Group 11
Sql_Admins

airbnb

Project Phase 1

(CMPG321)

Group Members:

38925958 – Michael Mathebula (Group Leader)[+27 62 217 3462]

34845577-TP Makgakga [+27 67 608 1802]

42132975 – O Zungu [+27 72 664 8856]

43706355 – D Saurombe [+27 67 733 6875]

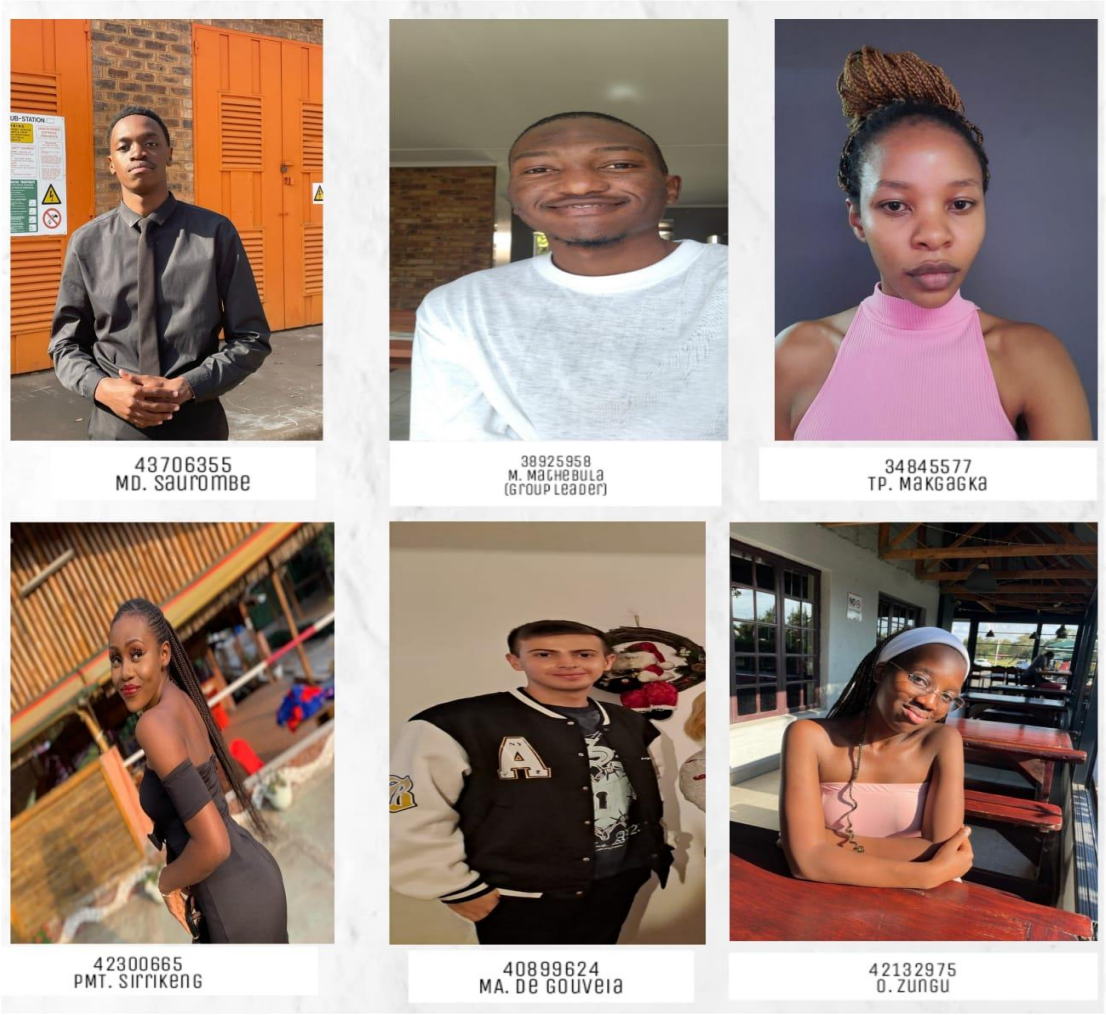
40899624 - M De Gouveia [+27 72 093 4903]

42300665 – P Sirrikeng [+27 84 575 1291]

1. CONTENTS

..... 2-1
Phase 1 – Focus Area and requirement analysis 2-1
2. Area of Focus 2-1
3. Research Questions 3-2
4. Dataset Description 4-5
Phase 2 4-9
5. Database Setup 5-9
6. ERD 6-11
6. SQL Query Development 6-12
7. Analysis and Reporting 6-19

Group Information



PHASE 1 – FOCUS AREA AND REQUIREMENT ANALYSIS

2. AREA OF FOCUS

Investment Opportunities: Determining the best areas for purchasing property based on Tourism Appeal, rental income, occupancy rates, and other factors:

Location

Tourism Appeal: Choose a location with a high volume of tourists or business travellers. Proximity to popular attractions, beaches, and scenic areas can significantly increase occupancy rates.

Crime rates: Take into consideration the crime rates in the surrounding areas

Market Research

Demand Analysis: Study the demand for short-term rentals in your chosen area. Look at occupancy rates, average daily rates (ADR), and seasonal trends. Most of this information is available on the Western Cape Government website.

Competition: Analyse existing Airbnb listings in the area. See what amenities they offer, how they're priced, and what their reviews say. Identify any gaps in the market you could fill.

Property Selection

Type of Property: Decide whether you want to invest in an apartment, a house, or a unique property (e.g., a treehouse, or boat). Different types attract different guests.

Condition of Property: Factor in renovation costs if the property needs updating. A well-maintained property will attract more bookings and better reviews.

Size: Consider the size of the property and the number of guests it can accommodate. Larger properties can command higher rates but may also come with higher maintenance costs.

Marketing

Listing Optimization: Write compelling descriptions, take professional photos, and highlight unique selling points.

Social media & Advertising: Promote your Airbnb on social media platforms, and consider running ads to reach a broader audience.

Customer Reviews: Encourage guests to leave positive reviews, as high ratings can significantly boost your property's visibility and desirability.

3. RESEARCH QUESTIONS

- Based on rental revenue vs property price, which areas with Airbnb are underperforming compared to their potential (property price), and what causes this?

    - This question highlight's locations of neighbourhoods of Airbnb's that are not performing based on the property value and the revenue made from guests booking with them. There may be many contributions to the underperformance such as seasonal bookings, competition, occupancy rate, etc, which can be determined. If the revenue for a certain time period is higher than the property price, this can be a good indicator of performance. This method can help identify those underperforming areas and the investor can make a decision based on that.

- Which neighbourhoods have the highest rental revenue from properties with the fewest reviews, suggesting that there may be little competition or great demand?
    - This research question helps identify neighbourhoods where despite the few reviews, the income is still high.

Generally, there are many reasons as to why an airbnb may have few reviews, such as booking in different channels, lack of encouragement from the host, or even the airbnb being at a high-demand location. This can also indicate that the prices for booking may be overcharged or there is a target of a very small clientele. Looking at these areas can help investors choose whether they can invest here or not, and also help find other areas that can bring a balance of clientele and "affordable" booking prices.

- What is the profit margin for Airbnb in this area?

  - (This question aims to determine the average profit margin for Airbnb's in Cape Town. The profit margin is a financial metric that indicates the percentage of revenue that turns into profit after all expenses are deducted. Specifically, it looks at how much profit an investor can make from investing in an Airbnb. To answer this question, we would have to analyze the occupancy rates, seasonality, fixed and variable costs, as well as the economic conditions of the country. Additional data will be collected from AirDNA, Airbnb or local property management companies. Understanding the competition in the area, typical pricing strategies, and average occupancy rates for similar properties will help the investor determine whether investing in an Airbnb is profitable or not.)

- Is investing in this area better than the other areas? Profit, amenities, tourist attractions, crime rates, activities?

  - (This question aims to evaluate whether the area we are considering for the Airbnb investment is more profitable or beneficial compared to other potential areas. The focus is on comparing the financial and possibly non-financial returns of investing in different locations. The goal is determining which area offers the best overall return on investment (ROI) for an Airbnb property. This question is dependent on the local economy and tourism, supply and demand for other Airbnb's in the surrounding areas, and, gathering the financial, market, and external factors data for each area, possibly using sources like real estate databases, market research reports, and tourism data. Most of the data can be found on the Western Cape Tourism website.

  Additionally, there are non-financial factors to consider, such as the surrounding tourism attraction, activities to do, guest preferences as well as local regulations.)

- Which listings in Cape Town offer the best value for money based on price, amenities, and location?

  - (This research question can be very valuable from an investor's perspective when considering investing in AirBnB properties in Cape Town. It can help in the following ways:

1. Identifying Profitable Opportunities:
   - By analyzing which listings offer the best value for money, investors can identify properties or areas that are currently undervalued or have high potential for return on investment.
2. Understanding Market Dynamics:

- The research will provide insights into what factors (price, amenities, location) are most important to guests in Cape Town, helping investors tailor their offerings to meet market demands.

3. Competitive Positioning:
    - Knowing what constitutes "best value for money" allows investors to position their properties competitively in the market.

4. Property Selection:

Investors can use the findings to guide their property selection process, focusing on areas or property types that tend to offer the best value for money.

5. Amenity Investment Decisions:

Understanding which amenities contribute most to perceived value can help investors make informed decisions about property improvements or renovations.)

- How does the quantity of reviews received by an Airbnb listing correlate with its average rating, and what are the underlying patterns that emerge from this relationship?

    - (Specifically, how does the average rating of Airbnb listings differ across various neighborhood groups within a city or region?)

- Furthermore, is there a statistically significant variance in average ratings between the different neighborhood groups, and how does this variance impact the frequency and volume of reviews each listing receives compared to one another?

    - (By analyzing the correlation between the number of reviews and several ratings, we can identify which Airbnb listing offers the best value for investors. By leveraging these insights, we can enhance property management for future investors by enabling investors to adopt the best practices to improve guest experiences. Additionally, understanding these patterns informs us of strategic pricing and marketing decisions, allowing investors to tailor their strategies for optimal performance.)

- What type of rooms do people book the most is it the Entire place, a Private room, or a shared room and why is it popular?

    - This question aims to identify the variety of rooms offered by the data. Room type indicates a preference for guests who frequent the Airbnb's in Cape Town and analysing this data will allow us to make useful decisions about what people desire in an Airbnb, therefore it is useful information to deduct what can be the best type for their money's worth and another advantage is investors will be able to use this information to carefully make their decisions as well. Information usable is room type, property type, price, and minimum/maximum night stays.

- What is the usual number of bedrooms, beds or bathrooms chosen?

    - This question aims to identify whether most travellers could be individuals or groups, therefore we can deduce if its friends, families or couples based on the number of beds and bathrooms selected. Understanding these preferences can assist investors in determining level of investment and influence their decisions on what travellers perceive as the best value for their money.
- What is the property type booked the most (Hotels or Bed and Breakfasts) and for how long?

    - Question is designed to identify the popular property types, if guests prefer hotels and Bed and breakfasts, from an investment view, it'll determine if a hotel is a better choice or an Airbnb.

- How do seasonal fluctuations affect Airbnb prices and the number of bookings in Cape Town(summer, winter, holidays)?

    - This question aims to help the investor understand the variation of profit during various seasons and help them make strategic business decisions when investing in AirBnBs in Cape Town. It also assists them in targeting their audience strategically during various seasons to maximise profit all year round and minimize loss.

- What is the impact of location on the pricing of Airbnb properties in South Africa?

    - This question aims to help both investors and customers. For example, Investors will analyze the variation in pricing of Airbnbs in Cape Town to evaluate which areas will generate more profit and the reasons why certain areas are more profitable than others.
      On the customer end, they will be able to evaluate which areas will give them the "BEST BANG FOR THEIR BUCK". This will also help them budget their spending with proper information.

4. DATASET DESCRIPTION

- **Room type**

  Data Type: Text

  Domain: Entire place, Private room, Shared room.

Description: Room offered in the property.

- **Property type**

  Data Type: Text

  Domain: Hotels and BnBs.

  Description: Type of property.

- **Bathrooms**

  Data type: numeric

  Domain: number of bathrooms

  Description: The number of bathrooms in the listing.

- **Bathroom text**: string

  Domain: number of bathrooms in text. E.g. (1 bath)

  Description: The number of bathrooms in the listing.

- **Bedrooms**

  Data type: integer

  Domain: number of bedrooms e.g. 1 to 10

  Description: The number of bedrooms in the listing.

- **Beds**

  Data type: integer

  Domain: number of beds e.g. 1 to 20.

  Description: The number of beds in the listing.

- **Price**

  Data type: Currency

  Domain: $635.00 to $1500

  Description: The rental price offered by properties.

- **Minimum nights**

  Data type: integer

  Domain: Minimum number of nights required for booking. E.g. ranging from 1 to 30

  Description: Minimum number of night stay for the listing.

- **Maximum nights**

  Data type: integer

  Domain: Maximum number of nights required for booking. E.g. ranging from 1 to 30

  Description: Maximum number of night stay for the listing.

- **Amenities**

  Data type: JSON
  Domain: Additional services
  Description: Additional services or facilities are provided to enhance the guest experience, such as Wi-Fi, swimming pools, fitness centers, and breakfast services**.**

- **Location**

  Data type: latitude/longitude
  Domain: Geographical location
  Description: Uses the World Geodetic System (WGS84) projection for latitude and longitude.

- **Neighbourhood**

  Data type: text
  Domain: Neighbourhood where the Airbnb is located
  Description: The neighborhood is geocoded using the latitude and longitude against neighborhoods as defined by open or public digital shapefiles.

- **Number of reviews**

  Data type: Integer

  Domain: Number of reviews that each Airbnb gets per month compared to other Airbnb's

  Description: The total amount of reviews each Airbnb has.

- **Reviews per month**
  Data type: numeric (2 decimal places)
  Domain: Comparing the number of reviews per month
  Description: Number of reviews each Airbnb receives per month

- **Date**
  Data type: date
  Domain: Date of each booking, review, and rating.
  Format: YYYY/MM/DD
  Description: Dates when every data about Airbnb is recorded into the system.

- **Listing ID**
  Data Type: integer, primary key
  Domain: The unique number given to each Airbnb
  Description: A unique number given to each owner of an Airbnb.

- **Airbnb ratings for each Airbnb**
  Data type: numeric
  Domain: Rating for each stay the user had in the past
  Description: The ratings of past stays people had when staying in a specific Airbnb.

- **Neighbourhood Group**
  Data type: text
  Domain: The location group each Airbnb is in
  Description: This is the group the Airbnb's are in the database.

- **Total revenue**
  Data type: currency
  Description: the generated revenue from each listing is from the price of booking for Airbnb.

- **Property price**
  Data type: currency
  Description: the value/price of the property in local currency.

PHASE 2

5. DATABASE SETUP

The following tables are created: LISTINGS, PROPERTIES, BOOKINGS, REVIEWS, & EXPENSES

LISTINGS/LISTINGZ TABLE

```
CREATE TABLE LISTINGS (
    listing_id INT PRIMARY KEY,
    listing_url VARCHAR2(255),
    title VARCHAR2(255),
    property_type VARCHAR2(50),
    listing_type VARCHAR2(50),
    last_scraped DATE,
    country VARCHAR2(50),
    city VARCHAR2(50),
    bedrooms INT,
    beds INT,
    bathrooms INT,
    max_guests INT,
    super_host CHAR(1),
    cancellation_policy VARCHAR2(50),
    cleaning_fee DECIMAL(10, 2),
    extra_people_fee DECIMAL(10, 2),
    check_in_time TIMESTAMP NULL,
    check_out_time TIMESTAMP NULL,
    minimum_nights INT,
    latitude NUMBER,
    longitude NUMBER,
    exact_location VARCHAR2(255),
    overall_rating NUMBER,
    cleanliness_rating NUMBER,
    communication_rating NUMBER,
    accuracy_rating NUMBER,
    check_in_rating NUMBER,
    location_rating NUMBER,
    value_rating NUMBER,
    amenities VARCHAR2(4000),
    license VARCHAR2(50),
    host_id VARCHAR2(255),
    host_name VARCHAR2(255),
    neighbourhood_group VARCHAR2(255),
    neighbourhood VARCHAR2(255),
    room_type VARCHAR2(50),
    price DECIMAL(10, 2),
    number_of_reviews INT,
    last_review DATE,
    reviews_per_month NUMBER,
    calculated_host_listing_count INT,
    availability_365 INT,
    number_of_reviews_ltm INT,
    neighbourhood_cleansed VARCHAR2(255),
    smart_location VARCHAR2(255),
    mean_price DECIMAL(10, 2),
    square_feet INT,
    guests_included INT
);
```

PROPERTIES TABLE

```
CREATE TABLE PROPERTIES (
    property_id INT PRIMARY KEY,
    area_type VARCHAR2(50),
    available_supply INT,
    avg_daily_rate DECIMAL(10, 2),
    occupancy_rate NUMBER,
    mean_stay_length INT,
    annual_revenue DECIMAL(10, 2),
    listing_id INT,
    CONSTRAINT fk_listing FOREIGN KEY (listing_id) REFERENCES LISTINGS(listing_id) ON DELETE CASCADE
);
```

REVIEWS TABLE

```
CREATE TABLE REVIEWS (
    listing_id INT,
    listing_date DATE,
    review_ratings INT,
    PRIMARY KEY (listing_id, listing_date),
    FOREIGN KEY (listing_id) REFERENCES LISTINGS(listing_id) ON DELETE CASCADE
);
```

BOOKINGS TABLE

Worksheet Query Builder

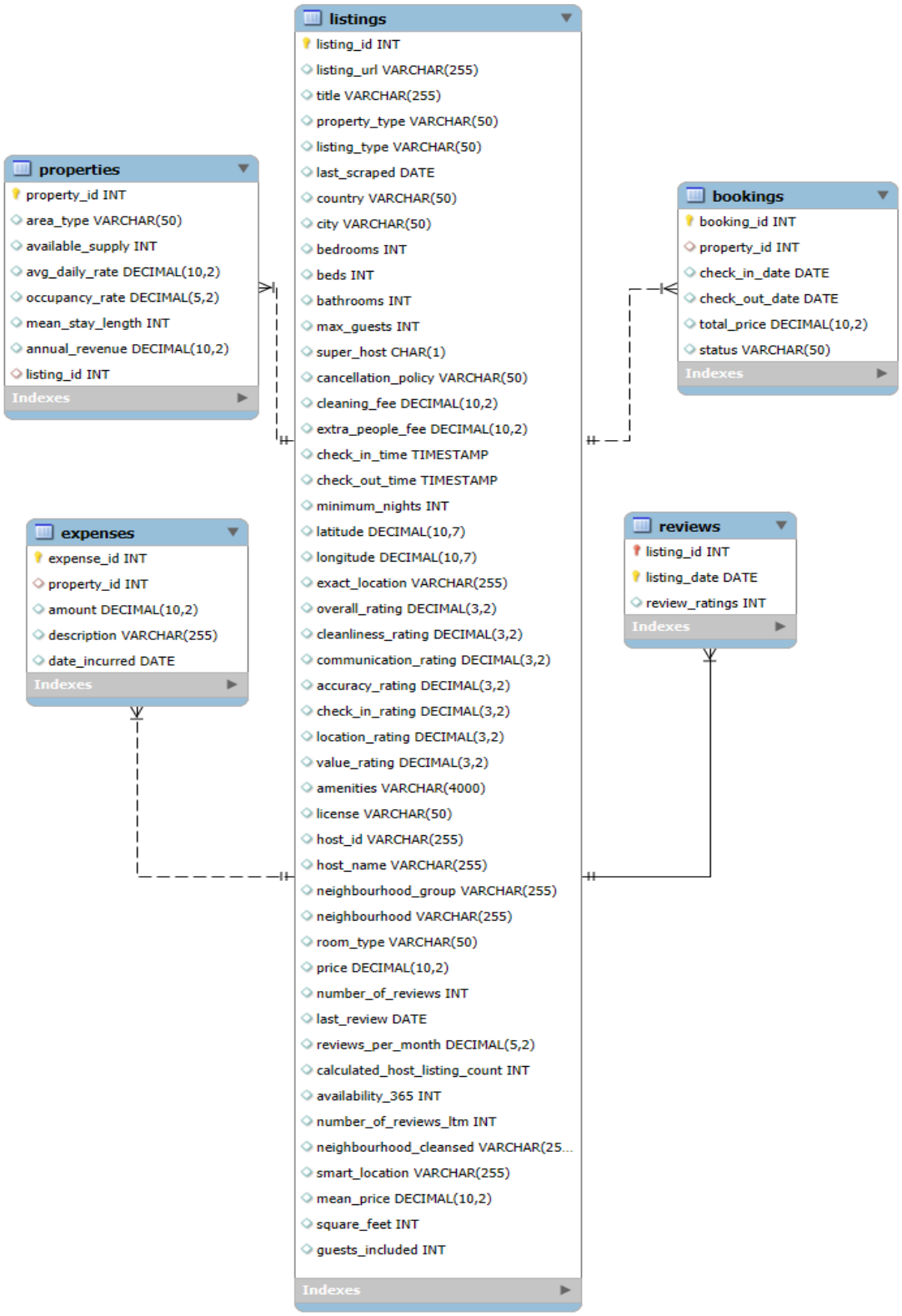
```
CREATE TABLE BOOKINGS (
    booking_id INT PRIMARY KEY,
    property_id INT,
    check_in_date DATE,
    check_out_date DATE,
    total_price DECIMAL(10, 2),
    status VARCHAR2(50),
    FOREIGN KEY (property_id) REFERENCES LISTINGS(listing_id) ON DELETE CASCADE
);
```

EXPENSES TABLE

Worksheet Query Builder

```
CREATE TABLE EXPENSES (
    expense_id INT PRIMARY KEY,
    property_id INT,
    amount DECIMAL(10, 2),
    description VARCHAR2(255),
    date_incurred DATE,
    FOREIGN KEY (property_id) REFERENCES LISTINGS(listing_id) ON DELETE CASCADE
);
```

6. ERD


listings
listing_id INT
listing_url VARCHAR(255)
title VARCHAR(255)
property_type VARCHAR(50)
listing_type VARCHAR(50)
last_scraped DATE
country VARCHAR(50)
city VARCHAR(50)
bedrooms INT
beds INT
bathrooms INT
max_guests INT
super_host CHAR(1)
cancellation_policy VARCHAR(50)
cleaning_fee DECIMAL(10,2)
extra_people_fee DECIMAL(10,2)
check_in_time TIMESTAMP
check_out_time TIMESTAMP
minimum_nights INT
latitude DECIMAL(10,7)
longitude DECIMAL(10,7)
exact_location VARCHAR(255)
overall_rating DECIMAL(3,2)
cleanliness_rating DECIMAL(3,2)
communication_rating DECIMAL(3,2)
accuracy_rating DECIMAL(3,2)
check_in_rating DECIMAL(3,2)
location_rating DECIMAL(3,2)
value_rating DECIMAL(3,2)
amenities VARCHAR(4000)
license VARCHAR(50)
host_id VARCHAR(255)
host_name VARCHAR(255)
neighbourhood_group VARCHAR(255)
neighbourhood VARCHAR(255)
room_type VARCHAR(50)
price DECIMAL(10,2)
number_of_reviews INT
last_review DATE
reviews_per_month DECIMAL(5,2)
calculated_host_listing_count INT
availability_365 INT
number_of_reviews_ltm INT
neighbourhood_cleansed VARCHAR(25...
smart_location VARCHAR(255)
mean_price DECIMAL(10,2)
square_feet INT
guests_included INT
Indexes
properties
property_id INT
area_type VARCHAR(50)
available_supply INT
avg_daily_rate DECIMAL(10,2)
occupancy_rate DECIMAL(5,2)
mean_stay_length INT
annual_revenue DECIMAL(10,2)
listing_id INT
Indexes
bookings
booking_id INT
property_id INT
check_in_date DATE
check_out_date DATE
total_price DECIMAL(10,2)
status VARCHAR(50)
Indexes
expenses
expense_id INT
property_id INT
amount DECIMAL(10,2)
description VARCHAR(255)
date_incurred DATE
Indexes
reviews
listing_id INT
listing_date DATE
review_ratings INT
Indexes

6. SQL Query Development

The SQL queries below answer the research questions given in phase 1 above.

- **Based on rental revenue vs property price, which neighbourhoods with airbnbs are underperforming compared to their potential rate, and what factors cause this?**

```sql
SELECT
    l.neighbourhood,
    AVG(p.annual_revenue) AS avg_revenue,
    (AVG(p.annual_revenue)) / AVG(p.avg_daily_rate) AS revenue_to_price_ratio
FROM
    listings l
JOIN
    properties p ON l.listing_id = p.listing_id
GROUP BY
    l.neighbourhood
HAVING
    (AVG(p.annual_revenue)) / AVG(p.avg_daily_rate) < (
        SELECT
            AVG(p2.annual_revenue) / AVG(p2.avg_daily_rate)
        FROM
            properties p2
            )
ORDER BY
    revenue_to_price_ratio;
```

- **Which neighbourhoods have the highest rental revenue from properties with the fewest reviews, suggesting that there may be little competition or great demand?**

```sql
SELECT
    l.neighbourhood AS neighbourhoods,
    AVG(l.number_of_reviews_ltm) AS avg_reviews,
    AVG(p.annual_revenue) AS avg_revenue
FROM
    listings l
JOIN
    properties p ON l.listing_id = p.listing_id
GROUP BY
    l.neighbourhood
HAVING
    AVG(l.number_of_reviews_ltm) < 12
ORDER BY
    avg_revenue DESC;
```

- **How do seasonal fluctuations affect Airbnb prices and the number of bookings in Cape Town(summer , winter ,holidays)?**

```sql
SELECT
    CASE
        WHEN month IN (12, 01, 02) THEN 'Summer'
        WHEN month IN (06, 07, 08) THEN 'Winter'
        WHEN month IN (03, 04, 05, 09, 10, 11) THEN 'Off-Season'
        ELSE 'Holiday'
    END AS season,
    AVG(price) AS average_price
FROM
    listings,bookings
GROUP BY
    CASE
        WHEN month IN (12, 01, 02) THEN 'Summer'
        WHEN month IN (06, 07, 08) THEN 'Winter'
        WHEN month IN (03, 04, 05, 09, 10, 11) THEN 'Off-Season'
        ELSE 'Holiday'
    END
ORDER BY
    season;
```

- **What is the impact of location on the pricing of Airbnb properties in South Africa?**

```sql
SELECT
     neighbourhood,
     AVG(price) AS average_price

FROM
     listings
     WHERE price IS NOT NULL
GROUP BY
     neighbourhood
ORDER BY
     neighbourhood ;
```

- **What is the impact of location on the pricing of Airbnb properties in South Africa?**

```sql
SELECT
     neighbourhood,
     AVG(price) AS average_price

FROM
     listings
     WHERE price IS NOT NULL
GROUP BY
     neighbourhood
ORDER BY
     neighbourhood ;
```

- **What is the profit margin for Airbnb in this area?**

```sql
SELECT
    l.exact_location AS location,
    SUM(b.total_price) AS total_revenue,
    COALESCE(SUM(e.amount), 0) AS total_expenses,
    (SUM(b.total_price) - COALESCE(SUM(e.amount), 0)) AS total_profit,
    ((SUM(b.total_price) - COALESCE(SUM(e.amount), 0)) / NULLIF(SUM(b.total_price), 0)) * 100 AS profit_margin
FROM
    BOOKINGS b
JOIN
    LISTINGS l ON b.property_id = l.listing_id  -- Adjusted to use listing_id
LEFT JOIN
    EXPENSES e ON l.listing_id = e.property_id  -- Adjusted to join on listing_id
WHERE
    b.check_in_date >= TO_DATE('2023-01-01', 'YYYY-MM-DD')
    AND b.check_out_date <= TO_DATE('2023-12-31', 'YYYY-MM-DD')
    AND b.status = 'completed'
GROUP BY
    l.exact_location;
```

- **Is investing in one area/ward better than the other areas? Profit, amenities, tourist attractions, crime rates, activities?**

```sql
SELECT
    neighbourhood_cleansed,
    AVG(mean_price) AS avg_price
FROM
    LISTINGS
GROUP BY
    neighbourhood_cleansed
ORDER BY
    avg_price ASC;
```

- **Which listings in Cape Town offer the best value for money based on price, amenities, and location?**

This query finds the listings that offer the best price relative to the number of amenities offered.

(NOTE: used in another DBMS. (sql lite))

```sql
SELECT id, name, price, neighbourhood, LENGTH(Amenities) - LENGTH(REPLACE(Amenities, ',', ''))
+ 1 AS num_amenities,
       price / (LENGTH(Amenities) - LENGTH(REPLACE(Amenities, ',', '')) + 1) AS price_per_amenity
FROM listings_with_amenities
ORDER BY price_per_amenity ASC
LIMIT 10;
```

This query targets listings in specific desirable areas(e.g., Ward 61) that offer value for money.

```sql
SELECT id, name, price, neighbourhood, Amenities
FROM listings_with_amenities
WHERE neighbourhood = 'Ward 61'
ORDER BY price ASC
LIMIT 10;
```

This query retrieves listings that offer many amenities but remain affordable.

```sql
SELECT id, name, price, LENGTH(Amenities) - LENGTH(REPLACE(Amenities, ',', '')) + 1 AS num_amenities
FROM listings_with_amenities
ORDER BY num_amenities DESC, price ASC
LIMIT 10;
```

This query groups listings by neighbourhood and finds the top listings with the best price per amenity in each area.

```sql
SELECT id, name, price, neighbourhood, Amenities,
       price / (LENGTH(Amenities) - LENGTH(REPLACE(Amenities, ',', '')) + 1) AS price_per_amenity
FROM listings_with_amenities
GROUP BY neighbourhood, id
ORDER BY neighbourhood, price_per_amenity ASC
LIMIT 10;
```

To find the listings near a specific location(like the City Center), use latitude and longitude. This query assumes a certain location(latitude = -33.92528, longitude = 18.42022) is the reference point.

```sql
SELECT id, name, price, neighbourhood, latitude, longitude,
       SQRT(POWER(latitude - (-33.92528), 2) + POWER(longitude - (18.42022), 2)) AS distance_from_center
FROM listings_with_amenities
ORDER BY distance_from_center ASC, price ASC
LIMIT 10;
```

This query filters listings with a set of popular amenities(e.g., Wi-Fi, Air Conditioning and Parking) and ranks them by price.

```sql
SELECT id, name, price, neighbourhood, Amenities
FROM listings_with_amenities
WHERE Amenities LIKE '%Wi-Fi%' AND Amenities LIKE '%Air Conditioning%' AND Amenities LIKE '%Parking%'
ORDER BY price ASC
LIMIT 10;
```

This query calculates the average price of listings in each neighbourhood, helping you to compare the relative affordability of different areas.

```
SELECT neighbourhood, AVG(price) AS avg_price
FROM listings_with_amenities
GROUP BY neighbourhood
ORDER BY avg_price ASC;
```

This query helps us to find listings that are priced below a specific value and still offer good amenities.

```
SELECT id, name, price, neighbourhood, LENGTH(Amenities) - LENGTH(REPLACE(Amenities, ',', '')) + 1 AS num_amenities
FROM listings_with_amenities
WHERE price < 1000
ORDER BY num_amenities DESC, price ASC
LIMIT 10;
```

- **What type of rooms do people book the most is it the Entire place, a Private room, or a shared room and why is it popular?**

```
SELECT ROOM_TYPE,
COUNT (*) AS NUMOFBOOKINGS
FROM LISTINGZ
GROUP BY ROOM_TYPE
ORDER BY NUMOFBOOKINGS DESC;
```

View and sequence created for the above sql query:

```
CREATE VIEW RoomTypeCount AS
SELECT
    ROW_NUMBER() OVER (ORDER BY COUNT(*) DESC) AS Sequence,
    ROOM_TYPE,
    COUNT(*) AS RoomCount
FROM
    LISTINGZ
GROUP BY
    ROOM_TYPE
ORDER BY
    COUNT(*) DESC;
```

Access the view as: **SELECT * FROM RoomTypeCount;**

The sequence is a ranking of room types based on the number of listings, and it provides a quick way to understand which room types are most prevalent.

- **What is the usual number of bedrooms, beds or bathrooms chosen?**

For the number of bedrooms:

```sql
SELECT DISTINCT
     property_type,Bedrooms,
    COUNT(*) AS NumberOfListings
FROM
    LISTINGS
GROUP BY
    Bedrooms, property_type
ORDER BY
    Bedrooms;
```

```sql
CREATE VIEW PropertyTypeBedroomCount AS
SELECT DISTINCT
    property_type,
    Bedrooms,
    COUNT(*) AS NumberOfListings
FROM
    LISTINGS
GROUP BY
    Bedrooms, property_type
ORDER BY
    Bedrooms;
```

Access view via: **SELECT * FROM PropertyTypeBedroomCount**;

The SQL query aims to find the most common number of bedrooms for each room type in the LISTINGZ table. It does this by counting how many times each combination of room type and number of bedrooms appears, then ranks the results for each room type. Finally, it selects the room type and corresponding number of bedrooms with the highest occurrence (ranked 1) for each room type.

```sql
WITH RankedBedrooms AS (
    SELECT
        ROOM_TYPE,
        BEDROOMS,
        COUNT(*) AS NUMOFOCCURRENCES,
        ROW_NUMBER() OVER (PARTITION BY ROOM_TYPE ORDER BY COUNT(*) DESC) AS rn
    FROM
        LISTINGZ
    GROUP BY
        ROOM_TYPE,
        BEDROOMS
)
SELECT
    ROOM_TYPE,
    BEDROOMS,
    NUMOFOCCURRENCES
FROM
    RankedBedrooms
WHERE
    rn = 1;
```

This SQL query retrieves the top 300 combinations of bathroom count and property type from the LISTINGZ table, ordered by the number of bookings. It groups the results by the number of bathrooms and property type, counts how many times each combination appears, and limits the output to the top 300 rows based on the highest number of bookings (NUMOFBOOKINGS).

```
SELECT * FROM (
    SELECT
        BATHROOMS,
        PROPERTY_TYPE,
        COUNT(*) AS NUMOFBOOKINGS
    FROM
        LISTINGZ
    GROUP BY
        BATHROOMS, PROPERTY_TYPE
    ORDER BY
        NUMOFBOOKINGS DESC
)
WHERE ROWNUM <= 300;
```

- **What is the property type booked the most (Hotels or Bed and Breakfasts) and for how long?**

```
SELECT DISTINCT
    PROPERTY_TYPE,
    MAX(MAXIMUM_NIGHTS - MINIMUM_NIGHTS + 1) AS MAXSTAYDURATION
FROM
    LISTINGZ
WHERE
    MAXIMUM_NIGHTS - MINIMUM_NIGHTS + 1 <= 365
GROUP BY
    PROPERTY_TYPE
ORDER BY
    MAXSTAYDURATION DESC;
```

View is:

```
CREATE VIEW PropertyStayDuration AS
SELECT DISTINCT
    PROPERTY_TYPE,
    MAX(MAXIMUM_NIGHTS - MINIMUM_NIGHTS + 1) AS MAXSTAYDURATION
FROM
    LISTINGZ
WHERE
    MAXIMUM_NIGHTS - MINIMUM_NIGHTS + 1 <= 365
GROUP BY
    PROPERTY_TYPE
ORDER BY
    MAXSTAYDURATION DESC;
```

Access view: **SELECT * FROM PropertyStayDuration;**

- **How does the quantity of reviews received by an Airbnb listing correlate with its average rating, and what are the underlying patterns that emerge from this relationship? Specifically, how does the average rating of Airbnb listings differ across various neighbourhood groups within a city or region?**

```
SELECT l.neighbourhood,
       AVG(listing_avg.avg_rating) AS avg_neighbourhood_rating,
       SUM(l.number_of_reviews) AS total_reviews
FROM (
    SELECT r.listing_id, AVG(r.review_ratings) AS avg_rating
    FROM reviews r
    GROUP BY r.listing_id
) listing_avg
JOIN listings l ON listing_avg.listing_id = l.id
GROUP BY l.neighbourhood
ORDER BY total_reviews DESC;
```

```
SELECT r.listing_id, AVG(r.review_ratings) AS avg_ratings, l.number_of_reviews,
l.neighbourhood, l.price, l.availability_365, l.minimum_nights,
FROM reviews r
JOIN listings l ON r.listing_id = l.id
WHERE l.number_of_reviews > 50
GROUP BY r.listing_id, l.number_of_reviews, l.neighbourhood, l.price, l.availability_365, l.minimum_nights,
HAVING AVG(r.review_ratings) >= 3
ORDER BY l.number_of_reviews DESC;
```

7. Analysis and Reporting

The following analysis is done after answering each research question with SQL queries.

- **Based on rental revenue vs property price, which neighbourhoods with airbnbs are underperforming compared to their potential rate, and what factors cause this?**

NEIGHBOURHOOD	REVENUE AVERAGE (R)	REVENUE_TO_PRICE_RATIO (R)
Neighborhood B	15000	146,34
Neighborhood G	20000	195,12
Neighborhood E	42500	236,11
Neighborhood D	40000	242,42
Neighborhood I	39500	243,08

The table shown above are the neighbourhoods that have a lower revenue to price ratio than the average revenue to price ratio, which can be a good indicator of underperformance. If an investor has to choose a good neighbourhood where the

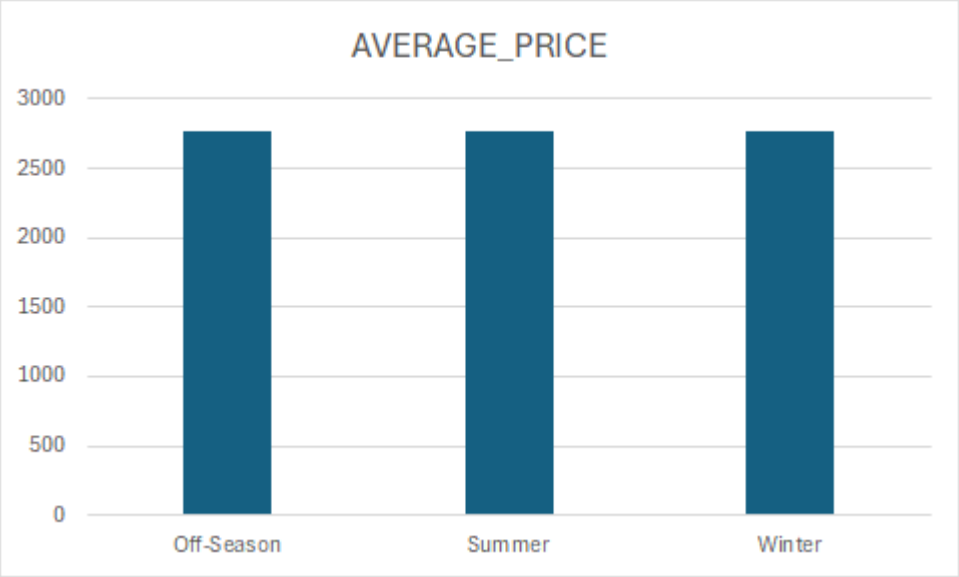
revenue to price ratio is above the average, they should not choose the ones listed above.

- **Which neighbourhoods have the highest rental revenue from properties with the fewest reviews, suggesting that there may be little competition or great demand?**

NEIGHBOURHOOD	AVERAGE NUMBER OF REVIEWS	AVERAGE REVENUE (R)
Neighborhood A	10	45000
Neighborhood C	8	47000

The table above shows the neighbourhoods that have a lower average of reviews yet have a higher value than its average. The factors of low reviews may include not booking on the official Airbnb app, booking through third-party apps, or booking face-to-face, and many more. These neighbourhoods are however good places because of their revenues generated on average.

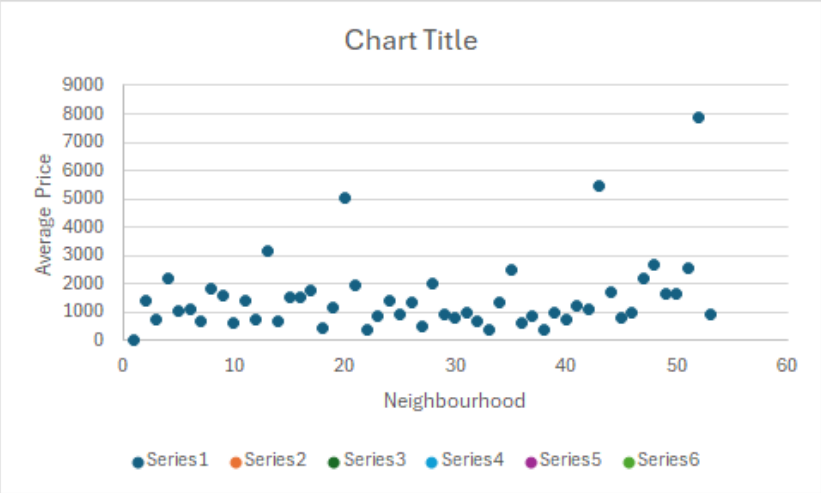
- **How do seasonal fluctuations affect Airbnb prices and the number of bookings in Cape Town (summer, winter, holidays)?**

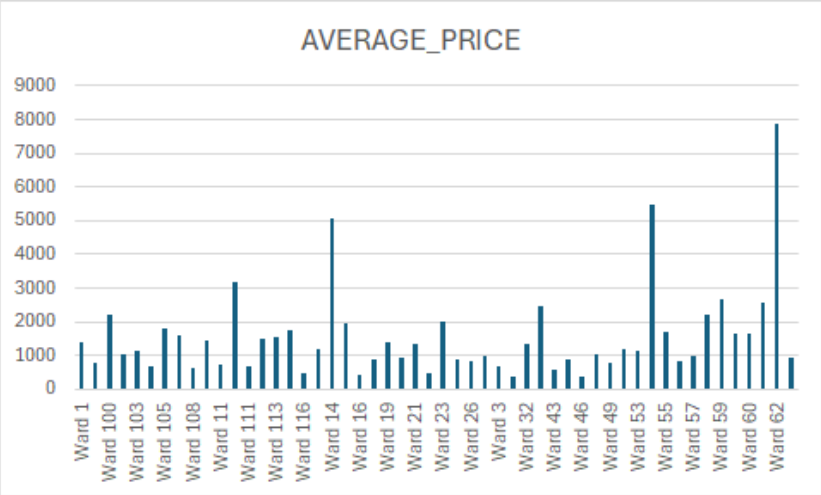


The average price remains constant over all seasons, this indicates that Airbnb prices do not fluctuate per season. This information provides insight to investors to understand how the profit margin for Airbnbs are distributed over the year which means there are no losses incurred just based on seasonal fluctuations.

- **What is the impact of location on the pricing of Airbnb properties in South Africa?**
  It is shown from the data that the average price of Airbnbs vary by location.

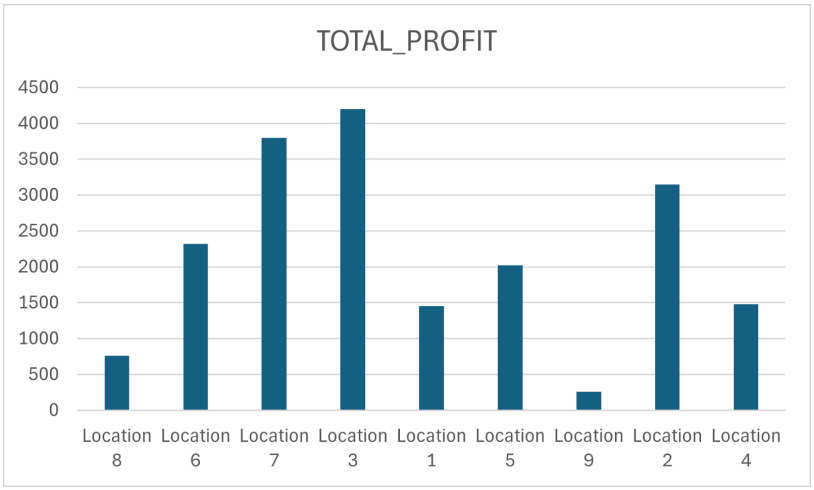
The prices of various Airbnbs in the same location are sparsely distributed from very high to very low, which indicates that location is not a deciding factor for the price of an Airbnb.
The most expensive Airbnbs are located in Ward 62 ,54,14 while the least expensive are located in Ward 46,4,22,16,116.





- **What is the profit margin for Airbnb in this area, & Is investing in one area/ward better than the other areas? Profit, amenities, tourist attractions, crime rates, activities??**

  Based on the sql queries above, this code will generate the wards with the highest and lowest prices within the wards. Hout Bay in Ward 74 and Camps Bay in Ward 54 are the wards with the most expensive listings. Kuilrivier has the lowest listings overall.



- Regarding the profit, this is dependent on many variables, some of which are qualitative. To ensure that expenses are accounted for, I generated general expenses to show that areas with the lowest expenses such as maintenance, wages and salaries would generally make more profit.
- However, it is important to note that Airbnb in areas that have access to public transportation, low crime rates, high ratings, and those in close proximity to tourist attractions are expected to generate more profit.

The second bar graph explains the difference between expense and the general profit of an airbnb. Based on the graph, areas in certain locations are operating at a loss, however, this could also depend on seasonality as well as the country's economic conditions.

- **Which listings in Cape Town offer the best value for money based on price, amenities, and location?**

This data analysis explained the results obtained when Airbnb listings in Cape Town where evaluated based on three key factors:

Price: The cost per night in South African Rand (ZAR).

Amenities: The number and type of amenities each listing offers.

Location: How the neighbourhood or proximity to a central point affects value.

The goal was to find out listings that offer the best value for money by balancing affordability with features and convenience.

Best Listings by Price-to Amenity Ratio

Key Findings:

- Listings with a higher number of amenities and lower price per amenity score highest
- Some top listings in this category:
    - Malleson Garden Cottage: ZAR 635 with 5 amenities (Wi-Fi, Kitchen, TV, Parking and Heating).
    - Stunning Sea View apt: ZAR 1500 with 4 amenities (Wi-Fi, Pool, Ocean view and Air Conditioning).

Best Listings In Prime Neighbourhoods

Neighbourhoods like ward 61(Cape Peninsula) tend to attract tourists due to their scenic views and proximity to popular attractions.

Key Listings in Ward 61:

- Stunning Sea View apt priced at R1500 stands out as offering both affordability and a prime location

Cheapest Listings

Focusing purely on cost, the cheapest listings provide a baseline for budget-conscious travellers.

Top 3 Cheapest Listings:

1. Malleson Garden Cottage- R635
2. Blaauwberg House- R3102(with inverter)
3. CityChic @the Piazza -R2000

While lower priced, some of these listings still offer core amenities such as Wi-Fi and kitchen facilities making them good value.

Listings with Proximity to City Center

Coordinates were used to assess listings proximity to the Cape Town city centre and found that:

- Listings near the city centre(e.g., CityChic @the Piazza) offer good access to central attractions at R2000 per night.

Listings farther from the city centre but still affordable, such as Malleson Garden Cottage in ward 57 provide a balance of peaceful surroundings and essential amenities.

Luxury Listings

For travellers willing to spend more, listings like Blaauwberg House offer luxury features at higher prices

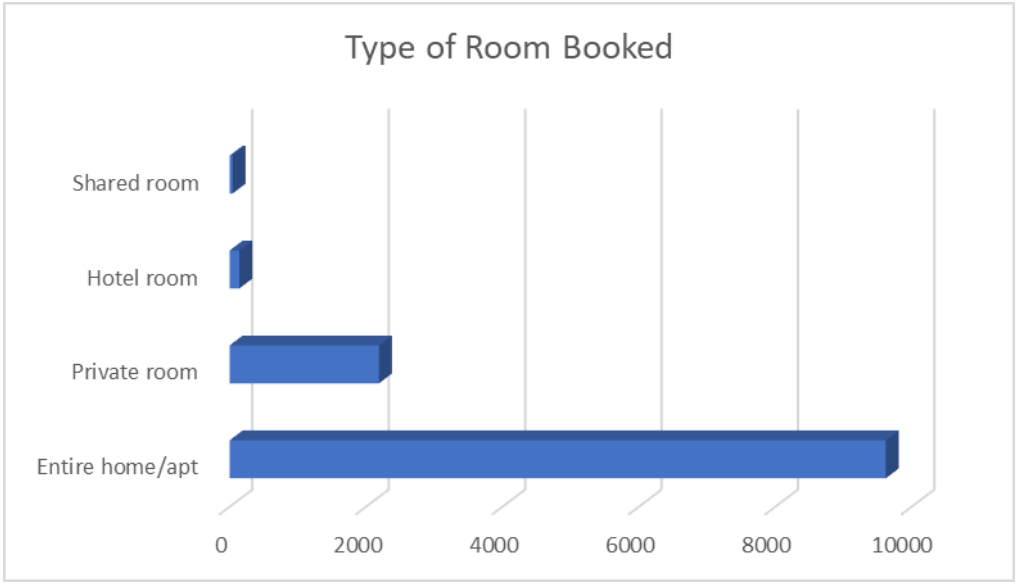
Most expensive Listings:

- Blaauwberg House: R3102
- Grande Bay: R2000

These listings offer high-end features like ocean views, pools and large living spaces.

Based on this analysis, the best value for money listings are those that offer a high number of amenities at affordable prices, especially in prime locations like Ward 61 or those close to the city centre. Listings like Malleson Garden Cottage and Stunning View apt strike a good balance between price, location and the range of amenities offered. For budget-conscious travellers. Listings like Malleson Garden Cottage(R635) are great choices, while those seeking luxury with more amenities should consider Blaauwberg House.

- **What type of rooms do people book the most is it the Entire place, a Private room, or a shared room and why is it popular?**



Explanation:

- ROOM_TYPE is the column indicating the type of room.
- COUNT(*) counts the number of bookings for each room type.
- GROUP BY ROOM_TYPE groups the results by room type.
- ORDER BY NUMOFBOOKINGS DESC orders the results to show the most popular room type at the top.

- **What is the usual number of bedrooms, beds or bathrooms chosen?**

  The query begins with a CTE called **RankedBedrooms**. This temporary result set is used to organize and rank data before the final selection.
  Inside the CTE:
  - Grouping: It groups the listings based on two criteria: ROOM_TYPE (such as "Entire home/apt" or "Hotel room") and BEDROOMS (the number of bedrooms).
  - Counting Occurrences: For each combination of room type and bedroom count, it counts how many listings exist. This is achieved with the COUNT(*) AS NUMOFOCCURRENCES statement.
  - Ranking: The ROW_NUMBER() function assigns a sequential integer (referred to as rn) to each row within each ROOM_TYPE group, ordered by the number of occurrences in descending order. This means the most frequent combination gets the rank of 1.

Main Query:

- The main part of the query selects specific columns from the RankedBedrooms CTE:
  - ROOM_TYPE: The type of accommodation.
  - BEDROOMS: The count of bedrooms in that type.
  - NUMOFOCCURRENCES: The total number of listings for that combination.
- The filter WHERE rn = 1 ensures that only the most common bedroom count for each room type is returned, eliminating any less frequent combinations.

Resulting Data:

- The output of the query shows:
  - Entire home/apt:
    - Bedrooms: 1
    - Occurrences: **3351** listings, indicating that one-bedroom entire homes or apartments are the most prevalent in the dataset.
  - Hotel room:
    - Bedrooms: 1
    - Occurrences: **110**, showing that single-bedroom hotel rooms are relatively fewer in comparison.
  - Private room:
    - Bedrooms: 1
    - Occurrences: **1468**, suggesting a significant number of listings for private rooms with one bedroom.
  - Shared room:
    - Bedrooms: 1
    - Occurrences: **31**, indicating that shared rooms are the least common option in the dataset.

The analysis highlights the most common configurations of bedrooms across different room types. For instance, the dominance of one-bedroom listings in "Entire home/apt" suggests a strong demand for compact accommodations. This information can be invaluable for property investors and managers, as it indicates market trends and potential areas for expansion or investment in one-bedroom properties.

The booking data provides valuable insights into customer preferences in the property rental market, highlighting trends in property types and their corresponding booking volumes.

Key Findings:

1. Top Property Types:
    - Entire Rental Units: Leading with 4,013 bookings, this category features an average of 6.5 bathrooms, indicating high demand for spacious accommodations.
    - Entire Homes: With 2,385 bookings and a maximum of 16 bathrooms, these properties cater to larger groups, emphasizing the importance of multiple bathroom facilities for convenience.
2. Mid-Tier Offerings:
    - Private Room in Home: Securing 822 bookings, this property type showcases a significant interest in more intimate settings, with an average of 8.5 bathrooms.
    - Entire Guest Suite: At 775 bookings, this option continues to attract customers looking for comfort without the need for full property rentals.
3. Smaller Properties:
    - Entire Condos and Villas: Both categories show solid performance, with 613 and 474 bookings, respectively, demonstrating a market for upscale options with amenities.
    - Private Rooms in Guesthouses and Rentals: These listings, with bookings of 314 and 306, illustrate the viability of smaller accommodations for budget-conscious travelers.
4. Niche Offerings:
    - Notably, the Private Room in Bed and Breakfast features 38 bathrooms but only 211 bookings, indicating that while large bathroom counts can be a unique selling point, they may not always translate to high demand.

Implications for Investors and Property Management:

Identifying Popular Property Types:

- Travelers can observe that "Entire rental units," with 4,013 bookings, are the most frequently booked option, indicating high quality and demand. Opting for such properties may enhance their overall experience.

Understanding Amenities:

- The average number of bathrooms in popular listings, like "Entire homes" with 16 bathrooms and 2,385 bookings, suggests that amenities significantly impact comfort. Customers should prioritize properties with sufficient bathrooms for their needs, especially when traveling in groups.

Tailoring Choices to Group Size:

- The data allows travelers to match their property choices to group size. For example, "Private room in home," with 822 bookings, is suitable for solo travelers or couples, while larger groups might prefer "Entire rental units."

Exploring Unique Offerings:

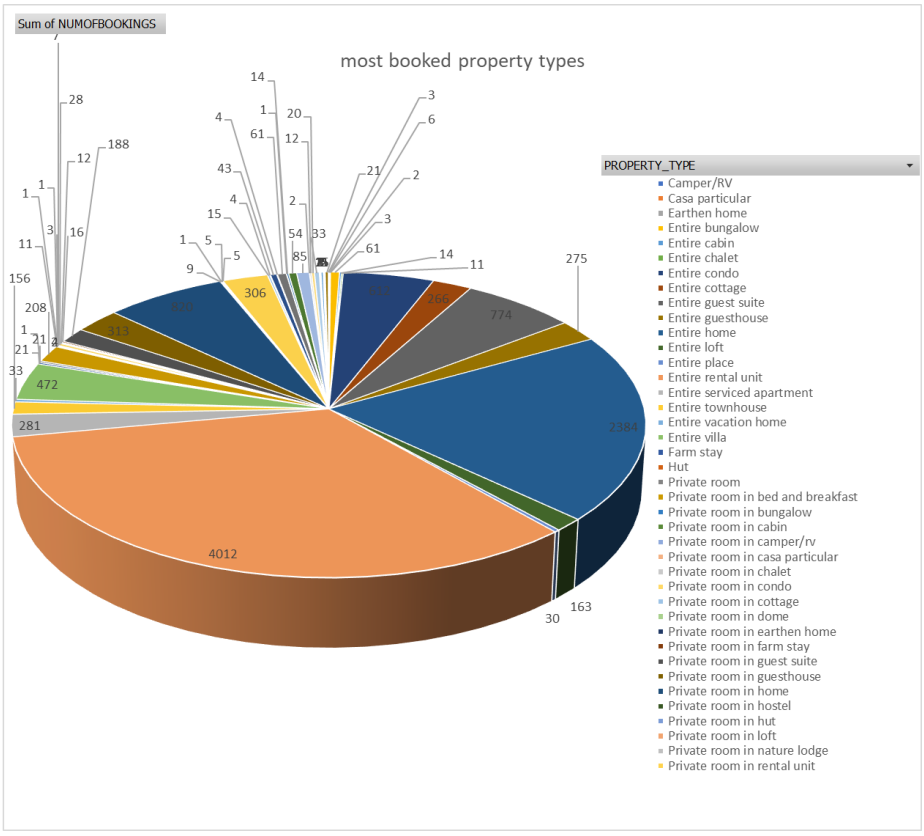
- Listings with distinct features, such as the "Private Room in Bed and Breakfast" with 38 bathrooms and 211 bookings, present unique options for travelers seeking specialized experiences or luxury accommodations.

Budget Considerations:

- Understanding that popular properties, like "Entire guest suites" with 775 bookings, reflect quality and customer satisfaction can help travelers justify higher prices, ensuring they receive good value for their money.
- **Complement with Bed Analysis:** To gain a comprehensive understanding of traveler preferences, it's crucial to analyze the number of beds alongside existing data on bathrooms. For instance, "Entire rental units," which have **4,**013 bookings and an average of 6.5 bathrooms, likely offer multiple beds, appealing to families or groups seeking spacious accommodations. In contrast, properties like "Private room in home," with 822 bookings, may feature fewer beds, indicating a preference for solo travelers or couples. This combined analysis helps identify market segments, such as those favoring shared accommodations or budget-friendly options, informing pricing strategies and targeted marketing campaigns. By understanding these dynamics, property owners and investors can tailor their offerings to better meet diverse traveler needs, ultimately enhancing booking rates and customer satisfaction.

Price and Booking Trends:

- Investors can consider investigating how pricing and booking trends correlate with the number of bathrooms can help to understand how much travelers are willing to pay for additional bathroom facilities.



- **What is the property type booked the most (Hotels or Bed and Breakfasts) and for how long?**

Max Stay Duration (365 Days)

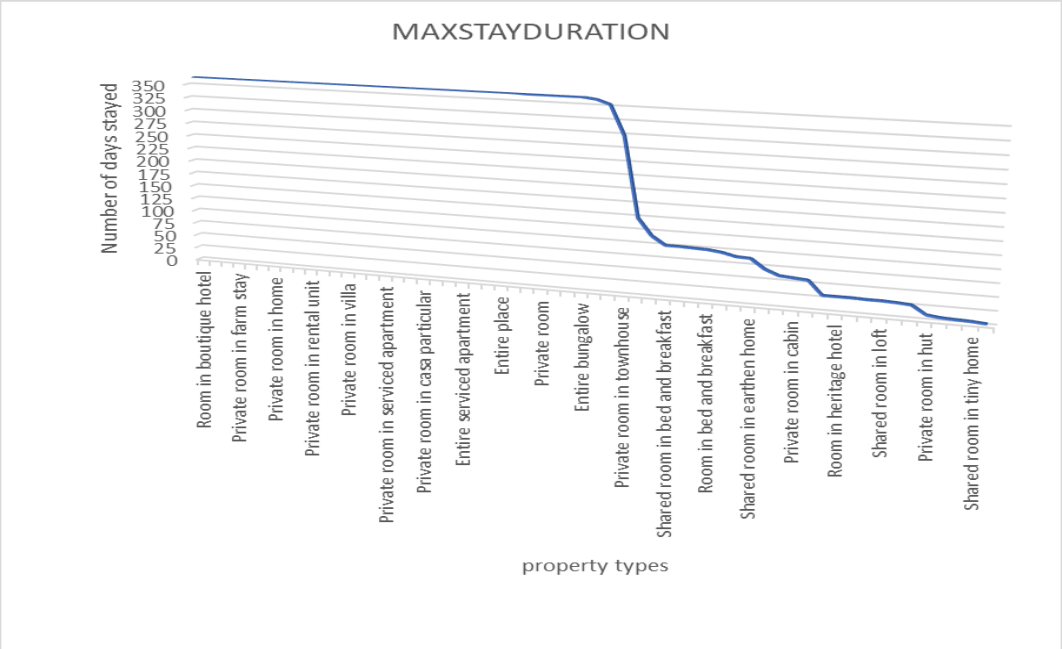
- Properties: "Room in boutique hotel," "Entire rental unit," "Private room in guest suite."
- Indicates strong demand for long-term rentals, appealing to expatriates and extended vacationers.
- Opportunity for steady income streams through long-term commitments.

Moderate Stay Duration (101 to 364 Days)

- Properties: "Entire townhouse" (364 days), "Shared room in rental unit" (353 days).
- Attracts travelers seeking flexibility without committing to a full year.
- Appeals to digital nomads and seasonal workers, creating diverse market opportunities.

Short-Term Stay (Less than 100 Days)

- Properties: "Room in hotel" (100 days), "Shared room in bed and breakfast" (100 days).
- Ideal for transient travelers like tourists and business professionals.
- Focus on attractive pricing strategies to appeal to budget-conscious guests.
- Hotels:
    - Property Types: Room in boutique hotel, Entire serviced apartment
    - Max Stay Duration: some allow up to 365 days.
- Bed and Breakfasts:
    - Property Types: Room in bed and breakfast, Shared room in bed and breakfast
    - Max Stay Duration: Ranges from 1 days to 365 days.



- **How does the quantity of reviews received by an Airbnb listing correlate with its average rating, and what are the underlying patterns that emerge from this relationship? Specifically, how does the average rating of Airbnb listings differ across various neighbourhood groups within a city or region?**

  By looking and identifying which listings in the Airbnb database have an average rating of 4 or higher with many reviews we can identify which Airbnb's are the most successful and determine what makes those Airbnb better than other Airbnb's. By looking at what amenities they provide, where are they located and what are their prices. We can also identify Airbnb's with low rating and many reviews to determine what not to do when investing in Airbnb's. By looking at the scatter chart we can also see how the number of reviews correlate with the average rating as we can see that the more reviews a Airbnb has the lower the rating which makes it more accurate. While Airbnb with higher average ratings but lower reviews can be less accurate.

  By looking at the data we can see there is 24 Airbnb with average ratings of 3 or higher and with more than 50 reviews. By reviewing each Airbnb we can see what neighbourhood group they are part of and what their location is and what amenities they have as well as their prices. For example we can see for listing 1058642 it has more than 289 reviews with a price of R2230 for 2 nights and it is only available 68 days of the entire year showing that it is very exclusive Airbnb in a very good area. For a more budget Airbnb we have listing 771176 which also has a average rating of higher than 3 and 275 reviews with a price of R827 and has more availability with the Airbnb being available for 291 days of the year with a minimum number of nights of 2.