

Applied Statistic and Data Visualization
Assignment

By
Syed Ali Murtaza

1. Part-01

Briefly summarises your chosen scenario as you understand it

In this part we have given different files with the name of ASDV_Data_Repository_2024. So from these datasets I have chosen the Scenario 1 - Population Dynamics. In this scenario-1 we have given 3 files

1. Population Figures & Projections.csv
2. Population Figures & Projections Metadata.xlsx
3. Country Groupings.xlsx

Now I will briefly summarize my understanding for each file.

Population Figures & Projections:-

This file contains information related to population over a time period.

- The column Country Name and Country code are used to identify different countries.
- Time and Time Code are used to identify the time period in Years.
- Population male & female are used to identify the total male and female population
- Population growth is used to identify the annual percentage growth in population.
- Rural and Urban population is used to identify the total rural and urban populations.
- Other factors are life expectancy, birth rate, and death rate.

Population Figures & Projections Metadata:-

This file includes description, definitions or units for the columns in the Population Figures & Projections dataset.

Country Groupings:-

This file provides a mapping of countries in terms of regions and income groups.

Initial exploration of data

When I briefly explore the dataset I found some notable observations which I am going to discuss here

- The data includes detailed information on population figures like male population and female population for each country over multiple years plus there is also a growth matrix with name of Population growth (annual%) which can be used to identify trends like period of growth, period of decline for specific countries and region as well.
- The data set includes the information on urban and rural population for each country which can be used to determine trends of shifting people from one place to another place globally or within specific region over a period of time.
- The dataset includes columns like life expectancy at birth, birth rate and death rate. These indicators provide us information that how was the quality of life changes across different countries and regions over a certain period of time.
- The “Country Grouping” file gives us information of grouping of countries across different regions, income level or other metrics. This information will enable us to compare analysis across different group of economies.

Potential Objectives for the Dashboard:-

- I will visualize population trend (including urban population, rural population and total population) over time using the line and stacked column to highlight the difference in population across top 5 most populated countries and other countries as well (for other countries use filter).
- I will compare urban and rural population growth using stacked area chart to identify urbanization patterns across time(year).
- I will plot line a chart for birth rate, annual population growth and death rate at different period of time(years) to identify the pattern that hows the quality of life and healthcare changes across different countries with the passage of time.
- I will introduce a map visualization that shows average life expectancy rate across different countries.
- I will introduce a card which will show the average of natural increase rate across different countries and time(Year).
- I will provide filters for users to explore trends by Year and Country Name, offering a dynamic and interactive experience.

Explains and justifies your choice of data visualisations.

Line and Stacked Column Chart

I used line and stacked column chart to visualize population trends for urban, rural, and total populations across the top five most populated countries.

Reference from Theory:-

Bar and column charts are very useful for comparing values across discrete categories and subcategories. The distinct height and color difference between each category enable us to quickly interpret comparisons without much effort.

Secondly **Temporal Trends**: Line charts are ideal for showing trends over time because humans can easily track the direction and steepness of a line.

Why Appropriate:-

By applying line and stacked column chart I provided a comprehensive view.

- The line in graph focuses on overall trend in population.
- The stacked column focuses on relative distribution of population.

Filled Map

I used the Filled Map to visualize the countries and analyse their life expectancy rate using the colour on a map interactively.

Reference from Theory:-

All the humans have the strong ability to process and interpret maps due to their alignment with their mental model of geography.

Color intensity is a pre-attentive characteristic through which user can quickly understand difference in life expectation rates across different countries without much effort.

I have encoded life expectancy through color gradients, the map will use Color intensity principle to give information about the area with extreme values like low, mid or high life expectancy.

This map uses the spatial positioning and color coding to convey the data quickly as compare to tabular or textual data.

Why appropriate:-

- It gives clear geographical insights.
- Interactive tooltip allow user to see the detail information of life expectancy without cluttering the map.
- The Color-coded shading facility will use allow to do comparisons between different countries.

Stacked Area Chart

I have used stacked area chart to show the difference in between urban and rural population across time(year).

Reference from Theory:-

In the theory we have studied that stacked area chart allow us to view multiple components within 1 view while highlighting their individual trends over different categories or time. It uses different colors to split the area under the line by another categorical variable.

Why appropriate:-

From this chart, viewer can easily differentiate the trends in urban and rural population. The main aim of this graph is to explore and compare urban and rural population trends across time (year). It uses color and stacking which provides clarity.

Card

I used card to represent the average of natural increase rate.

Reference from Theory:-

Card minimizes the visual clutter and visualizes the value without much effort. A card also isolate that specific value from the rest of dashboard so that it will be a main focal point for the viewer.

Why Appropriate:-

The natural increase rate is a important indicator that's why I represent it on a card.

Line chart

I have used the line chart to represent birth rate, death rate and annual population growth over the time.

Reference from theory:-

Multiple line chart are used to compare the data split out by another categorical variable and it is highly effective for showing trends over time. Distinct colors for lines enable quick recognition and differentiation of trends.

Why appropriate:-

This chart helps us to highlights the critical patterns such as period of high birth rates or death rates or population growth and vice versa which will provide actionable insights for analyses.

Explains and justifies your overall dashboard layout, formatting, and composition

Layout:-

The layout of this dashboard uses the top-bottom and left to right hierarchy. The graph of top 5 countries and life expectancy from map are on the top which align with principle of prioritizing key insights. The stacked area chart and line chart provide deeper analytics and are on the bottom left and bottom right.

The interactive filters for country name and year are added which align with best practice.

Formatting:-

- The color coding is used to differentiate b/w urban and rural population in area chart, and to emphasize trend in line chart. This approach is used for the purpose of clear segmentation.
- Titles, legends, and axes are clearly labeled across all visuals.
- Consistent font and style is used throughout the dashboard.

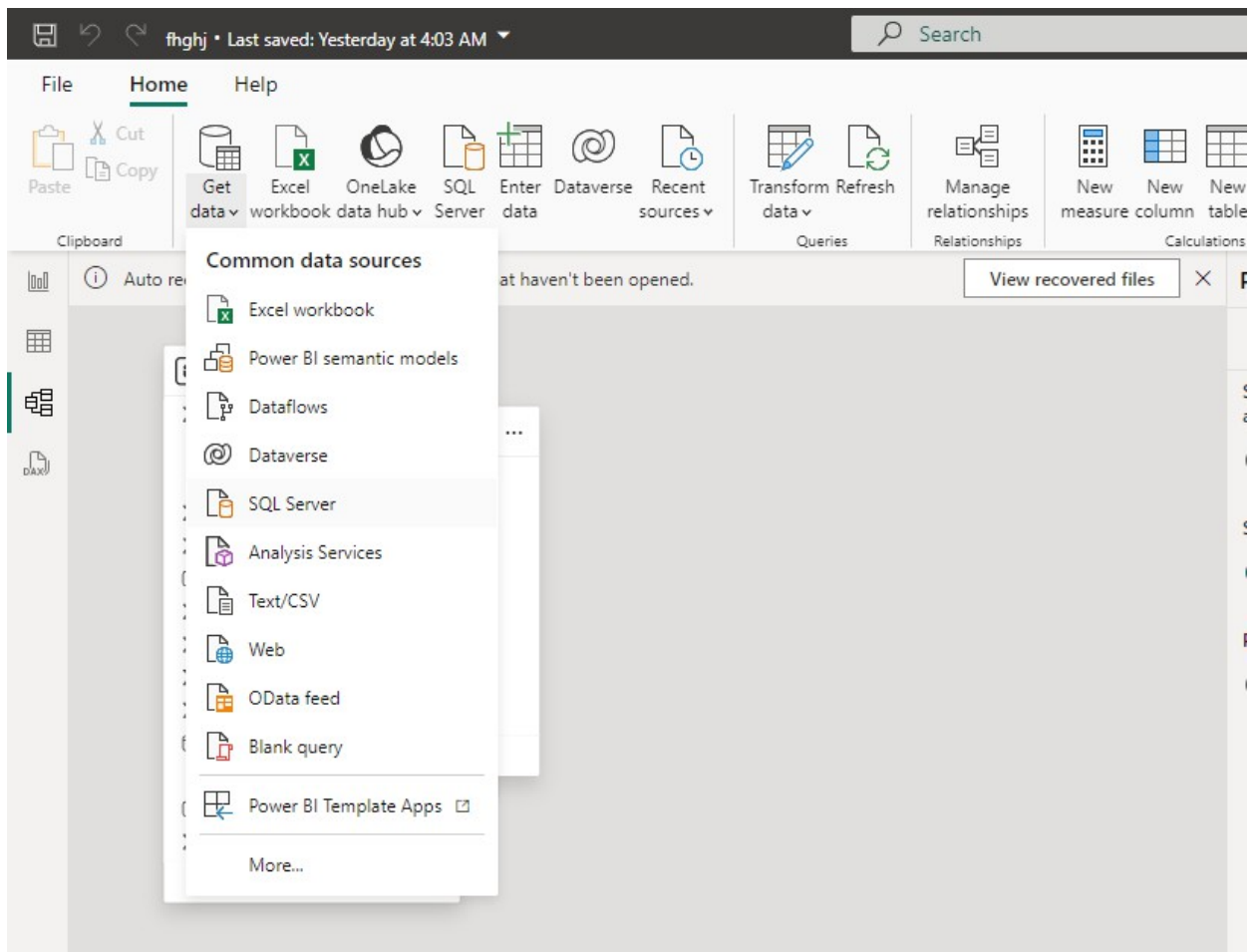
Composition:-

The dashboard effectively uses the chart types suited to their purpose like bar chart for comparing, line chart for trend visualization, stacked area chart for population breakdown. The prominent display of KPI “natural increase rate” focus on critical metrics, and all the graph adhere to best practice and theory.

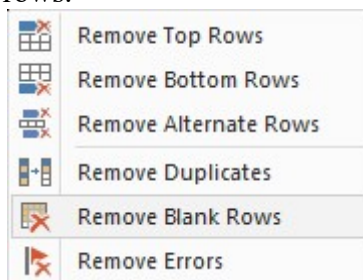
Provides a step-by-step overview of how the dashboard was built

Step by step overview:-

- Open the power Bi Desktop.
- In the Get data section of home ribbon, I select excel



- Then I navigate to the place where I saved the excel files.
- Then there is the option of load the data or you want to transform it. From there I click transform data.
- After viewing the data I see that there are some blank row in the country grouping files. So from the home tab there is an option to remove rows so from there I remove blank rows.



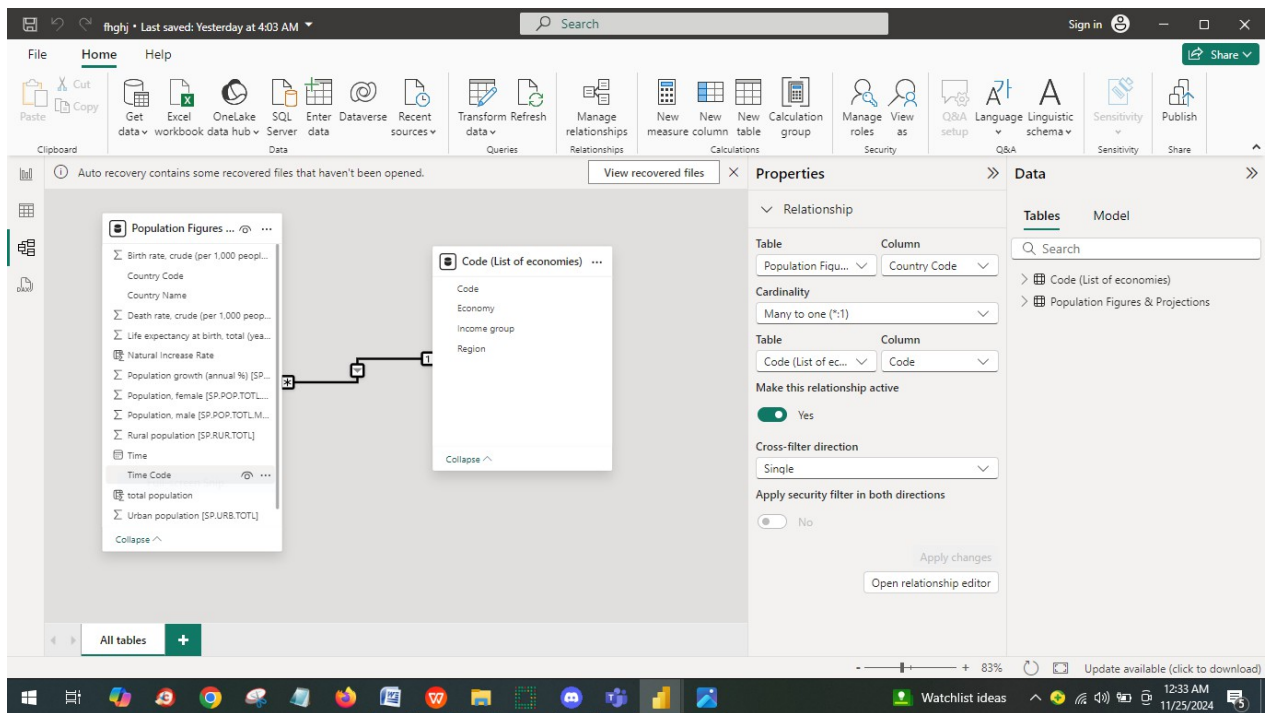
- Then I see that there are some columns with the name of null so I replaced them with the name of unclassified.

202	TKM	Turkmenistan	Europe & Central Asia	Upper middle income
203	TCA	Turks and Caicos Islands	Latin America & Caribbean	High income
204	TUV	Tuvalu	East Asia & Pacific	Upper middle income
205	UGA	Uganda	Sub-Saharan Africa	Low income
206	UKR	Ukraine	Europe & Central Asia	Upper middle income
207	ARE	United Arab Emirates	Middle East & North Africa	High income
208	GBR	United Kingdom	Europe & Central Asia	High income
209	USA	United States	North America	High income
210	URY	Uruguay	Latin America & Caribbean	High income
211	UZB	Uzbekistan	Europe & Central Asia	Lower middle income
212	VUT	Vanuatu	East Asia & Pacific	Lower middle income
213	VEN	Venezuela, RB	Latin America & Caribbean	unclassified
unclassified				

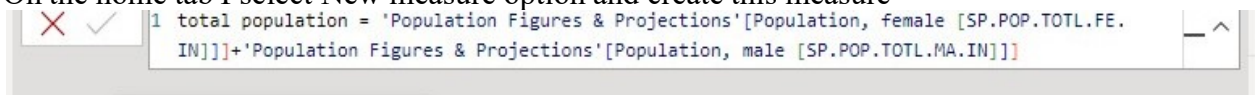
- Then I replace header name by the first column because that was not appropriate.
- Then in the main file there is a column of time. I changed its type from decimal to date type.

The screenshot displays the Microsoft Power BI Desktop interface. The main view shows a data table with the following columns: Country Name, Country Code, Time, Time Code, and Population, female [SP_POP_TOTL.FE.IN]. The 'Time' column is highlighted with a date icon, indicating its data type. The 'Query Settings' pane on the right shows the 'APPLIED STEPS' for the query, including 'Source', 'Promoted Headers', and 'Changed Type'. The 'Properties' pane on the right shows the query name 'Population Figures & Projections'.

- Then I load the data in power bi.
- In the model section I make the relationship between the time code of one file to the time code of another file. As you can see in this picture.

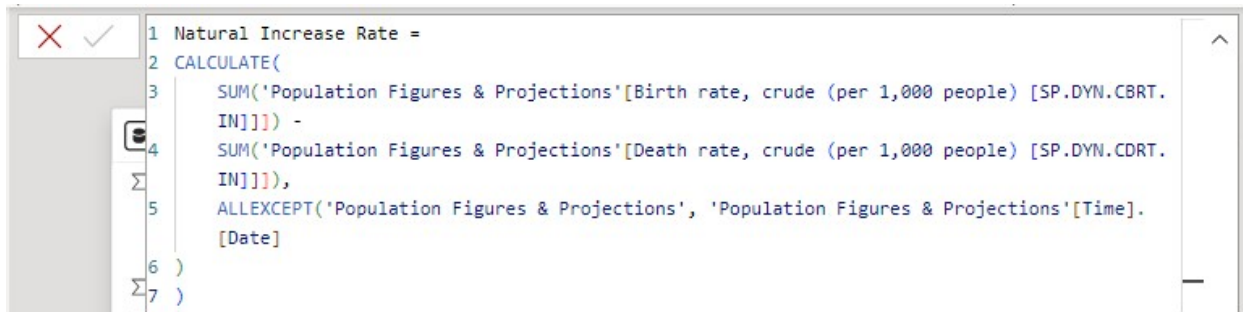


- On the home tab I select New measure option and create this measure



This measure is used to calculate the total population by summing up the female and male population.

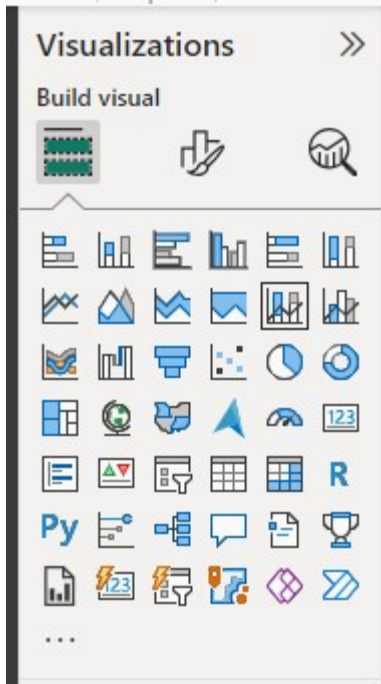
Then I create a new measure which is



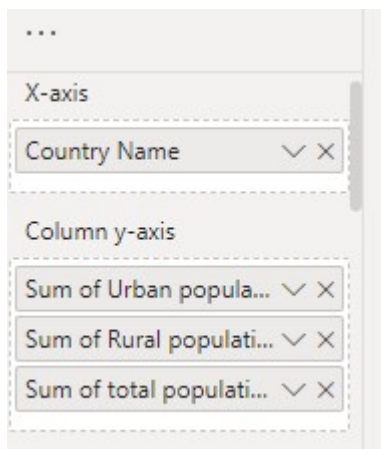
This measure is used to calculate the natural increase rate by subtracting the death rate from the birth rate.

Visualisation 1:-

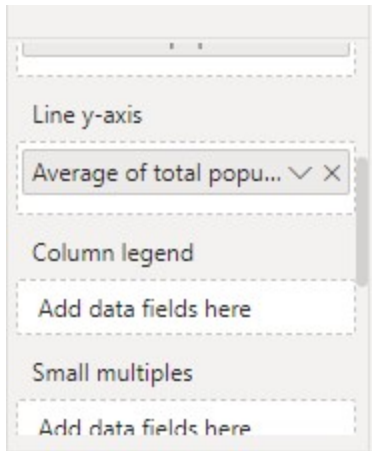
- From the field pane drag the line and stack column chart to a blank area on the canvas.



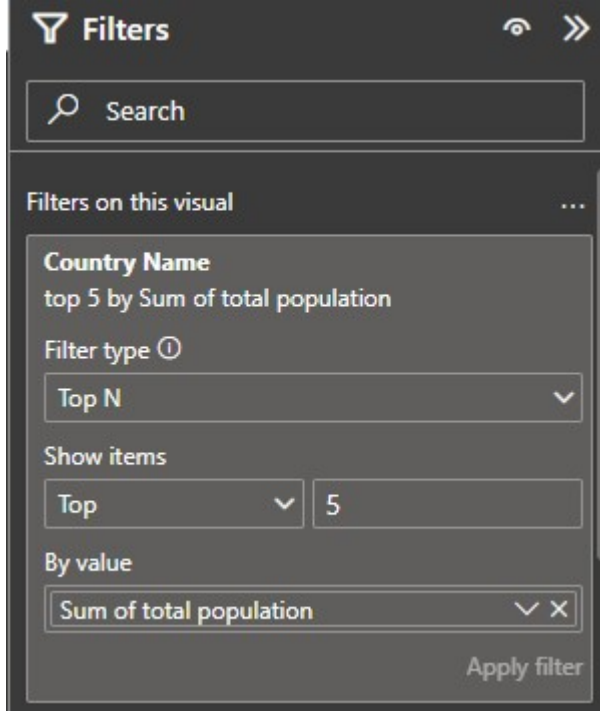
- Then I drag the Country names on the x-axis and for the y-axes I drag sum of urban population, sum of rural population and sum of total population.



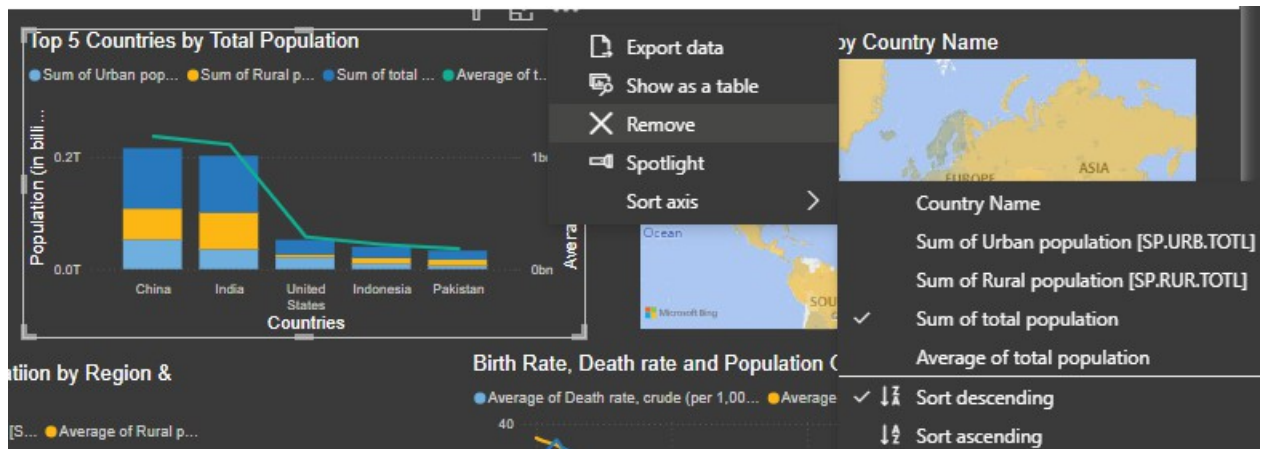
- Then drag the sum of total population column and place it on the line y-axis and then change sum to average for total population.



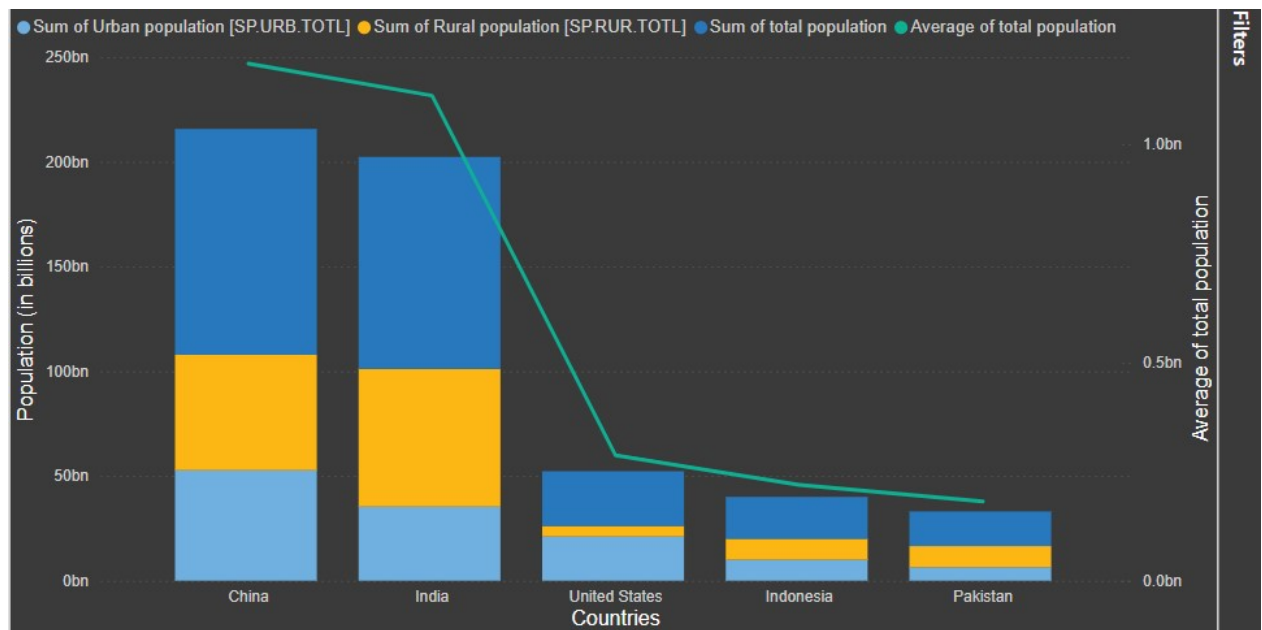
- Then I drag the Country Name field from the field pane and drop it into the top section of the Filter pane, and select Top N filter type and apply filter n=5 so that it will show only top 5 countries by the value sum of total population.



Then I select the ellipsis at the top right-hand corner of this visualisation and hover the cursor over Sort Axis. Then select sum of total population.

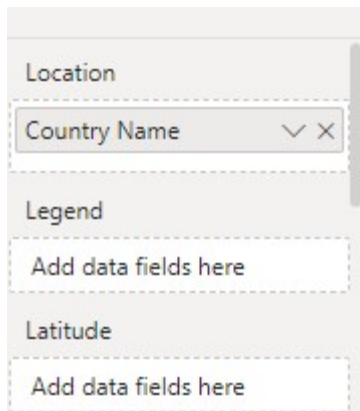


Then I rename the column names and my Overall visualization is

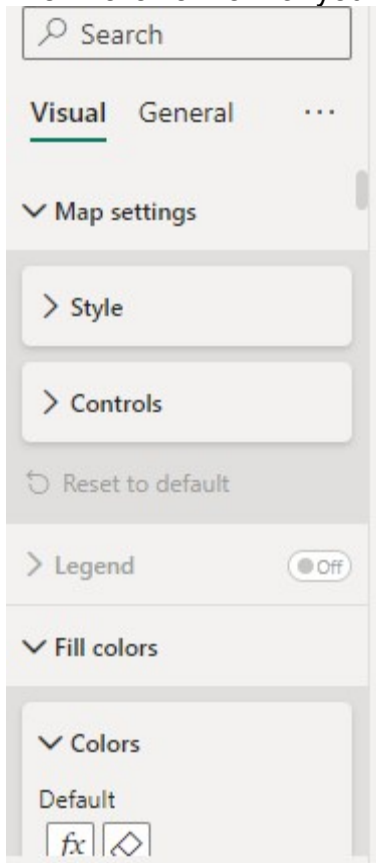


Visualization 2:-

- From the visualization pane drag the filled map and place it on the canvas.
- Add the country name in the location column.



- Then I click on format your visual and select color and select Fx.



- Then I select the gradient format style, select average of life expectancy at birth column as field on which this base and select donot format empty values and change the color of maximum value to purple and also add middle value.

Default color - Fill colors - Colors ✕

Format style
 Gradient ▼

What field should we base this on? Average of Life expectancy at birth, total ▼

Summarization Average ▼


How should we format empty values? Don't format ▼

Minimum Lowest value ▼ ■ ▼
 Enter a value

Center Middle value ▼ ■ ▼
 Enter a value

Maximum Highest value ▼ ■ ▼
 Enter a value

☒ Add a middle color



[Learn more about conditional formatting](#)

OK Cancel

- Then I drag the average of life expectancy column to the tooltip.

Longitude

Add data fields here

Tooltips

Average of Life expect... ▼ ✕

Drill through

Cross-report ● Off

- Name the visualization and the overall visualization becomes



Visualization 3:-

Drag the stacked area chart to the canvas

Then I add Time(Year) on x-axes and average of urban population, average of rural population on y-axes.

Time

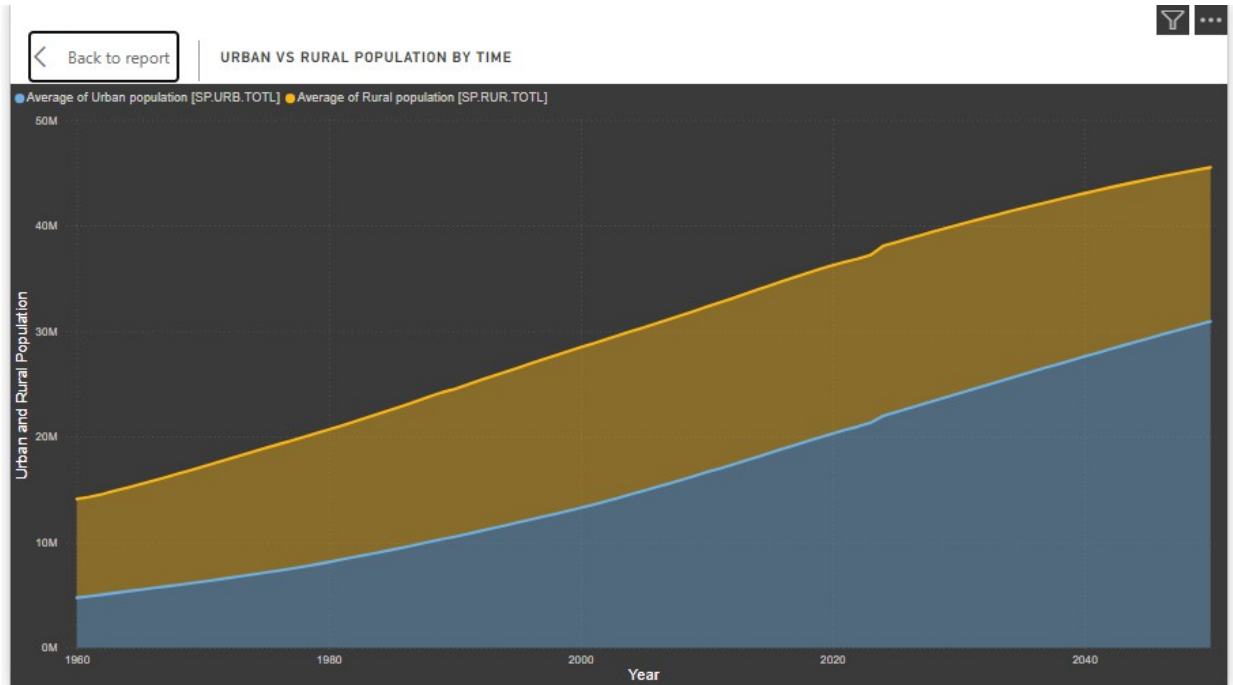
Year

Y-axis

Average of Urban po...

Average of Rural pop...

Now rename the title and column names



Visualization 3:-

Drag the card from visualization field to the canvas and click on natural increase rate from the data panel and rename the title.



Visualization 4:-

I drag the multiple line chart from visualization field to the canvas and apply time(years) on the x-axes and for the y-axes I add average of death rate and average of birth rate. Then I add average of annual population growth on the secondary y-axes.

X-axis

Time

Year

Y-axis

Average of Death rate...

Average of Birth rate, ...

Secondary y-axis

Average of Populatio...

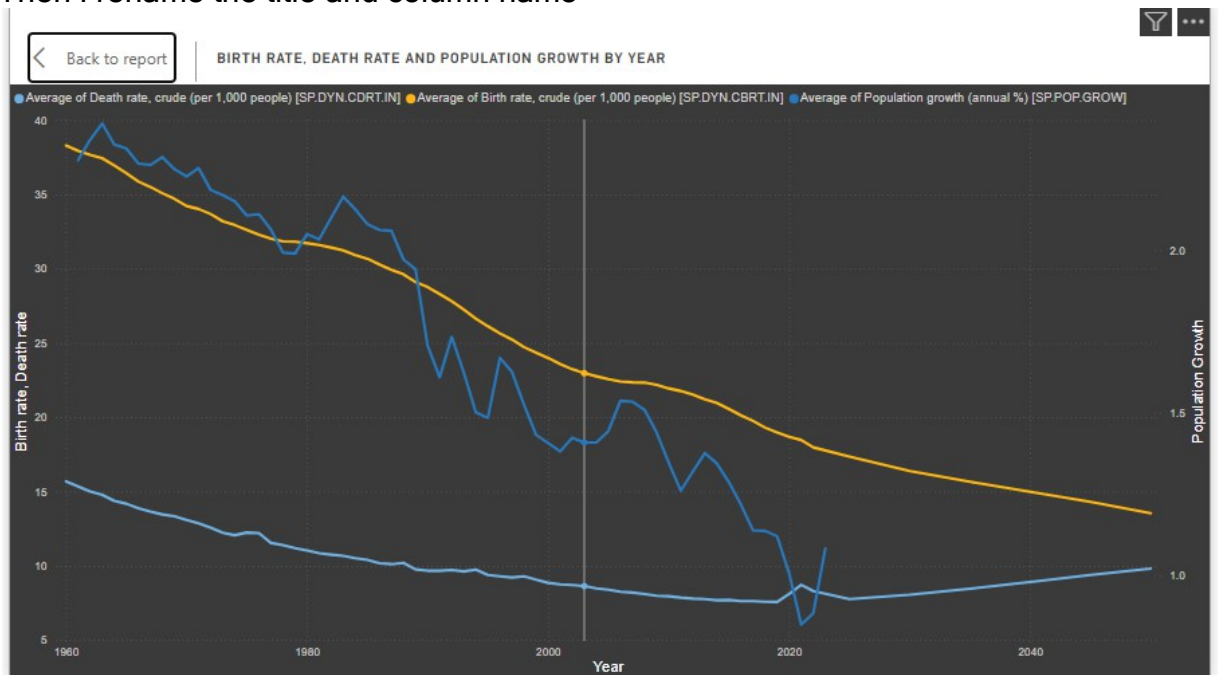
Legend

Add data fields here

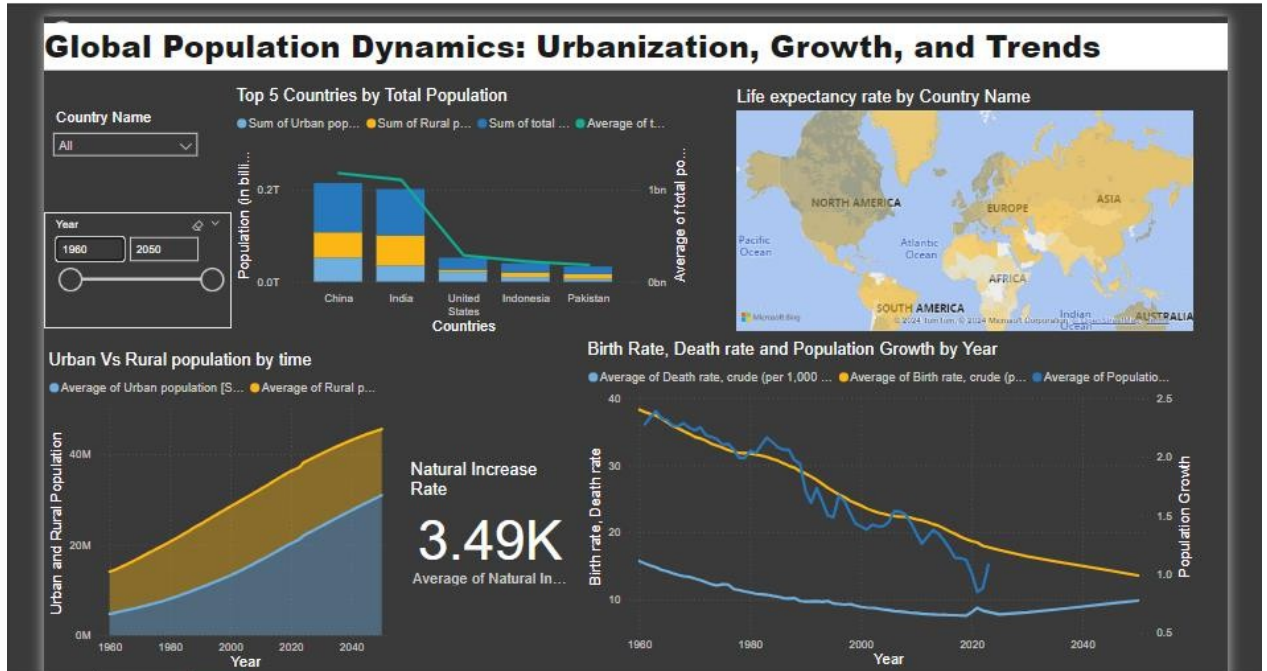
Small multiples

Add data fields here

Then I rename the title and column name



Overall Dashboard



Critical Evaluation of the Dashboard:-

Strenghts:-

- It represents population of top 5 countries through line and stacked column chart.
- It represent different trends of birth rate death rate and population annual growth through a line chart.
- It represent distribution of urban and rural population through stacked area chart.
- It represents life expectancy with respect to each geographic location through a filled map chart
- Filters for country and Year enable users for more specific information.
- Prominent display of natural increase rate as KPI in a card.

Limitations:-

- It does not have drill down option.
- The color patterns in this dashboard may not be fully accessible to color blind users.

