 다시 읽어 결과를 비교했습니다. 이것은 사용자의 다운로드 폴더 저장 완료나 엑셀 호환성을 확인한 검사가 아닙니다.

브라우저 테스트의 일부 파일 오류와 지연은 합성 File 객체로 재현했습니다. 실제 파일 선택은 별도의 Orca upload 명령으로 확인했습니다. 첫 화면 폭 검사는 DOM 경계 측정이었습니다. 2026-09-07 후속 검수에서는 Orca로 폭 390px·1280px의 시작 화면, 기준 선택, 결과, 펼친 행 비교 화면을 캡처해 직접 확인했습니다. 본문 표·이미지와 자료 안내도 같은 두 폭에서 확인했습니다. Chrome 계열 Orca 환경에서 검사했고 Safari·Firefox·실제 모바일 기기는 별도로 확인하지 않았습니다.

실패 예제와 실제 버그를 구분합니다

[learning/naive.cjs](#)는 CSV의 인용 처리가 필요한 이유를 보여 주려고 일부러 작성한 단순 구현입니다. 실제 AI가 그 코드를 생성했다는 주장이 아닙니다. 같은 [quotes.csv](#)로 실행하면 단순 구현은 물리 줄 5개를 3·4·2·2·3열로 나누고, 완성본은 데이터 3행·3열을 유지합니다. 결과는 [learning/failure-output.json](#)에 있습니다.

반면 '새 파일을 읽어도 이전 보존 순서가 남던 동작'은 제작 중 브라우저 검사에서 실제로 발견해 고친 문제입니다.

다음 검증

다른 브라우저와 실제 휴대폰에서 파일 선택·다운로드·줄바꿈을 확인하고, 처음 접한 독자가 기준 열과 제외 목록의 의미를 이해하는지 관찰해야 합니다. 자동 검사 통과를 독자 완주나 업무 적합성으로 대신하지 않습니다.