HOSPITAL ADMISSION ANSWERS TO BORROMEAN DIGITAL

Tuteja, Ashruti

SI 564 SQL & DATABASES Uniquename: tashruti

LETTER FROM

Kelly Davenport, CTO

This letter contains:

Segments	Topic
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 5	A .sql file containing your exported database.
Appendix	
Glossary	

PART 1

DATABASE DOCUMENTATION & ERD

About the Dataset

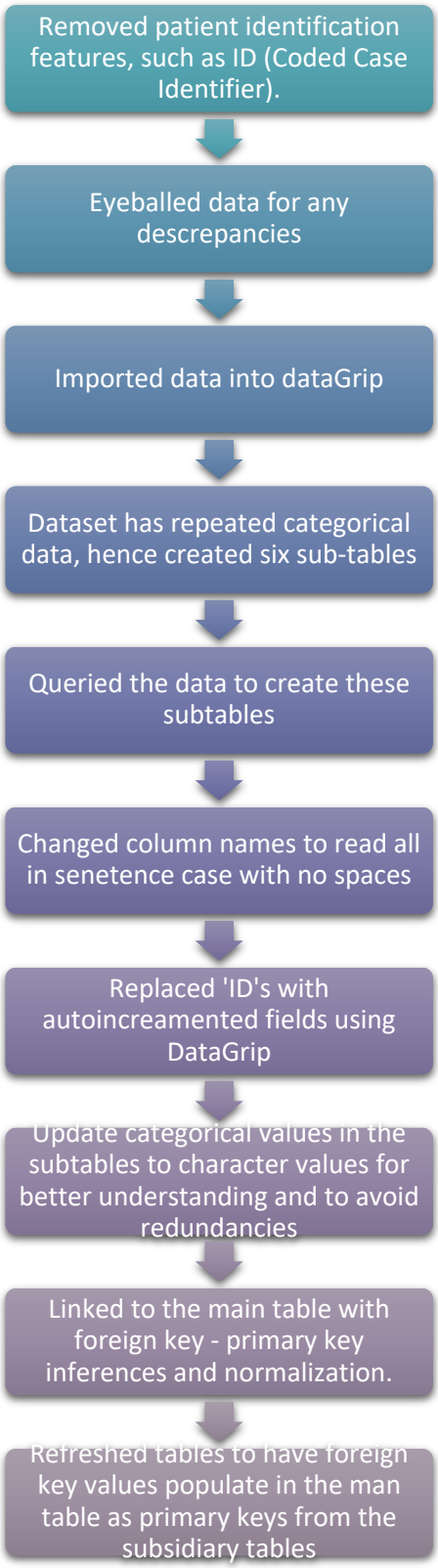
This is a case study from Open-source Canvas website of HS-853
https://umich.instructure.com/courses/38100/files/folder/Case_Studies/18_HospitalAdmissions?preview=5005926

It is a dataset with information on Hospital Admissions, which entails details like Patient ID, demographic information, average length of stay of the patient, diagnosis and outcome.
The dataset has interesting exploration insights like which is the most common cause of death, how long the patient stays in a hospital, or if diagnosis has an impact on patient death.
It contains of 58,863 rows (N) and 9 columns (k) referred to as features.

The dataset looks like:

N=58,863 (cases) and k=9 (features)	ID: Coded Case Identifier
	AdmissionLengthDays: Duration of hospital stay (in days)
	Death_1: Indicator of Death (1) survival (0)
	Admission_type: Type of admission (categorical)
	Insurance_Type: Type of Health insurance (string)
	EnglishLanguag_1: Primary Language English Indicator (1), 0 otherwise
	Religion_Type: Type of Religious Affiliation (string),
	Married_1: Indicator of marital status (Married=1)
	Race: Race, categorical (string)
	Dx: Diagnosis

Data Preparation (Appendix-1):



DATA TYPES (Appendix – 2)

hospital_admissions

```
mysql> desc hospital_admissions;
+-----------------------+--------+------+-----+---------+----------------+
| Field                 | Type   | Null | Key | Default | Extra          |
+-----------------------+--------+------+-----+---------+----------------+
| patient_id            | int    | NO   | PRI | NULL    | auto_increment |
| admission_length_days | double | YES  |     | NULL    |                |
| death                 | int    | YES  | MUL | NULL    |                |
| admission_type        | int    | YES  | MUL | NULL    |                |
| insurance_type        | int    | YES  | MUL | NULL    |                |
| english_language      | int    | YES  |     | NULL    |                |
| religion_type         | int    | YES  | MUL | NULL    |                |
| married               | int    | YES  | MUL | NULL    |                |
| race                  | int    | YES  | MUL | NULL    |                |
| diagnosis             | text   | YES  |     | NULL    |                |
+-----------------------+--------+------+-----+---------+----------------+
```

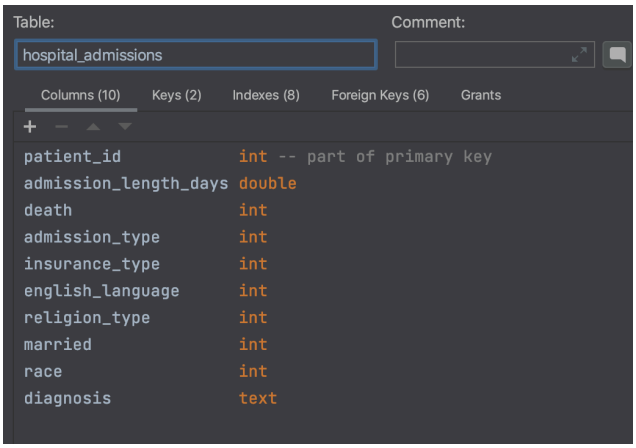
Other tables
Values: varchar
Patient_id: PRImary Key

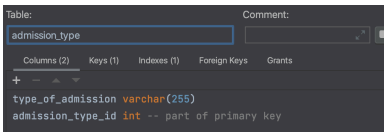
DATABASE DOCUMENTATION

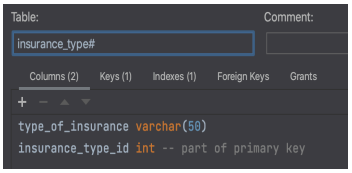
Tables in Database named FinalProject

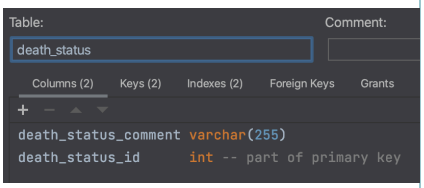
1. *hospital_admissions (main table)*
2. *admission_type*
3. *insurance_type*
4. *death_status*
5. *religion*
6. *race*
7. *marital_status*

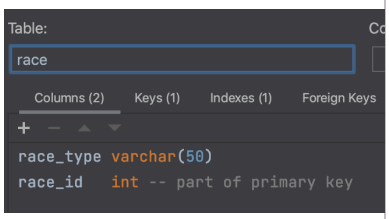
1. Table Name	Columns	Data Type	Description
hospital_admissions	patient_id (PK)	integer	unique patient identifier
	admission_length_days	float	Duration of stay in hospital
	death (FK)	boolean -> int	indicator of death (1) or survival (0)
	admission_type (FK)	str -> int	type of admission whether emergency, elective, urgent or newborn
	insurance_type (FK)	str -> int	type of health Insurance or Payer
	english_language	boolean -> int	English as primary language (1) or not (0)
	religion_type (FK)	str -> int	religion affiliation
	married (FK)	boolean -> int	indicator of marital status where married is 1, otherwise 0
	race (FK)	str -> int	race affiliation
	diagnosis	string	type of diagnosis

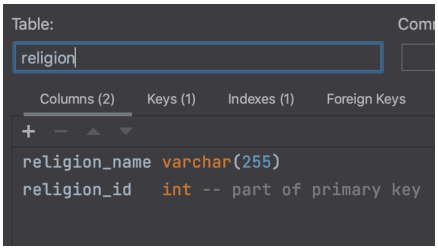


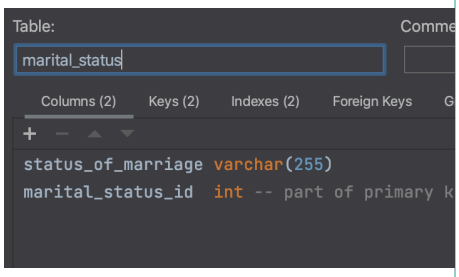
2. Table Name	Columns	Data Type	Description	
admission_type	admission_type_id (PK)	integer	unique identifier for patient admission type	
	type_of_admission	string	description of the patient's admission status when was admitted to the hospital	

3. Table Name	Columns	Data Type	Description	
insurance_type	insurance_type_id (PK)	integer	unique identifier for patient payer type	
	type_of_insurance	string	description of the type of insurance the patient carries	

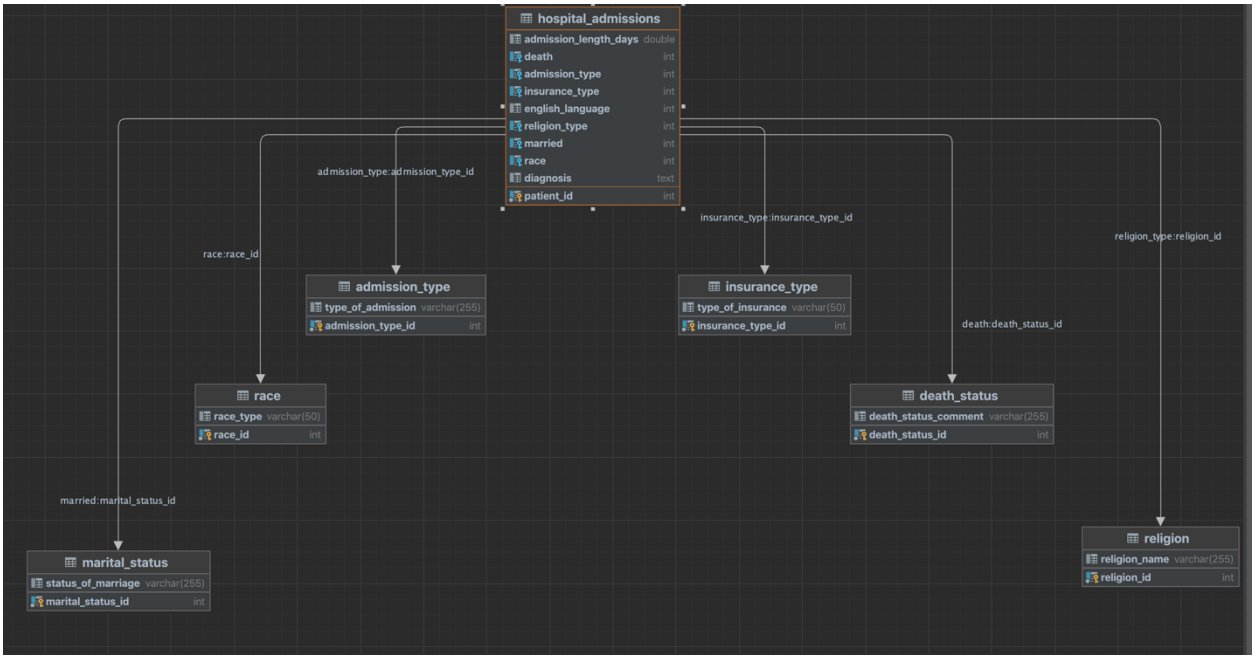
4. Table Name	Columns	Data Type	Description	
death_status	death_status_id (PK)	integer	unique identifier for the patient outcomes	
	death_status_comment	string	description on whether patient survived or not	

5. Table Name	Columns	Data Type	Description	
race	race_id (PK)	integer	unique identifier for the affiliation with race of the patient	
	race_type	string	description of the race of the patient	

6. Table Name	Columns	Data Type	Description	
religion	religion_id (PK)	integer	unique identifier for the affiliation with religion of the patient	
	religion_name	string	description of the religion of the patient	

7. Table Name	Columns	Data Type	Description	
marital_status	marital_status_id (PK)	integer	unique patient identifier whether married or not	
	status_of_marriage	string	description on patient marital status	

ERD



Note:

- 455 are blanks mentioned as <not null> in the database on 11th primary key

PART 2

LETTER FROM CTO

Hello Lawrence & Raj,

I am glad to share that our company got a big project with XYZ Hospital of Michigan. The client has shared with us their last quarter's dataset of patient admissions and had requested us to find out from this data that how can the patient outcomes be improved. I had a detailed meeting with the hospital stakeholders yesterday, and they had highlighted some important areas that our team can focus on, like:

- Average Length of Stay = Number of inpatient days of a patient divided by total number of inpatient days of all patients
- Outcomes = Deceased or Alive
- Diagnosis = patient condition or reason of being admitted to hospital
- Insurance Type = the type of healthcare insurance or payer type that the patient has
- Type of Admission whether planned or Emergency

The client is confident that data exploration around 'diagnosis' correlating to above mentioned areas will help improve patient outcomes, as well as give them better business solutions.

I want our DBA team along with research group to work together to find out the answers to the following questions:

Diagnosis - Mortality:

1. The most common diagnosis that results in death of a patient?
2. When admitted through Emergency, what diagnosis mostly leads to death of a patient?
3. What is the fatality rate (death of a newborn), when admission purpose is delivery of a pregnant woman?

Type of Insurance:

4. What is the insurance type patients carry when admitted to the hospital on the basis of type of admission they had?
5. Can you try to find out common diagnosis in these varied type of Insurance/Payer categories?

Average Length of Stay:

6.a Diagnosis pertaining to maximum Length of Stay on an average, if a patient is an Emergency Admission.

6.b. Diagnosis pertaining to minimum Length of Stay on an average, if a patient is an Emergency Admission.

7.a. Diagnosis pertaining to maximum Length of Stay on an average, if a patient is an Elective Admission.

7.b. Diagnosis pertaining to minimum Length of Stay on an average, if a patient is an Elective Admission.

8. Average length of stay when patient outcome is ‘death’.
9. Can we have a brief idea on maximum number of patient complaints on admissions (i.e. diagnosis)?

I recommend, Ashruti from DBA team to get guidance from Lawrence Summereset and report the answers to above questions within 2 weeks. We will organize a brief presentation with our client during the final week followed by a fun party for a huge collaboration and our future success. Looking forward to it.

Best Regards,
Kelly Davenport, CTO
Borromean Digital

PART 3

ANSWERS

BUSINESS QUESTION	ANSWER	MAJOR FOCUS AREA OF IMPROVEMENT FOR THE HOSPITAL
The most common diagnosis that results in death of a patient?	sepsis	'hospital infection' if readmission of the patient within 48 hours of discharge. Or to locate the reason of infection and patient history if patient admitted with a fresh complaint for sepsis.
When admitted through emergency, what diagnosis mostly leads to death of a patient?	sepsis	It would be interesting to know the underlying causes for sepsis or any correlated chronic condition that patient might have
What is the fatality rate (death of a newborn), when admission purpose is delivery of a pregnant woman?	0.079	0.079 deaths of newborns have occurred within first month of their birth (amongst 1000 births)
What is the insurance type patients carry when admitted to the hospital on the basis of type of admission they had?	Most of the Emergency patients had Medicare type of insurance, which is the highest in number amongst all admissions amounting up to 23590	Hospital can have more services in its 'scope' that entails under Medicare type of Health Insurance
Can you try to find out common diagnosis in these varied type of insurance/payer categories?	Nwborns are mostly selfpayers = 36, pneumonia patients are Medicare = 1124, Medicaid seems to be supporting newborns = 1236	Hospital can try to find out the reasons of why 36 deliveries were an out of pocket expenditure for some patients.
6.a diagnosis pertaining to maximum length of stay on an average, if a patient is an emergency admission.	thigh fluid collection	This could be confounded by a patient being diabetic, hence preventive measures on diabetes should be focused on. Like, organizing campaigns, awareness sessions, community announcements, etc.

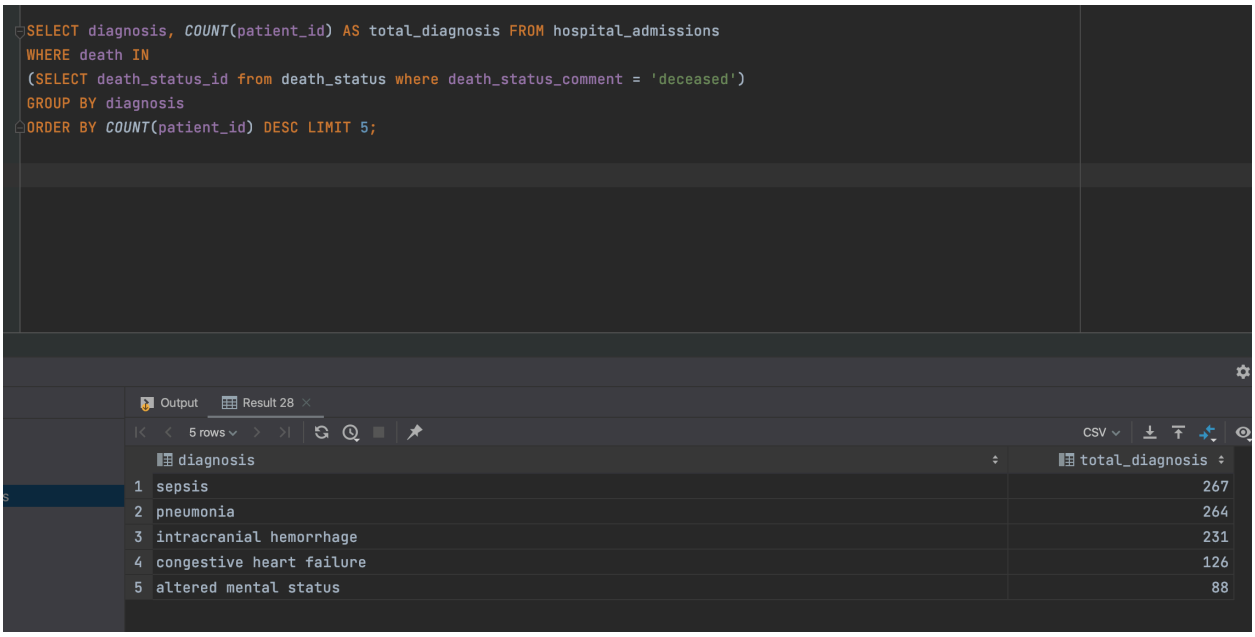
6.b. Diagnosis pertaining to minimum length of stay on an average, if a patient is an emergency admission.	intracranial hemorrhage	These patients could be sharing the numbers with death cases, as mostly intracranial cases are emergency and hence the shorter length of stay
7.a. Diagnosis pertaining to maximum length of stay on an average, if a patient is an elective admission.	abdominal fistula	invasive procedures like this could be encouraged more and scheduled well using electronic scheduling systems
7.b. Diagnosis pertaining to minimum length of stay on an average, if a patient is an elective admission.	post valve replacement	This is a day-care procedure and it shows that hospital has a good follow-up system in place. Could be enhanced by automated or AI generative messages to patients for regular reminders on their follow-up
Average length of stay when patient outcome is 'death'.	10.13	Hospital should focus on getting one bed being occupied for 10 days, especially when the patient is critical. Good TRIAGing practices should be in place and policies around having average length no more than 6 inpatient days.
Can we have a brief idea on maximum number of patient complaints on admissions (i.e. Diagnosis)?	Spondylolisthesis seems to be the diagnosis affecting most number of patients	This is evident due to change in today's generations lifestyle and could also be a COVID impact, as people are working from home and seem to have a sedentary lifestyle. Preventive measures and wellness activity camps would be best in such cases.

Queries and **SQL Outputs** are from the next page, *for your reference:*
(Appendix-3)

Diagnosis – Mortality

1. The most common diagnosis that results in death of a patient?

SELECT diagnosis, COUNT(patient_id) AS total_diagnosis FROM hospital_admissions
WHERE death IN
(SELECT death_status_id from death_status where death_status_comment = 'deceased')
GROUP BY diagnosis
ORDER BY COUNT(patient_id) DESC LIMIT 5;



2. When admitted through Emergency, what diagnosis mostly leads to death of a patient?

SELECT diagnosis, COUNT(patient_id) AS total_diag_death_ER FROM
hospital_admissionsWHERE ((death IN
(SELECT death_status_id from death_status where death_status_comment = 'deceased'))
AND (admission_type IN (SELECT admission_type_id FROM admission_type WHERE
type_of_admission = ‘emergency’)))
GROUP BY diagnosis
ORDER BY COUNT(patient_id) DESC LIMIT 5;

```
SELECT diagnosis, COUNT(patient_id) AS total_diag_death_ER FROM hospital_admissions
WHERE ((death IN
(SELECT death_status_id from death_status where death_status_comment = 'deceased'))
AND (admission_type IN (SELECT admission_type_id FROM admission_type WHERE type_of_admission = 'emergency')))
GROUP BY diagnosis
ORDER BY COUNT(patient_id) DESC LIMIT 5;
```

Output | Result 27

5 rows

	diagnosis	total_diag_death_ER
1	sepsis	266
2	pneumonia	263
3	intracranial hemorrhage	230
4	congestive heart failure	122
5	altered mental status	88

3. What is the fatality rate (death of a newborn), when admission purpose is delivery of a pregnant woman?

SELECT (SELECT COUNT(patient_id)
FROM hospital_admissions WHERE
((admission_type in (select admission_type_id from admission_type where type_of_admission = 'newborn'))
and (death IN (select death_status_id from death_status where death_status_comment = 'deceased'))))/COUNT(patient_id) AS fatality_rates
FROM hospital_admissions
WHERE admission_type IN (SELECT admission_type_id from admission_type WHERE type_of_admission = 'newborn');

```
SELECT (SELECT COUNT(patient_id)
FROM hospital_admissions WHERE
((admission_type in (select admission_type_id from admission_type where type_of_admission = 'newborn'))
and (death IN (select death_status_id from death_status where death_status_comment = 'deceased'))))/COUNT(patient_id) AS fatality_r
FROM hospital_admissions
WHERE admission_type IN (SELECT admission_type_id from admission_type WHERE type_of_admission = 'newborn');
```

Output | fatality_rates:int

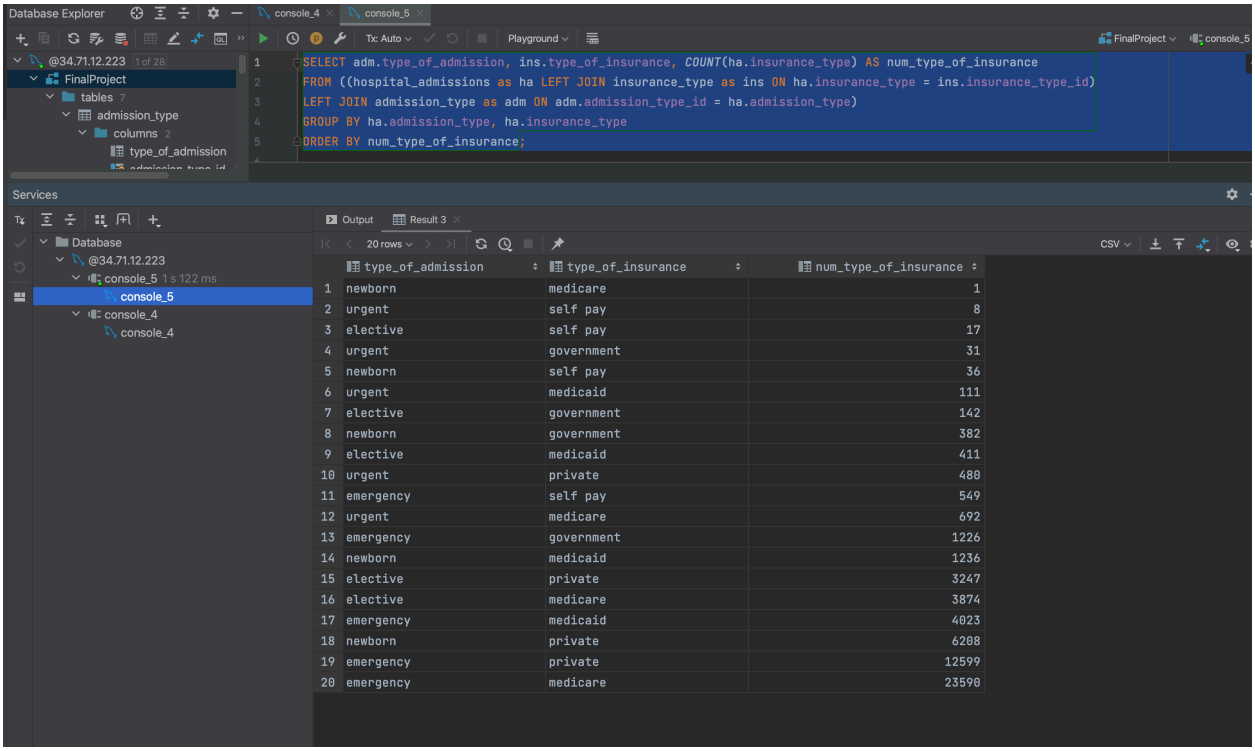
1 row

	fatality_rates
1	0.0079

Type of Insurance

4. What is the insurance type patients carry when admitted to the hospital on the basis of type of admission they had?

SELECT adm.type_of_admission, ins.type_of_insurance, COUNT(ha.insurance_type) AS num_type_of_insurance FROM ((hospital_admissions as ha LEFT JOIN insurance_type as ins ON ha.insurance_type = ins.insurance_type_id) LEFT JOIN admission_type as adm ON adm.admission_type_id = ha.admission_type) GROUP BY ha.admission_type, ha.insurance_type, ORDER BY num_type_of_insurance;



	type_of_admission	type_of_insurance	num_type_of_insurance
1	newborn	medicare	1
2	urgent	self pay	8
3	elective	self pay	17
4	urgent	government	31
5	newborn	self pay	36
6	urgent	medicaid	111
7	elective	government	142
8	newborn	government	382
9	elective	medicaid	411
10	urgent	private	480
11	emergency	self pay	549
12	urgent	medicare	692
13	emergency	government	1226
14	newborn	medicaid	1236
15	elective	private	3247
16	elective	medicare	3874
17	emergency	medicaid	4023
18	newborn	private	6208
19	emergency	private	12599
20	emergency	medicare	23590

5. Can you try to find out common diagnosis in these varied type of Insurance/Payer categories?

SELECT diagnosis, COUNT(patient_id) as medicare_common_diagnosis FROM hospital_admissions WHERE insurance_type
IN
(SELECT insurance_type_id FROM insurance_type WHERE type_of_insurance = “xxxxxxx”)
GROUP BY diagnosis
ORDER BY COUNT(patient_id)

SelfPay:

```
SELECT diagnosis, COUNT(patient_id) as SelfPay_common_diagnosis
FROM hospital_admissions WHERE insurance_type IN
(SELECT insurance_type_id FROM insurance_type WHERE type_of_insurance='self pay')
GROUP BY diagnosis
ORDER BY COUNT(patient_id) DESC LIMIT 5;
```

Output | Result 7 | 5 rows | CSV

	diagnosis	SelfPay_common_diagnosis
1	newborn	36
2	s/p fall	20
3	overdose	15
4	intracranial hemorrhage	13
5	altered mental status	13

51 ms
951 ms

Medicare:

```
SELECT diagnosis, COUNT(patient_id) as Medicare_common_diagnosis
FROM hospital_admissions WHERE insurance_type IN
(SELECT insurance_type_id FROM insurance_type WHERE type_of_insurance='medicare')
GROUP BY diagnosis
ORDER BY COUNT(patient_id) DESC LIMIT 5;
```

Output | Result 10 | 5 rows

	diagnosis	Medicare_common_diagnosis
1	pneumonia	1124
2	sepsis	814
3	congestive heart failure	727
4	chest pain	475
5	coronary artery disease	464

Medicaid:

```
SELECT diagnosis, COUNT(patient_id) as Medicaid_common_diagnosis
FROM hospital_admissions WHERE insurance_type IN
(SELECT insurance_type_id FROM insurance_type WHERE type_of_insurance='medicaid')
GROUP BY diagnosis
ORDER BY COUNT(patient_id) DESC LIMIT 5;
```

Output | Result 11

5 rows | CSV

	diagnosis	Medicaid_common_diagnosis
1	newborn	1236
2	pneumonia	126
3	diabetic ketoacidosis	123
4	altered mental status	119
5	sepsis	94

Average Length of Stay

6.a. Diagnosis pertaining to maximum Length of Stay on an average, if a patient is an Emergency Admission.

SELECT diagnosis, AVG(admission_length_days) AS LOS FROM hospital_admissions WHERE
((admission_type IN (SELECT admission_type_id from admission_type WHERE type_of_admission = 'emergency')) AND
Admission_length_days > 0)
GROUP BY diagnosis
ORDER BY LOS DESC LIMIT 8;

```
SELECT diagnosis, AVG(admission_length_days) AS LOS FROM hospital_admissions WHERE
((admission_type IN (SELECT admission_type_id from admission_type WHERE type_of_admission = 'emergency')) AND
Admission_length_days > 0)
GROUP BY diagnosis
ORDER BY LOS DESC LIMIT 8;
```

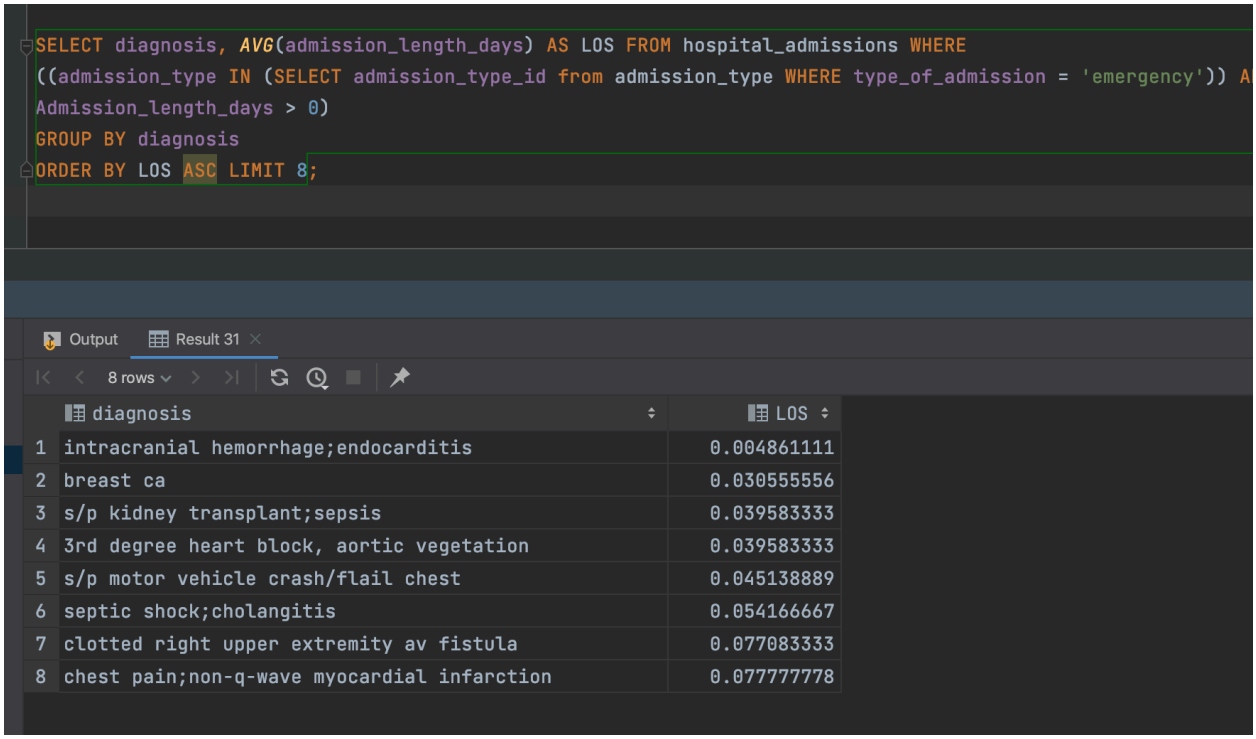
Output | Result 30

8 rows | CSV

	diagnosis	LOS
1	l thigh fluid collection	166.9069444
2	amc;fever	164.3659722
3	s/p liver tx-nausea-vomiting	137.9652778
4	aplastic anemia;pancytopenia	137.8916667
5	abdominal fistula	132.8951389
6	pulmonary embolism;subdural hematoma	131.9708333
7	acute myelogenous leukemia;chemotherapy	129.9604167
8	multiple myeloma;fever;neutropenia	129.0354167

6.b. Diagnosis pertaining to minimum Length of Stay on an average, if a patient is an Emergency Admission.

SELECT diagnosis, AVG(admission_length_days) AS LOS FROM hospital_admissions
WHERE
((admission_type IN (SELECT admission_type_id from admission_type WHERE
type_of_admission = 'emergency')) AND
Admission_length_days > 0)
GROUP BY diagnosis
ORDER BY LOS ASC LIMIT 8;



```
SELECT diagnosis, AVG(admission_length_days) AS LOS FROM hospital_admissions WHERE
((admission_type IN (SELECT admission_type_id from admission_type WHERE type_of_admission = 'emergency')) A
Admission_length_days > 0)
GROUP BY diagnosis
ORDER BY LOS ASC LIMIT 8;
```

Output | Result 31

8 rows

	diagnosis	LOS
1	intracranial hemorrhage;endocarditis	0.004861111
2	breast ca	0.030555556
3	s/p kidney transplant;sepsis	0.039583333
4	3rd degree heart block, aortic vegetation	0.039583333
5	s/p motor vehicle crash/flail chest	0.045138889
6	septic shock;cholangitis	0.054166667
7	clotted right upper extremity av fistula	0.077083333
8	chest pain;non-q-wave myocardial infarction	0.077777778

7.a. Diagnosis pertaining to maximum Length of Stay on an average, if a patient is an Elective Admission.

SELECT diagnosis, AVG(admission_length_days) AS LOS FROM hospital_admissions
WHERE
((admission_type IN (SELECT admission_type_id from admission_type WHERE
type_of_admission = 'elective')) AND
Admission_length_days > 0)
GROUP BY diagnosis
ORDER BY LOS DESC LIMIT 8;

```
SELECT diagnosis, AVG(admission_length_days) AS LOS FROM hospital_admissions WHERE
((admission_type IN (SELECT admission_type_id from admission_type WHERE type_of_admission = 'elective')) AND
Admission_length_days > 0)
GROUP BY diagnosis
ORDER BY LOS DESC LIMIT 8;
```

Output | Result 32

8 rows | CSV

	diagnosis	LOS
1	crohn's disease;abdominal fistula	294.6604167
2	aspiration; failure to thrive	191.4229167
3	chronic lymphocytic leukemia\bone marrow transplant	169.1875
4	myelodysplasia;bone marrow transplant/sda	126.2708333
5	large abdominal fistula	123.6875
6	biliary leak/sda	88.69722222
7	airway obstruction\ bronchoscopy with stent	84.08680556
8	multiple myeloma\bone marrow transplant	81.13460647666666

7.b. Diagnosis pertaining to minimum Length of Stay on an average, if a patient is an Elective Admission.

SELECT diagnosis, AVG(admission_length_days) AS LOS FROM hospital_admissions WHERE
((admission_type IN (SELECT admission_type_id from admission_type WHERE type_of_admission = 'elective')) AND
Admission_length_days > 0)
GROUP BY diagnosis
ORDER BY LOS ASC LIMIT 8;

```
SELECT diagnosis, AVG(admission_length_days) AS LOS FROM hospital_admissions WHERE
((admission_type IN (SELECT admission_type_id from admission_type WHERE type_of_admission = 'elective')) AND
Admission_length_days > 0)
GROUP BY diagnosis
ORDER BY LOS ASC LIMIT 8;
```

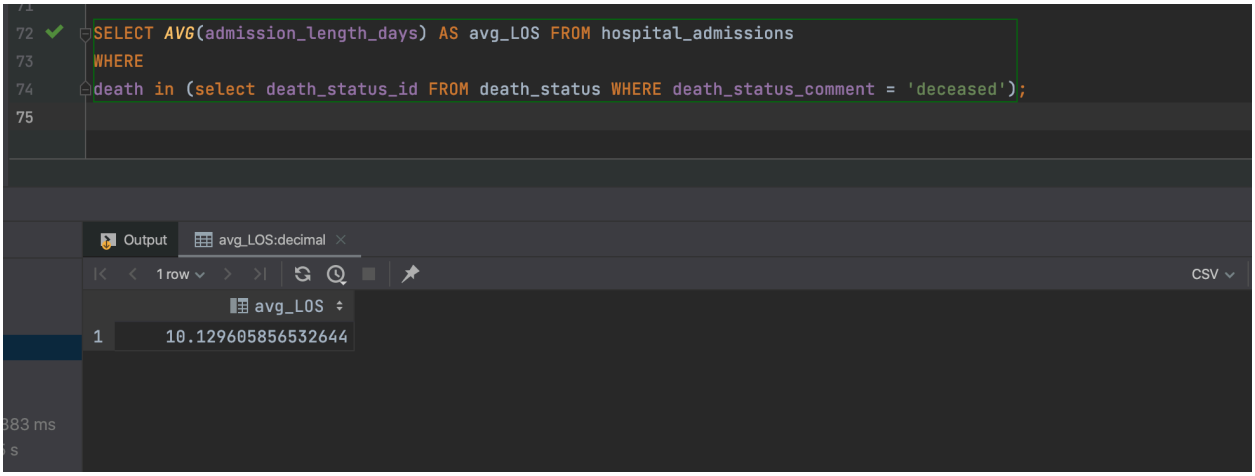
Output | Result 33

8 rows

	diagnosis	LOS
1	status post valve replacement\debridement ster…	0.069444444
2	hematuria/sda	0.096527778
3	mesh infection, umbilical wound/sda	0.153472222
4	right neck node/sda	0.15625
5	chf, left ventricular lead causing diaphragmat…	0.166666667
6	hydronephrosis & ureteral stent obstruction/sda	0.173611111
7	ulcer right leg/sda	0.177083333
8	open wound abdominal wall/sda	0.179861111

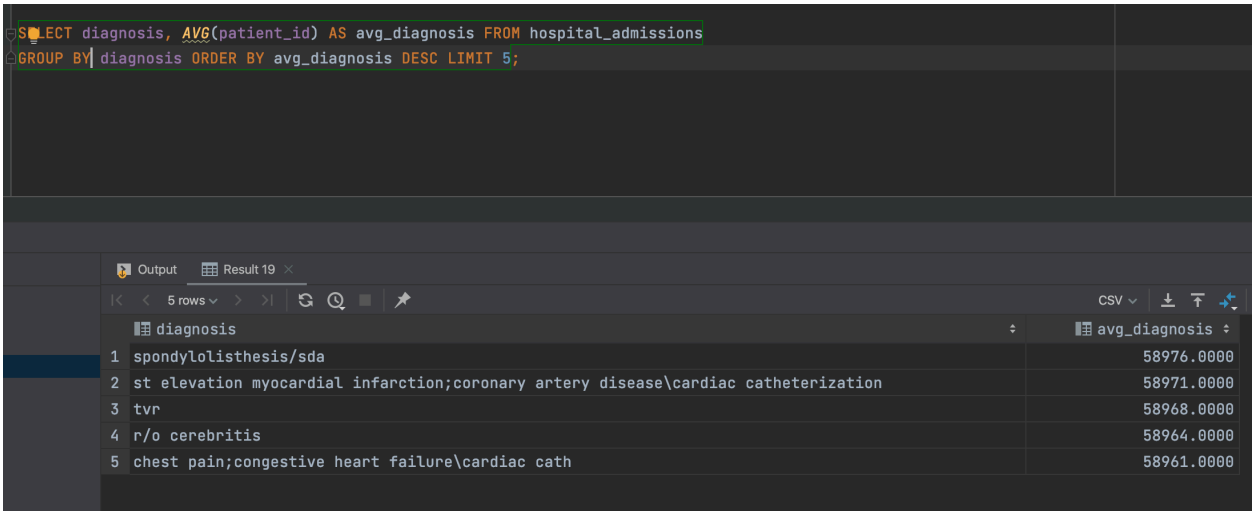
8. Average length of stay when patient outcome is ‘death’.

SELECT AVG(admission_length_days) AS avg_LOS FROM hospital_admissions
WHERE
death in (select death_status_id FROM death_status WHERE death_status_comment = 'deceased');



9. Can we have a brief idea on maximum number of patient complaints on admissions (i.e. diagnosis)?

SELECT diagnosis, AVG(patient_id) AS avg_diagnosis FROM hospital_admissions
GROUP BY diagnosis ORDER BY avg_diagnosis DESC LIMIT 5;



PART 4

DATABASE CHOICE

I am glad to share the what motivated to chose this dataset!
I always wanted to perform a case study to analyze the data and find answers to business questions. Especially coming from a healthcare background, it motivated me to pick ;hospital admissions dataset.'
Moreover the choice of dataset was also because this data had categorical variables which I assumed would be easier to normalize. And there were following reasons around the database choice:

- Use of datagrip made it easier to remove patient Identifiers and replace it with auto-incremental integer values that datagrip generated.
- The creation of subset tables removed the redundancies from the main table as all the subset tables were related with the main tables using foreign key restrictions.
- The main table had Foreign keys associated with all the subset tables, which made it easier to recognize the data in the main table referencing the PRImary keys from the subset tables. Hence, the data came out to be clean and organized.
- The subset tables were feeded with string values but the main table referred to the integer ~ primary keys of these values, which made querying the tables easier and organized.
- Moreover, healthcare industry related findings have a huge impact in the real world, and can change the patient outcomes as whole if answers to healthcare complications are made using analytical skills.
- The answers to complex healthcare questions play a vital role in reframing healthcare policies which have an impact on patient outcomes as a whole.
- In nutshell, I found it interesting, engaging to work with this dataset using SQL queries and datagrip.

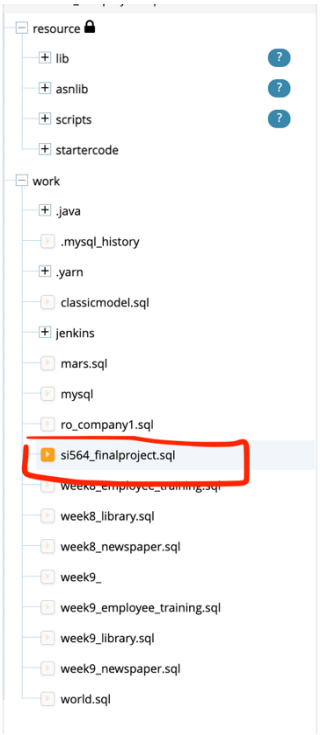
I tried my best to incorporate insert, update, drop while creating tables, and WHERE clause, subqueries to answer the above questions. Please feel free to let me know if you want me to walk you through any of these approaches.
Also,
I wanted to thank the Borromean Digital Solutions for helping me grow professionally and been given this opportunity to work with a new collaboration now. I am looking forward to the presentation followed by the part.
Thank you!

Regards,
Ashruti Tuteja | tashruti@umich.edu
Database Administrator
Borromean Digital Information

PART 5

DUMP FILE

Please find the file attached for your reference, with the name: **si564_finalproject.sql**



CREATING DUMP FILE

1. mysqldump -h 34.71.12.223 --port 10785 -u tashruti-rw ***--set-gtid-purged=OFF*** -p FinalProject > si564_finalproject.sql
2. In rw server, CREATE DATABASE si564_finalproject

```
labsuser@host:~$ mysqldump -h 34.71.12.223 --port 10785 -u tashruti-rw --set-gtid-purged=OFF -p Fina
lProject > si564_finalproject.sql
Enter password:
labsuser@host:~$ mysql -h 34.71.12.223 --port 10785 -u tashruti-rw -pf863af2d752fa0f075c9c80503e3d64
6
mysql: [Warning] Using a password on the command line interface can be insecure.
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 2141
Server version: 8.0.27 MySQL Community Server - GPL

Copyright (c) 2000, 2020, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> create database si564_finalproject
    -> create database si564_finalproject;
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your
MySQL server version for the right syntax to use near 'create database si564_finalproject' at line 2
mysql> create database si564_finalproject;
Query OK, 1 row affected (0.15 sec)

mysql>
```

3. mysql -h 34.71.12.223 --port 10785 -u tashruti-rw -p si564_finalproject < si564_finalproject.sql
4. Verify the data populated in si564_finalproject database

```
labsuser@host:~$ mysql -h 34.71.12.223 --port 10785 -u tashruti-ro -p5fad9a3b3674876c27d46cfcbbe5a744
mysql: [Warning] Using a password on the command line interface can be insecure.
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 2200
Server version: 8.0.27 MySQL Community Server - GPL

Copyright (c) 2000, 2020, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> show databases;
+----------------------------+
| Database                   |
+----------------------------+
| FinalProject               |
| bikes                      |
| classicmodels              |
| employee_training          |
| information_schema         |
| kubedb_system              |
| library                    |
| mars                       |
| mysql                      |
| newspaper                  |
| performance_schema         |
| ro_company1                |
| ro_company2                |
| ro_employees               |
| ro_query                   |
| ro_recipes                 |
| ro_research1               |
| ro_twitter                 |
| sakila                     |
| schema_name                |
| si564_finalproject         |
| sys                        |
| week8_employee_training    |
| week8_library              |
| week8_newspaper            |
| week9_employee_training    |
| week9_library              |
| week9_newspaper            |
| world                      |
+----------------------------+
29 rows in set (0.06 sec)

mysql> use si564_finalproject;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> show tables;
+-----------------------------+
| Tables_in_si564_finalproject |
+-----------------------------+
| admission_type              |
| death_status                |
| hospital_admissions         |
| insurance_type              |
```

APPENDIX

Appendix-1

Data Normalization

CREATING TABLES and ASSIGNING VALUES

```
2023-04-16 18:06:26] Connected
FinalProject> use FinalProject
[2023-04-16 18:06:27] completed in 70 ms
FinalProject> SELECT t.*
        FROM FinalProject.religion t
        LIMIT 501
[2023-04-16 18:06:27] 0 rows retrieved in 119 ms (execution: 97 ms, fetching: 22 ms)
FinalProject> SELECT t.*
        FROM FinalProject.religion t
        ORDER BY religion_id
        LIMIT 501
[2023-04-16 18:06:51] 0 rows retrieved in 103 ms (execution: 95 ms, fetching: 8 ms)
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('episcopalian')
[2023-04-16 18:09:38] 1 row affected in 72 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('catholic')
[2023-04-16 18:09:38] 1 row affected in 46 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('protestant quaker')
[2023-04-16 18:09:38] 1 row affected in 48 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('christian scientist')
[2023-04-16 18:09:38] 1 row affected in 49 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('other')
[2023-04-16 18:09:38] 1 row affected in 45 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('jewish')
[2023-04-16 18:09:38] 1 row affected in 45 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('not specified')
[2023-04-16 18:09:39] 1 row affected in 48 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('unobtainable')
[2023-04-16 18:09:39] 1 row affected in 48 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('greek orthodox')
[2023-04-16 18:09:39] 1 row affected in 48 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('buddhist')
[2023-04-16 18:09:39] 1 row affected in 47 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES (null)
[2023-04-16 18:09:39] 1 row affected in 48 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('baptist')
[2023-04-16 18:09:39] 1 row affected in 47 ms
```

```
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('jehovah\'s witness')
[2023-04-16 18:09:39] 1 row affected in 44 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('unitarian-universalist')
[2023-04-16 18:09:39] 1 row affected in 47 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('muslim')
[2023-04-16 18:09:39] 1 row affected in 45 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('romanian east. orth')
[2023-04-16 18:09:39] 1 row affected in 46 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('7th day adventist')
[2023-04-16 18:09:39] 1 row affected in 46 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('methodist')
[2023-04-16 18:09:39] 1 row affected in 45 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('hindu')
[2023-04-16 18:09:39] 1 row affected in 49 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('hebrew')
[2023-04-16 18:09:39] 1 row affected in 45 ms
FinalProject> INSERT INTO FinalProject.religion (religion_name) VALUES ('lutheran')
[2023-04-16 18:09:40] 1 row affected in 46 ms
FinalProject> SELECT t.*
        FROM FinalProject.religion t
        ORDER BY religion_id
        LIMIT 501
[2023-04-16 18:09:40] 21 rows retrieved starting from 1 in 108 ms (execution: 91 ms, fetching:
17 ms)
FinalProject> SELECT t.*
        FROM FinalProject.religion t
        ORDER BY religion_id
        LIMIT 501
[2023-04-16 18:13:16] 21 rows retrieved starting from 1 in 101 ms (execution: 93 ms, fetching:
8 ms)
```

QUERIES USED TO FIND VALUES FROM THE MAIN TABLE

```
select distinct Insurance_Type from hospital_admissions;
select distinct Admission_type from hospital_admissions;
select distinct Death_1 from hospital_admissions;
select distinct Married_1 from hospital_admissions;
select distinct Race from hospital_admissions;
select distinct Religion_Type from hospital_admissions;
```

MODIFIED COLUMN NAMES OF THE MAIN TABLE FOR EASE

use FinalProject
[2023-04-16 18:40:11] completed in 70 ms
FinalProject> alter table hospital_admissions
change ID patient_id int auto_increment
[2023-04-16 18:40:11] completed in 67 ms
FinalProject> alter table hospital_admissions
change AdmissionLengthDays admission_length_days double null
[2023-04-16 18:40:11] completed in 62 ms
FinalProject> alter table hospital_admissions
change Death_1 death int null
[2023-04-16 18:40:11] completed in 59 ms
FinalProject> alter table hospital_admissions
change Admission_type admission_type text null
[2023-04-16 18:40:12] completed in 92 ms
FinalProject> alter table hospital_admissions
change Insurance_Type insurance_type text null
[2023-04-16 18:40:12] completed in 59 ms
FinalProject> alter table hospital_admissions
change EnglishLanguage_1 english_language int null
[2023-04-16 18:40:12] completed in 57 ms
FinalProject> alter table hospital_admissions
change Religion_Type religion_type text null
[2023-04-16 18:40:12] completed in 62 ms
FinalProject> alter table hospital_admissions
change Married_1 married int null
[2023-04-16 18:40:12] completed in 58 ms
FinalProject> alter table hospital_admissions
change Race race text null
[2023-04-16 18:40:12] completed in 69 ms
FinalProject> alter table hospital_admissions
change Dx diagnosis text null
[2023-04-16 18:40:12] completed in 63 ms

CONNECTING TABLES

Updating Main table with values (primary keys) from the other subsidiary tables:

```
update hospital_admissions h set insurance_type = (select insurance_type_id
from insurance_type it where it.type_of_insurance = h.insurance_type);
update hospital_admissions h set race = (select race_id from race r where
r.race_type = h.race);
update hospital_admissions h set religion_type = (select religion_id from
religion rl where rl.religion_name = h.religion_type);
update hospital_admissions h set race = (select race_id from race r where
r.race_type = h.race);
```

```
update hospital_admissions h set admission_type = (select admission_type_id
from admission_type ad where ad.type_of_admission = h.admission_type);
```

This results in now all the values in our main table to be numeric, except the column values for 'diagnosis' column. And it makes it now easier for assigning Foreign keys.

ASSIGNING FOREIGN KEYS

Alter the tables from varchar() type -> integer type for all the connected columns like race, religion, insurance_type and admission_type

```
alter table hospital_admissions
        modify insurance_type int null
[2023-04-16 19:13:14] completed in 1 s 270 ms
FinalProject> alter table hospital_admissions
        modify religion_type int null
[2023-04-16 19:13:15] completed in 1 s 330 ms
FinalProject> alter table hospital_admissions
        modify race int null
```

Foreign Keys

```
use FinalProject
[2023-04-16 19:19:35] completed in 67 ms
FinalProject> alter table hospital_admissions
        add constraint hospital_admissions_race_race_id_fk
          foreign key (race) references race (race_id)
[2023-04-16 19:19:37] completed in 2 s 152 ms
FinalProject> alter table hospital_admissions
        add constraint hospital_admissions_religion_religion_id_fk
          foreign key (religion_type) references religion (religion_id)
[2023-04-16 19:20:39] completed in 2 s 911 ms
FinalProject> alter table hospital_admissions
        add constraint hospital_admissions_death_status_death_status_id_fk
          foreign key (death) references death_status (death_status_id)
[2023-04-16 19:22:22] completed in 2 s 858 ms
FinalProject> alter table hospital_admissions
        add constraint hospital_admissions_marital_status_marital_status_id_fk
          foreign key (married) references marital_status (marital_status_id)
[2023-04-16 19:23:33] completed in 2 s 879 ms
FinalProject> alter table hospital_admissions
        add constraint hospital_admissions_admission_type_admission_type_id_fk
          foreign key (admission_type) references admission_type
(admission_type_id)
```

Learnings from data Normalization

Marital Status and Death Status, i.e., the categorical variables does not require 'Auto_increament,' rather we need to specify what does the category mean. This could be done either by querying the data, or simply in Datagrip by typing the category description manually.
If query: SELECT
  CASE
    WHEN death_status = 0 THEN 'deceased'
    WHEN death_status = 1 THEN 'alive'
  END AS status
FROM
  hospital_admissions;

Appendix-2

Tables and Data Types

TABLES IN DATABASE

```
mysql> use FinalProject
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> show tables;
+------------------------+
| Tables_in_FinalProject |
+------------------------+
| admission_type         |
| death_status           |
| hospital_admissions    |
| insurance_type         |
| marital_status         |
| race                   |
| religion               |
+------------------------+
7 rows in set (0.07 sec)
```

Main Table: hospital_admissions

```
mysql> desc hospital_admissions;
+------------------------+--------+------+-----+---------+----------------+
| Field                  | Type   | Null | Key | Default | Extra          |
+------------------------+--------+------+-----+---------+----------------+
| patient_id             | int    | NO   | PRI | NULL    | auto_increment |
| admission_length_days  | double | YES  |     | NULL    |                |
| death                  | int    | YES  | MUL | NULL    |                |
| admission_type         | int    | YES  | MUL | NULL    |                |
| insurance_type         | int    | YES  | MUL | NULL    |                |
| english_language       | int    | YES  |     | NULL    |                |
| religion_type          | int    | YES  | MUL | NULL    |                |
| married                | int    | YES  | MUL | NULL    |                |
| race                   | int    | YES  | MUL | NULL    |                |
| diagnosis              | text   | YES  |     | NULL    |                |
+------------------------+--------+------+-----+---------+----------------+
```

Other tables in the database

```
mysql> desc admission_type;
+--------------------+--------------+------+-----+---------+----------------+
| Field              | Type         | Null | Key | Default | Extra          |
+--------------------+--------------+------+-----+---------+----------------+
| type_of_admission  | varchar(255) | YES  |     | NULL    |                |
| admission_type_id  | int          | NO   | PRI | NULL    | auto_increment |
+--------------------+--------------+------+-----+---------+----------------+
2 rows in set (0.06 sec)

mysql> desc insurance_type;
+--------------------+--------------+------+-----+---------+----------------+
| Field              | Type         | Null | Key | Default | Extra          |
+--------------------+--------------+------+-----+---------+----------------+
| type_of_insurance  | varchar(50) | YES  |     | NULL    |                |
| insurance_type_id  | int          | NO   | PRI | NULL    | auto_increment |
+--------------------+--------------+------+-----+---------+----------------+
2 rows in set (0.06 sec)

mysql> desc marital_status;
+---------------------+--------------+------+-----+---------+-------+
| Field               | Type         | Null | Key | Default | Extra |
+---------------------+--------------+------+-----+---------+-------+
| status_of_marriage  | varchar(255) | YES  |     | NULL    |       |
| marital_status_id   | int          | NO   | PRI | NULL    |       |
+---------------------+--------------+------+-----+---------+-------+
2 rows in set (0.06 sec)

mysql> desc death_status;
+----------------------+--------------+------+-----+---------+-------+
| Field                | Type         | Null | Key | Default | Extra |
+----------------------+--------------+------+-----+---------+-------+
| death_status_comment | varchar(255) | YES  |     | NULL    |       |
| death_status_id      | int          | NO   | PRI | NULL    |       |
+----------------------+--------------+------+-----+---------+-------+
2 rows in set (0.07 sec)

mysql> desc race;
+-----------+-------------+------+-----+---------+----------------+
| Field     | Type        | Null | Key | Default | Extra          |
+-----------+-------------+------+-----+---------+----------------+
| race_type | varchar(50) | YES  |     | NULL    |                |
| race_id   | int         | NO   | PRI | NULL    | auto_increment |
+-----------+-------------+------+-----+---------+----------------+
2 rows in set (0.06 sec)

mysql> desc religion;
+---------------+--------------+------+-----+---------+----------------+
| Field         | Type         | Null | Key | Default | Extra          |
+---------------+--------------+------+-----+---------+----------------+
| religion_name | varchar(255) | YES  |     | NULL    |                |
| religion_id   | int          | NO   | PRI | NULL    | auto_increment |
+---------------+--------------+------+-----+---------+----------------+
2 rows in set (0.06 sec)
```

Appendix-3

Tables and Data Types

QUERIES

```
SELECT adm.type_of_admission, ins.type_of_insurance,
COUNT(ha.insurance_type) AS num_type_of_insurance
FROM ((hospital_admissions as ha LEFT JOIN insurance_type as ins ON
ha.insurance_type = ins.insurance_type_id)
LEFT JOIN admission_type as adm ON adm.admission_type_id =
ha.admission_type)
GROUP BY ha.admission_type, ha.insurance_type
ORDER BY num_type_of_insurance;

SELECT diagnosis, COUNT(patient_id) as SelfPay_common_diagnosis
FROM hospital_admissions WHERE insurance_type IN
(SELECT insurance_type_id FROM insurance_type WHERE
type_of_insurance='self pay')
GROUP BY diagnosis
ORDER BY COUNT(patient_id) DESC LIMIT 5;

SELECT diagnosis, COUNT(patient_id) as Medicare_common_diagnosis
FROM hospital_admissions WHERE insurance_type IN
(SELECT insurance_type_id FROM insurance_type WHERE
type_of_insurance='medicare')
GROUP BY diagnosis
```

```
ORDER BY COUNT(patient_id) DESC LIMIT 5;

SELECT diagnosis, COUNT(patient_id) as Medicaid_common_diagnosis
FROM hospital_admissions WHERE insurance_type IN
(SELECT insurance_type_id FROM insurance_type WHERE
type_of_insurance='medicaid')
GROUP BY diagnosis
ORDER BY COUNT(patient_id) DESC LIMIT 5;

SELECT diagnosis, COUNT(patient_id) AS total_diagnosis FROM
hospital_admissions
WHERE death IN
(SELECT death_status_id from death_status where death_status_comment =
'deceased')
GROUP BY diagnosis
ORDER BY COUNT(patient_id) DESC LIMIT 5;

SELECT diagnosis, AVG(patient_id) AS avg_diagnosis FROM
hospital_admissions
GROUP BY diagnosis ORDER BY avg_diagnosis DESC LIMIT 5;

SELECT diagnosis, COUNT(patient_id) AS total_diag_death_ER FROM
hospital_admissions
WHERE ((death IN
(SELECT death_status_id from death_status where death_status_comment =
'deceased'))
AND (admission_type IN (SELECT admission_type_id FROM admission_type
WHERE type_of_admission = 'emergency')))
GROUP BY diagnosis
ORDER BY COUNT(patient_id) DESC LIMIT 5;


SELECT (SELECT COUNT(patient_id)
FROM hospital_admissions WHERE
((admission_type in (select admission_type_id from admission_type
where type_of_admission = 'newborn'))
and (death IN (select death_status_id from death_status where
death_status_comment = 'deceased'))))/COUNT(patient_id) AS
fatality_rates
FROM hospital_admissions
WHERE  admission_type IN (SELECT admission_type_id from admission_type
WHERE type_of_admission = 'newborn');

SELECT diagnosis, AVG(admission_length_days) AS LOS FROM
hospital_admissions WHERE
((admission_type IN (SELECT admission_type_id from admission_type
WHERE type_of_admission = 'emergency')) AND
Admission_length_days > 0)
GROUP BY diagnosis
ORDER BY LOS DESC LIMIT 8;

SELECT diagnosis, AVG(admission_length_days) AS LOS FROM
```

```
hospital_admissions WHERE
((admission_type IN (SELECT admission_type_id from admission_type
WHERE type_of_admission = 'emergency')) AND
Admission_length_days > 0)
GROUP BY diagnosis
ORDER BY LOS ASC LIMIT 8;

SELECT diagnosis, AVG(admission_length_days) AS LOS FROM
hospital_admissions WHERE
((admission_type IN (SELECT admission_type_id from admission_type
WHERE type_of_admission = 'elective')) AND
Admission_length_days > 0)
GROUP BY diagnosis
ORDER BY LOS DESC LIMIT 8;

SELECT diagnosis, AVG(admission_length_days) AS LOS FROM
hospital_admissions WHERE
((admission_type IN (SELECT admission_type_id from admission_type
WHERE type_of_admission = 'elective')) AND
Admission_length_days > 0)
GROUP BY diagnosis
ORDER BY LOS ASC LIMIT 8;

SELECT AVG(admission_length_days) AS avg_LOS FROM hospital_admissions
WHERE
death in (select death_status_id FROM death_status WHERE
death_status_comment = 'deceased');
```

THANK YOU