**Healthcare Project**



**By – Pavel Singh**

**Description:**

The healthcare data consists of six tables in star schema format.

**Table 1 : Encounters**

This table consists of the information about patients encounters in the hospital i.e. how the patients came to hospital (ambulatory, urgent care etc). Also it consists of cost for each encounter and total claim for encounter and insurance.

**Table 2: Conditions**

This table consists of the medical condition for which patient came to hospital.

**Table 3: Immunization**

It consists of the immunization procedure for every encounter along with the cost.

**Table 4: Patients**

This table consist of all the patient’s data such as name, age, gender, address, marital status, id, SSN, cost etc.

**Table 5: Supply**

This table consists of medication provided to each encountered patient.

**Table 6: Payers**

This table consists of the insurance company info that covered for the patients.

**Table 7: Procedures**

This table consists of the procedure a patient has undergone.

**Tools to be used:**

* EXCEL
* SQL
* Python

**Business Problems:**

1. Retrieve a count of patients by age group and gender.
2. Calculate the average age of patients in the dataset.
3. Identify the most common types of encounters (e.g., outpatient, inpatient, ER).
4. Calculate the average length of stay for inpatient encounters.
5. Analyse the distribution of encounters by month.
6. **Patients History:**

* Create a timeline of encounters and conditions for individual patients.
* Identify patients with a history of hospital readmissions.
* Analyse the association between certain conditions and subsequent encounters.

1. **Cost Analysis:**

* Calculate the average cost per encounter, including breakdowns for different cost components (e.g., medication, procedures).
* Analyse the average reimbursement rates from different payers for specific procedures or services.

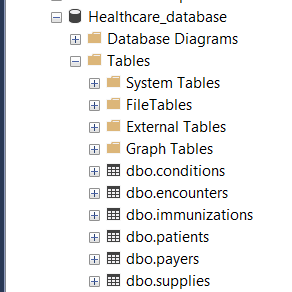
**Machine Learning Problem:**

Building a Classification model to Classify Encounter Class of a patient.

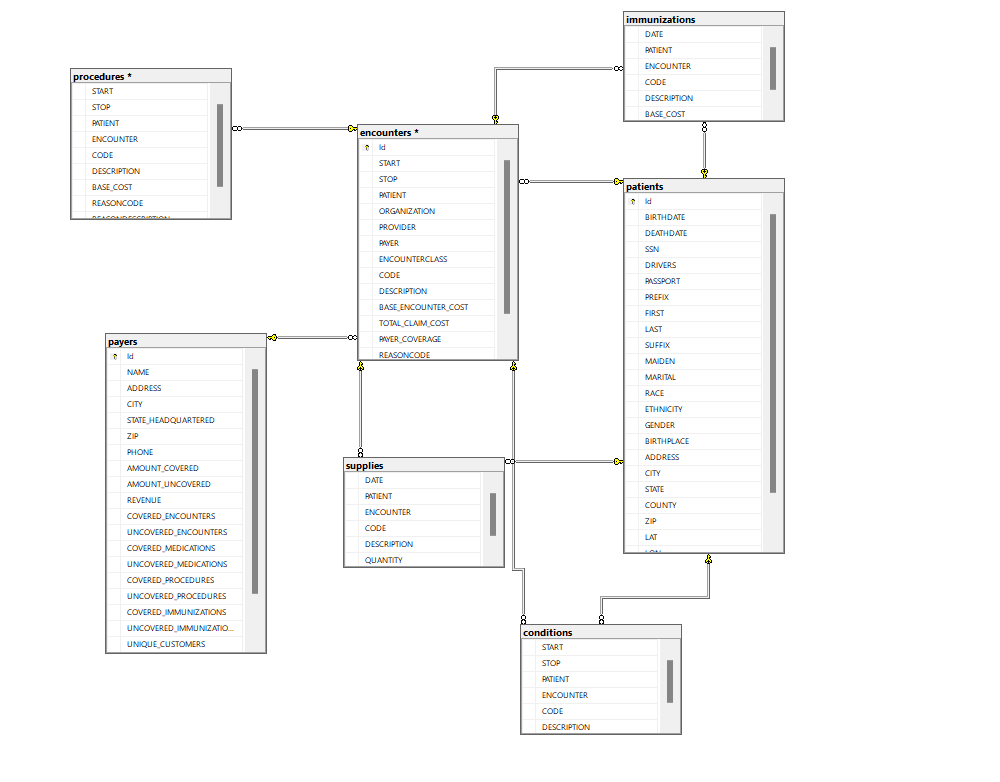
**SETTING HEALTHCARE DATABASE**

Importing all the 7 datasets into SQL server.

Setting up all **Primary Keys** and **Foreign Keys** to make a Relational database.



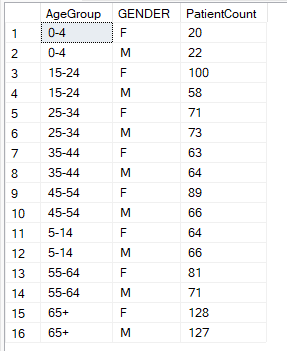
**ERD for Healthcare\_database:**



**Insights:**

1. The Healthcare Data have a total 1163 unique patients.

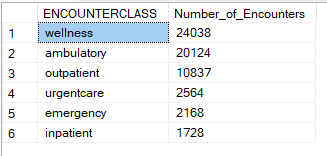
Out of which we have most patients which are Senior Citizens.



1. By analysing the data we can also see that the average age of the patients is near 43 years.



1. In the database there are mainly 5 encounters for which patients comes to hospital.



The most common occurring encounter is **Wellness** that includes patients coming for General Checkup procedure (stress, Hypertension, Anxiety, Anaemia etc.)

**Ambulatory:** This encounter refers to when a patient is brought up to hospital due to some urgent or emergency reason such as pregnancy, Covid-19 , Viral , Heart Attack etc.

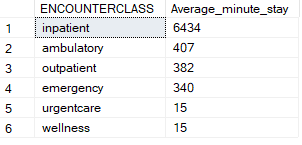
**Outpatient:** These are encounters in which a patient receives medical care without being admitted to a hospital. Outpatient encounters can include routine check-ups, specialist consultations, diagnostic tests, vaccinations etc.

**Urgent Care:** Urgent care encounters are for patients with medical issues that require prompt attention but are not life-threatening.

**Emergency**: Emergency encounters occur when a patient seeks immediate medical attention in the emergency department due to severe injuries, life-threatening conditions, or sudden and severe illnesses.

**Inpatient:** Inpatient encounters involve a patient being admitted to a hospital or healthcare facility for an extended period of care. This typically includes patients with serious or acute medical conditions that require continuous monitoring, treatment, and hospitalization.

1. So, the average time of stay in Hospital for **impatient** Encounter Class is approx. 6434 minutes.

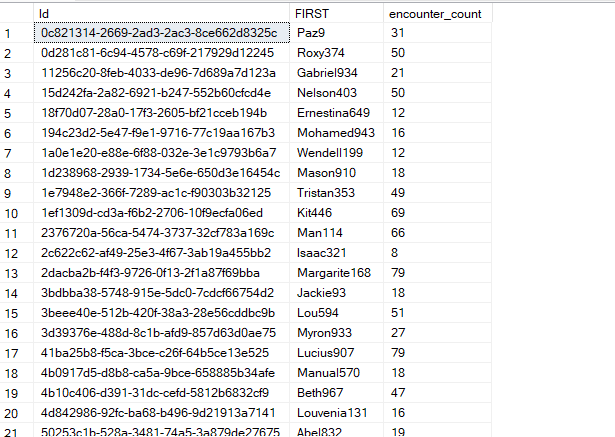


1. Identify patients with a history of hospital readmissions.

Here we can identify how many times a patient got admitted to the hospital.

For example:

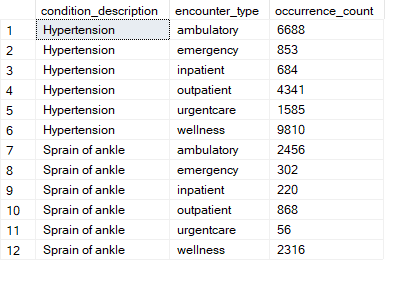
A patient named Paz9 encounter 31 times to the hospital.



Next now we can see some association between some conditions and encounters.

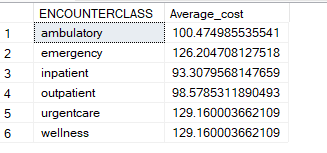
For example, for Conditions such as *Hypertension* and *Sprain in Ankle* we observe that most patients with Hypertension falls under wellness category because most of the patients have a high blood pressure and comes for regular checkup.

Also, we observe that Sprain of Ankle in a common Ambulatory encounter as it can happen accidentally to someone.



**Cost Analysis:**

1. Average Cost per encounter:



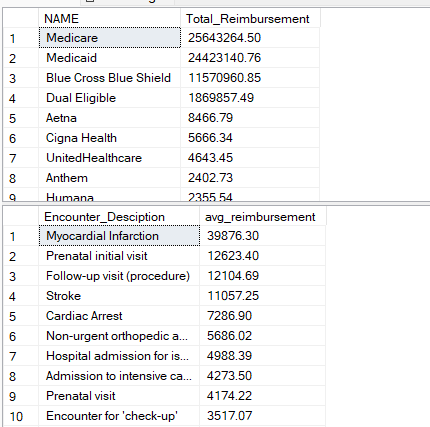
Lets have a look at top patients with highest Healthcare

expenses:

We can see Sung603 total heath expenses is around $13,197,859



1. Let’s have a look at top insurance company paying for patients and reimbursement for different procedures.



*Medicare healthcare Insurance* company have paid the highest Reimbursement Amount till now around $25 million

Also we see that *Myocardial Infarction* gets the highest Reimbursement followed by *Prenatal initial visit.*

**Machine Learning:**

The Machine Learning Model will Classify the Encounter Class (Ambulatory, wellness, impatient, outpatient, emergency) of Patients based on the other columns such as

*encounter\_code, BASE\_ENCOUNTER\_COST, condition\_code, IMMUNE\_CODE, Procedure\_Code*

We will be using **K Nearest Neighbours** for our Classification model.

**Step 1)**

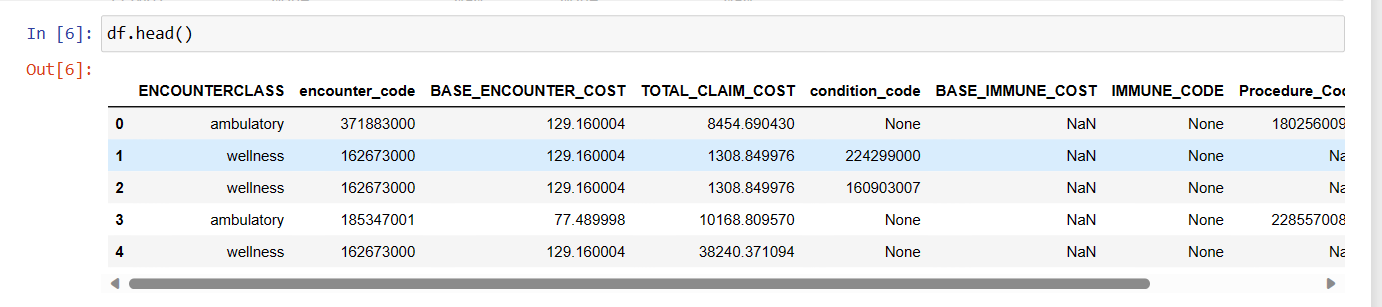
The first step is to prepare the Data for our model.

The Data has been pulled from Healthcare\_Database that we created for this Project.



**Step 2)**

Using the SQL queries in Our python we will prepare our Data by merging the tables from where we need necessary columns.



**Step 3)** Handling Missing Values:



**Step 4)** The Data is spilt into Train test split and using appropriate value for K we achieve accuracy score of 0.98 for our Model.

