**Dataset Description**

**1. Dataset Overview:**

* **Dataset Name:** Airbnb Listings - Cape Town, South Africa
* **Source:** Airbnb Listings
* **Date of Collection/Last Update:** August 2022
* **Format:** SQL database
* **Number of Records (Rows):** 114 Listings
* **Number of Attributes (Columns):** 5 columns

**2. Objective of the Dataset:**

* **Purpose:** The dataset provides detailed information on Airbnb listings in Cape Town, South Africa. It is intended for analysis of property availability, pricing, occupancy, ratings, and other metrics relevant to rental trends.

**3. Key Fields (Attributes):**

| **Field Name** | **Description** | **Data Type** | **Example Values** | **Domain/Range** |
| --- | --- | --- | --- | --- |
| id | Unique identifier for each listing | Integer | 12345 | Positive integers |
| name | Listing name/title | String | "Cozy Apartment near the Beach" | Text |
| host\_id | Unique identifier for the host | Integer | 54321 | Positive integers |
| neighbourhood | Area or suburb within Cape Town | String | "Sea Point", "Camps Bay" | Specific neighbourhood names |
| latitude | Latitude coordinate of the listing | Float | -33.9258 | Latitude range for Cape Town |
| longitude | Longitude coordinate of the listing | Float | 18.4232 | Longitude range for Cape Town |
| room\_type | Type of accommodation offered | String | "Entire home/apt", "Private room" | Set of predefined values |
| price | Price per night in local currency | Integer | 1200 | Positive integers (in ZAR, for example) |
| minimum\_nights | Minimum number of nights required for booking | Integer | 2 | Positive integers |
| number\_of\_reviews | Total number of reviews the listing has received | Integer | 45 | Non-negative integers |
| last\_review | Date of the most recent review | Date | "2023-11-15" | Date format (YYYY-MM-DD) |
| reviews\_per\_month | Average number of reviews received per month | Float | 2.5 | Positive floats |
| availability\_365 | Number of days the listing is available within a year | Integer | 200 | 0–365 |
| rating | Average user rating | Float | 4.8 | 1–5 (1 being worst, 5 being best) |
| amenities | List of amenities provided | String | "Wi-Fi, Air conditioning" | Text |

**4. Data Types and Format:**

* **String Fields:** Text-based information such as property names, room types, and neighbourhoods.
* **Numeric Fields:** Prices, number of reviews, and availability are all expressed as integers or floats.
* **Date Fields:** The dataset contains date fields for review dates, typically in the format YYYY-MM-DD.
* **Geospatial Data:** Latitude and longitude fields are provided for mapping purposes.

**5. Range/Domain of Key Attributes:**

* **Price:** [Min: 100 ZAR, Max: 20,000 ZAR]
* **Rating:** [Min: 1, Max: 5]
* **Availability:** [Min: 0 days, Max: 365 days]
* **Number of Reviews:** [Min: 0, Max: 1000+]

**6. Assumptions:**

* The dataset assumes that all entries are current and reflect accurate data for the time they were collected.
* Missing data is either represented by null values or specific placeholders such as "N/A."

**7. Limitations:**

* The dataset does not account for seasonal variations in booking rates.
* Certain fields may contain incomplete or outdated information due to the dynamic nature of Airbnb listings.
* External factors (e.g., economic changes or travel restrictions) that may influence trends are not reflected in this dataset.

**8. Data Cleaning Considerations:**

* Duplicate listings need to be identified and removed.
* Missing or erroneous data (e.g., extreme values in price or availability) should be handled appropriately.



















