

1 **Real time scalable data acquisition of COVID-19 in six continents through**
2 **PySpark - a big data tool**

3 **Tanvi S. Patel¹, Daxesh P. Patel² and *Chirag N. Patel³**

4 ¹Computer Science & Engineering, Gujarat Technological University, Ahmedabad, Gujarat
5 382424, India.

6 ²Department of Chemistry, School of Sciences, Gujarat University, Ahmedabad, 380009,
7 India.

8 ³Department of Botany, Bioinformatics, and Climate Change Impacts Management, School
9 of Sciences, Gujarat University, Ahmedabad, 380009, India.

10

11 **ORCID profiles**

12 Tanvi S. Patel (<https://orcid.org/0000-0002-9630-3123>)

13 Daxesh P. Patel (<https://orcid.org/0000-0002-9620-8660>)

14 Chirag N. Patel (<http://orcid.org/0000-0003-0777-7720>)

15

16 **Details of correspondence**

17 Dr. Chirag N. Patel

18 Department of Botany, Bioinformatics, and Climate Change Impacts Management, School of
19 Sciences, Gujarat University, Ahmedabad 380009, Gujarat, India.

20 Email: chiragpatel269@gmail.com

21

22 **Author Approvals**

23 All authors have seen and approved the manuscript, and that it hasn't been accepted or
24 published elsewhere

25

26 **Authors have no competing interests to declare**

27

28 **Real time scalable data acquisition of COVID-19 in six demographical continents**
29 **through PySpark - a big data tool**

30 **Abstract**

31 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was declared as a global
32 emergency in January 2020 due to its pandemic outbreak. To examine this Coronavirus
33 disease 2019 (COVID-19) effects various data are being generated through different
34 platforms. This study was focused on the clinical data of COVID-19 which relied on python
35 programming. Here, we proposed a machine learning approach to provide a insights into the
36 COVID-19 information. PySpark is a machine learning approach which also known as
37 Apache spark an accurate tool for the searching of results with minimum time intervals as
38 compare to Hadoop and other tools. World Health Organization (WHO) started gathering
39 corona patients' data from last week of the February 2020. On March 11, 2020, the WHO
40 declared COVID-19 a global pandemic. The cases became more evident and common after
41 mid-March. This paper used the live owid (our world in data) dataset and will analyse and
42 find out the following details on the live COVID-19 dataset. (1) The daily Corona virus
43 scenario on various continents using PySpark in microseconds of Processor time. (2) After
44 the various antibodies have been implemented, how they impact new cases on a regular basis
45 utilizing various graphs. (3) Tabular representation of COVID-19 new cases in all the
46 continents.

47

48 **Key words:** COVID-19; apache spark; PySpark; live data analysis; big data analysis;
49 continents.

50

51 **Real time scalable data acquisition of COVID-19 in six demographical continents**
52 **through PySpark - a big data tool**

53 **Introduction**

54 In early December 2019, an unidentified virus known as coronavirus was discovered in
55 Wuhan, China, among a group of patients suffering from severe tuberculosis, and it quickly
56 spread to other cities. The modern infection severe intense respiratory syndrome coronavirus
57 2 (SARS-CoV-2) was declared by the World Health Organization, and the illness caused by
58 this unused infection was called coronavirus disease 2019 (COVID-19)(You et al.
59 2020)(Patel et al. 2020).

60 In January 2020, the World Health Organization (WHO) declared a global emergency, and
61 the situation quickly escalated, prompting WHO to declare COVID-19 a global pandemic on
62 March 11, 2020 (Cucinotta and Vanelli 2020; Patel et al. 2021a). COVID-19 is spread
63 predominantly from person to person, most commonly through respiratory droplets from
64 coughing, sneezing, or talking. All the researchers, biologist, Pharmaceuticals and doctors
65 were trying to find the medicine, vaccines and treatment to control COVID-19 (Patel et al.
66 2021b). In December 2020, The U.S. Food and Drug Administration (USFDA) granted an
67 Emergency Use Authorization (EUA) for the Pfizer-BioNTech COVID-19 Vaccine. The
68 FDA granted an EUA for the use of the Moderna COVID-19 Vaccine on December 18, 2020.
69 The present study provides an insights of the COVID-19 new cases curve after antibody
70 vaccines were introduced using machine learning approach (Ong et al. 2020).

71 Since MS Excel has a 255-row limit for processing and visualizing data, we can use tableau,
72 Hadoop MapReduce, or Apache Spark for big data analysis (Gupta et al. 2017). However,
73 when compared to other machine learning tools, Apache Spark provides us with quick,
74 reliable, and accurate results. Apache spark can also handle Parquet, ORC, JSON, CSV, and
75 text format. There are many ways of data visualization and analysis, such as theoretical,
76 graphical and tabular forms. Graphs are striking psychological resources because they rapidly
77 and efficiently display information. Often, when provided by a graph, data can be better
78 interpreted than by a hypothesis since a pattern or comparison can be unveiled by the graph
79 and chart. The Python scripting language offers numerous libraries, such as pandas, to
80 construct very powerful graphs and charts (Gupta et al. 2017; Stancin and Jovic 2019).
81 PySpark includes comprehensive and quality ways to run zeta bytes and peta bytes of data on
82 parallel clusters a lot faster than conventional python applications for machine learning

83 applications. PySpark offers methods for statistical analysis and a ML (Machine Learning)
84 pipeline. In specific, Apache spark is emerged as popular open framework that support the
85 various data visualization techniques like different charts (Bar, line, pie, etc). Apache Spark
86 provides JAVA, Scala, R and python API's but among all of these, python is simple scripting
87 language and good for machine learning due to its versatile existence. Python's popularity in
88 the field of data science has reached new heights, especially in terms of freely accessible
89 frameworks and libraries. Among the many free libraries available, pandas, NumPy, and
90 Matplotlib are well-known and widely used in the field of data analysis (Ghaffar et al. 2015;
91 Assefi et al. 2017; Gupta et al. 2017; Magar et al. 2021).

92 In this COVID-19 challenge, everyday millions of data is being generated and these data are
93 continuously employed to established different preventions and control steps. COVID-19
94 datasets are increasing in size day by day as well as it is becoming sophisticated too. It is
95 necessary to build solutions using statistical methods to leverage this wealth of data. Due to
96 the rise in data and new data on a regular basis, new scripting codes every day need to be
97 analysed to create new visualizations of the data as well as new graphs. In order to solve this
98 issue, this paper incorporates live data analysis so that visualization is often updated every
99 day as data changes and there is no need for new scripts every day. This study focuses on the
100 daily cases and daily deaths in various continents due to COVID-19 and makes statistical
101 graphs for all continents. This work also highlights confirmed cases and death ratios in a
102 tabular format, which allows for easy comparison of the entire world's continents in a single
103 table. Using PySpark, compare all the Earth's continents in a graph and table.

104 **Methodology**

105 *Data curation*

106 In order to perform the statistical analysis, the data of COVID-19 collected from all over in
107 world. It includes confirmed cases, new cases, deaths, hospitalizations, testing, and
108 vaccinations, throughout the duration of the COVID-19 pandemic. COVID-19 cases and
109 death were collected from COVID-19 Data Repository by the Center for Systems Science
110 and Engineering (CSSE) at Johns Hopkins University (Badr et al. 2020; Magar et al. 2021).
111 European countries vaccination against COVID-19 details were obtained from the European
112 Surveillance System (TESSy) webpage (Ammon and Faensen 2009). Vaccinations against
113 COVID-19 data was retrieved by the Data team from official reports of all across the globe.
114 Other data was selected from a variety of sources like United Nations, World Bank, Global

115 Burden of Disease, Blavatnik School of Government and more. Web-based application was
116 used for the utilization of the COVID-19 data with live monitoring (Ghaffar et al. 2015).
117 Figure-1 depicts the workflow of proposed study which comprises of various data collection
118 and administered. The raw COVID-19 data was customized using text editor with
119 normalization with the filtered by continent accountability.

120 ***Machine learning Approach: Apache Spark with API python***

121 Apache Spark is a fast and trustable general-purpose cluster computing framework. This
122 framework offers API (application programming interfaces) in Java, Python, Scala and R
123 script. Spark is an Apache incubator cluster computing system that is focused in making data
124 processing easier with top agile application in coding as well as running. It is very fast at both
125 running and coding (Salloum et al. 2016; Assefi et al. 2017). This applicability was applied
126 for the current study which proposed novel data structure of COVID-19 using Spark core and
127 upper-level libraries such as Spark's MLlib for machine learning, GraphX for graph analysis,
128 Spark Streaming for stream processing, and Spark SQL for structure data processing make up
129 Apache Spark (Assefi et al. 2017; Gupta et al. 2017). Real-time data, such as streaming,
130 video, and social media data, was also analysed by Apache Spark. The promising and
131 accurate data processing was performed through Apache Spark which includes the 7500 rows
132 at a single instance or operation with meticulously. Figure-2 shows the Apache spark
133 architecture incorporates four APIs with correlation to COVID-19 dataset to analyse it on
134 case by case basis with numbers of deaths and vaccines are being employed till now. The
135 average speed of the Apache Spark is 75 thousand rows and 59 columns in 4.46 μ s of total
136 time and 589 μ s of wall time (Binti Hamzah et al. 2020).

137 Python's prominence is increasing, widely in the all areas like Bio informatics and data
138 analysis as well as in programming too (Latif et al. 2020). As a result, the number of free
139 frameworks and libraries for use is growing. Various Python big data research libraries, such
140 as pandas, matplotlib, NumPy, and seaborn, are also used in this analysis (Figure-3). Python
141 libraries are commonly used to visualize data, and visualization tool is becoming simpler with
142 the use of these free libraries.

143 ***Formulas***

$$\% \text{ of positive cases} = \left(\sum \text{daily confirm cases} / \sum \text{daily tests} \right) * 100$$

$$\% \text{ of death} = \left(\text{total deaths} / \text{total confirm cases} \right) * 100$$

144

145 *Script used to find charts*

146 The given script shows how a small amount of code can create a graph with real time data
147 rows of data in a short amount of time. Real time data set was obtained from
148 <https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/owid-covid-data>
149 .csv. PySpark Scripting code was uploaded on [https://github.com/pateltanvi2992/Data-](https://github.com/pateltanvi2992/Data-Analysis-of-Covid-19-using-Spark-)
150 [Analysis-of-Covid-19-using-Spark-](https://github.com/pateltanvi2992/Data-Analysis-of-Covid-19-using-Spark-) (Patel 2021).

151 **Results and Discussion**

152 The vast volume of daily produced data has pushed us beyond the limits of conventional
153 processing capacities, necessitating the adoption of advanced computing infrastructures that
154 can process data in parallel and distributed ways. The generated data consists of huge in size
155 and requires innumerable memory space, so Apache Spark was adopted to rectify this
156 problem. The Apache spark platform and its library provides the exploration of qualitative
157 and quantitative attributes collection. With the help of Python large-scale real-world data
158 experiments of COVID-19 were applied for this study. These data were represented in six
159 different continents.

160 Given demographic table-1 is describe the whole scenario of the COVID-19 in six continents.
161 The limitation of this study in our analysis we did not consider data from the seventh
162 continent Antarctica as there is not any permeant resident staying although there is at least 36
163 people confirmed with COVID infection (<https://www.barrons.com/>;
164 <https://www.coolantarctica.com/>).

165 **Asia**

166 Asia has a total of 23225625.0 COVID-19 positive cases and 22836076.0 recovered cases as
167 of May 15th, 2021. According to this, the mortality rate will be 1.6359 percent until May
168 15th, 2021. From February 24, 2020 to May 15th, 2021, the daily recover cases and daily
169 confirm cases are depicted in Figure-4 (B) (scatter graph). On September 16, 2020, the
170 highest number of new cases were registered in India, with 97894.0 (new cases) out of
171 1136613.0 (new tests). The COVID-19 cases peaked in September 2020, according to Figure-
172 5 (B) (histogram), and then started to decline. Total tests performed in Asia were
173 373625749.0, with 23225625.0 confirmed cases. As a result of this, 6.2163 percent of the
174 cases tested positive.

175

176 **Africa**

177 Africa has total COVID-19 positive cases till May 15th 2021 are 4027046.0 and recover are
178 3912657.0. On basis of that the death ratio is 2.6701 %. Figure-4 (E) is showing the daily
179 recover cases and positive cases. Figure-5 (E) is showing the positive cases going from 25th
180 February 2020 to May 15th 2021. Although Africa has 17.2% of the world's population, it
181 only has about 5% of COVID-19 cases diagnosed and 3% of COVID-19-related mortality
182 (Bamgboye et al. 2020).

183 **Europe**

184 Europe has total COVID-19 positive cases till May 15th 2021 are 36470789.0 and recover
185 cases are 35620243.0. On basis of that the death ration in Europe was 2.3321%. The
186 maximum cases reported on 2nd November 2020 in France which was 106091.0. Figure-4
187 (C) shows the recover cases and positive cases through the scatter graph. The total test occur
188 in Europe was 448349376.0. Figure-5 (C) is showing the graph of positive cases from 11th
189 March 2020 to 13th March 2021.

190 **North America**

191 North America has total COVID-19 positive cases till May 15th, 2021 are 33850076.0 and
192 recover cases are 33070581.0. The percentage of the death ratio in North America was
193 2.29889 %. Figure- 4 (A) (scattered graph) is showing the daily new cases and daily recover
194 cases from 24th February 2020 to May 15th, 2021. The maximum new confirm cases reported
195 in North America on 2nd January 2021 in United states was 300416.0 among 108349.0 new
196 tests. The total test occurred in North America till May 15th, 2021 was 363400300.0. So,
197 9.3148% positive cases among all tests. Figure-5 (A) (histogram) is showing the cases going
198 from February 24th, 2020 to May 15th, 2021 (Control 2021).

199 **South America**

200 South America has total COVID-19 positive cases till May 15th, 2021 are 19191051.0 and
201 recover cases are 18693169.0. Based on this the death ratio of the South America was
202 2.5853%. Figure- 4 (D) (scatter graph) is showing the daily new cases and daily recover
203 cases. The maximum new cases reported in South America 87843.0 in Brazil on 7th, January
204 2021. The total test occur in South America was 4245255498.0. Figure-5 (D) is showing the

205 graph going for daily positive cases in South America from 1st January 2020 to May 15th,
206 2021 (Control 2021).

207 **Oceania**

208 Oceania has total COVID-19 positive cases till May 15th, 2021 are 33454.0 and recover cases
209 are 31901.0. The death ratio of the COVID-19 in Oceania was 2.8636%. Figure-4 (F) shows
210 the recover cases and positive cases. The total test occur was 11088450.0 so the ration of
211 positive cases among tests are 0.3017%. Figure-5 (F) is showing the graph cases occur from
212 26th January 2020 to May 15th, 2021.

213 **Effect of COVID-19 Vaccine and out of vaccination campaigns**

214 The COVID-19 Vaccine Program in the United States was started on December 14, 2020.
215 About 87.49 million individuals, or 26.4% of the population of the United States, have
216 received at least one dose of vaccine as of March 26, 2021. Around 39.9 million people, or
217 12.07% of the US population, have received fully vaccination. Figure-6 (A) is showing that
218 still 73.6 % US population requires COVID-19 vaccination. Figure-6 (B) and (C) (line chart)
219 reveals that after Jan-2021 the chart of new cases is curve and new cases are decreasing day
220 by day. Mostly there are three kind of vaccines available in the USA namely Pfizer-
221 BioNTech, Moderna And J&J (Control 2021).

222 **Conclusion**

223 The current work represents a novel code - scripting big data analysis of COVID-19 using
224 Apache Spark and Python, as well as a machine learning approach that any researcher can
225 use.

226 The advantage of this study is real time data set rows were examined and endeavoured at the
227 speed of microseconds of time. These data comprised of six distinct continents of the world.
228 Based on this data we concluded the death ratio, positive COVID cases rate changes were
229 observed very thoroughly, Although Africa has the lowest death rate, it is not statistically
230 significantly lower than other continents. Interestingly, we observed that the Oceanian
231 continent has a higher death rate than the other continents, even though the total number of
232 COVID testing is very low. In Asia, 6% of the total COVID testing population tested positive
233 for COVID, whereas 9% of the total COVID testing population tested positive for COVID in
234 North America. Even though the total number of COVID testing population in North as well
235 South America is greater than either Asia or other continents. We also observed that after

236 vaccination, positive COVID cases begin to decrease dramatically, which is a result of the
237 ongoing successful vaccination campaign program in the United States and the North
238 American continent (Figure-6). In Conclusion, this Apache Spark and Python based a
239 machine learning approach implements the time dependent data availability with live
240 coverage of updates with personal hygiene and precautions towards COVID-19. Represented
241 machine learning-based PySpark analysis is also useful for healthcare research institutes in
242 clinical and biostatistics rapid analysis, allowing them to conduct a quick, efficient,
243 visualizing, and agile study on the COVID-19 outbreak.

244 **Compliance with Ethical Standards**

245 The authors declare that they have no known competing financial interests or personal
246 relationships that could have appeared to influence the work reported in this paper.

247 This article does not contain any studies with human participants or animals performed by
248 any of the authors.

249 This work relies on the online data available from Center for Systems Science and
250 Engineering (CSSE) at Johns Hopkins University.

251

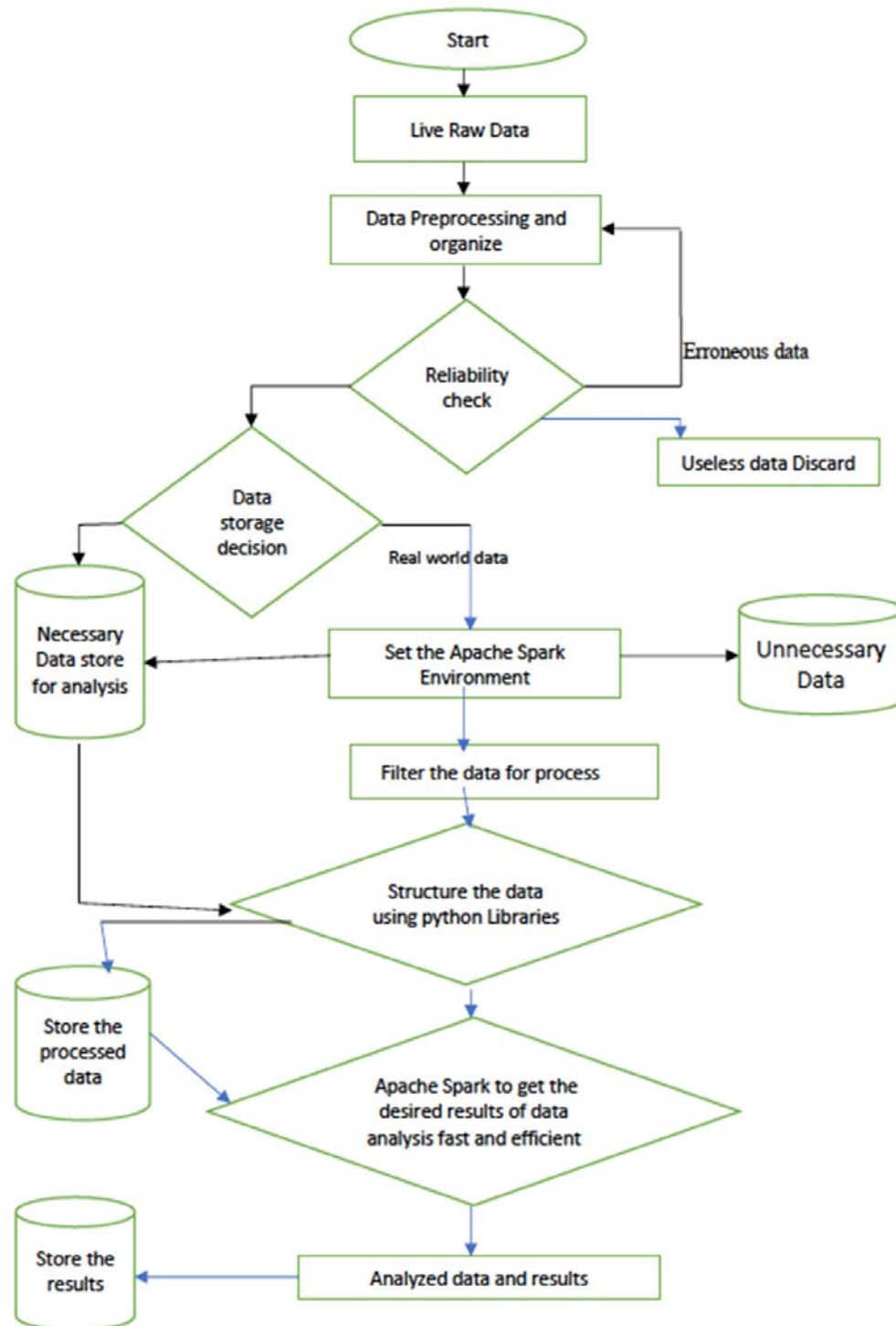
252 **References**

- 253 [1] M. You, Z. Wu, Y. Yang, J. Liu, D. Liu, Open Forum Infectious Diseases Spread of
254 Coronavirus 2019 From Wuhan to Rural Villages in the Hubei Province, Open Forum
255 Infect. Dis. (2020). <https://doi.org/10.1093/ofid/ofaa228>.
- 256 [2] C.N. Patel, S.P. Kumar, H.A. Pandya, R.M. Rawal, Identification of potential
257 inhibitors of coronavirus hemagglutinin-esterase using molecular docking, molecular
258 dynamics simulation and binding free energy calculation, Mol. Divers. (2020).
259 <https://doi.org/10.1007/s11030-020-10135-w>.
- 260 [3] C.N. Patel, D. Goswami, D.G. Jaiswal, R.M. Parmar, H.A. Solanki, H.A. Pandya,
261 Pinpointing the potential hits for hindering interaction of SARS-CoV-2 S-protein with
262 ACE2 from the pool of antiviral phytochemicals utilizing molecular docking and
263 molecular dynamics (MD) simulations., J. Mol. Graph. Model. 105 (2021) 107874.
264 <https://doi.org/10.1016/j.jm gm.2021.107874>.

- 265 [4] D. Cucinotta, M. Vanelli, WHO Declares COVID-19 a Pandemic, *Acta Biomed.* 91
266 (2020) 157–160. <https://doi.org/10.23750/abm.v91i1.9397>.
- 267 [5] C.N. Patel, D. Goswami, K. Sivakumar, H.A. Pandya, Repurposing of anticancer
268 phytochemicals for identifying potential fusion inhibitor for SARS-CoV-2 using
269 molecular docking and molecular dynamics (MD) simulations, (2021).
270 <https://doi.org/10.1080/07391102.2021.1902393>.
- 271 [6] E. Ong, M.U. Wong, A. Huffman, Y. He, COVID-19 coronavirus vaccine design using
272 reverse vaccinology and machine learning, *BioRxiv.* (2020).
273 <https://doi.org/10.1101/2020.03.20.000141>.
- 274 [7] A. Gupta, H.K. Thakur, R. Shrivastava, P. Kumar, S. Nag, A Big Data Analysis
275 Framework Using Apache Spark and Deep Learning, *IEEE Int. Conf. Data Min. Work.*
276 *ICDMW. 2017-Novem (2017)* 9–16. <https://doi.org/10.1109/ICDMW.2017.9>.
- 277 [8] I. Stancin, A. Jovic, An overview and comparison of free Python libraries for data
278 mining and big data analysis, 2019 42nd Int. Conv. Inf. Commun. Technol. Electron.
279 Microelectron. *MIPRO 2019 - Proc.* (2019) 977–982.
280 <https://doi.org/10.23919/MIPRO.2019.8757088>.
- 281 [9] A. Ghaffar, S.& Tariq, R. Soomro, A.G. Shoro, & Tariq, Big Data Analysis: Ap Spark
282 Perspective, *Type Double Blind Peer Rev. Int. Res. J. Publ. Glob. Journals Inc.* 15
283 (2015).
- 284 [10] M. Assefi, E. Behraves, G. Liu, A.P. Tafti, Big data machine learning using apache
285 spark MLlib, *Proc. - 2017 IEEE Int. Conf. Big Data, Big Data 2017.* 2018-Janua
286 (2017) 3492–3498. <https://doi.org/10.1109/BigData.2017.8258338>.
- 287 [11] R. Magar, P. Yadav, A. Barati Farimani, Potential neutralizing antibodies discovered
288 for novel corona virus using machine learning, *Sci. Rep.* 11 (2021) 5261.
289 <https://doi.org/10.1038/s41598-021-84637-4>.
- 290 [12] H.S. Badr, H. Du, M. Marshall, E. Dong, M.M. Squire, L.M. Gardner, Association
291 between mobility patterns and COVID-19 transmission in the USA: a mathematical
292 modelling study, *Lancet Infect. Dis.* 20 (2020) 1247–1254.
293 [https://doi.org/10.1016/S1473-3099\(20\)30553-3](https://doi.org/10.1016/S1473-3099(20)30553-3).
- 294 [13] A. Ammon, D. Faensen, Surveillance von Infektionskrankheiten auf europäischer
295 Ebene, *Bundesgesundheitsblatt - Gesundheitsforsch. - Gesundheitsschutz.* 52 (2009)

- 296 176–182. <https://doi.org/10.1007/s00103-009-0759-y>.
- 297 [14] S. Salloum, R. Dautov, X. Chen, P.X. Peng, J.Z. Huang, Big data analytics on Apache
298 Spark, *Int. J. Data Sci. Anal.* 1 (2016) 145–164. [https://doi.org/10.1007/s41060-016-](https://doi.org/10.1007/s41060-016-0027-9)
299 0027-9.
- 300 [15] F.A. Binti Hamzah, C.H. Lau, H. Nazri, D.C. Ligot, G. Lee, C.L. Tan, et al.,
301 CoronaTracker: World-wide Covid-19 outbreak data analysis and prediction, *Bull.*
302 *World Health Organ.* (2020) Submitted.
- 303 [16] S. Latif, M. Usman, W. Iqbal, J. Qadir, S. Manzoor, G. Tyson, I. Castro, A. Razi, M.N.
304 Kamel Boulos, A. Weller, J. Crowcroft, Leveraging Data Science To Combat COVID-
305 19: A Comprehensive Review TESCOON (Tools for Enforcement of Smart Contracts)
306 View project CONTRIVE View project Leveraging Data Science To Combat COVID-
307 19: A Comprehensive Review, 1 (2020) 85–103.
- 308 [17] T. Patel, GitHub - pateltanvi2992/Data-Analysis-of-Covid-19-using-Spark-: Data
309 Analysis of Covid-19(corona virus data) using Spark(Anaconda- Jupyter notebook),
310 (2021). <https://github.com/pateltanvi2992/Data-Analysis-of-Covid-19-using-Spark->
311 (accessed March 29, 2021).
- 312 [18] <https://www.coolantarctica.com/>, The Antarctic Population - Who lives in Antarctica?,
313 (n.d.). <https://www.coolantarctica.com/Antarctica> fact
314 file/science/can_you_live_in_antarctica.php (accessed March 29, 2021).
- 315 [19] <https://www.barrons.com/>, The Stock Market Is in a Manic Bull Market. How Long It
316 Will Last—and What to Buy. | Barron’s, (n.d.). [https://www.barrons.com/articles/the-](https://www.barrons.com/articles/the-stock-market-is-in-a-manic-bull-market-how-long-it-will-lastand-what-to-buy-51616760007)
317 [stock-market-is-in-a-manic-bull-market-how-long-it-will-lastand-what-to-buy-](https://www.barrons.com/articles/the-stock-market-is-in-a-manic-bull-market-how-long-it-will-lastand-what-to-buy-51616760007)
318 [51616760007](https://www.barrons.com/articles/the-stock-market-is-in-a-manic-bull-market-how-long-it-will-lastand-what-to-buy-51616760007) (accessed March 29, 2021).
- 319 [20] E.L. Bamgboye, J.A. Omiye, O.J. Afolaranmi, M.R. Davids, E.K. Tannor, S. Wadee,
320 A. Niang, A. Were, S. Naicker, COVID-19 Pandemic: Is Africa Different?, *J. Natl.*
321 *Med. Assoc.* (2020) 1–12. <https://doi.org/10.1016/j.jnma.2020.10.001>.
- 322 [21] D. Control, First Month of COVID-19 Vaccine Safety Monitoring — United States ,
323 70 (2021) 283–288.

324 List of Figures

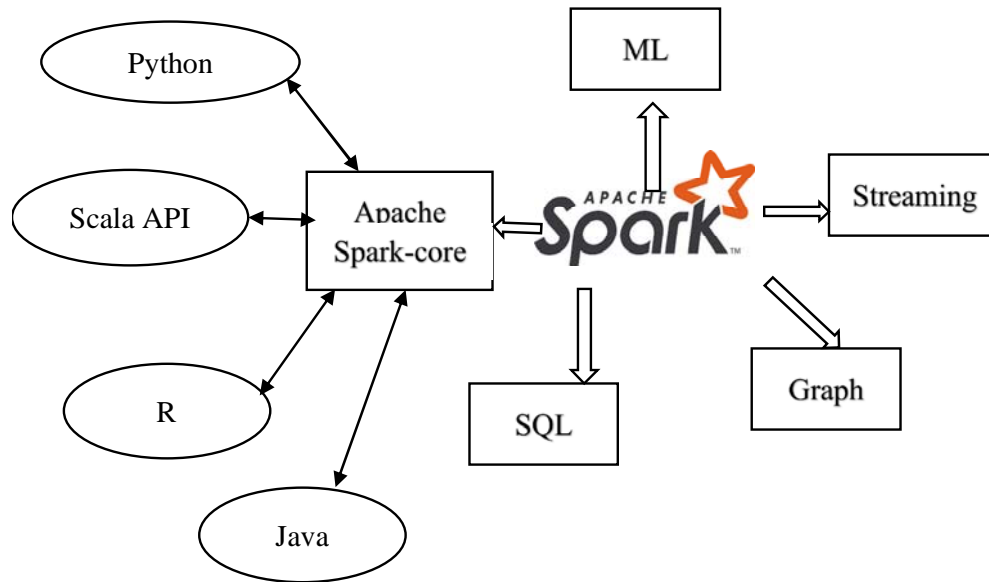


325

326

Figure-1 Flowchart of proposed study.

327

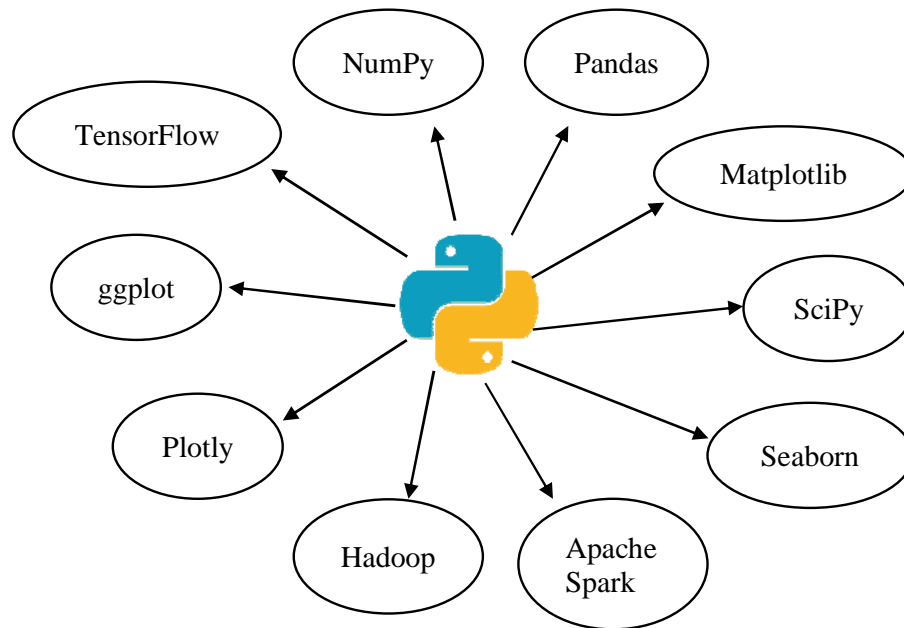


328

329

Figure-2 Apache Spark and API (Application Programming Interface).

330

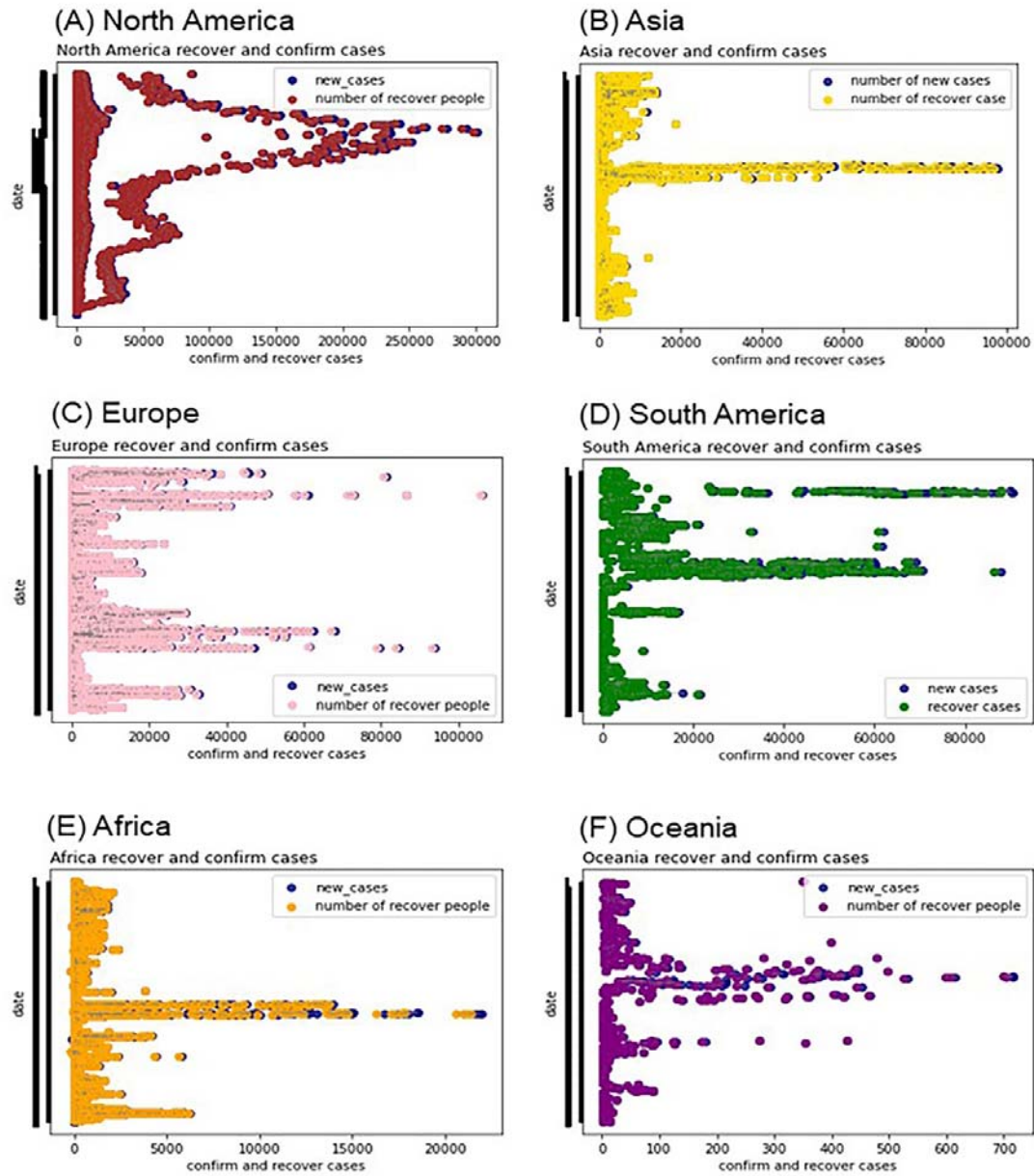


331

332

Figure-3 Python and it's used libraries.

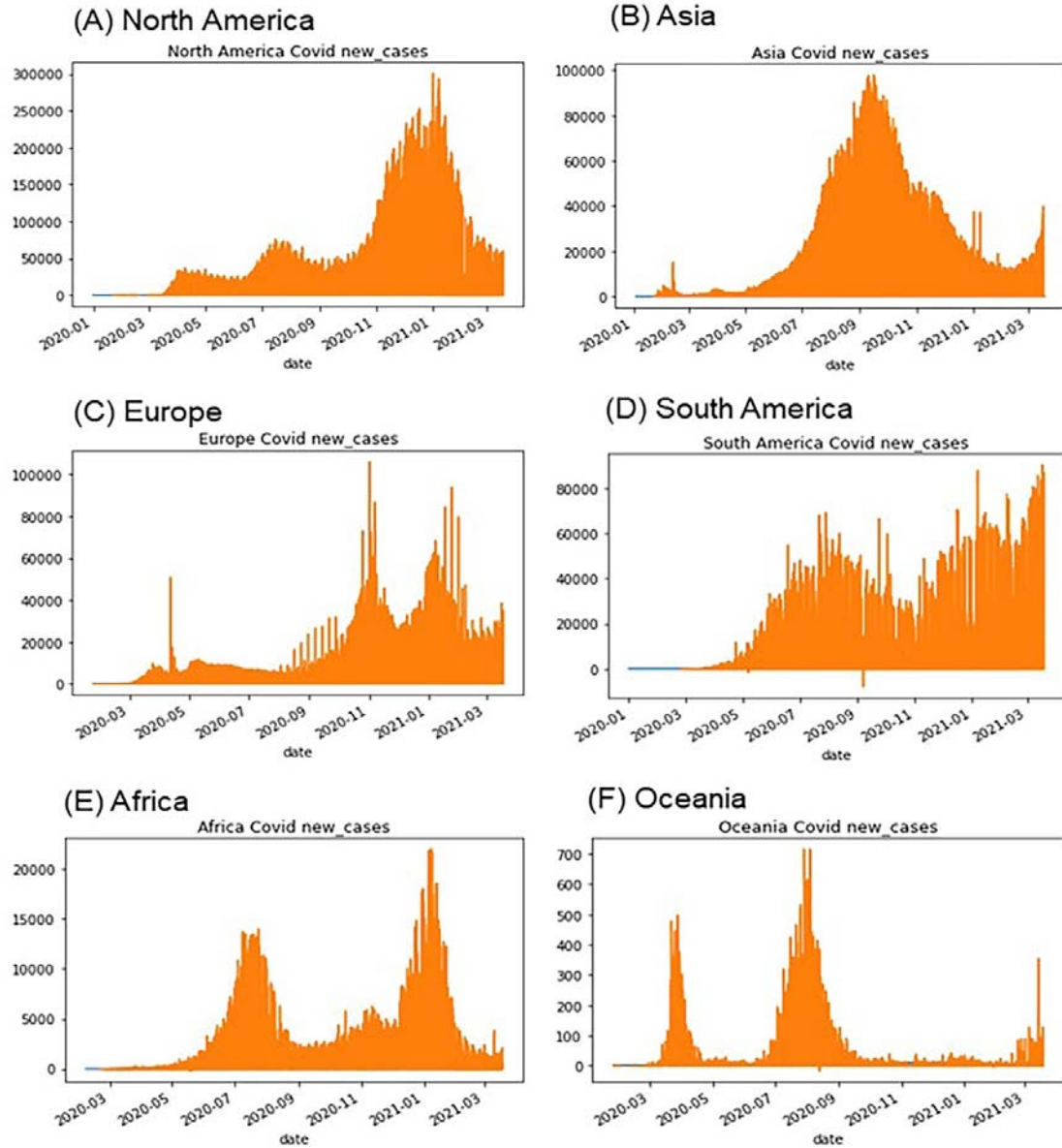
333



334

335 **Figure-4 COVID-19 cases distribution based on recovery rate from various continents.**

336



337

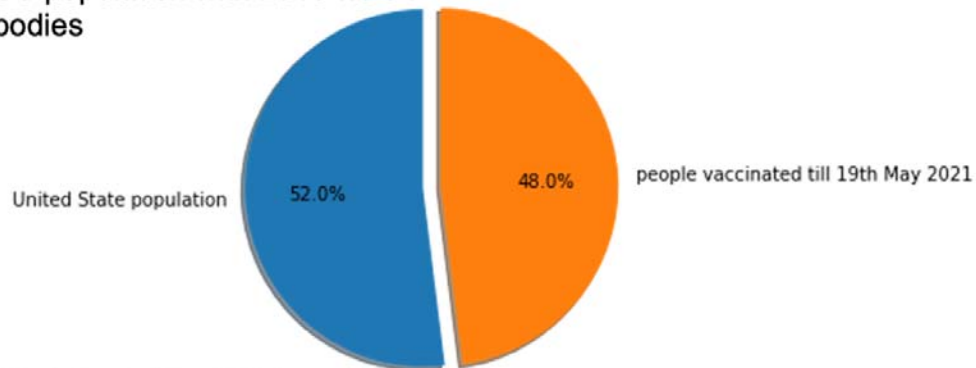
338

Figure-5 Histogrammic representation of COVID-19 updated cases in various regions.

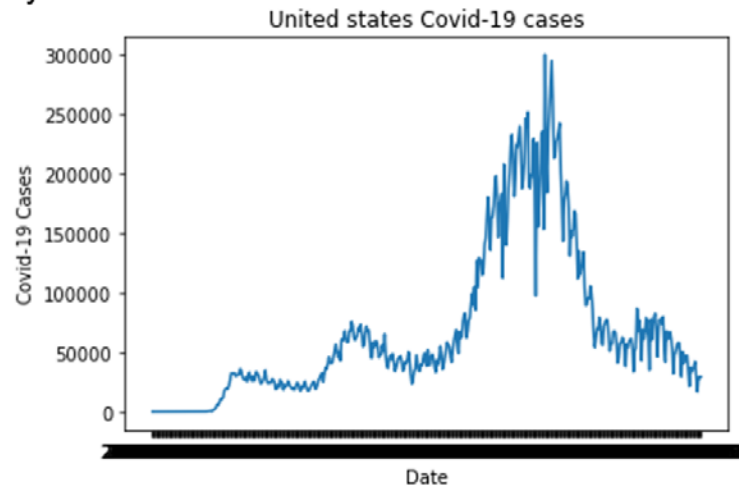
339

Fig 3.

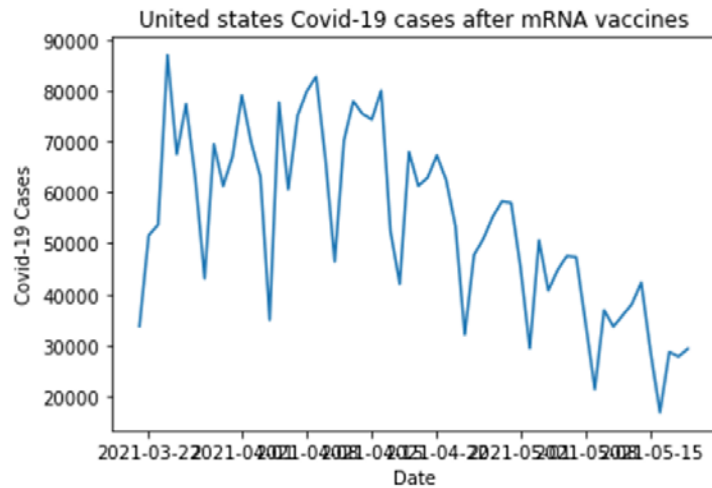
(A) US population with 2019-nCoV antibodies



(b) US daily 2019-nCoV scenario



(C) US daily 2019-nCoV scenario after antibodies



340

341

Figure-6 A comparison of case-based antibodies of COVID-19 observed in the U.S.

342

343 **Table-1 Covid-19 Tabular data (till May-15th-2021).**

344

Continents	Population	Total Tests	Positive cases	Total recover	People Vaccinated
Asia	4,669,823,825 (60.8%)	508,287,012	42,734,316	42,157,308	624,708,871
North America	368,869,647 (4.8%)	484,569,026	38,505,833	37,637,694	310,026,856
South America	433,340,178 (5.6%)	46,209,536	27,164,684	26,430,452	89,405,489
Europe	746,400,000 (9.7%)	662,528,505	46,740,089	45,688,120	255,813,054
Oceania	43,066,356 (0.56%)	12,149,407	46,922	45,851	3,501,157
Africa	1,364,509,774 (17.78%)	29,036,507	4,716,496	4,582,390	13,949,677

345

346

347

348

349

350

351

352