Effects of COVID-19 On Different Industries in Canada

Project Report

Emmanuella Bamgbose, Victor Nzemeke, Daniel Zhou,
Spencer Dow, Harman Kaur, Yves Nsoga

December 10, 2021

Table of Contents

Introduction and Project Description

The COVID-19 pandemic has drastically changed our way of life over the past two years. In this project we aim to compare how the pandemic has affected some of the industries in Canada. Many industries have suffered while others flourished during the pandemic. We will be examining the Tourism, Electricity, Retail, Trade, Manufacturing, Retail, Air travel, and Ground travel domains to see if there are interrelated trends between the industries as well as to the severity of pandemic. From this project we hope to provide an observation of trends that occur within the industries during the pandemic and see if in the current data there any signs of a return to normal. From analysing these trends, we hope to provide a greater understanding of the effect the pandemic has had on these important industries within Canada. Each group member is assigned their own domain and research questions that tie into our main COVID-19 dataset, these domains, research questions, and assigned group members are listed in the Research Questions section.

Research Questions by Domain

1. Tourism (Emmanuella Bamgbose)
    - What is the impact of Covid on expenditure for tourism demand? How is international and domestic tourism expenditure also affected?
    - Which category/industry was most hit by the effect of covid on tourism?
    - What is the effect of covid in employment generated by tourism?
    - Which category/industry experienced the most effect on employment?
    - Has there been any improvement in tourism expenditure and employment?
2. Air Travel (Victor Nzemeke)
    - What is the effect of Covid on Air Travel in Canada?
    - What is the effect of Vaccination on Aviation in Canada?
3. Ground Travel (Daniel Zhou)
    - For ground travel, we want to examine the effect of COVID-19 pandemic on international ground travel at the Canadian border, our research questions are:
        - Did COVID-19 affect the number of vehicles crossing the border between Canada and United States?
        - If so, how does the effect on international ground travel differ at a provincial level?
        - Has vaccine helped with international ground travel numbers recovery?
        - Do different bordering countries affect ground travel impact and recovery differently?
4. Electricity (Spencer Dow)
    - In this domain we hope to answer the following two research questions
        - Did Covid had a beneficial effect on our environment through the reduction of electricity generation?
        - Did the pandemic produce a change that impacted the source used to generate electricity?
5. Manufacturing (Harman Kaur)
    - Manufacturing is one of the major sectors contributing to the economy of the country. It provides a significant source of demand for services in the employment and trade sector.

Export of goods and services gives a lot of opportunity to the domestic companies to produce and manufacture more. We wish to understand how the pandemic has affected the manufacturing sector across the country. Few guiding questions that we wish to answer are given below:

- How were the manufacturing sales impacted during pandemic?
- Did manufacturing sector affected the international merchandise trade during the pandemic?
- What was the impact of manufacturing sector on the employment during pandemic?

6. Retail (Yves Nsoga)
    - For the impact of covid on retails we investigated the following research questions:
        - Which industries experienced growth during the pandemic and which industries experienced a drop in revenues?
        - What are the major changes in consumption habits brought by the pandemic?
        - What is the correlation between COVID'19 pandemic, and consumers habits?

Datasets / Data cleaning

The main dataset we will be using for connecting all the different sectors is a Canada COVID-19 dataset, below are the details for that dataset:

- COVID-19 (Canada)
    - The dataset we used for this is the Esri Canada COVID-19 Provincial Daily Totals dataset (Esri Canada, 2021a) found on Esri Canada's website.
    - It is an Open Dataset (Esri Canada, 2021b) and includes COVID-19 data from all the Canadian provinces and territories.
    - Those data include province, date, daily case count, death count, vaccinated count, etc. The primary feature we are interested in specifically is the daily case count as we are analysing the impact of COVID-19, however, we also made use of the vaccinated count column in some of our analysis to analyse recovery efforts.
    - The dataset is in CSV format as 1 file, the size of the file is 978KB, with a total of 9721 rows and 24 columns.
    - This dataset was chosen because it was the only open dataset that we found that had the most comprehensive COVID daily case/vaccinated counts for Canada.
    - The dataset was cleaned in Python before inserting into SQL, redundant data was taken out, for example, the dataset had both the national COVID case count and provincial case counts, we chose to remove the national one. Column with DATETIME was also cleaned up to a DATE format because the data points are daily. The spaces in column names were also taken out for ease of working in SQL.
    - The only challenge we encountered with this dataset is that province names were in all uppercase (ex. "ALBERTA"), this is drastically different from most of our other datasets which were from Statistics Canada which uses an upper/lowercase mixture (ex. "Alberta"). So, we decided to better match our other datasets from Statistics Canada by changing our column names to a similar format.

We also prepared an additional optional COVID-19 dataset for the United States, below are the details for that dataset:

- COVID-19 (United States)
    - The dataset we used for this is the COVID Tracking Project COVID-19 All States History dataset (The Atlantic Monthly Group, 2021a) found on the COVID Tracking Project's website.
    - It is a Public Dataset published under a Creative Commons CC BY 4.0 license (The Atlantic Monthly Group, 2021b) and includes COVID-19 data from all the U.S. States.
    - Like the Canada dataset, it includes data for state, date, daily case count, death count, etc. The primary feature we are interested in specifically is the daily case count because that is the one feature we will be using to compare to our Canadian dataset with in our analysis.
    - The dataset is in CSV format as 1 file, the size of the file is 2675KB, with a total of 20781 rows and 41 columns.
    - This dataset was chosen because it was the best public dataset that we found that shares a similar data structure as the Canada dataset.
    - The dataset was cleaned in Python before inserting into SQL, like the Canada dataset, column with DATETIME was also cleaned up to a DATE format because the data points are daily.
    - We did not encounter any notable challenges with this dataset.

The COVID datasets described above was then used by each group member (primarily using the Canada dataset as most of our other datasets were from Statistics Canada) to join with their own domain's datasets. The datasets for each domain will be explored separately in detail now.

1. Tourism (Emmanuella Bamgbose)
    - Datasets used for this domain is sourced from statistics Canada and is publicly licensed
    - Two (2) major datasets will be used for this section in addition to the primary covid dataset mentioned above:
        - Tourism demand expenditure (Statistics Canada, 2021a),
        - Tourism demand expenditure x1,000,000 (all inclusive)
        - Tourism demand expenditure x1,000,000 (International)
        - Tourism demand expenditure x1,000,000 (Domestic)
        - Employment generated by tourism x1,000 (Statistics Canada, 2021c).
    - The period in view for all datasets is 2015(Q1) – 2021(Q2)
    - Data frequency for both dataset is quarterly, and data is in csv format with an approximate size of 8kb per file
    - The choice of this dataset is because tourism was Canada's number one service export, and the fact that tourism also accounted for about accounting for 1.8 million direct and indirect jobs the dataset. We believe that both datasets are complimentary in investigating the effect of covid on the sector. The datasets provide information on tourism spending and Jobs across the following industries

- Transportation
  - Passenger air transport
  - Passenger rail transport
  - Interurban bus transport
  - Vehicle rental
  - Vehicle repairs and parts
  - Vehicle fuel
  - Other transportation
- Accommodation
- Food and Beverages
- Other tourism commodities
  - Recreation and entertainment
  - Travel agency services
  - Pre-trip expenditures
  - Convention fees

- Dataset preview is displayed below.
- Tourism Demand dataset (all inclusive). (Dataset for domestic and international tourism has the same format)
- All expenditure values are to the multiple of a million (x1,000,000). We decided to maintain this while loading the data in the database

	Expenditures	Q1 2015	Q2 2015	Q3 2015	Q4 2015	Q1 2016	Q2 2016	Q3 2016	Q4 2016	Q1 2017	...	Q1 2019	Q2 2019	Q3 2019	Q4 2019	Q1 2020	Q2 2020	Q3 2020	Q4 2020	Q1 2021	Q2 2021
0	Tourism expenditures	20,695	20,862	20,994	21,146	21,414	21,610	22,007	22,039	22,287	...	23,709	23,805	23,755	23,824	20,444	6,979	10,798	10,449	10,191	10,622
1	Total tourism commodities	17,502	17,649	17,764	17,875	18,113	18,295	18,682	18,701	18,930	...	20,218	20,294	20,237	20,318	17,406	5,879	9,050	8,752	8,526	8,929
2	Transportation	8,224	8,278	8,332	8,413	8,539	8,666	8,881	8,850	8,961	...	9,824	9,811	9,762	9,854	8,375	1,868	2,938	2,998	2,779	2,902
3	Passenger air transport	4,356	4,406	4,438	4,482	4,542	4,653	4,866	4,850	4,964	...	5,713	5,719	5,680	5,790	4,824	238	565	681	529	564

- Employment generated by tourism
  - All Job figures are to the multiple of a thousand (x1,000). We decided to maintain this while loading the data in the database

	Activities	Q1 2015	Q2 2015	Q3 2015	Q4 2015	Q1 2016	Q2 2016	Q3 2016	Q4 2016	Q1 2017	...	Q1 2019	Q2 2019	Q3 2019	Q4 2019	Q1 2020	Q2 2020	Q3 2020	Q4 2020	Q1 2021	Q2 2021
0	Tourism activities	640.7	647.0	650.5	652.2	658.6	662.6	669.3	670.3	671.3	...	690.1	690.0	691.8	695.9	652.5	371.4	465.7	470.8	448.7	453.2
1	Total tourism industries	516.8	524.0	526.9	528.3	533.4	537.0	544.0	544.4	545.5	...	560.3	559.9	562.0	566.3	533.0	288.6	365.9	368.3	347.9	351.7
2	Transportation	69.7	71.1	71.5	72.9	73.7	75.8	77.4	77.9	78.4	...	79.4	78.9	80.5	82.5	79.7	62.2	58.0	60.3	56.5	54.1
3	Air transportation	44.0	45.1	45.6	46.9	47.9	50.3	51.9	52.3	52.2	...	54.1	53.4	54.9	57.2	56.2	46.5	41.5	43.0	39.0	35.7
4	Other transportation	25.7	26.0	25.9	26.0	25.8	25.5	25.5	25.6	26.2	...	25.3	25.5	25.6	25.3	23.5	15.7	16.5	17.3	17.5	18.4

- To prepare the data for analysis, the format of all the datasets had to be modified. Data was provided with the year and quarter in columns, and this would have made it a little difficult to work with. The dataset was imported into a python data frame and further transformation was done
- Firstly, the tourism sub-category and tourism industry was extracted and loaded into tables in the database

	groupid	groupname
0	1	Accommodation
1	2	Food and beverage services
2	3	Other tourism commodities
3	4	Total other commodities
4	5	Transportation

	categoryid	category	groupid
0	1	Accommodation	1
1	2	Convention fees	3
2	3	Food and beverage services	2
3	4	Interurban bus transport	5
4	5	Other transportation	5

- All four dataset was also transposed to have the category, year, quarter, and expenditure/ jobs as the new column. Transposed data was uploaded to tables in the database. This gave us a total of 4 additional tables: (preview below)
- Script below was run for each dataset

```
#format data structure for international tourism data

rec1=[]
data1=[]
# iterating the columns
for q in range(1,len(df1.columns)):
    temparray1 = []
    # iterating the rows
    for i in range(1,len(df1)):
        #print(i)
        recid=  i
        test=df1[(df1.columns)[0]].loc[i]
        test2=(df1.columns)[q]
        test3=df1[(df1.columns)[q]].loc[i]
        temparray1=[recid,test,test2,test3]
        #print(temparray)
        rec1.append(temparray1)
   # rec.append(temparray)


#convert list to dataframe
records1=pd.DataFrame(rec1, columns = ['recordid','category', 'period', 'expenditure'])
display(records1.head())
```

- Tourism expenditure (all inclusive), Tourism expenditure (international), Tourism expenditure (domestic) and Tourism employment tables. (Preview of tourism expenditure(x1,000,000) table below):

	recordid	category	period	expenditure
0	1	Passenger air transport	Q1 2015	4,356
1	2	Passenger rail transport	Q1 2015	56
2	3	Interurban bus transport	Q1 2015	217
3	4	Vehicle rental	Q1 2015	266
4	5	Vehicle repairs and parts	Q1 2015	1,098

- Necessary updates were also done on the tables to ensure all required fields were properly captured. This included adding “year” column and updating it with values form the “period” column, adding and updating a column for the “category id” and formatting the “expenditure”/ “job” column to integer etc. (Preview below):

```
#add colums (quater,year,categoryid) to dataframe
engine.execute('ALTER TABLE tourism_expenditure ADD (quarter VARCHAR(20) NULL, year YEAR NULL, categoryid INT)')
engine.execute('ALTER TABLE tourism_expenditure_int ADD (quarter VARCHAR(20) NULL, year YEAR NULL, categoryid INT)')
engine.execute('ALTER TABLE tourism_expenditure_dom ADD (quarter VARCHAR(20) NULL, year YEAR NULL, categoryid INT)')
engine.execute('ALTER TABLE tourism_employment ADD (quarter VARCHAR(20) NULL, year YEAR NULL, categoryid INT)')
```

```
#update categoryid column
engine.execute('update tourism_expenditure set categoryid= (select categoryid from tourism_category where tourism_category.category=tourism_expenditure
engine.execute('update tourism_expenditure_int set categoryid= (select categoryid from tourism_category where tourism_category.category=tourism_expendi
engine.execute('update tourism_expenditure_dom set categoryid= (select categoryid from tourism_category where tourism_category.category=tourism_expendi
engine.execute('update tourism_employment set categoryid= (select categoryid from tourism_category where tourism_category.category=tourism_employment.c
#update quarter column
engine.execute('update tourism_expenditure set quarter= (select left(period,2) from tourism_category where tourism_category.category=tourism_expenditur
engine.execute('update tourism_expenditure_int set quarter= (select left(period,2) from tourism_category where tourism_category.category=tourism_expenc
engine.execute('update tourism_expenditure_dom set quarter= (select left(period,2) from tourism_category where tourism_category.category=tourism_expenc
engine.execute('update tourism_employment set quarter= (select left(period,2) from tourism_category where tourism_category.category=tourism_employment.
#update year column
engine.execute('update tourism_expenditure set year= (select right(period,4) from tourism_category where tourism_category.category=tourism_expenditure.
engine.execute('update tourism_expenditure_int set year= (select right(period,4) from tourism_category where tourism_category.category=tourism_expendit
engine.execute('update tourism_expenditure_dom set year= (select right(period,4) from tourism_category where tourism_category.category=tourism_expendit
engine.execute('update tourism_employment set year= (select right(period,4) from tourism_category where tourism_category.category=tourism_employment.ca
#correct column name in employment data, change expendirure to jobs
engine.execute('ALTER TABLE tourism_employment RENAME COLUMN expenditure TO jobs')
```

- We also had to aggregate the primary covid 19 datasets as all tourism data have a quarterly frequency compared to the daily frequency of the primary covid 19 datasets.
- Preview of primary covid dataset:

```
### Covid master dataset
query12 = pd.read_sql_query('SELECT * FROM ca_covid;', engine)
display (query12.head())
```

	OBJECTID	Province	Abbreviation	DailyTotals	SummaryDate	TotalCases	TotalRecovered	DailyRecovered	TotalDeaths	DailyDeaths	...	TotalHospitalized	DailyHospitalized
0	1	Alberta	AB	0	2020-01-25	0	0	0	0	0	...	NaN	NaN
1	2	Northwest Territories	NT	0	2020-01-25	0	0	0	0	0	...	NaN	NaN
2	3	Yukon	YT	0	2020-01-25	0	0	0	0	0	...	NaN	NaN
3	4	Saskatchewan	SK	0	2020-01-25	0	0	0	0	0	...	NaN	NaN
4	5	Prince Edward Island	PE	0	2020-01-25	0	0	0	0	0	...	NaN	NaN

- creation of new table for aggregated covid 19 datasets: focus was only on the total number of cases.

```
#create new table for quarterly covid data
engine.execute('create table eb_covid (Year year,Quarter varchar(20),TotalCases int)')
```

```
<sqlalchemy.engine.result.ResultProxy at 0x7fcd49aaefd0>
```

```
#insert quarterly agregated values in to table
engine.execute('insert into eb_covid (select year,quarter,sum(TotalCases) total from (Select year(SummaryDate) Year,month(SummaryDate) Month,\
              TotalCases, (Case When month(SummaryDate) in (1,2,3) Then "Q1" When month(SummaryDate) in (4,5,6) Then "Q2"\
              When month(SummaryDate) in (7,8,9) Then "Q3" When month(SummaryDate) in (10,11,12) Then "Q4" Else "NA" End) Quarter from ca_covid) a\
              group by year,quarter)')
```

```
<sqlalchemy.engine.result.ResultProxy at 0x7fcd49aaed00>
```

- new covid 19 data table structure:

```
#display table
query13 = pd.read_sql_query('SELECT *,concat(year,quarter) yearquarter FROM eb_covid order by Year;', engine)
display (query13.head())
```

	Year	Quarter	TotalCases	yearquarter
0	2015	Q2	0	2015Q2
1	2015	Q4	0	2015Q4
2	2015	Q3	0	2015Q3
3	2015	Q1	0	2015Q1
4	2016	Q4	0	2016Q4

- After sourcing our dataset for the tourism investigation, working with the dataset as-is appeared to be very straightforward. As we got into the data clean up and transformation, we discovered much more was required to make the data suitable for efficient handling and analysis. converting the data structure, creating the right tables, introducing the right keys to enable seamless joins between the tables was very tasking but this enabled us to have a better understanding of cleaning data in SQL and Python

2. Air travel (Victor Nzemeke)
    - When covid hit the world late 2019, the first step many governments across the world took was to impose travel bans and stop all form of non-essential international travel. The International Civil Aviation Organization (ICAO) reports that more than 371 billion dollars in revenue was lost because of covid in 2020 alone across the globe. It goes on to project a 329-billion-dollar loss for 2021.

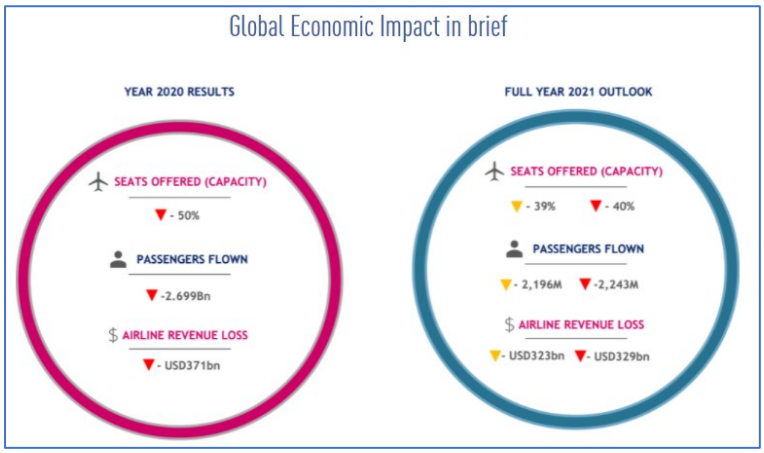


Figure 1 - ICAO Covid Impact 2020 and 2021 Outlook

    - For our project, we were interested in finding out the effect of air travel decline in Canada due to Covid. To achieve this, we would use some datasets provide by statistics Canada.
    - The primary dataset we will use for this is sourced from Statistics Canada "*Aircraft movements, by civil and military movements, airports with NAV CANADA towers, monthly* (Statistics Canada, 2021). Because this dataset does not include the passenger counts, we intend complementary datasets to gather passenger volume information. This data is hereinafter referred to as the "Flights Dataset".
    - The second dataset used for this project is the *Operating and financial statistics for major Canadian airlines, monthly* (Statistics Canada, 2021). This data reports metrics such as passenger counts and airline revenue for major Canadian airlines and is updated monthly by statistics Canada. This dataset is hereinafter referred to as the "Airlines Dataset".
    - Both datasets are open source and have public license.
    - The Flights Dataset provides a count of all flights into each airport every month from January 1997 to August 2021. The flights are divided into distinct categories, but we are most interested in the Civil commercial itinerant movements, Civil local movements, and Civil private and government itinerant movements. For the airports, we shall be using the main airports on the East and West Coasts of Canada namely Calgary International Airport, Vancouver International, Montreal International Airport and Toronto-Pearson International Airport.
    - The Flights Dataset is stored in csv format, and it is 14MB in size.
    - The Airlines Dataset provides operational and financial statistics for Canadian airlines from Jan 1981 to September 2021. The statistics provided are given below:

Operational and Financial Statistics	UOM
Passengers	Number
Passenger-kilometres	Number
Kilograms of goods	Kilograms
Goods tonne-kilometres	Tonne-kilometres
Hours flown	Number
Turbo fuel consumed	Litres
Total operating revenues	Dollars
Total operating expenses	Dollars
Total interest expenses	Dollars
Available seat-kilometres	Number
Load factor	Percent

- For this domain, we chose this dataset because they are directly relevant to the research questions for this section. They contain valid information that would help investigate the effects of covid on aviation in Canada.
- Data cleaning: For the Flights Dataset, the following steps were taken:
  - The following columns were dropped as they were not relevant to the analysis:
  - "'UOM', 'SCALAR_FACTOR','DGUID','SCALAR_ID', 'UOM_ID', 'VECTOR', 'TERMINATED', 'COORDINATE', 'STATUS', 'SYMBOL', 'DECIMALS', and 'GEO'.
  - The column "Civil and Military movements was renamed to "Flight Type" for simplicity
  - The "Date" Column was converted to a datetime object in python
  - A new column "ID" was created and populated as a serial number of each row using numpy.arange() function. The ID column would serve as primary key when uploading into the database.
  - Python code to accomplish data cleaning is given below:

```
#Preliminary data cleaning for the flight data:

#drop unnecessary columns:
flights.drop(['UOM', 'SCALAR_FACTOR','DGUID','SCALAR_ID', 'UOM_ID','VECTOR','TERMINATED',
              'COORDINATE', 'STATUS', 'SYMBOL', 'DECIMALS', 'GEO'], axis = 1, inplace = True)

#rename the "Civil and Military Movements' to Flight Type
flights.rename(columns = {'Civil and military movements':'Flight_Type'}, inplace= True)
flights.columns = map(str.lower, flights.columns)


#Convert the 'DATE' column to a `datetime` object in 24-hour time format,
flights.ref_date = pd.to_datetime(flights.ref_date, infer_datetime_format=True)


# Add a new column 'ID', which will serve as primary key
flights['id'] = np.arange(len(flights))


display(flights.head())
```

  - At the end of the cleaning, we had 89,229 rows and 5 columns. A brief description of the columns is also provided below:

	ref_date	airports	flight_type	value	id
0	1997-01-01	Total, all airports	Total civil movements	288679	0
1	1997-01-01	Total, all airports	Civil commercial itinerant movements	184686	1
2	1997-01-01	Total, all airports	Civil private and government itinerant movements	23616	2
3	1997-01-01	Total, all airports	Civil local movements	80377	3
4	1997-01-01	Total, all airports	Total military movements	6394	4

(89229, 5)

Column Name	Data Type	Description
ref_date	datetime	Date in scope
airports	string	Airport Name
flight_type	string	Flight movement type (military, civilian)
value	integer	Number of flights recorded for month in scope
id	integer	Index

- For the Airlines Dataset, the following steps were taken for data cleaning:
  - Columns that were not useful to the analysis were dropped
  - The “REF_DATE” column was converted to a python datetime object
  - For ease of reference, the “Operating and Financial Statistics” was renamed to “ops_stats”
  - A new column “ID” was created and populated as a serial number of each row using numpy.arange() function. The ID column would serve as primary key when uploading into the database.
  - Code to accomplish data cleaning is given below:

```
# Covid Data vs Flight numbers:
# Get the data for the covid flights
# 2021 - Vaccination numbers vs number of flights
#Import Table 23100079:
# Statistics Canada. Table 23-10-0079-01  Operating and financial statistics for
#                                          major Canadian airlines, monthly
# https://doi.org/10.25318/2310007901-eng

airline_data = pd.read_csv('23100079.csv')

#drop unnecessary columns:
airline_data.drop(['UOM','DGUID','SCALAR_ID', 'UOM_ID','VECTOR','TERMINATED', 'COORDINATE',
                   'STATUS', 'SYMBOL', 'DECIMALS', 'GEO'], axis = 1, inplace = True)


#Convert the 'Ref_date' column to a `datetime` object in 24-hour time format,
airline_data.REF_DATE= pd.to_datetime(airline_data.REF_DATE, infer_datetime_format=True)


#rename all columns to lowercase
airline_data.columns = map(str.lower, airline_data.columns)

#shorten col names:
airline_data.rename(columns= {"operational and financial statistics":"ops_stats"}, inplace=True)

#create new column to serve as primary key:
airline_data['id'] = np.arange(len(airline_data))


display(airline_data.head(10))
print (airline_data.shape)

```

- At the end of the data cleaning, we were left with 4,291 rows and 5 columns of data. Description is provided below.

	ref_date	ops_stats	scalar_factor	value	id
0	1981-01-01	Passengers	thousands	1895.0	0
1	1981-01-01	Passenger-kilometres	thousands	3411545.0	1
2	1981-01-01	Kilograms of goods	thousands	29091.0	2
3	1981-01-01	Goods tonne-kilometres	thousands	64084.0	3
4	1981-01-01	Hours flown	thousands	53.0	4
5	1981-01-01	Turbo fuel consumed	thousands	283396.0	5
6	1981-01-01	Total operating revenues	thousands	261969.0	6
7	1981-01-01	Total operating expenses	thousands	266295.0	7
8	1981-01-01	Total interest expenses	thousands	0.0	8
9	1981-02-01	Passengers	thousands	1852.0	9

(4291, 5)

Column Name	Data Type	Description
ref_date	datetime	Date in scope
ops_stats	string	Operational statistics such as passenger count, operating revenue, hours flown.
scalar_factor	string	The factor by which the value column is scaled. Example: 5 on the value column would mean 5,000 if the value on the scalar factor is 'thousands'
value	float	Value of the operational statistic being measured for the time in scope
id	integer	Index

  - Lessons learnt
    - Although learnt in hindsight, we found out that working in databases with single-word column names is much easier and less error-prone than working with long names. This informed the need to return and rename some of the columns the way we did.
  - Challenges
    - Generally, we found the data from statistics Canada to be already well-cleaned for general use. However, for the purpose of this project, many of the datasets we came across did not provide data to categorize as passenger numbers per airport, and passenger numbers by flight type which would have helped us gain a much better understanding of the effects of covid on international flights and local flights within Canada.

3. Ground travel (Daniel Zhou)
   - The datasets we used for this domain are sourced from Statistics Canada and Bureau of Transportation Statistics. Specifically, they are:
     - Statistics Canada. Table: 24-10-0002-01 Number of vehicles travelling between Canada and the United States (Statistics Canada, 2021l)

- U.S. Department of Transportation. Bureau of Transportation Statistics. Border Crossing Entry Data (U.S. Department of Homeland Security, 2021)

- We are free to use both datasets. Specifically, the dataset from Statistics Canada has an Open Licence (Statistics Canada, 2021l) and the dataset from U.S. Department of Transportation has a Public Domain License (U.S. Department of Homeland Security, 2021).
- Both datasets are structured datasets that includes similar data on what vehicles crossed the border and entered the respective country in any specific month, the primary features we are interested in are the year and month of entry as well as the number of vehicles entered.
- The datasets are stored in CSV format, 1 CSV file each, sized 143 MB and 30 MB respectively, the dimensions are 831361 x 17 and 368015 x 7 respectively.
- These datasets were chosen because they are comparable to each other in that both has vehicle counts entering their respective countries borders, and we felt like vehicle entry count through the border is an accurate measure of international ground travel between Canada and the US. The fact that they are both aggregated monthly is another bonus so that we can easily join the 2 datasets together.
- Both datasets were cleaned in Python before inserting into SQL. The U.S dataset was of excellent quality so no data cleaning was required other than to take out the spaces in the column names so that it's easier to work with in SQL. For the Canadian dataset, we had to do some extra work. First, we took out redundant data, for example, the dataset contained vehicle entry counts at both a national level and a provincial level, we chose to remove the national one to make it consistent with the COVID dataset (where it also aggregated data at a provincial level). There were also several columns with total values as well as values separated by categories which is redundant, specifically, that is the "Trip characteristics", "Length of stay", and "Mode of transportation" columns, we decided to take out the redundant values for those columns. Column with DATETIME was also cleaned up to a DATE format because the data points are daily. The spaces in column names were also taken out for ease of working in SQL. Please see figure below for some sample data from the 2 datasets, these samples are taken after data cleaning and SQL insertion.

RefDate	Province	TripCharacteristics	ModeOfTransport	Value
2021-01-01	Alberta	Total Canadian vehicles returning	Trucks	9299
2021-01-01	British Columbia	Total Canadian vehicles returning	Automobiles	11876
2021-01-01	British Columbia	Total Canadian vehicles returning	Trucks	40759
2021-01-01	British Columbia	Total United States vehicles entering	Automobiles	16197
2021-01-01	British Columbia	Total United States vehicles entering	Trucks	10692

portName	state	border	date	measure	value
Alcan	Alaska	US-Canada Border	2021-01-01	Personal Vehicle Passengers	690
Alcan	Alaska	US-Canada Border	2021-01-01	Personal Vehicles	444
Alcan	Alaska	US-Canada Border	2021-01-01	Trucks	416
Douglas	Arizona	US-Mexico Border	2021-01-01	Pedestrians	40886
Douglas	Arizona	US-Mexico Border	2021-01-01	Personal Vehicle Passengers	114575

- Our biggest learning from the data clearing step is that variable type is especially important for an SQL database, when we were initially inserting these tables into SQL, we

had some problems because DATETIME data were getting inserted as strings, so our takeaway was to declare each variable type ourselves before insertion instead of relying on SQL Alchemy to autodetect for us.

4. Electricity (Spencer Dow)
    - The data set used for this section is the "Electric power generation, monthly generation by type of electricity" dataset from Statistics Canada (Statistics Canada, 2021d)
    - This is an open dataset and has an open licence
    - In this domain we investigated the relationship between Covid 19 and the electricity industry. Our team thought that this would be an interesting topic to pursue because what we need to power has changed significantly with the Covid-19 pandemic. As people begin not working from the office and start working from home there is a shift in electricity usage observable in our everyday life. We wanted to see how these changes notable societal changes effected the electricity production industry.
    - First, we will explain what electricity does for us and the current landscape of the industry here in Canada. Electricity is a secondary energy source with many applications that include heating and lighting (Natural Resources Canada. 2021a). The electricity industry preforms three primary functions: generation, transmission, and distribution. For the purposes of this report, we will only be looking at the generation portion of the industry. To gain a better understanding of the use of electricity here in Canada the figure below shows how much electricity was used per sector in 2016.

Sector	Energy use (PJ)	% of the total
Residential	604.1	33.3%
Commercial	429.7	23.7%
Industrial	739.0	40.8%
Transportation	4.4	0.2%
Agriculture	34.8	1.9%
Total	**1,812.0**	**100%**

Table Retrieved from: Electricity facts (nrcan.gc.ca)

    - Another reason we wanted to investigate this topic was to discover the effects that the changes in energy production has had on the environment. The environmental impact of electricity generation is interesting for two reasons. The first being it is ahead of other industries in terms of GHG reduction. Over the past seventeen years the GHG emissions per year has steadily declined from just under 130 megatonnes in 2001 to just under 70 in 2018. This change has occurred because of the adoption of new ways to generate energy through non GHG emitting sources (Natural Resources Canada. 2021b). A yearly breakdown of GHG emissions by energy source is displayed in the figure below.

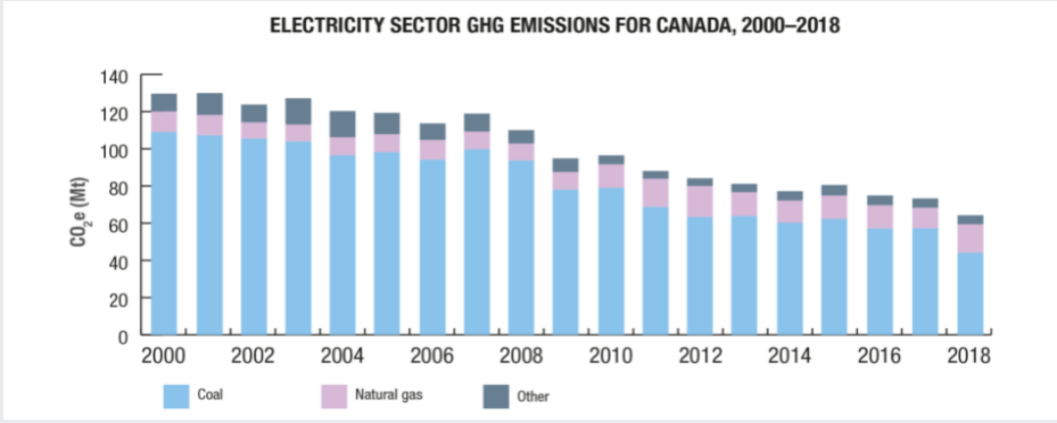


Figure retrieved from: [Energy and Greenhouse Gas Emissions (GHGs) (nrcan.gc.ca)]

- As evident in the chart the vast majority of GHG emissions come from the use of coal. This is true despite coal only making up 7% of the total electricity generation.
- In this section we hope to answer two key questions. The first being has the electricity industry experienced any major shifts in production due to the Covid-19 pandemic. The second question is having there been any shift in the source that we use to produce energy over the course of the pandemic.
- The data sets we will be using for this analysis are the main Covid-19 dataset used throughout the report as well as a data set from statistics Canada: Electric power generation, monthly generation by type of electricity (Statistics Canada. 2021d). This dataset provides a monthly breakdown electricity production by type. The only variables that we are interested in is the Geography, Type of electricity generation, Date and the Value of electricity produced. There are other identifier variables that will not be used in this analysis. The data was downloaded in 14 separate csv files (one for each province/territory and one for all of Canada). A sample of the data is provided below.

	REF_DATE	GEO	DGUID	Class of electricity producer	Type of electricity generation	UOM	UOM_ID	SCALAR_FACTOR	SCALAR_ID	VECTOR	COORDINATE	VALUE	STATUS	SYMBOL	TERMINATED	DECIMALS
0	2016-01	Canada	2016A000011124	Total all classes of electricity producer	Total all types of electricity generation	Megawatt hours	210	units	0	v44174609	1.1.1	64228215.0	NaN	NaN	NaN	0
1	2016-02	Canada	2016A000011124	Total all classes of electricity producer	Total all types of electricity generation	Megawatt hours	210	units	0	v44174609	1.1.1	57932188.0	NaN	NaN	NaN	0
2	2016-03	Canada	2016A000011124	Total all classes of electricity producer	Total all types of electricity generation	Megawatt hours	210	units	0	v44174609	1.1.1	57108230.0	NaN	NaN	NaN	0
3	2016-04	Canada	2016A000011124	Total all classes of electricity producer	Total all types of electricity generation	Megawatt hours	210	units	0	v44174609	1.1.1	49984993.0	NaN	NaN	NaN	0
4	2016-05	Canada	2016A000011124	Total all classes of electricity producer	Total all types of electricity generation	Megawatt hours	210	units	0	v44174609	1.1.1	47876205.0	NaN	NaN	NaN	0

- Data Cleaning and Preparation:
  - This data set was clean straight from the download but did require a few minor changes to make it easier to work with. The first step taken was to merge all 14 tables together. I opted to do this first because all the other data preparation would be the same for the tables and by merging them together it would save time in repeating the same steps 14 times. After combining the tables, the next step was to drop the unnecessary variables. I decided to leave these five variables in the table while the rest were removed: REF_DATE (The corresponding date), GEO (location), Class of electricity producer, Type of

electricity generation and VALUE (the amount of electricity generated). Next, I set cast the variables to the desired types

Variable	Type
REF_DATE	Date Time
GEO	String
Class of electricity producer	String
Type of electricity generation	String
VALUE	Float

- Lastly, I filled all the NULL values in the VALUE column to 0. The NULL values were originally placed in the dataset represent that no energy was produced by that source in a specified time frame. This was to make it easier for calculation purposes later in the project. After finishing the data cleaning, I exported it back to a csv file and uploaded it directly to MYSQL. Information about the final data and sample of the final data is provided below.
    - Data Sample:

	REF_DATE	GEO	Class of electricity producer	Type of electricity generation	VALUE
0	2016-01-01	Canada	Total all classes of electricity producer	Total all types of electricity generation	64228215.0
1	2016-02-01	Canada	Total all classes of electricity producer	Total all types of electricity generation	57932188.0
2	2016-03-01	Canada	Total all classes of electricity producer	Total all types of electricity generation	57108230.0
3	2016-04-01	Canada	Total all classes of electricity producer	Total all types of electricity generation	49984993.0
4	2016-05-01	Canada	Total all classes of electricity producer	Total all types of electricity generation	47876205.0

- Data Information:

```
RangeIndex: 8840 entries, 0 to 8839
Data columns (total 5 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   REF_DATE                        8840 non-null   datetime64[ns]
 1   GEO                             8840 non-null   object
 2   Class of electricity producer   8840 non-null   object
 3   Type of electricity generation  8840 non-null   object
 4   VALUE                           8840 non-null   float64
```

- The biggest challenge in working with this database is understanding and sorting out the information. It took a surprising amount of background information to be able to develop a plan for the analysis and to format the database accordingly.

5. Manufacturing (Harman Kaur)
    - The primary data used is the manufacturing data provided by Statistics Canada, which consists of the monthly sales data in thousand dollars from Jan 2019 to Sep 2021.This data includes 18 attributes and provides the total inventory value and the total value for the unfilled orders at the end of each month. The data is available for public use and is provided in structured, tabular csv files.

- The manufacturing data consisted of a total of 7 csv files with each file representing a different manufacturing sector, therefore, a significant amount of data cleaning was performed using pandas and several SQL queries. This consisted of filtering out the relevant information and dropping a few columns. The data was cleaned and combined into a single csv file .The processed data consisted of 640 rows and six columns.
- Another data set is the International Merchandise trade data provide by Statistics Canada which consisted of a total of 8 csv files, which were used to analyse how the international merchandise trade was impacted by the manufacturing sector. The data was cleaned and combined into a single csv file before uploading it to the database. The processed data consisted of 1057 rows and 8 data attributes.
- We also used the Employment data to analyse the impact of the manufacturing sector on the employment rate during covid-19. This dataset consists of 769 rows and 5 attributes. The data is open source and is published under the open license.
- All three datasets, I.e., manufacturing data, international merchandise trade data and employment data were uploaded on the MariaDB database.
- With most of data cleaning done in pandas, a few sets of SQL queries were also used for data cleaning. The SQL queries used are given below:

```
#Removing Null values from the Manufacturing sales table
q="Select count(*) from manufacturing_sales where `VALUE` IS NULL;"

query_1='DELETE FROM manufacturing_sales WHERE `VALUE` IS NULL;'
conn.execute(query_1)

q_rem="Select count(*) from manufacturing_sales where `VALUE` IS NULL;"
print(pd.read_sql_query(q_rem,engine))


#Removing Null values from the International Merchandize trade table
query_3='DELETE FROM international_trade WHERE `VALUE` IS NULL;'
conn.execute(query_3)

#Removing Null values form the Employment table
query_4='DELETE FROM employment WHERE `VALUE` IS NULL;'
conn.execute(query_4)

#Renaming columns from the manufacturing sales table
query_1="Alter table manufacturing_sales rename column `North American Industry Classification System (NAICS)` to category;"
query_2="Alter table manufacturing_sales rename column `VALUE` to manufacturing_value;"
conn.execute(query_1)
conn.execute(query_2)

#Renaming columns from the International merchandise trade data
query_3="Alter table international_trade rename column `North American Product Classification System (NAPCS)` to product_category;"
query_4="Alter table international_trade rename column `VALUE` to trade_value;"
conn.execute(query_3)
conn.execute(query_4)

#Renaming columns from the Employment table
query_5="Alter table employment rename column `North American Industry Classification System (NAICS)` to emp_sector;"
query_6="Alter table employment rename column `VALUE` to emp_value;"
conn.execute(query_5)
conn.execute(query_6)
```

6. Retail (Yves Nsoga)
    - To be able to carry out the impact of covid on retail and consumers shopping habit we are going to use two datasets
        - The main and primary dataset is the Retail trade sales by province and territory by statistic Canada, an open-source dataset provided by the government of Canada (Statistics Canada, 2021i)

    - The second dataset present the retail trade, total sales, and e-commerce sales in Canada from 2019 to 2021 used in the study published by authority of the Minister responsible for Statistics Canada exploring the ecommerce sales volumes. (Statistics Canada, 2021j)
  - Both datasets are tabular data on csv format, published under the open agreement license. The fact that both datasets contain records from the period pre covid (2019) to post covid is interesting.
  - We chose those datasets to understand the impact of covid pandemic on consumers shopping habits because these datasets present the variation of sales for a period covering the pre and post pandemic.
  - To make the data potable, we performed a set of operations. The first was to delete unnecessary columns and empty rows. Then we convert the date column into panda's date-time object. To homogenize the datasets, we modified the granularity of dataset to have monthly records. Having the same granularity for all datasets permit to merge those datasets into one data Frame.

Data Exploration and Discussion

The data exploration and discussion part of the project is done together by everyone for the COVID-19 dataset and separately by each member for their assigned domain with COVID-19 dataset as the common dataset.
We will separately discuss our exploration procedures, results, and findings for each domain in the individual sections that follow.

Tourism - Data Exploration and Discussion (Emmanuella Bamgbose)

After ensuring all required datasets have been well prepared and loaded into our database, we proceeded to investigate and provide answers to our research questions.

First research question, what is the impact of Covid on expenditure for tourism demand? How is international and domestic tourism also affected?

Our three (3) tourism expenditure tables (total, international and domestic) and our primary covid 19 data table were used to provide insight into this research question. The 3 tables were joined, and the “expenditure” column was summed per year and quarter and multiplied by a million (1,000,000) for each table. This is because the original dataset is provided in multiples of *1,000,000. The product of the join gave us a dataset that contained the total expenditure for every quarter from 2015 Q1 to 2021 Q2. The aggregated covid 19 data was also used to compare with the tourism expenditure trend

With this, we were able to produce a plot that shows the trend of expenditure over the years and the total number of covid cases for the same period.
Tourism expenditure script:

```
qtst = pd.read_sql_query('select a.yearquarter, total,international,domestic from\
 (select x.year,x.quarter,concat(x.year,x.quarter)yearquarter,\
 (sum(expenditure)*1000000) total  from tourism_expenditure x,\
(select a.*, b.groupname from tourism_category a, tourism_catgroup b where a.groupid=b.groupid)\
y where x.categoryid=y.categoryid group by x.quarter,x.year order by x.year,yearquarter) a,\
  (select x.year,x.quarter,concat(x.year,x.quarter)yearquarter,\
 (sum(expenditure)*1000000) international  from tourism_expenditure_int x,\
(select a.*, b.groupname from tourism_category a, tourism_catgroup b where a.groupid=b.groupid)\
y where x.categoryid=y.categoryid group by x.quarter,x.year order by x.year,yearquarter ) b,\
(select x.year,x.quarter,concat(x.year,x.quarter)yearquarter,\
 (sum(expenditure)*1000000) domestic  from tourism_expenditure_dom x,\
(select a.*, b.groupname from tourism_category a, tourism_catgroup b where a.groupid=b.groupid)\
y where x.categoryid=y.categoryid group by x.quarter,x.year order by x.year,yearquarter) c\
 where a.yearquarter= b.yearquarter and b.yearquarter=c.yearquarter;'
                     , engine)

display(qtst.head())
```

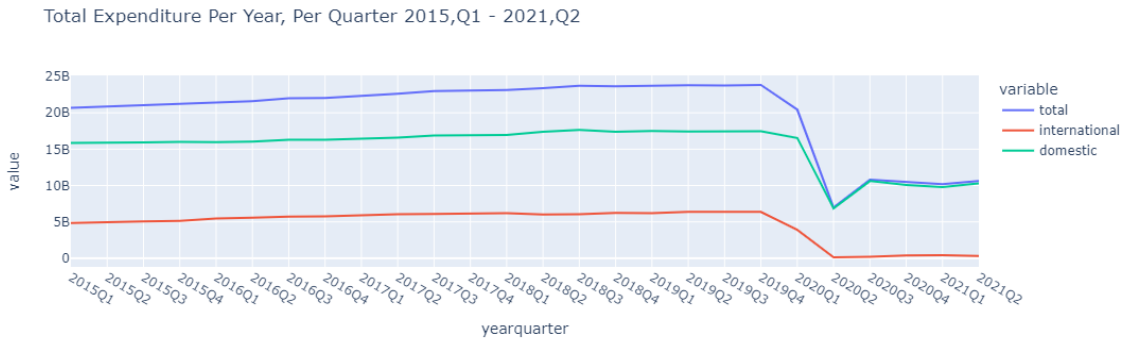
	yearquarter	total	international	domestic
0	2015Q1	2.069500e+10	4.858000e+09	1.583700e+10
1	2015Q2	2.086200e+10	5.013000e+09	1.584900e+10
2	2015Q3	2.099400e+10	5.080000e+09	1.591400e+10
3	2015Q4	2.114600e+10	5.142000e+09	1.600400e+10
4	2016Q1	2.141400e+10	5.462000e+09	1.595200e+10

Covid19 cases script:

```
#Covid data
q1= pd.read_sql_query('select "covid" groupname, year,quarter,concat(year,quarter) yearquarter, (totalcases/100) total\
from eb_covid order by year,yearquarter;',engine)
display(q1.head())

fig2 = px.line(q1, x="yearquarter", y="total", color="groupname",title='Number of Covid Cases')
fig2.show()
```

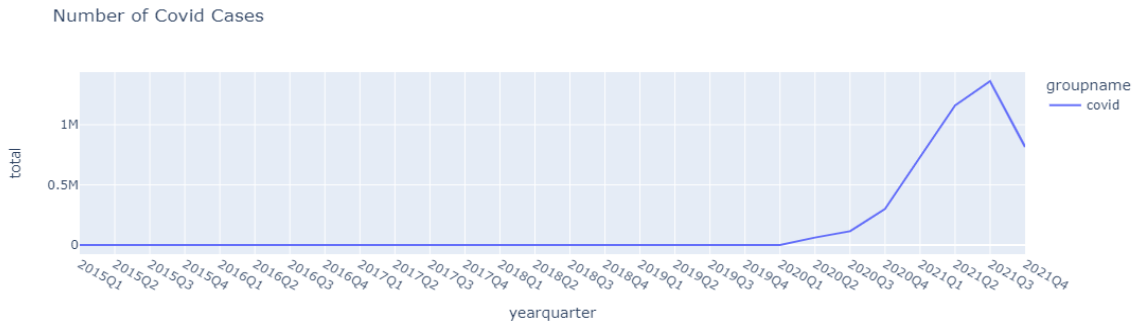
	groupname	year	quarter	yearquarter	total
0	covid	2015	Q1	2015Q1	0.0
1	covid	2015	Q2	2015Q2	0.0
2	covid	2015	Q3	2015Q3	0.0
3	covid	2015	Q4	2015Q4	0.0
4	covid	2016	Q1	2016Q1	0.0

Plots:



We can see a steady growth in tourism spending from 2015, expenditure of about $20B to 2019, expenditure of $23.8B, before a small decline in 2020, Q1 followed by a sharp decline by 2020, Q2.

all- time low of about $7B. Domestic tourism is seen to have picked up a bit compared to international tourism expenditure; this also caused a similar increase in the total tourism expenditure trend.



From the Covid 19 data, we can see the rise in covid cases from 2020, Q1 to 2021, Q3, this correlates with the year and quarter where a huge decline in tourism spending was experienced. Although the decline in tourist activities can also be attributed to government policies. We cannot deny the profound influence the pandemic had on government policies.

Proceeding to our second research question, we decided to investigate which tourism industry was most hit by the effect of covid on tourism.

We did a further dive into our tourism expenditure data. This time around, we summed expenditure by group (industry) and year, quarter for total covid expenditure, international covid expenditure and domestic covid expenditure separately to provide a better understanding of how the various industries are affected. This enabled us to plot a trend for the different industries for all three areas for total tourism expenditure

```
# get the average expenditure per group per quarter from year 2015 to 2021, is the effect of covid evident?
#for combined data
q = pd.read_sql_query('select y.groupname, x.year,x.quarter,concat(x.year,x.quarter)\
yearquarter, (sum(expenditure)*1000000) total  from tourism_expenditure x, \
(select a.*, b.groupname from tourism_category a, tourism_catgroup b where a.groupid=b.groupid)\
y where x.categoryid=y.categoryid group by x.quarter,x.year,y.groupname order by x.year ;'
                       , engine)
display (q.head(5))
fig1 = px.line(q, x="yearquarter", y="total", color="groupname",title='Total Expenditure per Group per Quarter(Year 2015 to 2021)')
fig1.show()
```

	groupname	year	quarter	yearquarter	total
0	Transportation	2015	Q1	2015Q1	8.224000e+09
1	Accommodation	2015	Q1	2015Q1	2.816000e+09
2	Food and beverage services	2015	Q1	2015Q1	3.088000e+09
3	Other tourism commodities	2015	Q1	2015Q1	3.374000e+09
4	Total other commodities	2015	Q1	2015Q1	3.193000e+09

for international tourism expenditure

```
#for international data
qi = pd.read_sql_query('select y.groupname, x.year,x.quarter,concat(x.year,x.quarter)\
yearquarter, (sum(expenditure)*1000000) total  from tourism_expenditure_int x, \
(select a.*, b.groupname from tourism_category a, tourism_catgroup b where a.groupid=b.groupid)\
y where x.categoryid=y.categoryid group by x.quarter,x.year,y.groupname order by x.year ;'
                       , engine)
display (qi.head(5))
fig1i = px.line(qi, x="yearquarter", y="total", color="groupname",title='International Expenditure per Group per Quarter(Year 2015 to 2021)')
fig1i.show()
```

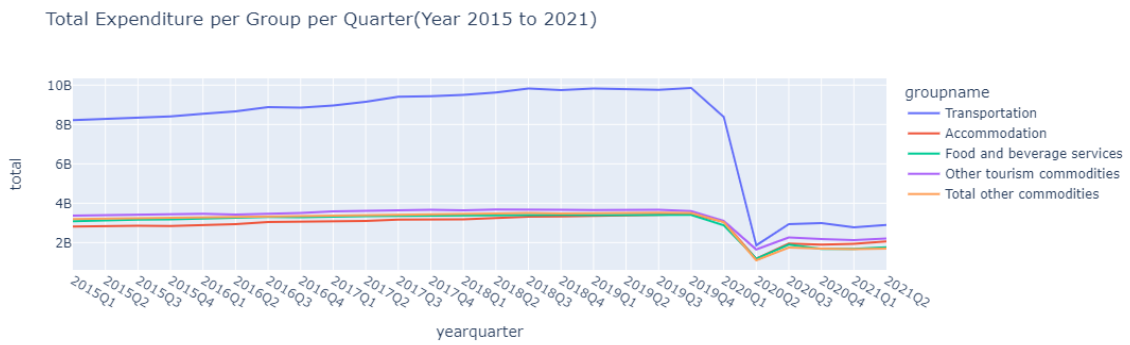
	groupname	year	quarter	yearquarter	total
0	Transportation	2015	Q1	2015Q1	1.668000e+09
1	Accommodation	2015	Q1	2015Q1	1.117000e+09
2	Food and beverage services	2015	Q1	2015Q1	6.890000e+08
3	Other tourism commodities	2015	Q1	2015Q1	5.290000e+08
4	Total other commodities	2015	Q1	2015Q1	8.550000e+08

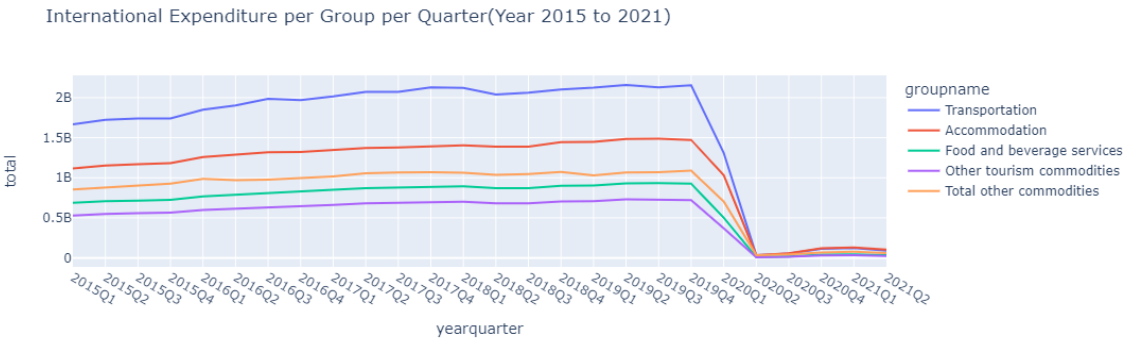
for domestic tourism expenditure

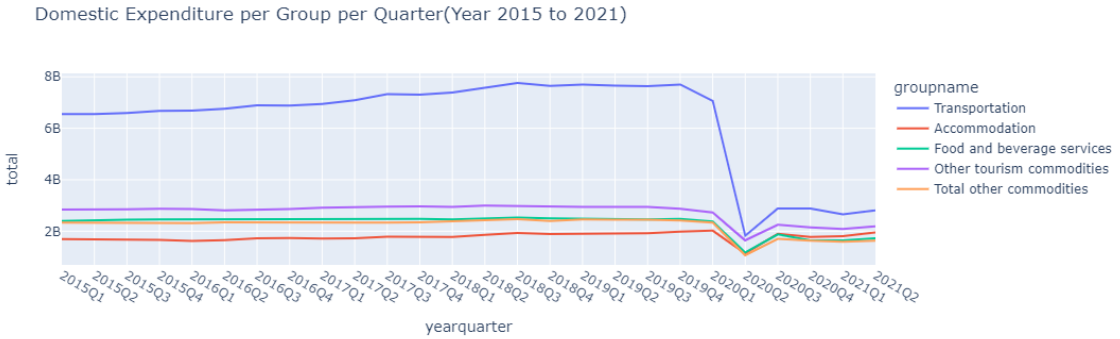
```
#for domestic data
qd = pd.read_sql_query('select y.groupname, x.year,x.quarter,concat(x.year,x.quarter)\
yearquarter, (sum(expenditure)*1000000) total  from tourism_expenditure_dom x, \
(select a.*, b.groupname from tourism_category a, tourism_catgroup b where a.groupid=b.groupid)\
y where x.categoryid=y.categoryid group by x.quarter,x.year,y.groupname order by x.year ;'
                       , engine)
display (qd.head(5))
fig1d = px.line(qd, x="yearquarter", y="total", color="groupname",title='Domestic Expenditure per Group per Quarter(Year 2015 to 2021)')
fig1d.show()
```

	groupname	year	quarter	yearquarter	total
0	Transportation	2015	Q1	2015Q1	6.556000e+09
1	Accommodation	2015	Q1	2015Q1	1.699000e+09
2	Food and beverage services	2015	Q1	2015Q1	2.399000e+09
3	Other tourism commodities	2015	Q1	2015Q1	2.845000e+09
4	Total other commodities	2015	Q1	2015Q1	2.338000e+09

Plots





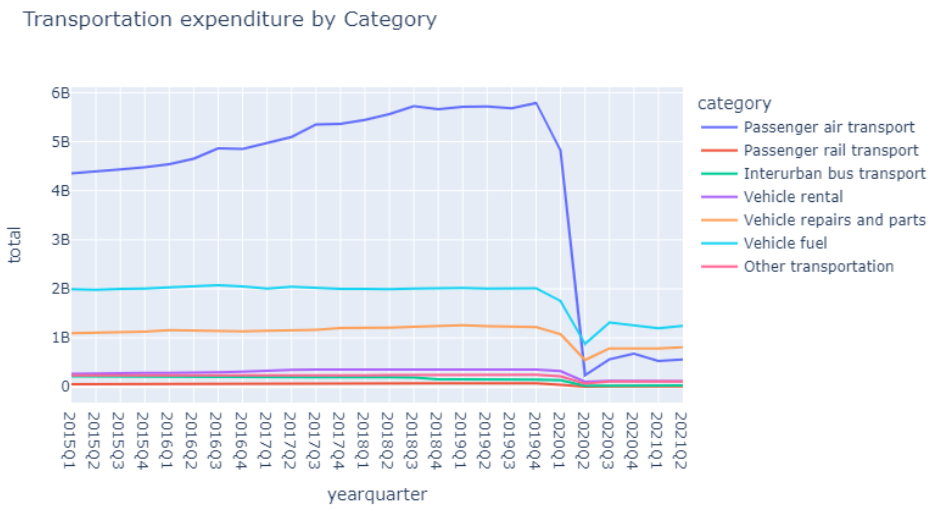


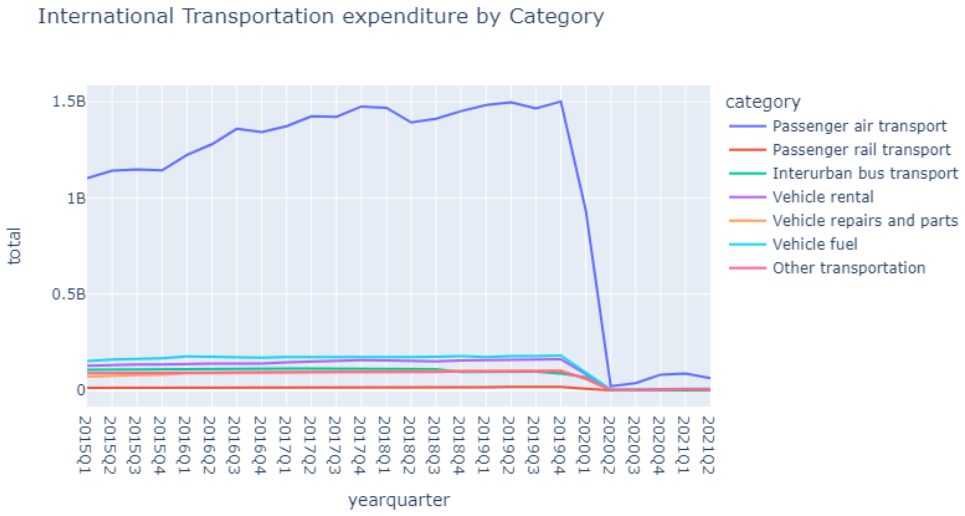
By further breaking down the tourism expenditure per tourism industry, we can see how they were and still are affected by the pandemic, the transportation industry being hit the hardest with a huge deep from about $10B to about$2B by 2020, Q2 following the covid19 pandemic. As seen in the covid 19 cases trend shared in the first research question, the decline in expenditure across all the industries started in 2020, Q1 just when the covid 19 cases started to rise.

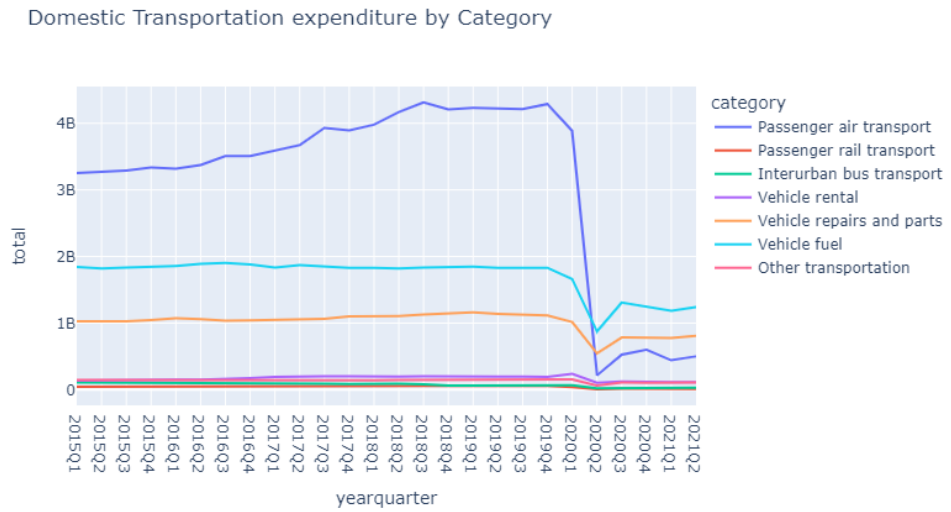
Although domestic tourism expenditure appears to rise from the sharp decline, the transportation sector is still vastly different from the expenditure as of 2019, Q3. International tourism is still the most affected as certain government policies are still in place to manage the spread of covid 19 and other variants.

We decided to look a bit deeper into some of the subcategories:

- Transportation

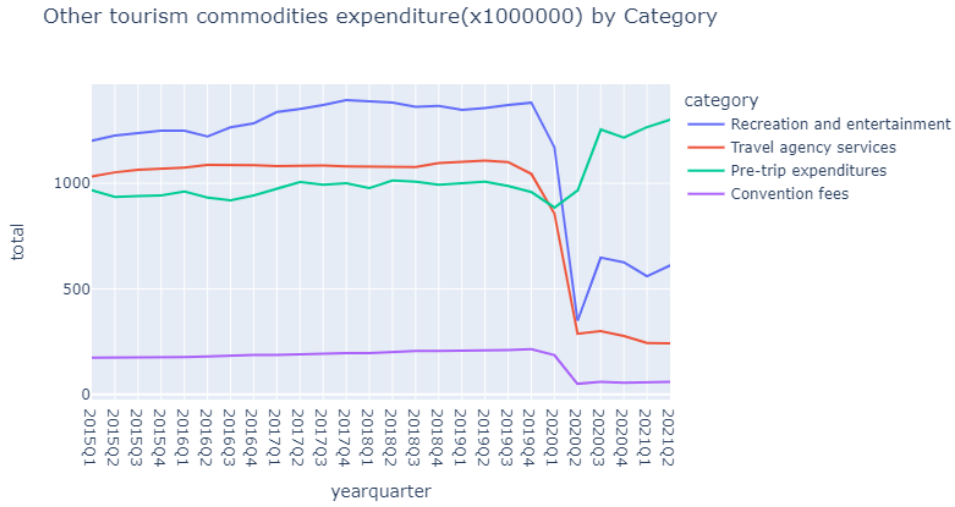


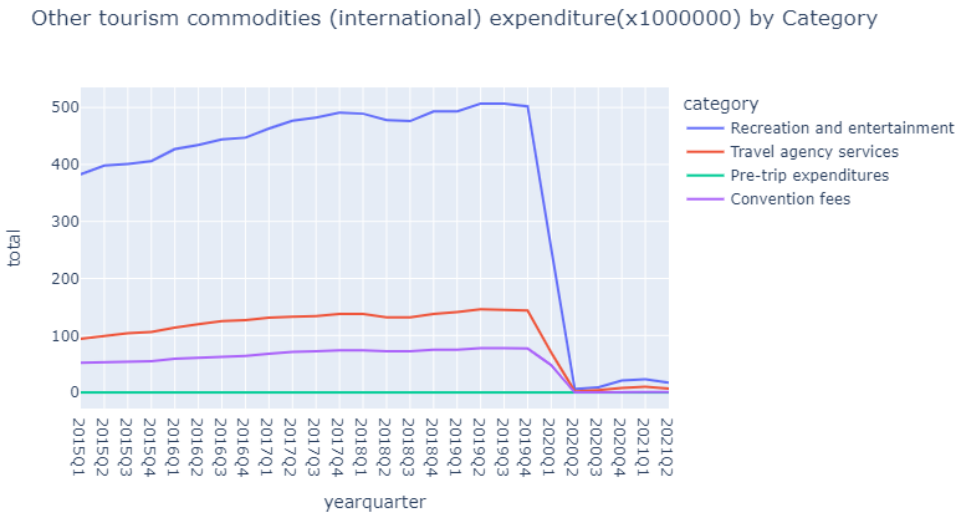


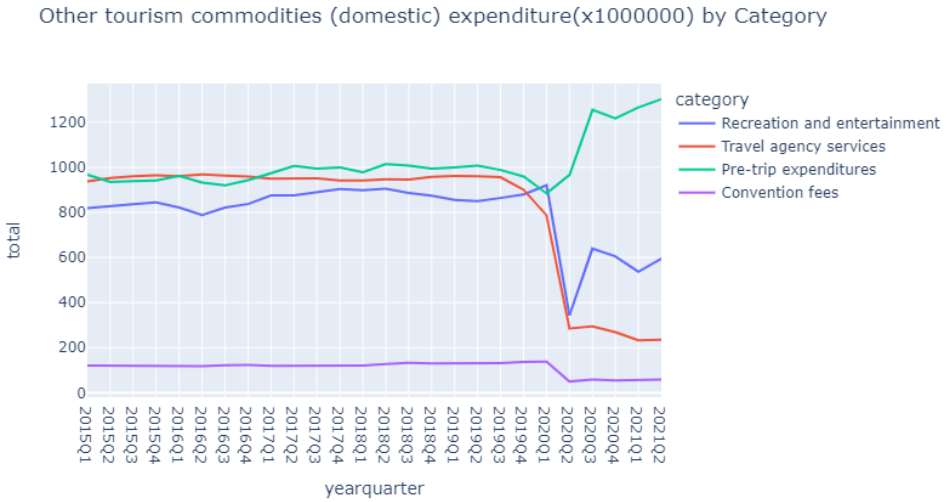


Within the transportation industry, the passenger air transport was most affected by the Covid 19 pandemic. Although there is a slow rise in the spending, this is still vastly different from the total transport expenditure from 2015-2019 before the pandemic.

Other tourism commodities (x1.000,000)







For the other commodities spending, the recreation and entertainment industry have the highest decline. Although, there was a rise in 2020, Q3 even when the covid cases was still rising, it

experienced another decline in 2021, Q1. The pre-trip expenditure appears to be the only sector not experiencing a decline

Moving to our investigation into the effect of covid on employment generated by tourism,
We used employment by tourism data to investigate the research question. We calculated the total number of jobs per year and quarter from 2015, Q1 to 2021, Q2. We compared the employment by tourism trend with the tourism expenditure per year and quarter and covid 19 cases across the same period.

Total employment generated by tourism per year and quarter

```
q9 = pd.read_sql_query('select year, quarter,concat(year,quarter) yearquarter,\
(sum(jobs)*1000) total  from tourism_employment group by quarter,year order by year, quarter ;'
                       , engine)
display (q9.head(5))
fig9 = px.line(q9, x="yearquarter", y="total",title='Total Jobs per Quarter(Year 2015 to 2021)')
fig9.show()
```

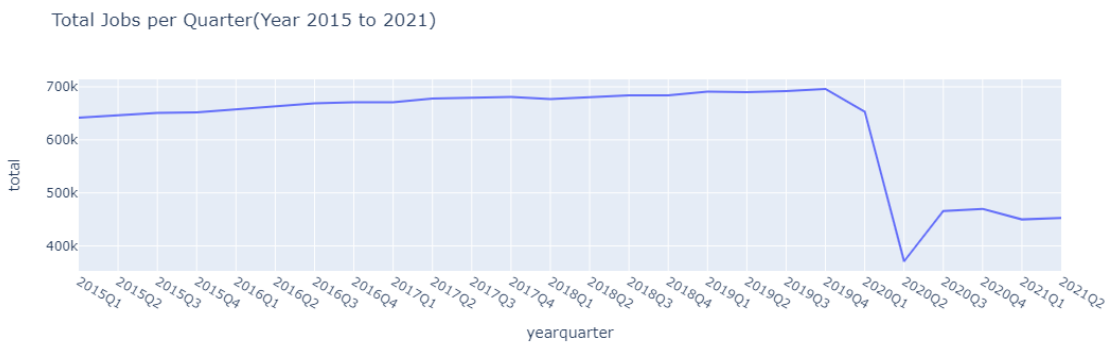
	year	quarter	yearquarter	total
0	2015	Q1	2015Q1	642000.0
1	2015	Q2	2015Q2	647000.0
2	2015	Q3	2015Q3	651000.0
3	2015	Q4	2015Q4	652000.0
4	2016	Q1	2016Q1	658000.0

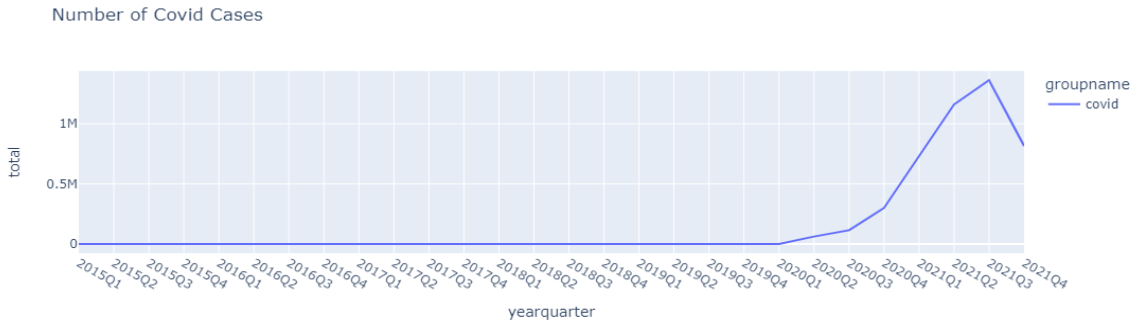
Total tourism expenditure per year and quarter

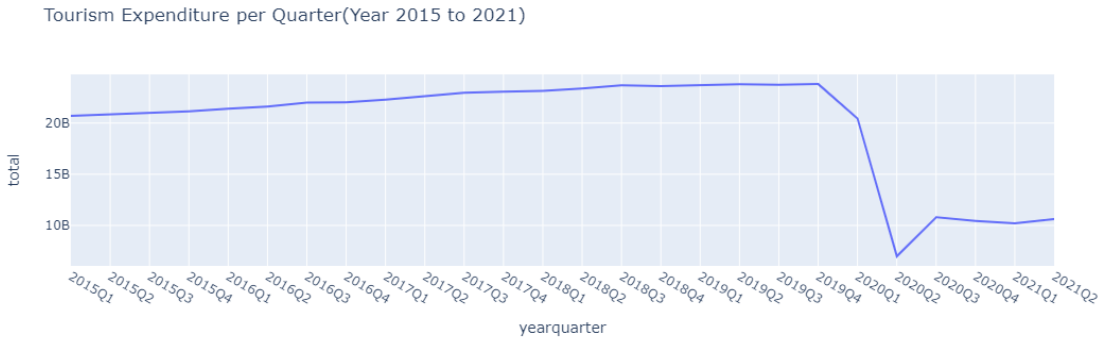
```
q8 = pd.read_sql_query('select year,quarter,concat(year,quarter)yearquarter, (sum(expenditure)*1000000) total\
from tourism_expenditure group by  year,quarter;', engine)
display (q8.head(5))
fig8 = px.line(q8, x="yearquarter", y="total", title='Tourism Expenditure per Quarter(Year 2015 to 2021)')
fig8.show()
```

	year	quarter	yearquarter	total
0	2015	Q1	2015Q1	2.069500e+10
1	2015	Q2	2015Q2	2.086200e+10
2	2015	Q3	2015Q3	2.099400e+10
3	2015	Q4	2015Q4	2.114600e+10
4	2016	Q1	2016Q1	2.141400e+10

Plots







The drop in employment appears to be proportional to tourism expenditure caused by the covid 19 pandemic. This shows that employment in tourism is affected by tourism expenditure also.
But we cannot rule out other factors caused by covid 19, such as the lockdown enforced by the government which was because of the covid 19 pandemic. This also had a direct effect on employment in the tourism sector.
A deeper dive into the various tourism industry will provide more insight as to how jobs were affected by the pandemic

Which category/industry experienced the most effect on employment?
We decided to get the total number of jobs, based on the different tourism industries provided by the data. This will help us identify which sector was most affected.
We also decided to compare this insight with the tourism expenditure per tourism industry.
Employment by tourism script:

```
q6 = pd.read_sql_query('select y.groupname, x.year,x.quarter,concat(x.year,x.quarter)\
yearquarter, (sum(jobs)*1000) total  from tourism_employment x, \
(select a.*, b.groupname from tourism_category a, tourism_catgroup b where a.groupid=b.groupid)\
y where x.categoryid=y.categoryid group by x.quarter,x.year,y.groupname order by x.year ;'
                       , engine)
display (q6.head(5))
fig6 = px.line(q6, x="yearquarter", y="total", color="groupname",title='Total Jobs per Group per Quarter(Year 2015 to 2021)')
fig6.show()
```

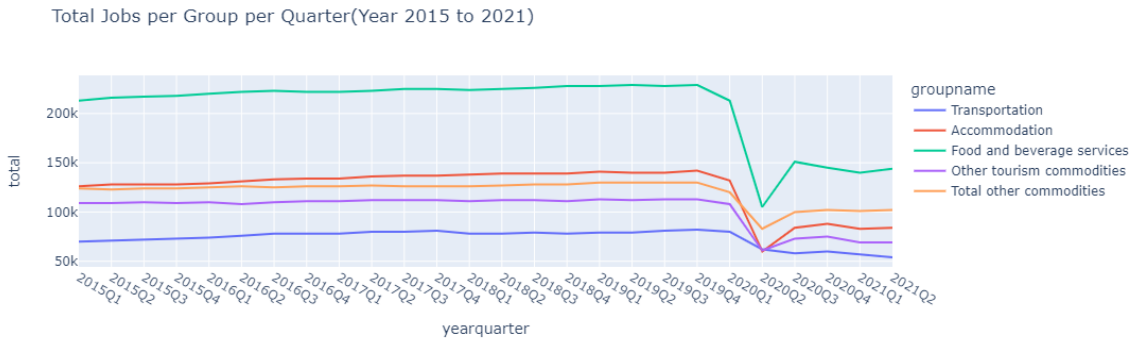
	groupname	year	quarter	yearquarter	total
0	Transportation	2015	Q1	2015Q1	70000.0
1	Accommodation	2015	Q1	2015Q1	126000.0
2	Food and beverage services	2015	Q1	2015Q1	213000.0
3	Other tourism commodities	2015	Q1	2015Q1	109000.0
4	Total other commodities	2015	Q1	2015Q1	124000.0

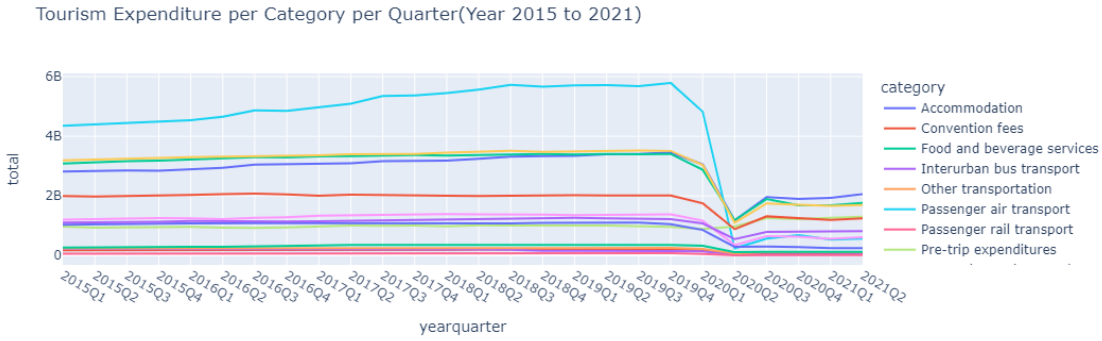
Tourism expenditure by sub-categories script

```
q8b = pd.read_sql_query('select category,year,quarter,concat(year,quarter)yearquarter, (sum(expenditure)*1000000) total\
from tourism_expenditure group by  year,quarter,category order by year,quarter;', engine)
display (q8b.head(5))
fig8b = px.line(q8b, x="yearquarter", y="total",color='category', title='Tourism Expenditure per Category per Quarter(Year 2015 to 2021)')
fig8b.show()
```

	category	year	quarter	yearquarter	total
0	Accommodation	2015	Q1	2015Q1	2.816000e+09
1	Convention fees	2015	Q1	2015Q1	1.730000e+08
2	Food and beverage services	2015	Q1	2015Q1	3.088000e+09
3	Interurban bus transport	2015	Q1	2015Q1	2.170000e+08
4	Other transportation	2015	Q1	2015Q1	2.360000e+08

Plots





The food and beverage industry had the highest number of jobs and experienced the biggest drop in employment, followed by the accommodation industry.
Although transportation expenditure was affected most by the pandemic, the drop in employment is minimal compared to the drop in employment for food and beverage services, accommodation, and other commodities but as we can see from the plot, Employment in the tourism transportation sector still appears to be on the decline even though it did not experience a very sharp drop in employment figures compared to the food and beverage services industry.

Has there been any improvement in tourism expenditure and employment?
Based on the data collected as of 2021, Q2, and as seen from the previous research questions, there has been no significant improvement in both employment and expenditure in tourism.

The domestic tourism experienced a small rise in expenditure after the decline, but this cannot be considered significant compared to the expenditure as of 2019.
The employment generated by tourism also has some slight improvement after the initial drop numbers but employment in the transportation sector as of 2021 Q2 was still on a decline.
If we could get more recent data on tourism activities, we could have provided a better conclusion to this research question.

In general, additional data for 2021 Q3 and Q4 could have provided a better insight into the tourism expenditure and employment after the relaxation of certain government policies. It would also have been good to also get more information on the timeline certain government policies were introduced and compare to the tourism expenditure and employment trend.

Air Travel - Data Exploration and Discussion (Victor Nzemeke)

Data setup in Database:
To explore the data, the data was uploaded into the database. The flight data was defined in the database using *mysql.connector* package while *SQL Alchemy* was used to load the datasets into the database. The code for accomplishing this is as follows:

Flight Dataset was uploaded to the database using code shown below:

```
create_statement = '''CREATE TABLE IF NOT EXISTS flights(
ref_date DATETIME(4),
airports VARCHAR(255),
flight_type VARCHAR(255),
value INT,
id INT(255),
PRIMARY KEY (id)
)'''

cursor = myconnection.cursor(buffered=True)
cursor.execute(create_statement)
cursor.close

#flights.to_sql(name ='flights', con = engine, if_exists='replace', index = False, chunksize = 500)
flights.to_sql(name ='flights', con = engine_group, if_exists='replace', index = False, chunksize = 500)
engine.dispose()
engine_group.dispose()
```

Figure 2 - Code for insertion of Flight Dataset into Database

For airline Dataset, the following code was used:

```
#create table in  database to store airline data:
create_statement = '''CREATE TABLE IF NOT EXISTS airline_data(
ref_date DATETIME,
ops_stats VARCHAR(255),
scalar_factor VARCHAR(255),
value FLOAT,
id INT,
PRIMARY KEY (id)
)'''

cursor = myconnection.cursor(buffered=True)
cursor.execute(create_statement)
cursor.close

#load data into database using engine
#airline_data.to_sql(name ='airline_data', con = engine, if_exists='replace', index = False, chunksize = 500)
airline_data.to_sql(name ='airline_data', con = engine_group, if_exists='replace', index = False, chunksize = 500)
engine.dispose()
engine_group.dispose()
```

Figure 3 - Code for insertion of Airline Dataset

1. **Effect of Covid on Flight Numbers:**
   The Flight Dataset provides information for many airports. Some of these airports only cater to local flights, some are merely airfields for aircraft enthusiasts. For this reason, for flight numbers, we would be taking four main airports in the East and West coasts of Canada to explore and visualize flight numbers since these are the main hubs that serve both international flights and local flights in the country. These airports are Montreal, Toronto-Pearson Airport,

Vancouver Airport and Calgary International Airport. We also believe that these airports can provide good representative data samples for air travel data for Canada. Also, we restricted the query to only show results for civil flights.

```
#Research Questions:

#How did Covid-19 affect Flight Numbers in Canada?
#To do this, pull the flight numbers for 22018 - 2021. Compare side-by-side

#Overall Effect

query1 = '''
SELECT
flight_type, monthname(ref_date) as 'Month', year(ref_date) as 'Year', sum(value) as "No. of Movements"
FROM
    flights
WHERE
    (airports LIKE '%Calgary International%'
        OR airports LIKE '%Pearson%'
        OR airports LIKE '%montreal/Pierre%'
        OR airports LIKE '%Vancouver International%')
        AND (ref_date BETWEEN '2018-01-01' AND '2021-12-31')
        AND (flight_type = 'Total Civil Movements')
GROUP BY month(ref_date), year(ref_date)
'''



#overall_effect = pd.read_sql(query1, engine)
overall_effect = pd.read_sql(query1, engine_group)
display(overall_effect.head())
overall_effect['Year'] = overall_effect['Year'].astype(str)

fig = px.bar(overall_effect, x = 'Month', y = "No. of Movements", color = 'Year', barmode= 'group' )
fig.show()
```

Figure 4 - Code for Extraction and Visualization of Flight Numbers

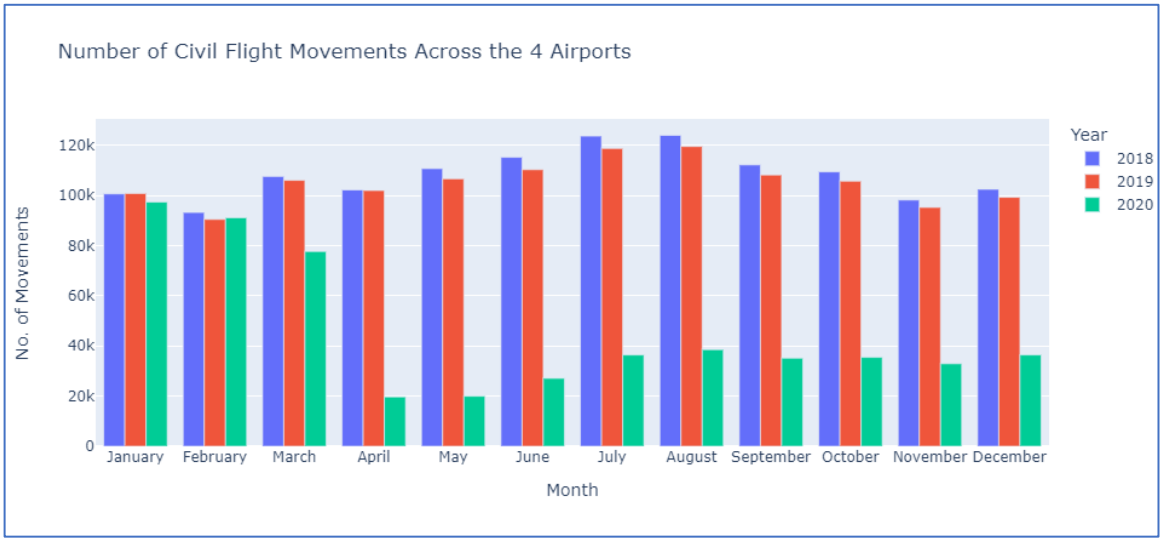


Figure 5 Bar chart Showing number of flight movements across 4 major airports

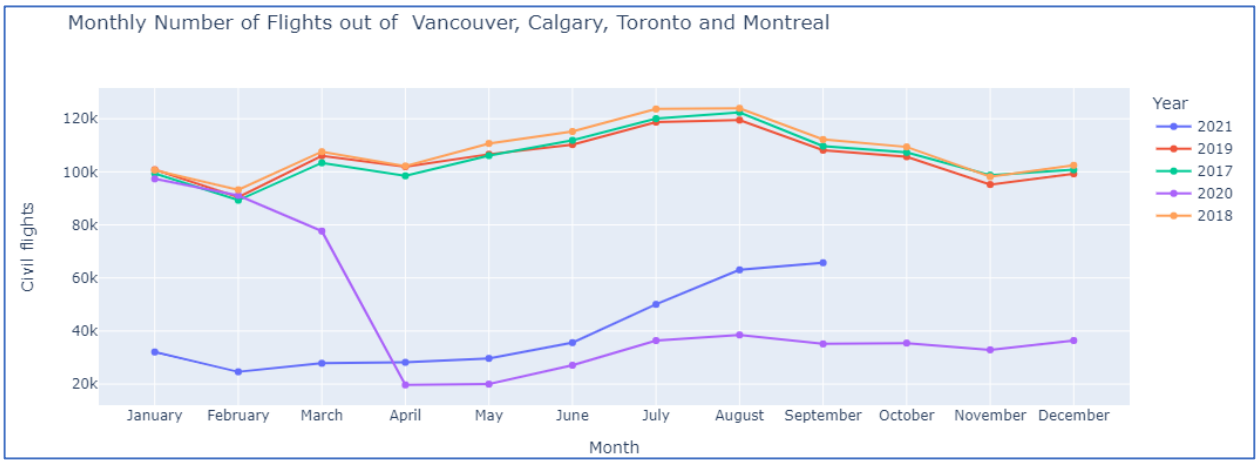
The same data obtained from the query can be presented in the form of a trendline. This time, we decided to go back a few years to get further confirmation that covid did affect flight numbers.

```
#Get the civil flights trend for the year (without covid data)


query2 = '''
SELECT
  monthname(ref_date) as 'Month', year(ref_date) as 'Year', sum(value) as 'Civil flights'
FROM
    flights
WHERE
    (airports LIKE '%Calgary International%'
        OR airports LIKE '%Pearson%'
        OR airports LIKE '%montreal/Pierre%'
        OR airports LIKE '%Vancouver International%')
        and
        (ref_date BETWEEN '2017-01-01' AND '2021-12-31')
        AND (flight_type = 'Total Civil Movements')
GROUP BY monthname(ref_date), year(ref_date)
ORDER BY month(ref_date)

'''

#Civil_flights = pd.read_sql(query2, engine)
Civil_flights = pd.read_sql(query2, engine_group)
engine.dispose()
engine_group.dispose()


display(Civil_flights.head())

fig = px.line(Civil_flights, y="Civil flights", x="Month", color='Year',
                labels = {"x": "Months", "Running Total": "Total Number of Active Hosts"},
                title = "Monthly Number of Flights out of  Vancouver, Calgary, Toronto and Montreal")
fig.update_traces(mode = 'markers+lines')
fig.show()




'''
import plotly.graph_objects as go
animals=['giraffes', 'orangutans', 'monkeys']

fig = go.Figure(data=[
    go.Bar(name='SF Zoo', x=animals, y=[20, 14, 23]),
    go.Bar(name='LA Zoo', x=animals, y=[12, 18, 29])
])
# Change the bar mode
fig.update_layout(barmode='group')
fig.show()
'''
```

Figure 6 - Code for Trendline from 2017 - 2021

	Month	Year	Civil flights
0	January	2019	100790.0
1	January	2017	99378.0
2	January	2020	97343.0
3	January	2018	100618.0
4	January	2021	32099.0

Figure 7 - Output of Code from Fig. 6



```
#Get the civil flights trend for the year (without covid data)


query2 = '''
SELECT
  monthname(ref_date) as 'Month', year(ref_date) as 'Year', sum(value) as 'Civil flights'
FROM
    flights
WHERE
    (airports LIKE '%Calgary International%'
        OR airports LIKE '%Pearson%'
        OR airports LIKE '%montreal/Pierre%'
        OR airports LIKE '%Vancouver International%')
        and
        (ref_date BETWEEN '2017-01-01' AND '2021-12-31')
        AND (flight_type = 'Total Civil Movements')
GROUP BY monthname(ref_date), year(ref_date)
ORDER BY month(ref_date)

'''

#Civil_flights = pd.read_sql(query2, engine)
Civil_flights = pd.read_sql(query2, engine_group)
engine.dispose()
engine_group.dispose()


display(Civil_flights.head())

fig = px.line(Civil_flights, y="Civil flights", x="Month", color='Year',
              labels = {"x": "Months", "Running Total": "Total Number of Active Hosts"},
              title = "Monthly Number of Flights out of  Vancouver, Calgary, Toronto and Montreal")
fig.update_traces(mode = 'markers+lines')
fig.show()
```

Figure 8 - Code for Extracting Flight Numbers Trends Across 4 Major Airports

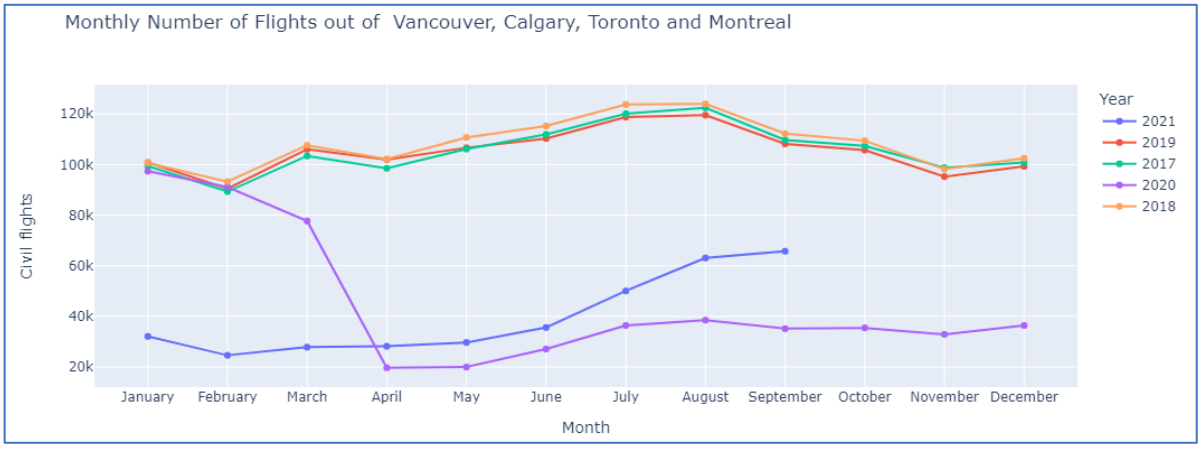


Figure 9 - Trendlines for Number of Flights between 2017 and 2021

Discussion of Results:

From both visualizations, we can tell that number of flights leaving and arriving the 4 major airports suffered a significant drop between March 2020 and April 2020. They remained historically low compared to the numbers from the previous 3 years. Taking April 2019, as a reference point, we see 81% drop from 107,528 to just 19,706 flight movements for the same month in 2020.

2. Effect of Covid on Passenger Numbers:

Similar to the number of flights per month explored in the previous section, we are also interested in exploring the effects of covid on passenger numbers. The following block of code was used to extract the Passenger numbers from the Airline Dataset:

```
#Explore and Visualize the number of passengers per month per year from the data:

passenger_trend = ''' SELECT
    MONTHNAME(ref_date) AS 'Month',
    Year(ref_date) as 'Year',
    SUM(value*1000) AS 'Passenger Numbers'
FROM
    airline_data
WHERE
    ops_stats = 'Passengers'
        AND (ref_date BETWEEN '2017-01-01' AND '2021-12-31')
GROUP BY monthname(ref_date), Year(ref_date)
ORDER BY month(ref_date)
'''

passenger_nos = pd.read_sql(passenger_trend, engine)
display(passenger_nos.head())
display(passenger_nos.tail())
engine.dispose()



trend = px.line(passenger_nos, x = 'Month', y = "Passenger Numbers", color = 'Year',
                title = "Monthly Number of Passengers")
trend.update_traces(mode = 'markers+lines')
trend.show()
```

✓ 0.6s

Figure 10 – Code to extract Data for Passenger Numbers Per Month before and during Covid

	Month	Year	Passenger Numbers
0	January	2017	6243000.0
1	January	2018	6594000.0
2	January	2019	6882000.0
3	January	2020	7092000.0
4	January	2021	797000.0

	Month	Year	Passenger Numbers
52	November	2020	781000.0
53	December	2017	6611000.0
54	December	2018	6928000.0
55	December	2019	7172000.0
56	December	2020	934000.0

Figure 111 - Output of Code from Figure 10.

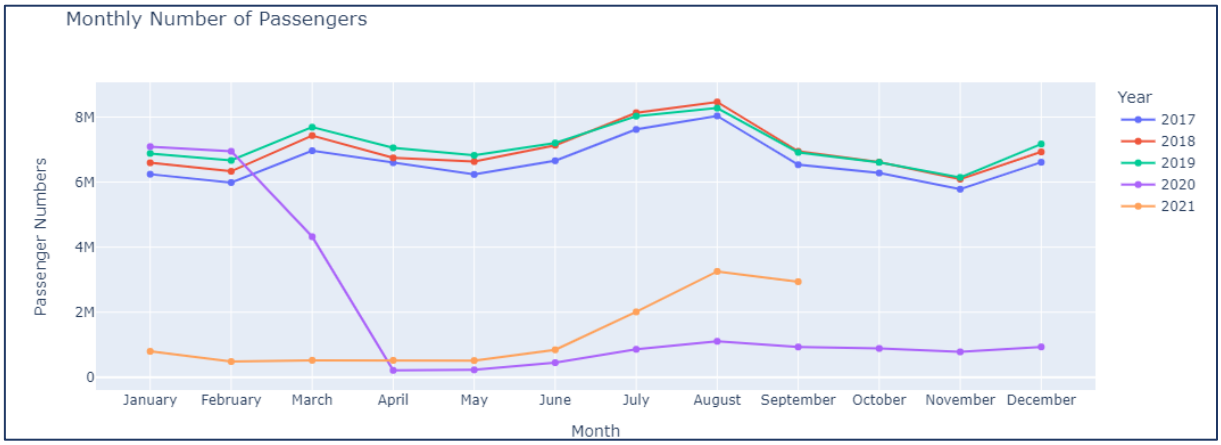


Figure 12 - Trendline for Number of Passengers Per Month

The passenger counts in 2020 took a nosedive like the number of flight movements. Again, taking April 2020 as reference to compare with the previous year, we see a **97.3% drop in passenger numbers**. From this perspective, we conclude that covid sincerely affected passenger numbers in Canada.

3. Year 2020 in Perspective

Since we have established that we see an overall impact on the aviation sector compared to previous years, we proceed to focus on the year 2020, to track the effect of covid as case numbers grew. To do this, we performed a database join statement to merge the base covid dataset with the Flight Dataset. Code is shown as follows:

```
#Comparing Year 2020 Passenger Counts and Number of Cases:

join_query = '''SELECT
    t2.monthNum,
    t1.Month,
    t2.Year,
    t1.No_of_Cases,
    t2.Passenger_Numbers
FROM
    (SELECT
        MONTHNAME(SummaryDate) AS 'Month',
            YEAR(SummaryDate),
            SUM(TotalCases) AS 'No_of_Cases'
    FROM
        ca_covid
    WHERE (SummaryDate between '2020-01-01' and '2020-12-31')
    GROUP BY MONTHNAME(SummaryDate)) t1
        JOIN
    (SELECT
        month(ref_date) as 'monthNum',
        MONTHNAME(ref_date) AS 'Month',
            YEAR(ref_date) AS 'Year',
            SUM(value*1000) AS 'Passenger_Numbers'
    FROM
        airline_data
    WHERE
        ops_stats = 'Passengers'
            AND (ref_date BETWEEN '2020-01-01' AND '2020-12-31')
    GROUP BY MONTHNAME(ref_date) , YEAR(ref_date)) t2 ON t1.Month = t2.Month
    order by monthNum
    '''
year_effect = pd.read_sql(join_query, engine)

year_effect.drop("Month", axis = 1)

display(year_effect)

✓ 0.9s
```

Figure 12 - Code to Join Passenger Counts and No. of Cases

Output:

	monthNum	Month	Year	No_of_Cases	Passenger_Numbers
0	1	January	2020	34.0	7092000.0
1	2	February	2020	494.0	6946000.0
2	3	March	2020	106502.0	4322000.0
3	4	April	2020	1832416.0	213000.0
4	5	May	2020	4656358.0	227000.0
5	6	June	2020	5932034.0	449000.0
6	7	July	2020	6805040.0	860000.0
7	8	August	2020	7589844.0	1106000.0
8	9	September	2020	8458061.0	931000.0
9	10	October	2020	12097028.0	886000.0
10	11	November	2020	18077654.0	781000.0
11	12	December	2020	28629700.0	934000.0

Figure 13 - Output of Code in Fig 12

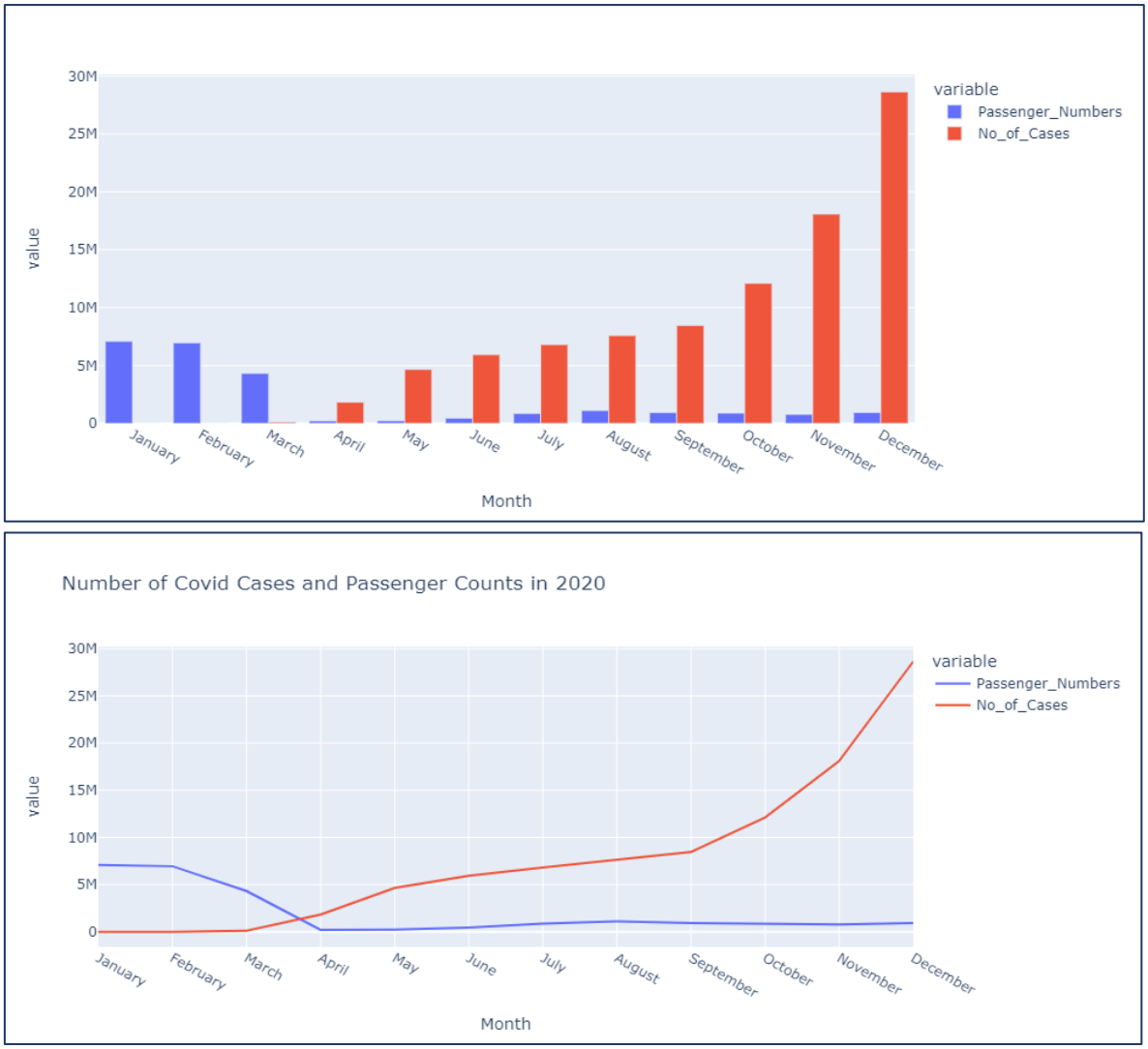


Figure 14 - Covid Numbers and Passenger Counts

The same information is presented in both figures above, showing that with the rise in covid cases in 2020, the passenger numbers reduced and remained low for the rest of the year.

4. Looking Ahead, 2021.

Since Vaccinations took off late 2020, we also decided to look at the effects of vaccinations on airline passenger numbers for 2021. The code for extracting the required data is as shown below:

```
#Comparing Year 2020 Passenger Counts and Vaccinations:

join_query_vacc = '''SELECT
    t2.monthNum,
    t1.Month,
    t2.Year,

    t1.TotalVaccinated,
    t2.Passenger_Numbers
FROM
    (SELECT
        MONTHNAME(SummaryDate) AS 'Month',
            YEAR(SummaryDate),
            (TotalVaccinated) AS 'TotalVaccinated'
    FROM
        ca_covid
    WHERE (SummaryDate between '2021-01-01' and '2021-12-31')
    GROUP BY MONTHNAME(SummaryDate)) t1
        JOIN
    (SELECT
        month(ref_date) as 'monthNum',
        MONTHNAME(ref_date) AS 'Month',
            YEAR(ref_date) AS 'Year',
            SUM(value*1000) AS 'Passenger_Numbers'
    FROM
        airline_data
    WHERE
        ops_stats = 'Passengers'
            AND (ref_date BETWEEN '2021-01-01' AND '2021-12-31')
    GROUP BY MONTHNAME(ref_date) , YEAR(ref_date)) t2 ON t1.Month = t2.Month
    order by monthNum
    '''

vacc_effect = pd.read_sql(join_query_vacc, engine)
vacc_effect.drop("Month", axis = 1)
display(vacc_effect)
```

Figure 15 - Code for Comparing Passenger Counts and Vaccinations

The Output of the query in Fig 15 is as shown below:

monthNum	Month	Year	TotalVaccinated	Passenger_Numbers
1	January	2021	NaN	797000.0
2	February	2021	5464000.0	483000.0
3	March	2021	7402000.0	522000.0
4	April	2021	21012000.0	520000.0
5	May	2021	28251000.0	514000.0
6	June	2021	32196000.0	844000.0
7	July	2021	37539000.0	2010000.0
8	August	2021	42032000.0	3255000.0
9	September	2021	44564000.0	2939000.0

Figure 16 - Output of code in Fig 15

Data obtained from the query is visualized in Fig. 16. There is a strong correlation between the increase in number of vaccinations administered in 2021 and the recovery of flight numbers for the same period.

Figure 17 - Effect of Vaccinations on Flight Numbers

Ground Travel - Data Exploration and Discussion (Daniel Zhou)

In our data exploration for the ground travel domain, we focus on answering our research questions. Like our other domains, this is done by combining our domain datasets with the COVID datasets in SQL. We chose a graphical approach to answer these questions, by first plotting out the data we get from our SQL queries and then drawing the conclusion from the plots.

Research Question 1: Did COVID-19 affect the number of vehicles crossing the border between Canada and United States?

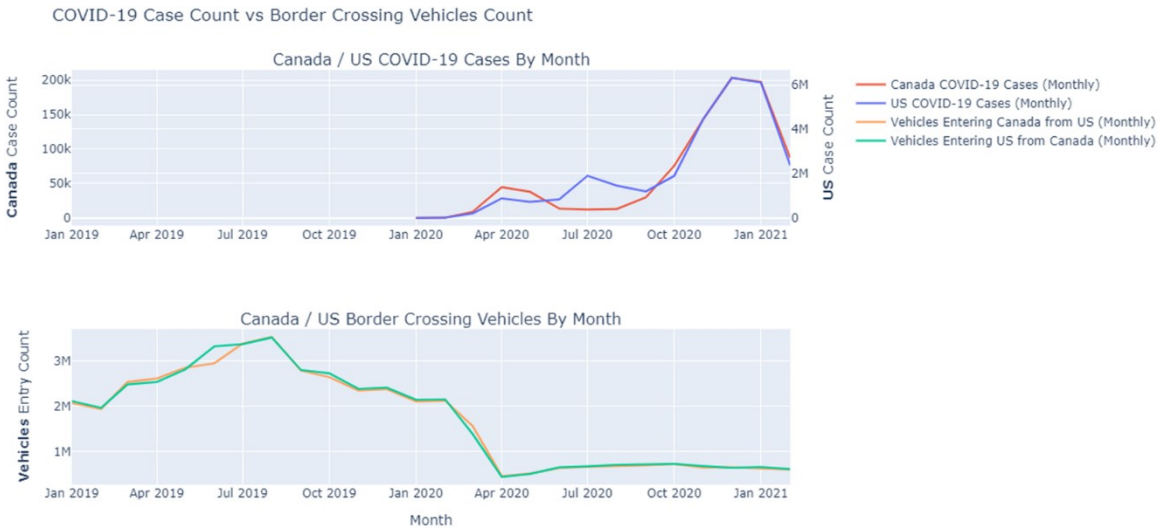
This is the research question that required the largest query. 4 tables were joined together to answer this question. The specific query we used is listed in the figure below.

```
SELECT DATE_FORMAT(STR_TO_DATE(CONCAT(ca_entry.year, ca_entry.month), '%Y%m'),'%Y%m') AS date,
ca_entry_count, us_entry_count, ca_covid_count, us_covid_count FROM
(
    SELECT YEAR(RefDate) as year, MONTH(RefDate) AS month, SUM(Value) AS 'ca_entry_count'
    FROM ca_entry
    WHERE(RefDate BETWEEN '2019-01-01' AND '2021-02-28')
    GROUP BY year, month
) AS ca_entry
LEFT JOIN
(
    SELECT YEAR(SummaryDate) as year, MONTH(SummaryDate) AS month, SUM(DailyTotals) AS 'ca_covid_count'
    FROM ca_covid
    WHERE(SummaryDate BETWEEN '2019-01-01' AND '2021-02-28')
    GROUP BY year, month
) AS ca_covid
ON ca_entry.year=ca_covid.year AND ca_entry.month=ca_covid.month
LEFT JOIN
(
    SELECT YEAR(date) as year, MONTH(date) AS month, SUM(value) AS 'us_entry_count'
    FROM us_entry
    WHERE(date BETWEEN '2019-01-01' AND '2021-02-28') AND(border='US-Canada Border')
    AND(Measure='Buses' OR Measure='Personal Vehicles' OR Measure='Trucks')
    GROUP BY year, month
) AS us_entry
ON ca_entry.year=us_entry.year AND ca_entry.month=us_entry.month
LEFT JOIN
(
    SELECT YEAR(date) as year, MONTH(date) AS month, SUM(positiveIncrease) AS 'us_covid_count'
    FROM us_covid
    WHERE(date BETWEEN '2019-01-01' AND '2021-02-28')
    GROUP BY year, month
) AS us_covid
ON ca_entry.year=us_covid.year AND ca_entry.month=us_covid.month
ORDER BY date asc;
```

This query returned a table containing the vehicle entering Canada / US and COVID count for Canada/US separated by month. A sample of the returned result is shown in the figure below.

date	ca_entry_count	us_entry_count	ca_covid_count	us_covid_count
202001	2108365	2147708	4	2
202002	2123151	2153502	16	16
202003	1564530	1390275	8571	196851
202004	458763	443168	44637	876279
202005	514823	507840	37711	718205

Because this is all done in Python with SQL Alchemy, we can import this table directly into a Pandas Data Frame and plot the results as line chart using Plotly Express as shown in figure below.

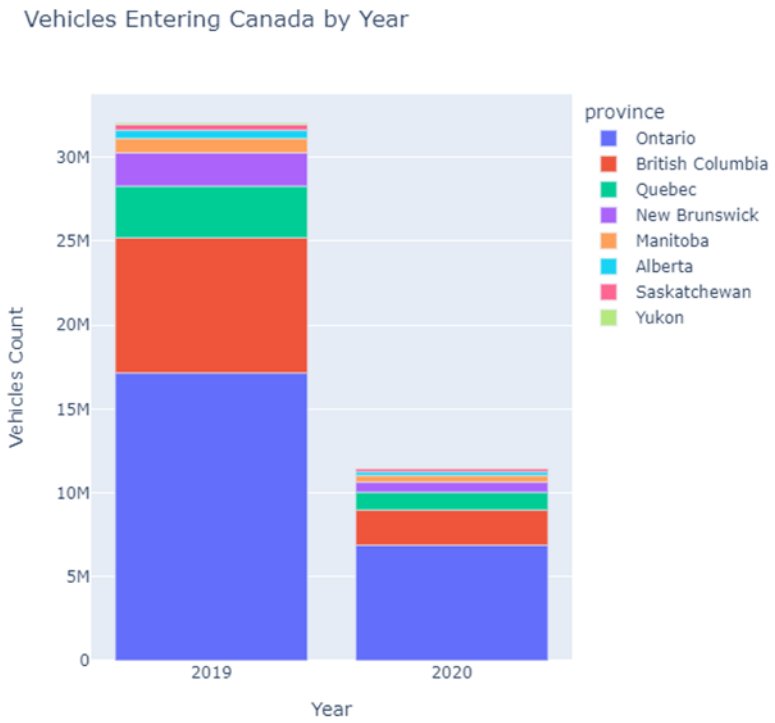


As we can see from the plot, US and Canada have mirrored case counts, although at different scales, the US had a max COVID case count of 6.3M vs Canada's 0.2M. Starting when COVID cases first peaked in both countries in March 2020, ground travel slowed down significantly, as we can see the Vehicles traveling across the border in both directions dropped.

One interesting thing we can see from the plot is that despite the US having significantly more cases, we do not see less vehicles going into the US from Canada as compared to coming into Canada from the US. In fact, the vehicles count going both directions are identical.

We also want to find the exact drop from 2020 compared to 2019. We sum up the travel numbers coming into Canada from 2019 and 2020 using an SQL query and plot out the results in Python as a bar chart shown in the two figures below.

```
SELECT DATE_FORMAT(STR_TO_DATE(ca_entry.year, '%Y'), '%Y') AS year,
ca_entry.Province AS province, ca_entry_count, ca_covid_count FROM
(
    SELECT Province, YEAR(RefDate) as year, SUM(Value) AS 'ca_entry_count'
    FROM ca_entry
    WHERE(RefDate BETWEEN '2019-01-01' AND '2020-12-31') AND Value > 0
    GROUP BY year, Province
) AS ca_entry
LEFT JOIN
(
    SELECT Province, YEAR(SummaryDate) as year, SUM(DailyTotals) AS 'ca_covid_count'
    FROM ca_covid
    WHERE(SummaryDate BETWEEN '2019-01-01' AND '2020-12-31')
    GROUP BY year, Province
) AS ca_covid
ON ca_entry.year=ca_covid.year AND ca_entry.Province=ca_covid.Province
ORDER BY year asc, ca_entry_count desc, Province asc;
```

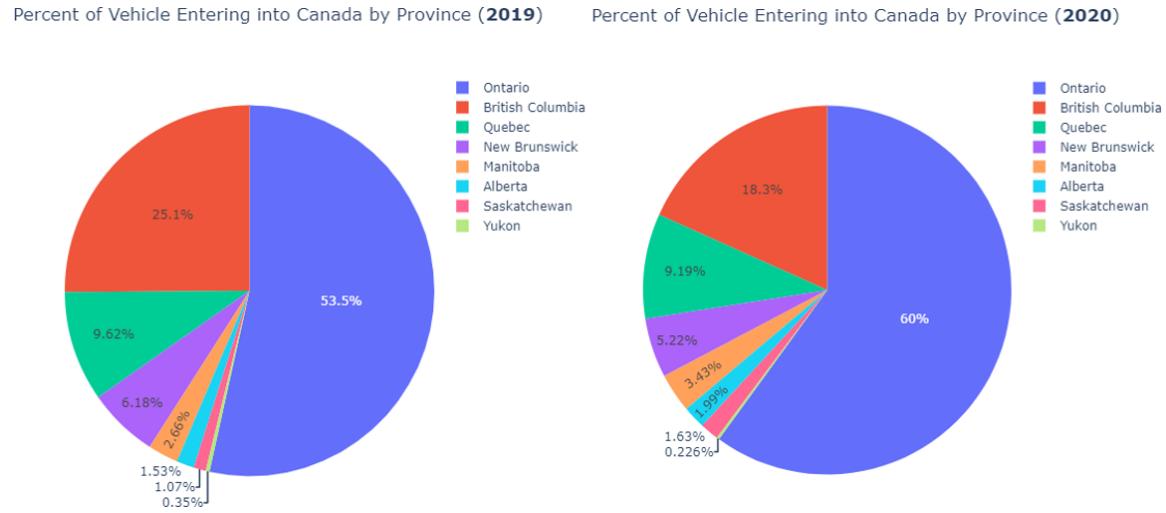


We see that the number of vehicles traveling into Canada dropped from 32.1M in 2019 to 11.5M in 2020, that is a 64% drop year over year.

For this research question, we conclude that there was significant impact from COVID on international ground travel between Canada and United States.

<u>Research Question 2:</u> How does the effect on international ground travel differ at a provincial level?

The result of the query / bar chart from the previous question has shown us that every province saw a drop in vehicle entry in 2020 as compared to 2019. Next, we will plot this out into two separate pie charts to see if the distribution of vehicles coming into Canada has changed between the provinces over these 2 years. The charts are shown below.



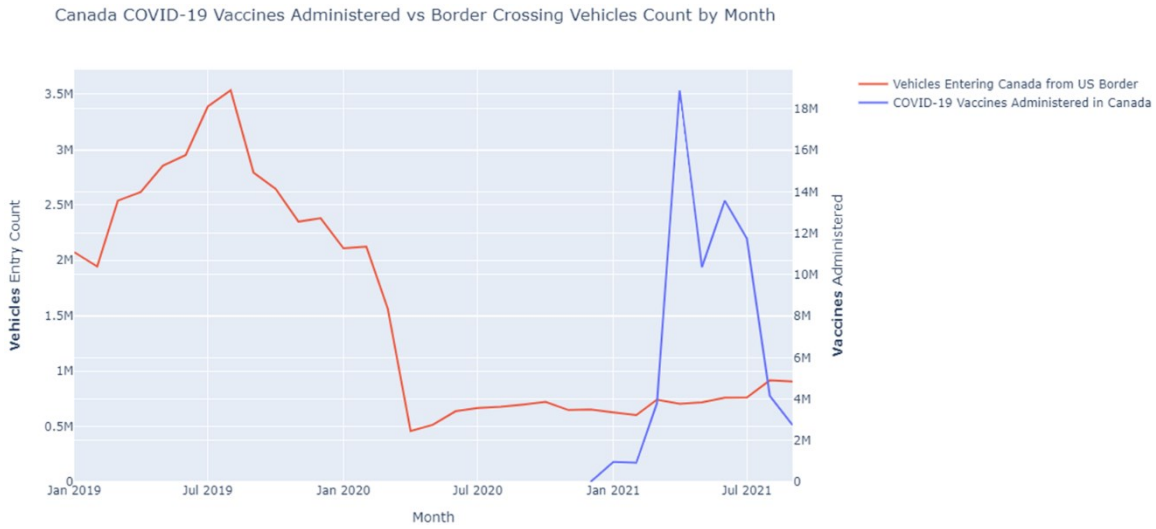
As we can see from the two pie charts above, compared to 2019, 2020 saw a larger proportion of vehicles entering Canada through prairie / central provinces (AB, SK, MB, ON) and a lesser proportion of vehicles entering Canada through the coastal / outer provinces (YT, BC, QC, NB). This may be because more commercial ground travel is done through the central provinces and more recreational ground travel is done through the outer provinces.

For this research question, we conclude that the coastal / outer provinces were impacted more than prairie / central provinces.

Research Question 3: Has vaccine helped with international ground travel numbers recovery?

To answer this question, we took the approach to query for the number of vehicles coming into Canada and compare to the number of COVID cases in that same period. To do this we had to join the Canada COVID table together with the Canada border vehicle entry table. The query and resulting line plot for this are shown below.

```
SELECT DATE_FORMAT(STR_TO_DATE(CONCAT(ca_entry.year, ca_entry.month), '%Y%m'),'%Y%m') AS date,
ca_entry_count, ca_vac_count, ca_dose1_count, ca_dose2_count FROM
(
    SELECT YEAR(RefDate) as year, MONTH(RefDate) AS month, SUM(Value) AS 'ca_entry_count'
    FROM  ca_entry
    WHERE(RefDate BETWEEN '2019-01-01' AND '2021-09-30')
    GROUP BY year, month
) AS ca_entry
LEFT JOIN
(
    SELECT YEAR(SummaryDate) as year, MONTH(SummaryDate) AS month, SUM(DailyVaccinated) AS 'ca_vac_count',
    SUM(DailyDose1) AS 'ca_dose1_count', SUM(DailyDose2) AS 'ca_dose2_count'
    FROM  ca_covid
    WHERE(SummaryDate BETWEEN '2019-01-01' AND '2021-09-30')
    GROUP BY year, month
) AS ca_covid
ON ca_entry.year=ca_covid.year AND ca_entry.month=ca_covid.month
ORDER BY date asc;
```



As we can see from the plot, ground travel numbers have been slowly recovering since it bottomed out during the first COVID wave in April 2020. Vaccinations started in Dec 2020 but the travel increase did not seem to speed up by that.

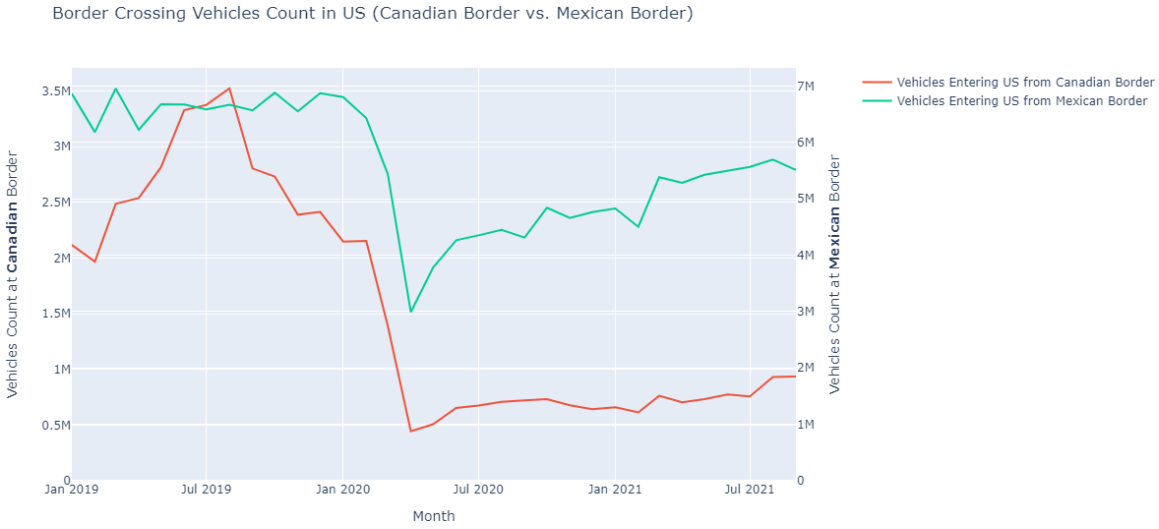
For this research question, we conclude that vaccination has negligible impact on ground travel recovery thus far.

Research Question 4: Do different bordering countries affect ground travel impact and recovery differently?

From our previous question, we saw that US-Canada border ground travel seem to be recovering very slowly. What about the US-Mexico border? Has it also been recovering slowly like the US-Canada border? Does different bordering countries affect the recovery.

Because our US border entry dataset includes data from both borders, we can query for that data and group them by border. Shown below is the query and the resulting line chart plotted in Python.

```
SELECT DATE_FORMAT(STR_TO_DATE(CONCAT(YEAR(date), MONTH(date)), '%Y%m'),'%Y%m') AS date,
border, SUM(value) AS 'us_entry_count'
FROM  us_entry
WHERE(date BETWEEN '2019-01-01' AND '2021-09-30')
AND(Measure='Buses' OR Measure='Personal Vehicles' OR Measure='Trucks')
GROUP BY date, border
ORDER BY date, border asc;
```



From the plot, we can see that the US-Mexico border vehicle count did not drop as drastically as the US-Canada border, it also recovered much faster.

For this research question, we conclude that international ground travel at the US-Mexico border recovered much quicker than at the US-Canada border and that different bordering countries do affect ground travel impact and recovery due to COVID.

Overall, what I learnt from this project is that having clean data is very important. For this domain, the data from the United States were much easier to work with than the data from Canada. The Canadian data had a lot of redundant and repetitive information that required a lot of data cleaning before we could do analysis with them.

As for technologies, I still believe that relational database would work best with our data, as dealing with this data in a NoSQL database would be very difficult due to the sheer magnitude of relationships, and since our data already has predefined columns, it fits very well in a relational database.

To extend the project in the future, I would look into getting more information on government policies during the COVID-19 pandemic to see exactly which policies impacted ground travel the most. Also, I would look into joining ground travel data with air travel data to do some analysis and see if the impact is mirrored.

Electricity - Data Exploration and Discussion (Spencer Dow)

With this data present in the data base, we can use SQL to pull the desired pieces out. For the electricity data we ran two distinct queries to pull out the information we needed. The first query would pull out the monthly values for the total energy production. This query would return the value, year, month, location and a combined year and month. This query found below.

```
energy = pd.read_sql("""
            SELECT value , YEAR(date), MONTH(date), geo, DATE_FORMAT(date, '%M %Y') AS 'Date'
            FROM energy_prod
            Where type = 'Total all types of electricity generation'
            Group By YEAR(date), MONTH(date), geo
            Order By YEAR(date),MONTH(date);
            """, con = credentials)
```

The next query was designed to analysis the change in electricity type over the years. This query would return the value, year, month, location, electricity generation type and a combined year and month. This query can be found below.

```
totalEnergy = pd.read_sql("""
            SELECT value , YEAR(date), MONTH(date), geo, type, DATE_FORMAT(date, '%M %Y') AS 'Date'
            FROM energy_prod
            Where type NOT LIKE '%Total%'
            Group By YEAR(date), MONTH(date), geo, type
            Order By YEAR(date),MONTH(date);
            """, con = credentials)
```

Now we need to pull some Covid data to pair with the electricity data. Unlike our electricity data there is no Canada wide data so we will need to group them ourselves. Since we are also interested in provincial data, we will need two queries. Both will return the year, month, the sum of the daily totals and a month year combination. The query below returns the results per province.

```
covid = pd.read_sql("""
            SELECT Province , YEAR(SummaryDate), MONTH(SummaryDate), SUM(DailyTotals)
            FROM ca_covid
            Where Province != 'Repatriated Cdn'
            Group By MONTH(SummaryDate), YEAR(SummaryDate), Province
            Order By YEAR(SummaryDate), month(SummaryDate)
            """, con = credentials)
```

The query below returns the Canada wide data.

```
totalCovid = pd.read_sql("""
            SELECT YEAR(SummaryDate), MONTH(SummaryDate), SUM(DailyTotals), DATE_FORMAT(SummaryDate, '%M %Y') AS 'Date'
            FROM ca_covid
            Where Province != 'Repatriated Cdn'
            Group By MONTH(SummaryDate), YEAR(SummaryDate)
            Order By YEAR(SummaryDate), month(SummaryDate)
            """, con = credentials)
```

We have opted to join the data using python over joining from the query because of the amount of group members we have. To join the data, we need to create a key because none of the variables given would be appropriate. The key we used to join them was a combination of the year, month, and province. The data for all of Canada was easier to join because we could just use the date. We now have Four panadas data frames that will be used for analysis. Samples of each can be found below.

Joining and key creation code:

```
key = []
for i in range(len(energy)):
    key.append(str(energy['YEAR(date)'][i]) + "-" + str(energy['MONTH(date)'][i]) + "-" + energy['geo'][i])

energy["key"] = key

key = []

for i in range(len(covid)):
    key.append(str(covid['YEAR(SummaryDate)'][i]) + "-" + str(covid['MONTH(SummaryDate)'][i]) + "-" + covid['Province'][i])

covid["key"] = key

key = []
for i in range(len(totalEnergy)):
    key.append(str(totalEnergy['YEAR(date)'][i]) + "-" + str(totalEnergy['MONTH(date)'][i]) + "-" + totalEnergy['geo'][i])

totalEnergy["key"] = key

natCovidEnergy = energy[energy['geo'] == 'Canada']
natCovidEnergyTypes = totalEnergy[totalEnergy['geo'] == 'Canada']

covidWithEnergy = energy.join(covid.set_index('key'), on='key')
covidWithTypesEnergy = totalEnergy.join(covid.set_index('key'), on='key')
natCovidWithEnergy = natCovidEnergy.join(totalCovid.set_index('Date'), on='Date')
natCovidWithTypesEnergy = natCovidEnergyTypes.join(totalCovid.set_index('Date'), on='Date')
```

National covid cases and Electricity generation by type:

	value	YEAR(date)	MONTH(date)	geo	type	Date	key	YEAR(SummaryDate)	MONTH(SummaryDate)	SUM(DailyTotals)
15	0.0	2016	1	Canada	Combustion turbine	January 2016	2016-1-Canada	NaN	NaN	NaN
16	0.0	2016	1	Canada	Conventional steam turbine	January 2016	2016-1-Canada	NaN	NaN	NaN
17	38747900.0	2016	1	Canada	Hydraulic turbine	January 2016	2016-1-Canada	NaN	NaN	NaN
18	0.0	2016	1	Canada	Internal combustion turbine	January 2016	2016-1-Canada	NaN	NaN	NaN
19	8924260.0	2016	1	Canada	Nuclear steam turbine	January 2016	2016-1-Canada	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...
6853	8070670.0	2021	8	Canada	Nuclear steam turbine	August 2021	2021-8-Canada	2021.0	8.0	68367.0
6854	14460.0	2021	8	Canada	Other types of electricity generation	August 2021	2021-8-Canada	2021.0	8.0	68367.0
6855	300715.0	2021	8	Canada	Solar	August 2021	2021-8-Canada	2021.0	8.0	68367.0
6856	0.0	2021	8	Canada	Tidal power turbine	August 2021	2021-8-Canada	2021.0	8.0	68367.0
6857	1994460.0	2021	8	Canada	Wind power turbine	August 2021	2021-8-Canada	2021.0	8.0	68367.0

National covid cases and Electricity generation total:

	value	YEAR(date)	MONTH(date)	geo	Date	key	key2	YEAR(SummaryDate)	MONTH(SummaryDate)	SUM(DailyTotals)
2	64228200.0	2016	1	Canada	January 2016	2016-1-Canada	1-Canada	NaN	NaN	NaN
16	57932200.0	2016	2	Canada	February 2016	2016-2-Canada	2-Canada	NaN	NaN	NaN
30	57108200.0	2016	3	Canada	March 2016	2016-3-Canada	3-Canada	NaN	NaN	NaN
44	49985000.0	2016	4	Canada	April 2016	2016-4-Canada	4-Canada	NaN	NaN	NaN
58	47876200.0	2016	5	Canada	May 2016	2016-5-Canada	5-Canada	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...
884	48555500.0	2021	4	Canada	April 2021	2021-4-Canada	4-Canada	2021.0	4.0	237305.0
898	48110200.0	2021	5	Canada	May 2021	2021-5-Canada	5-Canada	2021.0	5.0	162155.0
912	48940000.0	2021	6	Canada	June 2021	2021-6-Canada	6-Canada	2021.0	6.0	33702.0
926	49931700.0	2021	7	Canada	July 2021	2021-7-Canada	7-Canada	2021.0	7.0	15541.0
940	50334200.0	2021	8	Canada	August 2021	2021-8-Canada	8-Canada	2021.0	8.0	68367.0

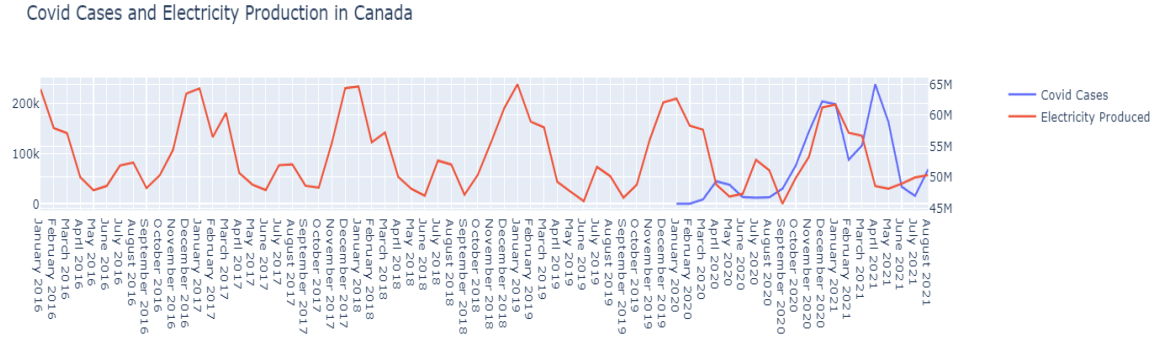
Provincial covid cases and Electricity generation by type:

	value	YEAR(date)	MONTH(date)	geo	type	Date	key	Province	YEAR(SummaryDate)	MONTH(SummaryDate)	SUM(DailyTotals)
0	0.0	2016	1	Alberta	Combustion turbine	January 2016	2016-1-Alberta	NaN	NaN	NaN	NaN
1	0.0	2016	1	Alberta	Conventional steam turbine	January 2016	2016-1-Alberta	NaN	NaN	NaN	NaN
2	143297.0	2016	1	Alberta	Hydraulic turbine	January 2016	2016-1-Alberta	NaN	NaN	NaN	NaN
3	0.0	2016	1	Alberta	Internal combustion turbine	January 2016	2016-1-Alberta	NaN	NaN	NaN	NaN
4	15725.0	2016	1	Alberta	Other types of electricity generation	January 2016	2016-1-Alberta	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...
6931	36872.0	2021	8	Yukon	Hydraulic turbine	August 2021	2021-8-Yukon	Yukon	2021.0	8.0	120.0
6932	0.0	2021	8	Yukon	Internal combustion turbine	August 2021	2021-8-Yukon	Yukon	2021.0	8.0	120.0
6933	0.0	2021	8	Yukon	Other types of electricity generation	August 2021	2021-8-Yukon	Yukon	2021.0	8.0	120.0
6934	0.0	2021	8	Yukon	Solar	August 2021	2021-8-Yukon	Yukon	2021.0	8.0	120.0
6935	0.0	2021	8	Yukon	Wind power turbine	August 2021	2021-8-Yukon	Yukon	2021.0	8.0	120.0

Provincial covid cases and Electricity generation total:

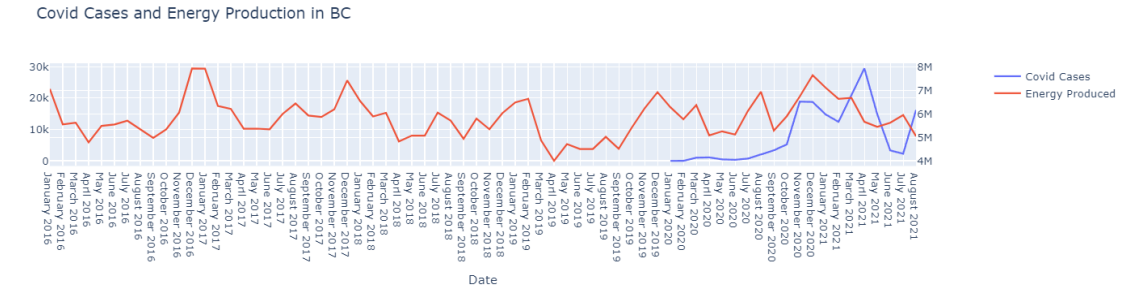
	value	YEAR(date)	MONTH(date)	geo	Date	key	Province	YEAR(SummaryDate)	MONTH(SummaryDate)	SUM(DailyTotals)
0	7689180.0	2016	1	Alberta	January 2016	2016-1-Alberta	NaN	NaN	NaN	NaN
1	7062370.0	2016	1	British Columbia	January 2016	2016-1-British Columbia	NaN	NaN	NaN	NaN
2	64228200.0	2016	1	Canada	January 2016	2016-1-Canada	NaN	NaN	NaN	NaN
3	3123600.0	2016	1	Manitoba	January 2016	2016-1-Manitoba	NaN	NaN	NaN	NaN
4	1514000.0	2016	1	New Brunswick	January 2016	2016-1-New Brunswick	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...
947	13795200.0	2021	8	Ontario	August 2021	2021-8-Ontario	Ontario	2021.0	8.0	15114.0
948	32829.0	2021	8	Prince Edward Island	August 2021	2021-8-Prince Edward Island	Prince Edward Island	2021.0	8.0	25.0
949	17113200.0	2021	8	Quebec	August 2021	2021-8-Quebec	Quebec	2021.0	8.0	11927.0
950	2007720.0	2021	8	Saskatchewan	August 2021	2021-8-Saskatchewan	Saskatchewan	2021.0	8.0	4343.0
951	38575.0	2021	8	Yukon	August 2021	2021-8-Yukon	Yukon	2021.0	8.0	120.0

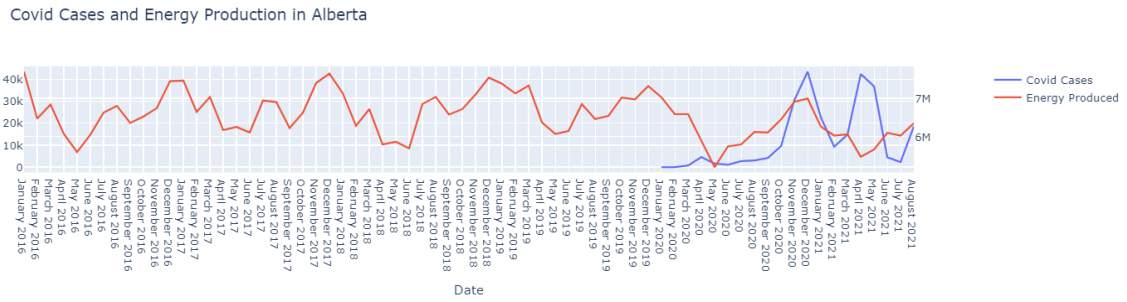
Now that we have the data in a usable format, we can start doing some investigation to answer our research questions. If goal of this domain is to see if Covid had a beneficial effect on our environment through the reduction of electricity generation and see if the source electricity has changed. We will start with analyzing if the pandemic has changed any trends in electricity generation

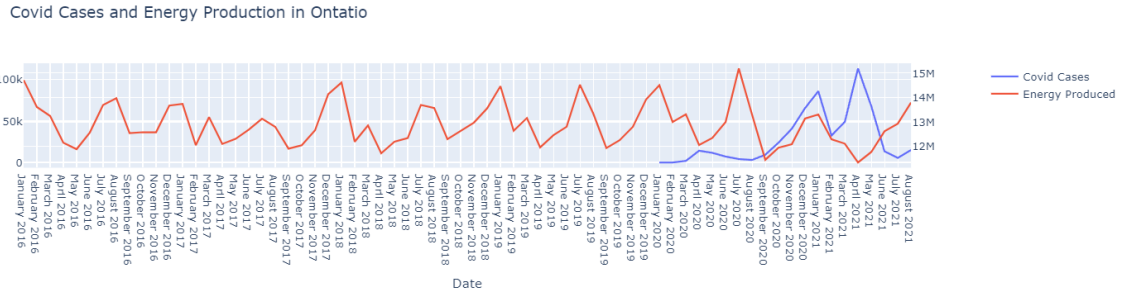


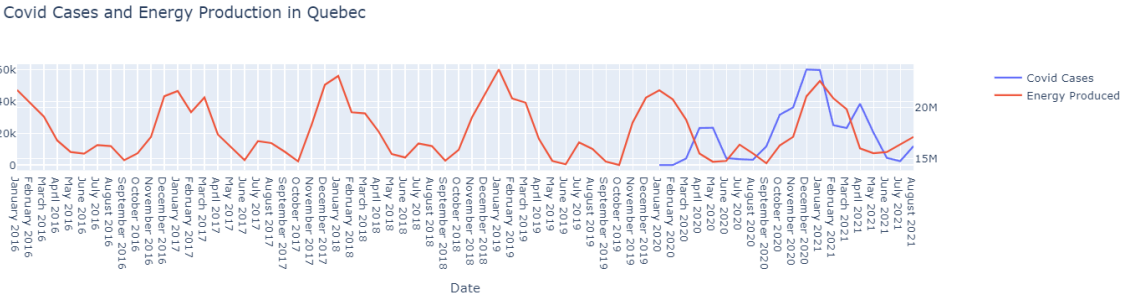
As you can see from the graph the trend is that Canada produces much more electricity in the winter months and has a small spike in the summer months starting around June and ending around September. From the chart we can see that covid does not seem to influence the trends of electricity production. Although one thing to consider is how near the end of the chart we do not experience the same dip and spike we would expect in June/July. It will be interesting to investigate this when more data is released. The amount produced is slightly lover in the peak winter months. In the chart below you can see the average monthly generation compared between pre and during covid.

As you can see there is very little difference between the two lines indicating that Covid 19 did not have a large impact on electricity generation. We will investigate a some of the bigger provinces to see if there was an individual impact.



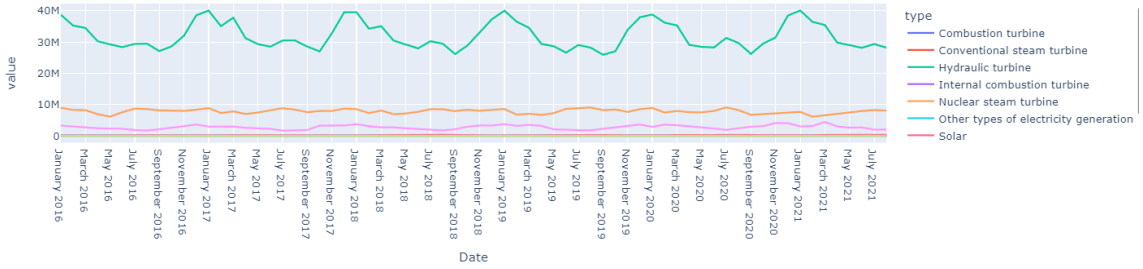




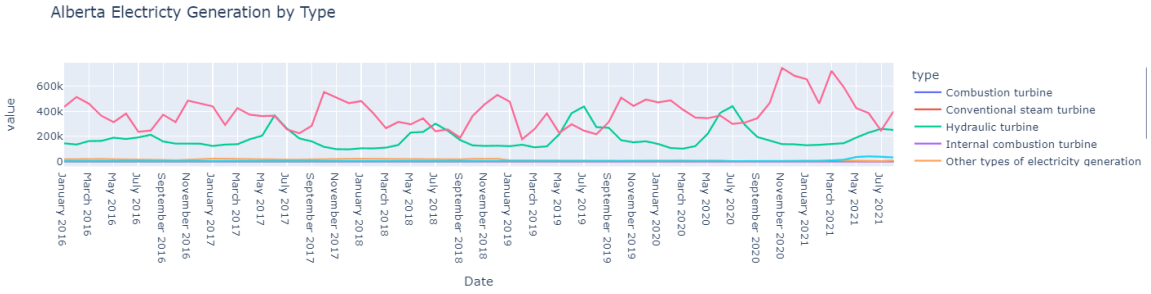


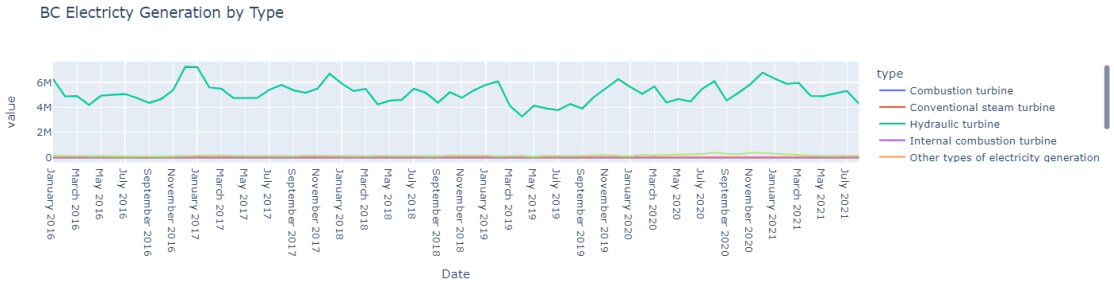
You can see from the trends above that two very important things about electrivity generation, That it is more volitle on a provincial scale. The trend we see for all of canada is heavly mimiced in Qubec chart as it is the largest electricty producer in Canada. It seems that Ontairo and Albertas electricity

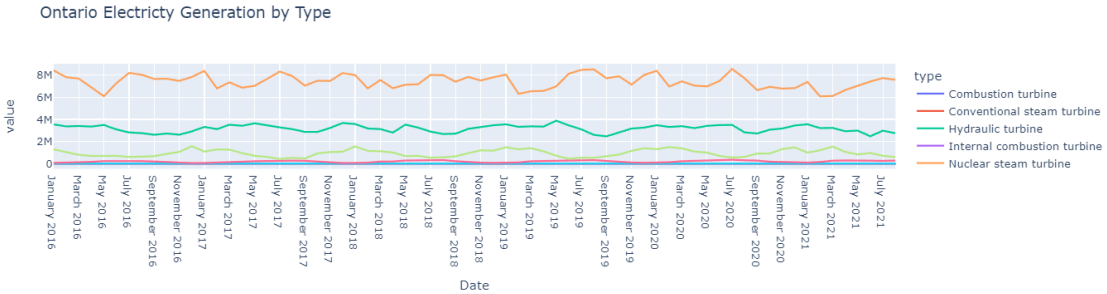
generation took were majorly effected during the pandemic. No we can investigate the if the type of electricy production has changed.

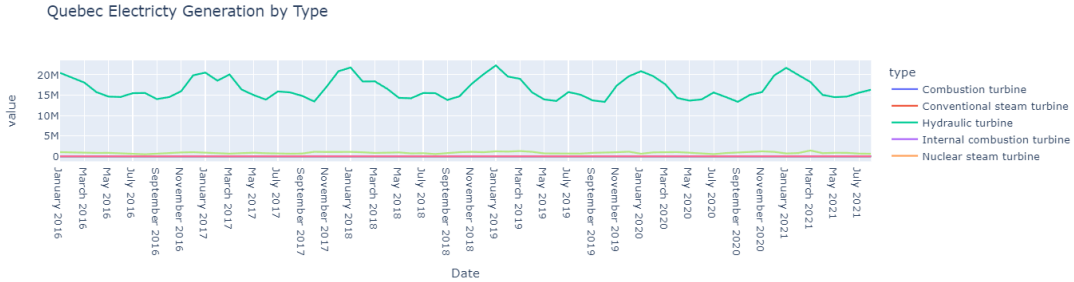


We can see from the chart above that most of Canadian electricity is generated via the Hydraulic turbine. We see no shift in trend at all past 2020 when the pandemic hit Canada. So, we can conclude that on a national scale there is no impact from the pandemic. We will now investigate this issue on a provincial scale.









Here we see the national trend continue down to the provincial level where there is not a lot of change in the source used to make electricity. We can say that even on a provincial level we cannot see an impact on the electricity generation source via Covid 19.

From this project I learned that it is very important to understand the data that you are investigating. For this project I spent a significant portion of the time just to be able to make the correct queries and interpret the results. In terms of technology, I feel like a relational data base was a great way to represent this data and because of the wide scope of our data would not fit well into any of the non-SQL databases we have learned about. For expanding this project in the future, I would change the purpose into a full investigation into the environmental effects of covid-19 in Canada. This would include looking at different industries like Oil and Gas as well as looking at other factors that affect the environment like technological improvements and recycling rates.

Manufacturing - Data Exploration and Discussion (Harman Kaur)

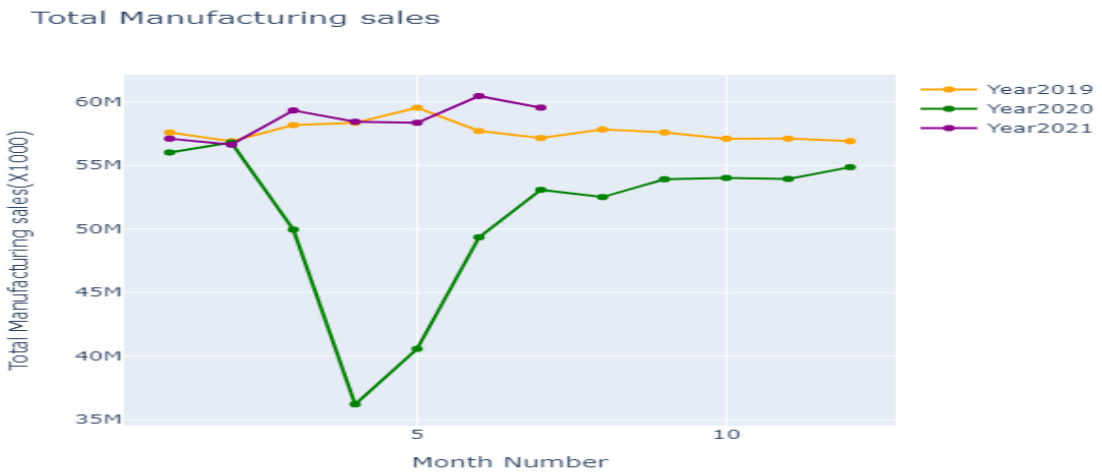
Before moving towards Data Exploration and Analysis, the significant step is to extract the relevant data from the database server. Several SQL queries were implemented to achieve this task. Tables were altered during the process to make sure that the data obtained was relevant for our analysis.

Research Question1: To answer the first research question about how the manufacturing sales were impacted during pandemic the approach was as follows:

- The monthly total manufacturing sales for the year 2019,2020 and 2021 was extracted from the database using the following SQL queries:

```
Select month_manufact,`Principal statistics`,category,manufacturing_value from merged_manufact where category='Manufacturing [31-33]'
and `Principal statistics`='Sales of goods manufactured (shipments)' and REF_DATE between '2019-01-01 00:00:00' and '2019-12-01 00:00:00';

Select month_manufact,`Principal statistics`,category,manufacturing_value from merged_manufact where category='Manufacturing [31-33]'
and `Principal statistics`='Sales of goods manufactured (shipments)' and REF_DATE between '2020-01-01 00:00:00' and '2020-12-01 00:00:00';

Select month_manufact,`Principal statistics`,category,manufacturing_value from merged_manufact where category='Manufacturing [31-33]'
 and `Principal statistics`='Sales of goods manufactured (shipments)' and REF_DATE between '2021-01-01 00:00:00' and '2021-08-01 00:00:00';
```

- A line graph was plotted to analyse the monthly trend of total manufacturing sales from Jan2019-Sep 2021, which showed that manufacturing sales in 2020 were comparatively less than 2019.

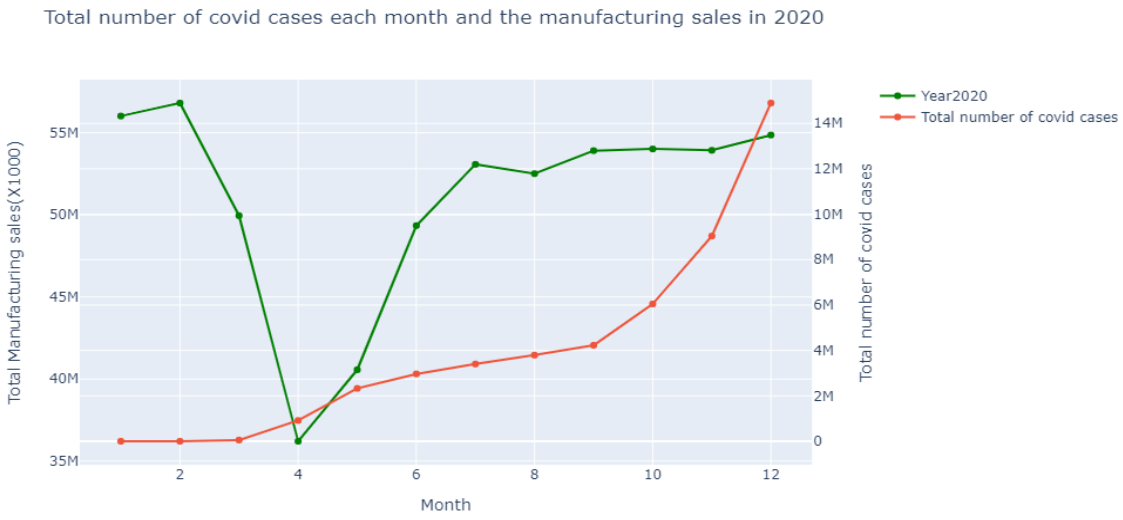


- It was also noticed that there was a sudden decrease in Manufacturing sales after Feb 2020, with the lowest value of the manufacturing sales in April 2020.The total manufacturing sales in April 2019 were around $58.35 Billions which reduce by 37% and was around $36.20 Billions in April 2020.

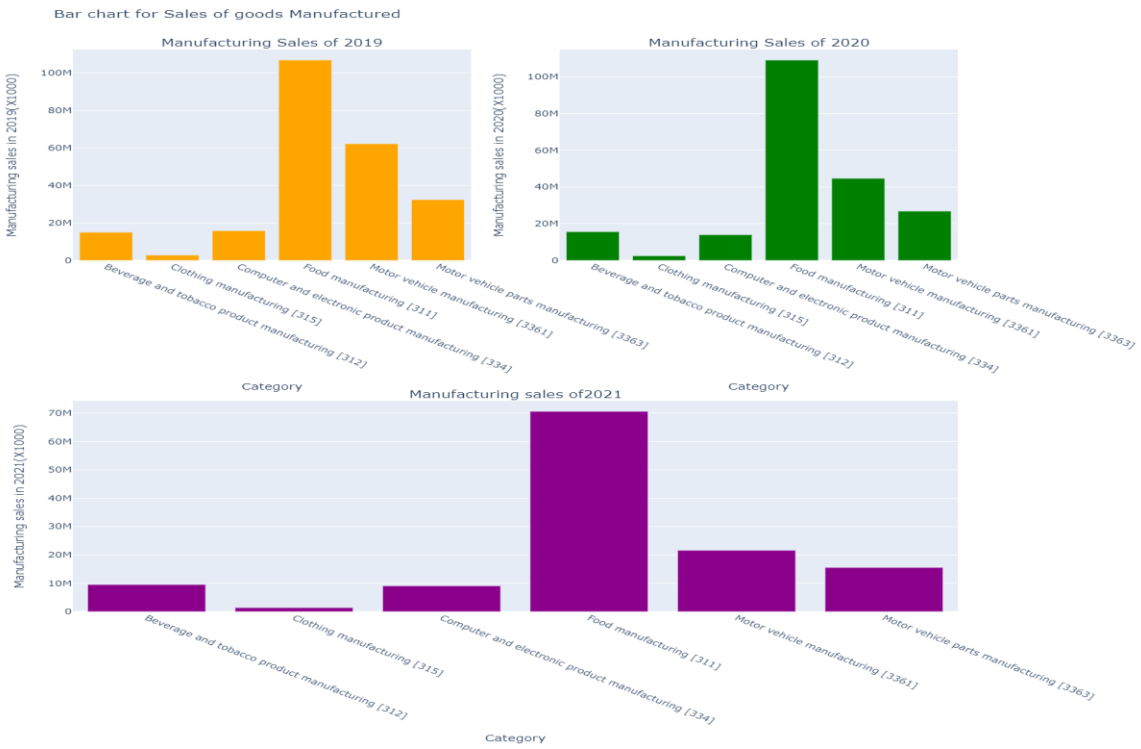
The SQL query given below was used to extract the total number of covid cases each month.

```
que_covid="Select Year(SummaryDate),month(SummaryDate),sum(TotalCases) from ca_covid
where Province!='Repatriated Cdn' group by Year(SummaryDate),month(SummaryDate);"
covid_data=pd.read_sql_query(que_covid,engine)
```

	Year(SummaryDate)	month(SummaryDate)	sum(TotalCases)
0	2020	1	17.0
1	2020	2	247.0
2	2020	3	53048.0
3	2020	4	915822.0
4	2020	5	2327776.0
5	2020	6	2965627.0
6	2020	7	3402117.0
7	2020	8	3794519.0
8	2020	9	4228453.0
9	2020	10	6048111.0
10	2020	11	9038437.0
11	2020	12	14895899.0
12	2021	1	21522163.0
13	2021	2	23142376.0
14	2021	3	28503085.0
15	2021	4	33012873.0
16	2021	5	40905166.0
17	2021	6	42084296.0
18	2021	7	44105523.0
19	2021	8	45218642.0
20	2021	9	46832780.0
21	2021	10	51954510.0
22	2021	11	29528319.0



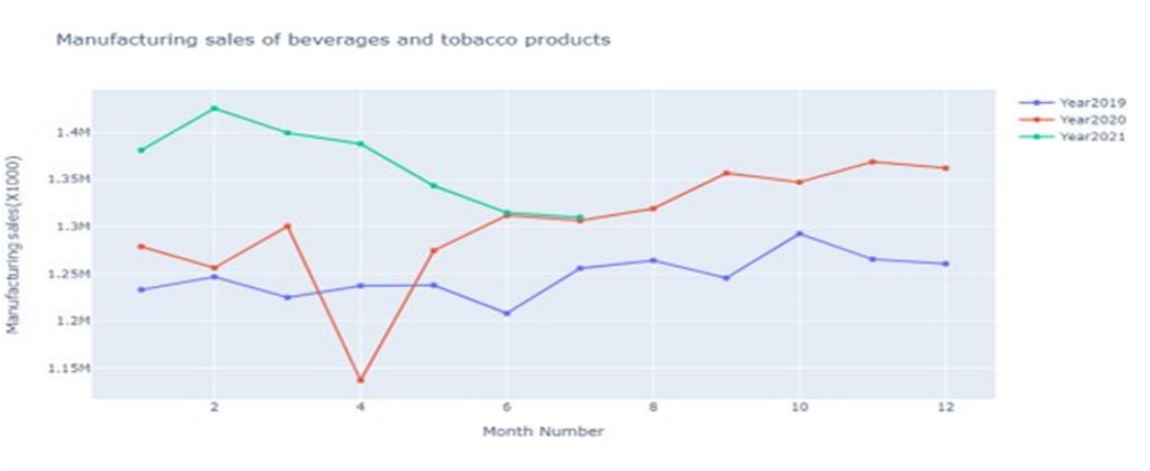
The given below bar graph shows the total manufacturing sales by different manufacturing industries for the year 2019,2020 and 2021.

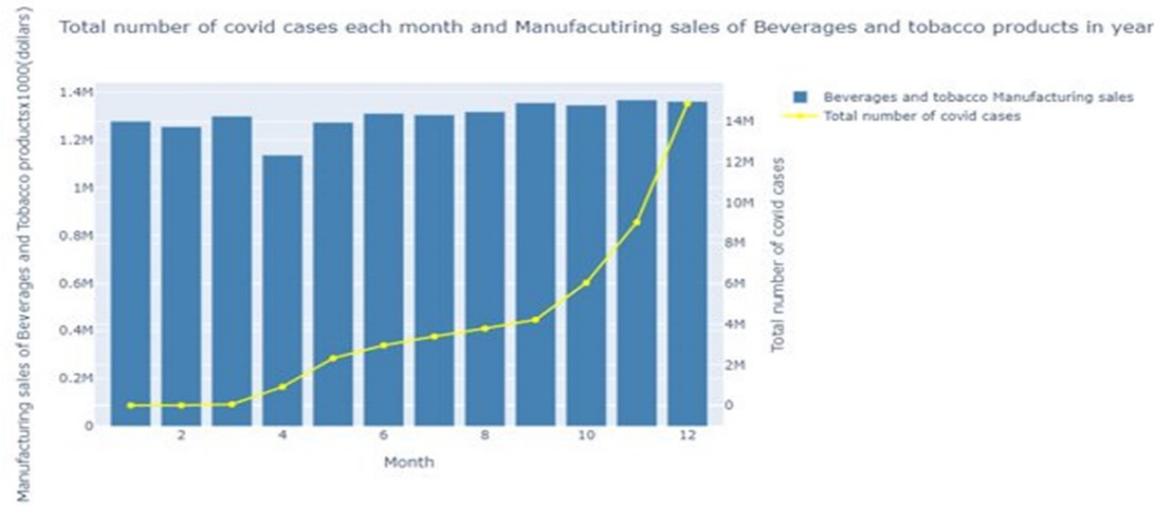


- It was noticed from bar graph that manufacturing sales of motor vehicles in 2019 was around $60 Billions, which reduced to around $42 Billion in 2020, which clearly indicates that sales in 2020 were affected due to pandemic.
- Surprisingly, the total manufacturing sales of beverages and tobacco products were around $14.97 Billions in 2019, which increased to $15.619 Billions in 2020.The food manufacturing sales in 2019 was around $106 Billion, which increased in 2020 to $109 Billion, which clearly indicates that manufacturing of food products increased in 2020,whereas the Clothing manufacturing sales were around $2.8 Billion in 2019 which reduced to $2.4 Billion in 2020.The Computer and Electronic products manufacturing sales were around $15.79 Billion in 2019, which reduced to $13.9 Billion in 2020.

For beverage and tobacco product manufacturing sales following SQL query was used to retrieve the required data.

```
#Sum of manufacturing sales in 2019 grouped by category(in thousands)
query_bt2019="Select * from merged_manufact where `Principal statistics`='Sales of goods manufactured (shipments)'and
 category='Beverage and tobacco product manufacturing [312]' and REF_DATE between '2019-01-01 00:00:00' and '2019-12-01 00:00:00';"
value19_btresult=pd.read_sql_query(query_bt2019,engine)

#Sum of manufacturing sales in 2020 grouped by category(in thousands)
query_bt2020="Select * from merged_manufact where `Principal statistics`='Sales of goods manufactured (shipments)'and
category='Beverage and tobacco product manufacturing [312]' and REF_DATE between '2020-01-01 00:00:00' and '2020-12-01 00:00:00';"
value20_btresult=pd.read_sql_query(query_bt2020,engine)


#Sum of manufacturing sales in 2021 grouped by category(in thousands)
query_bt2021="Select * from merged_manufact where `Principal statistics`='Sales of goods manufactured (shipments)'and
 category='Beverage and tobacco product manufacturing [312]' and REF_DATE between '2021-01-01 00:00:00' and '2021-08-01 00:00:00';"
value21_btresult=pd.read_sql_query(query_bt2021,engine)
```



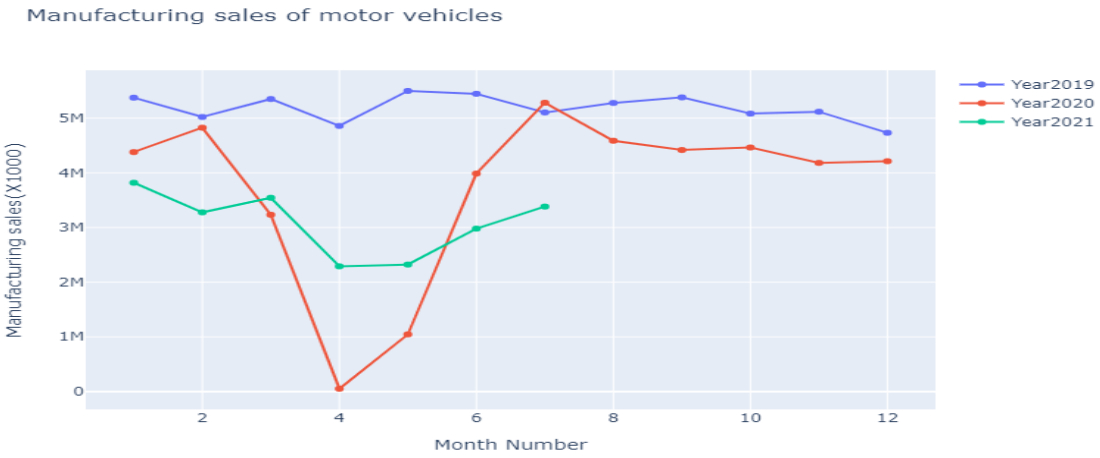


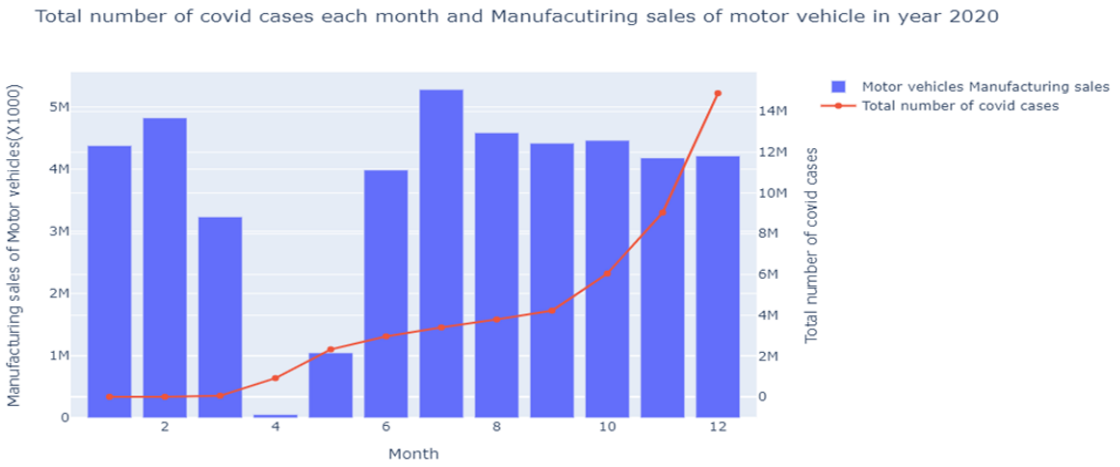
For the beverages and tobacco products, the manufacturing sales increased during pandemic. But for the year 2020, we saw a dip in the Month of April. One of the reasons was that this was the time when the first covid-wave hit Canada and there was a sudden increase in the number of covid cases, due to which lockdown was imposed in Canada.

The given below SQL query was used to extract the monthly manufacturing sales of Motor vehicles from the database.

```
#Sum of manufacturing sales in 2019 grouped by category(in thousands)
query_mv2019="Select * from merged_manufact where `Principal statistics`='Sales of goods manufactured (shipments)'and
 category='Motor vehicle manufacturing [3361]' and REF_DATE between '2019-01-01 00:00:00' and '2019-12-01 00:00:00';"
value19_mvresult=pd.read_sql_query(query_mv2019,engine)


#Sum of manufacturing sales in 2020 grouped by category(in thousands)
query_mv2020="Select * from merged_manufact where `Principal statistics`='Sales of goods manufactured (shipments)'and
 category='Motor vehicle manufacturing [3361]' and REF_DATE between '2020-01-01 00:00:00' and '2020-12-01 00:00:00';"
value20_mvresult=pd.read_sql_query(query_mv2020,engine)


#Sum of manufacturing sales in 2021 grouped by category(in thousands)
query_mv2021="Select * from merged_manufact where `Principal statistics`='Sales of goods manufactured (shipments)'and
category='Motor vehicle manufacturing [3361]' and REF_DATE between '2021-01-01 00:00:00' and '2021-08-01 00:00:00';"
value21_mvresult=pd.read_sql_query(query_mv2021,engine)
```

There was a significant difference observed in the manufacturing sales of motor vehicles in during 2019 and 2020.The manufacturing sales in April 2019 were around $2.2 Billion, which reduced to $51 Million in 2020.Overall the total motor vehicle manufacturing sales reduced by $18 Billion dollar in 2020 when compared with the total manufacturing sales in 2019.This clearly indicates that covid had a huge impact on the manufacturing sales of motor vehicles. The recovery is slow as the manufacturing sales had still not returned to the post-pandemic levels. The figure below shows monthly distribution of motor vehicle manufacturing sales from January 2019 to September 2021.



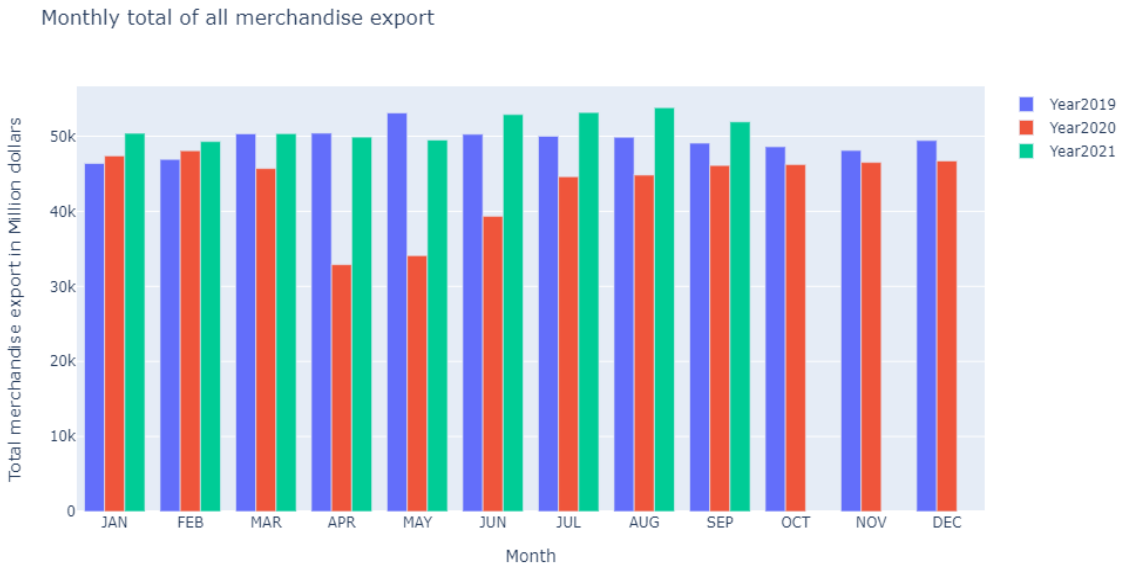


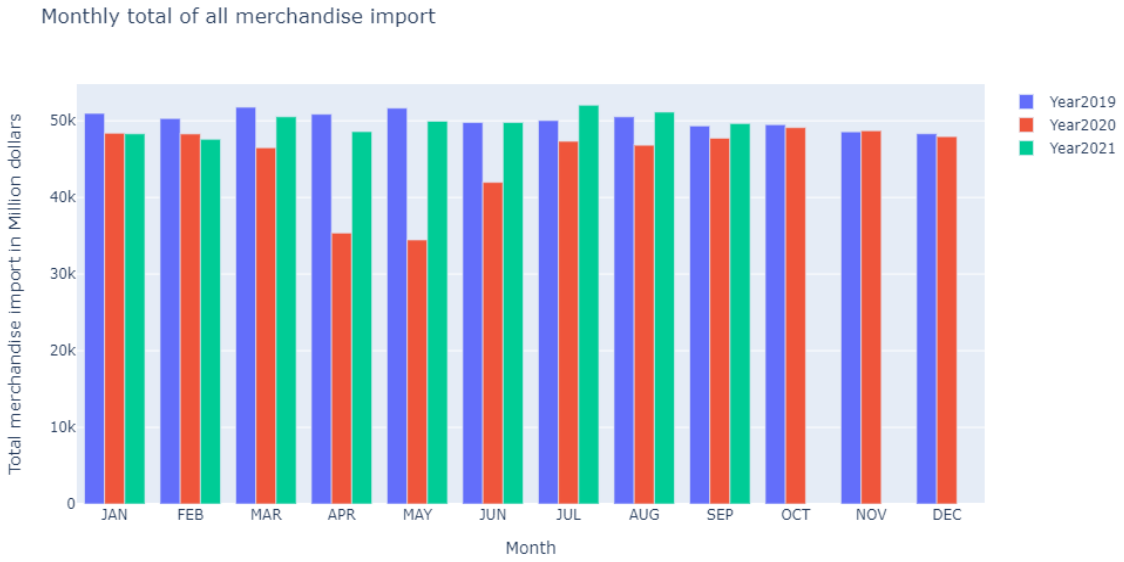
Another reason for decrease for decrease in the manufacturing sales is the shortage of semi-conductors that are used in the Car Manufacturing. This resulted in manufacturing less cars which resulted in impacting the manufacturing sales of Motor vehicles in the whole Country.

Research Question 2: To answer about how the manufacturing sector impacted the International Merchandise trade, we retrieved the total merchandise export and import data from the database using some set of SQL queries so that relevant information for analysis was used to answer the question.

```
#Monthly total import for year 2019
que_in19="Select * from international_trade where Basis='Customs' and product_category='Total of all merchandise' and Trade='Import' and YEAR(REF_DATE)='2019' group by MONTH(REF_DATE);"
df_monthin19=pd.read_sql_query(que_in19,engine)
print(df_monthin19)

#Monthly total import for year 2020
que_in20="Select * from international_trade where Basis='Customs' and product_category='Total of all merchandise' and Trade='Import' and YEAR(REF_DATE)='2020' group by MONTH(REF_DATE);"
df_monthin20=pd.read_sql_query(que_in20,engine)
print(df_monthin20)

#Monthly total import for year 2021
que_in21="Select * from international_trade where Basis='Customs' and product_category='Total of all merchandise' and Trade='Import' and YEAR(REF_DATE)='2021' group by MONTH(REF_DATE);"
df_monthin21=pd.read_sql_query(que_in21,engine)
print(df_monthin21)

#Monthly total international trade merchandise for year 2019
que_ein19="Select * from international_trade where Basis='Customs' and product_category='Total of all merchandise' and Trade='Export' and YEAR(REF_DATE)='2019' group by MONTH(REF_DATE);"
df_monthein19=pd.read_sql_query(que_ein19,engine)

#Monthly total international trade merchandise for year 2020
que_ein20="Select * from international_trade where Basis='Customs' and product_category='Total of all merchandise' and Trade='Export' and YEAR(REF_DATE)='2020' group by MONTH(REF_DATE);"
df_monthein20=pd.read_sql_query(que_ein20,engine)

#Monthly total international trade merchandise for year 2021
que_ein21="Select * from international_trade where Basis='Customs' and product_category='Total of all merchandise' and Trade='Export' and YEAR(REF_DATE)='2021' group by MONTH(REF_DATE);"
df_monthein21=pd.read_sql_query(que_ein21,engine)
```

It was noticed that total merchandise import was around $50 Billion in 2019, which reduced to $35 Billion dollar in April 2020.The distribution of the total monthly merchandise import value and export value is shown in the bar graph, where the merchandise trade value was highly impacted in 2020.





Surprisingly, the distribution of the total export of goods and services followed the similar pattern as that of the total manufacturing sales. It can be seen from the line graph that with decrease in the Manufacturing sales the total export of goods was also impacted, where April 2020 was the month with lowest manufacturing sales and the least value for the export of goods.

The analysis gave us sufficient results to suggest that the manufacturing sector was one of the major reasons which affected the total export of goods during pandemic. Several government and travel

restrictions during pandemic were also the reason which resulted in the lower export of goods. The pandemic had a huge impact on the manufacturing sector, which resulted in several domestic manufacturing companies to shut down or lower down their manufacturing sales, thus reducing the export of goods and services.

Research Question3: To analyse the impact of manufacturing sector on the employment rate, a significant step was to extract the total number of persons working in the manufacturing sector from Jan2019 to September 2021.

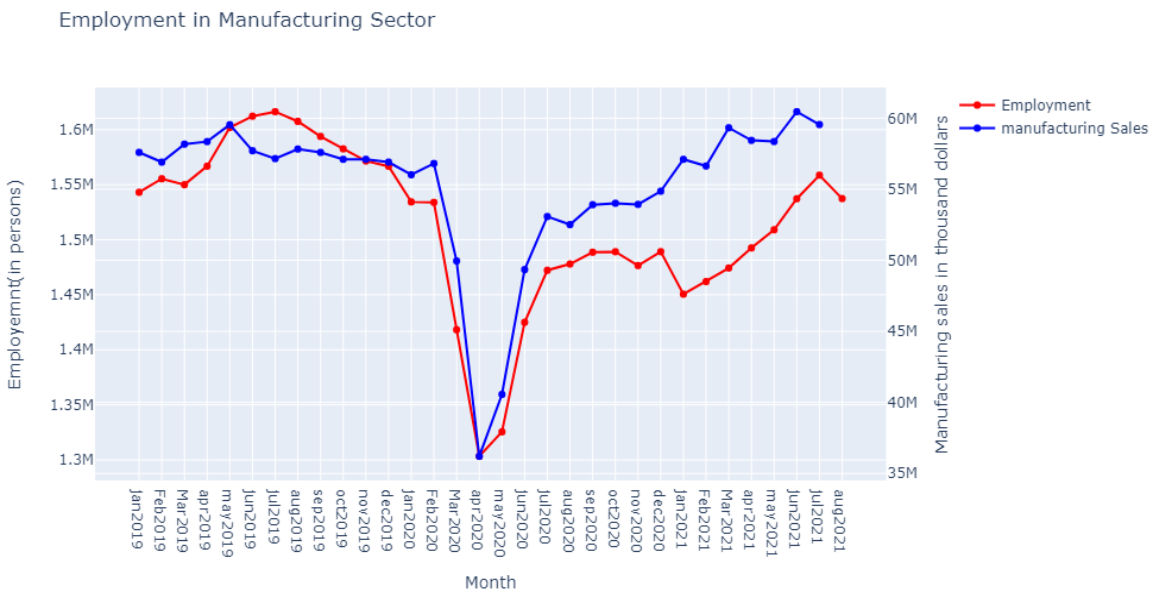
```
# Total number of people employed in manufacturing sector in 2019
que19="select YEAR(REF_DATA) AS emp_year,MONTH(REF_DATA) AS emp_month,emp_sector,emp_value from employment where emp_sector='Manufacturing [31-33]' and YEAR(REF_DATA)=2019;"
res_emp19=pd.read_sql_query(que19,engine)


# Total number of people employed in manufacturing sector in 2020
que20="select YEAR(REF_DATA) AS emp_year,MONTH(REF_DATA) AS emp_month,emp_sector,emp_value from employment where emp_sector='Manufacturing [31-33]' and YEAR(REF_DATA)=2020;"
res_emp20=pd.read_sql_query(que20,engine)


# Total number of people employed in manufacturing sector in 2021
que21="select YEAR(REF_DATA) AS emp_year,MONTH(REF_DATA) AS emp_month,emp_sector,emp_value from employment where emp_sector='Manufacturing [31-33]' and YEAR(REF_DATA)=2021;"
res_emp21=pd.read_sql_query(que21,engine)
```

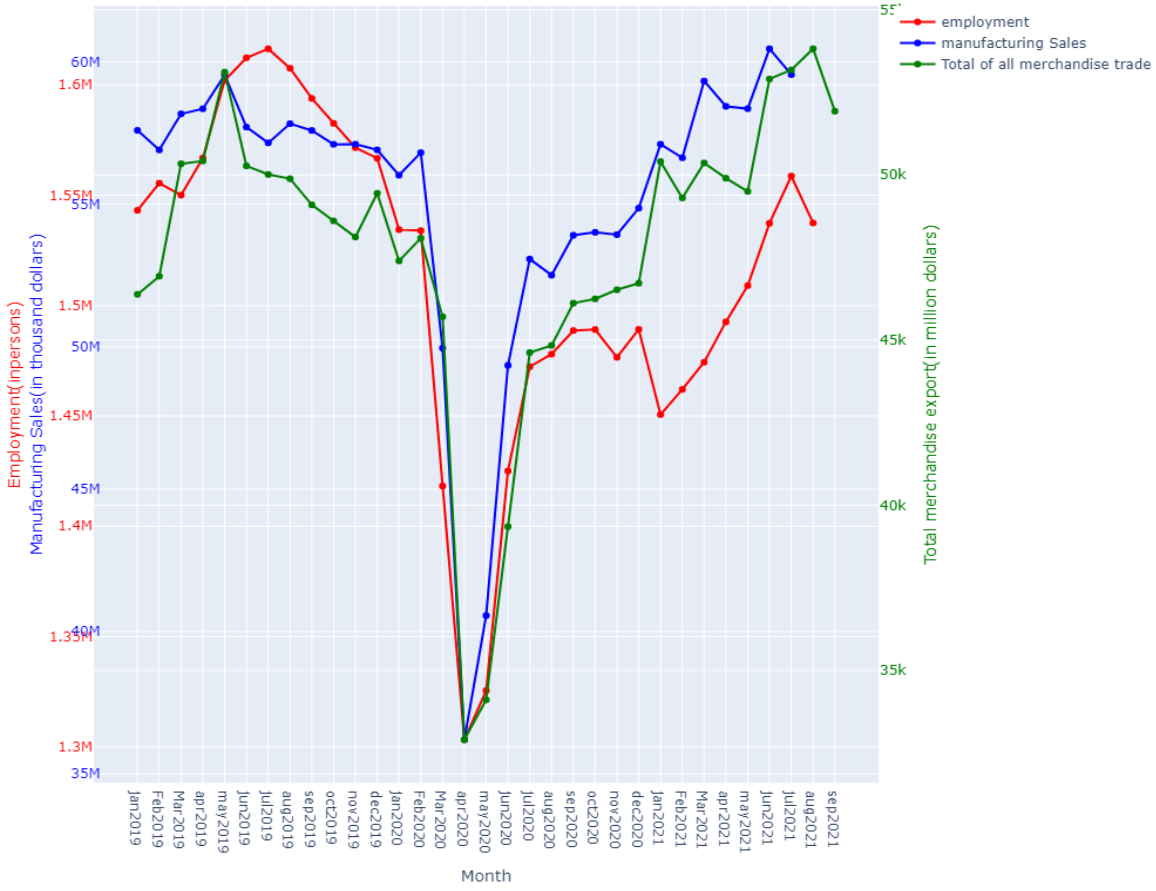
	emp_year	emp_month	emp_sector	emp_value
0	2019	1	Manufacturing [31-33]	1543149
1	2019	2	Manufacturing [31-33]	1555451
2	2019	3	Manufacturing [31-33]	1550054
3	2019	4	Manufacturing [31-33]	1566845
4	2019	5	Manufacturing [31-33]	1601992
5	2019	6	Manufacturing [31-33]	1612296
6	2019	7	Manufacturing [31-33]	1616370
7	2019	8	Manufacturing [31-33]	1607540
8	2019	9	Manufacturing [31-33]	1593877
9	2019	10	Manufacturing [31-33]	1582570
10	2019	11	Manufacturing [31-33]	1571554
11	2019	12	Manufacturing [31-33]	1566776

	emp_year	emp_month	emp_sector	emp_value
0	2020	1	Manufacturing [31-33]	1534424
1	2020	2	Manufacturing [31-33]	1533965
2	2020	3	Manufacturing [31-33]	1418230
3	2020	4	Manufacturing [31-33]	1303294
4	2020	5	Manufacturing [31-33]	1325521
5	2020	6	Manufacturing [31-33]	1425035
6	2020	7	Manufacturing [31-33]	1472287
7	2020	8	Manufacturing [31-33]	1477984
8	2020	9	Manufacturing [31-33]	1488679
9	2020	10	Manufacturing [31-33]	1489139
10	2020	11	Manufacturing [31-33]	1476568
11	2020	12	Manufacturing [31-33]	1489200

	emp_year	emp_month	emp_sector	emp_value
0	2021	1	Manufacturing [31-33]	1450606
1	2021	2	Manufacturing [31-33]	1462019
2	2021	3	Manufacturing [31-33]	1474345
3	2021	4	Manufacturing [31-33]	1492613
4	2021	5	Manufacturing [31-33]	1509060
5	2021	6	Manufacturing [31-33]	1537293
6	2021	7	Manufacturing [31-33]	1558685
7	2021	8	Manufacturing [31-33]	1537438

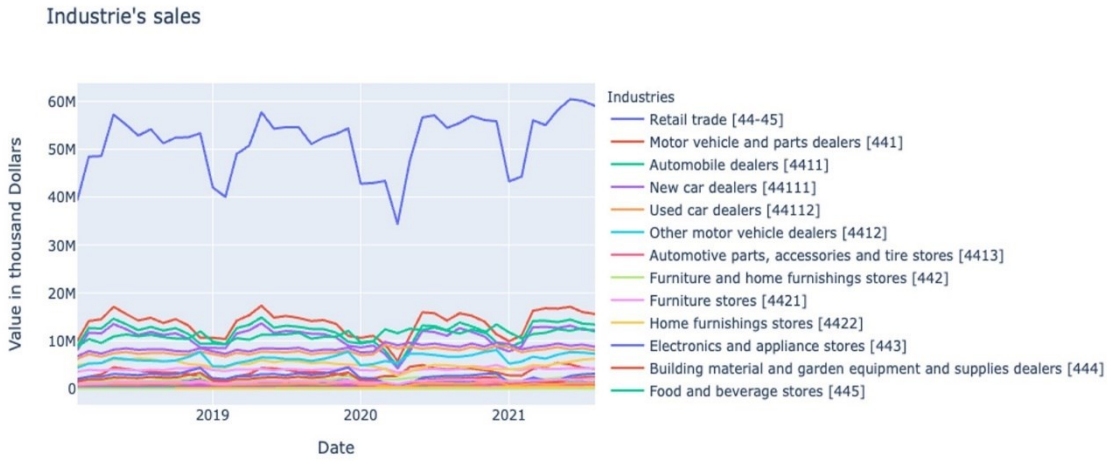
The line graph was plotted which showed some surprising results. The total number of people employed in the manufacturing sector was around 1.5 million in April 2019 which reduced to 1.3 million people in April 2020, which indicates that jobs of around 200 thousand people were impacted during the pandemic. The distribution of the total number of people employed is shown the graph, which clearly indicates that the employment rate was highly impacted during covid-19.



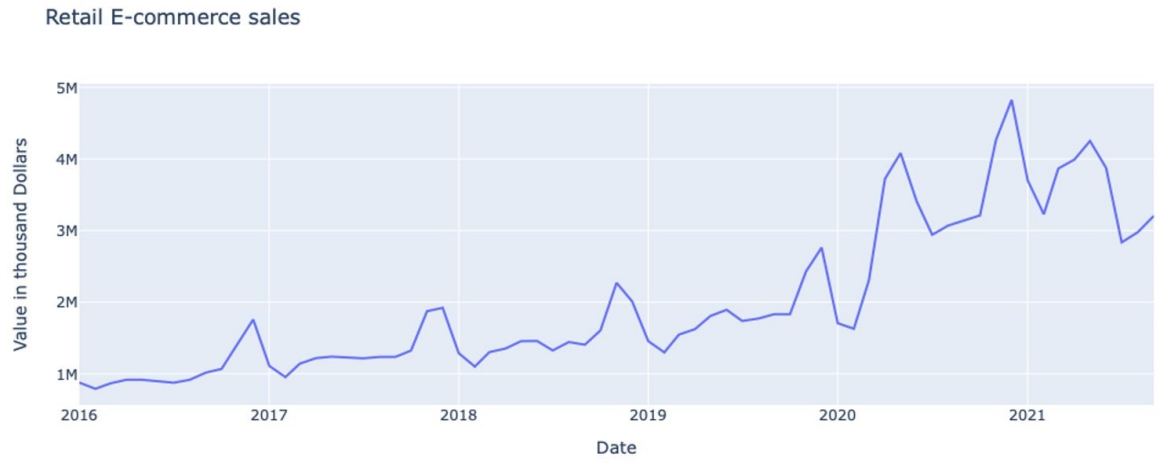


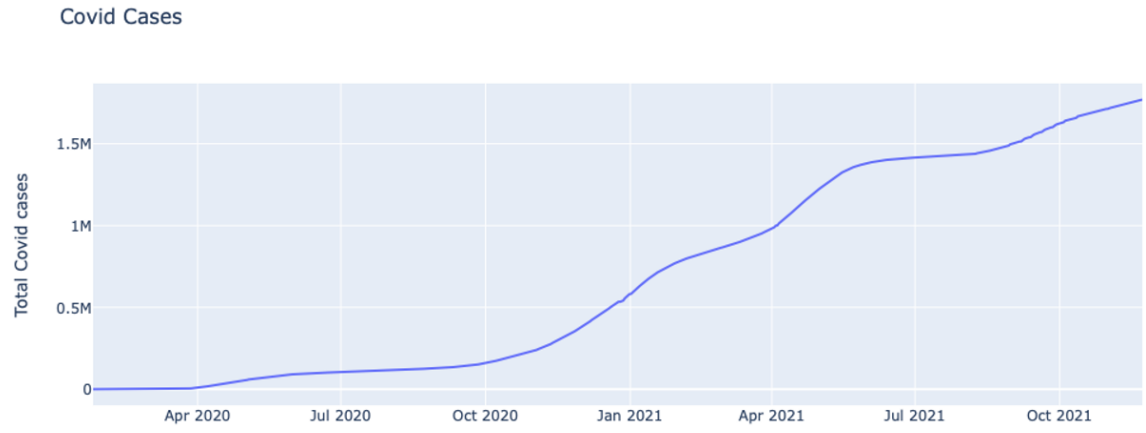
The distribution of the monthly manufacturing sales and the employment rate follows a similar pattern, which clearly indicates that manufacturing sector has a huge impact on the employment. Manufacturing provides a lot of services for the employment sector, therefore impact on the Manufacturing sales not only affects the Manufacturing sector, but it also has a huge impact on the employment and trade sector.

Retail - Data Exploration and Discussion (Yves Nsoga)



From the chart above, we realised that covid impacted industries relying on imported products. Because for those industries, in 2018 and 2019, we had the same seasonality, but in 2020 those industries recorded the lowest sales during. But industries relying on local productions, like industries providing basic needs products were more resilient.





For the retail industry, while this industry was recording low sales values during the pandemic, the e-commerce sales were exponentially increasing. And even when all the industries tend to come back to the normal trend, the revenue of sales made online are still increasing. The covid pandemic accelerate the integration of online shopping habit, so that even whit the restriction going down, people are more opened to purchase a product online than going to a store. Therefore, companies like amazon are making good profits during this period.

From these data we learn how it is important to have data with same granularity to perform a good analysis.

Conclusion

In conclusion, we found that most domain we explored were affected by COVID-19. The sectors related to transportation and movement of people were most negatively affected (aka. Tourism, Air Travel, Ground Travel, Motor Vehicle Manufacturing). Some were unaffected (Electricity Production) and some affected positively (e-commerce). Breaking them down in detail for each domain:

1. Tourism
    - there was significant negative impact from COVID on tourism expenditure. The international tourism expenditure is most affected by the pandemic as there has been very insignificant rise in spending as of 2021, Q2. This can also be attributed to safety measures placed by the government to manage the spread of covid 19 and its variants. Most noticeably in the transportation sector, and especially air travel
    - Pre-trip expenditure increased, the additional expense of PCR tests and mandatory quarantines may be a reason for the rise among other possibilities.
    - For the employment in tourism, there was also a significant negative impact from COVID, with the food and beverage sector having the highest drop in employment. This also shows that the employment is not only affected by the expenditure as the transportation sector was most affected in expenditure, but employment was not affected as much as the food and beverage services sector
2. Air Travel
    - There was a significant negative effect (more than 90%) on number of passengers, airline operating revenues and number of flights across Canada.
    - In 2021, we saw a rebound in the number of passengers flying in Canada. But they are nowhere near pre-pandemic levels.
3. Ground Travel
    - There was significant impact from COVID on international ground travel between Canada and United States. The number of vehicles traveling into Canada dropped by 64% in 2020 as compared to 2019.
    - Compared to 2019, 2020 saw a larger proportion of vehicles entering Canada through prairie / central Provinces (AB, SK, MB, ON) and a lesser proportion of vehicles entering Canada through the coastal / outer provinces (YT, BC, QC, NB).

- Vaccination has negligible impact on ground travel recovery thus far.
- International ground travel at the US-Mexico border recovered much quick than at the US-Canada border.

4. Electricity
    - Although there was a slight decrease, we can not conclusively say that Covid 19 had a significant impact on the monthly generation amount.
    - The impact was more visible in Alberta and Ontario. There was no notable impact from the pandemic on the source used to produce energy.
5. Manufacturing
    - The impact of the pandemic on the manufacturing sales varies according to different manufacturing industry.
    - The manufacturing sales of Food, beverages and tobacco products increased during pandemic, whereas the manufacturing sales of Clothing, Computers, Electronic products and Motor vehicles decreased significantly with highest impact being on the Motor vehicles Manufacturing sales.
    - Manufacturing had a huge impact on the International Merchandise trade where the export of good and services decreased by $18 billion dollar in April 2020 as compared to April 2019.The distribution of the monthly total Manufacturing sales and International Merchandise export showed a similar pattern from Jan 2019 to September 2021, which indicates that effect of covid on the Manufacturing sector also impacted the International Merchandise trade.
    - Employment in the Manufacturing sector was highly impacted during pandemic, where 1.5 million people were employed during April 2019 which reduced to 1.3 million people in April 2020. The recovery rate in the Employment sector is pretty good and has almost reached the post-pandemic level.
6. Retail
    - The pandemic accelerated the adoption of online shopping by the customers. The e-commerce industry is booming with the covid and even after the covid cases are going down. During 2020, E-commerce sales recorded an increase of 330% in value. The consumers developed the habit of shopping online and seem to be sticking to that habit.
    - The industries negatively impacted by the covid are industries relying on imported product, whereas industries providing local productions were more resilient.
    - Companies like amazon registered the highest revenue. The booming e-commerce industry led to the rise of companies such as Skip the Dishes, DoorDash, etc.

References

- Esri Canada. 2021a. *Provincial Daily Totals*. [online] Available at: <https://resources-covid19canada.hub.arcgis.com/datasets/covid19canada::provincial-daily-totals/about> [Accessed 4 November 2021].
- Esri Canada. 2021b. *FAQ: Responding to the COVID-19 Crisis with Free ArcGIS Hub Resources*. [online] Available at: <https://resources.esri.ca/news-and-updates/faq-responding-to-the-covid-19-crisis-with-free-arcgis-hub-resources> [Accessed 4 November 2021].
- The Atlantic Monthly Group, 2021a. *The Data.* [online] The COVID Tracking Project. Available at: <*https://covidtracking.com/data*> [Accessed 11 December 2021].
- The Atlantic Monthly Group, 2021b. *Data License.* [online] The COVID Tracking Project. Available at: <*https://covidtracking.com/about-data/license*> [Accessed 11 December 2021].
- Statistics Canada. 2021a, *Tourism demand in Canada, constant prices (x1,000,000)/* Available at: **<**https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610023001> [Accessed 4 November 2021].
- Statistics Canada. 2021c, *Employment generated by tourism (x 1,000)* Available at: <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610023201> [Accessed 4 November 2021].
- Statistics Canada. 2021d. *Electric power generation, monthly generation by type of electricity*. [online] Available at: < https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2510001501> [Accessed 4 November 2021].
- Statistics Canada. 2021e. *Statistics Canada Table 16-10-0119-01 Historical (real-time) releases of manufacturing sales, by North American Industry Classification System* [online] Available at: <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1610011901> [Accessed 4 November 2021].
- Statistics Canada. 2021f. *Employment by industry, monthly, unadjusted for seasonality*. [online] Available at: <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410020101> [Accessed 4 November 2021].
- Statistics Canada. 2021g. *Retail trade sales of motor vehicle and parts dealers, by industry*. [online] Available at: <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2010000803> [Accessed 4 November 2021].
- Statistics Canada. 2021h. *International merchandise trade by commodity, monthly*. [online] Available at: <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1210012101> [Accessed 4 November 2021].
- Statistics Canada 2021i. *Retail trade sales by province and territory*. [online/ Available at: <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2010000801> [Accessed 4 November 2021].
- Statistics Canada 2021j. *Retail trade, total sales, and e-commerce sales*. [online/ Available at: <https://open.canada.ca/data/en/dataset/89077c3a-cfe1-4b6c-a35f-015cb8688f47> [Accessed 4 November 2021].
- Statistics Canada, 2021k, *Aircraft movements, by civil and military movements, airports with NAV CANADA towers, monthly*. Available at https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2310000301#tables [Accessed 4 November 2021]
- Statistics Canada. 2021l. *Number of vehicles travelling between Canada and the United States*. [online] Available at: <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2410000201> [Accessed 4 November 2021].
- U.S. Department of Homeland Security. 2021. *Border Crossing Entry Data*. [online] Available at: <https://data.bts.gov/Research-and-Statistics/Border-Crossing-Entry-Data/keg4-3bc2> [Accessed 4 November 2021].
- Natural Resources Canada. 2021a. *Electricity facts*. [online] Available at: < https://www.nrcan.gc.ca/science-and-data/data-and-analysis/energy-data-and-analysis/energy-facts/electricity-facts/20068#L1> [Accessed 10 November 2021].
- Natural Resources Canada. 2021b. *Energy and Greenhouse Gas Emissions (GHGs)*. [online] Available at: < https://www.nrcan.gc.ca/science-and-data/data-and-analysis/energy-data-and-analysis/energy-facts/energy-and-greenhouse-gas-emissions-ghgs/20063> [Accessed 10 November 2021].