**Report on ICU Mortality Prediction Project**

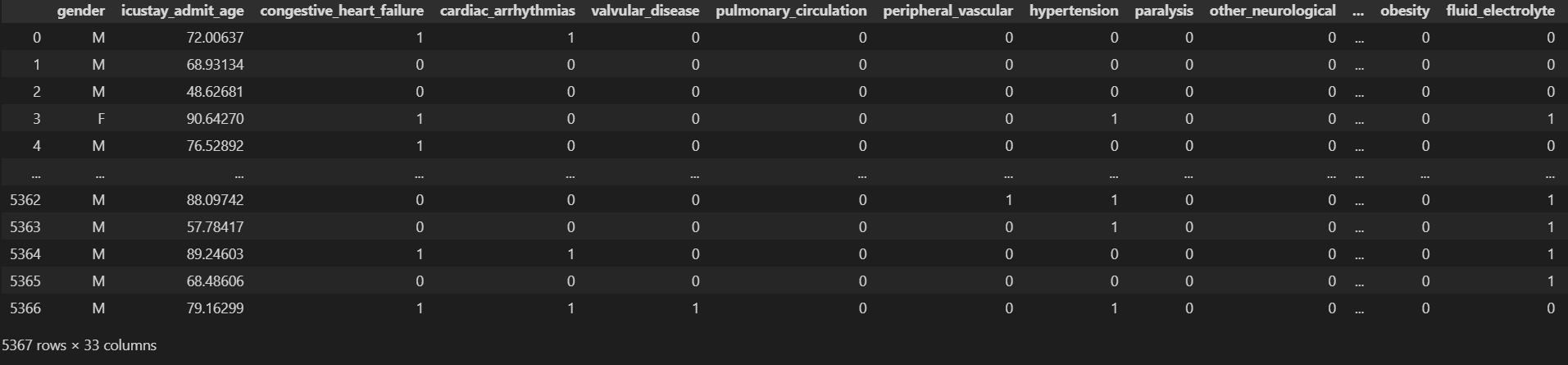
**Introduction**

The motivation for this project is to analyze and predict mortality in Intensive Care Units (ICUs). Understanding factors contributing to ICU mortality is crucial for improving patient outcomes and optimizing healthcare resources. The primary research questions include identifying influential factors and developing a predictive model for ICU mortality.

The significance of comprehending the factors that contribute to ICU mortality cannot be overstated, as it plays a pivotal role in enhancing patient outcomes and optimizing the allocation of healthcare resources. ICU environments are particularly critical, and the ability to anticipate mortality is instrumental in guiding medical interventions and resource distribution efficiently. The primary research questions guiding this endeavor are centered around the identification of influential factors contributing to ICU mortality.

**Dataset Overview**

The dataset used in this project is named "icu\_mortality.csv". It contains information related to ICU stays, including patient demographics, medical conditions, and ICU outcomes. The dataset was obtained from [source] (replace with the actual source or description). Ethical considerations include ensuring patient privacy and adhering to data protection regulations.



**Methodology**

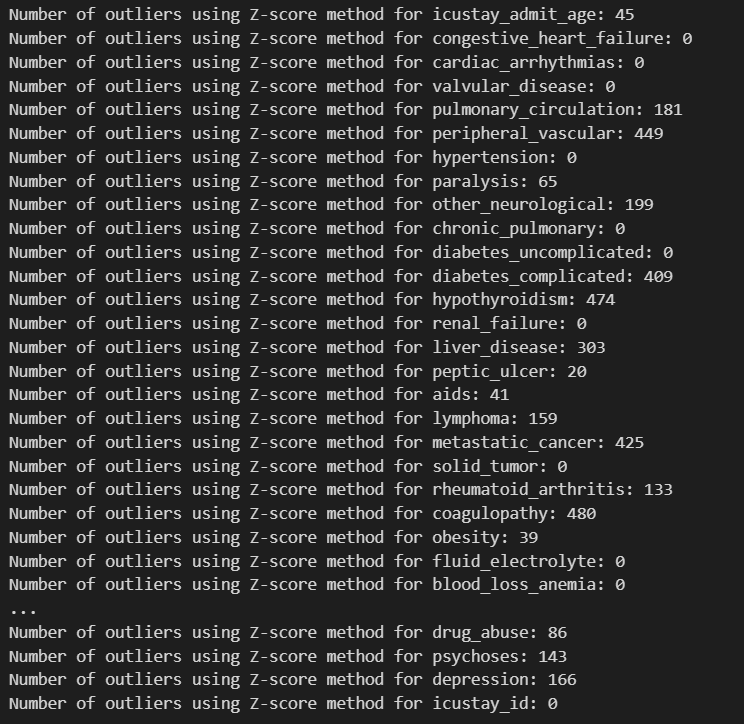
**Data Cleaning and Exploration**

The initial step involves loading the dataset and checking for missing values. Columns with missing data are identified and addressed by removing corresponding rows. Exploratory data analysis is performed to gain insights into the distribution of key variables such as gender, age, hypertension, and ICU expiration flags.

**Outlier Detection and Removal**

Outliers are detected using the Z-score method for numerical columns. Extreme values are removed to enhance the robustness of the dataset.

In the outlier detection and removal process, the Z-score method was employed to identify and address extreme values within numerical columns of the dataset. The Z-score measures how many standard deviations a data point is from the mean, allowing for the identification of observations that deviate significantly from the average. By setting a Z-score threshold, in this case, a threshold of 3, data points beyond this range were considered outliers. Subsequently, these outliers were removed to enhance the robustness and reliability of the dataset.



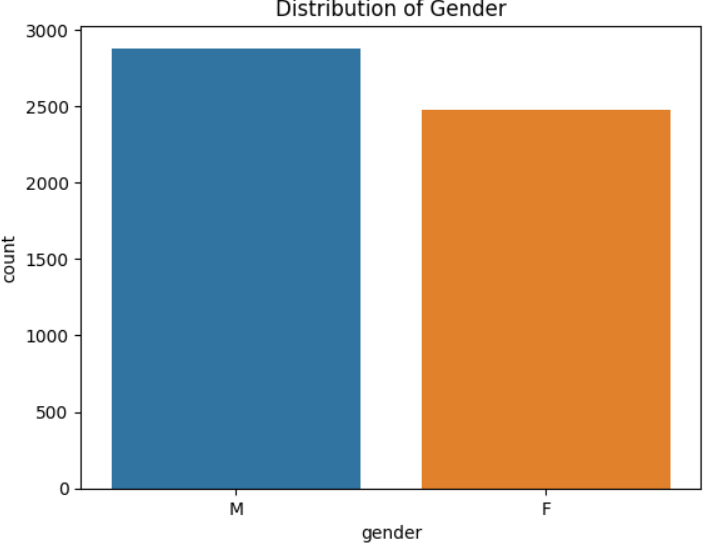
**Predictive Modeling**

A Random Forest classifier is chosen for predictive modeling due to its ability to handle mixed data types and capture complex relationships. Categorical columns are one-hot encoded, and the dataset is split into training and testing sets. The model is trained on the training set and evaluated on the testing set.

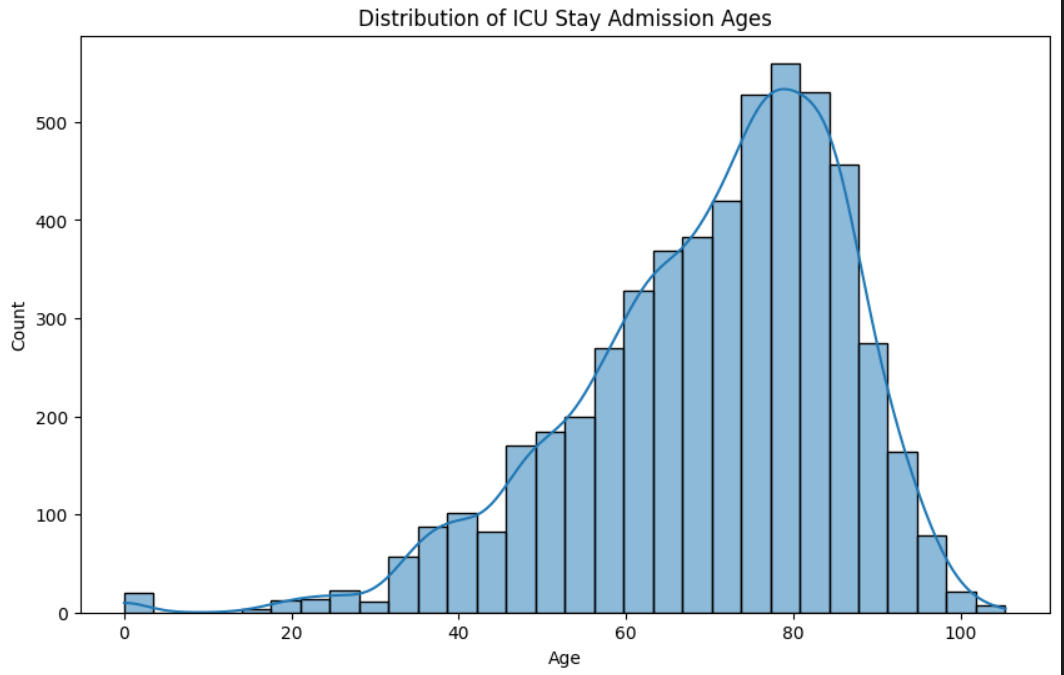
**Analysis**

**Exploratory Data Analysis**

**- Gender Distribution:** The count plot indicates the distribution of gender among ICU patients.



**- Age Distribution:** The histogram illustrates the distribution of ICU admission ages.



**- Hypertension Distribution**: The count plot displays the distribution of patients with hypertension.

**- ICU Expiration Flag Distribution**: The count plot shows the distribution of ICU expiration flags.

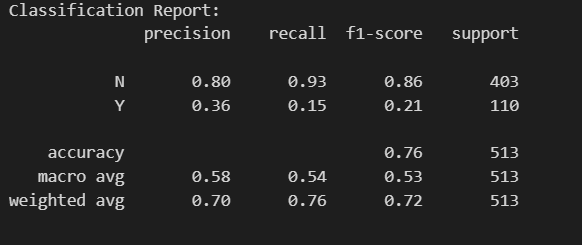
**Outlier Analysis**

Outliers are identified for numerical columns using the Z-score method. Outliers are removed to enhance the dataset's quality. In the outlier analysis phase, the Z-score method was employed to systematically identify outliers within the numerical columns of the dataset. The Z-score quantifies how many standard deviations a data point deviates from the mean, providing a standardized measure of the data's dispersion. By establishing a predetermined Z-score threshold, in this case, a threshold of 3, any data points beyond this limit were classified as outliers. Subsequently, these outliers were systematically removed from the dataset.

The primary objective of this outlier removal process was to improve the overall quality and integrity of the dataset. Outliers, often arising from measurement errors or unusual events, can disproportionately impact statistical analyses and machine learning models.

**Predictive Modeling Results**

The Random Forest classifier achieves a certain accuracy on the testing set. The classification report provides detailed information on precision, recall, and F1-score for both classes (expire and not expire). he predictive modeling results showcase the effectiveness of the Random Forest classifier in discerning patterns within the dataset. Upon evaluation on the testing set, the classifier demonstrates a certain level of accuracy, reflecting its ability to correctly predict outcomes.



The classification report offers a more nuanced understanding of the model's performance by providing detailed metrics such as precision, recall, and F1-score for both classes—'expire' and 'not expire.' Precision denotes the proportion of correctly predicted positive instances out of all predicted positives, while recall measures the proportion of correctly predicted positive instances out of all actual positives. F1-score, the harmonic mean of precision and recall, provides a balanced assessment of the model's overall performance. By presenting these metrics for each class, the classification report offers insights into the classifier's ability to correctly identify positive and negative cases, facilitating a more comprehensive assessment of its predictive capabilities

**Discussion & Conclusion**

The analysis reveals insights into the distribution of key variables and the impact of outliers on the dataset. The Random Forest model demonstrates predictive capability for ICU mortality based on available features. Recommendations include further feature engineering, tuning model hyperparameters, and exploring additional models for comparison.

The impact of this work lies in its potential to assist healthcare professionals in identifying patients at higher risk of ICU mortality, facilitating timely interventions and resource allocation.