****

Exploratory Data Analysis

**ON DESCRIPTIVE STATISTICS OF CERTAIN SPECIFIC PARAMETERS (Infant Mortality Rate, Birth Rate, Death Rate, GDP per Capita) FOR THE WORLD IN COMPARISION TO INDIA FOLLOWING A DETAILED DRILL DOWN ON INDIAN STATES**

* Manas Kumar Sarkar(A22022)
* Ashutosh Pandey (A22010)
* Atish Kumar Majee (A22012)
* Soumya Chatterjee (A22046)

Under the guidance of

Dr. Sayantani Roy Choudhury

STS-I Project

DS Fall Batch 2022

Motivation Behind the Project:

The aim of this project study is to explore the distribution and relation of infant mortality rate, death rate & birth rate with GDP per capita. With the help of exploratory data analysis, we intend to gain insights on the aforementioned parameters for the world, and compare them with India to see where it stands in the developing world. We further take a dig onto these data of Indian states and compare the data to get a holistic view of these measures across different geographies of our country.

Definitions:

* The **infant mortality rate** is **the number of infant deaths for every 1,000 live births**. In addition to giving us key information about maternal and infant health, the infant mortality rate is an important marker of the overall health of a society.
* **Birth rate** is **the number of individuals born in a population in a given amount of time**. Human birth rate is stated as the number of individuals born per year per 1000 in the population. For example, if 35 births occur per year per 1000 individuals, the birth rate is 35.
* **Mortality rate**, or death rate, is **a measure of the number of deaths (in general, or due to a specific cause) in a particular population, scaled to the size of that population, per unit of time**. Human death rate is stated as the number of individuals born per year per 1000 in the population.
* **Gross Domestic Product (GDP) per capita** **shows a country's GDP divided by its total population**

Data Collection:

The Data used for the project is for the year 2019. The data is collected in the form of secondary data from web sources which are listed below:

1. Infant Mortality Rate: <https://data.worldbank.org/indicator/SP.DYN.IMRT.IN>
2. Birth Rate:
3. Death Rate:
4. GDP per capita:
5. Birth Rate, Death Rate, Infant Mortality rates for Indian States: Economic Survey 2021-22
6. GPD per capita for Indian States: Wikipedia



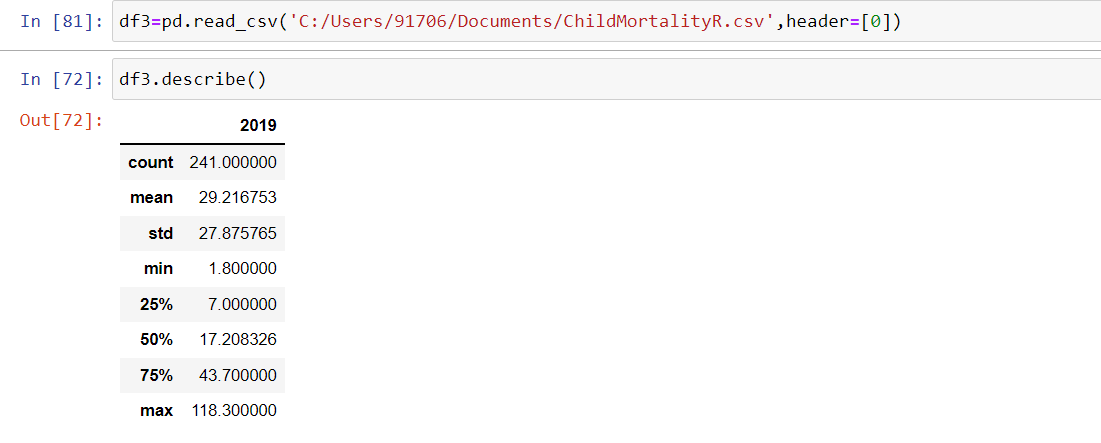
Analysis:

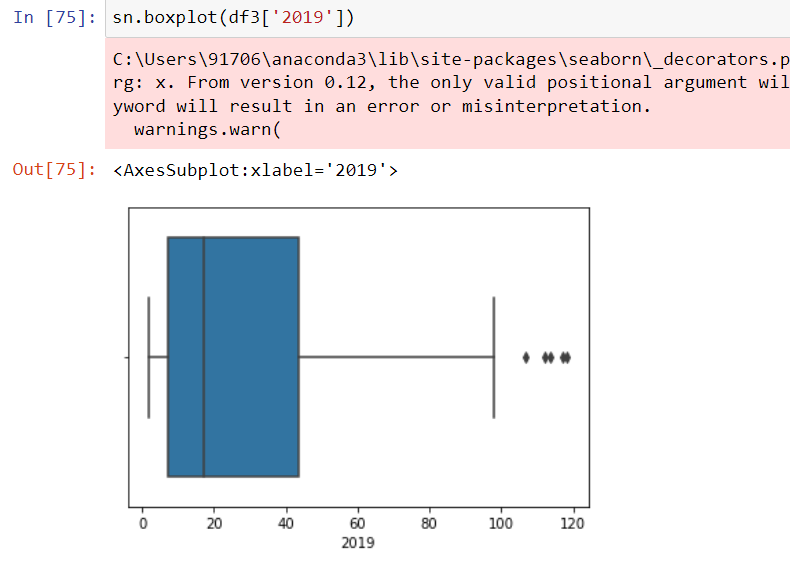
**TOOLS USED:**

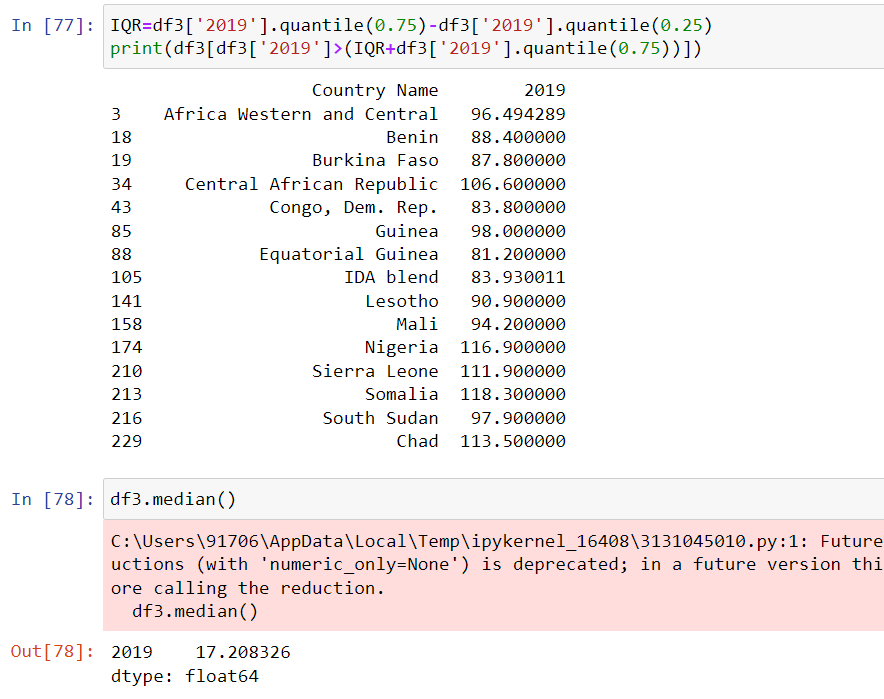
* **Microsoft Excel**
* **Python- Jupyter Notebook**
* **Power BI**
* **Tableau**

**The World with respect to India:**

1. **Infant Mortality rate:**

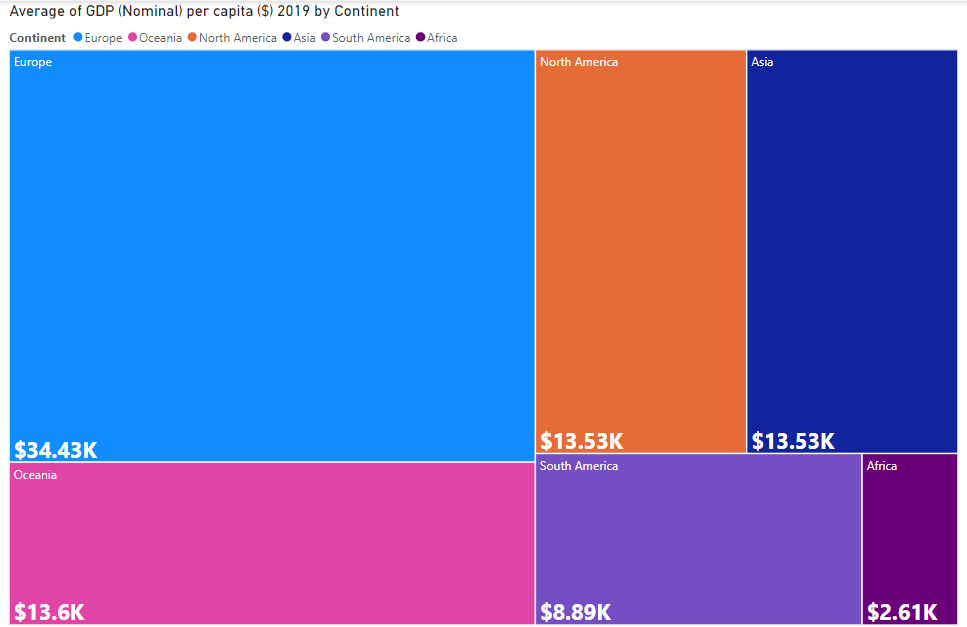
****

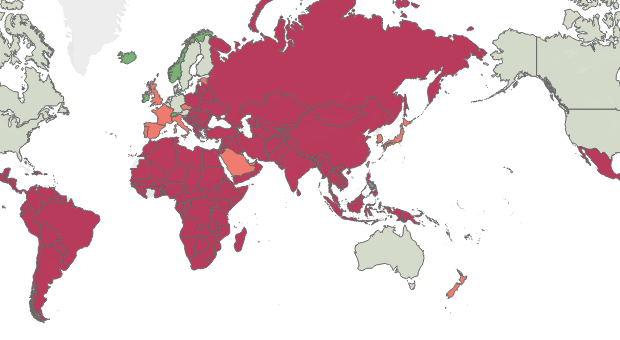
****

****

The infant mortality rate is rightly skewed with many outliers. The infant mortality rate for the world is 29.1 and median is 17.20. Average infant mortality rate for India is 27.01 which is lesser than the world average but still higher than the median infant mortality rate of the world.

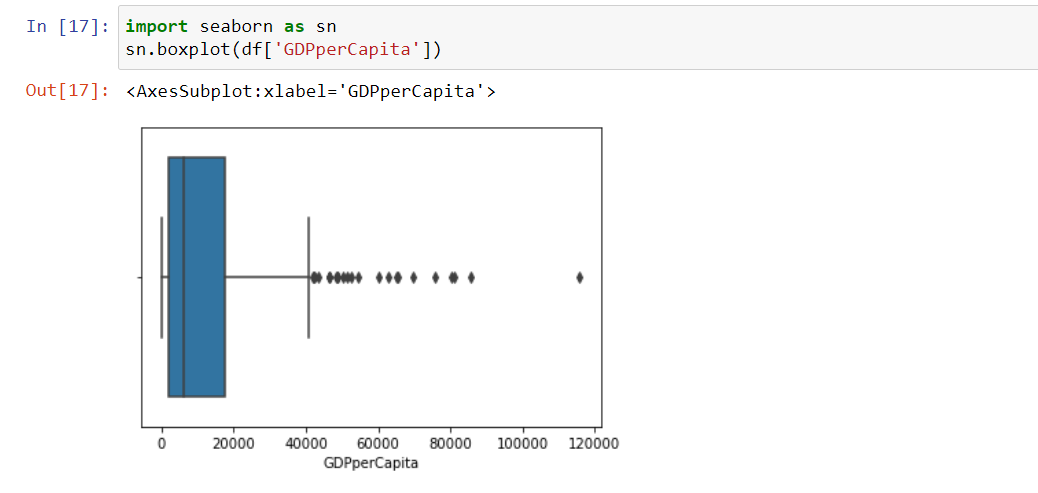
Among the developing countries, the average infant mortality rate is much higher than the average of the developing countries. It is a matter of concern & we need to emphasize on infant healthcare.

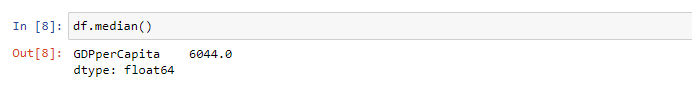
1. Timeline

   Description automatically generatedGDP per capita

The above plots give the distribution of GDP per capita for the world and the average GDP across the continents.

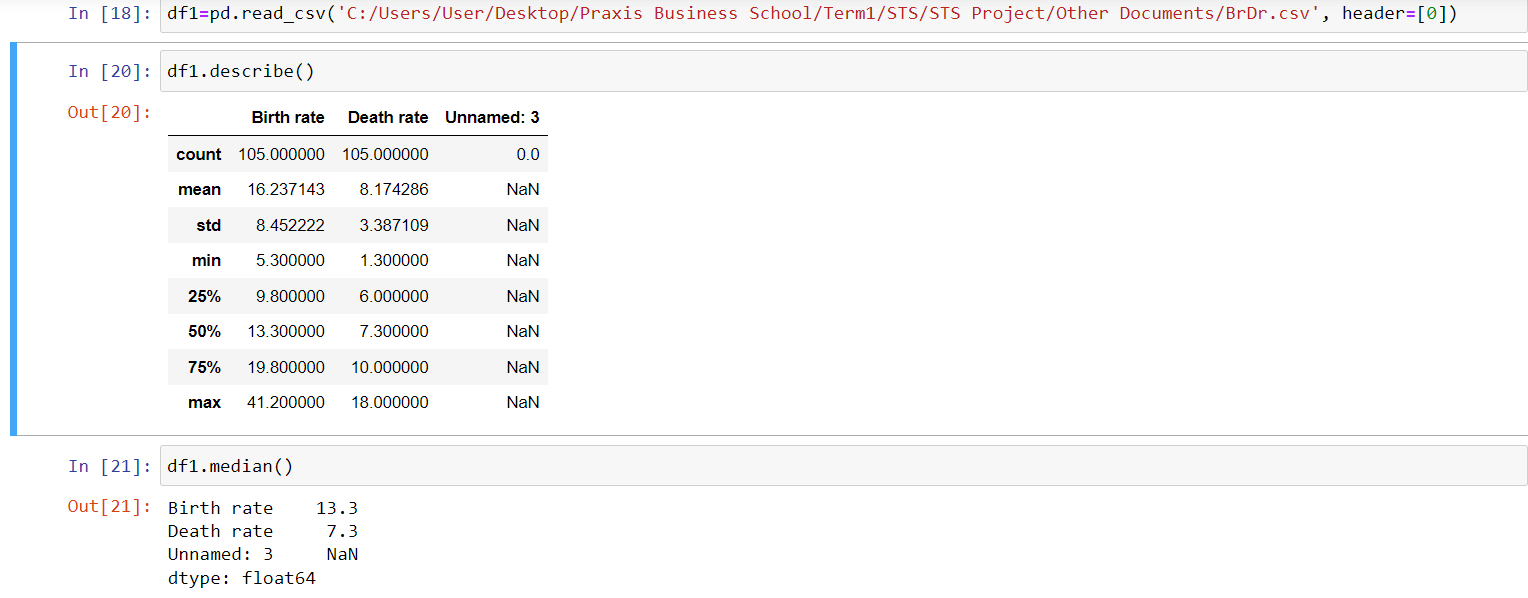


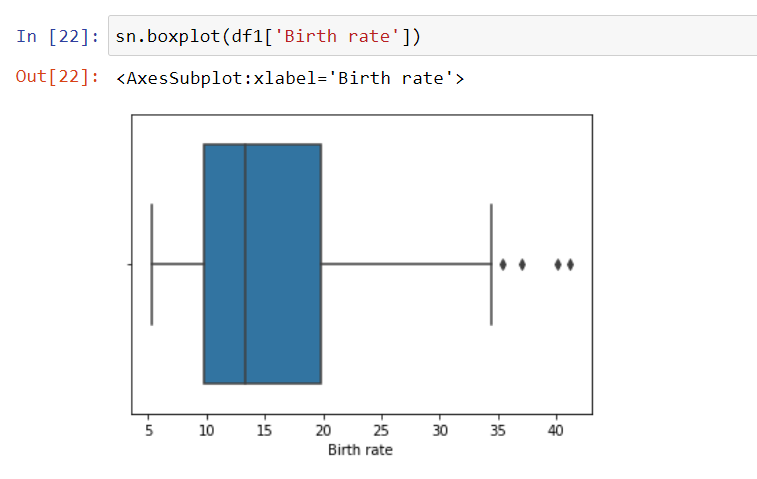


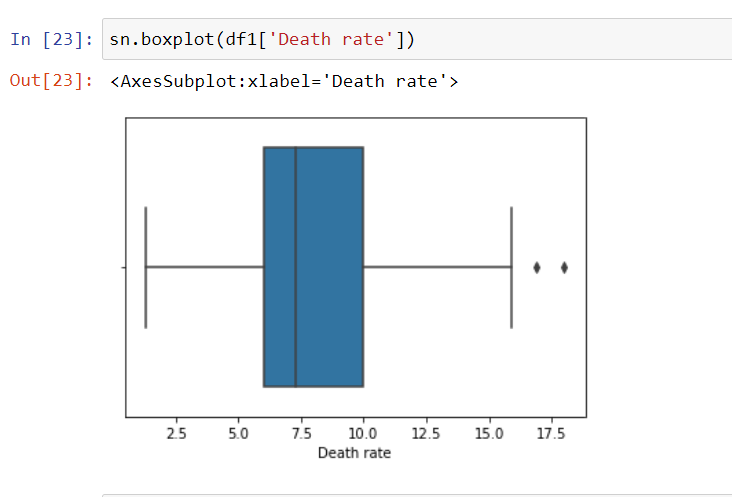


The distribution of GDP per capita across the world is rightly skewed & has several outliers which are mostly the advanced countries. The mean GDP per capita for the world is 14592.72 dollars whereas the median GDP is 6044 dollars. Per capita GDP for India is 2099 dollars which is very low even in the developing countries. It is even less than the average GDP per capita of Asia.

1. Birth Rate & 4) Death Rate





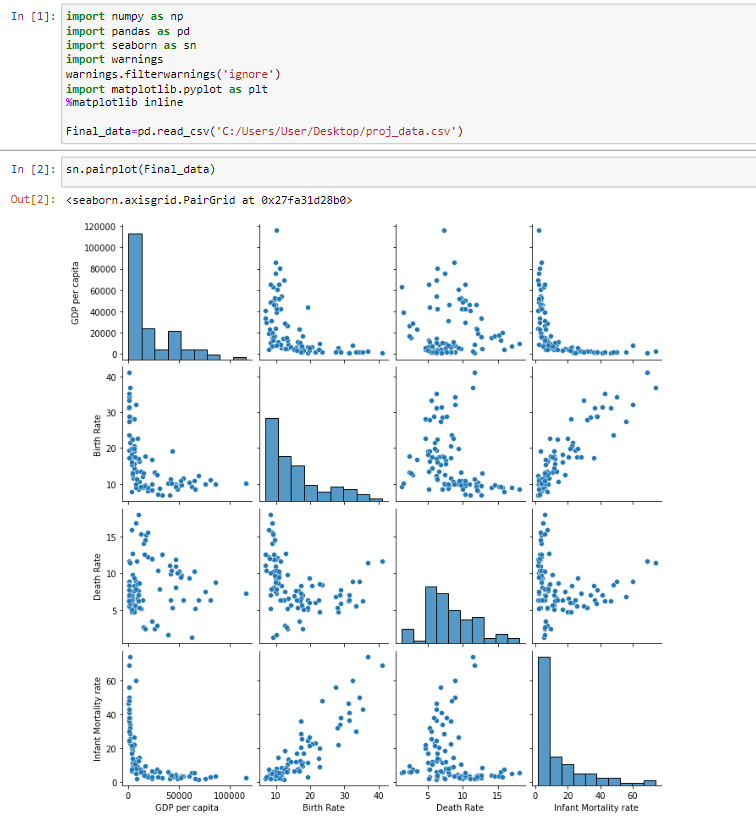


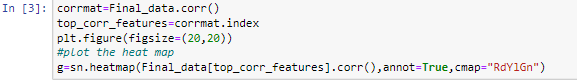
The death rate is rightly skewed with a few outliers. The average death rate for the world is 8.17 and median is 7.30. Average birth rate for India is 7.3 which is same as the median death rate of the world.

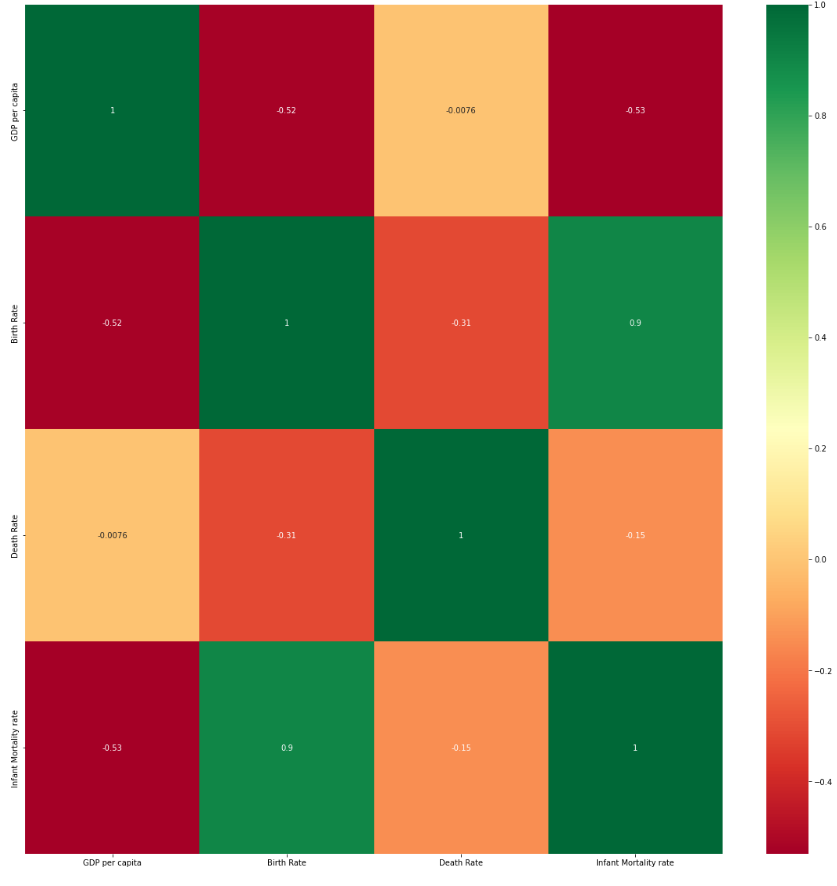
Among the developing countries, the average death rate of India is much higher than the average of the developing countries.

The birth rate is rightly skewed with a few outliers. The average birth rate for the world is 16.23 and median is 13.30. Average birth rate for India is 17.4 which is greater as well as the birth rate of the world.

Among the developing countries, the average birth rate of India is much higher than the average of the developing countries. That explains the more rise in population of India as compared to other countries.

**Pair Plots**

**Heat map**

****

From the **PAIR PLOT** and **HEAT MAP** we observe that:

1. **Birth Rate** has a **Moderate Negative** correlation with **GSDP** with r= -0.52
2. **Death Rate** has a **Weak Negative** correlation with **GSDP** with r= -0.0076
3. **Infant Mortality Rate** has a **Moderate Negative** correlation with **GSDP** with r= -0.53

**Inference:**

From the results we can infer that for the countries in the world, with the increase in GDP per capita the Birth rate & the infant mortality rate tend to decrease significantly whereas the Death Rate seems to be very weak negatively correlated with the GDP.

**Regression (best fit Lines):**

**Y** (Infant mortality rate) **= -0.00036X**(GDP per capita) **+ 22.48**

**Y** (Death rate) **= -1.09\*10-6 X**(GDP per capita) + **8.25**

**Y** (Birth Rate) **= -0.00018X** (GDP per capita) **+19.852**

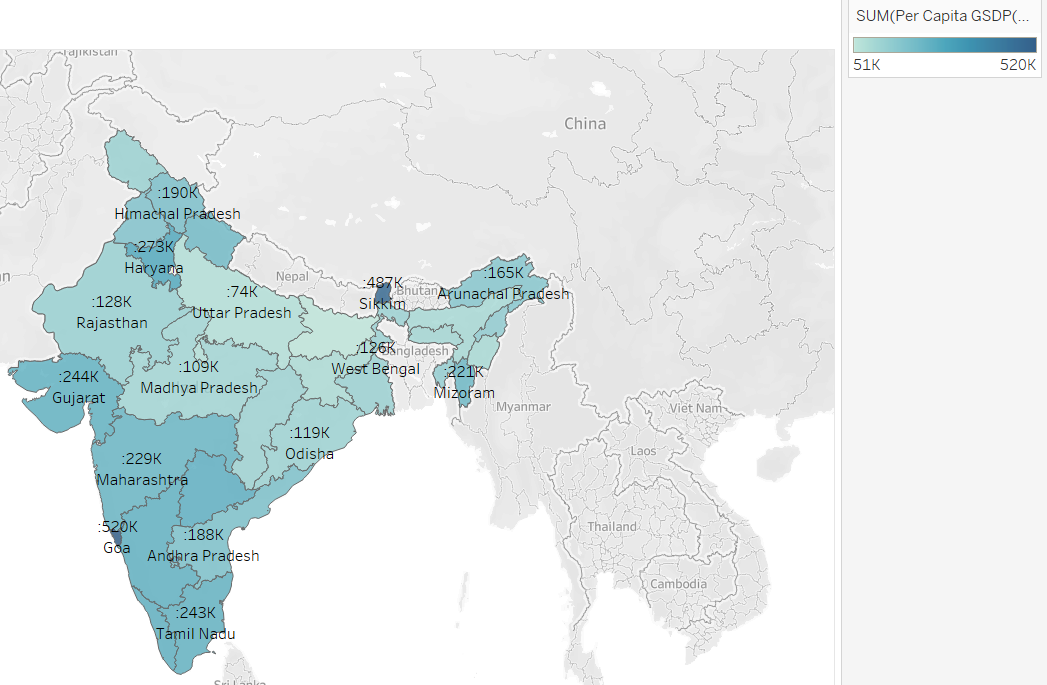
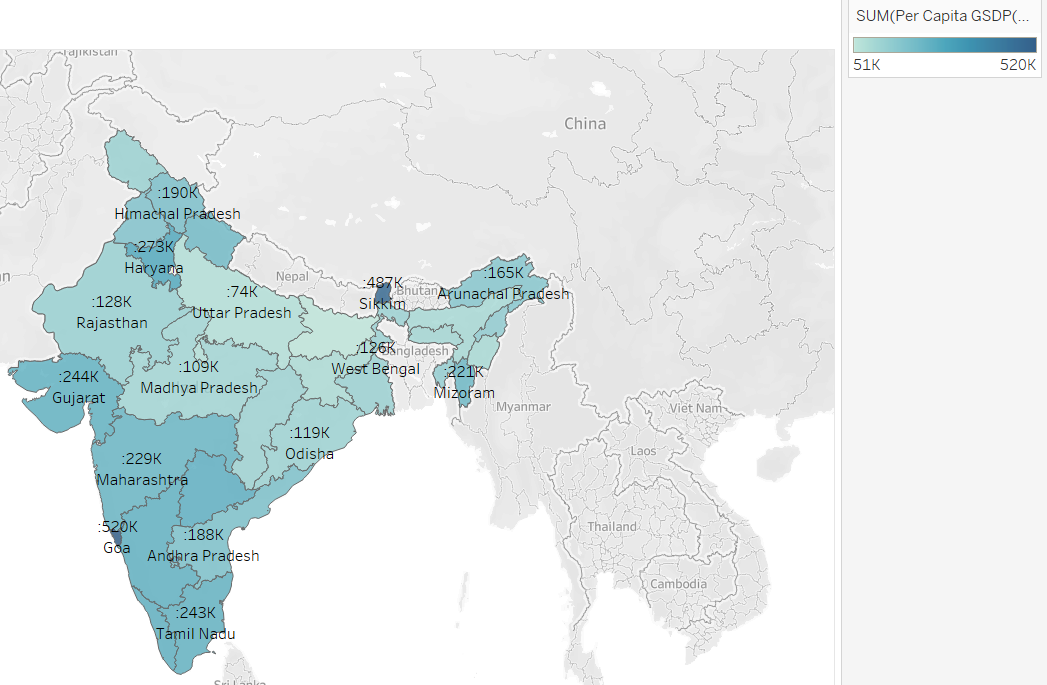
I

INDIA

We have collected these data for the year 2019 for different states in INDIA.

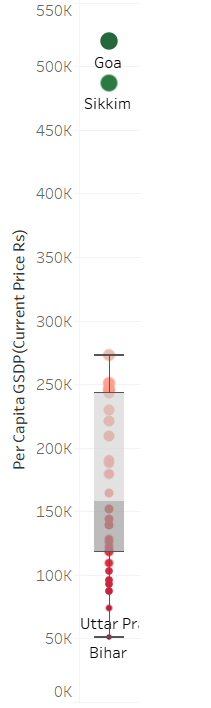
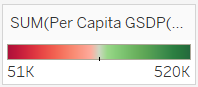
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STATE** | **Per Capita GSDP (Current Price Rs)** | **Birth Rate** | **Death Rate** | **Infant Mortality rate** |
| Andhra Pradesh | 188069 | 15.9 | 6.4 | 25 |
| Arunachal Pradesh | 164557 | 17.6 | 5.8 | 29 |
| Assam | 96224 | 21 | 6.3 | 40 |
| Bihar | 50735 | 25.8 | 5.5 | 29 |
| Chhattisgarh | 117700 | 22.2 | 7.3 | 40 |
| Goa | 520030 | 12.3 | 5.9 | 8 |
| Gujarat | 243761 | 19.5 | 5.6 | 25 |
| Haryana | 272884 | 20.1 | 5.9 | 27 |
| Himachal Pradesh | 190407 | 15.4 | 6.9 | 19 |
| Jharkhand | 87126 | 22.3 | 5.3 | 27 |
| Jammu & Kashmir | 121971 | 14.9 | 4.6 | 20 |
| Karnataka | 246419 | 16.9 | 6.2 | 21 |
| Kerala | 245323 | 13.5 | 7.1 | 6 |
| Madhya Pradesh | 109372 | 24.5 | 6.6 | 46 |
| Maharashtra | 229489 | 15.3 | 5.4 | 17 |
| Manipur | 92427 | 13.6 | 4.3 | 10 |
| Meghalaya | 102672 | 23.2 | 5.6 | 33 |
| Mizoram | 221384 | 14.5 | 4 | 3 |
| Nagaland | 144138 | 12.7 | 3.5 | 3 |
| Odisha | 119075 | 18 | 7.1 | 38 |
| Punjab | 179163 | 14.5 | 6.6 | 19 |
| Rajasthan | 128318 | 23.7 | 5.7 | 35 |
| Sikkim | 487201 | 16.5 | 4.2 | 5 |
| Tamil Nadu | 243189 | 14.2 | 6.1 | 15 |
| Telangana | 250920 | 16.7 | 6.1 | 23 |
| Tripura | 139540 | 12.8 | 5.5 | 21 |
| Uttar Pradesh | 74141 | 25.4 | 6.5 | 41 |
| Uttarakhand | 209116 | 17.1 | 6 | 27 |
| West Bengal | 126121 | 14.9 | 5.3 | 20 |

1. **GSDP per capita:**



1.From the Geolocation chart it can be seen that states located in the northern or the south-western part of the country are having higher GSDP(Per-capita).

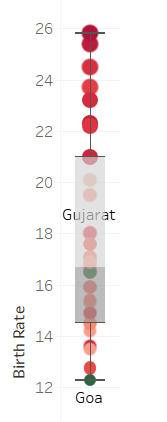
2.Goa is having the highest GSDP (per capita) whereas Sikkim has the 2nd highest in terms of GSDP(per capita).

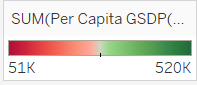


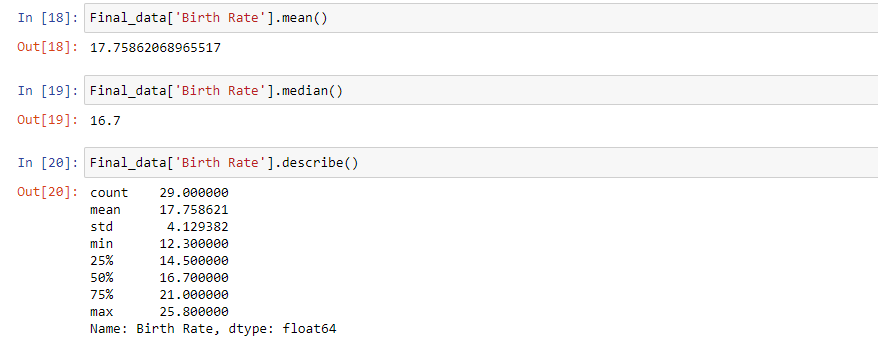


The GSDP distribution is positively skewed where Goa and Sikkim are outliers in the positive side. Also, eastern part of India seems to be having lesser GSDP per Capita as compared to the Northern and Western part of India. The mean GSDP per capita is 186257.65 rupees and median is 164557 rupees.

1. **Birth Rate**

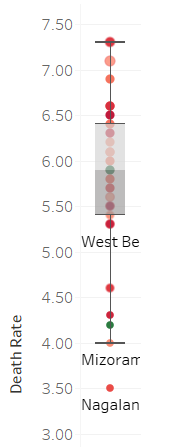


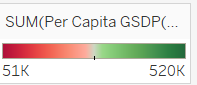


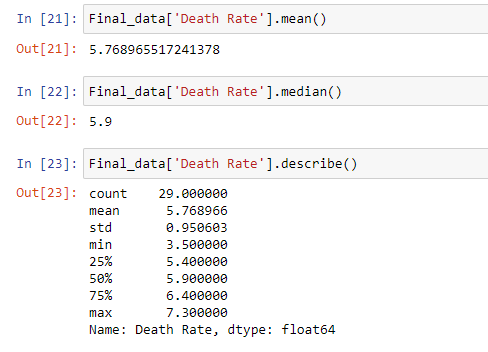


The Birth Rate distribution is positively skewed. The states with higher GSDP (Goa, Sikkim, Kerala etc) per capita are having lesser Birth Rate, while states with lower GSDP (Bihar, Uttarpradesh, Madhyapradesh etc) have higher Birth rates. The mean birth rate is 17.75 and the median values is 16.7.

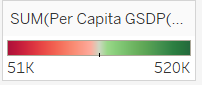
1. **Death Rate**

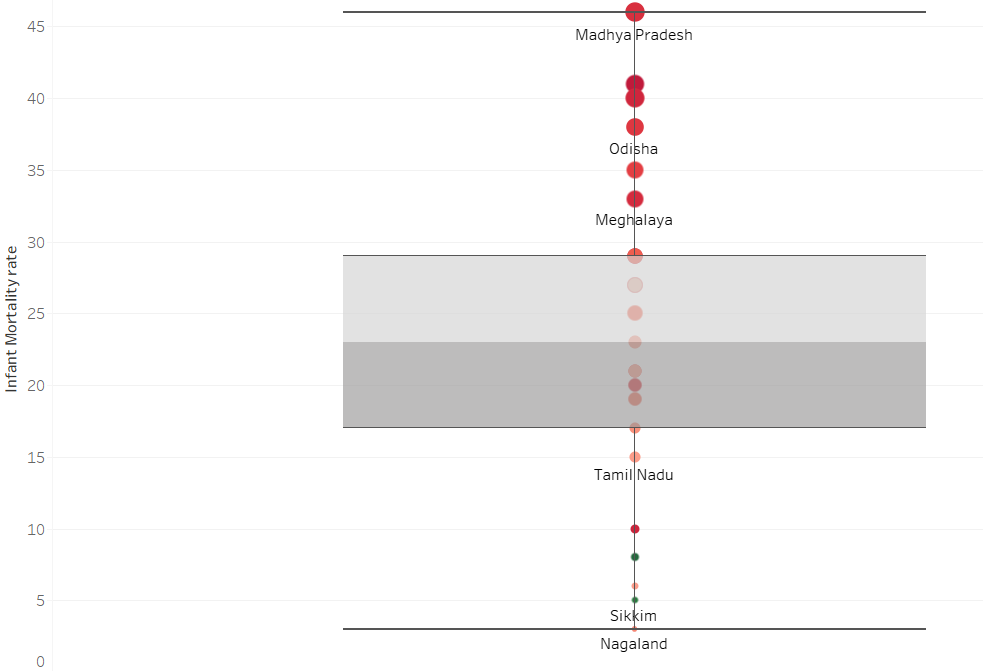


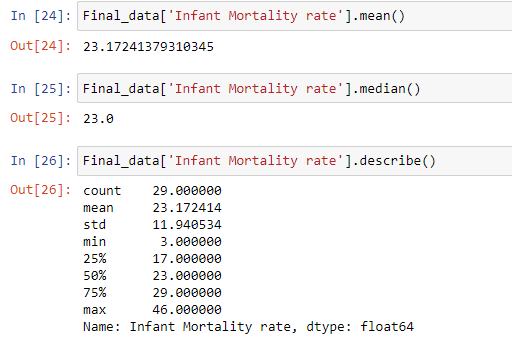




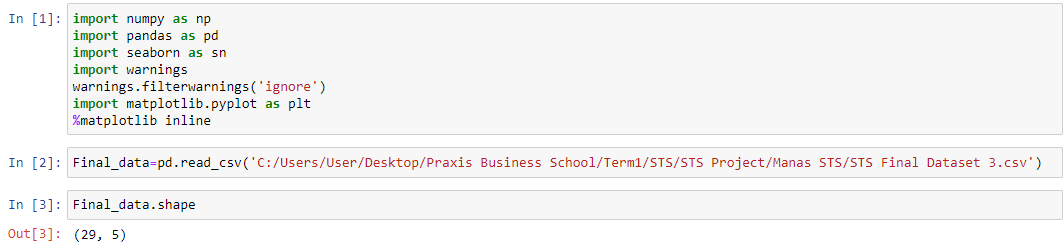
Death rate distribution is approximately normal. Only outlier is the state of Nagaland. The mean death rate is 5.76 and the median is 5.9

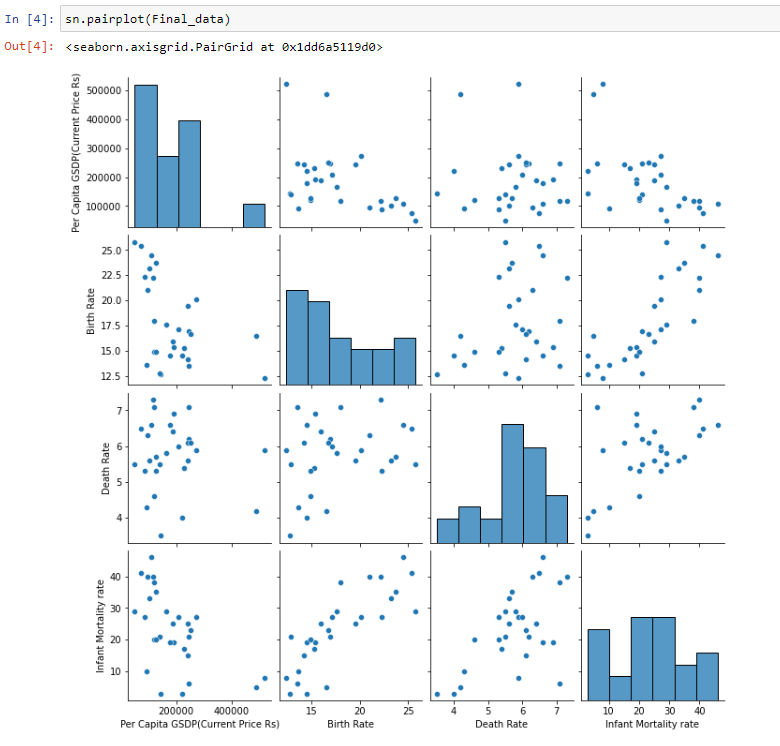
1. **Infant Mortality Rate**

****

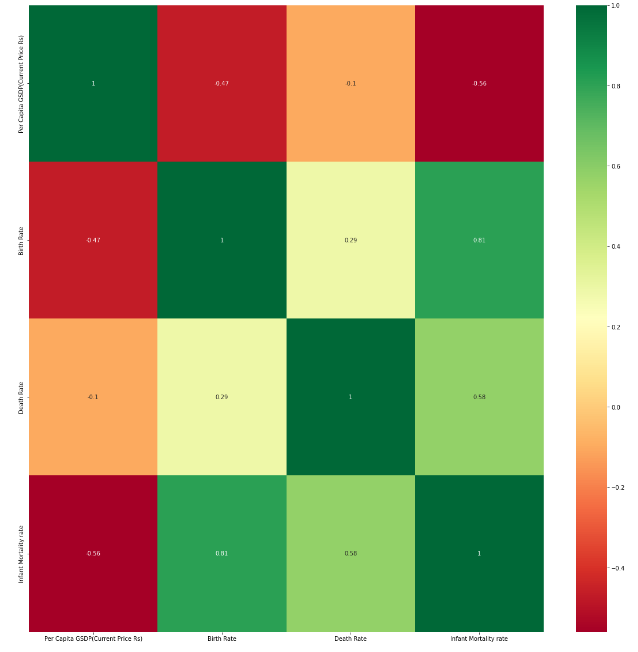


It can be seen that the states with low GSDP have high infant mortality rates and states with high GSDP has low infant mortality rates. The mean infant mortality rate across the states of India is 23.17 and median is 23.0

**Pair Plot**



**Heat Map**



From the **PAIR PLOT** and **HEAT MAP** we observe that:

1. **Birth Rate** has a **Moderate Negative** correlation with **GSDP** with r= -0.47
2. **Death Rate** has a **Weak Negative** correlation with **GSDP** with r= -0.1
3. **Infant Mortality Rate** has a **Moderate Negative** correlation with **GSDP** with r= -0.52

**Inference:**

From the results we can infer that for the states in India with the increase in GSDP per capita the Birth rate & the infant mortality rate tend to decrease significantly whereas the Death Rate seems to be very weak negatively correlated with the GSDP

**Regression (best fit Lines):**

**Y** (Birth Rate) **= -1.9\*10-5X** (GDP per capita) **+ 21.385**

**Y** (Death Rate) **= -1.2\*10-6X** (GDP per capita) **+ 6.03**

**Y** (Infant mortality rate) **= -6.4\*10-5X** (GDP per capita) **+ 35.25**

**CONCLUSION:**

In this project we can see how this Exploratory Data Analysis is playing an Essential part to get some incredibly valuable insights about the World Health Care as well as Indian Health care by considering Infant mortality Rate, Birth Rate, Death Rate with respect to GDP per capita. The rising cost of Health Care is considered a real concern and is directly linked to our income which affects our financial capability pertaining to various ailments and daily well-being & hence, the data of Infant mortality Rate, Birth Rate and Death Rate.

For our use case here, we used the basics statistical knowledge to generate various charts and investigated relationship of different variables and the spread of a certain variable across different countries, region & the states of India. The data set collected from web helped us find how much these parameters are linked to GDP with the help of correlation coefficient. Also, we found the best fit line to predict these parameters given the GDP per capita.

Finally, we hope that the entire effort given to draw a picture of the World Health Care with few of its parameters may be of some use to understand the need to uplift the living standards of the general public. We would like to thank Praxis Business School for providing us such an opportunity to present our work.

Thank you!

