SQL Project Queries - Vancouver

a. Analyze different metrics to draw the distinction between Super Host and Other Hosts:

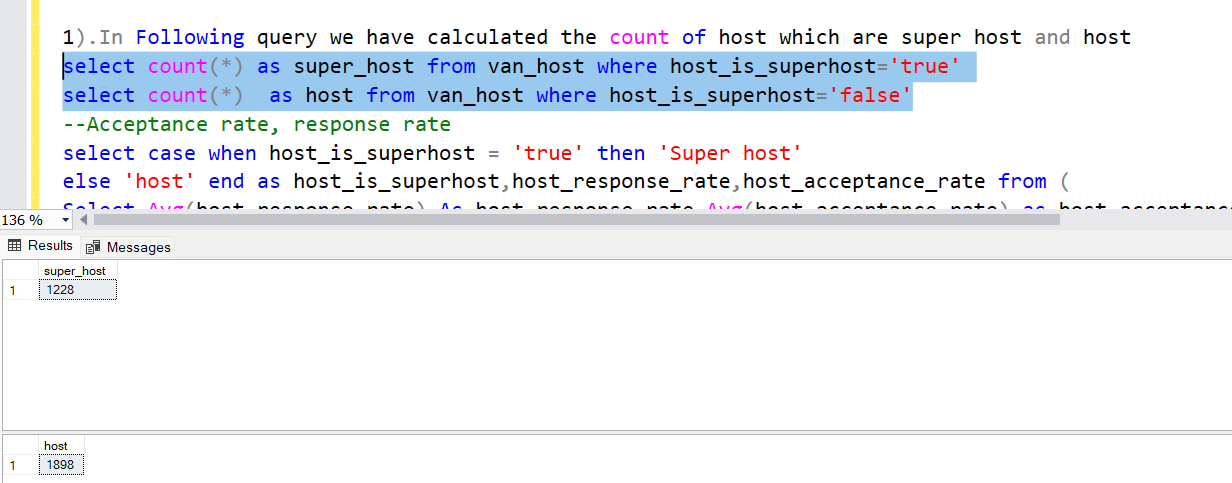
To achieve this, you can use the following metrics and explore a few yourself as well.

Acceptance rate, response rate, instant booking, profile picture, identity verified,

review review scores, average no of bookings per month, etc.

**1).In Following query we have calculated the count of host which are super host and host**

select count(\*) from van\_host where host\_is\_superhost='true'

select count(\*) from van\_host where host\_is\_superhost='false'  


--Acceptance rate, response rate  
**1).In Following query we have found the variation in response rate and acceptance rate   
2). For Analysis we found Average of response rate and acceptance rate from host table and then grouped them.**

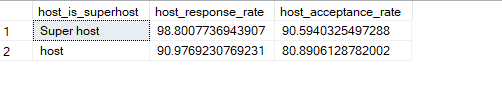
select case when host\_is\_superhost = 'true' then 'Super host'

else 'host' end as host\_is\_superhost,host\_response\_rate,host\_acceptance\_rate from (

Select Avg(host\_response\_rate) As host\_response\_rate,Avg(host\_acceptance\_rate) as host\_acceptance\_rate,

host\_is\_superhost from van\_host

where host\_response\_rate is not null or host\_acceptance\_rate is not null group by host\_is\_superhost) as mo



--profile picture  
1).for Analysis we found total number of super host and host who have profile picture .  
2) So from following output we can clearly see that 1225 out of 1228 super host have profile pic .

select case when host\_is\_superhost = 'true' then 'Super host'

else 'host' end as host\_is\_superhost,ct,cnt2 from (

select host\_is\_superhost,sum(true) as ct,sum(false)as cnt2 from(

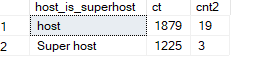
select host\_is\_superhost,[TRUE],[FALSE] from van\_host

pivot (count(host\_id) for

host\_has\_profile\_pic in ([TRUE],[FALSE]))a)a

where host\_is\_superhost is not null

group by host\_is\_superhost

) as mo  


--identity verified  
1).for Analysis we found total number of super host and host whos identity has verified .  
2) So from following output we can clearly see that 1184 out of 1228 super host have identity verified .

select case when host\_is\_superhost = 'true' then 'Super host'

else 'host' end as host\_is\_superhost,verified,not\_verified

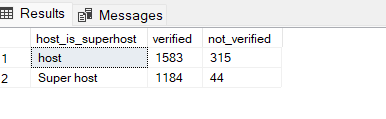
from (

select host\_is\_superhost,sum(true) as Verified,sum(false)as Not\_verified from(

select host\_is\_superhost,[TRUE],[FALSE] from van\_host

pivot (count(host\_id) for

host\_Identity\_verified in ([TRUE],[FALSE]))a)a where host\_is\_superhost is not null

group by host\_is\_superhost) as mo  
  


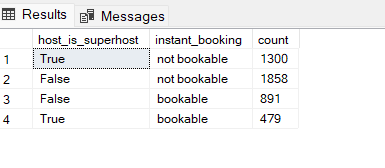
--instant booking  
1).for Analysis we found how many listing of super host and host are having option of instant booking   
2) Also in output if Value is true then then that hotel can be booked instantly.

select host\_is\_superhost, case when instant\_bookable='false' then 'not bookable'

else 'bookable' end as instant\_booking , count from (

Select B.Host\_is\_superhost,A.instant\_bookable,Count(A.HOST\_ID) AS count From van\_list A Inner Join van\_host B

ON A.HOST\_ID = B.HOST\_ID

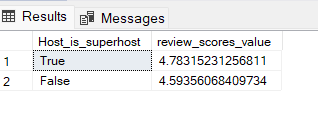
where b.host\_is\_superhost is not null Group by Host\_is\_superhost,instant\_bookable ) as mo   
  


--review review scores

1) we found average of review score for host and super host.  
2) Where true is for super host and false for host.  
3)we can see from output that average score for super host are slightly greater than that for host.

Select B.Host\_is\_superhost,avg(A.review\_scores\_value) as review\_scores\_value

From van\_list A Inner Join van\_host B

ON A.HOST\_ID = B.HOST\_ID Group by Host\_is\_superhost  
  


---Additional Analysis  
1st analysis

1). In following query we analyse the different ratings for listing under host and super host  
2). We found average rating for checkings,cleanliness,communication,location,score rating,  
3).Note that true is for super host and false for host

select b.host\_is\_superhost,avg(a.review\_scores\_checkin) as review\_scores\_checkin,

avg(a.review\_scores\_cleanliness) as review\_scores\_cleanliness,avg(a.review\_scores\_communication) as review\_scores\_communication

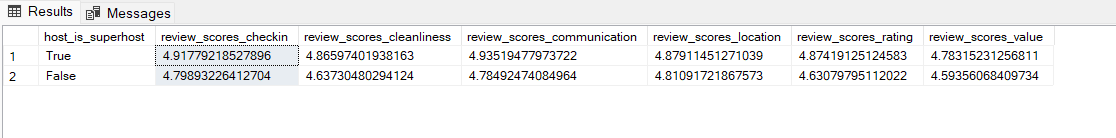
,avg(a.review\_scores\_location) as review\_scores\_location

,avg(a.review\_scores\_rating) as review\_scores\_rating ,avg(a.review\_scores\_value) as review\_scores\_value

from van\_list as a inner join van\_host as b on a.host\_id=b.host\_id

where b.host\_is\_superhost is not null

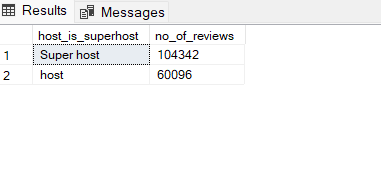
group by b.host\_is\_superhost;

  
  
**2nd analysis**  
1) here we found the total number of reviews for super host and host  
2) we can note from here that though super host count is less but they got higher reviews than host

select c.host\_is\_superhost,count(\*) from van\_review as a left join van\_list as b on a.listing\_id=b.id

left join van\_host as c on b.host\_id=c.host\_id where c.host\_is\_superhost is not null

group by c.host\_is\_superhost;

  
3rd Analysis   
1) in this analysis we found how average of bedroomas and beds vary between super host and host.  
2)from output we can see that there is no significant difference between the

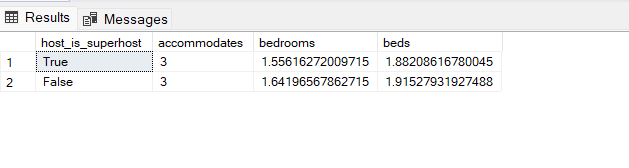
select b.host\_is\_superhost,avg(a.accommodates) as accommodates,avg(a.bedrooms) as bedrooms ,

avg(a.beds) as beds

from van\_list as a inner join van\_host as b on a.host\_id=b.host\_id

where b.host\_is\_superhost is not null

group by b.host\_is\_superhost;



---b Using the above analysis, identify top 3 crucial metrics one needs to maintain to become a Super Host and also, find their average values

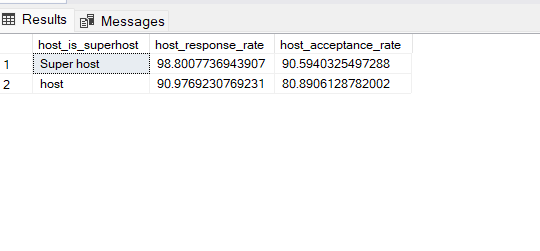
--Acceptance rate, response rate  
1). Using following parameter we can clearly see that if host improves these parameter he can become in super host

select case when host\_is\_superhost = 'true' then 'Super host'

else 'host' end as host\_is\_superhost,host\_response\_rate,host\_acceptance\_rate from (

Select Avg(host\_response\_rate) As host\_response\_rate,Avg(host\_acceptance\_rate) as host\_acceptance\_rate,

host\_is\_superhost from van\_host

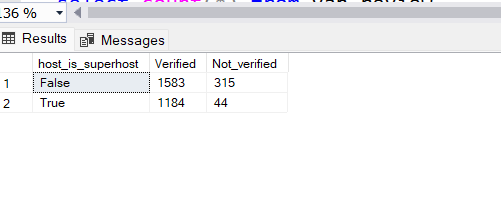
where host\_response\_rate is not null or host\_acceptance\_rate is not null group by host\_is\_superhost) as mo  


--identity verified

**select host\_is\_superhost,sum(true) as Verified,sum(false)as Not\_verified from(**

**select host\_is\_superhost,[TRUE],[FALSE] from van\_host**

**pivot (count(host\_id) for**

**host\_Identity\_verified in ([TRUE],[FALSE]))a)a group by host\_is\_superhost**  


--c Analyze how does the comments of reviewers vary for listings of Super Hosts vs Other Hosts(Extract words from the comments provided by the reviewers)  
1) For Analysis we took few Positive comments such as beautiful,tastefull,perfect,amazing,very clean,recommended, very welcoming, guest host.

2) For Analysis we took few Negative comments such as a.comments like slow wifi,canceled, not recommended,unprofessional,not good,rude

select count(\*) from van\_review

--Positive Comment

select case when host\_is\_superhost = 'true' then 'Super host'

else 'host' end as host\_is\_superhost,review\_count from (

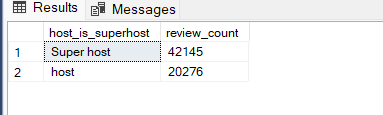
select c.host\_is\_superhost,count( a.reviewer\_id) as review\_count from van\_review as a left join van\_list as b on a.listing\_id=b.id

left join van\_host as c on b.host\_id=c.host\_id

where (a.comments like '%beautiful%'or a.comments like '%tastefull%' or a.comments like '%perfect%'

or a.comments like '%amazing%' or a.comments like '%very clean%' or a.comments like '%gracious%'or a.comments like

'%recommended%' or a.comments like '%very welcoming%' or a.comments like '%best host%')and c.host\_is\_superhost is not null

group by c.host\_is\_superhost ) as mm  


---Negative Comment

select case when host\_is\_superhost = 'true' then 'Super host'

else 'host' end as host\_is\_superhost,review\_count from (

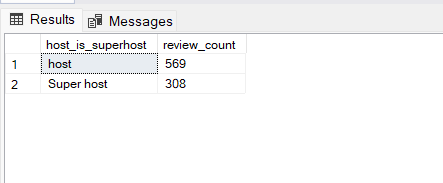
select c.host\_is\_superhost,count( a.reviewer\_id) as review\_count from van\_review as a left join van\_list as b on a.listing\_id=b.id

left join van\_host as c on b.host\_id=c.host\_id

where (a.comments like '%slow wifi%'or a.comments like '%canceled%' or a.comments like '%not recommended%'

or a.comments like '%unprofessional%' or a.comments like '%not good%' or a.comments like '%rude%'

) and c.host\_is\_superhost is not null

group by c.host\_is\_superhost ) as mo  


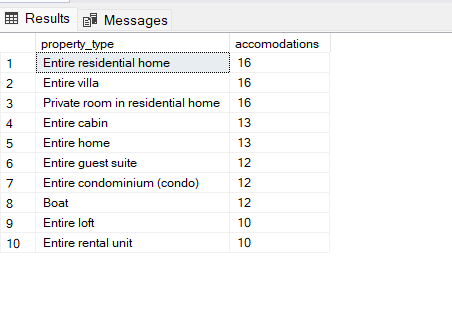
---d Analyze do Super Hosts tend to have large property types as compared to Other Hosts

1). In this analysis we sorted the property type on the basis of total accommodation  
2) We found average accommodation and selected the top 10 property type on basis of that.

select top 10 property\_type,max(accommodates) from van\_list

where accommodates > (select avg(accommodates) from van\_list)

group by property\_type

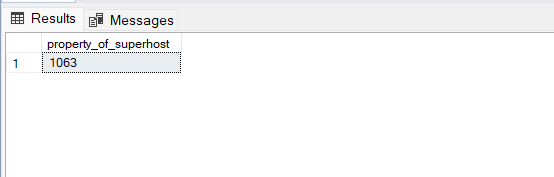
order by max(accommodates) desc  


select count(distinct a.host\_id) as property\_of\_superhost from van\_host as a inner join van\_list as b

on a.host\_id=b.host\_id where b.property\_type in ('Entire residential home','Entire villa',

'Private room in residential home','Entire cabin','Entire home','Entire guest suite',

'Entire condominium (condo)','Boat','Entire loft','Entire rental unit')

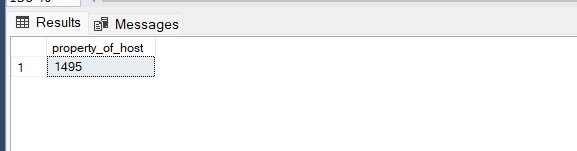
and a.host\_is\_superhost = 'true'  


(select count(distinct a.host\_id) as property\_of\_host from van\_host as a left join van\_list as b

on a.host\_id=b.host\_id where b.property\_type in ('Entire residential home','Entire villa',

'Private room in residential home','Entire cabin','Entire home','Entire guest suite',

'Entire condominium (condo)','Boat','Entire loft','Entire rental unit')

and a.host\_is\_superhost = 'false')  


select count(\*) from van\_host where host\_is\_superhost = 'false'

---e Analyze the average price and availability of the listings for the upcoming year between Super Hosts and Other Hosts

1) In this analysis we found average price for host and super host for this year and upcoming year  
2)Also we found average availability for host and super host

--van avg price

Select\* From(

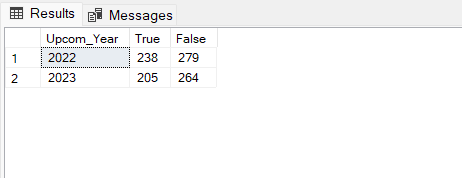
Select C.host\_is\_superhost ,Round(AVG(A.adjusted\_price),0) Avg\_Price,year(Date) Upcom\_Year

From van\_avail A inner join van\_list B on A.listing\_id = B.id

inner join van\_host C on B.host\_id = C.host\_id where A.available = 'true' and C.host\_is\_superhost is not null

Group By C.host\_is\_superhost,year(Date)) Z

Pivot(Avg(Avg\_price) For host\_is\_superhost in ([True],[False])) As Pvt;



---van avilable and not availavle list

Select\* From(

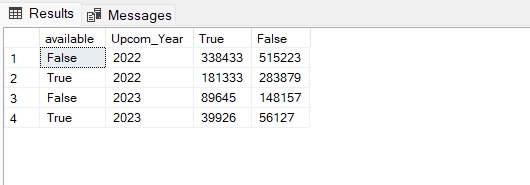
Select C.host\_is\_superhost,A.available ,Count(A.id) total\_Available,year(Date) Upcom\_Year

From van\_avail A inner join van\_list B on A.listing\_id = B.id

inner join van\_host C on B.host\_id = C.host\_id where C.host\_is\_superhost is not null

Group By C.host\_is\_superhost,A.available,A.listing\_id,(Date)) Z

Pivot(Sum(total\_Available) For host\_is\_superhost in ([True],[False])) As Pvt;



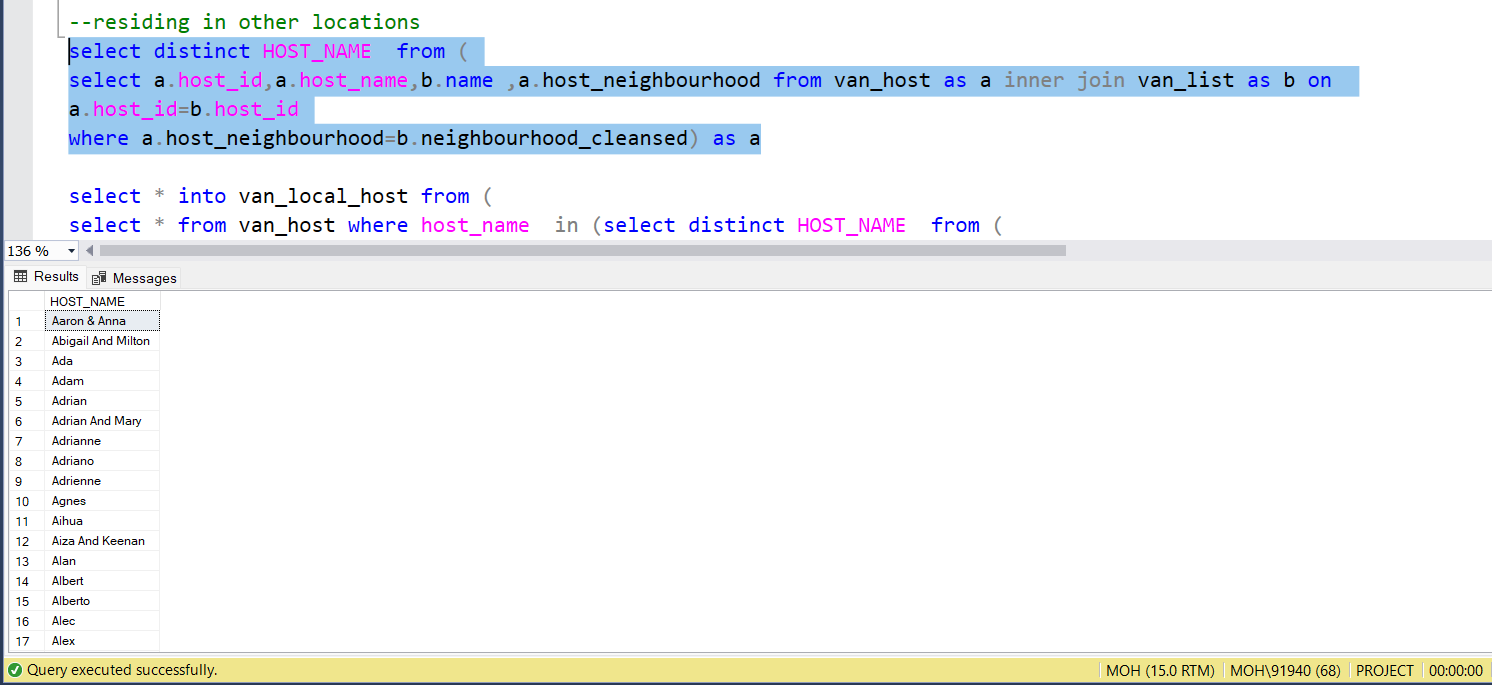
--f. Analyze if there is some difference in above mentioned trends between Local Hosts or Hosts  
1) For this analysis We considered local host whose location is same as neighbour location of its listing.  
2) And following analysis were made

--residing in other locations

select distinct HOST\_NAME from (

select a.host\_id,a.host\_name,b.name ,a.host\_neighbourhood from van\_host as a inner join van\_list as b on

a.host\_id=b.host\_id

where a.host\_neighbourhood=b.neighbourhood\_cleansed) as a  


select \* into van\_local\_host from (

select \* from van\_host where host\_name in (select distinct HOST\_NAME from (

select a.host\_id,a.host\_name,b.name ,a.host\_neighbourhood from van\_host as a inner join van\_list as b on

a.host\_id=b.host\_id

where a.host\_neighbourhood =b.neighbourhood\_cleansed order by a.host\_name) as a)) as mo

select \* into van\_other\_host from (

select \* from van\_host where host\_name not in (select distinct HOST\_NAME from (

select a.host\_id,a.host\_name,b.name ,a.host\_neighbourhood from van\_host as a inner join van\_list as b on

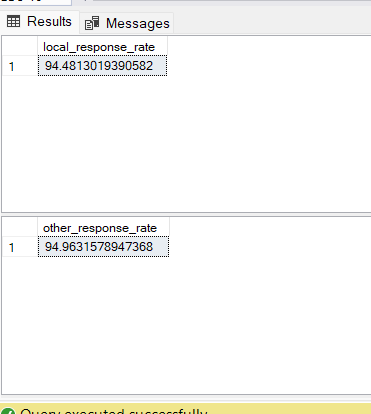
a.host\_id=b.host\_id

where a.host\_neighbourhood =b.neighbourhood\_cleansed) as a )) as mo

----

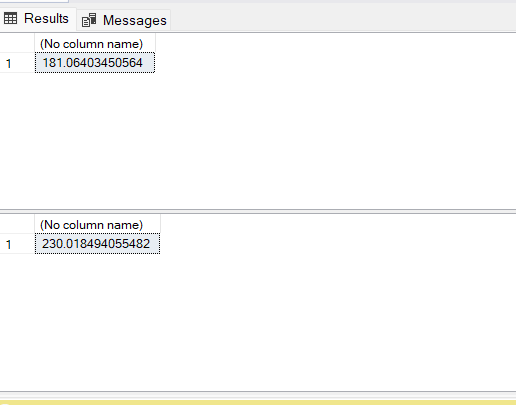
----- response rate

select avg(a.host\_response\_rate) from van\_local\_host

select avg(a.host\_response\_rate) from van\_other\_host   


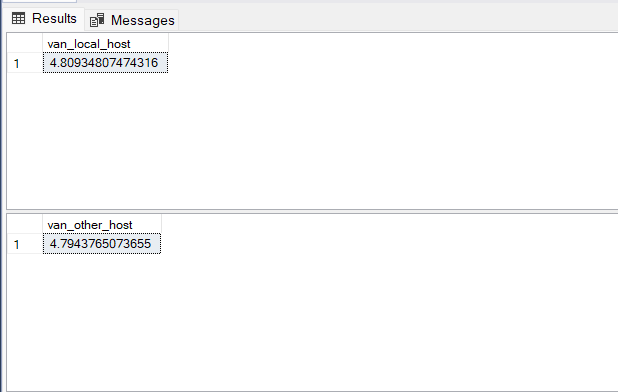
-----avg price

select avg(b.price) from van\_local\_host as a inner join van\_list as b on a.host\_id=b.host\_id

select avg(b.price) from van\_other\_host as a inner join van\_list as b on a.host\_id=b.host\_id  


---avg rating

select avg(b.review\_scores\_accuracy) from van\_local\_host as a inner join van\_list as b on a.host\_id=b.host\_id

select avg(b.review\_scores\_accuracy) from van\_other\_host as a inner join van\_list as b on a.host\_id=b.host\_id  


--supper host

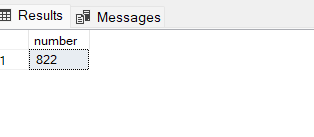
select count(\*) as number from (

select \* from van\_host where host\_name in (select HOST\_NAME from (

select a.host\_id,a.host\_name,b.name ,a.host\_neighbourhood from van\_host as a inner join van\_list as b on

a.host\_id=b.host\_id

where a.host\_neighbourhood =b.neighbourhood\_cleansed ) as a)) as mo

where host\_is\_superhost='true'  


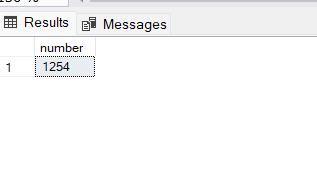
select count(\*) as number from (

select \* from van\_host where host\_name in (select HOST\_NAME from (

select a.host\_id,a.host\_name,b.name ,a.host\_neighbourhood from van\_host as a inner join van\_list as b on

a.host\_id=b.host\_id

where a.host\_neighbourhood =b.neighbourhood\_cleansed ) as a)) as mo

where host\_is\_superhost='false'  


--host

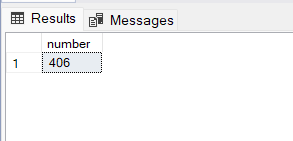
select count(\*) as number from (

select \* from van\_host where host\_name not in (select HOST\_NAME from (

select a.host\_id,a.host\_name,b.name ,a.host\_neighbourhood from van\_host as a inner join van\_list as b on

a.host\_id=b.host\_id

where a.host\_neighbourhood =b.neighbourhood\_cleansed ) as a)) as mo

where host\_is\_superhost='true'  


select count(\*) as number from (

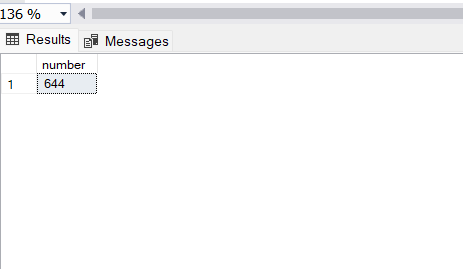
select \* from van\_host where host\_name not in (select HOST\_NAME from (

select a.host\_id,a.host\_name,b.name ,a.host\_neighbourhood from van\_host as a inner join van\_list as b on

a.host\_id=b.host\_id

where a.host\_neighbourhood =b.neighbourhood\_cleansed ) as a)) as mo

where host\_is\_superhost='false'



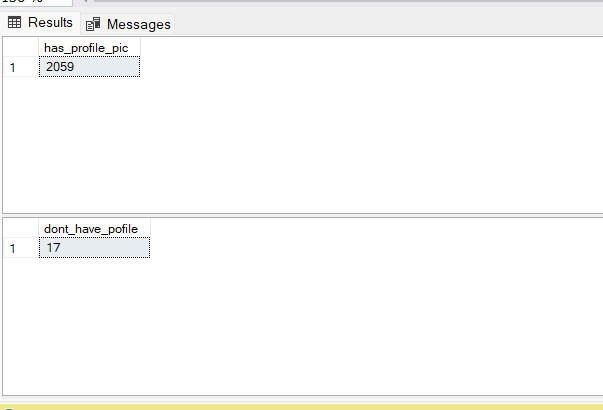
-----profile verified

---local

select \* from van\_local\_host

select count(\*) from van\_local\_host

select count(\*) from van\_local\_host where host\_has\_profile\_pic='true'

select count(\*) from van\_local\_host where host\_has\_profile\_pic='false'  


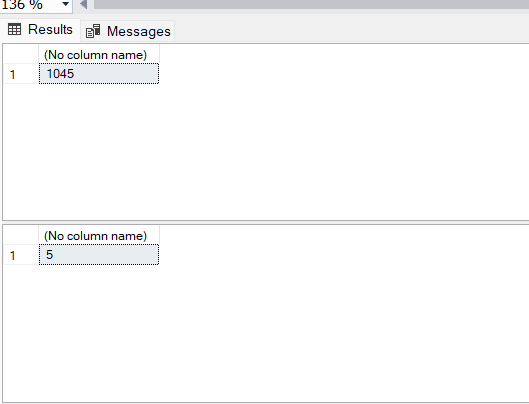
---other

select \* from van\_other\_host

select count(\*) from van\_other\_host

select count(\*) from van\_other\_host where host\_has\_profile\_pic='true'

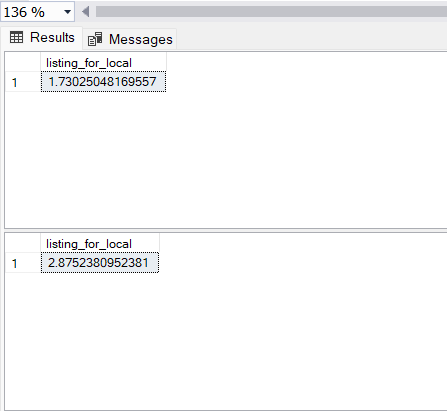
select count(\*) from van\_other\_host where host\_has\_profile\_pic='false'



---listing

select avg(host\_listings\_count) as listing\_for\_local from van\_local\_host

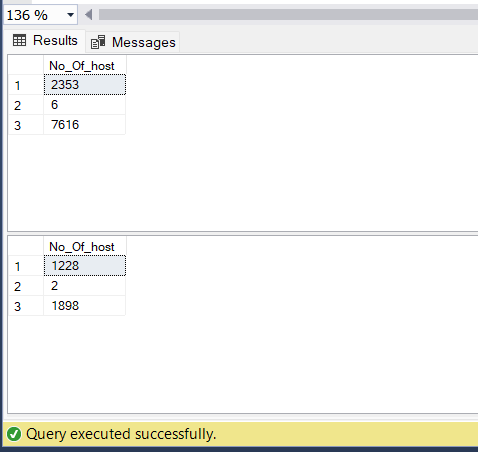
select avg(host\_listings\_count) as listing\_for\_local from van\_other\_host



---Q7 Analyze the above trends for the two cities for which data has been provided and provide insights on comparison

Select Count(host\_id) No\_Of\_host From toro\_host Group BY host\_is\_superhost

Select Count(host\_id) No\_Of\_host From van\_host Group BY host\_is\_superhost



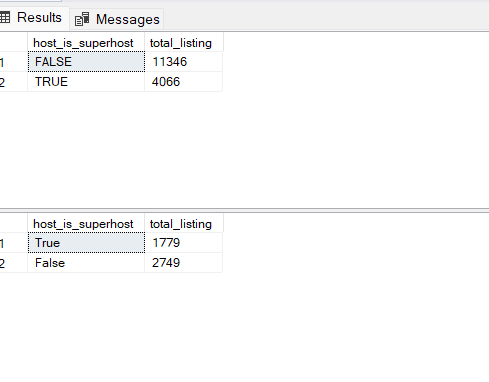
Select A.host\_is\_superhost,Count(B.id) total\_listing From toro\_host A inner join toro\_list B

on A.host\_id = B.host\_id where A.host\_is\_superhost is not null

Group By A.host\_is\_superhost

Select A.host\_is\_superhost,Count(B.id) total\_listing From van\_host A inner join van\_list B

on A.host\_id = B.host\_id where A.host\_is\_superhost is not null

Group By A.host\_is\_superhost  


--torr avg price

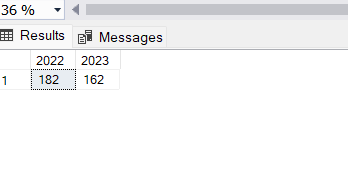
Select\*From(

Select Round(AVG(A.adjusted\_price),0) Avg\_Price,year(Date) Upcom\_Year

From toro\_avail A inner join toro\_list B on A.listing\_id = B.id

inner join toro\_host C on B.host\_id = C.host\_id where A.available = 'true' and C.host\_is\_superhost is not null

Group By year(Date))z

Pivot (Avg(Avg\_Price) For Upcom\_Year in ([2022],[2023])) pvt1  


--Van avg price

Select\*From(

Select Round(AVG(A.adjusted\_price),0) Avg\_Price,year(Date) Upcom\_Year

From van\_avail A inner join van\_list B on A.listing\_id = B.id

inner join van\_host C on B.host\_id = C.host\_id where A.available = 'true' and C.host\_is\_superhost is not null

Group By year(Date))z

Pivot (Avg(Avg\_Price) For Upcom\_Year in ([2022],[2023])) pvt2

