

DATA ANALYSIS

1. Data Overview:

- Number of unique metrics in the dataset: 30

The screenshot shows a SQL query editor with the following query: `1 SELECT count(distinct metric) FROM healthcare_data_clean;`. Below the query editor, there are tabs for 'Data Output', 'Messages', and 'Notifications'. The 'Data Output' tab is active, displaying a table with the results of the query. The table has two columns: 'count' (type: bigint) and a lock icon. The first row shows the count as 30.

| | count bigint | |
|---|-----------------|--|
| 1 | 30 | |

- Number of unique categories in the dataset: 7

The screenshot shows a SQL query editor with the following query: `1 SELECT count(distinct category) FROM healthcare_data_clean;`. Below the query editor, there are tabs for 'Data Output', 'Messages', and 'Notifications'. The 'Data Output' tab is active, displaying a table with the results of the query. The table has two columns: 'count' (type: bigint) and a lock icon. The first row shows the count as 7.

| | count bigint | |
|---|-----------------|--|
| 1 | 7 | |

- List of unique categories and their corresponding metrics.

The screenshot shows a SQL query editor with the following query: `1 select distinct category, metric from healthcare_data_clean order by category;`. Below the query editor, there are tabs for 'Data Output', 'Messages', and 'Notifications'. The 'Data Output' tab is active, displaying a table with the results of the query. The table has two columns: 'category' (type: text) and 'metric' (type: text). The results are ordered by category.

| | category text | metric text |
|----|---------------------------|-------------------------------|
| 1 | Access to Health Services | Dental Care |
| 2 | Access to Health Services | Prenatal Care |
| 3 | Chronic Health Conditions | Breast Cancer Deaths |
| 4 | Chronic Health Conditions | Heart Disease Deaths |
| 5 | Chronic Health Conditions | Diabetes Deaths |
| 6 | Chronic Health Conditions | All Cancer Deaths |
| 7 | Chronic Health Conditions | Lung Cancer Deaths |
| 8 | Chronic Health Conditions | High Blood Pressure |
| 9 | Chronic Health Conditions | Diabetes |
| 10 | Chronic Health Conditions | Adult Physical Inactivity |
| 11 | Chronic Health Conditions | Adult Obesity |
| 12 | Chronic Health Conditions | Cardiovascular Disease Deaths |
| 13 | Infectious Diseases | COVID-19 Deaths |
| 14 | Infectious Diseases | HIV-Related Deaths |

➤ Count of metrics per category.

```

1 SELECT category,count(distinct metric)
2 FROM healthcare_data_clean
3 group by category
4 order by category;

```

| | category text | count bigint |
|---|---------------------------------|-----------------|
| 1 | Access to Health Services | 2 |
| 2 | Chronic Health Conditions | 10 |
| 3 | Infectious Diseases | 3 |
| 4 | Life Expectancy and Deaths | 2 |
| 5 | Maternal and Child Health | 4 |
| 6 | Mental Health and Substance Use | 5 |
| 7 | Violence and Injury | 4 |

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

```

1 SELECT category, metric, round(AVG(value),2) AS avg_value, category_label
2 FROM healthcare_data_clean
3 GROUP BY category, metric, category_label
4 ORDER BY category;

```

| | category text | metric text | avg_value numeric | category_label text |
|----|----------------------------|-------------------------------|----------------------|---------------------------------------|
| 1 | Access to Health Services | Prenatal Care | 72.61 | Percent |
| 2 | Access to Health Services | Dental Care | 60.16 | Percent |
| 3 | Chronic Health Conditions | Diabetes | 10.80 | Percent |
| 4 | Chronic Health Conditions | Heart Disease Deaths | 166.51 | Deaths per 100,000 |
| 5 | Chronic Health Conditions | Adult Physical Inactivity | 25.55 | Percent |
| 6 | Chronic Health Conditions | Adult Obesity | 30.17 | Percent |
| 7 | Chronic Health Conditions | All Cancer Deaths | 159.27 | Deaths per 100,000 |
| 8 | Chronic Health Conditions | Cardiovascular Disease Deaths | 225.38 | Deaths per 100,000 |
| 9 | Chronic Health Conditions | High Blood Pressure | 30.21 | Percent |
| 10 | Chronic Health Conditions | Diabetes Deaths | 28.60 | Deaths per 100,000 |
| 11 | Chronic Health Conditions | Lung Cancer Deaths | 37.82 | Deaths per 100,000 |
| 12 | Chronic Health Conditions | Breast Cancer Deaths | 22.02 | Deaths per 100,000 females |
| 13 | Infectious Diseases | Pneumonia or Influenza Deaths | 14.95 | Deaths per 100,000 |
| 14 | Infectious Diseases | COVID-19 Deaths | 118.89 | Deaths per 100,000 |
| 15 | Infectious Diseases | HIV-Related Deaths | 6.24 | Deaths per 100,000 |
| 16 | Life Expectancy and Deaths | Premature Death (age < 75) | 6878.37 | Years per 100,000 population aged <75 |

3. COVID-19 Death Analysis:

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

```
1 SELECT state, city, round(AVG(value),2) AS avg_value, category_label
2 FROM healthcare_data_clean
3 WHERE metric = 'COVID-19 Deaths'
4 GROUP BY state, city, category_label
5 ORDER BY avg_value desc;
```

| | state text | city text | avg_value numeric | category_label text |
|---|---------------|---------------|----------------------|------------------------|
| 1 | New York | New York City | 229.73 | Deaths per 100,000 |
| 2 | Texas | El Paso | 210.12 | Deaths per 100,000 |
| 3 | Michigan | Detroit | 181.44 | Deaths per 100,000 |
| 4 | Indiana | Indianapolis | 165.14 | Deaths per 100,000 |

- States ranked by average COVID-19 death rate.

```
1 SELECT
2     state,
3     round(AVG(value), 2) AS avg_value,
4     category_label,
5     RANK() OVER (ORDER BY AVG(value) DESC) AS state_rank
6 FROM healthcare_data_clean
7 WHERE metric = 'COVID-19 Deaths'
8 GROUP BY state, category_label;
```

| | state text | avg_value numeric | category_label text | state_rank bigint |
|---|---------------|----------------------|------------------------|----------------------|
| 1 | New York | 229.73 | Deaths per 100,000 | 1 |
| 2 | Michigan | 181.44 | Deaths per 100,000 | 2 |
| 3 | Indiana | 165.14 | Deaths per 100,000 | 3 |
| 4 | Wisconsin | 160.79 | Deaths per 100,000 | 4 |

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  
        metric,  
        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;
```

| Data Output Messages Notifications | | | | | | | |
|------------------------------------|---------------------------|----------------------------------|------------------|-----------------------------------|------------------|--|--|
| | metric text | low_density_avg_value numeric | l_rank bigint | high_density_avg_value numeric | h_rank bigint | | |
| 1 | Prenatal Care | 72.1413635549900990 | 1 | 74.8134393063437500 | 1 | | |
| 2 | Dental Care | 59.4507836990595611 | 2 | 63.5681818181818182 | 2 | | |
| 3 | Adult Obesity | 31.0636363636363636 | 3 | 25.8742424242424242 | 4 | | |
| 4 | High Blood Pressure | 30.6040752351097179 | 4 | 28.3166666666666667 | 3 | | |
| 5 | Adult Physical Inactivity | 25.8144200626959248 | 5 | 24.2848484848484848 | 5 | | |
| 6 | Adult Smoking | 18.7912225705329154 | 6 | 17.2924242424242424 | 7 | | |

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 | l_rank bigint | large_population_value numeric | h_rank 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 | |
| 7 | Adult Diabetes | 17.361188811888118 | 7 | 17.3121212121212121 | 6 | |

- 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 | l_rank bigint | above_poverty_level_value numeric | h_rank 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.

```
SELECT date_label, round(AVG(value),2) AS avg_value
FROM healthcare_data_clean
WHERE metric = 'Diabetes Deaths'
GROUP BY date_label
ORDER BY date_label;
```

| Output | | Messages | Notifications |
|---|---|---|---|
|  |  |  |  |
|  |  |  | |
| date_label | avg_value | | |
| numeric | numeric | | |
| 2010 | 27.98 | | |
| 2011 | 28.27 | | |
| 2012 | 28.59 | | |
| 2013 | 28.29 | | |
| 2014 | 28.77 | | |
| 2015 | 28.28 | | |
| 2016 | 28.22 | | |
| 2017 | 27.52 | | |
| 2018 | 27.90 | | |
| 2019 | 29.64 | | |
| 2020 | 31.16 | | |

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.

```
SELECT
    metric,
    round(avg(CASE WHEN poverty = 'apl' THEN value END),1) AS above_poverty_level,
    round(avg(CASE WHEN poverty = 'bpl' THEN value END),1) AS below_poverty_level,
    category_label
FROM
    healthcare_data_clean
GROUP BY
    metric,category_label;
```

Output Messages Notifications

| metric text | above_poverty_level numeric | below_poverty_level numeric | category_label text |
|-------------------------------|--------------------------------|--------------------------------|----------------------------|
| Adult Binge Drinking | 16.0 | 18.2 | Percent |
| Adult Mental Distress | 16.3 | 12.8 | Percent |
| Adult Obesity | 37.2 | 28.4 | Percent |
| Adult Physical Inactivity | 30.7 | 24.3 | Percent |
| Adult Smoking | 24.3 | 17.1 | Percent |
| All Cancer Deaths | 180.1 | 154.2 | Deaths per 100,000 |
| Breast Cancer Deaths | 25.3 | 21.3 | Deaths per 100,000 females |
| Cardiovascular Disease Deaths | 265.5 | 215.6 | Deaths per 100,000 |
| COVID-19 Deaths | 140.9 | 113.5 | Deaths per 100,000 |
| Deaths from All Causes | 862.5 | 717.2 | Deaths per 100,000 |
| Dental Care | 54.3 | 61.6 | Percent |
| Diabetes | 13.8 | 10.0 | Percent |
| Diabetes Deaths | 31.6 | 27.9 | Deaths per 100,000 |

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.

```
SELECT category, metric, category_label, state, round(avg(value),2) as avg_value
FROM healthcare_data_clean
where category = 'Violence and Injury'
group by category, metric, category_label, state
order by state;
```

Output Messages Notifications

| category text | metric text | category_label text | state text | avg_value numeric |
|---------------------|----------------------|------------------------|----------------------|----------------------|
| Violence and Injury | Homicides | Incidents per 100,000 | Arizona | 12.65 |
| Violence and Injury | Injury Deaths | Deaths per 100,000 | Arizona | 61.40 |
| Violence and Injury | Motor Vehicle Deaths | Deaths per 100,000 | Arizona | 15.09 |
| Violence and Injury | Firearm Deaths | Deaths per 100,000 | Arizona | 19.97 |
| Violence and Injury | Firearm Deaths | Deaths per 100,000 | California | 12.65 |
| Violence and Injury | Homicides | Incidents per 100,000 | California | 11.86 |
| Violence and Injury | Injury Deaths | Deaths per 100,000 | California | 35.26 |
| Violence and Injury | Motor Vehicle Deaths | Deaths per 100,000 | California | 8.22 |
| Violence and Injury | Motor Vehicle Deaths | Deaths per 100,000 | Colorado | 10.53 |
| Violence and Injury | Firearm Deaths | Deaths per 100,000 | Colorado | 16.90 |
| Violence and Injury | Homicides | Incidents per 100,000 | Colorado | 12.78 |
| Violence and Injury | Injury Deaths | Deaths per 100,000 | Colorado | 54.04 |
| Violence and Injury | Motor Vehicle Deaths | Deaths per 100,000 | District of Columbia | 6.68 |
| Violence and Injury | Firearm Deaths | Deaths per 100,000 | District of Columbia | 17.79 |
| Violence and Injury | Injury Deaths | Deaths per 100,000 | District of Columbia | 42.88 |
| Violence and Injury | Homicides | Incidents per 100,000 | District of Columbia | 22.33 |

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:

```
select category, avg(normalized_value) as avg_value
from healthcare_data_clean
group by category
order by avg_value desc;
```

OutputMessagesNotifications

| category | avg_value |
|---------------------------------|------------------------|
| text | numeric |
| Access to Health Services | 70.4480929566846847 |
| Maternal and Child Health | 4.19117364537837769080 |
| Life Expectancy and Deaths | 3.90357355240782033097 |
| Mental Health and Substance Use | 2.22057006087258688638 |
| Chronic Health Conditions | 1.50586635754006427772 |
| Violence and Injury | 0.02405944303973076923 |
| Infectious Diseases | 0.01841124216922483510 |

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.