

Queries

<Hive Query for Task 5>

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

<Screenshot after executing Query>



	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

<Hive Query for Task 6>

select customer_id ,count(DISTINCT booking_id) as total_rides from booking_data group by customer_id order by customer_id asc;

<Screenshot after executing Query>

```

4 select customer_id ,count( DISTINCT booking_id) as total_rides from booking_data group by customer_id
5 order by customer_id asc;
6

```

customer_id	total_rides
1	10022393
2	10058402
3	10339567
4	10435129
5	10555335
6	10592274
7	10614890
8	10678994
9	11264797
10	11353346
11	11418437

<Hive Query for Task 7>

select count(b.button_id)/count(a.booking_id) from booking_data a full outer join
clickstream_data b on a.customer_id = b.customer_id;

<Screenshot after executing Query>

cab_bookings▼text▼⚙️?

```
1 select customer_id ,count( DISTINCT booking_id) as total_rides from booking_data group by customer_id order
2 by customer_id asc;
3
4 select customer_id ,count( DISTINCT booking_id) as total_rides from booking_data group by customer_id
5 order by customer_id asc;
6
7 select count(b.button_id)/count(a.booking_id) from booking_data a full outer join
8 clickstream_data b on a.customer_id = b.customer_id;
9
```

Query History 🔍📅Saved Queries 🔍Results (1+) 🔍📄

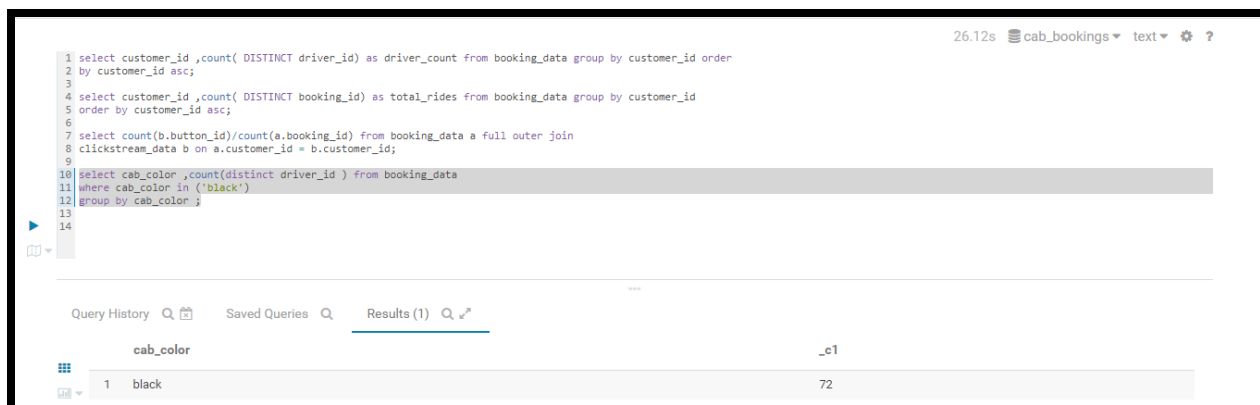
_c0

1	3.001
---	-------

<Hive Query for Task 8>

select cab_color ,count(distinct driver_id) from booking_data
where cab_color in ('black')
group by cab_color ;

<Screenshot after executing Query>



```

1 select customer_id ,count( DISTINCT driver_id) as driver_count from booking_data group by customer_id order
2 by customer_id asc;
3
4 select customer_id ,count( DISTINCT booking_id) as total_rides from booking_data group by customer_id
5 order by customer_id asc;
6
7 select count(b.button_id)/count(a.booking_id) from booking_data a full outer join
8 clickstream_data b on a.customer_id = b.customer_id;
9
10 select cab_color ,count(distinct driver_id ) from booking_data
11 where cab_color in ('black')
12 group by cab_color ;
13
14

```

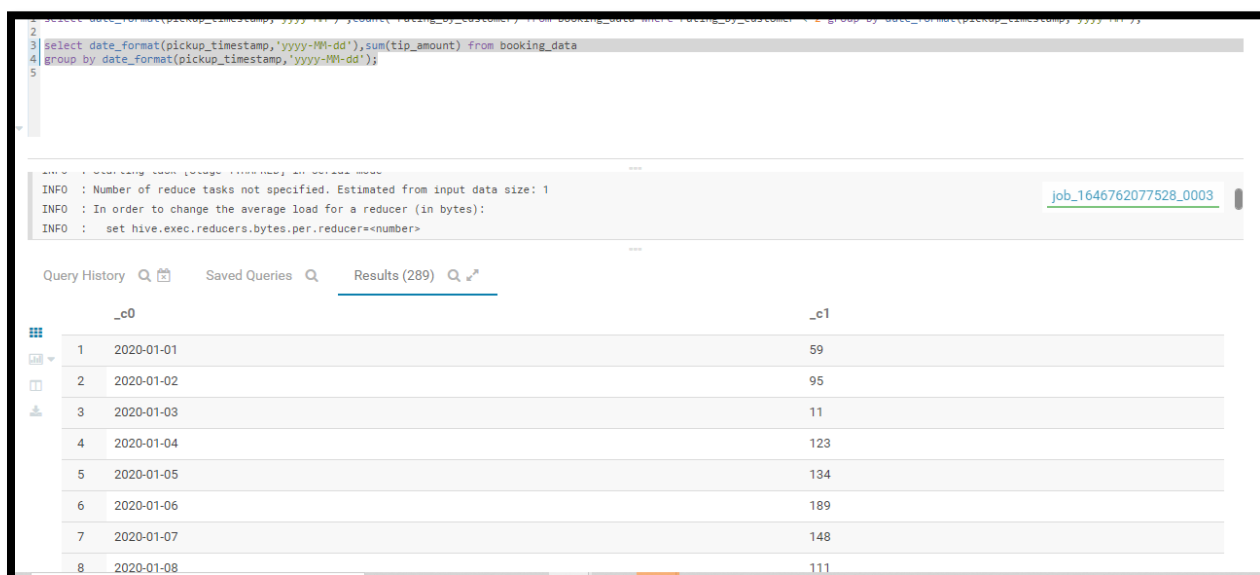
Query History | Saved Queries | Results (1)

cab_color	_c1
1 black	72

<Hive Query for Task 9>

```
select date_format(pickup_timestamp,'yyyy-MM-dd'),sum(tip_amount) from booking_data
group by date_format(pickup_timestamp,'yyyy-MM-dd');
```

<Screenshot after executing Query>



```

2
3 select date_format(pickup_timestamp,'yyyy-MM-dd'),sum(tip_amount) from booking_data
4 group by date_format(pickup_timestamp,'yyyy-MM-dd');
5

```

INFO : Number of reduce tasks not specified. Estimated from input data size: 1
 INFO : In order to change the average load for a reducer (in bytes):
 INFO : set hive.exec.reducers.bytes.per.reducer=<number>

Query History | Saved Queries | Results (289)

_c0	_c1
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

<Hive Query for Task 10>

```
select date_format(pickup_timestamp,'yyyy-MM') ,count( rating_by_customer) from
booking_data where rating_by_customer < 2 group by date_format(pickup_timestamp,'yyyy-
MM') ;
```

<Screenshot after executing Query>

```
INFO : Compiling command(queryId=hive_20220308185555_9b4acac6-c47e-4405-b226-741778211313) job_1646762077528_0002
ct date_format(pickup_timestamp,'yyyy-MM') ,count( rating_by_customer) from booking_data where ra
ting_by_customer < 2 group by date_format(pickup_timestamp,'yyyy-MM')
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:=c0, type:string, comment:nu
```

Query History 🔍 📅

Saved Queries

Results (10) 🔍 ↗

	_c0	_c1	
<div><div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div></div><div><div></div><div></div><div></div><div></div></div></div>	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

<Hive Query for Task 11>

```
select os_version ,count(distinct customer_id) from clickstream_data
where os_version in ('iOS')
group by os_version;
```

<Screenshot after executing Query>

27.12s

cab_bookings

text

```

1 select customer_id ,count( DISTINCT driver_id) as driver_count from booking_data group by customer_id order
2 by customer_id asc;
3
4 select customer_id ,count( DISTINCT booking_id) as total_rides from booking_data group by customer_id
5 order by customer_id asc;
6
7 select count(b.button_id)/count(a.booking_id) from booking_data a full outer join
8 clickstream_data b on a.customer_id = b.customer_id;
9
10 select cab_color ,count(distinct driver_id ) from booking_data
11 where cab_color in ('black')
12 group by cab_color ;
13
14 select os_version ,count(distinct customer_id) from clickstream_data
15 where os_version in ('ios')
16 group by os_version;
17
18

```

Query History

Saved Queries

Results (1)

os_version	_c1
1 iOS	1515