

파일	열	참고(조합해서 만든 변수)
patients.csv	isMale Anchor_age (real, group) dod (real, 존재 유무?)	
admissions.csv	race	def caculate_eGFR_MDRD_equation(cr, gender, eth, age): temp = 186 * (cr ** (-1.154)) * (age ** (-0.203)) if (gender == 'F'): temp = temp * 0.742 if eth == 'BLACK/AFRICAN AMERICAN': temp = temp * 1.21 return temp
labevents.csv	Creatinine WBC Hemoglobin Platelet Albumin sodium Potassium Chloride AST ALT BUN Total CO2 Bilirubin Calcium Glucose CK (3 cate) Lipase (3 cate) Troponin I (3 cate) eGFR	하루에 평균 1번 실시했지만, 하루 내에서는 크게 변동하지 않을 것으로 생각되어 하루에 여러번 한 환자는 그 횟수 상관없이 평균값으로 처리하면 좋을 것 같음  굳이 이 많은 걸 다 사용할 필요는 없을 것 같고, 상관관계를 보고 대표성을 지닌 것을 선택하거나 많이 검사한 상위 N개를 선택해 사용하면 좋을 것 같음
omr.csv	Weight Height Blood pressure (systolic, diastolic 나뉘져 있다고 하는데, 그냥 mean이 있으면 mean 쓰면 좋을 것 같음)	Bmi (weight, height로 가공) : bmi = weight / (height/100)**2
outputevents.csv		
chartevents.csv		

상위 10개 검사항목: lab

Item ID	검사 항목명	수행 횟수	설명
50912	Creatinine	1,111,185	신장 기능을 평가하기 위해 혈액에서 크레아티닌 수치를 측정하는 검사
50971	Potassium	1,099,162	혈액 내 칼륨 수치를 측정하여 전해질 균형 및 심장 건강 상태를 평가
51006	Urea Nitrogen	1,091,613	혈액 요소 질소(BUN)를 측정하여 신장 기능 및 대사 상태 평가
50983	Sodium	1,087,546	혈액 내 나트륨 수치를 측정하여 체액 및 전해질 균형 상태 평가
50902	Chloride	1,080,489	혈액 내 염화물 수치를 측정하여 산염기 균형 및 전해질 상태 평가
51221	Hematocrit	1,062,627	혈액 내 적혈구의 비율을 측정하여 빈혈 및 탈수 상태를 평가
50882	Bicarbonate	1,042,115	혈액 내 중탄산염 수치를 측정하여 산염기 균형 상태 평가
50868	Anion Gap	1,038,929	음이온 차이를 계산하여 산염기 상태 이상 평가
51265	Platelet Count	1,022,936	혈액 내 혈소판 수치를 측정하여 응고 기능 및 혈액 질환 평가
51222	Hemoglobin	1,010,467	혈액 내 헤모글로빈 수치를 측정하여 빈혈 상태를 평가

Vitals:

Mean blood pressure(있는지 확인 필요)

pulse rate(heart rate 있는지 확인 필요)

body temperature

Comorbidities기저질환: 유무로 01 coding OR 해당 기저질환이 몇 개 있는지 개수로 coding

Ischemic heart disease

Hypertension

Diabetes mellitus

Heart failure

Liver disease

Cerebrovascular disease

Chronic obstructive pulmonary disease

Chronic kidney disease

Previous AKI

cancer

HIV

CCI

medication: 유무로 01 coding OR 해당 기저질환이 몇 개 있는지 개수로 coding

NSAIDs

Diuretics

Acyclovir

Aminoglycoside

Amphotericin

Beta blocker

Calcium channel blocker  
Cisplatin  
Vancomycin  
Colistin  
Cyclosporin  
Renin angiotensin blocker  
vasopressors  
Statins

Departments (unique 개수 확인) OR ICU admissions (unique 개수 확인) → 둘 중 하나만 넣어도 될 것 같음

Mechanical ventilation(current/recent)

Surgery 유무

Surgery duration: 수술 시간

anesthesia(general or not)

- [procedures\\_icd.csv.gz](#) (수술 관련 정보)
- [chartevents.csv.gz](#) (수술 전후 상태 기록)
- [surgical\\_vitalsigns.csv](#) (특정 수술 관련 바이탈 사인 기록)

주진단, 부진단 (이건 어떤 파일에 있는지 확인 필요)

=> 애넌 어케 처리?


[ 완료한 것들 ]

- Stage별 AKI 진단 (0, 1, 2, 3)
  - 코드: AKI\_Cr\_Def.ipynb
  - 파일: aki\_results\_stage.csv
- BMI에 대한 건 이미 존재 (BMI/BMI 타임스탬프(초))
  - 코드: process\_omr.ipynb
  - 파일: omr.csv
- Comorbidities는 0과 1로 나누어야 하는 게 저 모든 리스트의 값을 각각 하면 컬럼 양이 너무 많아져서 어떻게 할까 상의 후 결정
  - 코드: process\_comorbidities.ipynb

- 파일:
- Medication도 이하동문
  - 코드: process\_medication.ipynb
- Vitals 세 항목 모두 존재, 모아놓음
  - 코드: process\_vital.ipynb
  - 파일: vitals\_chartevents.csv

#### [ 해야 할 것 ]

- Departments (unique 개수 확인) OR ICU admissions (unique 개수 확인) 둘 중 하나
- 0/1 어케 할 건지
- 피쳐 중요도 계산

 data.ipynb

## 1. ICU 관련 핵심 파일들 (mimic/icu 폴더)

이 파일들은 주로 중환자실(ICU)에서의 환자 데이터를 다룹니다.

- **chartevents.csv.gz:**
  - 가장 중요한 파일 중 하나로, 환자의 생체 징후 및 상태를 시간대별로 기록한 파일입니다.
  - 혈압, 심박수, 체온, 산소 포화도 등 환자의 실시간 데이터를 수집합니다.
- **icustays.csv.gz:**
  - 환자의 중환자실(ICU) 입원 정보를 기록한 파일로, 환자가 언제 ICU에 입원했고, 얼마나 머물렀는지에 대한 정보를 제공합니다.
- **inputevents.csv.gz:**
  - 환자에게 투여된 수액, 약물 등 입력 데이터가 기록되어 있습니다.
- **outputevents.csv.gz:**

- 환자의 배설(예: 소변량, 체액 배출 등)과 같은 출력 데이터를 포함합니다.
- **datetimeevents.csv.gz:**
  - 특정 시점에 발생한 이벤트(예: 검사, 투약 시간 등)를 기록한 파일입니다.
- **d\_items.csv.gz:**
  - ICU에서 사용되는 다양한 항목들의 **itemid**와 그 설명(**label**)을 포함한 파일입니다. 각 **itemid**는 생체 징후나 처치 항목을 정의합니다.

## 2. 병원 관련 핵심 파일들 (mimic/hosp 폴더)

이 파일들은 병원 전체에서 기록된 데이터를 포함하며, ICU 외에도 일반 병동에서의 기록도 다룹니다.

- **admissions.csv.gz:**
  - 환자의 병원 입원 정보가 기록된 파일입니다. 환자가 언제 병원에 입원했고, 퇴원 또는 사망했는지에 대한 정보를 제공합니다.
- **patients.csv.gz:**
  - 환자의 인구통계학적 정보(나이, 성별, 생년월일 등)가 기록된 파일입니다.
- **diagnoses\_icd.csv.gz:**
  - 환자의 진단 정보를 ICD(국제질병분류) 코드로 기록한 파일입니다. 각 환자가 입원 기간 동안 받은 진단을 포함하고 있습니다.
- **procedures\_icd.csv.gz:**
  - 환자가 받은 시술 및 수술 기록을 ICD 코드로 기록한 파일입니다.
- **prescriptions.csv.gz:**
  - 병원 내에서 환자에게 처방된 약물에 대한 정보를 포함한 파일입니다. 처방 시점과 약물의 용량, 경로 등이 기록되어 있습니다.
- **labevents.csv.gz:**
  - 환자의 혈액 검사, 소변 검사 등의 실험실 검사 결과가 기록된 파일입니다. 검사 항목, 결과 값, 단위 등이 포함됩니다.

## 3. 기타 유용한 파일들

- **services.csv.gz:**
  - 환자가 입원 중 어느 의료 서비스(예: 내과, 외과 등)를 받았는지에 대한 기록입니다.
- **transfers.csv.gz:**
  - 병원 내에서 환자의 병동 이동 정보를 기록한 파일입니다.

정리

MIMIC-IV에서 가장 중요한 파일들은 다음과 같습니다:

1. **chartevents.csv.gz:** 환자의 생체 징후와 상태 기록 (가장 방대한 데이터)
2. **icustays.csv.gz:** ICU 입원 정보
3. **admissions.csv.gz:** 병원 입원 정보

4. **patients.csv.gz**: 환자의 인구통계 정보
5. **diagnoses\_icd.csv.gz**: 환자의 진단 기록
6. **procedures\_icd.csv.gz**: 환자의 시술 및 수술 기록
7. **labevents.csv.gz**: 실험실 검사 결과
8. **prescriptions.csv.gz**: 약물 처방 정보