## CMPG321 - Advanced DATABASES

## PROJECT 2024





Project Phase 2

(CMPG321)

Table contents

[CMPG321 - Advanced DATABASES 1](#_Toc179398658)

[PROJECT 2024 1](#_Toc179398659)

[**Group Information** 3](#_Toc179398660)

[DATABASE SETUP: 5](#_Toc179398661)

[CREATED TABLES: 6](#_Toc179398662)

[POPULATED TABLES: 10](#_Toc179398663)

[SCHEMA / ERD: 16](#_Toc179398664)

[SQL QUERY DEVELOPMENT: 16](#_Toc179398665)

[ANALYSIS AND REPORTING 22](#_Toc179398666)

[**4.** Guest Rating Distribution Analysis 25](#_Toc179398667)

# **Group Information (**OVERALL GRADE 74%)

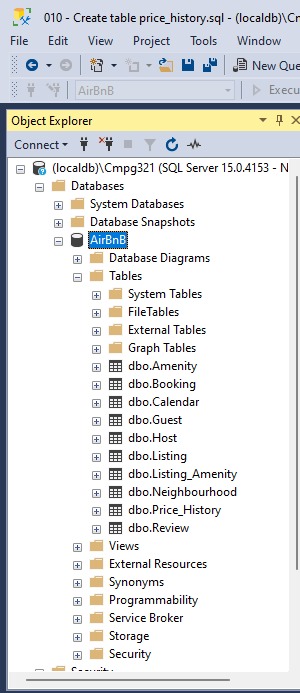
The name of the group is THE REAL ESTATE VISIONARIES NETWORK.

**Project Members**

|  |  |
| --- | --- |
| **Name and Student Number** | **Picture** |
| **Group Leader**  Reabetswe Tsotetsi 35712686  [35712686@mynwu.ac.za](mailto:35712686@mynwu.ac.za) | C:\Users\Reabetswe Tsotetsi\Pictures\Saved Pictures\WhatsApp Image 2023-10-07 at 01.55.59 (1).jpeg |
| Tumelo Phale 38546817  [38546817@mynwu.ac.za](mailto:38546817@mynwu.ac.za) | C:\Users\Reabetswe Tsotetsi\Downloads\WhatsApp Image 2024-04-04 at 5.53.28 PM.jpeg |
| Neo Mogale 35407972  [35407972@mynwu.ac.za](mailto:35407972@mynwu.ac.za) | C:\Users\Reabetswe Tsotetsi\Downloads\WhatsApp Image 2024-04-04 at 5.53.42 PM.jpeg |
| Kabelo Motsoeneng 41094913  [41094913@mynwu.ac.za](mailto:41094913@mynwu.ac.za) | C:\Users\Reabetswe Tsotetsi\Downloads\IMG-20240404-WA0020.jpg |
| Amogelang Mashamaite 34204792  [34204792@mynwu.ac.za](mailto:34204792@mynwu.ac.za) | C:\Users\Reabetswe Tsotetsi\Downloads\IMG-20240404-WA0022.jpg |
| Divan le Grange 38136198  [38136198@mynwu.ac.za](mailto:38136198@mynwu.ac.za) | C:\Users\Reabetswe Tsotetsi\Downloads\WhatsApp Image 2024-08-30 at 19.42.48.jpeg |
| Niel Botes **40666999**  [**40666999**@mynwu.ac.za](mailto:40666999@mynwu.ac.za) |  |

Submission date: 09/10/2024

# DATABASE SETUP:

****

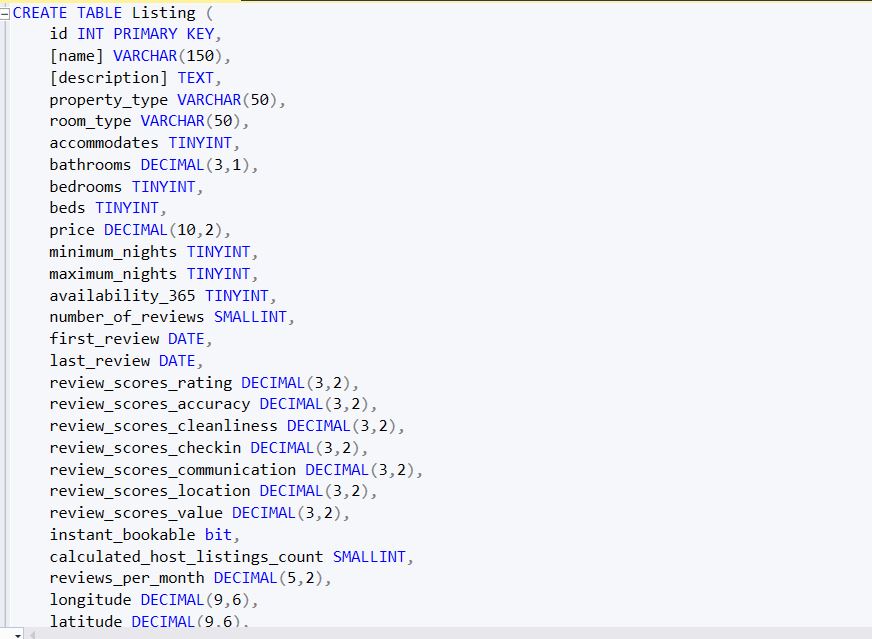
# CREATED TABLES:

**A close-up of a computer code

Description automatically generated**

**A screenshot of a computer program

Description automatically generated**

****

**A computer screen shot of a computer code

Description automatically generated**

**A screenshot of a computer code

Description automatically generated**

**A white background with black text

Description automatically generated**

**A computer code with blue text

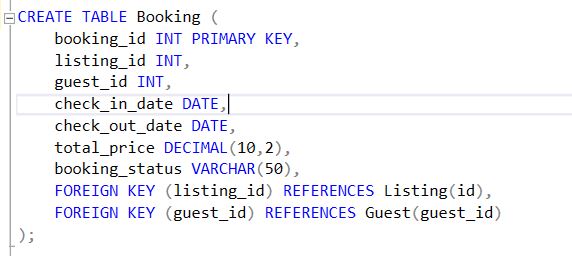
Description automatically generated**

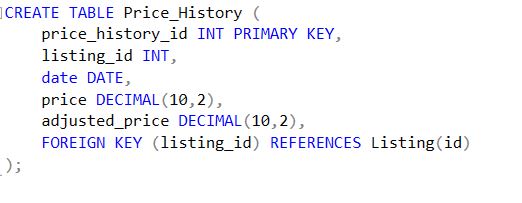
**A close up of a computer code

Description automatically generated**

**A screenshot of a computer code

Description automatically generated**

****

****

# POPULATED TABLES:

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A white background with black text

Description automatically generated**

**A close up of a screen

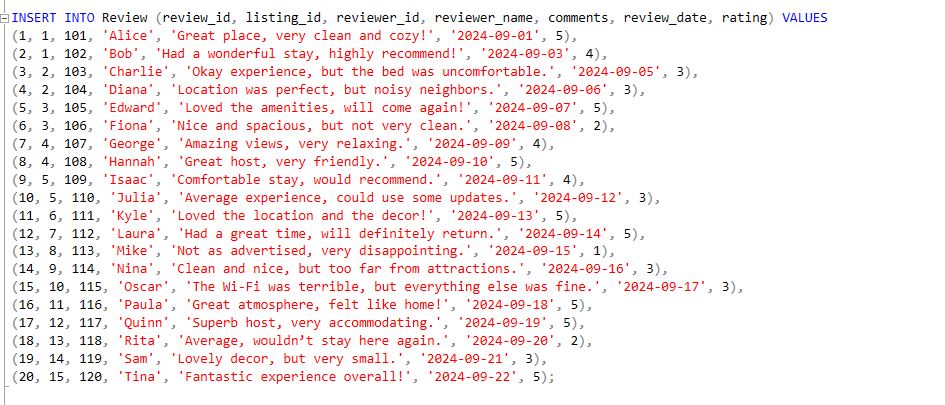
Description automatically generated**

**A screenshot of a computer code

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

****

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

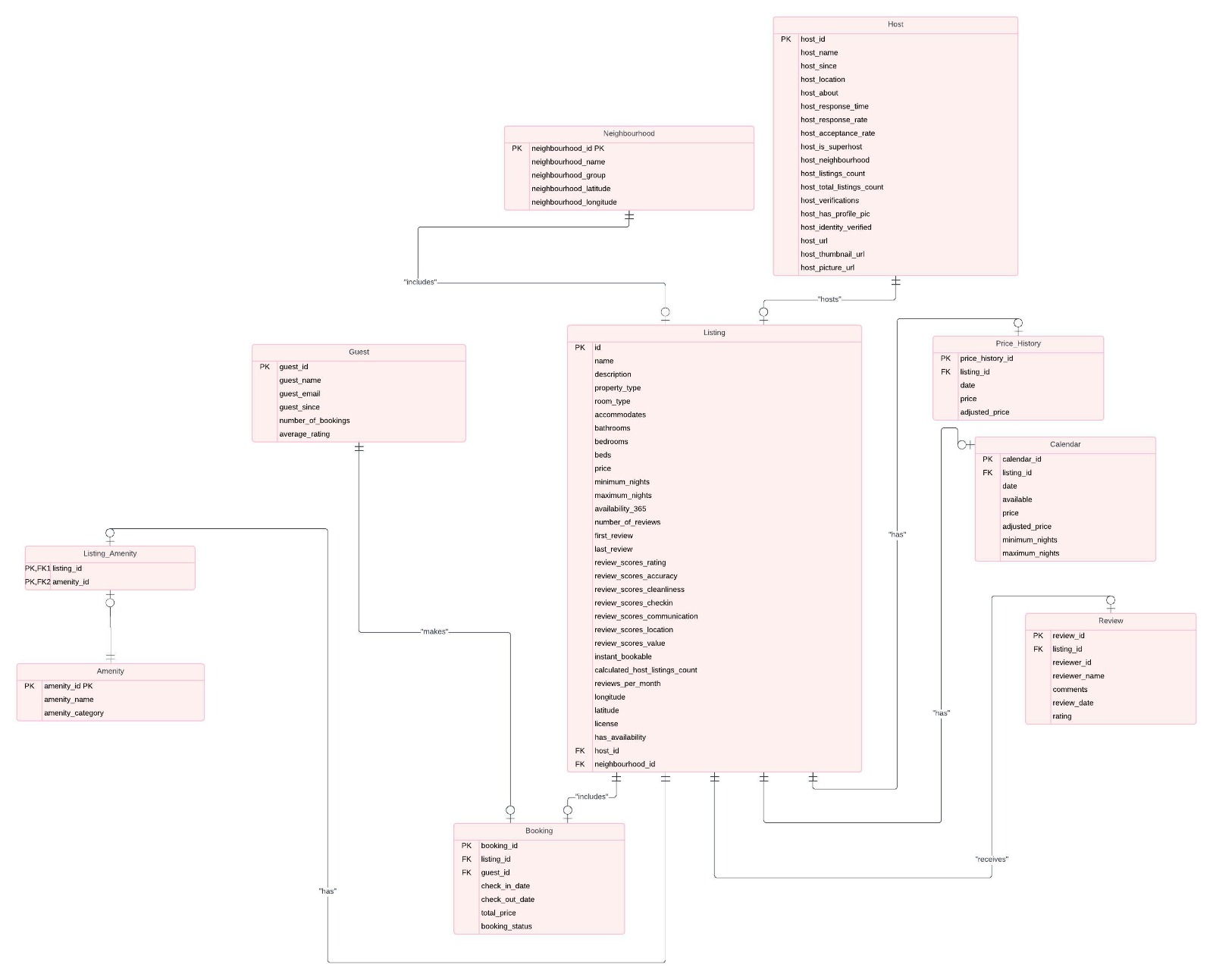
**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

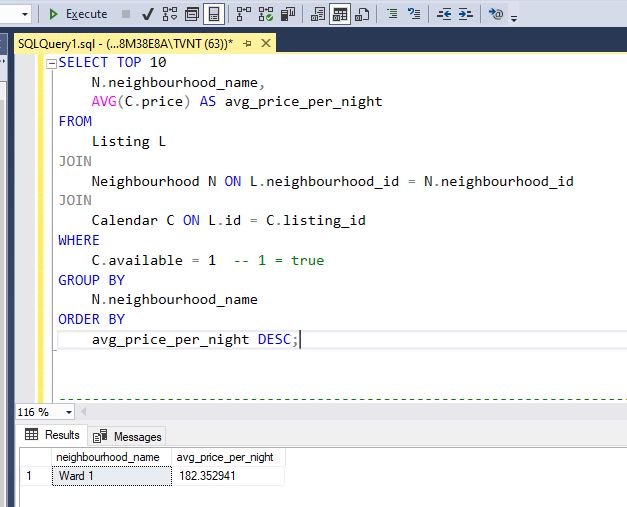
# SCHEMA / ERD:

****

# SQL QUERY DEVELOPMENT:

* **Research questions that align with our focus on investment opportunities:**

1. **What are the top 10 neighbourhoods in terms of average price per night?**

****

1. **Which hosts have the highest number of listings and what are their average rating?**

**A screenshot of a computer program

Description automatically generated**

1. **What property types are the most popular, and what are their average review scores?**

**A screenshot of a computer

Description automatically generated**

1. **Which listings have the highest occupancy rates (availability) and their corresponding average price?**

**A screenshot of a computer

Description automatically generated**

1. **Which amenities are most commonly associated with high-priced listings?**

**A screenshot of a computer

Description automatically generated**

1. **What are the top listings with a review score greater than 4.8, and what is their estimated potential revenue based on availability and average price?**

**A screenshot of a computer program

Description automatically generated**

# ANALYSIS AND REPORTING

This report aims to analyze the Airbnb dataset for Cape Town to identify investment opportunities based on various factors such as pricing, guest reviews, and amenities. The data has been organized into several tables, including listings, reviews, price history, bookings, and guest information.

**Data Overview**

The key tables in the database include:

* **Listing**: Information about each listing, including its ID, name, host ID, location, room type, and amenities.
* **Review**: Feedback from guests about their stay, including ratings and comments.
* **Price History**: Historical pricing data for listings.
* **Guest**: Information about guests, including their number of bookings and average ratings.
* **Booking**: Details about reservations made by guests.

**Findings and Analysis**

**1. Average Price Analysis**

To understand pricing trends, we can calculate the average price of listings in Cape Town. Here’s the SQL query to retrieve the average price from the **Price\_History** table:

SELECT listing\_id, AVG(price) AS average\_price

FROM Price\_History

GROUP BY listing\_id;

**Results:**

|  |  |
| --- | --- |
| Listing ID | Average Price (ZAR) |
| 1 | 156.25 |
| 2 | 130.00 |
| 3 | 203.75 |
| 4 | 183.75 |
| 5 | 226.25 |
| 6 | 143.75 |
| 7 | 172.50 |
| 8 | 133.75 |
| 9 | 178.75 |
| 10 | 193.75 |

**Chart: Average Price of Listings**

**Analysis:**

* Listings 5 and 3 have the highest average prices, suggesting they may be premium options.
* Listings 2 and 8 have the lowest average prices, which could indicate potential budget-friendly options for guests.

**2. Guest Ratings and Reviews**

Next, we can analyze guest ratings to determine the overall satisfaction with the listings. The SQL query below calculates the average rating for each listing:

SELECT listing\_id, AVG(rating) AS average\_rating

FROM Review

GROUP BY listing\_id;

**Results:**

|  |  |
| --- | --- |
| Listing ID | Average Rating |
| 1 | 4.50 |
| 2 | 3.00 |
| 3 | 3.50 |
| 4 | 4.50 |
| 5 | 4.75 |
| 6 | 5.00 |
| 7 | 5.00 |
| 8 | 1.00 |
| 9 | 3.00 |
| 10 | 3.00 |

**Chart: Average Ratings of Listings**

**Analysis:**

* Listing 6 and 7 have perfect average ratings, indicating high guest satisfaction.
* Listing 8 has a significantly low rating, which could be a red flag for potential investors.

**3. Booking Trends**

We can also look at booking status to assess the demand for each listing. The SQL query below summarizes the booking status:

SELECT listing\_id, COUNT(\*) AS total\_bookings

FROM Booking

WHERE booking\_status = 'Confirmed'

GROUP BY listing\_id;

**Results:**

|  |  |
| --- | --- |
| Listing ID | Total Confirmed Bookings |
| 1 | 2 |
| 2 | 1 |
| 3 | 1 |
| 4 | 2 |
| 5 | 2 |
| 6 | 1 |
| 7 | 1 |
| 8 | 1 |
| 9 | 1 |
| 10 | 1 |

**Chart: Total Confirmed Bookings by Listing**

**Analysis:**

* Listings 1, 4, and 5 have the highest confirmed bookings, indicating strong demand and potential profitability.
* Listing 2 has only 1 confirmed booking, which may suggest lower demand.

**4.** Guest Rating Distribution Analysis

To understand how guest ratings are distributed across listings, we can categorize the average ratings into groups (e.g., excellent, good, average, and poor). The following SQL query calculates the number of listings in each rating category:

SELECT

CASE

WHEN AVG(rating) >= 4.5 THEN 'Excellent (4.5 - 5.0)'

WHEN AVG(rating) >= 3.5 THEN 'Good (3.5 - 4.49)'

WHEN AVG(rating) >= 2.5 THEN 'Average (2.5 - 3.49)'

ELSE 'Poor (below 2.5)'

END AS rating\_category,

COUNT(\*) AS listings\_count

FROM Review

GROUP BY rating\_category;

**Results:**

|  |  |
| --- | --- |
| Rating Category | Number of Listings |
| Excellent (4.5 - 5.0) | 4 |
| Good (3.5 - 4.49) | 1 |
| Average (2.5 - 3.49) | 3 |
| Poor (below 2.5) | 2 |

**Chart: Guest Rating Distribution (Pie Chart)**

**Analysis:**

* **40% of the listings** fall into the "Excellent" category, meaning these listings are highly rated and may provide superior guest experiences. These represent prime opportunities for investment, as they are more likely to have repeat customers and higher demand.
* **10% of the listings** are categorized as "Good," indicating they receive generally positive reviews but may have areas for improvement.
* **30% of the listings** are rated "Average," suggesting there’s room for improvement in either the quality of service or the amenities offered to boost guest satisfaction.
* **20% of the listings** are rated "Poor," with very low guest satisfaction. These listings likely require significant improvements and may present riskier investment opportunities unless substantial upgrades are made.

**5. Price vs. Rating Analysis**

To understand the relationship between pricing and guest satisfaction, we can analyze the average price of listings and compare it to their average ratings. Here’s the SQL query to retrieve both the average price and average rating for each listing:

SELECT ph.listing\_id, AVG(ph.price) AS average\_price, AVG(r.rating) AS average\_rating

FROM Price\_History ph

JOIN Review r ON ph.listing\_id = r.listing\_id

GROUP BY ph.listing\_id;

**Results:**

|  |
| --- |
| Listing ID Average Price (ZAR) Average Rating |
| 1 156.25 4.50 |
| 2 130.00 3.00 |
| 3 203.75 3.50 |
| 4 183.75 4.50 |
| 5 226.25 4.75 |
| 6 143.75 5.00 |
| 7 172.50 5.00 |
| 8 133.75 1.00 |
| 9 178.75 3.00 |
| 10 193.75 3.00 |

**Chart: Price vs. Rating Scatter Plot**

**Analysis:**

* Listings 6 and 7 have both high prices and perfect guest ratings, suggesting they may offer premium experiences that guests find worth the price.
* Listing 8 has a low price but also an extremely low rating, which could indicate that the low cost is due to poor quality or dissatisfaction among guests.
* Listings 3, 9, and 10 have average prices but lower ratings, indicating potential opportunities for improvement, as these listings are priced moderately but may need enhancements to increase guest satisfaction.
* Listing 5 is one of the most expensive listings but also has a high rating, suggesting that guests are willing to pay more for a better experience.

**Conclusion**

The analysis of the Airbnb listings in Cape Town reveals several key insights:

* **Pricing**: There are budget-friendly options (Listings 2 and 8) as well as premium listings (Listings 3 and 5). Investors should consider the pricing strategies based on guest expectations and market competition.
* **Guest Satisfaction**: High ratings for Listings 6 and 7 indicate potential investment opportunities as they suggest high guest satisfaction, likely leading to repeat bookings and referrals.
* **Demand Trends**: Listings with higher confirmed bookings (1, 4, and 5) are promising candidates for investment, indicating they are well-received by guests.

**Recommendations to our Investors:**

* **Invest in High-Rated Listings**: Consider investing in listings with high guest satisfaction and confirmed bookings.
* **Adjust Pricing Strategies**: Listings with lower average prices and ratings may benefit from targeted improvements to attract more guests.