<u>LIFE EXPECTANCY DATA ANALYSIS – 2000 TO 2015</u>

CALIFORNIA STATE UNIVERSITY LOS ANGELES

SUBMITTED BY   **SUBMITTED TO:**

POOJA MADHUP   **Prof. SHILPA BALAN**

TABLE OF CONTENTS

A. __INTRODUCTION:__

Living conditions are a universal issue, with varying degree. For countries to focus their support and resources, it is imperative to identify the factors which are primarily affecting health standards. The Life expectancy Dataset originated from the Global Health Observatory which monitors the health status of the countries. The United Nations website was used to complete the economic data. Initially, the observation show that 'Developed' countries have better outcomes across most attributes in the dataset and have a reduced healthcare expenditure as this may be a result of a larger GDP. There are various attributes to analyze within the dataset. There are some of the primary one which will be used throughout the analysis such as status, Life expectancy, Infant deaths, and under-five deaths, Body Mass Index (BMI), Hepatitis B, Polio, and HIV/AIDS and so on. Based on this factor, we visualize the data of various countries over the 16 years from 2000 to 2015. In other words, we will study on the factors that are influencing the life expectancy globally. From this project, we will find out which country have highest life expectancy and lowest life expectancy. Moreover, the comparison between developed or developing countries in terms of life expectancy. We chose this topic because we wanted to know more about the health status of people with respect to their country. We would like to explore of the life expectancy of all the countries around the world with the help of world health organizations. The correlation of alcohol, polio, and AID/HIV, Hepatitis B with the life expectancy. Learning is a fundamental process of life, which has affected the past, present and will affect the future. It is important that we understand what variables change the way we learn, and for how long. We will investigate if there is a strong correlation between the economic output of a country and the life expectancy of its citizens. During this project, we will clean the data, analyze the data, prepare, and plot data, and seek to answer questions in a meaningful way. We will write the storytelling at the end.

B. DATASET DESCRIPTION:

DATASET URL:

https://www.kaggle.com/

https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who

WEBLINK:

Life Expectancy by Country and in the World (2022) - Worldometer (worldometers.info)

GHO | By category | Life expectancy and Healthy life expectancy - Data by WHO region

The Life Expectancy Dataset has over 16 years of all the countries around the world data extracted from the World Health Organization (WHO) data repository for the years 2000 to 2015. There are many factors in the dataset to get the life expectancy in percentage (2000-2015). Out of 193 countries, we analysis 104 countries due to the missing data in their most of the columns. Therefore, we analyze near about 1700 rows out of 2900 rows. Each life expectancy of a person has the following dataset attributes:

S.No.	Columns	Data description
1	Country (String)	It represents the name of the country
2	Year (Integer)	It represents the Year
3	Status (String)	It represents the country status such as developed or developing.
4	Life Expectancy (Integer)	It represents the age of a person in years
5	Diphtheria (Integer)	Diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year old (%)
6	Infant deaths (Integer)	Number of Infant deaths rate per 1000 population
7	Alcohol (Integer)	Alcohol intake recorded per capita (15+) consumption in liters
8	Percentage expenditure (Decimal)	It represents expenditure on health as a percentage of Gross Domestic product per capita
9	Hepatitis B (Integer)	Hepatitis B (HepB) immunization coverage among 1-year old in percentage
10	Measles (Integer)	Number of reported cases per 1000 population

11	BMI (Decimal)	Average Body Mass Index of the entire population
12	Under-five deaths (Integer)	Number of under-five deaths per 1000 population
13	Polio (Integer)	Polio immunization coverage among 1-year old in percentage
14	Total expenditure (Decimal)	It represents the general government expenditure on health as a percentage of total government expenditure.
15	Diphtheria (Integer)	Diphtheria tetanus toxoid and pertussis immunization coverage among 1-year old in percentage
16	HIV/AIDS (Decimal)	Number of deaths per 1000 live birth of HIV/AIDS over 0-4 years
17	GDP (Decimal)	Gross Domestic Product per capita in USD dollars. It shows how an increase in life expectancy is accompanied by an increase in GDP per capita income.
18	Population (Integer)	Population of the country
19	Thinness 1-19 years (Integer)	Prevalence of thinness among the children and adolescents for age group of 10 to 19 years in percent
20	Thinness 5.9 years (Integer)	prevalence of thinness among the children for age group of 5 to 9 years in percent
21	Income composition of resources (Integer)	Human Development Index as a income composition of resources ranging from 0 to 1
22	Schooling (Integer)	Number of years of schooling. This column explain how education affects life expectancy.

C. DATA CLEANING

1.) Missing Values

Pre-Cleaning

The column 'Alcohol' value is not present for the year '2015' for country 'Burundi' for 2015 years. For analyzing the life expectancy, we need the missing values of the alcohol for the selected year and country details.

1	Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol
417	Burkina Faso	2000	Developing	51	348	48	3.71
418	Burundi	2015	Developing	59.6	288	21	
419	Burundi	2014	Developing	59.1	297	22	0.01
420	Burundi	2013	Developing	58.6	33	22	0.01
421	Burundi	2012	Developing	58	312	22	0.01
422	Burundi	2011	Developing	57.4	321	22	Missing value 16
423	Burundi	2010	Developing	56.8	33	23	4.16

Post-cleaning

The values for Burundi country in the column Alcohol entered by referring to the WHO data for 2015. We searched the alcohol value for the highlighted country with the particular year from the World Health Organization website.

	A	B	C	D	E	F	G
1	Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol
417	Burkina Faso	2000	Developing	51	348	48	3.71
418	Burundi	2015	Developing	59.6	288	21	0.02
419	Burundi	2014	Developing	59.1	297	22	0.01
420	Burundi	2013	Developing	58.6	33	22	0.01
421	Burundi	2012	Developing	58	312	22	0.01
422	Burundi	2011	Developing	57.4	321	22	4.16

2.) **Illegal Values**

Pre-cleaning

Changing special characters to alphabets in the Country column 'CÃ´te d'Ivoire' from 2000 to 2015 year. These are also known as Ivory Coast. We replace this special character name to another name from the website listed in reference.

	A	B	C
1	Country	Year	Status
432	Burundi	2001	Developing
433	Burundi	2000	Developing
434	CÃ´te d'Ivoire	2015	Developing
435	CÃ´te d'Ivoire	2014	Developing
436	CÃ´te d'Ivoire	2013	Developing
437	CÃ´te d'Ivoire	2012	Developing
438	CÃ´te d'Ivoire	2011	Developing
439	CÃ´te d'Ivoire	2010	Developing
440	CÃ´te d'Ivoire	2009	Developing

Changing the special character into alphabet

Post-cleaning

Country Field - CÃ´te d'Ivoire has now been replaced with 'Ivory Coast' for 16 years from 2000 to 2015 years. The updated name is displayed on the below show with the highlighted section.

3.) **Inconsistent Year Format**

Pre-cleaning

The column 'Year' consist of inconsistent year formats for Benin country, the column year is showing '5/12/2005' for the Benin country with its status and other details which is not a correct format. We changed it to year 2005 from the incorrect way to a specific data.

Post-cleaning

As shown below the year format for Benin country with their country status have been changed to '2005' and made consistent throughout the dataset. The year format is displaying correctly under the year column.

4.) **Misspelling Name**

Pre-cleaning

Spelling of Country name 'Sudan' is incorrect for the 2015 year. As compared to the other years from 2000 to 2014 for the same country. We noticed the country name for 2015 year does not have same spelling. So, we would like to correct the name.

1	Country	Year	Status	Life expec	Adult Moi	infant dea
1379	Sri Lanka	2001	Developing	72.7	16	5
1380	Sri Lanka	2000	Developing	71.5	175	5
1381	Sudanaya	2015	Developing	64.1	225	58
1382	Sudan	2014	Developing	63.8	229	59
1383	Sudan	2013	Developing	63.5	232	60
1384	Sudan	2012	Developing	63.2	235	61
1385	Sudan	2011	Developing	62.7	241	61
1386	Sudan	2010	Developing	62.5	243	62
1387	Sudan	2009	Developing	62	248	63
1388	Sudan	2008	Developing	61.8	251	64
1389	Sudan	2007	Developing	61.4	254	65
1390	Sudan	2006	Developing	61	26	66
1391	Sudan	2005	Developing	67	261	66
1392	Sudan	2004	Developing	59.7	278	68
1393	Sudan	2003	Developing	59.6	278	69
1394	Sudan	2002	Developing	59.4	277	70
1395	Sudan	2001	Developing	58.9	283	71
1396	Sudan	2000	Developing	58.6	284	71

(annotation: **Spelling Incorrect** with upward arrow pointing to row 1381)

Post-cleaning

After the post-cleaning, the spelling of country 'Sudan' is correctly displayed for 2015 year. From 2000 to 2015 years, all the country name as Sudan is displaying as below. The country name should be displayed same over the 16 years.

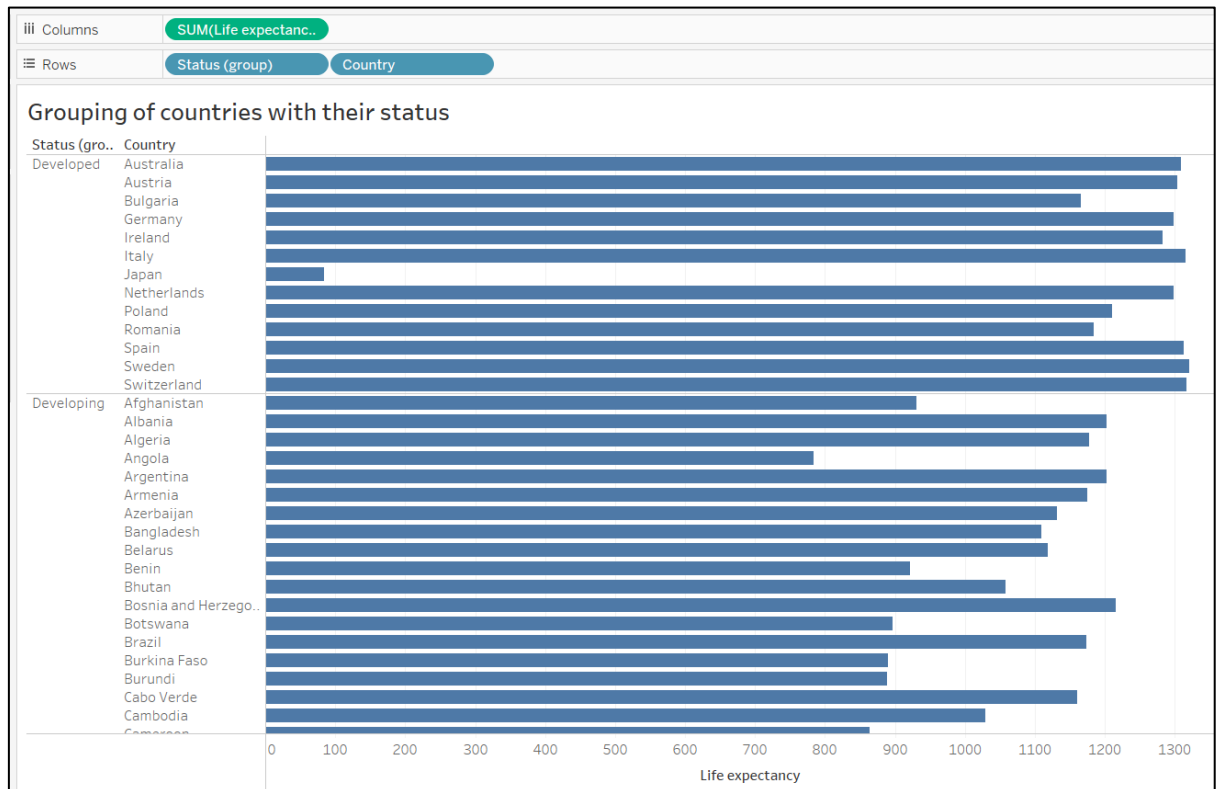| 1	Country	Year	Status	Life expe	Adult Mo	infant dea
1378	Sri Lanka	2002	Developing	73.7	149	5
1379	Sri Lanka	2001	Developing	72.7	16	5
1380	Sri Lanka	2000	Developing	71.5	175	5
1381	Sudan	2015	Developing	64.1	225	58
1382	Sudan	2014	Developing	63.8	229	59
1383	Sudan	2013	Developing	63.5	232	60
1384	Sudan	2012	Developing	63.2	235	61
1385	Sudan	2011	Developing	62.7	241	61
1386	Sudan	2010	Developing	62.5	243	62
1387	Sudan	2009	Developing	62	248	63
1388	Sudan	2008	Developing	61.8	251	64
1389	Sudan	2007	Developing	61.4	254	65
1390	Sudan	2006	Developing	61	26	66
1391	Sudan	2005	Developing	67	261	66
1392	Sudan	2004	Developing	59.7	278	68
1393	Sudan	2003	Developing	59.6	278	69
1394	Sudan	2002	Developing	59.4	277	70
1395	Sudan	2001	Developing	58.9	283	71
1396	Sudan	2000	Developing	58.6	284	71

5.) **Group**

Pre-cleaning

For categories the given countries, we need to group the countries with their developed and developing status for the further visualization. In the below chart, the status of country list is displaying for every country. For that, we will need country list under the developed and developing status.
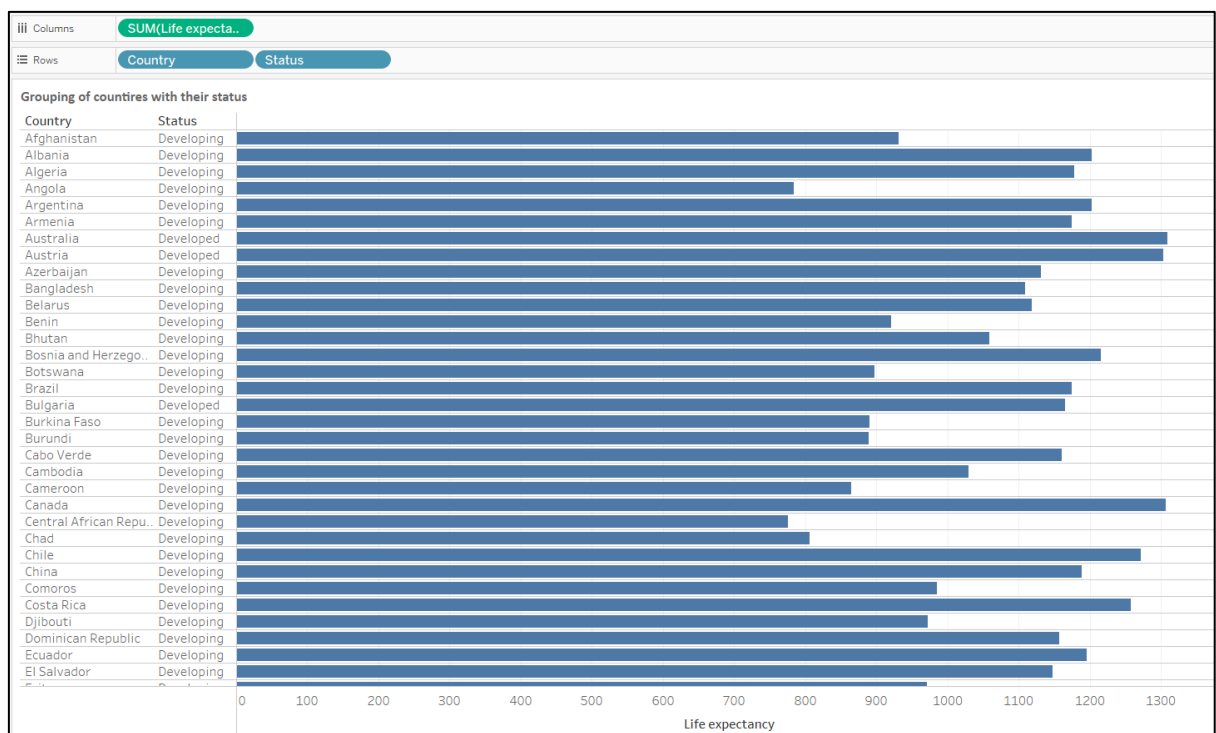
Post-cleaning

In the post-cleaning the countries are displayed group by their developed and developing

status for analyzing the life expectancy of the given data. Therefore, for more clarity for

their visualization, we have done the grouping of status for the given country list.

6.) **Duplicate Records**

Pre-cleaning

The Canada country for the year 2004 consists of duplicate dataset as shown in below.

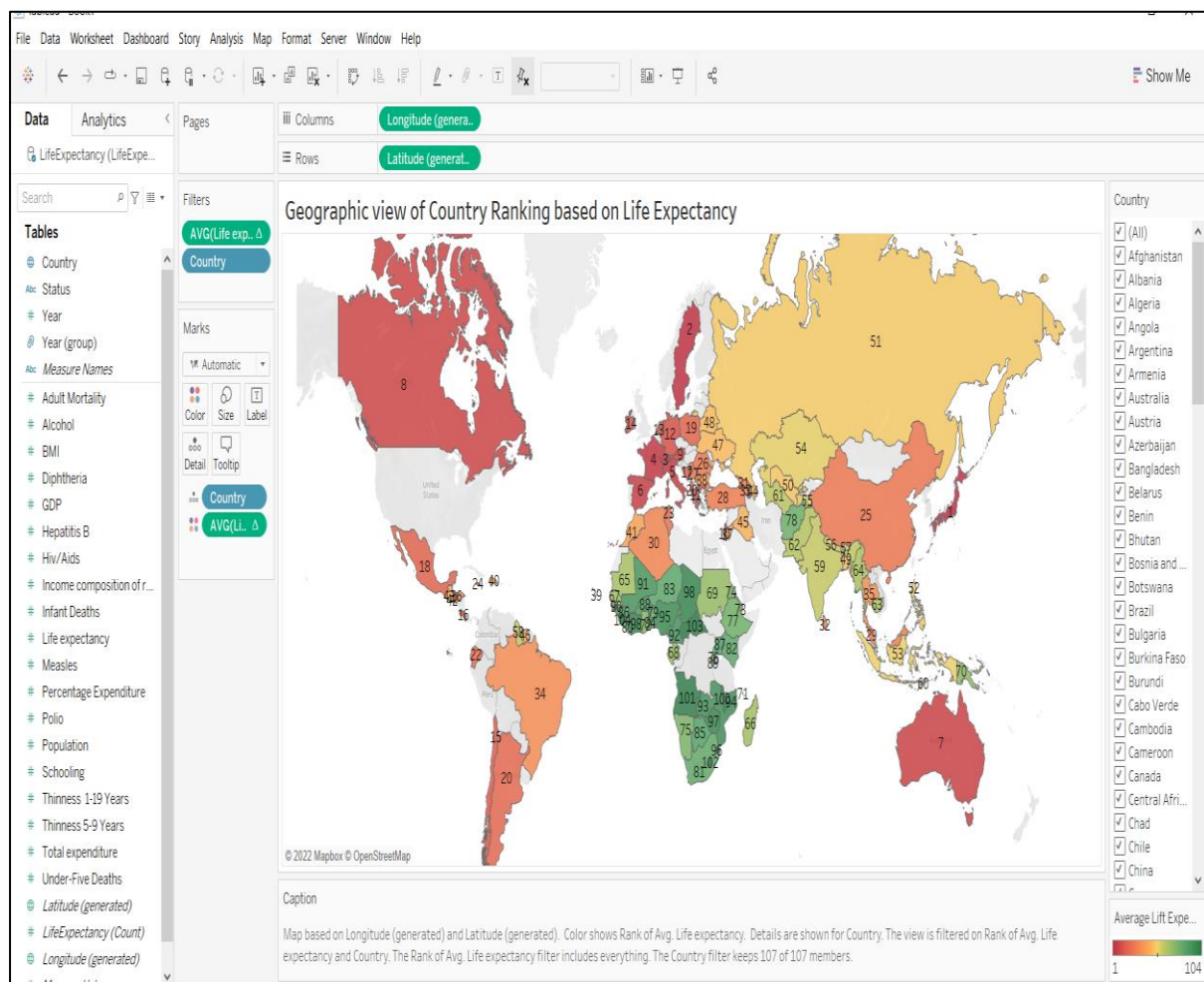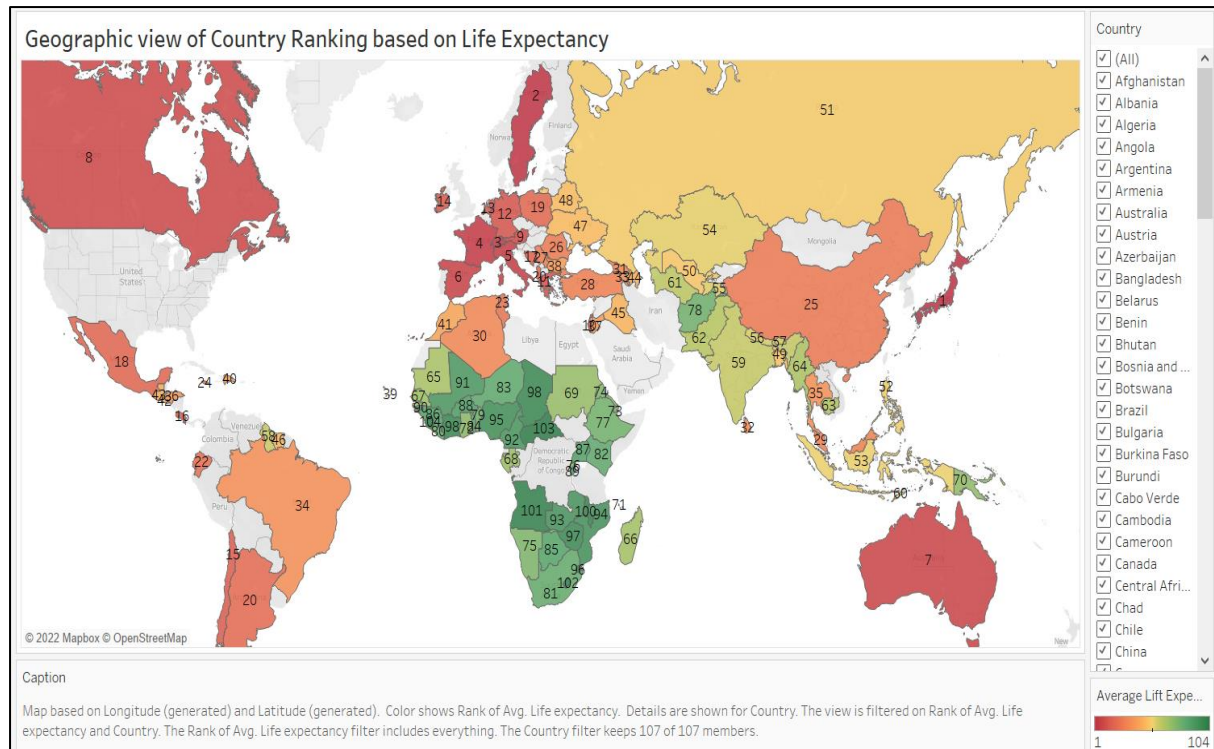This dataset creates confusion for analyzing the data.

Post-cleaning

Rows for the Canada country with duplicate values for the year 2004 have been discarded

after post-cleaning, it removes redundancy from the analysis.

	A	B	C	D	E	F	G	H
1	Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure
507	Canada	2006	Developing	85	75	2	8.2	771.1817686
508	Canada	2005	Developing	81	76	2	8	6333.17796
509	Canada	2004	Developing	80	77	2	7.8	5513.32992
510	Canada	2003	Developing	79.7	78	2	7.7	4687.84556
511	Canada	2002	Developing	79.5	79	2	7.7	3895.856308
512	Canada	2001	Developing	79.4	8	2	7.6	3665.089703
513	Canada	2000	Developing	79.1	82	2	7.6	3787.49456

D. DATA VISUALIZATION

1.) What is the country ranking based on the life expectancy?

Geographic view of Country Ranking based on Life Expectancy

Caption

Map based on Longitude (generated) and Latitude (generated). Color shows Rank of Avg. Life expectancy. Details are shown for Country. The view is filtered on Rank of Avg. Life expectancy and Country. The Rank of Avg. Life expectancy filter includes everything. The Country filter keeps 107 of 107 members.

[Application used: Geographic Maps & Rank]

This visualization illustrates the ranking of the countries based on the average life expectancy for the population. This is demonstrated with the help of intensity of color code given to each country. It also shows the rank in numbers annotated for each country. Moreover, some countries do not have any color coding because we do not select that country. We can see that the life expectancy of Japan is the highest among all. The Sweden country's population holds the second highest life expectancy age in years. Switzerland is in 3rd position; 4th is the France and Italy is the 5th country where people live more than other country population, while the Sierra Leona population has lowest life expectancy based on many different factors. This visualization helps us understand the average living age of the given countries globally from the year 2000 to 2015 respectively. Furthermore, using this map, we can find out the average life expectancy of any country with their ranking. This visualization also helps us to see about the healthy living population of a country.
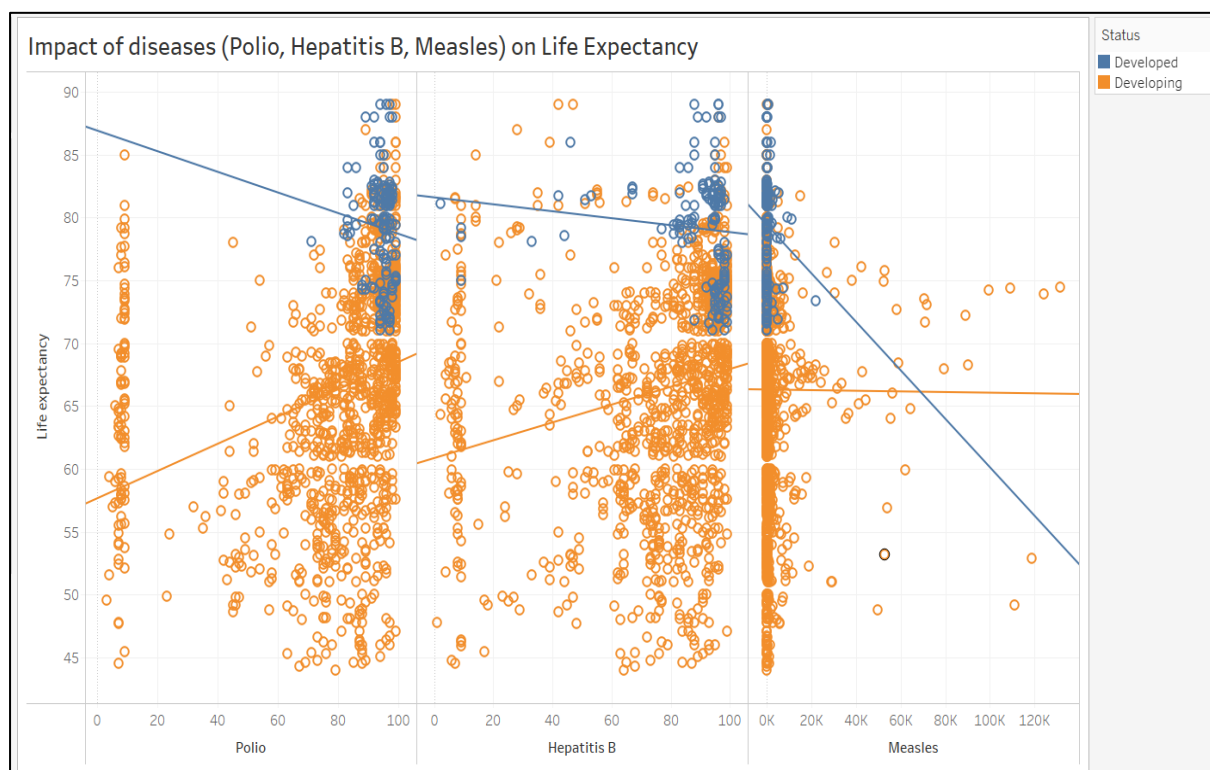
2.) <u>**How the Average Life Expectancy and average GDP shows over the year for**</u>

<u>**country status?**</u>
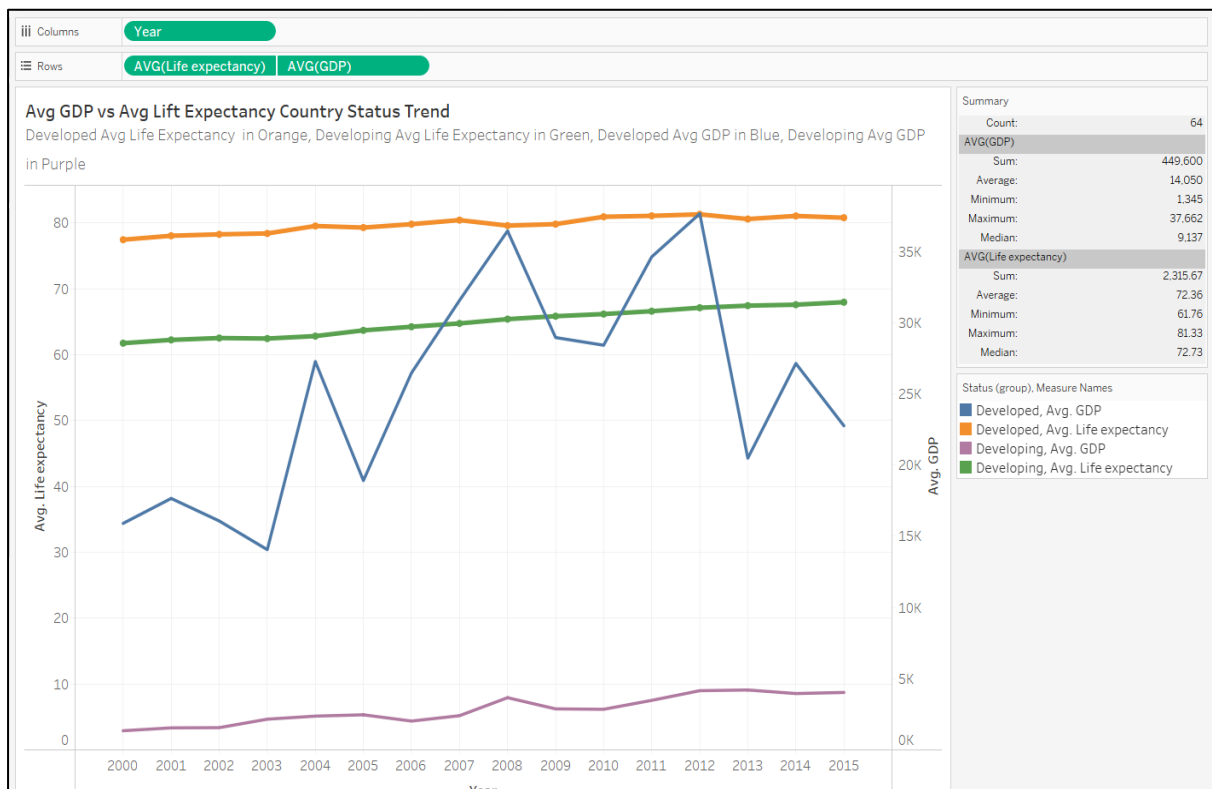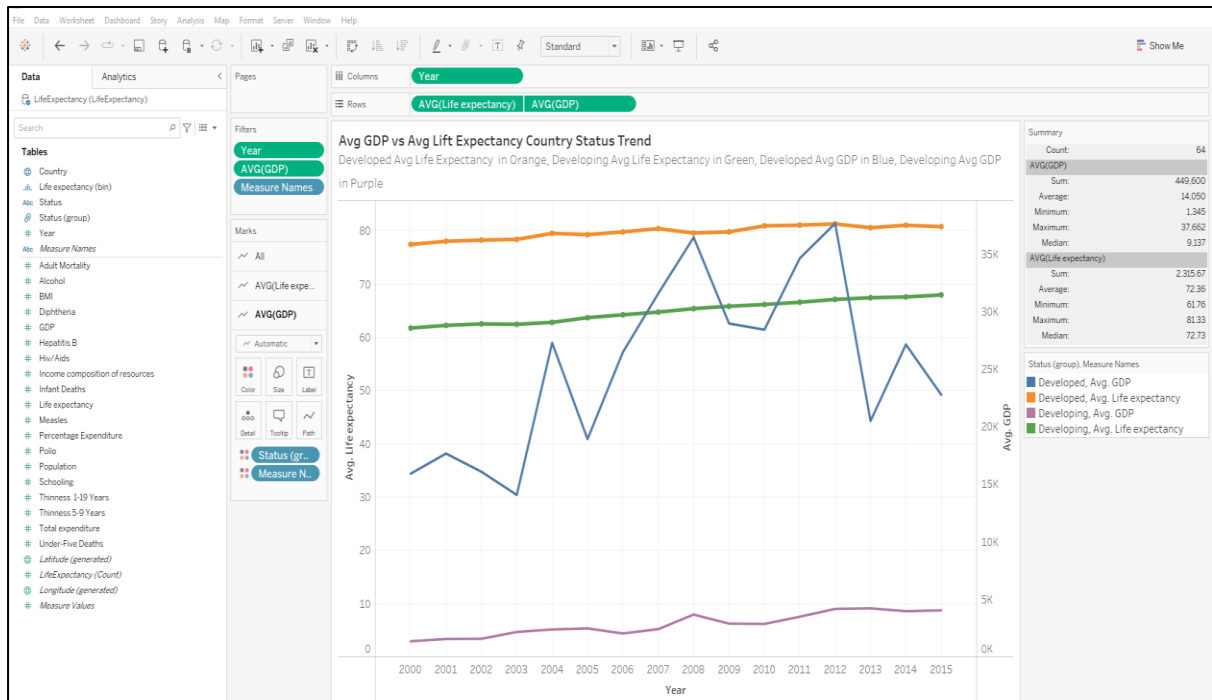
[Application used: Trend Line & Dual Axis chart]

This visualization shows the average GDP (Gross Domestic Product) and average life expectancy over the year of the developed and developing countries. We observe that the correlation between the average GDP over the years with respect to average life expectancy and observe the data trends using line plot. We are using average life expectancy and average GDP for the analysis. We also grouped the country list by their status, so that the curve would be clearer for analyze the given data. For the upper visualization, we can evaluate that average life expectancy of developed countries are higher than the average life expectancy of developing countries with the color coding over the year. Similarly, for the lower visualization, the average GDP of developed countries are fluctuated but higher on comparing with the average GDP of developing countries from 2000 to 2015 years. We can view the average life expectancy and average GDP based on the color coding on the right-hand side of the chart. Such as the dark blue color line indicate the developed countries average GDP while the purple color indicates the developing countries average GDP. Similarly, for average life expectancy, orange color represents the developed countries and green color represents the developing countries for the same. Based on the above information, we conclude that the average life expected with average GDP of developed countries are more increased year by year. The effect of GDP is also impacted on the life expectancy as the developing countries is less than the developed countries GDP. Also, the income ratio of developed countries is more in term of industries, business and so on, that's why the GDP of the developed countries are greater than the developing countries.

3.) **What is the vital impact of the Polio, Hepatitis B, Measles on Life Expectancy view?**

[Application used: Scatter Plot, Trend Line & Marks filter]

This visualization shows that the average life expectancy of population who have polio or

Hepatitis B or Measles with their country status (Developed or Developing). From the above

diagram, we can describe that if the developing country people are suffering more polio, hepatitis B and Measles than the developed country population. However, the average life expectancy is still more for the developed countries while comparing with the developing countries because developed countries have many medical facilities than developing country. Hence, this scatter plot shows that the average life expectancy is inversely proportional to the diseases which means that if the developing country people suffered more with the polio or hepatitis B or Measles than their life expectancy will be less. Similarly, if the developed country population suffered less with the given diseases than they are healthier and live more. We can find the highest and lowest life expectancy with the help of best fit line where R squared value and P value for different diseases. Also, while taking curser on any bubble they will give the details with the country name, disease details, life expectancy and their country status.

4.) <u>How the BMI with different countries and their status shows?</u>

Box & Whisker plot for BMI with different countries and their status
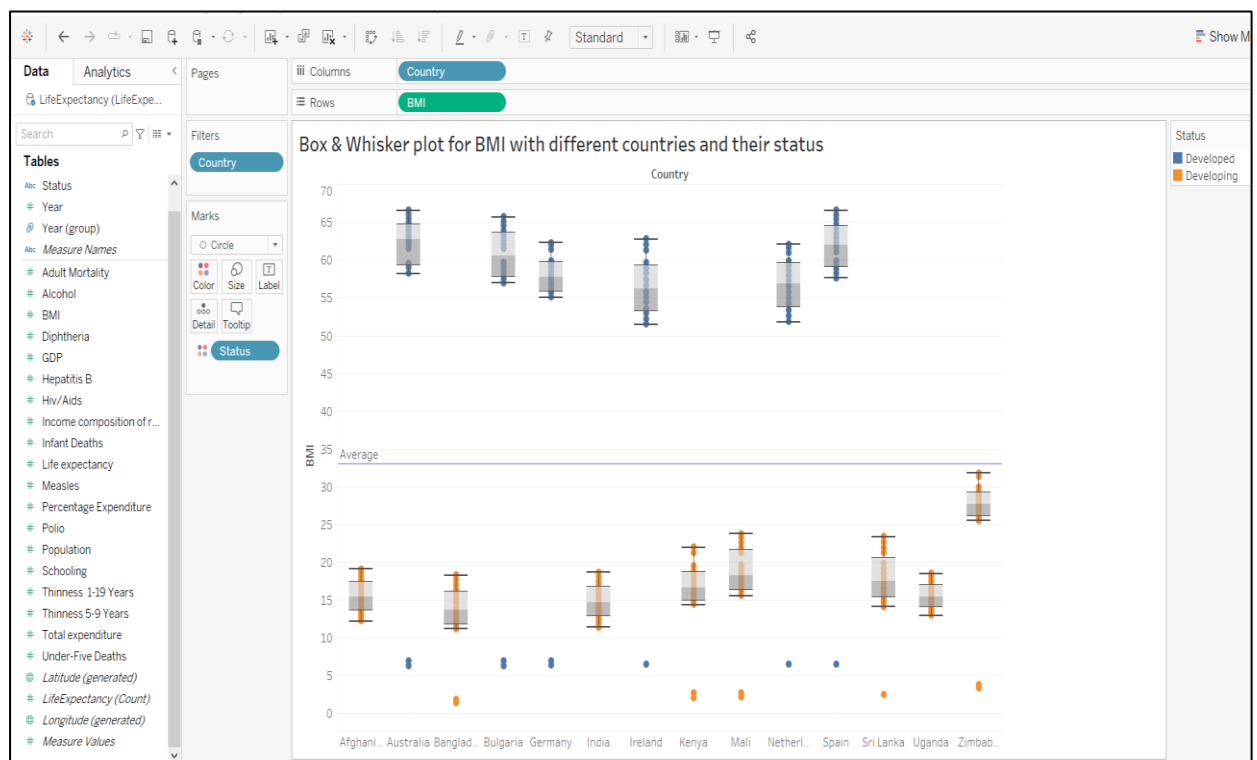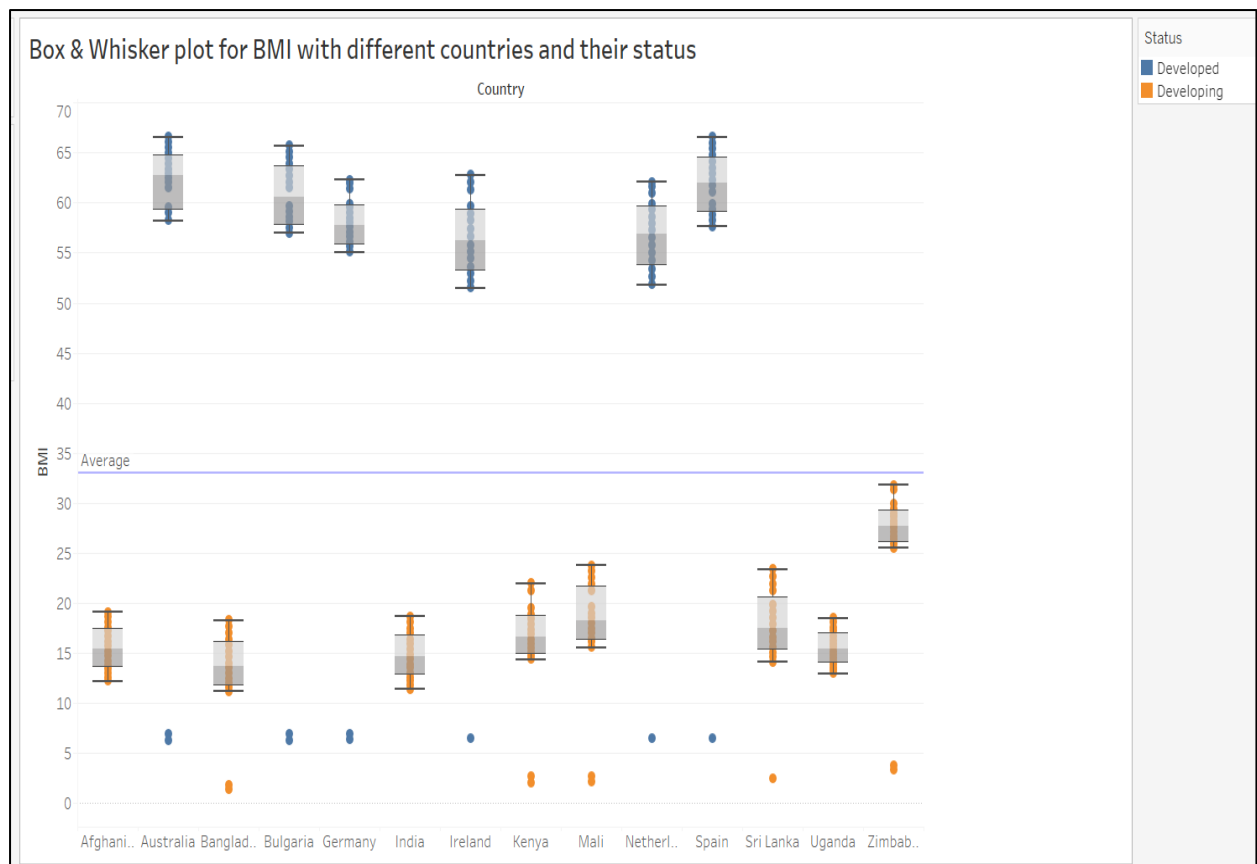
[Application used: Box and Whisker plot, Reference Line & Filters]

This visualization illustrates that the selected country wise Body Mass Index (BMI) and their country status using Box & Whisker plot. We selected country using the edit filter option on the right click on the country. The developed country and developing country status are differentiate by the different color on the right-hand side. There is a reference line which indicate the average value of body mass index over the fourteen countries. The plot shows that the developed countries have higher body mass index while the developing countries has lower body mass index. For example, the Australia is a developed country having higher body mass index is 66.60. On taking curser on the lower of the plot, it will give the detail about the upper whisker, upper hinge, median of the body mass index, lower hinge, lower whisker respectively. Australia, Bulgaria, Germany, Ireland, Netherland, Spain country population has more than average BMI while the Afghanistan, Bangladesh, India, Kenya,

Mali, Sri Lanka, Uganda, and Zimbabwe people has less body mass index. In other words,

wealthy countries have higher body mass index than the poor countries.

5.) <u>How we illustrate the distribution of life expectancy?</u>

Distribution of frequency of Life Expectancy
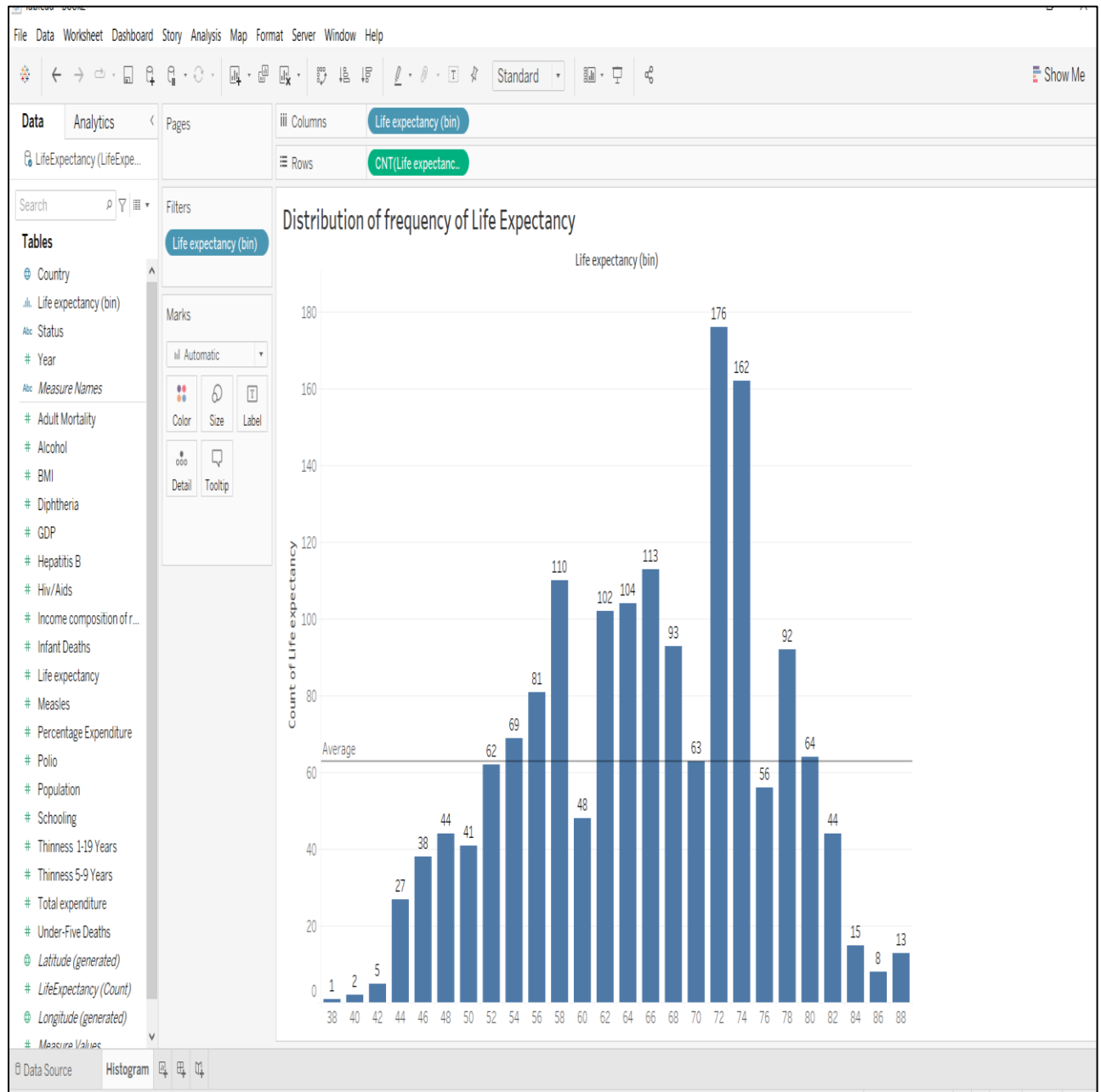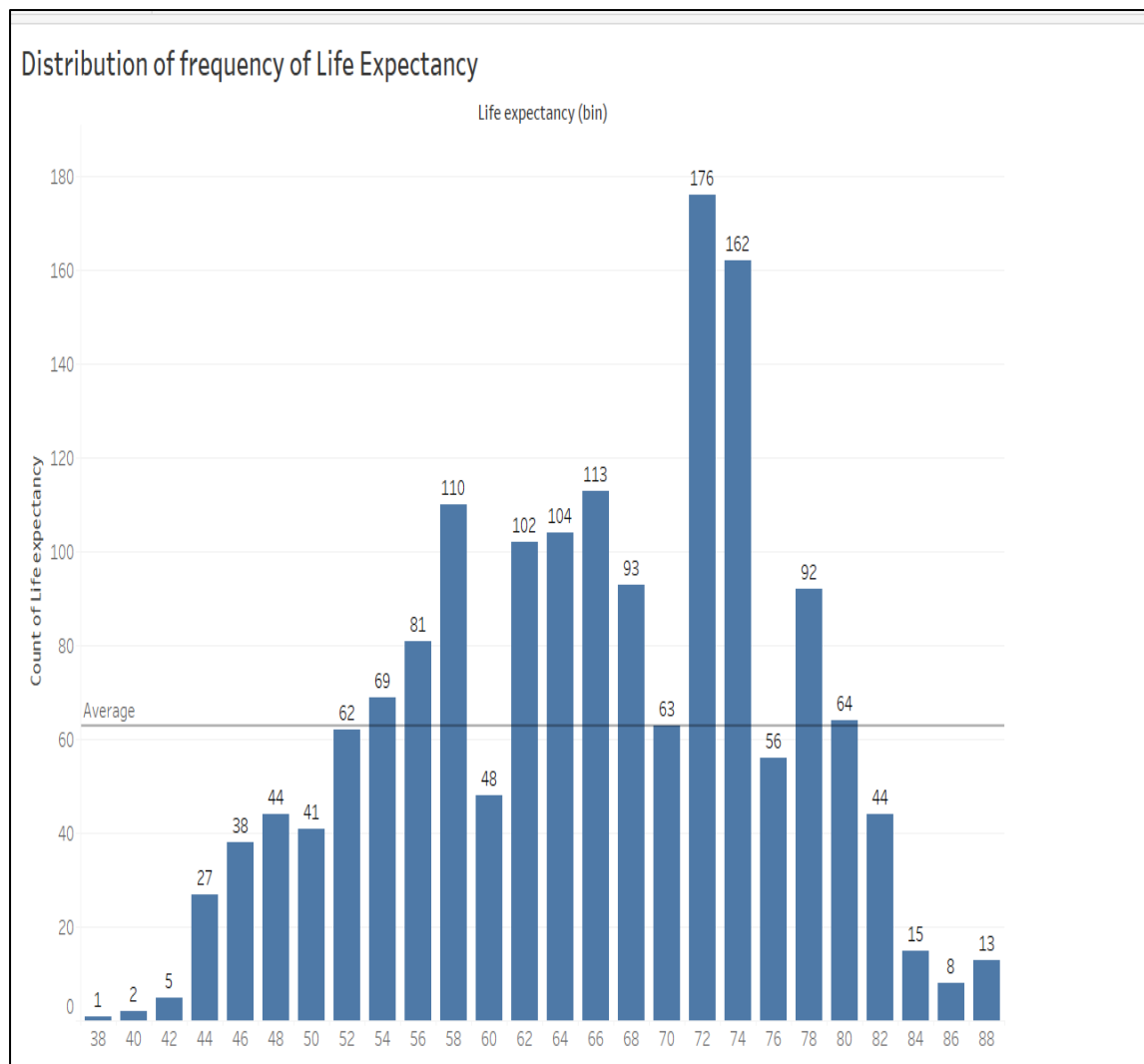
[Application used: Histogram, Reference Line & Filters]

This visualization shows that the distribution of count of life expectancy using Histogram. We create life expectancy (bin) using the create / bin option on the click of selected life expectancy. Meaning, the count of life expectancy of the given data with their age in years. For instance, 176 is the highest count of life expectancy for the 72 years age group for the given countries and 1 is the lowest count of life expectancy for the age of 38 years respectively. The average count of life expectancy is shown by the reference line. The count of life expectancy is shows in the bar type of chart. The count of life expectancy needed when we analyze life expectancy worldwide.

6.) **DASHBOARDS**
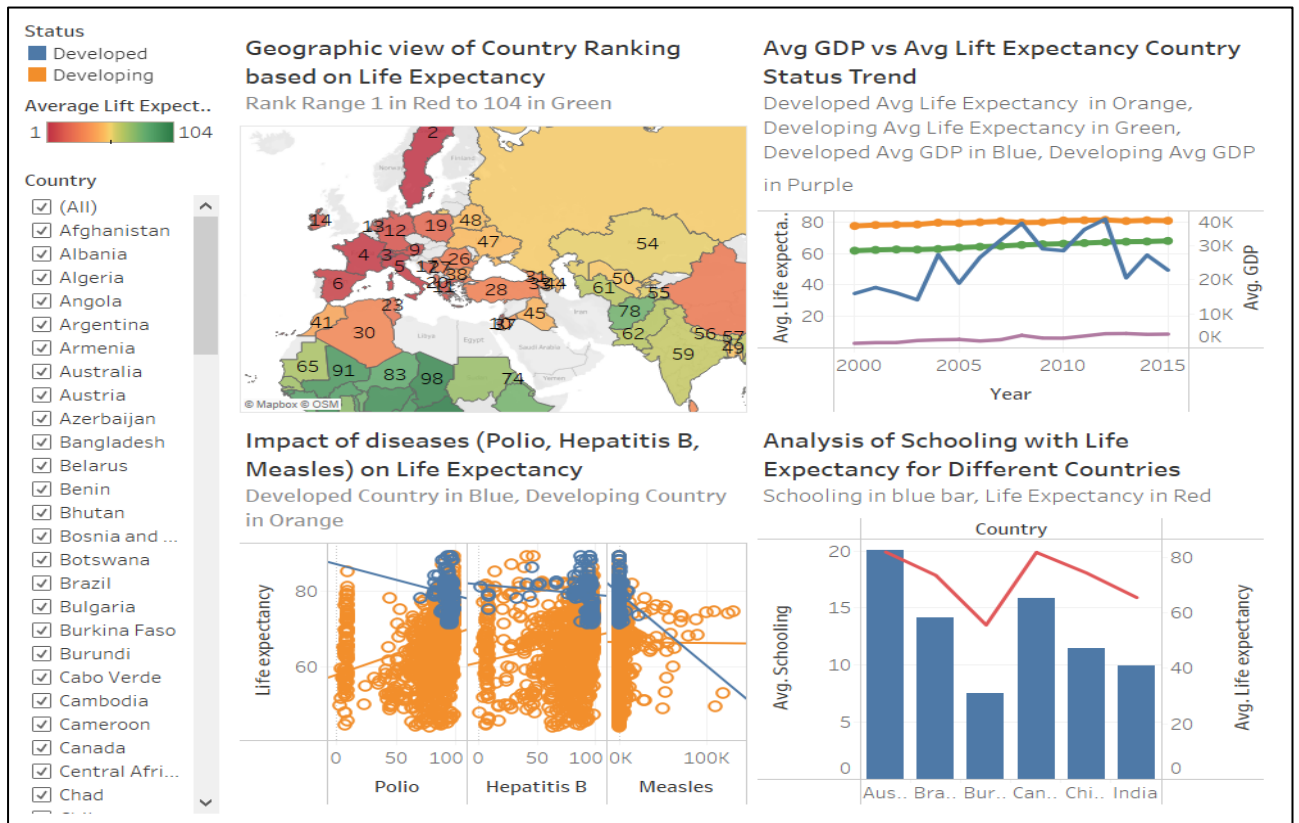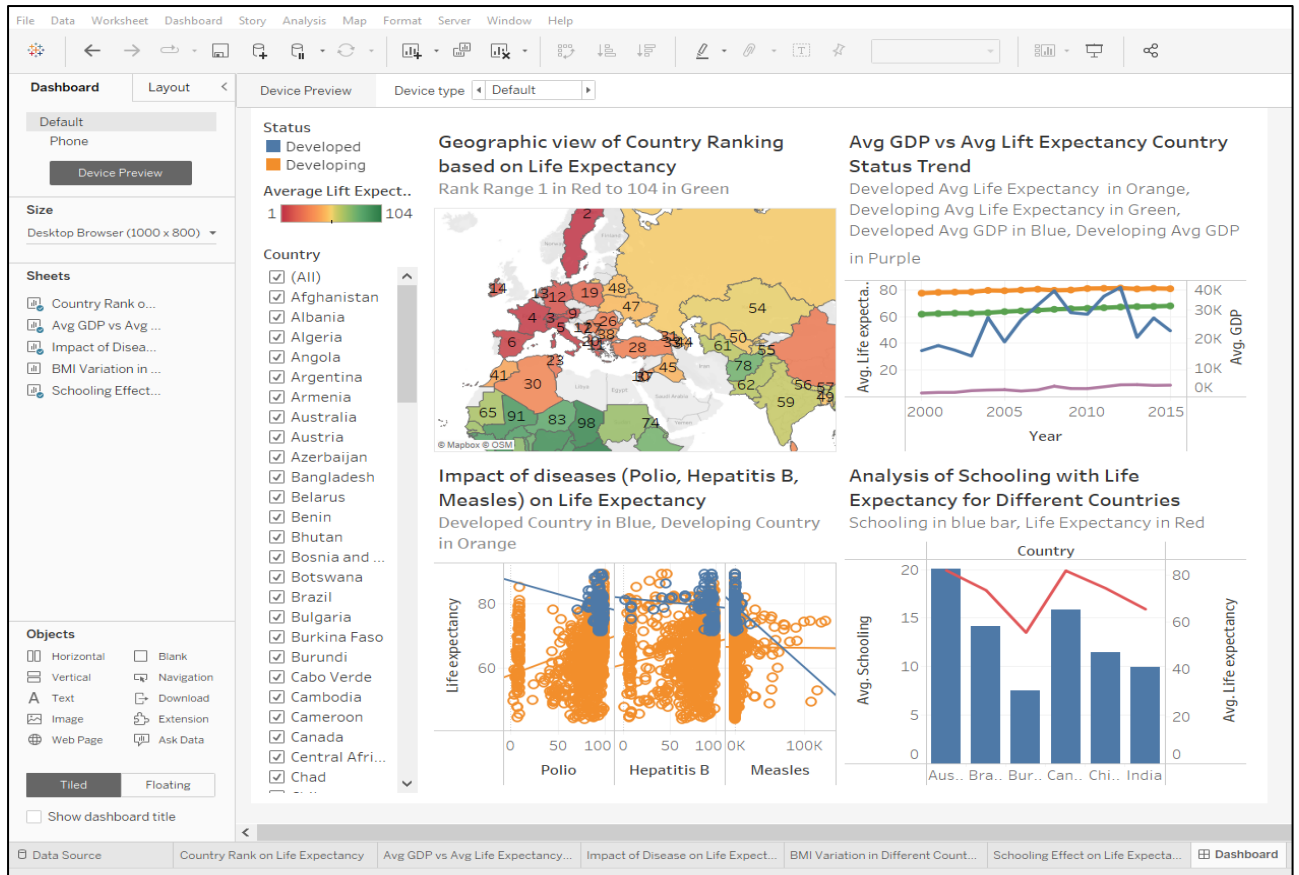
Dashboard demonstrates the collection of selected views of our data analysis of all worksheets in one place. It lets us compare and relate each analysis from the view. We have identified appropriate analysis worksheet to build a story and identify what the key incites of our data are describing further. The first is showing the geographic view of the country ranking based on the life expectancy. Here we initially start with identifying the life expectancy on a global scale by observing from the given country list. We can rank the country according to life expectancy in the ascending order such as first as highest life expectancy and so on. In this way we could have an overall view of the life expectancy with their country rank which could help us in our analysis. The second part explain the average GDP for the life expectancy from 2000 to 2015 year for the developed and developing countries. If the average GDP is increased over the past 16 years, then the average expectancy of the developed countries is increased. For developing countries, the average GDP and average life expectancy is slightly increased. However, the developed countries population have more life expectancy than developing countries. In the third part, they illustrate about the life expectancy curve after getting polio, hepatitis B, measles disease for the developed countries and developing countries. If the developed & developing country people suffered these three diseases then their life expectancy of the developed country is higher than the developing country population because they have a lots of hospital services regarding these diseases, from that they will get medical assistance in a very less time. That is why, the developed country population are living higher age than developing countries. When we click on any one part like country in the dashboard, then the other three part will be reflected and adjusted accordingly.

The fourth diagram shows that the analysis if schooling with the life expectancy for different countries, where the average schooling indicates in bar as a blue color and red line represents the average life expectancy.
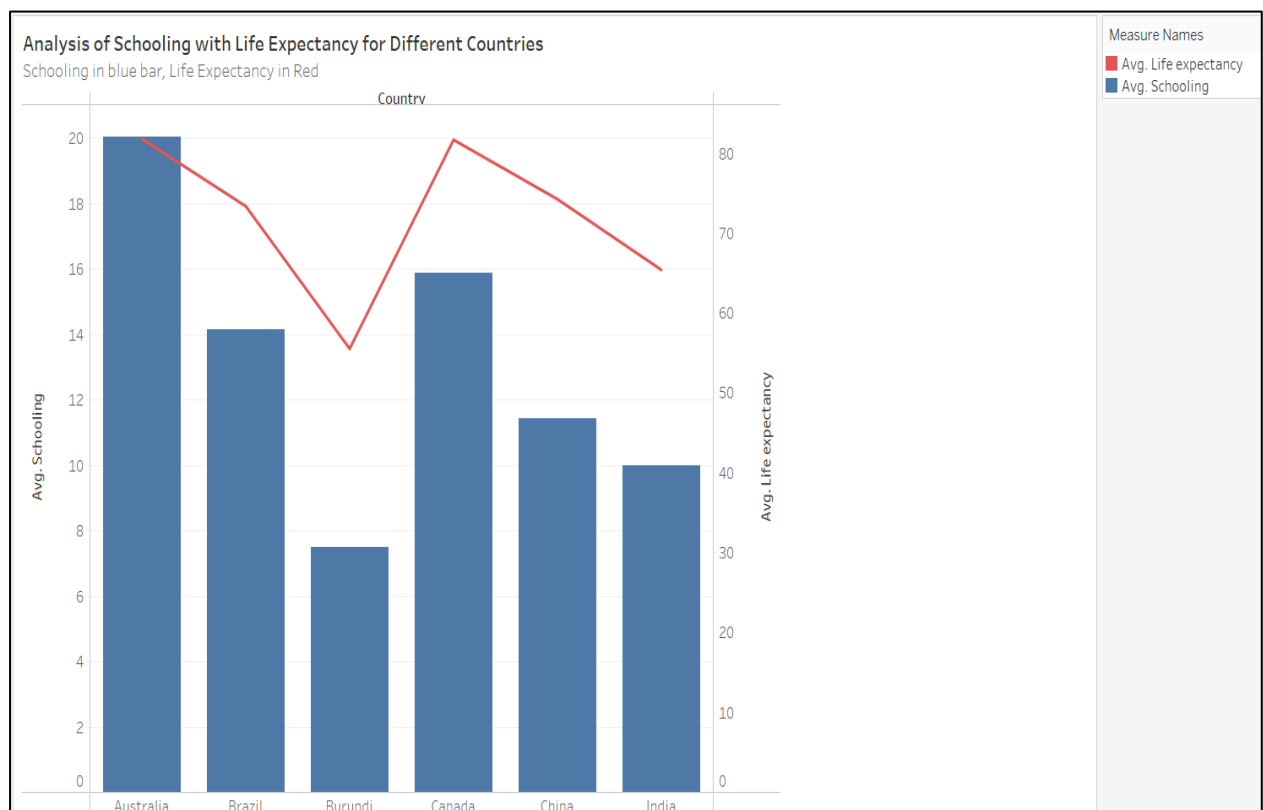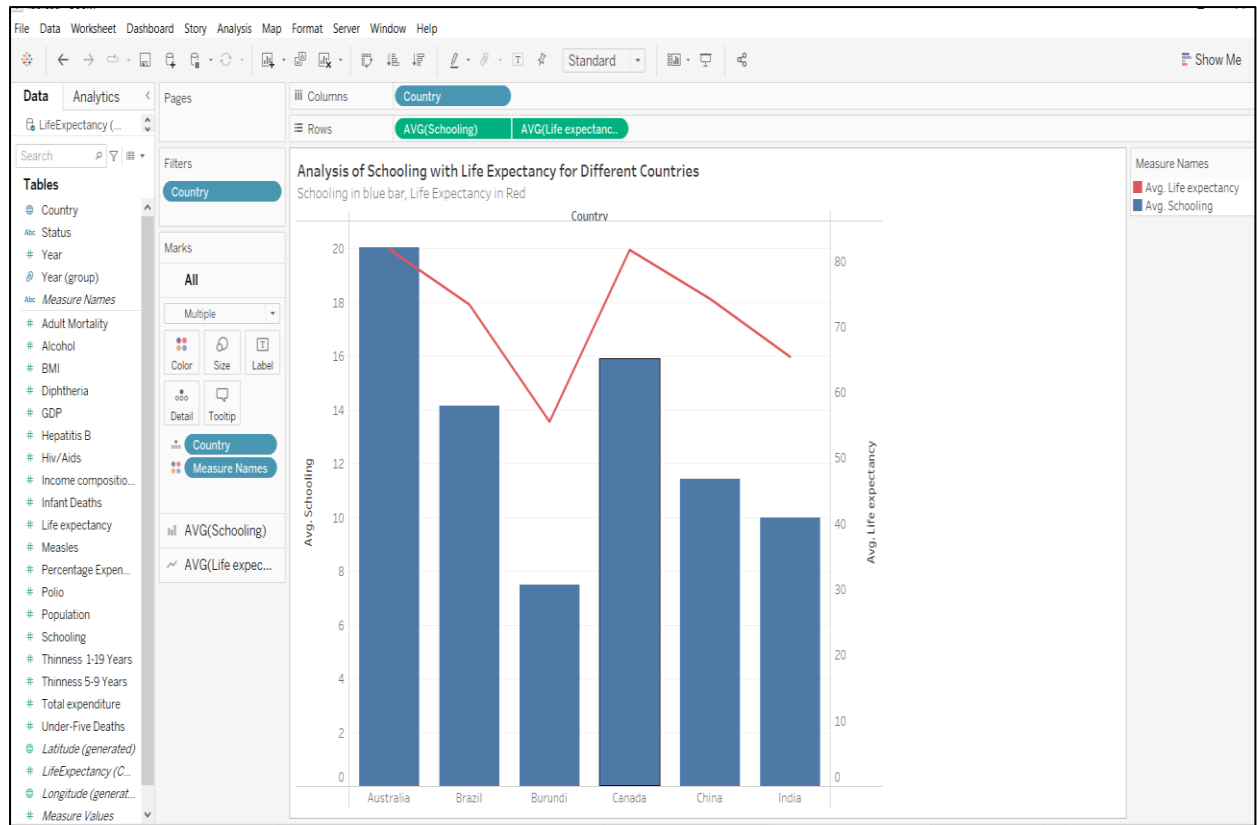
[Application used: Histogram, Line & Marks filter]

This visualization shows the impact of schooling people on their life expectancy for the given countries. The figure depicts Australia country has highest number of people done their schooling, so that their life expectancy is high. For the Brazil country, their literacy rate is higher than the Burundi country. That's why, the average life expectancy of Brazil is more than Burundi. For the last three countries, Canada population is increased in term of schooling than China & India population. Therefore, the Canadian people have better life expectancy. We can also visualize the other countries schooling level by selecting filter option in the country data and select countries as per the user's need. To conclude, higher the schooling of the country provides better average life expectancy. Based on the above analysis, we can anticipate that the life expectancy is higher for the people who are living in the developed country with various hospital facilities and school facilities. The healthier people live longer age.

7.) <u>STORYTELLING</u>

Life is the most important aspect of human being. We live it and try to cherish it every moment. Even though most of the people want to have it eternally, however, it is not possible right now. We can not live forever but we can extend our life by adopting a healthy lifestyle. The statistical measure of any organism lifespan is measured by Life Expectancy. The expectation of human life expectancy should uniform for everybody; however, it is not true. There are various factors that affect a human's lift expectancy. Although, it is extremely difficult to come up with a silver bullet that can maximize the global life expectancy, but we can analyze some of the factors that

are affecting the global lift expectancy with the available information. Here we are proposing

some of remedies to address the disparity in the life expectancy among various nations.

- **Continental View of Life Expectancy**
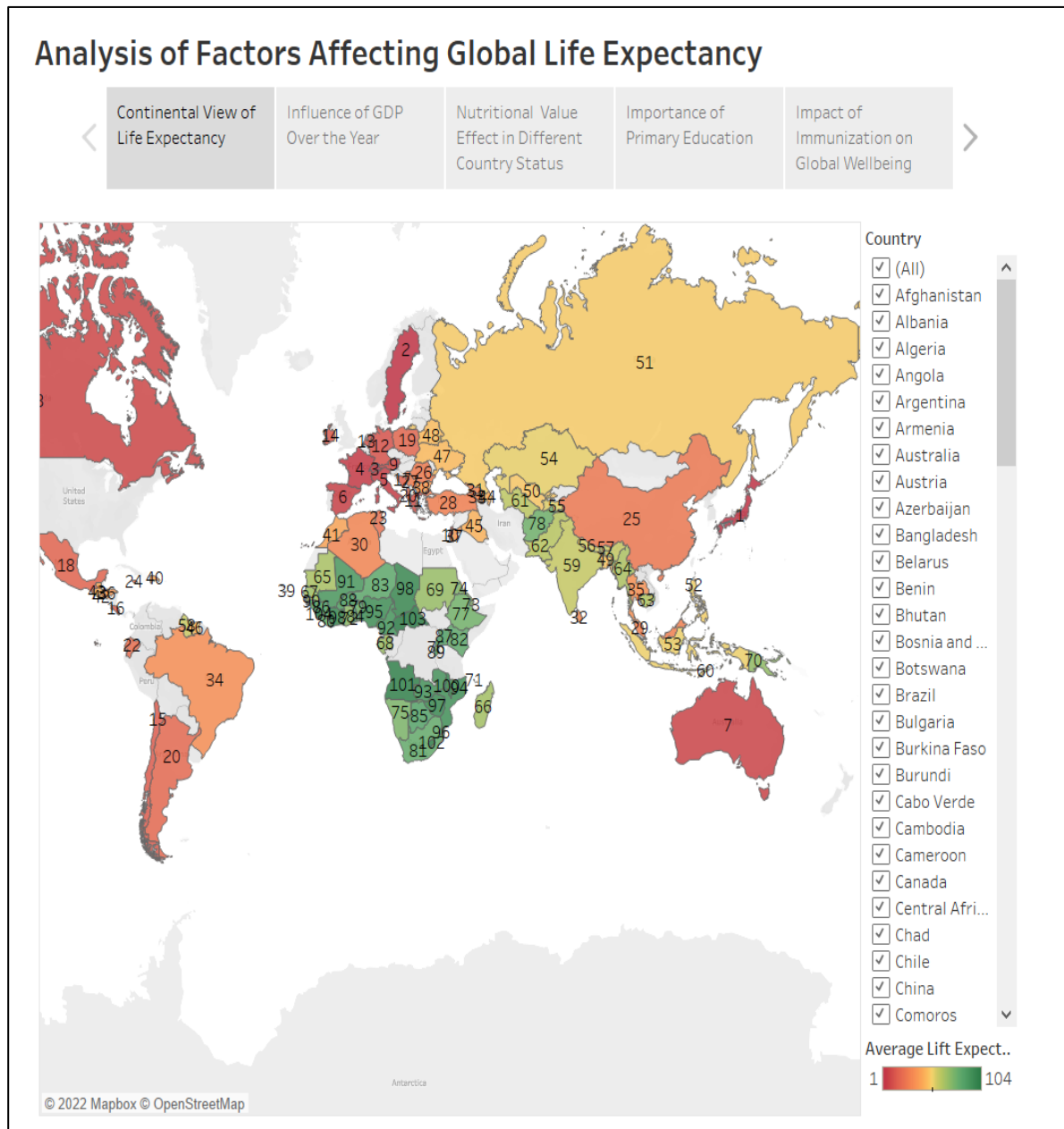
The dispersion of avg life expectancy over different continents can be view from the given

geographical ranks of different nations. The North and South American along with the

majority of Northern Europe have highest ranked avg GDP nations. It is also evident that most

of the higher life expectancy nations are clustered together. It provides the influence of nation

over its neighboring countries. Most Asian countries have mid-range avg GDP, Asia also houses approx. 60% of the global population and it shows that most of the population in the world have medium range Life Expectancy compared to the European and American countries. In the last we have African countries that have the lowest life expectancy. Even though human species evolved from the African continent the current situation of the majority of the African countries is very dire. This analysis provides us the insight that country of origin plays an important role in longevity of human. It is not surprising that we see global migration of people to more prosperous country to get a better healthy and long life. This analysis shows some nations need a greater help to overcome this shorter life expectancy problem.

- **Influence of GDP Over the Years**
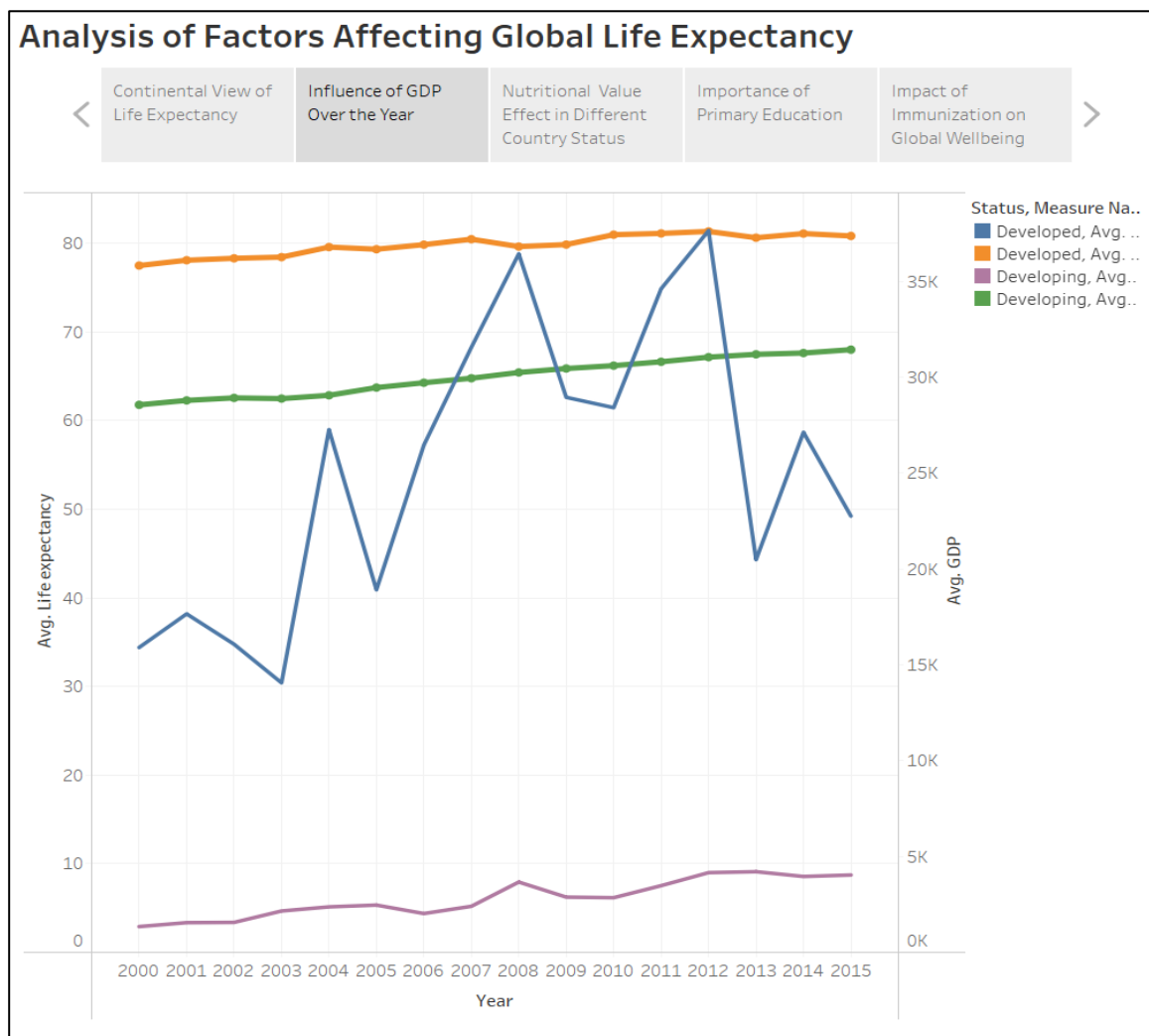
The Gross Domestic Product (GDP) is defined as the total monetary value of all the completed goods and services produced within a country. The GDP provides the economic prosperity of the country and can give a broad overview of the country's avg Life Expectancy. To simplify the analysis, we can divide the global countries into two statuses, developed and developing. This will help us to establish a relationship of avg GDP with avg Life Expectancy of developing and developed nations over the year from year 2000 to 2015. We can see that over the year both average GDP and avg Life Expectancy have increased. It is also noted that developed countries have much higher average GDP than the developing nation. Even though both average GDP and average Life Expectancy have increase but there is a wide gap in both statuses. The higher GDP nations have higher employment, better healthcare budget and people have access to more resources low GDP nation. These factor influences the life expectancy of a person. The economic propensity of a nation directly reflects to the lifespan of its citizen.

- **Nutritional Value Effect in Different Country Status**
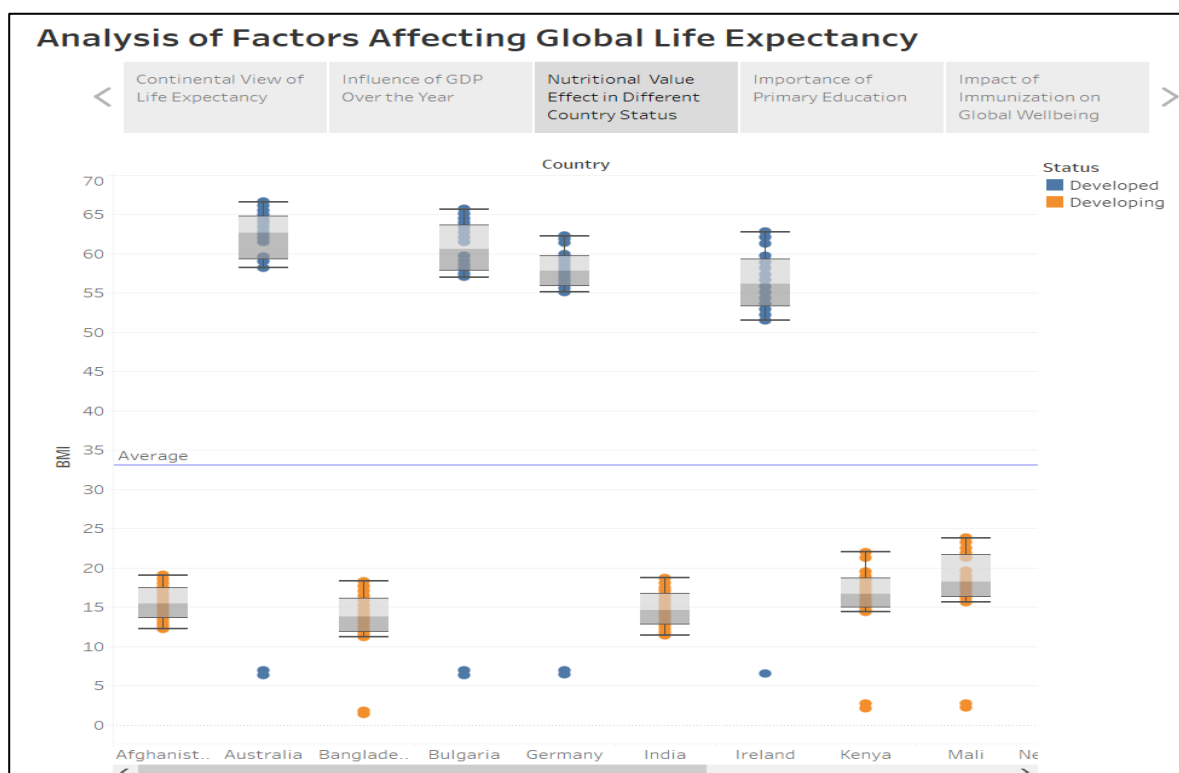
Food is essential for the survival of human being. However, nutritional value plays importance in longevity of the person. The human body need sufficient calorific value of food to maintain the healthy Body Mass Index (BMI). The analysis shows that the developed countries have higher BMI than the developing one. The box plot also shows that even thought the developed countries have high variance in BMI it is within the normal range. On the other hand, the developing countries have below average BMI. This indicates a lower access of balance diet that causes malnutrition. The deficiency of essential minerals and micronutrients affects the overall well beings. It showcases this stark contrast between these developing and developed countries.

- **Importance of Primary Education**
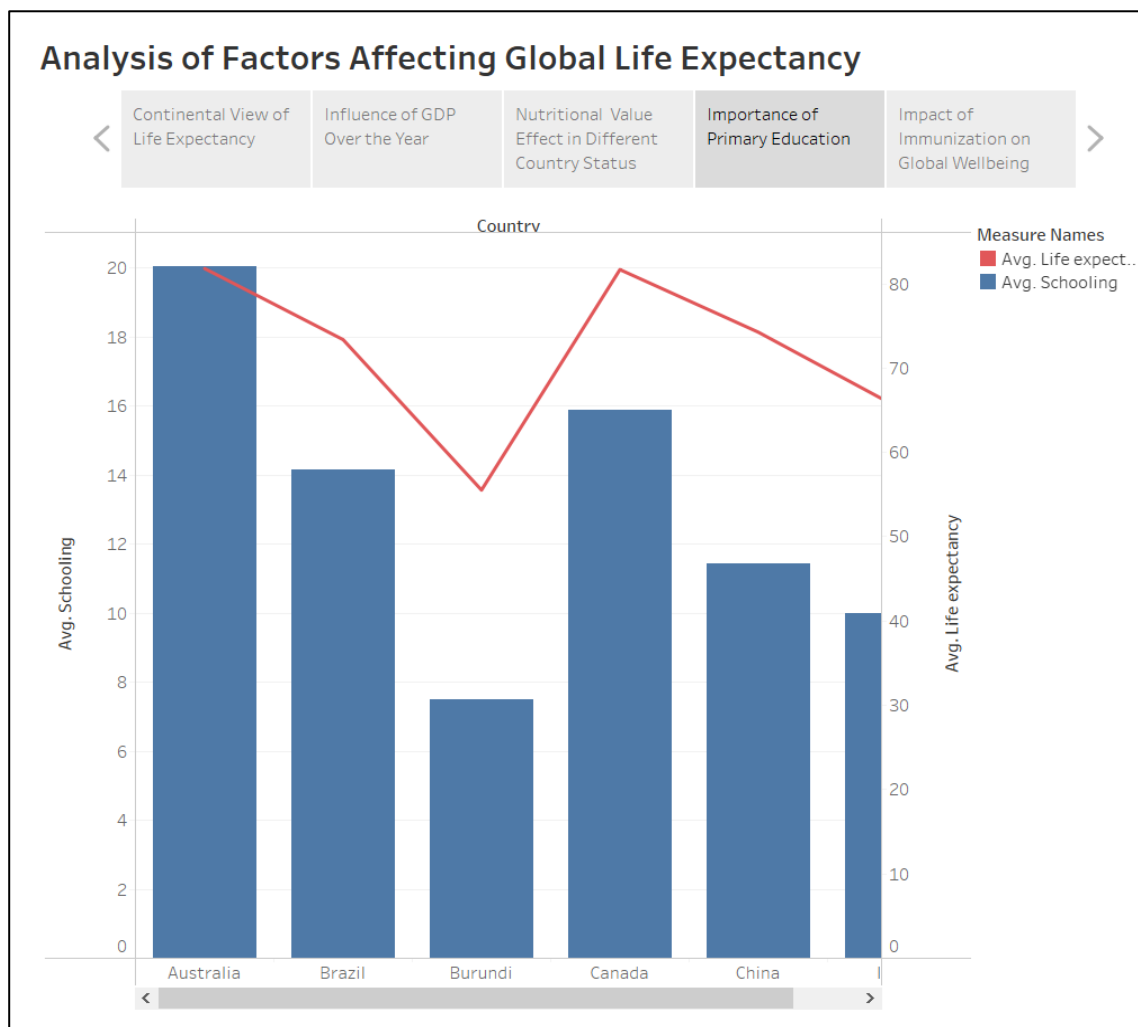
Education is a crucial part of human brain development. Knowledge empowers us to do not only better for ourselves but also for the society and nation. Education is important for learning and development of skills that help us to perform better work and take rights decisions. Someone not only can earn a reasonable wedge with education but can also access better healthcare where medical professionals are in sufficiently available. The primary education provides foundation of knowledge, and it is very clear from the plot that higher access to schooling relates with average higher life expectancy. The lack of schooling access affect on a broader level of country's lift expectancy, this provides an opportunity to nations to fix the educational sector so it can provide overall better outcomes for every citizen.

- **Impact of Immunization on Global Wellbeing**
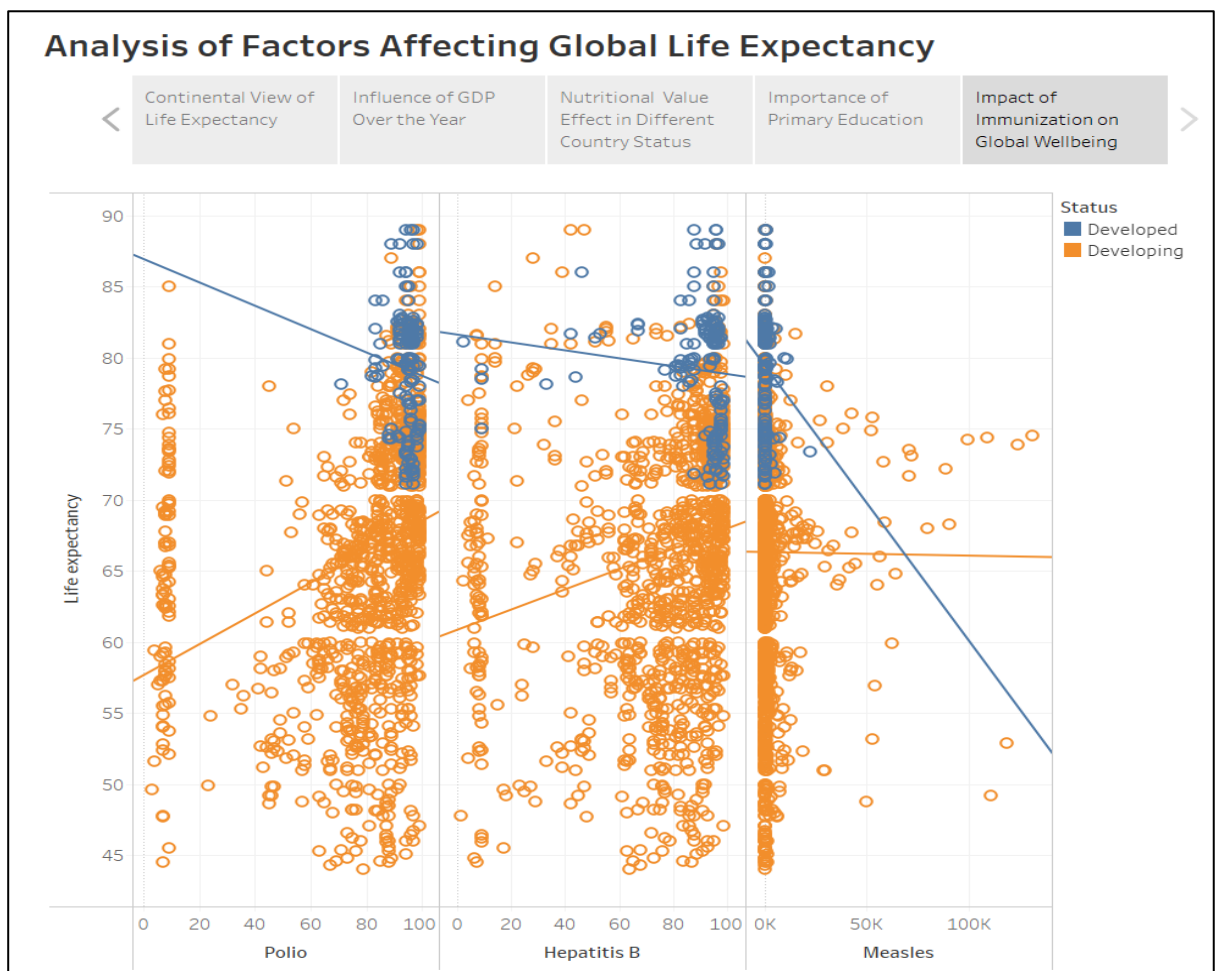
Infants are the one of the easiest targets of deadly diseases. There are many lethal diseases exists in the world and many of them are incurable once infected. Three of these major diseases are Polio, Hepatitis B and Measles. Not only these diseases are fatal but extremely contagious. Although these diseases are extremely dangerous, however, all of them are preventable by proper immunization. An early age vaccine provides highest level of immunity from these diseases for the lifetime. We have seen that the developed countries have high life expectancy due to low prevenance of these diseases in the country. We see from the trend lines that the life expectancy drops significantly with higher number of cases in developed countries. This is also one of the important factors to be addressed that the higher concentration of these disease affects the average life expectancy very drastically. There should a proper vaccination plans should be in placed to improve the overall of wellbeing for every nation.

CONCLUSION

In this report we have analyzed various factors that affect the Life Expectancy of a country. We started with the introduction about importance of analysis of Life Expectancy. We inspected the data for the any anomaly and inconsistency. A clean data is required for a credible analysis, and we applied many transformations to clean the data. We used Microsoft Excel for the data cleaning task. Once we cleaned the data, we analyzed all the important factors individually. We used Tableau software to perform the rest of the analysis. This analysis showed us that there are many variables the affect and correlates with the average lift expectancy of a country. We aggregated all those individual sheets into a interactive dashboard that provides us a wholistic view of report. We then used tableau story feature to synthesize our information into a story. This analysis showed us that we can control various parameter to increase the average human life expectancy. We can also help the underprivileged countries to increase their wellbeing. This report provides all the important details that can be used to elevate the overall life span of many people.

8.) <u>REFERENCES:</u>

1. [Exploratory Data Analysis - Life Expectancy by WHO - YouTube](#)

2. [Workbook: Life Expectancy WorldBank Data (tableau.com)](#)

3. [Recent trends in life expectancy across high income countries: retrospective observational study | The BMJ](#)

4. Preston SH. The changing relation between mortality and level of economic development. Popul Stud. 1975;29:231–48. [The Changing Relation between Mortality and level of Economic Development: Population Studies: Vol 29, No 2 (tandfonline.com)](#)

5. [Life Expectancy (WHO) Detailed EDA and Exploration | Kaggle](#)

6. [Life Expectancy WHO | Kaggle](#)

7. https://ourworldindata.org/grapher/life-expectancy-vs-gdp-per-capita