SQL PROJECT – HOST BEHAVIOR ANALYSIS FOR PROPERTY RENTAL COMPANY

Name- PawanKumar R. Bone {Student Code- (pd14_096)}

Name- Abhiraj s {student code-(pd14_053)}

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.

Query-

```
--Athens

select a.host_is_superhost,count(b.instant_bookable)
number_of_instant_booking,round(avg(b.review_scores_rating),2) as overall_avg_rating ,
round(avg(a.host_response_rate),2) as avg_of_Resp_rate,
COUNT(a.host_identity_verified) Number_of_verified_profile from host_athens_df as a
inner join listing_athens as b
on a.host_id=b.host_id
where a.host_is_superhost in (0,1)
group by a.host_is_superhost

--thessaloniki

select a.host_is_superhost,count(b.instant_bookable)
number_of_instant_booking,round(avg(b.review_scores_rating),2) as overall_avg_rating ,
round(avg(a.host_response_rate),2) as avg_of_Resp_rate,
COUNT(a.host_identity_verified) Number_of_verified_profile from host_thessaloniki as a
inner join listing_thessaloniki as b
on a.host_id=b.host_id
where a.host_is_superhost in (0,1)
group by a.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.

Query-

```
--Athens

select a.host_is_superhost,round(avg(a.host_acceptance_rate),2) avg_Accpt_rating
,round(avg(a.host_response_rate),2) avg_response_rate,
round(avg(b.review_scores_communication),2) as Communication_rating
from host_athens_df as a inner join listing_athens as b
on a.host_id=b.host_id
group by a.host_is_superhost
having a.host_is_superhost in (0,1)
```

```
--thessaloniki

select a.host_is_superhost,round(avg(a.host_acceptance_rate),2) avg_Accpt_rating
,round(avg(a.host_response_rate),2) avg_response_rate,
round(avg(b.review_scores_communication),2) as Communication_rating
from host_thessaloniki as a inner join listing_thessaloniki as b
on a.host_id=b.host_id
group by a.host_is_superhost
having a.host_is_superhost in (0,1)
```

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)

Query-

```
--Athens

select a.host_is_superhost,count(c.comments) as Good_comments
from host_athens_df a join listing_athens b on a.host_id = b.host_id
join review_athens c on b.id = c.listing_id
where c.comments like '%Very nice%'
group by a.host_is_superhost

select a.host_is_superhost,count(c.comments) as Bad_comments
from host_athens_df a join listing_athens b on a.host_id = b.host_id
join review_athens c on b.id = c.listing_id
where c.comments like '%Bad%'
group by a.host_is_superhost

--thessaloniki

select a.host_is_superhost,count(c.comments) as Bad_comments
from host_thessaloniki a join listing_thessaloniki b on a.host_id = b.host_id
join review_thessaloniki c on b.id = c.listing_id
where c.comments like '%Perfect%'
group by a.host_is_superhost



select a.host_is_superhost,count(c.comments) as Bad_comments
from host_thessaloniki a join listing_thessaloniki b on a.host_id = b.host_id
join review_thessaloniki c on b.id = c.listing_id
where c.comments like '%Bad%'
group by a.host_is_superhost
```

D. Analyze do Super Hosts tend to have large property types as compared to Other Hosts

Query-

```
--Athens

select a.host_is_superhost,count(distinct b.property_type) no_of_prop_type,
count(b.property_type) as counting_of_property_type from host_athens_df
as a inner join listing_athens as b
on a.host_id=b.host_id
where a.host_is_superhost in (0,1)
```

```
group by a.host_is_superhost

--After analysis Other host having large property types

--thessaloniki

select a.host_is_superhost,count(distinct b.property_type) no_of_prop_type,
count(b.property_type) as counting_of_property_type from host_thessaloniki
as a inner join listing_thessaloniki as b
on a.host_id=b.host_id
where a.host_is_superhost in (0,1)
group by a.host_is_superhost
```

E. Analyze the average price and availability of the listings for the upcoming year between Super Hosts and Other Hosts

Query-

```
##Average_price_and_availability_of_super_hosts

select date,id,listing_id,host_listings_count,avg(price) as
Average_Price,host_is_superhost from
df_vancouver_availability df
join host_vancouver_df hv on df.id = hv.host_id
where year(date)='2023' and host_is_superhost='1'
group by date,id,listing_id,host_listings_count,host_is_superhost
order by date;

##Average_price_and_availability_of_non_super_hosts

select date,id,listing_id,host_listings_count,avg(price) as
Average_Price,host_is_superhost
from df_vancouver_availability df
join host_vancouver_df hv on df.id = hv.host_id
where year(date)='2023' and host_is_superhost='0'
group by date,id,listing_id,host_listings_count,host_is_superhost
order by date;
```

F. Analyze if there is some difference in above mentioned trends between Local Hosts or Hosts residing in other locations

Query-

```
##Do_super_hosts_tend_to_have_large_property
##Vancouver: -
select property_type,room_type,accommodates,bedrooms,beds,host_neighbourhood from
host_vancouver_df hv
join listing_vancouver_df lv on hv.host_id=lv.host_id where host_is_superhost='0';

select property_type,room_type,accommodates,bedrooms,beds,host_neighbourhood from
host_vancouver_df hv
join listing_vancouver_df lv on hv.host_id=lv.host_id where host_is_superhost='1';

##Toronto
select property_type,room_type,accommodates,bedrooms,beds,host_neighbourhood from
host_toronto_df ht
join listing_toronto_df lt on ht.host_id=lt.host_id where host_is_superhost='0';
```

```
select property_type,room_type,accommodates,bedrooms,beds,host_neighbourhood from
host_toronto_df ht
join listing_toronto_df lt on ht.host_id=lt.host_id where host_is_superhost='1';

##Average_price_and_availability_of_super_hosts_and_hosts
##Vancouver
select date,id,listing_id,host_listings_count,avg(price) as Average_Price from
df_vancouver_availability df
join host_vancouver_df hv on df.id = hv.host_id
group by date,id,listing_id,host_listings_count
order by date;

##Toronto
select date,id,listing_id,host_listings_count,avg(price) as Average_Price from
df_toronto_availability df
join host_toronto_df hv on df.id = hv.host_id
group by date,id,listing_id,host_listings_count
order by date;
```

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

Query-

```
--Vancouver
select
date,price,minimum_nights,maximum_nights,host_id,host_name,host_response_time,host_res
ponse_rate,
host_is_superhost,host_neighbourhood
from df_vancouver_availability dv
join host_vancouver_df hv
on dv.id = hv.host_id
order by date;

select
date,price,minimum_nights,maximum_nights,host_id,host_name,host_response_time,host_res
ponse_rate,
host_is_superhost,host_neighbourhood
from df_toronto_availability dt
join host_toronto_df ht
on dt.id=ht.host_id
order by date;

--Toronto
select
lv.id,property_type,hv.host_location,host_neighbourhood,accommodates,review_scores_rat
ing,
review_scores_location,date
from listing_vancouver_df lv
join review_vancouver_df rv
on lv.id=rv.listing_id
join host_vancouver_df hv on lv.host_id=hv.host_id
order by review_scores_rating desc;

select
lt.id,property_type,host_location,host_neighbourhood,accommodates,review_scores_rating
,
review_scores_location,date
from listing_toronto_df lt
join review_toronto_df rt
on lt.id=rt.listing_id
```

```
join host_toronto_df ht on lt.host_id=ht.host_id
order by review_scores_rating desc;
```