Project

Host Behaviour Analysis For Property Rental Company

SUNIL SINGH

Queries used to derive insights for the project.

```
--athens
sELECT * FROM df_athens_availability
sELECT * FROM host_athens_df
sELECT * FROM listing_athens_df
sELECT * FROM review_athens_df
update host_athens_df set host_response_rate = 0 where
host_response_rate is null
update listing_athens_df set review_scores_value = 0 where
review_scores_value is null
--thessaloniki
sELECT * FROM df_thessaloniki_availability
sELECT * FROM host_thessaloniki_df
sELECT * FROM listing_thessaloniki_df
sELECT * FROM review_thessaloniki_df
update host_thessaloniki_df set host_response_rate = 0 where
host_response_rate is null
update listing_thessaloniki_df set review_scores_value = 0 where
review_scores_value is null


--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.


--athens

select host_is_superhost, avg(host_acceptance_rate) as avg_acceptance from
host_athens_df group by host_is_superhost
having host_is_superhost = 'true' or host_is_superhost = 'false';

select host_is_superhost, avg(host_response_rate) as avg_response from host_athens_df
group by host_is_superhost
having host_is_superhost = 'true' or host_is_superhost = 'false'

select a.host_is_superhost , avg(b.review_scores_value) as avg_review
from host_athens_df as a left join listing_athens_df as b on a.host_id=b.host_id group
by
host_is_superhost
having host_is_superhost = 'true' or host_is_superhost = 'false'

select host_is_superhost, count(host_has_profile_pic) got_profile_pic
from host_athens_df
where host_is_superhost = 'true' or host_is_superhost = 'false'
group by host_is_superhost
```

```sql
select host_is_superhost, count(host_identity_verified) got_identity_verified
from host_athens_df
where host_is_superhost = 'true' or host_is_superhost = 'false'
group by host_is_superhost


select h.host_is_superhost, count(L.instant_bookable) total_instant_book
from  host_athens_df h LEFT join listing_athens_df L on h.host_id = L.host_id
where h.host_is_superhost = 'true' or host_is_superhost = 'false'
group by H.host_is_superhost;

select top 10 a.host_is_superhost,datename(month,c.date) as
month_number,datename(year,c.date) as year_number,
count(c.id) total_booking
from host_athens_df as a inner join listing_athens_df as b on a.host_id=b.host_id
join df_athens_availability c on c.id = b.id
group by host_is_superhost, datename(month,c.date), datename(year,c.date)
having host_is_superhost = 'true' or host_is_superhost = 'false'
order by datename(year,c.date) desc;



--thessaloniki




select host_is_superhost, avg(host_acceptance_rate) as avg_acceptance from
host_thessaloniki_df group by host_is_superhost
having host_is_superhost = 'true' or host_is_superhost = 'false';

select host_is_superhost, avg(host_response_rate) as avg_response from
host_thessaloniki_df
group by host_is_superhost
having host_is_superhost = 'true' or host_is_superhost = 'false'

select a.host_is_superhost , avg(b.review_scores_value) as avg_review
from host_thessaloniki_df as a left join listing_thessaloniki_df as b on
a.host_id=b.host_id group by
host_is_superhost
having host_is_superhost = 'true' or host_is_superhost = 'false'

select host_is_superhost, count(host_has_profile_pic) got_profile_pic
from host_thessaloniki_df
where host_is_superhost = 'true' or host_is_superhost = 'false'
group by host_is_superhost

select host_is_superhost, count(host_identity_verified) got_identity_verified
from host_thessaloniki_df
where host_is_superhost = 'true' or host_is_superhost = 'false'
group by host_is_superhost

select h.host_is_superhost, count(L.instant_bookable) total_instant_book
from  host_thessaloniki_df h LEFT join listing_thessaloniki_df L on h.host_id =
L.host_id
where h.host_is_superhost = 'true' or host_is_superhost = 'false'
group by H.host_is_superhost;

select top 10 a.host_is_superhost,datename(month,c.date) as
month_number,datename(year,c.date) as year_number,
```

```
count(c.id) total_booking
from host_thessaloniki_df as a inner join listing_thessaloniki_df as b on
a.host_id=b.host_id
join df_thessaloniki_availability c on c.id = b.id
group by host_is_superhost, datename(month,c.date), datename(year,c.date)
having host_is_superhost = 'true' or host_is_superhost = 'false'
order by datename(year,c.date) desc;



--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)


--athens

select h.host_is_superhost,l.host_id,comments from host_athens_df h join
listing_athens_df l on h.host_id=l.host_id
join review_athens_df r on l.id=r.listing_id where h.host_is_superhost = 'True' and
comments like  '%good%' or comments like
'%excellent%' or comments like '%outstanding%';

--thessaloniki


select h.host_is_superhost,l.host_id,comments from host_thessaloniki_df h join
listing_thessaloniki_df l on h.host_id=l.host_id
join review_thessaloniki_df r on l.id=r.listing_id where h.host_is_superhost = 'True'
and comments like  '%good%' or comments like
'%excellent%' or comments like '%outstanding%';


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


--athens

 select h.host_id,
h.host_is_superhost,l.accommodates,l.bedrooms,count(l.property_type) as
cnt_propertytype,
 case
 when l.accommodates>7 then 'large'
 else 'small'
 end as property_size
 from  host_athens_df h join listing_athens_df l
 on h.host_id = l.host_id
 group by  h.host_id, h.host_is_superhost,l.accommodates,l.bedrooms


 --thessaloniki

 select h.host_id,
h.host_is_superhost,l.accommodates,l.bedrooms,count(l.property_type) as
cnt_propertytype,
 case
 when l.accommodates>7 then 'large'
 else 'small'
 end as property_size
 from  host_thessaloniki_df h join listing_thessaloniki_df l
 on h.host_id = l.host_id
```

```
 group by  h.host_id, h.host_is_superhost,l.accommodates,l.bedrooms



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


--athens

select  DATENAME(year,a.date) as years,available,host_is_superhost,avg(a.price) as
avg_price
from  df_athens_availability as a  join listing_athens_df as l on a.id=l.id join
host_athens_df h
on l.host_id=h.host_id  group by DATENAME(year,a.date) ,available,host_is_superhost
having DATENAME(year,a.date) = '2022'

--thessaloniki

select  DATENAME(year,a.date) as years,available,host_is_superhost,avg(a.price) as
avg_price
from  df_thessaloniki_availability as a  join listing_thessaloniki_df as l on
a.id=l.id join host_thessaloniki_df h
on l.host_id=h.host_id  group by DATENAME(year,a.date) ,available,host_is_superhost
having DATENAME(year,a.date) = '2022'



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

--athens

select avg(avg_price) as avg_price_local_host, avg(avg_acceptance) as
local_avg_acceptance,avg(avg_response) as local_avg_response
from(select H.host_location, datepart(year, A.date) upcoming_year, avg(A.price)
avg_price,
avg(H.host_acceptance_rate) avg_acceptance, avg(host_response_rate) avg_response
from host_athens_df h join listing_athens_df L on H.host_id = L.host_id
join  df_athens_availability A on L.id = A.id
where H.host_location like '%athens%'
group by H.host_location, datepart(year, A.date) )a

select avg(avg_price) as avg_price_other_location,
avg(avg_acceptance) as other_location_avg_acceptance,avg(avg_response) as
other_location_avg_response
from(select H.host_location, datepart(year, A.date) upcoming_year, avg(A.price)
avg_price,
avg(H.host_acceptance_rate) avg_acceptance, avg(host_response_rate) avg_response
from host_athens_df H join listing_athens_df L on H.host_id = L.host_id
join df_athens_availability A on L.id = A.id
where H.host_location not like '%athens%'
group by H.host_location, datepart(year, A.date) )a

--thessaloniki

select avg(avg_price) as avg_price_local_host, avg(avg_acceptance) as
local_avg_acceptance,avg(avg_response) as local_avg_response
from(select H.host_location, datepart(year, A.date) upcoming_year, avg(A.price)
```

```
avg_price,
avg(H.host_acceptance_rate) avg_acceptance, avg(host_response_rate) avg_response
from host_athens_df h join listing_athens_df L on H.host_id = L.host_id
join  df_athens_availability A on L.id = A.id
where H.host_location like '%athens%'
group by H.host_location, datepart(year, A.date) )a

select avg(avg_price) as avg_price_other_location,
avg(avg_acceptance) as other_location_avg_acceptance,avg(avg_response) as
other_location_avg_response
from(select H.host_location, datepart(year, A.date) upcoming_year, avg(A.price)
avg_price,
avg(H.host_acceptance_rate) avg_acceptance, avg(host_response_rate) avg_response
from host_athens_df H join listing_athens_df L on H.host_id = L.host_id
join df_athens_availability A on L.id = A.id
where H.host_location not like '%athens%'
group by H.host_location, datepart(year, A.date) )a


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


select host_is_superhost,avg(host_response_rate)
avg_host_response,avg(host_acceptance_rate) as avg_host_acceptance_rate,
avg(review_scores_value) as avg_review_scores_values from host_athens_df as h join
listing_athens_df as l
on h.host_id=l.host_id
where host_is_superhost = 'true' or host_is_superhost='false'
group by host_is_superhost;
```