

# DataEng S25: Project Assignment 3 Submission

**Team Name** - Data Foundry

**Team Members:**

Kiranmayi Thanikonda

Jayita Banerjee

Navya Sri Ambati

**Step-1:** Construct a table showing each day for which your pipeline successfully, automatically processed one complete day's worth of sensor readings for breadcrumb data only. The table should look like this:

Date	Day of Week	# breadcrumb records	# rows added to BreadCrumb table	# rows added to Trip table
05/12/2025	Monday	651224	650768	135
05/13/2025	Tuesday	643140	642659	133
05/14/2025	Wednesday	685645	684335	128
05/15/2025	Thursday	658948	657635	131
05/16/2025	Friday	629179	628558	137
05/17/2025	Saturday	431481	430772	88
05/18/2025	Sunday	445301	444769	102
05/19/2025	Monday	502788	501758	1039
05/20/2025	Tuesday	645513	644180	1371
05/21/2025	Wednesday	648141	647649	1340
05/22/2025	Thursday	659430	658594	1308
05/23/2025	Friday	633104	632667	1293
05/24/2025	Saturday	446040	445780	871
05/25/2025	Sunday	510846	509768	982
05/26/2025	Monday	699246	698440	1428

**Step-2:** Perform These Queries on Your DB and Report the Results

Query #1:

```
select tstamp::date, count(*)  
from breadcrumb  
group by tstamp::date  
order by tstamp::date ASC;
```

```
postgres=# select tstamp::date, count(*)  
from breadcrumb  
group by tstamp::date  
order by tstamp::date ASC;  
      tstamp      | count  
-----+-----  
2022-12-19 | 642929  
2022-12-20 | 659921  
2022-12-21 | 685035  
2022-12-22 | 679675  
2022-12-23 | 479900  
2022-12-24 | 499567  
2022-12-25 | 489186  
2022-12-26 | 511496  
2022-12-27 | 685128  
2022-12-28 | 705908  
2022-12-29 | 628167  
2022-12-30 | 716869  
2022-12-31 | 495559  
2023-01-01 | 483350  
2023-01-02 | 505935  
2023-01-03 | 638163  
2023-01-04 | 696300  
2023-01-05 | 677374  
2023-01-06 | 650748  
2023-01-07 | 499391  
2023-01-08 | 453120  
2023-01-09 | 648354  
2023-01-10 | 642705  
2023-01-11 | 684370  
2023-01-12 | 656734  
2023-01-13 | 627963  
2023-01-14 | 435618  
2023-01-15 | 444983  
2023-01-16 | 496836  
2023-01-17 | 649747  
2023-01-18 | 640100  
2023-01-19 | 659273  
2023-01-20 | 630202  
2023-01-21 | 19664  
(34 rows)
```

## Query #2:

```
SELECT
    b.timestamp::date AS trip_date,
    COUNT(DISTINCT t.vehicle_id) AS num_vehicles,
    COUNT(DISTINCT t.trip_id) AS num_trips,
    COUNT(DISTINCT t.route_id) AS num_routes,
    COUNT(*) AS num_breadcrumbs
FROM
    Breadcrumb b
JOIN
    Trip t ON b.trip_id = t.trip_id
WHERE
    b.timestamp::date > '2023-01-08'
GROUP BY
    b.timestamp::date
ORDER BY
    trip_date ASC;
```

```
postgres=# SELECT
    b.timestamp::date AS trip_date,
    COUNT(DISTINCT t.vehicle_id) AS num_vehicles,
    COUNT(DISTINCT t.trip_id) AS num_trips,
    COUNT(DISTINCT t.route_id) AS num_routes,
    COUNT(*) AS num_breadcrumbs
FROM
    Breadcrumb b
JOIN
    Trip t ON b.trip_id = t.trip_id
WHERE
    b.timestamp::date > '2023-01-08'
GROUP BY
    b.timestamp::date
ORDER BY
    trip_date ASC;


| trip_date  | num_vehicles | num_trips | num_routes | num_breadcrumbs |
|------------|--------------|-----------|------------|-----------------|
| 2023-01-09 | 145          | 1680      | 54         | 648354          |
| 2023-01-10 | 148          | 1725      | 54         | 642705          |
| 2023-01-11 | 150          | 1764      | 56         | 684370          |
| 2023-01-12 | 149          | 1668      | 58         | 656734          |
| 2023-01-13 | 150          | 1659      | 53         | 627963          |
| 2023-01-14 | 109          | 157       | 34         | 61876           |
| 2023-01-15 | 102          | 102       | 35         | 50448           |
| 2023-01-16 | 115          | 1018      | 41         | 446483          |
| 2023-01-17 | 148          | 1399      | 66         | 581837          |
| 2023-01-18 | 148          | 1346      | 65         | 574490          |
| 2023-01-19 | 146          | 1330      | 68         | 591463          |
| 2023-01-20 | 140          | 1310      | 64         | 565767          |
| 2023-01-21 | 122          | 902       | 40         | 412724          |
| 2023-01-22 | 132          | 998       | 38         | 463762          |
| 2023-01-23 | 143          | 1433      | 64         | 625262          |
| 2023-01-24 | 27           | 46        | 17         | 13842           |


(16 rows)
```

## Step-3: Visualizations

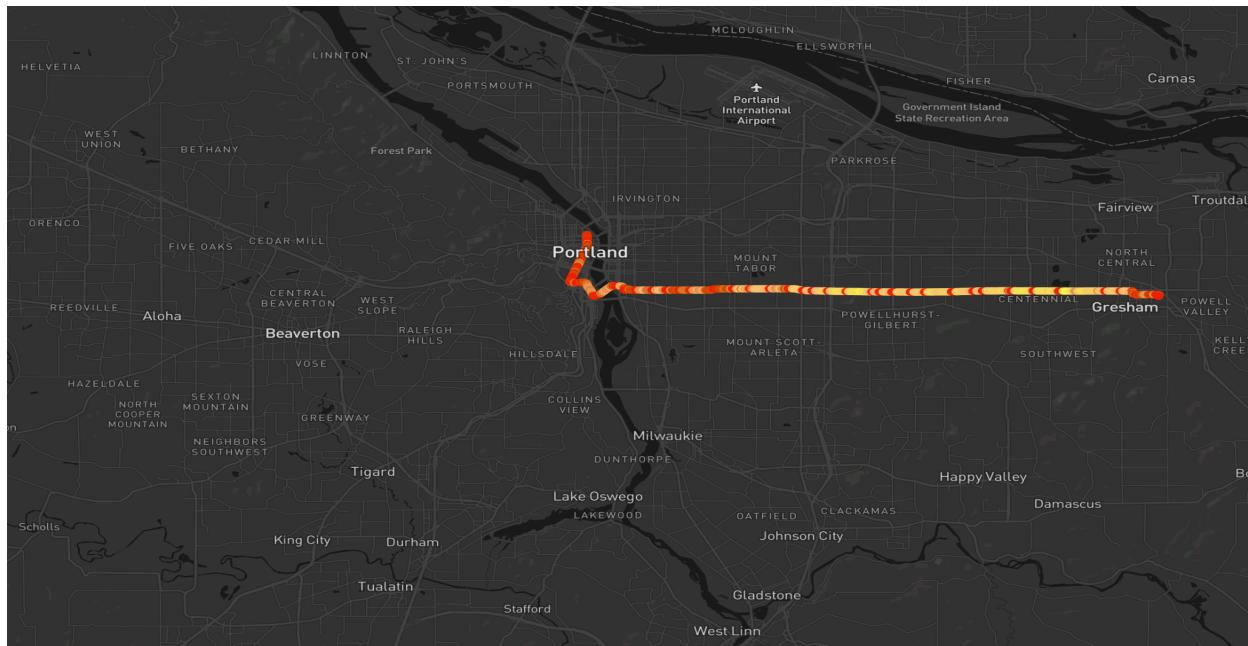
**Visualization 1:** A visualization of speeds for a single trip for any bus that crossed the Tilikum Crossing Bridge. You choose the day, time and route for your selected trip. To find a trip that traverses this bridge, consider finding a trip that includes breadcrumbs within the rectangle bounded by (45.505621, -122.668632), and (45.504416, -122.665738).

Query: SELECT DISTINCT trip\_id

```
FROM breadcrumb  
WHERE latitude BETWEEN 45.504416 AND 45.505621  
AND longitude BETWEEN -122.668632 AND -122.665738  
AND tstamp::date > '2023-01-08'  
LIMIT 10;
```

```
SELECT longitude, latitude, speed  
FROM breadcrumb  
WHERE trip_id = 231117393;
```

**Answer:** This visualization displays bus speeds for Trip ID 234408536, which crossed the Tilikum Crossing Bridge. Breadcrumbs were filtered using a bounding box around the bridge area. The result shows how vehicle speed varied as it crossed the bridge.



**Visualization 2:** Any single trip on route 20 (you choose which trip) between the hours of 4pm and 6pm (you choose which date).

**Query:** SELECT DISTINCT b.trip\_id, t.route\_id, t.vehicle\_id, b.tstamp

```
FROM breadcrumb b
JOIN trip t ON b.trip_id = t.trip_id
WHERE t.route_id = 20
AND b.tstamp::date = '2023-01-14'
AND b.tstamp::time BETWEEN '16:00:00' AND '18:00:00'
ORDER BY b.tstamp;
```

```
SELECT longitude, latitude, speed
```

```
FROM breadcrumb
```

```
WHERE trip_id = 238592021
```

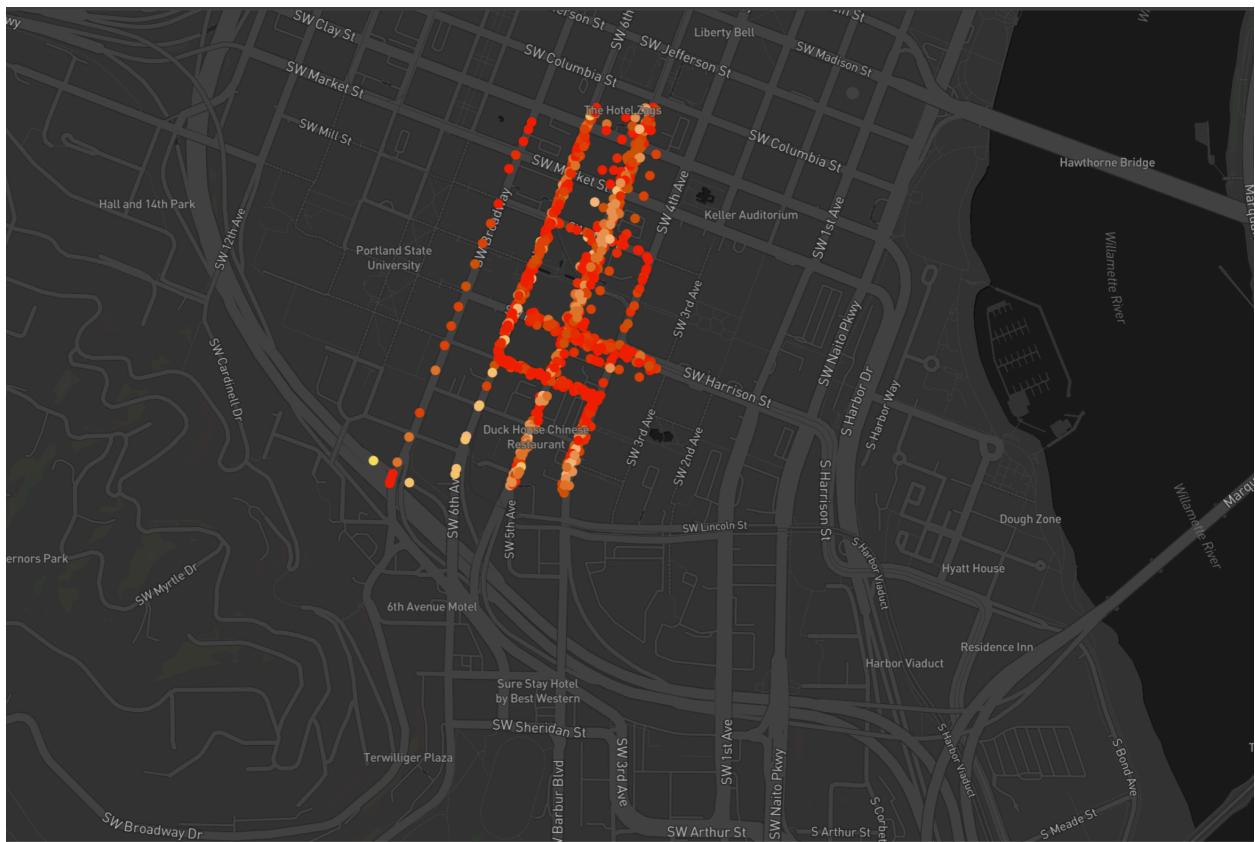
**Answer:** This visualization shows speeds for a single trip on Route 20 (Trip ID 238592021) taken between 4 PM and 6 PM on January 14, 2023. The map reveals typical evening traffic behavior on this route.



**Visualization 3:** All trips that traveled to and from PSU campus on any single weekday morning (you choose which date) between 9am and 11am.

```
Query: SELECT longitude, latitude, speed
FROM breadcrumb
WHERE latitude BETWEEN 45.5085 AND 45.5140
AND longitude BETWEEN -122.6860 AND -122.6800
AND tstamp::date = '2023-01-10'
AND tstamp::time BETWEEN '09:00:00' AND '11:00:00'
```

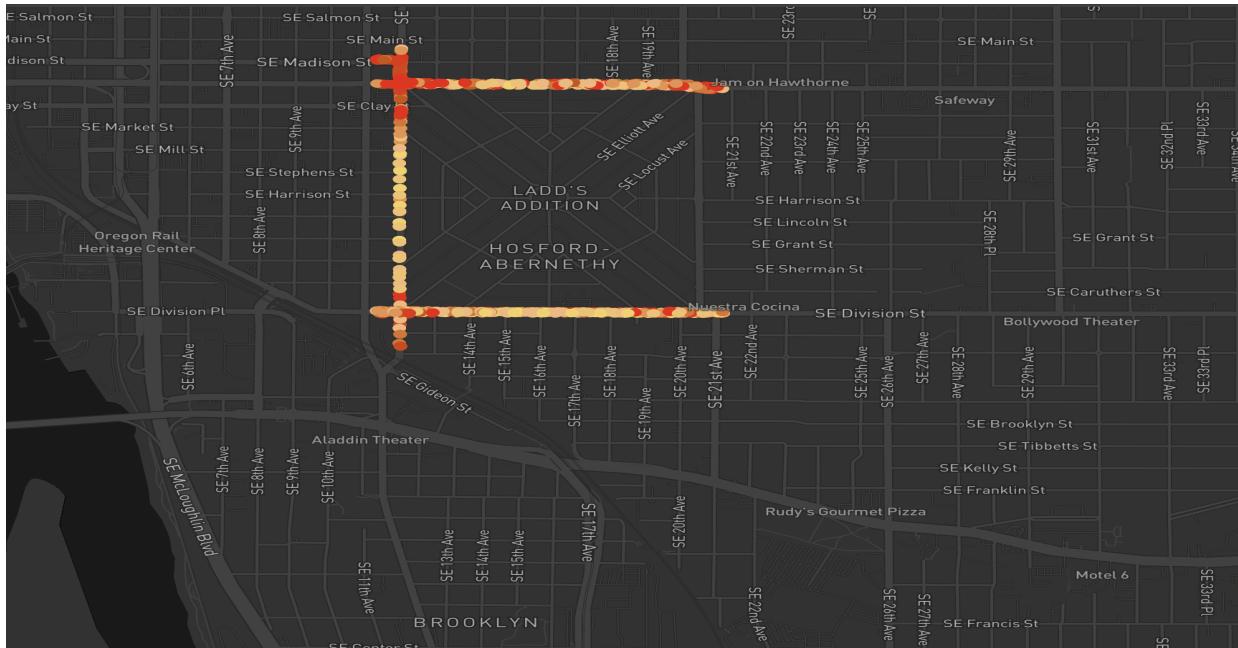
**Answer:** This visualization captures all trips that passed near PSU campus between 9:00 AM and 11:00 AM on January 10, 2023. The selected bounding box covers a dense downtown region, showing speeds through a high-traffic area.



**Visualization 4:** All trips that traveled within 1km of Ladd Circle Park (45.508537, -122.649434) on any selected morning before 11am (you choose which date).

```
SELECT longitude, latitude, speed  
FROM breadcrumb  
WHERE latitude BETWEEN 45.5035 AND 45.5135  
AND longitude BETWEEN -122.6544 AND -122.6444  
AND tstamp::date = '2023-01-10'  
AND tstamp::time < '11:00:00';
```

**Answer:** This visualization includes all bus activity near Ladd Circle Park on the morning of January 10, 2023. A bounding box around the park was used to isolate traffic before 11:00 AM.



**Visualization 5a:** Which trip recorded the highest average speed on a normal weekday after January 8, 2023 (selecting any one date)?

```
Query: SELECT trip_id, AVG(speed) AS avg_speed  
FROM breadcrumb  
WHERE tstamp::date = '2023-01-10'  
AND speed > 0  
GROUP BY trip_id  
ORDER BY avg_speed DESC
```

LIMIT 1;

```
SELECT longitude, latitude, speed  
FROM breadcrumb  
WHERE trip_id = 235573196;
```

**Answer:** This visualization shows the fastest trip by average speed on January 10, 2023. Trip ID 235573196 had an average speed of 26.68 mph, indicating efficient and uninterrupted travel.



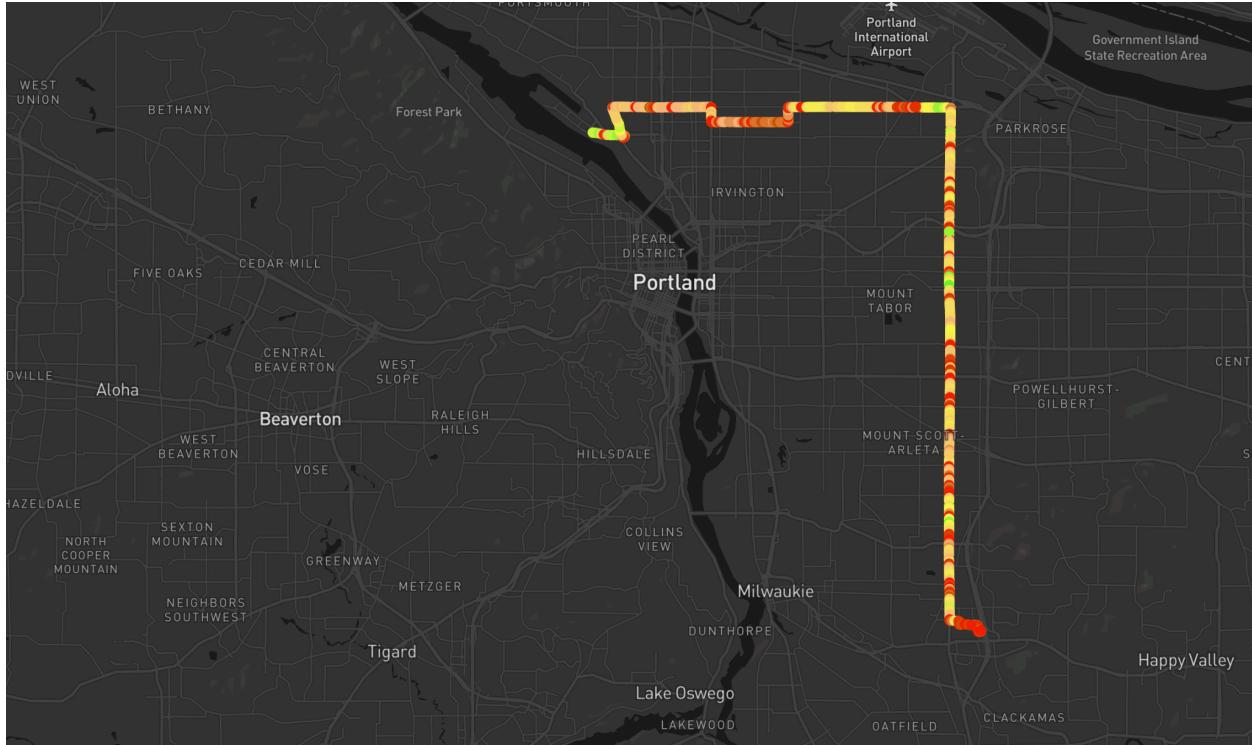
**Visualization 5b:** How do buses move across the Hillsboro city during the midday lull between the morning and evening rush hours?

**Query:**

```
SELECT DISTINCT trip_id  
FROM breadcrumb  
WHERE tstamp::date = '2023-01-12'  
AND tstamp::time BETWEEN '11:00:00' AND '13:00:00'  
ORDER BY trip_id  
LIMIT 1;
```

```
SELECT longitude, latitude, speed  
FROM breadcrumb  
WHERE trip_id = 236514460
```

**Answer:** This visualization highlights one trip ID 236514460 that occurred between 11:00 AM and 1:00 PM on January 12, 2023. It illustrates typical bus speed patterns during midday hours, a relatively low-traffic period.

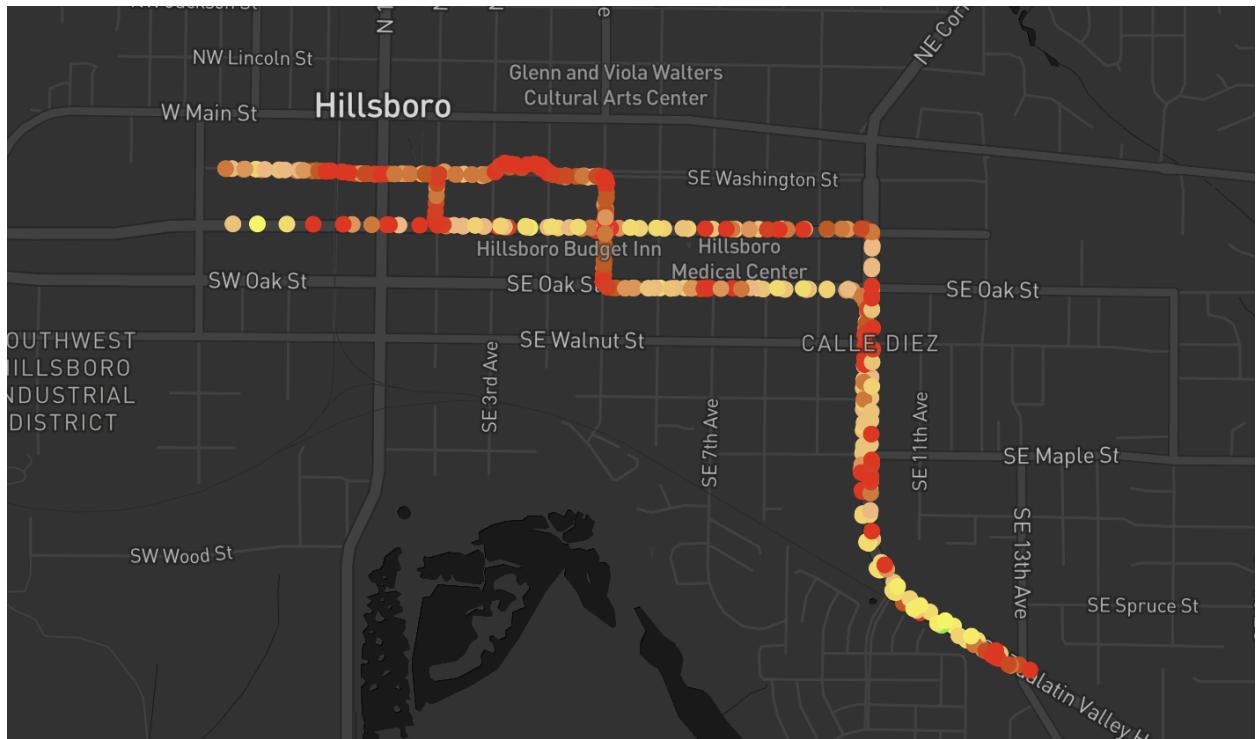


**Visualization 5c:** What does overall bus traffic look like in Hillsboro city during a typical evening period (6–8 PM)?

**Query:**

```
SELECT longitude, latitude, speed  
FROM breadcrumb  
WHERE tstamp::date = '2023-01-14'  
AND tstamp::time BETWEEN '18:00:00' AND '20:00:00'  
AND latitude BETWEEN 45.5100 AND 45.5400  
AND longitude BETWEEN -122.9950 AND -122.9350
```

**Answer:** This visualization displays all GPS points from trips in Hillsboro during the evening window on January 14, 2023. It highlights how traffic behaves toward the end of the day, with visible slowdowns and route-specific trends.



**Github Link:**

<https://github.com/kthanikonda/DataEngineering/tree/main/Project/ProjectAssignment3>