Executive Report on Comprehensive Analysis of Factors Influencing the Health Insurance Coverage in the United States

Introduction:

It is fundamental right of every individual to receive health-care treatment easily and efficiently, and health insurance plays a pivotal role in determining who has access to it fairly. Health insurance coverage is an important aspect in the present times as medical care is expensive, especially when hospitalization can burn a hole in one's pocket and derail finances. Enrollment in coverage supports the health and well-being of individuals & society.

This executive summary offers a high-level overview of our study's effort to look at the many variables affecting health insurance coverage in the US. This report is on creating a research dataset that is done by joining all the useful columns from the factor associated datasets that contribute to the likelihood to obtain or retain health insurance coverage.

Overview of socio-economic factors affecting health insurance coverage:

Social determinants of health can be grouped into 5 domains they are: economic stability, education access and quality, health care access and quality, environment, social and community context ⁽⁹⁾. Income was shown to be a determinant of the quality of care received ⁽⁸⁾. More than any other risk factor, being poor or nearly poor (earning less than 200 percent of the federal poverty line) imposed the largest burden and reduced quality-adjusted life expectancy ⁽¹⁰⁾.

Overview of datasets: The following datasets which are regarded as the above-mentioned factors contributing to the likelihood to obtain or retain health insurance coverage are used to create a final research dataset:

- 1. **Demographics:** This dataset ⁽²⁾ contains crucial information about the population composition that can be found in the demographics, which makes it possible to identify vulnerable groups and modify health insurance policies to better suit their needs.
- 2. **Poverty level:** By highlighting financial fragility, the poverty dataset ⁽⁷⁾ helps identify individuals who may find it difficult to obtain health insurance, enabling customized policy actions.
- 3. **Medicaid and CHIP (Children Health Insurance Program) enrollment:** This dataset ⁽³⁾ provides information about the range of coverage, allowing policymakers to evaluate the success of present initiatives and identify possibilities for future expansions to guarantee full health insurance coverage.
- 4. **Uninsured:** this dataset identifies at-risk populations, aiding targeted interventions for improved health insurance access. This is from aspe.hhs.gov ⁽⁴⁾.
- 5. **Unemployment:** This dataset ⁽¹⁾ demonstrates an economic strain that affects one's capacity to purchase health insurance and draws attention to people in need.
- 6. **Income and poverty:** This dataset ⁽⁵⁾ reveals the income gaps impacting health insurance accessibility and forming strategies for targeted coverage expansion.
- 7. **Measures of health:** This data set ⁽⁶⁾ provides information about a population's general health state, helping to identify health risks and guiding specific insurance programs and policies.

Description of the Final dataset:

• Schema: the final dataset contains a total of 34 useful columns, and they are:

Column name	Datatype	Description
State name	string	arranged in alphabetical order and represents states from the US.
FIPS code	integer	which has been taken from the income and poverty dataset and indicates the state's FIPS code ID.
County name	string	indicates specific area name in the particular state.
Population size	integer	indicates the count of population from each state.
Population density	integer	represents the average number of individuals per unit area in a given state.
Age 19 under	float	represents the population of age under 19 indicating children.
Age 19-64	float	denotes population in between the age groups 19 and 64.
Age 65-84	float	indicates people aged from 65 to 84.
Age 85 and over	float	shows the population of age 85 and over.
Number in poverty	integer	Indicates the count of population in poverty in specific area.
Percent in poverty	float	Explains the number of people in poverty range per 100 members in each state.
Poverty	float	Explains the number of people in poverty range per 100 members in each county.
Non-elderly	integer	Shows the number of young populations in each state.
population		
Non-elderly uninsured	integer	Indicates the number of uninsured young population in
population		each state which gives scope to find customers.
Uninsured age 0-18	integer	Indicates uninsured specific age group people from 0-18
Uninsured age 19-34	integer	Indicates uninsured and middle-aged people
Uninsured age 35-49	integer	Indicates uninsured and above middle-aged people
Uninsured age 50-64	integer	Represents uninsured and elderly people
Uninsured male	integer	Shows the number of uninsured people by gender
Uninsured female	integer	Shows the number of uninsured people by gender
Total Medicaid and chip enrollment	integer	Shows the number of Medicaid enrollees per each county
Total chip enrollment	integer	Shows the total number of chip enrollees per each area
Total chip enrollment	integer	Shows the total number of chip enrollees per each area
Employed 2022	integer	Shows the number of employed persons per each county
Unemployed 2022	integer	Shows the unemployed people per each county
Unemployment rate	integer	Shows number of unemployed persons per 10 members
2022	6	in each county
Median household	integer	Indicates the median income of households that
income percent of		contributes to income percent of state total per each
state total		county
Insurance coverage	Integer	Indicates the uninsured type of coverage

Number of persons in	Integer	Number of persons uninsured per each county
insurance coverage		
Deaths	integer	Average deaths per each county
Health status	integer	An indicator showing health of individuals in each area
Unhealthy days	integer	Average number of days with sickness per each county
Less than high school	integer	Indicates the education level of people per each state
High school diploma	Integer	The number of people with high school diploma
College grad	integer	The number of people who have completed college
		graduation

- **Number of rows:** the final dataset contains 77803005 rows that will process 21.2 GB of data.
- **Dataset's size:** the size of the final research dataset is 21.2GB which includes the information on the above specified columns. This dataset is a structured one as it is created by joining the useful columns and filtering out the useful conditions by writing a SQL query in BigQuery data warehouse platform.

Useful Research types of the dataset:

- **Policy Analysis:** The dataset can be used to evaluate how well the present healthcare laws and initiatives, such Medicaid and CHIP, are able to cover vulnerable groups for medical expenses. Researchers can assess how these policies affect health outcomes and poverty reduction.
- **Comparative Research:** By comparing data from several counties and states, researchers can find similarities and variations in several indicators, including population density, income levels, educational attainment, and health outcomes which results in an improved understanding of how these variables affect health insurance coverage.

Research questions:

- 1. How do variations in the distribution of Medicaid and CHIP (Children's Health Insurance Program) enrollment impact the overall rates of uninsured populations, and what are the associated implications for public health?
- 2. What is the relationship between educational attainment levels and health outcomes, and how does this impact the likelihood of obtaining health insurance coverage?

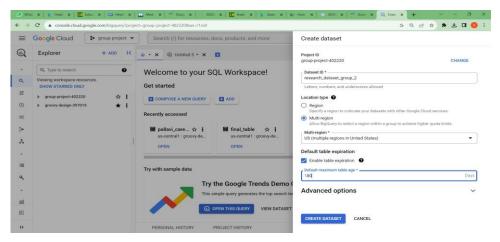
Conclusion:

Our research dataset focused on understanding the underlying factors that impact an individual's ability to secure adequate health insurance. This would be useful to make well-informed policy suggestions and focused actions to rectify inequalities by the stakeholders and policymakers in the healthcare field. By leveraging the dataset, we aimed to identify correlations and associations between various socioeconomic indicators and health insurance coverage.

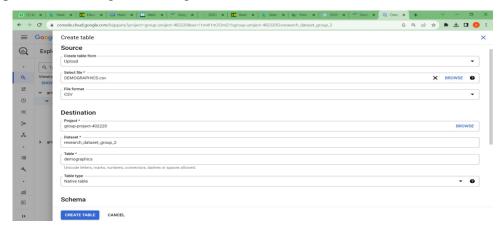
According to the findings, socioeconomic status has a significant influence on healthcare accessibility. Targeted initiatives are therefore necessary to alleviate insurance coverage gaps. To promote a more inclusive and equitable healthcare system and, eventually, improve health outcomes and reduce gaps in healthcare access among diverse communities, the dataset offers a strong basis for evidence-based policy recommendations.

Appendix:

• A picture indicating the creation of dataset named research dataset group 2.

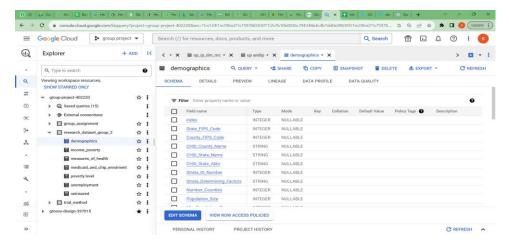


• A picture indicating the loading of table inside the created dataset.



The same way is followed for loading all the 7 tables inside the dataset.

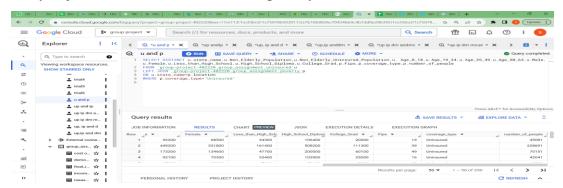
• A picture indicating the loaded tables inside a dataset named research dataset group 2 in BigQuery



• This dataset is created by joining the uninsured groups and education level columns from the uninsured dataset insurance coverage and number of uninsured columns from poverty dataset followed by joining poverty details such as percentage and population in poverty from income and poverty dataset then, columns from demographics like population, density, and age groups are joined to the existing table. Next, death counts, health status and unhealthy days columns from measures of health dataset are joined then, total enrollments in Medicaid and chip are joined to the table and filtered the condition where enrollees are in between 100000 to 1000000. Finally, the number of people employed and unemployed and unemployment rate, household income columns are joined by filtering out the null condition.

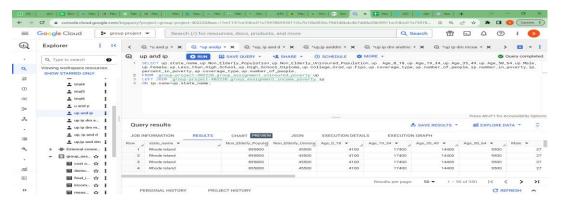
Queries: To combine all the 7 datasets mentioned in the report 6 queries have been written.

1. Query to join uninsured dataset with the poverty dataset.



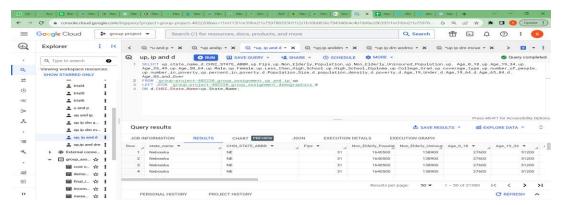
Explanation: In this SQL query, uninsured (u) and poverty (p) tables are joined, and rows are selected to extract state-specific uninsured population data and poverty data.

- Select all columns from the uninsured table (u) to obtain state name, non-elderly and uninsured population, age, gender, and education level.
- Left join the poverty table (p) on state name to filter "Uninsured" rows.
- FIPS code, coverage type, and population columns from each state's poverty table are added.
- We avoid duplicates by getting one row per state with the DISTINCT keyword.
- 2. Query to join the table formed previously in the above query with the income and poverty dataset.



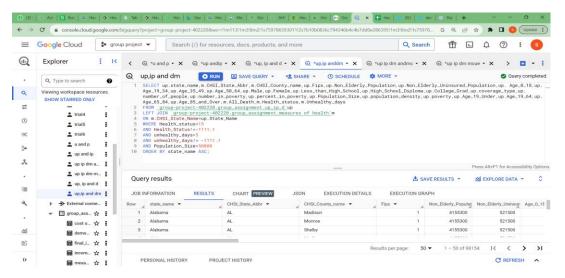
Explanation: This SQL query joins the above created table with the income and poverty table by specifying the columns number in poverty, percent in poverty in the SELECT keyword and then joining the columns using LEFT JOIN on the condition where the state name in both the tables matches with one another.

3. Query to join the previously formed table from the above query with the demographics dataset.



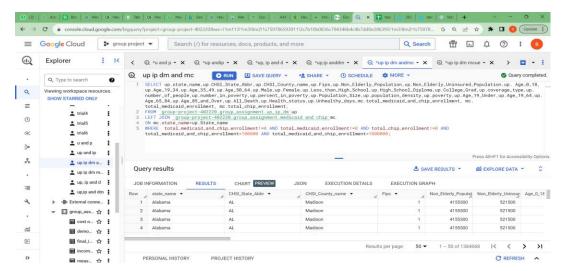
Explanation: This SQL query joins the prior table with the demographics dataset by specifying the population size, population density, poverty, age 19 under and age 19-64, age 65-84, age 85 and over using SELECT keyword and LEFT JOIN and specifying the condition where state name matches in both the tables.

4. Query to join the previously formed table from the above query with the measures of health dataset.



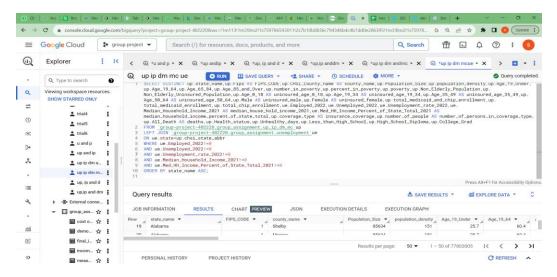
Explanation: this SQL query joins the previously created intermediate tables and joins with the measures of health table by specifying the SELECT keyword mentioning the columns like all deaths, health status and unhealthy days from the measures of health table. They are joined by using LEFT JOIN mentioning the condition where state name matches in both the tables and removing the redundancies by using not equals operator and where health status is less than 15 and unhealthy days are more than 5 which indicates the vulnerable groups.

5. Query to join the previously formed dataset from the above query with the Medicaid and CHIP enrollment.



Explanation: this SQL query joins the above created intermediate table with the medicaid and chip enrollment dataset by specifying the columns that are total Medicaid and chip enrollment, total Medicaid enrollment, total chip enrollment by using SELECT keyword and LEFT JOIN by specifying the condition where the state names match in both the tables. A condition is specified to represent the enrollees in Medicaid and chip in between 100000 and 1000000 and to remove the null values.

6. Query to join the previously formed table from the above query with the unemployment dataset.

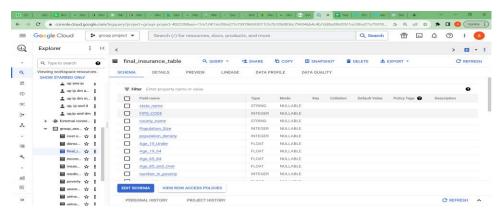


Explanation: this query joins the previously created intermediate tables with the unemployment dataset by specifying the columns like Employed_2022, Unemployed_2022, Unemployment_rate_2022, Median_Household_Income_2021, Med_HH_Income_Percent_of_State_Total_2021 using SELECT keyword and then LEFT JOIN using the condition where state abbreviation matches in both the tables. And cleaned the data by removing all the null values and selecting only distinct state names and then finally the order is in ascending order.

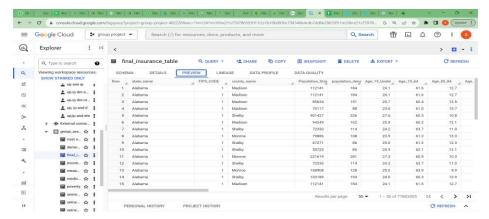
The final query:

```
SELECT DISTINCT up.state_name,up.Fips AS FIPS_CODE,up.CHSI_County_name AS
county_name, up.Population_Size, up.population_density, up.Age_19_Under, up.Age_19_64, up.A
ge_65_84, up.Age_85_and_Over, up.number_in_poverty, up.percent_in_poverty, up.poverty, up.N
on_Elderly_Population,up.Non_Elderly_Uninsured_Population,up.Age_0_18 AS
uninsured_age_0_18,up.Age_19_34 AS uninsured_age_19_34,up.Age_35_49 AS
uninsured_age_35_49,up.Age_50_64 AS uninsured_age_50_64,up.Male AS
uninsured_male, up.Female AS
uninsured_female,up.total_medicaid_and_chip_enrollment,up.total_medicaid_enrollment,up
.total_chip_enrollment,ue.Employed_2022,ue.Unemployed_2022,ue.Unemployment_rate_2022,u
e.Median_Household_Income_2021 AS
median_house_hold_income_2021,ue.Med_HH_Income_Percent_of_State_Total_2021 AS
median_household_income_percent_of_state_total,up.coverage_type AS
insurance_coverage,up.number_of_people AS
number_of_persons_in_coverage_type, up.All_Death AS
deaths, up. Health_status, up. Unhealthy_days, up. Less_than_High_School, up. High_School_Dipl
oma, up.College_Grad
FROM `group-project-402220.group_assignment.up_ip_dm_mc`up
LEFT JOIN `group-project-402220.group_assignment. unemployment`ue
ON ue.state=up.chsi_state_abbr
WHERE ue.Employed_2022!=0
AND ue.Unemployed_2022!=0
AND ue.Unemployment_rate_2022!=0
AND ue.Median_Household_Income_2021!=0
AND ue.Med_HH_Income_Percent_of_State_Total_2021!=0
ORDER BY state_name ASC;
```

Preview and schema of the final dataset: the schema shows the columns and indicates their datatypes.

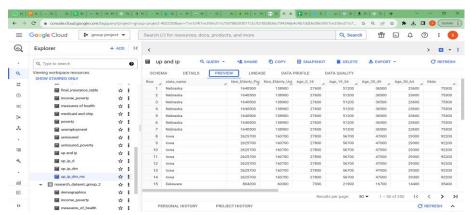


The preview shows the final table indicating each column's values and the number of rows, i.e., 77803005, at the bottom in the picture.

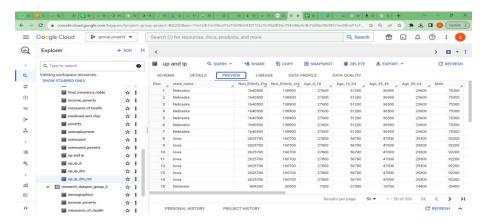


Preview and schemas of intermediate tables:

This picture indicates the preview of intermediate table created by joining uninsured, poverty and income-poverty datasets. Likewise, 5 intermediate tables are created to join 7 datasets where the last query yielded the final dataset which is mentioned above.



This picture indicates the schema of an intermediate table created by joining the uninsured, poverty, income-poverty, demographics, measures of health, Medicaid, and chip enrollment tables.



References:

- 1. USDA ERS County-level Data Sets: Download Data. Accessed October 19, 2023. https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data/.
- 2. Demographic trends and Health Outcomes in the U.S. Kaggle. Accessed October 19, 2023. https://www.kaggle.com/datasets/thedevastator/demographic-trends-and-health-outcomes-in-the-u.
- 3. Medicaid enrollment data collected through mbes. Medicaid. Accessed October 19, 2023. https://www.medicaid.gov/medicaid/national-medicaid-chip-program-information/medicaid-chip-enrollment-data/medicaid-enrollment-data-collected-through-mbes/index.html.
- 4. Health insurance coverage of the total population. KFF. October 28, 2022. Accessed October 19, 2023. <a href="https://www.kff.org/other/state-indicator/total-population/?dataView=1¤tTimeframe=0&selectedDistributions=medicaid-medicare--uninsured--total&selectedRows=%7B%22states%22%3A%7B%22all%22%3A%7B%7D%7D%7D&sortModel=%7B%22colId%22%3A%22Location%22%2C%22sort%22%3A%22asc%22%7D.
- 5. United States Census Bureau. Accessed October 19, 2023. https://www.census.gov/data-tools/demo/saipe/#/.
- 6. Demographic trends and Health Outcomes in the U.S. Kaggle. Accessed October 19, 2023. https://www.kaggle.com/datasets/thedevastator/demographic-trends-and-health-outcomes-in-the-u?select=SUMMARYMEASURESOFHEALTH.csv.
- 7. Table: Health insurance coverage type by poverty level: State Health Access Data Assistance Center. Table | Health Insurance Coverage Type by Poverty Level | State Health Access Data Assistance Center. Accessed October 19, 2023. <a href="https://statehealthcompare.shadac.org/table/23/health-insurance-coverage-type-by-poverty-level#1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52/5,4,1,10,86,9,8,6,35,36,32,37,38,33,34/37/45,46 .
- Cho HE, Wang L, Chen J-S, Liu M, Kuo C-F, Chung KC. Investigating the causal effect
 of socioeconomic status on quality of care under a universal health insurance system A
 marginal structural model approach BMC Health Services Research. BioMed Central.
 December 23, 2019. Accessed October 19, 2023.
 https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-019-4793-7.
- 9. Social Determinants of Health. Social Determinants of Health Healthy People 2030. Accessed October 19, 2023. https://health.gov/healthypeople/priority-areas/social-determinants-health.
- 10. HP-2022-12 addressing Social Determinants of Health: Examples of ... aspe. Accessed October 19, 2023.

 https://www.aspe.hhs.gov/sites/default/files/documents/e2b650cd64cf84aae8ff0fae7474af82/SDOH-Evidence-Review.pdf.