

Interactive Data Driven Visualization for COVID-19 with Trends, Analytics and Forecasting

Frincy Clement, Asket Kaur, Maryam Sedghi
Department of Computing Science
University of Alberta
Edmonton, Canada
 {frincy, asket, msedghi}@ualberta.ca

Deepa Krishnaswamy, Kumaradevan Punithakumar
Department of Radiology & Diagnostic Imaging
University of Alberta
Edmonton, Canada
 {deepa, punithak}@ualberta.ca

Abstract—Interactive dashboards process and present raw data in the form of visuals, graphs, and text along with various options for user interactions. The dashboards allow for extracting valuable information and showcase the data in an intuitive and easy to understand manner. As the world is battling with the COVID-19 pandemic, we developed an interactive data-driven dashboard to not only view the current trends, but to also display important analytics and projections for the upcoming week. Built using python modules Dash and Plotly for visualization, the proposed dashboard utilizes the data analytic capabilities of the Pandas python library to structure and organize the raw data efficiently. Our dashboard is lightweight and designed for optimal performance. It can update values dynamically and be loaded onto any web server. Moreover, our proposed solution performed the best when compared to three other COVID-19 solutions, in terms of performance and speed, page size, and the number of HTTP requests.

Index Terms—Dashboard visualization, interactive visualization, visual data analytics, COVID-19, temporal data

I. INTRODUCTION

COVID-19 was initially observed in Wuhan, China in late December 2019, where it later spread globally, affecting around 7 million people worldwide and causing 300,000 reported deaths by May 2020. A large amount of data has been collected from patients regarding their symptoms and history to help track the spread of the disease [1]. Understanding the trends in the spread and identifying patterns, key events across the countries, would help correlate the measures taken by them and its corresponding results. To make well-informed decisions, the visualizations should communicate the global, country-wide and state/province specific trends. It can also delve into useful analytics that are not easily visible by only examining the trends. It could also provide some projections into the near future that would help the government officials plan their immediate actions. For this, a robust visualization dashboard with useful interactions to drill down to the specific details is very important.

Currently, there are many forms of visualization available for COVID-19 through various research centers, federal and provincial websites, and visualization challenges. Most of these visualizations focus on a part of information. For instance, Johns Hopkins University (JHU) [2] focuses on the world and US trends, while the dashboard created by the World Health Organization (WHO) [3] focuses on trends across the

world. There is a need for a visualization that can provide more information than the trends, *i.e.*, analytics and forecasting for the upcoming weeks.

As a solution, in this study, we have created an interactive data-driven dashboard, which is a one-stop place with trends, analytics and forecast embedded together. The proposed visualization provides valuable and intuitive visuals with easy and flexible interactions. Using just a click, hover or dropdown, users can understand values for global, country or states/provinces for a number of different information presented. We have derived many useful information from the input data, which is rarely found in other solutions and performed non-traditional analysis and forecasting to equip users with better understanding of what future beholds. Our solution not only focuses on the variety of information presented, but also improving the web application performance. The dashboard we have created has a more robust performance compared to three other compared solutions in terms of web application performance measured using Google's PageSpeed scores and Yahoo's YSlow scores, page sizes and number of HTTP requests.

II. RELATED WORK

The available solutions for COVID-19 visualizations can be categorized into two main areas: (i) Analysis and visualizations by research centers, governments, individual contributors worldwide; (ii) Analysis and visualizations by consulting and financial institutions, and economists.

The first category includes visualizations that are presented in an easily consumable format. They are simple and clear, hosted on public web servers, dynamically updated, and can be viewed by the public. Visualizations in this category provide general information about the cases worldwide, as well as at the country and state/province level. Most of these solutions utilize some form of visuals involving maps, line, or bar charts to visualize the trends and a table showcasing the values. They are developed and maintained by research centers, universities, health organizations, governments, and individual contributors. Well-known examples have been developed by the WHO [3], European Centre for Disease Prevention and Control [4], JHU resource center [2], and by companies such as Google [5]. The federal [6], [7] as well as state/provincial [8] governments also

provide visualizations with specific details about their region. All of these platforms provide insights into the trends, however do not include intelligent analytics, predictions for the future, and do not have flexibility in interaction.

The second category of solutions are in the form of reports generated by leading research firms, such as IBM [9], Accenture [10], and Deloitte [11], to help businesses and their leaders understand the impact of COVID-19 on the economy and businesses. Some have graphical visualizations, while some others provide only documentations about their analyses. Some are open to public with visualizations, whereas some are specifically made for businesses.

Our proposed solution falls in the first category, where we started by combining the different elements of trend visuals that are common across all the solutions we have come across. Along with the trends, we added one module each for analytics and forecasting. This would make our approach a one-stop solution for anyone who would like to know about general COVID-19 trends, as well as understand a little more in-depth, experiencing flexible user interactions.

III. PROPOSED SOLUTION

We developed an interactive dashboard, designed to include global trends as well as Canada-wide trends. It also includes analytics and forecasting of cases that have rarely been seen in the most common visualization solutions available for COVID-19. Thus, we propose a solution with variety of visualisations that is intuitive and interactive, which is up-to-date by dynamic loading of data. Our dashboard has four tabs: global trends, Canada trends, analytics and weekly forecasting.

IV. IMPLEMENTATION DETAILS

A. Details of Packages used

- 1) **Dash**: Dash [12] is an open source framework which helps to create interactive, responsive and dynamic websites by using the Python or R programming languages. We used Dash core components to create tools to add inputs, sliders, drop down, graph, and other components to the web application to allow users to interact with the data. Dash's Hypertext Markup Language (HTML) components allows for composing the application layout using Python structures, rather than writing in HTML or using an HTML templating engine.
- 2) **Plotly**: Plotly's [13] Python graphing library allows for creating interactive and dynamic graphs. Dash is used to build the interface which is populated by the graphs that are created from and using Plotly. We have used Plotly's graphobjects, figurefactory and express modules for our application.
- 3) **Pandas**: Pandas [14] is a data analytics library in Python and commonly used for processing tabular data. Here we used Pandas to read the main data file, pre-process the unclean data, extract the necessary information and plot it using Plotly on Dash.
- 4) **statsmodel**: statsmodel [15] is a python library with statistical tools and is used in our application for importing Seasonal Autoregressive Integrated Moving Average (SARIMAX) [16],

[17] with eXogenous regressors model for performing forecasting on the cumulative cases added to our application.

The processes involved in the implementation leading to the development of the proposed application – COVID-19 Dashboard is shown in Fig. 1. The following subsections will explain each part in detail.

B. Data set

The most up-to-date data about the COVID-19 cases were obtained from the JHU resource center [18]. The dataset contains information about confirmed cases, death and recovered cases, country/region-wise and province/state-wise, from January 22, 2020. The fields which are of utmost importance are the confirmed cases, deaths, and recovered cases. From these three fields, we have derived the following:

- Active cases
- Fatality rate
- Daily new confirmed cases
- Daily new deaths
- Daily new recovered cases
- Cases per 100,000 people for Canada
- Estimated cases for the upcoming week
- 7-day moving average of new cases

C. Data Processing

Data processing is performed dynamically when the website is loaded. The input datasets in the excel format are each read into the application as a pandas DataFrame [14] and merged together. After removing the unused fields, new fields are derived by performing mathematical operations on the DataFrame. For example, *New Cases* for each day are derived using the difference operator on the rows of cumulative cases ordered by date.

D. COVID-19 Dashboard Application structure

Our application created on Dash was designed to include as much user interactions as possible. We chose Dash and Plotly as our visualization platform mainly because of the tabular format of the data, that could be efficiently processed using pandas DataFrame and easily added to Dash using Plotly library. The dashboard was created to allow users to have the maximum interactions as possible. A list of customized user actions are: *Text Input, Dropdowns, Radio button, Play/Pause, Slider, Click Data* and *Hover Data*. For each of these interactions the responses are triggered for data, colorscales, axis scales, axis titles, plot titles, hierarchy of sunburst charts etc.

Our dashboard is divided into four tabs: global trends, Canada trends, analytics and forecast. We would now discuss each tab in detail.

1) *Global Trends*: This tab, as seen in Fig. 2, gives the overview of COVID-19 trends across the globe.

- Part 1: It features a heatmap of the world map (Choropleth) with countries highlighted in different color intensities from a color scale, based on their respective figures of confirmed/death/recovered cases. The type of cases to be

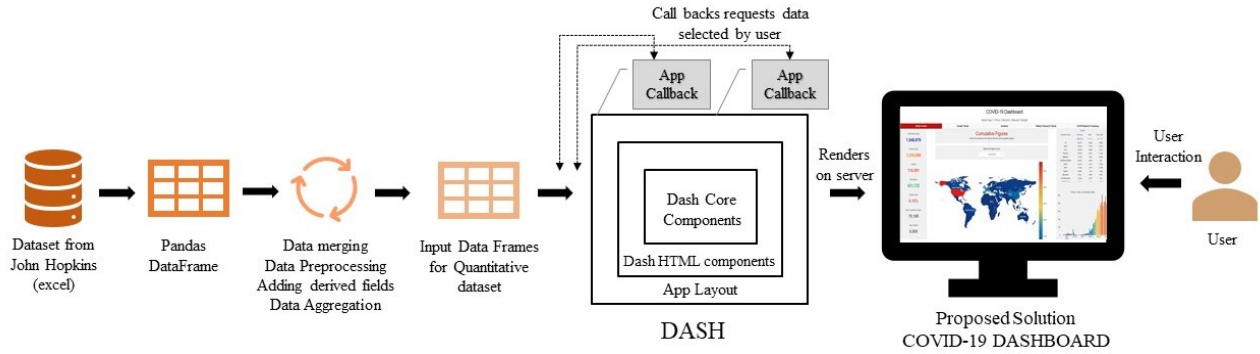


Fig. 1. System diagram showing different processes and workflow of the proposed COVID-19 dashboard implementation.

visualized can be selected from a drop down menu. This world heatmap is also click-activated; by clicking on a country, the country-specific values, state/province-wise (if captured in the dataset) and a bar chart showing the trend in the number of new cases is updated on the right panel.

- Part 2: The second half of the Global Trends tab presents a time-series animation of the COVID-19 pandemic spread using a Bubble map. The figures for the top 10 countries are also showcased along with a Doughnut chart and Line-plot to visualize their contribution and trend in the cumulative cases.

2) *Canada Trends*: This tab as shown in Fig. 3, gives information about the trends in Canada and its provinces. The topmost part of the tab has a highlights section, which presents the nation-wide trends in Canada.

This tab features a hover-activated Canada map with bubble plots, where the size indicates the cumulative confirmed and death cases. The line and bar plots change based on the hover data for respective provinces, which visualize the cumulative as well as new cases. The hover interaction on the bubble map changes the selected data, title and the graphs.

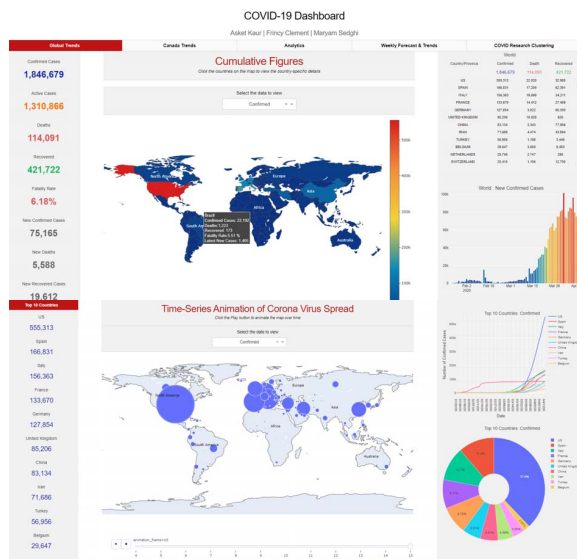


Fig. 2. Global Trends - First tab for our proposed COVID-19 Dashboard. Note: Both Choropleth and Bubble Map has custom hover over data

The user interaction in this tab includes click-data enabled on the Choropleth, hover information displayed on the Choropleth, play/pause button on the Bubble map and two drop downs to select the data to be viewed. The user has the option of changing the title, colorscale, selected data, and includes play/pause buttons for the animation and can update the graphs.

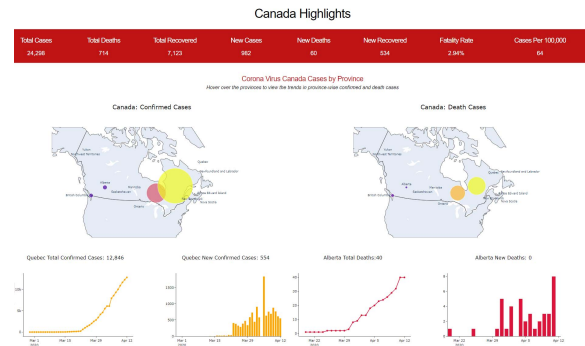


Fig. 3. Canada Trends - Second tab for the proposed COVID-19 Dashboard

3) *Analytics Tab*: This tab as shown in Fig. 4, provides information that is rarely seen in other solutions, as it is not readily available in the data set. It provides three types of information:

- Plot 1: The initial visual is a side-by-side comparison of the number of days since the first case was reported vs latest cumulative figures for top 20 or 30 countries. This analysis gives us an understanding as to how different countries have successfully fought back and evolved in the coronavirus spread. A horizontal bar chart and a line plot has been used for this visualization.
- Plot 2: Calendar heatmap of new cases (confirmed/death/recovered) for global or country-specific values is used to visualize the dates on which there has

been higher number of cases compared to others, as an intensity using different color scales.

- Plot 3: The hierarchical display graph sunburst chart shows the contribution of Country and provinces (if available) in the coronavirus spread.

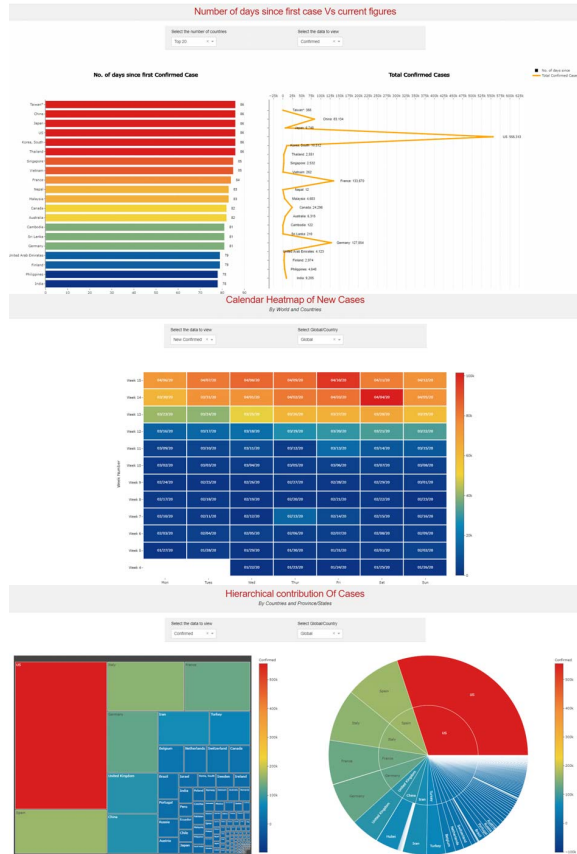


Fig. 4. Analytics Tab - Third tab for our proposed COVID-19 Dashboard

The user interactions available in this tab are mainly drop downs to select number of countries or country name and the data to be visualized. User interaction changes the axis title, annotations, color scales, data and graphs.

4) *Weekly Forecasting and Trends Tab*: The forecasting tab showcases two main visuals, one for weekly predictions and another for weekly moving average.

- Plot 1: Weekly forecast of cumulative cases for group of countries: Top 7, G7 and BRIC countries. This graph as shown in Fig. 5 provides information on the current figures and the predicted figures for next 7 days, which is modelled as a time series forecast using an auto regression algorithm provided by Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors (SARIMAX) model [15]–[17]. It takes into consideration of the trends and seasonal components present in the data to forecast the value for the future timeframe. The text annotation block displays the top three countries with highest forecasted values, based on the drop down for the selected countries.

- Plot 2: 7-day moving average of new cases vs cumulative cases for the top 9 countries give a different visualization compared to a time-series line or bar chart. On a linear scale, it shows the exponential curve of the COVID-19 spread. The same when visualized on a logarithmic scale results in a relatively linear plot at an approximate angle of 45 degrees, and the drop from that describes a more accurate descending curve. This could be defined as a recovery from the exponential growth of cases or that the curve is flattening.

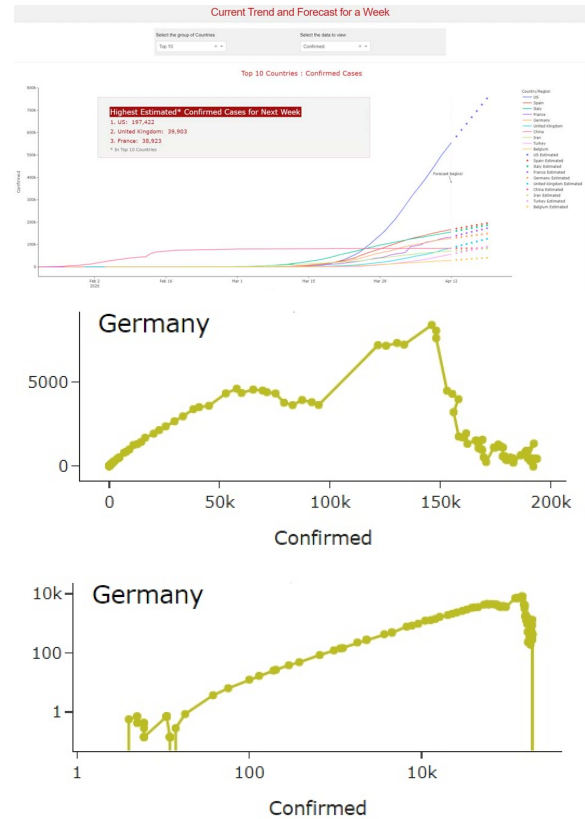


Fig. 5. Weekly Forecasting and Trends - Fourth tab for our proposed COVID-19 dashboard. Please note the difference in trend line, when the y-axis scale is changed from linear (second graph) to logarithmic (third graph) for the Weekly Moving average graphs

The user interaction in this tab includes drop down selection for countries, type of cases and changing between log and linear scale on y-axis. The changes are reflected on the y-axis scale, data to be selected for specific countries and cases, and the graph itself.

V. EVALUATION

A. Methodology

The proposed solution is evaluated by comparing against the most commonly used solutions available for COVID-19: Johns Hopkins World Map, World Health Organization (WHO) and the Canada Federal Government official dashboard. We present

results from two different tools GTmetrix¹ and Uptrends² from the plethora of different tools available, as it gives a comprehensive comparison on various metrics as listed below:

1) Website Performance Score: Two of the major metrics for website performance are Google's *PageSpeed*³ score and Yahoo's *YSlow*⁴ score. Each of the scores are calculated differently according to a set of rules by Google and Yahoo related to measuring website performance and speed. It mainly focus on the front-end performance which is based on load time of all of the resources including HTML, CSS, javascript and media (image, video etc.). A higher *PageSpeed* or *Yslow score* indicates that the web application is optimized for being rendered on that specific resolution it was tested on, in terms of a different set of criteria selected by Google and Yahoo.

2) Page Load Time (in secs) is the time taken to load the page and display its content in the browser window [19]. This metric keeps fluctuating with different tests and is not the best measure to compare the websites.

3) Page size (in MB) is the total size of the web pages being rendered in the browser window. Lower page size is considered better for an optimized web application.

4) Number of HTTP requests made during page load also helps to understand the performance of web application. Lower number of HTTP requests is usually considered to be better.

There is an option to choose the location of test for both tools. We have chosen Vancouver, Canada for GTmetrix and Los Angeles, USA for Uptrends. By default, the tests are run on a screen resolution of 1366×768, using Chrome browser for GTmetrix. For Uptrends, we have tested on HD+ 1600 × 900 Desktop screen and on Apple Iphone 8 Mobile screen using Chrome browser (version 83) at native network speed.

B. Results

The computational performance of the proposed dashboard is compared with three other publicly available visualization frameworks by JHU, WHO and the Government of Canada. The performance of web application was evaluated in terms of page loading scores by Google PageSpeed, and Yahoo YSlow as well as loading time in seconds, amount of web transfer in MB and number of HTTP requests. The evaluation results are reported in tables I using GTmetrix and in II using Uptrends. As per the results, our proposed dashboard performs much better than all other compared solutions for PageSpeed score and YSlow score. Our solution is lightweight and sends lesser number of HTTP requests for each page load. Thus, based on the evaluation, our proposed solution presents a variety of relevant visuals and analytics at a better performance than all compared solutions.

VI. CONCLUSION AND FUTURE WORK

In the time of data proliferation, interactive dashboards allows one to visualize information that is easier for the human

¹<https://gtmetrix.com>

²www.uptrends.com/tools/website-speed-test

³<https://developers.google.com/speed>

⁴<http://yslow.org>

brain to dissect and understand. We designed a comprehensive interactive data-driven visualization including trends, analytics and forecast as our contribution through this dashboard. We also added additional derived information along with the given information, which made some of the visualizations not just different, but more informative from the existing solutions.

Our solution has proved to be better than all of the other compared solutions, in terms of website performance score (PageSpeed and YSlow score), page size and number of HTTP requests. It presents more information in a lightweight framework, on a better performing platform, with less overhead.

In the immediate future, we would like to improve the aesthetics of the website and its rendering on mobile screens to make it more responsive. Even though our dashboard is on par with others with respect to page load, we plan to further optimize code for a faster page load. This would mean to reduce the dynamic calculations and predictions for each page load, which can be achieved by adding another worker to pre-calculate the required data for rendering the visuals at regular intervals. In this way, our main dashboard can run faster, without worrying about recalculating figures for each user interaction. We would also like to include additional visualizations like country-specific lock down information and its effect in the virus containment and more in-depth analytics involving characteristics of patients like age, gender and country-specific testing statistics.

VII. ACKNOWLEDGEMENT

We would like to thank Multimedia Research Centre at University of Alberta and Dr. Irene Cheng for the continued support and guidance. We would also like to thank all the researchers who have produced the current solutions, which were an inspiration for us while designing our dashboard.

VIII. RESOURCES

1. Source code: <http://tiny.cc/imfuqz>
2. Website hosted at : <https://covid19-ualberta.herokuapp.com/>

REFERENCES

- [1] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," *The Lancet infectious diseases*, vol. 20, no. 5, pp. 533–534, 2020.
- [2] "COVID-19 dashboard," Johns Hopkins University, <https://coronavirus.jhu.edu/map.html>, Accessed: 2020-04-18.
- [3] "COVID-19 dashboard," World Health Organization, <https://covid19.who.int>, Accessed: 2020-04-18.
- [4] "COVID-19 dashboard," European Centre for Disease Prevention and Control, www.ecdc.europa.eu/en/covid-19-pandemic, Accessed: 2020-04-18.
- [5] "COVID-19 situation dashboard," Google Data Studio, <https://datastudio.google.com>, Accessed: 2020-08-01.
- [6] "Coronavirus disease (COVID-19): Outbreak update," Public Health Agency of Canada, Government of Canada, <https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection.htm>, Accessed: 2020-08-01.
- [7] "Canada COVID-19 situational awareness dashboard," Public Health Agency of Canada, Statistics Canada and Natural Resources Canada, Government of Canada, <https://experience.arcgis.com/experience/2f1a13ca0b29422f9b34660f0b705043>, Accessed: 2020-04-18.
- [8] "COVID-19 Alberta statistics," Government of Alberta, <https://covid19stats.alberta.ca>, Accessed: 2020-04-18.

TABLE I
COMPARISON OF COVID-19 VISUALIZATIONS ON 1366×768 DESKTOP SCREEN USING GTMETRIX

	JHU (D)	WHO (D)	Canada Gov. (D)	Proposed (D)
PageSpeed Score (↑)	56%	82%	40%	88%
YSlow Score (↑)	62%	85%	97%	89%
Load Time in secs (↓)	28.5	13.6	24.1	19.7
Size in MB (↓)	6.2	4.75	7.93	1.28
HTTP Requests (↓)	348	55	619	39

TABLE II
COMPARISON OF COVID-19 VISUALIZATIONS ON HD+ 1600×900 DESKTOP AND APPLE IPHONE 8 MOBILE SCREEN USING UPTRENDS

	JHU (D)	JHU (M)	WHO (D)	WHO (M)	Canada Gov. (D)	Canada Gov. (M)	Proposed (D)	Proposed (M)
PageSpeed Score (↑)	25	16	7	5	19	17	61	40
Load Time in secs (↓)	3.6	4.2	7.4	5.8	9.4	9.3	4.5	2.9
Size in MB (↓)	27.4	34.1	17.9	17.7	21.2	22.6	4.5	4.5
HTTP Requests (↓)	259	164	65	33	522	515	21	23

Notes for both tables:

· JHU stands for Johns Hopkins University, WHO stands for World Health Organization

· (↑) means higher is better, (↓) means lower is better. (D) for results on desktop screen resolution and (M) for mobile screen resolution

- [9] “All COVID-19 analytics,” IBM Cognos Analytics, <https://bit.ly/3belySU>, Accessed: 2020-04-18.
- [10] “Coronavirus (COVID-19) business impact,” Accenture, www.accenture.com/ca-en/about/company/coronavirus-business-economic-impact, Accessed: 2020-04-16.
- [11] “COVID-19 insights collection by topic: Deloitte global,” Deloitte, www2.deloitte.com/global/en/pages/about-deloitte/articles/COVID-19-insights-collection-by-topic.html, Accessed: 2020-04-17.
- [12] A. Fuentes, *Hands-On Predictive Analytics with Python: Master the complete predictive analytics process, from problem definition to model deployment*. Packt Publishing Ltd, 2018.
- [13] C. Sievert, *Interactive Web-Based Data Visualization with R, plotly, and shiny*. CRC Press, 2020.
- [14] W. McKinney, *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. O’Reilly Media, Inc., 2012.
- [15] S. Seabold and J. Perktold, “Statsmodels: Econometric and statistical modeling with python,” in *Proceedings of the 9th Python in Science Conference*, vol. 57. Austin, TX, 2010, p. 61.
- [16] N. Elamin and M. Fukushima, “Modeling and forecasting hourly electricity demand by sarimax with interactions,” *Energy*, vol. 165, pp. 257–268, 2018.
- [17] A. Tarsitano and I. L. Amerise, “Short-term load forecasting using a two-stage sarimax model,” *Energy*, vol. 133, pp. 108–114, 2017.
- [18] “Novel coronavirus (COVID-19) cases data,” Johns Hopkins University, <https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases>, Accessed: 2020-04-18.
- [19] “What is page load time?” StackPath, <https://blog.stackpath.com/page-load-time>, Accessed: 2020-06-01.