

**Big Data Algorithms (Fall 2021) Project**

Doug Jih

Harica Bhogavalli Naga Lakshmi

Vaccine Adverse Event Reporting System

(VAERS)

**Table of Contents**

1. **Abstract**……………………………………………………………………………..….. 2
2. **Motivation**……………………………………………………………………………… 3
3. **Data Description**…………………………………………………………………….. 4
4. **Data Analysis/Algorithms**……………………………………………………….. 6
5. **Significance**………………………………………………………………………………
6. **Conclusion**…………………………………………………………………………………
7. **References**…………………………………………………………………………………
8. **Contributions**…………………………………………………………………………….

**Abstract**

Our project is based on analysing the Vaccine Adverse Event Reporting System (VAERS) data. The Vaccine Adverse Event Reporting System (VAERS) was created by the Food and Drug Administration (FDA) and Centers for Disease Control and Prevention (CDC) to receive reports about adverse events that may be associated with vaccines. No prescription drug or biological product, such as a vaccine, is completely free from side effects and though vaccines protect many people from dangerous illnesses they can cause side effects some which may be serious. VAERS researchers apply procedures and methods of analysis to help us closely monitor the safety of vaccines.

The VAERS data are in 3 types of CSV files and the dataset used here is of the year 2021, updated on November 28. The main data file contains individual reports including patient details such as state, age, sex, report date, previous medical history, allergies, symptoms and so on. The VAX file provides vaccine information such as name, lot number, type. The Symptoms file provides symptoms coded according to the MedDRA (Medical Dictionary for Regulatory Activities) dictionary. These three tables are correlated by the "VAERS\_ID" column as the primary/foreign key. The merged data from contains 993,374 records with 51 columns each.

Our goal in this project is to perform Exploratory Data Analysis on the dataset to get useful insights from data, uncover underlying structure of the dataset and detect outliers and anomalies if present. Building a dynamic visualizations using python libraries to show important patterns depending on geographical regions, gender, symptoms reported and vaccine manufacturers. With implementation of frequent itemset and association rule algorithms we have inferred insightful information and interesting patterns in the results.

**Motivation**

The Covid-19 pandemic, a major health disaster, has surrounded the world for nearly 2 years now. The pandemic has effected the lives of almost each and every person across the globe. With the fear of spread of the virus and its consequences, people have been eagerly waiting for a cure. The advancements in the field of invention of vaccine, gave a positive hope to everyone. But many people have reported of adverse effects and other health concerns post vaccination. It is a fact that different types of vaccines work in different ways to offer protection and at times result in side effects.

In our project we have used visualization of the covid vaccine adverse effects and patterns to transform data into an engaging story with details. Data visualization enables recognition of emerging trends. Such patterns make more sense when graphically represented; because visuals and diagrams make it easier for us to identify strongly correlated parameters. Humans can process visual images 60,000 times faster than text. Therefore, seeing a graph, chart, or other visual representation of data is more comfortable for the brain to process.

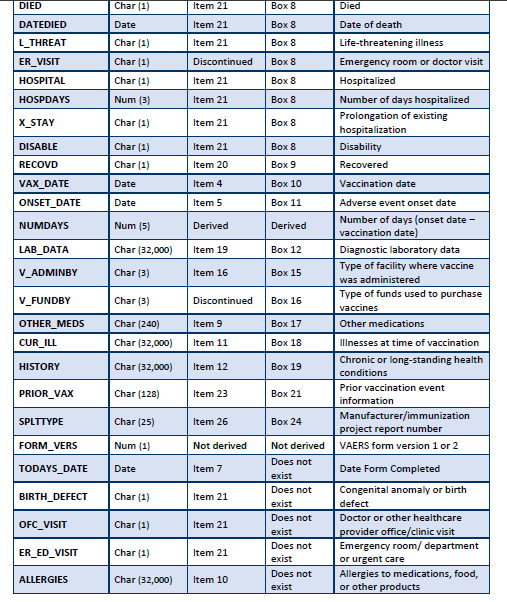
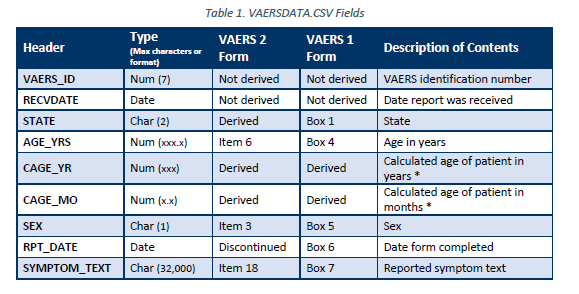
We have aimed to determine the number of adverse effects reported by people per each state of US, for each of the COVID vaccine manufacturer and the top 5 symptoms categorized based on gender and vaccine manufacturers. Apart from just data visualization, we have also tried to determine the correlation among the factors affecting the adverse effects of vaccine using FP-Growth algorithm and association rule generation.

**Data Description**

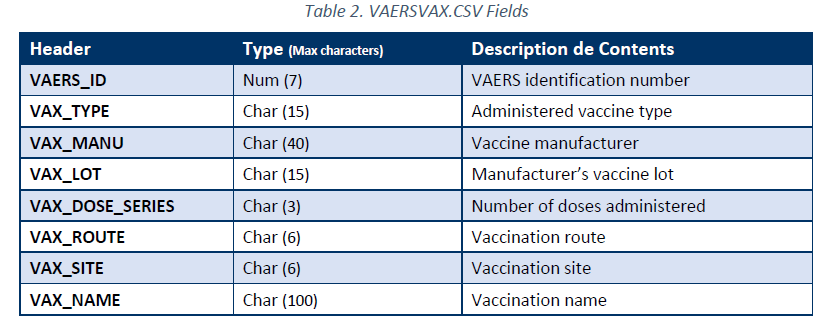
VAERS data is accessible by two mechanisms: by downloading raw data in comma-separated value (CSV) files for import into a database, spreadsheet, or text editing program, or by use of the CDC WONDER online search tool. The downloadable VAERS public data set consists of three separate data files namely:

* VAERSDATA.CSV
* VAERSVAX.CSV
* VAERSSYMPTOMS.CSV

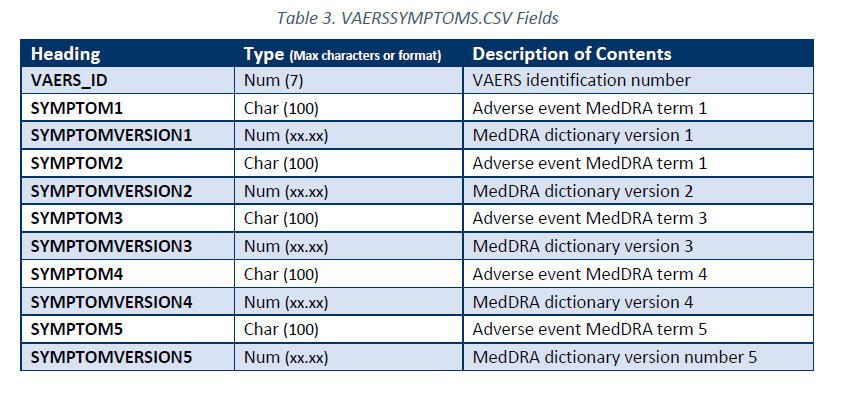
The VAERSDATA.CSV file contains patient details such as state, age, sex, report date, previous medical history, allergies, symptoms and so on.



The fields described in **VAERSVAX** table provide vaccine information (e.g., vaccine name, manufacturer, lot number, route, site, and number of previous doses administered), for each of the vaccines listed. There is a matching record in this file with the VAERSDATA file identified by VAERS\_ID.



The fields described in **VAERSSYMPTOMS** provides the adverse event coded terms utilizing the MedDRA dictionary. Each row in the CSV contains up to five MedDRA terms per VAERS ID; thus, there might be multiple rows per VAERS ID. For each of the VAERS\_ID’s listed in the VAERSDATA.CSV table, there is a matching record in this file, identified by VAERS\_ID.



**Data Analysis/Algorithms**

**References:**

1. <https://vaers.hhs.gov/about.html>
2. <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/different-vaccines/how-they-work.html>