A black and green rectangle

Description automatically generated

A black and green rectangle

Description automatically generated

|  |  |
| --- | --- |
| **Team name:** | Viz Visionaries |
| **Student name & ID:** | 1. Luwi Kah Wen 103828807 2. So Wai Ting 3. Liew E-Ling |
| **Tutorial day and time:** |  |
| **Year and Semester:** | July 2023, Semester 7 |
| **Word count:** |  |

<https://cos30045-group8.netlify.app/>

**Project Process Book**

MIGRATION OF INDIA DATA VISUALISATION

**Title Page**

Includes:

• descriptive title (e.g., ‘Data Visualisation Project’ is not acceptable)

• link to Mercury hosted website (must be on title page)

• team name and student names and IDs

• tutorial day and time

• year and semester

• word count

**Table of Contents**

[**1.** **Introduction** 4](#_Toc149517240)

[**1.1** **Background and Motivation** 4](#_Toc149517241)

[**1.2** **Visualisation Purpose** 4](#_Toc149517242)

[**1.3** **Project Schedule** 4](#_Toc149517243)

[**2.** **Data** 4](#_Toc149517244)

[**2.1** **Data Source** 4](#_Toc149517245)

[**2.2** **Data Processing** 5](#_Toc149517246)

[**3.** **Requirements** 6](#_Toc149517247)

[**3.1** **Must-Have Features** 6](#_Toc149517248)

[**3.2** **Optional Features** 6](#_Toc149517249)

[**4.** **Visualisation Design** 6](#_Toc149517250)

[**5.** **Validation [optional - Bonus Points]** 11](#_Toc149517251)

[**6.** **Conclusion** 11](#_Toc149517252)

[**7.** **References** 12](#_Toc149517253)

# **Introduction**

## **Background and Motivation**

Who will use, or be interested in, this visualisation (i.e., users)? What kind of tasks will they want to do? Why is it important?

The intended audience for our six data visualisations encompasses a diverse group, including individuals, policymakers, researchers, and organizations with a keen interest in comprehending the intricacies and ramifications of migration in India. This user base holds various roles and exhibits different interests, ranging from government officials and policymakers, researchers and analysts, and organisations to the general public.

For government officials and policymakers, our choropleth map and heatmap data visualisations, which delve into the unemployment rates of 2023 in different states of India and the wages disparities between India and other countries, hold immense significance. These visualisations are crucial for making well-informed decisions and addressing socio-economic challenges. They provide a valuable tool for formulating and evaluating policies related to immigration and economic development.

Similarly, researchers and analysts, tasked with studying migration trends and their impact on socio-economic factors, find our sunburst chart data visualisation particularly useful. This visualisation, which represents the population of India each year, contributes to academic research, informs studies on population dynamics, and supports social analysis.

Furthermore, organizations such as Non-Governmental Organizations (NGOs) can harness the insights from our cluster-stacked bar chart, which focuses on the relationship between diseases and migration. This data visualisation helps organizations understand migration patterns, enabling targeted interventions and assistance programs and contribute to designing initiatives that address the specific needs of migrant populations and contribute to the promotion of social welfare.

For the general public, our Pie Chart data visualisation provides accessible insights into the causes, consequences, and trends of migration in India. It serves as an informative tool for understanding the reasons behind Indian migration.

The importance of our data visualisations spans various critical aspects of socio-economic dynamics in India. Firstly, the unemployment rate visualisation, empowers policymakers to make informed decisions about employment strategies and interventions, crucial for addressing socioeconomic disparities. Also, by addressing unemployment through targeted strategies informed by the unemployment rate visualisation, individuals are more likely to find stable employment opportunities within their home country. Secondly, the reason for migration visualisation, encompassing gender-specific data, aids in crafting policies tailored to the diverse needs of migrant populations, fostering social inclusion and integration. Additionally, understanding the specific reasons for migration enables policymakers to implement measures that directly address the root causes, creating a more conducive environment for people to stay. Thirdly, wages disparities across countries, presented in our visualisations, offer insights into global economic trends, enabling policymakers to make informed decisions on labour-related policies for sustainable development. Also, the economic planning facilitated by wages disparities and demographic insights further aids in creating opportunities, improving living standards, and reducing the allure of seeking employment abroad. Moreover, our visualisations on international migration from India and the impact on metropolitan areas provide essential data for understanding the scale and implications of global diaspora, guiding diplomatic and urban planning decisions. Additionally, the age-group analysis from the population visualisation aids in demographic planning, influencing policies related to education, healthcare, and workforce development. The visualisation of disease prevalence affecting migration provides invaluable information for public health planning, allowing targeted healthcare strategies to address the unique health needs of migrants, thus enhancing overall well-being. In summary, our data visualisations play a pivotal role in fostering informed decision-making, socio-economic inclusivity, and sustainable development in India thereby reducing the chances of people emigrating.

## **Visualisation Purpose**

What questions will the user be able to answer with your visualisation? List the possible benefits of the completed visualisation.

## **Project Schedule**

Make sure that you plan your work so that you can avoid a big rush right before the final project deadline. Write this in terms of weekly deadlines.

**Week 3**

**Activity**

1. Conducted research on migration information.
2. Filtered out several suitable countries for our project and discussed with team members to decide which one was more suitable.
3. Read the Project Process Book instructions.

**Deadline for this week: 9 September 2023**

**Week 4**

**Activity:**

1. Discuss and finalize the project title.
2. Identify and retrieve datasets from online sources.
3. Determining the titles for the charts.
4. Search for suitable charts corresponding to each dataset.
5. Participated in and contributed to Stand-Up 1 meeting in class.

**Deadline for this week: 16 September 2023**

**Week 5**

**Activity:**

1. Research on Indian migration details to gain further insights and a deeper understanding of the various aspects of migration in India.
2. Research datasets that related to India.
3. Trying different research methods and sources to uncover diverse and valuable information that could contribute to the project's dataset collection.

**Deadline for this week: 23 September 2023**

**Week 6**

**Activity:**

1. Presented the datasets collected by each team member and discuss the which datasets are suitable for the project.
2. Identify the 6 datasets required for the project.
3. Renamed the dataset titles to enhance clarity and compatibility with the charts.

**Deadline for this week: 30 September 2023**

**Week 7**

**Activity:**

1. Determine and assign the types of charts for each dataset.
2. Generate 6 CSV files by extracting the essential data from the datasets we found.
3. Assign development responsibilities for each chart to every team member.

**Deadline for this week: 7 October 2023**

**Week 8**

**Activity:**

1. Research on the chart templates as references for development.
2. Create a new project in Figma to sketch the chart prototype: <https://www.figma.com/file/UWxd1ZE9F5eeS1g9glI1cY/Untitled?type=design&node-id=0%3A1&mode=design&t=lHaHPkCtoIW3EPjC-1>
3. Conduct research and commence the development of the charts.

**Deadline for this week: 14 October 2023**

**Week 9**

**Activity:**

1. Should filter out all necessary data and organize it into relevant CSV files.
2. Completed the development of at least one chart.
3. Should finish sketching the chart prototype in Figma.

**Deadline for this week: 21 October 2023**

**Week 10**

**Activity:**

1. Completed the development of at least one chart.
2. Identified and selected the cover page for the Project Process Book.
3. Participated in and contributed to Stand-Up 2 meeting in class.
4. Compiled a comprehensive list of all required features for each chart in point form within the Project Process Book.

**Deadline for this week: 28 October 2023**

**Week 11**

**Activity:**

1. Continued development on the remaining charts.
2. Completed the development of the landing page for the charts.
3. Progressed on the Project Process Book of the data source.

**Deadline for this week: 4 November 2023**

**Week 12**

**Activity:**

1. Completed the development of at least two charts.
2. Accomplished the task of finishing the design sketches through hand drawings.
3. Participated in and contributed to Stand-Up 3 meeting in class.

**Deadline for this week: 11 November 2023**

**Week 13**

**Activity:**

1. Should complete the development of the last two charts.
2. Should complete the addition of explanatory paragraphs to at least three charts, enhancing the understanding and context of the visualized data.
3. Complete the Project Process Book.

**Deadline for this week: 18 November 2023**

**Week 14**

**Activity:**

1. Check the Project Process Book to ensure accuracy and completeness.
2. Sumit the Project Process Book including all the necessary files.
3. Participated in and contributed to Stand-Up 4 meeting in class.

**Deadline for this week: 20 November 2023**

# **Data**

## **Data Source**

From where and how are you collecting your data? Provide a link to your data sources. What type of data set is it (e.g., table, network, field)? What are the attributes in your data set and what type of data are the values (i.e., categorial, ordinal, interval, ratio/quantitative)? Is there any data in the set that will not be included in your visualisation? Why?

NOTE: Make sure that the data can be used to answer the questions outlined in Section 1.2.

* + - 1. **Unemployment Rate of 2023 in India:**

We obtained the unemployment rate in India dataset from an article, and the data source is from [1]. The dataset is presented in tabular form within the article. So, we can input all the data into our CSV file by following the tabular form.

There are 2 attributes in the original dataset, and they are "State" and "% of Unemployment Rate in India". Both attributes are being utilized in our data visualisation. Also, to enhance clarity, we have renamed "% of Unemployment Rate in India" to "Unemployment\_Percentage\_in\_2023" and the “State” remains the same.

Our “State” cannot be measured in numerical terms, and it can be divided into categories, so it is qualitative data. Also, it is categorical data because it belongs to a specific category, and there is no inherent order or numeric significance to the states. Then, "Unemployment\_Percentage\_in\_2023" is quantitative because it involves numerical values that can be measured and compared. Moreover, it specifically ratio data, because it represents a percentage, and ratios in this context, have a clear definition of zero. For instance, a value of 0% indicates the complete absence of unemployment, making it a quantitative indicator with meaningful ratios for analysis and comparison.

<https://www.studyiq.com/articles/unemployment-rate-in-india/>

* + - 1. **Indian Migration Reason:**

We obtained the Indian migration reason dataset from an article, and the data source is from [2]. The dataset is presented in tabular form within the article. So, we can input all the data into our CSV file by following the tabular form.

There are 4 attributes in the original dataset, and they are "reason of migration", "female", "male" and "person". The selected attributes are “reason of migration", "female" and "male". The decision to exclude "person" is based on its lack of meaningful explanation in the articles under consideration.

“reason\_for\_migration” does not involve numerical measurements so it is qualitative data. Then, “reason\_for\_migration” is categorical data also because it represents different categories or reasons for migration. Moreover, both "male" and "female" are quantitative and ratio data due to their incorporation of numerical measurements expressed in percentages and the presence of a meaningful zero point further emphasizes their quantitative nature.

<https://pib.gov.in/PressReleasePage.aspx?PRID=1833854>

* + - 1. **Wages of Various Countries:**

We obtained the wages of various countries' dataset from a working paper website, and the data source is from [3]. The dataset is presented in a CSV file within the working paper website. So, we can download the CSV file then collect and filter the required data.

The original dataset comprises 63 attributes, and for data visualization, the selected attributes include "year," "country\_code," "country\_name," "hw3wl\_us," and "mw3wl\_us." For improved clarity, we have renamed "hw3wl\_us" to "hour\_wages" and "mw3wl\_us" to "monthly\_wages." The rationale for not selecting the other attributes is that they exhibit similarity and adhere to a different standard. For instance, attributes like "hw1wl\_current" represent the mean wages of hourly workers in standard data without calibration, specifically denoted in Lex, LCU (reported), while "hw2wl\_current" represents the mean hourly wages with country-specific calibration, also in Lex, LCU (reported).

The “year” is ordinal data because it indicates a chronological order. Next, "country\_code" and "country\_name" are categorical data because they are used to represent categories and labels without inherent numerical significance. Moreover, "monthly\_wages" and "hourly\_wages" are quantitative data and under the category of ratio data because they are numerical and can be measured.

<https://www.nber.org/research/data/occupational-wages-around-world-oww-database>

* + - 1. **Disease leads to migration dataset:**

We obtained the disease leads to migration dataset from an article, and the data source is from [4]. The dataset is presented in an image-tabulated form within the article. So, we can input all the data into our CSV file by following the image-tabulated form.

The original dataset comprises 16 attributes, and for data visualization, the selected attributes include "year," "Disease\_Category," "female," "male", "<40 years", ">=40 years" and "prevalence\_among\_migrants". For improved clarity, we have renamed ">=40 years" to "40\_years\_above" and "<40 years" to "39\_years\_below". The rationale for not selecting the “Prevalence among non-migrants” attributes is it is not related to our title. Also, given the large number of selected attributes, therefore excluded the “overall prevalence” attribute to simplify and focus our data visualisation.

<https://www.researchgate.net/publication/319659904_Determinants_of_internal_migrant_health_and_the_healthy_migrant_effect_in_South_India_A_mixed_methods_study>

* + - 1. **Immigration From India to Other Countries**

We obtained the immigration from India to other countries dataset from an article, and the data source is from [5]. The dataset is presented in an image-tabulated form within the article. So, we can input all the data into our CSV file by following the image-tabulated form.

There are 3 attributes in the original dataset, and they are "Metropolitan Area", "Immigrant Population from India" and "% of Metro Area Population". Three of the attributes are being utilized in our data visualisation. Also, to enhance clarity, renamed "% of Metro Area Population" to "percentage\_of\_metro\_area\_population" and the replaced spaces with underscores and converted to lowercase such as "metropolitan\_area" and "immigrant\_population\_from\_india".

"metropolitan\_area" is categorical data because it presents different metropolitan areas. "immigrant\_population\_from\_india" and "percentage\_of\_metro\_area\_population" are ratio data because they involve numerical values and they can be measured and compared.

<https://www.migrationpolicy.org/article/indian-immigrants-united-states>

* + - 1. **Population of India:**

We obtained the population of India dataset from the United Nations data bank, and the data source is from [6]. The dataset is presented in CSV files within the United Nations data bank. So, we can download the CSV file then collect and filter the required data.

The original dataset includes 12 attributes along with an age range from 0 to 100 attributes. Due to the multitude of attributes, we have consolidated the age range variables into a unified attribute named "age group" for simplification. So, the finalized set of attributes for data visualization comprises 3 attributes which are "Year", "Age\_Group" and "Value". The rationale for not selecting the other attributes is they are not related to our title.

"Year" is ordinal data because it represents different years and has a meaningful order but no consistent interval between them. "Age\_Group" is categorical data because it represents distinct groups into which classified based on age. "Value" is ratio data because it represents a numerical measure that has a true zero point and allows for meaningful ratios.

<https://population.un.org/wpp/Download/Standard/Population/>

## **Data Processing**

Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented? Will you be deriving any variables?

Describe clean up process that was implemented. Explanation and calculation of derived variables (if used).

Extensive data cleaning was conducted on the wages of various countries and the population of India due to the substantial volume of data that required clarification. For instance, focusing on the population of India, we filtered out non-Indian entries, selected data from the past 5 years, and aggregated 100 single-age population attributes into distinct age ranges, including Age 0-18, Age 19-30, Age 31-40, Age 41-50, Age 51-60, Age 60-70, Age 71-80, Age 81-90, Age 91-99 and Age 100+. Subsequently, we summed up the values within each age range. Additionally, unrelated attributes underwent cleanup for improved dataset clarity. Following these adjustments, the dataset now comprises a total of three attributes. The numeric value of 50 is a derived result from the data processing steps.

Subsequently, dealing with the wages of various countries, which presented extensive data, even after filtering for the required years, the dataset still comprised 36013 entries and 63 attributes. Furthermore, the dataset contains numerous similar attributes that can be cleaned. For instance, "mw1wl\_current" signifies monthly standard data without calibration in Lex, LCU (reported), while "mw3wl\_current" represents monthly country-specific calibration with imputation in Lex, LCU (reported), and so forth. To clean up the dataset for relevance to the project, a multi-step process was employed. Initially, numerous unrelated attributes were removed. Then, the dataset was filtered for attribute "y3," denoting industry code, selecting only JC. Subsequently, entries with the number 74 in attribute "y4," signifying occupation code, were filtered out. Lastly, the dataset was refined to include data from seven specific countries, including India. Following these filtering and cleaning steps, the dataset was ultimately condensed to retain only 5 attributes and 49 data points.

There is no extensive data cleaning conducted on the datasets of Unemployment Rate of 2023 in India and Immigration from India to Other Countries. There are some name replacements for those datasets.

In the Indian Migration Reason dataset, we eliminated one attribute that lacked a meaningful explanation in the articles, but the quantities of the data still remain the same.

In the Disease leads to migration dataset, we eliminated one attribute that lacked a meaningful explanation in the articles, but the quantities of the data still remain the same.

# **Requirements**

## **Must-Have Features**

These are features without which you would consider your project to be a failure. Were you able to deliver all the promised features? If not, explain why.

**Heatmap Chart: Wages of various countries**

**Idea:**

* Mouse over show the tooltip
* Have the color legend show the wages range
* Few countries button that allow to filter out their details in table form and after clicked the filter button the chart will also show the filter country and other countries will no show become transparent.

**Choropleth map: Unemployment Rate of India for the year 2023**

**Ideas:**

* Allow to zoom when clicking any states in map and zoom manually such as scroll up and down or double click.
* Can be filter the data using search bar or button
* After filter will also display the filtered statss in map.
* Have a color legend to show the range.
* After filter will display to table to show more clear data.
* Mouse over in the states will show the tooltip.
* Have click on function and display the tooltip function and fill the clicked state.

**Sunburst chart: Population of India**

**Ideas:**

* Have click on function on the year
* After click the year or age will show the result in center
* Mouse over to show to tooltip.

**Pie Chart: Reason of India people migrate**

**Ideas:**

* Have mouse over to show the tooltip
* Have 2 filter button for show the male and female migration reason.

**Clustered stacked bar chart**

**Ideas:**

* Mouse over will show the tooltip.
* Have color legend to display the categories.

**Bubble chart:**

**Ideas:**

* Mouse over will show the tooltip.
* Filter function.
* Have color legend to display the range.
* Search bar to filter the different number of bubbles.

In our pursuit of creating an effective Heatmap Chart for visualizing the wages of various countries, we aimed to enhance user experience through several specific features. First and foremost, we integrated mouse-over tooltips, providing users with immediate access to detailed information when hovering over specific data points. This feature aimed to enhance data visibility and user interaction. Additionally, we incorporated a color legend, strategically designed to illustrate the wage ranges. This color-coded representation not only added a visual element to the data but also facilitated easier interpretation for users. Furthermore, one of our key objectives was to introduce an interactive element through the integration of filter buttons. These buttons were intended to empower users to selectively focus on specific countries of interest. When a filter button corresponding to a particular country was clicked, the Heatmap Chart would respond dynamically by emphasizing the chosen country and rendering others transparent. The absence of these features would be deemed a failure for the project for us. Therefore, we are committed to delivering the functionalities as promised.

## **Optional Features**

Those features which you consider would be nice to have, but not critical. Were you able to deliver any of these extra features?

# **Visualisation Design**

How will you display your data? Provide some general ideas that you have for the visualisation design. Include sketches of your design. Include at lease 2-3 alternative ideas for your visualisation. Describe and justify your choice of visual encoding and idioms. Show the evolution of your design. How has it progressed? Justify the visualisation idioms you have chosen to represent your data.

Description (including screen shots) and explanation of final design.

[NOTE 1: You are encouraged to provide your own structure to this section (i.e., section headings etc).

NOTE 2: You MUST show evidence of iterative design (i.e., sketches of alternative and preliminary designs). ]

Include screenshots of final design.

**Figma link:** <https://www.figma.com/file/UWxd1ZE9F5eeS1g9glI1cY/Untitled?type=design&node-id=0%3A1&mode=design&t=S168Q81o0eQTJVIU-1>

**Draft of preliminary designs**

**Heatmap Chart: Monthly wages between few countries**

A paper with a chart and a graph

Description automatically generated with medium confidence

**Choropleth map: Unemployment Rate of India for the year 2023**

A white board with a black and white drawing

Description automatically generated with medium confidence

**Sunburst chart: Population of India**

A diagram of a sunburst chart

Description automatically generated

**Pie Chart: Reason of India people migrate**

A graph on a piece of paper

Description automatically generated

**Clustered stacked bar chart**

A graph of bar graph and bar graph

Description automatically generated

**Bubble chart:**

A drawing of circles and lines

Description automatically generated

# **Validation [optional - Bonus Points]**

Test your visualisation with users and report the results.

# **Conclusion**

Provide a summary of the project and what you learnt from doing it.

# **References**

# References

|  |  |
| --- | --- |
| [1] | manishsiq, "Unemployment Rate of India 2023, Complete State-wise List," 8 November 2023. [Online]. Available: https://www.studyiq.com/articles/unemployment-rate-in-india/. [Accessed 19 November 2023]. |
| [2] | P. Delhi, "Migration in India, 2020-2021," 14 June 2022 . [Online]. Available: https://pib.gov.in/PressReleasePage.aspx?PRID=1833854. [Accessed 19 November 2023]. |
| [3] | "Occupational Wages around the World (OWW) Database," National Bureau of Economic Research (NBER) , [Online]. Available: https://www.nber.org/research/data/occupational-wages-around-world-oww-database. [Accessed 19 November 2023]. |
| [4] | Warren Dodd, Sally Humphries, Kirit Patel and Shannon, "Determinants of internal migrant health and the healthy migrant effect in South India: A mixed methods study," September 2017. [Online]. Available: https://www.researchgate.net/publication/319659904\_Determinants\_of\_internal\_migrant\_health\_and\_the\_healthy\_migrant\_effect\_in\_South\_India\_A\_mixed\_methods\_study. [Accessed 19 November 2023]. |
| [5] | A. H. a. J. Batalova, "Indian Immigrants in the United States," 7 December 2022. [Online]. Available: https://www.migrationpolicy.org/article/indian-immigrants-united-states. [Accessed 19 November 2023]. |
| [6] | "Population Division," United Nations, 2022. [Online]. Available: https://population.un.org/wpp/Download/Standard/Population/. [Accessed 19 November 2023]. |