

ON THE IMPACT OF CLIMATIC, FISCAL AND ENVIRONMENTAL FACTORS ON COVID-19: A REGRESSION AND CORRELATION BASED STUDY

Koushika P¹, Mahima S¹, Shahina A¹, Nayeemulla Khan A²

¹Department of Information Technology

²School of Computer Science and Engineering

Koushikapadmanaban18051@it.ssn.edu.in, Mahimamahima18056@it.ssn.edu.in, shahinaa@ssn.edu.in, nayeemulla.khan@vit.ac.in

¹SriSivasubramaniyaNadar College of Engineering, Chennai, Tamil Nadu, India

²VIT Chennai, Tamil Nadu, India

Abstract

COVID-19 is an infectious disease that has been affecting millions of people across nations, leading to a pandemic. We study the impact of climatic conditions, fiscal and developmental indices, population parameters and environmental conditions of a country on COVID-19 spread. Using JHU CSSE COVID-19 time-series dataset, we find that while the north temperate zone shows the highest increase in confirmed cases and deaths, the tropical zone has shown the highest recovery rate. The polar zone is negligibly affected by the virus. The correlation between number of cases and gross domestic product is 0.91. The Cases Per Population (CPP) and the country's human development index show an unexpected positive correlation of 0.73. There is a negligible correlation of -0.11 between population density and the total cases. The correlation between urban population percent and CPP is 0.58. A weak correlation of 0.23 is obtained between urban particulate matter and total cases.

Keywords: COVID-19, zones, gross domestic product, human development index, population density, urban population percent, particulate matter, regression, correlation

1. Introduction

The COVID-19 pandemic is the defining global health crisis of our time and the greatest challenge we have faced since World War Two. Since its emergence in December 2019 in Asia, it has led to a dramatic loss of human life worldwide and presents an unprecedented challenge to public health, food systems and the world of work. This paper presents an exploratory data analysis conducted to find the effect of various factors such as a country's climatic zone, its Gross Domestic Product (GDP), population parameters, its Human Development Index (HDI) and pollution (particulate matter in the air) on their rising COVID-19 cases.

The following definitions will be helpful in understanding our paper better. The Earth has three main climatic zones: tropical, temperate and polar zones. We classify the

countries based on their latitudes into six different climatic zones, as given in Table 1. Gross Domestic Product (GDP) is the total monetary value of all the goods and services produced within the country in a year. We use the GDP value for the year 2019 as it would explain the country's economic state before the pandemic started. The Human Development index is a measure of the three important aspects required for human development: a long and healthy life, access to education and a decent standard of living [1]. Population density is the number of people per unit area in a region [2]. Urban population percent is the percentage of people living in urban areas out of the total population in a country. Particulate matter 2.5 (PM2.5) refers to tiny particles or droplets in the air that are two and one half microns or less in width. Exposure to high concentration of PM2.5 for a long period can be dangerous to a person's life. Here we use the PM2.5 % present in the

air quality for the year 2016 as no other recent dataset of all countries is available.

In our paper we have determined how the total number of COVID-19 cases worldwide vary by climatic zones. The countries have been split into six different climatic zones based on their latitudes as shown in Table 1 and Figure 1:

Table 1 shows zones and their latitudes

Climatic zone	Latitude
north temperate zone (NTE)	30 ⁰ N to 66.5 ⁰ N
south temperate zone (STE)	30 ⁰ S to 66.5 ⁰ S
north subtropical zone (NST)	23.5 ⁰ N to 30 ⁰ N
south subtropical zone (SST)	23.5 ⁰ S to 30 ⁰ S
tropical zone (TZ)	23.5 ⁰ N to 23.5 ⁰ S
polar zone (PZ)	66.5 ⁰ N to 90 ⁰ N and 66.5 ⁰ S to 90 ⁰ S

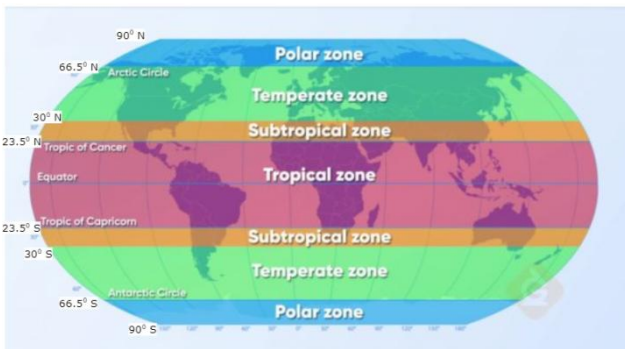


Figure 1 shows classification of zones

To get a better understanding of how all the external factors are related to the total number of COVID-19 cases, the most contributing countries alone were taken into consideration. These countries had their total cases above the average COVID-19 cases.

- The study demonstrates how the Gross Domestic Product (GDP) of each country is dependent on their total number of cases.
- The relation between total number of cases and population parameters country-wise was obtained.
- An analysis of how the particulate matter in the air of each country affects the spread of the virus was performed.
- We also analysed how the Human Development Index (HDI) affects the total number of cases per population in each country.

2. Literature Review

Malay Pramanik[3] studied the impact of climate on COVID-19 transmission over 228 cities around the world. Climatic factors average temperature, diurnal temperature change, and relative humidity were negatively associated whereas temperature seasonality was positively associated with the total number of cases. Through analysis they were able to observe the following as shown in Table 2:

Table 2 show zones and their significant climatic contributors

Zone	Significant climatic contributors of COVID-19 transmission
temperate	average temperature (42.9%), average relative humidity (25.9%)
tropical	mean diurnal range (52.2%) and temperature seasonality (30.8%)
subtropical	average temperature (61.7%) and relative humidity (17.5%)

Miguel B.Araújo [4] performed an analysis on the spread of COVID-19, and how it is constrained by climatic factors. They noted that the virus thrives in cool and dry conditions and detests extremely cold and very hot and wet conditions.

Priyanka Bajaj et.al [5] studied the SARS-CoV2 evolution in response to different latitudes and climatic conditions (Koppen's climate). When plotting per million active cases with respect to climatic zones, they got the results as seen in Table 3:

Table 3 shows zones and their percentage of per million active cases

Zone	Percentage of per million active cases
NTZ (30°N-66.5°N)	81%
NSTZ (23.5°N-30°N)	4%
TZ (23.5°N-23.5°S)	14%
STZ (30°S-66.6°S)	1%
Others	<0.5%

The two strains of the virus are G1 and G2. G1 strain favours Koppen's C climate, while G2 evolved to overcome the climatic constraints of G1 and is able to

survive in Koppen's D and A climate. As a result, SARS-CoV2 was able to spread to a large extent.

ArunavaBhadra et. al [6] found a moderate association between Covid-19 spread and population density after a detailed correlation and regression analysis of infection and mortality rates due to Covid-19 at the district level in India.

KushanTharukaLulbaddaet. al [7] explored the association between the spread of COVID-19 of 58 different countries and the temperature, population size, and median age and found that they are positively associated with the spreading rate of COVID-19. There is no evidence supporting that case counts of COVID-19 could decline in countries with better health care facilities.

DimitarValev[8] studied the relation between total COVID-19 cases and deaths per million populations with 10 demographic, economic and social indicators in 45 countries. Out of these 10 indicators, those of interest to us are:

1. Population Density
2. Human Development Index
3. GDP PPP per capita

The resulted obtained when the above indicators were analysed with Log cases are shown in Table 4:

Table 4 shows Indicator and their corresponding correlation coefficient

Indicator	Correlation coefficient
Human Development Index	0.724
Log GDP PPP per capita	0.687

Population Density	No statistically significant correlation
--------------------	--

Taking data from WHO reports, ZaherKhazaei et.al.[9] examined the relationship between the incidence rate of the disease and its relation to HDI in different countries. They obtained a significant positive correlation between the incidence of COVID-19 cases per million population and the HDI in the world ($r = 0.470$, $P < 0.0001$).

3. Datasets Used in this Study

The John Hopkins University's repository for the 2019 Novel Coronavirus Visual Dashboard is used to obtain the time series dataset[10]. We extract the country/region, their respective latitudes, and the number of confirmed and death cases from January 22nd 2020 to March 28th 2021. The data for the recovered cases is taken from the same from January 22nd 2020 until December 14th 2020 as there are multiple missing values present in the time series recovered data set. Confirmed cases, recovered cases and deaths due to COVID-19 country-wise are visualised in Figure 2, 3 and 4. Heat map of confirmed cases on 22/01/20 and 02/07/20 can be seen in Figure 5 and 6, respectively as well.

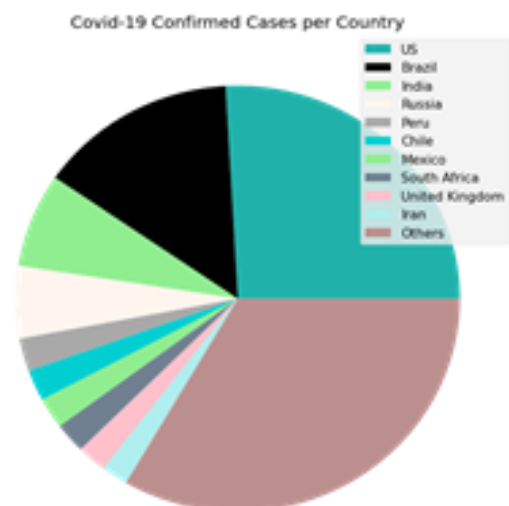


Figure 2 shows confirmed cases country-wise

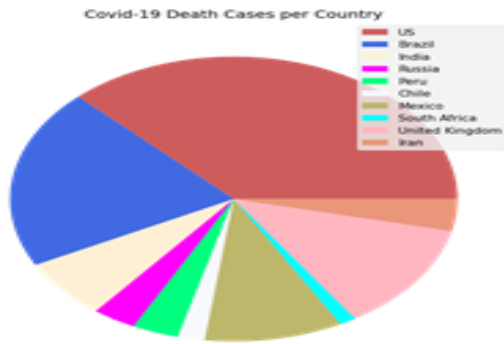


Figure 3 shows recovered cases country-wise



Figure 4 shows death cases country-wise

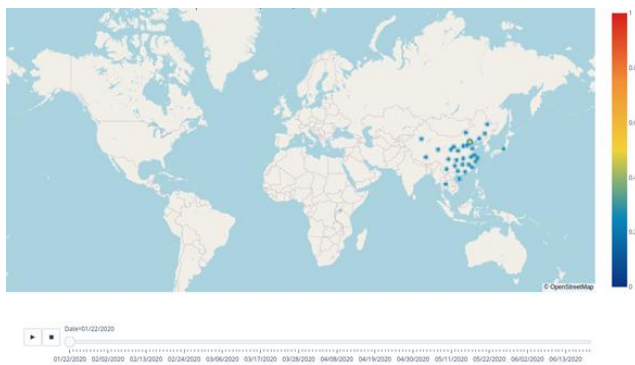


Figure 5 shows the heat map of cases on 22nd January 2020



Figure 6 shows the heap map of cases on 2nd June 2020

The Gross Density Product (GDP) dataset is taken from the World Bank [11]. Extracting only the country name and their GDP for the year 2019, the data is preprocessed by dropping unwanted columns and changing a few country names to match those present in the cases dataset. GDP of various countries can be seen in Figure 7 and its heat map can be visualized in Figure 8.

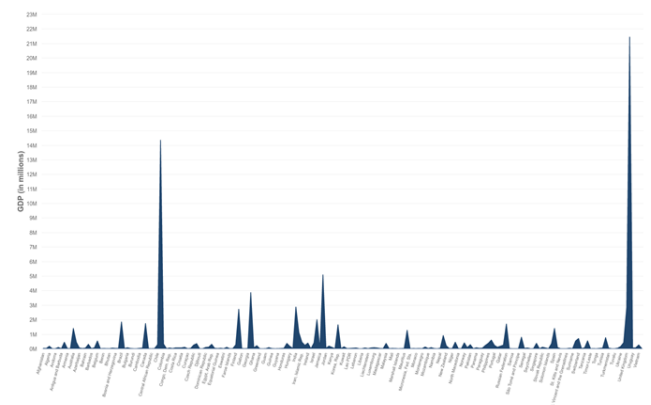


Figure 7 shows country-wise GDP

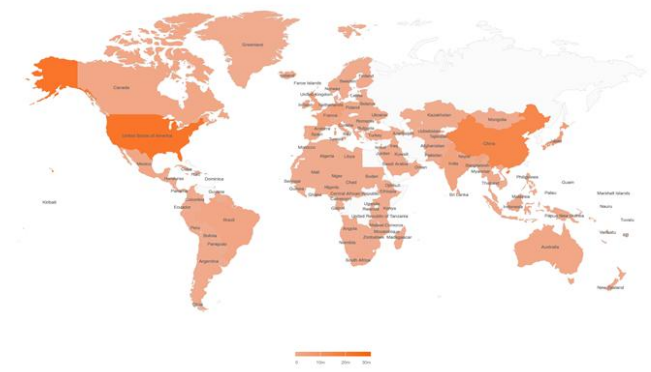


Figure 8 shows the heat of country-wise GDP

The 2020 Population density World Bank dataset [12] is used for calculating the relationship of the various population parameters to the total COVID-19 cases. Heat map and population country-wise can be seen in Figure 9 and 10 respectively. The population density, total population and the urban population percentage fields are utilized and the total cases per population for each country is found by dividing the total cases of that country by their population in the year 2020. The urban

population percent data is later right-stripped to remove the ‘%’ sign and the string variables are converted into integers.

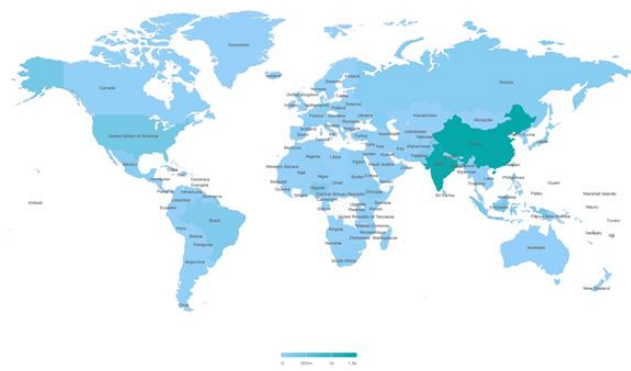


Figure 9 shows the heat map of population country-wise

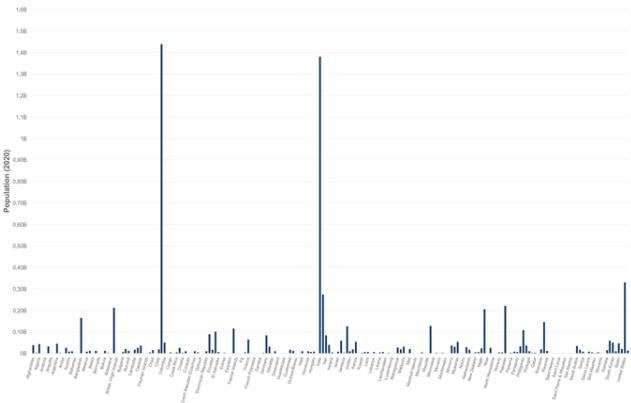


Figure 10 shows country-wise population

The data for the concentrations of fine particulate matter PM2.5 was obtained from the World Health Organization for the year 2016. The data was pre-processed by taking the average of the upper and lower limits and calculated for each country. Below in Figure 11 we observe the air quality of the various countries.

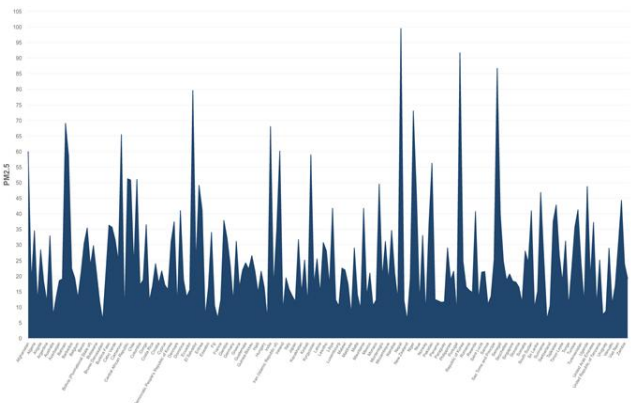


Figure 11 shows PM2.5 of each country

Human Development Index (HDI) value ranges from a scale of 0-1. Higher the value, the better the country in its overall key dimensions which include: a long and healthy life, access to education and a decent standard of living. The scores for the three HDI dimension indices are then aggregated into a composite index using geometric mean. The HDI dataset is obtained from United Nations Development Programme, Human Development Reports for the year 2019[13]. Each country’s HDI is visualised in Figure 12. This dataset is then preprocessed by left-stripping the space in the Country column and renaming a few countries to merge with other datasets, to study the relation between HDI and cases per population.

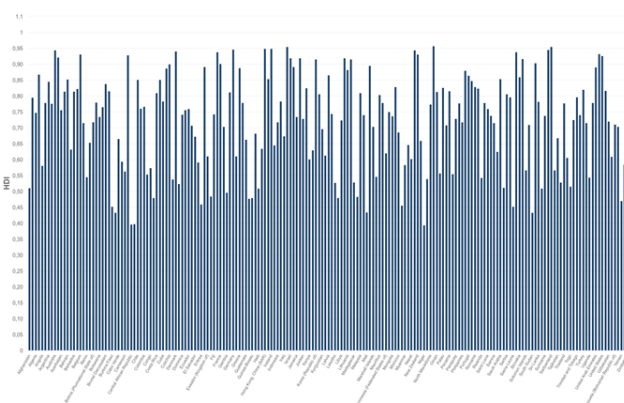


Figure 12 shows HDI of each country

The existing data is further filtered by taking only the major contributing countries which have their total cases above the average cases of the whole world. This list of countries is shown in Figure 13.

	Country/Region	Population	Population Density	Urban Pop %	Total cases	Zone	GDP (in millions)	Cases/Population	HDI
0	India	1382345085	454	35	12039644	TZ	2889929	0.008710	0.645
1	US	331341050	38	83	30264493	NTE	21433226	0.091339	0.926
2	Indonesia	274021804	151	56	1496085	TZ	1119151	0.005480	0.718
3	Brazil	212621698	25	88	12534688	TZ	1839758	0.058898	0.766
4	Mexico	129185028	85	84	2229550	NST	1286871	0.017238	0.779
5	Philippines	109830324	368	47	721892	TZ	376795	0.006573	0.718
6	Turkey	84495243	110	76	3208173	NTE	761425	0.037969	0.820
7	Germany	83830972	240	76	2784852	NTE	3881124	0.033217	0.947
8	United Kingdom	67948282	281	83	4370000	NTE	2829108	0.064314	0.932
9	Italy	60446035	208	69	3632067	NTE	2003578	0.058433	0.892
10	South Africa	59436725	49	67	1545431	STE	351432	0.026001	0.709
11	Colombia	50975248	45	80	2382730	TZ	323516	0.045742	0.787
12	Spain	45757980	94	80	3255324	NTE	1360491	0.069521	0.904
13	Argentina	45267449	17	93	2308597	STE	445445	0.050999	0.845
14	Ukraine	43886577	75	69	1694568	NTE	153781	0.038789	0.779
15	Iraq	40372771	93	73	832428	NTE	234004	0.020910	0.874
16	Poland	37839255	124	60	2250991	NTE	565858	0.056488	0.880
17	Canada	37799407	4	81	1040000	NTE	1736428	0.027514	0.929
18	Peru	33050211	26	79	1526882	TZ	226848	0.048290	0.777
19	Romania	19214508	84	55	939516	NTE	250077	0.048745	0.828
20	Chile	19144505	29	85	977243	STE	282318	0.051045	0.851
21	Belgium	11595451	383	98	870757	NTE	533067	0.075075	0.931
22	Portugal	10191409	111	66	820407	NTE	238785	0.080000	0.864
23	Sweden	10110233	25	88	780018	NTE	530884	0.077151	0.945
24	Israel	8878517	400	93	831924	NTE	394552	0.095880	0.919

Figure 13 shows the most contributing countries

After obtaining these datasets and preprocessing them, they are used for several analyses and their results can be observed under Experimental Results.

Correlation and Regression Analysis - The Experimental Study

In our paper we have employed correlation and regression to analyse the dependence of the various external factors to the increasing COVID-19 cases worldwide.

Correlation : The correlation coefficient is determined by dividing the covariance by the product of the standard deviations of the two variables [14] as given in (1).

$$\rho_{xy} = \frac{\text{Cov}(x,y)}{\sigma_x \sigma_y} \quad (1)$$

where,

ρ_{xy} = Pearson product-moment correlation coefficient

$\text{Cov}(x,y)$ = covariance of variables x and y

σ_x = standard deviation of x

σ_y = standard deviation of y

Standard deviation is a measure of the dispersion of data from its mean and covariance is a measure of how two random variables change together. By dividing covariance by the product of the two standard deviations, one calculates the correlation coefficient.

The correlation coefficient is a statistical measure which determines the strength and direction of the linear relationship between two variables. It ranges between -1.0 and 1.0. If it is greater than 1.0 or less than -1.0 indicates an error in the correlation measurement. A correlation of -1.0 shows a perfect negative correlation, meaning the variables move in opposite directions—for an increase in one variable, there is a decrease in the second variable or vice-versa. While a correlation of 1.0 shows a perfect positive correlation, for a positive increase in one variable, there is also a positive increase in the second variable. A correlation of 0.0 implies there

is no linear relationship between the two variables. Some of its limitations is that it cannot capture nonlinear relationships between two variables and cannot differentiate between dependent and independent variables.

The correlation coefficient is calculated between the total confirmed cases country wise and their GDP to find the relation between them. A high correlation coefficient would indicate that the country's GDP played a huge role in their total COVID-19 cases. To indicate how well the cases were dependent on the various population parameters the correlation coefficient is found between urban population percent and population density to the country's total cases per population. It is also calculated for total cases in the country to their total population for the year 2020. One can also understand how a country's overall education, standard of living and health, which is basically their HDI, impacts the total number of cases per population in that country.

Regression analysis is a statistical method that helps us to analyze and understand the relationship between two or more variables of interest[15]. This method helps us understand which factors influence each other, which can be ignored and which factors are the most important.

The variable that we are trying to understand or forecast is called the dependent variable. The factors that influence the analysis or target variable and provide us with information about the relationship between the variables and the target variable is called the independent variable.

Simple linear regression is a model that assesses the relationship between a dependent variable and an independent variable [16]. The simple linear model is expressed using (2):

$$Y = a + bX + \epsilon \quad (2)$$

where, Y – Dependent variable, X – Independent (explanatory) variable, a – Intercept, b – Slope and ϵ – Residual (error)

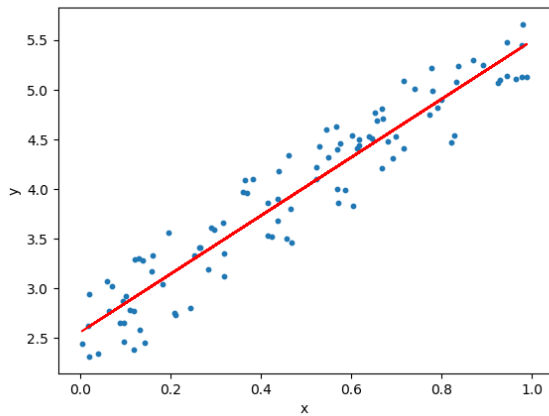


Figure 14 shows regression plot

Making use of a simple linear regression model, regression analysis is done to find the relation between the following parameters country-wise:

GDP and total cases

Population density and total cases

HDI and cases per population

Urban population percent cases per population

Here the former is the dependent variable and the latter is the independent variable.

4. Experimental Results

We obtain the following graphs when we plot the time series graph of confirmed and death cases from 22nd January 2020 to 28th March 2021, and recovered cases from 22nd January 2020 to 14th December 2020.

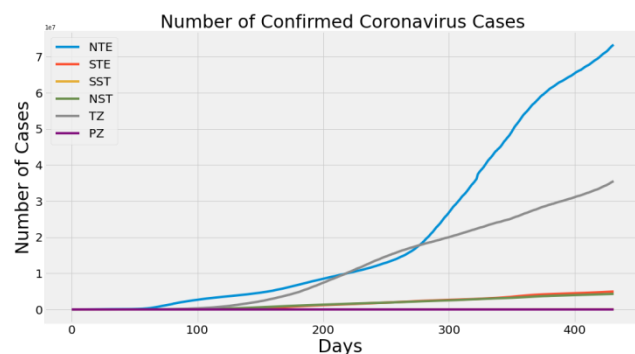


Figure 15 shows the time series graph for confirmed cases

The Northtemperate zone (NTE) has the highest number of total cases as well as the highest rate of growth. This is followed by tropical zone (TZ), south temperate zone (STE), north subtropical zone (NST) and finally at the bottom we have south subtropical zone (SST) and polar zone (PZ).

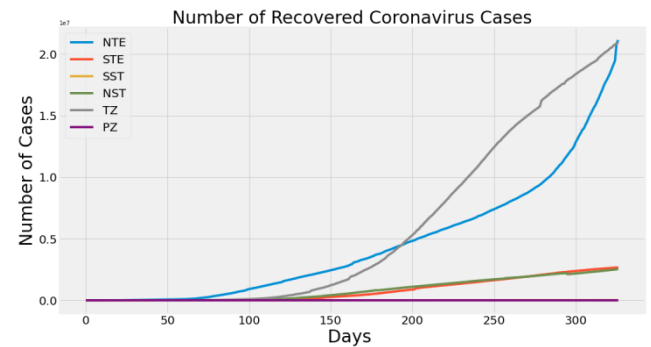


Figure 16 shows the time series graph for recovered cases

When it comes to recovery rate, TZ is in the lead despite the fact that NTE has the highest instances of COVID-19 cases. Most of the under-developing and

developing nations lie in TZ and in a study it was revealed that people from such low income countries tend to be exposed to more pathogens and as a result they have better immunity response[17]. This could explain why TZ has the highest recovery rate. Following NTE, we have STE, NST, SST and PZ.

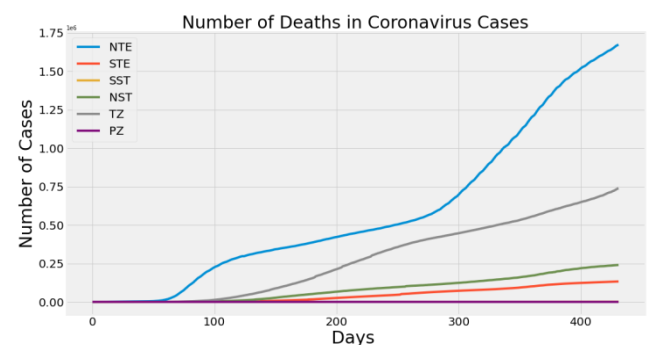


Figure 17 shows the time series graph for deaths due to COVID-19

For deaths due to COVID-19, highest casualty is observed in NTE, followed by TZ, NST, STE, SST and PZ.

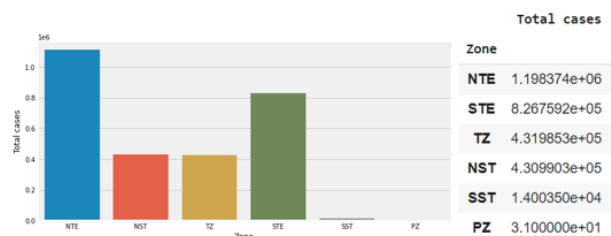


Figure 18 shows the average cases in each zone

In Figure 18,, we observe that NTE has the highest average cases in each zone, followed by STE, TZ, NST, SST and PZ.

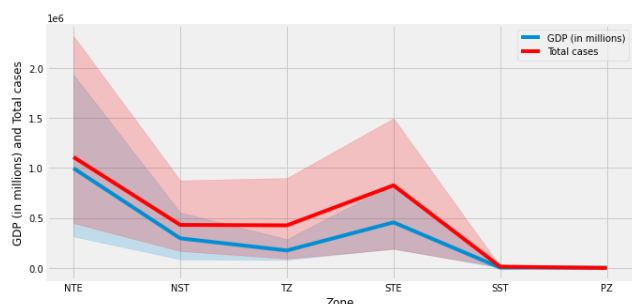


Figure 19 shows average cases vs average GDP in each zone

We notice here that GDP and total cases follow the same trend. This trajectory could explain the high instances of cases in NTE as it contains most of the developed nations implying high GDPs.

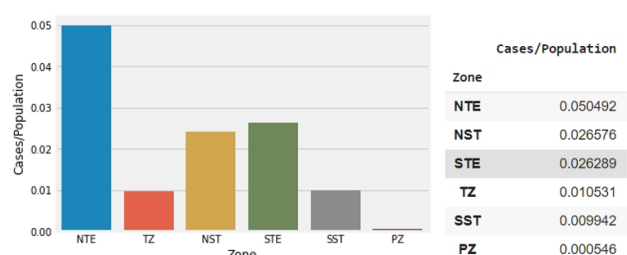


Figure 20 shows average cases per population in each zone

The plot above shows average cases per population in each zone and here we are able to visibly discern the difference between SST and PZ. This goes on to say at all times PZ is the least affected zone. This could be attributed to its harsh cold climate, where the average temperature is less than 10 °C.

Table 5 shows average total cases and cases per population zone wise in decreasing order

Average total cases	Average cases per population
NTE	NTE
STE	NST
TZ	STE
NST	TZ
SST	SST
PZ	PZ

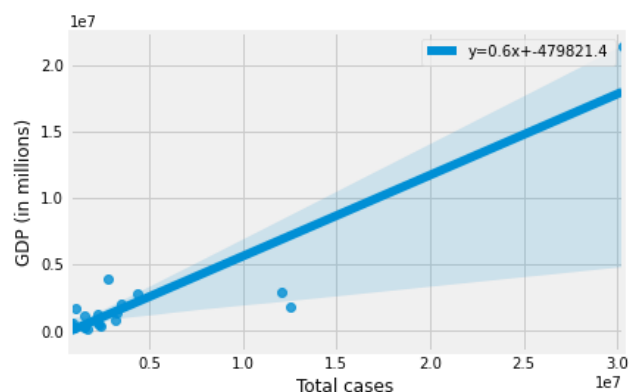


Figure 21 shows regression plot of GDP vs Total cases

	Total cases	GDP (in millions)
Total cases	1.000000	0.914468
GDP (in millions)	0.914468	1.000000

Figure 22 shows the correlation between total cases and GDP

Computing the correlation between GDP and the total cases in all the countries worldwide a correlation coefficient of 0.914 is obtained as seen in Figure 22. This

value indicates that countries with high GDP have seen more COVID-19 cases and countries with low GDP had fewer cases. This can be explained by the fact that countries with high GDP are able to conduct more tests and hence end up getting more confirmed cases. It can also be understood that since a country has high GDP it would have more economic activities which implies that more gatherings and crowds would have taken place which resulted in the further spreading of the virus. From Figure 21 we observe that the total cases and GDP have a linear relationship as the line is almost at 45°.

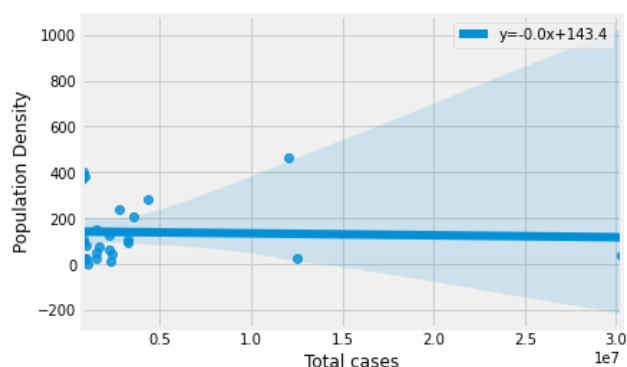


Figure 23 shows the regression plot of population density vs total cases

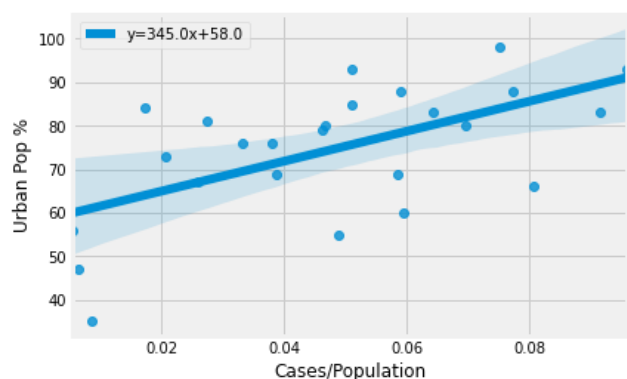


Figure 24 shows the regression plot of urban population percent vs cases per population

	Cases/Population	Urban Pop %
Cases/Population	1.000000	0.587612
Urban Pop %	0.587612	1.000000

Figure 25 shows correlation between cases/population and urban population percent

As we observe from Figure 25, the correlation between the total cases and population density of each country is achieved as -0.102. As the value is extremely small and almost near to zero we can conclude that there is no statistical relation between the population density and the COVID-19 cases present in a country. The Figure 23 also shows us that there is hardly any relation between the two. This may be due to the reason that the population density is calculated for the entire country, including all the dense and sparsely populated areas which include crowded cities, hilly mountainous regions, forests and so on. As this is the case, we then see how much the urban population percent is related to the total cases per population country wise. Achieving a correlation coefficient of 0.587 we understand that there is a relatively better relation between them, which is also visually depicted in Figure 24. As urban areas are the hotspots in the country and are usually the most busiest and crowded places, there is a higher possibility of people being infected by the virus.

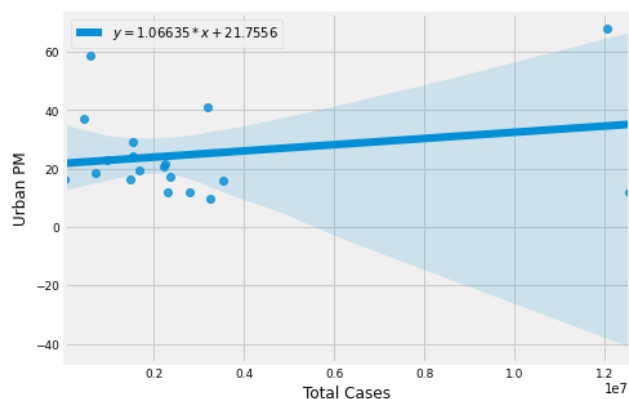


Figure 26 shows urban particulate matter vs total cases

To check if the particulate matter in the air of each country helps in the spreading of the virus we calculate the correlation coefficient. A value of just 0.23 could lead to a conclusion that the country's air quality alone does not affect the spread of the coronavirus, since transmission of the virus by air alone may not be the cause of the spread. From Figure 26 we observe that there is hardly any relation between the two. However, here too we have access to only the particulate matter of an entire country, and not specifically crowded urban areas.

Having an urban-rural divide in the data could throw better light on this.

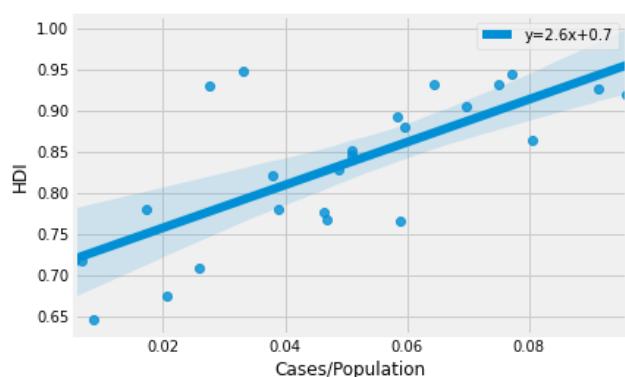


Figure 27 shows regression plot of HDI vs cases per population

	Cases/Population	HDI
Cases/Population	1.000000	0.731478
HDI	0.731478	1.000000

Figure 28 shows correlation between cases/population and HDI

While calculating the correlation coefficient for Human Development Index (HDI) and the cases per population an unexpected positive correlation of 0.731 is found as mentioned in Figure 28. This means that a country with a high HDI gives more COVID-19 cases which is visually represented by Figure 27. At first it might be surprising to see such a good positive relation but HDI is actually a cumulative index of three factors. A high HDI country implies that the people have better access to educational institutions, entertainment zones, supermarkets and transport facilities. The crowding and frequent visits of the people to the aforementioned areas could lead to the increase in the spread of the coronavirus.

5. Discussion on the results

Our study gives a novel insight into the spread of COVID-19 based on geographic, economic, demographic and health indicators. The findings of this study can be useful to provide constructive instructions for governments on making efficient and effective public health protection decisions. The knowledge about

different zones' susceptibility to this kind of viruses can allow drawing tailored preventive approaches based on such specificities to avoid, or reduce, future relapse in the COVID-19 pandemic or in future, somewhat similar conditions that might occur [18]. Owing to the result obtained from our research, that higher GDP countries have a higher number of cases, the countries with a high GDP must undertake appropriate measures to distribute the vaccines faster and throughout the country. In addition to that, these high income countries have more urban areas resulting in faster transmission of the virus, hence stricter lockdown measures must be employed.

6. Conclusion

The spread of COVID-19 has turned into a worldwide pandemic thus far [19]. Therefore analysing the various external factors such as GDP, population parameters and HDI could contribute to a better understanding of the increase in the cases worldwide. After studying the relation between climatic zones and total cases it was found that, the most affected and least affected zones are the north temperate zone and polar zones, respectively. The zone with highest recovery is the tropical zone. In addition to this, correlation between GDP and total number of cases is 0.914 and there exists a linear relationship between them. Surprisingly, there is no correlation between the population density and total number of cases in a country. However, urban population percent and cases per population in a country are associated and have a correlation coefficient of 0.587. As we took the particulate matter percentage of the entire country only for the year 2016 a very negligible correlation was found between the number of cases in that country to their PM%. Once again unexpectedly, HDI and cases per population are related by a correlation factor of 0.731.

Further improvements on our study can be done to predict the number of cases daily using state of the art machine learning models and strive to get a good accuracy. The time series analysis can be further studied to determine how effective were the lockdowns in keeping the spread of the virus under control. One can also further analyse how the coronavirus impacted the GDP and HDI of a

country. The impact of the air quality can also be accurately found using newer and cleaner data.

References

- [1] Max Roser "Human Development Index (HDI)". Published online at OurWorldInData.org. Retrieved from: '<https://ourworldindata.org/human-development-index>' [Online Resource] (2014)
- [2] Wikipedia contributors. Population density. In Wikipedia, The Free Encyclopedia. Retrieved, from https://en.wikipedia.org/w/index.php?title=Population_density&oldid=1012533226(2021).
- [3] Malay Pramanik et al. Climatic influence on the magnitude of COVID-19 outbreak: a stochastic model-based global analysis, International Journal of Environmental Health Research, DOI: 10.1080/09603123.2020.1831446(2020)
- [4] BabakNaimi and Miguel B. Araújo .Spread of SARS-CoV-2 Coronavirus likely constrained by climate. medRxiv 2020.03.12.20034728; doi: <https://doi.org/10.1101/2020.03.12.20034728>(2020)
- [5] Priyanka Bajaj and Prakash Chandra .Climatic-niche evolution of SARS CoV-2bioRxiv 2020.06.18.147074; doi: <https://doi.org/10.1101/2020.06.18.147074>Arya (2020)
- [6] Bhadra A. et al. Impact of population density on Covid-19 infected and mortality rate in India. Model. Earth Syst. Environ. 7, 623–629. <https://doi.org/10.1007/s40808-020-00984-7>(2021)
- [7] KushanTharukaLulbadda et alThe impact of temperature, population size and median age on COVID-19 (SARS-CoV-2) outbreak, Clinical Epidemiology and Global Health, Volume 9, 2021, Pages 231-236,ISSN 2213-3984, <https://doi.org/10.1016/j.cegh.2020.09.004>. (2021)
- [8] DimitarValev Relationships of total COVID-19 cases and deaths with ten demographic, economic and social indicators, medRxiv 2020.09.05.20188953; <https://doi.org/10.1101/2020.09.05.20188953>(2020)
- [9] Khazaei et al. COVID-19 Pandemic in the World and its Relation to Human Development Index: A Global Study. Archives of Clinical Infectious Diseases. In Press. 10.5812/archcid.103093. (2020)
- [10] Dong E, Du H and Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Inf Dis. 20(5):533-534. doi: 10.1016/S1473-3099(20)30120-1 (2020)
- [11] The World Bank, GDP (current US\$), Available at: <https://data.worldbank.org/indicator/NY.GDP.MKTP.CD>(2019)
- [12] The World Bank, Population density (people per sq. km of land area) , Available at:<https://data.worldbank.org/indicator/EN.POP.DNST>(2020)
- [13] Human Development Data Center Human Development Index Available at: <http://hdr.undp.org/en/indicators/137506>(2020)
- [14] Jason Fernando, Toby Walters, Correlation Coefficient, Available at: <https://www.investopedia.com/terms/c/correlationcoefficient.asp> (2021)
- [15] Great Learning Team What is Regression? Definition of Regression?, Available at: <https://www.mygreatlearning.com/blog/what-is-regression/> (2020)
- [16] CFI Education Inc , Regression Analysis, Available at: <https://corporatefinanceinstitute.com/resources/knowledge/finance/regression-analysis/>(2019)
- [17] Sakshi Post, Previous Why COVID Recovery Rate Is Highest In India, Available at: <https://english.sakshi.com/news/lifestyle/why-covid-recovery-rate-highest-india-125918>(2020)
- [18] Gangemi, S., Billeci, L. &Tonacci, A., Rich at risk: socio-economic drivers of COVID-19 pandemic spread. ClinMol Allergy 18, 12 <https://doi.org/10.1186/s12948-020-00127-4>(2020).
- [19] Coronavirus Disease (COVID-19), Events as They Happen, Mar. [online] Available at: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-asthey-happen> (2020)
- [20] T. Wang, K. Lu, K. P. Chow and Q. Zhu. "COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model," in IEEE Access, vol. 8, pp. 138162-138169, doi: 10.1109/ACCESS.2020.3012595. , (2020)
- [21] Comunian S, Dongo D, Milani C, Palestini P. Air Pollution and Covid-19: The Role of Particulate Matter in the Spread and Increase of Covid-19's Morbidity and Mortality. Int J Environ Res Public Health. 17(12):4487. doi: 10.3390/ijerph17124487. (2020)