HOST BEHAVIOURAL ANALYSIS PROJECT – QUERIES

-----(QUESTION 1) YEARWISE OVERALL REVIEW ANALYSIS------

CREATE VIEW Overall_Dallas_analysis AS SELECT Year,[0] AS Host,[1] AS Superhost FROM (SELECT

A.host_id,A.host_name,AVG(((B.review_scores_accuracy+B.review_scores_checkin+B.review_scores_cleanliness+B.review_scores_communication+B.review_scores_location+B.review_scores_rating+B.review_scores_value)/7)) AS AVG_OVERALL_RATING,

A.host_is_superhost, YEAR(C.date) AS Year

FROM host_dallas_df AS A

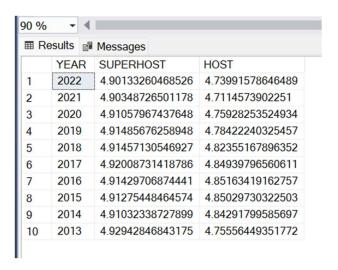
LEFT JOIN listing_dallas_df AS B ON A.host_id = B.host_id

LEFT JOIN review_dallas_df AS C ON B.id = C.listing_id

GROUP BY A.host_id, A.host_name, A.host_is_superhost, YEAR(C.date))C

PIVOT (AVG(AVG_OVERALL_RATING) FOR host_is_superhost IN ([0],[1])) AS pivot_table

SELECT YEAR,AVG(superhost) AS SUPERHOST,AVG(host) AS HOST FROM Overall_Dallas_analysis WHERE Year BETWEEN 2013 and 2022 GROUP BY YEAR ORDER BY Year DESC;



CREATE VIEW Overall_austin_analysis AS SELECT Year,[0] AS Host,[1] AS Superhost FROM (SELECT

A.host_id,A.host_name,AVG(((B.review_scores_accuracy+B.review_scores_checkin+B.review_scores_cleanliness+B.review_scores_communication+B.review_scores_location+B.review_scores_rating+B.review_scores_value)/7)) AS AVG_OVERALL_RATING,

A.host_is_superhost, YEAR(C.date) AS Year

FROM host austin df AS A

LEFT JOIN listing_austin_df AS B ON A.host_id = B.host_id

LEFT JOIN review_austin_df AS C ON B.id = C.listing_id

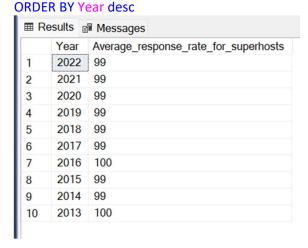
GROUP BY A.host_id, A.host_name, A.host_is_superhost, YEAR(C.date))C

PIVOT (AVG(AVG_OVERALL_RATING) FOR host_is_superhost IN ([0],[1])) AS pivot_table

SELECT YEAR,AVG(superhost) AS SUPERHOST,AVG(host) AS HOST FROM Overall_austin_analysis WHERE Year BETWEEN 2013 and 2022 GROUP BY YEAR ORDER BY Year DESC;

■ Results ■ Messages						
	YEAR	SUPERHOST	HOST			
1	2022	4.91615805796588	4.82438299094556			
2	2021	4.91723938090213	4.79109441300827			
3	2020	4.92017801403674	4.81477454324701			
4	2019	4.91964670123516	4.82237924969574			
5	2018	4.91909910832377	4.81661570434166			
6	2017	4.91537416838978	4.82820147914074			
7	2016	4.91156901040426	4.83085424740556			
8	2015	4.90511384202068	4.79572850637444			
9	2014	4.90213188613864	4.81382604099605			
10	2013	4.90511060089069	4.81343420951895			

SELECT YEAR(C.date) AS Year,ROUND(AVG(A.host_response_rate),0) AS Average_response_rate_for_superhosts FROM host_dallas_df AS A LEFT JOIN listing_dallas_df AS B ON A.host_id = B.host_id LEFT JOIN review_dallas_df AS C ON B.id = C.listing_id WHERE host_is_superhost = 1 AND YEAR(C.date) BETWEEN 2013 and 2022 GROUP BY YEAR(C.date)



SELECT YEAR(C.date) AS Year, ROUND(AVG(A.host_response_rate),0) AS Average_response_rate_for_hosts FROM host_dallas_df AS A LEFT JOIN listing_dallas_df AS B ON A.host_id = B.host_id LEFT JOIN review_dallas_df AS C ON B.id = C.listing_id WHERE host_is_superhost = 0 AND YEAR(C.date) BETWEEN 2013 and 2022 GROUP BY YEAR(C.date)

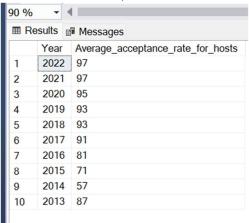


SELECT YEAR(C.date) AS Year,ROUND(AVG(A.host_acceptance_rate),0) AS Average_acceptance_rate_for_superhosts FROM host_dallas_df AS A LEFT JOIN listing_dallas_df AS B ON A.host_id = B.host_id LEFT JOIN review_dallas_df AS C ON B.id = C.listing_id WHERE host_is_superhost = 1 AND YEAR(C.date) BETWEEN 2013 and 2022 GROUP BY YEAR(C.date)

ORDER BY Year desc; Year Average_acceptance_rate_for_superhosts 2022 98 2 2021 97 3 2020 98 2019 97 5 2018 97 2017 96 7 2016 94 2015 91 9 2014 92 10 2013 94

SELECT YEAR(C.date) AS Year,ROUND(AVG(A.host_acceptance_rate),0) AS Average_acceptance_rate_for_hosts FROM host_dallas_df AS A LEFT JOIN listing_dallas_df AS B ON A.host_id = B.host_id LEFT JOIN review_dallas_df AS C ON B.id = C.listing_id WHERE host_is_superhost = 0 AND YEAR(C.date) BETWEEN 2013 and 2022 GROUP BY YEAR(C.date)

ORDER BY Year desc;



------Austin, Texas------

SELECT YEAR(C.date) AS Year,ROUND(AVG(A.host_response_rate),0) AS

Average_response_rate_for_superhosts FROM host_austin_df AS A

LEFT JOIN listing_austin_df AS B ON A.host_id = B.host_id

LEFT JOIN review_austin_df AS C ON B.id = C.listing_id

WHERE host_is_superhost = 1 AND YEAR(C.date) BETWEEN 2013 and 2022 GROUP BY YEAR(C.date)

ORDER BY Year desc

	Year	Average_response_rate_for_superhosts		
1	2022	99		
2	2021	99		
3	2020	99		
4	2019	99		
5	2018	99		
6	2017	99		
7	2016	99		
8	2015	99		
9	2014	99		
10	2013	100		

SELECT YEAR(C.date) AS Year,ROUND(AVG(A.host_response_rate),0) AS Average_response_rate_for_hosts FROM host_austin_df AS A LEFT JOIN listing_austin_df AS B ON A.host_id = B.host_id LEFT JOIN review_austin_df AS C ON B.id = C.listing_id WHERE host_is_superhost = 0 AND YEAR(C.date) BETWEEN 2013 and 2022 GROUP BY YEAR(C.date) ORDER BY Year desc;

SELECT YEAR(C.date) AS Year,ROUND(AVG(A.host_acceptance_rate),0) AS Average_acceptance_rate_for_superhosts FROM host_austin_df AS A LEFT JOIN listing_austin_df AS B ON A.host_id = B.host_id LEFT JOIN review_austin_df AS C ON B.id = C.listing_id WHERE host_is_superhost = 1 AND YEAR(C.date) BETWEEN 2013 and 2022 GROUP BY YEAR(C.date)

ORDER BY Year desc;

Year Average_acceptance_rate_for_superho			
1	2022	97	
2	2021	97	
3	2020	97	
4	2019	95	
5	2018	95	
6	2017	95	
7	2016	94	
8	2015	93	
9	2014	93	
10	2013	92	

SELECT YEAR(C.date) AS Year,ROUND(AVG(A.host_acceptance_rate),0) AS
Average_acceptance_rate_for_hosts FROM host_austin_df AS A
LEFT JOIN listing_austin_df AS B ON A.host_id = B.host_id
LEFT JOIN review_austin_df AS C ON B.id = C.listing_id
WHERE host_is_superhost = 0 AND YEAR(C.date) BETWEEN 2013 and 2022
GROUP BY YEAR(C.date)
ORDER BY Year desc;

	Year	Average_acceptance_rate_for_hosts
1	2022	96
2	2021	95
3	2020	94
4	2019	89
5	2018	88
6	2017	83
7	2016	79
8	2015	77
9	2014	74
10	2013	80

```
----(QUESTION 3) how does the comments of reviewers vary for listings of Super Hosts vs Other Hosts----
                           -----Dallas, Texas----
SELECT COUNT(*) AS count of good reviews for superhost FROM
(SELECT A.*, C.host is superhost FROM review dallas df AS A
LEFT JOIN listing dallas df AS B ON A.listing id = B.id
LEFT JOIN host dallas df AS C ON B.host id = C.host id
WHERE comments LIKE '%great host%' AND C.host_is_superhost = 1)C
      count_of_good_reviews_for_superhost
      4437
 1
SELECT COUNT(*) AS count_of_good_reviews_for_host FROM
(SELECT A.*, C.host_is_superhost FROM review_dallas_ df AS A
LEFT JOIN listing_dallas_df AS B ON A.listing_id = B.id
LEFT JOIN host dallas df AS C ON B.host id = C.host id
WHERE comments LIKE '%great host%' AND C.host_is_superhost = 0)C
      count of good reviews for host
      3427
SELECT COUNT(*) AS count of bad reviews for superhost FROM
(SELECT A.*, C.host is superhost FROM review dallas df AS A
LEFT JOIN listing_dallas_df AS B ON A.listing_id = B.id
LEFT JOIN host_dallas_df AS C ON B.host_id = C.host_id
WHERE comments LIKE '%poor%' AND C.host_is_superhost = 1)C
      count_of_bad_reviews_for_superhost
      52
 1
SELECT COUNT(*) AS count of bad reviews for host FROM
(SELECT A.*, C.host is superhost FROM review dallas df AS A
LEFT JOIN listing dallas df AS B ON A.listing id = B.id
LEFT JOIN host_dallas_df AS C ON B.host_id = C.host_id
WHERE comments LIKE '%poor%' AND C.host is superhost = 0)C
      count_of_bad_reviews_for_host
      216
 1
                                      -----Austin, Texas-
SELECT COUNT(*) AS count_of_good_reviews_for_superhost FROM
(SELECT A.*, C.host is superhost FROM review austin df AS A
LEFT JOIN listing austin df AS B ON A. listing id = B.id
LEFT JOIN host austin df AS C ON B.host id = C.host id
WHERE comments LIKE '%great host%' AND C.host is superhost = 1)C
     count_of_good_reviews_for_superhost
     15262
SELECT COUNT(*) AS count_of_good_reviews_for_host FROM
(SELECT A.*, C.host is superhost FROM review austin df AS A
LEFT JOIN listing_austin_df AS B ON A.listing_id = B.id
LEFT JOIN host austin df AS C ON B.host id = C.host id
WHERE comments LIKE '%great host%' AND C.host is superhost = 0)C
     count_of_good_reviews_for_host
     6243
```

```
(SELECT A.*, C.host is superhost FROM review austin df AS A
LEFT JOIN listing_austin_df AS B ON A.listing_id = B.id
LEFT JOIN host austin df AS C ON B.host id = C.host id
WHERE comments LIKE '%poor%' AND C.host is superhost = 1)C
       count_of_bad_reviews_for_superhost
SELECT COUNT(*) AS count_of_bad_reviews_for_host FROM
(SELECT A.*, C.host_is_superhost FROM review_austin_df AS A
LEFT JOIN listing austin df AS B ON A. listing id = B.id
LEFT JOIN host austin df AS C ON B.host id = C.host id
WHERE comments LIKE '%poor%' AND C.host is superhost = 0)C
      count_of_bad_reviews_for_host
      272
 1
-----(QUESTION 4) Do Super Hosts tend to have large property types as compared to Other Hosts -----
               -----Dallas, Texas-----
SELECT COUNT(property_type) TOTAL_large_properties FROM
(SELECT A.property_type FROM listing_dallas_df AS A
LEFT JOIN host_dallas_df AS B
ON A.host_id = B.host_id
WHERE host_is_superhost = 0)C;
      TOTAL_large_properties
      3909
 1
SELECT COUNT(*) TOTAL large property dealings by hosts FROM
(SELECT A.host_id FROM listing_dallas_df AS A
LEFT JOIN host dallas df AS B
ON A.host id = B.host id
WHERE room type = 'Entire home/apt' AND host is superhost = 0)C;
      TOTAL_large_property_dealings_by_hosts
      3346
SELECT COUNT(property_type) TOTAL_large_properties FROM
(SELECT A.property type FROM listing dallas df AS A
LEFT JOIN host_dallas_df AS B
ON A.host id = B.host id
WHERE host_is_superhost = 1)C;
      TOTAL_large_properties
 1
      1519
SELECT COUNT(*) TOTAL_large_property_dealings_by_Superhost FROM
(SELECT A.host id FROM listing dallas df AS A
LEFT JOIN host dallas df AS B
ON A.host id = B.host id
WHERE room_type = 'Entire home/apt' AND host_is_superhost = 1)C;
     TOTAL_large_property_dealings_by_Superhost
      1289
 1
```

SELECT COUNT(*) AS count_of_bad_reviews_for_superhost FROM

```
-----Austin, Texas-
SELECT COUNT(property_type) TOTAL_large_properties FROM
(SELECT A.property_type FROM listing_austin_df AS A
LEFT JOIN host austin df AS B
ON A.host id = B.host id
WHERE host is superhost = 0)C;
      TOTAL_large_properties
      8079
SELECT COUNT(*) TOTAL_large_property_dealings_by_hosts FROM
(SELECT A.host_id FROM listing_austin_df AS A
LEFT JOIN host austin df AS B
ON A.host id = B.host id
WHERE room type = 'Entire home/apt' AND host is superhost = 0)C;
     TOTAL_large_property_dealings_by_hosts
     6497
 1
SELECT COUNT(property_type) TOTAL_large_properties FROM
(SELECT A.property type FROM listing austin df AS A
LEFT JOIN host_austin_df AS B
ON A.host_id = B.host_id
WHERE host_is_superhost = 1)C;
     TOTAL_large_properties
     3890
SELECT COUNT(*) TOTAL large property dealings by Superhost FROM
(SELECT A.host id FROM listing austin df AS A
LEFT JOIN host_austin_df AS B
ON A.host id = B.host id
WHERE room_type = 'Entire home/apt' AND host_is_superhost = 1)C;
     TOTAL_large_property_dealings_by_Superhost
     3332
1
                         -----Small Property Dealings Analysis ------
                            -----Dallas, Texas-
SELECT COUNT(property_type) TOTAL_small_properties FROM
(SELECT A.property_type FROM listing_dallas_df AS A
LEFT JOIN host dallas df AS B
ON A.host_id = B.host_id
WHERE host is superhost = 0)C;
     TOTAL_small_properties
1
     3909
SELECT COUNT(*) TOTAL_small_property_dealings_by_superhosts FROM
(SELECT A.host_id FROM listing_dallas_df AS A
LEFT JOIN host_dallas_df AS B
ON A.host id = B.host id
WHERE room_type = 'private room' AND host_is_superhost = 0)C;
      TOTAL_small_property_dealings_by_superhosts
      445
```

```
(SELECT A.property type FROM listing dallas df AS A
LEFT JOIN host dallas df AS B
ON A.host id = B.host id
WHERE host_is_superhost = 1)C;
      TOTAL small properties
      1519
SELECT COUNT(*) TOTAL small property dealings by hosts FROM
(SELECT A.host_id FROM listing_dallas_df AS A
LEFT JOIN host_dallas_df AS B
ON A.host id = B.host id
WHERE room_type = 'private room' AND host_is_superhost = 1)C;
     TOTAL_small_property_dealings_by_hosts
 1
     227
                                  ----- Austin, Texas-
SELECT COUNT(property_type) TOTAL_small_properties FROM
(SELECT A.property_type FROM listing_austin_df AS A
LEFT JOIN host austin df AS B
ON A.host id = B.host id
WHERE host_is_superhost = 0)C;
      TOTAL small properties
      8079
SELECT COUNT(*) TOTAL_small_property_dealings_by_superhosts FROM
(SELECT A.host id FROM listing austin df AS A
LEFT JOIN host austin df AS B
ON A.host id = B.host id
WHERE room type = 'private room' AND host is superhost = 0)C;
      TOTAL small property dealings by superhosts
      1468
 1
SELECT COUNT(property type) TOTAL small properties FROM
(SELECT A.property type FROM listing austin df AS A
LEFT JOIN host_austin_df AS B
ON A.host_id = B.host_id
WHERE host_is_superhost = 1)C;
      TOTAL_small_properties
 1
      3890
SELECT COUNT(*) TOTAL small property dealings by hosts FROM
(SELECT A.host id FROM listing austin df AS A
LEFT JOIN host austin df AS B
ON A.host id = B.host id
WHERE room_type = 'private room' AND host_is_superhost = 1)C;
      TOTAL_small_property_dealings_by_hosts
       545
```

SELECT COUNT(property_type) TOTAL_small_properties FROM

(Question 5)Analyze the average price and availability of the listings for the upcoming year between

------Austin, Texas------

select year(B.date)as year,B.available,C.host_is_superhost,avg(A.price) AS Average_price_earned from listing_austin_df as A

left join df_austin_availability as B on A.id=B.listing_id left join host_austin_df as C on A.host_id=C.host_id

where C.host_is_superhost=0 AND YEAR(B.date) IS NOT NULL

group by B.available,C.host_is_superhost,year(B.date)

ORDER BY available DESC

	year	available	host_is_superhost	Average_price_earned
1	2023	1	0	421.671797181844
2	2022	1	0	406.507639786514
3	2023	0	0	285.385993818383
4	2022	0	0	291.615667507704

select year(B.date)as year,B.available,C.host_is_superhost,avg(A.price) AS Average_price_earned from listing_austin_df as A

left join df_austin_availability as B on A.id=B.listing_id left join host_austin_df as C on

A.host id=C.host id

where C.host_is_superhost=1

group by B.available,C.host_is_superhost,year(B.date)

ORDER BY available DESC

	year	available	host_is_superhost	Average_price_earned
1	2023	1	1	414.779478645291
2	2022	1	1	405.672400063008
3	2023	0	1	323.058993503345
4	2022	0	1	320.323632125231

------Dallas, Texas------

select year(B.date)as year,B.available,C.host_is_superhost,avg(A.price) AS Average_price_earned from listing dallas df as A

 $left\ join\ df_dallas_availability\ as\ B\ on\ A.id=B.listing_id\ left\ join\ host_dallas_df\ as\ C\ on$

A.host id=C.host id

where C.host is superhost=0 AND YEAR(B.date) IS NOT NULL

group by B.available,C.host_is_superhost,year(B.date)

ORDER BY available DESC

	year	available	host_is_superhost	Average_price_earned
1	2022	1	0	162.400564648388
2	2023	1	0	155.735644577067
3	2023	0	0	161.161850715154
4	2022	0	0	150.117857849874

select year(B.date)as year,B.available,C.host_is_superhost,avg(A.price) AS Average_price_earned from listing_dallas_df as A

 $left\ join\ df_dallas_availability\ as\ B\ on\ A.id=B.listing_id\ left\ join\ host_dallas_df\ as\ C\ on$

A.host_id=C.host_id

where C.host_is_superhost=1

group by B.available,C.host_is_superhost,year(B.date)

ORDER BY available DESC

	year	available	host_is_superhost	Average_price_earned
1	2023	1	1	201.683156197755
2	2022	1	1	198.538887917031
3	2023	0	1	163.926437061986
4	2022	0	1	160.11026391369