**SI 564 SQL & Databases Final Project**

**Team Members: Pratap Gude & Andy Higuera**

**Part 1:**

Your database documentation, including ERD diagrams.

**Part 2:**

A letter from someone at your fictional company outlining important questions they need answered from the dataset. This should have at 4-10 questions in it.

**Part 3:**

The answers to the questions posed in the letter (**including** your queries to get the result).

**Part 4:**

An outline of why you made the database design choices you made.

**Part 1:**

We chose a case study that includes the hospital admission dataset, which contains N=58,863 (cases) and k=9 (features). We removed any patient identification features, such as ID (Coded Case Identifier). Inserted autoincremented integers in their place. As the dataset has repeated categorical data, we created six sub-tables linking to the main table with foreign key - primary key inferences and normalization. Through the queries we ran in Part 3, we would include recommendations that the deidentified hospital could infer and work on to increase its business efficiency. We were mindful of the naming terminology to reduce redundancies.

**Reference:** <https://umich.instructure.com/courses/38100/files/folder/Case_Studies/18_HospitalAdmissions?preview=5005926>

**Queries used in updating the database:**

1. Update HospitalAdmissionsData H set Insurance\_Type = (select Insurance\_Type\_ID from Insurance\_Type\_Info where Insurance\_Type\_Comment = H.Insurance\_Type)
2. Update HospitalAdmissionsData H set Admission\_Type =

(select Admission\_Type\_ID from Admission\_Type where Admission\_Type\_Comment = H.Admission\_Type)

1. Update HospitalAdmissionsData H set Race =

(select Race\_ID from Race where Race\_Type = H.Race)

1. Update HospitalAdmissionsData H set Religion\_Type =

(select Religion\_ID from Religion where Religion\_Name = H.Religion\_Type)

***Tables in Database:***

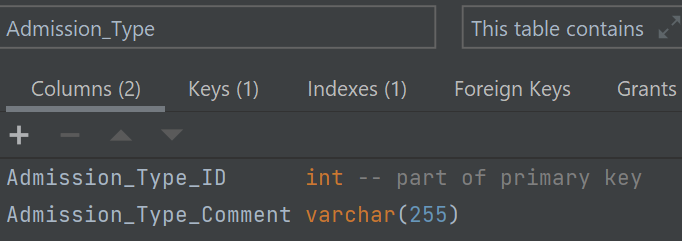
1. Admission\_Type
2. Death\_Status
3. HospitalAdmissions\_Data (Main table)
4. Insurance\_Type\_Info
5. Marital\_Status
6. Race
7. Religion

***Database Documentation:***

1. **Admission\_Type**

Admission\_Type\_ID: unique identifier for the type of admission type comment that the patient was enrolled in the hospital. (int) (PK)

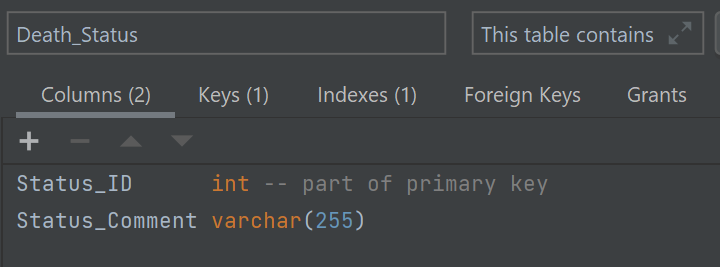
Admission\_Type\_Comment: description of the patient's admission status when was admitted to the hospital. (string)



1. **Death\_Status**

Status\_ID: Unique identifier for the description of the status of the patient. (int) (PK)

Status\_Comment: description of the status of the patient. (string)



1. **HospitalAdmissions\_Data (Main table)**

Unique\_ID: Autoincrement integers (deidentified) (int) (PK)

AdmissionLengthDays: Duration of hospital stay (in days) (float)

DeathStatus: Indicator of Death (1) survival (0) (int)

Admission\_Type: Type of admission (categorical) (string)

Insurance\_Type: Type of Health insurance (string)

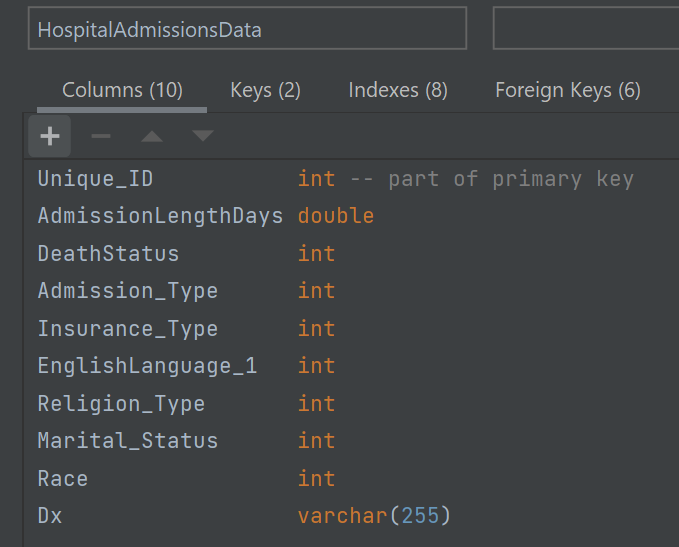
EnglishLanguage\_1: Primary Language English Indicator (1), 0 otherwise (int)

Religion\_Type: Type of Religious Affiliation (string)

Marrital\_Status: Indicator of marital status (Married=1) (int)

Race: Race, categorical (string)

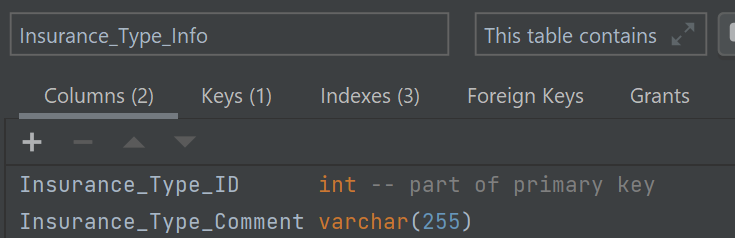
Dx: Diagnosis (string)



1. **Insurance\_Type\_Info**

Insurance\_Type\_ID: Unique identifier for the description of the status of the patient. (int) (PK)

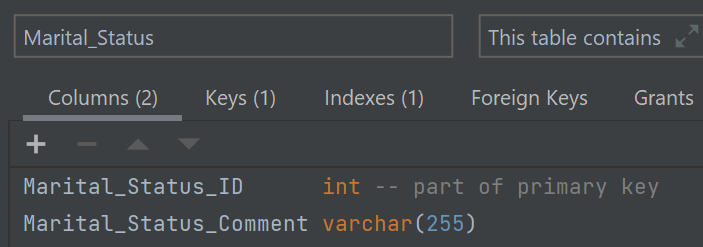
Insurance\_Type\_Comment: description of the type of insurance the patient carries. (string)



1. **Marital\_Status**

Marital\_Status: unique identifier for the marital status comment of the patient. (int) (PK)

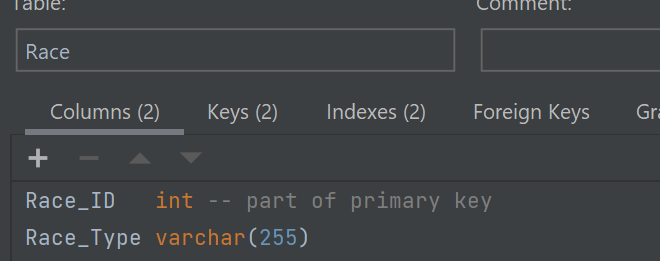
Marital\_Status\_Comment: description of patient marital status. (string)



1. **Race**

Race\_ID: unique identifier for the description of the patients' races. (PK)

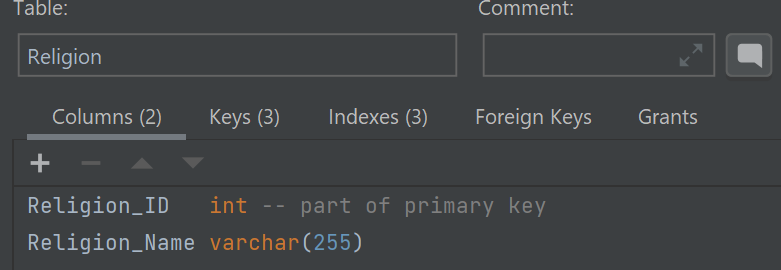
Race\_Type: description of the race of the patient.



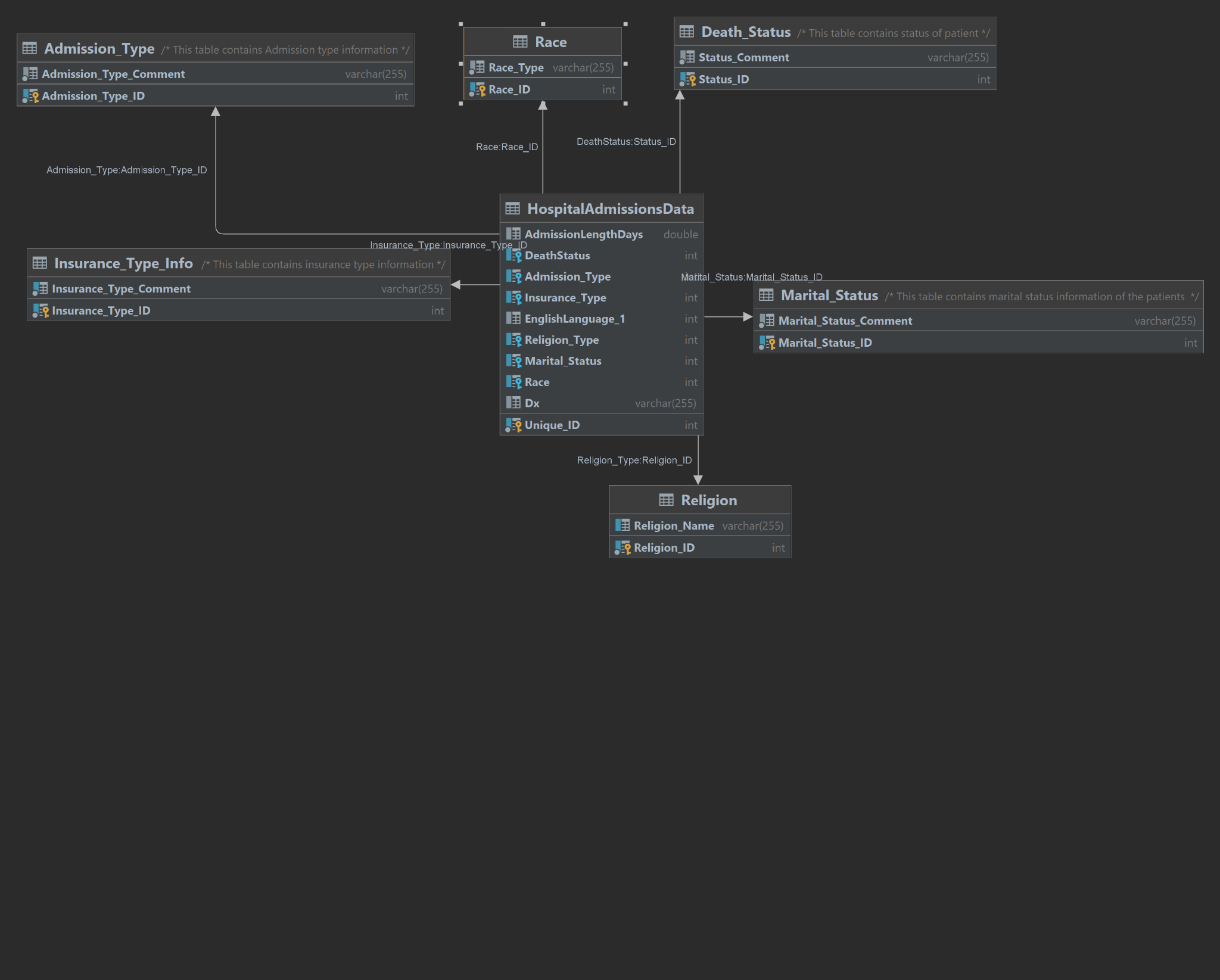
1. **Religion**

Religion\_ID: unique identifier for religion name. (int) (PK)

Religion\_Name: description of religion. (string)



**ERD Diagram:**

****

**Part 2:**

Hi Database Team,

This is Kelly Davenport, We have a client who is in Health Care, and they consulted us if we could help provide them with some recommendations based on some data they provided. Please use the data and create, update, read the Database and provide me with some meaningful recommendations. I will be working with other backend teams in the meanwhile to create a user-centric web platform for you. Please provide the code, references, ERD, and any additional documentation that ever seems fit for this case.

1. Based on the patient's admission, what is the insurance type they carry when grouped by patient admission type?
2. What is the most common diagnosis among patients when the insurance type is Medicare?
3. On average, what is the diagnosis that causes the most death among patients?
4. On average, what is the diagnosis that causes the most number of death among patients if they are admitted under emergency?
5. On average, what is the diagnosis with the least length of stay among patients admitted to emergency?
6. Among the admission type ‘newborn,’ what are the fatality rates?
7. When patient mortality occurs, what is the average length of admission?
8. What is the average length of stay for a patient who has a Dx of coronary artery disease?

Once these tasks are complete, please message me back so we can arrange a company-to-company mandatory fun day. We plan on throwing a large party in honor of our future collaborations together for the betterment of data science.

Best regards,

Kelly Davenport

CEO Borremean Digital

**Part 3**

1. Based on the patient's admission, what is the insurance type they carry when grouped by patient admission type?

**Ans:** Most of the patients in this dataset carry Medicare insurance type

**Query:**

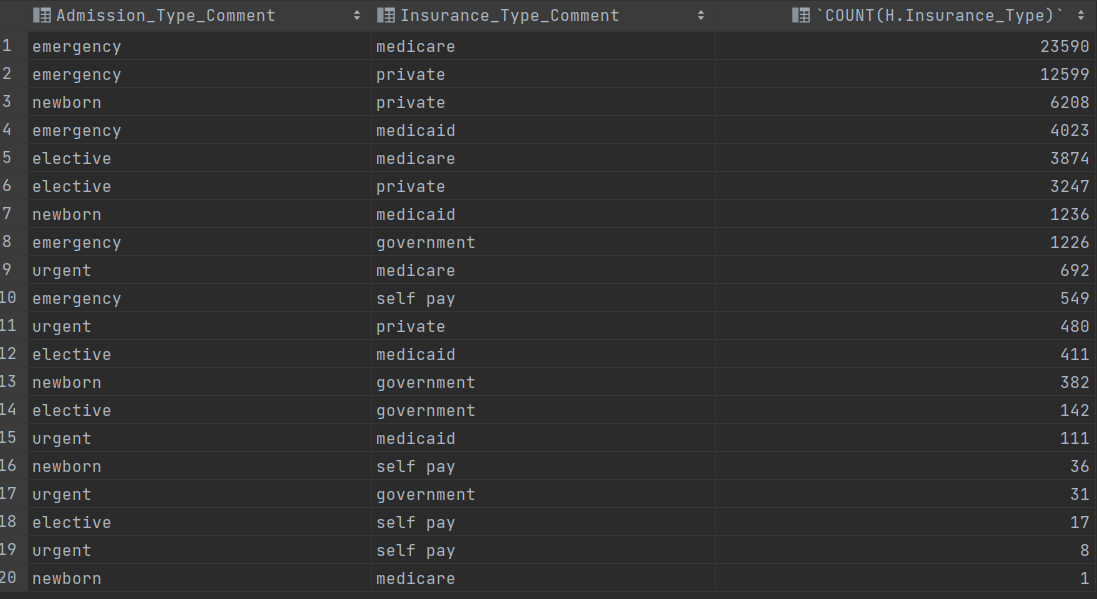
SELECT A.Admission\_Type\_Comment, ITI.Insurance\_Type\_Comment,COUNT(H.Insurance\_Type) FROM

((HospitalAdmissionsData H LEFT JOIN Insurance\_Type\_Info ITI on H.Insurance\_Type = ITI.Insurance\_Type\_ID) LEFT JOIN Admission\_Type A on A.Admission\_Type\_ID = H.Admission\_Type)

GROUP BY H.Admission\_Type, H.Insurance\_Type

ORDER BY COUNT(H.Insurance\_Type) DESC;

**Screenshot:**



**Recommendation:**

Given the aforementioned results. The hospital has more patient groups who pay through medicare. Please include services that seem to be covered under medicare to increase the revenue.

1. What is the most common diagnosis among patients when the insurance type is Medicare?

**Ans:** Pneumonia and Sepsis are the most frequent diagnostic groups which are seen through this data which are paid through medicare

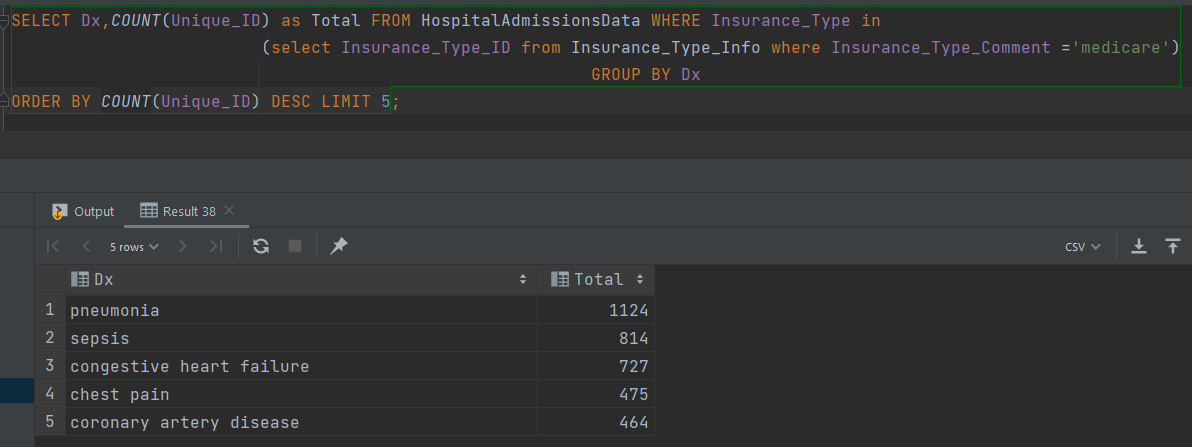
**Query:** SELECT Dx,COUNT(Unique\_ID) as Total FROM HospitalAdmissionsData WHERE Insurance\_Type in

(select Insurance\_Type\_ID from Insurance\_Type\_Info where Insurance\_Type\_Comment ='medicare')

GROUP BY Dx

ORDER BY COUNT(Unique\_ID) DESC LIMIT 5;

**Screenshot:**



**Recommendation:** We recommend placing a higher emphasis on care for these diagnoses and the departments that treat the diagnosis.

1. On average, what is the diagnosis that causes the most death among patients?

**Ans:** Newborn, Pneumonia & Sepsis are the most common cause of death among patients.

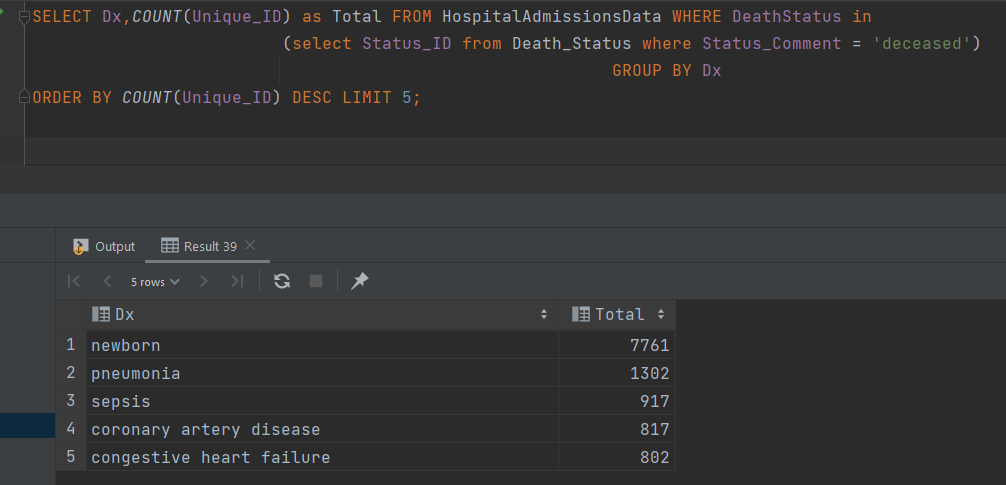
**Query:** SELECT Dx, COUNT(Unique\_ID) as Total FROM HospitalAdmissionsData WHERE DeathStatus in

(select Status\_ID from Death\_Status where Status\_Comment = 'deceased')

GROUP BY Dx

ORDER BY COUNT(Unique\_ID) DESC LIMIT 5;

**Screenshot:**



**Recommendation:**

Please look into workflows and staffing responsibilities in

* Neonatal Care & Maternity department
* Rheumatology department
* Infectious Diseases Department

1. On average, what is the diagnosis that causes the most number of death among patients if they are admitted under emergency?

**Ans:** Pneumonia and Sepsis are the most common cause of death when the patient is admitted to an emergency.

**Query:**

SELECT Dx,COUNT(Unique\_ID) as Total FROM HospitalAdmissionsData WHERE ((DeathStatus in

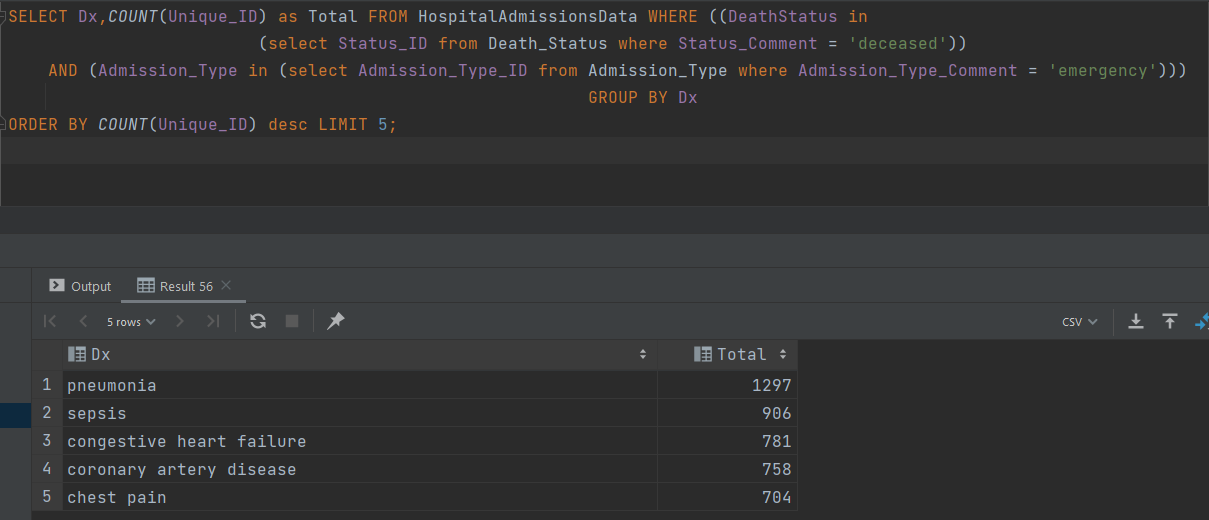
(select Status\_ID from Death\_Status where Status\_Comment = 'deceased'))

AND (Admission\_Type in (select Admission\_Type\_ID from Admission\_Type where Admission\_Type\_Comment = 'emergency')))

GROUP BY Dx

ORDER BY COUNT(Unique\_ID) desc LIMIT 5;

**Screenshot:**



**Recommendation:**

Emphasize on the service enhancements and process optimizations to provide better care for these diagnoses.

1. On average, what is the diagnosis with the least length of stay among patients admitted to emergency?

**Ans:** Intracranial hemorrhage and endocarditis diagnosis have the least length of stay.

**Query:**

SELECT Dx, AVG(AdmissionLengthDays) AS lOS FROM HospitalAdmissionsData WHERE (

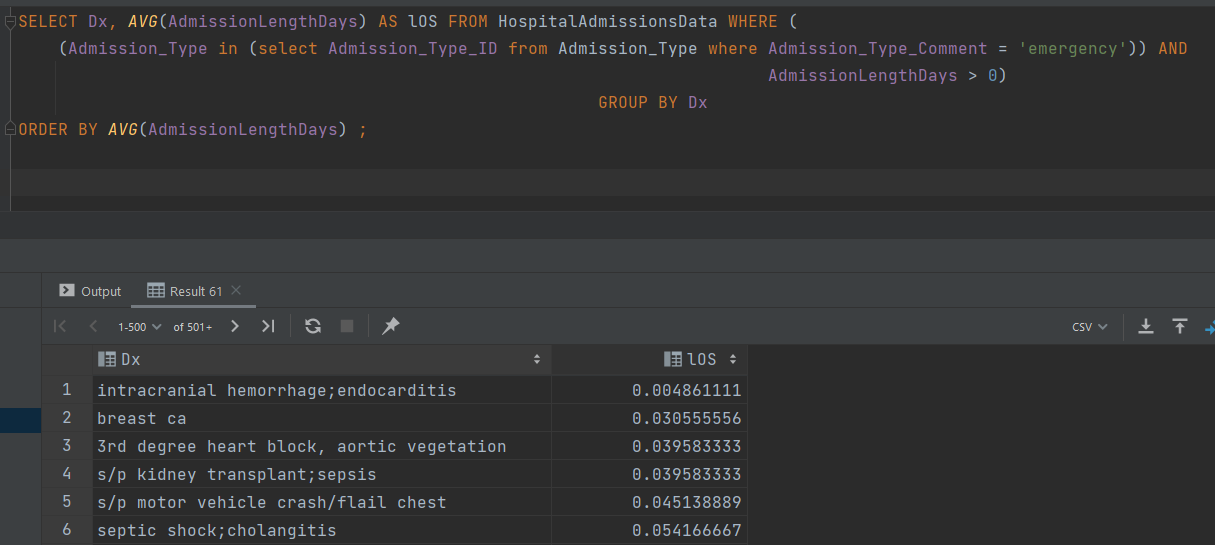
(Admission\_Type in (select Admission\_Type\_ID from Admission\_Type where Admission\_Type\_Comment = 'emergency')) AND

AdmissionLengthDays > 0)

GROUP BY Dx

ORDER BY AVG(AdmissionLengthDays);

**Screenshot:**



**Recommendation:**

Please consider looking into the outcome and prognosis for these diagnoses and consider metrics to reduce hospital readmissions.

1. Among the admission type ‘newborn,’ what are the fatality rates?

**Ans:** 99% fatality rates

**Query:**

SELECT (SELECT COUNT(Unique\_ID)

FROM HospitalAdmissionsData WHERE

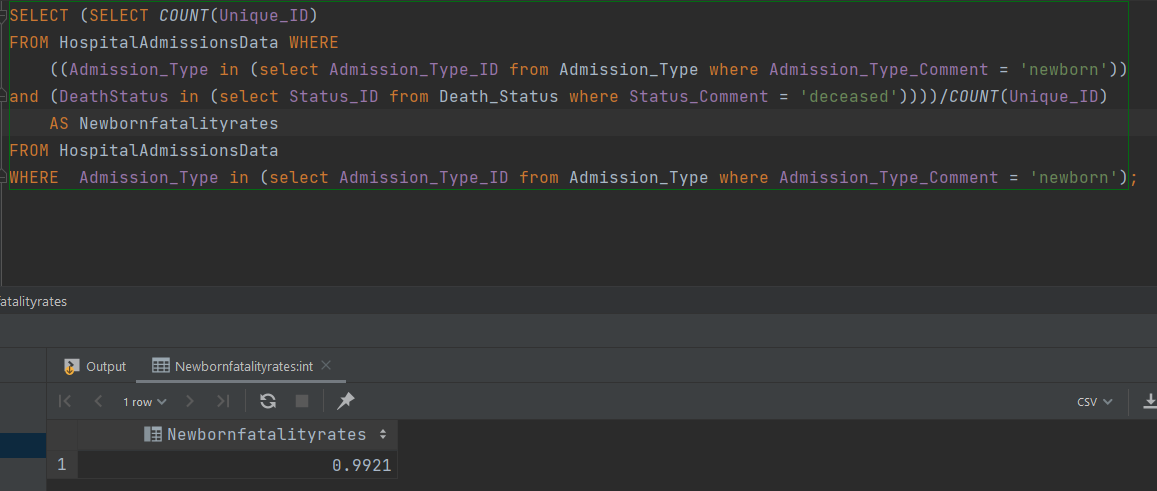
((Admission\_Type in (select Admission\_Type\_ID from Admission\_Type where Admission\_Type\_Comment = 'newborn'))

and (DeathStatus in (select Status\_ID from Death\_Status where Status\_Comment = 'deceased'))))/COUNT(Unique\_ID) AS Newbornfatalityrates

FROM HospitalAdmissionsData

WHERE Admission\_Type in (select Admission\_Type\_ID from Admission\_Type where Admission\_Type\_Comment = 'newborn');

**Screenshot:**



**Recommendation:**

Please consider installing the neonatal and maternal care facility in the hospital as the results seem to be not in the hospital's favor. Or please consider a referral system for other patient clinics that offers services.

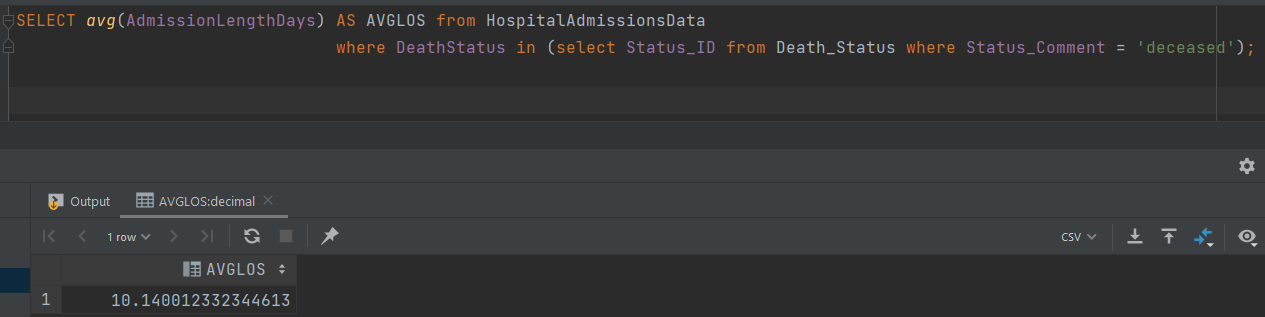
1. When patient mortality occurs, what is the average length of admission?

**Ans:** On average, 10 days is the length of stay when mortality occurs.

**Query:** SELECT avg(AdmissionLengthDays) AS AVGLOS from HospitalAdmissionsData

where DeathStatus in (select Status\_ID from Death\_Status where Status\_Comment = 'deceased');

**Screenshot:**



**Recommendation:**

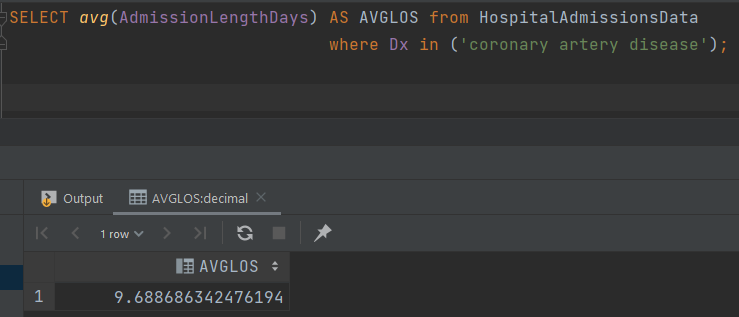
Consult your in-house clinic team to find various factors leading to mortality and prioritize them by weights to take necessary actions.

1. What is the average length of stay for a patient who has a Dx of coronary artery disease?

**Ans:** 9.6 days is the average length of stay with patients who are diagnosed with coronary artery disease.

**Query:** SELECT avg(AdmissionLengthDays) AS AVGLOS from HospitalAdmissionsData where Dx in ('coronary artery disease);

**Screenshot:**



**Recommendations:**

Provide palliative care and consider using digital check-ins on the patients who are the highest level of recovery to reduce their length of stays.

**Part 4:**

We made a choice on the database that we did because it had categorical data, and we thought it would be easier to normalize the data. At the start, we cleaned the data, removed some unnecessary columns that we thought would prove to be redundant, such as ID, and made our own by auto-incrementing the ID through the feature in Datagrip. There are some diagnoses that are in categories and have some special symbols in the data column. We included other choices we made in Part 1 of the assignment. Also, when it comes to recommendations, there is no external validation, and some inferences, such as newborn fatality rates, are surprisingly shocking. Seems like the data was not real but was created using simulation.