INDIAN LIVER PATIENT DASHBOARD

Group Project

ST3011

Dr. Nadeeka Dilini Basnayake

Department of Statistics

Group A

S13825 – Chamath De Silva

S13871 – Moditha Subasinghe

S14041 – Udani Chethana

S14058 – Eranga Harshani

S14089 – Hashan Maduraarachchi

S14108 – Dinushi Ranasinghe

S14119 – Shahan Gamage

I. CONTENT

I.I TABLE OF FIGURES

1. INTRODUCTION

A data dashboard is a tool that provides a centralized, interactive means of monitoring, measuring, analyzing, and extracting relevant insights from different datasets in key areas while displaying information in an interactive, intuitive, and visual way.

In the past, creating analytical web applications was a task for seasoned developers that required knowledge of multiple programming languages and frameworks. That's no longer the case. Nowadays, can make data visualization interfaces using pure Python. One popular tool for this is Dash.

Indian liver patients dataset has been used to create the dashboard. To compare response variable with explanatory variable here used bar graphs, pie charts, multiple bar graphs. To create those charts here it has been used python libraries and edit Jupiter notebook.

2. DATASET

Indian Liver Patient Data Set (ILPD) consists of 583 observations under 10 explanatory variables and one response variable.

2.1 VARIABLES TO BE USED IN THE DASHBOARD

Explanatory variables

Variable	Description	Data type
Age	Age of the patient	Numerical
Gender	Gender of the patient	Categorical
TB	Total Bilirubin	Numerical
DB	Direct Bilirubin	Numerical
alkphos	Alkaline Phosphotase	Numerical
sgpt	Alamine Aminotransferase	Numerical
sgot	Aspartate Aminotransferase	Numerical
TP	Total Protiens	Numerical
ALB	Albumin	Numerical
A_G	Albumin and Globulin Ratio	Numerical

Response variable

Variable	Description	Data type
Class	Liver patient (1) or not a liver patient (2)	Categorical

2.2 DATA PREPARATION

The data set have 10 explanatory variables and one response variable. 9 out of 10 explanatory variables are numerical. Other one is categorical. Response variable is also categorical. There are of 583 observations in this data set.
Here 4 missing values are identified in column A_G (Albumin and Globulin Ratio). There are no any missing values in other columns. To create dashboard dropped the entire rows of missing values. After that preprocessed data set has 579 observations.

3. OBJECTIVES

- To determine the distribution of liver patients with gender.
- To determine the age distribution of the liver patients (below 18 years, 19-50 years, over 50 years).
- To find-out overall count of liver patients and non-liver patients.
- To access the liver patients along with the gender with respect to their age range
- To graphically represent whether a liver patient or not and the total count, along with the ranges in results of liver abnormality tests.

4. METHODOLOGY & FRAMEWORK

4.1 Libraries and tools used:

The dashboard is created using Python programming language which is high-level and general-purpose programming language. Dash is the main python-based framework used to create this.

The codes are created and tested using Jupyter Notebook under the desktop GUI, Anaconda navigator.

Cascading Style Sheets (CSS) has used to create the style in dashboard.

- **Anaconda**

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda distribution that allows you to launch applications and easily manage conda packages, environments, and channels without using command-line commands. Navigator can search for packages on Anaconda.org or in a local Anaconda Repository. It is available for Windows, macOS, and Linux.

- **Jupyter notebook**

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text. Uses include data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

- **Dash**

Dash is an open-source Python framework used for building analytical web applications. It is a powerful library that simplifies the development of data-driven applications. It is especially useful for Python data scientists who are not very familiar with web development. Users can create amazing dashboards in their browser using dash.

Layout of the dashboard can be build using the dash_html_components and the dash_core_components library.

- **Dash Core Components**

The dash_core_components include a set of higher-level components like dropdowns, graphs, markdown, blocks and many more.

Like all other Dash components, they are described entirely declaratively. Every option that is configurable is available as a keyword argument of the component.

- **Dash HTML Components**

The dash_html_components library contains a component class for every HTML tag as well as keyword arguments for all the HTML arguments.

- **Pandas Framework**

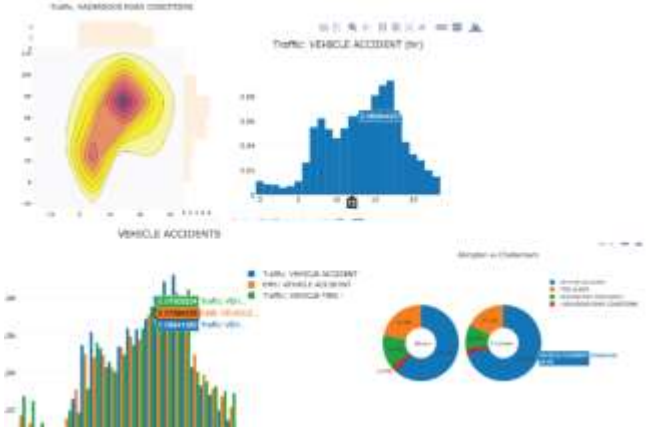
Pandas is an open-source, BSD-licensed Python library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language. Python with Pandas is used in a wide range of fields including academic and commercial domains.

- **NumPy**

NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, Fourier transform, and matrices. Arrays are very frequently used in data science, where speed and resources are very important.

- **Plotly**

Plotly allows users to import, copy and paste, or stream data to be analyzed and visualized. For analysis and styling graphs, Plotly offers a Python sandbox (NumPy supported), data grid, and GUI.



- **Plotly.Express**

Plotly Express is a new high-level Python visualization library which is a wrapper for Plotly.py that exposes a simple syntax for complex charts. It contains functions that can create entire figures at once. Every Plotly Express function uses graph objects internally and returns a plotly.graph_objects.Figure instance.

- **Dash.dependencies**

This module is used in order to make the Dash app interactive by importing input and output from dash.dependencies which is completely different from HTML input.

5. USAGE OF THE DASHBOARD

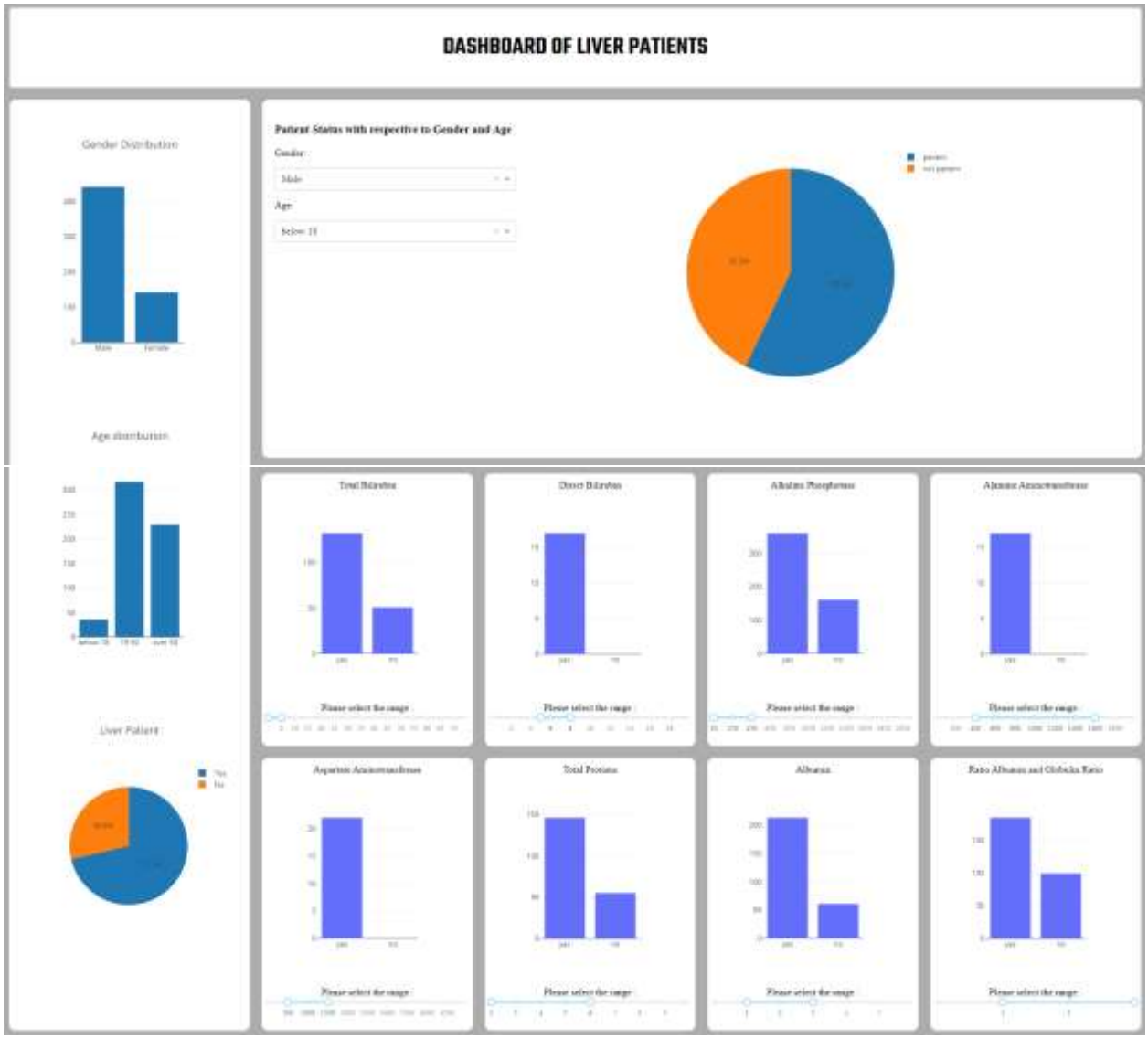


Figure 5.1

Upon observation, it can be seen that the dashboard of liver patients can be divided into 2 sub-topics; general analysis of the dashboard and advanced (further) analysis. As per the above, leftmost consists of the general analysis.

5.1 GENERAL ANALYSIS

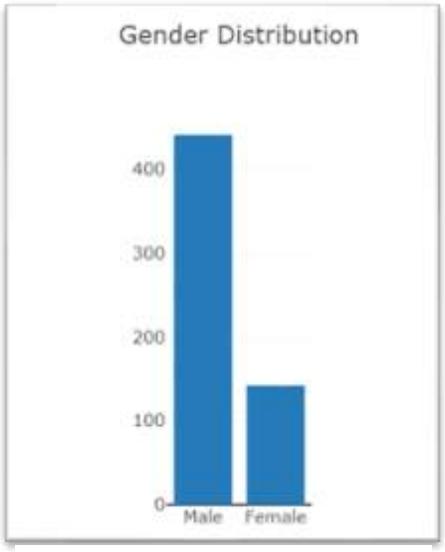


Figure 5.1.1

In this study, collected dataset is being used which contains 583 individuals which includes 441 males and 142 females.

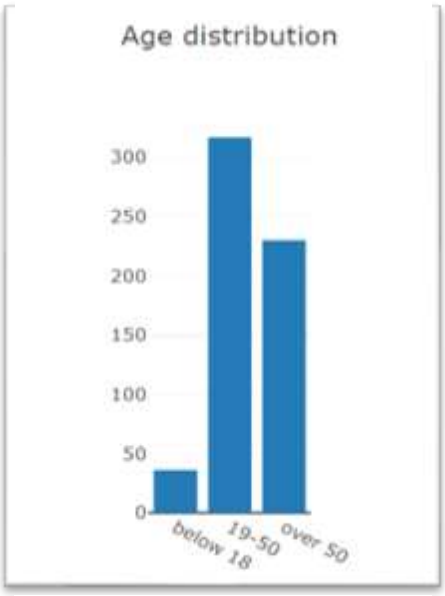


Figure 5.1.2

Bar graph of age distribution of the study shows there are 36 individuals who are below 18, 317 individuals who are between 19-50 and 230 individuals over 50 years have been used in the study.

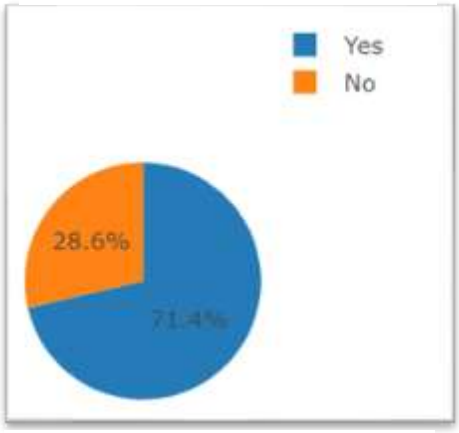


Figure 5.1.3

Pie chart of liver patients represent overview of the number of patients. According to that, 71.4% individuals who have been took part in the study are liver patients while 28.6% of individuals are not liver patients.

5.2 ADVANCED ANALYSIS

Patient Status with respective to Gender and Age

This is a demonstration of patient status with respect to Gender and Age using pie charts. Hence the impact of the gender and different age levels to the patient status can be observed through pie charts.

Two drop down boxes are used here, in order to make the selections of gender as well as different age groups.

The drop-down box for Gender is consisted of three categories namely “Male”, “Female” and “Both”. The drop-down box for Age is consisted of three categories namely “below 18”, “19-50”, “over 50” years.

Figure 5.2.1

When a selection is done by selecting one category form each drop-down boxes. Then the pie chart containing the patient status corresponding to selected information is displayed.

If no selection is done, then the default pie chart of patient status of Male below 18 years is displayed.

Consider an example: - When over 50-years category is selected along with the female category then the corresponding pie chart will be appeared on the screen as follows.

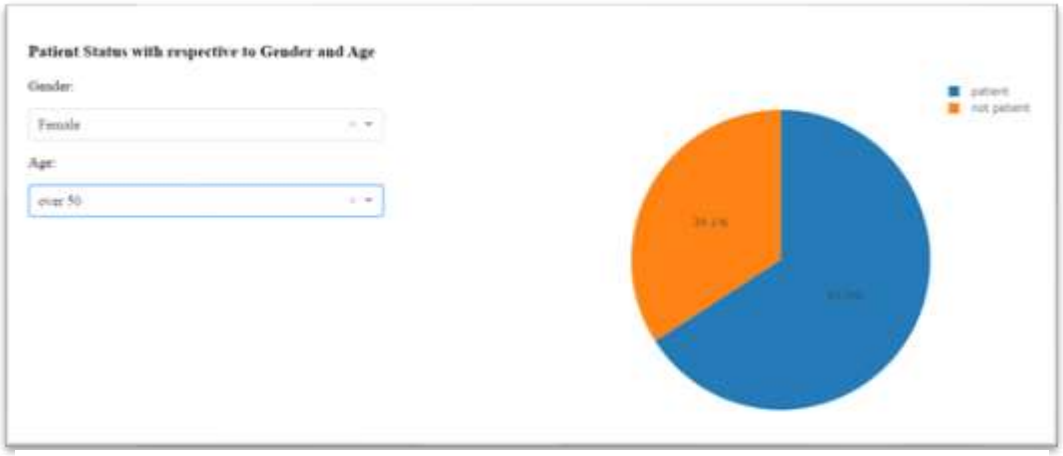


Figure 5.2.2

Figure 5.2.3

This bar graph shows the number of liver patients with respect to the variation of **Total Bilirubin (TB)**.

The y-axis represent the number of patients/ non patients and the x-axis represent whether he/she is a patient or not according to the Total Bilirubin range that can be selected from below slider.

It can select a range of TB or an exact value of TB and check the number of liver patients in that value by moving the curser on each bar.

As an example, in the range of TB 0-10, there are 366 liver patients and 167 non-liver patients as shown in the graph.

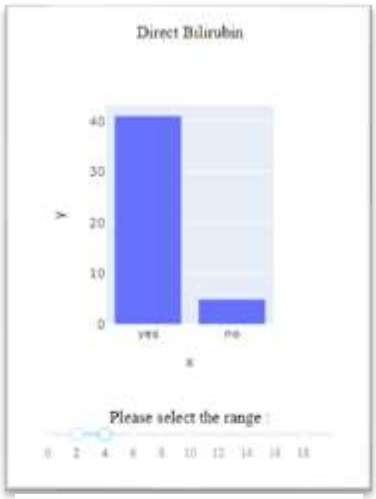


Figure 5.2.4

This bar graph shows the number of liver patients with respect to the variation of **Direct Bilirubin (DB)**.

The y-axis represent the number of patients/ non patients and the x-axis represent whether he/she is a patient or not according to the Direct Bilirubin range that can be selected from below slider.

It can select a range of DB or an exact value of TB and check the number of liver patients in that value by moving the curser on each bar.

As an example, in the range of DB 2-4, there are 41 liver patients and 5 non-liver patients as shown in the graph.

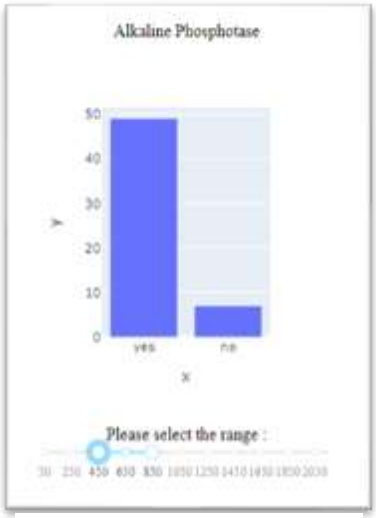


Figure 5.2.5

This bar graph shows the number of liver patients with respect to the variation of **Alkaline Phosphotase**.

The y-axis represent the number of patients/ non patients and the x-axis represent whether he/she is a patient or not according to the Alkaline Phosphotase range that can be selected from below slider.

It can select a range of Alkaline Phosphotase or an exact value of Alkaline Phosphotase and check the number of liver patients in that value by moving the curser on each bar.

As an example, in the range of Alkaline Phosphotase 450-850, there are 49 liver patients and 7 non-liver patients as shown in the graph.

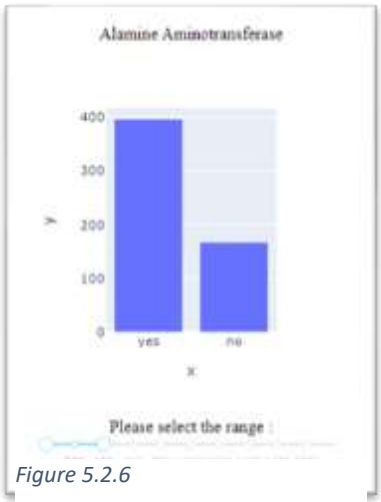


Figure 5.2.6

This bar graph shows the number of liver patients with respect to the variation of **Alamine Aminotransferase**.

The y-axis represent the number of patients/ non patients and the x-axis represent whether he/she is a patient or not according to the Alamine Aminotransferase range that can be selected from below slider. It can select a range of Alamine Aminotransferase or an exact value of Alamine Aminotransferase and check the number of liver patients in that value by moving the curser on each bar. As an example, in the range of Alamine Aminotransferase 0-400, there are 396 liver patients and 167 non-liver patients as shown in the graph.

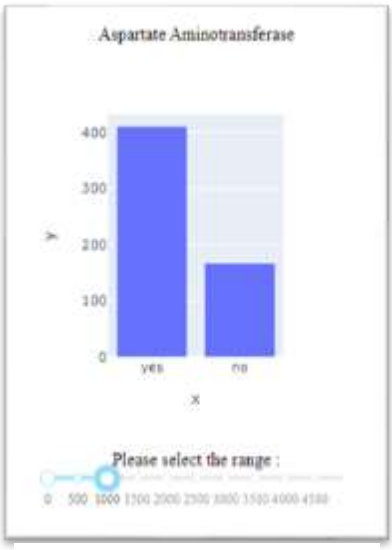


Figure 5.2.7

This bar graph shows the number of liver patients with respect to the variation of **Aspartate Aminotransferase**.

The y-axis represent the number of patients/ non patients and the x-axis represent whether he/she is a patient or not according to the Aspartate Aminotransferase range that can be selected from below slider.

It can select a range of Aspartate Aminotransferase or an exact value of Aspartate Aminotransferase and check the number of liver patients in that value by moving the curser on each bar..

As an example, in the range of Aspartate Aminotransferase 0-1000, there are 410 liver patients and 167 non-liver patients as shown in the graph.

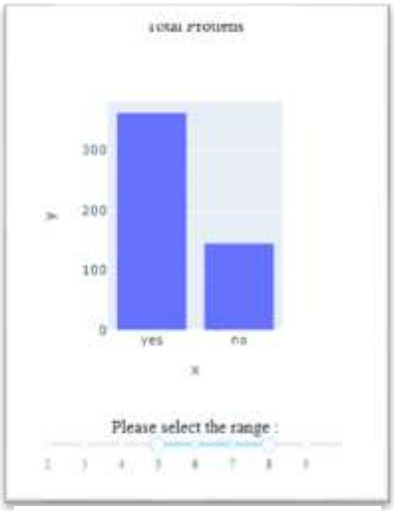


Figure 5.2.8

This bar graph shows the number of liver patients with respect to the variation of **Total Proteins**.

The y-axis represent the number of patients/ non patients and the x-axis represent whether he/she is a patient or not according to the Total Proteins range that can be selected from below slider.

It can select a range of Total Proteins or an exact value of Total Proteins and check the number of liver patients in that value by moving the curser on each bar. As an example, in the range of Total Proteins 5-8, there are 363 liver patients and 145 non-liver patients as shown in the graph.

Figure 5.2.9

This bar graph shows the number of liver patients with respect to the variation of **Albumin**.

The y-axis represent the number of patients/ non patients and the x-axis represent whether he/she is a patient or not according to the Albumin range that can be selected from below slider.

It can select a range of Albumin or an exact value of Albumin and check the number of liver patients in that value by moving the curser on each bar.

Figure 5.2.10

This bar graph shows the number of liver patients with respect to the variation of **Radio Albumin and Globulin Ratio**.

The y-axis represents the number of patients/ non patients and the x-axis represent whether he/she is a patient or not according to the Radio Albumin and Globulin Ratio range that can be selected from below slider.

It can select a range or an exact value of Radio Albumin and Globulin Ratio and check the number of liver patients in that value by moving the curser on each bar.

As an example, in the range of Radio Albumin and Globulin Ratio 1-3, there are 181liver patients and 99 non-liver patients as shown in the graph.

6. CONCLUSION

The use of dashboard will assist people who are in medical sector along with Doctors in improving their decision making process and hence better treatment level for liver patients. When used in conjunction with treatment evaluation, the key metrics can be set to monitor all important liver abnormalities tests and risk of being a liver patient. This should encourage regular review and action where required.

The dashboard report should align with implementation strategy and be easy to create, understand and explain. The dashboard also needs to be flexible to introduce new, more relevant metrics that would be find in the future to support the continual improvement of the liver diseases diagnostic process.

Dashboard should be developed with the due considerations of the needs of users.

Significant advantages to the liver abnormalities will be achieved when the dashboard is updated frequently and in a format that user can easily read. The use of visual aids is the most common form of dashboard reporting, as this provides a quick and concise presentation of key information.

Implementing the dashboard into the health care will ensure that the important metrics for treatment success are continually monitored and action upon.