

# Group 20

## Die Vrystaat Mielies



Project Phase 1

(CMPG321)

**Group Members:**

[42468000 - FC Venter] - (Group Leader)

[40887340 - JL Coetzee]

[34250174 - K Botha]

[31868444 - W Otto]

[37635883 - M Buitendag]

[41153324 - W Botha]

[41799305 – C Jansen]

[41080815 - M Fourie]

Submission date: 22/08/2024

# 1. TABLE OF CONTENTS

---

Group Information .....	1
.....	1
2. Phase 1 – Focus Area and Requirements Analysis .....	2
3. Area of Focus .....	2
4. Research Questions .....	2
5. Dataset Description.....	4

# DIE VRYSTAAT MIELIES

Meet the team!



**FC VENTER**

Team Leader

42468000

082 729 1986



**KAYLA BOTHA**

34250174

078 500 1276



**WIAN OTTO**

31868444

081 470 3105



**MICHAEL FOURIE**

41080815

073 665 6727



**CARLA JANSEN**

41799305

084 599 6531



**WIMPIE BOTHA**

41153324

066 249 1840



**MORGAN BUITENDAG**

37635883

071 509 3701



**JOHAN-LOUIS COETZEE**

40887340

072 674 6394

## 2. PHASE 1 – FOCUS AREA AND REQUIREMENTS ANALYSIS

### 3. AREA OF FOCUS

---

#### **Best value for money:**

We decided to focus on finding Airbnb listings that offer the best value for money. We chose this approach to give people an easy way to find ideal accommodations for their trips based on personalized criteria. We will use an extensive dataset containing relevant listing information, and analyse it with SQL using the research questions seen in this paper. When reviewing, we consider several key points.

- **Pricing:** Whether the listing is priced appropriately relative to similar properties in your neighbourhood.
- **Key Amenities:** Features and services of the listing.
- **Location:** The geographical location of the listing.
- **Ratings:** Reviews and ratings from guests on quality of listings and hosting experience.
- **Availability:** How many days the listing is available per year.
- **Family Friendliness:** Is the listing appropriate for families.
- **Group Accommodation:** Can the listing accommodate groups, and if so what size groups can be accommodated.

### 4. RESEARCH QUESTIONS

---

#### **1. What are the top 10 most affordable Airbnb listings in Cape Town that accommodate at least X people, offer at least X key amenities, and are priced under ZAR X per night?**

This will be answered by providing the user with the 10 most affordable listings that accommodate the user's preferred number of guests with the chosen necessary amenities that falls within a specific price.

#### **2. Which Airbnb listings for groups of X or more people have the highest guest ratings and fall within the X range price category?**

This will elucidate the Airbnb listings with the highest ratings for the user's indicated number of guests that fall within a specified price range.

#### **3. Which neighbourhoods in Cape Town have the highest concentration of budget-friendly listings that can accommodate X to X people and have positive ratings?**

This will indicate which cape town neighbourhoods have the greatest density of listings that are priced below average that meet the user's chosen range of guests and have ratings greater than 3 stars.

#### **4. What is the average nightly price for listings in each neighbourhood that can accommodate a group?**

This will provide the user with the average cost per night for Airbnb listings that can accommodate groups in various neighbourhoods, helping to identify areas with affordable group accommodations.

**5. Which listings offer the best value in terms of price per person per night for groups of X?**

This will highlight the Airbnb listings that offer the most cost-effective options, calculating the price per person per night for the user's specified group size.

**6. Which neighbourhoods offer the best value for families with listings under ZAR X per night, considering key amenities?**

This will identify the neighbourhoods where families can find affordable Airbnb listings under a specific price, ensuring that key amenities are included for a comfortable stay.

**7. What is the average cost of listings in top-rated areas that can accommodate X people, and how does this compare to similar listings in less popular areas?**

This will compare the average prices of Airbnb listings that can accommodate the specified number of guests in top-rated areas versus less popular areas, helping the user understand the price difference.

**8. Which listings for groups of X people have the most consistent positive guest feedback and are priced below the citywide average of ZAR X per night?**

This will identify the Airbnb listings that have consistently received positive feedback for groups of the specified size and are priced below the city's average, offering a balance of affordability and quality.

**9. What is the relationship between listing price and the number of amenities offered in the most popular neighbourhoods for large groups, focusing on listings under ZAR X per night?**

This will analyse how the price of listings correlates with the number of amenities provided in popular neighbourhoods, particularly for large groups, focusing on listings under a specified price.

**10. Which listings in under-serviced areas offer the best value for money in terms of guest ratings and price for groups of X people, particularly those under ZAR X per night?**

This will identify listings in areas with fewer options that still provide excellent value based on guest ratings and price, specifically for groups of the specified size and under a set budget.

**11. What is the order of the Airbnb listings in Cape Town sorted from highest to lowest by price?**

This will organize the Airbnb listings in Cape Town from the most expensive to the least expensive, helping the user see the range of prices available.

**12. What is the order of the Airbnb listings in Cape Town sorted from lowest to highest by price?**

This will arrange the Airbnb listings in Cape Town from the least expensive to the most expensive, enabling the user to find the most budget-friendly options first.

### 13. Which listings are rated X stars or higher?

This will filter the Airbnb listings to show only those that have received the user-specified star rating or higher, ensuring high-quality options are presented.

### 14. What is the order of the listings with the lowest price and highest rating?

This will rank the listings based on a weighted score, where price is given a 60% weighting and rating a 40% weighting, to find the best balance between affordability and guest satisfaction.

### 15. How likely is a listing to be available?

This will provide insights into the availability of a listing, helping the user understand the likelihood of securing a booking for their preferred dates.

### 16. Which Airbnb listings have the highest rating?

This will show the Airbnb listings with the highest overall guest ratings, allowing the user to identify the top-rated accommodations.

### 17. Which Airbnb listings have the lowest rating?

This will display the Airbnb listings with the lowest overall guest ratings, helping the user avoid poorly reviewed accommodations.

### 18. What are the listings based on room type?

This will categorize the Airbnb listings by room type, such as entire place, private room, or shared room, allowing the user to find options that match their preference.

## 5. DATASET DESCRIPTION

---

### Data Structure

- **Total Rows (Listings):** 23,564
- **Total Columns:** 18

### Key Attributes

#### 1. ID Fields:

- **id:** Unique identifier for each listing.
- **host\_id:** Unique identifier for each host.

#### 2. Property Information:

- **name:** The name of the listing, which typically describes the property.
- **room\_type:** Indicates the type of room being offered (e.g., Entire home/apt, Private room, Shared room).
- **neighbourhood:** The specific area or neighbourhood where the listing is located.

- **latitude & longitude:** Geographical coordinates of the property, useful for mapping the locations.

### 3. Pricing and Availability:

- **price:** The price per night for the listing, in the local currency (presumably South African Rand).
- **minimum\_nights:** The minimum number of nights required for booking.
- **availability\_365:** Number of days in the year that the listing is available for booking.

### 4. Review Data:

- **number\_of\_reviews:** Total number of reviews the listing has received.
- **last\_review:** Date of the most recent review.
- **reviews\_per\_month:** Average number of reviews the listing receives per month.
- **number\_of\_reviews\_ltm:** Number of reviews in the last 12 months.

### 5. Host Information:

- **host\_name:** The name of the host managing the listing.
- **calculated\_host\_listings\_count:** Number of listings the host manages.

### 6. Regulatory Information:

- **license:** License number associated with the listing, if available. This field is sparsely populated.

## Summary of Key Attributes

- **Price:**
  - Range: 160 to 999,538 (currency unit is not specified, but typically ZAR).
  - Median: 1,250.5
- **Minimum Nights:**
  - Range: 1 to 999 nights
  - Median: 2 nights
- **Number of Reviews:**
  - Range: 0 to 756 reviews
  - Median: 6 reviews
- **Availability:**
  - Range: 0 to 365 days
  - Median: 220 days



## Data Format and Domain

- **Format:** The dataset is in CSV format with a mix of numeric, categorical, and date fields.
- **Data Types:**
  - id, host\_id: Integer
  - name, host\_name, neighbourhood, room\_type, license: String
  - latitude, longitude, price, minimum\_nights, number\_of\_reviews, reviews\_per\_month, calculated\_host\_listings\_count, availability\_365, number\_of\_reviews\_ltm: Numeric
  - last\_review: Date

## Relevance to Our Application

For our application that focuses on finding the best places according to the number of people, price, and reviews:

- **price**, **number\_of\_reviews**, **room\_type**, and **availability\_365** are directly relevant.
- **latitude** and **longitude** can be used to filter or map listings by location.
- **minimum\_nights** might influence the suitability of a listing for different types of stays.
- **reviews\_per\_month** and **last\_review** can help gauge the recency and frequency of reviews, indicating popularity and reliability.

## Overview of the Dataset

The dataset contains various attributes related to Airbnb listings in Cape Town. Below is a detailed description of all the available data fields, including their format, range, and domain.

### Data Fields

1. **id**
  - **Format:** Integer
  - **Range:** Varies per listing (Unique ID for each listing)
  - **Domain:** Identifies each Airbnb listing uniquely.
2. **name**
  - **Format:** String
  - **Domain:** Descriptive name of the Airbnb listing, which often includes key features or location.
3. **host\_id**
  - **Format:** Integer
  - **Range:** Varies per host (Unique ID for each host)

- **Domain:** Identifies each host uniquely.
4. **host\_name**
- **Format:** String
  - **Domain:** The name of the person or entity hosting the listing.
5. **neighbourhood\_group**
- **Format:** String
  - **Domain:** This field is often used to group neighbourhoods, but it appears to be mostly NaN in this dataset.
6. **neighbourhood**
- **Format:** String
  - **Domain:** Specific area or neighbourhood where the listing is located (e.g., "Ward 57").
7. **latitude**
- **Format:** Float
  - **Range:** Approx. -34 to -33 (Cape Town coordinates)
  - **Domain:** Geographic coordinate for the latitude of the listing location.
8. **longitude**
- **Format:** Float
  - **Range:** Approx. 18 to 19 (Cape Town coordinates)
  - **Domain:** Geographic coordinate for the longitude of the listing location.
9. **room\_type**
- **Format:** String
  - **Domain:** Type of room offered (e.g., "Entire home/apt", "Private room", "Shared room").
10. **price**
- **Format:** Float
  - **Range:** 160 to 999,538
  - **Domain:** Price per night for the listing, likely in South African Rand (ZAR).
11. **minimum\_nights**
- **Format:** Integer
  - **Range:** 1 to 999
  - **Domain:** Minimum number of nights required to book the listing.

#### 12. number\_of\_reviews

- **Format:** Integer
- **Range:** 0 to 756
- **Domain:** Total number of reviews that the listing has received.

#### 13. last\_review

- **Format:** Date (YYYY-MM-DD)
- **Range:** Various dates, with the most recent around June 2024
- **Domain:** The date of the most recent review.

#### 14. reviews\_per\_month

- **Format:** Float
- **Range:** 0.01 to 15.18
- **Domain:** Average number of reviews the listing receives per month.

#### 15. calculated\_host\_listings\_count

- **Format:** Integer
- **Range:** 1 to 134
- **Domain:** Number of listings managed by the host.

#### 16. availability\_365

- **Format:** Integer
- **Range:** 0 to 365
- **Domain:** Number of days in the year that the listing is available for booking.

#### 17. number\_of\_reviews\_ltm

- **Format:** Integer
- **Range:** 0 to 178
- **Domain:** Number of reviews received in the last 12 months.

#### 18. license

- **Format:** String
- **Domain:** License number associated with the listing, if available. This field is mostly sparse.

### Summary of Key Attributes

- **Numerical Data:** Includes fields like price, minimum\_nights, number\_of\_reviews, reviews\_per\_month, calculated\_host\_listings\_count, availability\_365,

number\_of\_reviews\_ltm. These fields are mostly integers or floats and are critical for quantitative analysis.

- **Categorical Data:** Includes fields like name, host\_name, neighbourhood, room\_type, license. These fields are strings and provide descriptive or categorical information about the listings.
- **Geospatial Data:** Latitude and longitude fields allow for spatial analysis and mapping of listings.
- **Date Data:** last\_review provides temporal information regarding the recency of reviews.