The Data Prescription: An SQL Case Study on Healthcare Records

DATASET INFORMATION

Each column provides specific information about the patient, their admission, and the healthcare services provided, making this dataset suitable for various data analysis and modelling tasks in the healthcare domain. Here's a brief explanation of each column in the dataset –

- **Name:** This column represents the name of the patient associated with the healthcare record.
- **Age:** The age of the patient at the time of admission, expressed in years.
- **Gender:** Indicates the gender of the patient, either "Male" or "Female."
- **Blood Type:** The patient's blood type, which can be one of the common blood types (e.g., "A+", "O-", etc.).
- **Medical Condition:** This column specifies the primary medical condition or diagnosis associated with the patient, such as "Diabetes," "Hypertension," "Asthma," and more.
- **Date of Admission:** The date on which the patient was admitted to the healthcare facility.
- **Doctor:** The name of the doctor responsible for the patient's care during their admission.
- **Hospital:** Identifies the healthcare facility or hospital where the patient was admitted.
- **Insurance Provider:** This column indicates the patient's insurance provider, which can be one of several options, including "Aetna," "Blue Cross," "Cigna," "UnitedHealthcare," and "Medicare."
- **Billing Amount:** The amount of money billed for the patient's healthcare services during their admission. This is expressed as a floating-point number.
- **Room Number:** The room number where the patient was accommodated during their admission.
- **Admission Type:** Specifies the type of admission, which can be "Emergency," "Elective," or "Urgent," reflecting the circumstances of the admission.
- **Discharge Date:** The date on which the patient was discharged from the healthcare facility, based on the admission date and a random number of days within a realistic range.
- **Medication:** Identifies a medication prescribed or administered to the patient during their admission. Examples include "Aspirin," "Ibuprofen," "Penicillin," "Paracetamol," and "Lipitor."
- **Test Results:** Describes the results of a medical test conducted during the patient's admission. Possible values include "Normal," "Abnormal," or "Inconclusive," indicating the outcome of the test.

Tech-stack Used:

- Dataset: Healthcare Dataset (kaggle.com)
- Creating SQL Database
- Testing SQL Database
- Executing SQL query
- Perform Analysis
- Version – PostgreSQL v16.1

QUESTIONNAIRE:

1. Retrieve the total number of admissions for each medical condition.
2. Find the average billing amount for patients in each age group (e.g., 0-10, 11-20, etc.).
3. List the top 3 most common blood types among patients.
4. Find the 3 medical condition with the highest average billing amount.
5. Retrieve the names of doctors who have treated patients in more than one hospital.
6. Identify patients who have undergone both normal and abnormal tests.
7. List the hospitals where the majority of admissions are elective.
8. Identify patients with the longest and shortest lengths of stay in the hospital.
9. Find the top 5 hospitals where the billing amount has increased the most compared to the previous year.
10. List the medical conditions where the average age of patients is below the overall average age.
11. Identify the patients with the highest and lowest billing amounts within their respective medical conditions.
12. Find pairs of patients who share the same room number and calculate the average age of each pair.

Q1. Retrieve the total number of admissions for each medical condition.

Input SQL Query:

```
SELECT
      medicalcondition,
      count(*) as total_admissions
FROM
     health
GROUP BY
   medicalcondition
ORDER BY
   total_admissions DESC;
```

Output SQL Query:

	medicalcondition character varying (255)	total_admissions bigint
1	Asthma	1708
2	Cancer	1703
3	Hypertension	1688
4	Arthritis	1650
5	Obesity	1628
6	Diabetes	1623

Conclusion: The result reveals a varied spectrum of medical conditions, with a focus on respiratory issues like "Asthma." The dataset's uniform admission counts across conditions such as "Cancer" and chronic illnesses highlight consistent healthcare demands. This insight is valuable for resource planning and underscores the importance of a comprehensive healthcare approach to cater to diverse patient needs.

Q2. *Find the average billing amount for patients in each age group (e.g., 0-10, 11-20, etc.).*

Input SQL Query:

```
SELECT
      CASE
          WHEN age between 0 AND 10 THEN '0-10'
          WHEN age between 11 AND 20 THEN '11-20'
          WHEN age between 21 AND 30 THEN '21-30'
          WHEN age between 31 AND 40 THEN '31-40'
          WHEN age between 41 AND 50 THEN '41-50'
          WHEN age between 51 AND 60 THEN '51-60'
          else 'greater than 60'
       END AS Age_group,
       AVG(billingamount) as average_billingamount
FROM
    health
GROUP BY
    Age_group
ORDER BY
    Age_group desc
```

Output SQL Query:

	age_group text	average_billingamount double precision
1	greater than 60	25348.617353506223
2	51-60	25563.26120908465
3	41-50	25745.494119493913
4	31-40	25452.476735419266
5	21-30	25546.090811690054
6	11-20	26121.23155123046

Conclusion: The analysis of average billing amounts across age groups suggests a general increase with age, peaking in the "11-20" range. However, costs stabilize in adulthood and decrease in the elderly group. This insight can inform resource allocation strategies for addressing varying healthcare needs across different age demographics.

Q3. *List the top 3 most common blood types among patient*.

Input SQL Query:

```
SELECT
     bloodtype as Blood_group,
     COUNT(*) as Number_of_patirnts
FROM
   health
GROUP BY
      Blood_group
ORDER BY
      Number_of_patirnts desc
LIMIT 3
```

Output SQL Query:

	blood_group character varying (5)	number_of_patirnts bigint
1	AB-	1275
2	AB+	1258
3	B-	1252

Conclusion: The analysis of blood types among patients reveals that "AB-" is the most common blood type, followed closely by "AB+" and "B-." This information can be valuable for healthcare providers to ensure adequate blood supply and compatibility for patient care.

Q4. Find the 3 medical condition with the highest average billing amount.

Input SQL Query:

```
SELECT
      medicalcondition,
      avg(billingamount) as average_billingamount
from
    health
GROUP BY
     medicalcondition
ORDER BY
     average_billingamount desc
LIMIT 3
```

Output SQL Query:

	medicalcondition character varying (255)	average_billingamount double precision
1	Diabetes	26060.116129178103
2	Obesity	25720.84268283536
3	Cancer	25539.096132609557

Conclusion: The analysis of average billing amounts for different medical conditions indicates that patients with "Diabetes" have the highest average billing amount, followed by "Obesity" and "Cancer." This suggests that managing healthcare costs for these conditions may require specific attention and resource allocation.

Q5. Retrieve the names of doctors who have treated patients in more than one hospital.

Input SQL Query:

```
SELECT
      doctor,
      COUNT(DISTINCT hospital) as number_of_hospital
from
    health
GROUP BY
    doctor
HAVING
     COUNT(DISTINCT hospital)>1
LIMIT 10;
```

Output SQL Query:

	doctor character varying (255)	number_of_hospital bigint
1	Aaron Smith	2
2	Abigail Smith	2
3	Adam Martinez	2
4	Alyssa Morgan	2
5	Amanda Hernandez	2
6	Amanda James	2
7	Amanda Johnson	2
8	Amanda Jones	2
9	Amanda Shaw	2
10	Amanda Walker	2

Conclusion: The analysis indicates that certain doctors, including "Aaron Smith," "Abigail Smith," and others, have treated patients in more than one hospital.

Q6. Identify patients who have undergone both normal and abnormal tests.

Input SQL Query:

```sql
SELECT
    name,
    COUNT(DISTINCT CASE WHEN testresults = 'Normal' THEN TestResults END) AS NormalTests,
    COUNT(DISTINCT CASE WHEN testresults = 'Abnormal' THEN TestResults END) AS AbnormalTests
FROM
    Health
WHERE
    testresults IN ('Normal', 'Abnormal')
GROUP BY
    name
HAVING
    COUNT(DISTINCT CASE WHEN testresults = 'Normal' THEN TestResults END) > 0
    AND COUNT(DISTINCT CASE WHEN testresults = 'Abnormal' THEN TestResults END) > 0
LIMIT 10
```

Output SQL Query:

	name character varying (255)	normaltests bigint	abnormaltests bigint
1	Aaron Miller	1	1
2	Aaron Patel	1	1
3	Amanda Bailey	1	1
4	Amanda Lee	1	1
5	Amanda Williams	1	1
6	Angela Garcia	1	1
7	Ann Fisher	1	1
8	Anthony Butler	1	1
9	Anthony Steele	1	1
10	Anthony Thompson	1	1

Conclusion: The data indicates that patients such as "Aaron Miller," "Amanda Bailey," and others have undergone both normal and abnormal tests. This suggests a diverse range of medical conditions or testing scenarios for these patients, emphasizing the need for comprehensive healthcare monitoring and diagnostics tailored to individual patient needs.

Q7. List the hospitals where the majority of admissions are elective.

Input SQL Query:

```
SELECT
    Hospital,
    COUNT(*) AS TotalAdmissions,
    SUM(CASE WHEN AdmissionType = 'Elective' THEN 1 ELSE 0 END) AS ElectiveAdmissions,
    ROUND((SUM(CASE WHEN AdmissionType='Elective' THEN 1 ELSE 0 END)*100.0)/COUNT(*))AS PercentageElectiveAdmissions
FROM
    Health
GROUP BY
    Hospital
HAVING
    (SUM(CASE WHEN AdmissionType = 'Elective' THEN 1 ELSE 0 END) * 100.0) / COUNT(*) > 50
LIMIT 10
```

Output SQL Query:

	hospital character varying (255)	totaladmissions bigint	electiveadmissions bigint	percentageelectiveadmissions numeric
1	Wheeler, Flowers and Moore	1	1	100
2	Wise PLC	1	1	100
3	Gibson, Mueller and Rich	1	1	100
4	Wilson, Fernandez and Christian	1	1	100
5	Peterson, Banks and Walker	1	1	100
6	Mccoy, Harris and Freeman	1	1	100
7	Spencer-Rogers	1	1	100
8	Jones, Watson and Chen	1	1	100
9	White Ltd	6	4	67
10	Campbell-Parks	1	1	100

Conclusion: The analysis reveals hospitals such as "Wheeler, Flowers and Moore," "Wise PLC," and others where the majority of admissions are elective, accounting for 100% of total admissions. This suggests a specific focus on planned, non-emergency procedures in these healthcare facilities, emphasizing their role in providing elective medical services.

Q8. Identify patients with the longest and shortest lengths of stay in the hospital.

Input SQL Query:

```
(SELECT
    Name,
    Dateofadmission,
    DischargeDate,
    (DischargeDate - DateOfAdmission) AS LengthOfStay
FROM
    Health
ORDER BY
    LengthOfStay DESC
LIMIT 1)

UNION

(SELECT
    Name,
    DateOfAdmission,
    DischargeDate,
    (DischargeDate - DateOfAdmission) AS LengthOfStay
FROM
    Health
ORDER BY
    LengthOfStay ASC
LIMIT 1)
```

Output SQL Query:

	name character varying (255)	dateofadmission date	dischargedate date	lengthofstay integer
1	Antonio Frederick	2020-05-02	2020-05-03	1
2	Chad Byrd	2019-01-09	2019-02-08	30

Conclusion: The analysis of hospital stay durations reveals that "Chad Byrd" had the longest length of stay, spanning 30 days from admission on January 9th to discharge on February 8th. In contrast, "Antonio Frederick" had the shortest stay, with only 1 day from admission on May 2nd to discharge on May 3rd.

Q9. Find the top 5 hospitals where the billing amount has increased the most compared to the previous year.

Input SQL Query:

```sql
WITH BillingAmountChanges AS (
    SELECT
        Hospital,
        EXTRACT(YEAR FROM DateOfAdmission) AS AdmissionYear,
        SUM(BillingAmount) AS TotalBillingAmount
    FROM
        Health
    GROUP BY
        Hospital, EXTRACT(YEAR FROM DateOfAdmission)
)

SELECT
    Hospital,
    MAX(TotalBillingAmount) AS MaxBillingAmount,
    MIN(TotalBillingAmount) AS MinBillingAmount,
    MAX(TotalBillingAmount) - MIN(TotalBillingAmount) AS BillingAmountIncrease
FROM
    BillingAmountChanges
GROUP BY
    Hospital
ORDER BY
    BillingAmountIncrease DESC
LIMIT 5;
```

Output SQL Query:

	hospital character varying (255)	maxbillingamount double precision	minbillingamount double precision	billingamountincrease double precision
1	Smith Inc	165009.68569709788	8424.652221687034	156585.03347541086
2	Smith Ltd	184175.81078057154	44338.57425382923	139837.23652674232
3	Thomas and Sons	134738.30890051392	5100.061769330424	129638.2471311835
4	Jones and Sons	133838.90505085973	6515.100206365886	127323.80484449385
5	Williams LLC	147644.2294684391	29925.6613035495	117718.5681648896

Conclusion: The analysis of billing amounts across different hospitals reveals significant increases in billing from the previous year. The top 5 hospitals with the most notable billing amount increments are "Smith Inc," "Smith Ltd," "Thomas and Sons," "Jones and Sons," and "Williams LLC." These findings suggest varying financial dynamics among hospitals, emphasizing the importance of understanding and managing billing trends for effective financial planning and resource allocation in healthcare institutions.

Q10. List the medical conditions where the average age of patients is below the overall average age.

Input SQL Query:

```
WITH AverageAgeByCondition AS (
    SELECT
        MedicalCondition,
        ROUND(AVG(Age)) AS AverageAge
    FROM
        Health
    GROUP BY
        MedicalCondition
)

SELECT
    MedicalCondition,
    AverageAge
FROM
    AverageAgeByCondition
WHERE
    AverageAge < (SELECT AVG(Age) FROM Health);
```

Output SQL Query:

	medicalcondition character varying (255)	averageage numeric
1	Hypertension	51
2	Asthma	51

Conclusion: The medical conditions "Hypertension" and "Asthma" have average patient ages of 51, which are below the overall average age. This suggests that patients with these conditions tend to be younger on average compared to the entire patient population.

Q11. *Identify the patients with the highest and lowest billing amounts within their respective medical conditions.*

Input SQL Query:

```sql
WITH RankedBilling AS (
  SELECT
    Name,
    Medicalcondition,
    Billingamount,
    RANK() OVER (PARTITION BY Medicalcondition ORDER BY Billingamount DESC) AS Billing_Rank_Highest,
    RANK() OVER (PARTITION BY Medicalcondition ORDER BY Billingamount ASC) AS Billing_Rank_Lowest
  FROM
    Health
)
SELECT
  Name,
  Medicalcondition,
  Billingamount,
  'Highest' AS Billing_Rank_Type
FROM
  RankedBilling
WHERE
  Billing_Rank_Highest = 1

UNION

SELECT
  Name,
  Medicalcondition,
  Billingamount,
  'Lowest' AS Billing_Rank_Type
FROM
  RankedBilling
WHERE
  Billing_Rank_Lowest = 1;
```

Output SQL Query:

	name character varying (255)	medicalcondition character varying (255)	billingamount double precision	billing_rank_type text
1	Anna Adams	Obesity	1000.1808370217517	Lowest
2	Richard Jones	Asthma	49974.299137554735	Highest
3	Holly Clayton	Obesity	49974.16045847918	Highest
4	Daniel Hall	Hypertension	49995.902283221156	Highest
5	Ashley Novak	Arthritis	1009.4173268954232	Lowest
6	Teresa Buchanan	Cancer	49994.98474012479	Highest
7	Peggy Brooks	Hypertension	1084.422303472434	Lowest
8	Tonya Smith	Cancer	1020.337790368703	Lowest
9	Roy Beck	Arthritis	49985.97306779267	Highest
10	Joshua Crawford	Diabetes	1071.456127337255	Lowest
11	John Kim	Asthma	1032.2630866161537	Lowest
12	John Oneill	Diabetes	49954.968325041846	Highest

Conclusion: The provided data showcases patients with the highest and lowest billing amounts within their respective medical conditions. Notably, individuals such as "Richard Jones," "Holly Clayton," and "Daniel Hall" have the highest billing amounts in conditions like "Asthma," "Obesity," and "Hypertension," respectively. Conversely, patients like "Anna Adams," "Ashley Novak," and "Joshua Crawford" have the lowest billing amounts in their specified medical conditions. This information underscores the variability in healthcare costs among different medical conditions, emphasizing the need for tailored financial planning and patient support within each condition category.

Q12. Find pairs of patients who share the same room number and calculate the average age of each pair.

Input SQL Query:

```sql
SELECT
    P1.Name AS Patient1,
    P2.Name AS Patient2,
    P1.Roomnumber,
    ROUND(AVG((P1.Age + P2.Age) / 2))AS AverageAgeOfPair
FROM
   Health P1
JOIN
    Health P2 ON P1.Roomnumber = P2.Roomnumber AND P1.Name < P2.Name
GROUP BY
    P1.Name, P2.Name, P1.Roomnumber
ORDER BY
    P1.Roomnumber, P1.Name, P2.Name
LIMIT 10;
```

Output SQL Query:

	patient1 character varying (255)	patient2 character varying (255)	roomnumber character varying (10)	averageageofpair numeric
1	Alicia French	Angela Knox MD	101	59
2	Alicia French	Angie Peterson	101	63
3	Alicia French	Brianna Rose	101	40
4	Alicia French	Brittney Wells	101	58
5	Alicia French	Christopher Matthews	101	40
6	Alicia French	Collin Mcdonald	101	53
7	Alicia French	Cory Williams	101	41
8	Alicia French	Dorothy Ingram	101	63
9	Alicia French	Gary Hamilton	101	57
10	Alicia French	George Harris	101	52

Conclusion: The analysis of pairs of patients who share the same room number reveals various average ages. his diversity suggests that roommates may have different age profiles, emphasizing the importance of considering individualized care and support within shared healthcare accommodations. Understanding such dynamics can aid healthcare providers in tailoring services to accommodate varying needs and preferences among patients who share the same living space.