Healthcare Data Analysis using SQL

By- Ankush

www.linkedin.com/in/ankushsinghpec/

www. github.com/Ankush-Singh-1/Healthcare-Data-Analysis-SQL

Business Case Scenario

A mid-sized, multi-branch medical center handles thousands of patient admissions yearly. With rising healthcare costs and increasing patient volume, hospital management seeks to leverage data-driven insights to optimize operations, reduce expenses, and improve patient care outcomes. Currently, the hospital lacks clear visibility into billing trends, treatment patterns, and resource utilization across departments.

As a data analyst, my role is to analyze hospital data to uncover patterns in billing, length of stay, medication usage, and insurance behavior. The insights generated will guide strategic decisions such as renegotiating insurance contracts, streamlining treatment protocols, and improving patient turnover times, ultimately enhancing operational efficiency and patient care quality.

Steps taken to solve

->Data Exploration – 1.check entire data

2. check unique category

3. Summary stats (Min,Max,Avg)

->Data cleaning – 1. check null and remove if there is a null value.

2. check for duplicates and remove them or update.

->Data analysis – solve business problems asked by hospital.

Data Exploration

```sql
-- check entire data
select * from patient_data;
```

Data Output | Messages | Notifications

Showing rows: 1 to 1000 | Page No: 1 of 55

	name text	age integer	gender text	blood type text	medical condition text	date of admission date	doctor text	hospital text	insu text
1	brandon collins	77	Female	O+	Asthma	2019-07-19	Cameron Young	Lopez PLC	Cig
2	todd cooper	55	Female	A+	Diabetes	2022-12-10	Jason Price	Henderson-Taylor	Uni
3	matthew white	45	Female	AB+	Diabetes	2020-06-03	Ryan Ross	Hampton-Thomas	Me
4	chad huff	76	Female	AB+	Hypertension	2024-03-12	Bob Moyer	Clark-Tran	Aet

```sql
-- check unique category
select distinct gender as gender_category
from patient_data
select distinct "blood type" as blood_type_category
from patient_data
```

	gender_category text
1	Female
2	Male

	blood_type_category text
1	B+
2	O-
3	AB-
4	AB+

Total rows: 8 | Query com

```sql
-- Summary stats
select max(age) from patient_data
select min(age) from patient_data
select avg("billing amount") from patient_data
```

	max integer
1	89

	min integer
1	13

	avg double precision
1	25544.306284384216

```
----------------------------------------------------Data cleaning-----------------------------------------------------
select * from patient_data
where "name" IS NULL OR age IS NULL OR gender IS NULL OR "blood type" IS NULL OR
      "medical condition" IS NULL OR "date of admission" IS NULL OR doctor IS NULL OR
      hospital IS NULL OR "insurance provider" IS NULL OR "billing amount" IS NULL OR
      "room number" IS NULL OR "admission type" IS NULL OR "discharge date" IS NULL OR
      medication IS NULL OR "test results" IS NULL;
```

name text	age integer	gender text	blood type text	medical condition text	date of admission date	doctor text	hospital text	insurance provider text	billing amount double precision	room number integer

*No null value were present in data

```
--- check for duplicates
select "name", age,"blood type",gender,"insurance provider","discharge date","date of admission"
       ,"billing amount",count(*) as record_count
from patient_data
group by name,age,"blood type",gender,"insurance provider","discharge date","billing amount","date of admission"
Having count(*)>1;
```

*No ouput here has duplicates were removed when code was run for first time back then.

```sql
--- Remove Duplicates
With Dupli As(
     select ctid,
     Row_number() over (partition by "name",age
     order by "name",age) as rn
     from patient_data
)
delete from patient_data
where ctid in (select ctid from Dupli where rn>1)
```

```
DELETE 0

Query returned successfully in 672 msec.
```

```sql
--- standardize text column
UPDATE patient_data
SET name = LOWER(name);

---Extract year from date of admission
select Extract(year from "date of admission")
from patient_data

---Extract month from date of admission
select Extract(month from "date of admission")
from patient_data
```

```
UPDATE 54615

Query returned successfully in 1 secs 27 msec.
```

	extract numeric
1	2021
2	2020
3	2020
4	2023

	extract numeric
1	8
2	5
3	5
4	12

Business Questions

```sql
--- 1. What medical condition has highest average billing amount?

select "medical condition", count(*) as no_of_patient,avg("billing amount") as avg_billing_amount
from patient_data
group by "medical condition"
order by "avg_billing_amount" Desc
```

	medical condition text	no_of_patient bigint	avg_billing_amount double precision
1	Obesity	9077	25800.49775556358
2	Diabetes	9161	25648.132856771696
3	Asthma	9041	25636.30909668471
4	Arthritis	9165	25510.6031317171

```sql
--- 2. what is the average length of stay per admission type?

with stay_data as(
select
     "admission type",age("discharge date","date of admission") as length_of_stay
from patient_data)

select "admission type",Round(avg(EXTRACT(day FROM length_of_stay)),1) as avg_length_stay
from stay_data
group by "admission type";
```

	admission type text	avg_length_stay numeric
1	Elective	15.0
2	Urgent	14.9
3	Emergency	15.1

```sql
--- 3. Which insurance providers cover the highest cost cases?

select
"insurance provider" , Round(sum("billing amount")::numeric,2) as total_billing
from patient_data
group by "insurance provider"
order by total_billing desc
```

	insurance provider text	total_billing numeric
1	Cigna	282858083.67
2	Medicare	281018956.31
3	Blue Cross	278584390.98
4	UnitedHealthcare	277687200.34
5	Aetna	274781274.28

```sql
--- 4. what are the most common medication used per diagnosis?

select "medical condition", medication , count(*) as usage_count
from patient_data
group by "medical condition",medication
order by "medical condition" , usage_count desc
```

	medical condition text	medication text	usage_count bigint
1	Arthritis	Aspirin	1893
2	Arthritis	Paracetamol	1844
3	Arthritis	Penicillin	1837
4	Arthritis	Ibuprofen	1798
5	Arthritis	Lipitor	1793
6	Asthma	Paracetamol	1858
7	Asthma	Penicillin	1818
8	Asthma	Lipitor	1801
9	Asthma	Ibuprofen	1794

Total rows: 30 | Query complete 00:00:00.299

```sql
--- 5. what is the average time to discharge by hospital?
with time_data as (select hospital, age("discharge date", "date of admission") as discharge_time
from patient_data)
select hospital,avg(discharge_time) as avg_discharge_time
from time_data
group by hospital
order by avg_discharge_time desc;
```

	hospital text	avg_discharge_time interval
1	Johnson-Vargas	1 mon 2 days
2	Rollins, Vaughn and Carter	1 mon 2 days
3	Gonzalez-Hill	1 mon 2 days
4	Booth LLC	1 mon 2 days
5	Huff, Mendoza and Smith	1 mon 2 days
6	Bond-Gomez	1 mon 2 days
7	Barnes-Bradley	1 mon 2 days
8	Hall Gonzalez and Duarte,	1 mon 2 days
9	Harding-Adams	1 mon 2 days

Total rows: 39674 | Query complete 00:00:00.290

Conclusion

Through this analysis, we've uncovered valuable insights into hospital operations, highlighting opportunities to reduce costs, optimize care delivery, and enhance patient outcomes. By leveraging data on billing, treatment patterns, and insurance behavior, the hospital can make smarter, faster decisions—driving efficiency, improving care quality, and supporting long-term strategic goals.


Thank
You