Project Report for Phase Three

Team 10

Vasana Srinivasan (20221010123)
Nishita Kannan (2022101112)
Navdha Bansal (2022101055)

What we planned to accomplish in the first phase:

Our project's goal was to use data visualization to analyze hospital data and enhance healthcare.

We aimed to answer the following questions through the visualization:

1. What are the relationships between diseases and specific demographic groups, such as age and gender?
2. How does the patient's medical condition determine their admission type and their duration of stay?
3. How do the duration of the stay, medication, and insurance affect the billing amount?
4. How do medical conditions vary by season?
5. What is the hierarchy of medical conditions and associated medications, and how can we assess the superiority of two different medications for a given medical condition based on the duration of stay?
6. What is the relationship between blood type and disease, if any?
7. What is the relationship between disease and billing amount?
8. What are the trends between medication usage over time and season?

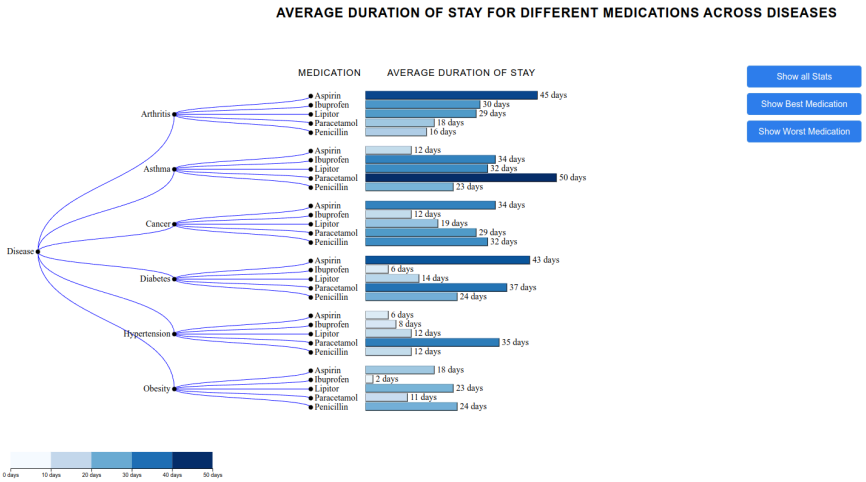
What was done in Phase 2

Since the visualizations 3 and 5 required a little more coding work than the others, we decided to implement them for phase 2.

Visualization 5:

The visualization for this question has been implemented through a *dendrogram* and *horizontal barplot*. There are 3 buttons along with the graph, using which the user can choose to:

- View all the statistics about the duration of stay
- View information about the best medication
- View information about the worst medication



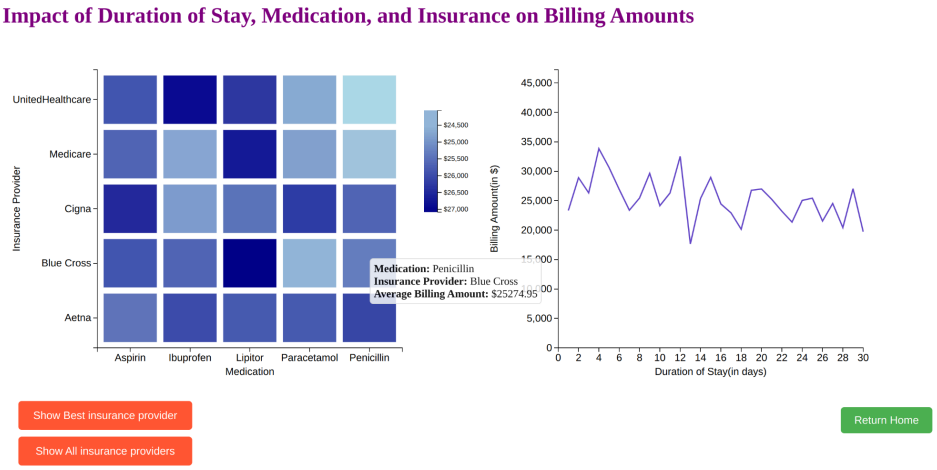
where the shortest length of stay is used to evaluate the best medication, and the longest stay is used to evaluate the worst medication.

Additionally, there is a tooltip that displays the details about the Medication, Average Duration of Stay, and Medical Condition when the user hovers over a bar. This redundancy has been implemented to make it very easy for the user to interpret the data in front of them.

Visualization 3:

The visualization for this question has been implemented through the use of a *heatmap* along with a *line graph*. There are 2 buttons along with the graph, using which the user can choose to:

- View all statistics about the insurance provider and medication
- Highlight the best insurance provider for each medical condition



The best insurance provider is determined based on lowest cost per day of stay.
Upon clicking on a rectangle on the heatmap a line graph is displayed, that shows the relationship between the billing amount and the duration of stay.
There is also a tooltip that is displayed while hovering over a rectangle on the heatmap that allows the user to view details about:

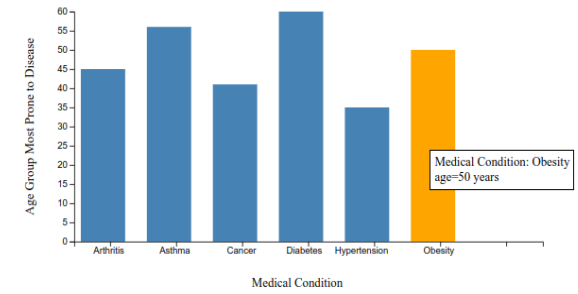
- Medication
- Insurance Provider
- Average Billing Amount per day of stay

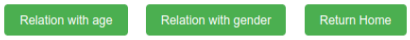
On hovering on the line graph, a second tooltip is displayed that allows the user to view details about:

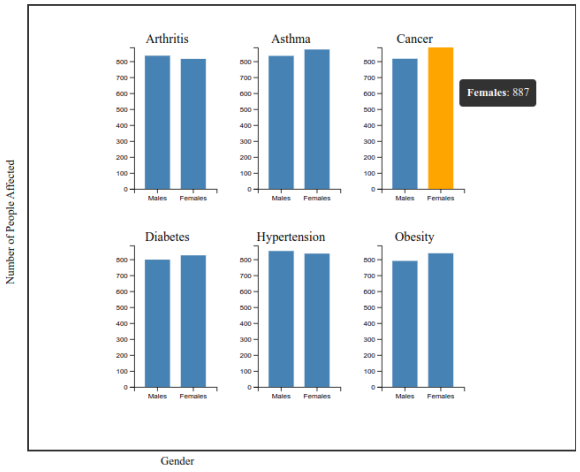
- Duration of Stay
- Billing Amount

Visualizations for Phase 3

Visualization 1:







Visual Encodings:

The aim of this visualization was to explore the relationship between medical conditions and the age groups and genders most susceptible to each condition.

To explore the relationship between medical conditions and age, a bar graph (hereby referred to as Graph 1) was used which offers a clear depiction of age-related trends in disease prevalence. Medical Conditions (categorical variable) were represented along the X-axis and age (quantitative variable) along the Y-axis.

The relationship between medical conditions and gender was visualized using a set of six small bar graphs neatly arranged within a box (hereby referred to as Graph 2). Each of these 6 graphs illustrates the distribution of a certain disease across gender categories (male and female), with the X-axis representing gender (categorical variable) and the Y-axis indicating the number of affected individuals (quantitative variable).

Navigation:

Three buttons have been provided to enable the user to navigate. Upon clicking the “Relation with Age” button, Graph 1 is displayed. Upon clicking the "Relation with Gender" button, Graph 2 is displayed. Upon completion of the analysis, users have the option to seamlessly return to the home page of HealthVue by clicking the "Return Home" button, facilitating ease of navigation and ensuring a user-friendly experience.

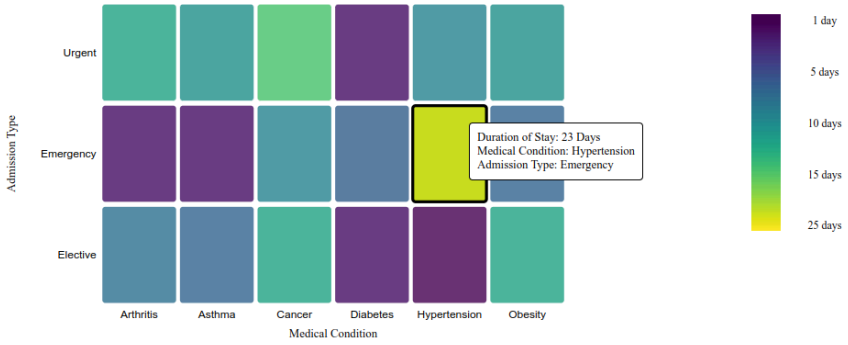
Animation and Colors:

Upon hovering over each bar on Graph 1, a color change highlights the selection, while tooltips provide additional context, revealing pertinent information such as the specific medical condition and the corresponding average age. Tooltips are also incorporated in Graph 2, displaying the exact count of male and female individuals affected by each disease. This multi-faceted approach enables users to discern nuanced patterns regarding gender-based disease susceptibility.

Visualization 2 :

Visual Encodings:

The aim of this visualization was to explore the relationship between Medical Conditions, Duration of Stay, and Admission type. A heatmap was chosen for this visualization since it effectively represents the relationship between two categorical variables (Medical Condition and Admission Type) and one quantitative variable(Duration of Stay). The heatmap shows the Medical Condition on the x-axis and the Admission Type on the y-axis.



Animation and Colors:

The heatmap uses colored rectangles to represent the Duration of Stay values, with the color intensity varying based on the magnitude of the value. A legend is displayed beside the graph to enhance the accessibility of the visualization for users with different visual abilities. Further, a tooltip is provided on hover, which shows the Medical Condition, Medication, and Duration of Stay of the square which allows the user to access more granular information without cluttering the main visualization. The Viridis color scheme was used because it is a perceptually uniform palette, making it easier to distinguish different values across the color range.

Navigation:

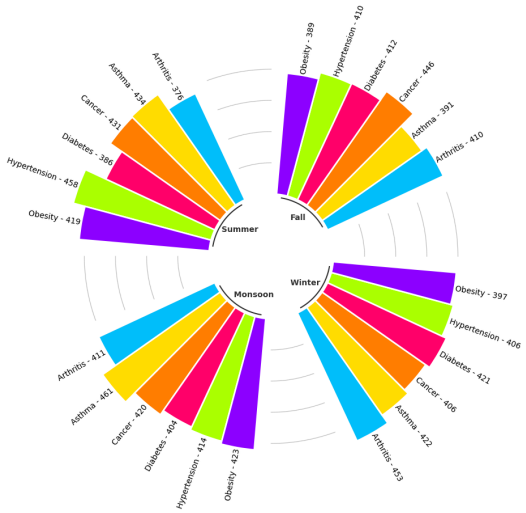
Upon completion of the analysis, users have the option to seamlessly return to the home page of HealthVue by clicking the "Return Home" button, facilitating ease of navigation and ensuring a user-friendly experience.

(Visualization 3 explained in the previous section.)

Visualization 4:

Visual Encodings:

The aim of this visualization was to explore the seasonal variation of Medical Conditions. A radial bar graph is capable of showing cyclical trends, which is why this graph was chosen - seasons occur in a cyclical pattern. The length of each bar in the graph is proportional to the frequency of the occurrence of the disease. There are labels that have been added to each bar in the graph that display the corresponding disease (the Categorical variable) and the frequency of occurrence of the disease in that particular season (the Quantitative variable). The bars have also been grouped according to the four seasons, and each group has also been labeled to allow for ease of interpretation.



Animation and Colors:

Each disease is color-coded - this redundancy was introduced for faster understanding. There is no user interaction/animation in the graph - it is static.

Navigation:

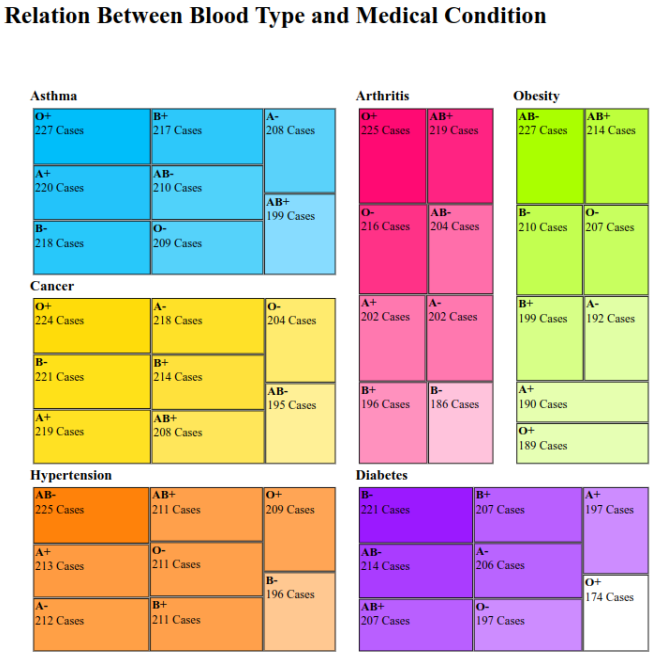
Upon completion of the analysis, users have the option to seamlessly return to the home page of HealthVue by clicking the "Return Home" button, facilitating ease of navigation and ensuring a user-friendly experience.

(Visualization 5 explained in the previous section)

Visualization 6 :

Visual Encodings:

The aim of this visualization was to explore the relationship between blood type (Categorical variable) and medical condition (Categorical variable). A treemap was utilized for this purpose which consists of rectangles for each medical Condition which are further divided into smaller rectangles representing the distribution of Blood Types. Each rectangle's area corresponds to the frequency of the blood type for the specified medical condition. A treemap was chosen since it is well-suited for representing hierarchical data, where there are parent-child relationships or nested categories. In this case, the data has a hierarchical structure with Medical Conditions as the top-level categories and Blood Types as subcategories or child nodes.



Animation and Colors:

Each Medical Condition rectangle is assigned a color. Within the rectangle, the same color is interpolated for the boxes, where darker shades represent higher values for the frequency of the blood type and lighter shades represent lower values. As each rectangle's area corresponds to the frequency of the blood type, as mentioned before, the channels of color and size work together to add redundancy, making the graph easier to interpret for the user. Moreover, upon hovering, the size of the rectangle expands, creating a pop-out effect to enrich the user's interaction.

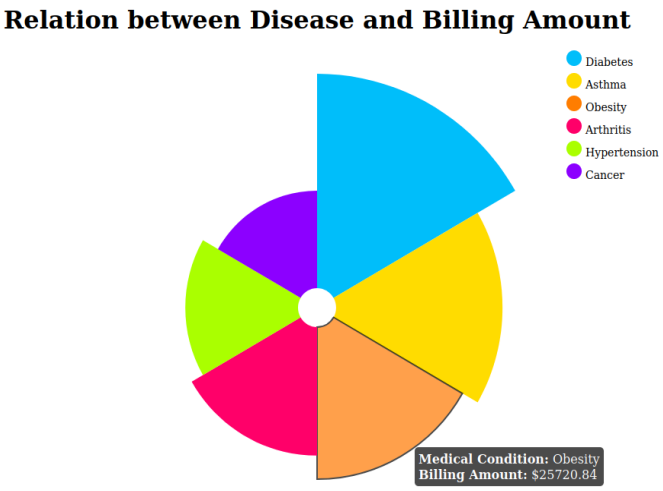
Navigation:

Upon completion of the analysis, users have the option to seamlessly return to the home page of HealthVue by clicking the "Return Home" button, facilitating ease of navigation and ensuring a user-friendly experience.

Visualization 7:

Visual Encodings:

The aim of this visualization was to explore the relationship between Medical Condition (the Categorical variable) and Billing Amount (the Quantitative variable). A polar area chart was used for this purpose, as it is effective for representing a parts-to-whole relationship. The chart is divided into segments, with each segment representing a different disease. Each segment's area represents its share of the total billing amount. The graph allows viewers to grasp the distribution of total healthcare spending across various medical conditions at a glance, offering insight into which medical conditions consume a larger portion of healthcare resources



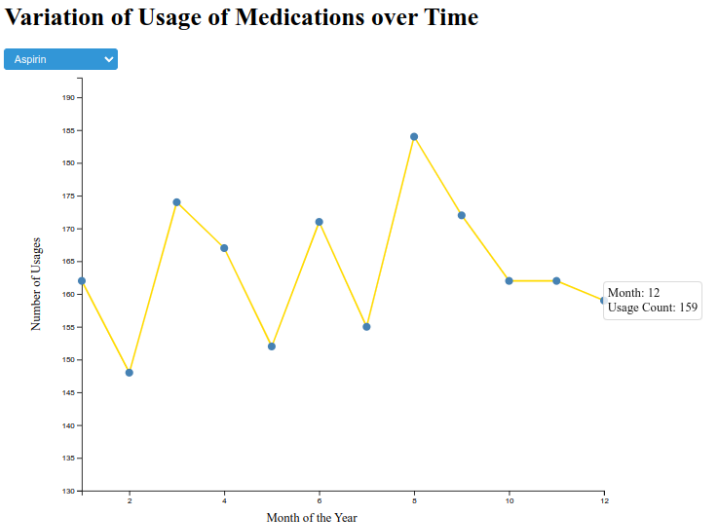
Animation and Colors:

The coloring of each segment is distinct, enabling users to identify and differentiate between the various disease categories by referring to the accompanying legend. Additionally, tooltips provide detailed information upon hovering over each segment, offering insights into specific disease categories and corresponding healthcare expenditures.

Navigation:

Upon completion of the analysis, users have the option to seamlessly return to the home page of HealthVue by clicking the "Return Home" button, facilitating ease of navigation and ensuring a user-friendly experience.

Visualization 8:



Visual Encodings:

The aim of this visualization was to explore the variation of Medical Conditions over seasons. A line chart was utilized for this purpose as it is an effective tool to communicate to a user about trends and patterns over time**.** The data is ideal for visualization in a line graph, where the X-axis represents the Month of the Year (Ordinal variable), and the Y-axis represents the Number of Usages (Quantitative variable). There are 5 line charts that each indicate the trend for a particular Medication.

Animation and Colors:

A drop-down menu is provided, using which users can select the Medication whose data they wish to see, which allows the user to navigate between the line charts that show the trends. Each medication is assigned a distinct color so that when switching the displayed graph, there's a noticeable color variation to highlight any changes or transitions. Furthermore, tooltips have been integrated at every data point, allowing users to precisely discern the frequency of usage for each medicine during a specific month.

Navigation:

Upon completion of the analysis, users have the option to seamlessly return to the home page of HealthVue by clicking the "Return Home" button, facilitating ease of navigation and ensuring a user-friendly experience.

Specific Tasks that can be done using our Visualization

1. Identify groups that may be at a higher risk of developing certain medical conditions.
2. Analyze how a patient's medical condition determines their admission type and their duration of stay and find patterns in ailments.
3. Analyze how the duration of stay, medication, and insurance affect the billing amount.
4. Analyze how medical conditions vary by season.
5. View the hierarchy of medical conditions and associated medications, and how can we assess the superiority of different medications for a given medical condition based on the duration of stay.
6. Explore the relationship between blood type and disease.
7. Analyze the relationship between disease and billing amount.
8. Analyze the trends in medication usage over time and season.

Who are the users, and why do they need to perform these tasks?

1. Healthcare providers (doctors, hospitals, clinics): They need to identify groups at higher risk for certain medical conditions (Task 1) and analyze how medical conditions determine admission type and duration of stay (Task 2) to better allocate resources, improve patient care pathways, and anticipate seasonal outbreaks (Task 4).

2. Healthcare administrators and policymakers: They need to analyze the relationship between duration of stay, medication, insurance, and billing amount (Tasks 3 and 7) to optimize healthcare costs, understand factors contributing to higher costs, and make informed decisions regarding insurance policies and medication selection.

3. Pharmaceutical companies and researchers: They need to assess the superiority of different medications based on duration of stay (Task 5) and explore the relationship between blood type and disease (Task 6) to inform drug development, personalized medicine, and further research into genetic predispositions.

4. Insurance providers: They need to analyze the relationship between duration of stay, medication, and billing amount (Task 3) to determine the most cost-effective insurance policies and negotiate better rates with healthcare providers.

5. Public health organizations and researchers: They need to analyze trends in medication usage over time and season (Task 8) to investigate potential causes, such as changes in pricing, manufacturing, or regulatory policies, and to plan for future healthcare needs.

6. Patients : Those with rare blood types or conditions with potential genetic links could explore the relationship between blood type and disease (Task 6) to better understand their risk factors and seek appropriate preventive care or genetic counseling. Patients

may also need to analyze the relationship between their own medical conditions and billing amounts along with the Insurance provider (Task 3 and 7) to understand their healthcare expenses, identify potential discrepancies, and ensure fair billing practices

Data sets on which the visualization was tested

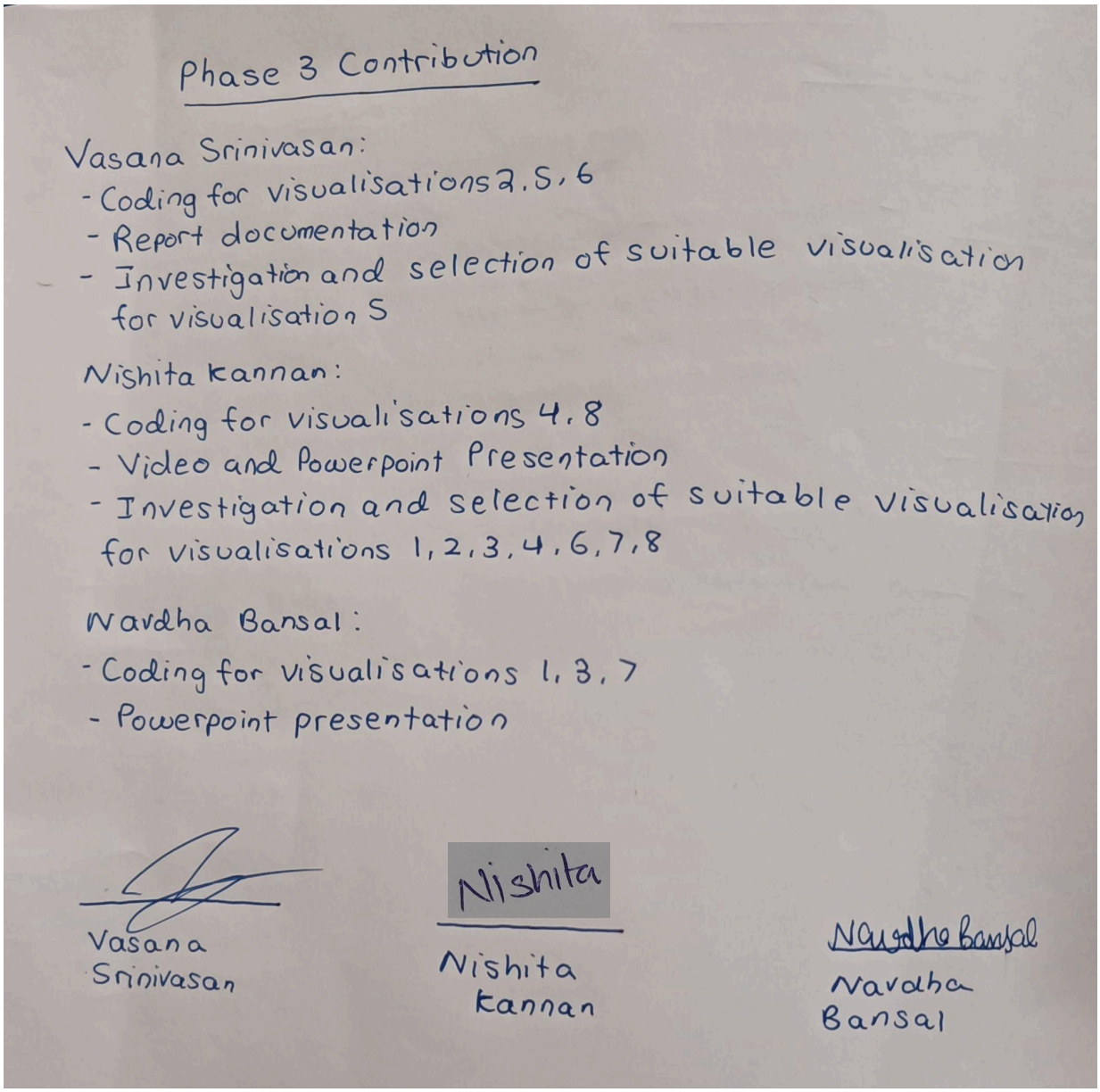
The dataset utilized in this analysis was sourced from Kaggle.
https://www.kaggle.com/datasets/prasad22/healthcare-dataset

The data extracted from the CSV file sourced from the website underwent modifications through the application of Python scripts to generate JSON or new CSV files containing only the data required for the relevant visualization. The submission includes the original Dataset CSV, Python Scripts, the JSON/New CSV files produced from the Python Scripts, and the HTML codes for the visualizations done in D3.js.

Limitations of our Visualization

1. Geographic limitations: The dataset is purely synthetic. Healthcare practices, regulations, and costs can vary significantly across different regions or countries, which could limit the generalizability of the insights derived from the visualizations.
2. Static data: As healthcare data is constantly evolving, the visualizations may become outdated over time, potentially limiting their relevance and applicability for decision-making.
3. Inaccessible for color blind users: The color schemes may not be accessible or interpretable for individuals with color blindness or color vision deficiencies. Although information has been provided in the form of text as well, color-blind users will not benefit from the additional impact that the color scales provide.

Contributions



Phase 3 Contribution

Vasana Srinivasan:
- Coding for visualisations 2, 5, 6
- Report documentation
- Investigation and selection of suitable visualisation for visualisation 5

Nishita Kannan:
- Coding for visualisations 4, 8
- Video and Powerpoint Presentation
- Investigation and selection of suitable visualisation for visualisations 1, 2, 3, 4, 6, 7, 8

Navdha Bansal:
- Coding for visualisations 1, 3, 7
- Powerpoint presentation

Vasana Srinivasan

Nishita

Nishita Kannan

Navdha Bansal

Navdha Bansal