# COVID-19 Vaccine Data Review and Reactions on Social Media

Jake Maupin
jmaupin7@gmu.edu
George mason University

Mazen Mohamed
mmohem40@gmu.edu
George mason University

Shova Bhandari
sbhanda7@gmu.edu
George mason University

***Abstract*—The USA is one of the hardest hit areas by COVID-19. As of March 13th, 2021, 29.5M people were infected and 534K have lost their lives [1]. The USA government has authorized Pfizer and Moderna mRNA COVID-19 vaccines for the prevention of Coronavirus. However, these vaccines are allocated to certain groups and are not available to the public yet. Vaccine's demand is skyrocketing as the United States of America is unable to contain the virus, and a new more contagious and deadlier Covid-19 variant is emerging. Given the impact of the COVID-19 pandemic, it is imperative to efficiently distribute vaccines to contain and eventually eradicate the virus. It is important to identify the hardest hit region/state that is still struggling to fight and protect their residence from viruses. The main question we are attempting to answer is how we can use COVID-19 data to identify focal points for COVID-19 breakout and formulate an efficient and speedy response. In this research we analyze COVID-19 vaccine trend.**

***Index Terms*—Machine learning, COVID-19, Data, vaccine**

## I. Problem Description and Motivation

Next step in fighting the COVID-19 is widespread acceptance of the covid-19 vaccines. Achieving the widespread uptake might be challenging and may be obstructed by the misinformation prevalent in social media. The social media platforms have become a common source of information and disinformation on vaccines. Vaccine's hesitancy is more prevalent in social media, especially Twitter.

Some of them believe that the Covid-19 vaccine is part of a conspiracy [1]. Some groups of individuals are still against the vaccination campaign, as they are trying to invalidate and undermine the confidence of vaccines in social media, especially Twitter. According to Jamison et al, "Vaccine opponents shared the greatest proportion of unreliable information topics including a mix of conspiracy theories, rumors, and scams. Vaccine proponents shared a much lower proportion of unreliable information topics" [2]. It is important to identify groups undermining the confidence of vaccines to enlighten the effectiveness of vaccines and provide them with the facts and numbers to disperse the mendacity.

Therefore, it is important to identify focal points of breakout, identify vaccine hesitancy and conspiracy theory groups and predict the vaccine demand to reduce the anticipated strain on the healthcare system and prevent more contagious COVID-19 variant transmission. Health professionals and public health organizations should take vaccine hesitancy groups into account and devise a plan to fight social media misinformation/disinformation from affecting people's perception.

## II. Detailed Literature Search

Given the gravity that pandemic events have on the world, many studies have been conducted that looked at the optimal strategies and timing for vaccine distribution. Effective distribution of vaccines is paramount for mitigation of the transmission of the virus. Some of these studies are:

- One such article, titled "Geographic prioritization of distribution pandemic influenza vaccines", did a study that evaluated four strategies of vaccine distribution to 15 counties of Arizona; pro rata (simultaneously proportional to population size) distribution; sequential distribution by population size; sequential distribution by estimated periods of pandemic peaks; and reverse sequential distribution by estimated order of pandemic peaks. They found that the two most effective policies mitigating the risks of the pandemic were the pro rata distribution and prioritization of communities expected to experience the latest outbreak. Many of the current governmental and institutional guidelines for vaccine allocation are based on prorated type strategy, while the most fair are not generally the most optimal strategy.
- A research article called "Optimal Vaccine Allocation for the Early Mitigation of Pandemic Influenza" expands upon this conclusion. The authors of this research article conclude that their results suggest that cooperative strategies where the resources are optimally distributed perform much better than strategies where the vaccine is equally distributed among the networks. However, the model used within this study used a closed network which did not consider any PAGE 2 flow of population. The model was quite conservative, as they did not have enough data to create a more realistic model, as the model they used was approximated by a simple life-death process and did not differentiate between a sick child and a sick adult. Emotion and sentiment analysis of COVID data provides more insight about people situation in the pandemic [3]–[10].
  In order to minimize the transmissions of coronavirus during the flights, a temperature screening was implemented at several airports. However, it was an ineffective method as approximately 45% of people would be detected by airports body screening and lower result in younger people.COVID-19 misinformation detection models help

to collect clean data for this research [11] [6].
Additionally, previous studies reported that coronavirus could be transmitted before symptoms appear [12], [13]. As people can be infected with coronavirus disease and show no symptoms, or the symptoms appear after a period of several days. Overall, *Khanh et al.* [14], *Bae et al.* [12], and *Choi et al.* [13] concluded that coronavirus could be transmitted on aircraft and consequently increase the infection risk.The study compared the ratio between the international imported cases and the internal cases on May 2020 and September 2020 in different countries [15]. Application of machine learning models on social media [16]–[32] can help to get valuable insight about COVID-19 data.
They also addressed that while their optimal strategy significantly reduced the overall attack rate (total number of new cases divided by the total population), it created an inequitable distribution, where some locations are no epidemic while others experienced a massive epidemic. They mentioned that the most optimal solution, while not the best, that the population might get behind is vaccinating the children first in a prorated fashion.

- Another article, "Optimizing Influenza Vaccine Distribution", written by Jan Medlock and Alison Galvani, backs up that assertion. Their analysis concludes that a model optimized for the following five outcome measures; deaths, infections, years lost in life lost, contingent valuation, and economic cost, that the optimal vaccination is achieved by prioritizing children in school and their parents, age 30 to 39. Their model dictates that the consideration of age-specific transmission dynamics is the most important to the optimal allocation of vaccines. While insight can be taken from these models, they lack current data involved with the COVID-19 pandemic, especially taking into consideration the political problems currently encompassing the United States with trying to open public schools without having a mandatory vaccination. Along with various groups, specifically social media groups, denouncing vaccines and spreading misinformation, new models and information need to be presented to the public to provide a trustworthy, optimized vaccine distribution plan that can be implemented with maximum speed that the public can get behind.

Despite widespread suffering from Covid-19, surveys indicate that the proportion of the U.S. population willing to be vaccinated has fluctuated from 72% in May to 51% in September and 60% in November; of the 39% of respondents who indicated that they probably or would not get the vaccine, only 46% said they might be open to vaccination once others start getting it and more information becomes available [33].

- An article from the New England Journal of Medicine attempts to address this by suggesting 12 key strategies for an effective vaccine-promotion effort, as seen on Table 1 [34]. They go on to describe how each tactic could be used to help boost the willingness and participation of the public in taking vaccines. The most important one they addressed was the issue of the base-rate fallacy, where people tend to ignore the baseline statistics and focus more on anecdotal cases. In the case of COVID vaccines, it is specifically an issue with cases of bad reactions or "Death after vaccination" that get the most media attention. The article however states that the strategies must be targeted towards a specific group or area to be effective.
- An article by natural Public Radio states that "Instagram recommended false claims about COVID-19, vaccines and the 2020 U.S. election to people who appeared interested in related topics, according to a new report from a group that tracks online misinformation." [35]. This was based off of Instagram's "suggested posts" feature, which is intended to keep users on the app for longer periods of time by recommending them content similar to their previous searches that they do not follow. The Center for Countering Digital Hate says Instagram should stop recommending posts and it should exclude posts about COVID-19 or vaccines from being recommended at all [6]. By continuing to feed false narratives into people's feeds from sources they have not "put their trust in".
- A Research paper called "Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA" [36] did a study of the effects of vaccine misinformation on people's intent to get vaccinated. It found that misinformation about COVID-19 vaccine lowered the intent to get vaccinated by 6.2% in the UK and 6.4% in the US in the surest category, or "Definitely". The participants were also asked to report if the image presented to them, misinformation, or information, was trustworthy. 25 % of the participants found that the misinformation given to them was trustworthy and agreed with it. However, since this was a PAGE 3 controlled experiment, done using participation polls, it is an unlikely representation of misinformation in a real-world social media setting. Individuals will not be exposed in the same volume or rate as the controlled experiment, as that will depend on their social media presence/preference. Studies have shown that brief exposure to misinformation can embed itself into long-term memory [37]. with how the real-world social media platform environment is complex and can promote echo chambers and selective reasoning, misinformation has been shown to spread faster than factually correct information. [37] Addressing the spread of misinformation will be a major component of a successful vaccination campaign.
- Another research paper, "Prevalence of health Misinformation on Social Media: Systematic Review" looks at the main health misinformation topics and their acceptance on different social media platforms [38]. They found that twitter was found to have the most amount of misinformation, especially when it comes to health misinformation, like COVID-19. They found that the information was often framed with scientific appearances against scien-

TABLE I
TIME PLAN

| # | Description | Due by |
|---|---|---|
| 1 | Analyzing social media data | 03/23 |
| 2 | Correlating social media data to vaccines data | 03/30 |
| 3 | Review results with more data | 04/26 |
| 4 | Finish project paper and presentation | 05/02 |

tific evidence. Those exposed to negative anti-vaccine information were more likely to spread negative concerns about the vaccine. They go on to say that "the greatest challenge lies in the difficulty of characterizing and evaluating the quality of the information on social media.". Future studies into this topic should be focused on developing evidence-based social policy action plans aimed at combating the public health problem through different social media platforms.

## III. PROPOSED APPROACH

Our approach is to analyze the COVID-19 cases/deaths in the US and compare the numbers of mRNA COVID-19 vaccines administered in each region/state with the number of cases/deaths, based on data obtained from Kaggle data repository and John Hopkins University, and track with time the effectiveness of the allocated doses in reducing the number of cases/deaths. The number of people vaccinated and people fully vaccinated are analyzed compared to the total number of vaccinations as well as relating people's reactions to the vaccine on the social media, and how the social media can play an effective role in decreasing/increasing the number of vaccinated people.

## IV. PRELIMINARY RESULTS

Our first step was to track the number of cases/deaths in the US since the start of the pandemic until now, shown in Fig. 1. The plot shows a significant decrease in the number of cases/deaths starting from January 2021, which is due to the vaccine distribution campaign. Then, we merged the cases/deaths data with the vaccinations data to come up with some useful insights. Our first conclusion is that the number of vaccines allocated daily to every state is most correlated with the number of daily deaths. Fig 2 and Fig 3 show the total number of vaccines allocated to every state and the total number of deaths since the distribution of vaccines, respectively

Our goal is to find the effect of social media on the number of people vaccinated or fully vaccinated. Fig 4 shows the number of vaccinations, people vaccinated, and people fully vaccinated for every state. We started by calculating the ratio between people vaccinated/fully vaccinated and total vaccinations and found out that some states like Arizona and Utah show a relatively lower ratio to the average ratio. On the other hand, a state like West Virginia shows a relatively higher ratio. Our next approach is to analyze the tweets on COVID-19, and how it can be related to the variance of the number of people vaccinated/fully vaccinated from the average ratio.

## REFERENCES


[1] N. Ray, "Twitter text analytics reveals covid-19 vaccine hesitancy tweets have crazy traction." https://dataconomy.com/2020/12/twitter-text-analytics-reveals-covid-19-vaccine-hesitancy-tweets-have-crazy-traction/, 2020, December 03.

[2] A. M. Jamison, D. A. Broniatowski, M. Dredze, A. Sangraula, M. C. Smith, and S. C. Quinn, "Not just conspiracy theories: Vaccine opponents and proponents add to the COVID-19 'infodemic' on twitter," *Harvard Kennedy School Misinformation Review*, Sept. 2020.

[3] S. Zad, M. Heidari, J. H. J. Jones, and O. Uzuner, "Emotion detection of textual data: An interdisciplinary survey," in *IEEE 2021 World AI IoT Congress,AIIoT2021*, 2021.

[4] A. Adekunle, M. Meehan, D. Rojas-Alvarez, J. Trauer, and E. McBryde, "Delaying the COVID-19 epidemic in australia: evaluating the effectiveness of international travel bans," *Australian and New Zealand Journal of Public Health*, vol. 44, pp. 257–259, July 2020.

[5] M. Heidari, S. Zad, B. Berlin, and S. Rafatirad, "Ontology creation model based on attention mechanism for a specific business domain," in *IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*, 2021.

[6] M. Bielecki, D. Patel, J. Hinkelbein, M. Komorowski, J. Kester, S. Ebrahim, A. J. Rodriguez-Morales, Z. A. Memish, and P. Schlagenhauf, "Air travel and COVID-19 prevention in the pandemic and peri-pandemic period: A narrative review," *Travel Medicine and Infectious Disease*, vol. 39, p. 101915, Jan. 2021.

[7] M. Heidari and J. H. Jones, "Using bert to extract topic-independent sentiment features for social media bot detection," in *2020 11th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, pp. 0542–0547, 2020.

[8] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. P. y Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. E. Halloran, I. M. Longini, and A. Vespignani, "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," *Science*, vol. 368, pp. 395–400, Mar. 2020.

[9] S. Zad, M. Heidari, J. H. J. Jones, and O. Uzuner, "A survey on concept-level sentiment analysis techniques of textual data," in *IEEE 2021 World AI IoT Congress,AIIoT2021*, 2021.

[10] M. Heidari and S. Rafatirad, "Using transfer learning approach to implement convolutional neural network to recommend airline tickets by using online reviews," in *IEEE 2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP 2020*, 2020.

[11] M. Heidari, J. H. J. Jones, and O. Uzuner, "Misinformation detection model to prevent spread of the covid-19 virus during the pandemic," 2021.

[12] S. H. Bae, H. Shin, H.-Y. Koo, S. W. Lee, J. M. Yang, and D. K. Yon, "Asymptomatic transmission of SARS-CoV-2 on evacuation flight," *Emerging Infectious Diseases*, vol. 26, pp. 2705–2708, Nov. 2020.

[13] E. M. Choi, D. K. Chu, P. K. Cheng, D. N. Tsang, M. Peiris, D. G. Bausch, L. L. Poon, and D. Watson-Jones, "In-flight transmission of SARS-CoV-2," *Emerging Infectious Diseases*, vol. 26, pp. 2713–2716, Nov. 2020.

[14] N. C. Khanh, P. Q. Thai, H.-L. Quach, N.-A. H. Thi, P. C. Dinh, T. N. Duong, L. T. Q. Mai, N. D. Nghia, T. A. Tu, L. N. Quang, T. D. Quang, T.-T. Nguyen, F. Vogt, and D. D. Anh, "Transmission of SARS-CoV 2 during long-haul flight," *Emerging Infectious Diseases*, vol. 26, pp. 2617–2624, Nov. 2020.

[15] T. W. Russell, J. T. Wu, S. Clifford, W. J. Edmunds, A. J. Kucharski, and M. Jit, "Effect of internationally imported cases on internal spread of COVID-19: a mathematical modelling study," *The Lancet Public Health*, vol. 6, pp. e12–e20, Jan. 2021.

[16] S. Chen, S. Owusu, and L. Zhou, "Social network based recommendation systems: A short survey," in *2013 International Conference on Social Computing*, pp. 882–885, 2013.

[17] S. Lin, C. Liu, and Z.-K. Zhang, "Multi-tasking link prediction on coupled networks via the factor graph model," in *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*, pp. 5570–5574, 2017.

[18] M. Heidari, J. H. J. Jones, and O. Uzuner, "Deep contextualized word embedding for text-based online user profiling to detect social bots on twitter," in *IEEE 2020 International Conference on Data Mining Workshops (ICDMW), ICDMW 2020*, 2020.

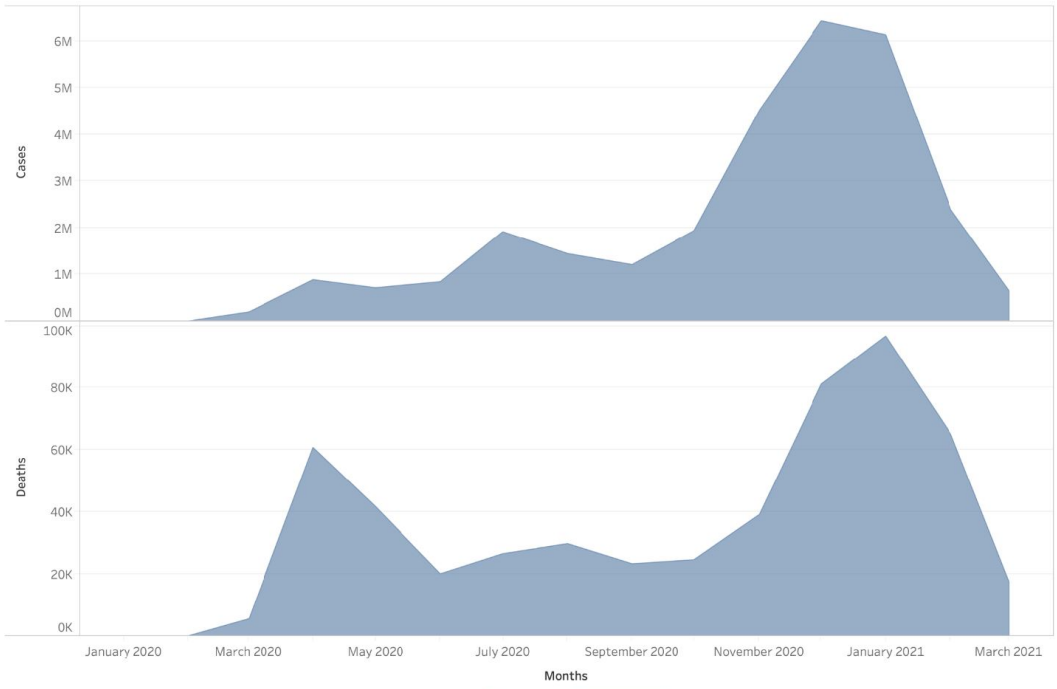


Figure 1: Number of cases/deaths across time

Fig. 1. Number of cases/deaths across time

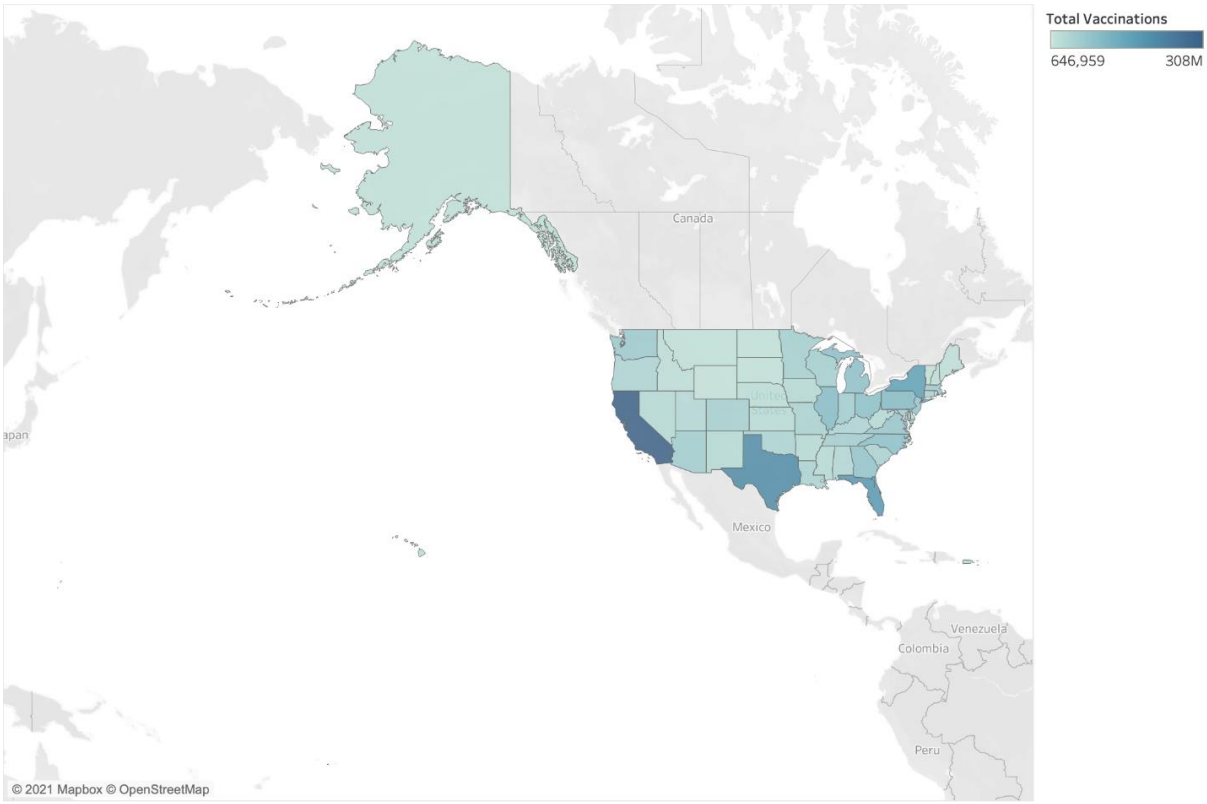


Fig. 2. Total vaccinations allocated to every state

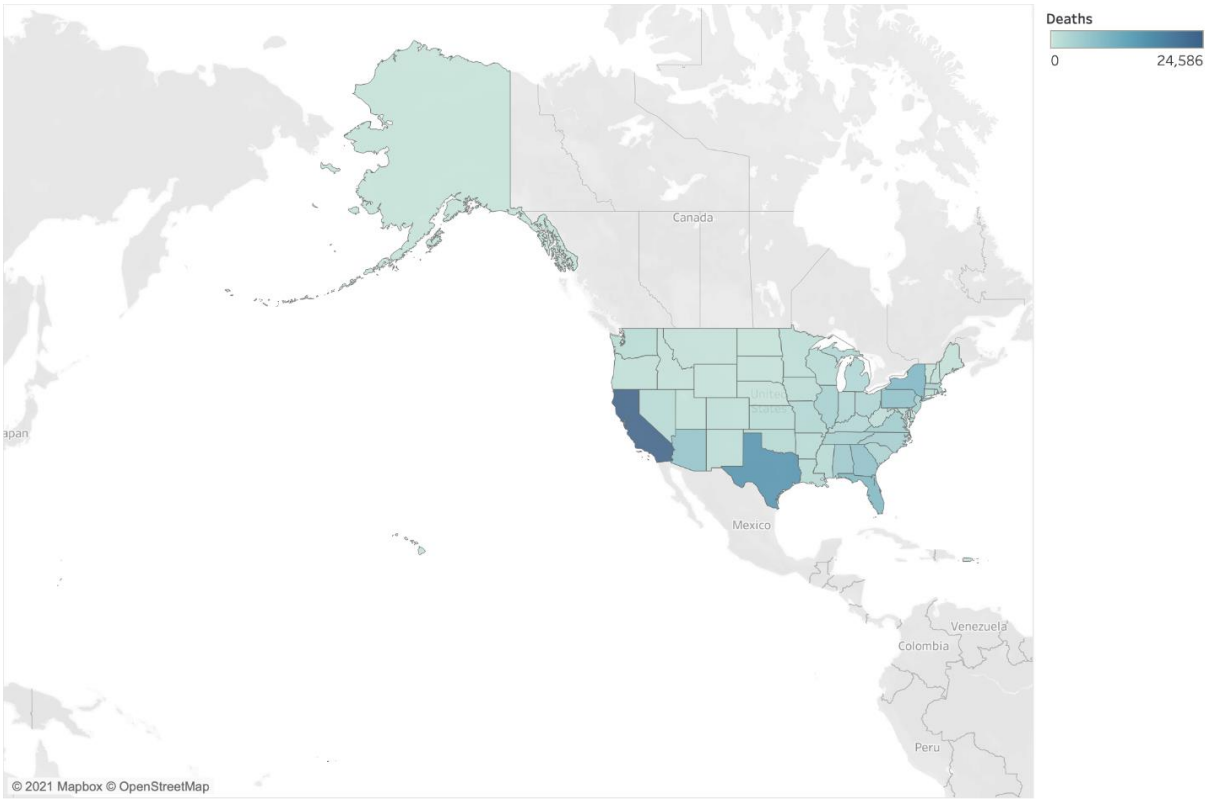


Fig. 3. Total deaths since vaccination campaign

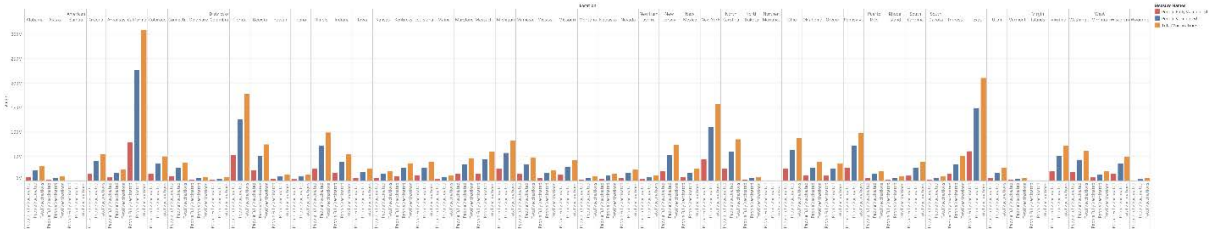

Fig. 4. Number of vaccinations, people vaccinated, and people fully vaccinated per state


[19] Y. Chu, F. Huang, H. Wang, G. Li, and X. Song, "Short-term recommendation with recurrent neural networks," in *2017 IEEE International Conference on Mechatronics and Automation (ICMA)*, pp. 927–932, 2017.

[20] C. Yang, X. Chen, T. Song, B. Jiang, and Q. Liu, "A hybrid recommendation algorithm based on heuristic similarity and trust measure," in *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, pp. 1413–1418, 2018.

[21] S. Ji and J. Liu, "Interpersonal ties and the social link recommendation problem," in *2019 6th International Conference on Systems and Informatics (ICSAI)*, pp. 456–462, 2019.

[22] M. Heidari and S. Rafatirad, "Bidirectional transformer based on online text-based information to implement convolutional neural network model for secure business investment," in *IEEE 2020 International Symposium on Technology and Society (ISTAS20), ISTAS20 2020*, 2020.

[23] J. Wang, H. Song, and X. Zhou, "A collaborative filtering recommendation algorithm based on biclustering," in *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, pp. 803–807, 2015.

[24] S. Chen, S. Owusu, and L. Zhou, "Social network based recommendation systems: A short survey," in *2013 International Conference on Social Computing*, pp. 882–885, 2013.

[25] M. Heidari and S. Rafatirad, "Semantic convolutional neural network model for safe business investment by using bert," in *IEEE 2020 Seventh International Conference on Social Networks Analysis, Management and Security, SNAMS 2020*, 2020.

[26] A. Gatzioura, J. Vinagre, A. M. Jorge, and M. Sànchez-Marrè, "A hybrid recommender system for improving automatic playlist continuation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 5, pp. 1819–1830, 2021.

[27] Z. Liao, Y. Song, Y. Huang, L.-w. He, and Q. He, "Task trail: An effective segmentation of user search behavior," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 12, pp. 3090–3102, 2014.

[28] M. Heidari, J. H. J. Jones, and O. Uzuner, "An empirical study of machine learning algorithms for social media bot detection," in *IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*, 2021.

[29] C.-Y. Chi, Y.-S. Wu, W.-r. Chu, D. C. Wu, J. Y.-j. Hsu, and R. T.-H. Tsai, "The power of words: Enhancing music mood estimation with textual input of lyrics," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pp. 1–6, 2009.

[30] A. Gatzioura, J. Vinagre, A. M. Jorge, and M. Sànchez-Marrè, "A hybrid recommender system for improving automatic playlist continuation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 5,

pp. 1819–1830, 2021.
[31] H. Yang, C. He, H. Zhu, and W. Song, "Prediction of slant path rain attenuation based on artificial neural network," in *2000 IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 1, pp. 152–155 vol.1, 2000.
[32] M. Heidari, S. Zad, and S. Rafatirad, "Ensemble of supervised and unsupervised learning models to predict a profitable business decision," in *IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*, 2021.
[33] C. Funk and A. Tyson, "Intent to get a covid-19 vaccine rises to 60% as confidence in research and development process increases." https://www.pewresearch.org/science/2020/12/03/intent-to-get-a-covid-19-vaccine-rises-to-60-as-confidence-in-research-and-development-process-increases/, December 3, 2020.
[34] S. Wood and K. Schulman, "Beyond politics — promoting covid-19 vaccination in the united states," *New England Journal of Medicine*, vol. 384, p. e23, Feb. 2021.
[35] S. Bond, "Instagram suggested posts to users. it served up covid-19 falsehoods, study finds," 2021, March 9.
[36] S. Loomba, A. de Figueiredo, S. J. Piatek, K. de Graaf, and H. J. Larson, "Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA," *Nature Human Behaviour*, vol. 5, pp. 337–348, Feb. 2021.
[37] B. Zhu, C. Chen, E. F. Loftus, Q. He, C. Chen, X. Lei, C. Lin, and Q. Dong, "Brief exposure to misinformation can lead to long-term false memories," *Applied Cognitive Psychology*, vol. 26, pp. 301–307, Dec. 2011.
[38] V. Suarez-Lledo and J. Alvarez-Galvez, "Prevalence of health misinformation on social media: Systematic review," *Journal of Medical Internet Research*, vol. 23, p. e17187, Jan. 2021.