Solutions to hosts of AIRBNB in Amsterdam by predicting rental prices

Introduction/Business Problem:

Airbnb is a marketplace for short term rentals, allowing you to list part or all of your living space for others to rent. The company itself has grown rapidly from its founding in 2008 to a 30 billion dollar valuation in 2016 and is currently worth more than any hotel chain in the world. One challenge that Airbnb hosts face is determining the optimal nightly rent price.

In many areas, renters are presented with a good selection of listings and can filter on criteria like price, number of bedrooms, room type, and more. Since Airbnb is a marketplace, the amount a host can charge on a nightly basis is closely linked to the dynamics of the marketplace.

The Challenge:

As hosts, if we try to charge above market price then renters will select more affordable alternatives. If we set our nightly rent price too low, we'll miss out on potential revenue.

Stakeholders :

The hosts of the Airbnb properties.

Data

While Airbnb doesn't release any data on the listings in their marketplace, a separate group named ‘Inside Airbnb’ has extracted data on a sample of the listings for many of the major cities on the website. In this report, I’ll be working with their data set on the listings from Amsterdam, the capital of the Netherlands. Each row in the data set is a specific listing that's available for renting on Airbnb in the Amsterdam area.

Strategy

One strategy I could use is to:

1. Find a few listings that are similar to mine,
2. Average the listed price for the ones most similar to mine,
3. Set my listing price to this calculated average price.

I’m going to build a machine learning model to automate this process using **k-nearest neighbors**.

Some of the more important columns:

1. *accommodates:* the number of guests the rental can accommodate
2. *bedrooms*: number of bedrooms included in the rental
3. *bathrooms*: number of bathrooms included in the rental
4. *beds*: number of beds included in the rental
5. *price*: nightly price for the rental
6. *minimum_nights*: minimum number of nights a guest can stay for the rental
7. *maximum_nights*: maximum number of nights a guest can stay for the rental
8. *number_of_reviews*: number of reviews that previous guests have left

In addition to the above, I will make use of the Foursquare location data to get information about the neighbourhood.

Methodolgy :

I have used the following key items to understand better and comprehensively how I can predict the price for a new listing to ensure it does not surpass 'maximum pricing' and also does not incur huge loss at the same time.

1. Number of people that can be accommodated
2. Number of bedrooms
3. Type of room : Private, Entire home, shared
4. Minimum stay
5. Reviews
6. Latitude
7. Longitude

For the type of room I have used the following identifiers:

0 for Private room
1 for Entire home
2 for Shared home

I have created a feature with the above 7 crucial factors.

Train/Test Split:

My data has **5614 rows**, I have done 0.7/0.3 split.

Why KNN?

The machine learning technique that I have used is KNN. To think back to my strategy of finding listings similar to mine, average their price and set my listing price as that average is basically the essence of KNN. Out of the 3929 rows in the training set, I noticed that if I ran the simulation with 100 neighbours I got the best accuracy with 65.

This meant that broadly I can divide the training data into 65 neighbourhoods and then calculate the distance of these neighbourhoods from my test data to make the predictions on the test data.

Results:

The training set of 3929 rows gave 65 as the best k value, I ued the test data to find the accuracy.

The best accuracy 0.0777448071217
F1 score 0.049674745299619391
Jaccard similarity score 0 .06943620178041543

Discussion:

For the new listings using the predict() function I can get accurate values of the new listings. If the features are increased by including those that further impact the target variable which is the price then accuracy will see an increase. Features that I can think of which can be useful would be

1. Distance from public transport
2. Proximity to hospital/medical help
3. Availability of internet

Conclusion:

It was a learning experience to understand how difficult it is for a new listing to find an optimal pricing on Airbnb. Big cities like Amsterdam have no shortage of more meaningful features that can increase accuracy of new listings.