Berlin's Short-term Rental Market Analysis

Jun Liu

July 29, 2021

1 Introduction

1.1 Background Information

England takes pleasure in London, the U.S. loves New York, France is slightly narcissistic about Paris and, following suit, Germany proudly claims Berlin. More than two decades since the fall of the Berlin Wall, the city basks in a cultural renaissance that boasts everything from museums and fashion to food and nightclubs. All these make Berlin the ideal destination for tourists.

According to the reports in 2014, Berlin had 788 hotels with 134,399 beds[1], and the city recorded 28.7 million overnight hotel stays and 11.9 million hotel guests[2]. Tourism figures have more than doubled within the last ten years and Berlin has become the third-most-visited city destination in Europe[3].

1.2 Business Problem

With the aforementioned prospect, various stakeholders (entrepreneurs, investors or individual hosts) may be interested to explore the opportunity of short-term accommodation market in Berlin. This data science project is thus carried out to help them answer the following question:

"What is the best price to rent out a space?"

The aim of this project is to propose a data-driven solution, by using machine learning to predict the rental price. For the current project, a location based predictor will be introduced to the model: the property's proximity to certain venues. This will allow the model to put an implicit price on things such as living close to a bar or a supermarket, along with the property characteristics, such as number of rooms, bathrooms and extra services on offer.

2 Data Sources

Airbnb is one of the top internet marketplaces for short-term home and apartment rentals. It allows people to, for example, rent out their home for a week while they are away, or rent

[1] Berlin Welcomes Record Numbers of Tourists and Convention Participants in 2014

[2] Record number of visitors in Berlin

[3] Berlin Is the 'Post-Tourist' Capital of Europe

out their empty bedrooms. However Airbnb does not release any data on the listings in its marketplaces. Luckily a separate group named `Inside Airbnb`[4] scrapes and compiles publicly available information about many cities Airbnb's listings from the Airbnb website. For this project, their data set scraped on June 22, 2021, on the city of Berlin, Germany, is used. It contains information on all Berlin Airbnb listings that were live on the site on that date (about 36,000). Here's the direct link.

Each row in the data set is a listing available for rental in Airbnb's site for Berlin. The columns describe different characteristics of each listing. Some of the more important features this project will look into are the following:

- accommodates: the number of guests the rental can accommodate
- bedrooms: number of bedrooms included in the rental
- bathrooms: number of bathrooms included in the rental
- beds: number of beds included in the rental
- price: nightly price for the rental
- minimum_nights: minimum number of nights a guest can stay for the rental
- maximum_nights: maximum number of nights a guest can stay for the rental
- number_of_reviews: number of reviews that previous guests have left

To model the spatial relationship between Airbnb rental prices and property proximity to certain venues, we use the `Foursquare API`[5] to access the city's venues and the street network, available through `OpenStreepMap (OSM)`[6].

[4] Inside Airbnb: http://insideairbnb.com/get-the-data.html
[5] Foursquare API: https://developer.foursquare.com
[6] OpenStreepMap (OSM): https://www.openstreetmap.org