# THE UNIVERSITY OF BURDWAN

**DEPARTMENT OF STATISTICS**

A Project Work On

# "Creating an Interactive Clinical Data Dashboard Using Python"

**Course Code: -** MSST 406

Submitted To The

**Department Of Statistics**

**The University Of Burdwan**

For The partial fulfillment of the Degree of M.Sc. in Statistics

By

**Xeviers Koner**

ROLL NO.: - BUR/ST/2021/026

REGISTRATION NO.: - 201801015980 Of 2018-19

M.Sc. SEMESTER: – IV EXAMINATION 2023

SESSION: – 2021-2023

# Certificate

This is to certify that **Xeviers Koner** (Roll no.- BUR/ST/2021/026) has prepared the project work entitled as "**Creating an Interactive Clinical Data Dashboard Using Python**" based on the survey of literatures in his area of interest for the partial fulfillment of the M.Sc. degree in Statistics from The University of Burdwan.

**Dated**: 14/08/2023

**Place**: Burdwan

**Dr. JAYABRATA BISWAS**

Assistant Professor

Department of Statistics

The University of Burdwan,

Burdwan

# Acknowledgement

I would like to express my sincere thanks and gratitude to my respected Assistant Professor **Dr. Jayabrata Biswas**, Department of Statistics, The University of Burdwan for his continuous guidance, support and for instilling 'Pure Thought' in carrying out of my project work. His continuous supervision, motivation and encouragement helped me a lot towards the explanation of results in understanding literature as available with me for the preparation of this project work. As a whole I have been able to give a concrete shape of this project.

I am also thankful to our respected Professor **Dr. Arindam Gupta** (HOD), Department of Statistics, The University of Burdwan for his continuous support and motivation towards the completion of my project work.

I also express my sincere thanks and gratitude to all my department professors Dr. Rabindranath Das, Dr. Ayan Pal and Prof. Pritam Sarkar for their constant encouragement and inspiration in carrying out my project work.

I am also thankful to the Librarian and staff members of the departmental library for their help and cooperation in searching books.

Also, I have taken the help of literatures as available in websites and e-books as cited in the references from where I quoted the results with a proper understanding to my knowledge and ability and with proper explanations and justifications to my capability.

# Table of Content

# Creating an Interactive Clinical Data Dashboard Using Python

## Introduction:

The primary objective of this project is to design and implement an interactive clinical data dashboard using the Python programming language. This dashboard aims to provide a user-friendly and dynamic platform for analyzing and visualizing clinical data. By leveraging Python's data manipulation and visualization capabilities, coupled with interactive features, this project seeks to enhance the accessibility and comprehensibility of complex clinical data. Through this project, we aim to address the growing need for efficient data analysis tools in healthcare and clinical research. The interactive dashboard will empower healthcare professionals, researchers, and stakeholders to gain valuable insights from clinical data, facilitating evidence-based decision-making and enhancing patient care. The integration of Python's libraries and frameworks, along with the utilization of dummy clinical data representative of real-world scenarios, will allow us to showcase the potential impact and benefits of such a dashboard in various clinical contexts. By the end of this project, we envision a tangible demonstration of how interactive data visualization can contribute to informed clinical practices and contribute to advancements in the field. In modern healthcare, the significance of data visualization and interactive tools cannot be overstated.

Data visualization transforms raw data into graphical representations that are intuitive to interpret. By rendering complex datasets into charts, graphs, and maps, data visualization allows for rapid identification of trends, correlations, and anomalies. Moreover, the interactivity aspect enables dynamic exploration, encouraging users to delve deeper into the data and uncover valuable insights.

## Importance of Data Visualization and Interactive Tools in Healthcare and Clinical Research:

Data visualization and interactive tools play a pivotal role in the healthcare and clinical research domains. The utilization of these tools goes beyond presenting raw data

I. **Enhanced Data Understanding:** Raw clinical data can be extensive and intricate, making it challenging to discern meaningful patterns and relationships. Data visualization transforms this data into visual representations such as graphs, charts, and maps. Interactive tools take this a step further by allowing users to manipulate and drill down into the data, uncovering hidden insights that might otherwise be overlooked.
II. **Improved Communication:** In healthcare settings, effective communication is paramount. Data visualization provides a universal language that simplifies the communication of complex medical information between healthcare professionals, patients, and even non-expert stakeholders. Interactive dashboards enable real-time discussions around data, fostering collaborative decision-making and ensuring everyone is on the same page.
III. **Transparency and Accountability:** Transparent data visualization fosters trust among stakeholders by providing clear and accessible insights. Interactive tools allow for

real-time tracking of healthcare metrics, ensuring accountability for quality of care, safety measures, and regulatory compliance.

IV. **Evidence-Based Decision-Making:** In both clinical practice and research, decisions need to be grounded in evidence. Data visualization allows for a visual representation of outcomes, treatment efficacy, and patient demographics, assisting clinicians in choosing the most effective interventions. Interactive tools enable dynamic exploration of data subsets, facilitating personalized treatment strategies based on patient-specific attributes.

## ❖ Objectives:

The primary objective of this project is to design and develop an interactive clinical data dashboard using Python, focused on analyzing and visualizing demographic and health-related data. The dashboard will provide a comprehensive and user-friendly interface for healthcare professionals and researchers to gain valuable insights from the provided clinical dataset. The specific objectives of this project are as follows:

1. **Creation of Group Variables:** Create new group variables for "AGE" and "Average Income" based on the available data. This involves segmenting the dataset to allow for meaningful analysis and visualization.
2. **Interactive Dashboard User Interface (UI):** Develop an interactive dashboard with a user interface (UI) that facilitates dynamic data exploration and visualization. The UI will serve as a platform for users to interact with the data through various filtering and visualization options.
3. **Implementation of Filtering Fields:** Incorporate filtering fields into the dashboard, including "Input Data," "Date," "Drug group," "Disease code," "Sex," "Age group," and "Smoking status." These fields will enable users to filter the data based on specific criteria, facilitating targeted analysis.
4. **Visualizations:** The following visualizations of my project are
    i. **Disease Code and Age Relationship:** Visualize the relationship between "Disease code" and "Age" using a bar plot. This visualization will provide insights into the distribution of diseases across different age groups.
    ii. **Drug Group and Disease Code Relationship:** Create a bar plot that illustrates the relationship between "Drug group" and "Disease code." This visualization will shed light on the connection between drug usage and specific diseases.
    iii. **Smoking Status and Cancer Affection:** Develop a stacked bar chart that depicts the relationship between "Smoking status" and the likelihood of being affected by cancer. This visualization will highlight the impact of smoking on cancer incidence.
    iv. **Age Distribution Visualization:** Generate a histogram to visualize the distribution of ages within the dataset. This visualization will provide insights into the age demographics of the patients.
    v. **Age and Average Income:** Create a bubble chart that visualizes the relationship between "Age" and "Average Income." This visualization will help identify any potential trends or patterns between age and income.
    vi. **Sex Distribution:** Visualize the distribution of "Sex" using a pie chart. This visualization will offer a clear representation of the gender distribution within the dataset.

## ❖ What is an Interactive Dashboard?

An interactive dashboard is a digital tool designed to present complex data in a dynamic and user-friendly manner. Unlike static visualizations, an interactive dashboard allows users to actively engage with the data, explore various perspectives, and derive insights through real-time manipulation of visual elements. It serves as a bridge between raw data and actionable insights, offering a seamless and intuitive interface for data analysis and decision-making.

**Definition and Explanation:** An interactive dashboard combines data visualization with interactivity, enabling users to navigate through visual representations and explore relationships within the data. It typically consists of multiple interactive components such as charts, graphs, maps, filters, sliders, and dropdowns. Users can adjust parameters, apply filters, and interact with graphical elements to reveal different dimensions of the data.

**Role in Presenting Complex Data:** The complexity of clinical data, often characterized by numerous variables and intricate relationships, can pose challenges in understanding and interpretation. An interactive dashboard addresses this challenge by visually summarizing the data, simplifying its complexity, and highlighting patterns and trends.

**Importance of Interactivity:** Interactivity lies at the heart of an interactive dashboard's effectiveness. It empowers users to tailor their exploration according to their specific needs and inquiries. By allowing users to filter, zoom, and drill down into data points, interactivity enables user-driven exploration.

**Dynamic Data Visualization:** Dynamic data visualization is a core feature of an interactive dashboard. It goes beyond static charts to provide a fluid experience where visualizations respond to user inputs in real-time. Users can change parameters, filter data subsets, and witness the immediate impact on visual outputs.

By harnessing the power of interactivity, the dashboard will bridge the gap between data complexity and user comprehension, ultimately contributing to informed decision-making in healthcare and clinical research.

## ❖ Literature Review:

**Importance of Data Visualization in Healthcare and Clinical Decision-Making:** Data visualization plays a crucial role in healthcare and clinical decision-making by transforming complex data into visually understandable insights. Visual representations simplify the communication of medical information to both expert and non-expert audiences, enhancing the accuracy of diagnoses, treatment plans, and patient education.

**Overview of Python Libraries for Data Visualization and Dashboard Creation:** Python has emerged as a powerful programming language for data visualization and dashboard creation. Libraries such as pandas, NumPy, and pandasql facilitate data manipulation, transformation, and analysis. More notably, Python's visualization libraries, including Matplotlib

enable the creation of a wide range of static visualizations. However, the dynamic nature of interactive dashboards necessitates specialized tools.

One such tool is Dash, a Python framework designed for building interactive web applications. Dash combines Python's data manipulation capabilities with Plotly's interactive visualizations. This framework enables developers to create responsive and user-friendly dashboards with minimal coding effort. Dash's capabilities extend beyond basic graphs to include complex data filtering, dynamic data updates, and interactive user interfaces.

## ❖ Data Description:

The dataset utilized for this project is a proprietary non-shareable dummy clinical dataset obtained from an alumnus working within an organization. Due to data privacy considerations, the dataset's specifics cannot be disclosed or shared externally. However, a general overview of the dataset's structure and attributes is provided below to contextualize the project's scope.

**Dataset Format and Attributes:** The dataset is formatted as a CSV (Comma-Separated Values) file, a common and widely used format for tabular data. Each row in the dataset corresponds to a unique entry, while each column represents a distinct attribute or variable related to clinical information. The following attributes are present in the dataset:

a) **ID:** A unique identifier assigned to each data entry.
b) **AGE:** The age of the individual at the time of data collection.(from 18 to 90)
c) **SEX:** The gender of the individual, categorized as male or female.(male=M, female=F)
d) **Smoking Status:** The smoking status of the individual, indicating whether they are a smoker or non-smoker. (0=non smoker, 1=current smoker, 2=smoke previously)
e) **Date:** The date of data collection, providing a temporal dimension to the dataset.
f) **Disease Code:** A code representing the specific disease or medical condition associated with the individual. There are 8 disease codes-ABCDEF1, ABCDEF2, ABCDEF3, ABCDEF4, ABCDEF5, ABCDEF6, ABCDEF7, ABCDEF8.
g) **Treatment Category:** The category under which the treatment falls, providing insight into the type of medical intervention. There are 3 treatment catagories-T1,T2,T3.
h) **Drug Group:** The group to which a specific drug or medication belongs. There are 4 drug groups-DRUG1, DRUG2, DRUG3, DRUG4.
i) **Average Income:** The average income level associated with the individual. (from 20000-100000).
j) **Affected to Cancer:** A binary variable indicating whether the individual has been affected by cancer. This is in binary format. (0=affected, 1=non affected).

## ❖ Data Preparation and Exploration:

Data preparation and exploration are fundamental stages in any data analysis project. In this section, we detail the process undertaken to prepare and structure the dummy clinical dataset, followed by an overview of the initial exploration conducted to glean insights from the data.

**Data Import and Group Variable Creation:** The project commenced with the importation of the dataset from an Excel file into the Jupyter Notebook environment. The pandas library, a

versatile tool for data manipulation and analysis, facilitated this importation. Subsequently, the pandasql library was employed to introduce the "sqldf" function, enabling SQL-like queries on the pandas Data Frame. Group variables were established to facilitate data segmentation for meaningful analysis. The "Age" variable was categorized into distinct age groups, such as '18-30', '31-43', '44-56', '57-69', '70-82', and '83-95'. Additionally, a similar approach was taken for the "Average Income" variable, yielding groups such as '20000-30000', '30001-40000', '40001-50000', and so forth.

**Python output:**

| | ID | AGE | SEX | Smoking status | date | disease code | Treatment Category | Drug group | Average Income | Affected to cancer |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | AA1 | 88 | M | 2 | 01-03-2012 | ABCDEF1 | T1 | DRUG1 | 47583 | 0 |
| **1** | BB1 | 69 | M | 1 | 26-09-2008 | ABCDEF2 | T2 | DRUG2 | 43495 | 1 |
| **2** | CC1 | 47 | M | 1 | 17-08-2014 | ABCDEF3 | T3 | DRUG3 | 61606 | 0 |
| **3** | AA2 | 47 | M | 0 | 18-03-2011 | ABCDEF4 | T1 | DRUG4 | 96705 | 0 |
| **4** | BB2 | 47 | M | 2 | 09-09-2013 | ABCDEF5 | T2 | DRUG1 | 87781 | 1 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **194** | CC65 | 26 | M | 0 | 23-02-2011 | ABCDEF7 | T3 | DRUG2 | 92222 | 0 |
| **195** | AA66 | 50 | F | 2 | 14-02-2011 | ABCDEF8 | T1 | DRUG2 | 93013 | 1 |
| **196** | BB66 | 40 | F | 2 | 24-04-2013 | ABCDEF1 | T2 | DRUG3 | 24309 | 0 |
| **197** | CC66 | 66 | F | 1 | 03-10-2011 | ABCDEF8 | T3 | DRUG4 | 57954 | 1 |
| **198** | AA67 | 58 | F | 0 | 08-06-2013 | ABCDEF1 | T1 | DRUG1 | 24926 | 0 |

199 rows × 10 columns

**Data Ordering:** A significant organizational step involved ordering the main dataset based on "AgeGroup" and "IncomeGroup." This arrangement streamlined the subsequent analysis andvisualization, aiding in the identification of patterns and trends within specific age and income cohorts.

**Python Output:**

| | ID | AGE | SEX | Smoking status | date | disease code | Treatment Category | Drug group | Average Income | Affected to cancer | AgeGroup | IncomeGroup |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | BB14 | 30 | M | 1 | 01-08-2014 | ABCDEF1 | T2 | DRUG2 | 23703 | 0 | 18-30 | 20000-30000 |
| **1** | BB56 | 24 | M | 2 | 20-01-2009 | ABCDEF1 | T2 | DRUG2 | 23318 | 1 | 18-30 | 20000-30000 |
| **2** | AA58 | 30 | M | 0 | 15-02-2014 | ABCDEF6 | T1 | DRUG1 | 24631 | 1 | 18-30 | 20000-30000 |
| **3** | BB30 | 40 | M | 2 | 21-09-2013 | ABCDEF5 | T2 | DRUG1 | 27729 | 0 | 31-43 | 20000-30000 |
| **4** | AA31 | 39 | F | 2 | 19-05-2013 | ABCDEF7 | T1 | DRUG3 | 21697 | 1 | 31-43 | 20000-30000 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **194** | AA63 | 71 | M | 1 | 07-04-2011 | ABCDEF7 | T1 | DRUG4 | 93394 | 1 | 70-82 | 90001-100000 |
| **195** | BB3 | 84 | M | 0 | 01-01-2013 | ABCDEF8 | T2 | DRUG4 | 90774 | 1 | 83-95 | 90001-100000 |
| **196** | AA22 | 86 | F | 1 | 18-06-2013 | ABCDEF8 | T1 | DRUG2 | 95379 | 1 | 83-95 | 90001-100000 |
| **197** | CC27 | 84 | F | 2 | 23-02-2011 | ABCDEF5 | T3 | DRUG4 | 98885 | 0 | 83-95 | 90001-100000 |
| **198** | BB53 | 85 | F | 2 | 18-06-2013 | ABCDEF8 | T2 | DRUG4 | 95192 | 1 | 83-95 | 90001-100000 |

199 rows × 12 columns

## ❖ Dashboard Design and Implementation:

The interactive clinical data dashboard was created using the Dash framework, a Python library renowned for its capability to develop interactive web applications for data visualization and analysis. The implementation encompasses the layout design, structure, and integration of various interactive components.

**Choice of Python Libraries/Frameworks:** The chosen framework for this dashboard is Dash, which seamlessly combines Python's data manipulation capabilities with Plotly's interactive visualizations. Dash's versatility and user-friendly syntax make it an ideal choice for building interactive web applications. The integration of Plotly Express (px) further enhances the visualization capabilities.

**Dashboard User Interface (UI):** The dashboard UI was meticulously structured using HTML and Dash components. The dashboard UI incorporates various visual elements, including headers, dropdowns, date pickers, buttons, and interactive graphs. The dashboard commences with a welcoming title and includes a tabbed structure that holds the primary data input tab.

**Interactive Components:** The dashboard presents users with multiple interactive components to facilitate data exploration:

- **File Upload:** Users can upload data files through a user-friendly drag-and-drop or selection interface. This step enables dynamic data input for analysis.
- **Dropdowns and Checklists:** Interactive dropdowns and checkboxes allow users to select parameters such as "Drug Group," "Disease Code," "Sex," "Age Group," and "Smoking Status." These dropdowns offer users the flexibility to customize their data visualization based on specific criteria.
- **Date Picker Range:** The date picker range enables users to specify a time period for data analysis. This feature supports time-based analysis and insights over specific intervals.
- **Filter Button:** The filter button serves as a trigger to apply selected filters and update the visualizations according to the chosen criteria.

**Visualization Types:** The dashboard incorporates a range of interactive visualization types to convey insights effectively:

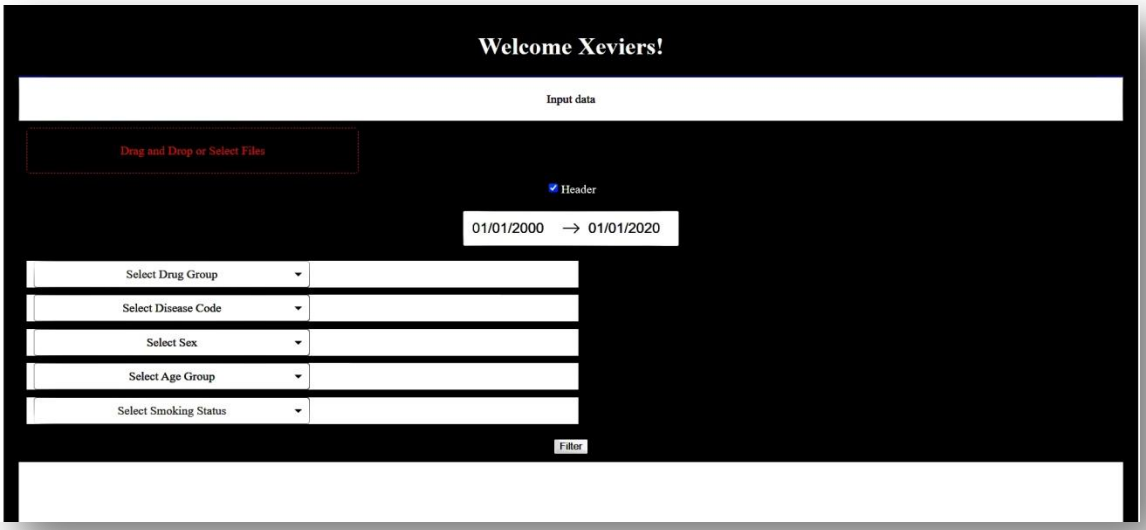
- **Bar Plots:** The dashboard includes two types of bar plots - one depicting the relationship between "Disease Code" and "Age", and another illustrating the correlation between "Drug Group" and "Disease Code."
- **Stacked Bar Chart:** The dashboard features a stacked bar chart that visualizes the interaction between "Smoking Status" and "Affected to Cancer." This graph enables users to observe the distribution of cancer cases based on smoking status.
- **Histogram:** A histogram portrays the distribution of ages within the dataset. This visualization aids in understanding the age demographics of the patients.
- **Bubble Chart:** The bubble chart demonstrates the connection between "Age" and "Average Income" while incorporating "Smoking Status" as bubble size. This interactive visualization enhances the identification of patterns and trends.
- **Box Plot:** A box plot illustrates the relationship between "Drug Group" and "Average Income," showcasing distribution and outliers in the data.
- **Pie Chart:** The pie chart showcases the distribution of "Sex" within the dataset, providing a clear gender-based representation.

Through the utilization of Dash, Python, and Plotly Express, this interactive clinical data dashboard successfully facilitates data exploration and visualization.

### ❖ Data Filtering:

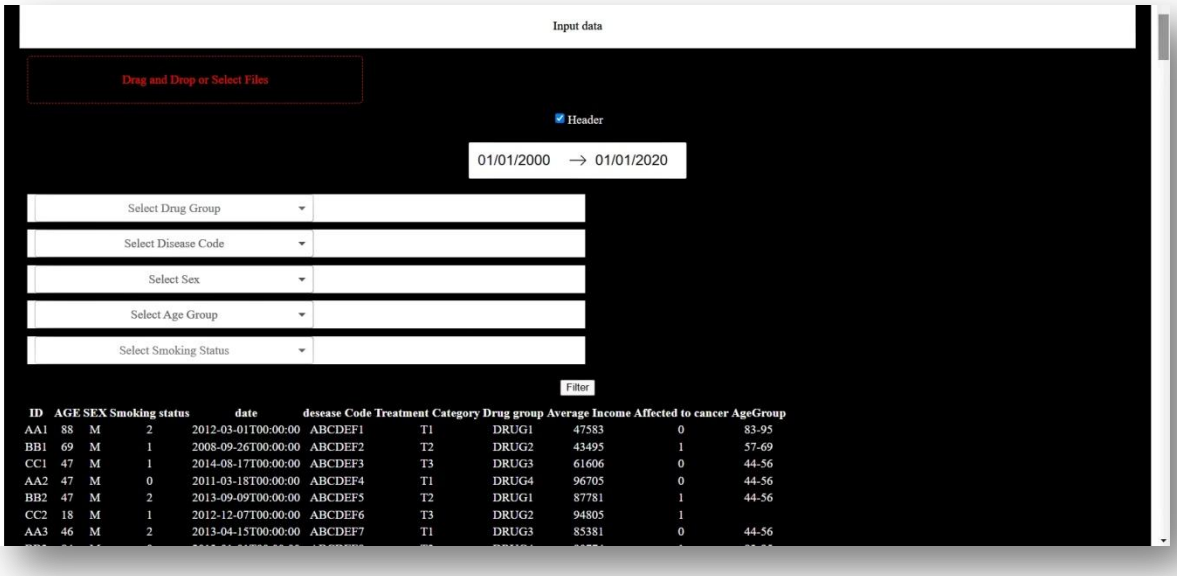
In this section we will show that the user will input the data and the user can filter the data corresponding with different dropdowns and the filtered data will be shown in our dashboard.

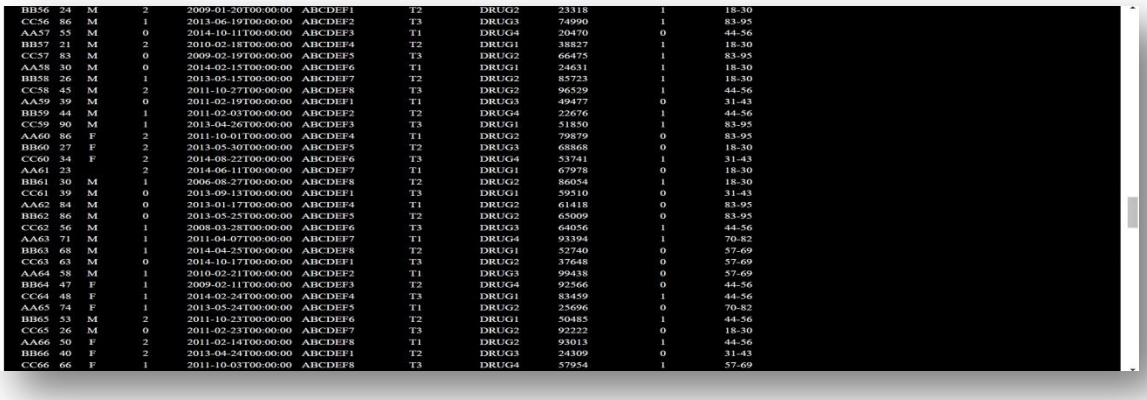
## Dashboard Screenshots and Illustrations:




This is our dashboard User Interface (UI). Here is a “Welcome Xeviers!” command at the top of our dashboard. Then we have “Drag or Drop or Select Files” option in red color, we have a date range where the user can chose date. Then we have interactive dropdowns such as: “Select Drug Group”, “Select Disease Code”, “Select Sex”, “Select Age group”, “Select Smoking Status” where the user can filter the data according to these dropdowns.

## Drag and Drop or Select Files:

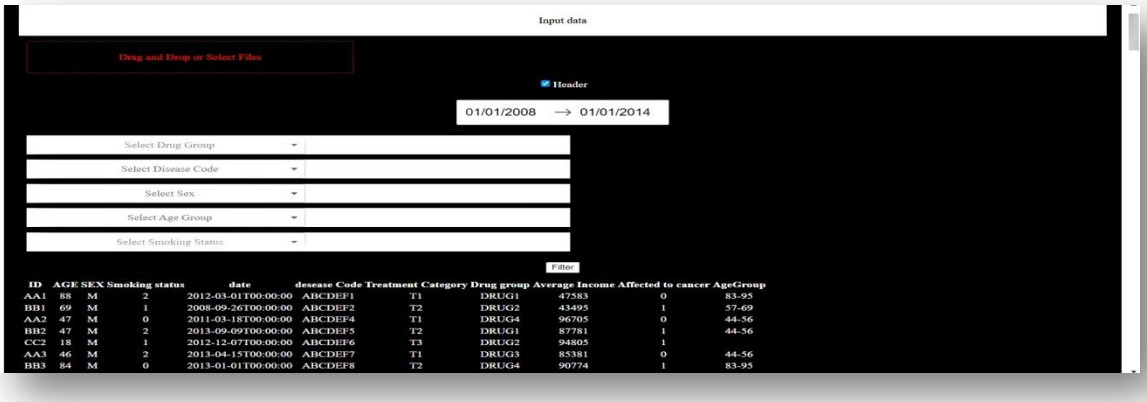




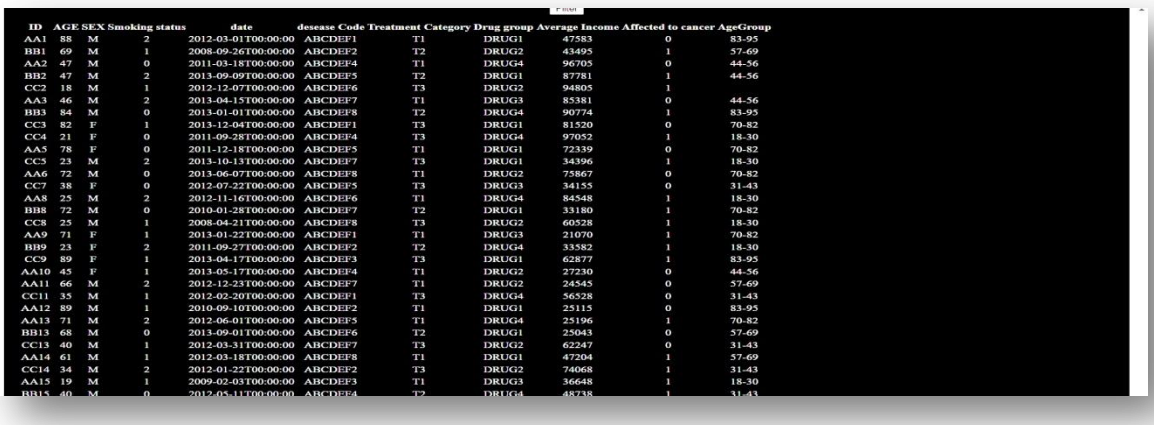
| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| BB56 | 24 | M | 2 | 2009-01-20T00:00:00 | ABCDEF1 | T2 | DRUG2 | 23318 | 1 | 18-30 | |
| CC56 | 86 | M | 1 | 2013-06-19T00:00:00 | ABCDEF2 | T3 | DRUG3 | 74990 | 1 | 83-95 | |
| AA57 | 55 | M | 0 | 2014-10-11T00:00:00 | ABCDEF3 | T1 | DRUG4 | 20470 | 0 | 44-56 | |
| BB57 | 21 | M | 2 | 2010-02-18T00:00:00 | ABCDEF4 | T2 | DRUG1 | 38827 | 1 | 18-30 | |
| CC57 | 83 | M | 0 | 2009-02-19T00:00:00 | ABCDEF5 | T3 | DRUG2 | 66475 | 1 | 83-95 | |
| AA58 | 30 | M | 0 | 2014-02-15T00:00:00 | ABCDEF6 | T1 | DRUG1 | 24631 | 1 | 18-30 | |
| BB58 | 26 | M | 1 | 2013-05-15T00:00:00 | ABCDEF7 | T2 | DRUG2 | 85723 | 1 | 18-30 | |
| CC58 | 45 | M | 2 | 2011-10-27T00:00:00 | ABCDEF8 | T3 | DRUG2 | 96529 | 1 | 44-56 | |
| AA59 | 39 | M | 0 | 2011-02-19T00:00:00 | ABCDEF1 | T1 | DRUG3 | 49477 | 0 | 31-43 | |
| BB59 | 44 | M | 1 | 2011-02-03T00:00:00 | ABCDEF2 | T2 | DRUG4 | 22676 | 1 | 44-56 | |
| CC59 | 90 | M | 1 | 2013-04-26T00:00:00 | ABCDEF3 | T3 | DRUG1 | 51850 | 1 | 83-95 | |
| AA60 | 86 | F | 2 | 2011-10-01T00:00:00 | ABCDEF4 | T1 | DRUG2 | 79879 | 0 | 83-95 | |
| BB60 | 27 | F | 2 | 2013-05-30T00:00:00 | ABCDEF5 | T2 | DRUG3 | 68868 | 0 | 18-30 | |
| CC60 | 34 | F | 2 | 2014-08-22T00:00:00 | ABCDEF6 | T3 | DRUG4 | 53741 | 1 | 31-43 | |
| AA61 | 23 | | 2 | 2014-06-11T00:00:00 | ABCDEF7 | T1 | DRUG1 | 67978 | 0 | 18-30 | |
| BB61 | 30 | M | 1 | 2006-08-27T00:00:00 | ABCDEF8 | T2 | DRUG2 | 86054 | 1 | 18-30 | |
| CC61 | 39 | M | 0 | 2013-09-13T00:00:00 | ABCDEF1 | T3 | DRUG1 | 59510 | 0 | 31-43 | |
| AA62 | 84 | M | 0 | 2013-01-17T00:00:00 | ABCDEF4 | T1 | DRUG2 | 61418 | 0 | 83-95 | |
| BB62 | 86 | M | 0 | 2013-05-25T00:00:00 | ABCDEF5 | T2 | DRUG2 | 65009 | 0 | 83-95 | |
| CC62 | 56 | M | 1 | 2008-03-28T00:00:00 | ABCDEF6 | T3 | DRUG3 | 64056 | 1 | 44-56 | |
| AA63 | 71 | M | 1 | 2011-04-07T00:00:00 | ABCDEF7 | T1 | DRUG4 | 93394 | 1 | 70-82 | |
| BB63 | 68 | M | 1 | 2014-04-25T00:00:00 | ABCDEF8 | T2 | DRUG1 | 52740 | 0 | 57-69 | |
| CC63 | 63 | M | 0 | 2014-10-17T00:00:00 | ABCDEF1 | T3 | DRUG2 | 37648 | 0 | 57-69 | |
| AA64 | 58 | M | 1 | 2010-02-21T00:00:00 | ABCDEF2 | T1 | DRUG3 | 99438 | 0 | 57-69 | |
| BB64 | 47 | F | 1 | 2009-02-11T00:00:00 | ABCDEF3 | T2 | DRUG4 | 92566 | 0 | 44-56 | |
| CC64 | 48 | F | 1 | 2014-02-24T00:00:00 | ABCDEF4 | T3 | DRUG1 | 83459 | 1 | 44-56 | |
| AA65 | 74 | F | 1 | 2013-05-24T00:00:00 | ABCDEF5 | T1 | DRUG2 | 25696 | 0 | 70-82 | |
| BB65 | 53 | M | 2 | 2011-10-23T00:00:00 | ABCDEF6 | T2 | DRUG1 | 50485 | 1 | 44-56 | |
| CC65 | 26 | M | 0 | 2011-02-23T00:00:00 | ABCDEF7 | T3 | DRUG2 | 92222 | 0 | 18-30 | |
| AA66 | 50 | F | 2 | 2011-02-14T00:00:00 | ABCDEF8 | T1 | DRUG2 | 93013 | 1 | 44-56 | |
| BB66 | 40 | F | 2 | 2013-04-24T00:00:00 | ABCDEF1 | T2 | DRUG3 | 24309 | 0 | 31-43 | |
| CC66 | 66 | F | 1 | 2011-10-03T00:00:00 | ABCDEF8 | T3 | DRUG4 | 57954 | 1 | 57-69 | |

Here we have dropped a dummy data in the “Drag and Drop or Select Files” option and the data is showing in our dashboard.

## Data filtering by Date:



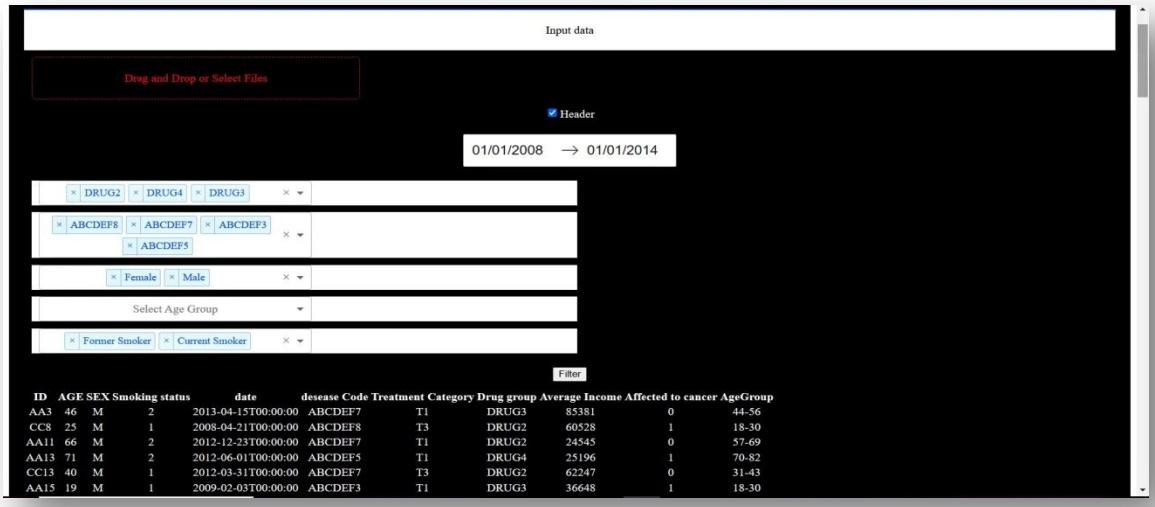

| ID | AGE | SEX | Smoking status | date | desease Code | Treatment Category | Drug group | Average Income | Affected to cancer | AgeGroup |
|---|---|---|---|---|---|---|---|---|---|---|
| AA1 | 88 | M | 2 | 2012-03-01T00:00:00 | ABCDEF1 | T1 | DRUG1 | 47583 | 0 | 83-95 |
| BB1 | 69 | M | 1 | 2008-09-26T00:00:00 | ABCDEF2 | T2 | DRUG2 | 43495 | 1 | 57-69 |
| AA2 | 47 | M | 0 | 2011-03-18T00:00:00 | ABCDEF4 | T1 | DRUG4 | 96705 | 0 | 44-56 |
| BB2 | 47 | M | 2 | 2013-09-09T00:00:00 | ABCDEF5 | T2 | DRUG1 | 87781 | 1 | 44-56 |
| CC2 | 18 | M | 1 | 2012-12-07T00:00:00 | ABCDEF6 | T3 | DRUG2 | 94805 | 1 | |
| AA3 | 46 | M | 2 | 2013-04-15T00:00:00 | ABCDEF7 | T1 | DRUG3 | 85381 | 0 | 44-56 |
| BB3 | 84 | M | 0 | 2013-01-01T00:00:00 | ABCDEF8 | T2 | DRUG4 | 90774 | 1 | 83-95 |



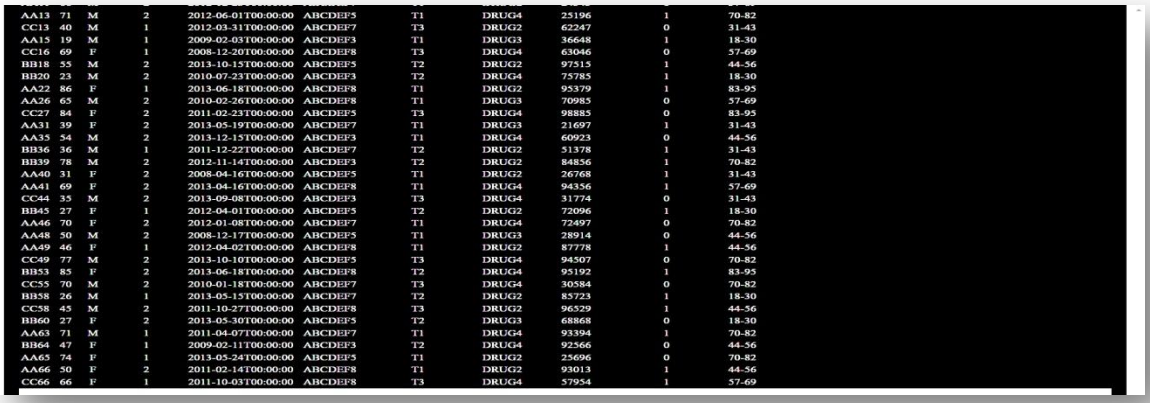
| ID | AGE | SEX | Smoking status | date | desease Code | Treatment Category | Drug group | Average Income | Affected to cancer | AgeGroup |
|---|---|---|---|---|---|---|---|---|---|---|
| AA1 | 88 | M | 2 | 2012-03-01T00:00:00 | ABCDEF1 | T1 | DRUG1 | 47583 | 0 | 83-95 |
| BB1 | 69 | M | 1 | 2008-09-26T00:00:00 | ABCDEF2 | T2 | DRUG2 | 43495 | 1 | 57-69 |
| AA2 | 47 | M | 0 | 2011-03-18T00:00:00 | ABCDEF4 | T1 | DRUG4 | 96705 | 0 | 44-56 |
| BB2 | 47 | M | 2 | 2013-09-09T00:00:00 | ABCDEF5 | T2 | DRUG1 | 87781 | 1 | 44-56 |
| CC2 | 18 | M | 1 | 2012-12-07T00:00:00 | ABCDEF6 | T3 | DRUG2 | 94805 | 1 | |
| AA3 | 46 | M | 2 | 2013-04-15T00:00:00 | ABCDEF7 | T1 | DRUG3 | 85381 | 0 | 44-56 |
| BB3 | 84 | M | 0 | 2013-01-01T00:00:00 | ABCDEF8 | T2 | DRUG4 | 90774 | 1 | 83-95 |
| CC3 | 82 | F | 1 | 2013-12-04T00:00:00 | ABCDEF1 | T3 | DRUG1 | 81520 | 0 | 70-82 |
| CC4 | 21 | F | 0 | 2011-09-28T00:00:00 | ABCDEF4 | T3 | DRUG4 | 97052 | 1 | 18-30 |
| AA5 | 78 | F | 0 | 2011-12-18T00:00:00 | ABCDEF5 | T1 | DRUG1 | 72339 | 0 | 70-82 |
| CC5 | 23 | M | 2 | 2013-10-13T00:00:00 | ABCDEF7 | T3 | DRUG1 | 34396 | 1 | 18-30 |
| AA6 | 72 | M | 0 | 2013-06-07T00:00:00 | ABCDEF8 | T1 | DRUG2 | 75867 | 0 | 70-82 |
| CC7 | 38 | F | 0 | 2012-07-22T00:00:00 | ABCDEF5 | T3 | DRUG3 | 34155 | 0 | 31-43 |
| AA8 | 25 | M | 2 | 2012-11-16T00:00:00 | ABCDEF6 | T1 | DRUG4 | 84548 | 1 | 18-30 |
| BB8 | 72 | M | 0 | 2010-01-28T00:00:00 | ABCDEF7 | T2 | DRUG1 | 33180 | 1 | 70-82 |
| CC8 | 25 | M | 1 | 2008-04-21T00:00:00 | ABCDEF8 | T3 | DRUG2 | 60528 | 1 | 18-30 |
| AA9 | 71 | F | 1 | 2013-01-22T00:00:00 | ABCDEF1 | T1 | DRUG3 | 21070 | 1 | 70-82 |
| BB9 | 23 | F | 2 | 2011-09-27T00:00:00 | ABCDEF2 | T2 | DRUG4 | 33582 | 1 | 18-30 |
| CC9 | 89 | F | 1 | 2013-04-17T00:00:00 | ABCDEF3 | T3 | DRUG1 | 62877 | 1 | 83-95 |
| AA10 | 45 | F | 1 | 2013-05-17T00:00:00 | ABCDEF4 | T1 | DRUG2 | 27230 | 0 | 44-56 |
| AA11 | 66 | M | 2 | 2012-12-23T00:00:00 | ABCDEF7 | T1 | DRUG2 | 24545 | 0 | 57-69 |
| CC11 | 35 | M | 1 | 2012-02-20T00:00:00 | ABCDEF1 | T3 | DRUG4 | 56528 | 0 | 31-43 |
| AA12 | 89 | M | 1 | 2010-09-10T00:00:00 | ABCDEF2 | T1 | DRUG1 | 25115 | 0 | 83-95 |
| AA13 | 71 | M | 2 | 2012-06-01T00:00:00 | ABCDEF5 | T1 | DRUG4 | 25196 | 1 | 70-82 |
| BB13 | 68 | M | 0 | 2013-09-01T00:00:00 | ABCDEF6 | T2 | DRUG1 | 25043 | 0 | 57-69 |
| CC13 | 40 | M | 1 | 2012-03-31T00:00:00 | ABCDEF7 | T3 | DRUG2 | 62247 | 0 | 31-43 |
| AA14 | 61 | M | 1 | 2012-03-18T00:00:00 | ABCDEF8 | T1 | DRUG1 | 47204 | 1 | 57-69 |
| CC14 | 34 | M | 2 | 2012-01-22T00:00:00 | ABCDEF2 | T3 | DRUG2 | 74068 | 1 | 31-43 |
| AA15 | 19 | M | 1 | 2009-02-03T00:00:00 | ABCDEF3 | T1 | DRUG3 | 36648 | 1 | 18-30 |
| BB15 | 40 | M | 0 | 2012-05-11T00:00:00 | ABCDEF4 | T2 | DRUG4 | 48738 | 1 | 31-43 |

Suppose we have picked date range from 2008 to 2014, then the data will be filtered out according to the date the user will choose.

## Data filtering by all the dropdowns:




| ID | AGE | SEX | Smoking status | date | desease Code | Treatment Category | Drug group | Average Income | Affected to cancer | AgeGroup |
|---|---|---|---|---|---|---|---|---|---|---|
| AA3 | 46 | M | 2 | 2013-04-15T00:00:00 | ABCDEF7 | T1 | DRUG3 | 85381 | 0 | 44-56 |
| CC8 | 25 | M | 1 | 2008-04-21T00:00:00 | ABCDEF8 | T3 | DRUG2 | 60528 | 1 | 18-30 |
| AA11 | 66 | M | 2 | 2012-12-23T00:00:00 | ABCDEF7 | T1 | DRUG2 | 24545 | 0 | 57-69 |
| AA13 | 71 | M | 2 | 2012-06-01T00:00:00 | ABCDEF5 | T1 | DRUG4 | 25196 | 1 | 70-82 |
| CC13 | 40 | M | 1 | 2012-03-31T00:00:00 | ABCDEF7 | T3 | DRUG2 | 62247 | 0 | 31-43 |
| AA15 | 19 | M | 1 | 2009-02-03T00:00:00 | ABCDEF3 | T1 | DRUG3 | 36648 | 1 | 18-30 |

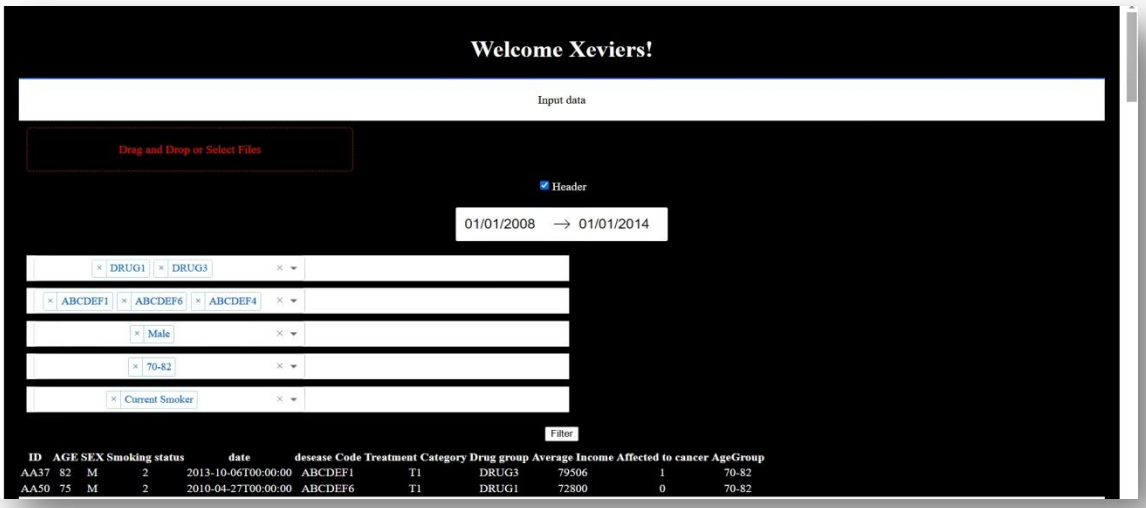


| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| AA13 | 71 | M | 2 | 2012-06-01T00:00:00 | ABCDEF5 | T1 | DRUG4 | 25196 | 1 | 70-82 |
| CC13 | 40 | M | 1 | 2012-03-31T00:00:00 | ABCDEF7 | T3 | DRUG2 | 62247 | 0 | 31-43 |
| AA15 | 19 | M | 1 | 2009-02-03T00:00:00 | ABCDEF3 | T1 | DRUG3 | 36648 | 1 | 18-30 |
| CC16 | 69 | F | 1 | 2008-12-20T00:00:00 | ABCDEF8 | T3 | DRUG4 | 63046 | 0 | 57-69 |
| BB18 | 55 | M | 2 | 2013-10-15T00:00:00 | ABCDEF5 | T2 | DRUG2 | 97515 | 1 | 44-56 |
| BB20 | 23 | M | 2 | 2010-07-23T00:00:00 | ABCDEF3 | T2 | DRUG4 | 75785 | 1 | 18-30 |
| AA22 | 86 | F | 1 | 2013-06-18T00:00:00 | ABCDEF8 | T1 | DRUG2 | 95379 | 1 | 83-95 |
| AA26 | 65 | M | 2 | 2010-02-26T00:00:00 | ABCDEF8 | T1 | DRUG3 | 70985 | 0 | 57-69 |
| CC27 | 84 | F | 2 | 2011-02-23T00:00:00 | ABCDEF5 | T3 | DRUG4 | 98885 | 0 | 83-95 |
| AA31 | 39 | F | 2 | 2013-05-19T00:00:00 | ABCDEF7 | T1 | DRUG3 | 21697 | 1 | 31-43 |
| AA35 | 54 | M | 2 | 2013-12-15T00:00:00 | ABCDEF3 | T1 | DRUG4 | 60923 | 0 | 44-56 |
| BB36 | 36 | M | 1 | 2011-12-22T00:00:00 | ABCDEF7 | T2 | DRUG2 | 51378 | 1 | 31-43 |
| BB39 | 78 | M | 2 | 2012-11-14T00:00:00 | ABCDEF3 | T2 | DRUG2 | 84856 | 1 | 70-82 |
| AA40 | 31 | F | 2 | 2008-04-16T00:00:00 | ABCDEF5 | T1 | DRUG2 | 26768 | 1 | 31-43 |
| AA41 | 69 | F | 2 | 2013-04-16T00:00:00 | ABCDEF8 | T1 | DRUG4 | 94356 | 1 | 57-69 |
| CC44 | 35 | M | 2 | 2013-09-08T00:00:00 | ABCDEF3 | T3 | DRUG4 | 31774 | 0 | 31-43 |
| BB45 | 27 | F | 1 | 2012-04-01T00:00:00 | ABCDEF5 | T2 | DRUG2 | 72096 | 1 | 18-30 |
| AA46 | 70 | F | 2 | 2012-01-08T00:00:00 | ABCDEF7 | T1 | DRUG4 | 72497 | 0 | 70-82 |
| AA48 | 50 | M | 2 | 2008-12-17T00:00:00 | ABCDEF5 | T1 | DRUG3 | 28914 | 0 | 44-56 |
| AA49 | 46 | F | 1 | 2012-04-02T00:00:00 | ABCDEF8 | T1 | DRUG2 | 87778 | 1 | 44-56 |
| CC49 | 77 | M | 2 | 2013-10-10T00:00:00 | ABCDEF5 | T3 | DRUG4 | 94507 | 0 | 70-82 |
| BB53 | 85 | F | 2 | 2013-06-18T00:00:00 | ABCDEF8 | T2 | DRUG4 | 95192 | 1 | 83-95 |
| CC55 | 70 | M | 2 | 2010-01-18T00:00:00 | ABCDEF7 | T3 | DRUG4 | 30584 | 0 | 70-82 |
| BB58 | 26 | M | 1 | 2013-05-15T00:00:00 | ABCDEF7 | T2 | DRUG2 | 85723 | 1 | 18-30 |
| CC58 | 45 | M | 2 | 2011-10-27T00:00:00 | ABCDEF8 | T3 | DRUG2 | 96529 | 1 | 44-56 |
| BB60 | 27 | F | 2 | 2013-05-30T00:00:00 | ABCDEF5 | T2 | DRUG3 | 68868 | 0 | 18-30 |
| AA63 | 71 | M | 1 | 2011-04-07T00:00:00 | ABCDEF7 | T1 | DRUG4 | 93394 | 1 | 70-82 |
| BB64 | 47 | F | 1 | 2009-02-11T00:00:00 | ABCDEF3 | T2 | DRUG4 | 92566 | 0 | 44-56 |
| AA65 | 74 | F | 1 | 2013-05-24T00:00:00 | ABCDEF5 | T1 | DRUG2 | 25696 | 0 | 70-82 |
| AA66 | 50 | F | 2 | 2011-02-14T00:00:00 | ABCDEF8 | T1 | DRUG2 | 93013 | 1 | 44-56 |
| CC66 | 66 | F | 1 | 2011-10-03T00:00:00 | ABCDEF8 | T3 | DRUG4 | 57954 | 1 | 57-69 |

Here we have chosen all the filtering dropdowns. We chose “DRUG2”,”DRUG4” and “DRUG3” in the “Select Drug Group” dropdown, we chose “ABCDEF8”,” ABCDEF7”,” ABCDEF3”,” ABCDEF5” in the “Select Disease Code” dropdown, then we chose both “Male” and “Female” in the “Select Sex” dropdown, then we kept the “Select Age Group” dropdown in its default format i.e., it will take all the age groups, and lastly we chose “Former Smoker” and “Current Smoker” which are represented as 2 and 1 in our data respectively, in the “Select Smoking Status” dropdown. Then if we filter our inputted data, we can see that the data has been filtered out corresponding to the Drug group, Disease code, Sex, Age group and Smoking status that have been selected by the user that you can see in the above 2 inserted pictures.

Now we will choose only “DRUG1” and “DRUG3” in the “Select Drug Group” dropdown, we will select “ABCDEF1”, “ABCDEF6”, “ABCDEF4” in the “Select Disease Code” dropdown, then we will select only “Male” in the “Select Sex” dropdown, then we will select “70-82” in the “Select Age Group” dropdown. And we will select “Current smoker” which is represented as 2 in our data, in the “Select Smoking

Status" dropdown. Then if we filter our data we can see that our data has been filtered out corresponding to the selected options in each dropdowns. Here you can see the out in the below picture:




After filtering the data here we can see that there are only 2 IDs in our filtered data whose both are Male(M), Smoking status=2, AgeGroup= 70-82, Disease code= ABCDEF1, ABCDEF6 and Drug group= DRUG3, DRUG1 respectively.

- **Data Visualization in Interactive way:**

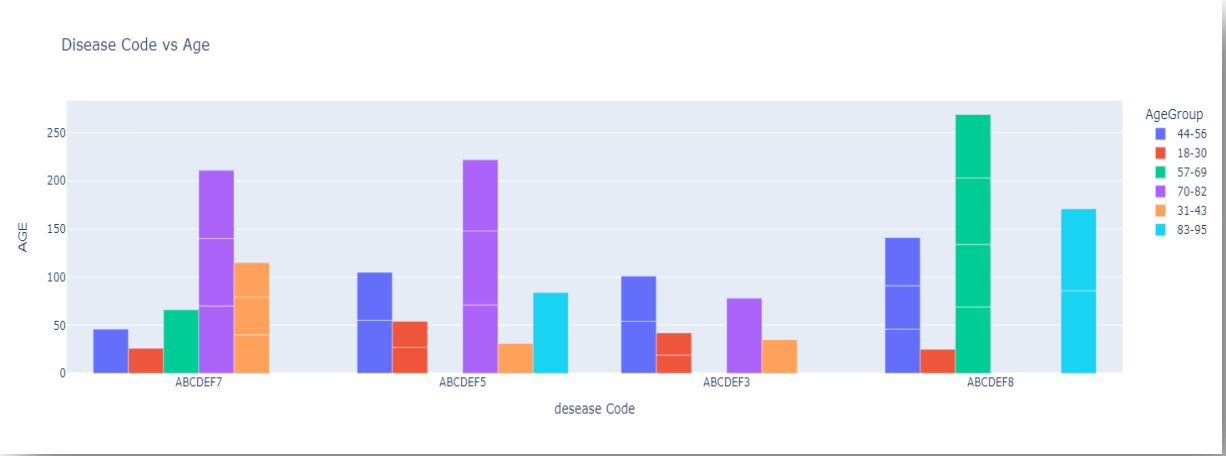
In this section, we present the visual representations of clinical data that have been crafted using the interactive dashboard created with the Dash framework.

Here we have chosen all the filtering dropdowns. We chose "DRUG2","DRUG4" and "DRUG3" in the "Select Drug Group" dropdown, we chose "ABCDEF8"," ABCDEF7"," ABCDEF3"," ABCDEF5" in the "Select Disease Code" dropdown, then we chose both "Male" and "Female" in the "Select Sex" dropdown, then we kept the "Select Age Group" dropdown in its default format i.e., it will take all the age groups, and lastly we chose "Former Smoker" and "Current Smoker" which are represented as 2 and 1 in our data respectively, in the "Select Smoking Status" dropdown:

## Bar Plot between "Disease Code" vs. "Age":

Here we can see the relationship between "Disease Code" and "Age". Here in the X axis we have the 4 selected Disease codes: ABCDEF3, ABCDEF5, ABCDEF7, ABCDEF8 and in the Y axis we have AGE.

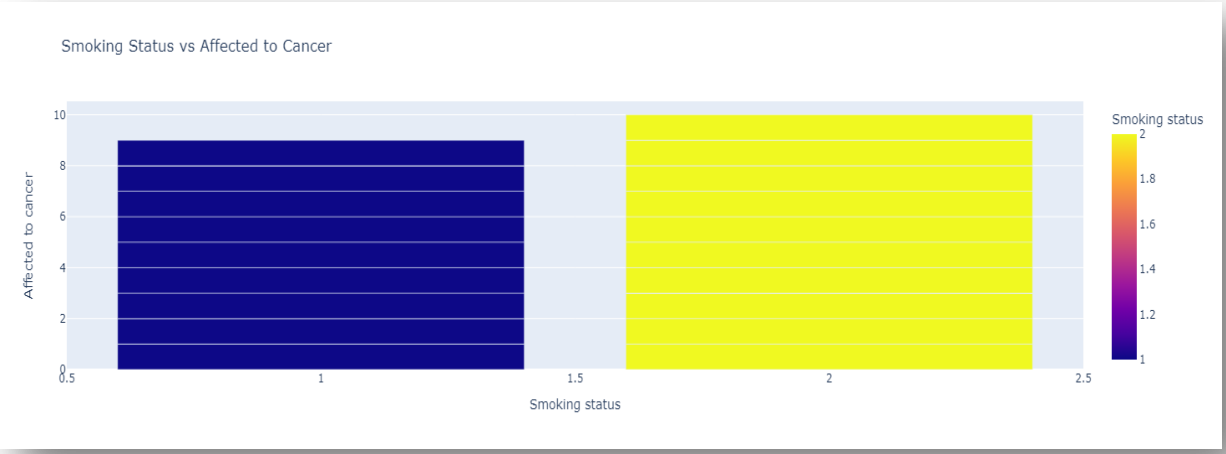
The bar plot between "Disease Code" and "Age" is in below:




Here different colors are representing different Age groups in our above bar plot. This is a group bar plot.

## Bar Plot between "Smoking Status" vs. "Affected to Cancer":

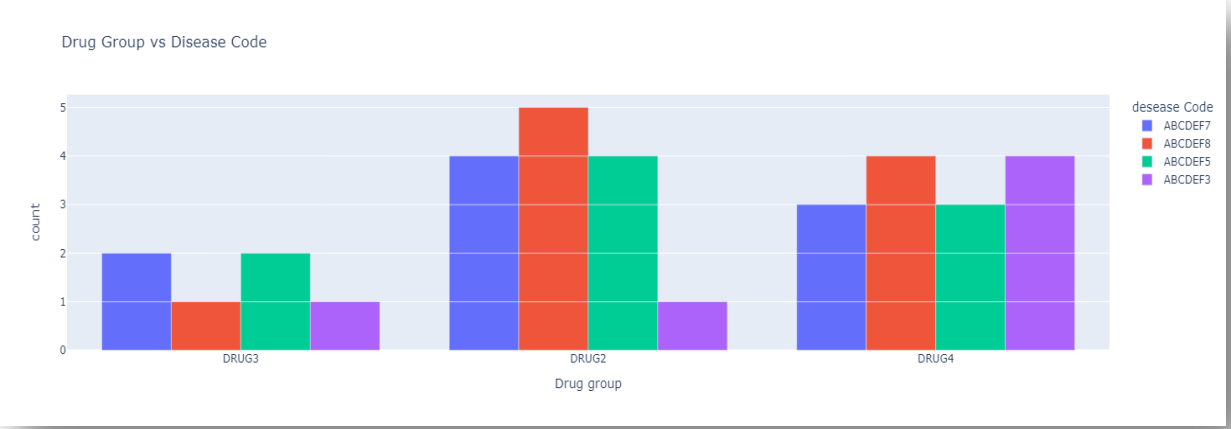
Here we can see the relationship between "Smoking Status" and "Affected to Cancer". The corresponding plot is below:




Here in the X axis we have "Smoking Status" where blue color=1 and yellow color=2 are representing Current smoker and Former smoker respectively. And in the Y axis have "Affected to cancer". This is a stacked bar plot.

## Bar Plot between "Drug Group" vs. "Disease Code":

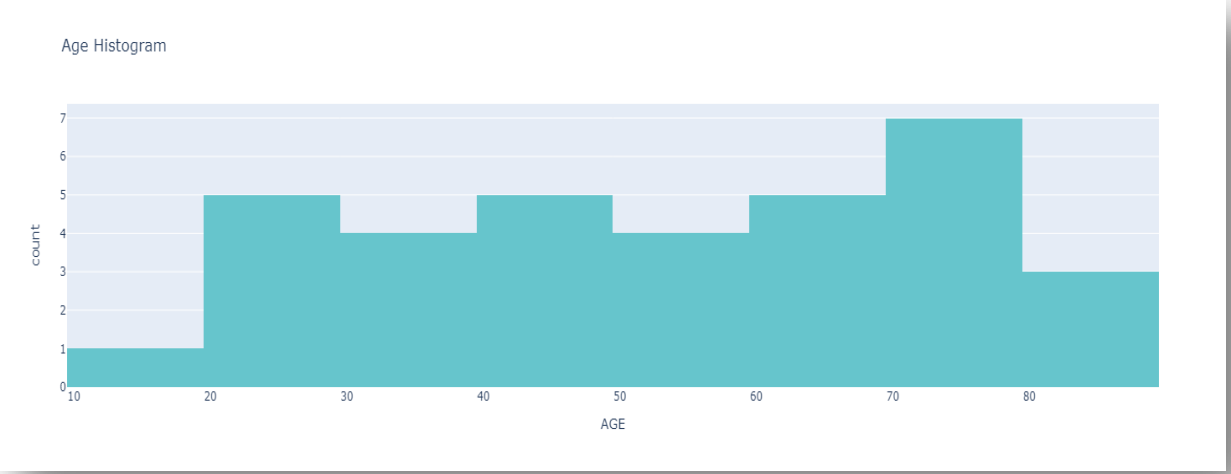
In this plot we can see the relationship between "Smoking Status" and "Disease Code". This is a group bar plot. The corresponding is below:




Here in the X axis we have 3 selected “Drug groups”= DRUG3, DRUG2, DRUG4 and in the Y axis we have the 4 selected “Disease codes” =ABCDEF7, ABCDEF8, ABCDEF5, ABCDEF3 where each colors are representing different disease codes. Blue= ABCDEF7, Orange= ABCDEF8, Green= ABCDEF5, Purple= ABCDEF3. This is a group bar plot.

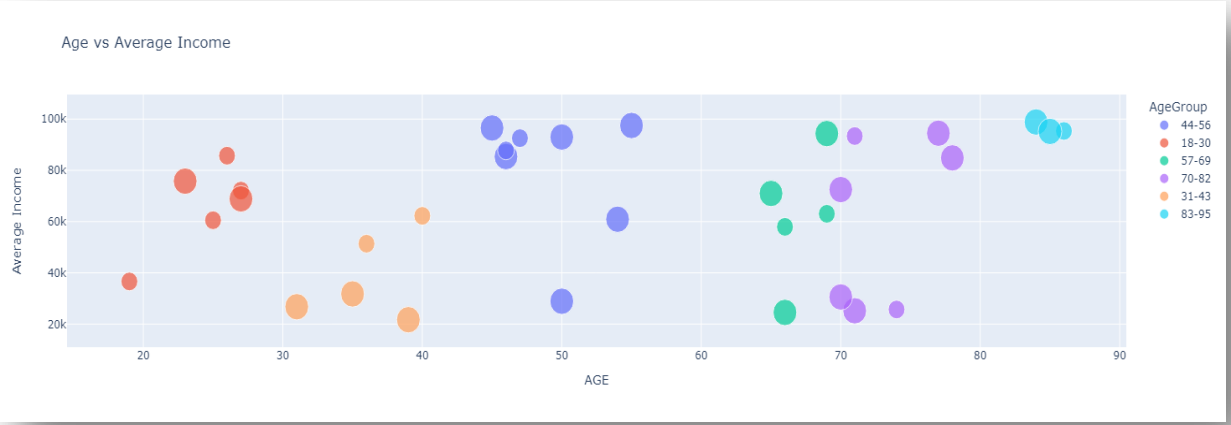
## Histogram of “AGE”:

In the below plot we can see the distribution of AGE by plotting a histogram. The corresponding plot is below:



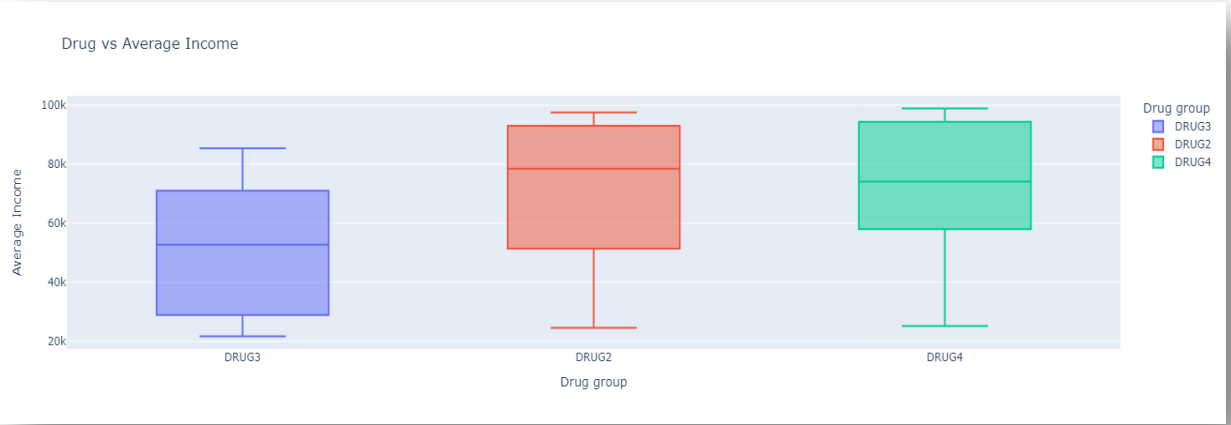

Here in the X axis we have “AGE” which starts from 10 to above 80. The histogram of "Age" distribution provides an insightful overview of the distribution of patients' ages within the clinical dataset. Each bar in the histogram represents a specific age range (bin) on the x-axis, while the height of the bar corresponds to the frequency or count of individuals falling within that age range on the y-axis.

**Bubble Chart between "Age" vs. "Average Income":** The bubble chart between "Age" and "Average Income" is an interactive visualization that provides a simultaneous representation of two continuous variables – the age of patients and their corresponding average income. In this chart, each individual is represented by a bubble placed at coordinates determined by their age and average income.




Here the horizontal position of each bubble along the x-axis represents the patient's age, while the vertical position along the y-axis corresponds to their average income. Bubbles located higher on the chart indicate higher average incomes, while those placed lower indicate lower average incomes. Each AgeGroup bubble is represented through different colors.

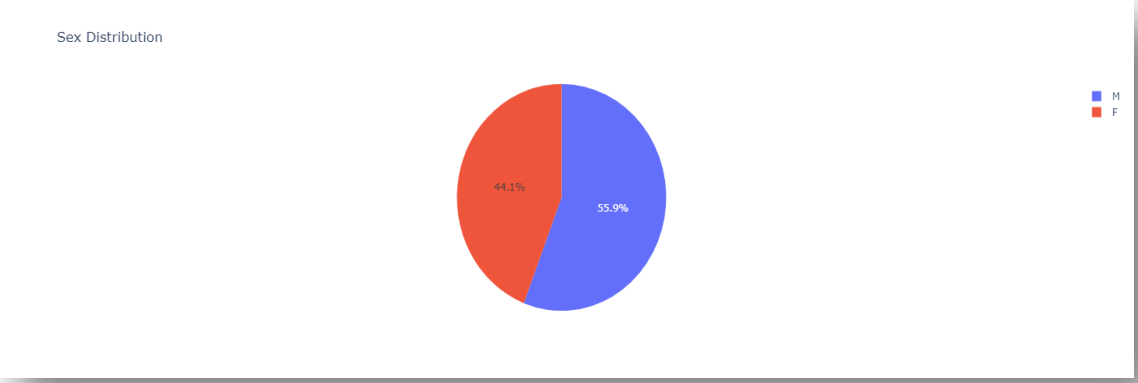
**Box plot between "Drug" vs. "Average Income":** The box plot depicting the relationship between "Drug Group" and "Average Income" provides insights into the distribution of average incomes across different drug categories. Each box represents the income distribution for a specific drug category




Here in the X axis we have 3 selected "Drug groups"=DRUG3, DRUG2, DRUG4 which are represented by blue, red and green color respectively and in the Y axis we have "Average Income" starting from 20k to 100k. The horizontal line within each box represents the median (50th percentile) average income for patients using the corresponding drug category.

## Pie Chart of SEX:

The pie chart representing the distribution of "Sex" provides a concise visual representation of the gender distribution. Each segment of the pie chart corresponds to a gender category, showing the proportion of male and female patients.




The pie chart clearly displays the relative distribution of male and female patients. Each segment's size is proportional to the number of patients belonging to that gender category. Here blue color is representing Male (M) and red color is representing Female(F) where M=55.9% and F=44.1% in our filtered data.

## ❖ Conclusion:

In summary, this project has successfully achieved the goal of designing and implementing an interactive clinical data dashboard using the Python programming language and the Dash framework. Through the integration of various Python libraries and the utilization of interactive components, the dashboard provides a dynamic platform for analyzing and visualizing clinical data in a user-friendly manner. The contributions of this project are multifold:

- **Enhanced Data Accessibility and Comprehensibility:** By leveraging the capabilities of interactive dashboards, healthcare professionals, researchers, and stakeholders gain the ability to explore and comprehend complex clinical data effortlessly. The intuitive interface and interactive features enable users to customize their analyses.
- **Informed Decision-Making:** The interactive dashboard empowers decision-makers with data-driven insights. Healthcare professionals can identify correlations between disease prevalence and patient attributes. Researchers can explore relationships between variables, leading to evidence-based research and discoveries.
- **Efficient Communication:** The visual nature of the interactive dashboards facilitates clear communication of insights to diverse audiences, regardless of their technical background. This is particularly valuable in interdisciplinary collaborations and presentations to stakeholders who may not have expertise in data analysis.

- **Demonstration of Python's Capabilities:** The project showcases the effectiveness of Python's libraries and frameworks, such as Pandas, Plotly Express, and Dash, in creating sophisticated and user-friendly data analysis tools. This serves as an exemplar for leveraging Python's capabilities in healthcare and clinical research.

The impact of interactive dashboards in clinical data analysis and decision-making is profound. It transforms raw data into actionable insights, promoting evidence-based practices and enabling healthcare providers and researchers to derive meaningful conclusions from complex datasets. In a rapidly evolving healthcare landscape, where data plays a pivotal role, the integration of interactive dashboards revolutionizes how we interact with and extract value from clinical data. The successful execution of this project underscores the potential of interactive visualization tools in catalyzing advancements in healthcare practices and research.

By combining data manipulation, visualization, and interactivity, this project underscores the transformative power of interactive dashboards in revolutionizing clinical data analysis and decision-making processes.

## ❖ References:

Here are some few references that helped me for constructing this project:


1. https://plotly.com/python/v3/create-online-dashboard-legacy/
2. https://www.topcoder.com/thrive/articles/creating-interactive-dashboards-using-plotly-dash
3. https://www.airplane.dev/blog/building-a-python-admin-dashboard-using-plotly-dash
4. https://chat.openai.com/
5. https://www.youtube.com/watch?v=34w-8DChx9g
6. https://pbpython.com/plotly-dash-intro.html
7. https://github.com/plotly/dash
8. https://github.com/meraki/dashboard-api-python
9. https://github.com/thu-vu92/python-dashboard-panel