Covid-19 Vaccine Reaction Analysis

Analyzing & evaluating risk factors using adverse events based on age & gender to predict life-threatening risk probability of receiving a COVID-19 vaccine.

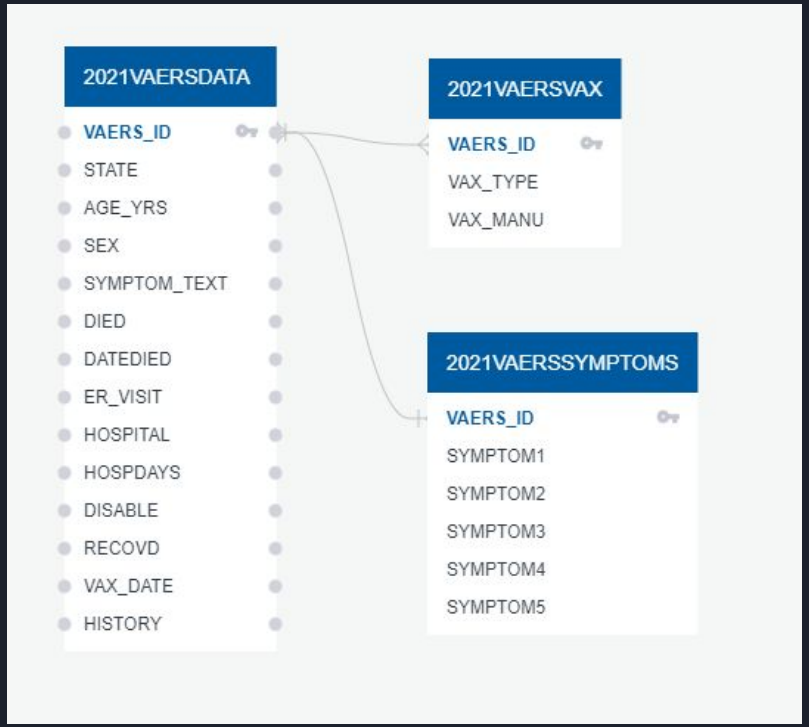
BACKGROUND & TOPIC SELECTION

- US Citizens are concerned about the risk factors of taking the COVID-19 vaccines. We are playing the role of a team of data scientists hired by the government to analyze and assess the risk factors of receiving one of the three Emergency-Use Authorized COVID-19 vaccines.
- What is VAERS?
- How VAERS Reporting System help CDC & FDA?
- Why did we chose this vaccine analysis?
- The outcome of the analysis is to help citizens make a more informed decision when taking the vaccine. We will use vaccine adverse event data provided by the government from the Vaccine Adverse Event Reporting System. We will analyze and assess risk factors of taking the COVID19 vaccine.

TECHNOLOGIES AND RESOURCES

- PostgresSQL
- Python
- Pandas Library
- Scikit-Learn Library - Machine Learning
- Tableau Public

DATA STRUCTURE & USER GUIDE



Header	Type	VAERS 2 Form	VAERS 1 Form	Description of Contents
VAERS_ID	Num(6)	✔	✔	VAERS Identification Number
RECVDATE	Date	✔	✔	Date report was received
STATE	Char(2)	Derived	Box 1	State
AGE_YRS	Num(xxx.x)	Item 6	Box 4	Age in Years
CAGE_YR	Num(xxx)	Derived	Derived	Calculated age of patient in years
CAGE_MO	Num(.x or 1)	Derived	Derived	Calculated age of patient in months
SEX	Char(1)	Item 3	Box 5	Sex
RPT_DATE	Date	Discontinued	Box 6	Date Form Completed
SYMPTOM_TEXT	Char(32,000)	Item 18	Box 7	Reported symptom text
DIED	Char(1)	Item 21	Box 8	Died
DATEDIED	Date	Item 21	Box 8	Date of Death
L_THREAT	Char(1)	Item 21	Box 8	Life-Threatening Illness
ER_VISIT	Char(1)	Discontinued	Box 8	Emergency Room or Doctor Visit
HOSPITAL	Char(1)	Item 21	Box 8	Hospitalized
HOSPDAYS	Num(3)	Item 21	Box 8	Number of days Hospitalized
X_STAY	Char(1)	Item 21	Box 8	Prolongation of Existing Hospitalization
DISABLE	Char(1)	Item 21	Box 8	Disability
RECOVD	Char(1)	Item 20	Box 9	Recovered
VAX_DATE	Date	Item 4	Box 10	Vaccination Date
ONSET_DATE	Date	Item 5	Box 11	Adverse Event Onset Date
NUMDAYS	Num(5)	Derived	Derived	Number of days (Onset date - Vax. Date)

DATA STRUCTURE & USER GUIDE

Header	Type	Description of Contents
VAERS_ID	Num(6)	VAERS Identification Number
VAX_TYPE	Char(15)	Administered Vaccine Type
VAX_MANU	Char(40)	Vaccine Manufacturer
VAX_LOT	Char(15)	Manufacturer's Vaccine Lot
VAX_DOSE_SERIES	Char (3)	Number of doses administered
VAX_ROUTE	Char(6)	Vaccination Route
VAX_SITE	Char(6)	Vaccination Site
VAX_NAME	Char(100)	Vaccination Name

Header	Type	Description of Contents
VAERS_ID	Num(6)	VAERS Identification Number
SYMPTOM1	Char(100)	Adverse Event MedDRA Term 1
SYMPTOMVERSION1	Num(XX.XX)	MedDRA dictionary version number 1
SYMPTOM2	Char(100)	Adverse Event MedDRA Term 1
SYMPTOMVERSION2	Num(XX.XX)	MedDRA dictionary version number 2
SYMPTOM3	Char(100)	Adverse Event MedDRA Term 3
SYMPTOMVERSION3	Num(XX.XX)	MedDRA dictionary version number 3
SYMPTOM4	Char(100)	Adverse Event MedDRA Term 4
SYMPTOMVERSION4	Num(XX.XX)	MedDRA dictionary version number 4
SYMPTOM5	Char(100)	Adverse Event MedDRA Term 5
SYMPTOMVERSION5	Num(XX.XX)	MedDRA dictionary version number 5

TABLEAU COVID-19 VAERS ANALYSIS

- Created a page to display the most frequent symptoms that were reported with life threatening risks.
- Total number of adverse events by vaccines.
- Most frequent symptoms reported by gender
- Most frequent hospitalizations reported by gender
- Total number of adverse events for ER visits, hospitalizations, life threatening events.

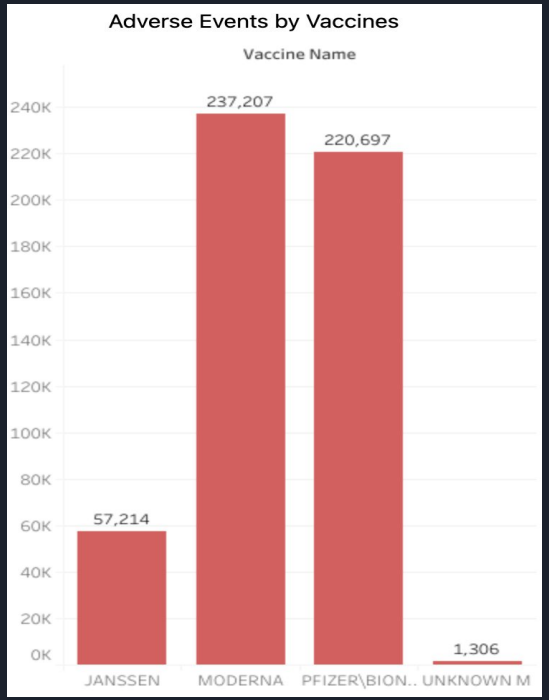
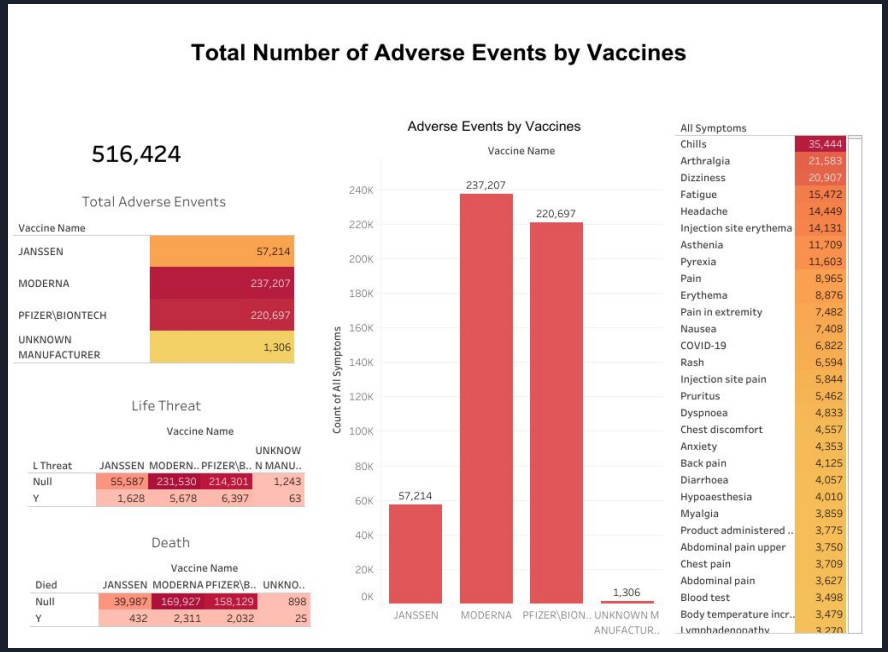


TABLEAU COVID-19 VAERS ANALYSIS

- Created a page to display the most frequent symptoms that were reported with life threatening risks.
- Total number of adverse events by vaccines.
- Most frequent symptoms reported by gender
- Most frequent hospitalizations reported by gender
- Total number of adverse events for ER visits, hospitalizations, life threatening events.



Can we predict the probability of a person over the age of 60 receiving a life threatening symptom?

From analyzing our data, we found a spike of life threatening symptoms in older patients. Using python's pandas and scikit-learn libraries, we can seek out the answer to our prediction.

CLEANING DATA USING PYTHON

For the 2021 VAERS Data file:

- Dropped the unnecessary columns.
- Replaced null values several columns to "N" to be fed into the machine learning.
- Null values on the age column were replaced with the median age for the provided gender.
- We created bins for ages to group them for the machine learning.

VAERS_ID	AGE_YRS	SEX	DIED	L_THREAT	ER_VISIT	HOSPITAL	DISABLE	BIRTH_DEFECT	AGE_BIN
916600	33.0	F	N	N	N	N	N	N	19-40
916601	73.0	F	N	N	N	N	N	N	60+
916602	23.0	F	N	N	N	N	N	N	19-40
916603	58.0	F	N	N	N	N	N	N	41-59
916604	47.0	F	N	N	N	N	N	N	41-59

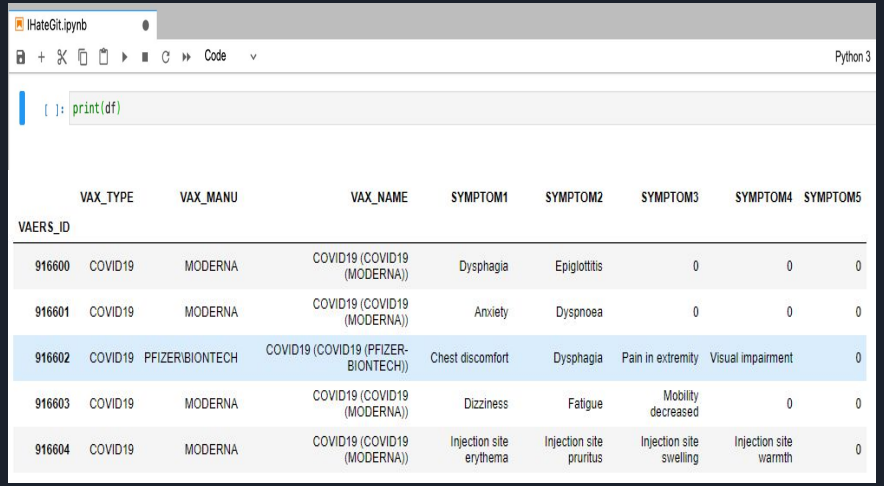
CLEANING DATA USING PYTHON

For the 2021 VAERS VAX file:

- Kept the columns with vaccine name and type in order to set as an unique ID and filter for COVID-19 adverse events only.

For the 2021 VAERS SYMPTOMS file:

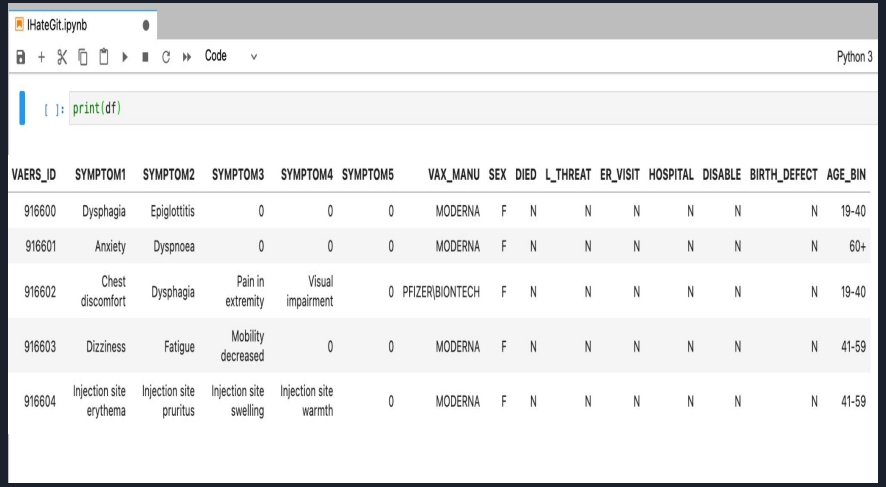
- We grouped the values of each of the symptom columns to remove the multiple rows for those IDs that had more than 5 symptoms.



VAERS_ID	VAX_TYPE	VAX_MANU	VAX_NAME	SYMPTOM1	SYMPTOM2	SYMPTOM3	SYMPTOM4	SYMPTOM5
916600	COVID19	MODERNA	COVID19 (COVID19 (MODERNA))	Dysphagia	Epiglottitis	0	0	0
916601	COVID19	MODERNA	COVID19 (COVID19 (MODERNA))	Anxiety	Dyspnoea	0	0	0
916602	COVID19	PFIZER\BIONTECH	COVID19 (COVID19 (PFIZER-BIONTECH))	Chest discomfort	Dysphagia	Pain in extremity	Visual impairment	0
916603	COVID19	MODERNA	COVID19 (COVID19 (MODERNA))	Dizziness	Fatigue	Mobility decreased	0	0
916604	COVID19	MODERNA	COVID19 (COVID19 (MODERNA))	Injection site erythema	Injection site pruritus	Injection site swelling	Injection site warmth	0

CLEANING DATA USING PYTHON

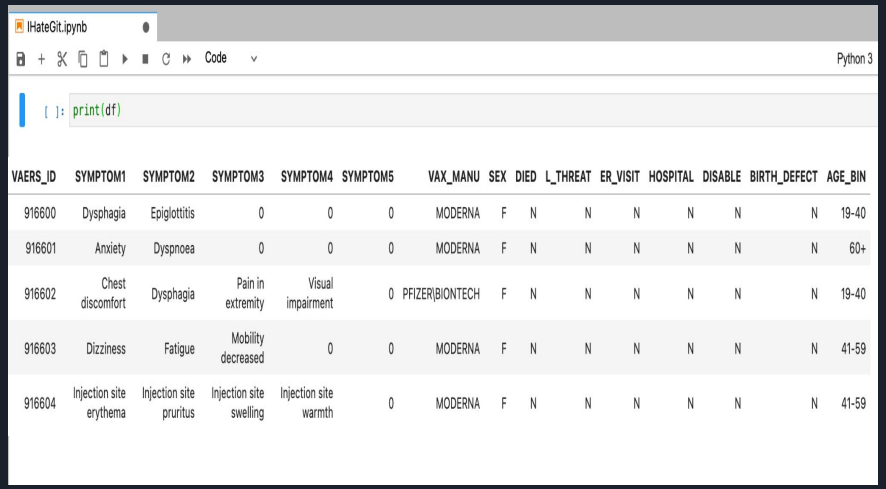
- We merged the Vax dataframe with the Symptoms dataframe.
- Filtered for COVID-19 vaccine type only.
- Finally, the Data dataframe was merged with the Vax and Symptoms dataframe and it was exported to be used on the machine learning model.



VAERS_ID	SYMPTOM1	SYMPTOM2	SYMPTOM3	SYMPTOM4	SYMPTOM5	VAX_MANU	SEX	DIED	L_THREAT	ER_VISIT	HOSPITAL	DISABLE	BIRTH_DEFECT	AGE_BIN
916600	Dysphagia	Epiglottitis	0	0	0	MODERNA	F	N	N	N	N	N	N	19-40
916601	Anxiety	Dyspnoea	0	0	0	MODERNA	F	N	N	N	N	N	N	60+
916602	Chest discomfort	Dysphagia	Pain in extremity	Visual impairment	0	PFIZER\BIONTECH	F	N	N	N	N	N	N	19-40
916603	Dizziness	Fatigue	Mobility decreased	0	0	MODERNA	F	N	N	N	N	N	N	41-59
916604	Injection site erythema	Injection site pruritus	Injection site swelling	Injection site warmth	0	MODERNA	F	N	N	N	N	N	N	41-59

CLEANING DATA USING PYTHON

- To find which symptoms had the most life threatening, we had to filter down the dataset by hospitalization, life threatening, and death
- We took the remaining symptoms from all 5 columns and listed out the most frequent symptoms.
- Finally, we have created a dataset that contains all ID's with only the most life threatening symptoms.
- The data frame is now ready for our machine learning model

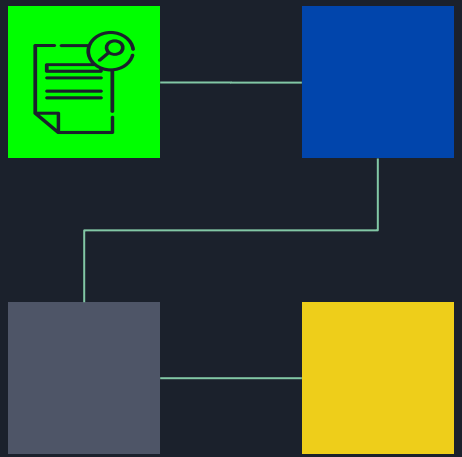


VAERS_ID	SYMPTOM1	SYMPTOM2	SYMPTOM3	SYMPTOM4	SYMPTOM5	VAX_MANU	SEX	DIED	L_THREAT	ER_VISIT	HOSPITAL	DISABLE	BIRTH_DEFECT	AGE_BIN
916600	Dysphagia	Epiglottitis	0	0	0	MODERNA	F	N	N	N	N	N	N	19-40
916601	Anxiety	Dyspnoea	0	0	0	MODERNA	F	N	N	N	N	N	N	60+
916602	Chest discomfort	Dysphagia	Pain in extremity	Visual impairment	0	PFIZER\BIONTECH	F	N	N	N	N	N	N	19-40
916603	Dizziness	Fatigue	Mobility decreased	0	0	MODERNA	F	N	N	N	N	N	N	41-59
916604	Injection site erythema	Injection site pruritus	Injection site swelling	Injection site warmth	0	MODERNA	F	N	N	N	N	N	N	41-59

MACHINE LEARNING IN PYTHON

IMPORT

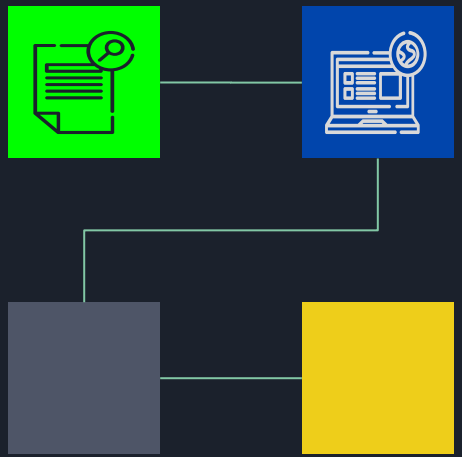
We imported the cleaned data file for our machine learning



MACHINE LEARNING IN PYTHON

IMPORT

We imported the cleaned data file for our machine learning



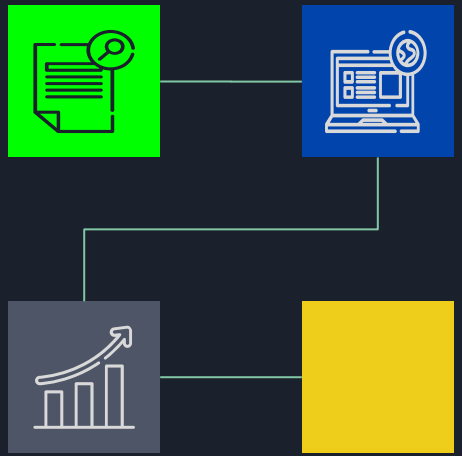
CONVERT

Next, we converted the symptom list into unique numbers and the rest of the categorical features into a binary column

MACHINE LEARNING IN PYTHON

IMPORT

We imported the cleaned data file for our machine learning.



CONVERT

Next, we converted the symptom list into unique numbers and the rest of the categorical features into a binary column.

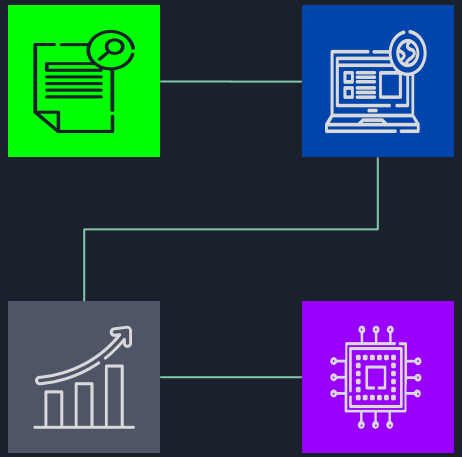
TRAIN AND TEST

The dataframe needed to be split into random train and test subsets, then scaled to unit variance.

MACHINE LEARNING IN PYTHON

IMPORT

We imported the cleaned data file for our machine learning.



CONVERT

Next, we converted the symptom list into unique numbers and the rest of the categorical features into a binary column.

TRAIN AND TEST

The dataframe needed to be split into random train and test subsets, then scaled to unit variance.

PREDICT

Using RandomForestClassifier, we were able to achieve a predictive accuracy with controlled over-fitting.

72%

Chance to get a symptom that was present during a life threatening adverse event if you are over the age of 60

Conclusion

Despite the heightened chance to get a life threatening symptom over the age of 60, our initial data found the death rate to be low. Even for the deaths reported, we lack the current resources to take pre-existing conditions into account. VAERS reports new data everyday, leaving more potential for unique questions to be asked, and more precise predictions to be answered.