
MIMIC III SQL JOE RICE (JR76683)





METHOD

- MIMIC III data was loaded locally into Parquet files
- The duckdb Python package was used to query the Parquet files using SQL
 - Some syntax used is specific to this implementation (e.g. RegEx use)
- **Queries will be in the presentation notes to preserve formatting and ensure enough room**
- In many cases, we limit row counts for readability

WHAT PATIENTS HAVE THE MOST RECORDS ACROSS ALL TABLES?

- In each table with a Subject ID, we count the records and rank by record count, where a rank of 1 means the most records in each table
- We union all these results together
- We then make aggregations from the unioned data, grouping by subject ID
- We filter by patients in excess of 30 years old
- We aggregate all of those results and sort by:
 - The minimum ranks observed across all tables, ascending
 - The average ranks across all tables, ascending
 - The maximum age of a subject at a hospital admission, descending

	SUBJECT_ID	age_at_admission	avg(record_rank)	min(record_rank)	max(record_rank)	sum(Count_)
0	17891	46.775342	1519.055556	1	16915	56384.0
1	27427	75.887671	1519.250000	1	25872	201833.0
2	29035	66.873973	1529.950000	1	27349	129820.0
3	25256	39.586301	1600.350000	1	23838	203090.0
4	12613	61.364384	2176.705882	1	11933	429868.0
5	11923	49.797260	2207.352941	1	11284	344043.0
6	13033	39.676712	2303.650000	1	16033	129104.0
7	55672	58.394521	2526.526316	1	34942	139529.0
8	78076	71.646575	2646.947368	1	40818	198711.0
9	55639	77.246575	13885.777778	1	40737	779432.0
10	11318	72.747945	913.333333	2	10699	114863.0
11	15716	69.832877	1093.117647	2	14857	179172.0
12	5060	39.019178	1356.200000	2	6978	71700.0
13	25225	32.884932	1831.700000	2	23810	170342.0
14	55973	62.000000	2065.052632	2	35026	75214.0

Query in notes below

```

select
src.subject_id,
datediff('day', subj.dob, cast(adm.admittime as date))/365 as age_at_admission,
avg(record_rank),
min(record_rank),
max(record_rank),
sum(Count_)
from (
  select
  'Admissions' as Table,
  subject_id,
  count(*) Count_,
  row_number() over (partition by null order by Count_ desc, subject_id)
Record_Rank
from 'ADMISSIONS*.parquet'
group by 2
union all
select

```

```

    'Callout' as Table,
    subject_id,
    count(*) Count_,
    row_number() over (partition by null order by Count_ desc, subject_id)
Record_Rank
    from 'CALLOUT*.parquet'
    group by 2
    union all
    select
    'Chartevents' as Table,
    subject_id,
    count(*) Count_,
    row_number() over (partition by null order by Count_ desc, subject_id)
Record_Rank
    from 'CHARTEVENTS*.parquet'
    group by 2
    union all
    select
    'Cptevents' as Table,
    subject_id,
    count(*) Count_,
    row_number() over (partition by null order by Count_ desc, subject_id)
Record_Rank
    from 'CPTEVENTS*.parquet'
    group by 2
    union all
    select
    'Datetimeevents' as Table,
    subject_id,
    count(*) Count_,
    row_number() over (partition by null order by Count_ desc, subject_id)
Record_Rank
    from 'DATETIMEEVENTS*.parquet'
    group by 2
    union all
    select
    'Diagnoses_icd' as Table,
    subject_id,
    count(*) Count_,
    row_number() over (partition by null order by Count_ desc, subject_id)
Record_Rank
    from 'DIAGNOSES_ICD*.parquet'

```

```

group by 2
union all
select
'Drgcodes' as Table,
subject_id,
count(*) Count_,
row_number() over (partition by null order by Count_ desc, subject_id)
Record_Rank
from 'DRGCODES*.parquet'
group by 2
union all
select
'Icustays' as Table,
subject_id,
count(*) Count_,
row_number() over (partition by null order by Count_ desc, subject_id)
Record_Rank
from 'ICUSTAYS*.parquet'
group by 2
union all
select
'Inputevents_cv' as Table,
subject_id,
count(*) Count_,
row_number() over (partition by null order by Count_ desc, subject_id)
Record_Rank
from 'INPUTEVENTS_CV*.parquet'
group by 2
union all
select
'Inputevents_mv' as Table,
subject_id,
count(*) Count_,
row_number() over (partition by null order by Count_ desc, subject_id)
Record_Rank
from 'INPUTEVENTS_MV*.parquet'
group by 2
union all
select
'Labevents' as Table,
subject_id,
count(*) Count_,

```

```

    row_number() over (partition by null order by Count_desc, subject_id)
Record_Rank
  from 'LABEVENTS*.parquet'
  group by 2
  union all
  select
    'Microbiologyevents' as Table,
    subject_id,
    count(*) Count_,
    row_number() over (partition by null order by Count_desc, subject_id)
Record_Rank
  from 'MICROBIOLOGYEVENTS*.parquet'
  group by 2
  union all
  select
    'Noteevents' as Table,
    subject_id,
    count(*) Count_,
    row_number() over (partition by null order by Count_desc, subject_id)
Record_Rank
  from 'NOTEEVENTS*.parquet'
  group by 2
  union all
  select
    'Outputevents' as Table,
    subject_id,
    count(*) Count_,
    row_number() over (partition by null order by Count_desc, subject_id)
Record_Rank
  from 'OUTPUTEVENTS*.parquet'
  group by 2
  union all
  select
    'Patients' as Table,
    subject_id,
    count(*) Count_,
    row_number() over (partition by null order by Count_desc, subject_id)
Record_Rank
  from 'PATIENTS*.parquet'
  group by 2
  union all
  select

```

```

    'Prescriptions' as Table,
    subject_id,
    count(*) Count_,
    row_number() over (partition by null order by Count_ desc, subject_id)
Record_Rank
    from 'PRESCRIPTIONS*.parquet'
    group by 2
    union all
    select
    'Procedureevents_mv' as Table,
    subject_id,
    count(*) Count_,
    row_number() over (partition by null order by Count_ desc, subject_id)
Record_Rank
    from 'PROCEDUREEVENTS_MV*.parquet'
    group by 2
    union all
    select
    'Procedures_icd' as Table,
    subject_id,
    count(*) Count_,
    row_number() over (partition by null order by Count_ desc, subject_id)
Record_Rank
    from 'PROCEDURES_ICD*.parquet'
    group by 2
    union all
    select
    'Services' as Table,
    subject_id,
    count(*) Count_,
    row_number() over (partition by null order by Count_ desc, subject_id)
Record_Rank
    from 'SERVICES*.parquet'
    group by 2
    union all
    select
    'Transfers' as Table,
    subject_id,
    count(*) Count_,
    row_number() over (partition by null order by Count_ desc, subject_id)
Record_Rank
    from 'TRANSFERS*.parquet'

```

```
    group by 2
  ) src
inner join 'PATIENTS*.parquet' subj
  on src.subject_id = subj.subject_id
inner join (
  select
    subject_id,
    max(admittime) as admittime
  from 'ADMISSIONS*.parquet'
  group by 1
) adm
  on src.subject_id = adm.subject_id
where age_at_admission > 30
group by 1, 2
order by 4,3, 2
```


DO THEY HAVE ANY DISEASES IN COMMON?

- Using the top 3 results from the initial query, we now examine
- To do this, we use the DIAGNOSES_ICD joined to D_ICD_DIAGNOSES and intersect results from 3 separate queries, one for each subject

ROW_ID	ICD9_CODE	SHORT_TITLE	LONG_TITLE
0	4473 4280	CHF NOS	Congestive heart failure, unspecified
1	5908 5849	Acute kidney failure NOS	Acute kidney failure, unspecified

Query in notes below

```
select
diag.*
from 'DIAGNOSES_ICD*.parquet' subj_diag
inner join 'D_ICD_DIAGNOSES*.parquet' diag
on subj_diag.ICD9_CODE = diag.ICD9_CODE

where subj_diag.SUBJECT_ID = 17891
intersect
select
diag.*
from 'DIAGNOSES_ICD*.parquet' subj_diag
inner join 'D_ICD_DIAGNOSES*.parquet' diag
on subj_diag.ICD9_CODE = diag.ICD9_CODE

where subj_diag.SUBJECT_ID = 27427
intersect
select
diag.*
from 'DIAGNOSES_ICD*.parquet' subj_diag
```

```
inner join 'D_ICD_DIAGNOSES*.parquet' diag  
on subj_diag.ICD9_CODE = diag.ICD9_CODE
```

```
where subj_diag.SUBJECT_ID = 29035  
order by 1
```

WHAT MEDICATIONS HAVE THEY HAD IN COMMON?

- Similar to the last step, we take 3 subqueries, one per subject using the PRESCRIPTIONS table
- However, we union the results and each subquery returns subject ID, drug, and number of admissions drug was used
- We aggregate the previous, filtering to drugs only used by all 3 using a HAVING clause
- We sort by descending admission counts and take the top 10 entries

	DRUG	usage
0	Insulin	23.0
1	Acetaminophen	22.0
2	Sodium Chloride 0.9% Flush	21.0
3	Heparin	21.0
4	Iso-Osmotic Dextrose	16.0
5	Furosemide	16.0
6	D5W	16.0
7	Lorazepam	15.0
8	D5 1/2NS	15.0
9	Pantoprazole	15.0

Query in notes below

```
select
drug,
sum(count_) as usage
from (
  select
  subject_id,
  drug,
  count(distinct hadm_id) as count_
  from 'PRESCRIPTIONS*.parquet'
  where SUBJECT_ID = 17891
  group by 1,2
  union
  select
  subject_id,
  drug,
  count(distinct hadm_id) as count_
  from 'PRESCRIPTIONS*.parquet'
  where SUBJECT_ID = 27427
  group by 1,2
```

```
union
select
subject_id,
drug,
count(distinct hadm_id) as count_
from 'PRESCRIPTIONS*.parquet'
where SUBJECT_ID = 29035
group by 1,2
) rx
group by 1
having count(distinct subject_id) = 3
order by 2 desc
limit 10
```

WHAT TYPES OF ICU ADMISSIONS HAVE THEY HAD?

- For each subject, we use the TRANSFERS to look at the number of ICU and Ward visits by subject, one column for number of ward stays and number of admissions involving that ward
- We group by the CURR_CAREUNIT if it's available, otherwise the CURR_WARDID
- We sort using a regular expression, where if the grouped by column has all text (i.e. is a care unit), we see those first, then followed by ward ID
 - Secondarily, we sort by descending total stays in each ward/care unit

Ward	S1_Stays_in_Ward	S2_Stays_in_Ward	S3_Stays_in_Ward	S1_Stays_by_Adm	S2_Stays_by_Adm	S3_Stays_by_Adm	Total_Stays_in_Ward	Total_Admissions_in_Ward
0 ICU	0.0	0.0	0.0	0	0	0	0.0	0
1 MICU	1.0	12.0	7.0	1	4	4	20.0	9
2 TIKCU	0.0	0.0	3.0	0	0	2	3.0	2
3 CCU	0.0	0.0	3.0	0	0	3	3.0	3
4 NICU	0.0	0.0	0.0	0	0	0	0.0	0
5 INWARD	0.0	0.0	0.0	0	0	0	0.0	0
6 CSRU	0.0	0.0	0.0	0	0	0	0.0	0
7 SAS	16.0	0.0	21.0	5	0	12	37.0	17
8 ICU	0.0	7.0	0.0	0	1	0	7.0	1
9 HCU	3.0	0.0	3.0	2	0	2	6.0	4

Query in notes below

```
select
coalesce(CURR_CAREUNIT, cast(CURR_WARDID as varchar(3))) as Ward,
sum(case when subject_id = 17891 then 1 else 0 end) as S1_Stays_in_Ward,
sum(case when subject_id = 27427 then 1 else 0 end) as S2_Stays_in_Ward,
sum(case when subject_id = 29035 then 1 else 0 end) as S3_Stays_in_Ward,
count(distinct Case when subject_id = 17891 then hadm_id end) as
S1_Stays_by_Adm,
count(distinct Case when subject_id = 27427 then hadm_id end) as
S2_Stays_by_Adm,
count(distinct Case when subject_id = 29035 then hadm_id end) as
S3_Stays_by_Adm,
S1_Stays_in_Ward + S2_Stays_in_Ward + S3_Stays_in_Ward as Total_Stays_In_Ward,
S1_Stays_by_Adm + S2_Stays_by_Adm + S3_Stays_by_Adm as
Total_Admissions_In_Ward
from 'TRANSFERS*.parquet'
where WARD is not null
group by 1
order by case when regexp_matches(WARD, '^[A-Z]+$') then 1 else 0 end desc,
Total_Stays_In_Ward desc
```

limit 10

WHAT CHARTING OCCURS WITH THESE SUBJECTS?

- Using the CHARTEVENTS table, we look at the number of chart events by TYPE
- We then look by each TYPE and for each subject ID, checking to see if the 3 subjects experienced that TYPE
- We filter out using a HAVING clause combined with a LEAST of the counts by subject ID, ensuring the least is greater than 0

	LABEL	Item Code	Count	Chart_Event_Subj_1st	Chart_Event_Subj_2nd	Chart_Event_Subj_3rd
0	Heart Rate	2	670	3277	2379	
1	Respiratory Rate	2	655	3256	2391	
2	SpO2	1	658	1167	301	
3	calprevflg	1	660	1122	290	
4	Arterial BP [Systolic]	1	510	1121	180	
5	Arterial BP [Diastolic]	1	510	1121	180	
6	Arterial BP Mean	1	503	1117	174	
7	Resp Alarm [Low]	1	564	937	272	
8	HR Alarm [High]	1	564	937	272	
9	HR Alarm [Low]	1	564	937	272	
10	Resp Alarm [High]	1	564	937	272	
11	SpO2 Alarm [High]	1	554	937	272	
12	SpO2 Alarm [Low]	1	554	937	272	
13	NBP Alarm [Low]	1	377	897	272	
14	NBP Alarm [High]	1	377	897	272	
15	ABP Alarm [Low]	1	427	722	165	
16	ABP Alarm [High]	1	427	722	165	
17	Previous WeightF	1	354	618	79	
18	Mean Airway Pressure	2	45	398	158	
19	CVP	1	435	61	90	

Query in notes below

```

select
--cv.ITEMID,
items.LABEL,
count(distinct cv.itemid) as Item_Code_Count,
count(distinct case when subject_id = 17891 then charttime else null end)
Chart_Event_Subj_1st,
count(distinct case when subject_id = 27427 then charttime else null end)
Chart_Event_Subj_2nd,
count(distinct case when subject_id = 29035 then charttime else null end)
Chart_Event_Subj_3rd
from 'CHARTEVENTS*.parquet' cv
inner join 'D_ITEMS*.parquet' items
on cv.ITEMID = items.ITEMID
where subject_id in (17891,27427,29035)
and cv.value is not null
group by 1
having least(Chart_Event_Subj_1st, Chart_Event_Subj_2nd, Chart_Event_Subj_3rd) >
0
order by chart_event_Subj_1st + chart_event_Subj_2nd + chart_event_Subj_3rd desc

```

limit 20

HOW DO THE VARIOUS CHART READINGS COMPARE TO EACH OTHER?

- We make 3 subqueries, one for each subject
- In those queries, we use ordered analytics and partitioning to identify the admission number any readings are associated with
- We then aggregate the top 6 readings from the previous, by subject and admission
- We do full outer joins by admission number; we do this since we want all data for all admissions and the admission numbers may be different for all
- Since the data are so wide, we only show one example of readings on the right

	Admission Number for Patient	Avg Heart Rate Subj1	Avg Heart Rate Subj2	Avg Heart Rate Subj3	Avg Respiratory Rate Subj1
0	1	65.292929	98.962963	69.290323	14.573477
1	2	89.331333	89.285135	68.712955	13.677081
2	3	107.635992	92.666920	65.285744	17.253589
3	4	93.352113	79.363333	84.083892	22.774648
4	5	NaN	86.787565	NaN	NaN
5	5	NaN	73.473684	NaN	NaN
6	6	NaN	NaN	67.572165	NaN
7	7	NaN	NaN	65.925210	NaN
8	8	NaN	NaN	64.906667	NaN
9	9	NaN	NaN	73.905660	NaN
10	10	NaN	NaN	68.822368	NaN
11	11	NaN	NaN	75.287356	NaN
12	12	NaN	NaN	76.638889	NaN
13	13	NaN	NaN	76.307692	NaN

Query in notes below

```

select
coalesce(subj1.admission_number, subj2.admission_number,
subj3.admission_number) as Admission_Number_for_Patient,
Subj1.Avg_Heart_Rate as Avg_Heart_Rate_Subj1,
Subj2.Avg_Heart_Rate as Avg_Heart_Rate_Subj2,
Subj3.Avg_Heart_Rate as Avg_Heart_Rate_Subj3,
Subj1.Avg_Respiratory_Rate as Avg_Respiratory_Rate_Subj1,
Subj2.Avg_Respiratory_Rate as Avg_Respiratory_Rate_Subj2,
Subj3.Avg_Respiratory_Rate as Avg_Respiratory_Rate_Subj3,
Subj1.Avg_SpO2 as Avg_SpO2_Subj1,
Subj2.Avg_SpO2 as Avg_SpO2_Subj2,
Subj3.Avg_SpO2 as Avg_SpO2_Subj3,
Subj1.Avg_calprevflg as Avg_calprevflg_Subj1,
Subj2.Avg_calprevflg as Avg_calprevflg_Subj2,
Subj3.Avg_calprevflg as Avg_calprevflg_Subj3,
Subj1.Avg_Arterial_BP_Diastolic as Avg_Arterial_BP_Diastolic_Subj1,
Subj2.Avg_Arterial_BP_Diastolic as Avg_Arterial_BP_Diastolic_Subj2,
Subj3.Avg_Arterial_BP_Diastolic as Avg_Arterial_BP_Diastolic_Subj3,

```

```

Subj1.Avg_Arterial_BP_Systolic as Avg_Arterial_BP_Systolic_Subj1,
Subj2.Avg_Arterial_BP_Systolic as Avg_Arterial_BP_Systolic_Subj2,
Subj3.Avg_Arterial_BP_Systolic as Avg_Arterial_BP_Systolic_Subj3,

from (
  select
    subject_id,
    hadm_id,
    min(charttime) as First_Chart_Event,
    avg(case when label = 'Heart Rate' then valuenum end) as Avg_Heart_Rate,
    avg(case when label = 'Respiratory Rate' then valuenum end) as
    Avg_Respiratory_Rate,
    avg(case when label = 'SpO2' then valuenum end) as Avg_SpO2,
    avg(case when label = 'calprevflg' then valuenum end) as Avg_calprevflg,
    avg(case when label = 'Arterial BP [Diastolic]' then valuenum end) as
    Avg_Arterial_BP_Diastolic,
    avg(case when label = 'Arterial BP [Systolic]' then valuenum end) as
    Avg_Arterial_BP_Systolic,

```

```

    Row_Number() over (partition by subject_id order by min(charttime)) as
    Admission_Number
  from 'CHARTEVENTS*.parquet' cv
  inner join 'D_ITEMS*.parquet' items
    on cv.ITEMID = items.ITEMID
  where items.LABEL in ('Heart Rate','Respiratory Rate','SpO2','calprevflg','Arterial BP
[Diastolic]','Arterial BP [Systolic]')
  and subject_id = 17891
  group by 1,2
) Subj1
full outer join (
  select
    subject_id,
    hadm_id,
    min(charttime) as First_Chart_Event,
    avg(case when label = 'Heart Rate' then valuenum end) as Avg_Heart_Rate,
    avg(case when label = 'Respiratory Rate' then valuenum end) as
    Avg_Respiratory_Rate,
    avg(case when label = 'SpO2' then valuenum end) as Avg_SpO2,
    avg(case when label = 'calprevflg' then valuenum end) as Avg_calprevflg,
    avg(case when label = 'Arterial BP [Diastolic]' then valuenum end) as
    Avg_Arterial_BP_Diastolic,
    avg(case when label = 'Arterial BP [Systolic]' then valuenum end) as

```

Avg_Arterial_BP_Systolic,

```
Row_Number() over (partition by subject_id order by min(charttime)) as
Admission_Number
from 'CHARTEVENTS*.parquet' cv
inner join 'D_ITEMS*.parquet' items
  on cv.ITEMID = items.ITEMID
where items.LABEL in ('Heart Rate','Respiratory Rate','SpO2','calprevflg','Arterial BP
[Diastolic]','Arterial BP [Systolic]')
and subject_id = 27427
group by 1,2
) Subj2
on Subj1.admission_number = subj2.admission_number
full outer join (
  select
    subject_id,
    hadm_id,
    min(charttime) as First_Chart_Event,
    avg(case when label = 'Heart Rate' then valuenum end) as Avg_Heart_Rate,
    avg(case when label = 'Respiratory Rate' then valuenum end) as
Avg_Respiratory_Rate,
    avg(case when label = 'SpO2' then valuenum end) as Avg_SpO2,
    avg(case when label = 'calprevflg' then valuenum end) as Avg_calprevflg,
    avg(case when label = 'Arterial BP [Diastolic]' then valuenum end) as
Avg_Arterial_BP_Diastolic,
    avg(case when label = 'Arterial BP [Systolic]' then valuenum end) as
Avg_Arterial_BP_Systolic,
```

```
Row_Number() over (partition by subject_id order by min(charttime)) as
Admission_Number
from 'CHARTEVENTS*.parquet' cv
inner join 'D_ITEMS*.parquet' items
  on cv.ITEMID = items.ITEMID
where items.LABEL in ('Heart Rate','Respiratory Rate','SpO2','calprevflg','Arterial BP
[Diastolic]','Arterial BP [Systolic]')
and subject_id = 29035
group by 1,2
) Subj3
on subj1.admission_number = subj3.admission_number
and subj2.admission_number = subj3.admission_number
order by 1
```

HOW MANY OTHER PEOPLE HAVE THESE DIAGNOSES?

- Using the DIAGNOSES_ICD table and D_ICD_DIAGNOSES, we find the applicable ICD9 codes for our 3 subjects and do a count distinct subject IDs for those codes

	ICD9_CODE	SHORT_TITLE	LONG_TITLE	Patients Affected
0	5849	Acute kidney failure NOS	Acute kidney failure, unspecified	7687
1	4280	CHF NOS	Congestive heart failure, unspecified	9843

Query in notes below

```
select
diag.*
from DIAGNOSES_ICD subj_diag
inner join D_ICD_DIAGNOSES diag
on subj_diag.ICD9_CODE = diag.ICD9_CODE

where subj_diag.SUBJECT_ID = 17891
intersect
select
diag.*
from DIAGNOSES_ICD subj_diag
inner join D_ICD_DIAGNOSES diag
on subj_diag.ICD9_CODE = diag.ICD9_CODE

where subj_diag.SUBJECT_ID = 27427
intersect
select
diag.*
from DIAGNOSES_ICD subj_diag
```

```
inner join D_ICD_DIAGNOSES diag
on subj_diag.ICD9_CODE = diag.ICD9_CODE

where subj_diag.SUBJECT_ID = 29035
order by 1
```

HOW MANY OTHER PEOPLE HAVE USED THE SAME MEDS?

- Like the previous, but we instead look at drugs instead

	DRUG	Patients_Administered_Rx
0	Sodium Chloride 0.9% Flush	29387
1	Acetaminophen	28821
2	Insulin	25235
3	Heparin	24444
4	D5W	21279
5	Iso-Osmotic Dextrose	21098
6	Furosemide	19738
7	Pantoprazole	17074
8	Lorazepam	15449
9	D5 1/2NS	7944

Query in notes below

```
select
drug,
count(distinct subject_id) as Patients_Administered_Rx
from 'PRESCRIPTIONS*.parquet'
where drug in (
select
  DRUG
from (
```

```
select
drug,
sum(count_) as usage
from (
  select
    subject_id,
    drug,
    count(distinct hadm_id) as count_
  from 'PRESCRIPTIONS*.parquet'
```

```

where SUBJECT_ID = 17891
group by 1,2
union
select
subject_id,
drug,
count(distinct hadm_id) as count_
from 'PRESCRIPTIONS*.parquet'
where SUBJECT_ID = 27427
group by 1,2
union
select
subject_id,
drug,
count(distinct hadm_id) as count_
from 'PRESCRIPTIONS*.parquet'
where SUBJECT_ID = 29035
group by 1,2
) rx
group by 1
having count(distinct subject_id) = 3
order by 2 desc
limit 10

)
)
group by 1
order by 2 desc
limit 10

```

HOW MANY TIMES HAVE THE SAME MEDS AND DISEASES OCCUR TOGETHER

- We join the DIAGNOSES_ICD and DRUG table to get the combinations
- We then down select using applicable subqueries for each subject, using queries that we used previously
- We then do a count distinct of the subject IDs to see how often they show up together
- With so many combinations, it's not surprising the value is so high

count(DISTINCT rx.subject_id)	
0	13199

Query in notes below

```
select
count(distinct rx.subject_id)
from (
  select
  distinct
  subject_id,
  drug
  from 'PRESCRIPTIONS*.parquet'
) rx
inner join (
  select
  distinct
  subject_id,
  icd9_code
  from 'DIAGNOSES_ICD*.parquet'
) diag
on rx.subject_id = diag.subject_id
where drug in (
```



```

select
  DRUG
  from (

select
drug
from 'PRESCRIPTIONS*.parquet'
where SUBJECT_ID = 17891
intersect
select
drug
from 'PRESCRIPTIONS*.parquet'
where SUBJECT_ID = 27427
intersect
select
drug
from 'PRESCRIPTIONS*.parquet'
where SUBJECT_ID = 29035

  )
)
and icd9_code in (
select
  ICD9_CODE
  from (

select
diag.*
from 'DIAGNOSES_ICD*.parquet' subj_diag
inner join 'D_ICD_DIAGNOSES*.parquet' diag
on subj_diag.ICD9_CODE = diag.ICD9_CODE

where subj_diag.SUBJECT_ID = 17891
intersect
select
diag.*
from 'DIAGNOSES_ICD*.parquet' subj_diag
inner join 'D_ICD_DIAGNOSES*.parquet' diag
on subj_diag.ICD9_CODE = diag.ICD9_CODE

where subj_diag.SUBJECT_ID = 27427
intersect

```

```
select
diag.*
from 'DIAGNOSES_ICD*.parquet' subj_diag
inner join 'D_ICD_DIAGNOSES*.parquet' diag
on subj_diag.ICD9_CODE = diag.ICD9_CODE

where subj_diag.SUBJECT_ID = 29035
order by 1

)
)
and rx.subject_id not in (17891,27427,29035)
```

WHAT ARE THE RAREST COMBINATIONS MOST RECENTLY RECORDED?

- We count drug/ICD9 combinations from joining the DIAGNOSES_ICD and DRUG table, joined by SUBJECT ID and ADMISSION ID
- We then sort by descending discharge date and ascending count
- It should be noted, the combinations do not necessarily mean a drug treat the diagnosis; just that they occur in the same hospital admission
- More analysis would have to be required to get that type of information

ICD9_CODE	SHORT_TITLE	LONG_TITLE	DRUG	Discharge Date	Patients Affected
0	4029 Bacterial pneumonia NOS	Bacterial pneumonia, unspecified	scopolamine base	2210-08-24 19:43:00	1
1	V463 Wheelchair dependence	Wheelchair dependence	Silver Sulfadiazine 1% Cream	2210-08-24 19:43:00	1
2	34540 Psychomotor epileptiform epilepsy	Localization-related (focal) (partial) epileps...	lacosamide	2210-08-24 19:43:00	1
3	V463 Wheelchair dependence	Wheelchair dependence	Nystatin Cream	2210-08-24 19:43:00	1
4	3341 Hereditary spastic paraplegia	Hereditary spastic paraplegia	Ranitidine	2210-08-24 19:43:00	1
5	3694 Legal blindness-usa def	Legal blindness, as defined in U.S.A.	Ascorbic Acid (Liquid)	2210-08-24 19:43:00	1
6	34540 Psychomotor epileptiform epilepsy	Localization-related (focal) (partial) epileps...	Silver Sulfadiazine 1% Cream	2210-08-24 19:43:00	1
7	V441 Gastrostomy status	Gastrostomy status	budesonide	2210-08-24 19:43:00	1
8	9341 Foreign body bronchus	Foreign body in main bronchus	Zonisamide	2210-08-24 19:43:00	1
9	51881 Acute respiratory failure	Acute respiratory failure	scopolamine base	2210-08-24 19:43:00	1

Query in notes below

```

select
  subj_diag.icd9_code,
  diag.short_title,
  diag.long_title,
  rx.drug,
  max(dischtime) as Discharge_Date,
  count(distinct subj_diag.subject_id) as Patients_Affected
from DIAGNOSES_ICD subj_diag
inner join D_ICD_DIAGNOSES diag
  on subj_diag.ICD9_CODE = diag.ICD9_CODE
inner join PRESCRIPTIONS rx
  on subj_diag.subject_id = rx.subject_id
  and subj_diag.hadm_id = rx.hadm_id
inner join ADMISSIONS adm
  on adm.subject_id = subj_diag.subject_id
  and adm.hadm_id = subj_diag.hadm_id
--where subj_diag.subject_id in (3386, 2187, 3417)
group by 1,2,3,4
order by 5 desc, 6

```

```
limit 10
```

BONUS – WHAT’S THE MOST FREQUENT MED/DIAGNOSIS COMBINATION?

- We join the DIAGNOSES_ICD and PRESCRIPTIONS table
- We count distinct admissions and patients observing that combination, but sorting by descending admission count
- Perhaps not surprising, many heart-related ailments are present

ICD9 CODE	SHORT TITLE	LONG TITLE	DRUG	Admissions_with_Combo	Patients_with_Combo
0	4019 Hypertension NOS	Unspecified essential hypertension	Sodium Chloride 0.9% Flush	55841	12796
1	4019 Hypertension NOS	Unspecified essential hypertension	Acetaminophen	15000	13170
2	4019 Hypertension NOS	Unspecified essential hypertension	Potassium Chloride	14927	13212
3	4019 Hypertension NOS	Unspecified essential hypertension	Magnesium Sulfate	13638	12320
4	4019 Hypertension NOS	Unspecified essential hypertension	Insulin	13632	12036
5	4019 Hypertension NOS	Unspecified essential hypertension	Docusate Sodium	12427	11142
6	4019 Hypertension NOS	Unspecified essential hypertension	Heparin	12033	10433
7	4019 Hypertension NOS	Unspecified essential hypertension	Furosemide	10814	9747
8	4019 Hypertension NOS	Unspecified essential hypertension	DSW	10780	9841
9	4019 Hypertension NOS	Unspecified essential hypertension	Morphine Sulfate	10270	9589
10	4019 Hypertension NOS	Unspecified essential hypertension	Iso-Osmotic Dextrose	10252	9186
11	4019 Hypertension NOS	Unspecified essential hypertension	SW	10087	9294
12	4019 Hypertension NOS	Unspecified essential hypertension	NS	10012	8983
13	42731 Atrial Fibrillation	Atrial Fibrillation	Sodium Chloride 0.9% Flush	9865	8058
14	4280 CHF NOS	Congestive heart failure, unspecified	Sodium Chloride 0.9% Flush	9638	7368
15	4019 Hypertension NOS	Unspecified essential hypertension	Calcium Gluconate	9605	9032
16	42731 Atrial Fibrillation	Atrial Fibrillation	Potassium Chloride	9604	8103
17	42731 Atrial Fibrillation	Atrial Fibrillation	Acetaminophen	9602	8079
18	4280 CHF NOS	Congestive heart failure, unspecified	Furosemide	9488	7576
19	4019 Hypertension NOS	Unspecified essential hypertension	Bisacodyl	9430	8693

Query in notes below

```
select
diag.icd9_code,
diag_text.short_title,
diag_text.long_title,
rx.drug,
count(distinct diag.hadm_id) as Admissions_with_Combo,
count(distinct diag.subject_id) as Patients_with_Combo
from DIAGNOSES_ICD diag
inner join D_ICD_DIAGNOSES diag_text
    on diag.icd9_code = diag_text.icd9_code
inner join PRESCRIPTIONS rx
    on diag.hadm_id = rx.hadm_id
group by 1,2,3,4
order by 5 desc
```