Purpose:

This document provides background on the data used to build the COVID-19 Tracking Metrics presented in this document/repository. It also presents the queries used for each metric along with the output of each query.

Data:

- Source: **https://covidtracking.com/data/download/all-states-history.csv**
- Data Last Refreshed: 2020-02-03 11:27pm MST
- Caveats about the data overall:
  - All dates and times are in US eastern time
  - The data includes data points from US states as well as US territories
    - Each metric will call out if both US states and US territories were used or not
- Caveats about the queries:
  - Please note that the queries shown below are built for the metrics as stated. If you were to run these queries on the data attached to this repository, then you would not get the same results because the last date listed in the data is 2020-02-03.

Database/SQL Used:

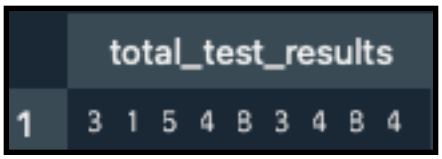
SQLite v3.31.0

Metric 1: Total Tests Performed in the United States

This metric provides the user with the total tests performed up until yesterday in the United States. The expected output for this metric is one number representing the total tests performed up until the previous day.

NOTE: The data used in calculating this metric uses data from the US States as well as US territories. As stated from the source, it should be noted that there is variation in how test reporting methods are conducted in each state and territory. So, at best, we are to take this metric as a best estimate.

Output:



	total_test_results
1	3 1 5 4 8 3 4 8 4

Query:

```
SELECT
      SUM(totalTestResults) AS total_test_results
FROM covid
WHERE date = date('now', '-1 day');
```

Metric 2: 7-Day Rolling Average of New Cases per Day

This metric provides the user with a 7-day rolling average of new cases per day for the last 30 days. The expected output for this metric is a date field and a numerical field. The numerical field represents the 7-Day Rolling Average.

NOTE: This metric uses data from US states and US territories.

Output:

	date	rolling_avg_7day
1	2021-01-04	179571.00
2	2021-01-05	199323.00
3	2021-01-06	216276.67
4	2021-01-07	230218.25
5	2021-01-08	243198.80
6	2021-01-09	247560.33
7	2021-01-10	244872.57
8	2021-01-11	247111.14
9	2021-01-12	246960.43
10	2021-01-13	243290.00
11	2021-01-14	236657.57
12	2021-01-15	229705.14
13	2021-01-16	222379.00
14	2021-01-17	216448.43
15	2021-01-18	210096.57
16	2021-01-19	199897.43
17	2021-01-20	194856.29
18	2021-01-21	189236.14
19	2021-01-22	181302.00
20	2021-01-23	175449.57
21	2021-01-24	169309.00
22	2021-01-25	166832.14
23	2021-01-26	166806.71
24	2021-01-27	161775.57
25	2021-01-28	157637.14
26	2021-01-29	154236.57
27	2021-01-30	150194.57
28	2021-01-31	146642.14
29	2021-02-01	144748.71
30	2021-02-02	140629.86

Query:

```
WITH dateagg(date, posInc) AS (SELECT date, sum(positiveIncrease) AS
posInc
      FROM covid
      GROUP BY date
      ORDER BY date)
SELECT
      date,
      printf("%.4f", AVG(posInc) OVER(ORDER BY date ROWS BETWEEN 6
PRECEDING AND CURRENT ROW)) AS rolling_avg_7day
FROM dateagg
WHERE date BETWEEN date('now', '-30 day') AND date('now', '-1 day');
```

Metric 3: Top 10 States for Highest Positivity Rate

This metric provides the user with the 10 states that have the highest positivity rate for tests performed in the last 30 days. The positivity rate is defined by positive test / tests performed. The expected output of this metric is a state identifier field and a numeric field. The numeric field represents the positivity rate for the particular state shown in decimal format.

NOTE: This data uses only US state data since the metric specifically calls out “Top 10 States”. US territories are excluded.

Output:

	state	positive_rate
1	ID	0.2696
2	SD	0.2688
3	KS	0.2349
4	AL	0.2170
5	IA	0.2136
6	MS	0.1724
7	GA	0.1413
8	TX	0.1372
9	UT	0.1328
10	OK	0.1220

Query:

```
SELECT
      state,
      printf("%.4f", (CAST(positive AS REAL) / CAST(totalTestResults AS
REAL))) AS positive_rate
FROM covid
WHERE date BETWEEN date('now', '-35 day') AND date('now', '-6 day')
      AND state NOT IN ('AS', 'DC', 'FM', 'GU', 'MH', 'MP', 'PW', 'PR',
'VI')
GROUP BY state
ORDER BY positive_rate DESC
LIMIT 10;
```