Airbnb Price Prediction Models and the Impact of NYC's Local Law 18 on Airbnb Prices

Christina Hyman

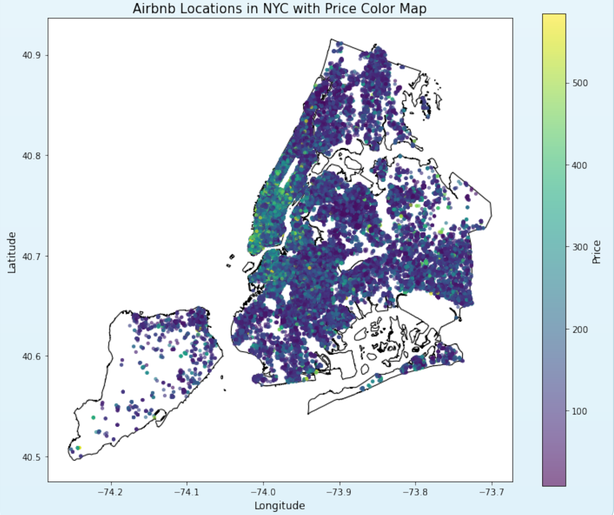
Seidenberg School of Computer Science & Information Systems

INTRODUCTION

- Airbnb is an online marketplace for short and long-term rental to provide a connection between hosts and guests, allowing hosts to list their accommodations and guests to book them given a price per night or month.
- In September 2023, Local Law 18, also known as the Short-Term Rental Registration Law, introduced strict guidelines for short-term rentals in NYC to address concerns related to illegal short-term rentals, ensure the safety of travelers, and alleviate pressure on the housing market in New York City
- Overall, price prediction models contribute to the efficacy and efficiency of the Airbnb marketplace, benefiting both hosts and guests.

OBJECTIVES

- Discover spatial patterns of Airbnb listings
- Provide insight to hosts on most influential features
- Build accurate machine learning predictive models that accurately predict Airbnb listing prices in NYC based on relevant features for enhanced user decision making
- Analyze the impact of Local Law 18 on Airbnb listing prices in NYC by conducting a Diff-in-Diff analysis



LITERATURE REVIEW

- Kalehbasti et al. (2021) - SVR with the RBF kernel; R^2 = 0.6901
- Thakur et al. (2022) - four-layered MLP to predict Airbnb prices in Rio de Janeiro, Brazil; R2 = 0.7443
- Lektorov et al. (2023) - Random Forests (R^2 = 0.95), Decision Trees (R^2 = 0.92), XGBoost Regressor (R^2 = 0.94)
- Masrom et al. (2022) - Decision Tree (R^2 = 0.997), Random Forest (R^2 = 0.998) to predict Airbnb prices in Singapore

DATA PREPROCESSING AND FEATUARE SELECTION

- Collected from Inside Airbnb between Feb-23 and April-24
- Duplicates were dropped by the subset {'id', 'host_name', 'latitude', 'longitude', 'price'}
- Missing values and outliers of prices were dropped
- Missing values and outliers for the remaining features were replaced by their respective medians
- Dropped redundant features (75 features → 19 features)
- Extracted features for bathrooms, amenities, and transit
- Selected features based on the Pearson coefficient being < -0.10 and > 0.10 with the price feature and confirmed selection with a p-value analysis at the α = 0.05 level
- 216039 observations and 23 features
- Applied logarithmic function to price target feature
- Applied StandardScaler( ) for the Support Vector Regressor

Category of Features	List of Features w. Data Type
Room Characteristics	longitude [-74.251907, -73.710870], accommodates (1 – 7), beds (0 – 3), minimum_nights (1 – 70), instant_bookable (0, 1), shared_bathroom (0, 1), bathrooms (1, 1.5, 2)
Type of Room	Entire home/apt (0, 1), Hotel room (0, 1), Private room (0, 1), Shared room (0, 1)
Amenities	has_ac (0, 1), has_essentials (0, 1), has_tv (0, 1), has_washer (0, 1), has_safety (0, 1), has_gym (0, 1), children_pet_friendly (0, 1), has_pool (0, 1)
Boroughs	Bronx (0, 1), Brooklyn (0, 1), Manhattan (0, 1), Queens (0, 1), Staten Island, (0, 1)

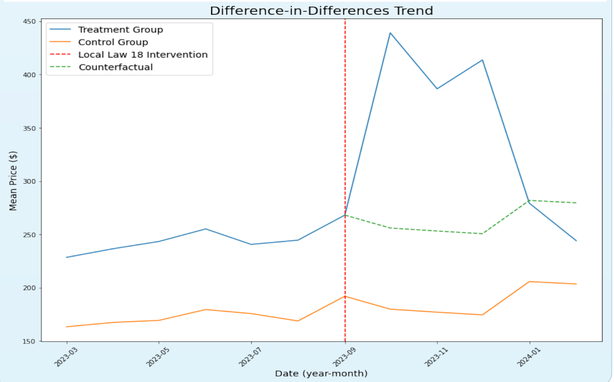
RESULTS AND FINDINGS OF PRICE PREDICTION MODELS

Model	Parameters	R^2	MSE	RMSE
Linear Regressor		0.5939	0.1819	0.4265
Support Vector Regressor	• kernel = 'rbf' • max_iter = 200000	0.7027	0.1332	0.3650
Decision Tree	• max_depth = 30 • max_leaf_nodes = 5000 • min_samples_leaf = 5	0.7373	0.1177	0.3431
Random Forest Regressor	• n_estimators = 100 • max_depth = 30 • max_leaf_nodes = 5000 • min_samples_leaf = 5 • random_state = 42	0.7736	0.1014	0.3184
XGBoost Regressor	• n_estimators = 100 • max_depth = 20 • eta = 0.05 • subsample = 0.875 • colsamplebytree = 0.875	0.8036	0.0880	0.2966

RESULTS AND FINDINGS OF DIFF-IN-DIFF ANALYSIS

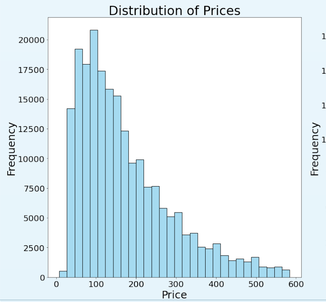
Difference-in-Differences (DiD) Regression Results

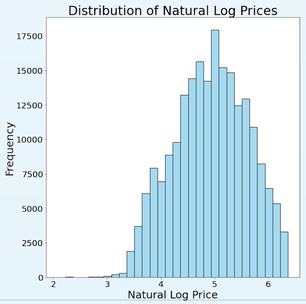
	coef	std err	t	P>|t|	[0.025	0.975]
const	174.8839	1.680	104.101	0.000	171.591	178.177
treatment	67.8111	2.826	23.993	0.000	62.272	73.351
post	13.1949	2.361	5.590	0.000	8.568	17.822
treatment_post	90.9741	6.840	13.301	0.000	77.569	104.379



CONCLUSIONS

- Applying the natural logarithmic function to the price feature increased the R^2 by 23.59% (from 0.4747 to 0.5867)
- Overall, I believe the Random Forest Regressor is the most superior price prediction model since it has the second highest test R2, it generalizes well with unseen data, and its computational resources are reasonable.
- The statistical significance of the `treatment_post` coefficient reveals that there was a substantial price change of short-term Airbnb listings in NYC after the implementation of Local Law 18.
- The findings suggest that average prices of short-term Airbnb listings are $90.97 higher than long-term Airbnb listings before the enforcement of Local Law 18.





REFERENCES

1. Airbnb. (2020, November 19). The amenities guests want - resource center.
2. Azmoudeh, A. (2022, August 1). Airbnb Open Data - New York Airbnb Open Data. Kaggle.
3. Deboosere, R., Kerrigan, D. J., Wachsmuth, D., & El-Geneidy, A. (2019). https://doi.org/10.1080/21681376.2019.1592699 \
4. Jiang, Y., Zhang, H., Cao, X., Wei, G., & Yang, Y. (2022). https://doi.org/10.1177/13548166221097585
5. Kalehbasti, P. R., Nikolenko, L., & Rezaei, H. (2021). https://doi.org/10.1007/978-3-030-84060-0_11
6. Lektorov, A., Abdelfattah, E., & Joshi, S. (2023). https://doi.org/10.1109/ccwc57344.2023.10099266
7. Masrom, S., Baharun, N., Razi, N. F. M., Rahman, R. A., & Rahman, A. S. A. (2022). https://doi.org/10.46338/ijetae0122_14
8. Short-Term Rental Registration and Verification by Booking Services. NYC Office of Special Enforcement (OSE). (2023). https://www.nyc.gov/site/specialenforcement/registration-law/registration.page
9. State of New York. MTA NYCT Subway Entrances and Exits: 2015 I State of New York. (updated August 2023). https://data.ny.gov/widgets/i9wp-a4ja
10. Thakur, N., Jain, R., Mahajan, A., & Islam, S. M. N. (2022). https://doi.org/10.1109/i2ct54291.2022.9824383
11. Voltes-Dorta, A., & Inchausti-Sintes, F. (2020). https://doi.org/10.1177/1354816619898075