



Airbnb stay

WITH

CAPECASA FINDERS

**RECOMMENDING THE BEST PRICE**

**&**

**STAY OPTIONS IN CAPETOWN**

# BUSINESS UNDERSTANDING

THIS PROJECT BUILDS A RECOMMENDER SYSTEM FOR CAPE TOWN AIRBNB LISTINGS TO HELP HOSTS OPTIMIZE PRICING, OCCUPANCY, AND GUEST SATISFACTION. USING DATA FROM INSIDE AIRBNB,

THE MODEL SUGGESTS OPTIMAL PRICES BASED ON LISTING FEATURES, GUEST SENTIMENT, AND OCCUPANCY PATTERNS, HELPING HOSTS SET COMPETITIVE AND PROFITABLE PRICES.



## DATA DESCRIPTION

KEY DATA FILES INCLUDE:

- LISTINGS.CSV: CONTAINS DETAILS ABOUT EACH LISTING, SUCH AS PROPERTY TYPE, LOCATION, AMENITIES, AND HOST INFORMATION.
- CALENDAR.CSV: SHOWS THE AVAILABILITY AND PRICING FOR EACH LISTING OVER TIME.
- REVIEWS.CSV: PROVIDES GUEST FEEDBACK, WHICH IS USED TO DERIVE SENTIMENT SCORES.



## KEY COLUMNS

- PROPERTY DETAILS: PROPERTY\_TYPE, ACCOMMODATES, BATHROOMS, BEDROOMS, AND BEDS.
- HOST INFORMATION: FIELDS LIKE HOST\_ID, HOST\_RESPONSE\_RATE, HOST\_IS\_SUPERHOST, AND HOST\_LISTINGS\_COUNT.
- PRICING AND OCCUPANCY: COLUMNS SUCH AS PRICE, AVAILABILITY, AND NUMBER\_OF\_REVIEWS.
- GUEST SENTIMENT: DERIVED FROM GUEST COMMENTS USING SENTIMENT ANALYSIS TO SCORE THE EMOTIONAL TONE OF EACH REVIEW.



# DATA PREPROCESSING

Process	Description
Data Cleaning:	<ul style="list-style-type: none"><li>Removed duplicate entries and irrelevant columns.</li><li>Addressed missing values through imputation for numerical columns and frequency encoding for categorical ones</li></ul>
Feature Engineering:	<ul style="list-style-type: none"><li>Created new features such as sentiment_score, extracted from guest reviews using sentiment analysis.</li><li>Encoded categorical variables, applying frequency encoding to columns like neighbourhood_cleansed and property_type for improved model performance.</li></ul>
Transformations:	<ul style="list-style-type: none"><li>Log-transformed the price column to reduce skewness and approximate a normal distribution.</li></ul>
Data Splitting:	Split the data into training and testing sets, ensuring each listing appeared only once in the analysis.

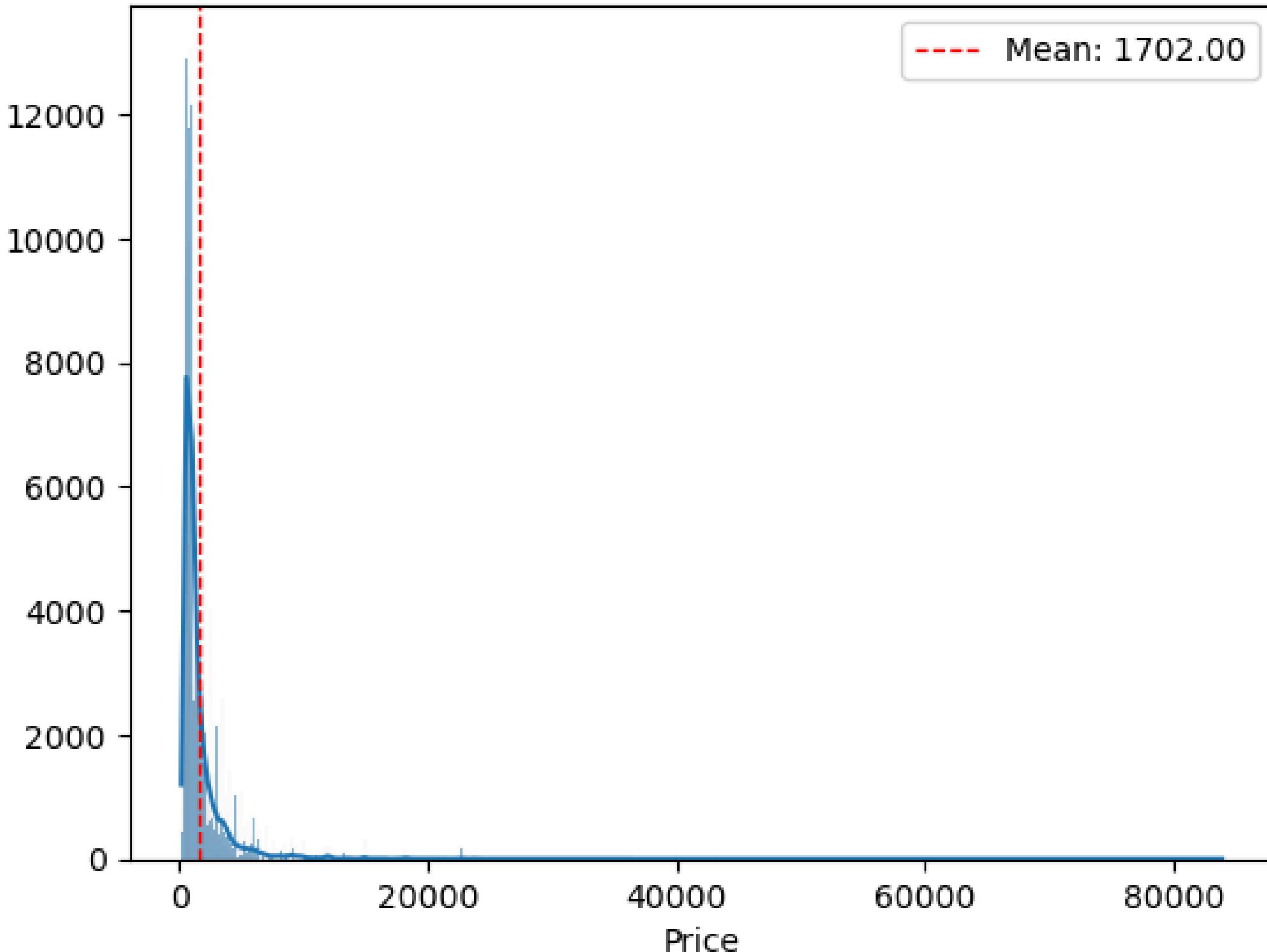
# EXPLORATORY DATA ANALYSIS

## INSIGHTS

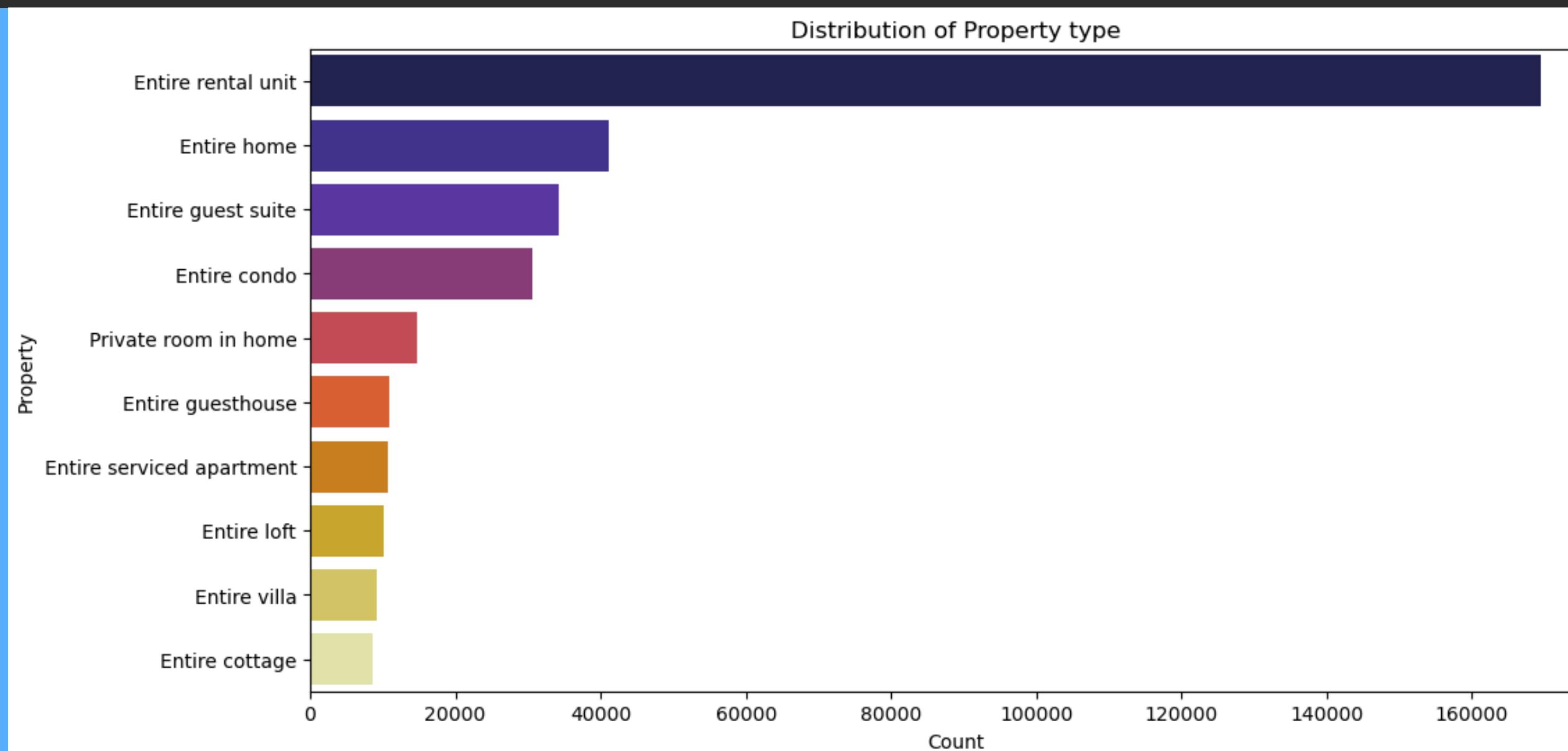
THE KDE OF THE HISTOGRAM ABOVE  
SHOWS THAT THE AVERAGE PRICE OF  
AIRBNBS IN CAPE TOWN IS 1702 ZAR.

Distribution of Price

Mean: 1702.00



# EXPLORATORY DATA ANALYSIS



## INSIGHTS

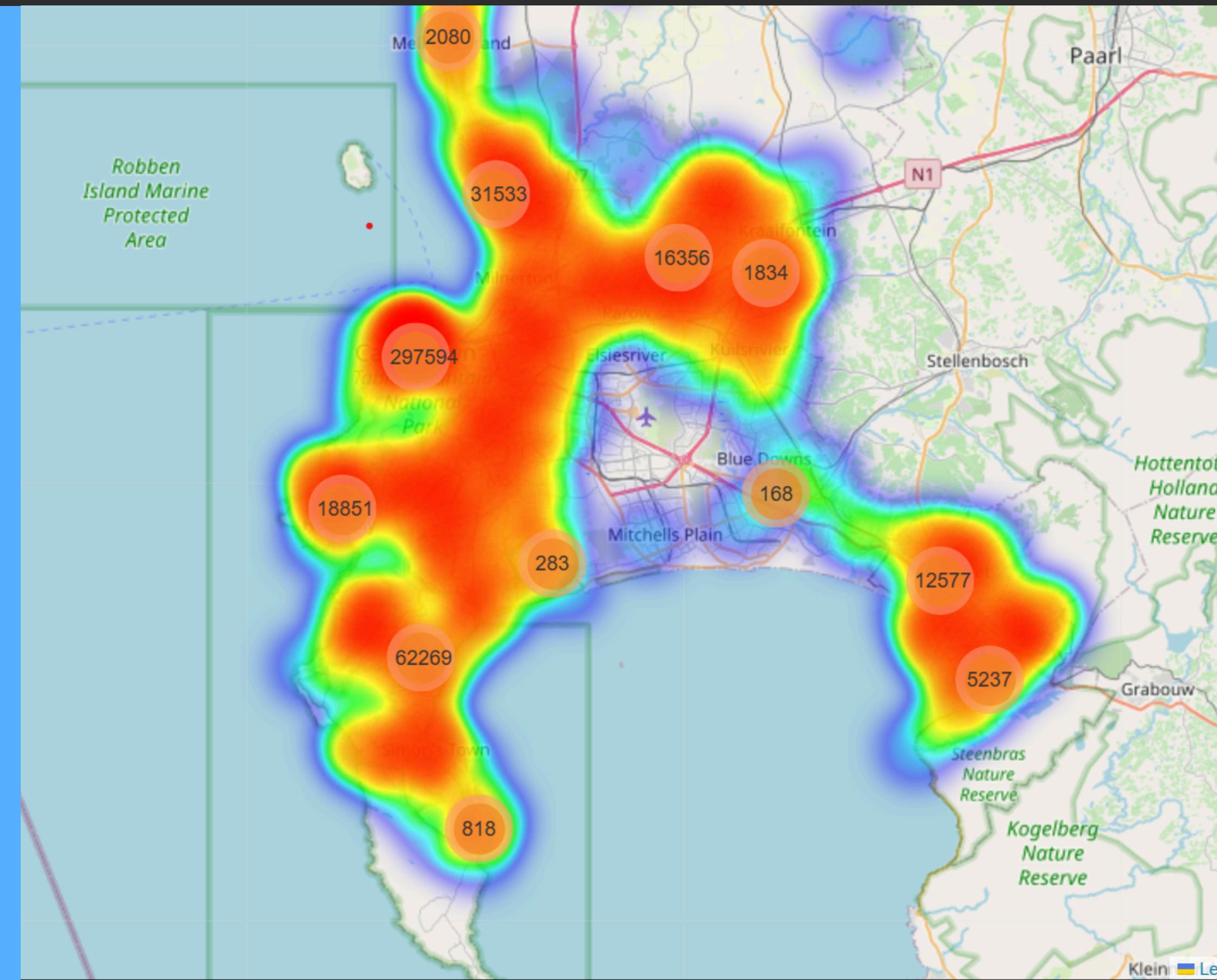
THE BAR CHART ABOVE SHOWS THAT THE MAJORITY OF AIRBNB LISTINGS IN CAPE TOWN ARE ENTIRE RENTAL UNITS (213788), FOLLOWED BY ENTIRE HOMES (48651), WHILE OTHER PROPERTY TYPES LIKE GUEST SUITES, CONDOS, AND PRIVATE ROOMS HAVE SIGNIFICANTLY LOWER REPRESENTATION.

## EXPLORATORY DATA ANALYSIS

### INSIGHTS

THE HEATMAP ABOVE SHOWS THAT THE HIGHEST CONCENTRATION (297594 LISTINGS) APPEARS TO BE IN THE CENTRAL CAPE TOWN AREA.

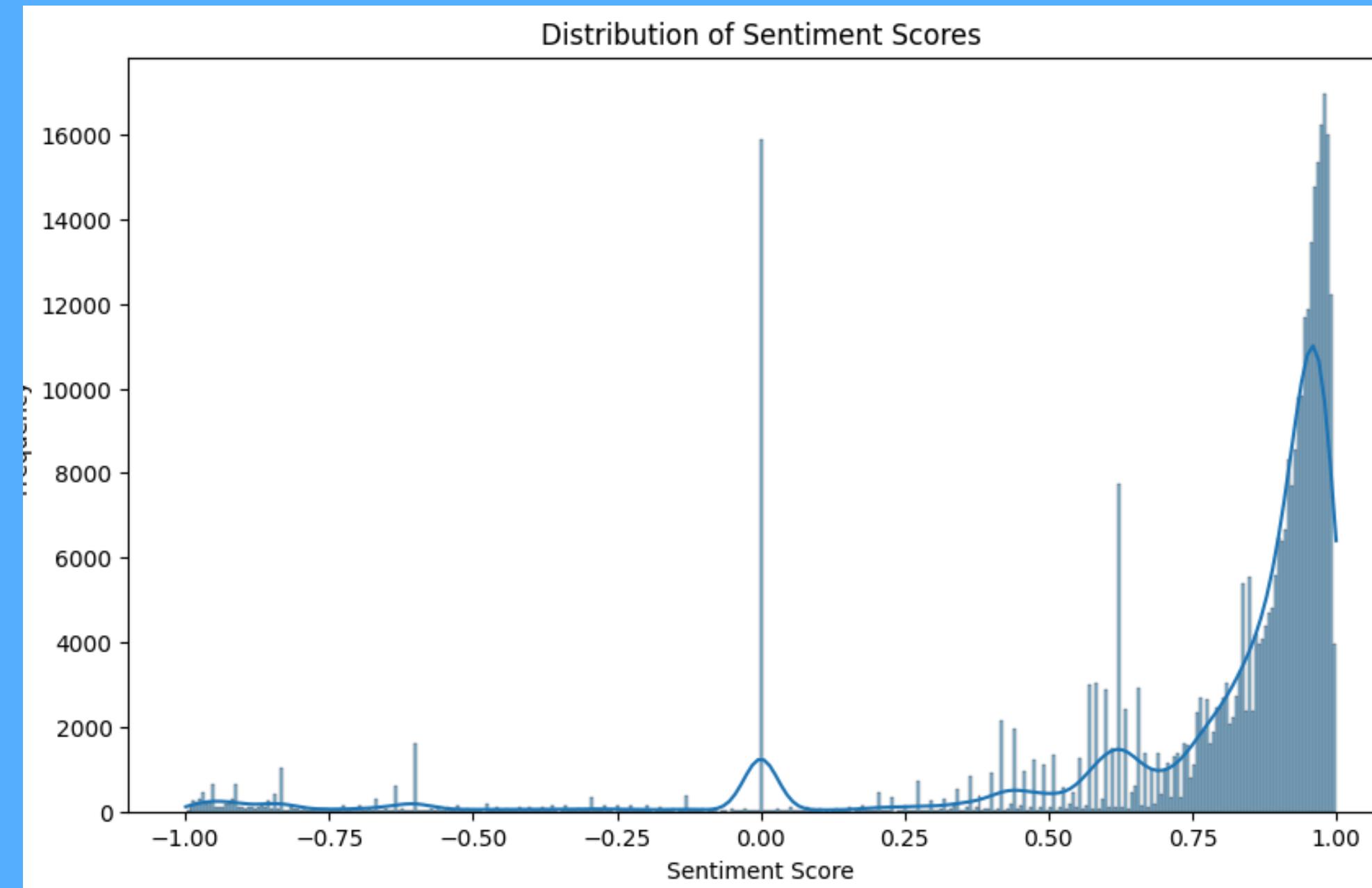
-AIRBNBS ARE HEAVILY CONCENTRATED IN THE TOURIST-FRIENDLY COASTAL AREAS AND CITY CENTER, WITH SIGNIFICANTLY LESS ACTIVITY IN THE INLAND REGIONS.



# EXPLORATORY DATA ANALYSIS

## INSIGHTS

THE PLOT REVEALS A SIGNIFICANT SKEW TOWARD POSITIVE SENTIMENT, WITH THE MAJORITY OF COMMENTS FALLING BETWEEN 0.5 AND 1. THIS INDICATES THAT MOST COMMENTS EXPRESS POSITIVE SENTIMENTS. A NOTICEABLE SPIKE AROUND THE 0 SCORE SUGGESTS THAT MANY COMMENTS ARE NEUTRAL OR LACK STRONG SENTIMENT. NEGATIVE SENTIMENT SCORES APPEAR FAR LESS FREQUENT, INDICATING RELATIVELY FEW STRONGLY NEGATIVE COMMENTS.



# EXPLORATORY DATA ANALYSIS

Model	Train RMSE	Test RMSE	R <sup>2</sup> Score
Baseline Model: Linear Regression	0.67	0.76	0.5961
Linear Regression (with PCA)	0.71	0.79	0.5721
Decision Tree Model	0.61	0.70	0.6412
Random Forest	0.17	0.57	0.7408
KNN Regression Model	0.56	0.68	0.6616

Model	Train RMSE	Test RMSE	R2 Score
Tuned KNN Model (with Grid Search)	0.54	0.68	0.6621
XGBoost Model	0.43	0.55	0.7579
LightGBM Model	0.32	0.54	0.7699
Neural Network Model	0.56	0.64	0.6900

**WITH THE FOLLOWING LINK CAPECASA RECOMMENDER SYSTEM HELPS YOU FIND THE BEST PRICE FOR STAY IN CAPE TOWN**

[HTTPS://CAPECASA-AIRBNB.STREAMLIT.APP/](https://capecasa-airbnb.streamlit.app/)



*the* *end*

The image shows a scenic coastal town built into a hillside overlooking a large bay. In the background, a range of mountains is visible under a clear blue sky. The word "the" is written in a white, cursive font, and the word "end" is written in a white, wavy, hand-drawn style, with an arrow pointing towards the right side of the image.