EXECUTIVE SUMMARY

Background :

The ongoing COVID-19 pandemic had a one true solution- a mass vaccination drive through vaccines created just a year ago. While the side effects of the vaccine are stated by medical experts, little is known about how the side effects vary across individuals, dosage, vaccine type and gender and what exactly do the people feel about the vaccine.

Objective :

This study analyzed the side effects and other major topics that people tweeted about after taking their first/second and booster doses of COVID-19 vaccine and aims to infer the health-related insights by identifying differences in side effects, medications to mitigate them and other major topics such as getting covid despite/after taking a vaccine and taking the vaccine with a pre existing chronic health condition across individuals, dosage and vaccine type. The study also aims to identify the response of people towards vaccines and trace patterns to deduce psychological perceptions.

Methods :

The data was gathered by scraping tweets via Snscrape from January 2021 to December 2022 for major vaccines such as Pfizer, Moderna, Johnson and Johnson, AstraZeneca etc. The final vaccines that would be considered for analysis were influenced by memory constraints to scrape the data, number of tweets and were selected to be Pfizer Jan 2022 to Dec 2022), AstraZeneca, Johnson and Johnson, Novavax (Jan 2021 to Dec 2022).

A sample was drawn from user tweets for users who have been vaccinated in this duration. Additionally, the next 7-day tweets for these users were extracted to gather responses against the vaccine doses. These textual tweets were processed using text mining to extract, dosage, vaccine type. These variables were then fed into a basic bag-of-words model containing health related keywords to segment only health related tweets. Finally, all these tweets were concatenated against the user to identify the topic strength for five identified topics using a Zero Shot Classifier, the five topics being, Arm related side effects, flu related side effects, medications to mitigate the side effects of the vaccine, getting covid-19 after taking the vaccine and taking the vaccine with a pre-existing condition.

Results:

Based on the user tweets, the following insights were obtained from the data prepared through the methods mentioned above.

1. Arm related side effects appear to be higher for Novavax for dose 1 and dose 2.
2. Pfizer had significantly higher topic scores for flu related side effects as compared to other vaccines.
3. The booster dose had higher topic scores for flu related side effects as opposed to dose 1 and dose 2 across all vaccines.
4. Novavax had a significantly higher topic score for "Getting covid after taking the vaccine" as opposed to other vaccines.
5. The topic scores for "Taking a Covid-19 Vaccine with a Pre-Existing Chronic Health Condition" remained almost comparable across all vaccines and doses.
6. An ANOVA was also conducted to check if these differences were statistically significant. ANOVA proved that these differences are not statistically significant and hence, further tests and analysis may be needed to affirm the same

Conclusion:

The COVID-19 vaccine has set a record for being the fastest vaccine ever developed and brought to market. While researchers may take some time to fully comprehend the vaccine's side effects in different populations due to the rapid testing conducted across the three phases, analyzing user responses via tweets is currently the most effective way to study the vaccine's side effects and related topics during the early phases of vaccination. This study serves as a valuable starting point for identifying the side effects of the vaccine based on vaccine types and doses. However, it could be further optimized with time and expertise to explore the side effects by age, race, and chronic conditions and gain deeper insights from people's responses.