

[TIL] 24/12/03_MIMIC III 주제 관련 지표 분석

의사 처방 패턴 분석 → 업무 자동화

- 특정 질환에 집중해서 패턴 분석을 해도, 비즈니스적인 지표가 잘 나올지 여부에 대해서 분석
- 사용할 테이블 정리
 - ADMISSIONS
 - PATIENTS
 - DIAGNOSES_ICD
 - PRESCRIPTIONS
 - (ICUSTAYS)
 - (LABEVENTS)

- 노트북

<https://colab.research.google.com/drive/14nQgj3IBEGxZGRE5VJNl6wUH5EWPPxCf>

- 지표
 - 결측치는 빠른 분석을 위해 모두 drop

```
[ ] 1 admissions_df.dropna(inplace=True)
    2 patients_df.dropna(inplace=True)
    3 diagnoses_df.dropna(inplace=True)
    4 prescriptions_df.dropna(inplace=True)
```

- 고혈압 데이터(401)를 중심으로 필터링하여 처방 데이터와 연결(+환자, 입원 데이터)

```
[ ] 1 # 특정 질환 (고혈압) 데이터 필터링 -> 401이 고혈압 분류 코드
    2 target_disease = diagnoses_df[diagnoses_df['ICD9_CODE'].str.startswith('401')]
    3 target_patients = target_disease[['HADM_ID', 'SUBJECT_ID']].drop_duplicates()

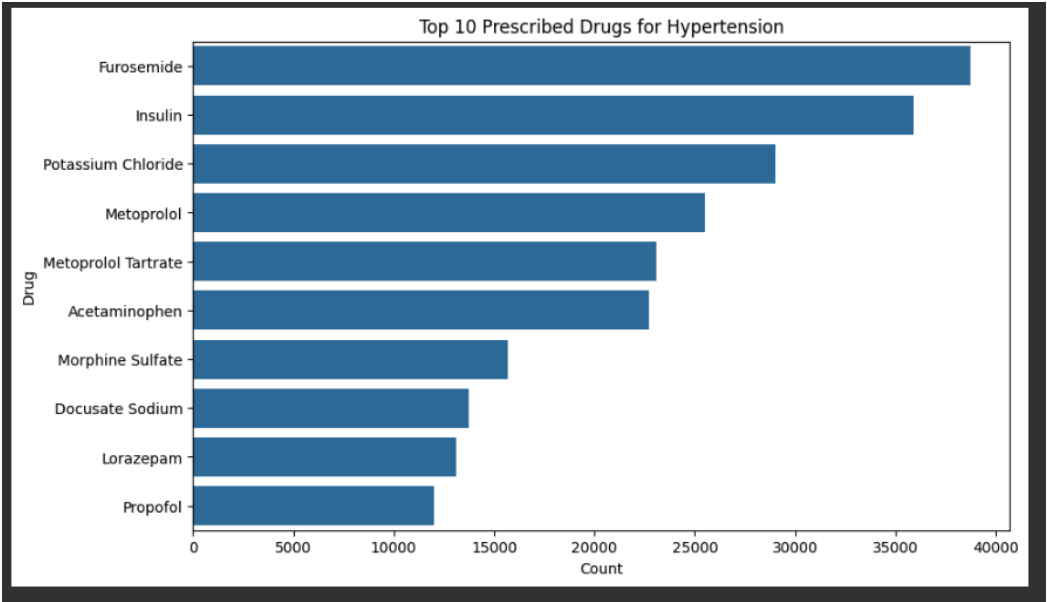
[ ] 1 # 처방 데이터와 연결
    2 target_prescriptions = prescriptions_df.merge(target_patients, on='HADM_ID', how='inner')

[ ] 1 target_prescriptions.head(5)
```

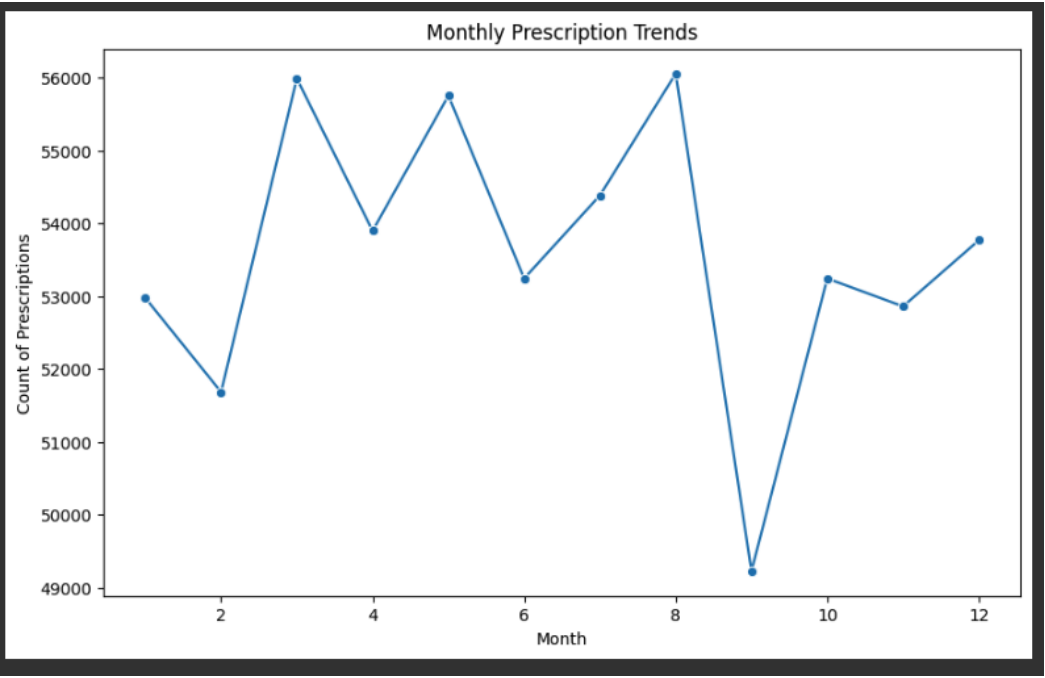
```
[ ] 1 # 환자 및 입원 데이터 추가 연결
    2 target_data = target_prescriptions.merge(admissions_df, on='HADM_ID', how='left')
    3 target_data = target_data.merge(patients_df, on='SUBJECT_ID', how='left')

[ ] 1 target_data.head(5)
```

- EDA
 - 처방 패턴 분석



- 처방 시간 분석



- FN
 - 환자 나이, 성별, 입원 기간 + 약물 투약 빈도 및 용량

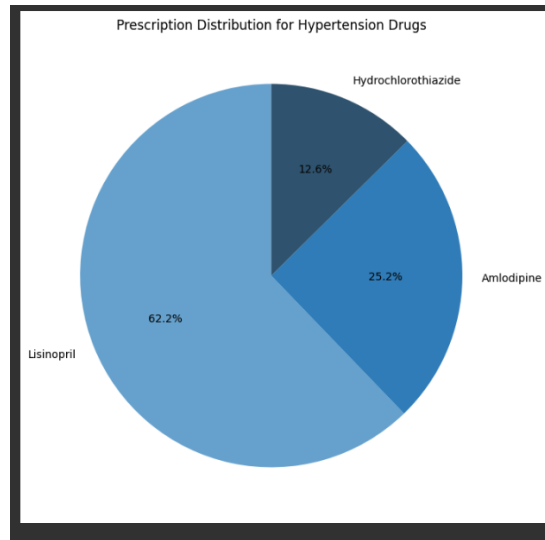
```
[ ] 1 target_data['ADMITTIME'] = pd.to_datetime(target_data['ADMITTIME'], errors='coerce')
2
3 # 나이 계산
4 target_data['AGE'] = target_data['ADMITTIME'].apply(lambda x: 2024 - x.year if pd.notnull(x) else None)
5
6 # DOSE_VAL_RX : 숫자 변환
7 target_data['DOSE_VAL_RX'] = pd.to_numeric(target_data['DOSE_VAL_RX'], errors='coerce')
8
9 # 약물 투약 빈도 및 용량
10 drug_features = target_data.groupby(['SUBJECT_ID', 'DRUG'])['DOSE_VAL_RX'].agg(['mean', 'count']).reset_index()
11 drug_features.rename(columns={'mean': 'Average_Dosage', 'count': 'Frequency'}, inplace=True)
12
13 # 입원 기간
14 target_data['DISCHTIME'] = pd.to_datetime(target_data['DISCHTIME'], errors='coerce')
15 target_data['LOS'] = (target_data['DISCHTIME'] - target_data['ADMITTIME']).dt.days
```

- 이전 투약 기록

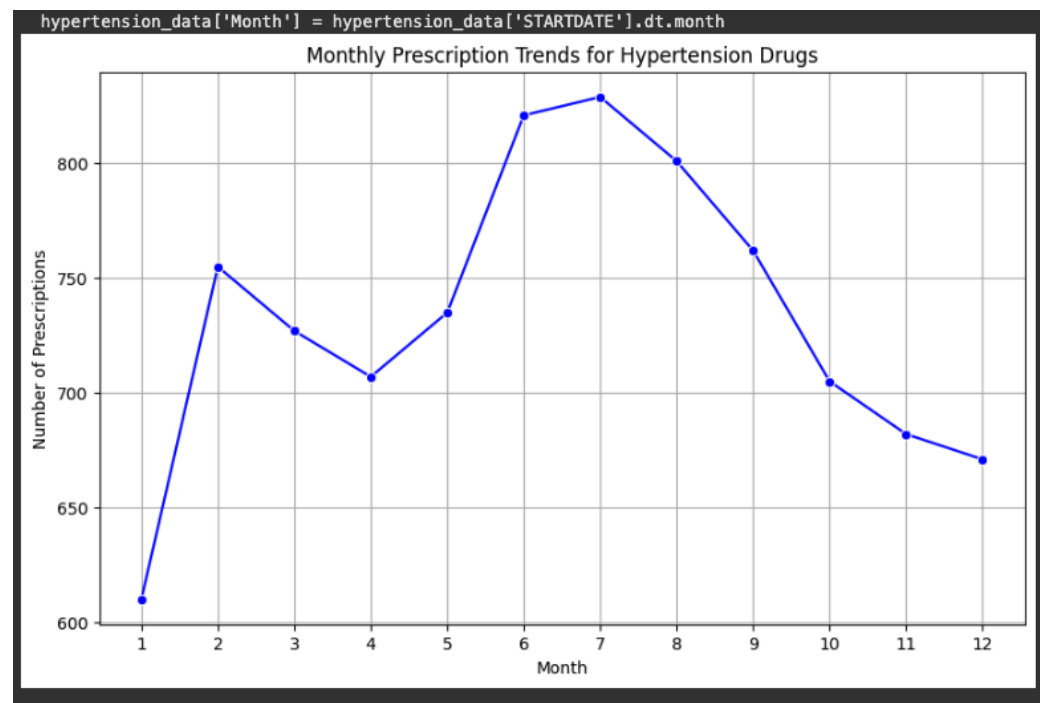
```
[ ] 1 # 이전 투약 기록 확인 (PRESCRIPTIONS와 ADMISSIONS 병합)
2 # 처방 데이터에 대한 날짜 변환
3 prescriptions_df['STARTDATE'] = pd.to_datetime(prescriptions_df['STARTDATE'])
4 admissions_df['ADMITTIME'] = pd.to_datetime(admissions_df['ADMITTIME'])
5
6 # 데이터 병합 (SUBJECT_ID와 HADM_ID를 기준으로)
7 merged_data = pd.merge(prescriptions_df, admissions_df, on=['SUBJECT_ID', 'HADM_ID'], how='inner')
8
9 # 이전 투약 기록 필터링 (STARTDATE가 ADMITTIME보다 이전인 경우)
10 previous_medications = merged_data[merged_data['STARTDATE'] < merged_data['ADMITTIME']]
11
12 # 'DOSE_VAL_RX' 열을 숫자형으로 변환 (문자열이 있을 경우 NaN으로 처리)
13 previous_medications['DOSE_VAL_RX'] = pd.to_numeric(previous_medications['DOSE_VAL_RX'], errors='coerce')
14
15 # 주요 컬럼 선택
16 previous_medications = previous_medications[['SUBJECT_ID', 'DRUG', 'STARTDATE', 'DOSE_VAL_RX']]
17
18 # 이전 투약 기록과 주요 분석 데이터 병합
19 target_data = pd.merge(
20     target_data,
21     previous_medications.groupby(['SUBJECT_ID', 'DRUG']).agg(
22         Previous_Frequency=('STARTDATE', 'count'),
23         Previous_Avg_Dosage=('DOSE_VAL_RX', 'mean')
24     ).reset_index(),
25     on=['SUBJECT_ID', 'DRUG'],
26     how='left'
27 )
28
29 # 결측치 처리: 이전 투약 기록이 없는 경우 0으로 대체
30 target_data['Previous_Frequency'].fillna(0, inplace=True)
31 target_data['Previous_Avg_Dosage'].fillna(0, inplace=True)
```

- 모델링
 - XGBoost 이용 → 약물 정보에 대한 라벨 인코딩이 필요하여 진행했으나 훈련 데이터와 실제 데이터 간의 불일치 정도가 높아 모델링 불가.

- 시각화
 - 고혈압 약물 처방 빈도

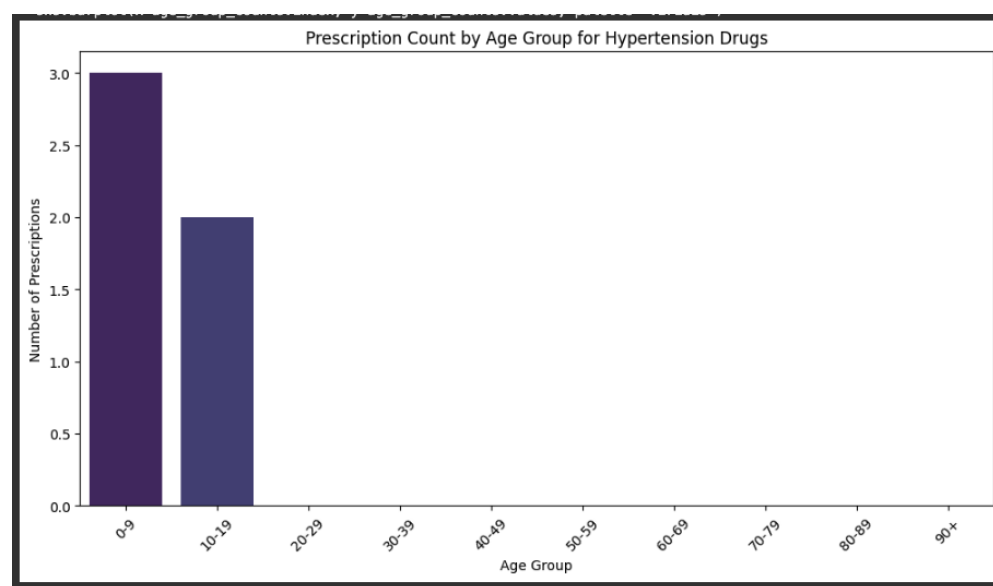


○ 고혈압 처방 시기 월별 패턴



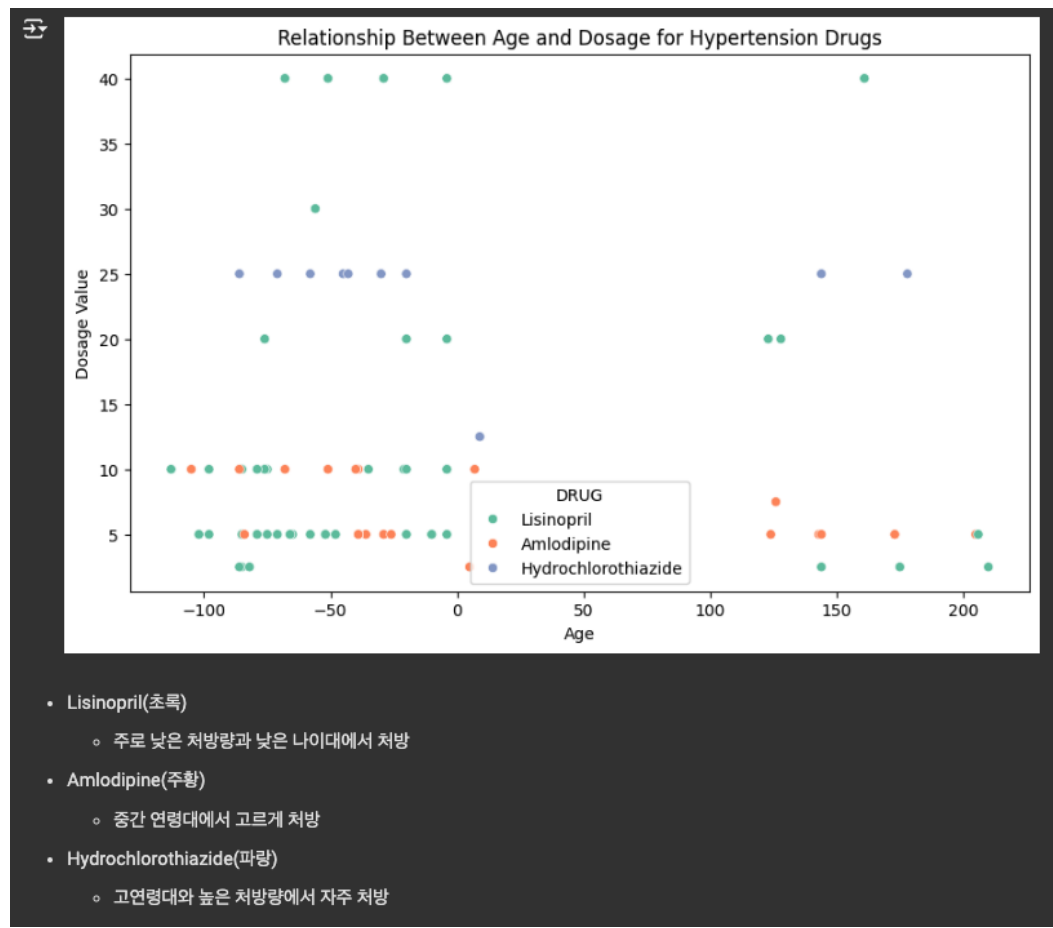
○ 고혈압 관련 약물 연령대별 처방

- 고령일수록 처방을 많이 받을 것으로 예상 → 반대의 결과가 나옴

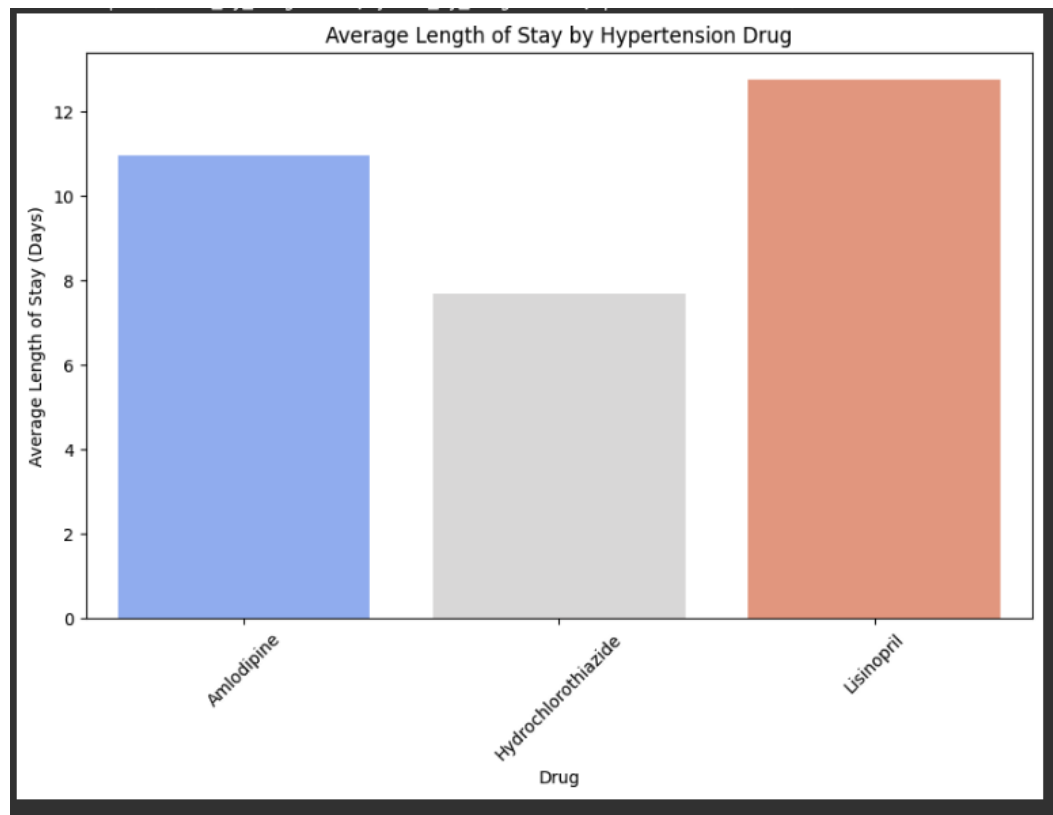


○ 나이와 처방량 간 관계

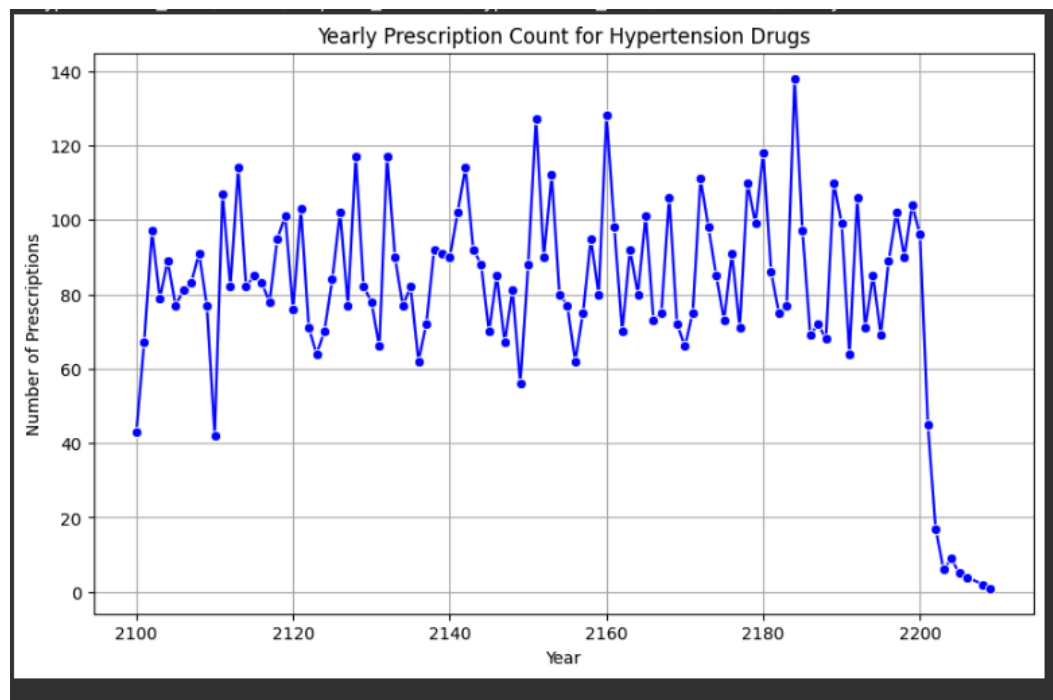
- 나이가 많을수록 고혈압 약물 처방량은 증가(처방 빈도와 처방량은 다름) + 일부 이상치가 있음



◦ 고혈압 관련 약물 처방에 따른 환자별 입원 기간



◦ 연령대와 성별에 따른 처방 빈도 계산



- 결론
 - 만약에 이러한 처방 패턴 분석을 한다면 다음의 정리가 필요함.
 - ICD-9 코드 정리
 - 질환 종류 파악 및 특정 질환 선정(거의 모든 질환을 해야하지 않을까 하는 문제..)
 - 그렇기 때문에, 특정 질환 지표보다는 한다면 질환은 전체로 전제를 두고, 나이대나 성별과 같은 환자의 다른 지표를 기준으로 하여 처방 패턴을 분석하는 것이 비즈니스 적으로 유의미한 결과가 나올 것으로 판단

ICD-9 컬럼 분석 및 필터링

- 테이블
 - D_ICD_DIAGNOSES
 - DIAGNOSES_ICD
 - PROCEDURES_ICD

- 노트북

https://colab.research.google.com/drive/10GyrmQiNFWA_3HoS1ik2_CKAHxUOMdct?usp=sharing

- 3가지 테이블의 ICD9_CODE 컬럼을 분석해 주요 지표 외 제외 시도
 - ICD-9 코드 목록 : https://ko.wikipedia.org/wiki/ICD-9_%EC%BD%94%EB%93%9C_%EB%AA%A9%EB%A1%9D

- 지표
 - E : 외상, V : 예방적 차원에서 치료(예방 접종, 건강 검진 등) → 제외
 - 9천개 이상에 속하는 지표만 사용하는 것이 좋아보임
 - 고혈압(4019) : 순환계
 - 심부전(4280) : 순환계
 - 부정맥(42731) : 순환계
 - 관상동맥질환(41401) : 순환계
 - 당뇨병(25000) : 내분비, 영양 및 대사 질환, 면역 장애

```
[ ] 1 icd9_counts_dicd = diagnoses_icd_df['ICD9_CODE'].value_counts()
    2 print(icd9_counts_dicd.head(40))

ICD9_CODE
4019    20703
4280    13111
42731    12091
41401    12429
5840     9119
25000     9058
2724     8690
51881     7497
5990     6555
53001     6326
2720     5930
V053     5779
V290     5519
2859     5406
2449     4917
486      4839
2051     4592
2762     4528
496      4431
99592     3912
V5861     3806
0389     3725
5070     3680
V3000     3566
5859     3435
311      3431
40390     3421
3051     3358
412      3278
2875     3065
V4581     3056
41071     3055
2761     3039
4240     2926
V1582     2811
V3001     2758
5119     2734
V4582     2725
40391     2630
78552     2586
Name: count, dtype: int64
```

- 이 테이블의 경우, ICD-9 코드의 챕터 3장에 해당하는 기타 치료로 대부분 분류됨
 - 호흡 치료, 정맥 주사, 수액 주입, 기계적 환기, 수혈 등

```
[ ] 1 icd9_counts_pcd = procedures_icd_df['ICD9_CODE'].value_counts()
    2 print(icd9_counts_pcd.head(40))
```

ICD9_CODE	
3893	14731
9604	10333
966	9300
9671	9100
9904	7244
3961	6838
9672	6048
9955	5842
8856	5337
3891	4737
3615	4401
9915	4244
8872	3548
3722	3311
3324	3269
3995	3254
4513	2919
9390	2727
3723	2711
9983	2402
5491	2217
331	2200
640	2123
8853	2113
9907	2028
3612	1903
3491	1834
3895	1825
311	1776
9920	1740
3404	1717
3897	1681
40	1656
8841	1579
4311	1536
3613	1504
3521	1419
66	1330
3322	1264
3606	1261

Name: count, dtype: int64

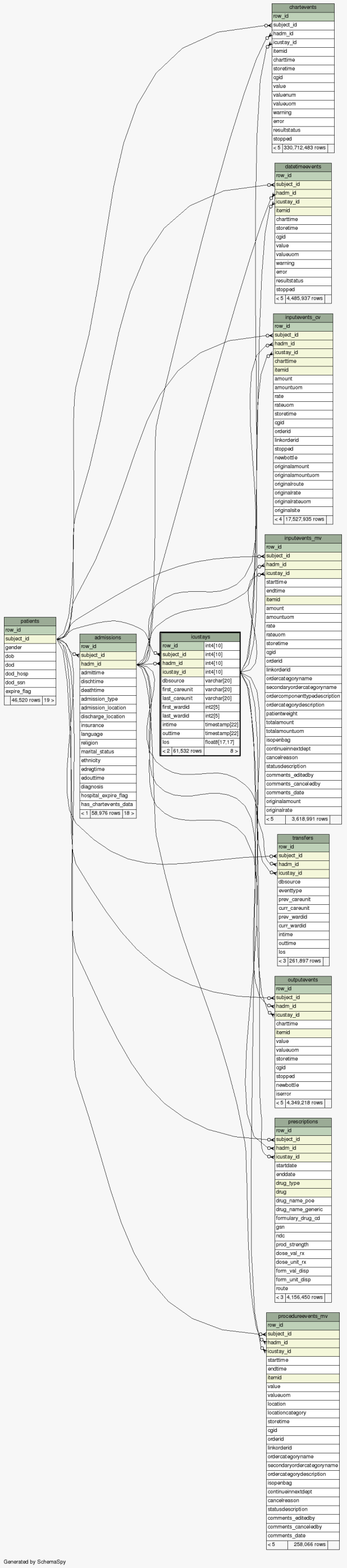
ITEM_ID 정리

- 일반적인 수술 시 사용하는 약물도 다수 포함되어 있어, 패턴 파악이 어려울 경우 → 매칭 작업 필요

[merged_data_v2.csv](#)

	A	B	C
	merged_data_v2		
1	ITEMID	COUNTS	LABEL
2	30013	2557507	D5W
3	30018	2392372	.9% Normal Saline
4	30131	924614	Propofol
5	30045	825758	Insulin
6	30025	813242	Heparin
7	30118	780555	Fentanyl
8	30128	554582	Neosyneprine-k
9	30124	505509	Midazolam
10	30120	476971	Levophed-k
11	30140	373023	
12	30026	291669	KCL
13	30112	259281	Amiodarone
14	30121	239268	Nitroglycerine-k
15	30187	227011	D10W
16	30021	214893	Lactated Ringers
17	30032	204714	TPN
18	30015	200221	D5/.45NS
19	30208	194741	Lipids 20%
20	30123	193403	Lasix
21	30056	188107	Po Intake
22	30043	173745	Dopamine
23	30141	173510	Ativan
24	30166	167025	Promote w/fiber
25	30303	165273	MCT
26	30051	165219	Vasopressin
27	30221	153406	Br Milk FS NG
28	30089	147518	Promod
29	30073	147333	Impact w/fiber
30	30144	141104	Gastric Meds
31	30302	137887	HMF
32	30125	132751	Milrinone
33	30196	129029	TPN D10.0
34	30050	112121	Nitroprusside
35	30296	98887	Sodium Chloride
36	30126	97439	Morphine Sulfate
37	30095	94752	TF Residual
38	30077	94128	Nepro
39	30297	85820	Potassium Chloride
40	30119	82889	Epinephrine-k
41	30090	78707	Tube Feeding
42	30304	75194	Polycose
43	30020	72177	.45% Normal Saline
44	30260	72082	Br Milk FS PO
45	30042	66775	Dobutamine

- 결론
 - chartevents 테이블을 제외한 테이블을 활용해 지표 분석하는 것으로 결론
 - 처방 약물 데이터를 활용



오늘의 결론



다제내성균

- 아래 내용에 따르면 메티실린으로 항생제가 국한되어 있는 것으로 보일 수 있으나, 원론적인 측면에서는 모든 항생제가 포함됨!

다제내성균 감염에 대한 학술자료

다제내성균과 의료관련감염 - 김미나 - 6회 인용

중환자실 간호사의 다제내성균 감염관리 수행에 영향을 ... - 김지희 - 6회 인용

종합병원 간호사의 다제내성균 감염관리 수행도에 영향 ... - 정림 - 5회 인용

사람의 피부나 구강인후 점막의 상재균인 황색포도알균 중 메티실린이라는 강력한 항생제에 내성을 나타내는 메티실린 내성 황색포도알균에 감염되면서 화농성 염증, 식중독, 패혈증 등 다양한 감염증으로 나타납니다.

국가건강정보포털
<https://health.kdca.go.kr> > health > gnrizHealthInfoView

다제내성균감염증 - 국가건강정보포털 - 질병관리청

<표> 다제내성능균(MDR) 국내현황

다제내성능균(MDR) 국내현황								
구분	2013년	2014년	2015년	2016년	2017년	2018년	2019년	
발생 건수	0.04	0.04	0.04	0.04	0.04	0.04	0.04	
발생 위 중재	0.04	0.02	0.06	0.07	0.07	0.06	0.06	

자료: 질병관리청, 다제내성능균(MDR) 국내현황
발생 건수: 2013년 0.04건, 2014년 0.04건, 2015년 0.04건, 2016년 0.04건, 2017년 0.04건, 2018년 0.04건, 2019년 0.04건
발생 위 중재: 2013년 0.04건, 2014년 0.02건, 2015년 0.06건, 2016년 0.07건, 2017년 0.07건, 2018년 0.06건, 2019년 0.06건
자료: 질병관리청, 다제내성능균(MDR) 국내현황

- 중환자실에서의 다제내성균(MDR) 감염 치료 효과 분석
 - 주제 설명: 다제내성균 감염 환자의 항생제 사용 실태와 치료 성공률 평가
특정 항생제의 사용이 치료 성공률에 미치는 영향을 분석
 - 활용 테이블
 - DIAGNOSES_ICD: MDR 감염 관련 진단 코드
 - PRESCRIPTIONS: 항생제 처방 데이터
 - LABEVENTS: 감염 지표(예: CRP, WBC) 변화
 - 모델링 방향
 - 입력: 항생제 사용량 + 실험실 검사 데이터
 - 출력: 감염 치료 성공 여부(0=실패, 1=성공)
 - 알고리즘: 로지스틱 회귀, XGBoost
- 위의 주제로 진행하기 위해 필요한 전제 조건 확인
 - 항생제가 1개일 경우, 진행 불가 → 감염지표, 항생제의 종류, 약물 사용 빈도수까지 파악하여 분석이 가능할지 여부에 대해서 파악해야 함.