

Covid-19: Creating an Aggregate Dataset

To the Department of Health and Human Services Judging Committee:

Recently in our lifetimes, we have gone through a global pandemic as a collective, which has impacted every aspect of our daily lives, our attitudes, the economy, and our sense of interconnectedness and resilience in the face of unprecedented challenges. We live in an era where we are able to communicate with each other across the globe about the nature of the virus and develop and dispatch vaccines at a rapid pace, which has helped to contain the spread of the disease. In the age of data, we have a vast collection of data that describes everything from vaccination patterns to socioeconomic changes during this pandemic. Through our research, we hope to understand many of these patterns.

"How do community-level COVID-19 cases, hospitalizations, vaccine hesitancy, influenza vaccination rates, allocation of provider relief funds, and weather patterns interrelate and influence the trajectory and management of the COVID-19 pandemic at a regional level?" This research question encompasses a wide range of factors that have a substantial impact on the spread and management of COVID-19. By examining these interconnected aspects, we aim to discern patterns, correlations, and potential causal relationships that can inform public health strategies, resource allocation, and targeted interventions to mitigate the spread of COVID-19 and enhance public health preparedness in the face of future pandemics.

To get started, we need a comprehensive dataset that we can explore. For this reason, we have prepared an aggregate dataset by joining several publicly available datasets. Each dataset contains valuable information that we can use to understand the complex interactions among these factors and their influence on the COVID-19 pandemic's regional dynamics.

Community covid levels, fips/zips converter:

The first dataset contains information on community covid levels by county and health service area. Notable features of this dataset include covid cases per 100k, covid hospital admissions and covid inpatient bed utilization. This dataset provides granular information on COVID-19 incidence at the community level, offering critical insights into the severity and geographic distribution of the pandemic. An additional dataset we have included in the aggregate dataset is a county level fips to zips converter, which has counties and their associated zip codes

and fips codes. This information will help to join together datasets with only zip codes and those with only fips codes.

Provider relief fund:

The next dataset contains information on the Provider Relief Fund Nursing Home Quality Incentive Program that allocated \$2 billion as an incentive to nursing home facilities to reduce both COVID-19 infection rates relative to their county and mortality rates against a national benchmark. Some notable features of this dataset include total payment per facility, total covid infections, county infection rate, infection performance score as well as mortality adjustment.

Flu vaccine coverage:

The next dataset contains information on influenza vaccination coverage for the general population at the national, regional, and state levels by age group and race/ethnicity. Although this dataset is not specifically related to COVID-19, it serves as a crucial piece of the puzzle by providing insights into influenza vaccination coverage. Understanding influenza vaccination rates during the pandemic is essential, as it allows us to differentiate between influenza and COVID-19 cases, enabling more accurate disease monitoring and resource allocation in the face of co-circulating respiratory viruses. Additionally, it aids in assessing the overall burden on healthcare systems during this challenging period.

Vaccine hesitancy:

The next dataset contains information on COVID-19 vaccine hesitancy rates by county. This dataset on COVID-19 vaccine hesitancy rates by county is a critical resource. It includes estimates of hesitancy, a social vulnerability index, and the percentage of the county population by race, offering a comprehensive view of vaccination challenges. Understanding vaccine hesitancy at the community level, considering social vulnerabilities and racial disparities, is essential for targeted public health campaigns and equitable vaccine distribution. This data enhances our research by illuminating the societal factors that influence vaccination decisions during the pandemic.

Covid19 weather, hospital beds and deaths:

The following three datasets, sourced from BigQuery's public datasets, are instrumental in our research effort. They provide valuable information on daily weather observations by county, aiding our understanding of the influence of weather on COVID-19 transmission and availability of hospital beds by county. Furthermore, a comprehensive summary of COVID-19

cases, including details on confirmed cases, active cases, recoveries, and fatalities by region, enables us to assess the pandemic's impact and the effectiveness of public health measures at different geographic levels. These datasets collectively enhance our ability to explore the multifaceted dynamics of the pandemic and inform evidence-based strategies for managing and mitigating its spread.

Description of final dataset and types of research it could help inform:

The dataset we've compiled from various sources holds significant potential for informing diverse research efforts. It allows researchers to analyze the complex interplay between COVID-19 incidence rates, vaccine hesitancy, influenza cases, and weather patterns, providing insights into disease dynamics and management at the regional level. This dataset also supports the evaluation of incentive programs' impact on COVID-19 infection rates and helps differentiate between influenza and COVID-19 cases for more accurate disease monitoring. Moreover, it offers the opportunity to investigate socioeconomic and racial disparities in vaccine acceptance, aiding in targeted public health interventions. Lastly, the inclusion of weather-related data facilitates the exploration of weather patterns' influence on COVID-19 incidence, contributing to our understanding of environmental factors in disease transmission.

APPENDIX

Link to corresponding presentation:

https://docs.google.com/presentation/d/1xa1cULEbszwOsKdj4klm1JaZsti29FWxK7gFBfoBBwU/edit?usp=share_link

Link to recording of presentation:

<https://drive.google.com/file/d/1LJQXmXRFmp8HjtS8w4rpF1uov1bztLS-/view?usp=sharing>

Link to other datasets used: Google Drive folder

https://drive.google.com/drive/folders/1FdiEGA0Qpfv-1hnI2hXqtX-zc_xgOdir?usp=share_link

Vendor Provided Datasets: BigQuery

- Covid_19_weathersource_com.county_day_history
- Covid19_aha.hospital_beds
- Covid19_jhu_esse.summary

Individual Links from where we downloaded the datasets provided in the Google Drive folder:

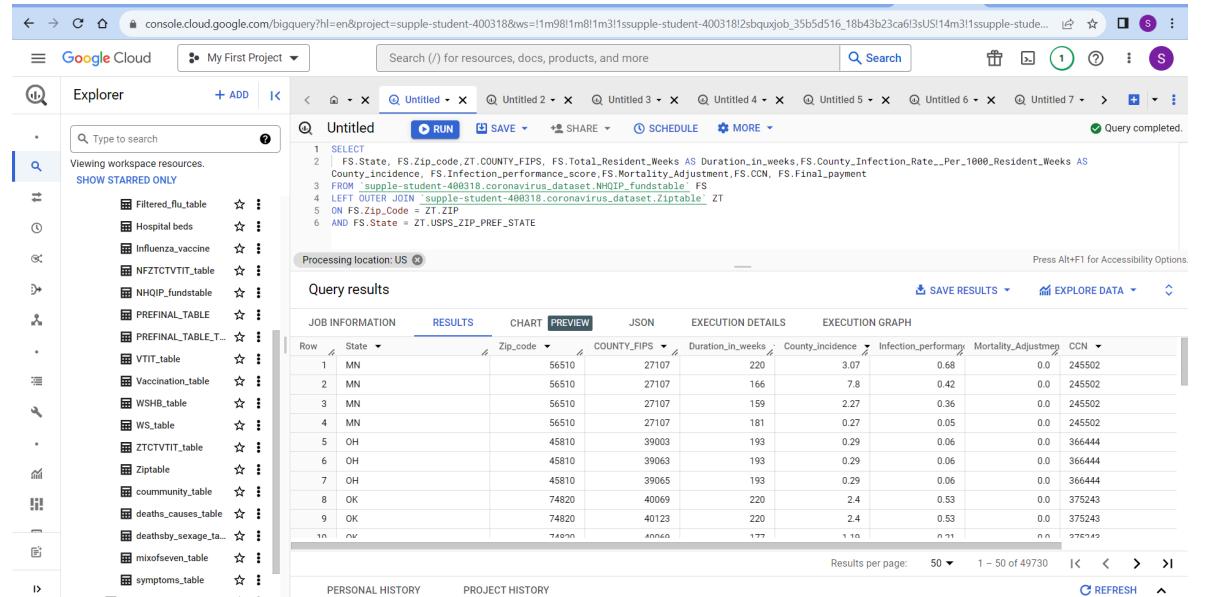
- 1) HUD USPS ZIP Code Crosswalk files. (n.d.).
<https://www.huduser.gov/apps/public/uspccrosswalk/home>
- 2) Centers for Disease Control and Prevention. (n.d.). *United States covid-19 community levels by county*. Centers for Disease Control and Prevention.
<https://data.cdc.gov/Public-Health-Surveillance/United-States-COVID-19-Community-L evels-by-County/3nnm-4jni>
- 3) Centers for Disease Control and Prevention. (n.d.-a). *Covid-19 county hesitancy*. Centers for Disease Control and Prevention.
<https://data.cdc.gov/Vaccinations/COVID-19-County-Hesitancy/c4bi-8ytd>
- 4) Centers for Disease Control and Prevention. (n.d.-b). *Influenza vaccination coverage for all ages (6+ months)*. Centers for Disease Control and Prevention.
<https://data.cdc.gov/Flu-Vaccinations/Influenza-Vaccination-Coverage-for-All-Ages-6-M ont/vh55-3he6>
- 5) Centers for Disease Control and Prevention. (n.d.-c). *Provider relief fund COVID-19 nursing home quality incentive program*. Centers for Disease Control and Prevention.
<https://data.cdc.gov/Administrative/Provider-Relief-Fund-COVID-19-Nursing-Home-Qu ality/bfqg-cb6d>

Resources to help write Queries:

- Writing subqueries in SQL: Advanced SQL - mode. Mode Resources. (2016, May 23).
<https://mode.com/sql-tutorial/sql-sub-queries/>

STEPS TO CREATE AN AGGREGATE DATASET, INCLUDING SCREENSHOTS OF SQL QUERIES:

1. Search public health data repositories: CDC, BigQuery, Snowflake, HealthData.gov, for publicly available datasets related to COVID-19.
 - a. We initially selected Cardiovascular Disease related datasets, but seeing as these datasets did not have common keys to join on, we switched our topic to COVID-19 due to the vast amount of data available as well as county level geographic markers that could be used as a primary key to join on.
2. Create an ongoing document with the name of each table and list the columns of each.
3. Find common column names to use as a key for joining. Highlight these column names.
4. Join 2 tables at a time and analyze the resulting intermediary table. Filter tables as necessary.
 - a. We performed a left outer join to integrate Provider relief funds with the county FIPS , Zip codes, using zip codes as the common key for the linkage as we wanted all the data from provider relief funds and only matched values from ZIP table then we named the resulting table "FSZT_table."



The screenshot shows the Google Cloud BigQuery interface. On the left is the Explorer sidebar listing various tables: Filtered_flu_table, Hospital_beds, Influenza_vaccine, NFZCTVTIT_table, NHQIP_fundstable, PREFINAL_TABLE, PREFINAL_TABLE_T..., VTIT_table, Vaccination_table, WSHB_table, WS_table, ZCTCTVTIT_table, Ziptable, community_table, deaths_causes_table, deathsby_sexage_ta..., mixofseven_table, symptoms_table. The main area shows an untitled query editor with the following SQL code:

```
1 SELECT
2   FS.State, FS.Zip_code,ZT.COUNTY_FIPS, FS.Total_Resident_Weeks AS Duration_in_weeks,FS.County_Infection_Rate_Per_1000_Resident_Weeks AS
3   County_incidence, FS.Infection_performance_score,FS.Mortality_Adjustment,FS.CCN, FS.Final_payment
4   FROM `supple-student-400318.coronavirus_dataset.NHQIP_fundstable` FS
5   LEFT OUTER JOIN `supple-student-400318.coronavirus_dataset.Ziptable` ZT
6   ON FS.Zip_Code = ZT.ZIP
7   AND FS.State = ZT.USPS_ZIP_PREF_STATE
```

The results table below has columns: JOB INFORMATION, RESULTS, CHART, PREVIEW, JSON, EXECUTION DETAILS, EXECUTION GRAPH. The PREVIEW tab is selected, showing the following data:

Row	State	Zip_code	COUNTY_FIPS	Duration_in_weeks	County_incidence	Infection_performance	Mortality_Adjustment	CCN
1	MN	56510	27107	220	3.07	0.68	0.0	245502
2	MN	56510	27107	166	7.8	0.42	0.0	245502
3	MN	56510	27107	159	2.27	0.36	0.0	245502
4	MN	56510	27107	181	0.27	0.05	0.0	245502
5	OH	45810	39003	193	0.29	0.06	0.0	366444
6	OH	45810	39063	193	0.29	0.06	0.0	366444
7	OH	45810	39065	193	0.29	0.06	0.0	366444
8	OK	74820	40069	220	2.4	0.53	0.0	375243
9	OK	74820	40123	220	2.4	0.53	0.0	375243
10	OK	74820	40069	177	1.10	0.71	0.0	375243

- b. Next we joined community levels by county with vaccine hesitancy based on fips code as key and named table “CTVT_table”

```

1 SELECT CT.county_fips, CT.county_population, CT.covid_inpatient_bed_utilization, CT.covid_hospital_admissions_per_100k, CT.covid_cases_per_100k, Social_Vulnerability_Index__SVI__, VT.SVI.Category, VT.Percent_adults_fully_vaccinated_against_COVID_19__as_of_6_10_21
2 FROM `supple-student-400318.coronavirus_dataset.community_table` CT
3 JOIN `supple-student-400318.coronavirus_dataset.Vaccination_table` VT
4 ON CT.county_fips = VT.FIPS_Code
    
```

Row	county_fips	county_population	covid_inpatient_bed	covid_hospital_adm	covid_cases_per_100k	covid_19_community_level	County_Name	Estimated_H
1	42047	29910	3.7	3.3	160.48	Low	Elk County, Pennsylvania	I
2	42047	29910	7.1	10.0	106.99	Medium	Elk County, Pennsylvania	I
3	42047	29910	2.6	0.0	60.18	Low	Elk County, Pennsylvania	I
4	42047	29910	3.6	3.3	45.81	Low	Elk County, Pennsylvania	I
5	42047	29910	1.0	0.0	43.46	Low	Elk County, Pennsylvania	I
6	42047	29910	0.5	3.3	33.43	Low	Elk County, Pennsylvania	I
7	42047	29910	0.5	0.0	16.72	Low	Elk County, Pennsylvania	I
8	42047	29910	0.5	3.3	3.34	Low	Elk County, Pennsylvania	I
9	42047	29910	0.0	0.0	13.37	Low	Elk County, Pennsylvania	I

- c. We filtered the influenza vaccine table using the following command, selecting data for survey years starting from 2019, and designated the resulting table as "Filtered_flu_table."

```

1 SELECT*
2 FROM `supple-student-400318.coronavirus_dataset.Influenza_vaccine` IV
3 WHERE Season_Survey_Year >= '2019';
    
```

Row	Vaccine	Geography_Type	Geography	FIPS	Season_Survey_Year	Month	Dimen
1	Seasonal Influenza	States/Local Areas	Connecticut	9	2019-20	7	Race a
2	Seasonal Influenza	States/Local Areas	Connecticut	9	2019-20	8	Race a
3	Seasonal Influenza	States/Local Areas	Connecticut	9	2019-20	9	Race a
4	Seasonal Influenza	States/Local Areas	Connecticut	9	2019-20	10	Race a
5	Seasonal Influenza	States/Local Areas	Connecticut	9	2019-20	11	Race a
6	Seasonal Influenza	States/Local Areas	Connecticut	9	2019-20	12	Race a
7	Seasonal Influenza	States/Local Areas	Connecticut	9	2019-20	1	Race a
8	Seasonal Influenza	States/Local Areas	Connecticut	9	2019-20	2	Race a
9	Seasonal Influenza	States/Local Areas	Connecticut	9	2019-20	3	Race a
10	Seasonal Influenza	States/Local Areas	Connecticut	9	2019-20	4	Race a

- d. Later we joined “CTVT” table with filtered_flu_table using county FIPS code as common key as we wanted all the values from CTVT table and matched values

from filtered_flu_table

```

1 SELECT CIVT, FIT.Vaccine, FIT.FIPS, FIT.Dimension, FIT.Dimension_Type, FIT.Season_Survey_Year, FIT.Estimate.....
2 FROM `supple-student-400318.coronavirus_dataset.FSZT_table` FSZT
3 LEFT OUTER JOIN `supple-student-400318.coronavirus_dataset.Filtered_flu_table` FIT
4 ON FSZT.country_fips = FIT.FIPS;
5

```

Row	Vaccine	FIPS	Dimension	Dimension_Type	Session_Survey_Year	Estimate
1	Seasonal Influenza	42047	≥18 Years	Age	2019-20	46.8
2	Seasonal Influenza	42047	≥18 Years	Age	2020-21	46.6
3	Seasonal Influenza	42047	≥18 Years	Age	2021-22	45.6
4	Seasonal Influenza	42047	Non-Medical Setting	≥18 Years	2021-22	47.3
5	Seasonal Influenza	42047	≥18 Years	Age	2019-20	46.8
6	Seasonal Influenza	42047	≥18 Years	Age	2020-21	46.6
7	Seasonal Influenza	42047	≥18 Years	Age	2021-22	45.6
8	Seasonal Influenza	42047	Non-Medical Setting	≥18 Years	2021-22	47.3
9	Seasonal Influenza	42047	≥18 Years	Age	2019-20	46.8
10	Seasonal Influenza	42047	≥18 Years	Age	2020-21	46.6

- e. Then “FSZT” was added to “CTVTIT” by using county_FIPS as a common key as we wanted all the values from left table and named it as “NFZTCTVTIT”

```

1 SELECT FSZT, CTVTIT
2 FROM `supple-student-400318.coronavirus_dataset.FSZT_table` FSZT
3 LEFT JOIN `supple-student-400318.coronavirus_dataset.CTVTIT_table` CTVTIT
4 ON FSZT.country_fips= CTVTIT.country_fips

```

Row	Final_payment	C.county_fips	county_population	covid_inpatient_bed	covid_hospital_adm	covid_cases_per_100k	covid_19_community_level	CTVTIT.CTVT_County_Nan
1	31625	37103	9419	3.1	2.3	42.47	Low	Jones County, North Carol
2	31625	37103	9419	0.8	0.8	21.23	Low	Jones County, North Carol
3	31625	37103	9419	0.9	1.8	10.62	Low	Jones County, North Carol
4	31625	37103	9419	2.2	4.9	95.55	Low	Jones County, North Carol
5	31625	37103	9419	4.1	10.7	63.7	Medium	Jones County, North Carol
6	31625	37103	9419	7.4	16.6	276.04	High	Jones County, North Carol
7	31625	37103	9419	7.7	10.2	212.34	High	Jones County, North Carol
8	31625	37103	9419	7.7	10.2	212.34	High	Jones County, North Carol
9	31625	37103	9419	4.6	6.9	254.8	Medium	Jones County, North Carol
10	31625	37103	9419	4.6	6.9	254.8	Medium	Jones County, North Carol
11	31625	37103	9419	3.2	5.4	233.57	Medium	Jones County, North Carol
12	31625	37103	9419	2.6	7.4	201.72	Medium	Jones County, North Carol

- f. Next we filtered covid19 weather by using the WHERE keyword as we wanted data from 2019 and named it as “WS_table”.

The screenshot shows the Google Cloud BigQuery interface. The top navigation bar includes the URL 'console.cloud.google.com/bigquery?hl=en&project=supple-student-400318&ws=!1m8!1m4!1m3!1ssupple-student-400318!2sbquxjob_35b5d516,18b43b23ca6!3sUS!1m4!1m3!1ssupple-st...'. Below the URL is a search bar 'Search (/) for resources, docs, products, and more' and a 'Search' button. The main workspace shows a sidebar with 'Google Cloud' and 'My First Project' selected. A search bar 'Type to search' is present. The main area displays a query editor with a completed query and its results.

Query completed.

```
1 SELECT WS.country, WS.date, WS.min_temperature_air_2m_f, WS.avg_temperature_air_2m_f, WS.max_temperature_air_2m_f, WS.min_humidity_relative_2m_pct, WS.max_humidity_relative_2m_pct, WS.min_humidity_specific_2m_gpk, WS.max_humidity_specific_2m_gpk
2 FROM `bigquery-public-data.covid19_weatherstable.com.county_day_history` WS
3 WHERE
4 DATE(WS.date) BETWEEN "2019-01-01" AND "2022-12-31";
```

Processing location: US

Query results

Row	country	date	min_temperature_air	avg_temperature_air	max_temperature_air	min_humidity_relative	avg_humidity_relative	max_humidity_relative	min_humidity_specific
1	US	2021-07-26	70.6	80.3	89.2	12	17	26	10.0
2	US	2021-07-26	61.7	75.1	89.4	11	19	28	10.0
3	US	2021-07-26	57.7	70.3	82.1	11	19	29	10.0
4	US	2021-07-26	68.7	77.4	86.0	14	21	29	10.0
5	US	2021-07-26	67.3	80.8	93.6	15	23	31	10.0
6	US	2021-07-26	72.8	86.2	99.6	10	20	34	10.0
7	US	2021-07-26	63.0	75.4	89.1	15	23	35	10.0
8	US	2021-07-26	75.5	87.4	101.9	11	22	36	10.0
9	US	2021-07-26	64.6	78.3	90.9	16	26	37	10.0
10	US	2021-07-26	63.9	81.3	97.7	8	20	37	10.0

RESULTS PAGE: 50 | 1 - 50 of 3664332

PERSONAL HISTORY PROJECT HISTORY

REFRESH

- g. We joined ‘WS_table’ with hospital_beds table based on county_fips_code as a common key as we wanted all the rows from left table and named resulting table as “WSHB_table”

The screenshot shows the Google Cloud BigQuery interface. The top navigation bar includes tabs for 'Login2_5', 'Using m...', 'My App...', 'Word |', 'Login2', 'Group A...', 'Covid D...', 'Covid-11', 'Covid-11', 'Covid-11', and 'coronavirus...'. The main title is 'console.cloud.google.com/bigquery?hl=en&project=supple-student-400318&ws=!lmp891m4!1m3!1ssupple-student-400318!2sbqujob_35b5d516_18b43b23ca6!3sUS!1m4!1m3!1ssupple-st...'. The left sidebar has sections for 'Explorer' and 'Viewing workspace resources. SHOW STARRED ONLY', listing datasets like 'coronavirus_dataset' containing tables such as 'CTVITI_Table', 'CTVTB_table', 'DCTDSAT_table', 'Deaths_table', 'FINAL_TABLE', 'FSZT_table', 'Filtered_flu_table', 'Hospital_beds', 'Influenza_vaccine', 'NFZCTVITI_table', 'NHQIP_fundstable', 'PREFINAL_TABLE', 'PREFINAL_TABLE_T...', 'VTIT_table', 'Vaccination_table', and 'WSHB_table'. The main area shows a query titled 'Untitled 11' with the following SQL code:

```
1 SELECT WS.HB.cbsa_code, HB.total_adult_beds, HB.acute_long_term_care_beds, HB.burn_care_beds, HB.cardiac_intensive_care_beds,
2 HB.gen_medical_surgical_adult_beds, HB.gen_medical_surgical_pediatric_beds, HB.intermediate_nursing_care_beds, HB.medical_surgical_intensive_care_beds,
3 HB.neonatal_intensive_care_beds, HB.neonatal_intermediate_care_beds, HB.obstetric_care_beds, HB.other_care_beds, HB.other_intensive_care_beds
4 FROM `supple-student-400318.coronavirus_dataset.WS_table` WS
5 LEFT JOIN `supple-student-400318.coronavirus_dataset.Hospital_beds` HB
6 ON WS.county_fips_code = HB.county_fips_code
```

The results section shows the following data:

Row	WS.country	WS.county_fips_code	WS.date	min_temperature	avg_temperature	max_temperature	min_humidity	avg_humidity
1	US	48141	2020-06-07	69.5	84.1	96.4	6	
2	US	04013	2020-06-07	70.5	82.9	94.9	7	
3	US	04019	2020-06-07	63.3	77.2	89.3	18	
4	US	18157	2020-06-07	58.1	73.9	86.2	26	
5	US	06047	2020-06-07	50.0	62.8	73.2	24	
6	US	53005	2020-06-07	45.0	58.2	68.9	24	
7	US	19163	2020-06-07	61.7	76.4	89.0	35	
8	US	42101	2020-06-07	63.2	70.5	78.0	34	
9	US	29510	2020-06-07	72.7	83.6	93.0	38	
10	US	24003	2020-06-07	62.3	73.1	80.5	37	

At the bottom, there are buttons for 'PERSONAL HISTORY' and 'PROJECT HISTORY', and a 'REFRESH' button.

- h. Later we filtered Covid19_jhu_csse.summary based on criteria and named it as deaths_table which was joined with PREFINAL_table.

The screenshot shows the Google Cloud BigQuery interface. The left sidebar lists various datasets and tables, including NFZTCTVTIT_table, PREFINAL_TABLE, PREFINAL_TABLE_TWO, and WSHB_table. The main area displays a query titled "Untitled 12" with the following SQL code:

```

1 SELECT
2   latitude,
3   longitude,
4   confirmed,
5   deaths,
6   fips
7 FROM `bigquery-public-data.covid19_jhu_csse.summary`
8 WHERE country_region = "US";
9 
```

The "Query results" section shows a table with 10 rows of data. The columns are latitude, longitude, confirmed, deaths, and fips. The data includes coordinates for various US locations and their corresponding values for confirmed cases, deaths, and fips. The table has headers for Row, latitude, longitude, confirmed, deaths, and fips. The results page indicates 3513278 total rows.

- i. NFZTCTVTIT was joined with WSHB to form a table named “FINAL_table”. We've limited the final dataset to 20 million rows and 57 columns, totaling 8.31 GB, to optimize query performance due to significant time delays.

The screenshot shows the Google Cloud BigQuery interface. The left sidebar lists various datasets and tables, including NFZTCTVTIT_table, PREFINAL_TABLE, PREFINAL_TABLE_TWO, and WSHB_table. The main area displays a query titled "Untitled 13" with the following SQL code:

```

1 SELECT NFZTCTVTIT.WSHB
2 FROM `supple-student-400318.coronavirus_dataset.NFZTCTVTIT_table` NFZTCTVTIT
3 LEFT JOIN `supple-student-400318.coronavirus_dataset.WSHB_table` WSHB
4 ON NFZTCTVTIT.ZCTVTIT.CTIVTIT.CTIVT.country_fips = CAST(WSHB.WS.county_fips_code AS INT64)
5 LIMIT 2000000
6
7 
```

The "Query results" section shows a table with 15 rows of data. The columns are NFZTCTVTIT.State, Total_Resident_West, County_Infection_Rate, Infection_Performance, Mortality_Adjustment, Final_Payment, NFZTCT...ZIP, N.county_fips, and col. The data includes information about different US states and their respective infection rates, mortality adjustments, and final payments. The table has headers for Row, NFZTCTVTIT.State, Total_Resident_West, County_Infection_Rate, Infection_Performance, Mortality_Adjustment, Final_Payment, NFZTCT...ZIP, N.county_fips, and col. The results page indicates 20000000 total rows.

5. Finally we changed the names of the columns as per our requirements.

```

SELECT mixofseven.WSHB_WS.country AS Country, mixofseven.NFZCTVITIT.State AS State, mixofseven.NFZCTVITIT.ZCTVITIT.CTVIT.County_Name AS Counties, deaths.Latitude AS Latitude, deaths.longitude AS longitude, mixofseven.WSHB_WS.county_fips_code AS FIPS_CODE, mixofseven.NFZCTVITIT.ZCTVITIT.ZIP, mixofseven.NFZCTVITIT.ZCTVITIT.CTVIT.Dimension_Type AS Age_Dimension_type, mixofseven.NFZCTVITIT.ZCTVITIT.CTVIT.Dimension AS Age_Dimension, mixofseven.NFZCTVITIT.ZCTVITIT.CTVIT.Season_Survey_Year AS Season_Survey_Year, mixofseven.NFZCTVITIT.ZCTVITIT.CTVIT.Percent_adults_fully_vaccinated_against_COVID_19_as_of_6_10_21 AS Adults_vaccinated, mixofseven.WSHB_WS.date AS Date, mixofseven.NFZCTVITIT.ZCTVITIT.CTVIT.Hospital_Admissions, mixofseven.NFZCTVITIT.ZCTVITIT.CTVIT.COVID_Cases_per_100K AS Covid_Cases, mixofseven.NFZCTVITIT.ZCTVITIT.CTVIT.COVID_Hospital_Admissions_per_100K AS Hospital_Admissions, mixofseven.NFZCTVITIT.ZCTVITIT.CTVIT.COVID_Inpatient_Bed_Utilization AS Covid_Inpatient_Beds, mixofseven.NFZCTVITIT.ZCTVITIT.CTVIT.COVID_Community_Level AS Covid_Community_Level, mixofseven.WSHB_WS.min_temperature.air_2m_f AS MIN_Temperature, mixofseven.WSHB_WS.avg_temperature.air_2m_f AS AVG_Temperature, mixofseven.WSHB_WS.min_humidity.relative_2m_pct AS MIN_Humidity, mixofseven.WSHB_WS.avg_humidity.relative_2m_pct AS AVG_Humidity, mixofseven.WSHB_WS.max_humidity.relative_2m_pct AS MAX_Humidity, mixofseven.WSHB_WS.max_humidity_specific_2m_gpk AS MAX_Humidity_specific, mixofseven.WSHB_WS.cbsa_code AS CBSA_Code, mixofseven.NFZCTVITIT.Total_Resident_Weeks AS Total_resident_weeks, mixofseven.NFZCTVITIT.Infection_Performance_Score AS Infection_Performance
  
```

Row	Country	State	Counties	Latitude	longitude	FIPS_CODE
1	US	PA	Adams County, Pennsylvania	39.87140411	-77.21610347	42001
2	US	PA	Adams County, Pennsylvania	39.87140411	-77.21610347	42001
3	US	PA	Adams County, Pennsylvania	39.87140411	-77.21610347	42001
4	US	TX	Nueces County, Texas	27.736286	-97.5433285	48355
5	US	TX	Sherman County, Texas	36.27751473	-101.8933156	48421
6	US	MN	Le Sueur County, Minnesota	44.37136597	-93.72844721	27079

Below is the screenshot of our final data set with its schema, size, number of observations and columns:

Schema

Field name	Type	Mode	Key	Collation	Default Value	Policy Tags	Description
Country	STRING	NULLABLE					
State	STRING	NULLABLE					
Counties	STRING	NULLABLE					
Latitude	FLOAT	NULLABLE					
longitude	FLOAT	NULLABLE					
FIPS_CODE	STRING	NULLABLE					
ZIP	INTEGER	NULLABLE					
County_Population	INTEGER	NULLABLE					
Age_Dimension	STRING	NULLABLE					
Age_Dimension_type	STRING	NULLABLE					
Adults_vaccinated	FLOAT	NULLABLE					
date	DATE	NULLABLE					
Season_Survey_Year	STRING	NULLABLE					

Google Cloud My First Project Search (/) for resources, docs, products, and more

FINAL_TABLE QUERY SHARE COPY SNAPSHOT DELETE EXPORT REFRESH

SCHEMA	DETAILS	PREVIEW	LINEAGE	DATA PROFILE	DATA QUALITY
Season_Survey_Year	STRING	NULLABLE			
Seasonal_Vaccine	STRING	NULLABLE			
Total_beds	INTEGER	NULLABLE			
Covid_cases	FLOAT	NULLABLE			
Hospital_admissions	FLOAT	NULLABLE			
Covid_inpatient_beds	FLOAT	NULLABLE			
Covid_Community_Level	STRING	NULLABLE			
MIN_temperature	FLOAT	NULLABLE			
AVG_temperature	FLOAT	NULLABLE			
MAX_temperature	FLOAT	NULLABLE			
MIN_humidity	INTEGER	NULLABLE			
AVG_humidity	INTEGER	NULLABLE			
MAX_humidity	INTEGER	NULLABLE			
MIN_humidity_specific	FLOAT	NULLABLE			
AVG_humidity_specific	FLOAT	NULLABLE			
MAX_humidity_specific	FLOAT	NULLABLE			

EDIT SCHEMA VIEW ROW ACCESS POLICIES PERSONAL HISTORY PROJECT HISTORY REFRESH

Google Cloud My First Project Search (/) for resources, docs, products, and more

FINAL_TABLE QUERY SHARE COPY SNAPSHOT DELETE EXPORT REFRESH

SCHEMA	DETAILS	PREVIEW	LINEAGE	DATA PROFILE	DATA QUALITY
Intermediate_nursing_care_devs	INTEGER	NULLABLE			
medical_surgical_intensive_care_beds	INTEGER	NULLABLE			
neonatal_intensive_care_beds	INTEGER	NULLABLE			
neonatal_intermediate_care_beds	INTEGER	NULLABLE			
obstetric_care_beds	INTEGER	NULLABLE			
other_care_beds	INTEGER	NULLABLE			
other_intensive_care_beds	INTEGER	NULLABLE			
Estimated_hesitant	FLOAT	NULLABLE			
Estimated_hesitant_or_unsure	FLOAT	NULLABLE			
Estimated_strongly_hesitant	FLOAT	NULLABLE			
Social_Vulnerability_Index	FLOAT	NULLABLE			
SVI_Category	STRING	NULLABLE			
confirmed	INTEGER	NULLABLE			
deaths	INTEGER	NULLABLE			
fips	STRING	NULLABLE			

EDIT SCHEMA VIEW ROW ACCESS POLICIES PERSONAL HISTORY PROJECT HISTORY REFRESH

Size:

Google Cloud My First Project Search (/) for resources, docs, products, and more

Untitled 16 RUN SAVE SHARE SCHEDULE MORE This query will process 3.48 GB when run.

```

1 SELECT NFZCTVIT1, WSHB
2 FROM `supple-student-400318.coronavirus_dataset.NFZCTVIT1_table` NFZCTVIT1
3 LEFT JOIN `supple-student-400318.coronavirus_dataset.WSHB_table` WSHB
4 ON NFZCTVIT1.ZTCVIT1.CIVIT1.county_fips = CAST(WSHB.WS.county_fips_code AS INT64)
5

```

Google Cloud My First Project Search (/) for resources, docs, products, and more

FINAL_TABLE QUERY SHARE COPY SNAPSHOT DELETE EXPORT REFRESH

SCHEMA DETAILS PREVIEW LINEAGE DATA PROFILE DATA QUALITY

Default rounding mode: ROUNDING_MODE_UNSPECIFIED
Case insensitive: false
Description:
Labels:
Primary key(s):

Storage info

Number of rows: 20,000,000
Total logical bytes: 8.31 GB
Active logical bytes: 8.31 GB
Long term logical bytes: 0 B
Total physical bytes: 494.71 MB
Active physical bytes: 494.71 MB
Long term physical bytes: 0 B
Time travel physical bytes:

PERSONAL HISTORY PROJECT HISTORY

Preview of our Final Dataset:

Google Cloud My First Project Search (/) for resources, docs, products, and more

FINAL_TABLE QUERY SHARE COPY SNAPSHOT DELETE EXPORT REFRESH

SCHEMA DETAILS PREVIEW LINEAGE DATA PROFILE DATA QUALITY

Row	Country	State	Counties	Latitude	Longitude	FIPS_CODE	ZIP
1	US	PA	Adams County, Pennsylvania	39.87140411	-77.21610347	42001	17325
2	US	PA	Adams County, Pennsylvania	39.87140411	-77.21610347	42001	17325
3	US	PA	Adams County, Pennsylvania	39.87140411	-77.21610347	42001	17325
4	US	TX	Nueces County, Texas	27.736286	-97.5433285	48355	78363
5	US	TX	Sherman County, Texas	36.27751473	-101.8933156	48421	79084
6	US	MN	Le Sueur County, Minnesota	44.37136597	-93.72844721	27079	56071
7	US	GA	Murray County, Georgia	34.78897608	-84.74993177	13213	30705
8	US	MD	Baltimore County, Maryland	39.45784712	-76.62911955	24005	21234
9	US	MN	Anoka County, Minnesota	45.27476015	-93.24604565	27009	55092
10	US	WV	Brooke County, West Virginia	40.27367263	-80.57915332	54009	26062
11	US	OH	Portage County, Ohio	41.16793482	-81.19735782	39133	44278
12	US	FL	Orange County, Florida	28.51367621	-81.31799498	12095	32757
13	US	FL	Holmes County, Florida	30.86747479	-85.81319222	12059	32440
14	US	MO	Franklin County, Missouri	38.41127108	-91.07522826	29071	63080
15	US	GA	Murray County, Georgia	34.78897608	-84.74993177	13213	30705
16	US	VA	Floyd County, Virginia	36.9302126	-80.36738849	51063	24141
17	US	VA	Suffolk city, Virginia	36.70155394	-76.63603554	51800	23434
18	US	PA	Adams County, Pennsylvania	39.87140411	-77.21610347	42001	17325

Results per page: 50 1 – 50 of 2000000