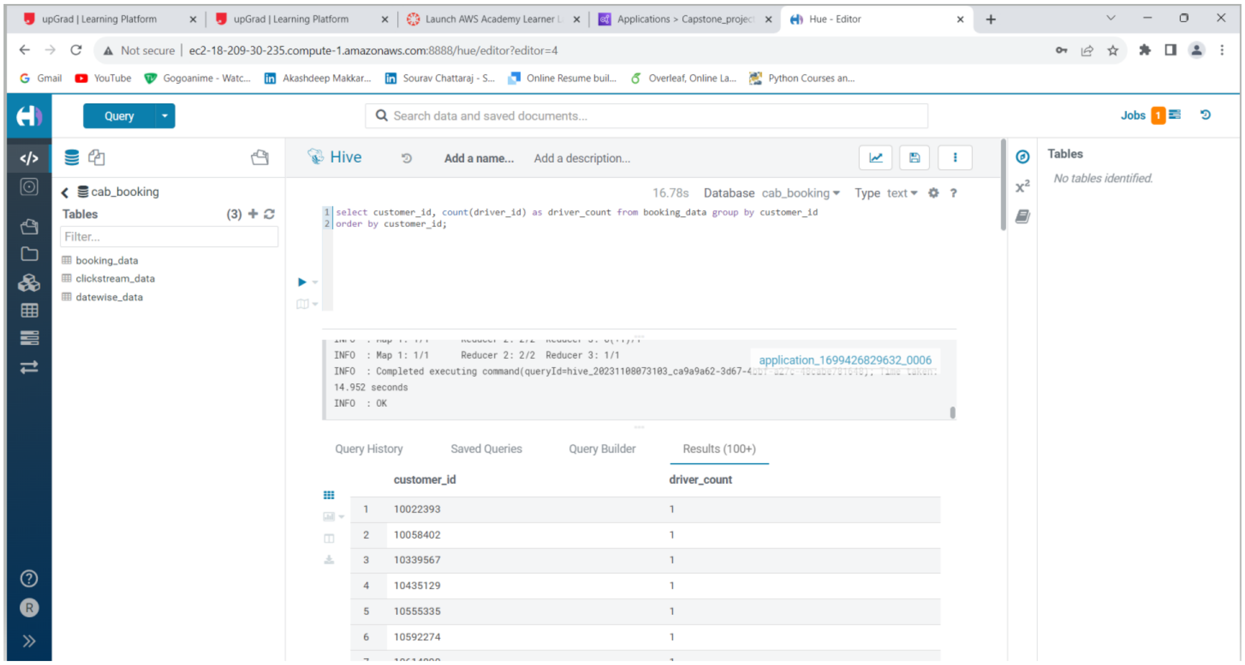
Queries

1. Hive Query for Task 5:- Calculate the total different drivers for each customer.

select customer_id, count(driver_id) as driver_count from booking_data group by customer_id order by customer_id;

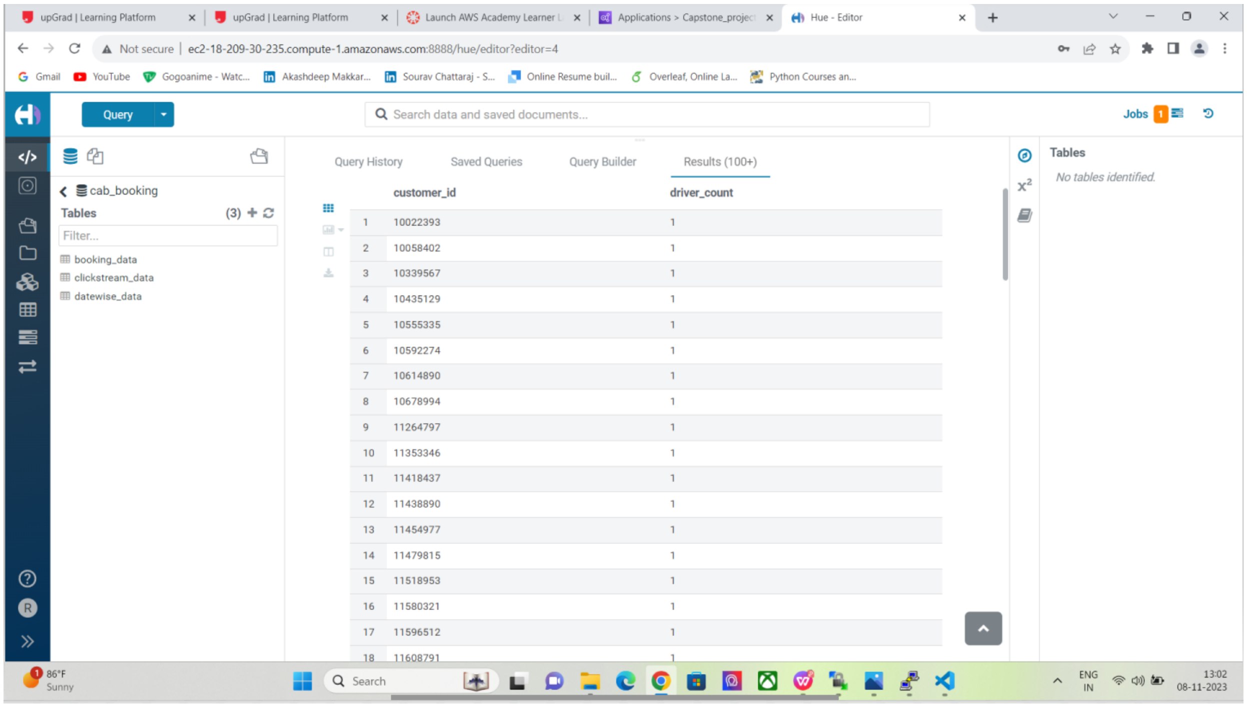
Screenshot after executing Query:-



```
select customer_id, count(driver_id) as driver_count from booking_data group by customer_id
order by customer_id;
```

```
INFO  : Map 1: 1/1      Reducer 2: 2/2  Reducer 3: 1/1
INFO  : Completed executing command(queryId=hive_20231108073103_ca9a9a62-3d67-4bbf-a27c-48cabe781648); Time taken:
14.952 seconds
INFO  : OK
```

	customer_id	driver_count
1	10022393	1
2	10058402	1
3	10339567	1
4	10435129	1
5	10555335	1
6	10592274	1

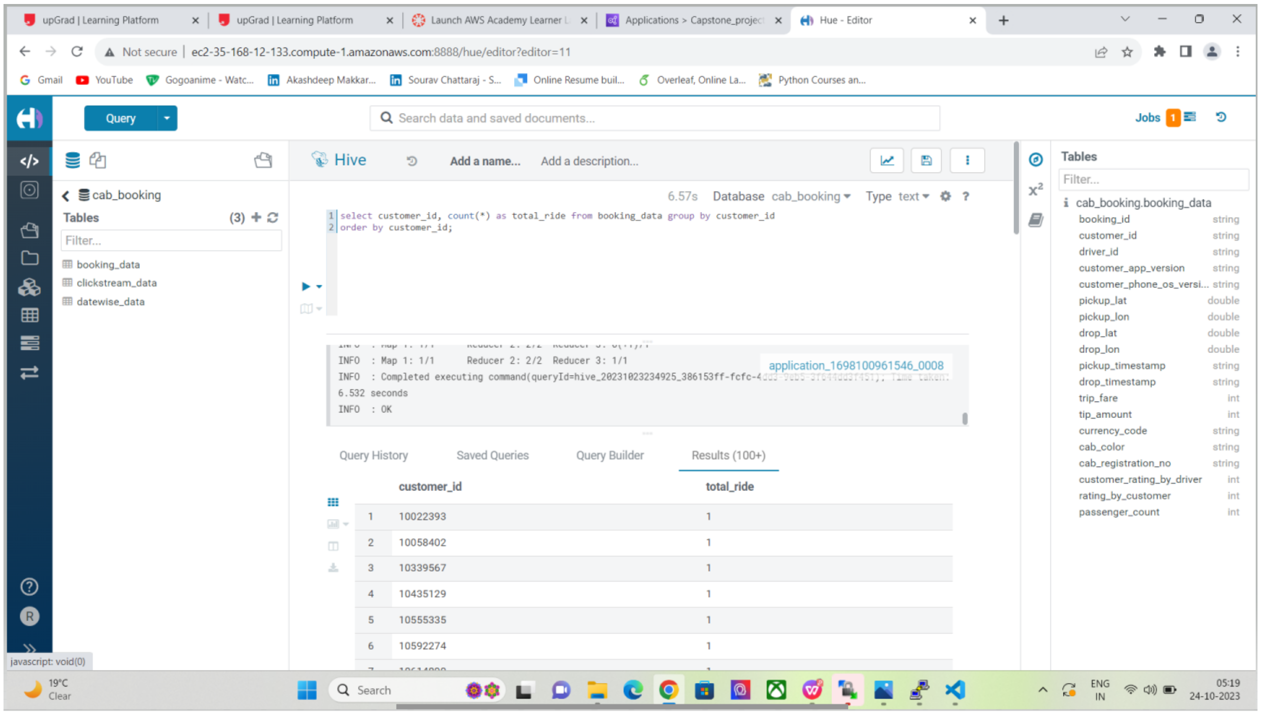


	customer_id	driver_count
1	10022393	1
2	10058402	1
3	10339567	1
4	10435129	1
5	10555335	1
6	10592274	1
7	10614890	1
8	10678994	1
9	11264797	1
10	11353346	1
11	11418437	1
12	11438890	1
13	11454977	1
14	11479815	1
15	11518953	1
16	11580321	1
17	11596512	1
18	11608791	1

2. Hive Query for Task 6:- Calculate the total rides taken by each customer.

select customer_id, count(*) as total_ride from booking_data group by customer_id order by customer_id;

Screenshot after executing Query:-



```
select customer_id, count(*) as total_ride from booking_data group by customer_id
order by customer_id;
```

```
INFO  : Map 1: 1/1      Reducer 2: 2/2  Reducer 3: 1/1
INFO  : Completed executing command(queryId=hive_20231023234925_386153ff-fcfc-4dd3-9eb5-3f644dd3f451); Time taken:
6.532 seconds
INFO  : OK
```

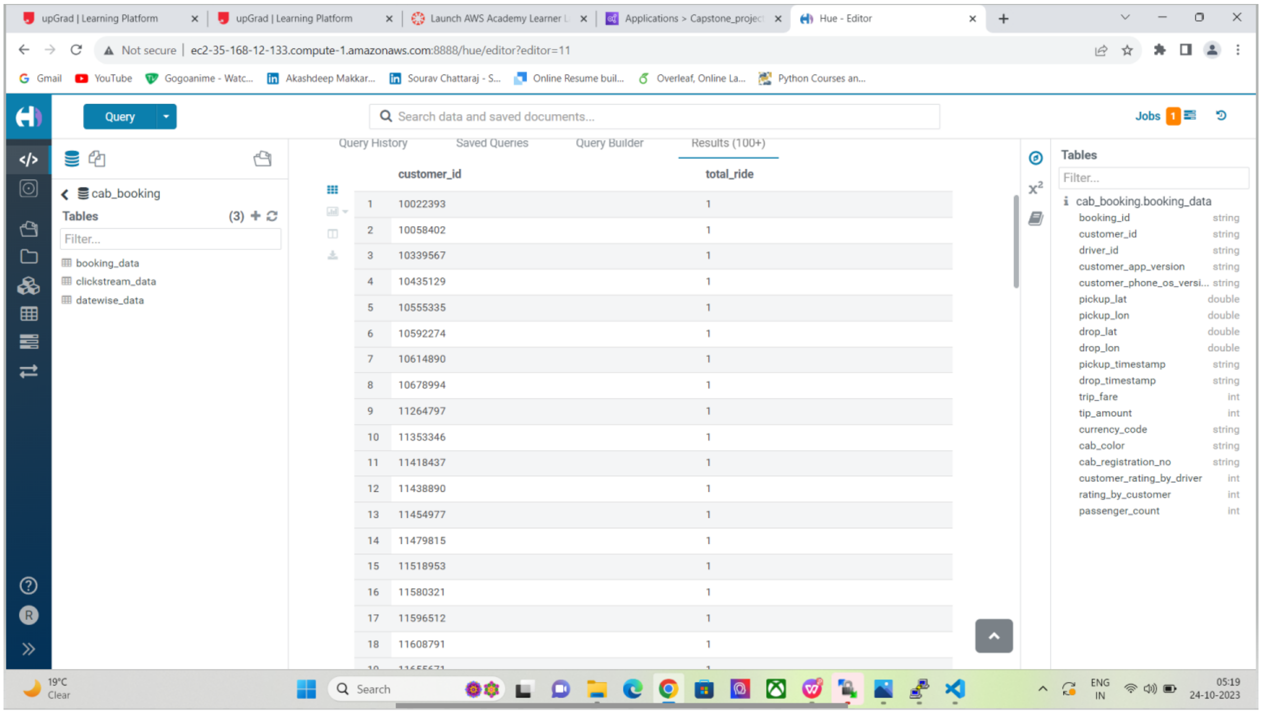
application_1698100961546_0008

	customer_id	total_ride
1	10022393	1
2	10058402	1
3	10339567	1
4	10435129	1
5	10555335	1
6	10592274	1

Tables

cab_booking.booking_data

booking_id	string
customer_id	string
driver_id	string
customer_app_version	string
customer_phone_os_versi...	string
pickup_lat	double
pickup_lon	double
drop_lat	double
drop_lon	double
pickup_timestamp	string
drop_timestamp	string
trip_fare	int
tip_amount	int
currency_code	string
cab_color	string
cab_registration_no	string
customer_rating_by_driver	int
rating_by_customer	int
passenger_count	int



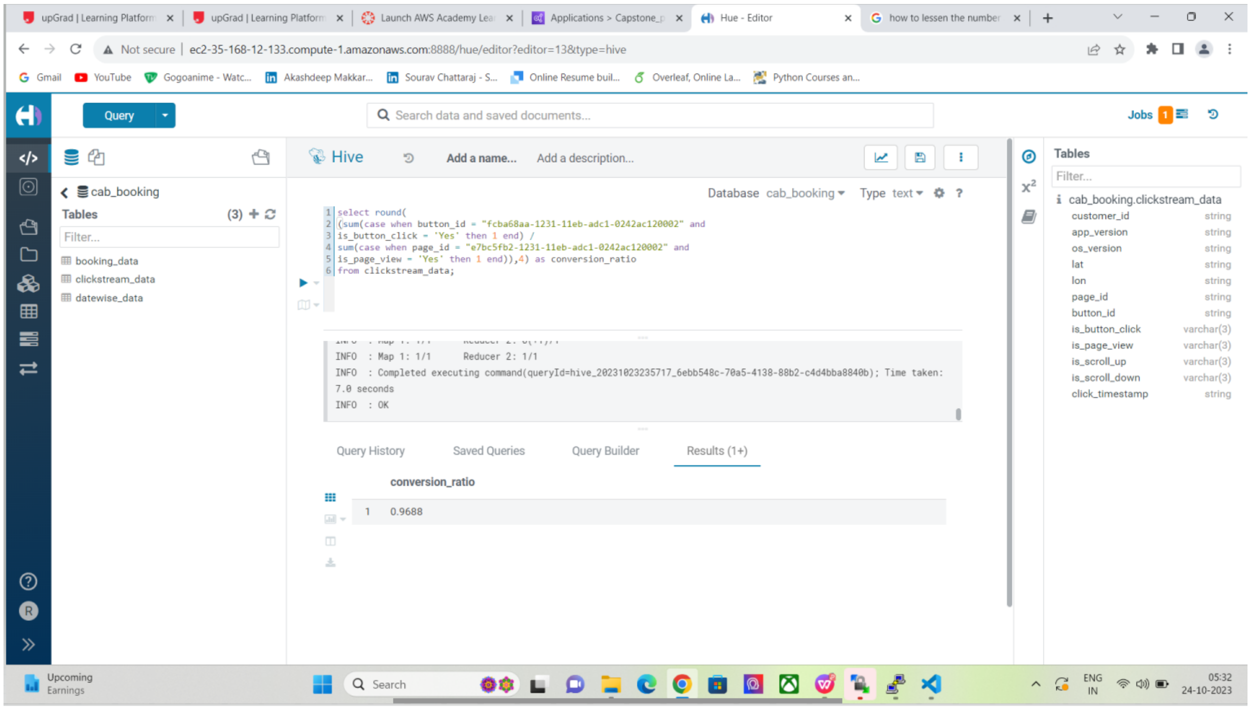
	customer_id	total_ride
1	10022393	1
2	10058402	1
3	10339567	1
4	10435129	1
5	10555335	1
6	10592274	1
7	10614890	1
8	10678994	1
9	11264797	1
10	11353346	1
11	11418437	1
12	11438890	1
13	11454977	1
14	11479815	1
15	11518953	1
16	11580321	1
17	11596512	1
18	11608791	1

3. Hive Query for Task 7:-

Find total visits by each customer on the booking page and total 'Book Now' button press. This can show the conversion-ratio.
Booking page id is 'e7bc5fb2-1231-11eb-adc1-0242ac120002'
Book Now button id is 'fcba68aa-1231-11eb-adc1-0242ac120002'.
Also, calculate the conversion ratio as described in the Tasks segment.

```
select round(
(sum(case when button_id = "fcba68aa-1231-11eb-adc1-0242ac120002" and
is_button_click = 'Yes' then 1 end) /
sum(case when page_id = "e7bc5fb2-1231-11eb-adc1-0242ac120002" and
is_page_view = 'Yes' then 1 end)),4) as conversion_ratio
from clickstream_data;
```

Screenshot after executing Query:-



```
select round(
(sum(case when button_id = "fcba68aa-1231-11eb-adc1-0242ac120002" and
is_button_click = 'Yes' then 1 end) /
sum(case when page_id = "e7bc5fb2-1231-11eb-adc1-0242ac120002" and
is_page_view = 'Yes' then 1 end)),4) as conversion_ratio
from clickstream_data;
```

```
INFO  : Map 1: 1/1      Reducer 2: 1/1
INFO  : Completed executing command(queryId=hive_20231023235717_6ebb548c-70a5-4138-88b2-c4d4bba8840b); Time taken: 7.0 seconds
INFO  : OK
```

	conversion_ratio
1	0.9688

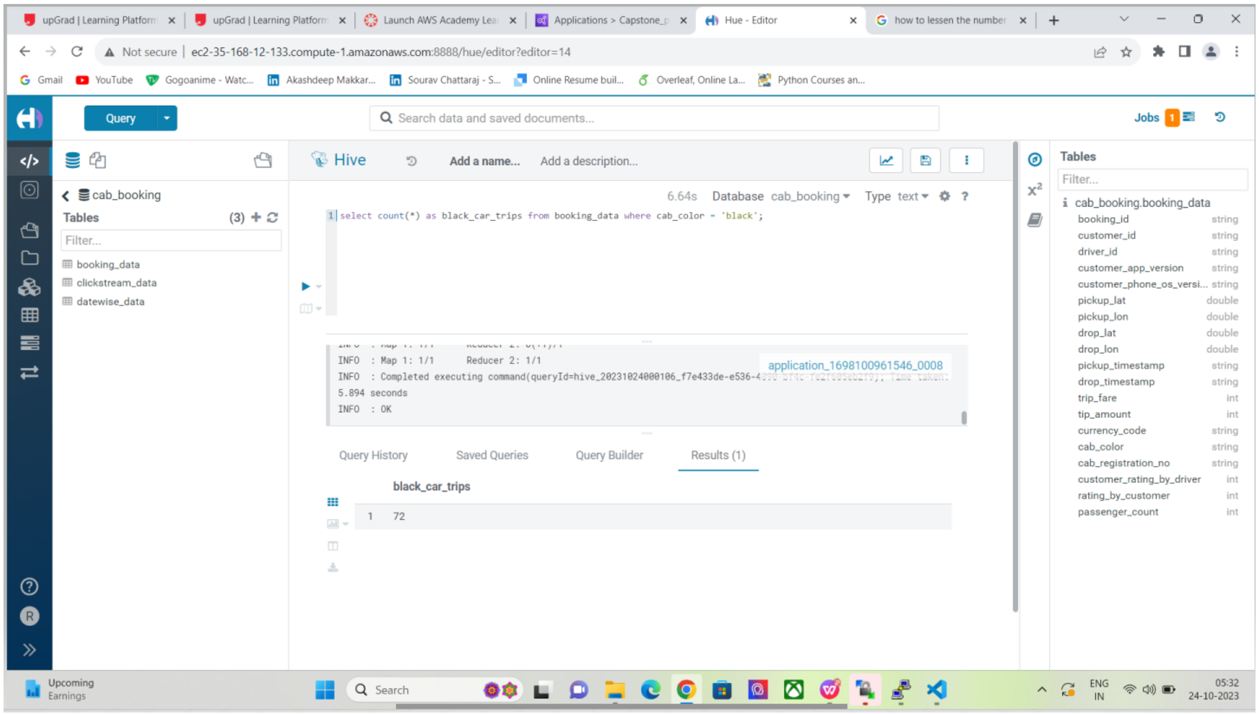
cab_booking.clickstream_data

customer_id	string
app_version	string
os_version	string
lat	string
lon	string
page_id	string
button_id	string
is_button_click	varchar(3)
is_page_view	varchar(3)
is_scroll_up	varchar(3)
is_scroll_down	varchar(3)
click_timestamp	string

4. Hive Query for Task 8:- Calculate the count of all trips done on Black cabs.

```
select count(*) as black_car_trips from booking_data where cab_color = 'black';
```

Screenshot after executing Query:-



6.64s Database cab_booking Type text

```
select count(*) as black_car_trips from booking_data where cab_color = 'black';
```

```
INFO  : Map 1: 1/1      Reducer 2: 1/1
INFO  : Completed executing command(queryId=hive_20231024000106_f7e433de-e536-4090-bf4c-fe2f605eb2f9); Time taken:
5.894 seconds
INFO  : OK
```

application_1698100961546_0008

	black_car_trips
1	72

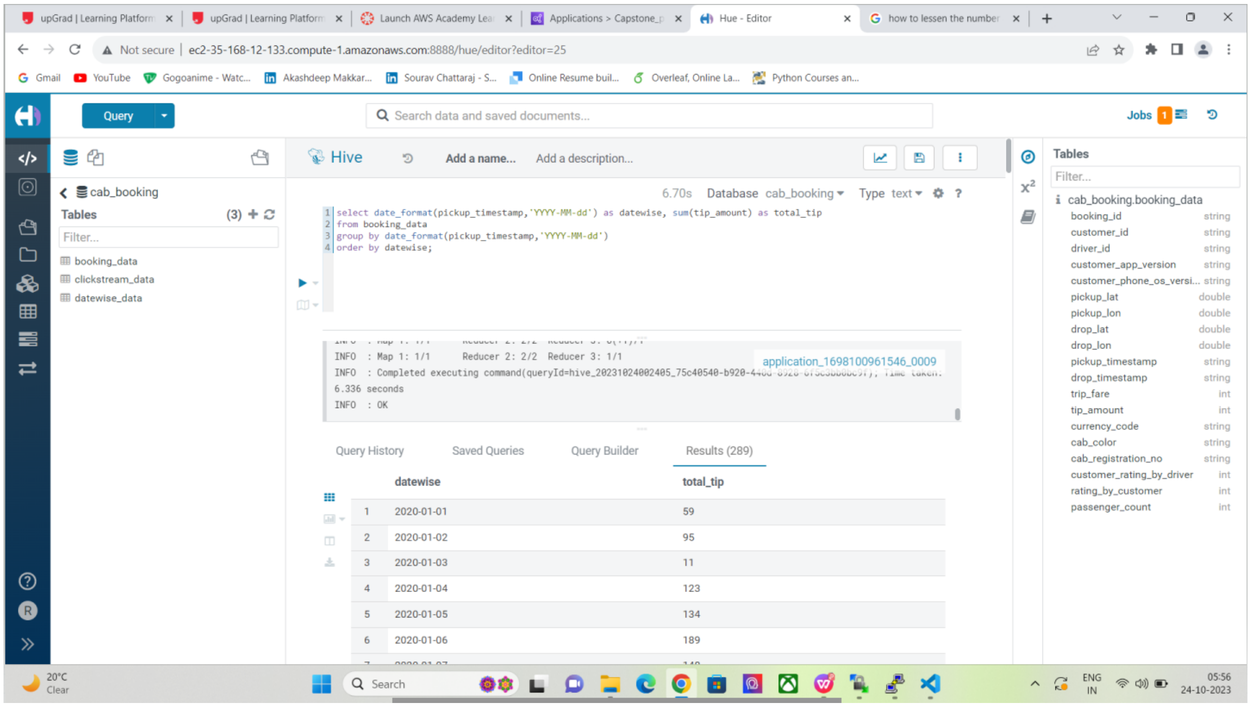
cab_booking.booking_data

booking_id	string
customer_id	string
driver_id	string
customer_app_version	string
customer_phone_os_versi...	string
pickup_lat	double
pickup_lon	double
drop_lat	double
drop_lon	double
pickup_timestamp	string
drop_timestamp	string
trip_fare	int
tip_amount	int
currency_code	string
cab_color	string
cab_registration_no	string
customer_rating_by_driver	int
rating_by_customer	int
passenger_count	int

5. **Hive Query for Task 9:- Calculate the total tip amount on a given date to all drivers by customers.**

select date_format(pickup_timestamp,'YYYY-MM-dd') as datewise, sum(tip_amount) as total_tip
from booking_data
group by date_format(pickup_timestamp,'YYYY-MM-dd')
order by datewise;

Screenshot after executing Query:-



Database cab_booking Type text

```
select date_format(pickup_timestamp,'YYYY-MM-dd') as datewise, sum(tip_amount) as total_tip
from booking_data
group by date_format(pickup_timestamp,'YYYY-MM-dd')
order by datewise;
```

```
INFO  : Map 1: 1/1      Reducer 2: 2/2  Reducer 3: 1/1
INFO  : Completed executing command(queryId=hive_20231024002405_75c40540-b920-440d-8928-6f5c3bb0bc9f); Time taken:
6.336 seconds
INFO  : OK
```

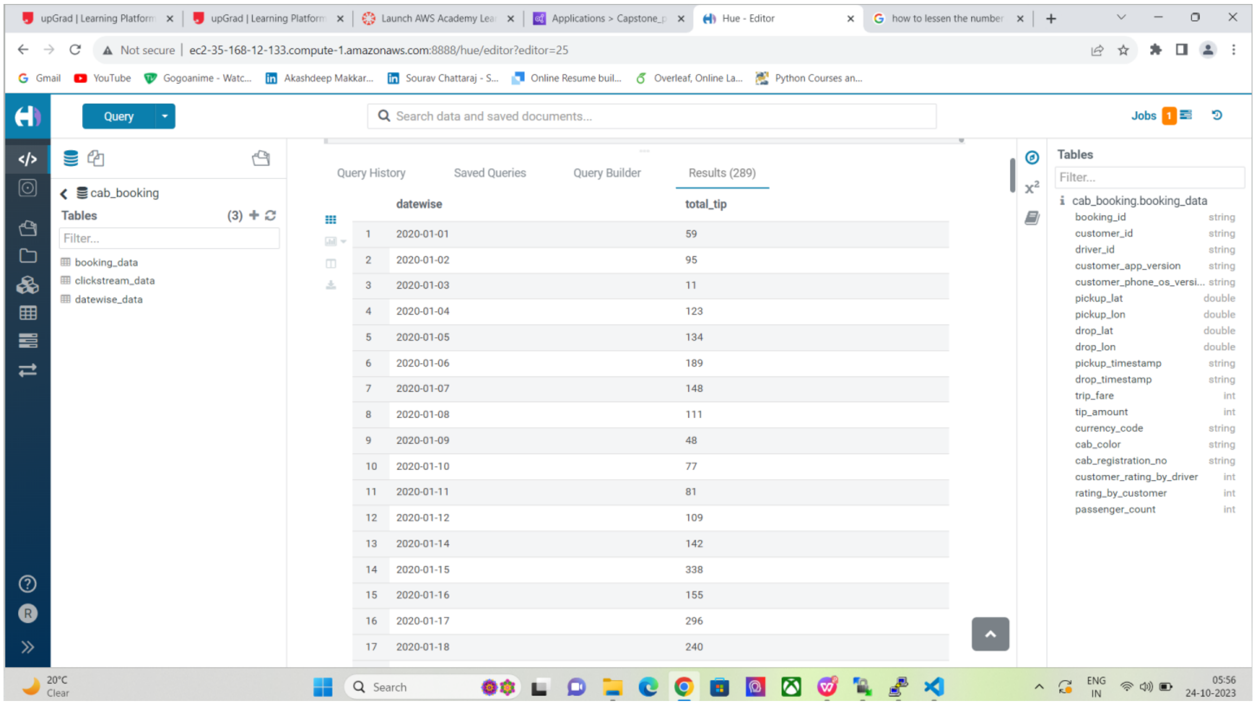
application_1698100961546_0009

Results (289)

	datewise	total_tip
1	2020-01-01	59
2	2020-01-02	95
3	2020-01-03	11
4	2020-01-04	123
5	2020-01-05	134
6	2020-01-06	189

cab_booking.booking_data

booking_id	string
customer_id	string
driver_id	string
customer_app_version	string
customer_phone_os_versi...	string
pickup_lat	double
pickup_lon	double
drop_lat	double
drop_lon	double
pickup_timestamp	string
drop_timestamp	string
trip_fare	int
tip_amount	int
currency_code	string
cab_color	string
cab_registration_no	string
customer_rating_by_driver	int
rating_by_customer	int
passenger_count	int



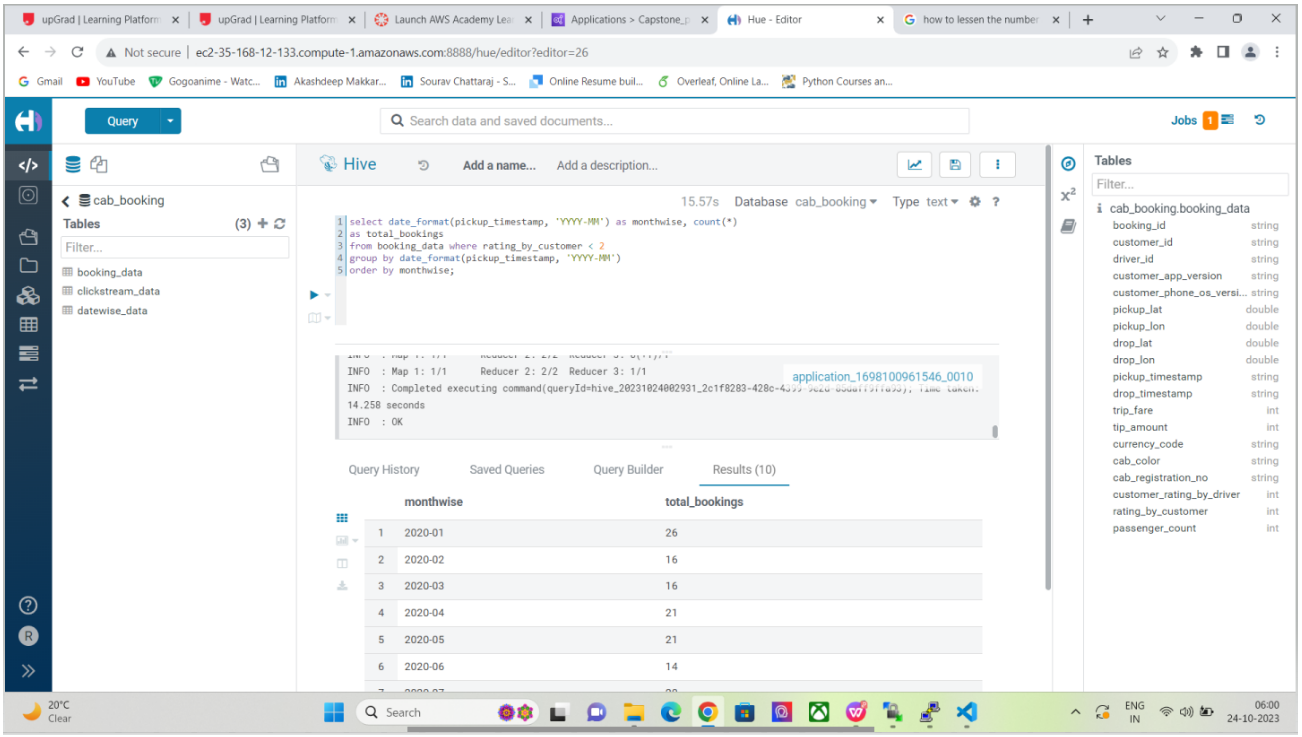
Results (289)

	datewise	total_tip
1	2020-01-01	59
2	2020-01-02	95
3	2020-01-03	11
4	2020-01-04	123
5	2020-01-05	134
6	2020-01-06	189
7	2020-01-07	148
8	2020-01-08	111
9	2020-01-09	48
10	2020-01-10	77
11	2020-01-11	81
12	2020-01-12	109
13	2020-01-14	142
14	2020-01-15	338
15	2020-01-16	155
16	2020-01-17	296
17	2020-01-18	240

6. Hive Query for Task 10:- Calculate the count of all the bookings with a rating below 2 in a particular month.

select date_format(pickup_timestamp, 'YYYY-MM') as monthwise, count(*)
as total_bookings
from booking_data where rating_by_customer < 2
group by date_format(pickup_timestamp, 'YYYY-MM')
order by monthwise;

Screenshot after executing Query:-



15.57s Database cab_booking Type text

```
select date_format(pickup_timestamp, 'YYYY-MM') as monthwise, count(*)
as total_bookings
from booking_data where rating_by_customer < 2
group by date_format(pickup_timestamp, 'YYYY-MM')
order by monthwise;
```

```
INFO  : Map 1: 1/1      Reducer 2: 2/2  Reducer 3: 1/1
INFO  : Completed executing command(queryId=hive_20231024002931_2c1f8283-428c-4399-9e2d-85daff9ffa93); Time taken:
14.258 seconds
INFO  : OK
```

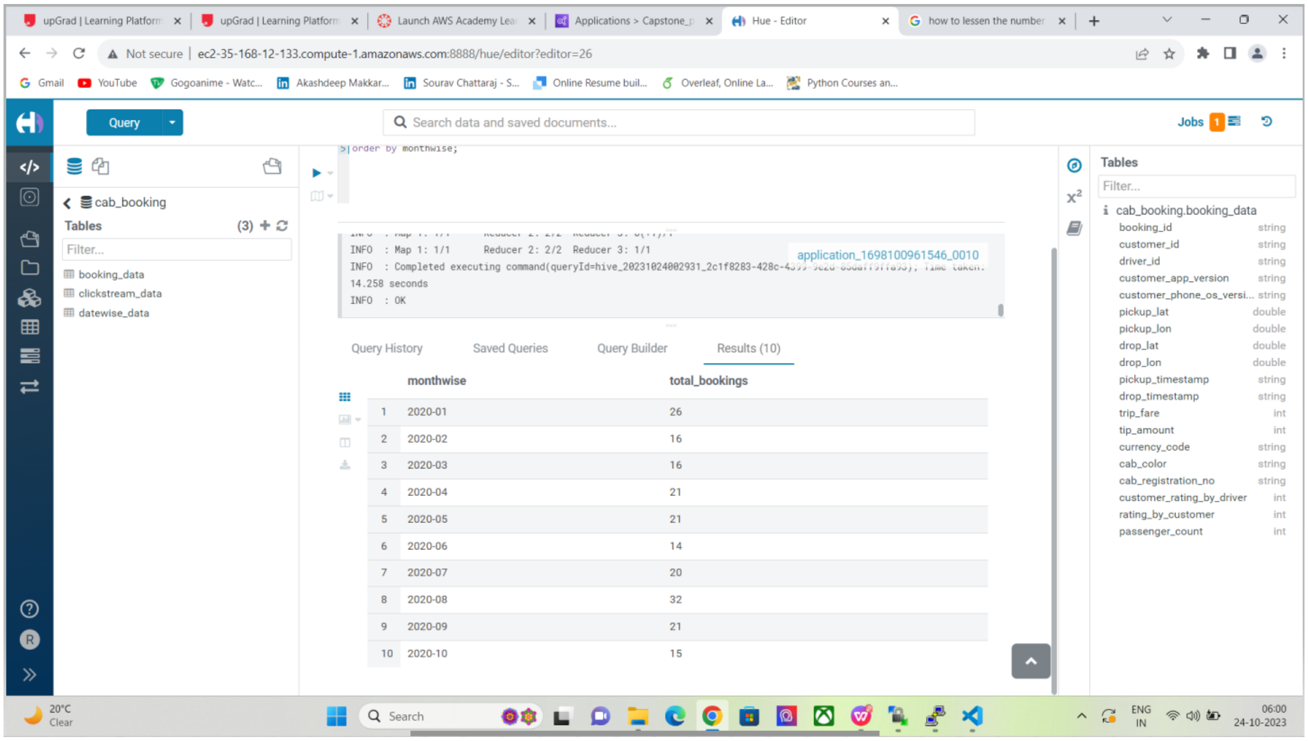
application_1698100961546_0010

Results (10)

	monthwise	total_bookings
1	2020-01	26
2	2020-02	16
3	2020-03	16
4	2020-04	21
5	2020-05	21
6	2020-06	14

cab_booking.booking_data

booking_id	string
customer_id	string
driver_id	string
customer_app_version	string
customer_phone_os_versi...	string
pickup_lat	double
pickup_lon	double
drop_lat	double
drop_lon	double
pickup_timestamp	string
drop_timestamp	string
trip_fare	int
tip_amount	int
currency_code	string
cab_color	string
cab_registration_no	string
customer_rating_by_driver	int
rating_by_customer	int
passenger_count	int



```
5 order by monthwise;
```

```
INFO  : Map 1: 1/1      Reducer 2: 2/2  Reducer 3: 1/1
INFO  : Completed executing command(queryId=hive_20231024002931_2c1f8283-428c-4399-9e2d-85daff9ffa93); Time taken: 14.258 seconds
INFO  : OK
```

	monthwise	total_bookings
1	2020-01	26
2	2020-02	16
3	2020-03	16
4	2020-04	21
5	2020-05	21
6	2020-06	14
7	2020-07	20
8	2020-08	32
9	2020-09	21
10	2020-10	15

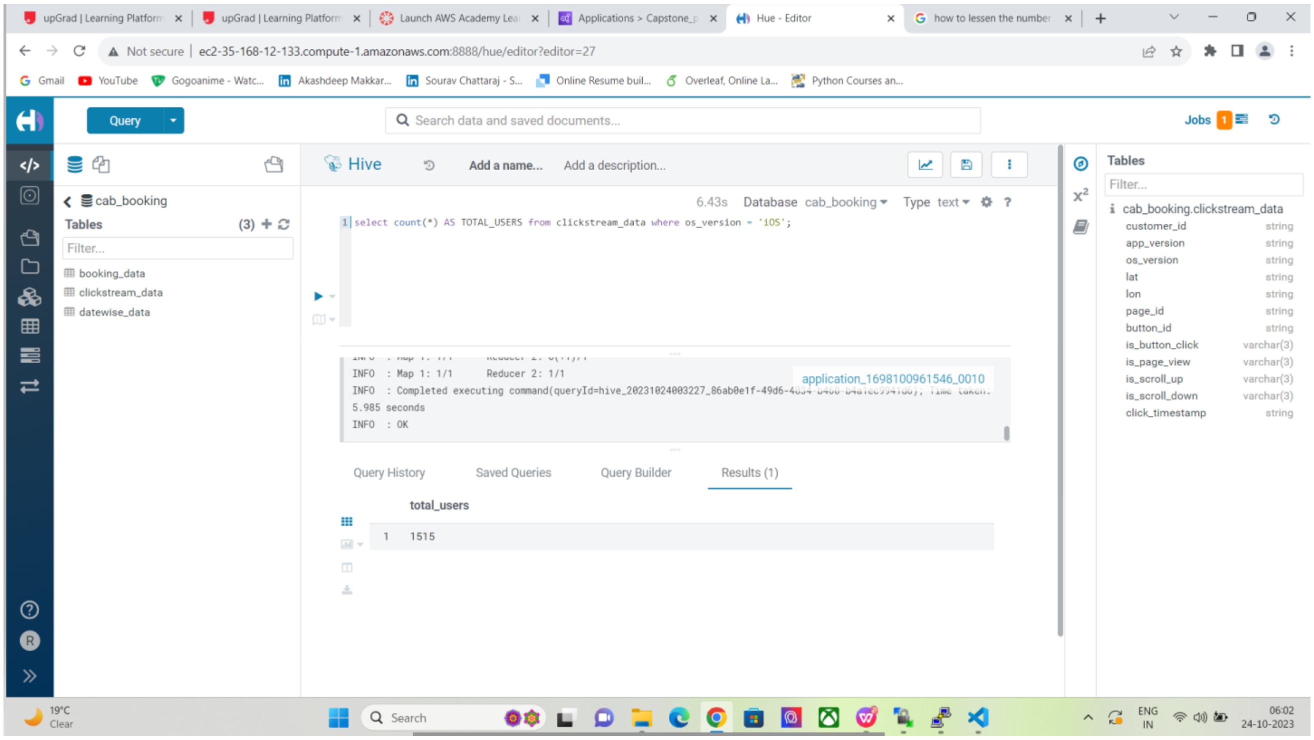
cab_booking.booking_data

booking_id	string
customer_id	string
driver_id	string
customer_app_version	string
customer_phone_os_versi...	string
pickup_lat	double
pickup_lon	double
drop_lat	double
drop_lon	double
pickup_timestamp	string
drop_timestamp	string
trip_fare	int
tip_amount	int
currency_code	string
cab_color	string
cab_registration_no	string
customer_rating_by_driver	int
rating_by_customer	int
passenger_count	int

7. Hive Query for Task 11:- Calculate the count of total iOS users.

select count(*) AS TOTAL_USERS from clickstream_data where os_version = 'iOS';

Screenshot after executing Query:-



```
select count(*) AS TOTAL_USERS from clickstream_data where os_version = 'iOS';
```

```
INFO  : Map 1: 1/1      Reducer 2: 1/1
INFO  : Completed executing command(queryId=hive_20231024003227_86ab0e1f-49d6-4034-b460-b4a1ec9941d6); Time taken:
5.985 seconds
INFO  : OK
```

	total_users
1	1515

Tables

cab_booking.clickstream_data

customer_id	string
app_version	string
os_version	string
lat	string
lon	string
page_id	string
button_id	string
is_button_click	varchar(3)
is_page_view	varchar(3)
is_scroll_up	varchar(3)
is_scroll_down	varchar(3)
click_timestamp	string