School of Information and Communication Technology

Griffith University

7821ICT

Green City Situation Awareness

User Manual - Volume

Proof of Concept Demonstrator for Sustainable Transport

18/10/2024 T2 2024

Industry Partner: City of Gold Coast

Client: Phillip Karfs

Team members:

Omkar Dhananjay Kulkarni (s5325810)

Vlesetty Vishal Kumar (s5314434)

Viet An Nguyen (s5321461)

Zhanrui Liao (s5290972)

Table of Content

Project Overview

This project involves analyzing vehicle volume data across the Gold Coast region, comparing data for July, August, and September of 2023 and 2024. Our objective is to understand changes in vehicle volume and detect patterns in private vehicle usage, particularly in response to external factors like public transport fare reductions.

Pre-requisites:

Prerequisites:

1. Open Snowflake: Login here
2. Install QGIS: Link to open/download **Steps:**

1. Access Snowflake using the provided credentials.
2. In QGIS, install required plugins (Quick Map Service).
3. Open QGIS and ensure the Quick Map Service layers for spatial data are activated.

Snowflake Analysis

In Snowflake:

The data is hosted in Snowflake, and the key tables used for analysis are:

- PAWS_DBT_RPT.MART_STREAMS_FCT_DETECTORSITE_AGG_DAILY (for vehicle volume data)
- PAWS_DBT_RPT.MART_STREAMS_DIM_DETECTORSITE_CURR (for site metadata like detector locations) **Steps to Load Data:**
- Run the SQL queries provided below to filter data from July 2023 to September 2024 and to focus on common links between the two years.

```
WITH Common_Links AS (
    -- Identify links that exist in both 2023 and 2024
    SELECT LINK_ID
    FROM PAWS_DBT_RPT.MART_STREAMS_FCT_DETECTORSITE_AGG_DAILY
    WHERE START_DATE_KEY LIKE '2023%' OR START_DATE_KEY LIKE '2024%'
    GROUP BY LINK_ID
    HAVING COUNT(DISTINCT SUBSTRING(START_DATE_KEY, 1, 4)) = 2
),
```

Analysis in Snowflake

1.**Vehicle VolumeAnalysis:**

- Calculate the average and total volume across multiple detectorsites for the months of July, August, and September of 2023 and 2024.

```
aggregated_volume AS (
    -- Aggregate data for the months of July, August, and September
    SELECT
        SUBSTRING(START_DATE_KEY, 1, 4) AS year,                -- Extract year
        CASE                                                    -- Convert month number to name
            WHEN SUBSTRING(START_DATE_KEY, 5, 2) = '07' THEN 'July'
            WHEN SUBSTRING(START_DATE_KEY, 5, 2) = '08' THEN 'August'
            WHEN SUBSTRING(START_DATE_KEY, 5, 2) = '09' THEN 'September'
        END AS month_name,                                      -- Return month name
        SUM(AGG_VOLUME) AS total_volume,                        -- Calculate total volume per month
        LINK_ID                                                 -- Include the LINK_ID for join
    FROM PAWS_DBT_RPT.MART_STREAMS_FCT_DETECTORSITE_AGG_DAILY
    WHERE
        (START_DATE_KEY BETWEEN '20230701' AND '20230930'       -- July to September 2023
        OR START_DATE_KEY BETWEEN '20240701' AND '20240930')    -- July to September 2024
        AND LINK_ID IN (SELECT LINK_ID FROM Common_Links)       -- Filter for common links only
    GROUP BY
        SUBSTRING(START_DATE_KEY, 1, 4),
        month_name,
        LINK_ID                                                -- Group by LINK_ID and month
)

```

1. **SELECT Clause**:
    - **agg.year** and **agg.month_name**: These columns indicate the year and month for which the data is aggregated. They allow grouping and identification of data trends over different periods.
    - **COALESCE(agg.total_volume, 0) AS total_volume**: This command ensures that any NULL values in the total_volume field are replaced with a zero. This prevents NULL values from skewing total volume calculations. ○ **geo.DETECTORSITE_NAME**: Retrieves the name of the detector site to provide additional context to the volume data.
    - **COALESCE(ST_X(geo.DETECTORSITE_GEOMETRY), 0) AS longitude** and **COALESCE(ST_Y(geo.DETECTORSITE_GEOMETRY), 0) AS latitude**: These commands extract the geographic coordinates (longitude and latitude) of each detector site. By handling NULL values with zero, we avoid issues with missing geographic information.
    - ○ **geo.DETECTORSITE_SUBURB_NAME** and **geo.DETECTORSITE_DIVISION_NAME**: These columns contain the suburb and division names for the detector sites. This information

is crucial for grouping data and performing spatial analysis at different levels.

2. **FROM Clause**:
    - **aggregated_volume_agg**: This alias references a previously created dataset that aggregates volume data based on common links for 2023 and 2024.
3. **JOIN Clause**:
    - **JOIN PAWS_DBT_RPT.MART_STREAMS_DIM_DETECTORSITE_CURR geo ON agg.LINK_ID = geo.LINK_ID**: This command performs an inner join between the aggregated volume data (agg) and the detector site metadata table (geo). The link is established based on the common LINK_ID field, which uniquely identifies each detector location.
4. **WHERE Clause**:
    - **geo.DETECTORSITE_SUBURB_NAME IS NOT NULL** and **geo.DETECTORSITE_DIVISION_NAME IS NOT NULL**: This filter removes records where the suburb or division name is missing, which ensures that all analysed data has a clear geographic identifier.
    - **geo.DETECTORSITE_SUBURB_NAME!='Unknown' geo.DETECTORSITE_DIVISION_NAME != 'Unknown'**: These additional filters exclude records where the suburb or division name is set as 'Unknown', which would be irrelevant for analysis.
5. **ORDER BY Clause**:
    - **ORDER BY agg.year, agg.month_name**: This command sorts the resulting records chronologically by year and month. This sorting makes it easier to identify trends or changes in vehicle volume over time

Output :

	YEAR	MONTH_NAME	TOTAL_VOLUME	DETECTORSITE_NAME	LONGITUDE	LATITUDE	DETECTORSITE_SUBURB_NAME	DETI
1	2023	August	385649.00000	5681/EB/RS	153.331048	-27.85309	Coomera	Divis
2	2023	August	213471.00000	5507/NB/LS	153.289896	-27.816083	Pimpama	Divis
3	2023	August	132666.00000	5507/EB/RS	153.289896	-27.816083	Pimpama	Divis
4	2023	August	132666.00000	5507/EB/LS	153.289896	-27.816083	Pimpama	Divis
5	2023	August	106230.00000	5671/SE/TS	153.300141	-27.822535	Pimpama	Divis
6	2023	August	93657.00000	5507/WB/LS	153.289896	-27.816083	Pimpama	Divis
7	2023	August	44827.00000	5404/NW/RS	153.303985	-27.997235	Nerang	Divis
8	2023	August	675245.00000	5670/EB/RS	153.296534	-27.820959	Pimpama	Divis
9	2023	August	398204.00000	5670/WB/TS	153.296534	-27.820959	Pimpama	Divis
10	2023	August	27625.00000	5657/EB/LS	153.301531	-27.833659	Pimpama	Divis
11	2023	August	124707.00000	5670/NB/RS	153.296534	-27.820959	Pimpama	Divis
12	2023	August	117851.00000	5657/NB/TQ	153.301531	-27.833659	Pimpama	Divis
13	2023	August	213471.00000	5507/NB/RS	153.289896	-27.816083	Pimpama	Divis
14	2023	August	264914.00000	5674/EB/TS	153.292703296	-27.823498307	Pimpama	Divis

Data Analysis in Snowflake: Volume Comparison by Weekday and Weekend

Step 1: Volume Comparison by Weekday and Weekend

- Use the following query to identifies all the LINK_ID values that exist in both 2023 and 2024. It ensures that the analysis considers only those links that have data in both years.

```sql
WITH Common_Links AS (
    -- Identify links that exist in both 2023 and 2024
    SELECT LINK_ID
    FROM PAWS_DBT_RPT.MART_STREAMS_FCT_DETECTORSITE_AGG_DAILY
    WHERE START_DATE_KEY LIKE '2023%' OR START_DATE_KEY LIKE '2024%'
    GROUP BY LINK_ID
    HAVING COUNT(DISTINCT SUBSTRING(START_DATE_KEY, 1, 4)) = 2
),
```

Using the following query to calculate the total vehicle volume for each day across all common links. It groups data by date, year, month, and weekday.

```sql
Summated_Volumes AS (
    -- Summate the aggregate volume for each day for the entire network (common links only)
    SELECT
        TO_DATE(START_DATE_KEY, 'YYYYMMDD') AS date_key,  -- Extract full date
        SUBSTRING(START_DATE_KEY, 1, 4) AS year,          -- Extract year
        SUBSTRING(START_DATE_KEY, 5, 2) AS month,         -- Extract month
        WEEK_DAY,                                         -- WEEK_DAY to identify weekdays/weekends
        SUM(AGG_VOLUME) AS total_volume                   -- Summate AGG_VOLUME for the entire network
    FROM
        PAWS_DBT_RPT.MART_STREAMS_FCT_DETECTORSITE_AGG_DAILY
    WHERE
        (START_DATE_KEY BETWEEN '20230701' AND '20230917'   -- Only consider data from July 1 to September 17, 2023
         OR START_DATE_KEY BETWEEN '20240701' AND '20240917')  -- Same period for 2024
        AND LINK_ID IN (SELECT LINK_ID FROM Common_Links)  -- Use only common links
    GROUP BY
        TO_DATE(START_DATE_KEY, 'YYYYMMDD'),
        SUBSTRING(START_DATE_KEY, 1, 4),
        SUBSTRING(START_DATE_KEY, 5, 2),
        WEEK_DAY
),
```

The next query calculates the average total volume for weekdays and weekends, grouping by year and month. It helps understand how traffic behaves on weekdays versus weekends.

```
Weekday_Weekend_Average AS (
    -- Calculate the average total volume for representative weekdays and weekends
    SELECT
        year,
        month,
        CASE
            WHEN WEEK_DAY BETWEEN 1 AND 5 THEN 'Weekday'    -- Weekdays are Monday (1) to Friday (5)
            ELSE 'Weekend'                                  -- Weekends are Saturday (6) and Sunday (7)
        END AS day_type,
        ROUND(AVG(total_volume), 3) AS avg_total_volume     -- Average total volume for representative day rounded to 3 decimal places
    FROM
        Summated_Volumes
    GROUP BY
        year,
        month,
        CASE
            WHEN WEEK_DAY BETWEEN 1 AND 5 THEN 'Weekday'
            ELSE 'Weekend'
        END
)
```

This query compares the average volume between 2023 and 2024 by matching months and weekday/weekend types. It calculates the percentage difference between the two years.

```
-- Compare the average representative weekday and weekend between 2023 and 2024
SELECT
    curr.year AS current_year,
    curr.month AS current_month,
    curr.day_type,
    curr.avg_total_volume AS current_avg_volume,
    prev.avg_total_volume AS previous_avg_volume,
    CASE
        WHEN prev.avg_total_volume = 0 THEN NULL
        ELSE ROUND((curr.avg_total_volume - prev.avg_total_volume) / prev.avg_total_volume * 100, 2)
    END AS percentage_difference
FROM
    Weekday_Weekend_Average curr
LEFT JOIN
    Weekday_Weekend_Average prev
ON
    curr.month = prev.month
    AND curr.day_type = prev.day_type
    AND prev.year = '2023'
WHERE
    curr.year = '2024'
    AND curr.month IN ('07', '08', '09')  -- Only consider July, August, and September
ORDER BY
    curr.year, curr.month, curr.day_type;

```

Output:

Results | Chart

	CURRENT_YEAR	CURRENT_MONTH	DAY_TYPE	CURRENT_AVG_VOLUME	PREVIOUS_AVG_VOLUME	PERCENTAGE_DIFFERENCE
1	2024	07	Weekday	155257489.565	143748211.762	8.01
2	2024	07	Weekend	127000687.625	115653086.100	9.81
3	2024	08	Weekday	154477726.500	145728474.522	6.00
4	2024	08	Weekend	129524827.222	117537491.750	10.20
5	2024	09	Weekday	143500797.333	148063037.000	-3.08
6	2024	09	Weekend	126310180.400	119534057.000	5.67

Data Visualization in QGIS

CSV FILE: Sept Aug and July 2023 vs 2024 agg volume.csv

Step 1: Load the CSV into QGIS

- Open QGIS and import the CSV file with the vehicle volume data.
    - Layer > Add Layer > Add Delimited Text Layer.
    - Choose the CSV file, and ensure the correct X (longitude) and Y (latitude) fields are selected for mapping.

Step 2: Set the Projection

- Ensure the projection is set to EPSG:4326 - WGS 84. ○ Right-click on the layer and go to Properties > Information to verify the Coordinate Reference System (CRS).

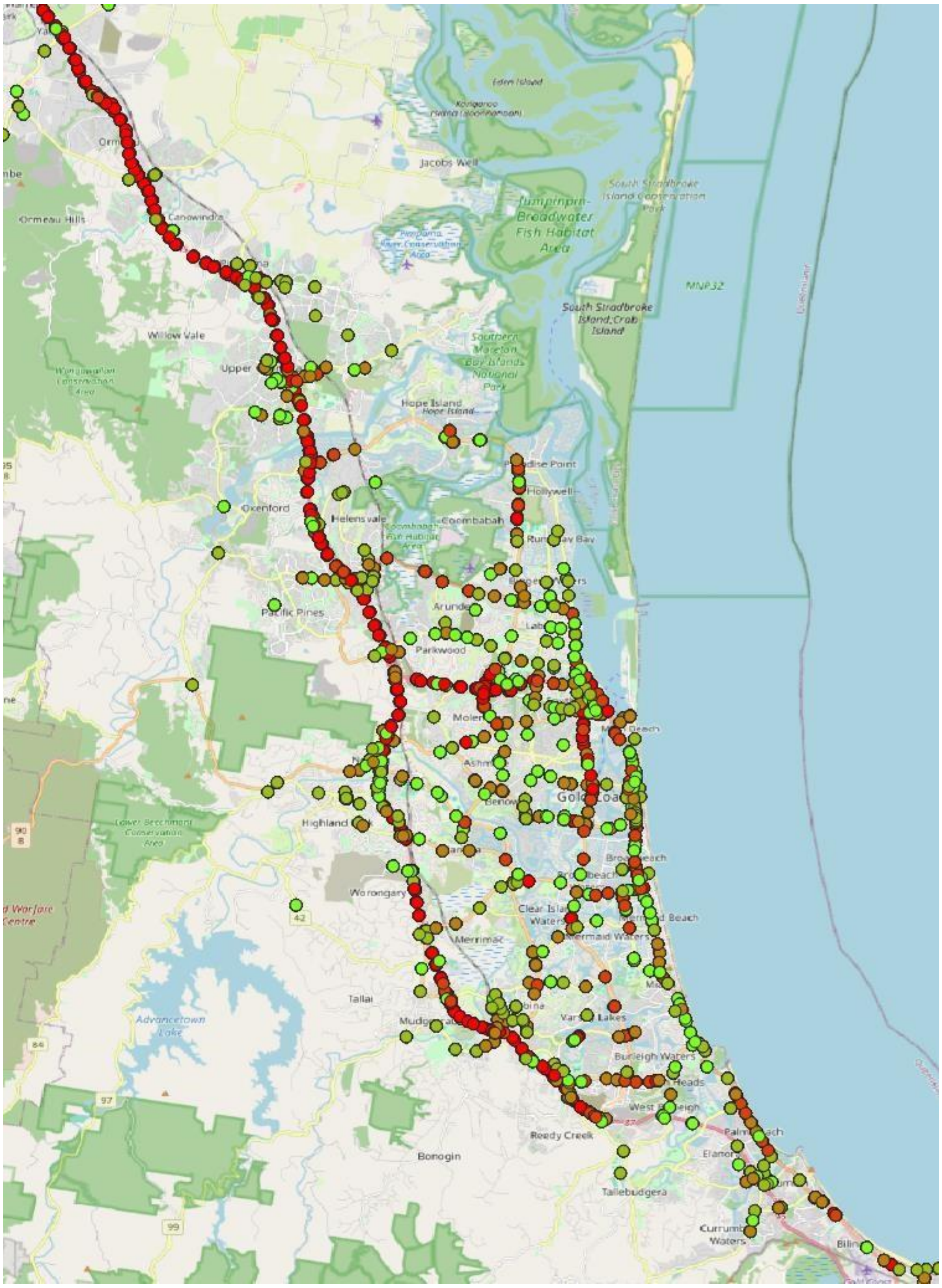
Step 3: Style the Layer

- Apply a graduated style to the vehicle volume data.
    - Right-click on Layer > Properties > Symbology.
    - Choose Graduated from the drop-down and select total_volume as the column. ○ Set the colour ramp to display green to red, with green indicating lower volumes and red indicating higher volumes.

Step 4: Apply Filters for Date Ranges

- Use the Query Builder in QGIS to filter the data for specific months (July, August, and September).
    - Right-click Layer > Filter and apply the filter:

NEXT PAGE FOR THE FINAL OUTPUT



Jacobs Well
Ormeau Hills
Willow Vale
Hope Island
Oxenford
Helensvale
Coombabah
Pacific Pines
Parkwood
Worongary
Merrimac
Tallai
Reedy Creek
Bonogin
Tallebudgera
Burleigh Waters
Advancetown Lake
South Stradbroke Island

Data Prediction

Interpretation of Results

- **Key Insights:**
  - The results indicated a slight decrease in private vehicle volumes in 2024 as compared to 2023, especially during weekdays. The public transport fare reduction may have contributed to this trend.
  - The percentage changes in volume, represented in graphs and maps, show how vehicle usage fluctuates across months.
- **Volume Summation and Averaging:**
  - We calculated the average volume per day for each detector site and performed a month-wise comparison of vehicle volumes, considering only common detector links between 2023 and 2024.

Tech Stack

- **Snowflake:** Used for querying large datasets.
- **QGIS:** For geospatial visualization of detector site volumes and trends.

Summary

This manual provides an overview of how the volume analysis was conducted for the Green City Situation Awareness Project, including details on Snowflake queries, QGIS plotting steps, and data interpretation. The analysis led to insights on changing vehicle usage patterns, particularly in response to public transport policy changes.