AirBnB Listings & Review Analysis with SQL

The data consist of information about reviews and listings of Airbnb's app. It features information about hosts, room and property types and also information related to bookings. The SQL tool I used to import the data was Microsoft SQL Server Management Studio (SSMS). The tables I created to analyze the data were divided into Fact and Dim Tables.

- Firstly, I created a fact table named FactReview which consists of information about the listing's reviews.

```
SELECT listing_id, id as review_id , date as review_date, comments,reviewer_id
INTO FactReview
FROM review;
```

- Second table I created was FactCalendar with data related to bookings, where I replaced null values with zero in the price and adjusted price columns. Also, I replaced with numeric values the column is_available .

```
SELECT listing_id,
FORMAT(date, 'yyyy-MM-dd') as booking_date,
ISNULL(price, 0) as price,
ISNULL(adjusted_price, 0) as adjusted_price ,
minimum_nights,maximum_nights,
CASE WHEN available= 'f' then 0
ELSE 1
END AS is_available
INTO Factcalendar
FROM calendars
```

- Third table I created was DimReviewer with data related to the person who made the review.

```
SELECT DISTINCT  reviewer_id,reviewer_name
INTO DimReviewer
FROM review
```

- Fourth table I created was FactListings. FactListings was the main table of the data analysis. In order to load the data correctly, I excluded all listing id that contained the E letter. Furthermore, since i tried to connect the Factlistings table with DimPropertyType and DimRoomType , I used the left join command.

```
SELECT
    CAST(id as bigint)as id ,CAST(name as nvarchar(max)) as listing_name,host_id,latitude,longitude,D.property_type_id,B.room_type_id,accommodates,bedrooms,
    beds,ISNULL(PRICE, 0) AS price,minimum_nights,maximum_nights,
    TRY_CAST(SUBSTRING(bathrooms_text, 1, CHARINDEX(' ', bathrooms_text + ' ') - 1) AS DECIMAL) AS bathrooms,
    CAST(ISNULL(neighbourhood, 'N/A') AS VARCHAR(150)) AS fact_neighbourhood,
    CAST(ISNULL(neighbourhood_cleansed, 'N/A') AS VARCHAR(150)) AS fact_neighbourhood_cleansed,
    CASE WHEN instant_bookable = 'f' THEN 0 ELSE 1 END AS is_instant_bookable,
    CASE WHEN has_availability = 'f' THEN 0 ELSE 1 END AS has_availability
INTO FactListings
FROM LISTING AS A
LEFT JOIN DimRoomType AS B ON A.room_type = B.room_type
LEFT JOIN DimPropertyType AS D ON A.property_type = D.property_type
WHERE id NOT LIKE '%E%';
```

- The two dimensional tables I've made to connect with Factlistings were the DimPropertyType table and the DimRoomType table. In order to connect those with the Fact table, I created two primary keys (one for each table) that start from one and increase by one for each new row.

```
CREATE TABLE DimRoomType (room_type_id int identity(1, 1) PRIMARY KEY  ,
room_type varchar(50)
)
INSERT INTO DimRoomType
SELECT distinct room_type
from listings
SELECT*from DimRoomType
CREATE TABLE DimPropertyType (property_type_id int identity(1, 1) PRIMARY KEY  ,
property_type varchar(50)
)
INSERT INTO DimPropertyType
SELECT distinct property_type
from listings
SELECT*from DimPropertyType
```

- Also, I created a fact table which I named as FactReviewsScores that contained the review scores for each listing.

```
SELECT id as listing_id, review_scores_accuracy,review_scores_checkin,
review_scores_cleanliness,review_scores_communication,review_scores_location,
review_scores_rating,review_scores_value
INTO FactReviewsScores
FROM listing
```

❖ Finally, I created a dim table named as DimHost that contained all information related to hosts.

```
SELECT DISTINCT host_id, host_name,host_since,host_response_time,host_acceptance_rate,host_response_rate,
ISNULL(host_location, 'N/A') AS dim_host_location,
ISNULL(host_neighbourhood,'N/A') AS dim_host_neighbourhood,
CASE WHEN host_is_superhost= 'f'then 0
ELSE 1
END AS is_superhost,
CASE WHEN host_has_profile_pic= 'f'then 0
ELSE 1
END AS has_profile_pic,
CASE WHEN host_identity_verified='f' then 0
ELSE 1
END AS identity_verified
INTO DimHost
FROM listing
```