Food Accessibility From an Individual, Travel Behavior Perspective: Studies from Seattle, Washington and Stockholm, Sweden

Kaitlyn Ng

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington

2024

Committee:
Cynthia Chen
Brian Lee

Program Authorized to Offer Degree:
Civil and Environmental Engineering

University of Washington

Abstract

Food Accessibility From an Individual, Travel Behavior Perspective: Studies from Seattle,
Washington and Stockholm, Sweden

Kaitlyn Ng

Chair of the Supervisory Committee:
Cynthia Chen
Civil and Environmental Engineering; Industrial and Systems Engineering

Food accessibility plays a key role in maintaining the diet-related health of a community. Traditionally, food accessibility is typically measured at static, aggregated levels centered around people's homes. This fails to consider the diverse preferences and travel behaviors among the population. This thesis explores an individual-level, travel-behavior informed perspective to understand how people travel to food through two studies. The first study examines how restaurant preferences are shaped by people's built environment, sociodemographics, and travel behavior in Seattle, Washington. Unique personas distinguished by characteristics such as number of children in a household, vehicle access, and age are formed. The second study develops a trajectory-based accessibility measure which considers connectivity, affordability and temporal alignment. A case study is performed in Stockholm, Sweden using transit smartcard data. Grocery store accessibility is measured at localized, 15-minute hubs along individual trajectories. Zones with few grocery stores but high transit connectivity to dense, opportunity-rich areas observe positive increases to accessibility when considering mobility. The methods developed in the second study can be used to test effects of policy on food access. Overall, this thesis provides two different perspectives to understand the intersection of transportation and food accessibility.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Number Page

ACKNOWLEDGMENTS

I would like to thank Cynthia Chen, my advisor and chair of my committee, for all her support throughout my master's program. I have learned how to be a researcher and writer under her guidance– thank you for letting me explore new ideas and for providing many opportunities to try new things. Another big thanks to Brian Lee for his support throughout my master's program and my thesis writing. I would also like to thank my fellow THINKLab members for their friendships (and academic collaborations) along the way– I will miss you all! I also had a lovely time finishing up my master's degree at the KTH Royal Institute of Technology in Sweden, thanks to the University of Washington Valle Scholarship and Scandinavian Exchange Program. Thank you to Erik Jenelius for his advising and all the support from my peers during my time at KTH.

During my time in graduate school, I have been funded by multiple organizations. These include the Center for Teaching Old Models New Tricks (TOMNET), the Center for Understanding Future Travel Behavior and Demand (TBD), and the National Science Foundation's Leading Engineering for America's Prosperity, Health, and Infrastructure (LEAP HI) project entitled "Re-Engineering for Adaptable Lives and Businesses". Thank you to these organizations for making this work possible.

Chapter 1

INTRODUCTION

Access to food is a human right. Of the United Nations' 17 Sustainable Development Goals, the second goal is zero hunger. By 2030, the UN aims to end hunger and ensure access by all people, especially those in vulnerable situations, to safe, nutritious, and sufficient food [1]. Across global food access studies, single-parent and low-income households are more likely to have lower food access [2, 3]. Those who lack access to affordable, healthy food are more susceptible to health issues such as diabetes, obesity, and heart disease [4]. Because food access has implications on health outcomes, it is important to study why access varies and how to accurately quantify access to guide food policy.

Food access can be defined in five dimensions: acceptability, accessibility, accommodation, affordability, and availability [5]. Acceptability captures residents' attitudes regarding quality and cultural relevance of food sold. Accessibility considers the relationship between location of food sources and residents, defined by travel time, distance, and transportation resources. Accommodation describes how food is organized to meet residents' needs, through store operating hours and physical condition of the venue. Affordability reflects the residents' perception of purchase food. Availability measures the volume of healthy food. These dimensions can be subjective (e.g., does a cleanliness of a store meet a customer's expectations?), though some metrics such as accessibility and affordability are easier to quantify. However, all of these metrics vary at the individual level. Based on where a person lives, how much money they make, and even personal taste, food accessibility can vary greatly from person to person.

From a transportation perspective, the focus of food access has traditionally been on the accessibility dimension. Transportation and accessibility are inherently intertwined: accessibility is defined as the way land use and transportation systems enable people to reach activities and destinations [6]. Therefore, many transportation perspectives on food access have been driven by measuring distance and travel time to grocery stores [7]. However, the five dimensions show that food access is dependent on individual characteristics in both the accessibility dimension and beyond.

Traditionally, food accessibility has been measured at an aggregate scale. For example, a "food desert", defined by the U.S. Department of Agriculture (USDA), is measured at a census tract-level in two dimensions: low access and low income. A census tract is a food desert if it has both a poverty rate of greater than 20% (low-income) and at least 33% of the population lives more than one mile away from a supermarket in urban areas (low-access) [8]. However, this makes generalizations of a tract population's true level of access. The USDA definition of access measures the straight-line distance between the census tract centroid and nearest grocery store [8]. Clearly, this may not reflect access for an individual who lives on the border of a census tract or those who may rely on public transportation to travel to the grocery store. Criticisms of the "food desert" definition have pointed out the need to consider smaller spatial units, other sources of food beyond grocery stores, and the use of travel time instead of distance [9]. Additionally, this definition also only considers people's access around home, where in reality, food may be acquired at other locations in their daily journeys [10, 9].

Considering travel behavior is one way to overcome the limitations in traditional food accessibility literature. A travel behavior perspective is able to capture how people travel to food and is suitable for assessing the access of individuals, rather than spatially aggregated groups. Constraints such as a travel time budget, travel mode, and existing physical transportation infrastructure can constrain the set of locations accessible to an individual [11]. Contextualizing travel within the built environment portrays a more realistic picture of accessibility.

The food access literature has begun to consider how the context of a trip may shape one's accessibility. The context of one's trip may reveal food access for purposes of leisure, socialization, or solely for convenience. For example, trip purposes have an impact on the restaurant choice. Purposes of socialization and celebration prioritize restaurant characteristics of food type and quality, while purposes of quick meals or convenience prioritize restaurant location [12, 13]. Demographics also influence people's restaurant choices. Households with children and higher incomes were found to be associated with a higher frequency of eating out [14]. Chapter 2 studies how individual preferences are related to people's restaurant choice, among built environment and sociodemographic factors using the household travel survey from Seattle, Washington. Personas, which are classes with distinct features, are formed using these characteristics and tailored healthy-eating interventions are suggested for each class. These personas provide a holistic overview of the context behind restaurant-related travel.

Another emerging direction of travel behavior-related food access literature is leveraging big data to gain detailed information about people's mobility. People often access food at locations beyond the home, such as at work or school [15]. Access along a trajectory provides a more realistic view of the opportunities people may have to interact with food sources outside the home. A study in Shenzhen, China created space-time prisms along individual trajectories to count the number of food service facilities across a user's prisms. Considering mobility improved the accessibility of those living in resource-poor areas [16]. Another study introduced the concept of a dynamic food environment and found that fast food exposure at the home differed from exposure along a trajectory [17]. Chapter 3 studies how individual-level trajectories can be used to measure grocery store accessibility by developing an accessibility metric which considers constraints of store operating hours, proximity and affordability. Smartcard data from Stockholm County, Sweden is used.

This thesis provides an individual-level, travel behavior-informed perspective on food accessibility. Two cases studies are presented, which focus on restaurant preference and trajectory-based grocery store access, respectively. Both small and big data are leveraged and studies are conducted in Seattle, Washington and Stockholm, Sweden. Dimensions of access beyond accessibility are considered. By taking a travel behavior perspective, these studies aim to provide insight to inform effective food access policy.

Chapter 2

UNDERSTANDING ACCESS TO RESTAURANTS THROUGH PERSONAS: A LATENT CLASS APPROACH INTEGRATING PREFERENCES AND TRAVEL BEHAVIOR

2.1 Introduction

Food access is a critical factor influencing community health. Traditionally, food access has been simplistically measured using straight-line distances between the home and nearest grocery store or by the percentage of healthy-food retailers in a given area [8, 18]. However, these approaches have two main limitations. First, the focus on grocery stores fails to consider other sources of food such as restaurants and their associated health concerns. As of 2021, the average American spent approximately 5.1% of their disposable income on food away from home and ate out more than six times a week. Fast food comprised a third of meals eaten away from home [19, 20]. Furthermore, food from restaurants tends to have larger portion sizes and contains high levels of fat, saturated fat, sodium, and calories. In excess, these are linked to obesity, hypertension, heart disease, and diabetes. Menus at fast food restaurants are saturated with fried, high-sodium items that are lacking in nutrients [21].

Second, other dimensions of accessibility at an individual level, particularly with regards to travel behavior, are not considered. Recent approaches to quantify individual level food access have considered daily travel paths, mode choice and associated travel time, and access to public transit [22, 23, 24]. Understanding the behavioral dimension of food access can serve as the basis to form healthy eating interventions, whether it be to encourage healthier food choices or active modes of travel to restaurants. In this paper, we study restaurant food access through its context in transportation by integrating individual-level restaurant preference, travel behavior and sociodemographic factors.

A holistic perspective integrating both restaurant preferences and travel behavior within the context of transportation and built environment constraints is motivated from the basic observation that these two aspects are intrinsically related with each other. The realization of preferences toward restaurants in terms of its type, price, rating and food diversity are constrained by the types of transportation trips that one can take: for example, going longer distances to a restaurant requires not only the will to do so but also the time available, while on the other hand, some food choices such as fast food may be simply due to a number of factors that act as constraints (lack of sufficient monetary and time resources make this choice appealing to certain population segments) [25, 26].

Built environment constraints such as availability of healthy foods also impact people's choices. Those who live in a USDA-defined ”food desert”, or census tracts which have both a poverty rate of greater than 20% (low-income) and at least 33% of the population that lives more than one mile away from a supermarket in urban areas (low-access), have limited food options due to their surrounding built environment [8]. These areas have been found to be associated with higher travel times to grocery stores, lower vehicle access rates and therefore more reliant on public transportation [27].

People's typical activity patterns also influence restaurant-related travel behavior. Eating out has been found to be anchored near individuals' homes, workplaces and work-related social ties [28].The incorporation of various aspects of travel behavior beyond simply the straight-line, home-based distances as done by the majority of the existing studies on food access provides a much more personalized understanding of accessibility, allowing us link various aspects of restaurant preferences with those of travel behaviors [29].

Beyond transportation-related constraints, restaurant-related travel behaviors and choices are shaped by sociodemographic characteristics. A survey of U.S. households found that as age increased or if a household ate out on weekends, the likelihood of eating at full-service restaurants also increased. On the other hand, larger household sizes, the presence of children between 6 to 12 years old, and living in a rural areas increased the likelihood of eating at quick-service restaurants [26, 14]. Younger adults, less-educated individuals, and those who value convenience such as workers on their lunch break also tend to visit fast food restaurants [25, 30]. Lower-income neighborhoods and neighborhoods with a larger Black population were doing to have higher access to fast food establishments [31]. Given the connections between travel behavior, restaurant preference, built environment and sociodemographic characteristics, we propose the use of personas to understand these relationships.

A persona is a real or fictional personality that represents a class of consumers of interest. In this case, a segmentation based upon travel behavior and restaurant preferences is performed to obtain “personas”, which are then defined by specific demographic characteristics of the individual and their environment. For transportation applications, personas have been used to test incentives for departure time changes, study commuting behavior of employees to understand how to improve public transit usage, and determine efficacy of transportation demand management programs through individuals’ willingness to shift modes [32, 33, 34].

In a food access context, personas can be used to understand how to encourage healthier choices. From the food-related travel literature, we can posit example personas. Imagine an individual who is a stay-at-home parent with multiple young children and lives in the suburbs. This individual may choose to drive to a fast-food restaurant during lunch time, potentially for a quick, inexpensive meal to feed their children. Compare them to a higher-income individual who has no children, lives in the central business district, and works full-time. This individual may choose to walk to a high-end restaurant for dinner to meet up with friends multiple times a week. These two individuals make distinct restaurant and travel decisions shaped by their personal preferences, socioeconomic characteristics, and built environment constraints and would therefore require unique interventions to encourage healthier choices. From the persona development, curated healthy-eating policies can be developed to target specific behaviors.

In sum, we aim to study food accessibility with an integrated perspective combining both restaurant preferences and travel behavior. Personas will be created to understand food accessibility patterns within a population. Our research questions are as follows:

1. What different types of eating-out personas, which are defined by travel behavior and restaurant preferences, exist in the Puget Sound region? How are these personas affected by individual and environmental sociodemographic characteristics?

2. Are there linkages between personas and environmental characteristics? For example, are low-income low-access neighborhoods associated with particular types of restaurant

choices (e.g., fast food)?

3. Does a relationship exist between an individual's eating-out persona and level of food access? In other words, how does low food access impact restaurant-related travel behavior? Food access will be evaluated using the USDA "food desert" definition.

2.2 Datasets and Data Preparation

This study used data from four counties in the Puget Sound region: King, Kitsap, Pierce, and Snohomish. These four counties are in the jurisdiction of the Puget Sound Regional Council (PSRC), the local metropolitan planning organization. The household travel survey (HTS) conducted by PSRC from 2017-2019 provided region-wide travel behavior and socioeconomic information, Yelp provided restaurant-level information and the American Community Survey and PSRC provided environmental demographic data. Table 2.1 presents summary statistics.

2.2.1 Travel Behavior Characteristics

The study sample was created from the 2017-2019 PSRC HTS [35]. The travel survey included 136,079 trips made by 14,112 unique individuals. Individuals were surveyed from one to seven days. Of all trips, 11,265 or 8.28% were for meal purposes. Our sample was created by filtering for individuals who made at least one restaurant trip. Long trips (travel time greater than two hours) were omitted from the sample as these were assumed to be an infrequent travel occurrence. Travel time, trip purpose, mode, time of day and number of restaurant trips were identified as key travel behavior variables. The trip information was supplemented by additional GPS data, which provided specific latitudes and longitudes of a person's origin and destination. If an individual made multiple restaurant trips, trip data was aggregated to the individual level to be representative of the most common travel behavior characteristics. For categorical variables, the most frequent value was used (e.g., if a person made four restaurant trips by bike and one restaurant trip by walking, their aggregate mode would be coded as 'bike'). For continuous variables, the median value was used. The study sample is made of $n = 3,150$ unique individuals who made 52,323 trips over a period ranging from 1 to 7 days.

2.2.2 Restaurant Characteristics

The Yelp Fusion API was leveraged to identify which specific restaurant an individual visited given their latitude and longitude [36]. A keyword of “food” was used to perform a broad restaurant search, and the latitude/longitude were used to identify the geographically nearest search result as the visited restaurant. To validate the accuracy of the query, a 5% random sample of the latitude/longitude data was drawn. An equal allocation stratified random sample was performed for low, medium, and high density areas. Restaurant data was validated against Google Maps and had an 86% accuracy rate. Relevant variables from the Fusion API query were restaurant name, price, rating, and restaurant categories. The API returned 267 unique restaurant category types which we aggregated into ten broad groups. Food diversity was calculated by identifying the number of unique, grouped restaurant types an individual visited. Fast food restaurants were identified by their category on Yelp, and was composed of mostly large chain fast-food restaurants, but also included local quick-service restaurants.

However, it should be noted that there were some uncertainties in using Yelp data. The latitude/longitude data did not provide accuracy information, which made it difficult to determine the exact establishment visited if it was located in a restaurant-dense area. Ultimately, since data was missing completely at random, the visited restaurant was randomly imputed from the choice set of restaurants located near the specified latitude/longitude. The Yelp API was queried based on proximity to the specified latitude/longitude; however, the API returned restaurants with more recent reviews, higher number of user ratings, and other quality metrics [36]. Therefore, the returned restaurants are well-established and are more likely to be visited.

2.2.3 Sociodemographic Characteristics

For each individual, sociodemographic characteristics were obtained from the PSRC HTS. Age, race, gender, household income, number of children, ratio of vehicles to number of household members, and occupation were identified as relevant variables from the literature.

2.2.4 Built Environment Characteristics

Environmental demographic data was also included from the 2019 ACS, USDA and PSRC [37, 8]. The restaurant census tract, rather than the home, was used to evaluate the "attractiveness" of the restaurant itself. Number of employees was used as a proxy of business density at the visited restaurant census tract. Variables also included median income of the restaurant census tract, distance of the restaurant from the home and if the home was located in a "food desert. Median income, employees per census tract, and distance from home were averaged across all restaurant trips taken by an individual.

Table 2.1: Descriptive statistics of individuals who made restaurant trips (n=3,150).

Variable	% (n)	Mean (sd)	Description
Restaurant Preferences			
Fast food		0.14 (0.22)	Percent of fast food trips from all restaurant trips
Price		1.73 (0.42)	Average number of "dollar signs" from Yelp
Rating		3.64 (0.54)	Average number of "stars" from Yelp
Food diversity		2.92 (1.55)	Number of different food categories visited
Travel Behavior			
Travel time		13.34 (8.74)	Average travel time to restaurant
Restaurant ratio		0.53 (0.59)	Ratio between restaurant trips and non-restaurant trips taken
Mode			Most frequently used mode
Carpool	32.63% (1028)		
Drive alone	23.81% (750)		
Walk	39.78% (1253)		
Transit	3.78% (119)		
Purpose			Most common trip purpose paired with restaurant trip
Shop	5.24% (165)		
Home	45.75% (1441)		
Errand	11.59% (365)		
School/Work	27.59% (869)		
Social	9.84% (310)		

Time of day			Breakfast (5a-10a), lunch (11a-2p), dinner (5p-9p), or other
Breakfast	5.46% (172)		
Lunch	36.29% (1143)		
Dinner	27.43% (864)		
Other	30.83% (971)		
Environmental Demographics			
Restaurant Tract			
Number of employees		10,574.55 (14,705.48)	Average number of employees in visited restaurant census tract
Median income		84,362.37 (26,878.22)	Median income of visited restaurant census tract
Distance from home		4.02 (5.29)	Average distance between visited restaurant and home
Home Tract			
Food desert	3.68% (116)		If an individual's home is located in a ”food desert”
Individual Demographics			
Number of children		0.41 (0.86)	Number of children in household
Vehicle ratio		0.68 (0.46)	Vehicle to household size ratio
Age			Age categories (≤15, 16-25, 26-45, 46-65, >65)
≤15 yrs	6.29% (198)		
16-25 yrs	6.44% (203)		
26-45 yrs	57.87% (1823)		
46-65 yrs	20.67% (651)		
>65 yrs	8.73% (275)		
Gender			Female or non-female (male and other)
Female	49.37% (1555)		
Non-female	50.63% (1595)		
Race			White, Asian, Hispanic, Black, or other
White	65.90% (2076)		
Asian	17.11% (539)		
Hispanic	5.97% (188)		
Black	3.11% (98)		
Other	7.90% (249)		
Income			Household income (<$35k, $35-75k, $75-150k, >$150k)
<$35k	10.00% (315)		
$35-75k	24.60% (775)		
$75-150k	37.71% (1188)		
>$150k	27.68% (872)		

Occupation		Full-time, part-time, homemaker, or unemployed
Full-time	62.51% (1969)	
Part-time	7.52% (237)	
Homemaker	2.95% (93)	
Unemployed	27.02% (851)	

2.3 Methods

2.3.1 Latent Class Analysis

In this study, latent class analysis (LCA) is used to identify subpopulations of individuals who share similar travel behavior and restaurant preferences. LCA, also known as finite mixture modeling, is a model-based clustering method that has traditionally been used in the social sciences to uncover hidden classes from observed data [38]. Subpopulations are formed through a categorical latent variable modeling approach and create conceptually meaningful typologies for intervention planning [39]. This method translates well to a market segmentation approach [40].

LCA has been used in travel behavior research to cluster similar individuals. Previous literature has used LCA to segment transit users by activity-travel patterns, understand multi-modal travel attitudes, and capture transit service preferences [41, 42, 43]. Other geometrical and density-based clustering methods (i.e., K-means, DBSCAN) have been used in travel behavior research, but are heuristic in nature and cannot be evaluated statistically [32, 44]. However, because LCA is a model-based method, statistical "goodness of fit" measures such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) can be used [44].

LCA models consist of two kinds of variables: indicators and covariates. The model indicators define the latent classes, while the covariates predict the probability of individual membership to a latent class [45]. In this study, it is hypothesized that individuals belong to various latent classes, which differentiates their restaurant preferences and travel behavior. In turn, these latent classes are a function of individual and environmental demographic characteristics. Nine indicators were selected: five of which represent travel behavior and four of which represent restaurant preferences. Eleven demographic covariates were selected.

Figure 2.1 shows the relationship between the variables.

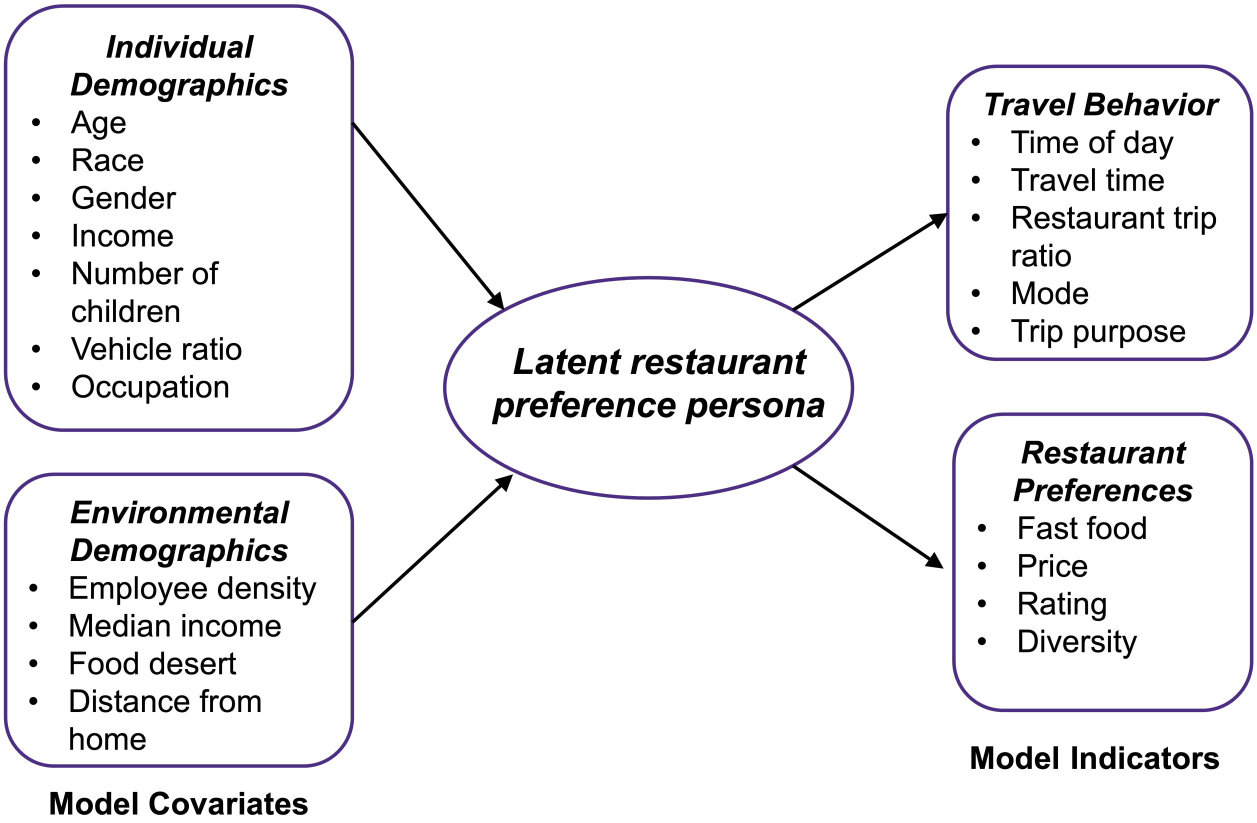


Figure 2.1: Latent class analysis model structure.

2.3.2 Model Formulation: LCA with Mixed Indicators

From Vermunt & Magidson, we adopt the following formulation for a latent class model [46]. Let $\mathbf{y_i}$ represent observations specific to individual i on a set of observed indicators denoted as $h = 1, .., H$; k is the latent variable with $k = 1, ..., K$ classes. Assuming independence between indicators within a class, the likelihood of the observed values on the H set of indicators for individual i can be expressed as:

$$f(\mathbf{y}_i) = \sum_{k=1}^{K} P(k) \prod_{h=1}^{H} f(\mathbf{y}_{ih}|k) \tag{2.1}$$

where $P(k)$ is the probability of belonging to a certain latent class and $f(\mathbf{y}_{ih}|k)$ is the conditional density function of a set of $\mathbf{y_i}$ values given a latent class k.

Equation (2.1) can be extended by including covariates in the model. Covariates are assumed to not influence the number of classes present in the model but can influence class membership [45]. Additionally, covariates provide further information about the individuals in each class [45]. We can thus write:

$$f(\mathbf{y}_i|\mathbf{z}_i) = \sum_{k=1}^{K} P(k|\mathbf{z}_i) \prod_{h=1}^{H} f(\mathbf{y}_{ih}|k, \mathbf{z}_i) \tag{2.2}$$

where $\mathbf{z_i}$ is the covariate vector for individual i.

Based on Equations (2.1) and (2.2) above, the log-likelihood function of the dataset can be written as:

$$LL = \sum_{i} log(\sum_{k=1}^{K} P(k|\mathbf{z}_i) \prod_{h=1}^{H} f(\mathbf{y}_{ih}|k, \mathbf{z}_i)) \tag{2.3}$$

The key item of interest is the posterior probability which can be found with the application of the Bayes' Theorem:

$$P(k|\mathbf{y_i}, \mathbf{z_i}) = \frac{P(k|\mathbf{z}_i) \prod_{h=1}^{H} f(\mathbf{y}_{ih}|k, \mathbf{z}_i)}{\sum_{k} P(k|\mathbf{z}_i) \prod_{h=1}^{H} f(\mathbf{y}_{ih}|k, \mathbf{z}_i)}. \tag{2.4}$$

Maximum Likelihood Estimation (MLE) can be used to estimate the latent class model. In particular, Expectation-Maximization (EM) algorithm was used, which was initialized with some initial guesses of the class membership k, followed by a fitting of the model for $\prod_{h=1}^{H} f(\mathbf{y}_{ih}|k, \mathbf{z}_i)$ (M step), and a recalculation of the posterior probability (E step) [38]. This process continued until little change was observed in the log-likelihood.

2.3.3 Model Estimation

The latent class model was estimated using a two-step approach using the software Mplus Version 8.9 [47]. In the first step, classes were estimated based off indicator data alone. Multiple models with 1 to K classes were fit in this step to enumerate the number of latent classes that underlay the data. Once the number of classes were finalized, the indicator model parameters were fixed, and a second model was estimated with the covariate data. This two-step model allowed for separation of covariates and indicators in the modeling process while minimizing classification errors [48].

During the estimation process, robust standard errors were used. Some variables were not normally distributed (i.e., time of day, which had peaks at lunch and dinner times; travel times had many short trips but fewer long trips). Observations were inherently non-independent among similar individuals, particularly those who belonged to the same household. Robust standard errors helped correct for these violations. Therefore, robust standard error estimation was used to account for non-normality and non-independence of observations [47, 49].

Class Enumeration and Formation

The first step in fitting the model was to enumerate the total amount of latent classes that underlay the data, using only indicators. Models with classes ranging from $k = 1$ to 6 classes were run, with each model of k classes compared to the model with $k - 1$ classes. Several goodness of fit measures were used to compare the models. The Bayesian Information Criterion (BIC), Sample-Adjusted BIC (SABIC) and Akaike Information Criterion (AIC) were used, for all of which a lower value represented the better relative model [50]. The Lo-Mendell-Rubin (LMR) test was used to determine if a more parsimonious model was better fitting. If a model with k classes had a non-significant LMR p-value, this suggested that the model with $k - 1$ classes was able to explain the data with fewer classes. Each model was also evaluated based on its entropy and size of its smallest class. A high entropy value was desired (entropy > 0.80), as it showed minimal uncertainty in the classification [51]. Class sizes greater than 50 samples were suggested to ensure that classes were generalizable [44]. Additionally, these results should be replicable as MLE solutions may get stuck in local minima or maxima [52]. In the event where log-likelihoods are unable to be replicated, the model may converge but yield different results each time. Unstable results such as these indicate a local, not global, solution. Table 2.2 below shows the fit statistics and Figure 2.2 compares the information criteria between $k = 1$ to 6 classes.

From the metrics used to evaluate the models, the data was best described by four or

Table 2.2: Class enumeration, with $k =$ 1 to 6 classes.[1]

k	Log-likelihood	AIC	BIC	SABIC	Entropy	LMR p-value	Class shares (%)
1	-35866.88	71777.75	71910.96	71841.06	-	-	100
2	-34336.42	68750.84	68986.99	68863.07	0.96	0.00	82, 18
3	-33491.61	67095.21	67434.30	67256.36	0.97	0.00	78, 19, 2
4	-32088.59	64323.19	64765.22	64533.26	0.98	0.00	65, 18, 15, 2
5	-31153.89	62487.78	63032.74	62746.77	0.99	0.00	62, 14, 11, 11, 2
6*	-30802.45	61818.90	62466.80	62126.82	0.97	0.00	57, 14, 11, 11, 5, 2

*Log-likelihood unable to be replicated.

1. Cluster robust standard errors were used during estimation.

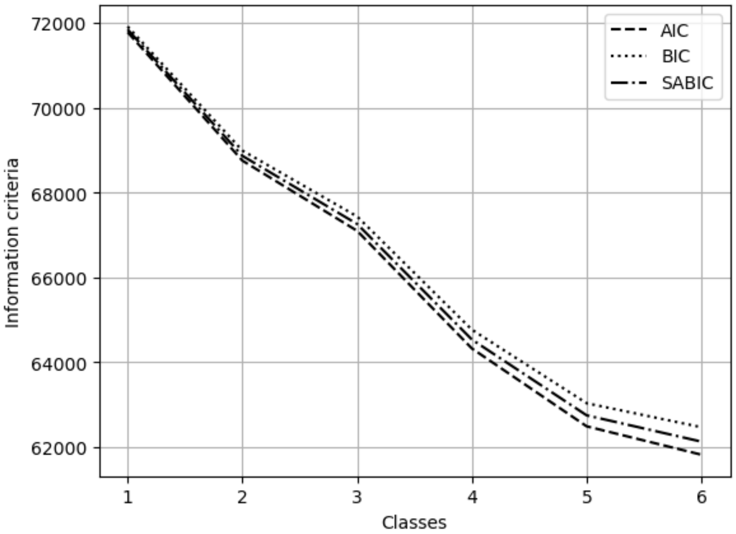


Figure 2.2: Information criteria comparison among models.

five classes. All information criteria (AIC, BIC, SABIC) showed decreasing values as class sizes increased; a greater number of classes explained the data better. All classes showed high entropy values and significant LMR p-values. A model was run with six class but could not replicate its log-likelihood value. This suggested that the six-class model attempted to over-extract classes from the data. Holistically, the four and five-class solutions were the most interpretable and were compared in the following step.

In the second step of the two-step approach, the four and five-class models were run again but with inclusion of covariates. The mean estimates for the indicator variables were fixed and only the covariates were estimated. The five-class model was unable to be estimated, suggesting the model was overly complex given the data. The four-class model was able to converge with the additional covariate data and was also more interpretable with only four classes. Thus, the four-class model was the best fit for the data and was used to interpret the final results.

2.4 Results

Four unique personas were found through the latent class model: 1) Convenient Eaters (14.64%), 2) Lunch Breakers (64.86%), 3) Restaurant Explorers (18.19%), and 4) Fast Food Enthusiasts (2.32%). The indicator results from the four-class solution are shown below in Table 2.3. These results provide the mean estimates of continuous indicator variables, and provide the odds ratio to show the most likely category for categorical variables. The covariate results are presented in an odds-ratio format in Table 2.4, comparing the odds of belonging in a class compared to Lunch Breakers for each variable. Table 2.5 displays the means and frequency values, showing the sociodemographic make-up of the assigned individuals and their environments for each class.

Across all classes, carpooling was the preferred mode, indicating a largely social nature of eating out. Home and work were the main anchors for restaurant trips. Variables at the restaurant location and gender did not significantly influence persona membership. Personas also did not display any geographical clustering within each class for home, work or trip locations. The four classes are comprised of a dominant segment, Lunch Breakers, and three smaller classes. The Lunch Breakers class represents a subset of workers who eat out during their lunch breaks. Lunch Breakers often ate at higher priced and well-rated restaurants for lunch, preferred to walk, and had home and work as the main anchors for restaurant activity. In the subsequent section, we focus on the three smaller segments with unique characteristics that were distinguished from the Lunch Breakers cluster. Figure 2.3 provides a comparison between the four classes.

2.4.1 Convenient Eaters (14.64%)

Individuals in the first class are defined by behaviors suggesting a preference for convenience. This class had moderately high fast food consumption, where 50% of restaurant

Table 2.3: Indicator mean estimates[1] and odds ratios for class definition, $k = 4$.

Indicators	1. Convenient Eaters	2. Lunch Breakers	3. Restaurant Explorers	4. Fast Food Enthusiasts
Class Size	$n = 461$ (14.64%)	$n = 2043$ (64.86%)	$n = 573$ (18.19%)	$n = 73$ (2.32%)
Restaurant Preferences[2]				
Fast food visits	0.50*	0.00*	0.22*	0.99*
Price (\$-\$\$\$\$)	1.44*	1.82*	1.70*	1.14*
Rating ($\star - \star\star\star\star\star$)	3.16*	3.79*	3.64*	2.49*
Food diversity	2.35*	2.76*	4.19*	1.18*
Travel Behavior				
Restaurant trip ratio	0.53*	0.52*	0.51*	0.79*
Travel time	14.03*	13.56*	11.76*	15.23*
Mode[3]				
Drive alone	0.75*	0.77*	0.59*	0.76
Walk	0.81	1.43*	0.95	1.06
Transit	0.10*	0.13*	0.08*	0.16*
Purpose[4]				
Shop	0.66	0.59*	0.32*	0.22*
Home	4.06*	4.85*	4.66*	3.06*
Errand	1.67*	1.14	0.97	1.15
School/Work	3.06*	2.94*	2.14*	2.66*
Time of day[5]				
Breakfast	0.20*	0.21*	0.15*	0.22*
Lunch	1.52*	1.30*	1.26*	1.39
Other	1.42*	1.08	1.02	1.38

* indicates a statistical significance at 5%.

[1] Mean estimates are provided for continuous variables (fast food through travel time). Odds ratios are provided for categorical variables (mode, purpose, time of day).

[2] Restaurant preference data was obtained from Yelp. Price and rating are the number of ”dollar signs” and ”stars” of a restaurant.

[3] Mode odds ratio value with baseline of carpooling.

[4] Purpose odds ratio value with baseline of social purpose.

[5] Time of day odds ratio value with baseline of ”dinner” mealtime.

Table 2.4: Covariate odds ratio[1] estimates with Lunch Breakers as a baseline, $k = 4$.

Covariates	Convenient Eaters vs. Lunch Breakers	Restaurant Explorers vs. Lunch Breakers	Fast Food Enthusiasts vs. Lunch Breakers
Environmental Demographics			
Restaurant Tract			
Number of employees	1.00	1.00	1.00
Median income ($)	1.00*	1.00*	1.00
Distance from home (mi)	1.01	1.01	1.02
Home Tract			
Food desert	1.23	2.04*	1.30
Individual Demographics			
Number of children	1.22*	1.04	1.04
Vehicle ratio	1.32*	1.09	0.71
Age[2]			
$\leq$ 15 yrs	1.24	1.95*	0.98
16-25 yrs	1.15	2.13*	0.91
26-45 yrs	1.42	1.90*	0.55
46-65 yrs	1.22	1.63	0.71
Gender[3]			
Female	1.03	0.88	1.16
Race[4]			
White	1.11	1.00	0.6
Asian	0.82	0.89	0.58
Hispanic	1.72*	1.34	0.52
Black	1.00	0.91	0.23
Income[5]			
<$35k	1.20	1.09	1.39
$35-75k	1.01	1.16	1.23
$75-150k	1.08	1.17	1.19
Occupation[6]			
Full-time	0.82	1.29	1.47
Part-time	0.85	1.47	1.28
Homemaker	0.94	1.14	3.17*

* indicates a statistical significance at 5%.

1. Odds ratio interpretation: "for individuals younger than 15, the odds of belonging to Restaurant Explorers are 1.95 times higher than belonging to Lunch Breakers".
2. Age odds ratio with baseline of 65 years.
3. Gender odds ratio with baseline of non-female.
4. Race odds ratio with baseline other race.
5. Income odds ratio with baseline of $150k.
6. Occupation odds ratio with baseline of unemployed.

Table 2.5: Covariate means and frequencies[1] of each class, $k = 4$.

Covariates	Convenient Eaters	Lunch Breakers	Restaurant Explorers	Fast Food Enthusiasts
Environmental Demographics				
Restaurant Tract				
Number of employees	11,080.93	10,641.94	9,842.18	11,239.50
Median income ($)	84,090.83	84,122.29	85,950.44	80,330.64
Distance from home (mi)	4.40	3.90	4.10	4.56
Home Tract				
Food desert	4.12%	2.99%	5.76%	4.11%
Individual Demographics				
Number of children	0.53	0.38	0.42	0.45
Vehicle ratio	0.71	0.67	0.68	0.60
Age				
≤15 yrs	8.24%	6.02%	5.58%	6.85%
16-25 yrs	5.21%	6.27%	7.68%	9.59%
26-45 yrs	57.92%	56.98%	61.95%	50.68%
46-65 yrs	20.17%	21.0%	19.72%	21.92%
65 yrs	8.46%	9.74%	5.06%	10.96%
Gender				
Female	50.98%	49.58%	46.77%	53.42%
Non-female	50.02%	50.42%	53.23%	46.58%
Race				
White	67.46%	65.83%	65.1%	64.38%
Asian	13.67%	18.06%	16.4%	17.81%
Hispanic	8.46%	5.19%	6.98%	4.11%
Black	3.04%	3.18%	3.14%	1.37%
Other	7.38%	7.73%	8.38%	12.33%
Income				
<$35k	10.63%	10.08%	8.9%	12.33%
$35-75k	22.99%	24.47%	26.0%	27.40%
$75-150k	38.61%	37.05%	39.44%	36.99%
$150k	27.77%	28.39%	25.65%	23.29%
Occupation				
Full-time	59.0%	61.92%	67.89%	58.9%
Part-time	6.94%	7.34%	8.55%	8.22%
Homemaker	3.47%	2.79%	2.62%	6.85%
Unemployed	30.59%	27.95%	20.94%	26.03%

---

1. Non-percentage numbers are the mean values for each class (i.e., Convenient Eaters visited restaurants an average of 4.40 miles from home). Percentages represent frequencies (i.e. 8.24% of Convenient Eaters are under 15 years old).

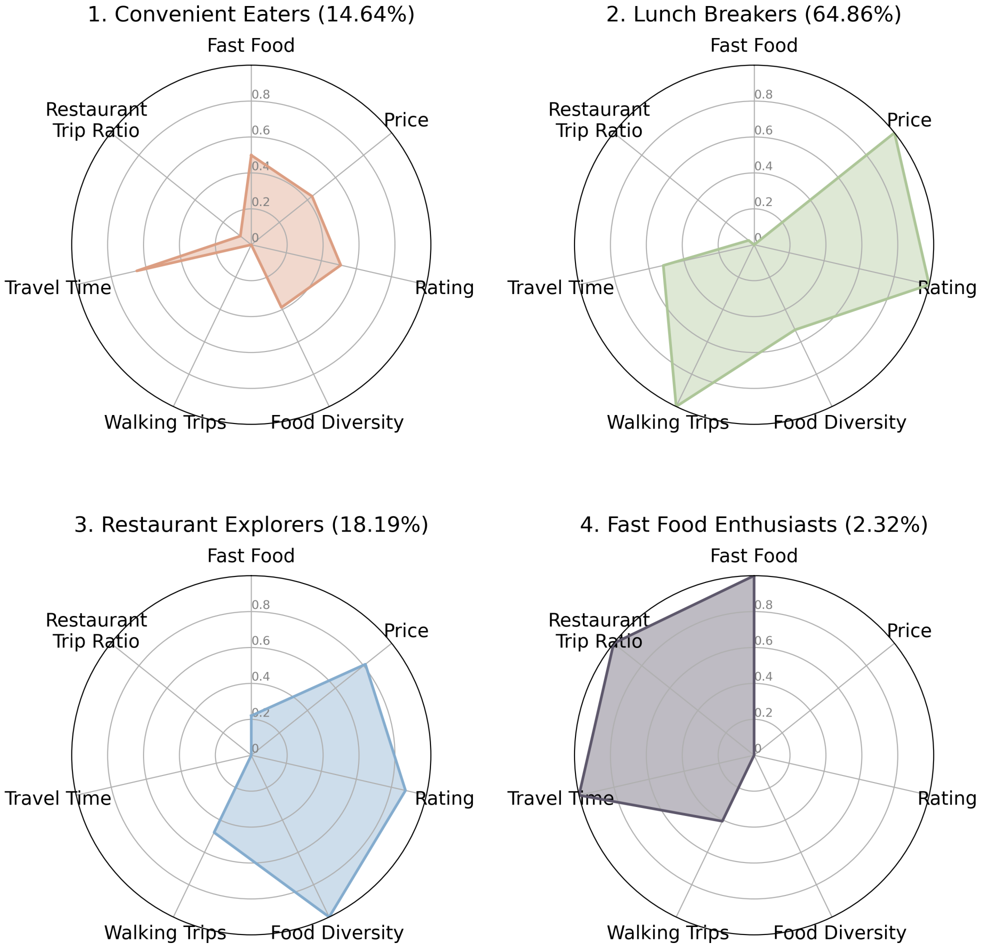


Figure 2.3: Persona comparison using normalized indicator estimates.

trips were fast food. Carpooling was preferred and driving alone was the second most likely mode. Convenient Eaters are the only class where errand trips were significantly more likely to be chained to restaurant trips than social trips (odds ratio of 1.67), though home and school/work were the most likely purposes before or after visiting a restaurant (odds ratio of 4.06 and 3.06, respectively). The odds of visiting a restaurant during lunch and non-meal times were 1.52 and 1.42 times greater than dinner times. Convenient eaters were more likely to have greater children per household and greater vehicles per household member. This class also had a greater proportion of ”unemployed” individuals, which includes those who are retired, students, and not currently employed. These individuals may have more temporal and spatial flexibility in their schedules.

These characteristics suggest a preference for convenience. Visiting fast food by driving may be a preferred choice for households with children or workers seeking a convenient spot for lunch, as fast food is quick and cheaper [14, 25]. Joint restaurant and errand trips suggest more efficient trip chaining behavior beyond eating out around traditional anchors such as home and work. Flexibility in meal times also supports the idea of a preference for convenience.

2.4.2 Restaurant Explorers (18.19%)

Individuals in this class were distinguished by their exploratory eating-out behaviors. Restaurant Explorers had the highest food diversity, with individuals trying an average of 4.19 different restaurant categories. This class also had the lowest travel times, were most likely to carpool to restaurants, and more likely to visit restaurants with higher prices and ratings. This class was more likely to be under 45 years old compared to the Lunch Breakers, and also had the highest proportion of full-time workers (67.89%) across all classes.

Interestingly enough, members of this class were almost twice as likely to live in a food desert compared to the Lunch Breakers (odds ratio of 2.04). 5.76% of Restaurant Explorers lived in a food desert, which was the highest proportion among all classes. However, this class also had the lowest proportion of individuals making less than $35k. These individuals had a similar frequency of eating out as Lunch Breakers and had the lowest travel time to restaurants across all classes. Limited grocery store access in their home census tract did not limit their restaurant choices or related travel behavior. This seemingly contradictory pattern may be an effect of rapid gentrification in the Seattle area. In 2019, Seattle was the third most gentrifying city in the U.S. [53]. ”Foodie” culture and gentrification are also correlated– newer, trendier restaurants attract younger, higher income residents [54]. Restaurant Explorers’ access to a range of restaurant types located near the home show that the ”food desert” definition is not representative of all types of food access, namely restaurants. Because ”food deserts” rely on aggregated population characteristics (i.e., census tracts with poverty levels greater than 20% and 33% of the population lives greater than one miles away from a supermarket), criticisms of the term argue that they do not capture individual nuances [7]. For example, the observed behavior of higher-income individuals in lower-income areas is something the "food desert" definition does not capture. These results support the arguments to shift away from using "food deserts" to measure food access as a whole [9].

2.4.3 Fast Food Enthusiasts (2.32%)

Individuals in the fourth class were defined by their high fast food visits. On average, this class visited fast food restaurants 99.6% of the time they ate out and had the highest frequency of eating out. This class also had the highest average travel time among all four classes of 15.23 minutes and were more likely to drive than to use any other mode. The average prices and ratings of restaurants visited were lower, which aligns with the fact that they ate out at mainly fast food establishments. These behavioral preferences show that Fast Food Enthusiasts were willing to travel longer to seek out fast food. This group also had the highest percent of low-income individuals (<\$35k) across all classes and also were significantly more likely to be a homemaker (odds ratio of 3.17) compared to Lunch Breakers.

Although this class made up a small share of the sample data, we argue that it highlights an important population segment with large public health concerns. Firstly, the average posterior probability of belonging to Fast Food Enthusiasts was 0.99 across all members of the class. This indicates high certainty that individuals belong to Fast Food Enthusiasts, and not any other class [45]. Secondly, this class is important to identify because of the extremely high frequency of fast food consumption. Fast food is typically less healthy than food at non-fast food restaurants, and is associated with higher body mass index [21, 55]. This persona identification is especially critical to identify the associated behaviors of those with unhealthy diets.

2.5 Discussion and Implications

The study identified four distinct eating-out personas who differ in terms of restaurant-related travel behavior, restaurant preferences, personal demographic, and environmental demographic characteristics. From these personas, we observe that preferences as well as travel behaviors play a role in the restaurant selection process. In particular, three segments are observed with characteristics distinct from Lunch Breakers, the segment of workers eating out during lunch time. Convenient Eaters demonstrated a preference for more efficient choices, choosing to chain trips, eat at various times of the day, and moderately consumed fast food. Restaurant Explorers were more likely to live in food deserts but preferred to visit different types of well-rated and higher priced restaurants. Fast Food Enthusiasts traveled longer to exclusively consume fast food.

This study analyzed eating-out behaviors from a transportation perspective. Convenient Eaters, Restaurant Explorers, and Fast Food Enthusiasts all preferred driving over any other mode. However, urban sprawl and subsequent physical inactivity lead to increased rates of obesity [56, 57]. Whether the observed driving behavior be from personal preference or environments not suited for walking or biking, encouraging active transportation is an approach to improving health. Increased mixed land uses by integrating retail into residential areas may be a way to promote more active lifestyles.

The personas also displayed distinct food choices. Namely, Convenient Eaters and Fast Food Enthusiasts often chose to eat at fast food establishments. Fast food is linked to negative health outcomes such as obesity [21, 55]. Additional research in conjunction with public health and behavioral science experts should be conducted to understand the decision-making processes of frequent fast food consumers. These personas may also be able to inform healthy-eating policy; personalized educational programs or food subsidies can be further explored alongside food policymakers.

There are factors that limit the generalizability of the persona results. First, the dataset only included individuals who made restaurant trips in the study period which represents 15.29% of all individuals covered by the PSRC HTS survey. This sample did not capture individuals who did not visit restaurant trips, omitting those who are unable to eat out due to financial, physical, or other constraints. Second, the accuracy of the geocoded restaurant locations is unable to be validated. However, several checks were used such as validating the identified restaurant and proximity to the provided GPS coordinate and prioritizing restaurants with a higher number of ratings. Lastly, the sample data is from before COVID-19, when eat-out behaviors changed drastically with the rise of food delivery services. The results from this study may differ from the behavior of current restaurant travel.

Future work can be done to expand the representativeness of the dataset. Eating-out behavior literature has pointed out gaps in studying the behavior of individuals who choose to eat-in [58]. The PSRC HTS did not provide information on ”eating-in” behaviors, as there was no detail on the activities performed at home. A further survey could be conducted to gather this data to capture cooking at home and the use of food delivery services. This would result in a more representative dataset, including both individuals who did and did not eat out. Grocery shopping data could also be incorporated to include all potential sources of food access and to improve comparison to grocery store access studies. Additionally, using methods such as natural language processing to obtain preference data from Yelp reviews may add further depth into understanding motivations behind restaurant selection [59].

Chapter 3

TRAJECTORY-BASED FOOD ACCESSIBILITY, AFFORDABILITY, AND CONNECTIVITY: FREQUENT TRANSIT USERS IN STOCKHOLM, SWEDEN

3.1 Introduction

Livable, accessible, and walkable neighborhood planning has long been a foundational goal in urban development, aiming to create environments which connect people essential amenities. Early 20th-century ideas such as the garden city and neighborhood units promoted functional, self-sufficient satellite areas [60, 61]. As public transport systems expanded, these ideas evolved to prioritize mobility along with accessibility. Concepts such as New Urbanism, pedestrian pockets, and transit villages promoted mixed use, dense neighborhoods centered around and interconnected by public transport [62, 63, 64]. These ideologies fall under a broader urbanist perspective of polycentric urbanism, a decentralized planning approach which emphasizes multiple, connected hubs [65].

In recent years, the 15-minute city model has revived the conversation about polycentric approaches to city planning. The 15-minute city focuses on access to amenities around the home: a person should be able to live, work, shop, go to school, access healthcare, and enjoy entertainment all in one neighborhood [66]. However, lessons regarding connectivity and mobility from the 15-minute city model's predecessors should be revisited to model accessibility in a realistic manner. Namely, the role of public transport in creating a polycentric city made of connected, walkable hubs should be considered. In this study, an accessibility measure is proposed which draws from the 15-minute city along with previous urbanist theories. By focusing on pedestrian accessibility around transit stations and mobility between hubs, accessibility is studied in the context of everyday travel patterns.

The 15-minute model is extremely decentralized, idealizing that people can conduct all necessary activities around the home; however, given the reality of existing urban planning, people often must travel outside their home neighborhoods to fulfill these needs [67, 68]. This requires a shift in understanding accessibility from a static, home-based to a dynamic, travel-sensitive perspective. For example, food access is often assessed by the distance or travel time between a grocery store and the home [10, 69]. However, a study in the U.S. found that individuals who live in low-income neighborhoods with limited food access were more likely to access grocery stores at work [15]. Therefore, people's travel patterns (human mobility) should be a key consideration when measuring accessibility. Accessibility approaches which consider human mobility exist, though they require very detailed data (e.g., travel diaries). For example, commute data was used to construct space-time prisms around home and a work commute to show time availability for grocery shopping. Accounting for commuting behavior improved accessibility for some residents compared to a home-based measure [70]. Emerging approaches measure access along daily activity sequences (”trajectory-based”) through passively-collected human mobility data [17, 71, 16]. A trajectory-based approach can capture travel patterns between urban hubs and therefore provide a more realistic understanding of accessibility.

Approaching accessibility from a trajectory-based approach requires consideration of region-wide transit, pedestrian, bike, and vehicle networks. As was reflected in the ideas of pedestrian pockets and transit villages, transit stations can serve as neighborhood centers and encourage dense, transit-oriented development [63, 64]. Additionally, transit users often complete the first and last mile of public transit journeys by foot or bike. If individuals are able to conduct their essential activities within 15-minute walks of transit stations, the transit network itself can serve as the backbone for a polycentric city [72]. However, limited work has been done on mobility-sensitive food accessibility of transit users. This population faces unique spatiotemporal constraints in their travel behavior due to fixed routes and timetables of transit and therefore have a more constrained choice set of sources of food [73, 74]. Few studies have considered human mobility in the context of food access via public transit. One study quantified accessibility by considering spatiotemporal and built environment constraints of a transit-reliant community, though it required detailed origin-destination flows of transit commuters at the TAZ level [75].

Leveraging big data can answer questions about individual, trajectory-based food accessibility for public transit users at a larger scale. In particular, smartcard data offers insight into transit users' mobility patterns and can quantify how well the transit system connects individuals to essential amenities such as grocery stores. Additionally, smartcard data is passively generated, which reduces the burden of data collection. This also means that many data points may be associated with each individual, which allows for mining of travel behavior patterns. However, several challenges with smartcard data exist. First, sociodemographic information is typically not available and information such as people's home or destination locations must be inferred [76]. Second, the amount of data is quite large and may capture behavioral outliers. The data must be efficiently processed to extract key information.

The above mentioned urban theories often center accessibility around proximity, though proximal locations may not necessarily be accessible. In the context of food access, affordability and operating hours of stores are important factors in addition to proximity. For example, dense city centers where 15-minute city goals are met may also have higher land values, resulting in higher priced amenities. These areas may have high proximity to amenities, but low affordability. Affordability should also be included in the conversation of the 15-minute city. Past efforts to incorporate affordability into food accessibility studies are often limited by spatially aggregated analysis. For example, an analysis performed between an inner city African American neighborhood and its mixed-race suburban neighbor in Chicago, Illinois found that the inner city neighborhood had competitively priced grocery stores, though the items were of poor quality [77]. Conclusions could not be drawn at the individual level. Additionally, temporal alignment with store operating hours should be considered, especially for transit users who are constrained by transit operating schedules. A study in Cincinnati, Ohio found that transit travel times between a census tract and its nearest supermarket varied at different times of the day [73].

Given the limitations of the accessibility literature, we propose a new trajectory-based accessibility measure which considers operating hour and affordability constraints. To do so, we aim to answer the following research questions:

1. How can we leverage smartcard big data to create trajectory-based accessibility measures?

2. Compared to traditional distance-based accessibility measures, to what extent does considering affordability and temporal availability in addition to proximity along a trajectory change the number of accessible grocery stores?

3. How well do people's distance-based accessibility and affordability-based accessibility align? Can we observe areas where distance-based accessibility may be high though affordability is low?

3.2 Accessibility Background

Broadly, accessibility is defined as the way land use and transport systems enable people to reach activities or destinations [6]. The utility of accessibility measures can address positive or normative needs. Normative accessibility prescribes an expectation of the world, in this context, how far people should travel. Positive accessibility describes observed behavior, in this case, how far people actually travel [78]. Distinguishing these two utilities is essential for selecting an appropriate measure of accessibility, which will be discussed in-depth in Section 3.3.

For both positive and normative approaches, accessibility can be measured from two main perspectives: place-based or person-based [79]. Place-based approaches quantify the amount and distribution of supply and/or demand of opportunities surrounding a given location. Typically, the number of locations in a defined reachable threshold are counted, and may be weighted by a distance-decay function from the point of interest. [6, 79]. However, place-based methods have been found to either underestimate or overestimate true accessibility, as neither individual-level behavior nor temporal variations in activities are considered.

Person-based approaches consider accessibility from an individual level, account for temporal constraints for activities, and are founded in travel behavior theories [80, 6]. One such person-based approach is space-time accessibility. Space-time accessibility is built from Hagerstrand's concept of space-time constraints, which include capability, coupling, and authority [81]. Capability considers the limits of human activity (e.g., people must sleep every day); coupling considers that people must be at certain locations at certain times (e.g., eight hours are spent at the workplace); authority considers rules set which limit activity (e.g., operating hours of the public transit system). Space-time accessibility uses these concepts to capture individual, transportation, and land use constraints at an individual level [11, 82, 83, 84]. Though more realistic than place-based approaches, these methods are much more data intensive, requiring information on transportation system travel times and individual travel diaries [85, 79]. With improvements in technology, passively-collected human mobility big data allows for easier collection of individual-level activities. This mobility data spans from GPS traces from cellphones to transit smartcard records [85].

There are emerging studies which leverage big data to assess person-based accessibility. A study in Shenzhen, China used mobile data to created space-time prisms along individual trajectories to count the cumulative number of food service facilities across a user's prisms [16]. Other studies using mobile data in the U.S. measured dynamic, trajectory-based environment exposure to fast food and assessed 15-minute city ideals by measuring access to essential amenities and 15-minute walking trips. [17, 86]. We build off these person-based frameworks to propose a novel accessibility measure and leverage human mobility big data to do so. In this study, we build from the idea of a fifteen-minute city and its predecessors to determine opportunities for food access of transit users while considering operating hours of amenities and affordability.

3.3 Methodology

3.3.1 Accessibility Measures

In this study, we present a trajectory-informed accessibility measure suitable for person-based interpretation that joins normative and positive ideologies. Accessibility is calculated given individual level mobility patterns and constraints from the built environment. Observed travel patterns ground this measure in reality, though an opportunistic perspective is used to measure the potential for access along visits in a trajectory. Two temporal constraints are considered:

1. **Maximum travel time budget**: An opportunity is accessible only if it can be reached within the maximum travel time budget of an individual. In compliance with a 15-minute city ideal, a walking time budget of 15 minutes is chosen. However, this budget

can be easily modified to meet x-minute city goals.

2. **Operating hours**: An opportunity is only accessible if it is visited during its operating hours.

For each individual, s, we observe a trajectory, g_s. Each trajectory is made up of N_s visits, where the n^{th} visit is written as $g_{s,n}$. A visit is defined as a point in both space and time. Let $t_{s,n}$ be the associated timestamp for $g_{s,n}$. We generate a travel shed (TS) for each visit along the trajectory, as activities can be conducted in the area surrounding each visit location. A travel shed is a spatial representation of the extent that an individual can travel, given some time budget. The travel shed at each visit is created with respect to the maximum travel time budget.

An opportunity, k, is said to be feasible for individual s if it is located within a travel shed of the individual and meets the operating hour constraint. Let l_k be the location of opportunity k and let $[t_{\text{k, open}}, t_{\text{k, close}}]$ be the operating hours. An indicator function, $I_{s,n}(k)$, determines if an opportunity k is feasible for $TS_{s,n}$:

$$I_{s,n}(k) = \begin{cases} 1 & \text{if } l_k \in TS_{s,n} \text{ and } t_{\text{k, open}} \leq t_{s,n} \leq t_{\text{k, close}} \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

Two indices are used to calculate accessibility at the person-level:

1. **Unique number of opportunities:**

   We identify all opportunities which satisfy $I_{s,n}(k) = 1$. The cardinality of all stores across all travel sheds is the number of unique opportunities A_s for individual s. In other words,

$$A_s = \left| \bigcup_{n=1}^{N} \{k \mid I_{s,n}(k) = 1\} \right| \tag{3.2}$$

2. **Weighted average of opportunities:** Accessible opportunities from each visit can also be weighted by the frequency by which a visit occurs. A visit with high frequency may be perceived as more familiar and therefore more likely to facilitate other activities

[87]. The weighted average of opportunities, WA_s, is weighted by the frequency of a visit occurring, $r_{s,n}$.

$$WA_s = \frac{\sum_{n=1}^{N} r_{s,n} \sum_k I_{s,n}(k)}{\sum_{n=1}^{N} r_{s,n}} \tag{3.3}$$

Both these measures are easily interpretable as numbers of opportunities. Equation 3.2 is interpreted as the total number of opportunities accessible to an individual along a daily trajectory. Equation 3.3 is interpreted as the average number of opportunities accessible to an individual at a single visit along their trajectory. Each of the measures captures different dimensions of travel behavior. Equation 3.2 is a function of how many different locations an individual visits; the more spatial areas a person covers in their trajectory, the higher the unique number of opportunities. Equation 3.3 considers behavioral regularity. Accessibility potential is weighted higher at frequent visit points (e.g., home or work).

3.3.2 Trajectory Processing

Mobility data can potentially be made up of many visits over the course of many days, requiring intensive computational power to analyze these points. However, because of potential regularity in an individual's behavior, these visits can be simplified [88]. Visits that are similar to other visits in both spatial and temporal characteristics can be clustered together and represented with a single visit. The clustered visit locations together form a "representative trajectory" which conveys a person's typical activity patterns. The size of each cluster represents the frequency of a visit occurring in space and time, also known as behavioral regularity. The visits in a representative trajectory can be weighted to place importance on visits with high regularity.

For example, a person may use transit to commute from home to work each day. Smartcard data observations may show visits at several different bus and metro stations near their home between 7 a.m. to 8:30 a.m. and visits at the metro station near their work from 4:30 p.m. to 6 p.m. This person may also sporadically take trips at lunch time, though it occurs less than once a week. Their representative trajectory would only show two simplified visits: one visit at the average location of bus stops near home at 7:45 a.m. and one visit at the metro station near work at 5:15 p.m.

Based on people's behavioral regularity, the construction of a representative trajectory identifies locations and times suitable for accessing additional opportunities which fit into pre-existing activity patterns. The representative trajectory also removes noise from travel patterns. This process mines key information from large datasets, which decreases computational time for subsequent trajectory analysis and also serves as a privacy preserving measure in human mobility analysis. It should be noted that each simplified visit location is an average of spatiotemporally similar visits. In the example above, each simplified visit in the representative trajectory is the average of the clustered transit stop locations and no longer represents the physical location of an actual stop.

Past efforts have used clustering methods to simplify people's trajectories from big data. Temporal, modal, and service regularity in London was identified by clustering smartcard entries [88]. A study in Seoul, South Korea used smartcard data to identify night shift workers by their commute patterns [89]. Both these studies employed DBSCAN, a density-based clustering algorithm. Unlike centroid-based clustering (e.g., k-means), DBSCAN identifies clusters by detecting regions of high point density [90]. This allows for clusters of non-spherical shapes. Another benefit of DBSCAN is that it does not require prior knowledge of the number of clusters, but rather relies on a distance threshold, ϵ, and minimum number of points to identify a point as belonging to a cluster [90].

To adapt DBSCAN to a spatiotemporal context, ST-DBSCAN was developed. ST-DBSCAN extends DBSCAN to include separate temporal and spatial ϵ parameters to set a joint threshold for spatialtemporal similarity [91]. Additionally, ST-DBSCAN allows for clusters of different densities. In a travel behavior context, ST-DBSCAN has been used to cluster social media data to infer traffic conditions and to identify activity regions using smartcard data [92, 93]. ST-DBSCAN requires the specification of four parameters: ϵ_1 and ϵ_2, the spatial and temporal distance parameters; min pts, the minimum number of points to form a cluster, and a distance function (e.g., Euclidean, Manhattan, Minkowski, etc.).

In this study, we run the ST-DBSCAN algorithm for each individual, s, on all of their observed visits, g_s. This produces N_s unique spatiotemporal clusters and the spatiotemporal centroid of each cluster is calculated, which we call $\mathbf{c}_{s,n}$. $\mathbf{c}_{s,n}$ is a 4-dimensional vector containing information on the location (latitude, longitude), timestamp, and size of cluster,

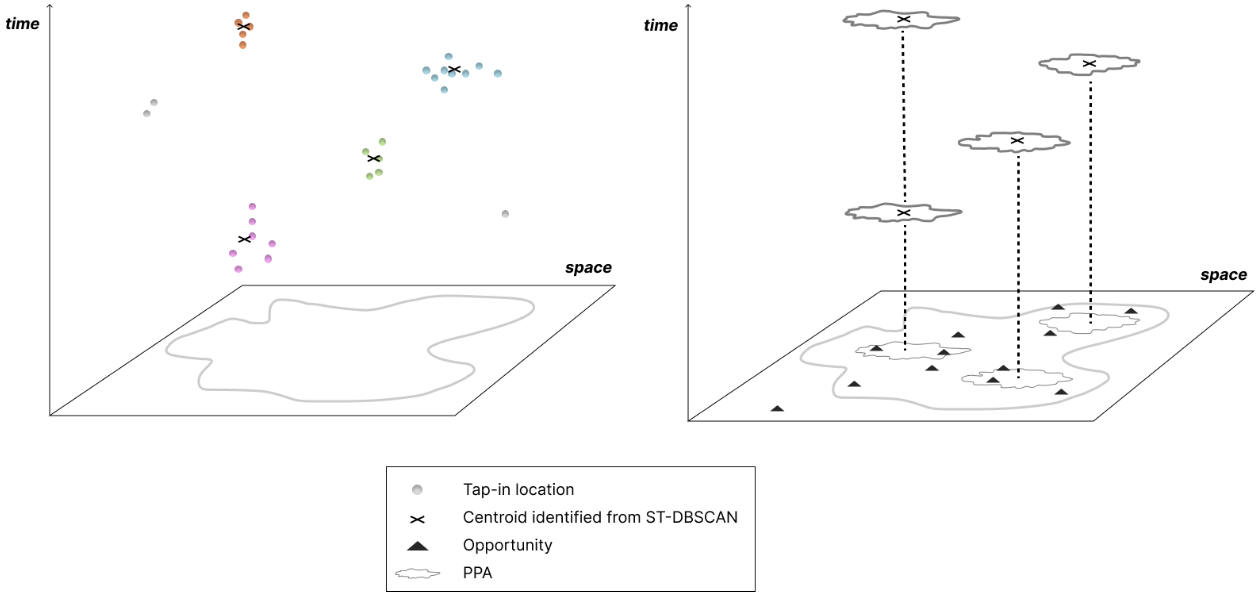


Figure 3.1: ST-DBSCAN clustering process (left) and travel shed generation (right).

or $\mathbf{c}_{s,n} = (\mathrm{lat}_{s,n}, \mathrm{lng}_{s,n}, \mathrm{t}_{s,n}, \mathrm{r}_{s,n})$. The set of all centroids then make up each person's representative trajectory, or g_s^*. Thus $g_s^* = \{\mathbf{c}_{s,1}, \mathbf{c}_{s,2}, \cdots, \mathbf{c}_{s,n}\}$. This representative trajectory can then be used as an input into the accessibility calculations from Section 3.3.1. Figure 3.1 shows how ST-DBSCAN is used to generate travel sheds.

3.4 Case Study: Grocery Store Access in Stockholm Region

This study uses transit smartcard data from Stockholm Region, Sweden to measure accessibility of grocery stores. We use Stockholm Region as a case study due to its high public transit usage rates and unique state of food affordability. We leverage the idea of a fifteen-minute city as an ideal, from the notion that essential goods should be within a fifteen-minute walking distance from a person's trip origin [66]. We employ our individual, mobility-sensitive, and temporally constrained accessibility measure and compare it to traditional home-based methods. Figure 3.2 provides an overview of the data processing and accessibility calculation workflow.

Stockholm Region is home to 2.46 million individuals. The majority of the population is concentrated in the Stockholm urban area; other cities and towns in the region are quite

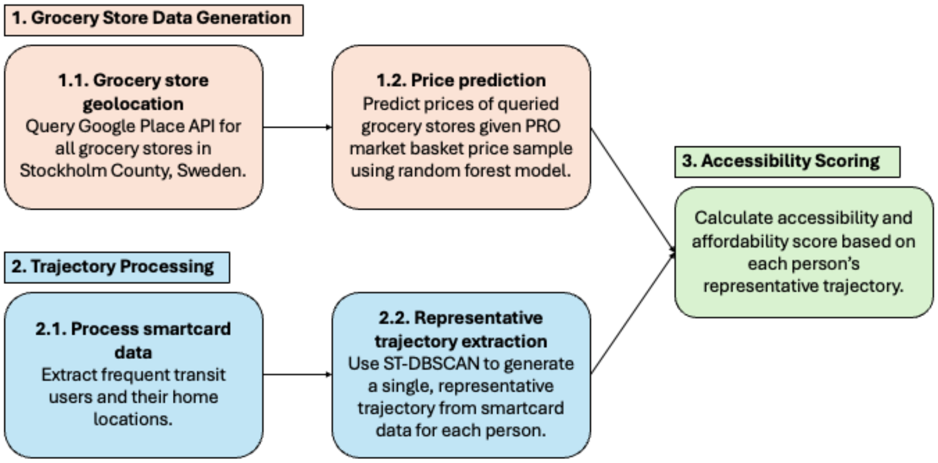


Figure 3.2: Data processing and accessibility calculation workflow.

isolated, separated by rural areas. Public transit is operated by Region Stockholm Transport Administration (SL) and includes bus, light rail, tram, metro, commuter rail and ferry. In the Stockholm urban area, the mode split is 32% public transit, 15% walking, 7% biking, and 46% driving [94]. A study region map with transit stops are shown in Figure 3.3. Transit services are clustered around Stockholm City, with connections to other urban areas by rail and bus.

With regards to food access, Sweden is a welfare state and food access has not historically been studied in-depth [95]. One study in Sweden found that disadvantaged groups such as low-income individuals, elderly, and single parents were found to live closer than average to the nearest food shop [96]. Availability of food is clearly not an issue due to Sweden's liberal food retail planning policy; however, affordability has become a rising issue. The Swedish food retail market is dominated by a few large chains: the three largest retailers comprise 90% of sales [97]. These chains therefore have significant control over prices and food costs have risen dramatically over the past few years [98]. Food prices have risen 25% from August 2021 to May 2024, though prices have recently stabilized [99].

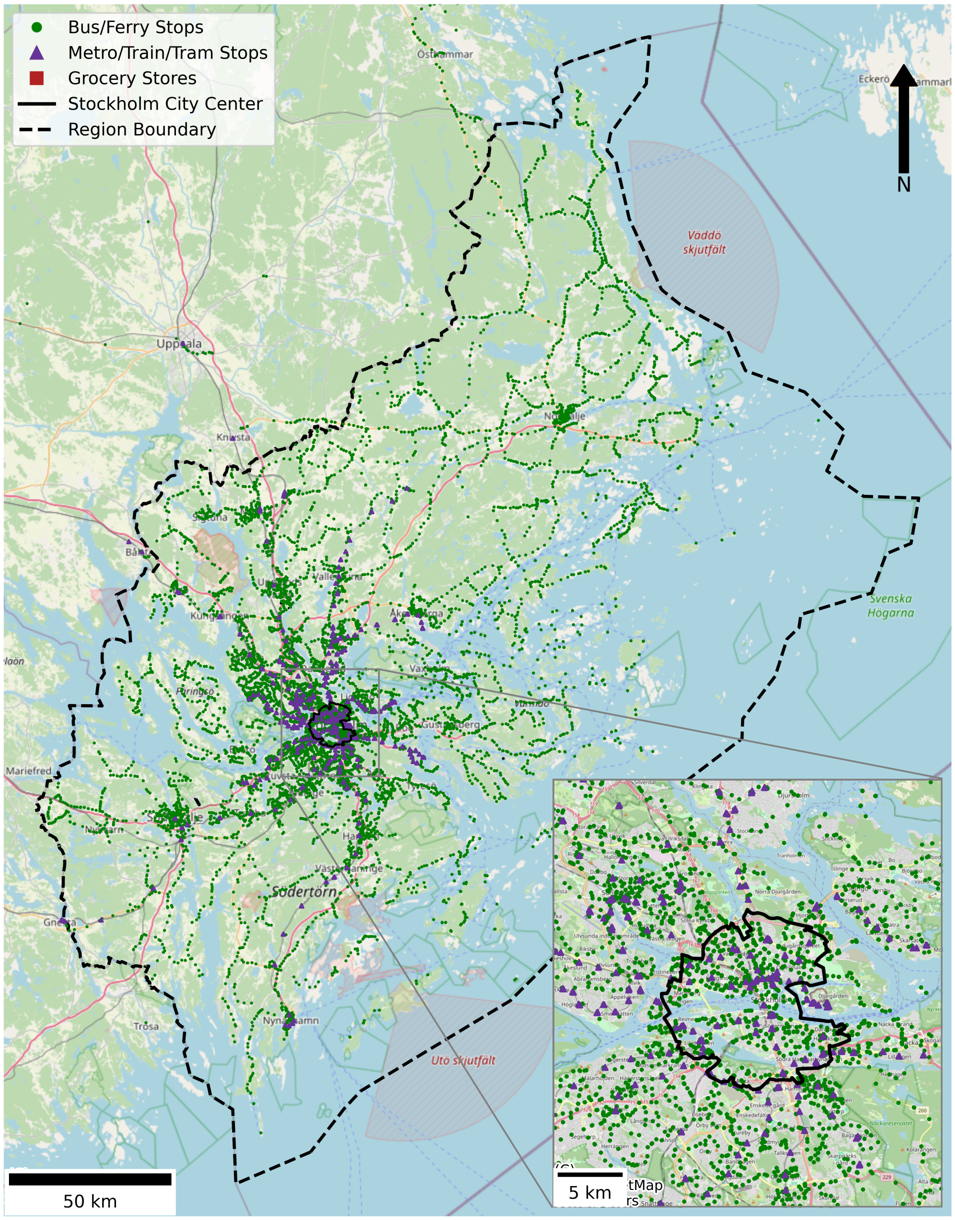


Figure 3.3: Transit station distribution throughout Stockholm Region.

3.4.1 Datasets

Smartcard Data

Transit smartcard data from SL is used in this analysis. Only tap-ins are recorded. The study period is a month-long period from September 11, 2023 to October 8, 2023, which does not have any public or school holidays. During this month, 2,276,011 unique users made a total of 48,273,937 trips. Frequent transit users, specifically those who used transit for 20 or more days per month, were extracted. Frequent users are more likely to be familiar with the areas surrounding transit stations and more likely to conduct activities such as grocery shopping via public transit. Additionally, travel patterns of frequent users can be identified due to their higher transit trip rates. Regularly visited locations and trip times can be extracted with higher confidence. During the study period, 272,584 frequent users (12.0% of total users) made 20,634,575 trips (42.7% of total trips), with a median of 70 trips per person during the study period.

Grocery Store Data

Grocery store location, price, and opening hours are obtained from a combination of the 2023 national survey on food prices conducted by the Swedish National Pensioner's Organization (PRO), and Google Place API. Each year, PRO conducts a market basket price survey at a sample of grocery stores throughout Sweden [100]. The most recent survey from 2023 samples a total of 46 items at 622 Swedish grocery stores, with 105 stores located in Stockholm Region. The PRO dataset is a rich source of food price information, with the total cost of each market basket along with the specific price of each item. The median price of a market basket is 1546 Swedish crowns (SEK), which is approximately $155 USD, and prices ranged from SEK 1352 to SEK 1657.

3.4.2 Grocery Store Data Generation

The limited PRO sample size poses a challenge to understanding the overall state of food accessibility and affordability in the Stockholm region since it does not capture the full choice set of stores for grocery shopping. The full set of store locations and prices is needed to accurately measure the true state of food accessibility. Thus, we leverage the Google Place API to obtain the location of all grocery stores in Stockholm Region, along with store characteristics, to aid in predicting market basket prices.

First, we query Google Place API for the location of chain grocery stores in Stockholm Region. The query area of Stockholm Region was buffered by 5 km, so that stores located slightly outside the county were also considered. This is important to avoid arbitrarily lower accessibility at county lines due to edge effects. A grid of overlapping circles of radii 3.1km are then generated to cover the buffered area, and queries using a search term of ”grocery store” are performed for each search area. After cleaning the queried stores (query results such as campgrounds and libraries were removed), 1096 unique stores were returned. Several sanity checks were performed: firstly, all stores in the PRO sample were successfully queried. Secondly, a comparison was made to King County, Washington in the U.S. King County has similar characteristics to Stockholm Region in terms of population and area– King County has a population of 2.27 million individuals and an area of 5,480 square kilometers [101]. Stockholm Region has a population of 2.46 million individuals and an area of 6,514 square kilometers [102, 103]. From their food inspection records, King County has a total of 878 grocery stores, which is similar to the number queried from Stockholm Region [104]. Figure 3.4 below shows the results of the API query.

We then further filter the dataset to only include chain grocery stores, which made up 49.8% ($n = 546$) of the queried stores. This was done for several reasons. First and most importantly, the Swedish retail food market is dominated by chains, which make up 90% of the market share [97]. Second, many of the non-chain stores queried from Google Place API, such as ethnic food stores, were much smaller in size and did not carry fresh produce or dairy. This means that those who shop at smaller stores may still need to visit larger grocery stores to obtain all needed food items. Third, the PRO price survey only sampled chain grocery stores. For greater confidence in predicting prices of grocery stores, we choose to focus solely on chain stores.

However, excluding non-chain grocery stores has implications on the subsequent results of the study. Swedish society has historically been racially and culturally homogeneous – in 2002 11.8% of the population had parents which were both Swedes. Naturally, the products

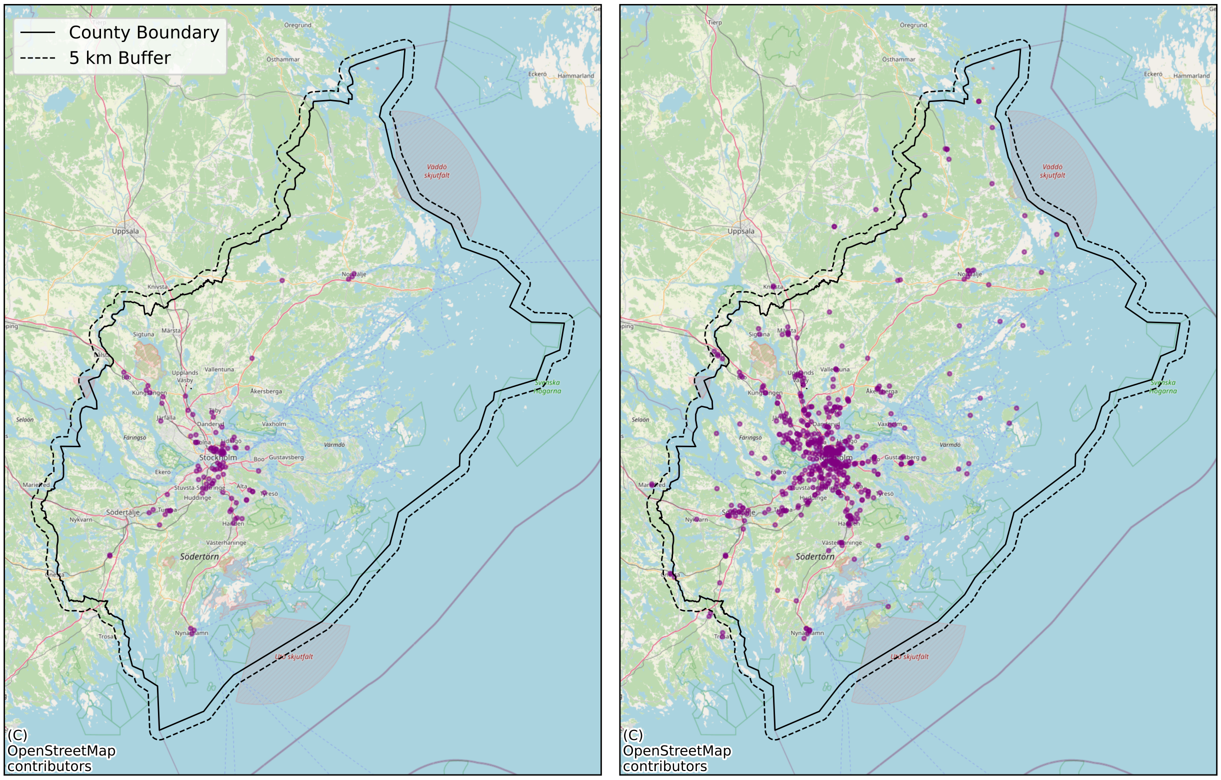


Figure 3.4: Chain grocery stores from PRO dataset (left, $n = 105$) and results from Google Place API query (right, $n = 546$).

sold in chain grocery stores meet the cultural needs of the dominant population. However, the immigrant population in Sweden has risen over the past two decades. In 2023, 20.6% of the population is non-European [103]. Of the non-chain stores excluded from the study, many were ethnic food stores which met the needs of the emerging immigrant population. Though these stores may not provide the full range of grocery items needed in a week, they are a resource for culturally-relevant food items. Excluding these stores as accessible options neglects to account for the cultural and personal preferences of a diversifying Swedish population.

Next, we use the sampled PRO market basket prices ($n = 105$) to predict the prices of the unsampled queried stores ($n = 441$). From the Place API query, we obtain information about the location of a store, the chain name, user rating, number of user ratings, and its Google tags (e.g., "grocery or supermarket", "pharmacy", "shopping mall", etc.). Chain store characteristics such as "discount", "hypermarket", and "supermarket" are also used for predictions. A random forest regression model is trained on these characteristics to predict the price of non-sampled stores. A random forest model is a tree of regression models trained with a subset of both model variables and bootstrapped samples of data [105]. This makes it robust to overfitting, which is especially important because of the limited size of our labeled dataset.

Model hyperparameters are tuned and result in an out of bag score, or average regression R^2 value, of 0.67. The root mean squared error between predicted and actual values was 57.91 Swedish krona (SEK). When normalized, the RMSE is 3.88% of the mean market basket price. The distribution of predicted prices is similar to those in the PRO sample, as can be seen in Figure 3.5. The median price of a market basket was SEK 1527, which is approximately \$150 USD, and prices ranged from SEK 1352 to SEK 1657.

3.4.3 Trajectory Processing

Smartcard data is processed to identify frequent transit users and their home locations. The home location is inferred as the most common first tap-in location of the day [76]. Once smartcard data for all frequent transit users is queried, a representative trajectory is generated for each person using ST-DBSCAN on their historical tap-in locations.

Ranges for the spatial and temporal ϵ are selected using domain knowledge. From DBSCAN applications using smartcard data, the spatial epsilon is suggested to represent two times the average distance between two transit stations [93]. The temporal epsilon should be reflective of the travel behavior regularity in the region [89, 93, 88]. Using these guidelines, we set ranges for both the spatial and temporal distance parameters. The spatial range is set by identifying the median distance between transit stations in the Stockholm region, which varied greatly between the highly dense city center and rural areas. The temporal range is set by analyzing behavioral regularity. Parameters are set to $eps_1 = [200m, 450m]$ and $eps_2 = [30 \text{ mins}, 120 \text{ mins}]$ in our study. The minimum number of points per cluster is set

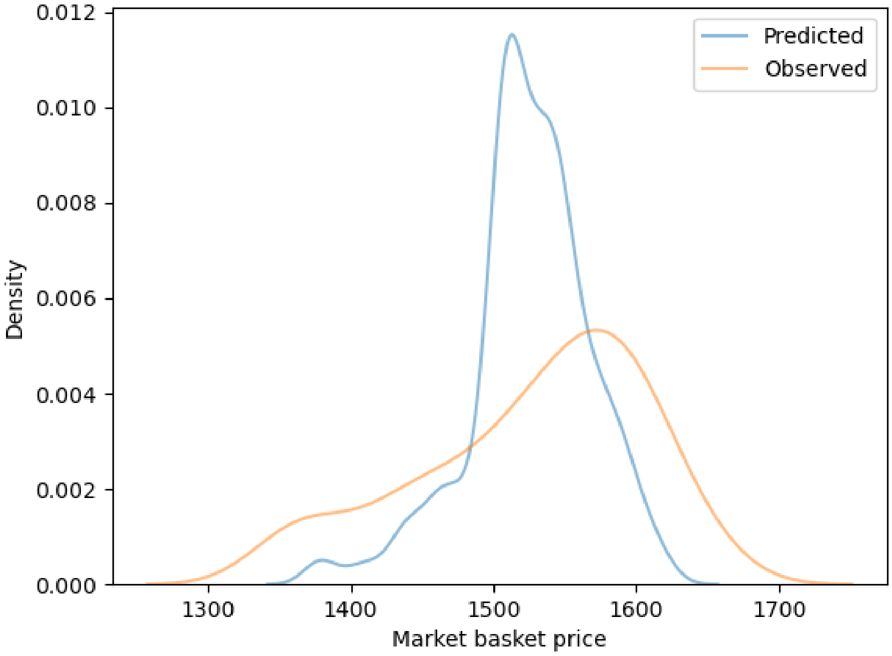
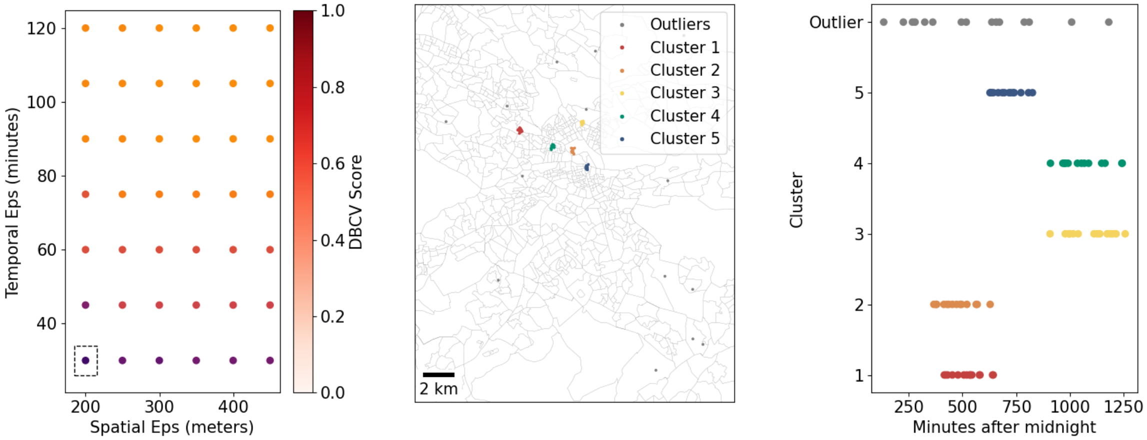


Figure 3.5: Comparison of price index distribution between observed and predicted datasets.

to 4, or a visit once per week. Euclidean distance is used to represent the ”closeness” of two smartcard tap-in locations.

Behavioral regularity varied greatly across individuals. Some individuals had high regularity (e.g., visiting the same transit stations at the same time of day) while others had low regularity (e.g., never re-visiting the same transit station and traveling at different times each day). Combinations of low spatial regularity and high temporal regularity (and vice versa) were also observed. For this reason, the spatiotemporal hyperparameters are then tuned within their respective ranges at the individual level. A grid search of the hyperparameters is performed and the combination that results in the highest density-based cluster validation (DBCV) score is chosen. DBCV scores range between (-1, 1), where an ideal score of 1 indicates high density within a cluster and low density between clusters [106]. A sample ST-DBSCAN hyperparameter tuning and clustering result is show below in Figure 3.6.



(a) Grid search to identify optimal $\epsilon_1 = 450$ and $\epsilon_2 = 120$.

(b) Spatial clustering result.

(c) Temporal clustering result.

Figure 3.6: Example hyperparameter tuning using grid search with ST-DBSCAN results.

3.4.4 Accessibility Measures

After running ST-DBSCAN for each person, we then calculate accessibility scores. For this case study, the two temporal constraints proposed in Section 3.3 are chosen as follows:

1. **Maximum travel time budget**: In the Stockholm region, transit and walking make up 47% of the mode share, and 84% of residents live in a 15-minute neighborhood [94, 107]. Building on Stockholm's 15-minute city and transit-oriented development goals, the maximum travel time budget was set to a 15-minute walk. In other words, grocery stores within a 15-minute walk of an individual's location are considered to be accessible.

2. **Operating hours**: Operating hours of grocery stores are obtained to determine temporal alignment. A grocery stores is accessible only if it is visited during its operating hours. Trips are assumed to be inherently constrained by public transit operating hours because of the nature of the dataset.

We also consider varying affordability. For each grocery store, k, we calculate an associated price index. The price index is a relative measure of affordability, as it defines affordability in comparison to the median price of a market basket. We use a market basket, which is a set of items typically bought at a grocery store, to avoid fluctuations in the price of a single item and to be more representative of a person's food budget. We define the price index at store k, PI_k, as:

$$PI_k = \frac{\text{cost of market basket, store } k}{\text{cost of market basket, median}} \tag{3.4}$$

where a $PI_k < 1$ is relatively affordable compared to the median, while $PI_k > 1$ is more expensive compared to the median. Though not explicitly incorporated into the accessibility measure, stores can be segmented by their relative affordability for both Equations 3.2 and 3.3.

The accessibility measures are then applied to measure food accessibility via public transit in the Stockholm region. For both the unique number of opportunities and the weighted average of opportunities, we measure four different categories of accessibility: all stores, stores which are open when visited, affordable stores, and stores which are both open and affordable. A store was considered to be affordable if its price index was greater than 1.

Accessibility measures are spatially aggregated by home zone location and reported at the DeSO level. DeSO zones are regional units at which demographic data is collected in Sweden. Median accessibility scores are reported at the DeSO level. The accessibility scores at the DeSO level reflect that of frequent transits users living in the zone, and not the entire zone population. Home-based and trajectory-based accessibility are compared. Home-based accessibility is measured by counting the number of stores reachable with in a 15-minute walk from the inferred home location.

3.5 Results

3.5.1 Accessibility distributions

Figure 3.9 shows the accessibility distributions of frequent transit users. Frequent transit users have a median of 5.18 stores reachable within a 15-minute walk at any visit along their trajectory and encounter a median of 15 unique stores along their entire trajectory. The operating hour constraint does not strongly constrain the number of opportunities– 95% of stores are open during the time of travel. However, the affordability constraint is much more limiting. Only 43% of all stores are affordable. Frequent transit users have a median of 2.31 affordable stores accessible at any visit along their trajectory and encounter a median of 6 affordable, unique stores along their entire trajectory.

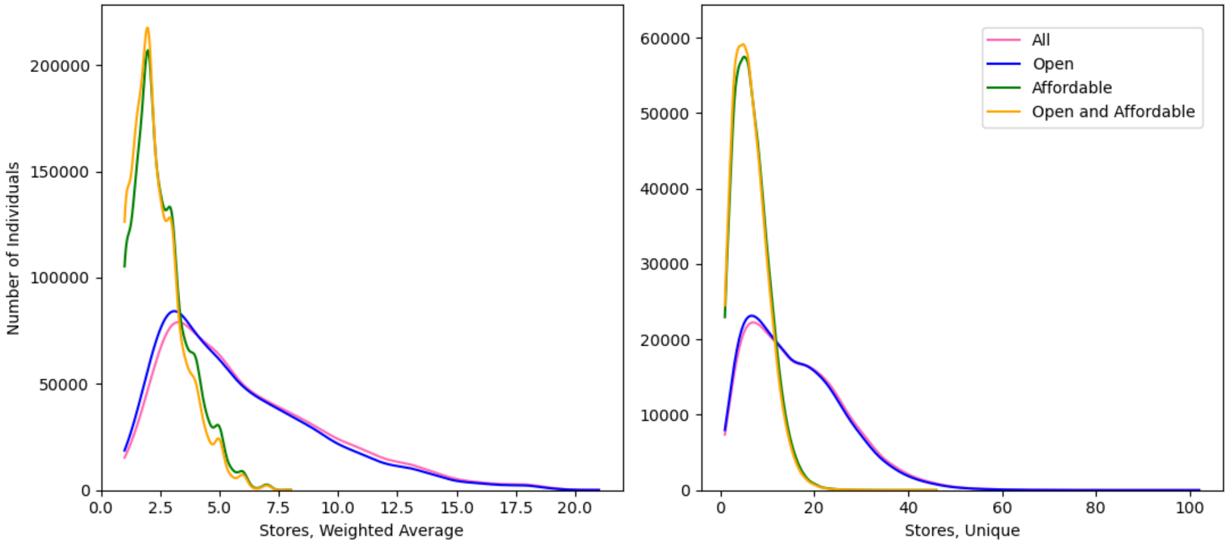


Figure 3.7: Number of stores accessible by 15-minute walk with operating hour and affordability constraints.

3.5.2 Sensitivity to x-minute city goals.

Various walkshed distances are tested to determine the sensitivity of accessibility measures. Walksheds of 5, 10, 15, 20, 25, and 30 minutes are tested. As can be seen in Figure 3.8, as walkshed sizes increase, so do the number of accessible grocery stores. Especially with the number of unique stores, diminishing returns are observed as the walkshed size grows. This indicates that with a large 30-minute walkshed, a person may be able to access the same grocery store from two different visits along their trajectory. Ultimately, a 15-minute walkshed is used for the rest of the study to align with 15-minute city development goals. 15 minutes is also deemed to be the maximum reasonable distance to walk to a grocery store. However, the appropriate walkshed size to measure accessibility will vary based on a region's built environment. Smaller walksheds may be appropriate for dense, urban areas such as

Stockholm City. Larger walksheds may be more appropriate for rural or suburban areas.

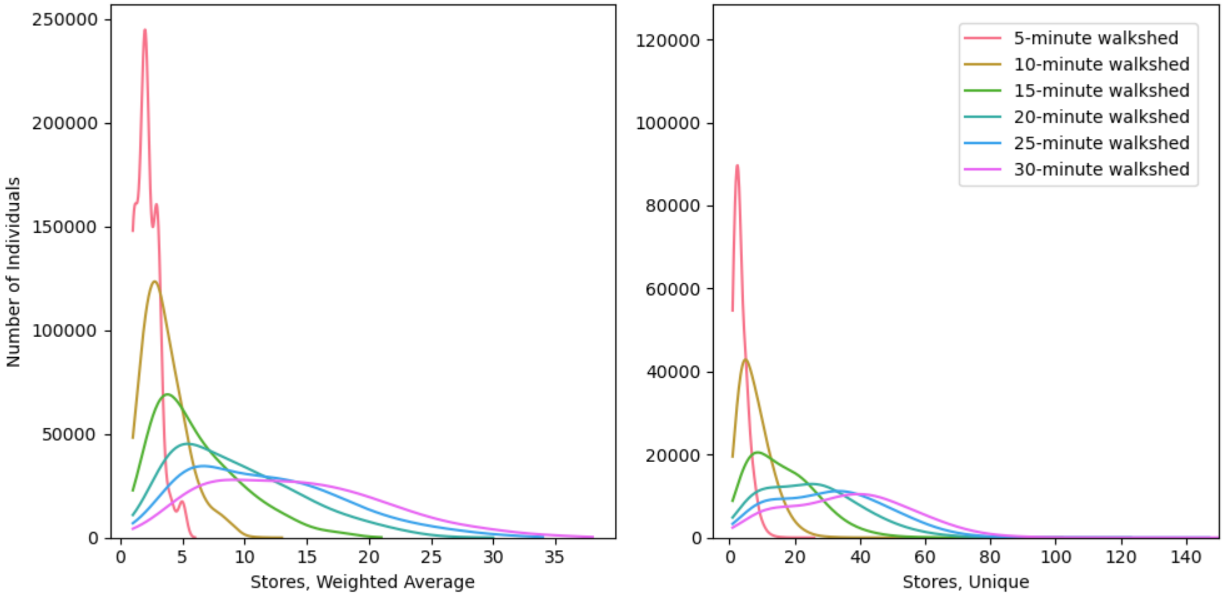


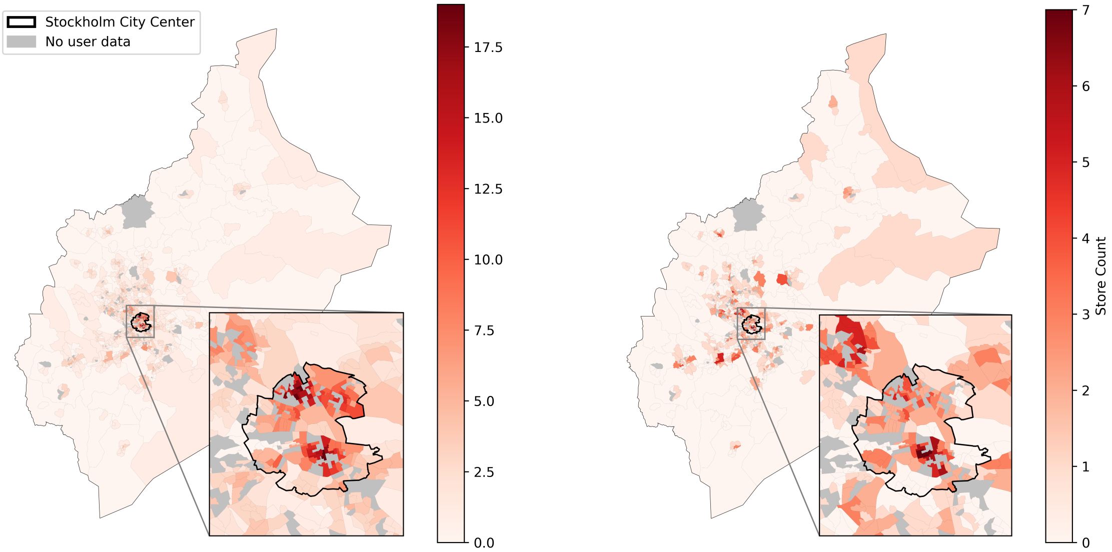
Figure 3.8: Sensitivity of trajectory-based accessibility measures to x-minute city goals.

3.5.3 Home-based accessibility

At the home-based level, we can observe the highest accessibility in Stockholm City Center, as seen in Figure 3.7. Clusters of high accessibility are observed in the other more isolated cities in Stockholm Region such as Södertalje, Norrtälje and Västerhaninge. However, the distribution of affordable stores differs. Affordable stores are concentrated in Solna and Haninge, which are municipalities located north and south of the city center, respectively. The southern portion of the city center has many stores which are both reachable within a 15-minute walk and affordable.

3.5.4 Trajectory-based accessibility

The trajectory-based accessibility measures capture people's mobility patterns. Because the mobility data is from frequent transit users, these patterns are reflective of public transport connectivity and usage. Figure 3.10 shows the distribution of open and affordable stores



(a) Total stores.

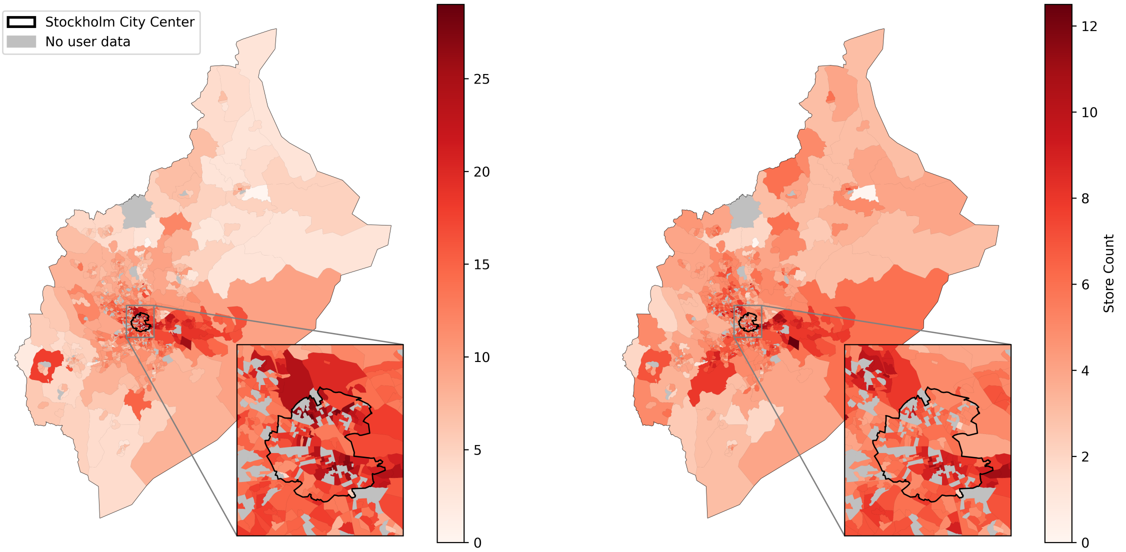
(b) Affordable stores.

Figure 3.9: Median number of stores accessible from 15 minute walk of home transit station.

based on the home location of frequent transit users. When considering people's mobility patterns, zones located in Stockholm City Center still have high levels of accessibility. However, other zones with high trajectory-based accessibility also emerge. For example, Södertäljie, which is located in the southwestern part of the region, has low home-based accessibility but high trajectory-based accessibility. This zone, and other areas of high trajectory-based accessibility far from the city center, are located near metro or commuter train stations. Frequent transit users who live in locations with low home-based accessibility may have high trajectory-based accessibility if they commute into the city or travel to other food-rich areas.

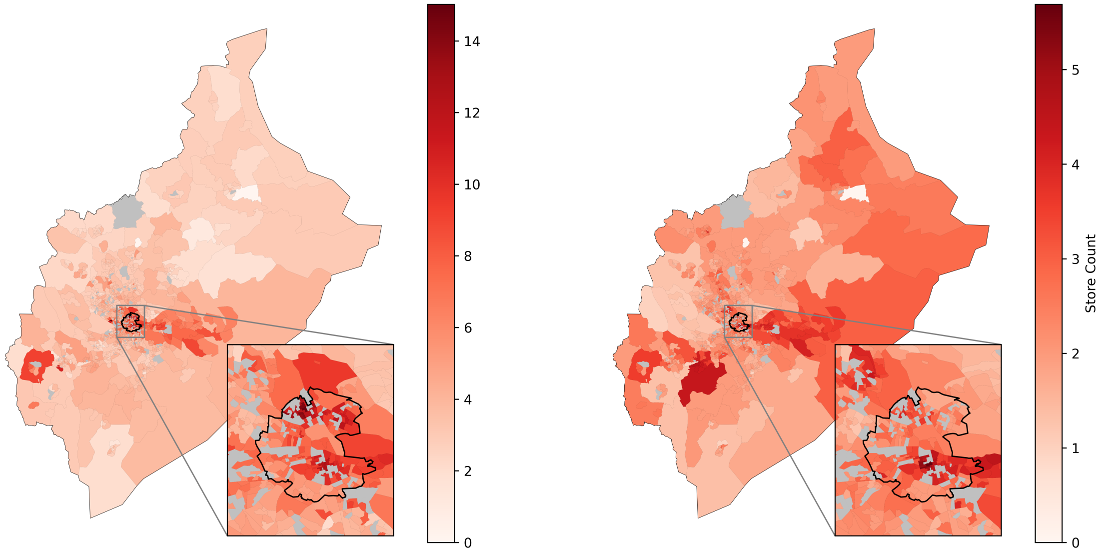
3.5.5 Comparing trajectory-based and home-based accessibility

Figure 3.11 shows the ratio of stores accessible along an entire trajectory to the stores accessible from home. Within Stockholm City Center, trajectory-based accessibility and home-based accessibility show little difference for frequent transit users. Those who live in the denser city center already have high accessibility at home, and do not observe a large "benefit" when accounting for mobility. However, frequent transit users who live in rural or suburban zones with low home-based access see a large positive effect when accounting



(a) Open stores, unique.

(b) Affordable stores, unique.



(c) Open stores, weighted average.

(d) Affordable stores, weighted average.

Figure 3.10: Median number of unique stores accessible within 15-minute walk of visits along trajectory.

for trajectory-based accessibility. Frequent transit users living near Gnesta, located by a commuter rail station in the southwestern portion of the region, encounter 35 times as many total unique stores and 17 times as many unique affordable stores when considering mobility. Using the trajectory-based measure increased the number of affordable stores by a median of 1.6 and 4.0 times for the weighted average and unique measures, respectively.

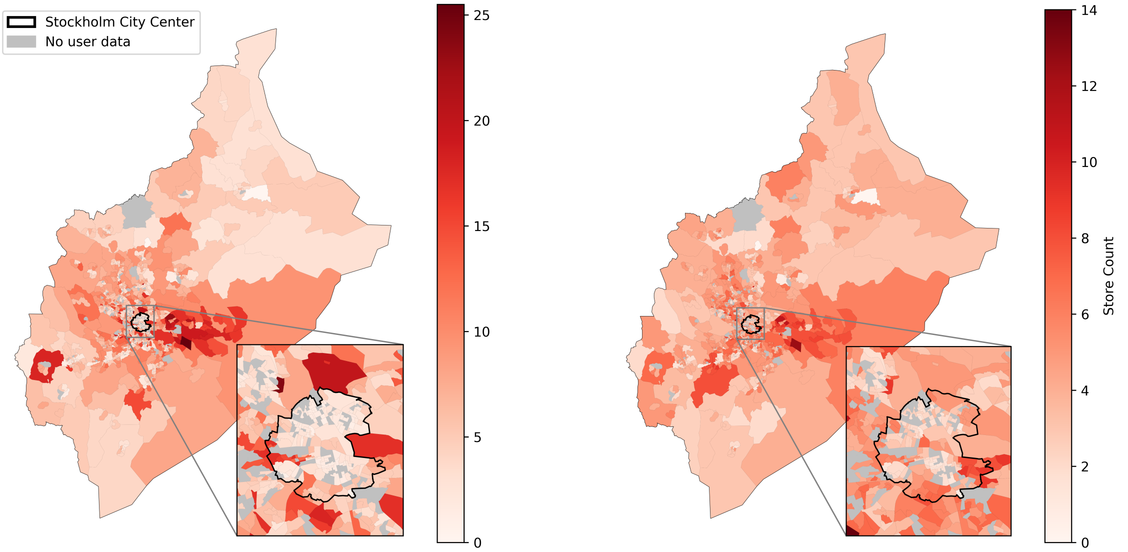
This indicates that studies should consider mobility to capture a fuller picture of true access, as it is known that food environments around the home can differ greatly than those along trajectories [17]. These findings align with a study in Shenzhen, China which found that those who lived in resource-poor areas saw positive increases from trajectory-based accessibility [16].

Typically, only low-home based zones are of concern when considering food policy. However, those with both low home-based and low trajectory-based access may be of highest priority when developing accessibility improvements. Figure 3.12 distinguishes two types of low home-based access of frequent transit users. Areas with high trajectory-based access but low home based access are located near or are connected by transit to urban areas. Zones with low trajectory-based access are located in rural areas which are not served by metro or commuter rail. An independent t-test is conducted to determine if a zone's proximity to transit is related to its trajectory-based food access. The unique accessibility measure is used. Proximity to transit is defined by the number of transit stops reachable within a 15-minute walk from a zone's centroid. Low home, high trajectory zones have significantly higher access to transit ($t(493) = -7.0, p < 0.05$). Again, this pattern emerges due to the nature of the dataset. Policy specific to improving transit connectivity in low home and low trajectory-based areas can be developed with regional planners.

3.5.6 Accessibility considering affordability

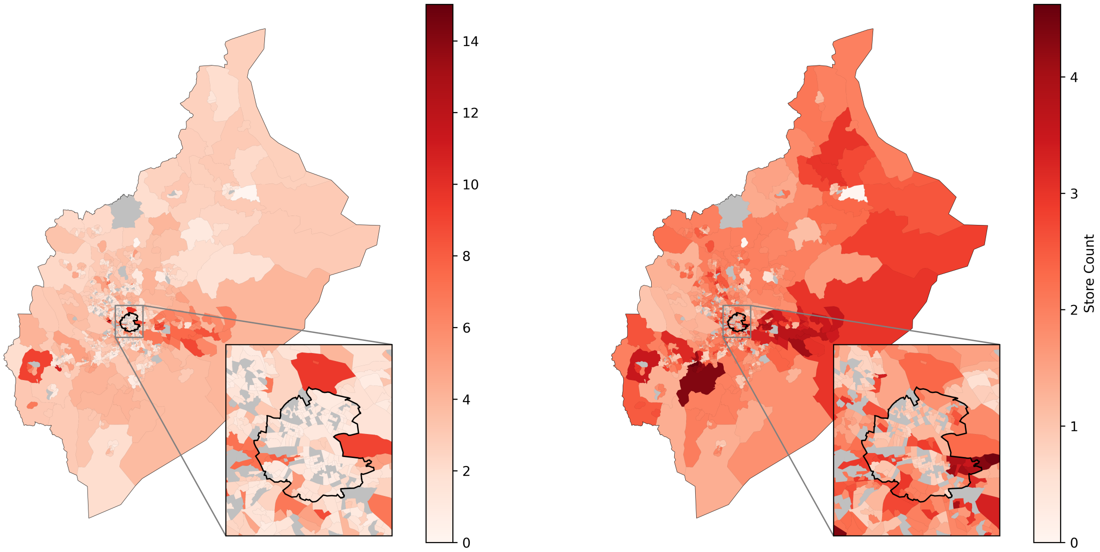
Trajectory and home-based accessibility are further compared by assessing four possible combinations of availability (number of grocery stores) and affordability. Relative accessibility is measured. Low availability is defined as a zone having less than the median number of accessible stores, aggregated by the median of all individuals living in a zone. Low affordability is defined as a zone having less than the median percentage of affordable stores, aggregated by the median of all individuals living in a zone. These combinations are:

- High availability, high affordability
- High availability, low affordability



(a) Open stores, unique.

(b) Affordable stores, unique.



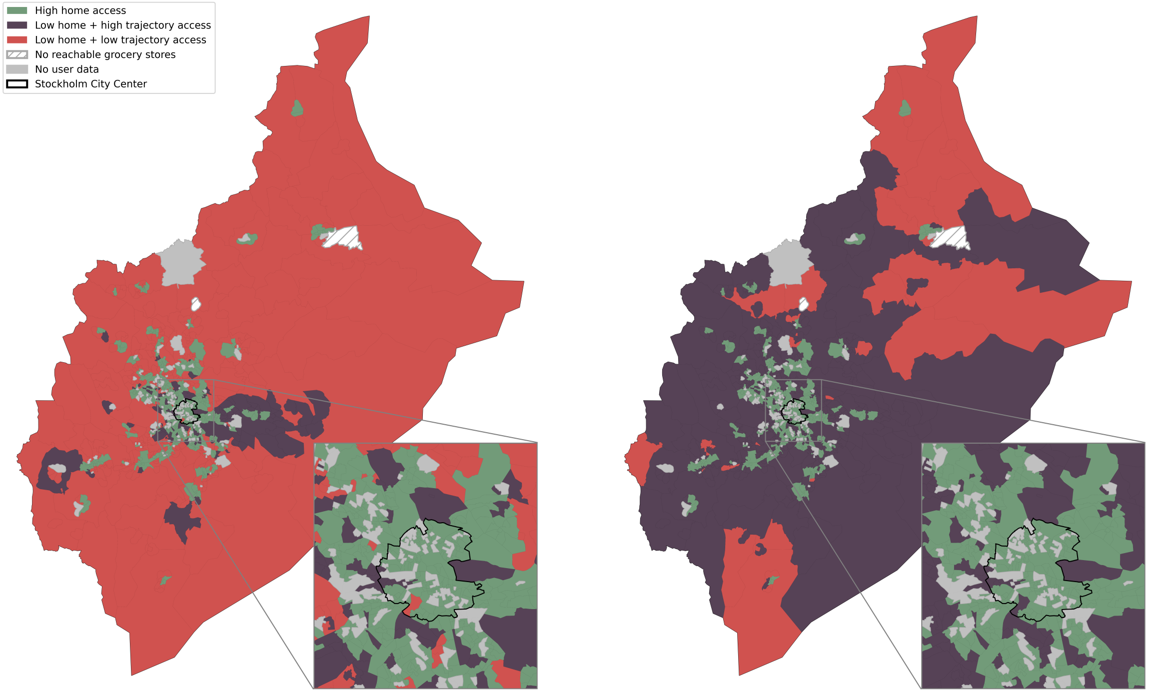
(c) Open stores, weighted average.

(d) Affordable stores, weighted average.

Figure 3.11: Ratio of trajectory-based access to home-based access.

- Low availability, high affordability
- Low availability, low affordability

Identifying high availability/low affordability areas may be of interest to policymakers,



(a) Unique.

(b) Weighted average.

Figure 3.12: Distinguishing locations with low home-based, low trajectory-based access and low home-based, high trajectory-based access.

as they reveal places that appear to have many options, but in reality may not be affordable for the people who live or pass through the area.

Figures 3.13 and 3.14 show the four combinations for both home- and trajectory-based accessibility based on frequent transit users' home zone. At the home-based level, Stockholm City has high availability of stores, though many stores are not affordable. The surrounding suburban areas have both high availability and high affordability. When considering trajectory-based accessibility, the high availability/low affordability areas grow to include other neighborhoods surrounding the Stockholm City Center. This indicates that the transit travel patterns of frequent transit users living in surrounding neighborhoods may often bring them into locations such as Stockholm City Center where food options are less affordable. Though the perceived supply of grocery stores is high, affordability limits the overall accessibility. Considering mobility also increases overall accessibility. Fewer frequent transit users have no reachable stores within a 15-minute walk. This indicates planning approaches focused on connectivity of urban hubs is a solution to improve accessibility for those who travel beyond their home neighborhoods, such as frequent transit users.

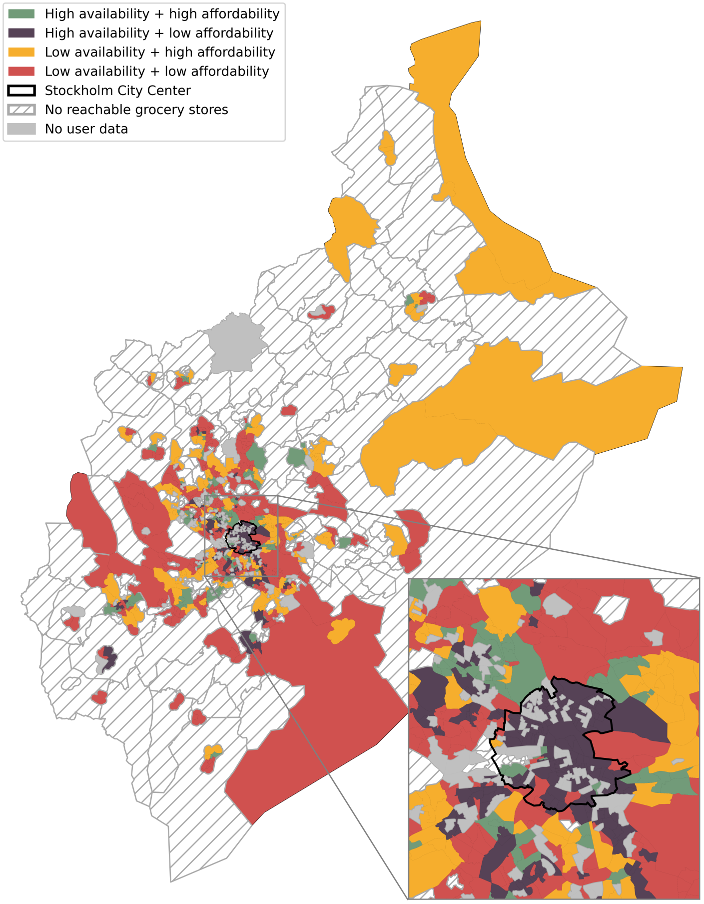
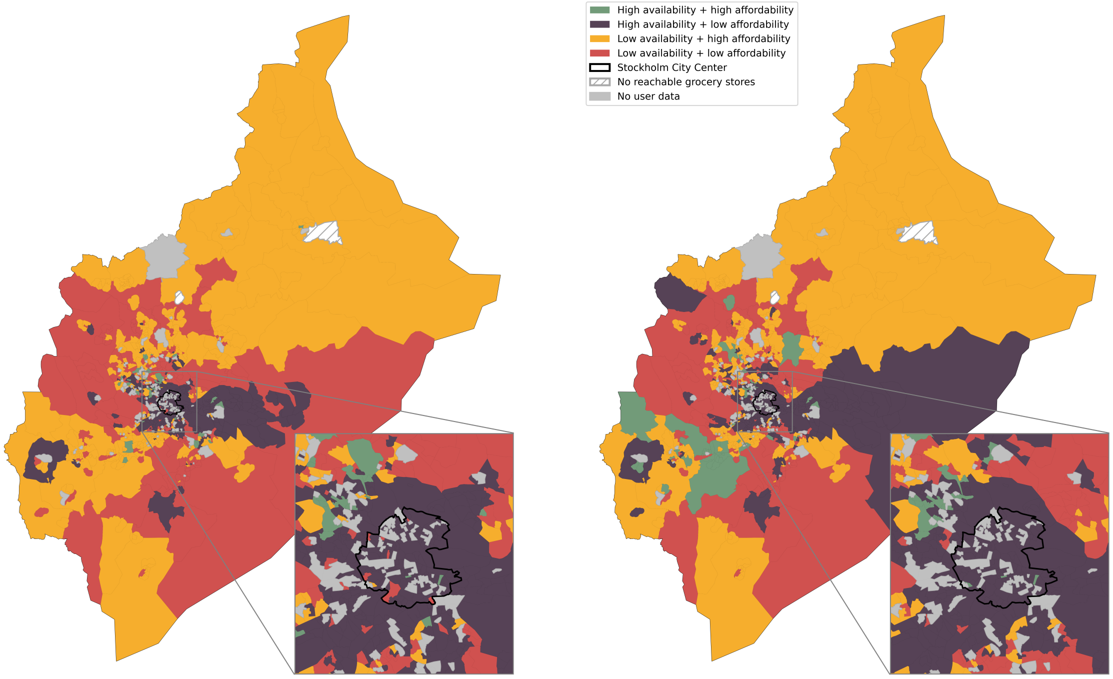


Figure 3.13: Home-based accessibility considering both availability and affordability.

3.6 Discussion

This study measures trajectory-informed grocery store access using smartcard data of frequent transit users in Stockholm Region, Sweden. A trajectory simplification method is introduced, which identifies opportunities for grocery shopping by extracting visits with high spatiotemporal frequency. Additionally, two person-level measures are developed to account for human mobility effects on accessibility. The methodology captures the potential to access opportunities, while still being grounded in people's actual travel behaviors.



(a) Unique.

(b) Weighted average.

Figure 3.14: Trajectory-based accessibility considering both availability and affordability

The applications of this method can help distinguish between true low accessibility and perceived low accessibility. Low home/low trajectory access and low home/high trajectory should be distinguished, as each group requires different interventions to improve access. For example, those with low accessibility both at the home and along their trajectories may benefit from improvements to local, home-based access or improved connectivity to high accessibility areas. On the other hand, those with low accessibility at home but high accessibility along their trajectories can benefit from improved accessibility at well-connected transport hubs (e.g., a central transit station). Similarly, high availability/high affordability and high availability/low affordability areas are separated to determine areas in need of affordable options.

The methods in this paper can also be used to test the effects of proposed policies. For example, impacts to trajectory-based accessibility can be modeled with the removal or addition of transit routes. Similarly, the methods presented can also quantify effects to accessibility from land use changes (e.g., new workplace locations, development of large multi-family housing,) which affect individual travel patterns.

It should be noted that trajectory-based accessibility should serve as a complement, not replacement, for home-based accessibility measures. Trajectory-based access shows a positive impact on those living in suburban and rural communities connected to urban areas. However, there may be individuals who live within these rural or suburban communities that work, shop, and conduct other activities within their home neighborhood. These individuals are not captured in trajectory-based accessibility. Additionally, communities which lack connectivity to resource-rich areas cannot benefit from trajectory-based accessibility. Therefore, local, home-based access should still be considered in food accessibility policy. Increasing connectivity by public transport or other modes could be a longer-term solution to improving food access. Stockholm region's 2050 plan sets goals for stronger links between urban and rural areas, which will help link people to food [108]. Additional collaboration with urban planners and food policy experts should be undertaken to integrate the transportation perspectives from our study with larger regional plans.

Throughout the paper, relative accessibility was used to showcase the capabilities of the trajectory-based accessibility measure. In other words, the results identified which groups were better or worse off compared to the overall population. However, absolute accessibility can also be measured to identify individuals falling below a certain threshold of access. The approach to measuring accessibility should be selected based on policy equity objectives. An egalitarian approach, which would ensure people have equal access to food, may target individuals with low home or trajectory-based access relative to the overall population. Relative accessibility measures, like the ones used in this study, are appropriate. However, a sufficientarian approaches are also valid, which ensure that all people have access to a minimum number of essential amenities [109]. Absolute measures should be used to meet sufficientarian policy goals.

3.7 Limitations and Future Work

Several limitations of the study should be acknowledged. First, the dataset only includes frequent transit users, or those who use transit 20 or more days a month. This likely means the dataset is more reflective of commuter behavior, which is not necessarily representative of the behavior of the general population. Additionally, since analysis is performed at the zone level, the results only represent the frequent transit users who live in the zones and not the entire zone population. Our study can be used as a framework for measuring trajectory-based accessibility, though developing policy from our specific results should be done with caution and with the above limitation in mind. Second, because of the anonymous nature of the smartcard dataset, individual characteristics of transit users are unknown. In particular, grocery shopping preferences such as convenience, brand loyalty, or price savings are unknown– these characteristics may strongly influence grocery shopping behaviors and cannot be inferred from the dataset. Last, weekday and weekend behavior are analyzed together. Because of the difference of activity purposes (e.g., commuting on weekdays and errand activities on weekends), grocery store-related travel patterns may vary greatly between weekdays and weekends.

Given the existing limitations, several lines of future work exist. Transit users with less frequent usage can be included in the analysis to make the results generalizable to a larger population. However, the representative trajectory generation process should be monitored to ensure it outputs accurate results, as regular spatiotemporal behavior may be more difficult to identify. Regression analyses can be conducted between zone-level accessibility and demographic characteristics. Although this analysis is still conducted at a spatially-aggregated level, it may help answer structural questions on access limitations based on socioeconomic status. Similarly, analysis can be done to understand the relationship between fare type (e.g., discount fares, single tickets, 30-day tickets, etc.) and accessibility levels. Weekend and weekday smartcard data can be separated to determine how accessibility differs on weekdays and weekends.

To conclude, the methods introduced in this study can be generalized to other sources of human mobility data and applied to measure the accessibility of amenities beyond grocery stores. This study helps bridge the gap between normative and positive ideologies of accessibility and leverages human mobility data to provide a more comprehensive understanding of accessibility.

Chapter 4

CONCLUSION

This thesis explores how individual-level travel behavior can be used to understand food accessibility. Two studies are conducted, each of which explores different concepts. Chapter 2 creates different population segments of ”eating-out personas” in the Puget Sound, Washington through clustering individuals by their restaurant-related travel behaviors, demographics, and built environment attributes. This study provides insight into the travel context behind restaurant trips. Characteristics such as the number of children in a household, vehicle access, and age affect restaurant-related travel, though living in a food desert does not constrain behavior. Personas are distinguished by travel time to restaurants, frequency of eating out and meal times. This study provides recommendations for healthy-eating interventions to target specific characteristics of each persona.

Chapter 3 develops trajectory-based accessibility measures using human mobility data. These measures are applied to measure grocery stores accessibility and affordability of transit users in Stockholm County, Sweden through smartcard data. Grocery store accessibility is measured at localized, 15-minute hubs along individual trajectories. The findings show that those living in neighborhoods located within the city center have high home-based, static accessibility, though stores are less affordable. When considering mobility, these individuals observe marginal improvements to accessibility. In contrast, individuals living in rural or suburban areas with high transit connectivity observe large positive impacts to accessibility. This method can be used to evaluate the effect of different policies on food accessibility (e.g., new transit routes, relocating a grocery store).

These studies approach food access and travel behavior from two different perspectives. Several comparisons can be made. First, the studies answer two conceptually distinct questions. The first study looks at the context in which restaurant-related trips occur, so that underlying motivations can be hypothesized. The findings of this study can inform behavioral interventions which encourage healthy eating. The second study creates a novel, trajectory-based accessibility measure which measures the opportunity to access amenities while being grounded in real travel patterns. The methods developed in the second study can be applied to quantify accessibility at a regional level. Regions in need of accessibility improvements can be identified and the effects of varying policies on accessibility can be tested.

Additionally, the studies use different types of data. The first study uses household travel surveys (”small data”), which is rich in information such as demographics and attitudinal question though limited in sample size. The individual-level, socioeconomic variables are key to understanding the context of a person’s trip in this first study. The second study uses smartcard data (”big data”), which has a much larger sample size and detailed spatiotemporal information but limited knowledge on the individual-level demographics. The ability to mine key travel characteristics based on observed mobility patterns is important to forming the trajectory-based accessibility measure in the second study. However, fusing big and small datasets is a promising line of future work. A fused dataset would provide additional depth to each of the studies. For example, in the first study, a fused dataset may allow for longitudinal studies to observe how restaurant-related behaviors change over time as food policy changes or healthy-eating interventions occur. In the second study, a fused dataset could answer specific demographic questions about individual-level characteristics and food access (e.g., can people afford the food prices in areas which they live or travel through?).

Lastly, the two studies presented in this thesis are conducted in different parts of the world. Though both are located in the Global North, the Puget Sound region in Washington, U.S. and Stockholm County, Sweden have varying built environments, food policies and broader cultural norms. The Puget Sound is dominated by urban sprawl and car-centric design, leading to physically isolated neighborhoods with poor food accessibility. Stockholm County is characterized by a much more polycentric urban design, with grocery stores typically reachable by public transit or walking in urban and suburban areas. Distance to food is much less of an issue in Stockholm, though affordability is a growing concern. As a welfare state, Sweden provides more social safety nets and food policy centers on healthy food regulations. In the U.S., food access is largely shaped by racial injustices and economic disparities, and food policy is much more focused on increasing the equity of access. Each study should be considered with respect to the context in which they take place.

Food is a powerful force that not only fulfills biological needs, but is a societal cornerstone. Food can provide comfort, foster connections, and build communities. Of the five dimensions of food access, many require an understanding of people's perceptions of their food environments. Though this thesis studies food access through a strictly quantitative lens, it is also important to recognize the importance of qualitative tools. Mixed methods studies can capture more of the nuances and inherent personal factors which influence food accessibility such as cultural relevance of foods and perception of store quality. By integrating behavioral approaches, context-specific environments, policy, and personal preferences, future work can provide holistic perspectives on food accessibility. Ultimately, a comprehensive understanding to food access is key to unlocking effective food policy strategies.

BIBLIOGRAPHY

[1] United Nations. The 17 Goals, 2015.

[2] Spiker M.L. Dai J. Tsend A.S. Bszkiewicz J.H. Beese S. Collier S.M. Otten, J.J. and A. Ismach. American community survey data. Food Security and Food Assistance in the Wake of COVID-19: A 4th Survey of Washington State Households. Washington State Food Security Survey. , 2023. Accessed July. 11, 2023.

[3] Anita Borch and Unni Kjærnes. The Prevalence and Risk of Food Insecurity in the Nordic Region: Preliminary Results. *Journal of Consumer Policy*, 39(2):261–274, June 2016.

[4] Renee E Walker, Christopher R Keane, and Jessica G Burke. Disparities and access to healthy food in the united states: A review of food deserts literature. *Health & place*, 16(5):876–884, 2010.

[5] Kareem M. Usher. Valuing All Knowledges Through an Expanded Definition of Access. *Journal of Agriculture, Food Systems, and Community Development*, 5(4):109–114, August 2015. Number: 4.

[6] Bert van Wee, Jan Anne Annema, and David Banister, editors. *The Transport System and Transport Policy: An Introduction.* Edward Elgar, Cheltenham, UK, 2013.

[7] Michele Ver Ploeg, Paula Dutko, and Vince Breneman. Measuring food access and food deserts for policy purposes. *Applied Economic Perspectives and Policy*, 37(2):205–225, 2015.

[8] USDA. Food access research atlas documentation. https://www.ers.usda.gov/data-products/food-access-research-atlas/documentation/, 2022. Accessed Dec. 9, 2022.

[9] Renee Johnson. *Defining Low-Income, Low-Access Food Areas (Food Deserts).* Congressional Research Service, 2021.

[10] Michele Ver Ploeg, Paula Dutko, and Vince Breneman. Measuring Food Access and Food Deserts for Policy Purposes. *Applied Economic Perspectives and Policy*, 37(2):205–225, 2015.

[11] Tijs Neutens, Tim Schwanen, and Frank Witlox. The Prism of Everyday Life: Towards a New Research Agenda for Time Geography. *Transport Reviews*, 31(1):25–47, January 2011.

[12] Gökhan Yilmaz and Selami GÜLTEKİN. Consumers and tourists' restaurant selections. *Global issues and trends in tourism*, 217, 2016.

[13] Bee-Lia Chua, Shahrim Karim, Sanghyeop Lee, and Heesup Han. Customer restaurant choice: an empirical analysis of restaurant types and eating-out occasions. *International journal of environmental research and public health*, 17(17):6276, 2020.

[14] Jeanette Eckert and Igor Vojnovic. Fast food landscapes: Exploring restaurant choice and travel behavior for residents living in lower eastside detroit neighborhoods. *Applied Geography*, 89:41–51, 2017.

[15] Michele ver Ploeg, Vince Breneman, Tracey Farrigan, Karen Hamrick, David Hopkins, Phillip Kaufman, Biing-Hwan Lin, Mark Nord, Travis A. Smith, Ryan Williams, Kelly Kinnison, Carol Olander, Anita Singh, and Elizabeth Tuckermanty, editors. *Access to Affordable and Nutritious Food: Measuring and Understanding Food Deserts and Their Consequences: Report to Congress*. Administrative Publication Number 036. 2009.

[16] Bi Yu Chen, Yafei Wang, Donggen Wang, Qingquan Li, William H. K. Lam, and Shih-Lung Shaw. Understanding the Impacts of Human Mobility on Accessibility Using Massive Mobile Phone Tracking Data. *Annals of the American Association of Geographers*, 108(4):1115–1133, July 2018.

[17] Bernardo García Bulle Bueno, Abigail L. Horn, Brooke M. Bell, Mohsen Bahrami, Burçin Bozkaya, Alex Pentland, Kayla de la Haye, and Esteban Moro. Effect of mobile food environments on fast food visits. *Nature Communications*, 15(1):2291, March 2024. Publisher: Nature Publishing Group.

[18] CDC. Census tract level state maps of the modified retail food environment index (mrfei). https://www.cdc.gov/obesity/downloads/census-tract-level-state-maps-mrfei$_TAG508.pdf, 2011. AccessedJan.12, 2023.$

[19] BLS. Consumer expenditure surveys. https://www.bls.gov/cex/tables.htm, 2021. Accessed Feb. 2, 2023.

[20] USDA. Food prices and spending. https://www.ers.usda.gov/data-products/ag-and-food-statistics-charting-the-essentials/food-prices-and-spending, 2023. Accessed July 14, 2023.

[21] Mary J Scourboutakos, Zhila Semnani-Azad, and Mary R L'Abbe. Restaurant meals: Almost a full day's worth of calories, fats, and sodium. *JAMA internal medicine*, 173(14):1373–1374, 2013.

[22] Xiang Chen and Mei-Po Kwan. Contextual uncertainties, human mobility, and perceived food environment: The uncertain geographic context problem in food access research. *American journal of public health*, 105(9):1734–1737, 2015.

[23] Michael J Widener, Steven Farber, Tijs Neutens, and Mark Horner. Spatiotemporal accessibility to supermarkets using public transit: an interaction potential approach in cincinnati, ohio. *Journal of Transport Geography*, 42:72–83, 2015.

[24] Kelsey C Abel and Kasey M Faust. Modeling food desert disruptors: Impact of public transit systems on food access. In *Construction Research Congress 2018*, pages 362–372, 2018.

[25] Heidi M Blanck, Amy L Yaroch, Audie A Atienza, Sarah L Yi, Jian Zhang, and Louise C Mâsse. Factors influencing lunchtime food choices among working americans. *Health Education & Behavior*, 36(2):289–301, 2009.

[26] Eun-Jung Kim and Loren V Geistfeld. Consumers' restaurant choice behavior and the impact of socio-economic and demographic factors. *Journal of Foodservice Business Research*, 6(1):3–24, 2003.

[27] Paula Dutko, Michele Ver Ploeg, and Tracey Farrigan. Characteristics and influential factors of food deserts. Technical report, 2012.

[28] Jerry Shannon and W Jay Christian. What is the relationship between food shopping and daily mobility? a relational approach to analysis of food access. *GeoJournal*, 82:769–785, 2017.

[29] Richard Casey Sadler, Jason Andrew Gilliland, and Godwin Arku. Theoretical issues in the 'food desert'debate and ways forward. *GeoJournal*, 81:443–455, 2016.

[30] Hayden Stewart, Noel Blisard, and Dean Jolliffe. Let's eat out: Americans weigh taste, convenience, and nutrition. 2006.

[31] Peter James, Mariana C Arcaya, Devin M Parker, Reginald D Tucker-Seeley, and SV Subramanian. Do minority and poor neighborhoods have higher access to fast-food restaurants in the united states? *Health & place*, 29:10–17, 2014.

[32] Ali Arian, Melrose Meiyu Pan, and Yi-Chang Chiu. Personas: a market segmentation approach for transportation behavior change. *Transportation research record*, 2675(11):172–185, 2021.

[33] Sinziana I Rasca, Karin Markvica, and Benjamin Biesinger. Persona design methodology for work-commute travel behaviour using latent class cluster analysis. *Multimodal Transportation*, 2(4):100095, 2023.

[34] Meiyu Pan and Alyssa Ryan. Segmenting the target audience for transportation demand management programs: An investigation between mode shift and individual characteristics. *International Journal of Sustainable Transportation*, 18(1):62–83, 2024.

[35] [dataset] PSRC. Household travel survey. https://household-travel-survey-psregcncl.hub.arcgis.com/, 2021. Accessed Feb. 12, 2023.

[36] Yelp. Getting started with the yelp fusion api. https://docs.developer.yelp.com/docs/fusion-intro., 2023. Accessed Mar. 23, 2023.

[37] [dataset] ACS. American community survey data. https://www.census.gov/programs-surveys/acs/data.html, 2019. Accessed May. 5, 2023.

[38] Daniel Oberski. Mixture models: Latent profile and latent class analysis. *Modern statistical methods for HCI*, pages 275–287, 2016.

[39] Daniel Spurk, Andreas Hirschi, Mo Wang, Domingo Valero, and Simone Kauffeld. Latent profile analysis: A review and “how to” guide of its application within vocational behavior research. *Journal of vocational behavior*, 120:103445, 2020.

[40] Sung Hoo Kim and Patricia L Mokhtarian. Finite mixture (or latent class) modeling in transportation: Trends, usage, potential, and future directions. *Transportation Research Part B: Methodological*, 172:134–173, 2023.

[41] Rezwana Rafiq and Michael G McNally. Heterogeneity in activity-travel patterns of public transit users: an application of latent class analysis. *Transportation Research Part A: Policy and Practice*, 152:1–18, 2021.

[42] Eric Molin, Patricia Mokhtarian, and Maarten Kroesen. Multimodal travel groups and attitudes: A latent class cluster analysis of dutch travelers. *Transportation Research Part A: Policy and Practice*, 83:14–29, 2016.

[43] Gamal Eldeeb and Moataz Mohamed. Quantifying preference heterogeneity in transit service desired quality using a latent class choice model. *Transportation Research Part A: Policy and Practice*, 139:119–133, 2020.

[44] Bridget E Weller, Natasha K Bowen, and Sarah J Faubert. Latent class analysis: a guide to best practice. *Journal of Black Psychology*, 46(4):287–311, 2020.

[45] Karen Nylund-Gibson and Andrew Young Choi. Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, 4(4):440, 2018.

[46] Jeroen K Vermunt and Jay Magidson. Technical guide for latent gold 5.0: Basic, advanced, and syntax. *Belmont, MA: Statistical Innovations Inc*, 2013.

[47] Linda K Muthén and Bengt O Muthén. Mplus user’s guide (version 8). *Los Angeles, CA: Muthen*, 2017.

[48] Zsuzsa Bakk and Jouni Kuha. Two-step estimation of models between latent classes and external variables. *Psychometrika*, 83:871–892, 2018.

[49] Christophe Croux, Geert Dhaene, and Dirk Hoorelbeke. Robust standard errors for robust estimators. *CES-Discussion paper series (DPS) 03.16*, pages 1–20, 2004.

[50] Sarah L Ferguson, E Whitney G. Moore, and Darrell M Hull. Finding latent groups in observed data: A primer on latent profile analysis in mplus for applied researchers. *International Journal of Behavioral Development*, 44(5):458–468, 2020.

[51] Jenn-Yun Tein, Stefany Coxe, and Heining Cham. Statistical power to detect the correct number of classes in latent profile analysis. *Structural equation modeling: a multidisciplinary journal*, 20(4):640–657, 2013.

[52] John R Hipp and Daniel J Bauer. Local solutions in the estimation of growth mixture models. *Psychological methods*, 11(1):36, 2006.

[53] Gene Balk. Seattle is the third most gentrifying us city-but that might not be as bad as you think, study finds. *The Seattle Times*, 24, 2019.

[54] Nevin Cohen. Feeding or starving gentrification: The role of food policy. *CUNY Urban Food Policy Institute*, 5:1–10, 2018.

[55] Shanthy A Bowman and Bryan T Vinyard. Fast food consumption of us adults: impact on energy and nutrient intakes and overweight status. *Journal of the american college of nutrition*, 23(2):163–168, 2004.

[56] Andrea Garfinkel-Castro, Keuntae Kim, Shima Hamidi, and Reid Ewing. Obesity and the built environment at different urban scales: examining the literature. *Nutrition reviews*, 75(suppl_1):51–61, 2017.

[57] Mikko Kärmeniemi, Tiina Lankila, Tiina Ikäheimo, Heli Koivumaa-Honkanen, and Raija Korpelainen. The built environment as a determinant of physical activity: a systematic review of longitudinal studies and natural experiments. *Annals of behavioral medicine*, 52(3):239–251, 2018.

[58] Angela J Haddad, Aupal Mondal, and Chandra R Bhat. Eat-in or eat-out? a joint model to analyze the new landscape of dinner meal preferences. *Transportation Research Part C: Emerging Technologies*, 147:104016, 2023.

[59] Boya Yu, Jiaxu Zhou, Yi Zhang, and Yunong Cao. Identifying restaurant features via sentiment analysis on yelp reviews. *arXiv preprint arXiv:1709.08698*, 2017.

[60] Ebenezer Howard. *Garden Cities of To-morrow.* Kessinger Publishing, 1902.

[61] Clarence Perry. The Neighborhood Unit. In *The Regional Plan of New York and its Environs*. Routledge, 7 edition, 1929. Num Pages: 13.

[62] Congress for the New Urbanism. The Charter of the New Urbanism, April 2015.

[63] Peter Calthorpe. The Pedestrian Pocket. In *The Pedestrian Pocket Book: New Suburban Design*. Princeton Architectural Press, 1989.

[64] Robert Cervero. Transit Villages: From Idea to Implementation. *ACCESS Magazine*, 1(5):8–13, September 1994.

[65] Robert C. Kloosterman and Bart Lambregts. Clustering of Economic Activities in Polycentric Urban Regions: The Case of the Randstad. *Urban Studies*, 38(4):717–732, April 2001. Publisher: SAGE Publications Ltd.

[66] Carlos Moreno, Zaheer Allam, Didier Chabaud, Catherine Gall, and Florent Pratlong. Introducing the "15-Minute City": Sustainability, Resilience and Place Identity in Future Post-Pandemic Cities. *Smart Cities*, 4(1):93–111, March 2021.

[67] Matteo Bruno, Hygor Piaget Monteiro Melo, Bruno Campanelli, and Vittorio Loreto. A universal framework for inclusive 15-minute cities. In *Nature Cities*, volume 1, pages 633–641, September 2024. ISSN: 2731-9997 Issue: 10 Journal Abbreviation: Nat Cities.

[68] Kostas Mouratidis. Time to challenge the 15-minute city: Seven pitfalls for sustainability, equity, livability, and spatial analysis. *Cities*, 153:105274, October 2024.

[69] Marcin Stepniak and Piotr Rosik. The Role of Transport and Population Components in Change in Accessibility: the Influence of the Distance Decay Parameter. *Networks and Spatial Economics*, 18(2):291–312, June 2018.

[70] Michael J. Widener, Steven Farber, Tijs Neutens, and Mark W. Horner. Using urban commuting data to calculate a spatiotemporal accessibility measure for food environment studies. *Health & Place*, 21:1–9, May 2013.

[71] Abigail L. Horn, Brooke M. Bell, Bernardo García Bulle Bueno, Mohsen Bahrami, Burçin Bozkaya, Yan Cui, John P. Wilson, Alex Pentland, Esteban Moro, and Kayla de la Haye. Population mobility data provides meaningful indicators of fast food intake and diet-related diseases in diverse populations. *npj Digital Medicine*, 6(1):1–12, November 2023. Publisher: Nature Publishing Group.

[72] Keunhyun Park, Reid Ewing, Sadegh Sabouri, Dong-ah Choi, Shima Hamidi, and Guang Tian. Guidelines for a Polycentric Region to Reduce Vehicle Use and Increase Walking and Transit Use. *Journal of the American Planning Association*, 86(2):236–249, April 2020.

[73] Steven Farber, Melinda Z. Morang, and Michael J. Widener. Temporal variability in transit-based accessibility to supermarkets. *Applied Geography*, 53:149–159, September 2014.

[74] Jerry Shannon. Beyond the Supermarket Solution: Linking Food Deserts, Neighborhood Context, and Everyday Mobility. *Annals of the American Association of Geographers*, 106(1):186–202, January 2016.

[75] Michael J. Widener, Steven Farber, Tijs Neutens, and Mark Horner. Spatiotemporal accessibility to supermarkets using public transit: an interaction potential approach in Cincinnati, Ohio. *Journal of Transport Geography*, 42:72–83, January 2015.

[76] Oded Cats, Isak Rubensson, Matej Cebecauer, Yaroslav Kholodov, Alex Vermeulen, Erik Jenelius, and Yusak Susilo. FairAccess How fair is the fare? Estimating travel patterns and the impacts of fare schemes for different user groups in Stockholm based on smartcard data. November 2019.

[77] Daniel Block and Joanne Kouba. A comparison of the availability and affordability of a market basket in two communities in the Chicago area. *Public Health Nutrition*, 9(7):837–845, October 2006.

[78] Antonio Páez, Darren M. Scott, and Catherine Morency. Measuring accessibility: positive and normative implementations of various accessibility indicators. *Journal of Transport Geography*, 25:141–153, November 2012.

[79] Ahmed M. El-Geneidy and David M. Levinson. Access to Destinations: Development of Accessibility Measures. May 2006. Accepted: 2007-08-08T21:00:22Z.

[80] Junghwan Kim. *What about People in Accessibility?* PhD thesis, University of Illinois, Urbana-Champaign, 2018.

[81] Torsten Hägerstrand. What about people in Regional Science? *Papers of the Regional Science Association*, 24(1):6–21, December 1970.

[82] Chandra Bhat, Susan Handy, Kara Kockelman, Hani Mahmassani, Qinglin Chen, Lisa Weston, and University of Texas at Austin. Center for Transportation Research. Development of an Urban Accessibility Index: Literature Review. Technical Report Report No. TX-01/7-4938-1, May 2000.

[83] Mei-Po Kwan. Space-Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point-based Framework. *Geographical Analysis*, 30(3):191–216, 1998.

[84] Harvey J. Miller. Modelling accessibility using space-time prism concepts within geographical information systems. *International Journal of Geographical Information Systems*, 5(3):287–301, January 1991.

[85] Harvey Miller. Place-Based versus People-Based Geographic Information Science. *Geography Compass*, 1(3):503–535, 2007.

[86] Timur Abbiasov, Cate Heine, Sadegh Sabouri, Arianna Salazar-Miranda, Paolo Santi, Edward Glaeser, and Carlo Ratti. The 15-minute city quantified using human mobility data. *Nature Human Behaviour*, 8(3):445–455, March 2024. Publisher: Nature Publishing Group.

[87] Mei Po Kwan and Xiao Dong Hong. Network-based constraints-oriented choice set formation using GIS. *Geographical Systems*, 5(1-2):139–162, 1998.

[88] Ed Manley, Chen Zhong, and Michael Batty. Spatiotemporal variation in travel regularity through transit user profiling. *Transportation*, 45(3):703–732, May 2018.

[89] Sungho Lim, Haesung Ahn, Seungchul Shin, Dongmin Lee, and Yong Hoon Kim. Investigating night shift workers' commuting patterns using passive mobility data. *Transportation Research Part A: Policy and Practice*, 181:104002, March 2024.

[90] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.

[91] Derya Birant and Alp Kut. ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data & Knowledge Engineering*, 60(1):208–221, January 2007.

[92] Wei Huang, Shishuo Xu, Yingwei Yan, and Alexander Zipf. An exploration of the interaction between urban human activities and daily traffic conditions: A case study of Toronto, Canada. *Cities*, 84:8–22, January 2019.

[93] Fehmi Can Ozer, Hediye Tuydes-Yaman, and Gulcin Dalkic-Melek. Increasing the precision of public transit user activity location detection from smart card data analysis via spatial–temporal DBSCAN. *Data & Knowledge Engineering*, 153:102343, September 2024.

[94] Deloitte. Deloitte City Mobility Index 2020 Stockholm.

[95] Stephanie Rost. Falling through the net An exploratory qualitative study of food poverty and insecurity in Sweden, June 2021.

[96] Jan Amcoff. Food deserts in Sweden? Access to food retail in 1998 and 2008. *Geografiska Annaler: Series B, Human Geography*, 99(1):94–105, January 2017. Publisher: Routledge _eprint: https://doi.org/10.1080/04353684.2016.1277076.

[97] Axfood. The Swedish food retail market, 2023.

[98] Marie Mannes. Sweden sees 'cause for concern' on retailer margins as food prices surge. *Reuters*, June 2023.

[99] Robin-Ivan Capar. Will the rise of food prices in Sweden slow down this year?, July 2024.

[100] PRO. Prisundersökningen 2023, 2023.

[101] U.S. Census Bureau. King County, Washington - Census Bureau Profile, September 2023.

[102] City Population. Stockholm (County, Sweden) - Population Statistics.

[103] Official Statistics of Sweden. Population per month by region, age and sex. Year 2000M01 - 2024M06, September 2024.

[104] King County Open Data. Food Establishment Inspection Data, September 2024.

[105] Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, October 2001.

[106] Davoud Moulavi, Pablo A. Jaskowiak, Ricardo J. G. B. Campello, Arthur Zimek, and Jörg Sander. Density-Based Clustering Validation. In *Proceedings of the 2014 SIAM International Conference on Data Mining*, pages 839–847. Society for Industrial and Applied Mathematics, April 2014.

[107] Stockholms Handelskammare. Making Stockholm Closer, March 2023.

[108] Region Stockholm. Regional Development Plan for the Stockholm Region 2050, November 2019.

[109] Andreas Bengtson and Lasse Nielsen. Relational Justice: Egalitarian and Sufficientarian. *Journal of Applied Philosophy*, 40(5):900–918, 2023.