G12: Towards understanding Airbnb rental pricing

Sayar Ghosh Roy (sayar3), Shradha Sehgal (ssehgal4), Aditya Sinha (as146)

Introduction: We aim to build multiple data mining models for understanding the underlying factors driving rental prices of Airbnb properties. Also, we aim to develop supervised models capable of accurately predicting such rental prices.

Motivation: Airbnb has served over 60M individuals looking for convenient yet affordable housing options. Present in over 34K cities, Airbnb allows homeowners to rent out their properties for short periods, thereby offering accommodation seekers an alternative to typical hotels. Thus, understanding the pricing dynamics for Airbnbs is of utmost importance, both for Airbnb hosts and potential guests. Note that our proposed modeling principles make use of pricing data conditioned on location and property type and, as such, would be applicable to any housing market (long-term rentals, properties for sale).

Data: The raw dataset can be found here. It contains 47,906 entries with features such as the neighborhood, latitude, longitude, room type, and price (the dependent variable for our predictive models). In a nutshell, the data describes Airbnb listing activity and associated metrics in 2019 New York City. We will present detailed statistics in our full report.

Plan of work: We aim to build data mining models to understand the factors that drive prices for short-term rentals in NYC's booming real estate market. We can divide our project goals into four broad sectors. (1) ***Prelims***: As a preliminary step, we perform fundamental Exploratory Data Analysis (EDA) steps to understand our dataset's measures of central tendencies and dispersion, inter-feature correlations & covariances, etc. (2) ***Predictive Models***: Our primary aim is to fit multiple regression models to predict the expected rent of a property based on a set of features (primarily related to its location and property type). Within that, we experiment with various algorithms, including decision trees, random forests, gradient boosting, Ordinary Least Squares (OLS), and neural networks. We evaluate our regression algorithms using metrics such as MSE, MAE, etc. Also, in this space, we explore the impact of feature selection and feature extraction through widely-used dimensionality reduction techniques (such as PCA). (3) ***Clustering***: We explore various clustering techniques to search for potential interesting insights. We experiment with popular unsupervised clustering approaches (K-means, t-SNE), considering the available features. (4) ***Visualizations***: We experiment with multiple visualization techniques (including heatmaps) to showcase observable trends in our dataset.

Related Work: [1, 2] utilize Machine Learning models, including Decision Trees, Random Forests, and Gradient Boosted Regression Trees for predicting warehouse rental prices, offering valuable predictive insights. [3, 4] attempt to study underlying price determining factors using OLS and quantile regression models for holiday-related travel. [5] use sentiment analysis to factor in the importance of customer reviews available on Airbnb. [6] propose a multi-modal setup with reviews, features, and geographical information to create more reliable forecasting models.

Challenges: We anticipate running into two main challenges — first, ensuring that our predictive models do not overfit or underfit and remain balanced in terms of bias and variance, and second, adding an interpretability element to our neural regression models.

References

[1] H. Yu and J. Wu, "Real Estate Price Prediction with Regression and Classification", 2016.

[2] Y. Ma, Z. Zhang, A. Ihler and B. Pan "Estimating warehouse rental prices using machine learning techniques", International Journal of Computers, Communications Control, 2018.

[3] D. Wang, J. L. Nicolau, "Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on airbnb.com", International Journal on Hospitality Management, 2017.

[4] L. Masiero, J. L. Nicolau and R. Law, "A demand-driven analysis of tourist accommodation price: A quantile regression of room bookings", International Journal of Hospitality Management, 2015.

[5] P. Kalehbasti, L. Nikolenko, H. Rezaei, "Airbnb Price Prediction Using Machine Learning and Sentiment Analysis", arxiv:1907.12665, 2019

[6] N. Peng, K. Li, Y. Qin, "Leveraging Multi-Modality Data to Airbnb Price Prediction", In Proceedings of Second International Conference on Economic Management and Model Engineering, 2020.