

**GROUP PROJECT (ST 3011)**

**DEPARTMENT OF STATISTICS**

**Group C**

**gROUP MEMBERS**

s13993

S13873

S14068

S14099

S14043

S14113

S14123

**Table of Contents**

[List of Figures and Tables 2](#_Toc62840405)

[1. Introduction 3](#_Toc62840406)

[2. Data 4](#_Toc62840407)

[2.1 Data preparation 4](#_Toc62840408)

[3. Dashboard Implementation 5](#_Toc62840409)

[3.1 Libraries Used to Implement Dashboard 5](#_Toc62840410)

[3.2 Interface 6](#_Toc62840411)

[3.3 Charts 9](#_Toc62840412)

[3.3.1 Pie chart 9](#_Toc62840413)

[3.3.2 Histogram 10](#_Toc62840414)

[3.3.3 Stacked bar chart 11](#_Toc62840415)

[3.3.4 Box plot 12](#_Toc62840416)

[4. Conclusion and Recommendation 13](#_Toc62840417)

[4.1 Conclusion 13](#_Toc62840418)

[4.2 Recommendations 13](#_Toc62840419)

[5. References 14](#_Toc62840420)

# List of Figures and Tables

[Table 2. 1 Re-coded Categorical variables 4](#_Toc62840437)

[Figure 3.2. 1 Dashboard Interface 6](file:///C:\Users\user\Desktop\group%20ST3011\Report%20-%20Group%20C.docx#_Toc62841101)

[Figure 3.2. 2 Dashboard Overview 1 - Drop down menu 7](file:///C:\Users\user\Desktop\group%20ST3011\Report%20-%20Group%20C.docx#_Toc62841102)

[Figure 3.2. 3 Dashboard Overview 2 8](file:///C:\Users\user\Desktop\group%20ST3011\Report%20-%20Group%20C.docx#_Toc62841103)

[Figure 3.3.1. 1 Pie Chart of diabetes 9](file:///C:\Users\user\Desktop\group%20ST3011\Report%20-%20Group%20C.docx#_Toc62841609)

[Figure 3.3.1. 2 Pie Chart of Death Event 9](file:///C:\Users\user\Desktop\group%20ST3011\Report%20-%20Group%20C.docx#_Toc62841610)

[Figure 3.3.2. 1 Histogram of Age 10](file:///C:\Users\user\Desktop\group%20ST3011\Report%20-%20Group%20C.docx#_Toc62841631)

[Figure 3.3.2. 2 Histogram of Serum Sodium 10](file:///C:\Users\user\Desktop\group%20ST3011\Report%20-%20Group%20C.docx#_Toc62841632)

[Figure 3.3.3. 1 Stacked bar chart of diabetes vs. death event 11](file:///C:\Users\user\Desktop\group%20ST3011\Report%20-%20Group%20C.docx#_Toc62841655)

[Figure 3.3.3. 2 Stacked bar chart of sex vs. death event 11](file:///C:\Users\user\Desktop\group%20ST3011\Report%20-%20Group%20C.docx#_Toc62841656)

[Figure 3.3.4. 1 Box plot of serum sodium vs. death event 12](file:///C:\Users\user\Desktop\group%20ST3011\Report%20-%20Group%20C.docx#_Toc62841663)

[Figure 3.3.4. 2 Box plot of platelets vs. death events 12](file:///C:\Users\user\Desktop\group%20ST3011\Report%20-%20Group%20C.docx#_Toc62841664)

# Introduction

Data has become the major aspect of this globalized world. Almost all organizations and business companies tend to focus on collecting relevant data and make decisions by analyzing those data. If those analyzed data can be presented in a much simpler, attractive and user-friendly way, it would be a great influence for each person to deal with this data even if they do not have any statistical background. In this context, dashboards have been already led the way and playing a major role.

A dashboard can be defined as a user-friendly graphical interface where all the important aspects are visualized in one frame. One of the major aspects of a dashboard is to monitor the whole process in one glance and give easy access to all the data and information which improves the efficiency of the system and also the decision making process and further forecasting. Apart from that, data visualization via dashboards helps views to understand and absorb the information quickly rather than going through numerous paragraphs of analyzed data.

Mainly dashboards are divided into three types as operational, strategic and analytical. Among these three types, an analytical dashboard is used for this study as it is needed to visualize and compare the data to make predictions.

The given dataset is based on the death events due to heart failures and the factors which can impact on heart failures. In this study, the statistical analysis was conducted to identify the distribution of variables, to get insights of the data set and to identify the association between each factor with the death event that occurred due to heart failures. Since the underlying problem is common among the general public, it is important to be aware of the causes which could lead to deaths due to heart failure. Then, they would be able to take preventive measures to avoid these factors. Therefore, using a dashboard in this study might be easier for people to select specific factors and view the results attractively and understandably rather than using a lengthy analysis report.

# Data

**Response variable** : **-death event**: if the patient deceased during the follow-up period (boolean)

**Explanatory variables**

- **age**: Age of the patient (years) - Quantitative

- **anaemia:** Decrease of red blood cells or hemoglobin  (Categorical - binary)

- **diabetes:** If the patient has diabetes  (Categorical - binary)

- **platelets:** Platelets in the blood (kiloplatelets/mL) - Quantitative

- **sex:** Male/Female (Categorical - binary)

- **serum creatinine:** Level of serum creatinine in the blood (mg/dL) - Quantitative

- **serum sodium:** Level of serum sodium in the blood (mEq/L) - Quantitative

- **smoking:** If the patient smokes or not (Categorical - binary)

* 1. **Data preparation**

The original dataset contains 299 observations with no missing values and no duplicates. Therefore all 299 observations were used for the analysis.

* The data type of variable age was ‘float’ type in the original data set and was converted into ‘int’ type.

|  |  |
| --- | --- |
| Death event | 1 - Dead , 0 - Alive |
| Sex | 1 -  Male , 0 - Female |
| Anaemia | 1 - Positive , 0 - Negative |
| Diabetes | 1 - Positive , 0 - Negative |

* The binary variables (Death event, sex, anaemia, diabetes) were coded as 1, 0 in the original dataset and the corresponding data type was integer for all those binary variables. Therefore, to improve the clearness of data visualizations, binary variables were re-coded as follows after changing the data type into ‘category’. *categorical variable*

Table 2. 1 Re-coded Categorical variables

1. **Dashboard Implementation**
   1. **Libraries Used to Implement Dashboard**

Recently, Python has become the most widely used programming language among the data science community as it is known as the most user friendly-programming language and also easier to learn even for beginners who do not have any programming background. As a result of this, **Dash** library in Python leads the way of creating dashboards by exceeding “Shiny” and “Flexdashboard” in R programming language. Currently, statisticians widely use dash not only because of the popularity of Python but also due to the many useful and versatile features carrying by Dash.

In Dash, it is not required any knowledge of HTML, JavaScript or CSS. Dash is purely written in Python and some extra features have been added to Dash in order to make the dashboards more attractive with the help of HTML, CSS and JavaScript. Therefore, some more libraries are imported after Dash library.

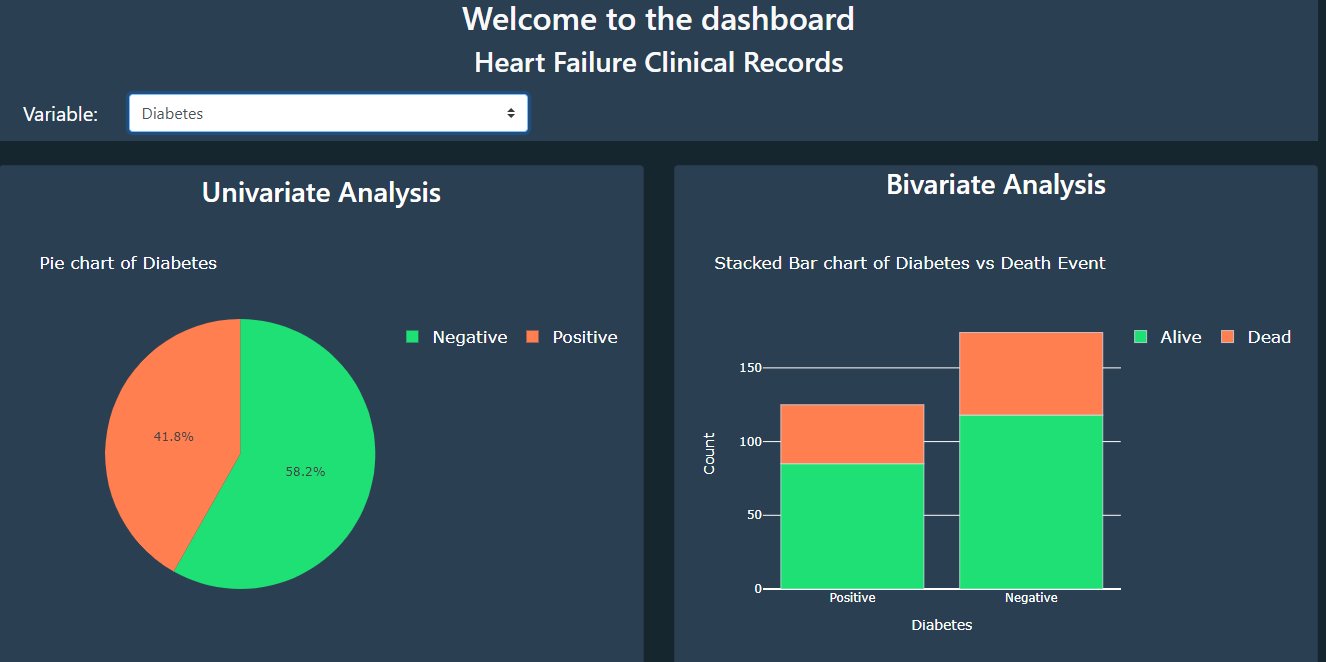
* **dash\_html\_components** – This library is imported as html. With the help of this library, layouts can be composed using Python structures instead of writing HTML codes. CSS and JavaScript styles can be also added to these HTML layouts.
* **dash\_bootstrap\_components –** This library is imported as dbc. Here, components are already made-up with styles. In this project, this library has been used to create a drop-down menu and navigation bar.
* **dash\_core\_components** – This library; which is imported as dcc, has been used to add graphs to the dashboard.
* **Pandas –** This is imported as pd and it is much useful for data analyzing and creating data frames.
* **plotly.express –** Plotly library is used to create attractive graphs to visualize the data and this library is imported as px.
  1. **Interface**
* **Layout of the app** - The layout consists of 3 main parts.
* Top layer – Headings and a menu to select a variable (dropdown list)
* Right layer – Bivariate graph
* Left layer – Univariate graph
* **Cards** - Two main cards are contained in the dashboard
* Card in the right layer – Contain the card title as ‘Bivariate Analysis’ and a graph for Bivariate Analysis.
* Card in the left layer – Contain the card title as ‘Univariate Analysis’ and a graph for Univariate Analysis.

Figure 3.2. 1 Dashboard Interface

* **Nav –** Two titles are created as “welcome to the dashboard ” and “heart failure clinical records '' using html.nav element. styles are added to them using external CSS styles.
* **Navigation bar –** A navigation bar is created to insert the **“**Variable” dropdown menu. It is created using dbc.navbar in the dash bootstrap components library.
* **Drop-down menu** – the dropdown menu in the navigation bar is created for selecting a variable. The dropdown menu is recreated using dbc.select in dash bootstrap components library.

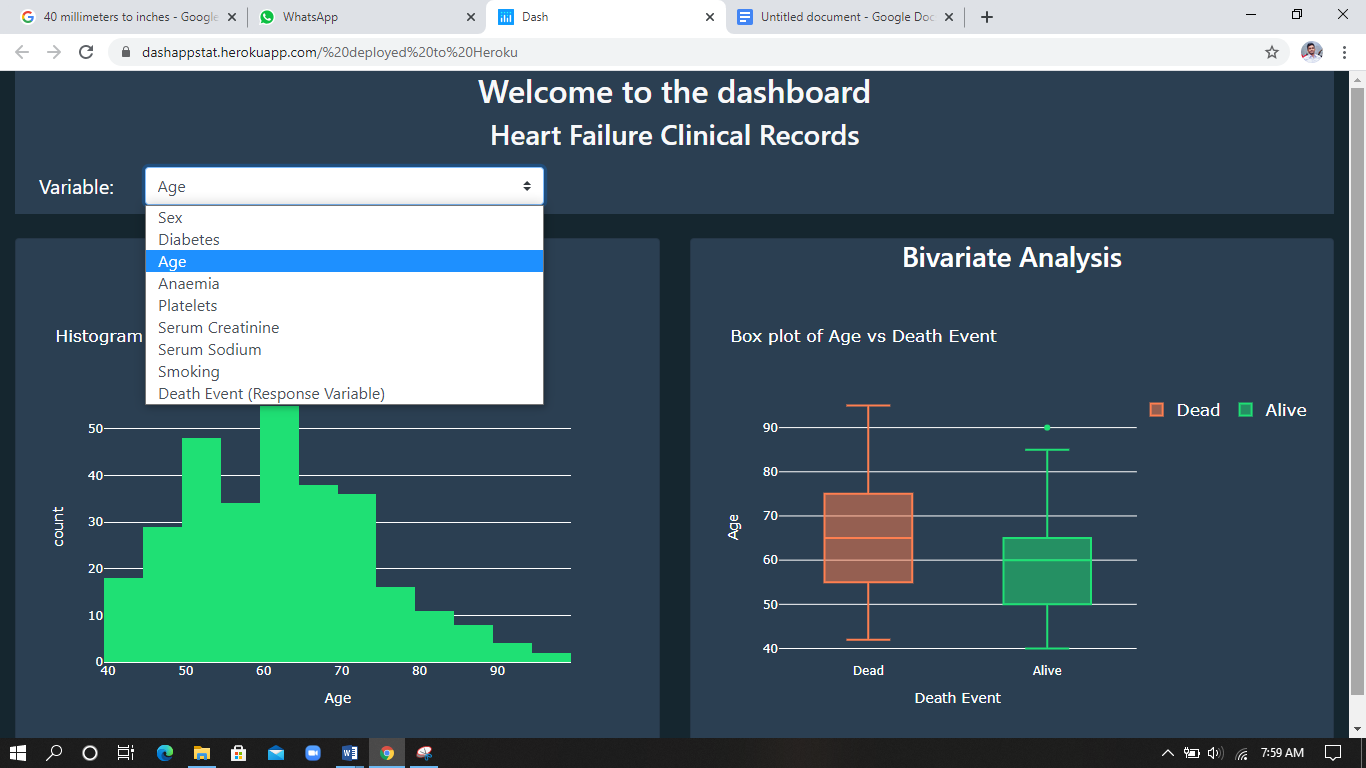


Figure 3.2. 2 Dashboard Overview 1 - Drop down menu

* **Graphs -** Control statements are used to draw the appropriate graphs according to the variable (quantitative or categorical) then the output is inserted to cards child property display the graphs in the dashboard. Plotly express is used for creating graphs.
* If the selected variable is ‘Death Event’, a pie chart for Death Event is displayed in the card in the left layer of the dashboard.
* If the selected variable is not ‘Death Event’ and the variable is categorical, matching pie chart and stacked bar chart are displayed in respectively Left and Right cards in the dashboard.
* If the selected variable is not ‘Death Event’ and the variable is quantitative, matching histogram and boxplot are displayed in respectively Left and Right cards in the dashboard.

All the style sheets which are used to customize the dashboard are created using CSS and JavaScript.

A folder named assets is created in the root of the app directory and included the custom CSS and JavaScript files in that folder. Dash will automatically serve all of the files that are included in this folder.

* **CSS** – colors, font styles, font sizes, headings, themes
* **JavaScript** –  When the variable ‘Death Event’ is selected in the menu there is only one graph to display because the variable ‘Death Event’ is the response variable of the data set. Therefore, in this case the place where bivariate graphs display should be hidden. Then the code for the above criteria is coded in a JavaScript file and saved as ‘app.js in the assets file.

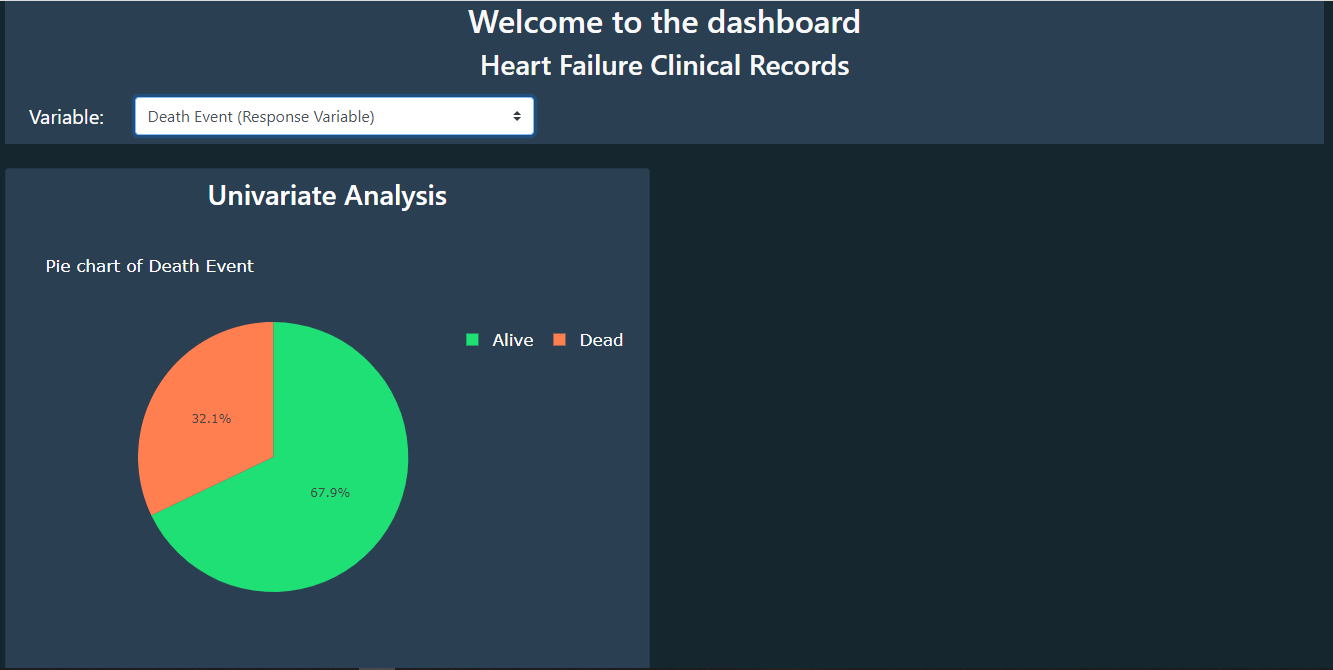
****

Figure 3.2. 3 Dashboard Overview 2

* 1. **Charts**

Four types of charts were mainly used to display the data in the dashboard.

They are pie chart, histogram, stacked bar chart and box plot.

* + 1. **Pie chart**

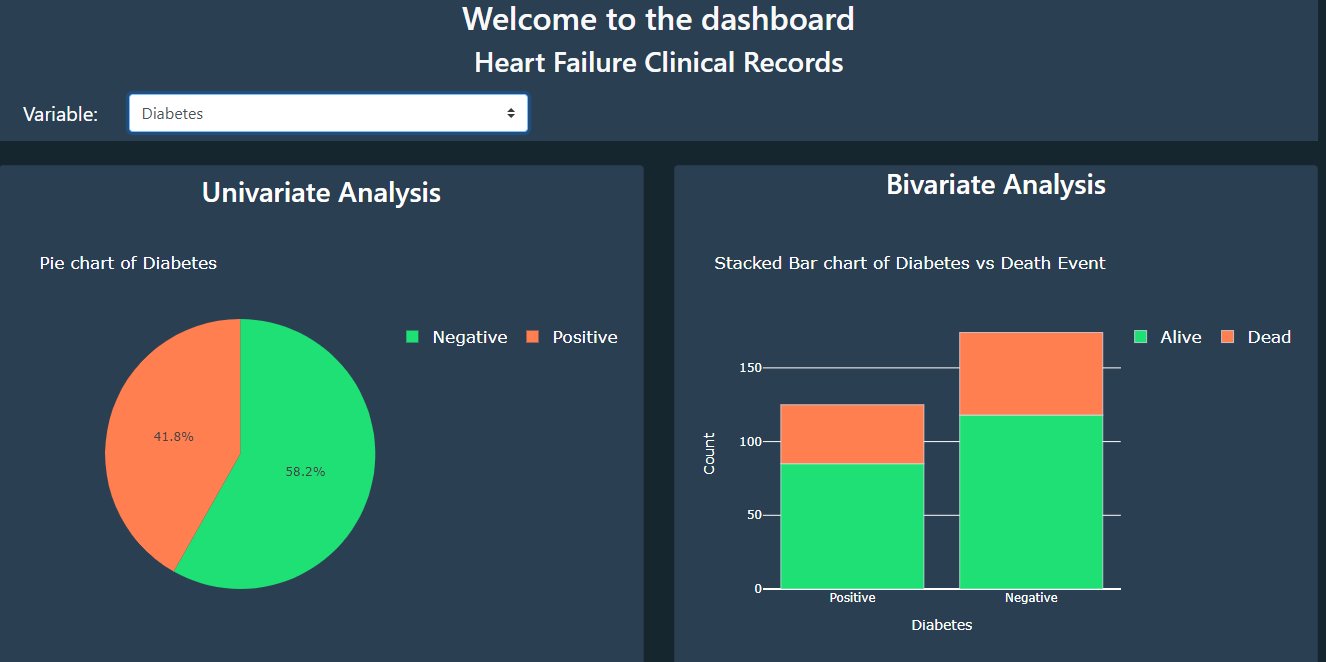
****

Figure 3.3.1. 1 Pie Chart of diabetes

There are five categorical

Variable in the data set.

Which are,

* death event
* aneamia
* diabetes
* sex
* smoking

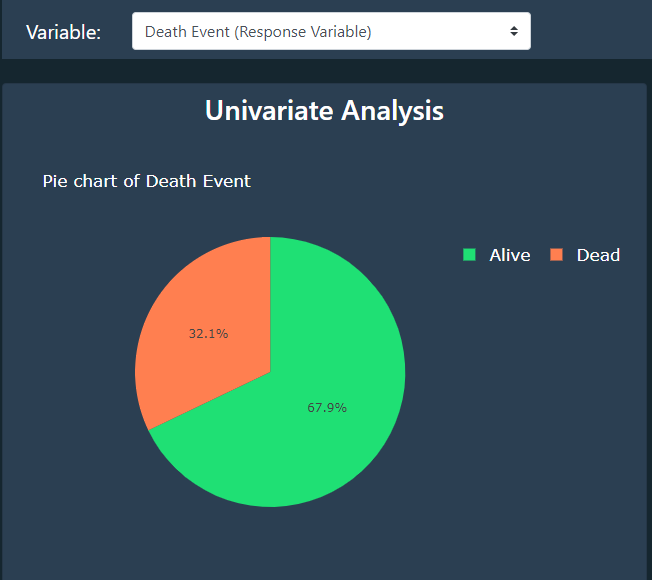


Figure 3.3.1. 2 Pie Chart of Death Event

Under univariate analysis, pie charts were implemented to display the distribution of those categorical variables.

* + 1. **Histogram**

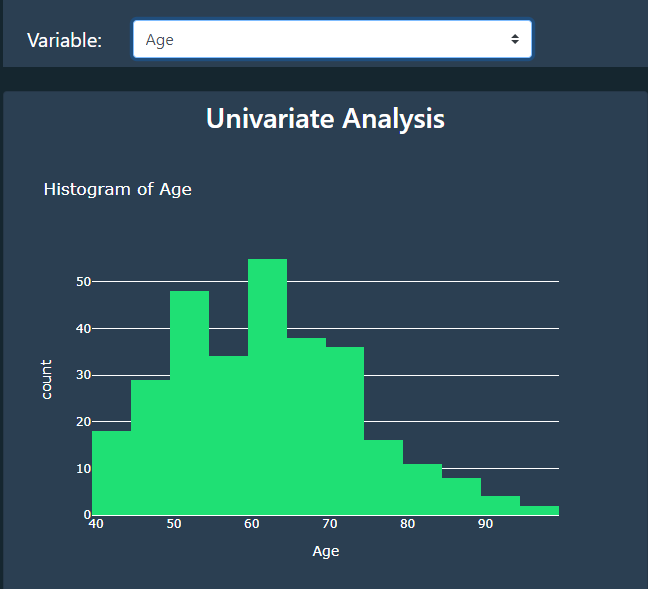
****

Figure 3.3.2. 1 Histogram of Age

There are four quantitative variables in the data set.

They are,

* age
* platelets
* serum creatinine
* serum sodium

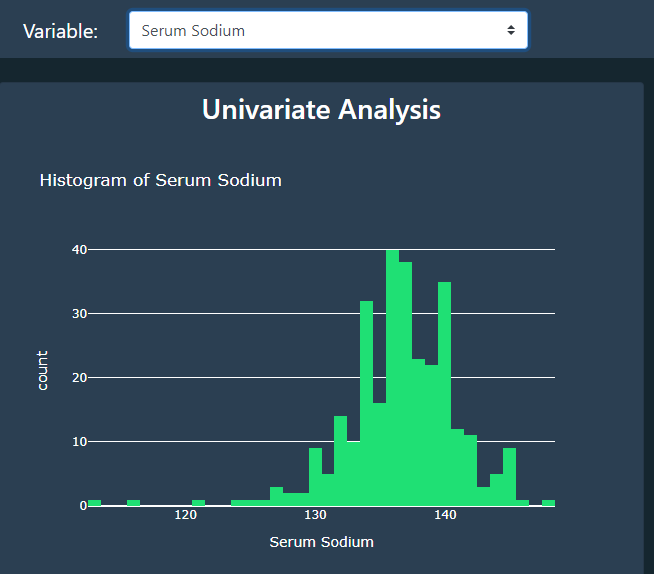
****

Figure 3.3.2. 2 Histogram of Serum Sodium

Under univariate analysis, histograms were implemented to display the distribution of those quantitative variables in the dashboard.

* + 1. **Stacked bar chart**

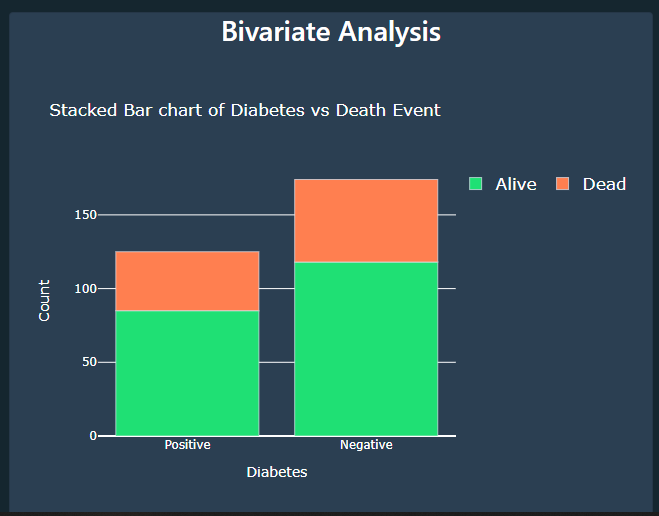
****

Figure 3.3.3. 1 Stacked bar chart of diabetes vs. death event

In bivariate analysis, all the explanatory variables were considered with the response variable (death event).

Since the response variable “death event” is a categorical variable,

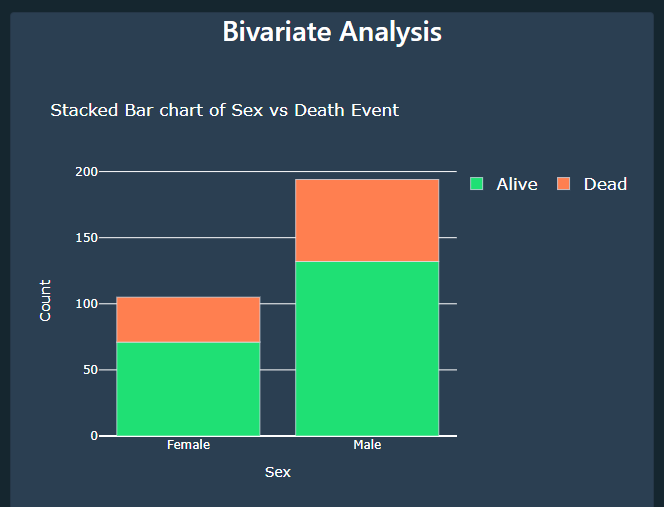
****Stacked bar charts were implemented to show the association between the response and each categorical explanatory variable. Therefore 4 stacked bar charts were implemented for the explanatory variables “anaemia”, “diabetes”, “sex” and “smoking” with respect to the response variable “death event”.

Figure 3.3.3. 2 Stacked bar chart of sex vs. death event

* + 1. **Box plot**

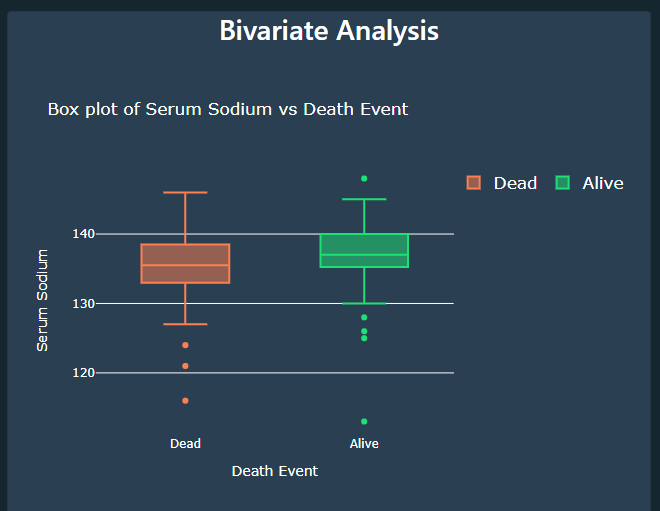
****

Figure 3.3.4. 1 Box plot of serum sodium vs. death event

In bivariate analysis, Box plots were implemented to show the association between the response and the quantitative explanatory variables in the data set.

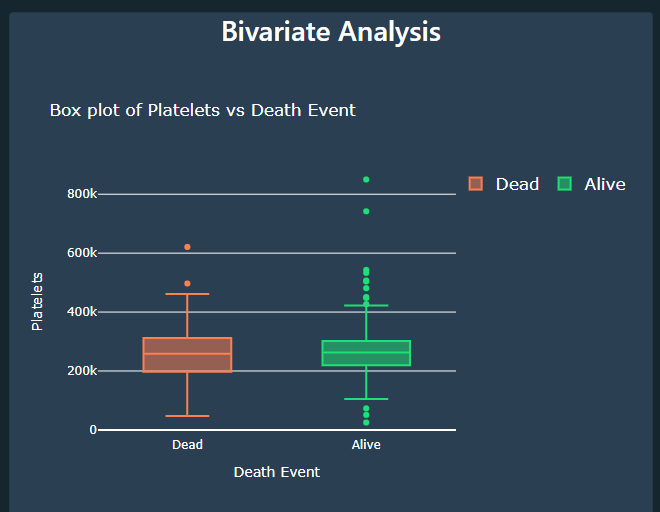
****

Figure 3.3.4. 2 Box plot of platelets vs. death events

Therefore, four box plots were implemented in the dashboard for the explanatory variables “age”, “platelets”, “serum creatinine” and “serum sodium” with respect to the response variable “death event”.

1. **Conclusion and Recommendation**
   1. **Conclusion**

* Basically, a dashboard combines charts, tables and graphs relevant to a particular data analysis on a single screen providing the reader with a big picture of the situation. The underlying dashboard represents the data relevant to death events that occurred due to heart failures.
* Even if this dashboard seems to be a simple one, it provides some informative data insights through the appropriate graphs.
* The univariate graphs indicate the distribution of each variable and bivariate graphs represent the associations between the response and each explanatory variable.
* Attractive graphical representations, higher accuracy level, self-service features ( no need for specific technical IT skills in implementing) and flexibility to change are the main advantages of this dashboard.
* Also, interactive data visualization which is presented on one screen is a major advantage of this dashboard over a lengthy statistical report.
  1. **Recommendations**
* To improve the efficiency of this dashboard and to make it more advanced;
* The number of charts displays on the screen at one time can be increased and some other types of graphs can be considered when visualizing data.
* Some descriptive statistics of variables can be presented.
* Consider adding an appropriate statistical model to predict the death events that occur due to heart failures**.**

1. **References**

* Plotly express graphs: <https://plotly.com/python/plotly-express/>
* Dash callbacks: <https://dash.plotly.com/basic-callbacks>
* Dash app Layout: <https://dash.plotly.com/layout>
* Dash Bootstrap components: <https://dash-bootstrap-components.opensource.faculty.ai/>
* App deployment: <https://dash.plotly.com/deployment>