

교통카드 데이터 전처리 코드

지역별 입력 차이를 처리하는 공통 실행 구조

이력서의 같은 제목 → 포트폴리오 PDF 2쪽 → 이 작업 기록

내 역할 R→Python 전환, 입력 표준화, 실행 설정 분리, 기존 결과 대조
도구 Python · pandas · argparse · Excel
기록 2026.07.13-14 개편 · 검증 기록

무엇을 했나

도시마다 다른 파일명과 항목 때문에 반복 수정이 필요했던 교통카드 분석 코드를 R에서 Python으로 전환했습니다.

어떻게 만들었나

Python·pandas로 입력 항목을 정리하고 지역별 설정과 계산 로직을 분리했습니다. 지역과 폴더를 지정해 같은 계산을 실행하도록 구성했습니다.

왜 이렇게 했나

기존 계산을 유지하면서 도시별 입력 차이만 바꾸기 위한 구조입니다. 필요한 파일이 없거나 선택이 모호하면 실행 전에 멈추도록 했습니다.

결과와 의미

원주 7일 자료에서 기존 산출물 49개의 행 수·열 순서·식별값·핵심 산출값을 대조하고 최종 혼잡도 결과의 일치를 확인했습니다. 알려진 입력 형식은 설정으로 재사용하고 새 지역은 표본 검증 후 적용하도록 정리했습니다.

확인된 개편·검증 기록: 2026.07.13-14 / 검증에 사용한 원주 자료 기간: 2023.10.16-22



웹 작업 기록

QR 또는 제목 링크로 상세 내용 열기

WORK LOG

작업 타임라인

해당 기간의 기록을 작업 순서로 정리했습니다. 검증 데이터는 2023.10.16-22 원주 자료입니다.

지역마다 달랐던 파일·항목 정리

동일한 의미의 항목이 노선명칭·노선명, 정류장순번·정류장순서처럼 다른 이름으로 들어왔습니다. 알려진 입력 차이를 지역 설정과 항목 대응표로 분리했습니다.

확인한 결과

지역별 코드를 복제하기보다 입력을 공통 형식으로 맞추는 구조를 구성했습니다.

Python 공통 계산과 입력 검사 연결

지역명과 입력 폴더를 받아 기존 하차 위치·시간 보정과 혼잡도 계산을 실행하도록 했습니다. 필수 파일 누락·지역 불일치·중복 후보를 실행 전에 검사했습니다.

지역·폴더 지정

파일·항목 확인

같은 계산 실행

이전 결과 대조

지역별 저장

확인한 결과

입력이 모호할 때 임의의 파일로 결과를 만들지 않고 원인을 확인한 뒤 다시 실행하도록 했습니다.

중간표 차이의 원인부터 분리

R 중간표의 7행 차이가 원본의 중복 식별값에서 비롯됐음을 확인하고, 처리 과정에서 새로 부여되는 일련번호는 비교에서 제외했습니다.

확인한 결과

원주 7일 자료의 산출물 49개(결측보정 14개·시간보정 21개·혼잡도 14개)를 단계별 기준에 맞춰 대조하고 최종 혼잡도 결과의 일치를 확인했습니다.

새 지역 적용 절차 정리

일부 체인 자료는 열 순서를 전제로 처리하므로 새 지역의 항목·예외 조건을 먼저 확인하도록 했습니다. 표본 실행과 결과 대조를 거쳐 적용하는 순서를 남겼습니다.

확인한 결과

최종 결과는 재사용 가능한 실행 구조와 검증 절차입니다. 모든 도시의 새로운 형식을 자동 지원한다는 의미는 아닙니다.

RESULT & LEARNING

최종 결과와 배운 점

원주 7일 자료에서 기존 산출물 49개의 행 수·열 순서·식별값·핵심 산출값을 대조하고 최종 혼잡도 결과의 일치를 확인했습니다. 알려진 입력 형식은 설정으로 재사용하고 새 지역은 표본 검증 후 적용하도록 정리했습니다.

전환 작업의 성공 기준을 코드 실행 여부보다 기존 업무 결과의 유지에 두었습니다. 입력 표준화와 비교 기준을 함께 정리해야 다음 지역에서도 사용할 수 있었습니다.

▼ 세부 구현·검산·평가 조건 읽기

항목 이름을 맞추고 기존 계산 결과와 비교

지역별 파일 경로와 예외 조건을 실행 설정으로 분리해 계산 코드를 복제하지 않고 사용할 수 있도록 했습니다. 알려진 컬럼 명칭을 변환하고 숫자 자료형을 맞췄으며, 입력이 모호할 때 임의의 파일을 선택하지 않도록 처리했습니다.

원주 검증에서는 중간 R 마스터의 중복과 재생성 식별키를 비교 기준에 맞춘 뒤 업무 컬럼을 대조했습니다. 최종 산출물은 동일함을 확인했지만, 모든 중간 파일의 원시 행이 처음부터 같았다는 의미는 아닙니다.

연속 이동을 연결한 일부 체인 자료는 정해진 열 순서를 전제로 처리합니다. 새 지역은 항목 순서와 예외 조건, 표본 결과를 확인한 뒤 적용해야 합니다.

알려진 입력 형식을 정리한 구조입니다. 새로운 형식은 항목 순서와 예외 조건을 먼저 확인해야 하며 회사 계산 코드는 공개하지 않습니다.

실행 전 확인 예시

필수 파일이 없으면 누락된 역할을 표시하고 중단합니다. 선택 지역과 파일의 지역이 다르면 불일치를 표시하며, 같은 역할의 파일 후보가 여러 개라 모호하면 사용자가 정리한 뒤 다시 실행하도록 합니다.

▼ 용어와 집계 단위

용어	이 자료에서의 의미
항목명·공통 계산	컬럼명은 표의 열 이름입니다. 같은 뜻의 명칭을 맞추고 지역 설정을 바꿔 같은 계산을 실행하도록 구성했습니다.
모호한 파일 선택	필수 자료의 같은 역할을 하는 파일이 여러 개면 어느 파일을 쓸지 임의로 결정하지 않습니다. 자료 안의 중복 행을 제거하는 것과 다른 검사입니다.

수행 결과와 확인하지 못한 범위를 함께 기록했습니다.

<https://zodia8393.github.io/career-portfolio/migration/>