Project Phase 3: OLAP Queries and BI Dashboard

Group：18
Student Name:
Chuhao Jia
Xiaohan Yu
Yiwen Liu

## Part 1. Standard OLAP operations

a. Drill down and roll up.

i. Total Number of resolved cases in your data mart; drill down to each week (week of year 2020), and drill down to a specific day.

SELECT COUNT(*), D.week_in_year, D.day, SUM(C.resolved)
FROM covid19_tracking_fact_table C, reported_date D
WHERE C.reported_date_surrogate = D.reported_date_surrogate_key
GROUP BY (D.week_in_year, D.day)
ORDER BY D.week_in_year, D.day
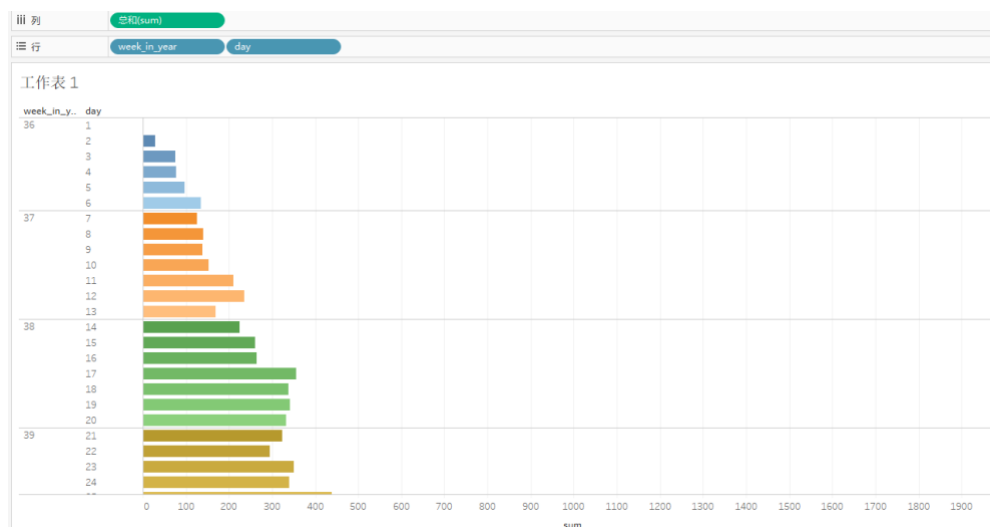
ii. All the unresolved cases in Toronto City, roll up to ON province in the data mart.

SELECT COUNT(*), COALESCE(P.province, 'Total') AS province, COALESCE(P.city, 'Total') AS city, SUM(C.unresolved)
FROM covid19_tracking_fact_table C, phu_location P
WHERE C.phu_surrogate = P.phu_surrogate_key
GROUP BY ROLLUP(P.province, P.city)
ORDER BY P.province, P.city
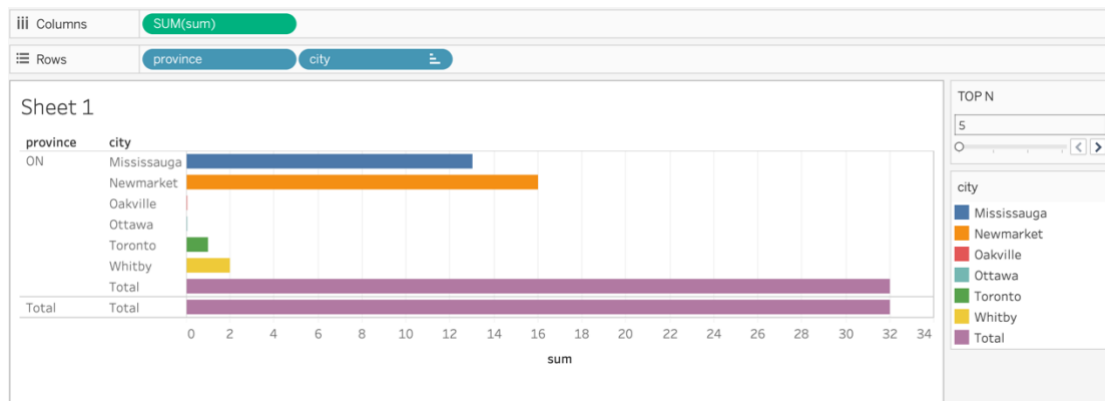
b. Slice, where only one dimension is selected.

i. The number of cases in a specific PHU (York Region Public Health Services), resolved, unresolved and fatal respectively

SELECT COUNT(*), P.phu_name, C.resolved, C.unresolved, C.fatal

FROM covid19_tracking_fact_table C, phu_location P

WHERE C.phu_surrogate = P.phu_surrogate_key AND P.phu_name = 'York Region Public Health Services'

GROUP BY (P.phu_name, C.resolved, C.unresolved, C.fatal)

ORDER BY P.phu_name
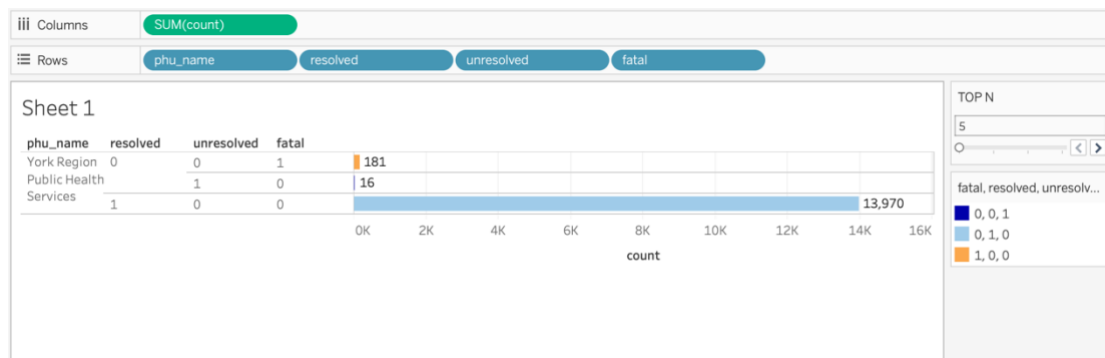
ii. The number of cases across the PHUs when a specific special measure (Grey Zone) was in place

SELECT COUNT(*), P.phu_name, SUM(C.resolved + C.unresolved + C.fatal), S.title

FROM covid19_tracking_fact_table C, phu_location P, special_measures S

WHERE C.phu_surrogate = P.phu_surrogate_key AND

C.special_measures_surrogate = S.surrogate_key AND S.title = 'Grey Zone'

GROUP BY (P.phu_name, S.title)

ORDER BY P.phu_name
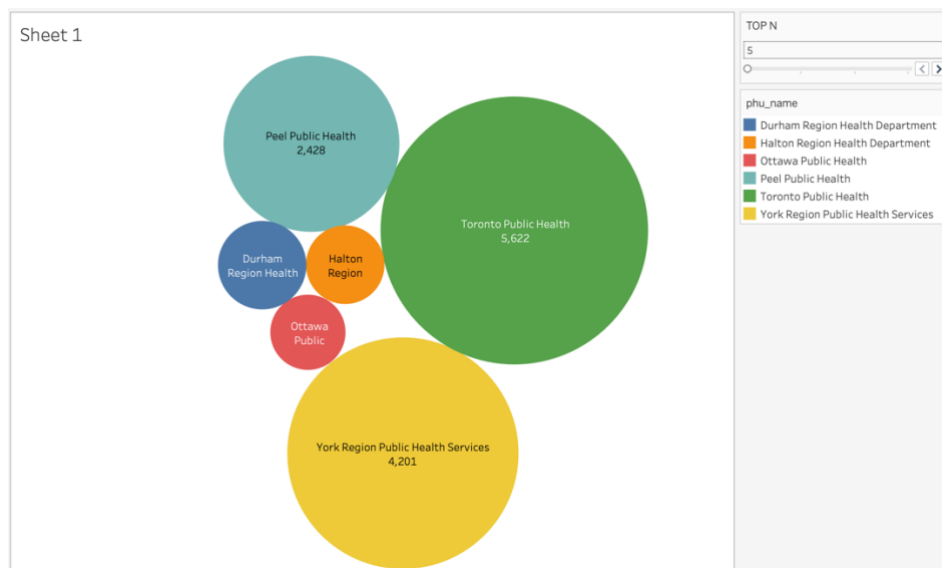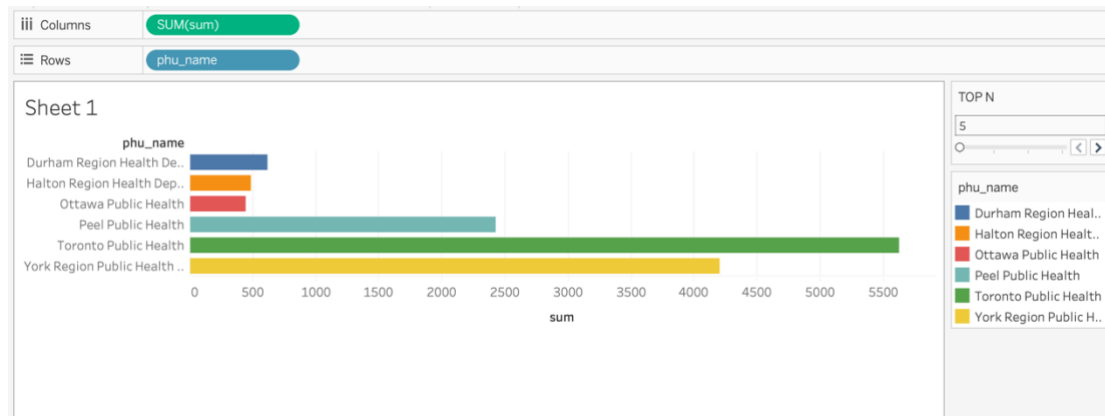
c. Dice, where one creates a sub-cube. – 2 queries

i. The number of fatal cases during a period of two months, e.g., Sep. and Oct., in Mississauga and Ottawa.

```
SELECT COUNT(*), D.month, P.city, SUM(C.fatal)
FROM covid19_tracking_fact_table C, reported_date D, phu_location P
WHERE C.reported_date_surrogate = D.reported_date_surrogate_key AND
C.phu_surrogate = P.phu_surrogate_key AND
D.month in ('September','October') AND
P.city in ('Mississauga','Ottawa')
GROUP BY (D.month, P.city)
```

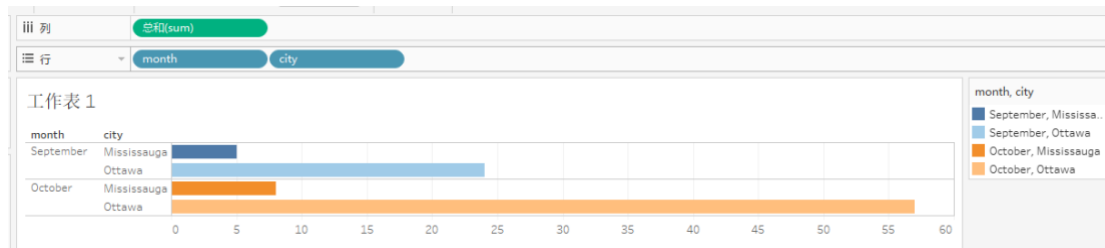## ii. The number cases across the PHUs when different special measures were in place

SELECT COUNT(*), P.phu_name, SUM(C.resolved + C.unresolved + C.fatal), S.title

FROM covid19_tracking_fact_table C, phu_location P, special_measures S

WHERE C.phu_surrogate = P.phu_surrogate_key AND

C.special_measures_surrogate = S.surrogate_key

GROUP BY (P.phu_name, S.title)

ORDER BY P.phu_name
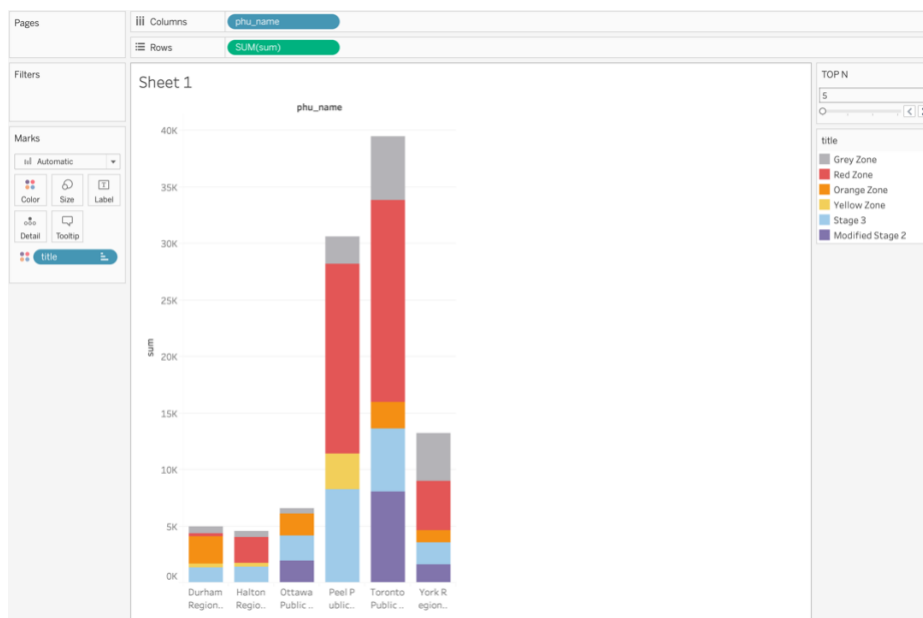
d. Combining OLAP operations.

i. The number of cases across the PHUs when a specific special measure (Grey Zone) was in place, roll up to city, and roll up to province, and roll up to all data in the data mart. (Roll up & Slice)

SELECT COUNT(*), SUM(C.resolved + C.unresolved + C.fatal), S.title, COALESCE(P.province, 'Total') AS province, COALESCE(P.city, 'Total') AS city, COALESCE(P.phu_name, 'Total') AS phu_name

FROM covid19_tracking_fact_table C, phu_location P, special_measures S

WHERE C.phu_surrogate = P.phu_surrogate_key AND

C.special_measures_surrogate = S.surrogate_key AND S.title = 'Grey Zone'

GROUP BY S.title, ROLLUP (P.province, P.city, P.phu_name)

ORDER BY P.city
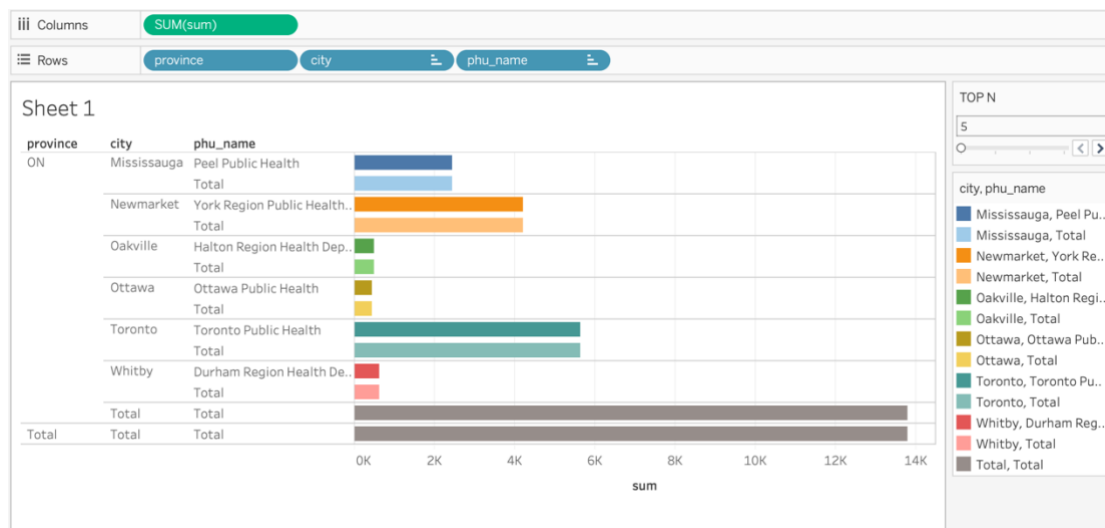
ii. The total number of resolved cases per month and day, when there is a special measure in that place, comparing the rain and snow precip in Toronto and Ottawa. (Drilling down & Dice)

SELECT COUNT(*), D.month, D.day, S.Title, P.city, W.total_rain_mm, W.total_snow_mm, SUM(C.resolved)

FROM covid19_tracking_fact_table C, reported_date D, phu_location P, weather W, special_measures S

WHERE C.phu_surrogate = P.phu_surrogate_key AND

C.reported_date_surrogate = D.reported_date_surrogate_key AND

C.weather_surrogate = W.surrogate_key AND

C.special_measures_surrogate = S.surrogate_key AND

P.city in ('Toronto','Ottawa')

GROUP BY (D.month, D.day, S.Title, P.city, W.total_rain_mm, W.total_snow_mm)

ORDER BY D.month, D.day, P.city
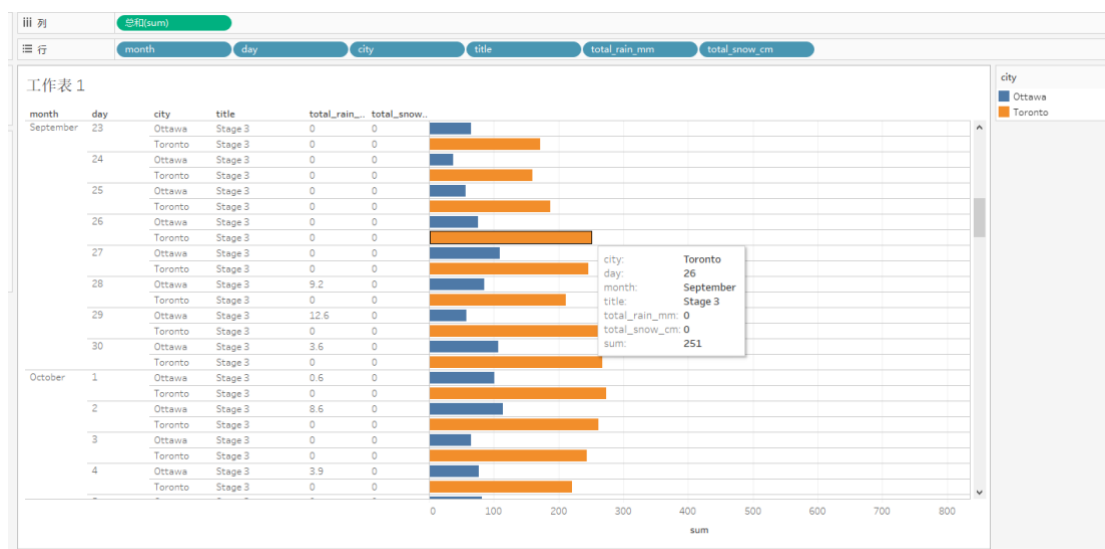
iii. The total number of cases when a specific special measure (Grey Zone) was in place, drill down based on location (data mart -> province -> city -> PHU). (Drilling down & Slice)

SELECT COUNT(*), SUM(C.resolved + C.unresolved + C.fatal), S.title, P.province, P.city, P.phu_name
FROM covid19_tracking_fact_table C, phu_location P, special_measures S
WHERE C.phu_surrogate = P.phu_surrogate_key AND
C.special_measures_surrogate = S.surrogate_key AND S.title = 'Grey Zone'
GROUP BY (S.title, P.province, P.city, P.phu_name)
ORDER BY P.city

Part 2. Explorative operation

a. Iceberg queries. Find the five days with the highest numbers of resolved outcomes

SELECT D.Date, SUM(C.resolved)
FROM covid19_tracking_fact_table C, reported_date D
WHERE C.reported_date_surrogate = D.reported_date_surrogate_key
GROUP BY(D.Date)
ORDER BY SUM DESC
LIMIT 5
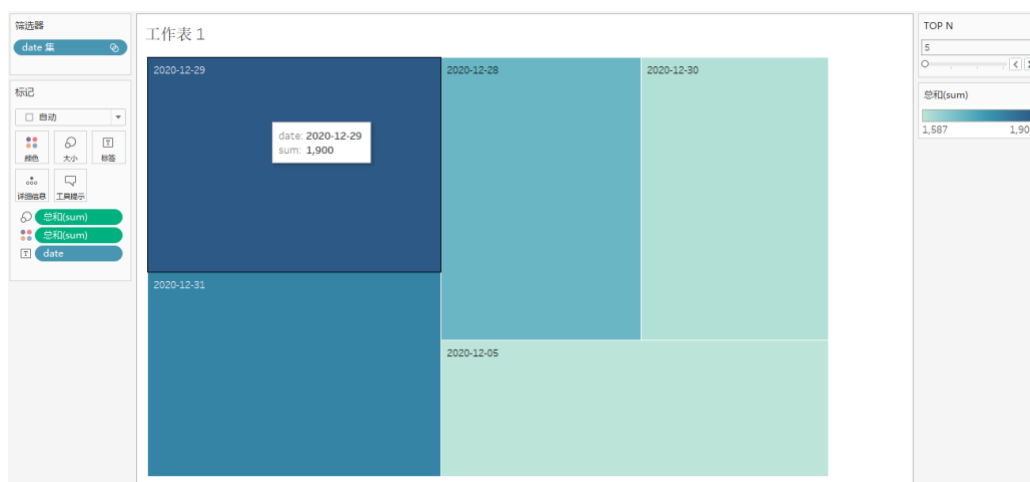
b. Windowing queries. Show the ranking of PHUs in terms of the number of cases per month

SELECT T.month, P.phu_name, P.city, SUM(C.resolved + C.unresolved + C.fatal),
RANK() OVER (PARTITION BY (T.month) ORDER BY SUM(C.resolved + C.unresolved + C.fatal) DESC)
FROM test_date T, phu_location P, covid19_tracking_fact_table C
WHERE   C.phu_surrogate   =   P.phu_surrogate_key   AND   C.test_date_surrogate   =
T.test_date_surrogate_key
GROUP BY (T.month, P.phu_name, P.city)

month	phu_name	city	sum	rank
December	Toronto Public Health	Toronto	17,200	1
December	Peel Public Health	Mississauga	13,911	2
December	York Region Public H...	Newmarket	6,683	3
December	Durham Region Healt...	Whitby	2,851	4
December	Halton Region Health...	Oakville	2,252	5
December	Ottawa Public Health	Ottawa	1,582	6
November	Toronto Public Health	Toronto	13,395	1
November	Peel Public Health	Mississauga	11,875	2
November	York Region Public H...	Newmarket	4,148	3
November	Durham Region Healt...	Whitby	1,541	4
November	Halton Region Health...	Oakville	1,495	5
November	Ottawa Public Health	Ottawa	1,381	6
October	Toronto Public Health	Toronto	8,880	1
October	Peel Public Health	Mississauga	5,002	2
October	Ottawa Public Health	Ottawa	2,681	3
October	York Region Public H...	Newmarket	2,557	4
October	Halton Region Health...	Oakville	891	5
October	Durham Region Healt...	Whitby	846	6
September	Toronto Public Health	Toronto	3,446	1
September	Peel Public Health	Mississauga	1,957	2

c. Window clause. Compare the number of resolved cases in Ottawa in four months.

SELECT COUNT(*), D.month, D.date, P.city, AVG(COUNT(*)) OVER W AS movavg

FROM covid19_tracking_fact_table C, reported_date D, phu_location P

WHERE C.reported_date_surrogate = D.reported_date_surrogate_key AND C.phu_surrogate = P.phu_surrogate_key AND P.city = 'Ottawa' AND C.resolved = 1

GROUP BY (D.month, D.date, P.city, C.resolved)

WINDOW W AS (PARTITION BY P.city

       ORDER BY D.date

       RANGE BETWEEN INTERVAL'1 MONTH' PRECEDING

       AND INTERVAL '1 MONTH' FOLLOWING)

count	month	date	city	movavg
1	September	2020/9/2	Ottawa	50.8710
10	September	2020/9/3	Ottawa	51.3125
15	September	2020/9/4	Ottawa	52.0909
12	September	2020/9/5	Ottawa	52.9706
16	September	2020/9/6	Ottawa	55.4000
25	September	2020/9/7	Ottawa	58.6111
22	September	2020/9/8	Ottawa	59.2973
13	September	2020/9/9	Ottawa	60.1579
26	September	2020/9/10	Ottawa	60.3077
59	September	2020/9/11	Ottawa	62.2750
46	September	2020/9/12	Ottawa	62.2439
30	September	2020/9/13	Ottawa	62.5000
57	September	2020/9/14	Ottawa	62.9767
39	September	2020/9/15	Ottawa	63.9318
40	September	2020/9/16	Ottawa	64.2444
55	September	2020/9/17	Ottawa	64.1739
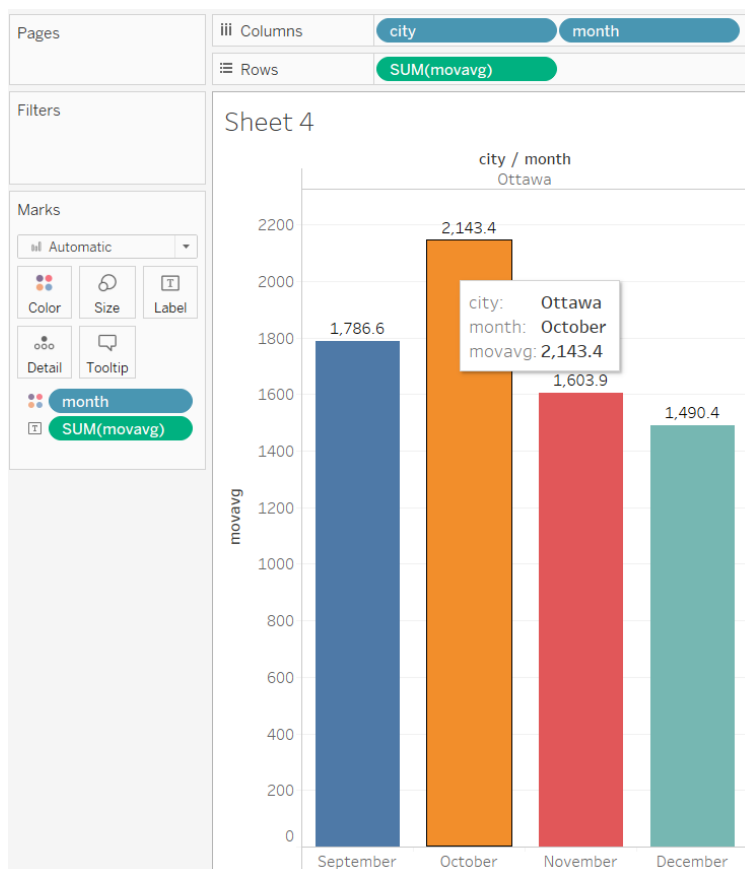
PART 3. BI dashboard and Information Visualization

For the graphs and charts related to Part 1 and Part 2, we put them above, so it is easier to match the queriers with the pictures.
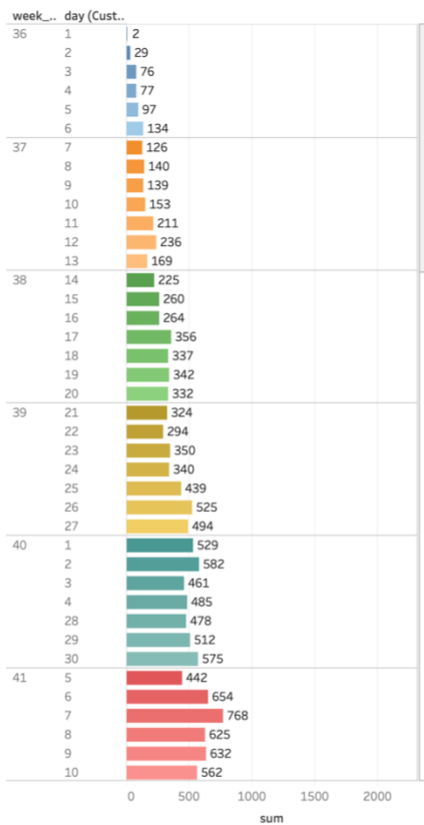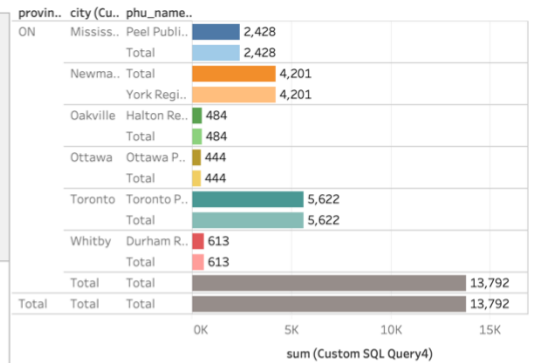
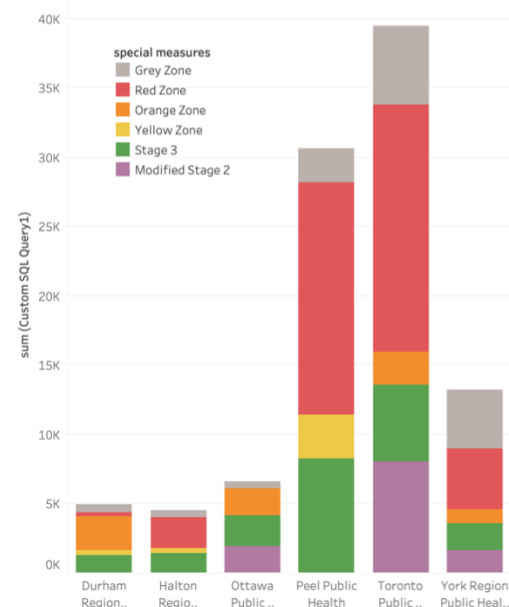Five days with the highest numbers of resolved outcomes

| December 29, 2020 1,900 | December 28, 2020 1,698 | December 30, 2020 1,600 |
| December 31, 2020 1,807 | December 5, 2020 1,587 | |

The number cases across the PHUs when different special measures were in place

phu_name (Custom SQL Query1)

special measures
- Grey Zone
- Red Zone
- Orange Zone
- Yellow Zone
- Stage 3
- Modified Stage 2
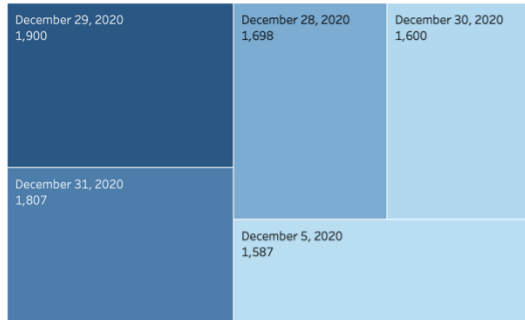
Total Number of resolved cases in the data mart; drill down to each week (week of year 2020), and drill down to a specific day.

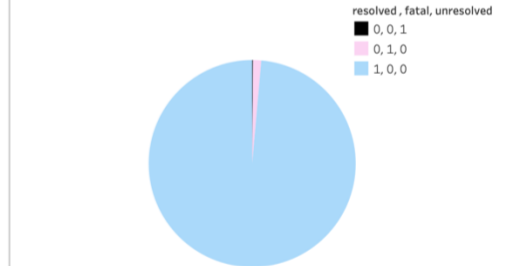| week_..	day (Cust..	sum
36	1	2
	2	29
	3	76
	4	77
	5	97
	6	134
37	7	126
	8	140
	9	139
	10	153
	11	211
	12	236
	13	169
38	14	225
	15	260
	16	264
	17	356
	18	337
	19	342
	20	332
39	21	324
	22	294
	23	350
	24	340
	25	439
	26	525
	27	494
40	1	529
	2	582
	3	461
	4	485
	28	478
	29	512
	30	575
41	5	442
	6	654
	7	768
	8	625
	9	632
	10	562

The number of cases across the PHUs when Grey Zone was in place, roll up to city, and roll up to province, and roll up to all data in the data mart
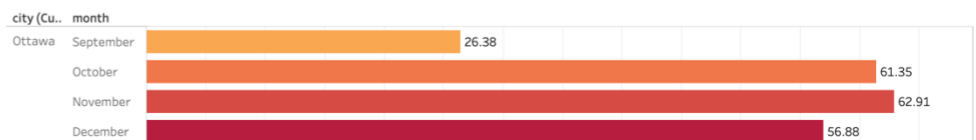
provin..	city (Cu..	phu_name..	sum (Custom SQL Query4)
ON	Mississ..	Peel Publi..	2,428
		Total	2,428
	Newma..	Total	4,201
		York Regi..	4,201
	Oakville	Halton Re..	484
		Total	484
	Ottawa	Ottawa P..	444
		Total	444
	Toronto	Toronto P..	5,622
		Total	5,622
	Whitby	Durham R..	613
		Total	613
	Total	Total	13,792
Total	Total	Total	13,792

The number of cases in a specific PHU (York Region Public Health Services), resolved, unresolved and fatal respectively

resolved , fatal, unresolved
- 0, 0, 1
- 0, 1, 0
- 1, 0, 0

Compare the number of resolved cases in Ottawa in four months.

city (Cu..	month	
Ottawa	September	26.38
	October	61.35
	November	62.91
	December	56.88

Task Divide & Meeting & Communication

As this project divided, each of us works on one part. After all three parts finished, we cross check the queries and graphs done by other team members.

We had two zoom meetings in total. In the first meeting, we went over all the queries and discussed the issues when writing those queries. During the second meeting, we discussed which graphs and charts should include in the dashboard.

Besides zoom meetings, we discussed problems using WeChat group chat.