

Preprocessing the data with following files in PowerBI:

Step1 :

Preprocessing on **Confirmed positive cases of COVID19 in Ontario**

1. Total Rows: 1720242 before transformation.
2. Loaded data in powerquery and start analysis of columns.
3. Outcome1 column blank values replaced with NON FATAL.
4. Age_Group column UNKNOWN row values removed, because there is no use of it. Total 561 entries removed.
5. Accurate_Episode_Date column, 11 row values removed because some dates have irrelevant years such as 1964, 1997, 2002 etc..
6. Test_Reported_Date column, 8 entries removed because some dates have irrelevant years such as 1964, 1987, 2011 etc..
7. Specimen_Date column 2 entries removed because it's irrelevant such as 2001, 2012 etc...
8. Created new index column and removed Row_ID column.
9. Total 592 rows removed.
10. Total Rows: 1719660 after transformation.

Step2 :

Preprocessing on **reporting date PHU seven day average testing metrics**

1. Total Rows: 51905 before transformation.
2. PHU_ID 35 removed because it's irrelevant with the data at the moment. Total 1483 rows removed.
3. Total Rows: 50422 after transformation

Step3 :

For bringing city column per phu_ID into **2021_HUID** table/csv file, Used Python, Pandas for quick transformation:

1. Loaded df1 : Confirmed positive cases of COVID19 in Ontario.csv
2. Loaded df2 : 2021_HUID.xls excel_sheet
3. Took unique_elements of df1:phu_id in s1
4. In dictionary dict1, received 34 pairs of phu_id:phu_city
5. Converted dict1 into df3
6. Joined/Merged with inner join 2 data frames df3 and df2 with inner join and exported joined table as csv to powrbi: as new : 2021_HUID with city column

Built the Report with metrics and measures as needed :

Step1 :

Building report with following pages as mentioned in Technical Assessment Word File :

Cases Stratified by fields:

Fulfill the requirement to analyze overall trends and reported cases stratified by age group, sex and public health unit as well as Total Covid cases.

Hospitalization Insight:

Analyze the insight into hospitalization status by vaccine status by fully vaccinated, partial vaccinated and unvaccinated people. Also, create a Donut chart which shows, icu and non icu category wise fully vaccinated, partial vaccinated and unvaccinated people.

Cases by Geography:

Analyze cases by Geographic location of PHU.

Ontario +ve Cases:

Analyze percentage of covid positive cases in province of Ontario by Year and month which shows percentage of covid positive cases, number of testing volume and number of positive cases PHU wise.

City wise +ve Cases:

Analyze percentage of covid positive cases in selected city by Year and month which shows percentage of covid positive cases, number of testing volume and number of positive cases.

Covid T2esting Vol by PHU pg1 & pg2:

Analyze covid testing volume by year and PHU wise

Percent of Covid+ve Population PHU wise pg1 & pg2:

Analyze the percentage of covid positive PHU wise and total covid positive cases out of population

Marg vs (Covid +ve):

Analyze how marginalization effect on covid cases PHU wise and effect on covid testing volume.

Extraas Pages Created as shown following:

1. Analyze marginalization in PHU location and created bins to divide in category 1,2,3,4,5.
2. Created a Graph which shows covid positive cases by actual & 7d average.

3. Compare all the dates for analysis purpose. Which date is reported first and last and reported date to PHU.

Step2 :

Created following Important Measures and Columns :

Tables names:

- 1) 1 Confirmed positive cases of COVID19 in Ontario
- 2) 2 Hospitalizations by vaccination status
- 3) 3 reporting date PHU seven day average testing metrics
- 4) 4 2021_HUID_after_python_pd_transformation
- 5) 5 2021_HUID_after_powerbi_bins_transformation

Important Measures :

- 1) Covid Positive Cases : for the count of cases

```
Covid Positive Cases = COUNT('1 Confirmed positive cases of COVID19 in  
Ontario'[Row_Index])
```

- 2) Created 3 measures from table 2 , to get Cumulative fully vac, part vac, unvac count
- 3) Created 3 measures : test_vol, covid_pos_cases, and percent of positive cases :
Number of Covid Positive Cases 7d avg,
Number of Testing Volume 7d avg,
Percent of Covid Positive Cases 7d avg

```
Percent of Covid Positive Cases 7d avg =  
100 * DIVIDE([Number of Covid Positive Cases 7d avg],[Number of Testing  
Volume 7d avg])
```

- 4) Created 3 measure for to create bins, to categorize marginalization from 1 to 5 scale rather than float_numbers, which measures not super critical to show here
- 5) PHU_populatio_measure : to get count of population corresponding to PHU_id

```
PHU_populatio_measure =  
SELECTEDVALUE('4 2021_HUID_after_python_pd_transformation'[pop2021])
```

- 6) Percent of Covid Positive Population : to get percentage of covid positive population
Percent of Covid Positive Population =

```
100 * DIVIDE( COUNT('1 Confirmed positive cases of COVID19 in  
Ontario'[Row_Index]), [PHU_populatio_measure])
```
- 7) Ontario_Cumulative_Per Capita Covid Cases : to get per capita Covid positive number dynamically as per selected criteria of PHU, date

Important Calculated Columns :

- 1) Age_Group1 : to convert elements into new categories :

```
Age_Group1 =  
if('1 Confirmed positive cases of COVID19 in Ontario'[Age_Group] = "<20" ,  
"Young (<20)", IF('1 Confirmed positive cases of COVID19 in  
Ontario'[Age_Group] = "20s", "Young Aged Adult (20-39)", IF('1 Confirmed  
positive cases of COVID19 in Ontario'[Age_Group] = "30s", "Young Aged Adult  
(20-39)", if('1 Confirmed positive cases of COVID19 in Ontario'[Age_Group] =  
"40s", "Middle Aged Adult (40-59)", if('1 Confirmed positive cases of  
COVID19 in Ontario'[Age_Group] = "50s", "Middle Aged Adult (40-59)", "Old  
(60+)")))))
```

- 2) Positive cases_7d_avg : to get positive_cases from multiplication of 2 columns,
percent_pos_7d_avg & test_vol_7d_avg

```
Positive cases_7d_avg =  
INT(ROUND('3 reporting date PHU seven day average testing  
metrics'[test_volumes_7d_avg] * '3 reporting date PHU seven day average  
testing metrics'[percent_positive_7d_avg],0))
```
