

ADVERSE REACTION OF VACCINATION : RISK PREDICTION AND SIDE EFFECTS CLUSTERING

Under the guidance: **Prof. Wang, Chaojie**

Presented By:
Sharath Srinivas
Chetan B Desai
Saideep Malgireddy



Problem Statement

An adverse event (AE) is a poor or damaging outcome that occurs after a patient receives medical care or therapy.

The First goal of this Project is to develop an User Interface **to differentiate between serious and non-serious adverse events (AE)** by implementing the **Machine learning and Natural Language Processing (NLP)** model as a preliminary screening tool for all incoming AE reports, to get an initial seriousness classification.

Secondly, creating an interface for the Patients to check the **co-related Symptoms** for the particular vaccine and based on patients other criteria and reported symptoms.



Do you Know!

1 million 'COVID-vaccine injuries' are reported in a CDC database

Most side effects from vaccination are mild and Serious side effects are rare, but may include seizure or life-threatening allergic reaction.

Table of Contents

01

Introduction

02

**EDA and
Preprocessing**

03

**Modelling &
Model Evaluation**

04

**Conclusions and
Future Work**



01

Introduction

Background
Dataset
Data Cleaning



Why Vaccine Adverse Event is Important?

Timely adverse event reporting is essential for signal event detection to minimize further patients receiving unsafe vaccines.

The most notable event occurred in 2010 when increased incidence of fever and seizures occurred following seasonal influenza vaccines

The vaccine was withdrawn 10 days later, Incident including a prolonged febrile seizure resulting in profound disability in a previously healthy 11-month-old child



Background



VAERS (FDA & CDC)

Reports of
vaccine-related
adverse events



Vaccination & Serious Incidents

85% to 90% modest events <
15% serious incidents

Initial Dataset

- Data Downloaded from The Vaccine Adverse Event Reporting System (VAERS)
- The Timeline of the Data Collected is from 1990 to 2022 (up to February)
- Data is stored in 3 different CSV
 - Patient details.
 - Vaccination Information
 - Symptoms Information
- Total have 52 feature in this data set. 45 Non Text Columns and 7 Text Columns

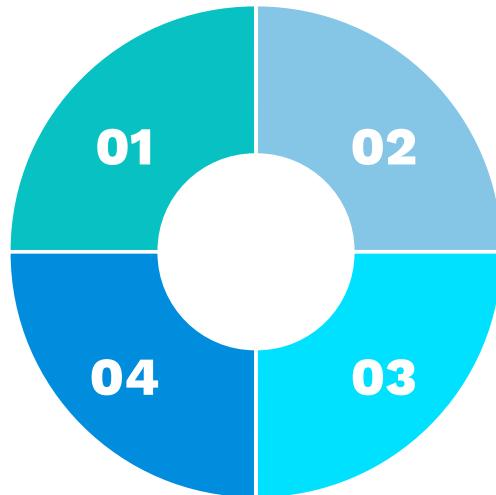
Data Cleaning

Filtering
Vaccines
Data

Cleaning
Text with
Regex

Removal of
Duplicates

Label
Encoding &
Removal of
Null Values



- Converted all Y/N columns to 1 or 0
- Selected Patients only between 18 to 100
- Removed Following Symptoms
 - Wrong drug administered
 - Incorrect route of drug administration
 - Incorrect dose administered
 - Product administered at inappropriate site
 - Incomplete course of vaccination
 - Incorrect route of product administration
 - Inappropriate schedule of drug administration
 - Inappropriate schedule of product administration



Target Variable



**Non
Serious**

Common Events
Like Fever,
cough, Pain etc



Serious

More Serious falls
into one of the
Serious criteria

Serious Criteria

1. Death

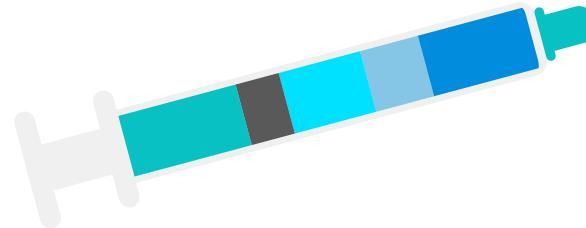
2. Life Threatening

3. Hospitalization

4. Disability or Permanent Damage

5. Congenital Anomaly or Birth Defect

6. Other Serious Important Medical Terminology



IME List

MedDRA Code	PT Name	SOC Name
10086470	Erythroid dysplasia	Blood and lymphatic system disorders
10087400	Paraneoplastic eosinophilia	Blood and lymphatic system disorders
10086663	Sarcoidosis of lymph node	Blood and lymphatic system disorders
10087091	Warm autoimmune haemolytic anaemia	Blood and lymphatic system disorders
10087237	Atrial standstill	Cardiac disorders
10087106	Chronic myocarditis	Cardiac disorders
10086740	Fascicular block	Cardiac disorders
10028650	Myopericarditis	Cardiac disorders
10086997	Pacing induced cardiomyopathy	Cardiac disorders

Target Variable Creation

Columns Representing Serious
criteria in data set



Serious



IME List



Columns
Representing
Symptoms



02

Exploratory Data Analysis

Pre-processing of Data
Data Visualization



We divided the EDA phase for all the Four Vaccines Separately



COVID19



varzos



HEPATITIS

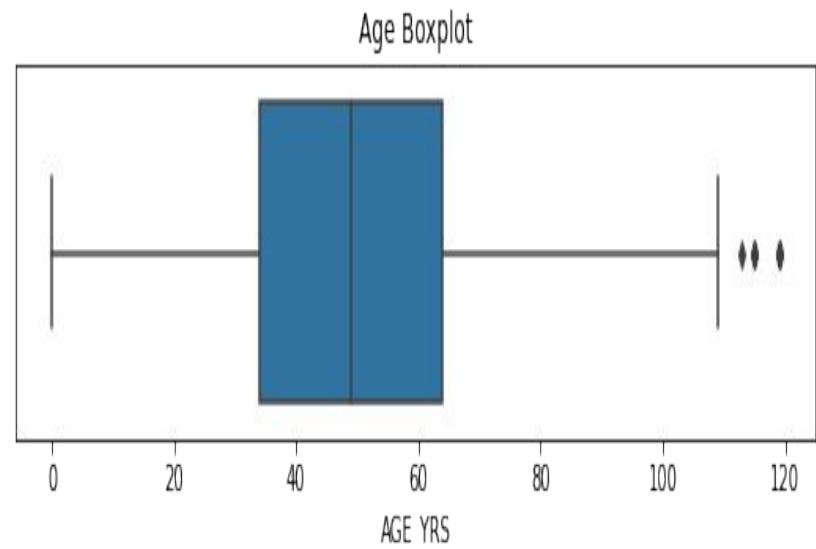
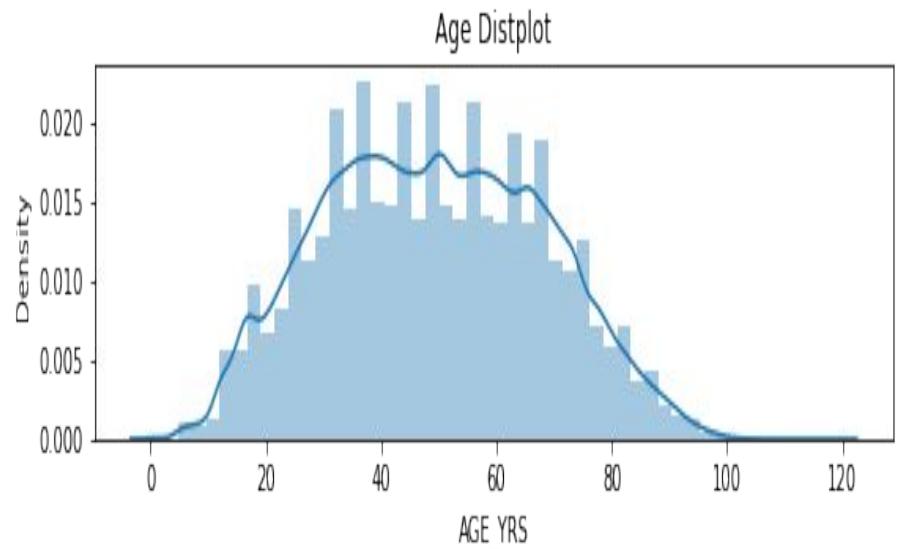


Flu

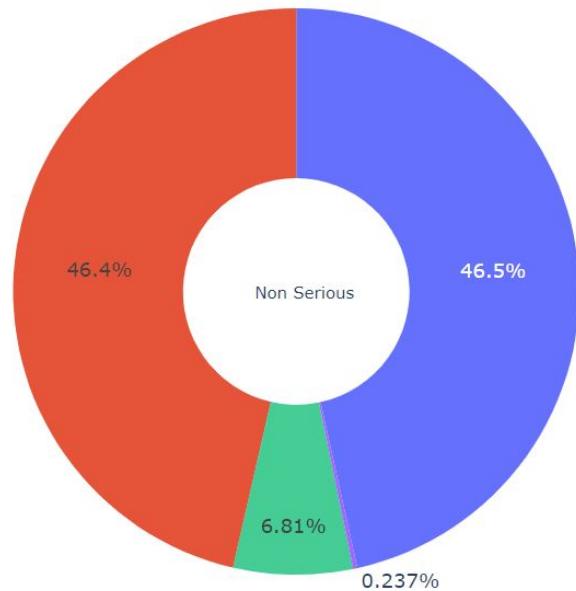
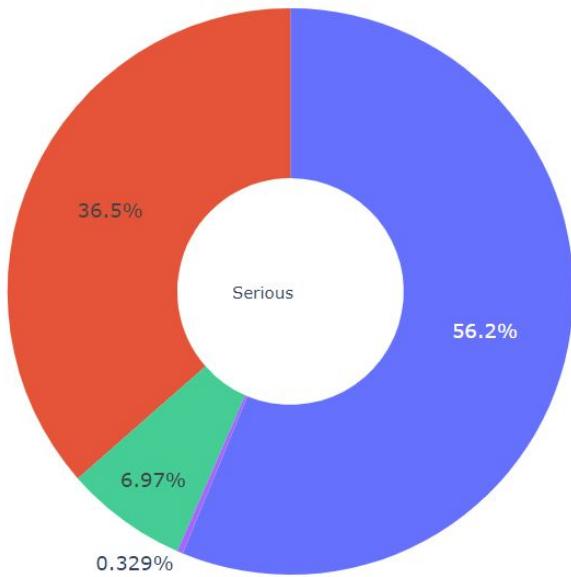
Covid 19



Plot describes age distributions and the box plot to identify the outliers in them, we considered the age group of vaccines to be between 18 to 100,

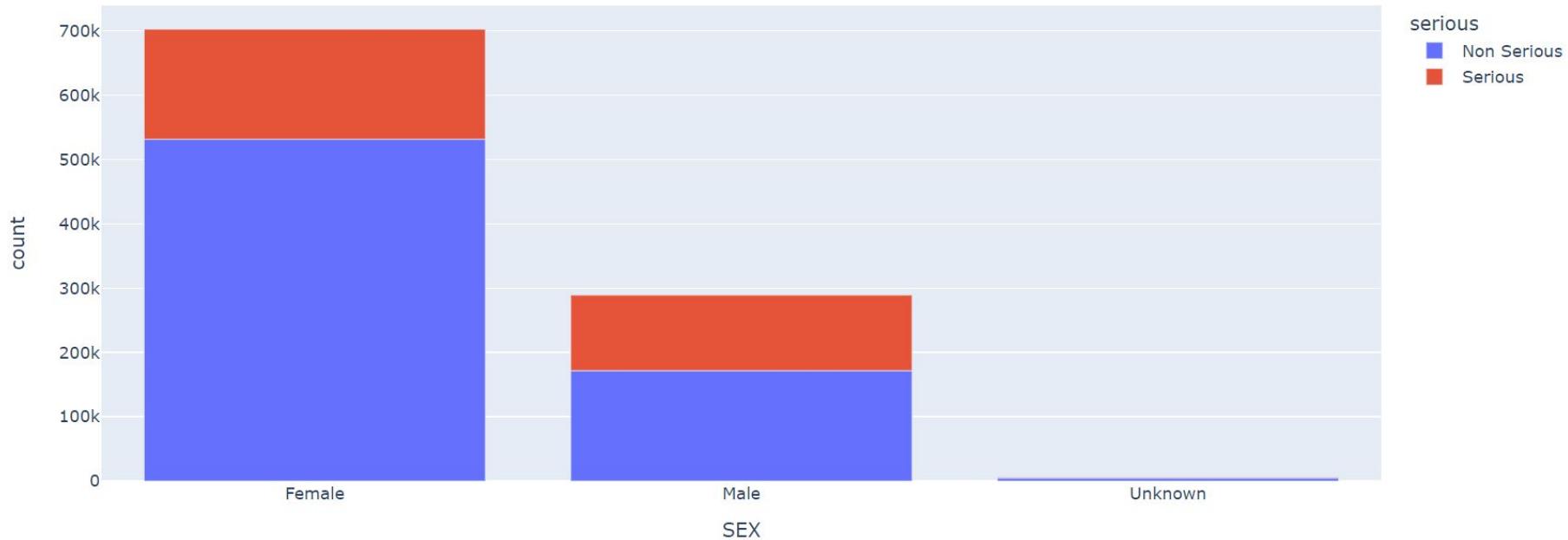


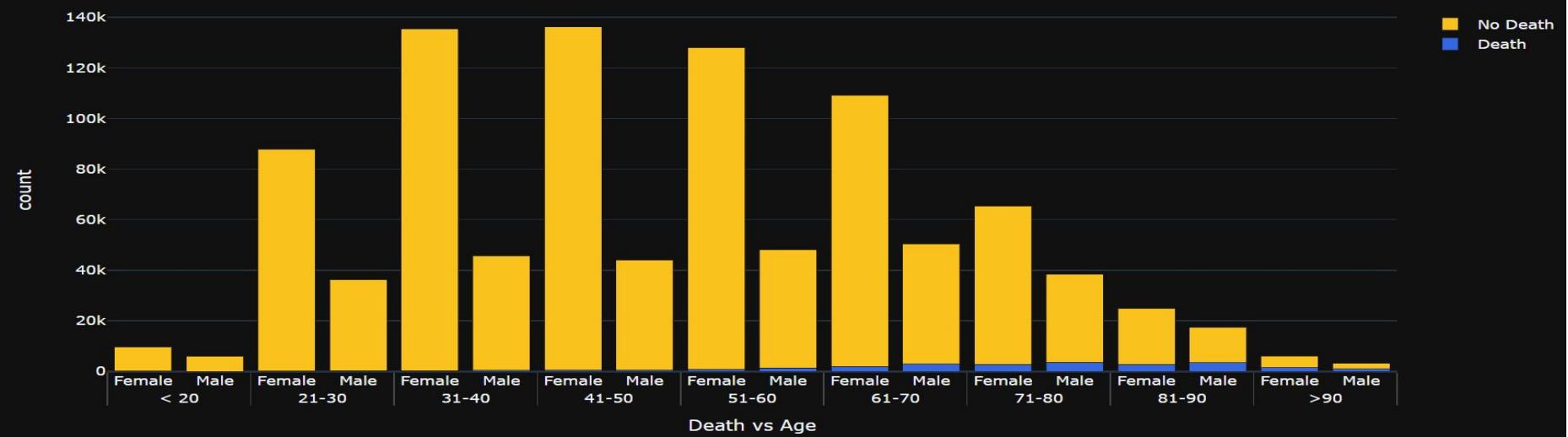
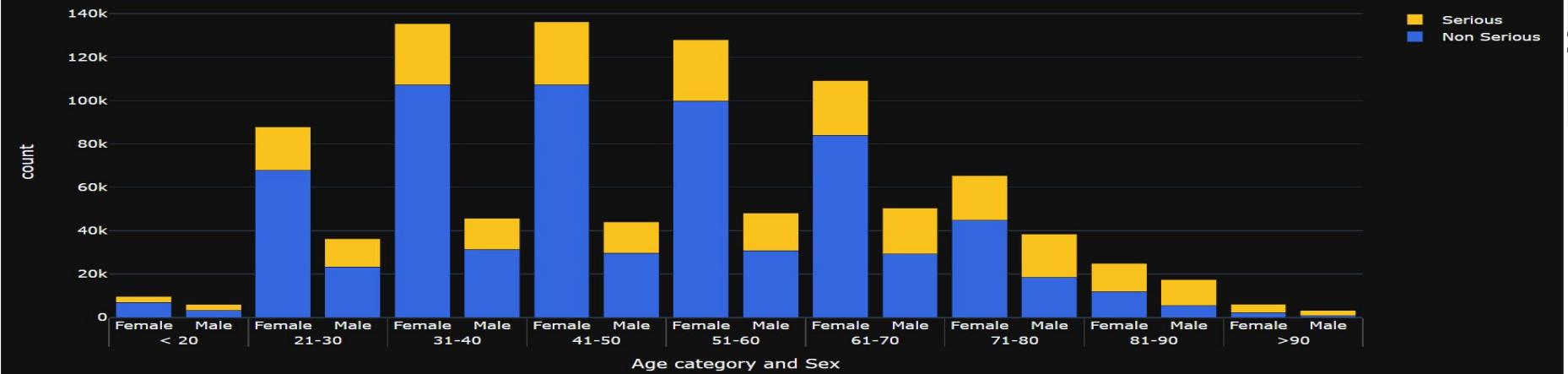
Seriousness of Vaccine Manufacturer



PFIZER\BIONTECH
MODERNA
JANSSEN
UNKNOWN MANUFACTURER

Gender seriousness comparison

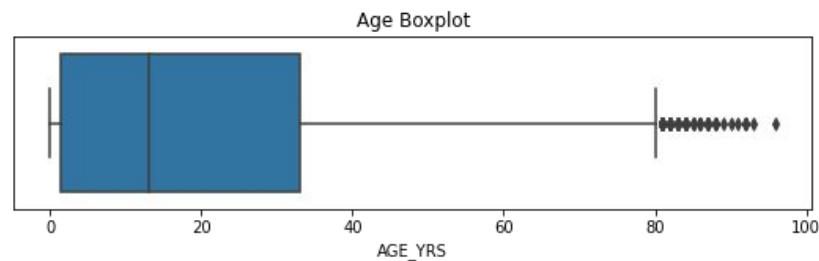
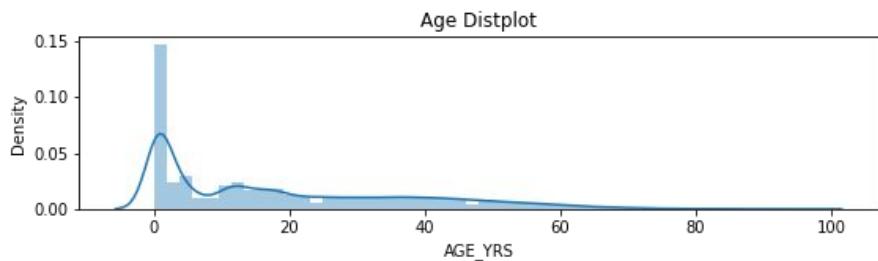




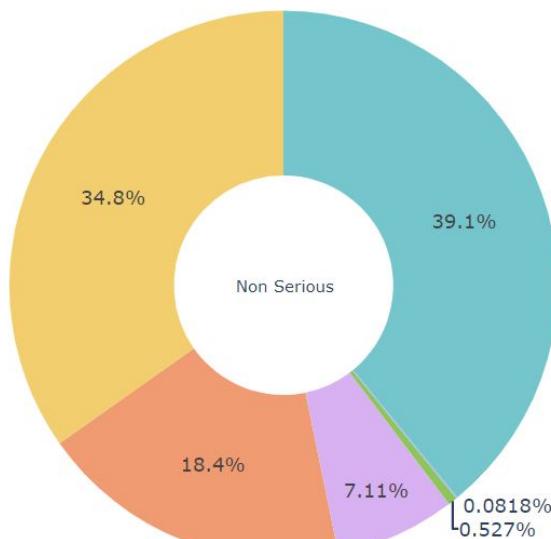
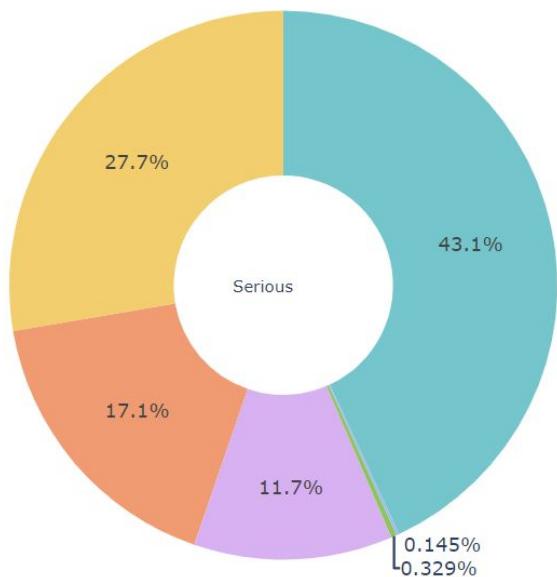
Hepatitis



Plot describes age distributions and the box plot to identify the outliers in them, we considered the age group of vaccines to be between 18 to 100,

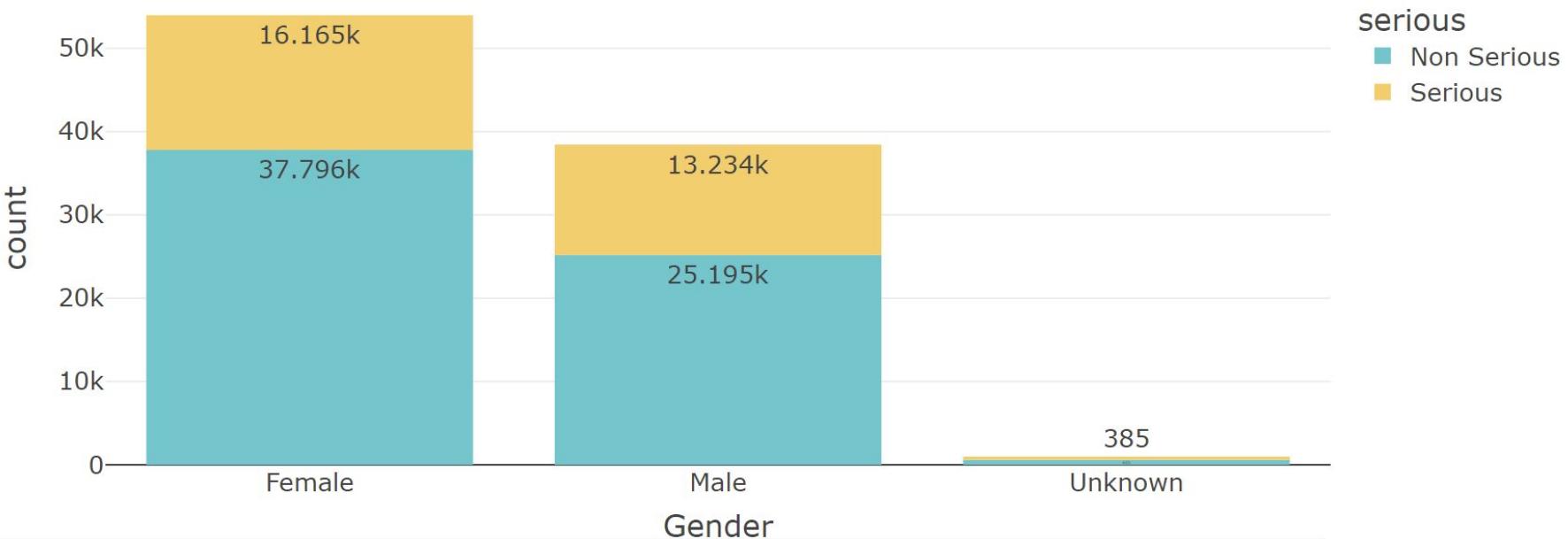


Seriousness of Vaccine Manufacturer



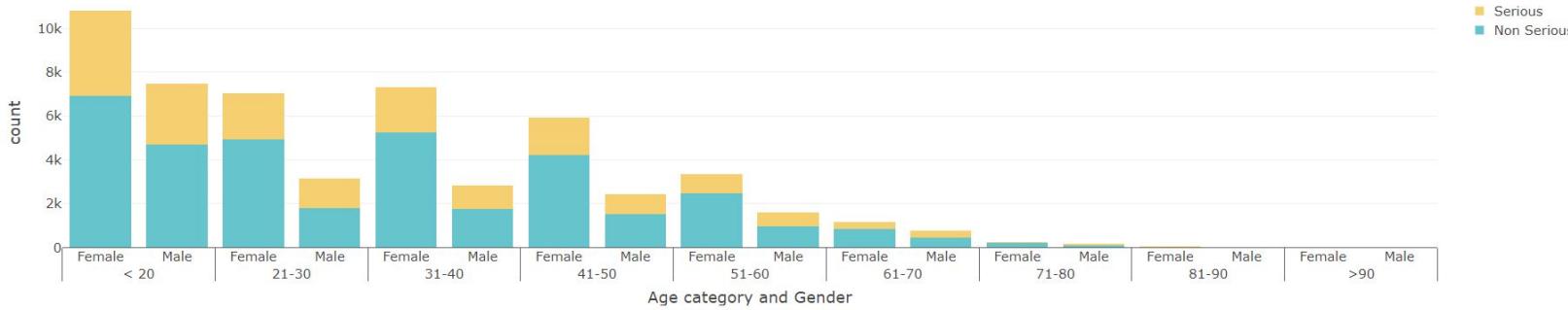
- GLAXOSMITHKLINE BIOLOGICALS
- MERCK & CO. INC.
- SMITHKLINE BEECHAM
- UNKNOWN MANUFACTURER
- DYNAVAX TECHNOLOGIES CORPORATION
- SANOFI PASTEUR

Gender seriousness comparison



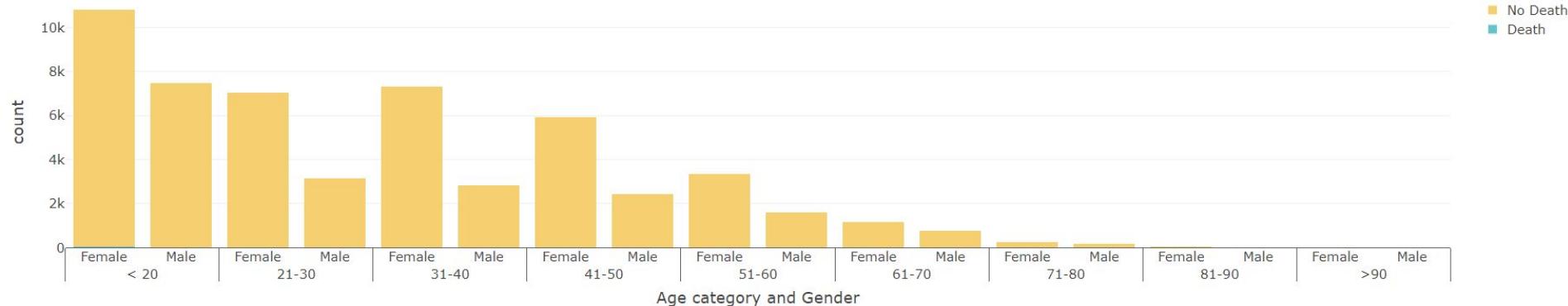
PLot of Age category vs Sex with serious and non serious

Age Category Vs Gender



PLot of Age category vs Sex with Deaths and no Deaths.

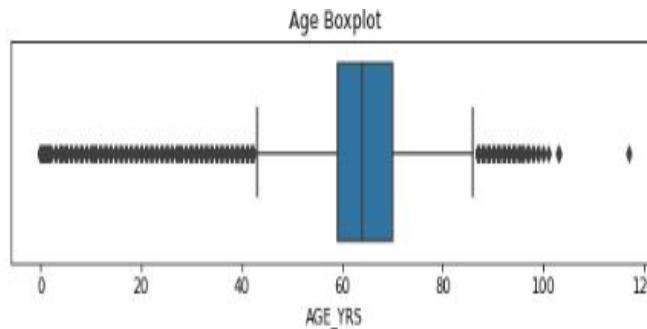
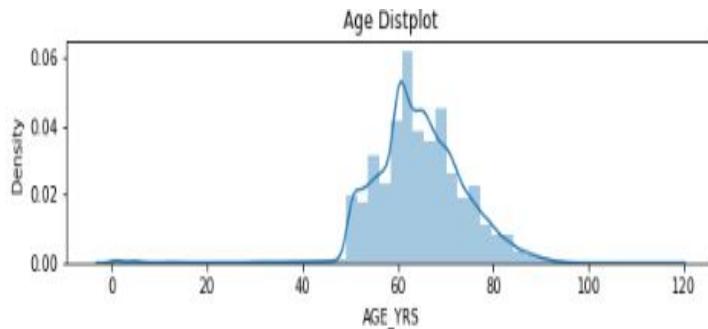
Age Category Vs Gender Vs Death/No Death



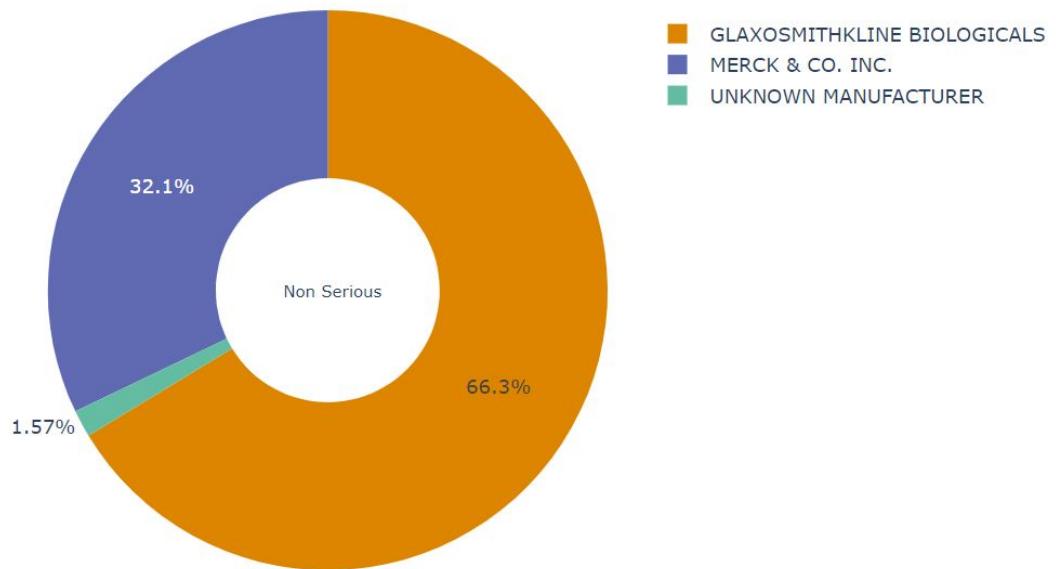
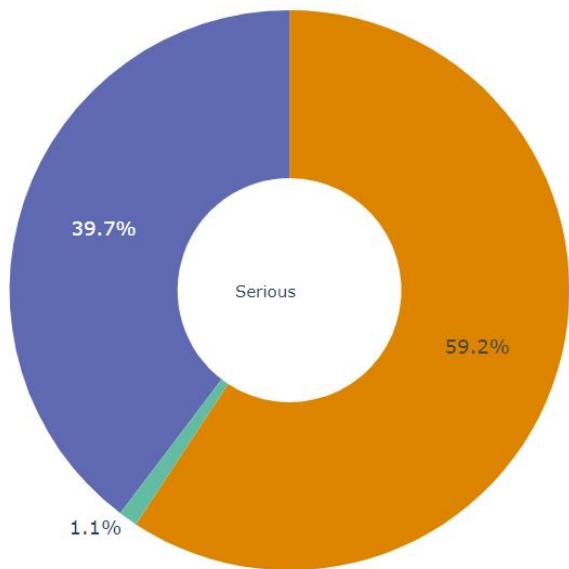
Varzos



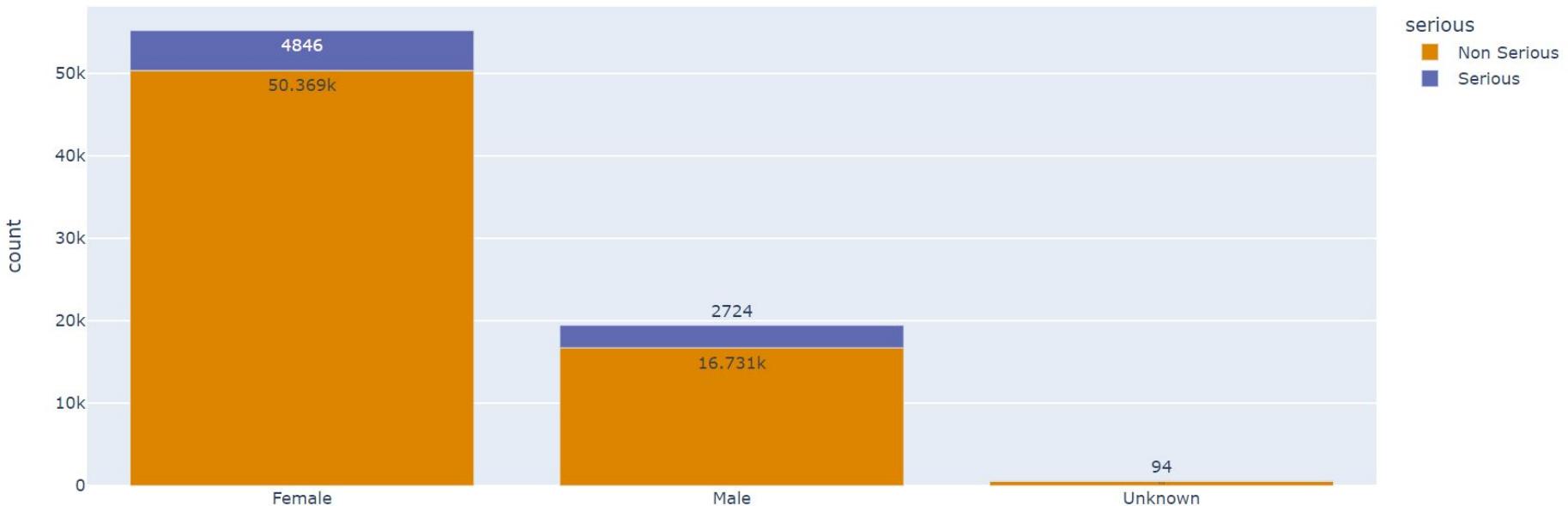
Plot describes age distributions and the box plot to identify the outliers in them, we considered the age group of vaccines to be between 18 to 100,



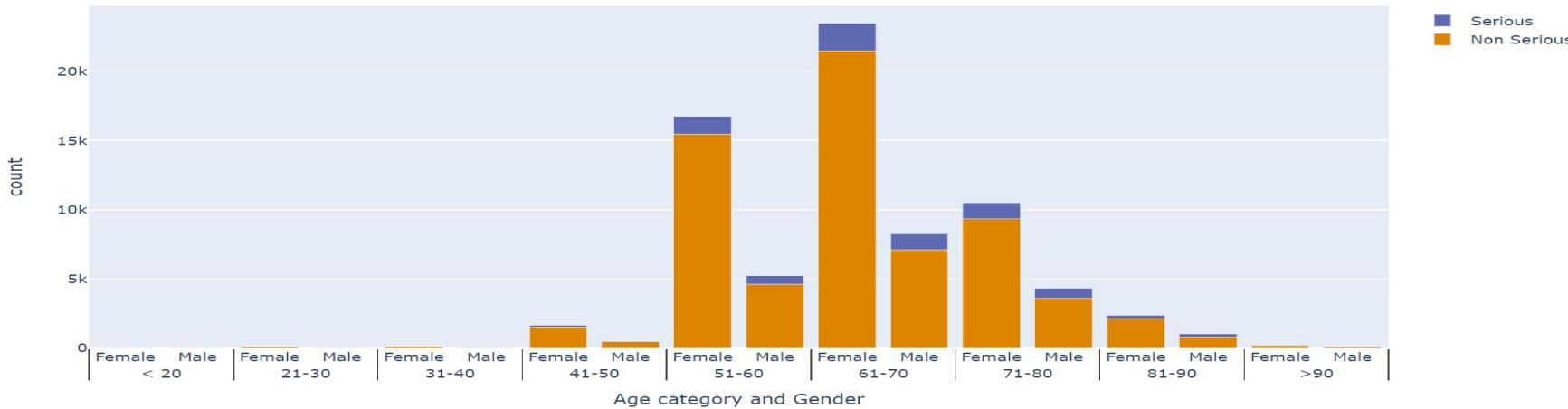
Seriousness of Vaccine Manufacturer



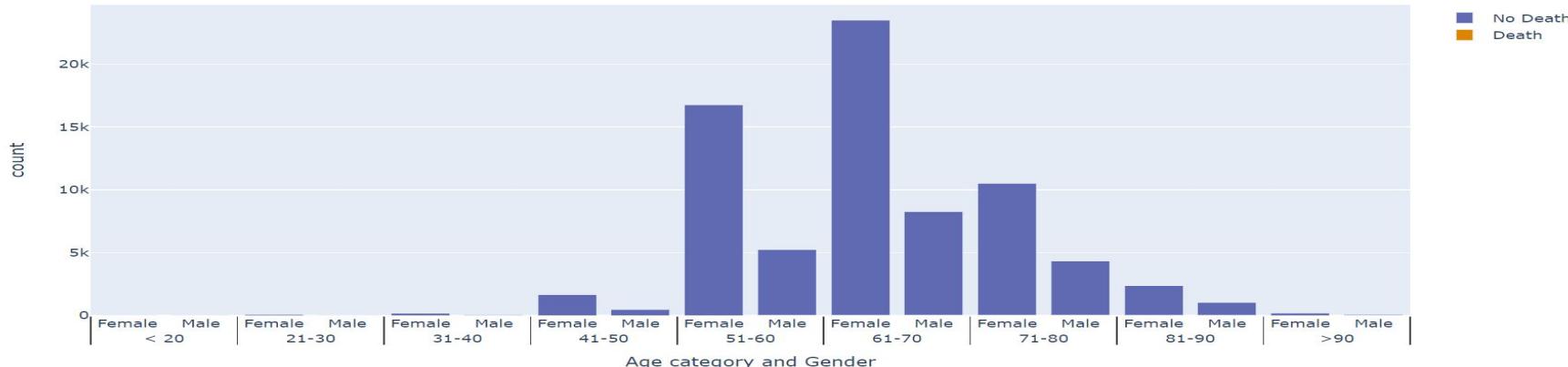
Gender seriousness comparison



Age Category Vs Gender



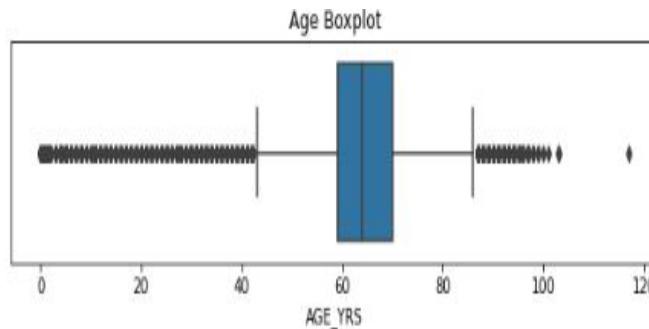
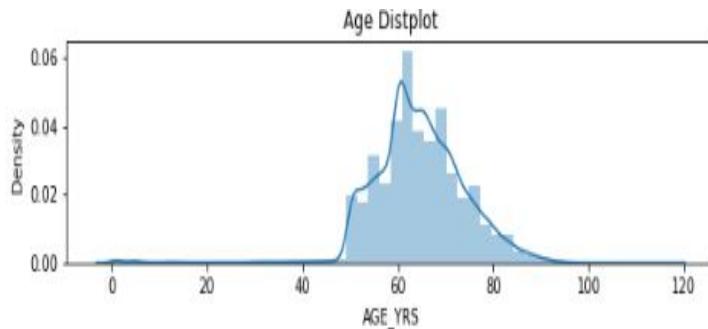
Age Category Vs Gender Vs Death/No Death



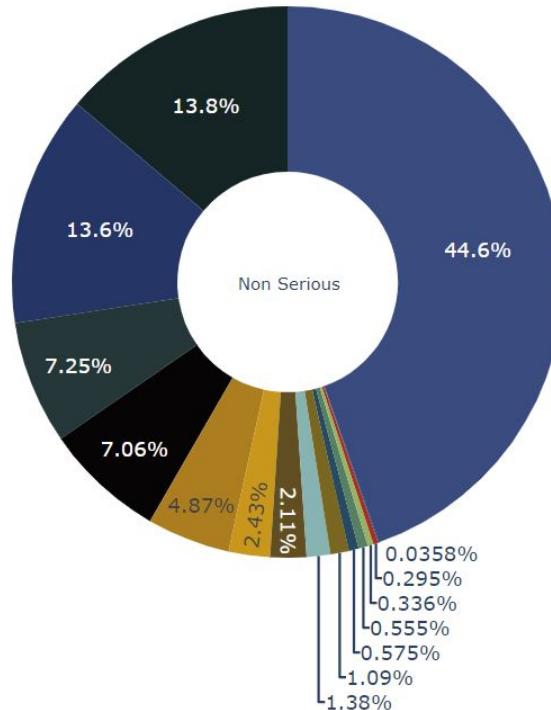
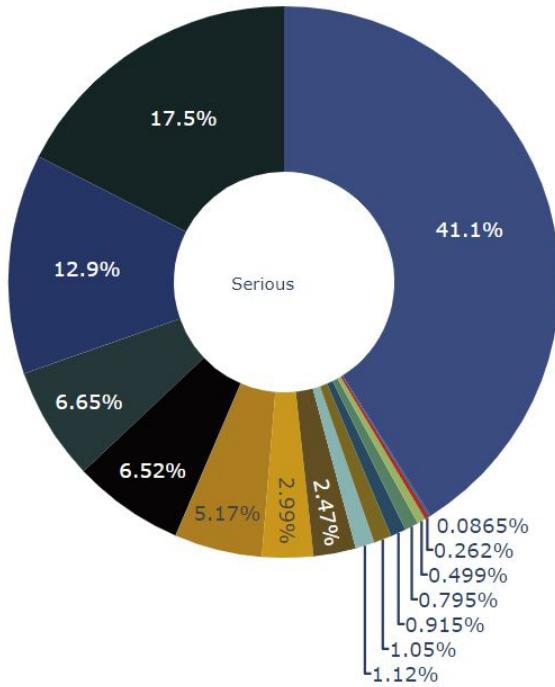
Flu



Plot describes age distributions and the box plot to identify the outliers in them, we considered the age group of vaccines to be between 18 to 100,

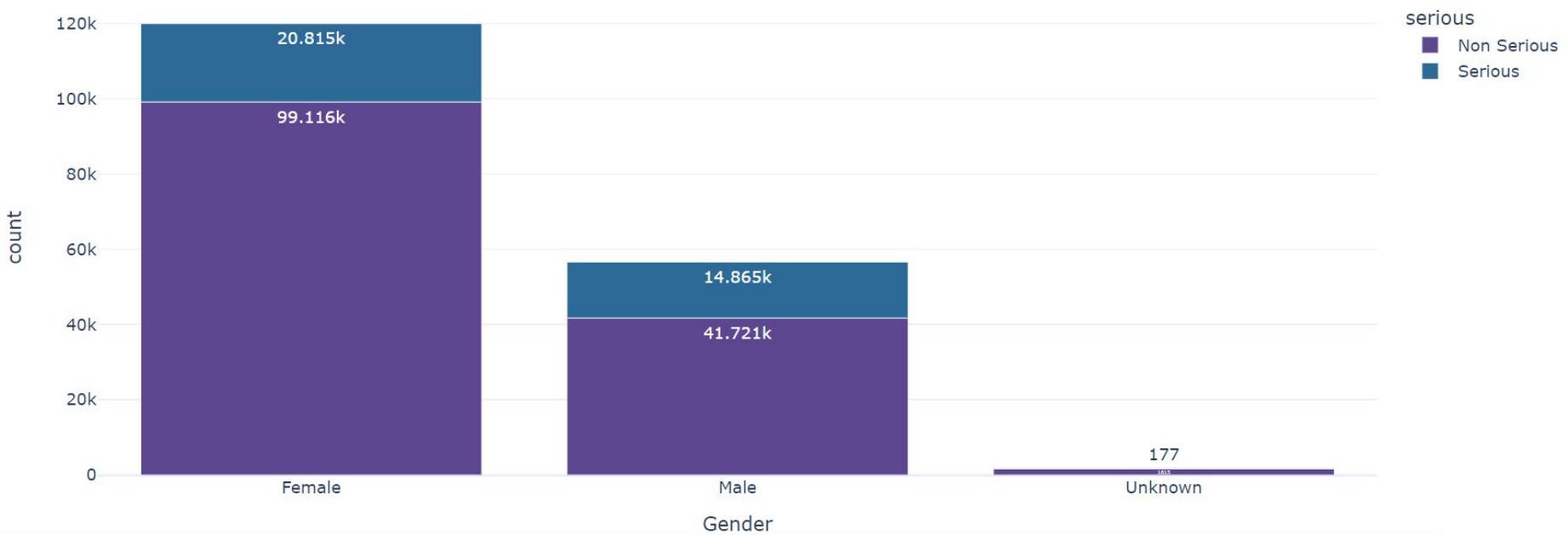


Serious prediction comparison with Manufacturers

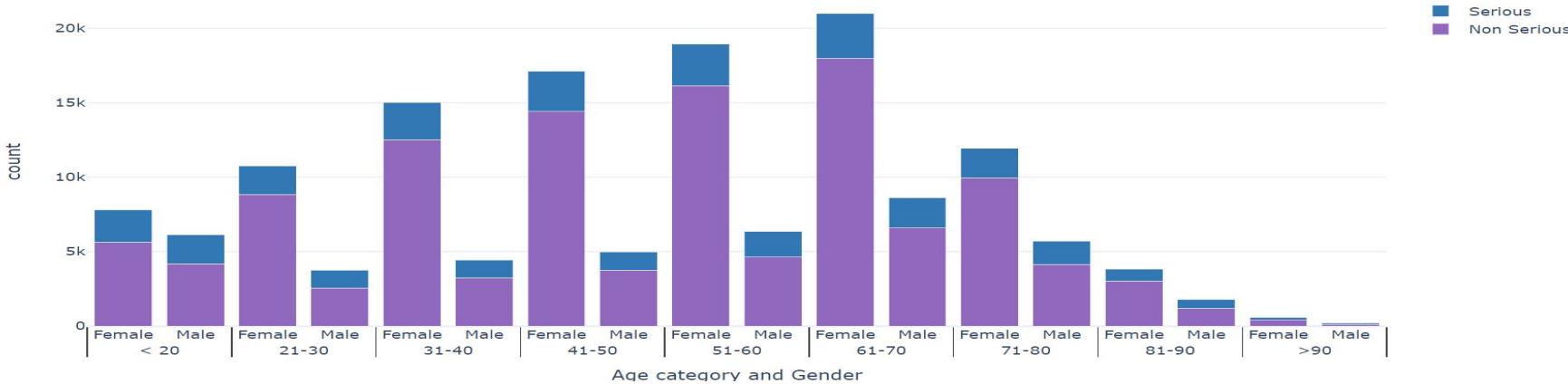


- SANOFI PASTEUR
- GLAXOSMITHKLINE BIOLOGICALS
- NOVARTIS VACCINES AND DIAGNOSTICS
- SEQIRUS, INC.
- MEDIMMUNE VACCINES, INC.
- CSL LIMITED
- CONNAUGHT LABORATORIES
- PFIZER\WYETH
- PROTEIN SCIENCES CORPORATION
- EVANS VACCINES
- MEDEVA PHARMA, LTD.
- PARKE-DAVIS
- AVENTIS PASTEUR
- PARKDALE PHARMACEUTICALS
- LEDERLE LABORATORIES

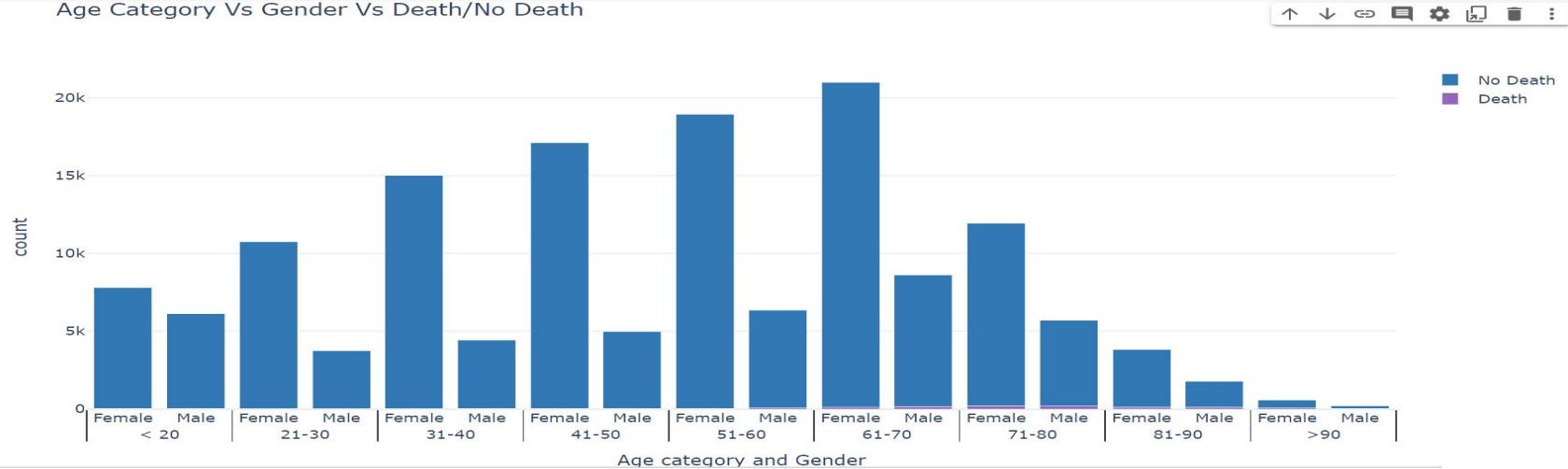
Gender seriousness comparison



Age Category Vs Gender



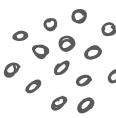
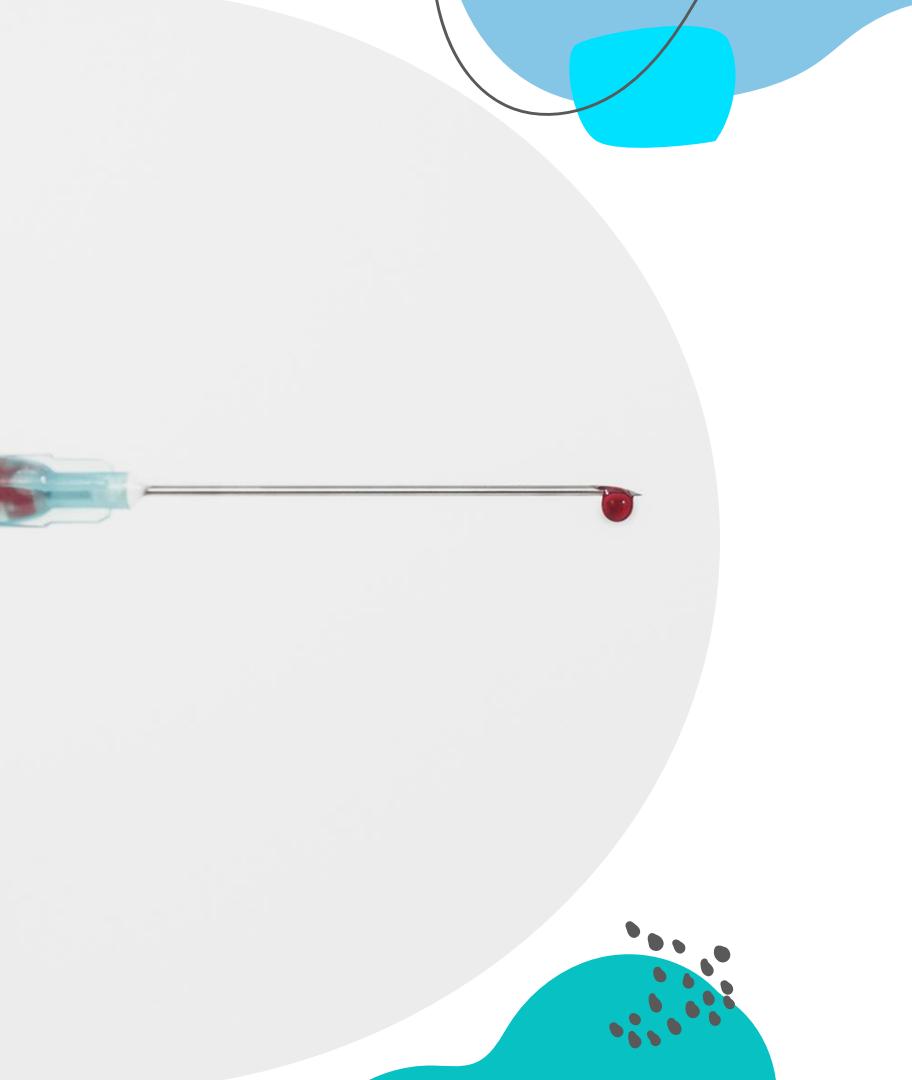
Age Category Vs Gender Vs Death/No Death



03

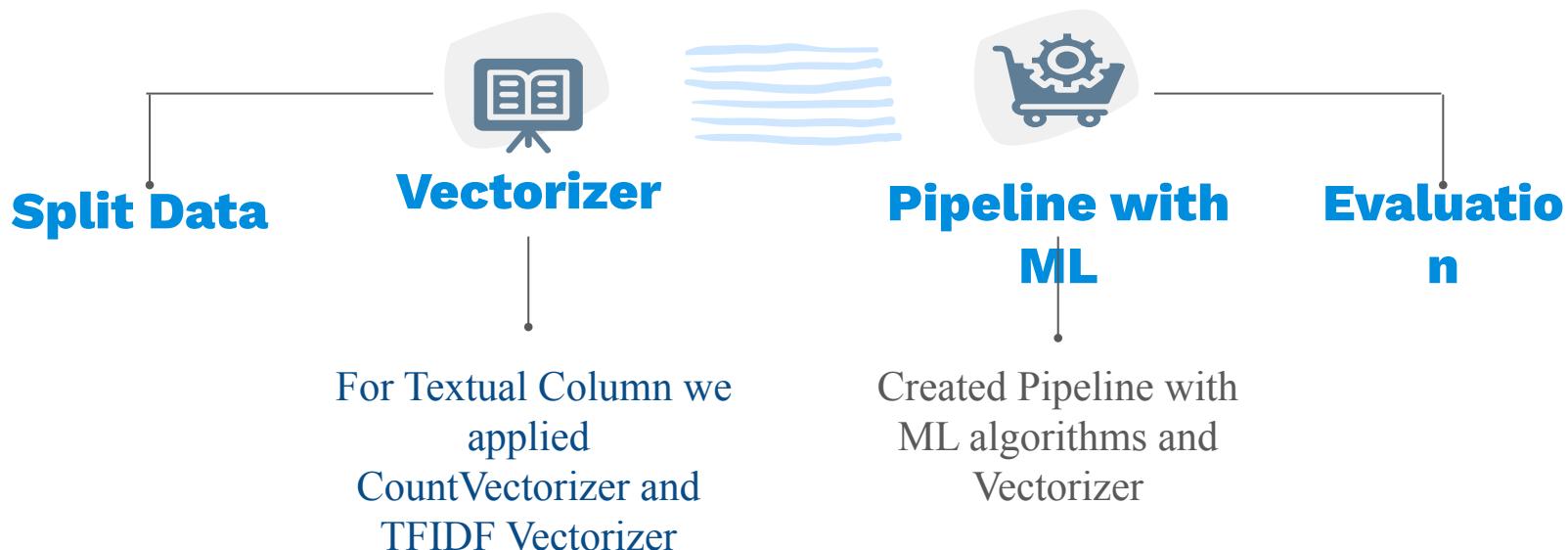
Modeling and Evaluation

Modeling and Optimization
Model Evaluation
User Interface



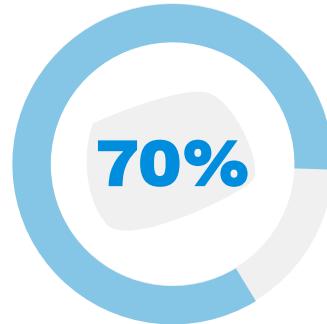
Methodology

Individual Modelling for Each Vaccine.

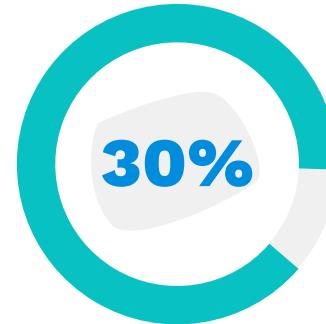


Train and Test Split

#	Column
0	AGE_YRS
1	SEX
2	SYMPTOM_TEXT
3	DIED
4	L_THREAT
5	HOSPITAL
6	HOSPDAYS
7	DISABLE
8	RECOVD
9	OTHER_MEDS
10	CUR_ILL
11	HISTORY
12	BIRTH_DEFECT
13	ALLERGIES
14	VAX_MANU



Train



Test

Models

**Logistic
Regression**

**Multinomial
NB**

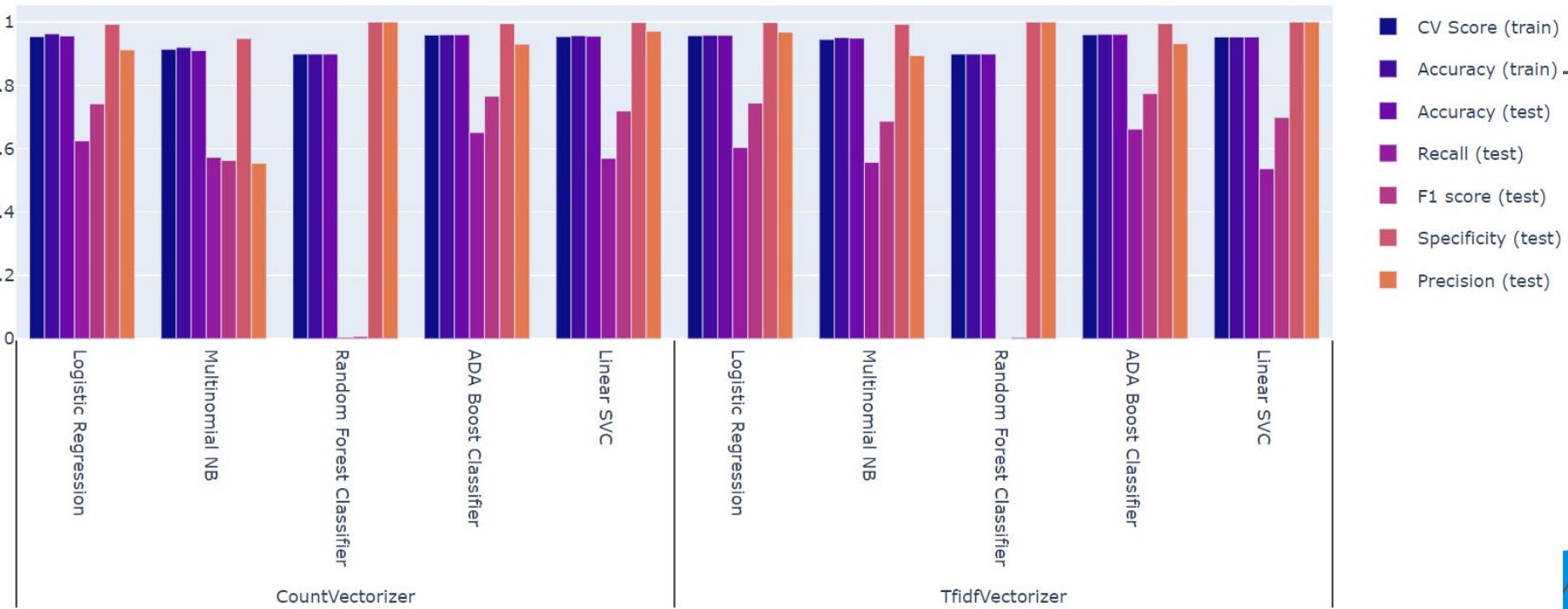
**Random Forest
Classifier**

**ADABoost
Classifier**

Linear SVC

**CountVectorizer &
TFIDF Vectorizer**

Models Results



Models Comparison

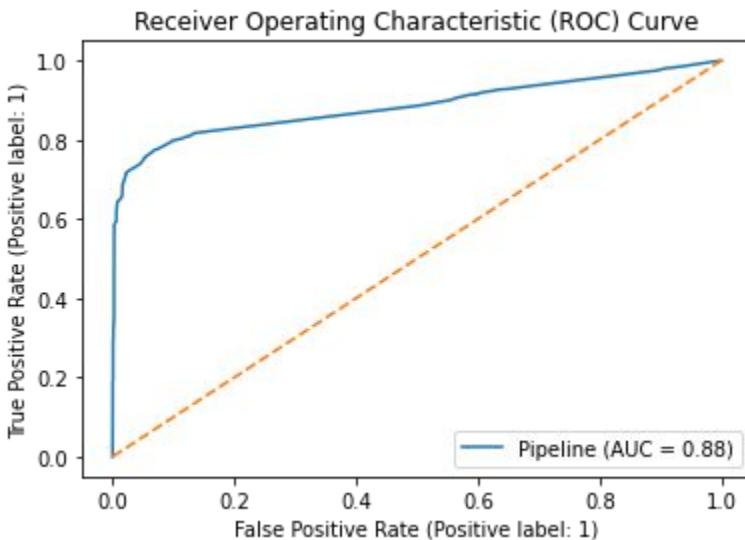
Top Three models

Word Vectorizer	Model	Accuracy	Recall	F1 Score	Precision
TfidfVectorizer	Logistic Regression	0.958	0.604	0.744	0.968
	ADABoost Classification	0.961	0.662	0.774	0.932
	SVM	0.953	0.537	0.699	0.971

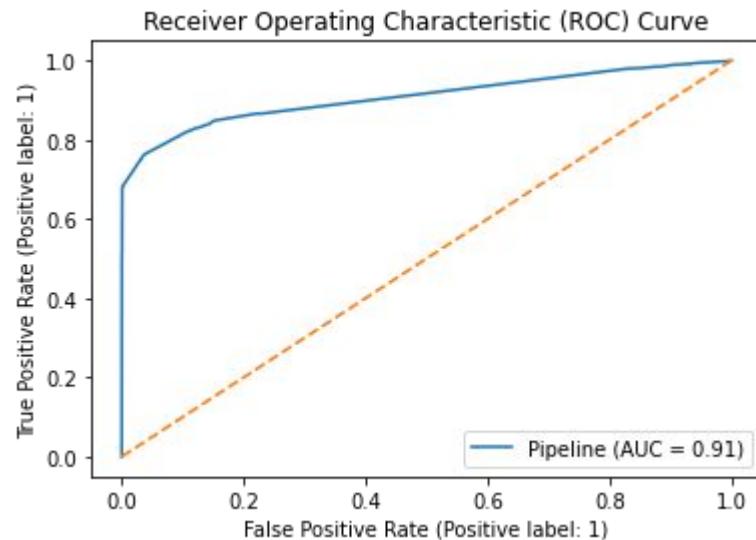
Final Model

ADABoost Classifier was selected as our final model with TFIDF Vectorizer final it lead to AUC of 0.85 to 0.90 between vaccines

Flu Vaccine



Hepatitis Vaccine



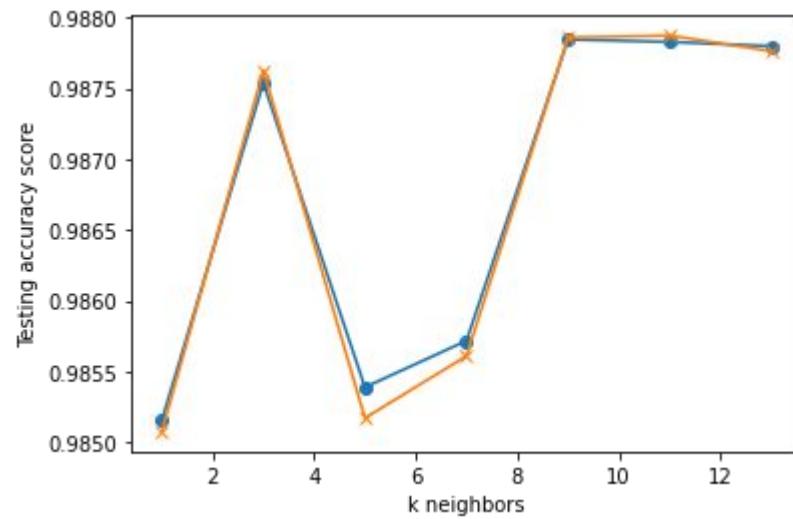
KNN Model for Symptoms Clustering

We Used KNN model to Predicted Life threatening for the patient below are the features used.

#	Column
0	SEX
1	BIRTH_DEFECT
2	DISABLE
3	OTHER_MEDS
4	CUR_ILL
5	HISTORY
6	ALLERGIES
7	< 20
8	21-30
9	31-40
10	41-50
11	51-60
12	61-70
13	71-80
14	81-90
15	>90
16	GLAXOSMITHKLINE
17	MERCK

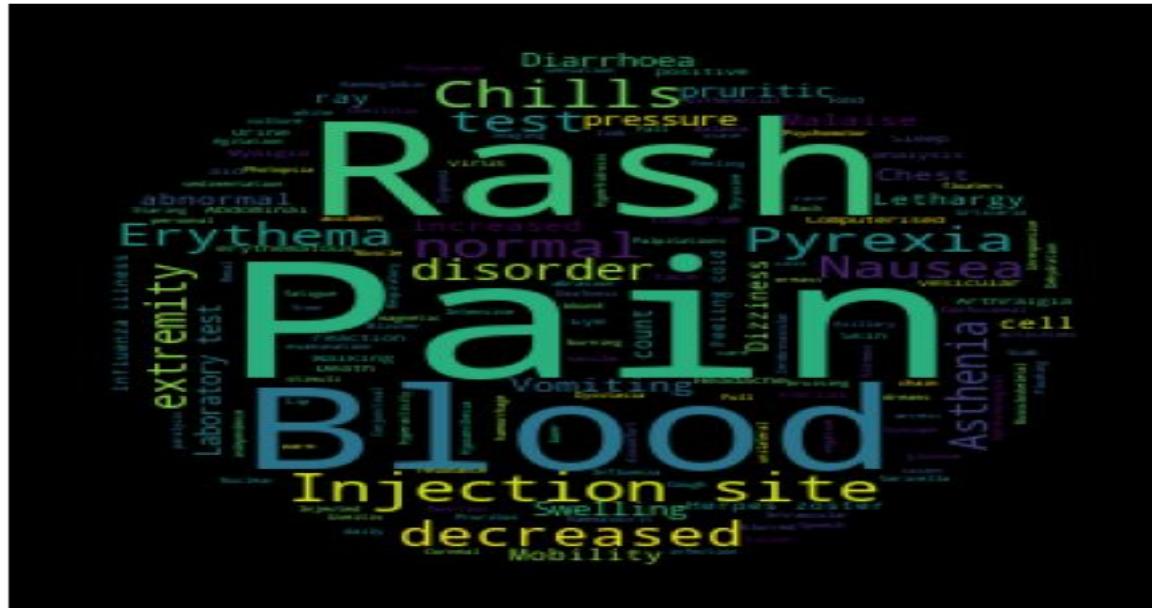
Model Results

k: 1, Train/Test Score: 0.985/0.985
k: 3, Train/Test Score: 0.988/0.988
k: 5, Train/Test Score: 0.985/0.985
k: 7, Train/Test Score: 0.986/0.986
k: 9, Train/Test Score: 0.988/0.988
k: 11, Train/Test Score: 0.988/0.988
k: 13, Train/Test Score: 0.988/0.988



KDTree

To get the related symptoms from user input we selected KDTree model from `sklearn.neighbors`



User Interface

Used Plotly dash for creating this interactive UI. In dash we used Dash core Components for creating Dropdowns, Graphs, input objects and Dash Bootstrap components for Creating layouts, buttons, tabs etc.



User Interface

Adverse Reaction of Vaccination : Risk Prediction and Side Effects Clustering

Seriousness Prediction

Predict Adverse Reactions

Age:

Vaccination Type:

Vaccine Manufacturer:

Sex: Male Female

Symptom:

Patient Died? Yes No

Is Symptom Life Threatening? Yes No

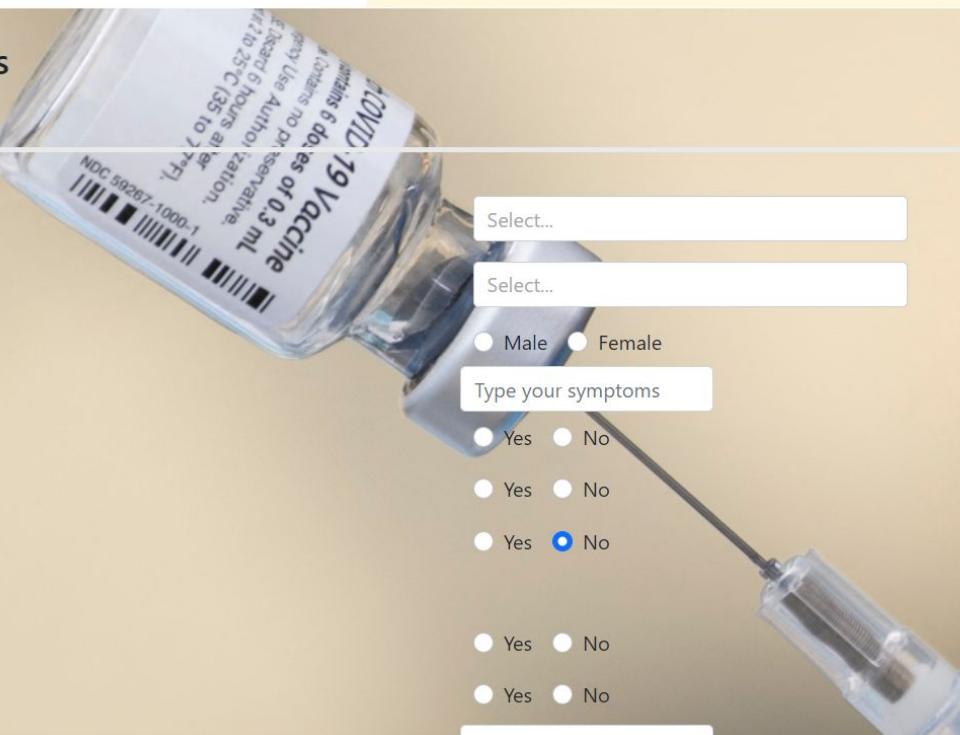
Is Patient Hospitalized? Yes No

How many days patient hospitalized?:

Is Patient having any Disabilities?: Yes No

Is Patient Recovered from Symptoms?: Yes No

Symptoms clustering



A photograph of a clear glass vial containing a white powder, labeled "COVID-19 VACCINE" and "3 ml". Below it is a clear plastic syringe with a needle.

User Interface

Adverse Reaction of Vaccination : Risk Prediction and Side Effects Clustering

Seriousness Prediction



Predict Adverse Reactions

Age Group: Select...

Vaccination Type: Select...

Vaccine Manufacturer: Select...

Sex:

- Male Female
- Yes No

Is Patient having any Disabilities?:

Any Other Medications?:

Is Patient having any Current Illness?:

Is Patient having any Medical History?:

Is Patient having any Allergies?:

Is Patient having any Birth Defect:

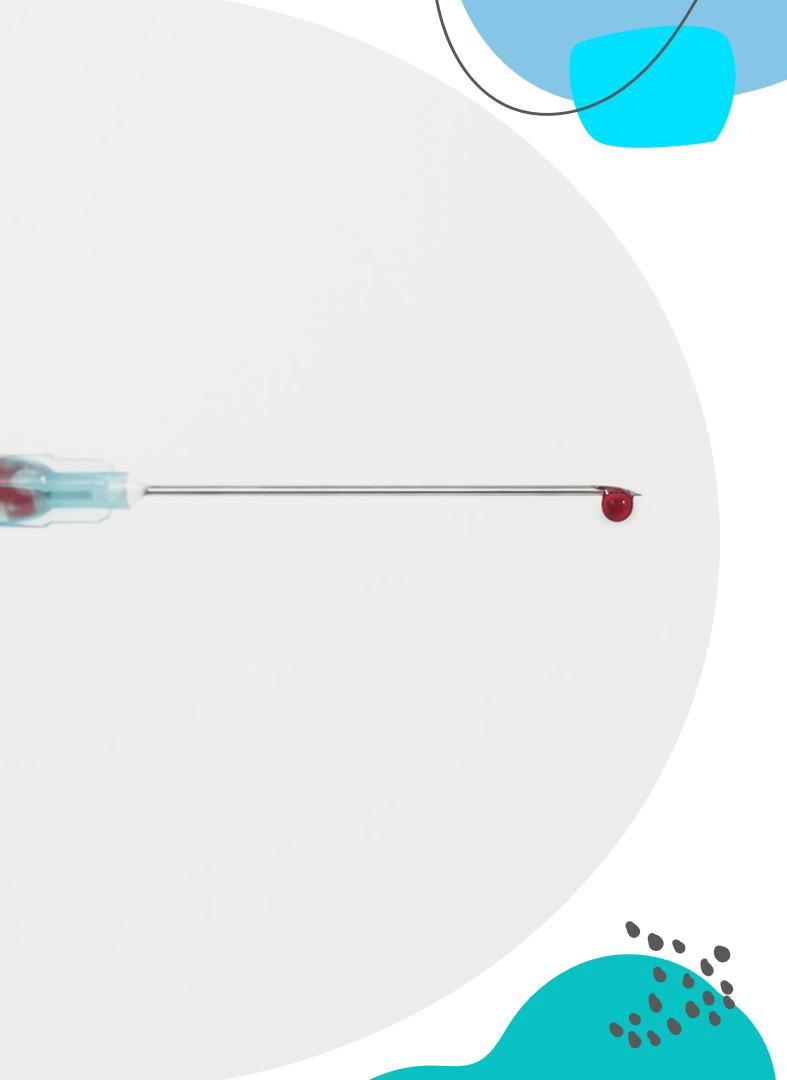
Symptoms clustering

Sample Results



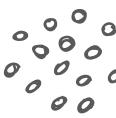
Sample Results





04

Conclusion and Future Work



Conclusions

AE reporting itself is less and most of vaccine consumers are not aware of the reporting. Additionally the minority class is imbalance and oversampling of minority class may give a true representation of the real life distribution and led to our model overfitting to the train so we decided not to move with SMOTE method. Overall we were able to achieve the problem statement by selecting suitable model. As data scientists, we have limited medical expertise and many medical symptom terms had to be researched, working with healthcare professional would had lead to better result.

Finally by developing user interface it will be helpful for both healthcare professional for preliminary screening of all incoming AE reports, to get an initial seriousness classification and for patient it will helpful to check symptoms prior to vaccination to understand about the vaccine

Future Implications

As there is low rate of reporting of AE. It would be help if we look for Adverse events from other sources like EHR and Social Media. Implemented Multi-Label Classification models to demonstrate and accurately predict the symptoms class labels with the usage of deep learning techniques. Usage of Transformers and pre-trained medical terminologies model on Textual Fields for better understanding of Medication and symptoms.

Our Team

Sharath Srinivas



Chetan B Desai



Saideep Malgireddy

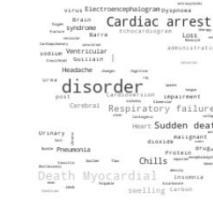
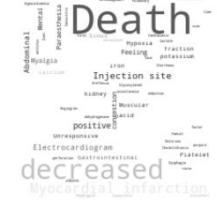


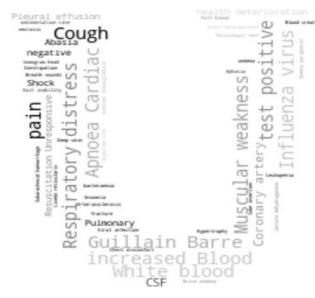
References

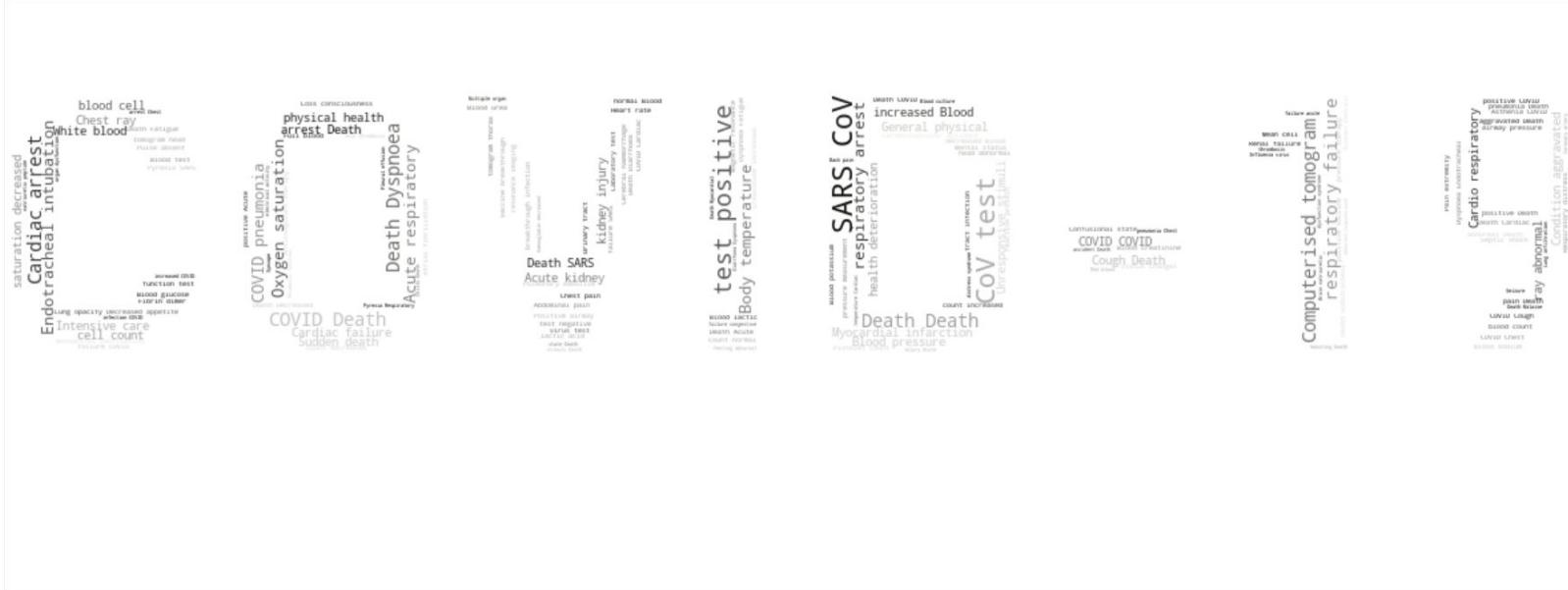
- [1]. <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/adverse-events.html>
- [2]. <https://pubmed.ncbi.nlm.nih.gov/15071280/>
- [3].<https://www.cdc.gov>
- [4].https://www.ema.europa.eu/documents/other/meddra-important-medical-event-terms-list-version-250_en.xlsx
- [5].<https://plotly.com/python/>
- [6].<https://www.historyofvaccines.org/content/articles/vaccine-side-effects-and-adverse-events>

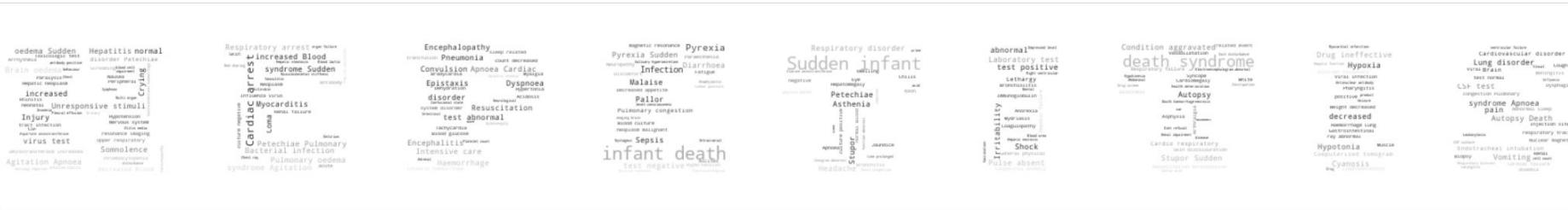
World Clouds











A circular inset on the left side of the slide shows a close-up of a hand wearing a blue nitrile glove. The hand is holding a metal test tube rack that contains eight test tubes, each filled with a bright blue liquid. The background of the slide features abstract blue and white organic shapes.

Thanks