Healthcare data analysis is a dynamic field that harnesses the power of data to extract meaningful insights, drive informed decision-making, and enhance the overall quality of healthcare delivery. In today's rapidly evolving healthcare landscape, the vast amount of data generated presents both challenges and opportunities. Analyzing this data allows healthcare professionals to uncover patterns, trends, and correlations that can inform critical decisions, improve patient outcomes, and optimize operational efficiency.

Structured Query Language (SQL) stands out as a fundamental tool in healthcare data analysis. With its ability to query and manipulate databases, SQL enables professionals to efficiently retrieve and analyze relevant information from healthcare datasets. From patient records and treatment histories to administrative and operational data, the application of SQL in healthcare data analysis facilitates a comprehensive understanding of the complex interactions within the healthcare system.

**Data Source: Synthea - Simulating Real world Healthcare Patterns**

Synthea, an open-source educational tool, provides a simulated patient database, emulating an electronic medical records system. The data generated by Synthea is designed to model real-world patterns observed in individuals dealing with various diseases and their corresponding hospital visits

**Dataset Overview: Females Health Data**

This specific dataset focuses on female health, comprising information on encounters, immunization, and conditions. The dataset includes:

1. Patients Details – Patient Id, Name, Marital status, Gender, Address, City
2. Immunization – Description of vaccines, code
3. Encounter – The dataset appears to represent healthcare encounters, with each row representing a specific encounter, costs, providers, payers, and start stop date
4. Condition – Diagnosis of patient’s condition, start and stop date of condition, patient id

**Exploration & Tools:**

I am using **SQL server Management tool** to uncover the objectives of this analysis. Here are the **SQL commands and statements** used in this project,

* DISTINCT
* HAVING
* Merging columns in tables with JOINS
* Combining rows with UNION
* SUBQUERIES
* CASE WHEN
* GROUP BY
* ORDER BY

**Problem statement** –

1. Calculate distinct marital statuses and their respective counts of the patients

select distinct MARITAL,count(\*) as count

from patients

group by MARITAL

order by count(\*) desc

result-

MARITAL count

M 4640

\N 3682

S 1549

D 1162

W 330

Key finding- The majority of female patients attending the hospital are married

1. Calculate the average age and count of patients based on their marital status

Select MARITAL, AVG(DATEDIFF(YEAR, BIRTHDATE, GETDATE()) -

CASE

WHEN (MONTH(BIRTHDATE) > MONTH(GETDATE()) OR (MONTH(BIRTHDATE) = MONTH(GETDATE()) AND DAY(BIRTHDATE) > DAY(GETDATE())))

THEN 1

ELSE 0

END) AS avg\_age,

COUNT(\*) AS count

from patients

GROUP BY MARITAL

ORDER BY count DESC;

result-

MARITAL avg\_age count

M 55 4640

\N 17 3682

S 57 1549

D 62 1162

W 65 330

Key finding- 4640 female patients attending the hospital are married

and they have an average age of 55

1. Types of Encounter class and number of patients in each class

select ENCOUNTERCLASS, count(ENCOUNTERCLASS) as Count

from encounters

group by ENCOUNTERCLASS

order by count(ENCOUNTERCLASS) desc

result-

ENCOUNTERCLASS Count

ambulatory 244148

outpatient 85849

wellness 77849

urgentcare 21479

emergency 17903

inpatient 3709

home 2318

snf 1221

hospice 750

virtual 709

Key finding- Most encounterclass in the hospital is ambulatory. An ambulatory hospital service refers to healthcare facilities and medical care that are designed to provide services on an outpatient basis, allowing patients to receive treatment or procedures without the need for an overnight stay. Ambulatory services typically include consultations, diagnostic tests, minor surgeries, and other medical interventions that can be performed within a single day.

4. Retrieve the description, encounter class, and count of occurrences for encounters where the encounter class is 'ambulatory

select DESCRIPTION, count(\*) as count

from encounters

Where ENCOUNTERCLASS= 'ambulatory'

group by DESCRIPTION

order by count(\*) desc

result-

DESCRIPTION ENCOUNTERCLASS count

Hospital Encounter with Problem ambulatory 124324

Prenatal Care ambulatory 39073

Hospital Encounter with Symptoms ambulatory 24197

Outpatient Clinic Procedure ambulatory 12312

Follow-up appointment ambulatory 9739

Initial Prenatal Care ambulatory 7020

Check-up ambulatory 6573

Follow-up appointment after childbirth ambulatory 4512

Postoperative Check-up ambulatory 3797

Telemedicine ambulatory 2599

Monitoring asthma progress ambulatory 2187

Monitoring Oversight ambulatory 2059

Procedure in Patient Encounter ambulatory 1846

Self-requested appointment ambulatory 1288

GYN service ambulatory 1122

Nursing Home ambulatory 492

Initial evaluation of allergic condition ambulatory 267

Office Care ambulatory 210

Follow-up Assessment of Allergic Conditions ambulatory 210

Hospitalization ambulatory 134

Discussion of treatment options ambulatory 111

Periodic Assessment and Oversight of ambulatory 66

Wellbeing for Individuals

Medical appointment with assessment ambulatory 4

and/or treatment

Follow-up appointment after surgery ambulatory 3

Outpatient Clinic Admission ambulatory 2

Psychiatric Assessment with Mental Health ambulatory 1

Assessment

Key insights- query shows the count of occurrences for different descriptions within the 'ambulatory' encounter class. The most common ambulatory encounter is "Hospital Encounter with Problem," indicating a significant number of patients seeking care for specific health issues. Strengthen initiatives related to preventive care, given the frequency of encounters like Prenatal Care and Check-ups. Promote regular health check-ups and screenings. The presence of Telemedicine indicates a utilization of technology for remote healthcare services, should explore and expand these services.

1. what the different types of insurance company which patients use and their count

select distinct Payer, count(\*) as Patient\_count

from encounters

group by Payer

order by count(\*) desc

result-

Payer Patient\_count

Medicare 125694

Humana 68657

Blue Cross Blue Shield 60911

Cigna 38736

Dual Eligible 31143

UnitedHealthcare 28624

No Insurance 26286

Anthem 26229

Medicaid 25072

Aetna 24583

Key insight- Medicare, Humana, and Blue Cross Blue Shield are the top three dominant payers with significantly higher patient counts. Dual Eligible patients represent a notable portion. Consider specific programs or services to address the unique healthcare needs of this population.

The count of patients with "No Insurance" suggests a population without coverage. Consider initiatives to improve access to healthcare services for individuals without insurance.

1. information about encounters, their average age, and count, categorized by encounter class

select

Distinct e.ENCOUNTERCLASS,

AVG(DATEDIFF(YEAR, p.BIRTHDATE, GETDATE()) -

CASE

WHEN (MONTH(p.BIRTHDATE) > MONTH(GETDATE()) OR (MONTH(p.BIRTHDATE) = MONTH(GETDATE()) AND DAY(p.BIRTHDATE) > DAY(GETDATE())))

THEN 1

ELSE 0

END) AS avg\_age,

COUNT(\*) AS count

from patients as p

JOIN encounters as e

ON p.Id = e.Patient

GROUP BY e.ENCOUNTERCLASS

ORDER BY count DESC

result –

ENCOUNTERCLASS avg\_age count

ambulatory 52 244148

outpatient 47 85849

wellness 40 77849

urgentcare 67 21479

emergency 50 17903

inpatient 57 3709

home 75 2318

snf 53 1221

hospice 72 750

virtual 50 709

key insights – The majority of encounters fall into the 'ambulatory,' 'outpatient,' and 'wellness' categories, indicating a significant utilization of these types of encounters. Average ages vary across different encounter classes, reflecting the diversity of patients seeking various types of healthcare services. Given the high count of 'ambulatory' encounters, consider tailoring and optimizing services in ambulatory settings to meet the diverse healthcare needs of patients.

1. list of medical conditions along with their respective counts

select Description,

count(\*) as count\_of\_condition

from conditions

group by Description

order by count\_of\_condition desc

result-

Description count\_of\_condition

Other psychological or physical stress, 58962

not elsewhere classified

Pregnant state, incidental 9872

Acute bronchitis 5684

Body Mass Index 30.0-30.9, adult 5461

Unemployment 4717

Anemia, unspecified 3704

Other specified anemias 3704….

Total -174 row

Key insight- Conditions related to stress and pregnancy (incidental or specified) have notable occurrences. Conditions like obesity (BMI 30.0-30.9) and unemployment indicate the inclusion of public health and lifestyle factors.

Implement preventive health programs targeting stress management, especially considering the high occurrence of stress-related conditions. Enhance prenatal care services and education to address both incidental and specified pregnancy-related conditions