

1) ADMISSIONS.csv

RangeIndex: 129 entries, 0 to 128

Data columns (total 19 columns):

#	Column	Non-Null Count	Dtype
0	row_id	129 non-null	int64
1	subject_id	129 non-null	int64
2	hadm_id	129 non-null	int64
3	admittime	129 non-null	object
4	dischtime	129 non-null	object
5	deathtime	40 non-null	object
6	admission_type	129 non-null	object
7	admission_location	129 non-null	object
8	discharge_location	129 non-null	object
9	insurance	129 non-null	object
10	language	81 non-null	object
11	religion	128 non-null	object
12	marital_status	113 non-null	object
13	ethnicity	129 non-null	object
14	edregtime	92 non-null	object
15	edouttime	92 non-null	object
16	diagnosis	129 non-null	object
17	hospital_expire_flag	129 non-null	int64
18	has_chartevents_data	129 non-null	int64

dtypes: int64(5), object(14)
memory usage: 19.3+ KB

ID	Sütun Adı	Anlamı	Proje Bağlantısı
row_id	Satır numarası (tekil ID)	Teknik amaçlı, modelde kullanılmaz	
subject_id	Hastaya özel benzersiz ID	Diğer tablolarda hasta bazlı birleştirme için kullanılır	
hadm_id	Hastane yatışı ID'si	ICUSTAYS , DIAGNOSES_ICD vb. tablolarla eşleştirmek için kullanılır	
admittime	Hastanın hastaneye yatış zamanı	Zaman serisi oluşturma, giriş süresi analizleri	
dischtime	Taburcu olma zamanı	Kalış süresi hesaplaması için gereklidir	
deathtime	Ölüm zamanı (varsa)	Hedef etiket (label) oluşturma için altın veridir	
admission_type	Yatış tipi (ACİL, ELEKTİF vb.)	Risk seviyesini etkileyebilecek ön bilgi	
admission_location	Hastanın geldiği yer (acil, transfer vb.)	İlk risk değerlendirmesi için önemli olabilir	
discharge_location	Taburcu olduğu yer (ev, bakım evi, ölmüş vb.)	Nihai durumu gösterir, risk sınıflandırmasında faydalı	
insurance	Sigorta türü (Medicare, Private vb.)	Sosyoekonomik etkenlerin analizinde kullanılabilir	
language	Hastanın konuştuğu dil	Çoğunlukla boş, ihmal edilebilir	
religion	Dini inancı	Genellikle modelde kullanılmaz	
marital_status	Medeni hali	Opsiyonel olarak sosyodemografik analizlerde değerlendirilebilir	
ethnicity	Etnik köken	Model için dikkatle kullanılmalı; etik sınırdadır	
edregtime	Acil servise kayıt zamanı	admittime ile karşılaştırılıp bekleme süresi hesaplanabilir	
edouttime	Acilden çıkış zamanı	Aynı şekilde klinik akış süresi ölçümü	
diagnosis	İlk tanı (serbest metin: örn. SEPSIS)	Ana etiketleme ve sepsis pozitif hasta belirlemede doğrudan kullanılır	
hospital_expire_flag	Hastanede ölüp ölmediği (1: evet, 0: hayır)	Modelin target sütunu olarak direkt kullanılabilir	
has_chartevents_data	Bu yatışta vital bulgu verisi var mı (1: evet)	Veri yeterliliği kontrolü için kullanılır	

- Bu tablo, **hastanın hastane yatışını anlamlandırmak için temel tablodur**.
- Sepsis olup olmadığını belirlemek için "diagnosis" içinde "SEPSIS" kelimesi aranabilir.
- Ölüm gerçekleşip gerçekleşmediği doğrudan "hospital_expire_flag" ve "deathtime" ile belirlenebilir.
- "admittime" ve "dischtime" ile hastanede kalış süresi hesaplanarak model girdisi yapılabilir.
- "admission_type" gibi kategorik veriler, modelin girişinde etkili özelliklere dönüştürülebilir (feature engineering).

	Eksik Değer Sayısı	Oran (%)
deathtime	89	68.99
language	48	37.21
edregtime	37	28.68
edouttime	37	28.68
marital_status	16	12.40
religion	1	0.78

📊 ADMISSION_TYPE değeri dağılımı:

admission_type
EMERGENCY 119
ELECTIVE 8
URGENT 2
Name: count, dtype: int64

🎯 ADMISSION_LOCATION değeri dağılımı:
admission_location
EMERGENCY ROOM ADMIT 81
TRANSFER FROM HOSP/EXTRAM 25
CLINIC REFERRAL/PREMATURE 12
PHYS REFERRAL/NORMAL DELI 10
TRANSFER FROM SKILLED NUR 1
Name: count, dtype: int64

🎯 DISCHARGE_LOCATION değeri dağılımı:
discharge_location
DEAD/EXPIRED 40
SNF 39
HOME 15
HOME HEALTH CARE 14
REHAB/DISTINCT PART HOSP 13
ICF 3
LONG TERM CARE HOSPITAL 2
HOSPICE-HOME 1
DISCH-TRAN TO PSYCH HOSP 1
HOME WITH HOME IV PROVIDR 1
Name: count, dtype: int64

🎯 INSURANCE değeri dağılımı:
insurance
Medicare 98
Private 24
Medicaid 6
Government 1
Name: count, dtype: int64

🎯 MARITAL_STATUS değeri dağılımı:
marital_status
MARRIED 60
SINGLE 26
NaN 16
WIDOWED 15
DIVORCED 6
UNKNOWN (DEFAULT) 5
SEPARATED 1
Name: count, dtype: int64

🎯 ETHNICITY değeri dağılımı:
ethnicity

WHITE	86	
HISPANIC/LATINO - PUERTO RICAN		15
UNKNOWN/NOT SPECIFIED		11
BLACK/AFRICAN AMERICAN		7
OTHER	3	
ASIAN	2	
HISPANIC OR LATINO		2
AMERICAN INDIAN/ALASKA NATIVE FEDERALLY RECOGNIZED TRIBE		2
UNABLE TO OBTAIN		1

Name: count, dtype: int64

```
# İlk tanıya göre sepsis pozitif etiket (SEPSIS kelimesi geçenler)
df_adm['sepsis_label'] = df_adm['diagnosis'].str.contains("SEPSIS", case=False, na=False).astype(int)

# Etiket dağılımı
df_adm['sepsis_label'].value_counts(normalize=True)
```

```
: sepsis_label
0    0.891473
1    0.108527
Name: proportion, dtype: float64
```

```
3]: # Hastane içi ölüm bilgisi
df_adm['hospital_expire_flag'].value_counts()

# Gerçek ölüm zamanı varsa (deathtime dolu), farklı bir etiket daha:
df_adm['death_flag'] = df_adm['deathtime'].notnull().astype(int)
df_adm[['hospital_expire_flag', 'death_flag']].value_counts()
```

```
8]: hospital_expire_flag  death_flag
0                        0           89
1                        1           40
Name: count, dtype: int64
```

ADMISSIONS.csv: Kullanılabilirlik Özeti ve Model İçin Önerilen Öznitelikler

Genel Amaç


Bu tablo, hastaların hastaneye kabul ve taburcu süreçlerini, ilk tanıılarını, sigorta durumlarını, etnik/demografik özelliklerini ve hastane içi ölüm durumlarını içerir. Modelde hedeflenen “sepsis riskini” veya “hastane içi ölüm riskini” tahmin etme görevleri için etiket oluşturmak ve bazı temel öznitelikleri çıkarmak açısından değerlidir.

✓ Modelde Kullanılabilecek Etiketler (Target)

- **sepsis_label**: **diagnosis** sütununda “SEPSIS” geçen hastalara 1, diğerlerine 0 atanarak oluşturuldu. Bu, *sepsis hastalarını sınıflandırmak* için kullanılabilir.
- **hospital_expire_flag**: Hastane içinde ölüm gerçekleşip gerçekleşmediğini gösterir. *Ölüm risk tahmin modelleri* için doğrudan hedef olarak kullanılabilir.
- **deathtime**: Ölümün gerçekleştiği zaman. Zaman serisi modellerinde hedef zamanlı etiket olarak anlamlı olabilir.



Model Girdisi (Feature) Olarak Kullanılabilir Öznitelikler

 Öznitelik	Kullanım Yeri	Açıklama
stay_duration_hours	Sayısal giriş	Kalış süresi, klinik yükseklik için önemli gösterge
admission_type	Kategorik giriş	“EMERGENCY”, “ELECTIVE”, “URGENT” — aciliyet seviyesi riski etkileyebilir
admission_location	Kategorik giriş	Nereden geldiği (acil, transfer) ön tanı hakkında fikir verir
discharge_location	Kategorik giriş	Durumun ciddiyeti (ölüm, rehabilitasyon, ev vs.) üzerine analiz yapılabilir
insurance	Kategorik giriş	Sosyoekonomik durumun dolaylı etkileri olabilir
marital_status	Kategorik giriş	Psikososyal destek düzeyine dair ipucu verebilir

<code>ethnicity</code>	Kategorik giriş	Etik hassasiyetle kullanılmalı, dikkatle yorumlanmalı
<code>hospital_expire_flag</code>	Target veya yardımcı etiket	Ölüm tahmini için kullanılabilir
<code>sepsis_label</code>	Target veya yardımcı etiket	Sepsis teşhisi tahmini için kullanılabilir

⚠ Kullanılmaması Önerilen veya Düşük Öneme Sahip Alanlar

- `row_id`, `subject_id`, `hadm_id` → yalnızca eşleştirme için teknik amaçlı kullanılmalı
- `language`, `religion` → yüksek eksiklik oranı ve düşük model katkısı nedeniyle hariç bırakılabilir
- `edregtime`, `edouttime` → ileri düzey zaman analizleri yapılmayacaksa çıkarılabilir

🔍 Veri Yapısı ve Yeterlilik Özeti

- Toplam örnek sayısı: 129
- Sepsis pozitif hasta oranı: %10.8
- Ölüm oranı (`hospital_expire_flag = 1`): %31.0
- Ciddi eksiklik oranı olan sütunlar: `deathtime`, `language`, `edregtime`, `edouttime`
- Veri seti küçük ama etiket ve giriş çeşitliliği açısından prototip geliştirme için yeterli.

Sonuç

ADMISSIONS.csv, hedef etiketlerin tanımlanması (sepsis ve ölüm durumu), temel kategorik ve sayısal özniteliklerin çıkarılması ve demografik yapının analiz edilmesi için temel yapı taşıdır. Diğer tablolarla (**ICUSTAYS**, **CHARTEVENTS**, **LABEVENTS**) birleştirildiğinde, zaman serisi tabanlı daha kapsamlı bir tahmin modelinin temelini oluşturur.

✓ Bu veri seti, erken uyarı sistemleri, sepsis tespiti ve ölüm risk tahmini projelerinde kullanılmak üzere güçlü ve temiz bir başlangıç sağlar.

2) CALLOUT.csv

RangeIndex: 77 entries, 0 to 76

Data columns (total 24 columns):

#	Column	Non-Null Count	Dtype
0	row_id	77 non-null	int64
1	subject_id	77 non-null	int64
2	hadm_id	77 non-null	int64
3	submit_wardid	77 non-null	int64
4	submit_careunit	9 non-null	object
5	curr_wardid	77 non-null	int64
6	curr_careunit	77 non-null	object
7	callout_wardid	77 non-null	int64
8	callout_service	77 non-null	object
9	request_tele	77 non-null	int64
10	request_resp	77 non-null	int64
11	request_cdifff	77 non-null	int64
12	request_mrsa	77 non-null	int64
13	request_vre	77 non-null	int64
14	callout_status	77 non-null	object
15	callout_outcome	77 non-null	object
16	discharge_wardid	71 non-null	float64
17	acknowledge_status	77 non-null	object
18	createtime	77 non-null	object
19	updatetime	77 non-null	object
20	acknowledgetime	74 non-null	object
21	outcometime	77 non-null	object
22	firstreservationtime	39 non-null	object
23	currentreservationtime	5 non-null	object

dtypes: float64(1), int64(11), object(12)
memory usage: 14.6+ KB

Tablo İncelemesi: `CALLOUT.csv` (Hasta Transfer ve Danışma İstekleri)

`CALLOUT.csv` tablosu, MIMIC-III veritabanında hastane içinde yapılan **hasta nakli, birim değişikliği veya uzman görüşü** (konsültasyon) gibi çağrı kayıtlarını içerir. Özellikle kritik bakım ünitelerine transfer planları, doktor değerlendirmeleri ve sonuçları gibi bilgilerle **hasta akış süreçlerini anlamak** açısından önemlidir.

Sütun Açıklamaları

ID	Sütun Adı	Anlamı	Proje Bağlantısı
	<code>row_id</code>	Satır numarası (tekil ID)	Teknik amaçlı, modelde kullanılmaz
	<code>subject_id</code>	Hastaya özel benzersiz ID	Diğer tablolarda hasta bazlı birleştirme için kullanılır
	<code>hadm_id</code>	Hastane yatışı ID'si	<code>ADMISSIONS</code> , <code>ICUSTAYS</code> ile eşleştirme için kullanılır
	<code>submit_wardid</code>	Çağrının yapıldığı servis ID	Transferin hangi servisten başlatıldığını gösterir
	<code>submit_careunit</code>	Başvuran birim adı	(çok eksik) — ihmal edilebilir
	<code>curr_wardid</code>	Çağrı sırasında hastanın bulunduğu servis	Mevcut yatış yerini belirtir
	<code>curr_careunit</code>	Mevcut bakım birimi (örn. CCU, SICU)	Kritik bakım ihtimalini yansıtır, önemli
	<code>callout_wardid</code>	Çağrının gönderildiği servis ID	Hastanın sevk edilmesinin istendiği servis
	<code>callout_service</code>	Talep edilen servis (örn. MED, SURG, CSURG)	Tıbbi danışmanın alanını belirtir
	<code>request_tele</code>	Telemetri ihtiyacı (0/1)	Hasta izleme ihtiyacı — önemli klinik gösterge

request_resp	Solunum desteęi ihtiyacı (0/1)	Hastanın solunum durumu hakkında bilgi sağlar
request_cdifff	Clostridium difficile riski (0/1)	Enfeksiyon kontrolü için önemli
request_mrsa	MRSA riski (0/1)	Enfeksiyon kontrolü için önemli
request_vre	VRE riski (0/1)	Enfeksiyon kontrolü için önemli
callout_status	Çaęrı durumu (örn. Inactive)	Tamamlanmış/iptal edilmiş çağrı kontrolü
callout_outcome	Çaęrı sonucu (örn. Discharged)	Hasta sevk edildi mi, taburcu mu oldu?
discharge_wardid	Sevkten sonraki hedef servis ID	Hastanın nereye sevk edildiğini gösterir
acknowledge_status	Çaęrı onay durumu (Acknowledged, Revised)	Çaęrının klinik ekip tarafından kabul durumu
createtime	Çaęrının oluşturulduğu zaman	Zaman analizleri için önemli
updatetime	Çaęrı detayının güncellendięi zaman	
acknowledgetime	Çaęrının onaylandığı zaman	
outcometime	Çaęrıya verilen nihai yanıt zamanı	
firstreservationtime	İlk rezervasyon zamanı	Çok eksik — çoęu boş
currentreservationtime	Güncel rezervasyon zamanı	Çok eksik — çoęu boş



Neden Önemli?

- Bu tablo, **hasta akış süreçleri**, **konsültasyon talepleri** ve **birimler arası transfer** analizleri için değerli bilgiler sunar.
- Özellikle **yoğun bakım (ICU)**, **cerrahi servis**, **kardiyoloji** gibi birimlere yapılan çağrılar, hastanın kritikleştiğine işaret edebilir.
- `callout_outcome`, `request_resp`, `curr_careunit`, `callout_service` gibi sütunlar, modelde **kritik duruma geçiş riski**, **erken müdahale ihtimali** gibi çıkarımlarda kullanılabilir.
- `acknowledgetime`, `createtime`, `outcometime` zaman farklarıyla **yanıt gecikmesi**, **kllinik karar süresi** gibi metrikler çıkarılabilir.

⚠ Sınırlamalar

- `submit_careunit`, `firstreservationtime`, `currentreservationtime` gibi sütunlarda ciddi oranda eksik veri bulunduğu için doğrudan model girişi olarak kullanılması önerilmez.
- Sadece **77 satır** veri içerdiği için bu tablo doğrudan eğitim verisi olarak değil, diğer tablolara **ek açıklayıcı özellikler** üretmek için kullanılmalıdır (örneğin: “bu hasta transfer edilmiş mi?”, “ne kadar süre sonra edilmiştir?” gibi).

🧠 Kullanılabilir Öznitelikler (Önerilen Features)

Öznitelik	Kullanım Önerisi
<code>curr_careunit</code>	Mevcut bakım birimi — risk düzeyi için kategorik giriş
<code>callout_service</code>	Talep edilen uzmanlık alanı — model girişine dönüştürülebilir
<code>request_resp</code> , <code>request_tele</code>	Sayısal binary giriş — hasta durumu hakkında doğrudan ipucu
<code>callout_outcome</code>	Hedef değişken değilse bile “taburcu edildi mi, sevk mi oldu?” anlamı taşır
<code>outcometime - createtime</code>	Çağrıya ne kadar sürede dönüş yapılmış? — zaman temelli feature

🧭 Sonuç

`CALLOUT.csv`, sepsis riski taşıyan hastaların **kllinik izlem süreçlerinde alınan kararları ve transfer taleplerini** anlamak için güçlü bir yardımcı kaynaktır. Modelin temel verisinden ziyade,

bu tablo **feature engineering** aşamasında klinik akışı destekleyen yardımcı değişkenler üretmek için idealdir.

→ Sonraki adımda bu tabloyu `ICUSTAYS.csv` ile birleştirip “çağrı sonrası yoğun bakım sevki” gibi kavramlar üzerinden bir analiz planı oluşturabiliriz.

	Eksik Değer Sayısı	Oran (%)
currentreservationtime	72	93.51
submit_careunit	68	88.31
firstreservationtime	38	49.35
discharge_wardid	6	7.79
acknowledgetime	3	3.90

🔗 SUBMIT_WARDID değeri dağılımı:

submit_wardid

52 15
50 14
23 12
7 10
33 10
57 9
14 4
15 2
12 1

Name: count, dtype: int64

🔗 SUBMIT_CAREUNIT değeri dağılımı:

submit_careunit

NaN 68
MICU 5
SICU 2
TSICU 1
CCU 1

Name: count, dtype: int64

🔗 CURR_CAREUNIT değeri dağılımı:

curr_careunit

MICU 41
SICU 19
CCU 10
TSICU 4
CSRU 3

Name: count, dtype: int64

🔗 CALLOUT_WARDID değeri dağılımı:

callout_wardid

1 60

17	5
36	3
31	2
3	1
55	1
28	1
40	1
48	1
2	1
54	1

Name: count, dtype: int64

🔍 CALLOUT_SERVICE değeri dağılımı:

callout_service

MED	47
-----	----

SURG	7
------	---

CCU	5
-----	---

TSURG	4
-------	---

NMED	3
------	---

CSURG	3
-------	---

OMED	3
------	---

NSURG	3
-------	---

ORTHO	1
-------	---

CMED	1
------	---

Name: count, dtype: int64

🔍 REQUEST_TELE değeri dağılımı:

request_tele

0	41
---	----

1	36
---	----

Name: count, dtype: int64

🔍 REQUEST_RESP değeri dağılımı:

request_resp

0	77
---	----

Name: count, dtype: int64

🔍 REQUEST_CDIFFF değeri dağılımı:

request_cdifff

0	71
---	----

1	6
---	---

Name: count, dtype: int64

🔍 REQUEST_MRSA değeri dağılımı:

request_mrsa

0	60
---	----

1	17
---	----

Name: count, dtype: int64

🔍 REQUEST_VRE değeri dağılımı:

request_vre

0 68

1 9

Name: count, dtype: int64

🌀 CALLOUT_STATUS değeri dağılımı:

callout_status

Inactive 77

Name: count, dtype: int64

🌀 CALLOUT_OUTCOME değeri dağılımı:

callout_outcome

Discharged 71

Cancelled 6

Name: count, dtype: int64

🌀 ACKNOWLEDGE_STATUS değeri dağılımı:

acknowledge_status

Acknowledged 74

Revised 3

Name: count, dtype: int64

🌀 CREATETIME değeri dağılımı:

createtime

2149-05-31 10:44:34 1

2117-03-28 10:28:53 1

2160-05-05 11:11:05 1

2162-01-17 10:17:57 1

2119-11-06 12:47:39 1

..

2184-08-06 11:11:06 1

2118-10-08 10:44:24 1

2163-11-25 09:40:40 1

2163-11-22 15:27:18 1

2170-12-19 11:50:22 1

Name: count, Length: 77, dtype: int64

🌀 UPDATETIME değeri dağılımı:

updatetime

2149-05-31 10:44:34 1

2117-03-28 11:01:51 1

2160-05-05 11:11:05 1

2162-01-17 10:17:57 1

2119-11-06 16:14:06 1

..

2184-08-06 11:11:06 1

2118-10-08 10:48:16 1

2163-11-25 09:40:40 1

2163-11-22 15:27:18 1

2170-12-19 13:58:10 1

Name: count, Length: 77, dtype: int64

🌀 ACKNOWLEDGETIME değeri dağılımı:

acknowledgetime

NaN 3

2117-03-28 11:02:57 1

2160-05-05 11:33:57 1

2162-01-17 10:26:27 1

2119-11-06 13:56:10 1

..

2159-11-20 10:20:53 1

2184-08-06 11:20:29 1

2118-10-08 12:12:22 1

2163-11-25 09:48:40 1

2170-12-19 14:03:19 1

Name: count, Length: 75, dtype: int64

🌀 OUTCOMETIME değeri dağılımı:

outcometime

2149-05-31 22:40:02 1

2117-03-28 22:55:05 1

2160-05-05 14:10:08 1

2162-01-17 14:55:03 1

2119-11-06 16:14:06 1

..

2184-08-06 17:40:03 1

2118-10-08 16:40:03 1

2163-11-25 20:55:04 1

2163-11-22 22:25:03 1

2170-12-19 19:55:04 1

Name: count, Length: 77, dtype: int64

🌀 FIRSTRESERVATIONTIME değeri dağılımı:

firstreservationtime

NaN 38

2130-02-06 10:58:34 1

2112-02-09 16:25:30 1

2127-07-30 12:55:08 1

2117-03-27 14:25:04 1

2117-03-28 11:10:03 1

2102-09-01 17:55:06 1

2135-10-25 14:55:05 1

2162-01-17 10:55:03 1

2160-05-05 11:55:09 1

2142-11-29 16:55:21 1

2128-11-05 13:25:06 1

2111-01-09 13:25:03 1

2111-01-14 13:25:04 1

2198-07-06 16:40:08 1

2176-07-16 21:40:05 1

2112-05-06 15:40:09	1
2112-05-24 17:55:07	1
2112-06-01 12:40:07	1
2131-07-28 12:10:05	1
2202-10-04 14:10:08	1
2202-09-18 16:25:16	1
2145-07-17 15:55:09	1
2130-08-15 21:10:04	1
2151-08-14 17:10:07	1
2193-10-17 10:55:03	1
2159-11-20 10:55:25	1
2144-07-19 11:40:09	1
2145-07-09 14:10:08	1
2145-07-16 09:25:05	1
2145-09-07 17:10:06	1
2202-05-02 18:26:13	1
2199-02-11 13:55:04	1
2200-03-22 16:25:24	1
2200-06-11 16:12:23	1
2201-05-13 17:27:09	1
2201-08-11 15:13:18	1
2202-01-01 12:27:18	1
2202-02-16 22:40:29	1
2123-12-25 12:10:04	1

Name: count, dtype: int64

🔍 CURRENTRESERVATIONTIME değeri dağılımı:

currentreservationtime

NaN	72
2159-11-20 10:55:25	1
2202-02-16 22:40:29	1
2111-01-14 13:25:04	1
2112-05-24 17:55:07	1
2123-12-25 12:10:04	1

Name: count, dtype: int64

📌 Kullanılabilirlik Özeti: **CALLOUT.csv** (Kritik Hasta Transfer ve Konsültasyon Kayıtları)

CALLOUT.csv, MIMIC-III veri setinde hastaların birimler arası transfer talepleri, danışmanlık istekleri ve enfeksiyon risklerine dair çağrı kayıtlarını içeren değerli bir tablodur. Özellikle **klini** **aksiyon zamanlamaları** ve **hasta kritikleşme süreçlerinin izlenmesi** açısından güçlü öznitelikler sunar.

✅ Güçlü Yönleri (Kullanılabilir Bilgiler)

Özellik	Açıklama
 <code>curr_careunit, callout_service</code>	Hastanın çağrı sırasında bulunduğu ve çağrının hedeflendiği birim. Klinik risk seviyesi için anlamlı.
 <code>createtime, acknowledgedtime, outcometime</code>	Çağrı oluşturma, kabul ve tamamlanma zamanları. Zaman farkları ile klinik karar süreci analiz edilebilir.
 <code>request_tele, request_resp, request_mrsa, request_cdifff, request_vre</code>	Enfeksiyon ve destek ihtiyaçlarına yönelik binary veriler. Model için güçlü sayısal özellikler.
 <code>callout_outcome</code>	Hastanın çağrı sonrası taburcu edilip edilmediği. Modelleme için etiket veya sınıf olabilir.
 <code>acknowledge_status</code>	Çağrının ne zaman ve nasıl kabul edildiği. Gecikme tespiti ve analizler için kullanılabilir.

Sınırlamalar


Sütun	Sorun
<code>submit_careunit</code>	%88.3 eksik veri — kullanılabilir değil
<code>currentreservationtime</code> , <code>firstreservationtime</code>	%93+ eksik veri — ihmal edilmeli
<code>callout_status</code>	Tüm kayıtlar "Inactive" — hiçbir bilgi taşıyor
Veri Sayısı	Sadece 77 satır — tek başına model eğitimi için yeterli değil, ek veri ile birleşmeli

Önerilen Feature Engineering Çıktıları

Yeni Özellik	Açıklama
<code>callout_duration_hr</code>	<code>createtime</code> ile <code>outcometime</code> farkı — toplam işlem süresi
<code>ack_delay_hr</code>	<code>createtime</code> ile <code>acknowledgetime</code> farkı — kabul gecikmesi
<code>decision_speed_hr</code>	<code>acknowledgetime</code> ile <code>outcometime</code> farkı — karar süresi

Sonuç

- `CALLOUT.csv`, doğrudan etiket üretimi değil, **kllinik aksiyon zamanlamaları** ve **enfeksiyon riski temelli destek ihtiyaçlarını** anlamak için kullanılır.
- Diğer tablolarla (örneğin `ICUSTAYS.csv`, `ADMISSIONS.csv`) `subject_id / hadm_id` üzerinden birleştirilerek, bu tabloya ait bilgiler **yardımcı öznitelikler** olarak makine öğrenmesi modellerine dahil edilebilir.
- `callout_outcome` sınıfı (Discharged / Cancelled), küçük veri setinde dahi prototip model hedefi olabilir.

 Bu tablo, sepsis gibi kritik durumların yönetiminde zamanlama temelli karar destek sistemlerine yardımcı olmak üzere, **özellikle klinik gecikme analizi** ve **enfeksiyon risk sınıflandırması** için güçlü bir ek veri kaynağıdır.

3) CAREGIVERS.csv

	row_id	cgid	label	description
0	2228	16174	RO	Read Only
1	2229	16175	RO	Read Only
2	2230	16176	Res	Resident/Fellow/PA/NP
3	2231	16177	RO	Read Only
4	2232	16178	RT	Respiratory

RangeIndex: 7567 entries, 0 to 7566

Data columns (total 4 columns):

Column Non-Null Count Dtype

```
---
0 row_id      7567 non-null int64
1 cgid        7567 non-null int64
2 label       7518 non-null object
3 description  5155 non-null object
dtypes: int64(2), object(2)
memory usage: 236.6+ KB
```

Tablo İncelemesi: CAREGIVERS.csv (Sağlık Personeli Tanımları)

CAREGIVERS.csv, MIMIC-III veritabanında yer alan tüm sağlık çalışanlarının sistem içindeki benzersiz tanımlarını, mesleki etiketlerini ve açıklamalarını içeren bir referans tablosudur. Bu tablo; hastane içindeki veri girişlerini kimlerin yaptığına dair meta bilgi sağlar.



Sütun Açıklamaları



Sütun Adı



Anlamı



Proje Bağlantısı

row_id

Satır numarası (teknik ID)

Modelleme için kullanılmaz

cgid

Caregiver ID (sağlık personeli ID'si)

Diğer tablolarda (CHARTEVENTS, NOTEEVENTS) ile eşleşme için kullanılır

label

Kısa rol etiketi (örn. "RO", "MD", "RN")

Sağlık çalışanının sistemsel unvanı — analizlerde sınıflandırma için kullanılabilir

description	Tam rol açıklaması (örn. "Resident", "Respiratory Therapist")	Daha detaylı mesleki bilgi — rollerin gruplandırılması veya filtrelenmesi için anlamlı
-------------	---	--

✓ Neden Önemli?

- Bu tablo doğrudan hasta verisi içermese de, **veri girişlerinin kim tarafından yapıldığını anlamak** için temel referans kaynağıdır.
- cgid sütunu, CHARTEVENTS, NOTEEVENTS, INPUTEVENTS_CV/MV gibi diğer zaman serisi tablolarında veri girişi yapan kişiyi tanımlar.
- Bazı projelerde veri kaynağının **doktor mu, hemşire mi, teknisyen mi** olduğu analiz edilirken label ve description sınıflamaları önemli hale gelir.

🧠 Veri Kalitesi ve Eksiklikler

Sütun	Eksik Değer Oranı
label	%0.65 (49 kayıt)
description	%31.9 (2412 kayıt)

- label sütunu neredeyse eksiksizdir, **kategorik analiz için güvenilir**dir.
- description sütunu daha detaylı bilgi sunsa da, eksiklik oranı yüksektir. Gerekirse sadece label üzerinden modelleme yapılabilir.

📋 Kullanılabilecek Kategorik Değerler (Örnek)

label	Açıklama (description)
RO	Read Only (Veri görüntüleme yetkisi)
MD	Medical Doctor
RN	Registered Nurse

Res Resident/Fellow/PA/NP

RT Respiratory Therapist

PA Physician Assistant

Bu etiketler, sağlık çalışanlarının rolleri arasında ayrım yapılmasını sağlar. Örneğin yalnızca **MD** ve **Res** etiketli girdiler incelenerek doktor kaynaklı veriler filtrelenebilir.

🧠 Projede Olası Kullanım Senaryoları

- cgid** kullanılarak **CHARTEVENTS** içindeki kayıtların **kimin tarafından girildiği** tespit edilebilir.
- Yalnızca belirli rollerin girdiği verilere göre **rol bazlı tahmin modelleri** geliştirilebilir (örn. sadece hemşirelerin gözlemlediği bulgularla sepsis tahmini).
- description** ve **label** üzerinden **çalışan türü dağılımı**, veri katkı yoğunluğu gibi istatistikler çıkarılabilir.

🏁 Sonuç

CAREGIVERS.csv tablosu, diğer klinik tablolarla birlikte kullanıldığında **veri kaynağı doğruluğunu artıran**, rollere göre veri filtreleme ve analiz yapılmasına imkân tanıyan önemli bir yardımcı tablodur.

Kendisi doğrudan model girdisi olmasa da, veri izleme ve sınıflandırma süreçlerinde büyük rol oynar.

Bu tablo, özellikle çok kaynaktan gelen verilerin analizinde "kim, ne zaman, neyi girmiş?" sorusuna cevap bulmak için kullanılır.

	Eksik Değer Sayısı	Oran (%)
description	2412	31.88
label	49	0.65

📊 LABEL değeri dağılımı:

label
RO 1658
MD 1380
Res 1238
RN 1185
MDs 226

...

NStude 1
di 1
CPhT 1
Std 1
MSV 1
Name: count, Length: 231, dtype: int64

DESCRIPTION değeri dağılımı:

description
NaN 2412
Read Only 1954
Resident/Fellow/PA/NP 1546
RN 600
Attending 189
Rehabilitation 173
PCT/NA 121
Respiratory 121
Pharmacist 85
Dietitian 82
UCO 71
Social Worker 71
Case Manager 70
Administrator 28
Pastoral Care 23
Research Assistant 11
IMD 10
Name: count, dtype: int64

Kullanılabilirlik Özeti: CAREGIVERS.csv (Sağlık Personeli Tanımları)

CAREGIVERS.csv, MIMIC-III veri setinde sağlık personelinin temsil eden ve veri girişlerinde hangi sağlık çalışanlarının rol aldığını belirlemeye yarayan bir referans tablosudur. Klinik gözlemlerin, ilaç uygulamalarının veya notların kimin tarafından kaydedildiği bu tablo aracılığıyla çözümlenebilir.

Veri Özeti

Özellik	Değer
 Toplam Kayıt Sayısı	7,567
 Benzersiz cgid Sayısı	7,567

🔑 Benzersiz `label` Değeri 231

📖 `description` Eksikliği %31.88 (2,412 kayıt)

📄 `label` Eksikliği %0.65 (49 kayıt)

🔍 `label` ve `description` Dağılımı

- **En Yaygın `label` Değerleri:**
 - `RO` (Read Only) – 1,658 kişi
 - `MD` (Medical Doctor) – 1,380 kişi
 - `Res` (Resident / Fellow / PA / NP) – 1,238 kişi
 - `RN` (Registered Nurse) – 1,185 kişi
- **En Yaygın `description` Değerleri:**
 - `Read Only` – 1,954 kayıt
 - `Resident/Fellow/PA/NP` – 1,546 kayıt
 - `RN` – 600 kayıt
 - `Attending` – 189 kayıt
 - `Rehabilitation`, `Respiratory`, `Pharmacist`, `Dietitian`, `Social Worker` vb. meslek grupları da yer almakta

✅ Modelleme Açısından Kullanılabilirlik

Amaç	Kullanım Yöntemi
Veri kaynağını sınıflandırma	<code>cgid</code> üzerinden <code>CHARTEVENTS</code> , <code>NOTEVENTS</code> , <code>INPUTEVENTS</code> tablolarındaki girişleri sınıflandırmak
Roller bazlı analiz	Sadece <code>MD</code> , <code>RN</code> , <code>Res</code> gibi belirli rollere odaklanarak daha güvenilir sinyal üretmek
Etki alanı analizi	Hangi sağlık çalışanı grubunun hangi tür verileri daha çok girdiğini ölçümlemek
Filtreleme	<code>RO</code> (Read Only) gibi yalnızca görüntüleme yetkisi olanları dışlamak

⚠️ Sınırlamalar

Alan

Açıklama

`description`

%32 oranında eksik — eğer detaylı analiz istenmiyorsa sadece `label` kullanılmalı

`label`

Bazı etiketler yalnızca 1-2 kayıt içeriyor — düşük frekanslı sınıflar gruplandırılmalı

Model girdisi olarak doğrudan kullanımı

Tavsiye edilmez, daha çok diğer verilerle birleştirildiğinde anlamlı hale gelir



Önerilen Kullanım Senaryoları

- Karmaşık zaman serisi tablolarındaki `cgid` değerleriyle eşleştirilerek, veri girişini yapan kişinin sağlık rolü üzerinden verilerin güvenilirliği veya önceliği analiz edilebilir.
- Örneğin sadece **doktorlar (MD, Res)** tarafından girilen `CHARTEVENTS` kayıtları üzerinden sepsis tahmini yapılabilir.
- Hemşireler ve teknisyenler tarafından girilen veriler ayrı analiz setlerine bölünebilir.



Sonuç

`CAREGIVERS.csv`, doğrudan model girdisi olmasa da, **veri izlenebilirliğini artırmak, girdi güvenilirliği sağlamak ve klinik personel rolleri üzerinden detaylı analiz yapmak** için güçlü bir destek tablosudur.



Özellikle `CHARTEVENTS.csv` ve `NOTEVENTS.csv` gibi büyük tablolarla `cgid` üzerinden eşleştirildiğinde, sepsis gibi kritik durumların takibinde veri giriş kaynağının sağlık personeli rolleriyle birlikte değerlendirilmesini mümkün kılar.

4) CHARTEVENTS.csv

	row_id	subject_id	hadm_id	icustay_id	itemid	charttime	storetime	cgid	value	valuenum	valueuom	warning	error	resultstatus	stopped
0	5279021	40124	126179	279554.0	223761	2130-02-04 04:00:00	2130-02-04 04:35:00	19085	95.9	95.9	?F	0.0	0.0	NaN	NaN
1	5279022	40124	126179	279554.0	224695	2130-02-04 04:25:00	2130-02-04 05:55:00	18999	2222221.7	2222221.7	cmH2O	0.0	0.0	NaN	NaN
2	5279023	40124	126179	279554.0	220210	2130-02-04 04:30:00	2130-02-04 04:43:00	21452	15.0	15.0	insp/min	0.0	0.0	NaN	NaN
3	5279024	40124	126179	279554.0	220045	2130-02-04 04:32:00	2130-02-04 04:43:00	21452	94.0	94.0	bpm	0.0	0.0	NaN	NaN
4	5279025	40124	126179	279554.0	220179	2130-02-04 04:32:00	2130-02-04 04:43:00	21452	163.0	163.0	mmHg	0.0	0.0	NaN	NaN

RangeIndex: 758355 entries, 0 to 758354

Data columns (total 15 columns):

Column Non-Null Count Dtype





```
-----
0 row_id      758355 non-null int64
1 subject_id  758355 non-null int64
2 hadm_id     758355 non-null int64
3 icustay_id  758274 non-null float64
4 itemid      758355 non-null int64
5 charttime   758355 non-null object
6 storetime   758355 non-null object
7 cgid        758355 non-null int64
8 value       725507 non-null object
9 valuenum    323884 non-null float64
10 valueuom    239855 non-null object
11 warning     382279 non-null float64
12 error       382279 non-null float64
13 resultstatus 21674 non-null object
14 stopped     374649 non-null object
dtypes: float64(4), int64(5), object(6)
memory usage: 86.8+ MB
```

Tablo İncelemesi: CHARTEVENTS.csv (Yoğun Bakım Gözlem Verileri)

CHARTEVENTS.csv, MIMIC-III veri setinin en büyük ve en kritik tablolarından biridir. Yoğun bakım ünitesinde (ICU) hastalar hakkında zaman serisi halinde toplanan fizyolojik ölçümler, gözlemler ve vital bulgular bu tabloda yer alır. Her bir kayıt, bir sağlık personeli tarafından belirli bir zaman diliminde kaydedilen klinik değeri temsil eder.



Genel Bilgiler

Özellik	Değer
 Toplam Kayıt	758,355
 subject_id Sayısı	100,000+ (hastalara ait benzersiz ID)
 Ölçüm Sayısı (itemid)	3,200+ farklı medikal ölçüm türü
 Tarih Sütunları	charttime, storetime (gerçek zaman ve kayıt zamanı)



Eksik Veri Durumu

Sütun	Eksik Sayısı	Oran (%)
icustay_id	81	%0.01
value	32,848	%4.33
valuenum	434,471	%57.29
valueuom	518,500	%68.37
warning, error	376,076	%49.59
resultstatus	736,681	%97.15
stopped	383,706	%50.61






🔑 Özellikle `valuenum`, `valueuom`, `resultstatus` ve `stopped` sütunlarındaki eksiklikler analizlerde dikkatle ele alınmalıdır.

🧠 Açıklamalı Öznitelik İncelemesi

Sütun Adı	Açıklama
row_id	Her satıra özel benzersiz satır kimliği
subject_id	Hastayı temsil eden benzersiz ID
hadm_id	Hastane yatışına özel ID
icustay_id	Yoğun bakım kalışına özel ID
itemid	Ölçülen değeri tanımlayan kod (örneğin kalp atışı, solunum sayısı vb.)
charttime	Ölçümün gerçekleştiği zaman
storetime	Ölçümün kayda geçtiği zaman

<code>cgid</code>	Ölçümü yapan sağlık personelinin kimliği (<code>CAREGIVERS.csv</code> ile eşleşir)
<code>value</code>	Ham ölçüm değeri (string olabilir)
<code>valuenum</code>	Sayısal olarak karşılığı olan ölçüm değeri
<code>valueuom</code>	Birim (örneğin bpm, mmHg, °C)
<code>warning</code>	Ölçüm sırasında sistem tarafından üretilmiş uyarı (1 ise problem var)
<code>error</code>	Ölçüm sırasında hata oluşmuş mu (1 ise hata var)
<code>resultstatus</code>	Test sonucu durumu (örneğin 'Normal', 'Abnormal', nadiren dolu)
<code>stopped</code>	Ölçümün iptal durumu (örneğin 'Stopped', 'D/C' gibi)


✓ Kullanım Potansiyeli

Kullanım Alanı	Açıklama
 Zaman serisi modelleme	Belirli <code>itemid</code> 'lerin (örn. kalp atışı, solunum, ateş) zaman içinde değişimi analiz edilebilir
 Sepsis ve ölüm tahmini	Kritik vital değerlerin zamanlaması ve eğilimleri sepsis/ölüm riski için güçlü sinyaller sunar
 Sağlık personeli analizi	<code>cgid</code> ile eşleşerek veri girişi yapan personelin tipi analiz edilebilir
 Eksik veri tamamlama	<code>valueuom</code> , <code>valuenum</code> uyumsuzlukları filtrelenerek temiz alt veri setleri çıkarılabilir
 Anomali tespiti	Çok yüksek / tutarsız değerler (<code>valuenum</code>) hata olabilir, kontrol edilmelidir

Zorluklar ve Uyarılar

- 5) `itemid` değerleri çözümlenmeden hangi ölçümün neyi ifade ettiği bilinemez → `D_ITEMS.csv` tablosu ile eşleştirme yapılmalı.
- 6) Birden fazla ölçüm birimi (`valueuom`) aynı `itemid` için mevcut olabilir → standardizasyon gerekebilir.
- 7) `valuenum` eksikse `value` üzerinden string parsing ile bazı analizler yapılabilir.
- 8) `resultstatus`, `stopped` gibi nadiren dolu sütunlar bazı analizler için göz ardı edilebilir.

	Eksik Değer Sayısı	Oran (%)
<code>resultstatus</code>	736681	97.14
<code>valueuom</code>	518500	68.37
<code>valuenum</code>	434471	57.29
<code>stopped</code>	383706	50.60
<code>warning</code>	376076	49.59
<code>error</code>	376076	49.59
<code>value</code>	32848	4.33
<code>icustay_id</code>	81	0.01

 CHARTTIME değeri dağılımı:

```
charttime
2144-07-19 00:00:00    295
2144-07-19 04:00:00    256
2123-11-25 08:00:00    243
2144-07-18 20:00:00    227
2171-07-17 07:00:00    226
...
2119-10-31 20:22:00     1
2119-11-01 00:45:00     1
2119-11-01 03:05:00     1
2119-11-01 05:50:00     1
2180-03-17 17:45:00     1
Name: count, Length: 30866, dtype: int64
```

 STORETIME değeri dağılımı:

```
storetime
2198-07-09 17:27:00    388
2147-02-23 14:56:00    360
2171-08-16 03:41:00    321
2171-07-15 17:11:00    314
2117-08-21 17:16:00    300
...
2147-10-21 22:09:00     1
2201-09-29 17:04:00     1
2201-11-18 06:07:00     1
2147-10-21 20:38:00     1
2123-12-01 18:18:00     1
Name: count, Length: 48532, dtype: int64
```

🎯 VALUE değeri dağılımı:

value

NaN	32848
1	22665
Yes	17178
No	10376
0	9412

...

9.509119987487793	1
112.697998046875	1
2013.4200439453125	1
845.07000732421875	1
495.0	1

Name: count, Length: 10862, dtype: int64

🎯 VALUEUOM değeri dağılımı:

valueuom

NaN	518500
mmHg	80918
BPM	33180
%	28478
insp/min	12210

...

years	1
umol/L	1
sum	1
#/uL	1
liters	1

Name: count, Length: 72, dtype: int64

🎯 WARNING değeri dağılımı:

warning

NaN	376076
0.0	374893
1.0	7386

Name: count, dtype: int64

🎯 ERROR değeri dağılımı:

error

0.0	381631
NaN	376076
1.0	648

Name: count, dtype: int64

🎯 RESULTSTATUS değeri dağılımı:

resultstatus

NaN	736681
Final	19879
Manual	1795

Name: count, dtype: int64

🔴 STOPPED değeri dağılımı:

stopped

NaN 383706

NotStopd 373405

D/C'd 1244

Name: count, dtype: int64



CHARTEVENTS.csv – Kullanılabilirlik Özeti

CHARTEVENTS.csv, MIMIC-III veri setinin en hacimli ve zengin klinik gözlem tablosudur. Zaman serisi yapısında, yoğun bakım hastalarına ait fizyolojik ölçümler ve sağlık personeli tarafından kaydedilen gözlemler içerir.



Genel Bilgiler

- 📦 Toplam Kayıt: 758,355
- 🧠 Değişken Sayısı: 15
- 🕒 Zaman Boyutlu: charttime, storetime (ölçüm ve kayıt zamanı)
- 🔗 Kimlik: subject_id, hadm_id, icustay_id, cgid (hasta, yatış, ICU kalışı, personel ID)



Eksik Değer Durumu

Sütun	Eksik Sayısı	Oran (%)
resultstatus	736,681	%97.14
valueuom	518,500	%68.37
valuenum	434,471	%57.29
stopped	383,706	%50.60
warning / error	376,076	%49.59
value	32,848	%4.33

icustay_id

81

%0.01

📌 Özellikle `resultstatus`, `valueuom`, `valuenum` ve `stopped` sütunları oldukça eksik. Bu nedenle bu sütunlar modele girmeden önce filtrelenmeli veya eksik veriler analitik olarak göz önüne alınmalıdır.



Değer Dağılımı Örnekleri

- **VALUE:** Sayısal ve kategorik içerikler iç içe (örneğin: `Yes`, `No`, `1`, `0`, gibi değerler karışık). → Numerik dönüşüm öncesi ayıklama yapılmalı.
- **VALUEUOM:** 72 farklı birim içeriyor. Örneğin `mmHg`, `BPM`, `%`, `insp/min`, `mL`. → Standartlaştırma ve birim uyumu gerekir.
- **CHARTTIME / STORETIME:** Çok geniş tarih aralığı içeriyor. En sık gözlemler belirli tarihlerde yoğunlaşmış. → Zaman serisi analizleri için uygun.
- **WARNING / ERROR:** Genelde `0.0`, nadiren `1.0`. → Anomali tespiti ve model hatası için işaret olabilir.
- **RESULTSTATUS / STOPPED:** `Final`, `Manual`, `NotStopd`, `D/C'd` gibi değerler az sayıda gözüküyor. → Filtrelenmiş analizlerde kullanılabilir.



Kullanım Önerileri

Alan	Açıklama
Vital Takibi	<code>itemid</code> bazlı kalp atımı, solunum, ateş gibi parametrelerin zamana göre takibi yapılabilir.
Trend Analizi	<code>valuenum</code> üzerinden sepsis, hipotansiyon, oksijen saturasyonu gibi trendler çıkarılabilir.
Anomali Tespiti	<code>warning</code> , <code>error</code> , <code>valuenum</code> alanları modellemede "sorunlu kayıt" işaretleyicisi olabilir.
Sağlık Personeli Analizi	<code>cgid</code> ile <code>CAREGIVERS.csv</code> eşleşerek hangi personel ne tür kayıt giriyor analiz edilebilir.
Veri Temizliği Gerekliliği	<code>value</code> alanında string-numeric karışıklığı var. Örneğin "Yes", "No", "1", "0", "95.9", vs. → dönüşüm/temizlik zorunluluğu mevcut.

🕒 Zaman Bazlı Örüntü

`charttime` üzerinden saatlik/günlük yoğunluk, gece/gündüz farkları, ölçüm sıklıkları incelenebilir.

⚠ Zorluklar

- `itemid` değerleri açıklamasızdır → Anlamlandırmak için `D_ITEMS.csv` ile eşleştirme zorunludur.
- `value`, `valuenum`, `valueuom` üçlüsü uyumlu olmayabilir → Önileme gerektirir.
- Birçok sütun yüksek eksiklik içerdiğinden modelleme öncesinde filtreleme/feature engineering gereklidir.

🔍 Not: Bu tablo model eğitimi, anomaly detection ve risk skor tahmini için en güçlü zaman serisi verilerden biridir. Ancak büyük boyutu ve eksik veri oranı nedeniyle sistematik temizlik şarttır.

9) CPTEVENTS.csv

row_id	subject_id	hadm_id	costcenter	chartdate	cpt_cd	cpt_number	cpt_suffix	ticket_id_seq	sectionheader	subsectionheader	description
0	4615	10117	105150	ICU	NaN	99254	99254	NaN	1.0 Evaluation and management	Consultations	NaN
1	4616	10117	105150	ICU	NaN	99231	99231	NaN	2.0 Evaluation and management	Hospital inpatient services	NaN
2	4617	10117	105150	ICU	NaN	90935	90935	NaN	3.0 Medicine	Dialysis	NaN
3	4618	10117	105150	ICU	NaN	99231	99231	NaN	4.0 Evaluation and management	Hospital inpatient services	NaN
4	7753	10111	174739	ICU	NaN	99253	99253	NaN	1.0 Evaluation and management	Consultations	NaN

```
<class 'pandas.core.frame.DataFrame'>
```

RangeIndex: 1579 entries, 0 to 1578

Data columns (total 12 columns):

```
#   Column          Non-Null Count  Dtype
```

```
---  ---
0  row_id          1579 non-null  int64
1  subject_id      1579 non-null  int64
2  hadm_id         1579 non-null  int64
3  costcenter      1579 non-null  object
4  chartdate       291 non-null   object
5  cpt_cd          1579 non-null  int64
6  cpt_number      1579 non-null  int64
7  cpt_suffix      0 non-null     float64
8  ticket_id_seq   1288 non-null  float64
9  sectionheader   1579 non-null  object
10 subsectionheader 1579 non-null  object
11 description     291 non-null   object
```

dtypes: float64(2), int64(5), object(5)





memory usage: 148.2+ KB



CPTEVENTS.csv – Kullanılabilirlik Özeti


`CPTEVENTS.csv`, hastaların aldığı tıbbi hizmetlere ilişkin **Current Procedural Terminology (CPT)** kodlarını içeren bir tablodur. Klinik bakımda faturalama, prosedür analizi ve hizmet çeşitliliği gibi konular için önemlidir.

Genel Bilgiler




-  Toplam Satır Sayısı: 1,579
-  Değişken Sayısı: 12
-  Kayıt Türü: Tıbbi işlem kodları ve faturalama bilgileri
-  Anahtar Değişkenler:
 - `subject_id`, `hadm_id`: Hasta ve hastane yatış ID'si
 - `cpt_cd`, `cpt_number`: CPT kodu bilgileri
 - `sectionheader`, `subsectionheader`: Kategorik sınıflandırma
 - `description`: İşlem tanımı (büyük oranda eksik)

Eksik Değer Durumu

Sütun	Eksik Değer Sayısı	Oran (%)
<code>chartdate</code>	1,288	%81.55
<code>description</code>	1,288	%81.55
<code>cpt_suffix</code>	1,579	%100.00
<code>ticket_id_seq</code>	291	%18.43

 `cpt_suffix` sütunu tamamen boş. Diğer önemli açıklayıcı sütun olan `description` da yalnızca %18 oranında dolu.

Veri Tipleri

-  Sayısal: `row_id`, `subject_id`, `hadm_id`, `cpt_cd`, `cpt_number`, `ticket_id_seq`
-  Metin: `costcenter`, `sectionheader`, `subsectionheader`, `description`
-  Tarih: `chartdate` (ama çoğu eksik)








Örnek Değer Dağılımları

- **costcenter**: Tamamı **ICU** → Bu tablo sadece Yoğun Bakım (ICU) kayıtlarını içeriyor.
- **sectionheader** ve **subsectionheader**:
 - **Evaluation and management**, **Medicine** gibi ana başlıklar mevcut.
 - Alt başlıklar: **Hospital inpatient services**, **Consultations**, **Dialysis** gibi spesifik hizmet alanları.
- **cpt_cd** & **cpt_number**: Tıbbi işlem kodları → CPT katalogları ile eşleştirilerek anlamlı hale getirilmeli.



Kullanım Önerileri

Alan	Açıklama
 Faturalama Analizi	CPT kodları ile işlem sıklığı, maliyet modelleme veya hizmet yoğunluğu analiz edilebilir.
 Hizmet Türü Dağılımı	sectionheader / subsectionheader üzerinden ICU'daki işlem çeşitliliği incelenebilir.
 Zaman Trendleri	chartdate dolu olan satırlar üzerinden işlem zamanına göre hizmet trendi oluşturulabilir.
 Kod Gruplama	cpt_number kodları benzer işlem gruplarına göre kümeleme yapılabilir.
 Veri Eksikliği Yönetimi	description ve chartdate alanları eksik olduğundan, dolu satırlar filtrelenerek alt küme kullanılmalı.



Veri Kalitesi Sorunları

- **cpt_suffix** tamamen boş → Veri işleme sırasında çıkarılabilir.
- **chartdate** ve **description** alanları çok büyük oranda eksik → zamansal analiz ve yorumlama sınırlı.
- Açıklayıcı metin (**description**) yerine sadece kodlar varsa, **CPT referans dosyasıyla** anlamlandırma gerekir.




Not: Bu tablo, özellikle Yoğun Bakım'daki tıbbi hizmetlerin sınıflandırılması ve CPT kodlarına göre gruplandırılması için uygundur. Ancak eksik açıklamalar ve tarih bilgisi nedeniyle

modelleme sürecinde sınırlı katkı sağlayabilir. Geniřletme için dıř referanslarla eřleřtirme önerilir.

	Eksik Deęer Sayısı	Oran (%)
cpt_suffix	1579	100.00
chartdate	1288	81.57
description	1288	81.57
ticket_id_seq	291	18.43

 COSTCENTER deęeri daęılımı:

costcenter
ICU 1288
Resp 291
Name: count, dtype: int64

 CHARTDATE deęeri daęılımı:

chartdate
NaN 1288
2192-05-04 00:00:00 2
2198-07-19 00:00:00 2
2198-07-18 00:00:00 2
2198-07-17 00:00:00 2
...
2171-07-17 00:00:00 1
2171-07-18 00:00:00 1
2171-07-19 00:00:00 1
2171-07-20 00:00:00 1
2155-03-10 00:00:00 1
Name: count, Length: 285, dtype: int64

 CPT_SUFFIX deęeri daęılımı:

cpt_suffix
NaN 1579
Name: count, dtype: int64

 SECTIONHEADER deęeri daęılımı:

sectionheader
Evaluation and management 1150
Medicine 351
Surgery 75
Radiology 3
Name: count, dtype: int64

 SUBSECTIONHEADER deęeri daęılımı:

subsectionheader
Hospital inpatient services
739

Critical care services
 326
 Pulmonary
 291
 Consultations
 77
 Dialysis
 50
 Cardiovascular system
 27
 Respiratory system
 18
 Digestive system
 14
 Musculoskeletal system
 8
 Hydration, therapeutic, prophylactic, diagnostic injections and infusions, and chemotherapy and
 other highly complex drug or highly complex biologic agent administration 5
 Cardiovascular
 5
 Follow-up inpatient consultations (deleted codes)
 4
 Diagnostic ultrasound
 3
 Nervous system
 3
 Integumentary system
 2
 Prolonged services
 2
 Confirmatory consultation for a new or established patient (deleted codes)
 1
 Mediastinum and diaphragm
 1
 Case management services
 1
 Hemic and lymphatic systems
 1
 Urinary system
 1
 Name: count, dtype: int64

🔍 DESCRIPTION değeri dağılımı:

description

NaN	1288
VENT MGMT;SUBSQ DAYS(INVASIVE)	259
VENT MGMT, 1ST DAY (INVASIVE)	26
VENT MGMT,1ST DAY (NIV)	3
VENT MGMT;SUBSQ DAYS(NIV)	3

Name: count, dtype: int64



CPTEVENTS.csv – Kullanılabilirlik Özeti

Bu tablo, hastalara uygulanan CPT (Current Procedural Terminology) kodlarını içermektedir. CPT kodları tıbbi işlemleri standartlaştırır ve faturalandırma, klinik analiz ve kaynak yönetimi açısından kritik öneme sahiptir.



Genel Bilgiler

- **Toplam Satır:** 1,579
- **Sütun Sayısı:** 12
- **Temel Sütunlar:**
 - `subject_id`, `hadm_id`: Hasta ve yatış bilgileri
 - `cpt_cd`, `sectionheader`, `subsectionheader`: Tıbbi işlem sınıflandırmaları
 - `description`: CPT açıklaması (büyük ölçüde eksik)



Eksik Veri Analizi

Sütun	Eksik Değer Sayısı	Oran (%)
<code>cpt_suffix</code>	1579	100.00%
<code>chartdate</code>	1288	81.57%
<code>description</code>	1288	81.57%
<code>ticket_id_seq</code>	291	18.43%



`cpt_suffix` sütunu tamamen boş. `chartdate` ve `description` gibi kritik sütunlar da büyük oranda eksik.



Sıklık Analizi



COSTCENTER

- ICU: 1,288 kayıt
- Resp: 291 kayıt
- Veriler büyük oranda Yoğun Bakım (ICU) ortamından geliyor.

SECTIONHEADER

Kategori	Sayı
Evaluation and management	1150
Medicine	351
Surgery	75
Radiology	3

SUBSECTIONHEADER (İlk 5)

Alt Kategori	Sayı
Hospital inpatient services	739
Critical care services	326
Pulmonary	291
Consultations	77
Dialysis	50

DESCRIPTION

Açıklama	Sayı
NaN	1288
VENT MGMT;SUBSQ DAYS(INVASIVE)	259

VENT MGMT, 1ST DAY (INVASIVE)	26
VENT MGMT,1ST DAY (NIV)	3
VENT MGMT;SUBSQ DAYS(NIV)	3

✓ Kullanılabilirlik Değerlendirmesi

Alan	Durum	Not
<code>cpt_cd</code> ve <code>cpt_number</code>	✓ Uygun	CPT kodlarının kendisi mevcut.
<code>sectionheader</code>	✓ Uygun	İşlem türü analizi yapılabilir.
<code>subsectionheader</code>	✓ Uygun	Detaylı işlem türü sınıflandırması yapılabilir.
<code>description</code>	⚠ Eksik	Sadece %18 dolu, metinsel analiz için yetersiz.
<code>chartdate</code>	⚠ Eksik	Zamansal analiz yapılamaz.
<code>ticket_id_seq</code>	⚠ Kısmen	Analitik faydası sınırlı.
<code>cpt_suffix</code>	✗ Kullanılamaz	Tüm değerler NaN.

📌 Öneriler

- **Veri filtreleme:** Sadece `description` ve `chartdate` dolu kayıtlar kullanılabilir.
- **Zaman analizleri** için `chartdate` içeren alt küme oluşturulmalı.
- **Kod anlamlandırma** için `cpt_cd` değerleri, harici bir CPT referans tablosu ile eşleştirilmeli.
- **Yoğun bakım analizleri** için `costcenter` üzerinden segmentasyon yapılabilir.
- **Karmaşık işlem gruplama:** `sectionheader` ve `subsectionheader` alanları birlikte kullanılmalı.



Sonuç

`CPTEVENTS.csv`, özellikle **ICU işlemleri**, **faturalama kodları**, ve **prosedürel çeşitlilik** açısından değerlidir. Ancak açıklayıcı alanların eksikliği nedeniyle **destekleyici referans tablolar** ve **filtreleme** gerektirir.

10) DATETIMEEVENTS.csv

	row_id	subject_id	hadm_id	icustay_id	itemid	charttime	storetime	cgid	value	valueuom	warning	error	resultstatus	stopped
0	208474	10076	198503	201006.0	5684	2107-03-25 04:00:00	2107-03-25 04:34:00	20482	2107-03-24 00:00:00	Date	NaN	NaN	NaN	NotStopd
1	208475	10076	198503	201006.0	5684	2107-03-25 07:00:00	2107-03-25 07:06:00	15004	2107-03-24 00:00:00	Date	NaN	NaN	NaN	NotStopd
2	208836	10076	198503	201006.0	5684	2107-03-26 04:00:00	2107-03-26 05:31:00	20834	2107-03-24 00:00:00	Date	NaN	NaN	NaN	NotStopd
3	208837	10076	198503	201006.0	5684	2107-03-26 08:00:00	2107-03-26 08:33:00	17480	2107-03-24 00:00:00	Date	NaN	NaN	NaN	NotStopd
4	208838	10076	198503	201006.0	5684	2107-03-26 16:00:00	2107-03-26 16:08:00	17480	2107-03-24 00:00:00	Date	NaN	NaN	NaN	NotStopd

RangelIndex: 15551 entries, 0 to 15550

Data columns (total 14 columns):

Column Non-Null Count Dtype

```

0 row_id      15551 non-null int64
1 subject_id  15551 non-null int64
2 hadm_id     15551 non-null int64
3 icustay_id  15539 non-null float64
4 itemid      15551 non-null int64
5 charttime   15551 non-null object
6 storetime   15551 non-null object
7 cgid        15551 non-null int64
8 value       15467 non-null object
9 valueuom    15551 non-null object
10 warning     10914 non-null float64
11 error       10914 non-null float64
12 resultstatus 0 non-null float64
13 stopped    4637 non-null object

```

dtypes: float64(4), int64(5), object(5)

memory usage: 1.7+ MB



DATETIMEEVENTS.csv – Tablo İncelemesi

Bu tablo, zaman damgası (timestamp) ile ilişkili olayları kaydeder. Genellikle belirli tıbbi işlemlerin, ölçümlerin veya girişlerin zaman bilgisiyle takip edilmesinde kullanılır.



Genel Bilgiler

- **Toplam Satır Sayısı:** 15,551
- **Toplam Sütun Sayısı:** 14
- **Veri Tipleri:**
 - `int64`: 5 sütun

- `float64`: 4 sütun
- `object`: 5 sütun (tarih/saat ve metinler)



Temel Sütunlar

Sütun Adı	Açıklama
<code>row_id</code>	Eşsiz satır kimliği
<code>subject_id</code>	Hasta kimliği
<code>hadm_id</code>	Hastane yatışı kimliği
<code>icustay_id</code>	Yoğun bakım kalış kimliği
<code>itemid</code>	İzlenen/ölçülen değer türü
<code>charttime</code>	Değerin ölçüldüğü zaman
<code>storetime</code>	Değerin sisteme girildiği zaman
<code>cgid</code>	Girişi yapan klinik grup kimliği
<code>value</code>	Girilen değer
<code>valueuom</code>	Birim (örn. Date, min, hours...)
<code>warning</code>	Uyarı değeri (varsa)
<code>error</code>	Hata bilgisi (varsa)
<code>resultstatus</code>	Sonuç durumu (her zaman NaN)
<code>stopped</code>	İşlem durdurulmuş mu? (örn. 'NotStopd')



Eksik Veri Durumu

Sütun	Eksik Değer	Oran (%)
icustay_id	12	%0.08
value	84	%0.54
warning	4,637	%29.82
error	4,637	%29.82
resultstatus	15,551	%100.00
stopped	10,914	%70.20

⚠ `resultstatus` tamamen boş. `stopped`, `warning` ve `error` alanlarında ciddi eksiklik bulunmakta.

✅ Kullanılabilirlik Değerlendirmesi

Sütun	Durum	Açıklama
<code>subject_id</code> , <code>hadm_id</code> , <code>itemid</code> , <code>cgid</code>	✅ Uygun	Kimliklendirme ve analiz için kullanılabilir.
<code>charttime</code> , <code>storetime</code>	✅ Uygun	Zaman serisi analizi için yeterli.
<code>value</code> , <code>valueuom</code>	⚠ Kısmen	<code>value</code> sütununda küçük eksikler mevcut.
<code>icustay_id</code>	✅	Neredeyse eksiksiz.
<code>stopped</code> , <code>warning</code> , <code>error</code>	⚠ Eksik	Anlamlı analiz yapılabilmesi için ek temizlik gerekebilir.
<code>resultstatus</code>	❌ Kullanılamaz	Tamamen boş, analiz dışı bırakılmalı.

📌 Öneriler

- `valueuom == "Date"` olan kayıtlar, **zaman damgası kaydı** olarak ayrıştırılabilir.
- `itemid` değerlerinin anlamlandırılması için **referans sözlüğü** gereklidir.
- `stopped`, `warning` ve `error` eksik değerlerine karşı `NaN` yönetimi planlanmalıdır.
- `charttime` üzerinden **zaman serisi görselleştirme ve yoğunluk haritaları** oluşturulabilir.

💡 Sonuç

Bu tablo, özellikle zaman damgalı işlem kayıtları, olay takibi ve yoğun bakım süreci modelleme için kullanılabilir. Ancak eksik alanların temizlenmesi ve `itemid` sözlük eşlemesi yapılması önerilir.

	Eksik Değer Sayısı	Oran (%)
resultstatus	15551	100.00
stopped	10914	70.18
warning	4637	29.82
error	4637	29.82
value	84	0.54
icustay_id	12	0.08

🎯 CHARTTIME değeri dağılımı:

```
charttime
2123-12-17 17:37:00    15
2123-12-18 04:10:00    15
2123-12-17 12:43:00    15
2123-12-18 08:22:00    15
2123-12-19 04:18:00    15
..
2145-12-03 00:03:00     1
2147-10-14 08:00:00     1
2145-12-03 04:34:00     1
2147-10-14 04:36:00     1
2144-10-18 04:11:00     1
Name: count, Length: 3819, dtype: int64
```

🎯 STORETIME değeri dağılımı:

```
storetime
2129-11-26 16:57:00    32
2171-07-18 23:29:00    24
2171-07-17 01:53:00    22
2192-05-11 13:07:00    16
2129-11-30 22:53:00    16
```

..
2147-10-17 00:54:00 1
2130-08-14 12:23:00 1
2147-10-16 20:30:00 1
2107-01-06 20:52:00 1
2142-11-27 07:56:00 1
Name: count, Length: 3844, dtype: int64

🔍 VALUE değeri dağılımı:

value
2123-11-24 00:00:00 284
2147-10-12 00:00:00 283
2171-07-14 00:00:00 276
2117-08-05 00:00:00 223
2192-05-01 00:00:00 221

...
2198-07-04 05:38:00 1
2031-08-12 00:00:00 1
2198-07-02 20:14:00 1
2151-08-12 12:55:00 1
2041-05-16 00:00:00 1
Name: count, Length: 945, dtype: int64

🔍 VALUEUOM değeri dağılımı:

valueuom
Date 12600
Date and Time 2951
Name: count, dtype: int64

🔍 WARNING değeri dağılımı:

warning
0.0 10914
NaN 4637
Name: count, dtype: int64

🔍 ERROR değeri dağılımı:

error
0.0 10914
NaN 4637
Name: count, dtype: int64

🔍 RESULTSTATUS değeri dağılımı:

resultstatus
NaN 15551
Name: count, dtype: int64

🔍 STOPPED değeri dağılımı:

stopped
NaN 10914
NotStopd 4569

D/C'd 68
Name: count, dtype: int64

✓ DATETIMEEVENTS.csv – Kullanılabilirlik Özeti

Bu tablo, tarihsel olayların kaydını içerir. Özellikle prosedürlerin veya girişimlerin zamanlamasıyla ilgili verileri barındırır. Kullanılabilirlik açısından aşağıdaki analiz yapılmıştır:

📊 Eksik Veri Durumu

Sütun	Eksik Değer	Oran (%)
resultstatus	15,551	100.00 % ✗
stopped	10,914	70.18 % ⚠
warning	4,637	29.82 % ⚠
error	4,637	29.82 % ⚠
value	84	0.54 % ✓
icustay_id	12	0.08 % ✓

📌 `resultstatus` sütunu tamamen boş olduğu için analiz dışı bırakılmalı. `stopped`, `warning` ve `error` sütunlarında yüksek oranda eksiklik mevcut.

📦 Kategorik Değişken Dağılımları

- **valueum:**
 - `Date`: 12,600 → sadece tarih bilgisi içerir.
 - `Date and Time`: 2,951 → hem tarih hem saat bilgisi vardır.
 - ✓ Analizlerde ayrıştırılarak kullanılabilir.
- **stopped:**
 - `NotStopd`: 4,569
 - `D/C'd`: 68
 - `NaN`: 10,914 → çoğunluk eksik.
 - ⚠ Sadece bazı olaylar için durdurulma durumu mevcut.

- **warning ve error:**
 - 10,914 kayıta 0.0, geri kalanı NaN.
 - ⚠️ Yalnızca belirli kayıtlar için uyarı/hata kaydı tutulmuş.

🧠 Değer Sütunu (value)

- Büyük oranda zaman değeri taşıyor: Örn. 2123-11-24, 2147-10-12
- Çoğunlukla Date formatında.
- ✅ Zaman analizleri veya prosedür takibi için uygun.

🕒 Zaman Sütunları

- **charttime ve storetime** oldukça dolu (eksiksiz).
- ✅ Zaman serisi ve işlem süresi analizleri için kullanılabilir.

🔍 Genel Değerlendirme

Sütun	Kullanılabilirlik	Açıklama
subject_id, hadm_id, itemid, cgid	✅	Kimliklendirme için uygun
charttime, storetime	✅	Zaman analizi için uygun
value, valueuom	✅	Çoğunlukla geçerli ve tarih biçimli
icustay_id	✅	%99.92 oranında dolu
warning, error	⚠️	Eksik ama anlamlı olabilir
stopped	⚠️	Eksik ancak sınırlı analiz yapılabilir
resultstatus	❌	Kullanılamaz durumda

✅ Sonuç

DATETIMEEVENTS.csv, zaman temelli analizler için oldukça değerli bir tablodur. Özellikle `value` alanındaki zaman verileri, `charttime` ile birlikte prosedür başlangıç-bitiş analizlerinde kullanılabilir. Eksik alanlar nedeniyle bazı sütunlar (örn. `resultstatus`) çıkarılmalı ya da özel işlemlerle doldurulmalıdır.