



Research Article

Descriptive Analysis of Pakistan's COVID Data

Mian Ibad Ali Shah*, Yasir Saleem Afridi and Laiq Hasan

Department of Computer Systems Engineering, University of Engineering and Technology, Peshawar, Khyber Pakhtunkhwa, Pakistan.

Abstract: Strategic information was requested to understand the novel coronavirus using a global call. In response to but not limited to the call and extending the work to similar global pandemics, this work has been conducted on a dataset acquired from Pakistan's COVID data. Analysis of the statistics and trends has been carried out and recommendations regarding the precautionary measures and implementation of controlling techniques for new cases using controllable factors have been provided. The results have been achieved using descriptive analysis of the time-series data. The goal was to check the effect of various parameters (like new tests, government implemented lockdown index, etc.) on the new/ active COVID-19 cases, so policies can be devised accordingly to control the spread of the disease. The data was first converted into summary numbers, followed by carrying out the statistical analysis that yielded meaningful results. Descriptive analysis in particular performs best to achieve this strategy. Descriptive analysis comprising of various methods and factors have been used to infer conclusions from data. Some important factors like, stringency index, number of tests, etc. have been utilized in this paper to achieve the aforementioned goals.

Received: November 23, 2020; **Accepted:** June 20, 2021; **Published:** June 28, 2021

***Correspondence:** Mian Ibad Ali Shah, Department of Computer Systems Engineering, University of Engineering and Technology, Peshawar, Khyber Pakhtunkhwa, Pakistan; **Email:** mianibad@outlook.com

Citation: Shah, M.I.A., Y.S. Afridi and L. Hasan. 2021. Descriptive analysis of Pakistan's COVID data. *Journal of Engineering and Applied Sciences*, 40(1): 15-23.

DOI: <https://dx.doi.org/10.17582/journal.jeas/40.1.15.23>

Keywords: COVID Pakistan, Coronavirus, Descriptive analysis, Time-series data analysis, Data analytics

Introduction

The Corona Virus Disease 2019 (COVID-19), the name COVID unknown before, first outbreak in the Wuhan City of China, in December 2019 (Chakraborty and Ghosh, 2020; Guan *et al.*, 2020). The matter straight away resulted in raising international concerns. In March 2020, the COVID was declared as a pandemic by the World Health Organization (WHO) (Spinelli and Pellino, 2020). 213 countries around the world reported COVID active cases, hence an international public health emergency was declared. As of 18th September 2020, the total number of reported COVID cases globally are 30,193,188 whereas, the total deaths are 946,490

(WHO, 2020). Pakistan in particular, has reported a total of 304,386 cases with a death toll of 6,408. The total number of recovered COVID patients are 291,683 (Hannah, 2020). The global data including the total number of cases, total active cases, total recovered patients and total deaths' data has since been collected on daily basis and added to various repositories (mostly open access) for researchers to work on a solution using their respective domains. This research comprehensively studies and analyzes Pakistan's COVID data from 31st December 2019 till 19th August 2020 by using various statistical techniques. The data set downloaded from the website (www.ourworldindata.org) (Hannah, 2020) has been discussed and analyzed in this paper. Data

visualization and descriptive statistics/ analysis has been conducted to achieve fruitful results. The trends of various features enable us to better understand the prevailing COVID situation. Other statistical features, discussed in section 2, also played a vital role in understanding the data. The results have shown a relationship between the effect of stringency index (Government's response on closure of places and travel bans) on new number of corona cases. For simplicity, the contributions of this research are as under:

- Analyzes Pakistan's COVID data by using various statistical techniques
- Descriptive analysis has been done to understand the trends and meanings of the data technically
- COVID-19 related parameters (mentioned in Table 1) have been analyzed to understand its impact on the spread of the virus
- Based on the results, important decisions for policy making have been advised

Literature review

COVID data, particularly of Pakistan, has been analyzed in many research papers in different ways. Saqlain *et al.* (2020) has worked on the attitude of health workers of Pakistan using the data. Shafi *et al.* (2020) has used Pakistan's COVID data to observe and analyze the pandemic effect on Small to Medium Enterprises (SMEs) that are operating in the country. Work done on sentiment analysis based on COVID-19 data has been discussed in (Ahmad *et al.*, 2021). It focuses on the sentiment analysis of users' reviews on the contact tracing of COVID mobile applications. After analysis, it shows how users react to such applications. Majority of reviews are concerned over location security and technical issues in the applications, hence less use of such applications. Likewise, in (Hamid *et al.*, 2020), the authors have worked on social media posts/ tweets authenticity detection of COVID-19 related posts. GNN algorithm has been used for this purpose. Olapegba *et al.* (2020) has conducted somewhat similar studies on Nigerian COVID dataset to study risk perceptions and precautionary behaviors of Nigerians using questionnaires. In this paper, the expected time of end the lockdown (stringency index in our case) was predicted and prescribed to end in May 2020, however, in June 2020 the rise in COVID cases was much higher than before. Hence, it infers that stringency index/ lockdown is one of the most important parameters to look at Qayyum *et al.* (2021) has analyzed the potential of intelligent

processing of clinical visual data of COVID-19. They have used machine learning for the purpose, which is the upcoming step of this research work (LSTM algorithm to be used). Ahmed (2020) has studied the impact of this pandemic on Pakistan's stock exchange using Pakistan's data. Sarkodie and Owusu (2020), data driven models have been created to study the linear relationship between number of deaths and number of confirmed cases. Saqlain *et al.* (2020) has concluded their work by showing results based on attitude and knowledge of nurses, doctors and pharmacists of Pakistan. Descriptive statistics has been conducted on geolocation data of Brazil by Peixoto *et al.* (2020) to achieve prediction of future spread (new cases) using moving dynamics. Hotspot analysis has been done using Pueyo predictive model in (Saeed *et al.*, 2021) to identify lockdown areas in Punjab province of Pakistan. Statistical analysis on Pakistan's COVID data using Auto-Regressive Integrated Moving Average method has been done in (Yousaf *et al.*, 2020) where new COVID cases and deaths for upcoming month has been predicted. However, they have not mentioned the reason for the spread. Our paper, in particular focuses on the reason for virus spread in Pakistan through descriptive study and analysis that will facilitate the strategic decision making for controlling the factors to mitigate the future pandemic spread.

Dataset and features definitions

The dataset of the COVID-19 number of cases/ deaths has been downloaded from an open access data source of our world in data's website (Hannah, 2020). The dataset in this website represents the worldwide data, updated on daily basis. This time series dataset has a total of 40 features and 238 records per country (based on the dates). The 40 features or variables in the dataset shows the condition and numbers of various parameters that are associated with COVID-19 disease in Pakistan. The numbers of the parameters are overall affected by the increase and decrease in the COVID dependent features. In simple words, if higher number of daily active/ critical new cases of COVID patients are recorded that will lead to an increase in the number of deaths (shown in results). The features are ISO code, Continent, Location, Date, Total cases, New cases, Total deaths, New deaths, Total cases per million, New cases per million, Total deaths per million, New deaths per million, Tests units, Total tests, New tests, New tests smoothed, Total tests per thousand, New tests

per thousand, New tests smoothed per thousand, Stringency index, Population, Population density, Median age, Aged 65 older, Age 70 older, GDP per capita, Extreme poverty, Cardiovascular death rate, Diabetes prevalence, Female smokers, Male smokers, Hand washing facilities, Hospital beds per thousand, Life expectancy. However, out of these only 15 significant features have been used in this research. These 15 features are defined in Table 1. The 15 significant variables that have been selected are because of their higher impact/ affect on new number of cases (that we have to predict/ forecast).

Table 1 includes all the relevant/ significant parameters that affects new number of cases or disease spread.

Descriptive analysis

Descriptive Analysis is the graphical and numerical set of techniques that are used to present and analyze a data set (Fisher and Marshall, 2009). Descriptive analysis is an important step for conducting the statistical analysis. It helps us understand the distribution of data, detects outliers, and identifies the relationships among various variables. In simple words, descriptive statistics provides you with a platform to conduct further statistical analyses. The descriptive analysis includes finding the frequency of data, minimum, maximum, mean, median, mode, data distribution (over 0%, 25%, 75% and soon), variance, standard deviation and skewness etc. Some of the statistical tests are also included in the descriptive analysis like z-tests and t-tests. The description of these methods can be seen in Table 2.

Table 1: Description of dataset features.

Feature	Description
Date	Date of observation
Total cases	Total confirmed cases of COVID-19
New cases	New confirmed cases of COVID-19
Total deaths	Total deaths attributed to COVID-19
New deaths	New deaths attributed to COVID-19
Total cases per million	Total confirmed cases of COVID-19 per 1M people
New cases per million	New confirmed cases of COVID-19 per 1M people
Total deaths per million	Total deaths attributed to COVID-19 per 1M people
New deaths per million	New deaths attributed to COVID-19 per 1M people
Total tests	Total tests for COVID-19
New tests	New tests for COVID-19
New tests smoothed	New tests for COVID-19 (7-day smoothed)
Tests per case	Tests conducted per new confirmed case of COVID-19, given as a rolling 7-day average
Positive Rate	The share of COVID-19 tests that are positive, given as a rolling 7-day average
Stringency index	Govt. Response: Measure based on 9 response indicators including school and workplace closure and travel bans (100 = strictest response)

Source: Ourworldindata.Org

Table 2: Description of descriptive statistical methods.

Statistical Technique	Description
Mode	A value with the greatest frequency
Median	The middle score of an ordered distribution
Mean	The average score
Frequency distribution	Number of cases in a category
Range	Distance between lowest and highest score
Interquartile range	The range within which the middle 50% of the scores fall
Variance	The dispersion of a set of data points around their mean
Standard deviation	Dispersion of a set of points around their mean (Square root of variance)
Skewness	Shows the data concentration on either side

Table 2 defines all the basic statistical parameters that have been used in this research to generate final research results. The formulas of some of these methods are as under:

$$\text{Median: (element at position } n + 1) / 2$$

$$\text{Population variance: } \sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

$$\text{Sample variance: } s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

$$\text{Population Standard deviation: } \sigma = \sqrt{\sigma^2}$$

$$\text{Sample Standard deviation: } s = \sqrt{s^2}$$

Materials and Methods

The time series data from the source (Hannah, 2020) was imported to a python notebook using pandas library. The data was first checked for missing/ corrupted values and outliers. The nulls or missing data was filled using different techniques. Some data was filled manually from news archives and sources like (WHO, 2020; Hannah, 2020). After that, the remaining nulls were filled using pandas library’s function *fillna*. The outliers (sensed by extremely high spikes or long tails in distribution) were removed using simple threshold values, based on the deviation from mean (horizontally and vertically). Then the count (number of entries), minimum, maximum, 25%, 50% and 75% of the data was recorded using pandas library. These calculations helped us finding the start, end and concentration of the data. This is one of the easiest way to know about the values of the features. Measures of central tendency and measures of symmetry were also calculated. The data was then visualized. Some of the features were visualized using python seaborn library. It helped to study the data visually using graphs and histograms for finding the outliers, Standard deviations and level of skewness,

etc. The techniques used and the results generated are discussed in detail in section 3.

Results and Discussion

Conducting the descriptive statistics on nine important features of the COVID data of Pakistan using python and various libraries, the descriptive statistics methods are calculated as in Table 3. This table shows in detail the maximum statistical outcome/ information about this research.

The results obtained in Table 3 reflect that the count (number of records) ranged from 224 to 234. The total count is then set to the minimum value later at the pre-processing stage, for predictive analysis, as the rows with nulls are dropped. Table 3 also shows that the minimum of every feature is 0. This is because in the initial days (till 26th February 2020) there were no reported cases in Pakistan. Hence zero new cases, tests and deaths and stringency index (no SOPs/ closures by Government of Pakistan (GOP)) etc. The mode could not be calculated for any feature, since there are a large number of modes. Following is an individual detailed discussion of each feature.

New cases

New cases are mapped from zero to maximum of 6825 date wise. By looking at the 25%, 50% and 75% values range, it can be said that the maximum value, including few others, are extreme values or outliers, as also seen in Figure 1. The mean and median values helped us in calculating the shape of the feature’s data, called skewness. Where mean is greater than median, it reflects that the dataset is positively skewed. We can verify skewness result in Table 3 i.e. 1.476 (a positive number). Standard deviation is at a value of 1644.77. This implies that our data is spread around the mean by 1644.77 points.

Table 3: Descriptive statistics result.

Feature	Count	Mini- mum	Maxi- mum	25 %	50 %	75%	Mean	Median	Mode	Variance	Standard Deviation	Skewness
New Cases	233	0	6825	0	587	1918	1256.5	587	NA	2705265	1644.769	1.476332
Total Cases	233	0	292765	2	15759	198883	88095.61	15759	NA	1.25E+10	111805.4	0.831841
New Deaths	233	0	153	0	14	39	26.75	14	NA	1214.839	34.85453	1.478029
Total Deaths	233	0	6235	0	346	4035	1841.54	346	NA	5562071	2358.404	0.872547
New Tests	229	0	41666	0	6816	21123	10174.12	6816	NA	1.08E+08	10409.05	0.51125
Total Tests	232	0	2389365	7.5	141256	1108312	578096	141256	NA	5.67E+11	753281	1.066712
Tests per case	224	0	292	0	7.1	12.2	10.15	7.115	NA	450.8604	21.23347	10.78381
Positive Rate	224	0	9.75	0	0.08	0.136	0.129	0.0805	NA	0.422525	0.650019	14.69646
Stringency Index	234	0	96.3	28.7	61.1	82.4	53.58	61.11	NA	1136.367	33.71004	-0.43014

ourworldindata.org

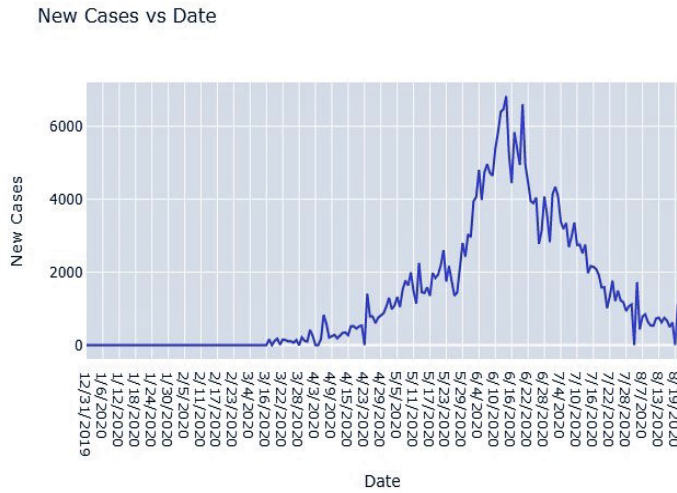


Figure 1: Visualization of new cases per date.

These descriptive results as well as data scatter plot can be observed in Figure 2. This figure shows that most of the data is concentrated in the lower half of the new cases as opposed by the smaller number of values in the upper half. From this, we can infer that new cases of the COVID-19 normally ranged from 0 to 3000, after excluding the outliers.



Figure 2: Stats of new cases.

Total cases

Total cases are mapped from zero to maximum of 292765 date wise. The feature is positively skewed as the mean is greater than median. We can verify skewness result in Table 3 i.e. 0.832 (a positive number). Our data is spread around the mean by 111805.42 points, based on standard deviation value. These descriptive results as well as data scatter plot can be observed in Figure 3. This figure shows that the data concentrated in the lower third and upper third of the total cases is more as opposed by the center third part of data, as it has a smaller number of entries. From this, we can infer that the new cases of the COVID-19 had a slow rise initially. Then in the middle days of the data, the data rose up at a very fast pace. Afterwards, the number of new cases reduced.

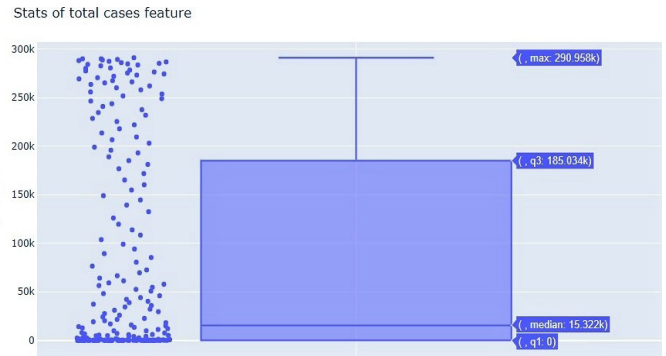


Figure 3: Stats of total cases.

New deaths

New deaths are mapped from zero to maximum of 153 date wise. By looking at the 25%, 50% and 75% values range, it can be said that there are upper outliers in the data, as also seen in Figure 4. The feature is positively skewed as the mean is greater than median. We can verify skewness result in Table 3 i.e. 1.478 (a positive number). Our data is spread around the mean by 34.85 points, based on standard deviation value.

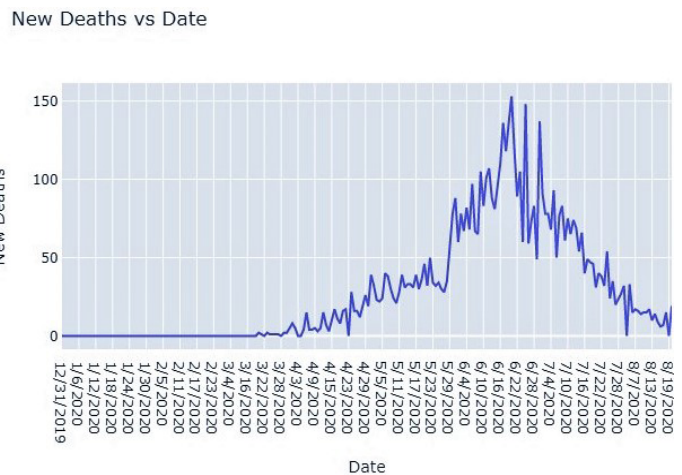


Figure 4: Visualization of new deaths per date.



Figure 5: Stats of new deaths.

These descriptive results as well as data scatter plot can be observed in Figure 5. This figure shows that data concentration in the first two quarters is high. Data concentration reduces in third quarter and reduces drastically in last quarter. From this, we can infer that

new deaths of the COVID-19 normally ranged from 0 to 100, after excluding the outliers.

Total deaths

Total deaths are mapped from zero to maximum of 6235 date wise, as seen in Figure 6. The feature is positively skewed as the mean is greater than median. We can verify skewness result in Table 3 i.e. 0.872 (a positive number). Our data is spread around the mean by 2358.40 points, based on standard deviation value.

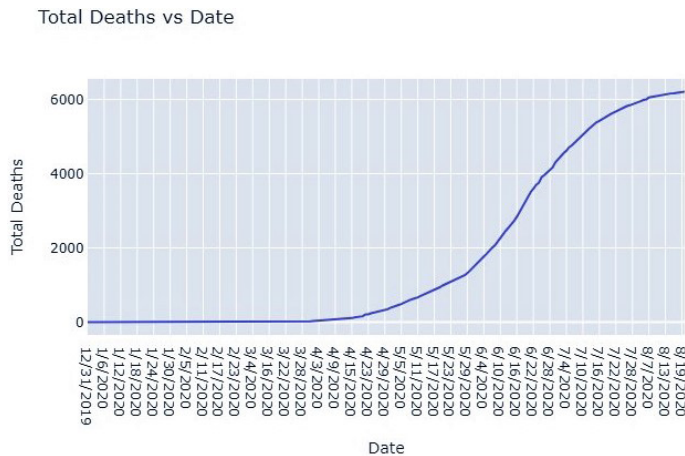


Figure 6: Visualization of total deaths per date.

These descriptive results as well as data scatter plot can be observed in Figure 7. This figure shows that more data is concentrated in the lower third and upper third of the total cases as opposed by the center third part of data, as it has less number of entries. From this, we can infer that the new deaths of the COVID-19 had a slow rise initially. Then in the middle days of the data, the data rose up at a very fast pace. Afterwards, the number of new deaths reduced.



Figure 7: Stats of total deaths.

New tests

New tests are mapped from zero to maximum of 41666 date wise. By looking at the 25%, 50% and 75% values range, it can be said that there are upper outliers in the data, as also seen in Figure 8. The feature is

positively skewed as the mean is greater than median. We can verify skewness result in Table 3 i.e. 0.511 (a positive number). Our data is spread around the mean by 10409 points, based on standard deviation value.



Figure 8: Visualization of new tests per date.

These descriptive results as well as data scatter plot can be observed in Figure 9. This figure shows that the data is scattered evenly. It ranges from 0 to 25,000. The news and literature sources reflect that the tests conducted initially were less (0 to 200). Afterwards, it increased slowly. Most of the data lies between 20,000 to 25,000. Whereas, the value 41,666 is an outlier.



Figure 9: Stats of new tests.

Total tests

Total tests are mapped from zero to maximum of 292765 date wise, as seen in Figure 10. The feature is positively skewed as the mean is greater than median. We can verify skewness result in Table 3 i.e. 1.066 (a positive number). Our data is spread around the mean by 753281 points, based on standard deviation value.

These descriptive results as well as data scatter plot can be observed in Figure 11. This figure shows that the data scatter is even. It ranges from 0 to 2.3 Million. It can be inferred that total tests were increasing at a constant pace.

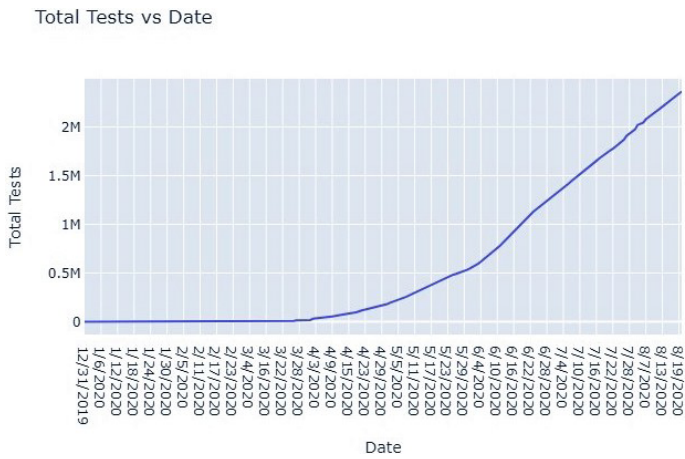


Figure 10: Visualization of total tests per date.

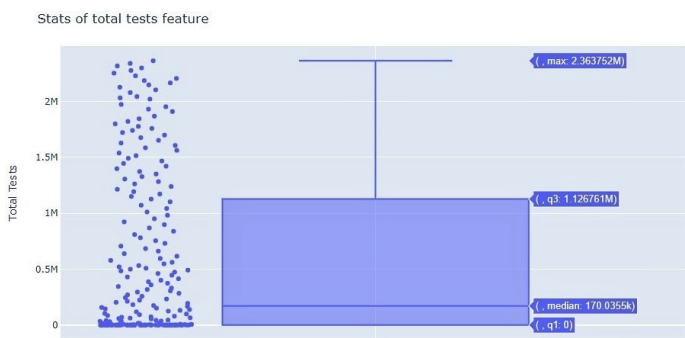


Figure 11: Stats of total tests.

Positive rate

Positive rate is mapped from zero to maximum of 9.75 date wise. By looking at the 25%, 50% and 75% values range, it can be said that there is an upper outlier in the data, as also seen in Figure 12. The feature is positively skewed as the mean is greater than median. We can verify skewness result in Table 3 i.e. 14.69 (a positive number). Our data is spread around the mean by 0.65 points, based on standard deviation value. These descriptive results as well as data scatter plot can be observed in Figure 13. The upper outlier was removed for further analysis.

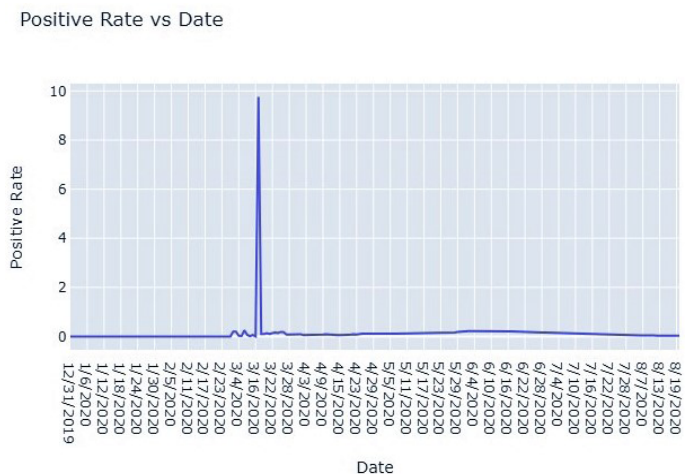


Figure 12: Visualization of positive rate per date.

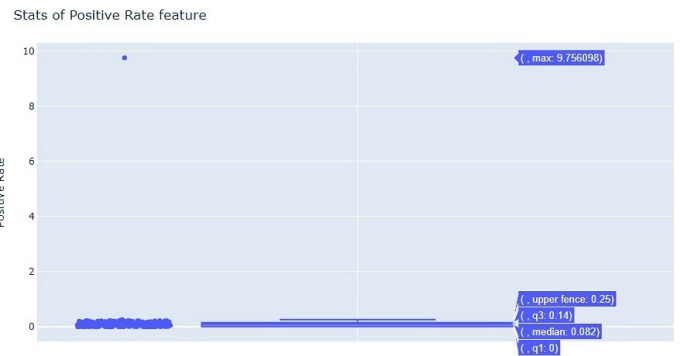


Figure 13: Stats of positive rate.

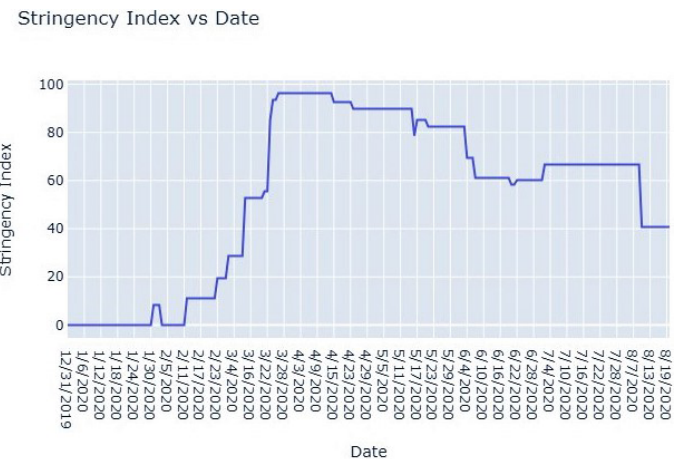


Figure 14: Visualization of stringency index per date.

Stringency index

It is mapped from zero to a maximum of 96.3, on orders of the government, as seen in Figure 14. This is a controllable feature. Figure 14 shows that the stringency index has constant straight lines throughout. The reason is that the government controls stringency. One stringency policy had to be implemented for a few days. After than a new stringency was decided. The feature is negatively skewed as the mean is lower than median. We can verify skewness result in Table 3 i.e. -0.43 (a negative number). Our data is spread around the mean by 33.71 points, based on the standard deviation value. These descriptive results as well as data scatter plot can be observed in Figure 15.

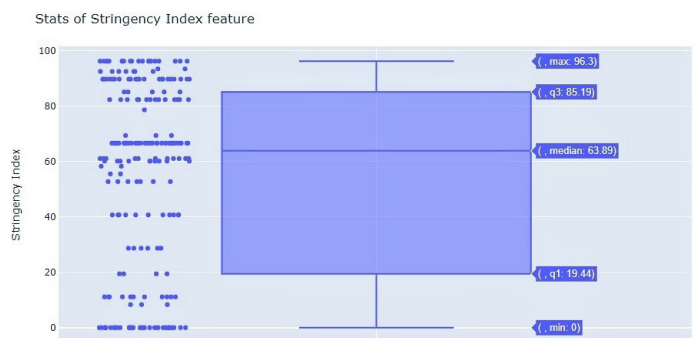


Figure 15: Stats of stringency index.

Key lessons learned

After analyzing the descriptive study of all these features together, it can be inferred that the number of new COVID cases are dependent on various factors. The first and obvious factor is the number of new tests. As the number of tests increased, new cases also increased. Likewise, when the number of new cases increased, so number of new deaths also increased proportionally. One important feature we can look at is that the values and statistics of stringency index feature. Once the response of the government towards the closures of public places and implementation of travel bans is strict (near to value 100), the number of new COVID cases are controlled. Initially, the rise of new cases was due to low value of stringency index. Looking at the cons of public meetings due to low stringency, the number of new cases added exponentially. After the stringency reaches above 90, in some course of time, new cases' value started declining.

Conclusion and Recommendations

Our study is strictly based on the numbers of the COVID data available at reputable resources like (Hannah, 2020; WHO, 2020). It is not undertaken by any clinical tests or resources. Hence, care must be taken to infer critical conclusions from the results. By looking at the outcome of descriptive analysis, it can be said that new COVID cases, new deaths and positive test rates are directly dependent on the number of new tests and more importantly, the stringency index. The implementation of right level of stringency (Govt. response using closures and bans on public proximity) can help control the number of new cases and hence number of deaths. Further analysis (predictive and prescriptive) are also being carried out, based on this dataset and study. To extend this research, we are using regression and deep learning algorithms to predict the number of new cases.

Novelty Statement

The techniques of descriptive analysis have been used to analyze the corona virus spread in Pakistan. For this purpose, peculiar features like stringency index have been utilized to achieve the goals.

Author's Contribution

Mian Ibad Ali Shah: Data fetching, processing and

analysis.

Yasir Saleem Afridi: Data analysis and paper writing.

Laiq Hasan: Project supervisor and did paper review.

Conflict of interest

The authors have declared no conflict of interest.

References

- Ahmad, K., Alam, F., Qadir, J., Qolomany, B., Khan, I., Khan, T., Suleman, M., Said, N., Hassan, S. Z. and Gul, A., 2021. Sentiment analysis of users' reviews on COVID-19 contact tracing apps with a benchmark dataset. ArXiv Preprint ArXiv: 2103.01196. <https://doi.org/10.2196/preprints.28371>
- Ahmed, S., 2020. Impact of COVID-19 on performance of Pakistan stock exchange. SSRN Electron. J., pp. 1–12. <https://doi.org/10.2139/ssrn.3643316>
- Chakraborty, T. and I. Ghosh. 2020. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis. Chaos, Solitons and Fractals, pp. 135. <https://doi.org/10.1016/j.chaos.2020.109850>
- Fisher, M.J. and A.P. Marshall. 2009. Understanding descriptive statistics. Aust. Crit. Care, 22(2): 93–97. <https://doi.org/10.1016/j.aucc.2008.11.003>
- Guan, W., Z. Ni, Y. Hu, W. Liang, C. Ou, J. He, L. Liu, H. Shan, C. Lei and D.S.C. Hui. 2020. Clinical characteristics of coronavirus disease 2019 in China. New Eng. J. Med., 382(18): 1708–1720. <https://doi.org/10.1056/NEJMoa2002032>
- Hamid, A., N. Shiekh, N. Said, K. Ahmad, A. Gul, L. Hassan and A. Al-Fuqaha. 2020. Fake news detection in social media using graph neural networks and NLP techniques: A COVID-19 use case. ArXiv Preprint ArXiv:2012.07517.
- Hannah, R., 2020. Coronavirus source data. <https://ourworldindata.org/coronavirus-source-data>
- Olapegba, P.O., S.K. Iorfa, S.O. Kolawole, R. Oguntayo, J.C. Gandi, I.F.A. Ottu and O. Ayandele. 2020. Survey data of COVID-19-related knowledge, risk perceptions and precautionary behavior among Nigerians. Data in Brief, 30: 0–5. <https://doi.org/10.1016/j.dib.2020.105685>
- Peixoto, P.S., D. Marcondes, C. Peixoto and S.M. Oliva. 2020. Modeling future spread of infections via mobile geolocation data and population dynamics. An application to COVID-19 in

- Brazil. PLoS One, 15(7 July), 1–23. <https://doi.org/10.1371/journal.pone.0235732>
- Qayyum, A., K. Ahmad, M.A. Ahsan, A. Al-Fuqaha and J. Qadir. 2021. Collaborative federated learning for healthcare: Multi-modal covid-19 diagnosis at the edge. ArXiv Preprint ArXiv:2101.07511.
- Saeed, U., K. Sherdil, U. Ashraf, I. Younas, H.J. Butt and S.R. Ahmad. 2021. Identification of potential lockdown areas during COVID-19 transmission in Punjab, Pakistan. Publ. Health, 190: 42–51. <https://doi.org/10.1016/j.puhe.2020.10.026>
- Saqlain, M., M.M. Munir, S.U. Rehman, A. Gulzar, S. Naz, Z. Ahmed, A.H. Tahir and M. Mashhood. 2020. Knowledge, attitude, practice and perceived barriers among healthcare workers regarding COVID-19: A cross-sectional survey from Pakistan. J. Hosp. Infect., 105(3): 419–423. <https://doi.org/10.1016/j.jhin.2020.05.007>
- Sarkodie, S.A. and P.A. Owusu. 2020. Investigating the cases of novel coronavirus disease (COVID-19) in China using dynamic statistical techniques. Heliyon, 6(4): e03747. <https://doi.org/10.1016/j.heliyon.2020.e03747>
- Shafi, M., J. Liu and W. Ren. 2020. Impact of COVID-19 pandemic on micro, small, and medium-sized Enterprises operating in Pakistan. Res. Glob., 2: 100018. <https://doi.org/10.1016/j.resglo.2020.100018>
- Spinelli, A. and G. Pellino. 2020. COVID 19 pandemic: Perspectives on an unfolding crisis. The British Journal of Surgery. <https://doi.org/10.1002/bjs.11627>
- WHO, 2020. Coronavirus disease (COVID-19) weekly epidemiological update and weekly operational update. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>
- Yousaf, M., S. Zahir, M. Riaz, S.M. Hussain and K. Shah. 2020. Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan. Chaos, Solitons Fractals, 138: 109926. <https://doi.org/10.1016/j.chaos.2020.109926>