



---

# SQL QUERIES - MIMIC-III

Cole Stokes

AI in Healthcare

The University of Texas at Austin

[Google Colab](#) - [GitHub](#)

---

---

# DESCRIPTION

This presentation will take you through SQL queries to the MIMIC-III dataset designed to gain insight and tell a story about Parkinson's disease. This was done by:

- Using a Google Colab Jupyter Notebook and Python.
- Loading and querying the data through Google BigQuery and SQL.
- Displaying the tables through Pandas.

These will show both demographic and medical takeaways pertaining to Parkinson's disease.

Note: This dataset is only limited to patients admitted to one hospital, the Beth Israel Deaconess Medical Center in Boston, Massachusetts.

---



---

# QUERY 1 – GENDER DISTRIBUTION OF PARKINSON'S PATIENTS

```
SELECT gender, COUNT(*) AS num_patients
FROM `physionet-data.mimiciii_clinical.patients`
WHERE subject_id IN (SELECT subject_id
FROM physionet-
data.mimiciii_clinical.diagnoses_icd WHERE
icd9_code = '3320')
GROUP BY gender;
```

	gender	num_patients
0	F	143
1	M	254



This query selects the counts of patients aggregated by gender from the patients table. It filters the data to only include subjects that have the 3320 (Parkinson's disease or paralysis agitans) icd9 code in the diagnoses\_icd table.



This table shows us that a much higher number of males were admitted to the hospital who are diagnosed with Parkinson's than females. This shows us the males are more likely to be diagnosed with Parkinson's than females.

---

# QUERY 2 – PARKINSON’S DISEASE AND ETHNICITY

```
WITH ethnicity_distribution AS (  
    SELECT ethnicity, COUNT(DISTINCT subject_id) AS num_patients  
    FROM physionet-data.mimiciii_clinical.admissions  
  
    WHERE subject_id IN (SELECT DISTINCT subject_id FROM physionet-data.mimiciii_clinical.diagnoses_icd  
    WHERE icd9_code = '3320')  
  
    GROUP BY ethnicity),  
total_ethnicity_admissions AS (SELECT ethnicity, COUNT(DISTINCT subject_id) AS total_admissions  
    FROM physionet-data.mimiciii_clinical.admissions  
  
    GROUP BY ethnicity)  
  
SELECT e.ethnicity, e.num_patients, t.total_admissions, (e.num_patients / t.total_admissions) AS  
percentage_for_ethnicity  
FROM ethnicity_distribution e  
  
JOIN total_ethnicity_admissions t ON e.ethnicity = t.ethnicity  
  
ORDER BY percentage_for_ethnicity DESC;
```

	ethnicity	num_patients	total_admissions	percentage_for_ethnicity
0	ASIAN - JAPANESE	1	7	0.142857
1	HISPANIC/LATINO - SALVADORAN	1	15	0.066667
2	ASIAN - CHINESE	6	231	0.025974
3	AMERICAN INDIAN/ALASKA NATIVE	1	45	0.022222
4	MULTI RACE ETHNICITY	2	112	0.017857
5	ASIAN - ASIAN INDIAN	1	58	0.017241
6	WHITE - OTHER EUROPEAN	1	71	0.014085
7	BLACK/HAITIAN	1	75	0.013333
8	HISPANIC/LATINO - PUERTO RICAN	2	156	0.012821
9	WHITE	317	32192	0.009847
10	UNABLE TO OBTAIN	7	794	0.008816
11	UNKNOWN/NOT SPECIFIED	27	4241	0.006366
12	ASIAN	6	1306	0.004594
13	OTHER	6	1316	0.004559
14	PATIENT DECLINED TO ANSWER	2	499	0.004008
15	BLACK/AFRICAN AMERICAN	14	3606	0.003882
16	HISPANIC OR LATINO	5	1362	0.003671

This query creates two new tables `ethnicity_distribution` and `total_ethnicity_admissions` to get the counts of patients with Parkinson’s aggregated by ethnicity and the total admission by ethnicity respectively. These tables are joined to show the number and percentage of Parkinson’s patients admitted to the hospital for each ethnicity.

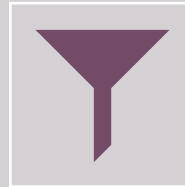
This table shows the number of patients with Parkinson’s compared to the total number of patients admitted and the percentage for each ethnicity. Since some ethnicities have small sample sizes and there are many ethnicity options that overlap, we may not be able to make representative assumptions about these.

---

# QUERY 3 – AGE DISTRIBUTION OF PARKINSON'S PATIENTS AT FIRST ADMISSION

```
WITH first_admission AS (  
    SELECT p.subject_id, DATE_DIFF(DATE(MIN(a.admtime)),  
    ARRAY_AGG(p.dob ORDER BY a.admtime LIMIT 1)[OFFSET(0)], YEAR)  
    AS age_at_first_admission  
    FROM physionet-data.mimiciii_clinical.patients p  
    JOIN physionet-data.mimiciii_clinical.admissions a ON p.subject_id =  
    a.subject_id  
    WHERE p.subject_id IN (SELECT subject_id FROM physionet-  
    data.mimiciii_clinical.diagnoses_icd WHERE icd9_code = '3320')  
    GROUP BY p.subject_id  
    HAVING age_at_first_admission <= 110)  
SELECT MIN(age_at_first_admission) AS min_age,  
APPROX_QUANTILES(age_at_first_admission, 4)[OFFSET(1)] AS Q1,  
AVG(age_at_first_admission) AS avg_age,  
APPROX_QUANTILES(age_at_first_admission, 4)[OFFSET(3)] AS Q3,  
MAX(age_at_first_admission) AS max_age  
FROM first_admission;
```

	min_age	Q1	avg_age	Q3	max_age
0	34	70	75.582857	83	89



This query defines a new table that takes each Parkinson's patient's first admission and gets their ages, removing unreasonable numbers. Then it takes this table and aggregates these ages to show the minimum, Q1, average, Q3, and maximum ages of the admitted Parkinson's patients.



This table shows that half of the admitted Parkinson's patients are between 70 and 83 years old. We can see that hospital admission for Parkinson's can happen for patients that are a little older and much younger. However, this distribution shows that the disease tends to seriously effect older populations more.

---

---

# QUERY 4 – PARKINSON’S READMISSION

```
WITH admission_data AS (  
    SELECT subject_id, COUNT(DISTINCT hadm_id) AS  
total_admissions  
    FROM physionet-data.mimiciii_clinical.admissions  
    WHERE subject_id IN (SELECT subject_id FROM  
physionet-data.mimiciii_clinical.diagnoses_icd WHERE  
icd9_code = '3320')  
    GROUP BY subject_id)  
SELECT COUNT(DISTINCT subject_id) AS  
total_patients, SUM(total_admissions) AS total_admissions,  
COUNT(CASE WHEN total_admissions > 1 THEN  
subject_id END) AS readmitted_patients, COUNT(CASE  
WHEN total_admissions >= 3 THEN subject_id END) AS  
patients_admitted_atleast_3_times  
FROM admission_data;
```

This query defines a new table `admission_data` that gets the `subject_id` and their counts of admissions for Parkinson’s patients. The table is then queried to show the counts of the total number of patients, the total number of admissions, the number of patients that were readmitted, and the number of patients admitted three or more times.

We can see from the results that there are about 1.5 admissions for every Parkinson’s patient, more than 1 in 4 patients get readmitted, and over 10% of them continue to get readmitted.

	total_patients	total_admissions	readmitted_patients	patients_admitted_atleast_3_times
0	397	595	109	43



---

# QUERY 5 – AVERAGE TIME BEFORE READMISSION FOR PARKINSON'S PATIENTS

WITH readmission\_times AS (

```
SELECT subject_id, TIMESTAMP_DIFF(LEAD(admittime)
OVER(PARTITION BY subject_id ORDER BY admittime),
disctime, DAY) AS days_before_readmission
```

```
FROM physionet-data.mimiciii_clinical.admissions
```

```
WHERE subject_id IN (SELECT subject_id FROM physionet-
data.mimiciii_clinical.diagnoses_icd WHERE icd9_code =
'3320'))
```

```
SELECT AVG(days_before_readmission) AS
avg_days_before_readmission
```

```
FROM readmission_times
```

```
WHERE days_before_readmission IS NOT NULL;
```

avg_days_before_readmission	
0	441.075758



This query defines a new table that gets the times before between the end of their admission and the start of their next if applicable. These days are then selected and averaged.



This result shows us that on average a Parkinson's patient will have over one year after an admission before something happens that causes them to be readmitted.

---

---

# QUERY 6 – LENGTH OF STAY OF PARKINSON’S PATIENTS AND OTHER PATIENTS

```
SELECT CASE WHEN subject_id IN (SELECT subject_id
FROM physionet-data.mimiciii_clinical.diagnoses_icd
WHERE icd9_code = '3320')
THEN 'Parkinsons Patients'
ELSE 'Other Patients'
END AS patient_group,
AVG(TIMESTAMP_DIFF(disctime, admittance, Day))
AS avg_length_of_stay_days
FROM physionet-data.mimiciii_clinical.admissions
GROUP BY patient_group;
```

	patient_group	avg_length_of_stay_days
0	Other Patients	10.095802
1	Parkinsons Patients	9.596639

This query classifies all patients into a Parkinson’s group or other group and calculates the time in days between discharge and admittance. This is aggregated by the classes and the average length of stay is given for both groups.

The table shows that there is not a significant difference between the length of stay for Parkinson’s patients and other patients.

---

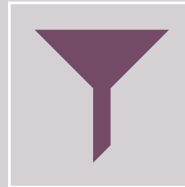


---

# QUERY 7 – TOP CO-OCCURRING DIAGNOSES WITH PARKINSON’S

```
SELECT d2.icd9_code, dicd.long_title, COUNT(DISTINCT d2.subject_id) AS co_occurrence_count
FROM physionet-data.mimiciii_clinical.diagnoses_icd d
JOIN physionet-data.mimiciii_clinical.diagnoses_icd d2 ON d.subject_id = d2.subject_id
JOIN physionet-data.mimiciii_clinical.d_icd_diagnoses dicd ON d2.icd9_code = dicd.icd9_code
WHERE d.icd9_code = '3320' AND d2.icd9_code != '3320'
GROUP BY d2.icd9_code, dicd.long_title
ORDER BY co_occurrence_count DESC
LIMIT 10;
```

	icd9_code	long_title	co_occurrence_count
0	4019	Unspecified essential hypertension	202
1	4280	Congestive heart failure, unspecified	132
2	5849	Acute kidney failure, unspecified	130
3	42731	Atrial fibrillation	128
4	5990	Urinary tract infection, site not specified	115
5	41401	Coronary atherosclerosis of native coronary ar...	104
6	5070	Pneumonitis due to inhalation of food or vomitus	103
7	51881	Acute respiratory failure	98
8	25000	Diabetes mellitus without mention of complicat...	86
9	2724	Other and unspecified hyperlipidemia	82



This query joins the diagnoses\_icd with itself where one is only Parkinson’s disease and the other is everything else. The d\_icd\_diagnoses table is also joined to get the titles of these codes. The ten highest counts of co-occurrences, their codes, and titles are given descending by count.



This query shows us that diagnoses such as hypertension, heart failure, kidney failure, and atrial fibrillation are common to occur alongside Parkinson’s.

---

---

# QUERY 8 – TOP MEDICATIONS GIVEN TO PARKINSON’S PATIENTS

```
SELECT pr.drug, COUNT(DISTINCT pr.subject_id) AS
num_patients_prescribed

FROM physionet-data.mimiciii_clinical.prescriptions pr

JOIN physionet-data.mimiciii_clinical.diagnoses_icd d ON pr.subject_id =
d.subject_id

WHERE d.icd9_code = '3320'

GROUP BY pr.drug

ORDER BY num_patients_prescribed DESC

LIMIT 10;
```

	drug	num_patients_prescribed
0	Sodium Chloride 0.9% Flush	315
1	Acetaminophen	308
2	Potassium Chloride	300
3	Heparin	294
4	Magnesium Sulfate	259
5	Insulin	255
6	Iso-Osmotic Dextrose	254
7	Docusate Sodium	241
8	NS	239
9	D5W	233

This query takes Parkinson’s patients from diagnoses\_icd and joins with the prescriptions table. This gets the drug names and the counts of patients that were prescribed each drug. This query is limited to 10 and is ordered descending by count.

This table shows that sodium chloride, acetaminophen, potassium chloride, heparin, and magnesium sulfate are commonly prescribed to Parkinson’s patients while in the hospital.

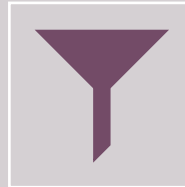
---

---

# QUERY 9 – MOST FREQUENT PROCEDURES TO PARKINSON’S PATIENTS

```
SELECT pr.icd9_code, dicd.long_title AS procedure_description,  
COUNT(DISTINCT pr.subject_id) AS num_patients  
FROM `physionet-data.mimiciii_clinical.procedures_icd` pr  
JOIN `physionet-data.mimiciii_clinical.diagnoses_icd` d ON pr.subject_id =  
d.subject_id  
JOIN `physionet-data.mimiciii_clinical.d_icd_procedures` dicd ON pr.icd9_code  
= dicd.icd9_code  
WHERE d.icd9_code = '3320'  
GROUP BY pr.icd9_code, dicd.long_title
```

	icd9_code	procedure_description	num_patients
0	3893	Venous catheterization, not elsewhere classified	143
1	9604	Insertion of endotracheal tube	106
2	966	Enteral infusion of concentrated nutritional s...	106
3	9671	Continuous invasive mechanical ventilation for...	90
4	9904	Transfusion of packed cells	84
5	9672	Continuous invasive mechanical ventilation for...	68
6	3891	Arterial catheterization	45
7	8856	Coronary arteriography using two catheters	43
8	3961	Extracorporeal circulation auxiliary to open h...	42
9	4311	Percutaneous [endoscopic] gastrostomy [PEG]	35



This query selects the icd9 codes, procedure titles, and counts of Parkinson’s patient’s that had each procedure. This is done by joining procedures\_icd, diagnoses\_icd, and d\_icd\_procedures. The procedures with the top ten counts are shown in descending order.



This table shows that venous catheterization, inserting endotracheal tubes, and others are common procedures preformed on Parkinson’s patients while in the hospital.

---

---

# QUERY 10 – IN-HOSPITAL MORTALITY RATE OF PARKINSON’S PATIENTS AND OTHER PATIENTS

```
SELECT CASE WHEN subject_id IN (SELECT subject_id FROM physionet-
data.mimiciii_clinical.diagnoses_icd WHERE icd9_code = '3320')
    THEN 'Parkinsons Patients'
    ELSE 'Other Patients'
END AS patient_group,
COUNT(DISTINCT subject_id) AS num_patients,
SUM(CASE WHEN hospital_expire_flag = 1 THEN 1 ELSE 0 END) AS
num_deaths,
SUM(CASE WHEN hospital_expire_flag = 1 THEN 1 ELSE 0 END) /
COUNT(DISTINCT subject_id) AS mortality_rate
FROM physionet-data.mimiciii_clinical.admissions
GROUP BY patient_group;
```

	patient_group	num_patients	num_deaths	mortality_rate
0	Other Patients	46123	5770	0.125100
1	Parkinsons Patients	397	84	0.211587

This query classifies patients to a Parkinson’s group and others group. It aggregates by this and gets the counts of patients, the number of deaths, and calculates the mortality rate for each group.

We can see that it is almost twice as likely for a Parkinson’s patient to pass while admitted to a hospital than other patients. This shows that over 1 in 5 Parkinson’s patients pass away during one of their admissions to the hospital.

---