Host Participation in Short-term Rental Markets: Geospatial and Socioeconomic Analysis of Airbnb in San Francisco

Avijit Sarkar
University of Redlands
avijit_sarkar@redlands.edu

James B. Pick
University of Redlands
james_pick@redlands.edu

Shaista Jabeen
University of Redlands
shaista_jabeen@redlands.edu

Abstract

This paper examines spatial patterns and socioeconomic influences on host participation in San Francisco's short-term home rental markets. Spatial distributions of Airbnb property densities are mapped in a Geographic Information System and clusters and outliers are identified. Unlike major Airbnb markets, Airbnb hotspots in San Francisco are not located in the city's core but are predominant in northeastern neighborhoods of the city located close to points of interest (POIs) frequented by visitors, have high proportions of hotel and lodging employment, and lower median household incomes. Regression results reveal that the dominant determinants of Airbnb property density are professional, scientific, and technical services, and hotel/lodging employment, Asian population, and POIs within located within 20 minutes of walking time from the centroid of the city's census tracts. Implications of these findings are discussed to understand supply-side motivations of Airbnb hosts for participating in the shared accommodation economy.

Keywords: Airbnb, Short-term rental market, GIS, Spatial Patterns, Regression

1. Introduction

The COVID-19 pandemic has had a profound influence on internet use in American households to participate in the sharing economy. In November 2021, over 94 million Americans used the internet to request services provided by other people, for example to hail a ride using Uber, order food and groceries using DoorDash or Instacart, reserve a room through Airbnb, or requesting a home repair through Angie's List. This represents 43 percent of all internet users, compared to 34 percent in 2019. Also in November 2021, almost 20 million American internet users offered services such as ridesharing and homesharing via platforms such as Uber and Airbnb, an increase of 3 million Americans since 2019 (NTIA, 2022). As such online activities intensify, they are likely to have a profound influence on the sharing economy, both on the supply and demand sides, necessitating research on the growth of platforms such as Uber and Airbnb and their influences on individuals, businesses, and communities, in light of the pandemic and otherwise.

In this study, we focus on Airbnb, a popular short-term home rental platform that has emerged as an important part of the shared accommodation economy. With its rapid proliferation in large, urban metropolitan areas, cities have grappled with depletion of their housing stock, increase in rental prices, and issues related to affordable housing, gentrification, and discrimination on the platform. This has necessitated research on understanding motivation of individuals who rent their homes on Airbnb. This is underpinned by where they live and who they are in terms of their demographic, social, economic, occupational outlooks, and attitudes towards trust and sustainability, all of which have been documented as key factors influencing host participation in short-term rental markets (STRs) (Sundararajan, 2016).

Within this context, this study focuses on the supply side of Airbnb for the city of San Francisco, Airbnb's birthplace. As a popular tourist destination, and a locus of technology innovation, Airbnb properties have diffused into almost all districts and neighborhoods of the city. However, the spatial patterns of this diffusion have not been studied previously. It is also unclear if the motivations of San Francisco's Airbnb hosts to rent their residences are the same as their counterparts in other major markets or if location-specific factors influence such choice. Furthermore, as discussed earlier, the effect of the pandemic on supply-side participation in San Francisco's STR markets is unknown at this time.

Given this context, this paper analyzes spatial patterns and socioeconomic influences on host participation in Airbnb in San Francisco. The research questions are: (1) What are the spatial patterns and variations of host participation in San Francisco's STR markets, measured by Airbnb property density during 2019-22? (2) What are the spatial clusters and outliers of Airbnb's STR markets in San Francisco and what

URI: https://hdl.handle.net/10125/107067
978-0-9981331-7-1

are their demographic, economic, and occupational attributes? (3) What are the influences of demographic, social, economic, occupational attributes, as well as hosts' attitudes toward trust and sustainability, and proximity to POIs on Airbnb property density during 2019-22?

The geographic unit of analysis in this paper is census tracts. This choice is motivated by the fact that San Francisco's neighborhoods are diverse in terms of their race/ethnic compositions, proximity to points of interest of cultural and tourist significance, as well as economic attributes. To account for this heterogeneity, we study Airbnb property density at the census tract level. Also, by virtue of being a standard census geography, robust demographic and socioeconomic data is available for census tracts in the U.S.

The remainder of this paper is organized into sections on literature review of geographic analysis of STR markets and key determinants of host participation in Airbnb, followed by the study's conceptual model, methodology and data, findings and their implications, culminating in conclusions and directions for future research.

2. Literature Review

Considerable prior work has studied the geographic distributions of Airbnb properties (Gutierrez et al., 2017; Rabiei-Dastjerdi et al. 2022; Sarkar, Koohikamali, and Pick, 2017), Airbnb hosts (Edelman & Luca, 2014; Franco et al., 2016; Ma et al., 2017), and determinants of Airbnb property densities (Gilheany et al., 2015; Sarkar, Koohikamali, and Pick, 2020; Zhang & Chen, 2019; Zhang and Fu, 2022). There is limited research on conceptual models relevant to Airbnb (Constantinou et al., 2017; Zhang & Chen, 2019). Airbnb is of keen interest because of its dominance in STR markets, its rapid growth, and the availability of large, worldwide datasets (InsideAirbnb.com, 2023). This literature review is organized into sections on characteristics and guest perception of hosts, spatial distribution of Airbnb property densities, and findings on determinants of Airbnb property densities.

Characteristics and guest perception of Airbnb hosts: There is considerable prior work on the behavioral characteristics, profiles, and guest perception of Airbnb hosts (Guttentag, 2019). Two relevant studies are discussed. The first concerned the prevalence and characteristics of Airbnb super hosts in the San Francisco-Oakland, Hayward metropolitan area (Gunter, 2018). A super host is a host who exhibits passion in making certain each guest has the best experience. They are chosen by the Airbnb platform firm based on four criteria reflecting high guest ratings and attentiveness (Gunter, 2018). They relate to this paper since they represent the best of hosts in the same geography as this paper. The paper considered the relative importance of four criteria to qualify as a super host using 40,600 listings spanning 2014-16. The most important criterion is excellent guest ratings, followed by host's reliability of cancellation policy, responsiveness, and achievement of high demand. Exogenous factors were also considered, with the findings that hosts were more important for property location in or near to San Francisco downtown, homes/apartments versus private/shared rooms, short-term rental periods, year-round availability, many photos on web listings, and instantaneous booking. The exogenous factor of central downtown location is consistent with the current research, as is the determinant of nearness to points of interest (POIs).

A second host behavioral study (Ma et al., 2017) focused on host's self-disclosure on Airbnb leading to guest perception of the trustworthiness of the host. This study performed detailed qualitative analysis of the process steps for hosts to be perceived by guests as trustworthy. Perceived trustworthiness was based on 1200 randomly selected host profiles, which served as an experimental dataset for Amazon Turk crowd workers to measure trust. The research found a wide variety of self-disclosure strategies among hosts, which suggests that platforms play an important role in facilitating sense of trust by sharing-economy guests. This study provided support for inclusion of attitude toward trust in the present study.

Spatial distribution of Airbnb property densities in San Francisco and other major US cities: Studies of the spatial distribution of Airbnb properties in New York City, Dublin, and Barcelona provide a backdrop of methods and findings that inform our study of the city of San Francisco. Research on the Airbnb's spatial distribution for NYC utilized Airbnb and US government data at the zip code level to spatially analyze host property density levels throughout the city limits (Sarkar et al., 2020). The cluster locations of high densities of host properties for years 2015-2017 were in South Manhattan, Northwest Brooklyn and West Queens, and revealed little longitudinal change. The high-density areas were identified by k-means cluster analysis and characterized with respect to independent socioeconomic factors. The two highest density zip codes in the southern parts of Manhattan and northern Brooklyn directly across the Hudson revealed especially high factors of professional, scientific, and technical employment and hotel/lodging employment. The pattern of central city location correspond to our present study, while the

characterization of the central clusters resembles the present study's findings.

Similar geographic results were confirmed in prior studies of Dublin metropolitan area (Rabiei-Dastjerdi et al., 2022) and municipality of Barcelona (Gutierrez et al., 2016), among others. For Dublin, not only were Airbnb locations identified but also text analysis of guest reviews, with subsequent machine learning analytics, as well as analysis of georeferenced social media data, yielded variables that were used to predict areas of emerging hotspots as well as identify locations in Dublin with elevated potential to experience gentrification (Rabiei-Dastjerdi et al. 2022). For Barcelona, the spatial distribution of Airbnb properties was mapped and compared to the spatial distribution of hotel properties. Findings indicated that Airbnb was highly concentrated in the central city sections of the city, while hotels were also concentrated in the center, but showed more dispersion to the periphery than Airbnb's. Furthermore, the Airbnb properties benefitted more than hotels from the proximity to tourist-favored POIs in or near the center (Gutierrez et al., 2016).

Findings on determinants of Airbnb property densities: There is an emerging body of work on determinants of Airbnb property concentrations and densities (Eugenio-Martin, 2014; Gilheany et al., 2015; Sarkar et al., 2020; Zhang & Chen, 2019; Adamiak et al., 2019; Zhang and Fu, 2022). The studies most relevant to the present one concern Los Angeles (Zhang & Fu, 2022) and NYC (Sarkar et al., 2020).

The spatial distribution of Airbnb supply of properties in Los Angeles city was analyzed from 2014 to 2019 (Zhang & Fu, 2022). This study included spatial autocorrelation to show global agglomerations and local indicators of spatial association (LISA) to examine local relationships of adjacent geographic units (Anselin, 1995). The results show a strong tendency for a center-periphery pattern. The study also examined determinants of Airbnb supply. The most significant determinants consistently were total population, total housing units, number of POIs, and distance to one of the polycentric centers.

The second study of NYC (Sarkar et al., 2020), already mentioned in the last section, performed stepwise OLS regression on 14 demographic, economic, social capital, green-consumption attitude, and trust variables, finding that the most important correlates were gender ratio, African American race, Asian and Hispanic ethnicities (inverse), dependency ratio (inverse) median household income (inverse), professional/scientific/technical occupation, and green-consumption attitude. These findings bear on the present research and detailed comparison with San Francisco, will be possible.

3. Conceptual Model

Conceptual models for the sharing economy are in flux and continue to be developed, with new opportunities in the emerging post-covid world. For this study, three models were considered, the Chaperone model (Constantiou et al., 2017), Convenience theory (Zhang & Chen, 2019), and the Spatially Aware Technology Utilization Model (SATUM) (Pick and Sarkar, 2016), with SATUM the choice, partly based on its uniqueness among the three in incorporation of spatial concepts. The other theories, which have advantages and are referred to in this research, will be briefly discussed.

The Chaperone model is high on emphasizing the competition between sharing economy providers, but low on the owner's control of the sharing economy platform. Airbnb is considered one of its best examples. Hosts have minimal control over the Airbnb platform. Airbnb property hosts are told by the Airbnb platform about the supply and demand situation of the market, but they can set their own prices. The platform thus encourages a host to be entrepreneurial and competitive with other hosts. Airbnb acts as a mentor to hosts but does not usually strictly control their business strategies, but rather only quantifies norms for basic services, such as maximum of 24-hour response times to customer inquiries. The chaperone has "boundary fluidity" i.e., the fluctuating tradeoff between being coddled by the platform and competing with other hosts (Constantiou et al., 2017). Although it provides a good understanding of the Airbnb market dynamics, it has no recognition of location of stakeholders, so is not suitable as the primary model.

Convenience theory focuses on the convenience of a product or service as a key to customer decisions to adopt (Berry, 2002). Convenience, a multi-dimensional construct, can be disaggregated into categories, such as an Airbnb in the city center, the customers of which experience convenience in shorter time and less effort (Zhang & Chen, 2019). Convenience theory can account for important decision-making tradeoffs made by Airbnb customers. Location clearly plays a role for Airbnb (Zhang & Chen, 2019), so this theory incorporates a limited perspective of location. However, the location is a customer perception, rather than actual locations enabling the advantages of spatial analysis to be applied. Hence, it is not adopted for the present study.

SATUM theory supports investigations that concern the combined influences of a variety of social, economic, social capital, trust, and geographical

proximities to understand the determinants of concentrations of STRs (Sarkar et al, 2020, 2017), measured as densities of Airbnb properties within a city. More specifically, the SATUM model posits that dependent factors (property densities), and independent variables -- social, economic, points of interest, social capital, and trust can, first of all, be explored using mapping, cluster analysis, and descriptive statistics. Then confirmatory analysis can be applied to estimate the determinants of the Airbnb dependent variables. For the present study, OLS stepwise regression analysis is utilized, with testing for spatial autocorrelation, will be discussed in the methodology section.

The specific independent and dependent variables for the SATUM model are shown in Figure 1. The remainder of this section will justify the choice of the 16 independent variables, organized to seven groups, and the dependent variable.

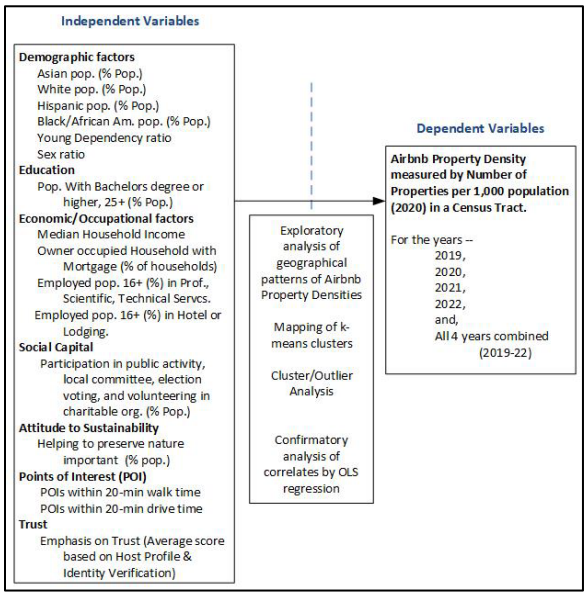

Figure 1. SATUM Conceptual Model.

Demographic factors: Racial/ethnic variables were utilized in a prior study of Airbnb densities in NYC, with proportion of Black population found to be significant (Sarkar et al., 2020). Asian and Hispanic ethnicities have been found to be significant in the same study.

Education: This variable was included in prior studies of Airbnb densities and found to be a significant positive influence (Zhang & Chen, 2022; Sarkar et al., 2017).

Economic/occupational factors: Employment was a positive influence on Airbnb densities in a study of the city of Chicago and New York City (Zhang & Chen, 2022), while median household income was an inverse factor for Airbnb densities in NYC (Sarkar et al., 2020).

Social capital: Although included in Airbnb studies infrequently, it was not significant for Airbnb densities in NYC (Sarkar et al., 2020), but has been a significant variable in studies of technology densities in US states (Pick and Sarkar, 2015).

Attitude to sustainability: This variable was included in a study of Airbnb densities in NYC and was significant for Airbnb densities of private rooms and shared rooms for the combined years of 2015-2017. We posit that it is related to Airbnb densities since the sharing economy is often considered a sustainable use of otherwise wasted resources (Sundararajan, 2016).

Points of interest (POI): This factor was significant in a study of Airbnb densities for three cities, NYC, Los Angeles, and Chicago (Zhang & Chen, 2022) and also for two polycentric downtowns with many POIs for Airbnb total rooms rented in areas of Los Angeles (Zhang and Fu, 2022).

Trust: Trust has been studied and validated as an important factor in a qualitative study of Trustworthiness of Airbnb host profiles (Ma et al., 2017). It was included, although not significant is a study of NYC (Sarkar et al., 2020).

4. Methodology and Data

The methodology consists of multiple steps. First the dataset for the study was compiled from a variety of data sources described later in this section. Data was collected for Airbnb property listings in San Francisco CA for the years 2019-22. Individual listings were aggregated at the census tract level for the city of San Francisco for each year 2019-22 as well as the for the overall 4-year period. While aggregating over 4 years at the census tract level, duplicate listings were removed. The aggregate number of listings was then divided by 1,000 population for the year 2020 to obtain a normalized measure of Airbnb property density at the census tract level. Post normalization, we observed that 18 census tracts had extremely high Airbnb property densities. Upon further examination, they were found to be either sparsely populated with a 2020 population of 500 inhabitants or fewer, or an extremely small number of listed Airbnb properties. These tracts were removed to reduce bias leaving a sample of n=223 census tracts for the study area. The Airbnb property density measures are the study's five dependent variables that are indicative of host participation in the STR marketplace in San Francisco and a gauge of supplier participation in the short-term shared accommodation economy.

Data was also compiled for 16 independent variables, attributes that are likely to explain the variations in Airbnb property density. These 16 variables are shown in Figure 1 and span demographic, educational attainment, economic and occupational attributes, indicators of social capital, attitude toward sustainability, trust, and points of interest. Correlation analysis was applied pairwise for the independent variables using the Pearson correlation coefficient to diagnose multicollinearity. No significant issues were observed.

Once all data was compiled and variables were normalized, descriptive statistics were calculated for the 5 dependent and 16 independent variables. To analyze spatial patterns of Airbnbs, the density dependent variables were mapped using a Geographic Information System (GIS). GIS mapping often provides useful visual cues about agglomerations, clusters and outliers in the dataset. Independent variables such as POI locations were overlaid to provide additional context to the spatial patterns of Airbnb densities. K-means cluster analysis was then applied to determine clusters of tracts that are similar in terms of their Airbnb property densities. The clusters were mapped and also characterized in terms of their demographic, economic, occupational, social, and other pertinent attributes such as POI locations.

Recognizing that K-means is an inherently non-spatial data mining method, spatial agglomeration of Airbnb property densities was analyzed by applying spatial autocorrelation analysis. Moran's Index, a diagnostic of global spatial autocorrelation was calculated for each dependent variable as follows:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{i=1}^{n} z_i^2}$$

z_i is the deviation of an attribute for feature i from its mean, $w_{i,j}$ is the spatial weight between features i and j, n is the total number of features, and S_0 is the sum of all spatial weights. Moran's I measures the extent of spatial autocorrelation of each dependent variable. Moran's I test is inferential; the null hypothesis is that the values of a variable are randomly distributed spatially. The test statistic ranges in value between –1 and +1. Moran's I statistic value close to 0 for a dependent variable (Airbnb property density) would indicate spatial randomness while values close to –1 and +1 indicate the presence of spatial bias for a dependent variable that needs to be accounted for while examining associations of independent variables with the dependent variable in question. Interpretation of Moran's I is performed using the p value for statistical significance (if p is not significant, the variable is randomly distributed spatially). Further, if the Z score is positive, the values of a variable are more geographically agglomerated (high values located near high ones and low values near low ones). If it is negative, the spatial pattern resembles a "checkerboard" pattern, in which high values are surrounded by low ones and vice versa (Moran 1950). Since Moran's I is unable to uncover spatial heterogeneity and identify spatial clusters and outliers, Local Indicators of Spatial Association (LISA) analysis was also conducted to determine spatial clusters and outliers of Airbnb density in the study area. Clusters and outliers are mapped in a GIS to visualize and subsequently contextualize statistically significant hotspots, coldspots, and outliers of Airbnb activity in San Francisco.

Finally, Ordinary Least Squares (OLS) regressions are conducted in stepwise manner with significance levels of equal or less than 0.05 to test the posited associations of independent variables with the dependent indicators of Airbnb density to understand host motivations for participation in San Francisco's STR markets. The Variance Inflation Factor (VIF) was calculated for each regression to diagnose multicollinearity, and a VIF cutoff of 5.0 was used. No multicollinearity problems were detected.

Airbnb property listing data were collected from InsideAirbnb.com, a data source that has been used in prior studies (Rabiei-Dastjerdi, 2022; Dudás et al., 2017). A total of 28,474 Airbnb listings data was collected for 4 years, 2019-2022 covering one year prior to the COVID-19 pandemic and three more years of data spanning the pandemic. For each year, listings data were collected for month of August for the sake of seasonal consistency. Data for a number of demographic, economic, occupational, and educational attainment independent variables were collected from Esri Demographics, an authoritative source of GIS data from the largest GIS software company in the world. Indicators for social capital and attitude toward sustainability was calculated using datasets from Esri/GfK MRI DoubleBase Survey. These data represent integrated information from 4 consumer surveys. GfK MRI surveys through interviews over 20,000 adults in the U.S. annually, with focus on summing over 6,000 products, attitudes, and lifestyles (Esri, 2020). POI data was collected from SafeGraph. SafeGraph's POI dataset includes places where consumers can spend money, time, or both, and include restaurants, grocery stores, malls, parks, hospitals, museums, art galleries, theaters, aquariums, and more. Finally, an indicator for trust was calculated using listings data from InsideAirbnb.com, which provides host profile and identity verification attributes for each listing.

All of the variables, their data sources, and descriptive statistics are in Table 1.

Table 1. Variables and Descriptive Statistics.

	Dependent Variable (*)	Source	Year of Data	Minimum	Maximum	Mean	Std. Dev.
Airbnb Property Listings per 1,000 pop. (2020)	All accommodations, all years	InsideAirbnb.com	2019-22	0.8308	483.4974	36.6863	40.2676
	All accommodations, 2022		2022	0.1880	125.0724	5.9970	9.2635
	All accommodations, 2021		2021	0.1452	75.8541	4.0103	5.3179
	All accommodations, 2020		2020	0.1763	65.4314	4.7342	5.6105
	All accommodations, 2019		2019	0.1995	66.5895	4.5977	5.0783
	Independent Variables*						
Demography	Young Dependency Ratio (Pop. 0-19 yr/ Pop. 20-64 yr)	Esri Demographics	2019-22	0.0024	0.7454	0.2069	0.1084
	Male Age 20+/Female Age 20+		2019-22	0.5715	2.5520	1.1010	0.3410
Race & Ethnicity	Black		2019-22	0.0000	0.5908	0.0492	0.0729
	Asian		2019-22	0.0386	0.9451	0.3287	0.1880
	Hispanic		2019-22	0.0000	0.5862	0.1436	0.1094
	White		2019-22	0.0382	0.9316	0.4731	0.2040
Education	Bachelors Degree		2019-22	0.0608	0.5267	0.2909	0.0976
Occupation & Economy	Median Household Income ($)		2019-22	18,650.00	200,001.00	123,896.93	46,704.68
	2020 Owner Households with a Mortgage/Total Households		2019-22	0.0000	0.6685	0.2567	0.1577
	Prof, Scientific, Technical Serv. Em		2019-22	0.0000	13.8078	0.1810	1.0360
	Hotel/Lodging Emp.		2019-22	0.0000	1.4878	0.0356	0.1643
Social Capital	Average of participation in public activity, serving on local committee, voting in election, and volunteering for charitable org	Esri/GfK MRI DoubleBase Survey	2019-22	0.2045	0.7250	0.3754	0.0898
Attitude for Greener Consumption	Helping to preserve nature very important		2019-22	0.4123	1.2802	0.6856	0.1350
Points of Interest	POI within 20 min Walk Time	Safegraph	2019-22	3.0000	1413.0000	383.4800	375.9880
	POI within 20 min Drive Time		2019-22	1774.0000	4028.0000	3634.2200	471.8840
Trust	Emphasis on Trust 2019-2022 - avg. of Emphasis on Trust 2019-2022	InsideAirbnb.com	2019-22	3.2881	7.9875	5.2784	0.6746
	* n = 223						

5. Spatial Patterns of Airbnb in San Francisco

Airbnb property densities were mapped using a Geographic Information System (GIS). GIS mapping provides visual cues about spatial patterns of property densities and reveals the extent of host participation in San Francisco's STR markets. For each of the years, 2019 to 2022, Airbnb property density is found to be the highest (120-over 160 listings, per 1,000 population) in the northern and eastern parts of the city. These areas (dark and light blue census tract polygons in Figure 2) are comprised of census tracts that are in Downtown and Tenderloin neighborhoods north of Market Street and east of the 101 Freeway near Union Square, which is a hotspot of hotel and lodging establishments. Moderately high concentrations of Airbnbs (teal) are in census tracts that are contiguous to the previous set of census tracts (very high Airbnb densities) in the Downtown/Civic Center sub-region as well as in the Haight Ashbury neighborhood near Golden Gate Park.

Moderate Airbnb property densities (light green) are found in the Northeast Waterfront Historic District that includes Sane Francisco's popular tourist attraction Pier 39. Moderate concentrations are also noticed in Chinatown north of Market Street, and in Mission District, Bernal Heights, Noe Valley, south of Market Street. Parts of the Haight Ashbury neighborhood, adjacent to Golden Gate Park are also moderate in Airbnb property density. Property densities per capita are low (off white color in Figure 2) in the rest of the city, with densities between 1-40 properties per 1,000 population (in 2022). These spatial patterns are largely consistent longitudinally, and as an aggregate over the 2019-22 period.

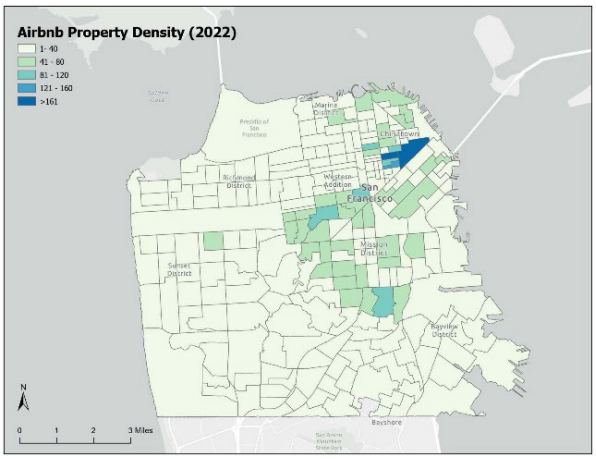

Figure 2. Airbnb Property Density, San Francisco, 2022.

5.1. Clusters of Airbnbs in San Francisco

K-means cluster analysis, an unsupervised data mining method was applied to determine the agglomerations that are similar in terms of Airbnb property densities in San Francisco. The clusters were mapped in a GIS and were characterized in terms of their demographic, economic, occupational attributes, and POIs. K-means clustering with different values of k = 3, 4, and 5 were tested. k=5 was selected based on the stability of cluster members.

K-means (k=5) cluster analysis reveals that of the 223 census tracts, a vast majority (143, or almost 64%) have low to very low property densities. These census tracts in Cluster 5 (Figure 3 and Table 2) are located in San Francisco's Sunset, Richmond, Marina, and Bayview Districts. Cluster 4's 65 census tracts located predominantly around the Mission District and Haight Ashbury and Bernal Heights neighborhoods have moderate-low property density. Most of the cluster 4 and 5 census tracts are spatially contiguous. Clusters 3 has 11 census tracts representing relatively high property density south of Market Street sub-region, in Mission District, Downtown, and in North Beach that includes Pier 39. Clusters 1 and 2 are comprised of 4 census tracts that have the high to very high property density located in San Francisco's Financial District, north of Market Street.

Unsurprisingly, these neighborhoods are in very close proximity to tourist attractions such as museums, theaters, and popular ethnic enclaves with cultural appeal such as Chinatown. They have almost five times as many POIs on the average (Table 2) within a 20-minute walk-time from the centroid of census tracts

within clusters 1 and 2 than their counterparts in cluster 5. The same is true for North Beach in cluster 3, which includes Fisherman's Wharf and Pier 39 inclusive of museums, retail and shopping establishments, restaurants, and other tourist attractions. Also, these clusters have much higher proportions of hotel and lodging employment (Table 2). Conversely, median household income in clusters 1 and 2 is comparatively much lower than cluster 5 and much lower than clusters 3 and 4. It may be likely that Airbnb hosts in clusters 1 and 2 require rental income from their Airbnb properties to supplement their regular income.

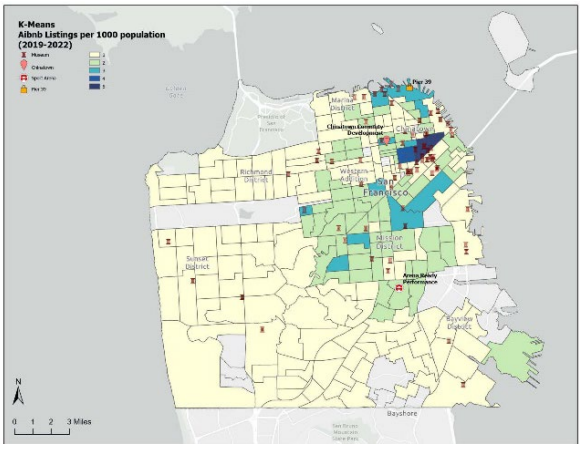

Figure 3. K-means Clusters of Airbnb Property Density, San Francisco, 2019-22.

Table 2. K-Means Cluster Characteristics, Airbnb Property Density, San Francisco, 2019-22.

	Attributes	Cluster 1 (Highest Property Density)	Cluster 2 (High)	Cluster 3 (Moderate)	Cluster 4 (Moderate-Low)	Cluster 5 (Low-Very Low Property Density)	Max/Min Ratio	Max/Min Ratio without Cluster 1
	Number of Census Tracts	1	3	11	65	143		
		Average Cluster Center						
	Airbnb Property Density (2019-22)	0.483	0.173	0.100	0.052	0.022	22.38	8.01
		Averages of Attributes						
Demography	Young Dependency Ratio (Pop. 0-19 yr/ Pop. 20-64 yr)	0.189	0.115	0.114	0.157	0.239	2.10	2.10
	Male Age 21+/Female Age 21+	1.052	1.036	1.342	1.304	0.992	1.35	1.35
Race Ethnicity	Asian (%)	0.401	0.308	0.224	0.214	0.389	1.87	1.81
	White (%)	0.407	0.473	0.619	0.584	0.412	1.52	1.50
	Hispanic (%)	0.088	0.172	0.110	0.150	0.143	1.95	1.55
	Black (%)	0.020	0.057	0.038	0.046	0.051	2.89	1.51
Education	Bachelors Degree (%)	0.358	0.266	0.373	0.334	0.265	1.41	1.41
Occupation and Economy	Median Household Income ($)	73,650	51,080	143,611	136,568	118,500	2.81	2.81
	2010-2014 ACS Owner Households with a Mortgage (%)	0.065	0.017	0.159	0.224	0.285	17.10	17.10
	Prof, Scientific, Technical Serv. Emp. (%)	13.808	0.068	0.163	0.154	0.101	203.63	2.40
	Hotel/Lodging Emp. (%)	1.356	0.637	0.095	0.014	0.019	100.31	47.12
Social Capital	Average of participation in public activity, serving on local committee, voting in election, and volunteering for charitable org (%)	0.324	0.418	0.455	0.424	0.347	1.40	1.31
Attitude for sustainability	Helping to preserve nature very important (%)	0.771	0.834	0.783	0.735	0.652	1.28	1.28
Points of Interest	Poi_20 min Walk	1337	1343	605	493	290	4.63	4.63
	Poi_20 min Drive	3874	3899	3827	3884	3499	1.11	1.11
Trust	Trust	5.010	5.164	5.228	5.443	5.212	1.09	1.05

Spatial autocorrelation analysis is used to examine the extent of spatial agglomeration of Airbnb property densities in San Francisco's census tracts. Moran's Index is calculated as discussed in Section 4. For each year, 2019 to 2022, Airbnb property density is found to be statistically significantly clustered in San Francisco, with the lowest Moran's Index value of .180, significant at the .001 level in 2020 and the highest index value of .225, also significant at the .001 level in 2022. When unduplicated properties are aggregated over the 4-year period, the densities are again clustered spatially, with Moran's I of .176, significant at .001 level. Moran's I values are reported in bottom in Table 3. Finally, spatial clustering seems to have a slight longitudinal trend, with Moran's I values increasing over time from 2020 to 2022. This positive and statistically significant spatial clustering of Airbnb properties has methodological implications when determining associations of independent variables with these dependent attributes.

Since spatial autocorrelation analysis using Moran's I is unable to uncover spatial heterogeneity and identify spatial clusters as well as outliers, we used Local Indicators of Spatial Association (LISA) analysis (Anselin, 1995) to capture local spatial autocorrelation. LISA analysis reveals that statistically significant hotspots of Airbnb properties are in the Financial District, north and south of Market Street, one of San Francisco's main thoroughfares. Statistically significant hotspots are also found in the Tenderloin, Nob Hill, Chinatown, and North Beach areas in the northeastern part of the city. Another hotspot is in the city's core, in the northwestern corner of Mission District, and also parts of the Western Addition and Haight-Ashbury neighborhoods. We note that the 101 Freeways, a gateway for visitors driving into San Francisco hugs many of these hotspots. Interestingly, statistically significant low-high outliers adjoin the hotspots and are located in these very same neighborhoods, predominantly in the northeastern periphery of the city. Coldspots of Airbnb are dominant in the west and south of San Francisco including Richmond and Sunset districts to the north and south of Golden Gate Park respectively.

In summary, unlike Los Angeles (LA), the spatial agglomeration of Airbnbs in San Francisco does not have a center-periphery pattern (Zhang and Fu, 2022). While Airbnb hotspots in Los Angeles are concentrated in the central areas of the city, those in San Francisco are in the northeastern periphery close to the Pacific Coast. On the other hand, coldspots in LA are in the city's northern and southern peripheral areas, while in San Francisco, they are in the western and southern parts of the city, with a much higher proportion of owner-occupied households with a mortgage but with much higher median household incomes than those in the hotspots. It can be inferred that due to higher incomes, homeowners in the city's

coldspots are less likely to rent their properties on Airbnb for supplemental income.

6. Determinants of Airbnb Property Density

Ordinary Least Squares (OLS) regressions were conducted to determine the associations of demographic, occupational and economic, social capital, attitude toward sustainability, proximity to POIs, and attitude toward trust on the dependent indicators of Airbnb property densities, estimated by Airbnb property densities, for years 2019, 2020, 2021, 2022, and all four years combined. Note that Airbnb property densities indicate the extent of host participation in San Francisco's short-term residential rental markets for the sample of n = 223 census tracts.

OLS regressions (in Table 3) reveal that the dominant correlates of Airbnb property densities are professional, scientific, technical services (PSTS) employment, Asian population, points of interested located within 20-minute walk times, and hotel and lodging employment. Longitudinally, these variables are consistently associated with Airbnb property densities in San Francisco. This is interesting and indicates that the factors influencing Airbnb hosts' participation in San Francisco's short-term rental markets did not fluctuate as a consequence of the covid-19 pandemic. That said, it is noticeable that the explanatory power of this quartet of demographic (race/ethnic), occupational, and POI variables increased from 51.9% in 2019 to 61.5% in 2021, with a slight decrease to 59.3% in 2022, as evident from the adjusted r-squared values in Table 3. for listings aggregated over the period 2019-22, the regression model explains 60% of the variation in property density. Overall, these coefficient of determination values indicate the robustness of the study's conceptual model. The Variance Inflation Factor (VIF) in each regression does not exceed 5.0 threshold; hence, multicollinearity is not a problem.

PSTS and hotel/lodging employment are both positively associated with Airbnb property densities. PSTS includes management, business, financial, computer, mathematical, healthcare, life, physical, and social science related occupations, as well as educational instruction, legal, and office and administrative support occupations. Those employed in PSTS are likely to be more proficient users of technology, making them more likely to use apps on laptops and mobile devices to participate in the sharing economy. The positive association of PSTS employment with Airbnb property densities was also found in a prior study focused on New York City (Sarkar et al., 2020). It is also consistent with reports that on-demand companies are ushering white-collar workers into the gig economy, taking advantage of their electronic skills and related digital privileges, and their willingness to leverage such skills to generate supplemental income (Rossa, 2015).

Table 3. Regression Results

	Independent Variables	Dependent Variables: Airbnb Property Listings Density (Listings/1,000 pop.)				
		2019 - 2022	2019	2020	2021	2022
Demography	Young Dependency Ratio (Pop. 0-19 yr/ Pop. 20-64 yr)					
	Male Age 21+/Female Age 21+					
Race / Ethnicity	Asian (%)	-0.288***	-0.359***	-0.279***	-0.256***	-0.245***
	White (%)					
	Hispanic (%)					
	Black (%)					
Education	Bachelors Degree (%)					
Occupation and Economy	Median Household Income					
	2010-2014 ACS Owner Households with a Mortgage/Total Households (%)					
	Prof, Scientific, Technical Serv. Emp	0.471***	0.564***	0.500***	0.466***	0.392***
	Hotel/Lodging Emp. (%)	0.238***		0.207**	0.269***	0.332***
Social Capital	Average of participation in public activity, serving on local committee, voting in election, and volunteering for charitable org (%)					
Attitude for sustainability	Helping to preserve nature very important (%)					
Points of Interest	Poi_20 min Walk	0.177***	0.213**	0.158***	0.177***	0.174***
	Poi_20 min Drive					
Trust	Trust					
	Adjusted R^2	0.596***	0.519***	0.576***	0.615***	0.593***
	VIF	2.244	1.112	2.244	2.244	2.244
	Spatial Autocorrelation (Morans' I)	0.176***	0.208***	0.180***	0.193***	0.225***
	Notes: * p <.05, ** p <.01, *** p <.001					

Hotel/lodging employment is also found to be positively associated with Airbnb property densities, consistently over the four years, as well as with the overall aggregate measure of property density. It has been documented that Airbnb supply often overlaps with established hotels in city tourism (Eugenio-Martin et al., 2019). A prior study has also found that Airbnb supply is associated with traditional tourism accommodation (Adamiak et al., 2019). We reason that the presence of hotel and lodging establishments makes it more likely that those employed in that sector in San Francisco are likely to be employed in the city's STR marketplace. Overall, the effect of the employment duo (PSTS and hotel/lodging) is a novel finding of this study that has not been documented previously for San Francisco.

POIs located within 20 minutes of walk-time from the centroid of census tracts is found to be positively associated with Airbnb property densities in San Francisco. A recent study has found a similar result, for Airbnb supply in Los Angeles. Interestingly, this study was also conducted at the census tract level (Zhang and Fu, 2022). Similar to Zhang and Fu, we reason that as POIs within 20-minute walk time become prevalent, the attractiveness of a neighborhood as a tourist destination increases making it more likely that hosts in such neighborhoods will list their homes for rent on Airbnb.

Finally, Asian population is found to be inversely associated with Airbnb property density in San

Francisco. The inverse association is longitudinally consistent as well including when property densities are aggregated over the period 2019-22. Prior studies that have examined racial discrimination on STR platforms such as Airbnb have documented that minority race/ethnic groups, particularly Asians hosts in Oakland and Berkeley are likely to earn significantly less income via rent than their White counterparts for similar rental properties (Gilheany et al., 2015). Another study has found that Hispanic and Asian Airbnb hosts, on average, have almost 10% lower list price relative to their white counterparts, in San Francisco, after controlling for neighborhood property values, user reviews and rental unit characteristics (Franco et al., 2016). Researchers have explained that guests may perceive rentals hosted by minorities to be situated in inferior locations, or that the quality of such properties may be inferior to those hosts belonging to majority race/ethnic groups (Edelman and Luca, 2014). The likelihood of lower incomes or other forms of discrimination are likely to be deterrents for Asians to list their properties on Airbnb offering one possible explanation for the inverse association. While this finding for San Francisco has not been previously documented to the best of our knowledge and is therefore novel, it merits further research.

Overall, in San Francisco, the key determinants of host participation in STR markets are race/ethnicity (Asian pop.), occupation (PSTS and hotel/lodging employment), and. These determinants did not vary over longitudinally between 2019 to 2022. Implications are discussed next.

7. Implications of Results

The research findings have implications for Airbnb hosts, potential guests, the public, and the city government. For Airbnb hosts, the high spatial density distribution of properties in San Francisco's financial district, Union Square, and Market Street towards the Bay corresponds to an affluent, high rent area, with many amenities, and points of interest including museums, famous hotels, and large, brand-name stores. An implication is that competition is strong for hosts and they would need to adopt strong marketing and technologies to improve their host profiles as well as make efforts to improve their perceived trustworthiness (Ma et al., 2017). This compact, centralized high-density concentration resembles the concentration of the densest areas in New York City in 2015-2017, located in North Brooklyn and mid to lower Manhattan on the west side. The Manhattan portion is similarly bordering the financial district, while parts of north Brooklyn are affluent. A difference is that the two parts are intersected by the East River. A second high Airbnb density area is Fisherman's Wharf neighborhood, which has Pier 39, museums, and other amenities. It would likewise imply a competitive, high rent area, although competition with nearby hotels would be lessened.

Another set of implications for these dense Airbnb areas in San Francisco is that "touristification," i.e. the authentic character of these city areas may be reduced by an overabundance of tourists, similar to Paris (Guttentag, 2019). Furthermore, housing prices may become inflated, and long-time locals may be priced-out and feel displaced by the growth in Airbnbs. At a certain point, this situation can boil over and lead to increased regulation and growth restraints.

The implication of these changes for the city government is that it might make efforts to decentralize the distribution of Airbnbs through regulations. It could put in incentives to favor the economic advantages to hosts to locate Airbnb properties in more peripheral parts of the city. This could also be done through transport improvements, so tourists could more easily access the most prominent points of interest in the central downtown area.

The implications of the regression analysis of correlates of Airbnb property density are, first of all, that Airbnbs are proximate to higher levels of PSTS employees, or vice-versa, PSTS employees prefer to live near higher densities of Airbnbs. San Francisco metropolitan area is well known for its strengths in high-tech employment, so high-tech workers may find the high Airbnb areas attractive, especially if they are in temporary job situations and could benefit by, and afford, renting in the downtown surrounded by POIs. Since Asians tend to have reduced correlation with high Airbnb areas, an implication is that they may feel less comfortable to live in the central downtown, but prefer the stability of peripheral suburban areas. This disinclination of Asians to participate in STRs is an area that future research could investigate. The strong association with hotel/lodging employment may imply that workforces are somewhat shared between hotels and Airbnbs. For instance, there could be common job markets that serve the two related industries in occupations such as cleaning, maintenance, tour guides, etc. This correlation calls for more research to confirm the reason for and extent of the apparent employment overlap. From the standpoint of city government, an implication is that it might combine Airbnb and hotels in planning for and forecasting employment needs in common occupations.

8. Conclusions and Limitations

This study analyzes spatial patterns and socioeconomic influences on San Francisco's short-term home rental markets. Geographic mapping of Airbnb property densities provides interesting insights about the spatial distributions of Airbnb property densities in the city. Unlike other major Airbnb markets such as Los Angeles in which Airbnb hotspots are in the city's core and coldspots are in the city's periphery, Airbnbs in San Francisco are predominantly clustered in neighborhoods in the city's northeast that are proximate to tourist attractions, have higher concentration of hotel/lodging employment, and have lower median household income. While this may allude that earning supplemental income might be a motivation for Airbnb hosts, regression findings reveal that the key variables influencing the variation of Airbnb property density are occupational (PSTS and hotel/lodging employment), minority population (Asian), and POIs that are likely to attract tourists. These influences were found to be longitudinally consistent over 2019-22 period indicating the host participation motivations in Airbnb's short-term rental markets were unlikely to be affected by the pandemic.

Due to ecological fallacy, a pitfall of geospatial data, it remains to be verified if these findings hold true for other geographic units of study, such as zip codes. In addition, this paper did not delve into specific types of properties such as entire home/apartments, private rooms, and shared rooms. These are some limitations of the study and present opportunities for further research. The findings have implications for the expansion of STR platforms in San Francisco, especially as cities are tightening regulations for such platforms to preserve their housing stock. It is also of interest to study how the findings may evolve in light of documented reports of remote working that has caused an exodus of tech-sector employees from the San Francisco-Bay Area.

9. References

Adamiak, C., Szyda, B., Dubownik, A., & García-Álvarez, D. (2019). Airbnb offer in Spain—spatial analysis of the pattern and determinants of its distribution. ISPRS *International Journal of Geo-Information*, 8(3), 155.

Anselin, L. (1995), Local indicators of spatial association – LISA, Geographical Analysis, 27(2), 93-115.

Constantinou, I., Marton, A. and Tuunainen, V.K. (2017), Four models of sharing economy platforms, *MIS Quarterly Executive*, 16(4), 231-251.

Dudás, G., Vida, G., Kovalcsik, T., & Boros, L. (2017). A socio-economic analysis of Airbnb in New York City. *Regional Statistics*, 7(1), 135-151.

Edelman, B. G., & Luca, M. (2014). Digital discrimination: The case of Airbnb. com. Harvard Business School NOM Unit Working Paper, (14-054).

Esri. (2020). Methodology Statement: 2020 Esri Market Potential. Technical Paper. Retrieved from https://downloads.esri.com/esri_content_doc/dbl/us/J9672_Market_Potential_DB_Methodology_Statement_2020.pdf.

Eugenio-Martin, J. L., Cazorla-Artiles, J. M., & González-Martel, C. (2019). On the determinants of Airbnb location and its spatial distribution. *Tourism Economics*, 25(8), 1224-1244.

Franco, J., Kakar, V., Voelz, J., & Wu, J. (2016). Effects of host race information on Airbnb listing prices in San Francisco. MPRA Paper No. 69974, Available online at https://mpra.ub.uni-muenchen.de/69974/.

Gilheany, J., Wang, D., & Xi, S. (2015). The model minority? Not on Airbnb. com: A hedonic pricing model to quantify racial bias against Asian Americans. *Technology Science*. Available online at https://techscience.org/a/2015090104/.

Gutierrez, J., Garcia-Palomares, J.C., Romanillos, G., and Salas-Olmedo, M.H.. (2017) Airbnb in tourist cities: Comparing spatial patterns of hotels and peer-to-peer accommodation. *Tourism Management*, 62, 278-291.

Guttentag, D. (2019). Progress on Airbnb: A literature review, *Journal of Hospitality and Tourism Technology*, 10(4), 814-844.

Ma, X., Hancock, J.T., Mingjie, K.L., and Naaman, M. (2017). Self-disclosure and perceived trustworthiness of Airbnb host profiles. Proceedings of Conference on Computer-Supported Cooperative Work and Society, Portland, Oregon, 2397-2409.

Moran, P. A. (1950). A test for the serial independence of residuals. *Biometrika*, 37(1/2), 178-181.

NTIA. (2022). Digital Nation Data Explorer. National Telecom & Information Admin, Washington, D.C.

Pick, J.B., & Sarkar, A. (2015). *The global digital divides: Explaining change*. Berlin: Springer.

Rabiei-Dastjerdi, H., McArdle, G., and Hynes, W. (2022). Which came first, the gentrification or the Airbnb? Identifying spatial patterns of neighborhood change using Airbnb data. Habitat International 125:102582.

Rossa, J. (2015), "The sharing economy, Bloomberg brief", https://newsletters.briefs.bloomberg.com/document/4vz1acbgfrxz8uwan9/front (accessed June 1, 2023).

Sarkar, A., Koohikamali, M., & Pick, J.B. (2020). Spatial and socioeconomic analysis of host participation in the sharing economy: Airbnb in New York City. *Information Technology & People*, 33(3), 983-1009.

Sarkar, A., Koohikamali, M., & Pick, J.B. (2017). Spatiotemporal patterns and socioeconomic dimensions of shared accommodations: The case of Airbnb in Los Angeles, California. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, 107-114.

Sundararajan, A. (2016). The sharing economy: The end of employment and the rise of crowd-based capitalism. MIT Press, Cambridge MA.

Zhang, Z., and Chen, R.J.C. (2019). Assessing Airbnb logistics in cities: Geographic information system and convenience theory. *Sustainability*, 11, 2462, 1-11.

Zhang, Z., and Fu, R.J. (2022). Spatial distribution of Airbnb supply in Los Angeles. *Tourism Analysis*, 27, 467-477.