# FINAL EXAM PROJECT

Using nlp and other machine learning algorithms on research articles to gain useful insights about covid-19 and how to improve vaccination efforts.

# EXPLORING THE DATASET OF RESEARCH ARTICLES

This graph shows that SARS-CoV-2 (COVID-19) strain is different from the other two strains. Symptoms like discharge, anxiety and depression are common with COVID-19 but cold is not that common compared to the other two strains.

Distribution of Symptoms for each Coronavirus Strains
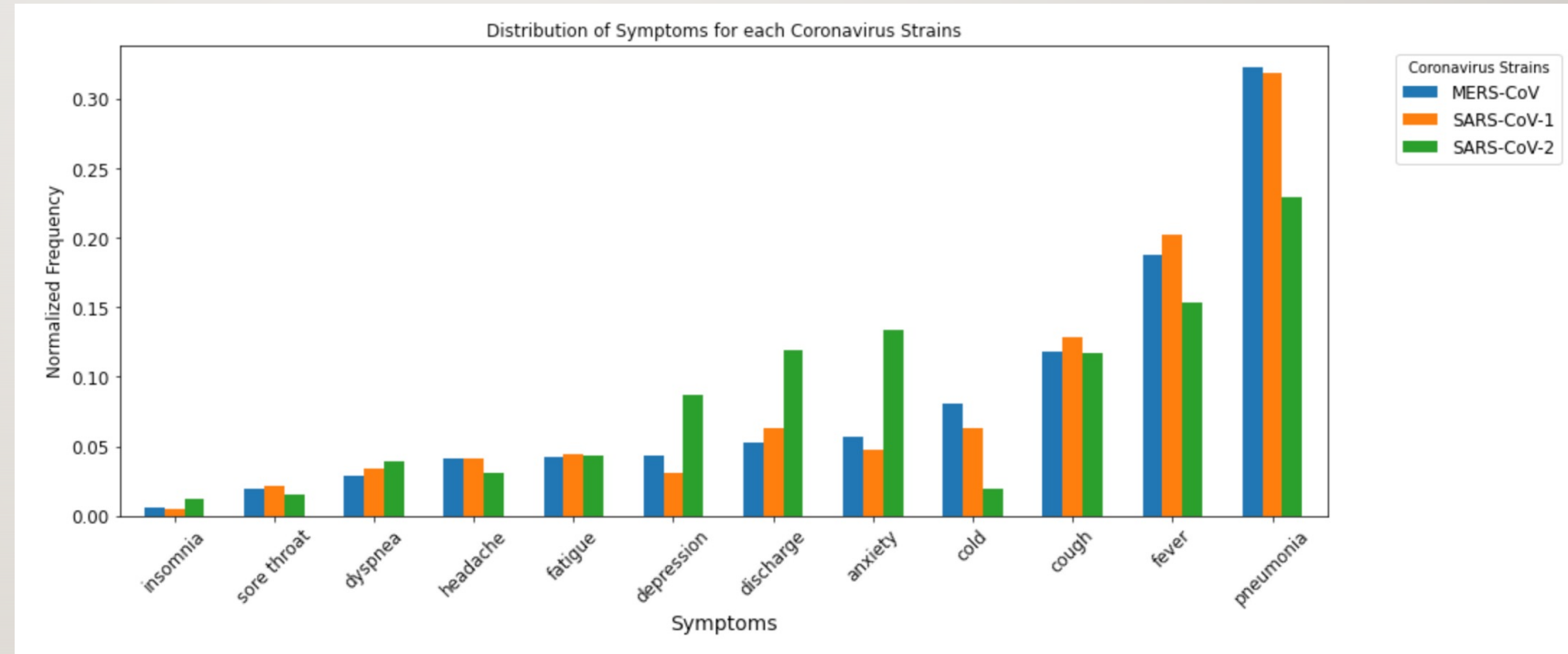
EXPLORATORY DATA ANALYSIS (EDA) CONTINUED…

The high co-occurrence of AstraZeneca vaccine with the three strains of COVID-19 shows that less attention has been given to other COVID-19 vaccines in tackling the spread of the virus.
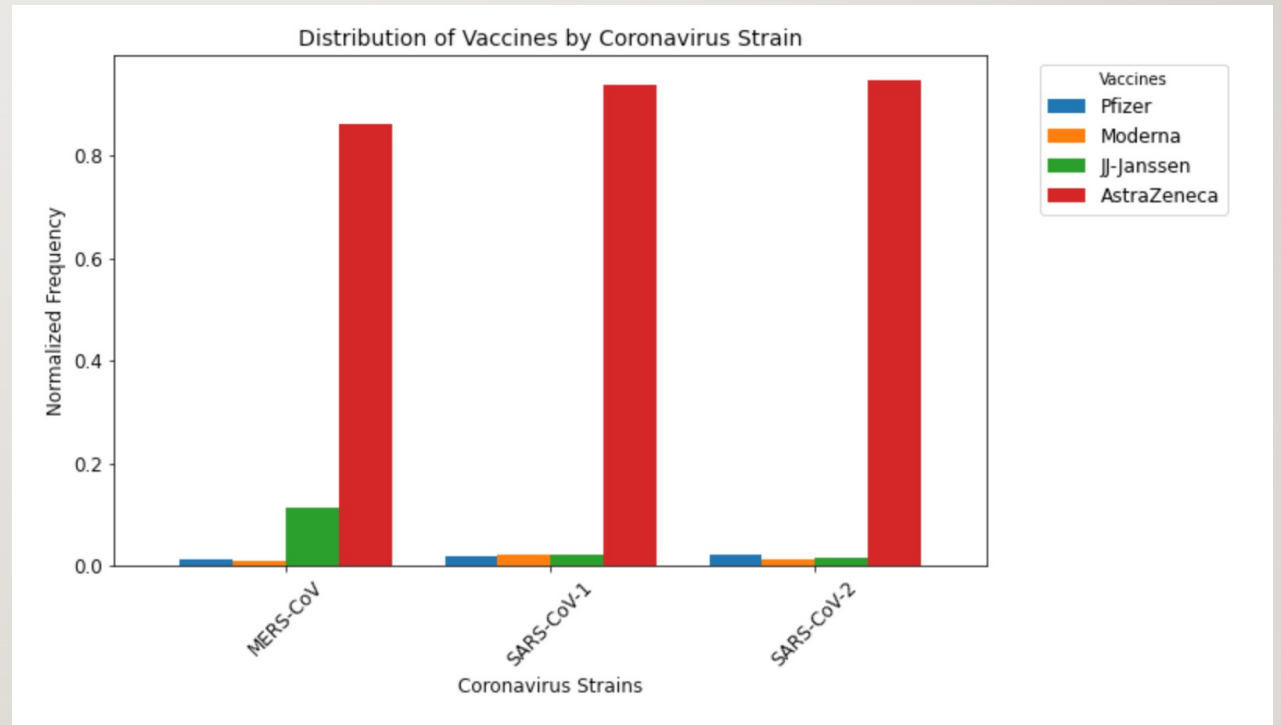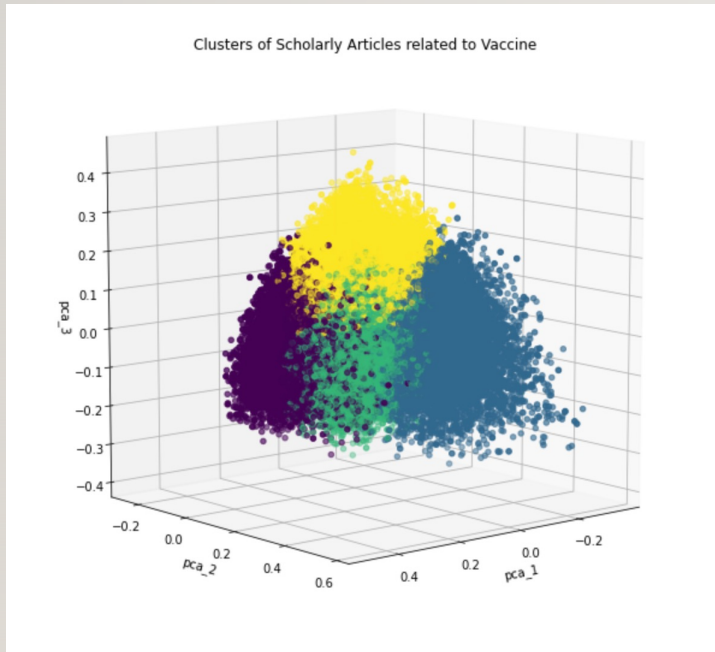
EDA CONTINUED…

AstraZeneca Vaccine might be a good vaccination option to control the spread of Covid-19 as it occur frequently with countries that were able to quickly tackle the spread of the virus.
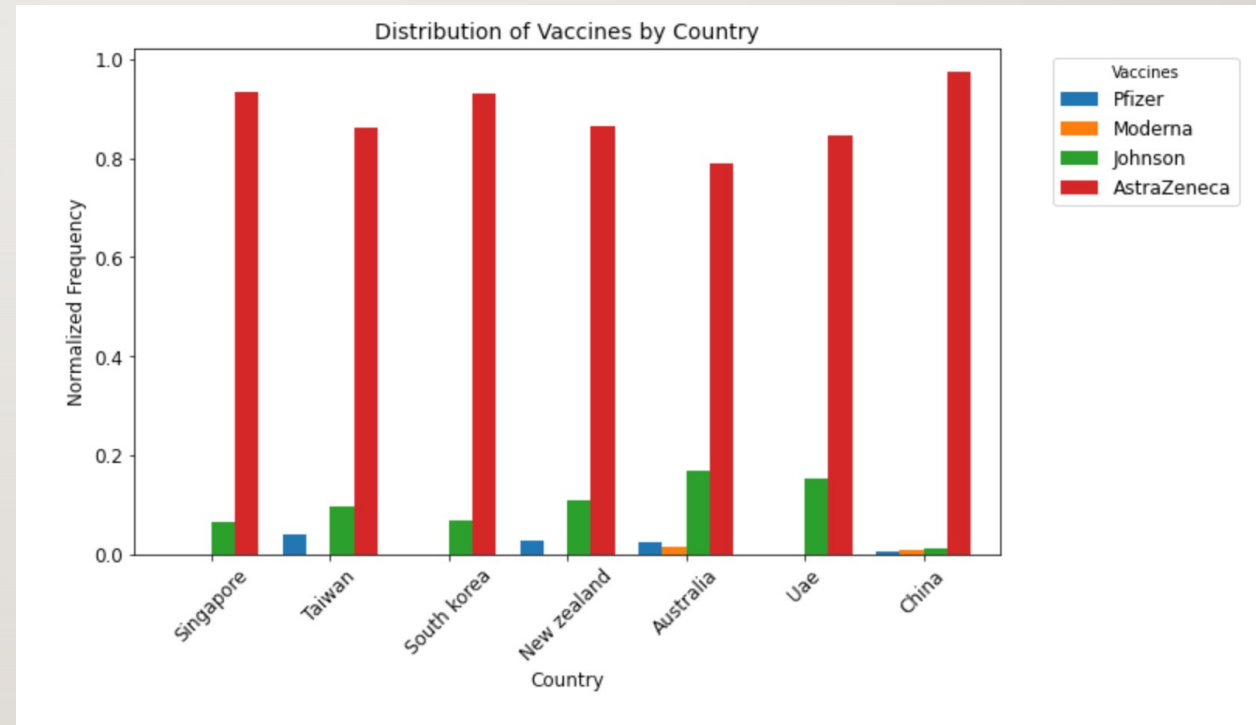
RESULT OF THE MODEL

This 3D plot shows the result of the clusters of vaccine-related articles that was generated by the KMeans model.

This table shows the top 15 topic keywords in cluster 2 obtained with the Latent Dirichlet Allocation (LDA) model. According to the keywords, the articles in this cluster will be useful for studies that are related to how the coronavirus propagates through the cell of humans and roles played by the antibodies.

Clusters of Scholarly Articles related to Vaccine

	Word 0	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10	Word 11	Word 12	Word 13	Word 14
Topic 0	sarscov2	covid19	protein	vaccine	infection	cell	coronavirus	drug	antibody	virus	disease	viral	respiratory	response	patient

RECOMMENDATION FOR HEALTHCARE SYSTEM

- The recommendation for our healthcare system is to revisit the creation of new vaccine putting into consideration only the COVID-19 strain or to explore the already existing vaccine like the Moderna, Pfizer and Johnson & Johnson Vaccine.