

서울시 지하철역 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 호선 정보 정제
 - 4.3 주소 데이터 처리
 - 4.4 호선 데이터 처리
 - 4.5 데이터 저장
- 5. 결과 및 요약
- 6. 향후 작업
- 7. 부록

1. 목적 및 배경

서울시 지하철역 데이터를 수집하고 정제하여, 데이터베이스에 저장 후 시각화 및 분석을 위한 준비를 목표
특히 각 지하철역의 위치 정보와 호선 정보를 관리하여 교통 접근성 분석에 활용

2. 데이터 개요

2.1 데이터 소스

- ./data/지하철/seoul_subway_stations.csv

2.2 주요 컬럼

컬럼명	설명
역이름	지하철역 이름
호선정보	지하철 호선 정보 (쉼표로 구분된 문자열)
위도	지하철역 위도
경도	지하철역 경도

2.3 대상 데이터

서울특별시의 모든 지하철역과 해당 역이 포함된 호선 정보를 포함

3. 전처리 목표

1. 주소 데이터 처리

지하철역 이름과 좌표 데이터를 기반으로 Address 테이블에 데이터를 저장.

2. 호선 데이터 정제

각 역의 호선 정보를 정리하여 SubwayStation 테이블에 저장.

3. 중복 데이터 방지

기존 데이터와 중복되지 않도록 관리.

4. 전처리 단계

4.1 데이터 로드

- 작업
 - 지하철역 데이터가 포함된 CSV 파일을 읽어 데이터프레임으로 로드.

4.2 호선 정보 정제

- 작업
 - 호선정보 컬럼의 쉼표로 구분된 문자열을 리스트로 변환.

4.3 주소 데이터 처리

- 작업
 - 역이름, 위도, 경도 정보를 기준으로 Address 테이블에 저장. 좌표 정보가 없는 기존 주소는 업데이트.

4.4 호선 데이터 처리

- 작업
 - 각 역의 호선 정보를 SubwayStation 테이블에 저장.

4.5 데이터 저장

- 작업
 - 정제된 데이터를 데이터베이스에 저장.

5. 결과 및 요약

- 처리 전 데이터
 - 서울시 지하철역 이름, 좌표, 호선 정보를 포함한 CSV 파일.
- 처리 후 데이터
 - Address 테이블에 지하철역 이름과 좌표 데이터 저장.
 - SubwayStation 테이블에 각 역의 호선 정보 저장.
 - 총 XX개의 지하철역 데이터와 관련 호선 정보를 데이터베이스에 적재.

6. 향후 작업

- 추가적인 지하철 데이터(운행 시간, 출입구 정보 등) 통합.
- 지하철역을 중심으로 한 접근성 분석 및 시각화.
- 실시간 데이터와의 통합을 통해 교통 혼잡도 예측 모델 개발.

7. 부록

- Address 테이블

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

- SubwayStation 테이블

컬럼명	설명
station_id	지하철역 ID
address_id	주소 ID (외래키)
line_info	호선 정보