# Predicting Sepsis Survival Using Clinical Data

**Zhaocheng Yang**

Instructor: Andras Zsom

TA: Mingjun Ma

Data Science Institute, Brown University

10/25/2024

- GitHub

# Introduction

| | |
|---|---|
| **What is Sepsis?** | A life-threatening condition triggered by an extreme immune response to infection. |
| **Importance** | 1. High mortality rate<br>2. Early prediction is critical for timely intervention and treatment |
| **Objective** | Build a classification model to predict patient survival using clinical features. |
| **Data Source:** | 1. UC Irvine Machine Learning Repository<br>2. Dataset from Norwegian hospital admissions (2011-2012) of patients with sepsis-related diagnoses. |

# EDA
# Dataset Overview

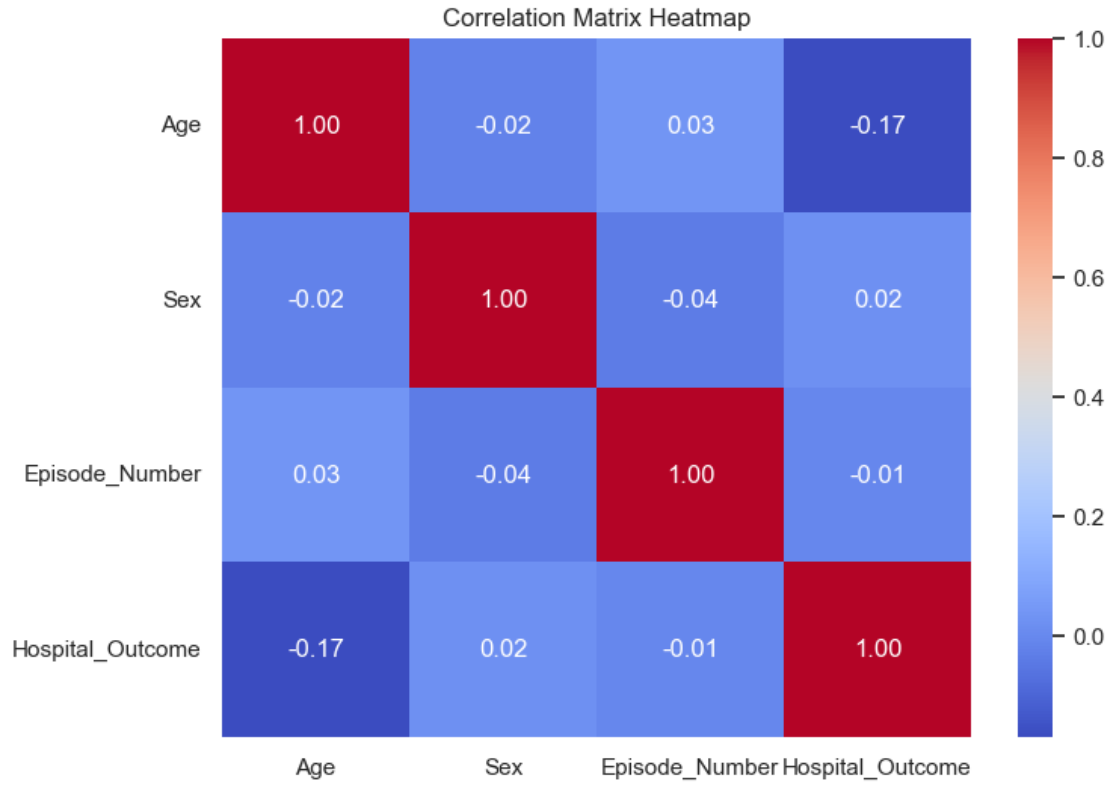
- Number of rows: 110,341
- Number of columns: 4
- Missing Values: NO

- Age: Age of the patient in years.
- Sex: Gender of the patient. (0: male, 1: female)
- Episode Number: Number of prior Sepsis episodes

  [1, 2, 3, 4, 5]

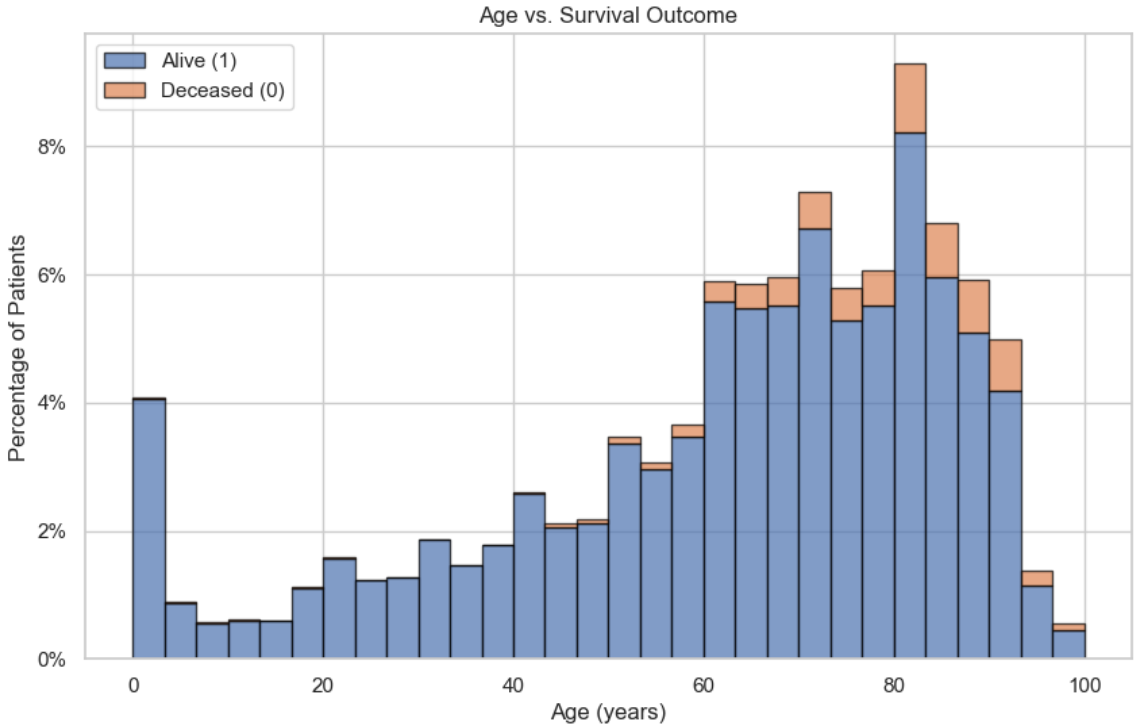
- Hospital_Outcome: Status of the patient after 9,351 days of being admitted to the hospital. (0: Decreased, 1: Alive)

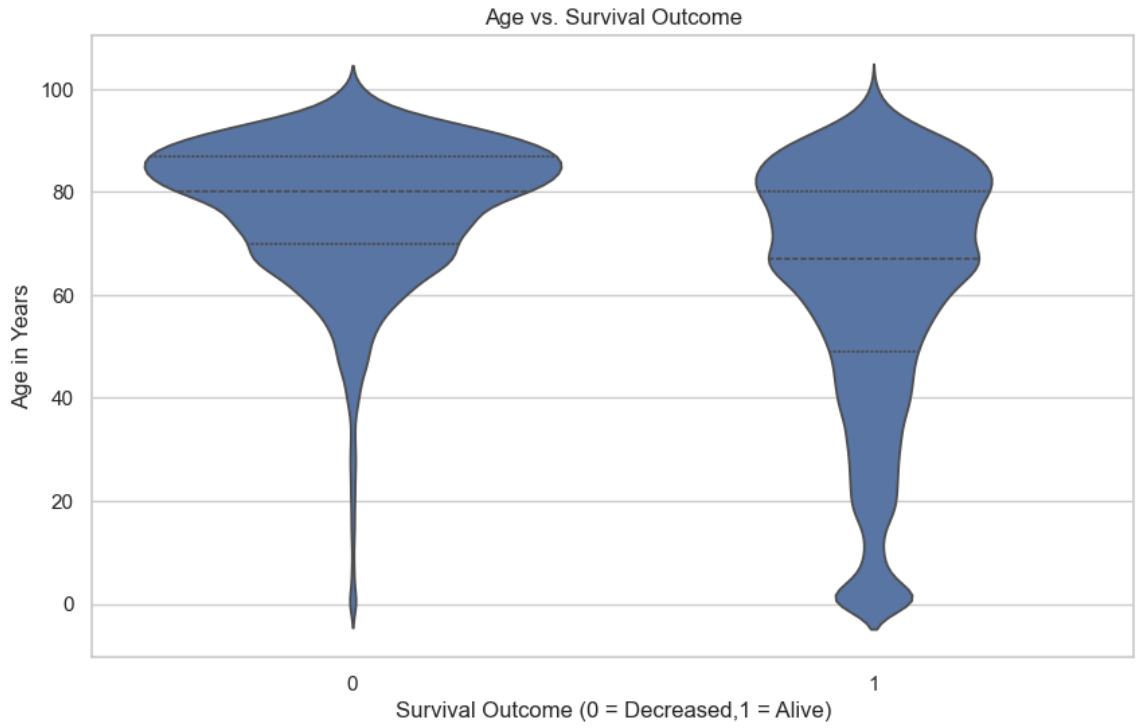
| | Age | Sex | Episode_Number | Hospital_Outcome |
|---|---|---|---|---|
| 0 | 21 | 1 | 1 | 1 |
| 1 | 20 | 1 | 1 | 1 |
| 2 | 21 | 1 | 1 | 1 |
| 3 | 77 | 0 | 1 | 1 |
| 4 | 72 | 0 | 1 | 1 |



# EDA
# Data Visualization 1

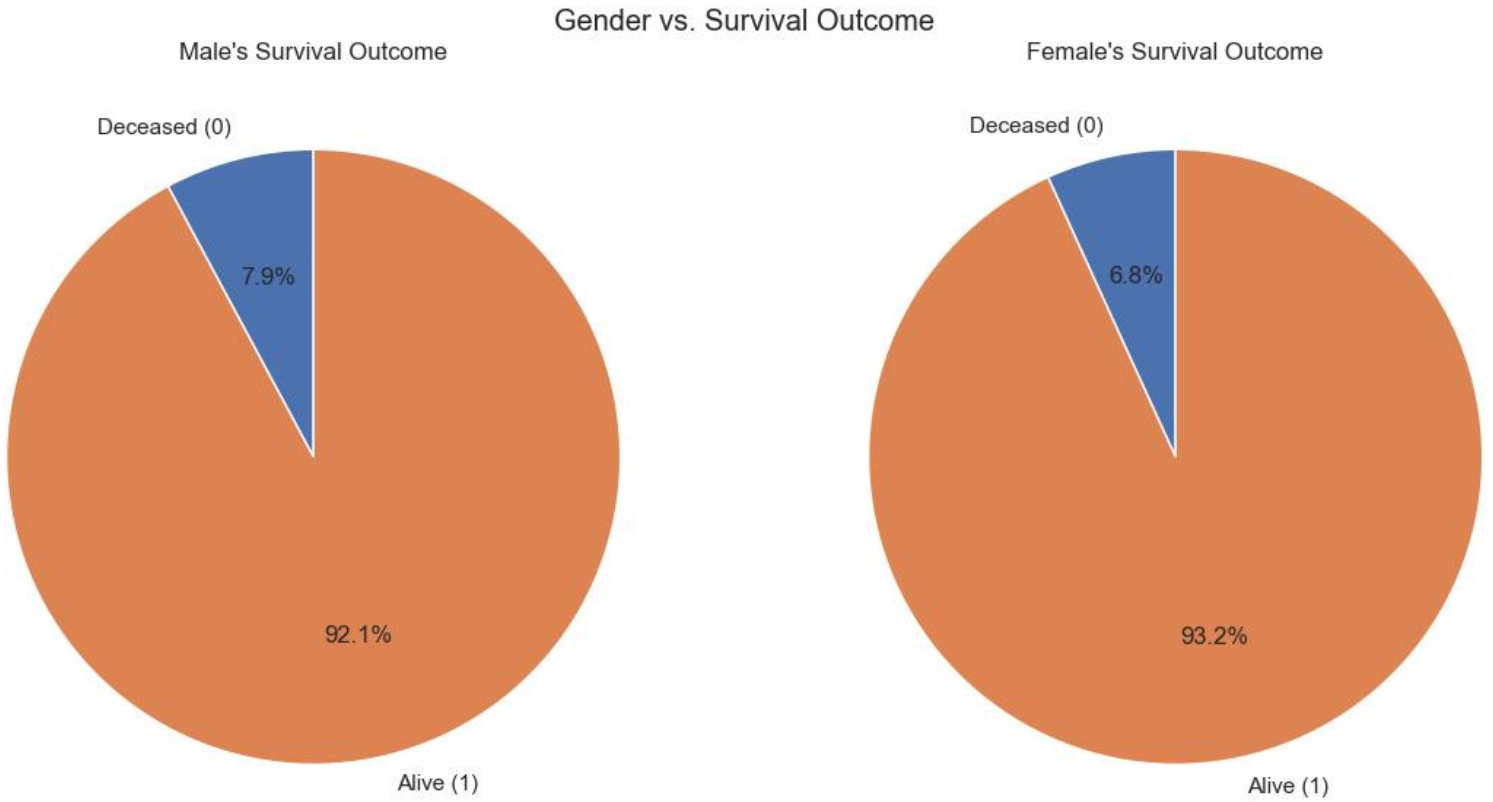
- Younger patients have higher survival rates.
- Mortality increases with age.



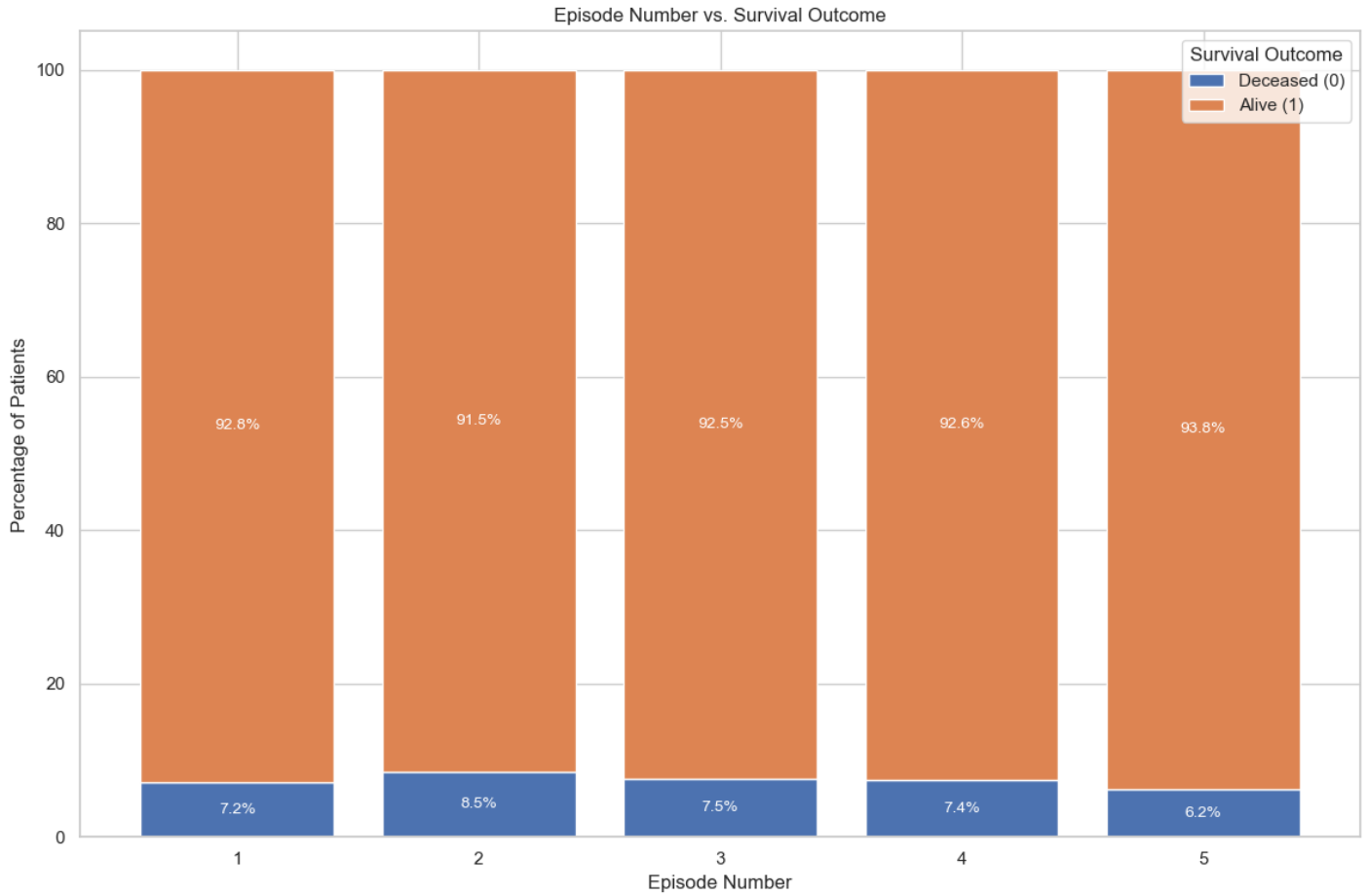
# EDA
# Data Visualization 2

- Slightly differences in survival rates between genders.



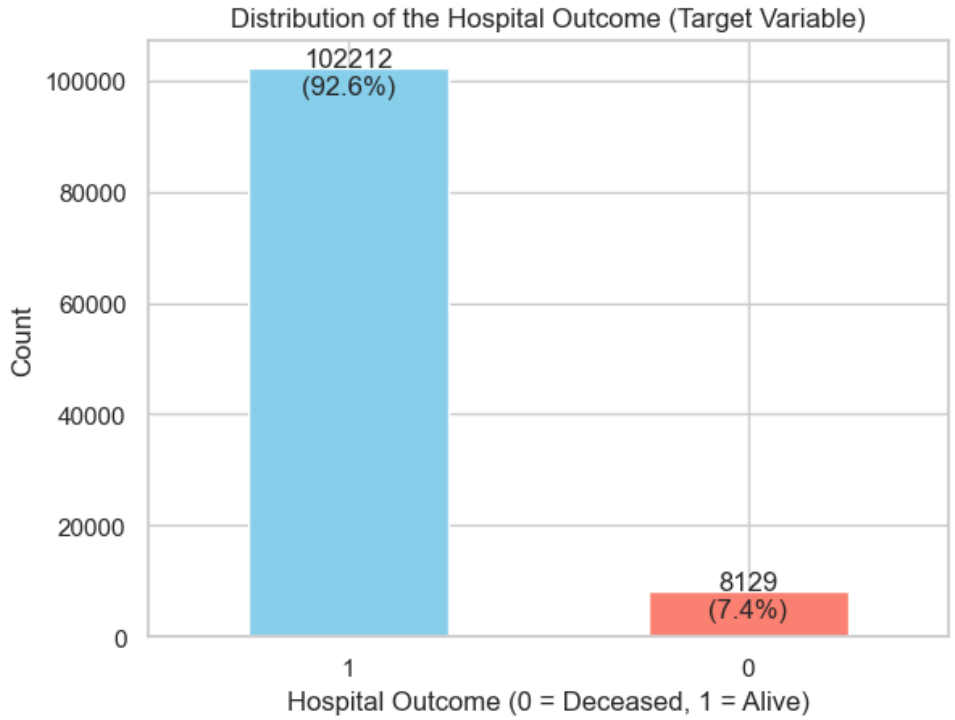
# EDA
# Data Visualization 3

- Unexpectedly, the mortality does not always increase with more episodes.



# Data Splitting

StratifiedKFold(n_splits=4)




## After the Splitting:

```
Train size: 66204 (60.00%)
Validation size: 22068 (20.00%)
Test size: 22069 (20.00%)
```

| | Train | Validation | Test |
|---|---|---|---|
| 1 | 92.6% (61326) | 92.6% (20443) | 92.6% (20443) |
| 0 | 7.4% (4878) | 7.4% (1625) | 7.4% (1626) |

The proportion of 0 and 1 remains consistent

# Data Preprocessing

| Feature | Type | Transformer |
|---|---|---|
| Age | Continuous | MinMaxScaler |
| Sex | Binary: 0 and 1 | \ |
| Episode_Number | Ordinal: 1,2,3,4,5 | OrdinalEncoder |

'X_train' Before preprocessing:

| | Age | Sex | Episode_Number |
|---|---|---|---|
| count | 66204.000000 | 66204.000000 | 66204.000000 |
| mean | 62.750378 | 0.473838 | 1.346112 |
| std | 24.090796 | 0.499319 | 0.745333 |
| min | 0.000000 | 0.000000 | 1.000000 |
| 25% | 51.000000 | 0.000000 | 1.000000 |
| 50% | 69.000000 | 0.000000 | 1.000000 |
| 75% | 81.000000 | 1.000000 | 1.000000 |
| max | 100.000000 | 1.000000 | 5.000000 |

'X_train' After preprocessing:

| | Age | Sex | Episode_Number |
|---|---|---|---|
| count | 66204.000000 | 66204.000000 | 66204.000000 |
| mean | 0.627504 | 0.473838 | 0.346112 |
| std | 0.240908 | 0.499319 | 0.745333 |
| min | 0.000000 | 0.000000 | 0.000000 |
| 25% | 0.510000 | 0.000000 | 0.000000 |
| 50% | 0.690000 | 0.000000 | 0.000000 |
| 75% | 0.810000 | 1.000000 | 0.000000 |
| max | 1.000000 | 1.000000 | 4.000000 |

# Thanks for Watching