BUTTERFLY DATA

www.butterflydata.co.uk/healthcare

COVID VACCINE HESITANCY

Being creative in the use of all available data to better understand the national perception to COVID-19 vaccines in the UK, in order to inform the UK National Health Service of hesitancy trends so that they may better target their health campaigns and improve vaccine uptake.

AT A GLANCE

Challenges

COVID-19 vaccination data was made public on the NHS(E) Azure Environment, including demographics of people who had been vaccinated, vaccination centre codes, and a number of 'no vaccine administered' cases. The data had to be wrangled and cleaned such that there were no biases or false positives.

Tools

- SAS Visual Analytics
- SAS Studio
- Jupyter Notebook
- Azure machine Learning
- Twitter feeds
- Kaggle(open source)

OBJECTIVES

NHS England (NHSE) required a proof of concept showing how AI/ML approaches can be applied to their stored COVID-19 vaccination data sets to deliver data insights to assist the targeting of vaccines to communities that will need it the most.
As a secondary objective, NHS England wanted to understand areas where vaccine hesitancy is high so that greater efforts can be made using community advocacy. The setting, e.g. major center, care home, GP practice.

SOLUTIONS

We worked with NHS(X) and their internal consultants to acquire access to their Azure platform and data via the Home Office Accelerated Collaboration environment. The following metrics were included in the COVID-19 data sets:

- Vaccine usage (doses given/consumption)
- Category of need based on advice and guidance from the Joint Committee on Vaccinations and Immunisation (JCVI)
- Population health needs, employing known risk factors such as ethnicity, age, or an underlying health condition such as diabetes mellitus

We created a model pipeline to combine and cluster the datasets. In order to further investigate these clusters, we used iterative feature engineering to create demographic criteria.

RESULTS

- Jupyter notebooks containing reproducible code were developed to extract and analyse NHS(E) data sets, as well as LITI syntax rules for focused text extraction and analysis in SAS VTA.
- We also produced an audience analysis based on Twitter data that indicated the primary demographic criteria for someone who is COVID-19 reluctant and utilises Twitter to voice their worries. Depending on the grouping, this clustering might then be leveraged for targeted efforts via social media marketing or 1:1 community advocacy.
- Although there was no discernible pattern in ethnicity, gender, or geographical deprivation to predict 'no vaccine uptake,' a link with age and family status was discovered.
- The solution was presented iteratively in weekly sprints, allowing the client to build and modify their requirements as well as raise new discovery questions