

## Create A Research Dataset

### Introduction

The late 2019 COVID-19 epidemic has had a significant effect on cultures all across the world. From January 2020 to December 2022, a period of three years, people around the world struggled to understand the intricate interactions between many circumstances that caused the occurrence of COVID-19. The number of cases, severity, and reactions to the virus during this time period varied drastically by area, and the reasons for these differences are complex. This study dives into the economical, political, and other influencing elements that have been crucial in deciding whether particular regions have had higher or fewer rates of COVID-19. Thus, the topic chosen to create a research dataset is to examine socioeconomic, political or other factors that contributed to either higher or lower incidences of covid 19 between Jan 2020 and Dec 2022.

Recording Video:

[https://drive.google.com/file/d/1lqsFoGvdQgkfn6DYJnZsdPq8Y3B66Wep/view?usp=share\\_link](https://drive.google.com/file/d/1lqsFoGvdQgkfn6DYJnZsdPq8Y3B66Wep/view?usp=share_link)

### Dataset Selection

To create a research dataset the following datasets have been selected:

1. COVID19 vaccine rate in the US by State.
2. COVID19 cases surveillance by state
3. Facility-level data for hospital utilization aggregated on a weekly basis
4. Testing of COVID 19 and outcomes for US States
5. GDP in the US by state from 2020 to 2022
6. Population in the US by state from 2020 to 2022

Google Drive Link:

<https://docs.google.com/spreadsheets/d/1nB3aztK1eVvrapERsmrmpzyVG17tpOTKRibstPqtXE/edit#gid=0>

From all the above data sets, the columns of choice are appended into a final research dataset which could be further analyzed to fetch any kind of information or perform research to derive answers.

### The process of integration

The "state\_testing\_and\_outcome" dataset, "COVID-19 cases surveillance" dataset, and "COVID-19 vaccine rate" dataset all include daily data that were recorded. Additionally, weekly data is kept in the file "Facility-level data for hospital utilization".

However, yearly statistics were kept on "GDP in the US by state" and "Population in the US by state" As a result, the initial step is to as closely as possible match all data with the month dimension of time. The "state\_testing\_and\_outcome", "COVID-19 cases surveillance", and "COVID-19 vaccine rate" datasets are combined by month. The "Population in the US by state" and "GDP in the US by state" data sets were shown by month. As a result, all of the tables could be pooled.

The second stage involves making every state name column in every dataset the same. As a result, we choose to convert all state fields for full names to abbreviation states. For instance, all datasets would change "Missouri" to "MO". We created a mapping table called "StateName" in this stage. It would be used to shorten the complete name of the state to its acronym.

Lastly, according to the number of records in each dataset, we decided "COVID19 cases surveillance" as the left table, and other datasets would be retrieved information to this left one. Below is the process of combination to final table.

```

1 create table `ores-groupass2.covid_dataset.V_Final_table`
2 as
3
4 SELECT distinct (CONCAT(cast (T2.Year as string), '-', cast (T2.Month as string), '-', T2.state)) AS T_id,
5 T2.Year, T2.MONTH, T2.state, sum(T2.case_total_count) AS cases_monthly_amount,
6 T6.yr2021 AS yr2021_population, T6.yr2022 AS yr2022_population, --T1.total_beds_7_day_avg,
7 T7.yr2021_GDP, T7.yr2022_GDP, T3.monthly_negative, T3.monthly_positive, T3.monthly_total_test_results,
8 T5.hospital_name, T5.total_beds_7_day_avg, T5.total_patients_hospitalized_confirmed_influenza_and_covid_7_day_coverage,
9 T1.people_fully_vaccinated_per_hundred
10
11 FROM
12 | `ores-groupass2.covid_dataset.V_COVID_19_Case` AS T2
13 left join `ores-groupass2.covid_dataset.V_vaccination_2` AS T1
14 ON cast(T1.Year as string) = T2.Year
15 AND cast( T1.MONTH as string) = T2.MONTH
16 AND T1.Postal = T2.state
17 left join `ores-groupass2.covid_dataset.V_state_testing_and_outcomes` AS T3
18 ON T2.Year = cast (T3.year as string)
19 AND T2.MONTH = cast (T3.month as string)
20 AND T2.state = T3.state
21 /*
22 Response too large to return. Consider specifying a destination table in your job configuration. For more details
23 */
24 left join `ores-groupass2.covid_dataset.V_COVID_19_HospitalCapacity_by_Facility` AS T5
25 ON T2.Year = cast (T5.year as string)
26 AND T2.MONTH = cast (T5.month as string)
27 AND T2.state = T5.state
28 left join `ores-groupass2.covid_dataset.V_population` AS T6
29 ON T2.state = T6.postal
30 left join `ores-groupass2.covid_dataset.V_GDP` AS T7
31 ON T2.state = T7.Postal
32 --where T2.age_group != '65+ years' AND monthly_negative is not null
33 group by T2.Year, T2.MONTH, T2.state, T6.yr2021, T6.yr2022, --T1.total_beds_7_day_avg,
34 T7.yr2021_GDP, T7.yr2022_GDP, T3.monthly_negative, T3.monthly_positive, T3.monthly_total_test_results,
35 T5.hospital_name, T5.total_beds_7_day_avg, T5.total_patients_hospitalized_confirmed_influenza_and_covid_7_day_coverage,
36 T1.people_fully_vaccinated_per_hundred

```

There is the result of the final table. It shows 75674 number of rows.

## Group assignment2 Group3

The screenshot shows the Google Cloud BigQuery interface. On the left is the Explorer pane with a tree view of workspace resources. The main pane displays a query result for 'V\_Final\_table'. The query is a SELECT statement from 'ores-groupass2.covid\_dataset.V\_Final\_table'. The results are shown in a table with columns: Row, T\_id, Year, MONTH, state, cases\_monthly\_amount, yr2021\_population, yr2022\_population, yr2021\_GDP, and yr2022\_GDP. The table contains 9 rows of data for different states and years.

Row	T_id	Year	MONTH	state	cases_monthly_amount	yr2021_population	yr2022_population	yr2021_GDP	yr2022_GDP
1	2020-01-AZ	2020	01	AZ	13	7264877	7399197	432279.8	475653.7
2	2020-01-CA	2020	01	CA	159	39142991	39029342	3416939.4	3641643.4
3	2020-01-FL	2020	01	FL	26	21828069	22244823	1292391.3	1439065.0
4	2020-01-GA	2020	01	GA	184	10788029	10912876	701606.1	767377.6
5	2020-01-NC	2020	01	NC	23	10565885	10698973	659529.3	715968.3
6	2020-01-NJ	2020	01	NJ	56	9267961	9261699	692227.3	754948.2
7	2020-01-NV	2020	01	NV	19	3146402	3177772	200127.3	222938.6
8	2020-01-NY	2020	01	NY	141	19857492	19677151	1911345.8	2048402.6
9	2020-01-TN	2020	01	TN	72	6864951	7051330	458180.0	484647.5

There is the schema and detail of the final table.

The screenshot shows the Google Cloud BigQuery interface with the schema of 'V\_Final\_table' displayed. The schema table lists fields with their names, types, modes, keys, collations, default values, policy tags, and descriptions.

Field name	Type	Mode	Key	Collation	Default Value	Policy Tags	Description
T_id	STRING	NULLABLE					
Year	STRING	NULLABLE					
MONTH	STRING	NULLABLE					
state	STRING	NULLABLE					
cases_monthly_amount	INTEGER	NULLABLE					
yr2021_population	INTEGER	NULLABLE					
yr2022_population	INTEGER	NULLABLE					
yr2021_GDP	FLOAT	NULLABLE					
yr2022_GDP	FLOAT	NULLABLE					
monthly_negative	INTEGER	NULLABLE					
monthly_positive	INTEGER	NULLABLE					
monthly_total_test_results	INTEGER	NULLABLE					
hospital_name	STRING	NULLABLE					
total_beds_7_day_avg	INTEGER	NULLABLE					
total_patients_hospitalized_confirmed_influenza_and_covid_7_day_coverage	INTEGER	NULLABLE					
people_fully_vaccinated_per_hundred	FLOAT	NULLABLE					

## Summary

This final dataset provides information about covid19 case surveillance, testing and outcomes of the disease, hospital facilities, vaccination, and GDP of the states. This dataset provides an opportunity to researchers, public healthcare officials to obtain valuable insights that can be helpful in knowing a few important questions like highest number of testing encounters by state, medical and hospital facilities by state, total population and number of people vaccinated and their effects, economic status etc..

## References

BEA Interactive Data Application. (n.d.).

https://apps.bea.gov/iTable/?reqid=70&step=1&acrdn=2#eyJhcHBpZCI6NzAsInN0ZXBzIjpbMSwyOSwyNSwzMSwyNiwzNywzMF0sImRhGEiOIjpbIiRhYmxiSWQiLCI5Ii0sWyJTdGF0ZSI6WyIwIiI1dFsiQXJlYSIsWyJYWCI6dXsxbiInOYXRpc3RpYyIsWyIxiI1dFsiVW5pdF9vZi9tZWZdXJllIiwzTG92ZWxzIi0sWyJZZWYyIixbIjIwMjliLCIyMDIxiwiMjAyMCJdXSBzIiIiYXJlZCZdYmIiIi0sWyJZZWYyX0VuZCI6Ii0sIi1dfQ==

COVID-19 Case Surveillance Public Use Data with Geography | Data | Centers for Disease Control and Prevention. (2023, October 6). <https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data-with-Geography/n8mc-b4w4>

COVID-19 reported patient impact and hospital capacity by facility | HealthData.gov.  
(n.d.). <https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/anag-cw7u/explore>

Google Cloud console. (n.d.).

<https://console.cloud.google.com/marketplace/product/bigquery-public-datasets/covid19-public-data-program?project=keerthana-bq-slu&pli=1>

USA COVID-19 vaccinations. (2023, September 19). Kaggle.

<https://www.kaggle.com/datasets/paultimothymooney/usa-covid19-vaccinations/>

US Census Bureau. (2023, June 13). State Population Totals and Components of Change: 2020-2022. Census.gov. <https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-total.html>

## Appendix

- I. Process of adjusting dataset" COIVD19 cases surveillance" and insert to new table " V\_COVID\_19\_Case"

```
1  /* Based on original column[case_month] to get the get the[Year] and [MONTH], then create the new columns in new table[V_COVID_19_Case].
2  In addition, in the new table, it already excluded the 'null' and 'Missing' value in [age_group], [sex]
3  */
4  CREATE TABLE `ores-groupass2.covid_dataset.V_COVID_19_Case`
5  AS
6  SELECT
7    case_month as case_y_m,
8    substring (case_month,1,4) as Year,
9    substring (case_month,6,7) as Month,
10   res_state as state,
11   age_group,
12   sex,
13   COUNT(*) AS case_total_count
14 FROM `ores-groupass2.covid_dataset.COVID_19_Case_Surveillance_Public_Use_Data_with_Geography__1_`
15 where age_group != 'Missing' and age_group != 'NA' and sex != 'NA'
16 GROUP BY case_month, res_state, age_group, sex
```

- II. Process of adjusting dataset" state\_testing\_and\_outcomes" and insert to new table " V\_state\_testing\_and\_outcomes"

```
21 /*Based on original column[date] to get the get the[year] and [month], then create the new columns in new table[V_state_testing_and_outcomes].
22 In addition, in the new table, it had already aggregate [death],[total_test_results], [negative] ,[positive],[positive_tests_antibody] by month
23 */
24 CREATE TABLE `ores-groupass2.covid_dataset.V_state_testing_and_outcomes`
25 AS
26
27 SELECT
28   EXTRACT(YEAR FROM date) AS year,
29   EXTRACT(MONTH FROM date) AS month,
30   state,
31   SUM(death) AS mothly_total_death,
32   SUM(total_test_results) AS monthly_total_test_results,
33   SUM(negative) AS monthly_negative,
34   SUM(positive) AS monthly_positive,
35   SUM(positive_tests_antibody) AS monthly_positive_tests_antibody
36 FROM `ores-groupass2.covid_dataset.state_testing_and_outcomes`
37 GROUP BY year, month, state
38 ORDER BY year, month, state;
39
40
41 ---drop table `ores-groupass2.covid_dataset.V_state_testing_and_outcomes`
```

- III. Process of adjusting dataset vacaccination" and insert to new table " V\_vaccination"

```

48 /* Explain: part1: use CTE to get [year], [month], [day] column
49 Based on original column[date] to get the [year],[month],and [day], then create the new columns in new table[V_state_testing_and_outcomes]. In addition, in the new table, it had already aggregate
[death],[total_test_results],[negative],[positive],[positive_tests_antibody] by month
50 */
51 --part2: use part1 result to calculate the last day of each month and mark it as 'Last Day', because it can show the monthly data rather than daily data.
52 CREATE TABLE 'ores-groupass2.covid_dataset.V_vaccination'
53 AS
54 AS
55 WITH VaccinationData AS (
56 SELECT
57     V.date,
58     EXTRACT(YEAR FROM V.date) AS year,
59     EXTRACT(MONTH FROM V.date) AS month,
60     EXTRACT(DAY FROM V.date) AS day,
61     V.location,
62     V.people_fully_vaccinated_per_hundred,
63     V.total_vaccinations_per_hundred,
64     V.total_boosters_per_hundred,
65     ROW_NUMBER() OVER (PARTITION BY EXTRACT(YEAR FROM V.date), EXTRACT(MONTH FROM V.date) ORDER BY V.date DESC) AS row_num
66 FROM 'ores-groupass2.covid_dataset.vaccination' AS V
67 )
68 SELECT
69     date,
70     year,
71     month,
72
73     CASE
74         WHEN row_num = 1 THEN 'Last Day'
75         ELSE 'Not Last Day'
76     END AS day_type,
77
78     --location,
79     T2.Postal,
80     people_fully_vaccinated_per_hundred,
81     total_vaccinations_per_hundred,
82     total_boosters_per_hundred
83 FROM VaccinationData AS T1
84 RIGHT JOIN 'ores-groupass2.covid_dataset.StateName' AS T2

```

#### IV. Process of adjusting dataset COVID\_19\_Hospital\_Capacity\_by\_Facility” and insert to new table

##### “V\_COVID\_19\_Reported\_Patient\_Impact\_and\_Hospital\_Capacity\_by\_Facility”

```

93 -- Based on [collection_week] to get the [year] and [MONTH], then create the new columns in new table[V_COVID_19_Hospital_Capacity_by_Facility]
94
95 Create table 'ores-groupass2.covid_dataset.V_COVID_19_Hospital_Capacity_by_Facility'
96 AS
97 SELECT K.collection_week,
98     cast (EXTRACT ( YEAR FROM K.collection_week) as string) as Year ,
99     cast (EXTRACT ( MONTH FROM K.collection_week) as string) as MONTH
100 , K.state, K.hospital_name, K.total_beds_7_day_avg, K.total_patients_hospitalized_confirmed_influenza_and_covid_7_day_coverage
101 FROM 'ores-groupass2.covid_dataset.COVID_19_Reported_Patient_Impact_and_Hospital_Capacity_by_Facility' AS K
102
103
104

```

#### V. Process of adjusting dataset “population ” and insert to new table “ V\_population”

```

108 --- to convert original table [location] to state abbreviations [Postal] by using mapping table [StateName], then create it as the new table [V_population]
109
110 CREATE TABLE 'ores-groupass2.covid_dataset.V_population'
111 AS
112 SELECT B.Postal, A.yr2020, A.yr2021, A.yr2022
113 FROM 'ores-groupass2.covid_dataset.population' AS A
114 join 'ores-groupass2.covid_dataset.StateName' AS B
115 On A.location = B.State
116

```

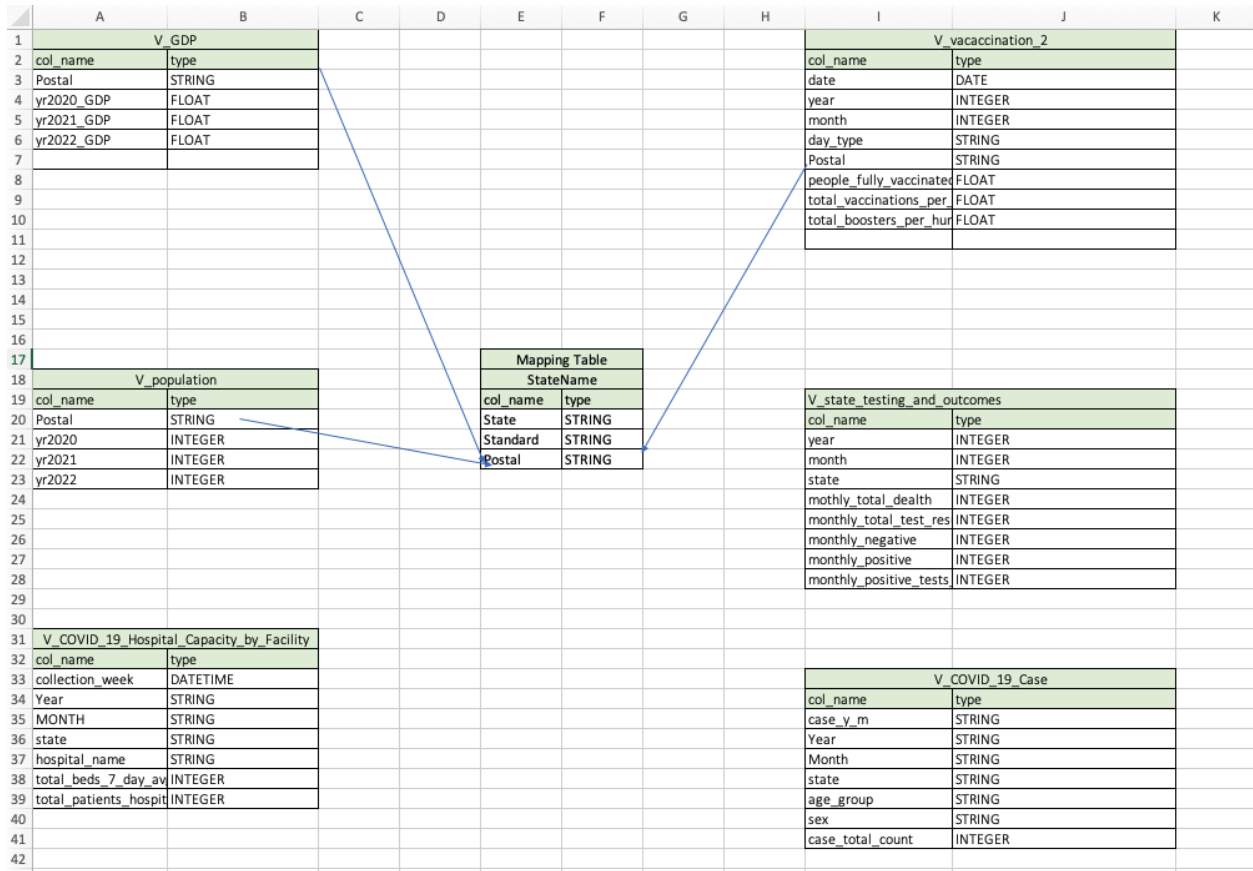
#### VI. Process of adjusting dataset “GDP\_By\_State ” and insert to new table “ V\_GDP”

```

120 --- to convert original table [GeoName ] to state abbreviations [Postal] by using mapping table [StateName], then create it as the new table [V_GDP]
121
122 Create Table 'ores-groupass2.covid_dataset.V_GDP'
123 AS
124 SELECT B.Postal, A.yr2020_GDP, A.yr2021_GDP, A.yr2022_GDP
125 FROM 'ores-groupass2.covid_dataset.GDP_By_State' AS A
126 join 'ores-groupass2.covid_dataset.StateName' AS B
127 On A.GeoName = B.State
128 --LIMIT 1000

```

## VII. State Name Mapping Diagram



## VIII. Result of table “V\_COVID\_19\_Case”

Google Cloud ORESGroupAss2

Search (/) for resources, docs, products, and more

Explorer

Q, Type to search

Viewing workspace resources.

SHOW STARRED ONLY

ores-groupass2

External connections

covid\_dataset

COVID\_19\_Case\_Surveillance\_Public\_Use\_Data\_with...

COVID\_19\_Reported\_Patient\_Impact\_and\_Hospital\_Ca...

GDP\_By\_State

StateName

V\_COVID\_19\_Case

V\_COVID\_19\_Hospital\_Capacity\_by\_Facility

V\_Final\_table

V\_GDP

V\_population

V\_public\_health\_measures

V\_state\_testing\_and\_outcomes

V\_vaccination

V\_vaccination\_2

population

public\_health\_measures

state\_testing\_and\_outcomes

vaccination

bigquery-public-data

Untitled 5

1 SELECT \* FROM `ores-groupass2.covid\_dataset.V\_COVID\_19\_Case` LIMIT 1000

Query completed.

Query results

SAVE RESULTS EXPLORE DATA

JOB INFORMATION RESULTS CHART PREVIEW JSON EXECUTION DETAILS EXECUTION GRAPH

Row	case_y_m	Year	Month	state	age_group	sex	case_total_count
1	2020-01	2020	01	NJ	65+ years	Female	25
2	2020-01	2020	01	NJ	65+ years	Male	19
3	2020-01	2020	01	GA	65+ years	Male	73
4	2020-01	2020	01	GA	65+ years	Female	53
5	2020-02	2020	02	WA	65+ years	Female	52
6	2020-02	2020	02	NY	65+ years	Male	163
7	2020-02	2020	02	WA	65+ years	Male	40
8	2020-02	2020	02	NY	65+ years	Female	199
9	2020-02	2020	02	CA	65+ years	Male	16

Results per page: 50 1 - 50 of 1000

PERSONAL HISTORY PROJECT HISTORY

REFRESH

## IX. Result of table “V\_state\_testing\_and\_outcomes”

The screenshot shows the Google Cloud BigQuery interface. The left sidebar displays the Explorer view with the project 'ores-groupass2' and the dataset 'covid\_dataset'. The main panel shows the query results for the table 'V\_state\_testing\_and\_outcomes'. The query is: `SELECT * FROM `ores-groupass2.covid_dataset.V_COVID_19_Case` LIMIT 1000`. The results are displayed in a table with columns: Row, case\_y\_m, Year, Month, state, age\_group, sex, and case\_total\_count.

Row	case_y_m	Year	Month	state	age_group	sex	case_total_count
1	2020-01	2020	01	NJ	65+ years	Female	25
2	2020-01	2020	01	NJ	65+ years	Male	19
3	2020-01	2020	01	GA	65+ years	Male	73
4	2020-01	2020	01	GA	65+ years	Female	53
5	2020-02	2020	02	WA	65+ years	Female	52
6	2020-02	2020	02	NY	65+ years	Male	163
7	2020-02	2020	02	WA	65+ years	Male	40
8	2020-02	2020	02	NY	65+ years	Female	199
9	2020-02	2020	02	CA	65+ years	Male	16

## X. Result of table “V\_vaccination”

The screenshot shows the Google Cloud BigQuery interface. The left sidebar displays the Explorer view with the project 'ores-groupass2' and the dataset 'covid\_dataset'. The main panel shows the query results for the table 'V\_vaccination'. The query is: `SELECT * FROM `ores-groupass2.covid_dataset.V_vaccination` LIMIT 1000`. The results are displayed in a table with columns: Row, date, year, month, day\_type, Postal, people\_fully\_vaccing, total\_vaccinations, and total\_boosters\_per\_h.

Row	date	year	month	day_type	Postal	people_fully_vaccing	total_vaccinations	total_boosters_per_h
51	2021-03-02	2021	3	Not Last Day	FL	8.26	23.64	null
52	2021-03-02	2021	3	Not Last Day	MN	8.45	25.6	null
53	2021-03-02	2021	3	Not Last Day	HI	10.15	27.74	null
54	2021-11-13	2021	11	Not Last Day	NV	53.72	122.91	9.66
55	2021-11-13	2021	11	Not Last Day	NJ	67.05	144.68	6.15
56	2021-11-13	2021	11	Not Last Day	WV	41.15	89.14	1.18
57	2021-11-13	2021	11	Not Last Day	OK	50.81	118.0	9.49
58	2021-11-13	2021	11	Not Last Day	NE	56.8	126.74	11.92

## XI. Result of table “V\_COVID\_19\_Reported\_Patient\_Impact\_and\_Hospital\_Capacity\_by\_Facility”



## Group assignment2 Group3

The screenshot shows the Google Cloud BigQuery Explorer interface. The left sidebar displays the project structure for 'ores-groupass2', including a 'covid\_dataset' with various tables. The main pane shows a query titled 'Untitled 9' with the following SQL:

```
1 SELECT * FROM `ores-groupass2.covid_dataset.v_COVID_19_Hospital_Capacity_by_Facility` LIMIT 1000
```

The query results are displayed in a table with the following columns: Row, collection\_week, Year, MONTH, state, hospital\_name, total\_beds\_7\_day\_ago, and total\_patients\_hospitalized\_7\_day\_ago. The results show data for various hospitals in Alaska, including Central Peninsula General, Bartlett Regional Hospital, and others.

Row	collection_week	Year	MONTH	state	hospital_name	total_beds_7_day_ago	total_patients_hospitalized_7_day_ago
24	2020-08-16T00:00:00	2020	8	AK	CENTRAL PENINSULA GENER.	null	0
25	2020-08-16T00:00:00	2020	8	AK	BARTLETT REGIONAL HOSPIT.	133	0
26	2020-08-16T00:00:00	2020	8	AK	PROVIDENCE KODIAK ISLAND	null	0
27	2020-08-16T00:00:00	2020	8	AK	PETERSBURG MEDICAL CENTER	12	0
28	2020-08-16T00:00:00	2020	8	AK	PROVIDENCE VALDEZ MEDICA.	null	0
29	2020-08-16T00:00:00	2020	8	AK	ST ELIAS SPECIALTY HOSPITAL	null	0
30	2020-08-16T00:00:00	2020	8	AK	PROVIDENCE ALASKA MEDICA.	null	0
31	2020-08-16T00:00:00	2020	8	AK	ALASKA REGIONAL HOSPITAL	260	0
32	2020-08-16T00:00:00	2020	8	AK	KETCHIKAN MEDICAL CENTER	25	0
33	2020-11-08T00:00:00	2020	11	AK	FAIRBANKS MEMORIAL HOSPI.	101	0
34	2020-11-08T00:00:00	2020	11	AK	PETERSBURG MEDICAL CENTER	12	7

### XII. Result of table "V\_population"

The screenshot shows the Google Cloud BigQuery Explorer interface. The left sidebar displays the project structure for 'ores-groupass2', including a 'covid\_dataset' with various tables. The main pane shows a query titled 'Untitled 10' with the following SQL:

```
1 SELECT * FROM `ores-groupass2.covid_dataset.V_population` LIMIT 1000
```

The query results are displayed in a table with the following columns: Row, Postal, yr2020, yr2021, and yr2022. The results show population data for various states, including NE, SD, NM, AZ, MD, IA, VT, MA, DC, TX, FL, NJ, MO, and CA.

Row	Postal	yr2020	yr2021	yr2022
1	NE	1962642	1963554	1967923
2	SD	887799	896164	909824
3	NM	2118390	2116677	2113344
4	AZ	7179943	7264877	7359197
5	MD	6173205	6174610	6164660
6	IA	3190571	3197689	3200517
7	VT	642893	646972	647064
8	MA	6995729	6989690	6981974
9	DC	670868	668791	671803
10	TX	29232474	29558864	30029572
11	FL	21589602	21828069	22244823
12	NJ	9271689	9267961	9261699
13	MO	6153998	6169823	6177957
14	CA	39501653	39142991	39029342

### XIII. Result of table "V\_GDP"

Query results

Row	Postal	yr2020_GDP	yr2021_GDP	yr2022_GDP
1	NM	100434.5	111730.6	125540.6
2	MT	53130.7	59996.7	67071.9
3	OH	692121.9	759626.2	825990.0
4	MA	589033.3	645434.0	691460.6
5	SD	56162.9	62607.1	68761.7
6	OR	251855.8	275444.0	297308.9
7	WI	343783.0	369032.4	396209.3
8	AZ	386443.5	432279.8	475653.7
9	CT	275801.9	295907.5	319344.8
10	KS	177720.6	191831.7	209326.1
11	SC	249482.2	271494.5	297546.3
12	AL	235118.3	257986.5	281569.0
13	KY	218755.2	237925.9	258981.2
14	VT	34470.8	37593.5	40830.8
15	WY	36675.5	42176.2	49080.6
16	NE	135285.0	149360.3	164933.9

#### IV. Process of joining to “V\_Final\_table”

```

1 create table `ores-groupass2.covid_dataset.V_Final_table`
2 as
3
4 SELECT distinct (CONCAT(cast (T2.Year as string), '-', cast (T2.Month as string), '-', T2.state)) AS T_id,
5 T2.Year, T2.MONTH, T2.state, sum(T2.case_total_count) AS cases_monthly_amount,
6 T6.yr2021 AS yr2021_population, T6.yr2022 AS yr2022_population, --T1.total_beds_7_day_avg,
7 T7.yr2021_GDP, T7.yr2022_GDP, T3.monthly_negative, T3.monthly_positive, T3.monthly_total_test_results,
8 T5.hospital_name, T5.total_beds_7_day_avg, T5.total_patients_hospitalized_confirmed_influenza_and_covid_7_day_coverage,
9 T1.people_fully_vaccinated_per_hundred
10
11 FROM
12 | `ores-groupass2.covid_dataset.V_COVID_19_Case` AS T2
13 left join `ores-groupass2.covid_dataset.V_vaccination_2` AS T1
14 ON cast(T1.Year as string) = T2.Year
15 AND cast( T1.MONTH as string) = T2.MONTH
16 AND T1.Postal = T2.state
17 left join `ores-groupass2.covid_dataset.V_state_testing_and_outcomes` AS T3
18 ON T2.Year = cast (T3.year as string)
19 AND T2.MONTH = cast (T3.month as string)
20 AND T2.state = T3.state
21 /*
22 Response too large to return. Consider specifying a destination table in your job configuration. For more details
23 */
24 left join `ores-groupass2.covid_dataset.V_COVID_19_Hospital_Capacity_by_Facility` AS T5
25 ON T2.Year = cast (T5.year as string)
26 AND T2.MONTH = cast (T5.month as string)
27 AND T2.state = T5.state
28 left join `ores-groupass2.covid_dataset.V_population` AS T6
29 ON T2.state = T6.postal
30 left join `ores-groupass2.covid_dataset.V_GDP` AS T7
31 ON T2.state = T7.Postal
32 --where T2.age_group != '65+ years' AND monthly_negative is not null
33 group by T2.Year, T2.MONTH, T2.state, T6.yr2021, T6.yr2022, --T1.total_beds_7_day_avg,
34 T7.yr2021_GDP, T7.yr2022_GDP, T3.monthly_negative, T3.monthly_positive, T3.monthly_total_test_results,
35 T5.hospital_name, T5.total_beds_7_day_avg, T5.total_patients_hospitalized_confirmed_influenza_and_covid_7_day_coverage,
36 T1.people_fully_vaccinated_per_hundred

```

#### V. Result of table “V\_Final\_table”

# Group assignment2

## Group3

Google Cloud ORES-GroupAss2

Search (/) for resources, docs, products, and more

Search

Untitled 12

SELECT \* FROM 'ores-groupass2.covid\_dataset\_v.Final\_table'

Query results

SAVE RESULTS EXPLORE DATA

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS		EXECUTION GRAPH									
Row	T_ID	Year	MONTH	state	cases_monthly_avg	yr2021_population	yr2022_population	yr2021_GDP	yr2022_GDP	monthly_negative	monthly_positive	monthly_total_test_pos	hospital_name	total_beds_7_day	total_patients	people_fully_vaccing
435	2020-10-AL	2020	10	AL	31539	5049846	5074296	257986.5	281569.0	33826150	5332670	38478012	NOLAND HOSPITAL ...	22	7	null
436	2020-10-AL	2020	10	AL	31539	5049846	5074296	257986.5	281569.0	33826150	5332670	38478012	USA HEALTH CHIL...	273	6	null
437	2020-10-AL	2020	10	AL	31539	5049846	5074296	257986.5	281569.0	33826150	5332670	38478012	SOUTHEAST HEALT...	null	6	null
438	2020-10-AL	2020	10	AL	31539	5049846	5074296	257986.5	281569.0	33826150	5332670	38478012	WASHINGTON COUN...	31	6	null
439	2020-10-AL	2020	10	AL	31539	5049846	5074296	257986.5	281569.0	33826150	5332670	38478012	ATHENS LIMESTON...	107	7	null
440	2020-10-AL	2020	10	AL	31539	5049846	5074296	257986.5	281569.0	33826150	5332670	38478012	ST VINCENT'S CHIL...	46	7	null
441	2020-10-AL	2020	10	AL	31539	5049846	5074296	257986.5	281569.0	33826150	5332670	38478012	NORTH BALDWIN IN...	71	7	null
442	2020-10-AL	2020	10	AL	31539	5049846	5074296	257986.5	281569.0	33826150	5332670	38478012	MEDICAL CENTER E...	106	7	null
443	2020-10-AL	2020	10	AL	31539	5049846	5074296	257986.5	281569.0	33826150	5332670	38478012	USA HEALTH CHIL...	273	7	null
444	2020-10-AL	2020	10	AL	31539	5049846	5074296	257986.5	281569.0	33826150	5332670	38478012	BULLOCK COUNTY H...	77	6	null
445	2020-10-AL	2020	10	AL	31539	5049846	5074296	257986.5	281569.0	33826150	5332670	38478012	OCHSNER CHOCTA...	35	6	null
446	2020-10-AL	2020	10	AL	31539	5049846	5074296	257986.5	281569.0	33826150	5332670	38478012	ST VINCENT'S EAST	null	0	null
447	2020-10-AL	2020	10	AL	31539	5049846	5074296	257986.5	281569.0	33826150	5332670	38478012	JACK HUGHSTON M...	53	5	null
448	2020-10-AL	2020	10	AL	31539	5049846	5074296	257986.5	281569.0	33826150	5332670	38478012	NORTHEAST ALABA...	null	7	null

Results per page: 50 401 - 450 of 75674

PERSONAL HISTORY PROJECT HISTORY

REFRESH