Bharatiya Vidya Bhavan’s

SARDAR PATEL INSTITUTE OF TECHNOLOGY

Advanced Data Visualization

Experiment no. 3

<u>**Submitted To**</u>

Prof. Pranav Nerurkar

<u>**Submitted By**</u>

Name: Aakriti Pathak

UID: 2021300094

Batch: BE Comps (Batch B)

1. Aim:

Design Interactive Dashboards and Storytelling using Tableau / Power BI / R (Shiny) / Python (Streamlit/Flask) / D3.js to be performed on the dataset - Disease spread / Healthcare

- Create interactive dashboard - Write observations from each chart given below
- (Advanced - Word chart, Box and whisker plot, Violin plot, Regression plot (linear and nonlinear), 3D chart, Jitter, Line, Area, Waterfall, Donut, Treemap, Funnel
- Basic - Bar chart, Pie chart, Histogram, Timeline chart, Scatter plot, Bubble plot)

2. Procedure Description:

Step-1: Dataset:

You can view the dataset from this link.

Step-2: Description:

The **Heart Disease UCI Dataset** is a widely used medical dataset designed to predict the presence of heart disease in patients based on a variety of clinical and physiological attributes. It consists of 303 rows with 14 attributes, including features like **age**, **sex**, **chest pain type (cp)**, **resting blood pressure (trestbps)**, **serum cholesterol levels (chol)**, **maximum heart rate achieved (thalach)**, and the presence of **exercise-induced angina (exang)**, among others. The target variable, **target**, indicates whether a patient has heart disease (1) or not (0). This dataset is instrumental in understanding correlations between patient attributes and heart disease, making it ideal for machine learning models and exploratory data analysis.

Step-3: MetaData:

- **age**: Age of the patient
- **sex**: Gender (1 = Male, 0 = Female)
- **cp**: Chest pain type
- **restbps**: Resting blood pressure
- **chol**: Serum cholesterol in mg/dl
- **thalach**: Maximum heart rate achieved
- **ca**: Number of major vessels colored by fluoroscopy
- **target**: Diagnosis of heart disease (1 = Disease, 0 = No Disease)

Step-4: Data Modeling - Star Schema:

In a **Star Schema**, data is organized into two types of tables:

1. **Fact Table**: The central table that contains quantitative data (i.e., measures or metrics) that are being analyzed, like sales, performance, or other aggregations. Fact tables usually have foreign key references to dimension tables.
2. **Dimension Tables**: These are smaller tables that provide context and descriptions for the facts. Dimension tables contain descriptive attributes like time, product details, customer information, etc.

The **Star Schema** is named for its structure, with the fact table at the center and dimension tables surrounding it, resembling a star.

Components of a Star Schema:

- **Fact Table**: Contains foreign keys that reference dimension tables and usually holds the numeric values to be analyzed (e.g., sales amount, quantity, cholesterol levels, etc.).
- **Dimension Tables**: Each dimension table contains information that helps describe the data in the fact table, such as age, sex, and chest pain type in this case.

Explanation of the Above Star Schema:

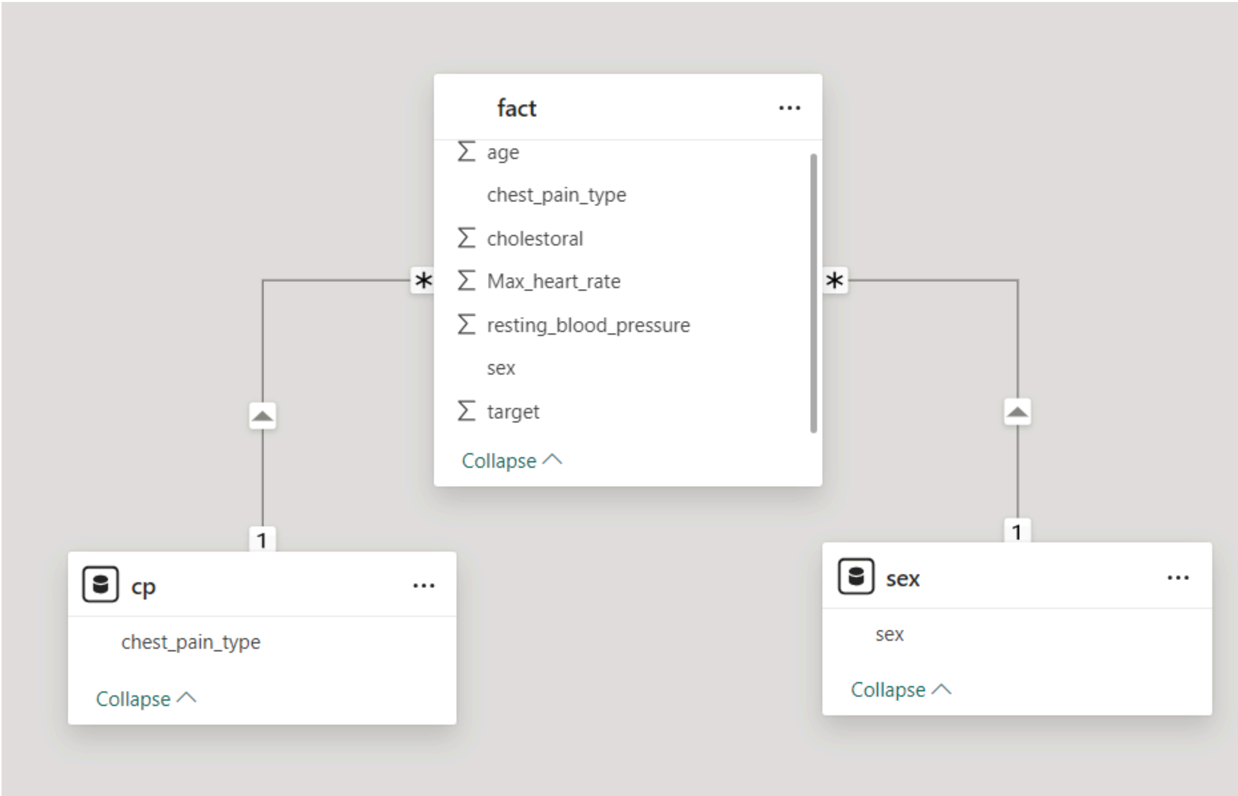
1. **Fact Table (`fact`)**:
   - This is the central table in the schema that stores the metrics or measures. In this case, it holds health-related metrics such as:
     - **age**: The age of individuals in the dataset.
     - **chest_pain_type**: The type of chest pain experienced by the individual (foreign key referencing the `cp` table).
     - **cholesterol**: The cholesterol level of the individual.
     - **Max_heart_rate**: The maximum heart rate recorded.
     - **resting_blood_pressure**: The resting blood pressure value of the individual.
     - **sex**: The sex of the individual (foreign key referencing the `sex` table).
     - **target**: The target variable (usually whether the individual has heart disease or not).
2. **Dimension Tables**:

- **Chest Pain Type (`cp`)**: This dimension table contains a description of different types of chest pain, such as:
  - **Asymptomatic**
  - **Atypical Angina**
  - **Non-Anginal Pain**
  - **Typical Angina**

3. The `fact` table references this table to describe the type of chest pain experienced by individuals.
   - **Sex (`sex`)**: This dimension table holds the gender of the individuals, with possible values like:
     - **Male**
     - **Female**
4. The `fact` table uses the `sex` column as a foreign key to refer to the `sex` dimension.

Schema Structure:

- The `fact` table is connected to the `cp` and `sex` tables by **foreign keys**, meaning each entry in the `fact` table corresponds to an entry in the `cp` table (for chest pain type) and the `sex` table (for the individual's gender). This allows users to analyze measures in the `fact` table (e.g., cholesterol, heart rate, etc.) based on these descriptive dimensions.



Step-5: Data Visualization Analysis:

This dashboard provides a comprehensive view of key factors related to heart disease risk, using various data visualizations to break down patterns in age, gender, cholesterol, chest pain type, and other health indicators. Each visualization contributes to a deeper understanding of the relationships between these variables and the target condition (heart disease), aiding in identifying potential trends, risk groups, and factors that may contribute to the diagnosis of heart disease.

Key Sections of the Dashboard:

1. **Sum of Target by Sex (Bar Chart & Donut Chart)**
   - **Purpose**: To highlight the distribution of the target condition (heart disease) across sex (male and female).
   - **Insight**: This shows that a higher percentage of males in the dataset are affected by the target condition. Gender may play an important role in heart disease prevalence, and this distribution could indicate that men are more prone to heart disease, at least within the scope of this data.

2. **Sum of Target by Chest Pain Type (Pie Chart)**
   - **Purpose**: To categorize individuals based on different types of chest pain and their association with the target condition.
   - **Insight**: "Typical Angina" and "Non-Anginal Pain" are the most common chest pain types among individuals with heart disease, highlighting these as potential diagnostic markers for cardiovascular conditions.
3. **Count of Age by Age (Histogram)**
   - **Purpose**: To show the age distribution of individuals in the dataset, with a focus on how many individuals fall into different age brackets.
   - **Insight**: The majority of individuals are between the ages of 50 and 60, indicating that this age group is more likely to be at risk of heart disease. This age distribution provides insights into the age at which individuals are most likely to experience heart disease symptoms or diagnosis.
4. **Funnel Chart: Sum of Target by Age**
   - **Purpose**: To illustrate the number of individuals across different age groups in the target population.
   - **Insight**: There is a gradual decline in the number of individuals affected by heart disease as age increases, with a significant focus on individuals aged 40-60. This suggests that middle-aged individuals are at a higher risk of heart disease compared to younger and older groups.
5. **Bubble Chart: Sum of Cholesterol, Resting Blood Pressure, and Age by Target**
   - **Purpose**: To explore how cholesterol, resting blood pressure, and age relate to the presence of heart disease (target).
   - **Insight**: Patients with high cholesterol and higher blood pressure are more likely to have heart disease, particularly in older age groups. This reinforces the idea that age, cholesterol, and blood pressure are key factors contributing to the risk of heart disease.
6. **Waterfall Chart: Sum of Target by Chest Pain Type**
   - **Purpose**: To break down the contribution of different chest pain types to the total number of individuals in the target group.
   - **Insight**: "Typical Angina" contributes the most to heart disease diagnosis, highlighting this specific chest pain type as a significant indicator. Other types, like "Asymptomatic" or "Atypical Angina," contribute less, suggesting they may not be as closely linked to heart disease.
7. **Tree Map: Sum of Target by Chest Pain Type and Fasting Blood Sugar**
   - **Purpose**: To show the relationship between chest pain type and fasting blood sugar in the target population.
   - **Insight**: This visualization reveals how chest pain type and fasting blood sugar levels interact, with the largest portion of individuals having "Typical Angina"

and lower fasting blood sugar levels. This combination might be useful in identifying patients at risk of heart disease.

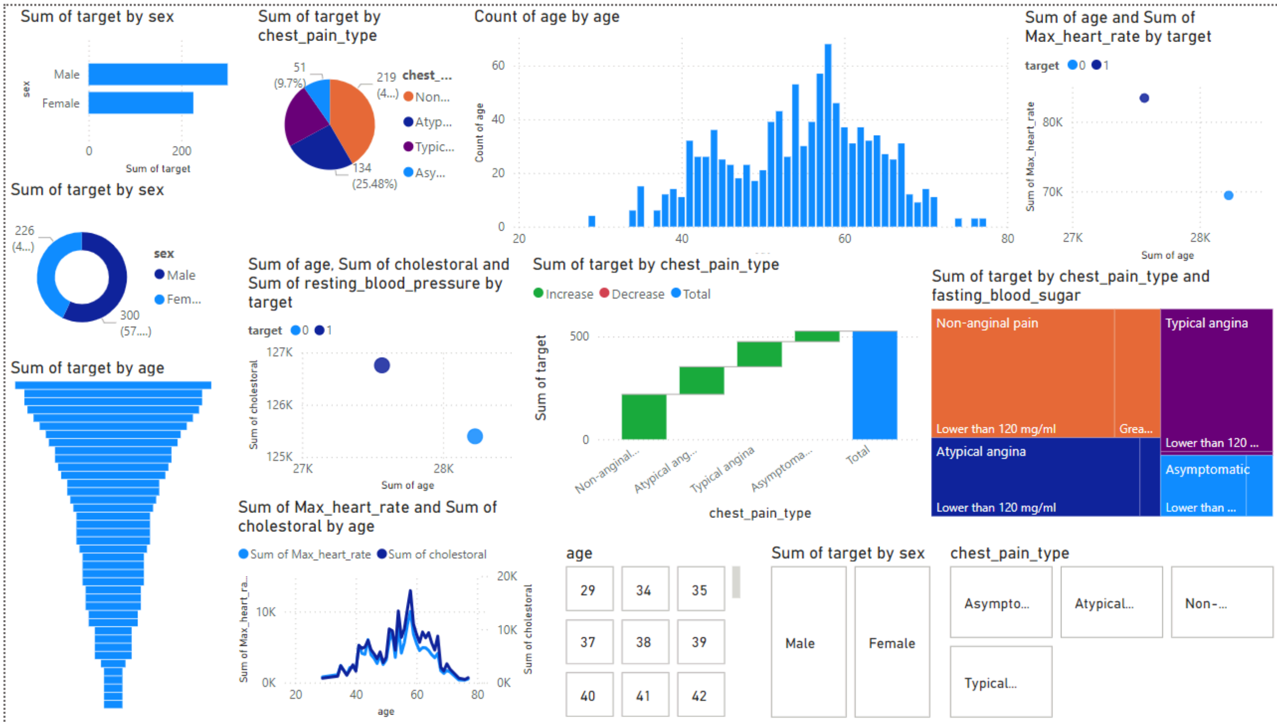
8. **Line Chart: Sum of Max Heart Rate and Cholesterol by Age**
    - **Purpose**: To display trends in maximum heart rate and cholesterol levels by age.
    - **Insight**: Maximum heart rate declines steadily with age, while cholesterol levels fluctuate across different ages. This trend may help doctors understand how cardiovascular function deteriorates with age and how cholesterol levels vary, potentially guiding personalized treatments.
9. **Scatter Plot: Sum of Max Heart Rate and Age by Target**
    - **Purpose**: To show the relationship between max heart rate and age in the target group.
    - **Insight**: There is a clear decline in max heart rate with age, and individuals with lower max heart rates and older ages are more likely to belong to the target group. This scatter plot suggests that max heart rate could be an important factor in identifying heart disease, especially in older individuals.

Dashboard Summary

This dashboard provides a multidimensional analysis of key health metrics—such as age, sex, chest pain type, cholesterol, and heart rate—and their relationships with heart disease. It visually communicates the distribution of these variables across different groups and highlights key factors that may increase the likelihood of heart disease. The insights derived from these visualizations can be used to guide medical professionals in diagnosing and treating heart disease more effectively.

- **Age and gender** appear to be important factors, with more males in the target group and a higher concentration of individuals in the 40-60 age range.
- **Cholesterol, max heart rate, chest pain type**, and **blood pressure** are critical health metrics that demonstrate strong correlations with the target group, indicating their importance in assessing heart disease risk.

By using this dashboard, medical practitioners or researchers can quickly identify patterns in the data and use these patterns to guide clinical decisions or further research.



The dashboard also includes slicers, which are interactive elements that allow users to filter and segment the data dynamically. The slicers used in this dashboard help viewers focus on specific variables such as **age**, **sex**, and **chest pain type**. These slicers are crucial for conducting a more granular analysis by narrowing down subsets of the population based on specific characteristics.

Slicers Included in the Dashboard:

1. **Age Slicer (Age Boxes)**
   - **Purpose**: This slicer allows users to filter the entire dataset by selecting specific age groups. Users can choose one or more age groups, which will automatically update all charts to reflect data only for the selected ages.
   - **Use Case**: For example, a user can select the 40-49 and 50-59 age groups to see the distribution of heart disease risks and related health metrics among middle-aged individuals. This allows for targeted analysis of risk factors prevalent in specific age groups.
2. **Sex Slicer (Male and Female Selection)**
   - **Purpose**: This slicer enables users to filter the dataset based on gender (male or female), updating all visualizations to reflect the data relevant to the selected gender.
   - **Use Case**: By selecting "Female," the user can isolate the analysis to view how heart disease impacts women specifically. This is useful for examining

gender-based trends in heart disease risk, such as whether cholesterol levels or chest pain types differ significantly between men and women.

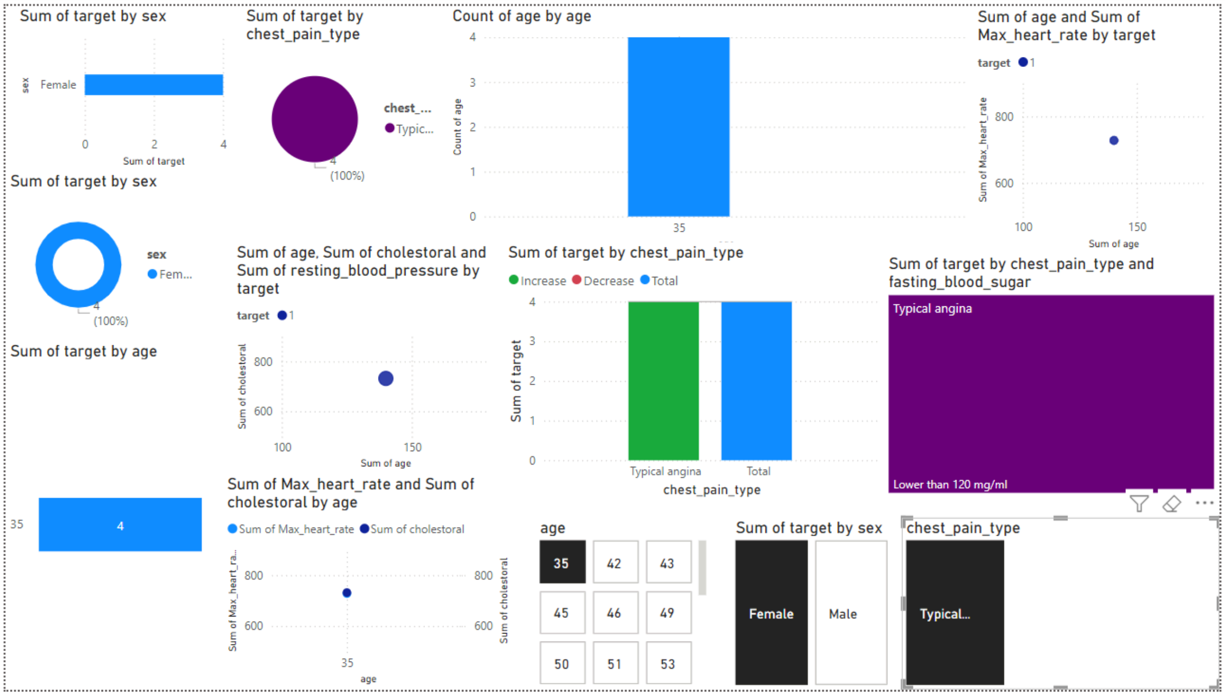
3. **Chest Pain Type Slicer (Asymptomatic, Atypical Angina, Non-Anginal Pain, Typical Angina)**
    - **Purpose**: This slicer lets users filter data based on different types of chest pain, updating visualizations to show how each type correlates with heart disease and other health metrics.
    - **Use Case**: A user can select "Typical Angina" to focus on individuals experiencing this type of chest pain and analyze whether it's more prevalent in certain age groups, genders, or those with specific cholesterol or blood pressure levels. It helps narrow down potential diagnostic markers.

Impact of Slicers on Dashboard Analysis

- **Granular Insights**: The slicers allow users to drill down into specific groups, enabling more focused analysis. For example, by selecting females aged 50-60, the dashboard can provide insights specifically about heart disease risk factors in that demographic, which could differ from the overall dataset.
- **Dynamic Filtering**: As users change slicer selections, the visualizations update dynamically, ensuring the most relevant data is being displayed. This helps in making comparisons between different groups (e.g., how heart disease factors vary between younger and older populations or between men and women).
- **User Interactivity**: The slicers add a layer of interactivity, making the dashboard more engaging. Users can interact with the data in real-time, exploring relationships between variables by adjusting the slicers.

Overall Slicer Summary

The slicers provide powerful control over the dashboard, making it easy for users to filter data based on key dimensions like **age**, **sex**, and **chest pain type**. This functionality enhances the depth of analysis by allowing users to explore specific subgroups in detail, uncovering hidden patterns or trends that might be missed in an aggregate analysis. In a medical context, these slicers enable health professionals to make data-driven decisions tailored to particular patient demographics.


Sum of target by sex
Female
Sum of target
Sum of target by chest_pain_type
(100%)
Count of age by age
Count of age
Sum of age and Sum of Max_heart_rate by target
target
Sum of Max_heart_rate
Sum of age
Sum of target by sex
sex
Fem...
(100%)
Sum of age, Sum of cholestoral and Sum of resting_blood_pressure by target
Sum of cholestoral
Sum of age
Sum of target by chest_pain_type
Increase
Decrease
Total
Sum of target
Typical angina
Total
chest_pain_type
Sum of target by chest_pain_type and fasting_blood_sugar
Typical angina
Lower than 120 mg/ml
Sum of target by age
Sum of Max_heart_rate and Sum of cholestoral by age
Sum of Max_heart_rate
Sum of cholestoral
age
age
35
42
43
45
46
49
50
51
53
Sum of target by sex
Female
Male
chest_pain_type
Typical...