



## Group 8



About:

Our goal for this project is to conduct a comprehensive analysis of the Airbnb dataset from Cape Town, South Africa. As a group, we will leverage our understanding of database management systems to explore the dataset, uncover important insights related to our chosen focus area, and present our findings.

**ZAHRA KARA – 41330595**

**MUHAMMED CAJEE – 43496385**

**ANGELINA RAMSUNAR – 41081269**

**DARIAN SCREUDER – 43552595**

**RIYA PATEL – 41914228**

**ROBERT FERREIRA – 33343136**

**ANDREW NARE - 31506543**

## Table of Contents

GROUP INFORMATION .....	3
PHASE 1:.....	5
1. Choose an area of focus .....	5
2. Define research questions .....	6
3. Dataset description.....	8
PHASE 2:.....	12
4. DB Design and Setup .....	12
5. Schema/ERD.....	16
6. SQL Query Development .....	17
Performance optimizing queries .....	29
7. Analysis and Reporting .....	30
Graphs and tables:.....	30
Final Report .....	36

## GROUP INFORMATION

---

ZAHRA KARA

(group leader)

41330595

076 874 9355



MUHAMMED CAJEE

43496385

081 516 1077

ANGELINA RAMSUNAR

41081269

071 162 4919



RIYA PATEL

41914228

084 811 2525

DARIAN SCHREUDER

43552595

076 548 4123



ROBERT FERREIRA

33343136

076 020 0509

ANDREW NARE

31506543

065 910 6522



# PHASE 1:

---

## 1. Choose an area of focus

Getting the most for their money is a top priority for customers searching for the greatest deal. The objective of our research is to conduct a comprehensive analysis of the Airbnb dataset from Cape Town, South Africa, focusing on identifying listings that provide the best value for money – commonly referred to as “Bang for your Buck”. Value, as seen by a client encompasses more than just a cheap rate, it also considers other elements such as amenities provided, accessibility to important sites, safety, and general guest satisfaction. By leveraging database management systems and data analytics techniques, we aim to explore key factors such as price, amenities, booking trends, and guest reviews to pinpoint which Airbnb listings offer the most rewarding experience at a fair price.

There are several reasons we feel our research is important. In a competitive market like Airbnb, where travelers have numerous options, identifying listings that offer the best value helps consumers make informed decisions, thus maximizing their satisfaction and experience during their stay. For Airbnb hosts, understanding what contributes to the value and gaining consumer feedback can greatly help them enhance their listings, set competitive prices, and attract more bookings. By assessing properties based on a broad spectrum of factors rather than just price, our research offers a more insightful understanding of what constitutes value in the short-term rental market.

Our research will enable us to gain the following insights: Understanding market dynamics, decision-making tools, seasonal occupancy insights, and identifying value trends. We can understand the market dynamics by gaining insights into how several factors such as amenities, proximity to tourist destinations, pricing strategies and guest reviews influence a listing's appeal and perceived value. Our research will allow us to create data-driven decision-making tools or models that can assist potential renters in finding the best value listings and help hosts optimize their offerings to attract more guests. We will also be able to determine the time of the year when rentals are most in demand and when they are less in demand. This research will allow us to identify patterns and trends in the data that reveal which listings provide the best balance of cost and quality, thereby offering travelers the best “bang for their buck”.

Overall, this research not only benefits consumers and hosts by optimizing the Airbnb experience, but it also contributes to a deeper understanding of the sharing economy and how value is perceived in the context of short-term rentals.

## 2. Define research questions

1. What is the price distribution for the listings with top rated amenities?
2. How does the frequency of reviews relate to the listing price and availability?
3. Which listings, when considering price, location, availability, and reviews, provide the most value for the money?
4. How do a host's general availability and number of ratings connect to the average price of their listings?
5. Are periods (certain days) when the price of listings is lower or higher?
6. What is the pricing of the listings with the most reviews in relation to those that are similar but have less reviews?
7. Do hosts with many listings provide a better deal than those with only one listing?
8. What is the difference in pricing between listings for different types of rooms (private room, shared room, complete home/apartment, etc.) in each neighbourhood?
9. What effect does the minimum night requirement have on the average listing price? Are properties with shorter minimum nights more costly?
10. What impact do reviews that were written within the previous month have on listing prices? Do freshly evaluated listings have a different price tag?
11. What seasonal differences exist in listing pricing, such as summer vs. winter? Are there particular seasons of the year when prices rise or fall noticeably?
12. Do older postings often cost more or less than newer listings, based on when they were originally reviewed?
13. How does the amount of reviews a host has received on all of their listings affect how much their experience is worth?
14. Is there a relationship between the number of amenities provided and the typical amount of favourable ratings a listing gets?
15. How many Airbnb listings are there overall, and how many are there in each neighbourhood?
16. Exist any neighbourhood's where the quantity of listings is either abnormally high or low?
17. Exist any neighbourhood's where the disparity in price is more noticeable?

18. What effect do unfavourable reviews have on listing prices and occupancy rates?
19. If a host receives a lot of negative feedback, can they bounce back? If yes, how?
20. Does the level of detail in guest reviews (e.g., long, descriptive reviews vs. short, generic ones) correlate with higher pricing or better value perceptions?
21. Are more detailed reviews associated with specific types of listings?
22. What criteria lead to repeat reservations and how frequently do customers return to the same listing?
23. How does a guest's place of origin affect the kinds of listings (area, kind of accommodation, etc.) that they like?
24. How much does it cost to offer a property that can accommodate varied party sizes (e.g., families vs single travellers)?
25. What changes have you noticed in the last year in Cape Town's Airbnb listing prices?
26. Are there any obvious price trends that correspond with demand throughout certain seasons or with economic factors?

### 3. Dataset description

#### Listing Details:

##### 1. **id**

- **Data Format:** Integer (Unique identifier)
- **Range/Domain:** A positive integer uniquely identifying each listing. This serves as the primary key for linking other data tables.

##### 2. **Listing\_url**

- **Data Format:** String (URL)
- **Range/Domain:** A valid URL string pointing to the listing's Airbnb page. It includes the complete URL structure (e.g., <https://www.airbnb.com/rooms/12345>).

##### 3. **Name**

- **Data Format:** String
- **Range/Domain:** Title of the listing, generally between 1 to 59 characters. It provides a brief name of the property.

##### 4. **Description**

- **Data Format:** String
- **Range/Domain:** Detailed description of the property provided by the host. The text varies in length and can be a few words to a whole paragraph, detailing the property's features and surroundings.

##### 5. **Host\_name**

- **Data Format:** String
- **Range/Domain:** Name of the host, typically 1 to 20 characters. It's often the first name or a nickname used by the host on Airbnb.

##### 6. **Host\_total\_listings\_count**

- **Data Format:** Integer
- **Range/Domain:** The total number of listings managed by the host, ranging from 1 (for single property owners) to larger numbers (for professional hosts with multiple properties).



## 7. Neighbourhood

- **Data Format:** String
- **Range/Domain:** General neighbourhood name where the listing is located, usually 1 to 50 characters. It provides an informal or colloquial name.

## 8. Neighbourhood\_cleansed

- **Data Format:** String
- **Range/Domain:** Standardized neighbourhood classification, typically more specific and consistent than Neighbourhood, 1 to 20 characters.

## 9. Property\_type

- **Data Format:** String
- **Range/Domain:** The type of property, such as “Apartment,” “House,” “Condo,” etc. Typically, it contains 1 to 30 characters and categorizes the listing based on its physical structure.

## 10. Accommodates

- **Data Format:** Integer
- **Range/Domain:** Number of guests the property can accommodate, typically ranging from 1 to 10+. This value helps potential guests determine the suitability of the property.

## 11. Amenities

- **Data Format:** String (List)
- **Range/Domain:** A comma-separated list of amenities such as “Wi-Fi,” “Kitchen,” “Parking.” The number and type of amenities vary widely between listings and could include a dozen or more items.

## 12. Number of reviews

- **Data Format:** Integer
- **Range/Domain:** Total number of reviews for the listing, typically ranges from 0 to 100+. This indicates the listing's popularity and level of engagement with guests.

## 13. Review\_scores\_rating

- **Data Format:** Float or Integer
- **Range/Domain:** Rating score, typically ranges from 0 to 5. It reflects the overall satisfaction of guests based on their stay experience

## **Review Data:**

### **1. Listing\_Id**

- **Data Format:** Integer (Unique identifier)
- **Range/Domain:** A positive integer. Serves as a foreign key from the listings table.

### **2. id**

- **Data Format:** Integer (Unique identifier)
- **Range/Domain:** A positive integer. Serves as a the primary key..

### **3. Date**

- **Data Format:** Date (MM-DD-YYYY)
- **Range/Domain:** Date when the review was posted. This typically falls within the timeframe covered by the dataset and may span several years.

### **4. Reviewer\_id**

- **Data Format:** Integer (Unique identifier)
- **Range/Domain:** A positive integer uniquely identifying each reviewer. This field links to the reviewer and their history.

### **5. Reviewer\_name**

- **Data Format:** String
- **Range/Domain:** Name of the reviewer, typically 1 to 20 characters. This can be a first name, nickname, or full name, depending on the reviewer's preference.

### **6. comments**

- **Data Format:** String
- **Range/Domain:** Text of the review left by the guest, which can vary from a few words to detailed feedback. It provides qualitative insights into the guest's experience.

## **Calendar Data:**

### **1. Listing\_id**

- **Data Format:** Integer (Unique identifier)
- **Range/Domain:** A positive integer linking the availability data to a specific listing. It corresponds directly to the Listing\_id in the listing details.

## 2. Date

- **Data Format:** Date (MM-DD-YYYY)
- **Range/Domain:** Dates on which availability and pricing are recorded. The date range typically spans a full year or more to capture booking trends.

## 3. Available

- **Data Format:** String (Boolean-like)
- **Range/Domain:** “Yes” or “No” indicating if the property is available on a specific date. This binary value is crucial for determining vacancy periods.

## 4. Price

- **Data Format:** Float (Currency format)
- **Range/Domain:** Price per night for the listing on that specific date. It typically ranges from \$10 to \$1000+, depending on factors like location, property type, and season.

## 5. Adjusted price

- **Data Format:** Float (Currency format)
- **Range/Domain:** Price after adjustments like discounts or special offers. It follows the same range as the Price but may be lower due to promotions.

## 6. Minimum nights

- **Data Format:** Integer
- **Range/Domain:** Minimum number of nights required to book, usually ranges from 1 to 6 nights. This field ensures that guests meet the host’s booking conditions.

## 7. Maximum nights

- **Data Format:** Integer
- **Range/Domain:** Maximum number of nights allowed for booking, typically ranges from 1 to 365+ nights. This cap is set by the host to manage the length of stays.

## Focus Area

The dataset we are working with provides a comprehensive view of Airbnb listings, covering essential aspects like property details, guest feedback, and availability. By focusing on key attributes, we can analyse factors such as pricing trends, guest preferences, neighbourhood popularity, and host reliability. These insights are crucial for making informed decisions and conducting further analysis based on our specific objectives.

## PHASE 2:

### 4. DB Design and Setup

#### LISTINGS TABLE

Create Table:

```
CREATE TABLE LISTINGS (  
    LISTING_ID NUMBER PRIMARY KEY, -- Primary key, cannot be null  
    LISTING_URL VARCHAR2(40), -- URL, optional  
    LISTING_NAME VARCHAR2(70) NOT NULL, -- Listing name, required  
    LISTING_DESCRIPTION CLOB, -- Listing description as a CLOB (for large text)  
    HOST_NAME VARCHAR2(50) NOT NULL, -- Host name, required  
    HOST_TOTAL_LISTINGS_COUNT NUMBER, -- Total listings count, optional  
    NEIGHBOURHOOD VARCHAR2(60), -- Neighbourhood, optional  
    NEIGHBOURHOOD_CLEANSSED VARCHAR2(10), -- Cleaned neighbourhood name, optional  
    PROPERTY_TYPE VARCHAR2(40), -- Property type, optional  
    ACCOMADATES NUMBER, -- Accommodates, optional  
    AMENITIES CLOB, -- Amenities, optional (assuming list stored as a string)  
    NUMBER_OF_REVIEWS NUMBER DEFAULT 0, -- Number of reviews, default is 0  
    REVIEW_SCORES_RATING NUMBER CHECK (REVIEW_SCORES_RATING BETWEEN 0 AND 5) -- Review rating range from 0 to 5  
);
```

Table View:

LISTING\_ID (Primary Key) (not null)

LISTING\_URL

LISTING\_NAME (not null)

LISTING\_DESCRIPTION

HOST\_NAME (not null)

HOST\_TOTAL\_LISTINGS\_COUNT

NEIGHBOURHOOD

NEIGHBOURHOOD\_CLEANSSED

PROPERTY\_TYPE

ACCOMADATES

AMENITIES

NUMBER\_OF\_REVIEWS (default value of 0)

REVIEW\_SCORES\_RATING (ranges between 0 and 5)

## REVIEW TABLE

Create Table:

```
CREATE TABLE REVIEW (  
    LISTING_ID NUMBER NOT NULL, -- LISTING_ID is a foreign key, references LISTINGS  
    REVIEW_ID NUMBER PRIMARY KEY, -- Primary key for the review table  
    REVIEW_DATE DATE NOT NULL, -- Review date, required  
    REVIEWER_ID NUMBER NOT NULL, -- Reviewer ID, required  
    REVIEWER_NAME VARCHAR2(40) NOT NULL, -- Reviewer name, max length 40, required  
    COMMENTS CLOB, -- Comments, optional (can be null)  
    CONSTRAINT fk_listing_review FOREIGN KEY (LISTING_ID) REFERENCES LISTINGS(LISTING_ID)  
);
```

Table View:

Schema: DATADragons

Name: REVIEW

Table Type: Normal

Search

Columns: name

PK	Name	Data Type	Size	Not Null	Default	Comment
	LISTING_ID	NUMBER		<input checked="" type="checkbox"/>		
	REVIEW_ID	NUMBER		<input checked="" type="checkbox"/>		
	REVIEW_DATE	DATE		<input checked="" type="checkbox"/>		
	REVIEWER_ID	NUMBER		<input checked="" type="checkbox"/>		
	REVIEWER_N...	VARCHAR2	40	<input type="checkbox"/>		
	COMMENTS	CLOB		<input type="checkbox"/>		

Constraints on Column:

Constraint Name	Constraint Type	Other Columns
SYS_C007636	Foreign Key	

LISTING\_ID (foreign key) (from LISTINGS table)

REVIEW\_ID (primary key) (not null)

REVIEW\_DATE (not null)

REVIEWER\_ID (not null)

REVIEWER\_NAME (not null)

COMMENTS

## CALENDAR TABLE

Create Table:

```
CREATE TABLE CALENDAR (  
    LISTING_ID NUMBER NOT NULL, -- Foreign key and primary key linking to LISTINGS  
    BOOKING_DATE DATE NOT NULL, -- Booking date, required  
    AVAILABLE CHAR(1), -- Availability optional  
    PRICE VARCHAR2(10), -- Price, optional (using 2 decimal places)  
    ADJUSTED_PRICE VARCHAR2(10), -- Adjusted price, optional (using 2 decimal places)  
    MINIMUM_NIGHTS NUMBER DEFAULT 1, -- Minimum nights, default is 1  
    MAXIMUM_NIGHTS NUMBER, -- Maximum nights, optional  
    CONSTRAINT fk_listing_calendar FOREIGN KEY (LISTING_ID) REFERENCES LISTINGS(LISTING_ID) -- Foreign key to LISTINGS  
);
```

Table View:

Edit Table

Schema: DATADragons

Name: CALENDAR

Table Type: Normal

Search

Columns

Constraints

Indexes

Storage

Comment

DDL

PK	Name	Data Type	Size	Not Null	Default	Comment
	LISTING_ID	NUMBER		<input checked="" type="checkbox"/>		
	BOOKING_DATE	DATE		<input checked="" type="checkbox"/>		
	AVAILABLE	CHAR	1	<input type="checkbox"/>		
	PRICE	VARCHAR2	10	<input type="checkbox"/>		
	ADJUSTED_PR...	VARCHAR2	10	<input type="checkbox"/>		
	MINIMUM_NIG...	NUMBER		<input checked="" type="checkbox"/>	1	
	MAXIMUM_NI...	NUMBER		<input type="checkbox"/>		

Data Type Constraints Indexes LOB Parameters Identity Column

Constraints on Column:

Constraint Name	Constraint Type	Other Columns
FK_CALENDAR_LISTING	Foreign Key	
SYS_C007637	Primary Key	

LISTING\_ID (primary key, foreign key) (not null)

BOOKING\_DATE (not null)

AVAILABLE (true or false)

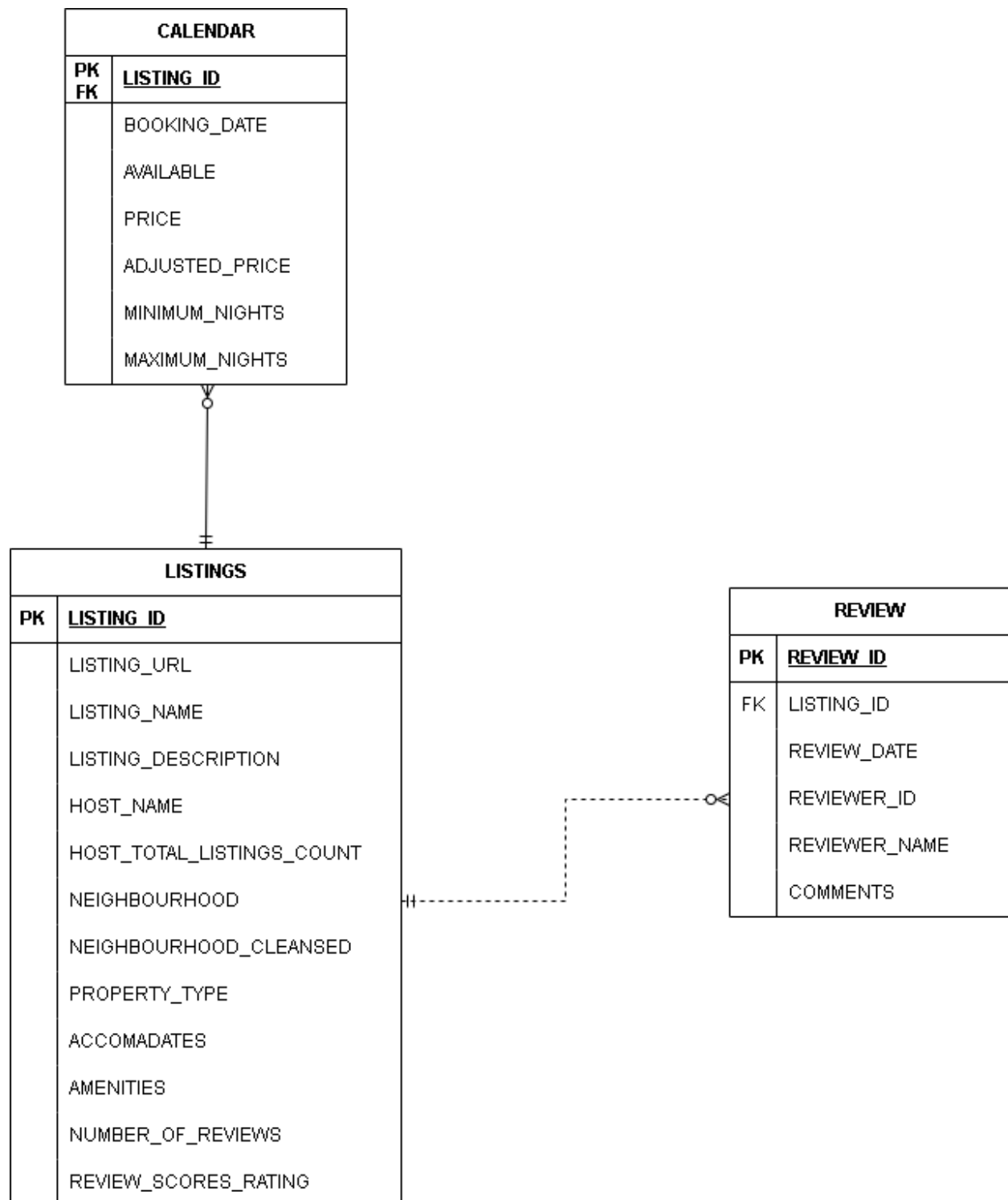
PRICE

ADJUSTED\_PRICE

MINIMUM\_NIGHTS (not null) (default value of 1)

MAXIMUM\_NIGHTS

## 5. Schema/ERD





## 6. SQL Query Development

### 1. What is the price distribution for the listings with top rated amenities?

```
--1.What is the price distribution for listings with top-rated amenities?--  
SELECT  
    CASE  
        WHEN TO_NUMBER(C.PRICE) < 5000 THEN 'Under 5,000'  
        WHEN TO_NUMBER(C.PRICE) BETWEEN 5000 AND 15000 THEN '5,000 - 15,000'  
        WHEN TO_NUMBER(C.PRICE) BETWEEN 15000 AND 30000 THEN '15,000 - 30,000'  
        ELSE 'Above 30,000'  
    END AS Price_Range,  
    COUNT(DISTINCT C.LISTING_ID) AS Listing_Count  
FROM CALENDAR C  
JOIN LISTINGS L ON C.LISTING_ID = L.LISTING_ID  
WHERE  
    L.AMENITIES LIKE '%Wifi%' AND L.AMENITIES LIKE '%Air conditioning%' AND L.AMENITIES LIKE '%Heating%' AND C.AVAILABLE = 't'  
GROUP BY  
    CASE  
        WHEN TO_NUMBER(C.PRICE) < 5000 THEN 'Under 5,000'  
        WHEN TO_NUMBER(C.PRICE) BETWEEN 5000 AND 15000 THEN '5,000 - 15,000'  
        WHEN TO_NUMBER(C.PRICE) BETWEEN 15000 AND 30000 THEN '15,000 - 30,000'  
        ELSE 'Above 30,000'  
    END;  
END;
```

Script Output x Query Result x

SQL | All Rows Fetched: 4 in 0.099 seconds

PRICE_RANGE	LISTING_COUNT
1 5,000 - 15,000	52
2 Under 5,000	215
3 15,000 - 30,000	8
4 Above 30,000	6

Categorizes listings into price ranges based on the presence of amenities like Wi-Fi, air conditioning, and heating.

### 2. How does the frequency of reviews relate to the listing price and availability?

```
--2. How does the frequency of reviews relate to the listing price and availability?--  
SELECT L.NUMBER_OF_REVIEWS, C.PRICE, C.AVAILABLE  
FROM LISTINGS L  
JOIN CALENDAR C ON L.LISTING_ID = C.LISTING_ID  
ORDER BY L.NUMBER_OF_REVIEWS DESC;
```

Script Output x Query Result x

SQL | Fetched 50 rows in 0.065 seconds

NUMBER_OF_REVIEWS	PRICE	AVAILABLE
1	644 645.00	f
2	637 850.00	t
3	487 12000.00	f
4	485 750.00	t
5	452 760.00	t
6	423 1295.00	t
7	401 650.00	t

Shows the relationship between the number of reviews, listing price, and availability.

### 3. Which listings, when considering price, location, availability, and reviews, provide the most value for the money?

```
--3.Which listings, considering price, location, availability, and reviews, provide the most value for the money?--
SELECT L.LISTING_ID, L.NEIGHBOURHOOD_CLEANSSED, C.PRICE, C.ADJUSTED_PRICE, L.NUMBER_OF_REVIEWS, L.REVIEW_SCORES_RATING,
       C.AVAILABLE,
       ROUND(L.REVIEW_SCORES_RATING / NULLIF(CAST(REGEXP_REPLACE(NVL(C.ADJUSTED_PRICE, C.PRICE), '^0-9.', '')) AS NUMBER), 0), 3)
       AS VALUE_FOR_MONEY
FROM LISTINGS L
JOIN CALENDAR C ON L.LISTING_ID = C.LISTING_ID
WHERE C.AVAILABLE = 't'
ORDER BY VALUE_FOR_MONEY DESC;
```

ript Output x Query Result x

SQL | Fetched 300 rows in 0.045 seconds

	LISTING_ID	NEIGHBOURHOOD_CLEANSSED	PRICE	ADJUSTED_PRICE	NUMBER_OF_REVIEWS	REVIEW_SCORES_RATING	AVAILABLE	VALUE_FOR_MONEY
23	1992333	Ward 23	70.00	(null)	60	4.82	t	0.069
24	6320435	Ward 54	70.00	(null)	68	4.79	t	0.068
25	53666830	Ward 54	78.00	(null)	29	4.97	t	0.064
26	9949005	Ward 49	80.00	(null)	1	5	t	0.063
27	32647089	Ward 23	79.00	(null)	26	5	t	0.063
28	31379142	Ward 77	74.00	(null)	6	4.67	t	0.063
29	24372654	Ward 74	75.00	(null)	40	4.73	t	0.063
30	24217479	Ward 115	79.00	(null)	156	4.87	t	0.062
31	53911755	Ward 55	77.00	(null)	7	4.71	t	0.061
32	20337058	Ward 74	84.00	(null)	94	4.85	t	0.058

Identifies the listings offering the best value by comparing price, availability, and reviews.

### 4. How do a host's general availability and number of ratings connect to the average price of their listings?

```
--4.How do a host's general availability and number of ratings connect to the average price of their listings?--
SELECT L.HOST_NAME, SUM(L.NUMBER_OF_REVIEWS) AS RATING_COUNT, ROUND(AVG(NVL(TO_NUMBER(C.PRICE), 0)), 2) AS AVG_PRICE,
       COUNT(CASE WHEN C.AVAILABLE = 't' THEN 1 END) AS AVAILABLE_DAYS -- Count only available days
FROM LISTINGS L
JOIN CALENDAR C ON L.LISTING_ID = C.LISTING_ID
GROUP BY L.HOST_NAME
ORDER BY RATING_COUNT DESC;
```

```
--5.Are there periods when the price of listings is lower or higher?--
```

ript Output x Query Result x

SQL | Fetched 150 rows in 0.033 seconds

	HOST_NAME	RATING_COUNT	AVG_PRICE	AVAILABLE_DAYS
4	Lance	1652	3184.3	6
5	Scott	1313	2067.33	5
6	Jayson	1177	1562.13	6
7	Sue	1036	1810.27	5
8	Uta	990	4800.86	4
9	Sally	969	1950.33	9
10	James	966	2533.59	6
11	Carol	958	2192.42	8

Examines how a host's overall availability and number of reviews affect the average price of their listings.

5. Are there periods (certain days) when the price of listings is lower or higher?

```
--5.Are there periods when the price of listings is lower or higher?--
SELECT TRUNC(BOOKING_DATE, 'MM') AS MONTH_START, -- Truncate to the first day of the month
       ROUND(AVG(NVL(TO_NUMBER(PRICE), 0)), 0) AS AVERAGE_MONTHLY_PRICE, MIN(NVL(TO_NUMBER(PRICE), 0)) AS MIN_MONTHLY_PRICE,
       MAX(NVL(TO_NUMBER(PRICE), 0)) AS MAX_MONTHLY_PRICE
FROM CALENDAR
GROUP BY TRUNC(BOOKING_DATE, 'MM') -- Group by the start of the month
ORDER BY MONTH_START;
```

Script Output x Query Result x

SQL | All Rows Fetched: 13 in 0.007 seconds

	MONTH_START	AVERAGE_MONTHLY_PRICE	MIN_MONTHLY_PRICE	MAX_MONTHLY_PRICE
1	01-JUN-24	2105	35	12400
2	01-JUL-24	2563	24	40000
3	01-AUG-24	2979	24	65000
4	01-SEP-24	2409	25	35000
5	01-OCT-24	2757	32	80000
6	01-NOV-24	2630	25	50000
7	01-DEC-24	2183	27	22000
8	01-JAN-25	2614	35	33696
9	01-FEB-25	2268	40	25000
10	01-MAR-25	2520	25	24960
11	01-APR-25	2825	30	77000
12	01-MAY-25	2428	21	49500

Analyses how listing prices vary by month to identify any seasonal trends.

6. What is the pricing of the listings with the most reviews in relation to those that are similar but have less reviews?

```
--6.What is the pricing of the listings with the most reviews in relation to those with fewer reviews?--
SELECT L.NUMBER_OF_REVIEWS, AVG(C.PRICE) AS AVERAGE_PRICE
FROM LISTINGS L
JOIN CALENDAR C ON L.LISTING_ID = C.LISTING_ID
GROUP BY L.NUMBER_OF_REVIEWS
ORDER BY L.NUMBER_OF_REVIEWS DESC;
```

Script Output x Query Result x

SQL | Fetched 50 rows in 0.017 seconds

	NUMBER_OF_REVIEWS	AVERAGE_PRICE
1	644	645
2	637	850
3	487	12000
4	485	750
5	452	760
6	423	1295
7	401	650
8	391	422

Compares the average price of listings with different numbers of reviews.

7. Do hosts with many listings provide a better deal than those with only one listing?

```
--7.Do hosts with many listings provide a better deal than those with only one listing?--
```

```
WITH HOSTLISTINGCOUNTS AS (
    SELECT L.HOST_NAME, COUNT(L.LISTING_ID) AS TOTAL_LISTINGS
    FROM LISTINGS L
    GROUP BY L.HOST_NAME
)
SELECT HLC.TOTAL_LISTINGS, ROUND(AVG(C.PRICE),0) AS AVERAGE_PRICE
FROM HOSTLISTINGCOUNTS HLC
JOIN LISTINGS L ON HLC.HOST_NAME = L.HOST_NAME
JOIN CALENDAR C ON L.LISTING_ID = C.LISTING_ID
GROUP BY HLC.TOTAL_LISTINGS
ORDER BY HLC.TOTAL_LISTINGS DESC;
```

Script Output x Query Result x

SQL | All Rows Fetched: 47 in 0.074 seconds

	TOTAL_LISTINGS	AVERAGE_PRICE
1	118	2688
2	74	10125
3	59	2528
4	56	970
5	54	2774

Determines whether hosts with more listings offer better prices compared to those with fewer listings.

8. What is the difference in pricing between listings for different types of rooms (private room, shared room, complete home/apartment, etc.) in each neighbourhood?

```
--8.What is the difference in pricing between room types in different neighborhoods?--
```

```
SELECT L.PROPERTY_TYPE AS ROOM_TYPE, L.NEIGHBOURHOOD_CLEANSSED AS NEIGHBORHOOD, ROUND(AVG(C.PRICE),0) AS AVERAGE_PRICE,
MIN(C.PRICE) AS MIN_PRICE, MAX(C.PRICE) AS MAX_PRICE
FROM LISTINGS L
JOIN CALENDAR C ON L.LISTING_ID = C.LISTING_ID
WHERE C.AVAILABLE = 't'
GROUP BY L.PROPERTY_TYPE, L.NEIGHBOURHOOD_CLEANSSED
ORDER BY ROOM_TYPE;
```

Script Output x Query Result x

SQL | Fetched 50 rows in 0.019 seconds

	ROOM_TYPE	NEIGHBORHOOD	AVERAGE_PRICE	MIN_PRICE	MAX_PRICE
1	Camper/RV	Ward 61	1200	1200.00	1200.00
2	Casa particular	Ward 26	412	412.00	412.00
3	Earthen home	Ward 69	800	800.00	800.00
4	Entire bungalow	Ward 107	1300	1300.00	1300.00
5	Entire bungalow	Ward 54	13025	12800.00	9300.00
6	Entire bungalow	Ward 61	5500	2000.00	9000.00
7	Entire bungalow	Ward 62	550	550.00	550.00
8	Entire bungalow	Ward 77	950	950.00	950.00

Compares the pricing of different property types across neighbourhoods.

9. What effect does the minimum night requirement have on the average listing price? Are properties with shorter minimum nights more costly?

```
--9.What effect does the minimum night requirement have on the average listing price?--
SELECT C.MINIMUM_NIGHTS, ROUND(AVG(C.PRICE),0) AS AVERAGE_PRICE
FROM CALENDAR C
GROUP BY C.MINIMUM_NIGHTS
ORDER BY C.MINIMUM_NIGHTS DESC;
```

Script Output x Query Result x

SQL | All Rows Fetched: 34 in 0.005 seconds

	MINIMUM_NIGHTS	AVERAGE_PRICE
29	6	2535
30	5	3218
31	4	2944
32	3	3132
33	2	1678
34	1	1719

Investigates the relationship between the minimum stay requirement and the average price of listings.

- 10.What impact do reviews that were written within the previous 6 months have on listing prices? Do freshly evaluated listings have a different price tag?

```
--10.What impact do recent reviews have on listing prices?--
SELECT
CASE
WHEN R.NUMBER_OF_REVIEWS = 0 THEN 'No Reviews'
WHEN R.NUMBER_OF_REVIEWS BETWEEN 1 AND 5 THEN '1-5 Reviews'
WHEN R.NUMBER_OF_REVIEWS BETWEEN 6 AND 10 THEN '6-10 Reviews'
ELSE 'More than 10 Reviews'
END AS REVIEW_CATEGORY, ROUND(AVG(C.PRICE), 0) AS AVERAGE_PRICE
FROM (
SELECT LISTING_ID, COUNT(REVIEW_ID) AS NUMBER_OF_REVIEWS
FROM REVIEW
WHERE REVIEW_DATE >= SYSDATE - 200 -- Filter for reviews in the last 30 days
GROUP BY LISTING_ID
) R
JOIN LISTINGS L ON R.LISTING_ID = L.LISTING_ID
JOIN CALENDAR C ON L.LISTING_ID = C.LISTING_ID
WHERE C.AVAILABLE = 't'
GROUP BY CASE WHEN R.NUMBER_OF_REVIEWS = 0 THEN 'No Reviews' WHEN R.NUMBER_OF_REVIEWS BETWEEN 1 AND 5 THEN '1-5 Reviews'
WHEN R.NUMBER_OF_REVIEWS BETWEEN 6 AND 10 THEN '6-10 Reviews' ELSE 'More than 10 Reviews'
END
ORDER BY REVIEW_CATEGORY;
```

Script Output x Query Result x

SQL | All Rows Fetched: 3 in 0.405 seconds

REVIEW_CATEGORY	AVERAGE_PRICE
1 1-5 Reviews	1784
2 6-10 Reviews	2293
3 More than 10 Reviews	1277

Analyses how the number of recent reviews affects the listing price.

11. What seasonal differences exist in listing pricing, such as summer vs. winter? Are there particular seasons of the year when prices rise or fall noticeably?

```
--11. What seasonal differences exist in listing pricing?--
SELECT EXTRACT(YEAR FROM C.BOOKING_DATE) AS YEAR,
CASE
WHEN EXTRACT(MONTH FROM C.BOOKING_DATE) IN (12, 1, 2) THEN 'Winter'
WHEN EXTRACT(MONTH FROM C.BOOKING_DATE) IN (3, 4, 5) THEN 'Spring'
WHEN EXTRACT(MONTH FROM C.BOOKING_DATE) IN (6, 7, 8) THEN 'Summer'
ELSE 'Autumn'
END AS SEASON, ROUND(AVG(C.PRICE),0) AS AVERAGE_PRICE
FROM CALENDAR C
JOIN LISTINGS L ON C.LISTING_ID = L.LISTING_ID
WHERE C.AVAILABLE = 't'
GROUP BY EXTRACT(YEAR FROM C.BOOKING_DATE),
CASE
WHEN EXTRACT(MONTH FROM C.BOOKING_DATE) IN (12, 1, 2) THEN 'Winter'
WHEN EXTRACT(MONTH FROM C.BOOKING_DATE) IN (3, 4, 5) THEN 'Spring'
WHEN EXTRACT(MONTH FROM C.BOOKING_DATE) IN (6, 7, 8) THEN 'Summer'
ELSE 'Autumn'
END
ORDER BY YEAR, SEASON;
```

Script Output x Query Result x

SQL | All Rows Fetched: 6 in 0.008 seconds

	YEAR	SEASON	AVERAGE_PRICE
1	2024	Autumn	2507
2	2024	Summer	2499
3	2024	Winter	1951
4	2025	Spring	2956

Explores how listing prices change across different seasons of the year.

12. Do older postings often cost more or less than newer listings, based on when they were originally reviewed?

```
--12. Do older listings cost more or less than newer listings?--
SELECT EXTRACT(YEAR FROM R.REVIEW_DATE) AS REVIEW_YEAR, ROUND(AVG(C.PRICE),0) AS AVERAGE_PRICE
FROM REVIEW R
JOIN CALENDAR C ON R.LISTING_ID = C.LISTING_ID
GROUP BY EXTRACT(YEAR FROM R.REVIEW_DATE)
ORDER BY REVIEW_YEAR;
```

Script Output x Query Result x

SQL | All Rows Fetched: 13 in 0.353 seconds

	REVIEW_YEAR	AVERAGE_PRICE
7	2018	1851
8	2019	1798
9	2020	2029
10	2021	1989

Compares the prices of older listings (with reviews from previous years) against newer listings.

13. How does the amount of reviews a host has received on all of their listings affect how much their experience is worth?

```
--13. How does the amount of reviews a host has received on all of their listings affect how much their experience is worth?--
SELECT L.HOST_NAME, SUM(L.NUMBER_OF_REVIEWS) AS TOTAL_REVIEWS, ROUND(AVG(C.PRICE),0) AS AVG_PRICE
FROM LISTINGS L
JOIN CALENDAR C ON L.LISTING_ID = C.LISTING_ID
WHERE C.AVAILABLE = 't' -- Ensure you only count available listings
GROUP BY L.HOST_NAME
ORDER BY TOTAL_REVIEWS DESC;
```

ript Output x Query Result x  
 SQL | Fetched 50 rows in 0.011 seconds

	HOST_NAME	TOTAL_REVIEWS	AVG_PRICE
1	Propr	2587	2630
2	Albertus	1540	3514
3	Lance	1221	858
4	Charmaine	970	1827
5	Jayson	879	1198

Evaluates whether hosts with more reviews charge higher prices.

14. Is there a relationship between the number of amenities provided and the typical amount of favourable ratings a listing gets?

```
--14. Is there a relationship between the number of amenities provided and the typical amount of favorable ratings a listing gets?--
SELECT AMENITIES_INFO.LISTING_ID, AMENITIES_INFO.NUMBER_OF_AMENITIES, AVG(L.REVIEW_SCORES_RATING) AS AVERAGE_RATING
FROM
(
SELECT L.LISTING_ID, REGEXP_COUNT(REPLACE(REPLACE(DBMS_LOB.SUBSTR(L.AMENITIES, 4000, 1), '[', ' '), ']', ' '), '^[^,]+')
AS NUMBER_OF_AMENITIES
FROM LISTINGS L
WHERE L.AMENITIES IS NOT NULL
) AMENITIES_INFO
JOIN LISTINGS L ON AMENITIES_INFO.LISTING_ID = L.LISTING_ID
GROUP BY AMENITIES_INFO.LISTING_ID, AMENITIES_INFO.NUMBER_OF_AMENITIES;
```

ript Output x Query Result x  
 SQL | Fetched 100 rows in 1.396 seconds

	LISTING_ID	NUMBER_OF_AMENITIES	AVERAGE_RATING
17	4012243	52	4.92
18	3265646	32	4.79
19	3277357	58	4.91
50	4043468	64	4.85
51	4045171	41	4.69
52	3358000	52	4.7

Investigates the connection between the number of amenities and guest ratings.

15. How many Airbnb listings are there overall, and how many are there in each neighbourhood?

```
--15. How many Airbnb listings are there overall, and how many are there in each neighborhood?--
SELECT COALESCE(NEIGHBOURHOOD_CLEANSSED, 'Total Listings') AS NEIGHBORHOOD, COUNT(LISTING_ID) AS TOTAL_LISTINGS
FROM LISTINGS
GROUP BY ROLLUP (NEIGHBOURHOOD_CLEANSSED);
```

NEIGHBORHOOD	TOTAL_LISTINGS
1 Ward 1	94
2 Ward 10	6
3 Ward 100	210
4 Ward 101	1
5 Ward 102	25

Shows the number of Airbnb listings in each neighborhood.

16. Exist any neighbourhood's where the quantity of listings is either abnormally high or low?

```
--16. Are there any neighborhoods where the quantity of listings is either abnormally high or low?--
SELECT NEIGHBOURHOOD_CLEANSSED, COUNT(LISTING_ID) AS TOTAL_LISTINGS
FROM LISTINGS
GROUP BY NEIGHBOURHOOD_CLEANSSED
HAVING
COUNT(LISTING_ID) > (SELECT AVG(LISTINGCOUNT) + 1.5 * STDDEV(LISTINGCOUNT)
FROM (SELECT COUNT(LISTING_ID) AS LISTINGCOUNT
FROM LISTINGS
GROUP BY NEIGHBOURHOOD_CLEANSSED))
OR
COUNT(LISTING_ID) < (SELECT AVG(LISTINGCOUNT) - 1.5 * STDDEV(LISTINGCOUNT)
FROM (SELECT COUNT(LISTING_ID) AS LISTINGCOUNT
FROM LISTINGS
GROUP BY NEIGHBOURHOOD_CLEANSSED))
ORDER BY TOTAL_LISTINGS DESC;
```

NEIGHBOURHOOD_CLEANSSED	TOTAL_LISTINGS
1 Ward 115	1848
2 Ward 54	1322
3 Ward 77	668
4 Ward 74	547
5 Ward 23	546
6 Ward 64	522

Identifies neighbourhoods with abnormally high or low numbers of listings.



## 17. Exist any neighbourhood's where the disparity in price is more noticeable?

```
--17. Are there any neighborhoods where the disparity in price is more noticeable?--
SELECT L.NEIGHBOURHOOD, MAX(C.PRICE) - MIN(C.PRICE) AS PRICE_DISPARITY
FROM LISTINGS L
JOIN CALENDAR C ON L.LISTING_ID = C.LISTING_ID -- Adjust the join condition based on your primary and foreign keys
WHERE C.AVAILABLE = 't' -- Consider only available listings
GROUP BY L.NEIGHBOURHOOD
ORDER BY PRICE_DISPARITY DESC;
```

Script Output x Query Result x

SQL | Fetched 50 rows in 0.013 seconds

NEIGHBOURHOOD	PRICE_DISPARITY
2 Gordons Bay, Western Cape, South Africa	1705
3 Fish Hoek, Western Cape, South Africa	1000
4 Cape Town, Western Cape, South Africa	899
5 Strand, Cape Town, Western Cape, South Africa	800
6 Scarborough, Western Cape, South Africa	0
7 Gardens, Cape Town, South Africa	0

Finds neighbourhoods with the greatest price disparity between the most and least expensive listings.

## 18. What effect do unfavourable reviews have on listing prices and occupancy rates?

```
--18. What effect do unfavorable reviews have on listing prices and occupancy rates?--
SELECT
CASE
WHEN AVG_REVIEW_SCORE < 3 THEN 'Unfavorable Reviews'
ELSE 'Favorable Reviews'
END AS REVIEW_TYPE, ROUND(AVG(TO_NUMBER(C.PRICE)),2) AS AVERAGE_PRICE, ROUND(AVG(CASE WHEN C.AVAILABLE = 't' THEN 1 ELSE 0 END),2)
AS AVERAGE_OCCUPANCY_RATE
FROM (
SELECT L.LISTING_ID, AVG(L.REVIEW_SCORES_RATING) AS AVG_REVIEW_SCORE
FROM LISTINGS L
JOIN REVIEW R ON L.LISTING_ID = R.LISTING_ID
GROUP BY L.LISTING_ID
) SUBQUERY
JOIN CALENDAR C ON SUBQUERY.LISTING_ID = C.LISTING_ID
GROUP BY
CASE
WHEN AVG_REVIEW_SCORE < 3 THEN 'Unfavorable Reviews'
ELSE 'Favorable Reviews'
END
ORDER BY REVIEW_TYPE;
```

Script Output x Query Result x

SQL | All Rows Fetched: 2 in 0.405 seconds

REVIEW_TYPE	AVERAGE_PRICE	AVERAGE_OCCUPANCY_RATE
1 Favorable Reviews	2210.81	0.53
2 Unfavorable Reviews	615	0.5

Shows how low review scores impact listing prices and occupancy rates.

19.If a host receives a lot of negative feedback, can they bounce back? If yes, how?

```
--19. If a host receives a lot of negative feedback, can they bounce back? If yes, how?--
SELECT L.LISTING_ID, L.REVIEW_SCORES_RATING, C.BOOKING_DATE, C.PRICE
FROM LISTINGS L
JOIN CALENDAR C ON L.LISTING_ID = C.LISTING_ID
WHERE L.REVIEW_SCORES_RATING < 3 -- Negative reviews threshold
ORDER BY C.BOOKING_DATE DESC; -- Ordering by booking date
```

Script Output x Query Result x

SQL | All Rows Fetched: 7 in 0.015 seconds

	LISTING_ID	REVIEW_SCORES_RATING	BOOKING_DATE	PRICE
1	35813782	1.5	29-JUN-25	880.00
2	42754324	1	12-JUN-25	412.00
3	21031952	1	06-MAR-25	700.00
4	52647190	1	03-JAN-25	1200.00

Examines whether hosts with low review scores manage to improve their listings' performance over time.

20.Does the level of detail in guest reviews (e.g., long, descriptive reviews vs. short, generic ones) correlate with higher pricing or better value perceptions?

```
--20.Does the level of detail in guest reviews correlate with higher pricing or better value perceptions?
SELECT CASE
    WHEN C.PRICE > 1000 THEN 'High-Priced' WHEN C.PRICE <= 1000 THEN 'Low-Priced'
END AS PRICE_CATEGORY,
ROUND(AVG(LENGTH(R.COMMENTS)),0) AS AVG_REVIEW_LENGTH -- Calculate average length of comments
FROM LISTINGS L
JOIN CALENDAR C ON L.LISTING_ID = C.LISTING_ID
JOIN REVIEW R ON L.LISTING_ID = R.LISTING_ID
GROUP BY CASE
    WHEN C.PRICE > 1000 THEN 'High-Priced' WHEN C.PRICE <= 1000 THEN 'Low-Priced'
END
ORDER BY PRICE_CATEGORY;
```

Script Output x Query Result x

SQL | All Rows Fetched: 2 in 0.356 seconds

	PRICE_CATEGORY	AVG_REVIEW_LENGTH
1	High-Priced	240
2	Low-Priced	223

Investigates whether longer guest reviews are associated with higher-priced listings.

21. Are more detailed reviews associated with specific types of listings?

```
--21. Are more detailed reviews associated with specific types of listings?--
SELECT PROPERTY_TYPE, ROUND(AVG(LENGTH(COMMENTS)),0) AS AVG_REVIEW_LENGTH
FROM LISTINGS L
JOIN REVIEW R ON L.LISTING_ID = R.LISTING_ID
GROUP BY PROPERTY_TYPE
ORDER BY AVG_REVIEW_LENGTH DESC;
```

Query Result	
SQL   Fetched 50 rows in 0.399 seconds	
PROPERTY_TYPE	AVG_REVIEW_LENGTH
6 Private room in vacation home	284
7 Private room in townhouse	280
8 Entire home	269
9 Private room	267
10 Private room in condo	259

Examines if certain types of listings tend to receive more detailed guest reviews.

22. What criteria lead to repeat reservations and how frequently do customers return to the same listing?

```
--22. What criteria lead to repeat reservations and how frequently do customers return to the same listing?--
SELECT L.LISTING_ID, L.PROPERTY_TYPE, R.REVIEWER_ID, COUNT(*) AS repeat_reviews
FROM LISTINGS L
JOIN REVIEW R ON L.LISTING_ID = R.LISTING_ID
GROUP BY L.LISTING_ID, R.REVIEWER_ID, L.PROPERTY_TYPE
HAVING COUNT(*) > 1
ORDER BY REPEAT_REVIEWS DESC;
```

Script Output x Query Result x

SQL | Fetched 50 rows in 0.402 seconds

	LISTING_ID	PROPERTY_TYPE	REVIEWER_ID	REPEAT_REVIEWS
1	33465092	Entire home	520797183	19
2	40917418	Entire cottage	103523096	14
3	4981886	Entire rental unit	36085743	11
4	10653572	Entire rental unit	21253271	9

Looks at how often customers return to the same listing and what criteria (e.g., property type) lead to repeat bookings.

23. How does a guest's place of origin affect the kinds of listings (area, kind of accommodation, etc.) that they like? (Was unable to make a query for this)

24.How much does it cost to offer a property that can accommodate varied party sizes (e.g., families vs single travellers)?

```
--24.How much does it cost to offer a property that can accommodate varied party sizes (e.g., families vs single travellers)?--
```

```
SELECT L.ACCOMADATES, MIN(C.PRICE) AS MIN_PRICE, MAX(C.PRICE) AS MAX_PRICE
FROM LISTINGS L
JOIN CALENDAR C ON L.LISTING_ID = C.LISTING_ID
WHERE L.ACCOMADATES IN (1, 2, 4, 6, 10) -- Specify the party sizes you are interested in
GROUP BY L.ACCOMADATES
ORDER BY L.ACCOMADATES;
```

ript Output x Query Result x

SQL | All Rows Fetched: 5 in 0.024 seconds

	ACCOMADATES	MIN_PRICE	MAX_PRICE
1	1	1000.00	887.00
2	2	100.00	998.00
3	4	100.00	9999.00
4	6	1000.00	9500.00
5	10	10600.00	9900.00

Analyzes how the price varies for listings that accommodate different group sizes.

25.What changes have you noticed in the last year in Cape Town's Airbnb listing prices?

```
--25.What changes have you noticed in the last year in Cape Town's Airbnb listing prices?--
```

```
SELECT EXTRACT(MONTH FROM BOOKING_DATE) AS MONTH, ROUND(AVG(TO_NUMBER(PRICE)),0) AS AVG_PRICE --
FROM CALENDAR
WHERE BOOKING_DATE BETWEEN ADD_MONTHS(SYSDATE, -12) AND SYSDATE -- Use ADD_MONTHS for a year ago
GROUP BY EXTRACT(MONTH FROM BOOKING_DATE)
ORDER BY MONTH;
```

ript Output x Query Result x

SQL | All Rows Fetched: 5 in 0.004 seconds

	MONTH	AVG_PRICE
1	6	2105
2	7	2563
3	8	2979
4	9	2409
5	10	2361

Tracks changes in listing prices in Cape Town over the last 12 months.

26.Are there any obvious price trends that correspond with demand throughout certain seasons or with economic factors?

```
--26.Are there any obvious price trends that correspond with demand throughout certain seasons or with economic factors?--  
SELECT EXTRACT(MONTH FROM BOOKING_DATE) AS month, ROUND(AVG(TO_NUMBER(PRICE)),0) AS avg_price -- Convert PRICE from VARCHAR2 to NUMBER  
FROM CALENDAR  
GROUP BY EXTRACT(MONTH FROM BOOKING_DATE)  
ORDER BY month;
```

Script Output x Query Result x

SQL | All Rows Fetched: 12 in 0.007 seconds

	MONTH	AVG_PRICE
1	1	2614
2	2	2268
3	3	2520
4	4	2825
5	5	2428
6	6	2688
7	7	2563
8	8	2979
9	9	2409
10	10	2757
11	11	2630
12	12	2183

Investigates how price trends align with seasonal demand or economic factors.

## Performance optimizing queries

```
--INDEXING FOR PERFORMANCE  
CREATE INDEX IDX_LISTING_ID_LISTINGS ON LISTINGS (LISTING_ID);  
CREATE INDEX IDX_LISTING_ID_CALENDAR ON CALENDAR (cLISTING_ID);  
CREATE INDEX IDX_LISTING_ID_REVIEW ON REVIEW (LISTING_ID);  
CREATE INDEX IDX_PROPERTY_TYPE ON LISTINGS (PROPERTY_TYPE);  
  
--UPDATING PRICE COLUMN TO REMOVE DOLLAR SIGNS AND COMMAS  
UPDATE CALENDAR  
SET ADJUSTED_PRICE = REPLACE(REPLACE(ADJUSTED_PRICE, '$', ''), ',', '');
```

## 7. Analysis and Reporting

The aim of this analysis is to assess Airbnb accommodations in Cape Town in order to identify those that offer the most value for the price. Key research questions involve analysing the price distribution of listings with top-rated amenities, recognizing booking rate patterns, and investigating the influence of seasonal trends on pricing and availability. Furthermore, we investigated the impact of guest reviews on perceptions of value, the most favoured property types, and the neighbourhoods that draw the highest number of bookings. Using SQL queries and data visualizations, our goal is to provide useful insights for renters and hosts, allowing them to make informed decisions considering market dynamics. Graphs and tables:

### Question 1

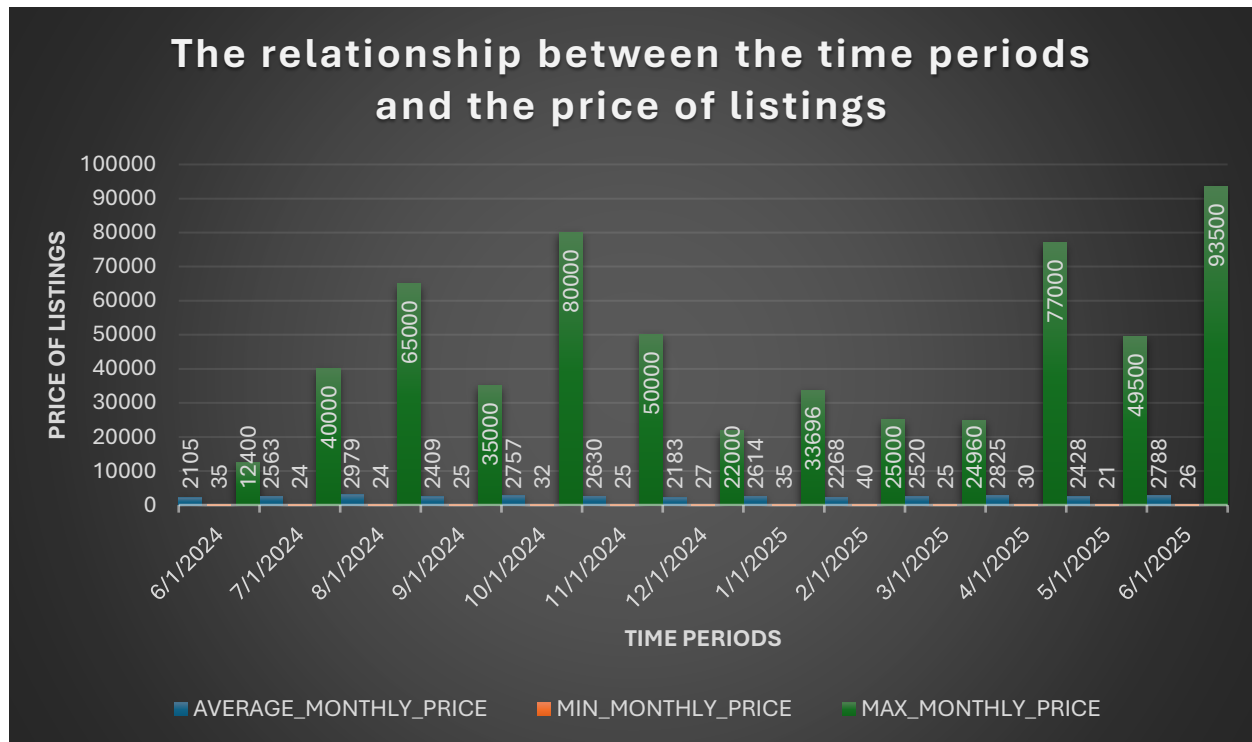
1. What is the price distribution for the listings with top rated amenities?



Based on the analysis of prices, most Airbnb listings with top-notch amenities are affordably priced, making them attractive options for budget-conscious visitors seeking quality. The decreased amount of listings in the more expensive price categories indicates that there is limited options for those willing to pay extra for top-tier services. This difference highlights opportunities for hosts to enhance their offerings in the mid-to-high price range, ultimately making them more attractive to visitors looking for the best value.

### Question 5

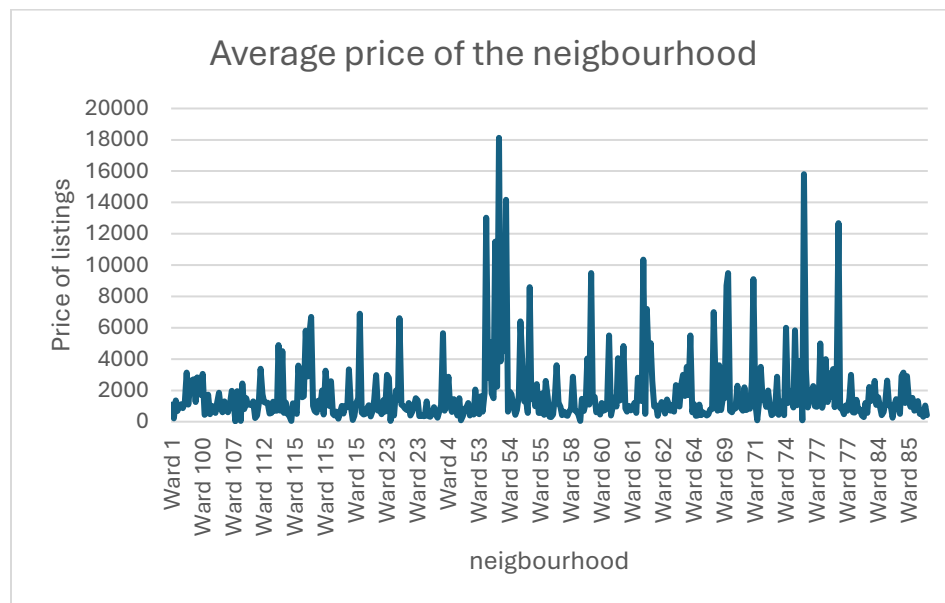
5. Are periods (certain days) when the price of listings is lower or higher?



The graph indicates that while the lowest and average prices have stayed steady, there has been fluctuation in the maximum prices, suggesting occasional spikes in high-end listings that significantly increase the top price. This could imply that the highest price is increased by premium or rare listings in certain months. In terms of getting the most out of your money, the consistent average prices indicate that the majority of listings offer dependable value, while occasional spikes in maximum prices present opportunities for wealthier tourists to explore luxurious choices without sacrificing overall value. Tenants can utilize this equilibrium to assist them in making informed choices according to their preferences and financial constraints.

### Question 8

8. What is the difference in pricing between listings for different types of rooms (private room, shared room, complete home/apartment, etc.) in each neighbourhood?



Certain wards, such as Ward 53 and Ward 74, boast notably higher average prices, whereas most other wards have lower prices. Differences in bar heights display the disparities in prices in various wards, highlighting regions with higher or lower-cost room offerings. In terms of getting the most value for your money, this study suggests that renters looking for affordability should look into neighbourhoods with lower average prices, as they may provide similar amenities and experiences at a lower cost. This chart assists in pinpointing pricing patterns and discrepancies among various regions, aiding in real estate evaluation and budgeting choices to help renters get the most out of their money.

### Question 10

10. What impact do reviews that were written within the previous month have on listing prices? Do freshly evaluated listings have a different price tag?

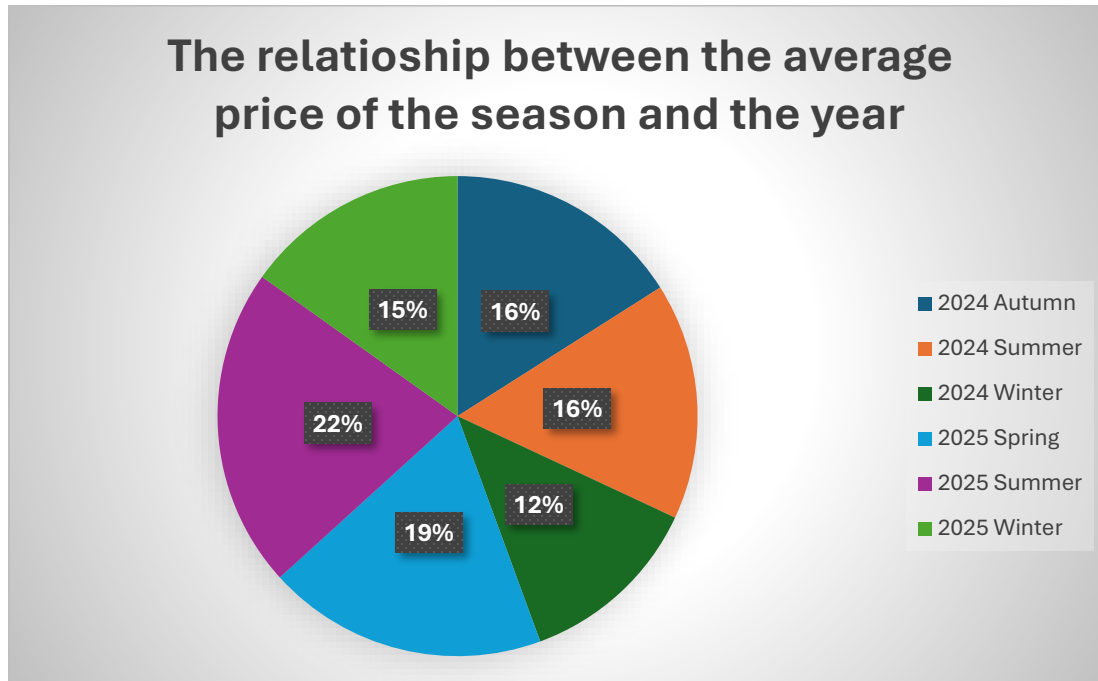
REVIEW_CATEGORY	AVERAGE_PRICE
1-5 Reviews	1784
6-10 Reviews	2293
More than 10 Reviews	1277

The evidence shows that accommodations with 1-5 reviews generally have a higher average cost than those with 6-10 or over 10 reviews. This pattern indicates that recent reviews can increase the perceived worth of listings and impact pricing changes. Listings that have received recent feedback could indicate higher quality or enhanced services, which can be appealing to customers looking for the best value.



### Question 11

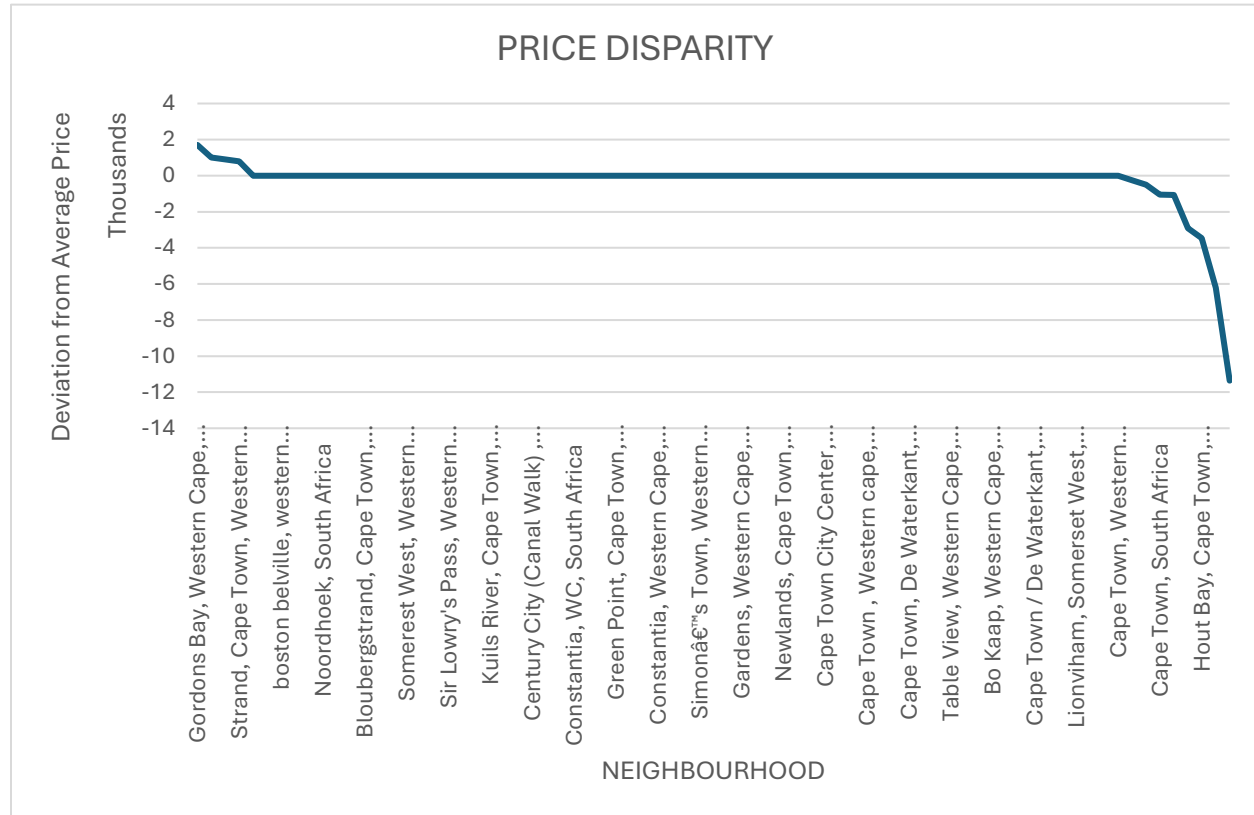
11.What seasonal differences exist in listing pricing, such as summer vs. winter? Are there particular seasons of the year when prices rise or fall noticeably?



In general, the chart shows fluctuations in average prices throughout the year, with prices usually increasing during the warmer months in comparison to winter. This pricing trend indicates that individuals might find more cost-effective options during low season, offering budget-conscious consumers the best value for their money.

### Question 17

17. Exist any neighbourhood's where the disparity in price is more noticeable?



The chart shows the difference in room prices in different neighbourhoods. The x-axis depicts various neighbourhoods, with the y-axis showing prices ranging from -15,000 to 15,000. This graphical display showcases the large differences in pricing across neighbourhoods, offering important information for renters seeking the most value based on location. Regions with lower costs could present more economical choices, whereas those with higher costs might signify high-end properties or desirable locales.

### Question 18

18. What effect do unfavourable reviews have on listing prices and occupancy rates?

REVIEW_TYPE	AVERAGE_PRICE	AVERAGE_OCCUPANCY_RATE
Favourable Reviews	2210.81	0.53
Unfavourable Reviews	615	0.5

The data compares average prices and occupancy rates based on review types. Listings with favorable reviews have a higher average price and a slightly higher occupancy rate than those with unfavorable reviews. This suggests that properties receiving positive feedback tend to attract more interest, allowing hosts to charge higher rates. Consequently, for renters seeking the best "bang for their buck," considering review types can be crucial, as listings with favorable reviews may provide a better overall value in terms of experience and pricing.

### Question 20

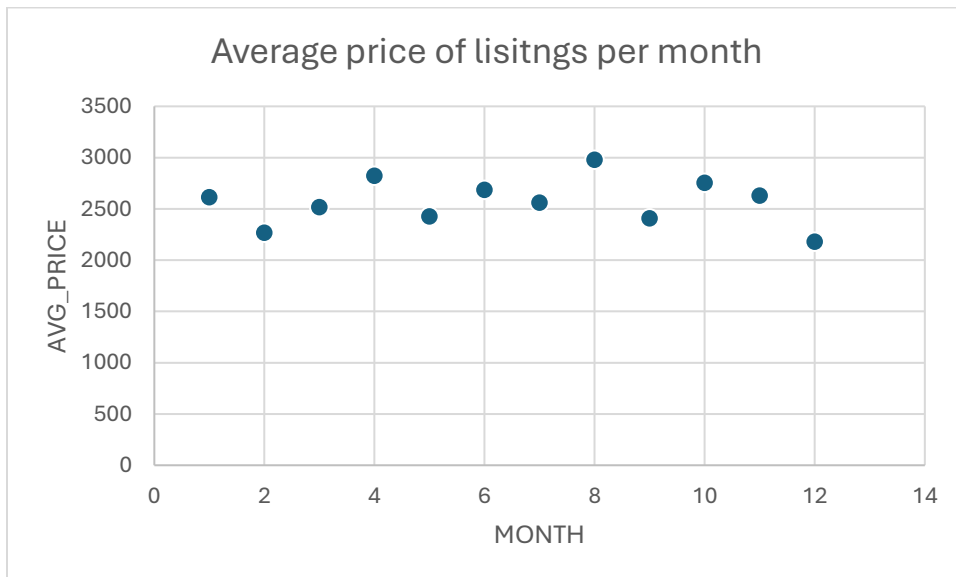
20. Does the level of detail in guest reviews (e.g., long, descriptive reviews vs. short, generic ones) correlate with higher pricing or better value perceptions?

PRICE_CATEGORY	REVIEW_LENGTH
High-Priced	240
Low-Priced	223

The information analyses mean prices and occupancy percentages according to different types of reviews. Listings that receive positive reviews tend to have a higher average price and slightly greater occupancy rates compared to those with negative reviews. This indicates that properties with positive feedback tend to draw more attention, enabling hosts to increase their fees. Hence, for tenants looking for the most value for their money, it's important to consider the types of reviews, as properties with positive feedback may offer a more worthwhile experience and pricing.

### Question 25

25. What changes have you noticed in the last year in Cape Town's Airbnb listing prices?



Accommodation costs increase during certain South African school breaks (March/April, June/July, September/October, and December/January) because more families go on vacation, driving up demand. Property owners take advantage of this by increasing prices, particularly for larger family-sized accommodations. December stands out from the rest, probably due to early deals, as numerous families choose it as their preferred time for a getaway. Identifying these patterns assists travellers in finding the most cost-effective choices and maximizing their decisions for better value during busy times.

## Final Report

### In conclusion

The goal of this Airbnb listings analysis in Cape Town was to pinpoint the ones offering the most bang for your buck by considering factors like pricing, booking habits, guest feedback, and seasonal fluctuations.

Our research shows that accommodations with highly rated features are usually priced reasonably, attracting budget-minded tourists. Nevertheless, there is a significant lack of luxury listings, indicating chances for hosts to improve offerings in the middle to high price bracket, ultimately drawing in more guests interested in getting great value for their money.

Price patterns show seasonal changes, with hotter months usually leading to higher prices because of a rise in demand. On the other hand, renting during non-peak times can offer more economic options, underlining the need for tenants to carefully schedule their visits for optimal savings. Furthermore, our investigation into variations in pricing between neighbourhoods highlights the importance of potential renters factoring in location as a key component in their assessment of value.

Furthermore, it was noted that recent reviews from guests have a positive impact on average prices, suggesting that properties with recent feedback are seen as more valuable. Properties that receive positive feedback not only can charge more money, but also have slightly higher occupancy rates, further demonstrating the connection between happy customers and successful rentals.

To sum up, our thorough examination provides useful information for both individuals looking to rent and Airbnb hosts about the market trends in Cape Town's short-term rental industry. Stakeholders can optimize their offerings and experiences in the competitive Airbnb market by comprehending pricing strategies, guest reviews impact, and seasonal trends.