# Data Analysis of Life Expectancy

Umar Shehzad (B1514261)

Big Data and Business Intelligence (CIS4008-N)

12/29/22

# Table of Contents


Section 1: Business Intelligence Report ..... 2

Executive Summary ..... 2

INTRODUCTION ..... 2

Objectives ..... 2

Data Source ..... 3

Dataset description ..... **Error! Bookmark not defined.**

Business Intelligence Question ..... 7

Finding based on analysis and evaluation ..... 7

Summary ..... 23

Recommendation ..... 24

Conclusion ..... 24

Section 2: Business Intelligence Design ..... 25

Data Pre-Processing or Data Cleansing ..... 25

Data Loading ..... 25

File origin ..... 28

Delimiter ..... 28

Data Cleaning ..... 30

BI Data Modelling ..... 33

Normalization of Data Set ..... 34

Query Merging ..... 37

Fact Table ..... 39

Dimension Tables ..... 40

DAX and M Language ..... 42

Dashboard ..... 44

References ..... 54

# Section 1: Business Intelligence Report

## Executive Summary

This report examines what are the average lifespan of humans being and what are the main factors that influence life expectancy across the globe by examining and analysing data from different countries and region across the planet, the data is from 1800-2020. In these 220 years life expectancy changes from 32 to 73. Life expectancy varies continent to continent, many factors are responsible for this positive change like focus of governments in health sector, access of public to smoking products, does dense population have any impact on human lifespan and alcoholics beverages, increase in income per capita of population. The sole purpose of this report is to examine what is the life expectancy for both gender during this timeline which we have mentioned and what are the important factor that are responsible for drastic change in life expectancy in different region of world and lastly what steps should be taken to increase life expectancy further.

## INTRODUCTION

According to many studies and researchers the typical lifespan of male is around 70 and female is roughly 75 years, which is huge variance in the domain of human life expectancy in terms of gender. Furthermore, according to many experts and researchers in this field speaks that there are numerous factors that are affecting human life expectancy, but the major factors are geography, demographic, wealth distribution, mortality rates, vaccinations, health related factors and human development.

Our analysis would be based on the parameters which are mentioned in the above paragraph to determine major questions regarding human lifespan across the globe.

### Objectives

The main objective of this business intelligence project is to provide an analytical and visual picture of how life expectancy across the globe is and what factors that are contributing between countries with high life expectancy and low life expectancy via the above parameter which we have mentioned that are playing a major part in influencing human life expectancy across the globe.

This project will help global organisation like NGO and WHO to get a pretty simple pictures of life expectancy across the globe, by helping and informing people what are the life expectancy factor drivers.

## Data Source

The has data been collect by Primary from two places which are World Bank Data Lakes which is kind of database the other source has also enormous amount of data and the data is reliable which called our world in data. Links are mentioned as follow.

https://databank.worldbank.org/reports.aspx?source=2&series=SP.POP.TOTL.FE.IN&country=#

**https://ourworldindata.org/life-expectancy**

I have downloaded all the data and will upload it with ICA in separate dataset file.

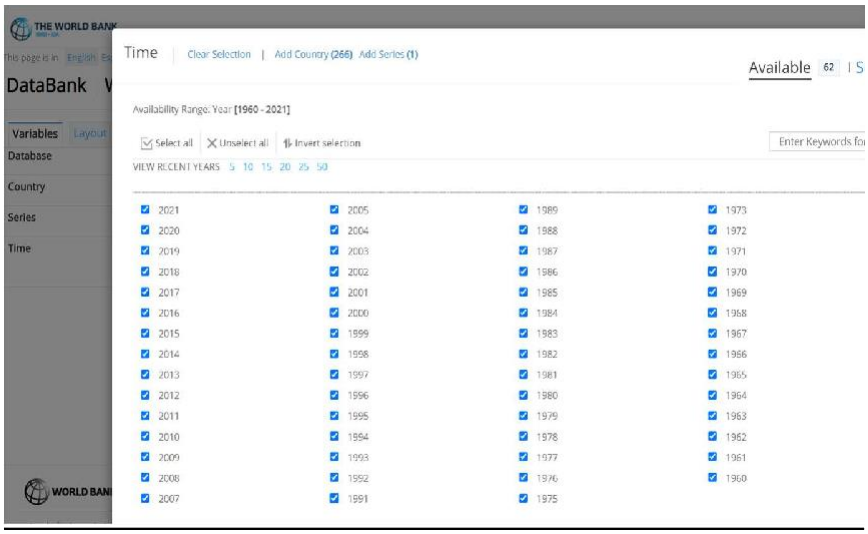



*Figure 1 Year Selection*

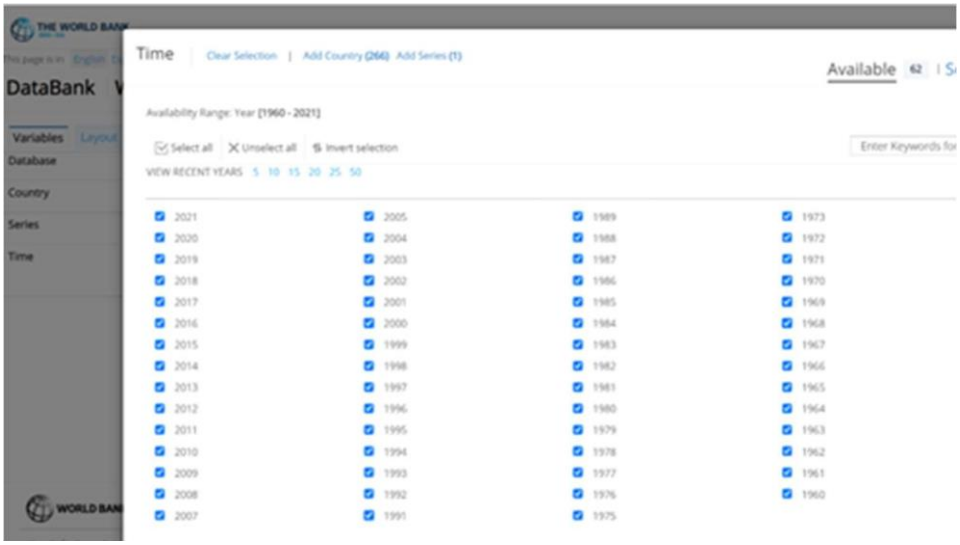



*Figure 2 Year selection*

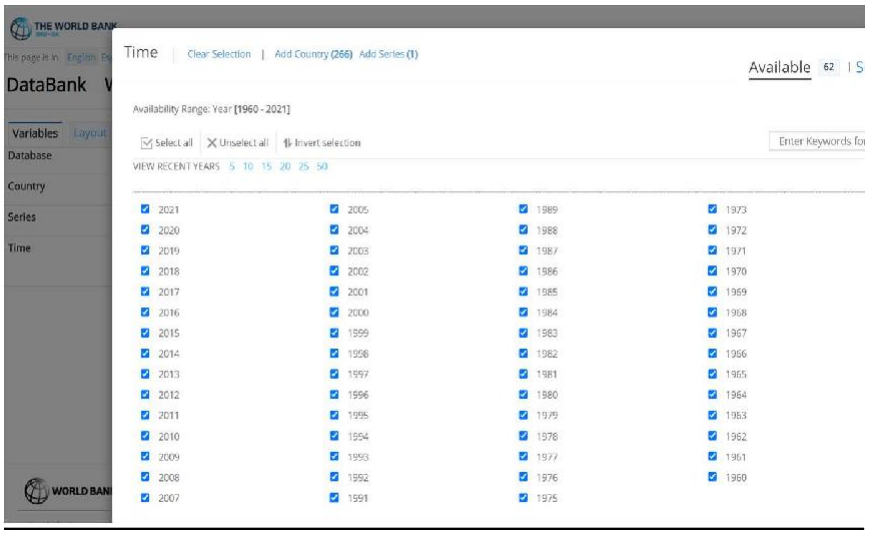
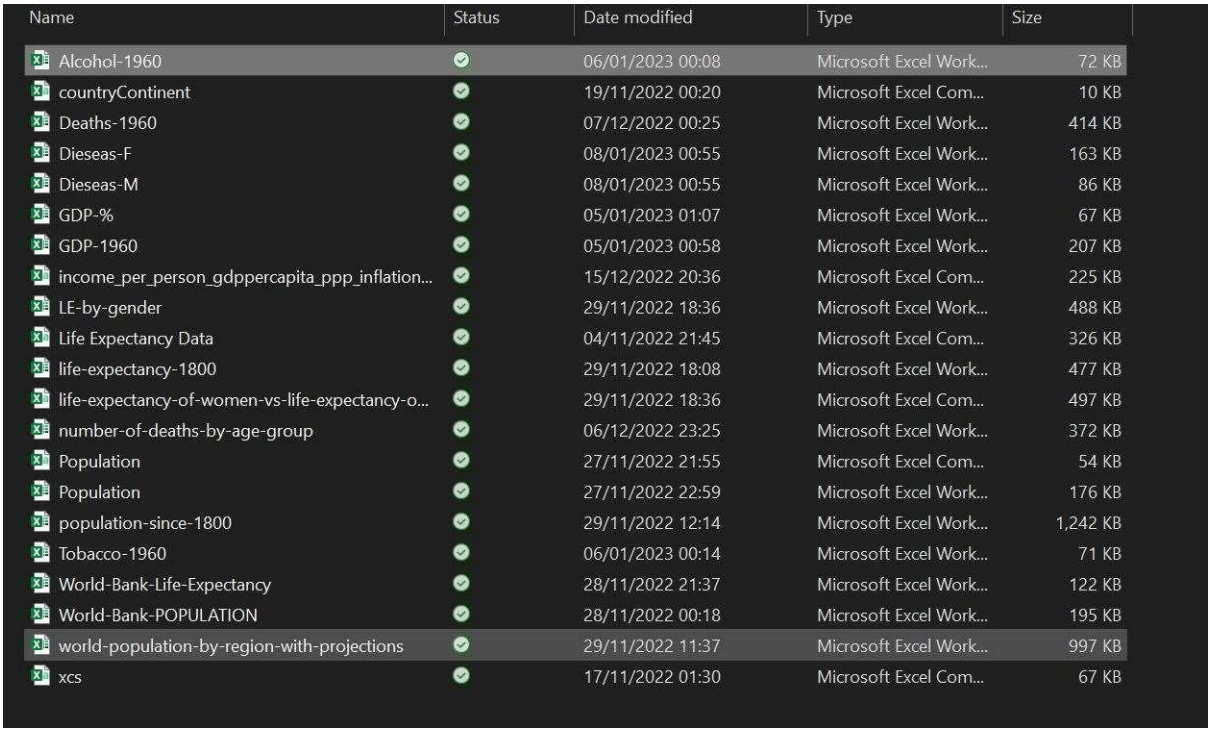
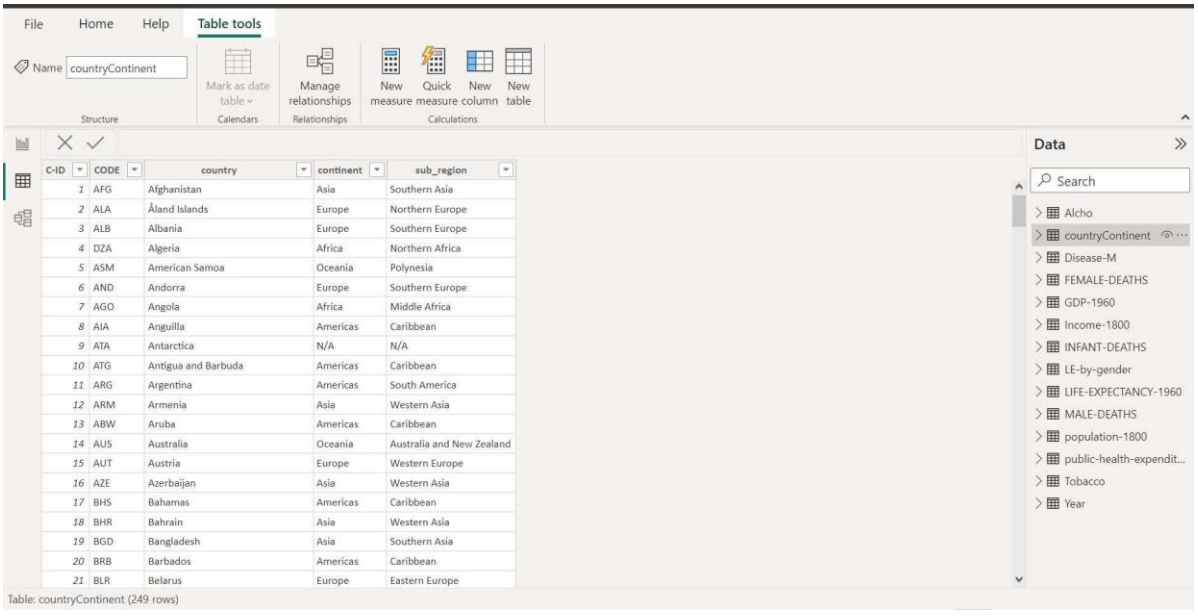



*Figure 3 Year Selection*



| Name | Status | Date modified | Type | Size |
|---|---|---|---|---|
| Alcohol-1960 | ✓ | 06/01/2023 00:08 | Microsoft Excel Work... | 72 KB |
| countryContinent | ✓ | 19/11/2022 00:20 | Microsoft Excel Com... | 10 KB |
| Deaths-1960 | ✓ | 07/12/2022 00:25 | Microsoft Excel Work... | 414 KB |
| Dieseas-F | ✓ | 08/01/2023 00:55 | Microsoft Excel Work... | 163 KB |
| Dieseas-M | ✓ | 08/01/2023 00:55 | Microsoft Excel Work... | 86 KB |
| GDP-% | ✓ | 05/01/2023 01:07 | Microsoft Excel Work... | 67 KB |
| GDP-1960 | ✓ | 05/01/2023 00:58 | Microsoft Excel Work... | 207 KB |
| income_per_person_gdppercapita_ppp_inflation... | ✓ | 15/12/2022 20:36 | Microsoft Excel Com... | 225 KB |
| LE-by-gender | ✓ | 29/11/2022 18:36 | Microsoft Excel Work... | 488 KB |
| Life Expectancy Data | ✓ | 04/11/2022 21:45 | Microsoft Excel Com... | 326 KB |
| life-expectancy-1800 | ✓ | 29/11/2022 18:08 | Microsoft Excel Work... | 477 KB |
| life-expectancy-of-women-vs-life-expectancy-o... | ✓ | 29/11/2022 18:36 | Microsoft Excel Com... | 497 KB |
| number-of-deaths-by-age-group | ✓ | 06/12/2022 23:25 | Microsoft Excel Work... | 372 KB |
| Population | ✓ | 27/11/2022 21:55 | Microsoft Excel Com... | 54 KB |
| Population | ✓ | 27/11/2022 22:59 | Microsoft Excel Work... | 176 KB |
| population-since-1800 | ✓ | 29/11/2022 12:14 | Microsoft Excel Work... | 1,242 KB |
| Tobacco-1960 | ✓ | 06/01/2023 00:14 | Microsoft Excel Work... | 71 KB |
| World-Bank-Life-Expectancy | ✓ | 28/11/2022 21:37 | Microsoft Excel Work... | 122 KB |
| World-Bank-POPULATION | ✓ | 28/11/2022 00:18 | Microsoft Excel Work... | 195 KB |
| world-population-by-region-with-projections | ✓ | 29/11/2022 11:37 | Microsoft Excel Work... | 997 KB |
| xcs | ✓ | 17/11/2022 01:30 | Microsoft Excel Com... | 67 KB |

The data has enourmas amount of rows and data so take screenshot of even will increase the length file as well as weste a lot of time ill upload the data as well with ICA and share a screenshot in power bi as well.




| C-ID | CODE | country | continent | sub_region |
|---|---|---|---|---|
| 1 | AFG | Afghanistan | Asia | Southern Asia |
| 2 | ALA | Åland Islands | Europe | Northern Europe |
| 3 | ALB | Albania | Europe | Southern Europe |
| 4 | DZA | Algeria | Africa | Northern Africa |
| 5 | ASM | American Samoa | Oceania | Polynesia |
| 6 | AND | Andorra | Europe | Southern Europe |
| 7 | AGO | Angola | Africa | Middle Africa |
| 8 | AIA | Anguilla | Americas | Caribbean |
| 9 | ATA | Antarctica | N/A | N/A |
| 10 | ATG | Antigua and Barbuda | Americas | Caribbean |
| 11 | ARG | Argentina | Americas | South America |
| 12 | ARM | Armenia | Asia | Western Asia |
| 13 | ABW | Aruba | Americas | Caribbean |
| 14 | AUS | Australia | Oceania | Australia and New Zealand |
| 15 | AUT | Austria | Europe | Western Europe |
| 16 | AZE | Azerbaijan | Asia | Western Asia |
| 17 | BHS | Bahamas | Americas | Caribbean |
| 18 | BHR | Bahrain | Asia | Western Asia |
| 19 | BGD | Bangladesh | Asia | Southern Asia |
| 20 | BRB | Barbados | Americas | Caribbean |
| 21 | BLR | Belarus | Europe | Eastern Europe |

On the right side of the above screenshot we can see the tables.

## Dataset description

All xslx downloaded from websites contains only 20 table some of the table and column description is bellow columns.

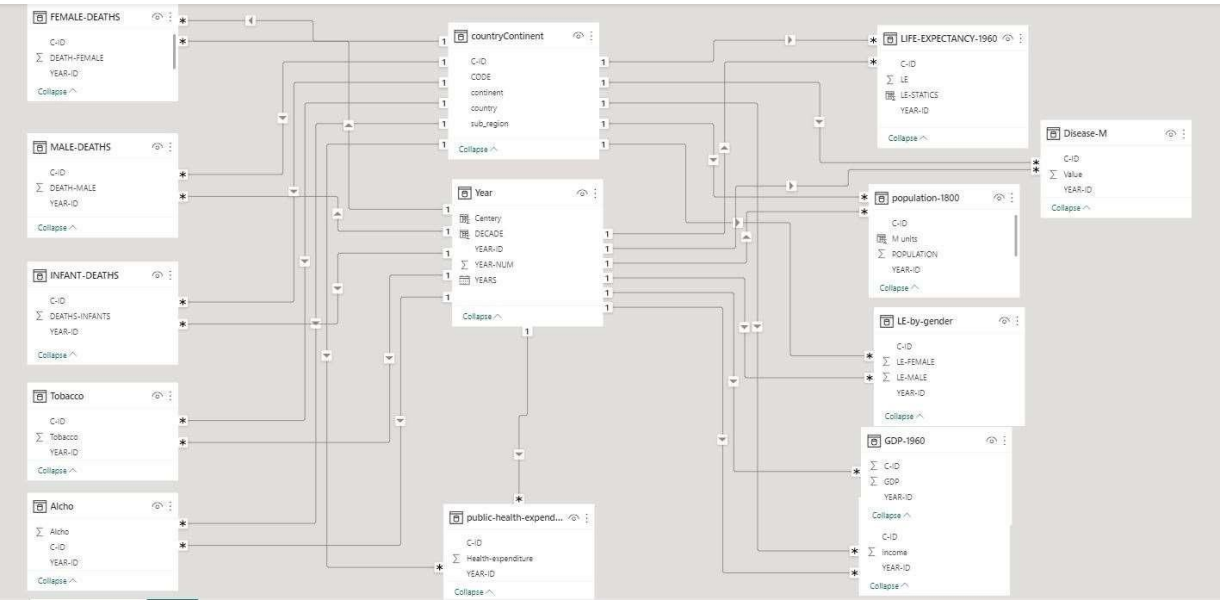
**countryContinent table**

| | **Column Name** | **Description** |
|---|---|---|
| 1 | C-ID | Id which common in every table |
| 2 | continent | Continent names |
| 3 | Country | Country name |
| 4 | Sub/ Region | Sub areas |

**Year table**

| | **Column Name** | **Description** |
|---|---|---|
| 1 | century | Convert years into century by using dax starting from 1800 |
| 2 | decade | Convert into decade |
| 3 | Year ID | Common in every table |
| 4 | years | Start from 0 adding the year from 1800 |
| 5 | Year_nun | Is the actual year that is record |

In the bellow screenshot we can see all the names of the tables and column used in this project the important fact table column I have mention above in the tables.




## Business Intelligence Question

The key question which we will be addressing in the analysis are mentioned below:

1. In which years did life expectancy start raising up?
2. What are the main factor that can increase human lifespan?
3. What was the effect of income per capta and GDP on life expectancy?
4. Can a country's healthcare spending have significant impact on life expectancy?
5. Is there a relationship between eating habits, lifestyle, exercise, smoking, and alcohol consumption that is either favourable or negative for human life expectancy?
6. How are life expectancies impacted by mortality rates?
7. Do countries with dense populations typically have shorter life expectancies?

# Finding based on analysis and evaluation

Different visualisations were created to answer the earlier stated data analysis questions and are detailed below:

The following are the findings of the business intelligence report's analysis of the above-mentioned business questions.

**Question 1: In which years did life expectancy start raising up?**

The line graph shows a life expectancy timeline, the three-line shown on this graph are, the main orange line shows the life expectancy trend in total across the years, while the pink line shows the female life expectancy and the last one which is light blue shows the male life expectancy. The x-axis in this picture shows years and y-axis shows the life expectancy.

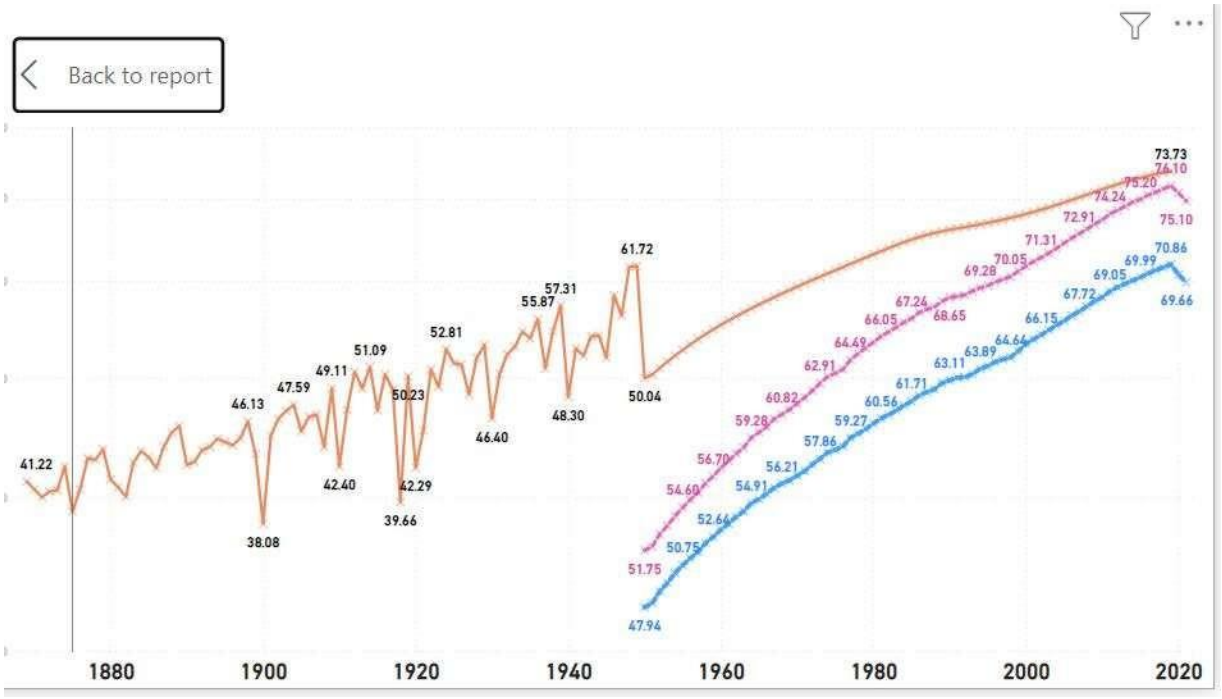



*Figure 4: Life expectancy line graph timeline*

In 1940, life expectancy decreased as a result of the World War 2 impact on the civilian population. For instance, a large number of soldiers were conscripted abroad. After the war, improvements in infant mortality and childhood immunisation contributed to the ongoing rise in life expectancy as well as GDP per capita.

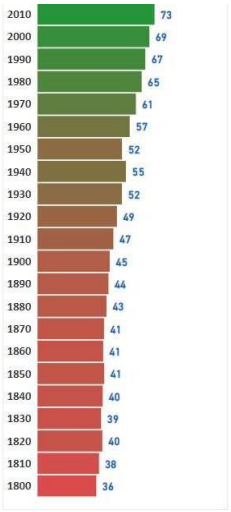



*Figure 5: horizontal bar life expectancy timeline by decade*

In Figure 5 shows that how life expectancy boomed after 1960' due to economical growth and opportunity also advancement in medicine and medical science and access to better healthcare. In the 19th century life expectancy was at its worst because of the reason mentioned in the start of the paragraph.

Life expectancy in the 1800s was generally lower than it is today due to a number of factors such as lack of medical knowledge and technology which resulted in diseases and injuries that were not treatable, poor sanitation and lack of clean water and proper sewage disposal which led to spread of infectious diseases, lack of nutrition due to poverty which made people vulnerable to illness and infection, high infant mortality due to lack of proper medical care and understanding of child health, wars and violent conflicts which led to significant loss of life, also climate changes such as famines and food scarcity. All these factors combined contributed to lower life expectancies in the 1800s compared to today.

**Question 2: What is the main factor that can increase human lifespan?**

Income is one of the main factor which life expectancy depends. If you have high income means you have high living standards access of food, medicine etc. the income is directly effect on the human life because if the person has not enough income to buy their food, medicine and their expenditures their life expectancy not raises. The government will give as much relief as it wants, until the person's income is increased, he will not get any benefit from the relief. Income is basic need of person to buy anything.

In this figure 1800 to 1900 we clear see that the average income of person is not too much therefore the result is African region has low life expectancy and other hand the countries which having high income these countries having high life expectancy.


*Figure 6 Countries with highest and lowest income*

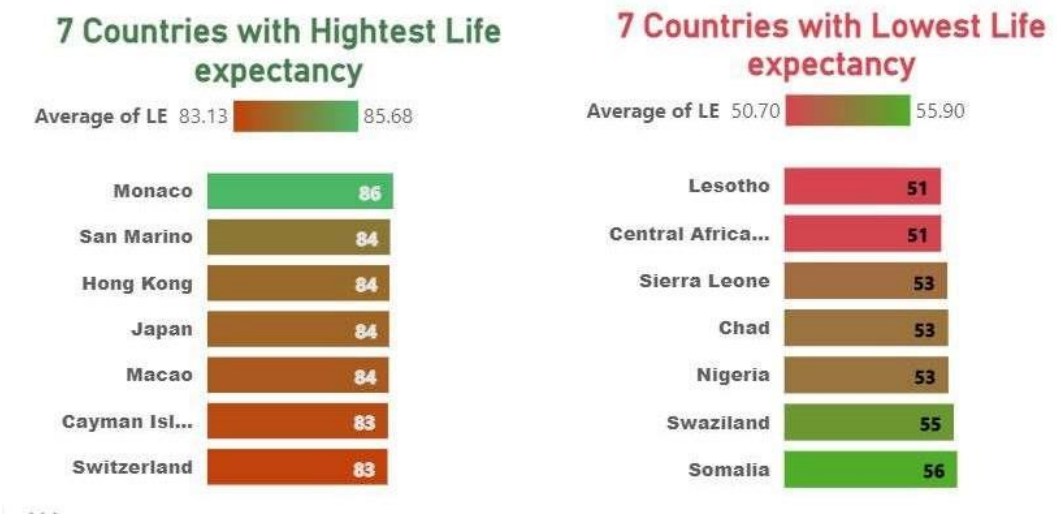



*Figure 7: countries with highest life expectancy*

In figure 6 and 7 , we clear see that the average income of person is not too much therefore the result is African region has low life expectancy and other hand the countries which having high income these countries having high life expectancy, so technically the more wealthy a country is the higher will be the life expectancy of it citizens we have applied tooltip on the graph so will give you complete detail one mouse arrow is hovered on the graph and we can adject the timeline as well.

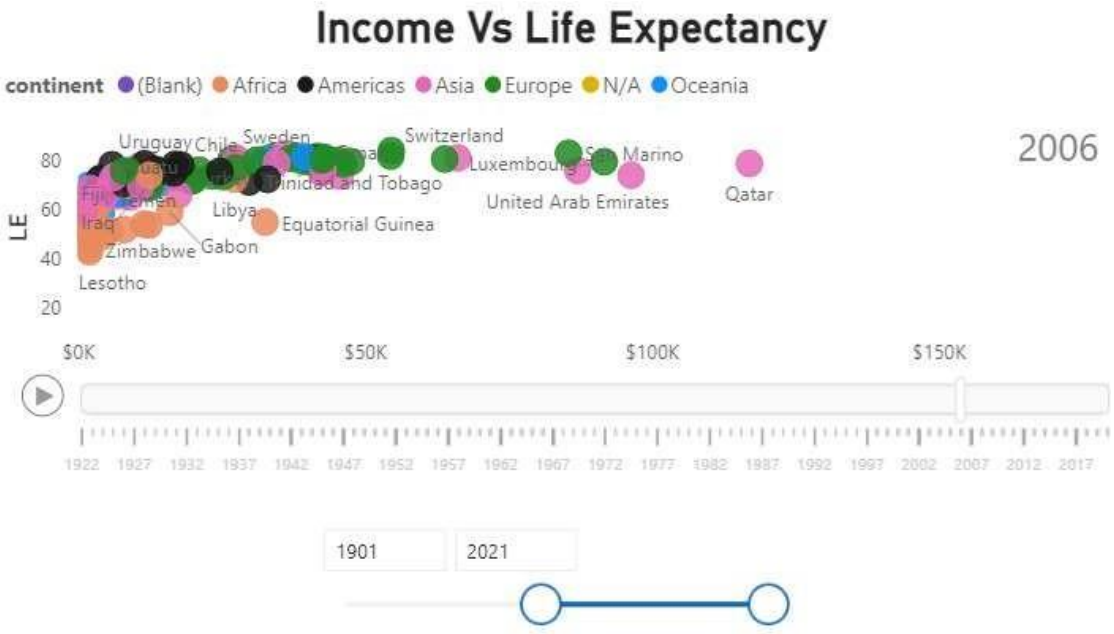



*Figure: 8 income vs life expectancy*

In figure 8, clear show that Asia Region and Europe region are high life expectancy as compared to other regions due to high income and after discovery of petroleum things in that region.

In the end all the top 5 counties which have low life expectancy are from African region. In Arab region the ratio of life expectancy is high because stable income.

The reason is Wealthy countries tend to have the highest life expectancies due to factors such as access to comprehensive and high-quality healthcare, better nutrition, education which leads to more informed choices about lifestyle, proper sanitation and clean water to reduce the spread of infectious diseases, lower infant mortality rate, low poverty rate and advanced technology in healthcare. These factors combined contribute to higher life expectancies in wealthy countries compared to less developed countries.

**Question 3: What was the effect of income per capita and GDP on life expectancy?**

Income per capita and GDP has major effect on life expectancy. The below figure 9 demonstrate the answer this question.

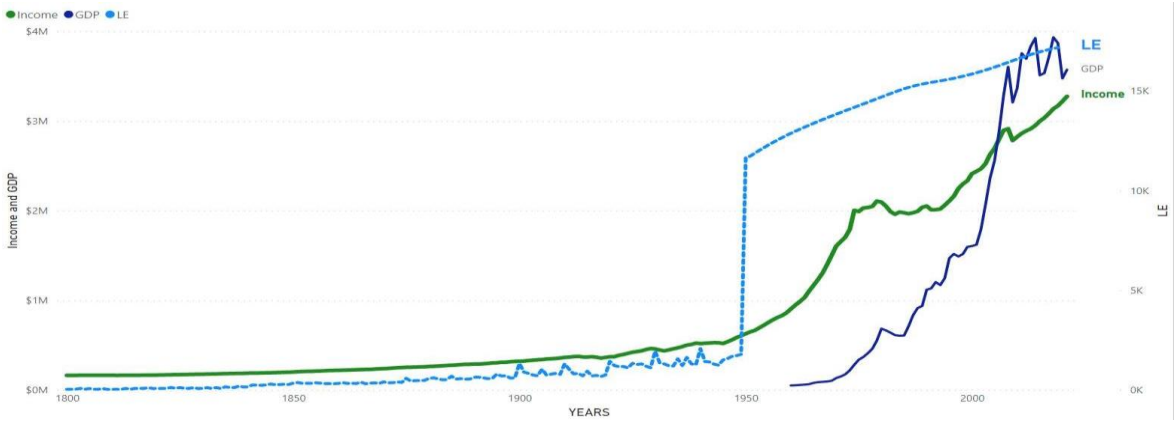



*Figure 9: GDP VS income VS Life expectancy*

The main blue line shows the life expectancy trend in total across the years, while the green line shows the income and the last one which is purple shows the GDP. The x-axis in this picture shows years and y-axis shows the Income and GDP. The life expectancy starts increasing after 1950 till this day now.

Gross Domestic Product (GDP) is a measure of a country's economic output and generally corresponds to the amount of goods and services produced in a country in a given time period. Income is the amount of money a person or household earns. Life expectancy is the average number of years a person is expected to live.

In general, countries with higher GDPs tend to have higher average incomes and longer life expectancies. This is because a strong economy allows for greater investment in healthcare, education, and social programs, which in turn improves access to healthcare, nutrition, and overall living conditions. Additionally, higher GDPs often mean that more people have access to higher paying jobs and better living conditions, leading to better overall health outcomes.


*Figure 10 income vs life expectancy*

In figure 10, on the x-axis shows the income with respect to years and y-axis shows the life expectancy. The bubbles show the different countries according to their continent like orange colour shows Africa region, black show Americas region, pink show Asia region, green show Europe region and blue show Oceania region. In this figure 10 we can clearly see that those countries having high income have a higher life expectancy their vice versa.

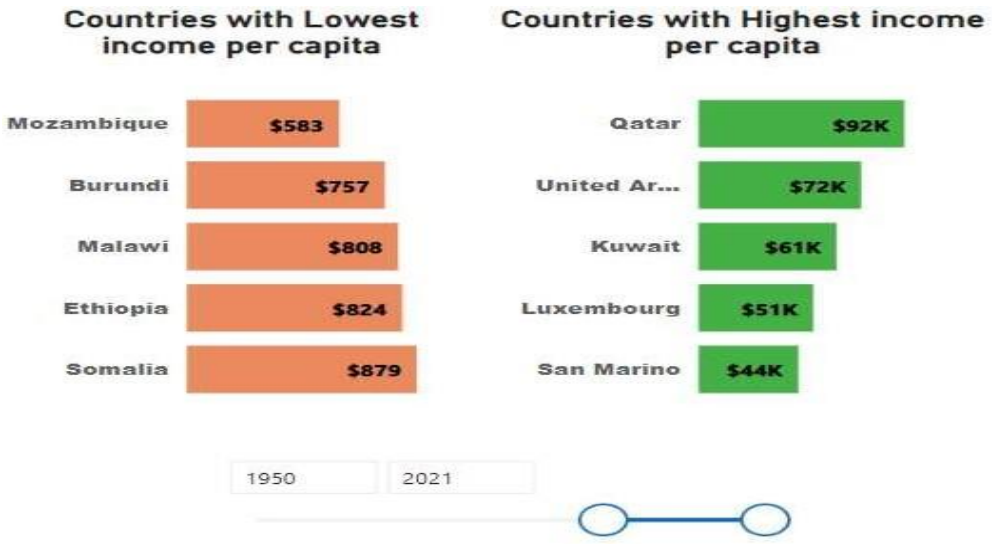



*Figure 11 Highest and lowest per capita countries*

Figure 11 shows the list of countries which have low income per capita. In figure 11 we can clearly see that Africa region countries are top 5 of list in low life expectancy from 1950s to this day. In this right side of figure 11, shows the list of countries which have highest income per capita. In which we can clearly see that among the top 5 countries 3 are Arab oil rich countries and 2 countries are from other region which have high life expectancy from 1950 to till now because of higher income per capita.

## Question 4: Can a country's healthcare spending have significant impact on life expectancy?

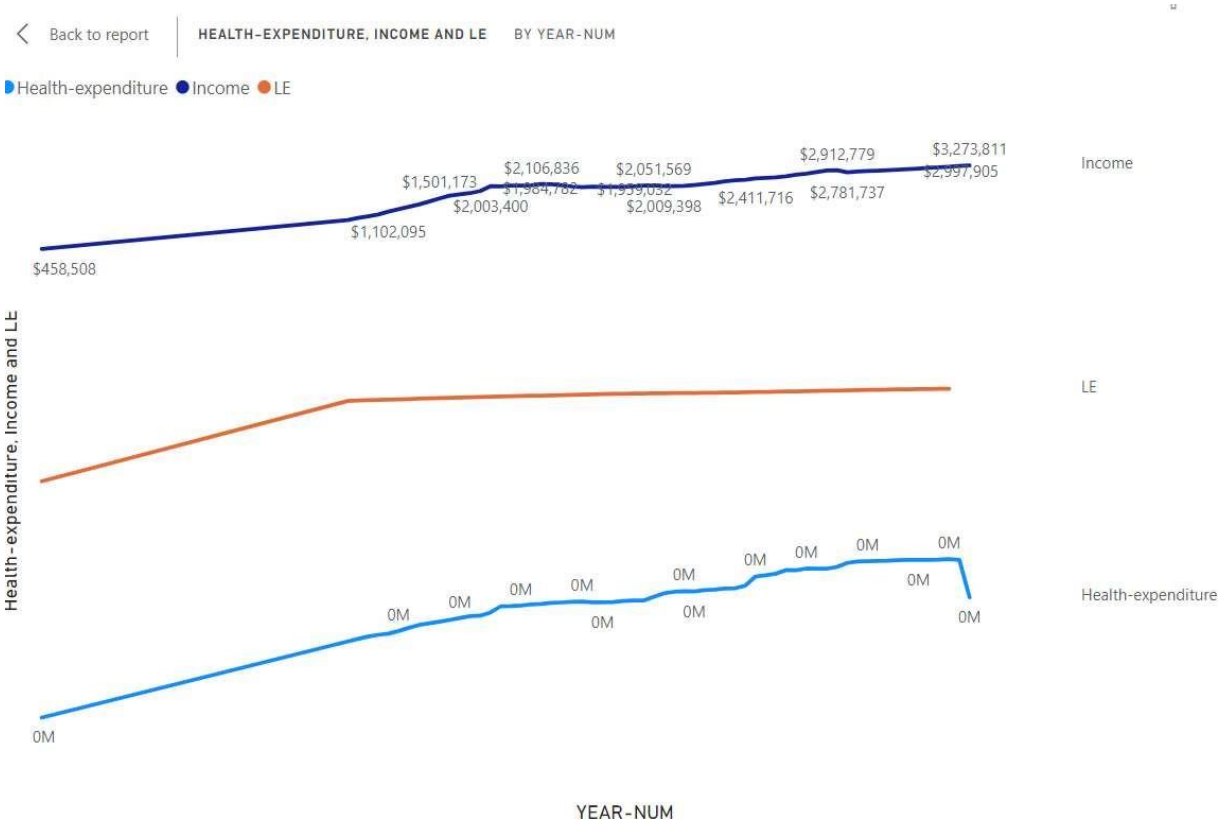



*Figure 12 health expenditure vs income vs life expectancy*

Here in figure 12, we can see the positive correlation between Government healthcare expenditure can play a role in increasing life expectancy in a country, with higher levels of government healthcare expenditure tend to have higher life expectancies. This is because government healthcare expenditure can provide more people with access to healthcare, which can lead to better health outcomes.

## Question 5: Is there a relationship between tobacco and alcohol consumption that is either favourable or negative for human life expectancy?

The answer is Yes, both Alcohol and tobacco consumption have different effect on human life expectancy. According to search unsure drinkers, no more than one drink a day, had benefits and could gain nearly 1 year in life expectancy, in contrast to a loss of nearly 7 years if drinking more than that.

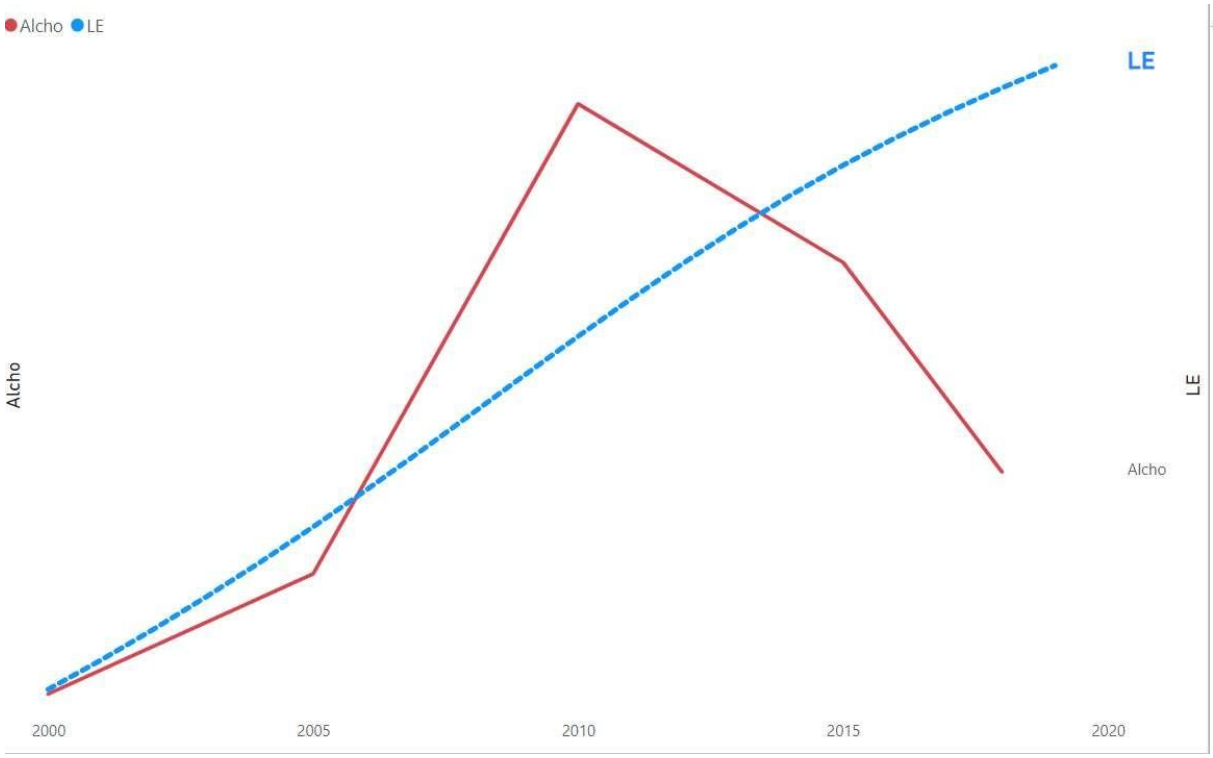



*Figure 13 Alcohol Vs Life Expectancy*

In figure 13, we can see two lines have been shown on this graph: the main blue line shows the life expectancy trend in total across the years, while the red line shows the alcohol. The x-axis in this picture shows years and y-axis shows the alcohol. The life expectancy starts increasing when use limit amount of alcohol or take as medical cure. The major take of alcohol is in non-Muslim countries. This graph show the Alcohol take down after 2010 when people know about the disadvantage of access amount.

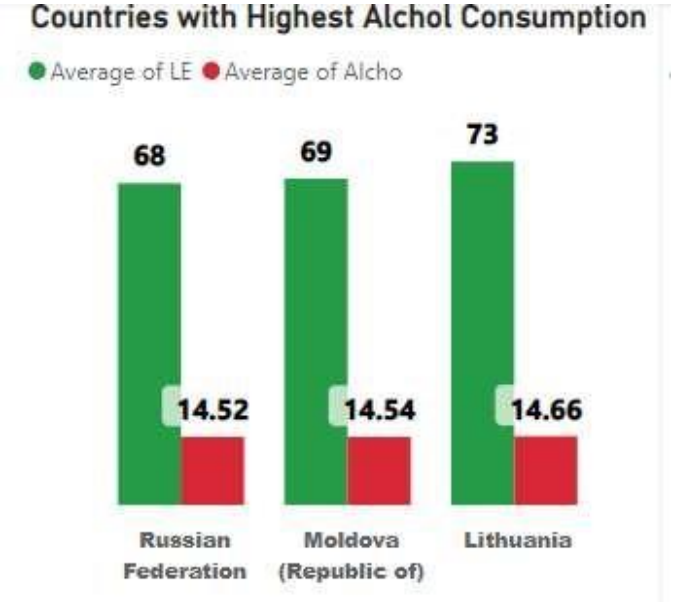



*Figure 14 Countries with hight consumption of alcohol*

In this figure 14 green bar show the average of life expectancy and red show average of alcohol. From this figure clear show that the top 3 countries which have high alcohol consumption are non-Muslim countries.

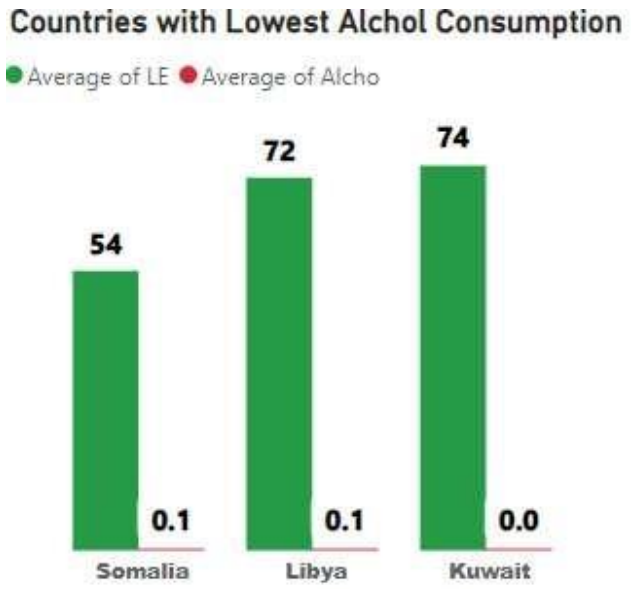



*Figure 15 countries with lowest alcohol consumption*

Figure 15 illustrate that green bar show the average of life expectancy and red show average of alcohol. From this figure clear show that the top 3 countries which have lowest alcohol consumption are Muslim countries.

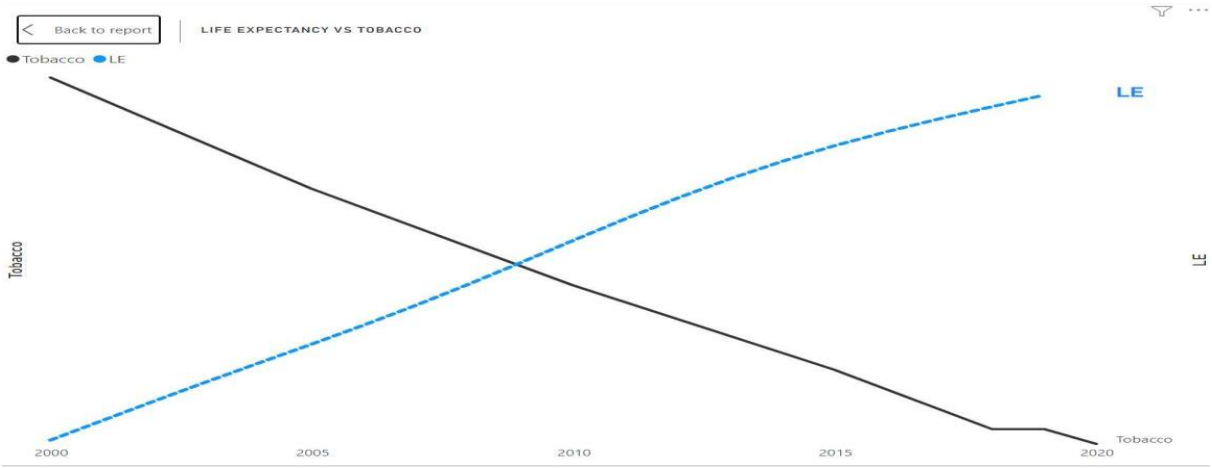



*Figure 16 LE Vs Tobacco*

Figure 16 graphs represent answer to our question related to does tobacco have any impact on life expectancy.

The two lines shown on this graph are: the main blue line shows the life expectancy trend in total across the years, while the grey line shows the tobacco. The x-axis in this picture shows years and y-axis shows the tobacco. The life expectancy starts increasing when decrease in intake of tobacco. The major take of tobacco is in Muslim countries. This figure 16 shows when tobacco take down the LE increases.

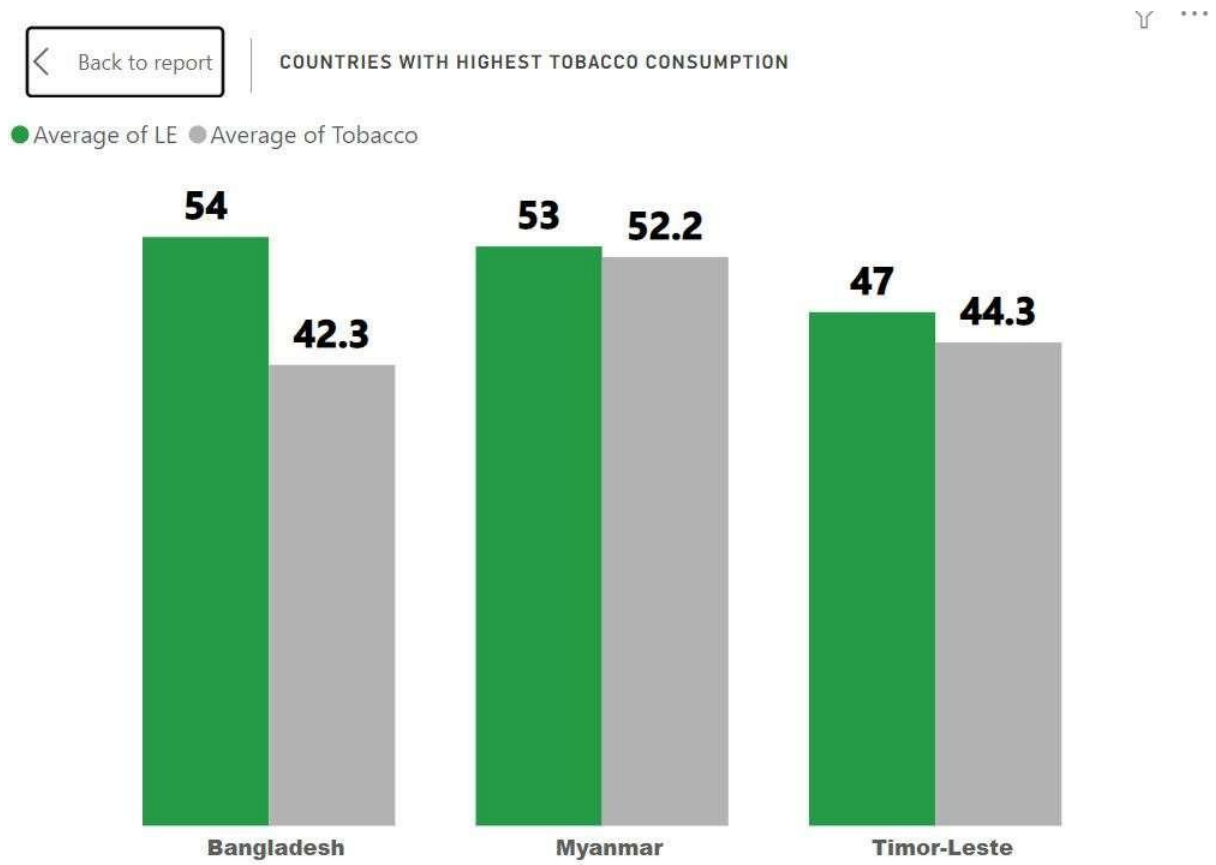



*Figure 17 countries with hight tobacco consumption*

In the above figure 17 green bar show the average of life expectancy and grey show average of tobacco. From this figure 17 its clearly shown that the top 3 countries which have high tobacco consumption have less life expectancy and these are mostly Muslim countries.

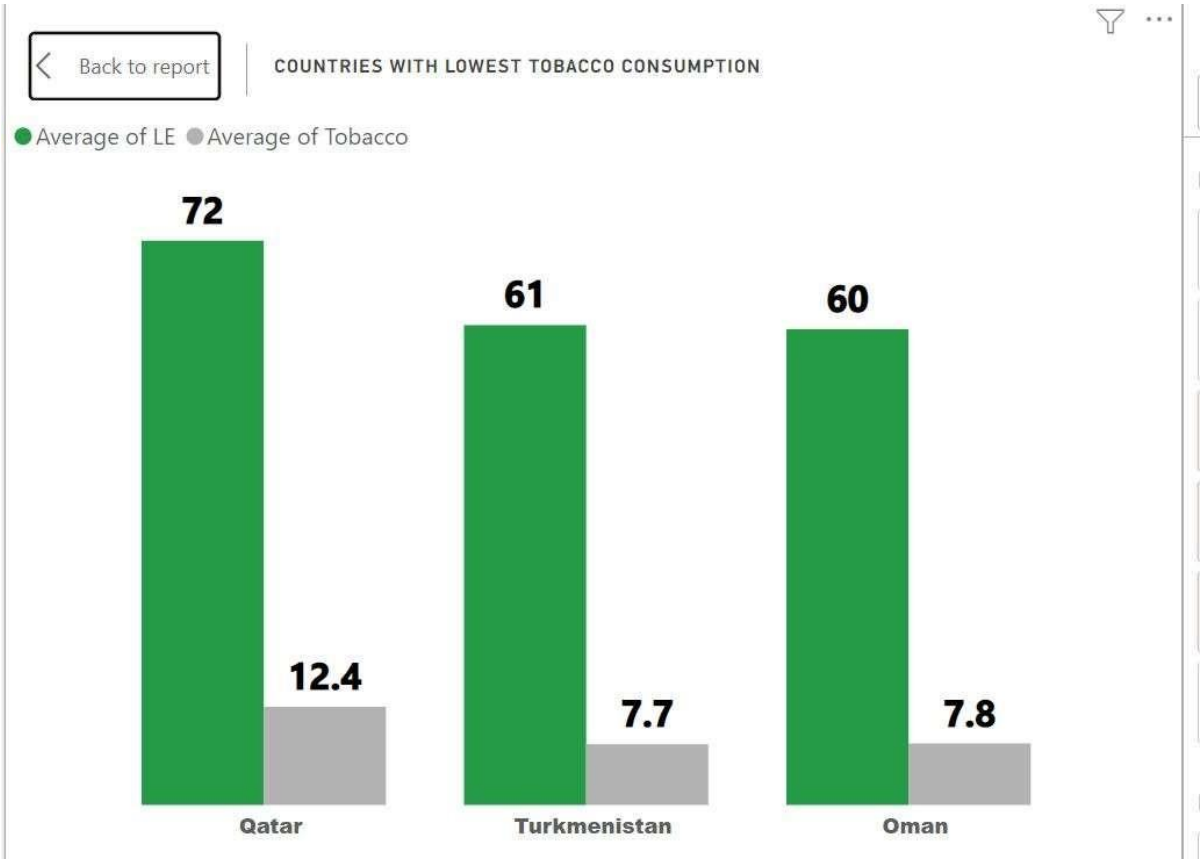



*Figure 18 with lowest tobacco consumption*

In the above figure 18 green bar show the average of life expectancy and grey show average of tobacco. From this figure we can clear understand that the top 3 countries which have lowest tobacco consumption are Muslim countries and they have high life expectancy as compared to figure 17 with high consumption of tobacco.

Overall, the highest smoking rates are found in Southeast Asia and the Balkan region of Europe. Western European countries and the Americas tend to have lower smoking rates.

## Question 6: How are life expectancies impacted by mortality rates?

Figure 19 provide a clear answer to our question.

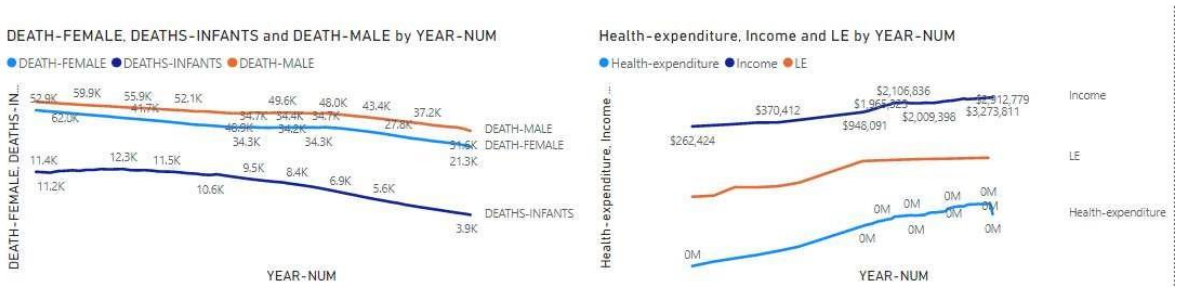



*Figure 19 mortality Vs health expenditure by the government*

The line chart on the left side of figure 19 show: the main orange line shows the male-death trend in total across the years-num, while the light blue line shows the female-death and the last one which is purple shows the death-infants. The y-axis in this picture shows male-death, female-death and infants-death and x-axis shows the year-num.

On the other hand, the line graph on the right side of figure graph 19 show: the main orange line shows the life expectancy trend in total across the years, while the light blue line shows the health expenditure and the last one which is purple shows the income. The y-axis in this picture shows health-expenditure, income and life expectancy and x-axis shows the year-num.

When the mortality rate is high, it means that more people are dying at an earlier age, which can lead to a lower life expectancy. Conversely, when the mortality rate is low, it means that fewer people are dying at an earlier age, which can lead to a higher life expectancy.

Therefore, a lower mortality rate, particularly among infants, children and young adults, generally results in a higher life expectancy. Mortality rate also varies by gender, with women usually having a lower mortality rate than men according to our insights of figure 19.

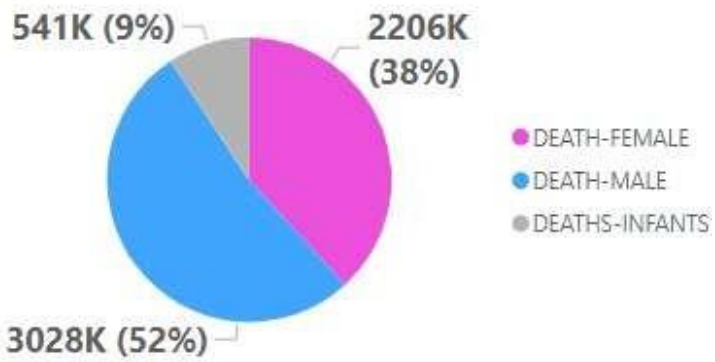



*Figure 20 pie chart for mortality*

Figure 20 illustrate the ratio of death of 2 centuries which depend on income and health-expenditure. This pie chart shows that in total percent of death the ratio of male death is more

than female and infants due because of income and different health things. The health issue ratio of male is more than women and due to no income or other problems the male death ratio is more. Total 52% out of 100% male death occur in last 2 centuries.

**Question 7: Do countries with dense populations typically have shorter life expectancies?**

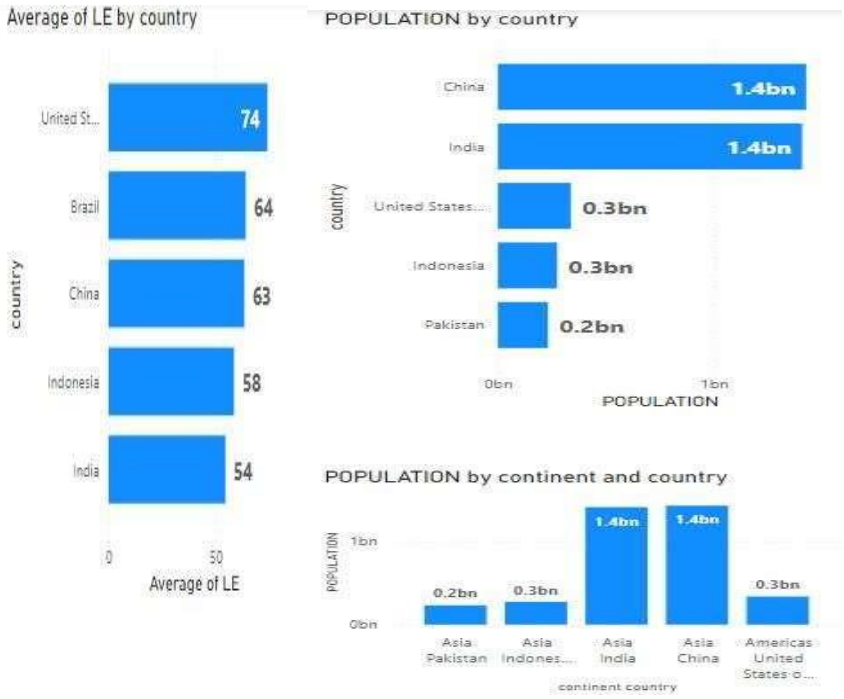



*Figure 21 dense population vs life expectancy*

In figure 21 we can clearly observe that high population counties don't have high life expectancy. For example, China has high population but no high average life expectancy, as compared to united states which have the highest average of life expectancy but don't have high population other country regarding other countries territory and resources. In this figure India have low average of life expectancy but India is the 2nd in population wise country around the world.

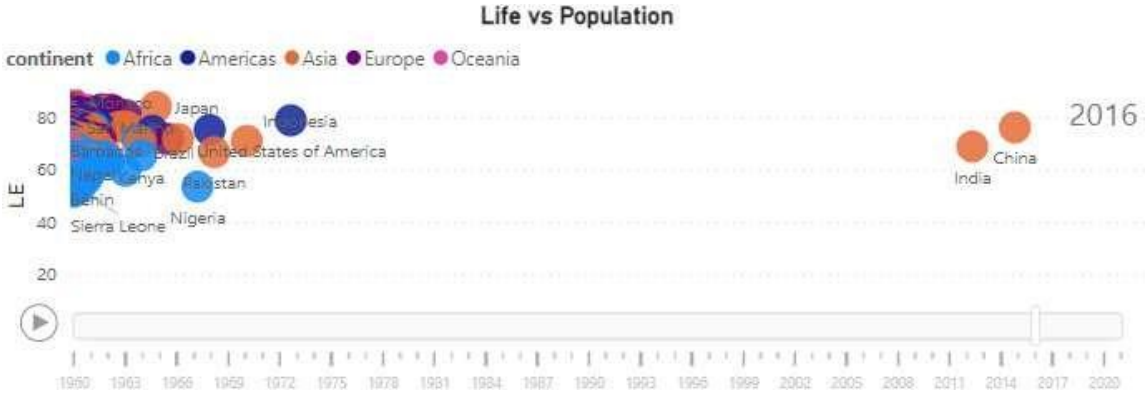



*Figure 22 population vs life expectancy*

In figure 22, we can see countries with less population and developed countries have higher life expectancy for exam Japan, America and most of European countries have have highest live expectancy majority of the countries.

## Summary

- Life expectancy began to increase significantly after 1950 due to, medical advances, such as the development of antibiotics and vaccines, which led to a reduction in deaths from infectious diseases.
- The expansion of access to healthcare and improvement of living conditions, which led to improvements in housing, sanitation, and access to clean water and food.
- A reduction in infant and child mortality.
- Low life expectancy is observed in African region and High life expectancy is observed in Europe region due to economical difference.
- Income is one of the most important factors which can increase life expectancy depends. If you have high income means you have high living standards access of food, medicine etc. as indicated in pervious point Africa has having low life expectancy.
- All the top 5 counties which have low life expectancy are from African region.
- In Arab region the ratio of life expectancy is high because stable income and wealthy government.
- The modest drinkers, no more than one drink a day, had benefits and could gain nearly 1 year in life expectancy, in contrast to a loss of nearly 7 years if drinking more than that.
- Life expectancy for smokers is **at least 10 years shorter than for non-smokers**. Quitting smoking before the age of 40 reduces the risk of dying from smoking-related disease by about 90%.
- European counties having high alcohol consumptions rates.
- Muslims counties having low alcohol consumption rates
- Overall, the highest smoking rates are found in Southeast Asia and the Balkan region of Europe. Western European countries and the Americas tend to have lower smoking rates.

- High population counties don’t have high life expectancy

### Recommendation

To improve overall global life expectancy, a multi-faceted approach is needed. One important aspect is to improve access to quality healthcare. This can include increasing funding for healthcare systems, training more healthcare professionals, and implementing policies to improve healthcare access and equity. Additionally, addressing social determinants of health such as income, education, and access to healthy food and clean water can have a significant impact on improving overall health and life expectancy.

Another important aspect is addressing major risk factors that can impact life expectancy, such as tobacco use and excessive alcohol consumption. Implementing policies to reduce tobacco use, such as increasing taxes on tobacco products, enacting bans on smoking in public places, and providing access to resources for smoking cessation can help lower the risk of premature death.

Another recommendation is encouraging a healthier lifestyle, promoting regular physical activity, healthy diet, and avoiding risky behaviours can contribute to lower mortality rates and increase life expectancy.

Furthermore, investing in research, data collection, and monitoring to better understand the drivers of health and longevity can help to inform and improve policies that aim to increase global life expectancy.

In summary, a comprehensive and multi-disciplinary approach is needed to improve global life expectancy, which includes addressing social determinants of health, reducing risk factors, encouraging healthy behaviours, and investing in research.

## Conclusion

After taking big data and business intelligence module and final using Power BI for creating my ICA project, I have a solid understanding of how to connect to and import data from various sources, as well as how to create and publish interactive visualizations and reports. I should be able to use the various tools and features in Power BI to analysed and communicate data insights effectively. It can be a powerful tool to help me discover insights, and share them with others, through interactive data visualization. Based on my continued use, I could gain proficiency in data modelling, building relationships and calculated tables, dashboards, and report creation, etc.

Secondly power BI will be the most useful tool I have in my day to day task in the field of data analytics.

# Section 2: Business Intelligence Design

## Data Pre-Processing or Data Cleansing

## Data Loading

From cleaning data to creating dashboard and reports, the first step is to load data in power BI without getting errors. Our source is excel sheet with size 326kB. We will use import mode to load data in BI. This will create copy of data in cache memory of BI. We perform the following steps in sequence.

i. Open Microsoft Power BI.

ii. Get Data.

iii. Select Text/CSV.

iv. Select file in open window.

200 rows data will be loaded for preview. From here we have two options either load the data and do transformation later, or direct get into power query mode for the transformation. We will select load option to view all data is uploaded successfully.

V. View data in Data Panel

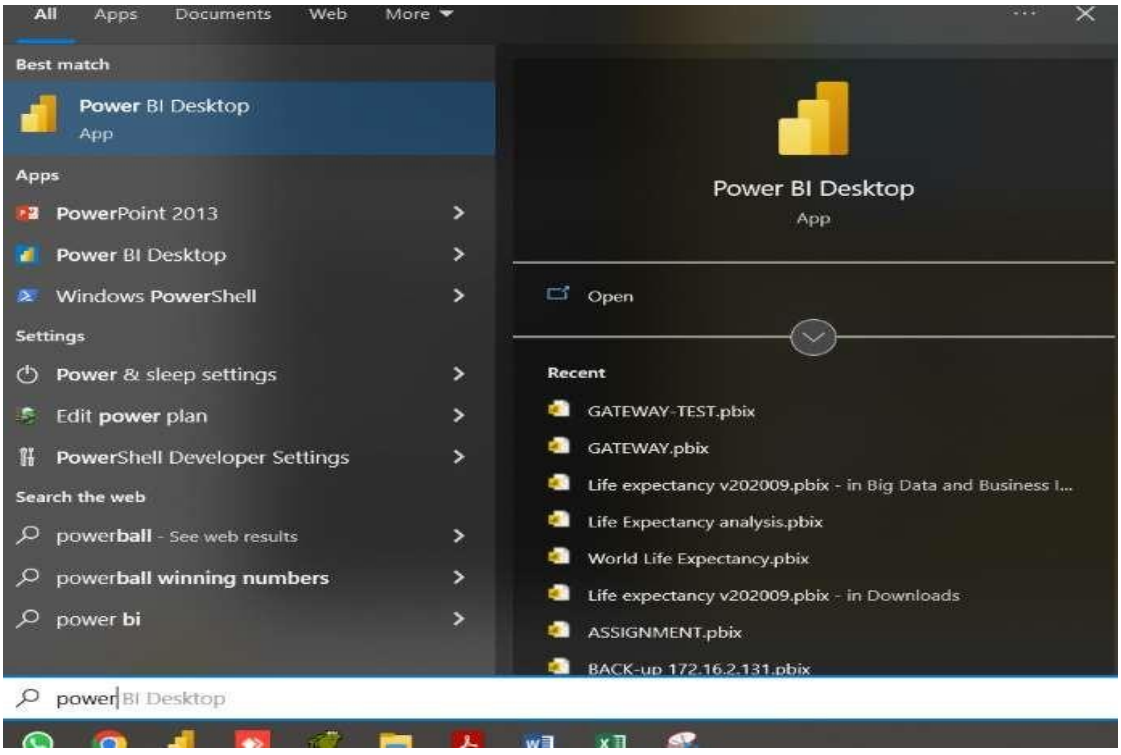



*Figure 23 Open Microsoft Power BI*

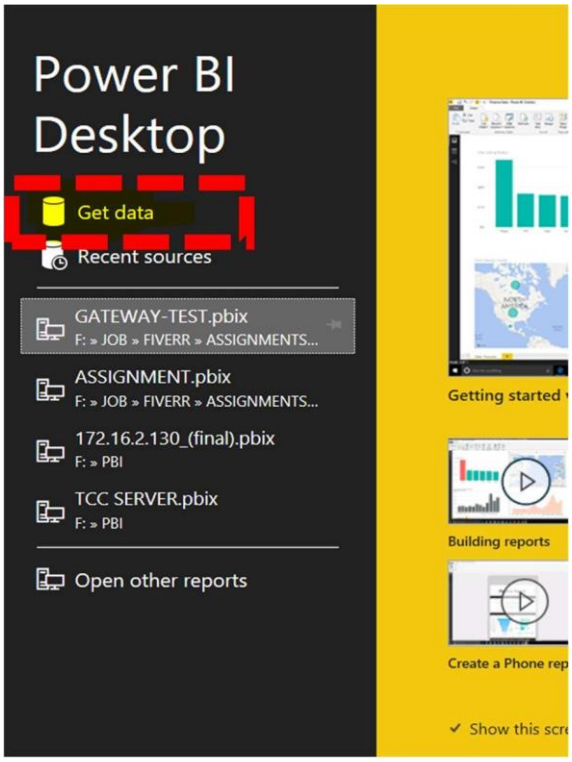



Figure 6 Get Data.

*Figure 24 Get data*

The dataset format is stored on the local system as a Comma separated file (csv). Click on "Text/CSV" and connect.

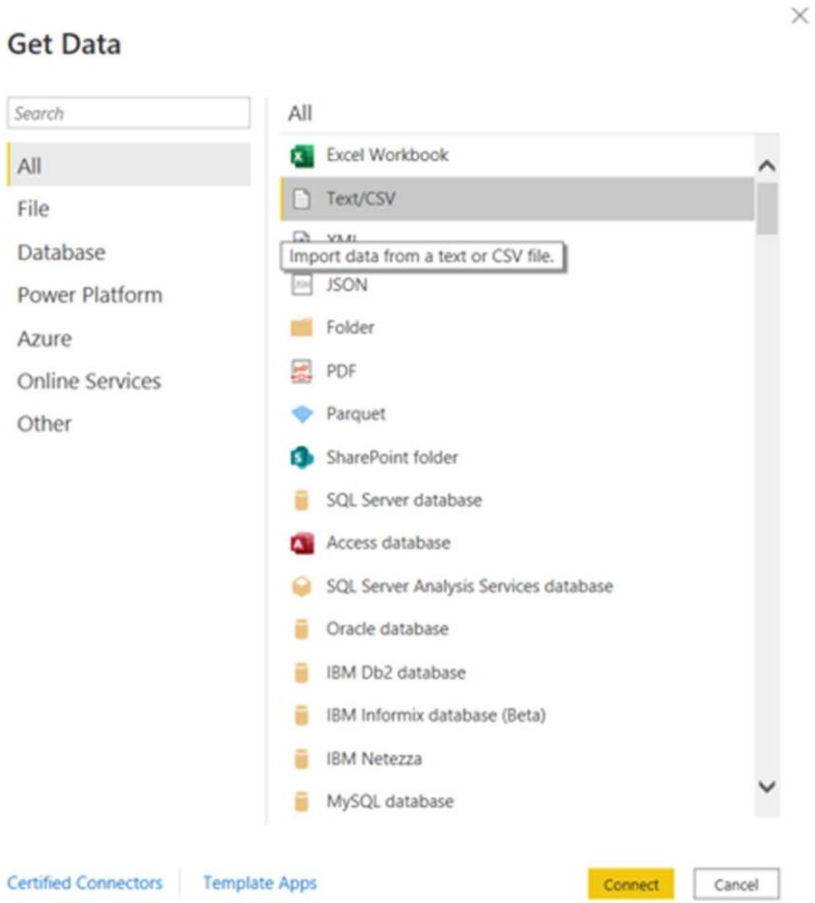



*Figure 25 select text/Csv*

Select the dataset from the location on the local system. After selecting the dataset, it pops up a dialog box to begin the loading process.

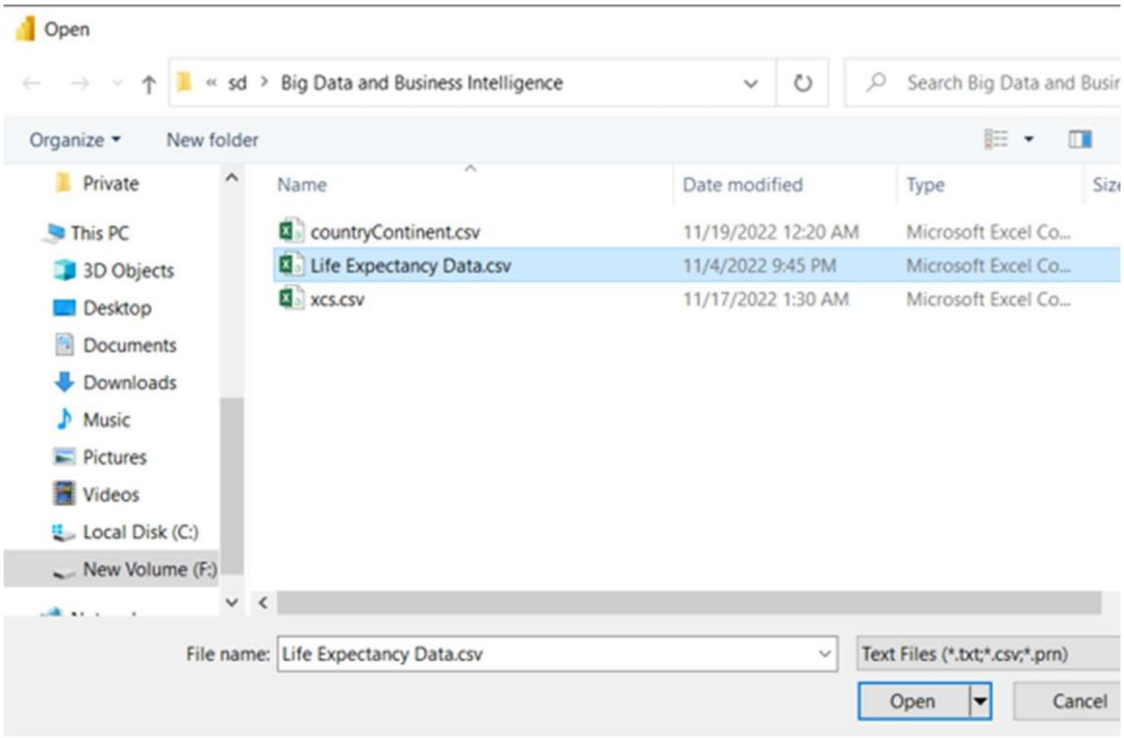



*Figure 26 Select file in open window*

During loading the data in, some basic info on the data sub window is as follow.

## File origin

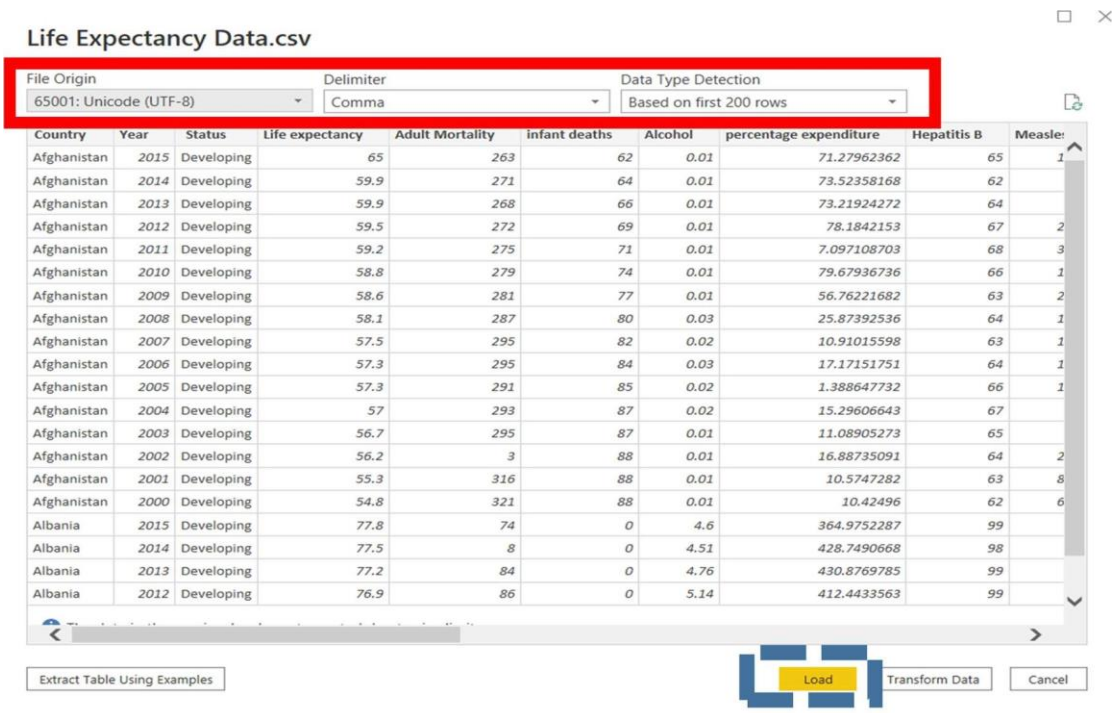
UTF-8 is a variable-length character encoding used for electronic communication. Defined by the Unicode Standard, the name is derived from Unicode Transformation Format – 8-bit. UTF-8 is capable of encoding all 1,112,064 valid character code points in Unicode using one to four one-byte code units. Power Bi auto picks the File origin encoding. (Figure 7) bellow.

## Delimiter

The delimiter for a csv file is commas.

## Data type Detection

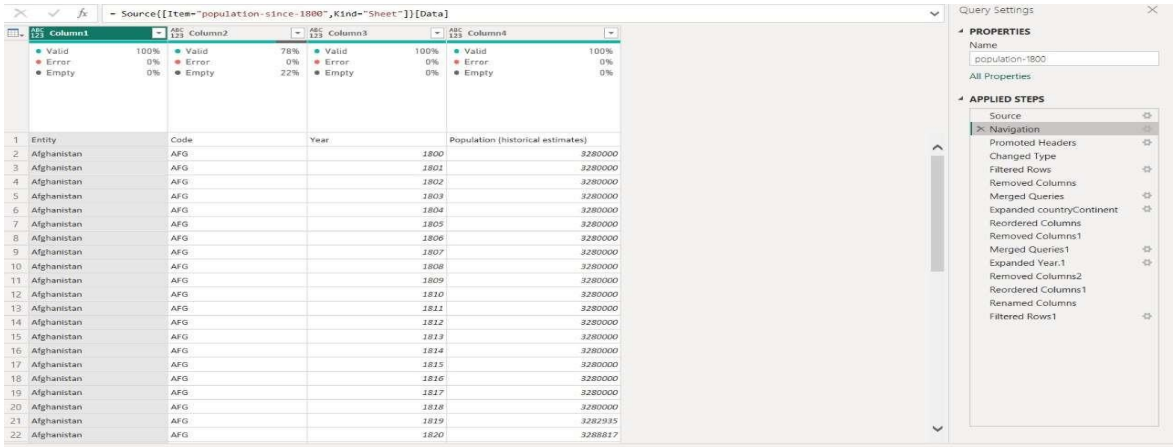
The data type is calibrated based on the first 200 rows by default



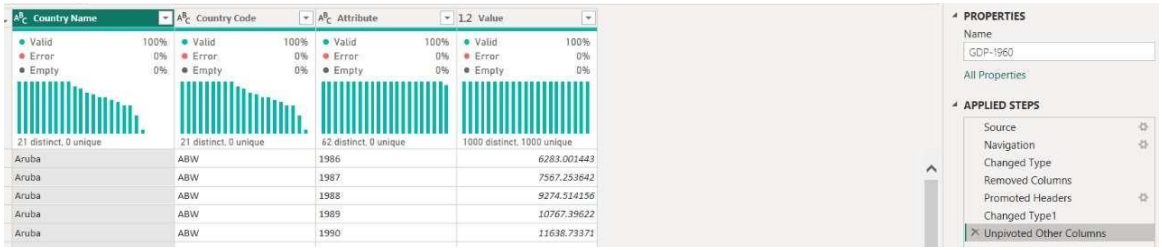

| Country | Year | Status | Life expectancy | Adult Mortality | infant deaths | Alcohol | percentage expenditure | Hepatitis B | Measle |
|---|---|---|---|---|---|---|---|---|---|
| Afghanistan | 2015 | Developing | 65 | 263 | 62 | 0.01 | 71.27962362 | 65 | 1 |
| Afghanistan | 2014 | Developing | 59.9 | 271 | 64 | 0.01 | 73.52358168 | 62 | |
| Afghanistan | 2013 | Developing | 59.9 | 268 | 66 | 0.01 | 73.21924272 | 64 | |
| Afghanistan | 2012 | Developing | 59.5 | 272 | 69 | 0.01 | 78.1842153 | 67 | 2 |
| Afghanistan | 2011 | Developing | 59.2 | 275 | 71 | 0.01 | 7.097108703 | 68 | 3 |
| Afghanistan | 2010 | Developing | 58.8 | 279 | 74 | 0.01 | 79.67936736 | 66 | 1 |
| Afghanistan | 2009 | Developing | 58.6 | 281 | 77 | 0.01 | 56.76221682 | 63 | 2 |
| Afghanistan | 2008 | Developing | 58.1 | 287 | 80 | 0.03 | 25.87392536 | 64 | 1 |
| Afghanistan | 2007 | Developing | 57.5 | 295 | 82 | 0.02 | 10.91015598 | 63 | 1 |
| Afghanistan | 2006 | Developing | 57.3 | 295 | 84 | 0.03 | 17.17151751 | 64 | 1 |
| Afghanistan | 2005 | Developing | 57.3 | 291 | 85 | 0.02 | 1.388647732 | 66 | 1 |
| Afghanistan | 2004 | Developing | 57 | 293 | 87 | 0.02 | 15.29606643 | 67 | |
| Afghanistan | 2003 | Developing | 56.7 | 295 | 87 | 0.01 | 11.08905273 | 65 | |
| Afghanistan | 2002 | Developing | 56.2 | 3 | 88 | 0.01 | 16.88735091 | 64 | 2 |
| Afghanistan | 2001 | Developing | 55.3 | 316 | 88 | 0.01 | 10.5747282 | 63 | 8 |
| Afghanistan | 2000 | Developing | 54.8 | 321 | 88 | 0.01 | 10.42496 | 62 | 6 |
| Albania | 2015 | Developing | 77.8 | 74 | 0 | 4.6 | 364.9752287 | 99 | |
| Albania | 2014 | Developing | 77.5 | 8 | 0 | 4.51 | 428.7490668 | 98 | |
| Albania | 2013 | Developing | 77.2 | 84 | 0 | 4.76 | 430.8769785 | 99 | |
| Albania | 2012 | Developing | 76.9 | 86 | 0 | 5.14 | 412.4433563 | 99 | |

Figure 7 Data Preview

Our dataset is successful load into the power bi as we can see the tables and column of our dataset on the right side of our screen.




| | Entity | Code | Year | Population (historical estimates) |
|---|---|---|---|---|
| 2 | Afghanistan | AFG | 1800 | 3280000 |
| 3 | Afghanistan | AFG | 1801 | 3280000 |
| 4 | Afghanistan | AFG | 1802 | 3280000 |
| 5 | Afghanistan | AFG | 1803 | 3280000 |
| 6 | Afghanistan | AFG | 1804 | 3280000 |
| 7 | Afghanistan | AFG | 1805 | 3280000 |
| 8 | Afghanistan | AFG | 1806 | 3280000 |
| 9 | Afghanistan | AFG | 1807 | 3280000 |
| 10 | Afghanistan | AFG | 1808 | 3280000 |
| 11 | Afghanistan | AFG | 1809 | 3280000 |
| 12 | Afghanistan | AFG | 1810 | 3280000 |
| 13 | Afghanistan | AFG | 1811 | 3280000 |
| 14 | Afghanistan | AFG | 1812 | 3280000 |
| 15 | Afghanistan | AFG | 1813 | 3280000 |
| 16 | Afghanistan | AFG | 1814 | 3280000 |
| 17 | Afghanistan | AFG | 1815 | 3280000 |
| 18 | Afghanistan | AFG | 1816 | 3280000 |
| 19 | Afghanistan | AFG | 1817 | 3280000 |
| 20 | Afghanistan | AFG | 1818 | 3280000 |
| 21 | Afghanistan | AFG | 1819 | 3282935 |
| 22 | Afghanistan | AFG | 1820 | 3288817 |




| Country Name | Country Code | Attribute | Value |
|---|---|---|---|
| Aruba | ABW | 1986 | 6283.001443 |
| Aruba | ABW | 1987 | 7567.253642 |
| Aruba | ABW | 1988 | 9274.514156 |
| Aruba | ABW | 1989 | 10767.39622 |
| Aruba | ABW | 1990 | 11638.73371 |

# Data Cleaning

To clean the data, go to home ribbon, click on transform data icon.

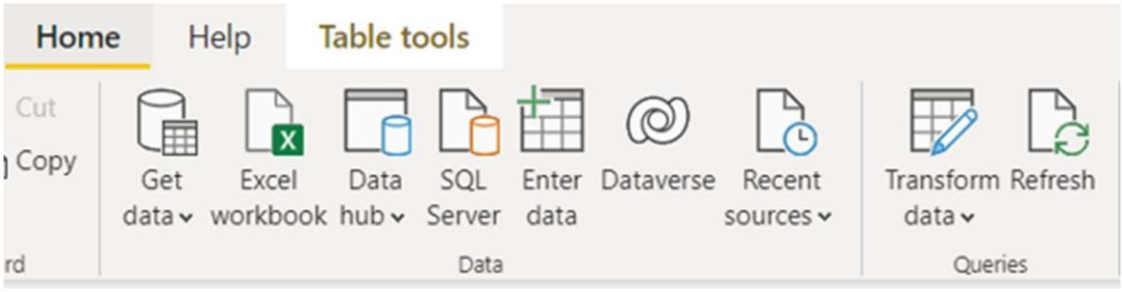



*Figure 27 Transform data*

To power Query Editor Window will be open, that's the platform that will be used for data cleaning and remodelling.

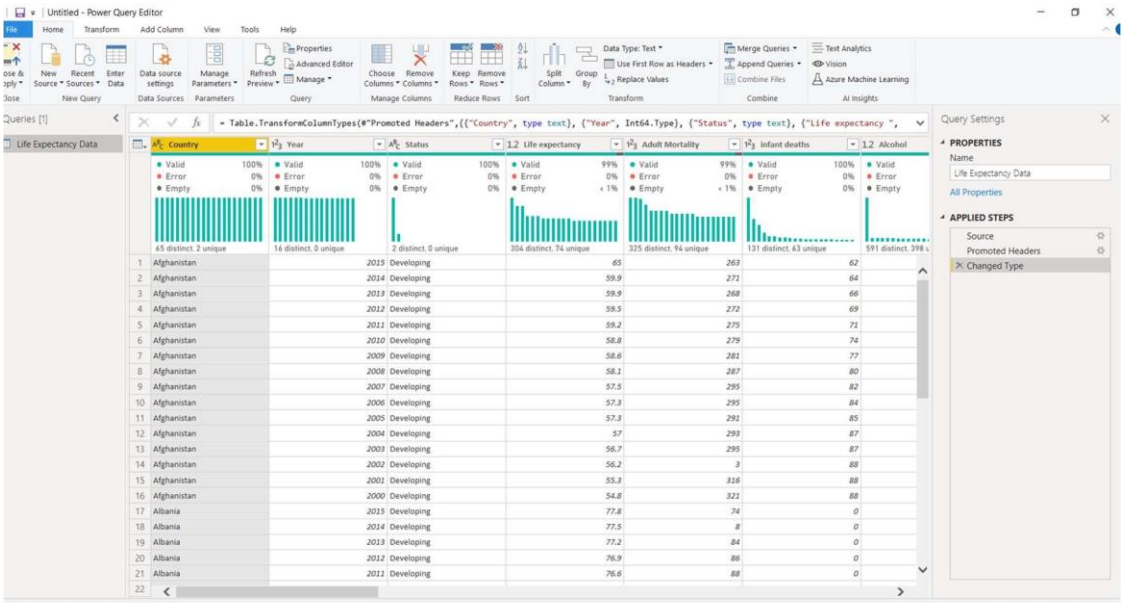



Figure 9 Power Query Mode

Above each column, an undersized summary about data is shown to get instant view about how many nulls values, unique or distinct values present in column along with data distribution.

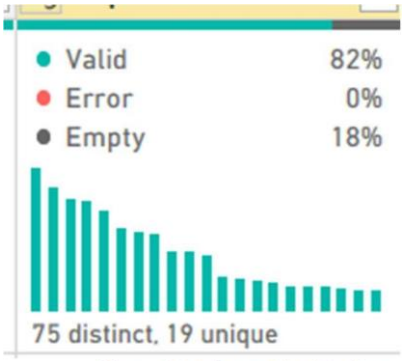



Figure 10 Column Distribution

## Null Values

Right click on the column which shows empty ratio values. A drop-down menu will appear. Click on replace values option. Replace null with 0. (Figure 11). Same procedure can be done by selecting multiple columns. Any anomalies like O instead of 0 can be fixed by this.

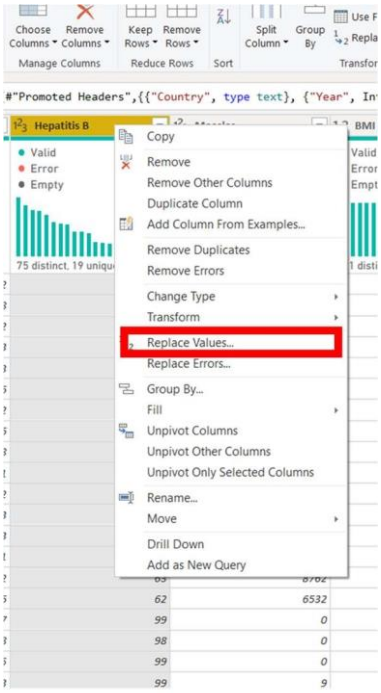



Figure 11 Replace Values

## Data Types

Most important thing is to check that data columns should be in the correct datatype. Data like primary keys, etc should be in whole number data type. Date and time columns should be in date/time format. Point values in decimal number format. Data type can be change by clicking on column name. or by right click and drop-down option A drop down menu will be displayed. Figure 12.

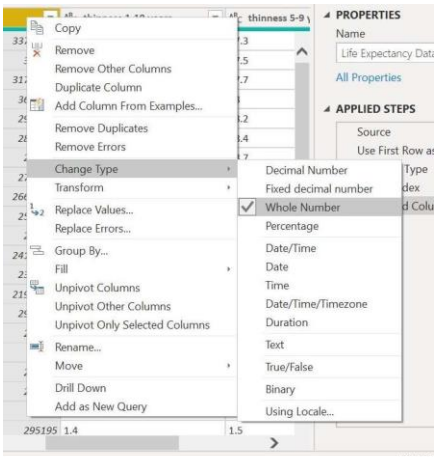



Figure 29 by Menu

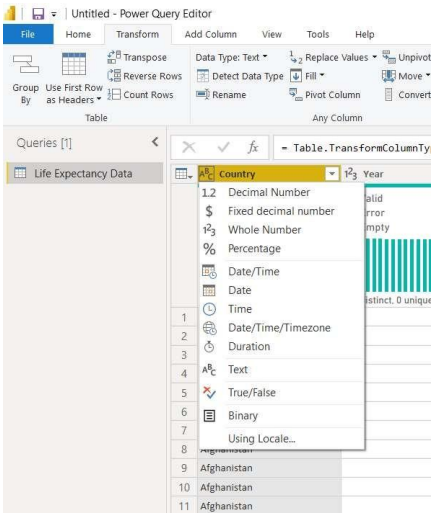



Figure 28 Data Type settings

A summary table is displayed for all the columns of dataset

| Sr | Column names | Column data | Data type |
|---|---|---|---|
| 1 | Country | Afghanistan | Text |
| 2 | Year | 2015 | Whole number |
| 3 | Country Code | Developing | Text |
| 4 | Value | 65 | Decimal number |

# BI Data Modelling

After uploading every dataset as separate data source, year and country is the only common key between every table. Created a snowflake scheme as shown below figure 30.

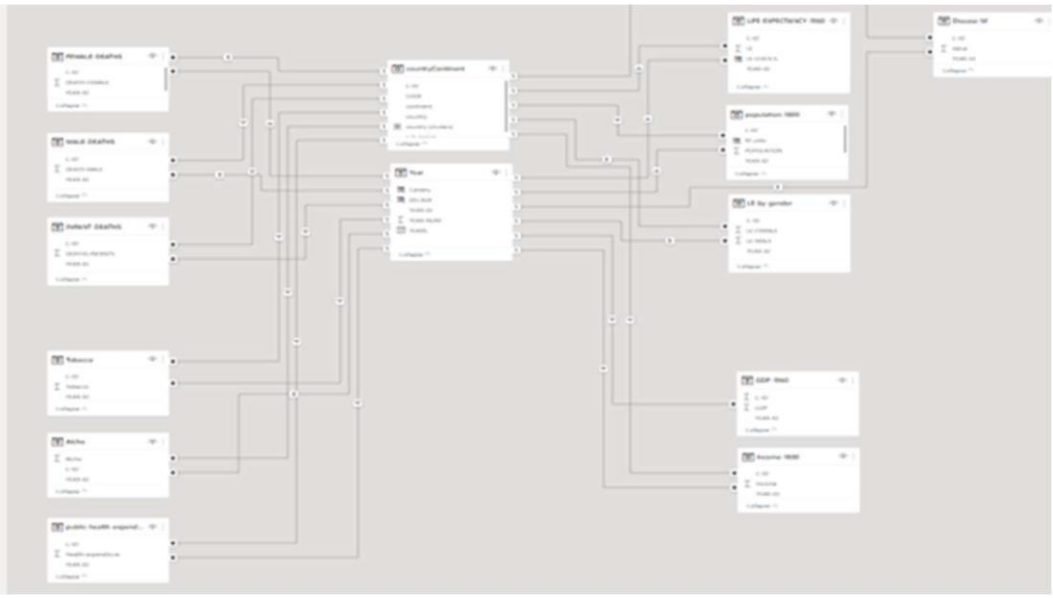


*Figure 30 Snowflake scheme*

Merge two queries be common key as below. The relationship between queries is one to many

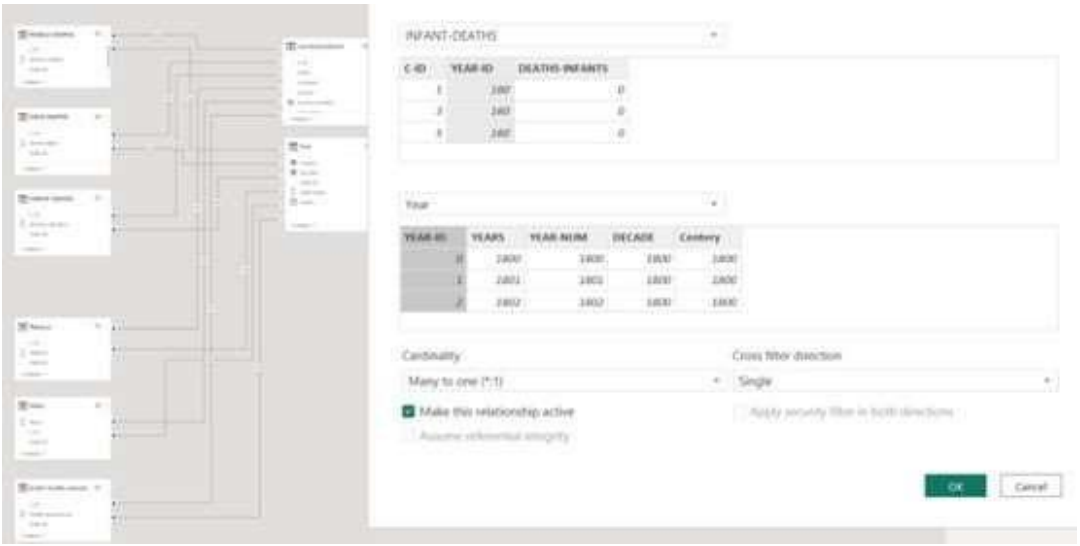


*Figure 31 Creating relation one-to-many*

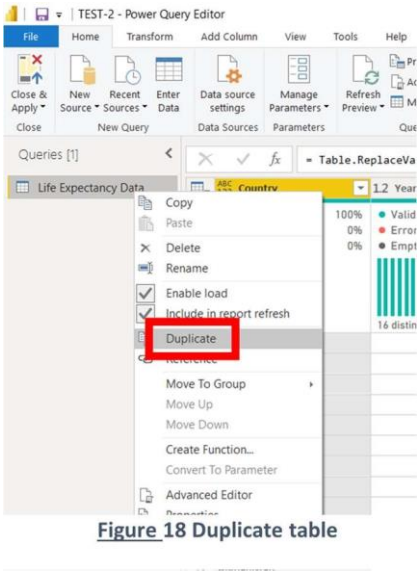



**Figure 18 Duplicate table**

Open Power Query Mode. And create the duplicate of original data table before applying changes.

## Normalization of Data Set

Create a duplicate table and perform the following steps:

1. Rename the query (table) name.
2. Select **country, status** and remove other columns.
3. Remove duplicate values.
4. Add index column, and rename it with C-ID.
5. Observe right side Applied panel, steps are added.

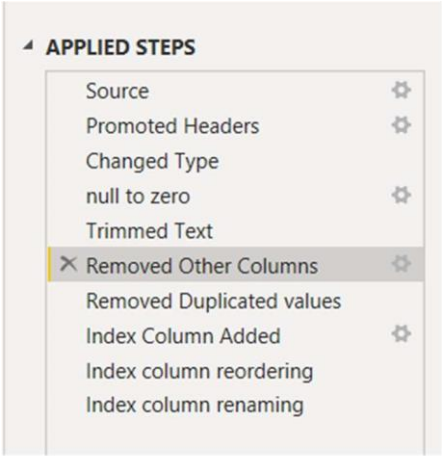



**Figure 19 Applied Steps**

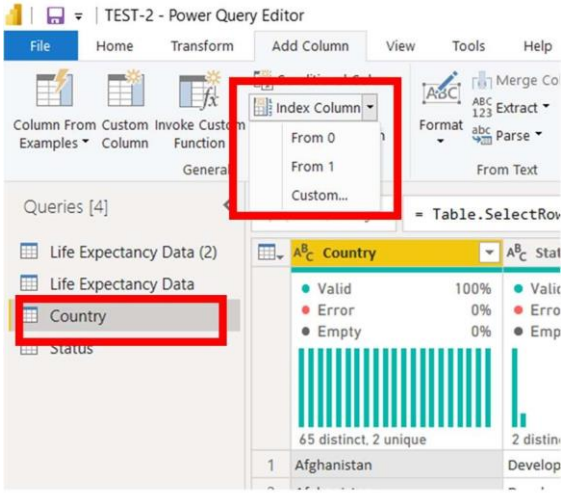



**Figure 20 Add Index Column and query renaming**

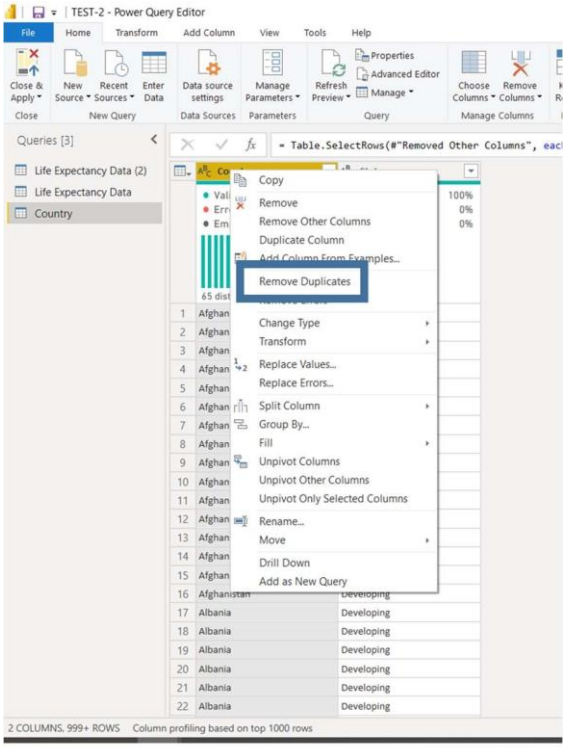



**Figure 21 Remove Duplicate values**

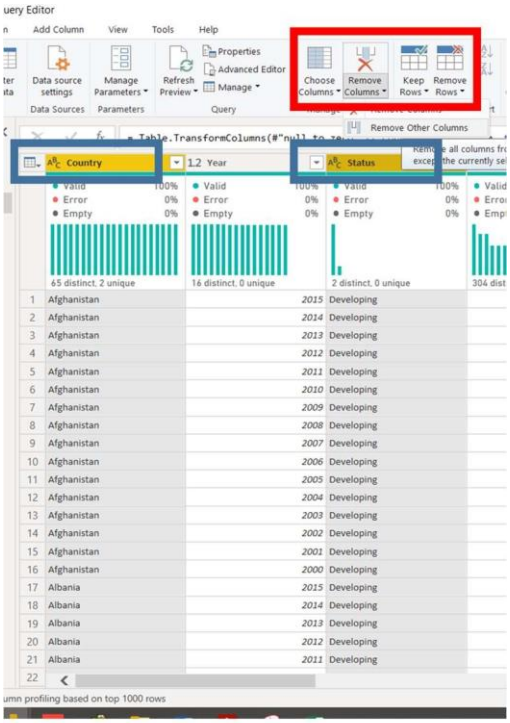



Figure 22 remove other columns

## Query Merging

We can see that in country table, status value is repeating, so create a separate table for status and merge this with country table by foreign key.

1. Select column with repeating values.
2. Add as new Query
3. Convert list to table (Figure 24).
4. Remove Duplicate Values
5. Add index column.
6. Rename column names.
7. Rename applied steps strings.
8. Go to Country query and select merge.
9. Select the Status column from country query and select the status table then status column of other.
10. Use Left outer join and click ok.
11. Status column will be added in country column.
12. Expend the column and only select primary key column.
13. Remove the status column.
14. Apply and save the power query mode.
15. Check model plane to see relationship. By default relation is one to many.

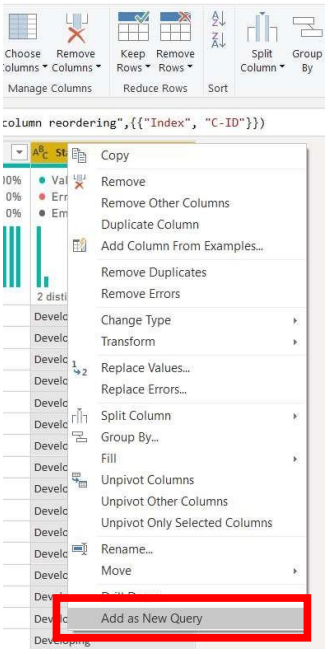



Figure 32 New Query

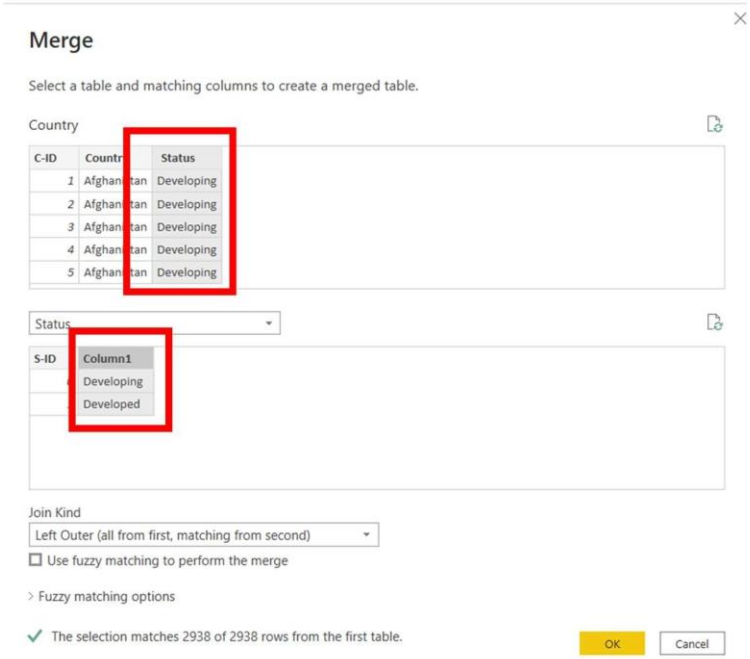



Figure 26 Merge Query

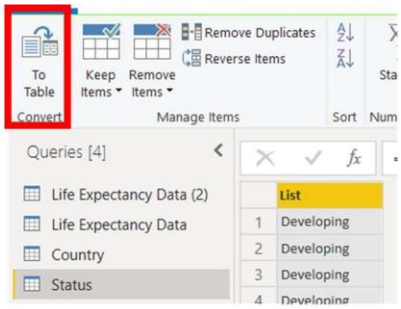



Figure 26List to Table

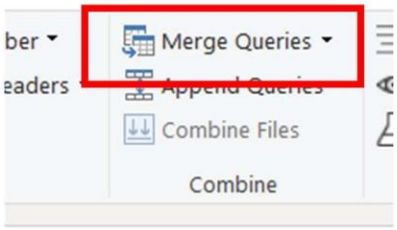



Figure 26 Merge Queries

*Figure 33 Merging table*

Same Procedure for the year table.

1. Select year column.
2. Add as new query.
3. Convert to table.
4. Remove duplicate.
5. Add index column.

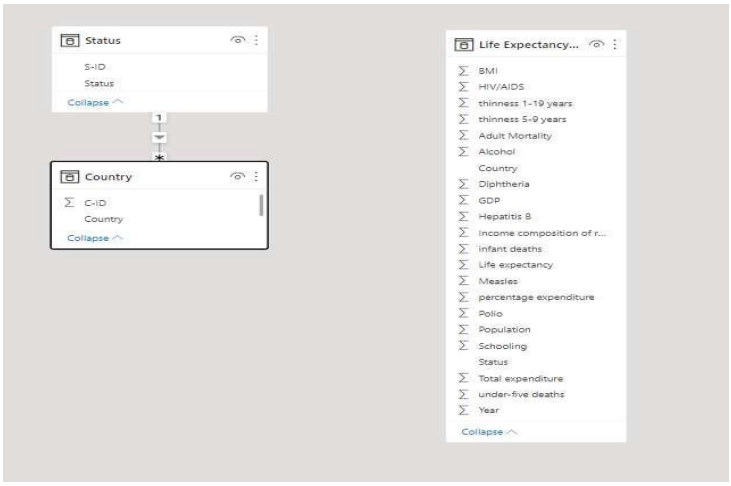



**Figure 34 Model Plane**

## Fact Table

We need to create a fact table of countries/years that should be connected to every other dimension table. From data analysis we realise that country and year is mandatory for any data distribution. So, we performed the following steps:

1. Create the copy of original query.
2. Rename the table as Data table.
3. Merge the county table with country column.
4. Merge the year table with table column.
5. Removed the country and year original column and replaced them by both tables' primary keys.
6. Add index column
7. Rename the index column with key-id
8. Create the copy of this table.
9. Rename this table as linking table.

Remove all columns except key-id column and country C-id column and year Y-Id column.

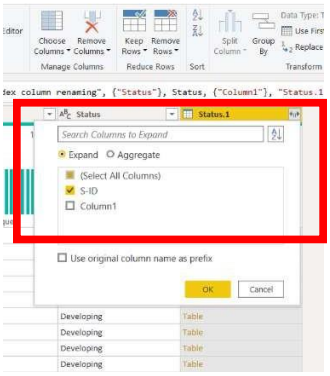



**Figure 35 Table extend and column selection**

After these steps we have data model as shown in below picture.

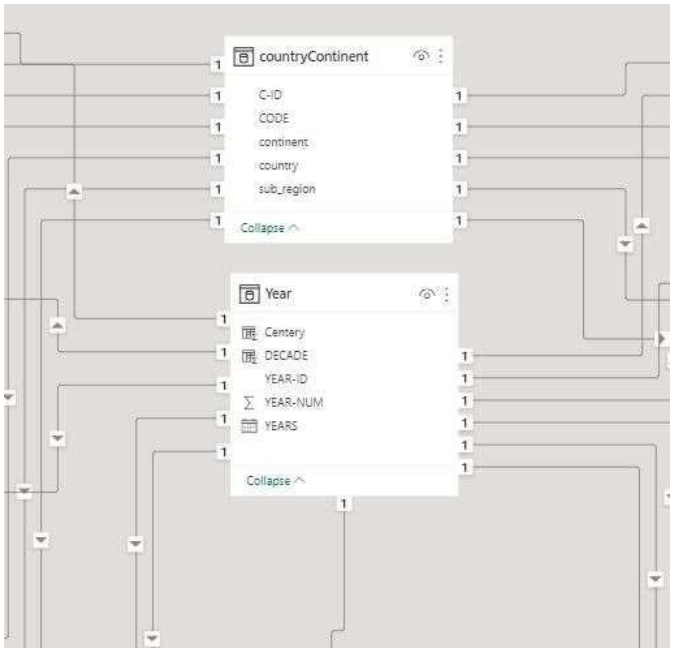



*Figure 36 Fact Table*

## Dimension Tables

As decided in above topics we will create dimensions' tables according to the given below figure WE will perform the following steps.

1. Create the copy of data table.
2. Except primary key and relatable data columns remove the others.
3. Rename the table according to the below figure.
4. Repeat the same steps for multiple dimension tables.
5. After all the creation of dimension tables, delete the key-id column form data table.
6. Apply and save.
7. Go the model table. Grab the key-ID inside a dimension table and place it on fact table ley-id column.
8. Click on relationship line and change the cardinality to one to many.


*Figure 37 dimension table*

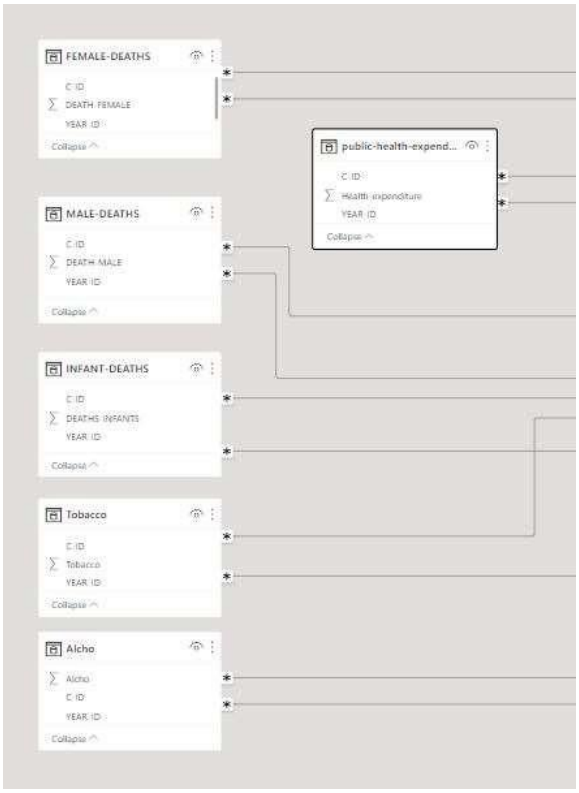



*Figure 38-dimension table*

# Final Model of relationship between our tables

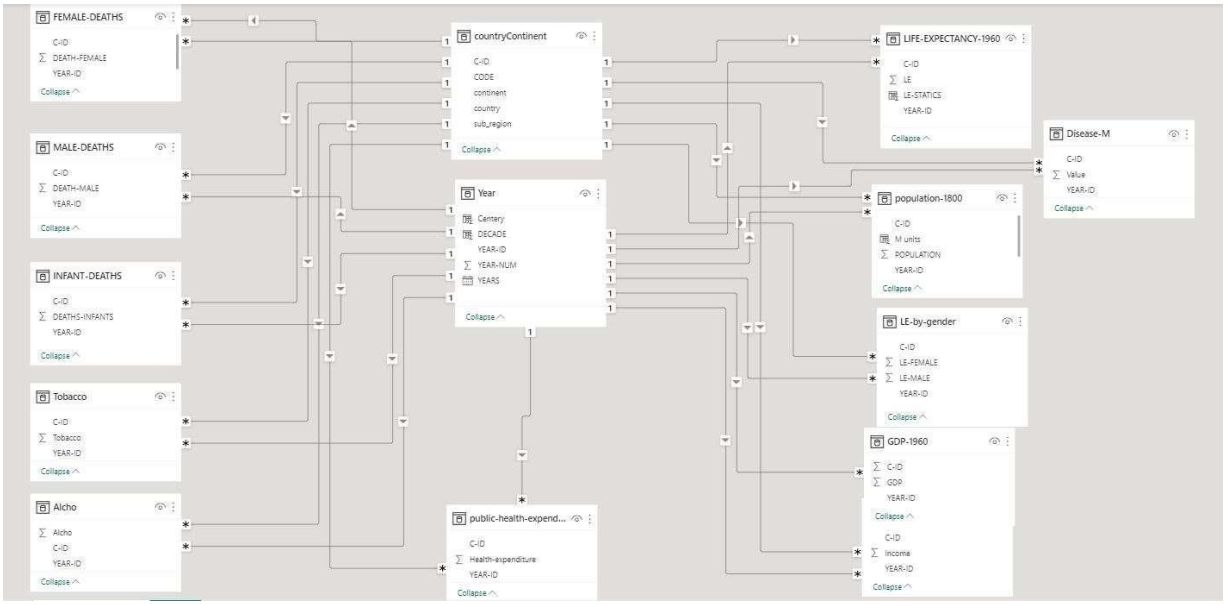



*Figure 39 Final model*

## DAX and M Language

We took data from multiple dataset so used most of mapping because a single data doesn't have the data we were looking for but I have use some DAX expression ill write bellow:

### Data Analysis Expressions

I. this Dax expression is used for a custom color scale for a chart based on a measure called "Life Expectancy for Chart." It uses a SWITCH statement with a series of conditions to assign different hex color codes to different ranges of values for the measure.

```
Life Expectancy Colour Coding For Chart =
SWITCH(
    TRUE(),
    SELECTEDVALUE('Year'[YEAR-NUM])<SELECTEDVALUE('Extra  Year'[Extra  year]),  "#dddddd",
    [Life Expectancy For Chart]<=30, "#D64550",
    [Life Expectancy For Chart]<=45, "#DE6A3E",
    [Life Expectancy For Chart]<=60, "#E89928",
    [Life Expectancy For Chart]<=70, "#E89928",
    [Life Expectancy For Chart]<=80, "#65606A",
    "#474A7D")
```

II. It takes the value from the column 'LIFE-EXPECTANCY-1960'[LE], divide it by 10 and then round it down to the nearest integer. And then the result is multiplied by 10, giving the value of the new measure LE-STATICS This formula can be used to bucket or group the value of Life Expectancy into bins of 10s, so that instead of having many different values, they are grouped together into ranges of 10.

```
LE-STATICS = INT('LIFE-EXPECTANCY-1960'[LE]/10)*10
```

III. The formula starts by declaring a variable called "a" and assigns the value of 'population-1800'[POPULATION] divided by 1000000, this basically converts the population from whatever unit it is in to million. Then the formula returns the value of the variable "a", which is the population in million units, this can be useful for making the data more readable or for comparing values in the same unit.

```
M units =
var a = 'population-1800'[POPULATION]/1000000
return a
```

IV. This is a DAX formula in Power BI that defines a new measure called "population(M)" by performing a mathematical operation on another measure called "population-1800[POPULATION]". It starts by using the SUM() function to sum all the values of the

column "population-1800[POPULATION]". Then it uses the DIVIDE() function which divides the sum of the population by 10000000, and this step reduces the population unit to million. The last parameter in the DIVIDE function "<1M" is optional, it is a custom format string that you can use to format the output of the division, and format it as <1M meaning less than 1 million, if the population is less than 1 million it will show "<1M" instead of a numeric value. This can be useful to make the data more readable. So the new measure "population(M)" gives the population in million units.

```
population(M)=
DIVIDE(SUM('population1800'[POPULATION]),10000000,"<1M")
```

V. It starts by using the SUMX() function which is an iterator function in Power BI, it iterates over the table "population-1800" and performs an expression on each row, in this case, it's adding the values of the columns 'population-1800'[C-ID] and 'population-1800'[YEAR-ID] and then summing the results of all the rows, in other words, it sums the values of C-ID and YEAR-ID columns for every row of the table population-1800.

```
test -2 = SUMX('population-1800','population-
    1800'[CID]+'population-1800'[YEAR-ID])
```

VI. This Dax formula is taking the value of the column "Year[YEAR-NUM]" which represents a year, divides it by 100, and rounds down to the nearest integer using the INT() function, this converts the year to century. Then it multiply the result of the operation with 100, which will give the beginning year of the century.

For example, the year 2018 will be converted to 20, by the division and then multiplied by 100 to give 2000, the beginning year of 21st century.

```
Centery = INT('Year'[YEAR-NUM]/100)*100
```

VII. This Dax formula taking the value of the column "Year[YEAR-NUM]" which represents a year, divides it by 10, and rounds down to the nearest integer using the INT() function, this converts the year to decade. Then it multiply the result of the operation with 10, which will give the beginning year of the decade.

For example, the year 2018 will be converted to 201, by the division and then multiplied by 10 to give 2010, the beginning year of 2010s decade.

```
DECADE = INT('Year'[YEAR-NUM]/10)*10
```

## M Language

Although power query is GUI based platform for handling data, but for every step like change type, rename etc. etc., M language coding take place in the background.

- M language is a Mashup query language

- M language is case-sensitive and functional language similar to F#, which can be used with Power Query in Excel, Get & Transform in Excel 2016, and Power BI Desktop.
- M language has following two main blocks: -
    - Let is used to define variables.
    - In is used to display output

Code can be seen in upper function bar.

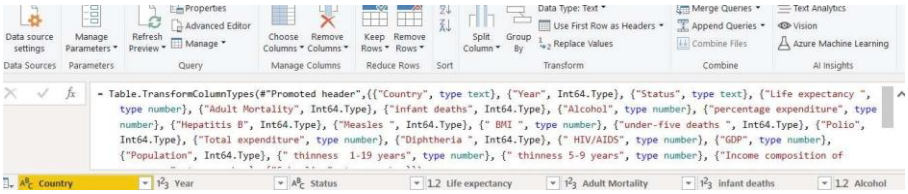



Figure 40 M Language Code

Following M functions are performed during data cleaning.

- **Promote headers:** = Table.PromoteHeaders (Source, [PromoteAllScalars=true])
- **Changed types:** = Table.TransformColumnTypes(#"Promoted Headers",{{"Entity", type text}, {"Code", type text}, {"Year", Int64.Type}, {"public_health_expenditure_pc_gdp", type number}})
- **Filtered row** = Table.SelectRows(#"Removed Columns2", each ([#"C-ID"] <> null))
- **Merged Queries** = Table.NestedJoin(#"Changed Type2", {"Country Code"}, countryContinent, {"CODE"}, "countryContinent", JoinKind.LeftOuter)

# Dashboard

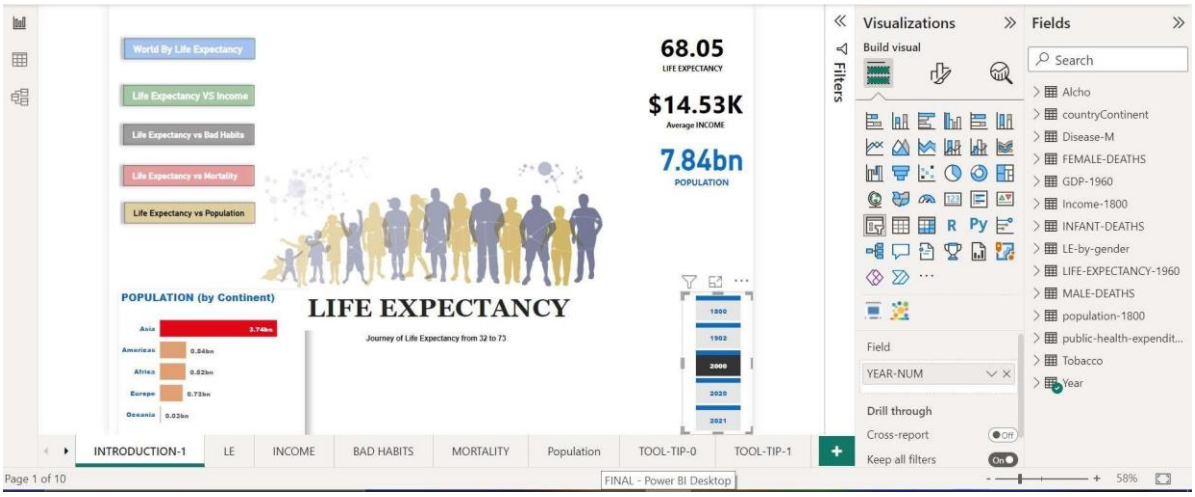



*Figure 41 Home page dashboard*

This is the introduction dashboard of LIFE EXPECTANCY. as we know life expectancy means life of a person expect to live in coming years. Well, it is based on average age of specific member. In figure 39 there is a chart which will explain the number of life expectancy of 2000s the results which we collect data of life expectancy is approximately 68.76. the average income of life per capita is almost $14.52k. total population might be 7.84bn. In this figure 39 we will explain the following some topic of LE.

- World by life expectancy
- Life expectancy vs income
- Life expectancy vs bad habit
- Life expectancy vs Mortality
- Life expectancy vs population

When we click on these dashboards url it will take us to the detail related to these topics which we chose.

Now we are going to discuss the population of life expectancy with respect to continents. i.e.

- Asia contains 3.74bn
- Europe contains 101.46bn
- Americans contains 0.84bn
- Africa contains 0.82bn
- Oceania contains0.03bn

The dashboard is customizable we just took for 2000s, you can choose according to your preference.

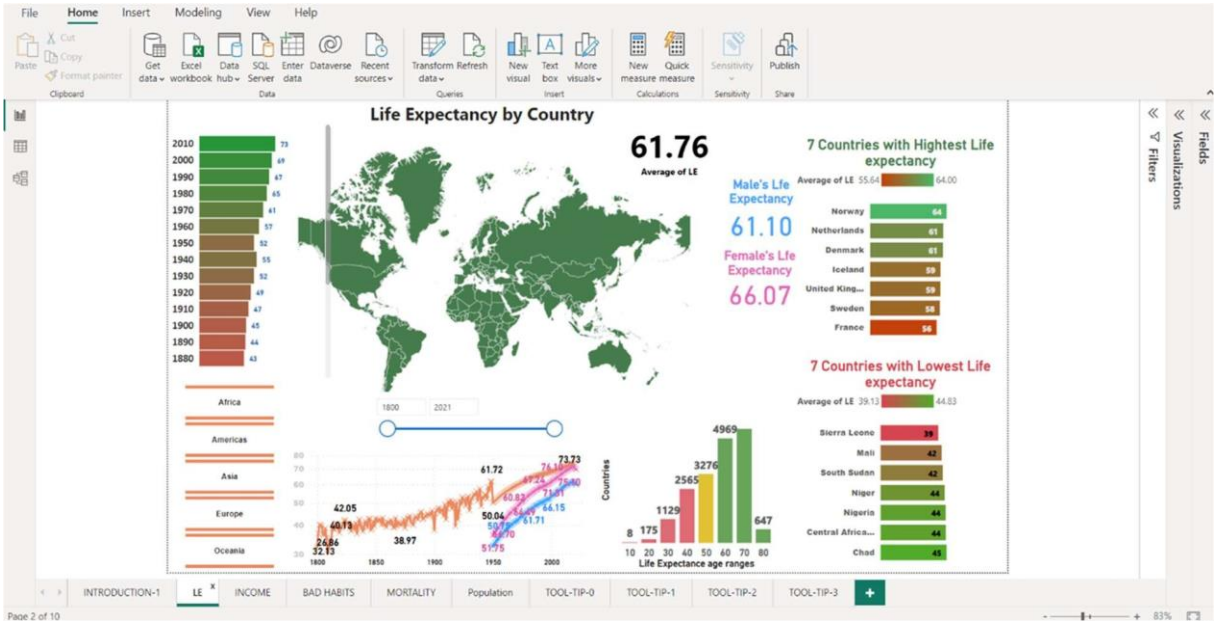



*Figure 42 Dashboard LE*

In figure 40 we will discuss about the life expectancy by country which will provide us detailed information about the country's average life. The graph on left corner tell us about the life expentancy till the 19th century by the ratio of ten years.

- Male's life expectancy is about 61.10.
- Female's expectancy is 66.07.

In during 1950s there is specific change in the graph of life expectancy. By this revolution in technology number of ages have quite different effect. Graph in the above figure will explain the fluctuations due to the revolution.

7 countries with highest life expectancy are

- Norway: 64.
- Netherlands: 61.
- Denmark: 61.
- Iceland: 59.
- United King: 59.
- Sweden: 58.
- France: 56.

7 countries with lowest life expectancy are

- Sierra Leone: 39.
- Mali:42.
- South Sudan:42.
- Niger:44.

- Nigeria:44.
- Central Africa:44.
- Chad:44

Furthermore, it will explain about the age ranges of the life expectancy. Figure 41 on bottom will provide us the information about the age ranges. From age ten to eighty the graph will provide us the numbers.

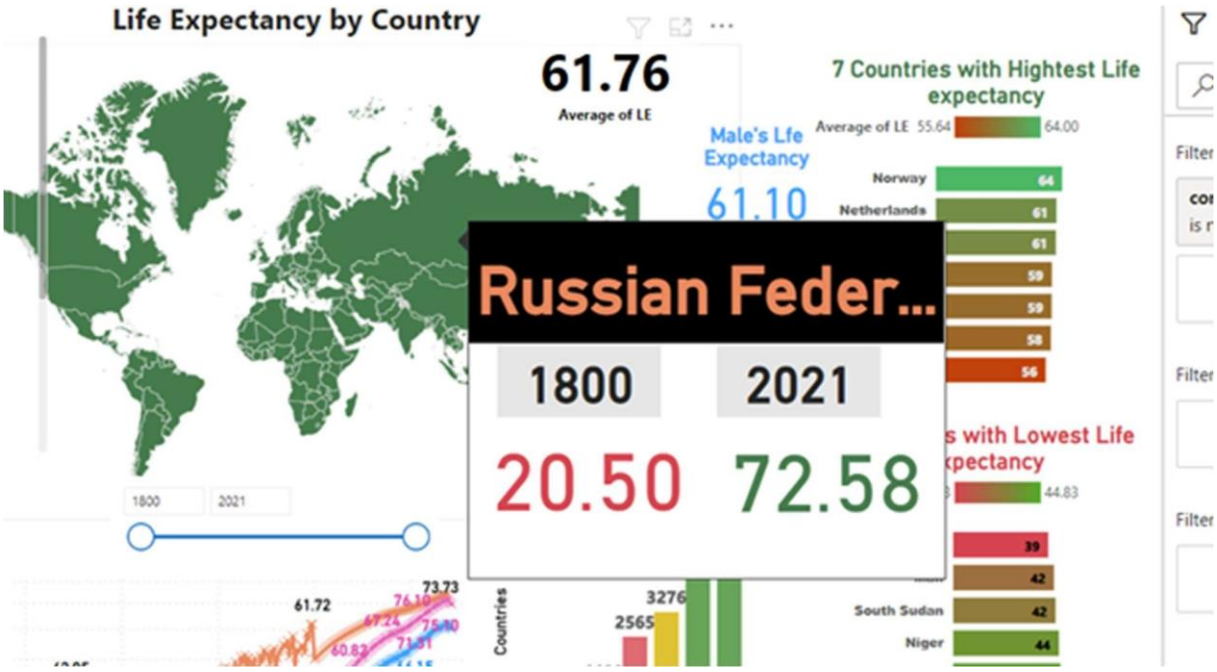



*Figure 43 insight on a country*

By clicking on the map if we select a specific country, we will get the data of this country. So the overall function is that, this dashboard will give selected countries information also. As seen above, numbers of selected country are shown.

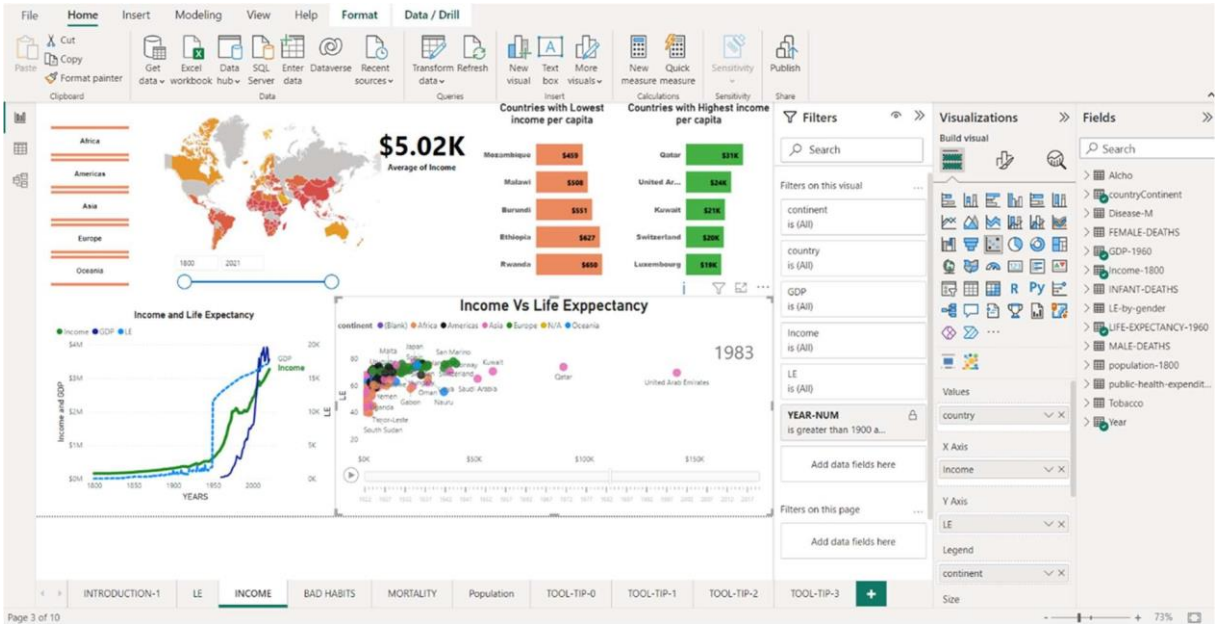



*Figure 44 Income Dashboard*

Average income of life expantancy according to this dashboard is $5.02k.

From last 2 centuries the income will show as per continent. On the lower left graph there is a graph which explains income and GDP in the x axis of graph and years in the y axis. Green line will explain the income and blue line will tell about the GDP. Due to industrial revolution in 1950 the graph of income, GDP and life expectancy increases because of the change in the industries. As seen in the figure there is a clear difference between lines of graph.

countries with highest income per capita are

- Qatar:$31k.
- UAE: $24k.
- Kuwait: $21k.
- Switzerland: $20k.
- Luxumbourg: $19k.

countries with lowest income per capita are

- Mozambique:$459k.
- Malawi: $508k.
- Burundi: $551k.
- Ethiopia: $627k.
- Rwanda: $650k.

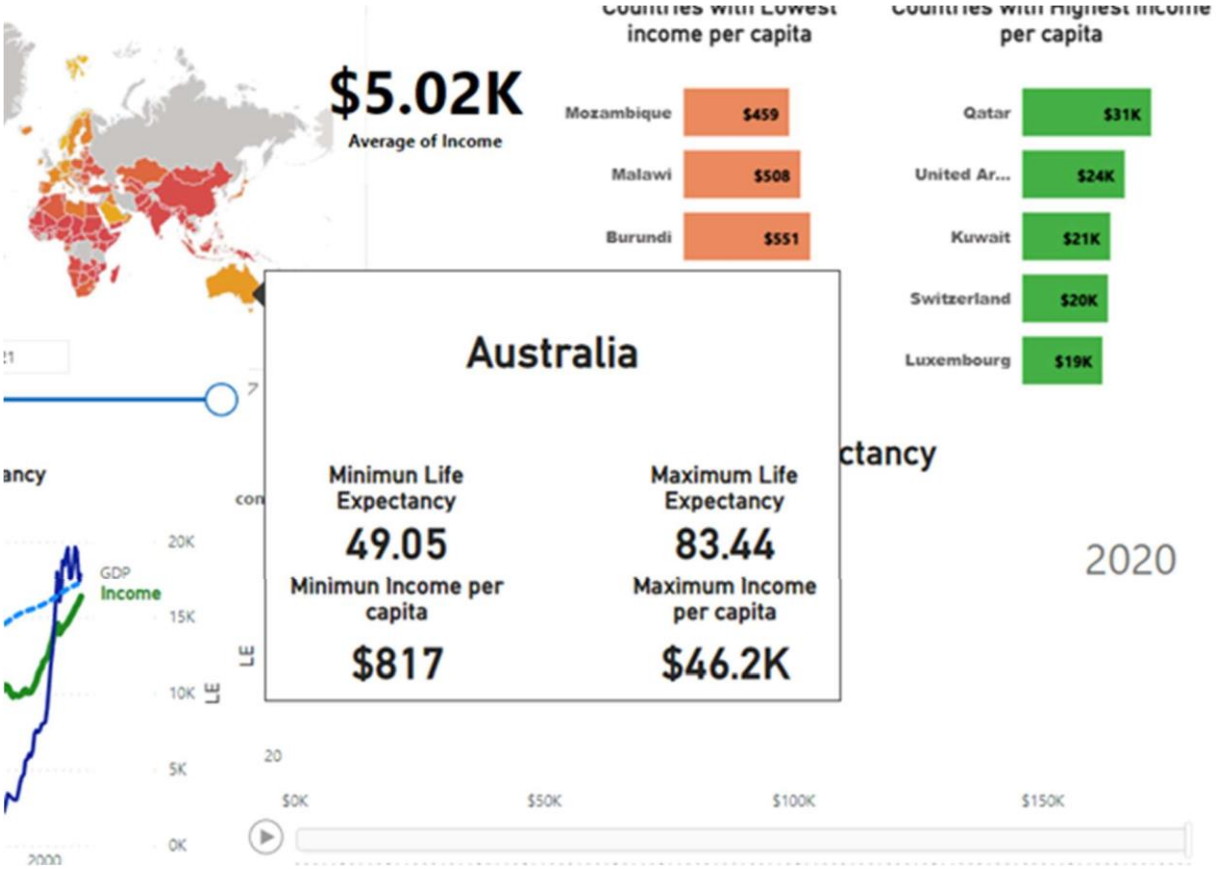



*Figure 45 income for a specific country*

There will be the Income vs Life expectancy graph which will explain the ratio of contnents ratio. It will tell us about the ratio of the continents. Same as the above graph it will give us option of finding the specific countries income per year. E.g seen in the figure, Australia's Minimum and maximum life

expectancy and Income per capita have been shown. As there is increase in both life expectancy and income per year because there is clear impact of income on the human's life expectancy.

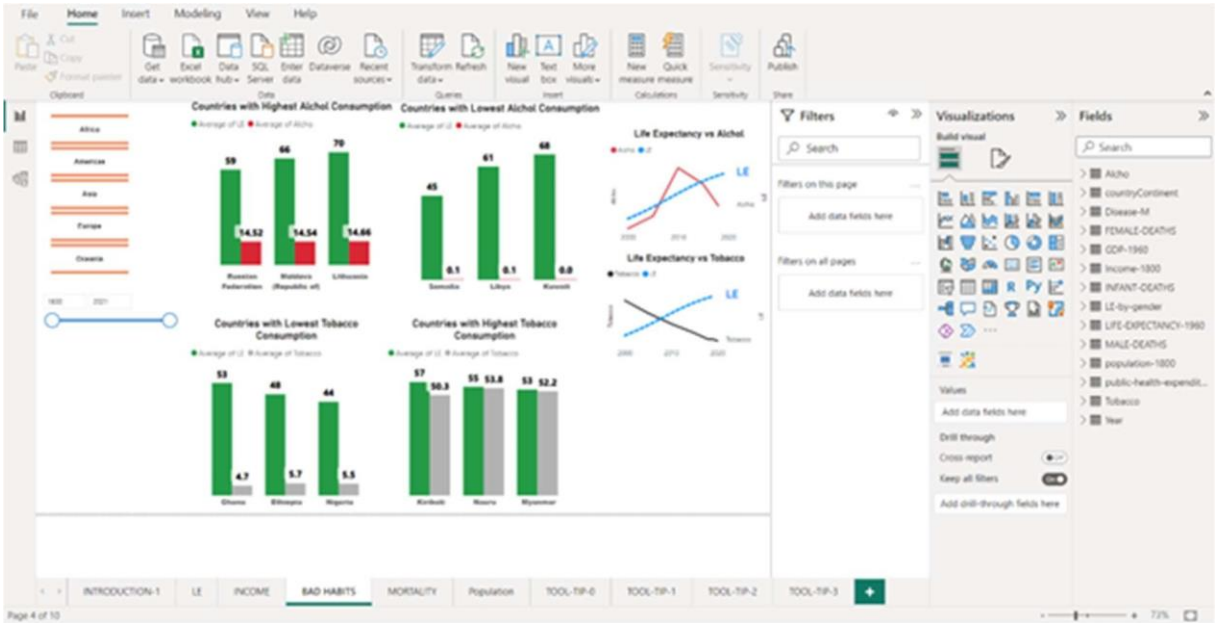



*Figure 46 Bed habit*

The dashboard in figure 44 will explain the bad habits of humans in specific ratio. As we seen consumption of alcohol is more in the European countries rather than the Asian and Arab countries.

**Countries with highest alcohol consumption, ratio by average life and alcohol consumption**

- Russian Federation : 59/ 14.52
- Moldova 66/ 14.54
- Lithuania 70/ 14.66

**Countries with lowest alcohol consumption, ratio by average life and alcohol consumption**

- Somalia: 45/0.1
- Libya: 61/0.1
- Kuwait68/0.0

**Countries with lowest Tobacco consumption, ratio by average life and alcohol consumption**

- Ghana: 53/4.7
- Ethiopia: 48/5.7
- Nigeria:44/5.5

**Countries with highest Tobacco consumption, ratio by average life and alcohol consumption**

- Kiribati: 57/50.03
- Nauru: 55/53.8

Myanmar: 53/52.2

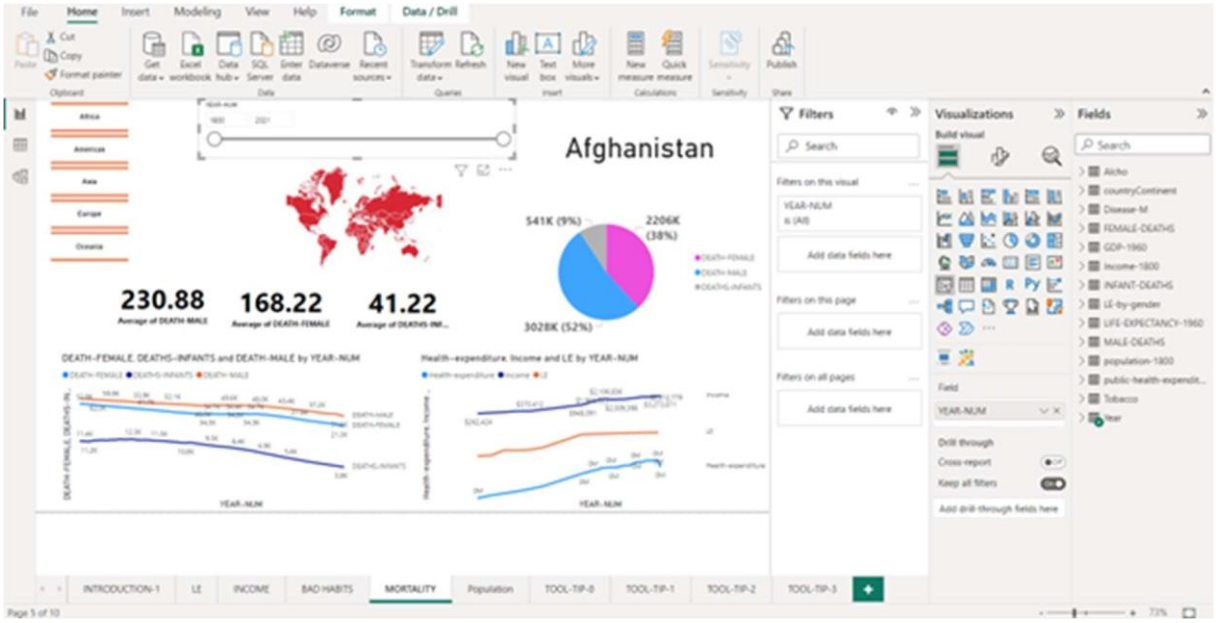



*Figure 47 mortality dashboard*

In this figure 45 we are going to discuss about the death ratio regarding life expectancy. Average death rate of following

- Death-Male:230.08
- Death-Female:168.22
- Death-Infants:41.22

Graphs on lower side of figure 45 show death rate of male, female, infants by year. Blue line here shows the Death-Female, Purple lines show Death-Infants, Orange line will show death-male. On x-axis it shows the death rate and on y-axis it will display the year number.

While on other graph, where we compare health expenditure, income, and LE By year number. We can see the difference of both graphs as health expenditure increase there is an increase in the life expectancy.

In this figure 45 it also shows the death rate, income GDP and life expectancy. In the left top corner, it gives us the information about the continent's death rate according to the health expenditure and income GDP rate.

**Death-Male**:

Color: blue

3028k (52%)

**Death-Female**:

Color: pink

2206k (38%)

**Death-Infants**:

Color: grey

541k (9%)

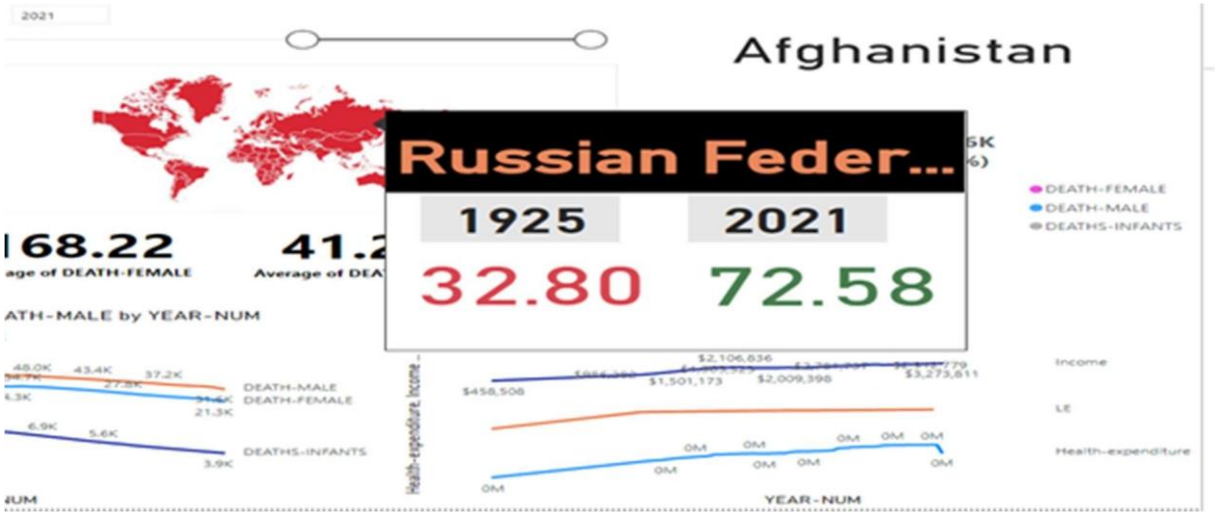



*Figure 48 data shown for Russia*

In the figure 46, when we click on specific part in map it will show the record of selected country, as it shows the life expectancy and health expenditure and death rate. As numbers are appeared of Russian Federation which show the data of death rates.

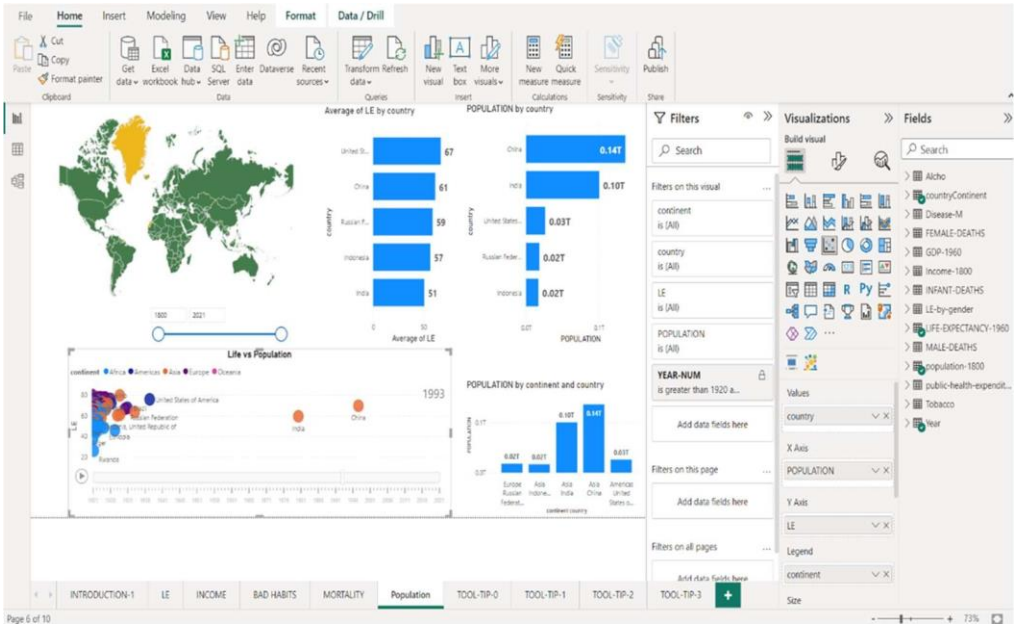



*Figure 49 Population dashboard*

In the above figure 47, there is a graph shown named as Life vs Population. It will show the fluctuation in numbers of different continent. As seen in the graph the highest population is in China, but it will not decide the life rate. It may be different from the population graph.

Africa: blue

Americans: Dark Blue

Asian: Orange.

Europe: Purple

Oceania: pink

There is another graph on lower right, where it shown as population by continent and country. On x - axis it shown as the Population and on y-axis it shows the continental country.

**Average LE by country**

- US :67
- China :61
- Russian: 59
- Indonesia: 57
- India :51

**Population by country**

- China :0.14T
- India: 0.10T

- US :0.03T
- Russian: 0.02T
- Indonesia: 0.02T

**Note: the will is shown in 0. trillion something because we are combining the overall population of 220 years dashboard is interactable if you on adjust the year btw 2019 and 2021 you will get the current information on population etc.**

# References


Anon., n.d. *DataBank World Development Indicators.* [Online]
Available at:
https://databank.worldbank.org/reports.aspx?source=2&series=SP.POP.TOTL.FE.IN&country=#
[Accessed 19 November 2022].

by Max Roser, E. O.-O. a. H. R., October 2019. *Our World in data.* [Online]
Available at: October 2019
[Accessed 12 November 2022].