HOST BEHAVIOUR ANALYSIS PROJECT

SQL QUERIES

**VENICE (ITALY)**

select \* from [venice\_availability (1)]

select \* from host\_venice\_df

select \* from listing\_venice\_df

select \* from review\_venice\_df

--count of host & superhost

select host\_is\_superhost,COUNT(host\_is\_superhost)

from host\_venice\_df

where host\_is\_superhost is not null

group by host\_is\_superhost

-- avg. acceptance rate & avg response rate

select

host\_is\_superhost,COUNT(host\_is\_superhost),

avg(host\_acceptance\_rate) as avg\_acceptance\_rate ,

avg(host\_response\_rate) as avg\_response\_rate

from host\_venice\_df

where host\_is\_superhost is not null

group by host\_is\_superhost

------ response time

select host\_is\_superhost,host\_response\_time,count(host\_is\_superhost)

from host\_venice\_df

where host\_is\_superhost is not null and host\_response\_time is not null

group by host\_is\_superhost,host\_response\_time

order by host\_response\_time desc

--instant booking

select b.host\_is\_superhost , a.instant\_bookable ,

COUNT(a.instant\_bookable)

from listing\_venice\_df as a join host\_venice\_df as b

on a.host\_id = b.host\_id

where b.host\_is\_superhost is not null

group by b.host\_is\_superhost,a.instant\_bookable

order by b.host\_is\_superhost

--profile picture

select b.host\_is\_superhost , b.host\_has\_profile\_pic

, count(b.host\_has\_profile\_pic)

from host\_venice\_df as b

where b.host\_is\_superhost is not null

group by b.host\_is\_superhost,b.host\_has\_profile\_pic

order by b.host\_is\_superhost

--identity verified

select b.host\_is\_superhost ,

b.host\_identity\_verified , count(\*)

from host\_venice\_df as b

where b.host\_is\_superhost is not null

group by b.host\_is\_superhost,b.host\_identity\_verified

order by b.host\_is\_superhost

--review scores

select host\_venice\_df.host\_is\_superhost ,

avg(listing\_venice\_df.review\_scores\_rating) as avg\_review\_scores\_of\_rating ,

avg(listing\_venice\_df.review\_scores\_accuracy) as

avg\_review\_scores\_of\_listing\_description\_accuracy,

avg(listing\_venice\_df.review\_scores\_cleanliness) as

avg\_review\_scores\_of\_cleanliness,

avg(listing\_venice\_df.review\_scores\_checkin) as

avg\_review\_scores\_of\_check\_in,

avg(listing\_venice\_df.review\_scores\_communication) as

avg\_review\_scores\_of\_host\_communication,

avg(listing\_venice\_df.review\_scores\_location) as

avg\_review\_scores\_of\_listing\_location,

avg(listing\_venice\_df.review\_scores\_value) as

avg\_review\_scores\_of\_value\_for\_money

from host\_venice\_df join listing\_venice\_df

on listing\_venice\_df.host\_id = host\_venice\_df.host\_id

where host\_venice\_df.host\_is\_superhost is not null

group by host\_venice\_df.host\_is\_superhost

-------- --- cancellation rate

select host\_is\_superhost , avg(100-host\_acceptance\_rate)as cancellation\_rate

from host\_venice\_df

where (100-host\_acceptance\_rate) is not null

group by host\_is\_superhost

-------- --- amount of positive comments given

select a.host\_is\_superhost,count(c.comments) from host\_venice\_df as a inner join listing\_venice\_df as b

on a.host\_id=b.host\_id

inner join review\_venice\_df as c on b.id=c.listing\_id

where (c.comments like '%fantastic%' or c.comments like'%great%' or c.comments like'%excellent%' or c.comments like '%wonderful%' or c.comments like'%lovely%'

or c.comments like'%nice%' or c.comments like'%warm%')

group by a.host\_is\_superhost

order by a.host\_is\_superhost;

**THESSALONIKI (GREECE)**

select \* from df\_thessaloniki\_availability

select \* from host\_thessaloniki\_df

select \* from listing\_thessaloniki\_df

select \* from review\_thessaloniki\_df

--count of host & superhost

select host\_is\_superhost,COUNT(host\_is\_superhost)

from host\_thessaloniki\_df

where host\_is\_superhost is not null

group by host\_is\_superhost

-- avg. acceptance rate & avg response rate

select

host\_is\_superhost,COUNT(host\_is\_superhost),

avg(host\_acceptance\_rate) as avg\_acceptance\_rate ,

avg(host\_response\_rate) as avg\_response\_rate

from host\_thessaloniki\_df

where host\_is\_superhost is not null

group by host\_is\_superhost

------ response time

select host\_is\_superhost,host\_response\_time,count(host\_is\_superhost)

from host\_thessaloniki\_df

where host\_is\_superhost is not null and host\_response\_time is not null

group by host\_is\_superhost,host\_response\_time

order by host\_response\_time desc

--instant booking

select b.host\_is\_superhost , a.instant\_bookable ,

COUNT(a.instant\_bookable)

from listing\_thessaloniki\_df as a join host\_thessaloniki\_df as b

on a.host\_id = b.host\_id

where b.host\_is\_superhost is not null

group by b.host\_is\_superhost,a.instant\_bookable

order by b.host\_is\_superhost

--profile picture

select b.host\_is\_superhost , b.host\_has\_profile\_pic

, count(b.host\_has\_profile\_pic)

from host\_thessaloniki\_df as b

where b.host\_is\_superhost is not null

group by b.host\_is\_superhost,b.host\_has\_profile\_pic

order by b.host\_is\_superhost

--identity verified

select b.host\_is\_superhost ,

b.host\_identity\_verified , count(\*)

from host\_thessaloniki\_df as b

where b.host\_is\_superhost is not null

group by b.host\_is\_superhost,b.host\_identity\_verified

order by b.host\_is\_superhost

--review scores

select host\_thessaloniki\_df.host\_is\_superhost ,

avg(listing\_thessaloniki\_df.review\_scores\_rating) as avg\_review\_scores\_of\_rating ,

avg(listing\_thessaloniki\_df.review\_scores\_accuracy) as

avg\_review\_scores\_of\_listing\_description\_accuracy,

avg(listing\_thessaloniki\_df.review\_scores\_cleanliness) as

avg\_review\_scores\_of\_cleanliness,

avg(listing\_thessaloniki\_df.review\_scores\_checkin) as

avg\_review\_scores\_of\_check\_in,

avg(listing\_thessaloniki\_df.review\_scores\_communication) as

avg\_review\_scores\_of\_host\_communication,

avg(listing\_thessaloniki\_df.review\_scores\_location) as

avg\_review\_scores\_of\_listing\_location,

avg(listing\_thessaloniki\_df.review\_scores\_value) as

avg\_review\_scores\_of\_value\_for\_money

from host\_thessaloniki\_df join listing\_thessaloniki\_df

on listing\_thessaloniki\_df.host\_id = host\_thessaloniki\_df.host\_id

where host\_thessaloniki\_df.host\_is\_superhost is not null

group by host\_thessaloniki\_df.host\_is\_superhost

----cancellation rate

select host\_is\_superhost , avg(100-host\_acceptance\_rate)as cancellation\_rate

from host\_thessaloniki\_df

where (100-host\_acceptance\_rate) is not null

group by host\_is\_superhost

----amount of positive comments

select a.host\_is\_superhost,count(c.comments) from host\_thessaloniki\_df as a inner join listing\_thessaloniki\_df as b

on a.host\_id=b.host\_id

inner join review\_thessaloniki\_df as c on b.id=c.listing\_id

where

(c.comments like '%fantastic%' or c.comments like'%great%' or c.comments like'%excellent%' or c.comments like '%wonderful%' or c.comments like'%lovely%'

or c.comments like'%nice%' or c.comments like'%warm%')

group by a.host\_is\_superhost

order by a.host\_is\_superhost;

**DALLAS(USA)**

SELECT \* FROM host\_dallas\_df;

select \* FROM listing\_dallas\_df;

select distinct(host\_is\_superhost) from host\_dallas\_df;

--count of host & superhost

select host\_dallas\_df.host\_is\_superhost,COUNT(host\_dallas\_df.host\_is\_superhost) as total\_host

from host\_dallas\_df

group by host\_dallas\_df.host\_is\_superhost;

-- avg. acceptance rate & avg response rate

select host\_dallas\_df.host\_is\_superhost,

COUNT(host\_dallas\_df.host\_is\_superhost),

avg(host\_dallas\_df.host\_acceptance\_rate) as avg\_acceptance\_rate,

avg(host\_dallas\_df.host\_response\_rate) as avg\_response\_rate

from host\_dallas\_df

group by host\_dallas\_df.host\_is\_superhost

--instant booking

select host\_dallas\_df.host\_is\_superhost , listing\_dallas\_df.instant\_bookable,COUNT(listing\_dallas\_df.instant\_bookable)

from listing\_dallas\_df join host\_dallas\_df

on listing\_dallas\_df.host\_id = host\_dallas\_df.host\_id

group by host\_dallas\_df.host\_is\_superhost,listing\_dallas\_df.instant\_bookable

order by host\_dallas\_df.host\_is\_superhost

--profile picture

select host\_dallas\_df.host\_is\_superhost , host\_dallas\_df.host\_has\_profile\_pic , count(host\_dallas\_df.host\_has\_profile\_pic)

from host\_dallas\_df

group by host\_dallas\_df.host\_is\_superhost,host\_dallas\_df.host\_has\_profile\_pic

order by host\_dallas\_df.host\_is\_superhost

--identity verified

select host\_dallas\_df.host\_is\_superhost , host\_dallas\_df.host\_identity\_verified , count(\*)

from host\_dallas\_df

group by host\_dallas\_df.host\_is\_superhost,host\_dallas\_df.host\_identity\_verified

order by host\_dallas\_df.host\_is\_superhost

--review scores

select host\_dallas\_df.host\_is\_superhost , avg(listing\_dallas\_df.review\_scores\_rating) as avg\_review\_scores\_of\_rating ,

avg(listing\_dallas\_df.review\_scores\_accuracy) as avg\_review\_scores\_of\_listing\_description\_accuracy,

avg(listing\_dallas\_df.review\_scores\_cleanliness) as avg\_review\_scores\_of\_cleanliness,

avg(listing\_dallas\_df.review\_scores\_checkin) as avg\_review\_scores\_of\_check\_in,

avg(listing\_dallas\_df.review\_scores\_communication) as avg\_review\_scores\_of\_host\_communication,

avg(listing\_dallas\_df.review\_scores\_location) as avg\_review\_scores\_of\_listing\_location,

avg(listing\_dallas\_df.review\_scores\_value) as avg\_review\_scores\_of\_value\_for\_money

from host\_dallas\_df join listing\_dallas\_df

on listing\_dallas\_df.host\_id = host\_dallas\_df.host\_id

group by host\_dallas\_df.host\_is\_superhost

----cancellation rate

select host\_is\_superhost , avg(100-host\_acceptance\_rate)as cancellation\_rate

from host\_dallas\_df

where (100-host\_acceptance\_rate) is not null

group by host\_is\_superhost

---good review count

select a.host\_is\_superhost, COUNT(c.comments) as good\_comment\_count

from host\_dallas\_df as a inner JOIN listing\_dallas\_df as bedrooms

on a.host\_id = b.host\_id

INNER join review\_dallas\_df as c

on b.id = c.listing\_id

where(c.comments like '%fantastic%' or c.comments like '%great%' or c.comments like '%excellent%' or c.comments like '%wonderful%' or c.comments like '%lovely%' or c.comments like '%good%' or c.comments like ‘%pleasant%’ or c.comments like ‘%comfortable%’)

group by a.host\_is\_superhost

order by a.host\_is\_superhost;

**BEIJING(CHINA)**

select \* from host\_beijing\_df

select \* from listing\_beijing\_df

select \* from [dbo].[df\_beijing\_availability]

select \* from [dbo].[review\_beijing\_df]

select distinct(host\_is\_superhost) from project2.dbo.host\_beijing\_df

--count of host & superhost

select host\_beijing\_df.host\_is\_superhost,COUNT(host\_beijing\_df.host\_is\_superhost)

from project2.dbo.host\_beijing\_df

group by host\_beijing\_df .host\_is\_superhost

-- avg. acceptance rate & avg response rate

select

host\_beijing\_df.host\_is\_superhost,COUNT(host\_beijing\_df.host\_is\_superhost),

avg(host\_beijing\_df.host\_acceptance\_rate) as avg\_acceptance\_rate ,

avg(host\_beijing\_df.host\_response\_rate) as avg\_response\_rate

from host\_beijing\_df

group by host\_beijing\_df .host\_is\_superhost

--instant booking

select host\_beijing\_df.host\_is\_superhost ,

listing\_beijing\_df.instant\_bookable,COUNT(listing\_beijing\_df.instant\_bookable)

from listing\_beijing\_df join host\_beijing\_df

on listing\_beijing\_df.host\_id = host\_beijing\_df.host\_id

group by host\_beijing\_df.host\_is\_superhost,listing\_beijing\_df.instant\_bookable

order by host\_beijing\_df.host\_is\_superhost

--profile picture

select host\_beijing\_df.host\_is\_superhost , host\_beijing\_df.host\_has\_profile\_pic

, count(host\_beijing\_df.host\_has\_profile\_pic)

from host\_beijing\_df

group by host\_beijing\_df.host\_is\_superhost,host\_beijing\_df.host\_has\_profile\_pic

order by host\_beijing\_df.host\_is\_superhost

--identity verified

select host\_beijing\_df.host\_is\_superhost ,

host\_beijing\_df.host\_identity\_verified , count(\*)

from host\_beijing\_df

group by host\_beijing\_df.host\_is\_superhost,host\_beijing\_df.host\_identity\_verified

order by host\_beijing\_df.host\_is\_superhost

--review scores

select host\_beijing\_df.host\_is\_superhost ,

avg(listing\_beijing\_df.review\_scores\_rating) as avg\_review\_scores\_of\_rating ,

avg(listing\_beijing\_df.review\_scores\_accuracy) as

avg\_review\_scores\_of\_listing\_description\_accuracy,

avg(listing\_beijing\_df.review\_scores\_cleanliness) as

avg\_review\_scores\_of\_cleanliness,

avg(listing\_beijing\_df.review\_scores\_checkin) as

avg\_review\_scores\_of\_check\_in,

avg(listing\_beijing\_df.review\_scores\_communication) as

avg\_review\_scores\_of\_host\_communication,

avg(listing\_beijing\_df.review\_scores\_location) as

avg\_review\_scores\_of\_listing\_location,

avg(listing\_beijing\_df.review\_scores\_value) as

avg\_review\_scores\_of\_value\_for\_money

from host\_beijing\_df join listing\_beijing\_df

on listing\_beijing\_df.host\_id = host\_beijing\_df.host\_id

group by host\_beijing\_df.host\_is\_superhost

----cancellation rate

select host\_is\_superhost , avg(100-host\_acceptance\_rate)as cancellation\_rate

from host\_beijing\_df

where (100-host\_acceptance\_rate) is not null

group by host\_is\_superhost

---good review count

select a.host\_is\_superhost, COUNT(c.comments) as good\_comment\_count

from host\_beijing \_df as a inner JOIN listing\_dallas\_df as bedrooms

on a.host\_id = b.host\_id

INNER join review\_beijing \_df as c

on b.id = c.listing\_id

where(c.comments like '%fantastic%' or c.comments like '%great%' or c.comments like '%excellent%' or c.comments like '%wonderful%' or c.comments like '%lovely%' or c.comments like '%good%' or c.comments like ‘%pleasant%’ or c.comments like ‘%comfortable%’)

group by a.host\_is\_superhost

order by a.host\_is\_superhost;

**SHANGHAI(CHINA)**

select \* from host\_shanghai\_df

select \* from listing\_shanghai\_df

select \* from [dbo].[df\_shanghai\_availability]

select \* from [dbo].[review\_shanghai\_df]

select distinct(host\_is\_superhost) from project2.dbo.host\_shanghai\_df

--count of host & superhost

select host\_shanghai\_df.host\_is\_superhost,COUNT(host\_shanghai\_df.host\_is\_superhost)

from project2.dbo.host\_shanghai\_df

group by host\_shanghai\_df .host\_is\_superhost

-- avg. acceptance rate & avg response rate

select

host\_shanghai\_df.host\_is\_superhost,COUNT(host\_shanghai\_df.host\_is\_superhost),

avg(host\_shanghai\_df.host\_acceptance\_rate) as avg\_acceptance\_rate ,

avg(host\_shanghai\_df.host\_response\_rate) as avg\_response\_rate

from host\_shanghai\_df

group by host\_shanghai\_df .host\_is\_superhost

--instant booking

select host\_shanghai\_df.host\_is\_superhost ,

listing\_shanghai\_df.instant\_bookable,COUNT(listing\_shanghai\_df.instant\_bookable)

from listing\_shanghai\_df join host\_shanghai\_df

on listing\_shanghai\_df.host\_id = host\_shanghai\_df.host\_id

group by host\_shanghai\_df.host\_is\_superhost,listing\_shanghai\_df.instant\_bookable

order by host\_shanghai\_df.host\_is\_superhost

--profile picture

select host\_shanghai\_df.host\_is\_superhost , host\_shanghai\_df.host\_has\_profile\_pic

, count(host\_shanghai\_df.host\_has\_profile\_pic)

from host\_shanghai\_df

group by host\_shanghai\_df.host\_is\_superhost,host\_shanghai\_df.host\_has\_profile\_pic

order by host\_shanghai\_df.host\_is\_superhost

--identity verified

select host\_shanghai\_df.host\_is\_superhost ,

host\_shanghai\_df.host\_identity\_verified , count(\*)

from host\_shanghai\_df

group by host\_beijing\_df.host\_is\_superhost,host\_shanghai\_df.host\_identity\_verified

order by host\_shanghai\_df.host\_is\_superhost

--review scores

select host\_shanghai\_df.host\_is\_superhost ,

avg(listing\_shanghai\_df.review\_scores\_rating) as avg\_review\_scores\_of\_rating ,

avg(listing\_shanghai\_df.review\_scores\_accuracy) as

avg\_review\_scores\_of\_listing\_description\_accuracy,

avg(listing\_shanghai\_df.review\_scores\_cleanliness) as

avg\_review\_scores\_of\_cleanliness,

avg(listing\_shanghai\_df.review\_scores\_checkin) as

avg\_review\_scores\_of\_check\_in,

avg(listing\_shanghai\_df.review\_scores\_communication) as

avg\_review\_scores\_of\_host\_communication,

avg(listing\_shanghai\_df.review\_scores\_location) as

avg\_review\_scores\_of\_listing\_location,

avg(listing\_shanghai\_df.review\_scores\_value) as

avg\_review\_scores\_of\_value\_for\_money

from host\_shanghai\_df join listing\_shanghai\_df

on listing\_shanghai\_df.host\_id = host\_beijing\_df.host\_id

group by host\_shanghai\_df.host\_is\_superhost

----cancellation rate

select host\_is\_superhost , avg(100-host\_acceptance\_rate)as cancellation\_rate

from host\_shanghai\_df

where (100-host\_acceptance\_rate) is not null

group by host\_is\_superhost

---good review count

select a.host\_is\_superhost, COUNT(c.comments) as good\_comment\_count

from host\_shanghai\_df as a inner JOIN listing\_shanghai\_df as bedrooms

on a.host\_id = b.host\_id

INNER join review\_shanghai\_df as c

on b.id = c.listing\_id

where(c.comments like '%fantastic%' or c.comments like '%great%' or c.comments like '%excellent%' or c.comments like '%wonderful%' or c.comments like '%lovely%' or c.comments like '%good%' or c.comments like ‘%pleasant%’ or c.comments like ‘%comfortable%’)

group by a.host\_is\_superhost

order by a.host\_is\_superhost;

**VANCOUVER(CANADA)**

select \* from host\_vancouver\_df

select \* from listing\_vancouver\_df

select \* from df\_vancouver\_availability

select \* from review\_vancouver\_df

select distinct(host\_is\_superhost) from host\_vancouver\_df

--count of host & superhost

select a.host\_is\_superhost,COUNT(\*)

from host\_vancouver\_df as a

where a.host\_is\_superhost is not null

group by a.host\_is\_superhost

-- avg. acceptance rate & avg response rate

select host\_is\_superhost ,COUNT(\*) as count\_, avg(host\_acceptance\_rate) as avg\_acceptance\_rate , avg(host\_response\_rate) as avg\_response\_rate

from host\_vancouver\_df

where host\_vancouver\_df.host\_is\_superhost is not null

group by host\_is\_superhost

--instant booking

select host\_vancouver\_df.host\_is\_superhost , listing\_vancouver\_df.instant\_bookable,COUNT(listing\_vancouver\_df.instant\_bookable)

from listing\_vancouver\_df join host\_vancouver\_df

on listing\_vancouver\_df.host\_id = host\_vancouver\_df.host\_id

where host\_vancouver\_df.host\_is\_superhost is not null

group by host\_vancouver\_df.host\_is\_superhost,listing\_vancouver\_df.instant\_bookable

order by host\_vancouver\_df.host\_is\_superhost

--profile picture

select host\_vancouver\_df.host\_is\_superhost , host\_vancouver\_df.host\_has\_profile\_pic , count(host\_vancouver\_df.host\_has\_profile\_pic)

from host\_vancouver\_df

where host\_vancouver\_df.host\_is\_superhost is not null

group by host\_vancouver\_df.host\_is\_superhost,host\_vancouver\_df.host\_has\_profile\_pic

order by host\_vancouver\_df.host\_is\_superhost

--identity verified

select host\_vancouver\_df.host\_is\_superhost , host\_vancouver\_df.host\_identity\_verified , count(\*)

from host\_vancouver\_df

where host\_vancouver\_df.host\_is\_superhost is not null

group by host\_vancouver\_df.host\_is\_superhost,host\_vancouver\_df.host\_identity\_verified

order by host\_vancouver\_df.host\_is\_superhost

--review scores

select host\_vancouver\_df.host\_is\_superhost , avg(listing\_vancouver\_df.review\_scores\_rating) as avg\_review\_scores\_of\_rating ,

avg(listing\_vancouver\_df.review\_scores\_accuracy) as avg\_review\_scores\_of\_listing\_description\_accuracy,

avg(listing\_vancouver\_df.review\_scores\_cleanliness) as avg\_review\_scores\_of\_cleanliness,

avg(listing\_vancouver\_df.review\_scores\_checkin) as avg\_review\_scores\_of\_check\_in,

avg(listing\_vancouver\_df.review\_scores\_communication) as avg\_review\_scores\_of\_host\_communication,

avg(listing\_vancouver\_df.review\_scores\_location) as avg\_review\_scores\_of\_listing\_location,

avg(listing\_vancouver\_df.review\_scores\_value) as avg\_review\_scores\_of\_value\_for\_money

from host\_vancouver\_df join listing\_vancouver\_df

on listing\_vancouver\_df.host\_id = host\_vancouver\_df.host\_id

where host\_vancouver\_df.host\_is\_superhost is not null

group by host\_vancouver\_df.host\_is\_superhost

--host listing count

select host\_id ,host\_listings\_count , host\_is\_superhost

from host\_vancouver\_df

where host\_is\_superhost is not null

group by host\_listings\_count,host\_is\_superhost,host\_id

order by host\_listings\_count desc

-------------------------------------------------

--Analyze how the comments of reviewers vary for listings of Super Hosts vs Other Hosts

select c.host\_is\_superhost , count(a.comments)

from review\_vancouver\_df as a join listing\_vancouver\_df as b ON a.listing\_id = b.id

join host\_vancouver\_df as c ON c.host\_id = b.host\_id

where(a.comments like '%great%' or a.comments like '%nice%' or a.comments like '%wonderful%' or a.comments like '%brilliant%'or

a.comments like '%great location%' or a.comments like '%good%' or a.comments like '%lovely%' or a.comments like '%friendly%'or

a.comments like '%perfect%' or a.comments like '%beautiful%' or a.comments like '%definetly stay%' or a.comments like '%excellent%'

or a.comments like '%highly recommended%')

AND c.host\_is\_superhost is not null

group by c.host\_is\_superhost

--- cancellation rate

select host\_is\_superhost , avg(100-host\_acceptance\_rate)as cancellation\_rate

from host\_vancouver\_df

where (100-host\_acceptance\_rate) is not null

group by host\_is\_superhost

------ response time

select host\_is\_superhost,host\_response\_time,count(host\_is\_superhost)

from host\_vancouver\_df

where host\_is\_superhost is not null and host\_response\_time is not null

group by host\_is\_superhost,host\_response\_time

order by host\_response\_time desc