Adverse Reaction of Vaccination : Side Effects Clustering and Risk and Vaccine Efficiency Prediction

Capstone

Under the guidance: Prof. Wang, Chaojie

Sharath Srinivas
Chetan B Desai
Saideep Malgireddy

Team and their presentation part

Sharat Srinivas

- Introduction
- Background
- Goals

Chetan Desai

- Initial Dataset
- Data Definitions
- Methods

Saideep Reddy

- Research Resources
- GUI
- EDA

TABLE OF CONTENTS

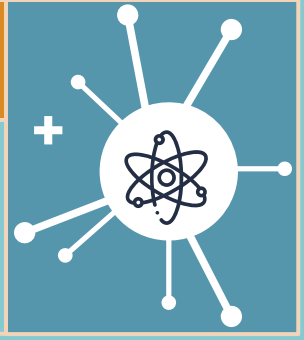
INTRODUCTION

- 543 million doses of COVID-19 vaccines were administered in the United States from December 14, 2020, through February 3, 2022
- During this time, VAERS received 12,122 preliminary reports of death (0.0022%) among people who received a COVID-19 vaccine
- It is important to analysis the data during this pandemic and be aware of the information regarding the vaccination.

INTRODUCTION

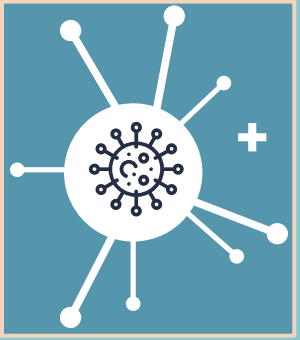
Drug Safety

Drug safety refers to the frequency of adverse drug effects that are treatment emergent. After regulatory approval of a drug, the ongoing process of post-market surveillance ensures continued safety of the product



Adverse Event

Harmful or negative outcome that occurs when a patient has been provided with medical care or treatment



INTRODUCTION

1. Product Approval
2. Product Administration
3. Adverse Event
4. Reporting
5. Signal Detection
6. Action Taken

BACKGROUND

VAERS

Vaccine Adverse Event Reaction System (VAERS)

FDA and CDC

Reports of vaccine-related adverse events

Formed

November 1, 1990 to the present

Monitoring

Monitor reports

Vaccination

Greater rate of incidents than expected

Serious Incidents

85% to 90% modest events

< 15% serious incidents

GOALS

Adverse Reaction

Can we Predict the Reactions after Immunization ? Is it Life threatening ?

Manufacturers

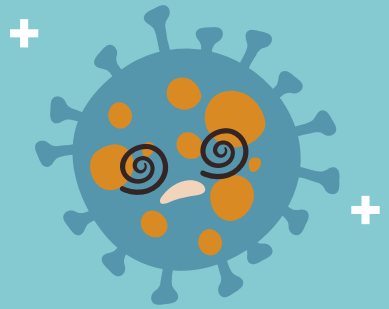
Which manufacture Vaccine will suits you based on various factors?

Adverse Event

Where more adverse events occurred after vaccinations?

Initial Dataset

Data sets



Data File containing Patient details.

VAERSDATA

Data Description

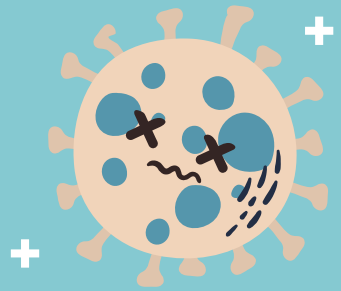
```
df_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1951645 entries, 0 to 1951644
Data columns (total 35 columns):
 #   Column          Dtype
---  ------          -----
 0   VAERS_ID        int64
 1   RECVDATE        object
 2   STATE           object
 3   AGE_YRS         float64
 4   CAGE_YR         float64
 5   CAGE_MO         float64
 6   SEX             object
 7   RPT_DATE        object
 8   SYMPTOM_TEXT    object
 9   DIED            object
 10  DATEDIED        object
 11  L_THREAT        object
 12  ER_VISIT        object
 13  HOSPITAL        object
 14  HOSPDAYS        float64
 15  X_STAY          object
 16  DISABLE         object
 17  RECOVD          object
 18  VAX_DATE        object
 19  ONSET_DATE      object
 20  NUMDAYS         float64
 21  LAB_DATA        object
 22  V_ADMINBY       object
 23  V_FUNDBY        object
 24  OTHER_MEDS      object
 25  CUR_ILL         object
 26  HISTORY         object
 27  PRIOR_VAX       object
 28  SPLTTYPE        object
 29  FORM_VERS       int64
 30  TODAYS_DATE     object
 31  BIRTH_DEFECT    object
 32  OFC_VISIT       object
 33  ER_ED_VISIT     object
 34  ALLERGIES       object
dtypes: float64(5), int64(2), object(28)
```

- Observations:1951645
- Features: 35

Initial Dataset

Data sets



Data provide the vaccine information

VAERSVAX

Data Description

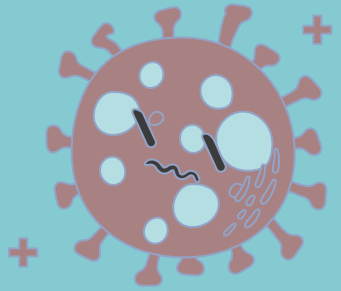
```
df_vax.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2420759 entries, 0 to 2420758
Data columns (total 8 columns):
 #   Column           Dtype 
---  ------           ----- 
 0   VAERS_ID         int64 
 1   VAX_TYPE         object
 2   VAX_MANU         object
 3   VAX_LOT          object
 4   VAX_DOSE_SERIES  object
 5   VAX_ROUTE        object
 6   VAX_SITE         object
 7   VAX_NAME         object
dtypes: int64(1), object(7)
memory usage: 147.8+ MB
```

- Observations:2420758
- Features: 08

Initial Dataset

Data sets



Data provide the adverse event coded terms utilizing the MedDRA dictionary

VAERSSYMTOMS

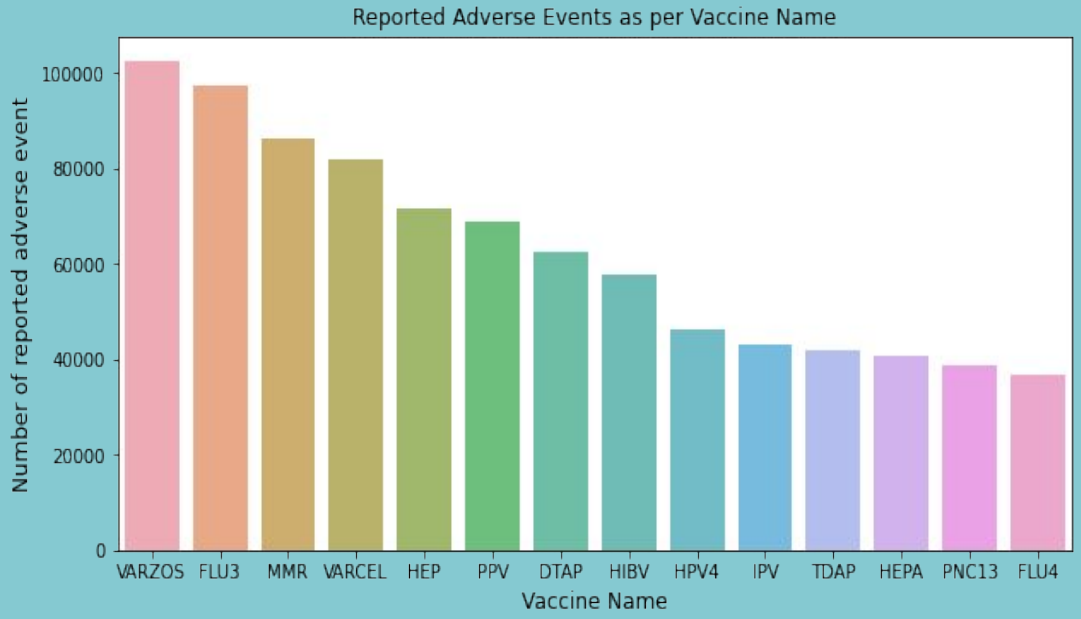
Data Description

```
df_symp.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2596826 entries, 0 to 2596825
Data columns (total 11 columns):
 #   Column            Dtype
---  ------            -----
 0   VAERS_ID          int64
 1   SYMPTOM1          object
 2   SYMPTOMVERSION1   float64
 3   SYMPTOM2          object
 4   SYMPTOMVERSION2   float64
 5   SYMPTOM3          object
 6   SYMPTOMVERSION3   float64
 7   SYMPTOM4          object
 8   SYMPTOMVERSION4   float64
 9   SYMPTOM5          object
 10  SYMPTOMVERSION5   float64
dtypes: float64(5), int64(1), object(5)
memory usage: 217.9+ MB
```

- Observations: 2596835
- Features: 11

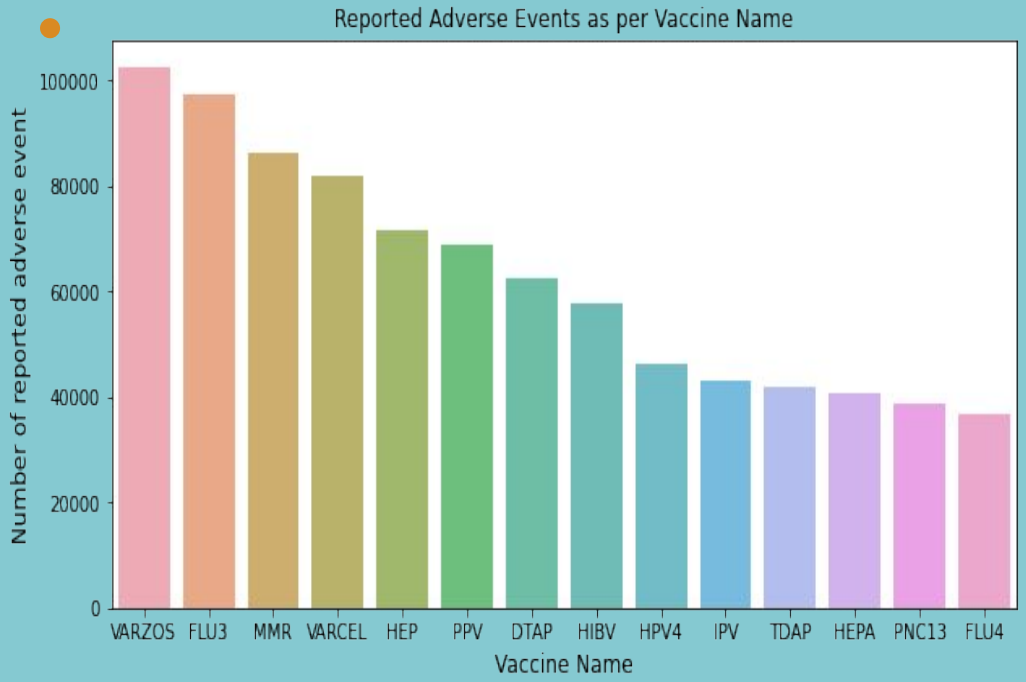
Vaccine?



FLU (Influenza) : Vaccine Protects against FLU and from its potential serious complications

HEPATITIS B : Vaccine that prevents inflammation of liver

Vaccine?



COVID 19 : Vaccine that Acute respiratory illness in humans caused by a coronavirus, capable of producing severe symptoms

VARICELLA-ZOSTER VACCINE (VARZOS) : Vaccine that reduces the incidence of herpes zoster (shingles), a disease caused by reactivation of the varicella zoster virus (VZV), which is also responsible for chickenpox.

In the above bar chart we can observe the number of reported cases for all vaccine. In those list we selected above 4 vaccine. FYI COVID is not listed above because it contains more cases for the visualization purpose we have not included

Data Definition

51 Columns

2596826 Rows

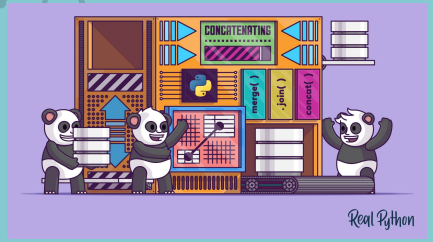
COVID 19  1088560

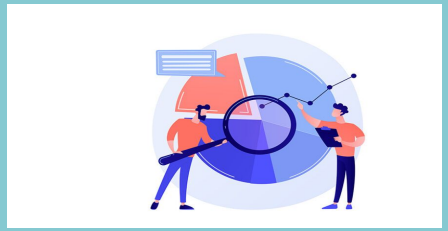
VARZOS  101037

FLU (Influenza)  169699

HEPATITIS B  113692

DATA CLEANING





- The First Step Includes concatenation of all the data frame into three main data frame df_vax, df_data, df_symp using glob and concat functions

- Performing Sanity Check to determining the unique count and the null counts of data

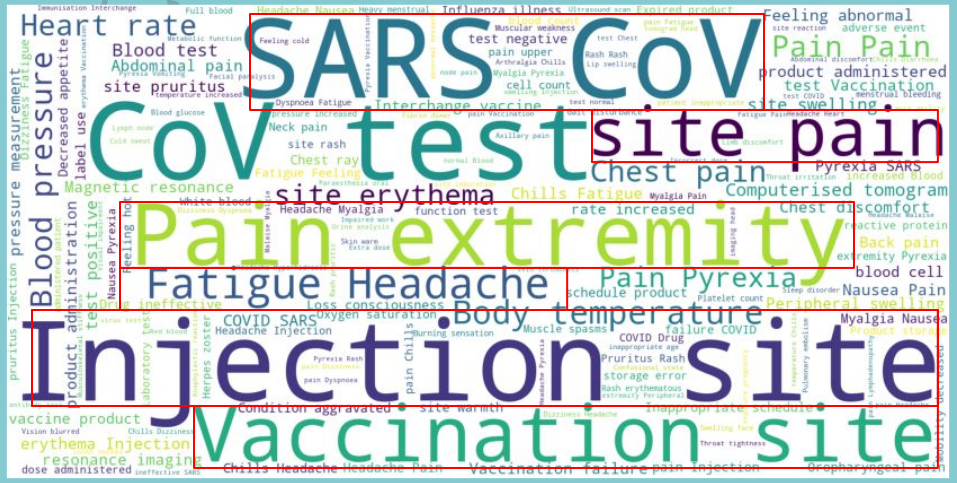
- Filtering Vaccines Data by name

DATA CLEANING

- Removal of Duplicates And Imputing Boolean or Removal of columns
- Converting Categorical values with binary parameters.
- Cleaning of Text Columns with Regex

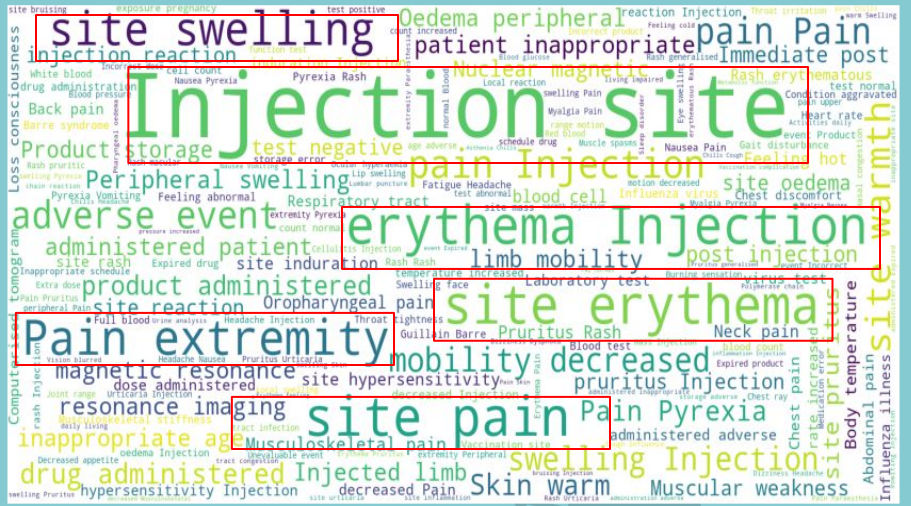
Word Cloud of Symptoms By vaccine

01 : Covid



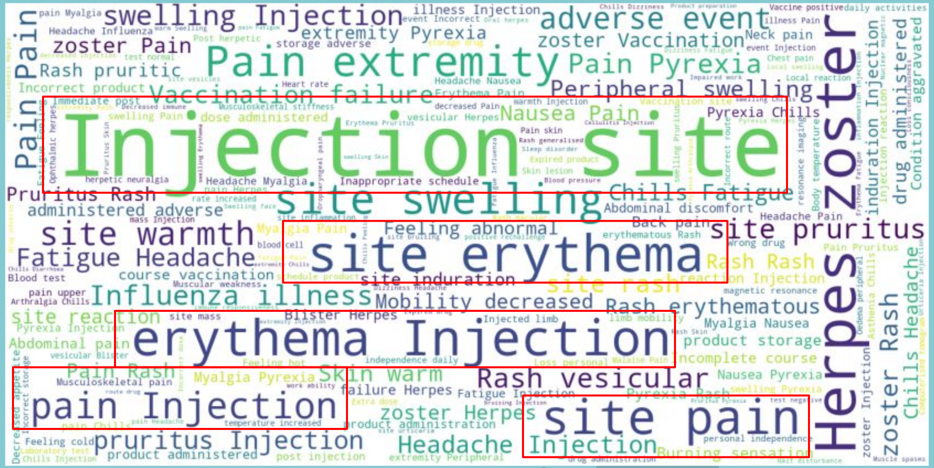
Injection-site reaction is considered to be any pain, swelling, rash, bleeding, or redness that occurs at the site of an injection and we can observe that these adverse event are common for any vaccination

02 : FLU Influenza



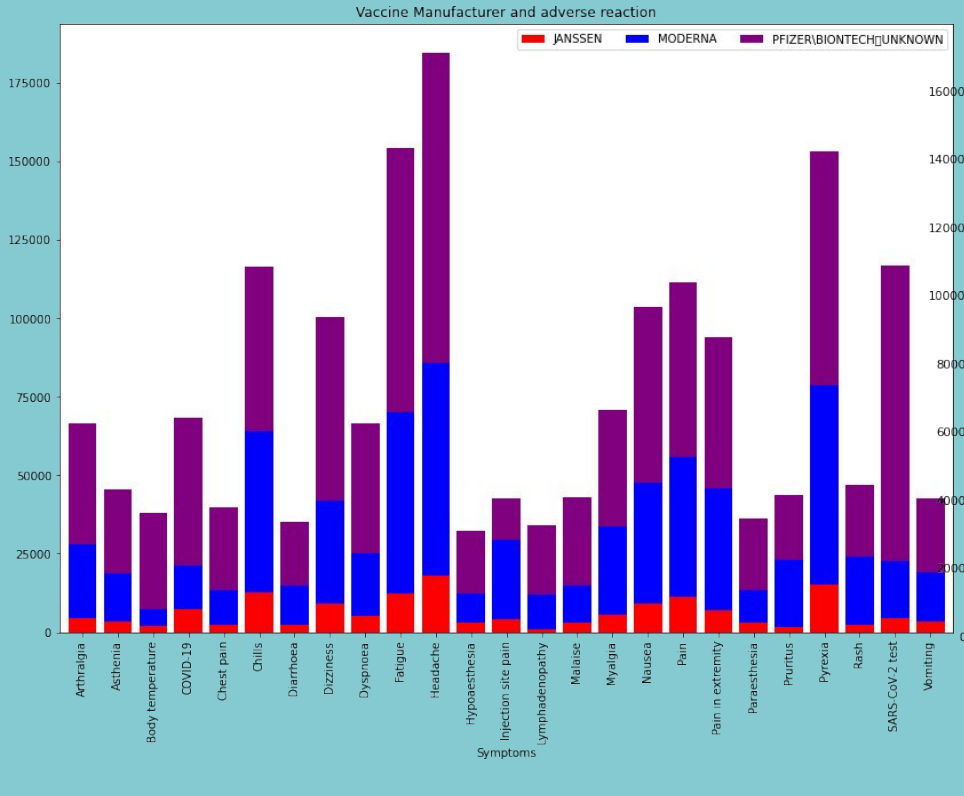
03 : Hepatities

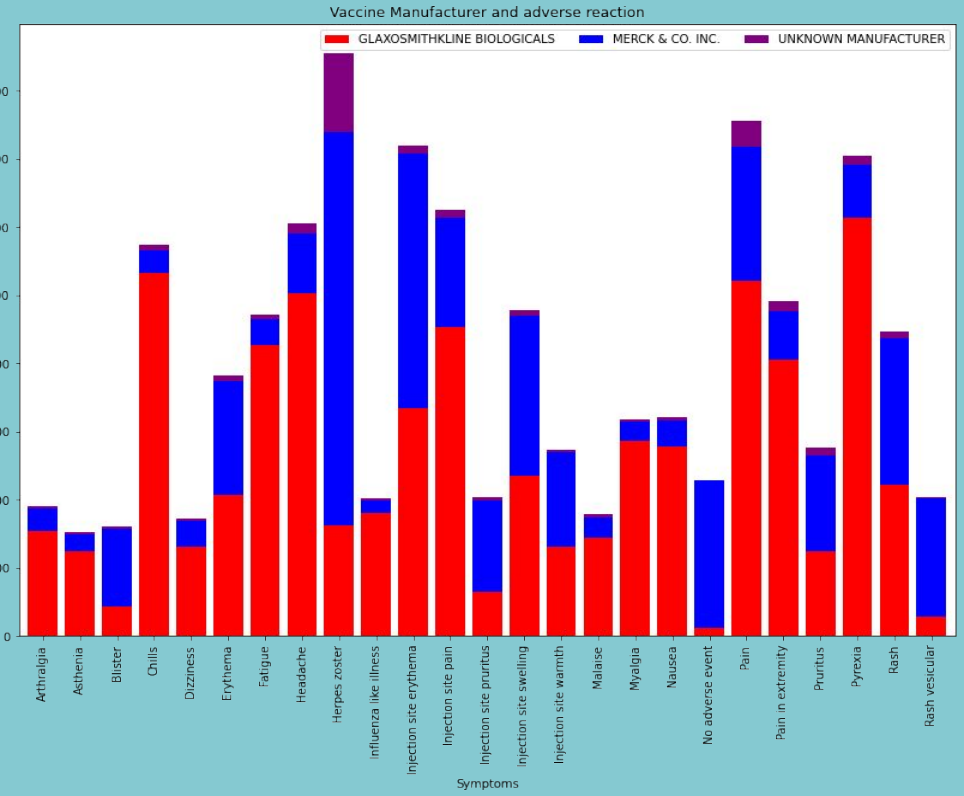
04 : Varzos



Injection-site, Site erythema, erythema Injection, pain injection, site pain and some adverse reactions are common for all 4 vaccine so we can ignore those reactions in the modelling

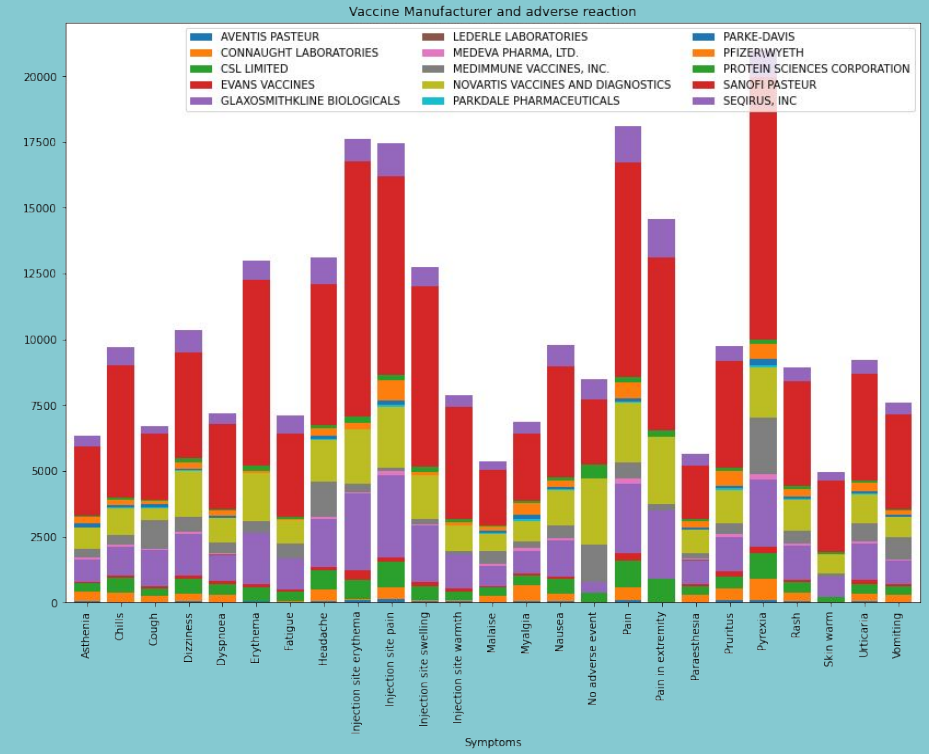
Vaccine Manufacturer and Adverse Reactions

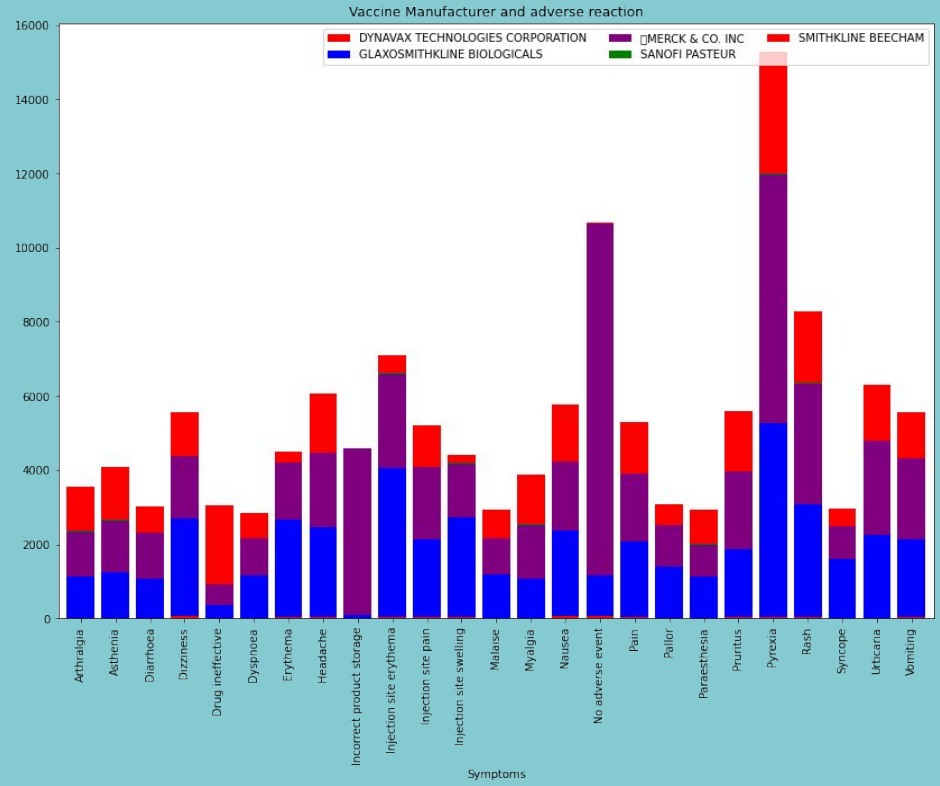




From the above graphs we can see the counts of adverse reactions with respect to the manufacturers.

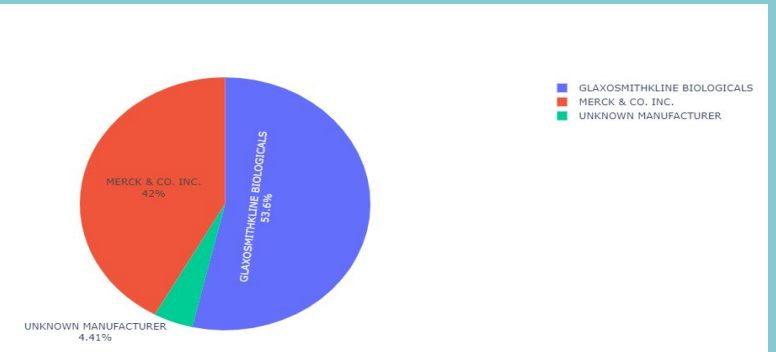
Vaccine Manufacturer and Adverse Reactions



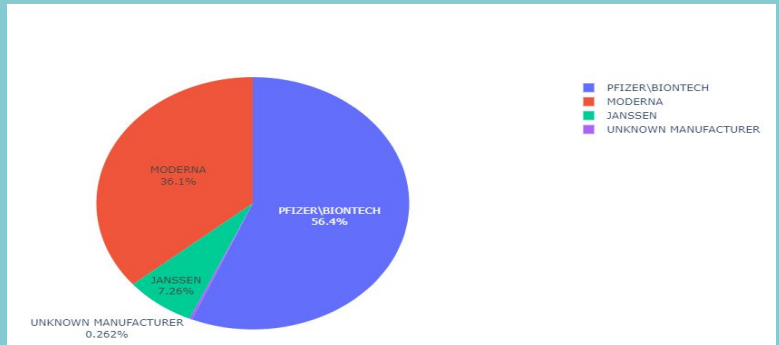


As we observed in worlcloud and above chart most of the symptoms reported are Fatigue, Headache ,Nausea,Pain, Injection site Chills Dizziness and no adverse event so for our further analysis we will be excluding common adverse reactions

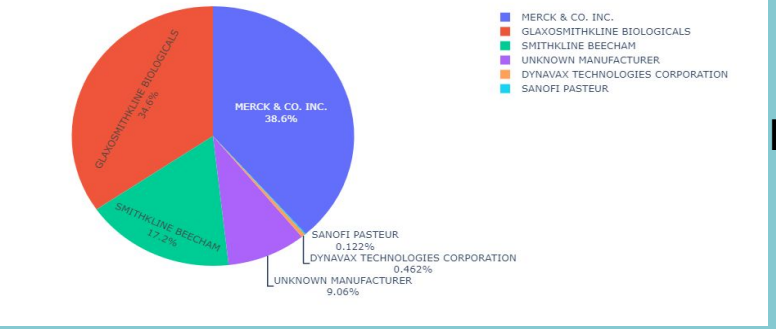
Vaccine Manufacturer ratio



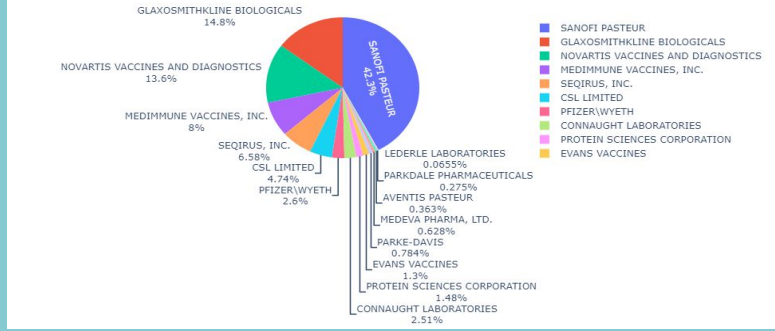
VARZOS



COVID 19



HEPATITIS

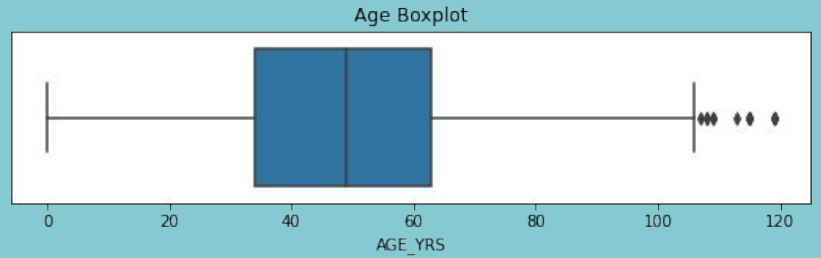


FLU

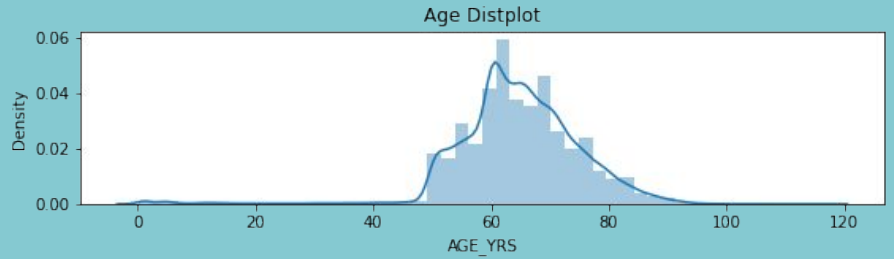
There are lot of adverse reactions where vaccine manufacturers details are unknown so for our further analysis we have removed unknown manufacturers details

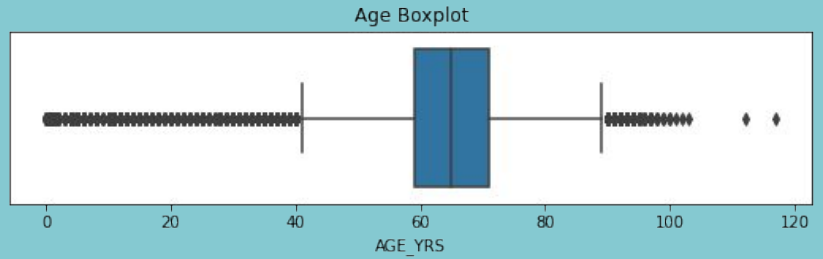
Age distribution plots for different vaccine

COVID 19



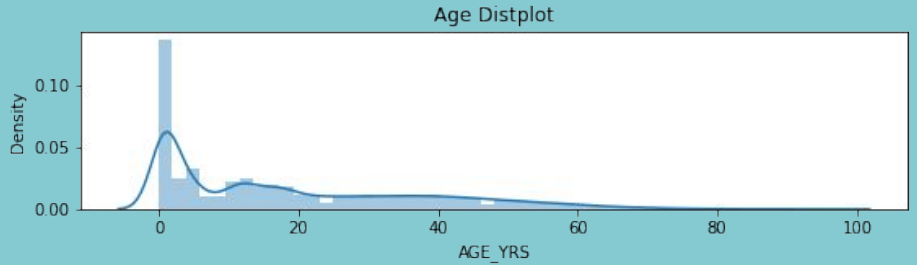
Varzos

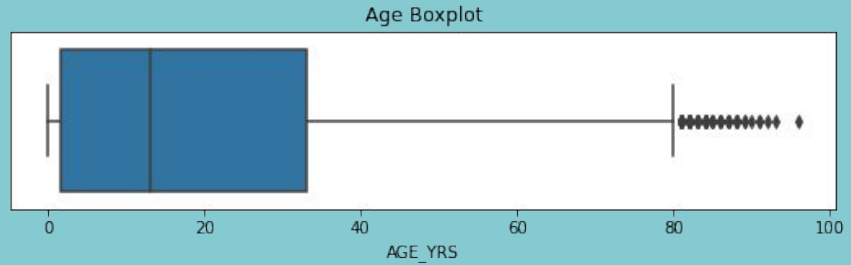




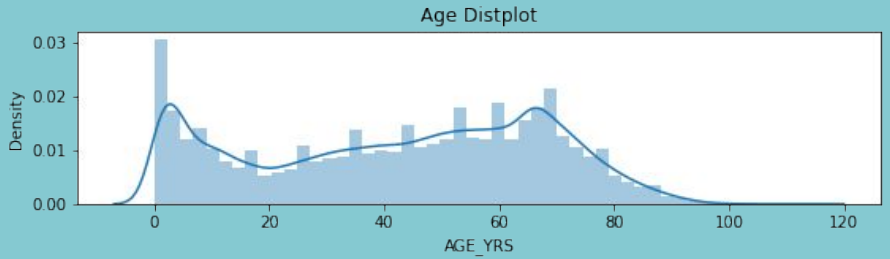
Age distribution plots for different vaccine

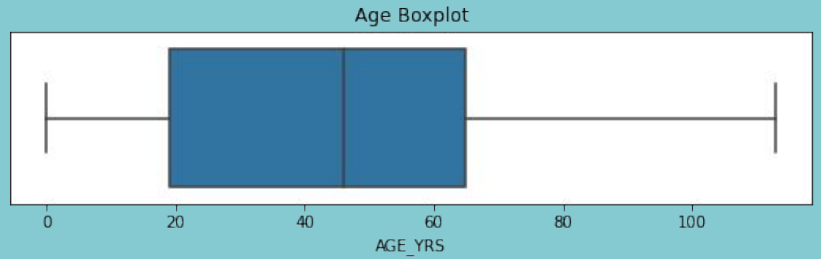
HEPATITIS





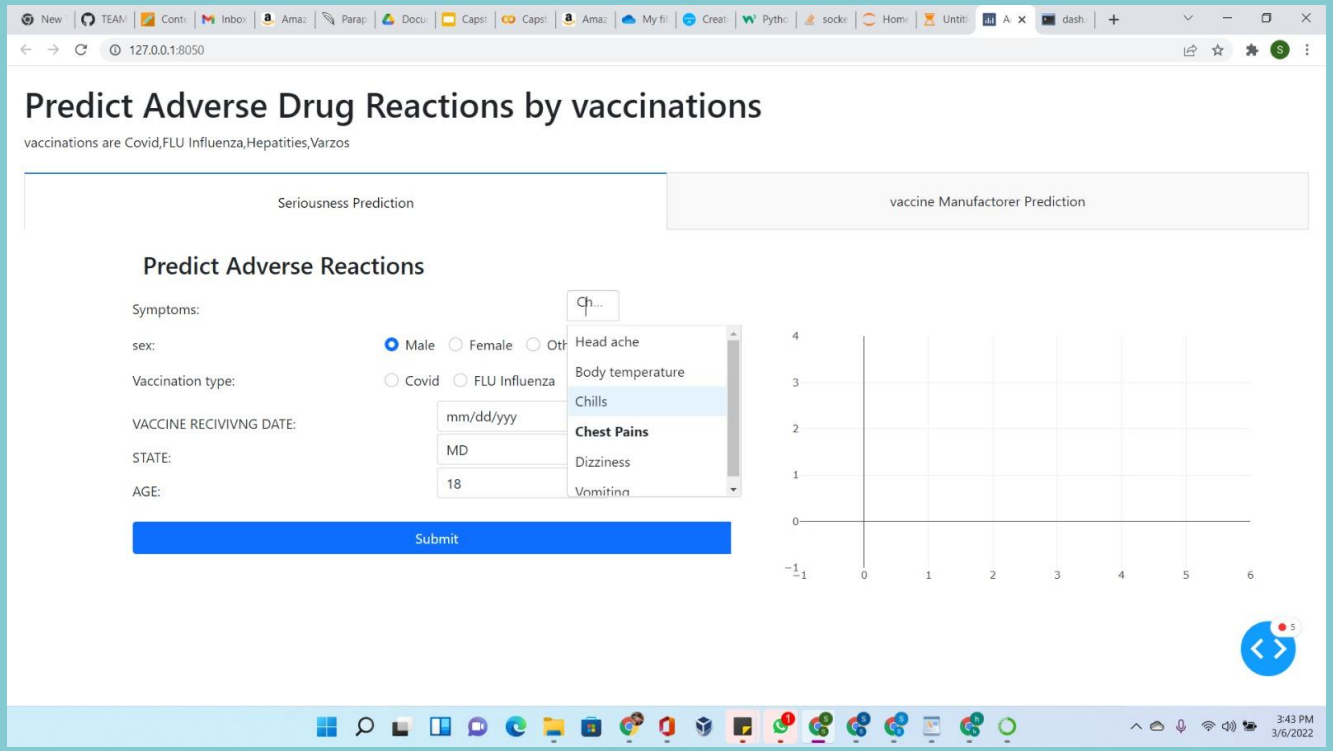
Flu





There are so many outliers in AGE as its one of the features used in modeling we will be removing the Outlier which are present

Graphical User Interface



Python Dash was used to create a User interface.

Observation and Future Modeling

- Data have more Sparsity
- We need to create a target variable serious and non Serious cases
- One Hot encoding symptoms, Allergies,Medical history and other medication
- Modeling with SMOTE and without SMOTE

References:

[1]. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/adverse-events.html

[2]. https://pubmed.ncbi.nlm.nih.gov/15071280/

[3].https://www.cdc.gov

[4].https://stackoverflow.com/questions/45787782/combine-multiple-columns-in-pandas-excluding-nans.

[5].https://stackoverflow.com/questions/17679089/pandas-dataframe-groupby-two-columns-and-get-counts

THANKS