# 

# **Findings on The Characteristics of Patients and Their Treatment in an Intensive Care Unit**

# **(MIMIC - III)**

# 

# **Final Project Report Group 4**

**Team Members:**

Sylvester Ntiamoah

Jose Cruz

Irene Ojukwu

Lavanya Kanagaraj

***Part 1: Acquiring and building the MIMIC III database***

**Our process of loading the MIMIC III dataset into MySQL:**

**Step 1**

Extract zip file contents into folder with multiple .csv files for each piece of the dataset

**Step 2**

There are 4 listed .sql files which will take care of importing the data from our local files and setting additional configurations. The only step here is to replace the file path locations with our local directories and run the scripts. Below are descriptions for what the .sql files do in setting up our local environments.

1-define-chartevents.sql - defines the chart events table, loads data

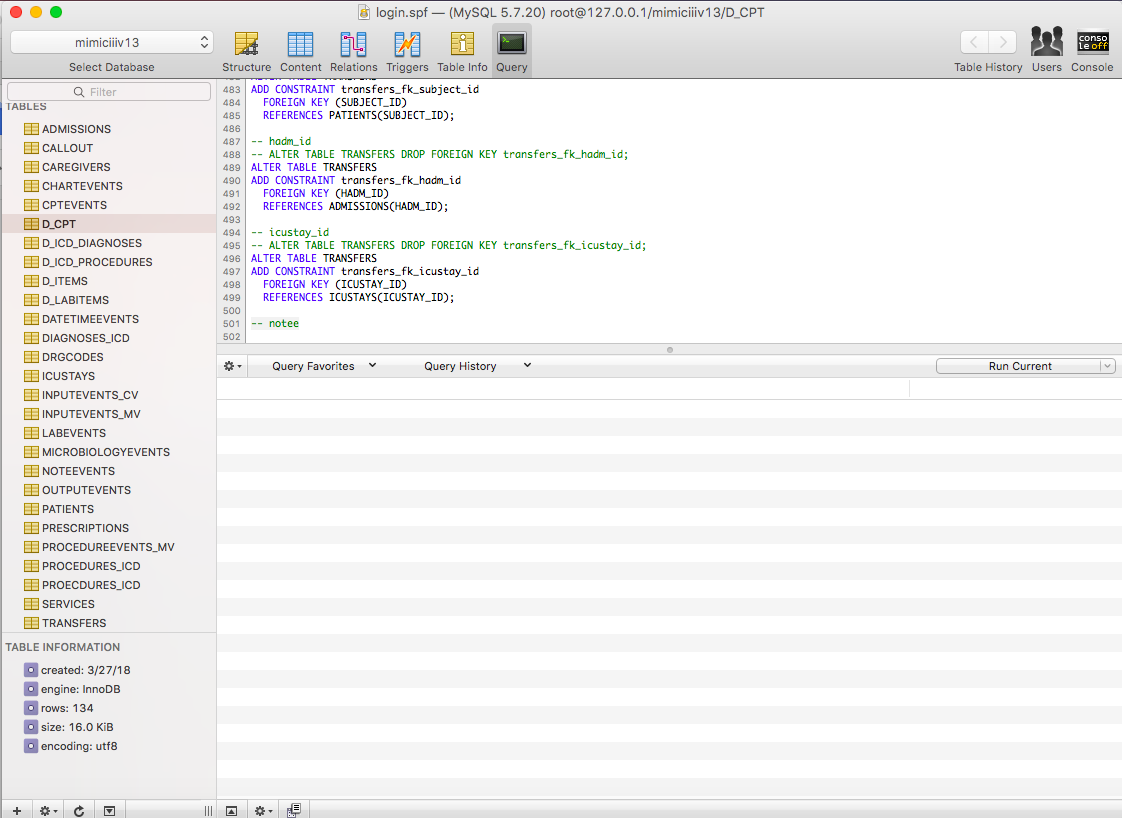
1-define.sql - defines the remaining tables, loads data

2-index.sql - sets up the indexes on the tables

3-constraints.sql - defines the constraints on the tables

**Step 3**

After running the .sql files, we have the database and tables in place, with constraints and indexing as well.



**Build process takeaways**

Configuring the database and tables took much longer than expected. With a database this size, loading the data took multiple hours. Configuring the tables (i.e. adding the constraints and keys) took even longer (upwards of 10 hours) longest since the script had to run across the entire dataset (some of these tables have upwards of millions of rows) for example when creating the individual primary/foreign keys on the dataset which is 1.8GB, which was very time consuming and resource intensive.

Following this process, we found out that increasing the buffer size and timeout configuration within MySQL speeds up the import process. We confirmed this as a team as everyone finished their database setup.

Building index on key attributes does reduce the query execution time. We verified it based on the query statistics before and after creating indexes on tables.

***Part 2: Understanding the MIMIC III database***

|  |  |  |
| --- | --- | --- |
| Table | Major Columns | Description |
| **ADMISSIONS** | **ADMITTIME**  Admission Time  **DIAGNOSIS**  Preliminary Diagnosis (textual information) | This is an entity table for admission records in that it keeps track of each time a subject is admitted into the hospital. This table is basically the record table for HADM\_ID, which is, that this table keeps a record of each admission using an ID for any (and all) patients. A patient could visit twice and have two unique HADM\_IDs in this table with resulting information. The table also includes information about the person, including admission/discharge time, insurance, and preliminary diagnosis. |
| **CALLOUT** | **CREATE\_TIME**  Creating date of the callout event  **CALLOUT\_STATUS**  Callout status | This table is an entity table for callout events in that it keeps track of each callout event for all patients. This table records the events of when a patient was ready to be discharged from ICU (intensive care unit) and when they were actually released. Neonates are not tracked as part of this table. Other attributes of the table include the ward the patient is in, and the status and acknowledgement of the callout event. |
| **CAREGIVERS** | **CGID**  Caregiver ID  **Label**  Caregiver’s Label  **Description**  The description of the caregiver e.g. Resident/Caregiver | This table is an entity table for caregivers in that it stores the record for each caregiver in the hospital along with defining information about them. It defines each individual caregiver along with a label and description for each. |
| **CHARTEVENTS** | **ITEM**  The ‘key’ value of an observation. This can be ‘stool analysis’ for example  **VALUE**  The ‘value’ of an ITEM observation | This table is an entity table for chart events in that it keeps track of each chart event for all patients. It defines the events recorded for each patient’s chart. It can be a numeric data such as blood pressure: 64, along with information about the caregiver that made the observation and when it was made. Additionally, it features some columns used in the Hospital’s software systems. |
| **CPTEVENTS** | **CPT\_CD**  Service code for the procedure serviced. This code can be looked up in the D\_CPT table  **CHARTDATE**  When the procedure occurred | This table is an association table for chart events and patients in that it resolves the relationship between the chart events and the text representation of a service code stored in D\_CPT. It stores information to help in the billing process for patients. It contains “Current Procedural Terminology” codes for services rendered to each patient, for each hospital admission. An example would be a record in the table containing information about the service ‘Evaluation and management’ rendered with a subsection of ‘Hospital Inpatient Services’ |
| **D\_CPT** | **CATEGORY**  Integer that defines the category of the CPT code  **SECTIONRANGE**  Defines the code range for a section  **SECTIONHEADER**  Defines the description of that range  **SUBSECTIONRANGE**  Defines the code for a subsection  **SUBSECTIONHEADER**  Defines the description of that subrange | This table is an entity table for chart events in that it is a reference table for CPT\_CD found in the CPTEVENTS table which are the services performed on a patient. Each record is a specific event that a patient can be serviced. It defines in descriptive language what a procedural code represents. It maps a range of procedural codes from CPTEVENTS to an overall purpose (a higher level definition) |
| **TRANSFERS** | **PREV\_CAREUNIT**  Previous care unit  **CURR\_CAREUNIT**  Current care unit  **INTIME**  For each current unit, gives the in time (time they entered)  **OUTTIME**  For each current current unit, gives the out time (time they left) | This table is an entity table for transfer events. It defines their physical location at different points during their hospital stay. It gives locations as well as times for when the transfers happened. |
| **NOTEEVENTS** | **CHARTDATE**  Date on which note entered  **CATEGORY**  Note type  **DESCRIPTION**  Details of the note  **CGID**  Caregiver ID  **ISERROR**  If note is error | This table is used to store the notes added to the patient records. We can see the description of the notes, the caregiver ID who charted the notes and the description of the notes. It can be linked to patient table, and admission table(if inpatient). |
| **OUTPUTEVENTS** | **SUBJECT\_ID**  Patient  **HADM\_ID**  Hospital admission ID  **ICUSTAY\_ID**  Patient ICU ID  **CHARTTIME**  Output event Time  **ITEMID**  Measurement type  **VALUE & VALUEUOM**  Amount of the substance  **STORETIME**  Input time  **CGID**  Caregiver ID  **ISERROR**  If error | This is CareVue and Metavision ICU databases that contains information on measurements as the a result of tests done on patients. It can include different details such as heart rate, blood pressure, weight and so on depending on their treatment.The type of test measured on a patient for ex., blood pressure has a unique item ID that can be linked with D\_ITEMS table for a definition on that particular Item. |
| **PATIENTS** | **SUBJECT\_ID**  Patient  **GENDER**  **DOB**  Date of Birth  **DOD**  Date of Death  **DOD\_HOSP**  DOD recorded at hospital  **DOD\_SSN**  DOD as in SSN DB  **EXPIRE\_FLAG**  Patient is dead/alive (binary) | This table can be used to know the lifetime details of a patient. Birthdate and Death date can be pulled from this table using the SUBJECT\_ID. If a patient expired at hospital, DOD\_HOSP will be recorded. If both DOD data are available, DOD\_HOSP is given high priority. |
| **PRESCRIPTIONS** | **SUBJECT\_ID**  Patient  **HADM\_ID**  Hospital admission ID  **ICUSTAY\_ID**  Patient ID at ICU  **DRUG\_TYPE**  Type of drug  **FORMULARY\_DRUG\_CD**  Drug code  **GSN**  Generic Seq number  **NDC**  National Drug code | It is the hospital provider order entry database. This table contains all the information on the drugs and medicine provided to the patient.  This table can be linked with both the inpatient and outpatients admitted in ICU using the composite key (SUBJECT\_ID,HADM\_ID,ICUSTAY\_ID). It includes the date the drug is prescribed, expiry date of the medicines, drug compositions, dosage details of prescribed drugs, drug type and so on. |
| **PROCEDUREEVENTS\_MV** | **SUBJECT\_ID**  Patient  **HADM\_ID**  Hospital admission ID  **ICUSTAY\_ID**  Patient ID at ICU  **STARTTIME**  **ENDTIME**  Procedure time  **ITEMID**  Measurement type  **CGID**  Caregiver ID  **CONTINUEINNEXTDEPT**  Approval to shift after procedure  **STATUSDESCRIPTION**  Patient’s status | It is the Metavision ICU database. This table contains information on procedures done to patients who are staying in ICU. The attributes (SUBJECT\_ID,HADM\_ID,ICUSTAY\_ID,ITEMID) works as composite key.  Each tuple is associated with a ORDERID that contains a unique identifier for the procedure performed for the patient. This table can be linked to INPUTEVENTS\_CV and INPUTEVENTS\_MV using the attributes ORDERID and LINKORDERID.  ITEMID can be used to track the particular item definition available in D\_ITEMS table  CONTINUEINNEXTDEPT is a binary variable which tells 0 or 1, meaning cannot move to next department and can move respectively. |
| **PROCEDURES\_ICD** | **SUBJECT\_ID**  Patient  **HADM\_ID**  Hospital admission ID  **SEQ\_NUM**  Procedures order  **ICD9\_CODE**  Procedure code | This table stores the ICD information for various procedures carried out for a patient. This is used by health insurers to classify medical procedures for billing purposes.  The attributes (SUBJECT\_ID,HADM\_ID) can be used as composite key to retrieve rows from this  table.  The definition of the ICD-9 code for the given procedure can be retrieved from D\_ICD\_PROCEDURES table which has ICD-9 code as the primary key. |
| **SERVICES** | **SUBJECT\_ID**  Patient  **HADM\_ID**  Hospital admission ID  **TRANSFERTIME**  Time transferred to other hospital/ department  **PREV\_SERVICE**  **CURR\_SERVICE**  Previous and current hospital/ department | This table will store the services the patient is receiving. The SERVICES table should be used if interested in identifying the type of service a patient is receiving in the hospital. For example, if interested in identifying surgical patients, the recommended method is searching for patients admitted under a surgical service.  The attributes (SUBJECT\_ID,HADM\_ID) can be used as composite key to retrieve rows from this table.  It will have information on the previous and current service the patient is receiving and also the time at which they are transferred from one to another. |
| **D\_ICD\_DIAGNOSES** | **ICD9\_CODE**  The international coding definitions version 9  **SHORT\_TITLE**  Short titles for diagnoses code  **LONG\_TITLE**  Long titles with more explanation of the diagnoses | This table is an entity table International Classification of Diseases Version 9 (ICD-9) codes for diagnoses. These codes are assigned at the end of the patient’s stay and are used by the hospital to bill for care provided. |
| **D\_ICD\_PROCEDURES** | **ICD9\_CODE**  International code for procedure | This table is an entity table International Classification of Diseases Version 9 (ICD-9) codes for procedures. These codes are assigned at the end of the patient’s stay and are used by the hospital to bill for care provided. They can further be used to identify if certain procedures have been performed |
| **D\_ITEMS** | **ITEMID**  Unique key for the each row  **ABBREVIATION**  Only use for Metavision and not CareVue  **LINKSTO**  provides the table name which the data links to | This table is an entity table is sourced from two *distinct* ICU databases. The main consequence is that there are duplicate ITEMID for each concept. |
| **D\_LABITEMS** | **LABEL**  The LABEL column describes the concept which is represented by the **ITEMID**  Fluid  **FLUID** describes the substance on which the measurement was made.  **Loinc\_code**  LOINC\_CODE contains the LOINC code associated with the given ITEMID. | This table is an entity table that contains definitions for all ITEMID associated with lab measurements in the MIMIC database. |
| **DATETIMEEVENTS** | **ICUSTAY\_ID**  ICUSTAY\_ID is unique to a patient ICU stay  **CHARTTIME**  records the time at which an observation was charted.  **STORETIME**  records the time at which an observation was manually input or manually validated by a member of the clinical staff.  **CGID**  is the identifier for the caregiver who validated the given measurement. | This table is an entity table contains all date measurements about a patient in the ICU. |
| **DIAGNOSES\_ICD** | **SEQ\_NUM**  Provides the order in which the ICD diagnoses related to the patient. | This table is an entity table Contains ICD diagnoses for patients, most notably ICD-9 diagnosis. |
| **DRGCODES** | **DRG\_type**  **DRG\_code**  **Description**  **DRG\_severity**  **DRG\_mortality** | This table contains the diagnostic codes and description for patients hospital visit. DRG\_code is used by the hospital for billing purposes. DRG stands for Diagnosis Related Group (DRG) |
| **MICROBIOLOGYEVENTS** | **SPEC\_itemID**  **SPEC\_type\_description**  **ORG\_itemID**  **ORG\_name**  **Chartdate**  **Charttime** | This tables stores microbiology measurements and sensitivities of patients from each hospital visit. |

***Part 3: Exploring major characteristics of the MIMIC III data***

***Exploratory Analysis on ICU Patients with Congestive Heart Failure***

Our major focus with this dataset characterization task was to analyze the population of patients who had been diagnosed with Heart Failure. Our approach was to use with exploratory data analysis on different aspects of the MIMIC-III database, looking at different aspects of the data such as gender, age, visits, diagnoses and procedures. Below we go into depth of each of the queries we pulled and the data in our results.

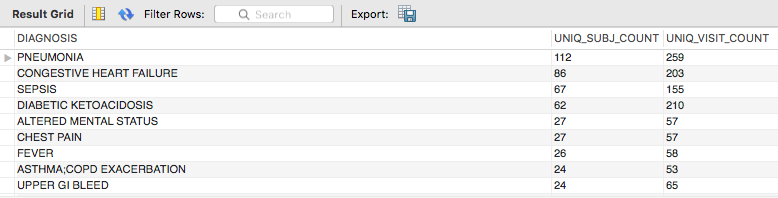
**Filter for Heart Failure Patients**:

(ADMISSIONS table) ‘DIAGNOSIS LIKE “%HEART%FAILURE%”’

**Queries**:

**Which diagnoses have seen more than one patient come back? (we consider diagnoses in which at least two patients have come back)** *[Refer to #1 in .sql]*

Here we analyzed all diagnoses with the number of unique subjects (patients) that experienced that diagnosis and the total number of unique visit counts. The subjects (patients) that are counted are only the patients who have come back (i.e. have more than one unique HADM\_ID). There are 73 diseases which have seen returning more than 1 returning patient in this dataset. The most prevalent of these is Pneumonia, which saw 112 unique patients come back at least once for re-admission. Since we are focusing on Heart Failure for this exploratory analysis, it is important to take note that it is the second ranked diagnosis among those that had at least two patients who each returned more than once.

**

**What is the gender distribution for ICU patients with Heart Failure?**

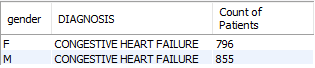
Ref--14.1 What is the gender distribution for patients with Heart Failure?

SELECT DISTINCT(DIAGNOSIS),gender, COUNT(\*) AS 'Count of Patients'

FROM patients JOIN admissions USING(SUBJECT\_ID)

WHERE diagnosis like '%Heart Failure%'

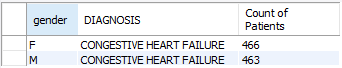
GROUP BY gender;

****

There are total of 1,651 patients with heart failure conditions. Out of this, 796 are females and 855 are males.

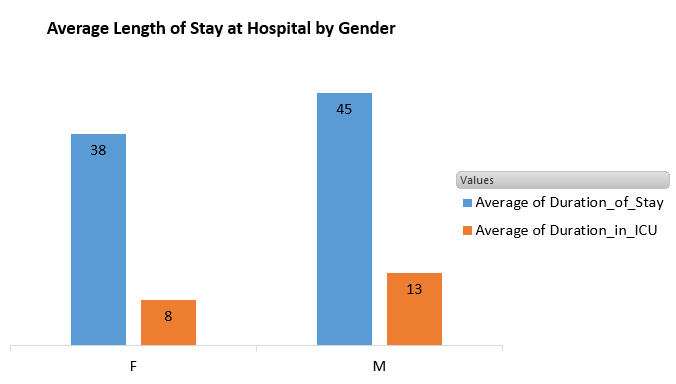
Out of the general number of people with heart failure, 929 had congestive heart failure (a form of heart failure).

-- 14.2 **Total count of patients with Congestive Heart Failure by gender?**

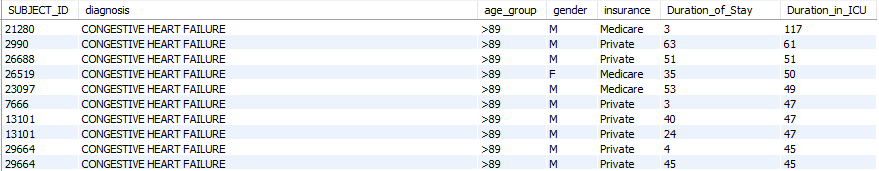


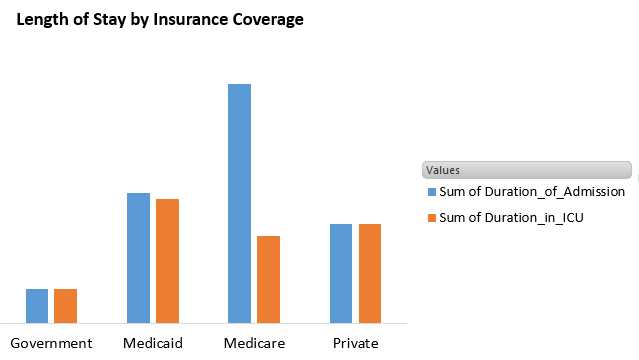
**Average number of days spent in ICU by patients of each gender with congestive heart failure?**

On average, males with congestive heart failure spent longer days in the admissions and ICU than females did. Females spent 7 fewer days in admissions and 5 fewer days in ICU than men did, which can mean faster recoveries for women.

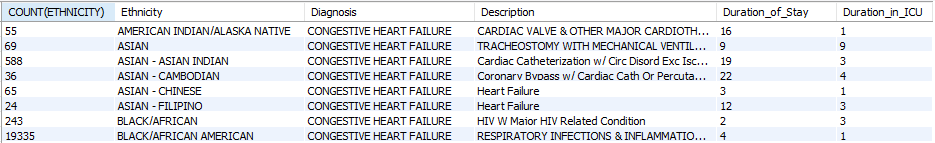


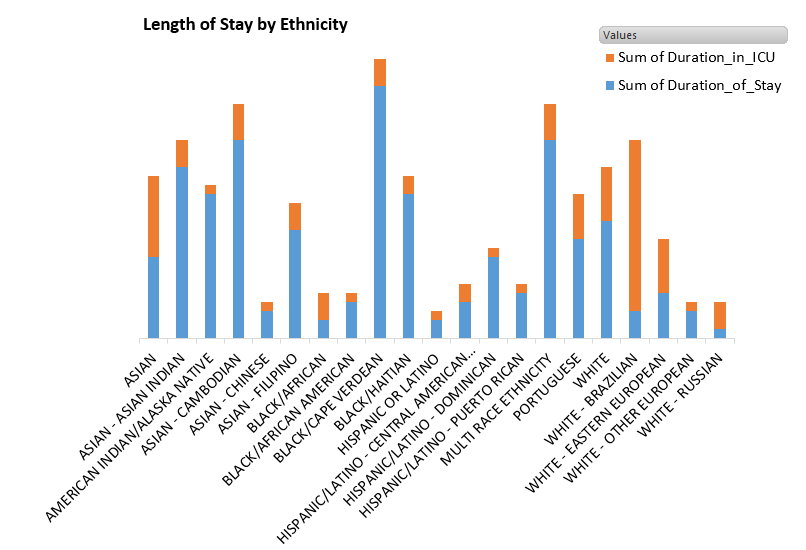
--Ref 14.3 **The Average Length of stay relationship with coverage of health care?**



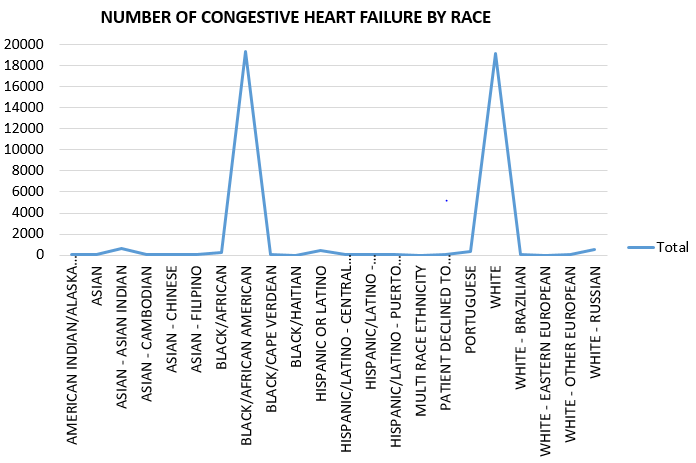


--Ref 14.4 **What is the ethnicity distribution for ICU patients with Heart Failure?**

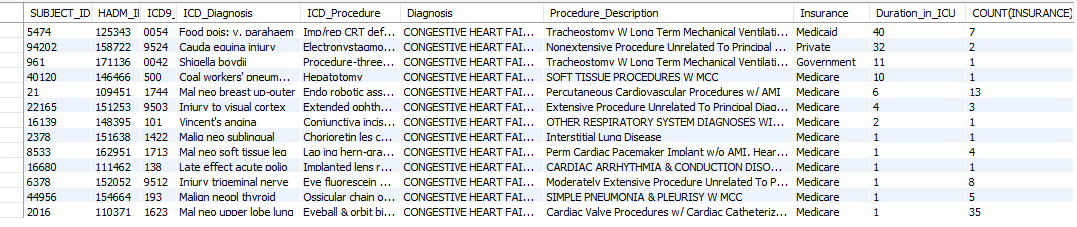




Blacks and Whites have substantially higher numbers of congestive heart failures (19,335 and 19,191 respectively). However, Blacks tend to spend longer periods in admissions and shorter periods in ICU. Whites, on the other hand tend to spends longer days in ICU and shorter days in admissions. ICU. What is causing this discrepancies? Is it lack of preventive care/ poverty/ early recovery rates? This is something that can be looked into more.



**#-14.6 General description of ICD Diagnosis, ICD\_Procedures, drugs prescription for diagnoses of patients admitted into ICU.**

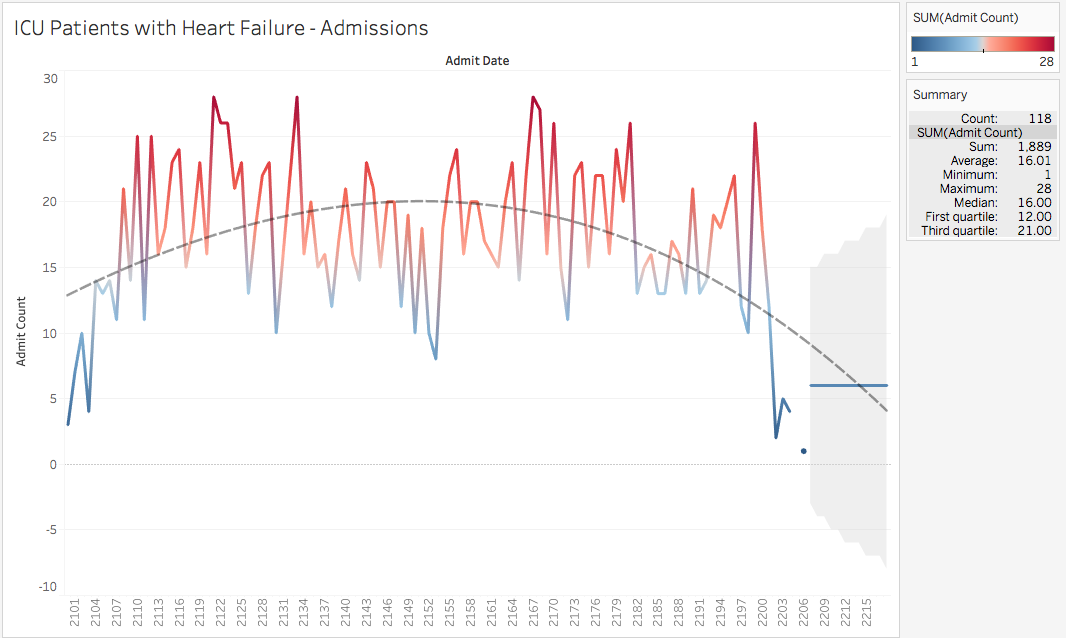


The medical procedure that will likely result in longer hospitalization in ICU is ‘Tracheostomy with long term mechanical ventilations’. Patients that undergo 'Cardiac Valve Procedures w/ Cardiac Catheterization' are among the class of procedures with fewer days of hospitalization in ICU.

**If we group the hospital admissions for ICU patients with Heart Failure by admit date, is there any sort of trend we can see of patients?** *[Refer to query #3 in .sql]*

Our analysis found that *there is a downward trend in admitted congestive heart failure cases over the span of the 106 years that we have admission data for*. Initially, the rate was increasing and stayed relatively steady for about 90 years at around 15-20 reported patients with heart failure per year. In the last 6 years we have collected data for, the trend is declining. Other stats include: The highest number of reported cases was 28, which happened a few years though not consecutively, with the lowest being 1 case also reported a few years. First quartile is 12 and third is 21. Average is 16.01.

The following tableau plot describes the trend of admissions for patients with heart failure. The far left are the admitted patients the beginning of when observations were recorded in the mimic database, and the far right shows the more recent admissions. The trendline is a polynomial function which charts the general trend showing that the patients increased over the years steadily but lately have been declining. Finally, the forecast shows that the estimate based on our polynomial trend line is that admit count is expected to be roughly around 6 per year in the next coming 10 years based on the latest information.

**

The following articles discuss how there is a general decrease in reports of heart failure in recent decades, though the cases may be going up as a whole because of the increase in population size. Our data does not capture population size but does show that the reported cases of Heart Failure are trending downward.

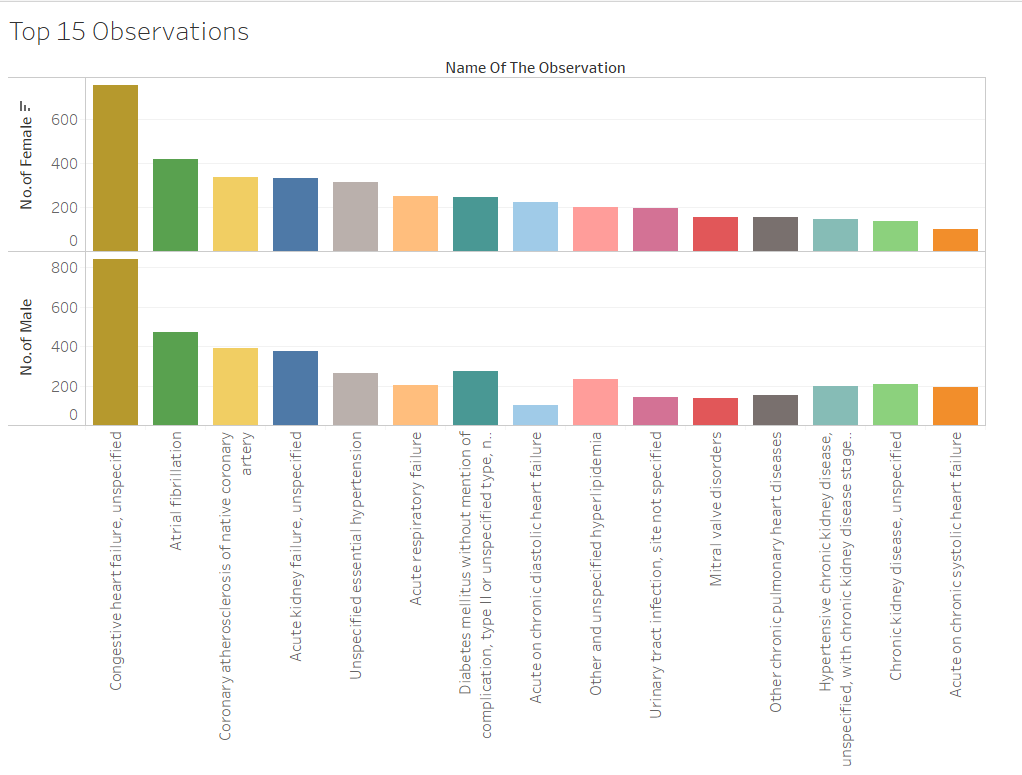
[*https://academic.oup.com/aje/article/178/8/1272/82398*](https://academic.oup.com/aje/article/178/8/1272/82398)

[*http://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(17)32520-5.pdf*](http://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(17)32520-5.pdf)

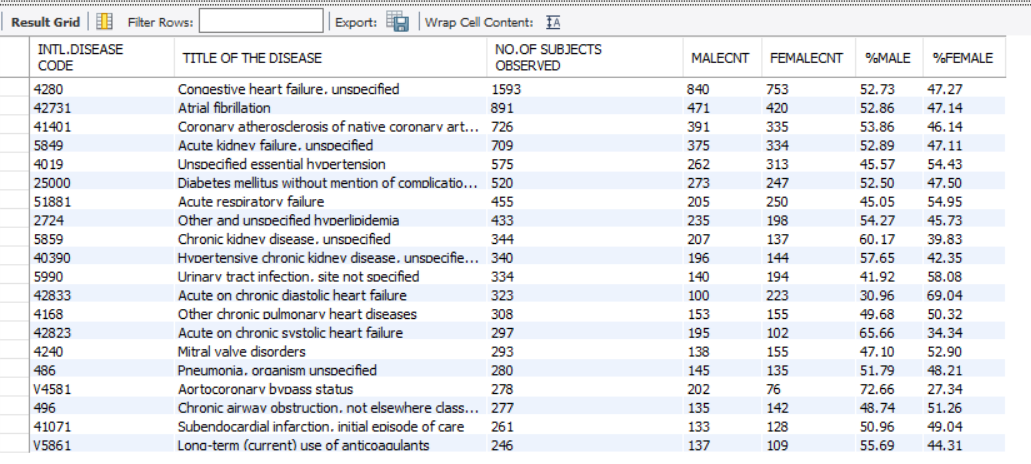
***“[...]from 2002 to 2014, heart failure incidence (standardised by age and sex) decreased, similarly for men and women, by 7% (from 358 to 332 per 100 000 person-years; adjusted incidence ratio 0·93, 95% CI 0·91–0·94). However, the estimated absolute number of individuals with newly diagnosed heart failure in the UK increased by 12% (from 170727 in 2002 to 190798 in 2014), largely due to an increase in population size and age”***

**What were the frequent observations on ICU patients with Heart Failure?**

The tableau chart shows the no.of subjects been reported for each observation with the gender distribution. *[Refer query #5A on ProjectReport.sql]*



The below is the results of mySQL query that returns top 20 observations.*[Refer query #5B on ProjectReport.sql]*

**

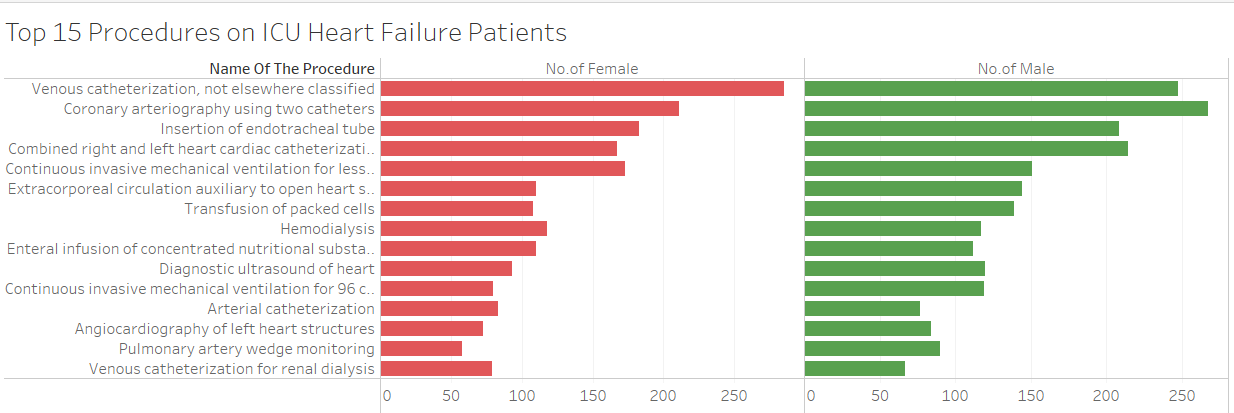
Findings:-

The query displays the frequent observations made on the heart failure patients and the percentage of male and female patients for each observation. Results indicate that the top observation is ‘Congestive heart failure. Unspecified’ in the MIMIC III database and no major difference in gender. The proportion of female patients is more for the observation ‘Acute on chronic diastolic heart failure’ and % of male patients show more of ‘Aortocoronary bypass status’.

**What was the procedures given distribution among ICU patients with Heart Failure?**

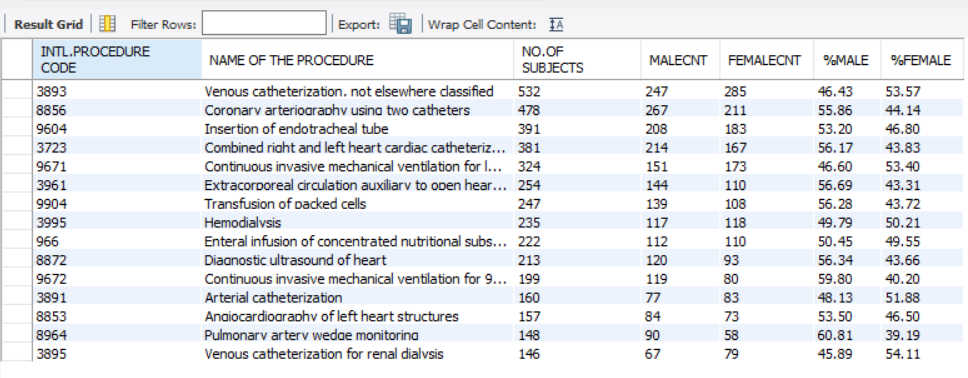
*[Refer query #6A on ProjectReport.sql]*

Below Tableau chart shows the frequent procedure made on patients in ICU with the gender classification.



The below is the results of mySQL query that returns top 15 procedures.

*[Refer query #6B on ProjectReport.sql]*



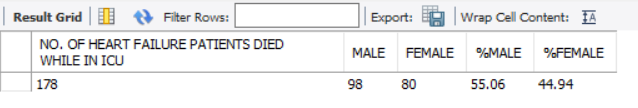
As per the results, ‘Venous catheterization’ being the common procedure carried out on patients as it is the most used method for administration of IV fluids. The second top is the Coronary arteriography using two catheters’ that is used to examine the coronary arteries.

There is no substantial gender difference on the procedures carried out on patients.

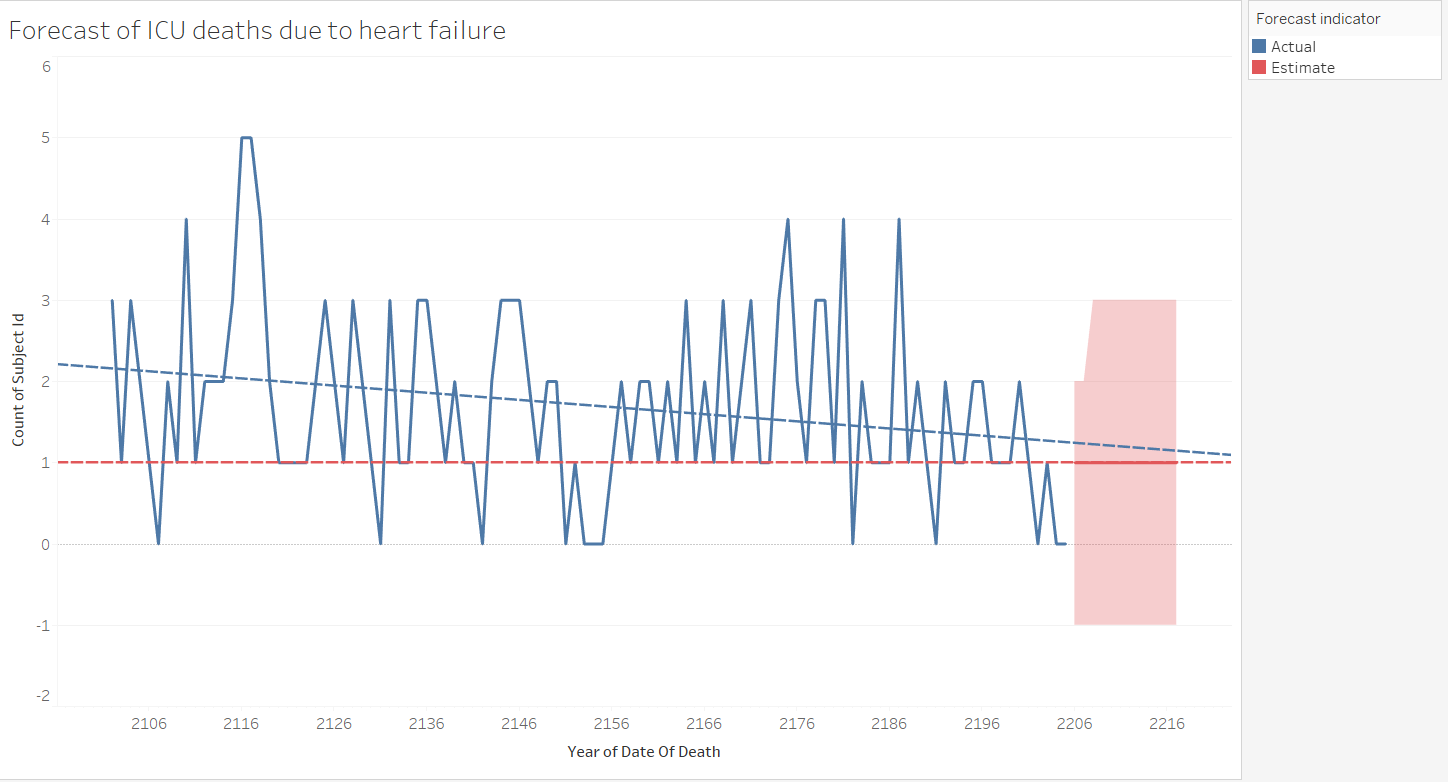
**What were the mortality rates among ICU patients with Heart Failure?**

To retrieve the mortality rate, queried the tables PATIENTS and ICUSTAYS and selected the subjects who have ICU out time same as Date of Death. *[Refer query #7A on ProjectReport.sql]*

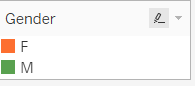
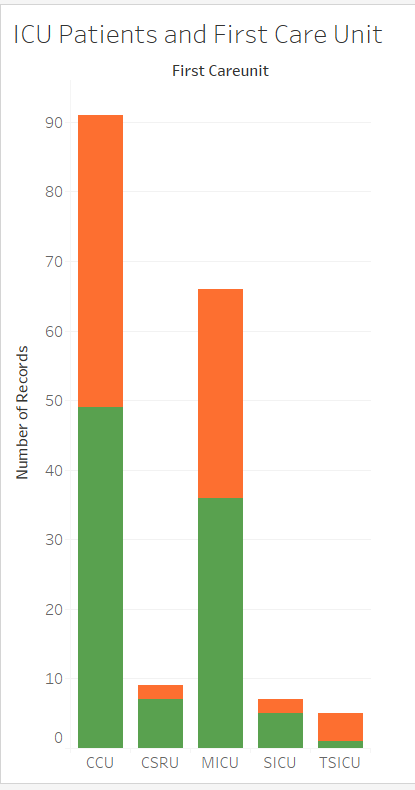
There are 178 patients in the MIMIC III database whose death date is recorded same as ICU out date. Below is the gender distribution of patients. *[Refer query #7B on ProjectReport.sql]*

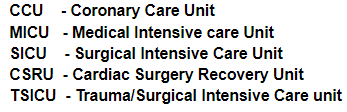


The below forecast from Tableau tells that the estimated count of deaths in future which is near to 1 based on the number reported in the past years.



Below is the chart that shows the number of patients who died in hospital with heart failure and their first care unit they admitted.

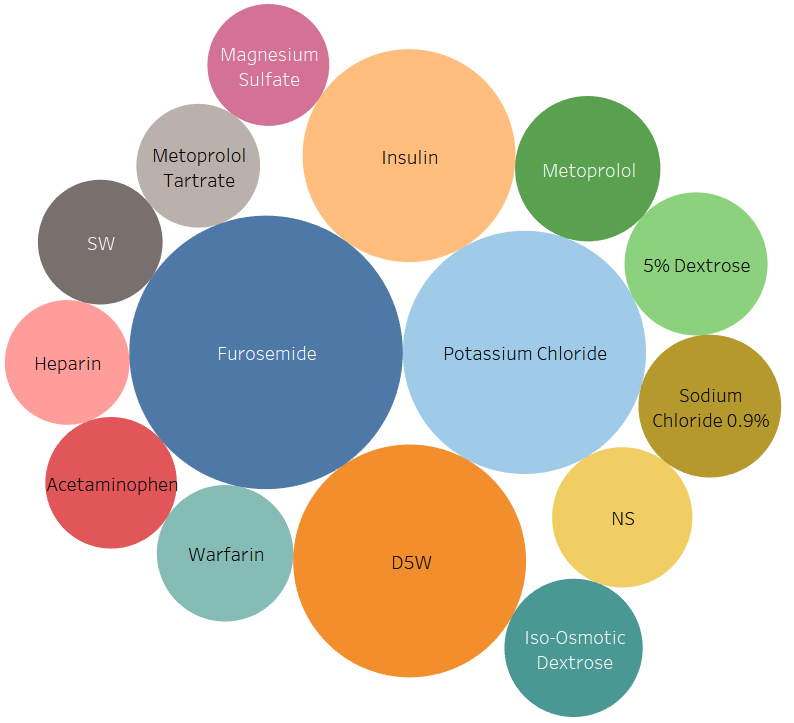




**What were the common drugs prescribed to ICU patients with Heart Failure?**

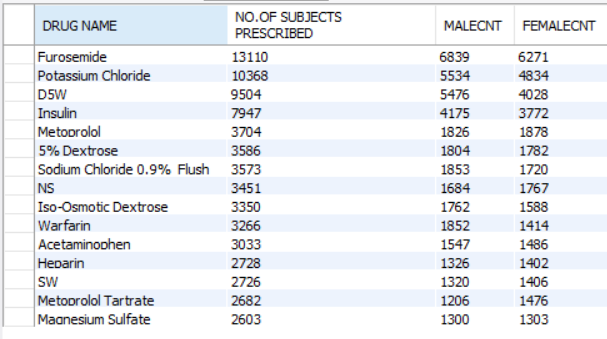
The top drug being ‘Furosemid’ is not surprising as it is used widely to treat people with edema caused by congestive heart failure. It is one of the famous diuretics used.

*[Refer query #13A on ProjectReport.sql]*

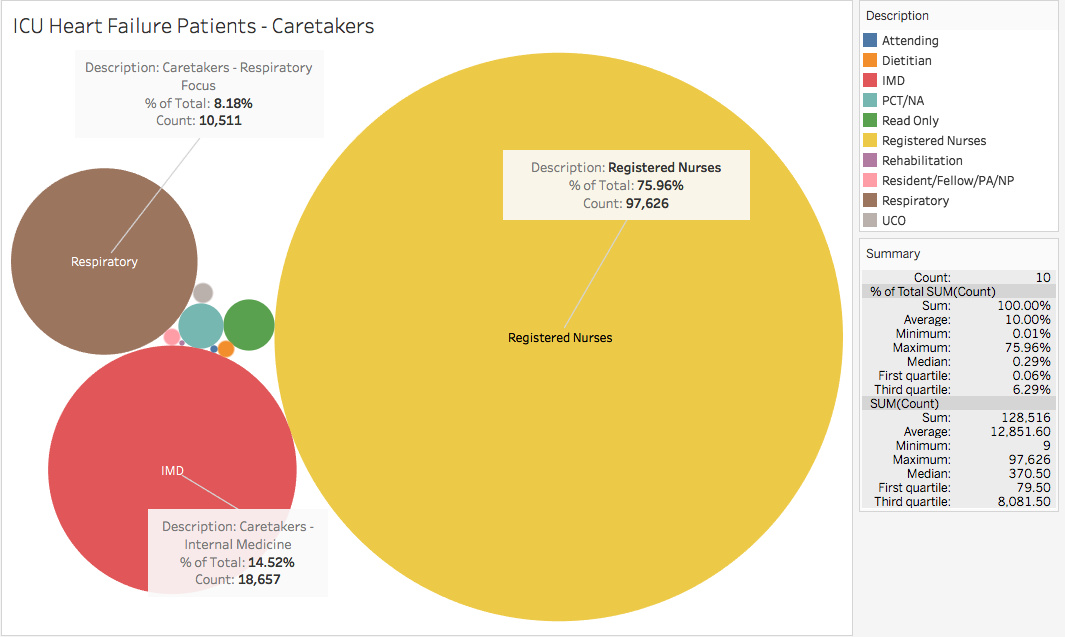


Below is the statistics from MySQL query of the top 15 drugs prescribed for heart failure patients.

*[Refer query #13B on ProjectReport.sql]*



**What kind of caregivers did ICU patients with Heart Failure interact with?** *[Refer to #9 in .sql]*

Not surprisingly, Registered Nurses were the most common in aiding patients with Congestive Heart Failure. They assumed close to 76% of the caretakers assigned to patients with Congestive Heart Failure with a count of roughly 97 thousand. Second were IMD caretakers associated with internal medicine. They assumed close to 15% of the caretakers assigned to these same patients. Statistics are shown along with the visualization - some notables are that the lowest was 9 - specifically ‘Rehabilitation’ nurses. Second to last place was the ‘Attending’ caretakers. The total number of caretakers assigned to patients with heart failure was 128,516.**

**Was the number of patients who went into ICU for Heart Failure higher than the number that were Non-ICU?** *[Refer to #10 in .sql]*

The results of this query were not surprising at all. This MIMIC database, is after all, a database containing information about patients who had intensive care visits. The one notable thing returned for this query is that for HADM\_ID IN (146310,163579,114310), there is no record of that admission being an ICU admission (no record in the ICUSTAYS table). This could just be a data error.

**Was there any report of Heart Failure in neonates (patients <1 year old)?**

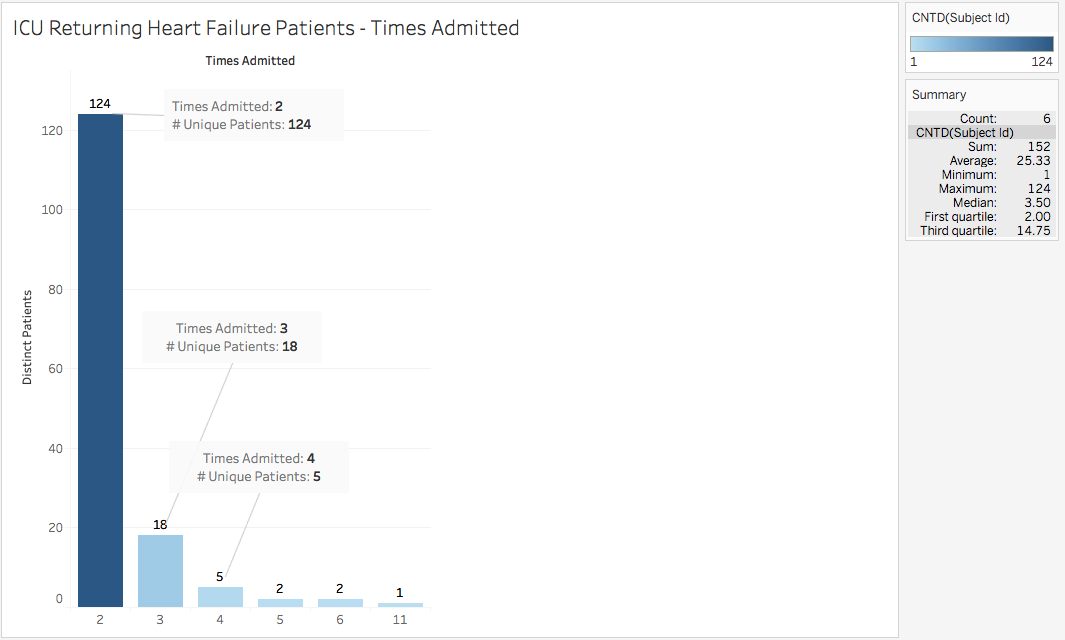
No neonates returned as a result. This shows that this dataset only includes Heart Failure Patients that are adults.

*[Refer query #8 on ProjectReport.sql]*

**What were the number of returning ICU patients with Heart Failure?** *[Refer to #11 in .sql]*

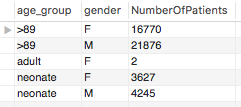
Very insightful information found when analyzing returning patients. Of the returning patients examined (total of 152 patients), 124 returned just once after their initial ICU visit. After that, there was a steep drop off where only 18 unique patients returned three times. We can infer that after a traumatic heart event, the 124 patients came back for possibly a complication that occured after their first admission. Patients who came back 3 or more times could have had a rare complication or other issues after their first visit.

The following visualization shows the number of unique subjects that came back more than once. As mentioned earlier, most patients came back only once, followed by a steep drop off in returning patients after that. The max number of visits was 11 (only one unique subject came back that many times).



**How many subjects were in each age group?** *[Refer to #13 in .sql]*

Our goal with this analysis was to figure out what our dataset looked like in terms patients of each age group and gender. In order to come up with our results, we had to look at the date each patient was first admitted into the hospital and take a calculation to determine what age group they fit in. Because of HIPAA, we are unable to pull the exact age for these patients. Below is the result of our findings. Most patients were over the age of 89 and male. There were only 2 ‘adults’ in our database and we also had more male neonates than female neonates.



**After how long from admitting to hospital the patient is put on ICU?**

*SELECT FLOOR(HOUR(i.intime - a.admittime) / 24) as time\_admit\_icu , a.subject\_id, a.hadm\_id, a.admittime, i.intime, a.dischtime, i.outtime*

*FROM admissions a INNER JOIN icustays i ON a.hadm\_id = i.hadm\_id*

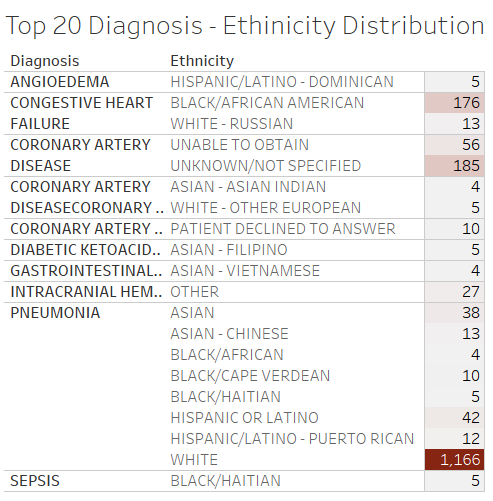
*WHERE FLOOR(HOUR(i.intime - a.admittime) / 24) IS NOT NULL AND FLOOR(HOUR(i.intime - a.admittime) / 24) <> 0*

*ORDER BY time\_admit\_icu asc;*

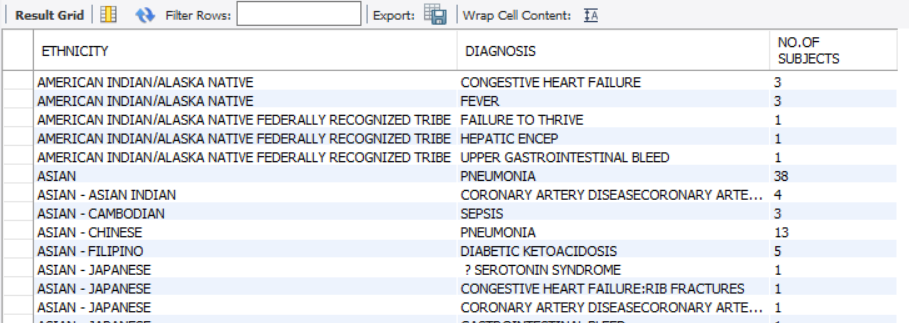
**The highest no.of subjects identified for diagnosis for each ethnicity**

This analysis is to classify if which diagnosis is prevalent in specific to a ethnicity. The top of the list is the White people diagnosed with Pneumonia. Black/African American people are high enough to be diagnosed next with heart failures.

Table below shows the top 20 ethnicities with highest number of subjects reported for each diagnosis

****

MySQL returning the highest number of subjects reported with a particular diagnosis for each ethnicity

**

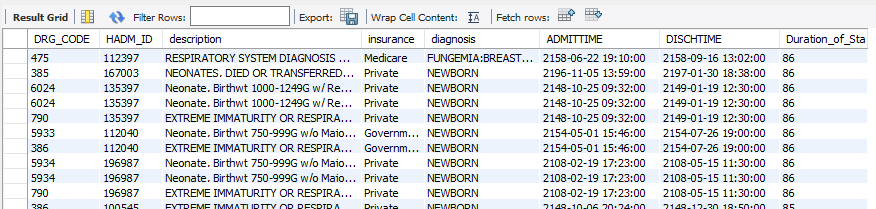
**Length of stay(admissions) at hospital based on types of insurance coverage. If we can find the average number of day covered by each insurance and one can better prepare for medical expenses in the future.**

*SELECT DC.DRG\_CODE,A.HADM\_ID, description,insurance,diagnosis,*

*ADMITTIME, DISCHTIME, datediff( DISCHTIME,ADMITTIME) AS 'Duration\_of\_Stay'*

*FROM admissions A JOIN drgcodes DC ON DC.HADM\_ID = A.HADM\_ID*

*ORDER BY Duration\_of\_stay DESC;*

****

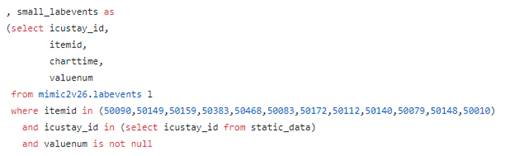
***Part 4: Preparing MIMIC data for analytical studies***

**Chapter 21**

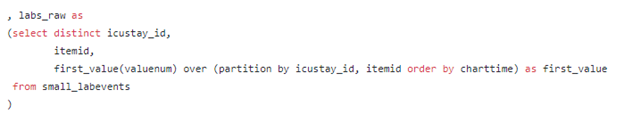
* In the first part of the query they want to find out the age of the people admit to ICU, the syntax “case” checks each row to see if the conditional statement is true, they wanted to know the age of adults, since the age of the patients are mask they made this statement for every age over 150 display 91.4. They use another conditional statement to find out if there is any hospital\_expire\_tag which means people that die, and they want to know less than 30 days, they also exclude every null value that contains information about ICU stays.



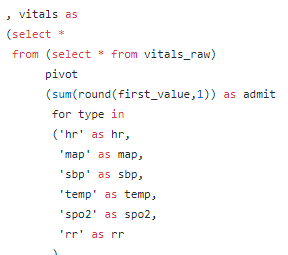
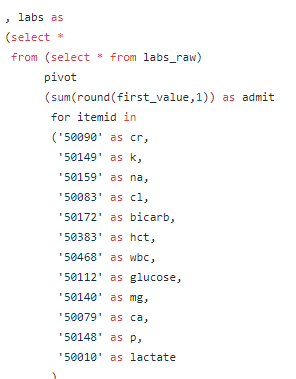
* In this query they created a table as Small\_labevents to find out of the ICU patients the measurements, the time in which the observation was made, and the physiologic scores by a certain itemid.



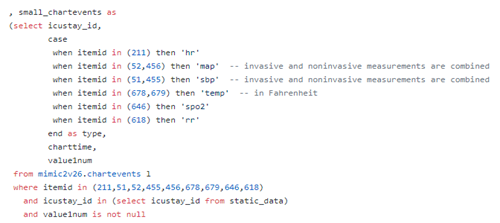
* In this query a lab\_raw table was created to divided two big table into one using the partition syntax.



* There are two table create as a sum of pivot table as lab results and each type of vitals.



* This query is also a conditional statement, to change some item ids into abbreviation definitions related to the number such as ‘rr’ = is the abbreviation to blood pressure, 'spo2'= the amount the oxygen in the blood.
* This query join vitals,labs,icustay tables together to transform text value into numeric value so it can be export to different software’s to do analysis.



There are some tables and columns in that MIMIC II is not the same in MIMIC III, in MIMIC II there is a column for age below there is a query showing how we would find the age in MIMIC III, we created a table and separated which gender into age group so it makes easier when building our queries;

*CREATE TABLE AAA\_FIRST\_ADMISSION\_TIME AS*

*SELECT*

*p.subject\_id, p.dob, p.gender, MIN(a.admittime) AS first\_admittime,MIN( ROUND( (cast(admittime as date) - cast(dob as date)) / 365.242,2) ) AS first\_admit\_age*

*FROM patients p*

*INNER JOIN admissions a*

*ON p.subject\_id = a.subject\_id*

*GROUP BY p.subject\_id, p.dob, p.gender*

*ORDER BY p.subject\_id;*

*CREATE TABLE AAA\_AGE AS*

*SELECT*

*subject\_id, dob, gender, first\_admittime, first\_admit\_age*

*CASE*

*-- all ages > 89 in the database were replaced with 300*

*-- we check using > 100 as a conservative threshold to ensure we capture all these patients*

*WHEN first\_admit\_age > 100 THEN '>89'*

*WHEN first\_admit\_age >= 14 THEN 'adult'*

*WHEN first\_admit\_age <= 1 THEN 'neonate'*

*ELSE 'middle'*

*END AS age\_group*

*FROM AAA\_FIRST\_ADMISSION\_TIME;*

*-- selecting*

*select age\_group, gender, count(subject\_id) as NumberOfPatients*

*from AAA\_AGE*

*group by age\_group, gender;*

* In addition, in MIMIC III there is not a table only for demographics the demographics for MIMIC III is into admissions and patients table. The admissions table contain information the whole hospital database, unlike the demographics on MIMIC II only contain information regarding patients staying in the ICU. In the query below will show how would you create a table for MIMIC III to extract anything specific to demographics;

*Create table Demographics as*

*Select language, religion, marital\_status, ethnicity, gender, DOB*

*From admissions a join patients p on a.subject\_id = p.subject\_id*

*Order by p.subject\_id*

**Chapter 23**

* Chapter 23 explains more on the propensity score analysis that will be used for statistical modeling on EHR based research.
* One case study on MIMIC II database used propensity score analysis in determining binary outcome based on several dependant factors and continuous outcome in determining the quantitative output.
* The queries are used to come up with results of adult patients admitted into ICU for the first time and diagnosed with Aﬁb with RVR. It also includes medical procedures and drugs given to patients based on the efficacy of treatment in a period. In the article, it used propensity score which is the statistical analysis of observational data. Propensity score matching (PSM), a statistical matching technique that attempts to estimate the effect of a treatment, policy, or other intervention by accounting for the covariates that predict the procedures patients receive.
* The queries for this chapter, extract the admission information on patients such as time of admission into ICU and discharge information from the hospital and ICU. The article states that waveform data allows easy and in-depth analysis of patients admitted for the first time to ICU. This data contains unstructured data such as nurses comments, and it's a small fraction of the database that allows a researcher to take full advantage of the information.

***Part 5 - Executive summary Findings and recommendations***

*Based on the data from MIMIC III we were able to find some insights on the hospital and ICU records in the hospital. For our data extraction we used Mysql and for visualization analysis we used tableau and excel . Our focus with this dataset characterization task was to analyze the population of patients who had been diagnosed with Congestive Heart Failure.Our approach was to use with exploratory data analysis on different aspects of the MIMIC-III database, looking at different aspects of the data such as Gender, Age, Visits, Procedures.*

*We chose this topic based on the MIMIC III data and conducted further research on heart failure. Base on the data the two diagnoses diseases with the highest number of patients was pneumonia and congestive heart failure. Research shows that “ heart failure affects nearly 6 million Americans. Roughly 670,000 people are diagnosed with heart failure each year”; with that we wanted to know more about people that was diagnosed with this disease and what are the findings we can make.*

*Our analysis on heart failure was necessary to gain insights into the demographics of patients and their diagnoses. Our study found out that patients with heart failure are 50.32% females and 49.68% males in the dataset. African- Americans and Whites had the most significant number of patients with heart failure. However, Blacks tend to spend more extended periods in admissions and shorter periods in ICU. On the other hand, Whites tend to spend longer days in ICU and shorter days in hospitalization.*

*Heart failure is a big problem for Americans, especially because studies shows that High blood pressure, high cholesterol, and smoking are key risk factors for heart disease. About half of Americans (47%) have at least one of these three risk factors. Per our findings in the MIMIC 3 database, it shows that the in the last six years the number of patients diagnosed with heart failure is decreasing. It is safe to say that this decrease is a result of healthy lifestyle choices by individuals and early preventive care for patients.*

***Roles and Contributions***

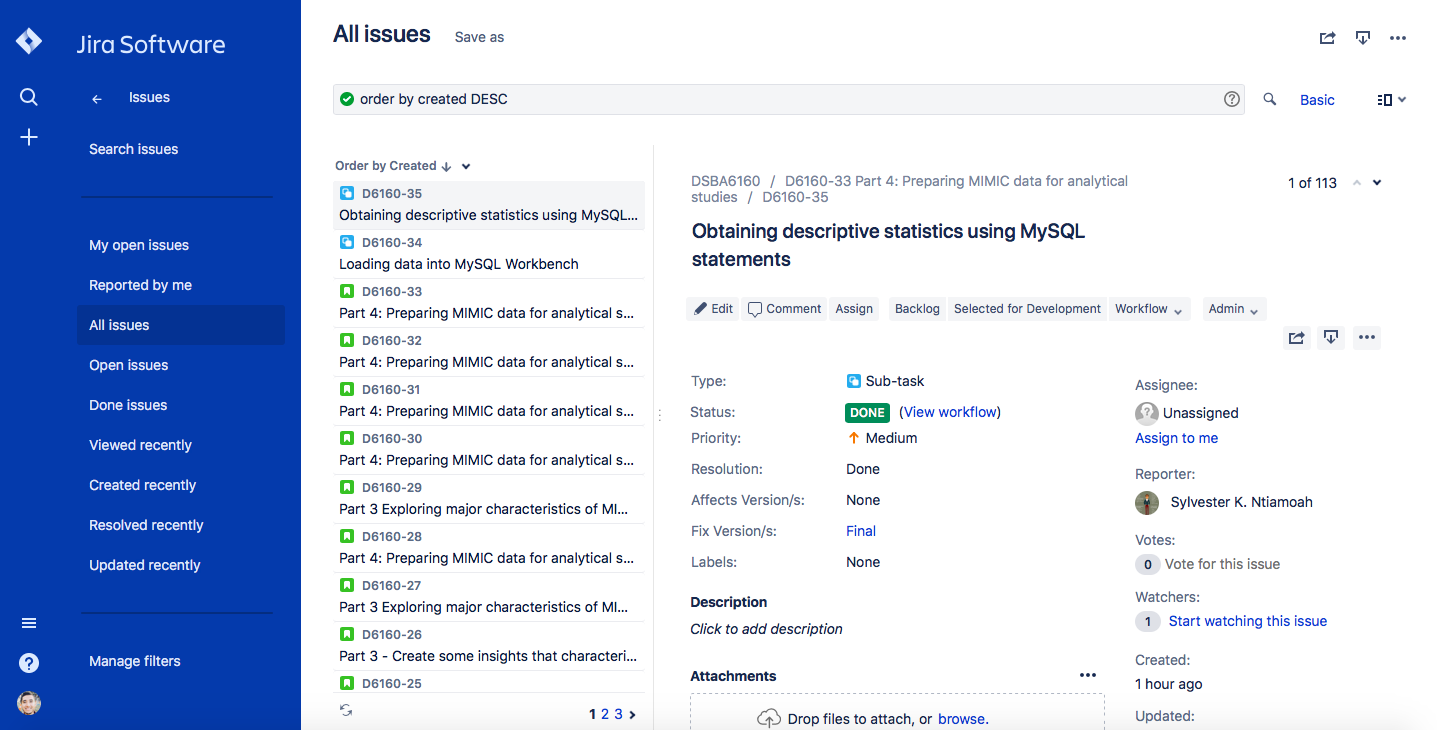
*Jose - My contribution was generating some of the queries that we used in analyzing heart failure patients. Along with my query analysis, I generated some tableau visualizations and gathered some statistics on the query results using the same software. I helped set up the JIRA project to track our progress and contributed to the slides.*

*Lavanya - I contributed to the project by writing MySQL queries that gives support for explaining several characteristics of patients admitted with Congestive Heart Failure in MIMIC III database. Exported the query results and created visualizations in tableau and took up the role as both developer and analyst.*

*Sylvester - I did some exploratory work on the data using MySQL. In addition, I transferred the results of queries into excel for further analysis and visualization of the some characteristics of the data. I did the final editing of the data to make sure it's clean and free of error.*

*Irene - Irene wrote the executive summary for the report and explained the requirements specified in the article "Secondary Analysis of Electronic Health Records” to give the reader a better understanding of the data.*

*As a team, we tracked our progress and efforts over WhatsApp and also using JIRA. We tracked the step by step process of our work on JIRA. Below are a screenshot and report of the stories and tasks that were completed over the course of this group project.*



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Issue Type | Issue key | Summary | Assignee | Status | Updated |
| Sub-task | D6160-35 | Obtaining descriptive statistics using MySQL statements | sntiamoa | Done | 4/22/18 21:31 |
| Sub-task | D6160-34 | Loading data into MySQL Workbench | sntiamoa | Done | 4/22/18 21:31 |
| Story | D6160-33 | Part 4: Preparing MIMIC data for analytical studies | sntiamoa | Done | 4/22/18 21:31 |
| Story | D6160-32 | Part 4: Preparing MIMIC data for analytical studies | Jose | Done | 4/22/18 21:31 |
| Story | D6160-30 | Part 4: Preparing MIMIC data for analytical studies | iojukwu | Done | 4/22/18 21:31 |
| Story | D6160-29 | Part 3 Exploring major characteristics of MIMIC critics data |  | Done | 4/22/18 21:31 |
| Story | D6160-28 | Part 4: Preparing MIMIC data for analytical studies | reachlavanyaa | Done | 4/22/18 21:31 |
| Story | D6160-27 | Part 3 Exploring major characteristics of MIMIC critics data | reachlavanyaa | Done | 4/22/18 21:31 |
| Story | D6160-26 | Part 3 - Create some insights that characterize the data - Jose | Jose | Done | 4/22/18 21:31 |
| Story | D6160-25 | Document tables in google doc - Lavanya | reachlavanyaa | Done | 4/22/18 21:31 |
| Story | D6160-24 | Document experience loading the database - Lavanya | reachlavanyaa | Done | 4/22/18 21:31 |
| Story | D6160-23 | Document experience loading the database - Irene | iojukwu | Done | 4/22/18 21:31 |
| Story | D6160-22 | Document experience loading the database - Sly | sntiamoa | Done | 4/22/18 21:31 |
| Story | D6160-21 | Document tables in google doc - Sly | sntiamoa | Done | 4/22/18 21:31 |
| Story | D6160-20 | Document tables in google doc - Irene | iojukwu | Done | 4/22/18 21:31 |
| Story | D6160-19 | Document tables in google doc - Jose | Jose | Done | 4/1/18 22:35 |
| Story | D6160-18 | Create Database and load MIMIC III dataset in MySQL | reachlavanyaa | Done | 4/22/18 21:31 |
| Story | D6160-17 | Document process in loading the scripts for MIMIC III Database | Jose | Done | 4/1/18 22:35 |
| Sub-task | D6160-16 | Upload Completion Report | sntiamoa | Done | 3/20/18 20:24 |
| Sub-task | D6160-15 | Complete Data or Specimen Only Research Course | sntiamoa | Done | 3/20/18 20:24 |
| Sub-task | D6160-14 | Enrolled in Data or Specimen Only Research Course | sntiamoa | Done | 3/20/18 20:24 |
| Sub-task | D6160-13 | Created CITI account | sntiamoa | Done | 3/20/18 20:24 |
| Sub-task | D6160-9 | Upload completion report | Jose | Done | 3/20/18 20:24 |
| Sub-task | D6160-8 | Complete Data or Specimens only course | Jose | Done | 3/20/18 20:24 |
| Sub-task | D6160-7 | Enrolled in the Data or Specimens only course | Jose | Done | 3/20/18 20:24 |
| Sub-task | D6160-6 | Created CITI account | Jose | Done | 3/20/18 20:24 |
| Story | D6160-5 | Load MIMIC III data into MySQL and document into a report | Jose | Done | 4/1/18 22:35 |
| Story | D6160-4 | Complete CITI Training and upload Completion Report to Canvas | reachlavanyaa | Done | 3/20/18 20:24 |
| Story | D6160-3 | Complete CITI Training and upload Completion Report to Canvas | iojukwu | Done | 4/1/18 22:35 |
| Story | D6160-2 | Complete CITI Training and upload Completion Report to Canvas | sntiamoa | Done | 3/20/18 20:24 |
| Story | D6160-1 | Complete CITI Training and upload Completion Report to Canvas | Jose | Done | 3/20/18 20:24 |