

Data Collection And Preprocessing

Date	10 July 2024
Team ID	740021
Project Title	Sepsis Survival Minimal Clinical Records
Maximum Marks	6 Marks

Data Exploration and Preprocessing Report

Dataset variable will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling ,and forming a strong foundation for insights and predictions.

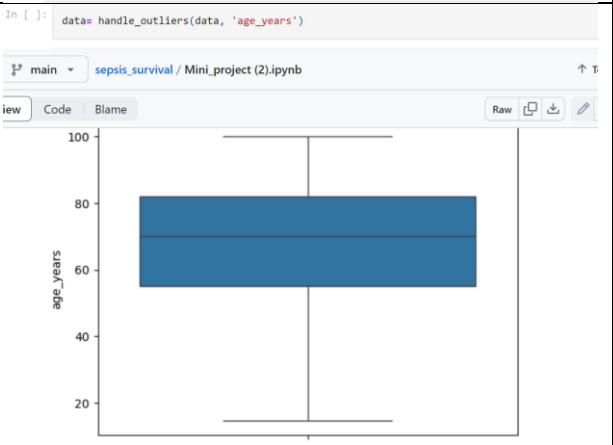
Section	Description
---------	-------------

Data overview

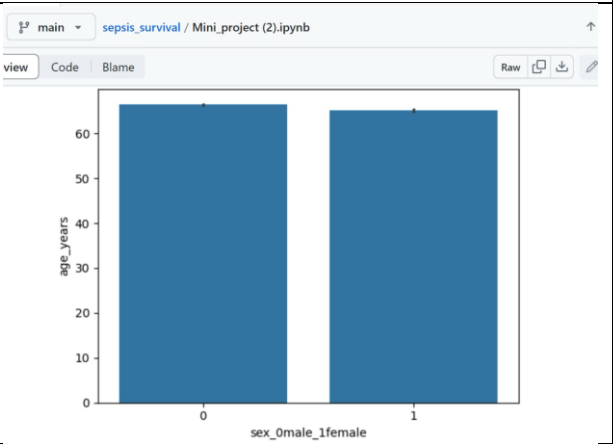
```
In [ ]: data.describe()
```

	age_years	sex_0male_1female	episode_number	hospital_outcome_1alive_0dead
count	110204.000000	110204.000000	110204.000000	110204.000000
mean	62.735255	0.473948	1.349379	0.926455
std	24.126806	0.499323	0.751799	0.261031
min	0.000000	0.000000	1.000000	0.000000
25%	51.000000	0.000000	1.000000	1.000000
50%	68.000000	0.000000	1.000000	1.000000
75%	81.000000	1.000000	1.000000	1.000000
max	100.000000	1.000000	5.000000	1.000000

Univariate Analysis



Bivariate Analysis



Multivariate Analysis



Data Preprocessing Code Screenshots

Loading	<pre> In []: data Out[]: age_years sex_0male_1female episode_number hospital_outcome_1alive_0dead 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 110199 0 0 1 1 110200 0 1 1 1 110201 70 1 1 1 110202 0 0 1 1 110203 0 0 1 1 </pre>
Handling Missing Data	<pre> In []: #checking for missing values data.isna().sum() Out[]: age_years 0 sex_0male_1female 0 episode_number 0 hospital_outcome_1alive_0dead 0 dtype: int64 </pre>
Data Transformation	-
Feature Engineering	Attached the codes in final submission
Saved Processed Data	-