# **Build and Optimize Data Warehouses with BigQuery: Challenge Lab**

## **GSP340**

## Copy paste DATASET NAME and TABLE NAME in the format of:

## <DATASET NAME>.<TABLE NAME> (it will be needed a lot in this lab)

## 

## **Task 1: Create a table partitioned by date**

Add Data → Explore Public Datasets → [oxford covid-19 government response tracker](https://console.cloud.google.com/bigquery?p=bigquery-public-data&d=covid19_govt_response&page=dataset)→ view dataset

Go to BigQuery → Click on 3 dots next to your project name → create dataset

Dataset name will be given to you

Leave other settings at default and click on create

Copy Paste in BigQuery Editor:

CREATE OR REPLACE TABLE <DATASET NAME>.<TABLE NAME>

PARTITION BY date

OPTIONS(

partition\_expiration\_days=360,

description="oxford\_policy\_tracker table in the COVID 19 Government Response public dataset with an expiry time set to 90 days."

) AS

SELECT

\*

FROM

`bigquery-public-data.covid19\_govt\_response.oxford\_policy\_tracker`

WHERE

alpha\_3\_code NOT IN ('GBR', 'BRA', 'CAN','USA')

## 

## **Task 2: Add new columns to your table**

ALTER TABLE <DATASET NAME>.<TABLE NAME>

ADD COLUMN population INT64,

ADD COLUMN country\_area FLOAT64,

ADD COLUMN mobility STRUCT<

avg\_retail FLOAT64,

avg\_grocery FLOAT64,

avg\_parks FLOAT64,

avg\_transit FLOAT64,

avg\_workplace FLOAT64,

avg\_residential FLOAT64

>

## **Task 3: Add country population data to the population column**

Query 1:

CREATE OR REPLACE TABLE <DATASET NAME>.pop\_data\_2019 AS

SELECT

country\_territory\_code,

pop\_data\_2019

FROM

`bigquery-public-data.covid19\_ecdc.covid\_19\_geographic\_distribution\_worldwide`

GROUP BY

country\_territory\_code,

pop\_data\_2019

ORDER BY

country\_territory\_code

Query 2:

UPDATE

`<DATASET NAME>.<TABLE NAME>` t0

SET

population = t1.pop\_data\_2019

FROM

`<DATASET NAME>.pop\_data\_2019` t1

WHERE

CONCAT(t0.alpha\_3\_code) = CONCAT(t1.country\_territory\_code);

**Task 4: Add country area data to the country\_area column**

UPDATE

`<DATASET NAME>.<TABLE NAME>` t0

SET

t0.country\_area = t1.country\_area

FROM

`bigquery-public-data.census\_bureau\_international.country\_names\_area` t1

WHERE

t0.country\_name = t1.country\_name

## **Task 5: Populate the mobility record data**

UPDATE

`<DATASET NAME>.<TABLE NAME>` t0

SET

t0.mobility.avg\_retail = t1.avg\_retail,

t0.mobility.avg\_grocery = t1.avg\_grocery,

t0.mobility.avg\_parks = t1.avg\_parks,

t0.mobility.avg\_transit = t1.avg\_transit,

t0.mobility.avg\_workplace = t1.avg\_workplace,

t0.mobility.avg\_residential = t1.avg\_residential

FROM

( SELECT country\_region, date,

AVG(retail\_and\_recreation\_percent\_change\_from\_baseline) as avg\_retail,

AVG(grocery\_and\_pharmacy\_percent\_change\_from\_baseline) as avg\_grocery,

AVG(parks\_percent\_change\_from\_baseline) as avg\_parks,

AVG(transit\_stations\_percent\_change\_from\_baseline) as avg\_transit,

AVG(workplaces\_percent\_change\_from\_baseline) as avg\_workplace,

AVG(residential\_percent\_change\_from\_baseline) as avg\_residential

FROM `bigquery-public-data.covid19\_google\_mobility.mobility\_report`

GROUP BY country\_region, date

) AS t1

WHERE

CONCAT(t0.country\_name, t0.date) = CONCAT(t1.country\_region, t1.date)

## 

## **Task 6: Query missing data in population & country\_area columns**

Query 1:

SELECT country\_name, population

FROM `<DATASET NAME>.<TABLE NAME>`

WHERE population is NULL

Query 2:

SELECT country\_name, country\_area

FROM `<DATASET NAME>.<TABLE NAME>`

WHERE country\_area IS NULL

Query 3:

SELECT DISTINCT country\_name

FROM `<DATASET NAME>.<TABLE NAME>`

WHERE population is NULL

UNION ALL

SELECT DISTINCT country\_name

FROM `<DATASET NAME>.<TABLE NAME>`

WHERE country\_area IS NULL

ORDER BY country\_name ASC

**CONGRATULATIONS YOU HAVE COMPLETED THE CHALLENGE LAB!!!**

**SAKURA SATIO**