How International Travel Has Been Affected By COVID-19

Taeyoung Eun (301365637)

Kenta Horie (301374650)

CMPT 353 D100 (Steven Bergner)

**Introduction**

COVID-19 affected everyone around the world in different ways. It has critically damaged various businesses and industries. Travel industry is one of the biggest victims of COVID-19 pandemic, for the harsh travel restrictions and the fear of catching COVID-19 has prevented or deterred international travel for many people. We will take a closer look on its impact. We will investigate how many fewer people traveled in and out of Canada due to COVID-19, whether or not the government set deterrents had an effect in the number of travelers, and how much more people will travel as the COVID-19 pandemic situation improves.

**Experimental Setup / Hypothesis**

We acquired our dataset from Statistics Canada website. It included how many people travelled every month, how they travelled, and their destinations/source (United States, Canada, other). We also used a COVID-19 intervention timeline in Canada to see one of the major deterrents of international travelers: The travel restrictions. We converted the dataset from Statistics Canada from pdf to a csv file, and the COVID-19 intervention timeline was provided as a csv. We used an external application called Able2Extract to convert the pdf file from Statistics Canada to a csv.

We changed the index and various fields of the pdf file, so that it can be readable by pandas in python and to be further cleaned and analyzed. The csv files were re-organized and separated to different files. Fields such as dates had to be changed to number of days from the start of the recording because dates directly could not be used for the machine learning techniques used.

The data set for this project has been separated into to two: one with pre-covid 19 data (2019, 2020, 2021), and one without. (2020 May - 2021). People could find the 2019 data will be the “normal values,” and therefore valuable to add to the dataset. However, it can also have a terrible effect in the dataset because it is like including a year’s worth of outlier values. We divided the data to: people travelling from Canada, people travelling from United states, anyone travelling from anywhere other than those two countries, and all the travelers together.

We hypothesized that the dataset without pre-covid 19 data will perform much better in the statistical test because the data from 2019 are so different from every single year. We also predict that having the pre-covid 19 data will cause the machine learning techniques to create extremely high values that will not make logical sense. In comparison, not accounting for the pre-covid 19 data will create a prediction of steady but slow growth which will be a more believable prediction.

**Results**

The figure 1 shows just how much effect COVID-19 had on travel. The number of people travelling sharply drops from January of 2020, and it stays at an incredibly low number for a very long time. You can see that it sees eventual climb back in 2021, mainly with climbing vaccination rates and loosening of the travel restrictions.

The dataset without the pre covid 19 data performs much better on normal tests and show a much more understandable trend without any background knowledge. Even when we predict the rates of travel for 2022 using linear regression, the dataset without pre covid 19 makes fewer radical predictions. The dataset with pre covid 19 data makes these predictions because it is aware of the change in travelers as the pandemic hit, and it accounts for the possible radical change. The dataset without covid 19 assumes bases its predictions on the slowly appearing growing rates, so it makes a steadier and trendier prediction.

Surprisingly, when performing the normal test with the new predicted values, the dataset with pre covid 19 data produced smaller values in some of the categories. This is likely the result of overall high variance of the dataset with the pre-covid 19 values. Data with huge gaps in between likely allowed the spectrum of values considered “normal” to be larger.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | 2019 – 2021  p-value | 2020 - 2021  p-value | 2019 - 2022 (prediction)  p-value | 2020 -2022 (prediction)  p-value |
| Canada | 3.6504 \* 10 ^  -13 | 0.0012 | 0.0020 | 0.0017 |
| US | 0.0144 | 3.2345 \* 10 ^ -5 | 0.0001 | 0.0514 |
| Other | 0.0092 | 2.0041 \* 10 ^ -6 | 5.0063 \* 10 ^ -5 | 0.0141 |
| Total | 0.0051 | 0.0002 | 0.1744 | 0.1488 |

**Discussion**

**Conclusion**

**References**

https://www150.statcan.gc.ca/n1/pub/66-001-p/2021009/tbl/tbl-1-eng.htm

https://www.cihi.ca/en/covid-19-intervention-timeline-in-canada