Data Science for Transportation (2026) 8:26
https://doi.org/10.1007/s42421-026-00167-3

RESEARCH

A Scenario Discovery Approach to Transit Network Optimization for Improved Access in Areas of Persistent Poverty

Efthymia Kostopoulou[1] · Tolu Oke[1] · Jimi Oke[1] · Eleni Christofa[1]

Received: 27 March 2026 / Revised: 26 May 2026 / Accepted: 3 June 2026

Abstract

Transportation is crucial for accessing essential destinations such as jobs. Lack of transportation access can increase economic polarization, health disparities, and food insecurity. This research develops an optimization model that maximizes a novel accessibility index by optimizing existing transit route headways. The proposed accessibility index highlights areas of low access to key destinations and high poverty density. Application of the model and scenario discovery are demonstrated through a case study of Amherst, Massachusetts, for access to healthcare and social assistance jobs. Since there is no explainable mapping between headway and access in the literature, an XGBoost regression model is used to estimate the impact of transit headway on access. Results are presented through scenario discovery and show that multiple combinations of optimal headways can substantially increase access to healthcare and social assistance jobs in low-income areas. The proposed optimization model accounts for social equity and is flexible in that it can be expanded to incorporate multiple destination types (e.g., education and food) and weights for the various destination types to address the specific needs of an area of interest.

Keywords Transportation access · Transit network optimization · Explainable AI · Decision-making · Scenario discovery

Introduction

Transportation is a social determinant of health, and such policies and management strategies should be targeted towards improving transportation access equity. Access to reliable transportation is a significant barrier for many people in the United States (U.S.), affecting their ability to reach essential services such as jobs, healthcare, and food. Studies by the Organization for Economic Co-operation and Development (OECD) demonstrate that equitable access to jobs can narrow income gaps and reduce systemic inequality (Organisation for Economic Co-operation and Development (OECD) xxx). To quantify the significance of transportation in jobs, for every $1 billion invested in public transportation, approximately 50,000 jobs are supported and created (American Public Transportation Association 2025). Specifically for healthcare and social assistance jobs, a policy report on commuting challenges for healthcare workers in New York found that in healthcare employees who rely on public transit had the longest median commute of any private-sector workers (Weintraub 2018). Ultimately, investigating access inequities to jobs is critical for the well-being of communities, especially those with high percentages of individuals with low or zero access to a car.

This paper proposes a mathematical model that optimizes transit service, i.e., transit headways, by maximizing access to key destinations, using readily available inputs. The novelty of our approach comes from the access measure that is maximized, which is a new accessibility index that highlights areas of high poverty density and low access, accounting for socioeconomically disadvantaged areas. Another contribution of this study is the framework used to solve the proposed optimization model. The framework includes a scenario discovery that provides sets of solutions rather than

✉ Efthymia Kostopoulou
ekostopoulou@umass.edu

Tolu Oke
toke@umass.edu

Jimi Oke
jimi@umass.edu

Eleni Christofa
echristofa@umass.edu

1 Department of Civil and Environmental Engineering, University of Massachusetts Amherst, 130 Natural Resources Road, Amherst, MA 01002, United States

Published online: 11 July 2026

unique best-found solutions, providing flexibility in decision-making. Lastly, even though the study evaluates accessibility by public transit, the framework is flexible and can be adapted to other transportation modes.

The optimization model contributes to the literature through its accountability of socioeconomic characteristics, therefore, promoting equitable access for all. The optimization framework is flexible in that it can be easily expanded to account for access to multiple destinations, different days (e.g., weekdays vs weekends), and times of day, when data is readily available, as well as incorporate other socioeconomic and demographic characteristics. While the framework is flexible in accounting for different destinations and socioeconomic characteristics, the methodology is presented here with a case study on access healthcare and social assistance jobs for low-income areas, using census block group as unit of resolution.

Literature Review

Transit Optimization

Various recent studies have developed optimization models for transit planning, with their primary contribution being improving the operational design of transit service. Park et al. (2022) proposed an optimization model that maximizes profit, minimizes unmet demand associated with trips requiring multiple transfers, and inequity, measured as the standard deviation of transit travel time relative to private vehicle travel time. Decision variables, among others, included headways, number and length of routes, and number of stations (Park et al. 2022). Other objectives included minimizing the total cost, including user access, waiting, and in-vehicle time costs as well as operator costs, by optimizing bus headway and vehicle size (Gayashani et al. 2026), as well as minimizing user and operator costs by optimizing bus headways and departure offsets (Ceylan and Ozcan 2018). However, these studies did not account for socioeconomic characteristics of the population that the tested network serves. In contrast, a number of other studies explicitly incorporate residents' socioeconomic status into the analysis. For instance, Camporeale et al. (2017) developed an optimization model that minimizes the user and operator costs as well as the unsatisfied demand, i.e., demand that cannot be served based on the current network, accounting for parameters such as income and unemployment, while Ruiz et al. (2017) optimized headways to maximize service level and social equity. However, these studies focus on assessing the quality of transit service between same origin–destination pairs, without accounting whether transit improvements meaningfully change access to destinations that involve employment activity or other essential services (Ruiz et al. 2017).

Few studies have developed optimization models to maximize transportation access to key destinations. Ding et al. (2023) optimized public transportation accessibility in the historical Beilin District of Xi'an, China, by minimizing the negative utility associated with walking, waiting, and in-vehicle travel times when using transit. The optimization included operational constraints, i.e., related to route length, network density, and line repetition factor, and applied a modified ant colony algorithm to generate the optimal layout of bus routes. The optimized transit network demonstrated improved performance across key indicators such as line length, stop spacing, and network density, and significantly enhanced accessibility, particularly in previously underserved districts. However, social equity was not addressed directly, since the socioeconomic background of the residents in that area was not considered in the optimization.

Another study developed a bi-objective optimization model that maximizes a weighted combination of (1) access from affordable housing units to jobs by public transportation and (2) an interaction-based impact measure, i.e., the effect created when two affordable housing developments are located close to each other, especially when they are large (Zhong et al. 2019). The authors used a weighted-sum method to solve the bi-objective problem and illustrated the trade-offs using Pareto fronts. The results showed that access to jobs by transit is initially low when the distribution of affordable housing is low, but small increases in the interaction-based impact measure lead to substantial gains in job access. As the interaction-based impact continues to increase, access to jobs changes only marginally.

Lastly, Li and Fan (2021) developed an optimization model that minimizes the Transit Gap Index (TGI). TGI incorporates demographic indicators along with spatial and temporal characteristics of transit service and uses General Transit Feed Specification (GTFS) data to quantify the mismatch between transit supply and transit demand at the block group level. The model minimizes the Transit Gap Index by optimizing the number of bus stops constructed in each census block group within a study area, subject to cost constraints. However, the proposed index does not account for people under poverty or any income-related variable, which substantially affects the accessibility experience of a user.

Access to Employment

Numerous studies have evaluated access to jobs using different metrics to identify access gaps. Cumulative opportunity measures such as the total number of services available within a distance or travel time threshold have been extensively used (Åslund et al. 2010; El-Geneidy et al. 2016; Boisjoly and El-Geneidy 2016; Slovic et al. 2019; Ermagun and Tilahun 2020; Hernandez et al. 2020; Tomasiello et al. 2020; Zhu and Shi 2022; Herszenhut et al. 2022).

While the cumulative opportunity metric accounts for the location of services and how difficult it is to access them, it does not account for the transportation impedance, i.e., the fact that as travel time increases, it is more difficult to reach an opportunity. Other measures used in job access include the gravity-based assessment of access that was proposed by Hansen (1959) and has been broadly used since to account for transportation impedance (Hess 2005; Osland 2010; Reggiani et al. 2011; Foth et al. 2013; Wang and Chen 2015; Hernandez 2018; Pritchard et al. 2019; McCahill et al. 2020). Cumulative opportunity gravity models rely on a distance decay function that places higher weights on the opportunities that are easier to reach. However, gravity-based assessment requires empirical survey data to define the distance decay function (Reggiani et al. 2011). Lastly, competition of labor demand and supply was incorporated in gravity models by Shen (1998) to account not only for the transportation impedance to a workplace but also how the number of accessible jobs decreases with the increase in the number of people seeking those jobs. Competition effects have been incorporated in various studies (Kawabata and Shen 2006; Pritchard et al. 2019; Barboza et al. 2021); however, these methods are quite complex for real-world implementation (Merlin and Hu 2017). This paper uses the cumulative opportunity metric due to data availability and its benefit to facilitate inclusion of various destinations simultaneously for future work.

These metrics have been incorporated in various studies that involve spatiotemporal analysis to reveal access gaps to employment. More specifically, studies have focused on exploring discrepancies in job access by transit versus personal vehicle specifically accounting for housing prices (Zhu and Shi 2022) or focusing on low-wage jobs (Yan et al. 2022). Results showed that access to jobs is fairer in areas with high housing prices (Zhu and Shi 2022), while jobs accessible by transit are only one-eighth of that accessible by personal vehicle (Yan et al. 2022). Apart from the role that travel time plays in access across different modes, other studies have investigated the role that competition plays in access to employment (Cheng and Bertolini 2013; Barboza et al. 2021). Results showed that the number of effectively accessible jobs dropped compared to traditional cumulative or gravity measures when incorporating competition (Cheng and Bertolini 2013). Lastly, studies have explored the temporal fluctuations of access to employment and other destinations (Lindner et al. 2024; Wang et al. 2024; Stepniak and Goliszek 2017; Yan et al. 2022), and found that the gap between transit and personal vehicle access is smaller during peak hours compared to evening hours or in the middle of the day (Yan et al. 2022). While these studies have revealed gaps across different modes, times of the day and demand patterns, further research is needed to inform transit changes that addresses them.

Equity in Accessibility

Several researchers have explicitly considered how access to destinations varies across population groups with different socioeconomic background. Ermagun and Tilahun (2020) explored how equitable access by transit is in Chicago accounting for jobs, parks, groceries, hospitals, schools, and libraries. They incorporated linear regression models with socioeconomic variables such as income, minority, and education. Results showed that the more the low-wage workers in an area, the lower the access is to the six tested destinations. Other studies developed metrics that measure access to low-income jobs by transit specifically for vulnerable residents, that are characterized by, e.g., high rates of unemployment and recent immigration (Deboosere and El-Geneidy 2018). Outcomes revealed that vulnerable populations are sensitive to any changes regarding job access. Overall, research up to date has concluded to the fact that worse socioeconomic conditions are associated with low access to jobs (Slovic et al. 2019; Allen and Farber 2020) that leads to unemployment (Hernandez et al. 2020) while low-income residents often experience longer travel time to work (Guzman et al. 2017). Therefore, further research is needed to present transit network improvements in vulnerable areas to address the issues.

Limitations of Existing Literature

The reviewed studies highlight limitations that should be addressed to ensure that transit improvements are designed to maximize access for all. Transit optimization studies typically focus on operational improvements but they do not account on the impact of those operational improvements on access to essential destinations and their impact to access equity by explicitly including access metrics in the optimization. The main objective in accessibility studies is to explore access gaps at various spatial and temporal levels given a transit network design and schedule and do not develop transit service improvement recommendations to address these gaps. Lastly, equity access studies incorporate comparisons in access that people from different socioeconomic background experience but again do not develop optimization methods to provide transit network changes to address them. There is a need to account for socioeconomic characteristics when optimizing transit network design to improve access equity.

To address the limitations in the literature, the objective of this research is to develop an optimization model for transit network improvements that advances equitable access to employment, specifically healthcare and social assistance jobs, subject to cost constraints in areas of persistent poverty. The primary contribution of the model is that it maximizes a novel accessibility index capable of identifying areas with low access to employment and high poverty density by optimizing only existing transit headways. This research improves upon existing studies by developing an optimization model that directly incorporates inequities caused by socioeconomic disparities and provides a scenario discovery framework that generates targeted transit improvement recommendations for improved transportation access equity.

Methodology

This section first defines the accessibility index that is maximized in the proposed mathematical model. Then, it presents the mathematical model that maximizes the accessibility index, followed by the overall optimization framework that shows step by step how the optimization model is solved.

Accessibility Index

The proposed Poverty Density Accessibility Index (PVDAI) is defined as the poverty density-weighted measured access of the census block group. It has been developed to account for differences in access to healthcare and social assistance jobs between areas of high and low poverty density as shown in Equation (1) for census block group i, generic mode, m and time period, t:

$$\mathrm{PVDAI}_{i,t,m} = \mathrm{PVD}_i \frac{A_{i,t,m} - \overline{A}_{t,m}}{\sigma_{A_{t,m}}} \tag{1}$$

where PVD_i is the poverty density of census block group i, $A_{i,t,m}$ the access for census block group i, during time period t, by mode m, $\overline{A}_{t,m}$ the mean access of all census block groups in the study area, during time period t, by mode m, and $\sigma_{A_{t,m}}$ the standard deviation of access for all census block groups in the study area, during time period t, by mode m.

Standardizing the measured access, before weighting it, ensures that areas with high poverty density but low access will present lower PVDAI scores than those with high poverty density and high access. If poverty density were to be multiplied by the measured access directly, then the index would present high values in census block groups that have very high poverty density but low access, veiling potential inequities.

The proposed index values and signs can be interpreted as follows: negative values with high magnitude indicate low job access and high poverty density while positive values with high magnitude indicate high job access and high poverty density. Amherst has medium to low access to healthcare jobs compared to the Pioneer Valley's average (red–orange–yellow colors) as shown in Fig. 1a, Fig. 1b shows that poverty density is high-medium (warm colors) in areas of Amherst. This is what leads to Fig. 1c, highlighting Amherst with negative PVDAI, i.e., low job access and high poverty density (red-orange colors). On the other hand, Springfield (highlighted with the yellow box) has very high access compared to the area's average value, and high poverty density, which lead to positive and high in magnitude PVDAI values in Fig. 1c (blue colors). While a standardized score centered around zero communicates relative advantage/disadvantage directly, index values allow for understanding where areas of high poverty density and low access versus high poverty density and high access are located. Other formulations such as normalized 0–1 index remove the interpretation of zero, which provides a meaningful midpoint. More specifically, index values around zero represent areas that are not considered vulnerable for the purposes of this manuscript, i.e., areas of low poverty density and low access or low poverty density and high access.

Mathematical Model

The mathematical model maximizes the Poverty Density Accessibility Index for all census block groups, i, in the area of interest, as shown in the objective function (2), which consists of the summation of the Poverty Density Accessibility Index values across all census block groups in the area of interest for a specific time period.

Constraint (3) ensures that the sum of the in-vehicle travel time t_i from each origin i plus the waiting time, $\frac{H_r}{2}$, add up at maximum to the travel time threshold τ.

Constraint (4) ensures that the total cost for all transit routes does not exceed a total budget B determined by the transit agency.

Constraint (5) defines upper and lower bounds for the transit headways.

Constraint (6) defines upper and lower bounds for the in-vehicle travel time.

The notation of the mathematical model is presented in Table 1.

$$\max_{H_r} \overline{\mathrm{PVD}}_i \cdot \frac{\overline{A}_i(H_r) - \overline{A}_j}{\sigma_{A_j}} \tag{2}$$

$$s.t.$$

$$t_i + \frac{H_r}{2} \le \tau \quad \forall i, r \tag{3}$$

$$\sum_r^R c_r \frac{Rt_r}{2 \cdot 60} \cdot \frac{60}{H_r} \cdot \mathrm{dr} \cdot \mathrm{op} \le \mathrm{B} \tag{4}$$

$$H_{\min,r} \le H_r \le H_{\max,r} \quad \forall r \tag{5}$$

$$0 \le t_i \le 60 \quad \forall i \tag{6}$$

A numerical example of constraint (4) for one route is presented below, with a headway of 20 min, runtime of 48.1 min, using a budget of \$285.7. Please note that this formulation

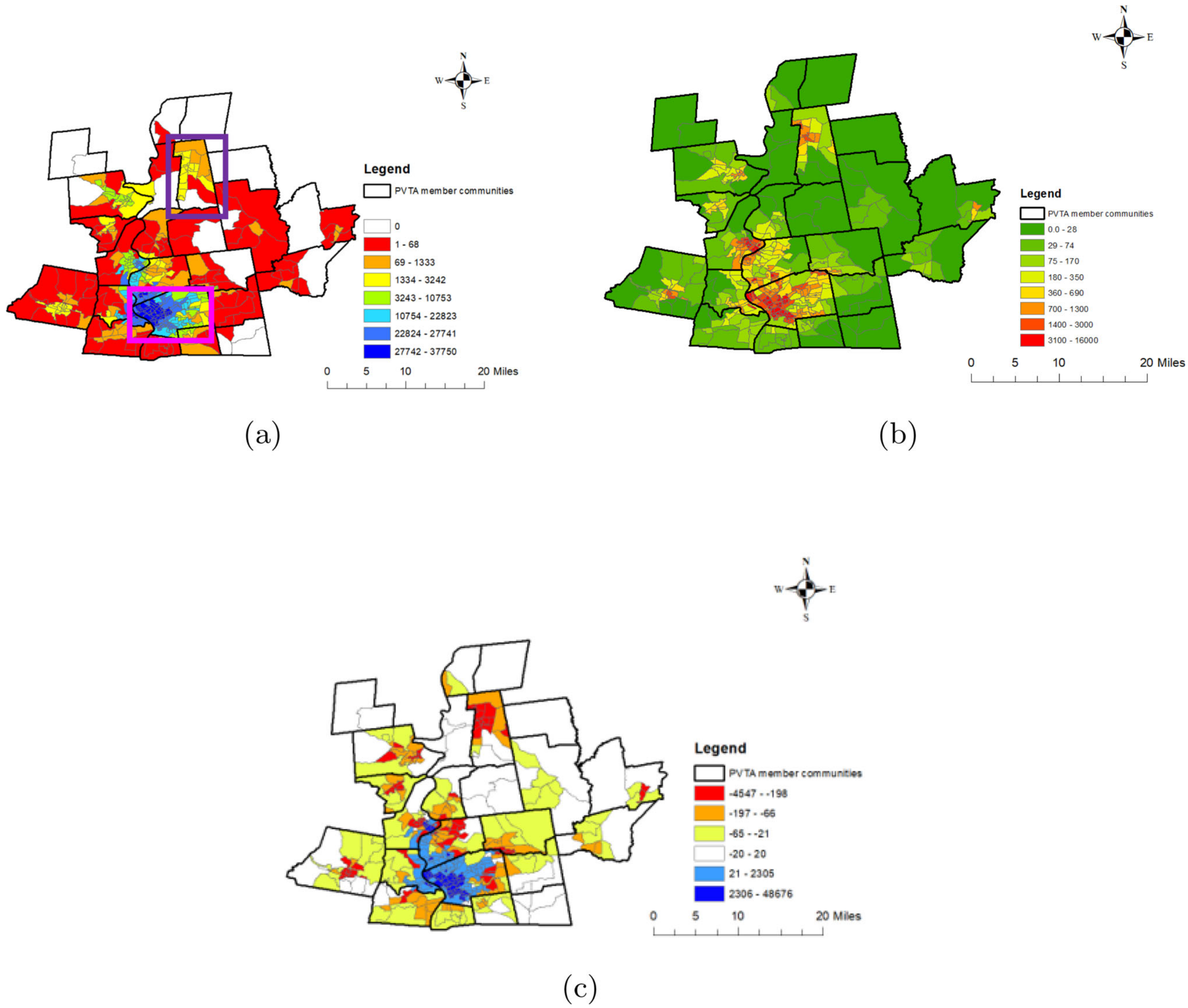


Fig. 1 **a** Measured access (number of healthcare and social assistance jobs), **b** Poverty density (people under poverty per sqmi) and **c** Poverty Density Accessibility Index (PVDAI) (Amherst is highlighted in the purple box and Springfield in pink box)

attempted to replicate the way costs are estimated in Remix, accounting for runtime, number of directions, hours of operation and headways and it is not the only way that a cost constraint can be built.

$$\left(100\,\frac{\$}{\text{hour}}\right)\cdot\left(\frac{48.1\text{ minutes of runtime}}{2\text{ directions}\times 60\text{ minutes of 1 hour}}\right)\cdot\left(\frac{60\text{ minutes of 1 hour}}{20\text{ minutes of headway}}\right)\cdot(2\text{ directions})\cdot(60\text{ minutes of operation})\leq 285.7\,\$$$

Case Study

The Pioneer Valley is a region in Western Massachusetts served by the Pioneer Valley Transit Authority (PVTA); see Fig. 2. PVTA operates a network of 41 local bus routes and 7 community routes, reaching a population of over 582,000; see Fig. 3 for the PVTA bus routes. Amherst, Massachusetts, was selected as the case study in this paper because it is one of the areas with a high number of vulnerable census block groups in terms of low income (Fig. 4) that at the same time have low access to healthcare and social assistance jobs compared to the rest of the study area (Fig. 5). The poverty density of the origin census block groups for the Amherst area ranges from 0 to 2,476 people under the poverty line per square mile (Fig. 4). The maximum poverty density of the region is 15,586 people under the poverty line per square mile and it appears in Springfield. Even though Springfield has very high poverty density as well (Fig. 4), it has the highest number of healthcare and social assistance jobs accessible in the PVTA service area (Fig. 5). Poverty density is defined as the

Table 1 Model notation

Category	Symbol	Description
Sets	$i \in \mathcal{I}$	Origins in the area of transit improvements (Census Block Groups of the area of interest)
	$j \in \mathcal{J}$	Origins in the overall study area (Census Block Groups in the area of service for the transit agency under consideration)
	$r \in \mathcal{R}$	Transit routes in area of interest
Decision variables	H_r	Headway of route r (min)
	t_i	In-vehicle travel time for route serving origin i within 60-min threshold (min)
Parameters	$\overline{A}_i$	Mean number of key destinations accessible from all origins i within 60 min
	$\overline{A}_j$	Mean access of all census block groups j in the area of service for the transit agency under consideration
	B	Budget ($), e.g., $2,000 per hour for all tested routes
	c_r	Operating cost per bus per hour for route r ($/hour/route)
	dr	Number of directions in trip (=2; =1 if circular route)
	op	Number of hours of operation (min)
	$\overline{PVD}_i$	Mean poverty density of origins i of the area of interest
	Rt_r	Runtime of route r (min)
	σ_{A_j}	Standard deviation of access for all census block groups j in study area
	τ	Travel time threshold (min), e.g., 60 min

number of residents living below the Federal Poverty Line based on the U.S. Department of Health and Human Services definition (U.S. Department of Health and Human Services 2025) (e.g., household income under $32,150 for a family of four in 2025) divided by the area of each of the 2020 U.S. census tracts. Census block groups that belong to the same census tracts were assigned the same value of poverty density, since poverty data were not available at the block group level. The routes whose headways were optimized are shown in Fig. 6 and the distribution of the destinations included in this analysis is presented in Fig. 5. In this study, Remix is used to edit General Transit Feed Specification (GTFS) files of the routes that serve Amherst, specifically routes 30, 31, 33, 38, 35, 46, and B43.

Optimization Framework

The framework used to solve the optimization model that maximizes the PVDAI by optimizing existing transit headways is presented in Fig. 7. The first step is data generation using Conveyal, which is a spatial analysis, web-based tool that is used to determine the number of healthcare and social assistance jobs that are accessible within a defined travel time threshold by transit mode from the census block group centroid. Conveyal builds transport networks from OSM street data and GTFS transit data, and is designed for multimodal accessibility analysis using transit, walking, biking, and driving networks (Conveyal 2026; Conveyal Network 2026). A key assumption is that accessibility is evaluated over a departure time window rather than at a single departure instant. For instance, in the present study where 7–8 a.m. is the window, Conveyal estimates access for every minute of the window and provides average estimate. Conveyal uses an open-source routing engine, *R*5 (Conveyal 2024), to estimate door-to-door travel times for the modes or interest. High-resolution grid cells (e.g., 251 x 177 feet), i.e., very small spatial units that, are used as destinations. When transit mode is enabled, Conveyal accounts for walking time to bus stop, waiting to board time, riding time, transfer time (walking), and walking time from a street network point to the center of the destination's grid cell. The default walking speed is 3.6 km/h (Conveyal 2026a). Conveyal does not include a "transfer penalty" parameter. However, it includes waiting time and extra walking time during transfers. Analysis interface also includes "Maximum transfers" as the upper limit on transfers considered in optimal trips (Conveyal 2026b). For scenario timetables, if "Times are exact" is unchecked, the timetable represents an assumed headway: vehicles depart at the specified frequency, but the exact offsets are unknown, so many possible schedules are simulated (Conveyal and Timetables 2026). If "Times are exact" is checked, Conveyal creates a fixed timetable with departures at the start time and then at the specified frequency until, but not including, the end time (Conveyal and Timetables 2026). The outcome of this tool is the number of accessible destinations within a set travel time threshold.

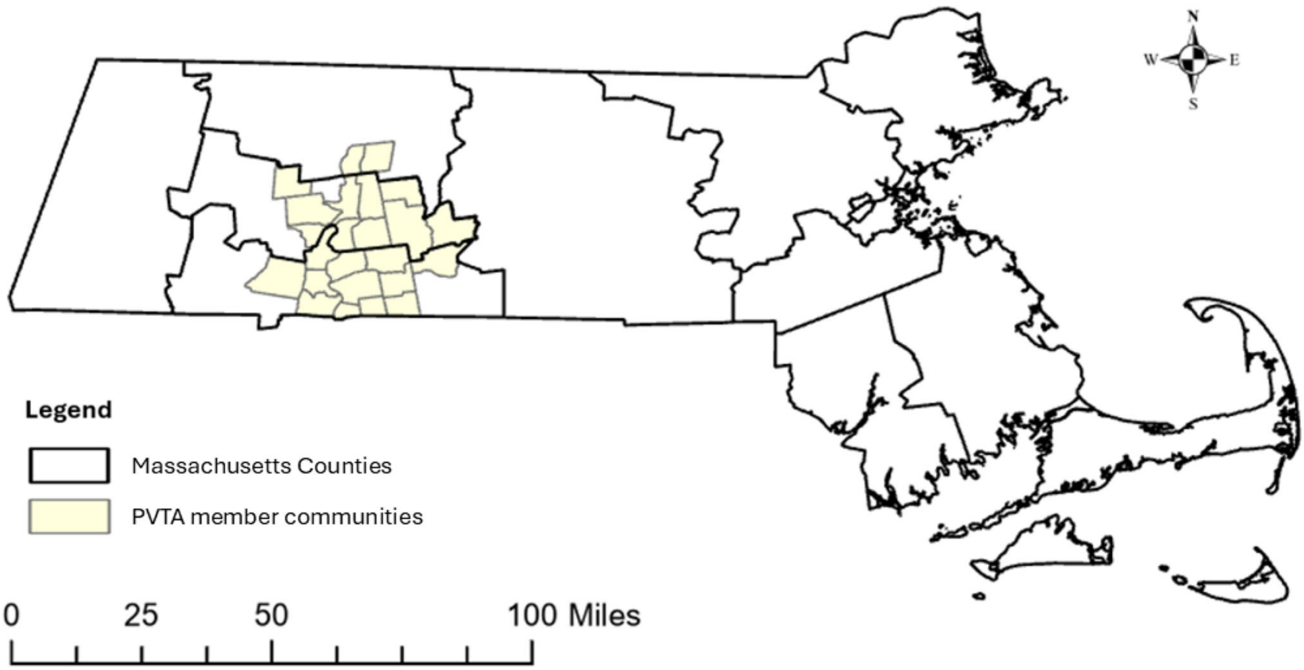


Fig. 2 Location of case study in Massachusetts

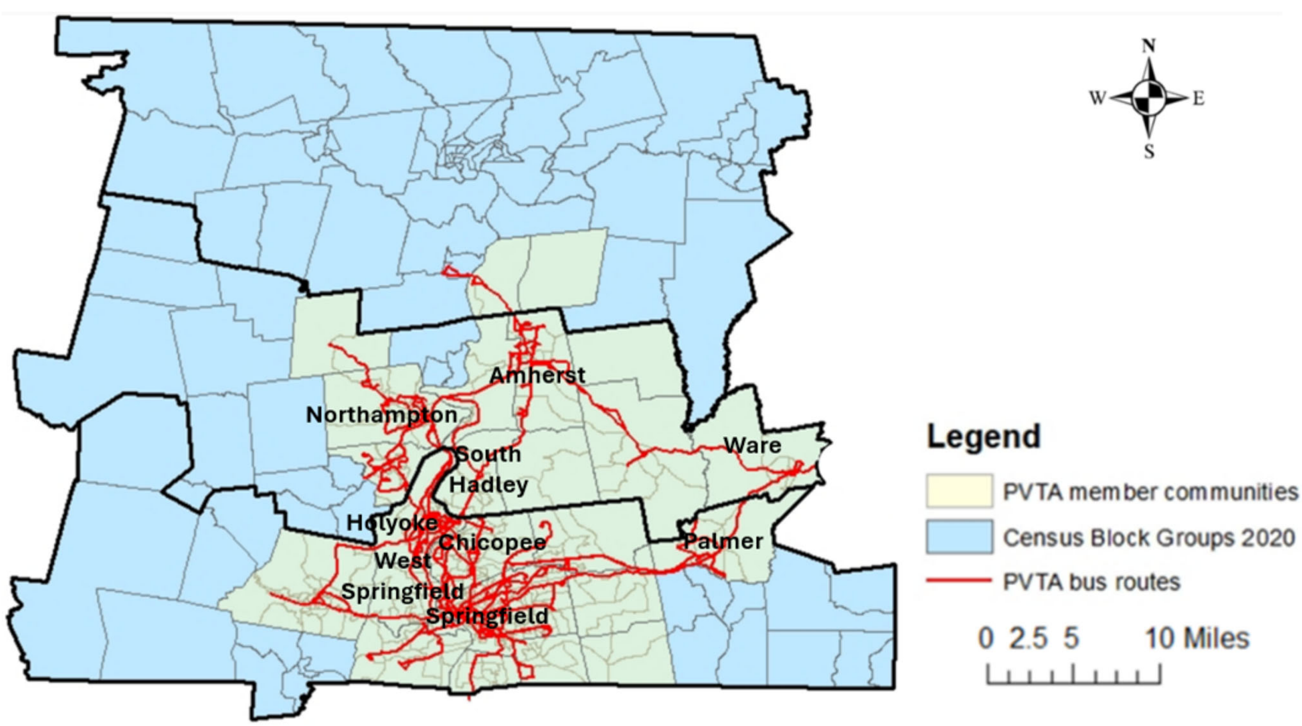


Fig. 3 Existing PVTA bus routes

The second step is to create the XGBoost regression model to predict access as a function of headway, which will be included in the objective function of the mathematical model. The need for this regression model came from the fact that there is no relationship in the literature that connects accessibility with transit headways. The third step is the state generation, where 1,000 different combinations of headways are produced for routes in the area of interest, using the lower and upper bounds for each route headway.

Fig. 4 Poverty density (ppl in poverty/sqmi) in PVTA member communities

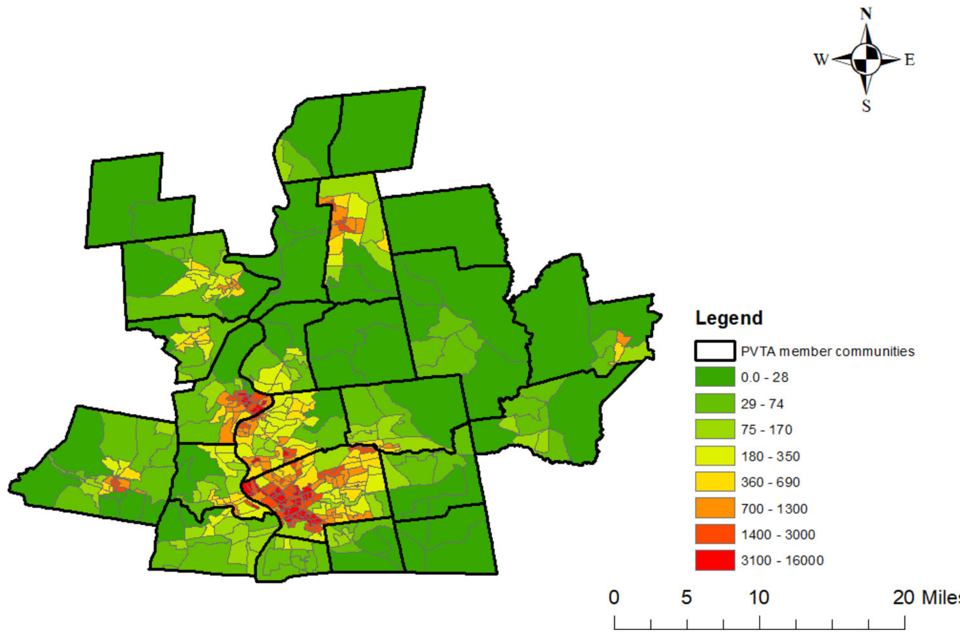
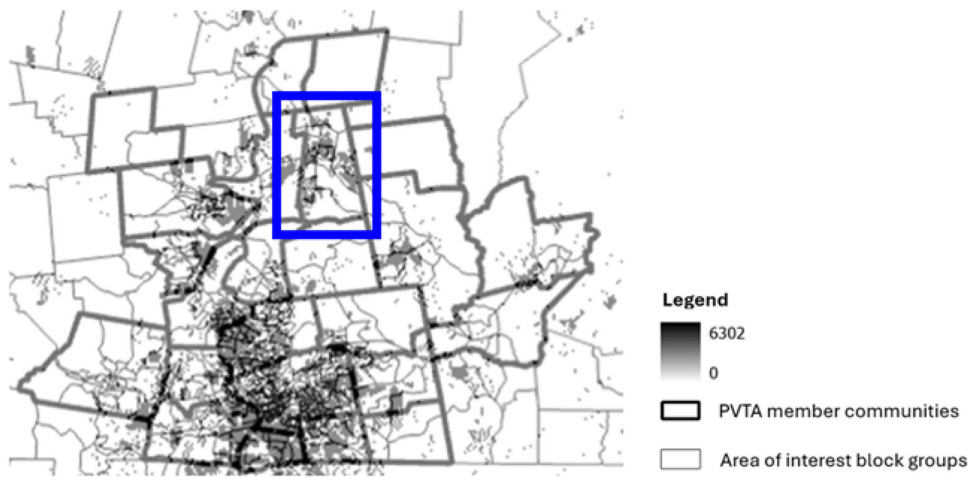


Fig. 5 Distribution of healthcare and social assistance jobs in Amherst (blue box) and Pioneer Valley area

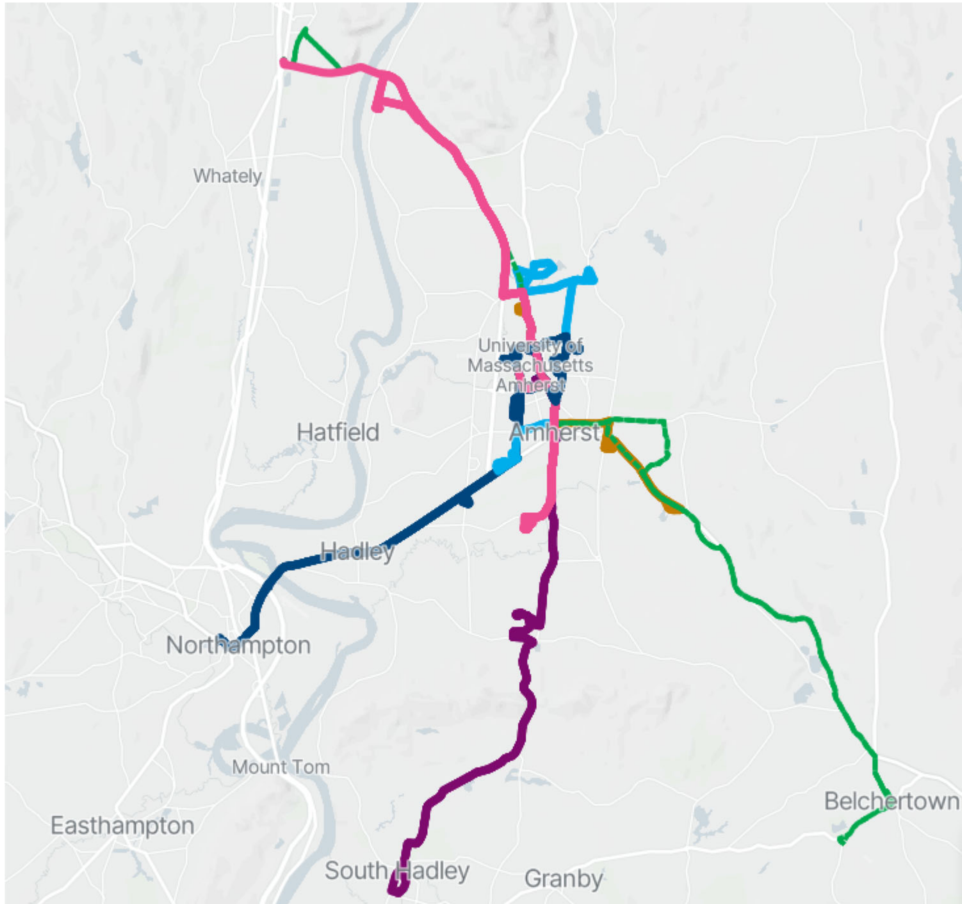


Next, the states are filtered based on the cost constraint, which is also a function of headways. For the remaining states, the access to healthcare and social assistance jobs is predicted using the XGBoost model that was trained in the second step. The PVDAI is then calculated by standardizing access and weighting it by poverty density. Since this is a maximization problem, we set a specific value as a threshold, i.e., that is the 0.75 percentile, and anything larger than this value is considered a satisficing solution. The remaining observations are feasible solutions and are assigned a label of 1 (success=1). The rest of the observations are assigned a label of 0. Based on this binary scoring, an XGBoost classification model is then built using the headway of the routes as features to see whether high or low headways of a specific route have a positive or negative impact on PVDAI, using the SHAP metric. Lastly, clustering based on headways is done on the successful solutions to obtain the range of headway values per route.

Data Generation

The first step is to obtain accessibility data, i.e., number of destinations within a travel time threshold, for various transit headway combinations. Given that there is no relationship in the literature that connects access to headway of transit routes, this study develops a regression model that predicts access based on various transit headways of existing routes. In this study, Remix (Via Transportation 2025) and Conveyal (Conveyal 2024) were used for this purpose. Remix is a cloud-based software used by cities, transit agencies, and planners to design, visualize, and manage public transit networks. In this research, Remix was used to edit General Transit Feed Specification (GTFS). Different combinations

Fig. 6 Transit routes in Amherst

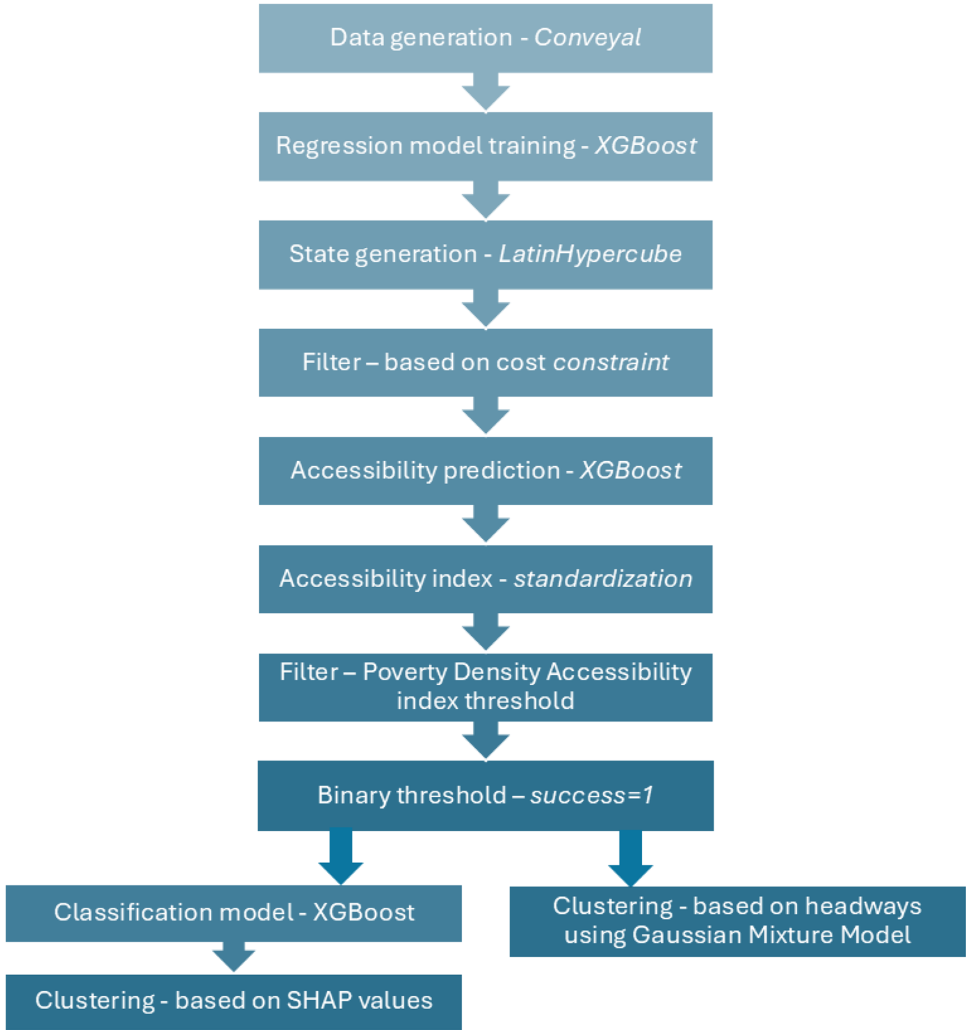


of transit headways were used as inputs in Remix and GTFS files were exported for each transit route headway combination. Note that the analytical framework does not depend on Remix itself; it depends on the ability to generate valid GTFS feeds representing alternative headway scenarios. Therefore, the same procedure could be implemented in another region using Remix, open-source GTFS editing tools, or scripts that directly modify GTFS frequency/schedule files. The GTFS files of each combination were then imported in Conveyal to obtain accessibility values. The transit travel time threshold in Conveyal was set to 60 min and the analysis was run for Friday, 21 October 2022. The analysis was run using healthcare and social assistance jobs as the destination type and the 7 –8 am period to reflect a typical morning peak hour in job access studies (Tomasiello et al. 2020; Slovic et al. 2019; Barboza et al. 2021). Healthcare jobs have been recognized as low-wage jobs in the literature (Yan et al. 2022), therefore selected in this study as point of interest since the main objective of this research is to address access in areas of persistent poverty. Each run took 110 s on average. It should be noted that even though healthcare workers may start their shifts at 7 am, i.e., their commute would have to start earlier than that, 7 am–8 am was used as the time period of analysis because no PVTA bus route in Amherst starts operating before 7 am. The output of the Conveyal analysis was the number of healthcare and social assistance jobs within 60 min by transit for each headway combination, and 100 different combinations were created in total. The proposed optimization model is simplified and does not account for operational parameters such as route alignment, stop locations, span of service, transfers, reliability, and network connectivity. To partially account for the effect of connectivity, before we select the range of values that headways can take as solutions (i.e., before we generate multiple headway combinations using the Latin hypercube), we test how each route individually affects access to healthcare jobs when headway changes, keeping all other routes in their regular schedule, as shown in Fig. 8. This figure shows that headways of 10 and 20 min appear in the daily schedules of all routes, except route B43. Therefore, conveyal analysis did not include all headway combinations, but different combinations of 10 and 20 min across all routes.

Accessibility Prediction Model

The XGBoost (eXtreme Gradient Boosting) regression model was trained next using the generated dataset in the previ-

Fig. 7 Optimization framework



ous step. This model used headway data to predict access, i.e., number of healthcare and social assistance jobs with 60 min by transit. XGBoost is an open-source, distributed machine learning library based on gradient-boosted decision trees (IBM 2025). XGBoost implements an advanced version of the gradient boosting framework, where successive weak decision trees are added to the model to correct the residual errors of previous trees, ultimately producing a strong predictive learner (IBM 2025). We used fivefold cross-validation during hyperparameter selection. The hyperparameters that were used to train the XGBoost model and the optimal values of them are presented in Table 2.The model has great performance, as depicted by its low mean absolute percentage error ($MAPE = 0.012$), low root mean square error ($RMSE = 782$) and high coefficient of determination ($R^2 = 0.96$) all of them being computed based on the test set (train/test set: 80/20). The fit of the model is also great as depicted in Fig. 9.

State Generation

Having a regression model that performs well, the next step was to generate a dataset of 1,000 observations, i.e., $s \in \mathcal{S}$, where $s = [H_{s1}, H_{s2}, \ldots H_{s1000}]$, and start filtering out observations based on the constraints of the optimization model, concluding with only observations that satisfy all the constraints while maximizing the PVDAI (feasible solutions). To randomly create the observations, a Latin hypercube (Mckay et al. 2000) was used. A Latin hypercube sample generated n points in $(\ell, u]^d$, where d is the dimension of the parameter space (SciPy Community 2025). Lower (ℓ) and upper bounds (u) were also set as presented in Table 3. Note that the upper bound is the headway of each route in the current bus schedule. The next step was to filter out any infeasible points, by applying the budget constraint (4) to ensure all solutions were feasible from a financial standpoint. Any points that remain are the budget-feasible states.

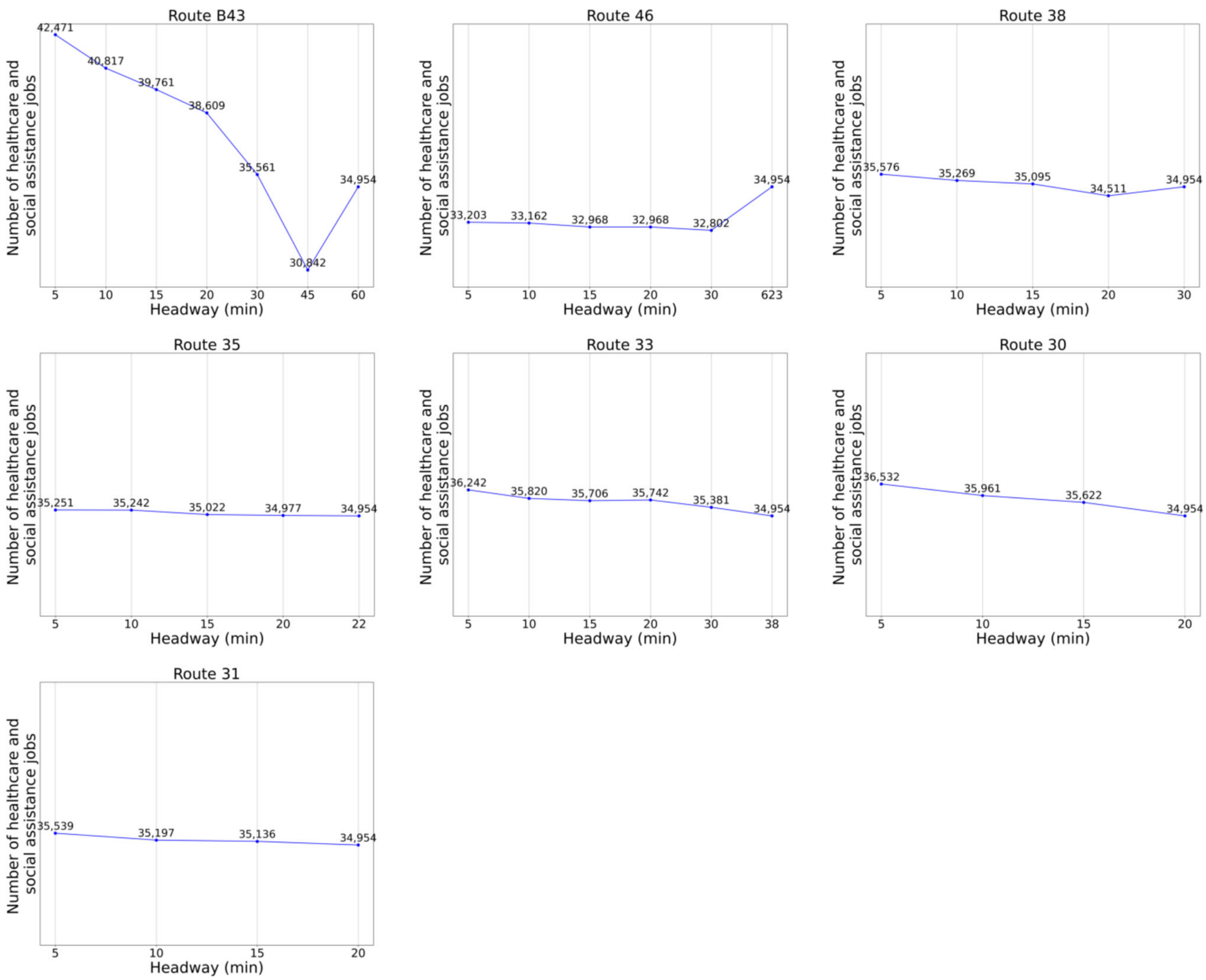


Fig. 8 Change in access to healthcare and social assistance jobs as headways vary across Amherst transit routes

Table 2 XGBoost hyperparameters used for regression model training

Parameter	Description	Tested values	Optimal value
max_depth	Maximum depth of each tree	{2,7}	2
min_child_weight	Minimum sum of instance weights to create a leaf	{5,10}	10
η	Learning rate controlling step size shrinkage	{0.01,1}	1
subsample	Fraction of observations sampled per tree	{0.5,0.7}	0.7
colsample_bytree	Fraction of features sampled per tree	{0.5,0.7}	0.5
λ	L2 regularization term (ridge penalty)	{5,10}	5
α	L1 regularization term (lasso penalty)	{1,5}	1
n_estimators	Number of boosted trees	{200,400}	200

Accessibility Prediction: XGBoost and Standardization

For the budget-feasible states, the developed XGBoost regression model was applied to predict accessibility, z, i.e., number of healthcare and social assistance jobs within 60 min by transit, where $z = f(s)$, $s \in \mathcal{S}$. Then, the PVDAI was created by standardizing the predicted access and weighting it by poverty density.

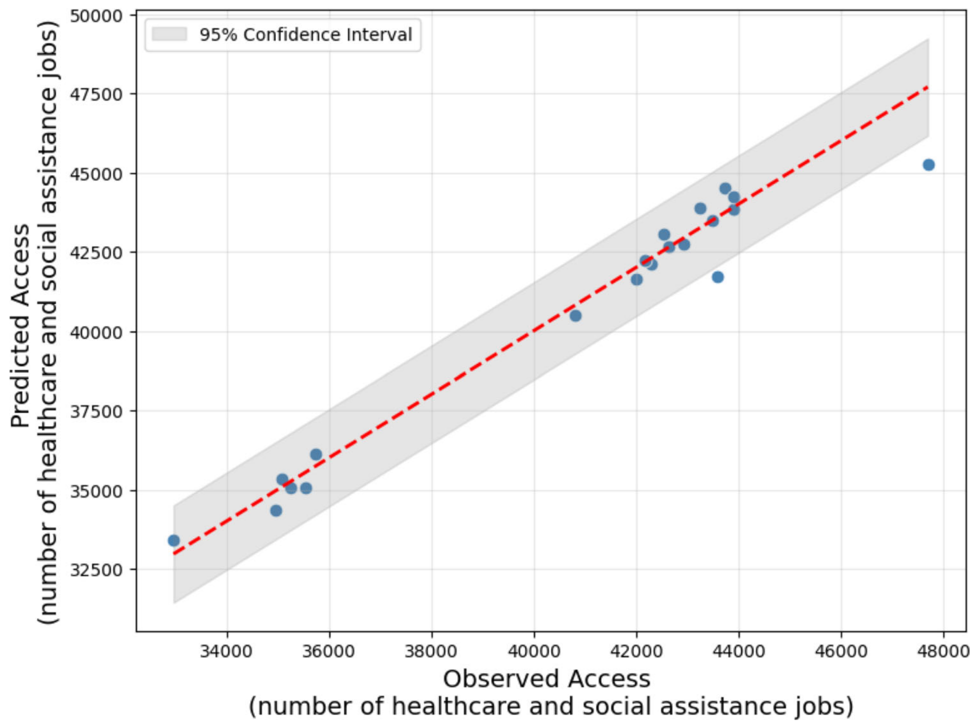


Fig. 9 Predicted vs observed access based on XGBoost regression model

Table 3 Lower and upper bounds on transit route headways

Route	Lower bound (min)	Upper bound (min)
30	5	20
31	5	20
33	5	38
35	5	22
38	5	30
46	5	623
B43	5	60

Filtering Poverty Density Accessibility Index Based on Threshold

To decide on the PVDAI threshold, the cumulative distribution function and a histogram of the index were plotted; see Fig. 10. The 0.75 quantile was chosen as the minimum threshold for maximizing PVDAI, which resulted in any PVDAI values higher than this threshold remaining in the set of PVDAI-retained solutions. Figure 10 shows that the distribution of the PVDAI is bimodal. The reason behind this pattern is more model-driven than spatial-driven. Since the XGBoost regression model predicts access within a wide range of headways (generated using Latin Hypercube sampling), some headway combinations produce low access to healthcare and social assistance jobs, while others high access, with some routes having more significant impact in determining access to healthcare and social assistance jobs than others.

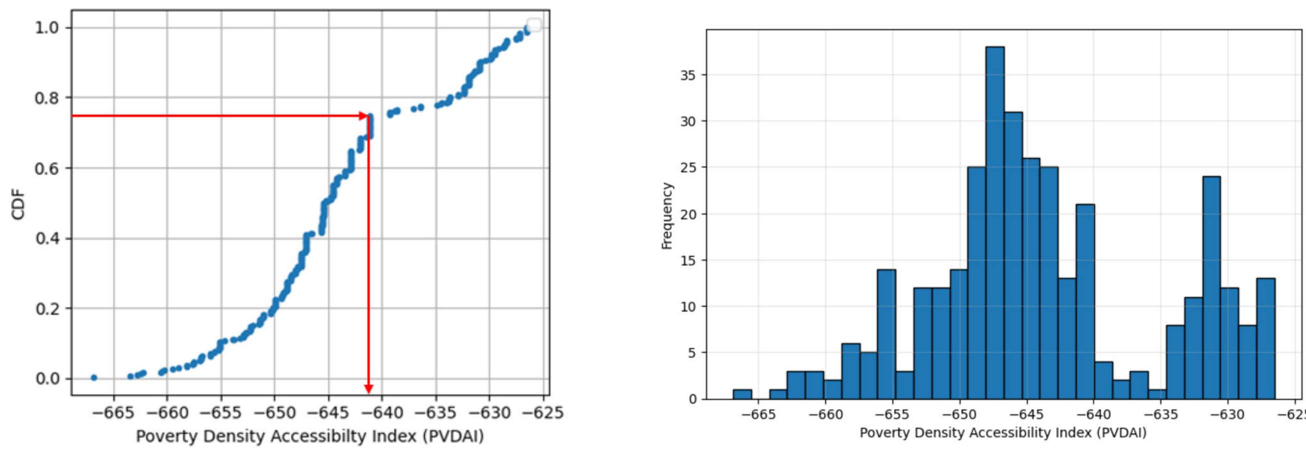
Success Thresholding and Classification Model

The PVDAI-retained solutions identified in the previous step were assigned a label of 1 ($success$=1). The rest of the observations are assigned a score of 0. Based on this binary scoring, an XGBoost classification model was then built using the transit route as features and was used to assess whether high or low headways of specific routes have a positive or negative impact on PVDAI. Table 4 shows the tested and optimal values of the hyperparameters in the training process of the XGBoost classification model and Table 5 summarizes the performance metrics of the classification model based on the test set.

Model Interpretation

The impact of each feature (i.e., headway of transit routes) was revealed using SHAP (SHapley Additive exPlanations) plots. SHAP is an *explainable AI* technique that assigns each feature an "importance" value, showing its contribution to a specific model prediction, rooted in cooperative game theory (Lundberg and Lee 2017). More specifically, the SHAP value, ϕ_h, shows the magnitude and direction that each independent variable, x_h, has on access to healthcare and social assistance jobs, and is estimated based on Eq. (4):

$$\hat{f} = \phi_0 + \sum_{h=1}^{M} \phi_h \tag{4}$$



(a) Cumulative Distribution Function (b) Distribution

Fig. 10 Statistics of poverty density accessibility index on the dataset of the optimization framework

Table 4 XGBoost hyperparameters used for classification model training

Parameter	Description	Tested values	Optimal value
max_depth	Maximum depth of each tree	{2,7}	7
min_child_weight	Minimum sum of instance weights to create a leaf	{5,10}	10
η	Learning rate controlling step size shrinkage	{0.01,0.1}	0.1
subsample	Fraction of observations sampled per tree	{0.5,0.8}	0.8
colsample_bytree	Fraction of features sampled per tree	{0.5,0.8}	0.8
λ	L2 regularization term (ridge penalty)	{5,10}	10
α	L1 regularization term (lasso penalty)	{0.1,1}	1
n_estimators	Number of boosted trees	{100,300}	200

$$\phi_h = \sum_{S \subseteq \{x_1,\ldots,x_H\} \setminus \{x_h\}} \frac{|S|!\,(|H|-|S|-1)!}{|H|!}\left[\hat{f}_{S\cup\{h\}}(x_{S\cup\{h\}}) - \hat{f}_S(x_S)\right] \tag{7}$$

where ϕ_h is the marginal contribution that each feature, x_h, has across all subsets S of input features, and $\hat{f}$ is the explanation model

The SHAP clustering was applied based on the K-means method using the SHAP value as variable. These plots were created using only the feasible solutions (success=1) and each dot on the plot is a feasible solution (i.e., pair of routes' headway and access). Therefore, a red dot means that high headway values have a positive impact on the model's outcome (success=1) if it is on the positive side on the x-axis, or a negative impact on the model's outcome (success=1) if it is on the negative side on the x-axis.

Table 5 Performance results for XGBoost classification model

Class	Support	Precision	Recall	F1-score	Accuracy
0 (Failure)	47	0.96	1.00	0.98	0.97
1 (Success)	22	1.00	0.91	0.95	

Discovering Data-Driven Scenarios via Clustering

We used the Gaussian Mixture Model (GMM) clustering algorithm to identify data-driven scenarios that represent similar accessibility outcomes of interest (i.e., from among the successful ones). The variables used to perform the clustering were the route-based headways. All types of covariance were tested, i.e., full, tied, diagonal and spherical and the spherical was selected since it provides a reasonable pattern (i.e., more clusters reduce error) and higher silhouette score values, given the low number of solutions (N=108). The silhouette score (scikit-learn developers 2024) was used to decide the number of clusters. The silhouette score evaluates clustering quality, measuring how similar a data point is to its own cluster (cohesion) compared to other clusters (separation), with scores varying from -1 to 1, with high silhouette score being beneficial. In addition to the silhouette score, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Log-likelihood (LL) were used to examine the optimal number of clusters. Low values of AIC and BIC are considered beneficial. These two metrics control for balance between goodness of fit and GMM model complexity (e.g., number of predictors). Lastly, the higher the LL values (i.e., less negative), the better the fit of GMM model. Six clusters were selected ($k = 6$) since this number led to the high silhouette score value along with low AIC and BIC and high LL values, as shown in Fig. 11. Throughout the remainder of this paper, we refer to these clusters of outcomes as *scenarios*.

Results

Scenario 1: Moderate Access and Variable Cost

This scenario is characterized by moderate PVDAI values compared to other scenarios and variable cost, as shown in Figs. 12 and 13, respectively. Figure 14a shows that moderate values of headways of route B43 can significantly and positively affect the increase of PVDAI (i.e., purple dots are on the positive side of the horizontal axis around value 5), and high values have a negative impact on PVDAI (i.e., red dots are on the left side). The median headway of B43 is 28 min (Fig. 15a) in this scenario. Routes 35 and 38 can affect the increase of PVDAI positively when operating with small headways and negatively when operating with large headways since blue dots are on the positive side of the horizontal axis and red dots are on the negative side of the same axis (Fig. 14a), with a median headway of 14 and 19 min, respectively (Fig. 15d and e). Routes 30, 31 and 33 have a smaller impact on increasing PVDAI compared to routes B43, 35 and 38 but still a positive impact at lower headways and negative at high headways (see Fig. 14a). Median headways of routes 30, 31, and 33 are 14, 13 and 22 min, respectively in this scenario. Lastly, large headways of route 46 can have a positive impact on increasing PVDAI, while in this scenario specifically, the median headway is 350 min.

Scenario 2: Steady Access and Variable Cost

This scenario has steady PVDAI values as shown in the low interquartile range (IQR), i.e., length of boxplot, in Fig. 12 and variable cost (Fig. 13). Figure 14b shows that various headways (i.e., low, moderate, high values) can have a positive impact on increasing PVDAI. The median highway for route B43 is 25 min in this scenario (Fig. 15f). In this scenario, routes 31, 35 and 38 can be beneficial in increasing PVDAI under low headway values and have a negative impact when operating with high headway values (Fig. 14b). The median headways for these routes are 13, 15 and 20 min, respectively (Fig. 15b, d and e). Routes 30 and 31 tend to increase PVDAI under moderate headway values (Fig. 14b) and their median headways are 14 and 22 min, respectively. Lastly, route 46 does not appear to a have significant impact on increasing PVDAI in this scenario (Fig. 14b) and its median headway is 250 min.

Scenario 3: Variable Access and Variable Cost

In this scenario, both access and cost are variable as depicted by the very large IQR in Fig. 12 and moderate IQR in Fig. 15. Figure 14c shows that only low headway values can be positively affecting PVDAI improvement and moderate and high values have clear negative impacts. The median headway in this scenario for route B43 is 27 min. Routes 38 and 35 are beneficial for increasing PVDAI at low headways and lead to reduction of PVDAI at moderate and high headways. The median headway values of these routes are 19 and 14 min, respectively. In this scenario, moderate headways of route 31 and low headway values of 30 and 33 can increase PVDAI Fig. 14c. The median headway values of these routes are 14, 13 and 26 min, respectively (Fig. 15b, a and c). Route 46 does not play a crucial role in this scenario (Fig. 14c), with a median headway being at 310 min.

Scenario 4: Variable Access and Low Cost

This scenario is characterized by variable PVDAI values, with PVDAI being within a smaller range of values compared to scenario *Variable Access and Variable Cost* (Fig. 12). This scenario has the lowest cost among all scenarios (Fig. 13). Both high and low headway values for route B43 can be effective in increasing PVDAI (Fig. 14d). Routes 30, 33, 35 and 38 improve PVDAI for low headway values only, while headways of routes 46 and 31 do not seem to have a substantial impact on PVDAI (Fig. 14d). The headway of route 33 varies more than in scenarios *Moderate Access and Variable Cost* and *Steady Access and Variable Cost* (as shown by the higher IQR in Fig. 15c), with a median headway of 24 min, lower than in scenario *Variable Access and Variable Cost*, but within a range that includes higher headway values.

Scenario 5: Variable Access and High Cost

This scenario has variable PVDAI values, with its maximum value being the highest among all scenarios (i.e., most improved PVDAI); see Fig. 15). However, it provides ranges of high cost values (Fig. 13). Route's B43 low headway values tend to substantially improve PVDAI and high/moderate values tend to reduce it but not significantly (Fig. 14e. In contrast to all the previous scenarios, route 46 plays a significant role in this scenario, as it can improve PVDAI in low headway values in this scenario and can decrease it under high headway values (Fig. 14e) . Route 46 has a median headway of 425 min. Moderate headways of route 33 can be beneficial for PVDAI in this scenario (Fig. 14e) with median headway

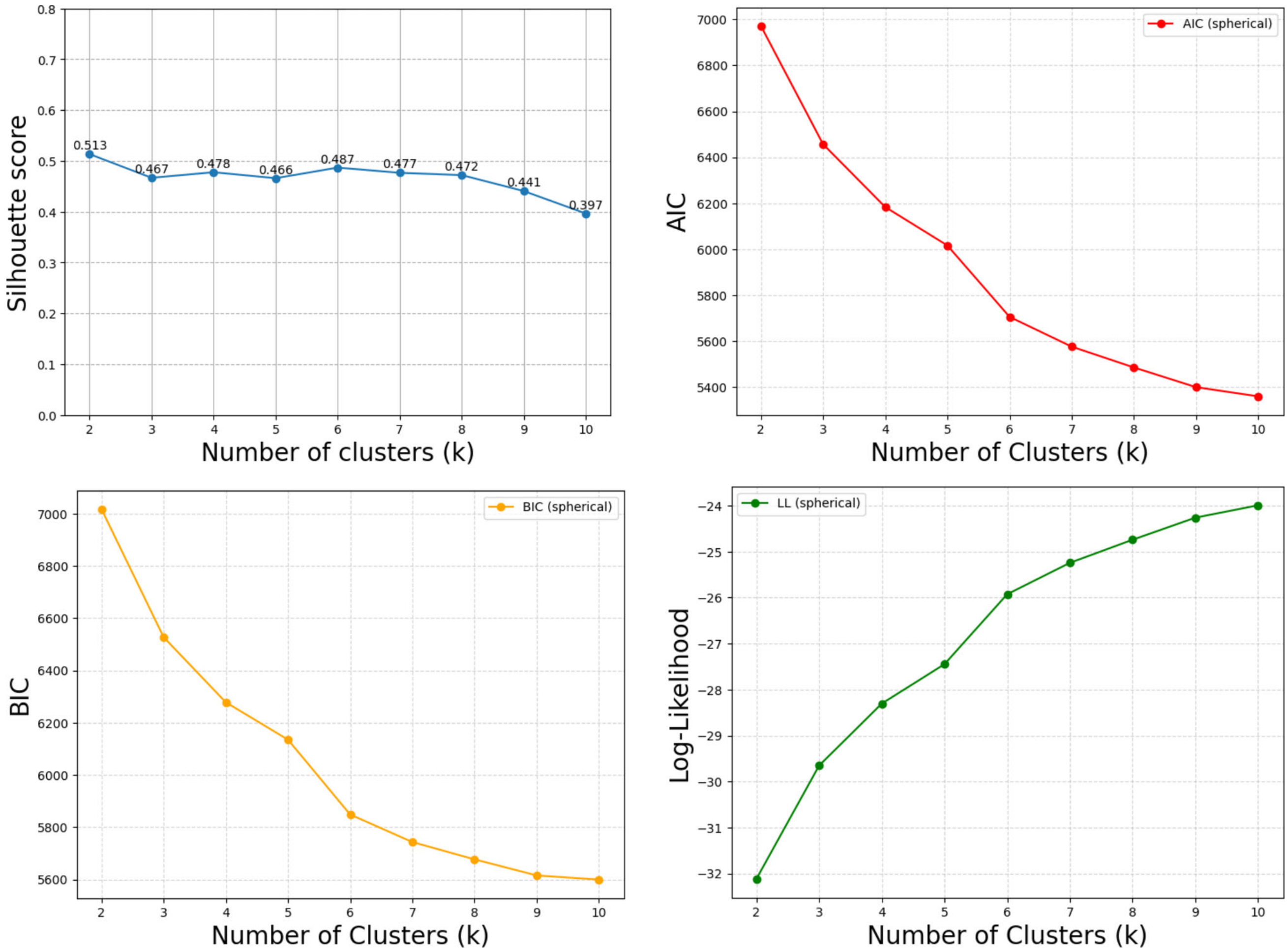


Fig. 11 Tested scores for selecting the optimal number of clusters

values being at 23 min. Low to moderate headway values on Routes 30, 31, and 35 (median 13 min) and on Route 38 (median 18 min) appear particularly effective in increasing PVDAI.

Scenario 6: Variable Access and Moderate Cost

This last scenario has the lowest median PVDAI together with scenario *Moderate Access and Variable Cost* and the second lowest cost after scenario *Variable Access and Low Cost*. Route B43 is steady in this scenario (as shown by the very low IQR in Fig. 15f), with solutions for headway being 27 or 28 min. Route 46 has the lowest headway among all scenarios, equal to 170 min (Fig. 15g). Routes 30, 31, 33, 35 and 38 can help maximize PVDAI in low/moderate headway values (Fig. 14f) and the ranges of their headways are similar to the ranges of the first four scenarios (Fig. 15a-e).

Practical Implications

A summary of scenarios along with practical implications is presented on Table 7. The proposed framework can be used by decision makers to support their goal on more equitable modifications in transit systems. First, the decision makers need to define their priorities such as improving access to essential destinations, by selecting equity indicators that will reflect transit changes, destinations types, time periods and the travel time threshold in which they want to improve access. For example, employment was the focus of this study while an agency can focus on supermarkets, childcare or healthcare services. In addition, similarly to poverty density other equity indicators might be older adults or zero-car households. Second, the agency needs to select specific routes that would be candidate for changes, minimum and maximum headways as well as defining their budget. Third, different headway combinations can be generated and the access these headways provide can be evaluated using the predictive accessibility model and the proposed index.

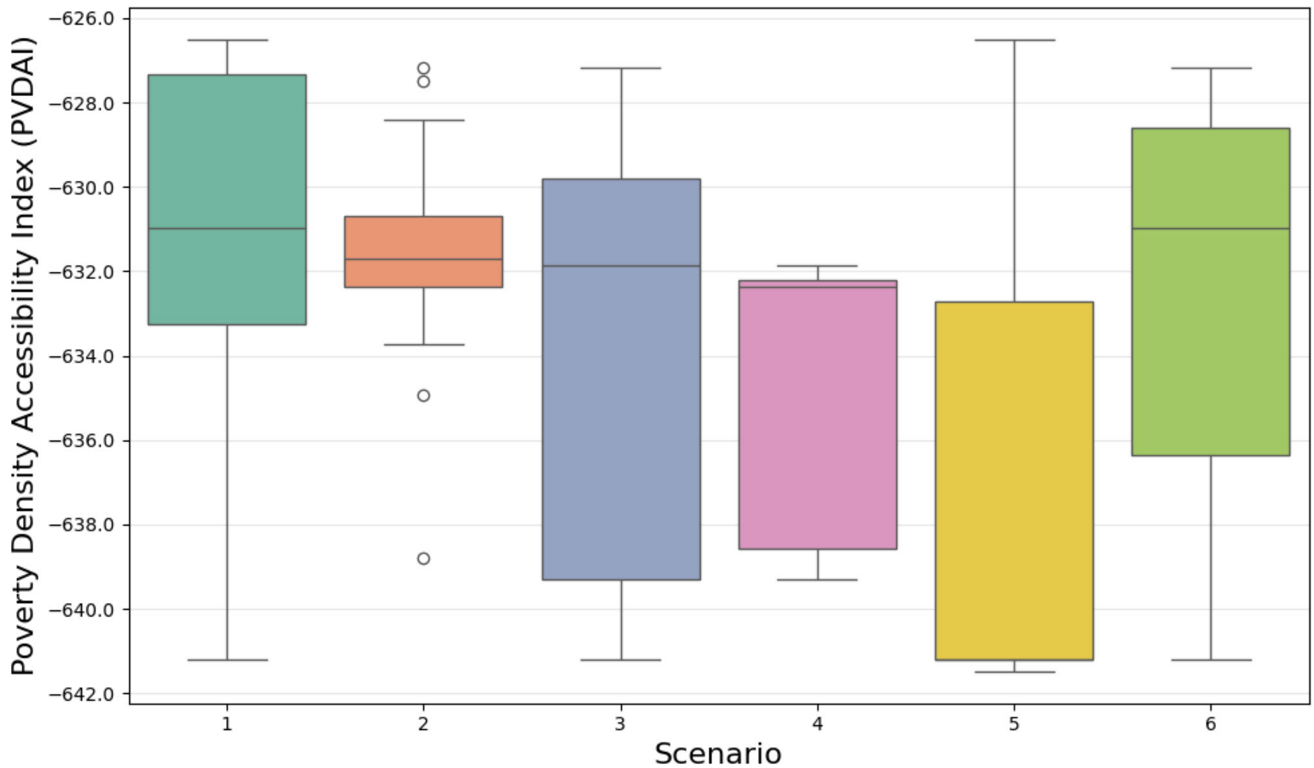


Fig. 12 Distribution of poverty density accessibility index by scenario (Baseline PVDAI = −660.67)

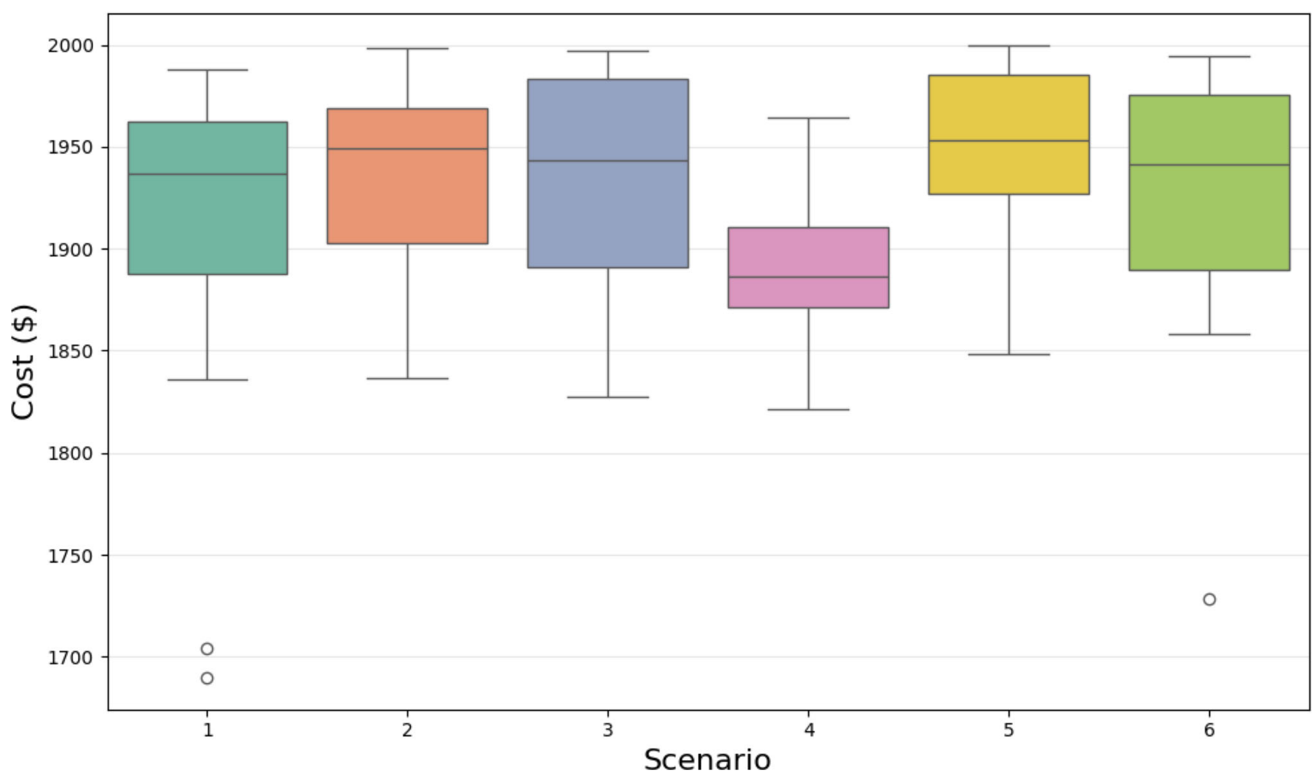


Fig. 13 Distribution of cost by scenario

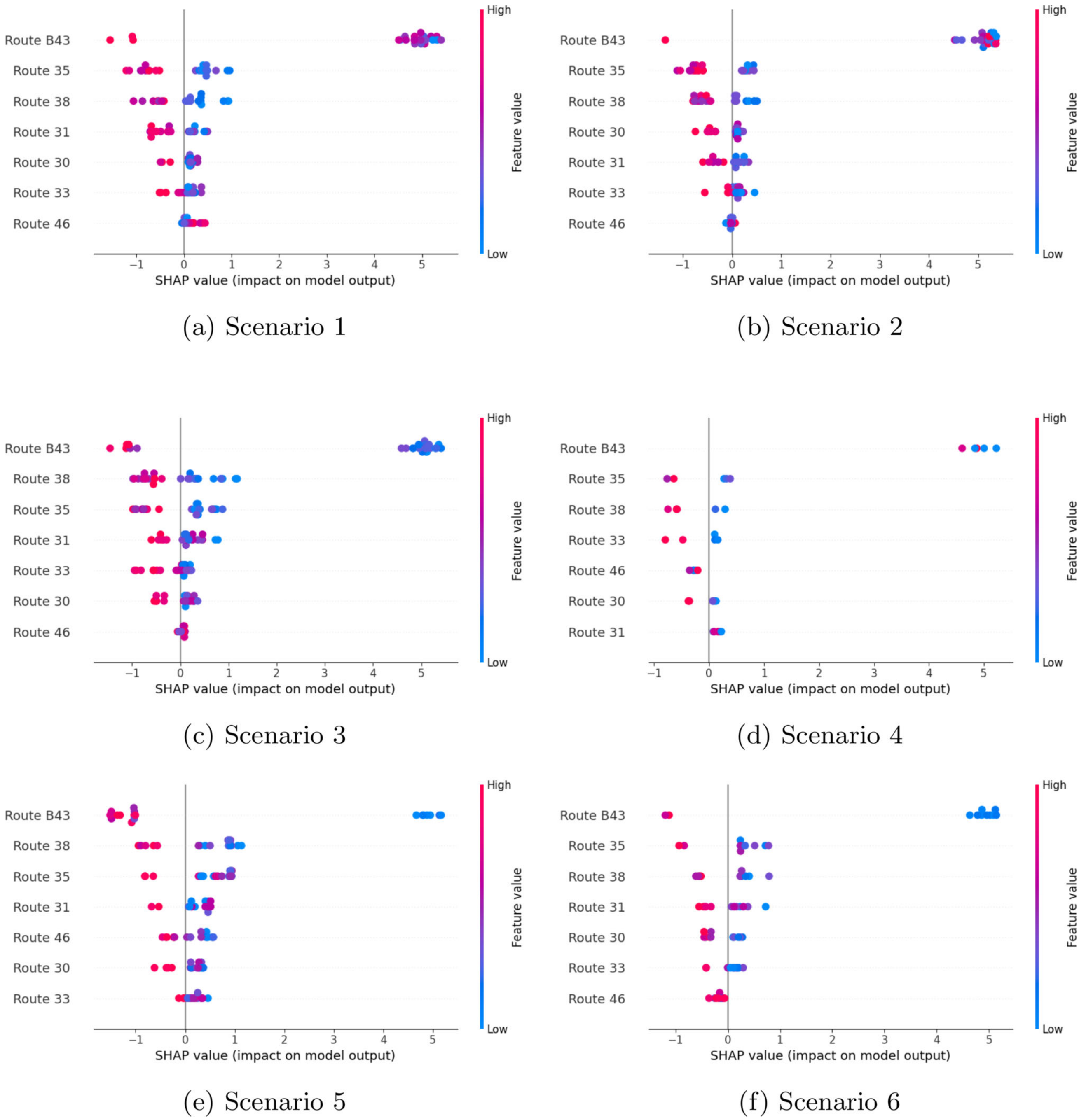
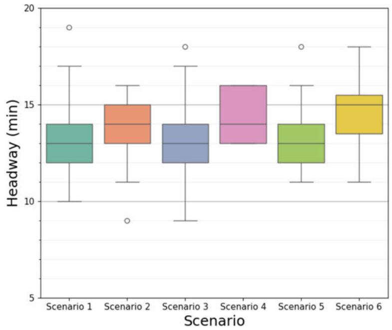


Fig. 14 SHAP plots for the six scenarios

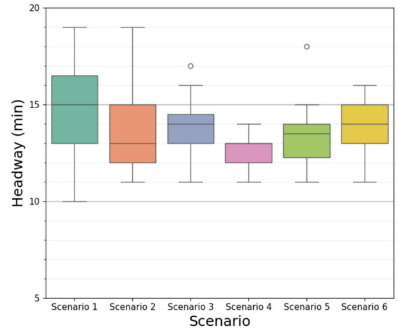
Fourth, decision makers should choose a success threshold. Finally, the satisficing solutions are grouped into scenarios with common characteristics in terms of main routes, access gains and cost. The last step is comparison across the scenarios. For example, if the main goal is to achieve the maximum equitable access, then a high cost scenario might be chosen. If a stable accessibility improvement is preferred, scenarios with low variation of PVDAI are preferable. In this way the proposed framework offers transparent comparison along candidate solutions and promotes equity in decision-making.

Sensitivity Analysis

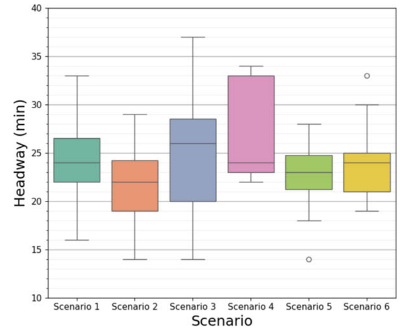
In this section, we provide sensitivity analysis based on the PVDAI threshold, and the number of clusters per threshold. Specifically, results are compared to 80th and 90th percentile of PVDAI. Number of clusters gets smaller as we shift to



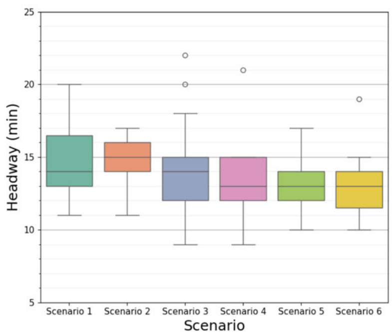
(a) Route 30



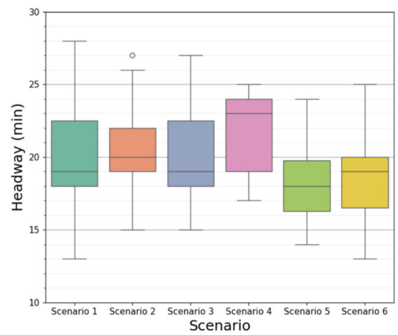
(b) Route 31



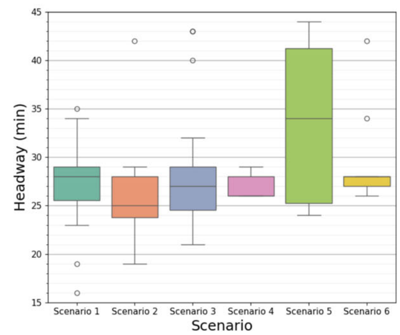
(c) Route 33



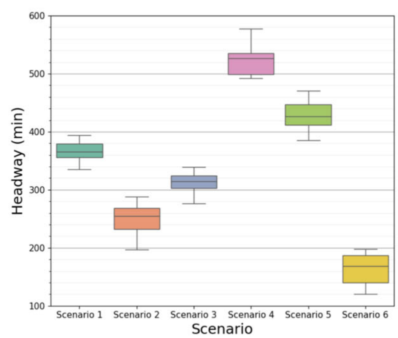
(d) Route 35



(e) Route 38



(f) Route B43



(g) Route 46

Fig. 15 Headway boxplots for each route across scenarios 1–6

stricter PVDAI thresholds, based on Figs. 16 and 17. However the number of successful solutions gets lower as we choose a stricter PVDAI threshold, as shown in Table 8. Therefore, any clustering results should be interpreted carefully because fewer observations are available to define the distinct scenarios.

Discussion

The results illustrate six distinct sets of solutions that improve PVDAI and, consequently, enhance access to healthcare and social assistance jobs in areas of persistent poverty. Compared to existing conditions, where the number of accessible healthcare and social assistance jobs is 34,954, all six sets of solutions substantially increase access. Specifically, comparing the median of each scenario to the existing conditions

Table 6 Summary of statistics by scenario

Statistic	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
			Access (number of healthcare and social assistance jobs)			
Median	43161.91	42957.57	42910.29	42775.95	40337.25	43161.91
Maximum	44394.30	44208.77	44208.77	42910.29	44394.30	44208.77
Minimum	40337.25	41002.23	40337.25	40862.11	40254.68	40337.25
			Cost ($)			
Median	1936.72	1948.94	1943.45	1886.14	1953.13	1941.36
Maximum	1987.68	1998.49	1997.37	1964.07	1999.93	1994.32
Minimum	1689.73	1836.81	1827.28	1821.41	1848.15	1728.34
			PVDAI			
Median	−630.96	−631.70	−631.87	−632.36	−641.19	−630.96
Maximum	−626.50	−627.17	−627.17	−631.87	−626.50	−627.17
Minimum	−641.19	−638.78	−641.19	−639.29	−641.48	−641.19

Table 7 Summary of scenarios and practical implications

Scenario	Main routes affecting PVDAI improvement	Access–cost trade-off	Practical implication
1	B43, 35, 38	Moderate PVDAI improvement with variable cost	It can be used when the goal is access gains without applying major changes to all routes
2	B43, 31, 35, 38	Steady PVDAI values with variable cost	It can be used when the goal is predictable accessibility improvement not necessarily maximum improvement
3	B43, 30, 31, 33, 35, 38	Variable PVDAI and variable cost	It can be used when testing broader frequency improvements with less predictable outcomes is acceptable
4	30, 33, 35, 38	Lowest cost among scenarios, with variable PVDAI improvements	It can be used when budget is the primary concern
5	B43, 35, 38, 46	Highest potential PVDAI improvement but highest cost	It can be used when the goal is to prioritize maximum equity and access gains even at higher operating cost
6	B43, 35, 38, 46	Moderate cost with relatively strong PVDAI improvements	It can be used when the goal is to have targeted improvements with lower cost than the highest-upside scenario

Table 8 Sensitivity analysis of PVDAI success threshold

Metric	Q75	Q80	Q90
PVDAI cutoff value	–642	–634	–631
Number of clusters	6	3	2
Satisficing solutions (N)	108	74	42
Maximum access across scenarios	44394.30	44394.30	44394.30
Minimum cost across scenarios ($)	1689.73	1704.35	1704.35
	Operational comparison		
Routes appear significant across scenarios	B43, 35, 38	B43, 46	B43, 30, 31, 35, 38

(i.e., transit route headways), access is increased by 23% in scenarios 1–4 and 6 and by 15% in scenario 5. All the sets of solutions improve the Poverty Density Accessibility Index given that the existing index is at −660.67, with the highest level of improvement appearing in Scenarios 1 and 5 (−626.50); see Table 6. In terms of cost, all six scenarios have similar cost ranges, with scenario 3 and 6 providing a wider cost range compared to the other scenarios, and scenario 4 providing the tightest range of cost values and the lowest cost among all, as shown in Fig. 13.

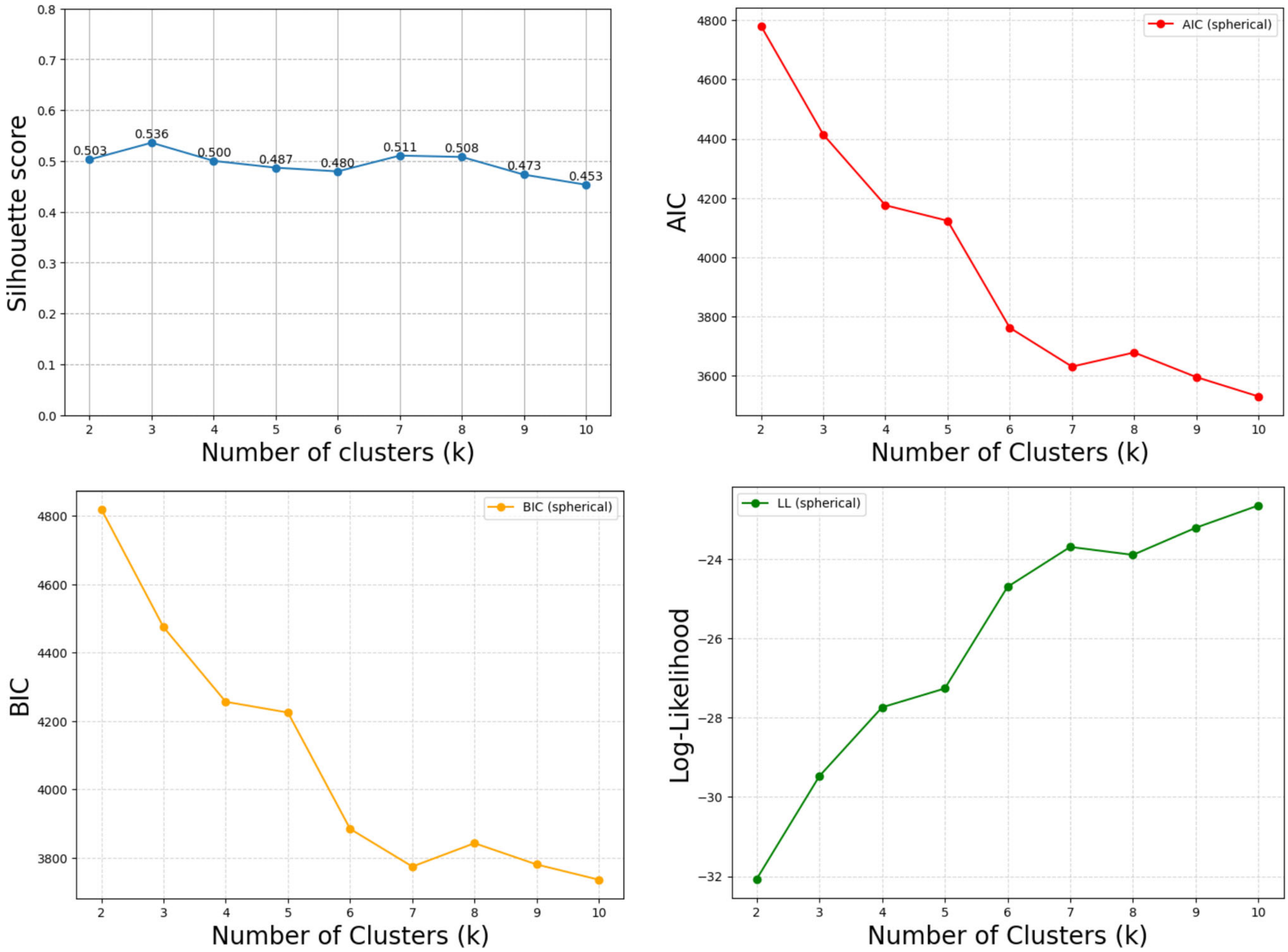


Fig. 16 Tested scores for selecting the optimal number of clusters when using PVDAI threshold of the 80th percentile.

Lastly, Fig. 12 shows the distribution of Poverty Density Accessibility Index by scenario. The variation across scenarios suggests that success (i.e., access improvements for areas of persistent poverty) can emerge under different PVDAI values, which is valuable for policy and operational interpretation. The presence of tight distributions (scenario 2) alongside wider ones (scenario 3, 5 and 6) hints at distinct success mechanisms, some tightly defined, while others being more flexible. Scenario 2 exhibits a tightly grouped PVDAI distribution, suggesting consistent access values among successful scenarios. In contrast, scenarios 3, 5 and 6 show wider spreads, indicating that success can occur under more varied access values.

Conclusion

This research develops an optimization model that maximizes the proposed Poverty Density Accessibility Index (PVDAI) by optimizing transit route headways subject to cost constraints. The optimization framework is applied to Amherst, Massachusetts, which is served by Pioneer Valley Transit Authority. Maximizing the PVDAI, which is a metric that highlights areas of high poverty density and low access to specific destinations, and in the case of this research healthcare and social assistance jobs, ensures that any proposed changes in transit routes account for the socioeconomic status of the people in the area. This is the first study that proposes an optimization model that improves access by solely optimizing transit route headways.

The study contributes to the literature by incorporating the optimization model into a scenario discovery framework, providing insights that traditional optimization approaches that target unique solutions can miss. Another significant contribution of this study is that the scenario discovery framework explicitly incorporates equity: all six scenarios are constructed not by simply maximizing access, but by maximizing the proposed Poverty Density Accessibility Index, therefore, prioritizing access for populations who need it most. Transit access is highly sensitive to operational factors

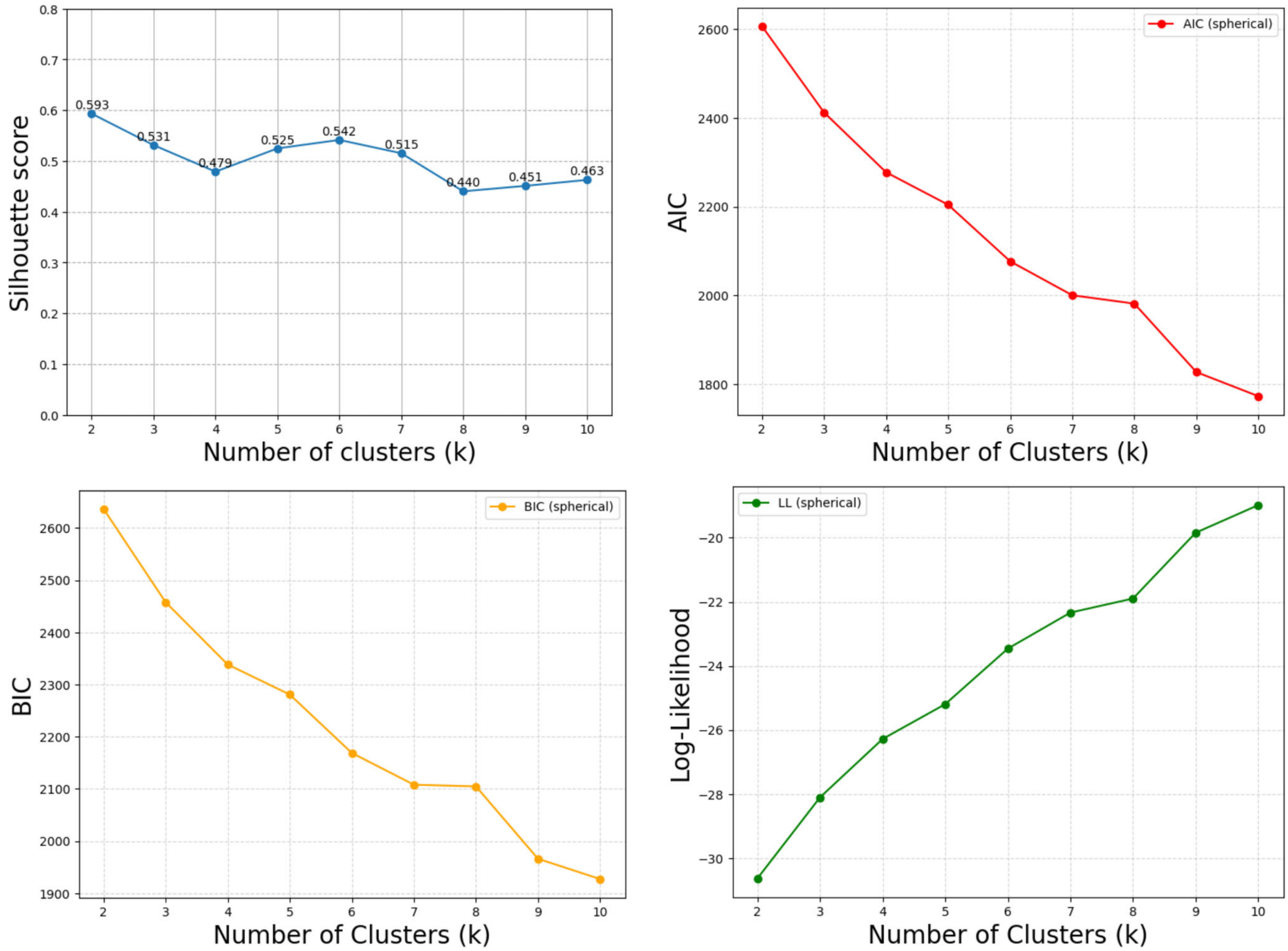


Fig. 17 Tested scores for selecting the optimal number of clusters when using PVDAI threshold of the 90th percentile

such as headways, runtimes, transfer reliability, and the connections between routes. Because these elements fluctuate in real-world systems, scenario discovery becomes essential for understanding how accessibility responds when any of these inputs shift. Rather than relying on a single optimized solution that reflects only one set of assumptions, scenario discovery allows decision makers to evaluate multiple plausible operating conditions by revealing how access changes across them. This is particularly important because access is a multidimensional outcome shaped by several interdependent components. In addition, the developed optimization framework is flexible in that it can account for multiple types of destinations and assign different weights based on the needs of a selected area. The proposed accessibility index, which incorporates poverty density, is also adaptable and can include any socioeconomic characteristic. This allows the model to maximize access based on different priorities, for example using population density or other socioeconomic criteria instead of poverty density.

Implementation of this optimization framework in Amherst showed that specific routes are important in all the scenarios and beneficial when operating under low or moderate headways (i.e., route B43), while other routes affect the improvement of PVDAI in some scenarios but not in others (i.e., route 46). In addition, there is a wide range of combinations of access improvements at poverty areas (i.e., PVDAI) and cost with trade-offs between improvements and cost (e.g., scenario *Variable Access and High Cost*). Therefore, the proposed methodology can guide transit agencies in identifying the routes whose schedule should be altered (i.e., by implementing low, moderate, or high headways) to advance their goals of increasing access to destinations in vulnerable areas, while understanding how costs differ across policy options.

The framework is transferable in larger network, taking into account some important considerations. First, the number of decision variables would increase substantially because many more routes, headways, transfer points, and service periods would need to be considered. As a result, the number of possible headway combinations would grow quickly, and a larger scenario sample would be needed to adequately cover the decision space. Second, transfer interactions would become more important. In Amherst, the analyzed network is relatively small, so accessibility changes can be interpreted mainly through route-level headway changes. In a larger network, however, changing the headway of one route may affect accessibility through transfer waiting times, missed connections, and network connectivity, in addition to affecting waiting times. Finally, instead of clustering all routes at once, it may be useful to group routes by corridor, service type, or subarea, or to first identify the routes with the strongest influence on accessibility.

Limitations of this study are related to estimating access using constant speeds throughout the day since those are the only speeds currently available in Conveyal. As a result, congestion and its impact on travel times and therefore, accessibility is not considered. However, the selected case study is in a rural area with low congestion and therefore, it is anticipated that the results are not substantially affected. In addition, the accessibility values are modeled outputs from Conveyal under hypothetical GTFS headway scenarios, not observed ridership or observed travel times. Similarly, the prediction of the PVDAI relies on the XGBoost modeling, not real-world data.

Access to several destination types, including manufacturing accommodation-food-retail jobs, healthcare jobs, construction jobs, supermarkets, healthcare services, and childcare centers, as well as multiple time periods, including wide span of 8 am–8 pm and 8 pm–12 am during weekdays and weekends has been assessed as part of this study. That broader screening showed that some areas repeatedly exhibited very low, i.e., negative PVDAI values across multiple destination types and timeframes. Amherst was one of these consistently vulnerable areas, which motivated its selection as the case study for the headway optimization framework. Future work will focus on accounting for multiple destinations of interest and timeframes as extra components in the objective function, while expanding the model by adding cost directly in its objective function subject to access thresholds. This will allow transit agencies to control directly for cost-effective solutions toward equitable access to key destinations while ensuring that any access changes are within acceptable thresholds.

Author Contributions The authors confirm their contribution to the paper as follows: study conception and design: EK, TO, JO, EC; data collection: EK; methodological implementation: EK; analysis and interpretation of results: EK, TO, JO, EC; draft manuscript preparation: EK, TO, JO, EC. All the authors reviewed the results and approved the final version of the manuscript.

Funding This research was partially supported by funding from the Pioneer Valley Transit Authority (PVTA). The authors gratefully acknowledge this support.

Data Availability Available with restricted access: The data that were generated for the current study required Conveyal license and so are not publicly available. The data are, however, available upon request and with the permission of the agency that provided the Conveyal license to the corresponding author.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval This study did not require ethical approval as it did not involve any human or animal subjects.

References

Allen J, Farber S (2020) Planning transport for social inclusion: an accessibility-activity participation approach. Transp Res Part D: Transp Environ 78:102212. https://doi.org/10.1016/j.trd.2019.102212

American Public Transportation Association. (2025). Public Transportation Facts. Accessed: 2024-10-03 https://www.apta.com/news-publications/public-transportation-facts/

Åslund O, Östh J, Zenou Y (2010) How important is access to jobs? old question-improved answer. J Econ Geogr 10(3):389–422. https://doi.org/10.1093/jeg/lbp040

Barboza MH, Carneiro MS, Falavigna C, Luz G, Orrico R (2021) Balancing time: using a new accessibility measure in rio de janeiro. J Transp Geogr 90:102924

Boisjoly G, El-Geneidy A (2016) Daily fluctuations in transit and job availability: a comparative assessment of time-sensitive accessibility measures. J Transp Geogr 52:73–81. https://doi.org/10.1016/j.jtrangeo.2016.03.004

Camporeale R, Caggiani L, Fonzone A, Ottomanelli M (2017) Quantifying the impacts of horizontal and vertical equity in transit route planning. Transp Plan Technol 40(1):28–44. https://doi.org/10.1080/03081060.2016.1238569

Ceylan H, Ozcan T (2018) Optimization of headways and departure times in urban bus networks: a case study of corlu, turkey. Adv Civil Eng 2018:7094504. https://doi.org/10.1155/2018/7094504

Cheng J, Bertolini L (2013) Measuring urban job accessibility with distance decay, competition and diversity. J Transp Geogr 30:100–109

Conveyal. Options and Configuration. Conveyal User Manual; accessed 2026-05-15 (2026). https://docs.conveyal.com/analysis/configuration

Conveyal. R5: Rapid realistic routing on real-world and reimagined networks. GitHub repository; accessed 2026-05-15 (2026). https://github.com/conveyal/r5

Conveyal. R5 (2024). https://github.com/conveyal/r5

Conveyal. R5 Source Code: ProfileRequest.java. Source code; accessed 2026-05-15 (2026). https://raw.githubusercontent.com/conveyal/r5/dev/src/main/java/com/conveyal/r5/profile/ProfileRequest.java

Conveyal. Timetables. Conveyal User Manual; accessed 2026-05-15 (2026). https://docs.conveyal.com/edit-scenario/timetable

Deboosere R, El-Geneidy A (2018) Evaluating equity and accessibility to jobs by public transport across canada. J Transp Geogr 73:54–63. https://doi.org/10.1016/j.jtrangeo.2018.10.006

U.S. Department of Health and Human Services. (2025). Poverty Guidelines. Accessed: 2025-03-27 https://aspe.hhs.gov/topics/poverty-economic-mobility/poverty-guidelines

Ding S, Yuan C, Chen J, Zhang C, Feng J (2023) Accessibility optimization of public transportation in historical districts-a study of bei'lin district, xi'an. CICTP 2023:330–339. https://doi.org/10.1061/9780784485040.033

El-Geneidy A, Levinson D, Diab E, Boisjoly G, Verbich D, Loong C (2016) The cost of equity: assessing transit accessibility and social disparity using total travel cost. Transp Res Part A Policy Pract 91:302–316. https://doi.org/10.1016/j.tra.2016.07.003

Ermagun A, Tilahun N (2020) Equity of transit accessibility across chicago. Transp Res Part D: Transp Environ 86:102461. https://doi.org/10.1016/j.trd.2020.102461

Foth N, Manaugh K, El-Geneidy AM (2013) Towards equitable transit: examining transit accessibility and social need in toronto, canada, 1996–2006. J Transp Geogr 29:1–10. https://doi.org/10.1016/j.jtrangeo.2012.12.008

Gayashani HS, Jayasooriya N, Kularathna AHTS (2026) Optimization of vehicle headway and capacity for cost minimization in feeder vehicle services for transit systems in developing countries. Transp Plan Technol. https://doi.org/10.1080/03081060.2026.2641775

Guzman LA, Oviedo D, Rivera C (2017) Assessing equity in transport accessibility to work and study: the bogotá region. J Transp Geogr 58:236–246. https://doi.org/10.1016/j.jtrangeo.2016.12.016

Hansen WG (1959) How accessibility shapes land use. J Am Inst Plann 25:73–76. https://doi.org/10.1080/01944365908978307

Hernandez D (2018) Uneven mobilities, uneven opportunities: social distribution of public transport accessibility to jobs and education in montevideo. J Transp Geogr 67:119–125. https://doi.org/10.1016/j.jtrangeo.2017.08.017

Hernandez D, Hansz M, Massobrio R (2020) Job accessibility through public transport and unemployment in latin america: the case of montevideo (uruguay). J Transp Geogr 85:102742

Herszenhut D, Pereira RHM, Silva Portugal L, Sousa Oliveira MH (2022) The impact of transit monetary costs on transport inequality. J Transp Geogr 99:103309. https://doi.org/10.1016/j.jtrangeo.2022.103309

Hess DB (2005) Access to employment for adults in poverty in the buffalo-niagara region. Urban Stud 42(7):1177–1200. https://doi.org/10.1080/0042098050012138

IBM. What is XGBoost? Accessed: 2026-01-03 (2025). https://www.ibm.com/think/topics/xgboost

Kawabata M, Shen Q (2006) Job accessibility as an indicator of auto-oriented urban structure: a comparison of boston and los angeles with tokyo. Environ Plann B Plann Des 33:115–130. https://doi.org/10.1068/b31144

Li Y, Fan DW (2021) Optimizing transit equity and accessibility of the city of charlotte, north carolina, by integrating transit gap index, a general transit feed specification data-relevant performance metric. J Transp Eng Part A Syst 147(4):04021005. https://doi.org/10.1061/JTEPBS.0000503

Lindner A, Kühnel F, Schrömbges M, Kuhnimhof T (2024) When to measure accessibility? temporal segmentation and aggregation in location-based public transit accessibility. Urban Sci. https://doi.org/10.3390/urbansci8040165

Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems, 30 (2017). https://papers.neurips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions

McCahill C, Jain S, Brenneis M (2020) Comparative assessment of accessibility metrics across the us. Transp Res Part D: Transp Environ 83:102328. https://doi.org/10.1016/j.trd.2020.102328

McKay MD, Beckman RJ, Conover WJ (2000) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 42(1):55–61. https://doi.org/10.1080/00401706.2000.10485979

Merlin LA, Hu L (2017) Does competition matter in measures of job accessibility? explaining employment in los angeles. J Transp Geogr 64:77–88. https://doi.org/10.1016/j.jtrangeo.2017.08.009

Conveyal. Baseline Network. Conveyal User Manual; accessed 2026-05-15 (2026). https://docs.conveyal.com/prepare-inputs

Organisation for Economic Co-operation and Development (2015) In it together: Why less inequality benefits all.OECD Publishing. https://doi.org/10.1787/9789264235120-en

Osland L (2010) Spatial variation in job accessibility and gender: an intraregional analysis using hedonic house-price estimation. Environ Plan A 42:2220–2237. https://doi.org/10.1068/a42422

Park SJ, Kang S, Byon Y-J, Kho S-Y (2022) Multiobjective approach to the transit network design problem with variable demand considering transit equity. J Adv Transp 2022:5887985. https://doi.org/10.1155/2022/5887985

Pritchard JP, Tomasiello D, Giannotti M, Geurs K (2019) An international comparison of equity in accessibility to jobs: London, são paulo and the randstad. Transport Findings. https://doi.org/10.32866/7412

Reggiani A, Bucci P, Russo G, Haas A, Nijkamp P (2011) Regional labour markets and job accessibility in city network systems in germany. J Transp Geogr 19(4):528–536. https://doi.org/10.1016/j.jtrangeo.2010.05.008

Ruiz M, Segui-Pons JM, Mateu-LLadó J (2017) Improving bus service levels and social equity through bus frequency modelling. J Transp Geogr 58:220–233. https://doi.org/10.1016/j.jtrangeo.2016.12.005

scikit-learn developers. Silhouette Coefficient. scikit-learn documentation (2024). https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html

SciPy Community. scipy.stats.qmc.LatinHypercube – SciPy v1.16.1 Manual. Accessed: 2026-01-03; documentation for SciPy version 1.16.1 (2025). https://docs.scipy.org/doc/scipy-1.16.1/reference/generated/scipy.stats.qmc.LatinHypercube.html

Shen Q (1998) Location characteristics of inner-city neighborhoods and employment accessibility of low-wage workers. Environ Plann B Plann Des 25(3):345–365. https://doi.org/10.1068/b250345

Slovic AD, Tomasiello DB, Giannotti M, Fatima Andrade M, Nardocci AC (2019) The long road to achieving equity: job accessibility restrictions and overlapping inequalities in the city of São Paulo. J Transp Geogr. https://doi.org/10.1016/j.jtrangeo.2019.06.003

Slovic AD, Tomasiello DB, Giannotti M, Fatima Andrade M, Nardocci AC (2019) The long road to achieving equity: job accessibility restrictions and overlapping inequalities in the city of são paulo. J Transp Geogr 78:181–193

Stepniak M, Goliszek S. (2017) Spatio-temporal variation of accessibility by public transport–the equity perspective, 241–261 https://doi.org/10.1007/978-3-319-45123-7_18

Tomasiello DB, Giannotti M, Feitosa FF (2020) Access: an agent-based model to explore job accessibility inequalities. Comput Environ Urban Syst 81:101462. https://doi.org/10.1016/j.compenvurbsys.2020.101462

Via Transportation, Inc. Remix Public Transit Planning Software. Accessed: 2026-01-03 (2025). https://ridewithvia.com/solutions/remix/planning

Wang CH, Chen N (2015) A gis-based spatial statistical approach to modeling job accessibility by transportation mode: case study of columbus, ohio. J Transp Geogr 45:1–11. https://doi.org/10.1016/j.jtrangeo.2015.03.015

Wang J, Kwan MP, Xiu G (2024) Assessing the fluctuations in job accessibility under travel time uncertainty. Appl Geogr 167:103296. https://doi.org/10.1016/j.apgeog.2024.103296

Weintraub A. (2018). An unhealthy commute: the transit challenges facing new york city's healthcare sector. Technical report, Center for an Urban Future Report examining transit barriers and commutes of healthcare workers in New York City. https://nycfuture.org/research/an-unhealthy-commute

Yan X, Bejleri I, Zhai L (2022) A spatiotemporal analysis of transit accessibility to low-wage jobs in miami-dade county. J Transp Geogr 98:103218

Zhong Q, Karner A, Kuby M, Golub A (2019) A multiobjective optimization model for locating affordable housing investments while maximizing accessibility to jobs by public transportation. Environ Plan B Urban Anal City Sci 46(3):490–510. https://doi.org/10.1177/2399808317719708

Zhu L, Shi F (2022) Spatial and social inequalities of job accessibility in kunshan city, china: application of the amap api and mobile phone signaling data. J Transp Geogr 104:103451. https://doi.org/10.1016/j.jtrangeo.2022.103451