/\*All analysis is explained in the final report in detail. Hence, we are only highlighting the main steps in comments here.\*/

/\*setting the library\*/

libname d "M:\datasets";

/\*Importing the 2021VAERSDATA.csv file\*/

**proc** **import** datafile = "M:\datasets\2021VAERSDATA.csv"

out=vdata

dbms=csv

replace;

**run**;

/\*Importing the 2021VAERSVAX.csv file\*/

**proc** **import** datafile = "M:\datasets\2021VAERSVAX.csv"

out=vvax

dbms=csv

replace;

**run**;

/\*Question 1\*/

/\*Duplicating vdata and creating new variable 'period' to extract the year from vaccination date\*/

**data** temp;

set vdata;

period = year(vax\_date);

**run**;

/\*joining both datasets to filter for only COVID-19 vaccines. \*/

/\*And looking up numdays(number of days between vaccination date and adverse symptoms onset date )and vaccination date\*/

**proc** **sql**;

create table temp2 as

select a.numdays, a.vax\_date

from temp a, Vvax b

where a.vaers\_id=b.vaers\_id and b.vax\_type= "COVID19"

order by a.numdays desc;

**quit**;

/\*looking at the summary of numdays\*/

**proc** **means**

data = temp2;

var numdays;

**run**;

/\*joining both datasets to filter for only COVID-19 vaccines and 2021 vaccination dates, to further study numdays.\*/

**proc** **sql**;

create table q1 as

select a.numdays

from temp a, Vvax b

where a.vaers\_id=b.vaers\_id and b.vax\_type= "COVID19" and period=**2021**

order by a.numdays desc;

**quit**;

/\*once again looking at the summary of numdays\*/

**proc** **means**

data = q1;

var numdays;

**run**;

/\*Part of Question 1 was completed in python and the resulting graphs and analysis are mentioned in the final report \*/

/\*Question 2 and Question 3 were completed in python and the resulting graphs and analysis are mentioned in the final report \*/

/\*Question 4\*/

/\*We wanted to know the relationship of other factors with the magnitude of adverse effects

so we chose "sex", "type of facilities", and "age" for our first 3 explorations.

/\*Exploration 1\*/

/\*First,we analyze sex and want to see the whether there is a difference between female and male.\*/

/\*joining datasets and grouping by sex and vax\_lot to view sum of adverse events by sex and vax\_lot\*/

**proc** **sql**;

create table a4\_1 as

select b.vax\_lot, a.SEX, count(a.vaers\_id) as total\_events, sum(a.numdays)

from Vdata a, Vvax b

where a.vaers\_id=b.vaers\_id and b.vax\_type= "COVID19" and b.vax\_lot ne "Unknown"

group by SEX, b.vax\_lot

having b.vax\_lot ne ""

order by total\_events desc;

**quit**;

/\*We made a graph to show the difference\*/

**proc** **sgplot** data=a4\_1;

scatter x=SEX y=total\_events;

title Comparison of Female and Male;

/\*Exploration 2\*/

/\*Then we analyze the type of facilities.\*/

/\*joining datasets and grouping by v\_adminby(Type of facility where vaccine was administered)

to view sum of adverse events by type of facility\*/

**proc** **sql**;

create table a4\_2 as

select a.V\_ADMINBY, count(a.vaers\_id) as total\_events, avg(a.numdays)

from Vdata a, Vvax b

where a.vaers\_id=b.vaers\_id and b.vax\_type= "COVID19"

group by V\_ADMINBY

order by total\_events desc;

**quit**;

/\*Then we made a graph of it\*/

**proc** **sgplot** data=a4\_2;

scatter x=V\_ADMINBY y=total\_events;

title Comparison of type of facility;

/\*Exploration 3\*/

/\*analysing age\*/

/\*joining datasets and grouping by cage\_yr(Calculated age of patient in years)

to view avg of adverse events by patient age\*/

**proc** **sql**;

create table t as

select a.CAGE\_YR,count(a.vaers\_id) as total\_events, count(a.AGE\_YRS) as Num\_age

from Vdata a, Vvax b

where a.vaers\_id=b.vaers\_id and b.vax\_type= "COVID19"

group by CAGE\_YR

order by total\_events desc;

**quit**;

/\*We need to calculate the average of the adverse events\*/

**proc** **sql**;

create table a4\_3 as

select \* from (select CAGE\_YR, total\_events,Num\_age,(total\_events/Num\_age)as avg\_events from t);

**quit**;

/\*The plot\*/

**proc** **sgplot** data=a4\_3;

scatter x=CAGE\_YR y=avg\_events;

title Comparison of Age;

/\*Exploration 4\*/

/\*Then we want to know which type of COVID19 vaccination is popular in all states and FL\*/

/\*To see in all states\*/

**proc** **sql**;

create table a4\_4all as

select count(a.STATE)as count\_STATE, b.VAX\_MANU

from Vdata a, Vvax b

where a.vaers\_id=b.vaers\_id and b.vax\_type= "COVID19" and STATE ne "XB" and STATE ne "AS" and STATE ne "MP"and STATE ne "MH"and STATE ne "FM"and STATE ne "QM"and STATE ne "XL"and STATE ne "XV"

group by VAX\_MANU ;

**quit**;

/\*To see in FL\*/

**proc** **sql**;

create table a4\_4fl as

select count(a.STATE)as count\_STATE, b.VAX\_MANU

from Vdata a, Vvax b

where a.vaers\_id=b.vaers\_id and b.vax\_type= "COVID19" and STATE = "FL"

group by VAX\_MANU ;

**quit**;

/\*The plot for the the vaccinations in FL\*/

**proc** **sgplot** data=a4\_4fl;

step x=VAX\_MANU y=count\_STATE;

title How popular of each type of accination in FL;