scientific reports

Check for updates

OPEN

Data-driven nutritional assessment of urban food landscapes: insights from Boston, London, and Dubai

Michele Tufano[1], Fábio Duarte[2]✉, Martina Mazzarello[2], Javad Eshtiyagh[2], Carlo Ratti[2,3] & Guido Camps[1]

Urban food landscapes significantly influence dietary habits and health outcomes, with disparities in food access contributing to obesity, particularly in socioeconomically disadvantaged neighborhoods. This study presents a data-driven approach to assess urban food landscapes using restaurant menu data from online delivery platforms in Boston, London, and Dubai. Machine learning matched menu items to the U.S. FoodData Central database, enabling the calculation of nutritional indices and neighborhood-level nutrient averages. The analysis revealed significant patterns between urban food landscapes, socioeconomic features, and obesity rates. In London and Boston, higher socioeconomic neighborhoods had better access to nutrient-rich foods, with dietary fibers showing a strong inverse association with obesity ($p = 0.001$, $p = 0.004$, respectively). In Dubai, due to limited health data, the analysis focused on food landscapes and rental prices as a proxy of a neighborhood's socioeconomic profile. This method offers a scalable alternative to traditional food environment studies and can guide policymakers in identifying neighborhoods at risk for obesity and lack of nutritious foods. Future research should extend this method to diverse regions and advocate for standardized, open-access nutritional data to implement targeted and evidence-based nutritional interventions.

Urban food landscapes—defined by the availability and accessibility of food options—play a pivotal role in shaping individuals' dietary habits and consequent health outcomes. Obesogenic environments, characterized by an abundance of fast-food outlets, aggressive marketing strategies, and affordable processed foods, contribute to rising obesity rates, disproportionately affecting low-income and minority communities[1–5]. Neighborhoods with lower median home values and high deprivation are characterized by a higher density of fast-food outlets[6–9]. In contrast, higher-income groups benefit from more healthy food environments with greater access to nutritious foods and more information on healthy eating[10,11]. Notably, the proximity to fast-food outlets is a risk factor for obesity, diabetes, cardiovascular and coronary heart diseases[12–19].

Assessing the nutritional health of food landscapes is a central issue in public health, which requires comprehensive data on food availability and quality. Previous research on food landscapes has primarily relied on observational methods such as diary logs and records to gather data on food availability, quality, and pricing. However, this approach is labor- and time-intensive. The logs and questionnaire require multiple audits of the food landscape to represent a typical condition, and adapting to different food settings is challenging and limited to food regionality[20,21].

Previous studies on food landscapes have primarily relied on Geographic Information Systems (GIS) network analysis, which, while valuable for modeling spatial accessibility, often relies on outdated business data and fails to capture nutritional content[22].

Here, we present data-driven and rapid nutritional quantification of the food landscape. Rather than classifying restaurants in a few typologies, we collect data from all restaurants' menus available on food delivery platforms during the summer of 2023. Our study utilizes machine learning techniques to match menu items with entries in the FoodData Central database, enabling the assignment of their corresponding nutritional values[23]. FoodData Central, the USDA's comprehensive database containing standardized nutritional information for over 375,000 food items, serves as our reference database for matching menu items with their corresponding nutritional values. We then calculate average nutrient levels, the Meal Balance Index (MBI)[24]and the Nutrient-Rich Food Index (NRF)[25,26] per neighborhood. The MBI is a composite score that assesses how well a meal meets daily nutritional recommendations across key nutrients to reflect overall balance, while the NRF score quantifies nutrient density by positively weighting beneficial nutrients and negatively weighting those to limit, thereby

[1]Division of Human Nutrition and Health, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands. [2]Senseable City Lab, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]ABC Department, Politecnico di Milano, Milan, Italy. ✉email: fduarte@mit.edu

providing a single measure of nutritional quality. A schematic overview of the main workflow is presented in Fig. 1.

Our method was applied to three cities across different continents—Boston, London, and Dubai—each offering unique characteristics to contribute to a comprehensive analysis of urban food landscapes.

Boston, London, and Dubai represent distinct urban contexts for examining food landscapes: Boston, as an American city, benefits from comprehensive food data within the FoodData Central database, providing a reliable benchmark for nutritional assessment; London, as a major Western city, reflects global dietary trends through its diverse and accessible food environment; and Dubai, with its non-Western cuisine and significant obesity burden[27] highlights the dietary challenges of a rapidly developing urban center. These cities were selected due to the availability of robust data for analysis. By comparing these diverse cities, our study provides valuable insights into the applicability of our food landscape assessment method across varying urban environments and socio-economic contexts, while also exploring its relationship with neighborhood characteristics and obesity rates.

In contrast, in Boston and London, the study explores the relationships between the food landscape, obesity rates, and neighborhood socioeconomic features; whereas in Dubai, the analysis focuses on quantifying the nutritional composition of the urban food landscape and its correlation with rental housing prices as a proxy of each neighborhood's socioeconomic profile.

Results

This study analyzed restaurant menu information from three cities to quantify the urban food landscapes and their association with obesity rates and neighborhood socioeconomic features. Data were collected from restaurants with a physical presence in each city through online food delivery platforms, encompassing thousands of restaurants and millions of menu items: London (18,002 restaurants and 3,108,134 menu items), followed by Dubai (9,288; 1,556,243, respectively), and Boston (2,375; 222,964). After filtering the menus to discard drink-, tobacco-, portion-related-, and sauce-related items, we calculated the nutritional indexes of MBI and NRF as

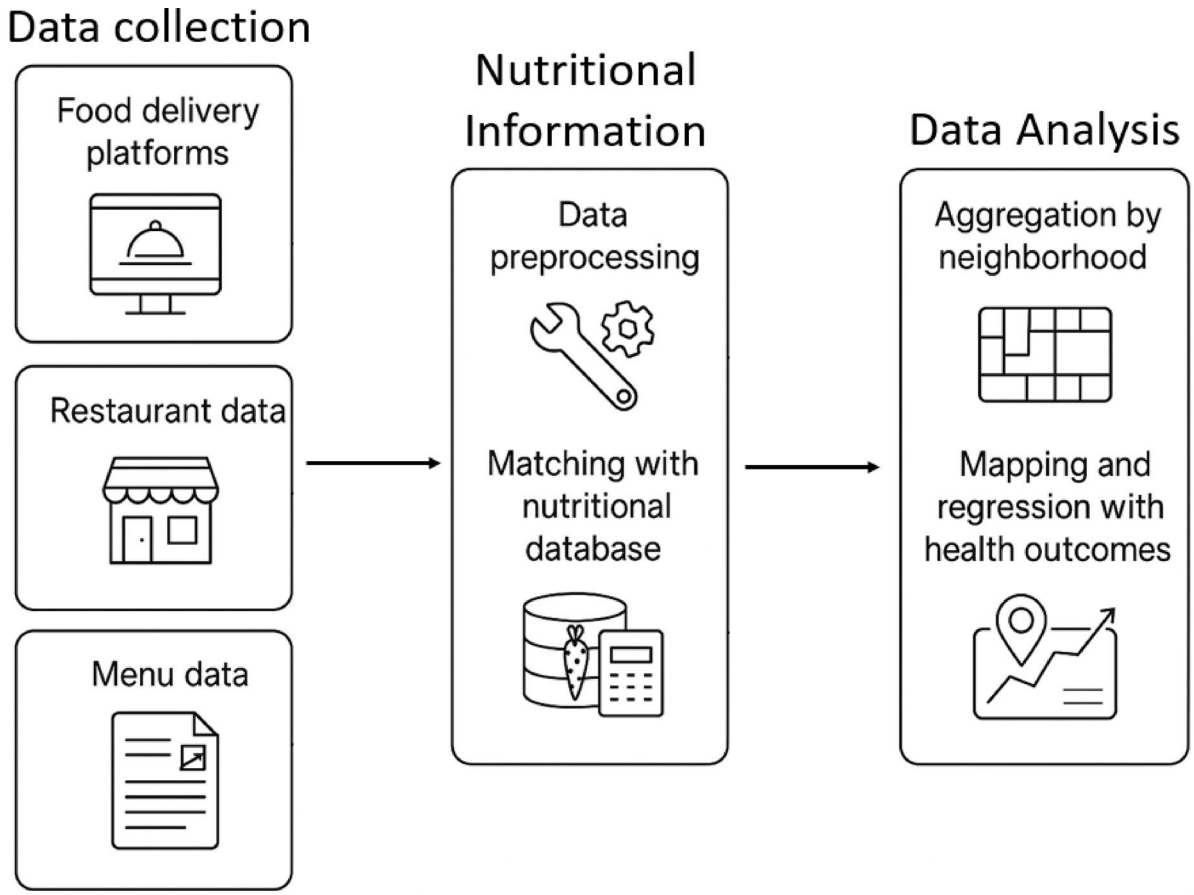

Fig. 1. Workflow for the collection, processing, and analysis of food delivery data for nutritional and geographic assessment.The left column illustrates data collection from food delivery platforms, including restaurant and menu information. The middle column represents data preprocessing and enrichment through matching menu items with a nutritional database to retrieve nutrient values and calculate indices (Meal Balance Index and Nutrient-Rich Foods Index). The right column shows the aggregation of nutritional data by neighborhood, followed by spatial mapping and multivariate regression analyses to investigate associations with obesity and socioeconomic features.

City	Restaurants, *N*	Total items, *N*	Restaurants after filtering, *N*	Foods after filtering, *N* (unique), %	Foods found in the database, *N* (unique), %	Foods with nutrients, *N* (unique), %
London	20,304	3,359,061	18,002	3,108,134 (485,016), 92.5%	1,746,424 (64,572), 56.1%	1,702,479 (62,686), 97.5%
Dubai	9529	1,775,769	9,288	1,556,243 (245,879), 87,6%	649,936 (43,142), 41.7%	640,551 (41,979), 98.5%
Boston	2857	264,269	2375	222,964 (100,796), 84.4%	158,295 (26,869) 71.1%	155,813 (26,150), 98.4%

Table 1. Overview of restaurant and menu data collected from online delivery platforms in london, dubai, and boston, including total items, filtered items, and the percentage of foods successfully matched to the fooddata central database. Nutrient assignments are also reported as a percentage of foods matched to the database.

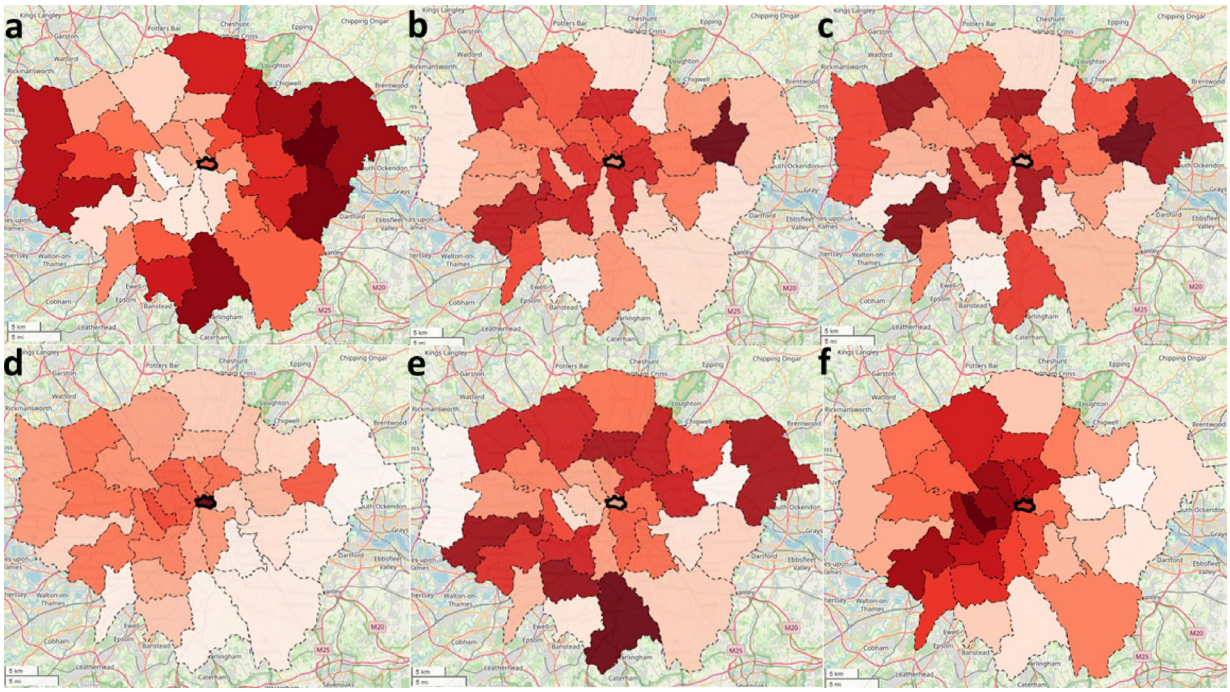

Fig. 2. London food landscape per neighborhoods, represented for (**a**) obesity prevalence in adults, (**b**) meal balance index, (**c**) nutrient-rich score, (**d**) dietary fibers, (**e**) potassium, and (**f**) housing prices. Areas with darker red gradients correspond to higher values. The City of London is marked in black at the center of the map. Implemented using Python (v3.10., https://www.python.org/downloads/release/python-3100/), geopandas (v0.8.1, https://geopandas.org/en/v0.8.1/reference.html), and Folium (v0.11.0, https://python-visualization.github.io/folium/version-v0.11.0/index.html).

indicative of the nutritional content of each food item. The nutrients were then aggregated per restaurant and neighborhood. For each neighborhood, we included socioeconomic features, such as housing prices and income. In London and Boston, we investigated the correlation between the food landscape and health outcomes. In Dubai, we limited the correlation between the food landscape and neighborhood socioeconomic features, due to the lack of health outcomes data. We begin by presenting the food landscapes of each city, followed by an analysis of the relationships between these food landscapes, obesity rates, and socioeconomic characteristics.

Food landscapes

The number of restaurants and foods with nutrients differed per city (Table 1). Foods from restaurant menus achieved the highest coverage in Boston, where 71.1% of the foods were found in the U.S.-based database. London and Dubai obtained lower database coverage with 56.1% and 41.7%, respectively. When a food item from a restaurant menu matched an entry in the FoodData Central database, its nutrient information was retrieved and assigned. This process successfully assigned nutrients in approximately 98% of cases. To investigate the inequalities of nutrients on a city level, we aggregate nutrients data by restaurant and neighborhood. The food landscapes depicted below for London (Fig. 2), Boston (Fig. 3), and Dubai (Fig. 4) show different concentrations of nutrient-rich foods, dietary fibers, MBI, NRF, and potassium across the cities. These indices were selected for their relevance in assessing overall meal quality, nutrient density, and potential health impacts at the community level. A detailed description of the nutritional values per neighborhood is provided in the Supplementary section online.

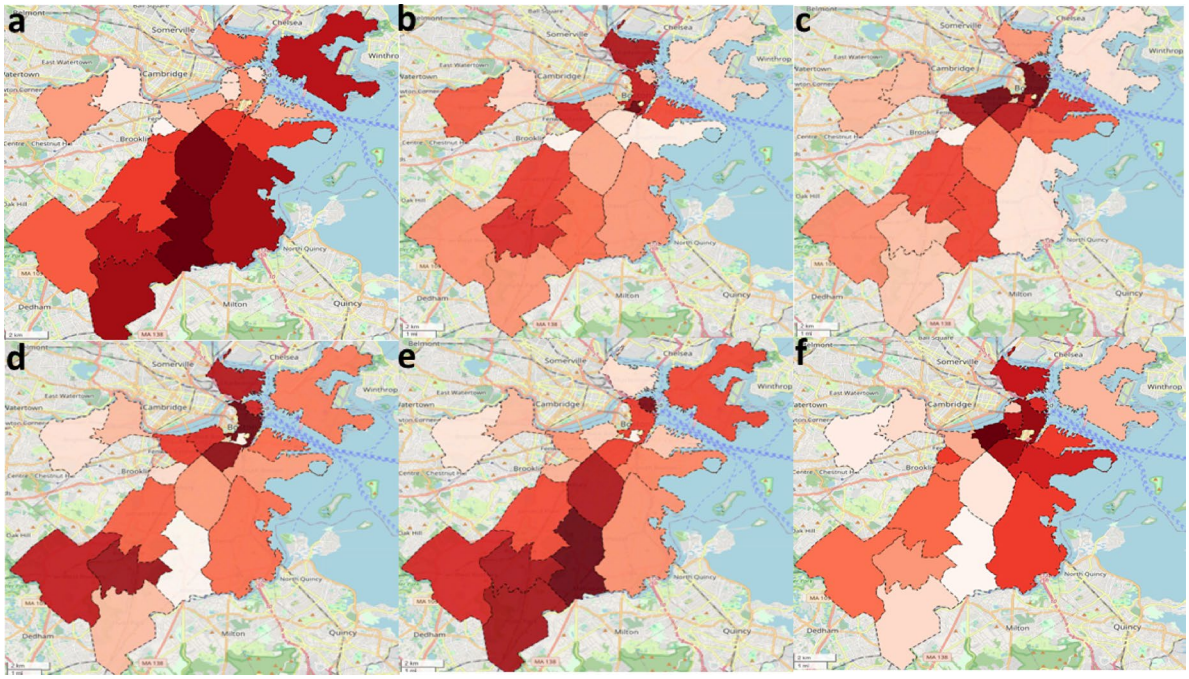

Fig. 3. Boston food landscape per neighborhoods, represented for (**a**) obesity prevalence in adults, (**b**) meal balance index, (**c**) nutrient-rich score, (**d**) dietary fibers, (**e**) potassium, and (**f**) housing prices. areas with darker red gradients correspond to higher values. Implemented using Python (v3.10., https://www.python.org/downloads/release/python-3100/), geopandas (v0.8.1, https://geopandas.org/en/v0.8.1/reference.html), and Folium (v0.11.0, https://python-visualization.github.io/folium/version-v0.11.0/index.html).

Fig. 4. Dubai food landscape per neighborhoods, represented for (**a**) meal balance index, (**b**) nutrient-rich score, (**c**) dietary fibers, (**d**) potassium, and (**e**) rental price. Areas with darker red gradients correspond to higher values. Implemented using Python (v3.10., https://www.python.org/downloads/release/python-3100/), geopandas (v0.8.1, https://geopandas.org/en/v0.8.1/reference.html), and Folium (v0.11.0, https://python-visualization.github.io/folium/version-v0.11.0/index.html).

Variable	Coefficient	Standard error	t-statistic	*p*-value	95% Confidence interval (lower)	95% Confidence interval (upper)	VIF
Intercept	19.7112	0.609	32.345	< 0.001	18.461	20.962	1
Meal balance index	0.5822	1.118	0.521	0.607	− 1.713	2.877	3.368
Nutrient-rich score	− 0.1999	1.065	− 0.188	0.853	− 2.385	1.986	3.055
Dietary fibers	− 2.5392	0.707	− 3.59	0.001	− 3.99	− 1.088	1.347
Potassium	− 1.9207	0.642	− 2.99	0.006	− 3.239	− 0.603	1.111

Table 2. Regression coefficients, standard errors, t-statistics, *p*-values, and 95% confidence intervals for predictors of obesity in adults in London.

Variable	Coefficient	Standard Error	t-statistic	*p*-value	95% Confidence Interval (Lower)	95% Confidence Interval (Upper)	VIF
Intercept	21.9992	0.741	29.697	< 0.001	20.443	23.556	1
Meal balance index	0.3811	0.916	0.416	0.682	− 1.543	2.305	1.528
Nutrient-rich score	− 1.3804	0.743	− 1.858	0.08	− 2.941	0.18	1.006
Dietary fibers	− 3.046	0.908	− 3.355	0.004	− 4.953	− 1.139	1.502
Potassium	2.7189	0.754	3.604	0.002	1.134	4.304	1.037

Table 3. Regression coefficients, standard errors, t-statistics, *p*-values, and 95% confidence intervals for predictors of obesity in adults in Boston.

Variable	Coefficient	Standard error	t-statistic	*p*-value	95% Confidence interval (lower)	95% Confidence interval (upper)	VIF
Intercept	569.0893	18.141	31.37	< 0.001	531.866	606.312	1
Eat 5-a-day fruit/vegetable - adults (%)	84.6118	18.753	4.512	< 0.001	46.134	123.089	1.069
Meal balance index	− 39.3576	32.98	− 1.193	0.243	− 107.027	28.312	3.305
Nutrient-rich score	3.7171	31.541	0.118	0.907	− 60.999	68.433	3.023
Dietary fibers	124.2368	20.692	6.004	< 0.001	81.78	166.694	1.301

Table 4. Regression results for predictors of housing prices in london: coefficients, standard errors, t-statistics, *p*-values, 95% confidence intervals, and variance inflation factors.

Food landscapes and obesity

We investigated the relationship between the food landscape and obesity in cities with available health data, specifically London (Table 2) and Boston (Table 3). Multivariate regression models revealed varying degrees of predictive power for obesity rates in both cities. Both models exhibited statistically significant relationships between predictor variables and obesity. However, the explained variance differed, with R-squared of 0.4 for London and 0.621 for Boston, suggesting a stronger predictive capability in the latter.

Notably, dietary fibers consistently showed a significant negative association with obesity in both cities (*p*-value: 0.001 for London; 0.004 for Boston). Potassium had differing effects in London and Boston. No significant association was found between obesity and MBI nor NRF in either city.

Food landscapes and neighborhood socioeconomic features

This section explores the relationship between neighborhood socioeconomic characteristics and the local food landscape.

The analysis considers key socioeconomic metrics, including median housing prices in London, median household income in Boston, and median rental prices in Dubai. Multivariate regression models are used to assess the associations between food-related variables and these socioeconomic factors. In London, the multivariate regression model demonstrated a strong predictive power for housing prices (R-squared = 0.703). Both dietary fiber intake and the daily consumption of five portions of fruits and vegetables were significantly and positively associated with higher housing prices ($p < 0.001$ for both). NRF showed no significant association with housing prices (Table 4).

In Boston, the multivariate regression model revealed robust predictive power for home values (R-squared = 0.725). Both NRF and dietary fiber intake were significantly and positively associated with higher home values ($p < 0.001$ for both). In contrast, the MBI was significantly and negatively associated with home values ($p = 0.021$) (Table 5).

In Dubai, the regression model accounted for only a small proportion of the variance in rental prices (R-squared = 0.046). Importantly, none of the dietary variables exhibited a statistically significant relationship with rental prices at the 0.05 level (Table 6).

Variable	Coefficient	Standard error	t-statistic	*p*-value	95% Confidence interval (lower)	95% Confidence interval (upper)	VIF
Intercept	779,700	24,600	31.695	< 0.001	728,000	831,000	1
Meal balance index	− 75,170	29,900	− 2.513	0.021	− 138,000	− 12,600	1.478
Nutrient-rich score	144,100	24,600	5.849	< 0.001	92,500	196,000	1.003
Dietary fibers	124,000	29,900	4.15	0.001	61,400	187,000	1.475

Table 5. Regression results for predictors of housing prices in boston: coefficients, standard errors, t-statistics, *p*-values, 95% confidence intervals, and variance inflation factors.

Variable	Coefficient	Standard error	t-statistic	*p*-value	95% Confidence interval (lower)	95% Confidence interval (upper)	VIF
Intercept	71.3771	2.644	26.993	< 0.001	66.145	76.609	1
Meal balance index	− 4.2895	3.34	− 1.284	0.201	− 10.899	2.32	1.596
Nutrient-rich score	− 1.3416	3.4	− 0.395	0.694	− 8.069	5.386	1.653
Dietary fibers	− 3.4064	2.707	− 1.259	0.21	− 8.762	1.949	1.048

Table 6. – Regression results for predictors of rental prices in dubai: coefficients, standard errors, t-statistics, *p*-values, 95% confidence intervals, and variance inflation factors.

Variable	Coefficient	Standard Error	t-statistic	*p*-value	95% Confidence interval (lower)	95% Confidence interval (upper)	VIF
Intercept	19.7112	0.506	38.96	< 0.001	18.676	20.746	1
Housing price	− 2.2092	0.593	− 3.727	0.001	− 3.421	− 0.997	1.372
Income (weekly)	− 1.2446	0.593	− 2.1	0.045	− 2.457	− 0.032	1.372

Table 7. Regression results for neighborhood socioeconomic predictors of obesity in london: coefficients, standard errors, t-statistics, *p*-values, 95% confidence intervals, and variance inflation factors.

Variable	Coefficient	Standard error	t-statistic	*p*-value	95% Confidence interval (lower)	95% Confidence interval (upper)	VIF
Intercept	21.9992	1.007	21.848	< 0.001	19.899	24.1	1
Income	− 1.3959	1.134	− 1.231	0.233	− 3.762	0.97	1.269
Housing price	− 1.4267	1.134	− 1.258	0.223	− 3.793	0.939	1.269

Table 8. Regression results for neighborhood socioeconomic predictors of obesity in boston: coefficients, standard errors, t-statistics, *p*-values, 95% confidence intervals, and variance inflation factors.

Obesity and neighborhood socioeconomic features

Lastly, we examine the associations between obesity rates and neighborhood socioeconomic features. Regression models for London and Boston were used to explore the correlation between obesity rates and key neighborhood socioeconomic factors, such as housing prices and income levels. In London, the regression model revealed significant negative associations between obesity rates and weekly income (Table 7), indicating that higher economic status is linked to lower obesity prevalence.

Similarly, in Boston, the regression results show limited ability to explain variability in obesity rates (R-squared = 0.223), with neither income ($p = 0.233$) nor home values ($p = 0.223$) showing statistically significant relationships with obesity (Table 8). The adjusted R-squared value of 0.145 suggests that a substantial portion of the variation in obesity rates is not explained by the included predictors.

Discussion

This study presents a quantitative nutritional analysis of urban food landscapes using menu data from restaurants available on online delivery platforms. Restaurant menus from Boston, London, and Dubai were matched to the FoodData Central database to retrieve food nutrients for each menu item, which were then aggregated per area and associated with obesity prevalence and neighborhood socioeconomic features.

Our findings demonstrate that leveraging online delivery platforms and AI-based nutritional analysis provides an efficient and scalable method for assessing urban food landscapes with potential global applications. This approach proved particularly effective in Boston and is likely applicable to other American cities, where a high match rate (71%) between restaurant foods and the standardized FoodData Central database was observed. This enabled the estimation of nutrient content and its assignment to online food items. Consequently, this method facilitated the visualization of food landscapes and the identification of neighborhoods or areas where access to healthy food may be limited, such as food deserts.

Nutritional data for each menu item were aggregated at the restaurant and neighborhood (or borough) levels to analyze the food landscape and examine variations across different areas within a city. Interestingly, dietary fibers from the food landscape were significantly associated with obesity rates and neighborhood socioeconomic features. Our findings highlight dietary fiber as a key nutrient consistently associated with lower obesity rates across both cities. In London, dietary fibers demonstrated a significant negative relationship with obesity ($p = 0.001$), and a similar association was observed in Boston ($p = 0.004$). This suggests that low concentrations of dietary fibers in the food landscape may serve as an indicator of higher obesity prevalence, consistent with epidemiological studies showing an inverse relationship between dietary fibers intake and body weight[28]. Neither the Meal Balance Index (MBI) nor the Nutrient-Rich Food (NRF) score showed significant associations with obesity in either location. This absence of correlation may reflect limitations in aggregate nutritional indices to capture the complex relationship between food landscapes and health outcomes.

Furthermore, the results highlight distinct patterns in the association between food landscape variables and neighborhood socioeconomic features across the three cities. Due to data availability, we employed diverse neighborhood socioeconomic features, such as median housing prices in London, median housing prices and median household income in Boston, and median rental housing prices in Dubai.

In London, the regression model demonstrated strong predictive power for housing prices (R-squared = 0.703), with significant associations observed for dietary fibers and fruit/vegetable consumption ($p < 0.001$ for both). Similarly, the model for Boston demonstrated strong predictive power for home values (R-squared = 0.725). Dietary fibers were significantly and positively associated with higher home values ($p < 0.001$). Surprisingly, the MBI was significantly negatively associated with home values ($p = 0.021$). This indicates that neighborhoods with higher socioeconomic conditions have better access to nutritious foods. These findings align with prior research indicating that higher-income neighborhoods generally have greater and more diverse access to nutritious foods[10,11]reflecting socioeconomic disparities in food availability. In contrast, the model for Dubai revealed a very low explanatory power for rental prices (R-squared = 0.046), with no significant associations observed for any of the dietary variables.

These findings suggest that data interpretability and local context will influence the ability to quantify the relationship between food landscapes and socioeconomic factors across cities. Strong associations in London and Boston indicate that dietary fiber and fruit/vegetable intake reflect socioeconomic disparities in food access. However, the weaker results observed in Dubai underscore the necessity for further research in non-Western urban contexts, particularly for cuisines or dishes that are not represented in the FoodData Central database. Furthermore, alternative sources of healthy food access, such as supermarket distribution patterns, may also contribute to the observed differences and require additional investigation.

The neighborhood socioeconomic features also had a relationship with obesity rates, regardless of the food landscape. In London, the regression model demonstrated significant negative relationships between obesity and both housing prices and weekly income. In Boston, the regression model showed low predicting power (R-squared = 0.223) and two non-significant negative relationships between obesity and both housing prices ($p = 0.223$) and income ($p = 0.233$). These results suggest that neighborhoods with higher socioeconomic status, as reflected by higher housing prices and income levels, are associated with lower obesity rates, and vice versa. Although the strong statistical significance of these associations in London ($p = 0.001$ for housing prices and $p = 0.045$ for income) underscores the link between neighborhood socioeconomic features and obesity, the Boston results suggest that these factors alone may not strongly predict obesity in this context, highlighting the need for further research.

These findings are consistent with previous studies indicating that higher income is associated with a reduced risk of obesity[29–31]. In contrast, residing in economically disadvantaged neighborhoods has been linked to a higher risk of obesity in both adults and adolescents[32,33]. In Dubai, the absence of obesity data prevented the analysis from being conducted.

Several limitations should be noted. A major limitation of this study is the varying coverage of menu items matched to the U.S. Department of Agriculture's FoodData Central across the three cities. The limited database coverage in Dubai highlights the difficulty of applying the U.S.-based food composition data to culturally different cuisines, despite the availability of Gulf-specific dietary guidelines and food composition Tables[34,35]. This limitation affects the generalizability of our approach, which depends on foods (1) listed on delivery platforms and (2) their compatibility with the U.S. nutrient database. The lower coverage in non-U.S. cities underscores the need for open-access, region-specific food composition databases to ensure comprehensive analysis—a limitation that contributes to the literature, highlighting bias in data availability. The lower database coverage in London (56%) and Dubai (42%) can be attributed to two main factors: differences in food terminology (e.g., "french fries" in the American database versus "chips" in British menus) and the absence of certain foods (e.g., common Arabic dishes are not included in the FoodUSDA database). This presents a limitation of the study, as not all cities and food items had nutritional values assigned consistently. While Natural Language Processing or Generative AI could, in the future, improve the semantic matching between menu items and food database entries, the lack of nutritional information for underrepresented foods remains a challenge. Therefore, region-specific nutritional datasets are crucial for accurate food landscape assessment. In spite of this limitation, the data-driven approach reaches a menu-item level of ingredient identification that far exceeds traditional studies that use coarser categorizations (fast vs. traditional restaurants, e.g.). In the future, other nutritional databases can be explored to include better descriptions of cuisines in all countries and foods from all ethnicities. This would enable easier access to nutritional information and more comprehensive assessments of food environments using these and other methods.

Another limitation of this study is data mining on food delivery platforms. Our datasets included all the foods served by restaurants with a physical presence that could be found online. Notwithstanding our data mining efforts, access to data was limited to a subset of food delivery platforms. As previously advocated by other

researchers, regulations on food delivery platforms are needed to ensure transparent nutrition information and accessibility[36,37].

The ecological, cross-sectional design of this study allows for the analysis of area-level patterns but does not account for individual-level variability. This limitation restricts the ability to infer causation or fully unravel complex associations, as city areas were used as the units of observation.

However, our approach enables rapid and automatic data collection of food availability, quality, and pricing in city restaurants. Unlike traditional food environment research, which is labor-intensive because it relies on observational scans and multiple audits[20–22]our method offers a more efficient alternative, although comparative validation remains limited. While traditional approaches such as in-person audits and food logs offer detailed individual dietary information, they are resource-intensive and geographically constrained. Our methodology enables comprehensive city-scale analysis with minimal manual intervention, though it primarily utilizes online menu data, it does not represent food consumption, and it might exclude individuals who cannot afford to purchase food regularly or who do not use online platform regularly. Future research should prioritize methodological validation through systematic comparisons with established techniques in selected urban regions, such as stratified random sampling, to evaluate consistency, identify discrepancies, and enhance accuracy. Our approach has potential applications in food policy by providing a quick and inexpensive way to identify areas for further study in urban health policy. The integration of ground-truthing protocols or hybrid methodologies combining digital and traditional data sources represents a promising approach to optimize reliability in urban nutritional landscape evaluation.

This study raises many opportunities for future research on the food environment using data-driven approaches, and for monitoring policy of obesity in the urban context. For example, the Al Ttay area, Dubai, showed the lowest average dietary fiber. Based on the correlation between nutrition data and obesity, our results indicate that this area might be at risk for obesity and lack of nutritious food—which could be used by Dubai to promote healthy food options to this community. Similarly, other cities may employ our data-driven approach to identify areas that may lack nutritious foods or that may be at increased risk for obesity and then intervene to counteract the obesogenic environment.

Restaurants were prioritized in our analysis, as they represent the most prevalent category on online delivery platforms, while grocery retail availability exhibits significant regional and platform-specific variation. Yet, nutritional information gathered with big data has already shown its potential for supermarket websites[38] and if coupled with the food landscape could represent a crucial innovation for monitoring nutrition and public health. To achieve this, we advocate for an increased number of open-source food databases per country, as well as more standardized food names on delivery platforms and across countries, ideally including nutritional information. This would enable our method to be applied in non-English-speaking cities and support global scalability. Menus collected from a new city could first be matched with the local food database, and any unmatched items could then be translated and queried against FoodData Central or other international databases.

In conclusion, our data-driven algorithm is specifically useful for assessing food landscapes based on restaurant data from online delivery platforms and food items identifiable in the US food system database (FoodData Central). Future research is needed to evaluate the algorithm's applicability in other contexts and with alternative food databases to extend its utility beyond the USA. Attention should be paid to neighborhoods that report a low average of fibers. This data-driven approach can serve as a valuable tool for policy makers to rapidly and automatically assess the urban food landscapes in a city and implement targeted interventions in neighborhoods at risk for lack of inadequate access to nutritious food and obesogenic environments.

Online methods

Data collection, processing and storage

Starting at the food delivery platform's website, the postcodes and addresses were used as delivery proxies including all neighborhoods for the analyzed cities. Boston, London, and Dubai were selected from a list of cities with attainable datasets on neighborhood socioeconomic features, health, and food menu items as representatives of three geographic regions with different dietary habits. The search was restricted to restaurant categories, therefore excluding grocery stores and other store types (e.g., liquor, pet, or flower stores). This focus was chosen because restaurants are the most prevalent category on the online delivery platforms used as our data source, whereas grocery stores are not consistently available across all cities and platforms. The time of delivery was changed to 'anytime' to include restaurants that did not deliver at the time of the analysis. The URL of each restaurant was retrieved and subsequently opened to fetch (where available) the name, address, latitude, longitude, postcode, city, geohash, cuisine tags, rating, and number of reviews. We assigned an identification number to each restaurant to link it with its menu information, which was stored separately. During the summer of 2023, we retrieved online menu information including (where available) the name of the menu item, description, price, image, and currency. Unique internal identifiers were used to identify individual instances of each menu item[39]. After removing duplicate restaurants, the data was aggregated per city and stored in a database using the open-source software for relational database management SQLite V.3.40[40]. All collection and processing code was written in Python V.3.10, using open-source Python Libraries request, BeautifulSoup, and selenium. These libraries were employed to send requests to website servers and manage HTML files or dynamically generated web pages of restaurants. The data collection code included built-in pauses to prevent overloading the website's servers. Multithreading was implemented to enhance efficiency.

Nutritional information

Data preprocessing was implemented on the dataset containing the dishes. Firstly, all food names and descriptions were standardized to lowercase, removing special characters, redundant single characters, and extra spaces. Drink-, size-, tobacco-, and sauce-related entries were removed using extensive lists of keywords to ensure a

focus on solid food items. To reduce the frequency of database queries, a dictionary was created to map the food items to restaurant ID numbers. The dictionary was sorted by the number of unique restaurant IDs. Further word replacements were used to standardize variations in food item naming. Certain words (e.g., specialty, combo) and characters (i.e., non-ASCII) were excluded to enhance data quality. Regular expressions were used for text replacement. To match the dishes from restaurant menus and the database entries, we used sentence transformers and FAISS index. The dishes and database entries were encoded using sentence transformers (model BAAI/bge-base-en-v1.5) using NVIDIA RTX A5500 Laptop GPU on Windows 10 Pro (12th Gen Intel(R) Core (TM) i9-12950HX, 2.30 GHz). The embeddings were moved from GPU to CPU, converted to an array, normalized, and an index was generated using the FAISS library[41]. After the semantic search, we queried a SQL database to retrieve the nutritional values of the best matches. SQL queries were created to extract detailed nutrient information for each menu item about protein, energy, total fat, fiber, potassium, calcium, iron, sugars, saturated fatty acids, sodium, vitamins, magnesium, zinc, and carbohydrate.

The nutrients were retrieved from a database created locally from Food DataCentral Table[23] (Supplementary Fig. 1). The Meal Balance Index (MBI) and the Nutrient-Rich Foods (NRF) index were calculated based on previously established methods[24–26],using functions that process nutritional values derived from databases.

The MBI and NRF scores quantify food nutritional balance and quality. These composite indices enhance our analysis by evaluating multidimensional nutritional properties across urban food landscapes, providing a more comprehensive assessment than single-nutrient measurements alone. The MBI was calculated to assess the nutritional balance of meals using a Python-based function. The MBI evaluates the proportion of key nutrients relative to the meal's total energy content and compares these proportions to established dietary recommendations. Each nutrient was normalized to a standard daily intake or energy content, scored against lower and upper intake limits, and assigned a weight based on its dietary significance. Nutrient scores were aggregated into an overall MBI by computing a weighted average. Higher MBI values indicate a more balanced meal composition, whereas lower scores suggest nutritional inadequacy or excesses.

The NRF was calculated to quantify the nutrient density of foods using a Python-based function. The NRF score incorporates nine encouraged nutrients (e.g., fiber, protein, calcium) and three limited nutrients (e.g., saturated fat, added sugars, sodium). Nutritional information for each food was normalized per 100 kcal to standardize comparisons. Each nutrient contribution was scaled relative to its daily value or maximum recommended value, with contributions capped at 100% to prevent overrepresentation. Positive contributions from encouraged nutrients were summed up, while negative contributions from limited nutrients were subtracted. The resulting NRF score reflects the overall balance of nutrient density, with higher scores indicating healthier food choices. Finally, the dataset was enriched with the calculated MBIs and NRFs scores.

Data analysis

The exploratory analysis was conducted using pandas[42]by extracting nutritional information from the dataset. The outliers were removed using interquartile ranges. For each numeric column, a statistical summary was generated including mean, median, standard deviation, and skewness. Histograms, box plots, density plots, and Q-Q plots were generated to inspect the characteristics and distribution of the dataset. The most common foods per city were plotted using plotly[43]. ChatGPT model 4 (OpenAI) was used for refactoring code and rephrasing text.

The food landscape analysis was performed by aggregating the nutritional and index values for all dishes associated with each restaurant and computing the averages per restaurant and per neighborhood (or borough, when applicable).

To map the nutrient distributions across the cities, geographical data were used to create polygons representing neighborhoods or boroughs using GeoPandas[44] and Folium[45]. Missing restaurants' latitude and longitude values were retrieved from the address using the Google Maps Geocoding API. Latitude and longitude were matched against the polygon list to assign restaurants to a specific area. The nutritional content was aggregated per area to visually represent the spatial data. Each nutritional metric was treated as a separate layer in the map, with custom style functions applied to each layer to assign specific colors and labels using branca and geopandas. After including layer controls, the maps were saved to HTML format and subsequently converted to PNG format for final presentation. Multivariate regression analysis with obesity rates was performed in London and Boston. The London latest data release from 2022 was retrieved from the London datastore[46]and from the Office for Health Improvement and Disparities (https://fingertips.phe.org.uk/). The Boston data were retrieved from the PLACES project from the Center for Disease Control and Prevention (CDC), using the 2023 PLACES: Local Data for Better Health, Census Tract Data 2023 release[47]. The nutritional data aggregated per neighborhood, or borough, was normalized and analyzed for multivariate regression with prevalence of obesity or neighborhood socioeconomic features. The Boston data were retrieved from https://data.boston.gov/. The Boston housing values corresponding to 31st December 2023 were obtained from https://www.zillow.com/research/data/. Due to absence of home price data, we collected rental prices in Dubai from www.bayut.com/ .

Data availability

The dataset associated with this manuscript, entitled "Restaurants with menus collected from online delivery Platforms in Boston, London, and Dubai", is available at https://doi.org/10.6084/m9.figshare.28302494.v1.

Code availability

The code will be made available upon reasonable request.

Received: 29 January 2025; Accepted: 19 June 2025

Published online: 08 July 2025

References

1. Egger, G. & Swinburn, B. An 'ecological' approach to the obesity pandemic. *BMJ* **315**, 477–480 (1997).
2. Cummins, S. & Macintyre, S. Food environments and obesity—neighbourhood or nation? *Int. J. Epidemiol.* **35**, 100–104 (2006).
3. Story, M., Kaphingst, K. M., Robinson-O'Brien, R. & Glanz, K. Creating healthy food and eating environments: Policy and environmental approaches. *Annu. Rev. Public. Health.* **29**, 253–272 (2008).
4. Pineda, E. & Mindell, J. Association of the food environment with obesity: A systematic review of geographical and statistical methods. *Lancet* **388**, S89 (2016).
5. Jones-Smith, J. C. et al. Obesity and the food environment: Income and ethnicity differences among people with diabetes. *Diabetes Care* **36**, 2697–2705 (2013).
6. Powell, L. M., Chaloupka, F. J. & Bao, Y. The availability of fast-food and full-service restaurants in the United States: associations with neighborhood characteristics. *Am. J. Prev. Med.* **33**, S240–245 (2007).
7. 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).
8. Kwate, N. O. A., Yau, C. Y., Loh, J. M. & Williams, D. Inequality in obesigenic environments: Fast food density in new York City. *Health Place* **15**, 364–373 (2009).
9. Pearce, J., Blakely, T., Witten, K. & Bartie, P. Neighborhood deprivation and access to fast-food retailing: A national study. *Am. J. Prev. Med.* **32**, 375–382 (2007).
10. Gordon, C. et al. Measuring food deserts in New York city's low-income neighborhoods. *Health Place* **17**, 696–700 (2011).
11. Storr, R., Carins, J. & Rundle-Thiele, S. Assessing support for advantaged and disadvantaged groups: A comparison of urban food environments. *Int. J. Environ. Res. Public. Health* **16**, 1135 (2019).
12. Kusuma, D. et al. Food environment and diabetes mellitus in South asia: A geospatial analysis of health outcome data. *PLOS Med.* **19**, e1003970 (2022).
13. Kanchi, R. et al. Longitudinal analysis of neighborhood food environment and diabetes risk in the veterans administration diabetes risk cohort. *JAMA Netw. Open.* **4**, e2130789 (2021).
14. Davis, B. & Carpenter, C. Proximity of fast-food restaurants to schools and adolescent obesity. *Am. J. Public. Health* **99**, 505–510 (2009).
15. Han, J., Schwartz, A. E. & Elbel, B. Does proximity to fast food cause childhood obesity? Evidence from public housing. *Reg. Sci. Urban Econ.* **84**, 103565 (2020).
16. Bodicoat, D. H. et al. Is the number of fast-food outlets in the neighbourhood related to screen-detected type 2 diabetes mellitus and associated risk factors? *Public. Health Nutr.* **18**, 1698–1705 (2015).
17. Walker, B. B. et al. The local food environment and obesity: Evidence from three cities. *Obes. Silver Spring Md.* **28**, 40–45 (2020).
18. Poelman, M. et al. Relations between the residential fast-food environment and the individual risk of cardiovascular diseases in the netherlands: A nationwide follow-up study. *Eur. J. Prev. Cardiol.* **25**, 1397–1405 (2018).
19. Ntarladima, A. M. et al. Associations between the fast-food environment and diabetes prevalence in the Netherlands: A cross-sectional study. *Lancet Planet. Health* **6**, e29–e39 (2022).
20. Lytle, L. A. & Sokol, R. L. Measures of the food environment: A systematic review of the field, 2007–2015. *Health Place* **44**, 18–34 (2017).
21. McKinnon, R. A., Reedy, J., Morrissette, M. A., Lytle, L. A. & Yaroch, A. L. Measures of the food environment: A compilation of the literature, 1990–2007. *Am. J. Prev. Med.* **36**, S124–133 (2009).
22. Gamba, R. J., Schuchter, J., Rutt, C. & Seto, E. Y. Measuring the food environment and its effects on obesity in the United States: A systematic review of methods and results. *J. Community Health.* **40**, 464–475 (2015).
23. FoodData Central. https://fdc.nal.usda.gov/.
24. Mainardi, F., Prozorovscaia, D., Sweeney, G. & Green, H. Development and validation of a meal quality index with applications to NHANES 2005–2014. *PLoS ONE* **15**, e0244391 (2020).
25. Fulgoni, V. L., Keast, D. R. & Drewnowski, A. Development and validation of the nutrient-rich foods index: A tool to measure nutritional quality of foods. *J. Nutr.* **139**, 1549–1554 (2009).
26. Drewnowski, A. Defining nutrient density: Development and validation of the nutrient rich foods index. *J. Am. Coll. Nutr.* **28**, 421S–426S (2009).
27. Mamdouh, H. et al. Prevalence and associated risk factors of overweight and obesity among adult population in dubai: a population-based cross-sectional survey in dubai, the United Arab Emirates. *BMJ Open.* **13**, e062053 (2023).
28. Slavin, J. L. Dietary fiber and body weight. *Nutrition* **21**, 411–418 (2005).
29. Grafova, I. B., Freedman, V. A., Kumar, R. & Rogowski, J. Neighborhoods and obesity in later life. *Am. J. Public. Health* **98**, 2065–2071 (2008).
30. Sheehan, C. M., Cantu, P. A., Powers, D. A., Margerison-Zilko, C. E. & Cubbin, C. Long-term neighborhood poverty trajectories and obesity in a sample of California mothers. *Health Place* **46**, 49–57 (2017).
31. Zick, C. D. et al. Running to the store? The relationship between neighborhood environments and the risk of obesity. *Soc. Sci. Med.* **1982**(69), 1493–1500 (2009).
32. : How entering, exiting, and remaining in poor and nonpoor neighborhoods is associated with obesity during the transition to adulthood - Adam M. Stuck in Unhealthy Places & Lippert (2016). https://journals.sagepub.com/doi/full/10.1177/0022146515627682.
33. Alvarado, S. E. Neighborhood disadvantage and obesity across childhood and adolescence: evidence from the NLSY children and young adults cohort (1986–2010). *Soc. Sci. Res.* **57**, 80–98 (2016).
34. Musaiger, A. O., Takruri, H. R., Hassan, A. S. & Abu-Tarboush, H. Food-based dietary guidelines for the arab gulf countries. *J. Nutr. Metab.* **2012**, 905303 (2012).
35. Bawajeeh, A. et al. Development of an Arabic food composition database for use in an Arabic online dietary assessment tool (myfood24). *J. Food Compos. Anal.* **102**, 104047 (2021).
36. Hetz, K. et al. The development and potential of a digital out of home food environment monitoring platform. *Nutrients* **15**, 3887 (2023).
37. Bates, S., Reeve, B. & Trevena, H. A narrative review of online food delivery in Australia: Challenges and opportunities for public health nutrition policy. *Public. Health Nutr.* **26**, 262–272 (2023).
38. Harrington, R. A., Adhikari, V., Rayner, M. & Scarborough, P. Nutrient composition databases in the age of big data: Fooddb, a comprehensive, real-time database infrastructure. *BMJ Open.* **9**, e026652 (2019).
39. Tufano, M. Restaurants with menus collected from online delivery Platforms in Boston, London, and Dubai. figshare. Dataset. (2024). https://doi.org/10.6084/m9.figshare.25976056.
40. Hipp, R. D. SQLite (2020). Available at: https://www.sqlite.org/index.html.
41. Douze, M. et al. The Faiss library. Preprint at (2024). http://arxiv.org/abs/2401.08281.
42. McKinney, W. et al. Data structures for statistical computing in python. in *Proceedings of the 9th Python in Science Conference* 51–56 (2010).
43. Plotly Technologies Inc. Collaborative data science. Montréal, QC (2015). https://plot.ly.

44. Kelsey, J. et al. Geopandas/geopandas: v0.8.1 (v0.8.1).*Zenodo*. https://doi.org/10.5281/zenodo.3946761 (2020).
45. Folium python-visualization. Version 0.11.0. (2020).
46. London, D. – Greater London Authority. https://data.london.gov.uk/
47. CDC. PLACES: Local Data for Better Health. Centers for Disease Control and Prevention (2023). https://www.cdc.gov/places/index.html.

Acknowledgements
The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the grant PROJ-01041/Province of Gelderland -Op Oost -EFRO InToEat. The authors wish to thank VLAG Graduate School from Wageningen University and Research for their support. J.E., M.M., F.D, and C.R. thank Dubai Future Foundation, UnipolTech, Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Volkswagen Group America, FAE Technology, Samoo Architects & Engineers, Shell, GoAigua, ENEL Foundation, Kyoto University, Weizmann Institute of Science, Universidad Autónoma de Occidente, Instituto Politecnico Nacional, Imperial College London, Universitá di Pisa, KTH Royal Institute of Technology, AMS Institute and all the members of the MIT Senseable City Lab Consortium for supporting this research. Grubhub data were sourced from the Grubhub website at https://www.grubhub.com/. JustEat's data were sourced from the JustEat website at https://www.just-eat.co.uk/. UberEats's data were sourced from the UberEats website at https://www.ubereats.com/. Talabat's data were sourced from the Talabat website at https://www.talabat.com. Deliveroo's data were sourced from the Deliveroo website at https://www.deliveroo.co.uk and https://www.deliveroo.ae. The authors wish to thank Lucrezia Bertoni and Karin Borgonjen for their nutritional advice.

Author contributions
M.T., F.D., M.M., J.E., and G.C. conceived the experiment(s). M.T., and J.E. created the primary data processing pipelines and performed the first data exploration. M.T., F.D., M.M., J.E., and G.C. created the final data presentation and interpretation. All authors reviewed and approved the final 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/10.1038/s41598-025-08098-9.

Correspondence and requests for materials should be addressed to F.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.