Gagannath Dalavai

U45739433

Business Analytics & Information Systems

Spring 2024

Final Project Report

Global Insights: COVID-19 Impact & Employment Dynamics

Data Visualization

Prof. Johannes Reichgelt, PhD

Introduction

The COVID-19 pandemic, which emerged in late 2019, has rapidly evolved into one of the most significant global health crises in recent history, profoundly impacting various aspects of society, including public health, the economy, and employment worldwide. As the pandemic continues to unfold, there is an urgent need for comprehensive analysis and understanding of its multifaceted dimensions to inform effective responses and strategies for mitigation and recovery.

The COVID-19 pandemic has also posed unprecedented challenges globally. With millions of confirmed cases and deaths reported worldwide, the pandemic has underscored the interconnectedness of nations and the need for coordinated, data-driven responses to mitigate its spread and mitigate its impact. Understanding the intricate relationship between the pandemic and employment trends is essential for devising effective policy responses and support mechanisms for affected individuals and communities. In this project, I have used Power BI to analyse covid data by linking it with employment data and population data from 2020-2023 to see if covid had any impact on employment worldwide. As part of my analysis, I have explored trends of COVID cases and deaths by year, employment count by occupation, correlations of employment, and population along with COVID data.

This project sets out an ambitious goal of gathering, linking, and analysing data from multiple sources to provide a comprehensive understanding of the COVID-19 pandemic's impact. The integration of disparate datasets, including COVID-19 case records, and employment figures, underscores the project's ambition to offer holistic insights into the pandemic's multifaceted dimensions and its implications for employment dynamics globally. Furthermore, the project's ambition is evident in its analytical approach. Rather than focusing solely on descriptive statistics, it delves into deeper analyses, such as correlations between COVID-19 incidence and employment trends and geographic disparities. This level of complexity demands advanced analytical techniques, including data normalization.

Here are the research questions that I have addressed in my project.

1. What proportion of global COVID-19 cases occurred in each continent from 2020 to 2023?

2. How does the median age of populations in the top 20 countries with the highest COVID-19 case counts vary across different continents, and what insights can this demographic analysis provide regarding vulnerability and susceptibility to the virus?
3. How do the trends in COVID-19 cases vary among different continents from 2020-2023?
4. How is employment distributed across various occupations and countries, and what insights can be gleaned from analysing employment trends by occupation type?
5. Is there a relationship between the number of employed individuals per 100,000 people and the incidence of COVID-19 cases per 100,000 people?
6. How has employment distribution varied across different age groups over the years?
7. How does population density correlate with the number of employed individuals per 100,000 people?
8. How do the rates of COVID-19 cases per 100,000 people differ among countries?
9. How does the interactive dashboard on COVID-19 analysis facilitate data exploration such as total cases, and deaths across different continents and periods?

Methodology

In this project, to generate the visualizations to answer the research questions I have used four different datasets taken from different sources. Each data source has its importance. Necessary cleaning operations are performed on all datasets for ease of analysis. To perform data cleaning I have used Power Query Editor for some datasets and others I have cleaned them within Excel and loaded them into Power BI directly. Here is the list of datasets that I have used for my analysis.

1.COVID-19 Data - **https://ourworldindata.org/covid-cases**

This dataset contains updated COVID-19 data from 2020-2024. I have taken data from 2020-2023. Some of the useful columns in this dataset are:

i. Continent – contains the list of continent names in the world.

ii. Country - contains the list of country names in the world.

iii. Date - The date on which the case is reported.

iv. Total Cases - Total cases by country as of that date.

v. New Cases - Number of new cases reported for that day.

vi. Total Deaths - Total deaths by country as of that date.

vii. New Deaths - Number of new cases reported for that day.

viii. Year-Country- This column was created manually so that it can be used to link with other tables. This column is created using the concatenate function on the country column and date column (YEAR (date)).

I have further separated this dataset into 3 different datasets. They are:

i. Daily Covid Data (original dataset – owid-covid-data (1).csv)

ii. Overall Covid Data – It has overall cases and deaths by country from 2020-2023 (owid-covid-data.csv)

iii. Covid Data by year – It has year year-wise count of cases and deaths from 2020-2023 (Covid Data_Year Wise 2020-2023.xlsx)

2.Employment Data - https://rshiny.ilo.org/dataexplorer38/?lang=en

(EMP_TEMP_SEX_AGE_OCU_NB_A-filtered-2024-04-22.csv)

This dataset (Employment by sex, age, and occupation(thousands) – Annual**)** is taken from the ILOSTAT website- International Labour Organization Statistics. The employed comprise all persons of working age who, during a specified brief period, were in one of the following categories: a) paid employment (whether at work or with a job but not at work); or b) self-employment (whether at work or with an enterprise but not at work). Some of the columns in this dataset are:

i. Country

ii. Age Group: Age group of employed individuals

iii. Gender

iv. Occupation Type

v. Value: Number of employed individuals

vi. Year- 2020-2023

Like the previous dataset, I performed data-cleaning operations on this too. For example, the Value column has decimal values which is number of employed in thousands. For ease of

analysis, I have created a new column with a value *1000 so that it is represented as a whole number instead of a decimal.

3.Population Dataset (World Population as of 2023.xlsx) – **World Population by Country 2023 (Live) (worldpopulationreview.com) (Historical Estimates of World Population (census.gov))**

The population dataset is taken from worldpopulationreview.com which in turn is sourced from census.gov. This dataset contains the population of countries worldwide as of 2023. The columns in this dataset are:

i. Country

ii. Population

iii. Population Density (per sq km) - a measure of the average number of people living within a defined area.

The reason why I took this dataset is because using this dataset I have created two calculated fields in the covid dataset and employment dataset. They are:

i. Covid cases per 100,000 – Total cases/Total population

ii. Employed individuals per 100,000 – Number of employed people/ Total Population

4.Vaccination Data - COVID-19 data (vaccination-data.csv)

This dataset is taken from the WHO website. It consists of vaccination details country-wise. Some of the columns in this dataset are:

i. COUNTRY: The name of the country

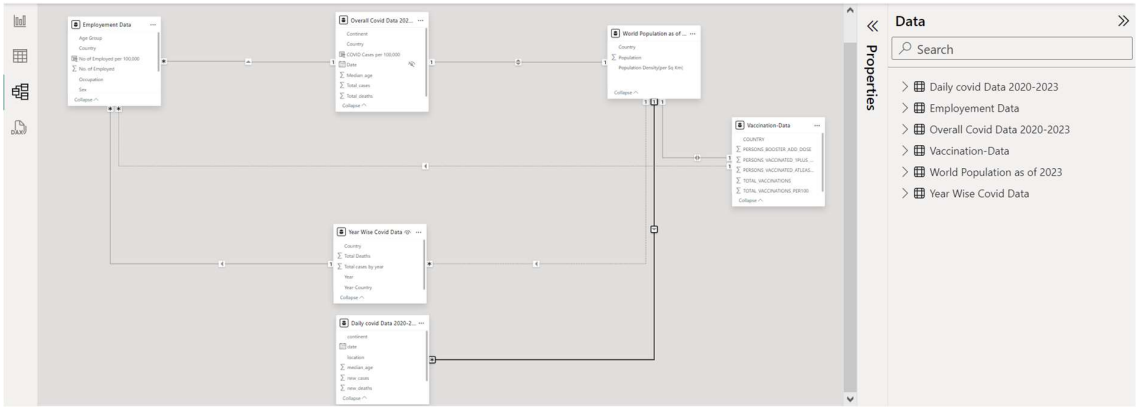
ii. TOTAL VACCINATIONS: Total Number of COVID-19 vaccine doses administered in the country.

iii. PERSONS VACCINATED 1PLUS DOSE: The number of individuals who have received at least one dose of the vaccine.

iv. TOTAL VACCINATIONS PER100: The total number of vaccine doses administered per 100 people in the population.

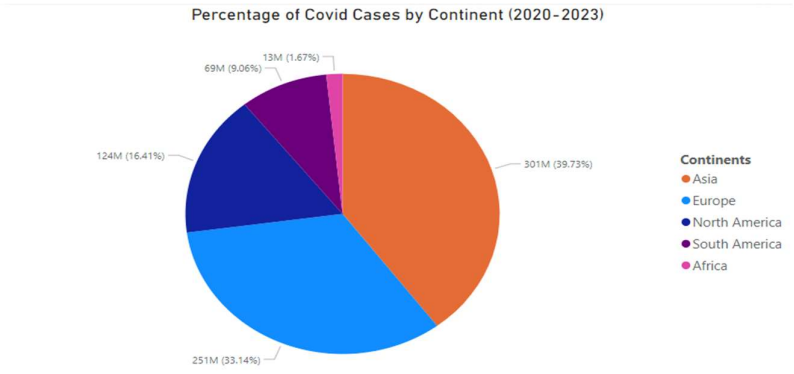
v. PERSONS BOOSTER ADD DOSE: The number of individuals who have received a booster or additional dose of the vaccine.

Here is the model view of my Power BI data source:



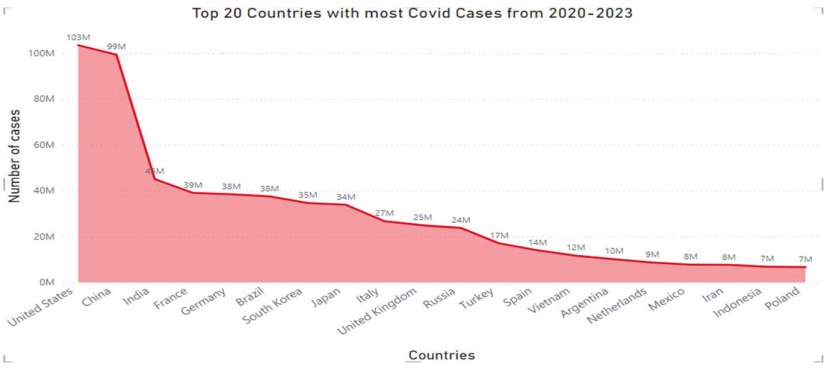
Analysis

1.What proportion of global COVID-19 cases occurred in each continent from 2020 to 2023?

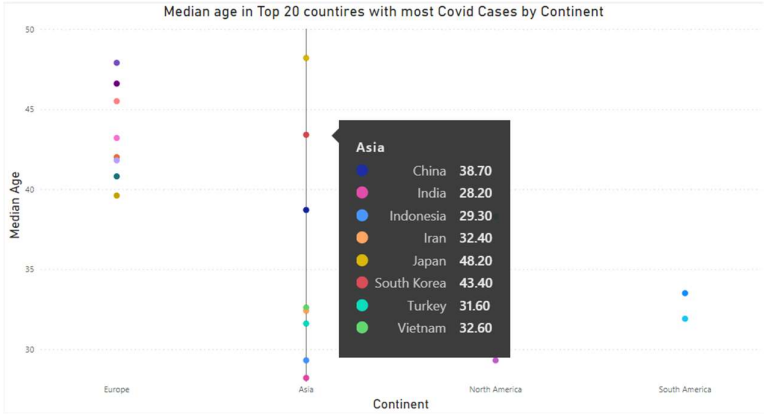


To answer the above research question, I have used a pie chart. From the pie chart we can observe that among the five continents, Asia has the highest proportion of COVID cases from 2020-2023 which is 39.73% i.e. a total of 301 million cases from 2020-2023. We can infer that countries in the Asia continent are highly impacted due to COVID-19.

2.How does the median age of populations in the top 20 countries with the highest COVID-19 case counts vary across different continents, and what insights can this demographic analysis provide regarding vulnerability and susceptibility to the virus?

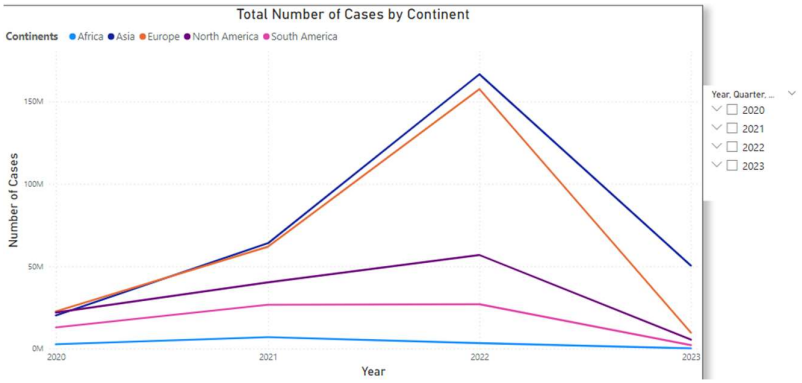


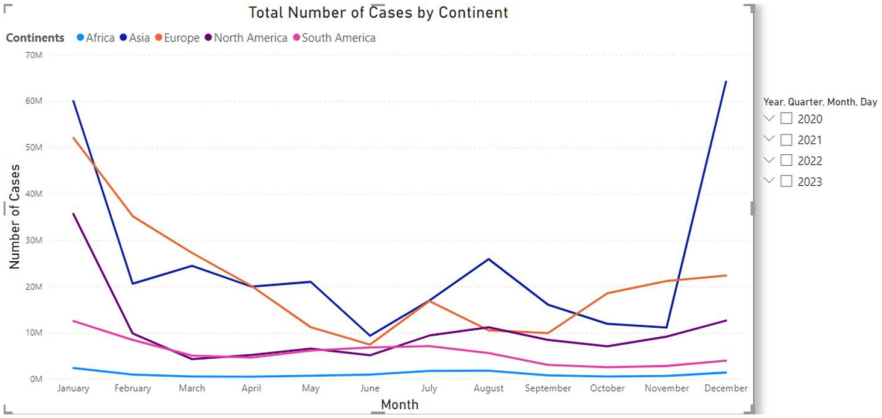
In order to address the above question, I have first generated an area chart that gives the top 20 countries with the most number of covid cases. It is clear that US has the most number of cases followed by China and India. Now using this chart, I have designed a line chart that shows the median age of people in all the above top 20 countries by continent.



From the above line graph, we can see the median age for countries in Asia. We can infer that in India people with a median age of 28.2 are highly infected with covid. Likewise in Japan, people in their late 40's are affected highly due to COVID-19.

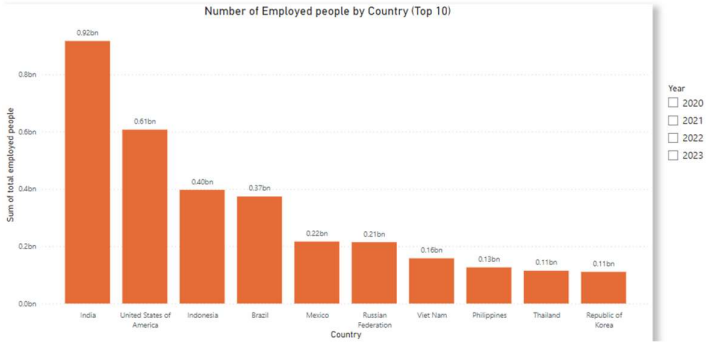
3.How do the trends in COVID-19 cases vary among different continents from 2020-2023?

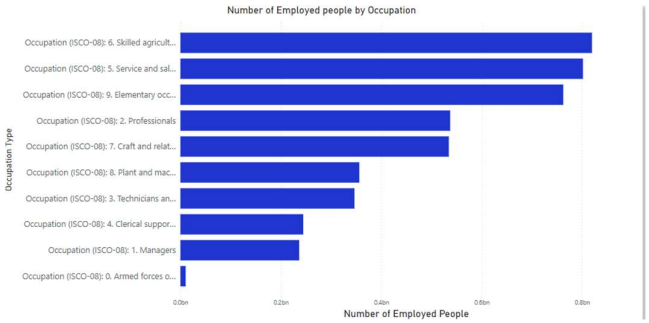




Here I have again used a line chart to know the trend of cases. I have put the year in the filter so that you can see the trend in COVID cases year-wise. The date is represented on the x-axis as a hierarchy. 1st picture shows the year hierarchy while in 2nd one is the month hierarchy. In 2020-2023 it is always Asia with the highest covid cases followed by Europe. When we look at by month, Europe recorded the most number of cases from Feb-Apr (2020-2023). If we look at the trends, from 2022 the cases started reducing in Asia and Europe.

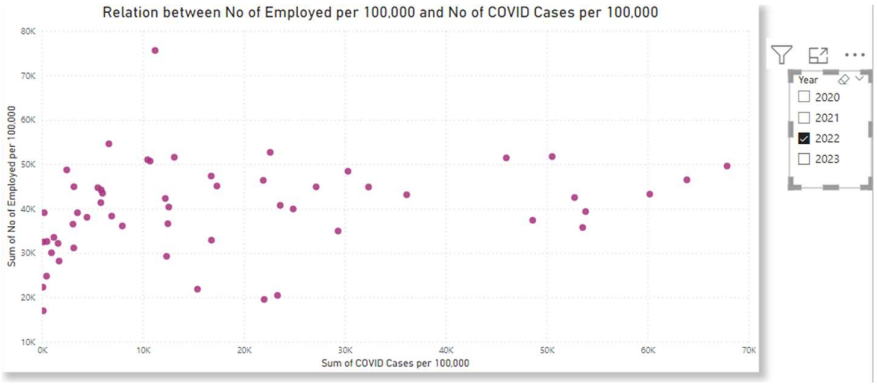
4.How is employment distributed across various countries, occupations, and what insights can be gleaned from analyzing employment trends by occupation type?





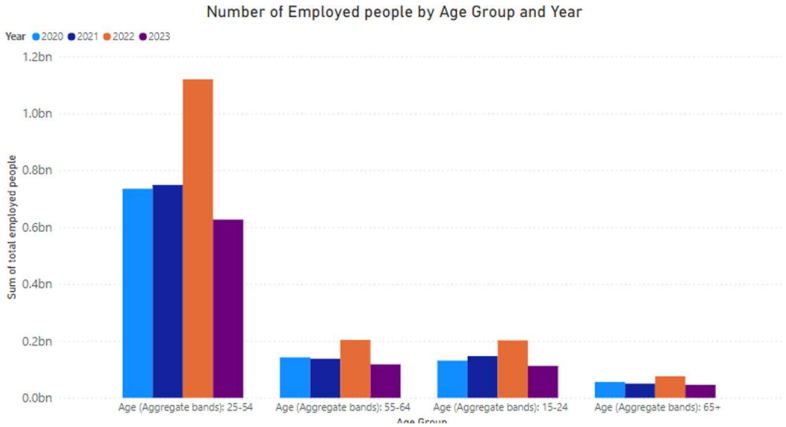
The column chart shows the number of employed by country and the horizontal bar chart shows the number of employed by occupation from 2020-2023. We can infer that overall India tops the list with 0.92 billion employed individuals and most of them belong to occupation type of skilled agriculture, forestry and fishery workers.

5.Is there a relationship between the number of employed individuals per 100,000 people and the incidence of COVID-19 cases per 100,000 people?



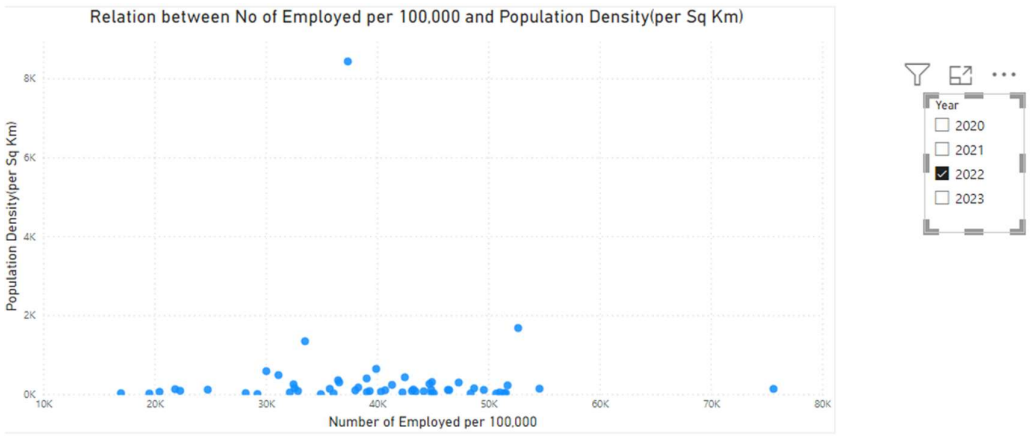
I used a scatter plot to address this research question. Covid cases are shown on the x-axis and the number of employed is shown on the y-axis. To have a more meaningful analysis I have normalised the covid cases and number of employed per 100,000 using the population dataset. For the year 2022 we can see the relation is weak positive which means there is not much impact on employment due to covid in 2022. Similarly, by choosing the year in the filter we can see the relation for other years also.

6.How has employment distribution varied across different age groups over the years?



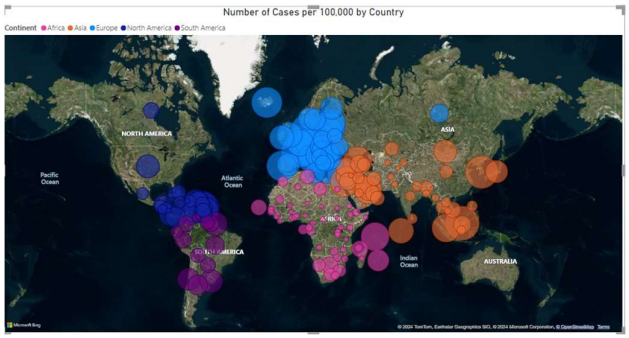
Here is the clustered column chart that shows the number of employed individuals by age group and by year. Most of the individuals who are employed belong to the age group of 25-54 and this continued for all the years. Also, for each age group, 2021 has seen the highest figures which says employment is better in 2021.

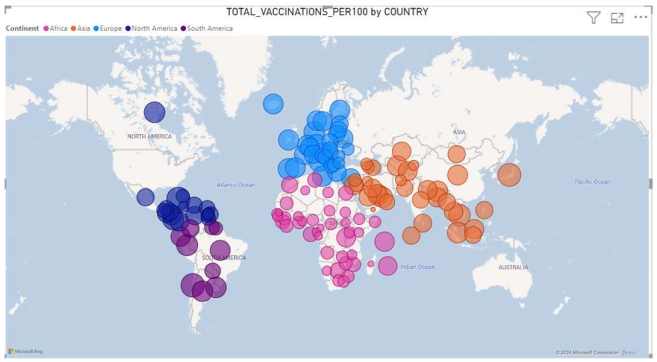
7.How does population density correlate with the number of employed individuals per 100,000 people?



The scatter plot shows the relation between number of employed per 100000 and population density (in sq km). The data points are not scattered neither linearly nor nonlinearly which means both are independent and has relation.

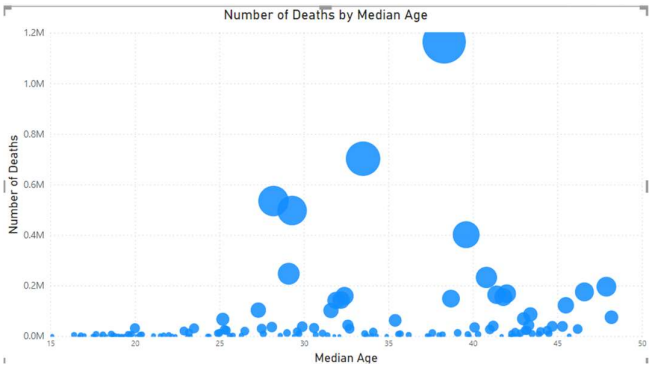
8. How do the rates of COVID-19 cases per 100,000 and total vaccinations per 100 differ among countries?





The map visualizations show number of cases per 100,000 and total vaccinations per 100 for countries. In the first map, the size of the bubble represents the number of COVID-19 cases per 100,000. The larger the size, the higher the cases in that country. Similarly, the 2nd map shows the total vaccinations per 100. The larger the size of the bubble, the higher the vaccinations in that country. For better understanding, I have differentiated each continent by adding them in legends in both maps.

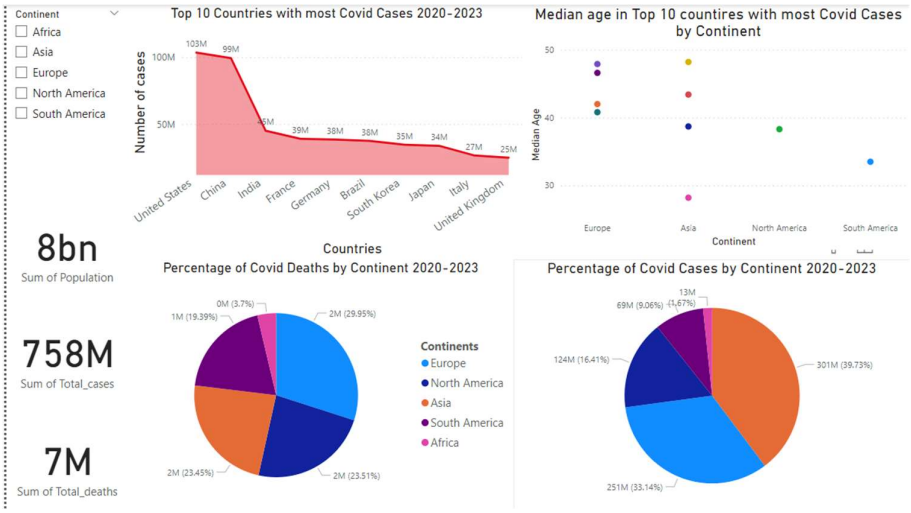
9.What is the global count of deaths by median age



The bubble size represents the number of deaths for that median age. So globally the individuals with a median age of 38.3 suffered the highest number of deaths. This shows people within that age limit are highly infectious to COVID-19 and should be more careful.

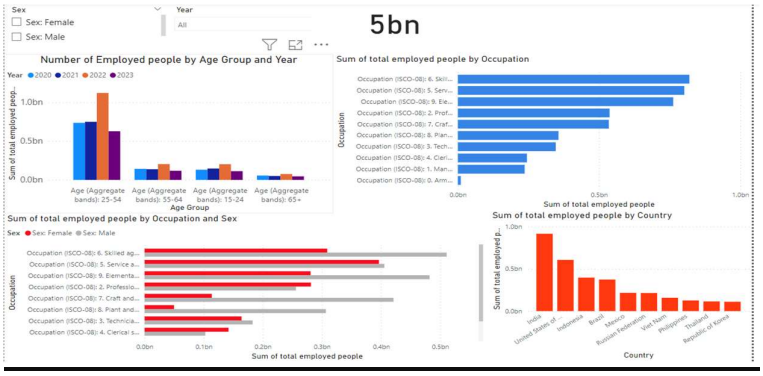
COVID-19 Dashboard

I have also developed a dashboard using the above visualizations so that it would be easy to multiple insights at same time.



This dashboard is responsive which means choosing any value in one chart reflects changes in other as well. It gives comprehensive analysis of covid deaths, cases, median age by continent and country.

Employment Dashboard



This dashboard gives overall insights on employment by age group, occupation type, country, and gender. You can choose the gender or year from filter and you can get the corresponding results.

Conclusion

This analysis provides valuable insights into various aspects of the COVID-19 pandemic and employment trends across different continents and demographics.

For the research questions regarding COVID-19, the visualizations highlighted the disproportionate impact of the virus across continents, with Asia bearing the highest proportion of cases from 2020 to 2023. Additionally, the examination of median age among countries with the highest case counts underscored the vulnerability of certain age groups, offering insights into demographic factors influencing susceptibility to the virus. In terms of employment trends, the analysis showcased the distribution of employment across different countries and occupations, shedding light on the diverse workforce compositions and employment opportunities globally. The scatter plot analysis further explored the potential relationship between COVID-19 incidence and employment rates, indicating a weak positive.

For this analysis, I have concentrated majorly on covid data (cases and deaths) and employment data. In the future, I would like to extend my analysis more on vaccination data. Here are some of the future research questions I have identified:

1.Explore how the post-pandemic economic landscape is shaping employment dynamics, including shifts in job sectors, employment opportunities, and workforce participation rates.

2.Evaluate the effectiveness of COVID-19 vaccines over time, considering factors such as vaccine types, dosing regimens, and population demographics

3.Explore the role of environmental factors, including air quality, temperature, humidity, and population density, in shaping COVID-19 transmission dynamics.

4. How do disparities in access to education and digital literacy skills affect employment opportunities and economic resilience during times of crisis?

5. What role do digital technologies and remote work arrangements play in mitigating the economic impacts of the pandemic on different sectors and industries?