ECE 356 Final Project Report
Group 23

Matthew Tang (20723384)
Naziba Haider (20661300)
Selina Ali (20722063)

Table of Contents

COVID-19 Database Project

Project Description

Project repository: https://github.com/naz321/356-Project

Our project supports a wide range of users that are looking into the effects of COVID-19. Our project is primarily focused on how COVID-19 is affecting Canadians and it's toll on hospitals. Although there are a wide variety of use cases of COVID-19 data such as how COVID-19 is affecting businesses and the economy, our group wanted to focus more on COVID-19 transmission, infection rate, mortality, testing, and recovery.

Our primary objective is to focus on the number of confirmed COVID-19 cases from September 2020 to October 2020. The data that we are using is from StatsCan and COVID-19 Canada. Users will be able to query the following:

- By region (Able to select individual provinces and territories or multiple at a time)
- By gender
- By age groups
- By occupation
- By calendar week

An example use case would be: Finding the number of **women** who were infected with COVID-19 from **Domestic Acquisition**, located in **British Columbia,** who are between the **ages 20-29**, and **recovered** from COVID-19 during the **Week of October 4th**.

Our secondary objective is to look at transmissions, symptoms, and hospital admission. Users will be able to find and edit the following:

- For confirmed cases, were they hospitalized and if so were they admitted to the ICU?
- For confirmed cases, when did their symptoms occur?
- For confirmed cases, were they asymptomatic?
- For confirmed cases, how did they get COVID-19? Was it through domestic acquisition or International Travel?

An example use case would be: Finding the number of **asymptomatic** patients that were infected with COVID-19 from **International Travel** and were **admitted into the hospital ICU**.

Our third use case would be looking at COVID-19 testing. Users will be able to find and edit the following:

- Find the weekly number of tests done by each region in Canada
- Find the total number of tests performed by each region given a certain week

An example use case would be: Finding the number of **COVID-19 tests** done in **Ontario** during the **Week of October 4th**.

Our fourth and final use case would be to use data mining to determine if we are able to predict the survival rate of future COVID-19 patients using inputs from our data sets. Predicting the survival rate of future COVID-19 patients would be extremely beneficial to hospitals as they could prioritize which patients to treat first.

Project Scope

For this project, our team used 2 online sources from Statistics Canada and COVID-19 Canada. The specific websites where data was retrieved are as follows:

- https://www150.statcan.gc.ca/n1/pub/13-26-0003/132600032020001-eng.htm
- https://resources-covid19canada.hub.arcgis.com/datasets/provincial-daily-totals

Using these datasets, we created a relational schema to organize our data. We created our database using information we found appropriate and relevant for our targeted users. Following this, our team put together a client application for users with various access rights to view and modify the datasets. Our team then used data-mining to determine the survival and mortality rate of those that were COVID-19 positive depending on various attributes. And finally, putting together test cases to ensure all aspects of our project were working as expected.

Client Application

Project Work Plan

Based on the outlined timelines of the project, throughout the semester our team focused on 5 main areas within this project, SQL code for relational schemas, the client application, data-mining and deliverables (report and demo video). All group members collaborated across all sections with one person being held responsible for the first three of these areas (SQL code, client application and data mining). Throughout the semester our team met weekly via Discord to discuss key action steps for the week (based on the individual roles and the project timeline) as well as any blockers we were facing. We also discussed questions and concerns we had and contacted our assigned TA (Raavi Soni) if further clarification was needed.

Work Plan and Responsibilities

Each group member was responsible for the following sections:
Selina Ali

- SQL code for the relational schema
    - Responsible for ensuring that the code for creating tables, views, constraints, etc. is written
    - Responsible for all code necessary to load data into the database (including the code for dealing with errors and the testing code for this schema)

Naziba Haider

- Client application
    - Responsible for ensuring that the code necessary for the client to be compiled and executed is functionally correct
    - Responsible for ensuring that the test case code for the application is properly written

Matthew Tang

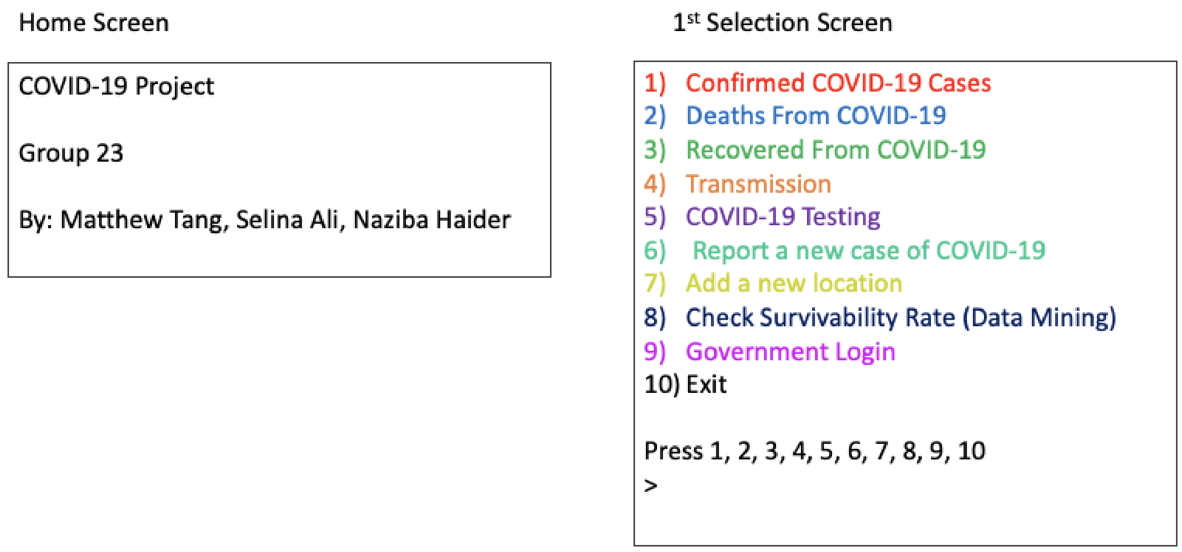
- Data-mining code
    - Responsible for ensuring that the code necessary for creating the desired data-mining technique is properly written (including all code needed to apply the data-mining technique to build a data model)
    - Responsible for ensuring that the code for validating the data model, as well as the test case code for testing the functionality of the data-mining system is correct

Assumptions

For this project, our team created assumptions surrounding the project that would allow us to create more specific data sets. Our first assumption was that a client may be located in any province in Canada. Our second assumption was that a person can be tested for COVID-19 at any time and this test could be either positive or negative. Our third assumption states that a person can recover from COVID-19 at any time and cases in which the person does not recover is considered a death. Our fourth and final assumption for this project is that the data used will be taken from records from September 2020 - October 2020 because of the limited range of data from StatsCan.

Client Experience - Viewing Data

At a high level, a person using our client application will be able to search for Canadian COVID-19 data based on various filters. Our client application is split into 2 main sections, users are able to both view data and modify existing data. A user will be able to choose from a variety of areas including confirmed COVID-19 cases, deaths, transmissions and weekly testing as well the last 3 (6,7, and 9) options allow users to modify the data, there are 2 types of users that can do so, regular users which are able to report a new case and add a location, and government users that can modify and delete existing data.


Home Screen

```
COVID-19 Project

Group 23

By: Matthew Tang, Selina Ali, Naziba Haider
```

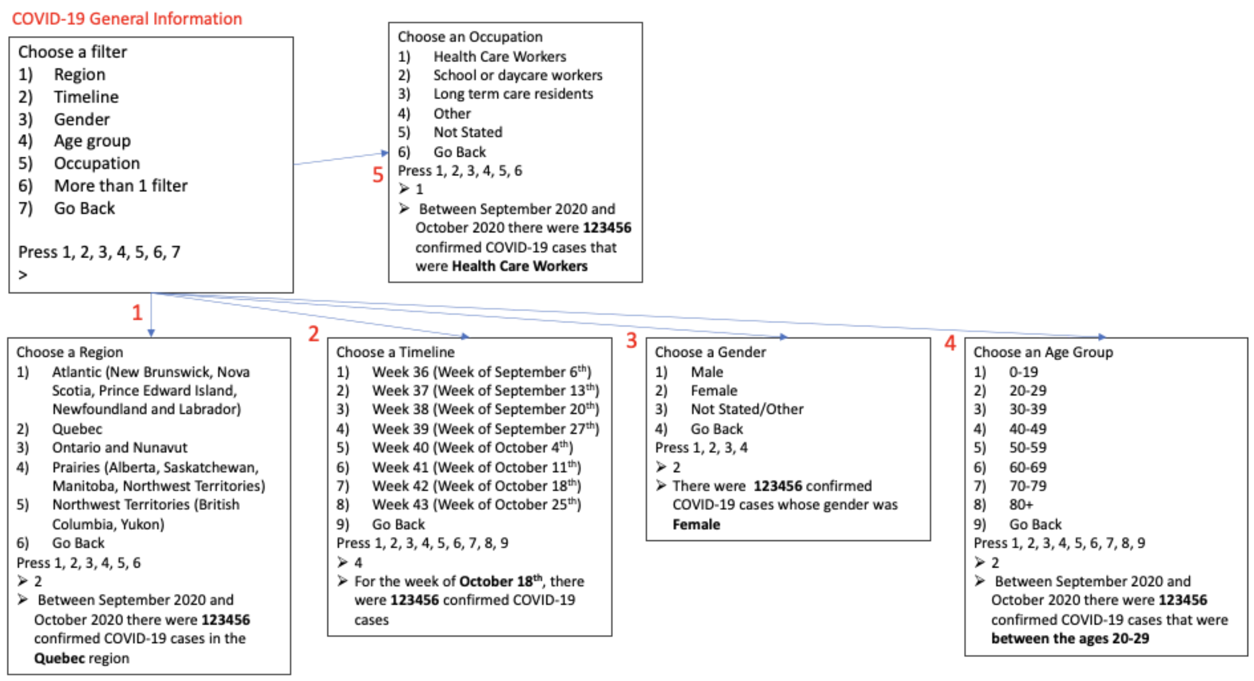
1st Selection Screen

```
1)  Confirmed COVID-19 Cases
2)  Deaths From COVID-19
3)  Recovered From COVID-19
4)  Transmission
5)  COVID-19 Testing
6)   Report a new case of COVID-19
7)  Add a new location
8)  Check Survivability Rate (Data Mining)
9)  Government Login
10) Exit

Press 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
>
```

Figure 1: Home Screen and First Selection Screen of Client Application

Based on the selection that the client picks, the application will provide the relevant information to query on each specific area.

The first selection details the confirmed COVID-19 cases and general background information about those that had gotten COVID-19. A user is able to filter all cases by region, timeline, gender, age group, occupation, and a combination of these filters. They can do so by selecting the corresponding numbers for the specific area they would like to query and the specific filter.



```
Choose a filter
1)  Region
2)  Timeline
3)  Gender
4)  Age group
5)  Occupation
6)  More than 1 filter
7)  Go Back

Press 1, 2, 3, 4, 5, 6, 7
>
```

```
Choose an Occupation
1)  Health Care Workers
2)  School or daycare workers
3)  Long term care residents
4)  Other
5)  Not Stated
6)  Go Back
Press 1, 2, 3, 4, 5, 6
➢ 1
➢ Between September 2020 and
  October 2020 there were 123456
  confirmed COVID-19 cases that
  were Health Care Workers
```

```
Choose a Region
1)  Atlantic (New Brunswick, Nova
    Scotia, Prince Edward Island,
    Newfoundland and Labrador)
2)  Quebec
3)  Ontario and Nunavut
4)  Prairies (Alberta, Saskatchewan,
    Manitoba, Northwest Territories)
5)  Northwest Territories (British
    Columbia, Yukon)
6)  Go Back
Press 1, 2, 3, 4, 5, 6
➢ 2
➢ Between September 2020 and
  October 2020 there were 123456
  confirmed COVID-19 cases in the
  Quebec region
```

```
Choose a Timeline
1)  Week 36 (Week of September 6th)
2)  Week 37 (Week of September 13th)
3)  Week 38 (Week of September 20th)
4)  Week 39 (Week of September 27th)
5)  Week 40 (Week of October 4th)
6)  Week 41 (Week of October 11th)
7)  Week 42 (Week of October 18th)
8)  Week 43 (Week of October 25th)
9)  Go Back
Press 1, 2, 3, 4, 5, 6, 7, 8, 9
➢ 4
➢ For the week of October 18th, there
  were 123456 confirmed COVID-19
  cases
```

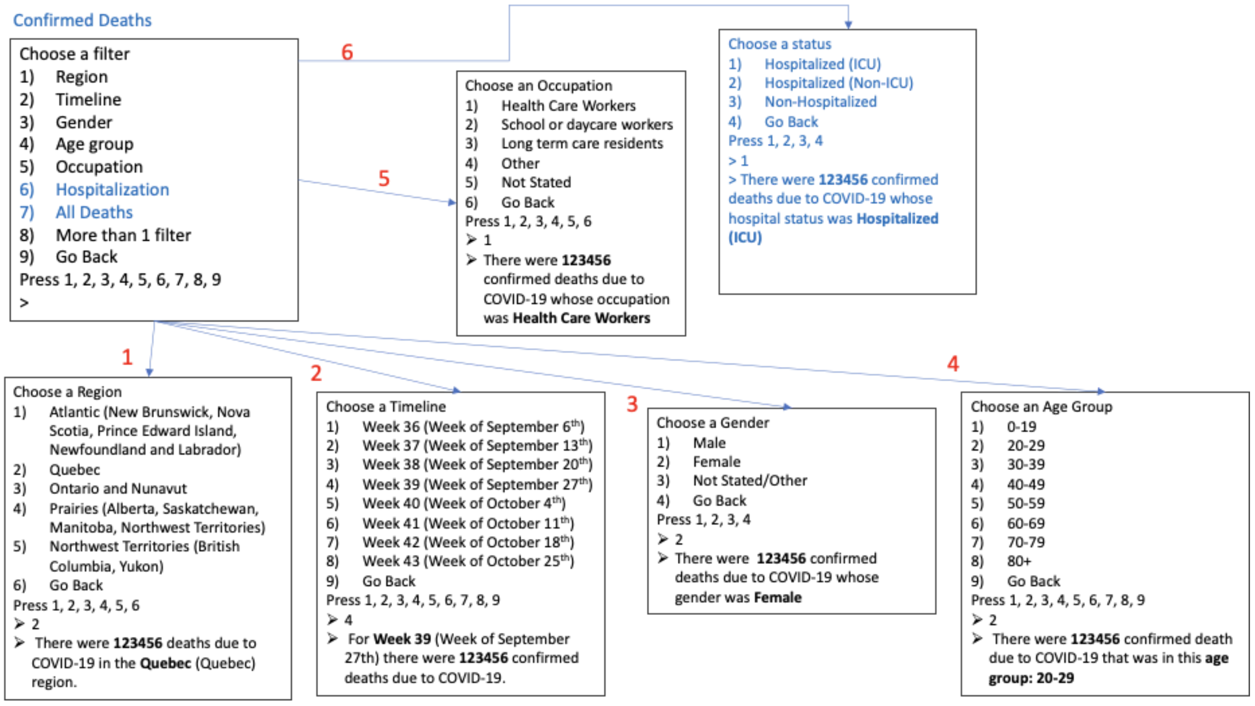
```
Choose a Gender
1)  Male
2)  Female
3)  Not Stated/Other
4)  Go Back
Press 1, 2, 3, 4
➢ 2
➢ There were 123456 confirmed
  COVID-19 cases whose gender was
  Female
```

```
Choose an Age Group
1)  0-19
2)  20-29
3)  30-39
4)  40-49
5)  50-59
6)  60-69
7)  70-79
8)  80+
9)  Go Back
Press 1, 2, 3, 4, 5, 6, 7, 8, 9
➢ 2
➢ Between September 2020 and
  October 2020 there were 123456
  confirmed COVID-19 cases that were
  between the ages 20-29
```

Figure 2: COVID-19 General Information Flow

The specific entries for each can be seen in the view above.

The next view available on the client application are the Deaths and Recovered flows. These views have similar information to that of the general information with the addition of a hospitalization field which will allow the client to filter by people who died and recovered from COVID-19 and whether or not they were hospitalized. This can be useful in researching the likeness of patients admitted and further information about them.



Choose a filter
1) Region
2) Timeline
3) Gender
4) Age group
5) Occupation
6) Hospitalization
7) All Deaths
8) More than 1 filter
9) Go Back
Press 1, 2, 3, 4, 5, 6, 7, 8, 9
>

Choose an Occupation
1) Health Care Workers
2) School or daycare workers
3) Long term care residents
4) Other
5) Not Stated
6) Go Back
Press 1, 2, 3, 4, 5, 6
➢ 1
➢ There were **123456** confirmed deaths due to COVID-19 whose occupation was **Health Care Workers**

Choose a status
1) Hospitalized (ICU)
2) Hospitalized (Non-ICU)
3) Non-Hospitalized
4) Go Back
Press 1, 2, 3, 4
> 1
> There were **123456** confirmed deaths due to COVID-19 whose hospital status was **Hospitalized (ICU)**

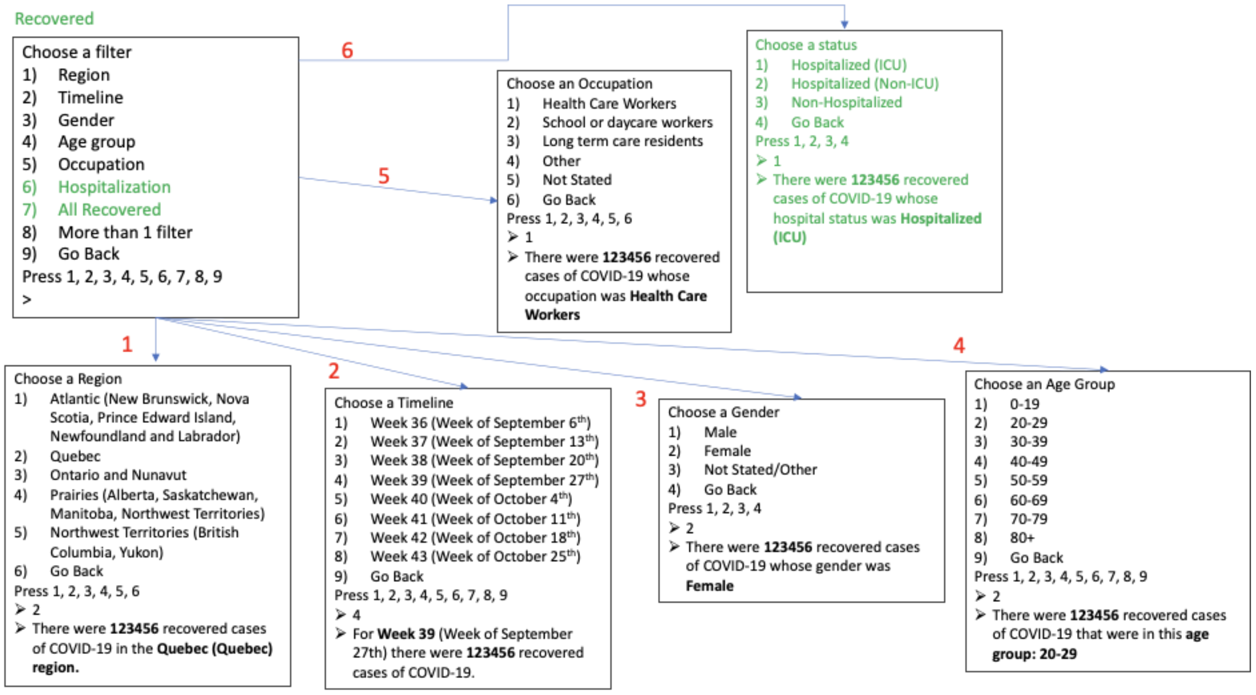
Choose a Region
1) Atlantic (New Brunswick, Nova Scotia, Prince Edward Island, Newfoundland and Labrador)
2) Quebec
3) Ontario and Nunavut
4) Prairies (Alberta, Saskatchewan, Manitoba, Northwest Territories)
5) Northwest Territories (British Columbia, Yukon)
6) Go Back
Press 1, 2, 3, 4, 5, 6
➢ 2
➢ There were **123456** deaths due to COVID-19 in the **Quebec** (Quebec) region.

Choose a Timeline
1) Week 36 (Week of September 6th)
2) Week 37 (Week of September 13th)
3) Week 38 (Week of September 20th)
4) Week 39 (Week of September 27th)
5) Week 40 (Week of October 4th)
6) Week 41 (Week of October 11th)
7) Week 42 (Week of October 18th)
8) Week 43 (Week of October 25th)
9) Go Back
Press 1, 2, 3, 4, 5, 6, 7, 8, 9
➢ 4
➢ For **Week 39** (Week of September 27th) there were **123456** confirmed deaths due to COVID-19.

Choose a Gender
1) Male
2) Female
3) Not Stated/Other
4) Go Back
Press 1, 2, 3, 4
➢ 2
➢ There were **123456** confirmed deaths due to COVID-19 whose gender was **Female**

Choose an Age Group
1) 0-19
2) 20-29
3) 30-39
4) 40-49
5) 50-59
6) 60-69
7) 70-79
8) 80+
9) Go Back
Press 1, 2, 3, 4, 5, 6, 7, 8, 9
➢ 2
➢ There were **123456** confirmed death due to COVID-19 that was in this **age group: 20-29**

Figure 3: Confirmed Deaths Flow



Choose a filter
1) Region
2) Timeline
3) Gender
4) Age group
5) Occupation
6) Hospitalization
7) All Recovered
8) More than 1 filter
9) Go Back
Press 1, 2, 3, 4, 5, 6, 7, 8, 9
>

Choose an Occupation
1) Health Care Workers
2) School or daycare workers
3) Long term care residents
4) Other
5) Not Stated
6) Go Back
Press 1, 2, 3, 4, 5, 6
➢ 1
➢ There were **123456** recovered cases of COVID-19 whose occupation was **Health Care Workers**

Choose a status
1) Hospitalized (ICU)
2) Hospitalized (Non-ICU)
3) Non-Hospitalized
4) Go Back
Press 1, 2, 3, 4
➢ 1
➢ There were **123456** recovered cases of COVID-19 whose hospital status was **Hospitalized (ICU)**

Choose a Region
1) Atlantic (New Brunswick, Nova Scotia, Prince Edward Island, Newfoundland and Labrador)
2) Quebec
3) Ontario and Nunavut
4) Prairies (Alberta, Saskatchewan, Manitoba, Northwest Territories)
5) Northwest Territories (British Columbia, Yukon)
6) Go Back
Press 1, 2, 3, 4, 5, 6
➢ 2
➢ There were **123456** recovered cases of COVID-19 in the **Quebec (Quebec) region.**

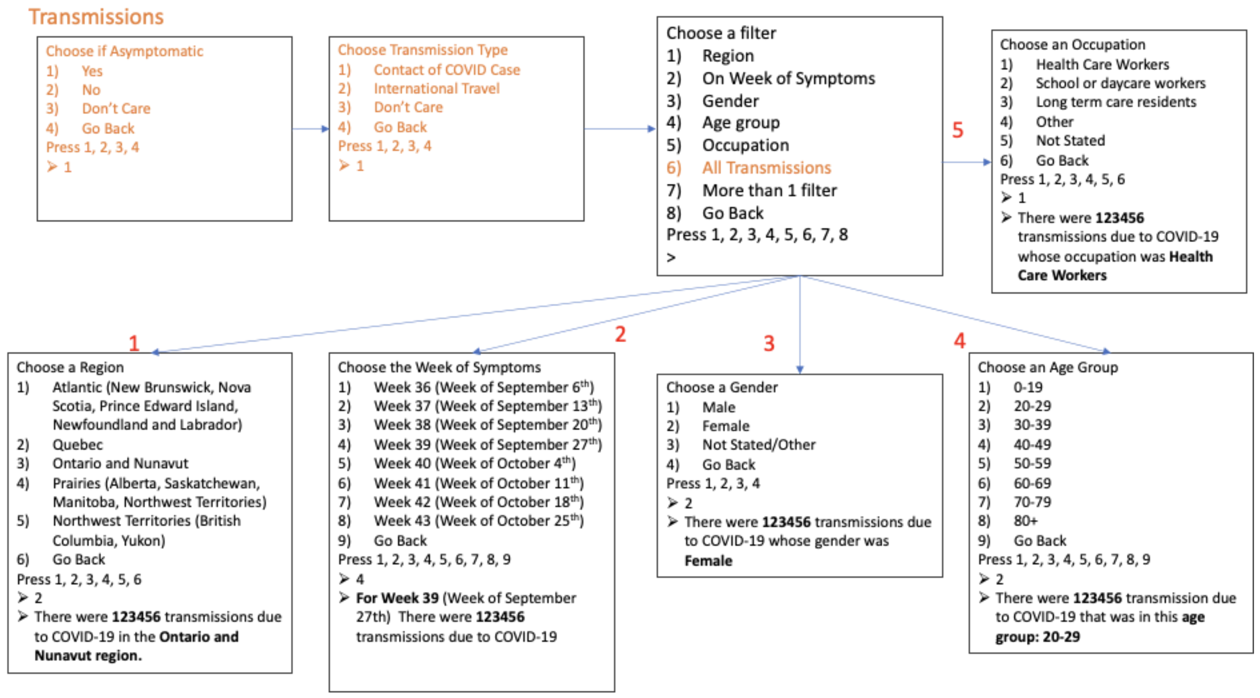
Choose a Timeline
1) Week 36 (Week of September 6th)
2) Week 37 (Week of September 13th)
3) Week 38 (Week of September 20th)
4) Week 39 (Week of September 27th)
5) Week 40 (Week of October 4th)
6) Week 41 (Week of October 11th)
7) Week 42 (Week of October 18th)
8) Week 43 (Week of October 25th)
9) Go Back
Press 1, 2, 3, 4, 5, 6, 7, 8, 9
➢ 4
➢ For **Week 39** (Week of September 27th) there were **123456** recovered cases of COVID-19.

Choose a Gender
1) Male
2) Female
3) Not Stated/Other
4) Go Back
Press 1, 2, 3, 4
➢ 2
➢ There were **123456** recovered cases of COVID-19 whose gender was **Female**

Choose an Age Group
1) 0-19
2) 20-29
3) 30-39
4) 40-49
5) 50-59
6) 60-69
7) 70-79
8) 80+
9) Go Back
Press 1, 2, 3, 4, 5, 6, 7, 8, 9
➢ 2
➢ There were **123456** recovered cases of COVID-19 that were in this **age group: 20-29**

Figure 4: Confirmed Recovery Flow

The next view on the client interface would be Transmissions. In this view, users are able to filter by the same criteria as general information but can filter by positive cases that are asymptomatic and their predicted transmission type as well. This can be useful information for users looking at how travelling and close contact affects different populations as well as ratios between symptomatic and asymptomatic groups.



```
Choose if Asymptomatic
1)    Yes
2)    No
3)    Don't Care
4)    Go Back
Press 1, 2, 3, 4
➢ 1
```

```
Choose Transmission Type
1)    Contact of COVID Case
2)    International Travel
3)    Don't Care
4)    Go Back
Press 1, 2, 3, 4
➢ 1
```

```
Choose a filter
1)    Region
2)    On Week of Symptoms
3)    Gender
4)    Age group
5)    Occupation
6)    All Transmissions
7)    More than 1 filter
8)    Go Back
Press 1, 2, 3, 4, 5, 6, 7, 8
>
```

```
Choose an Occupation
1)    Health Care Workers
2)    School or daycare workers
3)    Long term care residents
4)    Other
5)    Not Stated
6)    Go Back
Press 1, 2, 3, 4, 5, 6
➢ 1
➢ There were 123456
  transmissions due to COVID-19
  whose occupation was Health
  Care Workers
```

```
Choose a Region
1)    Atlantic (New Brunswick, Nova
      Scotia, Prince Edward Island,
      Newfoundland and Labrador)
2)    Quebec
3)    Ontario and Nunavut
4)    Prairies (Alberta, Saskatchewan,
      Manitoba, Northwest Territories)
5)    Northwest Territories (British
      Columbia, Yukon)
6)    Go Back
Press 1, 2, 3, 4, 5, 6
➢ 2
➢ There were 123456 transmissions due
  to COVID-19 in the Ontario and
  Nunavut region.
```

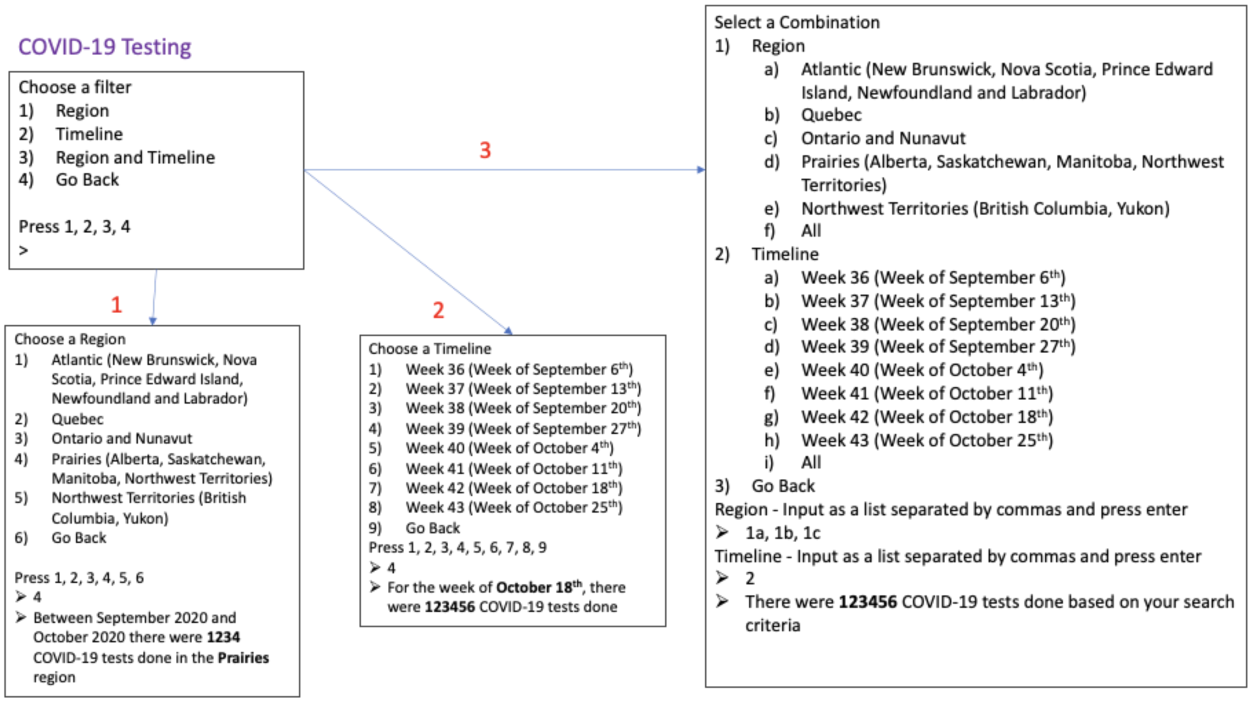
```
Choose the Week of Symptoms
1)    Week 36 (Week of September 6th)
2)    Week 37 (Week of September 13th)
3)    Week 38 (Week of September 20th)
4)    Week 39 (Week of September 27th)
5)    Week 40 (Week of October 4th)
6)    Week 41 (Week of October 11th)
7)    Week 42 (Week of October 18th)
8)    Week 43 (Week of October 25th)
9)    Go Back
Press 1, 2, 3, 4, 5, 6, 7, 8, 9
➢ 4
➢ For Week 39 (Week of September
  27th) There were 123456
  transmissions due to COVID-19
```

```
Choose a Gender
1)    Male
2)    Female
3)    Not Stated/Other
4)    Go Back
Press 1, 2, 3, 4
➢ 2
➢ There were 123456 transmissions due
  to COVID-19 whose gender was
  Female
```

```
Choose an Age Group
1)    0-19
2)    20-29
3)    30-39
4)    40-49
5)    50-59
6)    60-69
7)    70-79
8)    80+
9)    Go Back
Press 1, 2, 3, 4, 5, 6, 7, 8, 9
➢ 2
➢ There were 123456 transmission due
  to COVID-19 that was in this age
  group: 20-29
```

Figure 5: Transmissions Flow

The final standalone view we implemented was COVID-19 weekly testing which can be further filtered by region, timeline, and a combination of both.



```
Choose a filter
1)   Region
2)   Timeline
3)   Region and Timeline
4)   Go Back

Press 1, 2, 3, 4
>
```

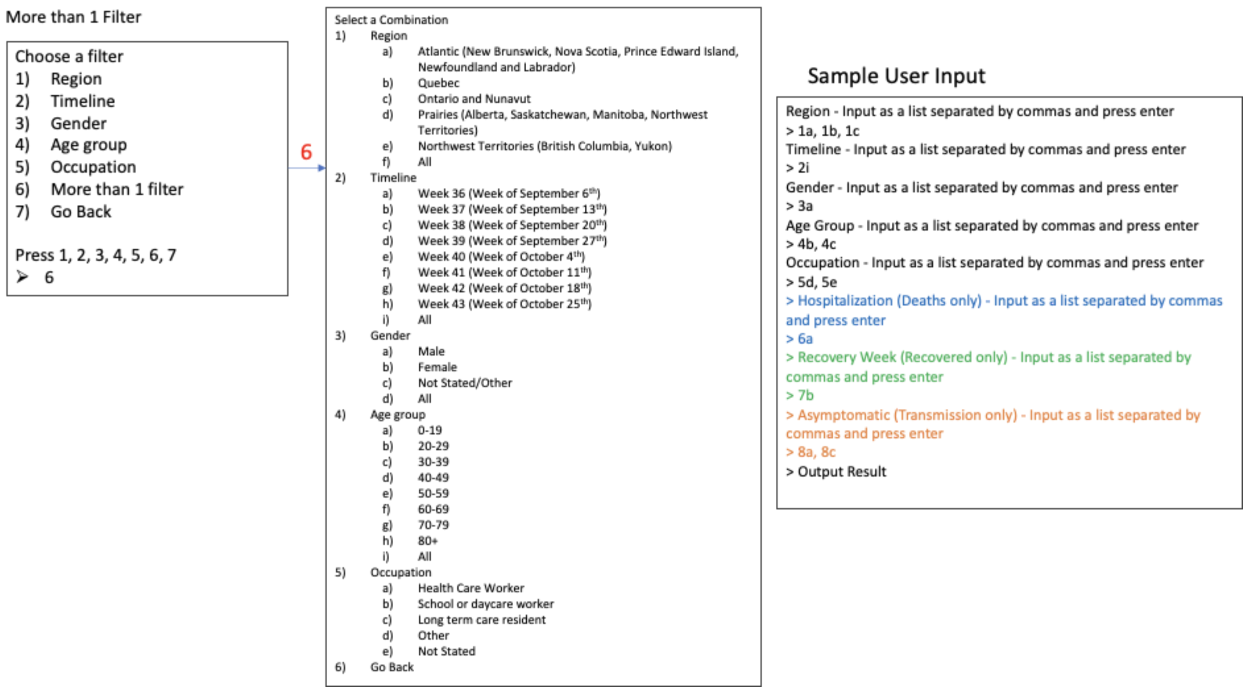
```
Choose a Region
1)   Atlantic (New Brunswick, Nova
     Scotia, Prince Edward Island,
     Newfoundland and Labrador)
2)   Quebec
3)   Ontario and Nunavut
4)   Prairies (Alberta, Saskatchewan,
     Manitoba, Northwest Territories)
5)   Northwest Territories (British
     Columbia, Yukon)
6)   Go Back

Press 1, 2, 3, 4, 5, 6
➢ 4
➢ Between September 2020 and
  October 2020 there were 1234
  COVID-19 tests done in the Prairies
  region
```

```
Choose a Timeline
1)   Week 36 (Week of September 6th)
2)   Week 37 (Week of September 13th)
3)   Week 38 (Week of September 20th)
4)   Week 39 (Week of September 27th)
5)   Week 40 (Week of October 4th)
6)   Week 41 (Week of October 11th)
7)   Week 42 (Week of October 18th)
8)   Week 43 (Week of October 25th)
9)   Go Back
Press 1, 2, 3, 4, 5, 6, 7, 8, 9
➢ 4
➢ For the week of October 18th, there
  were 123456 COVID-19 tests done
```

```
Select a Combination
1)   Region
     a)   Atlantic (New Brunswick, Nova Scotia, Prince Edward
          Island, Newfoundland and Labrador)
     b)   Quebec
     c)   Ontario and Nunavut
     d)   Prairies (Alberta, Saskatchewan, Manitoba, Northwest
          Territories)
     e)   Northwest Territories (British Columbia, Yukon)
     f)   All
2)   Timeline
     a)   Week 36 (Week of September 6th)
     b)   Week 37 (Week of September 13th)
     c)   Week 38 (Week of September 20th)
     d)   Week 39 (Week of September 27th)
     e)   Week 40 (Week of October 4th)
     f)   Week 41 (Week of October 11th)
     g)   Week 42 (Week of October 18th)
     h)   Week 43 (Week of October 25th)
     i)   All
3)   Go Back
Region - Input as a list separated by commas and press enter
➢  1a, 1b, 1c
Timeline - Input as a list separated by commas and press enter
➢  2
➢  There were 123456 COVID-19 tests done based on your search
   criteria
```

Figure 6: Weekly Testing Flow

Though it will be useful to filter COVID-19 data based on each view and a filter, our team found that it may also be useful to allow users to input a combination of filters for each view. With this flow, the user will be prompted using questions to determine the final query. They'll be able to choose multiple options from each field to determine the amount of cases of COVID-19 from our dataset.



More than 1 Filter

Choose a filter
1) Region
2) Timeline
3) Gender
4) Age group
5) Occupation
6) More than 1 filter
7) Go Back

Press 1, 2, 3, 4, 5, 6, 7
➢ 6

6

Select a Combination
1) Region
   a) Atlantic (New Brunswick, Nova Scotia, Prince Edward Island, Newfoundland and Labrador)
   b) Quebec
   c) Ontario and Nunavut
   d) Prairies (Alberta, Saskatchewan, Manitoba, Northwest Territories)
   e) Northwest Territories (British Columbia, Yukon)
   f) All
2) Timeline
   a) Week 36 (Week of September 6th)
   b) Week 37 (Week of September 13th)
   c) Week 38 (Week of September 20th)
   d) Week 39 (Week of September 27th)
   e) Week 40 (Week of October 4th)
   f) Week 41 (Week of October 11th)
   g) Week 42 (Week of October 18th)
   h) Week 43 (Week of October 25th)
   i) All
3) Gender
   a) Male
   b) Female
   c) Not Stated/Other
   d) All
4) Age group
   a) 0-19
   b) 20-29
   c) 30-39
   d) 40-49
   e) 50-59
   f) 60-69
   g) 70-79
   h) 80+
   i) All
5) Occupation
   a) Health Care Worker
   b) School or daycare worker
   c) Long term care resident
   d) Other
   e) Not Stated
6) Go Back

Sample User Input

Region - Input as a list separated by commas and press enter
> 1a, 1b, 1c
Timeline - Input as a list separated by commas and press enter
> 2i
Gender - Input as a list separated by commas and press enter
> 3a
Age Group - Input as a list separated by commas and press enter
> 4b, 4c
Occupation - Input as a list separated by commas and press enter
> 5d, 5e
> Hospitalization (Deaths only) - Input as a list separated by commas and press enter
> 6a
> Recovery Week (Recovered only) - Input as a list separated by commas and press enter
> 7b
> Asymptomatic (Transmission only) - Input as a list separated by commas and press enter
> 8a, 8c
> Output Result

Figure 7: Combined Filter Flow

Client Experience - Modifying and Deleting Data

Within the client application, certain users are also able to modify existing data. There will be 2 types of users that have the ability to do this, a user with a regular status and government status.

With a regular status, a login is not needed, these users are able to report a new COVID-19 case (INSERT) and add a new location to the database (INSERT). In the flow below, the user will be asked similar questions using the same filters for viewing data (ex: region, timeline, gender, etc). After inputting the following fields, the entry will be added to the BackgroundInfo, Deaths and Recovered tables appropriately with an incremented caseID.

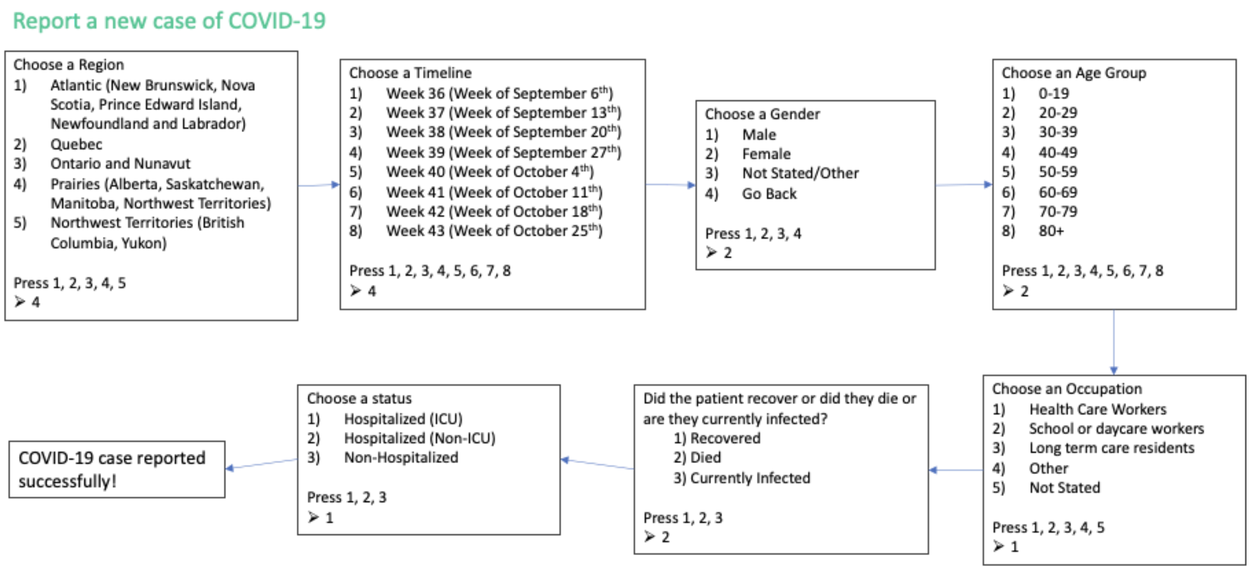

Report a new case of COVID-19
Choose a Region
1) Atlantic (New Brunswick, Nova Scotia, Prince Edward Island, Newfoundland and Labrador)
2) Quebec
3) Ontario and Nunavut
4) Prairies (Alberta, Saskatchewan, Manitoba, Northwest Territories)
5) Northwest Territories (British Columbia, Yukon)
Press 1, 2, 3, 4, 5
➢ 4
Choose a Timeline
1) Week 36 (Week of September 6th)
2) Week 37 (Week of September 13th)
3) Week 38 (Week of September 20th)
4) Week 39 (Week of September 27th)
5) Week 40 (Week of October 4th)
6) Week 41 (Week of October 11th)
7) Week 42 (Week of October 18th)
8) Week 43 (Week of October 25th)
Press 1, 2, 3, 4, 5, 6, 7, 8
➢ 4
Choose a Gender
1) Male
2) Female
3) Not Stated/Other
4) Go Back
Press 1, 2, 3, 4
➢ 2
Choose an Age Group
1) 0-19
2) 20-29
3) 30-39
4) 40-49
5) 50-59
6) 60-69
7) 70-79
8) 80+
Press 1, 2, 3, 4, 5, 6, 7, 8
➢ 2
Choose an Occupation
1) Health Care Workers
2) School or daycare workers
3) Long term care residents
4) Other
5) Not Stated
Press 1, 2, 3, 4, 5
➢ 1
Did the patient recover or did they die or are they currently infected?
1) Recovered
2) Died
3) Currently Infected
Press 1, 2, 3
➢ 2
Choose a status
1) Hospitalized (ICU)
2) Hospitalized (Non-ICU)
3) Non-Hospitalized
Press 1, 2, 3
➢ 1
COVID-19 case reported successfully!

Figure 8: Reporting a New Case Flow

The user is also able to add a new location to the database (INSERT). Upon entering a city and province name, this location will be added to the Location table and will show up when users try to view and modify data.

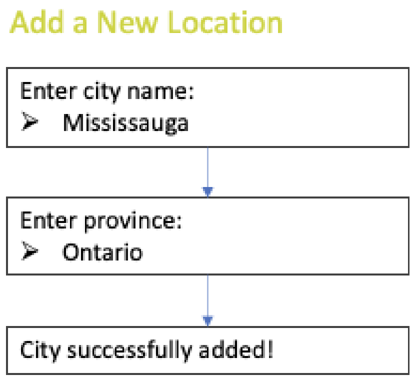
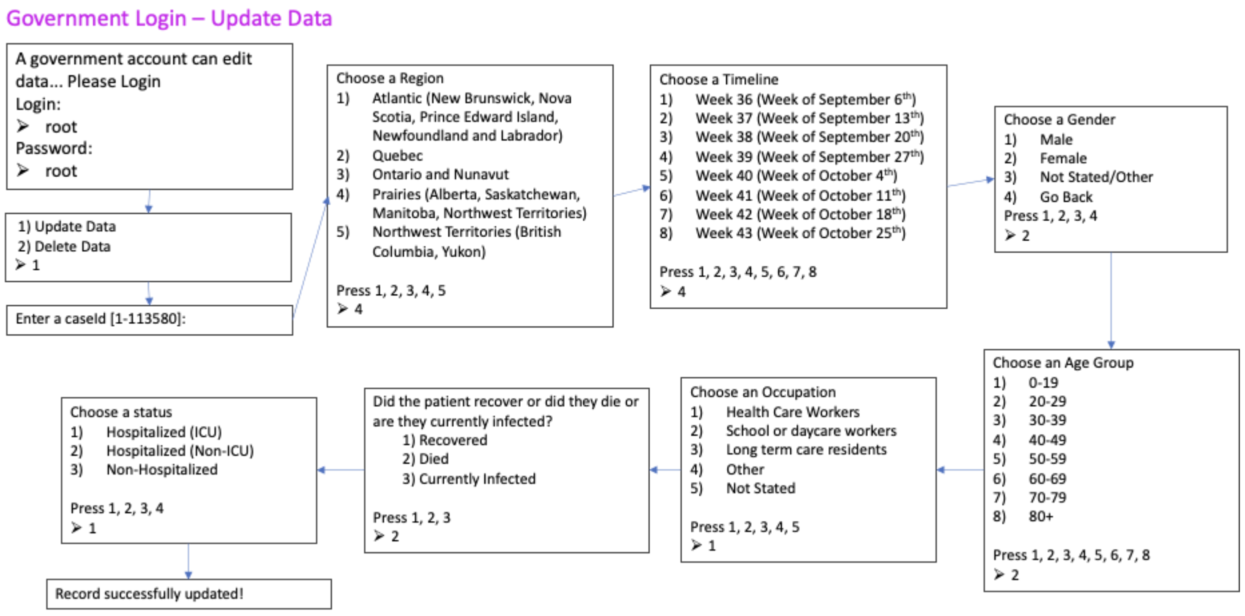


Figure 9: Adding a New Location Flow

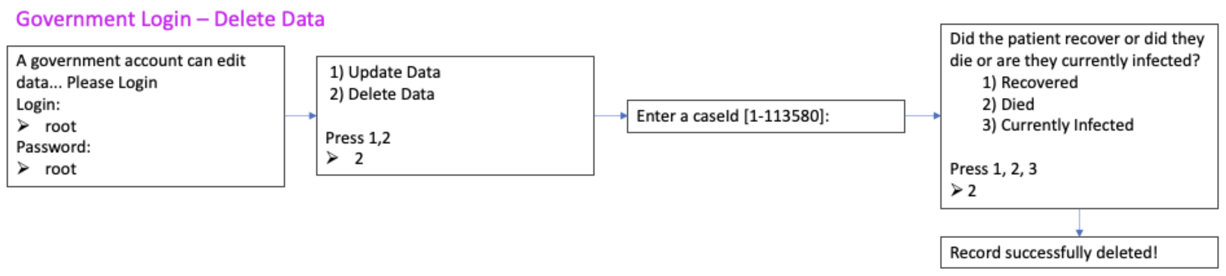
The second type of user would be those of government status. This type of user is able to modify and delete cases from the database. Upon entering the correct username and password ("root" and "root" in our case), if the user wants to update a COVID-19 case, they will specify the case they would like to change and re-enter data surrounding this caseID. Once all of the appropriate information has been entered, this entry will be reloaded into the database with the updated information.



```
Government Login – Update Data

A government account can edit
data... Please Login
Login:
➢ root
Password:
➢ root

1) Update Data
2) Delete Data
➢ 1

Enter a caseId [1-113580]:

Choose a Region
1)  Atlantic (New Brunswick, Nova
    Scotia, Prince Edward Island,
    Newfoundland and Labrador)
2)  Quebec
3)  Ontario and Nunavut
4)  Prairies (Alberta, Saskatchewan,
    Manitoba, Northwest Territories)
5)  Northwest Territories (British
    Columbia, Yukon)

Press 1, 2, 3, 4, 5
➢ 4

Choose a Timeline
1)  Week 36 (Week of September 6th)
2)  Week 37 (Week of September 13th)
3)  Week 38 (Week of September 20th)
4)  Week 39 (Week of September 27th)
5)  Week 40 (Week of October 4th)
6)  Week 41 (Week of October 11th)
7)  Week 42 (Week of October 18th)
8)  Week 43 (Week of October 25th)

Press 1, 2, 3, 4, 5, 6, 7, 8
➢ 4

Choose a Gender
1)  Male
2)  Female
3)  Not Stated/Other
4)  Go Back
Press 1, 2, 3, 4
➢ 2

Choose an Age Group
1)  0-19
2)  20-29
3)  30-39
4)  40-49
5)  50-59
6)  60-69
7)  70-79
8)  80+

Press 1, 2, 3, 4, 5, 6, 7, 8
➢ 2

Choose an Occupation
1)  Health Care Workers
2)  School or daycare workers
3)  Long term care residents
4)  Other
5)  Not Stated

Press 1, 2, 3, 4, 5
➢ 1

Did the patient recover or did they die or
are they currently infected?
    1) Recovered
    2) Died
    3) Currently Infected

Press 1, 2, 3
➢ 2

Choose a status
1)  Hospitalized (ICU)
2)  Hospitalized (Non-ICU)
3)  Non-Hospitalized

Press 1, 2, 3, 4
➢ 1

Record successfully updated!
```

Figure 10: Government Login and Updating Data Flow

The final flow that a user of government status has is the ability to remove data from the database. In this flow, a user can specify a caseID that they would like to remove from the database and this entry will be first removed from tables that have a foreign key to BackgroundInformation and then it is removed from BackgroundInformation.



```
Government Login – Delete Data

A government account can edit
data... Please Login
Login:
➢ root
Password:
➢ root

1) Update Data
2) Delete Data

Press 1,2
➢ 2

Enter a caseId [1-113580]:

Did the patient recover or did they
die or are they currently infected?
    1) Recovered
    2) Died
    3) Currently Infected

Press 1, 2, 3
➢ 2

Record successfully deleted!
```

Figure 11: Government Login and Deleting Data Flow

We chose the above flows to implement for our project as they best aligned with our goal to focus on COVID-19 affecting Canadians and hospitals. Allowing users to look up background information, deaths and recovered information and transmissions provides insights into the overall health of different populations. As well, allowing for different users to modify data ensures we have accurate additions and modifications to the data (only government access to modify existing data). Which will ensure that future viewing will result in an accurate representation of the health of Canadians through COVID-19.

ER Design

Our entity relational design is as follows. We start with a Background Information table that has an overview of a positive COVID-19 case. This data is connected to a Transmissions, Recovered, and Deaths table through a caseID primary key. It is also connected to a Location. Additionally, our team included weekly testing which has a primary key of region, abbreviation and episodeWeek. Looking at our relationship sets, we can see that Background Info and Transmissions are related via the Transmits relationship set. Background Info and Recovered as well as Background Info and Deaths are related via the Survives relationship set. This is also an example of a disjoint generation since a person cannot be both dead and have recovered, therefore the Survives relation will go into one of Recovered or Deaths, but never both. Finally, Background Info and Location as well as Weekly Testing and Location are related via the Located relationship set. We selected these entity and relationship sets mainly due to the lack of available COVID-19 data on StatsCan and other online sources in relation to the health of Canadians. Our team wanted to ensure that our data provided a good representation of all factors that can affect the severity of a COVID-19 case (gender, age, etc.) and all outcomes that could result.

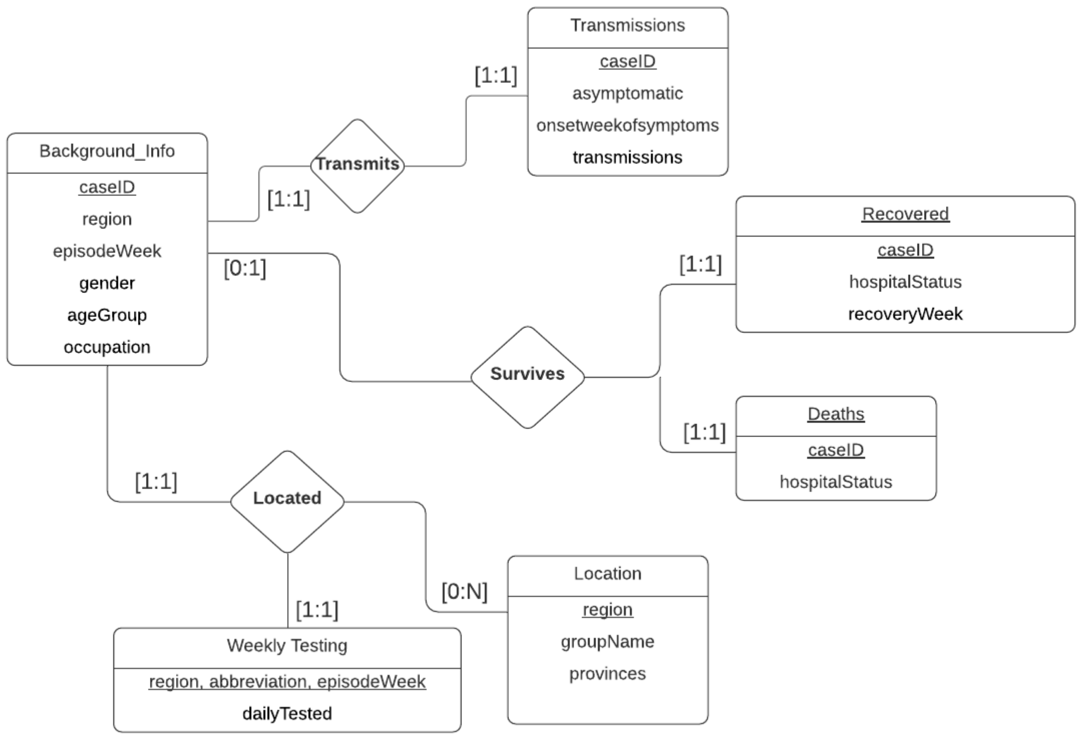


Figure 12: Entity Relationship Diagram for Current Dataset

Due to the limitations in the available online datasets for Canadian COVID-19 data and StatsCan, we have also included a more detailed ER design with hypothetical information coloured in red. Though we were unable to find accurate available data online for the hypothetical tables and attributes, we found that including these would allow us to further align with the project description and goals and make for a more complete project.

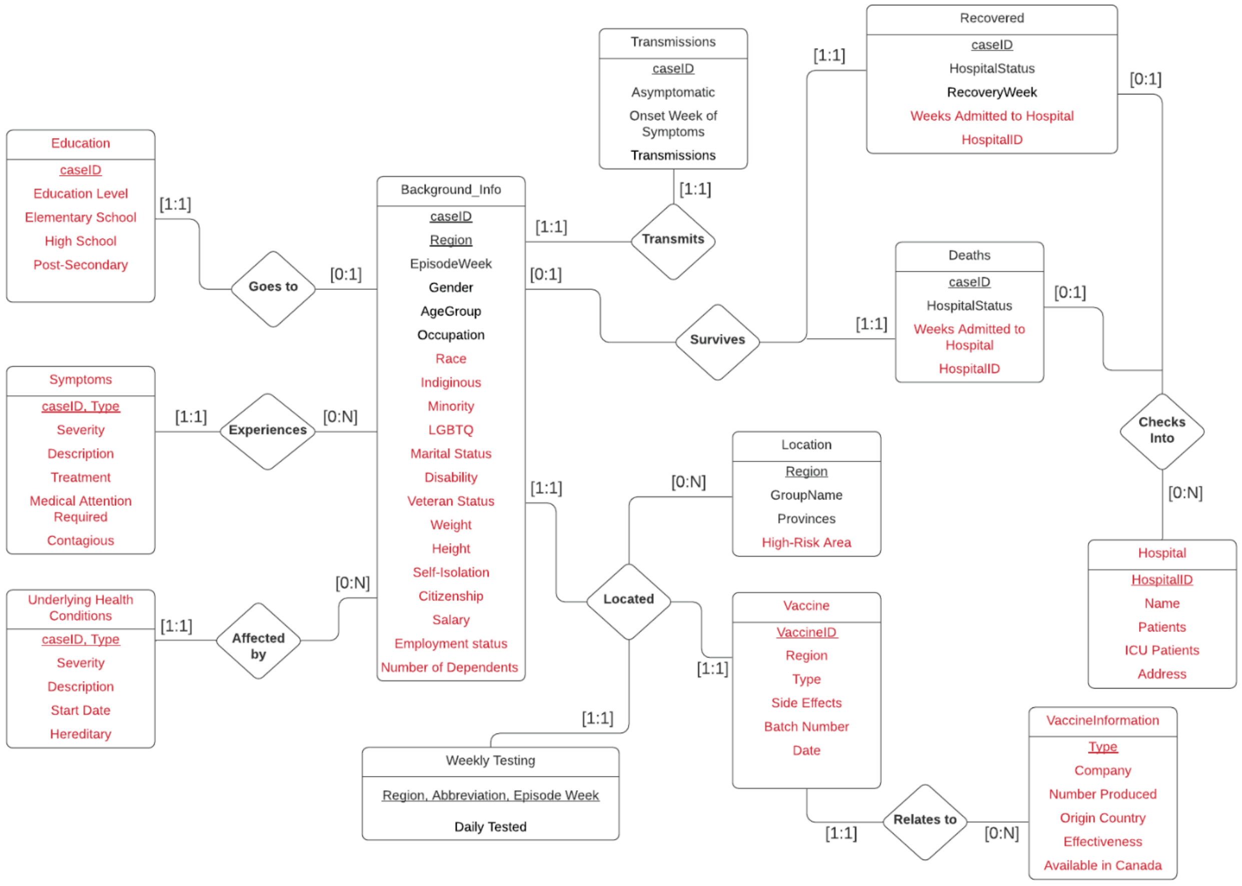


Figure 13: Entity Relationship Diagram for Hypothetical Additional Information

Relational Schema

In order to transform our ER model into a relational schema, we first reduced it to a relational schema with the same name and attributes, which is shown in the final relational schema diagram. This was done by creating a table for each entity and then mapping the entity's attributes to fields of the tables with their respective data types. For instance, we declared a Background Info table with the attributes: caseID, Region, EpisodeWeek, Gender, AgeGroup and Occupation where caseID was of type int and the rest were of type decimal. Finally, we declared our primary keys and foreign keys based on the ER design, which is displayed in the final schema table specification.

The code for the relational schema can be found here:
https://github.com/naz321/356-Project/tree/main/Create_Scripts

Final Relational Schema Diagram

The final relational schema diagram below outlines the blueprint of our COVID-19 database. It displays the structure of the tables and the relation between each table. The COVID-19 data has been organized into six different tables: Weekly Testing, Location, Background Info, Recovered, Deaths, and Transmissions. Furthermore, there is a zero-to-many relationship between Location and Weekly Testing since a location may have zero, 1 or many weekly testings. The same can be said about the Location/Background Info relation. On the other hand, Background Info has a zero-to-one cardinality constraint with both Recovered and Deaths. This is because a person does not necessarily have to recover or die from COVID-19 (since some of our data for this is not stated), however if they do, then they can recover or die from COVID-19 exactly once. The rest of the entities have a one-to-one relationship, since one occurrence of an entity relates to only one occurrence of another entity in the remaining relations.

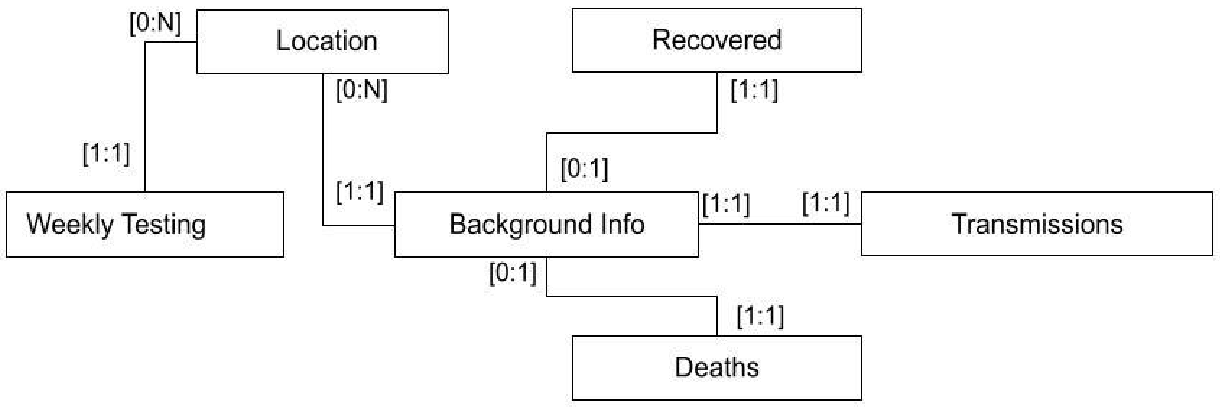


Figure 14: Relational Schema Diagram

Final Schema Table Specification

Please note that the primary keys have been underlined and the foreign keys are labelled as FK in the following diagram:

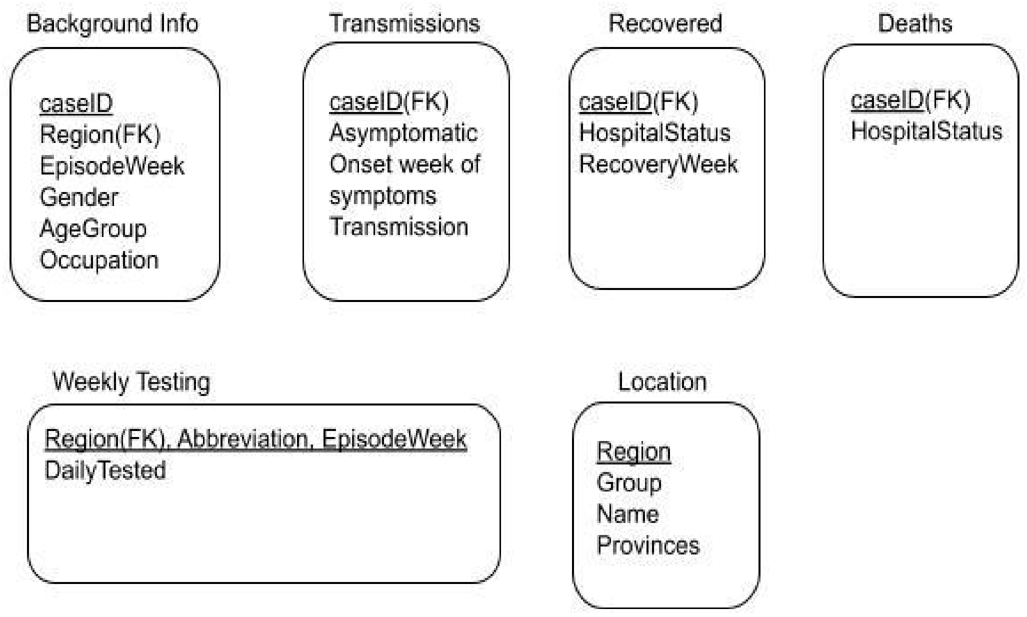


Figure 15: Final Schema Table Specification

Data Mining Investigation

Data mining question

The following scenario is happening all over the world right now and especially in Canada:

Hospital ICUs are completely overwhelmed and there is just not enough space. If there were two COVID-19 patients and only 1 ICU spot left, who would the hospital treat first? Would they treat the first patient in a first come first serve manner? Or perhaps they would treat the older patient first? How would they choose which patient to give the ICU to.

This following scenario was the backbone of our data mining investigation. From this, we were able to develop the following question: Given past records of COVID-19 patient data and their survival rate, would it be possible to predict the survival rate of future COVID-19 patients? If this is possible, this would have potentially large implications such as allowing a hospital to efficiently allocate resources to those who need it the most. Hospitals could also prioritize patients based on their chance of survival. This is what the data mining activity intends to do.

A user (ex: hospital) would input a hypothetical situation and the CLI will determine the rate of survivability and mortality. A user will be able to control 5 inputs:

1) Region

2) Week
3) Gender
4) Age group
5) Occupation

For example, a user could enter the following hypothetical situation:

1) Region = 2 (Quebec)
2) Week = 5 (Week of October 4th)
3) Gender = 1 (Male)
4) Age Group = 8 (80+)
5) Occupation = 4 (Not Stated)

And the following output will be shown to the user:

Given the following scenario, you have a 69.88% chance of surviving and a 30.12% of dying.

This would be extremely beneficial to hospitals. Given that they could enter 2 patient's background information, they could figure out who has the higher survival rate and make a resource decision based on that.

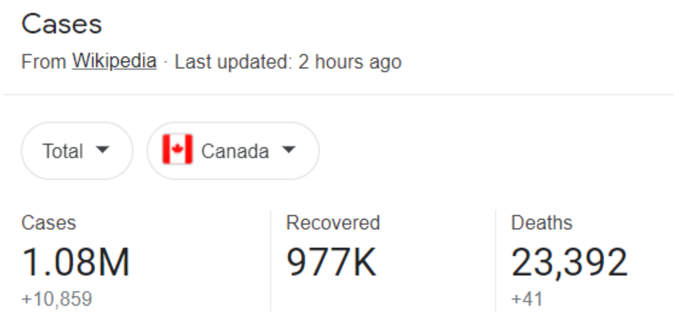
The proposed data-mining technique that was used would be a decision tree. More specifically, the decision tree machine learning model provided by scikit-learn library will be used. Decision trees are simple models for supervised classification. Each internal node performs a Boolean test on an inputted feature (in our situation, since each input feature test has more than 2 options, it'll be converted into a series of Boolean tests). Each edge in the tree is an answer to the boolean test, either true or false. The leaf nodes specifies the ending and determines if a patient has survived or died.

What data was required

The data that is required would be the patient's background data. For example, their location, their age, occupation, gender, etc. Fortunately the majority of this data is found already in the BackgroundInformation.csv

Preprocessing the data

The data that we performed data mining on is located in BackgroundInformation.csv. Since there were only ~114,000 records and only around ~1500 people died, the majority of the hypothetical scenarios that would be inputted would result in a higher survival chance than mortality rate. This relationship conforms with actual real life statistics from COVID-19 cases in Canada.


Cases

From Wikipedia · Last updated: 2 hours ago

Cases	Recovered	Deaths
1.08M	977K	23,392
+10,859		+41

Figure 16: Ratio of real life recovered vs. deaths follows the same relationship as our data

The data in the csv was mostly in the correct form needed for building a decision tree. However, some preparations and pruning was required to get it in the correct form.

1) There is a death column that indicates if the patient died or not. However, there are 3 values for it (1, 2, 9). 1 = Died, 2 = Recovered, 9 = Not Stated. Because living is a binary state, either living or dead, columns where death = 9 were removed.
2) There were a lot of other attributes that were available. However, the 5 choices that were chosen would best describe our data mining question. Also, if there were more inputs, the breadth and depth of the decision tree would grow exponentially.
3) For the age group column, there are 9 possible values, (1=0-19, 2=20-29, 3=30-39, 4=40-49, 5=50-59, 6=60-69, 7=70-79, 8=80+, 99=Not stated). The 99 would add noise and disturb the construction of the decision tree. Therefore, all columns where the age group was not stated was removed.

Unfortunately the data set was on the smaller side. To answer the question more accurately and to increase the certainty of the decision tree, we would require a much bigger data set.

Data mining results

The full output of the decision tree is available to be downloaded here:
https://github.com/naz321/356-Project/blob/main/Data_Mining/COVID_19_Decision_Tree.pdf

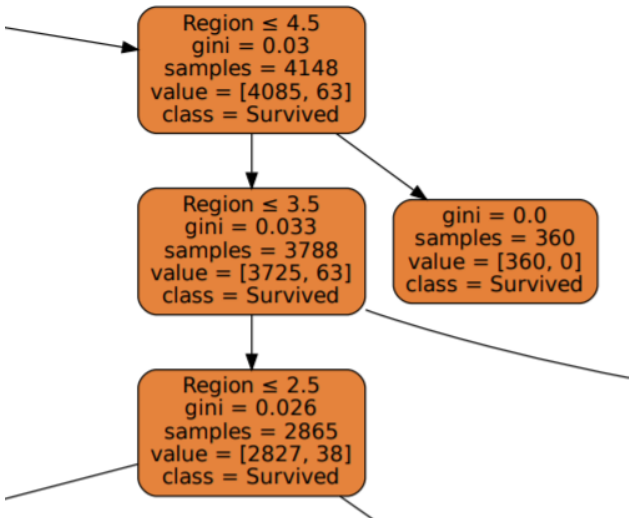


Figure 17: Snippet of the Decision Tree Created

In the diagram, if the region is less than 4.5 we go to the left branch. If the region is greater than 4.5 (ex: 5), we go to the right branch. Since the right branch has no more children it means that it is a leaf node and that is the end of the decision tree. In the node, it says class = "Survived", which means the person has survived.

It is important to note that when you read the decision tree and the nodes, most of the feature questions (ex: region <= 2.5) will show up as decimals instead of integer. This might appear wrong since only integer data was fed to train the model and integer data was inputted to create a prediction.

However, it is actually correct. The decision tree model is an example of a classification tree where the decision of each input feature is a yes or no. In other words, the model is a binary classifier and the construction of the tree is based on binary recursive partitioning. Since our input features test supports more than 2 options (ex: Pick a region between 1-5) it can't be answered by a simple yes or no. Rather, it is converted into a series of boolean tests. As a result, the nature of the data (not binary) causes the question to have decimal values.

There is no problem with the feature test question being a decimal as opposed to integer. We can treat a question such as (region <= 2.5) to traverse the left branch if region is [1,2] and traverse the right branch if region is [3,4,5].

Interesting observations from the data mining activity

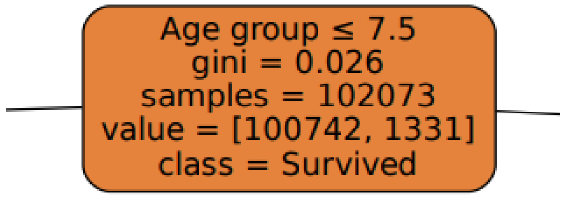


Figure 18: Root Node of the Decision Tree

The first question asked in the decision tree (root) is if the age group is <= 7.5. If the age group is [1,2,3,4,5,6,7] the decision tree would traverse left, and if the age group is [8], the decision tree would traverse right. The “value” attribute is a tuple with 2 values. The first value (100742) is the # of samples in the data set that has an age group <= 7.5. The second value (1331) is the # of samples in the data set that has an age group > 7.5. Taking a look at the entire left branch and its subtrees, all leaf nodes in the left subtree survived. Therefore, based on the training data, if a user inputs an age group <= 7.5 we can immediately say that the patient should survive. If it doesn’t, we would need to traverse the tree further and answer more feature questions to determine survivability.

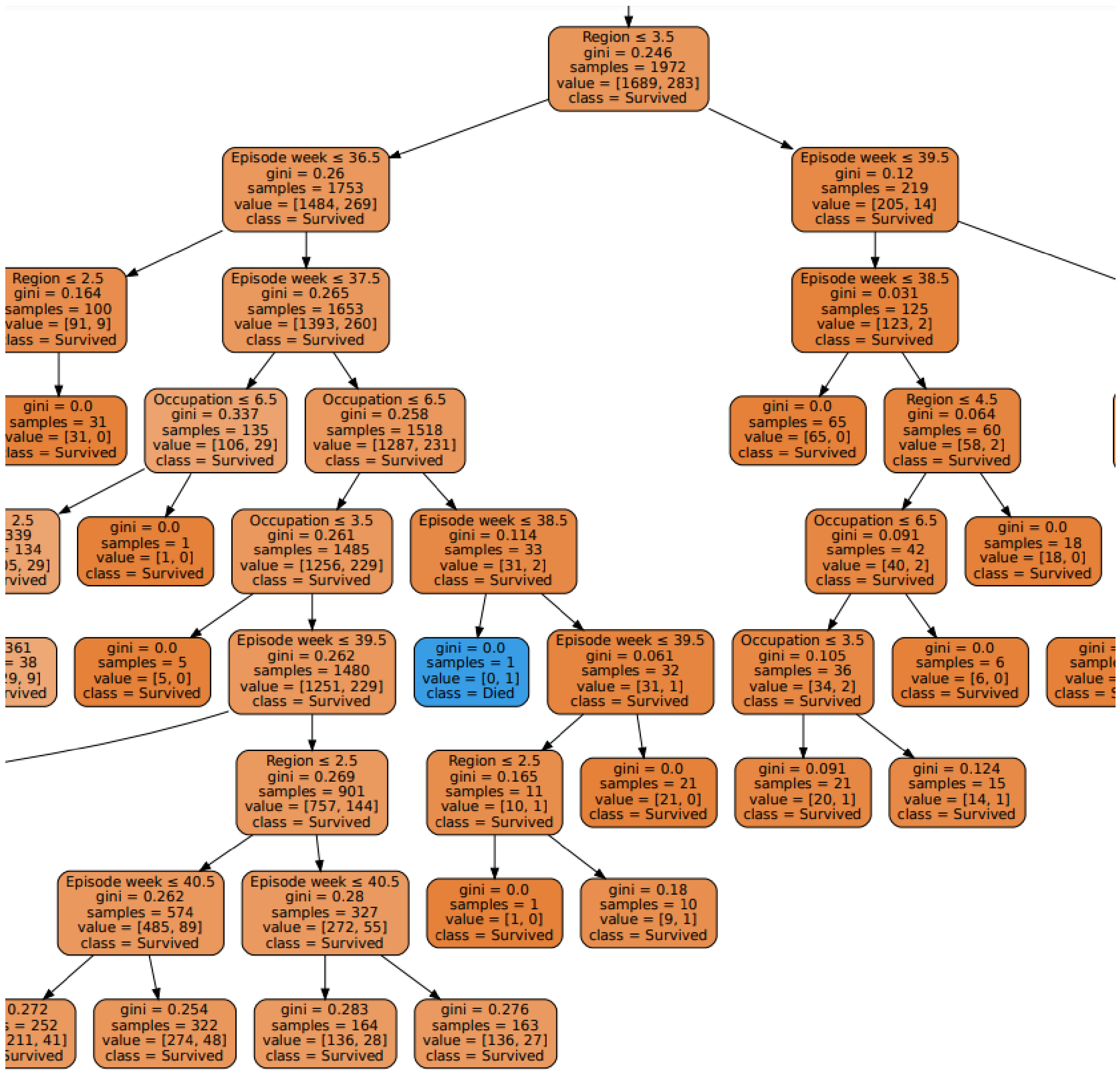

Region ≤ 3.5 gini = 0.246 samples = 1972 value = [1689, 283] class = Survived
Episode week ≤ 36.5 gini = 0.26 samples = 1753 value = [1484, 269] class = Survived
Episode week ≤ 39.5 gini = 0.12 samples = 219 value = [205, 14] class = Survived
Region ≤ 2.5 gini = 0.164 samples = 100 value = [91, 9] lass = Survived
Episode week ≤ 37.5 gini = 0.265 samples = 1653 value = [1393, 260] class = Survived
Episode week ≤ 38.5 gini = 0.031 samples = 125 value = [123, 2] class = Survived
gini = 0.0 samples = 31 value = [31, 0] lass = Survived
Occupation ≤ 6.5 gini = 0.337 samples = 135 value = [106, 29] class = Survived
Occupation ≤ 6.5 gini = 0.258 samples = 1518 value = [1287, 231] class = Survived
gini = 0.0 samples = 65 value = [65, 0] class = Survived
Region ≤ 4.5 gini = 0.064 samples = 60 value = [58, 2] class = Survived
gini = 0.0 samples = 1 value = [1, 0] class = Survived
Occupation ≤ 3.5 gini = 0.261 samples = 1485 value = [1256, 229] class = Survived
Episode week ≤ 38.5 gini = 0.114 samples = 33 value = [31, 2] class = Survived
Occupation ≤ 6.5 gini = 0.091 samples = 42 value = [40, 2] class = Survived
gini = 0.0 samples = 18 value = [18, 0] class = Survived
gini = 0.0 samples = 5 value = [5, 0] class = Survived
Episode week ≤ 39.5 gini = 0.262 samples = 1480 value = [1251, 229] class = Survived
gini = 0.0 samples = 1 value = [0, 1] class = Died
Episode week ≤ 39.5 gini = 0.061 samples = 32 value = [31, 1] class = Survived
Occupation ≤ 3.5 gini = 0.105 samples = 36 value = [34, 2] class = Survived
gini = 0.0 samples = 6 value = [6, 0] class = Survived
Region ≤ 2.5 gini = 0.269 samples = 901 value = [757, 144] class = Survived
Region ≤ 2.5 gini = 0.165 samples = 11 value = [10, 1] class = Survived
gini = 0.0 samples = 21 value = [21, 0] class = Survived
gini = 0.091 samples = 21 value = [20, 1] class = Survived
gini = 0.124 samples = 15 value = [14, 1] class = Survived
Episode week ≤ 40.5 gini = 0.262 samples = 574 value = [485, 89] class = Survived
Episode week ≤ 40.5 gini = 0.28 samples = 327 value = [272, 55] class = Survived
gini = 0.0 samples = 1 value = [1, 0] class = Survived
gini = 0.18 samples = 10 value = [9, 1] class = Survived
gini = 0.254 samples = 322 value = [274, 48] class = Survived
gini = 0.283 samples = 164 value = [136, 28] class = Survived
gini = 0.276 samples = 163 value = [136, 27] class = Survived

Figure 19: Small Section of the Decision Tree

Leaf nodes that are marked in blue represent a patient's death. Out of all of the leaf nodes in this decision tree, there are only 2 leaves that represent a patient's death. This makes sense because out of all the cases, 98.7% recovered and 1.3% died. The usefulness of a decision tree comes in the ability to visualize and traverse the tree in a simple manner. Since there were only 2 leaves that represent a patient's death, it is easy to traverse the tree to determine if the patient survived or died.

Validity of the decision tree

In order to test the validity of the model that was built, K-fold cross-validation was used. K-fold is a statistical method that is used to estimate the skill/accuracy of machine learning models. K-fold attempts to split the data set into training and testing sets and use the training set to train the model and testing set to test the model.

In this data mining activity, the given data set was split into a K number of sections/folds and each fold was used as a testing set at some point. Below is an example of how K-fold cross validation works when K=5.

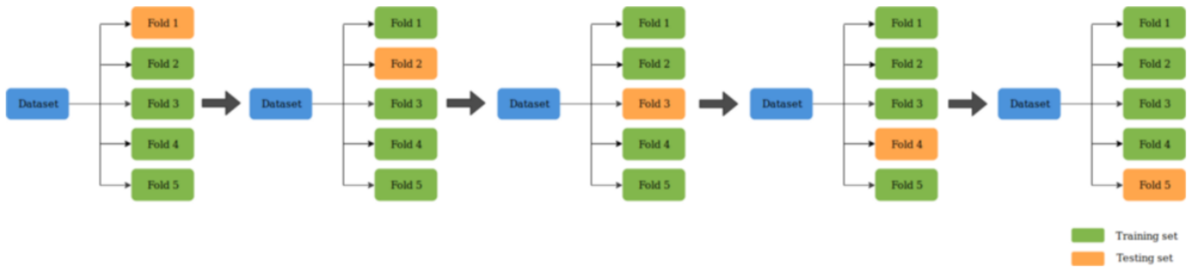


Figure 20: K Fold Visualization
Picture from: https://medium.datadriveninvestor.com/k-fold-cross-validation-6b8518070833

In the first iteration, the first fold is used to test the model and the rest of the folds are used to train the model. In the second iteration, the 2nd fold is used as the testing set while the rest of the folds serve as the training set. This process continues until all folds have been used as a testing set.

To validate our data mining model, K was set to 10. At every iteration, the accuracy of each model was calculated:

```
New accuracy:  0.9879210015870217
New accuracy:  0.984923293951684
New accuracy:  0.9870393228707459
New accuracy:  0.9868629871274908
New accuracy:  0.9879199365135349
New accuracy:  0.9885371660347412
New accuracy:  0.986773653117009
New accuracy:  0.9873027069923287
New accuracy:  0.9853628427828234
New accuracy:  0.988360814742968
Highest Value Accuracy:  0.9885371660347412
```

At the end of the k-folds, the model with the highest accuracy was `0.9885371660347412`. This model was then used to create predictions on hypothetical situations entered by the user. As a side note, this accuracy is extremely high. This is because the original data set was small.

Test-case Plan/How we tested the project

Test 1 (Simple Querying):
Given: The user wants to find the total number of COVID-19 Cases in Quebec.
When: The user selects
> Confirmed COVID-19 Cases
> Region
> Quebec
Then: The user should see the following message "There were 44054 confirmed cases of COVID-19 in the Quebec region."

Test 2 (Simple Querying part 2):
Given: The user wants to find the total number of people who died from COVID-19 that were between the ages 60-69.
When: The user selects
> Deaths From COVID-19
> Age Group
> 60-69
Then: The user should see the following message "There were 89 confirmed deaths due to COVID-19 that was in this age group: 60-69."

Test 3 (Complex Querying - Multi filtering menu selection):
Given: The user wants to find the total number of people who got COVID-19 that fit the following criteria:

- From the Atlantic or Quebec region
- During the week of September 6th 2020
- Male
- Between the ages 0-19
- Were Health Care Workers

When: The user selects the following:
> Confirmed COVID-19 Cases
> More than 1 Filter
> Atlantic (New Brunswick, Nova Scotia, Prince Edward Island & Newfoundland and Labrador), Quebec (Quebec)
> Week 36 (Week of September 6th)
> Male

> 0-19
> Health Care Workers
Then: The user should see the following message “Based on your query there were 4 confirmed cases of COVID-19.”

Test 4 (Reporting a new case of COVID-19 and seeing the results):
Given: The user originally queries the number of COVID-19 Cases in Quebec and figures out that there are **44054** number of cases.
When: The user reports a new case of COVID-19 and specifies that it is from the Quebec region. The user then queries the number of COVID-19 Cases in Quebec again.
Then: The user should see the following message “There were **44055** confirmed cases of COVID-19 in the Quebec region.” The number of cases in Quebec increased by 1 due to the user reporting a new case.

Test 5 (Reporting a new recovery from COVID-19 and seeing the results):
Given: The user originally queries the number of recoveries from COVID-19 that were between the ages 20-29 and figures out that there are **23927** recoveries.
When: The user reports a new recovery from COVID-19 and specifies that the patient is between the ages 20-29. The user then queries the number of recoveries from COVID-19 that were between the ages 20-29.
Then: The user should see the following message “There were **23928** recoveries from COVID-19 that were in this age group: 20-29.” The number of recoveries between the ages 20-29 increased by 1 due to the user reporting a new case.

Test 6 (Adding a new location):
Given: The user originally queries based on region. They should see the original 5 options (Atlantic, Quebec, Ontario and Nunavut, Prairies, Northwest Territories)
When: The user selects the “Add a new Location” option and adds a new location (City = Toronto, Province = Ontario) and the user queries again based on region.
Then: The user should now see 6 options (Atlantic, Quebec, Ontario and Nunavut, Prairies, Northwest Territories, **Toronto**). The new option that was added by the user is now a new region that can be queried.

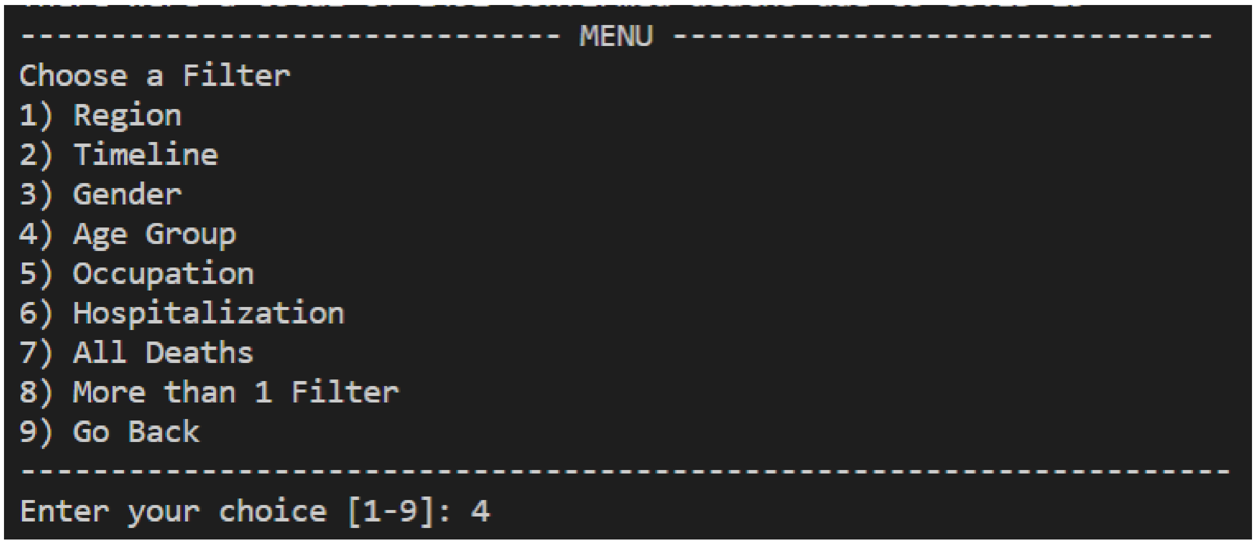
Test 7 (Government Login allows you to update existing data):
Given: A user who is not an admin queries the number of COVID-19 Cases in Quebec and figures out that there are **44054** cases. A new user who is an admin/government official and logs in via the “Government Login” option (Login = root, Password = root). They select the “Update Data” option.

When: The admin enters a caseID to be updated. The caseID is valid and exists in the database. The admin is prompted with questions to update the data and fills it in accordingly (ex: Admin changes the region of the caseID from 1->2).
Then: The record should be updated based on what the admin typed in and a new user who is not an admin queries based on the Quebec region again. They should observe that there are now **44055** cases. The number of cases in Quebec increased by 1 due to the admin updating the information in the database.

Test 8 (Government Login allows you to delete existing data):
Given: A user who is not an admin queries the number of COVID-19 Cases in Quebec and figures out that there are **44054** cases. A new user who is an admin/government official and logs in via the "Government Login" option (Login = root, Password = root). They select the "Delete Data" option.
When: The admin enters a caseId to be updated. The caseID is valid, exists in the database, and has the region set to Quebec. The admin then confirms deletion of the record.
Then: The record should be deleted and a new user who is not an admin queries based on the Quebec region again. They should observe that there are now **44053** cases. The number of cases in Quebec decreased by 1 due to the admin deleting the record in the database.

Other Additional Testing

Testing for wrong input:
Our CLIis a text based menu which is shown below:



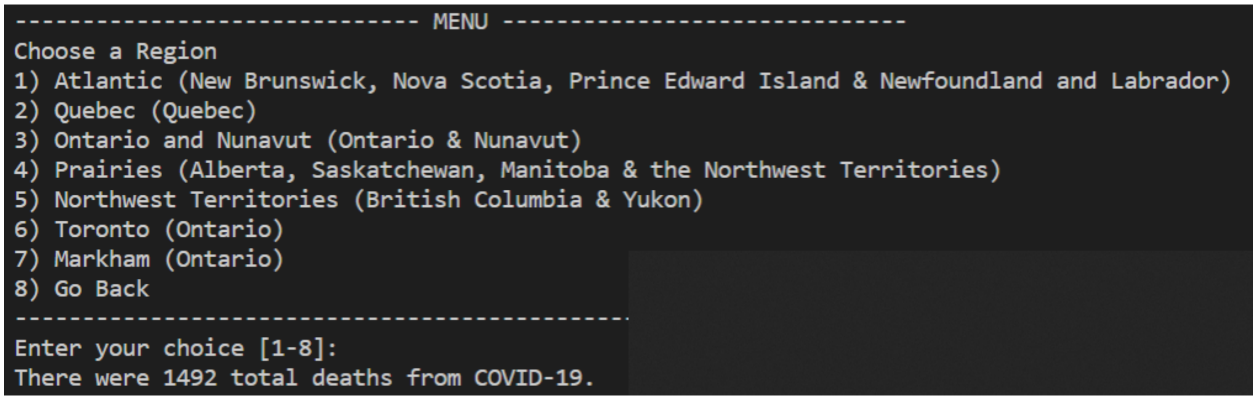
```
------------------------------ MENU ------------------------------
Choose a Filter
1) Region
2) Timeline
3) Gender
4) Age Group
5) Occupation
6) Hospitalization
7) All Deaths
8) More than 1 Filter
9) Go Back
--------------------------------------------------------------------
Enter your choice [1-9]: 4
```

As a result, the user has the freedom to type whatever they want. We do some error checking to prevent the user from typing wrong input. We check the following:

- The inputted value has to be a number
- The inputted value can only be 1 value (unless specified)

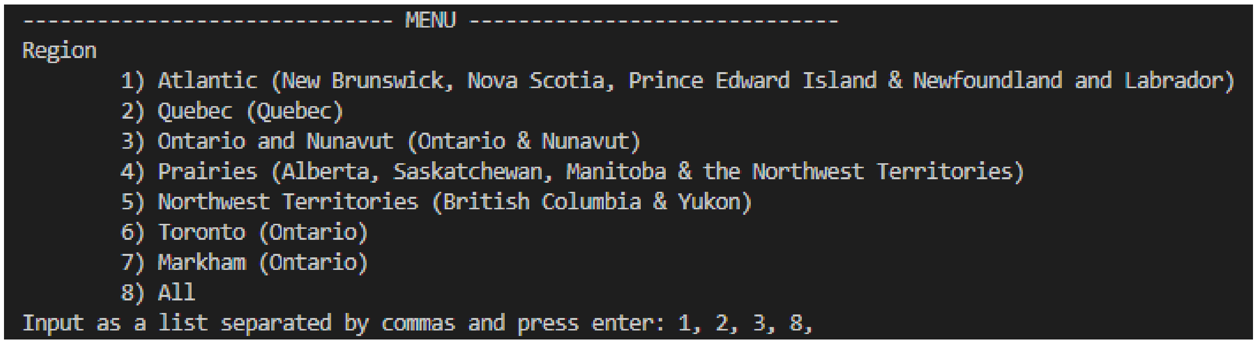
- The inputted value can only be within the range of the options shown

Some of our other menus have additional error checking logic such as hitting enter would just default to "ALL".



```
------------------------------ MENU ------------------------------
Choose a Region
1) Atlantic (New Brunswick, Nova Scotia, Prince Edward Island & Newfoundland and Labrador)
2) Quebec (Quebec)
3) Ontario and Nunavut (Ontario & Nunavut)
4) Prairies (Alberta, Saskatchewan, Manitoba & the Northwest Territories)
5) Northwest Territories (British Columbia & Yukon)
6) Toronto (Ontario)
7) Markham (Ontario)
8) Go Back
----------------------------------------------
Enter your choice [1-8]:
There were 1492 total deaths from COVID-19.
```

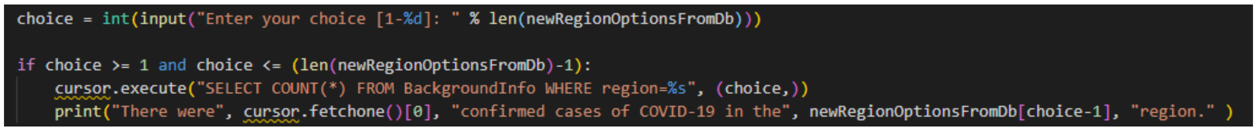
If we are filtering for more than 1 criteria, we do some preprocessing on the inputted values. For example, if the user accidentally types an extra comma, we remove it. Also if the user specifies certain options but also selects ALL, the ALL will override the specific option chosen.



```
 ------------------------------ MENU ------------------------------
 Region
        1) Atlantic (New Brunswick, Nova Scotia, Prince Edward Island & Newfoundland and Labrador)
        2) Quebec (Quebec)
        3) Ontario and Nunavut (Ontario & Nunavut)
        4) Prairies (Alberta, Saskatchewan, Manitoba & the Northwest Territories)
        5) Northwest Territories (British Columbia & Yukon)
        6) Toronto (Ontario)
        7) Markham (Ontario)
        8) All
 Input as a list separated by commas and press enter: 1, 2, 3, 8,
```

<u>SQL Injection</u>

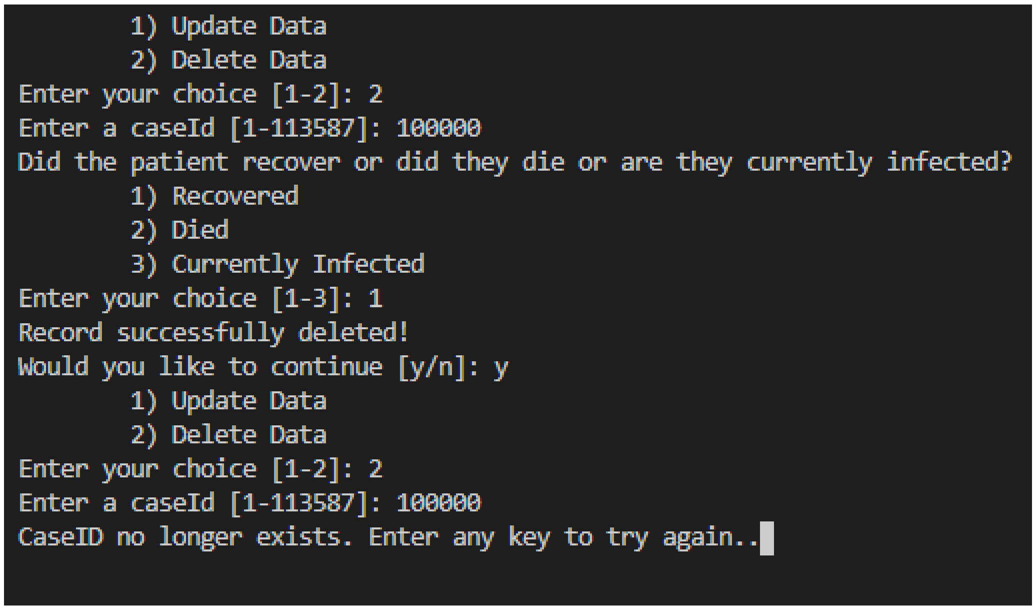
Although we are the only ones using the CLI application, we wanted to follow coding standards which include preventing SQL Injection when we are creating queries. Instead of appending variables onto our SQL query which is prone to SQL Injection we used parameterized queries whenever possible. Below is an example of how we used parameterized queries.



```python
choice = int(input("Enter your choice [1-%d]: " % len(newRegionOptionsFromDb)))

if choice >= 1 and choice <= (len(newRegionOptionsFromDb)-1):
    cursor.execute("SELECT COUNT(*) FROM BackgroundInfo WHERE region=%s", (choice,))
    print("There were", cursor.fetchone()[0], "confirmed cases of COVID-19 in the", newRegionOptionsFromDb[choice-1], "region." )
```

Modifying Data

There are cases where a deleted record is accessed again. This is the nature of a menu based CLI. However, we prevent users from updating a record that has already been deleted or deleting a record that has already been deleted. This is shown in the below code snippet:



```
        1) Update Data
        2) Delete Data
Enter your choice [1-2]: 2
Enter a caseId [1-113587]: 100000
Did the patient recover or did they die or are they currently infected?
        1) Recovered
        2) Died
        3) Currently Infected
Enter your choice [1-3]: 1
Record successfully deleted!
Would you like to continue [y/n]: y
        1) Update Data
        2) Delete Data
Enter your choice [1-2]: 2
Enter a caseId [1-113587]: 100000
CaseID no longer exists. Enter any key to try again..
```

Future Improvements

In order to improve the quality of our project, specifically the user-experience and the data-mining investigation, we've outlined some improvements that should be considered for the future.

1) Include more recent COVID-19 data. Since the original data gathered from StatsCan had a limited range that only went from September 2020 - October 2020, it was not possible to include COVID-19 data past October 2020. However, for the future it would be useful for users to view more recent data.
2) Allow users the ability to add and update COVID-19 testing data in order to further enhance the user-experience. While our project already allows users to report a new case of COVID-19, add a new location and update/delete data, it would be beneficial to add this feature due to the high number of COVID-19 tests being done every week. For instance, if a user received a COVID-19 test in Ontario, they should be able to add this information into the COVID-19 testing data set.
3) Include COVID-19 vaccination data. Considering how the number of vaccinations are rapidly increasing every day, this would give users easy-access to information regarding the vaccinations in Canada, including specific information about each type of vaccine (Pfizer, Moderna, Astrazeneca).

4) Include data surrounding the specific symptoms experienced by people who have died or recovered from COVID-19. Due to the limitations on StatsCan (which only had data on whether an individual experienced symptoms or not), our COVID-19 data did not include well-defined symptoms. However, for the future it would be useful to gather data on the exact symptoms experienced so that users have more insight on each COVID-19 case.
5) Add more attributes to the Background Info table. Currently, our table contains information about a person's location, gender, age group and occupation. In the future, it would be beneficial to include additional attributes such as their race, weight and height. This could expand the effectiveness of the data mining portion of our project, as it could result in a bigger data set which would increase the certainty of the decision tree.
6) Currently users can only add new locations. However, in the future we would like a government level account to be able to delete locations added in case users accidentally add new locations.