THE UNIVERSITY OF CHICAGO

# Heterogeneous Effects of Price Fluctuation and Crime on the Structure of the Short-term Rental Industry

By

Shuyi Zhang

June 2025

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts in Computational Social Science

Faculty Advisor: Ali Sanaei

Preceptor: Zhao Wang

## Abstract

This study investigates the heterogeneous effects of price fluctuations and crime happening nearby on listing performance in the short-term rental industry. Using a comprehensive dataset of Airbnb listings in Chicago from 2014 to 2024, the study examined how price surges and drops influence occupancy rates, Average Daily Rate (ADR), and revenue across different Airbnb property segments. Employing a combination of spatial analysis, temporal analysis, causal forest, and machine learning methodologies, the study identified four distinct property clusters with varying sensitivity to property characteristics and review ratings. The findings reveal a strong association between price surges and increased occupancy rates. While this counterintuitive result may reflect a signaling mechanism, we caution that causality cannot be conclusively established. It is also possible that elevated demand may precede price changes. Hosts raise prices in response to unobserved shifts in demand — suggesting that these outcomes reflect correlated dynamics rather than strictly causal relationships. Price increases may operate as signals of quality or scarcity that elevate perceived value. Conversely, price drops generally result in revenue losses, highlighting asymmetric market responses to price adjustments. Additionally, the research find that neighborhood safety characteristics moderate these effects, with crime incidents having heterogeneous impacts across property segments. This research contributes to revenue management theory by demonstrating that strategic price positioning can leverage quality signaling effects, where consumers interpret higher prices as indicators of superior value in certain market segments. The findings provide practical insights for hosts and platform operators seeking to optimize pricing strategies in increasingly complex short-term rental markets.


**Keywords:** Short-term Rental Market; Airbnb; Price Voliaty; Dynamic Pricing; Pricing Strategy; Revenue Management; Price Surge; Price Drop; Consumer Behavior; Listing Performance; Occupancy Rate; Revenue; Average Daily Rate

---

# Contents

# 1 Introduction

The short-term rental market, led by platforms like Airbnb, has experienced significant growth over the past decade, fundamentally altering accommodation options for travelers and creating new income streams for property owners. With global revenues exceeding $100 billion annually and over 7 million active listings worldwide (AirDNA, 2023), the platform economy has reshaped the hospitality industry. This growth has introduced complexities in pricing strategies and their effects on listing performance. Although hosts are equipped with algorithmic tools like ”smart pricing”, the ultimate responsibility for price positioning—adjusting listing prices dynamically in response to market shifts — rests with individual hosts (Kwok and Xie, 2019). Pricing and balancing the trade-off between lower daily revenue and higher occupancy rates have consequently become more complicated to understand for Airbnb hosts. This complexity is further amplified by the heterogeneity of property types, locations, and guest preferences that characterize the short-term rental ecosystem. In such a decentralized and competitive environment, understanding how price volatility affects performance metrics such as occupancy, revenue, and Average Daily Rate (ADR) becomes crucial.

Existing research has made important contributions by identifying static determinants of listing success, such as property attributes, location proximity to amenities, and reputation signals like reviews (Guttentag, 2019, Zervas et al., 2017). However, four significant gaps persist. First, prior studies focus largely on static price levels rather than dynamic pricing strategies and volatility. Second, heterogeneity in market response—how different types of listings react differently to pricing moves—remains underexplored. Third, much of the research identifies correlations without establishing causal relationships. Few studies account for the temporal evolution of market behavior as platforms and consumer expectations mature. These gaps have profound consequences: without deeper causal understanding and acknowledgment of heterogeneity, both academic models and practical revenue strategies risk being simplistic and misaligned with real-world dynamics.

One widespread assumption is that price increases inherently suppress demand due to price sensitivity. While this principle holds in traditional hospitality settings, it may not apply straightforwardly in the peer-to-peer short-term rental market, where price can serve as a signal of quality, exclusivity, or scarcity. Destabilizing this assumption is crucial: if higher prices under certain conditions actually correlate with higher occupancy and revenue, hosts and platforms could rethink revenue management strategies to emphasize perceived value rather than undercutting competitors.

At the same time, safety concerns, particularly neighborhood crime, introduce additional complexity. Crime may deter guests, but its effects could vary across market segments and interact with pricing strategies in unexpected ways. Thus, understanding how price changes and local safety factors jointly shape Airbnb performance requires a more sophisticated and context-sensitive approach.

This study addresses these challenges by analyzing a decade-long dataset of Airbnb listings in Chicago (2014–2024). Leveraging spatial analysis, machine learning-based clustering, and causal forest models, this research investigates the heterogeneous effects of price surges and drops, as well as nearby crime incidents, on key performance metrics. The use of causal forests, in particular, allows us to estimate heterogeneous treatment effects across different property types without imposing restrictive assumptions, thus partially addressing the challenge of endogeneity between price and demand shifts (Wager and Athey, 2018). By segmenting listings into distinct clusters based on review ratings, physical characteristics, and price volatility, the study captures important variations that aggregate models obscure. Furthermore, the incorporation of spatial dynamics highlights how the geographic dispersion of listings—and exposure to neighborhood crime—moderates pricing effects over time. Importantly, by introducing lagged variables for price changes, the analysis attempts to mitigate issues of reverse causality, although it acknowledges that definitive causal claims remain methodologically difficult.

Preliminary findings suggest that price surges are often associated with increases in occupancy and revenue, particularly among certain property clusters, challenging traditional elasticity models. This phenomenon likely reflects quality signaling effects, where higher prices elevate perceived value among guests. Conversely, price drops are associated with revenue declines, indicating asymmetric market responses. Moreover, the effects of nearby crime are not uniform: while some property segments suffer declines, others remain resilient, possibly due to compensating amenities or target customer differences.

The stakes of understanding these dynamics are high. For hosts, misjudging the effects of price changes or failing to account for heterogeneous guest responses can lead to substantial revenue losses. For platform operators and urban policymakers, a nuanced grasp of pricing dynamics and neighborhood factors can inform strategies to sustain market vibrancy without exacerbating housing inequities or safety concerns. More broadly, for scholars of the sharing economy, this research challenges overgeneralizations and highlights the need for contextually rich, causally informed models of market behavior.

In addressing an underexplored but critical problem in short-term rental economics, this study contributes both theoretically—to discussions of dynamic pricing and signaling—and practically, offering actionable insights for hosts navigating increasingly complex and competitive marketplaces.

# 2 Literature Review

## 2.1 Determinants of Airbnb Listing Performance

The concept of Average Daily Rate (ADR) has been widely used to measure pricing efficiency, alongside occupancy rate and revenue as key performance metrics (Zervas et al., 2017).

Past studies have extensively analyzed the determinants of Airbnb listing performance. Guttentag (2019) provides a comprehensive review of Airbnb research, highlighting pricing strategies as a critical factor in listing success. Besides, review ratings, location-related factors, amenities, host characteristics, etc. can all influence the revenue of Airbnb listings (Kirkos, 2022). These studies established important correlations, suggesting that consumers value different attributes depending on the property type and location context, supporting the notion of heterogeneous preferences in the short-term rental market. However, the prior research did not address the dynamic aspects of pricing over time.

### 2.1.1 Static Pricing Factors

Substantial differences in the operational and financial performance extist between professional and non-professional Airbnb hosts, which can be mainly explained by pricing inefficiencies (Li et al., 2019). In particular, properties managed by professional hosts earn 16.9% more in daily revenue, have 15.5% higher occupancy rates, and are 13.6% less likely to exit the market compared with properties owned by non-professional hosts.

In the context of Airbnb, price surging refers to the price raised as the consumption date approaches and is found to have a detrimental impact on the revenue of hosts (Leoni and Nilsson, 2021). Exacerbating or moderating variables are also found to make the price surge particularly detrimental to revenue, including distance to the main point of interest is generally associated with higher demand or performance (Arbel and Pizam, 1977; Tussyadiah, 2016; Y. Yang et al., 2018; Gunter and Önder, 2018). Similarly, the experience of the host is found to positively affect performance (Brouder and Eriksson, 2013;Gémar et al., 2016; K. Xie and Mao, 2017). Finally, the booking pace is a key determinant of revenue. Releasing an excessive percentage of the capacity with a high advance could have a negative effect on performance. On the contrary, releasing restrictions curb revenue dilution (Fyall et al., 2013).

For price drops, discounts and price promotions have been recognized as a significant factor influencing consumers' purchasing decisions (Haghighatnia et al., 2018). Prior researches conducted in the field of marketing has demonstrated that price reductions can have both advantageous and disadvantageous impacts on customers' assessments and buying patterns. Advantages can include increased sales volume, increased market share and clearance of old inventory; while disadvantages can consist of perception of low quality, reduced profit margins, difficulty in raising prices and damage to brand image (Lee and Chen-Yu, 2018; Darke and Dahl, 2003; Prakash et al., 2021; Koçaş and Bohlmann, 2008; Kwon et al., 2007; Z. Yang and Peterson, 2004). Liu et al.(2021) found that customers' total spending is influenced by product-specific price cuts in a U-shaped pattern. In other words, product-specific price discounts influence purchase quantity positively though at a decreasing rate. The utilization of order coupons has a beneficial impact on both total spending and purchase quantity, with a notable upward trend.

### 2.1.2 The Role of Reviews and Reputation

Customer reviews have been identified as particularly important in the sharing economy context. K. Xie and Mao (2017) demonstrate that positive reviews significantly enhance booking likelihood and allow hosts to command price premiums. This effect is especially pronounced for listings with professional hosts (i.e., Superhosts in the context of this study). Their findings suggest that reputation mechanisms serve as critical quality signals in markets characterized by information asymmetry.

Liang et al. (2017) specifically examined the impact of Airbnb's Superhost badge system on listing performance. Using a quasi-experimental approach, they found that acquiring Superhost status led to an average price premium of 8.2% and increased occupancy rates by approximately 5.1 percentage points. Their results highlight the economic value of reputation signals in peer-to-peer markets where quality uncertainty is high. Teubner et al. (2017) expanded this line of inquiry by examining multiple dimensions of host reputation, including review scores, number of reviews, and verified identities. They found that the relationship between reputation metrics and pricing power is non-linear, with diminishing returns beyond certain thresholds. This suggests that the value of additional reputation signals may vary across different market segments and contexts.

### 2.1.3 Location-based Factors

Location-based factors have also gained attention in recent literature. Zervas et al. (2017) found that proximity to attractions, public transportation, and local amenities significantly influences Airbnb listing performance. Additionally, the role of external factors, such as crime rates and economic conditions, has been explored by researchers like Metaxas and Romanopoulos (2023), who documented negative correlations between high crime rates and Airbnb pricing power. Zhang et al. (2017) applied geographically weighted regression to analyze spatial variation in Airbnb pricing determinants. Their findings revealed significant spatial non-stationarity in the effects of various factors on listing prices, highlighting the importance of geographic context in understanding market dynamics. They found that the influence of neighborhood characteristics such as restaurant density and transportation accessibility varied substantially across different urban areas. Gutiérrez et al. (2017) examined the spatial distribution of Airbnb listings in Barcelona, identifying patterns of concentration in tourist-centric areas and emerging clusters in residential neighborhoods. Their analysis suggested that Airbnb expansion follows distinct spatial logics related to tourist attractions, transportation networks, and housing market conditions. These patterns have implications for understanding how price effects might vary across different urban contexts.

#### 2.1.4 Temporal Dynamics and Seasonality

Research on temporal aspects of Airbnb performance has highlighted the importance of seasonality and dynamic pricing. Dogru et al. (2020) analyzed seasonal variations in Airbnb pricing across different urban markets, finding that price elasticity of demand varies substantially between peak and off-peak periods. Their research suggested that optimal pricing strategies should account for these temporal variations to maximize revenue. Magno et al. (2018) investigated the diffusion of dynamic pricing practices among Airbnb hosts, finding that adoption rates varied by host professionalism and market competition. Their longitudinal study revealed an increasing trend toward algorithmic pricing tools, particularly among multi-listing hosts. This trend highlights the growing sophistication of pricing strategies in the short-term rental market. Oskam et al. (2018) examined long-term trends in Airbnb market development across European cities, identifying distinct phases of market evolution characterized by changing host behaviors and platform dynamics. Their research suggested that as markets mature, pricing strategies become more sophisticated and heterogeneous across different host segments.

Considering the seasonal factors, the study set the month variable as the fixed effect.

### 2.2 Price Volatility in Shared Accommodation Markets

While static pricing factors have been well-studied, the impact of price volatility — defined as the frequency and magnitude of price changes over time — remains less understood. Dynamic pricing has become increasingly common in the hospitality industry, with algorithms adjusting prices based on demand forecasts, competitor behavior, and special events. However, the effects of such pricing strategies on consumer perception and booking behavior in peer-to-peer markets like Airbnb may differ from traditional hospitality contexts.

Several studies have explored pricing strategies in the short-term rental market. Wang and Nicolau (2017) identified determinants of price in sharing economy accommodations, finding that professional hosts tend to implement more sophisticated pricing strategies.

However, their analysis did not specifically address the impact of price volatility on listing performance. Similarly, Gibbs et al. (2018) examined price determinants across different types of Airbnb accommodations but did not focus on the temporal dimension of pricing. Kwok and Xie (2019) investigated the effects of price positioning relative to competitors on Airbnb listing performance. Their findings suggested that strategic price positioning significantly influenced both occupancy rates and revenue. Listings priced moderately below market averages achieved higher occupancy rates without sacrificing substantial revenue. Li (2016) analyzed the adoption of dynamic pricing among Airbnb hosts, finding that approximately 42% of hosts adjusted prices at least weekly in response to market conditions. Their research identified several factors associated with dynamic pricing adoption, including host experience, market competition, and property characteristics. However, both studies did not specifically examine the effects of price fluctuations over time. Recent work by Ma et al. (2020) has explored the relationship between price dispersion and demand in the short-term rental market, finding that excessive price volatility can create consumer uncertainty and potentially reduce booking likelihood. However, their research did not examine differential effects across property types or market segments, highlighting an important gap in our understanding of heterogeneous price effects.

### 2.3 Heterogeneity in Short-Term Rental Markets

A growing body of research has highlighted the heterogeneous nature of short-term rental markets. Lutz and Newlands (2018) used cluster analysis to identify distinct user segments in the Airbnb ecosystem, finding significant differences in preferences and behaviors across these segments. Their findings challenge the notion of a monolithic Airbnb market and suggest that pricing effects may vary across different user groups. Similarly, Adamiak (2022) documented substantial variation in Airbnb supply characteristics across 167 countries, identifying distinct patterns related to tourism development, urbanization, and regulatory environments. His research highlighted the importance of contextual factors in shaping market dynamics and suggested that findings from one geographic context might not generalize to others. Gunter and Önder (2018) examined heterogeneity in price determinants across different property types and locations. Using quantile regression, they demonstrated that the effects of various attributes on listing prices varied substantially across the price distribution. Their findings suggested that market segments at different price points respond differently to similar factors, supporting the notion of heterogeneous effects in short-term rental markets. Deboosere et al. (2019) investigated spatial inequality in Airbnb supply and performance across neighborhoods in New York City. Their analysis revealed substantial disparities in listing density, prices, and occupancy rates across different socioeconomic contexts. These findings suggest that the effects of pricing strategies may vary significantly across urban neighborhoods with different characteristics. Dolnicar and Zare (2020) documented significant shifts in Airbnb hosting patterns following the COVID-19 pandemic, with differential impacts across property segments and geographic regions. Their findings highlighted the dynamic nature of market segmentation and the need for contextually sensitive approaches to understanding price effects.

### 2.4 Methodological Approaches in Airbnb Research

Methodologically, research on Airbnb has evolved from descriptive and regression-based approaches to more sophisticated techniques. Recent studies have begun incorporating machine learning methods to capture complex relationships in Airbnb data. Lutz and Newlands (2018) used clustering techniques to segment Airbnb users, while Adamiak (2022) applied spatial analysis to understand the geographic distribution of listings.

Causal inference in Airbnb research remains relatively unexplored, with most studies establishing correlational rather than causal relationships. This study builds on these methodological advances by integrating machine learning techniques and causal forest methods to explore the impact of price volatility on Airbnb listing performance in Chicago over the past decade. The application of causal forests, a methodology developed by Wager and Athey (2018), represents a significant advancement in understanding heterogeneous treatment effects in pricing strategies. This approach allows researchers to identify subgroups that respond differently to interventions without imposing strong model assumptions, making it particularly suitable for examining complex market dynamics.

### 2.5 Research Goal

In conclusion, considerable prior research suggest that various factors—including price volatility, review ratings, and location factors — can influence occupancy rates and revenue. Despite the extensive literature on Airbnb pricing and performance, several important gaps remain. First, most studies have focused on static price levels rather than dynamic pricing strategies and volatility. Second, the heterogeneity of price effects across different market segments has received limited attention. Third, the causal impact of pricing strategies on listing performance remains underexplored, with most studies establishing correlational relationships. Fourth, the temporal evolution of price effects over time, as the Airbnb market matures, has not been thoroughly investigated. Using Chicago as the study area, the research leverages a comprehensive dataset spanning ten years to investigate how price fluctuations affect Airbnb listing performance across different market segments. Specifically, the study examines:

**Primary Research Question**: How do price fluctuations influence Average Daily Rate

(ADR), occupancy rates, and revenue across different types of Airbnb listings in Chicago?

**Sub-research Question**:

- Market Segmentation Effects: How do property attributes (has air conditioner/gym/pool/parking/hot tub/kitchen/allow pets or not), hosts' characteristics and strategies (max guests and minimum stay set by hosts, number of photos uploaded by hosts, response rate and response time of hosts, superhost or not), review ratings (number of reviews and review rating scores), location-related charateristics (crime happening nearby, longitude, latitude, zipcode) moderate the relationship between price volatility and listing performance?

- Spatial Dynamics: What spatial patterns emerge in the distribution of high-performing listings, and how have these patterns evolved over the past decade?

- Causality Assessment: To what extent can causal relationships be established between pricing strategies and economic outcomes in short-term rentals, considering potential reverse causality? Note: While this study employs causal forest methods to estimate treatment effects, the researcher recognizes the methodological limitations in asserting true causal direction. Reverse causality — where latent demand shifts precede price changes — remains a key challenge. Additional robustness checks and lagged treatments are used to probe, but not definitively resolve, this issue.

- Market Maturation: How have the effects of price volatility evolved as Chicago's Airbnb market has matured over the past decade?

By addressing these questions, this study contributes to the growing literature on the sharing economy and provides practical insights for stakeholders in the short-term rental ecosystem.

# 3 Data and Methods

## 3.1 Data Collection

- Airbnb Listing Data The study chooses Airbnb as the platform. AirDNA is a prominent supplier of short-term rental data, with a focus on providing market information for Airbnb and Vrbo. Provides valuable information for hosts, investors, and academics in the vacation rental business. AirDNA uses web scraping methodologies to collect publicly accessible data from platforms dedicated to short-term rentals. The coverage covers data from more than 10 million homes in 120,000 markets around the world. Historical data is generally available from 2014, but availability may differ depending on the market (AirDNA, 2024). Data of the selected Illinois are included in this research, covering the period from October 1st, 2014 to October 1st, 2024. The dataset includes daily information for all active listings in Illinois during the study period, which was aggregated to monthly for analysis. The dataset contains rich information on listing characteristics, performance metrics, and host attributes, including:
  - Property_ID: AirDNA's unique identifier for the property.
  - Airbnb_Property_ID: Unique property ID assigned by Airbnb. http://airbnb.com/rooms/(property_id) will bring up the Airbnb vacation rental listing.
  - Date: The last date AirDNA scrapers captured calendar information for the vacation rental listing. AirDNA scrapes active calendars every day.
  - ADR_USD: Average daily rate of booked nights in the currency specified in the currency field. ADR is calculated by dividing the total revenue by the booked nights. Cleaning fees are included.
  - Price_USD
  - Overall review rating: Web scrap from each Airbnb listing on the Airbnb official website.
  - Airbnb_Communication_Rating: Average communication rating of the property out of 10.
  - Airbnb_Accuracy_Rating: Average accuracy rating of the property out of 10.
  - Airbnb_Cleanliness_Rating: Average cleanliness rating of the property out of 10.
  - Airbnb_Checkin_Rating:Average checkin rating of the property out of 10.
  - Airbnb_Location_Rating: Average location rating of the property out of 10.

- Airbnb_Value_Rating: Average value rating of the property out of 10.
- Airbnb_Superhost: True or False depending if the host is a Superhost on Airbnb.
- Cleaning_Fee_USD: Cleaning fee charged per reservation in the currency specified on the Currency field.
- Count_Available_Days: Total number of listing calendar days that were classified as available for reservation, but not actually booked during the last twelve months.
- Count_Reservation_Days: Total number of listing calendar days that were classified as reserved during the last twelve months.
- Extra_People_Fee: Extra people fee in the currency specified on the Currency field.
- Longitude
- Latitude
- Max_Guests: Number of guests that the listing can accommodate.
- Minimum_Stay: The default minimum night stay required by host.
- Number_of_Photos: Number of photos in a vacation rental listing.
- Number of Reviews
- Occupancy_Rate: Monthly occupancy rate for the listing.
- Published_Monthly_Rate: Default monthly rate for a vacation rental listing in the currency specified on the Currency field.
- Published_Weekly_Rate: Default weekly rate for a vacation rental listing in the currency specified on the Currency field.
- Published_Nightly_Rate: Default nightly rate for a vacation rental listing in the currency specified on the Currency field.
- Response_Rate: The percentage of new inquiries and reservation requests a host responds to (by either accepting/pre-approving or declining) within 24 hours.
- Response_Time (min): The average host response rate for new inquiries and reservation requests from a guest.
- Has_Air_Con: If true, the listing has an air conditioner.
- Has_Gym: If true, the listing has a gym.
- Has_Hot_Tub: If true, the listing has a Hot Tub.
- Has_Kitchen: If true, the listing has a kitchen.

- Has_Parking: If true, the listing has parking space.
- Has_Pool: If true, the listing has a pool.
- Pets_Allowed: If true, the listing allows pets.
- Location_Type: Type of location of the listing. Classified into Destination/Resort - Mountains/Lake, Destination/Resort - Coastal, Small City/Rural, Large City - Suburban, Large City - Urban, or Mid-Size City. Can also be null.
- Metropolitan_Statistical_Area: Metropolitan statistical area of where the vacation rental property is located (US only). 929 total MSAs between metropolitan and micropolitan statistical areas.
- Price_tier: Average Daily Rate Price Tiers in MarketMinder segments listings within a market into different price points (Budget, Economy, Midscale, Upscale, and Luxury).
- Minimum_Stay: The default minimum night stay required by host.
- Revenue_USD: Last twelve months listing revenue. Includes cleaning fees but not other additional fees.
- Revenue_Potential_USD: Last twelve months (LTM) of potential listing revenue, in the currency specified on the Currency field. Includes cleaning fees but not other additional fees.
- Security_Deposit_USD: Security deposit in the currency specified on the Currency field.
- Zipcode: Zip code where the vacation rental property is located (US markets only).

- Chicago Crime Data: The study retrieves Chicago crime data from Chicago Data Portal, the City of Chicago's open data portal. The crime data includes incident reports categorized by crime type, location, and timestamp, allowing for the analysis of neighborhood safety characteristics and their impact on listing performance. Considering the influence of the Covid-19 pandemic, only data from from January 1st, 2022, to October 1, 2024 was selected in the causal forest and regression analysis. To select data from January 1st, 2022, to October 1, 2024, use the following link: Filtered Crime Data.

  The study follows the FBI's definition of violent crime to classify severe crimes. The FBI focuses on the most violent criminals and organizations that pose a major threat to American society. Significant violent crimes, including mass killings, gang violence, armed robberies, and drug trafficking, can paralyze communities and stretch state and local law enforcement resources to their limits. This classification ensures that our analysis captures meaningful safety indicators that might influence guest decision-making and host pricing strategies

### 3.2 Data Pre-processing

The research filtered listings with inconsistent, duplicate or missing values. New variables were created to facilitate the analysis, including:

- Guest_Group: Transfromed from Max_Guests. Small groups refer to 1 2 single guests/couples/friends; Medium groups refers 3 6 bunchs of friends/couples. Large groups refers to 7 10 guests. Party size refers to more than 10 people.
- price_rolling_mean: Calculate the rolling mean for each Airbnb_Property_ID.
- price_rolling_std: Calculate the standard deviation for each Airbnb_Property_ID.
- price_surge: True if the price increase in percentage in the past 3 months is higher than 1 s.d. from the mean of the same household in the past 12 months.
- price_drop: Price drop is the opposite of price_drop.
- price_surge_lag: Generate the lag of price surge for each Airbnb_Property_ID, which is whether the price surge happened in the past month.
- price_drop_lag: Generate the lag of price surge for each Airbnb_Property_ID, which is whether the price drop happened in the past month.
- price_change_pct: Price_change_pct is the percentage change in price from the rolling mean
- Price_Group: A categorical variable named Price_Group was created to segment listings based on their absolute price levels. Using quantile-based thresholds, listings were divided into three groups: ”high_price” (top 10% of prices), ”mid_price” (middle 80%), and ”low_price” (bottom 10%).
- crime_nearby: Whether crimes occurred (1) or not (0) within the radius (1.5 mile) and time window (6 months).

All datasets were merged for a holistic analysis of pricing dynamics. Apply log transformations to price-related variable (i.e., price_rolling_mean) to manage skewed distributions , which is a common approach in econometric analyses of price data (Kuo et al., 2025).

It is worth mentioning that the study kept outliers. Outliers in hospitality can indicate authentic diversities in the data rather than mistakes or irregularities. Retaining outliers in Airbnb data is a valid technique that can offer a more realistic depiction of the real-world scenario. In the context of Airbnb, extreme values probably represent to luxury or substantially discounted Airbnb listings, holding important meaning.

### 3.3 Exploratory Data Analysis (EDA)

Initial exploratory analysis aimed to identify patterns and relationships in the data. Table 1 presents descriptive statistics for key variables in the dataset.

Table 1: Descriptive Statistics of Key Variables in Airbnb Data

| Variable | Count | Mean | Std | Min | 25% | 50% | 75% | Max |
|---|---|---|---|---|---|---|---|---|
| Price (USD) | 2,419,974 | 169.34 | 472.78 | 0.00 | 64.93 | 105.00 | 179.00 | 135,718.60 |
| ADR (USD) | 1,056,465 | 123.21 | 184.27 | 0.00 | 0.00 | 80.15 | 171.37 | 4,812.21 |
| Accuracy Rating | 1,866,933 | 9.61 | 0.90 | 2.00 | 9.00 | 10.00 | 10.00 | 10.00 |
| Checkin Rating | 1,865,898 | 9.75 | 0.77 | 2.00 | 10.00 | 10.00 | 10.00 | 10.00 |
| Cleanliness Rating | 1,867,032 | 9.44 | 0.99 | 2.00 | 9.00 | 10.00 | 10.00 | 10.00 |
| Communication Rating | 1,866,871 | 9.74 | 0.79 | 2.00 | 10.00 | 10.00 | 10.00 | 10.00 |
| Location Rating | 1,865,720 | 9.64 | 0.81 | 2.00 | 9.00 | 10.00 | 10.00 | 10.00 |
| Value Rating | 1,865,606 | 9.43 | 0.95 | 2.00 | 9.00 | 10.00 | 10.00 | 10.00 |
| Overall Rating | 1,896,074 | 94.25 | 8.99 | 20.00 | 93.00 | 97.00 | 99.00 | 100.00 |
| Bathrooms | 2,374,841 | 1.38 | 0.85 | 0.00 | 1.00 | 1.00 | 1.50 | 115.00 |
| Bedrooms | 2,374,849 | 1.64 | 1.26 | 0.00 | 1.00 | 1.00 | 2.00 | 50.00 |
| Cleaning Fee (USD) | 2,078,256 | 71.16 | 83.52 | 0.00 | 15.00 | 50.00 | 100.00 | 3,066.00 |
| Extra People Fee | 1,388,711 | 12.21 | 20.73 | 0.00 | 0.00 | 5.00 | 20.00 | 300.00 |
| Max Guests | 2,376,163 | 4.11 | 2.97 | 1.00 | 2.00 | 3.00 | 6.00 | 100.00 |
| Occupancy Rate | 1,921,345 | 54.26 | 87.36 | 0.00 | 5.00 | 21.00 | 67.00 | 3,705.00 |
| Revenue (USD) | 2,378,085 | 975.57 | 2,370.69 | 0.00 | 0.00 | 0.00 | 1,099.00 | 140,380.00 |
| Security Deposit (USD) | 818,218 | 222.92 | 411.10 | 0.00 | 0.00 | 100.00 | 250.00 | 12,307.00 |

A correlation analysis was conducted to examine relationships among key variables, visualized through a heatmap (Figure 1).

Major findings from the exploratory analysis include:

- Positive Relations:
  - ADR: Number of bedrooms, Cleaning fee, Count_Available_Days_LTM & Count_Reservation_Da Max of Guests, Number of Photos, ADR & Published Monthly/Weekly/Nightly Rate, Revenue Potential, Security Deposit
  - ADR not highly correlated with all (sub-)rating & overall rating
  - Occupancy Rate: Max Guests, Number of Photos, Number of Reviews & Review Ratings, Revenue

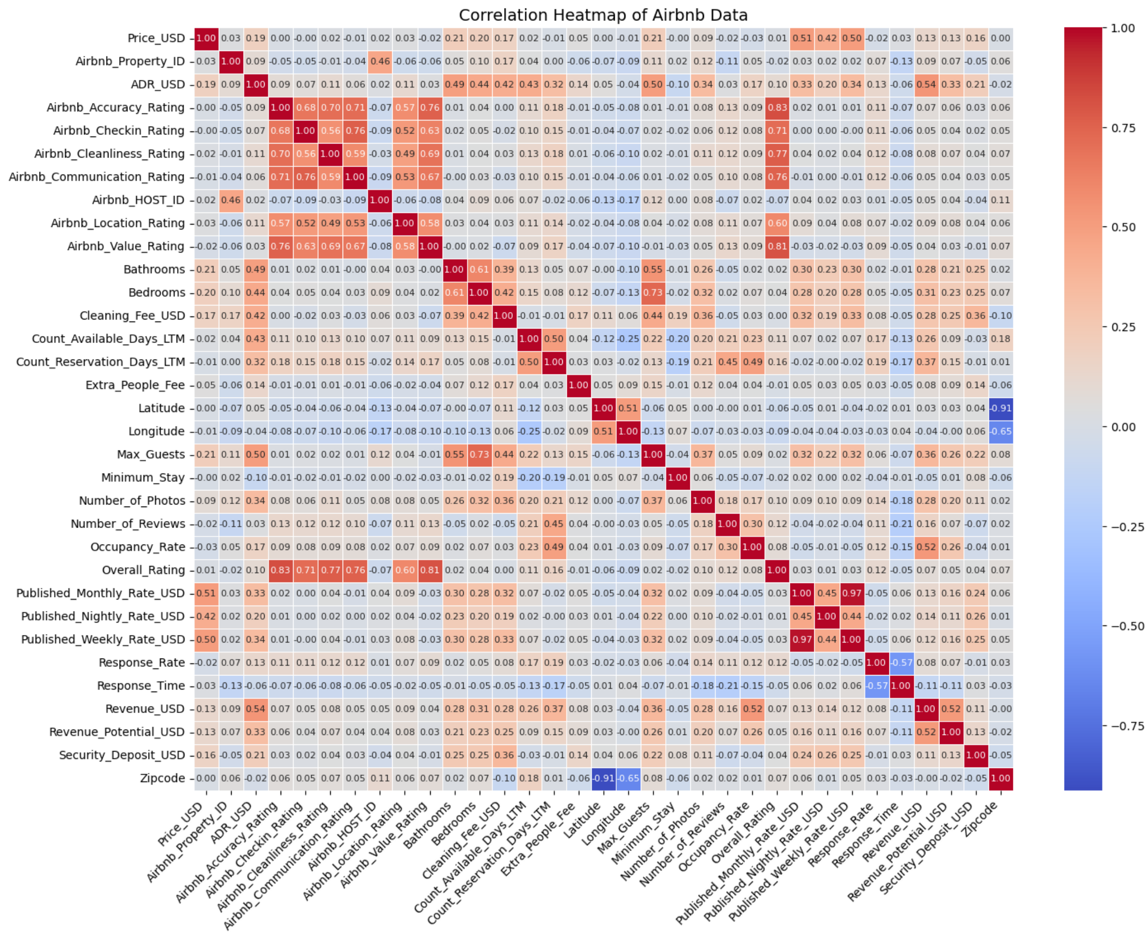


Figure 1: Correlation Heatmap of Airbnb Data

- Max Guests: Published Monthly/Weekly/Nightly Rate, Revenue

- Negative Relations:
  - Response Rate & Response Time
  - Response of time & Occupancy Rate

## 3.4 Analytical Methodology

This study employs a multi-method analytical approach to comprehensively examine the heterogeneous effects of price fluctuations and crime on Airbnb listing performance. The methodology integrates spatial econometrics, temporal analysis, machine learning techniques, and causal inference to address the research questions from complementary perspectives. Specifically, geographic information systems (GIS) analysis enables the visualization of spatial patterns and evolutionary trends in listing distribution across Chicago neighborhoods. Unsupervised machine learning techniques, including principal component analysis and K-means clustering, identify distinct property segments with varying characteristics and market behaviors. Causal forest models, a machine learning approach to causal inference developed by Wager and Athey (2018), estimate heterogeneous treatment effects of price changes and crime incidents on performance metrics while accounting for confounding variables. Finally, regression analysis quantifies relationships between listing characteristics, pricing strategies, and performance outcomes with fixed effects for temporal and spatial factors. This integrated methodological framework allows for robust investigation of both market-wide patterns and segment-specific dynamics in the short-term rental ecosystem.

- Cluster 0 **Location & Cleanliness Focused Urban Apartment Units**: This cluster represents compact homes with 1 bedroom and 1 bathroom, often units within larger apartment buildings. These homes typically include a gym, suggesting an emphasis on modern, urban living. Guests in this group prioritize location and cleanliness, indicating they seek convenience and well-maintained spaces. However, they are not concerned with lower ratings in other sub-categories, such as value or check-in experience.

- Cluster 1: **Minimalist 1B1B Rentals, General Quality Over Location**: Homes in this cluster are small units (1 bedroom, 1 bathroom) without a kitchen or gym, suggesting budget-friendly or short-term rental options. Guests in this group value overall review quality more, meaning they prefer well-rated listings in general. However, location does not play a major role in their decision, and cleanliness is not a deciding factor, indicating a focus on functionality over premium features.

- Cluster 2: **Kitchen-Centric Homes, Care about the general review quality but not cleaniness. Location has moderate effect**: This cluster consists of homes where the kitchen is a defining feature. Guests in this group highly value the overall review quality but are not overly concerned with cleanliness ratings. Location has a moderate influence on their choice, suggesting they balance convenience with home amenities.

- Cluster 3: **Luxury homes or entire apartment building for rent, Priotirizing the general review quality**: Homes in this cluster are also kitchen-focused, but both 1B1B and 2B2B layouts are acceptable to guests. Prefer 2b2b and the presence of gyms strongly. Care about the general review quality, don't require the singe sub-rating to be very high.

Perform GIS analysis on the four different clusters, with four distinct colors representing different clusters and heat map layer showing the distribution of Airbnb listings.