SQL Analytics Project – Ola Rides Dataset

Name: Prem Pushkar
Course: MBA (DTU)
Tools: PostgreSQL, pgAdmin 4
Date: [12-06-2025]

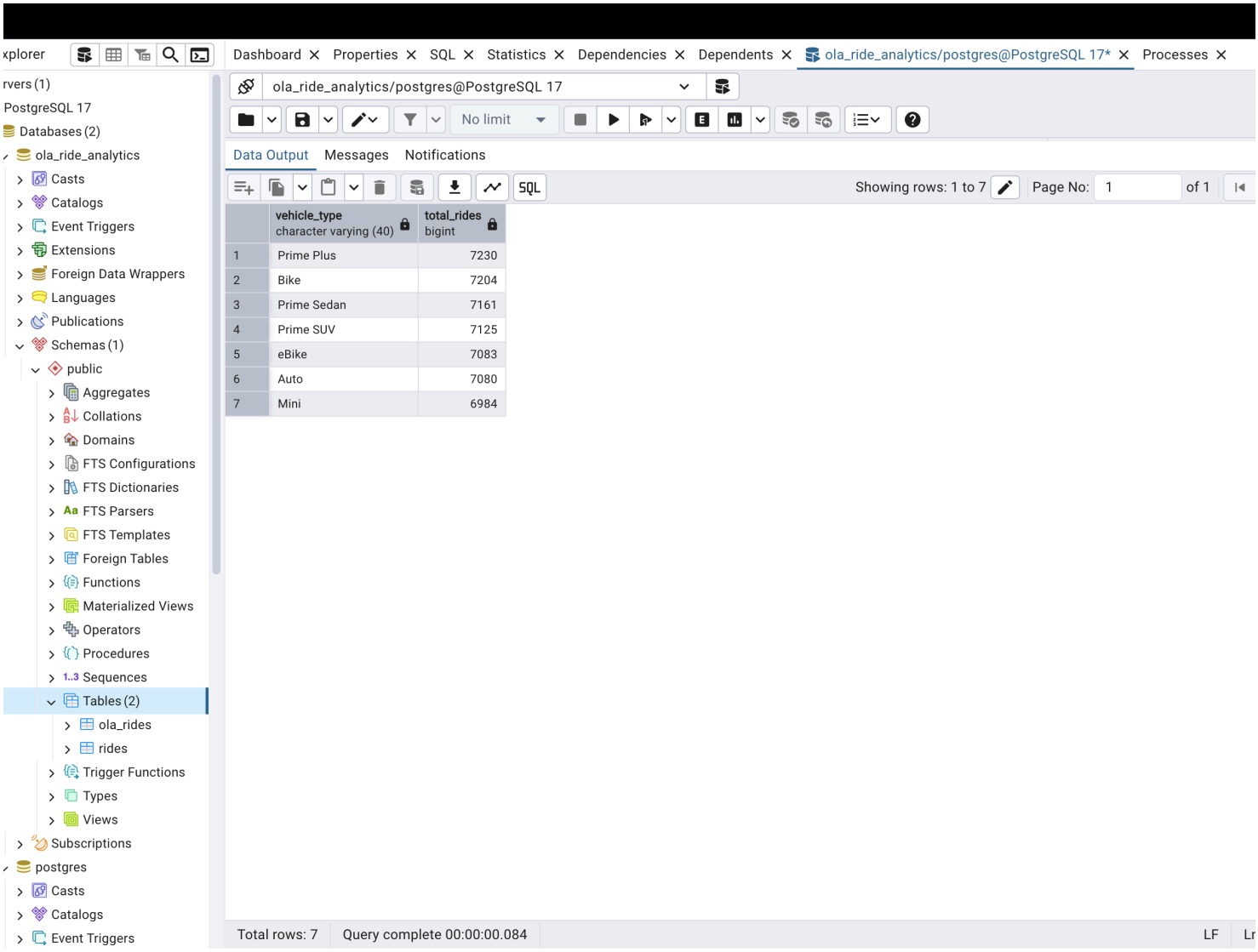
Dataset Overview

This dataset contains 50,000 Ola ride records with details like booking ID, ride time, vehicle type, pickup/drop locations, distance, fare, payment method, and customer/driver ratings. It is used to analyze ride trends, revenue, cancellations, and customer behavior.

Column Name	Data Type	Description
`booking_id`	VARCHAR(40)	Unique ID for each ride booking
`ride_timestamp`	TIMESTAMP	Combined date and time when the ride was scheduled or started
`booking_status`	VARCHAR(30)	Status of the booking (`Success`, `Cancelled`, `Incomplete`, etc.)
`customer_id`	BIGINT	Unique identifier for the customer who booked the ride
`vehicle_type`	VARCHAR(20)	Type of vehicle (`Auto`, `Mini`, `Bike`, `Prime Sedan`, etc.)
`pickup_location`	VARCHAR(50)	Starting location of the ride
`drop_location`	VARCHAR(50)	Ending location of the ride
`avg_vtat`	NUMERIC(6,2)	Average Vehicle Turnaround Time — time taken for the vehicle to start ride
`avg_ctat`	NUMERIC(6,2)	Average Customer Turnaround Time — time taken for the customer to board
`cancelled_by_customer`	SMALLINT	`1` if the ride was cancelled by the customer, otherwise `0`
`cancelled_by_driver`	SMALLINT	`1` if the ride was cancelled by the driver, otherwise `0`
`incomplete_rides`	SMALLINT	`1` if the ride was incomplete or dropped midway, otherwise `0`
`booking_value`	NUMERIC(10,2)	Amount paid or billed for the ride in INR
`payment_method`	VARCHAR(30)	Mode of payment (`Cash`, `Wallet`, `Card`, etc.)
`ride_distance`	NUMERIC(6,2)	Total distance of the ride in kilometers
`driver_rating`	NUMERIC(3,2)	Rating given by the customer to the driver (out of 5)
`customer_rating`	NUMERIC(3,2)	Rating given by the driver to the customer (out of

Question 1: What is the total number of rides by vehicle type?

```
SELECT vehicle_type, COUNT(*) AS total_rides
FROM rides
GROUP BY vehicle_type
ORDER BY total_rides DESC;
```



Caption

QUESTION 2: What is the total and average booking revenue per vehicle type?

QUERY-SELECT vehicle_type,

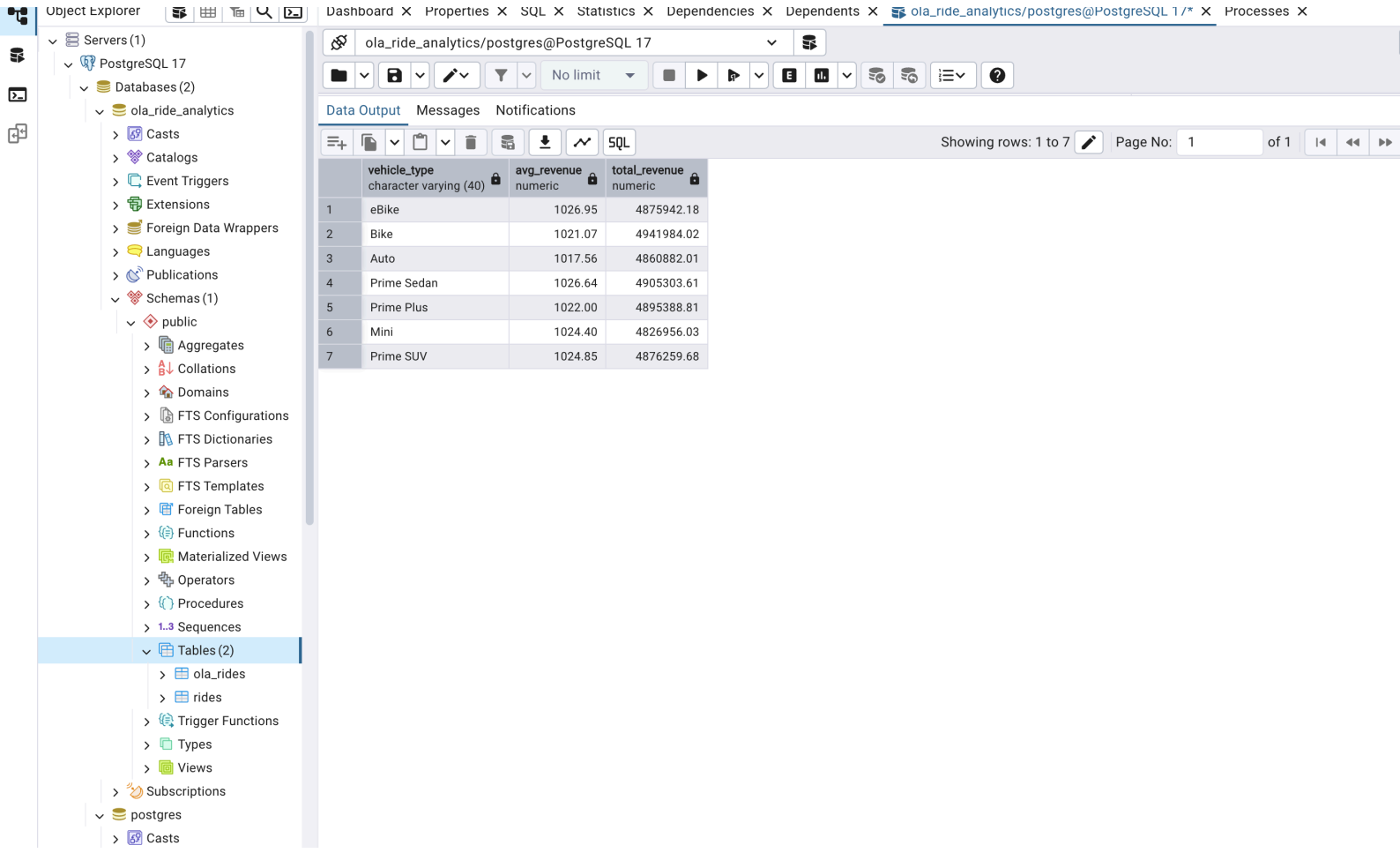
ROUND(AVG(booking_value),2) AS avg_revenue,

ROUND(SUM(booking_value),2) AS total_revenue

FROM rides

WHERE booking_status = 'Success'

GROUP BY vehicle_type;



Caption

QUESTION 3. What are the top 5 pickup-drop location pairs by frequency?

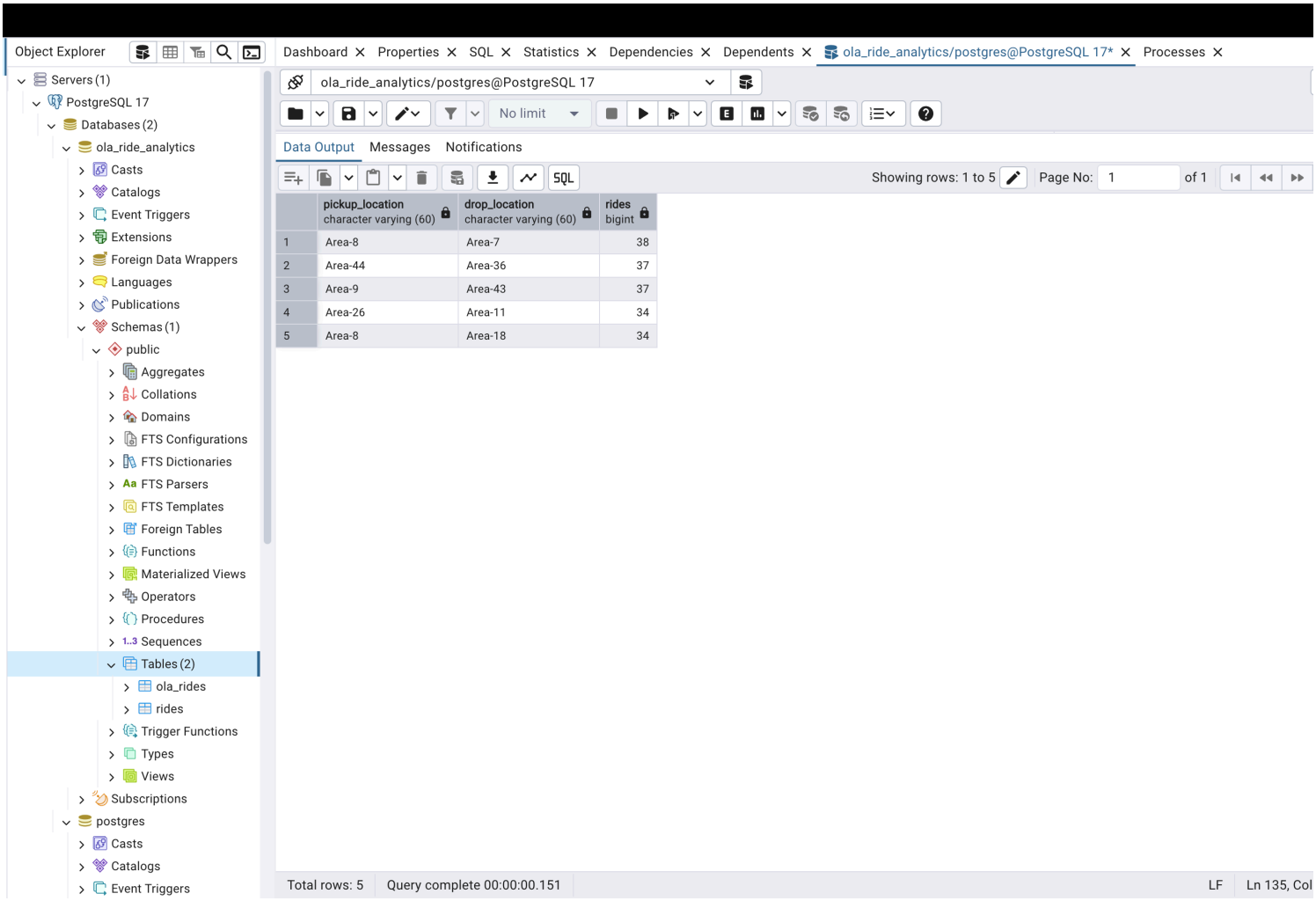
QUERY-SELECT pickup_location, drop_location, COUNT(*) AS rides

FROM rides

GROUP BY pickup_location, drop_location

ORDER BY rides DESC

LIMIT 5;



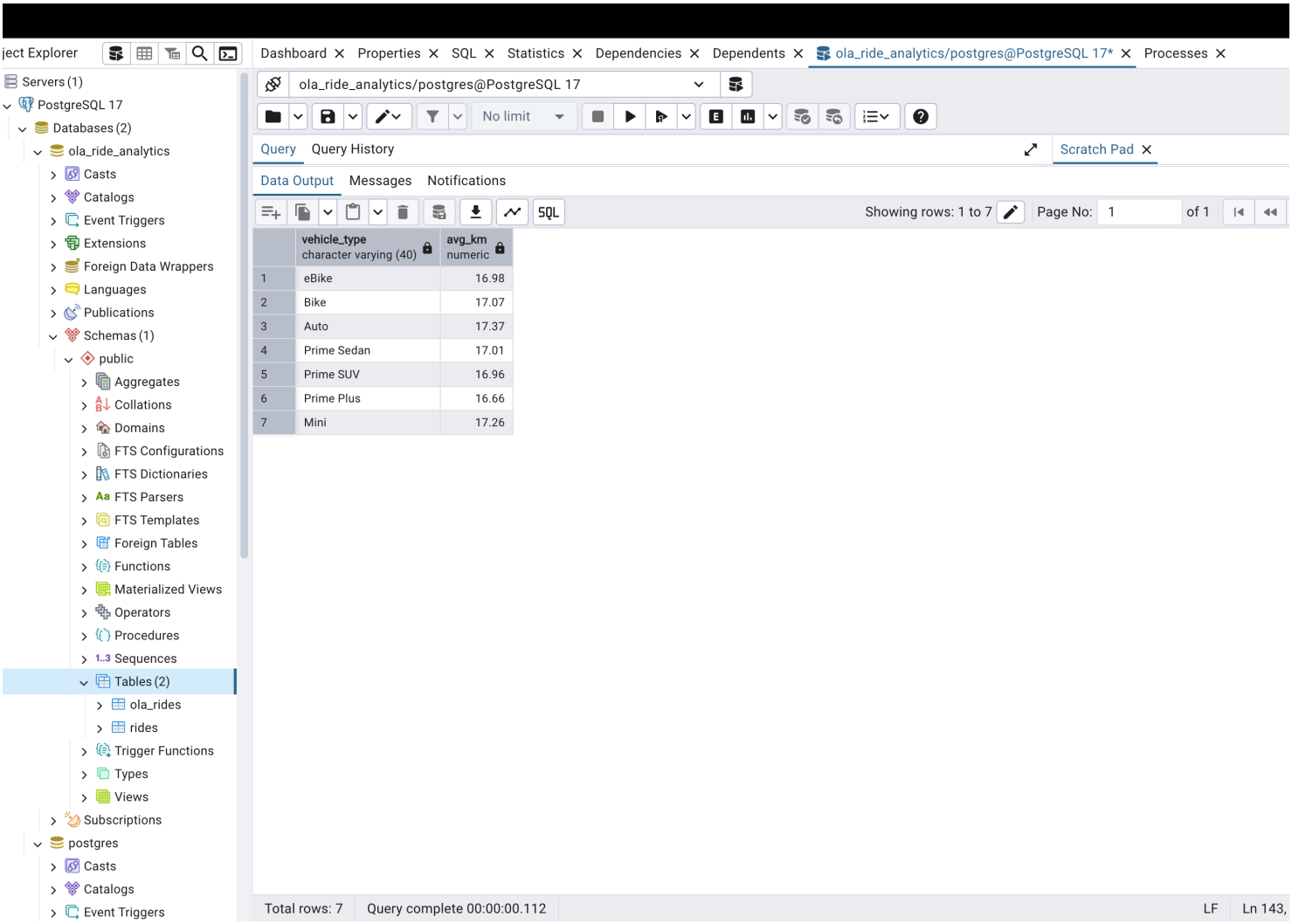
Caption

QUESTION 4: What is the Avgerage ride distance per vehicle?

QUERY: SELECT vehicle_type, ROUND(AVG(ride_distance),2) AS avg_km

FROM rides

GROUP BY vehicle_type;



Caption

QUESTION 5: Which hour(TOP 5) of the day has the highest number of completed rides?

QUERY- SELECT DATE_PART('hour', ride_timestamp) AS hr, COUNT(*) AS rides

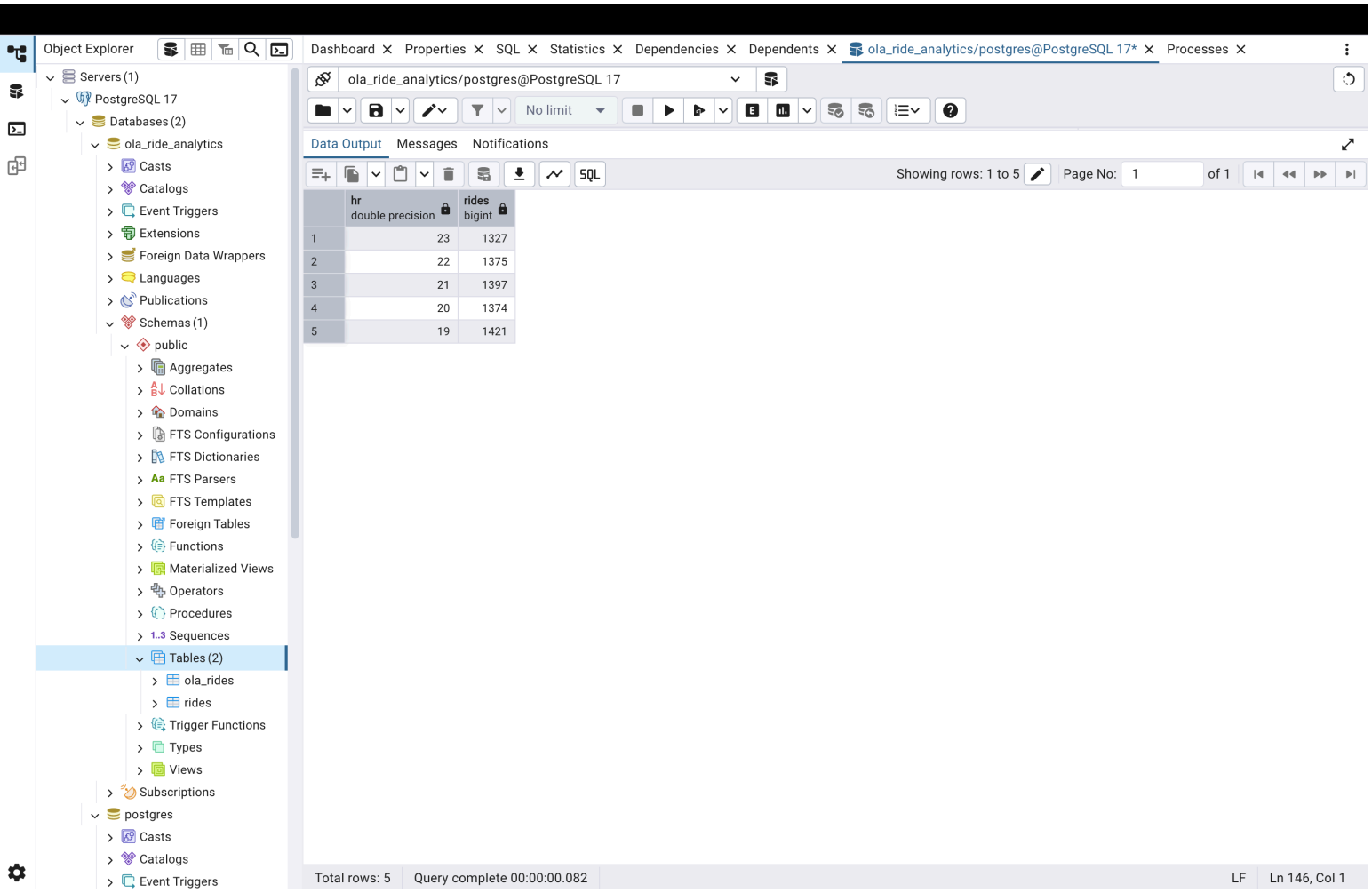
FROM rides

WHERE booking_status = 'Success'

GROUP BY hr

ORDER BY hr DESC

LIMIT 5;



Caption

QUESTION 6: What is the success vs cancellation rate for each vehicle type?

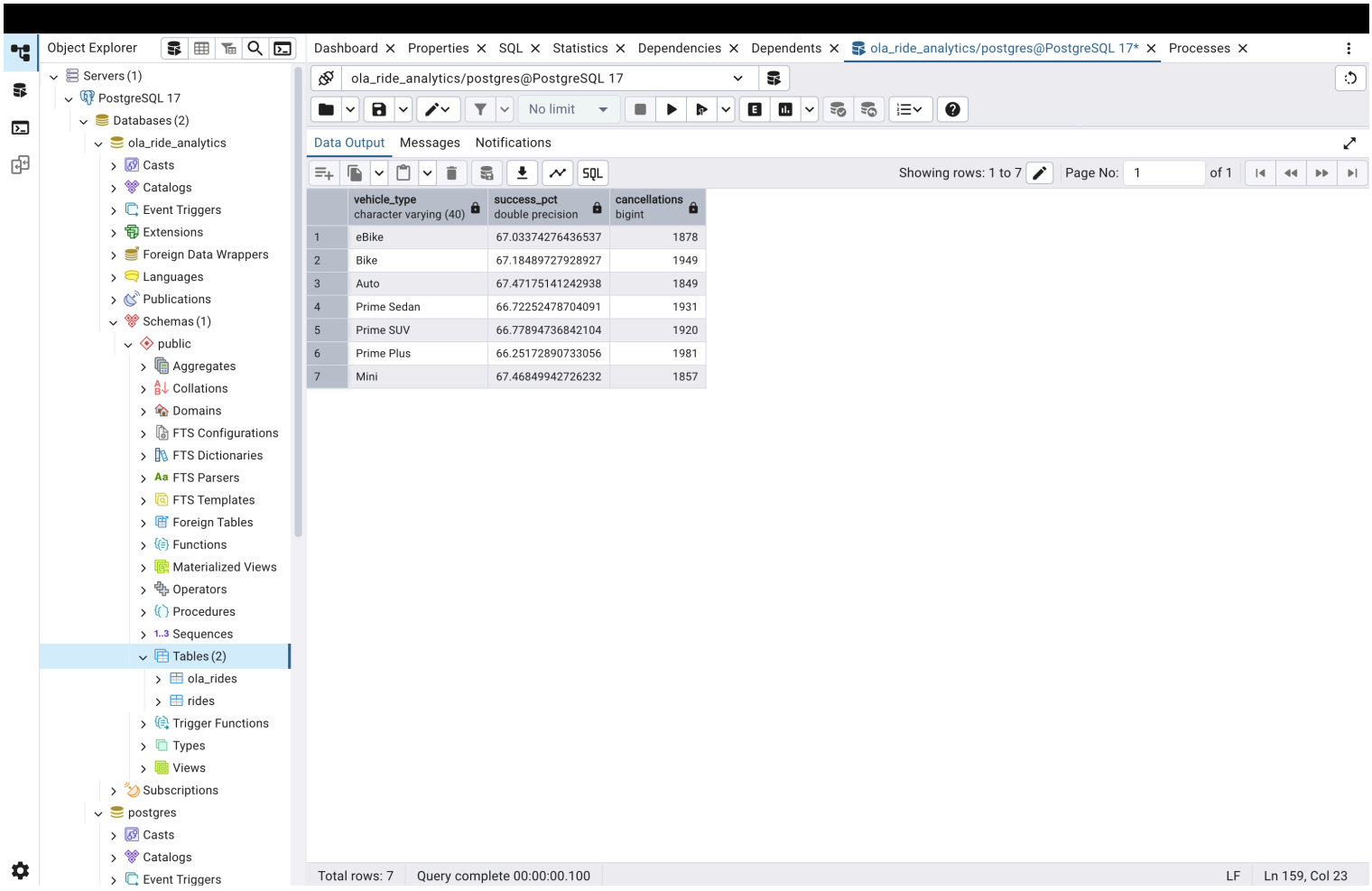
QUERY- SELECT vehicle_type,

SUM(CASE WHEN booking_status='Success' THEN 1 END)::float / COUNT(*) *100 AS success_pct,

SUM(cancelled_by_customer)+SUM(cancelled_by_driver) AS cancellations

FROM rides

GROUP BY vehicle_type;



Caption

QUESTION 7: Which locations have highest driver cancellations?

QUERY- SELECT pickup_location,

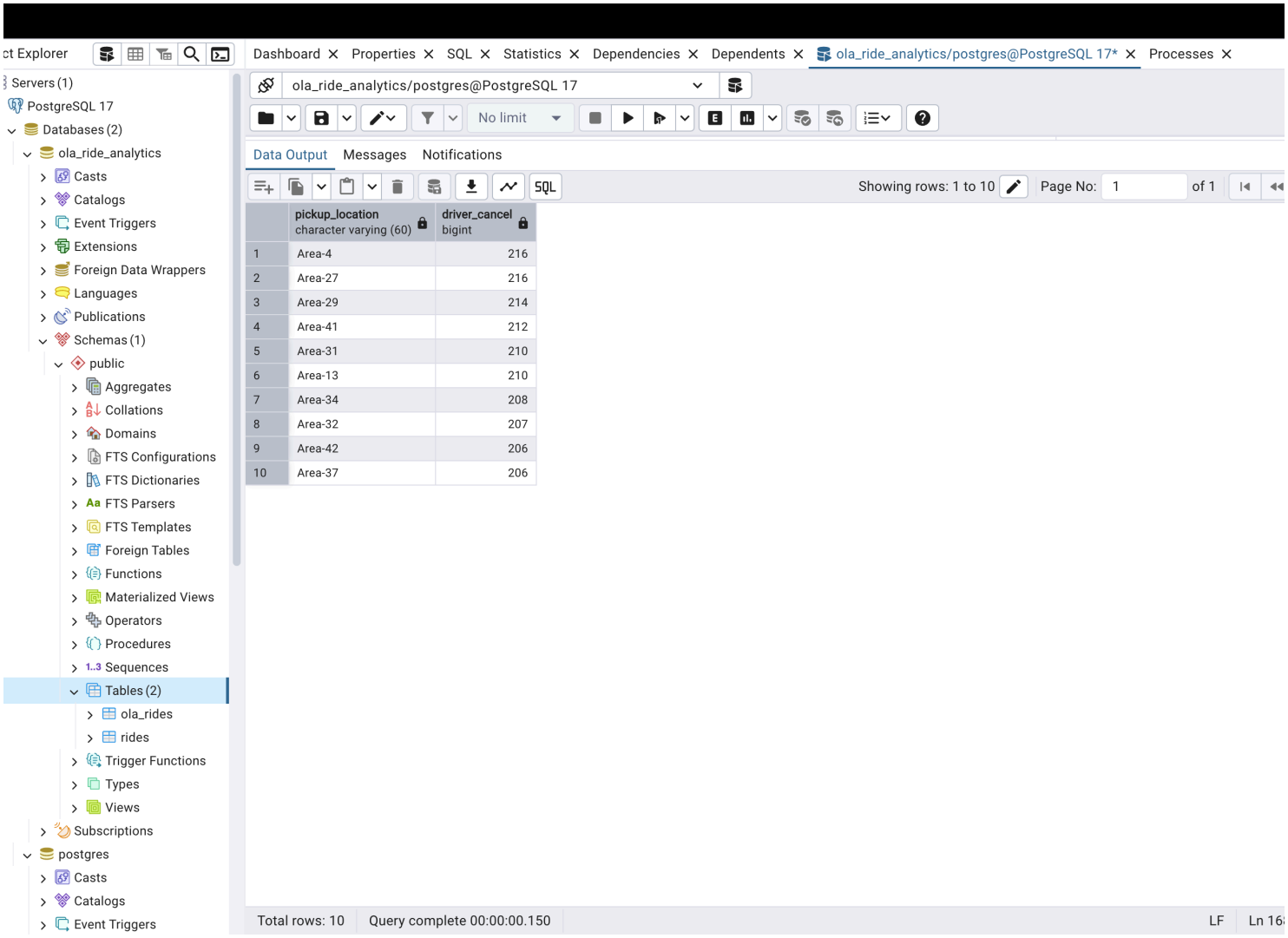
SUM(cancelled_by_driver) AS driver_cancel

FROM rides

GROUP BY pickup_location

ORDER BY driver_cancel DESC

LIMIT 10;

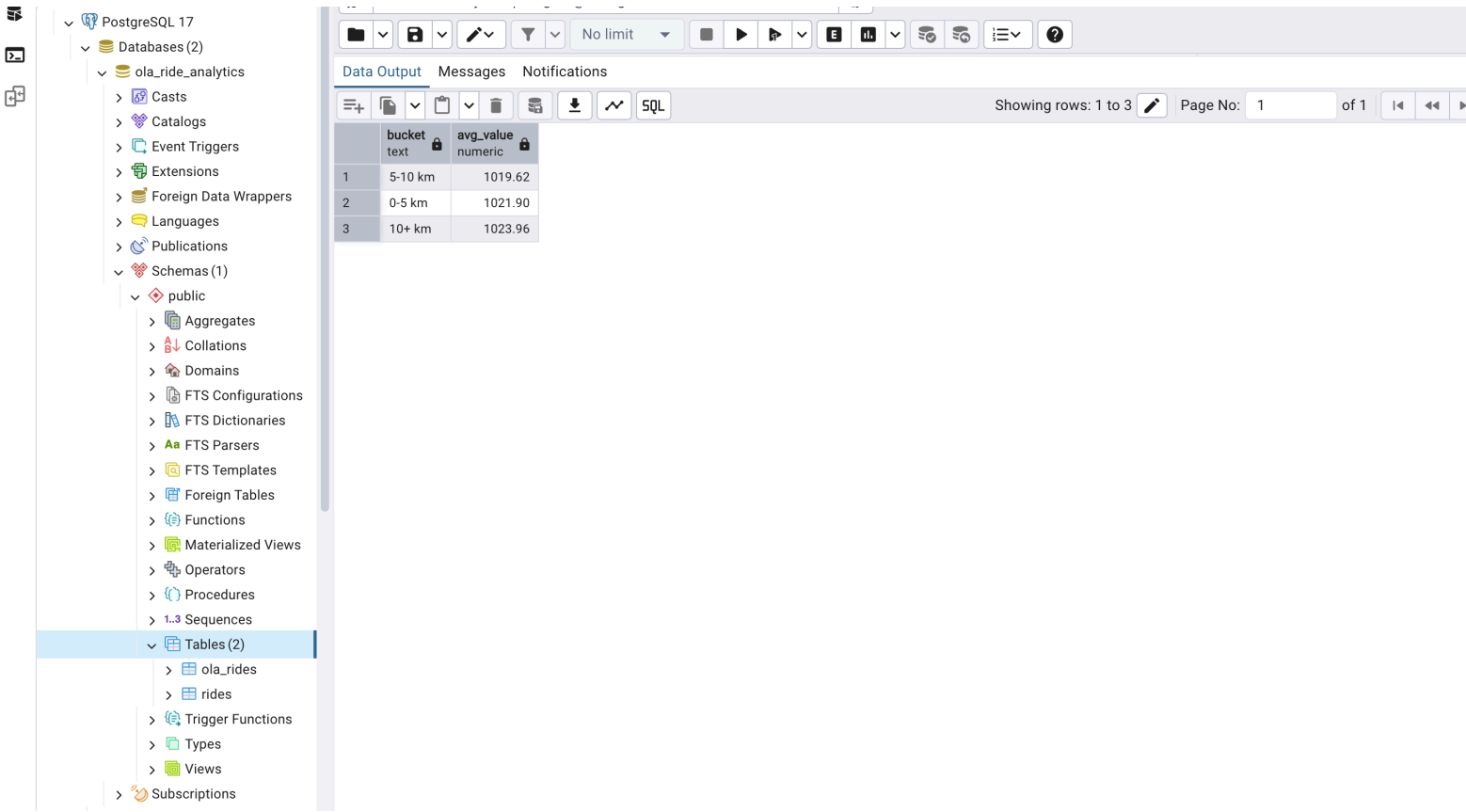


Caption

QUESTION 8: How does booking value vary by ride distance buckets (e.g., 0-5km, 5-10km)?

QUERY- WITH buckets AS (

```
 SELECT CASE
        WHEN ride_distance<5  THEN '0-5 km'
        WHEN ride_distance<10 THEN '5-10 km'
        ELSE '10+ km' END AS bucket,
       booking_value
 FROM rides
 WHERE booking_status='Success')
SELECT bucket,
     ROUND(AVG(booking_value),2) AS avg_value
FROM buckets
GROUP BY bucket;
```



Caption

QUESTION 9: What is the most common payment method for successful bookings?

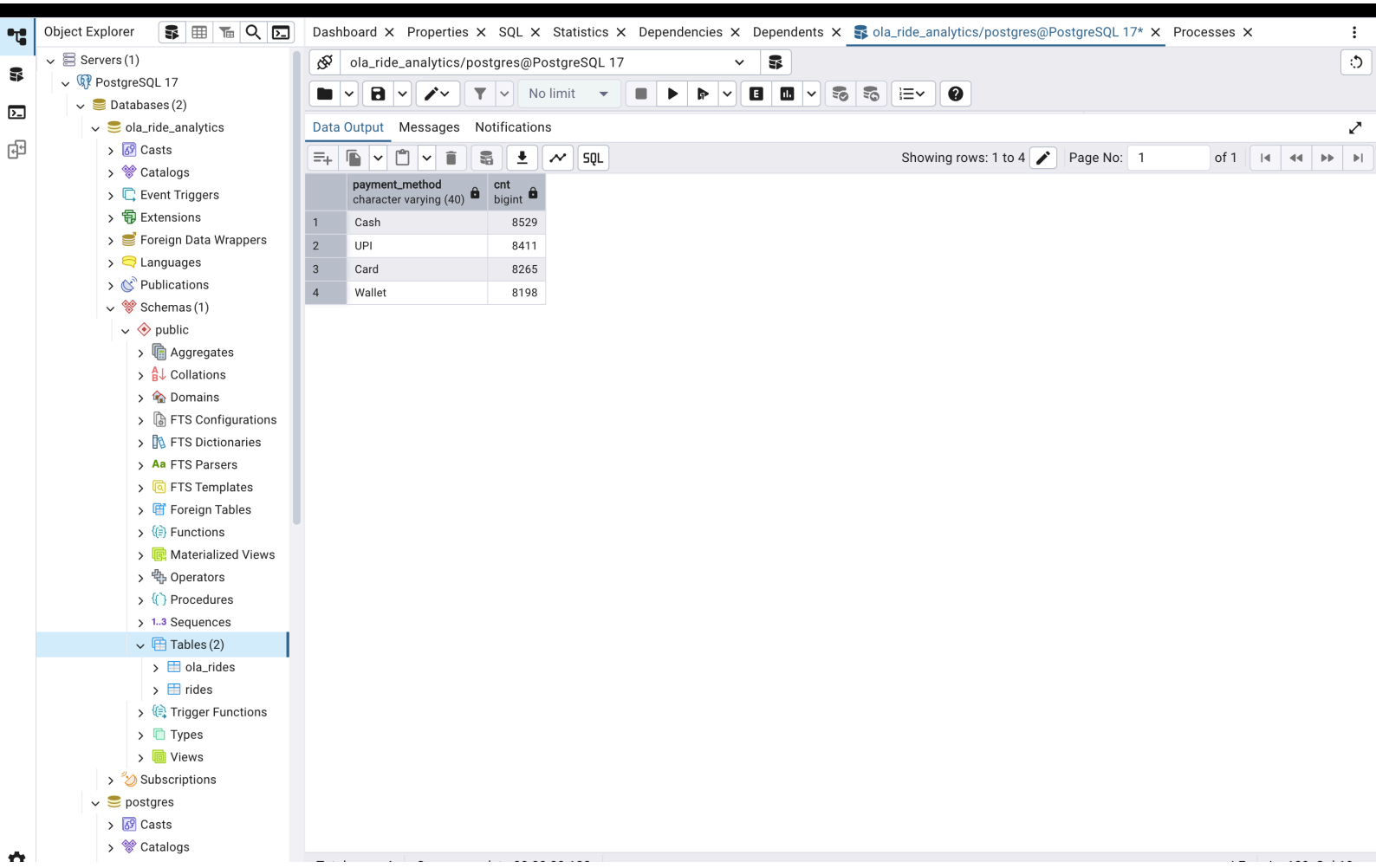
QUERY- SELECT payment_method, COUNT(*) AS cnt

FROM rides

WHERE booking_status='Success'

GROUP BY payment_method

ORDER BY cnt DESC;



Caption

QUESTION 10: Top 5 day which had the highest revenue generated ?

QUERY- SELECT CAST(ride_timestamp AS date) AS ride_day,

ROUND(SUM(booking_value),2) AS day_revenue

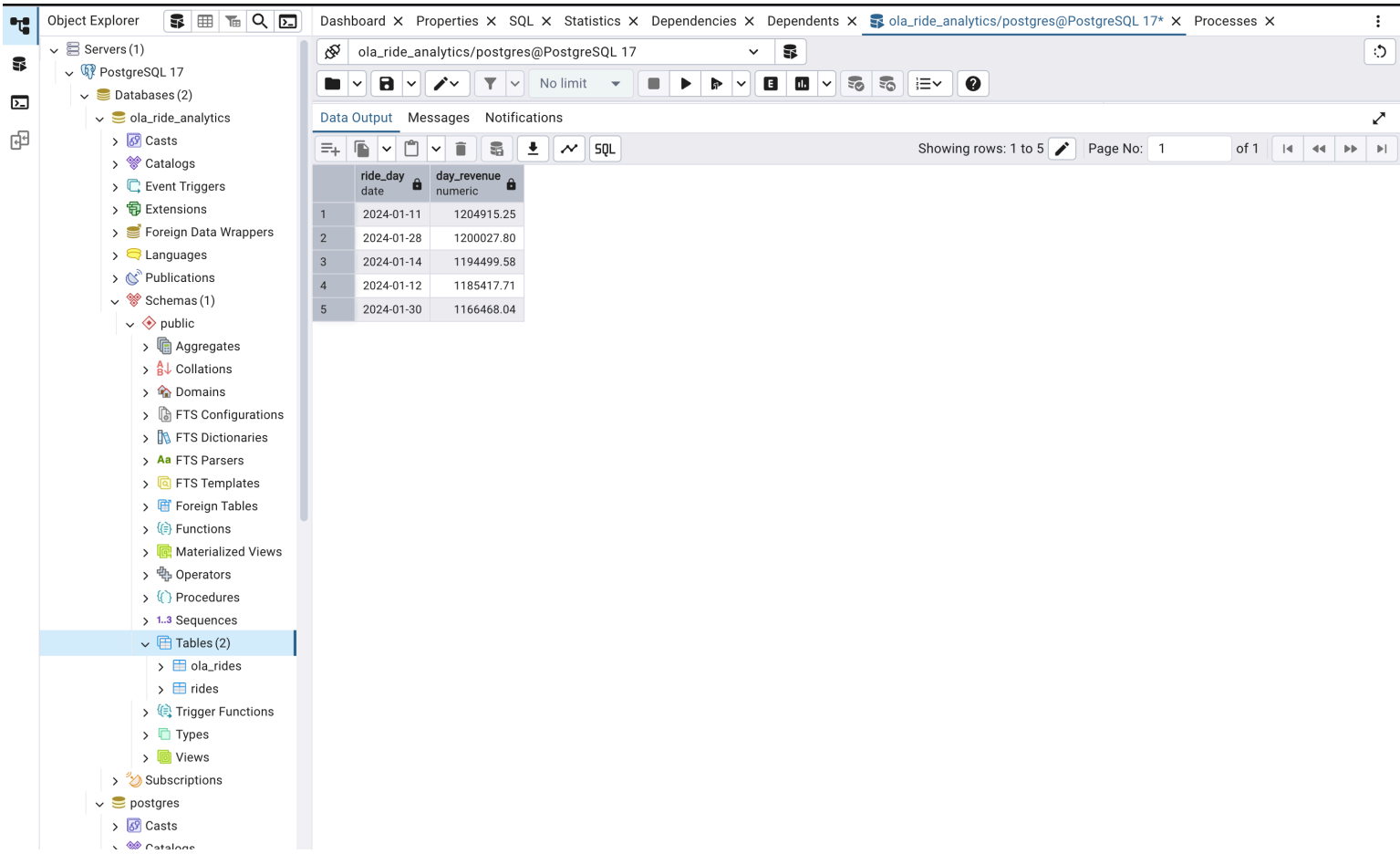
FROM rides

WHERE booking_status = 'Success'

GROUP BY ride_day

ORDER BY day_revenue DESC

LIMIT 5;



Caption

QUESTION 11. Which customers have the most successful rides?

QUERY-SELECT customer_id, COUNT(*) AS rides

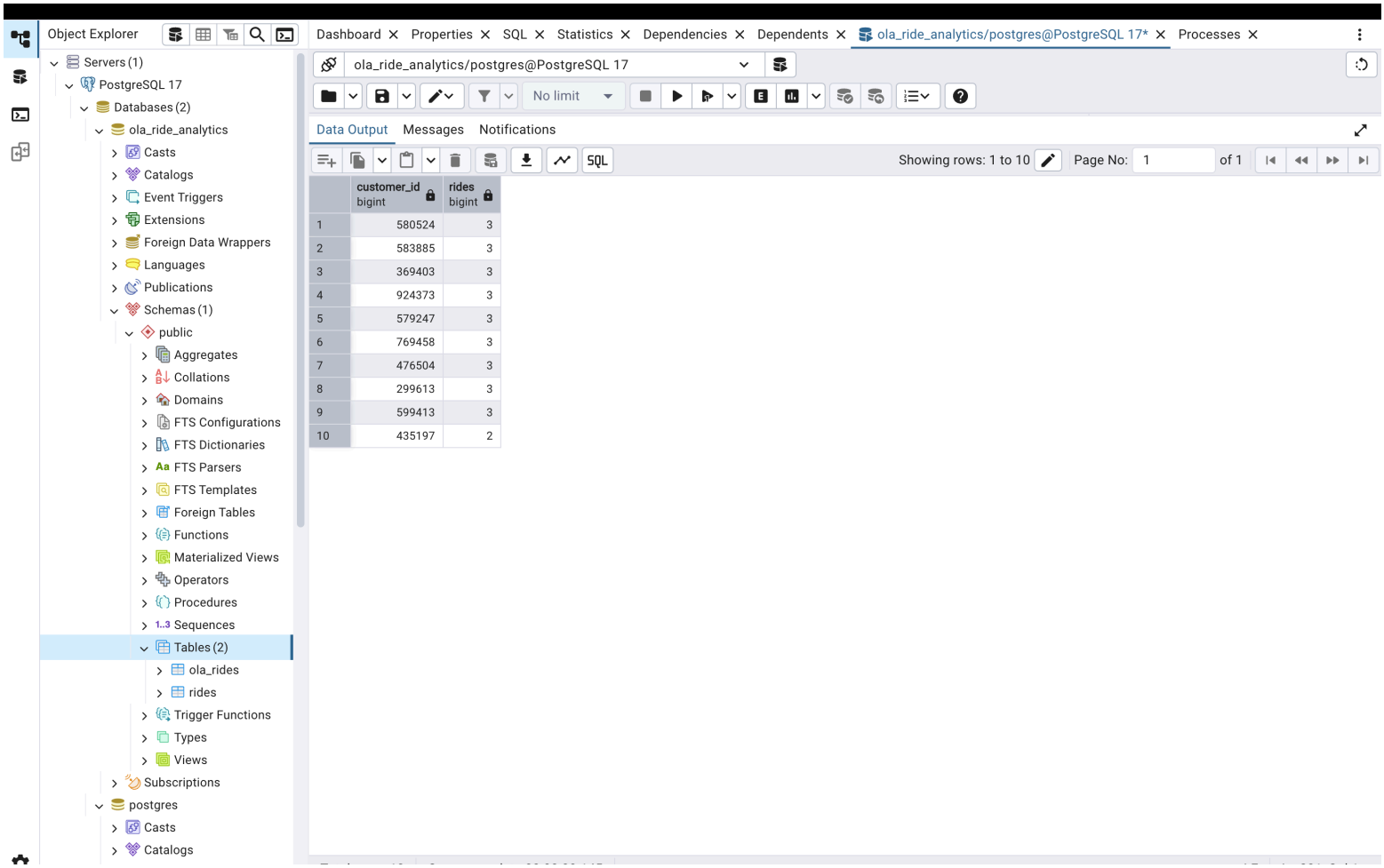
FROM rides

WHERE booking_status='Success'

GROUP BY customer_id

ORDER BY rides DESC

LIMIT 10;



Caption

QUESTION 12. What is the average customer rating by vehicle type?

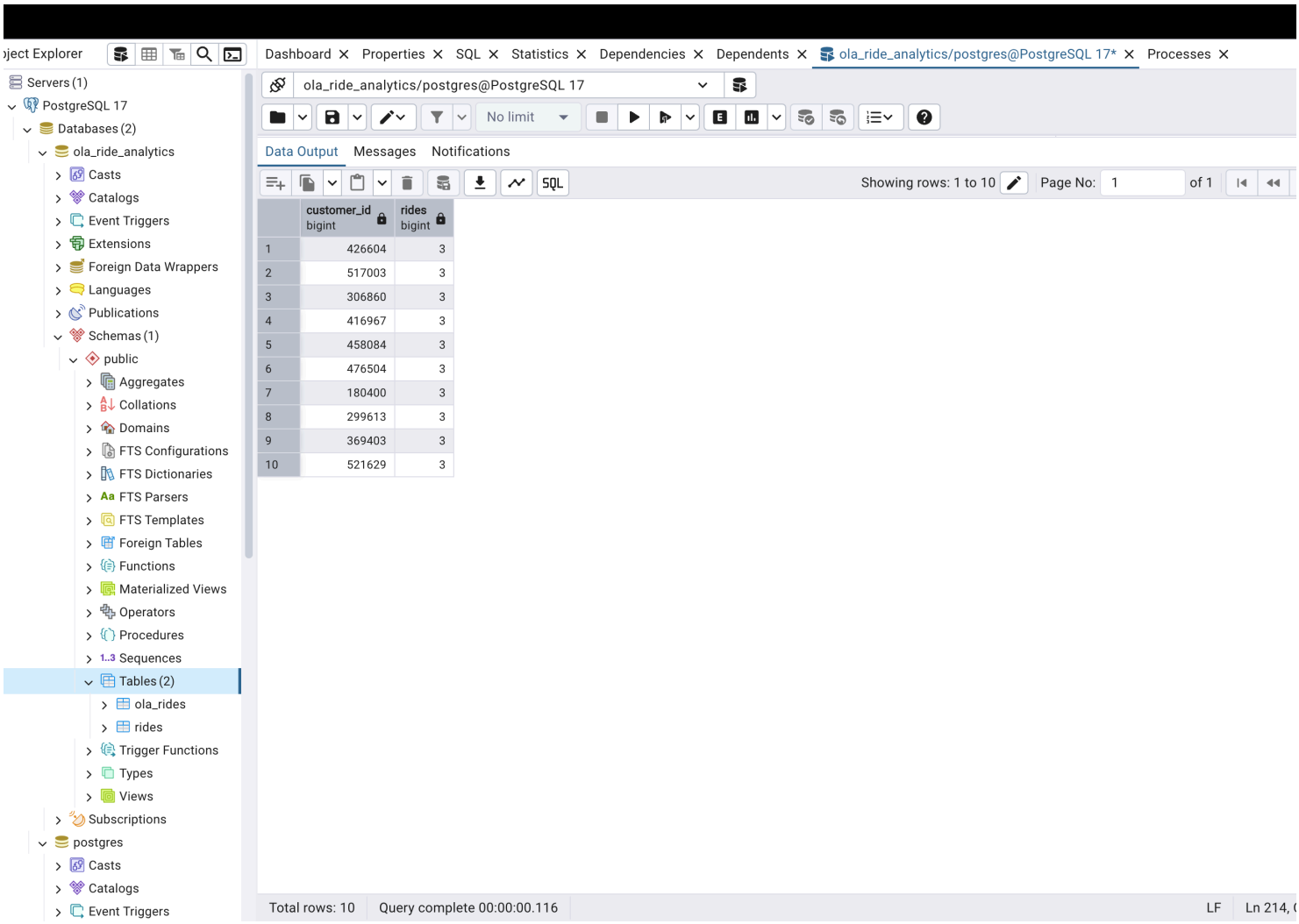
QUERY-SELECT customer_id, COUNT(*) AS rides

FROM rides

GROUP BY customer_id

ORDER BY rides DESC

LIMIT 10;



Caption

QUESTION 13.What is the average customer rating by vehicle type?

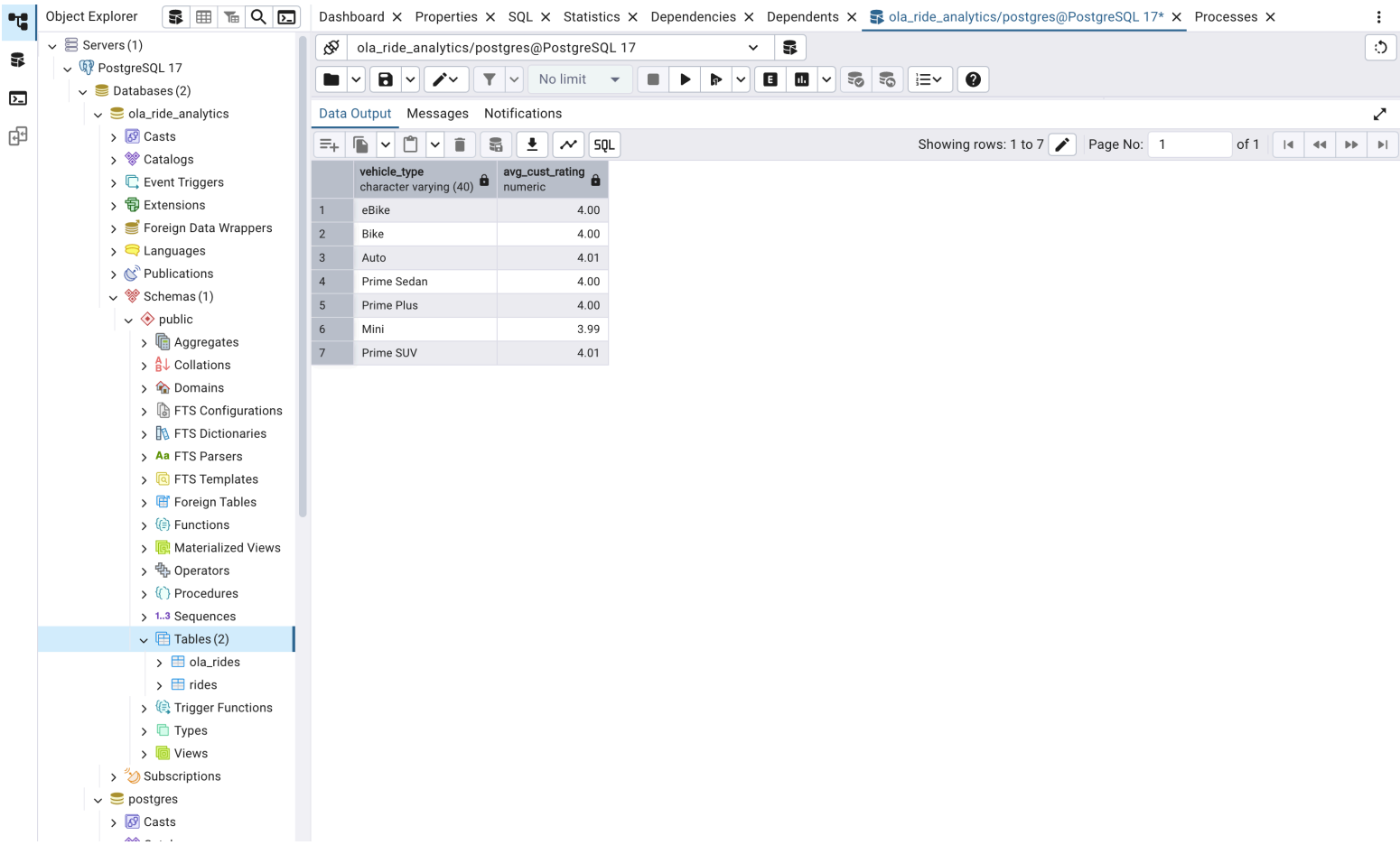
QUERY:- SELECT vehicle_type,

ROUND(AVG(customer_rating),2) AS avg_cust_rating

FROM rides

WHERE customer_rating>0

GROUP BY vehicle_type;



Caption

QUESTION 14: What is the distribution of driver ratings by vehicle type? (e.g., how many below 3.5?)

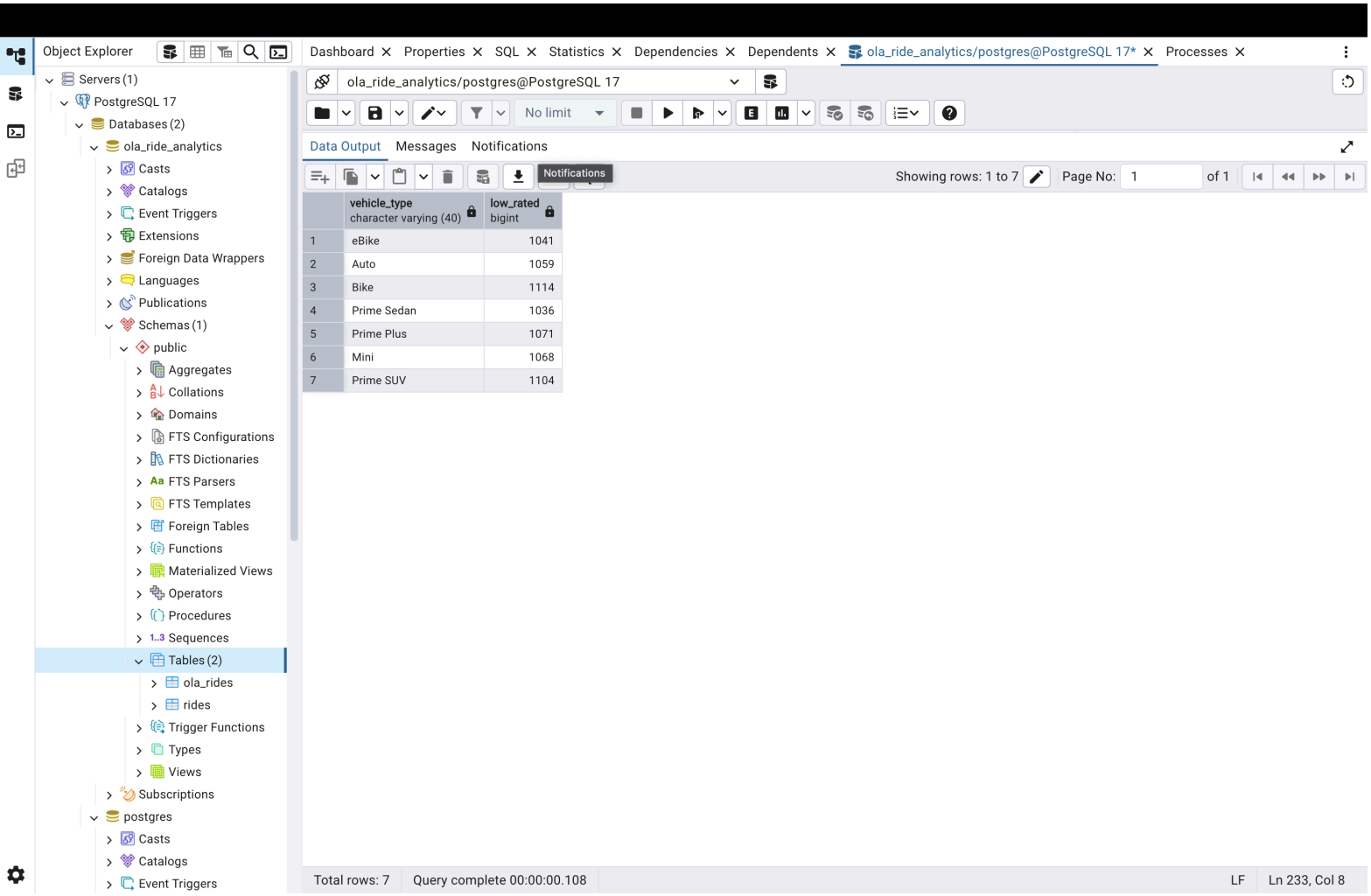
QUERY-SELECT vehicle_type, COUNT(*) AS low_rated

FROM rides

WHERE driver_rating < 3.5

AND driver_rating>0

GROUP BY vehicle_type;

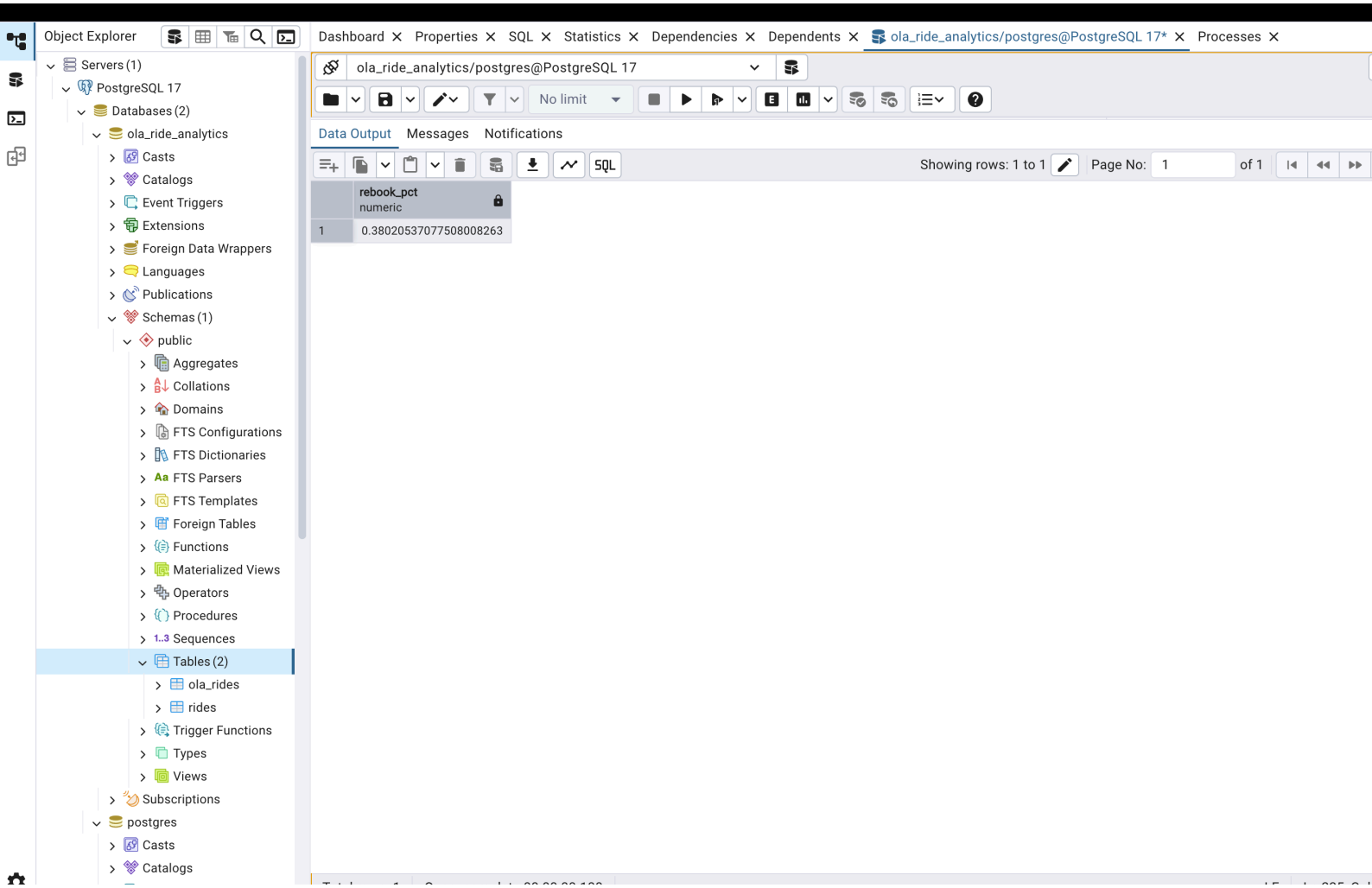


Caption

QUESTION 15: What is the rebooking rate of customers within 3 days of a ride?

QUERY-

```
WITH cte AS (
  SELECT customer_id, ride_timestamp,
       LEAD(ride_timestamp) OVER (PARTITION BY customer_id
                                  ORDER BY ride_timestamp) AS next_ride
                                                           FROM rides
  WHERE booking_status='Success')
SELECT COUNT(*) FILTER (WHERE next_ride IS NOT NULL
                        AND next_ride-ride_timestamp<=INTERVAL '3 day')
      *100.0 / COUNT(*) AS rebook_pct
FROM cte;
```



Caption

QUESTION 16: What is the gap between a customer's first and last ride?

QUERY- SELECT customer_id,

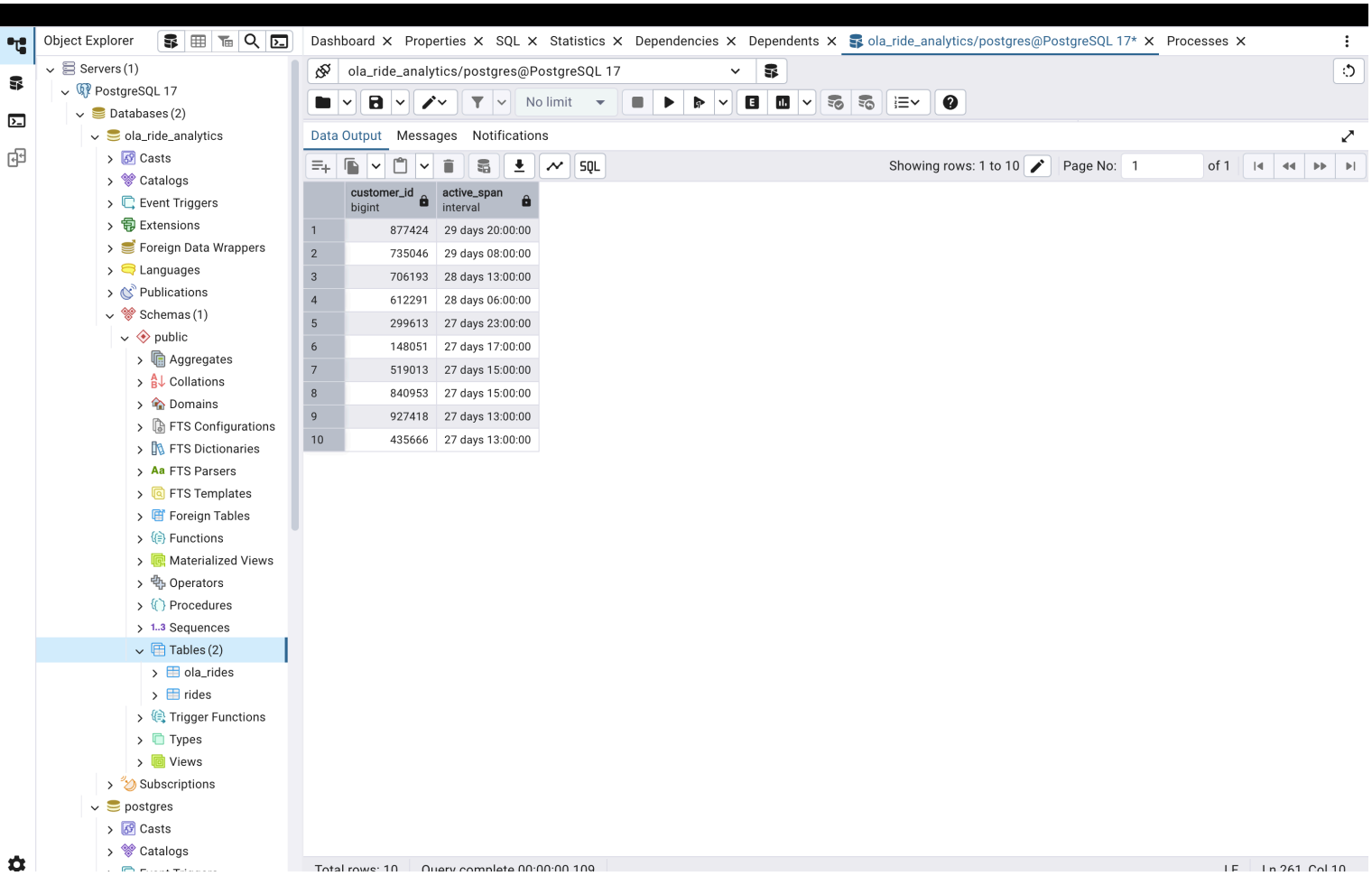
AGE(MAX(ride_timestamp), MIN(ride_timestamp)) AS active_span

FROM rides

GROUP BY customer_id

ORDER BY active_span DESC

LIMIT 10;



Caption

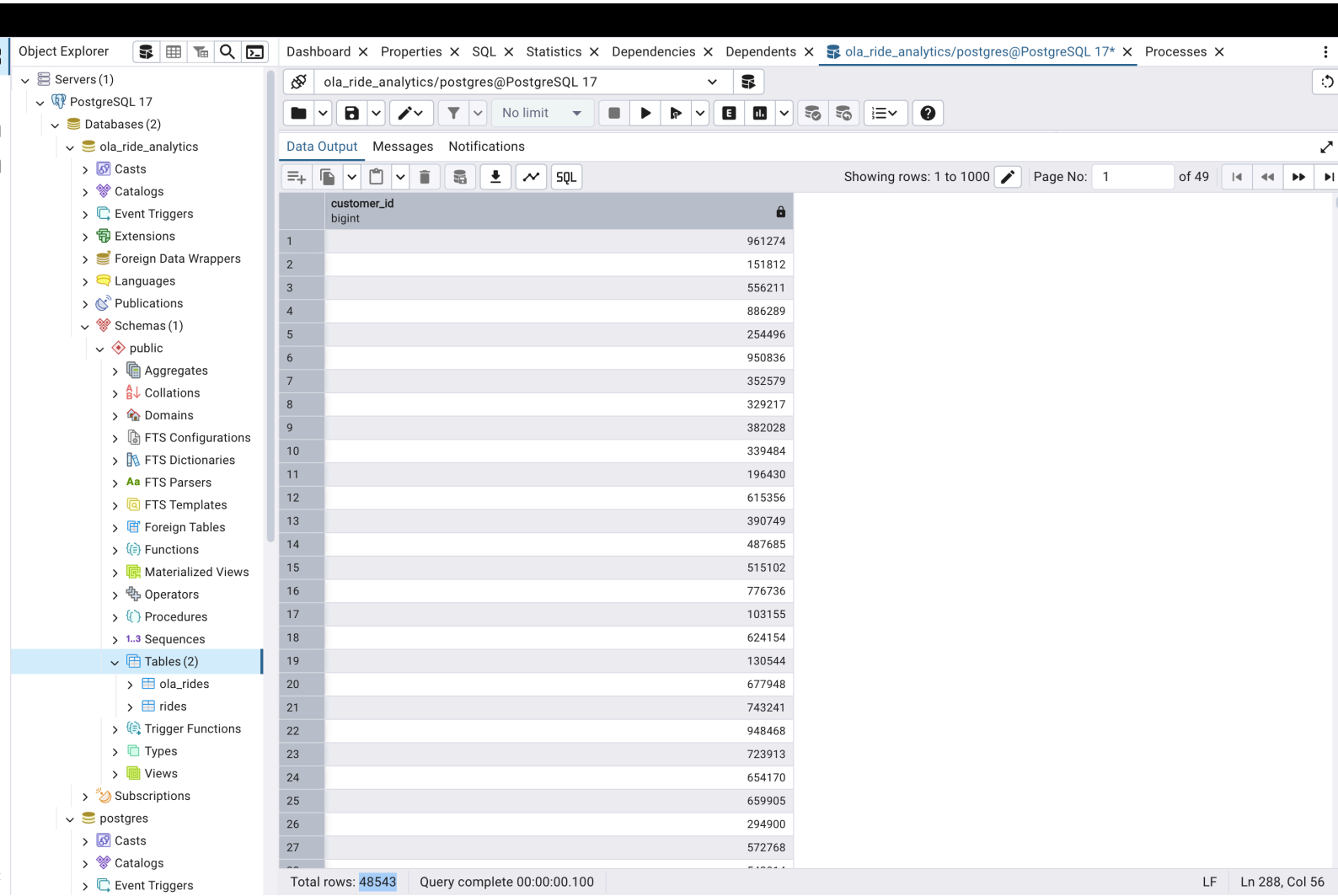
QUESTION 17:Who are the churned customers (no ride in last 60 days)?

QUERY- SELECT customer_id

FROM rides

GROUP BY customer_id

HAVING MAX(ride_timestamp) < NOW() - INTERVAL '60 day';



Caption

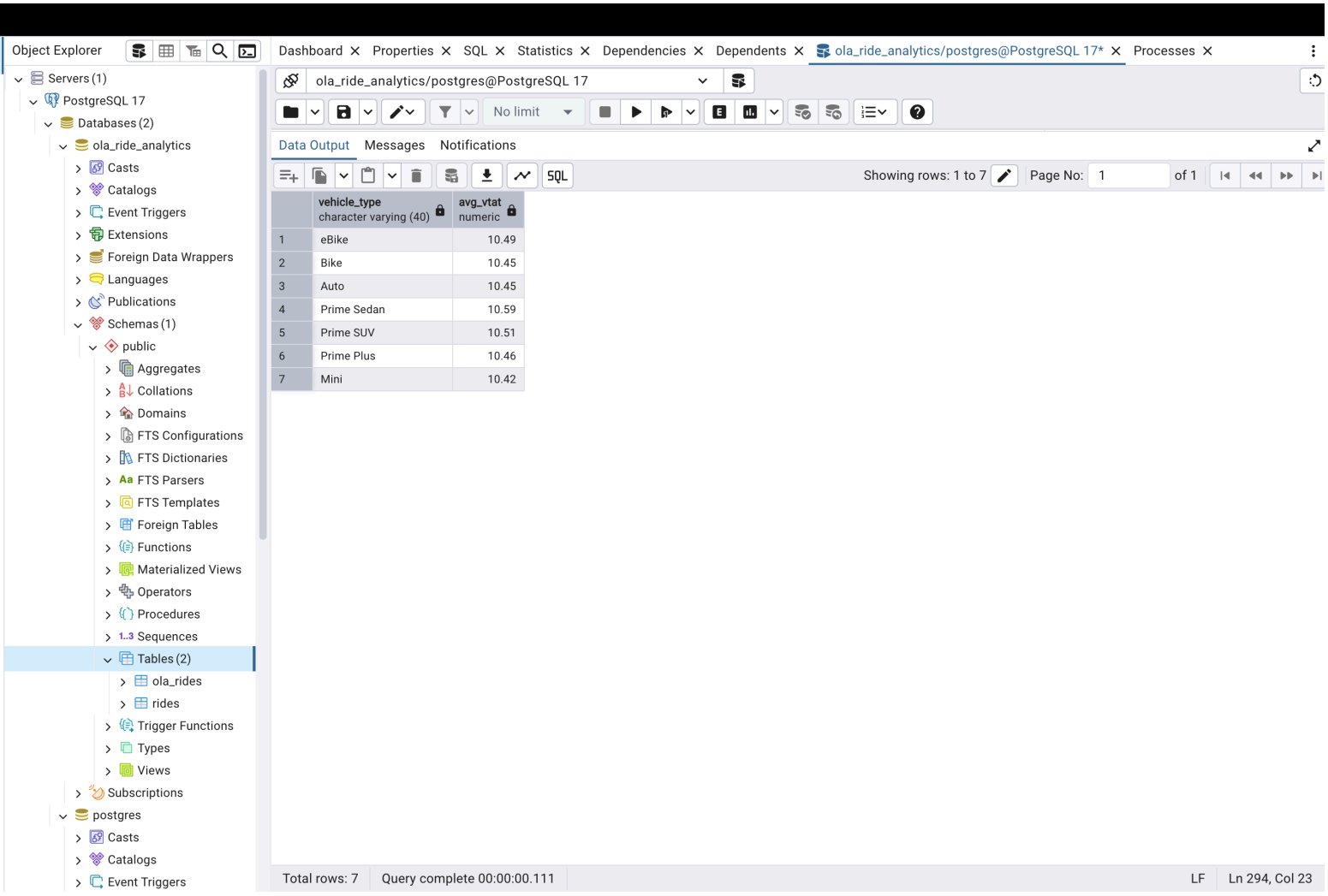
TOTAL- 48543 CUSTOMER

QUESTION 18: Calculate the average VTAT (Vehicle Turnaround Time) per vehicle type.

QUERY- SELECT vehicle_type, ROUND(AVG(avg_vtat),2) AS avg_vtat

FROM rides

GROUP BY vehicle_type;



Caption

QUESTION 19: Calculate average CTAT (Customer Turnaround Time) across time slots.(4 hr gap)

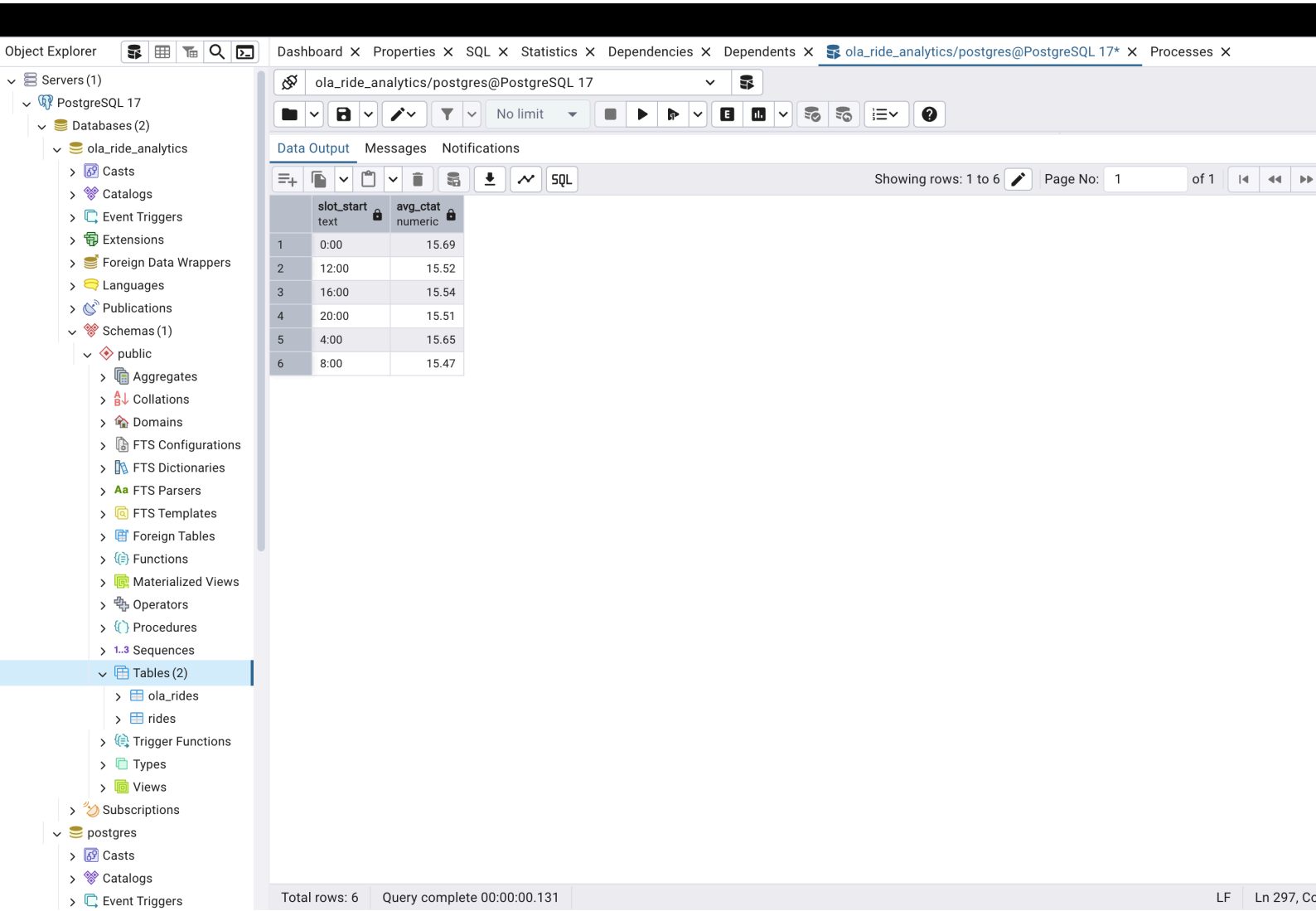
QUERY - SELECT CONCAT(FLOOR(EXTRACT(hour FROM ride_timestamp)/4)*4,':00') AS slot_start,

ROUND(AVG(avg_ctat),2) AS avg_ctat

FROM rides

GROUP BY slot_start

ORDER BY slot_start;



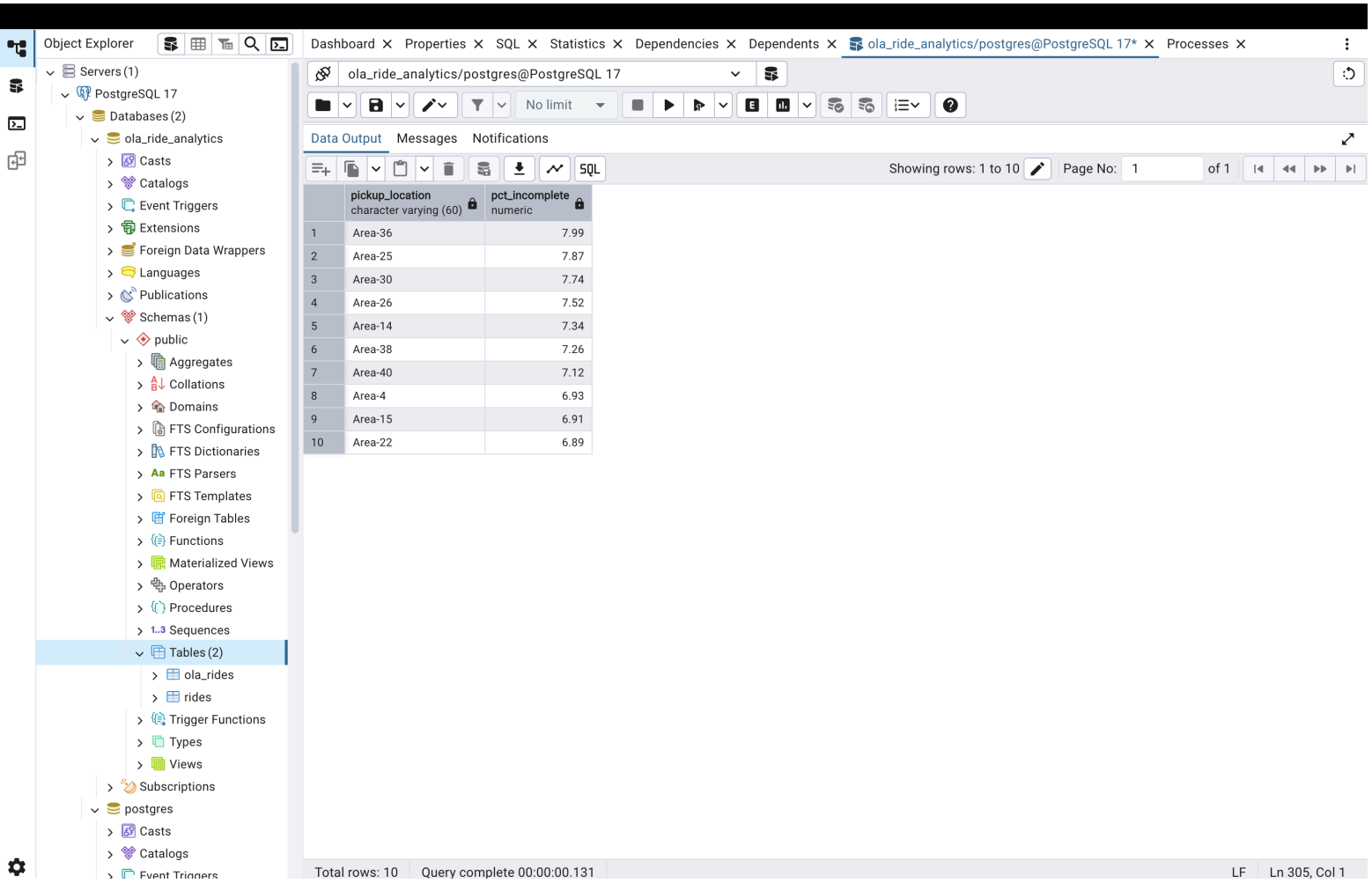
Caption

QUESTION 20: What is the percentage of incomplete rides overall and by location?

QUERY-

```
SELECT ROUND(SUM(incomplete_rides)::numeric / COUNT(*)*100,2) AS pct_incomplete
FROM rides;

SELECT pickup_location,
    ROUND(SUM(incomplete_rides)::numeric / COUNT(*)*100,2) AS pct_incomplete
FROM rides
GROUP BY pickup_location
ORDER BY pct_incomplete DESC
LIMIT 10;
```



Caption

QUESTION 21: Identify locations with highest ratio of failed to successful rides.

QUERY- SELECT pickup_location,

SUM(CASE WHEN booking_status!='Success' THEN 1 END)::float /

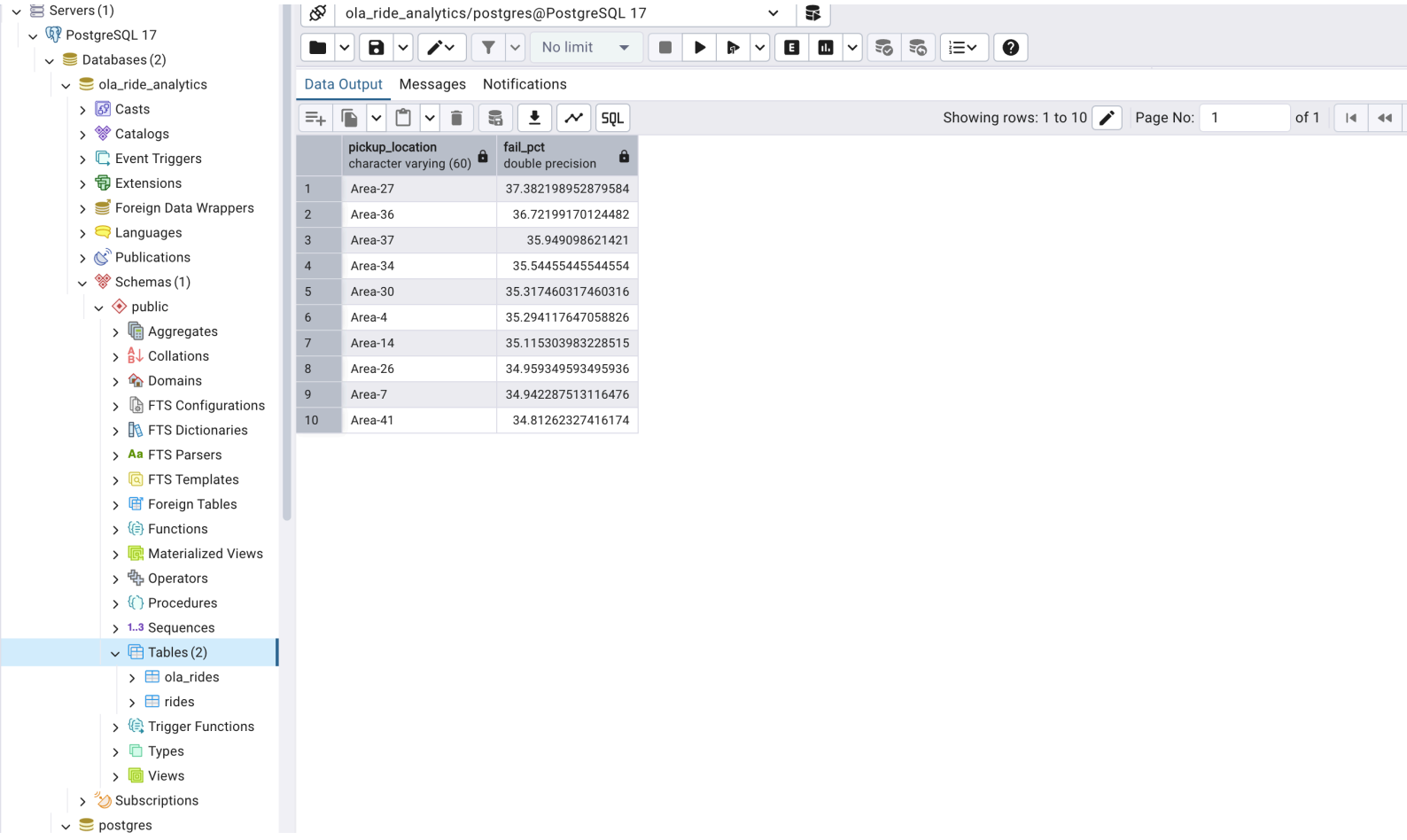
COUNT(*) *100 AS fail_pct

FROM rides

GROUP BY pickup_location

ORDER BY fail_pct DESC

LIMIT 10;



	pickup_location character varying (60)	fail_pct double precision
1	Area-27	37.382198952879584
2	Area-36	36.72199170124482
3	Area-37	35.949098621421
4	Area-34	35.54455445544554
5	Area-30	35.317460317460316
6	Area-4	35.294117647058826
7	Area-14	35.115303983228515
8	Area-26	34.959349593495936
9	Area-7	34.942287513116476
10	Area-41	34.81262327416174

Caption

QUESTION 22: Identify customers with revenue above the 90th percentile.

QUERY-WITH tot AS (

SELECT customer_id, SUM(booking_value) AS revenue

FROM rides

WHERE booking_status = 'Success'

GROUP BY customer_id),

percentile_90 AS (

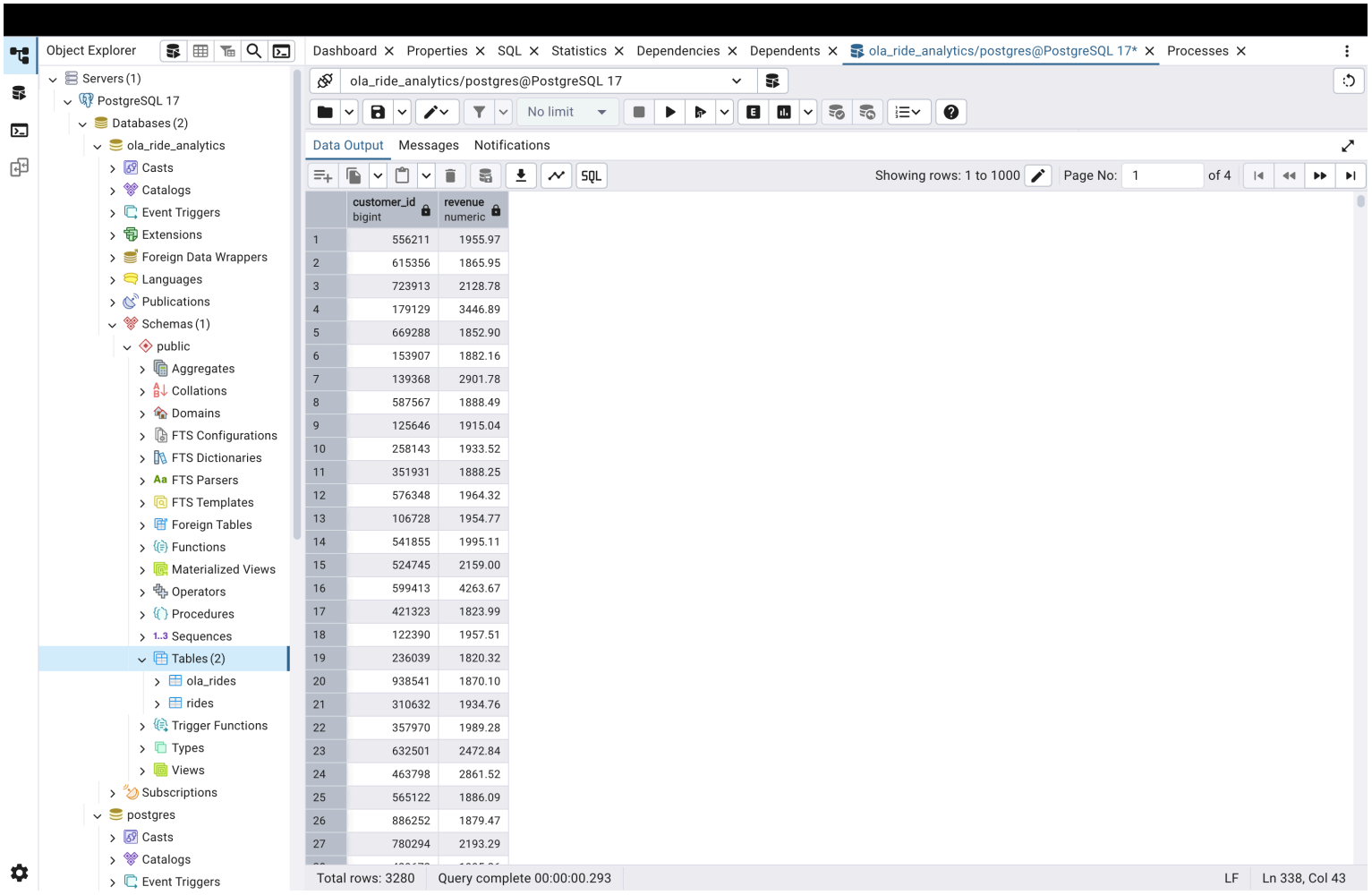
SELECT PERCENTILE_CONT(0.9) WITHIN GROUP (ORDER BY revenue) AS p90

FROM tot)

SELECT t.*

FROM tot t

JOIN percentile_90 p ON t.revenue > p.p90;



Caption

QUESTION 23:Find cumulative revenue day by day for the current month.

QUERY- WITH daily AS (SELECT CAST(ride_timestamp AS date) AS day,

SUM(booking_value) AS re

FROM rides

WHERE booking_status = 'Success'

GROUP BY day)

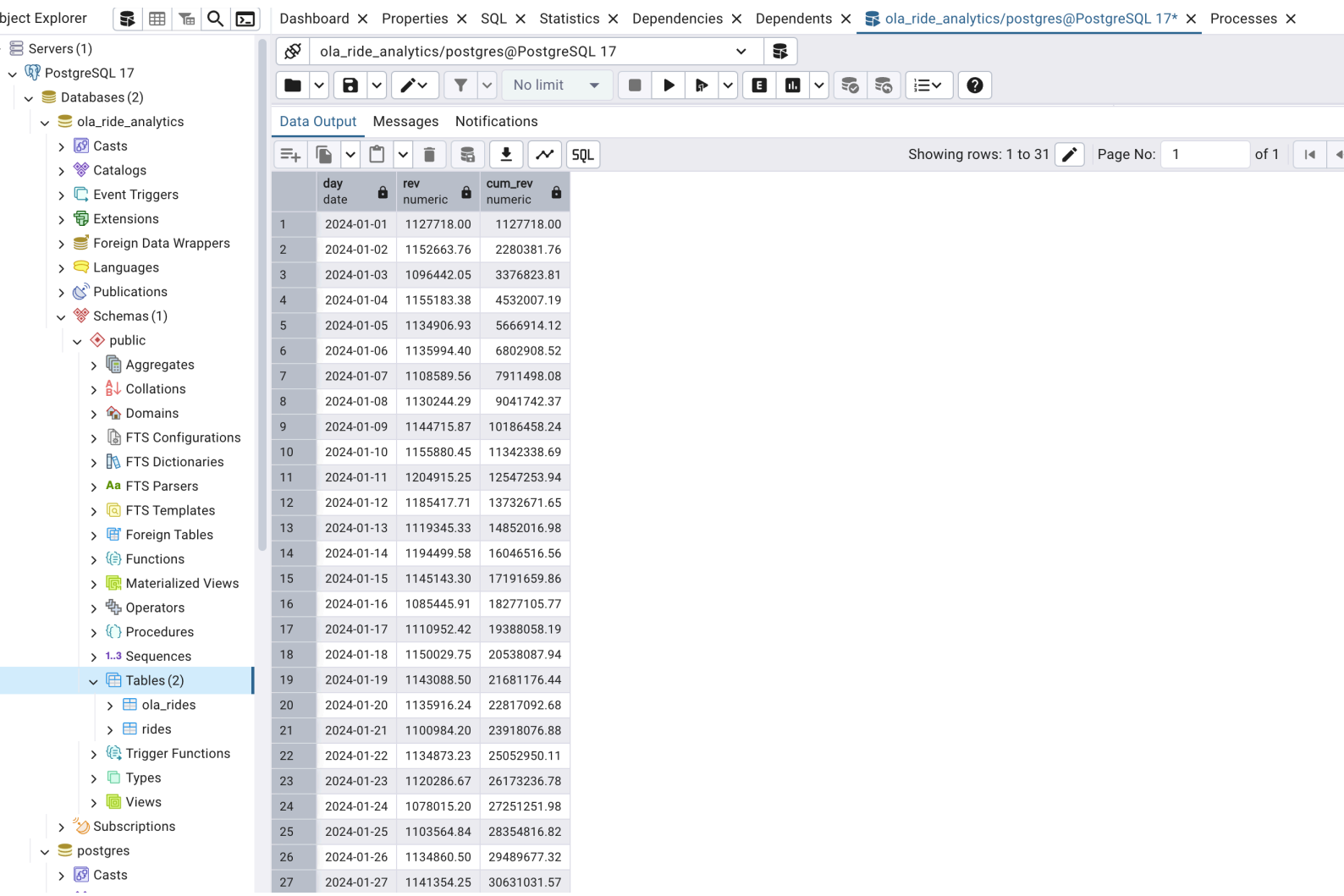
SELECT day,

rev,

SUM(rev) OVER (ORDER BY day) AS cum_rev

FROM daily

ORDER BY day;



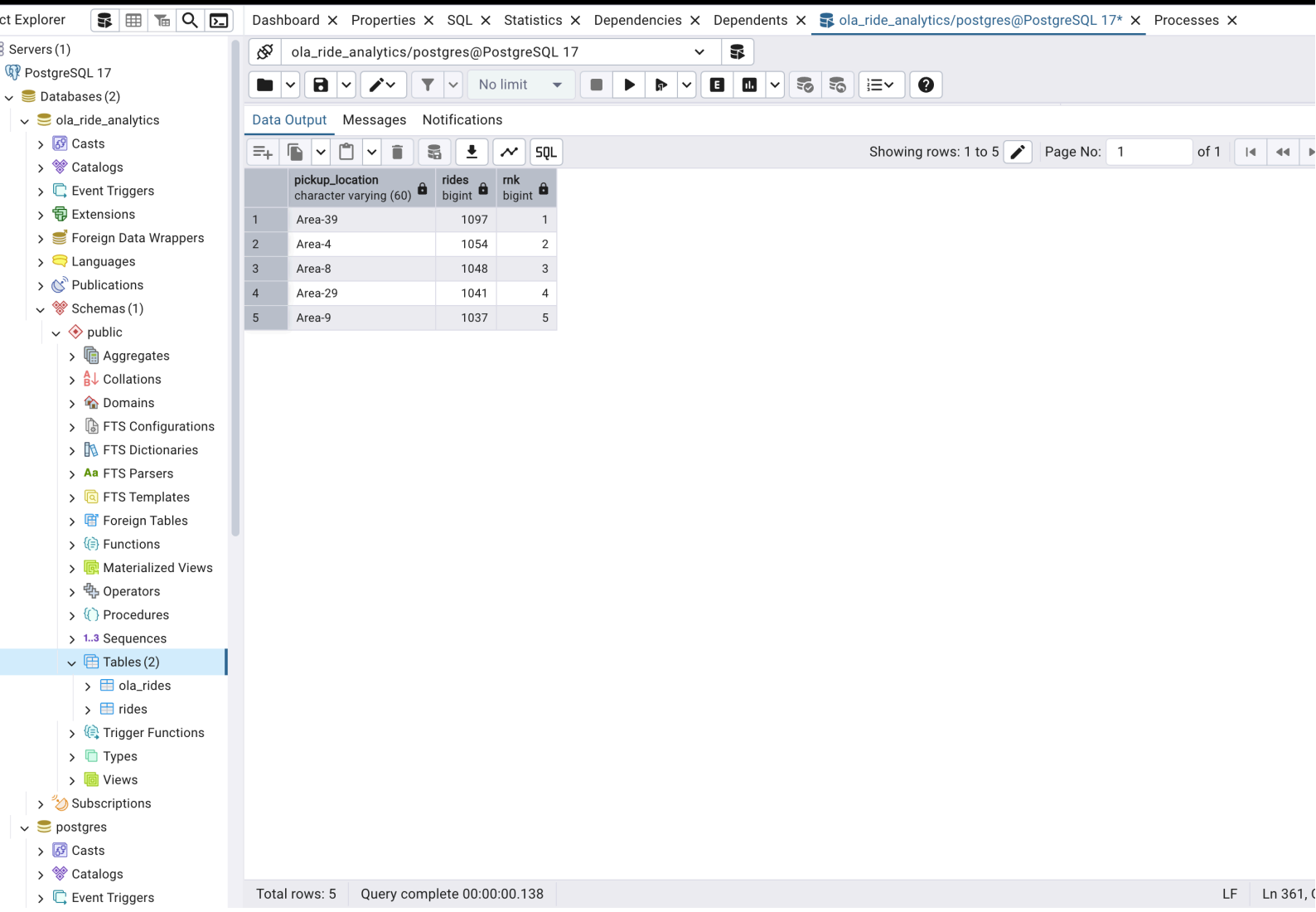
	day date	rev numeric	cum_rev numeric
1	2024-01-01	1127718.00	1127718.00
2	2024-01-02	1152663.76	2280381.76
3	2024-01-03	1096442.05	3376823.81
4	2024-01-04	1155183.38	4532007.19
5	2024-01-05	1134906.93	5666914.12
6	2024-01-06	1135994.40	6802908.52
7	2024-01-07	1108589.56	7911498.08
8	2024-01-08	1130244.29	9041742.37
9	2024-01-09	1144715.87	10186458.24
10	2024-01-10	1155880.45	11342338.69
11	2024-01-11	1204915.25	12547253.94
12	2024-01-12	1185417.71	13732671.65
13	2024-01-13	1119345.33	14852016.98
14	2024-01-14	1194499.58	16046516.56
15	2024-01-15	1145143.30	17191659.86
16	2024-01-16	1085445.91	18277105.77
17	2024-01-17	1110952.42	19388058.19
18	2024-01-18	1150029.75	20538087.94
19	2024-01-19	1143088.50	21681176.44
20	2024-01-20	1135916.24	22817092.68
21	2024-01-21	1100984.20	23918076.88
22	2024-01-22	1134873.23	25052950.11
23	2024-01-23	1120286.67	26173236.78
24	2024-01-24	1078015.20	27251251.98
25	2024-01-25	1103564.84	28354816.82
26	2024-01-26	1134860.50	29489677.32
27	2024-01-27	1141354.25	30631031.57

Caption

QUESTION 24:Rank pickup locations by ride frequency and show top 5.

QUERY-

```
SELECT pickup_location,
    COUNT(*) AS rides,
    RANK() OVER (ORDER BY COUNT(*) DESC) AS rnk
FROM rides
GROUP BY pickup_location
ORDER BY rnk
LIMIT 5;
```



Caption

QUESTION 25:Calculate monthly average booking value per customer.

QUERY- SELECT DATE_TRUNC('month',ride_timestamp) AS month,

customer_id,

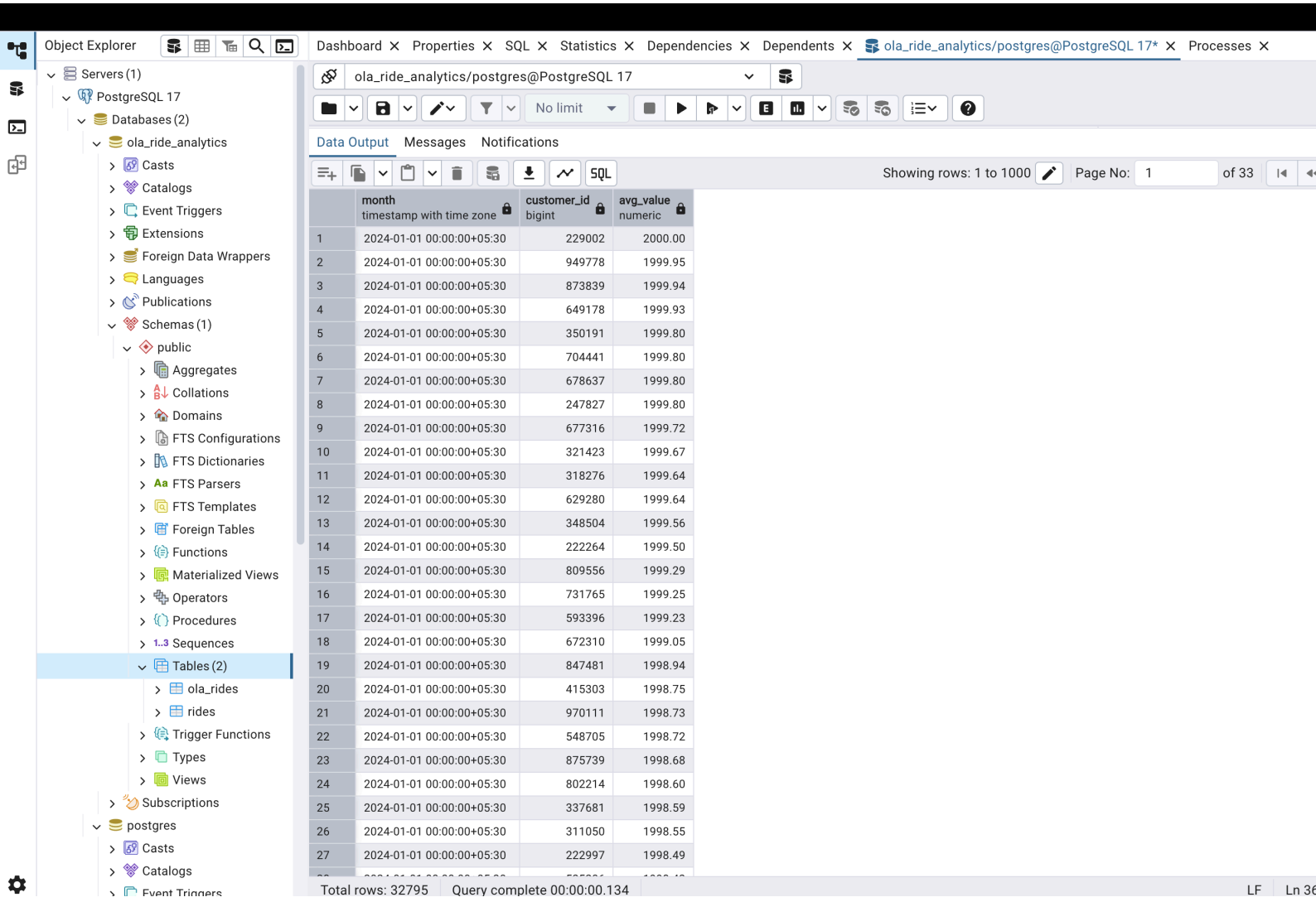
ROUND(AVG(booking_value),2) AS avg_value

FROM rides

WHERE booking_status='Success'

GROUP BY month, customer_id

ORDER BY month, avg_value DESC;



Caption

TOTAL ROWS =32795

QUESTION 26