

1 프로젝트 개요

현재 프로젝트는 대한민국 아파트 실거래가 데이터에 대한 하위지역 단위별 지니계수를 계산합니다.

2 설치 및 환경 구성

2.1 개발 환경/의존성

- Python 버전: 3.9+
- 주요 라이브러리: pandas, numpy, yaml 등

2.2 폴더 구조

```
project/
├── data/
├── src/
│   ├── election_processor.py
│   ├── preprocess.py
│   └── ...
├── main.py
├── config.yaml
└── requirements.txt
```

3 사용 방법

3.1 실행 방법

프로젝트를 실행하기 전, config.yaml 파일을 확인하고 필요한 설정(DB 경로, 매핑 폴더 등)을 기입합니다.

4 주요 모듈 설명

4.1 election_processor.py

- 역할: 전체 파이프라인(전처리, 매핑, 지니계수 계산, 결과 저장)을 총괄
- 주요 함수:
 - process_election_data(): 단일 선거 데이터 처리
 - process_and_save_all_elections(): 여러 선거 일괄 처리

4.2 preprocess.py

- 역할: 원본 데이터 전처리 (결측치, 형 변환 등)
- 클래스/함수:
 - DataProcessor: 전처리 로직 담당

4.3 matching.py

- 역할: 법정동/행정동 코드 매칭 담당
- 클래스/함수:
 - Matcher: gen_bdong(), conn_code() 등 시점별 코드 변환