

ETL - CASE STUDY SOLUTION

SQL queries to answer each business question

How does revenue compare between July and August?

```
SELECT import_month,  
       SUM(total_amount) AS total_revenue  
FROM dbt_kabdullah_mart.fact_combined_trips  
WHERE import_month IN ('2021-07', '2021-08')  
GROUP BY import_month;
```

What are the top 10 pickup zones by revenue in each month?

```
SELECT pickup_zone, SUM(total_amount) AS revenue  
FROM dbt_kabdullah_mart.fact_combined_trips  
GROUP BY pickup_zone  
ORDER BY revenue DESC  
LIMIT 10;
```

Which zones show the largest increase/decrease in revenue between months?

```
WITH revenue_per_zone AS (  
    SELECT import_month, pickup_zone, SUM(total_amount) AS revenue  
    FROM dbt_kabdullah_mart.fact_combined_trips  
    WHERE import_month IN ('2021-07', '2021-08')  
    GROUP BY import_month, pickup_zone  
)  
pivoted AS (  
    SELECT pickup_zone,  
           MAX(CASE WHEN import_month = '2021-07' THEN revenue END) AS july_revenue,  
           MAX(CASE WHEN import_month = '2021-08' THEN revenue END) AS august_revenue  
    FROM revenue_per_zone
```

```

GROUP BY pickup_zone
),
final AS (
    SELECT pickup_zone,
           COALESCE(august_revenue, 0) - COALESCE(july_revenue, 0) AS revenue_change
    FROM pivoted
)
SELECT *
FROM final
ORDER BY ABS(revenue_change) DESC;

```

How do average tips vary by drop-off zone and month?

```

SELECT import_month,
       dropoff_zone,
       AVG(tip_amount) AS avg_tip
FROM dbt_kabdullah_mart.fact_combined_trips
WHERE import_month IN ('2021-07', '2021-08')
GROUP BY import_month, dropoff_zone;

```

What's the average trip distance and fare per zone per month?

```

SELECT import_month,
       pickup_zone,
       AVG(trip_distance) AS avg_distance,
       AVG(fare_amount) AS avg_fare
FROM dbt_kabdullah_mart.fact_combined_trips
WHERE import_month IN ('2021-07', '2021-08')
GROUP BY import_month, pickup_zone;

```

Do certain pickup zones consistently generate higher tips or longer rides?

```

SELECT pickup_zone,

       AVG(tip_amount) AS avg_tip,

       AVG(trip_distance) AS avg_distance

FROM dbt_kabdullah_mart.fact_combined_trips

GROUP BY pickup_zone

ORDER BY avg_tip DESC, avg_distance DESC;

```

Is there a correlation between trip distance and tip amount across months?

```

SELECT import_month,

       CORR(trip_distance, tip_amount) AS distance_tip_correlation

FROM dbt_kabdullah_mart.fact_combined_trips

WHERE import_month IN ('2021-07', '2021-08')

GROUP BY import_month;

```

What zones saw the largest drop in average revenue per trip month-over-month?

```

WITH avg_revenue AS (

    SELECT import_month, pickup_zone, AVG(total_amount) AS avg_revenue

    FROM dbt_kabdullah_mart.fact_combined_trips

    WHERE import_month IN ('2021-07', '2021-08')

    GROUP BY import_month, pickup_zone

),

pivoted AS (

    SELECT pickup_zone,

           MAX(CASE WHEN import_month = '2021-07' THEN avg_revenue END) AS july_avg,

           MAX(CASE WHEN import_month = '2021-08' THEN avg_revenue END) AS aug_avg

    FROM avg_revenue

    GROUP BY pickup_zone

)

SELECT pickup_zone,

```

july_avg,

aug_avg,

(aug_avg - july_avg) AS avg_revenue_change

FROM pivoted

WHERE july_avg IS NOT NULL AND aug_avg IS NOT NULL

ORDER BY avg_revenue_change ASC

LIMIT 10;

DBT & Neon Connection

Loaded the seeds into Neon (raw, staging, and semantic Layers)

The screenshot shows the dbt IDE interface. On the left, the 'File explorer' pane displays the project structure, including 'analyses', 'macros', 'models', 'business', 'example', 'staging', 'src.yml', 'seeds', 'snapshots', and 'tests'. The 'src.yml' file is selected, showing its contents in the main editor. The 'src.yml' file defines the source schema 'dbt_kabdullah_raw' and lists the tables 'green_tripdata_2021_07', 'taxi_zone_lookup', and 'green_tripdata_2021_08'. Below the editor, a data lineage graph is displayed, showing the flow of data from 'raw_green_tripdata' and 'raw_taxi_zone_lookup' through 'stg_green_tripdata' and 'stg_taxi_zone_lookup' to 'fact_trips' and 'fact_combined_trips'. The graph is titled '2=models/staging/src.yml' and includes an 'Update Graph' button.

The screenshot shows the Neon console interface. The 'Tables' pane on the left lists the tables 'neondb', 'dbt_kabdullah_raw', 'fact_combined_trips', and 'fact_trips'. The 'fact_combined_trips' table is selected, and its data is displayed in the main pane. The table has 30 rows and 10 columns: 'trip_date', 'import_month', 'vendor_id', 'passenger_count', 'trip_distance', 'fare_amount', 'tip_amount', 'total', and two unnamed columns. The data is as follows:

trip_date	import_month	vendor_id	passenger_count	trip_distance	fare_amount	tip_amount	total		
2021-07-01	2021-07	1	1	1.2	6	0	7.3		
2021-07-01	2021-07	2	2	13.69	42	0	43.3		
2021-07-01	2021-07	2	1	0.95	6.5	2.34	10.1		
2021-07-01	2021-07	2	1	1.24	6.5	0	7.8		
2021-07-01	2021-07	2	1	1.1	7	0	8.3		
2021-07-01	2021-07	1	1	1.9	8	3	15.0		
2021-07-01	2021-07	2	1	0.66	5	0	6.3		
2021-07-01	2021-07	2	1	1.72	7	2.08	10.3		
2021-07-01	2021-07	2	1	1.37	7.5	0	8.8		
2021-07-01	2021-07	2	1	2.14	9	2.65	15.7		
2021-07-01	2021-07	2	1	0.01	13	0	13.3		
2021-06-07	2021-07	2	1	7.74	56	0	57.3		
2021-07-01	2021-07	2	1	25.08	65.5	0	73.3		
2021-07-01	2021-07	2	1	11.4	42	0	49.8		
2021-07-01	2021-07	2	1	9.88	31	0	38.8		
2021-07-01	2021-07	2	1	1.4	6.5	0	7.8		
2021-07-01	2021-07	2	1	5.76	19.5	4.16	24.9		
2021-07-01	2021-07	2	1	5.69	20.5	0	24.5		
2021-07-01	2021-07	2	1	5.65	16.5	0	17.8		
2021-07-01	2021-07	2	1	3.27	13	0	14.3		
2021-07-01	2021-07	2	1	1.04	6	0	7.3		
2021-07-01	2021-07	2	1	5.63	14	0	14.3		
2021-07-01	2021-07	2	1	9.81	28.5	0	29.8		
2021-07-01	2021-07	2	1	3.25	12	1.6	17.6		
2021-07-01	2021-07	2	1	1.12	5.5	1.36	8.16		
2021-07-01	2021-07	2	1	2.58	10.5	0	14.5		
2021-07-01	2021-07	2	1	2.19	9	0	10.3		
2021-07-01	2021-07	2	1	10.13	29	6.06	36.3		
2021-07-01	2021-07	2	1	3.5	13.5	0	14.8		
2021-07-01	2021-07	2	1	2.96	10	0	14.0		
2021-07-01	2021-07	2	1	0.31	3.5	0	4.8		