

Introduction

心衰和心梗【所有入院首诊为心衰心梗的病人，不限于ICU，若多次因为首诊心衰或者心梗入院，则提取第一次入院信息】：

- 1. 人口学基本信息和共病情况 data_info4
- 2. 实验室指标 data_chart & data_lab
- 3. 用药信息 data_pharmacy
- 4. 在icu停留的记录 data_icustay

Tasks

- 1. 把实验室指标、用药信息 长数据转成宽数据
- 2. 把上述表格合并成一个宽数据的表格

```
D:\Dev\anaconda3\envs\pytorch\lib\site-packages\IPython\core\interactiveshell.py:3457: DtypeWarning: Columns (5) have mixed
types.Specify dtype option on import or set low_memory=False.
exec(code_obj, self.user_global_ns, self.user_ns)
```

随机展示

chart表，实验室表，以及用药表的分析数据有:18504552, 3197442, 667606
Char表、实验表、用药表中的 住院号数量: 4297, 12257, 12352
Char表、实验表、用药表中的 病人数量: 4297, 12257, 12352

0. 基本信息

Out[4]:

	subject_id	hadm_id	admittime	disctime	admission_type	admission_location	discharge_location	insurance	language	marital_status	...	pepti
6198	14933260	22416120	13/3/2115 18:38:00	15/3/2115 17:00:00	URGENT	TRANSFER FROM HOSPITAL	HOME	Other	ENGLISH	MARRIED	...	
9789	17848221	21041287	17/5/2177 14:29:00	20/5/2177 12:55:00	EW EMER.	EMERGENCY ROOM	HOME	Other	ENGLISH	SINGLE	...	
587	10467087	27290760	26/12/2194 14:52:00	28/12/2194 13:15:00	EW EMER.	EMERGENCY ROOM	HOME	Other	ENGLISH	DIVORCED	...	
3382	12683040	23692421	8/1/2139 00:03:00	11/1/2139 14:02:00	EW EMER.	TRANSFER FROM HOSPITAL	HOME	Medicare	ENGLISH	MARRIED	...	
7825	16204626	25418859	26/6/2135 04:33:00	6/7/2135 16:00:00	OBSERVATION ADMIT	EMERGENCY ROOM	HOME HEALTH CARE	Other	ENGLISH	NaN	...	
3241	12584492	28436809	9/1/2171 21:35:00	14/1/2171 14:30:00	OBSERVATION ADMIT	EMERGENCY ROOM	HOME HEALTH CARE	Medicare	ENGLISH	MARRIED	...	
7708	16115972	22463163	1/11/2133 04:32:00	10/11/2133 16:16:00	EW EMER.	TRANSFER FROM HOSPITAL	HOME	Other	ENGLISH	MARRIED	...	
4028	13193465	23073217	5/10/2153 03:01:00	10/10/2153 17:21:00	URGENT	PHYSICIAN REFERRAL	HOME HEALTH CARE	Other	ENGLISH	MARRIED	...	
5704	14520921	20966920	8/5/2121 11:44:00	14/5/2121 16:15:00	EW EMER.	EMERGENCY ROOM	HOME HEALTH CARE	Medicare	ENGLISH	WIDOWED	...	
7609	16050234	25765521	29/9/2118 18:51:00	4/10/2118 16:27:00	EW EMER.	PROCEDURE SITE	HOME HEALTH CARE	Medicare	?	MARRIED	...	

10 rows × 31 columns

患有‘心衰’/‘心梗’的病人:12428
检查唯一的病号数量:12428
['subject_id', 'hadm_id', 'admittime', 'disctime', 'admission_type', 'admission_location', 'discharge_location', 'insuranc
e', 'language', 'marital_status', 'ethnicity', 'status', 'gender', 'age', 'myocardial_infarct', 'congestive_heart_failure',
'peripheral_vascular_disease', 'cerebrovascular_disease', 'dementia', 'chronic_pulmonary_disease', 'rheumatic_disease', 'pep
tic_ulcer_disease', 'mild_liver_disease', 'diabetes_without_cc', 'diabetes_with_cc', 'paraplegia', 'renal_disease', 'maligna
nt_cancer', 'severe_liver_disease', 'metastatic_solid_tumor', 'aids']

0.1 基本信息表中字段解释

info4 更多 ([https://mimic.mit.edu/docs/iv/modules/hosp/admissions/%5D\(https://mimic.mit.edu/docs/iv/modules/icu/chartevents/\)](https://mimic.mit.edu/docs/iv/modules/hosp/admissions/%5D(https://mimic.mit.edu/docs/iv/modules/icu/chartevents/)): 有关住院的详细信息。

入院表提供了关于病人入院的信息。由于病人的每一次独特的医院访问都被分配了一个独特的hdm_id，所以入院表可以被认为是hdm_id的定义表。可用的信息包括入院和出院的时间信息、人口统计信息、入院的来源等等。

- 数据来自医院的入院、出院和转院数据库（通常称为“ADT”数据）。

- 有时会为在医院死亡的患者创建器官捐献者账户。这些是不同的住院时间，住院时间很短，有时甚至是负数。此外，他们的死亡时间通常与较早入院患者的死亡时间相同。

- `subject_id`, `hadm_id` : 该表的每一行都包含一个唯一的 `hadm_id`，代表单个患者入院。`hadm_id`的取值范围是2000000-2999999，这张表`subject_id`可能重复，说明一个病人多次入院。可以使用 `subject_id` 将 `ADMISSIONS` 表链接到 `PATIENTS` 表。
- `admittime`, `dischtime`, (入院时间、离院时间、死亡时间): `admittime`提供病人入院的日期和时间，`dischtime`则提供病人出院的日期和时间。如果适用，`deathtime`提供病人在医院内死亡的时间。请注意，只有当病人在医院内死亡时才会出现`deathtime`，而且几乎总是与病人的`dischtime`相同。然而，由于打字错误，可能会有一些差异。

💡 `deathtime` 无

- `admission_type` : 用于对入院的紧迫性进行分类。有9种可能性。门诊观察"、"直接急诊"、"直接观察"、"择期"、"紧急观察"、"观察入院"、"当天手术入院"、"紧急"。
- `admission_location` : 提供病人到达医院之前的位置信息。请注意，由于急诊室在技术上是一个诊所，通过急诊室入院的病人通常将其作为入院地点。
- 同样，`discharge_location` 是患者出院后的处置情况。
- `insurance`, `language`, `marital_status`, `ethnicity` (保险、语言、婚姻状况、种族): 保险、语言、婚姻状况和种族列提供有关给定住院的患者人口统计信息。请注意，由于每次入院都会记录此数据，因此它们可能会因住院而异。

💡 `edregtime`, `edouttime` (暂无) 患者在急诊室登记和出院的日期和时间。

向外链接-Links to:

- `patients` on `subject_id`
- `admissions` on `hadm_id`
- `icustays` on `stay_id`
- `d_items` on `itemid`

0.2 ‘心衰’/‘心梗’患者

🏥 疾病的编码 (表Infor4)

- `subject_id`
- `status` :
- `gender` :
- `age` :

- `myocardial_infarct` : 心肌梗死 ✅
- `congestive_heart_failure` : 充血性心力衰竭' ✅
- `peripheral_vascular_disease` : 周边血管疾病
- `cerebrovascular_disease` : 脑血管疾病
- `dementia` : 痴呆症
- `chronic_pulmonary_disease` : 慢性肺部疾病
- `rheumatic_disease` : 风湿病
- `peptic_ulcer_disease` : 消化性溃疡病
- `mild_liver_disease` : 轻度肝脏疾病
- `diabetes_without_cc` : 无cc的糖尿病
- `diabetes_with_cc` : 有cc的糖尿病
- `paraplegia` : 截瘫
- `renal_disease` : 肾脏疾病
- `malignant_cancer` : 恶性肿瘤
- `severe_liver_disease` : 严重的肝脏疾病
- `metastatic_solid_tumor` : 转移性固体肿瘤
- `aids` : 艾滋病

Out[6]:

	subject_id	status	gender	age	myocardial_infarct	congestive_heart_failure	peripheral_vascular_disease	cerebrovascular_disease	dementia	chron
0	10000980	0	F	75	0	1	0	0	0	
1	10001492	0	F	71	1	0	0	0	0	
2	10001877	0	M	90	1	1	0	0	0	
3	10002013	0	F	57	1	1	0	0	0	
4	10002155	0	F	81	1	1	0	0	0	
...
12423	19996783	0	M	89	1	0	0	0	0	
12424	19997367	0	F	64	1	1	0	0	0	
12425	19997473	0	F	82	1	1	0	0	0	
12426	19998330	0	F	72	0	1	0	0	0	
12427	19998497	0	F	82	1	0	0	0	0	

12428 rows × 21 columns

1. Chart表

Out[8]:

	subject_id	hadm_id	charttime	storetime	itemid	value	valuenum
0	10464977	23034003	7/11/2140 20:58:00	7/11/2140 20:58:00	223956	Ventricular Demand	NaN
1	10464977	23034003	7/11/2140 20:58:00	7/11/2140 20:58:00	223961	Attached-Pacer	NaN
2	10464977	23034003	7/11/2140 20:58:00	7/11/2140 20:58:00	224390	No	NaN
3	10464977	23034003	7/11/2140 20:58:00	7/11/2140 20:58:00	224834	Yes	NaN
4	10464977	23034003	7/11/2140 20:58:00	7/11/2140 20:58:00	224835	Yes	NaN
...
18504547	10464977	23034003	3/11/2140 00:00:00	3/11/2140 03:12:00	227947	Confusion	NaN
18504548	10464977	23034003	3/11/2140 00:00:00	3/11/2140 03:12:00	227948	Soft limb	NaN
18504549	10464977	23034003	3/11/2140 00:00:00	3/11/2140 03:12:00	227949	Done	NaN
18504550	10464977	23034003	3/11/2140 00:00:00	3/11/2140 03:12:00	227950	Both arms	NaN
18504551	10464977	23034003	3/11/2140 00:00:00	3/11/2140 03:12:00	227951	Circulation adequate	NaN

18504552 rows × 7 columns

1.1 ICU表中字段解释

[chartevent 更多 \(https://mimic.mit.edu/docs/iv/modules/icu/chartevents/\)](https://mimic.mit.edu/docs/iv/modules/icu/chartevents/): 在 ICU 逗留期间发生的图表项目。包含 ICU 中记录的大部分信息。

chartevents 包含所有可用于患者的图表数据。在他们入住 ICU 期间，患者信息的主要存储库是他们的电子图表。电子图表显示患者的常规生命体征和与他们的护理相关的任何其他信息：**呼吸机设置**、**实验室值**、**代码状态**、**精神状态**等。因此，有关患者住院的大部分信息都包含在图表事件中。此外，即使在其他地方 (labevents) 捕获了实验室值，它们也经常在 chartevents 中重复。发生这种情况是因为希望在患者的电子图表上显示实验室值，因此将值从存储实验室值的数据库复制到存储图表事件的数据库。

- 指定患者的标识符： subject_id 对于患者是唯一的， hadm_id 对于患者住院是唯一的， stay_id 对于患者住院是唯一的。有关这些标识符的更多信息，请参见此处。
- charttime 记录进行观察的时间，通常是最接近实际测量数据的时间。 storetime 记录临床工作人员手动输入或手动验证观察结果的时间。
- 数据库中单一测量类型的标识符。与一个 ITEMID （如212）相关的每一行都对应于同一测量（如心率）的一个实例。
- value 包含为 ITEMID 标识的概念测量的值。如果此值为数字，则 valuenum 包含数字格式的相同数据。如果此数据不是数字，则 valuenum 为空。在某些情况下（例如 Glasgow Coma Scale、Richmond Sedation Agitation Scale 和 Code Status 等分数）， valuenum 包含分数， value 包含分数和描述分数含义的文本。

Links to:

- patients on subject_id
- admissions on hadm_id
- icustays on stay_id
- d_items on itemid

2. Lab表

Out[9]:

	subject_id	hadm_id	charttime	storetime	itemid	value	valuenum
0	10013653	26666796	26/10/2182 07:40:00	26/10/2182 09:02:00	50960	2.0	2.00
1	10013653	26666796	26/10/2182 00:57:00	26/10/2182 01:02:00	50802	-2	-2.00
2	10013653	26666796	26/10/2182 00:57:00	26/10/2182 01:02:00	50804	23	23.00
3	10013653	26666796	26/10/2182 00:57:00	26/10/2182 01:02:00	50818	37	37.00
4	10013653	26666796	26/10/2182 00:57:00	26/10/2182 01:02:00	50820	7.38	7.38
...
3197437	16221250	23282331	26/9/2140 22:22:00	26/9/2140 22:50:00	51508	Red	NaN
3197438	12741325	20860534	26/1/2121 12:35:00	26/1/2121 14:16:00	51508	PINK	NaN
3197439	19119896	28318588	2/4/2195 16:00:00	2/4/2195 16:54:00	51508	Red	NaN
3197440	19506871	21026986	11/4/2118 22:50:00	11/4/2118 23:30:00	51508	DKAMB	NaN
3197441	18426598	22940206	24/4/2173 01:30:00	24/4/2173 02:38:00	51508		NaN

3197442 rows × 7 columns

2.1 实验室表中字段解释

[labevents \(https://mimic.mit.edu/docs/iv/modules/hosp/labevents/\)](https://mimic.mit.edu/docs/iv/modules/hosp/labevents/): 来自患者样本的实验室测量。

labevents 表存储为单个患者进行的所有实验室测量的结果。这些包括血液学测量、血气、化学面板和不太常见的测试，如基因检测。

hadm_id 是使用转院表分配给靠近住院的实验室的。然而，这并不总是能完美地捕捉到与住院时间相近的实验室。具体来说，从v2.1版开始，通过连接入院病人的 subject_id、入院时间和 disctime，可以为59,327,830个观测值分配 hadm_id。然而，这些观测值中只有58,131,956个（98%）有一个 hadm_id 实际存储在labevents表中。希望确保捕获与住院时间相近的实验室的用户应该意识到这一点，并在必要时使用带有时间的连接。

- charttime: 绘制实验室测量值的时间。这通常是获取标本的时间，通常明显早于可进行测量的时间。
- storetime: 在实验室系统中提供测量值的时间。这是信息提供给护理提供者的时间。
- itemid: 一个标识符，唯一地表示实验室概念
- value & valuenum: 实验室测量的结果，如果是数字，则转换为数字数据类型的值。

Links to

- d_labitems(另外一张表) on itemid

3. 用药表

Out[10]:

	subject_id	hadm_id	starttime	stoptime	medication	frequency
0	15346117	24544765	13/1/2201 00:00:00	18/1/2201 14:00:00	Sodium Chloride 0.9% Flush	Q8H
1	15346117	24544765	13/1/2201 00:00:00	21/1/2201 19:00:00	Heparin	TID
2	15346117	24544765	13/1/2201 01:00:00	13/1/2201 16:00:00	Acetaminophen	Q6H:PRN
3	15346117	24544765	13/1/2201 01:00:00	21/1/2201 19:00:00	Amlodipine	DAILY
4	15346117	24544765	13/1/2201 01:00:00	21/1/2201 19:00:00	Aspirin	DAILY
...
667601	17614063	26391110	22/4/2177 14:00:00	22/4/2177 20:00:00	Fluzone Trival	ONCE
667602	12361593	28465050	16/7/2147 13:00:00	16/7/2147 19:00:00	Fluzone Trival	ONCE
667603	17050261	28079544	20/4/2184 15:00:00	20/4/2184 20:00:00	Fluzone Trival	ONCE
667604	18749775	26585465	22/7/2147 04:00:00	24/7/2147 12:00:00	COQ10 (300 mg)	Q 12H
667605	16112663	24076527	8/12/2167 08:00:00	8/12/2167 17:00:00	Tenecteplase	2X

667606 rows × 6 columns

3.1 用药表中字段解释

[pharmacy 更多指标解释 \(https://mimic.mit.edu/docs/iv/modules/hosp/pharmacy/\)](https://mimic.mit.edu/docs/iv/modules/hosp/pharmacy/): 处方药的处方、剂量和其他信息。

药房表提供了有关开始给患者的填充药物的详细信息。药房信息包括药物剂量、处方剂量数、给药频率、用药途径和处方持续时间。

- subject_id: 是一个唯一的标识符，它指定了一个单独的病人。任何与单个subject_id相关的行都与同一个人有关。

- `hadm_id`: 是一个整数标识符, 对每个病人的住院治疗是唯一的。
- `starttime` `stoptime`: 给定的处方药的开始和停止时间
- `medication`: 提供的药物名称。
- `frequency`: 应向患者给药的频率。许多常用的短手都用在频率栏中。Q#表示每隔#小时; 例如“Q6”或“Q6H”是每 6 小时一班。

Links to

- `poe` on `poe_id`
- `prescriptions` on `pharmacy_id`
- `emar` on `pharmacy_id`

4. icu停留 表

Out[11]:

	subject_id	hadm_id	stay_id	first_careunit	last_careunit	intime	outtime	los
0	11679839	29636865	30021881	Coronary Care Unit (CCU)	Coronary Care Unit (CCU)	19/6/2112 14:41:32	20/6/2112 21:50:49	1.298113
1	13787728	27446327	30024343	Coronary Care Unit (CCU)	Coronary Care Unit (CCU)	17/2/2123 18:52:51	18/2/2123 13:39:32	0.782419
2	19352034	21509557	30058866	Coronary Care Unit (CCU)	Coronary Care Unit (CCU)	18/4/2114 02:02:38	30/4/2114 21:41:14	12.818472
3	18817948	23379351	30061662	Coronary Care Unit (CCU)	Coronary Care Unit (CCU)	2/11/2142 22:18:49	4/11/2142 00:46:50	1.102789
4	14964123	25103078	30081334	Surgical Intensive Care Unit (SICU)	Surgical Intensive Care Unit (SICU)	15/6/2137 16:33:00	17/6/2137 18:28:59	2.080544
...
4821	19420059	22049257	39920422	Cardiac Vascular Intensive Care Unit (CVICU)	Cardiac Vascular Intensive Care Unit (CVICU)	17/1/2114 08:55:42	22/1/2114 13:16:30	5.181111
4822	12673141	21284928	39921830	Cardiac Vascular Intensive Care Unit (CVICU)	Coronary Care Unit (CCU)	19/4/2131 05:28:27	23/4/2131 21:17:05	4.658773
4823	10836444	25551438	39934059	Cardiac Vascular Intensive Care Unit (CVICU)	Cardiac Vascular Intensive Care Unit (CVICU)	9/12/2170 13:50:58	10/12/2170 17:23:26	1.147546
4824	17014465	25960753	39966638	Cardiac Vascular Intensive Care Unit (CVICU)	Cardiac Vascular Intensive Care Unit (CVICU)	5/2/2112 09:23:57	6/2/2112 17:00:33	1.317083
4825	12275003	22562812	39992247	Cardiac Vascular Intensive Care Unit (CVICU)	Cardiac Vascular Intensive Care Unit (CVICU)	15/8/2182 09:37:33	16/8/2182 17:25:44	1.325127

4826 rows × 8 columns

4.1 ICU停留表中字段解释

[ICU stays 更多 \(https://mimic.mit.edu/docs/iv/modules/icu/icustays/\)](https://mimic.mit.edu/docs/iv/modules/icu/icustays/): ICU 停留的跟踪信息, 包括入院和出院时间。

- `stay_id` 是一个生成的标识符, 不是基于任何原始数据标识符。医院和ICU的数据库没有内在的联系, 所以没有任何ICU遭遇标识符的概念。
- `icustays` 表是由 `transfer` 表衍生出来的。具体来说, 它排除了病房不是ICU的转院记录。
- `transfer` 表中多个连续的ICU住院条目被合并为 `transfer` 表中的一个条目。
- 心率测量被用来确定重症监护室住院是否有效。如果 `transfer` 表中的重症监护室停留的心率测量值不可用, 重症监护室停留将不包括在 `icustays` 表中 (~5%)。

- `subject_id` & `hadm_id`: 指定患者的标识符: `subject_id` 对于患者是唯一的, `hadm_id` 对于患者住院是唯一的, `stay_id` 对于患者住院是唯一的。
- `first_careunit` & `last_careunit`: `FIRST_CAREUNIT` 和 `LAST_CAREUNIT` 分别包含病人被护理的第一个和最后一个ICU类型。由于一个`stay_id`将所有在24小时内入院的重症监护室分组, 所以病人有可能从一种重症监护室转到另一种重症监护室而拥有相同的 `stay_id`。
- `intime` & `outtime`: `INTIME` 提供患者被转移到 ICU 的日期和时间。 `OUTTIME` 提供患者转出 ICU 的日期和时间。
- `los`: `LOS` 是患者在给定的 ICU 住院期间的住院时间, 可能包括一个或多个 ICU 病房。停留时间以小数天数计算。

Links to:

- `patients` on `subject_id`
- `admissions` on `hadm_id`

处理: Chart和Lab表中的itemId关联Label

任务: 由于Chart表&Lab表中的itemId没有具体的解释, 需要分别关联两张表itemId&Labels来获取具体的解释, 作者姑且认为是检查项目; 讨论的过程得知, Chart表是进ICU的检查(此观点仍需进一步确认).

1. 匹配Chart中itemId
2. 匹配Lab中itemId

特性：展示了Chart&Lab两张表的Label具体的分布情况（以做过该项检查病人数量为Y轴数据），同时展示了前10个检查项目

所有的Chart itemid类别1889

Out[12]: True

Out[13]:

	subject_id	hadm_id	charttime	storetime	itemid	value	valuenum	label
4610269	18092322	27951426	14/10/2149 23:00:00	14/10/2149 23:14:00	224687	5.2	5.2	Minute Volume
17646549	10717732	29250995	7/6/2191 22:00:00	7/6/2191 22:25:00	220277	100	100.0	O2 saturation pulseoxymetry
5502288	14259352	21914364	16/12/2147 17:57:00	16/12/2147 17:57:00	224226	Red/Purple	NaN	Impaired Skin Wound Base #1
7232605	10773491	23621570	1/8/2144 01:23:00	1/8/2144 01:33:00	228902	Patient will be able to maintain clear and pat...	NaN	Altered Respiratory Status NCP - Expected outc...
17362091	18457420	27709468	25/1/2132 22:00:00	25/1/2132 21:56:00	224089	Ace wraps	NaN	Anti Embolic Device
14182657	13877419	27295973	23/8/2161 13:38:00	23/8/2161 13:38:00	223907	Pinpoint	NaN	Pupil Size Right
5763597	16637805	27570229	27/3/2164 16:00:00	27/3/2164 18:40:00	224015	Indwelling Urinary Catheter	NaN	Urine Source
16613043	13414571	21643106	12/1/2144 11:41:00	12/1/2144 11:58:00	228894	Patient will maintain adequate volume output	NaN	Impaired Fluid Balance NCP - Expected outcomes
13365763	14768951	28438615	16/2/2125 11:45:00	16/2/2125 11:47:00	220052	93.0	93.0	Arterial Blood Pressure mean
16486937	12721193	20249497	2/12/2163 08:38:00	2/12/2163 08:39:00	224073	Medications	NaN	Education Topic

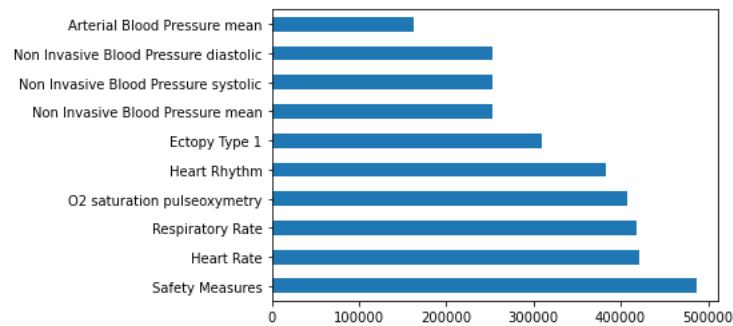


Chart-Label 分布情况统计



前10项ICU检查

Out[15]: <AxesSubplot:>



所有的实验室itemid类别655

Out[16]: True

Out[17]:

	subject_id	hadm_id	charttime	storetime	itemid	value	valuenum	label
906954	15135470	24506470	27/10/2142 07:35:00	27/10/2142 09:43:00	50970	4.2	4.2	Phosphate
1611814	15387461	22918339	5/8/2149 05:37:00	5/8/2149 12:21:00	50862	2.6	2.6	Albumin
2673625	15788134	24489664	7/9/2148 08:38:00	7/9/2148 11:33:00	51265	208	208.0	Platelet Count
1595883	15031111	27917559	6/2/2153 00:04:00	6/2/2153 01:20:00	50934	3	3.0	H
2002891	17091758	22364816	4/10/2126 04:02:00	4/10/2126 05:41:00	51221	29.6	29.6	Hematocrit
2748285	17936680	22901720	31/8/2195 07:33:00	31/8/2195 10:15:00	50882	22	22.0	Bicarbonate
2974448	10207476	23524758	20/12/2169 00:40:00	20/12/2169 02:05:00	51252	NaN	NaN	Microcytes
961324	16917696	26250558	24/7/2158 03:30:00	24/7/2158 04:00:00	51274	12.1	12.1	PT
2126677	10328270	29868079	17/8/2113 18:02:00	17/8/2113 23:13:00	50903	3.3	3.3	Cholesterol Ratio (Total/HDL)
1996547	16811765	23641208	13/5/2156 06:30:00	13/5/2156 07:44:00	50960	2.4	2.4	Magnesium

总计label(检查项目)种: 571

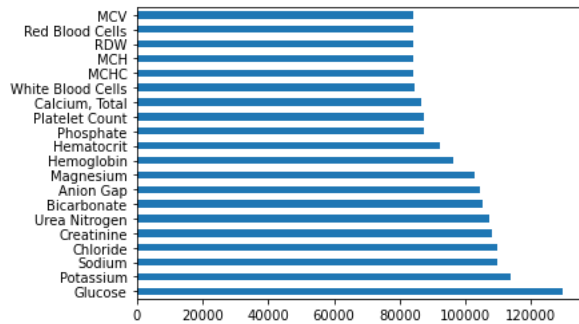


实验-Label 分布情况统计



前10项实验室检查

Out[20]: <AxesSubplot:>



处理: 合并ICU检查和实验室检查表

观察&思考

1. Chart表的数据条目比实验室Lab表要多 15million (15,307,110); 实验室条目: 3,197,442; chart条目: 18,504,552
2. Lab表中的病人为: 12,257; Chart表中病人为: 4,297
3. Chart表中的病人都能在Lab表中找到吗?

- 计算结果是 4288, ICU病人总计4,297; 有9个chart病人有chart检查没有Lab检查
- 4. 实验室中的itemid有655中, Chart中有1889种, 仍然符合chart多, 进ICU少 (lab) .

推测: chart检查的项目多,进ICU的病人少, 但是检查更多

5. ItemId以 5**** 开头的是lab表, 表示实验室检查; itemId以 2***** 开头为Chart表, 表示ICU检查
- 检查'5'开头的itemid有多少, 是否和Lab表的655同; 同理检查Chart表 (ICU,从上文得知ICU) 【检验, 正确】
6. ICU检查和实验室检查会重叠吗?
- 计算的结果说明这两个表的itemid没有发现重叠, 结果为0.

运算: $655 \times 1889 = 1237295$

- 说明总共的检查项目是 $655 + 1889$
- 7. 一个病人做的多项检查也如同 info4表中的长数据转变成宽数据吗, 也就是label值?

labels值有很长, 从实验室到ICU加起来($655+1889$).同一个人的实验室检查合成一条记录, ICU检查也合并成一条数据;

8. 以itemid为基础进行合并吗? 合并完成后以 subject_id 排序, 不同的检查对应了一个single patient.
9. value和valuenum如何处理呢

暂时不保留了?

合并的前提

- ☒ 根据观察5,将实验室表中vlue, valuenum, label重新命名; chart表同
- ☒ 因该是Concat,而不是Merge.

最后采取的操作是: '连接': 理由是itemid可以全部区分开; 两个表中的itemid没有交集

合并之后的检查项目表

Out[26]:

	subject_id	hadm_id	charttime	storetime	itemid	value	valuenum	label
20270947	10000980	29654838	4/1/2188 03:56:00	4/1/2188 05:25:00	50970	4.9	4.900	Phosphate
20270957	10000980	29654838	5/1/2188 06:09:00	5/1/2188 08:29:00	50912	1.4	1.400	Creatinine
20270958	10000980	29654838	5/1/2188 06:09:00	5/1/2188 08:29:00	50960	1.8	1.800	Magnesium
20270959	10000980	29654838	5/1/2188 06:09:00	5/1/2188 08:29:00	50970	5.3	5.300	Phosphate
20270960	10000980	29654838	5/1/2188 06:09:00	5/1/2188 08:29:00	50971	4.3	4.300	Potassium
...
19581388	19998497	29288061	2/7/2139 17:26:00	2/7/2139 21:17:00	51492	30	30.000	Protein
19581389	19998497	29288061	2/7/2139 17:26:00	2/7/2139 21:17:00	51493	2	2.000	RBC
19581390	19998497	29288061	2/7/2139 17:26:00	2/7/2139 21:17:00	51498	1.026	1.026	Specific Gravity
19581414	19998497	29288061	4/7/2139 02:40:00	4/7/2139 04:12:00	50960	2.2	2.200	Magnesium
19581387	19998497	29288061	2/7/2139 17:26:00	2/7/2139 21:17:00	51491	6.5	6.500	pH

21701994 rows × 8 columns

处理: 用药表&ICU停留表

- 1. 检查项目表 (Chart/Lab): uni_itmes_label (合并chart和lab)
- 2. 基本信息中的病人数量: 12428; 检查项目表的病人数量是 12,266; 用药表中的病人数量: 12,352; ICU停留表中的病人有: 4297例
- 3. 检查项目表表中的病人能够在用药表中找到的人数为 12,248
- 4. 用药表中病人号可以在基本信息中的找到的有: 12,352
- 5. 检查项目表的病人号可以在基本信息中的找到的有:
- 6. ICU的条目是4826

用药表

Out[41]:

	subject_id	hadm_id	medication	frequency
0	15346117	24544765	Sodium Chloride 0.9% Flush	Q8H
1	15346117	24544765	Heparin	TID
2	15346117	24544765	Acetaminophen	Q6H:PRN
3	15346117	24544765	Amlodipine	DAILY
4	15346117	24544765	Aspirin	DAILY
...
667601	17614063	26391110	Fluzone Trival	ONCE
667602	12361593	28465050	Fluzone Trival	ONCE
667603	17050261	28079544	Fluzone Trival	ONCE
667604	18749775	26585465	COQ10 (300 mg)	Q 12H
667605	16112663	24076527	Tenecteplase	2X

667606 rows × 4 columns

ICU停留表

Out[73]:

	subject_id	intime	los
0	11679839	19/6/2112 14:41:32	1.298113
1	13787728	17/2/2123 18:52:51	0.782419
2	19352034	18/4/2114 02:02:38	12.818472
3	18817948	2/11/2142 22:18:49	1.102789
4	14964123	15/6/2137 16:33:00	2.080544
...
4821	19420059	17/1/2114 08:55:42	5.181111
4822	12673141	19/4/2131 05:28:27	4.658773
4823	10836444	9/12/2170 13:50:58	1.147546
4824	17014465	5/2/2112 09:23:57	1.317083
4825	12275003	15/8/2182 09:37:33	1.325127

4826 rows × 3 columns