Overview

This study examines the correlation between Google searches for terms that may relate to COVID-19 cases, and actual COVID-19 cases within Canada. The study hypothesis is that Google searches would correlate to COVID-19 cases. For example, search for terms such as "cough" and "loss of smell" were examined for correlation with new COVID-19 case counts.

Data Sources

Google Trends[1] provides access to normalized trend data representing a sample of actual search requests made to Google.[2] The pytrends library for Python[3] was used to access the Google Trends API. Given keywords, a geography, and a timeframe, it provides search trends on a weekly basis. Our data can be found [here](#).

This study obtained COVID-19 case data for Canada from the Esri Canada COVID-19 Data Repository, using the ["Provincial Daily Totals" dataset](#)[4]. It provides a compilation of several data sources that include new daily case counts by geography ("DailyTotals"). The instance we used can be found [here](#).

Analysis

In this study, the COVID-19 and Google search trends data sets were plotted over time to help form an early picture of potential relationships and gather insights to focus further analyses. It was observed that "loss of smell" and "loss of taste" trend data grows at the same time as COVID-19 cases spiked in March, suggesting correlation.

This study explored correlation in three ways. The first was to seek correlation between search trends and the COVID-19 case counts within weekly intervals. The second was to seek correlation between search trends and case counts with a one and two-week lag. The last was to seek correlation between search trends and the rate of change.

In all cases, because of the variance between provinces, correlation was sought at a national level and the provincial and territory level. Before the analysis, the approach of seeking correlation with a lag was thought to be the most likely to show the strongest correlation.

[1] https://trends.google.com/trends/?geo=CA
[2] https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052
[3] https://pypi.org/project/pytrends
[4] https://resources-covid19canada.hub.arcgis.com/datasets/provincial-daily-totals

When correlation was calculated between search trends and cases within the same week, we found that there is correlation between searches for "loss of smell", "loss of taste", and "sleepy", and COVID-19 cases on a weekly basis in some cases. The lack of consistency across geographies do not enable a model that is reliable across data sets.

Higher correlation coefficients were found for Canada as a whole with a 2-week lag. The same terms were most correlated to COVID-19 cases, with Spearman coefficients of 0.71, 0.62, and 0.42 respectively. The R-squared value for an OLS regression for Canada using those terms was 0.539. These results support a moderate correlation of those search terms to COVID-19 cases in Canada, but the correlation is inconsistent across provinces.

Additional analysis was conducted after calculating the change in COVID-19 cases from week to week for each province. Normalizing values using rate of change sought to address the wide range of daily COVID-19 cases in the dataset.

Examining the correlation coefficients, significant correlation (.70 - .75) was found using the Pearson method for 'Fever', 'Cough', 'Shortness of Breath' and 'Sore throat' for Canada, and comparable values were found in some of the provinces and territories. The R-squared for the OLS regression using these terms for Canada is 0.574. These results support a moderate to strong correlation of Google searches for those terms with the rate of change of COVID-19 cases. Surprisingly, searches for these terms correlated more directly to the rate of change of COVID-19 cases than the terms that correlated to the actual case counts.

Introducing lag to the exploration of the rate of change showed that the correlation weakened with one week of lag, and weakened further with a two-week lag.

Conclusions

The study finds that there is some correlation between Google searches and COVID-19 cases, and the rate of change of COVID-19 cases at the national level.

Correlation was found between the search terms "loss of smell", "loss of taste", and "sleepy", and COVID-19 cases in Canada. The strongest correlation was found with a two-week lag between search and COVID-19 cases. The correlation is best described as moderate, and there was variability between provinces and territories over the same period.

Correlation was found between the search terms "Fever","'Cough", "Shortness of Breath" and "Sore throat" and the rate of change of COVID-19 cases. It is interesting that the terms are different than those that correlate to the case counts.

While correlation was found in this study, additional exploration may reveal stronger correlation and enable reliable models for prediction.