# **DATA ANALYSIS**

#### 1. Data Overview:

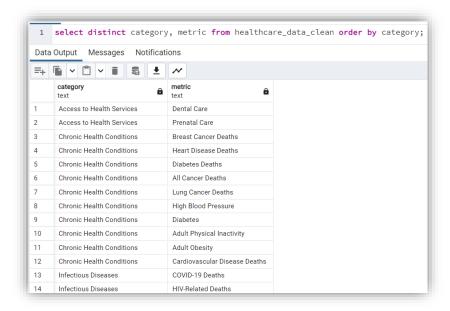
Number of unique metrics in the dataset: 30



Number of unique categories in the dataset: 7



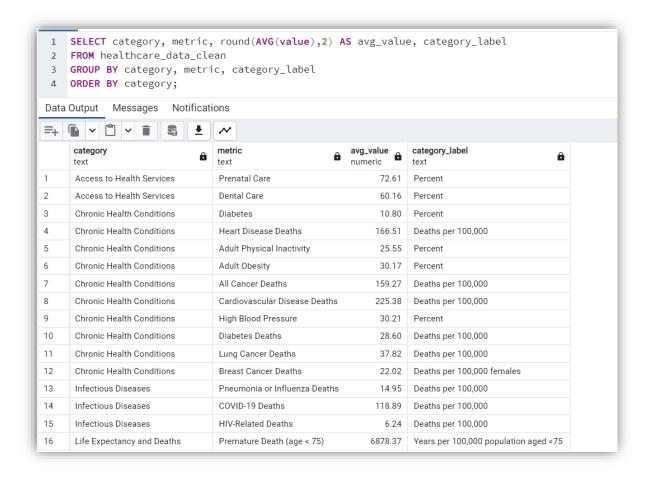
➤ List of unique categories and their corresponding metrics.



Count of metrics per category.

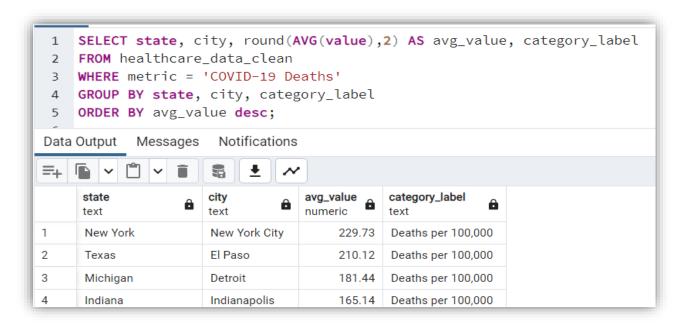


2. Average values for health metrics by grouping data based on metric.

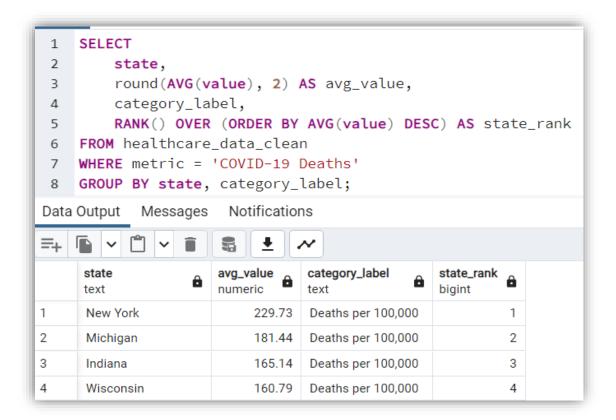


### 3. COVID-19 Death Analysis:

Average COVID-19 deaths by city, sorted in descending order.



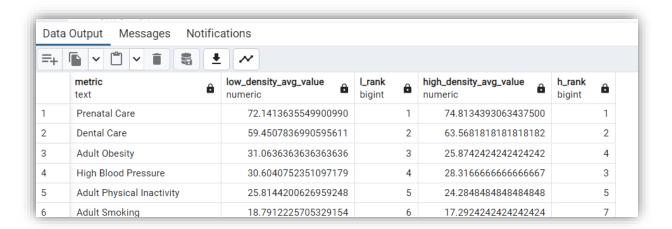
States ranked by average COVID-19 death rate.



The data retrieved by these queries helps in comparing and identifying COVID-19 deaths in different geographic areas. Additionally, we can apply similar analyses to different metrics to gain insights into how various health-related metrics vary across different geographic regions.

- 4. Rankings of health metrics based on normalized values.
  - Rankings based on average normalized values for lower and highest population density areas.

```
WITH ranked_metrics AS (
    SELECT
        AVG(normalized_value) AS low_density_avg_value
    FROM healthcare_data_clean
    WHERE population_density = 'lower'
    GROUP BY metric
),
ranked_metrics_2 AS (
    SELECT
        metric,
        AVG(normalized_value) AS high_density_avg_value
    FROM healthcare_data_clean
    WHERE population_density = 'highest'
    GROUP BY metric
)
SELECT
    r1.metric,
    low_density_avg_value,
    RANK() OVER (ORDER BY low_density_avg_value DESC) AS l_rank,
    high_density_avg_value,
    RANK() OVER (ORDER BY high_density_avg_value DESC) AS h_rank
FROM ranked_metrics AS r1
JOIN ranked_metrics_2 AS r2
ON r1.metric = r2.metric;
```



This SQL query calculates the average normalized values for health metrics in areas with lower and highest population densities. It ranks these metrics separately for lower and highest population density areas, providing insights into how health metrics vary based on population density.

Similar analyses can be performed for smaller and larger population areas, as well as for above and below poverty levels using the provided codes.

➤ Rankings based on average normalized values for health metrics in smaller and larger population areas.

```
WITH ranked_metrics AS (
    SELECT
        metric,
       AVG(normalized_value) AS smaller_population_value
    FROM healthcare_data_clean
    WHERE population = 'Smaller (<1.3 million)'
    GROUP BY metric
),
ranked_metrics_2 AS (
    SELECT
        metric,
        AVG(normalized_value) AS large_population_value
    FROM healthcare_data_clean
    WHERE population = 'Largest (>1.3 million)'
    GROUP BY metric
)
SELECT
    r1.metric,
    smaller_population_value,
    RANK() OVER (ORDER BY smaller_population_value DESC) AS l_rank,
    large_population_value,
    RANK() OVER (ORDER BY large_population_value DESC) AS h_rank
FROM ranked_metrics AS r1
JOIN ranked_metrics_2 AS r2
ON r1.metric = r2.metric;
```

| Data Output Messages Notifications |   |                                  |               |                                |                  |  |  |  |
|------------------------------------|---|----------------------------------|---------------|--------------------------------|------------------|--|--|--|
| '                                  | metric text                             | smaller_population_value numeric | I_rank bigint | large_population_value numeric | h_rank<br>bigint |  |  |  |
| 1                                  | Prenatal Care                           | 72.7381739941098901              | 1             | 72.2273592805957447            | 1                |  |  |  |
| 2                                  | Dental Care                             | 60.8807692307692308              | 2             | 58.0646464646464646            | 2                |  |  |  |
| 3                                  | High Blood Pressure                     | 30.4566433566433566              | 3             | 29.5050505050505051            | 4                |  |  |  |
| 4                                  | Adult Obesity                           | 30.1800699300699301              | 4             | 30.1565656565656566            | 3                |  |  |  |
| 5                                  | Adult Physical Inactivity               | 25.0223776223776224              | 5             | 27.0828282828282828            | 5                |  |  |  |
| 6                                  | Adult Smoking                           | 18.9573426573426573              | 6             | 17.3121212121212121            | 7                |  |  |  |
|                                    | A 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 | 17.7/11000111000110              | _             | 177161616161616161             |                  |  |  |  |

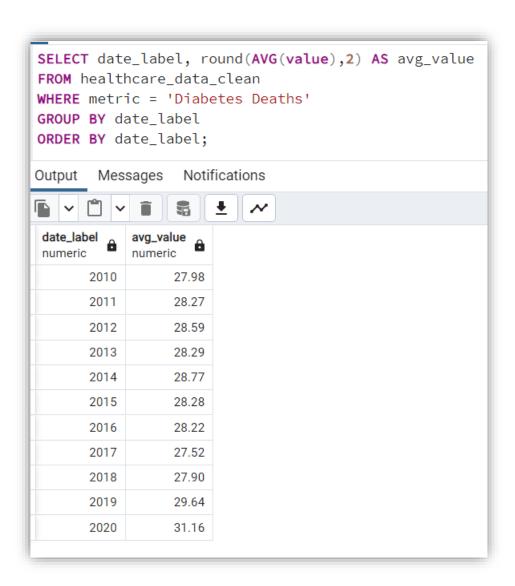
➤ Rankings based on average normalized values for health metrics in below poverty level and above poverty level areas.

```
WITH ranked_metrics AS (
    SELECT
        metric,
        AVG(normalized_value) AS below_poverty_level_value
    FROM healthcare_data_clean
   WHERE poverty = 'bpl'
   GROUP BY metric
),
ranked_metrics_2 AS (
    SELECT
        metric,
        AVG(normalized_value) AS above_poverty_level_value
    FROM healthcare_data_clean
   WHERE poverty = 'apl'
   GROUP BY metric
)
SELECT
    r1.metric,
    below_poverty_level_value,
    RANK() OVER (ORDER BY below_poverty_level_value DESC) AS l_rank,
    above_poverty_level_value,
    RANK() OVER (ORDER BY above_poverty_level_value DESC) AS h_rank
FROM ranked_metrics AS r1
JOIN ranked_metrics_2 AS r2
ON r1.metric = r2.metric;
```

| Data Output Messages Notifications |                           |                                   |                  |                                   |                  |  |  |
|------------------------------------|---------------------------|-----------------------------------|------------------|-----------------------------------|------------------|--|--|
| =+                                 |                           | . ~                               |                  |                                   |                  |  |  |
|                                    | metric text               | below_poverty_level_value numeric | I_rank<br>bigint | above_poverty_level_value numeric | h_rank<br>bigint |  |  |
| 1                                  | Prenatal Care             | 74.2142216375420875               | 1                | 65.7895635202571429               | 1                |  |  |
| 2                                  | Dental Care               | 61.6084415584415584               | 2                | 54.3493506493506494               | 2                |  |  |
| 3                                  | High Blood Pressure       | 28.4915584415584416               | 3                | 37.0935064935064935               | 4                |  |  |
| 4                                  | Adult Obesity             | 28.4233766233766234               | 4                | 37.1766233766233766               | 3                |  |  |
| 5                                  | Adult Physical Inactivity | 24.2756493506493506               | 5                | 30.6584415584415584               | 5                |  |  |
| 6                                  | Adult Binge Drinking      | 18.1775974025974026               | 6                | 16.0363636363636364               | 8                |  |  |

### 5. Year wise trend analysis

calculates the average values of diabetes deaths for various years.

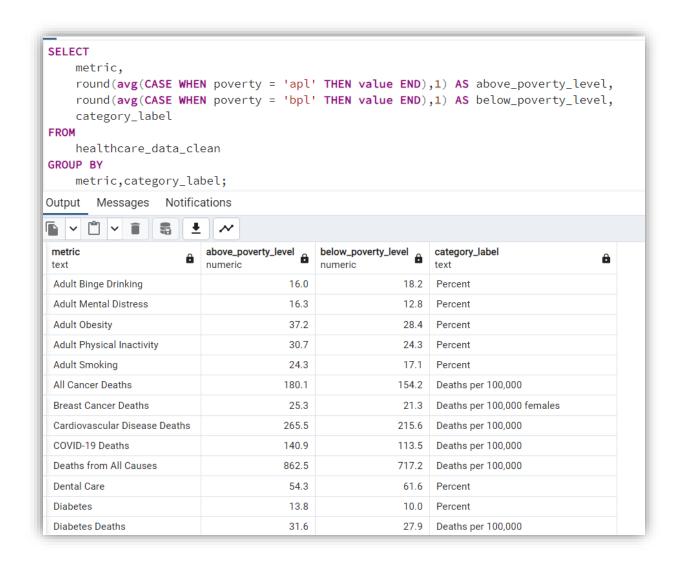


This query helps us see how the number of deaths related to diabetes has changed over the years. It calculates the average number of such deaths for each year, giving us insights into trends and variations. This information is important for understanding the impact of diabetes over time.

Similarly, we can perform this type of analysis for other health metrics such as COVID-19 deaths, adult obesity, and more to gain a comprehensive understanding of how various health factors have evolved over time.

### 6. Poverty Level Impact Analysis:

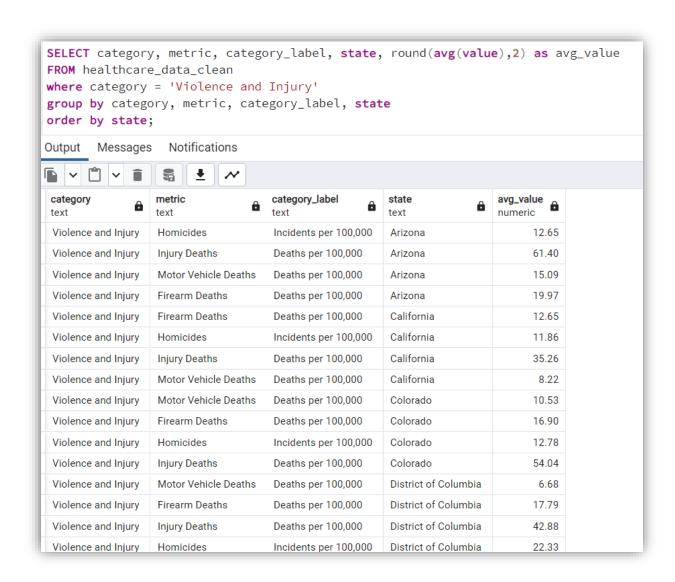
Calculate and compare the average values of health metrics for above and below poverty levels within various metrics.



This query calculates and compares the average values of health metrics, considering the above and below poverty levels. It provides insights into how the average values of these health metrics differ in areas with varying poverty levels, grouped by metric. This analysis helps identify the impact of poverty on health metrics across different categories, offering a broader perspective on the relationship between socioeconomic factors and health outcomes.

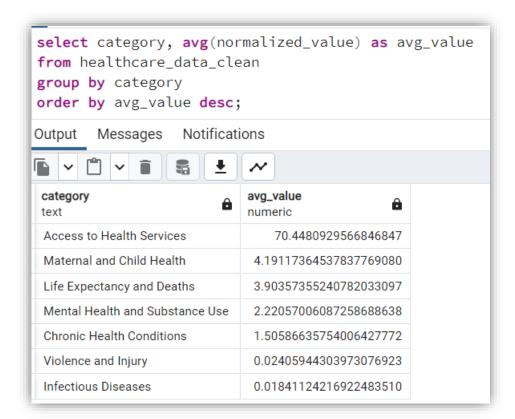
#### 7. State-wise Metric Value Analysis:

calculates the state-wise average values of all metrics within a specific category.



The query calculates the state-wise average values of all metrics within a specific category and provides the results grouped by the individual metrics, providing insights into how these metrics vary across different states. It allows for comparative analysis within the chosen category.

## 8. Category Average Analysis:



This query calculates the average normalized values within each category and presents them in descending order, offering an overview of which categories have higher average values. It helps identify categories with elevated health-related metrics.

\_\_\_\_\_

This analysis file provides an overview of the dataset, explores the relationships between various health metrics, and ranks health metrics based on normalized values in different contexts. The findings presented here are essential for further research and decision-making.

Please note that the outputs presented here are partial samples. For the full results, it is recommended to execute the queries locally to verify and access the complete findings.