서울시 범죄 통계 데이터

- 1. 목적 및 배경
- 2. 데이터 개요
 - 2.1 데이터 소스
 - 2.2 주요 컬럼
 - 2.3 대상 데이터
- 3. 전처리 목표
- 4. 전처리 단계
 - 4.1 데이터 로드
 - 4.2 인구 데이터 수집
 - 4.3 주소 데이터 처리
 - 4.4 주소 통계 데이터 처리
 - 4.5 범죄율 계산
 - 4.6 데이터 저장
- 5. 결과 및 요약
- 6. 향후 작업
- 7. 부록

1. 목적 및 배경 ⊘

서울시 범죄 통계 데이터를 수집, 정제 및 분석하여 데이터베이스에 저장하고, 시각화와 추가 분석을 위한 준비를 목표로 합니다.

2. 데이터 개요 ∂

2.1 데이터 소스 ⊘

- 범죄 통계: 로컬 CSV 파일(서울시_범죄_통계.csv)
- 생활 인구: 서울시 OpenAPI (http://openapi.seoul.go.kr:8088/)>

2.2 주요 컬럼 ∂

- 범죄 통계 csv
 - 범죄 대분류: 범죄의 주요 분류 (예: 폭력, 재산 범죄)
 - 범죄 중분류: 세부 범죄 유형 (예: 살인, 강도)
 - ∘ 서울 xx구: 구별 범죄 발생 건수
- 서울시 생활 인구 API
 - ADSTRD_CODE_SE: 행정동 코드
 - o TOT LVPOP CO: 구별 총생활인구

2.3 대상 데이터 ⊘

서울시 25개 구의 범죄 통계와 생활 인구 데이터를 통합하여 인구 10만 명당 범죄 발생률을 계산하고, 이를 기반으로 지역별 범죄 패턴과 트렌드를 분석

3. 전처리 목표 ⊘

1. 서울시 범죄 통계 데이터를 구별로 정리

- 2. 서울시 생활 인구 데이터를 OpenAPI를 통해 구별로 수집
- 3. 인구 10만 명당 범죄 발생률(crime_rate)을 계산
- 4. 정제된 데이터를 CrimeStats 테이블에 저장

4. 전처리 단계 ♂

4.1 데이터 로드 ⊘

- 작업
 - ∘ 범죄 통계 CSV 파일(서울시_범죄_통계.csv)을 읽어 데이터프레임으로 로드

4.2 인구 데이터 수집 ♂

- 작업
 - 서울시 OpenAPI를 호출하여 구별 총생활인구(TOT_LVPOP_CO) 데이터를 수집
 - API 응답에서 district_code와 일치하는 데이터를 추출

4.3 주소 데이터 처리 ⊘

- 작업
 - 데이터베이스에서 주소를 조회하거나 없는 경우 새 주소를 생성
 - ∘ 구 이름(서울XX구)을 기준으로 Address 테이블의 데이터를 관리

4.4 주소 통계 데이터 처리 ⊘

- 작업
 - 범죄 유형별 데이터를 정리하고 인구를 기준으로 범죄율을 계산
 - 。 CrimeStats 테이블에 데이터를 저장

4.5 범죄율 계산 ⊘

- 작업
 - ∘ 인구 10만 명당 범죄 발생률(crime_rate)을 계산
- 공식
 - 범죄율=(범죄 건수 / 총 인구) × 100,000

4.6 데이터 저장 ⊘

- 작업
 - ∘ 정제된 데이터를 CrimeStats 테이블에 저장

5. 결과 및 요약 ∂

- 처리 전 데이터
 - CSV 파일에서 서울시 25개 구의 범죄 통계를 로드.
- 처리 후 데이터
 - 총 13개 주요 범죄 유형에 대해 각 구의 범죄 발생 건수 및 범죄율을 계산.
 - ∘ 데이터베이스에 총 325개(25구 × 13범죄유형)의 데이터를 저장.

6. 향후 작업 ∂

- 위치 좌표 업데이트
 각 구의 위도 및 경도를 추가하여 지도 기반 시각화를 지원
- 2. 데이터 업데이트 자동화 서울시 API와 범죄 통계 파일의 주기적 업데이트 스크립트를 작성
- 추가 통계 계산 범죄율 외의 유의미한 통계 값을 추가 (예: 범죄 증감률).

7. 부록 ∂

• CrimeStats 테이블

컬럼명	설명
address_id	주소 ID (외래키)
reference_date	기준 날짜
total_population	총 인구
crime_category	범죄 대분류
crime_subcategory	범죄 중분류
incident_count	사건 발생 건수
crime_rate	인구 10만 명당 범죄율

• Address 테이블

컬럼명	설명
address_id	주소 ID
area_name	구 이름
latitude	위도
longitude	경도