Louie Balderrama
Springboard Data Science Career Track, January 2019 cohort

Capstone Project 1: Project Proposal

Problem Statement: Predicting the nightly rate of Airbnb listings in Austin, TX

Background

Airbnb has become the biggest disrupter of the hospitality and tourism industry. It offers a global platform for private individuals willing to share a space to transact with would-be guests at a nightly rate pre-set by the host. The study will take a look at the different aspects of Airbnb listings and see what combination of features will most accurately predict the nightly rate of a given listing. The study will focus on the city of Austin, the tech-capital of Texas, and will make use of the features already in the Airbnb listings dataset. As a supplement, the study will also incorporate data on the crimes reported within the areas covered by the Airbnb dataset.

Impact

A reliable model can benefit both hosts and guests because the model would be able to provide a standard fair price that is appropriate for what the host can offer. Hosts would have a good rule of thumb in pricing a nightly stay given the nature of the space (lodging type, number of rooms, amenities, etc.), the location (number of crimes reported in the area, serious crimes prevalent in the area), the nature of the booking offered (cancellation policy, fees for extra guests), and the details about the hosts themselves ("superhost" status, types of verifications provided). A host would therefore minimize opportunity cost by not missing out on bookings from guests who would have otherwise been offered a fair price. On the same token, guests would have a reliable indicator of what a fair deal would look like. Even the industry itself could see a positive impact from an optimized, competitive host-guest marketplace. The limitation of the study, however, other than a limited geographical scope, is that the surge in demand from popular events and its effect on price will not be covered. Conventional rates are assumed.

Datasets

a. Airbnb Listings

The dataset comes from [Inside Airbnb](), an independent, third-party website that has its own set of open-source tools that encourage users to explore the negative and positive impacts of Airbnb. They scrape Airbnb listings from different metropolitan areas around the world, strip personal identification (like the full name of hosts and guests), and share the raw data on the site. The dataset used in this study is the "listings.csv" for Austin, TX as compiled on November 2018. There are 11,919 entries which represent all the Airbnb listings in the city at the time. The dataset is both rich with information and quite raw and messy, having a total of 96 columns. Some of the useful details include the listing price, the listing type (private bedroom, shared bedroom, or entire home), the location of the listing, review score of the listing (aggregated guest ratings), review score of the host, etc.

b. Austin, TX Crime

The crime dataset was retrieved from the City of Austin's [official data portal](). This dataset details all the reported crimes that occurred in Austin during 2018 – the timeframe that coincides with the Airbnb dataset. Though there are a total of 102,663 entries, only around 36,000 crimes have real values (i.e. non NaNs) under their "Category Description" series. This classifies major crimes according to the following seven categories recommended by the FBI's Uniform Crime Reporting program: Theft, Auto Theft, Burglary, Robbery, Rape, Aggravated Assault, and Murder. This categorization will then help in getting a sense of what major crimes occur in which areas, while the dataset as a whole will provide the aggregate statistics of all crimes (both petty and serious) in Austin.

c. Zip Codes GeoJSON

This geospatial JSON file, also retrieved from Austin's [official data portal](), contains the boundary coordinates of every zip code within the city. When combined with the findings of the other datasets, this file will aid in the visualization of trends by location, including where the most expensive Airbnb listings are and where crimes are reported the most.