**Definition of Data Analysis:**

It is a crucial step in the process of making data-driven decisions, as it allows analysts to identify trends, outliers, and other important factors that can inform business strategies, scientific hypotheses, or policy decisions. The process of data analysis typically involves several stages, including data collection, data cleaning and pre-processing, exploratory data analysis, statistical modeling, and interpretation of results. Through these stages, data analysts can gain a deeper understanding of the data they are working with and use that understanding to make informed decisions.

**Cause to use data analysis:**

There are many different causes for using data analysis, as it has many different applications across a wide range of industries and fields. Here are some of the main causes for using data analysis:

1. Decision-making: Data analysis can help inform decision-making by providing insights and information that can help individuals and organizations make more informed choices.

2. Improved efficiency: Data analysis can help improve efficiency by identifying inefficiencies and areas for improvement within an organization or process.

3. Identification of opportunities: Data analysis can help identify new opportunities by identifying patterns and trends that may not be immediately apparent.

4. Risk management: Data analysis can help manage risk by identifying potential risks and predicting outcomes.

5. Research and development: Data analysis can help drive innovation by providing insights and information that can inform the development of new products, services, or technologies.

Overall, data analysis is a powerful tool that can provide valuable insights and information for individuals and organizations across a wide range of industries and fields. By analyzing data from various sources, data analysis can help inform decision-making, improve efficiency, identify opportunities, manage risk, and drive innovation.

In our application:

Causes of use data analysis:

As we have data about:

Doctors: Name, Phone, Address, Age.

Patients: National ID, Name, Phone, Address, Age, patients ID.

Patients: National ID, Name, Phone, Age, Address, Gender, his doctor id

Diseases: If the patient injured by diabetes, hypertension, cholesterol

Medical Result: Xray Result, Xray Date for each user.

Survey Result: Answers of survey questions in the application and it’s date.

So, we decided to make use of this data that we have, in:

1. Assist in diagnosing patients:

relying on my data and identify patterns within it. This will be after collecting many records and having many patients. Therefore, I will be able to assist in diagnosing patients based on the symptoms they exhibit.

The ability to diagnose patients based on their symptoms is a crucial skill for medical professionals. By using data to inform these diagnoses, healthcare providers can ensure that patients receive the most appropriate treatments and care. This can be especially important for patients with complex or rare conditions, who may require specialized knowledge and expertise to diagnose and treat effectively.

1. In marketing:

* We want people to use our application, and we can achieve this through data analysis. By analyzing user data, we can identify areas where users may be experiencing difficulties or friction points and take steps to improve the user experience. We can also use data analysis to identify features or content that are particularly popular or engaging and highlight these to encourage further usage. Additionally, analyzing user data can help us identify patterns and trends in user behavior, which can inform our marketing and outreach efforts to attract more users. Overall, data analysis is a powerful tool for increasing user adoption and engagement and can help us create a more successful and widely used application.
* In order to be able to locate the places where patients and doctors use the application, and therefore I can do marketing and expand more in the number of users.

1. Generate Ratios:

I will be able to use the ratios generated from data analysis and incorporate them into our application. These ratios can provide valuable insights into various aspects of the application, such as user behavior, engagement, and retention. For example, by analyzing the ratio of new users to returning users, we can gain insights into user loyalty and identify areas where we may need to improve the user experience. Similarly, by analyzing the ratio of clicks to conversions, we can identify areas where users may be experiencing difficulties or friction points and take steps to improve the user experience. Overall, incorporating data analysis ratios into our application can help us make data-driven decisions, improve the user experience, and ultimately drive business growth.

1. Display the charts in a dashboard: by creating a dashboard that displays the charts and insights from your data analysis. This can be a great way to provide a quick overview of the key findings in the data.
2. Use the charts to inform your application's user interface: You can use the insights from your data analysis to inform the design of your application's user interface. For example, if you discover that a particular feature of your application is not being used as much as you expected, you can modify the user interface to make it more prominent.

Using Python in Our Analysis:

Python is a popular programming language for data analysis for several reasons:

1. Ease of use: Python is a relatively easy language to learn and use. Its syntax is simple and easy to read, making it accessible to a wide range of users.

2. Large and active community: Python has a large and active community of developers who contribute to the language and develop tools and libraries specifically for data analysis. This community provides a wealth of resources and support for users of all skill levels.

3. Versatility: Python is a versatile language that can be used for a wide range of applications, including data analysis, machine learning, web development, and more. This makes it a useful language to learn for individuals and organizations looking to develop a variety of skills and applications.

4. Interoperability: Python can be easily integrated with other tools and languages commonly used in data analysis, such as SQL, R, and Excel. This allows users to leverage the strengths of multiple tools and languages to achieve their data analysis goals.

5. Powerful data analysis libraries: Python has several powerful data analysis libraries, such as Pandas, NumPy, and Matplotlib, that provide a wide range of functions and tools for data manipulation, cleaning, visualization, and statistical analysis.

Overall, Python is a popular language for data analysis due to its ease of use, large and active community, versatility, interoperability, and powerful data analysis libraries. These factors make it a useful language for individuals and organizations looking to perform data analysis tasks efficiently and effectively.

Data analysis typically involves several stages or processes, which can be summarized as follows:

1. Data collection:

2. Data cleaning and preprocessing

3. Data exploration

4. Data analysis

5. Data visualization and reporting

Overall, data analysis involves several stages or processes, including data collection, cleaning and preprocessing, data exploration, data analysis, and data visualization and reporting. By following these processes, data analysts can gain valuable insights and make informed decisions based on the data.

1. **Data collection**

The first step in data analysis is to collect the data that will be analyzed. This may involve gathering data from various sources, such as databases, sensors, or surveys.

Through the data collected from the application, when the patient enters the application and collects his data, it is stored in the data base, and when the image is uploaded, the user fills out a survey containing some questions related to the symptoms he has, as well as the doctor when he enters the application, his data is preserved, and with this The data used in the analysis are collected.

And we prepare the data for the application that we have stored. and that is through:

A picture containing text, screenshot, software, computer

Description automatically generated

1. **Data cleaning and preprocessing**

Once the data is collected, it must be cleaned and preprocessed to ensure that it is accurate, complete, and in a format that can be analyzed. This may involve removing outliers or missing values, converting data into appropriate formats, and normalizing data.

1. **Check completeness of data and handling any missing values**

A screenshot of a computer

Description automatically generated with medium confidence

But, what if the data will contain missing in the future?

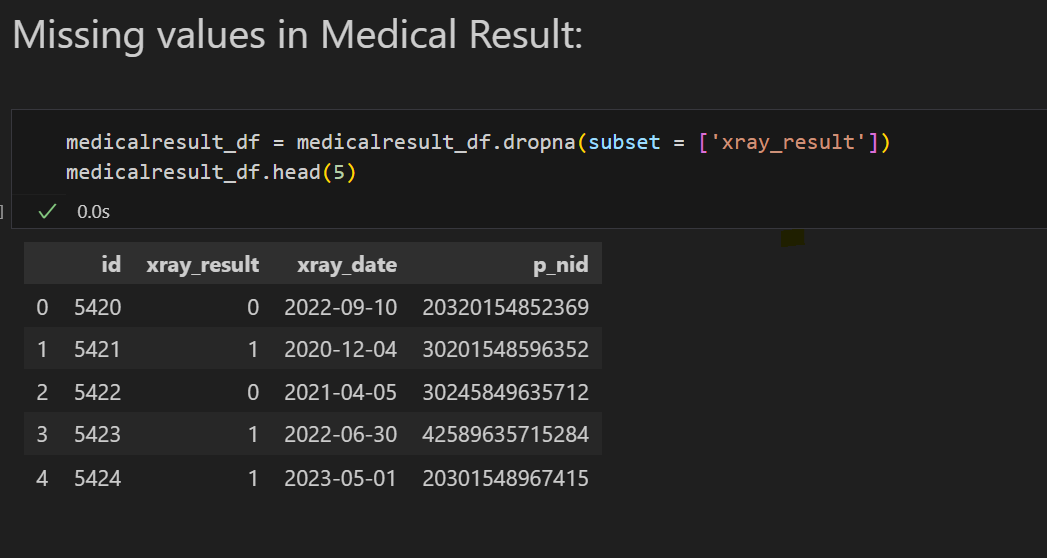
Handle of this is:

1. In Doctor table:

A screenshot of a computer program

Description automatically generated with medium confidence





A screenshot of a computer program

Description automatically generated with medium confidence

1. Datatypes of columns in each table, as done in:

A picture containing text, screenshot, font

Description automatically generated

A screenshot of a computer program

Description automatically generated with medium confidence

1. converting data into appropriate formats:

A screenshot of a computer program

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

1. Removing Outliers, improve the accuracy and reliability of your data analysis. This can be done using various statistical techniques, such as Z-score analysis, interquartile range (IQR) analysis, or box plots. These techniques help to identify data points that are significantly different from the rest of the dataset and remove them from the analysis.

* Z-score: used to identify and remove outliers from the dataset before performing any further analysis. A screen shot of a computer

  Description automatically generated with medium confidence
* A screenshot of a computer program

  Description automatically generated with medium confidence
* A screenshot of a computer

  Description automatically generated

A screen shot of a computer

Description automatically generated with low confidence

A screenshot of a computer

Description automatically generated

Cleaning data before analysis is important for several reasons:

1. Accuracy of analysis.

2. Consistency of analysis.

3. Efficiency of analysis.

4. Improved data quality.

5. Better decision-making.

Overall, cleaning data before analysis is an important step in the data analysis process. It helps ensure the accuracy, consistency, and efficiency of the analysis, and can lead to improved data quality and better decision-making.

3. **Data exploration** the next step is to explore the data to gain insights and identify patterns or trends. This may involve visualizing data using graphs or charts, calculating summary statistics, or applying machine learning algorithms to identify correlations or clusters.

A screenshot of a computer

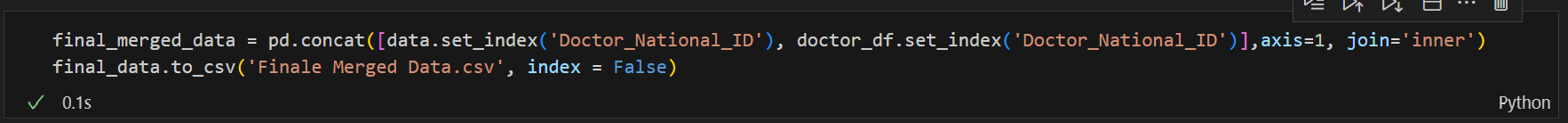
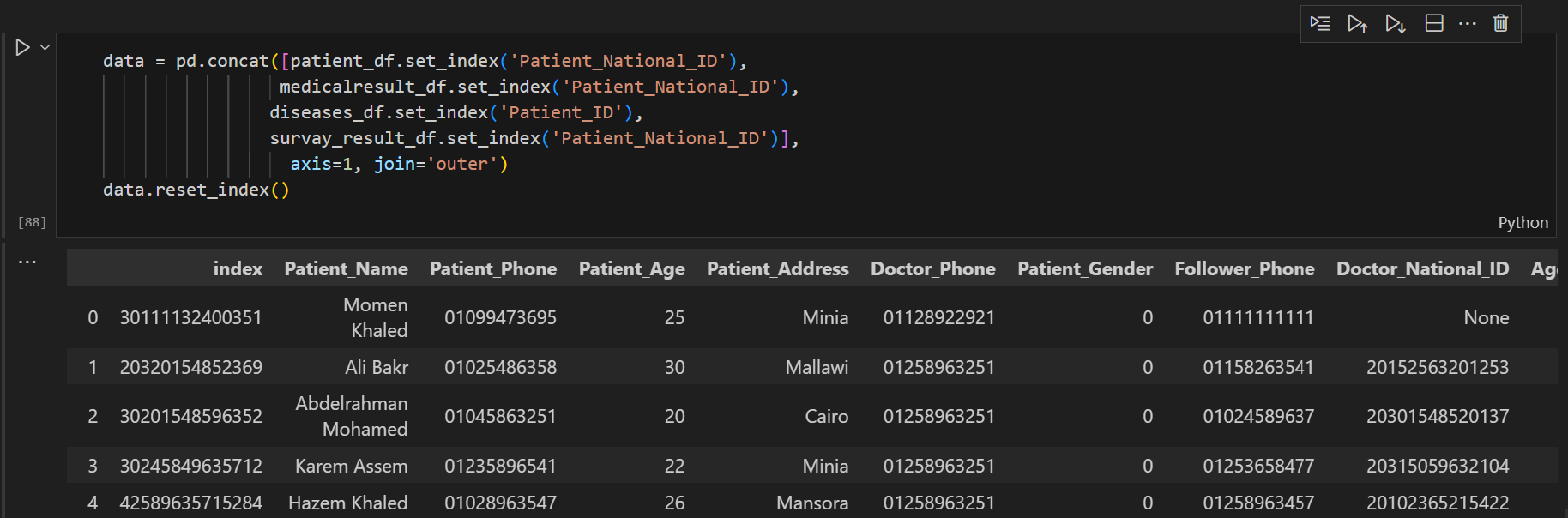
Description automatically generatedA screenshot of a computer

Description automatically generated with medium confidenceA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated with medium confidence

Marge all tables:



1. **Data analysis, Data visualization and reporting:**

Once the data has been explored, the next step is to perform the actual data analysis. This may involve applying statistical models or machine learning algorithms to the data or conducting hypothesis testing to determine the significance of observed patterns or trends.

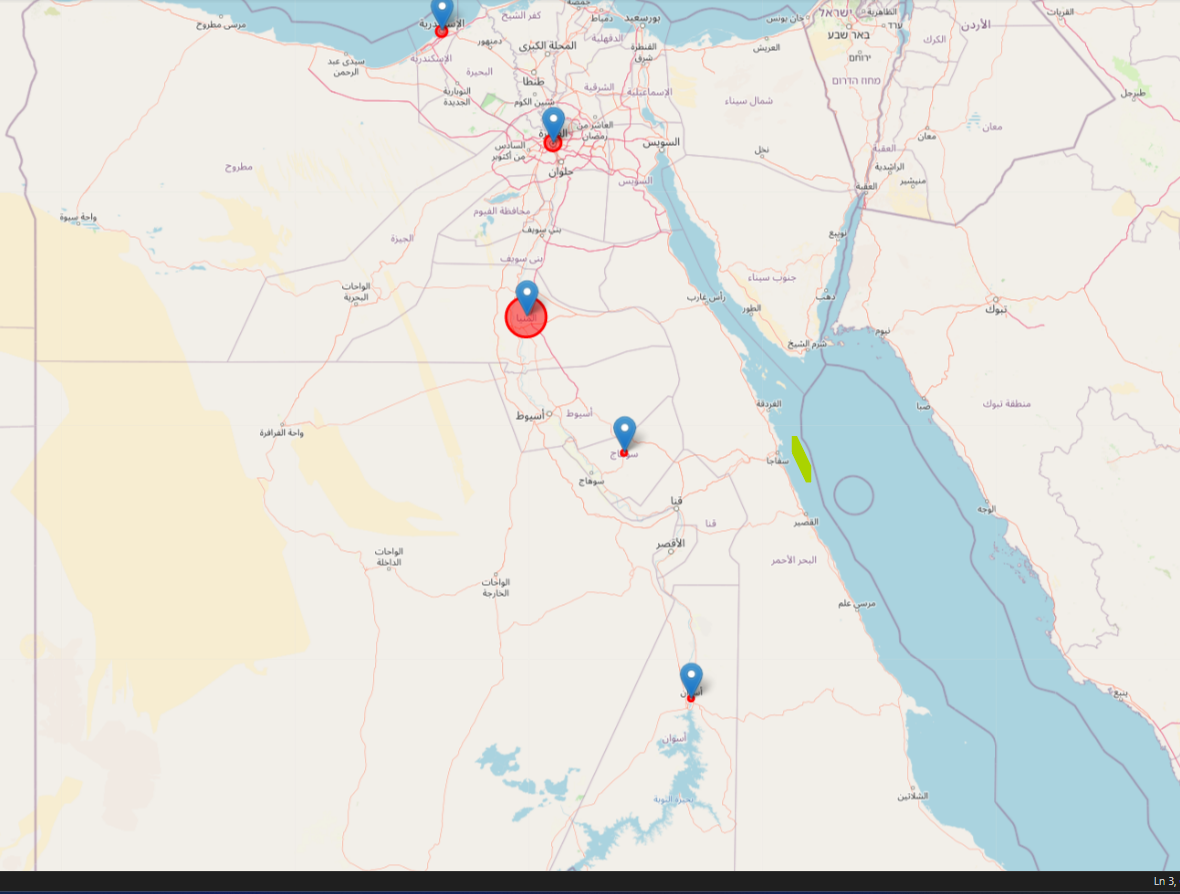
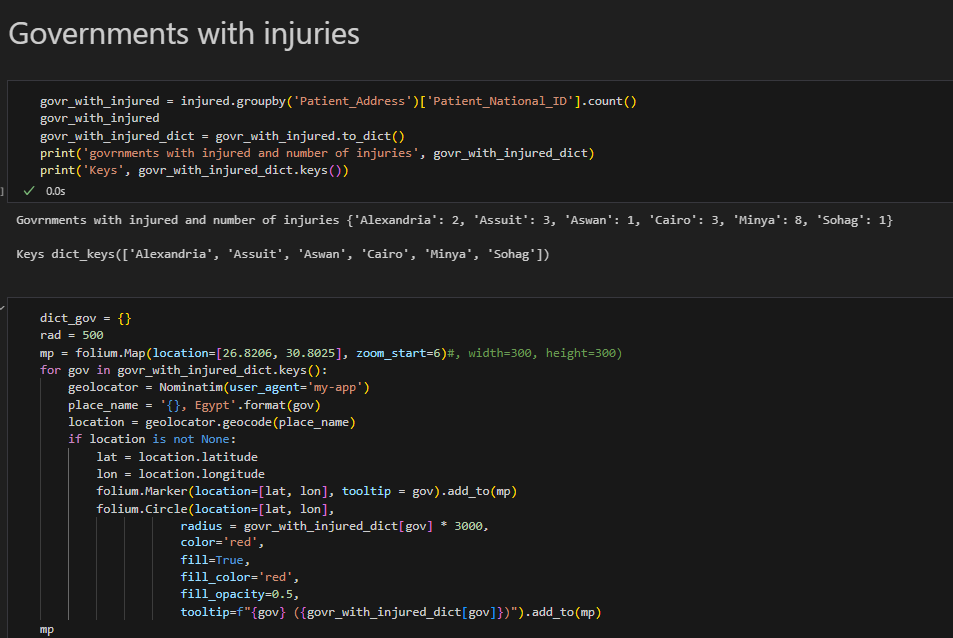
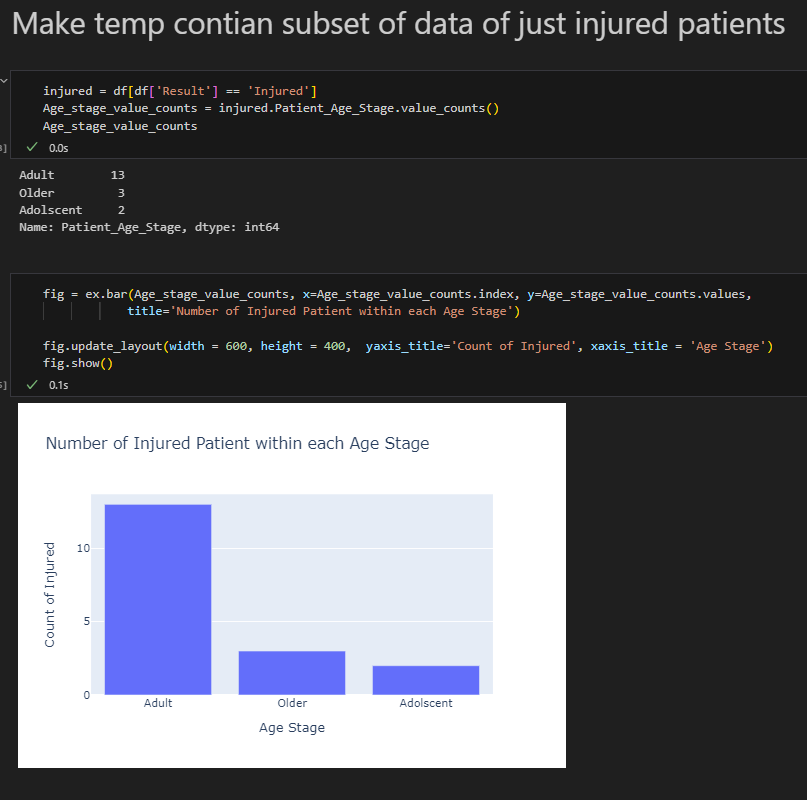
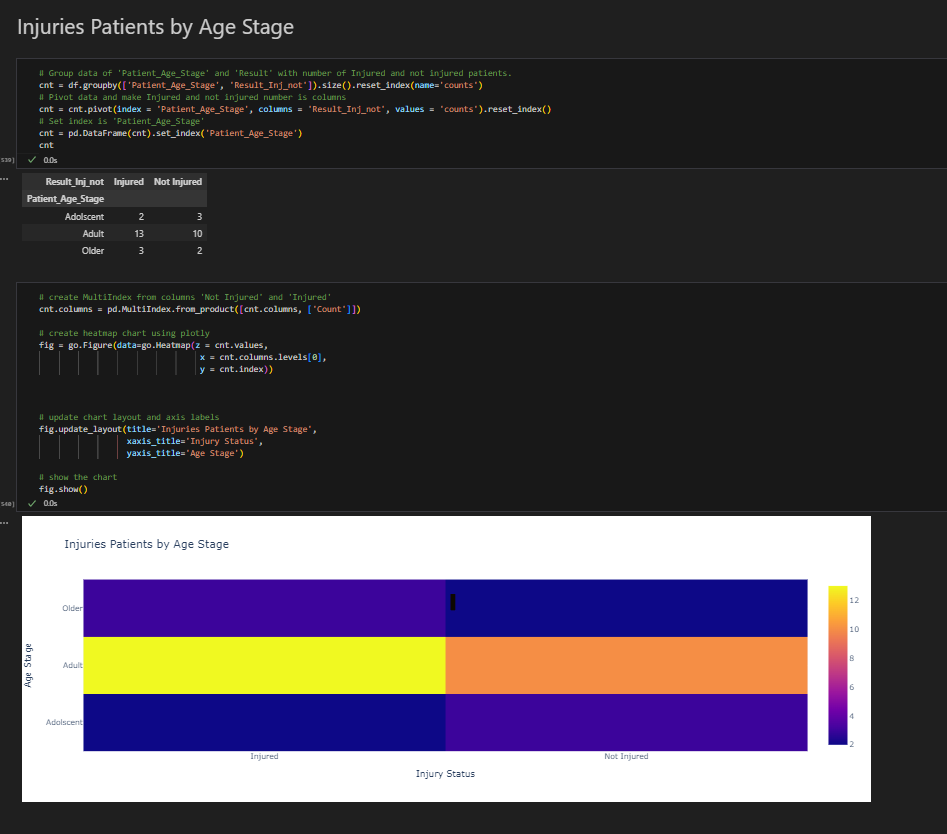
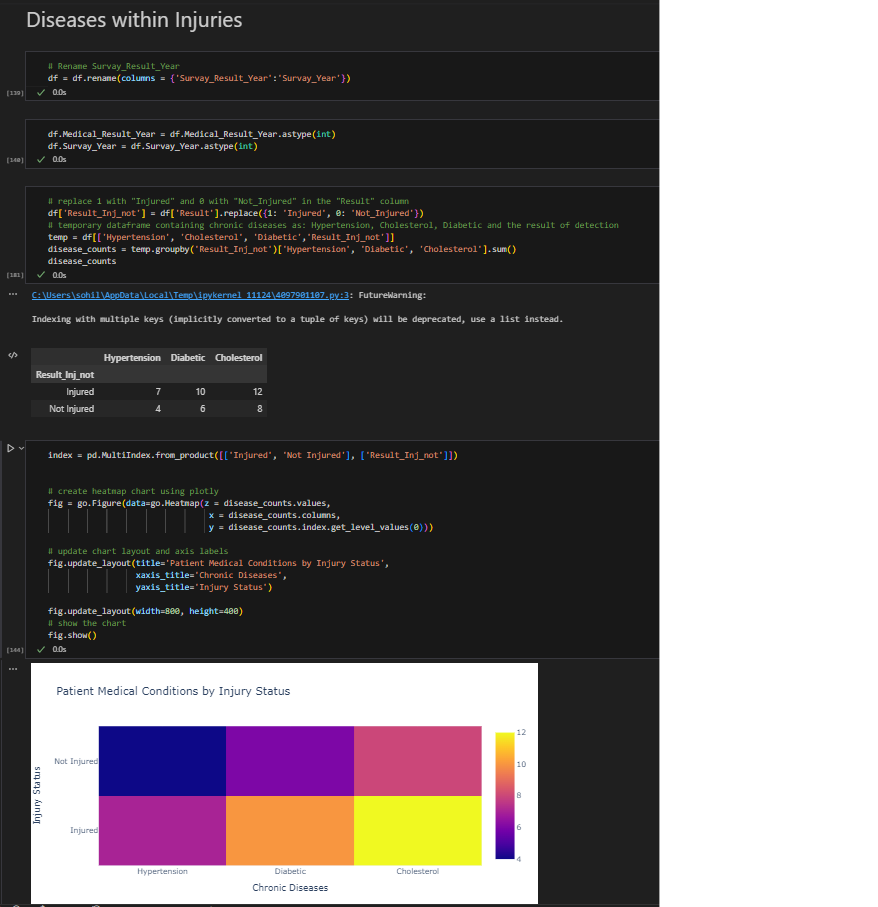
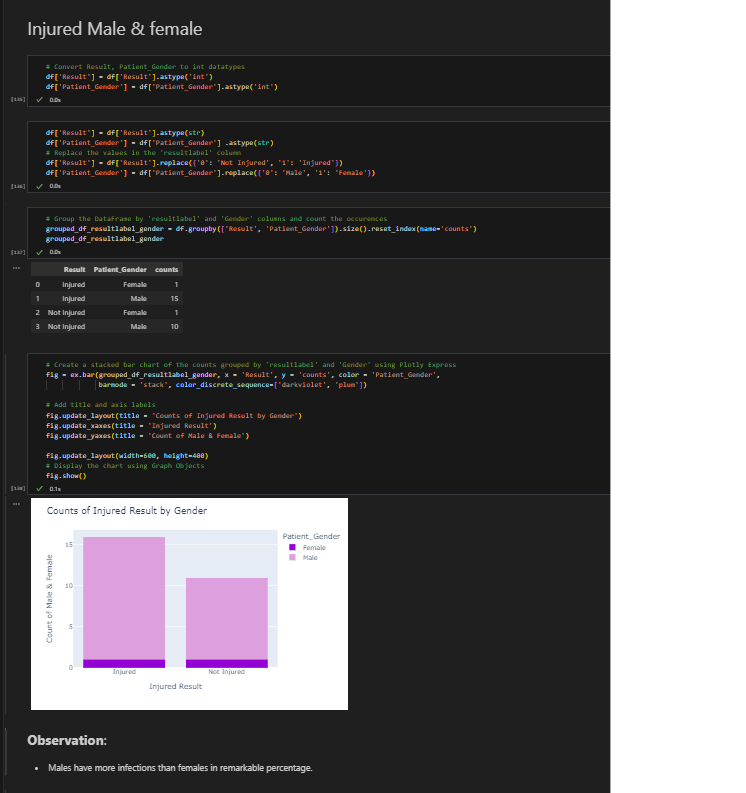
The choice of statistical models used in data analysis depends on the specific goals and questions being addressed. However, some common statistical models that are often used in data analysis include:

1. Linear regression: Linear regression is a statistical model used to establish a relationship between an independent variable and a dependent variable. It is often used to predict the value of a dependent variable based on the value of one or more independent variables.

A picture containing text, screenshot, multimedia software, software

Description automatically generated

* It is important to explore and understand the relationships between different variables in the dataset. One way to do this is through data visualization, which can help identify patterns and trends in the data that may not be immediately apparent from looking at the raw numbers. Some commonly used visualization techniques for data analysis include scatter plots, histograms, bar charts, and line graphs. These can help us to visualize the distribution of data, identify outliers or anomalies, and understand the relationships between different variables.
* The results of the data analysis must be communicated to stakeholders. This may involve creating visualizations or charts to help convey the results or writing reports or presentations to communicate the findings.

Some analysis and visualizations from code

In this chapter, we analyzed a dataset of Phenomena for our application. Our objective was to identify key factors that marketing, Generate Ratios, Display the charts in a dashboard.

As our data is still small, the results will change in increase of data and uses use the applications.