Customer Behavioral Shifts as a Result of the COVID-19 Pandemic: Are They "Sticky"?

Yilun Xu[1*], Mohsen Bahrami[2*] and Alex Pentland[2]

[1*]Laboratory for Innovation Science at Harvard, Harvard University, Allston, MA, USA.
[2]MIT Connection Science, Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA, USA.

*Corresponding authors: yixu@hbs.edu; bahrami@mit.edu;

Abstract

This study examines the COVID-19 pandemic's impact on customer preferences for shopping destinations and its persistence using a quantitative approach. It uses a modified version of the Huff gravity model to quantify temporal shifts in customer preferences when it comes to choosing a store for shopping purposes. The study uses large-scale mobility and place datasets and census information to analyze department stores in New York City. Using clustering techniques and statistical inference models, this study estimates the immediate and long-term impact of the COVID-19 pandemic on customer behavioral shifts, highlighting the heterogeneity of response among different socioeconomic communities. The findings of this research, show that the proposed model effectively captures the dynamics of shopping location decisions and temporal visit patterns, allowing managers and marketers to understand customer preferences and adapt strategies accordingly. Retailers should use quantitative methods to update assumptions based on new consumption patterns, rather than expecting a complete return to pre-pandemic consumer behavior.

Keywords: COVID-19 pandemic, behavioral shifts, consumer decision process, retail, shopping location decision

1 Introduction

The COVID-19 pandemic, as a worldwide public health emergency, has compelled individuals to alter their daily routines, particularly those residing in urban areas [1]. Mandates such as social distancing and government interventions aimed at reducing the transmission of the disease, such as shelter-in-place orders and the closure of non-essential businesses, coupled with the fear of contracting the virus and its uncertain consequences, prompted individuals to implement self-imposed precautions, disrupting their regular activities [2, 3]. The impact of these measures has been particularly significant in densely populated cities like New York City (NYC), which implemented strict city-wide lockdowns referred to as the "micro-cluster" strategy [4].

Much recent research has illuminated the dynamics of such behavioral changes in various contexts including human mobility [2, 5–7], physical activity (e.g., walking behavior) [8–10], spending patterns [11–16], shopping and consumption [13, 17–20], tendency to work from home [21–23], and adoption of online tools and new technologies [24]. More importantly, many of those studies have highlighted that the pandemic and resulting crisis have disproportionately impacted the lives of different socioeconomic groups, leading to heterogeneous behavioral shifts among them [7, 13, 25–29].

Certain behavioral changes have been documented as interim, temporary, and short-term in nature, resembling responses to exogenous shocks such as the impact of stimulus checks on shopping patterns [11, 12]. On the flip side, some behavioral changes are likely to be enduring and lead to the rise of a new normal even after re-openings and recovery from the economic downturn [30, 31]. Corresponding persisting behavioral shifts could result from enforced acceleration of widespread digital technology adoption (e.g. online shopping, telemedicine, remote working, and virtual meetings) or the emergence of new habits due to pandemic-induced experiences and talent discovery (e.g. cooking and exercising at home) [32, 33].

The impact of COVID-19 on the retail economy is characterized by widespread disruptions and substantial shifts in consumer behaviors across all retail sectors. The retail industry has experienced significant challenges, including disturbances in supply chains resulting in supply shocks, as well as changes in consumer behavior driven by demand shocks caused by panic buying, hesitancy, and financial constraints. [34, 35]. The adaptation of retailers to the pandemic varied between essential and non-essential retail businesses [36]. While non-essential retailers faced forced closures and significant declines in in-store sales, essential retailers faced challenges in meeting heightened demand while ensuring the safety of customers and employees. They were compelled to implement measures such as limiting customer numbers in-store, queuing systems, and restrictions on the number of products customers could purchase [37, 38]. The implementation of such restrictions, coupled with concerns about contamination and associated stress, have profoundly altered consumer behavior and decision-making processes, thereby introducing additional complexity to the already challenging trading conditions faced by the retail sector [39–41]. According to prior studies [2], it has been established that private, self-regulating behavior plays a significant role in explaining the majority of the decline in foot traffic across multiple industries, accounting for more than 75% of the overall decrease. Specifically, when considering essential retail, approximately half of the decrease in foot traffic can be attributed to the presence of restrictive regulations and temporary closures of complementary non-essential retail. In addition to the supply chain disruptions, the economic repercussions have rendered the retail sector less resilient and susceptible to potential recessions, similar to the significant ones it has experienced in the past, resulting in enduring impacts on consumer confidence [16].

While the pandemic posed significant challenges for retailers, it also served as a catalyst for innovation and adaptation, particularly within the department store, general merchandise, and supercenter sectors. These retailers, renowned for their diverse product offerings and immersive shopping experiences, encountered substantial turmoils rising from either temporary closures and decreased foot traffic, or surges in demand leading to operational complexities and an unprecedented pressure to ensure product availability, and maintain a safe shopping environment. Consequently, retailers swiftly embraced omnichannel strategies and enacted digital transformations in response to the rapid shift in consumer demand towards online channels and e-commerce [42]. They accelerated their investments in e-commerce platforms, improved their online shopping capabilities, and implemented contactless delivery options. This allowed customers to continue shopping while adhering to safety guidelines, underscoring that the ability to innovate and adapt has become critical for survival and long-term success of retailers in a post-pandemic retail landscape [43].

Although technology can cater to both hedonic and utilitarian needs, certain consumer needs, such as social needs including personal and psychological factors, remain unfulfilled [32]. For instance, in today's world, retail centers serve not only as commercial establishments but also as vibrant social spaces that facilitate interactions among people from diverse socioeconomic and demographic backgrounds [44, 45]. Moreover, certain services are inherently physical and cannot be adequately provided through virtual channels, underscoring the enduring importance of physical retail spaces despite the increased share of online sales even after the reopening of the economy [46, 47].

Furthermore, from a retailer's perspective, the expenses related to fixed and variable costs associated with physical retail spaces are significant. As a result, it becomes crucial for retailers to strategically optimize the utilization of their stores and facilities. This emphasizes the importance of understanding the dynamics and maximizing the potential of physical retail spaces, ultimately leading to enhanced operational efficiency and improved profitability [48]. Consequently, the significance of examining customer choice factors when making decisions about selecting physical shopping locations becomes evident. By understanding the factors that influence consumer behavior and decision-making processes, retailers can make informed choices about their store locations, layouts, and offerings, ultimately enhancing their competitiveness and meeting the evolving needs of their target customers.

Cruz-Cárdenas et al. (2021) conducted a comprehensive analysis of 70 studies examining consumer behavior and the impact of COVID-19 [49]. Their research emphasizes the significance of employing quantitative methods to estimate the extent of changes in consumer behavior from various angles. Similarly, Salon et al. (2021) collected responses through a questionnaire to assess consumer behavior before, during, and after the pandemic across different economic sectors. Their findings suggest that the changes in consumer behavior triggered by the pandemic have endured throughout the entire United States, emphasizing the significance of employing quantitative methods when preparing for a "new normal" [30].

In summary, the impact of the COVID-19 pandemic on the retail industry and consumer behavior has been widely acknowledged. While previous studies have highlighted the lasting changes in consumer behavior, there is a need for further investigation to measure and quantify these changes accurately. The physical space of retail establishments holds significance from various perspectives, including facilitating social interactions, supporting in-person services and shoppers, and optimizing its utility to mitigate the substantial costs incurred by retailers. In light of these factors, our research aims to quantitatively examine the dynamics and temporal shifts in customer preferences during the decision-making process for selecting a shopping location.

In order to comprehensively analyze and measure the dynamics of shopping location decisions, our paper introduces a new data-driven approach, built upon the well-established Huff gravity model [50]. By employing an enhanced multiplicative competitive interaction version of the Huff gravity model, we aim to understand and quantify the evolving patterns of consumer behavior in relation to shopping location choices, taking into account the impact of contextual factors such as the COVID-19 pandemic. This approach enables us to capture and measure the complex interplay between consumer preferences and various contextual influences, providing valuable insights into the dynamics of shopping location decisions.

Based on the existing body of literature and available data, we undertake a comprehensive analysis of customer shopping location preferences by considering six distinct factors. These factors include the distance between customers and stores, the floor area of the stores, the demographic similarity between customers and the residents of the store's neighborhood, customer loyalty to specific chains, the count and diversity of amenities in the vicinity of the stores. Our objective is to address the following research questions: 1) Did significant shifts occur in customer preferences for shopping locations between the years 2018 and 2021? 2) Are there variations in the responses of different socioeconomic and demographic groups within this context? 3) Do these preference shifts diminish as the pandemic recedes or are they enduring? 4) What is the relative effect of the pandemic on preference shifts for each studied factor, as well as for different customer groups? Answering these question will contribute to a better understanding of customer preferences, aiding businesses and policymakers in adapting strategies and making informed decisions in a rapidly evolving market landscape.

Our study contributes to the existing literature by presenting a new approach that utilizes an improved and modified version of the Huff gravity model that enables us to quantify the changes in consumer preferences regarding their choice of shopping locations, while simultaneously considering multiple attraction factors. We acknowledge that real-world decision-making problems are inherently complex, and relying solely on unidimensional descriptive analytics may lead to oversimplification, limiting our understanding of the dynamics of customer decision-making processes. In contrast, our proposed approach offers a comprehensive evaluation model that can effectively aggregate various factors, providing a more robust and nuanced understanding of these dynamics.

Moreover, our method has the potential to contribute valuable substantive insights into the factors shaping consumer shopping location decisions before, during, and after the COVID-19 pandemic through quantitative analysis. By utilizing our model, we are able to effectively capture the dynamics of shopping location choices and identify changes in visit patterns. These findings can provide managers and marketers with a deeper understanding of evolving consumer preferences, enabling them to adapt and enhance their strategies to remain competitive. Additionally, the validation of our model is demonstrated by the stable parameter values obtained during the pre-pandemic years, indicating the consistency of consumer preferences in the absence of exogenous shocks.

The remainder of this paper is structured as follows. In Section 2, we conduct a comprehensive review of existing literature on the impact of the pandemic on consumer behavior. Section 3 provides a detailed background of the model we used in this study. In Section 4, we outline our proposed methodology, describe the datasets utilized, and explain the analytical framework. We then proceed to investigate the case of retail business, focusing specifically on department stores in New York City (NYC), and present the empirical results and their implications in Section 5. We analyze the decision-making aspects of selecting a physical store for shopping purposes that gained more importance during the COVID-19 pandemic as individuals sought to balance their needs with the risk of infection. Additionally, we explore the heterogeneity in changes to store selection criteria among different socioeconomic communities, aiming to comprehend the persistence of these shifts. Subsequently, we validate the model results and examine the robustness of our approach. Finally, we summarize the key findings in Section 6, followed by a thorough discussion of the limitations and implications of this research.

2 Literature review

This section aims to provide a comprehensive overview of previous research and findings that are relevant to the research questions we are addressing. Previous studies have validated the impacts of various factors associated with the pandemic on consumer behavior, both tangible and intangible [41]. It has been observed that different consumer groups respond differently to consumption patterns during crisis events, often based on their individual value judgments [41]. While online platforms and non-contact services have experienced significant growth during the pandemic [51, 52], physical stores are still considered valuable due to their ability to provide an immersive shopping experience [53, 54]. Building on earlier studies that primarily focused on static comparisons at an aggregated level, our research aims to simulate the decision-making process of consumers faced with multiple store options.

Undoubtedly, consumers have embraced new behavioral patterns in light of the severe circumstances and economic upheavals triggered by the COVID-19 pandemic, backed by studies dating back as early as 2020. For example, Hall et al. (2020) spotted consumption behavior changes by region caused by the pandemic, where residents picked up a stronger preference for storage and hesitated to purchase hotel services [55]. In particular, these behaviors during the pandemic were shaped by two pivotal factors: strict preventive measures and self-protective mindsets marked by perceived risks encompassing both health and economic concerns [2, 41, 56]. In a study conducted by Sheth in 2020, it was noted that the pandemic caused significant disruptions to people's lives and work patterns, leading to changes in their behavior. The prevailing uncertainty led to panic buying and hoarding of essential items. Additionally, the physical constraints imposed by daily responsibilities discouraged people from engaging in consumer activities and made such activities less practical. However, the reduced consumption during the pandemic was expected to be followed by a subsequent increase in consumer spending. Furthermore, the rapid development of e-commerce technologies during the pandemic, offered consumers a wider range of choices. This phenomenon was predicted to have a lasting impact on consumers' preferences even after the pandemic subsided [32].

Various customer groups exhibited diverse behavioral changes. Some began frequenting larger stores due to their ability to facilitate social distancing, while, others conversely shifted their shopping preferences away from larger retailers, opting instead for smaller stores, local retail centers, and out-of-town shopping alternatives [57–59]. Eger et al. (2021) observed that consumers from various generations are likely to exhibit distinct changes in their consumption patterns when confronted with significant societal events [41]. This observation stems from the understanding that consumers' personal experiences and values, including their shopping motivations, play a role in shaping their consumption habits [60]. This insight motivated us to delve deeper into the analysis of heterogeneity in behavioral shifts among different consumer groups.

Simultaneously, there was a substantial increase in customers' preference for delivery and pick-up services [51, 61]. However, the most notable transformation in shopping behavior occurred through a widespread transition to e-commerce platforms, particularly during periods of lockdown restrictions. This shift significantly accelerated the already thriving growth of the e-retail sector [16, 52]. Several empirical studies have been conducted to demonstrate the growth of e-commerce in the retail and grocery sectors during the pandemic [62–64]. It is widely acknowledged that the pandemic served as a catalyst for the advancement of online shopping platforms [65, 66]. Furthermore, numerous studies have concentrated on comparing or examining the transition between online platforms and traditional shopping activities, with a particular emphasis on various regions worldwide [16, 64, 67–72].

As the pandemic promoted the prosperity of online shopping, people's inclination to visit physical stores, potentially driven by their desire to interact with products, was reaffirmed [54]. Consequently, consumers found themselves faced with the task of choosing which stores to visit when they intended to shop under social distancing measures. Alhaimer (2022) underscored that people's concerns during the pandemic were distinct from those in regular times, placing more emphasis on risk-related factors rather than the quality of products and services [73]. This study served as inspiration for our investigation into the impact of non-product-related factors on consumers' decisions to patronize brick-and-mortar retail stores during the pandemic. As a result, we selected department stores as an illustrative example. Furthermore, the presence of similar product offerings assists in minimizing the influence of the actual products on consumer behavior.

Although numerous studies have examined changes in customer behavior regarding the selection of physical retail store locations, the existing literature predominantly focuses on macro-level analysis. Related studies primarily utilize unidimensional analytics to explore collective behavioral patterns. Such studies employ longitudinal analysis to provide valuable insights into the changes in customer visit and spending patterns across different industries and geographic locations. Through diverse perspectives, these studies highlight several key findings, for example, the decline in in-store foot traffic, and increased customer in-store dwelling times [11, 74–76]. According to Rossetti (2022), customers tended to prioritize shopping locations with more effective social distancing regulations, even if it meant longer wait times. This highlights a decreasing emphasis on distance as a factor in consumers' decision-making process when selecting a store [77]. Additionally, other research has highlighted the significance of store uniqueness and shopping costs as influential factors in customer decision-making when choosing a shopping location, particularly in certain industries during the pandemic [72].

A series of studies conducted research using various regression techniques and choice models to examine different aspects of customer decision-making during the COVID-19 pandemic, considering restrictions and safety concerns. These studies aimed to determine the impact of such measures on people's travel patterns and their preferences for urban amenities. In a study by Rose et al. (2023), it was found that the implementation of lockdown measures had noticeable effects on the spatiotemporal patterns of customer visits [16]. In another study, Sevtsuk et al. (2021) employed a logistic choice model to analyze how visitors responded behaviorally to different attributes of amenity clusters (e.g., cluster area, number and diversity of establishments) during various phases of the pandemic. By comparing coefficient ratios of independent variables, the researchers sought to understand the trade-offs between cluster attributes within the context of customer preferences. The results of their analysis revealed that customers altered their travel patterns to amenity clusters and demonstrated a higher level of selectivity in their choice of destinations during the first six months of the pandemic. Furthermore, the importance of travel distance to stores diminished, with a preference emerging for larger essential stores. Additionally, neighborhood characteristics, such as the presence of parks and open spaces, played a role in shaping travel behavior during the pandemic. Areas with accessible outdoor amenities were more likely to generate trips [78]. This study holds significance as it simultaneously examines multiple factors linked to customers' preferences for selecting a shopping location. However, the chosen methodology falls short in capturing the magnitude of temporal shifts and the individual importance levels of these factors.

While previous research emphasizes the importance of using quantitative methods to measure and understand the impact of the pandemic on consumer behavior, there are few studies that specifically examine the decision-making processes of consumers when faced with multiple store options. To the best of our knowledge, this is also true about studies that have simulated the decision process from a computational standpoint, bringing multiple factors into account simultaneously, particularly when consumers have to choose among physical stores for shopping. Although some studies have employed psychological methodologies [79, 80], there is a need for new approaches to quantify changes in customer preferences when it comes to selecting a physical stores for shopping. In light of this, our research aims to propose a new method for measuring temporal shifts in customer preferences during the store selection process.

3 Theoretical background

Spatial marketing has a rich history, with its initial application dating back to the late 1920s and early 1930s, and it remains relevant today. Within spatial marketing, three fundamental concepts hold particular significance: attraction, gravitation, and spatial interaction [81]. The combination of these concepts gives rise to two prevalent model types: deterministic models [82, 83] and probabilistic models [84]. Probabilistic models offer a more flexible approach to understanding the impact of retail attraction, examining the likelihood of a customer choosing one store over another. Among these models, the original Huff gravity model is widely renowned for its simplicity and broad applicability, cementing its enduring influence in the field [50].

3.1 Huff gravity model

The Huff gravity model [50] is arguably the most famous probabilistic model in the spatial marketing domain. This model and its variants have been widely used and applied in many empirical studies for estimating market share in a competitive environment, mainly with the aim of determining optimal locations for new facilities or assessing a retail store's performance [85–88]. The original model proposed by David L. Huff (1964) estimates the probability of customer i choosing store j or another based on the utility value that is proportional to the store's attractiveness (A_j) and inversely proportional to customer-store distance (D_{ij}). Equation 1 formally defines the utility function. The parameters α and β are used to adjust the sensitivity of the model to the two factors. Equation 2 formalizes the probability that a customer i visits the store j, where J denotes the set of stores that a customer i could potentially visit.

$$U_{ij} = \frac{A_j^{\alpha}}{D_{ij}^{\beta}} \tag{1}$$

$$P_{ij} = \frac{U_{ij}}{\sum_{j' \in J} U_{ij'}} \tag{2}$$

The original Huff model primarily relies on the floor area of a facility to determine its attractiveness, while the Euclidean distance between a customer base and a store location serves as an approximation of customer-store proximity. However, numerous alternative factors have been suggested in the literature to gauge facility attractiveness from various perspectives. These factors encompass diverse aspects, such as store characteristics like product pricing and parking availability [89], demographic and purchase information including average customer income and brand loyalty [90], and more recently, neighborhood characteristics such as the abundance and variety of amenities surrounding a store [91]. These factors have been proposed as viable indicators of attractiveness for the respective stores, expanding the scope beyond the original measures utilized in the Huff model.

The consideration of attractions in the surrounding area of a store is founded on the principle that customers exhibit a preference for multipurpose trips over single-purpose trips [92–94]. Consequently, the quantity and variety of amenities near a store serve as an indicator of the various options available for customers to engage in multipurpose activities. The presence of a diverse range of amenities facilitates different combinations of choices for customers, enhancing the appeal of the store. Additionally, the distribution of business categories in a region can significantly influence customer behavior, and an imbalanced distribution may impact patronage patterns [95]. By incorporating these factors, a more comprehensive understanding of transient customer flows can be achieved, enabling a more accurate representation of the attractiveness of the store's location.

One major extension of the Huff model was proposed by Nakanishi and Cooper (1974). They proposed a multiplicative competitive interaction (MCI) model that replaces the store area with a product of multiple attractiveness factors, assigning tuning parameters for each factor in the product separately. Equation 3 formalizes the MCI model, where H is the set of store attractiveness factors h.

$$P_{ij} = \frac{\frac{\prod_{h \in H} A_{hj}^{\alpha_h}}{D_{ij}^{\beta}}}{\sum_{j' \in J} \frac{\prod_{h \in H} A_{hj'}^{\alpha_h}}{D_{ij'}^{\beta}}} \tag{3}$$

Another expansion to the Huff model was introduced by Dolega et al. (2016) [96]. Their objective was to extend the model's application from individual stores to encompass entire retail agglomerations. In their approach, they aimed to quantify the attractiveness of a destination within a specific town center (A_j). To determine this attractiveness, they proposed a linear combination of various factors, including the size of the center measured by the total number of outlets (S_j), the diversity of retail establishments within the center (RM_j), the proportion of leisure outlets (L_j), the proportion of the most appealing stores (An_j), and the number of vacant outlets (V_j). The inclusion of the number of vacant outlets aimed to account for the negative impact of empty stores on the perceived attractiveness of the retail agglomeration [97]. Equation 4 formalizes their proposed method for estimating the attractiveness of a destination.

$$A_j = (S_j - V_j) + RM_j + L_j + An_j \tag{4}$$

Further developments in the understanding of retail agglomerations and clustering have been pursued by researchers aiming to provide a comprehensive framework for classifying shopping and consumption spaces, along with a systematic analysis of the modern retail and consumer service landscape [98]. In a recent study conducted by Ballantyne et al. (2022), the authors employed the Hierarchical density-based spatial clustering of applications with noise (HDBSCAN) technique to define the boundaries of urban retail centers. This approach takes into account various characteristics and functions of retail centers, specifically focusing on the Chicago metropolitan statistical area [99].

Numerous extensions and adaptations of the Huff model have been put forth, employing diverse functions to capture the interaction of attractiveness measures, customer-store distance proxies (e.g., Haversine distance, road network distance, travel time, and mode), and distance decay functions (e.g., exponential, weighted dimensions) across various contexts [95, 100]. Nevertheless, in our study, we adhere to the MCI model originally proposed by Nakanishi and Cooper (1974).

3.2 Calibrating model parameters

The final step in preparing the model for application involves fitting or calibrating the parameters. Several techniques have been proposed and employed for parameter tuning, including Ordinary Least Squares (OLS) [101, 102], Generalized Linear Model (GLM) [103], Geographically Weighted Regression (GWR) [104], and the more recent Particle Swarm Optimization (PSO) [105].

The OLS approach aims to transform the non-linear model into a linear one by applying logarithmic transformation. The parameters are then optimized using OLS regression. However, this approach has limitations. It cannot handle zero observations, is sensitive to outliers, and may yield biased coefficients in the presence of heteroskedasticity. As an alternative, the linear regression is replaced by a GLM with a Poisson or Negative Binomial distribution, and parameters are fitted by maximizing the log-likelihood function. Due to the lack of a closed analytical solution for this maximization, a standard optimization technique such as gradient descent is employed for parameter tuning [103].

GWR aims to account for spatial non-stationarity in the model parameters. While local models can sometimes provide a better fit compared to global models, the focus on location in the GWR approach can lead to underrating other customer preferences, particularly in densely populated areas like NYC, where stores are located in close proximity to each other and consumers' homes. Additionally, GWR may produce less accurate and highly biased results when there are insufficient local observations.

PSO is a continuous nonlinear optimization technique inspired by the foraging behavior of bird flocks [106] and the swarming theory, which suggests that a flock of birds searching for food can benefit from the experiences of others. In PSO, a group of particles is randomly generated within a search space to explore the optimal position. PSO does not require derivative information for optimization, and it imposes few or no assumptions about the problem, making it suitable for finding solutions in a vast space of candidate solutions [107, 108]. These characteristics make PSO a problem-independent algorithm that only requires knowledge of the solution search space boundaries and the fitness evaluation metric to compare candidate solutions. As a result, it is commonly employed for model selection in machine learning [109] and more recently for parameter calibration in the Huff model [105]. The selection of an appropriate method for parameter calibration plays a crucial role in enhancing the accuracy and performance of the model, ensuring that the results accurately reflect reality. Based on these considerations, our study utilizes a modified version of the MCI model with PSO for parameter calibration, as described in the subsequent section.

4 Empirical strategy and data

In this section, we first introduce our proposed modified model for quantifying the temporal shifts in customer preferences for choosing a shopping location that builds on the Huff gravity model. We then describe the datasets utilized in this study and finally detail the analytical setting.

4.1 Model modifications

The primary aim of this research is to examine the changing patterns of customer preferences when selecting a shopping destination, comparing the pre-pandemic and post-pandemic periods. While the Huff gravity model and its variations were originally developed from a retailer's standpoint for trade area analysis and optimizing facility locations, we have adapted this model to measure the the change of customer preferences in choosing a store location for shopping. By employing this modified version, we can quantify the temporal shifts in these preferences and specifically evaluate the impact of the COVID-19 pandemic as a significant intervention in this context.

Since the MCI model allows taking multiple attractiveness measures into account simultaneously, we use it as the base model in this study. The first modification we make is normalizing all attractiveness factors into an interval of $[1, \Phi]$ using min-max normalization. This operation will map the minimum value of each attractiveness factor to 1, its maximum to Φ, and all other values are proportionally transformed into values within the normalization interval. Φ is tuned to optimize the model fit and can vary according to the distribution of attractiveness factors in different study cases. Details of the procedure for defining the optimal value for Φ are described in section 4.3.4. There are several reasons for implementing this normalization approach:

1. Eliminating the problem of 0 observations: In a multiplicative model, having a 0 value for any attractiveness factor can result in a numerator of 0, regardless of the magnitudes of other factors. For instance, if there are no other amenities and attractions near a particular store, a zero value can render the total attractiveness of the store as zero. Normalization addresses this issue by avoiding the presence of zeros in the calculation.

2. Enhancing optimization process speed and accuracy: By mapping all factors to the same interval, this approach improves the efficiency of the optimization process and provides better accuracy. It reduces the complexity of the objective function, especially when the existing factors span a wide range including values smaller and larger than 1. This modification preserves the increasing effect of the variable concerning an increase in the parameter value in the exponent.

3. Ensuring comparability and interpretability of parameters: The normalization approach facilitates comparability and interpretation of the parameters, which represent the exponents of the attractiveness factors. By having a common base range, the parameters become easier to compare and interpret, even when the attractiveness factors have significantly different ranges. For example, if the store area is measured in square meters or footage (with a large magnitude) and consumer loyalty is defined as a fractional number within the [0,1] interval, the parameters might become difficult or impossible to interpret without normalization. With this normalization approach, the parameters can reflect customer preference levels for each attractiveness factor accurately.

In the subsequent discussion, we will refer to the model parameters as “preference parameters” since, in the proposed modified version, they effectively reflect the customer preference levels associated with each attractiveness factor.

By utilizing this modified model, we can effectively measure the change in customer preference for each attractiveness factor by comparing the magnitude of the corresponding preference parameter across different time periods, while keeping the other attractiveness factors constant. However, it is crucial to note that interpreting the numerical values of the preference parameter changes independently may lead to misleading conclusions, as it disregards the significance of other attractiveness factors in the customer decision-making process. It is important to consider the collective impact of all attractiveness factors in order to fully understand the dynamics of customer preferences.

It is essential to acknowledge that the objective of this study is not to forecast future customer behavior and preferences, but rather to comprehend the significance of factors that have influenced their past decisions using historical records. This is in addition to the fact that extreme exogenous shocks, such as pandemics, can cause abrupt disruptions in mobility patterns. Such sudden perturbation in mobility patterns can significantly reduce the accuracy of the predictive models [110]. Therefore, all the methods we employ in our approach are aimed at achieving the best fit to the available data points. While overfitting can have negative implications for accuracy in the context of machine learning and forecasting, in this study, it helps us capture the dynamics of the customer decision-making process more effectively by quantifying their preferences regarding various decision-making factors.

4.2 Data

This study incorporates three distinct large-scale datasets at a granular level: mobility patterns, places, and census datasets.

The mobility patterns dataset is derived from the geolocation data of smartphones used by individuals who have voluntarily opted into specific applications. To safeguard user privacy, the data is anonymized and aggregated based on the users’ home census block groups (CBGs) on a weekly basis. CBGs are statistical areas defined by the United States Census Bureau and typically encompass a population ranging from 600 to 3,000 individuals. These CBGs serve as units for presenting demographic data and controlling block numbering. The dataset encompasses information on mobility patterns and visits made by users residing in CBGs inside NYC boundaries, spanning from January 2018 to December 2021.

The second dataset provides information about the places or points of interest (POIs) that the users visit. A POI commonly denotes a distinct geographic location or area that serves a specific purpose or provides particular amenities and attractions. These may include restaurants, gas stations, tourist attractions, transportation hubs, or notable landmarks. The dataset provided encompasses diverse information regarding the places, including their names, business categories, area measured in square footage, and geolocation coordinates. These details offer comprehensive insights into the characteristics and attributes of the included places.

It is important to note that both the mobility patterns and place datasets are geographically confined within the administrative boundaries of NYC. This encompasses 6,493 CBGs and 42,130 POIs. These datasets have been sourced from the SafeGraph COVID-19 Data Consortium, which is a trusted provider of accurate and curated POI and mobility data. Further information about SafeGraph and its datasets can be accessed through their website at www.safegraph.com.

In addition, we integrate census data obtained from the United States Census Bureau (www.census.gov) into our analysis. This dataset provides 5-year estimates of crucial demographic information at the level of census block groups (CBGs). It includes various demographic variables such as median household income, education level, median age, and the racial composition of residents. By utilizing this census data, we gain insights into the socioeconomic and demographic characteristics of the CBGs that are under examination. This information enhances our understanding of the contextual factors that may influence consumer behavior and preferences within these areas.

It is notable that previous research studies have highlighted the efficacy of these datasets in capturing consumer behavior from various viewpoints [78, 98, 111–113]. The richness of the data, which encompasses details on mobility patterns, visits, places, and demographic factors, offers a comprehensive perspective on consumer behaviors and preferences.

4.3 Analytical setting

In this section, we examine the specific context of retail stores in NYC, an urban center characterized by a high population (exceeding 8 million residents), advanced development, and modern infrastructure. NYC is renowned for its diverse demographic composition, encompassing residents from various backgrounds, and is home to numerous commercial brands striving to expand their market presence and gain a larger market share.

4.3.1 Data preprocessing

It is worth acknowledging that aggregating mobility and visit patterns from an individual level to a CBG level results in a loss of data granularity. However, we address this concern by implementing a preprocessing approach to ensure the representativeness of the sample utilized in our study. Through careful data preprocessing, we aim to mitigate any potential biases and maintain the integrity and reliability of our findings.

Several previous studies have extensively examined and discussed the representativeness of Safegraph mobility patterns at various levels of census statistical geographic areas, including counties, census tracts, and census block groups [7, 107, 112, 114–117]. Chang et al. (2021) determined that the aggregated trends in Safegraph mobility data align with the aggregated trends observed in Google mobility data across the entire United States [7]. In their recent comprehensive study, Li et al. (2023) extensively investigated the sampling bias within SafeGraph Patterns and analyzed it across five dimensions: spatial, temporal, urbanization, demographic, and socioeconomic [117]. Their research covered the entire United States from 2018 to 2022, providing valuable insights into the dataset's limitations. They demonstrated the dataset's reliability by establishing a strong correlation between the number of sampled devices in the SafeGraph dataset and the census population at the county level over the five-year period. The study revealed correlation coefficients exceeding 0.97 for urban counties and 0.91 for rural counties, indicating the dataset's robustness in capturing population dynamics at this level. However, the study also identified lower correlation values and higher sampling bias rates at the CBG level across the country. This suggests potential limitations in accurately representing population dynamics at a finer granularity.

Accordingly, to ensure the representativeness of our model's input data, we conducted similar analyses on the specific version of the available data for this research. We specifically examined the correlation between the tracked device counts in the mobility data panel and the population size at various geographic and socioeconomic levels, using the Pearson correlation coefficient (r). Detailed analysis results are available in Appendix A of this manuscript.

The findings of our analysis indicate that the data used in this study demonstrates a high level of representativeness at larger geographic levels beyond CBGs. However, at the CBG level, the correlation coefficient decreases to 0.463. As a result, we implemented the following data preprocessing steps to enhance the representativeness of our model's input data and ensure the generation of reliable and accurate results.

At the first step, to retain the robustness of the analyses, we filter and keep CBGs that: (i) have visit records during all four years (2018 to 2021); and (2) have demographic information available in the census dataset. From all 6,493 CBGs in the dataset, 6,078 of them have visit records over the course of the four years, but only 5,781 of them have complete demographic information available in the census dataset.

Subsequently, for each CBG, we calculate its sampling ratio and sampling bias using Equations 5 and 6, respectively. The sampling ratio of CBG_i, denoted as SR_{CBG_i} in Equation 5, is determined by the count of devices tracked in CBG_i represented by TD_{CBG_i} and the population size of CBG_i denoted by N_{CBG_i}. Additionally, Equation 6 defines the sampling bias of CBG_i as SB_{CBG_i}.

$$SR_{CBG_i} = \frac{TD_{CBG_i}}{N_{CBG_i}} \tag{5}$$

$$SB_{CBG_i} = \left(\frac{TD_{CBG_i}}{\sum_{j=1}^{5781} TD_{CBG_j}} - \frac{N_{CBG_i}}{\sum_{j=1}^{5781} N_{CBG_j}}\right) \times 100 \tag{6}$$

In the final step, we employ the Interquartile Range (IQR) method to identify and eliminate outlier Census Block Groups (CBGs) based on their sampling ratio and sampling bias values. This method establishes the normal value range as $Q1 - 1.5 \times IQR$ to $Q3 + 1.5 \times IQR$, where $Q1$ and $Q3$ denote the first and third quartile values, respectively. Additionally, IQR represents the interval length between $Q1$ and $Q3$, which can be calculated as $IQR = Q3 - Q1$. Any CBGs falling outside this normal range are considered outliers and subsequently removed from the dataset.

Following the removal of outliers using the IQR method, the resulting set of CBGs comprises more than 95% of the CBGs in NYC that had both visit records and demographic information. Specifically, out of the initial 5,781 CBGs, a total of 5,502 CBGs remain in the dataset. This high coverage indicates that the majority of CBGs, representing a significant proportion of the dataset, are retained for further analysis.

Furthermore, there is a strong correlation of 0.756 observed between the number of tracked devices in the mobility data panel and the corresponding populations of the remaining 5,502 CBGs. This correlation further reinforces the representativeness and reliability of the data after preprocessing, as it demonstrates a consistent relationship between the number of devices tracked and the population size of those CBGs.

Within the places dataset, POIs are categorized using the North American Industry Classification System (NAICS) 6-digit sector codes. For the purpose of this study, we focus on two specific categories within the retail industry: (1) General Merchandise Stores, which includes Warehouse Clubs and Supercenters; and (2) Department Stores. These two categories represent similar types of businesses and collectively consist of 282 stores located in the NYC area. To streamline the discussion, we will refer to the stores belonging to these two categories as “department stores” for simplicity.

4.3.2 Model building

By leveraging a multiplicative model as the foundation of our proposed approach, we are able to incorporate multiple attractiveness factors into the model concurrently. Taking into consideration the attractiveness factors identified in the existing literature (refer to Section 3) and the possibility of their extraction from our datasets, we have included a set of 6 variables in our analysis model.

1. Customer-Store distance (D_{ij}) The customer-store distance is a measure of the relative convenience of accessing a specific store from the customer’s location [50]. The attractiveness of a store to a customer decreases as the distance between the customer’s location and the store increases. We calculate the Haversine distance in miles between the store location and the centroid of CBG_i. The Haversine distance, which is the angular distance between two points on the Earth’s surface, is considered a more realistic measure of store-CBG distance compared to the Euclidean distance.

2. Store area (S_j) The incorporation of store size or floor area in the model is grounded on the principle that larger stores provide a wider and more comprehensive assortment of grocery products. The increased variety of choices and enhanced availability of products tends to be more enticing to consumers [50]. This factor gained additional significance during the COVID-19 pandemic as it could enable social distancing measures and accommodate a higher number of customers within the limitations of physical distancing requirements. The floor area for each store is available in the places dataset, and is provided in units of square feet.

3. Diversity of POI categories in a store's vicinity (E_j) As underscored by economic models, the abundance and variety of amenities play a crucial role as attractiveness factors in drawing people to a specific geographical region [118–120]. Previous studies have demonstrated the effectiveness of incorporating these factors in extensions of the Huff model. They can serve as proxies for area attractiveness and also contribute to attracting individuals for multipurpose trips. Following Shannon's entropy formula (Shannon, 1948), we calculate the diversity value based on the business categories of the POIs within the CBG where store j is located. In statistical physics, Shannon entropy quantifies the number of possible arrangements or states within a system. When applied to urban amenities, it quantifies the number of different ways these amenities can be combined. As a result, it serves as a metric of the variety of experiences that customers can obtain from this diverse mix of amenities.

4. Count of POIs in a store's vicinity (C_j) This factor serves as a measure of the abundance and availability of amenities within a geographic area (i.e., CBG). As described in item 3, this measure, along with the diversity of amenities, can serve as a proxy for the attractiveness of the area to a variety of customers. For each store j, we utilize the count of POIs within the CBG where the store is situated.

5. Store-Customer CBG demographic similarity (M_{ij}) This attractiveness factor quantifies the socioeconomic similarity between customers and the residents of the neighborhood where the store is situated. The inclusion of this factor in the Huff model has been grounded on the assumption that, given equal attractiveness and accessibility factors, customers are more inclined to select a store located in a neighborhood with socioeconomic characteristics similar to their own, compared to other alternatives [91]. To calculate this measure, we construct a seven-dimensional vector representing the demographic attributes of individuals residing in each CBG. These attributes include education level (percentage of residents with a college degree or higher), income level (median household income), age (median age), and the racial composition of residents (percentage of White, African-American, Asian, and Hispanic/Latino individuals). Next, we calculate the cosine similarity between the demographic vectors of each customer CBG (V_i) and store CBG (V_j) pair.

6. Customer chain loyalty (L_{ij}) The incorporation of this factor into the Huff model is founded on the notion that customer loyalty towards a particular chain or store can significantly influence their likelihood of selecting it for their future shopping visits [90, 91]. We define the customer chain loyalty as the proportion of total visits made by customers from CBG_i to all stores within a specific chain or the brand that owns store j during a given time period.

Equation 3 serves as the foundation for presenting the closed-form expression of the utility function employed in our study case, which is represented by Equation 7.

$$U_{ij} = \frac{S_j^{\delta} \times E_j^{\eta} \times C_j^{\zeta} \times M_{ij}^{\theta} \times L_{ij}^{\xi}}{D_{ij}^{\beta}} \tag{7}$$

It's important to note that in our analysis of the two variables, namely: diversity of POI categories and the count of POIs near a store, we define the vicinity of a POI as the CBG where it is located. Although other measures like circular buffers or travel time and cost cutoffs are used in the literature to define a POI's vicinity, given the small size of CBGs in the United States, we find no significant differences between the results from these different methods. The results strongly correlate, with correlation coefficients around 0.8 depending on the buffer size. We chose the CBG as the geographic vicinity mainly because other key factors in our analysis, such as census attributes, are defined at the CBG level. This approach ensures consistency in geographic boundaries across our variables without significantly changing the results.

4.3.3 Preference parameter calibration

As introduced in Section 3.2, four different methods have been proposed and utilized for preference parameter calibration in the literature, namely: OLS, GLM, GWR, and PSO. Previous research using OLS, GLM, and GWR has generated mixed results in various settings and applications [103, 104, 121]. However, recent studies have shown that the PSO technique outperforms the other three methods, with the calibrated parameters using PSO producing better fit to the input data in the Huff model [91, 105, 107]. Therefore, in this study, we employ the PSO technique for preference parameter calibration.

To determine the optimal parameter values, we choose Pearson's correlation between the estimated visit distribution calculated by the modified version of the Huff model and the actual visit distribution of a CBG as the objective to be maximized. The calibration process aims to minimize the cost, which is defined as one minus the correlation coefficient between the estimated and actual visits for each CBG in each considered year.

4.3.4 Normalization interval and PSO search space boundaries

The final step of this approach involves defining the upper end point of the normalization interval (Φ) and the limits of the search space for the PSO algorithm (L_{PSO} and U_{PSO}). The objective in this step is to maximize the model fit by adjusting these variables.

It is worth noting that changing the value of Φ can influence the set of optimal preference parameters and shift candidate solutions in a six-dimensional space, where six corresponds to the number of preference parameters in the proposed model. Additionally, due to the metaheuristic nature of PSO, there is no guarantee of converging to the global optima [108]. Hence, setting the boundaries of the search space $[L_{PSO}, U_{PSO}]$ is crucial to increase the likelihood of finding improved candidate solutions within our approach.

To maintain the increasing effect of the preference parameters on the utility value, we set the lower bound of the search space to 1 (i.e., $L_{PSO} = 1$), ensuring that the parameters have a positive impact on the utility. The upper bound of the search space (U_{PSO}) is then determined to find the optimal combination with Φ that maximizes the model fit. This approach aims to enhance the efficiency of the PSO algorithm by reducing the size of the search space while maximizing the model fit.

At this stage, we encounter two main challenges. Firstly, due to the multistep design of our approach, it is not possible to define a closed-form mathematical expression that can map Φ and U_{PSO} to the output cost of the PSO algorithm that should be minimized. Thus, gradient-based optimization methods are not applicable. Secondly, based on the large-scale design of our study case - a six-dimensional PSO should run once for each CBG with visit pattern records in that year for each of the four years of study - exhaustive search or brute-force approaches are inefficient and computationally expensive.

To overcome these challenges, we employ a simulated annealing algorithm (SA) [122]. SA is a metaheuristic optimization algorithm for approximating global optima in a discrete search space. This algorithm does not make any assumptions about the objective function and is problem-independent. These properties makes it suitable for our needs as (1) the variables Φ and U_{PSO} are predefined to take integer values in a discrete search space, and (2) there is no functional form that can relate these two variables (i.e., Φ and U_{PSO}) with the PSO output cost values.

We assign the median of the PSO cost values computed for the CBGs in our study as the state cost, denoted as CS_n. The objective of the SA algorithm is to minimize this state cost, which we represent as CS_{min}. We then proceed to follow the steps outlined in Algorithm 16.

The algorithm begins by initializing the temperature of the initial state as T and setting a decay rate $0 < k < 1$. We then assign random integer values to Φ and U_{PSO}, representing the initial state as $S_o = (\Phi_o, U_{PSO_o})$. Next, we compute the corresponding state cost as CS_o. We initialize the minimum initial cost as $CS_{min} = CS_o$ and proceed with the algorithm.

Each state S_o in the algorithm has four neighbors, denoted as $S_{o1}, S_{o2}, S_{o3}, S_{o4}$, which are obtained by incrementing or decrementing the corresponding state's Φ or U_{PSO} value by one unit. The algorithm continues iterating until two conditions are met: (1) T reaches a small enough value where the transition probability P_T approaches zero. This is determined by comparing the value of T with a predefined small positive number ($\epsilon > 0$) at each iteration, and (2) the cost of the current iteration state is lower than the costs of all its neighboring states.

Algorithm 1 Using SA to find the optimal combination of Φ and U_{PSO}

---

while $(T > \epsilon)$or$(CS_o \neq min(CS_o, CS_{o1}, CS_{o2}, CS_{o3}, CS_{o4}))$ **do**
  $S_n \leftarrow$ **a random neighbor of the current state** S_o
  $CS_n \leftarrow$ **cost of the state** S_n
  if $CS_n < CS_o$ **then**
    $S_o \leftarrow S_n$
    $T \leftarrow k \times T$
    $CS_{min} \leftarrow CS_n$
  else
    $P_T \leftarrow e^{(\frac{CS_o - CS_n}{T})}$
    $P \leftarrow$ **a random number in** $[0, 1]$
    if $P_T \leq P$ **then**
      $S_o \leftarrow S_n$
      $T \leftarrow k \times T$
    end if
  end if
end while

During each iteration of the algorithm, we maintain a record of the SA trajectory, which includes the states traversed and their respective costs. This allows us to compare the cost of the terminated state with the minimum cost observed throughout the iterations. Additionally, keeping a record of the states and their corresponding cost values reduces computational time in subsequent runs, as there is no need to recompute the cost for states that were previously traversed.

4.4 Overview of the proposed method

Figure 1 depicts a flowchart illustrating the schema of our approach, offering an overview of the inputs and analysis steps involved in calibrating the optimal set of preference parameters.

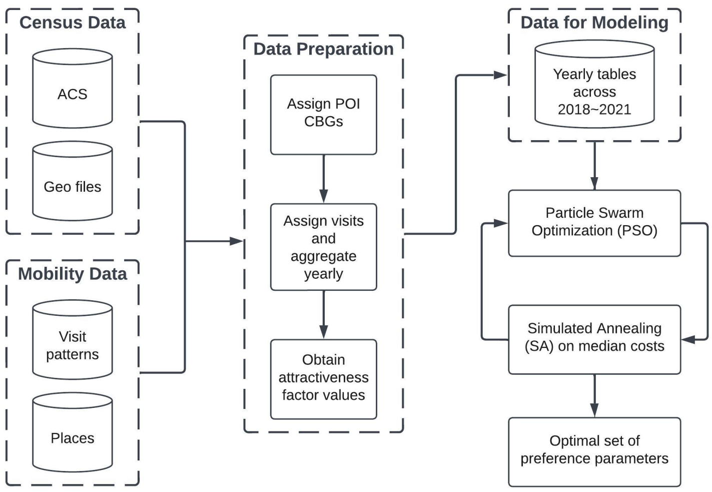


Fig. 1: **Preference parameter calibration process flowchart.**

5 Results

This section begins by presenting and discussing the results of the conducted case study, followed by the validation and assessment of the robustness of the derived analytics. The case study focused on retail stores, specifically General Merchandise Stores and Department Stores, in NYC. To obtain the results, we fitted the model parameters at a CBG level on an annual basis from 2018 to 2021, resulting in six preference parameters for each CBG-year pair, corresponding to the six attractiveness factors incorporated into our model.

The fitted model demonstrated significant correlations at the chain level for the years 2018, 2019, 2020, and 2021, with values of 0.964, 0.956, 0.742, and 0.544, respectively. These correlation values were computed by comparing the actual yearly visits to chains with the estimated visits obtained using the fitted parameters produced by PSO at its optimal point.

While the model initially exhibited a strong alignment with the data from 2018 and 2019, its performance diminishes in subsequent years. This decline can be attributed to governmental interventions that imposed limitations on consumer choices as well as self-protective mindsets marked by perceived health risks that impacted customer decision process. Nevertheless, when evaluated on an annual basis, the performance of the proposed model in this paper remains reasonably comparable, if not superior, to that of the original model, particularly when considering essential retailers in NYC [107, 123].

It is worth noting that although the stores analyzed in this study were temporarily closed due to COVID-19 prevention measures, the impact on the model's performance is likely limited as the model considers visit proportions rather than actual visit numbers. However, it is possible that an extreme decline in visit counts by residents within a CBG could affect the model's performance for the corresponding CBGs. To mitigate this issue, as explained in section 4.3.1, we addressed the problem by including only CBGs that exhibited visits to such stores throughout the entire study timeframe.

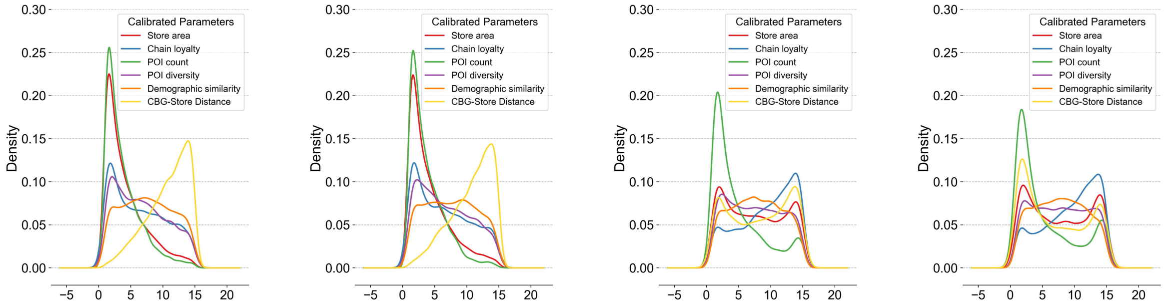
The resulting preference parameters and their distributions were utilized to investigate the dynamics of customer preferences during the four-year period. For each year, we used the preference parameter set that produced the highest correlation between the model estimated visits and actual visits from the data. The obtained result indicate that while there are no significant changes in the preferences of customers in the store selection from 2018 to 2019, with the outbreak of the COVID-19 pandemic, the parameters of the customer's preference change significantly; therefore, in what follows, we focus primarily on understanding the effects of the pandemic on such preference shifts.

5.1 The general effect of the COVID-19 pandemic on customer preference

Figure 2 represents the yearly distributions of the calibrated preference parameter values from 2018 to 2021. Values do not change significantly from 2018 to 2019, while substantial changes are observed from 2019 to 2020 as well as smaller-scale shifts from 2020 to 2021. For instance, chain loyalty apparently weighed more from 2019 to 2020, but its calibrated-value distribution in 2021 is similar to the one in 2020. On the contrary, while the importance of CBG-store distance decreased from 2019 to 2020, a similar trend was also observed between 2020 and 2021.

To validate the observations illustrated in Figure 2, we performed a Kolmogorov-Smirnov test [124] to compare the distributions of preference parameter values across consecutive years in the 5,502 CBGs where visits were captured consistently over the four-year period. The resulting p-values can be found in Table 1. Under a two-sided 95% confidence level, it is evident that there are no significant changes in any of the parameter values between 2018 and 2019. However, with the onset of the COVID-19 pandemic in 2020, all parameters, except for the preference for demographic similarity, exhibit significant changes. Yet, in 2021, despite the commencement of the vaccination program, complete reopenings, and the gradual relaxation of various restrictions as medical concerns diminished, parameter value distributions bear more resemblance to those in 2020, rather than returning to their pre-pandemic levels.

The comparison of the aforementioned findings with the temporal trends and global impact of the COVID-19 pandemic, validates the accuracy of our model in capturing shifts in customer preferences when selecting department stores. It also highlights the significant influence of the pandemic on altering the selection preferences of New Yorkers when choosing a retailer. In essence, during special occasions, customers prioritize certain factors while forsaking others.



(a) 2018 Distributions (b) 2019 Distributions (c) 2020 Distributions (d) 2021 Distributions

Fig. 2: **Yearly distribution of calibrated parameter values.** Each sub figure depicts the distribution of preference parameters for each of the six attractiveness factors in the corresponding year.

Table 1: Kolmogorov–Smirnov test p-values in consecutive years

Variable	2018∼2019	2019∼2020	2019∼2021	2020∼2021
Store area	0.91	9.407e-295***	7.222e-293***	9.835e-03**
Chain loyalty	0.851	5.294e-200***	2.990e-207***	0.312
POI count	0.439	5.150e-37***	5.652e-87***	3.965e-13***
POI diversity	0.574	4.311e-19***	1.240e-30***	2.140e-02*
Demographic similarity	0.734	2.020e-02*	0.063	0.946
CBG-Store Distance	0.622	5.138e-160***	0.000e+00***	4.195e-49***

Note: This table provides the Kolmogorov–Smirnov test p-values for the distribution of each preference parameter in a pair of years. Statistical significance is denoted by: $*p < 0.05, **p < 0.01, ***p < 0.001$.

Figure 3 illustrates the percentage change in mean for each preference parameter from 2019 to 2020, 2019 to 2021, and 2020 to 2021. Notably, while the mean values of the demographic similarity parameter exhibit slight percentage changes in each year pair, customers demonstrated a greater emphasis on store area, chain loyalty, POI count, and POI diversity during the COVID-19 shock in 2020.

To elaborate, during the pandemic, customers displayed a preference for retail locations with larger floor areas, potentially motivated by a desire to reduce the risk of infection and maintain social distancing measures. Additionally, customers exhibited increased loyalty towards the chain they selected for their purchases, making fewer exploratory visits. This behavior can be attributed to heightened search costs resulting from elevated health risks [125].

Simultaneously, customers showed a preference for stores located in areas with additional amenities and a more diverse range of businesses. This preference was potentially aimed to minimize the number of trips required and satisfy their needs through multipurpose visits. Interestingly, customers became less sensitive to the distance factor.

However, the changes in customer preferences towards the 6 parameters from 2020 to 2021 did not exhibit any indications of reverting to their pre-pandemic distribution in 2019. This phenomenon revealed that the impact of the pandemic on store selection preferences towards department stores in NYC is likely to exhibit sustained behavior and may not fully revert to pre-pandemic levels in the short term. In other words, New Yorkers had modified their consumption habits influenced by the persistent effect created by the pandemic.

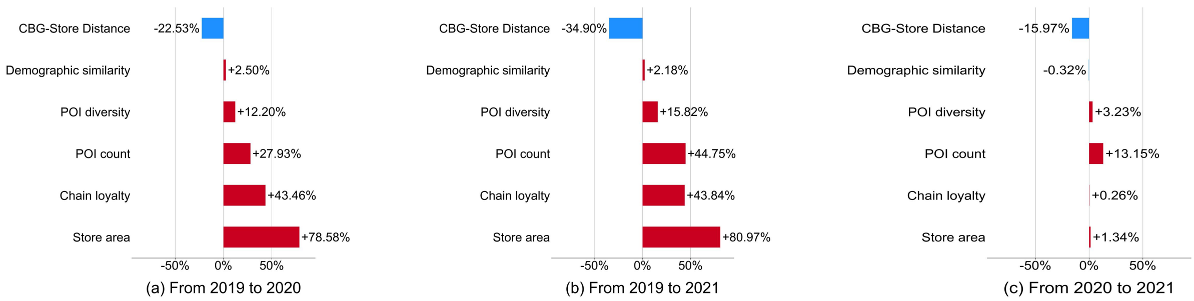


Fig. 3: **Percentage change in preference parameters' values at the population level from 2019 to 2021.** This figure shows the percentage change in the mean value of preference parameters between two different years. For example, in sub-figure (a) the mean value of CBG-store distance parameter in 2020 is 22.53% less than the same parameter value in 2019.

5.2 Heterogeneity among socioeconomic and demographic groups

The existing body of research has shed light on the unequal impact of the pandemic on various socioeconomic groups and the diversity of their response to COVID-19 containment measures [7, 13, 25–29]. This emphasizes the need to explore the heterogeneity in how different communities have made shopping location choices amidst and after the pandemic.

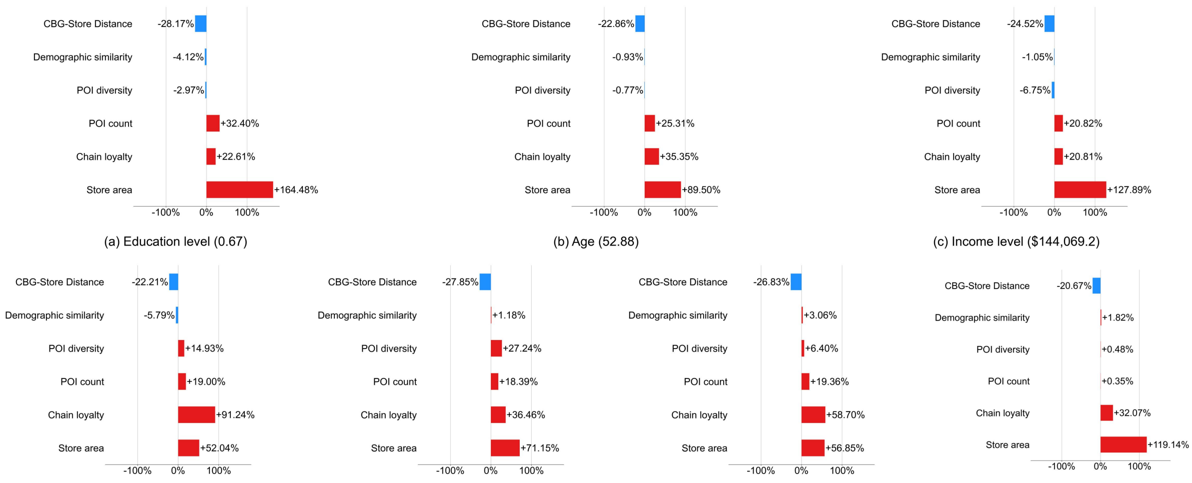
In this study, we focus on examining changes in preference parameter values from 2019 to 2020, as well as 2020 to 2021, specifically in communities where a single socioeconomic and demographic attribute ranks in the top 5% among all CBGs with records in NYC. The socioeconomic attributes of interest include education level (percentage of residents holding a college degree or higher), income level (median household income), age (median age), and racial composition of residents (percentage of White, African-American, Asian, and Hispanic/Latino residents). To facilitate the subsequent analysis, Table 2 presents the preference parameter values for the overall NYC population and the CBG groups ranking in the top 5% for each socioeconomic attribute in 2019.

Table 2: **Preference parameter values for the population and top 5% socioeconomic communities in 2019.**

Parameter	Population	Age	Education	Income	White	Black	Asian	Hispanic/Latino
Store area	4.179	3.644	2.674	3.141	3.436	5.119	3.915	5.164
Chain loyalty	6.435	6.489	8.247	8.127	6.678	7.058	5.027	5.839
POI count	3.779	3.925	4.077	4.14	4.2	3.712	4.26	3.898
POI diversity	6.725	7.473	7.574	7.894	7.524	5.48	7.301	6.578
Demographic similarity	7.882	8.34	8.516	8.261	8.087	7.573	8.142	7.879
CBG-Store Distance	10.651	10.453	10.65	10.412	10.23	10.783	10.373	11.134

Note: The values provided in this table show the average values of the preference parameters for the six attractiveness factors at the NYC population level, as well as the seven CBG groups that rank in the top 5% of each single socioeconomic attribute as of 2019.

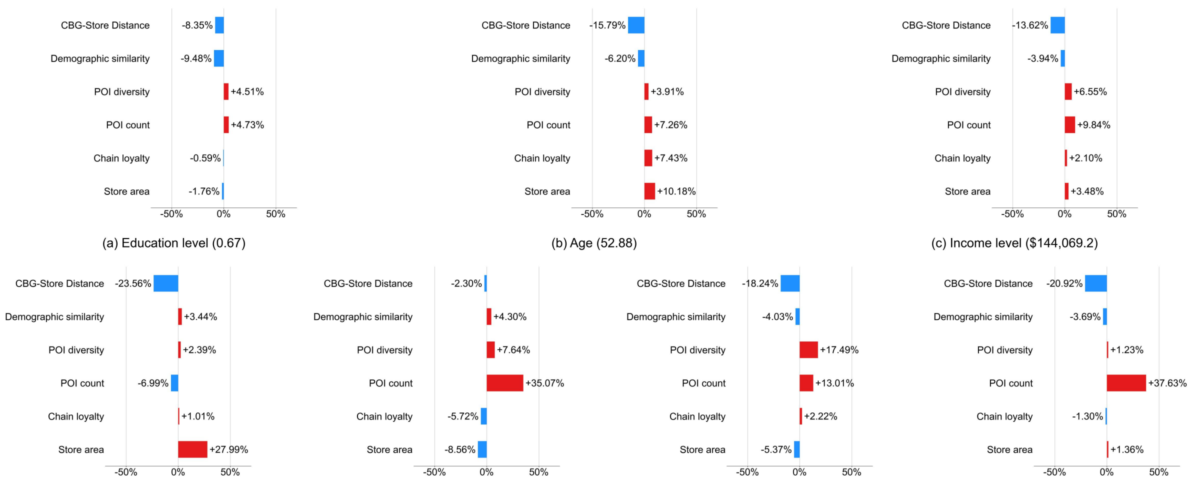
Figures 4 and 5 display the percentage change in parameter values among the CBG groups ranking in the top 5% for each individual socioeconomic attribute in consecutive years. According to Figure 4, a noticeable trend is observed across all seven CBG groups: the importance of CBG-store distance decreases, while the significance of store area increases during the pandemic. Among the six communities, except for the CBGs with White residents' proportions ranking in the top 5%, the count of POIs becomes increasingly critical. The perception of demographic similarity remains relatively consistent across all groups, while higher chain loyalty stands out among Asian residents, followed by Hispanics and residents in older age groups. This diversity in the change of preferences towards chain loyalty among different groups may be attributed to variations in their perceived search costs.



(a) Education level (0.67) (b) Age (52.88) (c) Income level ($144,069.2)

(d) Asian (0.55) (e) African-American (0.89) (f) Hispanic/Latino (0.78) (g) White (0.92)

Fig. 4: **Percentage change in preference parameters' values of Top 5% Socioeconomic Communities from 2019 to 2020.** In this figure, each sub-figure presents how the value of each preference parameter changed from 2019 to 2020 among the top 5% communities based on different socioeconomic factors. The numbers in parentheses show the minimum value of each socioeconomic feature among the top 5% CBGs. For instance, the number in parentheses in sub-figure (a) indicates that in the CBGs of the top 5% education level group, 67% of the residents hold a college degree or higher.



(a) Education level (0.67) (b) Age (52.88) (c) Income level ($144,069.2)

(d) Asian (0.55) (e) African-American (0.89) (f) Hispanic/Latino (0.78) (g) White (0.92)

Fig. 5: **Percentage change in preference parameters' values of Top 5% Socioeconomic Communities from 2020 to 2021.**

Although in 2021, as shown in Figure 5, the CBG groups ranking in the top 5% for each socioeconomic characteristic demonstrated a consistent trend of decreasing attention towards CBG-store distance, their considerations of the other attractiveness factors showcased varied shifts. Consequently, it is necessary to analyze the shifts from the perspective of each individual community, and thus, in the following, we provide an explanation of the shifts observed in each community. Since there is a similarity in the trend of preference changes among the top 5% communities in terms of education level, median age, and median income, we consolidate them and describe their behavioral shifts under a unified perspective.

The top 5% CBGs in terms of age, income, and education level The trends of preference changes among these three communities exhibit a remarkable similarity, which can be attributed to notable pairwise correlations between age, income, and education level within communities. Specifically, their department store selection preferences remained largely unchanged from 2020, with a slight increase in attention towards factors such as POI count, POI diversity, chain loyalty, and store area. It can be inferred that individuals with higher age, income, and/or education levels, despite the gradual introduction of vaccination in 2021, continued to maintain their store selection preferences that were formed during the pandemic. Moreover, there was an even stronger emphasis on store area as the impact of the pandemic began to diminish within these communities.

The top 5% Asian resident proportion CBGs Within this group of CBGs, there is a continued emphasis on store area among residents, while their attention towards POI count decreased in 2021 after an initial increase from 2019 to 2020. Based on their consistent outlook towards the other five attractiveness factors, it can be inferred that Asian residents have a preference for larger department stores while potentially disregarding the surrounding amenities and commuting costs. This hypothesis suggests that Asian individuals prioritize store loyalty based on the availability of desired products or services, regardless of the store's location.

The top 5% African-American resident proportion CBGs The residents within this group of CBGs demonstrated a partial decrease in their emphasis on store area, which had previously increased significantly during the pandemic. Instead, they showed a greater attraction towards stores surrounded by a diverse range of facilities. Additionally, the distance from home factor became moderately less important for them.

The top 5% Hispanic/Latino resident proportion CBGs The observed patterns of change in preference among the residents in this group are similar to those of CBGs within the top 5% proportion of African-American residents. It suggests that the environment of department store locations appears more inviting to them, leading to similar shifts in their preferences.

The top 5% White resident proportion CBGs It is evident that the residents of CBGs in this group paid more attention to the number of POIs in their selection of department stores. This shift in focus from distance suggests that as they gradually eased out of pandemic-related restrictions, the importance of proximity became less significant to them.

Along with the findings explained above, it is deduced that while some impacts brought by COVID-19 started diminishing in the post-pandemic era, some residents strengthened their preference on a few parameters at the price of compromising on the rest of the aspects. In conclusion, this section has provided insights into the heterogeneity of shifts in store selection preferences from the perspective of different communities identified in terms of each single socioeconomic factor.

Furthermore, we have compared the differences in temporal store selection preferences between communities and the general NYC population, focusing on the direction of change from each attractiveness factor's perspective. For details about the analysis and results, please refer to Appendix B.

To get a better understanding of the direction in temporal preference shifts, we examine the change in changes of preference parameter values. For a preference parameter ω and a year t this value is denoted by $\Delta\omega_t$ and computed as shown in Equation 8.

$$\Delta\omega_t = (\omega_t - \omega_{t-1}) - (\omega_{t-1} - \omega_{t-2}) \tag{8}$$

Figures 6 and 7 show the average $\Delta\omega_t$ values for the seven top 5% distinctive communities in each single socioeconomic feature, compared with the population average $\Delta\omega_t$ values as of 2020 and 2021 separately. Since there are no significant shifts in preference parameters from 2018 to 2019, the $\Delta\omega_t$ values for 2020 are very similar to change in the parameter values from 2019 to 2020, which makes the

Figures 4 and 6 convey a similar message. Therefore, in the following, we compare the differences in temporal store selection preferences between communities and the general NYC population from the perspective of each attractiveness factor with a greater focus on values of $\Delta\omega_t$ for 2021, as shown in Figure 7.

Store area The community with the top 5% Asian resident proportion showed limited interest in shifting their focus to store area during both 2020 and 2021, indicating strong loyalty to their preferred chains throughout the pandemic and afterward. In contrast, other communities, including those with higher education levels, median incomes, and older age groups, displayed an approximately 20% increased emphasis on store area during the pandemic. This suggests the presence of homogeneous subgroups with socioeconomic features where citizens valued store area more amid the COVID-19 shock. other communities paid approximately 20% more attention to store area during the pandemic. Interestingly, in 2021, Asian residents seemed indifferent to paying more attention to store area compared to other communities, while communities with higher education levels and median incomes still showed a 20% increase in attention to store area. Additionally, CBGs with residents in the top 5% based on median age exhibited a significant 15% increase in attention to this feature, even greater than the initial increase when the COVID-19 pandemic started to spread. Conversely, African-American residents demonstrated a higher value for store area in their department store selection after the introduction of vaccination and relaxation of social restrictions.

Chain loyalty It is observed that chain loyalty among Asian residents grew faster with and even after the pandemic, while higher income CBGs were most indifferent to this factor.

POI count More educated, and African-American residents acknowledged the value of POI count in department store selection during COVID-19. Whereas, in 2021, the CBGs with a higher proportion of Asian residents, higher educated people, and those ranked higher based on the median age of their residents, favored the stores with more POIs in their proximity.

POI diversity African-American residents were leading in emphasizing the POI diversity both during the pandemic and when the society started to be freed from the crisis in 2021.

Demographic similarity Residents from higher education, and higher income CBGs were more likely to choose the stores located in neighborhoods with greater demographic similarities to them. This phenomenon enlightened us that possibly the residents' decision process toward department store selection could be assimilated by the community they belong to.

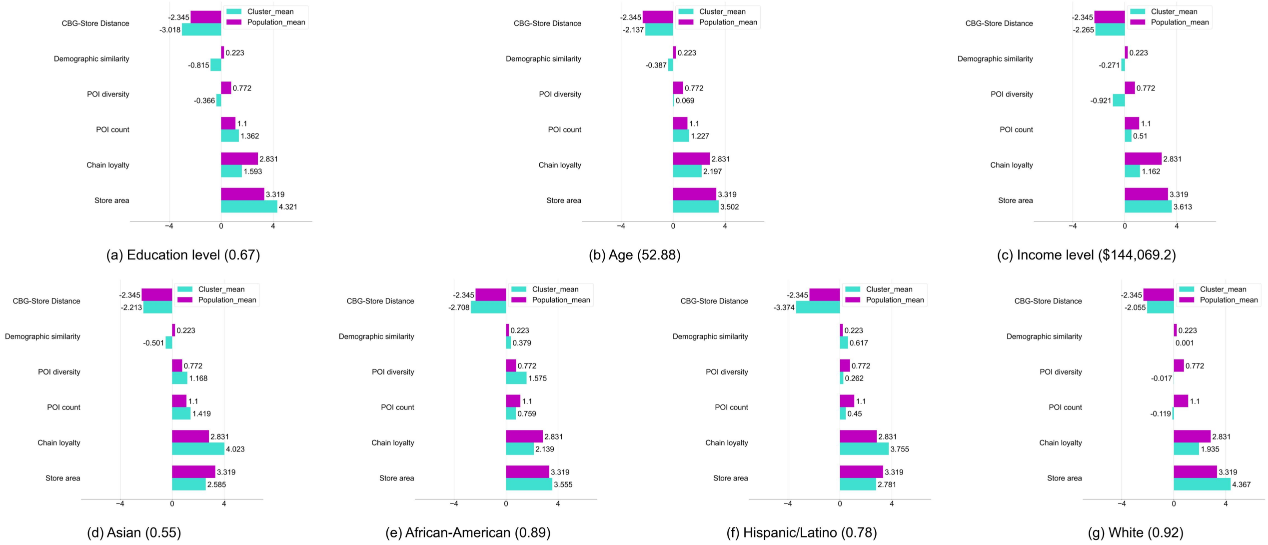
CBG-Store distance When the society was stepping into the post-pandemic era, CBGs with higher education levels, a higher African-American population and a higher age population increased their attention to this factor more than the general population.

5.3 Validation and robustness checks

In this section, our objective is to validate the findings presented in Sections 5.1 and 5.2. Firstly, we conduct a comparison between the model results and ground truth data. Additionally, we employ an unsupervised learning approach, specifically the K-means clustering method, to investigate the presence of heterogeneity in the response among clusters. This unsupervised technique allows us to explore and analyze potential variations in customer preferences without relying on predefined labels or categories. Finally, we employ a Difference-in-Differences method to estimate the impact of the pandemic on various aspects of the retail store selection preference shifts among customers.

5.3.1 Validating model results with ground truth

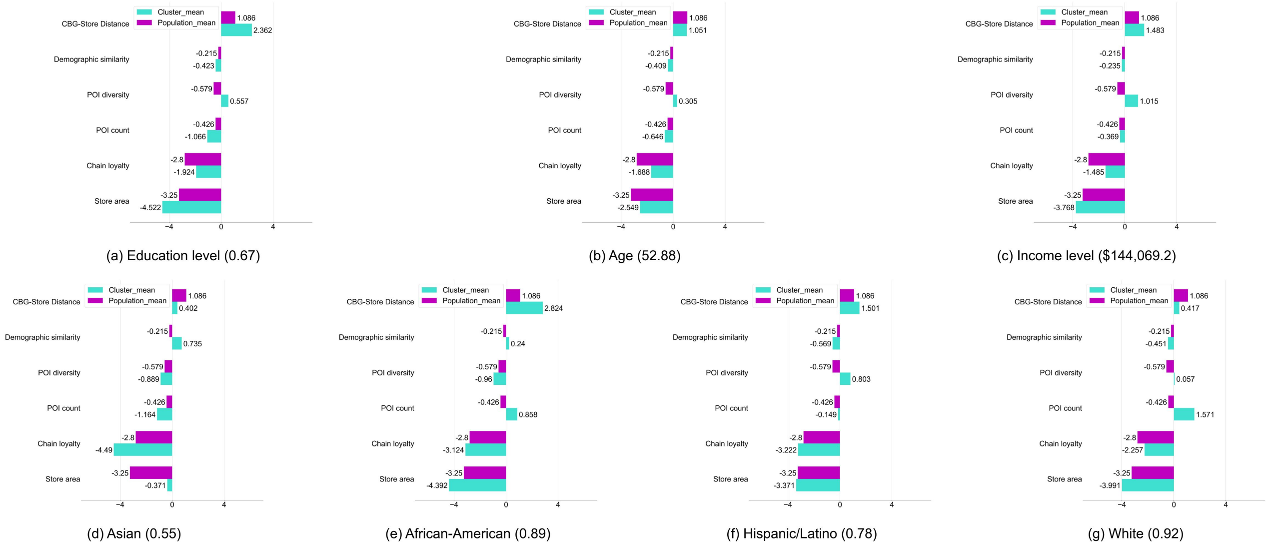
To evaluate and validate the information derived from our experimental results in Section 5.1, we conduct an analysis of store visits over different years. Utilizing the model framework that considers multiple attractiveness factors concurrently, we compare stores with similar characteristics across all six factors, except for a difference in one specific factor. This approach enables us to simplify the comparisons and facilitate easier interpretation of the results.



(a) Education level (0.67) (b) Age (52.88) (c) Income level ($144,069.2)

(d) Asian (0.55) (e) African-American (0.89) (f) Hispanic/Latino (0.78) (g) White (0.92)

Fig. 6: **2020 Mean $\Delta\omega_t$ Value Comparisons in Top 5% Socioeconomic Communities and Population.** This figure shows the average change in changes of preference parameter values ($\Delta\omega_t$) for the top 5% Socioeconomic Communities in each sub-figure. The community $\Delta\omega_t$ values are shown by blue bars and for comparison purposes the population level $\Delta\omega_t$ values are shown by purple bars. $\Delta\omega_t$ values are computed using Equation 8.



(a) Education level (0.67) (b) Age (52.88) (c) Income level ($144,069.2)

(d) Asian (0.55) (e) African-American (0.89) (f) Hispanic/Latino (0.78) (g) White (0.92)

Fig. 7: **2021 Mean $\Delta\omega_t$ Value Comparisons in Top 5% Socioeconomic Communities and Population**

Considering the basic statistics obtained from the attractiveness factors, it is observed that the store area exhibits the largest variance among the stores in our dataset. Consequently, we aim to select a pair of stores from each chain that fulfill certain criteria: they should be situated in CBGs with similar POI count and diversity, be geographically close to each other, or possess a comparable average distance to the population. By selecting stores from the same chain, we ensure that they share the same loyalty value for that chain, leaving only the store area as the variable attractiveness factor in our model. This approach allows us to isolate and analyze the impact of store area on customer preferences more effectively.

According to our results in Section 5.1, during years 2020 and 2021, customers exhibit a heightened preference for store area when compared to 2019, while controlling for the other factors. Hence, if two stores belonging to the same chain are located in CBGs with similar POI counts and diversities, the store with a larger floor area will draw a higher share of customers in 2020 and 2021 in comparison to 2019.

Table 3 displays the ground truth examples for store pairs from the same chains that validate our model results. In our dataset, the chains including Dollar Tree, Family Dollar Stores, Dollar General, T.J. Maxx, and Target own the highest numbers of stores in descending order. For each chain, we select one pair of stores with similar POI count and diversity in their vicinity (CBG), but evidently different scales of store size (i.e., floor area). Stores with a larger floor area have been shown to attract more visitors from 2019 to 2020 with the outbreak of COVID-19. When the society started to go back to the normal track from 2020 to 2021, the attraction for visitors faded but the persistence of COVID-19 impact on the importance of store area remained. In 2021, the larger stores continued to draw more customers than the smaller ones of the same chain. Table 3 shows detailed information on the factors taken into account for each store. The last three columns provide the values of yearly visit counts and their corresponding visit share percentages in parentheses for the selected stores.

Table 3: Validation of model results using ground truth examples from store pairs within the same chains

Store Brand	Store ID	POI count	POI Diversity	Area	Weighted Distance	2019 Visits	2020 Visits	2021 Visits
Dollar Tree	s6g-kmk	24	2.556827168	4,147	8.374391	5,126(53.5%)	5,865(44.7%)	1,117(21.6%)
Dollar Tree	s6f-w49	24	2.556827168	9,199	8.248515	4,448(46.5%)	7,266(55.3%)	4,047(78.4%)
Family Dollar	s94-mhq	9	2.043191871	2,257	6.870538	3,605(50.2%)	3,989(48.8%)	2,027(38.6%)
Family Dollar	s8x-3bk	10	2.025326221	14,301	7.136968	3,582(49.8%)	4,177(51.2%)	3,228(61.4%)
Dollar General	s6b-8vz	14	2.397895273	6,516	9.421959	3,606(50.8%)	3,963(40.1%)	3,254(50.6%)
Dollar General	wg9-ty9	14	2.205598359	8,080	8.226718	3,491(49.2%)	5,920(59.9%)	3,175(49.4%)
T.J. Maxx	s7m-h5z	50	2.902884061	901	7.028166	2,942(42.7%)	5,048(35.2%)	67(30.5%)
T.J. Maxx	wdh-swk	45	2.92211547	6,020	9.798352	3,956(57.3%)	9,288(64.8%)	153(69.5%)
Target	s8k-sdv	39	2.724920504	30,554	7.3432	6,750(61.7%)	5,043(36.7%)	2,542(33.8%)
Target	517E*	35	2.924102826	169,356	7.742251	4,185(38.3%)	8,712(63.3%)	4,972(66.2%)

* Note: We noticed that three POIs with different Store IDs refer to the same location. Therefore, the information of those three POIs was aggregated and assigned to a new combined Store ID of 517E. The values in Area and Weighted Distance columns are provided in square feet and miles respectively. POI Diversity is computed using the Shannon entropy formula.

Although the ground-truth examples could validate the results of the study case achieved by our proposed model, such insights may not be attainable through analyzing the unidimensional descriptive statistics. To illustrate this limitation, let us consider the example of the importance of store area in shopping location decisions. Figure 8 showcases the cumulative distribution functions (CDF) of visit share based on store area for the years 2018 to 2021. If we examine the CDFs in Figure 8 and solely focus on store area as an isolated and independent attractiveness factor, one could mistakenly conclude that customers preferred small and medium-sized stores during the pandemic. However, such conclusions drawn solely from unidimensional analytics may fail to explain the reality and can be misleading. This occurs due to oversimplification of the customer decision process to a single dimension, while real-life decisions are influenced by complex interactions among multiple preference factors within a high-dimensional space. The subset of ground truth examples provided in Table 3 demonstrates that customers' preference for store area increased in conjunction with the other attractiveness factors, challenging the notion that customers preferred small and medium-sized stores during the pandemic.

5.3.2 Heterogeneity among clusters by combined socioeconomic features

In addition to the supervised cluster analysis in Section 5.2, where we explored communities that are more identifiable considering a single socioeconomic feature, we utilize an unsupervised clustering method in order to further investigate the spatial distribution of clusters based on combined socioeconomic features and validate our findings regarding the heterogeneity of shopping location preferences in

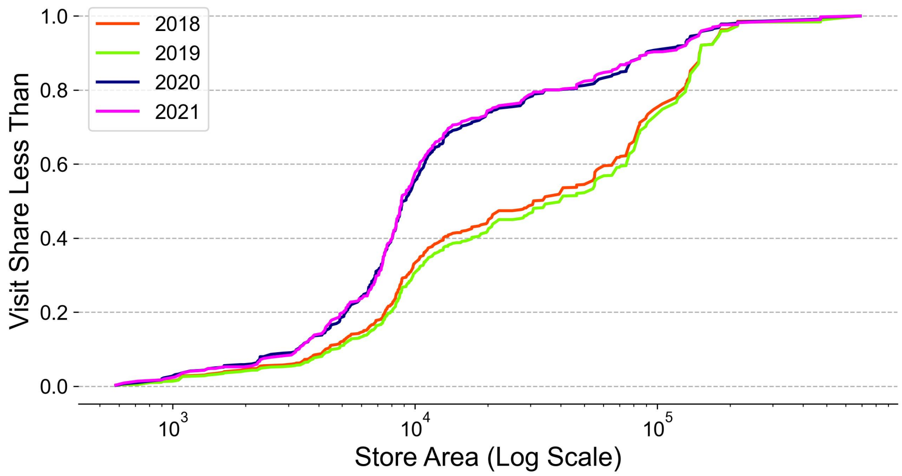


Fig. 8: Cumulative distribution of visit frequency by store area

Section 5.2. This approach allows us to assess the robustness of our previous finding and gain insights into the variations in shopping location preferences among different clusters.

To achieve this, we employed the K-means clustering method to partition NYC's CBGs into five subregions, considering all seven socioeconomic attributes. The trained K-means model yielded a between-to-total sum of squares ratio (BSS/TSS) equal to 0.63, indicating a satisfactory separation of clusters. Table 4 presents the centroids of the clusters for each socioeconomic attribute, with the percentile indicated within parentheses alongside each centroid.

Table 4: K-means cluster centers by socioeconomic factors in NYC

Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Median Age	39.889 (62.9%)	33.566 (29.8%)	37.662 (52.4%)	41.048 (68.5%)	41.356 (69.5%)
Education level	0.311 (69.6%)	0.117 (22.3%)	0.188 (43.4%)	0.226 (53.3%)	0.636 (93.4%)
Median Income	$72969 (61.9%)	$41313 (22.4%)	$60493 (46.8%)	$63187 (50.3%)	$140215 (94.6%)
White	0.72 (77.1%)	0.26 (39.0%)	0.112 (20.4%)	0.294 (42.3%)	0.795 (83.7%)
African-American	0.056 (43.6%)	0.247 (64.6%)	0.749 (88.6%)	0.052 (42.5%)	0.039 (38.2%)
Asian	0.119 (63.2%)	0.058 (48.1%)	0.035 (37.8%)	0.509 (93.8%)	0.106 (60.7%)
Hispanic/Latino	0.194 (49.9%)	0.644 (87.7%)	0.144 (40.8%)	0.188 (48.9%)	0.095 (29.0%)

Note: The values provided in parentheses show the percentile score of each attribute's centroid in the corresponding cluster.

Based on the centroids, the clusters can be interpreted as follows.

Cluster 1 Predominantly White residents with average income levels.

Cluster 2 Financially limited young residents, lowest income and education levels, with a majority being Hispanic/Latino.

Cluster 3 Less educated, middle-aged residents with average income levels, and a majority being African-American.

Cluster 4 Highest in terms of median age, mid-level education, and common-income residents with the most racial diversity, half are Asian.

Cluster 5 Most educated and highest-income residents, with a majority being White.

The spatial distributions of the five clusters, as described earlier, are visualized in Figure 9, utilizing the Geographical and Spatial Data Analysis Software (GeoDa), developed by the Center for Spatial

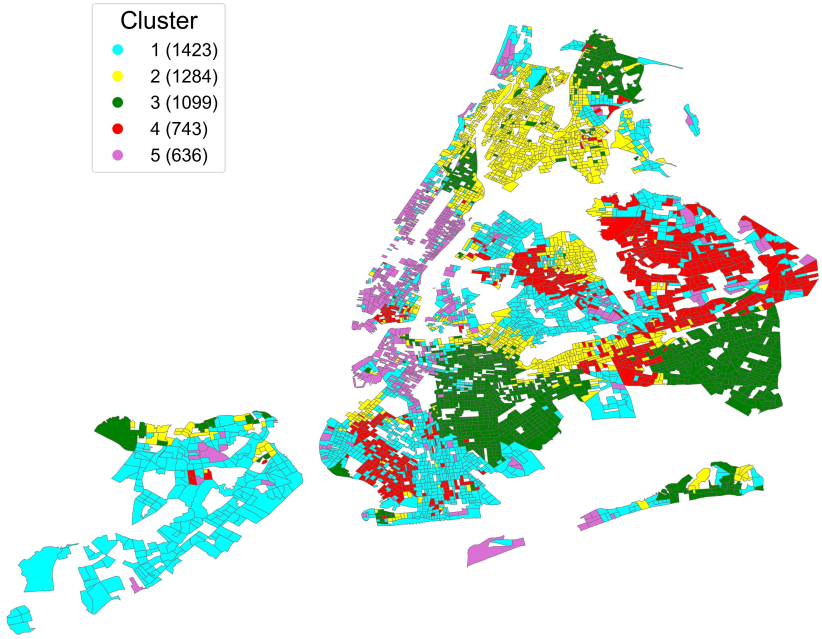


Fig. 9: **Spatial distribution of the five K-means clusters in NYC determined by combined socioeconomic attributes.** In this figure, each polygon represents a CBG and the colors indicate the cluster a CBG belongs to. The numbers inside parentheses in the figure legend, show the number of CBGs that belong to each corresponding cluster.

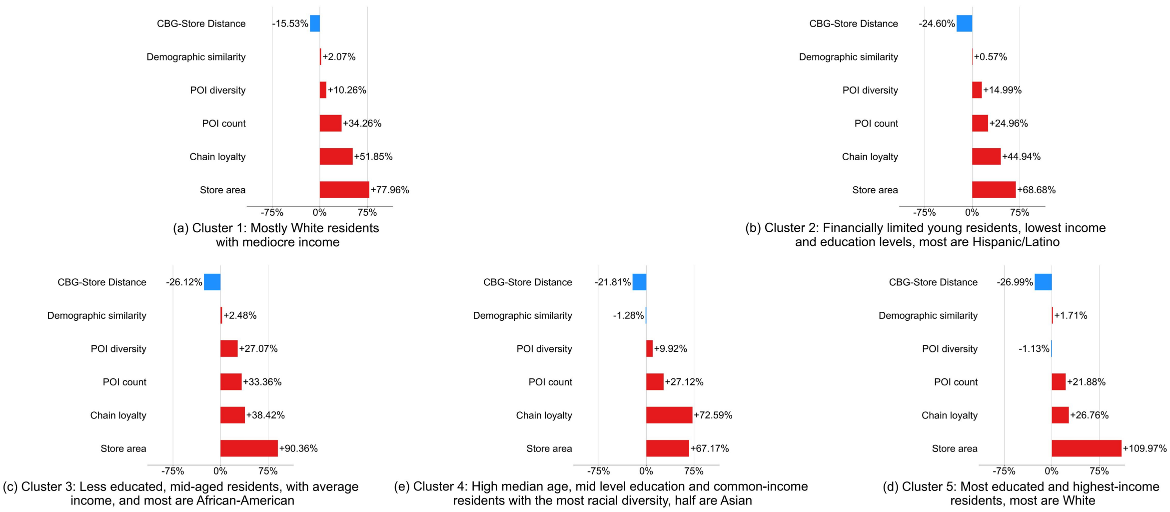
Data Science at the University of Chicago [126]. GeoDa offers a range of tools and functionalities for spatial data analysis, spatial statistics, and the exploration of spatial patterns and relationships.

The collective trends in residents' preferences across the five clusters from 2019 to 2020 and from 2020 to 2021 are illustrated in Figures 10 and 11 respectively. During the transition from 2019 to 2020, as depicted in Figure 10, residents across all clusters exhibited a decreased emphasis on distance and instead showed a preference for stores with larger floor areas, and/or with a greater abundance and variety of nearby POIs. Furthermore, residents displayed an increased level of loyalty towards the stores they selected for in-person shopping.

Figure 11 reveals that residents in Cluster 3 experienced a notable shift in their preferences between 2020 and 2021, as they significantly decreased their emphasis on the store area. In contrast, the other communities demonstrated a reverse trend, placing increased importance on the store area when visiting department stores in 2021. Additionally, all five communities showed a decreased emphasis on distance as a factor influencing their preferences. However, it is important to note that the shifts in preferences towards other attractiveness factors among the communities were relatively small and non-significant, despite the differences in directions.

For a more in-depth analysis and detailed results, please refer to Appendix B, where we compare the differences in temporal store selection preferences between clusters and the general NYC population from the perspective of each attractiveness factor, with a particular emphasis on the direction of change.

Figures 12 and 13 show changes in the value of $\Delta\omega_t$ between 2020 and 2021. It is abundantly clear that from 2020 to 2021, $\Delta\omega_t$ changes reversed roughly among each community and the entire population. People's earlier increased interest in store size, chain loyalty, POI count, and POI diversity from 2019 to 2020 began to decline from 2020 to 2019, while their indifference to CBG-store distance persisted in 2021. The magnitudes of the changes in 2021, nonetheless, are not as large as those in 2020. This occurrence confirms our earlier observation that following the outbreak, the decision patterns of



(a) Cluster 1: Mostly White residents with mediocre income

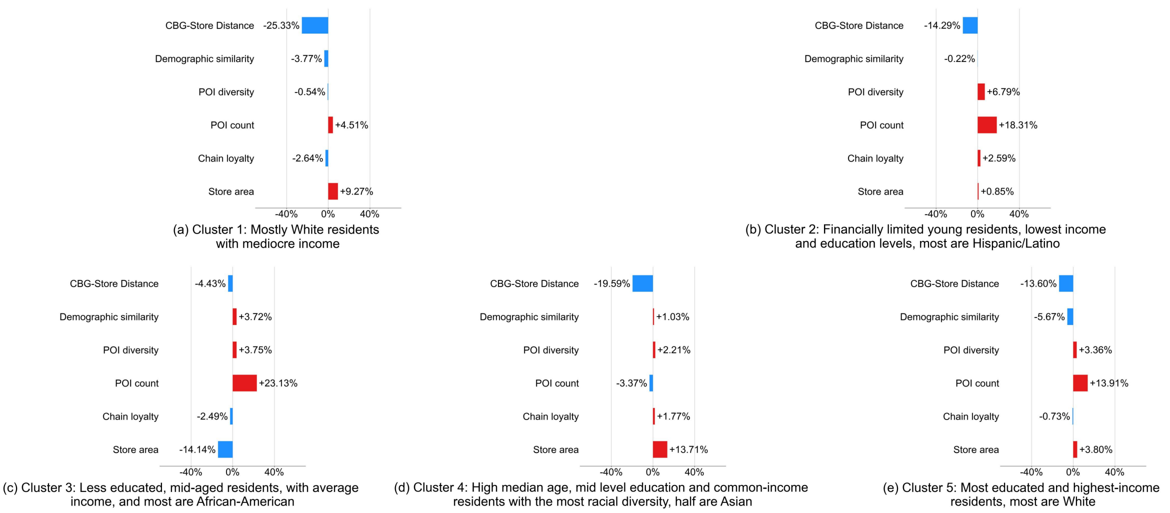
(b) Cluster 2: Financially limited young residents, lowest income and education levels, most are Hispanic/Latino

(c) Cluster 3: Less educated, mid-aged residents, with average income, and most are African-American

(e) Cluster 4: High median age, mid level education and common-income residents with the most racial diversity, half are Asian

(d) Cluster 5: Most educated and highest-income residents, most are White

Fig. 10: **Percentage change in preference parameters' values of K-means clusters from 2019 to 2020**



(a) Cluster 1: Mostly White residents with mediocre income

(b) Cluster 2: Financially limited young residents, lowest income and education levels, most are Hispanic/Latino

(c) Cluster 3: Less educated, mid-aged residents, with average income, and most are African-American

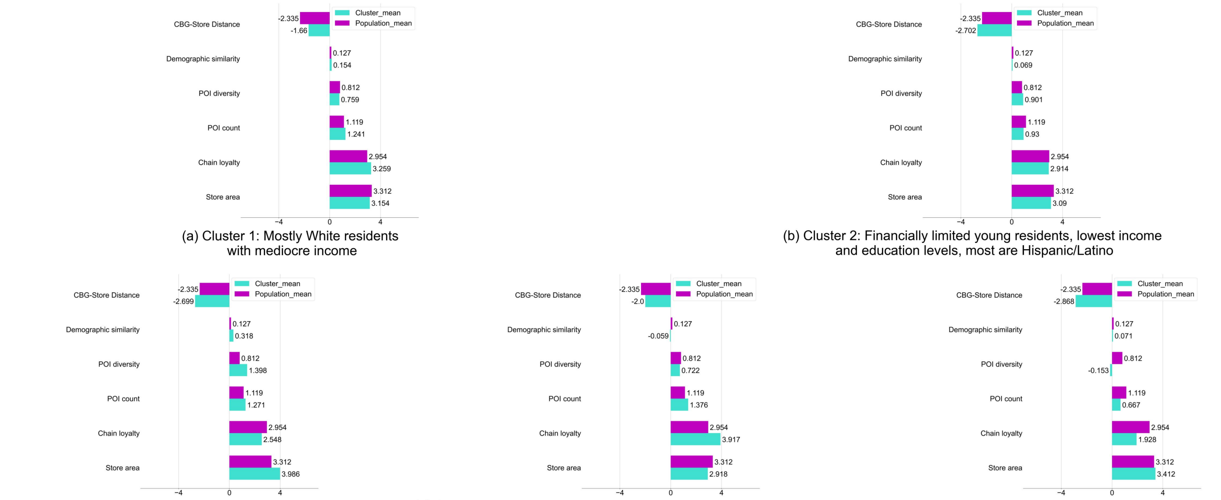
(d) Cluster 4: High median age, mid level education and common-income residents with the most racial diversity, half are Asian

(e) Cluster 5: Most educated and highest-income residents, most are White

Fig. 11: **Percentage change in preference parameters' values of K-means clusters from 2020 to 2021**

store choice of NYC residents changed from their pre-epidemic behavior and the impact of COVID-19 has not been fully recovered by 2021.

It has also been observed that changes in $\Delta\omega_t$ due to store area, chain loyalty and CBG-store distance are apparently greater than those in POI count, POI diversity and demographic similarity in 2020, while in 2021 changes in $\Delta\omega_t$ in CBG store distance are similar to those of POI count and diversity. Consequently, it has been signaled that since the outbreak of COVID-19, the products and services that a store can provide have caught more attention instead of the nearby commercial service and the cost of commuting. Overall, our findings using unsupervised clustering confirm that the pandemic response was heterogeneous among different communities.



(a) Cluster 1: Mostly White residents with mediocre income

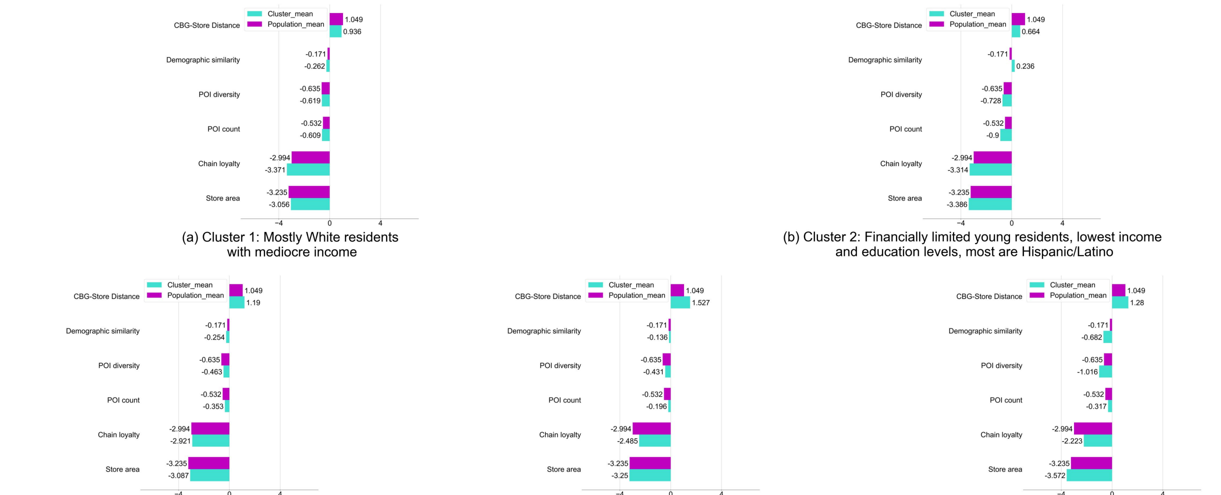
(b) Cluster 2: Financially limited young residents, lowest income and education levels, most are Hispanic/Latino

(c) Cluster 3: Less educated, mid-aged residents, with average income, and most are African-American

(d) Cluster 4: High median age, mid level education and common-income residents with the most racial diversity, half are Asian

(e) Cluster 5: Most educated and highest-income residents, most are White

Fig. 12: **Mean $\Delta\omega_t$ Value Comparisons in Clusters by K-means as of 2020**



(a) Cluster 1: Mostly White residents with mediocre income

(b) Cluster 2: Financially limited young residents, lowest income and education levels, most are Hispanic/Latino

(c) Cluster 3: Less educated, mid-aged residents, with average income, and most are African-American

(d) Cluster 4: High median age, mid level education and common-income residents with the most racial diversity, half are Asian

(e) Cluster 5: Most educated and highest-income residents, most are White

Fig. 13: **Mean $\Delta\omega_t$ Value Comparisons in Clusters by K-means as of 2021**

5.3.3 Estimating the impact of the pandemic on behavioral shifts

The rise of the COVID-19 pandemic in early 2020 as an exogenous shock, along with access to historical visits data, enables us to investigate the impact of the pandemic on changes in consumer behavior in a quasi-experimental setting. Using a Difference-in-Differences (DID) method, we compare changes in preference parameters from 2018 to 2019 (control group) with changes from 2019 to 2020 (treatment group), considering the pandemic and its subsequent interventions as treatment. For comparison, we implemented DID on the entire NYC population, and also the subgroups defined by supervised and unsupervised clustering tools.

Figure 14 displays the coefficients obtained from the DID tests. Notably, significant changes are observed in the preferences towards the store area and chain loyalty, while distance shows an opposite trend. These shifts in preference are particularly prominent among the seven communities identified based on each individual socioeconomic factor. A noteworthy observation is the significant increase in the importance of chain loyalty compared to other parameters among the top 5% groups with the highest proportion of Asian and Hispanic populations. This trend indicates a stronger preference for loyalty to specific store chains among these communities, while exhibiting less exploratory behavior.

For the general population and all communities, there was a noticeable increase in residents' focus on chain loyalty and the count of POI in department store selection during the pandemic. Specifically, among the top 5% of the White population, there was a stronger inclination towards chain loyalty, with less emphasis on the surrounding POI counts and diversity. Moreover, the DID analysis results suggest that the location and physical characteristics of the store itself had a greater impact on attracting customers than the attractiveness of its nearby commercial or leisure districts.

Furthermore, the analysis of the DID coefficients reveals that all groups examined in the comparison maintained a consistent preference for demographic similarity. However, there were variations in the significance levels among communities when it came to the POI aspects. Finally, regardless of the communities, it is evident that the overall ranking of the parameters among the significant coefficients are as follows: store area, chain loyalty, POI features, and CBG-store distance. These findings highlight the importance of store size, and the surrounding POI environment in reshaping customers' preferences during the COVID-19 pandemic.

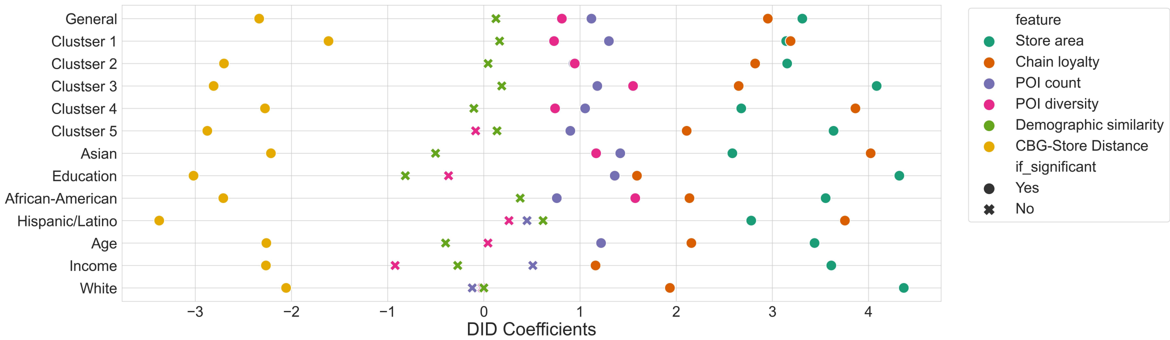


Fig. 14: **DID analysis results for preference parameter value changes.** In this figure, the x axis shows the range of DID coefficients obtained from the comparison between the preference parameter value changes from 2018 to 2019 and from 2019 to 2020. The y axis, lists different communities that we specifically investigated the impact of COVID-19, including the whole population of NYC residents, the 5 clusters identified by k-means clustering, and the top 5% society in terms of each single socioeconomic factor. As the legend box shows, the significant coefficients at the two-sided 5% level are presented as dots, while the insignificant ones are visualized by the $\times$ marks.

6 Discussion and conclusion

6.1 Summary and theoretical implications

The COVID-19 pandemic, and the subsequent restrictions imposed by authorities in order to contain the virus, resulted in a large-scale and significant shift in customer decision making process and shopping habits. While a large portion of customer demand has been satisfied through the online sales channel, the importance of the physical space for in-person shopping did not diminish, especially in the retail sector. Retail stores, and in particular department stores, performed better in maintaining their customer foot-traffic volume among other similar businesses during the pandemic, emphasizing the importance of their facilities. However, given the change in customer preferences towards in-person shopping location selection, the pre-pandemic operational strategies may no longer be as effective in the post-pandemic era. Therefore, for retail managers and marketers, in order to survive in a highly competitive environment, it is of crucial importance to be able to measure and evaluate such changes in customer preferences.

In response to this need, our study introduces a new approach to measure the changing customer decision factors in selecting department stores for in-person shopping over time. We employ a modified version of the Huff gravity model, incorporating its MCI extension, which allows for the simultaneous consideration of multiple decision-making factors. By leveraging this modified model, our proposed method effectively captures the dynamic nature of customer preferences for shopping locations. It achieves this by analyzing the intricate interplay between factors such as customer socioeconomic status, risk perception, store attributes, and the surrounding environment.

In our study, we focus on examining department stores in New York City (NYC) using extensive datasets on mobility, places, and census information. To analyze the data effectively and derive reliable insights, we utilize mixed computation and optimization techniques. Considering the available data, we incorporate six key factors of attractiveness in our analysis. These factors include the distance between customers and stores, the size of the stores, the demographic similarity between customers and residents of the CBG where the store is located, customer loyalty to specific chains, and the count and diversity of points of interest (POIs) in the vicinity of the stores.

The findings of our study reveal that the residents of NYC have made adjustments in their department store selection choices in response to the COVID-19 pandemic. These behavioral changes are established to be associated with the pandemic through a quasi-experimental setting. Furthermore, following the pandemic, various governmental non-pharmaceutical and pharmaceutical (e.g., vaccination) interventions, continue to influence these transformations. It is important to note that the post-pandemic behavioral adjustments aim to synchronize people's purchasing patterns with the habits they developed during the crisis, rather than reverting back to their pre-pandemic behaviors.

In 2021, as society gradually returned to its pre-pandemic state, New Yorkers exhibited minor changes in their department store choices across various aspects, except for the distance between the stores and customers' locations, which remained relatively consistent compared to 2020. Interestingly, despite the easing of official restrictions, customers' emphasis on store distance continued to decline. Instead, their attention shifted towards factors such as store size, chain loyalty, the number of POIs, and the variety of POIs in the store's vicinity, which were more influenced by the pandemic. These trends were not isolated incidents but rather reflected general patterns observed at the population level. As the pandemic spread globally, the trained distance parameter underwent a shift towards lower values in the parameter distribution. Prior to the pandemic, the distribution of distance parameter values exhibited a right-skewed pattern in 2018 and 2019. However, starting in 2021, it began to shift towards the left. This implies that during the pandemic, customers were more concerned with stores being able to meet their demands rather than their geographical proximity to their location.

We further examined the heterogeneity in behavioral and preference shifts among different communities. Initially, we analyzed the behavioral characteristics of CBGs that ranked in the top 5% for each individual socioeconomic attribute. The distinct dissimilarities observed among these top 5% groups prompted us to employ an unsupervised clustering technique, specifically K-means clustering, to identify different clusters within NYC using all the socioeconomic features combined.

Regardless of whether they belonged to the top 5% single demographic clusters or the clusters identified by K-means, consumers across these clusters exhibited a similar shift in their preferences from 2019 to 2020. They placed a higher value on store size, as well as one or more factors among chain loyalty, POI count, and POI diversity, while their emphasis on customer-store distance decreased. This trend aligns with the observed broader trend at the population level.

As society began to recover from the outbreak in 2021, the occupants of these clusters continued to devalue the importance of distance when selecting department stores. However, their preferences for other aspects experienced a variety of changes. It is worth noting that the magnitudes of these preference changes from 2020 to 2021 were notably smaller compared to the changes observed from 2019 to 2020, with the exception of the top 5% community consisting of individuals with the highest proportion of white population, whose preferences regarding POI abundance exhibited a more substantial shift.

Motivated by the aforementioned findings, we employed the Difference-in-Differences (DID) technique to estimate the impact of the pandemic on the temporal shifts in customer preferences across the entire population and various subgroups. Through our analysis, we confirmed that chain loyalty exhibited the strongest positive preference shift among Asian community. Moreover, the outlook towards each parameter varied among different communities to different degrees, except for demographic similarity. Nevertheless, when evaluating the seven demographic groups separately, we consistently observed the patterns above.

Our DID analysis further revealed that the most substantial changes induced by the pandemic, across different clusters, population subgroups, and the overall population of NYC, were evident in customers' preferences for store area and chain loyalty. Remarkably, the shift in preference for customer-store distance stood out as a noteworthy finding. While being significant and robust, it displayed a reversed trend compared to the other parameters examined.

6.2 Managerial implications

Our analytical results demonstrate that the unprecedented global public health crisis significantly reshaped consumer behavior. However, as the crisis gradually subsides, customers have begun to reassess their reactions. Nonetheless, they did not completely abandon all their prior behaviors when new consumption patterns emerged. Consequently, due to the enduring nature of customer preference shifts, retailers should not anticipate a complete return to pre-pandemic customer behavior.

The findings of this study have important implications for retail managers and marketers in adapting to the post-pandemic retail landscape. The analysis of customer preference shifts provides valuable insights that can guide decision-making and strategic planning. We propose the following as key managerial implications to be considered:

Strategic adaptation for retailers Quantitative methods, like the one proposed in this paper, enable retailers to update their understanding of customer behavior and adapt their operational and marketing strategies accordingly. By identifying the directions and magnitudes of customer preference shifts, managers can align their approaches with the evolving market conditions.

Optimize resource allocation With changes in customer preferences, retailers should carefully evaluate their resource allocation. This includes assessing the allocation of physical space and marketing budgets. For instance, with the increased significance of store size as well as the abundance and richness of attractions in the vicinity, retailers may need to allocate resources to expand their store space or carefully evaluate the variety and abundance of the nearby amenities when selecting the location for a new store. This strategic consideration can help retailers better cater to changing customer preferences and enhance their competitive advantage.

Enhance customer satisfaction By gaining insights into the factors that influence customer decision-making when it comes to choosing a shopping location, retailers can prioritize strategies to enhance customer satisfaction. Given that customers have reduced their exploratory behavior and exhibit increased store loyalty, managers can capitalize on this opportunity to gain a deeper understanding of their frequenting customer base. This understanding enables retailers to adapt to customer preferences, improve the store environment, optimize product assortment, and provide tailored services that meet the evolving needs of their loyal customers. By focusing on customer satisfaction, retailers can foster stronger relationships and cultivate customer loyalty.

Monitor and analyze customer behavior The shifts in customer preferences observed in this study emphasize the need for continuous monitoring and analysis of customer behavior. Retailers should invest in data collection and analysis tools to gain insights into customer preferences and behaviors. By tracking temporal changes and staying updated on customer preferences, managers can make timely adjustments to their strategies and stay ahead in a competitive market.

Embrace an omnichannel approach This study emphasizes the significance of considering multiple sales channels for retailers who may not have the flexibility to modify their store's physical space (e.g., floor area) or location. Retailers should explore and leverage various sales channels, including online platforms, to cater to changing customer preferences. Adopting an omnichannel approach can provide customers with more convenience and flexibility, ultimately leading to a higher level of customer satisfaction and loyalty.

In summary, our study results underscore the importance of understanding the changing consumer trends to enable businesses to effectively adapt their strategies and offerings to meet evolving consumer needs in the post-pandemic era.

6.3 Limitations, strengths, and future directions

This study represents the first attempt, to the best of our knowledge, to quantify the temporal shifts in customer preferences when selecting a retail location for shopping. The findings provide a potential avenue for future research to further investigate and quantify these preference shifts in customer retail location decisions. Our proposed approach in utilizing a modified and enhanced version of the MCI variation of the Huff gravity model, combined with quantitative methods and optimization techniques, maximizes the model's fit to the large-scale historical records of customer mobility and visit patterns.

However, it is essential to acknowledge that due to the limited availability of data, we could only incorporate six attractiveness factors out of the various factors introduced in the literature, into our study. While the model still yields more informative results compared to unidimensional statistics, it presents limitations in capturing the complete picture. Future research could explore alternative proxies for store attractiveness that have the potential to enhance the model's performance. Additionally, investigating different approaches for determining the optimal parameters of the model could be a worthwhile pursuit. Determining the most suitable factors to incorporate and identifying readily available ones present opportunities for further investigation and future research. Moreover, the possibility of replacing the current proxies for store attractiveness, based on the count and diversity of POIs in the store's vicinity, with more comprehensive factors derived from assessing retail agglomerations, as proposed by Dolega et al. (2016), warrants further investigation.

While the mobility dataset utilized in this study demonstrates a level of representativeness, it is important to acknowledge that the proposed model may face limitations in accurately predicting outcomes in geographies that are under-represented in the dataset, particularly in rural areas. Moreover, the reduction in observations stemming from the scarcity of data in these under-represented geographies, coupled with the closure of businesses as a result of COVID-19 prevention measures, can have a detrimental effect on the model's performance.

Additionally, although mobility models can accurately predict human mobility patterns during normal times, their predictive performance declines during exogenous shocks and crises that significantly disrupt the daily lives of citizens. This is particularly relevant in the case of the pandemic, where both the disaster itself and the implemented interventions restrict customer decision-making. Therefore, our experimental design primarily focused on maximizing the model's fit to past data to understand customer decisions in the past, rather than utilizing the model to predict future behavior.

Furthermore, our study is limited by the absence of data on spending and transactions across various sales channels in our dataset. This limitation impedes our ability to establish any association between the utilization of different purchase channels and store location choice. Consequently, future research should prioritize investigating the relationship between the utilization of diverse sales channels and the decision-making process regarding store location. Moreover, an important area for exploration would be the analysis of the heterogeneity in utilization of different purchase channels among different socioeconomic groups, along with an examination of the underlying dynamics that drive these patterns.

Availability of data and code

The code and pre-processed data used for the analyses in this study are publicly available in the project repository for complete replication and/or further future analysis purposes. The repository is available at:
https://osf.io/phbqz/?view_only=0ec27466b42b4d398e959b5d5dec4a71

Acknowledgements

The authors are grateful to SafeGraph for making their data available for this research.

Declaration of competing interest

The authors declared that they have no conflicts of interest in this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Authors' contributions

Y.X. and M.B. were involved in idea generation, design and implementation of experiments, drafting the article, writing, and creating the project repository. Whereas A.P. was involved in idea generation, results, discussion, feedback, and final revision and drafting of the article.

Funding information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A Mobility data representativeness assessment

Table A1 displays the Pearson correlation coefficients (r) between the tracked device counts in the mobility data panel and the population size across various geographic, socioeconomic, and demographic levels. The table also provides additional information, such as the number of observations or groups available for each aggregation. Furthermore, it includes the sum of absolute sampling bias (sum(AB)), median absolute sampling bias (median(AB)), median sampling bias (median(B)), and standard deviation of sampling bias values (sd(B)) for each aggregation, offering further insights into the representativeness of data considering the magnitude and variability of the sampling bias.

The final row of Table A1 presents information about the dataset utilized in the case study of this research. It confirms that the data is reasonably representative, as indicated by the strong correlation between the tracked device counts and the population size in addition to the low level of variation in bias. This suggests that the dataset is reliable and capable of producing meaningful insights about the population under investigation.

Table A1: Basic statistics of mobility data sampling ratio and bias at different geographic, and socioeconomic levels

Aggregation	Observations	r	sum(AB)	median(AB)	median(B)	sd(B)
Race	7	1.000	2.926	0.1833	-0.0793	0.6486
Household income	16	0.971	7.35	0.4298	-0.1607	0.5627
Educational attainment	24	0.998	5.524	0.1379	-0.0559	0.4112
State	29	0.994	11.297	0.1778	-0.0421	0.6506
County	1,176	0.976	18.108	0.0042	-0.0019	0.0580
CBG	61,437	0.463	45.192	0.0004	-0.0002	0.0025
CBGs preprocessed	5,502	0.756	24.291	0.0035	-0.0029	0.0046

Note: All required data and code for attaining the results shown in this table are available in the repository of this research project: https://osf.io/phbqz/?view_only=0ec27466b42b4d398e959b5d5dec4a71

Appendix B Analysis of change in direction of preference shifts

B.0.1 Top 5% socioeconomic communities

In order to gain deeper insights into the direction of temporal preference shifts, we analyze the change in changes of preference parameter values. This change is represented by $\Delta\omega_t$, where ω denotes a preference parameter and t represents a specific year. The calculation of $\Delta\omega_t$ is illustrated in Equation B1, which provides a quantitative measure for assessing the magnitude and direction of preference changes over a two years period.

$$\Delta\omega_t = (\omega_t - \omega_{t-1}) - (\omega_{t-1} - \omega_{t-2}) \tag{B1}$$

Figures B1 and B2 illustrate the average values of $\Delta\omega_t$ for the top 5% distinctive communities in each of seven single socioeconomic feature, in comparison to the population average. Details regarding the definitions of these distinctive communities can be found in Section 5.2. $\Delta\omega_t$ values as of 2020 and 2021 separately. Since there are no significant shifts in preference parameters from 2018 to 2019, the $\Delta\omega_t$ values for 2020 are very similar to change in the parameter values from 2019 to 2020, which makes the Figure 4 in the main text and Figure B1 here, convey a similar message. Therefore, in the following, we compare the differences in temporal store selection preferences between communities and the general NYC population from the perspective of each attractiveness factor with a greater focus on values of $\Delta\omega_t$ for 2021, as shown in Figure B2.

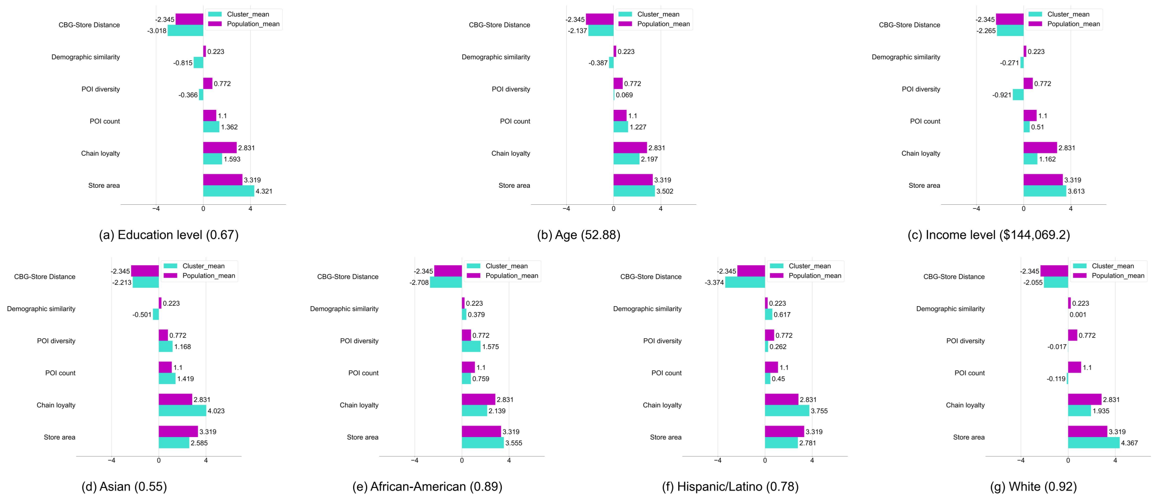


Fig. B1: **Mean $\Delta\omega_t$ value comparisons in Top 5% socioeconomic communities and population as of 2020.** This figure shows the average change in changes of preference parameter values ($\Delta\omega_t$) for the top 5% Socioeconomic Communities in each sub-figure. The community $\Delta\omega_t$ values are shown by blue bars and for comparison purposes the population level $\Delta\omega_t$ values are shown by purple bars. $\Delta\omega_t$ values are computed using Equation B1.

Store area The community with the top 5% Asian resident proportion showed limited interest in shifting their focus to store area during both 2020 and 2021, indicating strong loyalty to their preferred chains throughout the pandemic and afterward. In contrast, other communities, including those with higher education levels, median incomes, and older age groups, displayed an approximately 20% increased emphasis on store area during the pandemic. This suggests the presence of homogeneous subgroups with socioeconomic features where

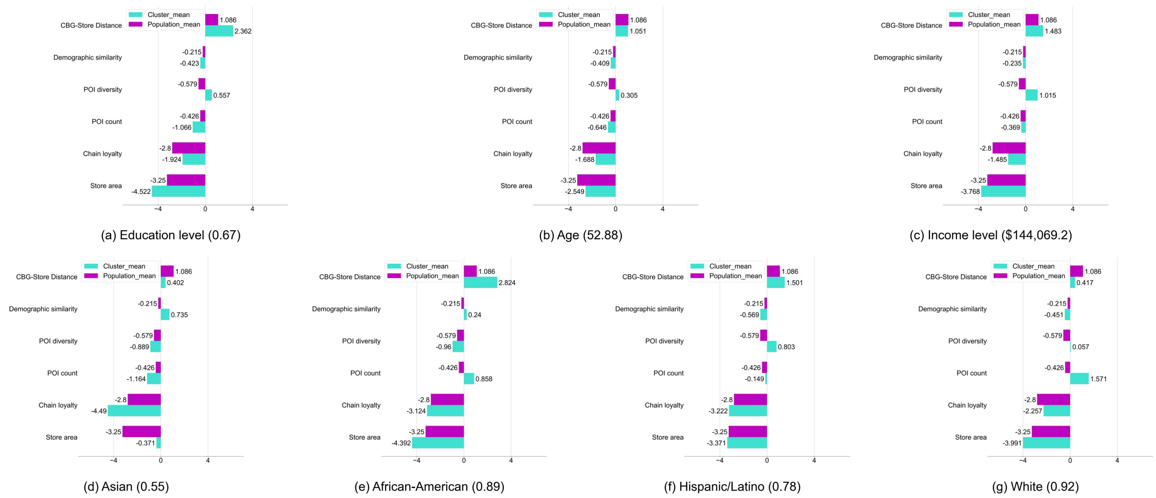


Fig. B2: **Mean $\Delta\omega_t$ value comparisons in Top 5% socioeconomic communities and population as of 2021.**

citizens valued store area more amid the COVID-19 shock. other communities paid approximately 20% more attention to store area during the pandemic. Interestingly, in 2021, Asian residents seemed indifferent to paying more attention to store area compared to other communities, while communities with higher education levels and median incomes still showed a 20% increase in attention to store area. Additionally, CBGs with residents in the top 5% based on median age exhibited a significant 15% increase in attention to this feature, even greater than the initial increase when the COVID-19 pandemic started to spread. Conversely, African-American residents demonstrated a higher value for store area in their department store selection after the introduction of vaccination and relaxation of social restrictions.

Chain loyalty The analysis reveals a notable trend among Asian residents, as their preference for chain loyalty experienced a significant increase during and after the pandemic. On the other hand, higher income CBGs displayed a relatively lower level of interest or importance placed on chain loyalty, suggesting a greater degree of indifference towards this factor in their shopping location decisions.

POI count During the COVID-19 pandemic, residents with higher levels of education, and African-American residents recognized the importance of the abundance of POIs in their decision-making process for selecting department stores. However, in 2021, CBGs with a higher proportion of Asian residents, higher levels of education, and a higher median age among their residents showed a preference for stores that had a greater number of POIs in close proximity. This suggests a shift in the importance placed on the availability and variety of nearby amenities as the pandemic situation evolved.

POI diversity During the pandemic and as society began to recover in 2021, African-American residents emerged as the leading group in placing a strong emphasis on the diversity and richness of POIs. This indicates their preference for retail locations that offer a wide range of amenities and services to meet their diverse needs and interests.

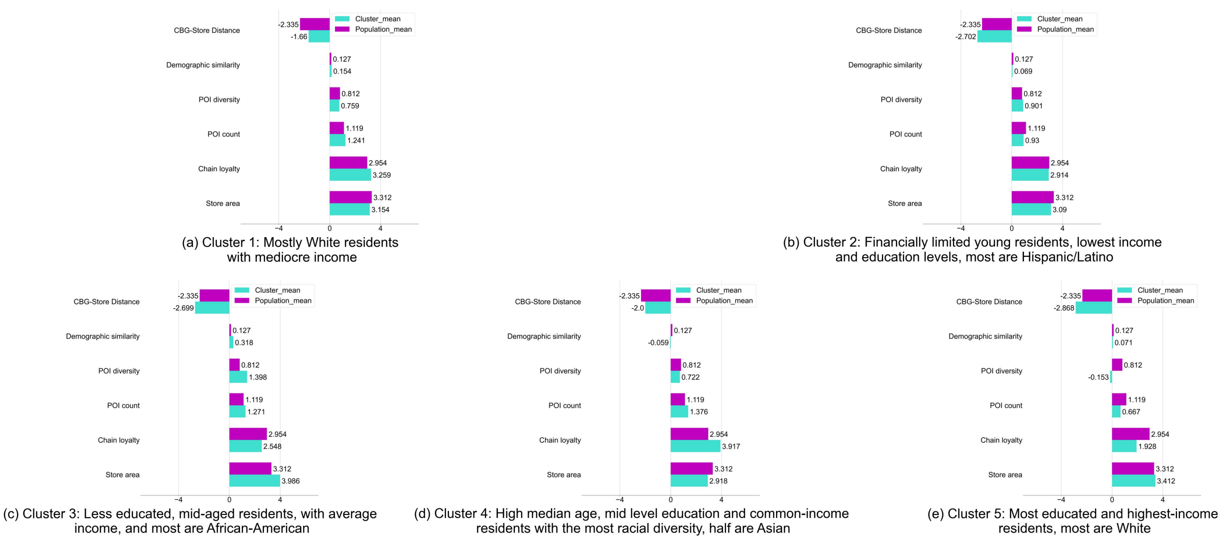
Demographic similarity Residents residing in CBGs with higher levels of education and income displayed a stronger inclination towards selecting stores located in neighborhoods that shared greater demographic similarities with their own community. This observation suggests that the decision-making process of residents regarding department store selection may be influenced by the characteristics and preferences of the community they belong to.

CBG-Store distance As society transitioned into the post-pandemic era, it was observed that CBGs characterized by higher education levels, a larger African-American population, and an older population demonstrated a greater increase in their attention towards this factor compared to the general population.

B.0.2 Clusters by combined socioeconomic characteristics: Five K-means clusters

This section aims to analyze and compare the temporal store selection preferences among five clusters that have been defined using the K-means algorithm and are based on combined socioeconomic characteristics. We also include the general NYC population for comparison. Specifically, we examine the differences in preferences for each attractiveness factor and place a particular emphasis on understanding the direction of change over time. Details regarding the definitions of these distinctive communities can be found in Section 5.3.2.

Figures B3 and B4 show changes in the value of $\Delta\omega_t$ between 2020 and 2021. It is abundantly clear that from 2020 to 2021, $\Delta\omega_t$ changes reversed roughly among each community and the entire population. People's earlier increased interest in store size, chain loyalty, POI count, and POI diversity from 2019 to 2020 began to decline from 2020 to 2019, while their indifference to CBG-store distance persisted in 2021. The magnitudes of the changes in 2021, nonetheless, are not as large as those in 2020. This occurrence confirms our earlier observation that following the outbreak, the decision patterns of store choice of New Yorkers changed from their pre-epidemic behavior and the impact of COVID-19 has not been fully recovered by 2021.



(a) Cluster 1: Mostly White residents with mediocre income

(b) Cluster 2: Financially limited young residents, lowest income and education levels, most are Hispanic/Latino

(c) Cluster 3: Less educated, mid-aged residents, with average income, and most are African-American

(d) Cluster 4: High median age, mid level education and common-income residents with the most racial diversity, half are Asian

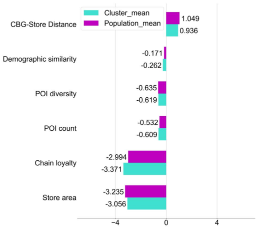
(e) Cluster 5: Most educated and highest-income residents, most are White

Fig. B3: **Mean $\Delta\omega_t$ value comparisons in clusters by K-means as of 2020**

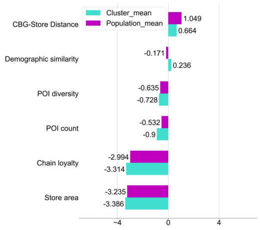
It has also been observed that changes in $\Delta\omega_t$ due to store area, chain loyalty and CBG-store distance are apparently greater than those in POI count, POI diversity and demographic similarity in 2020, while in 2021 changes in $\Delta\omega_t$ in CBG store distance are similar to those of POI count and diversity. Consequently, it has been signaled that since the outbreak of COVID-19, the products and services that a store can provide have caught more attention instead of the nearby commercial service and the cost of commuting. Overall, our findings using unsupervised clustering confirm that the pandemic response was heterogeneous among different communities.

Appendix C Abbreviations used in this paper

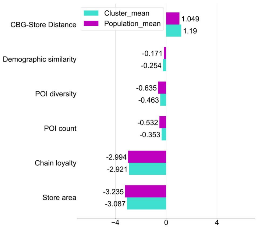
CBG: Census Block Group
CDF: Cumulative Distribution Function



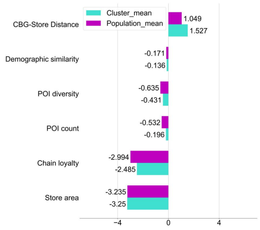
(a) Cluster 1: Mostly White residents with mediocre income



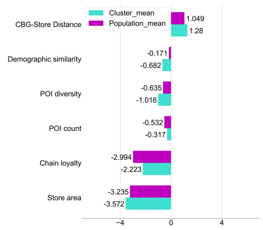
(b) Cluster 2: Financially limited young residents, lowest income and education levels, most are Hispanic/Latino



(c) Cluster 3: Less educated, mid-aged residents, with average income, and most are African-American



(d) Cluster 4: High median age, mid level education and common-income residents with the most racial diversity, half are Asian



(e) Cluster 5: Most educated and highest-income residents, most are White

Fig. B4: **Mean $\Delta\omega_t$ value comparisons in clusters by K-means as of 2021**

DID: Difference-in-Differences
GeoDa: Geographical and Spatial Data Analysis Software
GLM: Generalized Linear Model
GWR: Geographically Weighted Regression
MCI: Multiplicative Competitive Interaction
NAICS: North American Industry Classification System
NYC: New York City
OLS: Ordinary Least Squares
POI: Point of Interest
PSO: Particle Swarm Optimization
SA: Simulated Annealing

References

[1] Bavel, J.J.V., Baicker, K., Boggio, P.S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M.J., Crum, A.J., Douglas, K.M., Druckman, J.N., *et al.*: Using social and behavioural science to support covid-19 pandemic response. Nature human behaviour **4**(5), 460–471 (2020)

[2] Cronin, C.J., Evans, W.N.: Private precaution and public restrictions: what drives social distancing and industry foot traffic in the covid-19 era? Technical report, National Bureau of Economic Research (2020)

[3] Zion, S.R., Louis, K., Horii, R., Leibowitz, K., Heathcote, L.C., Crum, A.J.: Making sense of a pandemic: Mindsets influence emotions, behaviors, health, and wellbeing during the covid-19 pandemic. Social Science & Medicine **301**, 114889 (2022)

[4] Birge, J.R., Candogan, O., Feng, Y.: Controlling epidemic spread: Reducing economic losses with targeted closures. Management Science (2022)

[5] Lee, M., Zhao, J., Sun, Q., Pan, Y., Zhou, W., Xiong, C., Zhang, L.: Human mobility trends during the early stage of the covid-19 pandemic in the united states. PLoS One **15**(11), 0241468 (2020)

[6] Chan, H.F., Skali, A., Savage, D.A., Stadelmann, D., Torgler, B.: Risk attitudes and human

mobility during the covid-19 pandemic. Scientific reports **10**(1), 1–13 (2020)

[7] Chang, S., Pierson, E., Koh, P.W., Gerardin, J., Redbird, B., Grusky, D., Leskovec, J.: Mobility network models of covid-19 explain inequities and inform reopening. Nature **589**(7840), 82–87 (2021)

[8] Dwyer, M.J., Pasini, M., De Dominicis, S., Righi, E.: Physical activity: Benefits and challenges during the covid-19 pandemic. Scandinavian journal of medicine & science in sports **30**(7), 1291 (2020)

[9] Hunter, R.F., Garcia, L., Sa, T.H., Zapata-Diomedi, B., Millett, C., Woodcock, J., Pentland, A., Moro, E.: Effect of covid-19 response policies on walking behavior in us cities. Nature communications **12**(1), 1–9 (2021)

[10] Stockwell, S., Trott, M., Tully, M., Shin, J., Barnett, Y., Butler, L., McDermott, D., Schuch, F., Smith, L.: Changes in physical activity and sedentary behaviours from before to during the covid-19 pandemic lockdown: a systematic review. BMJ open sport & exercise medicine **7**(1), 000960 (2021)

[11] Chetty, R., Friedman, J.N., Hendren, N., Stepner, M., et al.: The economic impacts of covid-19: Evidence from a new public database built using private sector data. Technical report, national Bureau of economic research (2020)

[12] Mulay, N., Bishnoi, V., Katyal, Y., Bahrami, M., Moro, E., Saraswat, A., Pentland, A.: Effects of stimulus checks on spending patterns of different economic groups. In: 2021 International Conference on Data Mining Workshops (ICDMW), pp. 873–877 (2021). IEEE

[13] Liu, Q., Shen, Q., Li, Z., Chen, S.: Stimulating consumption at low budget: evidence from a large-scale policy experiment amid the covid-19 pandemic. Management Science **67**(12), 7291–7307 (2021)

[14] Yang, Z., Choe, Y., Martell, M.: Covid-19 economic policy effects on consumer spending and foot traffic in the us. Journal of Safety Science and Resilience **2**(4), 230–237 (2021)

[15] Chawla, A., Mulay, N., Bahrami, M., Bishnoi, V., Katyal, Y., Moro, E., Saraswat, A., Pentland, A.: Post-pandemic economic transformations in the united states of america. In: 2022 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 1–5 (2022). IEEE

[16] Rose, N., Rowe, F., Dolega, L.: How consumer behaviours changed in response to covid-19 lockdown stringency measures: A case study of walmart. Applied Geography **154**, 102948 (2023)

[17] Grashuis, J., Skevas, T., Segovia, M.S.: Grocery shopping preferences during the covid-19 pandemic. Sustainability **12**(13), 5369 (2020)

[18] Moon, J., Choe, Y., Song, H.: Determinants of consumers' online/offline shopping behaviours during the covid-19 pandemic. International journal of environmental research and public health **18**(4), 1593 (2021)

[19] Lucchini, L., Centellegher, S., Pappalardo, L., Gallotti, R., Privitera, F., Lepri, B., De Nadai, M.: Living in a pandemic: changes in mobility routines, social activity and adherence to covid-19 protective measures. Scientific reports **11**(1), 1–12 (2021)

[20] Park, I., Lee, J., Lee, D., Lee, C., Chung, W.Y.: Changes in consumption patterns during the covid-19 pandemic: Analyzing the revenge spending motivations of different emotional groups. Journal of Retailing and Consumer Services **65**, 102874 (2022)

[21] Phillips, S.: Working through the pandemic: Accelerating the transition to remote working. Business Information Review **37**(3), 129–134 (2020)

[22] Klerk, J.J.d., Joubert, M., Mosca, H.F.: Is working from home the new workplace panacea? lessons from the covid-19 pandemic for the future world of work. SA Journal of Industrial Psychology **47**(1), 1–14 (2021)

[23] Jones, C., Philippon, T., Venkateswaran, V.: Optimal mitigation policies in a pandemic: Social distancing and working from home. The Review of Financial Studies **34**(11), 5188–5223 (2021)

[24] Lowenthal, P., Borup, J., West, R., Archambault, L.: Thinking beyond zoom: Using asynchronous video to maintain connection and engagement during the covid-19 pandemic. Journal of Technology and Teacher Education **28**(2), 383–391 (2020)

[25] Coven, J., Gupta, A.: Disparities in mobility responses to covid-19. Unpublished Manuscript. https://static1. squarespace. com/static/56086d00e4b0fb7874bc2d42 **5**, 1589583893816 (2020)

[26] Holtz, D., Zhao, M., Benzell, S.G., Cao, C.Y., Rahimian, M.A., Yang, J., Allen, J., Collis, A., Moehring, A., Sowrirajan, T., *et al.*: Interdependence and the cost of uncoordinated responses to covid-19. Proceedings of the National Academy of Sciences **117**(33), 19837–19843 (2020)

[27] Heroy, S., Loaiza, I., Pentland, A., O'Clery, N.: Covid-19 policy analysis: labour structure dictates lockdown mobility behaviour. Journal of the Royal Society Interface **18**(176), 20201035 (2021)

[28] Maleki, M., Bahrami, M., Menendez, M., Balsa-Barreiro, J.: Social behavior and covid-19: Analysis of the social factors behind compliance with interventions across the united states. International journal of environmental research and public health **19**(23), 15716 (2022)

[29] Gür, M.: Post-pandemic lifestyle changes and their interaction with resident behavior in housing and neighborhoods: Bursa, turkey. Journal of Housing and the Built Environment **37**(2), 823–862 (2022)

[30] Salon, D., Conway, M.W., Silva, D., Chauhan, R.S., Derrible, S., Mohammadian, A., Khoeini, S., Parker, N., Mirtich, L., Shamshiripour, A., *et al.*: The potential stickiness of pandemic-induced behavior changes in the united states. Proceedings of the National Academy of Sciences **118**(27), 2106499118 (2021)

[31] Oblander, E.S., McCarthy, D.: Persistence of consumer lifestyle choices: Evidence from restaurant delivery during covid-19. Columbia Business School Research Paper Forthcoming (2022)

[32] Sheth, J.: Impact of covid-19 on consumer behavior: Will the old habits return or die? Journal of business research **117**, 280–283 (2020)

[33] Kohli, S., Timelin, B., Fabius, V., Veranen, S.M.: How covid-19 is changing consumer behavior–now and forever. McKinsey & Company, 1–2 (2020)

[34] Chen, T., Jin, Y., Yang, J., Cong, G.: Identifying emergence process of group panic buying behavior under the covid-19 pandemic. Journal of Retailing and Consumer Services **67**, 102970 (2022)

[35] Alekseev, G., Amer, S., Gopal, M., Kuchler, T., Schneider, J.W., Stroebel, J., Wernerfelt, N.: The effects of covid-19 on us small businesses: evidence from owners, managers, and employees. Management Science **69**(1), 7–24 (2023)

[36] Berg, C.J., Callanan, R., Johnson, T.O., Schliecher, N.C., Sussman, S., Wagener, T.L., Meaney,

M., Henriksen, L.: Vape shop and consumer activity during covid-19 non-essential business closures in the usa. Tobacco control **30**(e1), 41–44 (2021)

[37] Showrav, D.G.Y., Hassan, M.A., Anam, S., Chakrabarty, A.K.: Factors influencing the rapid growth of online shopping during covid-19 pandemic time in dhaka city, bangladesh. Academy of Strategic Management Journal **20**(2), 1–13 (2021)

[38] Budd, C., Calvert, K., Johnson, S., Tickle, S.: Assessing risk in the retail environment during the covid-19 pandemic. Royal Society Open Science **8**(5), 210344 (2021)

[39] Cox, N., Ganong, P., Noel, P., Vavra, J., Wong, A., Farrell, D., Greig, F., Deadman, E.: Initial impacts of the pandemic on consumer behavior: Evidence from linked income, spending, and savings data. Brookings Papers on Economic Activity **2020**(2), 35–82 (2020)

[40] Grimmer, L.: Lessons from the covid19 pandemic: The case of retail and consumer service firms. Journal of Retailing and Consumer Services **68**, 103012 (2022)

[41] Eger, L., Komárková, L., Egerová, D., Mičík, M.: The effect of covid-19 on consumer shopping behaviour: Generational cohort perspective. Journal of Retailing and Consumer Services **61**, 102542 (2021)

[42] Khaliq, M.: Impact of the coronavirus (covid-19) pandemic on retail sales in 2020. Office for (2021)

[43] Pantano, E., Pizzi, G., Scarpi, D., Dennis, C.: Competing during a pandemic? retailers' ups and downs during the covid-19 outbreak. Journal of Business research **116**, 209–213 (2020)

[44] Erkip, F.: The shopping mall as an emergent public space in turkey. Environment and planning A **35**(6), 1073–1093 (2003)

[45] Kunc, J., Križan, F., Novotná, M., Bilková, K.: Social dimension of shopping centers operation: Managerial perspectives. Sustainability **14**(2), 709 (2022)

[46] Foroohar, R.: Small business: A canary in the us economic coal mine. Financial Times (2020)

[47] Beckers, J., Weekx, S., Beutels, P., Verhetsel, A.: Covid-19 and retail: The catalyst for e-commerce in belgium? Journal of Retailing and Consumer Services **62**, 102645 (2021)

[48] Zhang, Z., Wen, X.: Physical or virtual showroom? the decision for omni-channel retailers in the context of cross-channel free-riding. Electronic Commerce Research, 1–27 (2023)

[49] Cruz-Cárdenas, J., Zabelina, E., Guadalupe-Lanas, J., Palacio-Fierro, A., Ramos-Galarza, C.: Covid-19, consumer behavior, technology, and society: A literature review and bibliometric analysis. Technological forecasting and social change **173**, 121179 (2021)

[50] Huff, D.L.: Defining and estimating a trading area. Journal of marketing **28**(3), 34–38 (1964)

[51] Chenarides, L., Grebitus, C., Lusk, J.L., Printezis, I.: Food consumption behavior during the covid-19 pandemic. Agribusiness **37**(1), 44–81 (2021)

[52] Nanda, A., Xu, Y., Zhang, F.: How would the covid-19 pandemic reshape retail real estate and high streets through acceleration of e-commerce and digitalization? Journal of Urban Management **10**(2), 110–124 (2021)

[53] Puccinelli, N.M., Goodstein, R.C., Grewal, D., Price, R., Raghubir, P., Stewart, D.: Customer

experience management in retailing: understanding the buying process. Journal of retailing **85**(1), 15–30 (2009)

[54] Sayyida, S., Hartini, S., Gunawan, S., Husin, S.N.: The impact of the covid-19 pandemic on retail consumer behavior. Aptisi Transactions on Management (ATM) **5**(1), 79–88 (2021)

[55] Hall, M.C., Prayag, G., Fieger, P., Dyason, D.: Beyond panic buying: consumption displacement and covid-19. Journal of service management **32**(1), 113–128 (2020)

[56] Wang, X., Shi, W., Yuen, K.F.: A synthesised review of pandemic-driven consumer behaviours and applied theories: towards a unified framework. International Journal of Disaster Risk Reduction **77**, 103104 (2022)

[57] Sundström, M., Lundberg, C., Ziakas, V.: Episodic retail settings: A sustainable and adaptive strategy for city centre stores. Sustainability **13**(5), 2482 (2021)

[58] Lashgari, Y.S., Shahab, S.: The impact of the covid-19 pandemic on retail in city centres. Sustainability **14**(18), 11463 (2022)

[59] Ballantyne, P., Singleton, A., Dolega, L.: Using unstable data from mobile phone applications to examine recent trajectories of retail centre recovery. Urban informatics **1**(1), 21 (2022)

[60] Koch, J., Frommeyer, B., Schewe, G.: Online shopping motives during the covid-19 pandemic—lessons from the crisis. Sustainability **12**(24), 10247 (2020)

[61] Jiang, Y., Lai, P.-L., Yang, C.-C., Wang, X.: Exploring the factors that drive consumers to use contactless delivery services in the context of the continued covid-19 pandemic. Journal of Retailing and Consumer Services **72**, 103276 (2023)

[62] Aryani, D.N., Nair, R.K., Hoo, D.X.Y., Hung, D.K.M., Lim, D.H.R., Chew, W.P., Desai, A., *et al.*: A study on consumer behaviour: Transition from traditional shopping to online shopping during the covid-19 pandemic. International Journal of Applied Business and International Management (IJABIM) **6**(2), 81–95 (2021)

[63] Naseri, R.N.N., *et al.*: Issues and challenges of online shoppingactivities on the impact of corona pandemic: A study on malaysia retail industry. Turkish Journal of Computer and Mathematics Education (TURCOMAT) **12**(10), 7682–7686 (2021)

[64] Gomes, S., Lopes, J.M.: Evolution of the online grocery shopping experience during the covid-19 pandemic: Empiric study from portugal. Journal of Theoretical and Applied Electronic Commerce Research **17**(3), 909–923 (2022)

[65] Alfonso, V., Boar, C., Frost, J., Gambacorta, L., Liu, J.: E-commerce in the pandemic and beyond. BIS bulletin **36**(9), 1–9 (2021)

[66] Szasz, L., Balint, C., Csiki, O., Nagy, B.Z., Racz, B.-G., Csala, D., Harris, L.C.: The impact of covid-19 on the evolution of online retail: The pandemic as a window of opportunity. Journal of Retailing and Consumer Services **69**, 103089 (2022)

[67] Gungordu Belbag, A.: Impacts of covid-19 pandemic on consumer behavior in turkey: A qualitative study. Journal of Consumer Affairs **56**(1), 339–358 (2022)

[68] Timotius, E., Octavius, G.S.: Global changing of consumer behavior to retail distribution due to pandemic of covid-19: a systematic review. Journal of Distribution Science **19**(11), 69–80 (2021)

[69] Aditantri, R., Mahliza, F., Wibisono, A.D.: Urban planning and e-commerce: Understanding the impact during pandemic covid-19 in jakarta. International Journal of Business, Economics, and Social Development **2**(3), 135–142 (2021)

[70] Wieland, T.: Identifying the determinants of store choice in a multi-channel environment: a hurdle model approach. Papers in Applied Geography **7**(4), 343–371 (2021)

[71] Wieland, T.: A micro-econometric store choice model incorporating multi-and omni-channel shopping: The case of furniture retailing in germany. Geographical Analysis **55**(1), 3–30 (2023)

[72] Wieland, T.: Spatial shopping behavior during the corona pandemic: insights from a micro-econometric store choice model for consumer electronics and furniture retailing in germany. Journal of Geographical Systems **25**(2), 291–326 (2023)

[73] Alhaimer, R.: Fluctuating attitudes and behaviors of customers toward online shopping in times of emergency: The case of kuwait during the covid-19 pandemic. Journal of Internet Commerce **21**(1), 26–50 (2022)

[74] Dou, Z., Stefanovski, D., Galligan, D., Lindem, M., Rozin, P., Chen, T., Chao, A.M.: The covid-19 pandemic impacting household food dynamics: A cross-national comparison of china and the us (2020)

[75] Boyle, P., Bond, R., Martinez Carracedo, J., Simmons, G., Mulvenna, M., Hollywood, L.: The impact of the covid-19 pandemic on grocery shopper behaviour: Analysis of shopper behaviour change using store transaction data. Journal of Consumer Behaviour **21**(2), 259–271 (2022)

[76] Aversa, J., Quach, X., Hernandez, T., Vatrapu, R.: The effect of covid-19 on customer traffic: A case study of food and beverage stores in erie county, new york (2022)

[77] Rossetti, T., Yoon, S.-Y., Daziano, R.A.: Social distancing and store choice in times of a pandemic. Journal of Retailing and Consumer Services **65**, 102860 (2022)

[78] Sevtsuk, A., Hudson, A., Halpern, D., Basu, R., Ng, K., Jong, J.: The impact of covid-19 on trips to urban amenities: Examining travel behavior changes in somerville, ma. PloS one **16**(9), 0252794 (2021)

[79] Yuen, K.F., Wang, X., Ma, F., Li, K.X.: The psychological causes of panic buying following a health crisis. International journal of environmental research and public health **17**(10), 3513 (2020)

[80] Naeem, M.: Understanding the customer psychology of impulse buying during covid-19 pandemic: implications for retailers. International Journal of Retail & Distribution Management **49**(3), 377–393 (2021)

[81] Cliquet, G.: From geomarketing to spatial marketing. Spatial Economics Volume II: Applications, 277–305 (2021)

[82] Converse, P.D.: New laws of retail gravitation. Journal of marketing **14**(3), 379–384 (1949)

[83] Reilly, W.J.: The law of retail gravitation. Knickerbocker Press (1931)

[84] Huff, D.L.: A probabilistic analysis of shopping center trade areas. Land economics **39**(1), 81–90 (1963)

[85] Huff, D.L.: Parameter estimation in the Huff model. Esri, ArcUser, 34–36 (2003)

[86] Drezner, T., Drezner, Z.: Finding the optimal solution to the Huff based competitive location model. Computational Management Science **1**(2), 193–208 (2004)

[87] Kim, P.-J., Kim, W., Chung, W.-K., Youn, M.-K.: Using new Huff model for predicting potential retail market in south korea. African Journal of Business Management **5**(5), 1543–1550 (2011)

[88] De Beule, M., Poel, D., Weghe, N.: An extended Huff model for robustly benchmarking and predicting retail network performance. Applied Geography **46**, 80–89 (2014)

[89] Bozkaya, B., Yanik, S., Balcisoy, S.: A gis-based optimization framework for competitive multi-facility location-routing problem. Networks and Spatial Economics **10**, 297–320 (2010)

[90] Seenivasan, S., Sudhir, K., Talukdar, D.: Do store brands aid store loyalty? Management Science **62**(3), 802–816 (2016)

[91] Bahrami, M., Xu, Y., Tweed, M., Bozkaya, B., Pentland, A.: Using gravity model to make store closing decisions: A data driven approach. Expert Systems with Applications, 117703 (2022)

[92] Gripsrud, G., Horverak, Ø.: Determinants of retail patronage. a 'natural'experiment. International Journal of Research in Marketing **3**(4), 263–272 (1986)

[93] Arentze, T.A., Timmermans, H.J.: Deriving performance indicators from models of multipurpose shopping behavior. Journal of Retailing and Consumer Services **8**(6), 325–334 (2001)

[94] Netto, C.F.S., Bahrami, M., Brei, V.A., Bozkaya, B., Balcisoy, S., Pentland, A.P.: Disaggregating sales prediction: A gravitational approach. Expert Systems with Applications, 119565 (2023)

[95] France, S.L., Ghose, S.: Marketing analytics: Methods, practice, implementation, and links to other fields. Expert Systems with Applications **119**, 456–475 (2019)

[96] Dolega, L., Pavlis, M., Singleton, A.: Estimating attractiveness, hierarchy and catchment area extents for a national set of retail centre agglomerations. Journal of Retailing and Consumer Services **28**, 78–90 (2016)

[97] Wrigley, N., Dolega, L.: Resilience, fragility, and adaptation: new evidence on the performance of uk high streets during global economic crisis and its policy implications. Environment and Planning A **43**(10), 2337–2363 (2011)

[98] Dolega, L., Reynolds, J., Singleton, A., Pavlis, M.: Beyond retail: New ways of classifying uk shopping and consumption spaces. Environment and Planning B: Urban Analytics and City Science **48**(1), 132–150 (2021)

[99] Ballantyne, P., Singleton, A., Dolega, L., Credit, K.: A framework for delineating the scale, extent and characteristics of american retail centre agglomerations. Environment and Planning B: Urban Analytics and City Science **49**(3), 1112–1128 (2022)

[100] Singleton, A.D., Wilson, A., O'Brien, O.: Geodemographics and spatial interaction: an integrated model for higher education. Journal of Geographical Systems **14**, 223–241 (2012)

[101] Nakanishi, M., Cooper, L.G.: Simplified estimation procedures for mci models. Marketing Science **1**(3), 314–322 (1982)

[102] Huff, D., McCALLUM, B.M.: Calibrating the Huff model using arcgis business analyst. ESRI White Paper, 1–33 (2008)

[103] Beiró, M.G., Bravo, L., Caro, D., Cattuto, C., Ferres, L., Graells-Garrido, E.: Shopping mall attraction and social mixing at a city scale. EPJ Data Science **7**(1), 1–21 (2018)

[104] Suárez-Vega, R., Gutiérrez-Acuna, J.L., Rodríguez-Díaz, M.: Locating a supermarket using a locally calibrated Huff model. International Journal of Geographical Information Science **29**(2), 217–233 (2015)

[105] Suhara, Y., Bahrami, M., Bozkaya, B., Pentland, A.: Validating gravity-based market share models using large-scale transactional data. Big Data **9**(3), 188–202 (2021)

[106] Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of ICNN'95-international Conference on Neural Networks, vol. 4, pp. 1942–1948 (1995). IEEE

[107] Liang, Y., Gao, S., Cai, Y., Foutz, N.Z., Wu, L.: Calibrating the dynamic Huff model for business analysis using location big data. Transactions in GIS **24**(3), 681–703 (2020)

[108] Freitas, D., Lopes, L.G., Morgado-Dias, F.: Particle swarm optimisation: a historical review up to the current developments. Entropy **22**(3), 362 (2020)

[109] Escalante, H.J., Montes, M., Sucar, L.E.: Particle swarm model selection. Journal of Machine Learning Research **10**(2) (2009)

[110] Li, W., Wang, Q., Liu, Y., Small, M.L., Gao, J.: A spatiotemporal decay model of human mobility when facing large-scale crises. Proceedings of the National Academy of Sciences **119**(33), 2203042119 (2022)

[111] Zhou, H., Kim, G., Wang, J., Wilson, K.: Investigating the association between the socioeconomic environment of the service area and fast food visitation: A context-based crystal growth approach. Health & Place **76**, 102855 (2022)

[112] Fe, H., Sanfelice, V.: How bad is crime for business? evidence from consumer behavior. Journal of urban economics **129**, 103448 (2022)

[113] Bahrami, M.: A Perspective Towards Using Aggregated Data in Research. OSF Preprints (2023). https://doi.org/10.31219/osf.io/k268q

[114] Athey, S., Ferguson, B.A., Gentzkow, M., Schmidt, T.: Experienced segregation. Technical report, National Bureau of Economic Research (2020)

[115] Jay, J., Bor, J., Nsoesie, E.O., Lipson, S.K., Jones, D.K., Galea, S., Raifman, J.: Neighbourhood income and physical distancing during the covid-19 pandemic in the united states. Nature human behaviour **4**(12), 1294–1302 (2020)

[116] Klise, K., Beyeler, W., Finley, P., Makvandi, M.: Analysis of mobility data to build contact networks for covid-19. PLoS One **16**(4), 0249726 (2021)

[117] Li, Z., Ning, H., Jing, F., Lessani, M.N.: Understanding the bias of mobile location data across spatial scales and over time: a comprehensive analysis of safegraph data in the united states. Available at SSRN 4383333 (2023)

[118] Ogawa, H.: Preference for product variety and city size. Urban Studies **35**(1), 45–51 (1998)

[119] Glaeser, E.L., Kolko, J., Saiz, A.: Consumer city. Journal of economic geography **1**(1), 27–50 (2001)

[120] Adamson, D.W., Clark, D.E., Partridge, M.D.: Do urban agglomeration effects and household amenities have a skill bias? Journal of Regional Science **44**(2), 201–224 (2004)

[121] Chong, S.K., Bahrami, M., Chen, H., Balcisoy, S., Bozkaya, B., *et al.*: Economic outcomes predicted by diversity in cities. EPJ Data Science **9**(1), 17 (2020)

[122] Kirkpatrick, S., Gelatt Jr, C.D., Vecchi, M.P.: Optimization by simulated annealing. science **220**(4598), 671–680 (1983)

[123] Boz, H.A., Bahrami, M., Balcisoy, S., Bozkaya, B., Mazar, N., Nichols, A., Pentland, A.: Investigating neighborhood adaptability using mobility networks: a case study of the covid-19 pandemic. Humanities and Social Sciences Communications **11**(1), 1–11 (2024)

[124] Massey Jr, F.J.: The kolmogorov-smirnov test for goodness of fit. Journal of the American statistical Association **46**(253), 68–78 (1951)

[125] Morozov, I., Seiler, S., Dong, X., Hou, L.: Estimation of preference heterogeneity in markets with costly search. Marketing Science **40**(5), 871–899 (2021)

[126] Anselin, L., Syabri, I., Kho, Y.: Geoda: an introduction to spatial data analysis. In: Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications, pp. 73–89 (2009)