JSC270H1 – Assignment 1

Background:

This assignment investigates the COVID-19 Open Data Repository from Google, which contains data on epidemiology and Google search trends around the world during the COVID-19 pandemic. The data have been collected for each day from early 2020 to September 2022, and the sources include Wikidata, the World Bank, and Google. The five specific datasets chosen are from the US states of California, Florida, Massachusetts, Pennsylvania, and Washington.

Motivation:

This assignment aims to determine if the popularity of search terms of common COVID symptoms correlates with the number of tested individuals. The underlying assumption is that as more individuals contact the disease, more would experience various symptoms associated with it. These individuals might be tested for the disease, and/or they might search their symptoms on Google. In addition, the popularity of a search term in Google is based on how many times the term is searched in a certain period (Google, 2023). As such, a first hypothesis could be that as more individuals are tested for the disease, more would also search their symptoms on Google, and thus the search terms for the symptoms become more popular. Thus, the key question of the visualization is whether certain Google search terms corresponding to COVID symptoms become more popular over time as more individuals are tested for the disease.

Methods:

The five datasets chosen to be investigated are the ones collected in the US states of California, Florida, Massachusetts, Pennsylvania, and Washington. They are chosen since data on Google search trends are available in those regions. The datasets have the same feature variables and are concatenated into one single data frame for the visualization.

The search term investigated is ‘fever’, which is a common symptom of COVID (Hollimon, 2020). Relative to other Google search terms during the pandemic, ‘fever’ is more popular: it has a popularity factor fluctuating around 5 compared to 1 or 2 for other terms. Thus, the search term ‘fever’ would potentially be valuable in investigating the key question of the visualization. The feature variable corresponding to this is *search_trends_fever,* which is a continuous numerical variable. The other variable investigated is *new_tested*, which is the number of newly tested individuals in the region in each day and which is a discrete numerical variable.

Since data on newly tested individuals have not been collected for the five states after March 7, 2021, the visualization only includes data from before that date.

Conclusion:

Overall, based on the visualization, there is likely to be little correlation between the number of newly tested individuals and the popularity of the search term ‘fever’. The spikes in popularity for ‘fever’ occurred in early 2020 for all five states, during which time the number of new tests were relatively low. Importantly, the number of newly tested individuals peaked at various points in late 2020, when the relative popularity of the search term declined and plateaued. Thus, contrary to the hypothesis, the number of tests and the popularity of ‘fever’ is unlikely to correlate, according to the visualization.

This counterintuitive result may be explained by several factors. First, when news of COVID first emerged in early 2020, many individuals became anxious of contacting the disease. As a result, they could have been more proactive in searching up potential symptoms of COVID, even if they did not experience them. Second, as the pandemic continued, this panic subsided as more people became familiar with the concept of COVID, so that even if a person did contact the disease, they were less interested in confirming their symptoms by Google. This likely resulted in the decrease of the popularity of ‘fever’ as the pandemic continued and more individuals were tested.

References

COVID-19 and your health. (2022, October 26). Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html

FAQ about Google trends data. (2023). Google Help. https://support.google.com/trends/answer/4365533?hl=en

Hollimon, N. (2020, March 18). *Signs of coronavirus (COVID-19)*. WebMD. https://www.webmd.com/covid/covid-19-symptoms