

CONTRIBUTION REPORT – Heart Care Cardiovascular Trends Analysis and Interactive Dashboard

Shivansh Bhatnagar
x23237252
MSCDAD_A
National College of Ireland, Dublin

Other Team Members:- Malav Naik (x23271779) / Pratik Sunar(x23292512)
(Dataset Name:- “Heart_Care_and_Hospital_Data.json”)

I. Summary of Contribution

As a Part of Team5, I have worked on the Healthcare dataset which contains 10,000 records, 10 categorical columns and 6 Numerical columns. I was responsible for performing Data Processing, EDA, Data Cleaning, Data Visualization and further analytics steps. I have started by **integrating data** from **MongoDB**, Data Exploring, Data Cleaning, Data Transformation, pulling relationships between variables, and creating visualizations to showcase meaningful and beautiful insights. Additionally, I have contributed to Statistical Analytics to provide actionable insights by providing decision-making insights for stakeholders by creating **Interactive Dashboard** by combining all three dataset's insights.

II. Tasks and Responsibilities

a) Data Selection

Dataset Used:- Heart_Care_and_Hospital_Data.json

Data Source:- [Heart_Care_and_Hospital_Data](#)

The rationale for selecting this dataset:- The dataset was chosen because it has various type of healthcare parameters just like BMI, Blood Pressure, Glucose level, Cholesterol, and treatment, also dataset contains demographics information and treatment data including USA statewide records. Which helped me in performing various visualizations and extracting meaningful insights from it.

b) Data Preprocessing and Transformation

- Extracted Dataset from MongoDB and converted into dataFrames using Pandas.
- Explored Dataset using Python features like dtypes, info.
- Extracted Month and Year from variable “treatment_date” for temporal analysis.
- Created new Feature by dividing age group into category for further analysis ('0-18', '19-30', '31-45', '46-60', '61-99').
- Created New Feature Health_Score to calculate Health Average per state by using formulae $(\text{row}[\text{'blood_pressure'}] + \text{row}[\text{'cholesterol'}] + \text{row}[\text{'bmi'}] + \text{row}[\text{'glucose_level'}]) / 4$
- Checked for Missing Value, Null Value and For duplicate values.
- Filtered and Transformed the Data for yearly health Analysis.

Tools Used:- MongoDB, Python, Pandas, Numpy, Matplotlib, Dash

c) Database Integration

Database Used:- MongoDB

- Connected to MongoDB server using pymongo to retrieve the data, Framed successful pipeline to process the flow of data and then transform it to DataFrames in jupyter using Pandas. Data was Successfully processed to python for further analytics.

d) Analysis and Data Visualization

1. Created Correlation Heatmap to identify the relationships between the target variables.

Figure 1 is demonstrating relationships between all health parameters, we can see that BMI has good correlation with Blood Pressure (0.52) and Cholesterol (0.36). Similarly Heart_Disease in Figure 2 demonstrating strong correlation with Blood_pressure(0.45) and BMI(0.46).

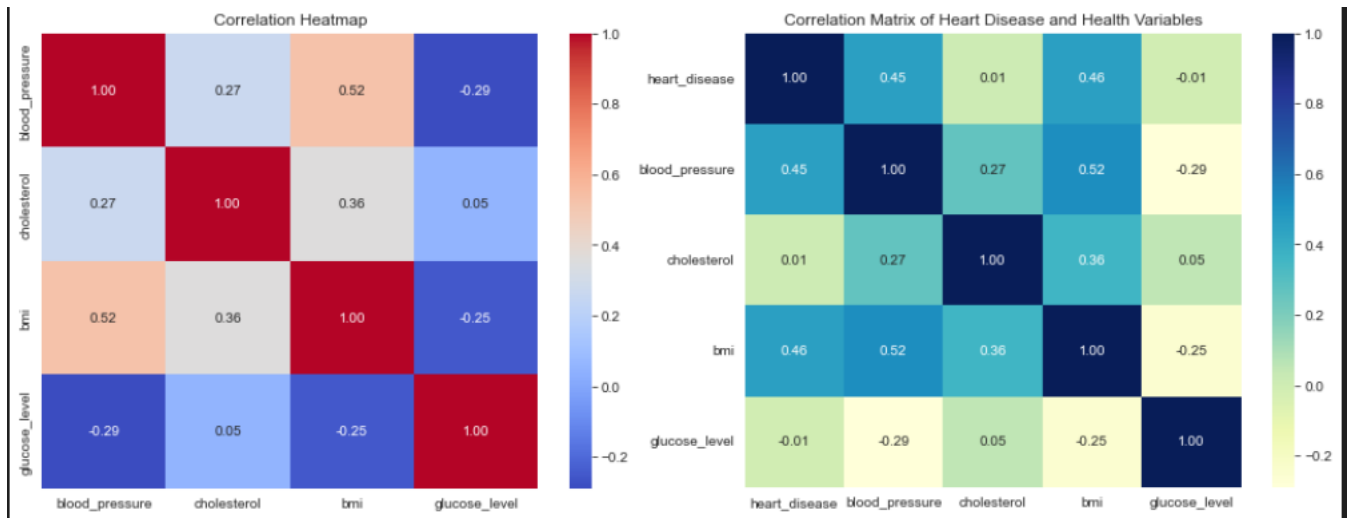


Fig1:- Correlation Heatmap 1

Fig2:- Correlation Heatmap 2

2. Plotted chart to know relationships between numerical variables

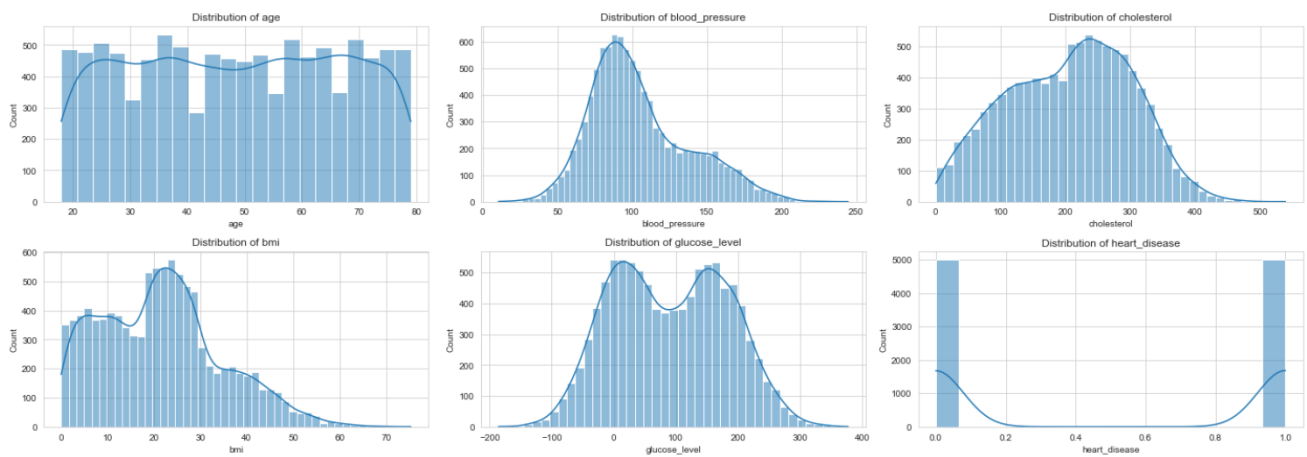


Fig 3:- Numerical Feature Distribution

3. Plotted Chart to know relationships between categorical variables

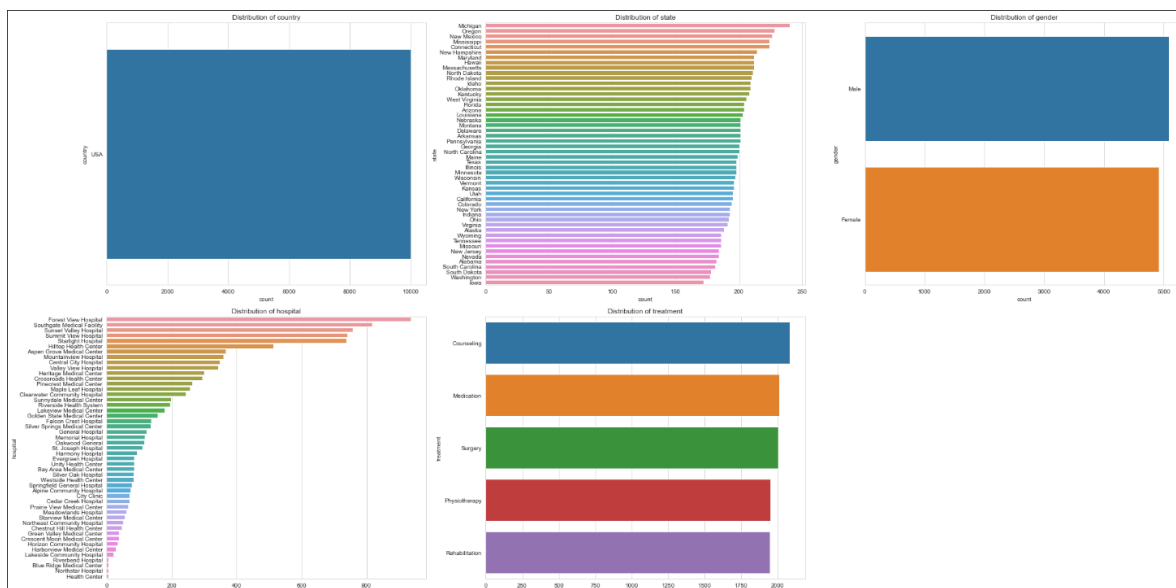


Fig 4:- Categorical Feature Distribution

- Plotted statewise heart disease cases by implementing groupby function on basis of states.

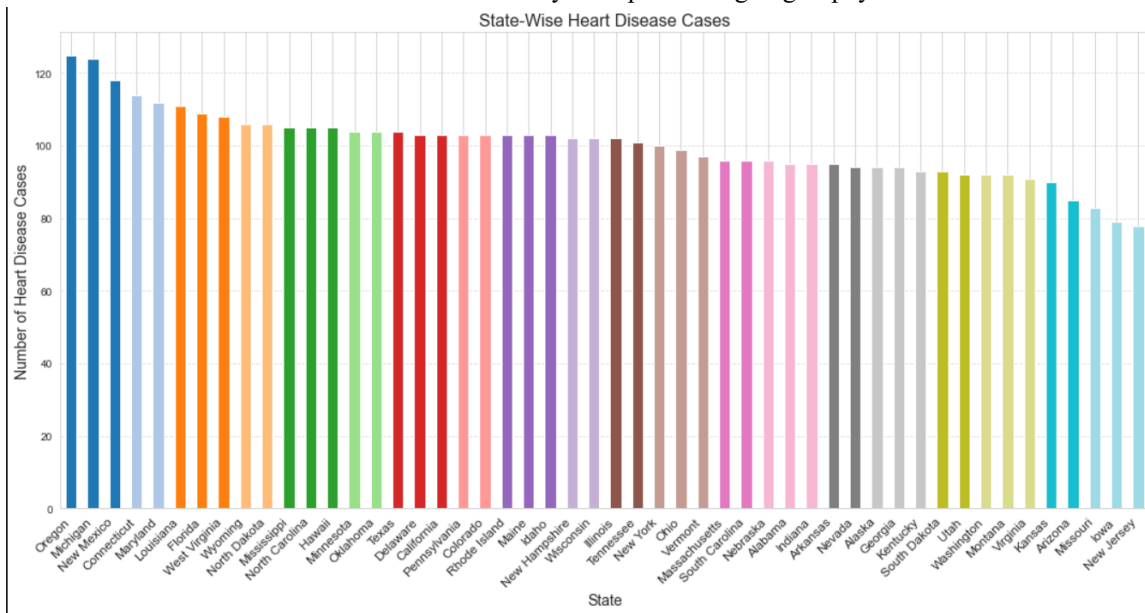


Fig5: Statewise Heart Disease Cases

- Plotted Statewise Average of Health Variables for Heart Disease Cases Heatmap. Where Dark Color Signifies that parameter average is high in respective state.

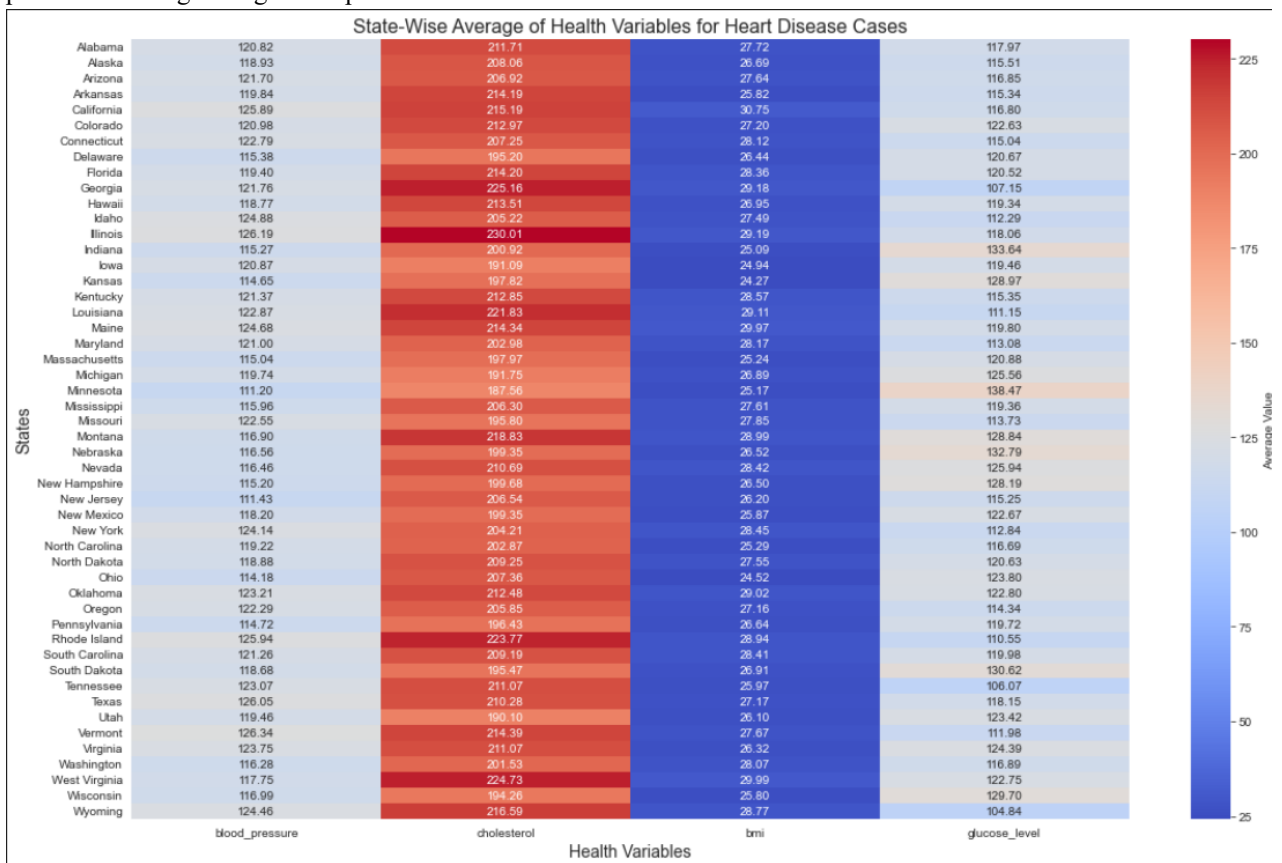


Fig6: Statewise average of health variables for heart disease cases.

6. Plotted patient segmentation graph distributed by blood pressure issue by age using Kmeans module of sklearn library.

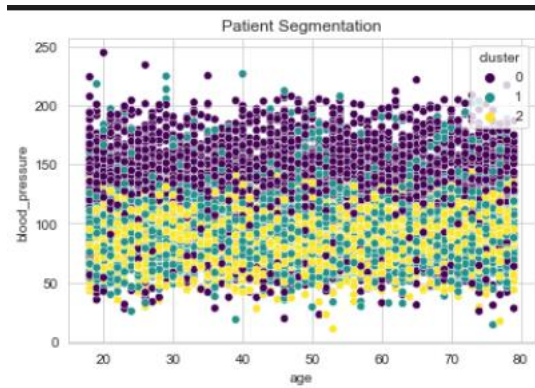


Fig 7:- Patient Segmentation

e) Documentation and Reporting

- I have followed the complete Data Processing Methodology in which I have Loaded and Fetched Data from MongoDB, Performed EDA(Exploratory Data Analysis), Performed All significant Data Cleaning, and transformation steps and In the end created an Interactive dashboard.
- Constructed Interactive dashboard by coming both three data sets 1. Emissions Dataset 2. Pollution Dataset 3. Healthcare Dataset. Visualized trends and analysis on the filter of State and sub-filter of year and fuel consumption to show from where that emissions are coming. I have generated a storyline from this data and interpreted it inside the dashboard. First I started by showing Emissions by type of fuel then showcased what are the sector from where these fuels are emitted. Then showed seasonal pollution trends and the average health scores of persons and treatments available in that state. All These statistical insights are working statewide. I have used “Dash” and “Plotly” to create this dashboard.

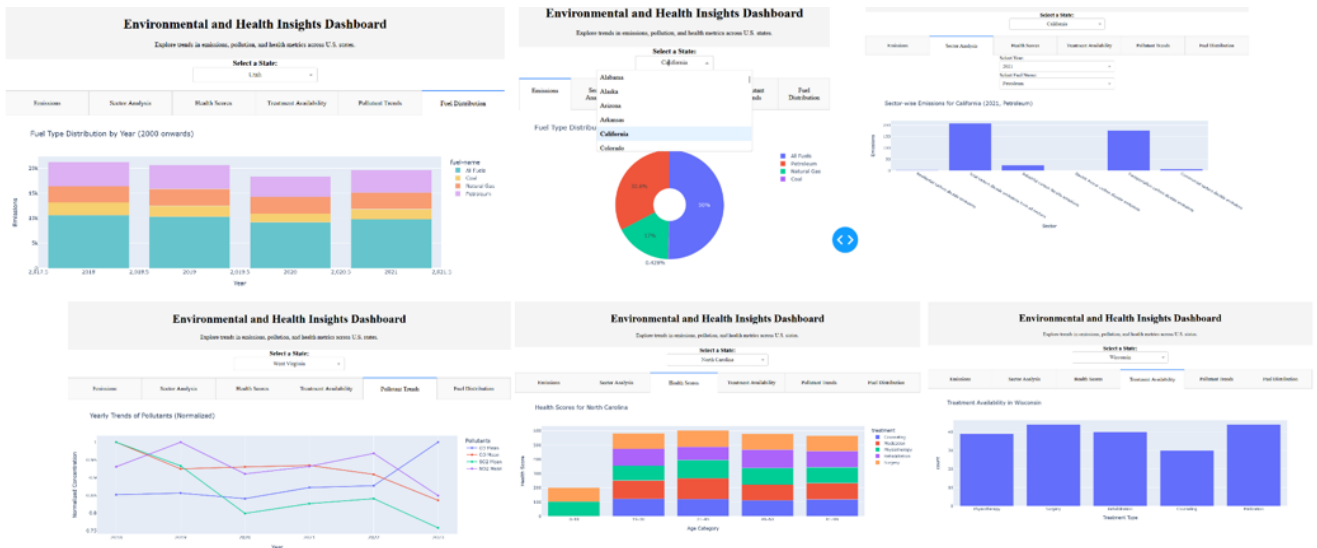


Fig8: Environmental and Health Insights Interactive Dashboard

- Contribution to Final Report:- I have taken care of all formatting of report in IEEE formats and read some reference papers on Interactive Dashboard working in Python and Healthcare in the USA. Also, I have written the Abstract, Introduction, Healthcare Data part, Dashboard part, and Conclusion part in Final Report on my own.

III. Key Challenges and Solutions

Challenges:- High Cardinality in categorical columns (State, hospital and treatment)

Solution:- Aggregated data feature on the basis of main variable and clustering should be implemented for segmentation.

IV. Reflection and Learnings

- Gained and Learned to Make connections between MongoDB->Python->SQL and also learned to generate and create data pipeline using dagster library in python.
- Enhanced Visualization skills and learned to create an Interactive Dashboard using Python only.
- Improved my working skills in a team by coordinating various steps with team members.
- The project allowed me to work on real-time healthcare data and merge insights with Pollution and Emission data, Which enhanced my ability to derive actionable insights from various data sets and also improved my visualization skills.