**Problem Statement**  
Recent Covid-19 Pandemic has raised alarms over one of the most overlooked area to focus: Healthcare Management. While healthcare management has various use cases for using data science, patient length of stay is one critical parameter to observe and predict if one wants to improve the efficiency of the healthcare management in a hospital.  
This parameter helps hospitals to identify patients of high LOS risk (patients who will stay longer) at the time of admission. Once identified, patients with high LOS risk can have their treatment plan optimized to miminize LOS and lower the chance of staff/visitor infection. Also, prior knowledge of LOS can aid in logistics such as room and bed allocation planning.  
Suppose you have been hired as Data Scientist of HealthMan – a not for profit organization dedicated to manage the functioning of Hospitals in a professional and optimal manner.  
The task is to accurately predict the Length of Stay for each patient on case by case basis so that the Hospitals can use this information for optimal resource allocation and better functioning. The length of stay is divided into 11 different classes ranging from 0-10 days to more than 100 days.

**Data Description**  
train\_data.csv – File containing features related to patient, hospital and Length of stay on case basis  
train\_data\_dictonary.csv – File containing the information of the features in train file

Test Set  
test\_data.csv – File containing features related to patient, hospital. Need to predict the Length of stay for each case\_id

Sample Submission:

case\_id: Unique id for each case

Stay: Length of stay for the patient w.r.t each case id in test data

PROJECT QUESTIONS

**INNER JOIN**

QuestionsQ1:

List the case IDs and corresponding hospital types for all cases that are in both sample\_sub and test\_data tables.

Q2:

Find the case IDs and admission deposits for all cases that appear in both sample\_sub and train\_data tables.

**LEFT JOIN Questions**

Q3:

List all case IDs from sample\_sub and their corresponding hospital codes from test\_data, including those without a match in test\_data.

Q4:

Display the hospital region code and bed grade for each case in train\_data, even if there is no corresponding case in sample\_sub.

**FULL OUTER JOIN Questions**

Q5:

Show a list of all case IDs from both sample\_sub and test\_data, including those that do not have a match in the other table.

Q6:

Combine all records from train\_data and test\_data, showing the hospital code and patient ID from both tables.

**CROSS JOIN Questions**

Q7:

Create a combination of each case\_id from sample\_sub with every Hospital\_code from test\_data.

Q8:

Pair every Department in train\_data with every Ward\_Type in test\_data.

**SELF JOIN Questions**

Q9:

Find pairs of cases in test\_data that are from the same city and have different hospital codes.

Q10:

Identify cases in train\_data that have the same department but are of different severity of illness.