OBJECTIVE

The objective of this analysis is to explore and derive insights from ride, driver, rider, and payment data. The analysis answers eight key business questions to help management understand performance trends, driver consistency, cancellation patterns, and identify top-performing drivers eligible for bonuses.

DATA OVERVIEW

- **drivers_cleaned:** This gives us individual details of the drivers such as driver_id, name, city, signup_date, signup_time, and average rating

- **Riders_cleaned:** This gives us details of customers such rider_id, name, city, signup_date, signup_time, and email

- **rides_cleaned:this** This gives us details of the ride by riders such as ride_id, rider_id, driver_id, pickup_city, dropoff_city, distance_km, status, fare, request_date, request_time, pickup_date, pickup_time, dropoff_date, dropoff_time

- **payments_cleaned:** payment_id, ride_id, amount, method, paid_date, paid_time

METHODOLOGY

Data was cleaned and analyzed using SQL. Each business question was addressed with an independent SQL query. The results were validated to highlight insights.

DATA CLEANING AND TRANSFORMATIONS

These were the steps taken in PostgreSQL:

- Creation of tables for the 4 dataset with the right data types
- Inputting the data into the created tables
- Missing values were checked - None were found
- Duplicates were checked - None were found
- Extra spaces were removed
- Cities names were standardize
- Ensure that the rate columns are 1-5
- Formatted the date column by
    - Creating two new column (Date and time)
    - Convent and Populate the new columns
    - removed the original date column.
- Filter the tables to ensure the dates are between 2021-06-01 and 2024-12-31

INSIGHTS

- **Top ten drivers by distance**

Driver_1774 and Driver_1886 are the joint top drivers by distance with 30km with the latter going from Calgary to Los Angeles while the former going from Ottawa to Vancouver. We have 12 drivers with 29.99km of distance covered as seen in image below. I suggest we use other metrics to determine the remaining 8 e.g. revenue, consistency, etc.

```sql
SELECT
    d.name AS driver_name,
    ri.name AS rider_name,
    r.pickup_city,
    r.dropoff_city,
    r.distance_km,
    p.method AS payment_method
FROM rides_cleaned r
JOIN drivers_cleaned d
    ON r.driver_id = d.driver_id
JOIN riders_cleaned ri
    ON r.rider_id = ri.rider_id
JOIN payments_cleaned p
    ON r.ride_id = p.ride_id
ORDER BY r.distance_km DESC
LIMIT 14;
```

Data Output | Messages | Notifications

Showing rows: 1 to 14 | Page

	driver_name character varying (50)	rider_name character varying (50)	pickup_city character varying (100)	dropoff_city character varying (100)	distance_km numeric (10,2)	payment_method character varying (50)
1	Driver_1886	Rider_6131	Ottawa	Vancouver	30.00	voucher
2	Driver_1774	Rider_292	Calgary	Los Angeles	30.00	voucher
3	Driver_958	Rider_4433	Los Angeles	Vancouver	29.99	paypal
4	Driver_932	Rider_3532	San Francisco	Chicago	29.99	paypal
5	Driver_1343	Rider_9618	Calgary	New York	29.99	cash
6	Driver_1752	Rider_2994	Ottawa	Chicago	29.99	card
7	Driver_1379	Rider_1599	Boston	San Francisco	29.99	card

- **How many riders who signed up in 2021 still took rides in 2024?**

The company has had 1236 (Out of 1729) drivers who have signed up in 2021 and are still with the company by 2024.

```sql
SELECT
    COUNT(DISTINCT r.rider_id) AS active_2021_riders_in_2024
FROM riders_cleaned r
JOIN rides_cleaned rd
    ON r.rider_id = rd.rider_id
WHERE EXTRACT(YEAR FROM r.signup_date) = 2021
  AND EXTRACT(YEAR FROM rd.request_date) = 2024;
```

Data Output | Messages | Notifications

	active_2021_riders_in_2024 bigint
1	1236

- **Quarter with the biggest YoY growth**

Q2 2022 had the highest Quarterly Y0Y growth with little over 200% growth rate. What this means is that they were over 200% growth in revenue in Q2 2022 from Q2 2021. It is also necessary to point out quarters with negative growth - Q1, 2023; Q3, 2023 through Q2, 2024. This also shows drop in revenue in 2023 and first half of 2024

Query | Query History

```
WITH quarterly_revenue AS (
    SELECT
        EXTRACT(YEAR FROM paid_date) AS year,
        EXTRACT(QUARTER FROM paid_date) AS quarter,
        SUM(amount) AS total_revenue
    FROM payments_cleaned
    WHERE paid_date BETWEEN '2021-01-01' AND '2024-12-31'
    GROUP BY EXTRACT(YEAR FROM paid_date), EXTRACT(QUARTER FROM paid_date)
),
yoy_growth AS (
    SELECT
        curr.year,
        curr.quarter,
        curr.total_revenue,
        prev.total_revenue AS prev_year_revenue,
        ROUND(
```

Data Output | Messages | Notifications

	year numeric	quarter numeric	total_revenue numeric	prev_year_revenue numeric	yoy_growth_percent numeric
5	2022	2	101159.23	33530.12	201.70
6	2022	3	105208.17	104763.20	0.42
7	2022	4	102919.00	100161.41	2.75
8	2023	1	99687.35	101460.49	-1.75
9	2023	2	101360.86	101159.23	0.20
10	2023	3	99462.76	105208.17	-5.46
11	2023	4	97181.49	102919.00	-5.57

- **Top 5 drivers with the highest consistency (most rides per active month)?**

The drivers who has been consistency in terms of number of ride per month are as followed Driver_1722, Driver_1690, Driver_1733, Driver_800, and Driver_1876. Over time, they have shown consistency in having more riders than other drivers on average.

Query | Query History

```sql
WITH driver_activity AS (
    SELECT
        r.driver_id,
        COUNT(r.ride_id) AS total_rides,
        MIN(d.signup_date) AS signup_date,
        MAX(r.pickup_date) AS last_ride_date
    FROM rides_cleaned r
    JOIN drivers_cleaned d ON r.driver_id = d.driver_id
    WHERE r.status = 'completed'
    GROUP BY r.driver_id
),
driver_months AS (
    SELECT
        driver_id,
        total_rides,
        signup_date,
        last_ride_date,
        GREATEST(
            (DATE_PART('year', AGE(last_ride_date, signup_date)) * 12
```

Data Output | Messages | Notifications

	driver_id integer	driver_name character varying (50)	avg_monthly_rides numeric	total_rides bigint	active_months double precision
1	1722	Driver_1722	13.00	13	1
2	1690	Driver_1690	9.00	18	2
3	1733	Driver_1733	9.00	9	1
4	800	Driver_800	8.50	17	2
5	1876	Driver_1876	8.50	17	2

- **Cities with the most cancellation rate**

As seen below, Chicago is the city with the highest cancellation rate with 19.26%. Toronto and San Francisco are second and third with 19.08% and 18.47% respectively. Boston has the least cancellation of 17.76%.

Query Query History

```sql
WITH city_cancellation AS (
    SELECT
        pickup_city AS city,
        COUNT(*) AS total_rides,
        COUNT(*) FILTER (WHERE LOWER(status) = 'cancelled') AS cancelled_rides
    FROM rides_cleaned
    GROUP BY pickup_city
)
SELECT
    city,
    total_rides,
    cancelled_rides,
    ROUND((cancelled_rides::NUMERIC / NULLIF(total_rides, 0)) * 100, 2) AS cancellation_rate
FROM city_cancellation
ORDER BY cancellation_rate DESC;
```

Data Output Messages Notifications

SQL Showing rows: 1

	city character varying (100)	total_rides bigint	cancelled_rides bigint	cancellation_rate numeric
1	Chicago	4513	869	19.26
2	Toronto	4524	863	19.08
3	San Francisco	4494	830	18.47
4	Calgary	4525	834	18.43
5	Montreal	4437	805	18.14
6	New York	4415	799	18.10
7	Los Angeles	4463	800	17.93
8	Vancouver	4385	784	17.88

- **Riders with 10+ rides but never paid with cash**

Only Rider_7823 and Rider_3890 didn't receive cash payment and had 10+ rides. These drivers should be encouraged to do more cashless transactions.

Query Query History

```sql
WITH rider_ride_counts AS (
    SELECT
        rider_id,
        COUNT(DISTINCT ride_id) AS total_rides
    FROM rides_cleaned
    GROUP BY rider_id
),
riders_no_cash AS (
    SELECT
        r.rider_id
    FROM payments_cleaned p
    JOIN rides_cleaned r ON p.ride_id = r.ride_id
    GROUP BY r.rider_id
    HAVING SUM(CASE WHEN LOWER(p.method) = 'cash' THEN 1 ELSE 0 END) = 0
)
SELECT
    rc.rider_id,
    rd.name AS rider_name,
    rc.total_rides
```

Data Output Messages Notifications

	rider_id integer	rider_name character varying (50)	total_rides bigint
1	7823	Rider_7823	14
2	3890	Rider_3890	11

- **Top 3 drivers in each city by total revenue earned between June 2021 and Dec 2024**

Find below top3 for each cities

city	driver_id	driver_name	total_revenue	city_rank
Boston	1176	Driver_1176	448.4	1
Boston	286	Driver_286	326.58	2
Boston	1141	Driver_1141	315.88	3
Calgary	1980	Driver_1980	476.91	1
Calgary	1059	Driver_1059	346.86	2
Calgary	404	Driver_404	338.8	3
Chicago	413	Driver_413	449.45	1
Chicago	1410	Driver_1410	421.9	2

Chicago	1459	Driver_1459	315.48	3
Los Angeles	761	Driver_761	433.12	1
Los Angeles	448	Driver_448	373.29	2
Los Angeles	287	Driver_287	334.24	3
Montreal	163	Driver_163	377.87	1
Montreal	1328	Driver_1328	341.06	2
Montreal	541	Driver_541	337	3
New York	681	Driver_681	338.41	1
New York	1708	Driver_1708	318.65	2
New York	1910	Driver_1910	303.01	3
Ottawa	418	Driver_418	358.1	1
Ottawa	76	Driver_76	353.81	2
Ottawa	1393	Driver_1393	306.59	3
San Francisco	286	Driver_286	354.75	1
San Francisco	13	Driver_13	352.62	2
San Francisco	1626	Driver_1626	352.13	3
Toronto	988	Driver_988	380.56	1
Toronto	372	Driver_372	363.52	2
Toronto	383	Driver_383	322.63	3
Vancouver	1924	Driver_1924	365.35	1
Vancouver	508	Driver_508	358.06	2
Vancouver	578	Driver_578	329.28	3

Query Query History

```sql
WITH city_revenue AS (
    SELECT
        r.pickup_city AS city,
        r.driver_id,
        d.name AS driver_name,
        SUM(p.amount) AS total_revenue
    FROM rides_cleaned r
    JOIN payments_cleaned p ON r.ride_id = p.ride_id
    JOIN drivers_cleaned d ON r.driver_id = d.driver_id
    WHERE r.pickup_date BETWEEN '2021-06-01' AND '2024-12-31'
    GROUP BY r.pickup_city, r.driver_id, d.name
),
ranked_drivers AS (
    SELECT
```

Data Output Messages Notifications

SQL

	city character varying (100)	driver_id integer	driver_name character varying (50)	total_revenue numeric	city_rank bigint
1	Boston	1176	Driver_1176	448.40	1
2	Boston	286	Driver_286	326.58	2
3	Boston	1141	Driver_1141	315.88	3
4	Calgary	1980	Driver_1980	476.91	1
5	Calgary	1059	Driver_1059	346.86	2
6	Calgary	404	Driver_404	338.80	3
7	Chicago	413	Driver_413	449.45	1
8	Chicago	1410	Driver_1410	421.90	2
9	Chicago	1459	Driver_1459	315.48	3

Total rows: 30 Query complete 00:00:00.170

- **Top 10 drivers qualified for bonus**

Based on drivers that has 30+ rides, greater that 4.5 rating, and cancellation rate of less than 5%, these are the top 10 drivers.

Query | Query History

```sql
SELECT
    d.driver_id,
    d.name AS driver_name,
    COUNT(r.ride_id) AS total_rides,
    d.rating AS avg_rating,
    ROUND(
        (SUM(CASE WHEN r.status = 'Cancelled' THEN 1 ELSE 0 END) * 100.0) / COUNT(r.ride_id),
        2
    ) AS cancellation_rate
FROM
    rides_cleaned r
JOIN
    drivers_cleaned d
```

Data Output | Messages | Notifications

Showing rows: 1 to

	driver_id integer	driver_name character varying (50)	total_rides bigint	avg_rating numeric (2,1)	cancellation_rate numeric
1	1137	Driver_1137	38	5.0	0.00
2	837	Driver_837	36	5.0	0.00
3	1483	Driver_1483	34	5.0	0.00
4	553	Driver_553	33	5.0	0.00
5	252	Driver_252	33	5.0	0.00
6	257	Driver_257	33	5.0	0.00
7	774	Driver_774	33	5.0	0.00
8	421	Driver_421	33	5.0	0.00
9	1309	Driver_1309	32	5.0	0.00
10	213	Driver_213	32	5.0	0.00

Total rows: 10 | Query complete 00:00:00.352

Summary/Conclusion

This analysis provided a comprehensive overview of ride, driver, and rider performance across multiple cities between June 2021 and December 2024. The results revealed that business performance peaked in Q2 2022 with over 200% year-over-year revenue growth, followed by a noticeable slowdown through 2023 and early 2024.

Top-performing drivers such as Driver_1774, Driver_1886, and Driver_1722 demonstrated consistent performance, covering long distances and maintaining high ride frequencies — key indicators of reliability and operational excellence. Meanwhile, Chicago, Toronto, and San Francisco recorded the highest ride cancellations, signaling potential service or logistics issues that may require deeper investigation.

From a revenue perspective, the top three drivers per city consistently outperformed peers, with strong representation from markets like Calgary, Boston, and Los Angeles. In addition, ten standout drivers qualified for bonuses based on their ride volume, high ratings, and low cancellation rates — confirming strong alignment with company performance targets.

Overall, the findings highlight both growth opportunities and operational challenges: improving customer retention in high-cancellation cities, recognizing consistent drivers through incentives, and replicating best-performing regions' success patterns. These insights can guide management in rewarding excellence, improving service quality, and driving sustainable revenue growth in future quarters.