//* Athens*//

```
/*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 scores, average no of bookings per month, etc.*/

/*Total No. of Host and Superhost*/
select case
when host_is_superhost=1 then 'SuperHost'
when host_is_superhost=0 then 'Host' end as host_superhost,No_of_host from
(select host_is_superhost,count(*) as No_of_host from host_athens_df where
host_is_superhost=1 or host_is_superhost=0 group by host_is_superhost)a

/*Response Rate*/

select case
when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as
host_Superhost,AvgResponseRate,MinResponseRate,MaxResponseRate from
(select host_is_superhost,avg(host_response_rate) as AvgResponseRate, min
(host_response_rate) as MinResponseRate , max(host_response_rate) as maxResponseRate
from host_athens_df where host_is_superhost = 1 and host_response_rate !=0 or
host_is_superhost = 0 and host_response_rate!=0 group by host_is_superhost)c

/*No of host having response rate > avg(response rate)*/

select case
when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as
host_Superhost,No_Of_Host from
(select host_is_superhost,count(host_id) as No_OF_Host from host_athens_df where
host_response_rate >
(select avg(host_response_rate) as Avg_Response_rate from host_athens_df) group by
host_is_superhost)c

/*Listing Count*/

select case
when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as
host_Superhost,AvgListingCount,MinListingCount,MaxListingCount from
(select host_is_superhost,avg(host_listings_count)as Avglistingcount,
min(host_listings_count) as Minlistingcount,
max(host_listings_count) as Maxlistingcount
from host_athens_df where host_is_superhost=1 or host_is_superhost=0
group by host_is_superhost) aa

/*Listing Count > Avg. listing count*/

select case
when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as
host_Superhost,No_Of_Host from
(select host_is_superhost,count(host_id) as No_OF_Host from host_athens_df where
host_listings_count >
(select avg(host_listings_count) as Avg_Listing_rate from host_athens_df) group by
host_is_superhost)c
```

```sql
/*Response Time*/

select case
when host_is_superhost=0 then 'Host'
when host_is_superhost=1 then 'Superhost' end as
host_Superhost,host_response_time,TotalHost from
(select       host_is_superhost,host_response_time,count(host_id) as TotalHost from
host_athens_df where
host_is_superhost=0 and host_response_time is not null or host_is_superhost=1 and
host_response_time is not null
group by host_is_superhost,host_response_time ) ccc order by host_Superhost

/*Acceptance Rate*/

select case
when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as
host_Superhost,AvgAcceptanceRate,MinAcceptanceRate,MaxAcceptanceRate from
(select host_is_superhost,avg(host_acceptance_rate) as
AvgAcceptanceRate,min(host_acceptance_rate) as MinAcceptanceRate,
max(host_acceptance_rate) as MaxAcceptanceRate from
host_athens_df where host_is_superhost = 1 or host_is_superhost = 0 group by
host_is_superhost)c

/*No of host having acceptance rate > avg(acceptance rate)-*/

select case
when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as host_Superhost,No_Of_Host from
(select host_is_superhost,count(host_id) as No_OF_Host from host_athens_df where
host_acceptance_rate >
(select avg(host_acceptance_rate) as Avg_Acceptance_rate from host_athens_df) group by
host_is_superhost)c

/*profile Pic-*/

select case
when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as
host_Superhost,host_has_profile_pic,TotalHost from
(select host_is_superhost,host_has_profile_pic,count(host_id) as TotalHost from
host_athens_df
where host_is_superhost =0 and host_has_profile_pic is not null or host_is_superhost
=1 and host_has_profile_pic is not null
group by host_is_superhost,host_has_profile_pic ) ccc order by host_Superhost

/*Identity Verified*/

select case
when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as host_Superhost,
case when host_identity_verified = 0 then 'NO'
when host_identity_verified = 1 then 'YES'end as host_identity_verified,TotalHost from
(select host_is_superhost,host_identity_verified,count(host_id) as TotalHost from
host_athens_df
where host_is_superhost =0 and host_identity_verified is not null or host_is_superhost
=1 and host_identity_verified is not null
group by host_is_superhost,host_identity_verified ) ccc order by host_Superhost


/*Instant Booking*/
```

```sql
select case when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as host_Superhost,NO_of_host from
(select b.host_is_superhost,count(distinct a.host_id) as NO_of_host from
listing_athens_df a
inner join host_athens_df b on a.host_id=b.host_id
where instant_bookable = 1 group by b.host_is_superhost)c where host_is_superhost=0 or
host_is_superhost=1


/*Review*/

select avg(a.review_scores_rating) as review_scores_rating ,avg(b.host_is_superhost)
from listing_athens_df a
inner join host_athens_df b on a.host_id = b.host_id
where a.review_scores_rating!=0 and b.host_is_superhost=0 or a.review_scores_rating!=0
and b.host_is_superhost=1
group by b.host_is_superhost


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

/*From the Above Analysis The Top 3 Crucial Metrics Are*/

1.Response Rate
2.Response Time
3.Acceptance Rate

/*c. Analyze how the comments of reviewers vary for listings of Super Hosts vs Other
Hosts
(Extract words from the comments provided by the reviewers)*/

/*d. Analyze do Super Hosts tend to have large property types as compared to Other
Hosts*/

/*Large property Type*/

select case when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as host_Superhost,No_Of_property from
(select host_is_superhost , count(property_type) as No_Of_Property from
(select a.property_type,b.host_is_superhost from listing_athens_df a inner join
host_athens_df b on a.host_id=b.host_id
where a.property_type like '%entire%' or
a.property_type like '%RV%' or a.property_type like '%Private%' or a.property_type
like '%Sahred%' or a.property_type like '%Room%'
or a.property_type like '%Cycladic%' or a.property_type like '%Boat%' or
a.property_type like '%Floor%' or a.property_type like '%Tiny%'
or a.property_type like '%Earth%')c
where host_is_superhost=0 or host_is_superhost=1 group by host_is_superhost ) cc;

select distinct property_type from listing_athens_df;
```

//*Thessaloniki*//

```
/*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 scores, average no of bookings per month, etc.*/

/*Total No. of Host and Superhost*/

select case
when host_is_superhost=1 then 'SuperHost'
when host_is_superhost=0 then 'Host' end as host_superhost,No_of_host from
(select host_is_superhost,count(*) as No_of_host from host_thessaloniki_df where
host_is_superhost=1 or host_is_superhost=0 group by host_is_superhost)a

/*Response Rate*/

select case
when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as
host_Superhost,AvgResponseRate,MinResponseRate,MaxResponseRate from
(select host_is_superhost,avg(host_response_rate) as AvgResponseRate, min
(host_response_rate) as MinResponseRate , max(host_response_rate) as maxResponseRate
from host_thessaloniki_df where host_is_superhost = 1 and host_response_rate !=0 or
host_is_superhost = 0 and host_response_rate!=0 group by host_is_superhost)c

/*No of host having response rate > avg(response rate)*/

select case
when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as
host_Superhost,No_Of_Host from
(select host_is_superhost,count(host_id) as No_OF_Host from host_thessaloniki_df where
host_response_rate >
(select avg(host_response_rate) as Avg_Response_rate from host_thessaloniki_df) group
by
host_is_superhost)c

/*Listing Count*/

select case
when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as
host_Superhost,AvgListingCount,MinListingCount,MaxListingCount from
(select host_is_superhost,avg(host_listings_count)as Avglistingcount,
min(host_listings_count) as Minlistingcount,
max(host_listings_count) as Maxlistingcount
from host_thessaloniki_df where host_is_superhost=1 or host_is_superhost=0
group by host_is_superhost) aa

/*Listing Count > Avg. listing count*/

select case
when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as
host_Superhost,No_Of_Host from
(select host_is_superhost,count(host_id) as No_OF_Host from .host_thessaloniki_df
where
```

```
host_listings_count >
(select avg(host_listings_count) as Avg_Listing_rate from .host_thessaloniki_df) group
by
host_is_superhost)c

/*Response Time*/

select case
when host_is_superhost=0 then 'Host'
when host_is_superhost=1 then 'Superhost' end as
host_Superhost,host_response_time,TotalHost from
(select       host_is_superhost,host_response_time,count(host_id) as TotalHost from
host_thessaloniki_df where
host_is_superhost=0 and host_response_time is not null or host_is_superhost=1 and
host_response_time is not null
group by host_is_superhost,host_response_time ) ccc order by host_Superhost

/*Acceptance Rate*/

select case
when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as
host_Superhost,AvgAcceptanceRate,MinAcceptanceRate,MaxAcceptanceRate from
(select host_is_superhost,avg(host_acceptance_rate) as
AvgAcceptanceRate,min(host_acceptance_rate) as MinAcceptanceRate,
max(host_acceptance_rate) as MaxAcceptanceRate from
host_thessaloniki_df where host_is_superhost = 1 or host_is_superhost = 0 group by
host_is_superhost)c

/*No of host having acceptance rate > avg(acceptance rate)-*/

select case
when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as host_Superhost,No_Of_Host from
(select host_is_superhost,count(host_id) as No_OF_Host from host_thessaloniki_df where
host_acceptance_rate >
(select avg(host_acceptance_rate) as Avg_Acceptance_rate from host_thessaloniki_df)
group by host_is_superhost)c

/*profile Pic-*/

select case
when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as
host_Superhost,host_has_profile_pic,TotalHost from
(select host_is_superhost,host_has_profile_pic,count(host_id) as TotalHost from
host_thessaloniki_df
where host_is_superhost =0 and host_has_profile_pic is not null or host_is_superhost
=1 and host_has_profile_pic is not null
group by host_is_superhost,host_has_profile_pic ) ccc order by host_Superhost

/*Identity Verified*/

select case
when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as host_Superhost,
case when host_identity_verified = 0 then 'NO'
when host_identity_verified = 1 then 'YES'end as host_identity_verified,TotalHost from
(select host_is_superhost,host_identity_verified,count(host_id) as TotalHost from
host_thessaloniki_df
where host_is_superhost =0 and host_identity_verified is not null or host_is_superhost
=1 and host_identity_verified is not null
```

```
group by host_is_superhost,host_identity_verified ) ccc order by host_Superhost

/*Comments*/

select q.hosts,Avg(count_comments) as avg_comments from
(select h.host_id,h.host_is_superhost as hosts,count(r.comments) as count_comments
from host_thessaloniki_df h
join listing_thessaloniki_df l on l.host_id = h.host_id
join review_thessaloniki_df r on r.listing_id = l.id
join df_thessaloniki_availability a on a.listing_id = l.id
group by h.host_id,h.host_is_superhost) q
where q.hosts is not null
group by q.hosts

/*Avg. Monthly Booking-**/

select case when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as
host_Superhost,Average_montly_Booking from
(select h.host_is_superhost , avg (g.Avg_booking) as Average_montly_Booking from
(select host_id,avg(Total_booking) as Avg_booking from
(select distinct b.host_id,month(a.date) as month,count(a.available) as Total_Booking
from df_thessaloniki_availability a
inner join listing_thessaloniki_df b on a.listing_id=b.id where a.available = 'False'
group by host_id,month(a.date))c
group by host_id )g inner join host_thessaloniki_df h on g.host_id=h.host_id where
h.host_is_superhost=0 or h.host_is_superhost =1
group by h.host_is_superhost)cccc

/*Instant Booking*/

select case when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as host_Superhost,NO_of_host from
(select b.host_is_superhost,count(distinct a.host_id) as NO_of_host from
listing_thessaloniki_df a
inner join host_thessaloniki_df b on a.host_id=b.host_id
where instant_bookable = 1 group by b.host_is_superhost)c where host_is_superhost=0 or
host_is_superhost=1

/*Review*/

select avg(a.review_scores_rating) as review_scores_rating ,avg(b.host_is_superhost)
from listing_thessaloniki_df a
inner join host_thessaloniki_df b on a.host_id = b.host_id
where a.review_scores_rating!=0 and b.host_is_superhost=0 or a.review_scores_rating!=0
and b.host_is_superhost=1
group by b.host_is_superhost

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

/*From the Above Analysis The Top 3 Crucial Metrics Are*/

1.Response Rate
2.Response Time
3.Acceptance Rate

/*c. Analyze how the comments of reviewers vary for listings of Super Hosts vs Other
Hosts
(Extract words from the comments provided by the reviewers)*/
```

```sql
select case
when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as host_Superhost,
No_of_Positive_comment from
(select host_is_superhost,sum(case when
comments like '%great%' or comments like '%nice%' or comments like '%wonderful%' or
comments like '%brilliant%'or
comments like '%great location%' or comments like '%good%' or comments like '%lovely%'
or comments like '%friendly%'or
comments like '%perfect%' or comments like '%beautiful%' or comments like '%definetly
stay%' or comments like '%excellent%'
or comments like '%highly recommended%'
then 1 else 0 end) as No_of_Positive_comment from
(select a.comments,c.host_is_superhost from review_thessaloniki_df a
inner join listing_thessaloniki_df b on a.listing_id=b.id
inner join host_thessaloniki_df c on b.host_id = c.host_id)dd where
host_is_superhost=0 or host_is_superhost=1 group by host_is_superhost)ee

/*d. Analyze do Super Hosts tend to have large property types as compared to Other
Hosts*/


/*Large property Type*/

select case when host_is_superhost = 0 then 'Host'
when host_is_superhost = 1 then 'SuperHost' end as host_Superhost,No_Of_property from
(select host_is_superhost , count(property_type) as No_Of_Property from
(select a.property_type,b.host_is_superhost from listing_thessaloniki_df a inner join
host_thessaloniki_df b on a.host_id=b.host_id
where a.property_type like '%entire%' or
a.property_type like '%RV%' or a.property_type like '%Private%' or a.property_type
like '%Sahred%' or a.property_type like '%Room%'
or a.property_type like '%Tower%' or a.property_type like '%Earth%' or a.property_type
like '%Dome%' or a.property_type like '%Boat%' )c
where host_is_superhost=0 or host_is_superhost=1 group by host_is_superhost ) cc;
```