Leveraging Place Reviews to Identify the Effects of COVID-19 on Canadian Tourism

Grant McKenzie

Platial Analysis Lab, McGill University, Montréal, Quebec, Canada

The emergence of the COVID-19 pandemic disrupted travel world-wide and substantially impacted tourism in most countries. Though many governmental agencies and tourism boards have published data on the impact of the pandemic, in Canada, the vast majority of these data are reported at the national level or sparsely within individual regions. In this preliminary work, we leverage user-contributed tourist attraction reviews to better understand the nuanced changes in travel behavior resulting from the COVID-19 pandemic. We examine the regional impacts as well as the effects on different categories of tourism within Canada. The purpose of this short paper is to demonstrate the value of analyzing place-based user-generated tourism data and highlight some of the ways it can be leveraged by policy experts and tourism agencies.

Keywords: Tourism, COVID-19, Tripadvisor, Canada, User-generated content, Travel

1 Introduction

In 2019, Canada's tourism sector generated an estimated $104.4 billion dollars in revenue, contributing roughly 2% of the country's gross domestic product (GDP) (Transport Canada, 2021). In late 2019, the SARS-CoV-2 virus emerged and by early 2020, the COVID-19 pandemic was disrupting virtually every sector of the Canadian economy. By the end of 2020, tourism in Canada had dropped by 48.1% with the tourism contribution to the GDP falling to just 1.1% (Statistics Canada, 2021). Canadians were not alone in this experience. Many countries were similarly affected as international borders were closed, businesses shuttered, and lock-down policies enacted.

While COVID-19's impact on tourism has been well documented at a national level, data and spatial analysis comparing the effects of the pandemic *between* sub-national regions in Canada has been limited. For instance, Nhamo et al. (2020) investigated the impact of travel advisories related to COVID-19, but only at the national level whereas Lapointe (2020) discussed the pandemic's effects on tourism in the province of Québec. In fact, much of the research in this domain has focused either at the global level (Gössling et al., 2020), nationally (Liu, 2020), or within individual subnational regions (Roy et al., 2021). Similarly, little research has comparatively investigated the pandemic's impact on categories of tourism (e.g., Transportation, Shopping, Sight-seeing, etc.). Some studies have highlighted opportunities for sub-sectors of the tourism industry to reduce the impacts (Dube et al., 2021; Sigala, 2020), but little comparative analysis has been done. Much of this is likely owing to the fact these types of data are often collected by local tourism boards, municipal government agencies, or commercial companies. The heterogeneity, varied-resolutions, and accessibility of these data, make it difficult to compare across regions and tourism categories.

To address these limitations and investigate the effects of the COVID-19 pandemic on Canadian tourism at a higher spatial and categorical resolution, we leverage tourism review data contributed

to the travel platform, TripAdvisor (TA). TripAdvisor is an American-based online travel platform that operates globally. It's primary draw is that visitor's to physical tourist attractions contribute reviews, photos, and recommendations through the online platform, allowing readers to make informed decisions on where to travel. The company claims to have more than 859 million reviews on 8.6 million establishments world-wide (TripAdvisor, 2021). These data have been used in previous spatial analyses to identify similarities between tourist attractions (McKenzie and Adams, 2018), understand individual hotel preferences (O'connor, 2008), and for sentiment analysis tasks (Valdivia et al., 2017). These data offer an unprecedented opportunity to understand tourism behavior, and specifically investigate how this behavior changed with the emergence of COVID-19. With this in mind, we aim to address the following research questions (RQ).

RQ1 *Tourism in which provinces has been most affected by the COVID-19 pandemic? Furthermore, visits from which country were most impacted by the pandemic?*

RQ2 *Which categories of tourist attractions were most affected by the pandemic? In addition, did these impacts vary by province?*

While the primary goal of this work is to address both of the research questions above, this short paper is a report on our preliminary analysis. Our intention is that this whets the reader's appetite, facilitates discussions, and sparks ideas for future work. Our secondary goal, is to briefly highlight the value of place-based user-generated content and demonstrate some of the analytical possibilities in working with tourism reviews.

2   Data & Analysis

In October 2020, attractions (places of interest) and reviews were accessed through the TA web platform. In total, 19,831 attractions from across Canada and 1,309,204 reviewers were accessed contributed by 479,062 unique users. Each attraction consists of a name, geographic coordinates, category label, and list of reviews. Each review consists of a unique user identifier, the location of the reviewer, the date of the review, the date of experience, and the review text itself. Reviews were accessed for the English and French version of the platform (TripAdvisor.com & TripAdvisor.fr). All reviews between January 2015 and September 2020 were used in this analysis. The vast majority of reviews report temporal resolution at the level of month and year. We use *date of experience* when available and *date of review* in the 2.7% of reviews where date of experience was not provided. A reviewer's location is manually entered by the review contributor. The means by which this information is entered changed over time with early users offered free-text entry and recent users being limited to selecting from a predefined set of places around the world. Google's Geolocation API was used to successfully identify the home country of 79.3% of users. A total of 213 unique home countries or territories were identified.

2.1 Reviews as a proxy for visits

In this work, we argue for using reviews as a proxy for physical visits. We acknowledge that most tourists do not post on TripAdvisor but make the assumption that on aggregate, the count distribution of reviews mirrors the distribution of physical visits. In analyzing our data, the volume of reviews between $2015 - 2019$ were averaged by month to produce a robust *baseline* on which to compare the first nine months of 2020 data. Figure 1a shows a comparison of the monthly review contributions nationally split into 2020 and pre-2020 (4 year average). Error bars show standard error. At the national level we clearly see that 2020 visits experienced a significant decline. Up to September, review volume in 2020 was 18.7% of the average review volume over the previous 5 years with May 2020 being just 4.4% of the previous May average.

To support our argument that user-contributed reviews are a suitable proxy for tourist visits, we compared our review counts to the national-level, international traveler data as reported by Statistics Canada. Reviewers that listed a home location in Canada were removed, since Statistics Canada does not report this information. Furthermore, only travel data between January-August were reported from Statistics Canada, therefore our review data were restricted to the same period. Figures 1b and 1c show a comparison between 2020, and pre-2020 travelers to Canada and the same two time periods for TA reviewers. Notably, the data from Statistics Canada represent people enter the country for a variety of reasons, only one of which is tourism.
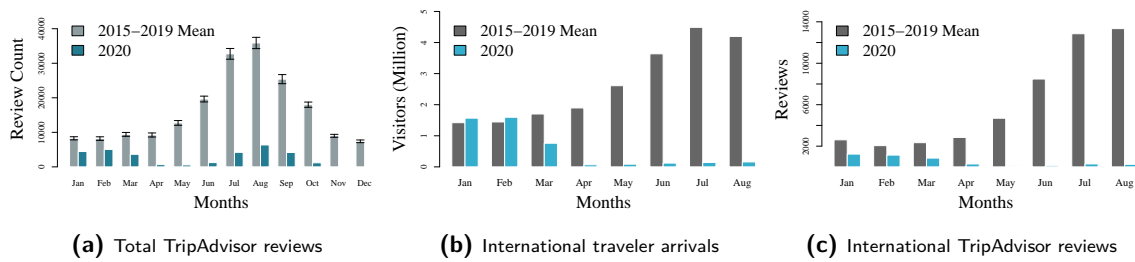
(a) Total TripAdvisor reviews

(b) International traveler arrivals

(c) International TripAdvisor reviews

Figure 1: (a) shows total user-generated reviews contributed to Canadian attractions on the TripAdvisor platform, split by month. (b & c) show a comparison of international traveler data from Statistics Canada to international TripAdvisor reviews over the same time period.

We subtract the yearly distributions from one another within each dataset to produce a *difference distribution* for each. We then calculated the similarity between the two difference distributions using both the *Jensen-Shannon distance* and *cosine similarity*. The results were 0.0397 (0 is identical) and 0.9725 (1 is identical), respectively indicating a high degree of similarity over the 8 month time period. This suggests that the TA data does reflect actual international travel mobility.

To further demonstrate the validity and value of these TA data, we compared review data to visitor data at a city level. Destination BC (2021) reported the volume of visitors to the city of Vancouver per month by country of origin for the same time period as our TA data (2015-2020). While not all countries were reported, we aggregated visitors into three groups, namely Canada, United States, and Other. We compared the averaged 2015-2019 Destination BC data to the same averaged time period of TA review volume for all attractions in Vancouver and did the same for the 2020 datasets. As shown in Figure 2, there is a high degree of similarity between the two datasets. Notably, Destination BC reported a higher percentage of Canadian visitors as compared to TA reviews in 2019, with TA reporting a higher percentage of international reviews. This is likely due to the fact that Destination BC primarily reported on visitors, not tourists and therefore included those traveling to visit family, for business, etc. By comparison, TA exclusively reports tourist activity.
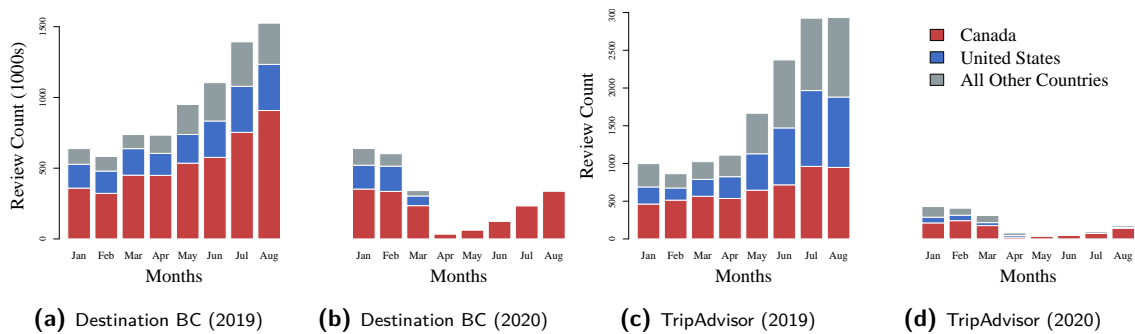

(a) Destination BC (2019)

(b) Destination BC (2020)

(c) TripAdvisor (2019)

(d) TripAdvisor (2020)

Figure 2: Comparing visitor and review volume by country of origin in Vancouver, British Columbia, split by year. Note that only data for January-August 2020 were reported by Destination BC at time of analysis.

2.2 Regional variations

Though national level travel and tourism data is published by Statistics Canada, these data are not available at the provincial level, nor do they report the home country of the tourist. Our TA data, however, allows us to analyze review volume at varying spatial resolutions as well as by the home country of the reviewer. We grouped all attractions by their province and then calculated the average number of reviews for each province for 2015-2019 and 2020 independently. We then calculated the percentage change in number of reviews split by province (Table 1). Reviews contributed by *All* visitors are shown as well as being split by the top home countries of reviewers. Notably, data on Canada's northern territories of Nunavut, Yukon, and Northwest Territories are not reported due to lack of data.

In a typical year, tourists visiting the provinces of Ontario, British Columbia, and Québec account for roughly 75% of all tourism in Canada. On average, these provinces reported a 78% decrease in

Table 1: Percentage change in number of TripAdvisor attraction reviews between 2015-2019 (average) and 2020, split by province of attractions and home country of reviewer.

Province	All	Canada	USA	UK	Australia	France	Others
Alberta	-77.6%	32.2%	-16.9%	-8.5%	-4.8%	-1.1%	-0.9%
British Columbia	-78.3%	28.4%	-14.7%	-6.8%	-2.9%	-0.8%	-3.2%
New Brunswick	-84.7%	23.6%	-20.5%	0.7%	0.2%	-2.2%	-1.8%
Newfoundland and Labrador	-89.8%	15.1%	-6.4%	-3.8%	-1.6%	-0.3%	-3.0%
Nova Scotia	-86.1%	26.5%	-18.6%	-5.3%	-0.3%	-0.5%	-1.7%
Manitoba	-72.1%	12.8%	-8.6%	-2.8%	-0.5%	-0.3%	-0.6%
Ontario	-76.3%	21.8%	-14.5%	-4.0%	-0.7%	-1.0%	-1.7%
Prince Edward Island	-91.6%	13.6%	-10.0%	-2.5%	0.2%	-0.2%	-1.2%
Quebec	-80.7%	30.0%	-16.9%	-2.9%	-1.1%	-4.9%	-4.1%
Saskatchewan	-77.7%	6.9%	-6.8%	-1.3%	0.9%	-0.1%	0.5%

* Values for Yukon, Nunavut, and Northwest Territories are not reported due to data sparsity.

tourism. While the smallest province, Prince Edward Island, contributes a much smaller percentage to the Canadian GDP, tourism accounted for roughly 3.5% of its local GDP pre-pandemic (Statistics Canada, 2021). The province experienced the largest decrease of all provinces with one of the smallest increases in percentage of tourism attributed to domestic travel. This can partially be attributed to the establishment of the *Atlantic Bubble* where the Atlantic provinces barred entry to visitors from all other provinces . Alberta and Québec saw the largest increases in percentage of domestic tourism with Québec seeing a notable decrease in visitors from France. Alberta and British Columbia saw their largest non-U.S. decreases from the United Kingdom and Australia. The percentage of visitors from the United States, the largest source of international tourists, decreased substantially across all provinces.

2.3 Category of attraction variations

We were also able to use the review data from TripAdvisor to identify changes in tourism behavior towards different categories of tourist attractions. The number of reviews were counted and split by the category of the attraction to which they were contributed. This, again, was done for 2015-2019 and 2020 independently. We then calculated the percentage change in number of reviews between the two datasets. The results are shown in Figure 3. The category *Other* showed the largest decrease with *Transportation*, *Sights & Landmarks*, and *Museums* showing similarly large decreases, all above 80%. The category *Spas & Wellness* presented the smallest decrease of 35.0% followed by *Outdoor Activities* (48.5%) and *Food & Drink* (58.4%).
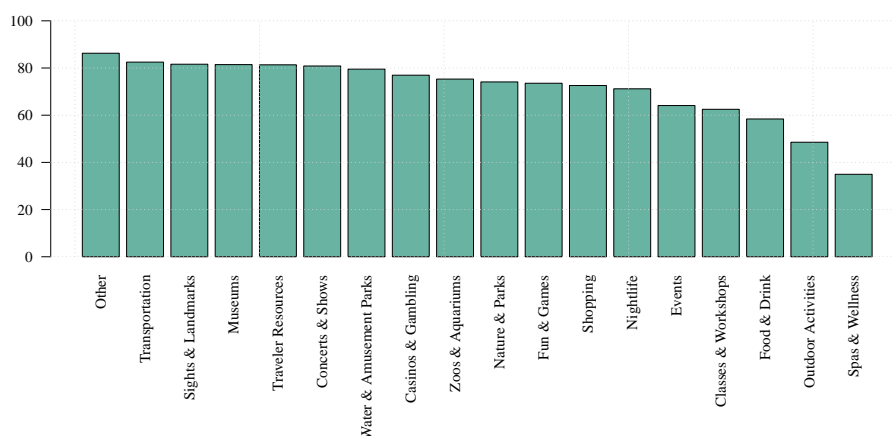

Figure 3: Percentage change in number of TripAdvisor attraction reviews between 2019 and 2020, split by category of attraction.

Exploring this data by province we found that *Outdoor Activities* had the largest variance in percentage change across all provinces with Prince Edward Island seeing the largest change (96.8%) and Saskatchewan experiencing the lowest (12.5%). *Nightlife* was another notable category with

decreases of over 85% in Ontario and Nova Scotia but under 50% for Québec and British Columbia. Categories that showed the highest percentage decreases such as *Transportation*, and *Museums* also showed the lowest variance across provinces.

3   Discussion

In this short paper we highlight the value of place-based user-generated content in understanding nuanced regional and categorical impacts due to the COVID-19 pandemic. This initial analysis demonstrates that tourism reviews can be used (within reason) in place of authoritative data when such data is not available at the required resolution. There are a number of limitations to using such data. For instance, there is a significant bias towards affluent and English-speaking populations. Since TripAdvisor is a U.S.-based company, it's largest user base is also in the United States with many other Western nations contributing large amounts of reviews. While contributions from inhabitants of smaller, developing, and non-English speaking nations are present on the platform, they are undoubtedly under-represented. Fake contributions are also a concern in analysis of any user-generated content and while efforts are made to reduce these contributions, their presence must be acknowledged.

In this preliminary work, we did not discuss the political or social factors contributing to a change in tourism behavior during this time period. For instance, provincial and municipal lock-downs, business closures, and travel advisories all varied in severity and temporal duration. Fear of virus transmission, case counts, and media coverage also play an important and complex role in tourism and travel behavior. All of these aspects, while significant, were outside the current scope of this paper and will be investigated in future work. Next steps will also involve comparative analysis with other countries and datasets. Our overarching goal is to leverage these types of user-generated data to provide actionable insight to policy experts, tourism boards, and public health agencies.

References

Destination BC: *International Visitor Arrivals | Destination BC*, 2021. URL https://www.destinationbc.ca/research-insights/type/industry-performance/international-visitor-arrivals, [Online; accessed 20. Oct. 2021]

Dube, Kaitano; Nhamo, Godwell; and Chikodzi, David: *Covid-19 cripples global restaurant and hospitality industry*. Current Issues in Tourism, 24(11), 2021, 1487–1490

Gössling, Stefan; Scott, Daniel; and Hall, C Michael: *Pandemics, tourism and global change: a rapid assessment of covid-19*. Journal of Sustainable Tourism, 29(1), 2020, 1–20

Lapointe, Dominic: *Reconnecting tourism after covid-19: The paradox of alterity in tourism areas*. Tourism Geographies, 22(3), 2020, 633–638

Liu, Huju: *The Economic Impact of Travel Restrictions on the Canadian Economy due to the COVID-19 Pandemic*. Statistics Canada= Statistique Canada, 2020

McKenzie, Grant and Adams, Benjamin: *A data-driven approach to exploring similarities of tourist attractions through online reviews*. Journal of Location Based Services, 12(2), 2018, 94–118

Nhamo, Godwell; Dube, Kaitano; and Chikodzi, David: *Counting the cost of COVID-19 on the global tourism industry*. Springer, 2020

O'connor, Peter: *User-generated content and travel: A case study on tripadvisor. com*. In: *ENTER*. 2008, vol. 2008, 47–58

Roy, Hiran; Gupta, Vikas; Faroque, Anisur R; and Patel, Alpa: *The impact of covid-19 on the foodservice industry in vancouver, british columbia, canada*. Anatolia, 32(1), 2021, 157–160

Sigala, Marianna: *Tourism and covid-19: Impacts and implications for advancing and resetting industry and research*. Journal of business research, 117, 2020, 312–321

Statistics Canada: *The Daily — Provincial and territorial tourism satellite account, 2017*, 2021. URL https://www150.statcan.gc.ca/n1/daily-quotidien/210429/dq210429b-eng.htm, [Online; accessed 19. Oct. 2021]

Transport Canada: *The Government of Canada supporting Prince Edward Island's tourism sector*. Government of Canada, 2021. URL https://www.canada.ca/en/transport-canada/news/2021/07/the-government-of-canada-supporting-prince-edward-islands-tourism-sector.html

TripAdvisor: *IN Press Center | About TripAdvisor*, 2021. URL https://tripadvisor.mediaroom.com/IN-about-us

Valdivia, Ana; Luzón, M Victoria; and Herrera, Francisco: *Sentiment analysis in tripadvisor*. IEEE Intelligent Systems, 32(4), 2017, 72–77