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)