Assessing the Impact of Airbnb's Plus Program: A Fixed Effects Analysis with Cross-Sectional and Time-Series Dimensions

Amir Eid, Leander Tim Engel, Aurel Roedern and Federico Scandizzo

Introduction

This study builds on prior research exploring the Airbnb Plus program's impact on **booking_rate**, which found no significant effect. Addressing limitations such as unobserved factors and missing fixed effects, it uses a Fixed Effects model to estimate the program's causal effect by comparing treated cities (Nashville, New Orleans, Washington DC, Denver) to controls, accounting for **employment_rate**, **price_mean**, and **year_built_to_now**. Additionally, the study examines the impact of the program on **listing_avg_review**, testing whether it increases customer satisfaction in treated cities compared to controls, after controlling for city-specific and time-specific effects.

1. Refining Hypotheses

The **first hypothesis** is, therefore, the following:

- H_0 : The introduction of the Airbnb Plus program does not significantly affect booking rates in treated cities compared to control cities, after accounting for city-specific and time-specific effects (i.e., *β3 = 0*, where β3 is the coefficient of the interaction term).
- H_a : The introduction of the Airbnb Plus program does significantly affect booking rates in treated cities compared to control cities, after accounting for city-specific and time-specific

effects (i.e.,$\beta 3 \neq 0$).

$$\text{booking\_rate}_{it} = \alpha_i + \gamma_t + \beta_3(Plus_i \times Treated_i) + \epsilon_{it}$$

The **second hypothesis** is formulated as follows:

- H_0 : The introduction of the Airbnb Plus program does not increase listing average reviews in treated cities compared to control cities, after accounting for city-specific and time-specific effects (i.e., $\beta 3 \leq 0$, where β3 is the coefficient of the interaction term).
- H_a : The introduction of the Airbnb Plus program increases listing average reviews in treated cities compared to control cities, after accounting for city-specific and time-specific effects (i.e., $\beta 3 > 0$).

$$\text{listing\_avg\_review}_{it} = \alpha_i + \gamma_t + \beta_3(Plus_i \times Treated_i) + \epsilon_{it}$$

Where:

- α_i: City fixed effects
- γ_t: Time fixed effects, affecting all cities equally
- **group**: Indicator variable for whether cities are in the **treated** or **control** group
- $\beta_3(\textbf{Plus} \times \textbf{Treated})$: Interaction term capturing the DiD effect
- ϵ_{it}: Error term, capturing random noise or unobserved factors varying across cities and over time

2. Dataset Overview, Data Quality & Model-Free Investigation

As the Airbnb Plus program is implemented in five cities, five others will serve as controls. San Francisco, part of the pilot phase, is excluded to avoid bias. The dataset includes 14604 observations across 99 variables, covering 773 zip codes in 11 U.S. cities from August 2017 (`timeperiod` = 8) to October 2019 (`timeperiod` = 34). Key variables include **performance metrics** (`booking_rate`, `listing_avg_reviews`) and **secondary variables** (`zipcode`, `timeperiod`, `city_number`, `policy_entry`, `employment_rate`, `price_mean`, `year_built_to_now`). Additional variables created for analysis are:

- **`city_name`**: maps `city_number` to city names or *Others*.
- **`date`**: converts `timeperiod` to year-month format.
- **`group`**: distinguishes treated and control cities.
- **`time`**: flags pre- and post-Plus program periods.

To begin assessing the quality of the data-set, the following shows a quick inspection, showing the structure of the data.

Table 1: Preview of the data

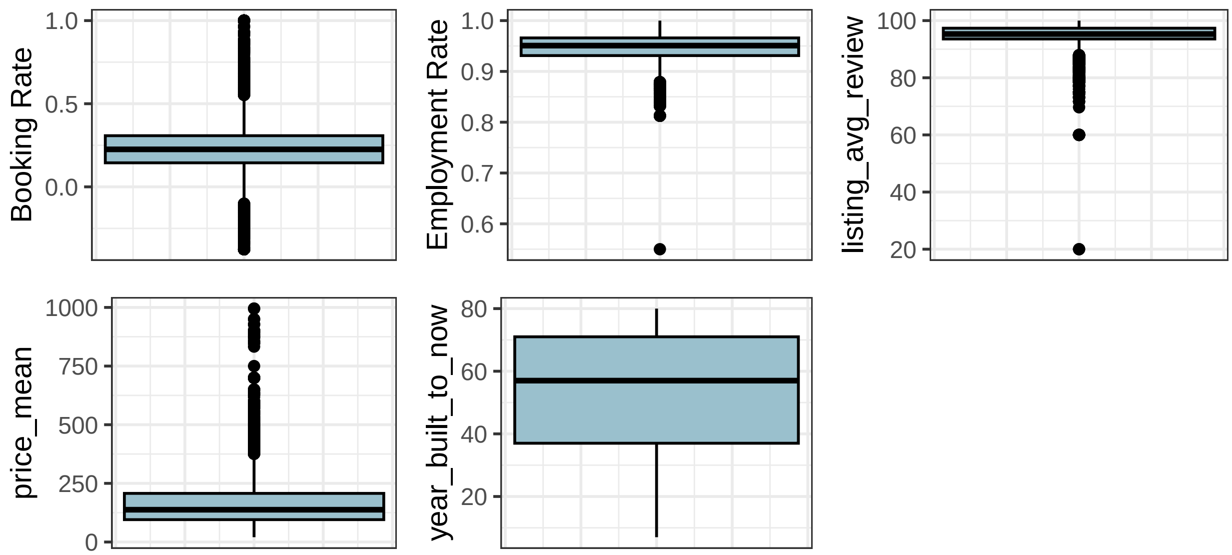
booking_rate	zipcode	timeperiod	date	city_number	city_name	policy_entry	employment_rate	price_mean	listing_avg_review	year_built_to_now
NA	1217	23	2018-11-01	1	Others	0	NA	132.5000	-1	NA
0.7500000	1217	24	2018-12-01	1	Others	0	NA	158.0000	-1	NA
NA	1217	25	2019-01-01	1	Others	0	NA	179.5000	-1	NA
-0.1250000	1217	26	2019-02-01	1	Others	0	NA	212.0000	-1	NA
0.8666667	1217	27	2019-03-01	1	Others	0	NA	207.3333	-1	NA
NA	1217	28	2019-04-01	1	Others	0	NA	207.0000	-1	NA

Missing Values

During the data cleaning process, a total of 1760 missing values were identified across various variables, including **booking_rate**, **employment_rate**, **price_mean**, **listing_avg_review**, **year_built_to_now**. For **booking_rate**, both zero and negative values were retained, as the variable reflects the ratio of bookings (positive values) to cancellations (negative values). For **listing_avg_review**, negative values (-1) were treated as an indicator for a booking having no reviews yet due to being new. These were treated as non-random missing values. For this reason, instead of being replaced with the mean, they were dropped to not introduce bias to the research. Instead, actual missing values were treated as random ones and were replace with the mean of the city. Similarly, for **price_mean**, zeroes were found and were dropped as these were likely input because there were no listing prices in that specific zip code to compute the price mean. The other variables did not have any zeroes or negative values, their actual missing values were treated as random missing data and replaced with their city mean. This approach was chosen to prevent the removal of rows, which could have introduced gaps in time for certain zip codes and cities. By using city-specific means, we minimized bias and retained as much relevant data as possible.

Each variable was carefully examined, and it was found that most missing values were NAs (rather than NaN or zeros). The imputation strategy allowed for the preservation of data integrity, ensuring minimal disruption to the overall data set. The cleaned data was then summarized, confirming the successful handling of missing values and providing a solid foundation for further analysis.

Outliers



All variables show significant of outliers, except for **year_built_to_now**. **Listing_avg_review** and **employment_rate** outliers will be replaced with the mean for their respective city number since they're likely caused by either a very negative review or a random error, like an employment rate below 60% for 11430 which is the zipcode for Queens County, NY. After a brief online research, employment rates in this area were found to be unlikely to drop below this threshold. Furthermore, listing average reviews and booking rate will be truncated using 5th and 95th percentiles, retaining only 90% of the data which will be enough to test the hypotheses. Price mean instead will be log-transformed to deal with its outliers.

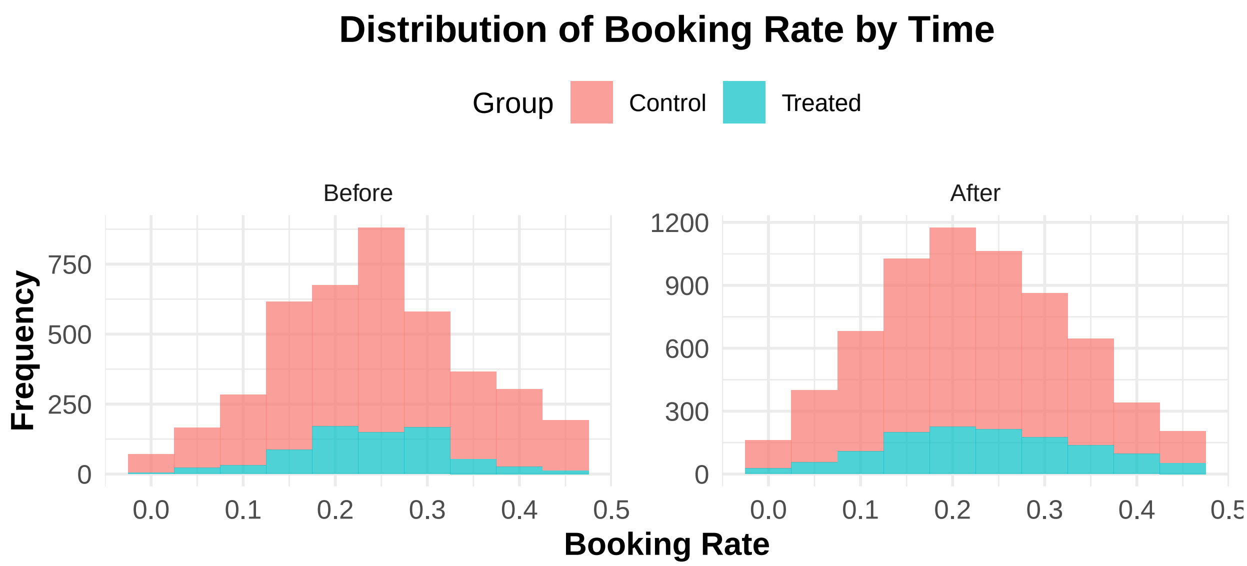
Table 2: Summary Statistics for Final Data Set

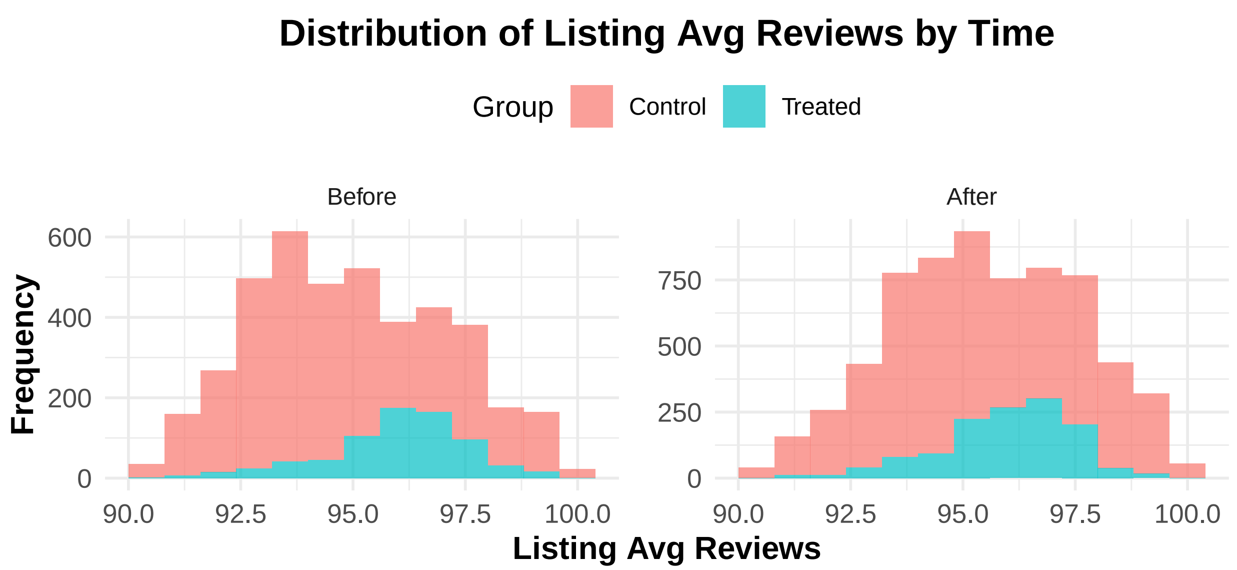
Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
booking_rate	10,702	0.230	0.105	0.000	0.159	0.228	0.300	0.473
zipcode	10,702	37,303.850	29,160.620	2,026	11,103	37,013	55,418	97,317
timeperiod	10,702	23.441	6.872	8	18	24	29	34
city_number	10,702	6.964	2.688	1	6	7	10	11
policy_entry	10,702	0.121	0.326	0	0	0	0	1
employment_rate	10,702	0.944	0.030	0.813	0.930	0.950	0.965	1.000
listing_avg_review	10,702	95.317	2.114	90.500	93.727	95.218	97.000	99.833
year_built_to_now	10,702	54.534	20.380	7.000	38.000	58.000	73.000	80.000
log_price_mean	10,702	4.969	0.484	3.624	4.573	4.937	5.332	6.745

A panel data set was created for the DiD regression analysis, allowing to compare changes before and after the introduction of the Airbnb Plus program across multiple cities. The two dimensions of the data are **city_name** and **time**, where the former represents the entities and the latter denotes the periods under study. The panel data is balanced, meaning each city has data for all time periods.

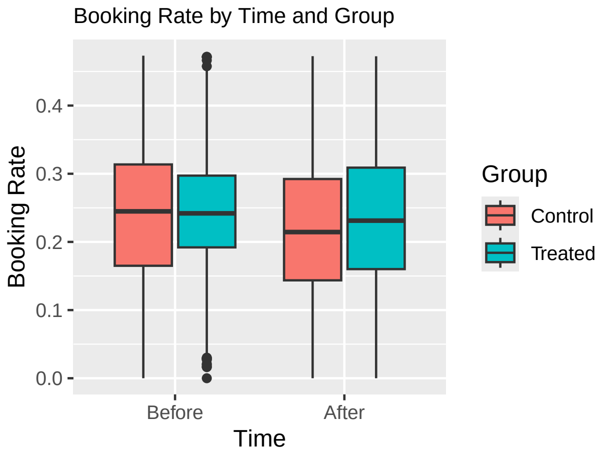
Model Free investigation

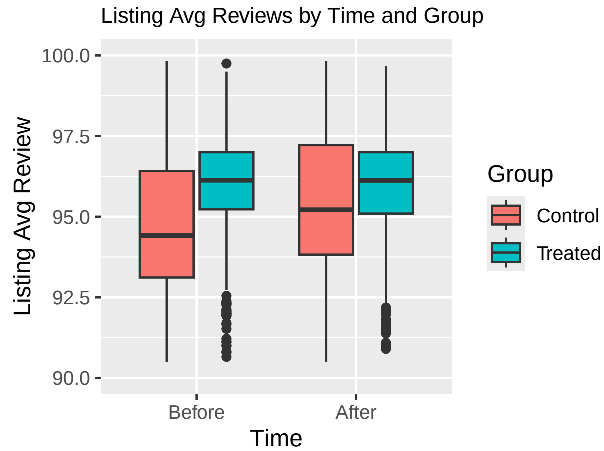
This section presents a model-free investigation using a Difference-in-Differences (DiD) approach to assess the impact of Airbnb Plus. We visualize booking rates and listing reviews before and after the program's introduction for treated and control cities. This analysis offers an intuitive understanding of the program's effects.



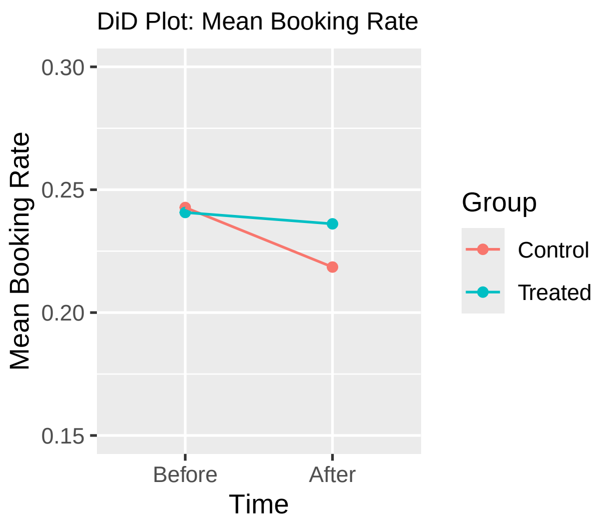


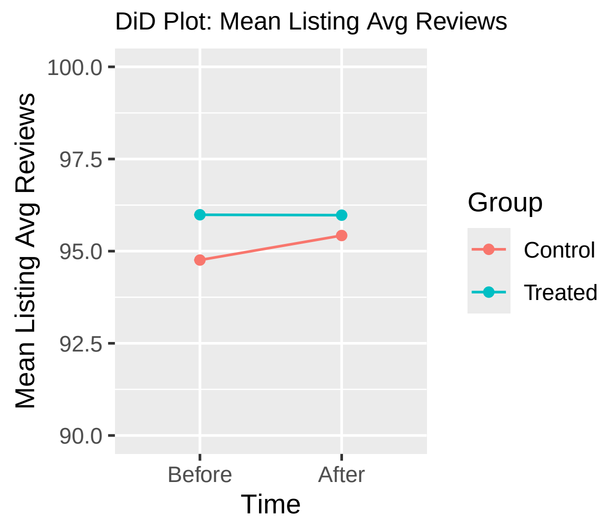
The distributions of **booking_rate** and **listing_avg_review** remain approximately normal before and after implementing the program. Listing avg reviews appear to be slightly left-skewed. In both cases, the control groups have more observations than the treated group. The differences in group sizes may lead to heteroscedasticity, impacting the precision of the interaction term estimates.





The boxplots provides a detailed view of the distributions before and after the program's introduction, showing that booking rates in treated cities maintain similar median levels with less variance compared to the control group, while average review ratings for treated cities consistently remain higher but with limited post-treatment changes.





The line graphs show the trends over time for booking rates and average review ratings. The first graph highlights a divergence in booking rates where treated cities stabilize while control cities decline. The graph for the average reviews shows the opposite as the gap between the two groups is narrowing. Together, these visualizations suggest that the Airbnb Plus program's impact is more noticeable in stabilizing booking rates than in influencing average reviews.

3. Regression Analysis: Estimating the Impact of Airbnb Plus

In this section, we apply regression analysis to estimate the impact of the Airbnb Plus program, controlling for city-specific factors using fixed city effects. By incorporating fixed effects, we account for unobserved heterogeneity across cities that could influence the outcome variables, such as booking rate and average reviews. This approach helps isolate the effect of Airbnb Plus from other city-level influences, providing a more robust estimation of the program's impact. The Fixed Effects results for **booking_rates** show that the Plus program has a marginally significant effect on booking rates for treated cities after its implementation. After including control variables, the interaction term loses its significance, meaning that the Plus program's effect on booking rates may be explained by factors like property characteristics (*year_built_to_now*) or specific rental market conditions (*log_price_mean*), rather than the program itself. For instance, it can be observed that for every unit increase in *price_mean*, *booking_rate* decreases by 0.023 units, while every unit increase in *year_built_to_now* leads to 0.01 pp increase in *booking_rate*. Ultimately, there is no evidence that the interaction effect is different from zero, so we **fail to reject** the null hypothesis. The Fixed Effects results for **listing_avg_reviews** show a **significantly negative** interaction term before and after including the control variables. Contrary to expectations, the Plus program may have had an adverse effect on average reviews in treated cities, potentially due to unintended consequences or negative customer experiences associated with the program. The observed effect seems to not be driven by confounding factors, such as differences in listing characteristics (*log_price_mean* or *year_built_to_now*) or economic conditions (*employment_rate*). Since the interaction term is statistically different from zero, we **fail to reject** the null hypothesis. The control variables provide further insights into factors influencing listing average reviews. For every unit increase in *price_mean*, listing average reviews increase by 0.75 units, suggesting that higher prices may be associated with higher perceived quality or guest satisfaction. For every unit increase in *year_built_to_now* (indicating older properties), listing average reviews decrease by 0.025 units, possibly reflecting a preference for older, more established properties. For every unit increase in *employment_rate*, listing average reviews increase by 14.8 units, highlighting the potential influence of local economic conditions on guest experiences. While the Plus program might not have achieved its intended effects, other property and market characteristics play a significant role in shaping outcomes.

Table 3: Fixed Effects Regression Results

	Dependent variable:			
	Booking Rate		Listing Avg Review	
	BR (no controls)	BR (with controls)	LAR (no controls)	LAR (with controls)
	(1)	(2)	(3)	(4)
timeAfter	-0.024^{***}	-0.015^{***}	0.665^{***}	0.314^{***}
	(0.002)	(0.002)	(0.045)	(0.041)
employment_rate		0.009		14.891^{***}
		(0.038)		(0.689)
log_price_mean		-0.023^{***}		0.752^{***}
		(0.002)		(0.042)
year_built_to_now		0.001^{***}		-0.025^{***}
		(0.0001)		(0.001)
timeAfter:groupTreated	0.010^{*}	0.003	-0.768^{***}	-0.524^{***}
	(0.005)	(0.005)	(0.106)	(0.095)
Observations	10,702	10,702	10,702	10,702
R^2	0.011	0.080	0.020	0.227
Adjusted R^2	0.011	0.080	0.020	0.226
F Statistic	61.482^{***} (df = 2; 10695)	187.090^{***} (df = 5; 10692)	109.945^{***} (df = 2; 10695)	626.761^{***} (df = 5; 10692)
Note:				*p<0.1; **p<0.05; ***p<0.01

4. Assumptions and Model Diagnostics

1. Linearity

For the scope of this research, linearity of the panel data model was assumed. Future work should assess linearity to validate model assumptions, though achieving it may be challenging due to demographic and economic variability across cities.

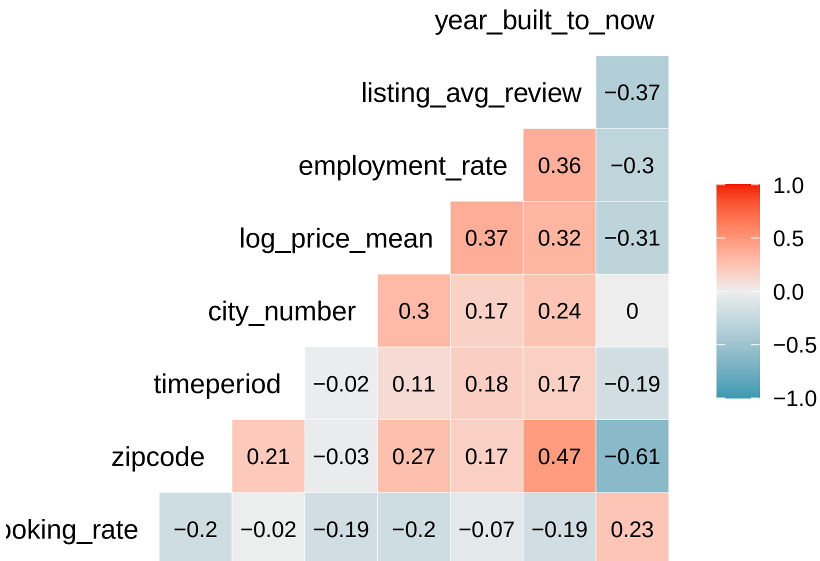
2. Random Sampling

The non-random selection of treated cities, geographic bias, and staggered program implementation challenge the random sampling assumption. Efforts to address this included imputing missing data, treating outliers in listing_avg_review and employment_rate, truncating booking_rate and reviews at the 5th and 95th percentiles, and log-transforming price_mean, but significant biases likely remain.

3. Multicollinearity

While no severe multicollinearity issues are evident, moderate correlations (year_built_to_now and listing_avg_review vs. zipcode) may still affect the model's interpretation. The Variance Inflation

Factor (VIF) diagnostics confirm this assumption was not violated as all VIF results are below 5, indicating low predictors' multicollinearity.



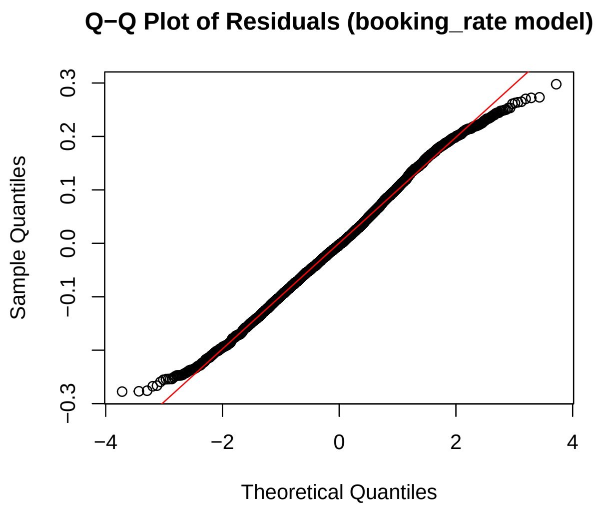
4. Zero conditional mean

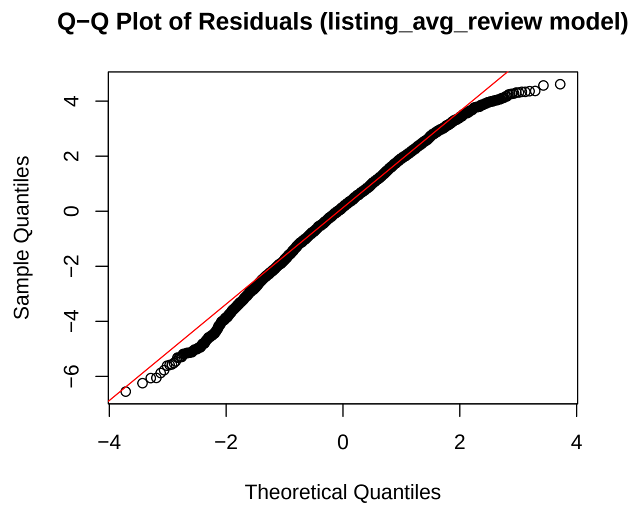
The assumption of zero conditional mean is challenged by omitted variable bias, as unobserved factors likely influence booking rates and average reviews beyond the study's scope. Potential measurement errors were noted in **employment_rate**, which may not reflect monthly fluctuations; **price_mean**, affected by price volatility and promotions; and **year_built_to_now**, susceptible to inaccurate or misreported property records. Simultaneity is also a concern, as dependent variables like booking rates and reviews may influence independent factors such as employment, housing construction, and pricing. These issues highlight the complexity of fully isolating causal relationships in this analysis.

5. Homoscedasticity

The studentized Breusch-Pagan test checks if the error term has the same variance given any values of the explanatory variable. The results of the test showed very small p-values for all models, indicating a violation of the homoscedasticity assumption. This violation can be attributed to inherent differences between cities, such as variations in population size, economic conditions, tourism activity, and other contextual factors.

6. Normality of Residuals





The Q-Q plots for both models, booking_rate and listing_avg_review, show that the residuals in the central area (approx. between -2.5 and +2) are close to the normal distribution. In the extreme areas, the points deviate slightly from the line - for very negative values they are slightly above, for very positive values slightly below. Such deviations are common in large samples and hardly affect the analysis. The Shapiro-Wilk tests resulted in very small p-values ($p < 0.0001$), which indicates statistically significant deviations from the normal distribution. However, the W values are close to 1 (0.99688 for booking_rate and 0.99342 for listing_avg_review), which means a good approximation to the normal distribution. Significant p-values are to be expected for large samples, as the test is very sensitive to small deviations. The normality assumption of the residuals is fulfilled in the central area, where the majority of the data lies. The slight deviations in the extreme ranges are tolerable and do not significantly affect the estimates of the regression coefficients or the statistical inference. Overall, the models are reliable and the results remain valid.

5. Discussion and Business Implications

This study examined the impact of the Airbnb Plus program on booking rates and customer satisfaction across various U.S. cities. The analysis revealed no statistically significant impact of the program on booking rates in treated cities compared to control cities. While treated cities experienced stabilization in booking rates, these changes appear to be driven more by external factors such as market conditions and property characteristics than the introduction of the Plus program itself. Interestingly, the program was associated with a slight decrease in average reviews, suggesting potential mismatches between customer expectations and actual experiences. Higher pricing for Plus-certified listings likely led to elevated customer expectations, making them more critical in reviews when expectations were unmet. Higher prices negatively influenced booking rates, which is in line with business expectations. However, higher prices correlated with improved customer satisfaction, possibly due to better property matching and higher quality expectations. Older buildings tended to have higher booking rates, possibly reflecting unique charm or historical value. Conversely, these older properties were associated with lower customer satisfaction, potentially due to reliability or maintenance issues. Cities with higher employment rates showed significantly higher booking rates, indicating economic factors play a critical role in driving demand. This study faced several limitations, including potential non-linear relationships in the data, geographic and temporal biases from the non-random selection of treated cities, and omitted variable bias due to unobserved factors affecting booking rates and customer satisfaction. Measurement errors in variables such as employment rate and average price, as well as heteroscedasticity caused by city-specific variations, further impacted the analysis. However, residuals closely approximated normality in the central ranges, enhancing the reliability of the regression coefficients and statistical inferences. Future research could address these limitations by exploring non-linear modeling techniques and accounting for more unobserved factors to improve explanatory power. However, building on previous research, this study concludes that the Airbnb Plus program is not a worthy investment for the company. If further research on the topic were to be conducted, this study suggests exploring the impact of booking rates across each treated city, highlighting which socio-economic and seasonal factors enhance the program's success. By doing so, the program could be implemented only in cities that meet these requirements, thereby boosting Airbnb's revenue. These cities are expected to be large central business districts (CBDs) with high employment rates, relatively low listing prices, and a good number of historical buildings.

Disclaimer

This research paper had its code deliberately debugged with the help of AI, making sure that the inevitable human errors were swiftly corrected by our digital assistant. Rest assured, no robots were harmed in the process, and the human touch remained essential. Other than debugging, AI was also used for coding suggestions and alternatives.