

Heart Disease Analysis

MENTOR NAME: SRI LANKA LAKSHMINARAYANA SIR

FACULTY MENTOR NAME: K. RATNA KUMARI MADAM

TEAM ID: LTVIP2026TMIDS58782

TEAM LEADER



NAME : GANTA NAVEEN

ID : SBAP0052563

Email : naveen232214@gmail.com

TEAM MEMBER



NAME : BADAM SRI RAMA SAI NARASIMHAMURTHY

ID: SBAP0052639

EMAIL: badamnarsimha405@gmail.com

TEAM MEMBER



NAME : RONGALA NITEESH

ID : SBAP0052592

EMAIL: rongalaniteesh000@gmail.com

TEAM MEMBER



NAME : SHAIK ABDULLA
ID : SBAP0052616
EMAIL: sk0003118@gmail.com

TEAM MEMBER



NAME : KADIYALA LOHITH SESHA VEERA SURYA VAMSI
ID : SBAP0052620
EMAIL: lohithk2211@gmail.com

TEAM MEMBER



NAME : BANDIREDDY MAHESH DURGA RAJU
ID: SBAP0052626
EMAIL: maheshdurgaraju@gmail.com

INTRODUCTION

HEART DISEASE ANALYSIS

Heart disease remains one of the leading causes of mortality worldwide, and its impact continues to grow due to lifestyle changes, poor dietary habits, and lack of physical activity. Despite advances in medicine, prevention and early detection remain critical in reducing the risk of severe outcomes. However, analyzing large-scale health data related to heart disease—such as patient demographics, medical history, lifestyle choices, and clinical indicators—requires advanced tools for extracting insights.

In this project, we aim to use Tableau as a powerful data visualization and business intelligence tool to analyze heart disease data. The goal is to transform raw data into meaningful dashboards, highlight key risk factors, and identify correlations that can support better decision-making for healthcare providers, policymakers, and individuals. By leveraging Tableau's interactive visualizations, the project seeks to uncover hidden trends, compare patient groups, and tell stories through data that aid in preventive care and awareness.

SCENARIO 1

Dr. Sharma is a senior cardiologist working at a metropolitan hospital. She wants to understand which lifestyle factors contribute most to the increase in heart disease cases among middle-aged patients. Using Tableau dashboards, she can analyze patient data segmented by age, gender, BMI, cholesterol levels, and smoking habits. This helps her identify high-risk groups and design targeted awareness campaigns, such as advising patients on weight management and smoking cessation programs.

SCENARIO 2

Ramesh works with a government health department and is tasked with developing preventive health policies. He uses Tableau dashboards to study trends in heart disease prevalence across different regions, comparing rural and urban populations. By analyzing correlations between sedentary lifestyle indicators and disease rates, he can recommend policies such as fitness programs in workplaces, stricter tobacco regulations, and subsidies for healthier food options. Tableau helps him present these findings in interactive dashboards for decision-makers.

SCENARIO 3

Anita, a 45-year-old professional with a family history of heart disease, wants to monitor her health risks. With simplified Tableau dashboards provided by her healthcare provider, she can visualize her risk factors such as cholesterol levels, blood pressure, and lifestyle habits compared to healthy benchmarks. The dashboard highlights actionable steps like increasing physical activity or reducing fat intake. This empowers Anita to make informed decisions about her lifestyle and proactively reduce her risk of developing heart disease.

PROJECT FLOW

To accomplish this, we have to complete all the activities listed below,

- Define Problem / Problem Understanding
 - Specify the business problem
 - Business requirements
 - Literature Survey
 - Social or Business Impact.
- Data Collection & Extraction from Database
 - Collect the dataset,
 - Storing Data in DB
 - Perform SQL Operations
 - Connect DB with Tableau
- Data Preparation
- Prepare the Data for Visualization
- Data Visualizations
 - No of Unique Visualizations
- Dashboard
 - Responsive and Design of Dashboard
- Story
 - No of Scenes of Story
- Performance Testing
 - Amount of Data Rendered to DB
 - Utilization of Data Filters
 - No of Calculation Fields
 - No of Visualizations/ Graphs
- Web Integration
 - Dashboard and Story embed with UI With Flask
- Project Demonstration & Documentation
 - Record explanation Video for project end to end solution
 - Project Documentation-Step by step project development procedure

MILESTONE 1 : DATA COLLECTION & EXTRACTION FROM DATABASE

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes and generate insights from the data.

ACTIVITY 1 : DOWNLOADING THE DATASET

Acquire the finalized dataset required for the Tableau project, ensuring it is clean, relevant, and aligned with the defined problem. Validate data integrity and readiness for analysis and visualization tasks.

Link: https://drive.google.com/file/d/190Qmq27LeZZ_nWricP3Obl7ys_5otEsp/view

ACTIVITY 2 : STORING DATA IN DB & PERFORM SQL OPERATIONS

Store the collected dataset in a structured database system. Use SQL operations to clean, filter, transform, and prepare the data for seamless integration with Tableau for visualization and analysis.

ACTIVITY 3 : CONNECT DB WITH TABLEAU

Establish a secure and reliable connection between the database and Tableau. Ensure real-time or scheduled data access for creating interactive dashboards and performing dynamic data analysis within the Tableau environment.

MILESTONE 2 : DATA PREPARATION

Clean, transform, and organize the connected data to ensure consistency and accuracy. Create calculated fields, handle null values, and structure the data appropriately for effective visualization and insightful analysis in Tableau.

ACTIVITY 1 : PREPARE THE DATA FOR VISUALIZATION

In this step, we focus on preparing the dataset for visualization in Tableau. Fortunately, the dataset we're working with has already been pre-cleaned, meaning major cleaning steps such as handling missing values, removing duplicates, and correcting inconsistencies have already been taken care of.

However, even with a clean dataset, it's still essential to go through a brief review process to ensure it's truly ready for analysis:

- Data Review & Exploration
While the dataset is clean, it's good practice to explore it briefly—checking data types, value ranges, and distributions. This helps us understand the structure, identify any potential outliers, and gain familiarity with the data we'll be visualizing.

- Filtering and Structuring for Purpose
Depending on the business question, we may still need to filter the data to focus on specific subsets—such as certain time periods, regions, or product categories. Structuring the data to match the visualization goal helps ensure relevance and clarity.
- Field Renaming & Final Formatting
To enhance clarity in Tableau, we ensure field names are intuitive and consistent. We also check for proper data types (e.g., date fields, numeric values) and relationships if the dataset spans multiple tables.
- Optional Calculated Fields
If needed, we can create calculated fields (e.g., profit margin, growth rate) to support deeper analysis. Even with a clean dataset, these additions can make our visualizations more insightful.
- Validation for Accuracy
Lastly, a quick validation against the source or summary metrics ensures everything is accurate. This final step helps maintain trust in the insights generated.

MILESTONE 3 : DATA VISUALIZATION

Data visualization is the process of creating graphical representations of data to help people understand and explore the information. The goal of data visualization is to make complex data sets more accessible, intuitive, and easier to interpret. By using visual elements such as charts, graphs, and maps, data visualizations can help people quickly identify patterns, trends, and outliers in the data.

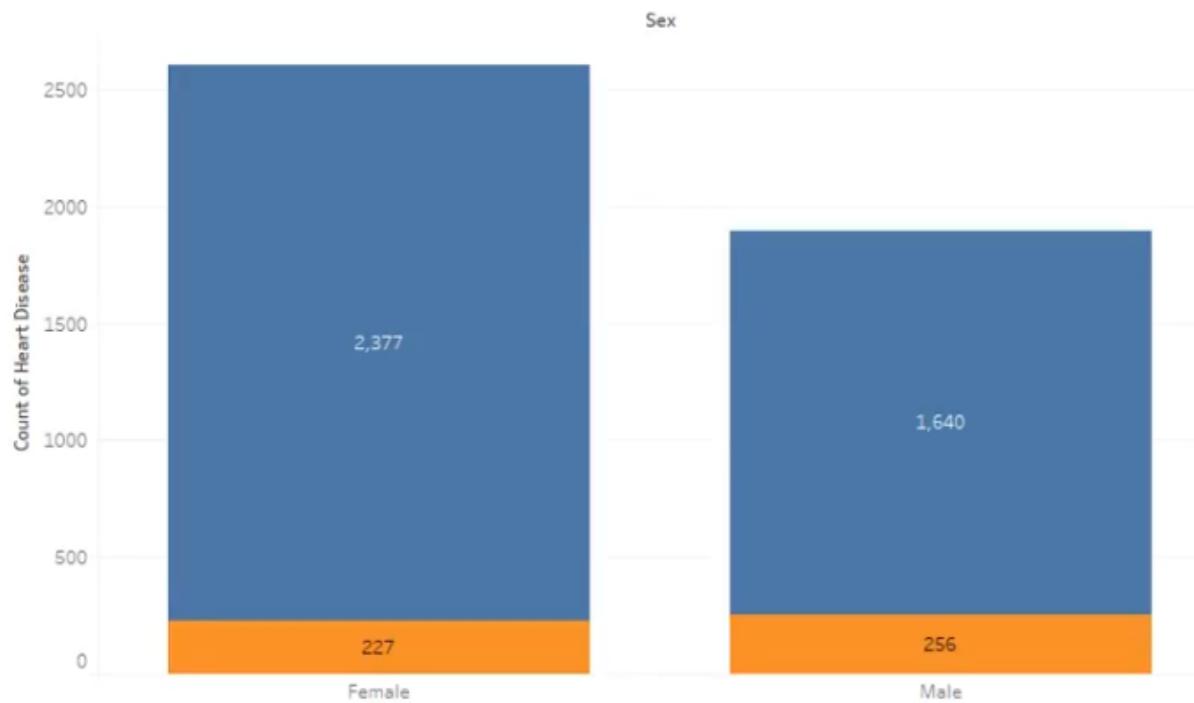
ACTIVITY 1 : NO OF UNIQUE VISUALIZATIONS

This project focuses on analyzing heart disease data using Tableau by creating 8–10 unique visualizations including trend charts, heat maps, comparative bar graphs, scatter plots, and dashboards. The aim is to uncover risk factors like obesity, smoking, and lifestyle habits, enabling healthcare professionals, policymakers, and patients to gain actionable insights and make informed, preventive decisions against cardiovascular diseases.

1. Gender Vs Heart Disease

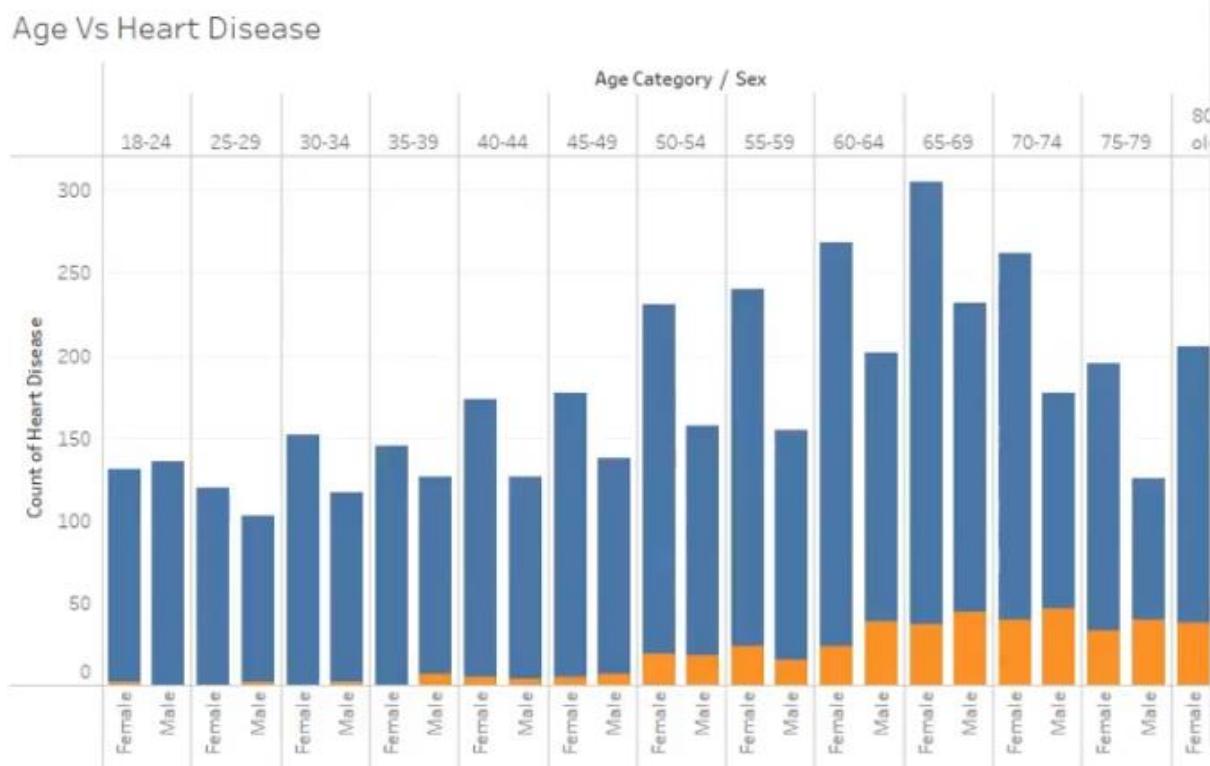
[Demo Link](#)

Gender vs Heart disease



2. Age vs Heart Disease

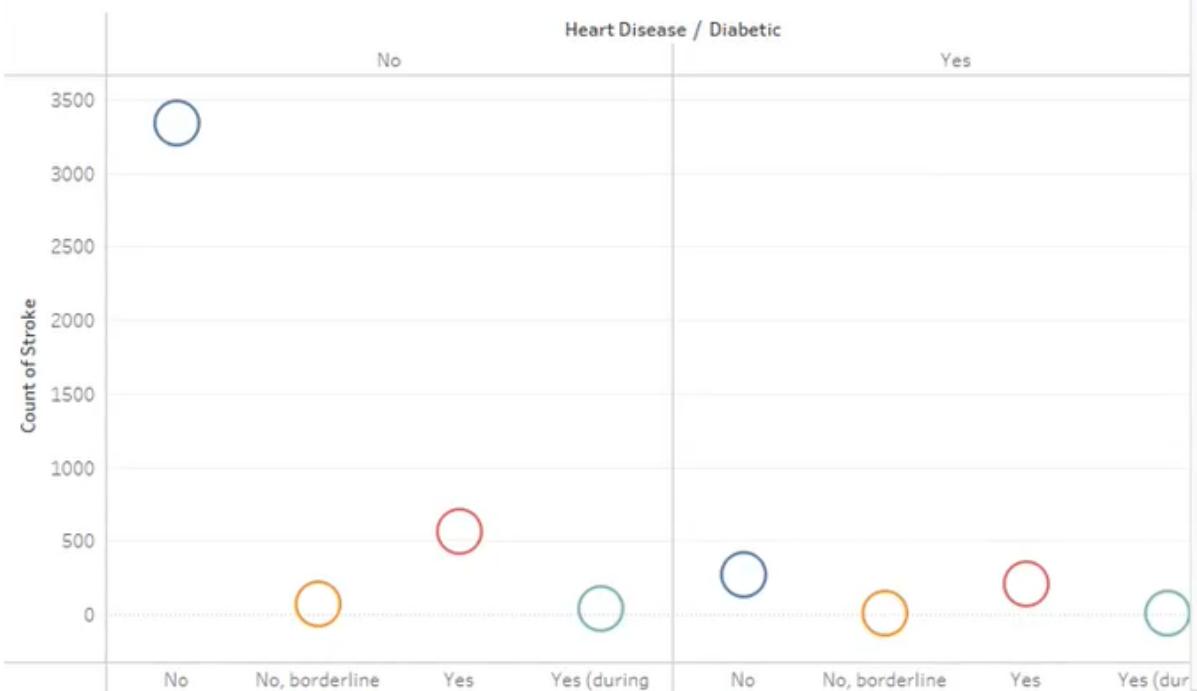
[Demo Link](#)



3. Diabetic vs Stroke

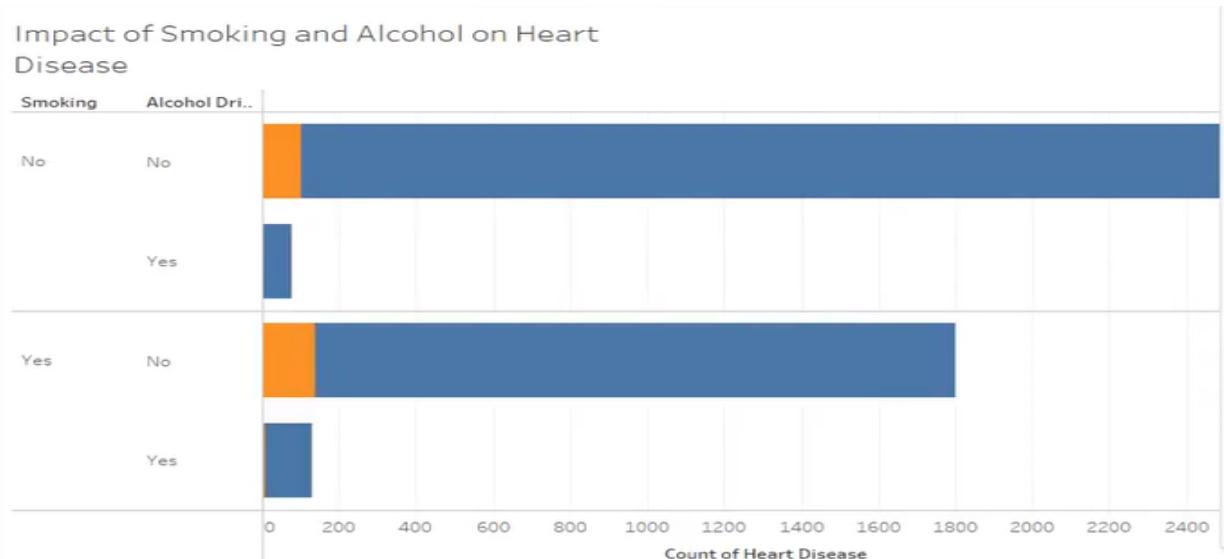
[Demo Link](#)

Diabetic vs Stroke



4. Impact of Smoking and Alcohol on Heart Disease

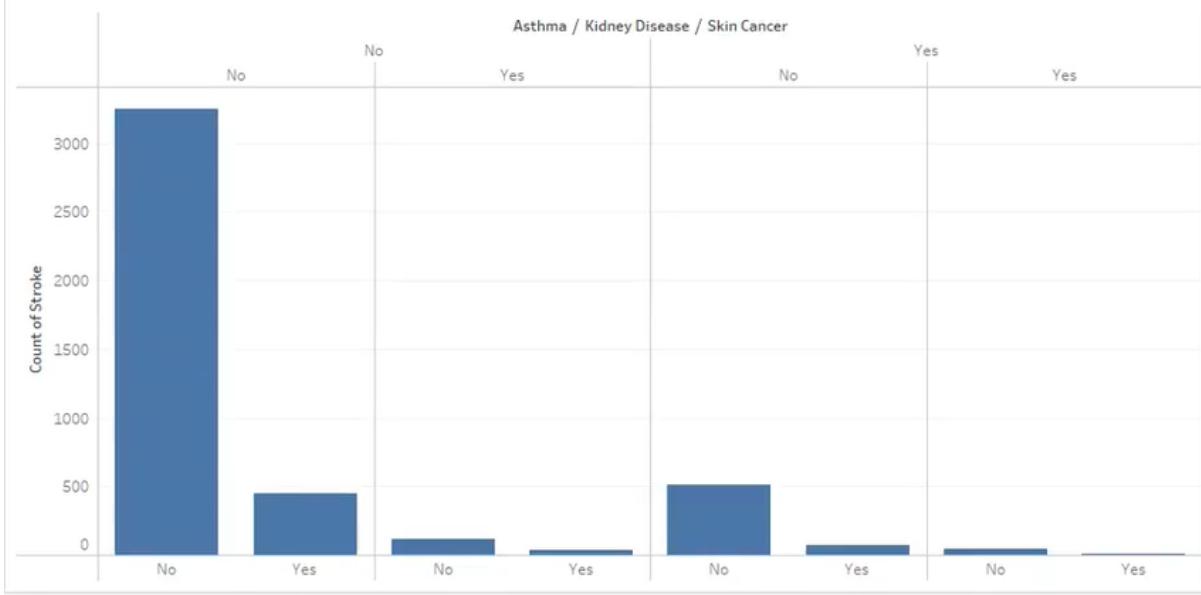
[Demo Link](#)



5. Other Heart Disease vs Stroke

[Demo Link](#)

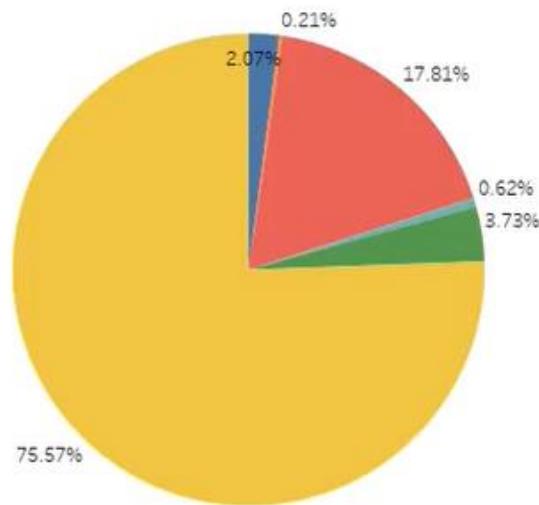
Stroke vs Other Disease



6. Race wise Heart Disease

[Demo Link](#)

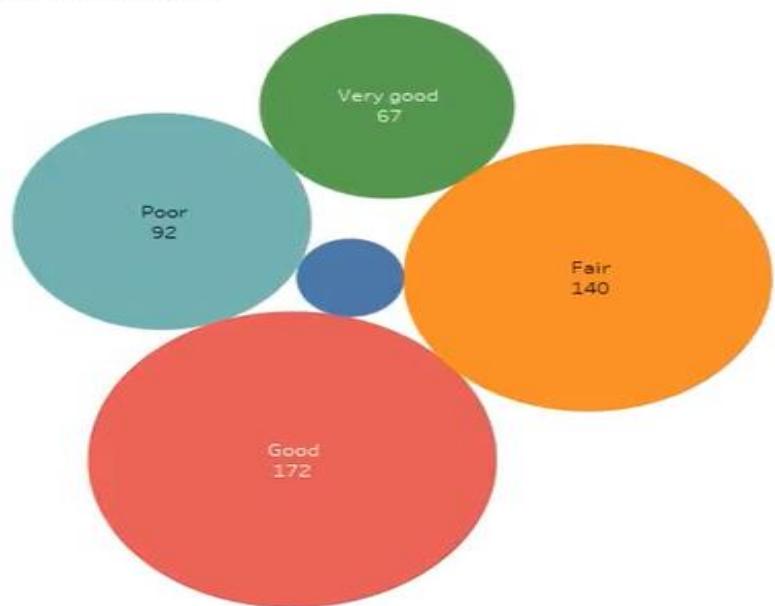
Race wise Heart Disease



7. General Health vs Heart Disease

[Demo Link](#)

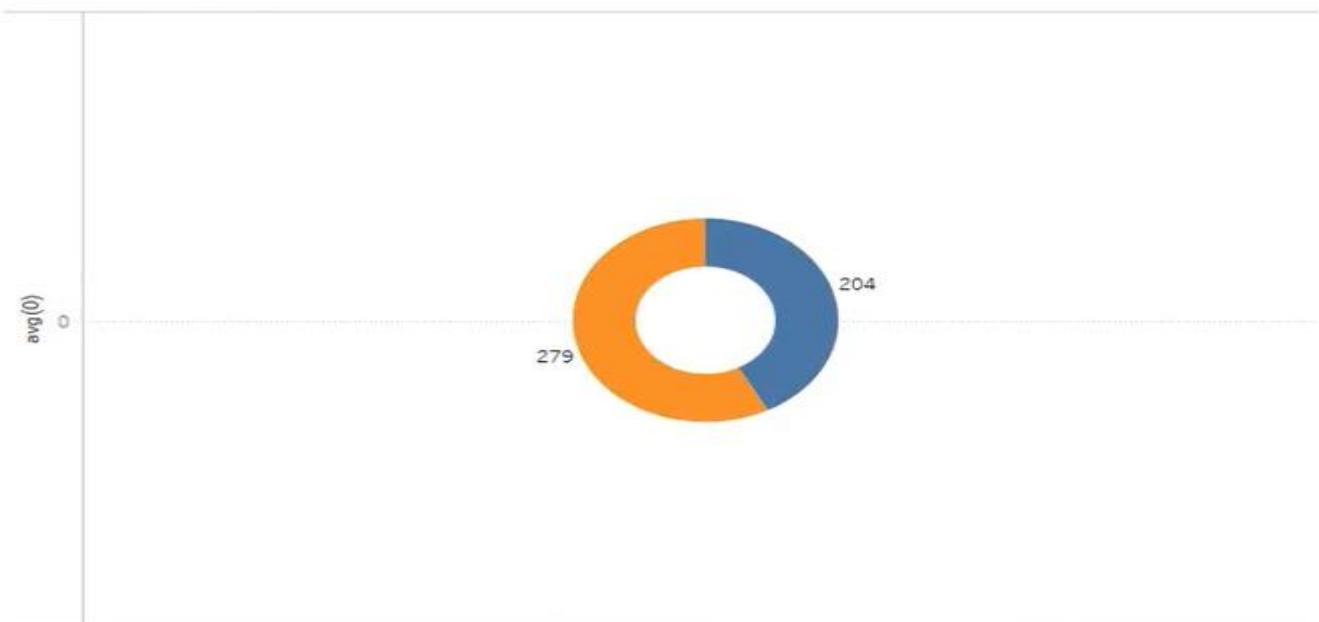
General Health vs Heart Disease



8. Physical Activity vs Heart Disease

[Demo Link](#)

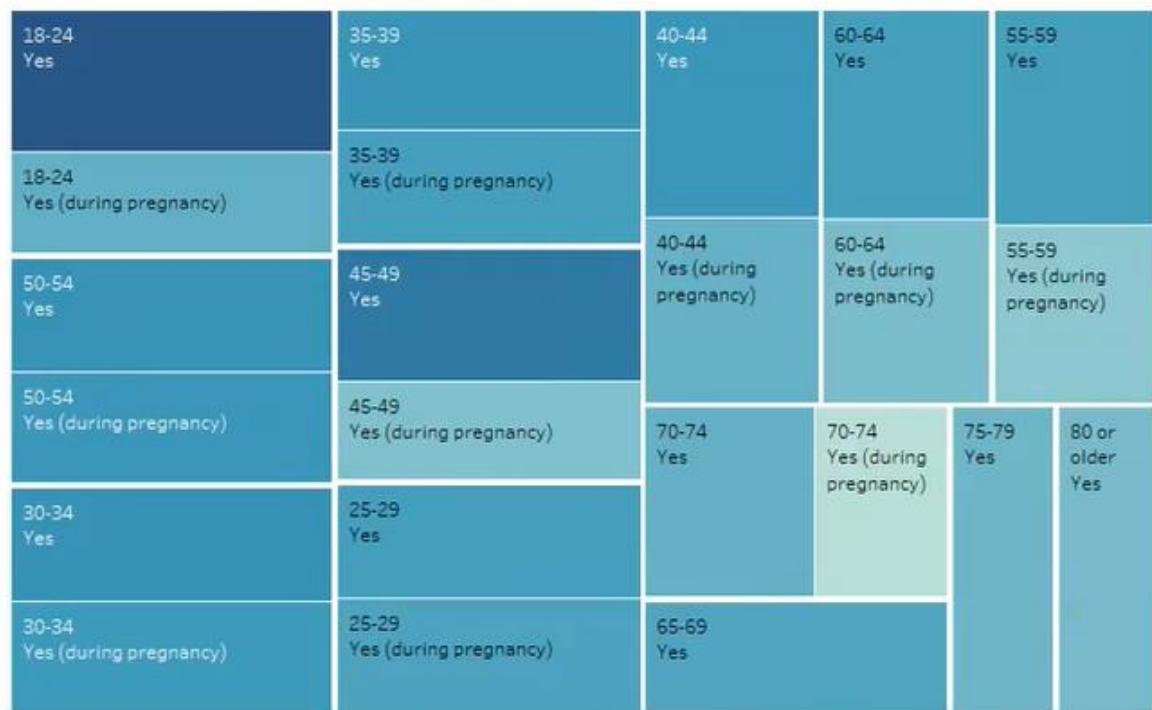
Physical Activity Vs Heart Disease



9. Age vs BMI vs Diabetic

[Demo Link](#)

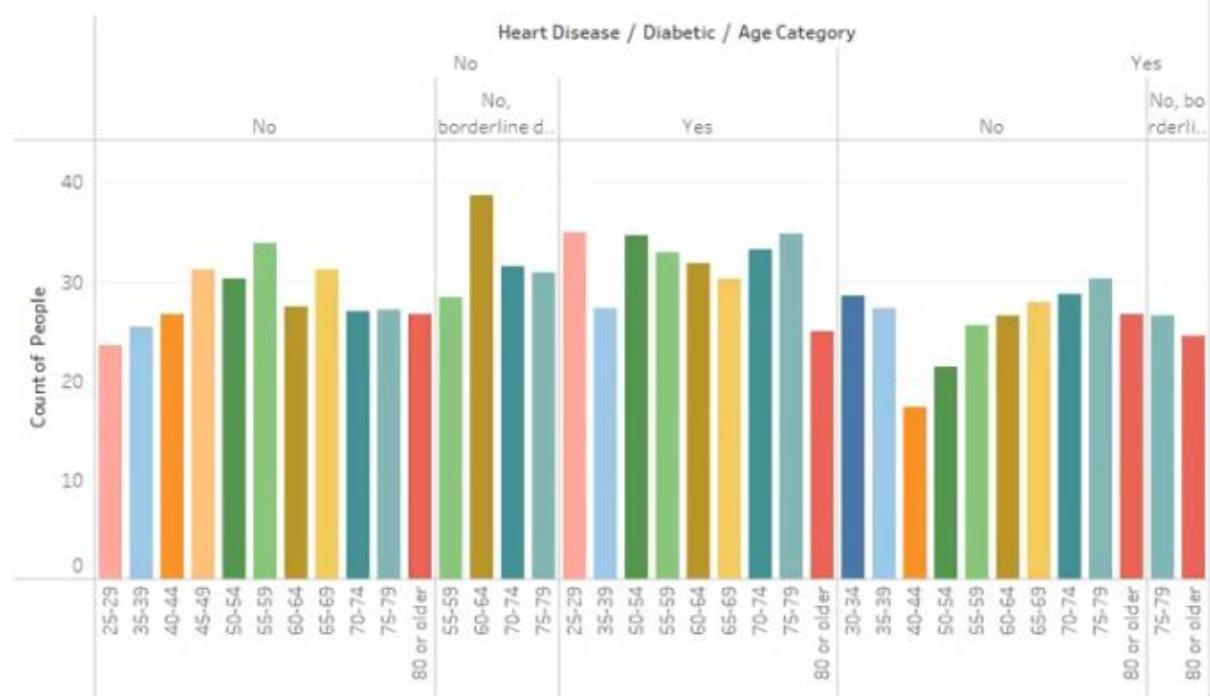
Age vs BMI vs Diabetic



10. People got stroke suffering from Heart Disease and Diabetic

[Demo Link](#)

People Got Stroke Suffering from Diabetes and Heart Disease

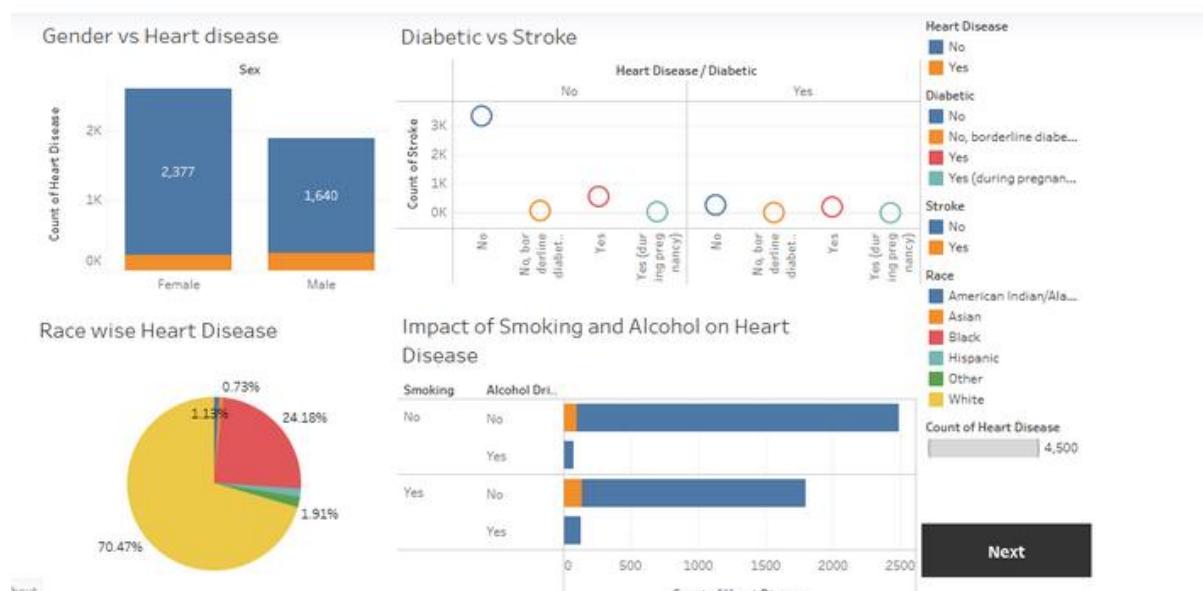


MILESTONE 4 : DASHBOARD

A dashboard is a graphical user interface (GUI) that displays information and data in an organized, easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data and are typically designed for a specific purpose or use case. Dashboards can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables.

ACTIVITY 1 : RESPONSIVE AND DESIGN OF DASHBOARD

A responsive dashboard adapts to different screen sizes—desktop, tablet, or phone—so it looks good and is easy to use everywhere. Use flexible layouts, simplify visuals for small screens, keep fonts and colors clear, and test on multiple devices. This ensures everyone can view and interact with your data smoothly.

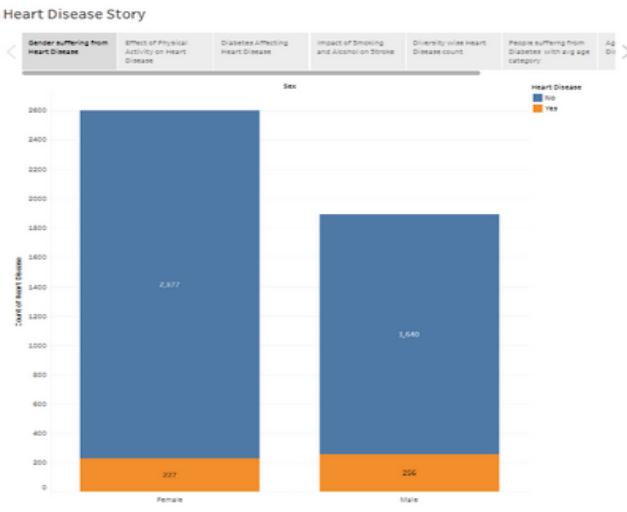


MILESTONE 5 : STORY

A data story is a way of presenting data and analysis in a narrative format, intending to make the information more engaging and easier to understand. A data story typically includes a clear introduction that sets the stage and explains the context for the data, a body that presents the data and analysis logically and systematically, and a conclusion that summarizes the key findings and highlights their implications. Data stories can be told using a variety of mediums, such as reports, presentations, interactive visualizations, and videos.

ACTIVITY 1 : NO OF SCENES OF STORY

The number of scenes in a storyboard for a data visualization analysis of the Heart disease will depend on the complexity of the analysis and the specific insights that are trying to be conveyed. A storyboard is a visual representation of the data analysis process and it breaks down the analysis into a series of steps or scenes.



MILESTONE 6 : PERFORMANCE TESTING

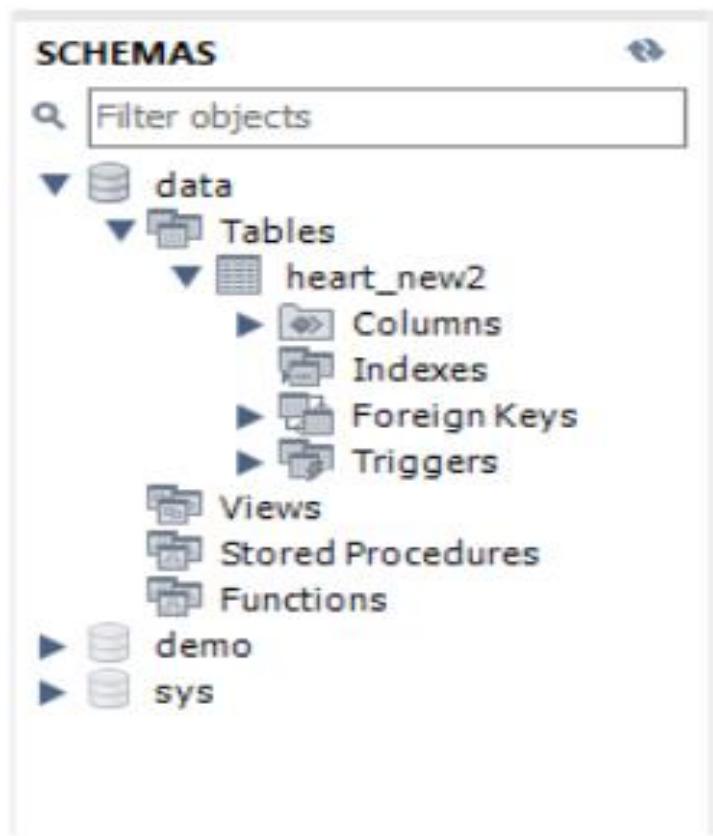
Performance testing involves assessing the volume of data rendered from the database, the impact of data filters on system responsiveness, and the complexity introduced by the number of visualizations. Optimizing these factors ensures the dashboard operates efficiently, providing timely and reliable insights.

ACTIVITY 1 :AMOUNT OF DATA TO DB:

Monitor the volume of data being pulled and rendered from the database to ensure queries are optimized and not overloading the system.

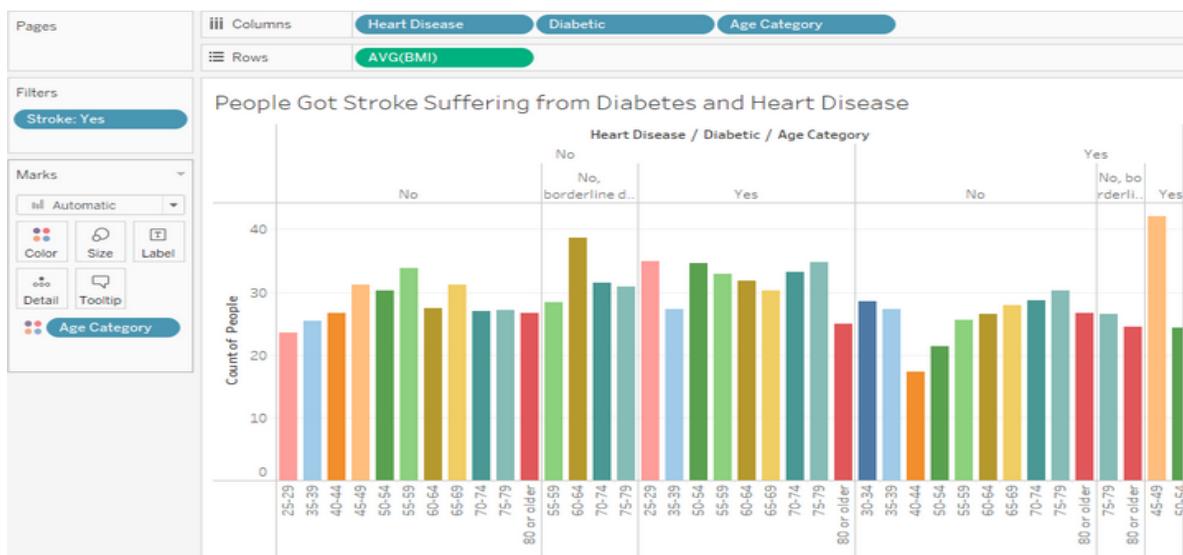
The amount of data that is rendered to a database depends on the size of the dataset and the capacity of the database to store and retrieve data.

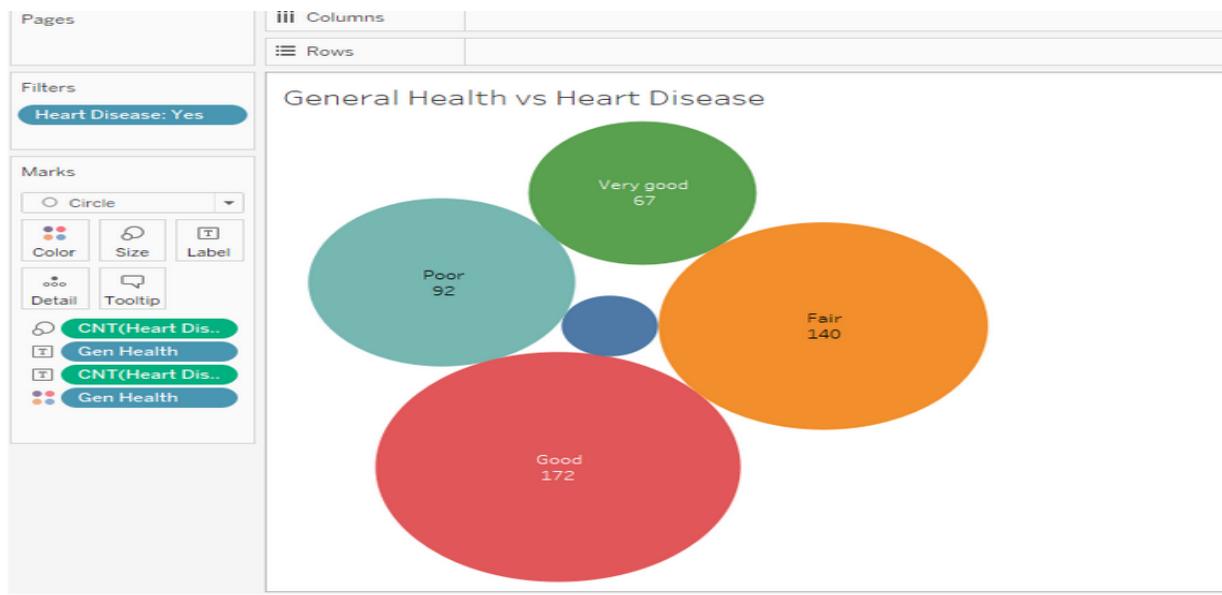
Open the MySQL Workbench, go to the database then click to expand the tables,select the table and click on (i) button to get the information related to table such as column count,table rows etc.



ACTIVITY 2 : UTILIZATION OF DATA FILTERS

Utilization of data filters refers to the effective implementation and management of filtering mechanisms within the Project to refine and focus the dataset. Proper use of filters enhances performance by limiting the volume of data processed and displayed, thereby improving responsiveness. It also enables users to interactively explore specific segments of data, leading to more targeted and meaningful insights.





Attachment

No Story Attachments

No story attachments have been added yet. Please upload attachments to view them here.

ACTIVITY 3 : NO OF CALCULATION FIELDS

In Tableau, a Calculated Field is a custom field you create using formulas to perform computations on your data. It allows you to go beyond the raw fields in your dataset and generate new metrics or dimensions. For example, you might calculate BMI categories, risk scores, or percentage changes.

You can create a calculated field in Tableau by:

1. Right-clicking on the data pane and selecting Create ? Calculated Field.
2. Giving it a name.
3. Writing your formula (e.g., IF [Cholesterol] > 200 THEN "High" ELSE "Normal" END).
4. Clicking OK, after which the new field appears in your data pane.

In our project, we have not created any calculated fields, but you can create them according to the specific insights you want, such as risk group classifications, ratios, or derived health metrics.

ACTIVITY 4 : NO OF VISUALIZATIONS/ GRAPHS

1. Gender wise Heart Disease
2. Age wise Heart Disease

3. People Suffering from Diabetic and Stroke
4. Impact of Smoking and alcohol drinking on heart disease
5. Other Diseases vs Stroke
6. Race wise Heart disease
7. General Health vs Heart Disease
8. Physical activity vs heart disease
9. Age and BMI vs Heart disease
10. People got stroke suffering from Diabetes and Heart disease

MILESTONE 7 : WEB INTEGRATION

Web integration of a Tableau Dashboard Story involves embedding interactive visualizations into a website or web application. This allows users to explore data insights directly within a web interface, enhancing accessibility and engagement. It supports real-time updates, user filtering, and seamless navigation for a dynamic data storytelling experience.

ACTIVITY 1 : PUBLISHING

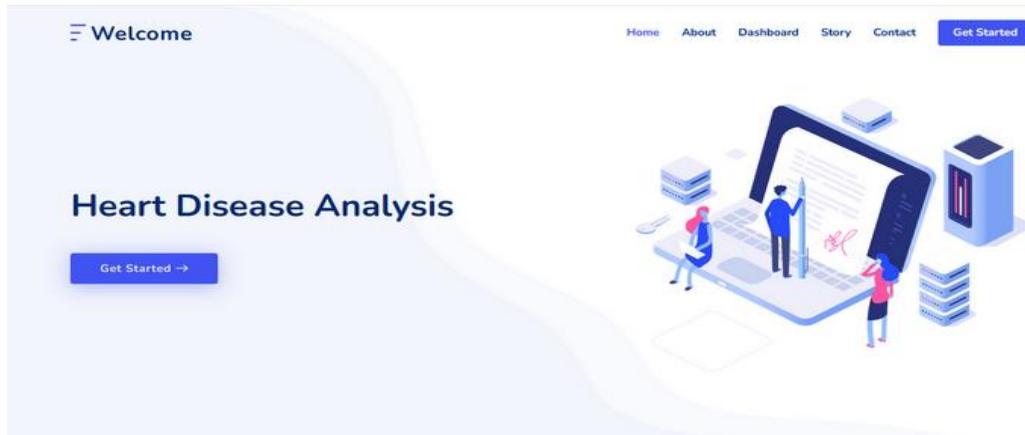
Publishing helps us to track and monitor key performance metrics and to communicate results and progress. help a publisher stay informed, make better decisions, and communicate their performance to others.

1. Prepare Your Dashboard or Story
 - Ensure your dashboard or story is complete and working as expected.
 - Clean up any unnecessary sheets or data to reduce file size.
2. Sign in to Tableau Public
 - In Tableau Desktop, go to File > Save to Tableau Public.
 - If you're not already signed in, a login window will appear.
 - Enter your Tableau Public credentials or sign up if you don't have an account.
3. Save and Publish
 - After logging in, you'll be prompted to name your workbook.
 - Click Save – Tableau will upload the workbook to your Tableau Public profile.
4. View Your Published Dashboard/Story
 - After uploading, your browser will open the published workbook on your Tableau Public profile.

- Here you can:
 - Share the link
 - Embed it into a website
 - Set the workbook to public or hidden

ACTIVITY 2 : DASHBOARD AND STORY EMBED WITH UI WITH FLASK

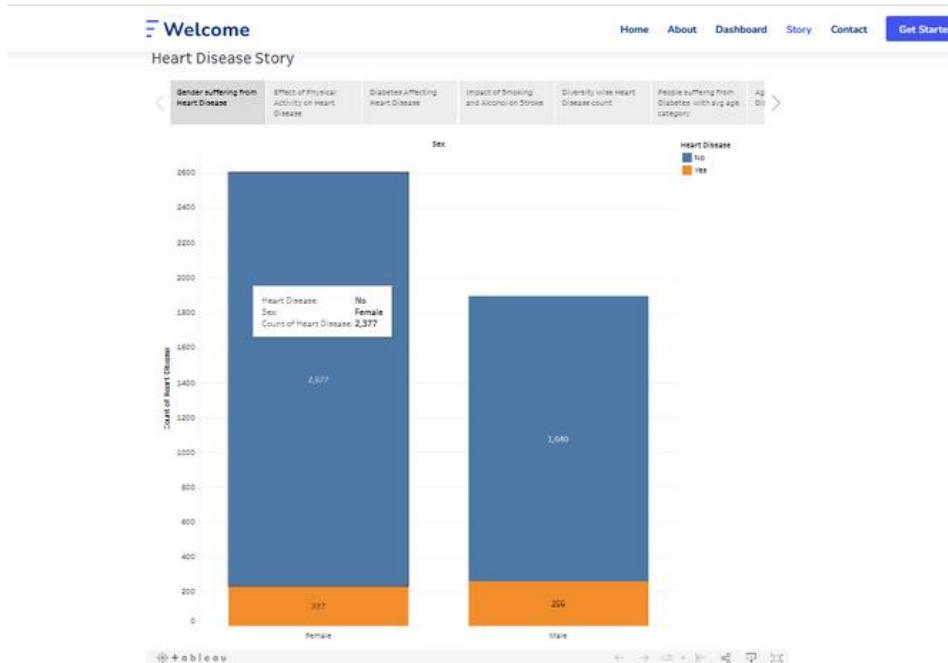
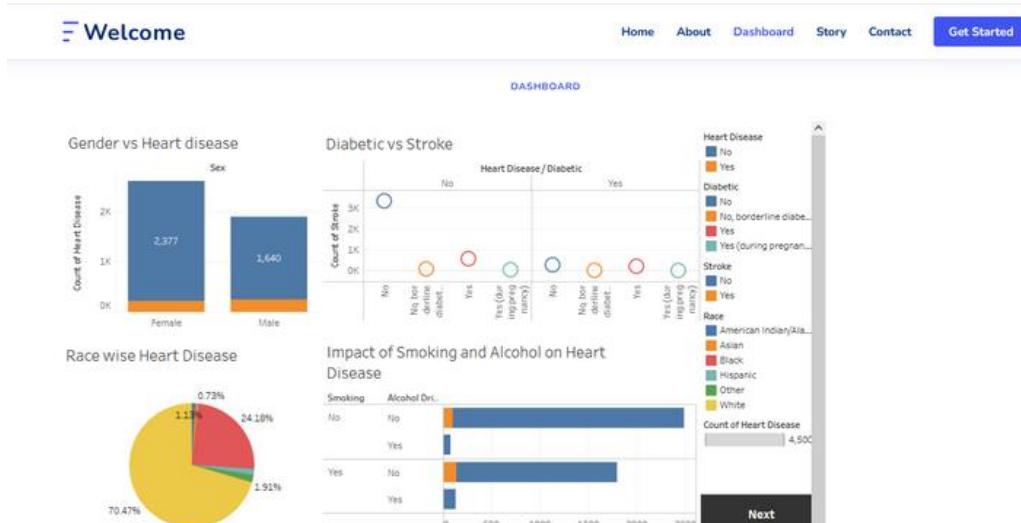
This section demonstrates how to embed Tableau Dashboards and Stories within a Flask web application to provide interactive, real-time access to heart disease insights. By integrating Tableau's powerful visualization capabilities with Flask's lightweight framework, users can explore dashboards directly in a browser, making the analysis more accessible, shareable, and actionable for healthcare professionals, policymakers, and individuals.



Welcome

Heart disease (heart disease) is a group of diseases related to cardiovascular diseases, manifested by a violation of the normal functioning of the heart. May be caused by damage to the epicardium, pericardium, myocardium, endocardium, valvular apparatus of the heart, heart vessels.

According to the National Heart, Lung and Blood Institute in Framingham (USA), the most important factors in the development of cardiovascular disease in humans are obesity, sedentary lifestyle and smoking.



MILESTONE 8 : PROJECT DEMONSTRATION & DOCUMENTATION

Project Demonstration & Documentation involves presenting the project's functionality, features, and outcomes while providing clear written records, diagrams, and explanations to ensure understanding, usability, and reproducibility for stakeholders and future reference.

Below mentioned deliverables to be submitted along with other deliverables

Activity 1:- Record explanation Video for project end to end solution

Activity 2:- Project Documentation-Step by step project development procedure

Create document as per the template provided

OUTPUT :

Clean Data from Excel, CSV, PDF, and Google Sheets with Data Interpreter

Applies to: Tableau Cloud, Tableau Desktop, Tableau Server

When you track data in Excel spreadsheets, you create them with the human interface in mind. To make your spreadsheets easy to read, you might include things like titles, stacked headers, notes, maybe empty rows and columns to add white space, and you probably have multiple tabs of data too.

When you want to analyze this data in Tableau, these aesthetically pleasing attributes make it very difficult for Tableau to interpret your data. That's where Data Interpreter can help.

Tip: Though Tableau's Excel add-in is no longer supported, Data Interpreter can help you reshape your data for analysis in Tableau.

What does Data Interpreter do?

Data Interpreter can give you a head start when cleaning your data. It can detect things like titles, notes, footers, empty cells, and so on and bypass them to identify the actual fields and values in your data set.

It can even detect additional tables and sub-tables so that you can work with a subset of your data independently of the other data.

After Data Interpreter has done its magic, you can check its work to make sure it captured the data that you wanted and identified it correctly. Then, you can make any necessary adjustments.

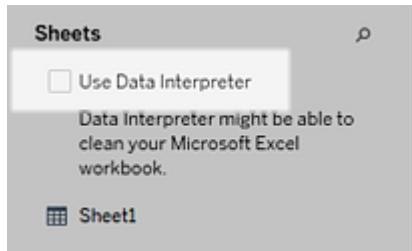
After you select the data that you want to work with, you might also need to do some additional cleaning steps like pivoting your data, splitting fields, or adding filters to get the data in the shape you want before starting your analysis.

Note: If your data needs more cleaning than what Data Interpreter can help you with, try [Tableau Prep](#)(Link opens in a new window).

Turn on Data Interpreter and review results

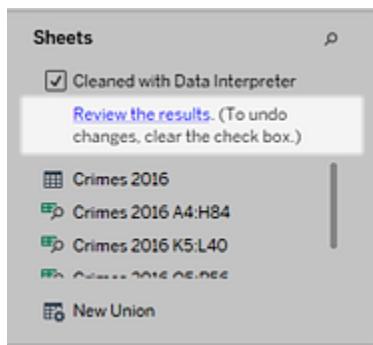
1. From the Connect pane, connect to an Excel spreadsheet or other connector that supports Data Interpreter such as Text (.csv) files, PDF files or Google sheets.

2. Drag a table to the canvas (if needed), then on the Data Source page, in the left pane, select the Use Data Interpreter check box to see if Data Interpreter can help clean up your data.



Note: When you clean your data with Data Interpreter, Data Interpreter cleans all the data associated with a connection in the data source. Data Interpreter does not change the underlying data.

3. In the Data pane, click the Review the results link to review the results of the Data Interpreter.



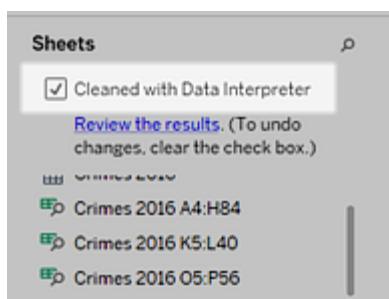
A copy of your data source opens in Excel on the Key for the Data Interpreter tab. Review the key to find out how to read the results.

A	B	C	D	E	F	G	H	I	J	K	L	M	N
1													
2													
3													
4													
5													
6													
7													
8													
9													
10													
11													
12													
13													
14													
15													
16													
17													
18													
19													
20													
21													
22													
23													
24													
25													
26													
27													
28													
29													
30													

4. Click each tab to review how Data Interpreter interpreted the data source.

If Data Interpreter found additional tables, also called found tables or sub-tables, they are identified in the <sheet name>_subtables tab by outlining their cell ranges. A separate tab is also included for each sub-table, color coded to identify the header and data rows.

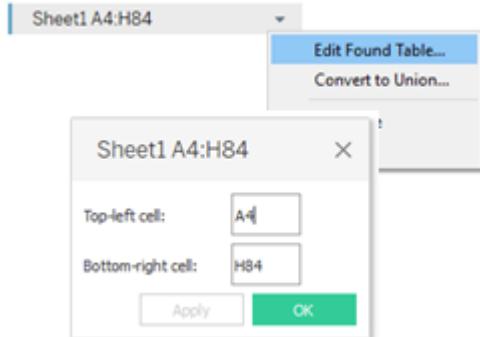
If Data Interpreter does not provide the expected results, clear the Cleaned with Data Interpreter check box to use the original data source.



5. To replace the current table with any of the found tables, drag the current table off the canvas and then drag the found table that you want to use to the canvas.

If Data interpreter has misidentified the range of the found table, after you drag the found table to the canvas, click the drop-down arrow on that table, and then

select Edit Found Table to adjust the corners of the found table (the top-left cell and bottom-right cell of the table).



- After you have the data that you want to work with, you can apply any additional cleaning operations to your data so that you can analyze it.

Data Interpreter Example

In this example we are connecting to an Excel spreadsheet with violent crime data by city and state for the year 2016. This spreadsheet includes multiple tables on one sheet and some extra formatting.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Violent Crimes in 2016 In the United States by City and State																
2																	
3																	
4	Location		Months														
5	city	state	Apr	Jun	Jul	Aug	Sep	Oct									
6	Albuquerque	New Mexico						46									
7	Anaheim	California							4								
8	Anchorage	Alaska								26							
9	Arlington	Texas								17							
10	Atlanta	Georgia									85						
11	Aurora	Colorado									16						
12	Austin	Texas									28						
13	Bakersfield	California									22						
14	Baltimore	Maryland										230					
15	Boston	Massachusetts										28					
16	Buffalo	New York										38					
17	Chandler	Arizona										3					
18	Charlotte-M	North Carolina									25						
19	Chicago	Illinois										536					
20	Chula Vista	California										1					
21	Cincinnati	Ohio										50					
22	Cleveland	Ohio										89					
23	Colorado Sp.	Colorado										15					
24	Columbus	Ohio										70					
25	Corpus Chris	Texas										9					
26	Dallas	Texas										118					
27	Denver	Colorado										33					
28	Detroit	Michigan										5					
29	Durham	North Carolina											221				
30	El Paso	Texas											30				
31	Fort Wayne	Indiana											14				
32	Fort Worth	Texas										7					
33	Fresno	California											19				
34	Greensboro	North Carolina												20			
														Total Crimes 2016			
														State	Population 2016		
														Alabama	4860545		
														Alaska	741522		
														Arizona	6908642		
														Arkansas	2588231		
														California	39296476		
														Colorado	5530105		
														Connecticut	3587685		
														Delaware	952698		
														District of Co	684336		
														Florida	20656589		
														Georgia	10313620		
														Hawaii	1428683		
														Idaho	1680026		
														Illinois	12835726		
														Indiana	6634007		
														Iowa	3130869		
														Kansas	2907731		
														Kentucky	4436113		
														Louisiana	4686157		
														Maine	133032		
														Maryland	6024752		
														Massachuset	6823721		
														Michigan	9933445		
														Minnesota	5525050		
														Mississippi	2985415		
														Missouri	6091176		
														Montana	1038656		
														Oregon	1407603		
														Pennsylvania	2593254		
														Nevada			

- A. Title
- B. Merged header cells
- C. Extra white space
- D. Sub-tables

The extra formatting in this spreadsheet makes it difficult for Tableau to determine what the field headers and values are.

Instead, it reads the data vertically and assigns each column the default value F1, F2, F3 (Field 1, Field 2, Field 3) and so on. Blank cells are read as null values.

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21	F22	F23	F24	F25	F26
Violent Crimes in 2016	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null
Location	null	Marine	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null
City	State	Apr	Jun	Jul	Aug	Sep	Oct	state	Total Crimes 2016	State	Population 2016														
Albuquerque	New Mexico	null	null	null	46	null	Alabama	12	Alabama	4360545															
Anchorage	California	1	null	null	26	null	Alaska	26	Alaska	741522															
Anchorage	Alaska	1	null	null	26	null	Arizona	132	Arizona	6906642															
Arlington	Texas	null	null	null	17	null	California	515	Arkansas	2988231															
Atlanta	Georgia	null	null	null	85	null	Colorado	64	California	39296476															

To see if Data Interpreter can help clean this data set, we select Use Data Interpreter.

Data Interpreter detected the proper headings for the fields, removed the extra formatting and found several sub-tables. The sub-tables are listed in the Sheets section in the Data pane and are named using the original sheet name and the cell ranges for each sub-table.

In this example there are three sub-tables: Crimes 2016 A4:H84, Crimes 2016 K5:L40, and Crimes 2016 O5:P56.

Location City	Location State	Month Apr	Month Jun	Month Jul	Month Aug	Month Sep	Month Oct	State	Total Crimes 2016	State	Population 2016	
Albuquerque	New Mexico	null	null	46	null	Alabama	12	Alabama	4360545			
Anchorage	California	1	4	null	null	null	26	Alaska	741522			
Anchorage	Alaska	1	null	null	26	null	Arizona	132	Arizona	6906642		
Arlington	Texas	null	null	17	null	null	California	515	Arkansas	2988231		
Atlanta	Georgia	null	null	85	null	Colorado	64	California	39296476			
Boulder	Colorado	null	null	16	null	126	Colorado	126	Colorado	830,126		
Austin	Texas	null	null	28	null	null	Florida	210	Connecticut	3,537,685		
Bakersfield	California	null	22	null	null	null	Georgia	85	Delaware	952,698		
Baltimore	Maryland	null	null	230	null	230	Hawaii	9	District of Columbia	684,336		

To examine the results of the Data Interpreter more closely, we click the Review the results link in the Data pane to view an annotated copy of the spreadsheet.

Here we see a copy of the original data, color coded to identify which data was identified as header data and which data was identified as field values.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1 Violent Crimes in 2016 in the United States by City and State																
2																
3																
4																
5 city	Location state	Location	Months	state	Total Crimes 2016	State	Population	Header	Header	Header						
6 Albuquerque	New Mexico									Alabama	12	Alabama	4860545	Data		
7 Anaheim	California									Alaska	26	Alaska	741522	Data		
8 Anchorage	Alaska									Arizona	132	Arizona	6908642	Data		
9 Arlington	Texas									California	515	Arkansas	2988231	Data		
10 Atlanta	Georgia									Colorado	64	California	39296476	Data		
11 Aurora	Colorado									D.C.	105	Colorado	5530105	Data		
12 Austin	Texas									Florida	210	Connecticut	3587685	Data		
13 Bakersfield	California									Georgia	85	Delaware	952698	Data		
14 Baltimore	Maryland									Hawaii	9	District of Columbia	684336	Data		
15 Boston	Massachusetts									Illinois	536	Florida	20656589	Data		
16 Buffalo	New York									Indiana	151	Georgia	10313620	Data		
17 Chandler	Arizona									Kansas	10	Hawaii	1428683	Data		

The next tab shows us the sub-tables that Data Interpreter found, outlined by the cell ranges.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1 Violent Crimes in 2016 in the United States by City and State																
2																
3																
4																
5 city	Location state	Location	Months	state	Total Crimes 2016	State	Population	Header	Header	Header						
6 Albuquerque	New Mexico									Alabama	12	Alabama	4860545	Data		
7 Anaheim	California									Alaska	26	Alaska	741522	Data		
8 Anchorage	Alaska									Arizona	132	Arizona	6908642	Data		
9 Arlington	Texas									California	515	Arkansas	2988231	Data		
10 Atlanta	Georgia									Colorado	64	California	39296476	Data		
11 Aurora	Colorado									D.C.	105	Colorado	5530105	Data		
12 Austin	Texas									Florida	210	Connecticut	3587685	Data		
13 Bakersfield	California									Georgia	85	Delaware	952698	Data		
14 Baltimore	Maryland									Hawaii	9	District of Columbia	684336	Data		
15 Boston	Massachusetts									Illinois	536	Florida	20656589	Data		
16 Buffalo	New York									Indiana	151	Georgia	10313620	Data		
17 Chandler	Arizona									Kansas	10	Hawaii	1428683	Data		
18 Charlotte	- North Carolina									Kentucky	95	Idaho	1680026	Data		
19 Chicago	Illinois									Louisiana	127	Illinois	12835726	Data		
20 Chula Vista	California									Maryland	230	Indiana	6634007	Data		
21 Cincinnati	Ohio									Massachusetts	28	Iowa	3130869	Data		
22 Cleveland	Ohio									Michigan	221	Kansas	2907731	Data		
23 Colorado	: Colorado									Minnesota	26	Kentucky	4436113	Data		
24 Columbus	Ohio									Missouri	223	Louisiana	4686157	Data		
25 Corpus Christi	Texas									Nebraska	29	Maine	1330232	Data		
26 Dallas	Texas									Nevada	128	Maryland	6024752	Data		
27 Denver	Colorado									Massachusetts	66	Michigan	6823721	Data		
28 Detroit	Michigan									New Mexico	46	Michigan	9933445	Data		
29 Durham	North Carolina									New York	290	Minnesota	5525050	Data		
30 El Paso	Texas									North Carolina	82	Mississippi	2985415	Data		
31 Fort Wayne	Indiana									Ohio	217	Missouri	6091176	Data		

In this example the first sub-table, Crimes 2016 A4:H84, has the main data that we want to work with. To use this table as our data table, we can simply drag the original table off the canvas and then drag the new table to the canvas.

The screenshot shows the Microsoft Power BI Data Editor interface. On the left, the 'Connections' pane displays a single connection named 'Crimes_2016' from Microsoft Excel. Below it, the 'Sheets' pane lists several sub-tables: 'Crimes 2016', 'Crimes 2016 A4:H84', 'Crimes 2016 K5:L40', and 'Crimes 2016 A4:H84 (crimes_2016)'. A checked checkbox indicates the data was cleaned with the Data Interpreter. The main workspace shows a table titled 'Crimes 2016 A4:H84' with the following schema:

Location city	Location state	Months Apr	Months Jun	Months Jul	Months Aug	Months Sep	Months Oct
Albuquerque	New Mexico	null	null	null	null	46	null
Anaheim	California	null	4	null	null	null	null
Anchorage	Alaska	1	null	null	null	26	null
Arlington	Texas	null	null	null	17	null	null
Atlanta	Georgia	null	null	null	null	85	null
Aurora	Colorado	null	null	null	null	16	null
Austin	Texas	null	null	null	28	null	null
Bakersfield	California	null	22	null	null	null	null
Baltimore	Maryland	null	null	null	null	null	230
Boston	Massachusetts	null	null	null	null	28	null
Buffalo	New York	null	null	null	null	38	null
Chandler	Arizona	null	null	null	null	3	null

Once we have the data that we want to work with in the canvas, we can do some additional clean up on the data. For example we can:

- Change the field names so that they represent city, state, and month names.
- Pivot the months fields.
- Drag in the third sub-table Crimes 2016 o5:P56 and join it to our first sub-table on the State field to include state populations for our analysis.
- Hide any duplicate fields that were added as a result of the join.

The results might look something like this:

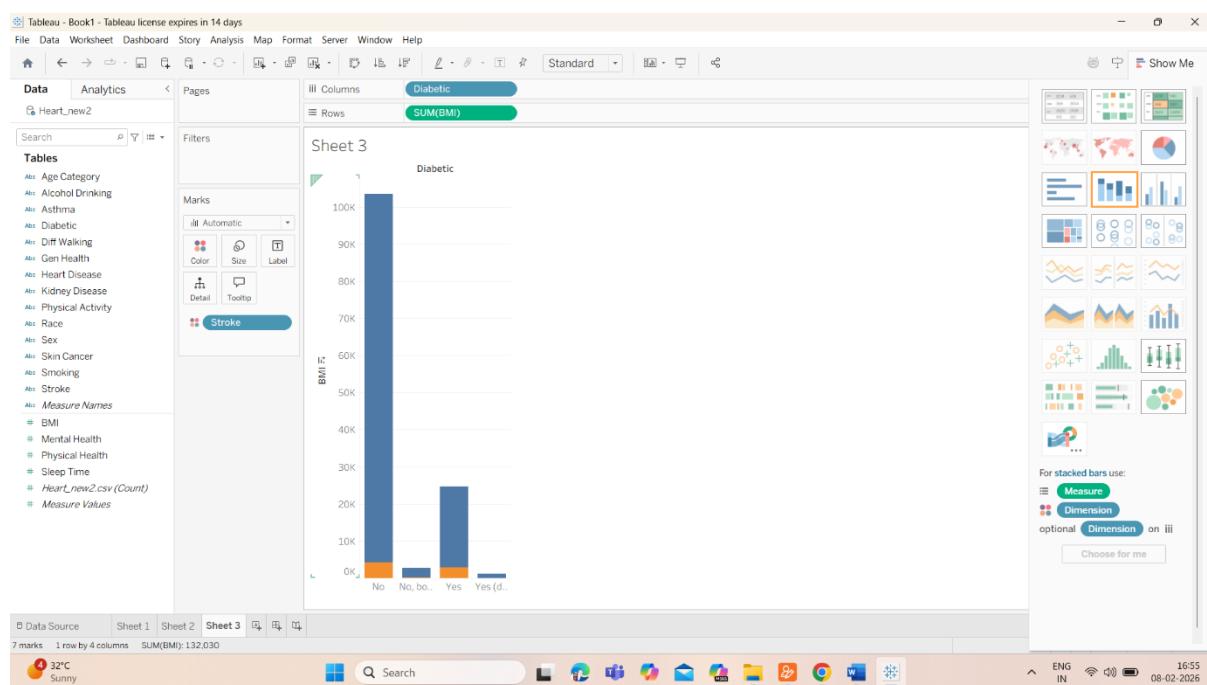
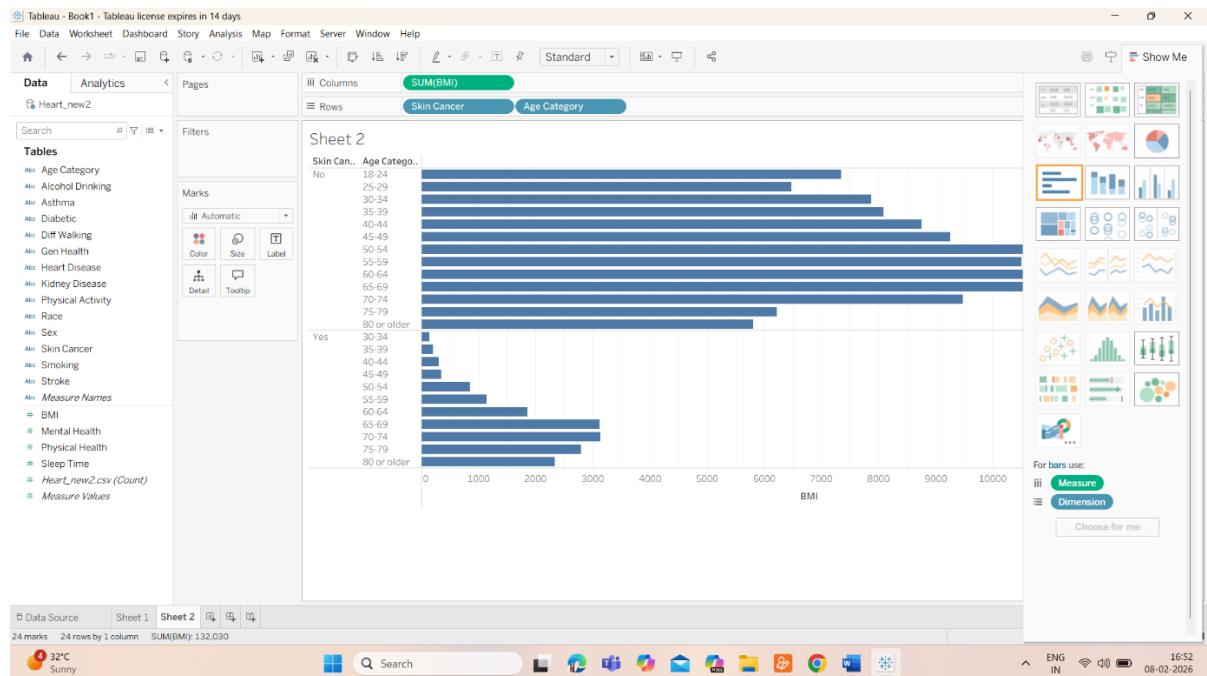
City	State	Population 2016	Month	Crimes
Phoenix	Arizona	6,908,642	August	111
Pittsburgh	Pennsylvania	12,787,085	August	null
Plano	Texas	27,904,862	August	5
Portland	Oregon	4,085,989	August	null
Raleigh	North Carolina	10,156,689	August	null
Riverside	California	39,296,476	August	7
Sacramento	California	39,296,476	August	null
San Antonio	Texas	27,904,862	August	null
San Diego	California	39,296,476	August	30
San Francisco	California	39,296,476	August	null
San Jose	California	39,296,476	August	35
Santa Ana	California	39,296,476	August	null
Seattle	Washington	7,280,934	August	14
St. Louis	Missouri	6,091,176	August	133
St. Petersburg	Florida	20,656,589	August	14

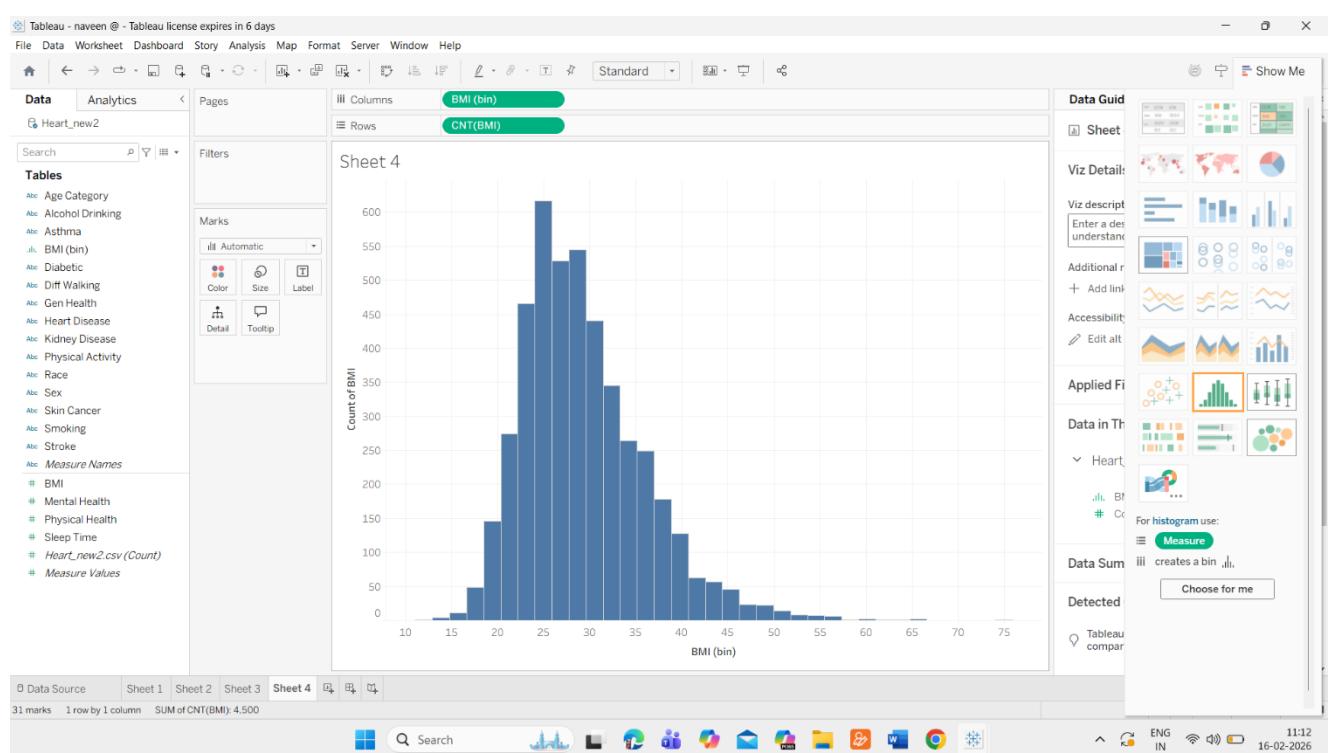
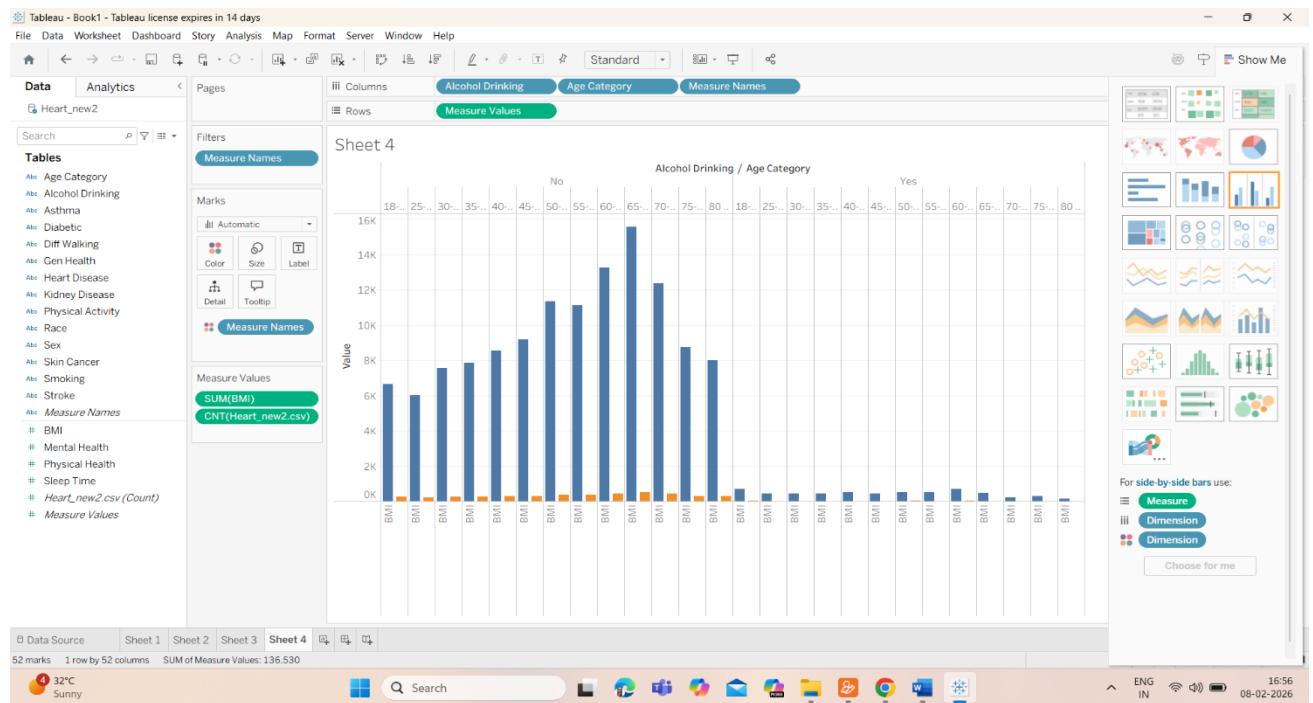
Now we are ready to start analyzing our data in Tableau.

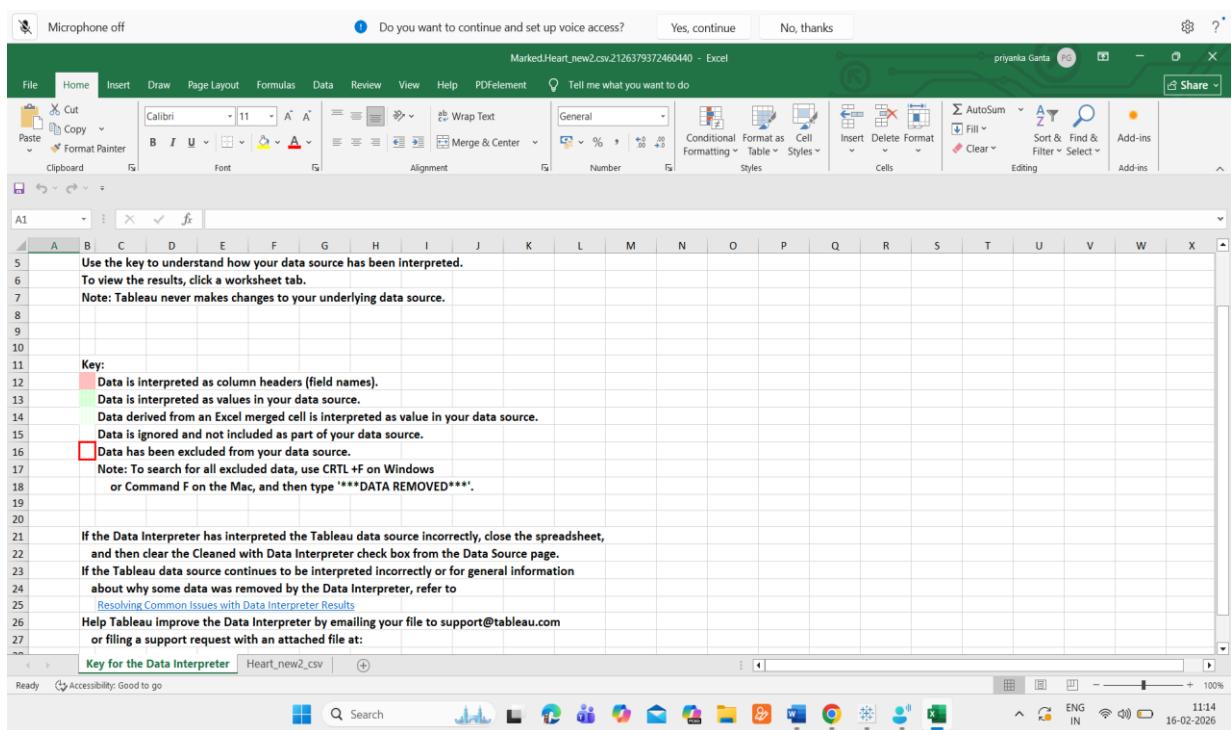
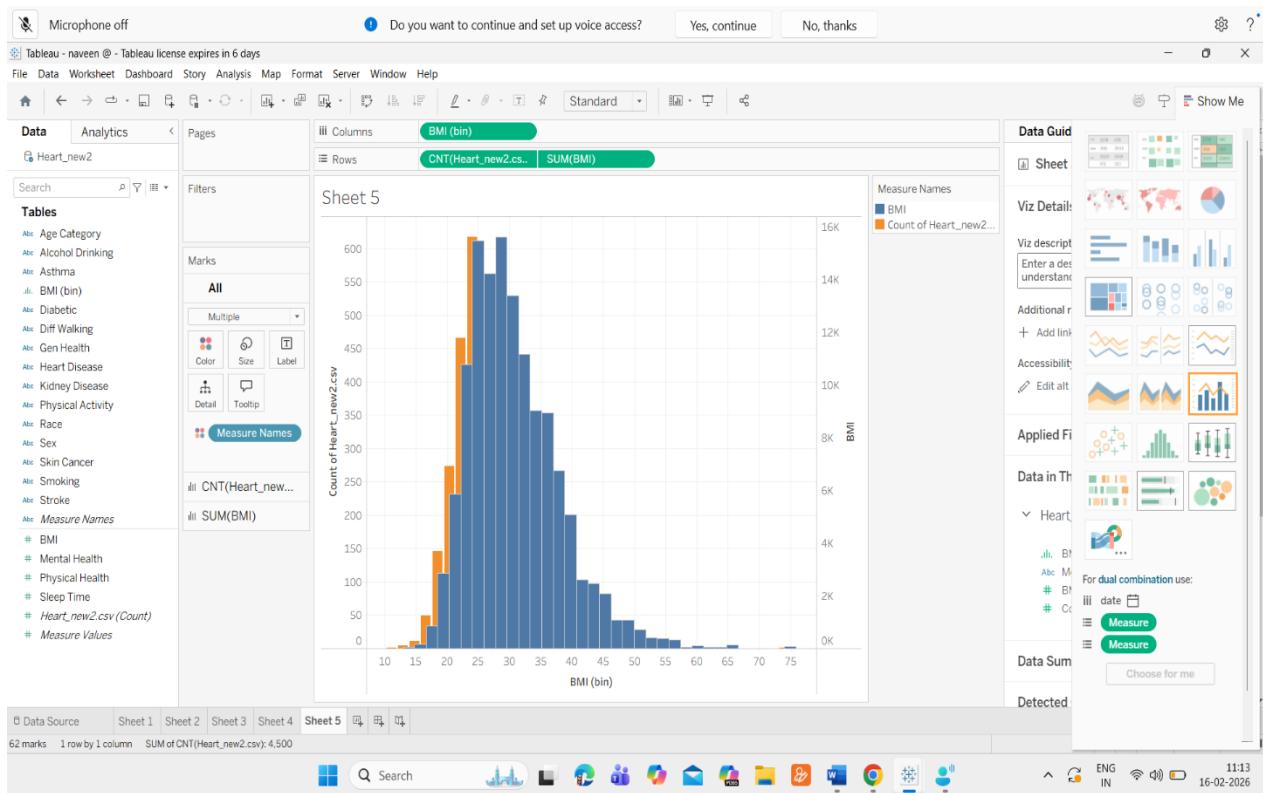
When Data Interpreter is not available

The Data Interpreter option might not be available for the following reasons:

- The data source is already in a format that Tableau can interpret: If Tableau Desktop doesn't need extra help from Data Interpreter to handle unique formatting or extraneous information, the Data Interpreter option is not available.
- Many rows or many columns: The Data Interpreter option is not available when your data has the following attributes:
 - Data contains more than 2000 columns.
 - Data contains more than 3000 rows and more than 150 columns.
- The data source is not supported: Data Interpreter is only available for Microsoft Excel, Text (.csv) files, PDF files and Google Sheets. For Excel, your data must be in the .xls or .xlsx format.







GROUP PHOTO



DEMO LINK :

https://drive.google.com/file/d/1_NhdIjF4HXr1sTDElq3QrBiXuvF-6wbN/view?usp=drivesdk

