scientific reports

Check for updates

OPEN

Utilizing large-scale human mobility data to identify determinants of physical activity

Giorgos Ioannou[1], George Pallis[1], Marios Dikaiakos[1] & Christos Nicolaides[2]✉

Analyzing the habits of exercisers is crucial for developing targeted interventions that can effectively promote long-term physical activity behavior. While much of existing literature has focused on individual-level factors, there is a growing recognition of the importance of examining how broader determinants impact physical activity. In this study, we analyze large-scale human mobility data from over 20 million individuals to investigate how visits to various locations, such as cafes and restaurants, influence visits to fitness centers. In particular, we (i) rank categories of locations that exercisers prefer to visit, (ii) compare visiting patterns between individuals who visit fitness centers and those who do not, (iii) investigate how exercisers replace fitness visits on non-exercise days, and (iv) identify location categories mainly visited before or after fitness sessions. We show that individuals engaging in physical exercise prefer to visit "Non-Alcoholic Beverage Bars" (e.g., Starbucks) in conjunction with their exercise sessions. On their rest days, they often substitute exercise with visits to full-service restaurants and parks. Moreover, they tend to visit grocery stores immediately after their exercise session. Our findings can help public health policy towards a more targeted promotion of exercise and well-being.

Keywords Exercise habits, Human mobility, GPS devices, Social networks

Identifying the factors and underlying causes that influence participation in physical activity is crucial for evidence-based planning of public health interventions[1,2]. Several interventions are deemed insignificant, in light of the incentives provided not being meaningful to exercisers[3–5]. For example, a recent study testing 53 different interventions found that the most effective intervention increased the weekly fitness visits by just 0.403 visits on average[6]. Predictions made from recruited participants, professors of public health and practitioners in applied behavioural science gave a 9.1 times more optimistic results than what was actually observed. Another limitation of many studied interventions is their lack of generalizability due to small sample sizes[7] and potential sampling biases depending on the participants[6].

In this research, we expand current knowledge regarding the correlates of physical activity, with the end goal of optimizing future intervention design. We provide novel insights into how behavioral cues, related to the types of locations people visit, affect their exercise behavior. Specifically, we analyze the visits of more than 20 million individuals from the U.S. to various location categories, such as cafes and restaurants, to uncover how these visits influence the decision to visit fitness centers and engage in physical activity. Analyzing which location categories are frequently visited on exercise days, compared to those mainly visited on non-exercise days, reveal trends that highlight exercisers' priorities, habits, and potential barriers to maintaining consistent exercise routines. These insights can inform the development of targeted interventions aimed at encouraging more regular physical activity by addressing barriers or promoting routines that seamlessly integrate exercise into daily life. Given that most exercisers drop out within 6 months[8], tailored interventions can help both newcomers and current exercisers build and sustain long-term exercise habits.

Our results are derived from patterns observed from millions of individuals within our networked society. We do not merely provide results, but also a comprehensive methodology we developed, which can be applied to similar datasets from different countries to extract habitual patterns shared among exercisers.

We propose the relevant research questions we address, along with our contributions for tackling them.

1. "How do visits to different location categories impact same-day fitness visits?" We propose an Ordinary Least Squares (OLS) regression framework which quantifies how visits to various location categories influence same-day fitness visits. We also account for potential biases due to users' U.S. state and the day of the week.

[1]Department of Computer Science, University of Cyprus, 2109 Aglantzia, Nicosia, Cyprus. [2]School of Economics and Management, University of Cyprus, 2109 Aglantzia, Nicosia, Cyprus. ✉email: nicolaides.christos@ucy.ac.cy

2. "How do the visiting patterns of both exercisers and non-exercisers differ on days when neither group exercised?" We identify fitness center visitors over the course of a week and calculate weekly descriptive statistics for their visits to different categories on their rest days, comparing them with corresponding statistics from non-exercisers.
3. "How do exercisers substitute their workouts on rest days?" We group fitness center visitors by their number of consecutive-day visits and quantify how each group substitutes their fitness visits on a non-exercise day.
4. "How important is exercise for visitors of other categories?" We construct importance networks to measure how various location categories value exercise, while avoiding bias toward frequently visited categories.
5. "Do exercisers visit specific categories before or after their exercise session?" We analyze the visits of users to determine what categories they visited before and after their fitness center visits, and aggregate the results to obtain overall preferences for when they visit different categories.

Related work

Research on the correlates of healthy behaviors enhance the development of more effective interventions[9]. Several studies have explored the factors influencing user engagement in physical activity, aiming to refine strategies for future interventions.

To gain deeper insights into the factors influencing physical activity, researchers have hugely relied on recruiting participants and conducting population-based surveys. These studies aim to identify key demographic, behavioural, and environmental determinants. Some common findings showed that age is negatively associated with physical activity, self-efficacy is a key correlate, and the neighbourhood environment is linked to the level of physical activity [2,10]. Moreover, the prevalence of insufficient physical activity in high-income countries is more than twice that in low-income countries[11]. However, most of these studies rely on self-reported data, which, while cost-effective, are prone to recall errors and social desirability bias[12]. These limitations suggested a need for more objective and accurate data collection methods, such as wearable technology or mobile-based tracking, to improve the reliability of future research.

The emergence of mobile applications and wearable devices enabled a variety of new methods to study the behaviors of exercisers. Janssen et al.[13] identified consumer profiles of runners concerning their utilization of running apps, sports watches, and demographic characteristics, aiming to refine exercise targeting through personalized approaches. Moreover, mobile applications played a pivotal role for identifying that physical activity is socially contagious. Social contagion is a ubiquitous process through which information and behaviors spread among individuals through their social contacts[14]. For example, individuals are more likely to gain weight if their friends do[15], and they are also more likely to sleep less when their friends sleep less[16]. Application notifications received when a friend completed a run caused peers to run more, even when they were physically distant[17]. Althoff et al.[18] analyzed the number of steps from over 700,000 people across 111 countries, and found that inequality in activity distribution within countries predicts obesity prevalence better than average activity levels. Pontin et al.[19] found that apps incentivizing physical activity are more likely to be used from populations in areas of lower socioeconomic status. Mejova and Kalimeri[20] analyzed self-reported data and mobile app usage patterns from 15k users. Their analysis established robust associations between physical activity and various demographic groups, and identified that exercisers are strongly linked to the value of happiness and stimulation. However, they acknowledged a limitation in their work related to self-reports of physical activity behaviors and encouraged future researchers to explore unobtrusive methods for identifying exercisers, to enable more sophisticated interventions.

Human mobility data holds significant potential for analyzing the habitual patterns of millions of exercisers to identify their needs and enable the formation of optimized interventions for long-term results. In recent years, their utilization experienced a notable increase with the onset of COVID-19, as researchers employed them to examine the creation and influence of effective government regulations[21–24]. They were also utilized for designing dynamic intervention policies in financial networks[25,26], and in identifying relationships between different locations for accurately predicting socioeconomic indicators[27]. Consequently, human mobility data has informed targeted interventions in various fields, mitigating sampling biases inherent in studies with limited participants. To the best of our knowledge, this is the first study that utilizes human mobility data for guiding the promotion of physical activity.

Dataset

We leverage the "Visits" dataset by Veraset[28], which was collected in the U.S. It comprises records from over 20 million anonymised devices, recording the places individuals visited at specific dates and times. It covers a wide range of 4 million points of interest. The anonymized devices opted-in to provide access to their location through applications and Software Development Kits (SDKs) with Veraset. Previous studies utilized the dataset to assess the impacts of COVID-19 and provide valuable insights for shaping public policies[29–31]. The human mobility datasets provided to aid in the fight against COVID-19 were shown to be unbiased, with the sampling highly correlated to the true census populations[32].

Each record of the dataset is associated with an *identifier* of the user; a *timestamp* recorded at the beginning of the visit; the *location name* visited; the *top category* and the *subcategory* in which that location belongs (e.g. McDonald's, top category: "Restaurants and Other Eating Places", subcategory: "Limited-Service Restaurants"); the *minimum duration* of the visit and the *U.S. state* of the visit. In our analyses, we utilize the local time of each user to divide the days instead of the universal UTC timestamp that the original dataset does. Our objective is to explore the relationship between different categories of locations and physical activity on the same day for individuals users, rather than using a uniform time reference for all users.

Date	Subcategory	#Visits	#Visitors	Median visit time	#Users
	Full-Service Restaurants	3,479,657	2,697,295	31 min	
Mon, Oct 7, 2019	Limited-Service Restaurants	1,942,840	1,645,506	12 min	18,644,818
	Fitness and Recreational Sports Centers	1,524,812	1,284,867	44 min	
	Full-Service Restaurants	3,984,960	3,016,637	36 min	
Thu, Dec 5, 2019	Limited-Service Restaurants	2,438,376	2,036,696	12 min	17,737,402
	Merchandise Stores	1,660,269	1,403,844	20 min	
	Full-Service Restaurants	3,644,729	2,820,750	34 min	
Mon, Feb 24, 2020	Limited-Service Restaurants	2,335,340	1,956,438	13 min	20,162,495
	Fitness and Recreational Sports Centers	1,613,587	1,350,410	47 min	

Table 1. The most visited location subcategories on 3 example days, and some statistics for those.

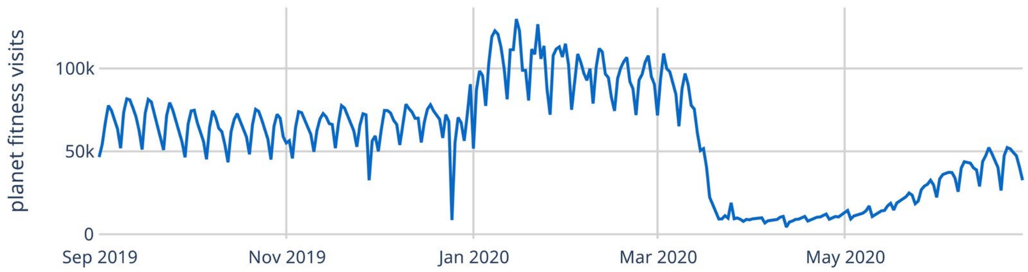

Fig. 1. Daily visits to planet fitness, which is the most visited fitness chain in the U.S. according to human mobility data from September 2019 to June 2020.

The top category of fitness locations is “Other Amusement and Recreation Industries”, which is broad and captures playgrounds, which adults visit them without necessarily engaging in physical activity. So, we focus on the subcategory of that top category, which is the “Fitness and Recreational Sports Centers”. On a typical week day, approximately 6% of the total visits are to a fitness center. Table 1 shows aggregated statistics we have calculated for the most visited subcategories on 3 example days. The number of visitors is less than the number of visits, because some users visited a particular subcategory multiple times in a single day. Supplementary Table 1 shows aggregated statistics we have calculated about the most visited locations that lie within the subcategory “Fitness and Recreational Sports Center” on 3 example days, which are consistently “Planet Fitness”, “CrossFit” and “Anytime Fitness”. We repeated each of our subsequent analyses using data from one week per month, spanning September 2019 to February 2020, to ensure robust and reliable outcomes and to assess potential seasonal differences.

Results

Seasonal trends in fitness center attendance

Figure 1 shows the number of daily visits to Planet Fitness, the most visited U.S. fitness chain, over a 9-month period. Consistent with previous research, the local minima that are visits on Sundays indicate reduced activity on weekends[33]. The two global minima in 2019 correspond to Thanksgiving (November 28) and Christmas, while the onset of COVID-19 diminished the visits from early March.

From the beginning of January, there is a noticeable upward trend in visits, aligning with the most popular New Year’s Resolutions that are related to physical health (33%) followed by weight loss (20%)[34]. This pattern emerges despite these months being colder compared to autumn, contrasting with prior studies suggesting higher physical activity during months with more daylight[19,33,35]. However, it is important to note that previous research primarily tracked outdoor activities like steps or running, while our focus is on fitness center visits, representing indoor exercise. Some outdoor exercisers may choose to move their activities indoors during colder weather. As time progresses, the January rise gradually declines, highlighting that while many individuals intend to engage in physical activity, they struggle to maintain the habit over time.

Complementarities of physical exercise

This section addresses research question 1, quantifying the influence of a visit to a specific location subcategory on the likelihood of a same-day fitness visit through the use of OLS regression analysis (see Methods).

Table 2 shows the results of three OLS models for a week in December. For validation, we present results from one week per month spanning September 2019 to February 2020 (Supplementary Tables 2–6). The nearly identical ordering of subcategories across these periods indicates that the relationship between visiting specific categories and fitness centers on the same day is consistent and unaffected by seasonality.

The OLS regression provides a quantitative measurement of the impact of a single visit to a subcategory S on a same-day visit to a fitness facility, while accounting for potential biases related to the day of the week and the

	Monday–Tuesday*	Wednesday–Thursday*	Friday–Sunday*
	Dec 2–3, 2019	Dec 4–5, 2019	Dec 6–8, 2019
Const	0.2136	0.2009	0.1720
Department Stores	− 0.0285	− 0.0227	− 0.0088
Sporting Goods Stores	− 0.0348	− 0.0292	− 0.0103
Snack and Nonalcoholic Beverage Bars	− 0.0367	− 0.0306	− 0.0103
Pet and Pet Supplies Stores	− 0.0347	− 0.0298	− 0.0163
Convenience Stores	− 0.0427	− 0.0383	− 0.0235
Used Merchandise Stores	− 0.0431	− 0.0393	− 0.0224
Pharmacies	− 0.0499	− 0.0437	− 0.0248
Commercial Banking	− 0.0494	− 0.0474	− 0.0328
Limited-Service Restaurants	− 0.0596	− 0.0528	− 0.0310
Gas Stations	− 0.0573	− 0.0529	− 0.0341
Hardware Stores	− 0.0563	− 0.0530	− 0.0399
Automotive Parts and Accessories Stores	− 0.0563	− 0.0535	− 0.0397
Grocery Stores	− 0.0637	− 0.0591	− 0.0420
Full-Service Restaurants	− 0.0716	− 0.0659	− 0.0430
Merchandise Stores	− 0.0715	− 0.0653	− 0.0472
Hotels and Motels	− 0.0801	− 0.0716	− 0.0397
Religious Organizations	− 0.0848	− 0.0820	− 0.0632
Colleges and Universities	− 0.0882	− 0.0830	− 0.0599
Nature Parks and Other Similar Institutions	− 0.0951	− 0.0865	− 0.0561

Table 2. OLS regression results. The dependent variable is whether a user visited a fitness location on a particular day. In all models, the day of the visit as well as the state of the user were used as fixed effects, to avoid any variable bias. The subcategories are ordered in descending order based on their average coefficient obtained from all three models. *All values significant at the 1% level

user's U.S. state of residence. The coefficients derived show the impact of an additional visit to a subcategory on fitness visits. For example, during Monday and Tuesday, someone who has not visited any other category has a probability of 21.36% for visiting a fitness center (the constant). A single visit to a convenience store decreases that probability by 4.27%.

All coefficients except from the constants are negative, because each visit to any subcategory decreases the probability of an individual visiting a fitness center as well. This occurs because spending time elsewhere during the day leaves less time available for a visit to a fitness center. In fact, on Tuesday, December 3rd, 2019, from the 1,147,929 individuals who visited a fitness location, just 350,208 visited an additional subcategory. What is important for a valid comparison between different subcategories is to identify the least negative coefficients. The ordering of subcategories remains nearly consistent across the week, indicating that user behaviour regarding same-day visits to fitness locations does not significantly depend on specific days.

The subcategory which has the smallest impact on reducing the likelihood of a same-day visit to fitness locations is the "Department Stores", with only a $\sim$2% decrease in the same-day visit percentage. Other subcategories which individuals tend to combine their fitness visits with (small decrease) are the "Sporting Goods Stores", the "Snack and Nonalcoholic Beverage Bars" (e.g. Starbucks), and the "Pet and Pet Supplies Stores".

On the contrary, visiting a location categorized under "Colleges and Universities" is associated with a significant decrease in the probability of visiting a fitness location on the same day (7.70% on average). One factor that could influence this is that visits to such locations are time-consuming. For instance, on Tuesday, December 3rd, 2019, the median visit time of the subcategory was 65 min. Another subcategory which causes a major decrease in the same-day fitness visits is "Nature Parks and Other Similar Institutions" (7.92% on average). If some of these people engage in physical activities in parks, they may not need to visit fitness locations.

Across all six weeks analyzed, the coefficient of the constants for Fridays and weekends are smaller than those for Mondays-Thursdays, reflecting fewer average visits to fitness centers. For example, if no other subcategories were visited, the probability of visiting a fitness center was 0.2136 on Monday-Tuesday (Dec 2–3, 2019) but dropped to 0.1720 on Friday-Sunday (Dec 6–8, 2019). Additionally, the coefficients for all subcategories are smaller on Fridays and weekends, as more available time during these days means visiting another location reduces the likelihood of visiting a fitness center less.

In Supplementary Table 14, we present results from OLS models focusing on visit durations rather than visits. These results show that spending more time in all types of stores reduces the time spent in fitness centers. This suggests that while an average visit to a department store does not significantly impact same-day fitness visits, users who spend longer durations at stores have less time available for fitness center visits.

Exercisers vs non-exercisers

In this section, we address research question 2 by analyzing visiting patterns to compare individuals who engage in physical activity with those who do not, on days when neither group exercised. For this analysis, we define an 'exerciser' as someone who visited a fitness location at least once within a week (See Methods).

Figure 2 shows noteworthy results during the week of December 2–8, 2019, from some of the most visited subcategories. The black lines at the top of each bar denote the 99% confidence interval. The narrowness is consistent, owing to the substantial number of visits from each group's population observed over the entire week. The observed patterns in different months are identical, indicating that the visiting behaviours of exercisers and non-exercisers are consistent across different seasons (Supplementary Figs. 1–5).

A notable observation is the opposite outcome between full-service restaurants and limited-service restaurants. Full-service restaurants contain establishments where customers place their orders and are served while seated, and they settle their bill after eating. In limited-service restaurants, customers order and pay before eating. The latter primarily consists of fast-food establishments; on December 3rd, 2019, its most visited locations were McDonald's and Subway. Exercisers show a clear preference for visiting full-service restaurants over non-exercisers, whereas the limited-service restaurants are almost equally visited. Full-service restaurants show an opposite effect when compared to the results of the OLS model. This indicates that exercisers are less inclined to visit a full-service restaurant on their exercise day, but are more willing to do so on their rest day. Exercisers show a preference for visiting "Snack and Nonalcoholic Beverage Bars" compared to non-exercisers. There are no notable differences between the two groups when it comes to "Pharmacies". Non-exercisers dedicate a higher percentage of their visits to "Gas Stations" and "Merchandise Stores".

Substitutions of physical exercise

In this section, we examine how individuals with different number of consecutive days visiting fitness centers substitute their fitness-related visits on a rest day, to address research question 3.

Figure 3 shows the percentage of visits across different subcategories for the three groups, both on days with fitness visits (prior December 5, 2019), and on a day without (December 5, 2019). The groups are created to determine if individuals who exercised for several consecutive days exhibit different patterns of substituting fitness visits compared to those who exercised for fewer consecutive days.

The notable observation is that fitness center visitors substitute their fitness visits with visits to full-service restaurants and parks on their rest day. Individuals who visited fitness centers for at least two consecutive days generally visit parks and full-service restaurants even more. When exercisers rest (not visit a fitness location), visits to limited-service restaurants decrease, while visits to beverage bars show almost no change, indicating a preference to combine these visits with their exercise routines.

For validation, we repeated this analysis across different time periods (Supplementary Figs. 6–11). While the substitution of full-service restaurants remained consistent, the substitution to parks varied significantly with temperature. On colder non-exercise days, no substitution behavior was observed. The clearest example is shifting the data in Fig. 3 by one day, where the temperature dropped by 5°C on average for the non-exercise day, resulting in decreased park visits (Supplementary Fig. 11). On warmer days, park visits increased, suggesting that some individuals substituted fitness center visits with outdoor exercise. This aligns with previous research showing that park visits are higher on warmer days[36].

Importance of physical exercise across categories

In this section, we address research question 4 in identifying the importance that visitors of different categories attribute towards exercise. The importance is determined by addressing the question "Given the users that visited subcategory X, what is their percentage of visits to fitness locations on the same day?"

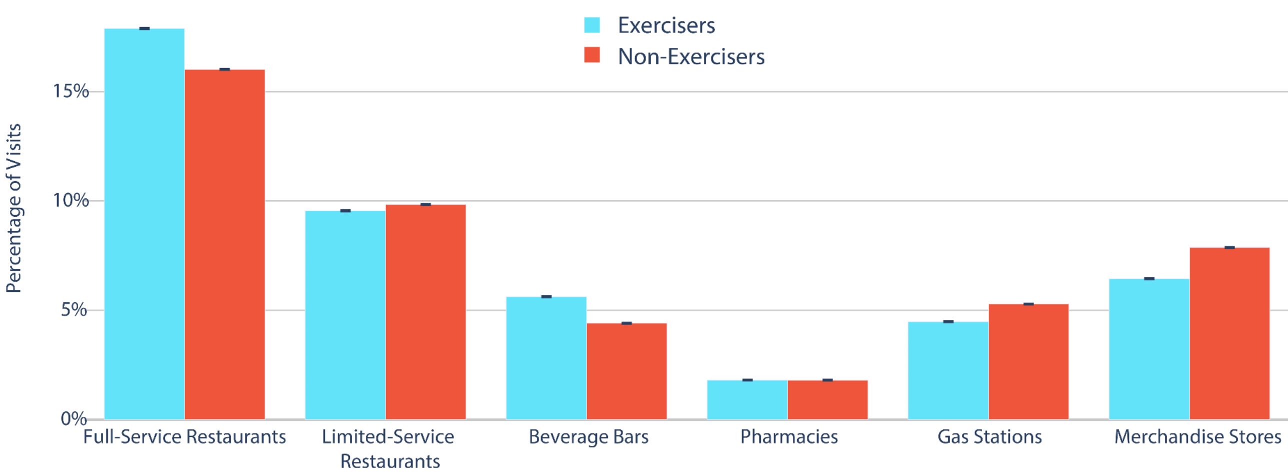

Fig. 2. Comparison of the visiting patterns between exercisers and non-exercisers towards highly-visited subcategories over the course of the week Dec 2–Dec 8, 2019. The black lines on top of each bar denote the 99% confidence interval.

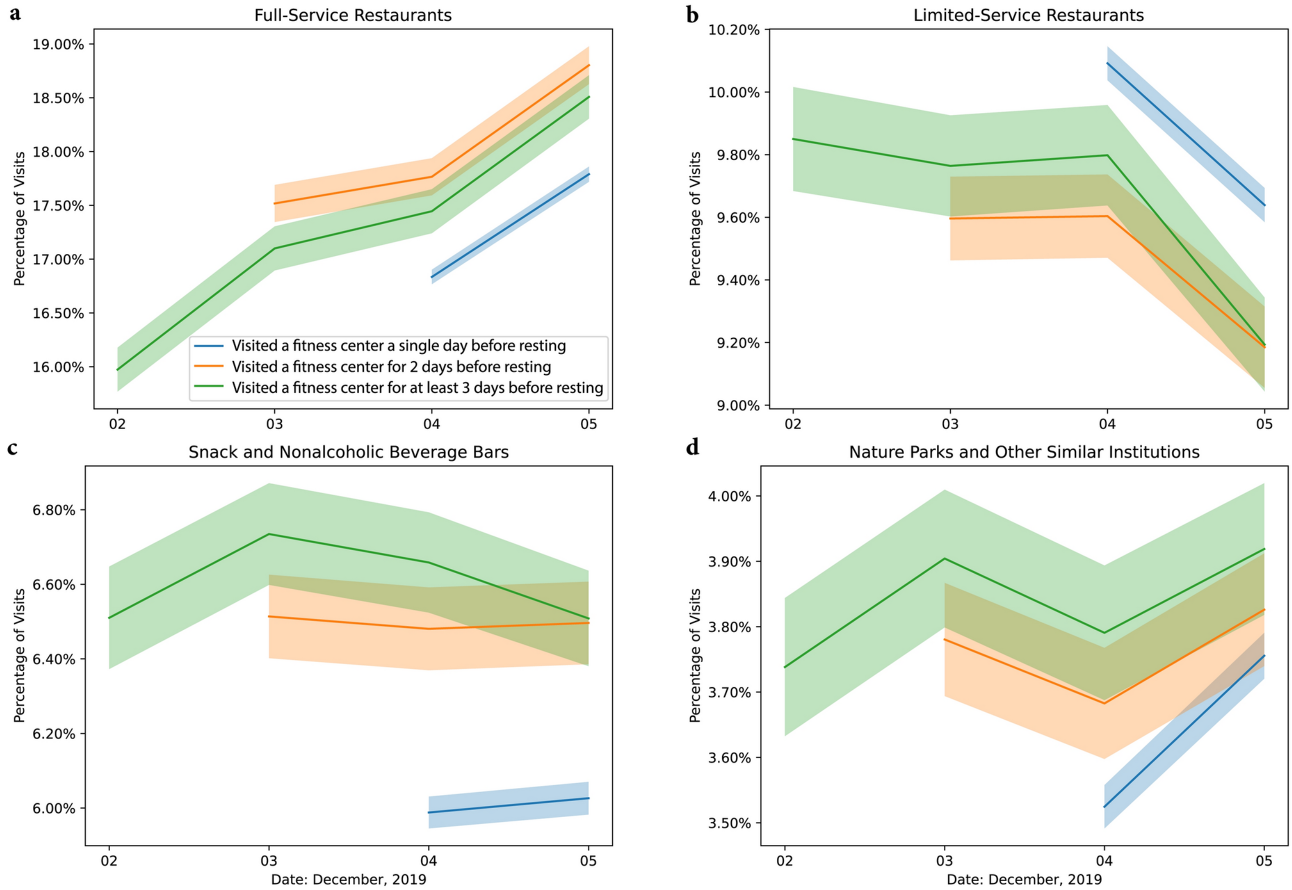

Fig. 3. Percentages of visits to different subcategories for three groups of users that visited fitness centers. The first group visited on December 4, 2019 and have not visited on December 3 and December 5 (blue), the second group visited on December 3 and 4 but not on December 2 and December 5 (orange), and the third group visited on December 2, 3 and 4 but not on December 5 (green). The light colors around the lines display the 95% confidence intervals.

Subcategory	Importance*	95% C.I.*	Ranking*	Importance**	95% C.I.**	Ranking**
Exam Preparation and Tutoring	6.83%	(6.67%; 6.98%)	1	5.69%	(5.42%; 5.95%)	1
Snack and Nonalcoholic Beverage Bars	6.20%	(6.19%; 6.22%)	2	5.38%	(5.36%; 5.40%)	4
Investment Advice	6.10%	(6.00%; 6.19%)	3	5.09%	(4.90%; 5.29%)	10
Full-Service Restaurants	6.09%	(6.09%; 6.10%)	4	5.42%	(5.41%; 5.43%)	3
Musical Instrument and Supplies Stores	6.07%	(6.03%; 6.11%)	5	5.82%	(5.28%; 5.39%)	5
Amusement and Recreation Industries	5.98%	(5.91%; 6.05%)	6	5.32%	(5.22%; 5.41%)	6
Nature Parks and Other Similar Institutions	5.93%	(5.92%; 5.95%)	7	5.14%	(5.11%; 5.16%)	8
Golf Courses and Country Clubs	5.92%	(5.85%; 6.00%)	8	5.12%	(5.03%; 5.22%)	9
Libraries and Archives	5.78%	(5.73%; 5.83%)	9	4.93%	(4.84%; 5.02%)	11
Photography Studios	5.76%	(5.60%; 5.92%)	10	4.84%	(4.62%; 5.07%)	13

Table 3. Location subcategories for which fitness is an important subcategory to visit on the same day. The importance denotes the percentage of co-visits a subcategory attributed towards fitness locations. *Presenting results from weekdays spanning from December 2 to December 6, 2019. **Presenting results from the weekend spanning from December 7 to December 8, 2019

Table 3 shows the subcategories that gave the most importance towards exercise on weekdays (December 2–6, 2019), and the weekend (December 7–8, 2019). In the results, we excluded any subcategory that had fewer than 50,000 co-visits with fitness locations on weekdays and 20,000 on the weekend. For validation, we replicated the method for 5 different weeks (Supplementary Tables 7–11), where there is great similarity among the top-10 entries.

Some of the subcategories, like "Exam Preparation and Tutoring" and "Amusement and Recreation Industries" (e.g., Trampoline Parks), achieve high rankings, potentially because they appeal less to seniors, who might show lower preferences for physical activity. According to a study of 2,593 participants over the age of 65, only 16% met the World Health Organisation's physical activity guidelines[37]. For park visitors, fitness centers are important throughout all 6 months, but become especially popular in autumn, suggesting that exercisers may prefer a mix of indoor and outdoor activities during that season. The ranking of subcategories remains consistent between weekdays and the weekend, indicating that this factor does not significantly affect which subcategories prioritize physical activity.

Parks and full-service restaurants are among the top-10 in this analysis, whereas in the OLS model, they had some of the poorest results. This happens because these two subcategories have very few co-visits when compared with their total number of visits. Typically, when someone visits these subcategories, they tend not to visit many others. We investigated this on December 3, 2019. When ignoring the small subcategories with fewer than 10,000 co-visits, there are 101 subcategories remaining. Among these, nature parks have the 2nd lowest co-visit to visit ratio (1.13) and full-service restaurants the 7th (1.23). This implies that a visit to these subcategories would reduce the likelihood of visiting any other location on the same day. However, this does not affect the importance, as it provides the percentage of co-visits directed towards fitness locations, given that the individuals visited parks or full-service restaurants on the same day respectively. The co-visit to visit ratio for fitness centers on December 3, 2019, is 1.29, obtaining the 11th lowest ranking. This suggests that visiting fitness locations reduces the probability of visiting numerous other locations on the same day.

The beverage bars follow results similar to Fig. 2, highlighting the importance of physical activity for their visitors. Combined with the substitutability results of Fig. 3, a preference exists for visiting beverage bars on the same day. Other subcategories for which fitness locations are an important destination are the "Investment Advice" (6.10% of co-visits on weekdays), "Musical Instrument and Supplies Stores" (6.07%), "Golf Courses and Country Clubs" (5.92%) and "Libraries and Archives" (5.78%).

Before or after exercise

In this section, we analyze whether users visit particular subcategories before their exercise session or afterwards, to address research question 5. We include only subcategories with many co-visits with fitness locations for confident results (see Methods).

Table 4 presents results for locations primarily visited before or after exercise sessions, along with some of the most visited subcategories. Supplementary Tables 12 and 13 show results for 4 more weeks. Despite minor differences between different months, the overall trend remains consistent. Most of the subcategories are visited after the exercise session rather than before. "Grocery Stores" ranks first in 5 out of 6 weeks having the most visits after the exercise session (October: 59.1%, December: 61.1% of users visited after exercise). This trend was even more pronounced on weekends, with percentages of 63.0% in December and 61.4% in October respectively. The "Merchandise Stores" (October: 58.8%, December: 58.5%) and the "Department Stores" (October: 58.7%, December: 56.2%) are two other subcategories that are visited mostly after the exercise sessions. The "Full-Service Restaurants" are also mostly visited after exercising (October: 55.7%, December: 56.1%). The subcategories that are visited before the exercise session the most are the "New Car Dealers" (October: 52.4%, December: 52.7% of users visited before fitness) and the "Colleges and Universities" (October: 59.2%, December: 60.0%).

We further analyzed the time of fitness visits to the nearest hour across a week in October and December (Supplementary Fig. 12). The time of visit contains an almost identical pattern, but with December seeing slightly more visits at 18:00 and 19:00, and October having slightly more visits at 20:00.

Subcategory	Fitness first*	Ranking*	Fitness first**	Ranking**
Grocery Stores	59.1%	1	61.1%	1
Merchandise Stores	58.8%	2	58.5%	2
Department Stores	58.7%	3	56.2%	6
Family Clothing Stores	58.6%	4	55.7%	9
Game Stores	57.8%	5	57.2%	4
Full-Service Restaurants	55.7%	10	56.1%	8
Sporting Goods Stores	55.1%	12	54.3%	12
Snack and Nonalcoholic Beverage Bars	49.0%	32	50.8%	24
New Car Dealers	48.6%	34	47.3%	35
Colleges and Universities	40.8%	36	40.0%	36

Table 4. Percentage of users exercising before visiting these subcategories and the corresponding ranking (including only 36 subcategories that had more than 50,000 co-visits in both weeks with "Fitness and Recreational Sports Centers"). A percentage below 50% means that users tend to exercise after visiting the corresponding subcategory. One entry is considered from one user visiting both fitness and subcategory X on a day. Users can have entries on multiple days of the week. *Presenting results from October 7–13, 2019. **Presenting results from December, 2–8, 2019

Discussion

The results identified empirical and robust patterns common among individuals engaged in physical activity through visits to fitness centers, laying a solid foundation for tailoring efficient reward schemes and public policies based on their needs, thereby enhancing motivation and combating global inactivity. Bauman et al.[1] highlighted the importance of understanding various influences of physical activity to inform successful interventions and identify potential mediators. While they covered various dimensions impacting physical activity, including demographics and environment, our analysis provides a valuable addition to existing knowledge. The utilization of mobility datasets provides real-time feedback on activity levels, empowering policymakers to build societies where physical activity is not only healthy but also convenient and enjoyable[38]. In light of our multifaceted results, several targeted intervention policies can be developed by domain experts. To elucidate, we offer a selection of illustrative examples.

On exercise days, and to a lesser extent on rest days, exercisers tend to visit locations like Starbucks more, both before (possibly for an energy boost) and after (possibly for rewarding themselves) their workout. This pattern suggests an opportunity for constructing a reward system through discount bundling, potentially fostering long-term physical activity habits. Numerous studies underscore the effectiveness of using rewards as a strategic method to promote healthy habits, such as engaging in physical activity[39–41]. A notable example is Sweatcoin[42], an app partnered with the UK's NHS that rewards users with discounts for steps taken, reaching over 170 million downloads and increasing step count by 19.5% within six months post-download[43]. Except from national bodies like the NHS, reward-based schemes to promote healthy habits has also been leveraged by regional policymakers, such as New York City's Department of Health and Mental Hygiene, which introduced the 'Health Bucks' program[44], offering coupons for fresh fruits and vegetables to encourage healthier eating. Introducing discount bundles tailored to meet the needs of exercisers facilitates the establishment of stable location-specific cue formation. Several gym chains already offer perks for frequent visits, but their effectiveness could be enhanced by partnering with coffee shops to provide rewards relevant for exercisers, and allow the development of location-specific cues, which is highly effective in sustaining long-term physical activity habits[39]. Additionally, engaging in physical activity with a partner significantly reduces the likelihood of missing sessions compared to individual efforts, and implementing a shared reward system can facilitate the formation of groups, enhancing adherence and motivation[45].

When exercisers take rest days, particularly those who visited fitness centers for consecutive days, they often opt for visits to full-service restaurants. Conversely, there is a noticeable decrease in visits to fast-food establishments on these days. This decline may stem from their preference against such venues, while on days of exercise, time constraints may lead them to compromise with fast-food outlets. Although the preference for healthier food for exercisers is evident, this creates disparities among populations. Fast-food restaurants are predominantly observed in low-income neighborhoods, strategically targeting potential consumers within these communities[46]. It was shown using large-scale mobility data that a 10% rise in exposure to fast-food outlets results in a 20% increase in the likelihood of individuals visiting them[47]. Moreover, in rural areas, there is limited availability of healthier options[48]. During the COVID-19 pandemic, while fast-food visits in urban counties saw a 4% reduction, they increased by 10% in rural counties[49]. To mitigate disparities among populations and ensure equal opportunities for a healthy lifestyle, a substantial investment in affordable and healthy options seems warranted.

Exercisers show a preference for first engaging in physical activity and then visiting grocery stores. Figure 4 illustrates the time interval between leaving a fitness venue and visiting a grocery store on December 8, 2019. The distribution exhibits characteristics of power-law, with many exercisers visiting a grocery store immediately after leaving the fitness center. This tendency may arise from exercisers' urge to efficiently manage their time. In specific neighborhoods, individuals may lack convenient access from their fitness venues to nearby grocery stores, potentially leading to time constraints and missed exercise sessions. The location of a fitness center is crucial for both users and owners, with a study of 193 fitness centers in Madrid showing that those in less walkable areas offered extra services to stand out from competitors[50]. Previously, it was shown that the surrounding environment is important for enhancing physical activity[51], with the Centre for Disease Control and Preventions (CDC) emphasizing the importance of active-friendly environments for increasing physical activity[52]. Improving accessibility between grocery stores and fitness locations could incentivize individuals to integrate these activities, promoting healthier lifestyles.

Our study focuses on patterns specific to visits to fitness centers and sports clubs, representing a subset of overall physical activity. While people also exercise in parks, at home, or during commutes, our emphasis on fitness facility visits addresses an important gap in the literature. Prior studies often concentrate on activity in parks or step counts, whereas behaviors associated with fitness centers remain underexplored. Given that access to recreational facilities is a well-established environmental correlate of physical activity[1], our study provides valuable insights into this significant yet less examined domain.

Conclusion

In this work, we analyzed visiting patterns collected from over 20 million GPS devices to identify interrelations of physical activity. We identified exercisers from visits to fitness centers and investigated their visits to other locations. We trained OLS models to quantitatively evaluate the probability of visiting a fitness location based on the rest of the same-day visits to other subcategories. We analyzed the visiting patterns of individuals who engage in exercise and those who do not, and identified significant differences in their preferences. We further found that exercisers tend to replace their fitness center visits on their rest day with visits to full-service restaurants and parks. Moreover, we constructed importance networks to identify subcategories for which fitness locations are important destinations on weekdays and weekends. Additionally, we identified subcategories which are

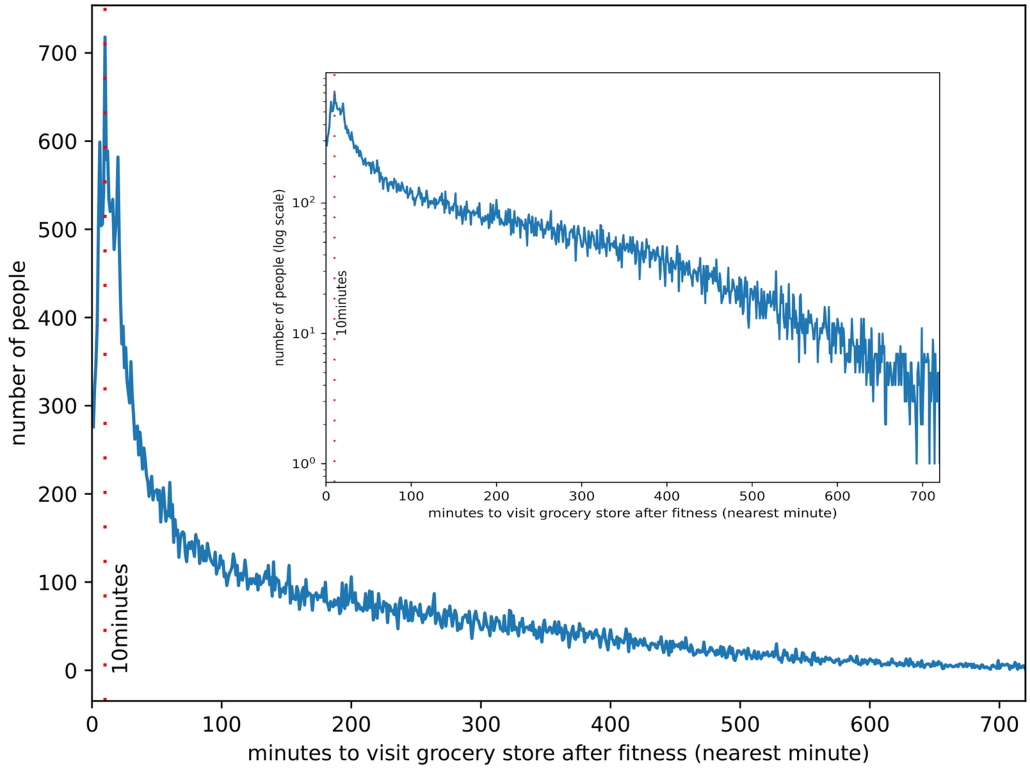

Fig. 4. Time interval between leaving a fitness venue and visiting a grocery store venue on December 8, 2019. Both figures depict the same data, but the y-axis (number of people) in the inner figure is displayed on a logarithmic scale.

visited mainly before or mainly after an exercise session. Our findings can serve as a tool for developing effective decision-making strategies aimed at promoting physical activity and facilitating the formation of prolonged exercise habits. Such policies have the potential to reduce global inactivity and encourage the adoption of a healthier lifestyle. Future research could further investigate trip chaining, particularly in relation to post-work exercise patterns, and consider the duration of exercise sessions to develop more personalized approaches.

Methods

Regression framework

The OLS regressions utilized in our analysis to explore the complementarities of physical exercise yield results based on the following equation:

$$F_{it} = \alpha + \sum_{s}^{S} \beta_{it}^{s} + \gamma_t + \delta_{it} \tag{1}$$

F_{it} is the dependent variable, which checks whether user *i* visited a fitness center on day *t*. β_{it}^{s} checks whether user *i* on day *t* visited subcategory *s*, using *S* number of predefined subcategories. We chose *S* to contain the nineteen most visited subcategories. For each user who visited at least one of the *S* subcategories, there exists a distinct entry in the model for each different day. γ_t represents the fixed effect corresponding to the day of the week, while $\delta_{i,t}$ represents the fixed effect specific to each user *i*, indicating the US state in which they performed most of their visits on day *t*.

We utilize OLS due to its interpretability, as demonstrated in similar studies[6], aligning with our analytical objectives. For example, the coefficients derived from OLS show the impact of an additional visit to a subcategory on fitness visits. To demonstrate the suitability of OLS, we performed a 5-fold stratified cross-validation, which yielded an average AUC of 0.703 (0.7023, 0.7027, 0.7031, 0.7033, 0.7027) on October 7–8, 2019, and an average AUC of 0.680 (0.6802, 0.6794, 0.6801, 0.6804, 0.6810) on December 2–3, 2019. Moreover, we tested the model by splitting the dataset into 80% training and 20% testing, and found that the mean squared errors for both sets were consistently close. For example, on December 2–3, the training error was 0.1081, and the testing error was 0.1079. We further tested for multicollinearity using the Variance Inflation Factor[53]. The highest value amongst the different subcategories used as independent variables was found to be consistently less than 1.5, showing that the independent variables are not correlated. To obtain quantitative insights, we trained the models on the entire datasets without keeping train and tests datasets, for which the results are shown in table 2 and in Supplementary Tables 2–6. The millions of entries that were used for fitting the models produced robust and confident results. For example, on Monday, December 2nd, 2019, there were 8,433,215 entries, and on Tuesday, December 3rd, 2019, there were 8,743,306 entries, giving a combined total of 17,176,521 entries for that corresponding OLS model. All obtained coefficients for all models are significant at the 1% level.

In the OLS models of Supplementary Table 14, the only differences in the equation solved is that F_{it} shows the amount of time a user *i* has spent at a fitness center on day *t*, and β_{it}^{s} shows the amount of time a user *i* has spent on subcategory *s* on day *t*.

Exercisers vs non-exercisers

We define an 'exerciser' as someone who has visited a fitness location at least once within a week. The remaining users of that particular week are categorized as 'non-exercisers'. Although different definitions of an exerciser would still provide meaningful results (e.g., exercised twice in a week), our decision was informed by observed user behavior, as the majority of exercisers ($\sim 62\%$) in our dataset demonstrated a tendency to visit fitness facilities only once a week. Additionally, we considered the mean number of weekly visits from a dataset containing ground-truth values of enrolled individuals (n = 61,293), which was 1.27 visits per week[6].

For a fair comparison, when analyzing results for each day of the week for the group of exercisers, we exclude individuals who exercised on the specific day under investigation and only consider those who exercised on the remaining days of the week. Our approach compares the visit patterns of exercisers and non-exercisers, on days where neither group engaged in physical activity. When a user exercises on a specific day, the visits to fitness locations make up a significant percentage of the total visits, thereby reducing the percentage of visits to other subcategories, compromising the fairness of the comparison. The objective of this comparison is to identify differences in visits between exercisers on their rest days and non-exercisers, while the OLS and "importance of fitness" analyzes focus on assessing same-day effects.

Substitutions of physical exercise

We create three groups of users:

1. Visited a fitness location on a single day and rested the following day.
2. Visited a fitness location for exactly two consecutive days and rested the following day.
3. Visited a fitness location for at least three consecutive days and rested the following day.The day in which the above groups rested is chosen to be the same one as a reference point (December 5, 2019). The same method is repeated for 4 days in different months (Supplementary Figs. 6–11). Each user can be assigned to only one of the three groups. For each group and for each day in the above schemes, we calculate the percentage of visits to different subcategories and we display the results in the figures. To ensure comparability in visit percentages between the three exercise days and the non-exercise one, we exclude fitness visits made on exercise days when calculating the percentages. If we include those visits, the percentage of visits for each subcategory increases on the non-exercise day due to the absence of fitness visits that were present on previous days.

Importance of physical exercise across categories

We initially construct a network that illustrates relationships between different location subcategories. Subsequently, we modify this initial network to create an importance network, removing bias related to the frequency of visits in larger subcategories.

Network construction

We calculate the total number of co-visits between the different subcategories. We define $n^{i,t}$ to be equal to 1 if user *n* has visited subcategory *i* on day *t*, and 0 otherwise. Then, the number of co-visits $C^{i,j,t}$ between subcategories *i* and *j* on day *t* is calculated using the following equation, with N the set of users on day *t*:

$$C^{i,j,t} = \sum_{n \in N} n^{i,t} n^{j,t} \tag{2}$$

We calculate this for all pairwise combinations of subcategories. If a user has visited *m* different subcategories on a single day, all *m* subcategories will contain *m-1* co-visits. At the same time, if a user only visited a single subcategory on a particular day, no co-visits exist. We repeat this methodology for a week in each of our analyzed periods and aggregate the total co-visits between each pair of subcategories, to get more robust results. Using the final co-visit values, we construct a network in which nodes are the different subcategories of locations, and the edges have weights equal to the number of co-visits between the subcategories. We present some interesting insights derived from a network constructed using co-visits from five consecutive days, Tuesday to Saturday, December 3rd to December 7th, 2019. The 4 most co-visited subcategories (full-service restaurants, limited-service restaurants, merchandise stores, gas stations) are involved in 57.5% of the total co-visits. Fitness locations are the sixth most co-visited subcategory with 10,185,198 co-visits. Based on the total visits during this period, fitness ranks as the fourth most visited subcategory. This difference (fourth in visits, sixth in co-visits) occurs since people that visit fitness locations tend to not visit many more subcategories on the same day, giving less total number of co-visits. The total co-visits between all subcategories in this period were 107,362,622.

Importance network
The subcategories that have the highest number of visits tend to be the ones most frequently co-visited with fitness locations. However, having the most co-visits does not necessarily mean that a subcategory is specifically important for fitness locations. In order to overcome this bias, we normalize the weights of the network, so that the sum of weights going out of each node is equal to 1. Each weight from node *i* to node *j* on day *t* gets adapted using the following equation:

$$w^{i,j,t} \leftarrow \frac{w^{i,j,t}}{\sum\limits_{k \in K \setminus \{i\}} w^{i,k,t}} \tag{3}$$

where K is the set of subcategories. The introduced network is a directed one and the importance given from a subcategory *i* towards fitness locations is determined by the weight of the edge moving from *i* to fitness. This weight essentially represents the percentage of co-visits of subcategory *i* to fitness, which is defined as 'importance' in Table 3.

Before or after exercise
For each subcategory *i*, we only keep the users who visited both *i* and fitness locations on a particular day. Then, for each user, we identify which visit among the two happened first. In each of our provided analyses, we have aggregated the results from an entire week. We retained results only from subcategories that received a minimum of 50,000 co-visits with fitness locations throughout both weeks of the comparison. This enhances the likelihood of the results being confident.

Data availability
The data supporting the findings of this study are available from Veraset, Inc. (provided upon request submitted at https://www.veraset.com). The data were collected through a CCPA compliant framework and utilized for research purposes. They were shared under contracts through their academic collaborative program, in which they provide access to de-identified and privacy-enhanced mobility data for academic research. The data were subjected to processing and analysis under a non-disclosure agreement, preventing any further sharing of data or attempts to re-identify individuals. The code is available at https://github.com/GeoIoan98/DeterminantsOfPhysicalActivity.

Received: 4 June 2024; Accepted: 15 January 2025
Published online: 27 January 2025

References
1. Bauman, A. E. et al. Correlates of physical activity: why are some people physically active and others not?. *Lancet* **380**, 258–271 (2012).
2. Choi, J., Lee, M., Lee, J.-K., Kang, D. & Choi, J.-Y. Correlates associated with participation in physical activity among adults: a systematic review of reviews and update. *BMC Public Health* **17**, 1–13 (2017).
3. Mitchell, M. S. et al. Financial incentives for exercise adherence in adults: systematic review and meta-analysis. *Am. J. Prev. Med.* **45**, 658–667 (2013).
4. Johnson, D. et al. Gamification for health and wellbeing: A systematic review of the literature. *Internet Interv.* **6**, 89–106 (2016).
5. Munson, S. A. & Consolvo, S. Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. In *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (Pervasivehealth) and Workshops* 25–32 (IEEE, 2012).
6. Milkman, K. L. et al. Megastudies improve the impact of applied behavioural science. *Nature* **600**, 478–483 (2021).
7. Coughlin, S. S. & Stewart, J. Use of consumer wearable devices to promote physical activity: a review of health intervention studies. *J. Environ. Health Sci.* **2** (2016).
8. Biddle, S. & Mutrie, N. *Psychology of Physical Activity: Determinants, Well-being and Interventions* (Routledge, 2007).
9. Sallis, J. F., Owen, N. & Fotheringham, M. J. Behavioral epidemiology: a systematic framework to classify phases of research on health promotion and disease prevention. *Ann. Behav. Med.* **22**, 294–298 (2000).
10. Saelens, B. E. et al. Neighborhood environment and psychosocial correlates of adults' physical activity. *Med. Sci. Sports Exerc.* **44**, 637–646 (2012).
11. Guthold, R., Stevens, G. A., Riley, L. M. & Bull, F. C. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1.9 million participants. *Lancet Global Health* **6**, e1077–e1086 (2018).
12. Sylvia, L. G., Bernstein, E. E., Hubbard, J. L., Keating, L. & Anderson, E. J. A practical guide to measuring physical activity. *J. Acad. Nutr. Diet.* **114**, 199 (2014).
13. Janssen, M., Scheerder, J., Thibaut, E., Brombacher, A. & Vos, S. Who uses running apps and sports watches? Determinants and consumer profiles of event runners' usage of running-related smartphone applications and sports watches. *PLoS One* **12**, e0181167 (2017).
14. Hodas, N. O. & Lerman, K. The simple rules of social contagion. *Sci. Rep.* **4**, 4343 (2014).
15. Christakis, N. A. & Fowler, J. H. The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* **357**, 370–379 (2007).
16. Mednick, S. C., Christakis, N. A. & Fowler, J. H. The spread of sleep loss influences drug use in adolescent social networks. *PLoS One* **5**, e9775 (2010).
17. Aral, S. & Nicolaides, C. Exercise contagion in a global social network. *Nat. Commun.* **8**, 14753 (2017).
18. Althoff, T. et al. Large-scale physical activity data reveal worldwide activity inequality. *Nature* **547**, 336–339 (2017).
19. Pontin, F., Lomax, N., Clarke, G. & Morris, M. A. Socio-demographic determinants of physical activity and app usage from smartphone data. *Soc. Sci. Med.* **284**, 114235 (2021).
20. Mejova, Y. & Kalimeri, K. Effect of values and technology use on exercise: Implications for personalized behavior change interventions. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization* 36–45 (2019).

21. Coleman, N., Gao, X., DeLeon, J. & Mostafavi, A. Human activity and mobility data reveal disparities in exposure risk reduction indicators among socially vulnerable populations during covid-19 for five us metropolitan cities. *Sci. Rep.* **12**, 15814 (2022).
22. Alba, C. et al. Covid-19's impact on visitation behavior to us national parks from communities of color: evidence from mobile phone data. *Sci. Rep.* **12**, 13398 (2022).
23. Benzell, S. G., Collis, A. & Nicolaides, C. Rationing social contact during the covid-19 pandemic: Transmission risk and social benefits of us locations. *Proc. Natl. Acad. Sci.* **117**, 14642–14644 (2020).
24. Holtz, D. et al. Interdependence and the cost of uncoordinated responses to covid-19. *Proc. Natl. Acad. Sci.* **117**, 19837–19843 (2020).
25. Papachristou, M., Banerjee, S. & Kleinberg, J. Dynamic interventions for networked contagions. *Proc. ACM Web Conf.* **2023**, 3519–3529 (2023).
26. Papachristou, M. & Kleinberg, J. Allocating stimulus checks in times of crisis. *Proc. ACM Web Conf.* **2022**, 16–26 (2022).
27. Zhou, Z., Liu, Y., Ding, J., Jin, D. & Li, Y. Hierarchical knowledge graph learning enabled socioeconomic indicator prediction in location-based social network. *Proc. ACM Web Conf.* **2023**, 122–132 (2023).
28. Veraset. Veraset visits. https://www.veraset.com/solutions/visits (2022).
29. Fan, C., Lee, R., Yang, Y. & Mostafavi, A. Fine-grained data reveal segregated mobility networks and opportunities for local containment of covid-19. *Sci. Rep.* **11**, 16895 (2021).
30. Chen, M. K., Chevalier, J. A. & Long, E. F. Nursing home staff networks and covid-19. *Proc. Natl. Acad. Sci.* **118**, e2015455118 (2021).
31. Ioannou, G., Marchioro, T., Nicolaides, C., Pallis, G. & Markatos, E. Evaluating the utility of human mobility data under local differential privacy. In *2024 25th IEEE International Conference on Mobile Data Management (MDM)* 67–76 (IEEE, 2024).
32. Squire, R. F. What about bias in the safegraph dataset? https://medium.com/safegraph/what-about-bias-in-the-safegraph-dataset-b95b7208893c (SafeGraph, 2019).
33. Pontin, F., Lomax, N., Clarke, G. & Morris, M. A. Characterisation of temporal patterns in step count behaviour from smartphone app data: An unsupervised machine learning approach. *Int. J. Environ. Res. Public Health* **18**, 11476 (2021).
34. Oscarsson, M., Carlbring, P., Andersson, G. & Rozental, A. A large-scale experiment on new year's resolutions: Approach-oriented goals are more successful than avoidance-oriented goals. *PLoS One* **15**, e0234097 (2020).
35. Arnardottir, N. Y. et al. Comparison of summer and winter objectively measured physical activity and sedentary behavior in older adults: age, gene/environment susceptibility reykjavik study. *Int. J. Environ. Res. Public Health* **14**, 1268 (2017).
36. Wilkins, E. J. & Horne, L. Effects and perceptions of weather, climate, and climate change on outdoor recreation and nature-based tourism in the united states: A systematic review. *PLoS Clim.* **3**, e0000266 (2024).
37. Rydwik, E. et al. Adherence to physical exercise recommendations in people over 65—the SNAC-Kungsholmen study. *Eur. J. Public Health* **23**, 799–804 (2013).
38. Hallal, P. C. et al. Global physical activity levels: surveillance progress, pitfalls, and prospects. *Lancet* **380**, 247–257 (2012).
39. Hagger, M. S. Habit and physical activity: Theoretical advances, practical implications, and agenda for future research. *Psychol. Sport Exerc.* **42**, 118–129 (2019).
40. Wood, W. & Rünger, D. Psychology of habit. *Annu. Rev. Psychol.* **67**, 289–314 (2016).
41. Yuan, Y., Nicolaides, C., Pentland, A. & Eckles, D. Promoting physical activity through prosocial incentives on mobile platforms. Available at SSRN 3960420 (2021).
42. Ltd, S. Sweatcoin. https://sweatco.in/ (2024).
43. Derlyatka, A., Fomenko, O., Eck, F., Khmelev, E. & Elliott, M. T. Bright spots, physical activity investments that work: Sweatcoin: a steps generated virtual currency for sustained physical activity behaviour change. *Br. J. Sports Med.* **53**, 1195–1196 (2019).
44. Payne, G. H. et al. Peer reviewed: implementing a farmers' market incentive program: perspectives on the New York City health bucks program. *Prev. Chron. Dis.* **10** (2013).
45. Burke, S. M., Carron, A. V., Eys, M. A., Ntoumanis, N. & Estabrooks, P. A. Group versus individual approach? a meta-analysis of the effectiveness of interventions to promote physical activity. *Sport Exerc. Psychol. Rev.* **2**, 19–35 (2006).
46. Block, J. P., Scribner, R. A. & DeSalvo, K. B. Fast food, race/ethnicity, and income: a geographic analysis. *Am. J. Prev. Med.* **27**, 211–217 (2004).
47. García Bulle Bueno, B. et al. Effect of mobile food environments on fast food visits. *Nat. Commun.* **15**, 2291 (2024).
48. Ko, L. K. et al. Food availability and food access in rural agricultural communities: use of mixed methods. *BMC Public Health* **18**, 1–9 (2018).
49. Banerjee, T., Nayak, A. & Zhao, H. A county-level study of the effects of state-mandated covid-19 lockdowns on urban and rural restaurant visits using consumers' cell phone geo-location data. *J. Public Health* 1–10 (2021).
50. León-Quismondo, J., Bonal, J., Burillo, P. & Fernández-Luna, Á. Walkability and fitness center prices, opening hours, and extra services: The case of Madrid, Spain. *Int. J. Environ. Res. Public Health* **17**, 5622 (2020).
51. Zhang, Y. et al. The impact of interventions in the built environment on physical activity levels: a systematic umbrella review. *Int. J. Behav. Nutr. Phys. Act.* **19**, 156 (2022).
52. Piercy, K. L. et al. The physical activity guidelines for Americans. *JAMA* **320**, 2020–2028 (2018).
53. Daoud, J. I. Multicollinearity and regression analysis. In *Journal of Physics: Conference Series*, Vol. 949, 012009 (IOP Publishing, 2017).

Acknowledgements
We thank Veraset for the dataset. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813162. This project was partly funded by European Union's Horizon Europe research and innovation program under Grant Agreement No: 101135826.

Author contributions
G.I. and C.N. designed the research. G.I. performed the analysis. All authors wrote the paper, discussed the results and provided feedback on the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/1

0.1038/s41598-025-87017-4.

Correspondence and requests for materials should be addressed to C.N.