



Analysis on Side Effects of Covid 19 Vaccination *

Pravallika Vallapuri

Master of Science in Data Analytics

Spring 2021

San Jose State University

pravallika.vallapuri@sjsu.edu

Nikita Balani

Master of Science in Data Analytics

Spring 2021

San Jose State University

nikita.balani@sjsu.edu

Nirvisha Garara

Master of Science in Data Analytics

Spring 2021

San Jose State University

nirvisha.garara@sjsu.edu

Jinu Rose Mathew

Master of Science in Data Analytics

Spring 2021

San Jose State University

jinurose.mathew@sjsu.edu

Vineeth Reddy Chinthala

Master of Science in Data Analytics

Spring 2021

San Jose State University

vineethreddy.chinthala@sjsu.edu

Abstract—COVID-19 is spread by the SARS-CoV-2 virus (CoronaVirus Infectious Disease-19). In December of 2019, the first case was identified. The COVID-19 pandemic is wreaking havoc on people all over the world right now. Because of the virus's scope and magnitude, as well as the lack of a clear clinical range, vaccine development has accelerated at an exponential pace. Vaccine portals and methods are being studied in a variety of ways. The aim of our investigation is to figure out which vaccine caused the most harm, which age demographic was most affected by the vaccine's side effects, and how vaccination affects patients who are already infected.

I. INTRODUCTION

The Food and Drug Administration (FDA) and the Centers for Disease Control and Prevention (CDC) developed the Vaccine Adverse Event Reporting System (VAERS) to gather details on vaccine-related adverse effects. There is no such thing as a side effect-free prescription medication or biological device, such as a vaccine. Many people are protected from harmful diseases by vaccinations, but vaccines, like medications, may have side effects, some of which are severe. VAERS is used to continuously track data to see whether every drug or vaccine lot has a higher number of events than predicted.

Doctors and other vaccine suppliers are advised to monitor adverse effects, even though they are unsure if the injection is to blame. Since it's impossible to tell the difference between a coincidental occurrence and one triggered by a vaccine, the VAERS archive will have all kinds of events.

Furthermore, it is common for more than one vaccine to be given, making it impossible to pinpoint which vaccine was responsible for the incident. When reviewing individual cases, analysts look at the medical records about the incident and, if possible, get more detailed information from the media. The monitoring patterns for vaccinations and vaccination lots are

examined.

Fevers, redness, and swelling at the injection site account for between 85-90 percent of vaccine adverse effect reports. The remaining records (less than 15 percentage) detail critical incidents including hospitalizations, life-threatening injuries, or casualties. Severe case accounts are of the utmost importance to VAERS, and they are scrutinized with the utmost caution. Researchers at VAERS use protocols and techniques of study to help them keep a close eye on vaccine safety.

II. SIGNIFICANCE TO THE REAL WORLD

When vaccines square measure prepared for tests in humans, they're tested on thousands to tens of thousands of volunteers. However, even this huge variety isn't continually enough to search out rare aspect effects, like a one in-a-million aspect impact. Sometimes, it's solely when a vaccine has been approved and used loosely that rare aspect effects are detected by observation systems like VAERS. So, VAERS is required to perpetually search for attainable aspect effects that may not be detected antecedently.

There is currently no consensus on how to treat a covid infection. There is also a misunderstanding about how effective vaccines are in combating the virus. People are generally distrustful of healthcare facilities, so as a result, these specifics can be used sed in future messaging. By engaging in these behaviors, we reduce our chances of being ill or contracting the illness. COVID-19 is a human-infectious virus. We will increase our chances of contracting COVID-19 and spreading it by engaging in such behaviors. The initiative will advise pharmaceutical providers, insurance companies, and consultants based on indirect relationships and concrete actions to be taken in terms of planning and operation for the public

benefit by creating an ambitious course of action in terms of decision-making and support.

III. LITERATURE REVIEW

- While going through all the datasets available for covid 19 on various sources like kaggle,data.gov etc. we came across the Vaccine Adverse Effect Reporting System(VAERS) website. It is a forecasting system to detect safety problems in US-licensed vaccines.
- VAERS identifies and detects the adverse effects of any vaccination used in the country. VAERS is co-managed by Central Disease Control(CDC) and US Food and Drug Administration(FDA).
- Amidst pandemic situations around the world due to COVID-19 disease, the vaccine for covid-19 was developed in almost 1-1.5 years instead of the 10-15 years that usual vaccine development takes. After the development of covid-19 vaccine, some people were hesitant but most of the people started taking the vaccine around the world. Currently there are 3 vaccines in USA Pfizer/BioTech, Moderna and Johnson And Johnson and VAERS is collecting the data of the people who got side effects after getting these vaccinations.
- We realized that studying the negative effects of covid-19 vaccination would aid in understanding the efficacy of various vaccines.

IV. PROJECT DEVELOPMENT METHODOLOGY

- This project was created using the agile methodology of pair programming. It is a software development method.
- Using the pair programming has improved the time management for the coding and testing part of the project. Also establishing teamwork amongst the team was comparatively easy by using this approach. We can say that so far the project outcome has been affected very positively by using a pair programming approach alongside the development.
- To follow pair programming methods during pandemic we arranged google meet to arrange team's driver-observer meetups, used cloud technologies like AWS, google doc, google sheet to work together. Usually the driver will share the screen while coding and the observer will observe the code and catch up the error.

V. DATABASE STRUCTURE

The database consists of five tables namely **patient_info**, **patient_history**, **vaccine_details**, **symptoms**, **patient_outcomes**. Patient_info is the main table. The other tables are joined to the Patient_info.

A. *patient_info*

- **Patient_ID**: A serially allocated number that is used for recognition.
- **report_date**: The date that the VAERS form data was received at processing center.
- **age_yrs**: Age of vaccine recipient's recorded in years.
- **gender**: Vaccine recipient's sex.
- **vaccine_date**: Vaccinated date.

B. *patient_history*

- **Patient_ID**: A serially allocated number that is used for recognition.
- **other_meds**: Any prescribed or over-the-counter drugs that the vaccine patient was taking at the time of vaccination.
- **current_illness**: Any ailments at the time of vaccination.
- **med_history**: This attribute stores the previously diagnosed medical conditions.
- **prior_vacc**: This attribute includes details about previous vaccine events.
- **allergies**: Any allergies that were present at the time of vaccination.

C. *vaccine_details*

- **Patient_ID**: A serially allocated number that is used for recognition.
- **vaccine_site**: This attribute indicates location of the vaccination.
- **vaccine_route**: This attribute specifies the route of vaccine.
- **vaccine_name**: The brand name of the vaccine administered.
- **dose_series**: This attribute specifies the vaccine dose.

D. *symptoms*

- **Patient_ID**: A serially allocated number that is used for recognition.
- **symptoms_1**: Initial Symptom.
- **symptoms_2**: Second Symptom.
- **symptoms_3**: Third Symptom.
- **symptoms_4**: Fourth Symptom.
- **symptoms_5**: Fifth Symptom.

E. *patient_outcomes*

- **Patient_ID**: A serially allocated number that is used for recognition.
- **died**: "Yes" is used if the vaccine recipient died; otherwise, the letter "No" is used.
- **date_of_death**: If the vaccinated person expired, this field will just be blank to write down the date of death; otherwise, it will be filled in with the date of death.
- **life_threat**: A "Yes" is used if the vaccinated person experienced a harmful event as a result of the vaccination;

otherwise, a "No" is used.

- **hospitalized:** If the vaccinated person is hospitalized then "Yes" else "No"
- **hospital_days:** It records no of days the vaccinated person is hospitalized .
- **prolonged_stay:** If vaccinated person extends his stay in hospital then "Yes" otherwise "No".
- **disabilities:** If the vaccine person became disable because of vaccine enter a "Yes" here; otherwise, enter a "No".
- **recovered:** A "Yes" is entered in the field if the vaccinated person is recovered,otherwise "NO".

VI. ENTITY RELATIONSHIP DIAGRAM

An entity-relationship diagram depicts the relationships between entity sets contained in a database (ERD). In this way, an individual, or a data unit, is an entity. The term "entity package" refers to a collection of similar entities. Attributes are to specify properties for these entities. We can see from the ERD that Patient Info is a table with several relationships, including vaccine details, patient outcomes, symptoms, and patient history, as well as the establishment of a foreign key for the same.

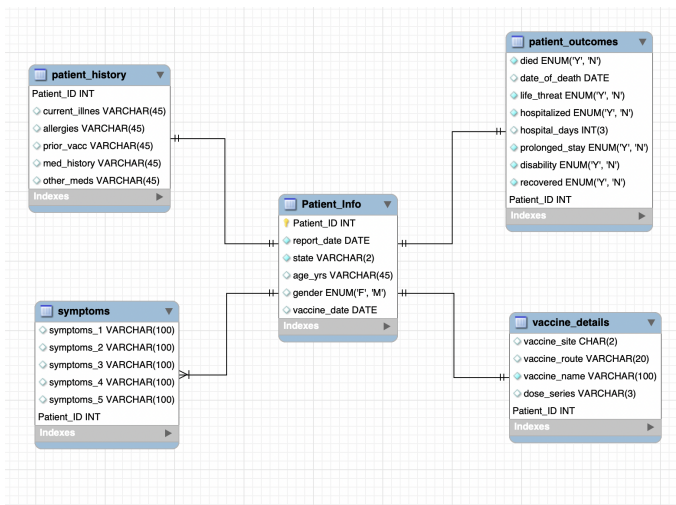


Fig. 1. Entity Relationship Diagram

VII. ANALYSIS

- Each table contains 58,172 records who faced adverse reactions after taking the vaccination.
- Out of them 42,109 are female and 13,938 are male and 2,125 are others.

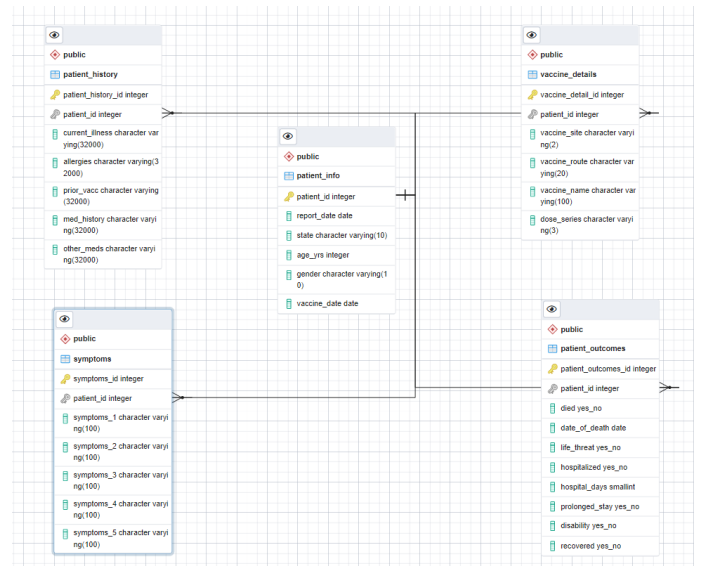


Fig. 2. Entity Relationship Diagram generated using Postgres

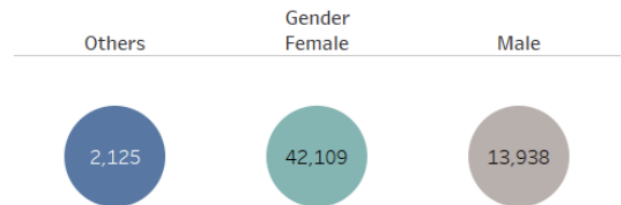


Fig. 3. Gender Distribution

- From the above analysis we could predict that in females the adverse reactions are more when compared to males and others.
- From the analysis we could see that many people got vaccinated to Moderna are facing more adverse reactions when compared to Pfizer and Janssen.
- From the analysis it is shown that Moderna has more effect on the people in their first dosage of the vaccination when compared to Pfizer. But in second dosage the effect of Pfizer is more when compared to Moderna.
- 4.603 percent people who faced reactions after taking Moderna are died.
- 4.65 percent people who faced reactions after taking Pfizer are died.
- 1.06 percent people who faced reactions after taking Janssen are died.
- Janssen's death rate was comparatively very low when compared to other two vaccines.
- Eventhough people effected rate of Moderna is high death rate of both
- People from California State has faced more adverse reactions in United States than any other state.

VIII. DESIGN STEPS

No	Task	Tool Used
1	Project Scope	Team meeting on Google meets, Brainstorming
2	Project Proposal	Google doc, Team meeting on Google Meet
3	Covid-19 Adverse Effect Dataset	Kaggle, VAERS
4	Data Model	MySQL Workbench, postgresSQL
5	Physical Schema	PostgreSQL using pgAdmin
6	Raw Data Storage	Excel, Amazon Simple Storage Services(S3)
7	Data Cleaning/ETL	Python Pandas, Excel, AWS Glue
8	Loading Data(CSV) on Tables	Using pgAdmin, Aws Glue Job
9	Data Analysis	postgresql(pgAdmin) , Python, Amazon redshift database, Jupyter Notebook, postgresql Connector
10	Data Visualization	Tableau, Python seaborn Library
11	Code integration and testing	postgresql, Python, Amazon Redshift
12	Project Report	Google Doc, Excel Sheet, Latex
13	Project PPT	Google Slides
14	Version Control	Github Repo

Fig. 4. Design Steps

IX. TECHNICAL DIFFICULTY

A. With Postgres

- Our dataset contains many features and we were facing difficulties in loading all of them in one table, we solved this by splitting the data into different tables and establishing relationships between them.
- We are using tableau for visualization so while connecting postgres to tableau we faced difficulty in making connections.

B. With AWS

- While making connections from s3 to glue to redshift we faced issues with endpoints, which was then resolved by creating VPC endpoints.

C. With Pandas

- File encoding issue with the dataset of 2021vaersdata , so we changed the encoding from csv to iso-8859-1 while reading pandas.
- And while loading the database we changed the encoding of the csv file manually using excel to utf8.

X. DATA VISUALIZATION

- Data visualization is a graph, diagram, or other graphic representation of data or information. It conveys relationships by combining details and images. This is important because it facilitates the detection of trends and patterns. When big data becomes more common, we must be able to access increasingly bigger batches of data.

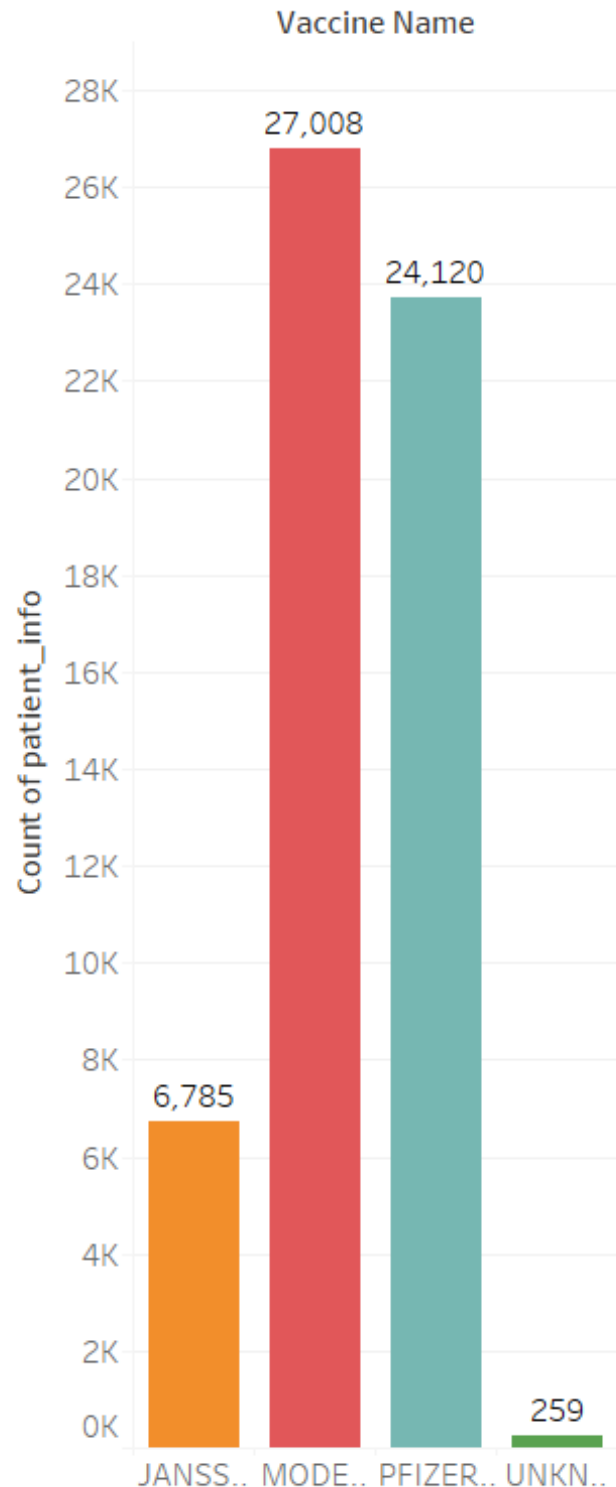


Fig. 5. Patient Count-Vaccine

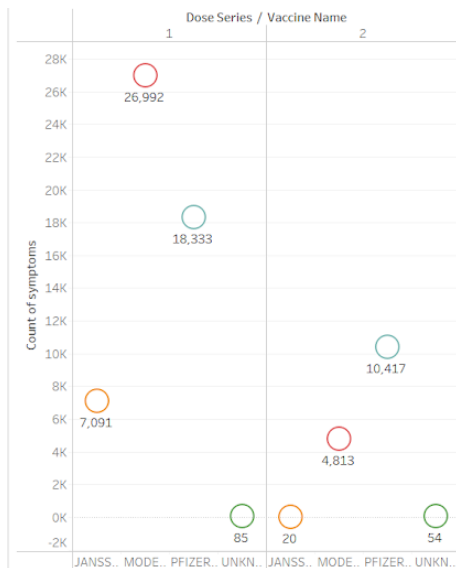


Fig. 6. Dosage effect on People

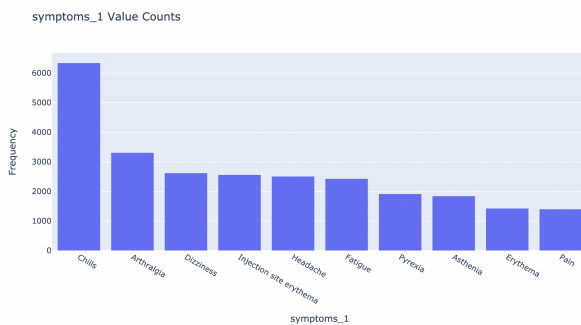


Fig. 7. Top-10 Symptoms

- Tableau is a popular data visualization platform in the business intelligence field that is rapidly increasing in popularity. It aids in the translation of raw data into an understandable format. Tableau makes it easier to create facts that researchers at all levels of an enterprise can understand.
- Seaborn is a Python data visualization library built on matplotlib. It comes with a high-level graphical user interface for producing visually pleasing and informative statistical graphics.

XI. AWS IMPLEMENTATION

In AWS, we loaded the data into s3 buckets and then created a VPC End point and from s3 we moved the data to crawler which is stored in Glue Catalog and then used to Glue job to load the data from S3 to Redshift and then used nosql data base Dynamodb for table creation of one table and then loaded the data to s3 using Lambda functions.

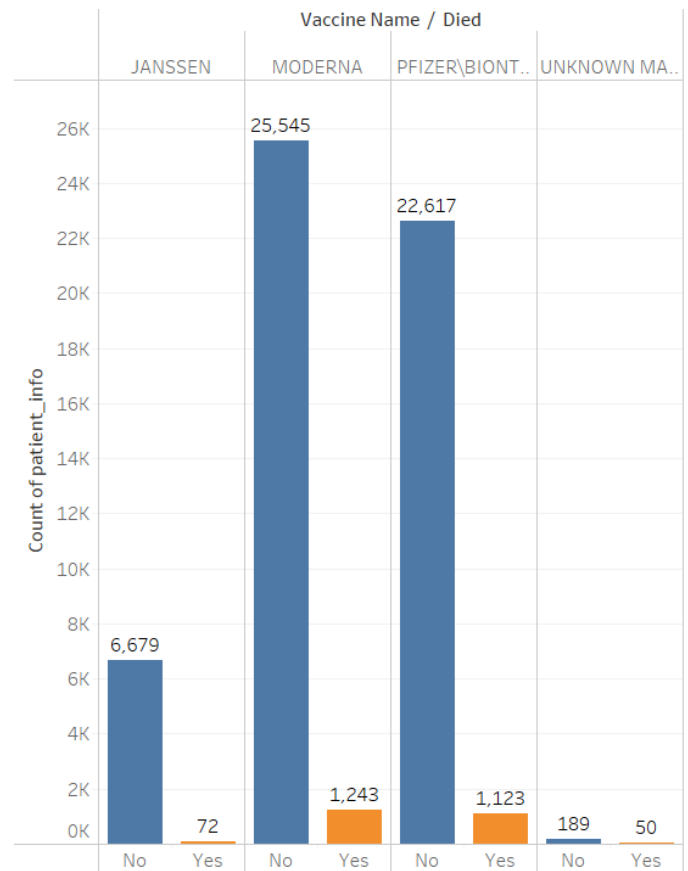


Fig. 8. Count of people who expired after taking vaccination based on vaccine name

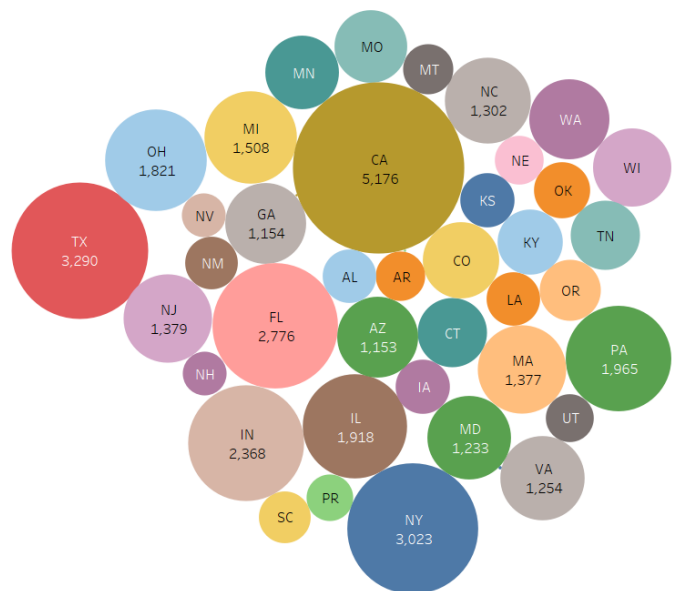


Fig. 9. State wide effected count

XII. LESSONS LEARNED

- Designing the tables (or Data Modelling) is the most important step and should be done in the early stage of the project.
- We have loaded data in the postgres database from csv files. Before loading the data the files should include cleaned data only otherwise it become tedious task to clean and load at every step after an error is thrown by database
- Data cleaning with an ETL tool like AWS glue is much easier than python pandas.

XIII. AWS IMPLEMENTATION

In AWS, we loaded the data into s3 buckets and then created a VPC End point and from s3 we moved the data to crawler which is stored in Glue Catalog and then used to Glue job to load the data from S3 to Redshift and then used nosql data base Dynamodb for table creation of one table and then loaded the data to s3 using Lambda functions.

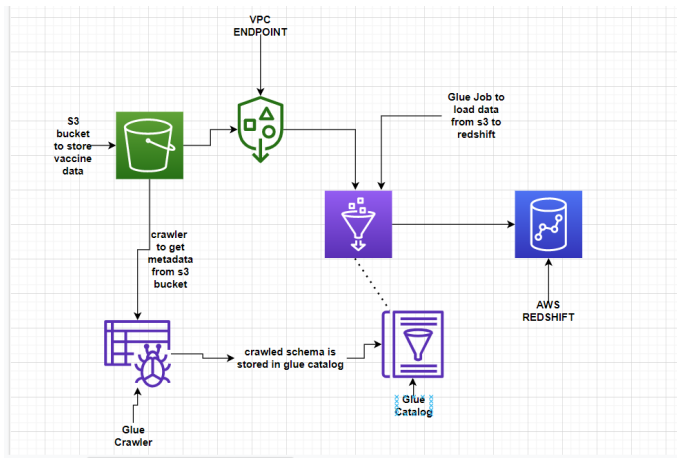


Fig. 10. AWS Implementation-I

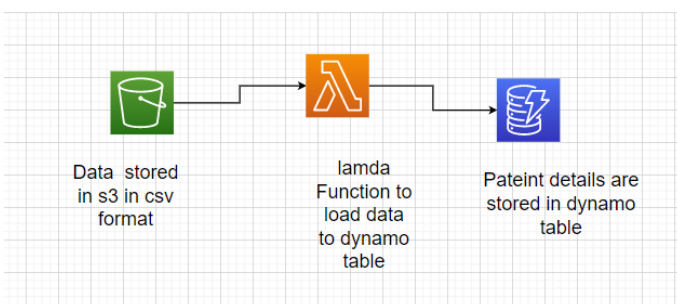


Fig. 11. AWS Implementation-II

XIV. ESSENTIAL OUTCOMES

- According to our analysis on the data that we have collected from VAERS, the maximum number of patients had taken Moderna.

- Maximum number of people belong to states like California, Texas and New York.
- There are very few patients who have died due to the severe side effects of covid vaccine. Total of 4.3
- Total 8.9 percentage people were hospitalized out of all the patients. Out of all patients, 5004 people were hospitalized from which 64 had the prolonged_stay which is over 20 days.
- According to our analysis on the data that we have collected from VAERS, the maximum number of patients had taken Moderna.
- Maximum number of people belong to states like California, Texas and New York.

XV. VERSION CONTROL

We uploaded our work to github and shared it with the public, which allows anyone to keep track of our progress and easily explore the changes we've made, whether it's data, coding scripts, notes, or anything else.

- Link - <https://github.com/nirvishagarara/Data225-CovidVaccineEffect-Project>

XVI. CONCLUSION

In conclusion, the analysis on covid-19 vaccines' side effects will help in preventing the future casualties also. If the disease-prevention effects of vaccines outweigh the dangers of a previously discovered side effect, the vaccine can still be recommended. The packaging insert for the vaccine, which lists safety details, is updated with information about recently discovered side effects.

REFERENCES

- [1] <https://vaers.hhs.gov/index.html>.
- [2] <https://www.healthline.com/health-news/heres-how-it-was-possible-to-develop-covid-19-vaccines-so-quicklyHard-work-and-luck-played-a-part>
- [3] <https://www.medrxiv.org/content/10.1101/2021.05.06.21256282v1>
- [4] <https://www.reuters.com/article/factcheck-vaers-deaths/fact-check-vaers-data-does-not-prove-thousands-died-from-receiving-covid-19-vaccines-idUSL1N2LV0NY>

XVII. TERM PROJECT RUBRIC

Youtube Video Link- <https://www.youtube.com/watch?v=36cjwWr18g>

