Table of Contents	
Appendix A	Python Notebooks
Appendix B	Datasets
Appendix C	ANOVA Results
Appendix D	Topic Scores Visualizations

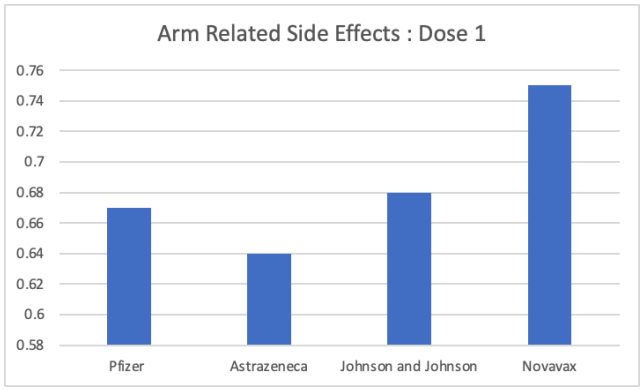
Appendix A (.ipynb)

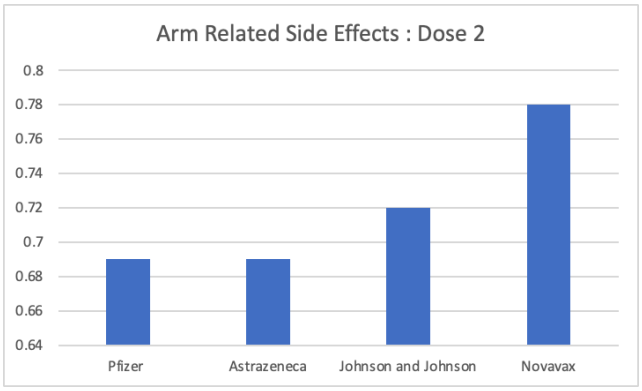
- **Pfizer_2022_by_user**
  - The following script utilizes the Snscrape package to extract Twitter data.
  - Initially, it scrapes tweets from Twitter related to Pfizer for January 2022.
  - Subsequently, the code is executed to extract tweets for the same user for a period of 10 days.
  - This process is repeated for other dates and vaccines to collect the required tweets. The output of this script, for each month and vaccine, is stored in the **User_tweets** folder.
- **Zero Shot Classifier**
  - The following iPython file performs data pre-processing and cleaning, as well as executing the zero-shot classifier.
  - **Input data** from the **User_Tweets** directory for all 12 months, one vaccine at a time.
  - The pre-processing steps and zero-shot classifier execution described in the respective slides are carried out in this code.
  - **Output** of this code is stored in the **Topic_Score** folder.
- **Final Topic Scores**
  - Input Data: All the topic score .csv files for a particular vaccine (includes both years).
  - Combines all the topic scores into a single dataframe. Drops Irrelevant tweets. For each dose, calculates the mean score for each topic.
- **ANOVA**
  - The folder contain python notebooks with ANOVA test for all 4 vaccines and 5 Topic Scores.

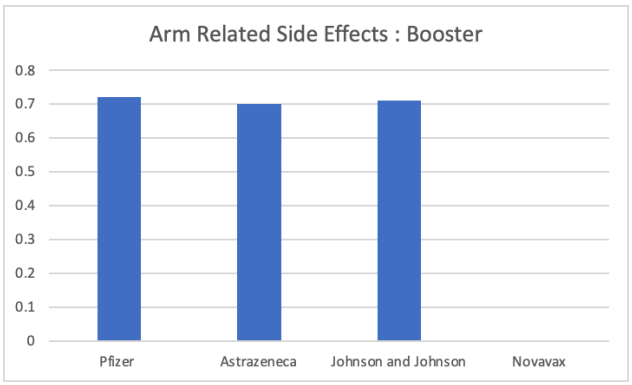
Appendix B (Datasets)

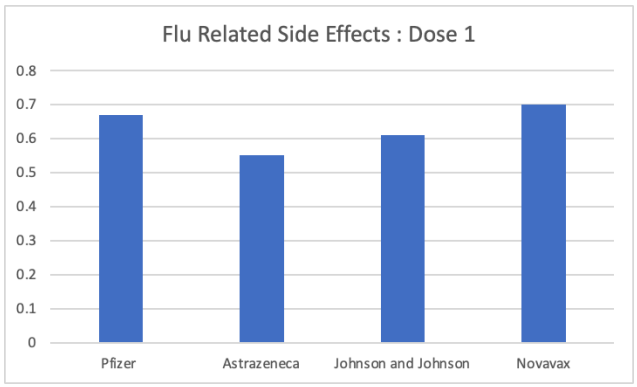
- User_Tweets (83 .csv files)
  - These are the raw files that were generated by running the ‘Pfizer_2022_by_user’ code.
    Each .csv file has a single month of tweets for a particular vaccine.
    These files serve as input for the ‘Zero Shot Classifier.ipynb’ notebook
- Topic_Score (76 .csv files)
  - These files are generated by running the ‘Zero Shot Classifier.ipynb’ notebook.
  - Contains the results of the zero-shot classifier.
  - Each .csv file has a month of topic scores for a particular vaccine.

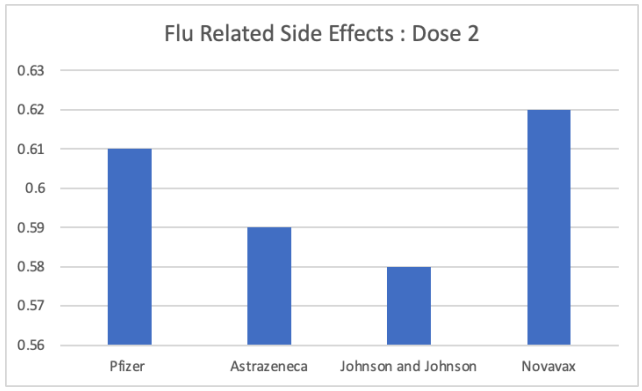
Appendix C

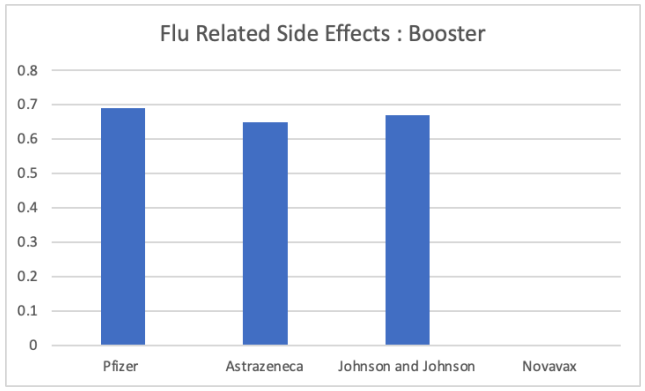

Arm Related Side Effects : Dose 1
0.76
0.74
0.72
0.7
0.68
0.66
0.64
0.62
0.6
0.58
Pfizer
Astrazeneca
Johnson and Johnson
Novavax

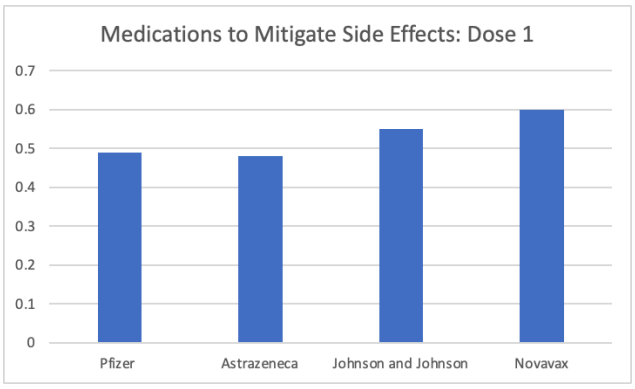

Arm Related Side Effects : Dose 2
0.8
0.78
0.76
0.74
0.72
0.7
0.68
0.66
0.64
Pfizer
Astrazeneca
Johnson and Johnson
Novavax

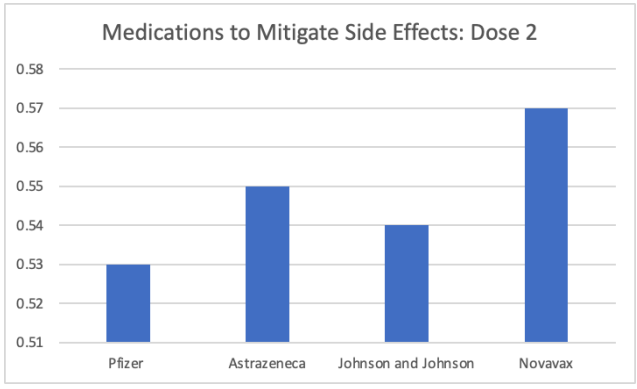

Arm Related Side Effects : Booster
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Pfizer
Astrazeneca
Johnson and Johnson
Novavax

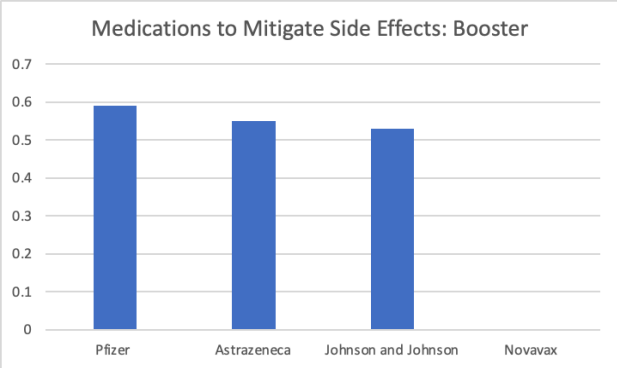

Flu Related Side Effects : Dose 1
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Pfizer
Astrazeneca
Johnson and Johnson
Novavax

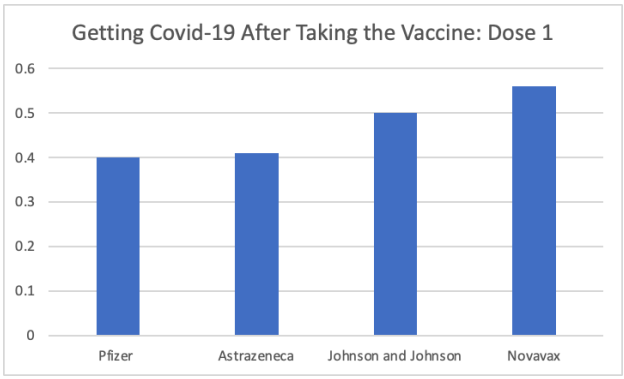

Flu Related Side Effects : Dose 2
0.63
0.62
0.61
0.6
0.59
0.58
0.57
0.56
Pfizer
Astrazeneca
Johnson and Johnson
Novavax

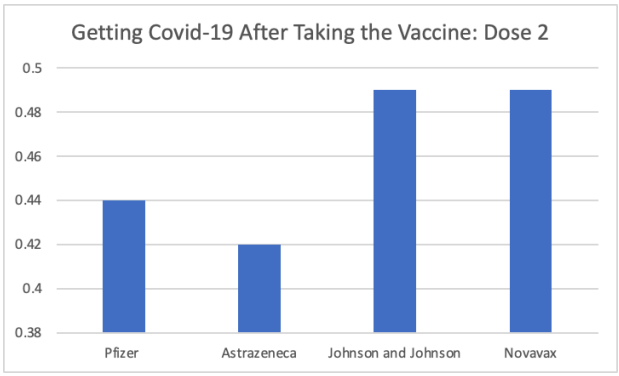

Flu Related Side Effects : Booster
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Pfizer
Astrazeneca
Johnson and Johnson
Novavax

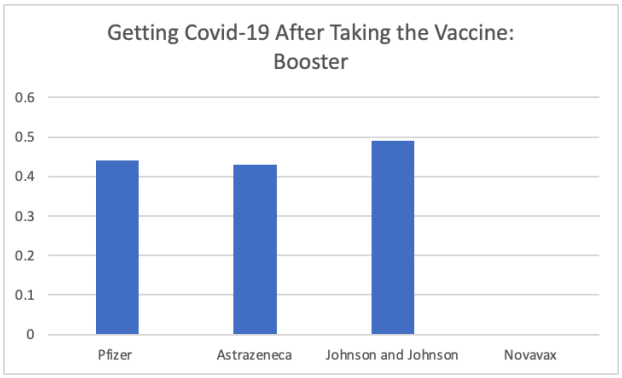

Medications to Mitigate Side Effects: Dose 1
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Pfizer
Astrazeneca
Johnson and Johnson
Novavax

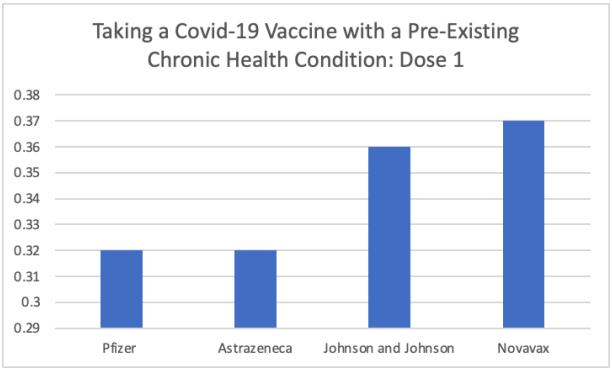

Medications to Mitigate Side Effects: Dose 2
0.58
0.57
0.56
0.55
0.54
0.53
0.52
0.51
Pfizer
Astrazeneca
Johnson and Johnson
Novavax

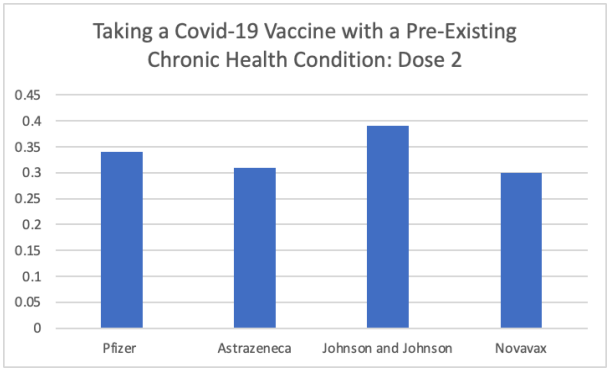

Medications to Mitigate Side Effects: Booster
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Pfizer
Astrazeneca
Johnson and Johnson
Novavax

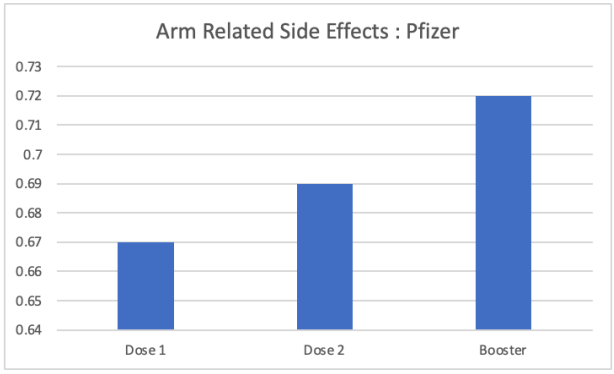

Getting Covid-19 After Taking the Vaccine: Dose 1
0.6
0.5
0.4
0.3
0.2
0.1
0
Pfizer
Astrazeneca
Johnson and Johnson
Novavax

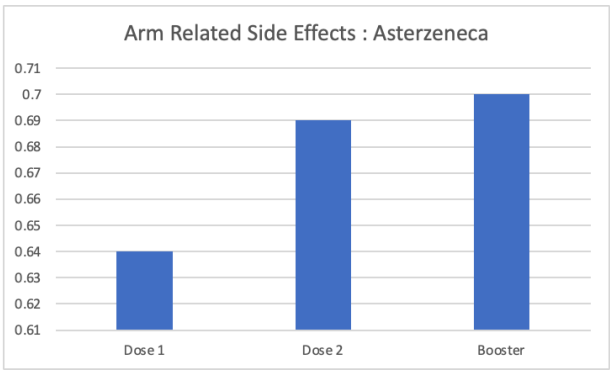

Getting Covid-19 After Taking the Vaccine: Dose 2
0.5
0.48
0.46
0.44
0.42
0.4
0.38
Pfizer
Astrazeneca
Johnson and Johnson
Novavax

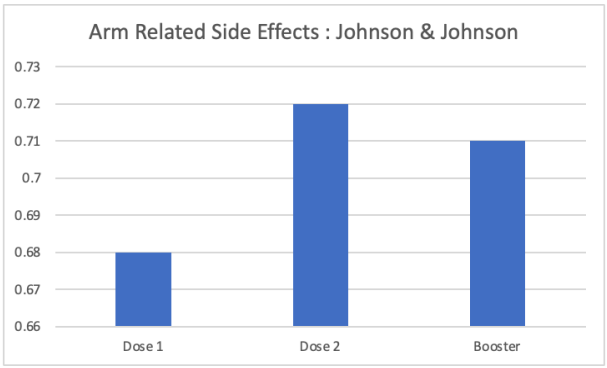

Getting Covid-19 After Taking the Vaccine: Booster
0.6
0.5
0.4
0.3
0.2
0.1
0
Pfizer
Astrazeneca
Johnson and Johnson
Novavax

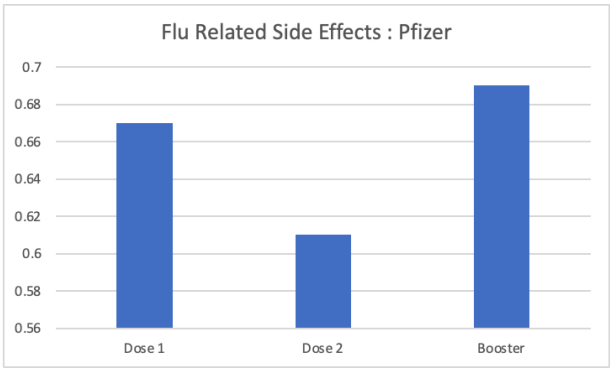

Taking a Covid-19 Vaccine with a Pre-Existing Chronic Health Condition: Dose 1
0.38
0.37
0.36
0.35
0.34
0.33
0.32
0.31
0.3
0.29
Pfizer
Astrazeneca
Johnson and Johnson
Novavax

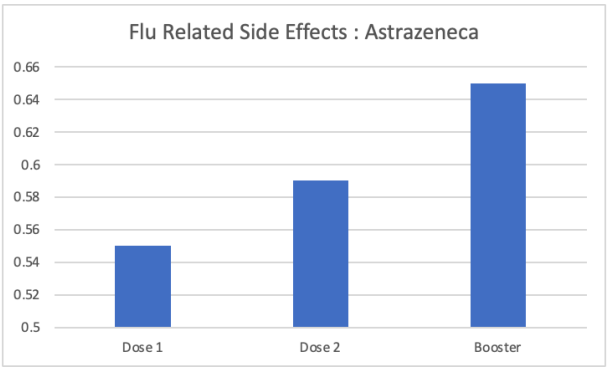

Taking a Covid-19 Vaccine with a Pre-Existing Chronic Health Condition: Dose 2
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Pfizer
Astrazeneca
Johnson and Johnson
Novavax

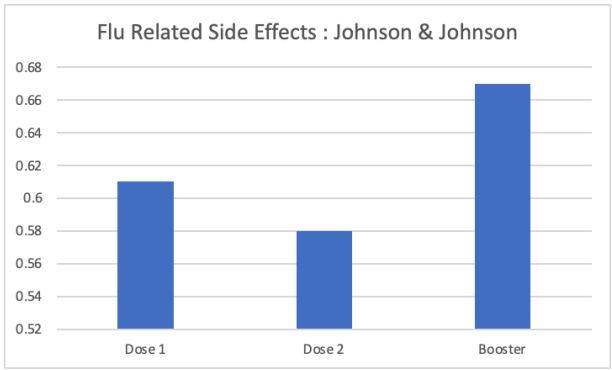

Arm Related Side Effects : Pfizer
0.73
0.72
0.71
0.7
0.69
0.68
0.67
0.66
0.65
0.64
Dose 1
Dose 2
Booster

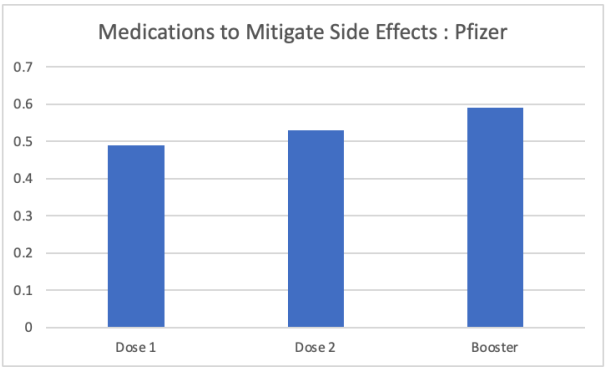

Arm Related Side Effects : Asterzeneca
0.71
0.7
0.69
0.68
0.67
0.66
0.65
0.64
0.63
0.62
0.61
Dose 1
Dose 2
Booster

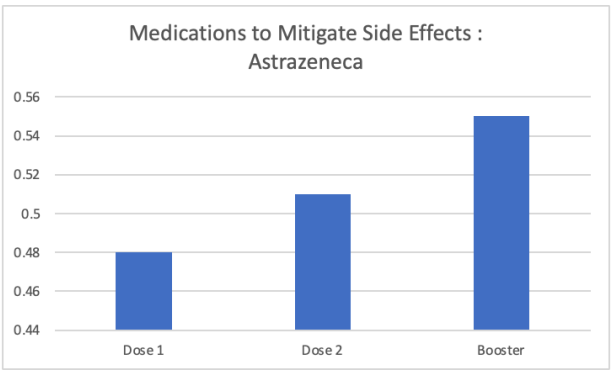

Arm Related Side Effects : Johnson & Johnson
0.73
0.72
0.71
0.7
0.69
0.68
0.67
0.66
Dose 1
Dose 2
Booster

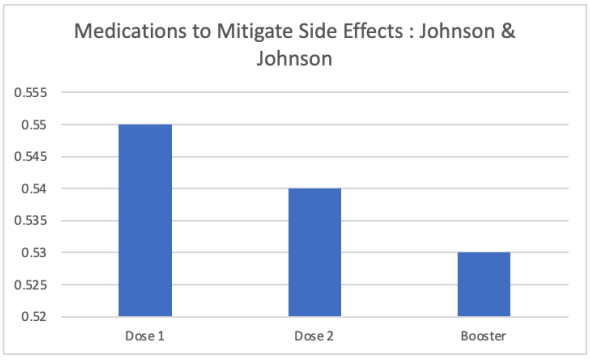

Flu Related Side Effects : Pfizer
0.7
0.68
0.66
0.64
0.62
0.6
0.58
0.56
Dose 1
Dose 2
Booster

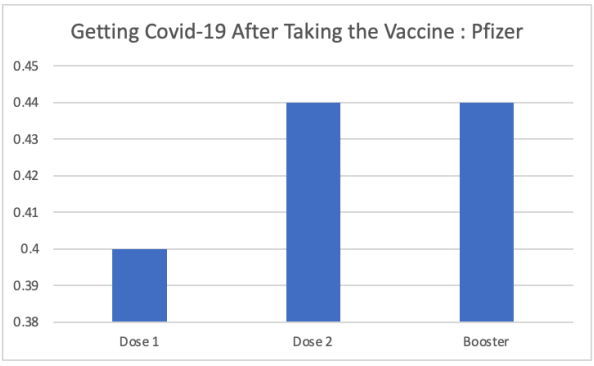

Flu Related Side Effects : Astrazeneca
0.66
0.64
0.62
0.6
0.58
0.56
0.54
0.52
0.5
Dose 1
Dose 2
Booster

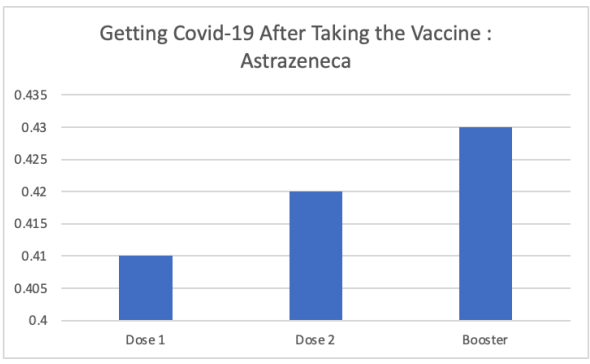

Flu Related Side Effects : Johnson & Johnson
0.68
0.66
0.64
0.62
0.6
0.58
0.56
0.54
0.52
Dose 1
Dose 2
Booster

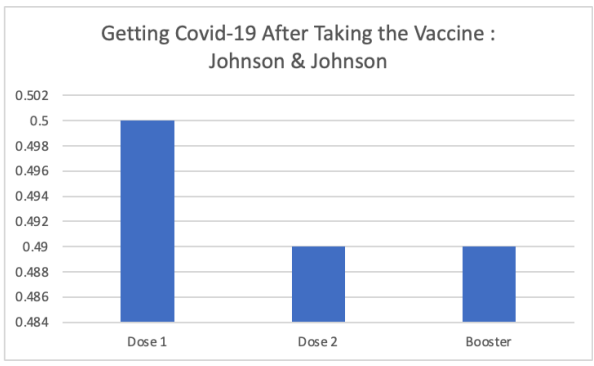

Medications to Mitigate Side Effects : Pfizer
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Dose 1
Dose 2
Booster

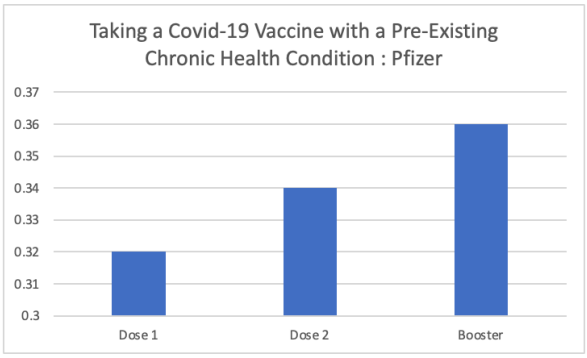

Medications to Mitigate Side Effects : Astrazeneca
0.56
0.54
0.52
0.5
0.48
0.46
0.44
Dose 1
Dose 2
Booster

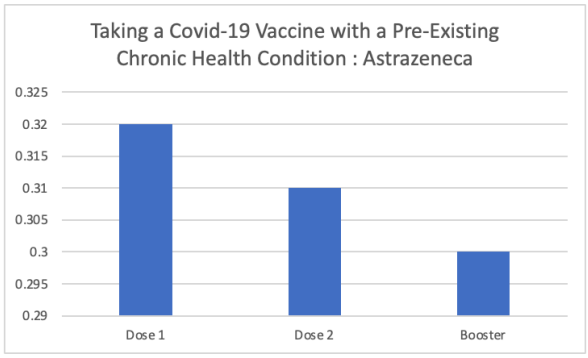

Medications to Mitigate Side Effects : Johnson & Johnson
0.555
0.55
0.545
0.54
0.535
0.53
0.525
0.52
Dose 1
Dose 2
Booster

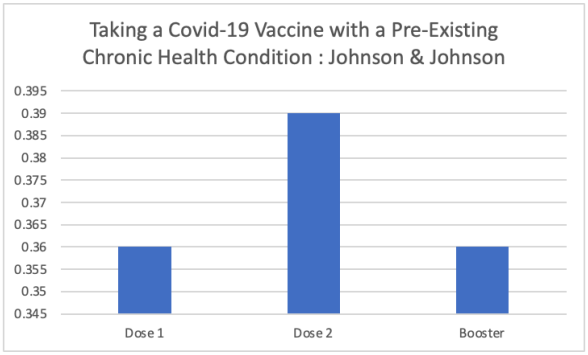

Getting Covid-19 After Taking the Vaccine : Pfizer
0.45
0.44
0.43
0.42
0.41
0.4
0.39
0.38
Dose 1
Dose 2
Booster


Getting Covid-19 After Taking the Vaccine : Astrazeneca
0.435
0.43
0.425
0.42
0.415
0.41
0.405
0.4
Dose 1
Dose 2
Booster


Getting Covid-19 After Taking the Vaccine : Johnson & Johnson
0.502
0.5
0.498
0.496
0.494
0.492
0.49
0.488
0.486
0.484
Dose 1
Dose 2
Booster







Appendix D

ANOVA TO CHECK IF THERE IS ANY DIFFERENCE BETWEEN THE VACCINES FOR EACH OF THE 5 TOPICS IDENTIFIED

TOPIC 1 : DOSE 1

	df	sum_sq	mean_sq	F	PR(>F)
C(vaccine)	3.0	6.500000e-03	0.002167	0.0	NaN
Residual	0.0	2.465190e-31	inf	NaN	NaN

No statistically significant difference

TOPIC 1 : DOSE 2

	df	sum_sq	mean_sq	F	PR(>F)
C(vaccine)	3.0	9.000000e-04	0.0003	0.0	NaN
Residual	0.0	9.860761e-32	inf	NaN	NaN

No statistically significant difference

TOPIC 1 : BOOSTER

	df	sum_sq	mean_sq	F	PR(>F)
C(vaccine)	2.0	2.000000e-04	0.0001	0.0	NaN
Residual	0.0	2.465190e-32	inf	NaN	NaN

No statistically significant difference

TOPIC 2 : DOSE 1

	df	sum_sq	mean_sq	F	PR(>F)
C(vaccine)	3.0	1.327500e-02	0.004425	0.0	NaN
Residual	0.0	2.834969e-31	inf	NaN	NaN

No statistically significant difference

TOPIC 2 : DOSE 2

	df	sum_sq	mean_sq	F	PR(>F)
C(vaccine)	3.0	1.000000e-03	0.000333	0.0	NaN
Residual	0.0	2.218671e-31	inf	NaN	NaN

No statistically significant difference

TOPIC 2 : BOOSTER

	df	sum_sq	mean_sq	F	PR(>F)
C(vaccine)	2.0	1.866667e-03	0.000933	0.0	NaN
Residual	0.0	1.232595e-32	inf	NaN	NaN

No statistically significant difference

TOPIC 3 : DOSE 1

	df	sum_sq	mean_sq	F	PR(>F)
C(vaccine)	3.0	9.400000e-03	0.003133	0.0	NaN
Residual	0.0	2.157042e-31	inf	NaN	NaN

No statistically significant difference

TOPIC 3 : DOSE 2

	df	sum_sq	mean_sq	F	PR(>F)
C(vaccine)	3.0	1.875000e-03	0.000625	0.0	NaN
Residual	0.0	1.602374e-31	inf	NaN	NaN

No statistically significant difference

TOPIC 3 : BOOSTER

	df	sum_sq	mean_sq	F	PR(>F)
C(vaccine)	2.0	1.866667e-03	0.000933	0.0	NaN
Residual	0.0	1.232595e-32	inf	NaN	NaN

No statistically significant difference

TOPIC 4 : DOSE 1

	df	sum_sq	mean_sq	F	PR(>F)
C(vaccine)	3.0	1.747500e-02	0.005825	0.0	NaN
Residual	0.0	9.552613e-32	inf	NaN	NaN

No statistically significant difference

TOPIC 4 : DOSE 2

	df	sum_sq	mean_sq	F	PR(>F)
C(vaccine)	3.0	3.800000e-03	0.001267	0.0	NaN
Residual	0.0	6.779273e-32	inf	NaN	NaN

No statistically significant difference

TOPIC 4 : BOOSTER

	df	sum_sq	mean_sq	F	PR(>F)
C(vaccine)	2.0	2.066667e-03	0.001033	0.0	NaN
Residual	0.0	6.162976e-33	inf	NaN	NaN

No statistically significant difference

TOPIC 5 : DOSE 1

	df	sum_sq	mean_sq	F	PR(>F)
C(vaccine)	3.0	2.075000e-03	0.000692	0.0	NaN
Residual	0.0	4.622232e-32	inf	NaN	NaN

No statistically significant difference

TOPIC 5 : DOSE 2

	df	sum_sq	mean_sq	F	PR(>F)
C(vaccine)	3.0	4.900000e-03	0.001633	0.0	NaN
Residual	0.0	1.232595e-32	inf	NaN	NaN

No statistically significant difference

TOPIC 5 : BOOSTER

	df	sum_sq	mean_sq	F	PR(>F)
C(vaccine)	2.0	2.400000e-03	0.0012	0.0	NaN
Residual	0.0	6.162976e-33	inf	NaN	NaN

No statistically significant difference

ANOVA TO CHECK IF THERE IS ANY DIFFERENCE BETWEEN DOSES OF THE SAME VACCINE FOR THE 5 TOPICS IDENTIFIED

PFIZER : TOPIC 1

	df	sum_sq	mean_sq	F	PR(>F)
C(dose)	2.0	0.001267	0.000633	NaN	NaN
Residual	0.0	0.000000	NaN	NaN	NaN

No statistically significant difference

PFIZER : TOPIC 2

	df	sum_sq	mean_sq	F	PR(>F)
C(dose)	2.0	0.003467	0.001733	NaN	NaN
Residual	0.0	0.000000	NaN	NaN	NaN

No statistically significant difference

PFIZER : TOPIC 3

	df	sum_sq	mean_sq	F	PR(>F)
C(dose)	2.0	5.066667e-03	0.002533	0.0	NaN
Residual	0.0	1.540744e-32	inf	NaN	NaN

No statistically significant difference

PFIZER : TOPIC 4

	df	sum_sq	mean_sq	F	PR(>F)
C(dose)	2.0	1.066667e-03	0.000533	0.0	NaN
Residual	0.0	1.540744e-32	inf	NaN	NaN

No statistically significant difference

PFIZER : TOPIC 5

	df	sum_sq	mean_sq	F	PR(>F)
C(dose)	2.0	8.000000e-04	0.0004	0.0	NaN
Residual	0.0	9.244464e-33	inf	NaN	NaN

No statistically significant difference

ASTRAZENECA : TOPIC 1

	df	sum_sq	mean_sq	F	PR(>F)
C(dose)	2.0	0.002067	0.001033	NaN	NaN
Residual	0.0	0.000000	NaN	NaN	NaN

No statistically significant difference

ASTRAZENECA : TOPIC 2

```
          df        sum_sq   mean_sq    F  PR(>F)
C(dose)   2.0  5.066667e-03  0.002533  0.0     NaN
Residual  0.0  2.465190e-32       inf  NaN     NaN
```

No statistically significant difference

ASTRAZENECA : TOPIC 3

```
          df        sum_sq   mean_sq    F  PR(>F)
C(dose)   2.0  2.466667e-03  0.001233  0.0     NaN
Residual  0.0  1.540744e-32       inf  NaN     NaN
```

No statistically significant difference

ASTRAZENECA : TOPIC 4

```
          df        sum_sq  mean_sq    F  PR(>F)
C(dose)   2.0  2.000000e-04   0.0001  0.0     NaN
Residual  0.0  1.540744e-32      inf  NaN     NaN
```

No statistically significant difference

ASTRAZENECA : TOPIC 5

```
          df  sum_sq  mean_sq   F  PR(>F)
C(dose)   2.0  0.0002   0.0001 NaN     NaN
Residual  0.0  0.0000      NaN NaN     NaN
```

No statistically significant difference

J&J : TOPIC 1

```
          df    sum_sq   mean_sq   F  PR(>F)
C(dose)   2.0  0.000867  0.000433 NaN     NaN
Residual  0.0  0.000000       NaN NaN     NaN
```

No statistically significant difference

J&J : TOPIC 2

```
          df        sum_sq  mean_sq    F  PR(>F)
C(dose)   2.0  4.200000e-03   0.0021  0.0     NaN
Residual  0.0  1.232595e-32      inf  NaN     NaN
```

No statistically significant difference

J&J : TOPIC 3

	df	sum_sq	mean_sq	F	PR(>F)
C(dose)	2.0	2.000000e-04	0.0001	0.0	NaN
Residual	0.0	1.232595e-32	inf	NaN	NaN

No statistically significant difference

J&J : TOPIC 4

	df	sum_sq	mean_sq	F	PR(>F)
C(dose)	2.0	6.666667e-05	0.000033	0.0	NaN
Residual	0.0	3.081488e-33	inf	NaN	NaN

No statistically significant difference

J&J : TOPIC 5

	df	sum_sq	mean_sq	F	PR(>F)
C(dose)	2.0	0.0006	0.0003	NaN	NaN
Residual	0.0	0.0000	NaN	NaN	NaN

No statistically significant difference

NOVAVAX : TOPIC 1

	df	sum_sq	mean_sq	F	PR(>F)
C(dose)	1.0	2.450000e-03	0.00245	0.0	NaN
Residual	0.0	1.232595e-31	inf	NaN	NaN

No statistically significant difference

NOVAVAX : TOPIC 2

	df	sum_sq	mean_sq	F	PR(>F)
C(dose)	1.0	3.200000e-03	0.0032	0.0	NaN
Residual	0.0	6.162976e-32	inf	NaN	NaN

No statistically significant difference

NOVAVAX : TOPIC 3

	df	sum_sq	mean_sq	F	PR(>F)
C(dose)	1.0	4.500000e-04	0.00045	0.0	NaN
Residual	0.0	1.109336e-31	inf	NaN	NaN

No statistically significant difference

NOVAVAX : TOPIC 4

	df	sum_sq	mean_sq	F	PR(>F)
C(dose)	1.0	2.450000e-03	0.00245	0.0	NaN
Residual	0.0	7.703720e-32	inf	NaN	NaN

No statistically significant difference

NOVAVAX : TOPIC 5

	df	sum_sq	mean_sq	F	PR(>F)
C(dose)	1.0	2.450000e-03	0.00245	0.0	NaN
Residual	0.0	3.081488e-32	inf	NaN	NaN

No statistically significant difference