***MYSQL***

***EXPLORATORY DATA ANALYSIS***

***On***

***Hospitals DataSet***

***By***

***Aman Rajbhar***

***Data Cleaning***

*DATABASE NAME------USE hospital\_db;*

***Renaming/Atering The Table Column Names***

*ALTER TABLE facility*

*RENAME COLUMN `Facility.Name` TO Name,*

*RENAME COLUMN `Facility.City` TO City,*

*RENAME COLUMN `Facility.State` TO State,*

*RENAME COLUMN `Facility.Type` TO Type;*

*ALTER TABLE procedures*

*RENAME COLUMN `Procedure.Heart Attack.Cost` TO Heart\_Attack\_Cost,*

*RENAME COLUMN `Procedure.Heart Attack.Quality` TO Heart\_Attack\_Quality,*

*RENAME COLUMN `Procedure.Heart Attack.Value` TO Heart\_Attack\_Value,*

*RENAME COLUMN `Procedure.Heart Failure.Cost` TO Heart\_Failure\_Cost,*

*RENAME COLUMN `Procedure.Heart Failure.Quality` TO Heart\_Failure\_Quality,*

*RENAME COLUMN `Procedure.Heart Failure.Value` TO Heart\_Failure\_Value,*

*RENAME COLUMN `Procedure.Hip Knee.Cost` TO Hip\_Knee\_Cost,*

*RENAME COLUMN `Procedure.Hip Knee.Quality` TO Hip\_Knee\_Quality,*

*RENAME COLUMN `Procedure.Hip Knee.Value` TO Hip\_Knee\_Value,*

*RENAME COLUMN `Procedure.Pneumonia.Cost` TO Pneumonia\_Cost,*

*RENAME COLUMN `Procedure.Pneumonia.Quality` TO Pneumonia\_Quality,*

*RENAME COLUMN `Procedure.Pneumonia.Value` TO Pneumonia\_Value;*

*ALTER TABLE rating*

*RENAME COLUMN `Rating.Effectiveness` TO Effectiveness,*

*RENAME COLUMN `Rating.Experience` TO Experience,*

*RENAME COLUMN `Rating.Imaging` TO Imaging,*

*RENAME COLUMN `Rating.Mortality` TO Mortality,*

*RENAME COLUMN `Rating.Overall` TO Overall,*

*RENAME COLUMN `Rating.Readmission` TO Readmission,*

*RENAME COLUMN `Rating.Safety` TO Safety,*

*RENAME COLUMN `Rating.Timeliness` TO Timeliness;*

***# Update Invalid Ratings (e.g., -1)***

*UPDATE rating*

*SET*

*rating.Overall = NULL*

*WHERE*

*rating.Overall = - 1;*

***#Remove Rows with Zero Procedure Costs***

*DELETE FROM procedures*

*WHERE*

*procedures.Heart\_Attack\_Cost = 0*

*OR procedures.Heart\_Failure\_Cost = 0*

*OR procedures.Hip\_Knee\_Cost = 0*

*OR procedures.Pneumonia\_Cost = 0;*

***#Remove Rows with Zero Ratings***

*DELETE FROM rating*

*WHERE*

*Overall = 0*

*OR Mortality = 0*

*OR Safety = 0*

*OR Readmission = 0*

*OR Experience = 0*

*OR Effectiveness = 0*

*OR Timeliness = 0*

*OR Imaging = 0;*

*The ratings column that contains values such as 'above', 'below', 'same', and 'none',*

*i have* ***coverted them into ordinal values*** *like 3 for 'above', 2 for 'same', 1 for 'below', and 0 for 'none'*

*foe example:-*

*UPDATE rating*

*SET*

*rating.Mortality = 0*

*WHERE*

*rating.Mortality = 'None';*

*------*

*UPDATE rating*

*SET*

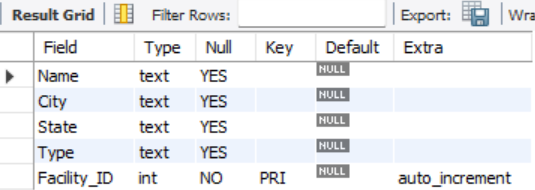
*rating.Mortality = 3*

*WHERE*

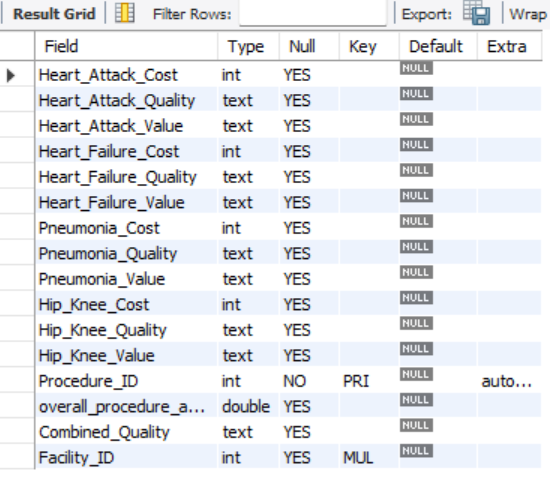
*rating.Mortality = 'above';*

***Data Exploration***

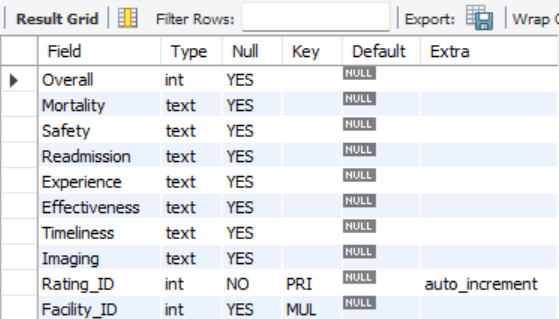
**DESCRIBE facility;**



**DESCRIBE procedures;**



**DESCRIBE rating;**



These queries provide the structure of the facility, procedures, and rating tables, listing columns, their data types, and constraints.

***#Count Total Rows and Distinct Facilities***

*SELECT*

*COUNT(\*) AS Total\_Rows,*

*COUNT(DISTINCT facility.Name) AS Distinct\_Facility*

*FROM*

*facility;*



This query counts the total number of rows in the facility table and the number of distinct facilities by name.

***#Check Distribution of Facility Types***

*SELECT*

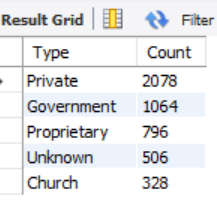
*facility.Type, COUNT(\*) AS Count*

*FROM*

*facility*

*GROUP BY facility.Type*

*ORDER BY count DESC;*



This query groups the facility table by Type and counts the number of facilities for each type, ordered from most to least frequent.

***# Summary Statistics for All Procedure Costs***

*SELECT*

*ROUND(AVG(procedures.Heart\_Attack\_Cost)) AS avg\_heart\_attack\_cost,*

*MAX(procedures.Heart\_Attack\_Cost) AS max\_heart\_attack\_cost,*

*MIN(procedures.Heart\_Attack\_Cost) AS min\_heart\_attack\_cost,*

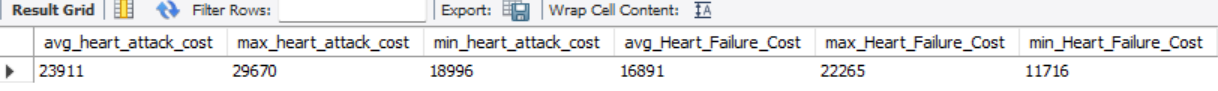
*ROUND(AVG(procedures.Heart\_Failure\_Cost)) AS avg\_Heart\_Failure\_Cost,*

*MAX(procedures.Heart\_Failure\_Cost) AS max\_Heart\_Failure\_Cost,*

*MIN(procedures.Heart\_Failure\_Cost) AS min\_Heart\_Failure\_Cost*

*FROM*

*procedures;*



-----

*SELECT*

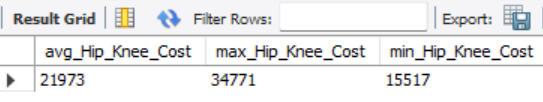
*ROUND(AVG(procedures.Hip\_Knee\_Cost)) AS avg\_Hip\_Knee\_Cost,*

*MAX(procedures.Hip\_Knee\_Cost) AS max\_Hip\_Knee\_Cost,*

*MIN(procedures.Hip\_Knee\_Cost) AS min\_Hip\_Knee\_Cost*

*FROM*

*procedures;*



------

*SELECT*

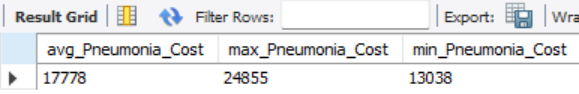
*ROUND(AVG(procedures.Pneumonia\_Cost)) AS avg\_Pneumonia\_Cost,*

*MAX(procedures.Pneumonia\_Cost) AS max\_Pneumonia\_Cost,*

*MIN(procedures.Pneumonia\_Cost) AS min\_Pneumonia\_Cost*

*FROM*

Procedures;



This query calculates the average, maximum, and minimum costs for heart attack,heart failure procedures,Hip Knee & Pneumonia.

***#Count of Ratings***

*SELECT*

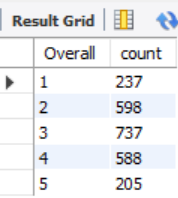
*Overall, COUNT(\*) AS count*

*FROM*

*rating*

*GROUP BY Overall*

*ORDER BY Overall;*



***Data Integration***

***#Create a New Column for Average Procedure Cost***

*alter table procedures*

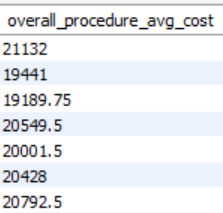
*add column overall\_procedure\_avg\_cost real;*

*UPDATE procedures*

*SET*

*overall\_procedure\_avg\_cost = ((procedures.Heart\_Attack\_Cost + procedures.Heart\_Failure\_Cost*

*+ procedures.Hip\_Knee\_Cost + procedures.Pneumonia\_Cost) / 4);*



***# Combine Quality Ratings into a Single Column***

*ALTER TABLE procedures*

*ADD COLUMN Combined\_Quality TEXT;*

*UPDATE procedures*

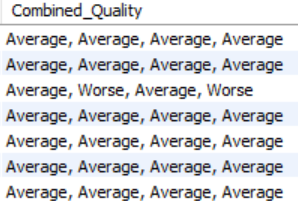
*SET Combined\_Quality = concat(*

*procedures.Heart\_Attack\_Quality,', ',*

*procedures.Heart\_Failure\_Quality,', ',*

*procedures.Hip\_Knee\_Quality,', ',*

*procedures.Pneumonia\_Quality);*

**

***Statistical Analysis***

***#Correlation Between Overall Rating and Procedure Costs***

*SELECT*

*rating.Overall,*

*AVG(procedures.Heart\_Attack\_Cost) AS avg\_Heart\_Attack\_Cost,*

*AVG(procedures.Heart\_Failure\_Cost) AS avg\_Heart\_Failure\_Cost,*

*AVG(procedures.Hip\_Knee\_Cost) AS avg\_Hip\_Knee\_Cost,*

*AVG(procedures.Pneumonia\_Cost) AS avg\_Pneumonia\_Cost*

*FROM*

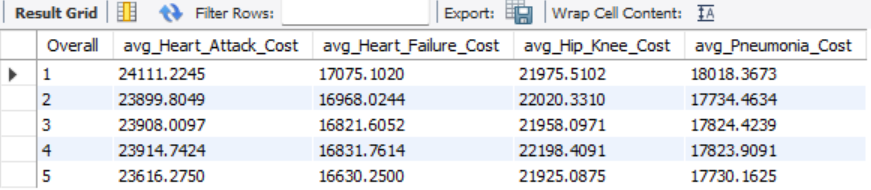
*procedures*

*JOIN*

*rating ON procedures.Facility\_ID = rating.Facility\_ID*

*GROUP BY rating.Overall*

*ORDER BY rating.Overall;*



***#Distribution of Ratings by Facility Type***

*SELECT*

*facility.Type, rating.Overall, COUNT(\*) AS count*

*FROM*

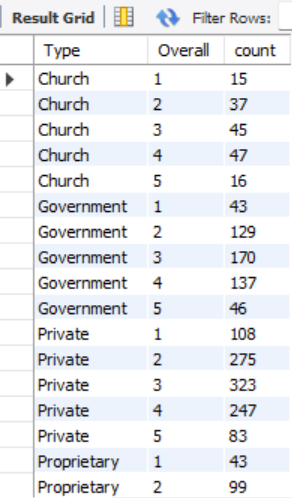
*facility*

*JOIN*

*rating ON facility.Facility\_ID = rating.Facility\_ID*

*GROUP BY facility.Type , rating.Overall*

*ORDER BY facility.Type , rating.Overall;*



***Heart Attack Quality Distribution***

*SELECT*

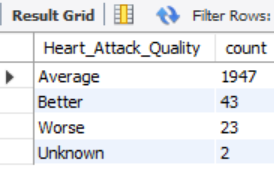
*Heart\_Attack\_Quality, COUNT(\*) AS count*

*FROM*

*procedures*

*GROUP BY Heart\_Attack\_Quality*

*ORDER BY count DESC;*



***#Top 10 Hospitals by Overall Rating***

*SELECT*

*f.Name, f.City, f.State, r.Overall*

*FROM*

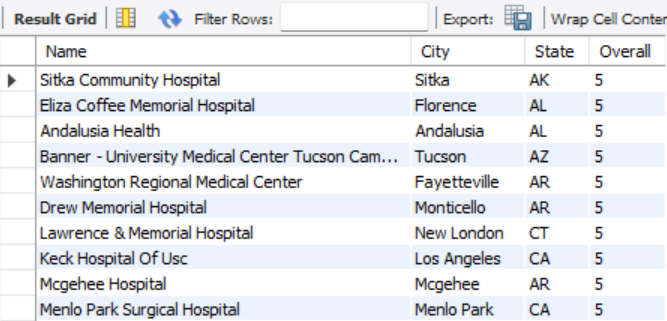
*facility f*

*JOIN*

*rating r ON f.Facility\_ID = r.Facility\_ID*

*ORDER BY r.Overall DESC*

*LIMIT 10;*



***Track Patient Records***

***#Total Number of Patients Per Facility***

*SELECT*

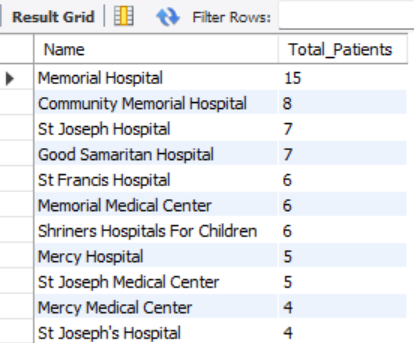
*Name, COUNT(\*) AS Total\_Patients*

*FROM*

*facility*

*GROUP BY Name*

*ORDER BY Total\_patients DESC;*



***#Patient Distribution by State:***

*SELECT*

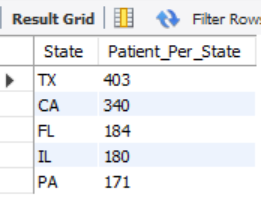
*State, COUNT(\*) AS Patient\_Per\_State*

*FROM*

*facility*

*GROUP BY State*

*ORDER BY Patient\_Per\_State DESC LIMIT 5*;



#HERE TX STAND FOR TEXAS WITH 403 TOTAL PATIENTS

--------------------------------------------------------------------------------------

***Evaluating Healthcare Metrics***

***#Average Rating of Facilities***:

*SELECT*

*facility.Name, AVG(rating.Overall) AS Avg\_Overall\_Rating*

*FROM*

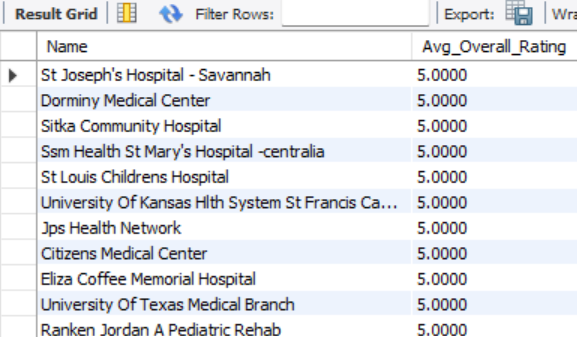
*facility*

*JOIN*

*rating ON facility.Facility\_ID = rating.Facility\_ID*

*GROUP BY facility.Name*

*ORDER BY Avg\_Overall\_Rating DESC;*



#***Identify Facilities with Low Safety Ratings***:

*SELECT*

*facility.Name,*

*facility.State,*

*rating.Safety AS low\_safety\_ratings*

*FROM*

*facility*

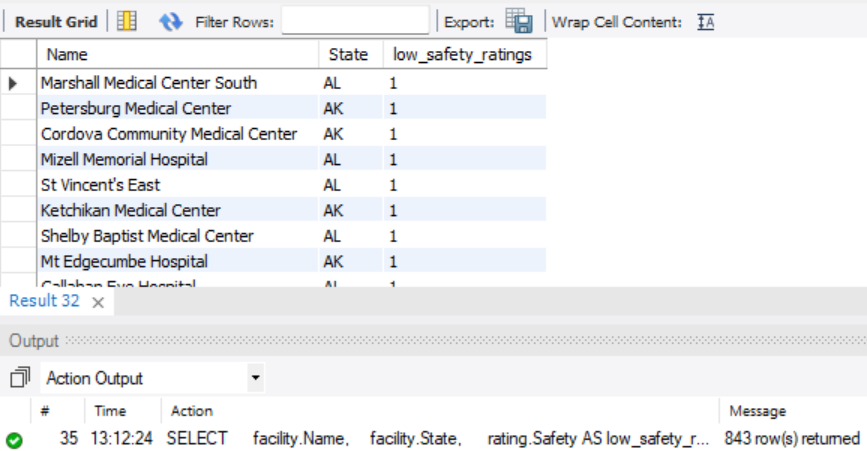
*JOIN*

*rating ON facility.Facility\_ID = rating.Facility\_ID*

*WHERE*

*Safety = 'Below'*

*GROUP BY facility.Name , facility.State , low\_safety\_ratings*;



***#Calculate Average Cost of Each Procedures***:

*SELECT*

*f.Name, AVG(p.Heart\_Attack\_Cost) AS avg\_heart\_attack\_cost*

*FROM*

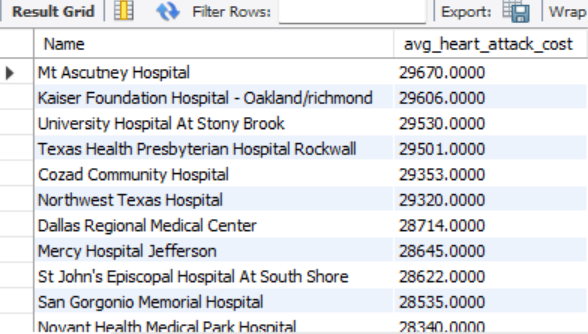
*facility f*

*JOIN*

*procedures p ON f.Facility\_ID = p.Facility\_ID*

*GROUP BY f.Name*

*ORDER BY avg\_heart\_attack\_cost DESC;*



---------

*SELECT*

*f.Name, AVG(p.Heart\_Failure\_Cost) AS avg\_Heart\_Failure\_Cost*

*FROM*

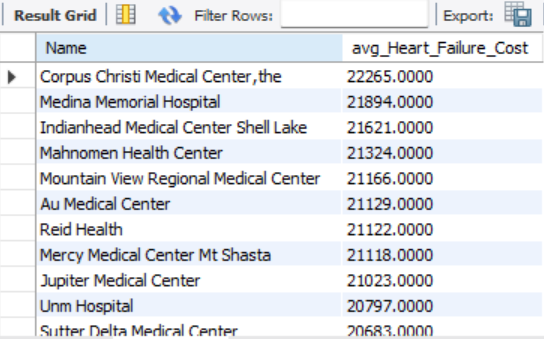
*facility f*

*JOIN*

*procedures p ON f.Facility\_ID = p.Facility\_ID*

*GROUP BY f.Name*

*ORDER BY avg\_Heart\_Failure\_Cost DESC;*



----------

*SELECT*

*f.Name, AVG(p.Hip\_Knee\_Cost) AS avg\_Hip\_Knee\_Cost*

*FROM*

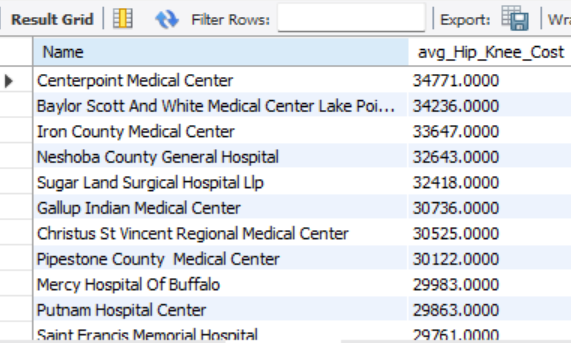
*facility f*

*JOIN*

*procedures p ON f.Facility\_ID = p.Facility\_ID*

*GROUP BY f.Name*

*ORDER BY avg\_Hip\_Knee\_Cost DESC;*



-----------

*SELECT*

*f.Name, AVG(p.Pneumonia\_Cost) AS avg\_Pneumonia\_Cost*

*FROM*

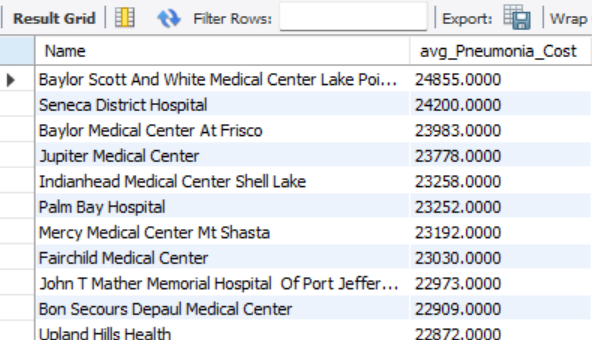
*facility f*

*JOIN*

*procedures p ON f.Facility\_ID = p.Facility\_ID*

*GROUP BY f.Name*

*ORDER BY avg\_Pneumonia\_Cost DESC;*



***Identify Areas for Improvement***

***#Facilities with High Readmission Rates:***

*SELECT DISTINCT*

*f.Name, f.State, r.Readmission, r.Overall AS Overall\_Rating*

*FROM*

*facility f*

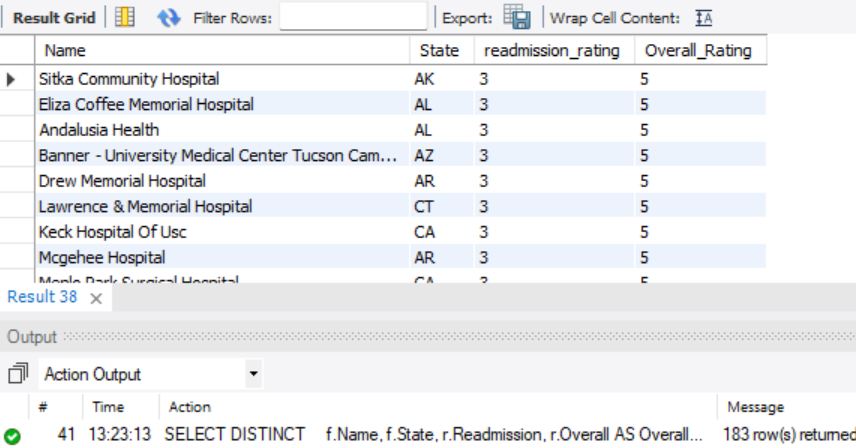
*JOIN*

*rating r ON f.Facility\_ID = r.Facility\_ID*

*WHERE*

*Readmission = 'Above' AND r.Overall = 5*

*ORDER BY Overall\_Rating DESC;*

* 🡪 As you can see there are 183 rows with the oveall rating of 5 & readmiiision rating 3 on the scale of (1-3)*



***#Patient Distribution by State(TOP 5):***

*SELECT*

*f.State, COUNT(f.State) AS Patient\_from\_each\_state*

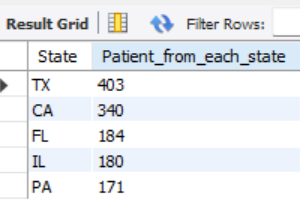
*FROM*

*facility f*

*GROUP BY f.State*

*ORDER BY Patient\_from\_each\_state DESC*

*LIMIT 5;*

**

***Ranking Facilities by Procedure Cost***

***#Rank Facilities by Heart Attack Procedure Cost***

*SELECT*

*f.Name,*

*f.City,*

*f.State,*

*p.Heart\_Attack\_Cost,*

*RANK() OVER (ORDER BY p.Heart\_Attack\_Cost DESC) AS Cost\_Rank*

*FROM*

*facility f*

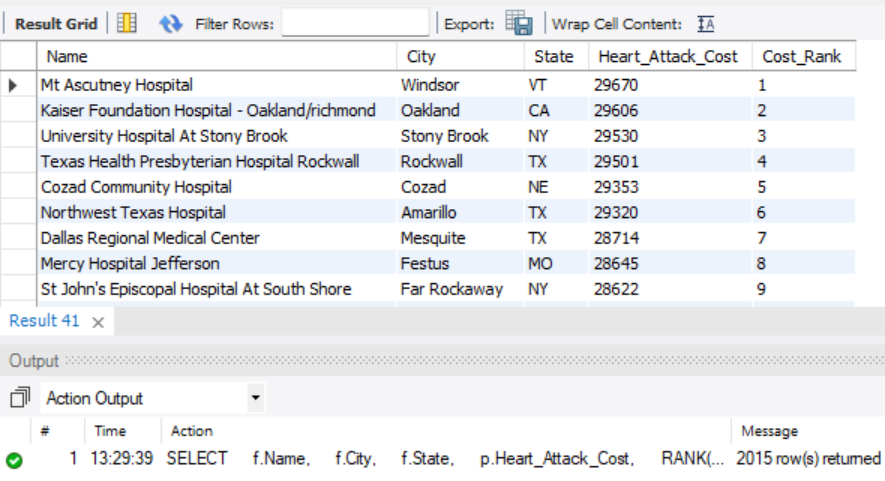
*JOIN*

*procedures p*

*ON f.Facility\_ID = p.Facility\_ID*

*ORDER BY*

*Cost\_Rank;*

**

***#Ranking Facilities by Overall Rating within Each State***

*SELECT*

*f.Name,*

*f.State,*

*r.Overall AS overall\_rating,*

*RANK() OVER (PARTITION BY f.State ORDER BY r.Overall) AS State\_Rank*

*FROM*

*facility f*

*JOIN*

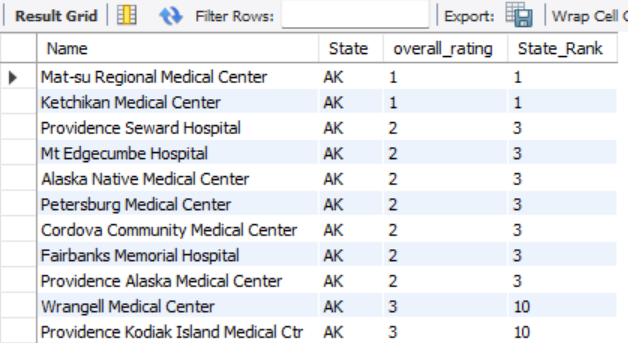
*rating r*

*ON f.Facility\_ID = r.Facility\_ID*

*ORDER BY*

f.State,

State\_Rank;



***Calculating Running Totals and Averages***

***#Running Total of Heart Attack Procedure Costs***

*SELECT*

*f.Name,*

*f.City,*

*f.State,*

*p.Heart\_Attack\_Cost,*

*SUM(p.Heart\_Attack\_Cost) OVER (ORDER BY p.Heart\_Attack\_Cost) AS Running\_total\_cost*

*FROM*

*facility f*

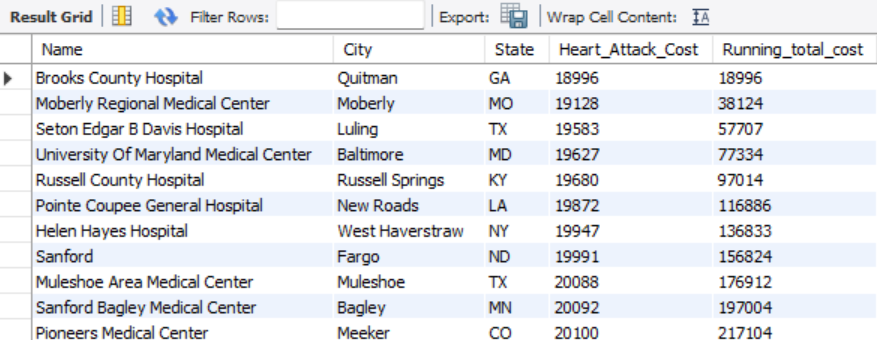
*JOIN*

*procedures p*

*ON f.Facility\_ID = p.Facility\_ID*

*ORDER BY*

*Running\_total\_cost;*



***#Moving Average of Heart Failure Procedure Cost***

*SELECT*

*f.Name,*

*f.City,*

*f.State,*

*p.Heart\_Failure\_Cost,*

*AVG(p.Heart\_Failure\_Cost) OVER (ORDER BY p.Heart\_Failure\_Cost) AS Moving\_Avg\_cost*

*FROM*

*facility f*

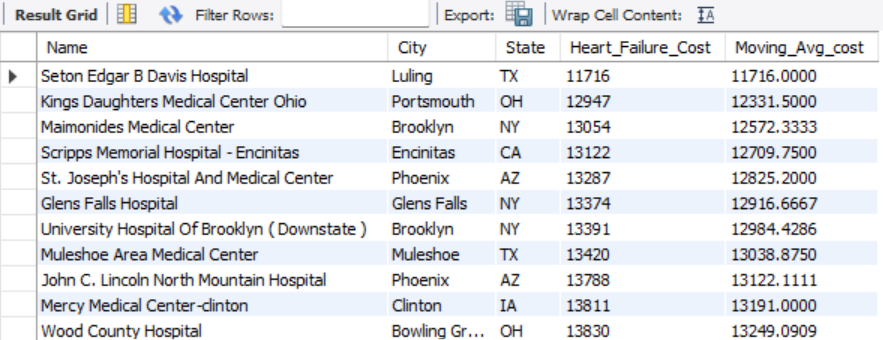
*JOIN*

*procedures p*

*ON f.Facility\_ID = p.Facility\_ID*

*ORDER BY*

*Moving\_Avg\_cost;*



***Cumulative Metrics***

***#Cumulative Count of Facilities by Rating***

*SELECT*

*f.Name,*

*f.City,*

*f.State,*

*r.Overall,*

*COUNT(\*) OVER (ORDER BY r.Overall) AS cumulative\_count*

*FROM*

*facility f*

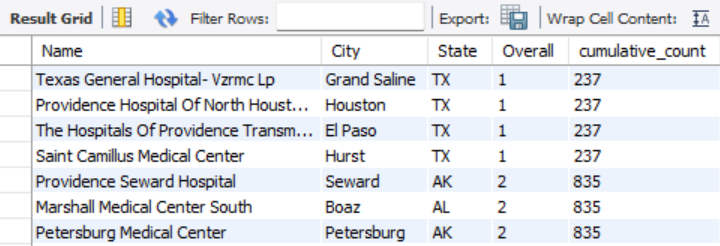
*JOIN*

*rating r*

*ON f.Facility\_ID = r.Facility\_ID*

*ORDER BY*

*cumulative\_count;*



***#Cumulative Average of Procedure Costs by Facility Type***

*SELECT*

*f.Type,*

*f.Name,*

*p.Pneumonia\_Cost,*

*AVG(p.Pneumonia\_Cost) OVER (PARTITION BY f.Type ORDER BY f.Name) AS cumulative\_avg\_cost*

*FROM*

*facility f*

*JOIN*

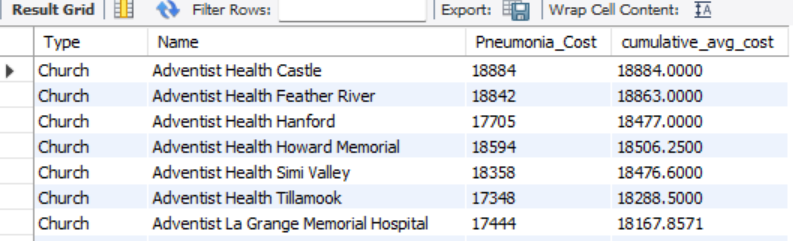
*procedures p*

*ON f.Facility\_ID = p.Facility\_ID*

*ORDER BY*

*f.Type,*

*f.Name;*



***Identifying Top and Bottom Performers***

***#Top 5 Facilities by Hip/Knee Procedure Cost within Each City***

*WITH RankedFacilities AS (*

*SELECT*

*f.Name,*

*f.City,*

*f.State,*

*p.Hip\_Knee\_Cost,*

*RANK() OVER (PARTITION BY f.City ORDER BY p.Hip\_Knee\_Cost DESC) AS City\_cost\_rank*

*FROM*

*procedures p*

*JOIN*

*facility f ON p.Facility\_ID = f.Facility\_ID*

*)*

*SELECT*

*Name,*

*City,*

*State,*

*Hip\_Knee\_Cost,*

*City\_cost\_rank*

*FROM*

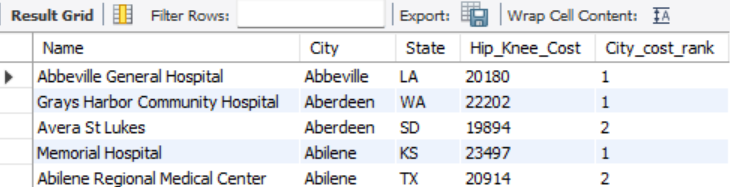
*RankedFacilities*

*WHERE*

*City\_cost\_rank <= 5*

*ORDER BY*

*City, City\_cost\_rank limit 5;*



***#Bottom 5 Facilities by Overall Rating***

*WITH RankedFacilities AS (*

*SELECT*

*f.Name,*

*f.City,*

*f.State,*

*r.Overall,*

*RANK() OVER (ORDER BY r.Overall asc) AS rating\_rank*

*FROM*

*rating r*

*JOIN*

*facility f ON r.Facility\_ID = f.Facility\_ID*

*)*

*SELECT*

*Name,*

*City,*

*State,*

*Overall,*

*rating\_rank*

*FROM*

*RankedFacilities*

*WHERE*

*rating\_rank <= 5*

*ORDER BY*

*City, rating\_rank limit 5;*

