Predicting nightly rates for airbnb properties in New York

Capstone Two: Project Proposal by Deepti Jindal

Problem statement

What are key factors used in predicting the nightly rate of Airbnb listings in New York, NY that affect nightly rates.

Context

Airbnb has disrupted the hospitality and tourism industry. Airbnb is a platform that allows individuals to provide a space at a nightly rate pre-set by the host. This 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 New York city, and will make use of the features already in the Airbnb listings dataset as well as the zipcode the property is located.

Hosts would have a good model for 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. Conventional rates are assumed.

Criteria for success

Find the most important features that highly influence the nightly rate and how certain features relate to one another as well as to our target variable.

Scope of solution space

Identify what are the factors that affect the nightly rates for airbnb properties and then predict a property’s nightly rate based on the features.

Constraints

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. In addition, the information given is currently a single csv file. The dataset used in this study is the “listings.csv.gz” for New York, NY as compiled on October 1, 2023.

Impact

A reliable model can provide key information for both the hosts and guests because the model would provide a fair price for what the host can offer.

Key Data sources

The data source for this analysis is from InsiderAirbnb. This is an independent third-party website that encourage users to explore the negative and positive impacts of Airbnb.