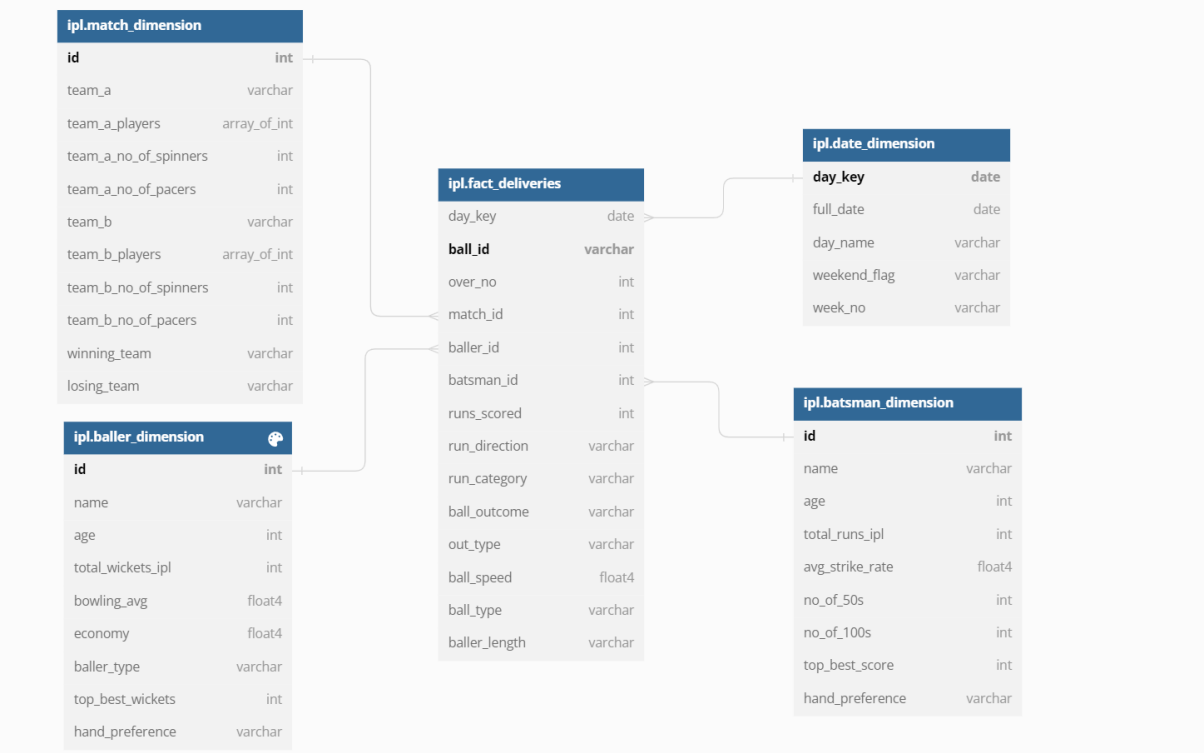
Q1. IPL tournament

Ans:

DWH diagram (Star Schema)



List of Dimensions and their definition:

All the dimensions shall be maintained in SCD-Type1 fashion

Table ipl.batsman_dimension {

id int [pk]

name varchar

age int

total_runs_ipl int

avg_strike_rate float4

no_of_50s int

no_of_100s int

top_best_score int

```
  hand_preference varchar
}

Table ipl.baller_dimension {
  id int [pk]
  name varchar
  age int
  total_wickets_ipl int
  bowling_avg float4
  economy float4
  baller_type varchar
  top_best_wickets int
  hand_preference varchar
}


Table ipl.match_dimension {
  id int [pk]
  team_a varchar
  team_a_players array_of_int
  team_a_no_of_spinners int
  team_a_no_of_pacers int
  team_b varchar
  team_b_players array_of_int
  team_b_no_of_spinners int
  team_b_no_of_pacers int
  winning_team varchar
  losing_team varchar
}
```

```
Table ipl.date_dimension {
  day_key date [pk]
  full_date date
  day_name varchar
  weekend_flag varchar
  week_no varchar
}
```

Facts and their definition:

```
Table ipl.fact_deliveries {
  day_key date [ref: > ipl.date_dimension.day_key]
  ball_id varchar [pk]
  over_no int
  match_id int [ref: > ipl.match_dimension.id]
  baller_id int [ref: > ipl.baller_dimension.id]
  batsman_id int [ref: > ipl.batsman_dimension.id]
  runs_scored int
  run_direction varchar
  run_category varchar
  ball_outcome varchar
  out_type varchar
  ball_speed float4
  ball_type varchar
  baller_length varchar



}
```

Insightful reports:

1. **How much runs were scored in each direction (legside, offside, longoff, longon, thirdman) in a particular match?**

Ans:

```
Select sum(runs_scored), run_direction
 from ipl.fact_deliveries d
Inner join match_dimension m
on d.match_id= m.id
and m.team_a='Rajasthan Royals' and m.team_b='Mumbai Indians'
group by run_direction
```

2. **What type of runs (1's, 2's, 3's etc) were scored on each baller length (yorker, length, good_length, short_length) MATCHWISE?**

```
Ans: Select count(case when d.runs_scored=1 then '1's'
                When d.runs_scored=2 then '2's'
                When d.runs_scored=3 then '3's'
                When d.runs_scored=4 then '4's'
                When d.runs_scored=6 then '6's'
                End) as run_category_count,
case when d.runs_scored=1 then '1's'
                When d.runs_scored=2 then '2's'
                When d.runs_scored=3 then '3's'
                When d.runs_scored=4 then '4's'
                When d.runs_scored=6 then '6's'
                End as run_category
, d.baller_length,
m.team_a, m.team_b
 from ipl.fact_deliveries d
left join match_dimension m
```

on d.match_id= m.id

group by case when d.runs_scored=1 then ‘1’s’

When d.runs_scored=2 then ‘2’s’

When d.runs_scored=3 then ‘3’s’

When d.runs_scored=4 then ‘4’s’

When d.runs_scored=6 then ‘6’s’

End,

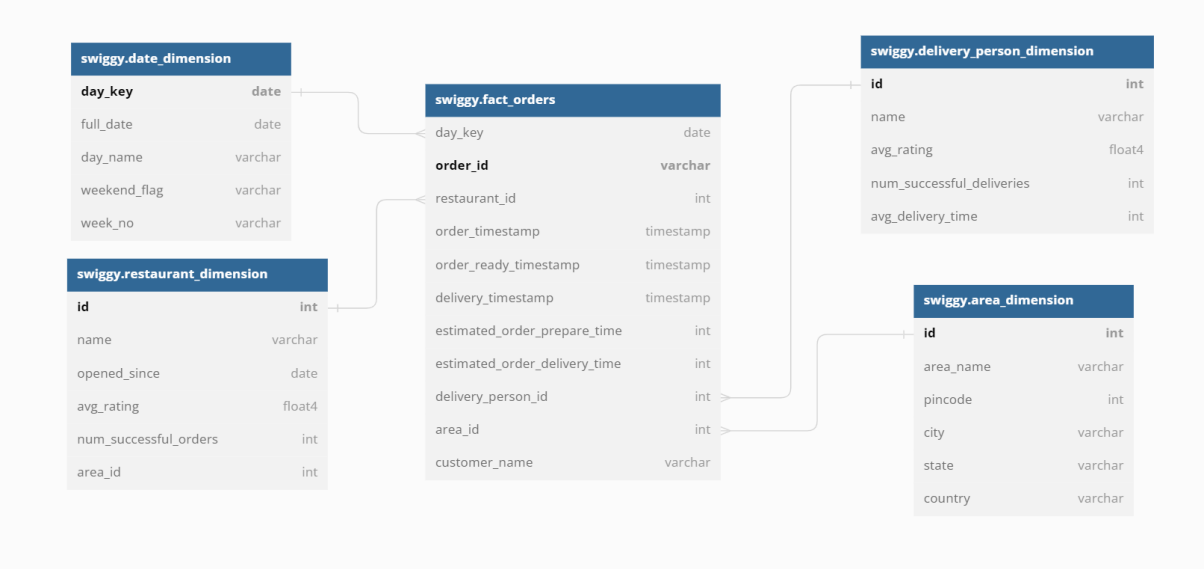
, d.baller_length,

m.team_a, m.team_b

Q2. Swiggy DWH

Ans:

DWH diagram (Star Schema)



List of Dimensions and their definition:

All the dimensions shall be maintained in SCD-Type1 fashion

```
Table swiggy.restaurant_dimension {
  id int [pk]
  name varchar
  opened_since date
  avg_rating float4
  num_successful_orders int
  area_id int

}

Table swiggy.area_dimension {
  id int [pk]
  area_name varchar
  pincode int
  city varchar
  state varchar
  country varchar
}


Table swiggy.delivery_person_dimension {
  id int [pk]
  name varchar
  avg_rating float4
  num_successful_deliveries int
  avg_delivery_time int
}

Table swiggy.date_dimension {
```

day_key date [pk]

full_date date

day_name varchar

weekend_flag varchar

week_no varchar

}

Facts and their definition:

Table swiggy.fact_orders {

day_key date [ref: > swiggy.date_dimension.day_key]

order_id varchar [pk]

restaurant_id int [ref: > swiggy.restaurant_dimension.id]

order_timestamp timestamp

order_ready_timestamp timestamp

delivery_timestamp timestamp

estimated_order_prepare_time int

estimated_order_delivery_time int

delivery_person_id int [ref: > swiggy.delivery_person_dimension.id]

area_id int [ref: > swiggy.area_dimension.id]

customer_name varchar

}

Insightful reports:

1. **What is the area wise percentage of orders delivered on time?**

Ans:

with cte as

(

select a.area_name, a.pincode, count(*) ontime_deliveries from swiggy.area_dim a

inner join

```
swiggy.fact_orders f

on a.id= f.area_id

where TIMESTAMPADD(MINUTE,estimated_order_delivery_time,order_timestamp) <=
order_timestamp

group by a.area_name, a.pincode


),

cte1 as

(

select a.area_name, a.pincode, count(*) total_deliveries from swiggy.area_dim a

inner join

swiggy.fact_orders f

on a.id= f.area_id

group by a.area_name, a.pincode

)

select cte.area_name,cte.pincode, (ontime_deliveries*100/total_deliveries)
area_wise_percentage

from cte

inner join cte1

on cte.area_name=cte1.area_name and cte.pincode=cte1.pincode;
```

2. **What is the area from which most orders are received?**

Ans:

```
with cte as

(

select a.area_name, a.pincode, count(*) total_deliveries from swiggy.area_dim a

inner join

swiggy.fact_orders f

on a.id= f.area_id

group by a.area_name, a.pincode
```

)

Select * from cte

Order by total_deliveries desc limit 1;

3. **Who's fault was there in a particular area for delayed deliveries (restaurant or delivery person)?**

Ans:

select a.area_name, a.pincode,

case

when TIMESTAMPADD(MINUTE, estimated_order_prepare_time,order_timestamp) > order_ready_timestamp then 'Restaurant's fault'

when TIMESTAMPADD(MINUTE, estimated_order_delivery_time,order_timestamp) > delivery_timestamp then 'Delivery person's fault'

end Delay_Reason

from swiggy.fact_orders f

inner join

swiggy.area_dim a

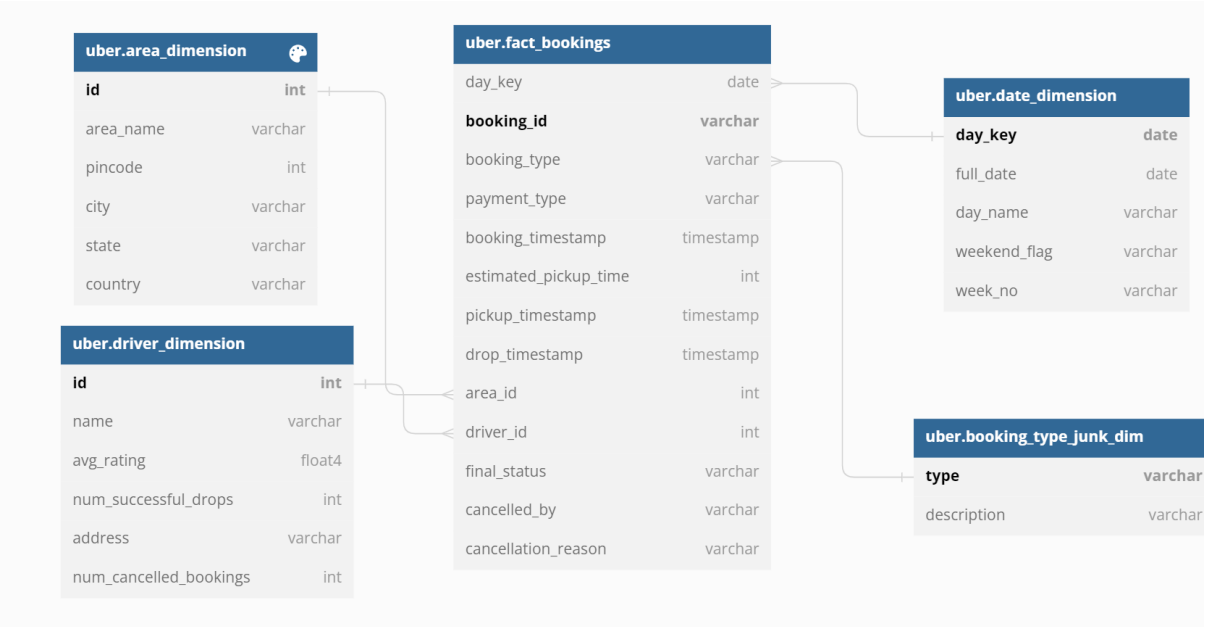
on a.id= f.area_id

where a.area_name='Kondapur' and a.pincode=500084

Q3.

Ans:

DWH diagram (Star Schema)



List of Dimensions and their definition:

All the dimensions shall be maintained in SCD-Type1 fashion

Table uber.area_dimension {

id int [pk]

area_name varchar

pincode int

city varchar

state varchar

country varchar

}

Table uber.driver_dimension {

id int [pk]

```
  name varchar
  avg_rating float4
  num_successful_drops int
  address varchar
  num_cancelled_bookings int
}

Table uber.date_dimension {
  day_key date [pk]
  full_date date
  day_name varchar
  weekend_flag varchar
  week_no varchar
}
```

Facts and their definition:

```
Table uber.fact_bookings {
day_key date  [ref: > uber.date_dimension.day_key]
booking_id varchar [pk]
booking_type varchar
payment_type varchar
booking_timestamp timestamp
estimated_pickup_time int
pickup_timestamp timestamp
drop_timestamp timestamp
area_id int [ref: > uber.area_dimension.id]
driver_id int [ref: > uber.driver_dimension.id]
final_status varchar
cancelled_by varchar
```

cancellation_reason varchar

}

Insightful reports:

1. **What is the area wise count of ride bookings on a particular day?**

Ans:

```
select count(*) num_of_rides,f.booking_type , a.area_name, a.city, a.state from
uber.fact_bookings f left join uber.area_dimension a
on f.area_id=a.id
where f.day_key='2022-01-05'
group by f.booking_type , a.area_name, a.city, a.state;
```

2. **Percentage of rides which have abnormally long pickup time daywise?**

Ans:

```
with cte as
(
select f.day_key, count(*) delayed_pickup_num from uber.fact_orders f
inner join
uber.area_dim a
on a.id= f.area_id
where TIMESTAMPADD(MINUTE,estimated_pickup_time,booking_timestamp) >
pickup_timestamp
group by f.day_key

),
cte1 as
(
select f.day_key, count(*) total_drops from uber.fact_orders f
inner join
```

```
uber.area_dim a
on a.id= f.area_id
where f.final_status='Dropped successfully'
group by f.day_key
)
select cte.day_key, (cte.delayed_pickup_num*100/cte1.total_drops) day_wise_percentage
from cte
inner join cte1
on cte.day_key=cte1.day_key;
```

3. **Daywise, Areawise, peak time of the day (in terms of hour of the day)?**

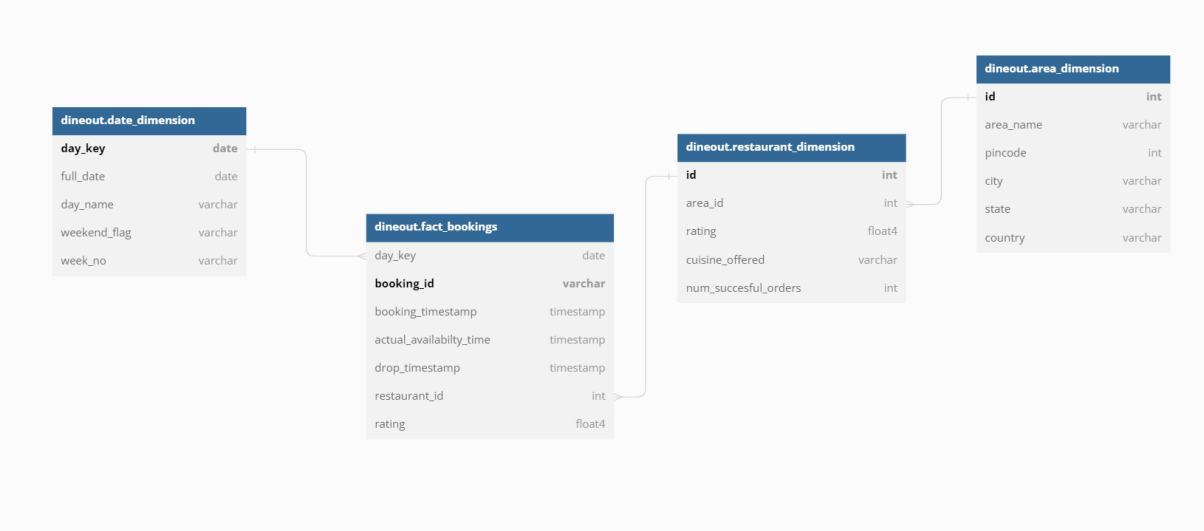
Ans:

```
with cte as
(
select f.day_key, hour(f.booking_timestamp) hour_of_booking , a.area_name, a.city,
a.state,
count(*) num_of_rides,
from
uber.fact_bookings f
inner join uber.area_dimension a
on f.area_id=a.id
group by f.day_key , hour(f.booking_timestamp), a.area_name, a.city, a.state
order by count(*) desc limit 1
)
select * from cte
```

Q4. Dineout Datawarehouse

Ans:

DWH diagram (Snowflake Schema)



List of Dimensions and their definition:

All the dimensions shall be maintained in SCD-Type1 fashion

Table dineout.restaurant_dimension {

id int [pk]

area_id int [ref: > dineout.area_dimension.id]

rating float4

cuisine_offered varchar

num_succesful_orders int

}

Table dineout.area_dimension {

```
  id int [pk]
  area_name varchar
  pincode int
  city varchar
  state varchar
  country varchar
}



Table dineout.date_dimension {
  day_key date [pk]
  full_date date
  day_name varchar
  weekend_flag varchar
  week_no varchar
}
```

Facts and their definition:

```
Table dineout.fact_bookings {
day_key date  [ref: > dineout.date_dimension.day_key]
booking_id varchar [pk]
booking_timestamp timestamp
est_availability_time timestamp
actual_availabilty_time timestamp
drop_timestamp timestamp
restaurant_id int [ref: > dineout.restaurant_dimension.id]
rating float4
final_status varchar
```

}

Insightful reports:

1. **What is the area wise percentage of tables provided to the customer on time?**

Ans:

```
with cte as
(
select a.area_name, a.city,a.state, count(*) ontime_bookings from dineout.fact_bookings f
inner join
dineout.area_dim a
on a.id= f.area_id
where actual_availabilty_time <= est_availabilty_time
group by a.area_name, a.city,a.state
),
cte1 as
(
select a.area_name, a.city,a.state, count(*) total_bookings from dineout.fact_bookings f
inner join
dineout.area_dim a
on a.id= f.area_id
where f.final_status='success'
group by a.area_name. a.city,a.state
)
select a.area_name, a.city,a.state,  (cte.ontime_bookings *100/cte1.total_bookings)
area_wise_percentage
from cte
inner join cte1
on cte.area_name =cte1. area_name
```

```
and cte.city =cte1.city
and cte.state =cte1.state
```

2. **How many bookings (percentage) were cancelled because of delayed table availabilty, area wise?**

Ans:

```
with cte as
(
select a.area_name, a.city,a.state, count(*) late_bookings from dineout.fact_bookings f
inner join
dineout.area_dim a
on a.id= f.area_id
where actual_availabilty_time > est_availabilty_time
and final_status='cancelled'
group by a.area_name, a.city,a.state
),
cte1 as
(
select a.area_name, a.city,a.state, count(*) total_bookings from dineout.fact_bookings f
inner join
dineout.area_dim a
on a.id= f.area_id
group by a.area_name. a.city,a.state
)
select a.area_name, a.city,a.state,  (cte.late_bookings  *100/cte1.total_bookings)
area_wise_percentage
from cte
inner join cte1
on cte.area_name =cte1. area_name
and cte.city =cte1.city
```

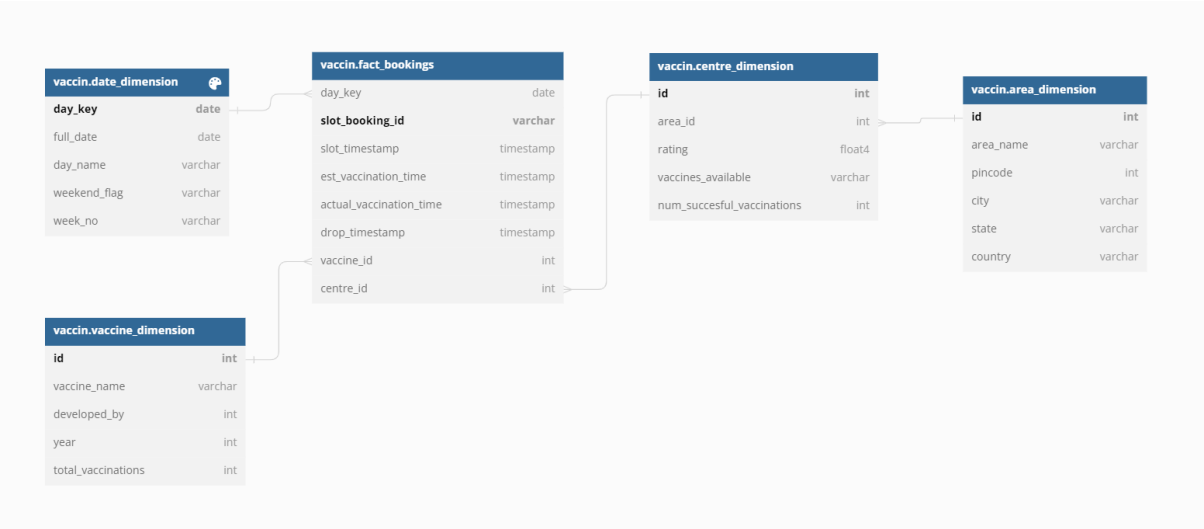
and cte.state =cte1.state

;

Q4. Covid Vaccination Datawarehouse

Ans:

DWH diagram (Snowflake Schema)



List of Dimensions and their definition:

All the dimensions shall be maintained in SCD-Type1 fashion

Table vaccin.centre_dimension {

id int [pk]

area_id int [ref: > vaccin.area_dimension.id]

rating float4

vaccines_available varchar

num_succesful_vaccinations int

}

```
Table vaccin.area_dimension {
  id int [pk]
  area_name varchar
  pincode int
  city varchar
  state varchar
  country varchar
}

Table vaccin.date_dimension {
  day_key date [pk]
  full_date date
  day_name varchar
  weekend_flag varchar
  week_no varchar
}

Table vaccin.date_dimension {
  day_key date [pk]
  full_date date
  day_name varchar
  weekend_flag varchar
  week_no varchar
}
```

Facts and their definition:

```
Table vaccin.fact_vaccinations {
```

day_key date [ref: > vaccin.date_dimension.day_key]

slot_booking_id varchar [pk]

slot_timestamp timestamp

est_vaccination_time timestamp

actual_vaccination_time timestamp

drop_timestamp timestamp

vaccine_id int [ref: > vaccin.vaccine_dimension.id]

centre_id int [ref: > vaccin.centre_dimension.id]

}

Insightful reports:

1. What is the area wise percentage of vaccinations done on time?

Ans:

```
with cte as
(
select a.area_name, a.city,a.state, count(*) ontime_vaccinations from
vaccin.fact_vaccinations f
inner join
vaccin.centre_dimension c
on c.id= f.centre_id
inner join
vaccin.area_dimension a
on a.id= c.area_id
where actual_vaccination_time <= est_vaccination_time
group by a.area_name, a.city,a.state
),
cte1 as
(
select a.area_name, a.city,a.state, count(*) total_vaccinations from vaccin.fact_vaccinations
f
```

```
inner join

vaccin.centre_dimension c

on c.id= f.centre_id

inner join

vaccin.area_dimension a

on a.id= c.area_id

group by a.area_name. a.city,a.state

)

select a.area_name, a.city,a.state,  (cte.ontime_vaccinations *100/cte1.total_vaccinations)
area_wise_percentage

from cte

inner join cte1

on cte.area_name =cte1. area_name

and cte.city =cte1.city

and cte.state =cte1.state
```

2. What was the peak time of slot bookings, area wise , day wise?

Ans:

```
with cte as

(

select a.area_name, a.city,a.state,f.day_key,f.day_name,hour(f.slot_timestamp) hour,

count(*) over (partition by a.area_name,
a.city,a.state,f.day_key,f.day_name,hour(f.slot_timestamp)) total from
vaccin.fact_vaccinations f

inner join vaccin.date_dimension d

d.day_key=f.day_key

inner join

vaccin.centre_dimension c

on c.id= f.centre_id

inner join
```

```
vaccin.area_dimension a

on a.id= c.area_id

),

cte1 as

(

select

cte.area_name, cte.city,cte.state,cte.day_key,cte.day_name,cte.hour hour,

dense_rank() over(partition by cte.area_name, cte.city,cte.state,cte.day_key,cte.day_name order by cte.total desc) rnk

from cte

)

select *

from cte1

where rnk=1
```