

라인웍스 채용 과제를 위한 데이터

Synthea를 이용하여 가상으로 생성한 EMR 데이터를 Common Data Model(이하 CDM)로 변환한 데이터가 제공됩니다.

그 중 일부 테이블만 제공되며, 제공되는 목록은 다음과 같습니다.

- person
- visit_occurrence
- condition_occurrence
- drug_exposure
- death
- concept

기본 설명

제공되는 데이터를 말로 풀어 설명하면

환자(person)가 **방문(visit_occurrence)**을 하고, 해당 방문에서 어떤 **진단(condition_occurrence)**과 **처방(drug_exposure)**을 받았는지, 그리고 사망을 했다면 언제 **사망(death)**했는지에 대한 정보

입니다.

각 테이블간을 연결하기 위해 중요하게 사용되어야할 공통 Key는 person_id(환자 id)와 visit_occurrence_id(방문 id)입니다.

CDM은 concept라는 개념을 사용하고 있습니다. 대부분의 카테고리(범주)형 자료들은 concept_id로 표현이 되어있습니다. 이에 대한 정보(이름 등)를 concept 테이블이나 그와 매칭되는 source_value 컬럼을 확인해야합니다.

CDM에 대한 더 자세한 설명은 다음 링크를 참조 바랍니다.

<https://github.com/OHDSI/CommonDataModel/wiki>

테이블 별 설명

테이블 별로 해당 테이블이 어떤 정보를 포함하고 있는지에 대해 설명합니다. 우선적으로 확인해야할 컬럼에 설명만 작성하였습니다. 모든 컬럼에 대한 정보는 위 CDM 상세 설명 링크 참조 바랍니다.

person

환자에 대한 정보를 담고 있습니다.

- person_id : 환자 id
- gender_concept_id : 성별
- birth_datetime : 생년월일
- race_concept_id : 인종
- ethnicity_concept_id : 민족

visit_occurrence

방문에 대한 정보를 담고 있습니다.

- visit_occurrence_id : 방문 id
- person_id : 환자 id
- visit_concept_id : 방문 유형
 - 9201 : Inpatient Visit (입원)
 - 9202 : Outpatient Visit (외래)
 - 9203 : Emergency Room Visit (응급)
- visit_start_datetime : 방문 시작 일시
- visit_end_datetime : 방문 종료 일시

condition_occurrence

진단(병명)에 대한 정보를 담고 있습니다.

- person_id : 환자 id
- condition_concept_id : 진단(병명)
- condition_start_datetime : 진단 시작 일시
- condition_end_datetime : 진단 종료 일시
- visit_occurrence_id : 방문 id

drug_exposure

의약품 처방에 대한 정보를 담고 있습니다.

- person_id : 환자 id
- drug_concept_id : 처방 의약품
- drug_exposure_start_datetime : 처방 시작 일시
- drug_exposure_end_datetime : 처방 종료 일시
- visit_occurrence_id : 방문 id

concept

여러 테이블에서 사용되는 concept들의 정보를 담고 있습니다.

- concept_id : concept id
- concept_name : concept 이름
- domain_id : concept가 주로 사용되는 도메인(카테고리)

death

환자의 사망 정보를 담고 있습니다.

- person_id : 환자 id
- death_date : 사망일