Patient Mortality Prediction (ICU Patients)

# Dataset Overview:

* **Entries and Columns**: There are 37 rows, each corresponding to a patient record, across 28 different columns.
* **Data Types**: The columns contain a mix of data types:
  + **int64** for integer values (4 columns) such as **subject\_id**, **hadm\_id**, **age**, and **hospital\_expire\_flag**.
  + **float64** for floating-point numbers (14 columns) covering various measurements like **icu\_los**, **heart\_rate\_mean**, and **temperature\_fahrenheit**, among others.
  + **object** for text or mixed data types (10 columns), including **edadmittime**, **age\_bucket**, **gender**, and **ethnicity**, to name a few.
* **Non-Null Counts**: All columns have non-null values for all 37 entries, except for **height\_cm** (22 non-null values), **admission\_weight\_lbs** (34 non-null values), **mixed\_venous\_o2\_sat** (8 non-null values), and **ph\_dipstick** (12 non-null values), indicating some missing data in these areas.
* **Unique Identifiers**: The **subject\_id** and **hadm\_id** columns likely serve as identifiers for patients and their admission instances, respectively.
* **Temporal Data**: The **edadmittime** column, which appears twice in the summary (possibly due to a duplication or parsing issue), represents times of admission.
* **Clinical Measurements**: Several columns record clinical measurements and assessments, such as heart rate, GCS (Glasgow Coma Scale) scores, blood pressure, respiratory rate, and oxygen saturation levels.
* **Demographics and Admission Details**: Demographic information (e.g., age, gender, ethnicity) and admission details (e.g., admission type, care unit) are included, providing insights into the patient population and their initial care settings.
* **Outcome Indicator**: The **hospital\_expire\_flag** column is a binary indicator (likely 0 or 1), signifying whether the patient died during their hospital stay.

# Exploratory Data Analysis (EDA):

Please refer to files:

1. report.html
2. sweetvizReport.html

# Models & Evaluation Metrics:

1. Random Forest:
   1. Model Accuracy - 0.9387248136903119
2. Random Forest – Hyper Tuned:
   1. Best accuracy found: 0.9353347135955832
   2. Accuracy with best parameters: 0.9383107921611924
3. XG Boost:
   1. Accuracy: 0.9314104333425338
4. XG Boost – Hyper Tuned:
   1. Accuracy: 0.94

## Pickle Files:

Finally, we considered to go ahead with XG Boost Model after hyper tuning to create pkl file.