DATA ANALYSIS

I made a list of initial questions I was interested in:

- **Which are the top five Greek neighborhoods with the highest average price per night?**

  As a first step of the analysis, we want to identify the areas in Greece with the highest number of Airbnb listings. There are 7335 distinct neighborhoods with Airbnb listings in Greece and the average price is $121.54.

```
SELECT DISTINCT COUNT(Fact_neighbourhood) as Count_neighbourhoods,
avg(PRICE_1) as Average_Neighbourhood_price
FROM FactListings
```

  And here are the top five listed neighborhoods with the highest average price:
  The coding process:

```
SELECT TOP 5 avg([price])as Average_price,
[fact_neighbourhood]
FROM FactListings
GROUP BY [fact_neighbourhood]
ORDER BY avg([price]) desc;
```

  The results: **Acropolis** is the neighborhood with the highest average price per listing with the average price being estimated at **415$.**

	Average_price	Fact_neighbourhood
1	415	Acropolis, Athens, Greece
2	411.5	Athina, plaka, Greece
3	321	Athina, 0, Greece
4	258.75	Koukaki, athens, Greece
5	250	????a ??f?? , ?tt????, Greece

- **What is the number of accommodates for the highly-priced neighborhood of Athens?**

  In order to find the most expensive neighborhood, I created a CTE expression that calculates the highest average price of listings in Athens. Then, I joined the FactListings table with the CTE expression to filter listings to only those in the most expensive neighborhood.

  The coding process:

```
WITH max_price as (SELECT TOP 1 fact_neighbourhood
FROM FactListings
WHERE fact_neighbourhood like '%Athens%'
GROUP BY fact_neighbourhood
ORDER BY avg([price]) desc)
SELECT FL.accommodates,
AVG(FL.[price]) as average_price
FROM FactListings as FL,max_price
WHERE FL.fact_neighbourhood = max_price.fact_neighbourhood
GROUP BY FL.accommodates
ORDER BY FL.accommodates;
```

  The results: A listing in the Acropolis neighborhood, which has the highest average price, accommodates 8 people, with the average listing price reaching $415.

accommodates	average_price
8	415

- **Which are the top five neighborhoods with the highest revenue potential based on listing data?**

  To calculate the revenue potential, we multiply the price of the listing by the maximum number of nights a traveler can stay. We do not take into account listings with N/A values in our data.

  The coding process:

```
SELECT TOP  5 fact_neighbourhood,
price* maximum_nights as monthly_revenue
FROM FactListings
WHERE fact_neighbourhood not like '%N/A%'
GROUP BY fact_neighbourhood,price * maximum_nights
ORDER BY Sum(price * (30-maximum_nights)) DESC ;
```

The results: We can notice from the following table that neighborhoods in Athens tend to exhibit a higher revenue potential reaching the number of 48490.

	fact_neighbourhood	monthly_revenue
1	Athina, Greece	48490
2	Athina, Greece	2800
3	Athina, Greece	300
4	Athina, Greece	2514
5	Athina, Greece	8456

- **Do super hosts or regular hosts have the most properties?**

The coding process:

```
SELECT CASE WHEN is_superhost = 1 then 'Superhost' when is_superhost=NULL then 'Unknown' else 'Regular Host' end as host_type,
COUNT(*) AS property_type_count
FROM DimHost
GROUP BY is_superhost;
```

The results: As we can see from the property count, regular hosts seem to own more properties than super hosts. This may be true for some hosts who are in the process of becoming superhosts or who are new to the platform.

	host_type	property_type_count
1	Regular Host	3640
2	Superhost	2134

- **What percentage of superhosts and non-superhosts have a profile picture?**

The coding process:

```
SELECT SUM(CASE WHEN has_profile_pic = 1 then 1 else 0 end)*100.00 / count(*) as percentage
FROM DimHost
WHERE is_superhost=1;

SELECT SUM(CASE WHEN has_profile_pic = 1 then 1 else 0 end)*100.00 / count(*) as percentage
FROM DimHost
WHERE is_superhost=0;
```

The results: 98% of superhosts on Airbnb have a profile picture, compared to 94.3% of non-superhosts, showing a 3.7% higher prevalence among superhosts.

	percentage_of_superhosts
1	98.3598875351452

	percentage_of_non_superhosts
1	94.2582417582417

- **What is the most common room type among superhosts and non-superhosts?**

The coding process:

In order to find the property count for each room type among superhosts and non-superhosts, I used the Inner Join command to join the DimRoomType table and the DimHost table to the FactListings table.

```
SELECT TOP 1  B.room_type,count(*)as room_type_count_superhosts
FROM FactListings as FL
INNER JOIN DimRoomType as B on FL.room_type_id= B.room_type_id
INNER JOIN DimHost as DM on FL.host_id=DM.host_id
WHERE DM.is_superhost=1
GROUP BY B.room_type
ORDER BY room_type_count_superhosts desc;

SELECT TOP 1  B.room_type,count(*)as room_type_count_non_superhosts
FROM FactListings as FL
INNER JOIN DimRoomType as B on FL.room_type_id= B.room_type_id
INNER JOIN DimHost as DM on FL.host_id=DM.host_id
WHERE DM.is_superhost=0
GROUP BY B.room_type
ORDER BY room_type_count_non_superhosts desc;
```

The results: The majority of superhosts and non-superhosts prefer to list entire apartments on the Airbnb platform, indicating it as the ideal accommodation choice.

	room_type	room_type_count_superhosts
1	Entire home/apt	3102

	room_type	room_type_count_non_superhosts
1	Entire home/apt	3418

- **What is the average number of beds and bathrooms in listings hosted by superhosts and non-superhosts?**

The coding process:

```
SELECT AVG(beds) as num_beds, AVG(bathrooms)as num_bathrooms_superhosts
FROM FactListings as FL
INNER JOIN DimHost as DM on FL.host_id= DM.host_id
WHERE DM.is_superhost=1;

SELECT AVG(beds) as num_beds, AVG(bathrooms)as num_bathrooms_non_superhosts
FROM FactListings as FL
INNER JOIN DimHost as DM on FL.host_id= DM.host_id
WHERE DM.is_superhost=0;
```

The results: The average number of beds and bathrooms in listings is similar for both superhosts and non-superhosts, averaging around two beds and approximately one bathroom.

	num_beds	num_bathrooms_superhosts
1	2	1.306267

	num_beds	num_bathrooms_non_superhosts
1	2	1.252690

- **Which neighborhoods have the top fifteen highest average review scores for properties?**

The coding process:

```
SELECT TOP 15 AVG(review_scores_rating)as average_review_score, fact_neighbourhood
FROM FactListings
GROUP BY fact_neighbourhood
ORDER BY average_review_score desc;
```

The results: As we can see from the results presented below, neighborhoods in Athens, along with some neighborhoods in the countryside, are aiming for high ratings. Kolonaki, Koukaki, Petralona, Thissio, and Kypriadou are some of the top-rated neighborhoods, with Acropolis also ranking in the top 15 list.

	average_review_score	fact_neighbourhood
1	5	Kypriadou, Greece
2	5	Athens, Victoria Square, Greece
3	5	Athens, PETRALONA, Greece
4	5	????a ??f?? , ?tt????, Greece
5	5	Attiki, Kolonaki, Athens, Greece
6	5	Athina, Famagusta, Greece
7	5	Athina , Greece, Athina, Koukaki , Greece
8	5	????a, ??? ??????, Greece
9	5	????a, ????a, Greece
10	4.99	Athina, Thissio, Greece
11	4.98	Athens , Athens, Greece
12	4.98	Athens, Acropolis, Greece
13	4.98	Greece
14	4.975	Athina, Peloponnisos Dytiki Ellada ke Ionio, Gre...
15	4.97	Ampelokipoi, Athens, Greece

- **Who are the top five hosts according to the number of reviews?**

The coding process:

```
SELECT TOP 5 DH.host_name, count(*) as review_count
FROM FactReview FR
JOIN FactListings FL on FR.listing_id = FL.id
JOIN DimHost DH on DH.host_id = FL.host_id
GROUP BY host_name
ORDER BY count(*) desc;
```

The results: According to the results, Toni seems to have received the highest numbers of reviews with the number reaching the 13505.

	host_name	review_count
1	Toni	13505
2	George	11514
3	Konstantinos	10466
4	Maria	9679
5	Evan&Anetta	8615

- **Find the listings who have received the highest number of reviews?**

The coding process:

```
SELECT TOP 5 FL.listing_name, COUNT (FR.review_id) as total_reviews
FROM FactListings as FL
INNER JOIN FactReview FR on FL.id = FR.listing_id
GROUP BY FL.listing_name
ORDER BY total_reviews DESC;
```

The results: The highest number of reviews per accommodation seems to escalate to the maximum number of 5034.

	listing_name	total_reviews
1	Rental unit in Athina · ?4.88 · 1 bedroom · 1 be...	5034
2	Rental unit in Athina · ?4.93 · 1 bedroom · 1 be...	4332
3	Rental unit in Athina · ?4.92 · 1 bedroom · 2 be...	4186
4	Rental unit in Athina · ?4.92 · 1 bedroom · 1 be...	3944
5	Rental unit in Athina · ?4.86 · 1 bedroom · 1 be...	3897

- **In which months do we have the most bookings?**

The coding process:

```
SELECT TOP 5 datename (month,booking_date) as booking_month, count(booking_date) as total_bookings
FROM Factcalendar
WHERE is_available=0
GROUP BY  datename (month,booking_date)
ORDER BY count(booking_date) desc;
```

The results: The month with the highest number of bookings seems to be the month of October. Also, September is high in the ranking.

	booking_month	total_bookings
1	October	57770
2	September	51018
3	August	43218
4	July	43192
5	June	42453

- **Find the listings with the most available dates**

The coding process:

```
WITH available_dates_per_listing as (
SELECT listing_id, count(is_available) as available_days
FROM Factcalendar
WHERE is_available=1
GROUP BY listing_id)
, max_avail_dates as (
SELECT max(available_days) as mx
FROM available_dates_per_listing
)
SELECT available_dates_per_listing.*
FROM available_dates_per_listing
JOIN max_avail_dates on available_days = mx;
```

The results: Almost 30 listings seem to reach the maximum number of available days someone can rent an Airbnb.

	listing_id	available_days
1	23260377	365
2	3150921	365
3	14034839	365
4	780733	365
5	3314818	365
6	26217775	365
7	4246787	365
8	27589300	365
9	13766522	365
10	784414	365
11	9768981	365
12	11566709	365
13	8835412	365
14	28068269	365
15	15539713	365
16	13455532	365
17	19077015	365
18	27958339	365
19	14818761	365
20	4812115	365
21	11542772	365
22	3461034	365
23	10808869	365
24	27958734	365
25	10473236	365
26	20023385	365
27	767272	365
28	10184538	365
29	1315027	365
30	25023777	365

- **Find the hosts with the highest number of bookings**

The coding process:

```
SELECT TOP 5 DM.host_name, count (booking_date) as total_bookings from Factcalendar as FC
INNER JOIN FactListings FL on FC.listing_id= FL.id
INNER JOIN DimHost DM ON DM.host_id= FL.host_id
WHERE FC.is_available=0
GROUP BY DM.host_name
ORDER BY count (booking_date) desc;
```

The results: According to the results, George seems to have received the highest numbers of bookings with the number reaching the 14663.

	host_name	total_bookings
1	George	14663
2	Toni	11419
3	Maria	9581
4	Konstantinos	7942
5	Nikos	7686

- **Find the average price, average minimum nights, and average maximum nights per room type.**

The coding process:

```
SELECT AVG(price) as average_price , AVG(FC.minimum_nights) as average_minimum_nights,
avg(FC.maximum_nights) as average_maximum_nights,DRT.room_type
from Factcalendar as FC
INNER JOIN FactListings FL on FC.listing_id= FL.id
INNER JOIN DimRoomType DRT on DRT.room_type_id=FL.room_type_id
GROUP BY FL.room_type_id,DRT.room_type
ORDER BY AVG(price) desc;
```

The results: The average minimum nights for Airbnb rentals show minimal variation across room types, ranging from 1 to 6. On the other hand, average maximum nights vary more noticeably, from 714 for private rooms to 1037 for hotel rooms.

	average_price	average_minimum_nights	average_maximum_nights	room_type
1	1146.54268493151	1	1037	Hotel room
2	142.473025220842	6	714	Private room
3	107.984666833354	4	912	Entire home/apt
4	33.7732876712329	3	846	Shared room