

DSC520 Week8/9 Exercise 8.3

Anjani Bonda

February 12th 2022

Project: Impact of AirBnB on rental home prices in Columbus,OH.

Introduction

Airbnb is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. Airbnb does not own any of the listed properties; instead, it profits by receiving commission from each booking. The company has been criticized for possibly driving up home rents and a direct correlation between listings and rental prices and creating nuisances for those living near leased properties.

I would like to address the impact of Columbus(OH) Airbnb rentals on the nearby rental prices using Data Science algorithm in predicting the rental prices of Columbus and understand the correlation between them.

Research Questions

1. How are the Airbnb rental prices for different neighborhoods in Columbus ?
2. What's the correlation between Airbnb prices and Columbus neighborhood rentals ?
3. What kind of homes are most rented on Airbnb ?
4. What are the average rental prices on Airbnb by neighborhood ?
5. What are the average rental prices by Columbus neighborhood ?
6. How much profit does Airbnb possibly make ?

Approach

Analyze data to discover correlations and patterns and create a model to predict how the Airbnb rentals affect the nearby neighborhood prices in Columbus using various key factors like zip code, county/neighborhood, prices, reviews, rental units etc.

1. Find the key predictors for the regression model.
2. Calculate R2, Adjusted R2, p-value, betas etc
3. Check confidence intervals if the model is likely to be representative of the true population.
4. Perform variance, residuals, leverage, cooks distance, covariance ratio analysis on the model(s).
5. Check if model is unbiased and use it for predictions.

How this approach addresses (fully or partially) the problem

This approach is based on the learnings in the course. Focus is to get enough data points to be able to address the problem. The predictors for the model and then the various calculations involved in the approach will give enough data points to analyze and draw conclusions on the correlations, model performance and if the model is unbiased to be able to use effectively for the predictions.

Data (Minimun of 3 datasets - but no requirement on number of fields/rows)

1. Airbnb listing data for Columbus,OH as of December 23, 2021. Data Source: <http://insideairbnb.com/get-the-data.html>
2. Columbus Housing Data Data Source: <https://opendata.columbus.gov/>
3. Average rent Columbus, OH. Data Source: <https://www.zumper.com/rent-research/columbus-oh>

Required Packages

Some of these packages might be required for data transformations, regression diagnostics, different statistical tests, plotting, visualizations and evaluations.

1. dplyr
2. purrr
3. QuantPsyc
4. lm.test
5. car
6. ggplot2
7. qqplotr
8. rcmdr

Plots and Table needs

1. Histogram
2. Scatter Plot
3. QQ Plot
4. Density Plot

Questions for future steps

1. How to ensure the data quality for the analysis ?
2. Can a different model be used for predictions ?