Business questions

1. What are the most popular neighborhoods for short-term rentals in New York City?

```
select borough, count(listing_id) AS num_of_listings

from prices

group by 1

order by 2 desc
```

2. What is the average rental price for short-term rentals in New York City, and how does it vary by neighborhood and property type?

Assumptions > all property types are included no matter whether they are fully occupied (i.e. booked_days_365 is excluded)

```
with t1 AS (

  select p.listing_id, room_type

from prices p

inner join room_types r

on p.listing_id = r.listing_id)

select p.borough, t1.room_type, round(avg(p.price))

from prices p

left outer join t1

on p.listing_id = t1.listing_id

group by 1, 2

order by 1,2,3 desc
```

3. What are the most commonly rented property types on Airbnb in New York City, and how does this vary by neighborhoods.

We assumed that the most commonly rented property types are those occupied fully for the whole year (defined as number of listings and booked_days_365).

```
select borough, room_type, sum(num_listings) as total_num
from (
select borough, r.room_type as room_type,
    count(p.listing_id) AS num_listings
from prices p
inner join room_types r
on p.listing_id = r.listing_id
where p.listing_id in (
  select listing_id
  from reviews
  where booked_days_365 = 365)
group by 1,2) t1
group by 1,2
```

4. What is the average length of stay for short-term rentals in New York City, and how does this vary by neighborhood and property type?

   Answer: The given datasets did not have length of stay data for computation.

5. How has demand for short-term rentals in New York City changed over time, and are there any seasonal trends that could impact business decisions?

   Answer: No data available in the datasets to fully answer this question.

Analyze the data

1. **What is the most common room type in NYC Airbnb listings?**

```
SELECT room_type, count(listing_id)

FROM room_types

group by 1

order by 2 DESC
```

2. **What is the average price of a listing by room type?**

```
select room_type, round(avg(price))

from room_types r

inner join prices p

on p.listing_id=r.listing_id

group by 1

order by 2 DESC
```

3. **Which borough has the highest average price per month?**

```
select borough, round(avg(price_per_month))

from prices

group by 1

order by 2 desc

limit 1
```

4. **How many listings of each room type are in each borough?**

   Here we include all listings (no matter whether they are occupied throughout the whole year).

```
SELECT borough, room_type, count(r.listing_id)

FROM prices p

INNER JOIN room_types r
```

```
ON p.listing_id = r.listing_id

GROUP BY 1,2

ORDER BY 1,2
```

5. **How many listings in each room type category have a price of over $500 per night?**

```
select room_type, count(listing_id) as num_listing

from room_types

where listing_id in (

select listing_id

from prices

where price > 500)

group by 1

order by 2 desc
```

6. **What is the distribution of listing prices by neighborhood?**

```
with t1 AS (

SELECT borough, min(price), max(price),

     (max(price) - min(price)) as price_diff,

     round(avg(price)) as avg_price

from prices

group by 1

order by 3 desc, 4 desc)


select borough,price_diff,avg_price
```

```
from t1
```

7. **What is the estimated amount of revenue generated by hosts in each borough?**

```
select borough, sum(revenue)
from (
select p.borough, r.host_name, r.listing_id,
    (booked_days_365 * p.price) AS revenue
from reviews r
inner join prices p
on p.listing_id = r.listing_id
group by 1,2,3,4) sub
group by 1
order by 2 desc
```

8. **What is the average price per month for listings in each neighborhood?**

```
select borough,
    round(avg(price_per_month)) as avg_price
from prices
group by 1
order by 2 desc
```

9. **How many listings have no reviews?**

All listings have reviews. Rather, we found 7 listing_id did not have prices or borough records.

20333471
20523843
20639628
20639792
20639914
21291569
21304320

```
select listing_id
from reviews
where listing_id not in (
      select listing_id from prices)
```

10. **How do the estimated book days correlate with the price of an Airbnb listing in New York City?**

```
select corr(price, booked_days_365)
from prices p
inner join reviews r
on p.listing_id = r.listing_id
```

Bonus Questions:

What is the average price per room type for listings that have at least 100 reviews and are available more than 200 days a year?

```
select room_type, round(avg(p.price)) AS avg_price

from room_types r

inner join prices p

on r.listing_id = p.listing_id

where p.listing_id in (

select listing_id

from reviews

where number_of_reviews >= 100

    and availability_365 > 200)

group by 1

order by 2 desc
```

How many hosts have more than one listing, and what's the maximum number of listings by a single host name?

There are 3119 hosts that have more than one listing. The maximum number of listings by a single host name is 327.

```
select count(host_name), max(calculated_host_listings_count)

from (

SELECT distinct host_name, calculated_host_listings_count

from reviews

where calculated_host_listings_count > 1) t1
```

Determine the top 5 hosts who have the highest price_per_month for their listings, considering only hosts who have at least 10 listings.

```
select r.host_name, round(max(price_per_month)) AS max_price

from prices p

inner join reviews r

on p.listing_id = r.listing_id

where r.host_name in (

SELECT distinct host_name

FROM reviews

where calculated_host_listings_count >= 10)

group by 1

order by 2 desc

limit 5
```

Find the neighborhood(s) that have the highest variance in listing prices.

```
select neighbourhood, diff

from (

select neighbourhood, (max(price) - min(price)) AS diff

from prices

group by 1

order by 2 desc) sub

limit 5
```

Calculate the average price_per_month for each neighborhood, taking into account only listings where the host has a minimum_nights value that is higher than the average minimum_nights value across all listings.

```
select borough,

     round(avg(price_per_month)) AS avg_price_per_month

from prices p

left outer join reviews r

on p.listing_id = r.listing_id

where p.listing_id in (

select listing_id

from reviews

where minimum_nights > (

select round(avg(minimum_nights))

from reviews))

group by 1

order by 2 desc
```