

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green. They are positioned diagonally, with the blue one partially covering the green one.

Air BnB Occupancy Rate Optimization

By Stephen Waweru



Overview

Milkah Petso asked me to find the optimum neighbourhood to open an Air Bnb in Western Cape Region in Cape Town, South Africa.

She also would like to know the factors that most affect her occupancy rate to assist in making the best marketing strategy for her Air BnB.

The data i used in this project was sourced from The [Inside Air BnB](#) platform covering an average of 12 months in the year 2024.



Data Glossary

Listing : a property listed on the Air BnB platform

number_of_reviews : Number of reviews in the 12 month period

occupancy rate: this is calculated as the percentage of days the listing was occupied capped at 70%. The rate also depends on the number of reviews and the minimum nights as imposed by the metrics of the Air BnB algorithm

Minimum nights : the minimum number of nights listing can be taken

Room_type : this is the type of room on the listed property e.g single private room, entire home or apartment, studio etc



Data Cleaning and Feature Engineering

Data cleaning involves eliminating or filling null values from the dataframe to avoid corrupting our results.

- Dropping null values
- Imputing the null values with median and mean values

Feature engineering involves configuring the categorical values such as room type to hold numerical values that represent them in the models we will use to make the predictions.

- One Hot Encoding
- Label Encoding
- Categorizing our occupancy rate into Low, Medium and High



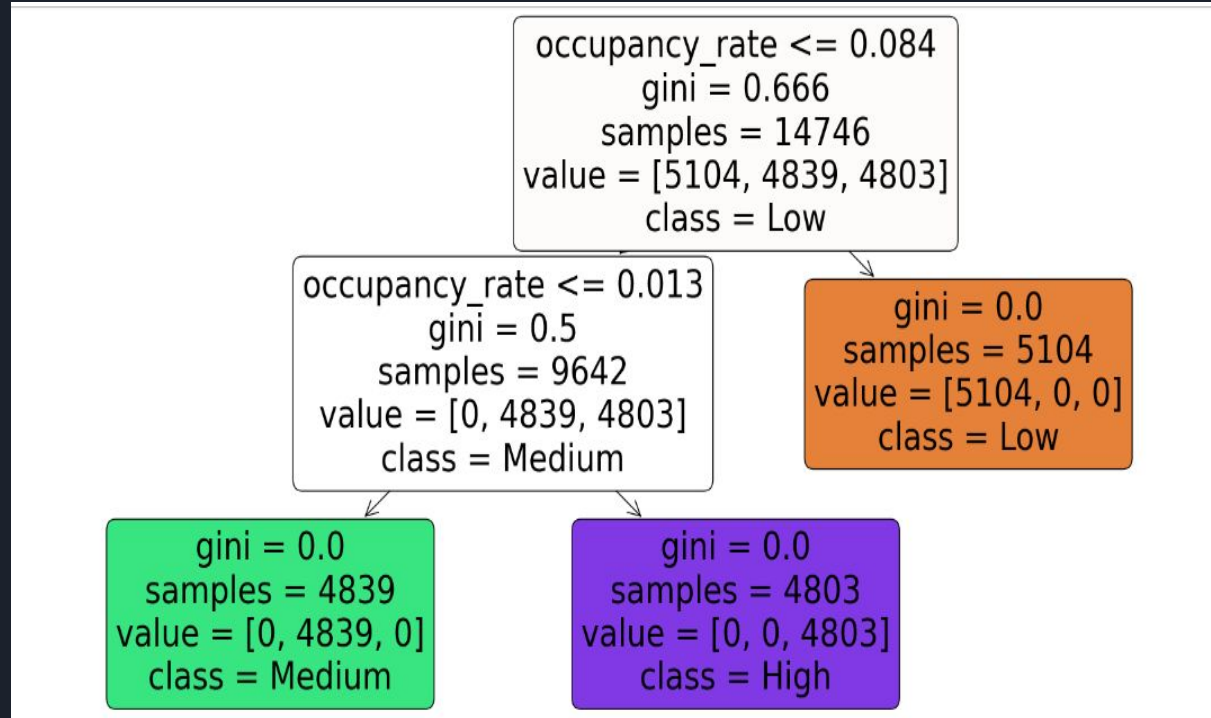
Models Used

1. **Decision Tree Classifier** :- A model that predicts outcomes by splitting data into branches based on feature values, forming a tree-like structure.
2. **Logistic Regression**: - A model that predicts the probability of a binary outcome using a logistic function based on input features.
3. **Random Forest Classifier**: - An ensemble model that combines multiple decision trees to improve predictive accuracy and reduce overfitting.

Decision Tree Classifier

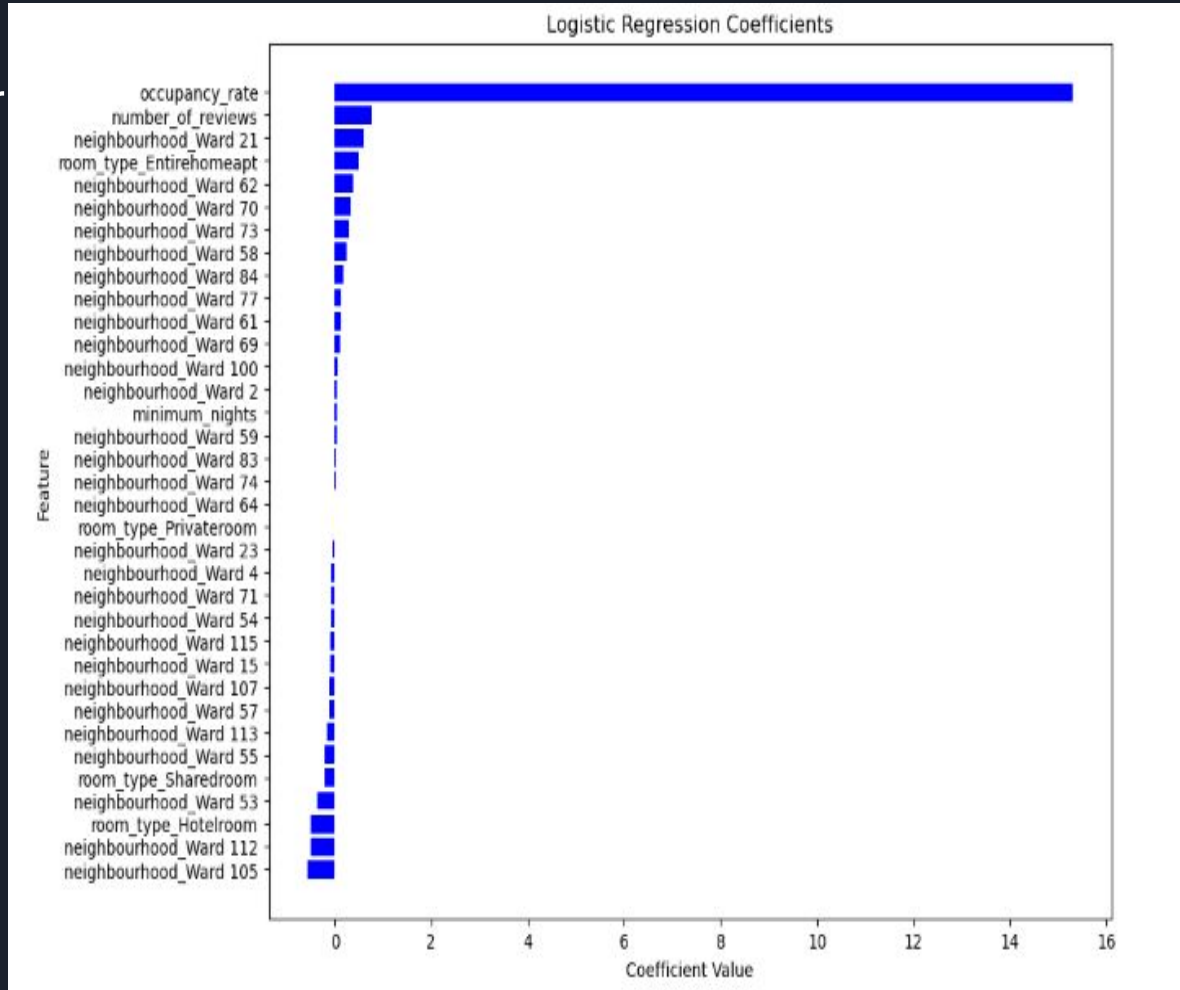
The decision tree model visualised shows the distribution of the occupancy rate among the listings in the dataset.

High and Medium are evenly matched and Low seems to have the most listings indicating a competitive and thereby growing market in Western Cape



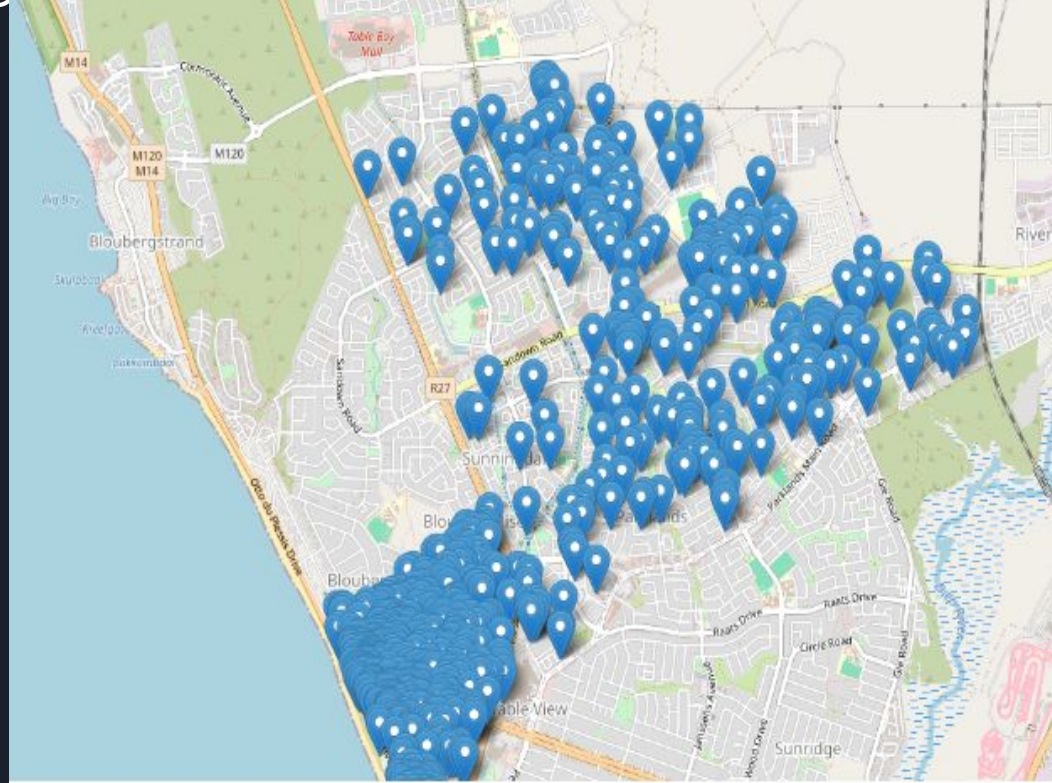
Logistic Regr

The logistic regression coefficient plot shows the feature importance when it comes to occupancy rate and indicates that number of reviews and the neighbourhood are the most essential determinants of a high occupancy rate



Random Forest Classifier

The Random Forest Classifier contains several decision trees and was used to predict the optimum neighbourhood to create a listing within Western Cape and Ward 107 was picked to have a 'High' occupancy rate i.e a suppressed mean of 28% consistently for the next 12 months. The illustration shows 107 with all the listings in its vicinity. Note the crowding near the beach.





Recommendations

1. Encourage reviews especially positive ones by the guests to maximise occupancy push by the Air BnB algorithm
2. List a property within the neighbourhood indicated as Ward 107 preferably an entire apartment or home
3. Dynamic pricing, target marketing, loyalty programs, partner with neighbouring businesses for discounted services, advertise on several channels and constantly remodel with maintenance of amenities.
4. Conduct further modelling with additional data in the coming months to keep abreast with any needed changes.



Question And Answer Section

Thank you for viewing my presentation.

My LinkedIn Profile: [Stephen.Waweru](#)