

# Airbnb Price Regression Modeling

Bárbara Flores

## Introduction

The objective of this project is to generate a model that allows setting prices for Airbnb ads in Asheville, North Carolina.

To achieve this goal, we have a database of various Airbnb rental listings in the city, which contains detailed information about the listings, such as price, number of rooms, amenities, number of bathrooms, property location, etc. This database was extracted in June 2023 by the company Inside Airbnb and contains information about 3239 rental listings in Asheville, North Carolina. Therefore, the conclusions drawn are based on this time period.

To develop a model that enables us to set prices for Airbnb ads in Asheville, we will use a linear regression approach. Linear regression is a statistical technique used to model the relationship between a dependent variable and one or more independent variables.

## Linear regression

Linear regression is a statistical technique used to model the relationship between a dependent variable and one or more independent variables. In this context, the dependent variable would be the price of Airbnb ads, while independent variables may include features such as the number of rooms, amenities, number of bathrooms, and property location, among others.

## Conclusion

throomthroom

# Airbnb Price Regression Modeling

Bárbara Flores

## Introduction

The objective of this project is to generate a linear regression model that allows setting prices for Airbnb ads in Asheville, North Carolina.

## Dataset

We have a database of various Airbnb rental listings in the city, obtained through web scraping, which contains detailed information about the listings, such as price, number of rooms, amenities, number of bathrooms, property location, etc. The original database includes **75 variables**. This dataset was extracted in June 2023 by the company Inside Airbnb and contains information about **3239** rental listings in Asheville, North Carolina. Therefore, the conclusions drawn are based on this time period.

The database and dictionary can be downloaded from [listings.csv](#) and [Inside Airbnb Data Dictionary](#)

## Data Cleaning

First, the database was cleaned to be used for a linear regression model. Variables that did not contribute information to the model were excluded, such as: “id”, “listing\_url”, “scrape\_id”, “last\_scraped”, “source”, “name”, “description”, “neighborhood\_overview”, “picture\_url”, “host\_id”, “host\_url”, “host\_name”, “host\_location”, “host\_about”, etc.

Secondly, certain transformations were performed on some variables to handle them within the model, such as “host\_response\_time”, “host\_response\_rate”, “host\_acceptance\_rate”, “host\_is\_superhost”, “neighbourhood”, “latitude”, “longitude”, “property\_type”, “room\_type”, “bathrooms\_text”, “amenities” and “price”.

It is worth mentioning that the ‘**amenities**’ variable considered combinations of 2115 different amenities, so the following transformation was performed: new binary variables were generated to group the most frequent amenities. For example, ‘**security**’ considers ‘smoke and fire alarms’, ‘fire extinguisher’, and ‘first aid kit’. ‘**kitchen\_amenities**’ includes ‘dishes and silverware’, ‘microwave’, ‘kitchen’, ‘refrigerator’, ‘cooking basics’, ‘coffee’, ‘freezer’, ‘wine glasses’, ‘oven’, ‘toaster’, and ‘dining table’.

**Conclusion**

**Conclusion**