

Assignment 1 Report

Visualization of Google Search Patterns related to COVID-19 vaccination

Part 1: Background

The anonymized “COVID-19 Vaccine Search Insights” dataset aggregates data from Google search data and Google Map. It contains weekly information about Google **search interest** related to **COVID-19 vaccination** in different regions in several countries (but I only select data from Canada), from Jan 2021 to present. All searches are divided into 3 categories:

1. **All COVID-19 vaccination**; hence, it includes the 2 categories below.
2. Searches that show users’ **intent** for being vaccinated, i.e. searches related to availability, eligibility of vaccines
3. **Safety and side effects**

For each category, the value of search interest is a relative interest as opposed to an absolute one, enabling researchers to compare search interest across different regions, or same region over time. This dataset is used in real life to help public health officials target and design vaccination education campaigns.

Part 2: Motivation

My visualization is aimed at providing insight into **where and when** to invest more health care resources:

1. By presenting COVID-19 vaccination search interest of each search category for different provinces over time, public health organizations can compare search interest across different regions, and decide **in what provinces** they should focus on targeting vaccination campaigns.
2. Also, if combined with other dataset (e.g. financial situation, COVID-19 daily new cases) for different provinces, one may predict the search trends, and decide **when** to put more effort in advertising vaccination campaigns.

Part 3: Methods

1. **Data Cleaning:**

- **Only data from Canada is selected**, so as to focus on search interest across provinces in Canada.
- Also, there are 3 null values and are all **replaced by mean value of that column (search category) before 2021-02-08**, because a) missing values are all before this date, b) search interest in all regions before 2021-02-08 is relatively flat and close, so it is reasonable to assume the missing data is close to other data before this date.

2. Visualization:

A **multiple line graph** is used to show the COVID-19 vaccination search interest in different provinces of Canada from Jan 2021 to present. **Two widgets** are added: one for selecting one of the three search categories, and the other one for selecting one or more regions (one can press command/shift to select multiple regions). The graph type makes it much easier to **analyze search interest in a region over time**. Together with the widgets, it is much convenient and intuitive to **compare search interest across different provinces** in Canada.

Part 4: Conclusion

For all provinces, the general search interest and intent search interest **fluctuates** over time. Roughly speaking, all provinces peak at close time. As all provinces react in a similar way, this possibly indicates nation-wide COVID-19 events/policies are drawing all Canadians attention. If combining with other policy/event data, the government may gain insight into **people's (including people who don't publicly speak up) reactions towards COVID-19 situations/policies**.

On the contrary, the safety search interest only has a high peak at April 2021 - June 2021, and it generally **declines** afterwards. Hence, we may draw the conclusion that **Canadians may not be as worried as vaccines first released, and may have more intent to get vaccinated over time**. However, high intent search may not mean people are actually interested in getting vaccinated, so in order to draw a precise conclusion, one should combine with actual vaccination data and fully analyze what intent search really indicates.

As for comparing across provinces, **Alberta has generally lower search interest in intent search and high interest in safety search, so government may need to focus on Alberta**. However, since no province has extremely low search interest, we should be careful if Alberta is reluctant to get vaccinated.

Further improvement: As mentioned above, combine this dataset with other dataset (such as COVID-19 daily cases, policy releasing date) to draw more precise conclusions.