Spring 5-13-2023

Equity in Transportation: Data Driven Analysis of Transportation Services and Infrastructures

Javad Jomehpour Chahar Aman
jjomehpour@smu.edu

EQUITY IN TRANSPORTATION: DATA DRIVEN ANALYSIS OF TRANSPORTATION SERVICES AND INFRASTRUCTURES

Approved by:

Prof. Janille Smith-Colin

Prof. Khaled Abdelghany

Prof. Barbara Minsker

Prof. Eric Larson

Prof. Wenwen Zhang

EQUITY IN TRANSPORTATION: DATA DRIVEN ANALYSIS OF TRANSPORTATION SERVICES AND INFRASTRUCTURES

A Dissertation Presented to the Graduate Faculty of the

Lyle School of Engineering

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

by

Javad Jomehpour Chahar Aman

May 13, 2023

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Dr. Janille Smith-Colin, for her unwavering support and guidance throughout my PhD journey. Without her exceptional mentorship and continuous encouragement, these years of research would not have been possible. I consider myself fortunate to have been mentored by someone as knowledgeable and supportive as Dr. Smith-Colin, and I am proud to be her first graduated PhD student. Dr. Smith-Colin's guidance has been instrumental in helping me navigate through challenges and explore new ideas. I am immensely grateful for the opportunity to learn from her.

I am grateful for the guidance provided by my advisory committee members, Professor Khaled Abdelghany, Professor Barbara Minsker, Professor Eric Larson, and Professor Wenwen Zhang.

I value and appreciate the support, friendship, and camaraderie of my research group members, Myriam Zakhem, Collin Yarbrough, Mohammad Maleki, and Owen Li, throughout our journey together.

I am immensely grateful for the love, support, and guidance of my parents, Fatemeh and Rasool. Their sacrifices, encouragement, and belief in me have shaped me into the person I am today, and I will forever cherish their presence in my life.

I am grateful to have such supportive individuals in my life, including Dornaz Niknehzad, Mohammad Reza Hasanzade Gorakhki, Amir Arsalan Mehrara Molan, Ehsan Yahyazade Rineh, Masoud Zafar Hoseini, Mahdi Heidarizad, Soheil Sohrabi, Arefeh Safaie Moghadam, Hedieh Ashrafi, Hamed Nikfarjam, and Nadereh Mansouri.

Jomehpour Chahar Aman, Javad

Equity In Transportation:
Data Driven Analysis of Transportation Services and Infrastructures

Advisor: Dr. Janille Smith-Colin

Doctor of Philosophy conferred May 13, 2023

Dissertation completed May 8, 2023

Achieving equity in transportation is an ongoing challenge, as transportation options still vary tremendously when it comes to marginalized populations. This dissertation addresses this challenge by conducting a comprehensive review of existing transportation equity literature and identifying two critical gaps: the lack of data-driven approaches to studying spatial mismatch between transportation supply and demand, and limited information on women's perceptions and expectations towards emerging transportation services.

Chapter two introduces the concept of transportation "deserts," specifically transit deserts and walking deserts, and develops data-driven frameworks to identify and investigate neighborhoods with limited transportation service supply but high demand. The frameworks compare mobility demand and supply for active transportation modes and utilize statistical modeling techniques to reveal the inequitable distribution of transportation services. The identification of transportation deserts provides valuable insights for investment and redevelopment, highlighting areas of underinvestment.

Chapter three focuses on gender equity and the lack of understanding about transportation user preferences, particularly for women. Through a gender-sensitive analysis of online reviews using text-mining techniques, the chapter presents an empirical analysis of rider satisfaction with scooter

services. The study utilizes online data from app store reviews and employs machine learning techniques to uncover factors that influence overall satisfaction across genders. The findings enhance our understanding of gendered differences in micromobility rider sentiment and satisfaction.

In conclusion, this dissertation offers a comprehensive examination of transportation equity from multiple perspectives. It identifies critical gaps in existing literature and employs innovative analytical methodologies to address these gaps. The research findings have important policy implications for city planners, transportation managers, urban authorities, and decision-makers striving to create inclusive and vibrant urban spaces that benefit all members of society. By addressing these gaps, policymakers can promote equitable transportation services and ensure access to safe, reliable, and affordable transportation options for all individuals.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

1.1 Introduction

Over the past decade, researchers and decision-makers in the transportation field have become increasingly interested in issues of equity in transportation [1]. Theoretically, equity is defined by the fairness of the distribution of impacts among populations [2]. Accordingly, equity in transportation refers to accessible and affordable transportation for every person in the community. Equity in transportation results in the equitable allocation of transportation infrastructure benefits, costs, impacts, and services across groups of different income, gender, needs, ability, and other factors affecting transportation choice and impact. An equitable transportation system supports an ideal condition in which no individual or group is underprivileged due to a lack of access to opportunities needed for a meaningful and respectable life [3].

However, transportation options vary tremendously, particularly when it comes to people of color, low-income individuals, residents with disabilities, elderly people, women, and youth [4]. Current evidence on differences in personal vehicle ownership rates, trip planning, mode choice behavior, activity engagements, and experiencing environmental burdens across various population groups has vividly revealed the disparities in access to resources and benefits.

1.2 Definitions of Equity

The word “equity” is often used interchangeably with other terms such as equality [5], fairness, justice [6], environmental justice (EJ), and social exclusion. Although the various terms generally convey the same concept, the slight nuances between them could be misleading. For instance, while equity refers to a “subjective” distribution of resources based on “moral judgment” informed by recipient need, equality expresses a uniform allocation of resources among people without considering their needs and condition [5]. This omission could prove costly and impractical.

Therefore, a distribution could be equitable and still unequal [7], and conversely, distributions can be equal yet inequitable.

Uniform need or horizontal equity has been another way to express equality. Horizontal equity was concerned with the uniform distribution of resources among people [8] regardless of their needs, expectations, and quality of life. Horizontal equity, also refers to as egalitarianism [9], is a philosophical viewpoint that emphasizes the equal right of the population to be treated the same [10]. Vertical equity, also called social equity [11], social justice, environmental justice, and social inclusion [7], [9], distributes resources and benefits to groups based on need or ability or membership in a social class group categorized as disadvantaged [2]. As an example, the tax system illustrates vertical equity where low-income populations pay lower taxes [8]

1.3 Challenges in Transportation Equity Analysis

Achieving equity in transportation has been an ongoing challenge. Implementing policy is merely the first stage in an endless cycle of equitable transportation planning that alternate between establishing a goal, placing policy and programs into practice, addressing detected issues, evaluating results, and iterating the program. Implementation of a policy intended to promote equity has not included an assessment of how successfully (or how poorly) the program would accomplish its goals. Understanding relative successes is helpful for identifying best practices moving forward and valuable for determining the extent to which the program iteration addresses detected issues [12].

Setting explicit goals – such as equity goals informed by community mobility demand – paired with innovative data-gathering approaches and systematic data-driven evaluation frameworks could enable equity analysis. This ongoing process helps urban governments ensure that programs and policies would remove systemic barriers to opportunities and benefits for underserved population groups. By removing barriers, urban governments cab develop policies and programs that equitably deliver resources and benefits [13].

Various studies have tried to consider the role of equity in transportation planning and decision-making. However, there has been little consensus about the most effective approach to measure equity [14]. Within the context of transportation, challenges in the measurement of equity have been attributed to the lack of a standard definition of transportation equity [15]. In addition to impacts and outcomes that differ based on the selected definition of equity [9], often a vague relationship existed among measurement methods, defined indicators, and final goals [6].

In recent years, significant advancements have been achieved in literature that addressed fundamental scientific and engineering challenges related to transportation equity. However, inequities are still deeply entrenched in many transportation systems [2]. Planning for transportation has historically led to long-lasting inequities for marginalized groups like Black and Hispanic populations, low-income people, and women [16]. Data concerning the needs of marginalized groups have been lacking for a more comprehensive data-driven equity analysis. The lack of this data has been a major barrier to promoting transportation equity [17]. Although previous research has attempted to integrate data analytics into transportation planning and decision-making to support equity analysis [12], [18], [19], numerous gaps still exist in equity analysis performed at the community or neighborhood scale [4].

This study focuses on the concept of vertical equity. As mentioned previously, in vertical equity resources are provided based on need and ability. Special effort is made to offer resources to those who are disadvantaged and often excluded. As such, in this study, our data analysis was targeted toward socially excluded persons and marginalized populations. The data was used to identify and comprehend the requirements, expectations, mobility patterns, and gaps in the transportation services that marginalized populations receive.

1.4 Research Gaps in Literature

In the review of transportation equity literature, three gaps were identified: 1) equity in access and the lack of systematic data-driven approaches to study spatial mismatch between transportation supply and demand, and 2) gender equity and the limited information on women's attitudes and expectations toward emerging transportation services. The primary goal of this dissertation was to explore these gaps using quantitative methodologies. This research advanced knowledge in the field by investigating equity in transportation from two different perspectives on equity: equity in access, gender equity, and mobility equity. Each gap perspective is introduced as follows and explicitly described in the sections that follow.

1.4.1 Equity Access: Spatial mismatch between demand and supply

Equity has been concerned with the fair distribution of impacts (benefits and burdens) based on the needs of the recipients. In the case of transportation, inadequate or inequitable access to transportation services may have been the cause of health, education, and financial challenges for disadvantaged individuals [20]. However, traditionally, transportation studies have failed to take equity into account, instead using economic impact assessment as the primary yardstick for measures of fairness in policymaking [14]. Several forms of equity, various types of impacts, differing approaches to equity measurement, and many social groupings to consider in the analysis have made equity evaluation a challenging endeavor. The literature is lacking in comprehensive, data-driven frameworks to support equity analysis and practical ways to incorporate equity goals into transportation planning and decision-making [2], [21]. With a specific focus on walking and public transit modes, this research has identified and suggested alternatives to resolve a gap in the transportation literature. The literature has overlooked the importance of measuring disparities in access to opportunities and the resultant social exclusion experienced by populations at risk.

The focus of analysis within the gap of equity of access, (Chapter 2) explores the spatial mismatch between the demand for and supply of infrastructure that supports walking and public transit modes. This gap was the primary focus of analysis within this chapter.

The equity of access chapter introduced the concepts of transportation ’deserts,’ or more specifically transit deserts [22] and walking deserts. The data from the public transit and walking deserts were captured and analyzed in structured, data-driven frameworks. These frameworks compared mobility demand and supply for active transportation modes (i.e., public transit and walking). The identification of transportation deserts ultimately contributed to practice by revealing areas in need of investment and (re)development while also revealing areas of underinvestment.

The equity of access chapter aimed to achieve the following goals:

- assess whether different population groups equitably receive public transit and walking services.
- develop a data-driven structured framework to identify and investigate neighborhoods with limited transportation service supply and considerable demand.

1.4.2 Gender Equity: Examine the application of text mining in gender-sensitive analysis

Gender, in general, refers to the socially constructed roles, behaviors, and identities of male, female, and gender-diverse people [23]. Office et al. (2000) defined ‘gender equity’ as the fair treatment of women and men [and non-binary persons] according to their respective needs [24]. The first step in gender-sensitive planning for promoting gender equity was gender analysis [25]. Gender analysis was described as a method to identify distinctions between men and women in terms of their unique activities, circumstances, needs, access to development benefits, and decision-making. Therefore, gender equity included gathering sex-disaggregated statistics and gender-sensitive information about the populations in question.

However, gender differences are often overlooked in transportation research, design, study implementation, and reporting, resulting in adverse consequences. This is not unexpected as the transportation industry historically is a male-dominated sector, both from the employment point of view and for the embedded values [26]. It is necessary to have a systematic understanding of women's demands, concerns, and expectations in order to make transportation policy more responsive to such needs. This lack of gendered understanding is further deepened when it comes to emerging technologies and modes of transportation. For example, very little has been done to understand women's perceptions toward shared micromobility systems, Mobility as a Service (MaaS), or autonomous vehicles.

Moreover, gender equity studies in transportation to date most frequently used traditional datasets like travel survey data or questionnaires. The studies have not prominently utilized new information sources such as big data [27]. Urban big data stored in various forms (text, image, etc.) and in unprecedented quantities allowed researchers to address questions that were previously impossible to investigate [28]. Novel datasets (e.g., social media and app reviews) were able to reveal information about user experiences and expectations while offering opportunities to study less investigated groups, particularly women [29]. These studies brought to the forefront often unheard concerns about the 'default male' urban system [30], shed light on attitudes toward mobility services, and provided a systematic understanding of marginalized groups' needs. Consequently, these concerns made transportation policy more responsive to 'gendered' travel concerns.

To address existing gaps in the literature, research for this dissertation applied text-mining methods in a gender-sensitive analysis of micromobility reviews and comments. Chapter three presents a first-of-its-kind empirical analysis using online data to examine rider satisfaction with scooter services and factors impacting overall satisfaction and use. App store reviews from two major micromobility companies, including Lime and Bird, were investigated using machine learning techniques to identify the factors that influence rider satisfaction across gender.

The Gender Equity section aimed to achieve the following goals:

- understand the differences in perceptions and experiences toward shared electric scooter programs across gender.

  investigate how levels of rider sentiment and satisfaction towards e-scooters vary across gender.

1.5 Dissertation Organization

This dissertation has been organized as follows:

- Chapter one, Introduction, included an introduction to the primary gaps in literature that this dissertation discusses.
- Chapter two, Equity Access, included two sections that investigate equity in access through the concepts of walking deserts and transit deserts.
- Chapter three, Gender Equity, investigated gender equity by analyzing the differences between women and men riders' perceptions toward the shared electric scooter program.
- Chapter four provides the conclusion, summarizing the key findings, implications, and contributions of this research.

CHAPTER TWO: EQUITY IN ACCESS

2.1 Introduction

This study introduces and applies the concept of transportation ‘deserts’, a structured, data-driven framework, which captures and compares accessibility demand and supply. Consequently, the study of transportation deserts reveals areas in need of investment and (re)development (Figure 2-1). A transportation desert, or the gap between demand and supply, is often the result of the inequitable distribution of resources and services. This distribution gap results in service that fails to meet the accessibility needs of communities through reliable, functional, efficient, and effective service. The general principles of supply and demand are thus applied to study walking deserts and transit deserts as subsets of a broader transportation desert.

To capture the level of supply and demand, various indicators are introduced. The indicators are aggregated to reveal a holistic overview of transportation (walking and transit) supply and demand across neighborhoods. Demand indicators evaluate the concentration of disadvantaged populations, identified based on sociodemographic characteristics of residents (e.g., low-income or minority populations or active transport commuters including bikers and transit users). Supply indicators evaluate transportation service features of each mode; for instance, assessing transit deserts connectivity to the network, connectivity to destinations, service frequency, flexibility, and time efficiency.

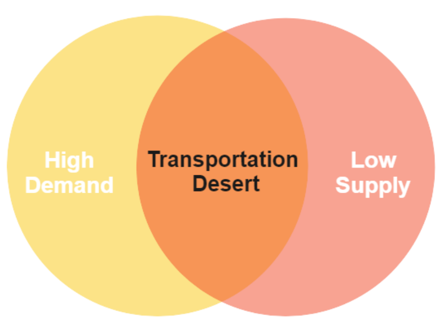


Figure 2-1 Where are transportation deserts?

The proposed framework sought to resolve a gap in the transportation literature that overlooked the importance of measuring disparities in access to opportunities. Also overlooked in literature was the mismatch between the demand and supply for public transportation services and walking infrastructure among the transportation disadvantaged, including persons often socially excluded. The suggested data-driven methodology could assist local and national transportation agencies in implementing a holistic, transparent, reliable, and straightforward equity analysis of access to services and opportunities. This study proposed a methodology acknowledging the need for the equitable distribution of accessibility based on identified population needs. The methodology applied the idea of 'deserts' as a practical approach for policymakers and urban planners to monitor the equity of infrastructure investments across cities. The outputs of the proposed methodology can increase decision-makers and urban planners' understanding of the transportation needs of disadvantaged individuals. The findings could support future policy actions and ongoing efforts to deliver equitable transportation systems to underserved populations. The following two sections (2.2 and 2.3) described the methodology for calculating demand and supply as they related to 1) walking deserts and 2) transit deserts. Both types of deserts were explored in the context of the City of Dallas and the City of Austin, TX.

2.2 Walking Deserts

2.2.1 Introduction

Walking deserts are defined as areas with high levels of walking demand (i.e., a concentration of walking-dependent populations) and low levels of walking supply. To capture levels of demand and supply, two specific indices were introduced in this study: Overall Demand Index (ODI) and Walking Supply Index (WSI). ODI evaluated the concentration of walking-dependent populations, who were identified based on sociodemographic characteristics of residents (e.g., low-income or minority populations and active transport commuters (bikers and transit users). WSI was evaluated by incorporating the network connectivity, sidewalk availability, and gravity-based walking accessibility of medical facilities, grocery stores, and schools. Using the walking desert concept, this study proposed a framework to assess and display the gap between walking-dependent population demand and walking access to essential destinations. This study aimed to:

- develop indices to assess walking supply and demand,
- examine statistical associations between supply and demand,
- develop a methodology to capture and display walking deserts,
- implement the method, and map walking deserts across the City of Austin, Texas.

2.2.2 Background

Walking supply could be defined as the level of network connectivity, sidewalk availability, and walking accessibility available within a neighborhood. Generally, walking accessibility meant to “ensure the possibility of access, approximation and use of any environment” [31] via walking mode. Conversely, walkability was an indicator of how walking-friendly the urban environment is [32]. In other words, in defining walking accessibility, the concept of reaching destinations and addressing basic human needs (the need for recreation, education, food, etc.) was a key focus, while the concept of walkability was merely concerned with walking quality.

Existing literature suggested a good walking environment with a well-designed infrastructure could reduce transport disparities for elderly and disabled populations [33]. Furthermore, as a mode available to all income groups and most abled-bodied residents, walking was an equitable transport mode. However, disinvestment in walking infrastructure and inconsistency across active transportation plans has created an inequitable walking supply across cities [34]. Equity in the analysis of walking accessibility was a topic receiving little attention in the literature; similarly, these concepts received inconsistent treatment in practice [35].

Various studies have investigated the characteristics of a walkable environment and proposed indicators to measure walkability [36], [37]. Yet, very few studies highlighted the associations among the level of walking accessibility and infrastructure availability and race, income, and social class or social equity factors in walking studies. These associations are historically under-researched [36], [38], [39]. Moreover, no study has proposed a framework to capture and display disparities between walking supply and walking demand across different sociodemographic populations. The overall goal of this work was to contribute to solving these issues by introducing walking deserts.

The City of Austin was used as a case demonstration of relevant methodologies in this work. Austin's current mode share is 70% drive-alone to work, meaning only 30% of commuters in the City of Austin bike, walk, carpool, use public transit, or work from home [40]. Through the adoption of the Austin Strategic Mobility Plan in 2019, Austin's city council set a goal of a 50/50 mode share by 2039. This goal expressed a desire to have a 50% use of modes other than a personal vehicle by 2039 [40]. Accordingly, there plans have been made to increase the share of walking across the city from 2% to 4% by 2039 [40].

2.2.3 Method

Figure 2-2 provides an overview of the methodology applied in this study, starting with data collection. Raw datasets were preprocessed, and required indicators for both demand and supply were created. In this study, walking supply indicators were used to support the overall assessment of walking accessibility including levels of: 1) access to hospitals, 2) access to grocery stores, 3) access to schools, and 3) sidewalk availability. Demand was assessed using nine socioeconomic indicators (race, age, income, etc.). Scores for the ODI and WSI indices were then calculated for all neighborhoods using the identified indicators. Linear regression models were generated to reveal associations between demand and supply indicators. The overall demand index (ODI) and walking supply index (WSI) were introduced as measures of the level of demand and supply. Finally, walking deserts were explored and mapped by subtracting the level of supply from demand.

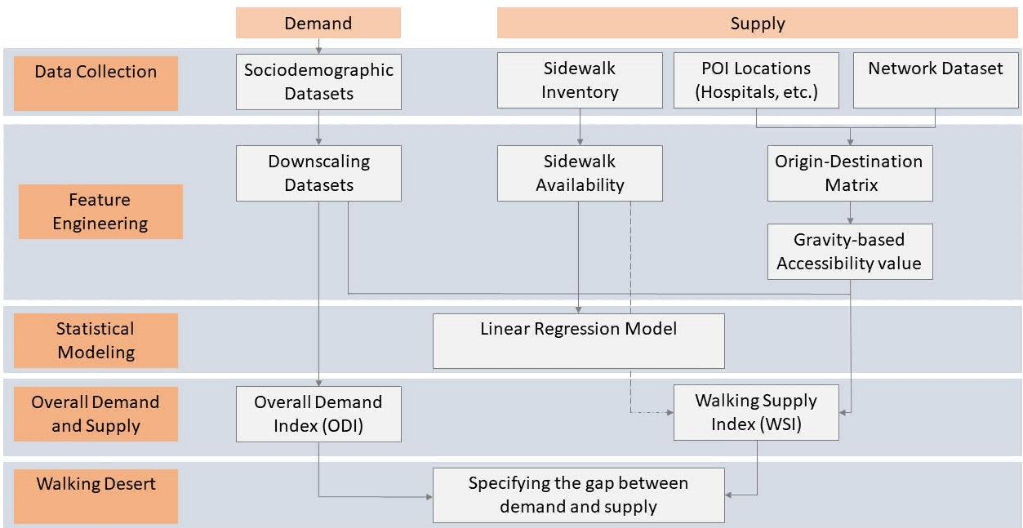


Figure 2-2 Walking desert methodology overview

2.2.3.1 Datasets

American Community Survey (ACS) datasets were used to obtain sociodemographic data [41] for the study area. Road network centerline and sidewalk availability shapefiles were exported from the City of Austin's open data platform. The shapefiles dataset indicated the areas with absent sidewalks throughout the city [42]. Points of interest (i.e., grocery stores, schools, and medical facilities) were imported from SafeGraph [43].

2.2.3.2 Walking Demand Measurement

In this study, demand was considered a function of the number of walking-dependent residents within the study area. The literature has classified walking-dependent populations into two distinct groups based on "income and social class" and "mobility needs and ability." In this study, nine different indicators were considered to identify walking-dependent populations and, thus, neighborhoods. The first group of indicators consisted of household income, education level, and vehicle ownership in addition to race/ethnicity-related characteristics (e.g., Black or Hispanic), and age (i.e., seniors older than 60 and children between 10 and 19 years old). The second group of indicators included physical impairment and the number of active transportation users (i.e., individuals who walk, bike, or use public transit to go to work) as mobility and ability measures.

Sociodemographic census block-group of datasets were spatially restructured into a grid of 4490 equally-sized hexagons (0.3 square kilometers). The grid covered the entire surface area enclosed by the City of Austin boundaries. Hexagon cells were used to reduce sampling bias from edge effects and were more appropriate for analyzing spatial phenomena where connectivity was important [44], [45]. The centroid of each hexagon cell was intersected with the census block groups to determine spatial collocation and the socioeconomic data allocated to their related hexagon. Census block groups were usually larger than the hexagons; therefore, the socioeconomic datasets had to be shared or split into multiple hexagons [46]. For the income level indicator, we allocated the exact income level to all hexagons of the related census block. Yet, for population-related indicators (e.g., count of the Black population), the count of the population was divided by the number of the hexagons within the census block. For instance, if the total count of the Black population in a specific census block was 100, the median income of the household was $50,000, and there were 20 distinct hexagons within the census block group, the Black population of each hexagon would be considered 5 with a median income of $50,000. This method, offered by Mayaud et al. [46] was straightforward and easy to apply. However, it did not consider the location of the residential areas and assumed populations were located uniformly across the regions [46], [47].

To calculate the ODI, all demand indicators were used to rank each hexagon cell in the city. Each cell was assigned a set of ranked scores, one for each of the indicators mentioned above (e.g., Black or Hispanic population). To score cells, for the population indicators (namely, Black, Hispanic, and low education) a percentage of population groups were calculated. For instance, for the percentage of Black population, the number of Black residents in each hexagon was divided by the total number of residents in the hexagon assessed. The hexagon cells were placed into 10 quantiles by indicator value (quantiles categorized hexagons into groups with equal numbers of records), each containing 10 percent of the total hexagon cells. The cells were then ranked from 1 to 10, depending on their quantile position. A ranked score of 1 was assigned to block groups with the lowest relative demand. A ranked score of 10 was assigned to cells with the highest demand value with middle scores in between. For example, if a cell had a relatively high percentage of Hispanic persons, it might receive a rank score of nine for that Hispanic population indicator. If a cell had a relatively low percentage of the Black population, it might receive a score of 2 for that indicator. The ODI score was then calculated as the sum of all demand score indicators, as shown in Equation 1-1:

$$ODI_i = \sum_{i,k} D_{ik} \qquad \text{Equation 2-1}$$

ODI_i= Overal Demand Index score for cell i
D_{ik}= Value of demand indicator k in cell i

2.2.3.3 Walking Supply Measurement

Walking Supply Index (WSI) was evaluated using four levels of indicators: 1) access to hospitals, 2) access to grocery stores, 3) access to schools, and 4) sidewalk availability To evaluate access to amenity, an O-D matrix between all origins and destinations across the city was developed using ArcGIS Pro. Hexagon centroids were considered origins, and points of interest (i.e., grocery stores, hospitals, and schools) were defined as destinations. Gravity-based methods were then used to measure the level of access to destinations. This method utilized a travel impedance function to reflect the "friction" of the distance separating origins and destinations [48], [49]. Gravity-based measures follow Newton's Law of Gravity and are represented as shown in Equation 2-2:

$$A_i = \sum_j O_j \, f(C_{ij}) \qquad \text{Equation 2-2}$$

A_i: Accessibility of hexagon i;
O_j: Opportunities located at j;
C_{ij}: Cost (distance) between $i \; and \; j$;
$f\left(C_{ij}\right)$: The impedance function implemented on cost between $i \; and \; j$.

The selected impedance function highly influences the calculated level of access to destinations [49]. Vale and Periera [48] measured and compared various impedance functions proposed by the literature (e.g., power, exponential, Gaussian, and cumulative). The authors finally promoted the cumulative–Gaussian method to measure walking access to destinations. This method explicitly considered travel tolerance and was a robust measure of data variability [48]. The Cumulative-Gaussian method was a combined method. This method merged the rectangular function used in cumulative opportunity measures with a Gaussian function. The merged function then considered the influence of distance beyond an acceptable distance and up to a maximum accessible distance (Figure 2-3). In this study, the maximum accessible distance is considered 2500 meters, as shown in Equation 2-3:

$$\begin{cases} f(C_{ij}) = 1 & C_{ij} < 450 \\ f(C_{ij}) = e^{\frac{(C_{ij}-1)^2}{v}}, & C_{ij} \geq 450 \end{cases}$$ Equation 2-3

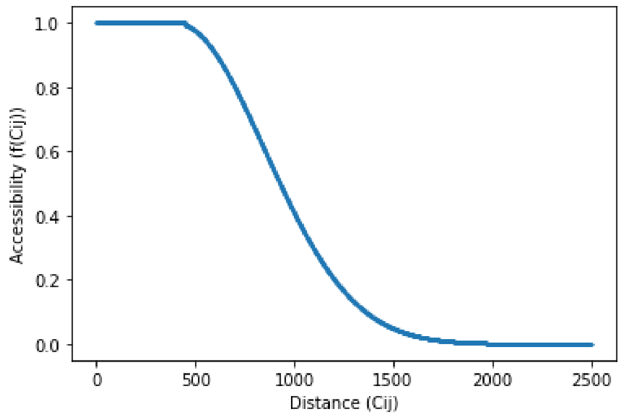


Figure 2-3 Combined Cumulative-Gaussian function

The O-D distance matrix was calculated for all origins and destinations, and the cumulative-Gaussian decay function was applied to distances. Then, the level of access to destinations was calculated. As a next step in calculating supply indicators, the percentage of sidewalk availability for each hexagon was calculated. Sidewalk availability was formulated as equal to one minus the ratio of absent sidewalk length. Finally, to calculate the WSI, all hexagons were ranked based on the four supply indicators (i.e., access to hospitals, schools, grocery stores, and sidewalk availability) and according to the process explained for ODI. The overall supply score (WSI) is calculated using Equation 2-4:

$$WSI_i = \sum_i Sidewalk_i + (Hospital_i + School_i + Grocery_i)/3$$ Equation 2-4

WSI_i: Walking Supply Index score

$Hospital_i, School_i, and\ Grocery_i$: Rank of the level of access to destinations for cell i

$Sidewalk_{ik}$: Rank of the sidewalk availability for cell i

Different aggregation rules are possible, such as linear aggregation (e.g., averaging or summing up) or geometric aggregation (multiplication), each of which implies different assumptions and consequences. For instance, linear aggregation rewards sub-indicators proportionally to the weights, while geometric aggregation rewards more those with higher scores [1]. However, the averaging method was chosen to calculate the walking supply for the following reasons.

In the multiplication method, each indicator is multiplied together to give an overall score. This means that the relative importance of each indicator is often not clear. It can thus be difficult to determine which indicators are driving the overall score. Additionally, resulting scores may vary widely depending on the values of each individual indicator, making it difficult to compare scores across different neighborhoods or locations [50].

Moreover, the EPA has defined a National Walkability Index score that incorporates transportation network characteristics, such as intersection density and proximity to transit stops, using a similar averaging method. The used of the averaging method in the walking deserts analysis thus provides a useful benchmark for walking deserts allowing for direct comparisons with other studies such as the EPA analysis that have used similar methods [51].

2.2.3.4 Regression Models

To understand the relationship between supply and the social characteristics of residents, two different models were generated. Various studies in transportation have generated geographically weighted regression (GWR) and ordinary least squares (OLS) models and compared the results. The studies often reported that GWR models yielded better fitting results than the OLS [52].

The regression model was developed first (Equation 2-5). In this model, we assume that residuals are independently distributed. This assumption can be tested using the Moran index.

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + \cdots + x_n\beta_n + \varepsilon \quad \text{Equation 2-4}$$

y: Dependent variable (walking supply indicator),

β_0: Intercept,

x_i: Social characteristics,

β_0: Regression coefficient of social characteristic i, and

ε: Error term.

The joint F-statistic, joint Wald statistic, and Jarque-Bera and Koenker (BP) statistic are commonly used to test the overall significance of a regression model. The joint F-statistic is the ratio of the mean squared error of the model to the mean squared error of the residuals. The statistic tested the null hypothesis that all regression coefficients were equal to zero, meaning that the independent variables had no effect on the dependent variable. A significant p-value for the joint F-statistic indicated that the null hypothesis could be rejected, and the model as a whole was a good fit for the data. The joint Wald statistic tested the null hypothesis that all regression coefficients were simultaneously equal to zero. A significant p-value for the joint Wald statistic indicated that at least one of the coefficients was significantly different from zero, and the model as a whole was a good fit for the data. The Koenker (BP) statistic is a test for heteroscedasticity; that is when the variance of the residuals is not constant across the range of the dependent variables. A significant p-value for the Koenker (BP) statistic indicated the presence of heteroscedasticity, and the model might need to be modified to address this issue. The Jarque-Bera (JB) statistic is a goodness-of-fit test, commonly used to test whether a set of data followed a normal distribution. The JB test is based on the skewness and kurtosis of the data and then compared the data to what would be expected under the assumption of a normal distribution. If the JB test returned a p-value below a certain significance level (e.g., 0.05), then the null hypothesis of normality is rejected, and it was concluded that the data were not normally distributed [53].

If the Moran index revealed a significant spatial autocorrelation in the residuals, a GWR model should be implemented to avoid violating the independence assumption. While generating a model for the GWR, there was a need to consider the autocorrelation of the observations and the nature of spatial data. GWR took into consideration the spatial aspect of the dataset. Unlike traditional statistical models, GWR provided a different coefficient for each social characteristic in different locations as defined by its coordinates (u_i, v_i). The dependent variable is estimated according to Equation 2-6:

$$y = \beta_0 + x_1\beta_1(u_1, v_1) + x_2\beta_2(u_2, v_2) + \cdots + x_n\beta_n(u_n, v_n) + \varepsilon \qquad \text{Equation 2-6}$$

Cardozo et al. (2012) highlighted the advantages of the GWR models as follows: 1) GWR models were more detailed and accurate, 2) they helped researchers investigate local spatial patterns, and 3) estimation errors were often lower than models generated using OLS [54].

2.2.3.5 Moran Index

Moran's Index is a measure of spatial autocorrelation that is commonly used in geography and spatial statistics. It measures the degree to which similar values of a variable are clustered together in space. The index ranges from -1 to 1, where a positive value indicates clustering, a negative value indicates dispersion and a value of 0 indicates randomness (Equations 2-7 and 2-8) [3]. The formula for calculating Moran's Index is:

$$I = \frac{n}{S_0}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{i=1}^{n} z_i^2} \qquad \text{Equation 2-7}$$

z_i: deviation of an attribute for feature i from its mean.

$w_{i,j}$: spatial weight between feature i and j

n: total number of features

S_0: aggregate of all the spatial weight

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} \quad \text{Equation 2-8}$$

The numerator of formula (2-7) represents the sum of the spatial cross-products of the variable, while the denominator represents the sum of the squared deviations from the mean. The spatial weight matrix specifies the relationship between the locations and determines which locations are considered neighbors for the purpose of calculating spatial autocorrelation. In this study, the Moran Index is evaluated using ArcGIS Pro software

2.2.3.6 Walking Deserts

After calculating ODI and WSI scores for all hexagons, walking deserts were identified. Walking deserts were defined as the areas with the highest ODI (demand) and the lowest WSI (supply) scores. The walking desert score is calculated using Equation 2-9:

$$WD_i = ODI_i - WSI_i \quad \text{Equation 2-9}$$

WD_i: Walking desert score of hexagons i

As ODI and WSI both ranged from 1 to 10, WD scores ranged between -9 and 9. This study considered scores between 6 and 9 to represent extreme walking deserts.

The threshold for extreme walking deserts is set between WD scores ranging from 6 to 9, representing areas that require immediate attention and intervention to improve walking, and prioritizing the top third tier of areas with the most severe gap between demand and supply. The choice of threshold impacts the study results by providing an objective measure to identify areas that require immediate attention for improving walking services. However, this threshold may not be appropriate for all studies depending on the research goals and context. It is possible that a threshold of 6 may not be generalizable to other contexts or regions and may need to be adjusted based on local conditions and stakeholder input. Therefore, researchers should carefully consider the specific context and goals of their study when choosing a threshold for identifying walking deserts.

2.2.4 Results

2.2.4.1 Supply

Walking deserts are identified as areas with high walking demand accompanied by considerably low supply. After calculating walking demand, the next step was to estimate walking supply. This study introduced the WSI score to measure walking supply across the city. WSI incorporated sidewalk availability, network connectivity, and walking accessibility to essential destinations, representing walking supply across different neighborhoods. To indicate sidewalk availability of the hexagon, the percentage of functional sidewalks throughout the cell was calculated. Moreover, gravity-based measurements of accessibility with respect to schools, medical facilities, and grocery stores were calculated using a cumulative-Gaussian model. An impendence factor was applied to distances that needed to be traveled. The level of network connectivity and proximity of the opportunities of each hexagonal cell were then designated.

Figure 2-4 displays, based on distances, the spatial distribution of sidewalk availability in addition to the level of walking access to grocery stores, hospitals, and schools, Centralized distribution of grocery stores and schools across the City of Austin resulted in concentrated regions of high accessibility in the downtown core, inner north, and inner south of the city. Results indicated that in the City of Austin there was higher school accessibility than grocery store accessibility, even though there were almost equal numbers of both facilities located across the city. This finding seemed to be the result of a more even distribution of schools across neighborhoods and the less even distribution of grocery stores. This distribution could be expected given the neighborhood school model employed broadly across the U.S. There were multiple clusters of accessible hospitals scattered across the city, mostly concentrated around the downtown. There were almost no accessible medical facilities in the eastern and southeastern parts of the city where mostly Black and Hispanic populations were located.

In comparison, the pattern of sidewalk availability differed from the spatial distribution of essential services. Unlike access to essential services, sidewalks were primarily available in the outer east areas of the city where racial minorities were located and the outer west of the city.

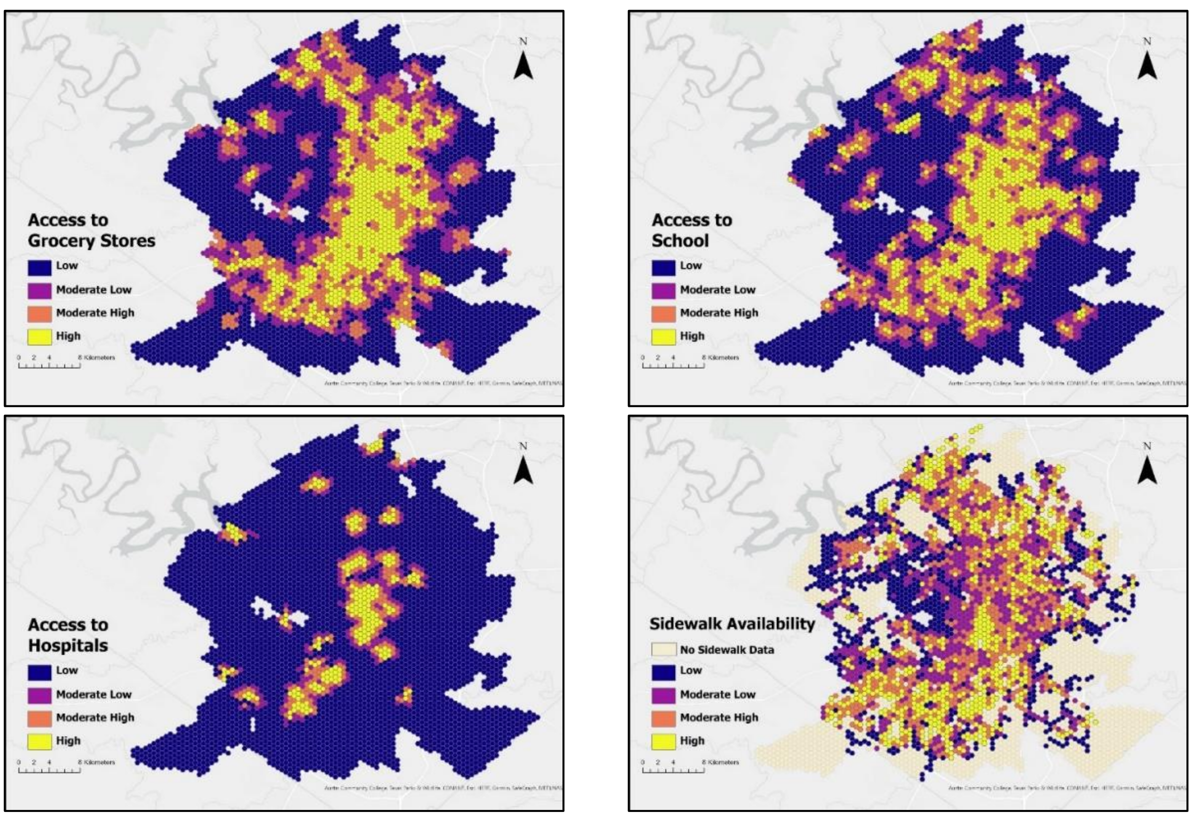


Figure 2-4 Spatial distribution of walking access to essential destinations in the City of Austin

Populations that have limited or no access to essential services and opportunities often experience social exclusion. Social exclusion happens when a group of the population is prevented from participating in activities or achieving opportunities [55]. As one of the cheapest and most accessible forms of transportation, walking presents a pathway for reducing barriers created by social exclusion. Yet, the graphs displayed in Figure 2-4 show that there are vast areas with zero access to amenities. In other words, there are individuals that are spatially excluded from opportunities. To understand the rate of exclusion for each different population group in the City of Austin, we calculated the percentage of individuals in each social group with zero access to amenities. The proportion of spatially-excluded individuals living in Hispanic, Black, and low-education areas presented an interesting nuance to the issue of walking accessibility and equity.

Figure 2-5 shows over 70 percent of Black, Hispanic, and senior neighborhoods in the City of Austin have zero access to hospitals. Additionally, over 20 percent of Black, Hispanic, and senior neighborhoods have zero access to grocery stores, and over 20 percent of the low-education population has zero access to schools. Hospitals have the highest levels of social exclusion across the board (i.e., across all social groups, hospitals are the least accessible).

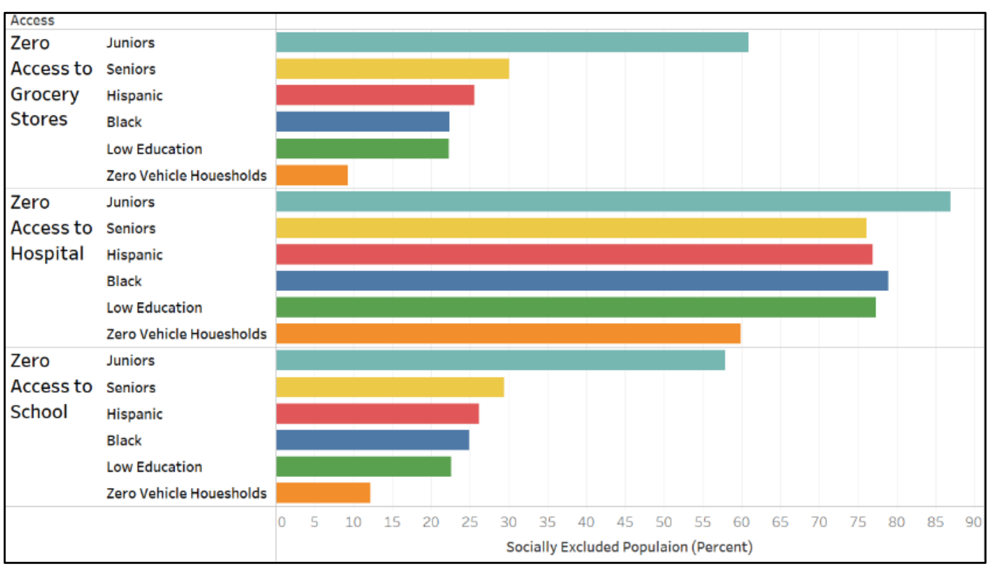


Figure 2-5 Socially excluded populations

2.2.4.2 Demand

The demand analysis measured the concentration of populations across the city, their associated socioeconomic characteristics, and their spatial distribution. This type of demand analysis was not exclusively done in support of a walking accessibility assessment but was crucial for establishing a baseline assessment of needs across populations to finally explore the walking deserts.

The spatial distribution of Black, Hispanic, walkers (individuals who walk from home to their work location), low education populations (adults with no more than a high school diploma), and zero vehicle households are mapped in Figure 2-6. All indicators are classified into quartiles. Dark blue represents the highest concentration, and light yellow represents the lowest concentration. The city showed a considerable bunching of the population across almost all social indicators. The central parts of the city were the densest areas in terms of the total population. Black populations mostly resided in the northeastern and eastern parts of the city. The western part of the city had the lowest concentration of Black people. Hispanic populations were concentrated in the southeastern portions of the city. Southeastern neighborhoods also showed the lowest levels of educational attainment across the city or the highest concentration of adult populations with low education (i.e., no more than a high school diploma). Walkers and zero-vehicle households mostly lived in the inner/central areas of the city where downtown and the University of Texas at Austin are also located.

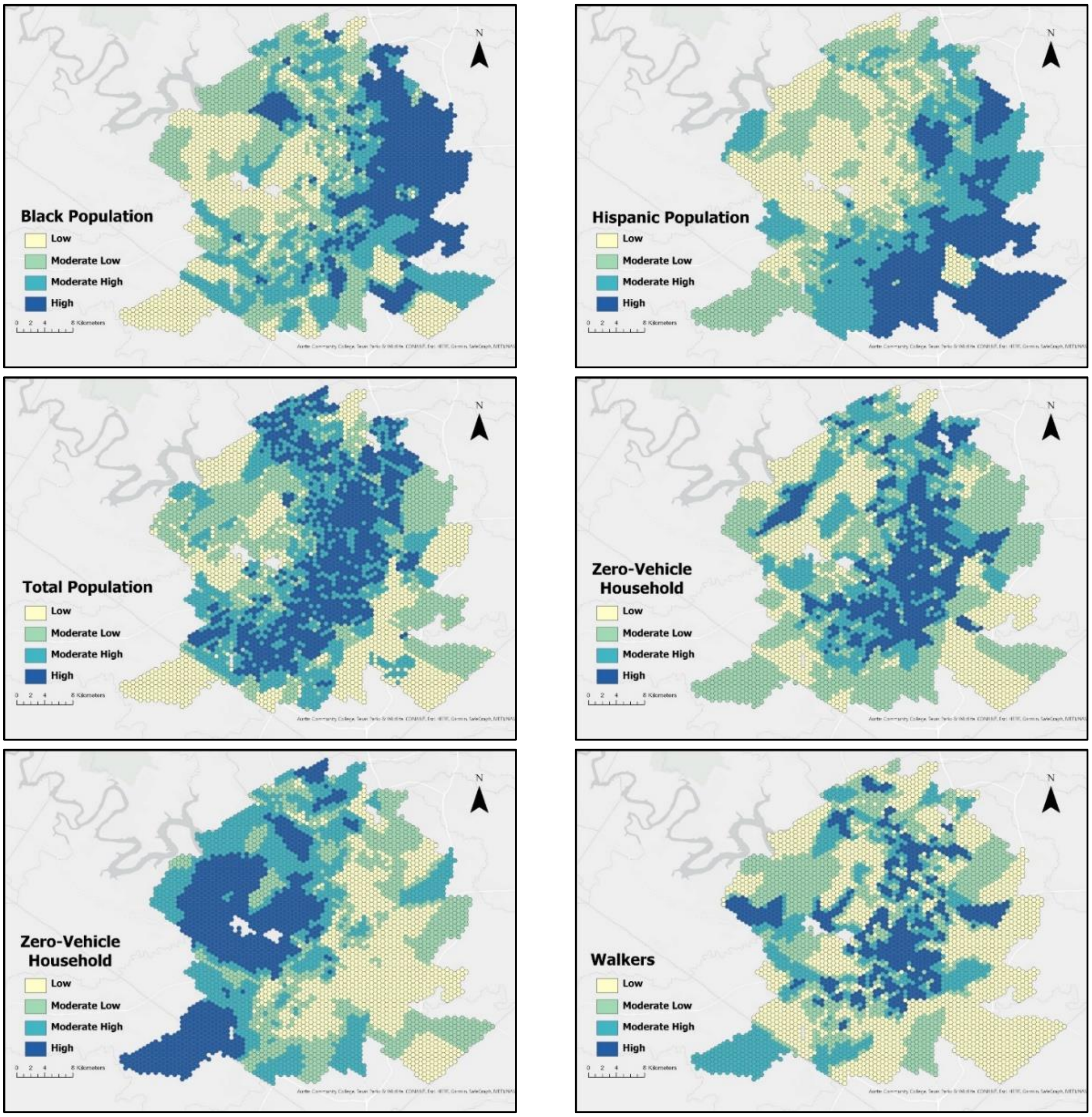


Figure 2-6 Spatial distribution of total population and vulnerable population groups

2.2.4.3 Ordinary Least Squares Model

Linear regression models were developed to understand the associations between the supply-side and demand-side indicators(Table 2-1). Prior to developing the models, bivariate correlations between potential candidates for the regression analysis were estimated to examine multicollinearity among the variables. All correlation coefficients were lower than the "danger level" of 0.70 [56]. Therefore, no threat of including these indicators in the modeling process was determined.

Four separate models were created for access to grocery stores, schools and hospitals, and sidewalk availability. Backward elimination, a feature selection method commonly used in regression analysis, was used to identify the most relevant variables that could be utilized to predict the outcome of interest while eliminating unnecessary variables that did not contribute much to the prediction accuracy of the model. The least significant predictors were removed iteratively from the model until a final model with only significant predictors was obtained**.**

The results of all four models showed relatively low R-squared scores (grocery store: 0.38, hospital: 0.10, school: 0.25, and sidewalk: 0.14). These scores indicate that the independent variables used in the models explain only a small proportion of the variance in the dependent variable. This also suggests that the global OLS models did not fully capture the complex relationships between the variables, and that other factors not included in the models may also be influencing the walking supply in the study area.

Results of this analysis showed a strong positive Moran Index, indicating that there are spatial patterns in the distribution of the variables, suggesting that the values of each variable in a given location are dependent on the values of neighboring locations. This can lead to biased parameter estimates and inaccurate results if not properly addressed in the modeling process.

Therefore, local OLS models may not perform well to depict the relationships between walking supply and demand, as they assume that the relationships are constant across space and do not account for spatial heterogeneity. To address this issue, a GWR model was used to capture the spatial variability in the relationships by estimating local regression coefficients at each location.

This allows for estimates of the relationships between walking supply and demand while accounting for the spatial autocorrelation in the data.

Table 2-1 Summary of Diagnosis of the Linear Regression Models Coefficient for Supply

	Grocery Store			Hospital			School			Sidewalk		
	Coeeficient	Probability	Robust Probability	Coeeficient	Probability	Robust Probability	Coeeficient	Probability	Robust Probability	Coeeficient	Probability	Robust Probability
Intercept	0.94	0.00*	0.00*	0.10	0.00*	0.000*	0.27	0.00*	0.00*	43.98	0.00*	0.00*
Black	-	-	-	-0.01	0.01*	0.00*	-	-	-	-	-	-
Asian	-	-	-	-	-	-	-	-	-	-0.075668	0.05*	0.04*
Hispanic	-	-	-	-	-	-	-0.001	0.00*	0.00*	-	-	-
Public Transit User	-	-	-	0.01	0.00*	0.00*	0.01	0.00*	0.00*	-	-	-
Biker	0.21	0.00*	0.00*	0.03	0.00*	0.00*	0.17	0.00*	0.00*	2.21	0.00*	0.00*
Walker	0.07	0.00*	0.00*	0.01	0.00*	0.00*	-	-	-	-	-	-
Unemployed	-0.02	0.00*	0.00*	-0.02	0.00*	0.00*	-0.01	0.00*	0.00*	-0.69	0.00*	0.00*
Over 65	-0.02	0.00*	0.00*	-0.03	0.00*	0.00*	-	-	-	-0.72	0.00*	0.00*
Under 18	-	-	-	-	-	-	-	-	-	-	-	-
Low Income Household	-	-	-	-0.03	0.00*	0.00*	0.02	0.00*	0.00*	-	-	-
Zero Vehicle Household	0.05	0.00*	0.00*	-	-	-	0.03	0.00*	0.00*	1.54	0.00*	0.00*
Number of Observations	4490			4490			4490			4490		
Multiple R-Squared[d]	0.381144			0.10			0.25			0.14		
Joint F-Statistic[e]				77.5			256.93			157.40		
Joint Wald Statistic[e]				113.3			628.63			922.31		
Koenker (BP) Statistic[f]				232.6			176.29			170.96		
Jarque-Bera Statistic[g]	27472.30			1293757.5			63695.51			653.78		
Akaike's Information Criterion (AICc)[d]	11371			-224.81129			8538			43319		
Adjusted R-Squared[d]	0.38			0.10			0.25			0.14		
Prob(>chi-squared)	0.00*			0.00*			0.00*			0.00*		
Moran's Index	0.624			0.591			0.594			0.472		
Expected Index	-0.000223			-0.000223			-0.000223			-0.000223		
Variance	0.000093			0.000091			0.000092			0.000093		
z-score	64.954514			61.959025			61.802322			49.011509		
p-value	0			0			0			0		

2.2.4.4 GWR Models

The GWR model selected explanatory variables using a stepwise procedure similar to that of a general linear regression. Variables were added to the GWR model one by one based on their significance and AICc values. However, a challenge arose when checking the significance of explanatory variables in GWR. That is, a variable might be significant in some neighborhoods but not in others. To address this issue, we selected variables that were significant in over 50 percent of areas. This process was repeated until the model with the lowest AICc value was obtained. This model was considered the best GWR model (Table 2-2). Kim and Nicholls (2017) utilized a similar framework to investigate the spatial variation of access to public beaches among various social compositions using GWR. Their work has been also used as a framework to report and discuss the results of GWR models.

The GWR model is used to account for spatial heterogeneity or variations in the relationship between dependent and independent variables across different geographic areas. Spatial clustering in demand variables, as indicated by the Moran Index, can lead to variations in the strength of the relationship between independent and dependent variables across different areas.

The GWR model is used to account for spatial heterogeneity or variations in the relationship between dependent and independent variables across different geographic areas. Spatial clustering in demand variables, as indicated by Moran's Index, can lead to variations in the strength of the relationship between independent and dependent variables across different areas.

In general, the discussion of results focused on areas with acceptable R-squared levels to explain the relationship between dependent and independent variables. As R-squared values resulting from the GWR models remained low in some areas, though statistically significant, further discussion of the relationship between dependent and independent variables focused on the areas where R-squared levels were between 0.3 and 0.65. In comparison to other studies applying GWR to understand access and availability, results presented in this dissertation are within an acceptable range for R^2 values of over 0.3. For example, a study conducted by Yu et al. (2014) examining the relationship between sociodemographic indicators and their potential accessibility to community pharmacies resulted in models with an average R-squared value of 0.35 using GWR [57]. Similarly, a study conducted by Khanyile and Fatti (2022) analyzed how socioeconomic characteristics are related to park access measures, using GWR model, resulting in an average R-squared of 0.53 and 0.61 [58]. Moreover, a study by Andersson (2017) examining the relationship between public transport accessibility and car use in Lund and Malmö, Sweden, found the R-squared values between 0.22 and 0.55 using GWR [59].

- **Access to Grocery Store**

The coefficients of the model indicated the direction and strength of the relationship between each predictor and the outcome variable. The R-squared value of 0.55 indicated that the model was a moderately good fit for the data. Table 2-1 indicates that there was no significant association between Black population and access to grocery stores in the OLS model. However, both positive and negative correlations are observed in Figure 2-7 and Table 2-2. The local coefficients for Black range from -0.04 to 0.13. These coefficients indicate that the relationship between access to grocery stores and proportion of Black population is not consistent across the study area.

Specifically, there were strong negative correlations in the northeastern neighborhoods (where the local R-square is also the highest (i.e., 0.39-0.64)) of the city, where there is a higher concentration of Black neighborhoods. These correlations suggested inequitable access to grocery stores for this population. Figure 2-7 and Table 2-2 demonstrate a mix of positive and negative correlations. The local coefficients for Walker varied from -0.21 to 0.92 (Figure 2-8). Positive correlations, implying equitable access to grocery stores for the residents who walk to work, were primarily observed in the central and downtown areas of the city (where the local R-square level is also acceptable). In contrast, negative correlations, suggesting unfair access, were predominantly found in the suburbs and were distant from downtown.

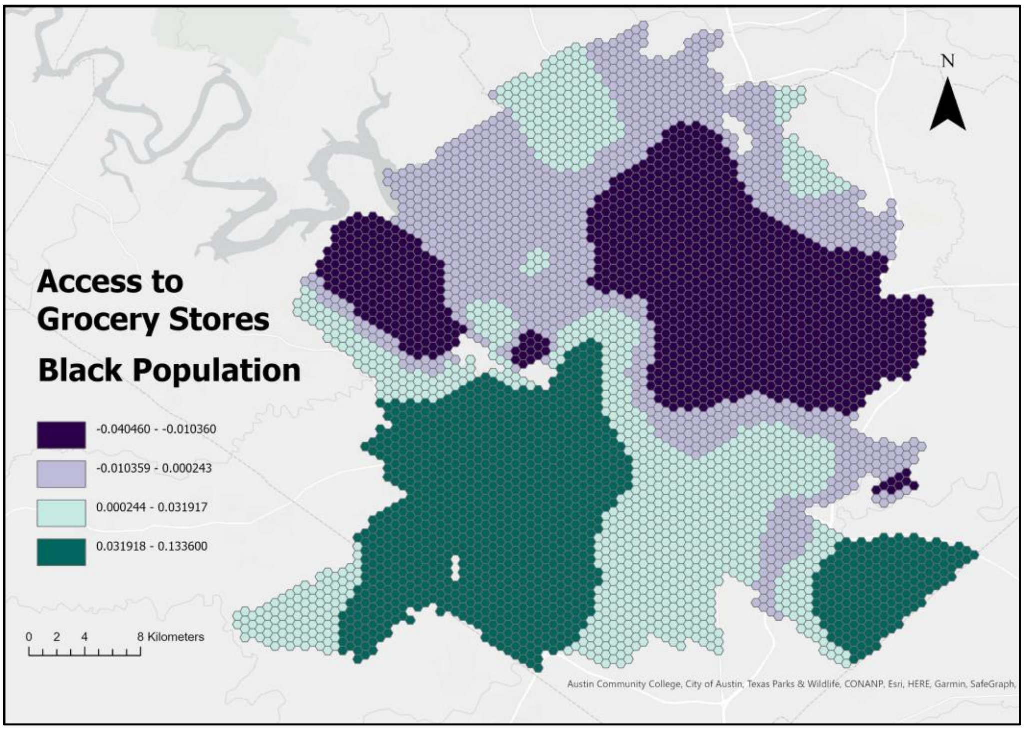


Figure 2-7 Spatial distribution of local parameter estimates for proportion of Black population

Figure 2-8 Spatial distribution of local estimates for the proportion of walker population

The R^2 value at the global level was 0.38, but the R^2 value varied locally across the study area, ranging from 0.01 to 0.64 (mean: 0.55) (Figure 2-9). It is important to note that conclusions cannot be reliably drawn in areas where the R-squared value is low (lower than 0.3). More than 50% of neighborhoods had higher local R^2 values than the global value. The local model performed well to explain the level supply in central, eastern, and southwestern neighborhoods. However, the local model's explanatory power was limited in some northwestern areas. This limitation suggested that the set of explanatory variables did not adequately explain the level of supply in these areas. These results indicated that the performance of the local model varied spatially across the study area.

Figure 2-9 Spatial distribution of local R^2

- **Access to Hospitals**

Table 2-2 shows the results of a GWR model that examines the relationship between various predictors and walking access to hospitals in a geographic area. The GWR model suggests that the relationship between walking access to hospitals and the indicators varied spatially across the study area, and the coefficients for each indicator differed in different parts of the area.

As shown in Figure 2-10 and Table 2-2, both positive and negative correlations were observed. The local coefficients for over 65 ranged from -0.018 to 0.015 with strong positive correlations. The local coefficients indicated fair access to hospitals in the northwestern neighborhoods where there was a high concentration of seniors. In contrast, strong negative correlations that indicated unequal access were observed in the northern and eastern areas of the city.

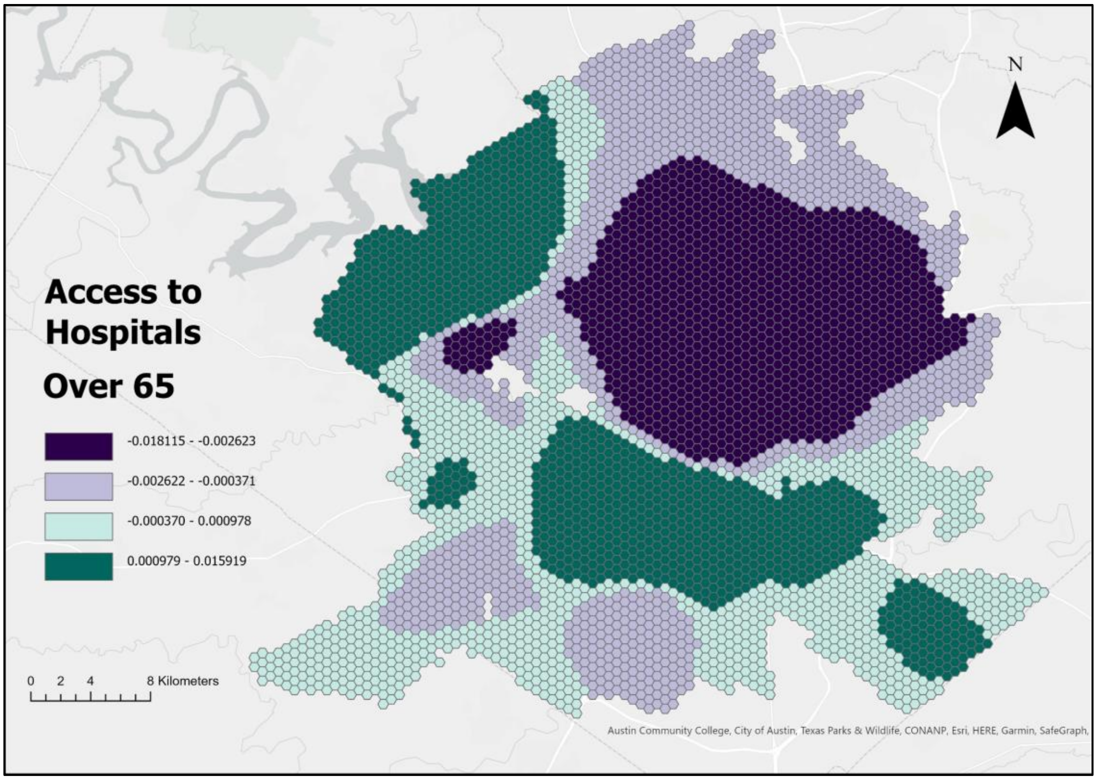


Figure 2-10 Spatial distribution of local parameter estimates for proportion of over 65

Figure 2-11 Spatial distribution of local R^2

The global value of R^2 was 0.10, but the local value of R^2 varied over the study area from 0.00 to 0.36 (Figure 2-11). Almost half of the neighborhoods had local R2 values greater than the global value. However, in general the model did not perform well in showing the variations of the target variable.

- **Access to Schools**

The GWR model for walking access to school showed that there were variations in the coefficients for different indicators across different spatial areas. The R-squared value for the model was 0.44. The GWR model also showed that there were variations in the coefficients across different spatial areas. For instance, the coefficient for Black ranged from -0.11 to 0.09. This coefficient indicated that the relationship between Black population and walking access to schools varied across different parts of the area. Similarly, the coefficient for Public Transit that ranged from -0.36 to 0.31 showed a wide range of correlations. Overall, the GWR model for walking access to schools indicated that there were variations in the relationships between different indicators and walking access to schools across different spatial areas.

In the OLS model, under 18 variable was not found statistically significant in interpreting the level of access to school (Table 2-1). However, Figure 2-12 and Table 2-2 show that both positive and negative correlations occur. The global value of R2 was 0.25, but the local value of R^2 varied over the study area from 0.00 to 0.49 (Figure 2-13).

Figure 2-12 Spatial distribution of local R^2

Figure 2-13 Spatial distribution of local parameter estimates for proportion of Over 65

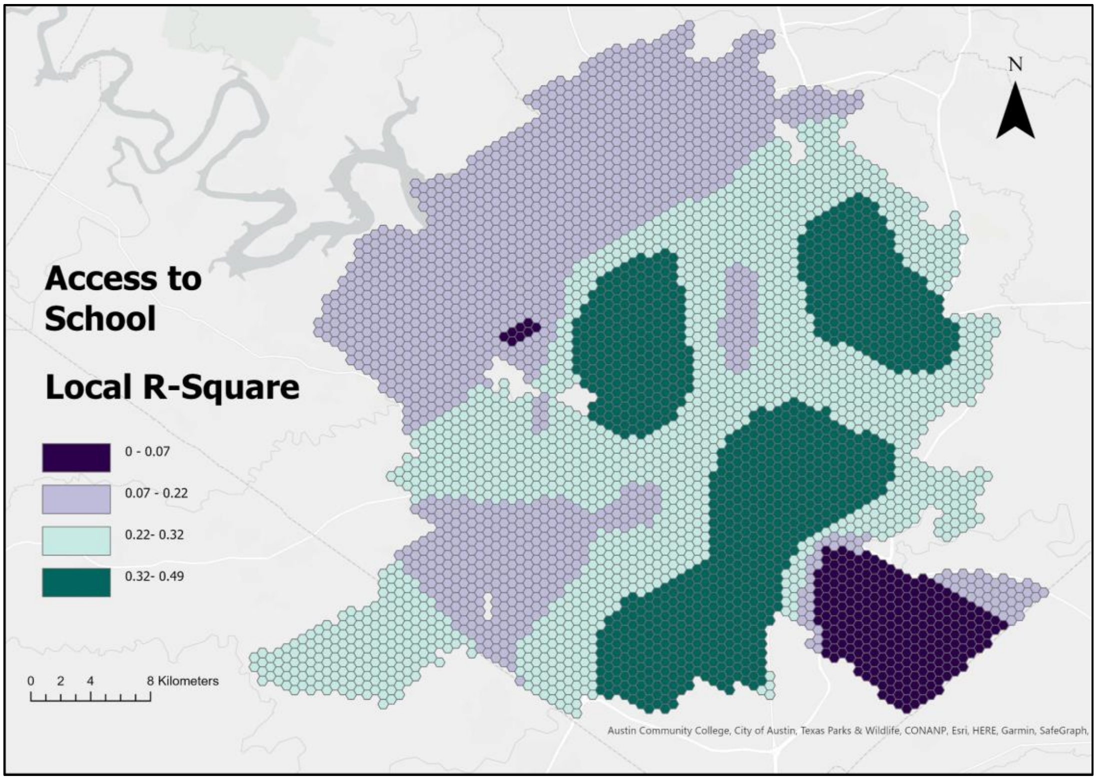


Figure 2-14 Spatial distribution of local R^2

- **Sidewalk Availability**

The findings suggested that although the overall OLS coefficient for Black and sidewalk distribution was positive and statistically significant, the relationship between the variables was not stationary and varied across different areas of the city. Figure 2-15 and Table 2-2 show that strong negative correlations, indicating equitable access to sidewalks, were observed in northern parts of the city. However, strong positive correlations, indicating equitable access, emerged in southwestern and southeastern parts of the city. The model performed the best in the central areas of the city in showing the relationship between dependent and independent values (2-16).

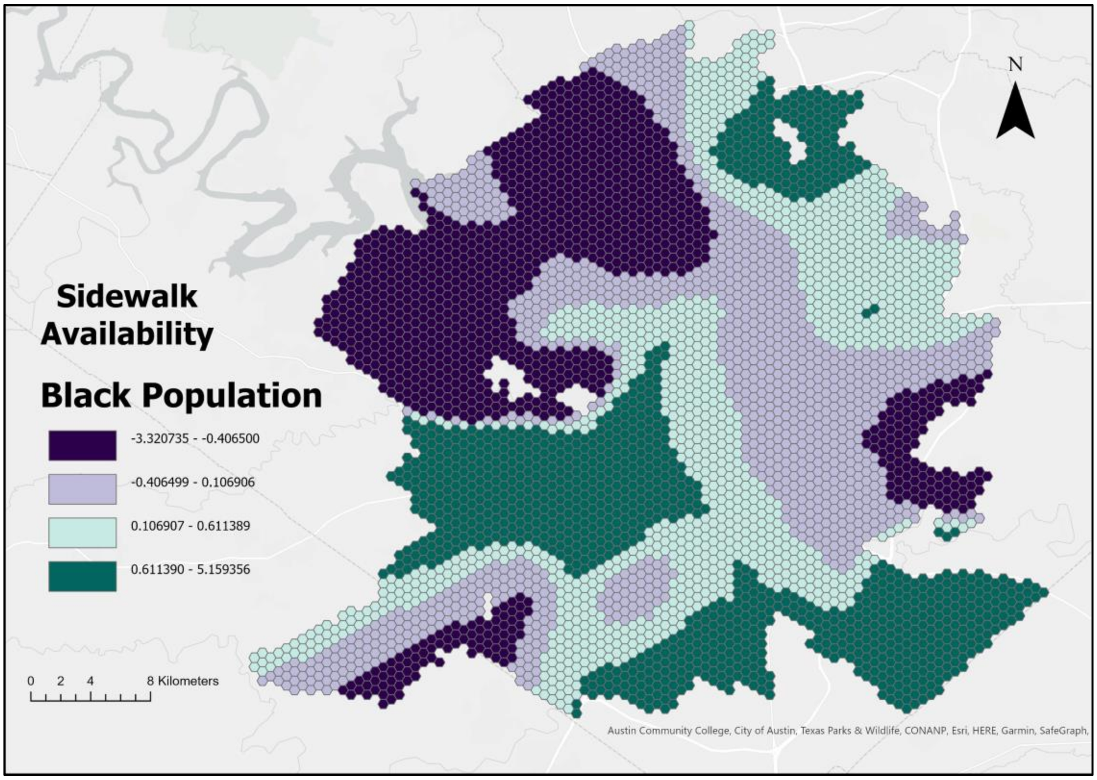


Figure 2-15 Spatial distribution of estimates for a proportion of Black

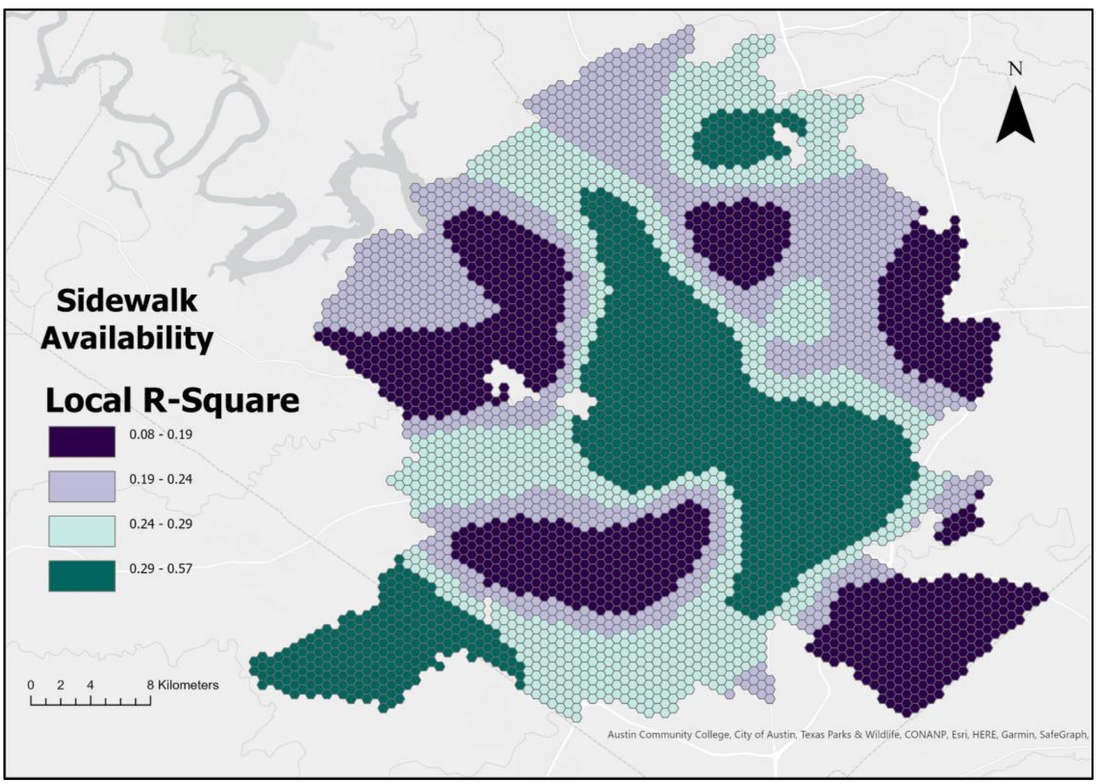


Figure 2-16 Spatial distribution of local R^2

Table 2-2 Summary of Diagnosis of the GWR Coefficient for Supply

	Access to Grocery				**Hospital**				**School**				**Sidewalk**			
	Mean	**Median**	**Max**	**Min**	**Mean**	**Median**	**Max**	**Min**	**Mean**	**Median**	**Max**	**Min**	**Mean**	**Median**	**Max**	**Min**
Intercept	0.9532	0.8614	2.4965	-0.7354	0.0630	0.0251	0.5891	-0.1546	0.2777	0.2256	2.2300	-0.7458	40.7355	39.8506	105.8710	-14.5126
Black	0.0131	0.0002	0.1336	-0.0405	-0.0007	-0.0007	0.0270	-0.0127	0.0043	0.0034	0.0985	-0.1119	0.1893	0.1068	5.1594	-3.3207
Hispanic	0.0052	0.0030	0.0624	-0.0182	0.0000	0.0000	0.0083	-0.0064	0.0007	0.0003	0.0767	-0.0350	0.1472	0.2083	3.7352	-1.2392
Public Transit	0.0336	0.0210	0.8091	-0.2958	0.0033	0.0009	0.1388	-0.0918	0.0161	0.0108	0.3146	-0.3699	2.0387	0.7772	19.9112	-13.7599
Biker	0.1031	0.0846	2.2705	-0.4258	0.0054	0.0009	0.1902	-0.1541	0.1126	0.1055	0.9977	-0.3681	4.9676	3.1291	78.2610	-25.6777
Walker	0.0383	-0.0004	0.9269	-0.2157	0.0028	0.0023	0.0456	-0.0348	0.0334	0.0172	0.8520	-0.1361	0.8187	0.4493	45.2436	-17.2686
Over 60	-0.0095	-0.0116	0.1054	-0.0750	-0.0007	-0.0004	0.0159	-0.0181	-0.0035	-0.0051	0.0632	-0.0862	-0.3676	-0.5703	5.9915	-2.0080
Low Income	0.0062	0.0084	0.0974	-0.2245	-0.0003	-0.0006	0.0144	-0.0166	0.0150	0.0173	0.1079	-0.2183	-0.3968	-0.1829	6.0225	-13.2489
Zero Vehicle Household	0.0310	0.0280	0.1333	-0.0403	0.0024	0.0014	0.0195	-0.0164	0.0134	0.0100	0.0896	-0.0297	0.8892	0.6912	7.2938	-3.2858
Under 18	-0.0307	-0.0313	0.0158	-0.1038	-	-	-	-	-0.0071	-0.0029	0.0377	-0.1236	-0.6070	-0.6420	1.9366	-5.2550
R Square			0.55		R Square			0.21	R Square			0.44	R Square			0.39
Adjusted R Square			0.53		Adjusted R Square			0.18	Adjusted R Square			0.42	Adjusted R Square			0.37
Sigma Squared			0.56		Sigma Squared			0.05	Sigma Squared			0.3	Sigma Squared			663
Sigma Squared MLE			0.54		Sigma Squared MLE			0.05	Sigma Squared MLE			0.29	Sigma Squared MLE			643
Effective Degree of Free Dom			4352		Effective Degree of Free Dom			43663	Effective Degree of Free Dom			4352	Effective Degree of Free Dom			4352
Adjusted Critical Value of Pseudo-t Statistics			2.83		Adjusted Critical Value of Pseudo-t Statistics			2.83	Adjusted Critical Value of Pseudo-t Statistics			2.83	Adjusted Critical Value of Pseudo-t Statistics			2.83

2.2.4.5 Where are the Walking Deserts?

As mentioned above, walking deserts were areas with considerable walking demand (i.e., the concentration of walking-dependent populations) and comparatively less availability of walking supply (i.e., lack of walking supply). Having evaluated the level of supply (WSI) and demand (ODI) for areas across the city, walking deserts were then determined. Walking deserts were subsequently defined as areas where the lowest walking supply (WSI) scores overlapped with the highest demand (ODI) scores. Walking deserts are displayed in orange in Figure 2-17. The blue spectrum colors show the concentration of supply (access to destinations and sidewalk availability), whereas dark blue indicates the highest level of supply. Similarly, the orange spectrum offers a level of demand (i.e., location of walking-dependent populations). As can be seen in Figure 2-17, walking deserts are mainly located in the eastern and southeastern parts of the city. These areas also overlap with areas of high Black, Hispanic, and low-education populations, as shown in Figure 2-6 presented earlier in the demand analysis. In contrast, areas with the lowest demand (i.e., low walking-dependent populations) and low supply are in the western and southwestern parts of the city (shown in light beige in Figure 2-17). The central and inner areas of the city were found to offer adequate levels of supply, given the population demand.

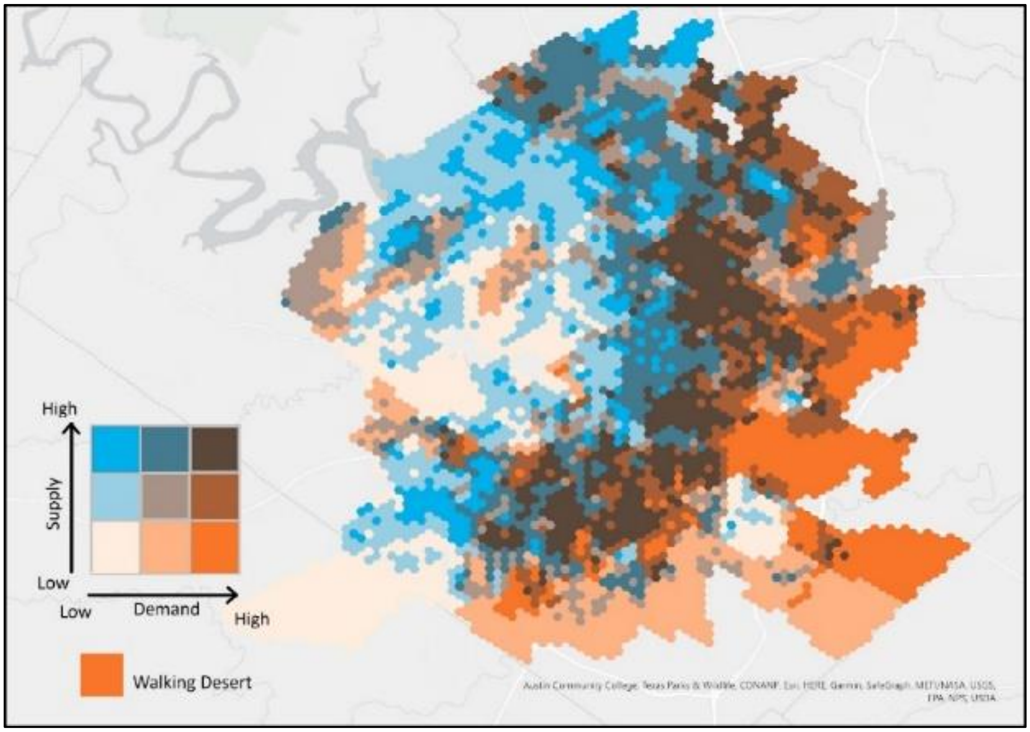


Figure 2-17 Spatial distribution of walking deserts

2.2.5 Conclusion and Policy Implications

National and local policymakers have increasingly advocated for active transportation modes, particularly walking, to expand multi-modal transportation options and to shift away from the car-centric atmospheres that exist in many cities. There also has been high demand for active transportation modes like walking among specific segments of the population including those with low mobility and physical ability, carless populations, and low-income minorities. The coupling of these factors created the potential for a feedback loop, where investments in walking supply (infrastructure and services) in walking-dependent neighborhoods could support sustainability, address unmet needs, and respond to public demands for multi-modal transportation. To support equitable investments, there has been a growing need for innovative approaches that would allow decision-makers to effectively monitor performance and efficiently manage resources.

In this study, we introduced and applied walking deserts, a structured, data-driven framework, that captured and compared walking demand and supply. Consequently, the framework revealed areas in need of investment and (re)development. Results from the regression analysis demonstrated that the availability of walking supply within different neighborhoods was significantly related to race, educational attainment, and income. For instance, our analysis showed that a quarter of Hispanic and Black populations, who often rely more heavily on public transportation services and active modes than other population groups, received zero walking access to schools or grocery stores. The disparity in access to medical facilities was even larger for these walking-dependent populations when compared to other populations. Our results were consistent with previous research that found associations between sociodemographic characteristics, disadvantaged populations, and walking services [60], [61].

This study also showed that geographically weighted regression (GWR) is a practical method for assessing the level of equity in the distribution of access resources. The study was noteworthy for being one of the first to use GWR in the context of walking accessibility and provide significant contributions to both the methodological and practical aspects of the literature. The results demonstrated that the GWR models greatly outperformed the OLS models. The GWR models showed that local variations might exist in the relationships between residents' demographic and socioeconomic status and level of supply. Furthermore, these variations could diminish the ability of global models to explain the data [61]. These findings provided strong evidence that GWR models were more suitable than OLS models for assessing the spatial distribution of access to destinations and opportunities. This conclusion was consistent with previous equity studies of locally unwanted land uses and urban parks [44], [60]–[62].

As previously reported by Kim and Nicholls (2017), the use of GWR also made it possible to widen the study question's focus. Finding "who gets what" in the context of environmental or spatial equity has historically been the primary objective of equity-related research in the literature on the urban environment and transportation planning [63]. Finding "who gets what" in the context of environmental or territorial justice has historically been the primary objective of equity-related research in the literature on urban service delivery. However, this study broadened the focus from "who gets what" to "who gets what, where, to what extent, and how much more considerably." Such information could direct state and municipal urban organizations, whose objectives include ensuring equitable urban services, by identifying the communities and locations that require improved walking service and infrastructure. In addition, this information could help minority communities, civic organizations, and local advocacy groups in their efforts to create or obtain fair access to walking services [61].

The walking desert framework ultimately allowed for focused attention on the mobility demands of vulnerable social groups. The mixed-method approach could help public and private organizations perform transparent, performance-based reviews of available infrastructure and services. Consistent implementation of the framework and tracking of outcomes could assist decision-makers by identifying policies and practices that work effectively and ones in need of improvement. However, despite its merits and contributions to walking accessibility research, this study had limitations that should be acknowledged. The assessment of walking access and social exclusion was a complex task. Many factors not considered here are, in fact, relevant to the analysis. For example, there may be a spatial mismatch between the location of certain services and a population's interest in using or accessing those services or facilities that are relevant to these findings. The use and availability of more fine-grained datasets on the everyday experiences faced by specific populations in their efforts to access services and use walking infrastructure could further enrich the spatial and statistical analysis presented here. Finally, the models used in this study were linear regression models and only considered linear interpolation. Therefore, they reflected certain limitations. Future work may seek to investigate the impact of other social factors, such as gender, and built environment characteristics, such as condition. Additionally, it is important to acknowledge that the limitations of this study include the fact that the R-squared values obtained indicate the presence of other unmodeled factors or potentially more complex relationships, underscoring the need for future research to delve deeper into these aspects and further investigate their implications.

The WSI formulation has limitations that should be acknowledged. First, sidewalk availability and access to POIs are weighted equally in the formula, assuming equal importance for each indicator in determining the overall score. However, in different contexts, indicators may have differing levels of importance for measuring walking supply, and WSI weights can be adjusted to reflect these conditions. Second, The WSI formulation always assumes some level of walking supply, even when one of the indicators is zero. For example, if sidewalk availability is zero, the averaging method still assigns a score based on the non-zero indicators, which may mask the specific challenges experienced while walking.

To address the limitations of the WSI formulation, several steps can be taken. First, a contextual analysis should be performed to determine which factors are more important for the specific neighborhood or location being assessed. This could involve surveying local residents and experts or conducting focus groups to identify the most pressing needs for walking infrastructure. Second, the weighting of different factors in the WSI formula could be adjusted based on the results of this analysis, so that factors that are more important in a given context are given greater weight. In general, addressing the limitations of the WSI formulation requires a nuanced and context-specific approach that takes into account the unique needs and characteristics of each neighborhood or city, and the flexibility provided by the averaging method allows for this.

2.3 Transit Deserts

2.3.1 Introduction

This section describes a method to identify transit deserts by quantifying and comparing transit accessibility (supply) and transit dependency (demand) using an equity lens that prioritized the most vulnerable among us. While previous studies have only used proximity or spatial indicators to describe the level of accessibility, this study utilized a more comprehensive set of spatial and temporal measures to estimate the level of accessibility across different census tracts.

Reviewing the transportation literature, there has been little focus on developing a holistic framework to assess the quality of transit from an equity perspective that moves beyond the narrative of access (i.e., cost and proximity to service). Traditionally, transportation studies have overlooked the equity considerations of transportation projects, and economic impact assessment has been used as the main measure of fairness in policy-making [14]. Moreover, current studies, despite the importance of the topic, have failed to sufficiently take into account equity considerations in transportation planning.

A more comprehensive framework for measuring equity was vital for examining whether transportation policies lead to a fairly distributed transportation system [64] or, in fact, create transit deserts. The significant difference between this study and existing studies dealing with accessibility and equity was the introduction of a public transit accessibility (PTA) index. The PTA was used to identify areas where the user needs were not being met related to service, coverage, and route efficiency. The remainder of this section is arranged as follows. First, a brief review of the literature on equity, accessibility, and transit deserts is provided, the study methodology is outlined, the case study analysis is presented, and a summary of the results is discussed.

2.3.2 Method

The first step in identifying transit deserts was to determine the demand for transit – that is, the location of transit-dependent populations. In this study, 10 different demographic factors were considered as measures of transit dependency. These factors focused on the mobility, need, and ability of users as well as characteristics of income and social class. The next step involved the identification of the supply. Transit supply was estimated as a function of the spatial and temporal components of the transit service. Thus, transit supply provided a holistic view of the level of accessibility afforded by the transit system. This study extended the concept of transit deserts beyond notions of infrastructure and service availability (i.e., the density of bus stops or high-frequency bus arrivals) to include broader ideas of good transit and the service-demand needs of the transit-dependent as discussed as follows. Finally, the PTA score was introduced as a holistic measure of accessibility.

2.3.3 Demand

Demand was estimated as a function of the number of transit-dependent populations present across the city. To identify transit-dependent populations, previous studies have grouped indicators of transit dependency into two distinct groups based on income and social class, and mobility need and ability, as outlined in Table 2-2. In this study, ten different indicators were considered to identify transit-dependent census tracts (CTs)

Table 2-2 Transit dependency indicators

Equity Indicators	Income and Social Class							Mobility Need and Ability					
	Income	Car ownership	Unemployed	Housing rent	Emigration status	Job level	Education	Youth	Seniors	Socially isolated	Distance from CBD	Disabled	Transit users
Frequency	7	4	3	2	3	3	2	3	5	1	1	2	2
[64]	■												
[65]									■				
[66]	■	■	■			■	■	■	■		■	■	
[9]	■	■						■	■				
[15]	■		■	■	■								
[6]	■				■	■	■						■
[67]	■	■							■				
[68]		■											
This study	■	■	■		■		■	■	■	■		■	■

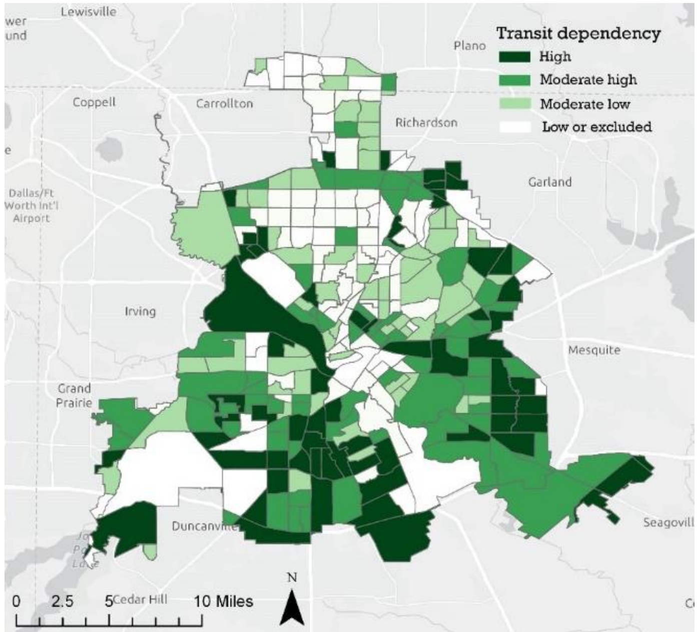


Figure 2-18 Transit-dependency level

2.3.4 Supply (PTA)

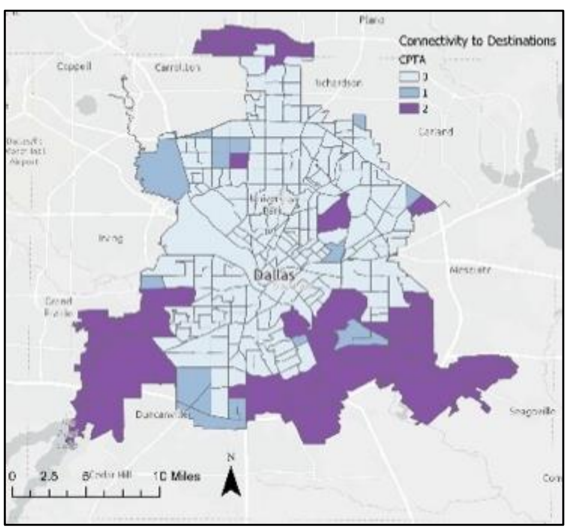
PTA estimated the level of supply across the city. To assess how good transit was distributed, relevant service characteristics were captured. Four factors were adopted in this study to illustrate the characteristics of good transit – connectivity to the network, connectivity to destinations, flexibility and time efficiency, and service frequency (Table 2-3). These factors captured the level of transit system accessibility from different temporal and spatial views to provide a more holistic understanding of transit service distribution. Each factor also had an associated set of indicators, shown in Table 2-3, that allowed for data collection and measurement.

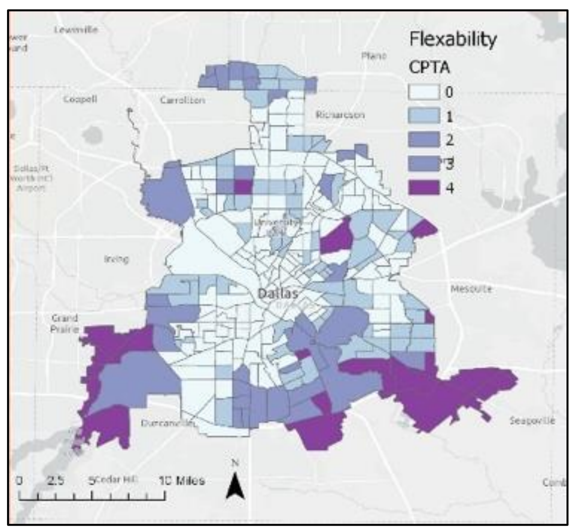
Table 2-3 Good Transit Factors and Indicators

Factor	**Indicator**	**Description**	
Connectivity to the network	Bus stop service coverage area	$SC_{CTi} = \frac{\sum SC_r}{A_{CTi}}$	SC: service coverage, CTi: ith census tract, A_{CTi}: total area of the census tract, r: number of the bus stops in a CT
	Density of bus stops	$DBS_{CTi} = \frac{r}{A_{CTi}}$	DBS_{CTi}: density of the bus stops in census tract, A_{CTi}: total area of the census tract, r: number of the bus stops in a census tract.

Factor	Indicator	Description
	Route coverage	$R_{CTi} = \frac{\sum L_j}{\sum St_i}$ R: route coverage, L: length of the jth route in CTi, and St: length of the street network throughout the census tract i.
	ADA accessibility	Number of bus stops with ADA accessibility
Connectivity to destinations	Accessible jobs in one hour	Good transit takes system users to their destinations. It is recognized that a travel time of 1-hour is not considered ideal for a one-way daily commute. However, this indicator assesses connectivity to destinations regardless of travel time.
	Accessible medical facilities in one hour	
Service frequency	Bus arrivals during weekdays between 6:00 am - 10:00 am	Good transit should provide reliable service throughout the day.
	Bus arrivals during weekends between 9:00 am - 4:00 pm	
	Bus arrivals during weekdays overnight between 12:00 am - 5:00 am	
Flexibility and time efficiency	Number of routes	Good transit provides users with options for where they can go and gets them there in a reasonable amount of time.
	Accessible jobs in 30 minutes	
	Accessible medical facilities in 30 minutes	
	Accessible grocery stores in 30 minutes	

The outlined methodology identified areas where transit supply was inadequate for each of the four-factor areas: connectivity to the network, connectivity to destinations, service frequency, flexibility, and time efficiency. Areas of inadequate transit supply are shown in Figure 2-19.





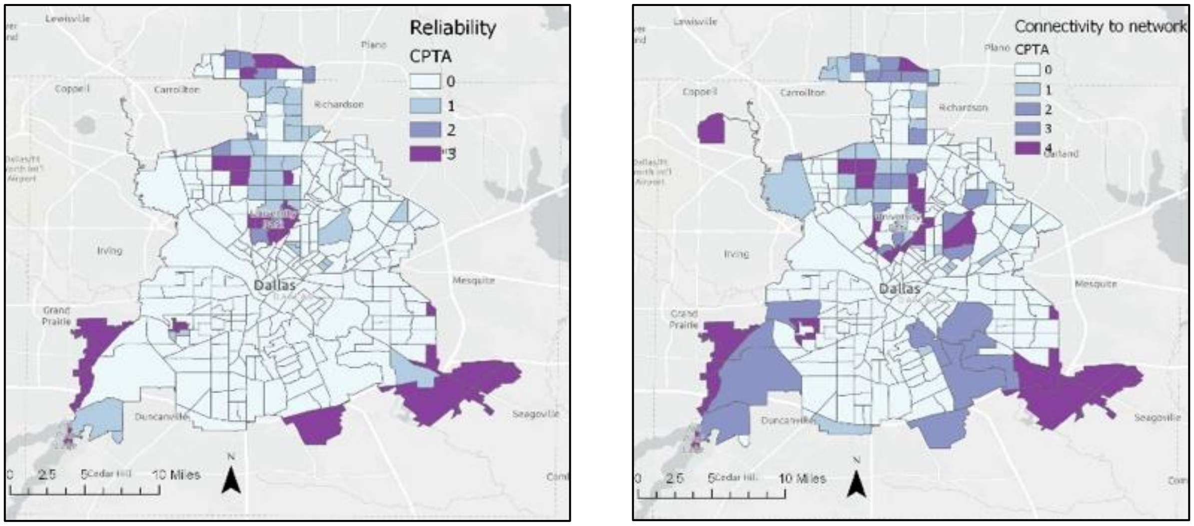


Figure 2-19 Estimation of supply – Factors of good transit

PTA score was calculated for each census tract. PTA score equaled the number of times a CT's indicators values were in the lowest quantile across all CTs (maximum PTA score was 13, where the CT was among the worst CTs in all 13 indicators). CTs that were based on their overall PTA score, were grouped by quartiles labeled as high to low PTA. Figure 2-20 shows the overall results of the PTA score by census tract.

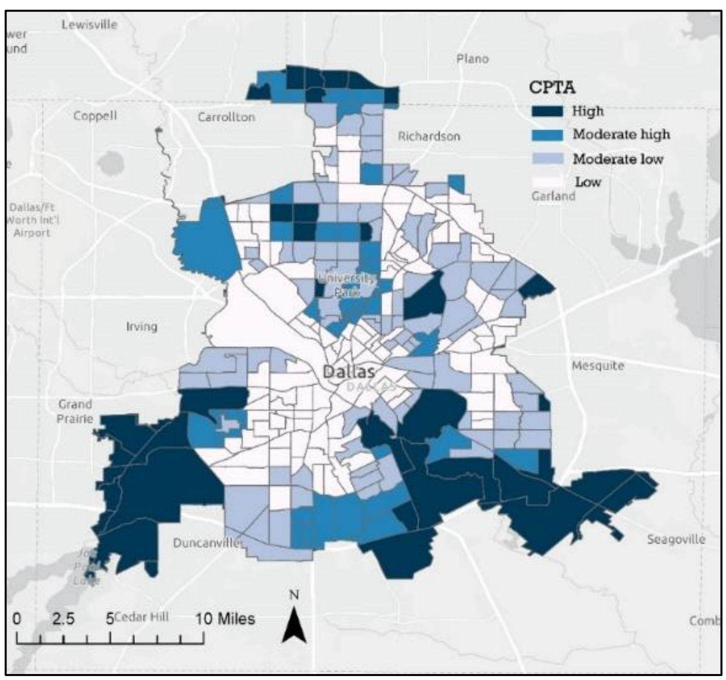


Figure 2-20 Estimation of transit supply

2.3.5 Transit Deserts

Transit deserts were identified as areas of overlapping high transit dependency and low transit supply. Figure 2-21 shows these areas or census tracts. Figure 2-21 shows that Transit deserts were mainly located in southern parts of the city and were in general areas where the PTA score reflected low supply across all transit supply factors. The highest PTA score scores were found in the northernmost and southeastern census tracts of the city. In these census tracts, levels of transit supply were found to be low across all four factors (connectivity to a network, connectivity to destinations, service frequency, flexibility, and time efficiency). These areas could thus be viewed as having the greatest need for access to transit while offering the least amount of transit supply. Figure 2-21 identifies transit desert areas by three levels or categories. Extreme transit deserts were areas with high transit demand and low supply. High transit deserts were areas where demand or need was high and supply was high. Moderate transit deserts were areas where demand and supply were moderately high and moderately low, respectively.

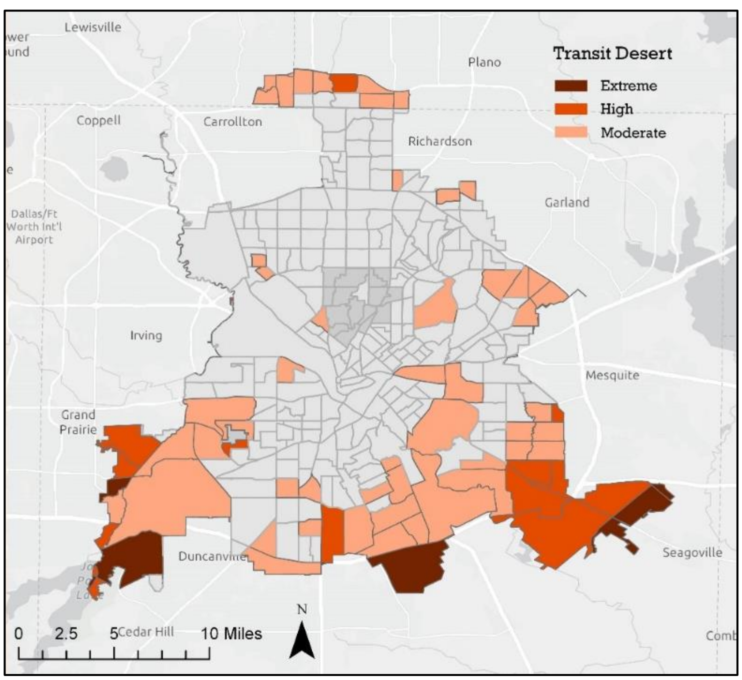


Figure 2-21 Transit Desert levels

2.4 Conclusion and Policy Implications

This study introduced a method to identify city-level transit deserts using a more comprehensive public transit accessibility (PTA) index that examined supply indicators such as level of connectivity to destinations, service flexibility, service frequency, and time efficiency of transit service and compares them to the needs of transit-dependent populations. In cities across the U.S., resources, wealth, and services are not fairly distributed among the population. Equity requires that services be distributed to groups who do not have access to opportunities or cannot afford opportunities based on their determined level of need. Race, income, social class, mobility needs, and disabilities are characteristics that limit access to opportunities and socially exclude individuals. Those characteristics help to identify the transit-dependent. Measures of supply, such as connectivity to the network, connectivity to destinations, service frequency, and flexibility, help to determine the distribution of transit supply among different communities and groups. Areas, where disadvantaged and transit-dependent populations are provided with inadequate amounts of transit supply, can be labeled transit deserts.

In this study, a structured framework was developed to identify transit deserts across the city. Identifying transit deserts could provide valuable insights for decision-makers in urban planning, serving as a powerful tool to reveal transportation needs and challenges within underserved communities. This information could help prioritize investments, promote equity, make data-driven decisions, engage stakeholders, and monitor and evaluate transportation services in these areas:

1) Targeted Investment – By identifying areas that lack adequate public transportation options, decision-makers could prioritize targeted investments in those areas to improve access and connectivity. This prioritization could involve investing in new transit infrastructure or providing subsidies for on-demand transit services, as well as improving pedestrian and cycling infrastructure to enhance mobility. 2) Improved equity – transit desert concept could help decision-makers to promote greater equity in transportation

planning by ensuring that transportation investments and services are distributed fairly across all communities. Fair distribution could help reduce social and economic inequalities and improve overall access to essential services. 3) Data-driven decision-making– The transit desert framework could provide decision-makers with data-driven insights into transportation patterns and needs in underserved areas. This could help inform planning decisions and ensure that resources are used efficiently to address the most pressing transportation challenges. 4) Stakeholder engagement – transit desert concept could facilitate greater engagement and collaboration between decision-makers, community organizations, and transit providers. By working together, these stakeholders could develop more effective transportation solutions that meet the needs of underserved communities. 5) Monitoring and evaluation – this framework could help decision-makers to monitor and evaluate the impact of transportation investments and services in underserved areas. This framework could help identify areas of success, areas for improvement, and lead to more informed and effective transportation planning in the future.

Additionally, in this exploratory study, demographic data and GTFS open data were used to determine the level of transit demand or transit dependency. Our empirical analysis of transit deserts highlighted how comprehensive spatial indicators (connectivity to the network, connectivity to destinations, service frequency, flexibility, and time efficiency) could provide a more holistic analysis of transit deserts. Empirical analysis highlighted the limitations of more traditional frameworks that utilized one-dimensional indicators such as connectivity to the network. Traditional studies often considered proximity to bus stops and routes (or in this study connectivity to the network) as the only accessibility indicator.

This study used the concept of transit supply to represent a broader definition of accessibility than just the availability of transit. This study considered, for example, that bus arrivals differ by time of day and day of the week, and that different communities have different levels of access to opportunity via transit. This study integrated these differences into the analysis. Accessibility or transit supply, as characterized in this study were more than just a function of access to the network. Transit supply included access to destinations, service frequency, and flexibility. This study showed how various temporal (time of day or week, frequency, and span) and spatial (proximity and density of infrastructure, opportunities, and service area coverage) indicators could be incorporated to identify transit deserts.

Multiple indicators were used to determine the level of transit supply available to census tracts within Dallas, Texas, the case study presented in this research. To address transit deserts in Dallas, TX., several practices could be implemented:

Transit Infrastructure and Services – consider the distribution of transit supply across transit deserts, increasing the frequency and hours of service for existing public transportation routes could improve access for people in transit deserts. This study considered adding more buses to existing routes, or extending service hours to accommodate people who work late or have other obligations. Moreover, improving existing public transportation infrastructure such as bus stops, and transit hubs, in addition to developing new transit corridors or extending existing transit lines to serve underserved areas could make public transportation more accessible and convenient for people in transit deserts.

Shared micromobility –could be a cost-effective solution to provide last-mile and first-mile services or even as a public transportation alternative in transit deserts. Shared micromobility would be a viable alternative from investing in new bus routes that would require higher investments and would be highly dependent on population density. Policymakers could create incentives for shared micromobility companies to better serve disadvantaged neighborhoods by reducing e-scooter application and permit fees, as evidenced by the positive impact on overall e-scooter usage in cities like Portland, OR [69]. However, to ensure affordability for low-income households, it would be important to provide optimal incentives and subsidies to encourage their use. Shared micromobility services could significantly enhance accessibility in underserved areas. However, it is crucial to consider the socio-economic factors of the target population to ensure equity and inclusivity in the transportation system.

On-Demand Transit Services – could help address transit deserts by providing more flexible and convenient transportation options to people living in underserved areas. These services could offer customized routes, better access, improved frequency, and cost-effective solutions. The use of innovative technology solutions, such as mobile apps and real-time tracking, could improve the overall rider experience and make public transportation more appealing to people in transit deserts. On-demand transit services could help reduce barriers to access essential services and opportunities for people living in transit deserts. On-demand transit services could make public transportation more equitable and sustainable. Decision makers could address transit deserts using on-demand services by taking a data-driven and community-led approach, defining service parameters, choosing a service model, developing a marketing and outreach plan, and monitoring and evaluating service performance.

CHAPTER THREE: GENDER EQUITY

3.1.1 Introduction

Gender disparity in transportation has been an unsolved problem, and to achieve equity, data has been needed [12]. The lack of sex-disaggregated mobility data is the central problem when addressing women's access to city facilities. [70]. Evidence-based, gender-informed public policies are vital to enhancing women’s experience with transportation services, their overall wellbeing, and urban mobility. However, it is impossible to develop public policies that are focused on the requirements of users without accurate data that provides insights on who, how, and for what purposes transportation services are being used, which is especially true for women. This data gap is particularly troubling given that the available data that suggests that women experience the most hardship using transportation [70]. Historically, a significant gender gap exists in cycling. This gap has been one of the most glaring examples of gender inequality in transportation, and this dynamic has continued in city after city. However, there has been very little information available about scooters. This study has aimed to help close this gap by performing a gender-sensitive analysis to reveal differences in men and women’s perceptions and experiences toward shared electric scooters.

This study conducted a first-of-its-kind empirical analysis using text mining of online data to examine rider satisfaction with scooter use and factors impacting overall satisfaction. App-store reviews from two major micromobility companies, including Lime and Bird, were investigated using machine learning techniques to identify the factors that influenced rider satisfaction across gender. Specifically, this study aimed to:

- extract general topics discussed within the reviews,
- identify how different topics coexisted in app reviews,
- examine how sentiments varied across topics for men and women, and
- investigate how topics were associated with rider satisfaction across gender.

The research methodology consisted of four primary steps. We first used term frequency - inverse document frequency (tf-idf) and the latent Dirichlet allocation (LDA) model, a widely used topic modeling technique, to discover the hidden topical patterns in e-scooter app store reviews. We then examined sentiments across topics by applying polarity analysis to labeled reviews under each extracted topic. Third, we developed topic-coexistence networks to identify how different review topics coexisted. We also adopted a novel approach, i.e., name-based gender prediction, to ascertain user gender for a subsample of app reviews and to determine whether review topics and sentiments varied across gender. Finally, we developed logistic regression models to investigate the key factors contributing to rider satisfaction across genders.

Research findings contributed to the existing literature in several ways. Limited research, to date, has utilized app store reviews to investigate the opinions of e-scooter users. Factors that influenced rider satisfaction were still misunderstood as there had been limited use of traditional techniques (e.g., surveys) to examine e-scooter rider satisfaction or sentiment. Moreover, very few studies have investigated the impact of gender as an attribute relevant to riding experience and satisfaction. Assessing factors that contributed to user or rider satisfaction, therefore, provided the ability to evaluate current and future barriers to e-scooter access, rider needs, and willingness to ride. These results ultimately will have provided companies, planners, and policymakers with the information needed to implement a consistent, effective, and integrated strategy for improving the e-scooter experience.

3.2 Background

While scooter companies have launched a battle for space within urban transportation corridors, scooters have yet to receive sufficient attention. Though scooter studies have been on the increase, much remains to be learned about scooter interactions with pedestrians and bikes or the increasing complexity of the transportation systems [71], [72]. Research to date has investigated topics such as e-scooter safety [73]–

[77], micromobility systems (design, regulations, and policies) [71], [78]–[80], parking [81], [82], temporal and spatial distributions [83]–[86], business models[87], and social and cultural issues [88].

Nevertheless, no study has yet investigated factors impacting scooter-rider satisfaction. Whereas, a growing body of literature has examined factors affecting bike rider or transit user satisfaction [89]–[92], prior scooter studies have addressed safety, spatial-temporal distributions, and policies. On the other hand, only very few studies have attempted to investigate rider behavior. Thus, scooter rider characteristics, expectations, and demands have still remained unknown. Moreover, there have been relatively few studies that have focused on text mining as a tool to investigate rider satisfaction.

In general, little work has been done in the mining of app store reviews [93]–[96]. App store review-mining studies were mostly undertaken for app development purposes (e.g., bug removal and feature requirements). The app reviews were also undertaken to inform business decision-making (e.g., user willingness to pay for specific paid apps or services). The textual content available on app stores, like other social media platforms, has provided public beliefs, sentiments, and needs about services and products. However, few studies have used this data source to understand user needs and experiences. A review of the literature revealed that text mining was frequently used as a tool to investigate transportation issues. Text mining techniques have been used in various areas of transportation studies. Yang et al. (2020) investigated 167 news reports related to scooter crashes and highlighted the distinctive characteristics of these crashes [77]. Das et al. (2015) examined the popularity of shared bikes in Washington, D.C., by performing a sentiment analysis of user comments on Twitter. Finally, a case study in Korea reported that using text-mining approaches to understand factors influencing bike sharing service quality contributed to improvement in public service [97]. Advances have been made in the use of text mining in transportation literature, although very few studies were found that focused on app store reviews as a source for textual data.

Predicting gender through online data has had an important role in analyzing the gender gap [98]. Historically, it has been difficult to obtain individual gender data. The anonymity and privacy policy of social media has made it even more difficult or impossible to acquire ender information from social media. In response, gender-prediction methods have been widely used for analyzing gender disparities in science in many published articles. These methods could be enhanced by choosing the most suitable prediction method for a given purpose with optimal parameters. The validation studies could then use a reliable data source. However, without having a suitable prediction method, most prior gender research with social media was made by using small-sized or large-sized data. Gender information could be collected without optimal parameters set. If unavailable, gender information had to be manually annotated by researchers, or simple estimation approaches were adopted based on relevant cues such as names (Table 3-1)[99].

Table 3-1 Examples of name-based gender prediction studies

Previous Work	**Purpose**	**data**
[100]	Hurricane evacuation behavior modeling using social media	Twitter
[101]	Gender prediction	Reuters newsgroup dataset
[102]	Gender prediction	-
[103]	The Geo-temporal Demographics of Twitter Usage	Twitter
[104]	Gender analysis (gender was estimated and used for studying the impact of gender on both linguistic style and social networks in social media text)	Twitter
[105]	Understanding the demographics of Twitter users	Twitter
[106]	Online Sentiment during Hurricane Irene	Twitter

3.3 Methodology

To identify the factors of satisfaction, first, topic modeling was implemented to extract hidden topics discussed within the reviews. In the next step, logistic regression was used to determine the dominant factors that influenced satisfaction. Additionally, name-based gender prediction enabled us to shed light on the differences between male and female riders' sentiments and factors of satisfaction.

3.3.1 Data Extraction

The textual data used in this study was associated with two of the major private micro-mobility companies in the U.S. (Lime and Bird) and extracted from reviews on Google Play Store and Apple App Store. Each review consisted of the title, text or review content, rating, date, username, and app version. For this study, 12,026 unique reviews from May 2019 to January 2020 were mined.

3.3.2 Gender Prediction

One of the goals of this study was to investigate how satisfaction levels expressed in the reviews varied across gender. However, user gender was not specified within the extracted dataset. Therefore, in order to predict gender, we implemented name-based classification methods. Name-based methods classify user usernames to predict gender. As not all usernames contain "actual" names, this method was only able to predict a portion of the dataset (almost one-third in this study). First names were matched with historical databases. These databases provided the probability of whether names belong to men or women, and the gender with the higher probability match was assigned to a review [107], [108]. In this study, the "gender" open-source R package, which utilized the United States social security and census databases, was used to predict gender [109]. Gender prediction accuracy was identified as the average probability of specified genders. Gender prediction was challenging as many usernames did not contain actual names (e.g., FakeGenie1 and Freeotz) or often reflected names that were not available in the databases consulted. To evaluate the prediction method, name-based gender prediction was implemented on two other social media datasets. Both test datasets were open source and available for download and use. Both datasets were labeled and, therefore, appropriate for use as ground truth datasets (the result of the test can be found in the appendix).

3.3.3 Topic Modeling

One of the main objectives of this study was to uncover the common topics (factors of satisfaction) within the reviews. Topics were the service-related factors frequently discussed or mentioned within user reviews (e.g., pricing, safety, or parking). Topics were also assumed to inform how services were rated or the level of satisfaction/dissatisfaction expressed by users. In topic modeling, Brett (2012) described topics as "recurrent patterns of co-occurring words." Topic modeling was an unsupervised machine learning approach that scanned sets of documents, and discovered clusters that were hidden related groups of words (i.e., topics). In situations where pre-categorized and labeled documents were not available or expensive to achieve, unsupervised learning approaches were used to classify datasets. The following sections elaborate the steps we used to extract latent topics in app reviews.

3.3.4 Preprocessing

Preprocessing transformed the original raw texts into a "data-mining-ready" dataset [110]. Preprocessing started with tokenization. This method divided the content of the text into a list of single characters called tokens [111]. Next, all uppercase tokens were converted into lowercase. Then, non-informative terms such as stop words (e.g., "the", "to" or "me"), and common words (e.g., "scooter" or "ride") in addition to numbers and punctuations were removed. To correct typos, words with minor errors were replaced with the correct ones using the "hunspell" R-package [112]. Alterations were suggested by looking up a similar token in the dictionary. Furthermore, inflected words were transformed into their roots (lemmatization). The modified words had an equal part of speech label, the same semantic meaning, but a different syntax [113]. Finally, parts of speech tags (POS) were assigned to each token. POS tagging assisted in the topic modeling step, where only adjectives and nouns were used to identify and extract features. Albeit different studies have used different POS tags for this matter. For instance, Benamara et al. (2007) suggested using adverbs and adjectives in semantic analysis, and Lucini et al. (2020) used only nouns for topic modeling [114], [115]. An example of preprocessing results is shown below:

- Before preprocessing: "Brake was in poor condition. So expensive. BTW, why the app continues to attempt to grab my GPS location?
- After preprocessing: "brake be poor condition expensive why app continue attempt grab gps location"

3.3.5 Feature Extraction

In this paper, term frequency-inverse document frequency (tf-idf) was used to identify and extract features. Features (keywords) were a collection of important words in a text that were able to provide a significant depiction of the content [116]. The tf-idf statistical technique not only considered the frequency of the term use but the importance of terms within the text. In other words, this technique helped to identify important words that did not appear frequently within the content [117].

3.3.6 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA), has been a frequently used, popular generative probabilistic approach that was frequently utilized for fitting topic models [118]. The LDA approach relied on two basic principles: first, each document contained a collection of topics that were sampled for that document; second, each topic was a collection of words (Equation 3-1). Krestel et al. (2009) described the modeling process of LDA as:

$$P(w_i|d) = \sum_{j=1}^{K} P(w_i|\, k_i = j)P(k_i = j|d) \qquad \text{Equation 3-1}$$

$P(w_i|d)$: Probability of the ith word considering a given document.

$P(w_i|\, k_i = j)$: Probability of the w_i in topic k.

$P(k_i = j|d)$: Probability of choosing a word from topic k in the document.

The objective of using the LDA approach was to discover a collection of topics in each document (i.e., $P(k|d)$). Each individual topic was defined by words based on another probability distribution (i.e., $P(w|d)$). LDA discovered $P(w|d)$ and $P(k|d)$ by utilizing Dirichlet priorities for the required distributions in addition to the pre-selected number of topics (K) [119]. After fitting topic models, researchers generally interpreted the top-twenty most related words per topic and assigned a "descriptive label" to each topic for presentation purposes. Debortoli et al. (2016) warned that by following topic modeling, researchers were likely to face low-quality topics that were "mixed" (i.e., contained subsets of words belonging to more than one topic), "identical" (i.e., semantically the same words in two topics), or "nonessential" (e.g., irrelevant topics). Therefore, Debortoli et al. recommended that low-quality topics be split, merged, or excluded [120].

3.3.7 Polarity Analysis

Dictionary-based polarity classification was conducted to interpret user sentiment (positive, neutral, or negative) toward identified factors of satisfaction. Polarity classification categorized opinions based on the feelings conveyed within the text. Polarity analysis could be performed on reviews of different levels, such as a word, a sentence, or a document [121]. Dictionary-based classification has been one of the most popular sentiment analysis approaches. This method has relied on predefined labeled lists of words (lexicons) that specify the associated sentiment level or value of the words [120], [122]. Polarity analysis was performed by the "sentimentr" R package [123]. Methods presented by Rinker (2019) were adopted in this work for polarity value calculation.

3.3.8 Logistic Regression Analysis

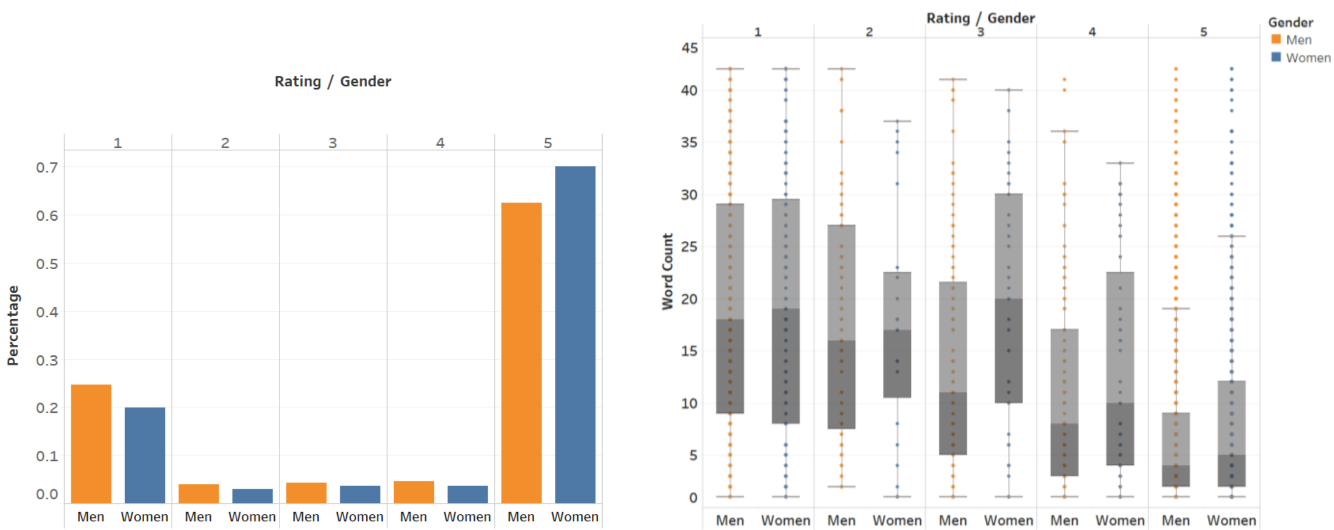
The unsupervised character of topic models provided no "ground truth" to test the outputs [120]. However, previous studies [115], [120], [124] have proposed regression analysis to quantify the impact of the extracted topics (independent variables) on user satisfaction (dependent variable). As discussed previously, app stores enabled users to rate their overall satisfaction level based on a five-star Likert scale. The star rating used an ordinal scale, where one and two stars represented high and relative dissatisfaction, three stars implied a neutral stance, and four and five stars indicated a relative and high degree of satisfaction. Consequently, in this section, a one-star review was considered as overall dissatisfaction, two, three, and four-star reviews were discarded, and five-star reviews were deemed as overall satisfaction. Two (and four) star reviews were dismissed, due to their mixed feelings, because a degree of satisfaction (dissatisfaction) could be found in their content. Therefore, considering the binary nature of the dependent variable (satisfaction or dissatisfaction), the logistic regression approach was identified as being reasonable for modeling. In the regression process, independent variables were represented by the weights of satisfaction factors identified using LDA, and the dependent variable was whether the user was satisfied (1) or not (0).

3.4 Results

In this study, statistical and machine learning techniques were utilized to investigate topic trends and explore the patterns within the reviews. In the sections that follow, detailed results regarding descriptive analysis, polarity classification, topic modeling, and topic validation are presented.

3.4.1 Gender Prediction and Descriptive Analysis

As usernames did not necessarily contain actual names, name-based analysis was only applicable to 37 percent of the username reviews. The results showed that, from the predictable observations, 71 percent of users were men (i.e., 3197 reviews), and 29 percent were women (i.e., 1256 reviews). The average accuracy of the gender prediction score was 0.96. There were significantly more male reviewers than female reviewers. The gender of the reviewers was consistent with the existing literature that micro-mobility was not gender-neutral [125], with substantially more male early adopters [126]. The percentage of App Store ratings generated by female and male users, as shown in Figure 3-1-a, suggests that most users submitted ratings of the lowest (1) star or highest (5) stars. A J-shaped distribution in a five-star online rating was not an uncommon phenomenon (Debortoli et al., 2016). Pearson's chi-square test revealed a significant association between gender and rating $(\chi^2(4) = 22.713, pvalue < 0.001)$. Women were more likely to submit higher ratings than men. The average rating from men was 3.7 (Std Dev: 1.72) and 4.01 from women (Std Dev: 1.62). Figure 3-1-b shows the word-count distribution by review (i.e., the total number of words in a review). What stands out is that the lower-rated (1 and 2 star) reviews were noticeably longer (higher word-count) and potentially more informative than the higher-rated reviews (4 and 5 stars). On the other hand, highly satisfied riders (5-star ratings) used considerably fewer words to describe their experience.



a- Rating distribution

b- Word count distribution

Table 3-2 Word count distribution app rating among men and women

3.4.2 Topic Modeling

Topics ranging from 1 through 30 were investigated to determine the optimal number of topics. Given the results of prior researcher and researcher knowledge, 10 - 15 topics were selected as a reasonable number of optimal topics. Each set of topics was qualitatively assessed for cohesiveness and consistency. The final analysis resulted in 12 topics, shown in Table 2. As the process of topic modeling was an unsupervised learning approach, human involvement was required to label the topics [127]. Each topic needed a label to illustrate the content of the topic clearly. Labels were tagged, one by one, by a first researcher, then a second, and a third researcher verified the results [128]. Each topic consisted of a set of closely related words. The labeling process was performed using the set of 20 words with the highest probability in each topic. For instance, the Customer Service topic was generally about contacting customer service, as both “customer” and “service” appeared as words on the topic list. Moreover, other words such as “phone,” “report,” “help,” and “reply” were also semantically linked to the act of contacting customer service. Similarly, the frequent appearance of terms in a particular topic such as “price,” “expensive,” “cheap,” and “cent” suggested that the topic was related to oricing or the monetary value of the service.

3.4.3 Topics Distribution and Coexistence

In addition to considering topics as a mixture of words, LDA was also able to estimate each review as a mixture of topics. For each review, we had estimated review and topic probabilities (also called “gamma”), which were the expected fraction of terms or words within a review that were generated from different topics [129]. Table 4-1 illustrates a sample review and its corresponding gamma value for each topic. For example, in the first row of Table 4-1, a gamma equal to 0.29 denotes that 29 percent of the words in that review were generated from the refund topic. It should be noted that one review could belong to more than one topic. Gamma equal to 0.125 was specified as the threshold for assigning reviews to the topics. In other words, reviews with gamma lower than 0.125 were removed for further analysis. As a result, a review could be assigned to more than one topic. The topic was then reusable, because users frequently discussed more than one topic in a single review. This threshold was first manually checked by authors to examine if there was a semantic connection between reviews and the label of topics. Next, using co-occurrence analysis, a sensitivity analysis was performed to understand how different values of gamma assign reviews to the topics. Gammas lower than 0.125 provided inaccurate assignment, and higher thresholds resulted in an excessively strict designation (coexistence networks for gammas equal to 0.10 and 0.15 are provided in

Appendix A). Figure 3-2 indicates the coexistence network of topics considering gamma equal to 0.125 for men, women, and total reviews. Thicker edges described a higher probability of coexistence of the topics in one review. As expected, the following topics often existed in one review: payment, refund and customer service; safety (speed and scooter lane) and safety technical; unlock and payment; pricing and payment; customer service and lock or unlock; and app issues and map. There was a higher coexistence of payment and app issues topics in women's reviews. Ease of use had the lowest coexistence among all topics in men and women.

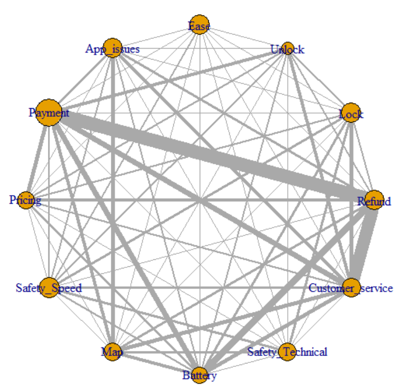
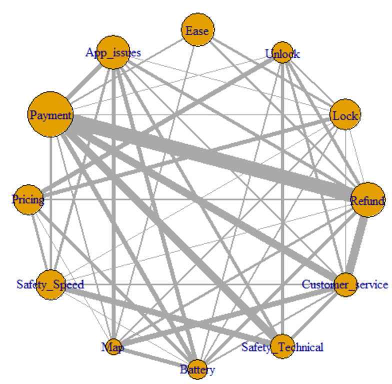
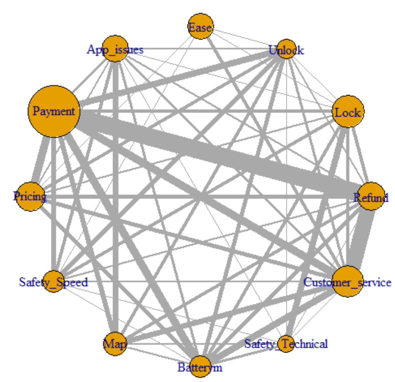
Total	Women	Men
		

Figure 3-1 Topic coexistence network

Table 3-3 Topic descriptions

Topic		Description	Frequent Relevant Words	Example	Gamma
1	Refund	Customers usually asked for refunds in exchange for the return of purchased credit, wrong charges, or when service was unsatisfactory or unfulfilled.	Money, account, refund, support, call, multiple, contact, terrible, month, unable, response.	…. I tried to get a refund of the unused money they were holding in my account goes by increments of 5 I think but after contacting customer service more than once I was still unsuccessful….	0.292
2	End of trip/ lock	To end the trip, and not be charged anymore, scooters should be locked. Reviews often were concerned with finding the proper spot, broken device, or unfamiliarity with the process.	Park, lock, bad, zone, slow, free, wait, spot, push.	…we cruised by finding a parking spot was irritating because the gps wouldn't calibrate to precise spot.	0.206
3	Ease of use	Users usually shared their positive renting and riding experiences in their review.	Easy, fast, pretty, system, access, destination, super, smooth, friendly.	easy to set up easier to ride quite fast as well will definite get some friends to ride	0.248
4	Unlock/ start	To start the trip, users usually first needed to scan the barcode and unlock the scooter. Companies usually charged riders even for unlocking the scooter.	Start, unlock, friend, brake, move, die, low, rate, foot, horrible, wrong, expect.	First time using the app the scooter didn't work [..] Tried to start it I immediately cancelled the ride at least [other company] has a much better app you can report broken ones and they [..] wave fees when you cancel within a few seconds of unlocking.	0.206
5	App issues	User complaints also contained functional errors, feature requests, and app crashes.	Email, sign (in/out), load, login, crash, screen, button, mode.	the app keeps getting stuck on the launch screen I was able to sign in the very first time I installed it but all subsequent times it was stuck on the splash [..] tried reinstalling it but same issue	0.353
6	Payment	Under this topic users usually expressed challenges regarding payments with credit/debit cards, their account balance, or asked for alternative payment methods, such as PayPal or their preferred credit card.	Pay, card, credit, option, add, payment, worth, balance, update, amount, auto, feature, bank.	please add options to pay with debit card maestro ban contact PayPal i don t have a credit card.	0.360
7	Pricing	Customers provided their preferences and opinions about pricing and often compared different alternatives.	Time, duration (hour, minute, or second), price, expensive, cheap, change, check, cent, top, ridiculous.	I was having so much fun riding these but they raised the prices its 31 cents a minute now unless they change it back to 15 cents, I won't be using these anymore.	0.264
8	Safety (Speed, scooter driving lane)	This topic considered safety concerns related to speeding, riding conditions, and riding lanes.	Drive, quick, safe, street, transportation, road, mph (mile per hour), light, sidewalk, fall, license.	scooter slows down to a walking pace in random places it considers pedestrian zones even they are not busy why is it illegal to ride on the sidewalk. this seriously endangering riders who are trying to keep up with traffic in the bike lane or on the street or while crossing the street as was the case for me.	0.378
9	Map/ juicer	Map and juicer was a mixed topic related to the app maps used for finding available scooters and juicers. Juicers were contact employees who picked up scooters and charged them at their homes, earning money for each charge. Juicers used a real-time map that specified ready-to-charge scooters.	Company, day, map, fix, location, pick, night, drop, close, bring, juicer, respond, reserve.	app has a frustrating zoom feature if you zoom out too far city view harvesting picking up and serving dropping off locations disappear you have to tediously zoom in on really confined areas of the map to find chargeable scooters and areas to drop them off …	0.438
10	Battery	Battery charge level was represented by a bar graph on scooter screens (this level highly influenced the rider's trip duration).	Battery, charge, delete, error, bar, message, finally, offer.	… the battery percentage would be easier than battery points the overall mileage will change based on acceleration and braking so does the battery points but battery percentage will help us to know whether to pick or not [..] [moreover] being able to interchange the battery with our own is always a boon.	0.248
11	Safety (Technical issues)	This topic included technical issues that contained subtopics such as complaints about scooter parts (tires, brakes, etc.) as well as reports of damages and accidents.	Break, car, hit, tire, dangerous, damage, brake	These things are dangerous traffic passes you within inches of handlebars wheels are so small and non-air filled yet have no flex steering is scary loose never again do not ride f ing dangerous	0.27
12	Costumer service	Users often described their experience by contacting the company via customer service.	Service, customer, phone, week, happen, business, report, hold, receive, product, help, reply.	…worst costumer service [..] was assured I would be receiving a call back regarding a lost item …	0.25

Figure 3-3 illustrates the percentage of reviews in the dataset related to each topic and presents the distribution of topics across gender. Figure 3-3 shows that payment is the most discussed topic within the reviews for both men and women. More than 14 percent of total reviews address the payment process. The least discussed topic is map and juicer for women, and safety (technical issues) for men. Although there was no significant difference between male and female users in the distribution of topics, it seems men were more concerned than women with payment and customer service. Compared to men, women's reviews were more associated with ease of use and safety-related technical issues.

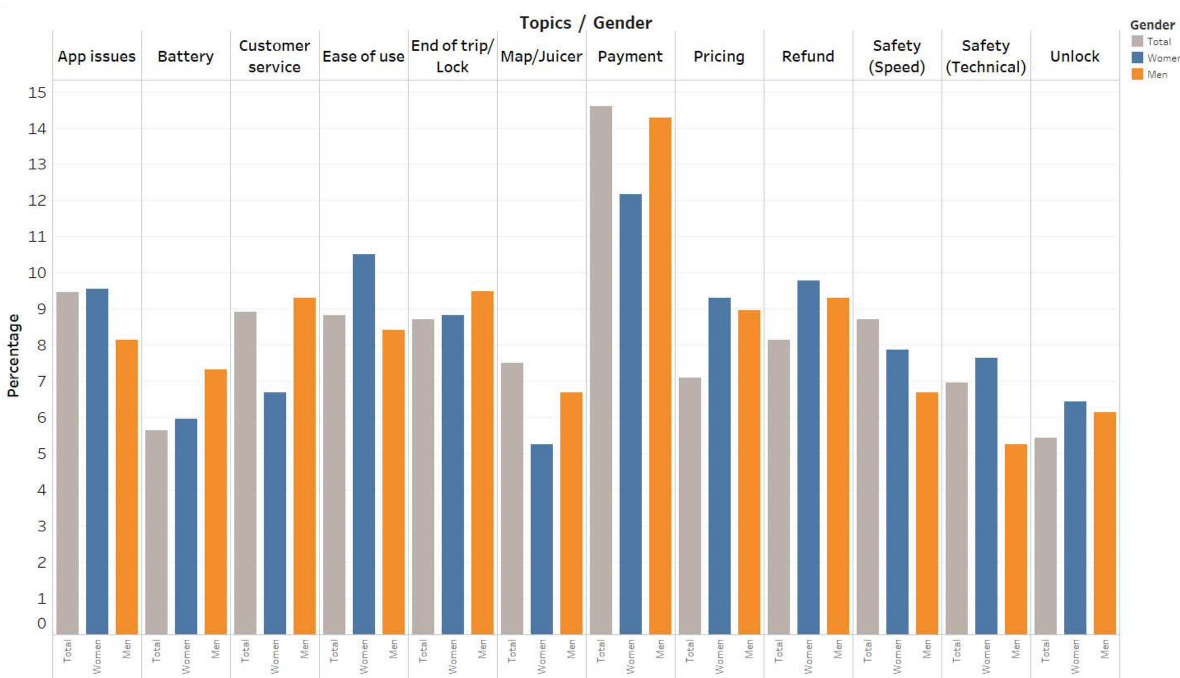


Figure 3-2 Distribution of topics in reviews

3.4.4 Polarity Analysis

Figure 3-4 shows the distribution of reviews based on polarity value for each topic. As can be seen, topics usually contain positive sentiment. However, topics, such as ease of use, app issues, safety (speed and riding lane), and safety technical issues are comparatively more positive than others, demonstrated by the height of the bar in each section of the graph. Women's reviews were relatively more negative than men regarding map and juicer, and end of trip and lock. Previous studies have reported that women left more positive comments on online platforms than men [130]. Likewise, our findings also showed that women generally exhibited more positive sentiment than male riders. These findings were in agreement with accessed online comments.

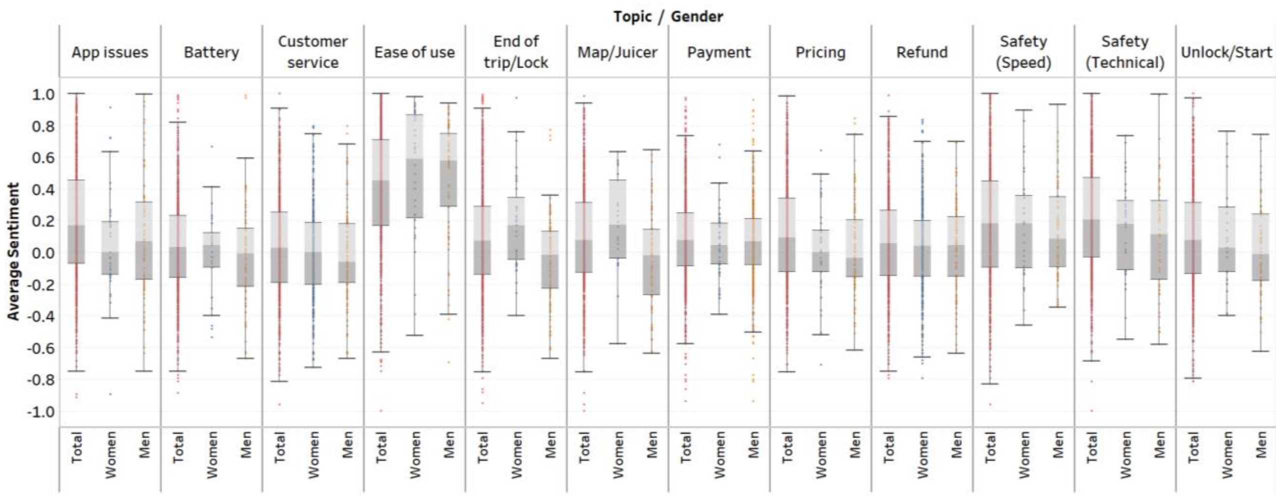


Figure 3-3 Distribution of reviews polarity value in each topic across gender

3.4.5 Logistic Regression

In this section, the influence of identified factors on user satisfaction was examined by applying the logistic regression model. Logistic regression analysis was conducted to examine potential significant predictors of rider satisfaction. Table 3-2 summarizes the estimated coefficient of factors in the logistic regression models for total reviews, and men and women' reviews. In total reviews, refund, payment, battery, and customer service had the highest coefficient factors. The results indicated that these factors significantly influenced overall rider satisfaction. For men, the most important factors were refund, ease of use, payment, and pricing. For women, the highest coefficients were refund, payment, and pricing. Moreover, there was also a significant difference between men and women in the refund coefficient. Safety technical issues were not significant in any of the three models. Safety (speed and riding lane) was not significant for the men and women's models. Map and unlock were not significant in the women's model.

Table 3-4 Coefficient of factors in the logistic regression model

Topics	Total			Men			Women		
	Coefficient	Standard	Signify	Coeff.	Std.	Signify	Coeff.	Std.	Signify
(Intercept)	1.14	0.07	***	0.51	0.14	***	1.08	0.19	***
Refund	-14.35	1.13	***	-16.22	2.86	***	-12.21	2.90	***
Payment	-14.05	0.98	***	-10.98	1.71	***	-16.84	3.38	**
Ease of use	10.04	1.40	***	11.49	1.40	***	8.51	3.56	*
Pricing	-10.62	0.99	***	-10.37	2.06	***	-15.10	3.33	***
Battery	-13.37	1.17	***	-8.94	2.26	***	-10.50	3.56	**
Customer Service	-10.39	0.97	***	-10.26	1.98	***	-9.95	3.52	**
Lock	-7.55	0.86	***	-6.24	1.60	***	-6.95	2.36	**
Map	-8.23	0.93	***	-7.24	1.95	***	-	-	-
Unlock	-7.67	1.08	***	-7.65	1.60	***	-	-	-
App issues	-4.37	0.80	***	-4.33	1.61	**	-	-	-
Safety (Speed and riding lane)	-1.55	0.79	*	-	-	-	-	-	-
Safety (Technical issues)	-	-	-	-	-	-	-	-	-
Log Likelihood	-1755.5			-426.6			-165.9		
Sample Number	3335			834			330		
Significance codes: 0 '***', 0.001 '**', 0.01 '*'									

Odds ratios in Table 4 for each variable indicate how the odds change with a one-percent increase in a topic weight while holding other variables constant. In total reviews, the refund odds ratio for satisfaction decreased by 0.866 (97.5% CI = 0.84 to 0.88) for each step-increase of the variables weight. For men, refund and payment had significantly larger odds ratio for dissatisfaction than other factors. For women, payment and pricing played the same role.

Table 3-5 Odds ratios (OR) with 97.5% confidence interval

Topics	Total		Men		Women	
	OR	97.5% CI	OR	97.5% CI	OR.	97.5% CI
Refund	0.866	(0.84, 0.88)	0.850	(0.79, 0.89)	0.885	(0.83, 0.93)

Payment	0.869	(0.85, 0.88)	0.896	(0.86, 0.92)	0.845	(0.78, 0.89)
Ease of use	1.106	(1.07, 1.13)	1.122	(1.07, 1.18)	1.089	(1.02, 1.17)
Pricing	0.899	(0.88, 0.91)	0.901	(0.86, 0.93)	0.860	(0.80, 0.91)
Battery	0.875	(0.85, 0.89)	0.914	(0.87, 0.95)	0.900	(0.83, 0.96)
Customer Service	0.901	(0.88, 0.91)	0.903	(0.86, 0.93)	0.905	(0.84, 0.96)
Lock	0.927	(0.91, 0.94)	0.940	(0.90, 0.96)	0.933	(0.88, 0.97)
Map	0.921	(0.90, 0.93)	0.930	(0.89, 0.96)		
Unlock	0.926	(0.90, 0.94)	0.926	(0.88, 0.96)		
App issues	0.957	(0.94, 0.97)	0.958	(0.92, 0.98)		
Safety (Speed and riding lane)	0.985	(0.97, 1.00)				

3.5 Discussion and Policy Implications

Rider satisfaction studies have helped to communicate user satisfaction with a transportation service or technology. These studies have helped city officials and transportation companies understand perceptions of service quality and helped to identify factors important for overall rider satisfaction. Previous studies have highlighted the importance of user-generated content, such as social media posts via Instagram, Facebook, or Twitter in understanding public opinions towards micro mobility services and related policy actions. This study demonstrated the value of app review comments for clarifying public perceptions, including individual feelings towards services and factors influencing opinions. This process of examining public opinions and perceptions towards services might thus prove useful for shaping public policy actions that would respond directly to the needs of the public and improve rider satisfaction. This study identified 12 topics of public concern expressed via App Store reviews: payment, app issues, customer service, ease of use, end of trip/lock, safety (speed and riding lane), refund, map/juicer, pricing, safety (technical issues), battery, and unlock. Payment was the most discussed topic. Each of these topics represented an opportunity to improve rider satisfaction and experience through policy interventions. These policy actions might be implemented as rules or requirements, put in place for e-scooter vendors defining actions that "must" or "shall" be taken, actions that "might be taken" and actions that "shall not be taken."

Unexplained charges and automatic payments from, for example, parking in restricted zones likely led to customer complaints about refund and payment within the reviews. Moreover, scooter companies required debit or credit cards for payment or collateral while scooters were in use. This requirement made payment systems an integral part of the e-scooter rider experience. The prerequisite of a debit or credit card for scooter operation also has formed a serious barrier for users, especially unbanked and underbanked households. This barrier raised serious equity concerns [79], as equity was mainly concerned with the fair distribution of opportunities based on the demands and characteristics of the recipients. For instance, it was reported that 57% of the low-income persons of color did not have a credit card [131]. Discontentment with existing payment systems has prompted users to ask for alternative sources of payment. These requests had implications for the policy approach to e-scooter use, because payment systems were often found to be a source of inequity in the management of new mobility technologies. Public policy actions and regulations could be used to facilitate public-private partnerships that enable more payment options. For example, options such as transit smart cards or passes could leverage the need for and use of multimodal travel options. The results in general demonstrated an urgent need to integrate e-scooter or micromobility services into current land use and transportation planning systems.

The topic modeling results suggested that "expensive," "minute," and "price" were the most common words used in the pricing topic. Scooter riding prices needed to be low enough to satisfy riders, yet high enough for private companies to maintain a sustainable business model, as such companies offer various pricing models [84]. Renting an available electric scooter often costs $1 to start/unlock and then $0.15 per each minute of riding. Espinoza et al. (2019) found that shared scooter pricing was relatively high when compared to transit pricing. They found that although typically available around transit stations, scooters were still not used often as a last-mile trip mode. Further demonstrating the impact of pricing, Fitt and Curl (2020) asked riders why they started using scooters, and only 7% of the respondents reported that "it was cheaper than the alternative." Shaheen and Cohen (2019) have proposed discounted or subsidized plans for underprivileged individuals to overcome cost-efficiency challenges. Public-private partnerships could also be used to negotiate equitable (pricing) structures including mileage based, frequency-based, as well as need-based. Discount programs and passes similar to those instituted for traditional transit modes might also be an option.

The quality of the user interface, and technical shortcomings formed a considerable share of user reviews focused on app issues and customer service. Having to download the app from App Stores, the registration process, finding the location of available scooters from the map (Map), scanning the code, unlocking the scooter (Unlock/Start), and locking the scooter afterward (End of trip/lock), all required a certain degree of knowledge and technological skill on the part of riders, as well a straightforward and user-friendly app interface. Several studies have reported that not being able to use the online interface of shared mobility technologies created a considerable barrier their use [131], [132]. For instance, according to Fitt and Curl (2020), 13% of their study sample believed that using a scooter app could be challenging at first. Other studies have reported that individuals have refrained from using shared mobility services due to a lack of digital knowledge and perceptions that the micromobilty apps are cumbersome and difficult to use. In our study, app related reviews also exhibited user difficulties. It should further be noted that these reviews did not capture the concerns of those who in general had difficulty using micromobility. Concerns about the app have public policy implications when considering requirements under the Americans with Disabilities Act (ADA) to ensure access for persons of all abilities. If state and local governments pursue micromobility as an option to expand transportation alternatives, concerns about physical and cognitive barriers to access might require policy intervention to respond to user concerns: how could scooter vendors improve app functionality to reduce barriers?

How could policies provide an incentive for micromobility behavior or deemphasize existing app limitations?

Related to app issues was customer service. More than nine percent of the reviews contained issues related to customer service. In this study, topic coexistence analysis indicated that reviews under the topic of customer service mostly co-occurred with reviews related to refund, payment, and map/juicer topics. Customer service described the assistance or advice received by riders from the e-scooter company about the product or services being used. Numerous studies have investigated the relationship between customer service and user satisfaction. Higher quality customer service was found to lead to higher user satisfaction [133]. Here again full integration might prove useful of e-scooter or micromobility services into current land use and transportation planning systems and policies that require specialized training for customer service operators. Additionally, understanding perceptions of app-related functions and creating policies that expect improvements to app-related technology might help to increase ridership of micromobility services.

Reviews show that the lack of clear safety regulations regarding e-scooter use has been confusing for riders. Reviews, specifically, demonstrated the confusion of users in choosing the proper path for riding scooters. Some reviews mentioned the risk of riding on streets near motor-vehicles, while other reviews expressed frustration about riding on sidewalks alongside pedestrians as well as incidents of aggression initiated by pedestrians. Our understanding of e-scooter rider expectations regarding infrastructure has remained mixed. Unlike bike riders, scooter riders usually prefer to ride on the sidewalk as opposed to the street [134]. In contrast, the latest e-scooter routing study suggested that e-scooter users tended to favor bicycle-friendly infrastructure over pedestrian-oriented infrastructures [135]. Although some cities have forbidden riding on sidewalks, riders frequently ride and park scooters on sidewalks, often in direct conflict with pedestrians. Additionally, regulations and policies about proper speeds, minimum age to ride (driver license), riding under the influence (RUI), and helmet requirements vary from state to state and local jurisdiction to local jurisdiction. These issues have generated numerous complaints and questions at the App Stores. Transportation departments within cites and other location jurisdictions might consider partnering with e-scooter companies to offer clear guidance on riding rules and regulations, including wrong-way riding, right-of-way, speeding, and enforcement. To enhance the micromobility options in the transportation network, it is crucial to plan modifications to the existing bicycle-and-pedestrian infrastructure. This can be achieved by incorporating specialized lanes for micromobility vehicles, allocating dedicated e-scooter zones on sidewalks, or establishing shared-use paths.

In this study, coexistence analysis indicated that app-related concerns tended to be discussed together with map/juicer and safety (speed) for men, and map/juicer, payment and battery for women. Very few researchers have addressed the differences between the male and female e-scooter riding experience. Although our analysis showed that women exhibited more positive sentiments and were slightly more satisfied than men, it was recognized that micromobility was not a gender-neutral mode of transportation [125]. Men were viewed widely as the dominant user group [136]. Gauquelin (2020) considered cultural factors as a reason behind the lower usage of scooters by women. He invited shared-mobility stakeholders to become more aware of this gender gap. Some cultural factors that may contribute to this disparity include gender norms and stereotypes surrounding acceptable modes of transportation for women. For instance, certain transportation modes like bikes or motorcycles may be associated with masculinity, making them less appealing or stigmatized as "unfeminine" choices for women [137], [138]. On the other hand, modes like cars or public transport may be seen as more "appropriate" for women. However, an interesting finding of this paper was that women’s satisfaction (unlike men) was not associated with the technical features of the apps or vehicles including lock, map/juicer, app issues, or safety-related factors. Concerns were mostly related to payment, pricing, and refund. These findings provided unique marketing and policy opportunities for cities and local jurisdictions. For instance, policies or actions that prioritize transparent pricing, flexible payment options, and clear refund policies may help to increase satisfaction with the services and potential uptake of micromobility services among women.

Recent research suggests that promoting active travel through micromobility services requires a more nuanced and gender-sensitive approach. For instance, Ng and Acker (2018) reported that there are distinct differences in travel patterns and behaviors between women and men. Women often have more complicated schedules involving multiple non-commuting short trips to various destinations during off-peak hours, when public transport services may not be available when needed. Therefore, more flexible transportation modes such as shared micromobility services and ridesharing may be better suited to their needs [139]. In order to promote micromobility among women, existing research indicates that a gendered approach to travel is needed.

A gendered approach to travel-related policy should be explored as an approach to improve rider satisfaction and experience. A gender-sensitive approach to travel-related policy would involve considering the needs and experiences of individuals of all genders when designing transportation systems and services. This would include recognizing that people's transportation needs and preferences may vary based on their gender identity. This study adds to ongoing work to advance gender neutrality in transportation policies and programs by recognizing and addressing the diverse needs and experiences of men and women in micromobility services. However, it is important to conduct further research to understand the unique needs and experiences of other marginalized groups, such as non-binary individuals.

Overall, the lack of a policy focusing on scooter rider satisfaction might lead to the loss of users. A positive association has been reported between user satisfaction and loyalty to the services [140]. Highly satisfied users were more prone to recommend the services to other individuals. On the other hand, dissatisfied users usually shared their negative experiences with other people including potential riders [141]. Additionally, the use of text mining as a tool increases the chance of gaining a broader set of opinion results, thus increasing the potential effectiveness of service improvements. Overlooking factors that influence rider satisfaction could perpetuate negative perceptions about micromobility services and consequently lead to a loss of riders. By not using text mining as a tool for understanding user satisfaction transportation agencies have been neglecting a lesser cost, potentially less biased, and more informative approach to data analysis. It could be recommended that a policy approach remain focused on improving rider satisfaction and experience. A policy approach could deliver service improvements that would incorporate opinion mining as a methodology.

3.6 Conclusion

Traditionally, studies have implemented qualitative or quantitative surveys to understand rider satisfaction. Typically, the surveys were expensive to perform and lacked consistent measurement factors. This research used a text mining methodology to explore, identify, and validate hidden factors of scooter rider satisfaction utilizing App Store reviews. The major barrier in analyzing online textual data was their substantial noise. Therefore, advanced text mining and machine learning approaches were employed in this study to clearly understand attributes that influenced rider satisfaction. In this study, we also used a name-based gender prediction algorithm to enable gender-based analysis.

This study identified twelve topics in app reviews using the Latent Dirichlet Allocation (LDA) topic model. Identified topics covered different attributes of scooter riding, including payment, app issues, customer services, and safety, to name a few. The most discussed topic was the lack of options for payment or payment method. The method of payment issues tended to coexist with app issues, customer services, and refunds, as well as unexpected costs incurred through parking in restricted zones. There were also concerns regarding safety and right of way among users. Some of the user concerns could be addressed directly by e-scooter companies (e.g., enhancing app design, diversifying payment options, enhancing the map/juicers design, and, improving the technical aspects of e-scooters). However, the results also suggested that some issues could only be mitigated through public private partnership (e.g., extending payment options to include transit passes, clear guidance regarding parking zones and the use of e-scooters in the public right of way, more public investments in bicycle-friendly infrastructures, and policy intervention and incentives). The gender-based analysis showed that e-scooters remained a male dominated travel mode. The results were consistent with the existing literature. Female riders were found to have slightly more positive reviews than male riders.

Logistic regression results suggested user satisfaction levels also tended to vary across topics and gender groups. Ease of use, safety (speed and riding lane), as well as app issues appeared to have a major influence on user satisfaction, based on the odds-ratios calculated using the logistic regression results. There was no significant difference in the odds ratio of different topics in the women and men’s logistic model results. However, several topics, such as map, unlock, and app issues were found to be significant in the model for men, while insignificant in the model for women. These results might have been due to the fact that the female model had a more limited sample when compared to the numbers of reviews that generated the male model. The study offers a valuable resource for micromobility companies, officials, and decision-makers who are grappling with the rise of shared e-scooters. It provides insights into key factors that impact rider satisfaction, including payment options, pricing, app functionality, and safety concerns. The study's insights can guide the development of policies that enhance the adoption of micromobility services. For instance, policymakers could focus on improving payment options, pricing transparency, and addressing safety concerns to increase user satisfaction and loyalty to micromobility services. To address payment options, various options such as cashless payments and pre-paid plans. For pricing, discounted rates could be given for off-peak travel or for users who frequently take shorter trips. Improving app functionality is another important aspect to consider. Making the user interface of micromobility apps more user-friendly and accessible to women can lead to increased usage. This could be achieved through conducting user testing with women to identify areas for improvement and implementing changes based on their feedback.

In addition to these measures, safety concerns can also be addressed by implementing safety measures such as providing helmets, increasing lighting in areas with high ridership, and establishing an online reporting system for incidents of harassment or assault. Micromobility companies' staff could also receive training on how to respond to such incidents. Furthermore, the use of text-mining and opinion-mining methodologies can help companies and public officials better understand user feedback and make informed decisions about service improvements.

Results of this study have raised several issues regarding interventions for shared electric scooters. In particular, polices about e-scooters have still remained unclear and have continued to evolve in many cities. The analysis identified several key areas that present opportunities for improving scooter rider satisfaction and experience through policy interventions. These include addressing issues related to app functionality, payment methods (such as the mandatory use of credit cards), high pricing, safety concerns, and ambiguity around rider behavior (such as helmet use and riding on sidewalks). By implementing policies that address these areas, officials and decision-makers can work to enhance the overall quality and safety of micromobility services, thereby promoting their wider adoption as a sustainable transportation option. These policy actions could be implemented as rules or requirements, put in place for e-scooter vendors defining actions that "must" or "shall" be taken, actions that "may be taken" and actions that "shall not be taken." Discontentment with existing payment systems has prompted users to ask for alternative sources of payment. Public policy actions and regulations could be used to facilitate public-private partnerships that enable more payment options, such as transit smart cards or passes. Discounted or subsidized plans for underprivileged individuals to overcome cost-efficiency challenges and public-private partnerships could be used to negotiate equitable pricing structures for e-scooters.

Furthermore, understanding riders perceptions of app-related functions and regulating app-related technology improvements could potentially increase the ridership of micromobility services. As they relate to micromobility infrastructure, policy-oriented steps should also be taken to plan for revisions to existing bicycle and pedestrian infrastructure. Policy could ensure incorporation of the appropriate micromobility infrastructure in the transportation network. Micromobility could be micromobility lanes, dedicated e-scooter space on the sidewalk, or shared use paths.

Finally, recent research has emphasized the importance of a nuanced approach to encourage active travel, particularly through micromobility. A gendered approach to travel-related policy should be explored to ensure the needs of all riders are met. Studies have shown a positive correlation between rider satisfaction and loyalty to micromobility services. Without a focus on improving scooter-rider satisfaction, cities risk increasing discontentment with available micromobility services amongst users (and non-users), as well as the potential for increased requests to have service bans within their jurisdictions. It is recommended that cities incorporate text mining and other opinion-mining methodologies, to prioritize rider feedback in decision-making related to the growth and sustainability of micromobility services.

The main limitation of this study was that app store reviews were used as the only source of data. Although app store reviews often discuss the scooter-rider experience and provide valuable information, the content of reviews were in general related to the user experiences with the app as opposed to the riding experience. Therefore, topic modeling algorithms could have made a biased allocation by overemphasizing scooter topics related to app performance and app user experience while overlooking the rider experience. Additionally, service delivery topics were often not discussed in the context of the app store reviews. For instance, equity or inequities in micromobility facilities is a topic of great significance that was not mentioned directly within the app store reviews. Moreover, using app store reviews (or other forms of social media information) as a research tool might have caused selection bias.

samples were drawn from individuals typically younger, affluent, or more engaged with technology. Reviews were also typically written by users who had at least tried to rent a scooter by downloading the app. Therefore, reviews were the voice of a fraction of riders and hardly the voice of those individuals who had not or could not ride scooters (e.g., individuals who did not own smartphones, those who lived where scooters were unavailable, or those with physical or cognitive disabilities that limited scooter use). Therefore, caution should be made in generalizing the results. However, it should be restated that app store data also reduced bias that presented itself in traditional methods. Furthermore, dissatisfied users were more eager to write longer reviews. Therefore, shorter, more technical reviews, in this case functions and features that delighted users, might have been overlooked. In future research, studies might use other sources of data (such as tweets and surveys) to more accurately capture user-satisfaction factors. Thus, a broader database could potentially integrate additional topics relevant to the scooter rider experience, expectations, and satisfaction. Finally, methods used for name-based gender prediction provided a source of uncertainty. Given that app-store users often used pseudo names, or that riders might have used someone else's phone to book a ride or write a review, there existed the potential for incorrectly predicting the gender of users. The area of gendered travel analysis in shared micromobilty, therefore, presented opportunities for further study and improved methodologies. For instance, using other attributes such as a profile picture, or the text itself could probably strengthen gender prediction accuracy.

CHAPTER FOUR: CONCLUSION

Achieving equity in transportation is an ongoing challenge, as transportation options still vary tremendously when it comes to people of color, low-income individuals, residents with disabilities, elderly people, women, and youth. To tackle this problem and promote more equitable transportation services and infrastructure, a comprehensive review of the existing transportation equity literature was conducted, which led to the identification of two critical gaps that must be addressed: 1) the lack of systematic data-driven approaches to studying spatial mismatch between transportation supply and demand, which impedes equity in access, and 2) limited information on women's attitudes and expectations toward emerging transportation services, which hinders gender equity.

Chapter two introduced transportation "deserts," specifically transit deserts and walking deserts. The outcome of the chapter was developing data-driven structured frameworks that identify and investigate neighborhoods with limited transportation service supply and considerable demand. The frameworks compared mobility demand and supply for active transportation modes (i.e., public transit and walking) and used statistical modeling (OLS and GWR respectively) to reveal the inequitable distribution of transportation services. The identification of transportation deserts contributes to practice by revealing areas in need of investment and redevelopment, while also exposing areas of underinvestment. Research findings further contribute to transportation planning and analysis research by developing a framework to identify areas with the lowest transportation supply and highest transportation demand.

Analysis using GWR demonstrates that the availability of walking supply within different neighborhoods was significantly related to race, educational attainment, and income. For instance, a quarter of Hispanic and Black populations, who often rely more heavily than other groups on public transportation services and other active travel modes, received zero walking access to schools or grocery stores. The disparity in access to medical facilities was even larger for these walking-dependent populations. The transportation desert framework thus allows for focused attention to be placed on the mobility demands of vulnerable social groups.

Despite the contributions made, the transportation desert assessment had limitations. As discussed, there were limitations in the formulation of the WSI. Moving forward it will be important to consider specific neighborhood contexts when defining WSI factors. Limitations of the WSI include the equal weighting of indicators, and the use of thresholds to define deserts. To overcome these limitations, future studies can explore alternative methods for weighting supply indicators such as using the analytical hierarchy process (AHP) and related methodologies to incorporate expert judgment. Additionally, to provide a more nuanced assessment of neighborhood-level walking supply researchers can incorporate additional indicators or measures such as street connectivity, safety, land use mix, and aesthetics. Continued work to assess disparities between transportation supply and demand ultimately helps to promote more equitable, livable, and healthy communities.

Chapter three addressed gender equity and specifically a lack of gendered understanding about transportation user preferences. Findings revealed information about user experiences and expectations, particularly for women, by applying text-mining methods in a gender-sensitive analysis of online reviews and comments. The chapter presented a first-of-its-kind empirical analysis using online data to examine rider satisfaction with scooter services while uncovering factors impacting overall satisfaction. App store reviews from two major micromobility companies, Lime and Bird, were also investigated using machine learning techniques.

The analysis identified factors that influenced rider satisfaction across genders. Findings provided a deeper understanding of how micromobility rider sentiment and satisfaction vary across gender.

For example, the study found that e-scooters continue to be a male-dominated mode of travel. However, female riders tended to have slightly more positive reviews than male riders. Logistic regression results showed that user satisfaction levels varied across topics and gender groups, with ease of use, payment, pricing, safety, and app issues having a major influence on user satisfaction. Payment was found to be the most discussed topic, with unexplained charges and automatic payments leading to customer complaints about refunds and payments. Pricing was also a common concern, specifically among women. In general, scooter riding prices need to be low enough to satisfy riders, yet high enough for companies to maintain a sustainable business model. The quality of the user interface and technical shortcomings were also a significant share of male user reviews which focused on app issues and customer service. Study results highlighted the need for regulatory action supporting for example public-private partnerships that expand payment options, deliver equitable pricing structures, and improve app functionality.

It is important to acknowledge the limitations of the micromobility rider satisfaction study. First, app store reviews may not fully capture the rider experience, as there is potential for selection bias given the demographics of the individuals who write reviews. Additionally, the method used for name-based gender prediction introduces some uncertainty. Future research should explore other sources of data, such as tweets and surveys, to capture a more diverse range of user experiences and opinions. Attributes like profile pictures may also be incorporated to improve gender prediction accuracy.

Despite analytical limitations gendered travel analysis in shared micromobility and transportation desert analysis of active transportation modes both present opportunities for advancing methodologies and practice around transportation equity. This dissertation thus makes a meaningful contribution to ongoing and future work in transportation equity.

In conclusion, this dissertation provides a comprehensive examination of transportation equity from multiple perspectives. The research identified critical gaps in existing literature and addressed these gaps through a diverse set of innovative analytical methodologies. Research findings have important policy implications for city planners, transportation managers, urban authorities, and decision-makers charged with creating more inclusive and vibrant urban spaces that benefit all members of society. Addressing these gaps is crucial for promoting equitable transportation services, and delivering infrastructure that ensures all individuals have access to safe, reliable, and affordable transportation options.

REFERENCE

[1] R. H. M. Pereira, T. Schwanen, and D. Banister, “Distributive justice and equity in transportation,” *Transp. Rev.*, vol. 37, no. 2, pp. 170–191, Mar. 2017, doi: 10.1080/01441647.2016.1257660.

[2] T. Litman, “Evaluating transportation equity,” *World Transp. Policy Pract.*, vol. 8, pp. 50–65, Jan. 2002.

[3] A. Karner, J. London, D. Rowangould, and K. Manaugh, “From Transportation Equity to Transportation Justice: Within, Through, and Beyond the State,” *J. Plan. Lit.*, vol. 35, no. 4, pp. 440–459, 2020, doi: 10.1177/0885412220927691.

[4] R. H. Pereira and A. Karner, “Transportation equity,” *Int. Encycl. Transp.*, vol. 1, pp. 271–277, 2021.

[5] P. R. Carleton and J. D. Porter, “A comparative analysis of the challenges in measuring transit equity: definitions, interpretations, and limitations,” *J. Transp. Geogr.*, vol. 72, pp. 64–75, 2018, doi: //doi.org/10.1016/j.jtrangeo.2018.08.012.

[6] K. Manaugh and A. El-Geneidy, “Who Benefits from New Transportation Infrastructure? Using Accessibility Measures to Evaluate Social Equity in Transit Provision,” *Access. Anal. Transp. Plan. Chall. Eur. N. Am.*, 2012, doi: 10.4337/9781781000106.00021.

[7] B. Wee and K. Geurs, “Discussing Equity and Social Exclusion in Accessibility Evaluations,” *Eur. J. Transp. Infrastruct. Res.*, vol. 11, no. 4, 2011, doi: 10.18757/ejtir.2011.11.4.2940.

[8] T. F. Welch and S. Mishra, “A measure of equity for public transit connectivity,” *J. Transp. Geogr.*, vol. 33, pp. 29–41, 2013, doi: //doi.org/10.1016/j.jtrangeo.2013.09.007.

[9] A. Delbosc and G. Currie, “Using Lorenz curves to assess public transport equity,” *J. Transp. Geogr.*, vol. 19, no. 6, pp. 1252–1259, 2011, doi: //doi.org/10.1016/j.jtrangeo.2011.02.008.

[10] A. El-Geneidy, D. Levinson, E. Diab, G. Boisjoly, D. Verbich, and C. Loong, “The cost of equity: Assessing transit accessibility and social disparity using total travel cost,” *Transp. Res. Part Policy Pract.*, vol. 91, pp. 302–316, 2016, doi: //doi.org/10.1016/j.tra.2016.07.003.

[11] Y. Shiftan, “Transport equity analysis AU - Di Ciommo, Floridea,” *Transp. Rev.*, vol. 37, no. 2, pp. 139–151, 2017, doi: 10.1080/01441647.2017.1278647.

[12] A. Brown, “From aspiration to operation: ensuring equity in transportation,” *Transp. Rev.*, vol. 42, no. 4, pp. 409–414, Jul. 2022, doi: 10.1080/01441647.2022.2064527.

[13] S. Mookerjee, “How data can help reduce transportation inequity,” *GCN*, 2021. https://gcn.com/data-analytics/2021/08/how-data-can-help-reduce-transportation-inequity/316197/ (accessed Aug. 08, 2022).

[14] L. Caggiani, R. Camporeale, and M. Ottomanelli, “Facing equity in transportation Network Design Problem: A flexible constraints based model,” *Transp. Policy*, vol. 55, pp. 9–17, 2017, doi: //doi.org/10.1016/j.tranpol.2017.01.003.

[15] N. Foth, K. Manaugh, and A. M. El-Geneidy, “Towards equitable transit: examining transit accessibility and social need in Toronto, Canada, 1996–2006,” *J. Transp. Geogr.*, vol. 29, pp. 1–10, 2013, doi: //doi.org/10.1016/j.jtrangeo.2012.12.008.

[16] A. Karner and D. Niemeier, “Civil rights guidance and equity analysis methods for regional transportation plans: a critical review of literature and practice,” *J. Transp. Geogr.*, vol. 33, pp. 126–134, Dec. 2013, doi: 10.1016/j.jtrangeo.2013.09.017.

[17] Z. Chen, K. Long, A. Stuart, F. Mannering, and X. Li, *The Promise of Big Data for Transportation Equity*. 2022.

[18] Y. Guo, Z. Chen, A. Stuart, X. Li, and Y. Zhang, “A systematic overview of transportation equity in terms of accessibility, traffic emissions, and safety outcomes: From conventional to emerging

technologies,” *Transp. Res. Interdiscip. Perspect.*, vol. 4, p. 100091, Mar. 2020, doi: 10.1016/j.trip.2020.100091.

[19] A. Lubitow, R. Liévanos, J. McGee, and E. Carpenter, “Developing Data, Models, and Tools to Enhance Transportation Equity,” *TREC Final Rep.*, Sep. 2019, doi: 10.15760/trec.239.

[20] S. Sohrabi and H. Khreis, “Transportation and Public Health: A Burden of Disease Analysis of Transportation Noise,” *J. Transp. Health*, vol. 14, p. 100686, 2019.

[21] N. Chen and C.-H. Wang, “Does green transportation promote accessibility for equity in medium-size U.S. cites?,” *Transp. Res. Part Transp. Environ.*, vol. 84, p. 102365, Jul. 2020, doi: 10.1016/j.trd.2020.102365.

[22] J. J. C. Aman and J. Smith-Colin, “Transit Deserts: Equity analysis of public transit accessibility,” *J. Transp. Geogr.*, vol. 89, p. 102869, 2020, doi: https://doi.org/10.1016/j.jtrangeo.2020.102869.

[23] S. Heidari, T. F. Babor, P. De Castro, S. Tort, and M. Curno, “Sex and Gender Equity in Research: rationale for the SAGER guidelines and recommended use,” *Res. Integr. Peer Rev.*, vol. 1, no. 1, p. 2, May 2016, doi: 10.1186/s41073-016-0007-6.

[24] I. L. Office, I. L. Organization, and O. internationale du travail, *ABC of Women Workers’ Rights and Gender Equality*. International Labour Organization, 2000.

[25] L. Thompson, “Gender equity and corporate social responsibility in a post-feminist era,” *Bus. Ethics Eur. Rev.*, vol. 17, pp. 87–106, Dec. 2007, doi: 10.1111/j.1467-8608.2008.00523.x.

[26] UNECE, “Gender and transport,” 2009. https://unece.org/gender-and-transport (accessed Aug. 11, 2022).

[27] P. Zhao and Y. Cao, “Commuting inequity and its determinants in Shanghai: New findings from big-data analytics,” *Transp. Policy*, vol. 92, pp. 20–37, Jun. 2020, doi: 10.1016/j.tranpol.2020.03.006.

[28] M. Cai, "Natural language processing for urban research: A systematic review," *Heliyon*, vol. 7, no. 3, p. e06322, Mar. 2021, doi: 10.1016/j.heliyon.2021.e06322.

[29] J. J. C. Aman, J. Smith-Colin, and W. Zhang, "Listen to E-scooter riders: Mining rider satisfaction factors from app store reviews," *Transp. Res. Part Transp. Environ.*, vol. 95, p. 102856, Jun. 2021, doi: 10.1016/j.trd.2021.102856.

[30] C. C. Perez, *Invisible Women: Data Bias in a World Designed for Men*. Abrams, 2019.

[31] J. P. Lima and M. H. Machado, "Walking accessibility for individuals with reduced mobility: A Brazilian case study," *Case Stud. Transp. Policy*, vol. 7, no. 2, pp. 269–279, Jun. 2019, doi: 10.1016/j.cstp.2019.02.007.

[32] S. Abley and S. Turner, *Predicting walkability*, vol. 452. NZ Transport Agency Wellington, 2011.

[33] J. Arellana, V. Alvarez, D. Oviedo, and L. A. Guzman, "Walk this way: Pedestrian accessibility and equity in Barranquilla and Soledad, Colombia," *Res. Transp. Econ.*, vol. 86, p. 101024, May 2021, doi: 10.1016/j.retrec.2020.101024.

[34] Wilson, "Why Every City Needs to Learn the Three A's of Equitable Pedestrian Planning," *Streetsblog USA*, Aug. 21, 2020. https://usa.streetsblog.org/2020/08/21/why-every-city-needs-to-learn-the-three-as-of-equitable-pedestrian-planning/ (accessed Jul. 28, 2022).

[35] R. J. Lee, I. N. Sener, and S. N. Jones, "Understanding the role of equity in active transportation planning in the United States," *Transp. Rev.*, vol. 37, no. 2, pp. 211–226, Mar. 2017, doi: 10.1080/01441647.2016.1239660.

[36] M. Weng *et al.*, "The 15-minute walkable neighborhoods: Measurement, social inequalities and implications for building healthy communities in urban China," *J. Transp. Health*, vol. 13, pp. 259–273, Jun. 2019, doi: 10.1016/j.jth.2019.05.005.

[37] F. Moura, P. Cambra, and A. B. Gonçalves, “Measuring walkability for distinct pedestrian groups with a participatory assessment method: A case study in Lisbon,” *Landsc. Urban Plan.*, vol. 157, pp. 282–296, Jan. 2017, doi: 10.1016/j.landurbplan.2016.07.002.

[38] J. Knight, R. Weaver, and P. Jones, “Walkable and resurgent for whom? The uneven geographies of walkability in Buffalo, NY,” *Appl. Geogr.*, vol. 92, pp. 1–11, Mar. 2018, doi: 10.1016/j.apgeog.2018.01.008.

[39] W. Riggs, “Inclusively walkable: exploring the equity of walkable housing in the San Francisco Bay Area,” *Local Environ.*, vol. 21, no. 5, pp. 527–554, May 2016, doi: 10.1080/13549839.2014.982080.

[40] City of Austin, “Austin Strategic Mobility Plan,” City of Austin, Austin, TX, 2022.

[41] U.S. Census Bureau, “American Community Survey (ACS),” *Census.gov*. https://www.census.gov/programs-surveys/acs (accessed Jul. 25, 2022).

[42] City of Austin, “Austin’s Open Data Portal,” *Austin’s Open Data Portal*, 2022. https://data.austintexas.gov/ (accessed Jul. 25, 2022).

[43] SafeGraph, “SafeGraph,” *Home*, 2022. https://www.safegraph.com (accessed Jul. 25, 2022).

[44] C. P. D. Birch, S. P. Oom, and J. A. Beecham, “Rectangular and hexagonal grids used for observation, experiment and simulation in ecology,” *Ecol. Model.*, vol. 206, no. 3–4, pp. 347–359, Aug. 2007, doi: 10.1016/j.ecolmodel.2007.03.041.

[45] R. H. M. Pereira, “Transport legacy of mega-events and the redistribution of accessibility to urban destinations,” *Cities*, vol. 81, pp. 45–60, Nov. 2018, doi: 10.1016/j.cities.2018.03.013.

[46] J. R. Mayaud, M. Tran, and R. Nuttall, “An urban data framework for assessing equity in cities: Comparing accessibility to healthcare facilities in Cascadia,” *Comput. Environ. Urban Syst.*, vol. 78, p. 101401, Nov. 2019, doi: 10.1016/j.compenvurbsys.2019.101401.

[47] J. R. Mayaud, M. Tran, R. H. M. Pereira, and R. Nuttall, "Future access to essential services in a growing smart city: The case of Surrey, British Columbia," *Comput. Environ. Urban Syst.*, vol. 73, pp. 1–15, Jan. 2019, doi: 10.1016/j.compenvurbsys.2018.07.005.

[48] D. S. Vale and M. Pereira, "The influence of the impedance function on gravity-based pedestrian accessibility measures: A comparative analysis," *Environ. Plan. B Urban Anal. City Sci.*, vol. 44, no. 4, pp. 740–763, Jul. 2017, doi: 10.1177/0265813516641685.

[49] M.-P. Kwan, "Space-Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point-based Framework," *Geogr. Anal.*, vol. 30, no. 3, pp. 191–216, 1998, doi: 10.1111/j.1538-4632.1998.tb00396.x.

[50] M. Nardo, M. Saisana, A. Saltelli, and S. Tarantola, "Tools for Composite Indicators Building," *JRC Publications Repository*, Dec. 07, 2005. https://publications.jrc.ec.europa.eu/repository/handle/JRC31473 (accessed May 01, 2023).

[51] O. US EPA, "National Walkability Index User Guide and Methodology," May 17, 2021. https://www.epa.gov/smartgrowth/national-walkability-index-user-guide-and-methodology (accessed May 01, 2023).

[52] B. Zhu, C.-W. Hsieh, and Y. Zhang, "Incorporating spatial statistics into examining equity in health workforce distribution: an empirical analysis in the Chinese context," *Int. J. Environ. Res. Public. Health*, vol. 15, no. 7, p. 1309, 2018.

[53] D. C. Howell, *Statistical Methods for Psychology*, 8th edition. Belmont, CA: Cengage Learning, 2012.

[54] X. Ma, J. Zhang, C. Ding, and Y. Wang, "A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership," *Comput. Environ. Urban Syst.*, vol. 70, pp. 113–124, Jul. 2018, doi: 10.1016/j.compenvurbsys.2018.03.001.

[55] S. Özkazanç and F. N. Özdemir Sönmez, “Spatial analysis of social exclusion from a transportation perspective: A case study of Ankara metropolitan area,” *Cities*, vol. 67, pp. 74–84, Jul. 2017, doi: 10.1016/j.cities.2017.04.013.

[56] W. A. V. Clark and P. L. Hosking, *Statistical methods for geographers*. New York: Wiley, 1991.

[57] D. Yu, C. M. Morton, and N. A. Peterson, “Community pharmacies and addictive products: sociodemographic predictors of accessibility from a mixed GWR perspective,” *GIScience Remote Sens.*, vol. 51, no. 1, pp. 99–113, Jan. 2014, doi: 10.1080/15481603.2014.886457.

[58] S. Khanyile and C. C. Fatti, “Interrogating park access and equity in Johannesburg, South Africa,” *Environ. Urban.*, vol. 34, no. 1, pp. 10–31, Apr. 2022, doi: 10.1177/09562478221083891.

[59] Johanna Andersson, “Using Geographically Weighted Regression (GWR) to explore spatial variations in the relationship between public transport accessibility and car use : a case study in Lund and Malmö, Sweden | Lund University,” Lund Uiversity, 2017. Accessed: May 02, 2023. [Online]. Available: https://www.lunduniversity.lu.se/lup/publication/8919808

[60] J. L. Mennis and L. Jordan, “The Distribution of Environmental Equity: Exploring Spatial Nonstationarity in Multivariate Models of Air Toxic Releases,” *Ann. Assoc. Am. Geogr.*, vol. 95, no. 2, pp. 249–268, Jun. 2005, doi: 10.1111/j.1467-8306.2005.00459.x.

[61] J. Kim and S. Nicholls, “Using Geographically Weighted Regression to Explore the Equity of Public Open Space Distributions,” *J. Leis. Res.*, vol. 48, no. 2, pp. 105–133, Apr. 2016, doi: 10.18666/jlr-2016-v48-i2-6539.

[62] A. Maroko, J. A. Maantay, and K. Grady, “Using Geovisualization and Geospatial Analysis to Explore Respiratory Disease and Environmental Health Justice in New York City,” *Geospatial Anal. Environ. Health*, pp. 39–66, 2011.

[63] E. Talen and S. Shah, “Neighborhood Evaluation Using GIS: An Exploratory Study,” *Environ. Behav.*, vol. 39, no. 5, pp. 583–615, Sep. 2007, doi: 10.1177/0013916506292332.

[64] S. N. Adli, S. Chowdhury, and Y. Shiftan, “Justice in public transport systems: A comparative study of Auckland, Brisbane, Perth and Vancouver,” *Cities*, vol. 90, pp. 88–99, 2019, doi: //doi.org/10.1016/j.cities.2019.01.031.

[65] Y. Chen, A. Bouferguene, Y. Shen, and M. Al-Hussein, “Assessing accessibility-based service effectiveness (ABSEV) and social equity for urban bus transit: A sustainability perspective,” *Sustain. Cities Soc.*, vol. 44, pp. 499–510, 2019, doi: //doi.org/10.1016/j.scs.2018.10.003.

[66] G. Currie, “Quantifying spatial gaps in public transport supply based on social needs,” *J. Transp. Geogr.*, vol. 18, no. 1, pp. 31–41, 2010, doi: //doi.org/10.1016/j.jtrangeo.2008.12.002.

[67] A. M. Ricciardi, J. Xia, and G. Currie, “Exploring public transport equity between separate disadvantaged cohorts: a case study in Perth, Australia,” *J. Transp. Geogr.*, vol. 43, pp. 111–122, 2015, doi: //doi.org/10.1016/j.jtrangeo.2015.01.011.

[68] J. Jiao, “Identifying transit deserts in major Texas cities where the supplies missed the demands,” *J. Transp. Land Use*, vol. 10, no. 1, Jan. 2017, doi: 10.5198/jtlu.2017.899.

[69] Y. Zhang, P. Chen, and Y. Guo, “Identifying Multi-Modal Deserts: A Multivariate Outlier Detection Approach,” Feb. 2021, Accessed: Feb. 20, 2023. [Online]. Available: https://trid.trb.org/view/1922720

[70] Antoniassi, “Gender Data Series: Mobility, Accessibility and the Gender Data Gap in Urban Transportation Planning,” *Data-Pop Alliance*, Mar. 04, 2022. https://datapopalliance.org/gender-data-series-mobility-accessibility-and-the-gender-data-gap-in-urban-transportation-planning/ (accessed Oct. 13, 2022).

[71] S. Gössling, “Integrating e-scooters in urban transportation: Problems, policies, and the prospect of system change,” *Transp. Res. Part Transp. Environ.*, vol. 79, p. 102230, 2020, doi: https://doi.org/10.1016/j.trd.2020.102230.

[72] H. Younes, Z. Zou, J. Wu, and G. Baiocchi, “Comparing the Temporal Determinants of Dockless Scooter-share and Station-based Bike-share in Washington, D.C.,” *Transp. Res. Part Policy Pract.*, vol. 134, pp. 308–320, 2020, doi: https://doi.org/10.1016/j.tra.2020.02.021.

[73] J.-P. Allem and A. Majmundar, “Are electric scooters promoted on social media with safety in mind? A case study on Bird’s Instagram,” *Prev. Med. Rep.*, vol. 13, pp. 62–63, 2019.

[74] A. Dormanesh, A. Majmundar, and J.-P. Allem, “Follow-Up Investigation on the Promotional Practices of Electric Scooter Companies: Content Analysis of Posts on Instagram and Twitter,” *JMIR Public Health Surveill.*, vol. 6, no. 1, p. e16833, 2020, doi: 10.2196/16833.

[75] L. J. Mayhew and C. Bergin, “Impact of e-scooter injuries on Emergency Department imaging,” *J. Med. Imaging Radiat. Oncol.*, vol. 63, no. 4, pp. 461–466, 2019, doi: 10.1111/1754-9485.12889.

[76] M. Nellamattathil and I. Amber, “An evaluation of scooter injury and injury patterns following widespread adoption of E-scooters in a major metropolitan area,” *Clinical Imaging*, vol. 60, no. 2. pp. 200–203, 2020. doi: https://doi.org/10.1016/j.clinimag.2019.12.012.

[77] H. Yang, Q. Ma, Z. Wang, Q. Cai, K. Xie, and D. Yang, “Safety of micro-mobility: Analysis of E-Scooter crashes by mining news reports,” *Accid. Anal. Prev.*, vol. 143, p. 105608, 2020, doi: https://doi.org/10.1016/j.aap.2020.105608.

[78] K. Anderson-Hall, B. Bordenkircher, R. O’Neil, and S. C. Scott, “Governing micro-mobility: A nationwide assessment of electric scooter regulations,” in *Transportation Research Board 98th Annual Meeting*, 2019.

[79] S. Shaheen and A. Cohen, “Shared Micromoblity Policy Toolkit: Docked and Dockless Bike and Scooter Sharing,” 2019.

[80] S. Tuncer and B. Brown, “E-scooters on the Ground: Lessons for Redesigning Urban Micro-Mobility,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–14.

[81] A. Brown, N. J. Klein, C. Thigpen, and N. Williams, “Impeding access: The frequency and characteristics of improper scooter, bike, and car parking,” *Transportation Research Interdisciplinary Perspectives*. p. 100099, 2020. doi: https://doi.org/10.1016/j.trip.2020.100099.

[82] K. Fang, “Do They Block the Way in San Jose? Where Do Riders Park Dockless, Shared Electric Scooters and Their Implications for Sidewalk Users,” *J. Transp. Health*, vol. 14, p. 100787, Sep. 2019, doi: 10.1016/j.jth.2019.100787.

[83] Y.-W. Chen, C.-Y. Cheng, S.-F. Li, and C.-H. Yu, “Location optimization for multiple types of charging stations for electric scooters,” *Appl. Soft Comput.*, vol. 67, pp. 519–528, 2018, doi: https://doi.org/10.1016/j.asoc.2018.02.038.

[84] W. Espinoza, M. Howard, J. Lane, and P. V. Hentenryck, “Shared E-scooters: Business, Pleasure, or Transit?,” *ArXiv Prepr. ArXiv191005807*, 2019.

[85] G. McKenzie, “Spatiotemporal comparative analysis of scooter-share and bike-share usage patterns in Washington, D.C.,” *Journal of Transport Geography*, vol. 78. pp. 19–28, 2019. doi: https://doi.org/10.1016/j.jtrangeo.2019.05.007.

[86] R. Zhu, X. Zhang, D. Kondor, P. Santi, and C. Ratti, “Understanding spatio-temporal heterogeneity of bike-sharing and scooter-sharing mobility,” *Comput. Environ. Urban Syst.*, vol. 81, p. 101483, 2020.

[87] J. Degele *et al.*, “Identifying E-Scooter Sharing Customer Segments Using Clustering,” IEEE, 2018, pp. 1–8. doi: 10.1109/ICE.2018.8436288.

[88] H. Fitt and A. Curl, “The early days of shared micromobility: A social practices approach,” *J. Transp. Geogr.*, vol. 86, p. 102779, 2020, doi: https://doi.org/10.1016/j.jtrangeo.2020.102779.

[89] L. Eboli and G. Mazzulla, “Service quality attributes affecting customer satisfaction for bus transit,” *J. Public Transp.*, vol. 10, no. 3, p. 2, 2007.

[90] Y. Guo, J. Zhou, Y. Wu, and Z. Li, “Identifying the factors affecting bike-sharing usage and degree of satisfaction in Ningbo, China,” *PloS One*, vol. 12, no. 9, p. e0185100, 2017.

[91] H. Han, B. Meng, and W. Kim, “Bike-traveling as a growing phenomenon: Role of attributes, value, satisfaction, desire, and gender in developing loyalty,” *Tour. Manag.*, vol. 59, pp. 91–103, 2017, doi: https://doi.org/10.1016/j.tourman.2016.07.013.

[92] G. Manzi and G. Saibene, “Are they telling the truth? Revealing hidden traits of satisfaction with a public bike-sharing service,” *null*, vol. 12, no. 4, pp. 253–270, 2018, doi: 10.1080/15568318.2017.1353186.

[93] B. Fu, J. Lin, L. Li, C. Faloutsos, J. Hong, and N. Sadeh, “Why people hate your app: Making sense of user feedback in a mobile app store,” in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2013, pp. 1276–1284.

[94] J. J. C. Aman and J. Smith-Colin, “Application of Crowdsourced Data to Infer User Satisfaction with Mobility as a Service (MaaS),” *Transp. Res. Interdiscip. Perspect.*, 2021.

[95] J. J. C. Aman and J. Smith-Colin, “Leveraging Social Media to Understand Public Perceptions of Micromobility Policies: The Dallas Scooter Ban Case,” *Transp. Find.*, 2021.

[96] S. Das, J. J. C. Aman, and M. A. Rahman, “Content Analysis on Homelessness Issues at Airports by News Media Mining,” *Transp. Res. Rec. J. Transp. Res. Board*, p. 036119812211081, Jul. 2022, doi: 10.1177/03611981221108151.

[97] N. R. Kim and S. G. Hong, “Text mining for the evaluation of public services: the case of a public bike-sharing system,” *Serv. Bus.*, vol. 14, no. 3, pp. 315–331, 2020, doi: 10.1007/s11628-020-00419-4.

[98] H. Zhao and F. Kamareddine, “Advance Gender Prediction Tool of First Names and its Use in Analysing Gender Disparity in Computer Science in the UK, Malaysia and China,” in *2017*

International Conference on Computational Science and Computational Intelligence (CSCI), Dec. 2017, pp. 222–227. doi: 10.1109/CSCI.2017.35.

[99] J. H. Suh, “Machine-Learning-Based Gender Distribution Prediction from Anonymous News Comments: The Case of Korean News Portal,” *Sustainability*, vol. 14, no. 16, Art. no. 16, Jan. 2022, doi: 10.3390/su14169939.

[100] D. Kumar and S. V. Ukkusuri, “Enhancing demographic coverage of hurricane evacuation behavior modeling using social media,” *J. Comput. Sci.*, vol. 45, p. 101184, Sep. 2020, doi: 10.1016/j.jocs.2020.101184.

[101] N. Cheng, R. Chandramouli, and K. P. Subbalakshmi, “Author gender identification from text,” *Digit. Investig.*, vol. 8, no. 1, pp. 78–88, Jul. 2011, doi: 10.1016/j.diin.2011.04.002.

[102] S. Das and J. H. Paik, “Context-sensitive gender inference of named entities in text,” *Inf. Process. Manag.*, vol. 58, no. 1, p. 102423, Jan. 2021, doi: 10.1016/j.ipm.2020.102423.

[103] P. A. Longley, M. Adnan, and G. Lansley, “The Geotemporal Demographics of Twitter Usage,” *Environ. Plan. Econ. Space*, vol. 47, no. 2, pp. 465–484, Feb. 2015, doi: 10.1068/a130122p.

[104] D. Bamman, J. Eisenstein, and T. Schnoebelen, “Gender identity and lexical variation in social media,” *J. Socioling.*, vol. 18, no. 2, pp. 135–160, 2014, doi: 10.1111/josl.12080.

[105] A. Mislove, S. Lehmann, Y.-Y. Ahn, J.-P. Onnela, and J. Rosenquist, “Understanding the Demographics of Twitter Users,” *Proc. Int. AAAI Conf. Web Soc. Media*, vol. 5, no. 1, Art. no. 1, 2011.

[106] B. Mandel, A. Culotta, J. Boulahanis, D. Stark, B. Lewis, and J. Rodrigue, “A Demographic Analysis of Online Sentiment during Hurricane Irene,” in *Proceedings of the Second Workshop on Language in Social Media*, Montréal, Canada: Association for Computational Linguistics, Jun. 2012, pp. 27–36. Accessed: Oct. 13, 2022. [Online]. Available: https://aclanthology.org/W12-2104

[107] C. Blevins and L. Mullen, “Jane, John... Leslie? A Historical Method for Algorithmic Gender Prediction.,” *DHQ Digit. Humanit. Q.*, vol. 9, no. 3, 2015.

[108] K. Wais, “Gender Prediction Methods Based on First Names with genderizeR.,” *R J*, vol. 8, no. 1, p. 17, 2016.

[109] L. Mullen, “gender: Predict Gender from Names Using Historical Data,” vol. R package version 0.5.2. 2018.

[110] V. Srividhya and R. Anitha, “Evaluating preprocessing techniques in text categorization,” *Int. J. Comput. Sci. Appl.*, vol. 47, no. 11, pp. 49–51, 2010.

[111] C. Silva and B. Ribeiro, “The importance of stop word removal on recall values in text categorization,” in *Proceedings of the International Joint Conference on Neural Networks, 2003.*, IEEE, 2003, pp. 1661–1666.

[112] J. Ooms, “hunspell: High-Performance Stemmer, Tokenizer, and Spell Checker for R. R package version 2.6.” 2017. [Online]. Available: https://CRAN.R-project.org/package=hunspell

[113] E. Guzman and W. Maalej, “How Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews,” 2014, pp. 153–162. doi: 10.1109/RE.2014.6912257.

[114] F. Benamara, C. Cesarano, A. Picariello, D. R. Recupero, and V. S. Subrahmanian, “Sentiment analysis: Adjectives and adverbs are better than adjectives alone.,” *ICWSM*, vol. 7. pp. 203–206, 2007.

[115] F. R. Lucini, L. M. Tonetto, F. S. Fogliatto, and M. J. Anzanello, “Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews,” *Journal of Air Transport Management*, vol. 83. p. 101760, 2020.

[116] V. Gupta and G. S. Lehal, “A survey of text mining techniques and applications,” *J. Emerg. Technol. Web Intell.*, vol. 1, no. 1, pp. 60–76, 2009.

[117] K. Ghag and K. Shah, "SentiTFIDF-Sentiment classification using relative term frequency inverse document frequency," *Int. J. Adv. Comput. Sci. Appl.*, vol. 5, no. 2, 2014.

[118] H. Jelodar *et al.*, "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," *Multimed. Tools Appl.*, vol. 78, no. 11, pp. 15169–15211, 2019, doi: 10.1007/s11042-018-6894-4.

[119] R. Krestel, P. Fankhauser, and W. Nejdl, "Latent dirichlet allocation for tag recommendation," in *Proceedings of the third ACM conference on Recommender systems*, 2009, pp. 61–68.

[120] S. Debortoli, O. Müller, I. Junglas, and J. vom Brocke, "Text mining for information systems researchers: An annotated topic modeling tutorial," *Commun. Assoc. Inf. Syst.*, vol. 39, no. 1, p. 7, 2016.

[121] V. Jha, S. R, P. D. Shenoy, V. K. R, and A. K. Sangaiah, "A novel sentiment aware dictionary for multi-domain sentiment classification," *Comput. Electr. Eng.*, vol. 69, pp. 585–597, 2018, doi: https://doi.org/10.1016/j.compeleceng.2017.10.015.

[122] S. Park and Y. Kim, "Building thesaurus lexicon using dictionary-based approach for sentiment classification," in *2016 IEEE 14th International Conference on Software Engineering Research, Management and Applications (SERA)*, IEEE, 2016, pp. 39–44.

[123] T. W. Rinker, "sentimentr: Calculate Text Polarity Sentiment," vol. 2.7.1. 2019. [Online]. Available: http://github.com/trinker/sentimentr

[124] X. Zhang, D. R. Jeske, J. Li, and V. Wong, "A sequential logistic regression classifier based on mixed effects with applications to longitudinal data," *Comput. Stat. Data Anal.*, vol. 94, pp. 238–249, 2016, doi: https://doi.org/10.1016/j.csda.2015.08.009.

[125] A. Gauquelin, "The Gender Gap in Shared Micromobility," vol. 2020, no. June 15,. 2020. [Online]. Available: https://urbanmobilitydaily.com/

[126] N. McGuckin and E. Murakami, “Examining trip-chaining behavior: Comparison of travel by men and women,” *Transp. Res. Rec.*, vol. 1693, no. 1, pp. 79–85, 1999.

[127] K. Bastani, H. Namavari, and J. Shaffer, “Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints,” *Expert Syst. Appl.*, vol. 127, pp. 256–271, 2019.

[128] Y. Guo, S. J. Barnes, and Q. Jia, “Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation,” *Tourism Management*, vol. 59. pp. 467–483, 2017. doi: https://doi.org/10.1016/j.tourman.2016.09.009.

[129] J. Silge and D. Robinson, *Text Mining with R*, 1st ed. Sebastopol: O’Reilly, 2017. [Online]. Available: https://ebookcentral.proquest.com/lib/[SITE_ID]/detail.action?docID=4876883

[130] A. Pettit, “Identifying the Real Differences of Opinion in Social Media Sentiment,” *Int. J. Mark. Res.*, vol. 55, no. 6, pp. 757–767, 2013, doi: 10.2501/IJMR-2013-065.

[131] N. McNeil, J. Broach, and J. Dill, “Breaking barriers to bike share: Lessons on bike share equity,” *Inst. Transp. Eng. J.*, vol. 88, no. 2, pp. 31–35, 2018.

[132] R. Dowling and J. Kent, “Practice and public–private partnerships in sustainable transport governance: The case of car sharing in Sydney, Australia,” *Transp. Policy*, vol. 40, pp. 58–64, 2015, doi: https://doi.org/10.1016/j.tranpol.2015.02.007.

[133] A. B. Steven, Y. Dong, and M. Dresner, “Linkages between customer service, customer satisfaction and performance in the airline industry: Investigation of non-linearities and moderating effects,” *Transp. Res. Part E Logist. Transp. Rev.*, vol. 48, no. 4, pp. 743–754, 2012, doi: https://doi.org/10.1016/j.tre.2011.12.006.

[134] N. Sikka, C. Vila, M. Stratton, M. Ghassemi, and A. Pourmand, “Sharing the sidewalk: A case of E-scooter related pedestrian injury,” *Am. J. Emerg. Med.*, vol. 37, no. 9, p. 1807. e5-1807. e7, 2019.

[135] W. Zhang, R. Buehler, A. Broaddus, and T. Sweeney, “What type of infrastructures do e-scooter riders prefer? A route choice model,” *Transp. Res. Part Transp. Environ.*, vol. 94, p. 102761, 2021, doi: https://doi.org/10.1016/j.trd.2021.102761.

[136] K. J. Krizek and N. McGuckin, “Shedding NHTS Light on the Use of’Little Vehicles’ in Urban Areas,” *Transp. Find.*, 2019.

[137] L. P. Hof, “Teaching Girls How to Ride a Bicycle: Gender and Cycling in Lunsar, Sierra Leone,” Master Thesis, 2016. Accessed: May 01, 2023. [Online]. Available: https://studenttheses.uu.nl/handle/20.500.12932/24972

[138] M. Shearlaw, “Harassment, stigma and fatwas: what is it like to cycle as a woman?,” *The Guardian*, Jun. 13, 2017. Accessed: May 01, 2023. [Online]. Available: https://www.theguardian.com/cities/2017/jun/13/harassment-stigma-fatwas-cycle-woman

[139] W.- Ng and A. Acker, “Understanding urban travel behaviour by gender for efficient and equitable transport policies,” no. 2018–01, Feb. 2018, Accessed: Apr. 29, 2023. [Online]. Available: https://trid.trb.org/View/1505620

[140] M. Söderlund, “Customer satisfaction and its consequences on customer behaviour revisited,” *Int. J. Serv. Ind. Manag.*, vol. 9, no. 2, pp. 169–188, May 1998, doi: 10.1108/09564239810210532.

[141] L. U. Tatikonda, “The hidden costs of customer dissatisfaction,” *Manag. Account. Q.*, vol. 14, no. 3, p. 34, Mar. 2013.