Marketing Science Institute Working Paper Series 2022

Report No. 22-102

Balancing Consumer Privacy with Marketing Insights in Mobile Location Data

Tal Shoshani, Peter Pal Zubcsek and David A. Schweidel

Balancing Consumer Privacy with Marketing Insights in Mobile Location Data

Tal Shoshani

Peter Pal Zubcsek

David A. Schweidel*

* Tal Shoshani is a PhD student at Marshall School of Business, University of Southern California. Peter Pal Zubcsek is Senior Lecturer in Marketing at Coller School of Management, Tel Aviv University. David A. Schweidel is Rebecca Cheney McGreevy Endowed Chair and Professor of Marketing at Emory University's Goizueta Business School. The authors thank the Marketing Science Institute and the Coller Foundation for their financial support, and Mogean for providing the data for this research.

Balancing Consumer Privacy with Marketing Insights in Mobile Location Data

Abstract

Mobile location data offers granular detail into consumers' mobility patterns. The data reveal where consumers spend their time, which may be indicative of consumer preferences and ultimately behavior. The detailed nature of the data, however, can reveal sensitive information about consumers based on the places that they visit such as healthcare facilities, social connections or educational institutions. At the same time, such data offers an opportunity for marketers to leverage detailed location data for real-time and contextually relevant marketing. Can the desires of both marketers seeking actionable insights and individuals preferring privacy be accommodated? Using device-level data from a location data provider, we examine the extent to which predictive performance is affected by grouping individuals into homogeneous clusters that afford increased privacy. In our empirical context, we find that some level of aggregation results in higher predictive accuracy. Our results also reveal that reliance on the locations of commercial activity yields results that are at least as good as home locations. Taken together, our findings offer guidance to data providers who must balance service to their clients with consumers' expectations of privacy, as well as providing regulators with insight into the data granularity that marketers require for their operations.

INTRODUCTION

Location data derived from mobile devices has been a boon for marketers, enabling increased precision to support targeted marketing. Such data has been used to support the delivery of mobile promotions, taking advantage of users' proximity to a focal retailer (e.g., Luo, Andrews, Fang and Phang 2014; Fong, Fang and Luo 2015), and individuals' location trajectories have been used to further refine the accuracy of predicting users' responsiveness to mobile advertising (Ghose, Li and Liu 2019; Zubcsek, Katona and Sarvary 2017). Despite the potential use of mobile location data, however, concerns have been raised about the extent to which the collection and analysis of mobile location data invades consumer privacy (e.g., Bengio et al. 2020; Thompson and Warzel 2019).

While legislation such as the European Union's General Data Protection Regulation (GDPR) have only recently thrust data privacy and consumers' privacy preferences into the spotlight, academic research has long explored this question in the context of targeted digital advertising. Goldfarb and Tucker (2012) examined how consumers' privacy concerns have evolved over time, finding that consumers were becoming increasingly likely to refuse to reveal information in the early 2000s. In her investigation of social network advertising, Tucker (2014) showed that users were more likely to click on personalized advertisements after they perceived an increase in their control over the privacy of their data. Bleier et al. (2020) identified factors that mitigate consumer privacy concerns, including trust, information sensitivity and transparency. Finally, in addition to research on consumer preferences regarding privacy, research has also developed new mechanisms to preserve privacy (e.g., Schneider et al. 2017; 2018).

Though the notion of privacy and its potential impact on marketing is well-established (e.g., Mayer and White 1969), it has received increased scrutiny as it pertains to location data. De Montjoye et al. (2013) demonstrated that as little as four timestamped locations are sufficient to identify 95% of individuals. Moreover, the authors showed that coarsening the spatial or temporal resolution with which the data are collected offers little benefit in terms of user privacy. More recently, investigative reports have brought increased attention to the threat that location data poses to privacy due to the level of precision with which data are collected. Based on the sequence of locations visited by one device out of more than 10 million in the dataset, Thompson and Warzel (2019) inferred that the device belonged to someone who travels with the U.S. President, suggesting that mobile location data not only pose a threat to consumer privacy but also to national security. Location data from mobile devices also enabled the identification of individuals who participated in the January 6, 2021 riot at the U.S. Capitol (Warzel and Thompson 2021). In light of the privacy concerns that have been raised, technology companies have begun to take positions with regards to consumer privacy (e.g., Snider 2021; Laziuk 2021). However, as for the practice of collecting location data to aid in contact tracing during the COVID-19 pandemic, while concerns were expressed about the importance of privacy (e.g., Bengio et al. 2020), consumers were found to be more willing to trade privacy in exchange for the social good (Ghose et al. 2020).

One way in which research leveraging location data has addressed privacy concerns is by aggregating data across devices. Such data sources recently have been used to investigate the spread of COVID-19 and related policy decisions. For instance, Liu, Thomadsen and Yao (2020) modeled the spread of COVID-19 based on assumptions of social distancing policies. Chiou and Tucker (2020) examined differences in mobility behavior related to income and high-speed

Internet access, finding that users are more likely to remain in their homes if they have high income or access to high-speed internet access.

In light of the availability of location data from mobile devices and increased attention to the ensuing privacy concerns, the question arises as to how marketers' interests can be balanced with consumers' desire to maintain privacy. In the context of mobile advertising, Rafieian and Yoganarasimhan (2020) showed that as targeting increases (thereby reducing privacy), most advertisers stand to benefit, suggesting that privacy and the interest of marketers are at odds. In this research, we investigate whether consumers' desire for privacy and marketers' ability to derive predictive insights from mobile location data are at odds with each other, and, if so, what options are available to regulators.

Consider a brand that wants to target customers with advertising. While a brand may incorporate location-based services into its own mobile app, which would allow it to target those who choose to download it, they may desire to reach a broader audience. One way in which the brand may embark on such an effort would be to identify its current customers or its competitors' customers by using historic mobile location data offered by a third-party data provider, enabling the marketer to target these devices with advertising (Goldfarb and Tucker 2020). While this enables a brand to engage in targeted marketing, such data poses a privacy concern because it could enable the identification of individuals' homes and workplaces, as well as locations they have visited that they might not wish known by others such as healthcare or religiously affiliated locations (e.g., Macha et al. 2021). Could the brand's objective of identifying potential customers for targeting be accomplished without creating the risk that individuals could be identified from the underlying data? This is the fundamental question that we address in this research.

One way in which data providers and brands could reduce such risk would be to only make use of brand-related information. Rather than knowing the specific Starbucks location that is visited, it may be sufficient to summarize the behavior of the device as the number of times it has visited any Starbucks location in a given time period. By restricting the data employed to brand visits rather than requiring the identification of the specific location that is visited, we reduce the likelihood that an individual could be reidentified from the data. However, such information may not be sufficient for marketers to use as the basis for targeted advertising. Marketers may want to focus their efforts on residents of specific neighborhoods, perhaps due to underlying demographic factors or proximity to brick-and-mortar locations. To enable such targeting, marketers would need to link devices to these neighborhoods, which raises concerns about the ability to link a device to a specific individual.

Another avenue to reduce the risk by which an individual could be identified from location data would be through aggregation, which is being explored in the context of digital marketing as a means of preserving consumer privacy while still enabling targeted advertising (Bohn 2021). Grouping devices into homogeneous clusters could enable marketers to identify desirable clusters while not being able to link a device to a particular individual.

In this research, we consider different mechanisms by which individuals can be aggregated into homogeneous clusters of varying sizes. While detailed, individual-level location histories may be predictive of future behavior, such granularity jeopardizes consumer privacy. Conversely, aggregating consumers into large heterogeneous segments may compromise the predictive value of the data. Our analysis will shed light on the impact of partitioning consumers into segments of different sizes, allowing us to determine if there is a resolution that preserves both user privacy and predictive accuracy.

We anticipate that predictive performance will initially improve as the size of segments diminishes. That is, as we move from a single heterogeneous segment for all individuals to smaller, more homogeneous segments, we would expect that predictive ability increases. However, at some resolution, splitting segments further may result in over-fitting, yielding poorer predictive performance.

In addition to exploring the impact of segment resolution on predictive accuracy, we also probe the basis by which homogeneous segments are formed. One approach is to form segments on the basis of historic brand visitation data. Under such an approach, individuals who visit the same brands would be grouped into the same segments. Another option would be to cluster individuals based on their "home" locations. As mobile device location data consists of a device identifier, timestamp, and latitude and longitude, the home location of a device can be inferred, such as by using the most visited latitude/longitude during overnight hours (e.g., Alesandretti et al. 2018). To the extent that individuals live around like-minded individuals (e.g., McPherson et al. 2001), the use of home location to form clusters of individuals may implicitly incorporate brand preferences. Additionally, individuals who live in the same areas are confronted with similar distances to different brands' locations, which may affect the frequency with which they visit such locations.

Despite the intuitive appeal of partitioning based on an individual's home location, this runs counter to efforts to afford consumers more privacy. Is there a way of capturing the information contained in one's home location without infringing on one's privacy to such a large extent? Toward this end, we propose the use of a "pseudo-home" location that we construct as the centroid of the businesses visited by a device. An appealing aspect of clustering individuals in this way is that it only requires that data be collected when individuals are detected at business

locations. As location data generated in residential areas is superfluous to the analysis, it need not be collected, allowing this approach to provide consumers with an enhanced degree of privacy.

To the best of our knowledge, our work is among the first to probe location data from the standpoint of data privacy, examining whether this is at odds with marketers' goals or if privacy can be achieved without compromising business performance. By examining the segment resolution and the basis for partitioning individuals, our analysis offers guidance to location data providers and their clients about how such data can be utilized to produce marketing insights in a privacy-friendly fashion. That is, rather than viewing privacy and marketing insights as being inherently at odds with each other, we posit that this is a false dichotomy and demonstrate how both consumers' desire for privacy and marketers' ability to target individuals can be achieved. In doing so, our research sheds light on potential considerations that regulators may wish to take into account as they grapple with data privacy.

In the next section, we describe the data we use to conduct our empirical analysis. We then detail our empirical approach and discuss our findings. We conclude with a discussion of recommendations for the use of mobile location data for marketing insights.

DATA

Data have been provided by Mogean, a marketing firm that specializes in the collection and analysis of location data collected through a network of mobile apps. Our data contain mobile device activity in a state in the southeastern United States spanning a ten-week period from January 2020 through early March 2020. We focus on this time period to avoid the potential

impact of COVID-19, as statewide regulations affecting mobility behavior were enacted in late March 2020.

The data were collected from mobile apps based on the permissions set by the device user. Each record contains the identifier and the GPS location of the observed device, along with the timestamp of the observation. In addition, for observations in the immediate vicinity of retailers, we observe the brand whose location the device was the closest to.

We divide our data into three consecutive stages. We use the first two weeks of data – the "forming" stage – to partition devices into segments based on criteria described below. Weeks three to six of our data constitute the "training" stage. We use brand visitation data from this period to train a model that predicts the future brand visits of *each device* based on *segment-level* aggregations of past behavior. Finally, the last four weeks of our data constitute the "test stage." Using segment-level brand visitation data as inputs, at this stage we evaluate the predictive accuracy of our models. We compare the predictive performance, assessed by AUC, across different sets of data being used to form segments, for consumer partitions of varying granularity.

To avoid double-counting, we merge consecutive observations at the same location into one observation for each device. We treat each "branded" observation in the resulting dataset as *visit*s to the corresponding brand. We restrict our attention to consumers who visited at least one location of one of the top 200 most visited brands in our data during each week of the forming stage of our analysis. The resulting sample for the full ten weeks contains 1,668,395,906 observations that reveal 8,851,898 brand visits by 180,674 devices.

EMPIRICAL ANALYSIS

Overview

To investigate the potential tradeoff between predictive ability and consumer privacy, we operationalize privacy in terms of the minimum segment size $z > 0$. This operationalization is consistent with the notion of *k*-anonymity (Sweeney 2002), which holds that an individual in the data cannot be distinguished from at least *k*-1 other individuals. We propose two approaches for clustering mobile devices using the data collected during the forming stage. The first is to cluster individuals based on the brands they have visited previously. In this approach, we select a set of brands *B*, and we use individuals' visitation frequencies at each brand in *B* to derive clusters of homogeneous individuals. A second approach we test makes use of the device's "home" location, which we infer from the most common nighttime (8PM-4AM) location (Bettini et al. 2005). By including the GPS coordinates of home location in the underlying feature space, we can cluster individuals based on the similarity of their brand visitation behaviors and/or the proximity of their homes to each other. To account for the different scale of home location coordinates while preserving the proportions of brand visitation counts, we iterate over a range of different weights that are placed on brand visitation and home location data, respectively. This approach allows us to empirically assess how much additional predictive insight the incorporation of home location provides beyond brand visitation data.

We vary the size of the segments (i.e., the value for *z* which dictates the extent of privacy) and assess the predictive performance of the classifiers we train as a function of the degree of privacy enabled by the given resolution. To form groups, we use the constrained K-means clustering algorithm (Bradley et al. 2000), that allows us to restrict the minimum group size to any $0 < z \leq N$ for a sample of *N* devices. Thus, we can set the minimum number of

devices and their location observations to be grouped together, ensuring a higher degree of privacy by design. In contrast, alternative clustering methods such as (Hierarchical) Density-Based Spatial Clustering of Applications ((H)DBSCAN, Ester et al. 1996; Li and Xi 2011) may generate groups that contain as few as one device. This could allow for the reidentification of individuals, which we strive to avoid. To maximize this benefit of the constrained K-means algorithm, we restrict the minimum group size z to be equal to N / K (i.e., the number of devices in the sample divided by the number of groups to be formed), where K is a divisor of N, ensuring that the resulting groups each have the same size.

At the training and test stages, we average each feature – the weekly brand visitation counts – across all devices within each group, and we assign the averaged features from the cluster to which the device belongs as independent variables that correspond to each device. Using the training data, we then train a model to predict, using the group-level average features for week t, whether a given device will visit (any location of) a given brand on week $t+1$. That is, the dependent variable is whether or not (1 or 0) device i was observed at brand b in week $t+1$. While the predictors used in our analysis are the same for all individuals within a given segment, the dependent variable is the behavior of the individual device, not the average of the group. We assess model performance on a holdout sample during the test stage. That is, for weeks 7, 8, and 9, respectively, we calculate the average features for each group of devices, and use the model to predict the device-level visitation behavior (0 or 1) at the brands of interest on weeks 8, 9, and 10, respectively. For each of the 200 brands we include in this analysis, we evaluate the predictive accuracy of each model considered using the Area Under the Curve (AUC) metric.

At the predictive modeling (training and test) stage, we consider a range of methods including logistic regression and other machine learning methods such as tree-based models.

Given the similar performance that we observe across the different methods, we focus our discussion on the results derived using logistic regression.[1] We assess accuracy based on the ability to predict whether or not an individual visits specific brands during the forecasting period based on the average brand visitation data from the cluster to which the device belongs. We first do so using the clusters derived based solely on individuals' brand visitation behavior during the forming stage. We then evaluate the extent to which predictive performance is improved by including devices' home locations in forming the clusters, evaluating different weights placed on the brand visitation and home location data at the forming stage. Finally, we examine how predictive performance is affected by replacing each device's home location with the pseudo-home location. Comparing the predictive performance resulting from the use of home and pseudo-home locations, we probe the accuracy with which an individual's home location can be identified from their pseudo-home location, providing an assessment of the reidentification threat posed by our approach.

In doing so, we provide guidance to data providers and their clients as to the level of cluster resolution necessary to derive marketing insights from location data. Combined with our assessment of the ability to identify individuals' true home locations, we offer guidance to regulators concerned about the extent to which marketing efforts erode consumer privacy.

Analysis and Results

We begin by discussing the results when brand choices are the basis for aggregating mobile devices and degree of privacy z granted by a partitioning of devices into homogeneous clusters is varied. We let z take on values between 1 and N such that z is a divisor of N, where N

[1] We present the full set of estimated results in the Web Appendix.

is the number of devices in the sample. Maintaining privacy requires aggregating a larger number of mobile devices into a smaller number of more heterogeneous groups. That is, as *(z-1)-anonymity* increases, the number of clusters *K* diminishes. At the same time, however, one may expect that the use of a larger value of *z* results in poorer predictive power for future behavior. Firms must therefore decide if providing consumers with increased privacy can justify the inferior predictive performance. It remains an empirical question if the relationship between predictive accuracy and the number of segments into which devices are partitioned exhibit a positive, monotonic relationship, or if there is a point at which the addition of more segments adversely affects predictive performance.

We draw a random sample of N=18,000 devices from our full sample described earlier, and infer their home location as the most common nighttime (8PM-4AM) location throughout the forming period. We then define the set *B* to include the top 25 brands, and extract the features describing brand visitation behavior (for each week, the number of times the given device was observed at any store of the given brand) for each device in the sample.[2] We iterate $1 \leq z \leq N$, and perform the following computation for each *z* that divides *N*. First, we perform the mini-batch K-means algorithm for $K = N / z$ and minimum group size *z* using the weighted features, and cluster the sampled devices into *K* groups. We then average brand visitation counts within each cluster for each week of the training stage, and, *for each* of the top 200 brands (i.e., including brands outside *B*), we train a binary logit model to predict brand visitations on week *t* (for $t \in \{4, 5, 6\}$), using the averaged brand visitation counts at each of the top 200 brands on week *t-1*. [3] Finally, we evaluate the predictive performance of our models using data from the

[2] We report the number of visitations at each brand in *B* in Table 1 in the Appendix.
[3] In other words, we train 200 binary classifiers – one for each brand – each using the same 200 (continuous) predictors.

test stage (i.e., for $t \in \{8, 9, 10\}$) using the Area Under the Curve (AUC) score, and average the results across the top 200 brands to obtain the predictive performance for different values of z, which reflected by the solid line presented in Figure 1.

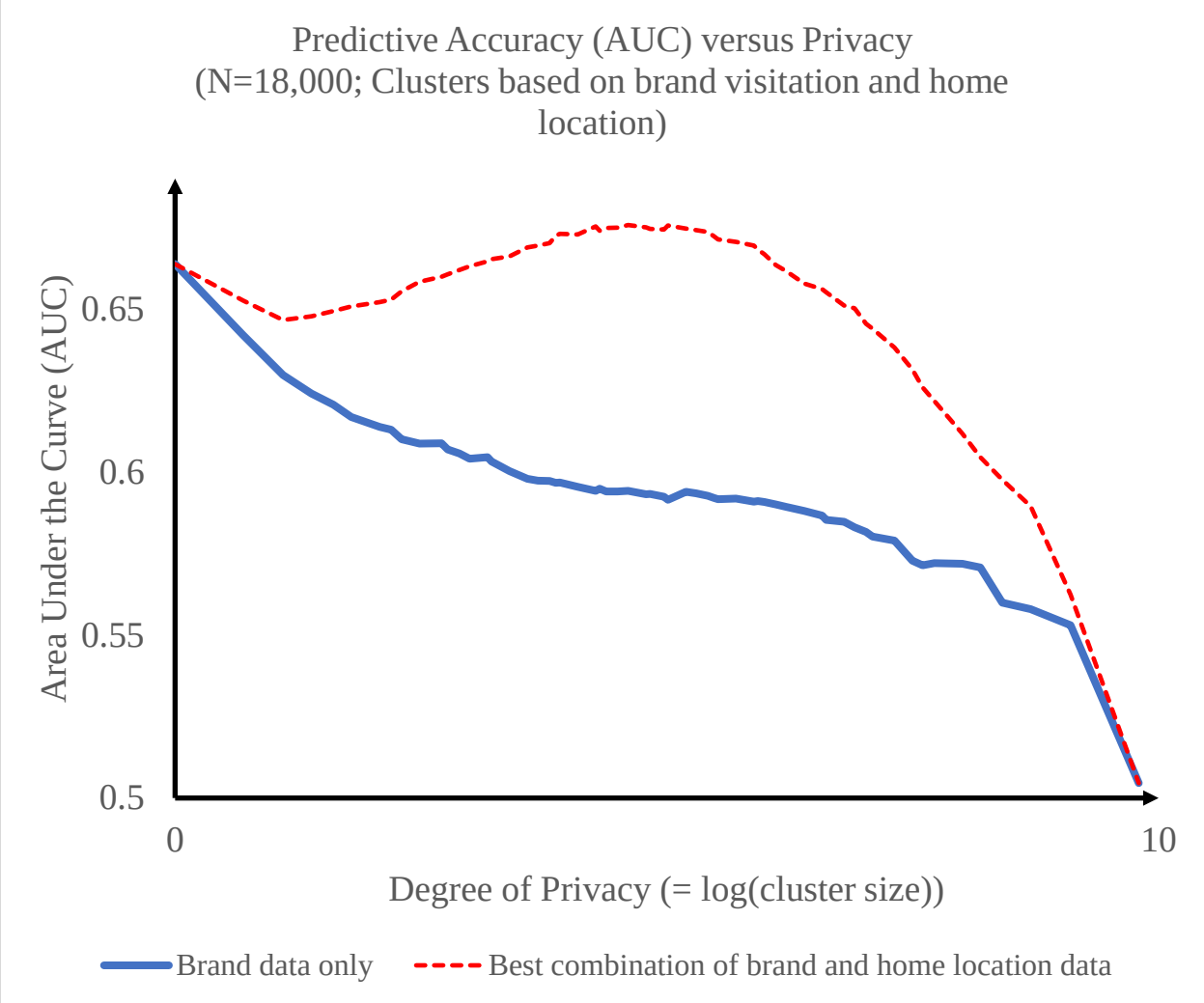

Figure 1 - How incorporating home location affects the privacy - predictive accuracy trade-off.

As Figure 1 reveals, predictive performance deteriorates as cluster size increases, which provides consumers with increased privacy. Relying entirely on brand visitation data, this suggests that marketing insights and consumer privacy are fundamentally at odds with each other.

Incorporating Home Location

While historic brand visits provide insight into future brand visits, we next investigate the extent to which incorporating location data at the forming stage improves the predictive ability of our models. For each value of $w \in \{0, .05, .1, \ldots, 1\}$, we average the brand visitation counts for the two weeks of the forming period, and weight them by w, while we weight the longitude and latitude coordinates of the inferred home location by $(1-w)$. For any given value of $1 < z < N$, we explore the range of $w \in \{0, .05, .1, \ldots, 1\}$ to find the optimal relative weights for the two types of input data.[4] While $w = 1$ puts full weight on the brand visitation data, assuming $w=0$ puts full weight on the home location and ignores the brand visitation data.

For each value of z, we calculate the maximum AUC under values of $w \in \{0, .05, .1, \ldots, 1\}$. The dashed line in Figure 1 plots the resulting AUC. The incorporation of home location data, as illustrated in Figure 1, reveals two key insights. First, for non-trivial values of z (i.e., $1 < z < N$), the incorporation of home location data for cluster formation indeed provides superior predictive performance compared to the model that relies only on brand visitation data. Second, in contrast to the AUC from the model that uses only brand visitation data, for a value of z such that $3 \leq z \leq 100$, the AUC from the model that combines brand visitation and home location data is not monotonically decreasing in z. When device clusters are formed using a combination of brand visitation data and home location data, there is a range of z values ($24 \leq z \leq 400$) capable of both allowing for privacy through aggregation and offering superior predictive performance compared to the brand visitation model that makes use of device-level data.

"Where You Shop" vs. "Where You Sleep"

[4] For values of $z=1$ and $z=N$, the clustering is trivial.

The above analysis demonstrates that it is possible to achieve superior predictive performance while using homogeneous clusters rather than device-level data. One significant limitation of the analysis, however, is its reliance on home location. Though aggregating devices into homogeneous clusters affords consumers some degree of privacy, the very use of home location data poses a threat to consumer privacy: home location data must be collected for it to be used as a basis on which to form the homogeneous clusters. Whereas a firm's collection of brand visitation data may not reveal the identity of individuals, an inference of home location could be made from mobile location data to identify the device owner (e.g., Macha et al. 2021).

To alleviate such concerns, we propose an additional means to preserving privacy. In this approach, we use the same procedure described above, but instead of inferring the most common nighttime location for each device, we derive a "pseudo-home" location that we define as the centroid of all identified points of interest (POI; e.g., branded locations, parks and other identifiable non-residential locations) at which the device was detected, weighted by visitation frequency. In contrast to the inferred home location that is based on "where you sleep," we base the pseudo-home location on "where you shop." Though perhaps subtle, it is a key distinction. Whereas clustering based on the inferred home location will reflect proximity among individuals who reside in the same neighborhood, the use of the pseudo-home location as a basis for clustering will group individuals who frequent nearby brick-and-mortar locations into the same cluster. By relying on detection at commercial locations rather than the likely home location, we hope to capture the same information contained in home locations such as the convenience and affinity for shopping at particular brands, while avoiding potential concerns regarding identifying individual device owners.

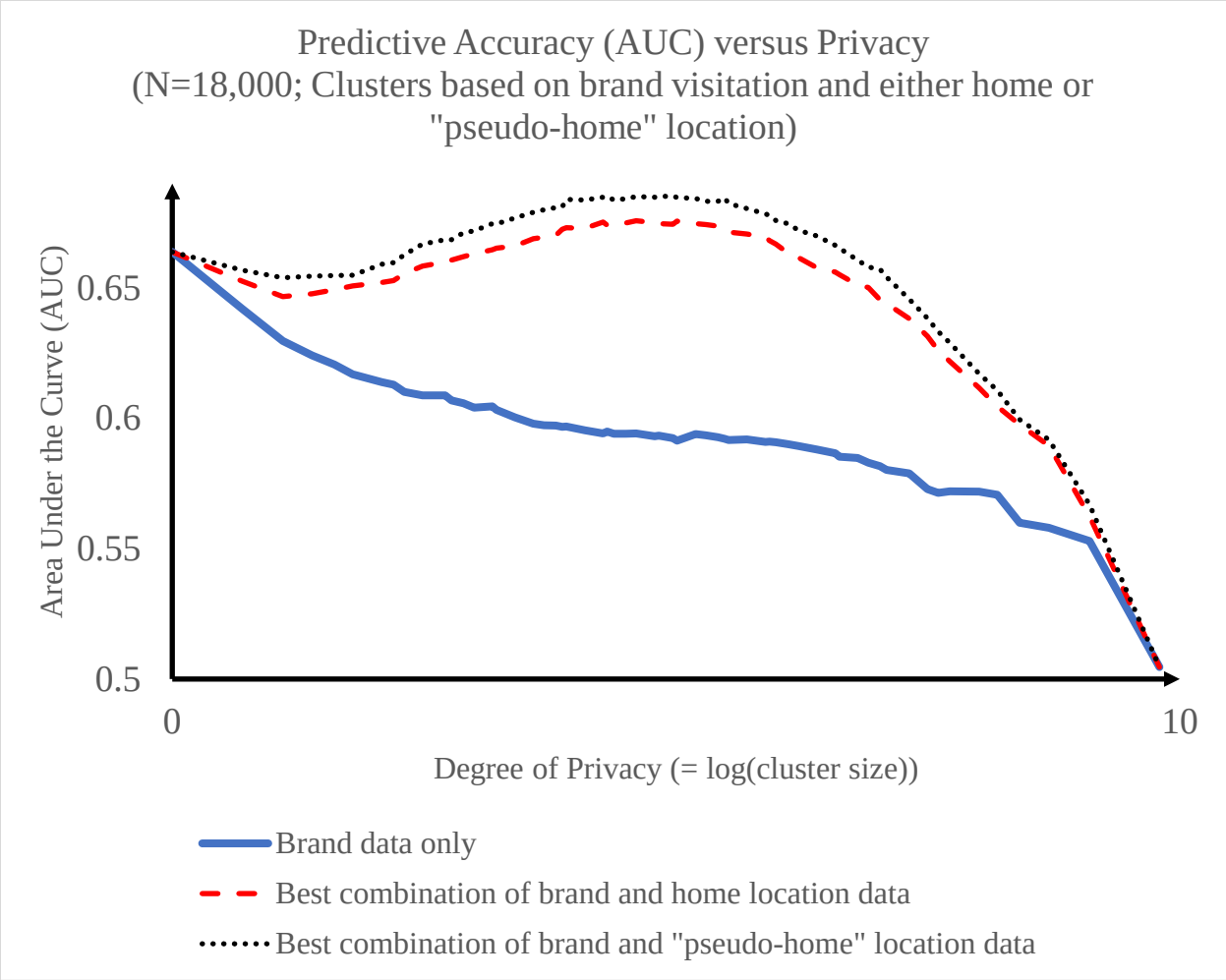

Figure 2 - Clusters based on "pseudo-home" data improve predictive accuracy irrespective of the clustering resolution.

Figure 2 illustrates the average AUC values for the analysis wherein clusters are formed based on brand visitations and the "pseudo-home" location of each device. Importantly, while the AUC for different values of *z* closely resemble those obtained using the actual home location of individuals, we find that clustering devices based on their pseudo-home location and brand visitation counts improves the predictive performance of our model for all $1 < z < N$. Our analysis thereby demonstrates that it is not necessary to collect a device's home location in order to achieve the highest possible predictive accuracy. By relying on data collected from commercial locations enables us cluster devices in such a way as to capture homogeneity in

terms of both geographic proximity and brand preferences, we can forego the collection and storage of home location data, further alleviating privacy concerns.

To further underscore the privacy preserving nature of the use of pseudo-home, we evaluate the accuracy with which a device's home location can be identified from the pseudo-home location. In Figure 3, we illustrate the distribution of distances between the inferred real and the "pseudo" home location of each device.

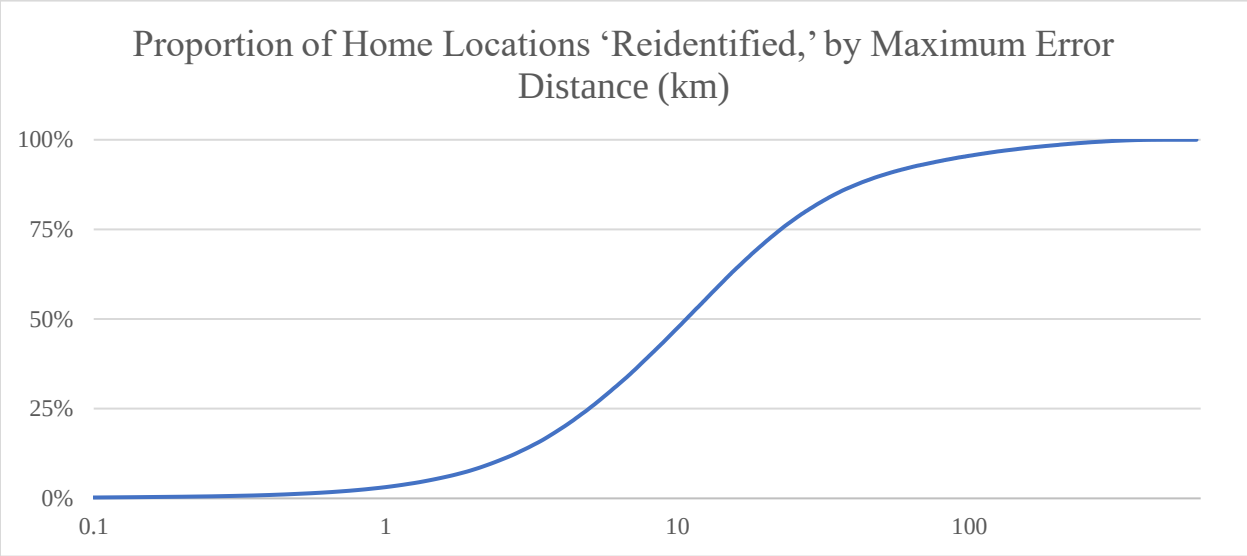

Figure 3 - Cumulative distribution of the distance between real and pseudo-home locations.

The median distance across all 180 thousand devices is 10.71km (6.65 miles), far higher than the predictive accuracy measured by Macha et al. (2021), who concluded that, using unobfuscated mobility data, "an individual's home address can be accurately predicted within an average radius of 2.5 miles." This result suggests that our design, which only considers brand visitation data and the latitude and longitude coordinates of where one shops, can provide a greater degree of consumer privacy than a design that uses raw mobile location data (Ghose et al. 2019; Macha et al. 2021).

DISCUSSION

Consumer mobility data collected from smartphone devices has created tremendous value through enabling marketers to target their promotional messages at greater precision. However, as location data may reveal not only brand preferences but also sensitive information about consumers, there have been increasing concerns regarding the extent to which collecting and storing dynamic location data may invade consumer privacy.

In this research, we investigate whether marketers' desires for actionable insights and consumers' need for greater privacy can both be simultaneously accommodated. To this end, we study the common setting in which a location data platform mediates between advertisers and consumers. We propose that location data platforms consider increasing consumer privacy by clustering devices into cohorts, and sharing only cohort-level average features (including the estimated likelihood of a consumer visiting the offline retail stores of a given brand within a week).

Our results suggest that marketers may not have to choose between predictive insights and consumer privacy: By clustering devices into cohorts based on similarities between where they shop (including the number of times they visit each brand plus the geographic centroid of all their brand visits), advertisers gain predictive accuracy relative to using device-level data and/or the inferred location of consumers' home.

Moreover, we demonstrate that home location, which consumers may have a reluctance to share, is superfluous in our empirical application. Restricting location data collection to commercial locations could be one means by which consumer privacy could be protected for a large number of consumers. That is, relying on data from non-residential locations could provide

a means by which the collection and use of location data can accommodate the objectives of marketers with compromising consumer privacy to an unacceptable degree. Future research could investigate the feasibility of such approaches, including identifying the granularity of data the data that must be collected from a mobile device and the data that need not leave the mobile device (Liu et al. 2019). We hope that these insights will inform the behavior of regulators as well as all players in the mobile location data industry alike.

REFERENCES

Alessandretti, Laura, Piotr Sapiezynski, Vedran Sekara, Sune Lehmann, and Andrea Baronchelli (2018), "Evidence for a conserved quantity in human mobility," *Nature Human Behaviour*, 2 (7), 485-491.

Andrews, Michelle, Xueming Luo, Zheng Fang, and Anindya Ghose (2016), "Mobile ad effectiveness: Hyper-contextual targeting with crowdedness," *Marketing Science*, 35 (2), 218-233.

Bengio, Yoshua, Richard Janda, Yun William Yu, Daphne Ippolito, Max Jarvie, Dan Pilat, Brooke Struck, Sekoul Krastev, and Abhinav Sharma (2020), "The need for privacy with public digital contact tracing during the COVID-19 pandemic," *The Lancet Digital Health*. 2 (7), e342-e344.

Bettini Claudio, X. Sean Wang, Sushil Jajodia (2005), "Protecting Privacy Against Location-Based Personal Identification." In: Jonker W., Petković M. (eds) Secure Data Management. SDM 2005. Lecture Notes in Computer Science, vol 3674, 185-199. Springer, Berlin, Heidelberg.

Bleier, Alexander, Avi Goldfarb and Catherine Tucker (2020), "Consumer privacy and the future of data-based innovation and marketing," *International Journal of Research in Marketing*, 37 (3), 466-480.

Bohn, Dieter (2021), "Privacy and Ads in Chrome Are About to Become Flocing Complicated," *The Verge*, March 30, accessed at https://www.theverge.com/2021/3/30/22358287/privacy-ads-google-chrome-floc-cookies-cookiepocalypse-finger-printing.

Bradley, Paul S., Kristin P. Bennett, and Ayhan Demiriz (2000) "Constrained k-means clustering," *Microsoft Research, Redmond*, 20(0), 0.

Brzezinski, Adam, Valentin Kecht and David Van Dijcke (2020), "The Cost of Staying Open: Voluntary Social Distancing and Lockdowns in the US," Economics Series Working Papers 910, University of Oxford, Department of Economics., Available at SSRN: https://ssrn.com/abstract=3614494.

Chiou, Lesley, and Catherine Tucker (2020), "Social distancing, internet access and inequality," No. w26982, National Bureau of Economic Research.

De Montjoye, Yves-Alexandre, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel (2013), "Unique in the crowd: The privacy bounds of human mobility," *Scientific Reports*, 3 (1), 1-5.

Ehrenberg, Andrew SC, Gerald J. Goodhardt, and T. Patrick Barwise (1990), "Double jeopardy revisited," *Journal of Marketing*, 54 (3), 82-91.

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, No. 34, pp. 226-231).

Fong, Nathan M., Zheng Fang, and Xueming Luo (2015), "Geo-conquesting: Competitive locational targeting of mobile promotions," *Journal of Marketing Research*, 52 (5), 726-735.

Ghose, Anindya, Beibei Li, and Siyuan Liu (2019), "Mobile targeting using customer trajectory patterns," *Management Science*, 65 (11), 5027-5049.

Goldfarb, Avi, and Catherine Tucker (2012), "Shifts in privacy concerns," *American Economic Review*, 102 (3), 349-53.

Goldfarb, A., & Tucker, C. (2020). *Which retail outlets generate the most physical interactions?* NBER Working Paper 27042.

Laziuk, Estelle (2021), "Daily iOS 14.5 Opt-In Rate," Flurry, April 29, https://www.flurry.com/blog/ios-14-5-opt-in-rate-att-restricted-app-tracking-transparency-worldwide-us-daily-latest-update/, accessed on May 13, 2021.

Li, L., & Xi, Y. (2011, October). Research on clustering algorithm and its parallelization strategy. In *2011 International conference on computational and information sciences* (pp. 325-328). IEEE.

Liu, Meng, Raphael Thomadsen and Song Yao (2020), "Forecasting the spread of COVID-19 under different reopening strategies," *Scientific Reports*, 10 (20367).

Liu, Yi-Ning, Yan-Ping Wang, Xiao-Fen Wang, Zhe Xia, and Jing-Fang Xu (2019), "Privacy-preserving raw data collection without a trusted authority for IoT," *Computer Networks*, 148, 340-348.

Macha, Meghanath, Beibei Li, Natasha Zhang Foutz and Anindya Ghose (2021), "Perils of Location Tracking? Personalized and Interpretable Privacy Preservation in Consumer Mobile Trajectories," working paper, accessed at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3628640.

Mayer, Charles S., and Charles H. White, Jr (1969), "The law of privacy and marketing research," *Journal of Marketing*, 33 (2), 1-4.

McPhee, William N. (1963), *Formal Theories of Mass Behavior*. New York: The Free Press.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook (2001), "Birds of a feather: Homophily in social networks," *Annual Review of Sociology*, 27 (1), 415-444.

Rafieian, Omid, and Hema Yoganarasimhan (2021), "Targeting and privacy in mobile advertising," *Marketing Science*, 40 (2), 193-218.

Schneider, Matthew J., Sharan Jagpal, Sachin Gupta, Shaobo Li, and Yan Yu (2017a), "Protecting customer privacy when marketing with second-party data," *International Journal of Research in Marketing*, 34 (3), 593-603.

Schneider, Matthew J., Sharan Jagpal, Sachin Gupta, Shaobo Li, and Yan Yu (2018), "A flexible method for protecting marketing data: An application to point-of-sale data," *Marketing Science*, 37 (1), 153-171.

Snider, Mike (2021), "Apple's privacy changes to iOS have arrive. What do you do with your Facebook app?" *USA Today*, April 26, accessed at https://www.usatoday.com/story/tech/2021/04/26/facebook-apple-iphone-ipad-privacy-ios-software-update/7368248002/.

Sweeney, Latanya (2002), "k-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10 (05), 557-570.

Thompson, Stuart A. and Charlie Warzel (2019), "How to Track President Trump," *The New York Times*, December 20, accessed at https://www.nytimes.com/interactive/2019/12/20/opinion/location-data-national-security.html.

Tucker, Catherine E. (2014), "Social networks, personalized advertising, and privacy controls," *Journal of Marketing Research*, 51 (5), 546-562.

Warzel, Charlie and Stuart A. Thompson (2021), "They Stormed the Capitol. Their Apps Tracked Them," *New York Times*, February 5, accessed at https://www.nytimes.com/2021/02/05/opinion/capitol-attack-cellphone-data.html.

Valentino-DeVries, Jennifer, Natasha Singer, Michael E. Keller and Aaron Krolik (2018), "Your Apps Know Where You Were Last Night, and They're Not Keeping it Secret," *New York Times*, December 2018, accessed at https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html.

APPENDIX: TOP BRANDS IN THE DATA

Brand	Number of visitations
McDonald's	459,784
Walmart	294,109
Chick-fil-A	275,057
Subway	256,628
Kroger	206,025
SmartStyle	205,607
Starbucks	171,040
Circle K	165,079
Waffle House	164,716
QuikTrip	162,166
CVS	147,314
Wendy's	145,122
Home Depot	142,729
Jackson Hewitt Tax Service	133,983
Burger King	128,147
Zaxby's	123,117
Publix	117,758
Dollar General	108,254
Walgreens	103,686
Chevron	102,368
RaceTrac	100,310
H&R Block	93,314
Taco Bell	92,611
Suntrust	90,381
Shell	81,809

Table 1: Number of visitations at the top 25 brands.