Airbnb Price Prediction

11/27/2024

—

Tina Nosrati
Fall 2024
CMPS 240
Professor Lawrence D'Antonio

Overview

This project focuses on analyzing Airbnb prices in Athens using visualizations and descriptive statistics to uncover key insights. Additionally, hypothesis testing and regression analysis will be conducted to explore and understand relationships within the data.

Problem Definition

This project explores what drives Airbnb prices in Athens during weekends, aiming to uncover the factors that matter most to pricing. Using the Athens Weekends dataset, which includes information about properties, locations, hosts, and guest feedback, we'll dive into the data with visualizations and simple statistics to spot patterns. We'll also use hypothesis testing and regression analysis to dig deeper into questions like how room type, location, or Superhost status affects prices. The goal is to provide helpful insights for hosts to set competitive prices and for travelers to better understand what influences costs in Athens.

Research Questions

1. What factors influence Airbnb prices in Athens during weekends?
2. How do property features, such as room type or number of bedrooms, affect pricing?
3. How do host characteristics, such as Superhost status or response rate, impact pricing?
4. Which attributes are the strongest predictors of listing prices?

Data [Link]

This dataset focuses on weekend Airbnb prices in Athens, analyzing factors like room type, guest satisfaction, bedrooms, distance from the city center, and Superhost status to uncover key pricing determinants.

Column name	Description
realSum	The total price of the Airbnb listing. (Numeric)
room_type	The type of room being offered (e.g. private, shared, etc.). (Categorical)
room_shared	Whether the room is shared or not. (Boolean)
room_private	Whether the room is private or not. (Boolean)
person_capacity	The maximum number of people that can stay in the room. (Numeric)
host_is_superhost	Whether the host is a superhost or not. (Boolean)
multi	Whether the listing is for multiple rooms or not. (Boolean)
biz	Whether the listing is for business purposes or not. (Boolean)
cleanliness_rating	The cleanliness rating of the listing. (Numeric)
guest_satisfaction_ov erall	The overall guest satisfaction rating of the listing. (Numeric)
bedrooms	The number of bedrooms in the listing. (Numeric)
dist	The distance from the city center. (Numeric)
metro_dist	The distance from the nearest metro station. (Numeric)
lng	The longitude of the listing. (Numeric)
lat	The latitude of the listing. (Numeric)