

# **DATA VISUALISATION**

**Project Title:-** Scope and salaries of Various IT roles during(from) 2020 to 2022

**Author Name:-** Tirumalesh Nagothi

## Table of contents

Section	Page No.
1. Introduction	3
2. Design	4
3. Implementation	8
4. User guide	13
5. Conclusion	16
6. Bibliography	17
7. Appendix	18

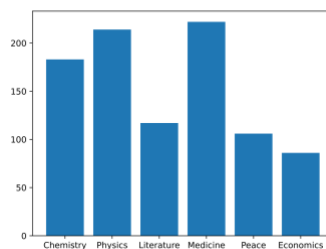
## Introduction

This project is made to keep the public in mind who wants to know how covid 19 affected the trend of IT roles. This narrates the findings on the salaries for various IT roles, mostly related to the data science field during the three covid stages in the years 2020, 2021 and 2022. I wanted to tell the audience about the average salary paid to the employee, which is based on job role, location of the company(country name), and year. A map is visualised with the geo-location point of the country, which describes the country name, average covid 19 cases in that year and average salary. In addition to this map, there is a bar chart plotted which visualises the average wage of different job roles paid in a particular country. Both the map and bar chart is entirely interactive, and the project is implemented using shiny.

## Design

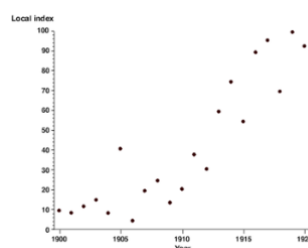
- The bar chart for tabular data or data frame; referring to making sense of tabular data.
- Use of spatial data in the leaflet map and the pop-ups; referring to the marker concept in leaflet with R(making sense of spatial data)
- Visual elements include location pointer and the pop-ups in the leaflet map; referring to the visual attention, and working memory.

Firstly, in the Ideas sheet(figure 5.1 in appendix). Here I've taken the eight different maps. The first four maps, such as



**Figure 1.1(bar-chart)**

**bar-chart**(this is simple and easy to understand the overview of data),



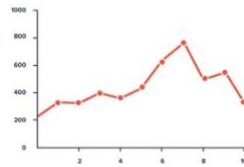
**Figure 1.2(scatter plot)**

**scatter-plot**(makes the audience to clearly focus on a particular point and bit consisted to understand),



**Figure 1.3(pie-chart)**

**pie-chart**(this is perfect to visualize but looks more consisted to visualize more number of unique labels in the data) and



**Figure 1.4(line chart)**

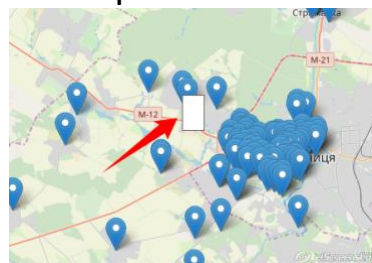
line chart (this is great for the timeline visualisation).

Coming to the second cluster of visualisation, i.e. maps such as



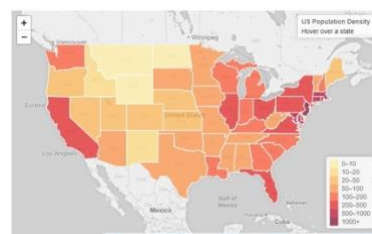
**Figure 2.1(Network map)**

**network map**(to visualize the point outs to the various jobs in a list),



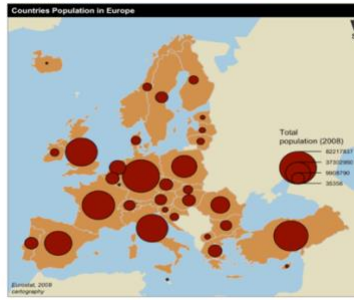
**Figure 2.2(Marker map)**

**marker map**(to mark the geo points of the countries using appropriate markers),



**Figure 2.3(color map)**

**color map**(Choropleth map to visualize the covid 19 cases using range of single color),



**Figure 2.4(Bubble map chart)**

**bubble map chart**(to visualise the covid 19 cases with a circle using the radius).

In the second sheet, the first model is designed(figure 5.2 in appendix). They choose the country names and the year using the select input tool to plot the salary based on job role using a bar chart and pie chart. Also, a plain marker map is used to point out the country's location, representing the average salary in that country. But, here, a single and straightforward chart should be used to describe the salary data.

In the third sheet, the second model is designed(figure 5.3 in appendix). Choosing the country names uses the select input tool, and the year is selected using radio buttons. And the data is visualised to plot the salaries using a scatter plot and line charts for covid 19 trend. A colour map represents the covid 19 cases using the geometry values for a country. When the marker on a country is clicked on, a tooltip describing the country name, average salary and the average covid19 cases is displayed.

In the fourth sheet, the third model is designed(figure 5.4 in appendix). Using a line chart, choose the country names using the select input tool to visualise the covid 19 trend. A year is selected using the radio button. To plot the map, a button(Go) should be clicked. When a country is clicked, a tooltip displays the overview of jobs, salaries and the remote ratio.

In the fifth sheet, the final model is designed (figure 5.5 in appendix). There will be two tabs named 'Main' and 'Explore' where 'Main' will visualise the colour map with markers(\$), and the 'Explore' tab will visualise the bar chart representing the salaries of different job roles. There are radio buttons at the beginning of the main tab to select the year and a unique input tool below the map.

## Implementation

This project is implemented using R programming and shiny. There are five different libraries used to run the application. There is a need of internet to plot the map and to download the packages if they are not existed in the user's local machine. When the code is executed, the program will check whether the required libraries exist or not. If not existed, then the libraries will be downloaded.

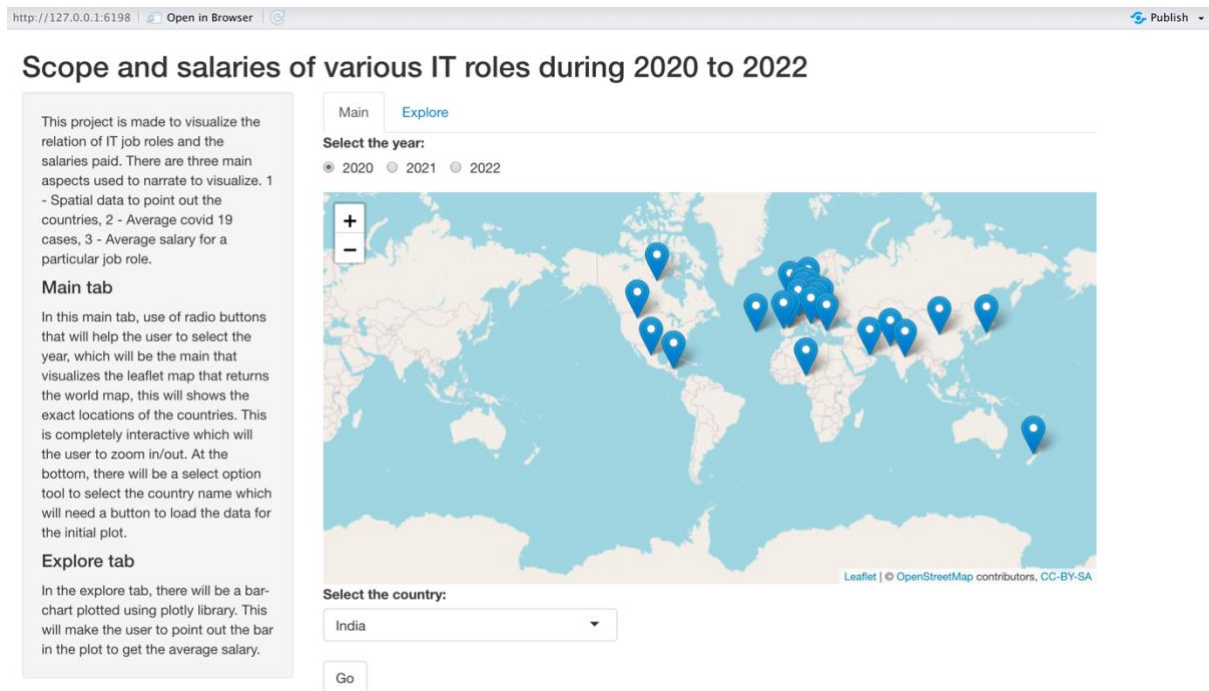
Libraries used:-

- Shiny(This library is used to provide an interface and make the multiple data sources meet at one place to visualise and make it interactive).
- dplyr(This library filters and manipulates the data based on our choices, such as including specific columns and rows to answer a question).
- Leaflet(This library is used to visualise the maps and to make them interactive for the users to handle).
- Plotly(This library is used to plot the interactive charts).
- Tidyverse(This library is used to manipulate the content inside the dataframe)

Data sources used:-

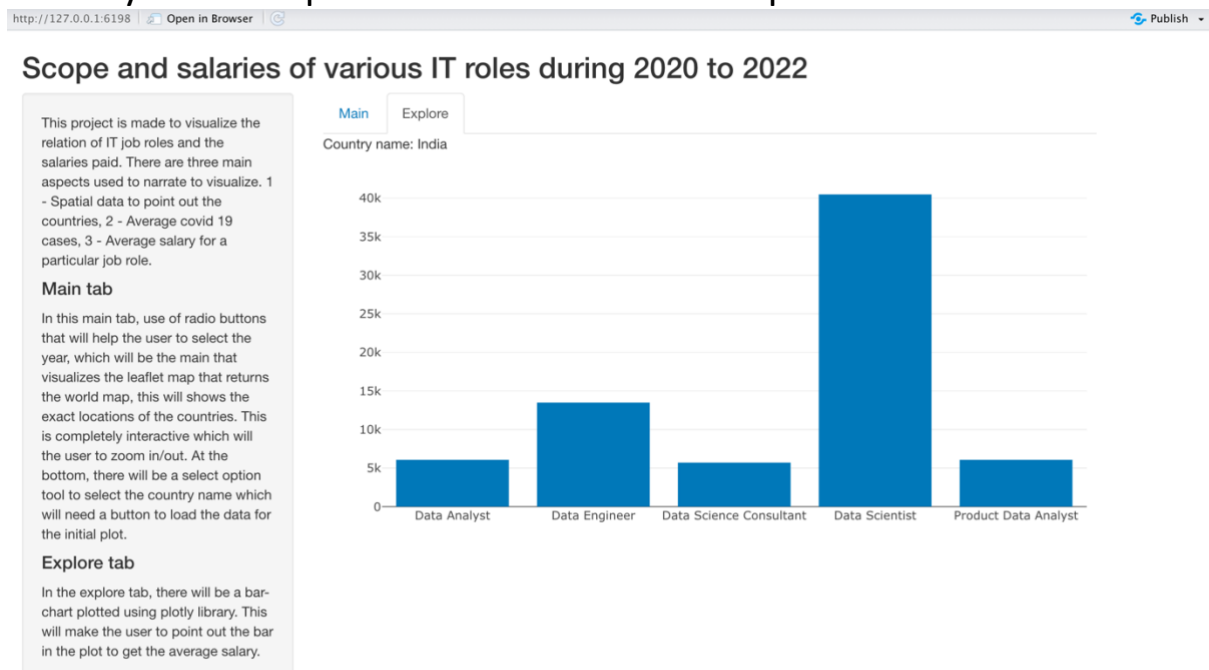
- Ds\_salaries.csv:- This data is used to get the salary for each role and the company location.
- Salary\_data\_each\_year.csv (Filtered and derived from ds\_salaries.csv and covid 19 datasets):- This data contains the average salary and average covid cases each year from 2020 to 2022.
- World-administrative-boundaries.csv:- This dataset is used to get the geo point locations to place the markers in the leaflet map.





**Figure 3.1(Main tab)**

Here in the main tab, we can see the map with the markers of the countries and the tools to select the year for plotting the map and country name to plot the bar chart in the explore tab.



**Figure 3.2(Explore tab)**

Here in the explore tab, we can see a country with different job roles bar chart.

The standard type of features in the salary\_data\_each\_year.csv and world-administrative.csv is the alpha two country code with different column names.

The three columns extracted from the world-administrative.csv data include geo point location, alpha two country code and the English name of the country code. The alpha two country codes in both datasets are matched and merged.

The data is filtered into three multiple data based on the years and where the average salary is 0.

The years are selected using the radio buttons options, which will plot the map with a tooltip that displays the country name, average covid19 cases and the average salary. Bottom to the map, there will be an appropriate select input option showing the countries with the average salary in that particular year. There will be a button for the initial visualisation, allowing the explore tab to display the bar chart using **plotly** that shows the average salary for the job roles available in that country.

For each year, the multiple filtered data mentioned above will be used to plot the map data. The markers will be added to the leaflet map with the latitude and longitude from the filtered data. The select options are updated based on the year selected with the radio buttons. The plotly map is rendered by getting the country code from the merged data using the chosen English name in the particular option tool. Then, the year data was obtained by matching the year value selected in the radio button tool and the country code extracted. Finally, getting the modified data will get the average salary of each job role and store it in a new column. Then there will be a bar chart

plotted where the job roles will be at the x-axis and the average salaries are given at the y-axis.

All the interactive happened with the `interactive()` function with the if condition.

There is a single change between my final design model and the implemented model, i.e. not using the colour map to visualise the covid19 data. Because I couldn't find the geometry data with the country, if such data exists, I would map the country name with the English name available in the merged data and apply the colour map in the leaflet. No scatter points are available at the top of the bar chart, which I didn't feel would be helpful.

In my project, there is more need for the integrity of multiple sources, which will help wrangle, clean and merge the data for filtering based on the problem to solve. To visualise the findings, data manipulation takes place. Each kind of plot is based on the different tool's output. Salary salaries based on the covid 19 cases show how the covid19 impacted IT jobs. Multiple tabs provide a clear picture of how each plot can be plotted separately to understand.

#### References:

- [1] For handling the button, 23.10.2022. URL: <https://shiny.rstudio.com/reference/shiny/latest/observeEvent.html>
- [2] Displaying the radio button in a single row, 23.10.2022. URL: <https://stackoverflow.com/questions/31974588/displaying-radio-button-in-elements-in-shiny-in-a-horizontal-order-instead-of-de>
- [3] Grouping the data using dplyr library, 24.10.2022. URL: <https://www.geeksforgeeks.org/group-by-function-in-r-using-dplyr/>
- [4] Updating the select input options, 25.10.2022. URL: <https://shiny.rstudio.com/reference/shiny/1.2.0/updateSelectInput.html>
- [5] Linking the plotly charts at the server side, 26.10.2022. URL: <https://plotly-r.com/linking-views-with-shiny.html>

[6] Working with the geo points in the leaflet as markers, 26.10.2022.  
URL: <https://leafletjs.com/reference.html#marker>

[7] Dataset for geospatial data, 26.10.2022. URL:  
<https://public.opendatasoft.com/explore/dataset/world-administrative-boundaries/export/>

## User guide

The user is directly prompted to the application with the initial load of the main tab. Here the main-tab consists of a radio button tool, a world map and the select option tool.

- To start with, the user needs to select the year to get the countries available in that year with the salaries.

### Select the year:

☒ 2020   ☐ 2021   ☐ 2022

**Figure 4.1**

The initial year will be 2020(Figure 4.1). There will be changes in countries for each year, i.e., not all the countries are available with the data.



**Figure 4.2**

- When the user clicks on the markers available in each country, a tooltip pop-up will give information about the country name, average covid19 cases and the average salary in USD\$. Refer to figure 4.2.
- The user can zoom in/out to see the countries available.

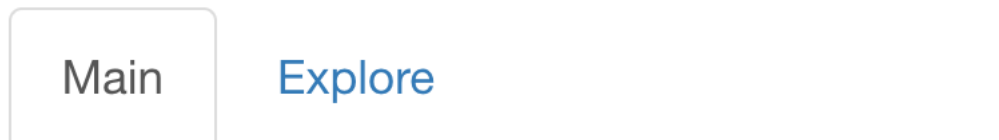
## Select the country:

India ▼

Go

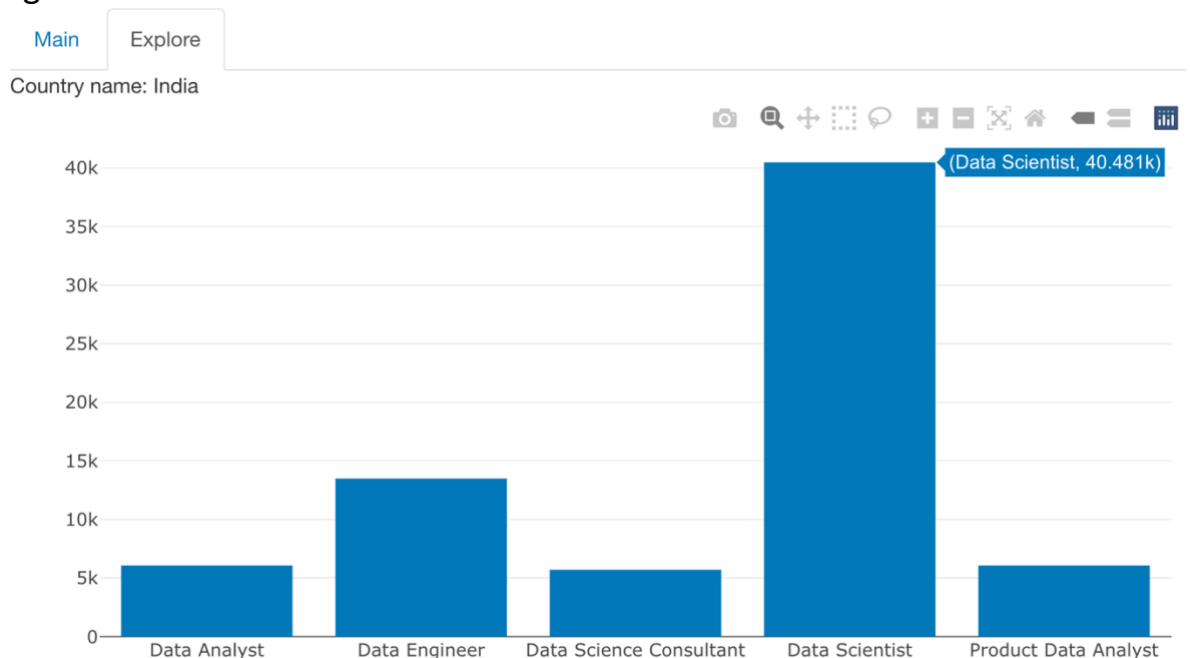
**Figure 4.3**

- Bottom of the map, there is an option to select the country names, where some country names will be in the initial place based on the year. Refer to figure 4.3
- For the initial load of the bar chart, the user needs to click on the Go button. Refer to figure 4.3



**Figure 4.4**

- Then, the user needs to switch to the Explore tab(refer to figure 4.4), where a bar chart will be plotted. Here, the user can interact with the graph by moving and pointing out the bar of a particular job role. Refer to figure 4.5.



**Figure 4.5**

- If the user wants to select another country, then the user needs to choose the country in the main tab select option tool, and there is no need to

click on the Go button. After selecting the country, the plot will be visualised directly in the explore tab, and the user can interact with it.

- There is also an overview of the tasks provided in this project given in the sidebar panel in the application. Refer to figure 4.6.

This project is made to visualize the relation of IT job roles and the salaries paid. There are three main aspects used to narrate to visualize. 1 - Spatial data to point out the countries, 2 - Average covid 19 cases, 3 - Average salary for a particular job role.

### **Main tab**

In this main tab, use of radio buttons that will help the user to select the year, which will be the main that visualizes the leaflet map that returns the world map, this will shows the exact locations of the countries. This is completely interactive which will the user to zoom in/out. At the bottom, there will be a select option tool to select the country name which will need a button to load the data for the initial plot.

### **Explore tab**

In the explore tab, there will be a bar-chart plotted using plotly library. This will make the user to point out the bar in the plot to get the average salary.

**Figure 4.6**

## Conclusion

To conclude the findings, not all countries have the same number of job roles available. So, for some pair choices, only the job role will be functional in that country. That implies the impact of IT or the availability of IT jobs in that country. For example, UAE has only one job role in a particular year, and India has multiple job roles. This indicates that the IT market in India is more massive than in the UAE. Both the covid19 and salary are derived as the average because the average calculation looks better for understanding. I've learnt how to implement visualisation using frameworks and integration of various libraries to make a single application—making the visualisations interactive and user-friendly. Also, on how to make simple and easy-to-understand plots. I also got the idea of using more from the five design sheets.



## Bibliography

[1] Javier Canales, Luna (2022), *Article shows 'Data science salary expectations in 2022*, 18.05.2022. URL: <https://www.datacamp.com/blog/data-science-salaries> [Re-Accessed on: 21.10.2022].

# Appendix

## Five Design Sheets

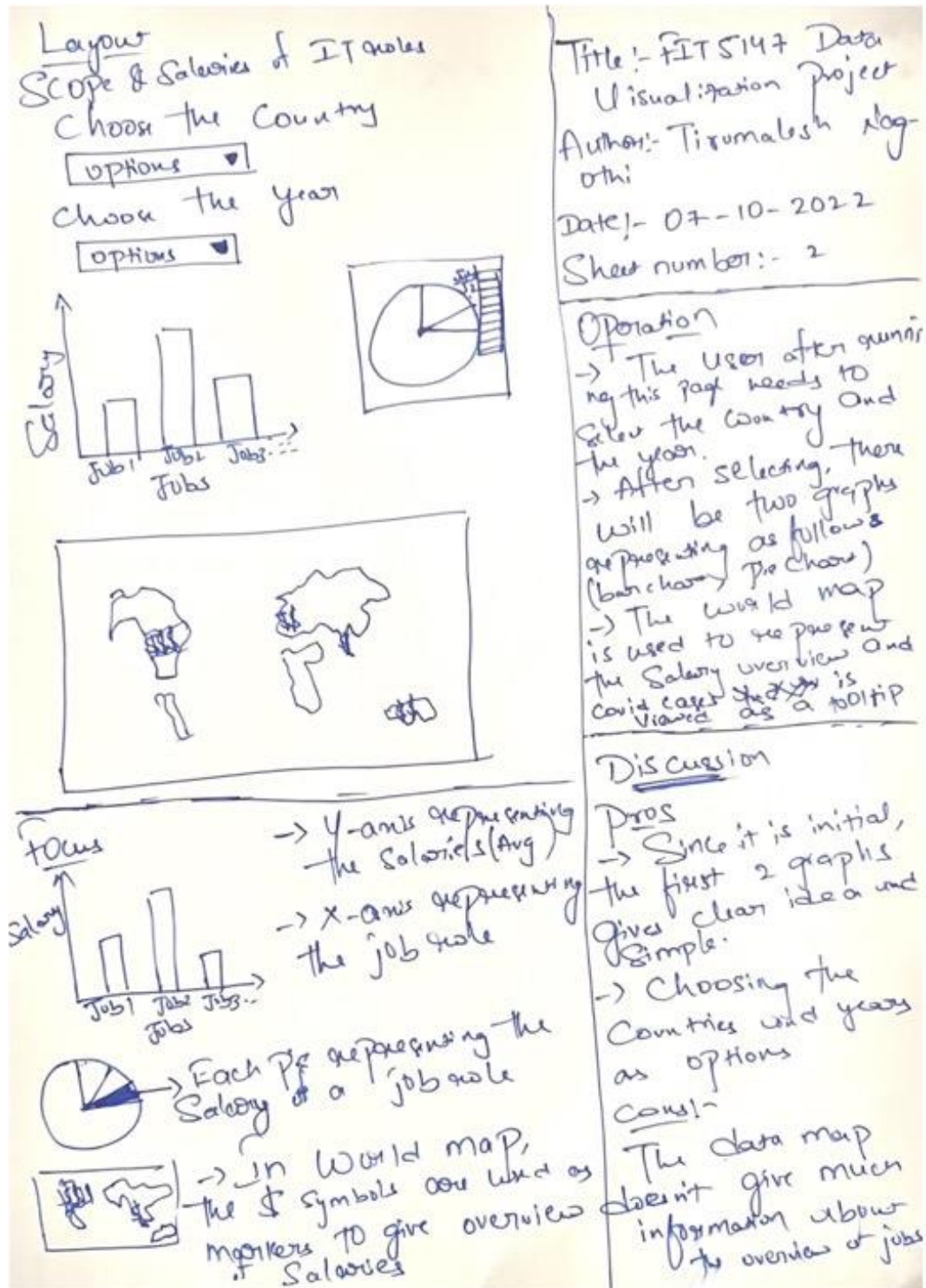


Figure 5.1(Ideas sheet)

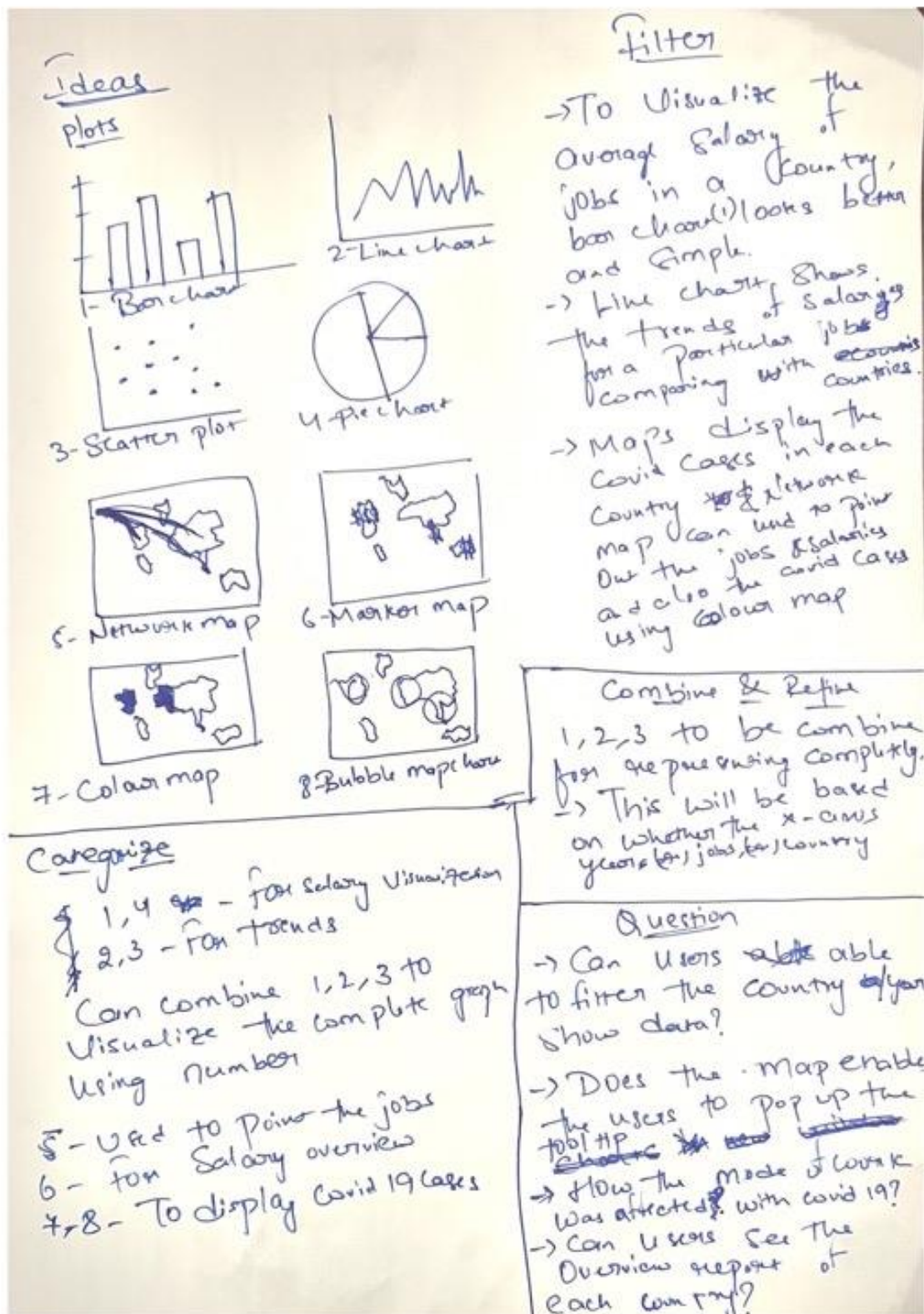


Figure 5.2(Design 1)



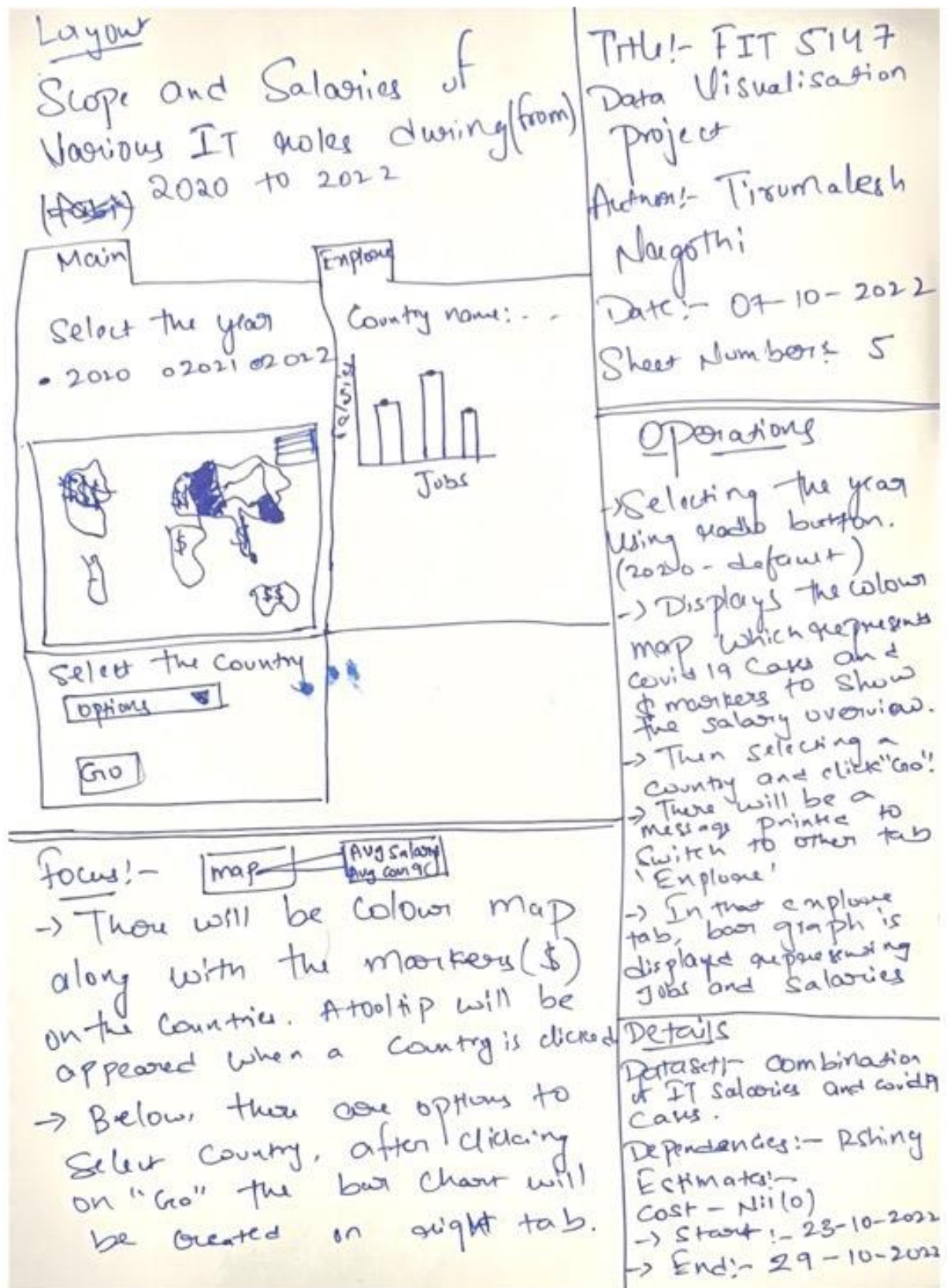


Figure 5.3(Design 2)

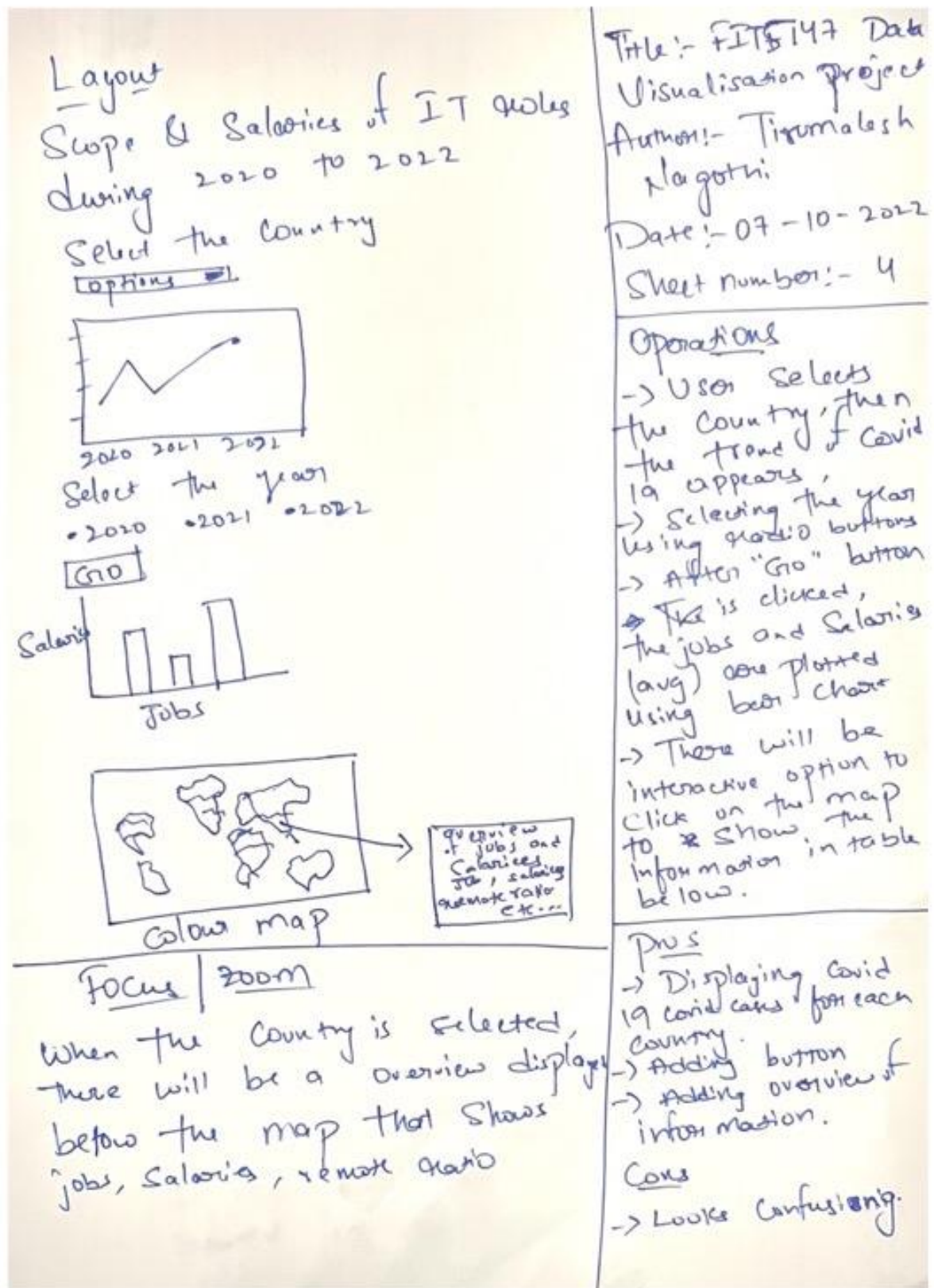


Figure 5.4(Design 3)



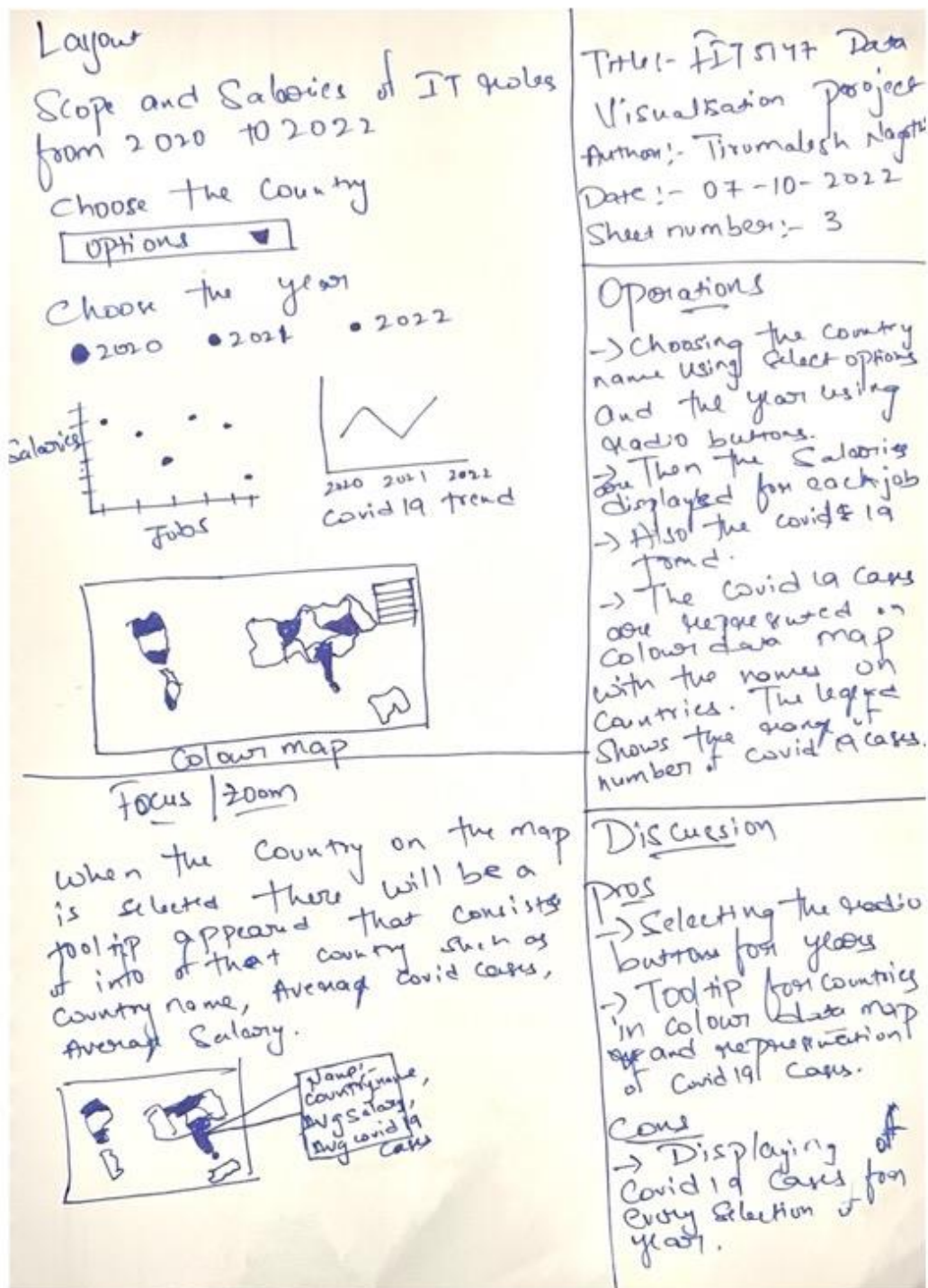


Figure 5.5(Final design)