The Impact of Covid-19 in Open-Source Software Development

Aysun Öğüt   aysuno@sabanciuniv.edu

Computer Science and Engineering - Faculty of Engineering and Natural Sciences, Junior

Bengisu Tepe   bengisutepe@sabanciuniv.edu

Computer Science and Engineering - Faculty of Engineering and Natural Sciences, Junior

Doruk Can Metiner   dorukmetiner@sabanciuniv.edu

Computer Science and Engineering - Faculty of Engineering and Natural Sciences, Junior

Emre Genç   gemre@sabanciuniv.edu

Industrial Engineering - Faculty of Engineering and Natural Sciences, Sophomore

Emirhan Böge   emirhanboge@sabanciuniv.edu

Computer Science and Engineering - Faculty of Engineering and Natural Sciences, Sophomore

Supervised by

Anıl Koyuncu

Computer Science and Engineering - Faculty of Engineering and Natural Sciences, Faculty Member

Abstract

Software developers have been affected by COVID-19 pandemic. Before the pandemic a lot of developers have been working physically however, due to the COVID-19, developers had to change their working conditions from physical to virtual. This rapid change may also affect the commit numbers of the developers if we compare before and after pandemic. In this paper we try to analyze commit numbers of the developers during COVID-19 pandemic. We use GitHub to clone a python project and after that analyze this project's commit history and visualize the data that we get. We also observe COVID-19 cases graphs and commit numbers of graphs to show correlation between these two. We found that there is no considerable change in commit numbers before and during pandemic. However, we believe that with more detailed observation and using more projects the result that we found may change. Our findings can be used to see what the correlation between COVID-19 and number is of commits in open-source projects.

Keywords: COVID-19, GitHub, Open-Source Development

1. Introduction

COVID-19 pandemic has global impact in different areas and one of the areas that may be affected by COVID-19 is open-source software. Open-Source Software is the process that provides public access to software projects and allows developers to contribute, improve and change it. During the pandemic software developers' attitude, productivity and working hours are also affected by COVID-19 and these changes may lead to increase or decrease in code efficiency.

GitHub is a platform that allows developers to collaborate and contribute to each other and it facilitates group working. Individual change revisions are named as commits and this helps developers to understand what other developers have done. Number of commits that are made by developers can be used as a parameter of how developers' contributions change during a pandemic. This commit history can be obtained through GitHub using 'git clone' and 'git log'. 'Git clone' is used to make a clone of a repository to a new directory and after that 'git log' is used to obtain commit history which also includes author name, commit date, and commit message.

Our research question is "Are there any generalizable characteristics for the working standards of open-source developers? If there is any, how is the impact of COVID-19 on these standards?" (**RQ**) and in this research, we collect a number of projects from GitHub and examine the commit history of these projects. We divide these commits into years and after that convert it to a graph which shows the number of commits according to years. In addition, we also analyze commit numbers according to months of the year to understand whether COVID-19 influences commit numbers or not. To get more detailed

information about COVID-19 effect, we also consider the commit number in the sense of days of the months and hours of the days. While we are interpreting, to make clear determination we put the graphs of the 2019 and 2020 one after another. We also examine COVID-19 cases and deaths graphs to show correlation between COVID-19 and developers' contribution rates by looking at their commit numbers.

When we compare pre-COVID and post-COVID situations, even though we found some differentiation in commit numbers in the year that COVID-19 broke out, we could not observe any remarkable changes between pre-COVID and post-COVID since there is other variables that may also affect these changes in projects, we cannot directly state that COVID-19 affects open software developers working standards.

2.   Analysis and Research

2.1.   Observation of Time-Dependent Changes

In order to conduct an analysis on the commit characteristics of the project, the time-dependency of the commits were examined. For this purpose, commit times are separated as years, months, days, and hours and each parameter is examined to see the changes in the number of commits. For the evaluation of hours, the local hour of the committed location is taken into consideration. This is done in order to better evaluate whether the users are choosing to commit in the mornings, around noon or at night. Figure 1, depicted above, visualizes the number of commits per year since the initiation of the Python project.
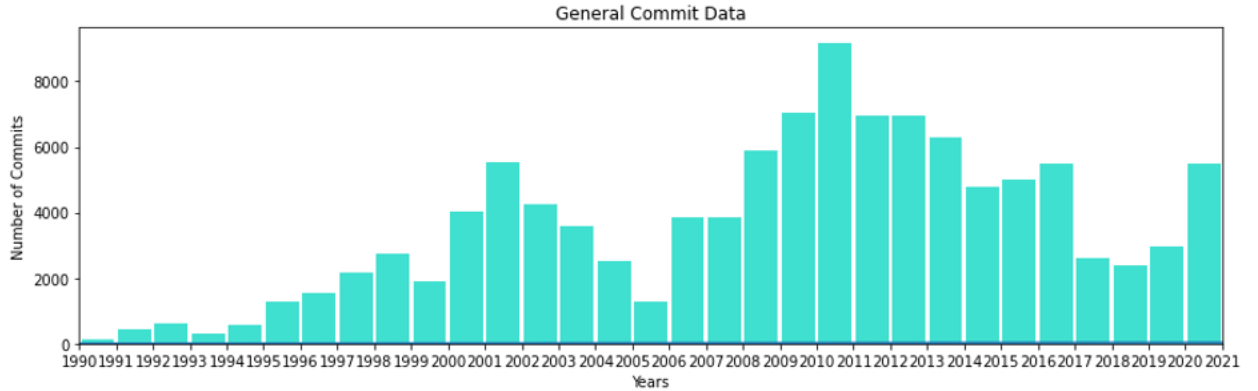
Figure 1: The yearly distribution graph of the commits of Python project, ranging from 1990 to 2021

According to Figure 1, the number of commits display an increasing trend starting from 1990 until 2001, followed by a sharp decrease until 2006 and another sharp increase starting after 2006. Although in a decreasing trend, the number of commits is still high when compared to before 2006 trends. It is seen that the trends in the number of commits throughout the years is highly correlated with the increasing popularity of Python itself.
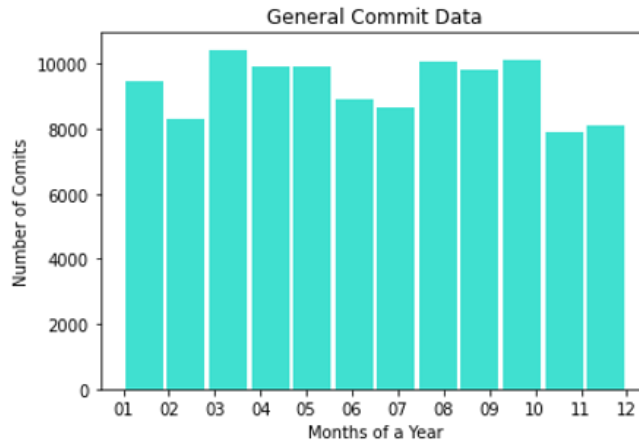
Figure 2: The monthly distribution graph of the commits of Python project

Continuing with the monthly analysis of the number of commits, it is seen that although generally following a similar characteristic within a year, a slight decrease in the

number of commits can be seen especially towards the end of a year and in the summer periods, in June, July, November and December to be specific. In order to better analyze the possible reasons behind these decreases, the major release dates of Python are illustrated in Figure 3.
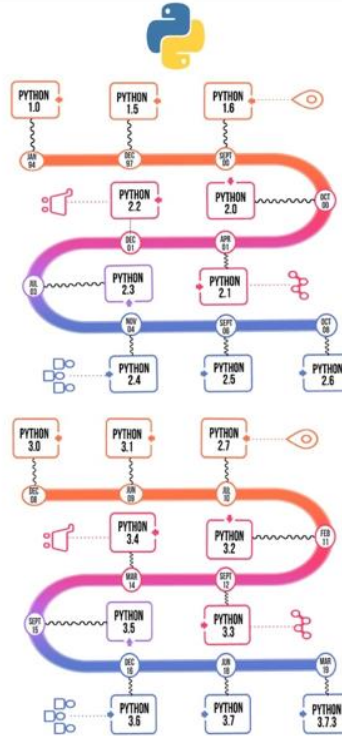

Figure 3: Illustration of Python version with corresponding release dates. (Pramanick, 2019)

As depicted in Figure 3, most major releases of Python are made towards the end of a year, around October, November, and December or towards the start of the summer period, around June and July. The slight decreases in the number of commits may be correlated with after-commit periods, where programmers are more flexible with the improvements to the Python language.
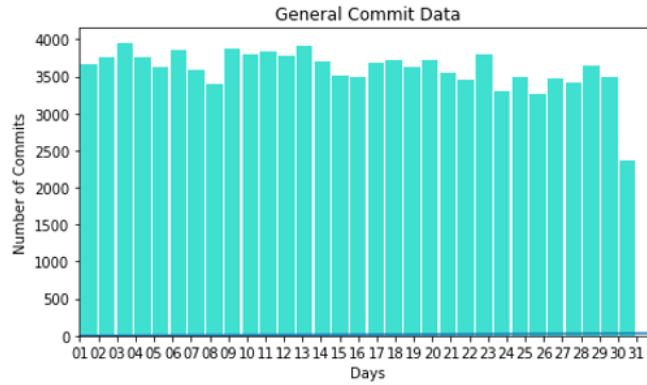
Figure 4: The daily distribution graph of the commits of Python project

Continuing with Figure 4, the daily distribution graphs are visualized. Although a decrease towards the end of a month was expected, it was not observed in the graph. The number of commits kept a similar trend over a month. This implies that developers are keeping up with the effort they provide to the project disregarding the month end comfort or Monday blues. Following the daily analysis on the Python project, hourly analysis was conducted.
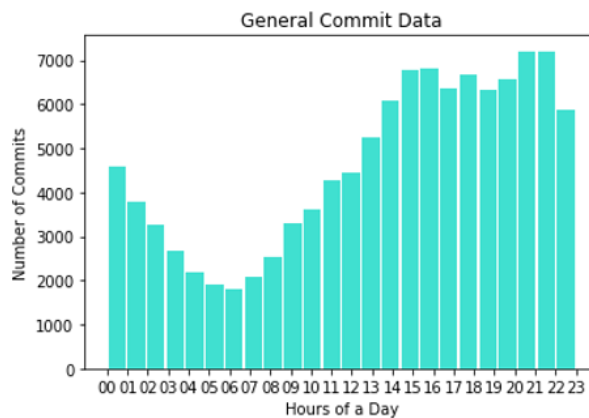

Figure 5: The hourly distribution graph of the commits of Python project

Following the daily analysis, hourly analysis on the number of commits were conducted. The results were revealing that the developers are choosing to work at the

second half of the day. Surprisingly, the most active time period developers committed were afternoons, evenings, and nights. The results suggested that the most active commit periods are 15.00-23.00 in a day for the developers. Finalizing the analysis of the time-dependent commit characteristic on the Python open-source coding language project, Figure 6 and Figure 7 depict the same characteristics explained above in a line graph for further and better evaluation.

Figure 6: The line graph of the commits per year

Figure 7: The line graph of the commits per month

2.2 Impact of COVID-19 on Software Projects

2.2.1   Data Analysis of COVID-19 Cases

For the purpose of better understanding on what scale COVID-19 affected the software development world, it is crucial to first analyze COVID-19 Cases since the beginning of the worldwide pandemic up until today. The graphs given below demonstrate certain statistics about the global pandemic such as the number of daily cases and daily deaths. The timeline of the data commences from January 22, 2020, and lasts until 21 December 2021.
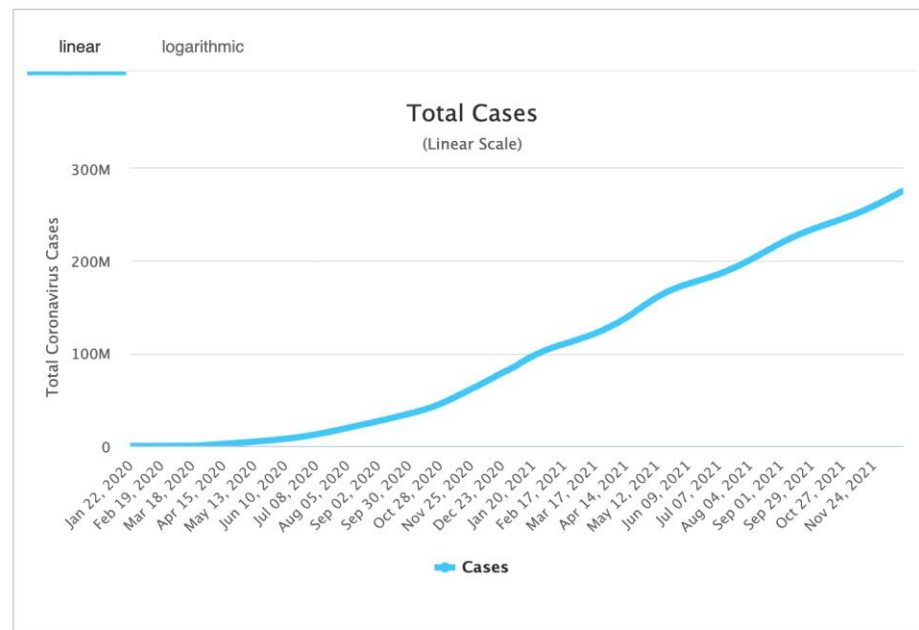
Figure 8: Total number of COVID-19 cases (Worldwide)

Daily Cases (worldwide)

On April 3, the French Government reported 17,827 additional cases and 532 additional deaths from nursing homes that had not been reported previously. On April 2, it had reported 884 additional deaths.

On February 12, China reported 51,152 additional new cases due to a change in how cases were diagnosed and reported. More details
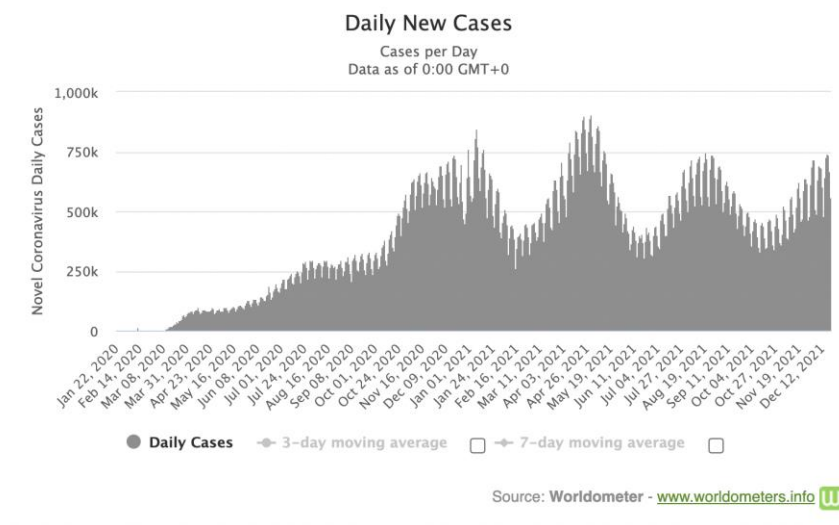
Daily New Cases
Cases per Day
Data as of 0:00 GMT+0

● **Daily Cases**   ◆ 3-day moving average   ☐   ◆ 7-day moving average   ☐

Source: **Worldometer** - www.worldometers.info

Figure 9: Daily COVID-19 cases (Worldwide)

Daily Deaths

Daily Deaths
Deaths per Day
Data as of 0:00 GMT+0

● **Daily Deaths**   ◆ 3-day moving average   ☐   ◆ 7-day moving average   ☐
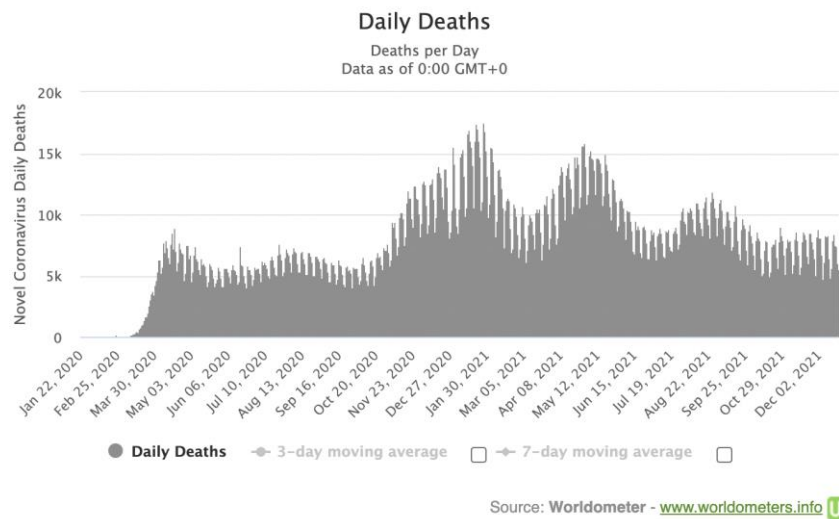
Source: **Worldometer** - www.worldometers.info

Figure 10: Daily COVID-19 cases (Worldwide)

By analyzing the graphs of daily cases and daily deaths, it can be concluded that there were instances of sudden spikes and declines during the aforementioned timeframe. In the case of the former, one can observe that starting from October 1, 2020, the number of daily cases started to increase in a rather short period of time, that is reaching approximately 750.000 daily cases (triple the number of cases at the start of October 2020) worldwide at the start of January 2021. Furthermore, it seems that the number of cases decreased dramatically, during the first two months of Spring 2021, then immediately skyrocketed to a record of approximately 850.000 COVID-19 cases daily at the beginning of May 2021. Moreover, it can be seen that there is a certain trend of sudden augmentations and declines after May 2021, that is the number of cases decrease drastically in the following two months and then skyrockets in the next two months. This trend continues until the end of our time frame, which is Dec 12, 2021.

As for the latter being the number of daily deaths worldwide, it differentiates itself from the former in terms of two certain aspects. Firstly, one can see that at the commencement of the global pandemic, since Covid-19 was not yet exhaustively understood by scientific authorities, the number of deaths was rather high from the very beginning. However, starting from May 12, 2021, it can be observed that this number has been decreasing steadily up until the start of December 2021, which is probably due to the fact advancements in vaccination and medicine were proven to be effective in terms of lowering the rate of mortality.

2.2.2 Pre-COVID-19 and Post-COVID-19

In order to examine the effect of the new regulations brought by the COVID-19 pandemic on software development, data exploration was carried out according to the years of the selected project. The years 2019 and 2020 were considered to analyze this project. In order to understand the changes in software development just before and during the pandemic, and whether COVID-19 has an effect on these changes, the analyzes of these two years were based on different metrics such as commit numbers in different time periods. In this subsection, we will visualize the data we have taken and talk about where the changes are.

First of all, the commits of the project cloned via GitHub were turned into a JSON file and the data in this JSON file was turned into a data frame with python. Then, some graphs were drawn to understand the changes of the project with this data obtained.
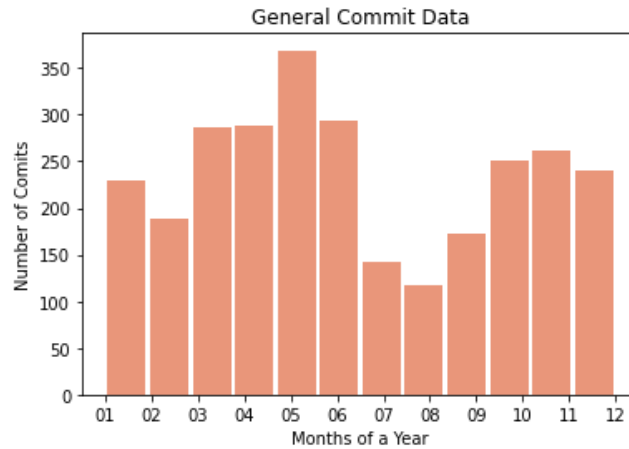

Figure 11: Total number of commits in 2020 by months

Figure 11 shows that the distribution according to the months of the year 2020 is drawn. Looking at this graph, which is the year the COVID-19 pandemic broke out, it can

be observed that there is a decrease in the number of commits in the summer months. Although it is difficult and not correct to make an inference just by looking at the graph directly, this is the value we could read from this graph. It can be thought that there are different parameters affecting this change in the number of commits.
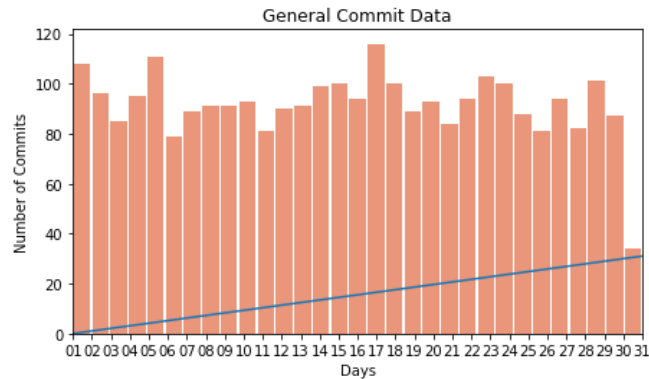

Figure 12: Total number of commits in 2020 by days

Figure 12 shows how many commits were made on which days of the month during 2020. A decrease is observed at the end of the month.
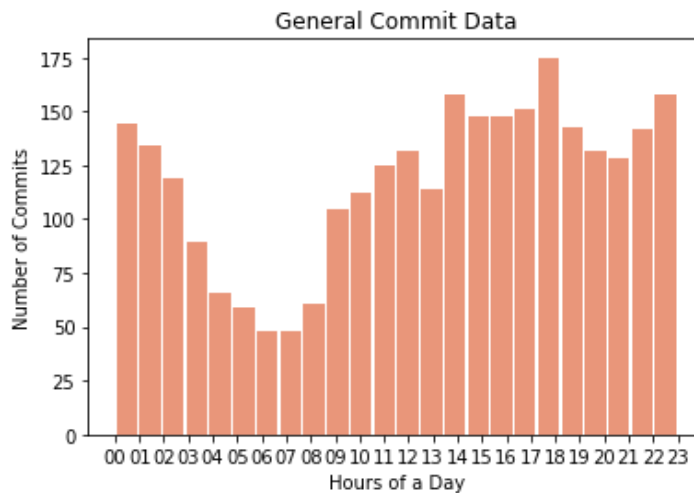

Figure 13: Total number of commits in 2020 by hours of a day

The chart in Figure 13 shows how many commits were made during the year 2020, by dividing the hours of the day. Here, a decrease is observed in the morning hours. In addition, we thought that there might be a decrease in the end of the official working hours but as can be seen from the table, there is no decrease compared to other hours. Our assumption here is that with the work from home model that comes with COVID-19, there is no such decrease in the evening time, as there are times when people do not waste time on the road and work hours are stretched.
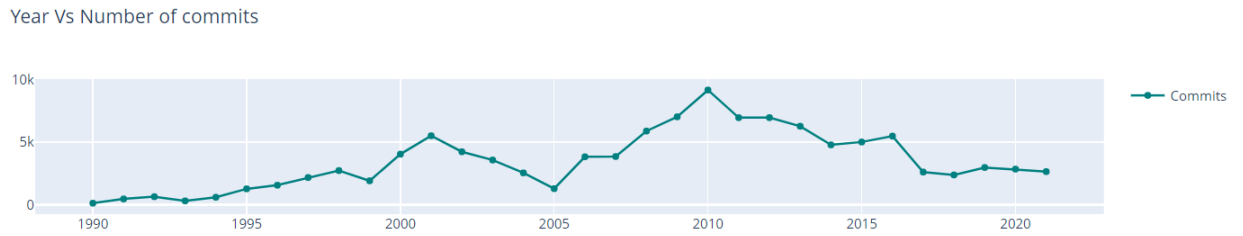
Year Vs Number of commits

Figure 14: Total number of commits by year

Figure 14 shows how many commits have been made over the years for this project that we have examined. Since Python is an increasingly popular programming language, it is normal to expect a graph with a constantly increasing number of commits. Here, we see a decline in recent years, but we expect an increase. It is not very reliable to say that this decrease is directly due to COVID-19, because as can be seen, the decrease started before 2020.

The graphic that is most relevant to the subject that forms the basis of this project is to see the graph of the commit change before and during COVID-19 (Figure 15). In this chart, the commit numbers in the pre-COVID-19 (2019) and COVID-19 (2020) period are

examined. In this research, it is aimed to observe whether COVID-19 has an effect on the work of software developers and open-source software development.
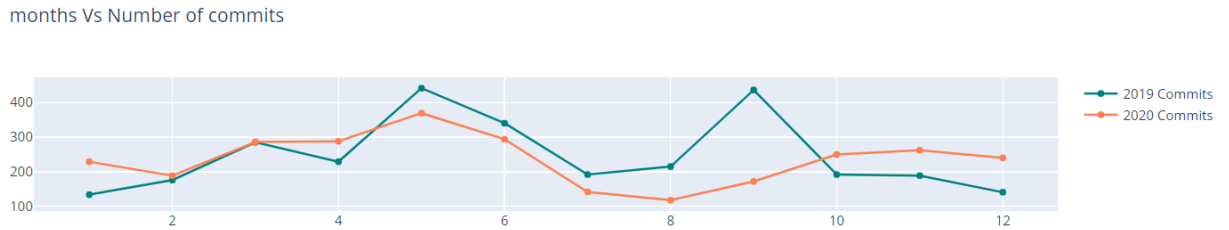
months Vs Number of commits

Figure 15: Comparison between the number of commits in 2019 and 2020

With the values we can read from this table, although there was a decrease in the summer of the year when COVID-19 broke out, it can be seen that even more commits were made towards the end of the year than in 2019. However, as we mentioned in the previous parts of our article, we do not ignore the possibility of other parameters and directly linking this to COVID-19 may not be correct. Of course, this is a situation that can change from project to project. In our case, we could not observe any remarkable changes between pre-COVID and post-COVID.

3   Discussion and Conclusion

Being one of the major crises of the 21st century, Covid-19 had its impact in almost all aspects of our lives. Unpredictable future, the number of covid cases and deaths which have been announced every day and a wide range of news deeply affected humanity. In this research, the effect of the pandemic on open-source development and thus software developers were questioned, and answers are tried to be found. For this purpose, general time-dependent commit characteristics were analyzed as a beginning step. Graphs are

created based on yearly, monthly, daily, and hourly commit numbers on a well-known open-source project, Python Programming Language project. Yearly commit numbers are found to be correlated with the popularity of the language while monthly commit numbers are found to be dependent on major release dates. Daily and hourly commits are kept as a control group data to be compared with after covid data.

By comparing the commit numbers just before COVID-19 and during COVID-19, it was examined whether the COVID-19 pandemic, the work-from-home model, would contribute to the development of the project. In the selected project, metrics such as commit numbers in 2019 and 2020 were compared and it was observed whether there was a difference in the year of COVID. As a result, we concluded that there is no remarkable change, and that the changes cannot be directly linked to COVID-19. There may be different parameters affecting the development of the project. Therefore, for further research, studies such as examining different metrics, surveying the developers doing the development and finding the difference between their own ideas and what they do can be done.

References

Coronavirus worldwide graphs. Worldometer. (2021, December 12). Retrieved December 12, 2021, from https://www.worldometers.info/coronavirus/worldwide-graphs/

Ford, D., Storey, M.-A., Zimmermann, T., Bird, C., Jaffe, S., Maddila, C., Butler, J. L., Houck, B., & Nagappan, N. (2022). A tale of two cities: Software developers working from home during the COVID-19 pandemic. *ACM Transactions on Software Engineering and Methodology*, *31*(2), 1–37. https://doi.org/10.1145/3487567

Pramanick, S. (2019). History of Python. photograph, GeeksforGeeks. Retrieved December 22, 2021, from https://media.geeksforgeeks.org/wp-content/uploads/20190502023317/TIMELINE.jpg.

Rodeghero, P., Zimmermann, T., Houck, B., & Ford, D. (2021). Please turn your cameras on: Remote onboarding of software developers during a pandemic. *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. https://doi.org/10.1109/icse-seip52600.2021.00013