Individual Project Part 1

Report #2: Task 5 INSY 661

NOTE: All required SQL statements for Task 5 are saved in the plain text file: "Individual Project Part1_Report2_-Task5.sql"

There is also a .sql file called "Exported Tables from phpMyAdmin" that was included in my submission, it contains the table structure for tasks 1 to 5.

5) EXTERNAL DATA INTEGRATION AND RELATED QUERIES

CONTEXT AND PROCEDURE

In this section, external data from the province of Quebec on the number of COVID-19 cases per day in the province will be integrated into the La Ronde database. The COVID-19 data from Quebec is publicly available HERE. For the sake of demonstration, the COVID-19 data will be used to draw insights about how the number of cases per day in the province affected ticket sales at La Ronde. The new COVID-19 data will also be used to illustrate how La Ronde's management could adjust aspects of the park (e.g., facility capacity) based on the daily number of COVID-19 cases.

The COVID-19 data is real, but please note that since the La Ronde data is artificial, any trends that appear in the below analysis will not be representative of how people at La Ronde actually behaved during the pandemic. Additionally, it will be assumed that La Ronde remained open to customers during the pandemic. Below is a description of how the COVID-19 data was prepared and then integrated into the La Ronde database:

1. The COVID-19 data was downloaded as a csv file called QC_COVID_CASES, and each row

was assigned a record ID. See below:

4		_			_	
1	record_ID	record_date	in_lab	by_epi_link	active_cases	
2	R1	2020-12-08 0:00	148229	8008	15897	
3	R2	2020-12-07 0:00	146590	7991	15360	
4	R3	2020-12-06 0:00	144993	7963	15108	
5	R4	2020-12-05 0:00	143480	7940	14775	
6	R5	2020-12-04 0:00	141855	7925	14306	
7	R6	2020-12-03 0:00	140110	7904	13878	

Column descriptions:

- **in_lab** represents the number of individuals in the province who are having COVID-19 test results processed.
- **by_epi_link** represents the number of individuals who have had exposure to a confirmed positive case or the same exposure as a confirmed positive case.
- **active_cases** is the total number of individuals who are COVID-19 positive in the province.

The decision was made to only analyse COVID-19 cases records from 2020-12-08 to the beginning of when Quebec started recording COVID-19 case data. This was because the most recent ticket sale recorded in the La Ronde database is 2020-12-08 (see below screen shot), and the goal is to investigate how COVID-19 cases could affect ticket sales.

т Ориона				
←T→				purchase_date > 1
	<i> </i>	⊋	Delete	2020-12-08
		≩ с Сору	Delete	2020-12-08
	⊘ Edit	<u>a copy</u>	Delete	2020-12-08

Please see "Individual Project Part1 Report2 - Task5.sql" for SQL statements

2. A csv file called CASES_TICKETS was created to hold the data for the bridge entity between QC_COVID_CASES and Tickets (see ERD below for illustration). This csv file contains the following columns:

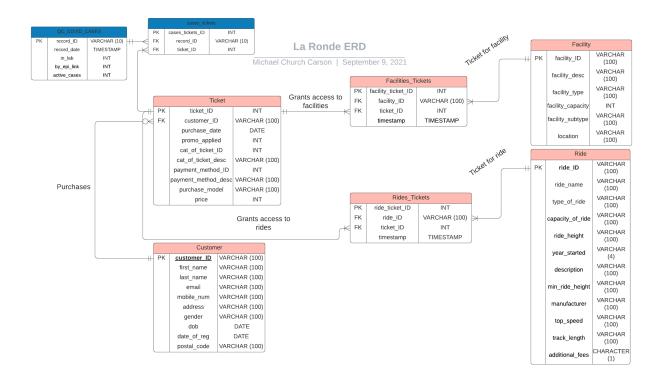
<u> </u>		
cases_ticket_ID	record_ID	ticket_ID
1	R1	189
2	R1	212
3	R1	580
4	R1	677
5	R1	946
6	D2	166

As the above screen shot demonstrates, this file is organized similarly to the FACILIITIES_TICKETS and RIDES_TICKETS files. There is an id column (cases_ticket_ID) and data from the QC_COVID_CASES and TICKET files. Note that each record_ID corresponds to a date (based on the record_date column value in the QC_COVID_CASES csv). For example, R1 corresponds to 2020-12-08. The ticket_IDs are the ID of the tickets that were purchased on that same date. Therefore, you can see in the above screenshot that there were five tickets (189, 212, 580, 677, and 946) that were purchased on 2020-12-08.

3. The exact same importing and populating procedure that was used in Report #1 for the La Ronde data was performed again with this new external data.

Please refer to "Individual Project Part1_Report2_-Task5.sql" to see the new DDL and SQL statements used to populate the new tables.

NEW ERD:



NEW RELATIONAL MODEL:

Customer (<u>customer_ID</u>, first_name, last_name, email, mobile_num, address, gender, dob, date_of_reg, postal_code)

Facility (facility_ID, facility_desc, facility_type, facility_capacity, facility_subtype, location)

Ride (<u>ride_ID</u>, ride_name, type_of_ride, capacity_of_ride, ride_height, year_started, description, min_ride_height, manufacturer, top_speed, track_length, additional_fees)

Ticket (<u>ticket_III</u> customer_ID) purchase_date, promo_applied, cat_of_ticket_ID, cat_of_ticket_desc, payment_method_ID, payment_method_desc, purchase_model, price)

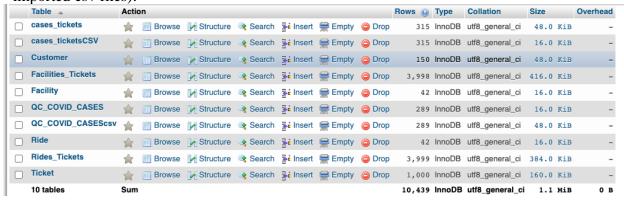
Rides_Tickets (ride_ticket_ID, ride_ID, ticket_ID, timestamp)

Facilities_Tickets (facility_ticket_ID_facility_ID_ticket_ID_timestamp)

QC_COVID_CASES (<u>record_ID</u>, record_date, in_lab, by_epi_link, active_cases)

cases_tickets (cases_tickets_ID_record_ID_ticket_ID)

The below screenshot shows the database with the four new tables (two relational and two from imported csv files):



The below screenshot shows all 8 tables in the new relational schema.



5 QUERIES

Query 1:

Objective: How many daily tickets were purchased when the daily active case count was over 10,000?

Code:

CREATE VIEW highactive cases AS

SELECT DISTINCT Ticket.ticket_ID as purchased_tickets, Ticket.purchase_date, Ticket.cat of ticket desc, QC COVID CASES.active cases

FROM Ticket, cases tickets, QC COVID CASES

WHERE Ticket.ticket ID = cases tickets.ticket ID AND

cases tickets.record ID = QC COVID CASES.record ID AND

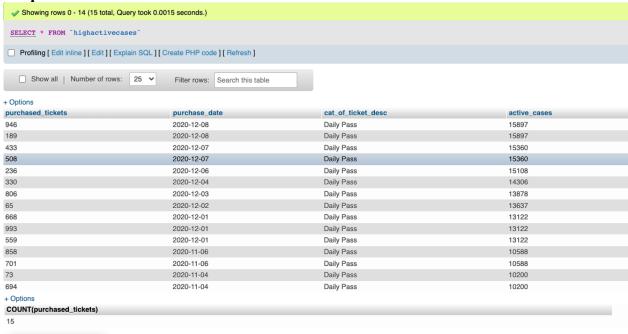
QC COVID CASES.active cases > 10000 AND

Ticket.cat of ticket desc = "Daily Pass"

ORDER BY QC COVID CASES.active cases DESC;

SELECT COUNT(purchased_tickets) FROM highactivecases

Output Screenshot:



From the above screenshots we can see that 15 daily tickets were purchased when the daily active case count was over 10,000.

Query 2:

Objective: Were customers still using the dinning facilities when the daily active case count was over 10,000?

Code:

SELECT DISTINCT Ticket.ticket_ID, Ticket.purchase_date, Ticket.cat_of_ticket_desc, QC_COVID_CASES.active_cases, facility_type, Facilities_Tickets.timestamp
FROM Ticket, cases_tickets, QC_COVID_CASES, Facilities_Tickets, Facility
WHERE Ticket.ticket_ID = cases_tickets.ticket_ID AND
cases_tickets.record_ID = QC_COVID_CASES.record_ID AND
Ticket.ticket_ID = Facilities_Tickets.ticket_ID AND
Facilities_Tickets.facility_ID = Facility.facility_ID AND
Facility.facility_type = "Dinning" AND
QC_COVID_CASES.active_cases > 10000;

Output Screenshot:

✓ Showing	g rows 0 - 24 (157 total, Query too	ok 0.0024 seconds.)			
Facilitie cases_tic	es_Tickets.timestamp FROM ckets.ticket_ID AND cases	Ticket.purchase_date, Ticket.c: Ticket, cases_tickets, QC_COVII _tickets.record_ID = QC_COVID_C: Facility.facility_ID AND Facility	D_CASES, Facilities_Ticke ASES.record_ID AND Ticket	ts, Facility WHERE Tic	cket.ticket_ID = es_Tickets.ticket_ID AND
Profiling	[Edit inline] [Edit] [Explain SQI	L] [Create PHP code] [Refresh]			
1 V	> >> Show all N	umber of rows: 25 V Filter rows	: Search this table		
atom In	and the same states	and the state of t			
cket_ID	purchase_date	cat_of_ticket_desc	active_cases	facility_type	timestamp
70	2020-12-04	Annual pass	14306	Dinning	2020-12-04 16:28:00
70 92	2020-12-04 2020-11-05	Annual pass Annual pass	14306 10363	Dinning Dinning	2020-12-04 16:28:00 2020-11-05 17:42:00
70 92 12	2020-12-04 2020-11-05 2020-12-08	Annual pass Annual pass Parking ticket	14306 10363 15897	Dinning Dinning Dinning	2020-12-04 16:28:00 2020-11-05 17:42:00 2020-12-08 01:03:00
70 92 12 3	2020-12-04 2020-11-05	Annual pass Annual pass Parking ticket Daily Pass	14306 10363 15897 10200	Dinning Dinning Dinning Dinning	2020-12-04 16:28:00 2020-11-05 17:42:00 2020-12-08 01:03:00 2020-11-04 09:58:00
70 92 112 3	2020-12-04 2020-11-05 2020-12-08 2020-11-04	Annual pass Annual pass Parking ticket	14306 10363 15897	Dinning Dinning Dinning	2020-12-04 16:28:00 2020-11-05 17:42:00 2020-12-08 01:03:00
70 92 12 3 22	2020-12-04 2020-11-05 2020-12-08 2020-11-04 2020-12-03	Annual pass Annual pass Parking ticket Daily Pass Annual pass	14306 10363 15897 10200 13878	Dinning Dinning Dinning Dinning Dinning	2020-12-04 16:28:00 2020-11-05 17:42:00 2020-12-08 01:03:00 2020-11-04 09:58:00 2020-12-03 23:21:00
70 92 12 3 3 22 08	2020-12-04 2020-11-05 2020-12-08 2020-11-04 2020-12-03 2020-12-07	Annual pass Annual pass Parking ticket Daily Pass Annual pass Daily Pass	14306 10363 15897 10200 13878 15360	Dinning Dinning Dinning Dinning Dinning Dinning Dinning	2020-12-04 16:28:00 2020-11-05 17:42:00 2020-12-08 01:03:00 2020-11-04 09:58:00 2020-12-03 23:21:00 2020-12-07 10:28:00
770 992 112 33 222 008 660 441	2020-12-04 2020-11-05 2020-12-08 2020-11-04 2020-12-03 2020-12-07 2020-12-02	Annual pass Annual pass Parking ticket Daily Pass Annual pass Daily Pass Annual pass	14306 10363 15897 10200 13878 15360 13637	Dinning Dinning Dinning Dinning Dinning Dinning Dinning Dinning Dinning	2020-12-04 16:28:00 2020-11-05 17:42:00 2020-12-08 01:03:00 2020-11-04 09:58:00 2020-12-03 23:21:00 2020-12-07 10:28:00 2020-12-02 02:00:00
cket_ID 70 992 112 33 222 08 60 41 06 46	2020-12-04 2020-11-05 2020-12-08 2020-11-04 2020-12-03 2020-12-07 2020-12-07 2020-12-02	Annual pass Annual pass Parking ticket Daily Pass Annual pass Daily Pass Annual pass Annual pass	14306 10363 15897 10200 13878 15360 13637 10363	Dinning	2020-12-04 16:28:00 2020-11-05 17:42:00 2020-12-08 01:03:00 2020-11-04 09:58:00 2020-12-03 23:21:00 2020-12-07 10:28:00 2020-12-02 02:00:00 2020-11-05 18:46:00
70 70 70 70 70 70 70 70 70 70	2020-12-04 2020-11-05 2020-12-08 2020-11-04 2020-12-03 2020-12-07 2020-12-02 2020-11-05 2020-12-03	Annual pass Annual pass Parking ticket Daily Pass Annual pass Daily Pass Annual pass Annual pass Annual pass Daily Pass	14306 10363 15897 10200 13878 15360 13637 10363	Dinning	2020-12-04 16:28:00 2020-11-05 17:42:00 2020-12-08 01:03:00 2020-11-04 09:58:00 2020-12-03 23:21:00 2020-12-07 10:28:00 2020-12-02 02:00:00 2020-11-05 18:46:00 2020-12-03 04:13:00

The above screenshot shows that even when the active case count was over 10,000, customers were still using the dinning facilities. In fact, tickets were scanned 157 times to enter dinning facilities when the active case count was over 10,000.

Ouerv 3:

Objective: There were high daily active case counts in the most recent two months of the imported COVID-19 data (i.e., 2020-10-08 to 2020-12-08). As this period of high active cases progressed, did more customers decide to drive to La Ronde (i.e., purchase parking tickets)? Compare the number of parking tickets purchased in the first month of this period of high active cases, to the number of parking tickets purchased in the second month of this period of high active cases. Assumption: customers are aware of the high number of active cases and considering this when deciding how to travel to La Ronde.

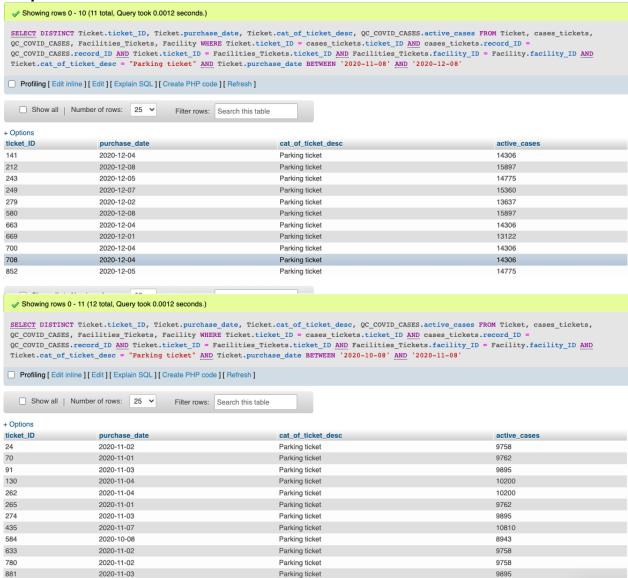
Code:

```
SELECT DISTINCT Ticket.ticket_ID, Ticket.purchase_date, Ticket.cat_of_ticket_desc, QC_COVID_CASES.active_cases
FROM Ticket, cases_tickets, QC_COVID_CASES, Facilities_Tickets, Facility
WHERE Ticket.ticket_ID = cases_tickets.ticket_ID AND
cases_tickets.record_ID = QC_COVID_CASES.record_ID AND
Ticket.ticket_ID = Facilities_Tickets.ticket_ID AND
Facilities_Tickets.facility_ID = Facility.facility_ID AND
Ticket.cat_of_ticket_desc = "Parking ticket" AND
Ticket.purchase_date BETWEEN '2020-11-08' AND '2020-12-08';
```

```
SELECT DISTINCT Ticket.ticket_ID, Ticket.purchase_date, Ticket.cat_of_ticket_desc, QC_COVID_CASES.active_cases
FROM Ticket, cases_tickets, QC_COVID_CASES, Facilities_Tickets, Facility
WHERE Ticket.ticket_ID = cases_tickets.ticket_ID AND
```

cases_tickets.record_ID = QC_COVID_CASES.record_ID AND Ticket.ticket_ID = Facilities_Tickets.ticket_ID AND Facilities_Tickets.facility_ID = Facility.facility_ID AND Ticket.cat_of_ticket_desc = "Parking ticket" AND Ticket.purchase date BETWEEN '2020-10-08' AND '2020-11-08'

Output Screenshot:



From the above outputs we can see that 12 parking tickets were purchased in the first month of this period of high active cases, and similarity, 11 were purchased in the second month. Therefore, over the course of this period of high active cases, customers do not seem to have decided to drive more often to La Ronde. (Remember: La Ronde data is artificial and therefore it is not actually going to reflect the preferences of people during this period of high active cases in Quebec.)

Query 4:

Objective: Quebec public health officials are investigating the impact that La Ronde had on the early stages of COVID-19 transmission in Quebec. To do so, they would like the names and contact information of all the customers who visited La Ronde during the early stages of the spread of COVID-19 in Quebec so that they can compare them to the names of the individuals who they know tested positive during this period.

Assumption: Let early stages of the spread of COVID-19 be when daily active cases were between 0 and 3000.

Code:

SELECT Customer.customer_ID, Ticket.purchase_date, Ticket.cat_of_ticket_desc, QC_COVID_CASES.active_cases, Customer.first_name, Customer.last_name, Customer.email, Customer.mobile_num

FROM Ticket, cases_tickets, QC_COVID_CASES, Customer

WHERE Customer.customer_ID = Ticket.customer_ID AND

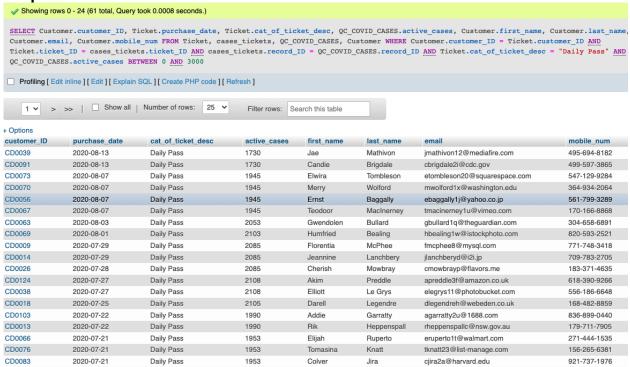
Ticket.ticket_ID = cases_tickets.ticket_ID AND

cases_tickets.record_ID = QC_COVID_CASES.record_ID AND

Ticket.cat_of_ticket_desc = "Daily Pass" AND

QC_COVID_CASES.active_cases BETWEEN 0 AND 3000;

Output Screenshot:



The above screenshot shows a partial output with the names and contact information of customers who visited La Ronde when daily active cases were between 0 and 3000.

Query 5:

Objective: Suppose that on days when the total active number of cases in Quebec is over 10,000, the government requires indoor capacity to be reduced by 20%. Show how many of La Ronde's facilities would have had to reduce their capacity by 20% and then what their capacity would have become after the 20% reduction.

Assumption: Facility capacity refers to the maximum number of people who can use the indoor part of a facility at one time.

Code:

CREATE VIEW capacity_rule AS

SELECT QC_COVID_CASES.record_date as highcases, QC_COVID_CASES.active_cases,
Facility.facility_capacity as cap, Facility.facility_desc

FROM QC_COVID_CASES, cases_tickets, Ticket, Facilities_Tickets, Facility

WHERE Ticket.ticket_ID = cases_tickets.ticket_ID AND

cases_tickets.record_ID = QC_COVID_CASES.record_ID AND

Ticket.ticket_ID = Facilities_Tickets.ticket_ID AND

Facilities_Tickets.facility_ID = Facility.facility_ID AND

QC_COVID_CASES.active_cases > 10000

ORDER BY QC_COVID_CASES.record_date DESC;

SELECT COUNT(DISTINCT facility_desc) FROM capacity rule;

SELECT DISTINCT facility_desc, cap*0.8 FROM capacity_rule

Output Screenshot:

Showing rows 0 - 24 (193 total, Que	ery took 0.0128 seconds.)					
#CREATE VIEW capacity_rule AS SELECT QC_COVID_CASES.record_date as highcases, QC_COVID_CASES.active_cases, Facility_facility_capacity as cap, Facility_facility_desc FROM QC_COVID_CASES, cases_tickets, Ticket, Facilities_Tickets, Facility_WHERE Ticket.ticket_ID = cases_tickets.ticket_ID AND cases_tickets.record_ID = QC_COVID_CASES.record_ID AND Ticket.ticket_ID = Facilities_Tickets.ticket_ID AND Facilities_Tickets.facility_ID = Facility_facility_ID AND QC_COVID_CASES.active_cases > 10000 ORDER BY QC_COVID_CASES.record_date DESC						
Edit inline] [Edit] [Create PHP code]						
1 > >> Show all	Number of rows: 25 V Filter rows: So	earch this table				
highcases	active_cases	cap	facility_desc			
2020-12-08 00:00:00	15897	20	Amir			
2020-12-08 00:00:00	15897	20	Popcorn & Cie			
2020-12-08 00:00:00	15897	20	Yo! Chine			
2020-12-08 00:00:00	15897	20	Subway			
2020-12-08 00:00:00	15897	20	Popcorn & Cie			
2020-12-08 00:00:00	15897	20	Marchand Du Village			
2020-12-08 00:00:00	15897	20	Pizza Ronde 1			
2020-12-08 00:00:00	15897	20	Rolopan 2			
2020-12-08 00:00:00	15897	20	Au Comptoir Frais			
2020-12-08 00:00:00	15897	20	Au Comptoir Frais			
2020-12-08 00:00:00	15897	20	Halte Gourmande - Au Bol			
2020-12-08 00:00:00	15897	20	Marchand Du Village			
2020-12-08 00:00:00	15897	20	La Centrale Burgers & Frites			
COUNT(DISTINCT facility_desc)						
37						

+ Options	
facility_desc	cap*0.8
Emporium	16.0
Marchand Du Village	16.0
Photo Goliath	16.0
Carrousel Du Bonbon	16.0
Château Du Bonbon	16.0
Boutique Du Far West	16.0
Boutique Spirale	16.0
Boutique Splash	16.0
Yol Chine	16.0
Fines Poutines Express	16.0
Fines Poutines Plus	16.0
La Centrale Burgers & Frites	16.0
Pizza Ronde 1	16.0
Pizza Ronde 2	16.0
Pizza Ronde 3	16.0
Poulet Etc.	16.0
Restaurant Lafleur	16.0
Subway	16.0
Dippin' Dots	16.0
Moozoo 1	16.0
Moozoo 2	16.0
Moozoo 3	16.0
Queues De Castor 1	16.0
Queues De Castor 2	16.0
Queues De Castor 3	16.0
- Specific	
facility_desc	cap*0.8
Rolopan 1	16.0
Rolopan 2	16.0
Au Comptoir Frais	16.0
Halte Gourmande - Au Bol	16.0
Halte Gourmande - Le Marché	16.0
Sabra	16.0
Amir	16.0
Mucho Nacho Churros	16.0
Park Opening Celebration	160.0
Saloon	16.0
Popcorn & Cie	16.0
Bar Rouge	16.0

The first screenshot shows a VIEW that was created of all the facilities that were visited by customers (ticket scanned) on days when the active case count in the province was over 10,000. The second screenshot counts the number of distinct facilities in the VIEW and demonstrates that 37 (or all) of La Ronde's facilities would have had to reduce their capacity by 20% at some point during the period captured by the COVID-19 data. The last two screenshots show that reducing capacity by 20% would mean that every facility would have had a capacity of 16, expect for the Park Opening Celebration facility, which would have had a capacity of 160.