Uber ML Hackathon 2019:
Uber as a Platform

Problem Statement & Scoring Document

Uber

ITSBerkeley
Institute of Transportation Studies

BERKELEY LAB
Lawrence Berkeley National Laboratory

Executive Summary

The inaugural Uber Prize, to be launched in the Spring 2019 is a machine learning transportation challenge intended to create plausible, implementable and useful insights for cities to better understand the challenges of their transportation network to improve urban mobility. The inaugural Uber Prize harnesses advances in the machine learning and data science fields to enable data-driven transportation demand modeling and policy analysis at a resolution and scale that can empower and engage city officials, transportation system managers, the private sector, academics, and citizens to understand, analyze, and collaboratively plan for the rapidly evolving transportation realities shaping urban areas worldwide. Contestants will be challenged to design a set of changes to existing transportation systems that bring about the greatest improvements across important indicators of transportation system performance. To do this, they will be provided a transportation modeling and data analytics platform, BISTRO, based on BEAM, an agent based modeling software developed at the *Lawrence Berkeley National Laboratory* (LBNL).

The Uber as a Platform, Transportation Policy Optimization ML hackathon will be used to inform the *Uber Prize*. For this hackathon (and Phase I launch when the Uber Prize is launched), contestants are asked to optimize the transportation network for a sample of citizens from a mock city: Sioux Faux (a virtual instantiation of Sioux Falls, a real city in South Dakota). The city's 157,000 citizens travel between activities using either their personal automobiles, buses provided via a public transit system, on-demand rides, walking, or a combination of multiple modes in accordance with their preferences. To participate, contestants will optimize a set of **inputs** that represent possible city-wide policies regarding transportation: policies that control *operational* and *financial* aspects of mass transit as well as *incentives* that influence the use of mass transit and on-demand ride services. An agent-based simulation of the city takes in the contestant-defined inputs and produces a set of **output** metrics and **score components** that evaluate system-wide level of service, congestion, environmental sustainability, and the net costs of the policy interventions. The metrics evaluated from contestant submissions will be judged in comparison to the corresponding metrics from the **business as usual**, or *do nothing*, scenario. The **total score** for a contestant submission is computed as the average of the standardized score components.

To read about the problem statement, **Part I** of this document gives a full description of the models of Sioux Faux, the typical control inputs at the disposal of the city (hence the competitors), their effects on the city and the corresponding outputs. The outputs (to be computed numerically by BISTRO) are used in the scoring of the solutions of the competitors. Part I is important to understand how the Uber Prize will be of relevance for real-world planning. While it is recommended to read it entirely, some parts can be skipped upon first reading, to get started faster.

To get started, competitors can use the Starter Kit on GitHub.[1] To facilitate participation, the Uber Prize Working Group, a team of Uber employees, Berkeley researchers and doctoral students, and LBNL scientists have provided a suite of resources for contestants.

Scoring and judging is explained in detail in **Part II** of this document. This part provides several scenarios that can be used by the contestants as a starter point. In particular, a *business as usual* (BAU) scenario provides the current modus operandi of Sioux Faux, and can be considered as neutral. Solutions with scores better than BAU (which improve Sioux Faux) are presented. In particular corresponding inputs and outputs are plotted, along with interpretations in terms of policies for the city, to help contestants gain insights on their solutions.

Winners and prizes will be determined by a panel of judges based on a combination of (1) **submission scores**, (2) **presentations** (to the panel), and (3) **citizenship** (contributions to the Uber Prize by hacks, tutorials, or feedback). This is explained in the last section of Part II of this document, which provides a detailed and quantitative description of the judging criteria and corresponding rubrics. The **three winning teams** will be invited for a dinner with Dara. The **First Prize** winning team will be awarded $10,000.

Good luck, have fun, and help Uber and the world improve future urban mobility by accelerating the fusion of machine learning, optimization, data science, modeling and simulation!

[1] https://github.com/vgolfier/Uber-Prize-Starter-Kit

Part I

Problem Statement

PART I

Uber Prize: Sioux Faux Problem Statement

Jessica Lazarus*†, Teddy Forscher*†, Sidney Feygin†, Valentine Golfier-Vetterli†, Abhishek Gupta†, Andrew Salzberg†, Emily Strand†, Rashid Waraich†‡, Colin Sheppard‡, Alexandre M. Bayen*†‡

* Institute of Transportation Studies, University of California, Berkeley, US

† Uber, Inc., US

‡ Lawrence Berkeley National Laboratory, US

February 19, 2019

Contents

1 Introduction

The inaugural *Uber Prize* harnesses advances in the machine learning and data science fields to enable data-driven transportation demand modeling and policy analysis at a resolution and scale that can empower and engage city officials, transportation system managers, the private sector, academics, and citizens to understand, analyze, and collaboratively plan for the rapidly evolving transportation realities shaping urban areas worldwide.

Uber Prize contestants will envision a set of changes to existing transportation systems that bring about the greatest improvements across important indicators of transportation system performance in terms of system-wide level of service, congestion, and sustainability.

The first round of the Uber Prize challenges contestants to optimize the transportation network of a benchmark city: Sioux Faux. The city's 157,000 citizens travel between activities using either their personal automobiles, buses provided via a mass transit system, on-demand rides, walking, or a combination of multiple modes in accordance with their preferences. Considering the overall transportation system costs and revenues in Sioux Faux, contestants will compete to produce the best outcomes given the trade-offs between the metrics defined in Section 5.3.

2 Background

2.1 Related work

As the available transportation modes are proliferating and changing on a monthly basis or faster (e.g., micromobility, shared mobility, automated mobility, etc.), understanding the behavioral responses of users and the subsequent transportation outcomes merit investigation. Forward-looking studies have begun to look at the possible transportation and environmental benefits offered by future and emerging shared modes, such as work conducted by the International Transport Forum[1] and the Lawrence Berkeley National Laboratory.[2]

In the U.S. context, many *metropolitan planning agencies* (MPOs) are equipped to analyze the long-range transportation effects of new modes in their regions, but just as many are not. The purpose of the *Uber Prize* is to provide a platform on which contestants can envision near-term implementable operational changes to a region's transportation system. These changes are intended to complement existing operational and infrastructure expenditure processes. Additionally, the outputs of the platform can provide a basis for discussions across many sets of stakeholders.

2.2 Agent-Based Simulation

Agent-based simulation of travel demand, is a method by which to evaluate the network-wide effects of modifications to a transportation system. Agent-based travel demand microsimulation realizes the daily activity schedules and transportation choices of a socio-demographically heterogeneous population of citizens on a virtual representation of physical road networks.[3] This methodology enables an informative resolution of feedback loops and spatio-temporal constraints operating between travel purposes, road network congestion, household vehicle availability, and the levels of service provided by infrastructure and available transportation modes.

Person agents represent simulated individuals who make decisions about what transportation mode(s) to use to travel to and from their daily activities. During the simulation, person agents make one or more *tours* of travel to sequential activities, starting and ending each tour at home. Each *trip* in a tour represents travel from one activity to the next. Trips may consist of one or more *legs* of travel, each using a particular *mode* of transportation. A *mode choice model* characterizes the transportation mode preferences of agents by accounting for the sensitivity of the agent to the attributes of each alternative, such as wait time, in-vehicle travel time, and trip cost.[4] The simulator uses a realistic representation of the transportation network and

[1] For more information, see: https://www.itf-oecd.org/sites/default/files/docs/shared-mobility-liveable-cities.pdf

[2] For more information, see https://www.nature.com/articles/nclimate2685

[3] While the population and its plans are synthetic, econometric modelling techniques using census data and travel surveys together with calibration against observed mode splits and network volumes ensure that the simulation represents typical daily traffic conditions.

[4] For more information about mode choice models, see: https://eml.berkeley.edu/books/choice2.html

a *routing algorithm* to determine the generalized cost of routing vehicles on the network as a function of the expected travel-time on links taking into account congestion (i.e., movement slower than the maximum allowable speed due to the number of vehicles on a link exceeding capacity).

The inputs to one instance of the simulation include a representative population of synthetic agents together with their typical daily activity plans. A virtual road network, transit schedule, and parking infrastructure define the transportation supply. The simulation proceeds iteratively: evaluating the plans on the physical network and then permitting agents to replan components of their schedule in response to the generalized costs of travel. Once agents have settled on a set of plans that collectively maximize the average utility of their set of evaluated plans, the simulation engine reaches a fixed point or equilibrium condition.

Each simulation run produces outputs of the actual paths and travel times realized by each person agent and each vehicle, as well as a host of other data, further detailed in Section 4. In practice, outputs of agent-based simulations may be used to communicate policy alternatives to stakeholders. For example, visualizations of congested roadways with millions of agents behaving independently can provide a concise method to communicate the effects of infrastructure interventions.

Agent-based simulation allows for the evaluation of counterfactual *scenarios*. A scenario is a simulation that implements a unique set of circumstances that differs in some way from a *base case*. The base case is calibrated using data representing the current state of the transportation system being simulated. Examples of scenarios include alteration of the population configuration representing population or employment growth, alteration of the transportation network such as unexpected road network restrictions due to sporting events, inclement weather or traffic accidents, as well as the introduction of new modes of transportation such as autonomous vehicles. Well-calibrated simulations of transportation systems, such as those just described, allow stakeholders to better understand the implications of policy proposals in hypothetical travel environments.

2.3 BISTRO

The *Berkeley Integrated System for Transportation Optimization* (BISTRO) is the engine through which the *Uber Prize* will be run. BISTRO is an analysis and evaluation framework that works in concert with an agent-based simulation: *Behavior, Energy, Autonomy, and Mobility* (BEAM) to enable the open-sourced development and evaluation of transportation optimization methods in response to given policy priorities. For more information on BEAM, see https://github.com/vgolfier/Uber-Prize-Starter-Kit/blob/master/docs/Introduction_transportation_problem.md#what-is-beam.

2.4 Sioux Faux

This round of the *Uber Prize* challenges contestants to use ABS to optimize the transportation network of the benchmark city Sioux Faux. Like many cities around the world, Sioux Faux is urbanizing rapidly. All of this growth has led to a precipitous rise in congestion. Roads that were once pathways to jobs and opportunity are now clogged with idling vehicles polluting the air. Fauxians are in need of improved mobility– particularly given the population expansion predicted to take place over the next two decades. The city has put out a call for support, and is engaging the foremost minds in the fields of transportation and data science to help them understand the problem, explore solutions, and set this booming metropolis back on the path to sustainable growth.

2.4.1 Sioux Falls, South Dakota

Sioux Falls is South Dakota's largest city with a population of about 187,200. It has an estimated metropolitan area population of 259,094 and urban area population of 156,777. Sioux Falls is the 47th fastest growing city in the United States and covers 73.47 square miles, of which 72.96 square miles is land and 0.51 square miles is water. It is located near the Minnesota border, at the junction of Interstates 90 and 29. Within the city, there are several taxi companies operating, as well as local public transit, Sioux Area Metro, that runs 16 bus lines. The city also has long-distance bus routes to Minneapolis, Omaha, and Kansas City. Sioux Falls is conveniently located and provides easy access to most Midwestern cities, hence its motto, "the Heart of America." The rest of this document, while inspired by the real city of Sioux Falls, is only a numerical model of a virtual city, *Sioux Faux*;[5] the conclusions and outcomes of the Hackathon are thus not meant to influence the policy or politics of the city of Sioux Falls.

(a) Sioux Falls City Hall *(b) Falls Park, Sioux Falls* *(c) Sioux Falls BRT*

Figure 1: Sioux Falls, South Dakota.

2.4.2 Population synthesis

In order to simulate the activity and travel behavior of the citizens of Sioux Faux, a population of virtual agents and households was generated such that the sociodemographic attributes of these virtual entities are spatially distributed in accordance with real-world census data. In order to provide realistic distributions of household and individual attributes for Sioux Faux, we expanded publicly-available survey data for the city of Sioux Falls, South Dakota.[6]

In practice, we use the Doppelganger library[7], a state-of-the-art population synthesis framework developed in Python. Specific inputs to Doppelganger used to generate the Sioux Faux population included household and individual Public Use Microdata Sample (PUMS) data for South Dakota from the 2012-2016 (5-year) American Community Survey (ACS), which is conducted annually by the US Census. [8] The Public Use Microdata Area (PUMA) for Sioux Falls constrains the state-wide survey data to our general area of interest. The PUMS data as well as a set of configurable parameters are used to train a set of Bayesian networks, which are structured such that drawing samples from these networks reproduces empirically-observed relationships between attributes that define household and individuals. The population-level characteristics of synthetic agents and households are further constrained using convex optimization to match summary statistics (also known as *marginals*) from other census data aggregates for the Sioux Falls PUMA. Household attributes that act as controls in the allocation optimization algorithm include the size of households (number of residents),

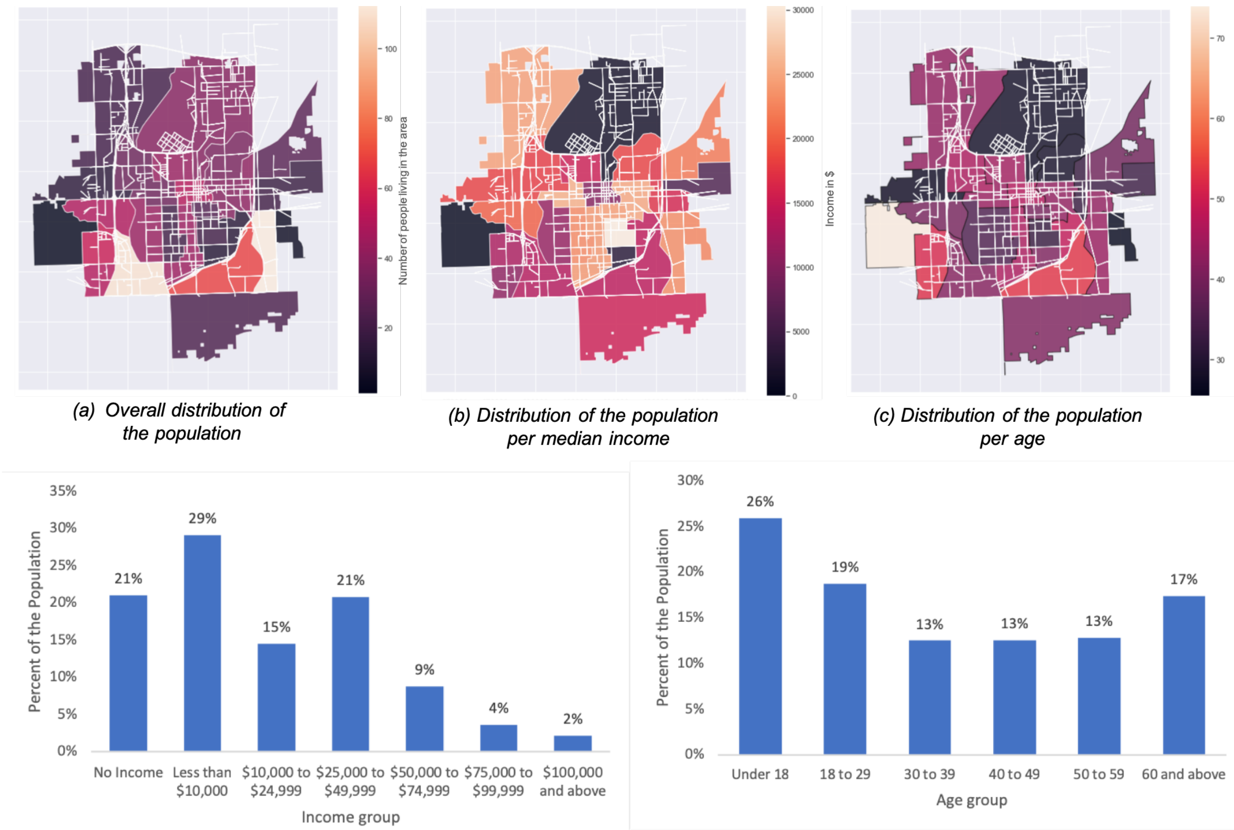
[5] We have made the distinction, to be clear, "faux" in French meaning "not real."

[6] The commonly referred "Sioux Falls" scenario has often been used as a benchmark in agent-based simulation research, hence its use in the initial phase of the *Uber Prize*, see https://github.com/bstabler/TransportationNetworks/tree/master/SiouxFalls

[7] For more information about the Doppelganger library, see https://github.com/sidewalklabs/doppelganger

[8] The 5-year PUMS comprises a 5% sample of the US population. It is computed as an aggregate of 1-year samples, which themselves aim to survey 1% of the US population

number of vehicles per household, and household income. Similarly, individual marginals include age, gender, and income.



(a) Overall distribution of the population

(b) Distribution of the population per median income

(c) Distribution of the population per age

(d) Income distribution

(e) Age distribution

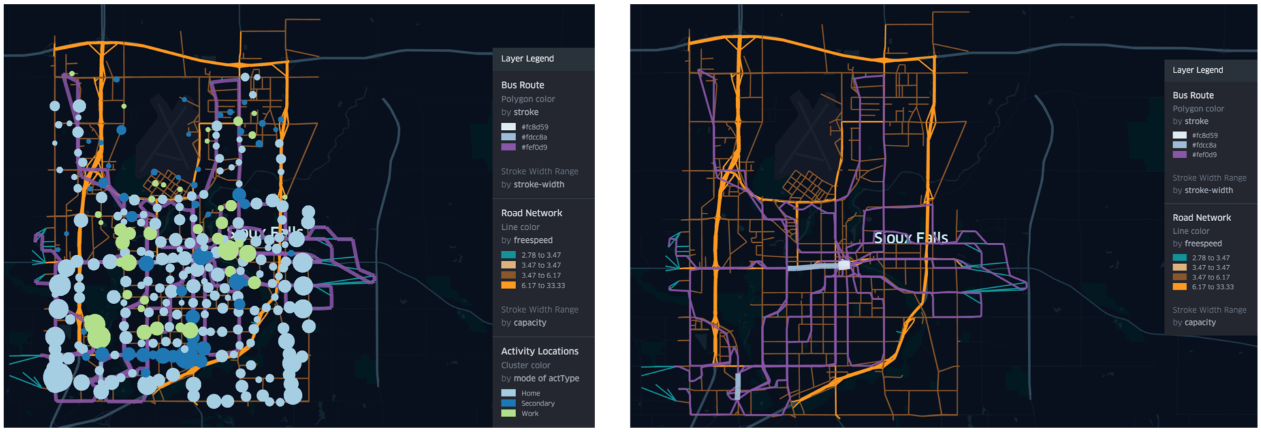
Figure 2: Demographics of Sioux Faux (per census tract).

2.4.3 Daily activity-travel plan synthesis

An existing set of agent plans for Sioux Falls developed for MATSim simulations was used as the basis for our expanded Sioux Faux population. For more information on the Sioux Falls scenario, see `https://www.ethz.ch/content/dam/ethz/special-interest/baug/ivt/ivt-dam/vpl/reports/901-1000/ab978.pdf`. After initial pilot runs and evaluations to determine trade-offs between population size, behavioral realism, and computational complexity, we took a 15% sub-sample of the full synthetic population (approximately 15,000 agents). We used a spatially-constrained sampling mechanism in order to allocate plans to agents in accordance with household locations and census tract household and individual attribute distributions. The subsampling mechanism also enforces logical assumptions such as "agents under the age of 18 should not have a work activity" and "agents under the age of 16 should not be allowed to drive".

2.4.4 Transportation network configuration

The transportation network for Sioux Faux is derived from the Open Street Maps (OSM) data regarding the street network of Sioux Falls, South Dakota. The data provides the configuration of the road network in the form of links (road segments) and nodes, which are encoded as a directed graph in BEAM. The physical dynamics of each link are determined based on the road type classification provided by the OSM data (see Table 1). These include the length (in meters), the number of lanes per direction of travel, the free-flow travel speed, and the capacity (in vehicles per hour) of each lane for every link[9].



(a) Sioux Faux' activities locations *(b) Sioux Faux' road network*

Figure 3: Activities locations and road network in Sioux Faux.

The information regarding the bus routes in Sioux Faux is presented in the Generalized Transit Feed Specification (GTFS) format. GTFS characterizes scheduled transit movements in a standardized fashion, using linked text files to encode agency operation characteristics such as transit stops, routes, trips, fares, and other schedule data. GTFS can also accommodate frequency-based dispatching as opposed to scheduled arrival and departure times from stops; for more information on this type of dispatching, refer to Section 3 of the Starter Kit Input Schema.

The final component of the transportation network configuration is the allocation of residential, public, and workplace parking facilities across Traffic Analysis Zones (TAZ). For the purposes of this competition, there is unlimited parking in Sioux Faux. Person agents seek parking mid-trip, at some time during a leg of their trip. Agents will pick the cheapest and closest parking alternative to their destination with attributes which match their use case (e.g., public if parking at at a transit station or residential if going home).

2.4.5 Vehicle routing and parking

The PhysSim module within BEAM simulates traffic on the road network. The underlying simulation engine is based on the Java Discrete Event Queue Simulator (JDEQSim) from the MATSim framework. [10] The JDEQSim simulates traffic flow through the system using a queue-based model in which link travel times increase as the number of vehicles on the link increases. In each iteration of the simulation, JDEQSIM updates the routing engine with new network travel times for use in subsequent iterations.

BEAM uses the R5 routing engine to accomplish multi-modal routing. Agents' trip requests are input to the router, which computes the shortest path (based on travel time estimates from JDEQSim) for the corresponding mode(s) available to the agent for the trip. [11] The results of the routing calculation are then transformed into objects that are used as inputs to the mode choice model, described in the following subsection.

2.4.6 On-Demand ride services

[9]The values of the parameters used for each link classification type can be found in the MATSim documentation, see https://github.com/matsim-org/matsim/blob/0.10.x/matsim/src/main/java/org/matsim/core/utils/io/OsmNetworkReader.java

[10]see https://www.researchgate.net/publication/239925133_Performance_Improvements_for_Large_Scale_Traffic_Simula-_tion_in_MATSim

[11]For detailed information about the R5 router, see https://github.com/conveyal/r5 and https://doi.org/10.3141/2653-06

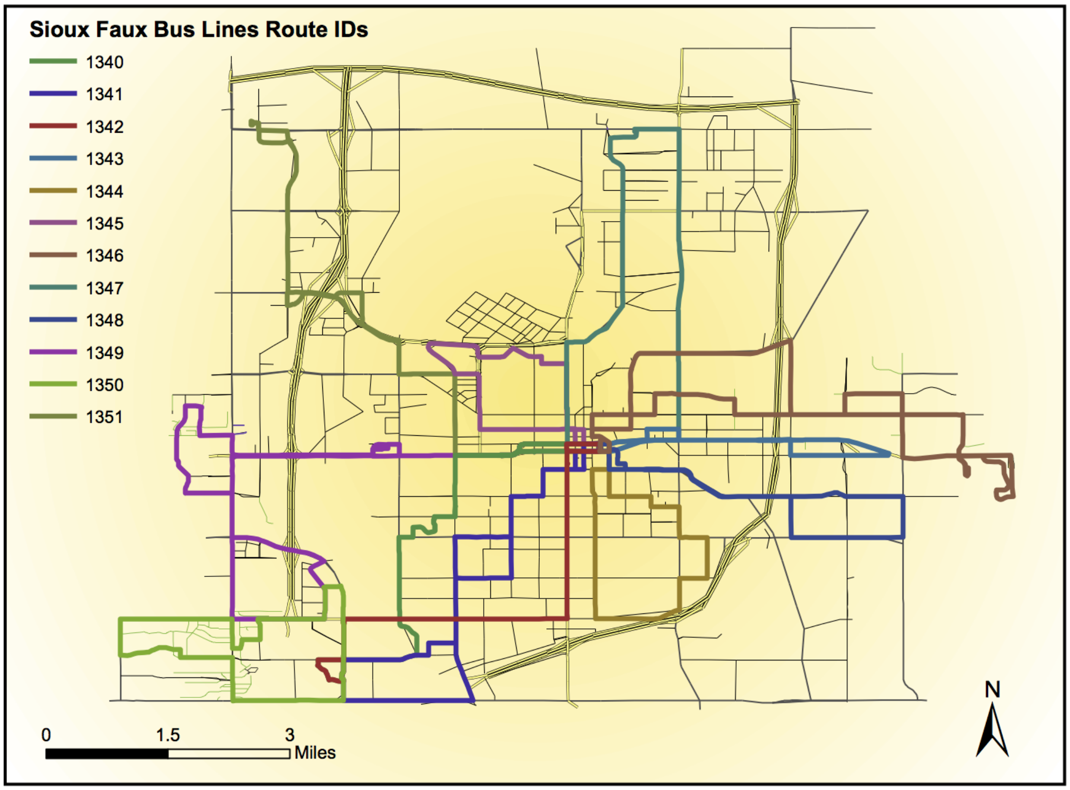


Figure 4: Map of the Sioux Faux Bus Lines (SFBL) Routes

Nomenclature To provide a usable working definition and to remain as consistent as possible with current academic[12] and industry[13] literature practices, for the purposes of the *Uber Prize*, on-demand for-hire automobility services are referred to in the following ways below:

- **On-demand rides** refer to the private use of a for-hire automobile (in which the user is being driven by a another individual), usually booked via an online platform or mobile app. While the number of riders and number of stops can vary, the trip is booked and controlled by one individual. Such modes have also been referred to as sequentially shared.

  To remain consistent with model configuration, technical documentation may refer to on-demand rides as **ridehailing**.

- **On-demand pooled rides** refer to the shared or pooled use of a for-hire automobile, usually booked via an online platform or mobile app. This is distinguished from on-demand rides as the trip is shared between multiple users who have each booked their own ride. Such modes have also been referred to as concurrently shared.

[12]For more information see: https://escholarship.org/uc/item/68g2h1qv#main
https://www.nap.edu/catalog/25020/private-transit-existing-services-and-emerging-directions
http://onlinepubs.trb.org/onlinepubs/sr/sr319AppendixA.pdf
https://escholarship.org/uc/item/46p6n2sk
[13]For more information see: http://apalosangeles.org/sidewalk-wars-the-infrastructure-of-micro-mobility/
https://www.populus.ai/micro-mobility-2018-july

Other forms of shared automobility, such as carpooling, incidental pooling (or casual carpooling), and ridesharing (or vanpooling)–which typically do not involve the repeated use of an online platform or mobile app–are not included with on-demand rides.

On-demand ride service operations On-demand ride service is delivered by a population of on-demand ride vehicles that are randomly distributed at the start of a simulation run. On-demand ride vehicles undergo three phases of service:

1. empty: the vehicle is not carrying a passenger and is available to be reserved for service;
2. fetch: the vehicle is reserved and is en-route to pickup a passenger;
3. fare: a passenger is in the vehicle traveling to its requested destination.

The price of on-demand rides is fixed, consisting of a distance-based and a duration-based component ($1.00/mile and $0.50 per minute, respectively for Sioux Faux). On-demand rides cannot be shared across multiple trips in the Sioux Faux scenario. On-demand pooled rides will be implemented in a later phase of the competition.

The size of the on-demand ride service's fleet is determined by a configuration parameter to be a proportion of the total number of agents in the simulation. In accordance with a parameter of 0.10, at the start of each simulation, a random sample of 1,500 agents is selected. Each on-demand driver agent begins each iteration of a simulation run located a random distance (within a specified radius of 500 meters) from the *household* of a selected agent. Note that more than one agent from a single household may be selected such that multiple on-demand ride service drivers are initialized near the household. The random seed for each simulation run is fixed such that the locations of drivers are determined only once for all simulations run by all contestants.

A repositioning algorithm specifies how agents behave when not carrying passengers. Repositioning in BEAM specified that a driver in an empty state (even one in motion) may be re-routed to a new customer. While several repositioning algorithms are available in BEAM, for computational and evaluation reasons, we have decided to specify a default repositioning mechanism for Sioux Faux. The default behavior causes driver agents who have just dropped off a passenger to stop driving and wait at the drop-off point until receiving a subsequent request.

2.4.7 Mode choice modeling

One of the key features of travel demand simulations (including BEAM) is that they represent agent decision-making behavior in accordance with discrete choice theory[14] Agents in BEAM are endowed with the ability to trade off time and money to choose the mode of travel that maximizes their enjoyment of important activities. This subsection explains the basic concepts behind how incentives and level of service changes input to BISTRO influence the behavior of agents in BEAM.

Discrete choice modeling is a probabilistic modeling approach to predict the choices of users between a finite number of two or more orthogonal (or mutually exclusive) alternatives, among other applications. Discrete choice has its roots in utility theory, and in general, discrete choice models can vary based upon functional form, the ways in which individual and alternative characteristics are modeled, and the ways in which stochasticity is handled.

In transportation, discrete choice models are often used to explain, predict, and analyze user decisions regarding what mode of transportation to utilize, hence the name mode choice model. One prevalent example is *multinomial logit modeling* (MNL), which models the decision of individuals between a set of greater than two mode choice alternatives, assuming that the alternatives are not correlated. In MNL, each individual $n \in N$ is assigned a probability of choosing a mode $i \in J$ between an exhaustive set of alternatives $j \in J$ (i.e., an individual cannot choose something outside the set). The probability is represented as follows:

$$P_{ni} \equiv \Pr(\text{Person } n \text{ chooses alternative } i) = G(x_{ni}, x_{nj,j\neq i}, s_n, \eta)$$

where:

[14]Pioneering models of travel mode choice in the disciplines of Econometrics, Psychology, and Transportation Science have been developed and repeatedly, empirically validated by Kenneth Train, Daniel McFadden, and Moshe Ben-Akiva, among others.

- G is a function that returns the probability that person i chooses alternative n given the attributes of the alternatives, person characteristics, and model parameters.
- x_{ni} is a vector of attributes of alternative i.
- $x_{nj,j\neq i}$ is a vector of attributes of all other alternatives
- s_n is a vector of characteristics (e.g., demographics, value of time, etc.) of person n.
- η is a set of parameters delineating the effects of the attributes on choice probabilities, which are estimated statistically.

Each individual has a utility function which characterizes the welfare resulting from making a mode choice decision. Individuals are assumed to be utility maximizing–they will choose the alternative with the highest expected utility, given their attributes and the attributes of each alternative. The utility functions have the form:

$$U_{ni} = \beta z_{ni} + \epsilon_{ni}$$

where:

- U_{ni} is the utility associated with choosing alternative i.
- β is a vector of coefficients corresponding to observed variables related to alternative i and person n.
- z_{ni} is a vector of observed variables for alternative i and person n that may depend on x_{ni}, and/or s_n.
- ϵ_{ni} is an error term that captures the influence of unobserved factors that affect person n's utility from choosing alternative i. For the purposes of MNL, ϵ_{ni} is assumed to be independently and identically distributed (iid) following the extreme value distribution.

The probability the person n chooses alternative i, in this case, can be calculated in the following way:

$$P_{ni} = \frac{\exp(\beta z_{ni})}{\sum_{j=1}^{J} \exp(\beta z_{nj})}$$

This type of formulation is sometimes referred to as the Softmax function. The probabilities for every alternative $i, j \in J$ must sum to 1, and it is assumed that every person n chooses the alternative with the highest associated probability.

The parameters used for the MNL model within BEAM, and thus BISTRO, are shown below in Table 1. Figure 6 displays the mode share in the business as usual scenario of Sioux Faux.

Description	**Variable Type**	**Value**
Value of Time	$s_{n_{VOT}}$	18 [$/hr]
Transfer Coefficient	$\beta_{transfer}$	-0.6
Car Coefficient	β_{car}	10.562
Walk to Transit Coefficient	$\beta_{walk-transit}$	-10.0
Drive to Transit Coefficient	$\beta_{drive-transit}$	0
On-demand Ride Coefficient	$\beta_{on-demand-ride}$	-0.124
Walk Coefficient	β_{walk}	-11.0

Table 1: Mode Choice Parameters

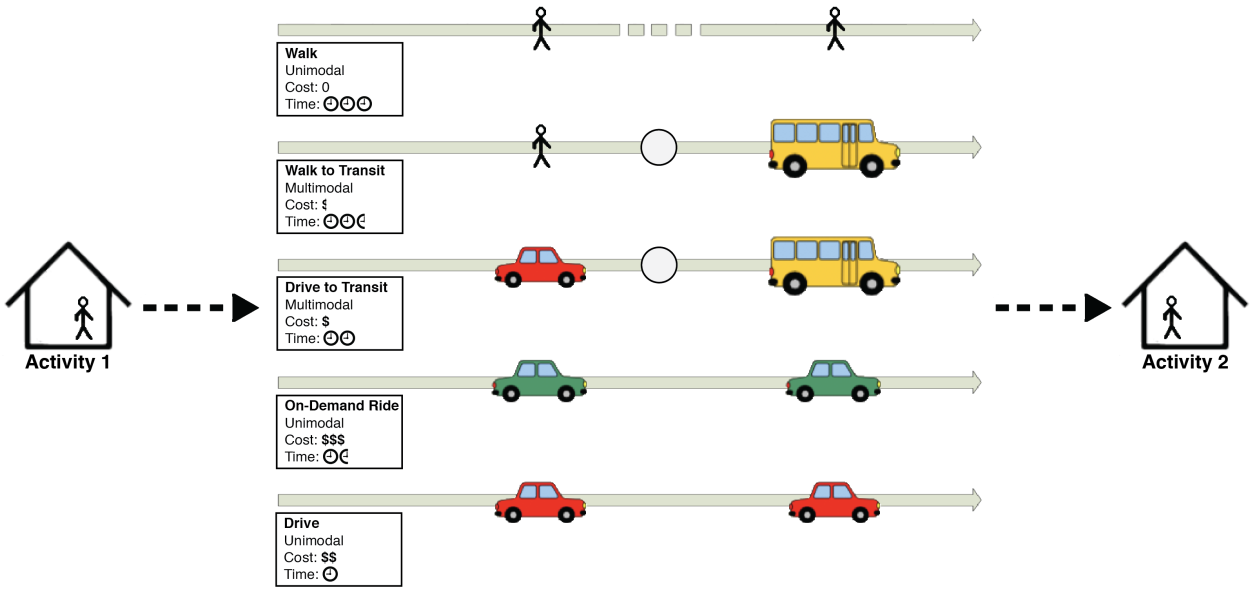


Figure 5: The transportation modes available to Sioux Faux' population.

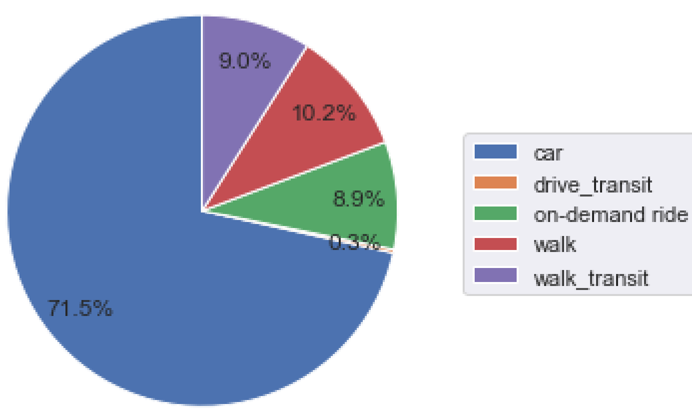


Figure 6: Mode Share for the Sioux Faux BAU Scenario

3 Task Description

Uber Prize contestants are tasked with assisting the *Sioux Faux Department of Transportation* (SFDOT) to reduce congestion and improve overall accessibility in Sioux Faux using an agent-based simulation of the city's transportation demand. The suite of inputs under contestants' control represent possible city-wide policies regarding transportation: policies that control *operational* and *financial* aspects of mass transit as well as *incentives* that influence the use of mass transit and on-demand ride services. Representatives from SFDOT have provided *guidelines* and *evaluation metrics* for the policies under consideration.

A fully specified transportation policy consists of a set of *contestant-defined input variables*: bus fleet vehicle composition, bus service frequency, and the details of a newly agreed upon partnership between Sioux Faux Bus Lines and on-demand ride service providers operating within the city boundaries that provides incentives for on-demand ride and transit trips.

Solutions will be evaluated based on a scoring function that weights performance metrics computed from the simulation outputs, including measures of accessibility and the level of service of the transportation network, congestion, the net costs to the SFDOT, and environmental sustainability. The scoring and evaluation criteria are detailed in Section 4.

3.1 Mass transit operations

SFDOT is exploring the possibility of providing *Sioux Faux Bus Lines* (SFBL) with additional funding to improve the operations of existing bus services. As the highest-capacity transportation option available in Sioux Faux, mass transit offers an efficient means by which to service travel along corridors under high demand. Unfortunately, the *mode share* of mass transit in Sioux Faux is rather low, at about nine percent. According to a user survey, travelers avoid mass transit due to a perceived lack of control over departure timing (both in frequency and variability), habit or inertia that induces automobile dependency, as well as indicators of comfort such as crowding.

Currently, SFBL operates a small fleet of buses on 12 routes in Sioux Faux. Contestants may impact the level of service provided by the bus system by modifying the transit vehicle fleet composition, altering the bus route headways (service frequencies), and impacting the fares of existing bus routes by altering the fares and/or providing incentives for transit use by particular demographic groups.

Contestants will tune the parameters listed below:

1. **Vehicle fleet composition:** originally purchased as a group, each bus currently in the SFBL fleet possesses identical attributes of seating capacity, fuel consumption, operations and maintenance cost, etc. SFBL is considering optimizing bus type in order to improve the level of bus service by better matching the specific demand characteristics of each route. Four types of buses are available from its supplier, each of them with different technical properties and cost characteristics. Contestants may choose to purchase additional vehicles to include in the bus fleet for any of the existing bus routes, specifying which vehicle types are to be used for each route.

2. **Route schedules:** a bus route schedule is determined by the hours of service, or the *service period*, and the *headway*, or time between vehicles traveling along the route. At the start of a service period of a route, a bus is dispatched from the route origin to travel a predefined route between stops. An additional bus departs from the origin after each interval of time defined by the route headway. Contestants may choose to define multiple service periods for each route, each with varying headways. The number of buses used to service a route will be dependent on the headways across all service periods, thus requiring contestants to co-optimize the vehicle fleet composition with route scheduling.

3. **Transit fares:** contestants may alter the *fare*, or cost to a passenger of traveling on a particular transit route. Passengers pay one flat fare each time they board a bus. Contestants may choose to alter the fare for any of the SFBL routes and may choose to segment the fares based on passenger age groups.

3.2 Transit and on-demand ride incentives

In an effort to provide a viable alternative to private automobiles for as wide a population as possible, SFBL is considering providing incentives to promote mass transit in Sioux Faux. SFBL is exploring options for citizens lacking access to quality transit or means to pay fares, including defraying the cost of certain qualified transit trips and/or on-demand rides. While the exact details of the incentives are yet to be determined, SFBL staff are keenly interested in seeing the trade-offs between service provision, operational costs, Fauxian mobility, and sustainability.

In order to specify the incentive policy, contestants will use a structure that provides reimbursement to qualifying individuals based on age and income. Contestants may choose to defray the cost of on-demand rides and/or transit based on either age group, income group, or both. To do so, the range of qualifying socio-demographic characteristics and value of the incentive provided to each group must be defined for passengers using each of the following modes of transportation to complete a trip: on-demand rides, walk to/from transit, or drive to/from transit.

4 Evaluation and Scoring

Solutions will be judged based on a weighted combination of measurable outcomes from the simulation that emulate common operational and social goals considered by cities when evaluating the broader impacts of transportation policy and investment. The judging criteria are derived from a discrete set of output variables produced for each simulation run.

Each simulation run takes as input a set of configuration and contestant-defined input variables. The configuration variables define the geographic and physical constraints of the transportation network, the characteristics of vehicles in the transit, private, and on-demand ride fleets, and the instantiation of each agent in the population, including their socio-demographic characteristics and activity plans. The configuration variables may not be altered in any way by the challenge contestants.

The contestant-defined inputs specify the type of vehicles to be added to the bus fleet, the assignment of vehicle types to specific routes in the SFBL service, the headway, service periods, and fare of each route, and finally, transit and on-demand ride incentive policies.

The outputs of the simulation produced by contestants' solutions will determine the values of key performance metrics of the impact of the solutions on the accessibility, level of service, and congestion of the transportation network in Sioux Faux, as well as the resource constraints and environmental sustainability of the resulting network-wide travel equilibrium.

The following subsections detail the relevant person agent and vehicle movement outputs in the simulation as well as the scoring criteria used for evaluation.

4.1 Person output

For each trip taken by a person agent in the simulation, the following data is produced as output:

1. **Transportation mode(s):** for each trip, a person agent chooses one of the modes available to them. Person agents may use one or more modes to travel from their origin to their desired destination, as they may transfer between modes along the way.
   (a) **mode(s) available:** the mode(s) available to person agents to use for each trip, including: walk, personal car, on-demand ride, bus, drive to/from bus, and walk to/from bus.
   (b) **mode choice:** the mode chosen for each trip, as determined by the process described in Section 2.4.7.
2. **Travel time:** the time spent by the person agent in the act of traveling during each leg of a trip. Travel time has several components, including:
   (a) **In-vehicle travel time:** the time spent in a vehicle by an agent while traveling to an activity.
   (b) **Wait time:** the time spent by an agent waiting for the arrival of a vehicle. Wait time may include time spent at a bus stop or time spent waiting for the arrival of an on-demand ride vehicle after the ride has been reserved.
   (c) **Transfer time:** time spent walking from one transportation mode to another while completing a trip. Transfer time may include walking from the bus stop of one bus route to a bus stop of another bus route.

3. **Travel expenditure:** the cost incurred by a person agent during a trip. The net cost of travel incurred may include:
   (a) Mass transit fares
   (b) On-demand ride fares
   (c) Gas consumption by a personal vehicle
   (d) Applicable incentives
4. **Incentives:** the amount of monetary incentive available (based on the modes available) for each leg of a trip and the amount of incentive consumed by an agent during each leg of a trip.
5. **Trip purpose:** the nature of the primary activity to which a person agent is traveling during a trip. Trip purpose is segmented into two mutually exclusive categories: work trips and secondary trips.

4.2 Vehicle output

Every vehicle movement during the simulation produces the following outputs:

1. **Origin-Destination-Time (ODT) record:** the origin location, destination location, time of departure, and time of arrival of a vehicle movement.
2. **Path:** an ordered list of the links traversed on the path from the origin to the destination of a vehicle movement.
3. **Fuel consumption:** the amount of fuel consumed by a vehicle during a movement.
4. **Vehicle occupancy:** the number of passengers in a vehicle during a movement.

4.3 Scoring criteria

The overall scoring function will appropriately weigh multiple aggregate measures of the the simulation outputs that evaluate system-wide accessibility, level of service, congestion, resource constraints, and sustainability. These five categories of criteria are all affected by the user-defined input variables, and care must be taken when optimizing to understand the interactions between metrics.

- There are three types of metrics: 1) those that involve system users only (e.g., incentives used, average travel expenditure), 2) those that involve users and the system (e.g., accessibility, delay), and 3) those that involve only the system (e.g. operational costs, revenue)
- We are treating these three types of metrics simultaneously and comparing the user-produced results to the BAU scenario
- This can be thought of similarly to a costs and benefits approach, however it differs from many transportation implementations as user and agency-specific scores can both be thought of as benefits, or costs, depending upon the metric.

4.3.1 Accessibility

While the term *accessibility* takes on a variety of meanings in different contexts, in an urban transportation planning setting, accessibility has often been defined as a measure of the ease and feasibility with which opportunities or points of interest can be reached via available modes of travel. Although there are many ways to measure accessibility, it will be quantified as the average number of points of interest (of a specific type of activity) reachable within a given duration of time. More specifically, during Phase 1 accessibility will be measured as the sum of the average number of points of interest reachable from network nodes by any mode using the road network, (i.e., on-demand rides, buses, and cars), within a specified amount of time during the peak hour as shown below.

1. **Work-based trips** The sum of the average number of work locations accessible from each node by automotive modes within 15 minutes during the AM peak (7-10 am) and PM peak (5-8 pm) periods.
2. **Secondary trips** The sum of the average number of secondary locations accessible from each node by automotive modes within 15 minutes during the AM peak (7-10 am) and PM peak (5-8 pm) periods.

Figures 7 and 8 display the accessibility for work-based and other trips, respectively, in the Sioux Faux BAU scenario.

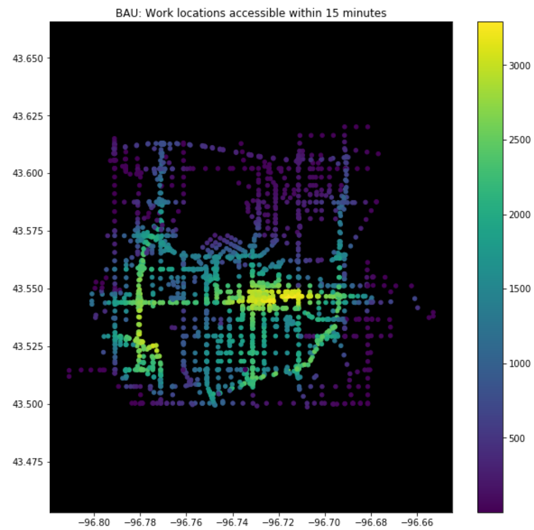


Figure 7: Accessibility to work locations in the Sioux Faux BAU scenario

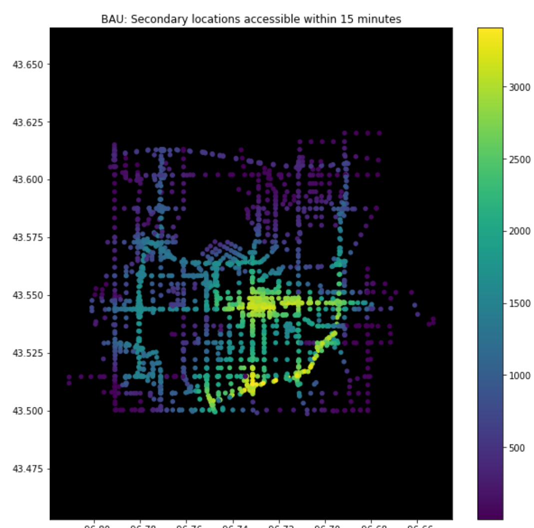


Figure 8: Accessibility to secondary locations in the Sioux Faux BAU scenario

4.3.2 Level of service (LoS)

The level of service (LoS) provided by the transportation network will be evaluated by assessing the travel cost, wait times, and comfort - measured as crowding on transit vehicles. The LoS scoring criteria will be assessed on average, across all agent trips.

1. **Average Travel Expenditure**: the average over all trips of the total cost of travel incurred by all person agents during the simulation. Travel expenditures for a person agent trip are calculated as the sum of all bus fares, on-demand ride fares, and the cost of gas consumed subtracted by the total incentives received during the trip. Average travel expenditure will be assessed for work trips and secondary trips separately.

2. **Average Bus Crowding Experienced**: measured in agent hours, this is the average time spent per agent trip in buses occupied above their seating capacity.

4.3.3 Congestion

Congestion is measured in two primary ways: the total sum of miles traveled by all modes on the network, and by the delay incurred. Delay–calculated as the difference between actual and free flow time–is presented both as a sum across all vehicle movements in the simulation and as an average delay per agent trip. Using these three measures of congestion provides insight into the destination- or opportunity-independent level of mobility on a network, the overall network performance, and efficiency.

1. **Total Vehicle Miles Traveled (VMT)**: total miles traveled by all motorized vehicles of the system during the simulation.

2. **Average Vehicle Delay per Person Agent Trip**: the average across agent trips of vehicle hours of delay experienced by all vehicles while occupied by one or more passengers during the simulation.

4.3.4 Mass transit LoS intervention: costs and benefits

The costs and benefits of mass transit interventions will be considered on aggregate, as the sum of the operational costs of bus service and the total incentives used subtracted by the total bus revenue collected.

1. **Operational Costs**: total costs incurred by SFBL operations including the cost of fuel consumed, and hourly variable costs. Hourly variable costs include estimated labor, maintenance and operational costs. The rates for each of these factors is specified in the vehicle fleet configuration variables (see Table 2).

Vehicle type, $c \in C$	**Fuel type**	**Fuel consumption rate ($Joule/meter$)**	**Operational cost ($\$/hour$)**	**Seating capacity**	**Standing capacity**
CAR	gasoline	3655.98	n/a	4	0
BUS-DEFAULT	diesel	20048	89.88	37	20
BUS-SMALL-HD	diesel	18043.2	90.18	27	10
BUS-STD-HD	diesel	20048	90.18	35	20
BUS-STD-ART	diesel	26663.84	97.26	54	25

Table 2: Vehicle Types

Fuel type, $f \in F$	**Fuel cost ($\$/MJoule$)**
Gasoline	0.03
Diesel	0.02
Electricity	0.01
Biodiesel	0.01

Table 3: Fuel Types

2. **Incentives Used**: total incentives used by agents.
3. **Revenue**: sum of total bus fares collected.

4.3.5 Sustainability

Sustainability metrics provide a sense of the local externalities resulting from any transportation interventions. Using criteria pollutants, specifically particulate matter running exhaust emission factors, presents a mileage-based measure of local air quality impacts based upon vehicle type. This metric provides a complementary addition to fuel/energy consumption-based metrics, which are captured elsewhere. For more information on the methodology followed to develop this metric, please refer to the California Air Resources Board documentation [15].

1. **Particulate ($PM_{2.5}$) Emissions**: total $PM_{2.5}$ emissions produced by all motorized vehicles during the simulation.

5 Variable and Scoring Function Specification

The scoring criteria are defined explicitly as functions of the input and output variables of a simulation run. For ease of understanding, the variable notation reflects the meaning of each category of variable in the following manner: all network and vehicle configuration input variables are denoted with a Z, all population configuration input variables are denoted with a N, all contestant-defined input variables are denoted with a D, all person agent-level output variables are denoted with an X, and all vehicle movement output variables are denoted with a Y. The indices identifying the meaning of each variable will be defined in the following sections in addition to the corresponding units of measurement.

5.1 Input Variable Specification

5.1.1 Network and fleet configuration

Network and fleet characteristic variables are defined during the configuration of a simulation, and remain static for each scenario.

The road network is a directed graph, $\mathcal{G}(L, W)$, comprised of W nodes connected by L links. Road network configuration variables are indexed by the link identifier $l \in L$. A link instance is defined by the Boolean variable, $Z^l_{i,j}$, which indicates the origin and destination nodes connected by link l.

$$Z^l_{i,j} = \begin{cases} 1 \text{ if link } l \text{ connects node } i \text{ to node } j \\ 0 \text{ otherwise} \end{cases}$$

Each link is given a length (in miles), Z^l_{length}, capacity (in vehicles per hour), $Z^l_{capacity}$, and free-flow speed (in miles per hour), $Z^l_{speed-limit}$. These variables are input to the routing operator during the simulation to determine the travel time of vehicle movements as a function of the number of vehicles on a link at a particular time.

The transit network configuration follows a *General Transit Feed Specification* (GTFS) format. The transit network is a subgraph of the road network comprised of a set of transit facilities (bus stops) $F \subset W$ and a set of T routes. The location of each bus stop, $f \in F$, is defined by the Boolean variable Z^f_w such that

$$Z^f_w = \begin{cases} 1 \text{ if bus stop } f \text{ is located at node } w \\ 0 \text{ otherwise} \end{cases}$$

The assignment of transit facilities to routes is defined by the Boolean variable $Z^f_{route,t}$ such that

$$Z^f_{route,t} = \begin{cases} 1 \text{ if bus stop } f \text{ is on bus route } t \\ 0 \text{ otherwise} \end{cases}$$

[15]https://www.arb.ca.gov/cc/capandtrade/auctionproceeds/cci$_emissionfactordatabase_documentation.pdf?_ga = 2.94247453.1690201828.1547860553 - 1364631033.1545190476$

In the base case, the fare structure for SFBL is age-based and identical across all routes in the agency. The fares in the base case are defined by the set $Z_{fare} = \{Z_{fare,a}\}_{a=0,1,\ldots,100}$ such that $Z_{fare,a}$ denotes the fare for passengers of age $a \in \{0, 1, \ldots, 100\}$.

Each route $t \in \{1, 2, ..T\}$ is defined by an ordered set of p_t links, $Z^t_{path} = \{z^t_l\}_{l\in\{1,2,\ldots,p_t\}}$. The schedules for each bus route are defined by the GTFS input, which provides the headway, $Z^t_{headway}$, and the set of all trips, $Z^t_{trips} \in \mathbb{R}^{r_t}$, to be completed for each route. For the purposes of this phase of the *Uber Prize*, each trip on a particular bus route is assigned to a unique bus id. Thus, we denote the set of all buses to be used on route t as $V_{bus,t} \in \mathbb{R}^{r_t}$, and the set of all buses in the base case as $V_{bus} = \cup^T_{t=1} V_{bus,t} \in \mathbb{R}^{Tr_t}$. Each bus, $v \in V_{bus,t}$, is given a start and end time (in seconds from the start of the simulation) for servicing the corresponding route, t, $Z^v_{start-time}$ and $Z^v_{end-time}$, respectively. Thus, the configured service period for each route, t, is the tuple $(Z^t_{start-time}, Z^t_{end-time})$ where $Z^t_{start-time}$ corresponds to the start time of the bus making the first trip on the route and $Z^t_{end-time}$ corresponds to the end time of the bus making the last trip on the route.

The set of all vehicles, V, is comprised of the union of the set of personally owned vehicles, $V_{personal}$, the set of buses, V_{bus}, and the set of on-demand vehicles, $V_{on-demand}$. The number and type of vehicles in the private and on-demand fleets is defined during the configuration of the simulation and remains independent of contestant-defined input. However, the number of vehicles in the bus fleet is dependent on the service periods and headways for each route, which can be altered by contestants. The type of vehicles in the transit fleet may also be altered by the contestants.

Across all vehicles, the set of possible vehicle types is the set C. Each vehicle type, $c \in C$, is defined by a seating and standing capacity (in number of person agents), $Z^c_{seating}$ and $Z^c_{standing}$, a fuel type, $Z^c_{fuel-type}$, fuel consumption rate (in units of fuel consumed per mile), $Z^c_{fuel-consumption}$, and a variable operational cost (in dollars per hour traveled), $V^c_{var-op-cost}$. The set of fuel types in this phase of the *Uber Prize* is $F = \{gasoline, diesel\}$. For each fuel type, $f \in F$, $Z^f_{fuel-cost}$ defines the cost of consumption (in dollars per unit of fuel consumed).

At the start of a simulation run, each vehicle is instantiated with an identifier, $v \in V$, and a Boolean identifier, $Z^v_{veh-type,c}$, that denotes the vehicle type of the vehicle such that

$$Z^v_{veh-type,c} = \begin{cases} 1 \text{ if vehicle } v \text{ is of vehicle type } c \\ 0 \text{ otherwise} \end{cases}$$

Additionally, all vehicles in the private vehicle fleet are identified by the household to which they belong using the Boolean identifier Z^v_h, defined in the following subsection.

5.1.2 Population configuration

At the start of the simulation, a synthetic population is generated as described in 2.4.2. The population consists of a set of households, H. Each household $h \in H$ owns a number of vehicles, $N^h_{vehicles} \geq 0$. The Boolean identifier, Z^v_h denotes whether vehicle $v \in V_{private}$ is owned by household h such that

$$Z^v_h = \begin{cases} 1 \text{ if vehicle } v \text{ is owned by household } h \in H \\ 0 \text{ otherwise} \end{cases}$$

Each person agent in the population, $n \in N$, is a member of one household, denoted by the Boolean identifier $N^n_{hhd,h}$, with home location denoted by $N^h_{home} \in W$ such that

$$N^n_{hhd,h} = \begin{cases} 1 \text{ if agent } n \text{ is a member of household } h \in H \\ 0 \text{ otherwise} \end{cases}$$

Additionally, each person agent is assigned fixed socio-demographic attributes, denoted by Boolean identifiers. These include variables identifying age, $N^n_{age,a}$, gender, $N^n_{gender,g}$, and income, $N^n_{income,i}$ such that

$$N^n_{age,a} = \begin{cases} 1 \text{ if agent } n \text{ is of age } a \in \{0, 1, \ldots, 100\} \\ 0 \text{ otherwise} \end{cases}$$

$$N^n_{gender,g} = \begin{cases} 1 \text{ if agent } n \text{ is of gender } g \in \{\text{female, male, other}\} \\ 0 \text{ otherwise} \end{cases}$$

$$N^n_{income,i} = \begin{cases} 1 \text{ if individual } n \text{ is in the income group } i \in I \\ 0 \text{ otherwise} \end{cases}$$

where H is the set of all households and $I = \{$no income, less than \$10,000, \$10,000 to \$24,999, \$25,000 to \$49,999, \$50,000 to \$74,999, \$75,000 to \$99,999, \$100,000 and above$\}$.

Finally, each person is instantiated with a plan, N^n_{plan}, which is an ordered set of activities starting and ending at home. During this phase of the *Uber Prize*, plans will be simplified, with just one activity outside of the home. Thus, all agents have a plan of the form

$$N^n_{plan} = \{home, N^n_{primary-activity}, home\}$$

where $N^n_{primary-activity} \in \{\text{work, secondary}\}$. Each activity in a person's plan, $p \in N^n_{plan}$, has a desired start and end time, denoted (in seconds from the start of the simulation) by $N^{n,p}_{start}$ and $N^{n,p}_{end}$, respectively.

5.1.3 Contestant-defined input

As described in section 3, contestants may redefine the transit vehicle fleet composition, bus service schedules, and fares, as well as incentive amounts for particular demographic groups to use bus and/or on-demand ride services.

In the base case, the bus fleet is homogeneous, with all vehicles in the fleet configured with the same vehicle type:

$$Z^v_{veh-type,c} = Z^{v'}_{veh-type,c} \ \forall v, v' \in V_{bus}, c \in C$$

Contestants may choose to alter the vehicle type servicing each transit route, $t \in \{1, 2, \ldots, T\}$, by changing the value of the variable $D^t_{veh-type,c}$, which denotes the vehicle type for route t. Thus

$$D^t_{veh-type,c} = D^v_{veh-type,c} = D^{v'}_{veh-type,c} \ \forall v, v' \in V_{bus,t}, c \in C$$

Contestants may redefine the service period of a bus route by appending one or more tuples to the mutable, ordered array of tuples, $D^t_{service-period}$, corresponding to the bus route t. Thus, a bus route with s contestant-defined service periods is of the form:

$$D^t_{service-period} = \{(D^t_{start,1}, D^t_{end,1}), (D^t_{start,2}, D^t_{end,2}), \ldots (D^t_{start,s}, D^t_{end,s})\}$$

where

$$D^t_{start,i} > D^t_{end,i} \ \forall i \in \{1, 2, \ldots s\}$$

and

$$D^t_{end,i} \geq D^t_{start,i+1} \ \forall i \in \{1, 2, \ldots s-1\}$$

Contestants may redefine the headway of a route using the ordered array of s variables, $D^t_{headway} = \{D^t_{headway,i}\}_{i=1,2,\ldots,s}$, corresponding to each service period. All route headways must be greater than or equal to a minimum, $Z_{headway-min}$, defined in the network configuration.

The fare for riding on route t can be redefined by the set of contestant-defined variables, $D^t_{fare} = \{D^t_{fare,a}\}_{a=0,1,\ldots,100}$. All fares must be nonnegative.

Finally, contestants may choose to allocate incentives for on-demand ride, walk to/from bus, and/or drive to/from bus trips, using the variables

$$D^{on-demand}_{incentive} = \{D^{on-demand}_{incentive,a,i}\}_{a=0,1,\ldots,100; i \in I}$$

$$D^{walk-transit}_{incentive} = \{D^{walk-transit}_{incentive,a,i}\}_{a=0,1,\ldots,100; i \in I}$$

$$D^{drive-transit}_{incentive} = \{D^{drive-transit}_{incentive,a,i}\}_{a=0,1,\ldots,100; i \in I}$$

where the indices a, i correspond to the age and income group of riders, respectively. All incentive values must be nonnegative.

5.2 Output Variables Specification

5.2.1 Person output

The person agent output reports the choices, movements, and expenditures of each person agent in the population during a simulation run. Since all plans in Sioux Faux consist of one activity outside of the home, each person agent, $n \in N$ takes $R_n \in \mathbb{R}^2$ trips during the simulation.

The mode(s) available to person agent n for trip $r \in R_n$ are output as a set of Booleans,

$$X^{n,r}_{available} = \{X^{n,r}_{available,m}\}_{m=\text{ \{walk, car, on-demand, walk-bus, car-bus \}}}$$

where

$$X^{n,r}_{available,m} = \begin{cases} 1 \text{ if mode } m \text{ is available to person } n \text{ for trip } r \in R_n \\ 0 \text{ otherwise} \end{cases}$$

The mode ultimately chosen by the agent is output as a set of Booleans,

$$X^{n,r}_{choice} = \{X^{n,r}_{choice,m}\}_{m=\text{walk, car, on-demand, walk-bus, car-bus}}$$

where

$$X^{n,r}_{choice,m} = \begin{cases} 1 \text{ if mode } m \text{ is chosen by person } n \text{ for trip } r \in R_n \\ 0 \text{ otherwise} \end{cases}$$

Trips may include multiple legs. Thus, the number of legs, $G_{n,r}$, included in trip $r \in R_n$ of agent $n \in N$, is

$$G_{n,r} \begin{cases} = 1 \text{ if } X^{n,r}_{choice,walk} = 1 \text{ or } X^{n,r}_{choice,drive} = 1 \text{ or } X^{n,r}_{choice,on-demand} = 1 \\ >= 2\text{otherwise} \end{cases}$$

Bus trips include an access and egress leg as well as one or more bus legs, depending on the use of one or more bus routes during the trip. The leg used during each leg is denoted by the Boolean

$$X^{n,r,g}_{mode,m} = \begin{cases} 1 \text{ if mode } m \text{ is used by person } n \text{ for leg } g \text{ of trip } r \\ 0 \text{ otherwise} \end{cases}$$

The output variable $X^{n,r,g}_{route} \in \{0, 1, 2, \ldots, T\}$ records the bus route used by agent n during leg g of trip r such that $X^{n,r,g}_{route} = 0$ if the agent did not use a bus during leg g.

Each leg of a trip traverses a path, recorded as an ordered list of $p_{n,r,g}$ links traversed,

$$X^{n,r,g}_{path} = \{X^{n,r,g}_{path,l}\}_{l \in \{1,2,\ldots,p_{n,r,g}\}}$$

where $X^{n,r,g}_{path,l} \in L$ is a link in the network.

The distance traveled (in miles) by agent n during leg g of trip r is

$$X^{n,r,g}_{distance} = \sum_{l \in X^{n,r,g,}_{path}} Z^{X^{n,r,g}_{path,l}}_{length}$$

The duration (in seconds) of each leg is a result of the traffic dynamics of the simulation and is recorded by the ouput variable $X^{n,r,g}_{duration} \geq 0$. The expenditure (in dollars \$) incurred by a person agent during a trip is a product of the mode use identifier, and the corresponding input variables denoting the applicable fare(s) for each leg and incentive(s) for the trip. The bus fare incurred by agent n during a trip leg is

$$X^{n,r,g}_{fare,bus} = X^{n,r,g}_{mode,bus} \sum_{a=0}^{100} N^{n}_{age,a} D^{X^{n,r,g}_{route}}_{fare,a}$$

The on-demand ride fare incurred by agent n during any trip leg is

$$X^{n,r,g}_{fare,on-demand} = X^{n,r,g}_{mode,on-demand}(Z^{on-demand}_{fare,base} + X^{n,r,g}_{distance} Z^{on-demand}_{fare,distance} + \frac{1}{60} X^{n,r,g}_{duration} Z^{on-demand}_{fare,duration})$$

The cost of fuel consumed by agent n during any trip leg is

$$X^{n,r,g}_{fuel-cost} = X^{n,r,g}_{mode,car} X^{n,r,g}_{distance} \sum_{h \in H} \sum_{v \in V_personal} \sum c \in C N^n_{hhd,h} Z^v_h Z^v_{veh-type,c} Z^{Z^c_{fuel-type}}_{fuel-consumption} Z^{Z^c_{fuel-type}}_{fuel-cost}$$

The incentive available to agent n during trip r is given by the output variable:

$$X^{n,r}_{incentive} = \sum_{a=0}^{100} \sum_{i \in I} N^n_{age,a} N^n_{income,i} (X^{n,r}_{choice,on-demand} D^{on-demand}_{incentive,a,i} + X^{n,r}_{choice,walk-bus} D^{walk-bus}_{incentive,a,i} + X^{n,r}_{choice,drive-bus} D^{drive-bus}_{incentive,a,i})$$

Thus, the total expenditure incurred per person agent trip is captured by the output variable,

$$X^{n,r}_{exp} = \left(\sum_{g=1}^{G_{r,n}} (X^{n,r,g}_{mode,bus} X^n_{fare,bus} + X^{n,r,g}_{mode,drive} X^{n,r,g}_{fuel-cost} + X^{n,r,g}_{mode,on-demand} X^n_{fare,on-demand}) - X^{n,r}_{incentive} \right)_+$$

where

$$(x)_+ = \begin{cases} x \text{ if } x \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

In the event that the incentive amount available to an agent for a particular trip exceeds the total fare of the trip, the agent receives an incentive amount equal to the total fare(s) incurred.

5.2.2 Vehicle output

The vehicle output reports the movements of all vehicles during a simulation run. Each vehicle, $v \in V$, makes $Q_v \geq 0$ movements. For buses, a movement consists of travel between two bus stops. For personal vehicles, a movement consists of travel between two parking facilities. Finally, for on-demand ride vehicles, a movement consists of the travel from origin to destination during any one of the three phases of service: empty, fetch, and fare (see 2.4.6).

The path, fuel consumption, and occupancy of each vehicle v is recorded upon every vehicle movement. Similar to paths for person agent legs, the path traversed by vehicle v during a movement, $q = \{1, 2, ..., Q_v\}$, is recorded as an ordered list of $p_{v,q}$ links traversed,

$$Y^{v,q}_{path} = \{Y^{v,q}_{path,l}\}_{l \in \{1,2,...,p_{v,q}\}}$$

where $Y^{v,q}_{path,l} \in L$ is a link in the network.

Thus, the distance traveled by vehicle v during movement q is given by:

$$Y^{v,q}_{distance} = \sum_{l \in Y^{v,q}_{path}} Z^{Y^{v,q}_{path,l}}_{length}$$

The duration of the movement (in seconds) is recorded by the output variable $Y^{v,q}_{duration} \geq 0$, and the fuel consumed is:

$$Y^{v,q}_{fuel-consumed} = Y^{v,q}_{distance} \sum_{c \in C} Z^v_{veh-type,c} Z^{Z^c_{fuel-type}}_{fuel-consumption}$$

Thus the cost of fuel consumed by vehicle v during movement q is

$$Y^{v,q}_{fuel-cost} = Y^{v,q}_{fuel-consumed} \sum_{c \in C} Z^v_{veh-type,c} Z^{Z^c_{fuel-type}}_{fuel-cost}$$

Finally, the number of passengers in vehicle v during movement q is recorded by the output variable $Y^{v,q}_{pax} \geq 0$.

5.3 Scoring Function Specfication

All scoring functions will be assessed in comparison to the corresponding measure in the base scenario; that is, the scenario without any inputs. The following sections detail each of the scoring functions included in the composite score, which is explained in section 5.4.

5.3.1 Accessibility

Accessibility measurements utilize the network resulting for any given submission (i.e., links $l \in L$ are weighted with average travel-times during periods of interest). Accessibility is then calculated as the sum of the average number of points of interest (work or secondary) reachable from all nodes $w \in W$ by any road network mode within a specified amount of time. Pandana is used to aggregate over the network[16]. Points of interest are taken from activities in the 15k scenario.

1. **Work-based trips:**

$$F_{accessibility,work} = \frac{1}{2\sum_{w \in W}} \sum_{p \in P} \sum_{i \in W} \sum_{j \in W} I_{primary-activity=work}[d_{ij} < \tau]_p$$

where d_{ij} is the *shortest path directed network distance* on $\mathcal{G}(L, W)$ from node i to node j as follows:

$$d_{ij} = \sum_{l \in L} l_{ij}\ travel - time_{l_p}$$

where l_{ij} is a Boolean indicator regarding link l being on the shortest path from i to j:

$$l_{ij} = \begin{cases} 1 \text{ if } l \text{ is in } shortest - path_{ij} \\ 0 \text{ otherwise} \end{cases}$$

and $travel - time_{l_p}$ is the average link travel-time for each hour in the specified time period of interest, p. This average travel-time is based upon all modes using the link and could be a combination of private automobile, bus, and on-demand ride travel-times. The time periods of interest are defined as:

with the time periods of interest defined as:

- **Morning Peak** 7:00:00 am to 10:00:01 am
- **Evening Peak** 5:00:00 pm to 8:00:01 pm

2. **Secondary trips:**

$$F_{accessibility,secondary} = \frac{1}{2\sum_{w \in W}} \sum_{p \in P} \sum_{i \in W} \sum_{j \in W} I_{primary-activity=secondary}[d_{ij} < \tau]_p$$

where all variables except for $I_{primary-activity=secondary}$ are equivalent to those in the calculation for work-based trips.

5.3.2 Measures of LoS

The LoS of the transportation system will be evaluated on average, across person trips. Each person, $n \in N$ completes R_n trips from one activity to another throughout a simulation run. Each trip, $r \in R_n$ includes $G_{n,r}$ legs, each using a single mode of transportation.

[16]For more information, see: https://udst.github.io/pandana/

1. **Average Travel Expenditure**: As detailed in 4.1:3, the total travel expenditure, $X^{n,r}_{exp}$, for a person agent, $n \in N$, during any trip, $r \in R_n$, may include bus and/or on-demand ride fares, and the cost of fuel consumed less any applicable incentives.

   We consider the average travel expenditure by trip purpose: work trips and secondary activity trips. Thus the average travel expenditure for work trips is

$$F_{expenditure,work} = \frac{1}{\sum_{n\in N}\sum_{r\in R_n} I_{(N^n_{primary-activity}=work)}} \sum_{n\in N}\sum_{r\in R_n} I_{(N^n_{primary-activity}=work)} X^{n,r}_{exp}$$

   where the Boolean indicator $I_{(N^n_{primary-activity}=work)}$ is defined as follows

$$I_{(N^n_{primary-activity}=work)} = \begin{cases} 1 \text{ if the primary activity of agent } n \text{ is work} \\ 0 \text{ otherwise} \end{cases}$$

   Similarly, the average travel expenditure for secondary trips is

$$F_{expenditure,2ndary} = \frac{1}{\sum_{n\in N}\sum_{r\in R_n} I_{(N^n_{primary-activity}=secondary)}} \sum_{n\in N}\sum_{r\in R_n} I_{(N^n_{primary-activity}=secondary)} X^{n,r}_{exp}$$

2. **Average Bus Crowding Experienced**:

$$F_{bus-crowding} = \frac{1}{G_{bus}} \sum_{v\in V_{bus}} \sum_{q=1}^{Q_v} \left(Y^{v,q}_{duration} Y^{v,q}_{pax} \sum_{c=1}^{C} D^v_{veh-type,c} I_{(Y^{v,q}_{pax} > Z^c_{seating})} \right)$$

   where G_{bus} is the total number of trip legs in which a bus was used such that

$$G_{bus} = \sum_{n\in N}\sum_{r\in R_n}\sum_{g=1}^{G_{n,r}} X^{n,r,g}_{mode,bus}$$

   $I_{(Y^{v,q}_{pax} > Z^c_{seating})}$ indicates whether vehicle in the bus fleet $v \in V_{bus}$ is occupied above its seating capacity during movement q.

$$I_{(Y^{v,q}_{pax} > Z^c_{seating})} = \begin{cases} 1 \text{ if } Y^{v,q}_{pax} > Z^c_{seating} \\ 0 \text{ otherwise} \end{cases}$$

   .

5.3.3 Measures of congestion

The first two measures of congestion, total VMT and total vehicle hours of delay, are assessed on aggregate across all vehicle movements in a simulation run.

1. **Total VMT**:

$$F_{vmt} = \sum_{v\in V}\sum_{q=1}^{Q_v} Y^{v,q}_{distance}$$

   where $Y^{v,q}_{distance}$ is the total distance (in miles) traveled by vehicle v during movement q.

2. **Average Vehicle Delay per Passenger Trip**:

$$F_{pax-trip-delay} = \frac{1}{\sum_{n\in N} R^{motorized}_n} \sum_{n\in N}\sum_{r\in R_n} \left(\frac{1}{3600} X^{r,n}_{duration} - \sum_{g=1}^{G_{n,r}} \sum_{l\in X^{n,r,g}_{path}} \frac{Z^{X^{n,r,g}_{path,l}}_{length}}{Z^{X^{n,r,g}_{path,l}}_{speed-limit}} \right)$$

   where $R^{motorized}_n$ indicates the number of trips taken by agent n using a motorized mode:

$$R^{motorized}_n = \sum_{r\in R_n} (X^{n,r}_{car} + X^{n,r}_{on-demand} + X^{n,r}_{walk-bus} + X^{n,r}_{car-bus})$$

5.3.4 Mass transit LoS intervention: costs and benefits

Any intervention on the transportation system is likely to result in costs and benefits for the operation of mass transit in Sioux Faux. The costs/benefit scoring component is computed as the difference between the total revenue collected through bus fares $F_{revenue}$ and the net cost of operation buses $F_{op-cost}$ and incentives used to encourage greater adoption of mass transit $F_{incentives-used}$:

$$F_{C/B} = F_{revenue} - (F_{op-cost} + F_{incentives-used})$$

1. **Operational costs**: Operational costs include fixed costs, variable hourly costs, and fuel costs for mass transit operations.

$$F_{op-cost} = \sum_{c \in C} \sum_{v \in V_{transit} \cup V_{additional}} Z^{v}_{veh-type,c} \Big(\sum_{q \in Q_v} \frac{1}{3600} Y^{v,q}_{duration} Z^{c}_{var-cost} + Y^{v,q}_{fuel-consumed} Z^{c}_{fuel-cost} \Big)$$

where $Z^{v}_{veh-type,c}$ denotes the vehicle type,c, of vehicle, v, $X^{c}_{variable}$ is the variable hourly costs of operating a vehicle of type c, $Z^{c}_{fuel-cost}$ is the cost per unit of fuel consumed by vehicle type c, and $Y^{v,q}_{fuel-consumed}$ is the amount of fuel consumed by vehicle v during movement q.

2. **Incentives used**

$$F_{incentives-used} = \sum_{n \in N} \sum_{r \in R_n} \Bigg(\sum_{g=1}^{G_{r,n}} \big(X^{n,r,g}_{mode,bus} X^{n}_{fare,bus} + X^{n,r,g}_{mode,drive} X^{n,r,g}_{fuel-cost} + X^{n,r,g}_{mode,on-demand} X^{n}_{fare,on-demand} \big) - X^{n,r}_{exp} \Bigg)$$

where $X^{n,r}_{exp}$ is the expenditure by agent n during tour leg r, $X^{n,r}_{exp} \geq 0$ as defined in section 5.2.1. Incentives may only be used by qualifying agents, as defined by contestant inputs.

3. **Revenue**:

$$F_{revenue} = \sum_{n \in N} \sum_{r \in R_n} \sum_{g \in G_{n,r}} X^{n,r,g}_{choice,bus} X^{n,r,g}_{fare}$$

where $X^{n,r,g}_{choice,bus}$ indicates that agent n chose to use the bus for leg g of trip r. $X^{n,r,g}_{fare}$ is the fare paid, and $X^{n,r,g}_{incentive}$ is the incentive amount received by agent n for leg g of trip r.

5.3.5 Measures of sustainability

Sustainability will be assessed as the total particulate matter $PM_{2.5}$ running exhaust (RUNEX) emissions from all vehicle movements in a simulation run. $PM_{2.5}$ emissions vary by mode and by fuel type; in Phase 1 there are only two possible vehicle-fuel type combinations: gasoline auto and diesel bus, simplifying the summation to:

1. **Total $PM_{2.5}$ Emissions**:

$$F_{sust} = \sum_{c \in C} \sum_{v \in V} \sum_{f \in F} \sum_{q=1}^{Q_v} Y^{v,q}_{distance} PM^{c,f}_{2.5}$$

where $Y^{v,q}_{distance}$ is the total distance (in miles) traveled by vehicle v during movement q and PM_c is defined as follows:

$$PM^{c,diesel}_{2.5} = \begin{cases} 0.259366648 \; grams/mile \text{ if the vehicle is a bus} \\ 0.018403666 \; grams/mile \text{ if the vehicle is a car} \\ 0 \text{ otherwise} \end{cases}$$

$$PM^{c,gas}_{2.5} = \begin{cases} 0.002517723 \; grams/mile \text{ if the vehicle is a bus} \\ 0.001716086 \; grams/mile \text{ if the vehicle is a car} \\ 0 \text{ otherwise} \end{cases}$$

5.4 Composite Score

The composite score is a function of the relative improvement in each of the scoring metrics achieved by contestant submissions. The functional form of the composite score is additive, taking as inputs the normalized ratios of the scores for each metric of the contestant submission to the business as usual case. In order to normalize the ratios, 757 samples of randomized input variables were generated to produce the mean and standard deviation of the ratio for each metric. These values are presented in Table **??**. Furthermore, the objective is to minimize the composite score function, since an increase in many of the scoring metrics actually represents a scenario that is worse than business as usual (e.g., decreasing VMT over BAU results in a lower unscaled score than increasing VMT). To maintain consistency in this regard, we take the reciprocal of several of the scoring components that are positively related to desirable outcomes (e.g., improvements in accessibility). Thus, the composite scoring function is as follows:

$$\Phi\left(C_s, \mathbb{P} = \left(\vec{F}, \vec{\sigma}, \vec{\mu}, \vec{\tau}\right)\right) = \sum_{\{F_i, \sigma_i, \mu_i, \tau_i\} \in \mathbb{P}} \left(\frac{1}{\sigma_i}\left(\frac{F_i(C_s)}{F_i(C_{BAU})}\right)^{\tau_i} - \mu_i\right)$$

Where $\vec{F}$ is the vector of all scoring metrics defined previously.C_s and C_{BAU} are the sets of all scoring function parameters of a contestant submission and the business as usual scenario, respectively. Thus $\vec{F}(C_s)$ and $\vec{F}(C_{BAU})$ are the values of the metric vector for outputs of the submission and business as usual simulations, respectively. The parameters $\vec{\sigma}$ and $\vec{\mu}$ are vectors of the standard deviation and mean of each ratio of metrics obtained from random sampling of inputs, as described in PART II. The parameter $\vec{\tau} \in \{-1, 1\}$ determines whether the metric ratio will be taken as is, or as a reciprocal, as follows:

$$\tau_i = \begin{cases} -1 \text{ if it is desirable for metric } i \text{ to increase} \\ 1 \text{ otherwise} \end{cases}$$

6 Appendix A: Variable Specification

6.1 Notation

· **H** households

· **N** individual agents

· **P** agent plans (tours)

· **S** types of activities

· **V** vehicles

· **C** vehicle types

· **M** transportation modes

· **T** transit routes

· **F** transit facilities (bus stops)

· **L** links

· **W** nodes

· **Z** geographic zones

Scoring Document

Part II

PART II
BISTRO inputs, outputs and sample scenarios: Hackathon rules, scoring, judging and prizes

Sidney Feygin[†], Valentine Golfier-Vetterli[†], Teddy Forscher[*†], Jessica Lazarus[*†], Rashid Waraich[‡†], Colin Sheppard[‡], Alexandre M. Bayen[*‡†]

[*] Institute of Transportation Studies, University of California, Berkeley, US

[†] Uber, Inc., US

[‡] Lawrence Berkeley National Laboratory, US

February 19, 2019

Part II presents a benchmark scenario for Sioux Faux, as well as four sample scenarios aimed at optimizing specific metrics of the city. It can serve as a starter point for the teams participating in the Uber Prize, and can provide illustrative results aimed at giving competitors intuition about the meaning of the results. Additionally, it includes judging rules to determine the winners of the hackathon.

This Hackathon will be used to inform the Uber Prize: A machine learning transportation challenge intended to create plausible, implementable and useful insights for cities to better understand the challenges of their transportation network to improve urban mobility. Optimization that promotes the mobility and welfare for all citizens, mimics social realities and acceptable norms (i,e., not making senior citizens pay more than others), and improves the sustainability of the city are desirable.

For this first Phase of the competition, contestants are asked to optimize the transportation network for a sample of citizens from a mock city: Sioux Faux. The city's 157,000 citizens travel between activities using either their personal automobiles, buses provided via a public transit system, taxis enabled via an on-demand car-sharing company, active modes such as walking, or a combination of multiple modes in accordance with their preferences.

Section 2 explains the computation of the outputs and gives an interpretation to the inputs chosen by competitors to achieve specific goals. Section Section 4 explains how specific outputs are rewarded. In this section, the competitors can see how specific ways to score the solutions can reward specific policies more (for example a "transit friendly" policy, vs. a "transit adverse" policy. Gaining a good understanding how the scoring presented in this section will help the competitors win the Prize. In Section 5, we briefly explain how subsampling can be used to accelerate the search over solutions in the exploratory phase of the contest. Finally the **judging criteria and scoring rules** are covered in Section 6, which explains in detail how the scoring is done, and how it relates to the leaderboard.

Contents

1 Introduction

This document provides a full walk through of how the *Berkeley Integrated System for Transportation Optimization* (BISTRO) system is used in the context of the Uber Prize. BISTRO is an extension of BEAM, an LBNL-developed agent-based simulation architecture, which incorporates the MATSim algorithm for dynamic traffic assignment.

The walkthrough simultaneously highlights the rationale and methodology for transformation of Uber Prize score components via *standardization*. The components themselves represent improvements (or deterioration) in metrics representing *Key Performance Indicators* (KPI) of policy controls (referred to here as inputs) used to characterize the transportation system of Sioux Faux. The utility of standardization is illustrated using the results of simulations run using several input samples.

2 Rationale and Methodology

2.1 Goal of the Uber Prize

The overall goal of the first round of the Uber Prize is to develop an algorithm that finds a policy that will best improve several indicators of the quality of the transportation system in Sioux Faux. The Uber Prize uses policy to refer to a **combination of inputs** (as .csv files) representing changes in:

- Mass transit vehicle fleet composition;
- Bus frequencies;
- Distribution of incentives for agents using on-demand carsharing and/or mass transit;
- Mass transit fares.

All of these could improve the cost and experience of using multimodal transport for commuters.

Policies are tested via the BISTRO engine, by simulating the daily travels of a synthetic population of individuals called *Agents*, each with their own sociodemographic characteristics.

The quality of the new policy-based transportation system is evaluated based upon a comparison against the *business as usual* (BAU) scenario. The BAU scenario represents the baseline or "current" status-quo of the Sioux Faux transportation system. In other words, this is a "do-nothing" approach. This comparison answers the following question: **How will the new policy improve over the current state of the transportation system?**

The performance of simulated policies over the baseline is measured using a scoring function. The scoring function is comprised of quantitative metrics or **outputs** that assess how well policies addressed the following questions:

- How much congestion did agents and the system as a whole experience during the day?
- What level of service did the transportation system provide to users?
- Did benefits to the transit company exceed its operational costs?
- Has accessibility to important opportunities or points of interest improved?
- Did the city achieve its sustainability goals?

A total of nine metrics will be evaluated to compute the total submission scores. Each metric, $i = 1, \ldots 9$ is a function $F_i(C_s)$ of a set of outputs, C_s, from a BISTRO simulator run. The score corresponding to the metric is a function of the ratio of the value of the metric evaluated for the submission, $F_i(C_s)$, to the value for the same metric evaluated for the BAU scenario, $F_i(C_{BAU})$. The total submission score from a policy is a weighted sum of each standardized score component, $\frac{F_i(C_s)}{F_i(C_{BAU})}$. Standardization procedures are described in detail in Section 2.2.2. A total submission score less than 1 indicates that, under the evaluated policy portfolio, the system is performing better in comparison to the BAU scenario. To win, competitors need to make scores as small as possible, as they represent negative externalities to be reduced.

2.2 Methodology

2.2.1 What differentiates a good score from a bad score?

A dataset of inputs to the model was randomly generated according to pre-specified bounds on all input variables (i.e., frequency adjustment, mode incentives, mass transit fares, and transit vehicle fleet mix). Output statistics from the BISTRO system runs based on these inputs are used to compute quantitative metrics (such as *vehicle miles traveled* (VMT)) representing KPIs of the Sioux Faux transportation system. In order to provide a standard of comparison, the simulation was previously run without providing any additional inputs, a scenario that is representative of the status quo or Business as Usual (BAU).

The *unscaled score* for a particular metric C is computed as the ratio between its output value given simulation using a set of inputs, $F_i(C_s)$, to its output value for the BAU simulation, $F_i(C_{BAU})$:

$$S_{unscaled,i} = \left(\frac{F_i(C_s)}{F_i(C_{BAU})} \right)^{\tau_i}. \tag{1}$$

This score can be interpreted as the improvement of $F_i(C_s)$ over $F_i(C_{BAU})$ where the smaller $S_{unscaled,i}$, the better. In other words, the simulation run under the set of inputs results in an N times improvement for scoring component i over the "do nothing" scenario.

The objective of the Prize is a minimization since an increase in many of the score components represents a situation that is worse than BAU (e.g., increasing VMT over BAU is bad, hence the objective should seek values of S less than 1). To maintain consistency in this regard, we take the reciprocal of the unscaled score for several of the score components that are positively related to desirable outcomes (e.g., improvements in accessibility). The value of $\tau_i \in \{-1, 1\}$ determines whether the metric ratio will be taken as is, or as a reciprocal. Values of τ_i are provided in Table 1. In subsequent figures, one may therefore interpret $S_{unscaled,i} < 1$ (implying $\tau_i = 1$) as better than BAU and $S_{unscaled,i} > 1$ (implying $tau_i = -1$) as worse than BAU.

2.2.2 Standardizing the score components

When appraising alternative scenarios based on multiple quantitative criteria, ensuring that metrics are compared on the same scale is desirable from both a policy perspective as well as for the purposes of optimization. By applying a numerical ranking or preference scheme for metrics, urban planners and other stakeholders can emphasize local needs or regulatory requirements. These weights can be used in our optimization setting, as dimensionless score components are readily combined with these preference rankings (or relative importances) using a weighted average (or weighted sum if positive weights are normalized to sum to 1) to compute a total score. However, combining *unscaled* score components is not advisable in the present setting due to differences in the measurement scale of metrics. Figure 1 shows the typical output ranges for score components computed from simulation runs using a large, randomly sampled dataset of inputs. As can be seen from Figure 1, the spread of some metrics (for example H) can dwarf some others (for example A), hence the need for normalization.

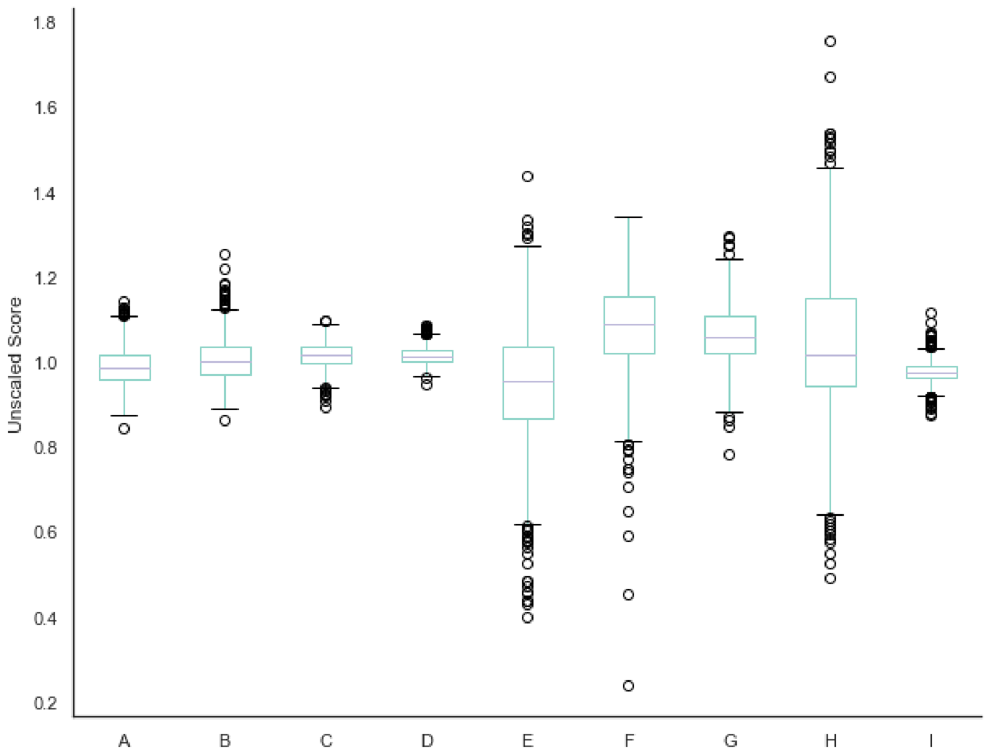


Figure 1: Boxplot of the aggregated, unweighted scoring components.

Key	**Description**
A:	Accessibility: Number of secondary locations accessible within 15 minutes
B:	Accessibility: Number of work locations accessible within 15 minutes
C:	Congestion: average vehicle delay per passenger trip
D:	Congestion: total vehicle miles traveled
E:	Level of service: average bus crowding experienced
F:	Level of service: average trip expenditure - secondary
G:	Level of service: average trip expenditure - work
H:	Mass transit level of service intervention: costs and benefits
I:	Sustainability: Total PM 2.5 Emissions

Statistics from the random sample may be computed to evaluate the applicability of several transformations commonly used to rescale features of data in optimization and machine learning applications. Ultimately, it was found that *standardization* of score components yields numeric scores that are most directly comparable to each other, yet robust to outliers. Standardization assumes that each score component is randomly distributed such that its z-score, z_i, may be computed as follows:

$$z_i = \frac{S_{unscaled,i} - \mu_i}{\sigma_i} \tag{2}$$

where μ_i and σ_i are the mean and standard deviation of, $S_{unscaled,i}$, respectively, computed over its values from the random input sample.

Computing the z-score (as just described) transforms each score component to be normally distributed around 0 with a standard deviation of 1.0 (i.e., each component is rescaled to the unit normal distribution). This transformation is applied to new submissions by subtracting the mean and dividing by the standard deviation of the scoring component distributions generated using the random input sample. Values of μ_i and σ_i are tabulated in Table 1, while Figure 2 illustrates the distribution of scores from the random sample after applying standardization. The result of this process is shown below in Figure 3.

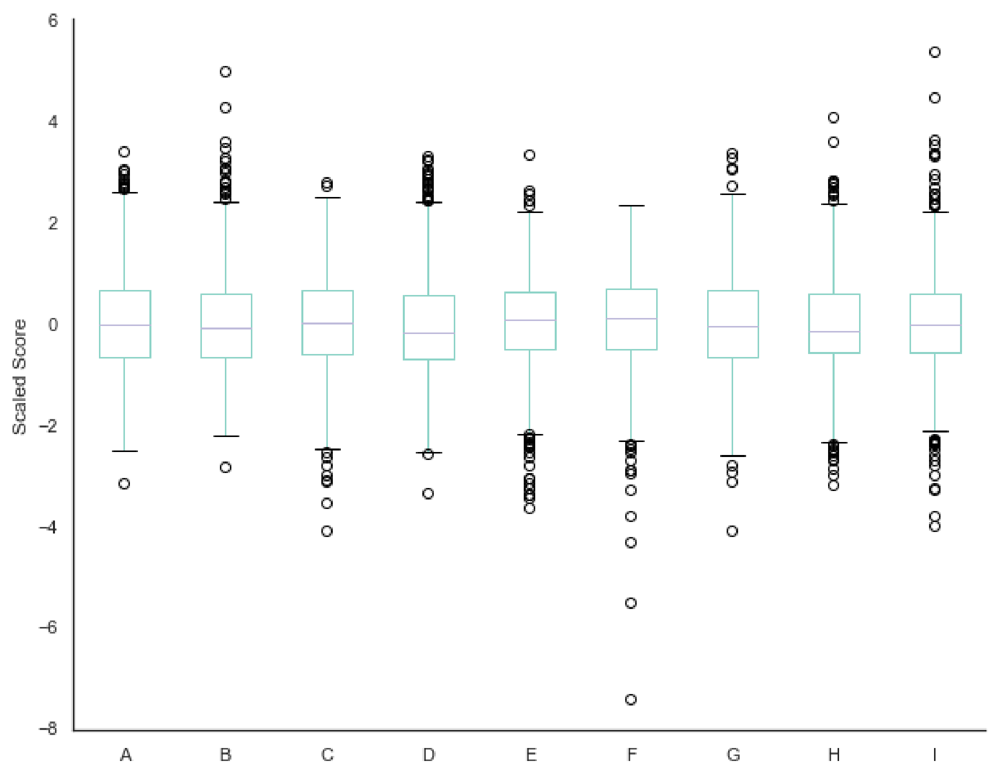


Figure 2: Boxplot of the standardized, unweighted scoring components. (see Key table in Section 2.2.2 above for the description of letters references

Table 1: Statistics computed from random search (757 samples) used for standardization of submission scores

Score Component, $F_i \in \vec{F}$	**Mean, $\mu_i \in \vec{\mu}$**	**Standard Deviation, $\sigma_i \in \vec{\sigma}$**	**Reciprocal, $\tau_i \in \vec{\tau}$**
Accessibility: Work trips	1.004	0.050	-1
Accessibility: Secondary trips	0.988	0.046	-1
LoS: Average travel expenditure - work trips	1.064	0.069	1
LoS: Average travel expenditure - secondary trips	1.077	0.113	1
LoS: Average bus crowding	0.942	0.149	1
Congestion: Total VMT	0.979	0.023	1
Congestion: Average vehicle delay per passenger trip	1.105	0.030	1
Mass transit LoS intervention: Costs and benefits	1.047	0.174	-1
Sustainability: Total $PM_{2.5}$ emissions	0.976	0.026	1

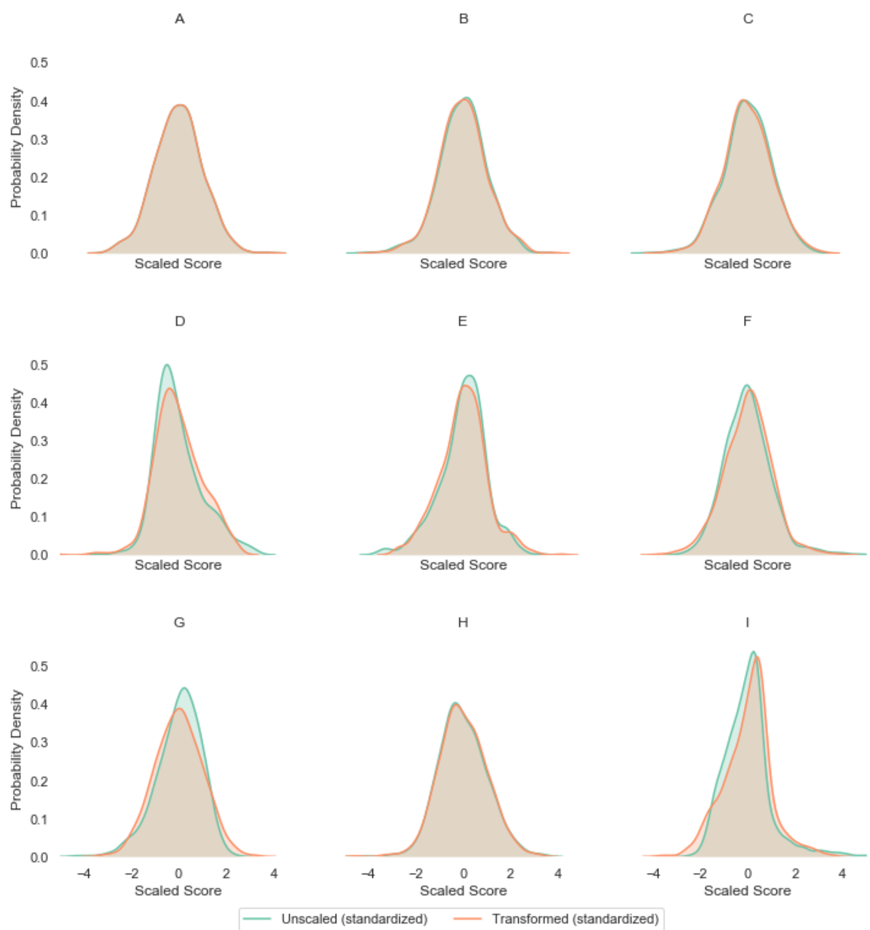


Figure 3: Note: This graphic is presented for illustration purposes only. Standardized score component distributions: unscaled (in blue) and after applying the Box-Cox transformation[a] (in orange). Based on Figure 2, it appears that several of the score components have significant skew and/or kurtosis when standardized (e.g., “Level of service: average trip expenditure - secondary” and “Mass transit level of service intervention: costs and benefits”), which is potentially indicative of a heteroscedastic distribution of scores. The Box-Cox transformation was applied in order to attempt to rescale score component values closer to a normal distribution. However, in the analysis presented and for the time being, we use z-score standardization without first applying the Box-Cox transformation.

[a]Box, George EP, and David R. Cox. “An analysis of transformations.” Journal of the Royal Statistical Society. Series B (Methodological) (1964): 211-252

3 Sample Input Evaluation and Scoring Analysis

The following sections illustrate the performance of two sample inputs (selected from the random input data) that result in appreciably dissimilar transportation outcomes, as well as two additional sets of inputs that were designed to determine the performance of the simulations when only one or two sets of inputs were varied. These are shown as illustrative examples of what Teams could generate to compete in the Prize. The four sets of inputs are meant to be used for illustration purposes only; two of them were chosen based upon their score performance, and the other two are used to see the score performance of policy-driven inputs (these policies are not reflective of what might be termed a "good" or "thoughtfully crafted" policy, rather it is our goal to have their intentions be easily interpretable). Further inputs and outputs from the BAU scenario are included to further assist in comparison and interpretation of inputs, outputs, and score components for the other scenarios. Table 2 presents each set of inputs and the rationale for choosing them.

Table 2: Input set names and reasons for inclusion.

Input Set Name (number)	**Motivation**
BAU (1)	To present a baseline for comparison
"Transit Adverse" [1] (2)	Designed to produce the worst quality bus service. We specify 2 hr vehicle headways and maximal fares for all age groups.
"Redistributive Effects (Transit and On-Demand)" (3)	Designed to increase transit fares for adults (aged 18-65) and offer transit and on-demand incentives to anyone earning less than $20,000/year.
"VMT Reducing" (4)	Achieved the best reduction in VMT score within the random search
"Transit Operating Cost Inefficient" (5)	Achieved the worst Cost-Benefit score within the random search

[1] Here the phrase "Transit Adverse" is not used in a strict literal sense, but instead to imply that this set of inputs was designed to produce the lowest quality bus service possible, without removing it entirely.

3.1 Sample Inputs

Inputs for Phase I target improving transit operations, vehicle fleet assignment, and incentives to use non-motorized or on-demand transport.

Teams may choose to optimize any and all of the following inputs:

- The bus fleet composition;
- The frequency of buses on routes;
- The distribution of incentives for agents using on-demand car-sharing and/or mass transit;
- Mass transit fares.

3.1.1 Vehicle fleet mix

Assumptions

- Sioux Faux has 12 bus lines operating in the city (see Figure 4 below).
- In the BAU scenario, all buses are `DEFAULT` buses.
- Contestants can change the bus fleet for each bus line, selecting vehicles from four types of buses, described in Table 3 below, each of them with different technical properties.

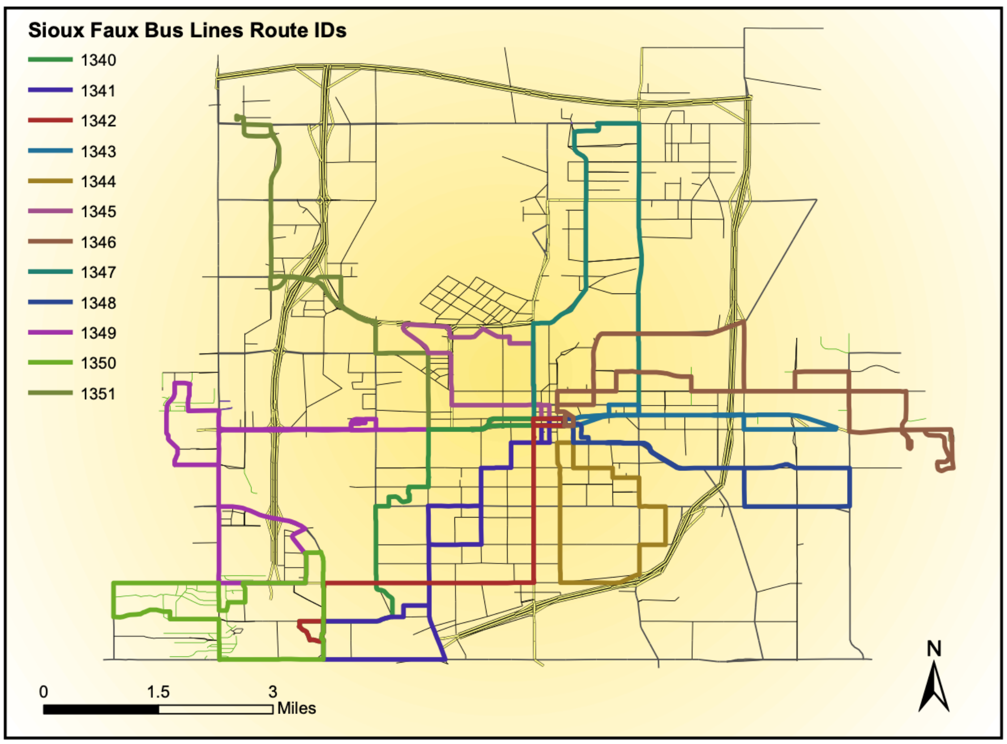


Figure 4: Sioux Faux Bus Lines' (SFBL) Route Map.

Input constraints

- Each route is assigned only one type of bus.
- All trips on a route will use the selected bus for that route

Table 3: Bus vehicle types

Vehicle type, $c \in C$	**Fuel type**	**Fuel consumption rate ($Joule/meter$)**	**Operational cost ($\$/hour$)**	**Seating capacity**	**Standing capacity**
CAR	gasoline	3655.98	n/a	4	0
BUS-DEFAULT	diesel	20048	89.88	37	20
BUS-SMALL-HD	diesel	18043.2	90.18	27	10
BUS-STD-HD	diesel	20048	90.18	35	20
BUS-STD-ART	diesel	26663.84	97.26	54	25

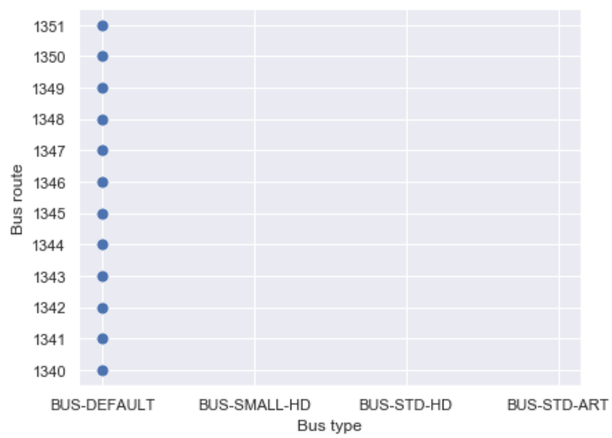
- This input will be applied after the frequency adjustment input below

Visualization of the control variables

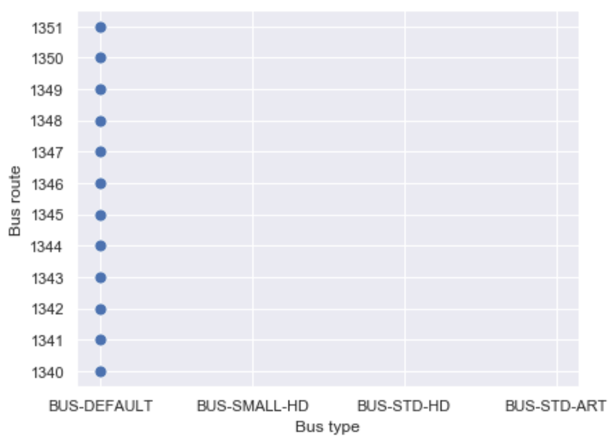
- For each bus line (y-axis), a blue point indicates which vehicle type (x-axis) is operating on the line.

Interpretation of the graphs

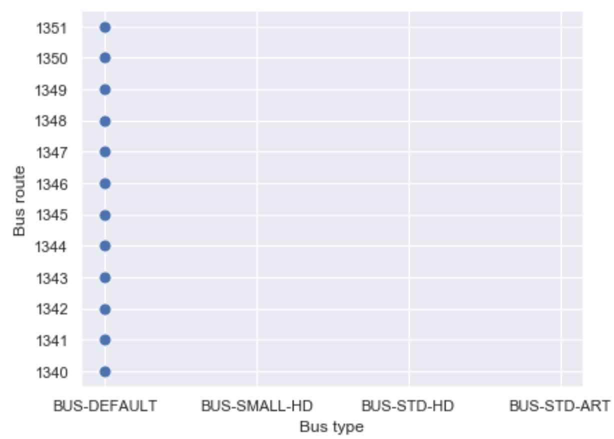
- The “Transit Adverse” and “Redistributive for Transit and On-Demand Rideshare” scenarios’ bus fleets stay unchanged from the BAU scenario.
- For “VMT Reducing”, lines 1346 and 1343 operate with `BUS-STD-HD` buses. All other lines operate with `DEFAULT` buses.
- For “VMT Reducing”, lines 1341 and 1344 operate with `BUS-STD-HD` buses. All other lines operate with `DEFAULT` buses.



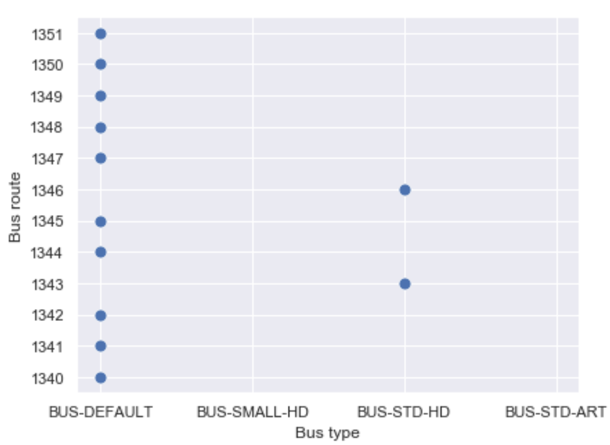
(1) Vehicle Fleet Mix for the BAU scenario.



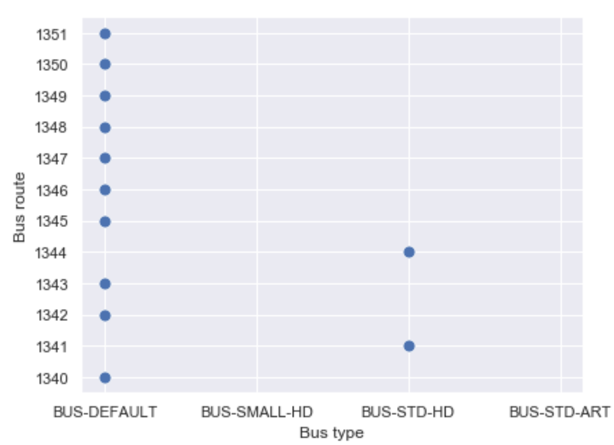
(2) Vehicle Fleet Mix for the “Transit Adverse” scenario.



(3) Vehicle Fleet Mix for the “Redistributive for Transit and On-Demand Rideshare” scenario.



(4) Vehicle Fleet Mix for the “VMT Reducing” scenario.



(5) Vehicle Fleet Mix for the “Transit Operating Cost Inefficient” scenario.

Figure 5: Vehicle fleet mix input for (1) the baseline scenario, (2) the “Transit Adverse” scenario, (3) the “Redistributive for Transit and On-Demand Rideshare” scenario, (4) the “VMT Reducing” scenario, (5) the “Transit Operating Cost Inefficient” scenario.

3.1.2 Mass transit fare

Assumptions

- The contestants can change the bus fare structure and its distribution among age groups.
- In the BAU scenario, the fare structure is distributed as follows: Children 5 years and under and persons over 65 years are not charged any fare; riders 6 to 10 years pay a fare of $0.75, while riders between the ages of 10 and 65 pay a fare of $1.50.

Input constraints

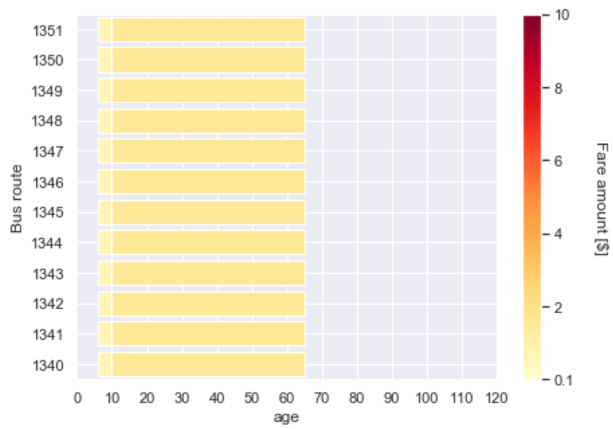
- The fares must be greater than 0 and and may not exceed a maximum of $10/person-trip.
- The bus fares may not isolate a single age; fares must be defined in bins no smaller than five years in range.

Visualization of the control variables

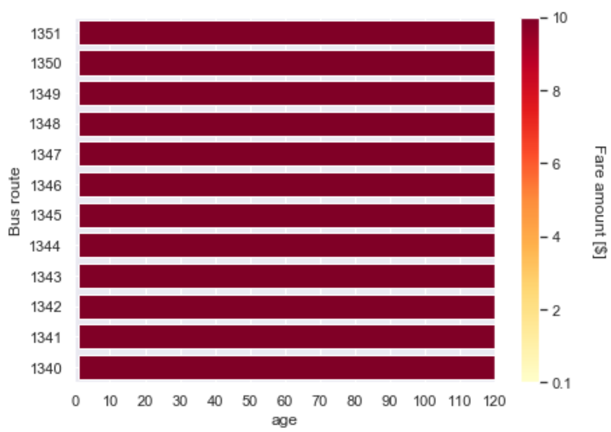
- For each bus line (y-axis), the fare amount is represented by the color of the bar (color scale). The length of the bar defines the age group (x-axis) to which the fare applies.

Interpretation of the graphs

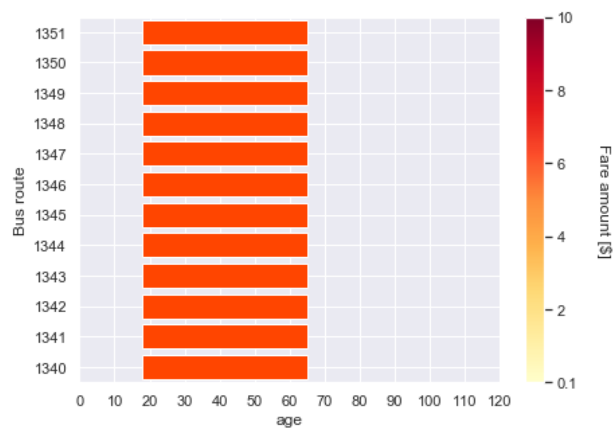
- When there is no bar or color, it means that taking the bus is free.
- The “Transit Adverse” scenario applies a maximum $10 fare to everyone.
- The "Redistributive for Transit and On-Demand Rideshare" scenario imposes a $6 fare for people 20 to 65 years old.
- For the “VMT Reducing” and “Transit Operating Cost Inefficient” scenarios, the bus fare amount and to which age group it applies differ for each bus route. For instance in the “VMT Reducing” scenario, only people 50 to 55 years pay a fare ($6) to take the 1349 bus.



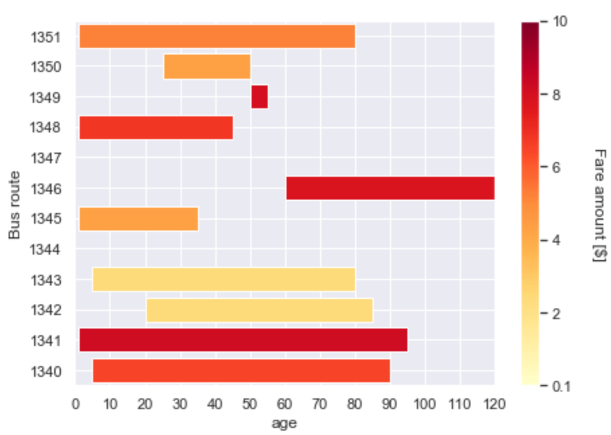
(1) Mass Transit Fare for the BAU scenario.



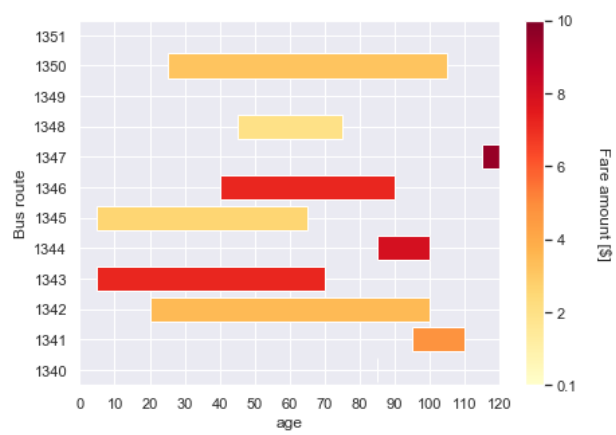
(2) Mass Transit Fare for the "Transit Adverse" scenario.



(3) Mass Transit Fare for the "Redistributive for Transit and On-Demand Rideshare" scenario.



(4) Mass Transit Fare for the "VMT Reducing" scenario.



(5) Mass Transit Fare for the "Transit Operating Cost Inefficient" scenario.

Figure 6: Horizontal bar graph of the Mass Transit Fare input (1) the baseline scenario, (2) the "Transit Adverse" scenario, (3) the "Redistributive for Transit and On-Demand Rideshare" scenario, (4) "VMT Reducing" scenario, (5) "Transit Operating Cost Inefficient" scenario.

3.1.3 Mode Incentives

Assumptions

- The contestants can specify the incentive amounts and distribution (among ages and incomes) for the following travel modes:
  - `walk_transit`: walking as an access/egress mode to/from transit.
  - `drive_transit`: use of the personal car as an access/egress mode to/from transit.
  - `OnDemand_ride`: use of on-demand rideshare as the main transport mode for the trip.
- In the BAU scenario, no incentives are provided.

Input constraints

- The incentive may vary from $0 to $50.
- If the incentive is greater than the cost of travel, Agents are credited incentives after their travel (a rebate, rather than cash up-front).
- Incentives must be defined in bins no smaller than 5 years for an age range and $5,000 for an income group.

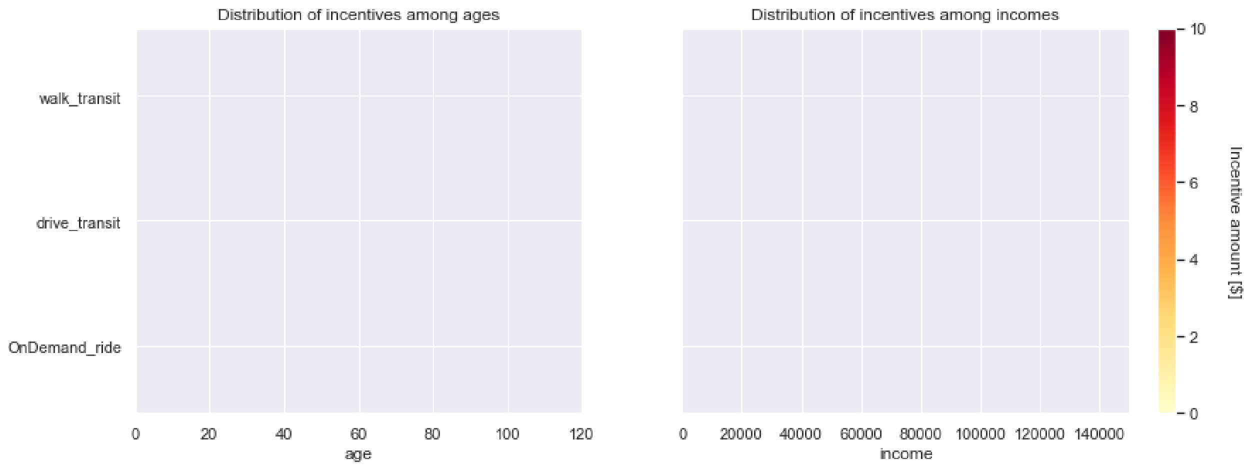
Visualization of the control variables

- The y-axis represents the modes that can get an incentive: `walk_transit`, `drive_transit` and `OnDemand_ride`
- For each scenario, a pair of graphs is representing the incentives structure: their distribution among age groups (x-axis on the left) and among income groups (x-axis on the right).
- The colors (color scale) represent the amount of the incentive provided and the length of the horizontal bars describes the incentive distribution among age and income groups (x-axis).
- When there is no bar or color in front of a mode (y-axis), it means that there is no incentive for this mode.
- The two graphs are cumulative, meaning that for each mode, people getting an incentive must satisfy both age and income qualifications.
- People who meet the qualifications for multiple incentives accumulate them. It means that one could receive an incentive greater than the monetary cost of travel. Agents are credited incentives after their travel (a rebate, rather than cash up-front).

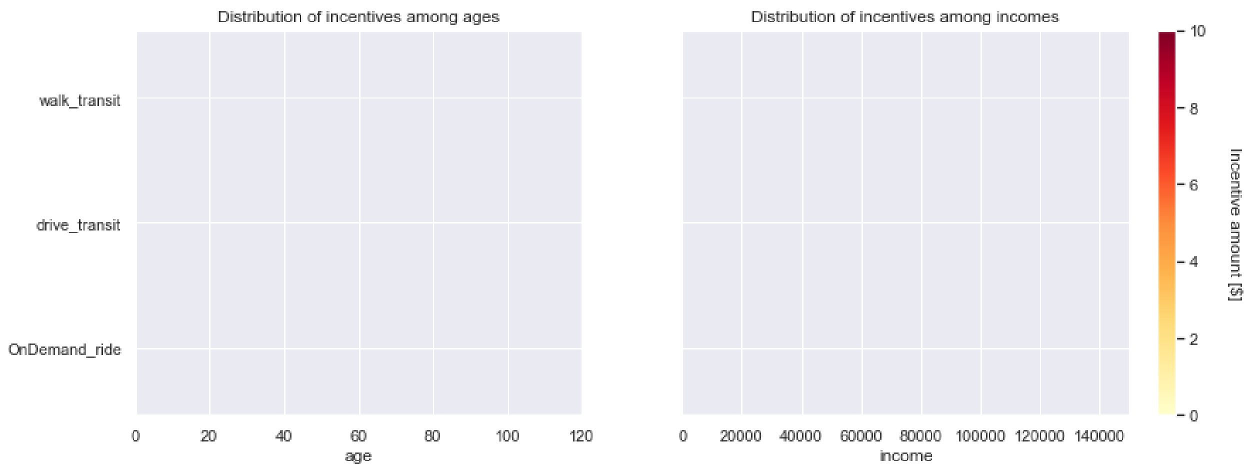
Interpretation of the graphs

- As for the BAU scenario, the "Transit Adverse" scenario does not provide any incentives.
- The "Redistributive for Transit and On-Demand Rideshare" scenario gives a $6 incentive for all three modes to anyone earning between 0 and $20,000 per year.
- For the "VMT Reducing" scenario:
  - People older than 10 and up to 45 years of age and earning between $35,000 and $60,000/year as well as people older than 45 and up to 100 years of age and earning

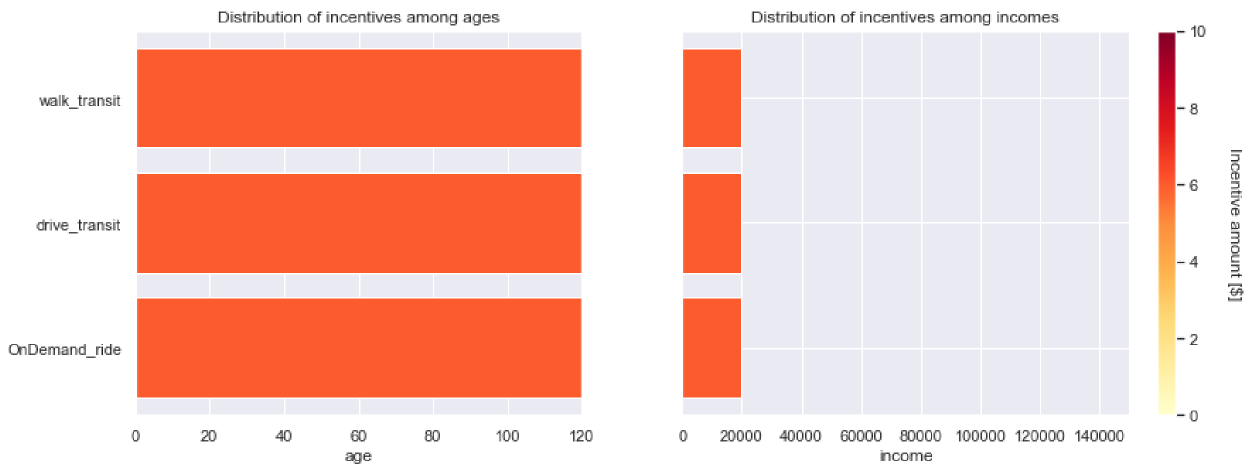
between \$95,000 and \$115,000/year receive respectively a \$7 and \$5 incentive for `walk_transit`.
  - People older than 5 and up to 50 years if age and earning between \$20,000 and \$90,000/year receive a \$5 incentive for `OnDemand_ride`.
- For the "Transit Operating Cost Inefficient" scenario:
  - People older than 10 and up to 90 years of age and earning between \$60,000 and \$130,000/year receive a \$4 incentive for `walk_transit`.
  - Children older than 5 and up to 25 years of age and earning between \$45,000 and \$90,000/year receive a \$4 incentive for `drive_transit`.
  - People 75 years old or younger and earning between \$125,000 and \$135,000/year receive a \$4 incentive for `OnDemand_ride`.



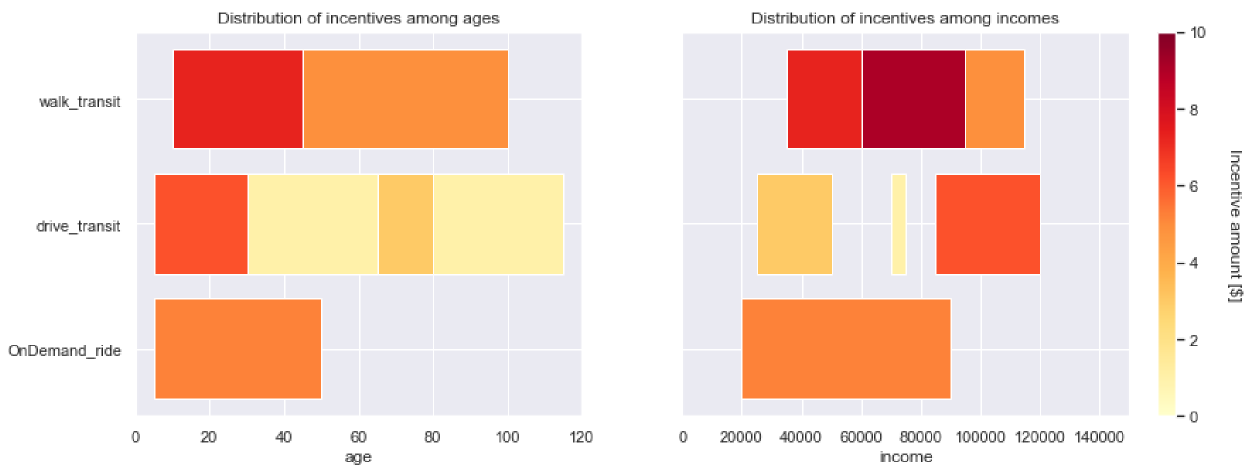
(1) Mode Incentives for the BAU scenario.



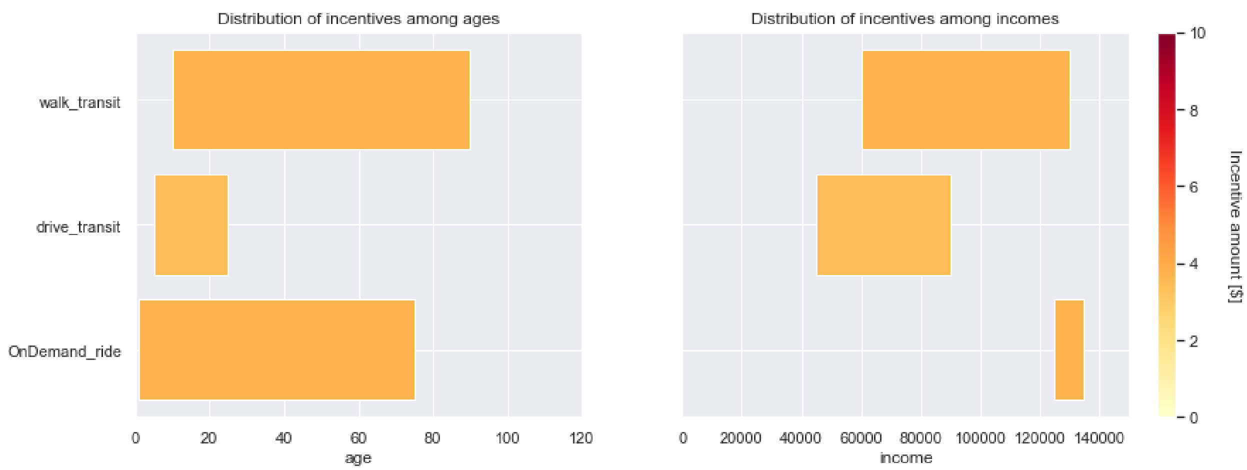
(2) Mode Incentives for the “Transit Adverse” scenario.



(3) Mode Incentives for the “Redistributive for Transit and On-Demand Rideshare” scenario.



(4) Mode Incentives for the "VMT Reducing" scenario.



(5) Mode Incentives for the "Transit Operating Cost Inefficient" scenario.

Figure 7: Horizontal bar graph of the Mode Incentives input for (1) the baseline scenario, (2) the "Transit Adverse" scenario, (3) the "Redistributive for Transit and On-Demand Rideshare" scenario, (4) "VMT Reducing" scenario, (5) "Transit Operating Cost Inefficient" scenario.

3.1.4 Mass Transit Frequency Adjustment

Assumptions

- The BAU scenario, bus schedules are *non-frequency schedules*, meaning that buses do not respect a regular headway but arrive and depart stops on trips as specified in the schedule found in a recent (2018) release of the Sioux Area Metro's *Generalized Transit Feed Specification* (GTFS) data acquired from the Transitland site.
- Any modification to the baseline in the input file (i.e. the definition of a fixed headway for certain service periods) for a route will clear all initial non-frequency schedules on the route and impose a *frequency-based schedule* for the specified service periods. It means that for each modified route, no bus is operating during the time periods outside of the newly defined service periods. It is up to the contestant to decide if the new service periods should connect to provide a continuous bus service or not.

Input constraints

- There can be no more than five distinct bus service periods per route (this mimics the typical delineation: am peak, midday, pm peak, evening, late night/early morning).

Visualization of the control variables

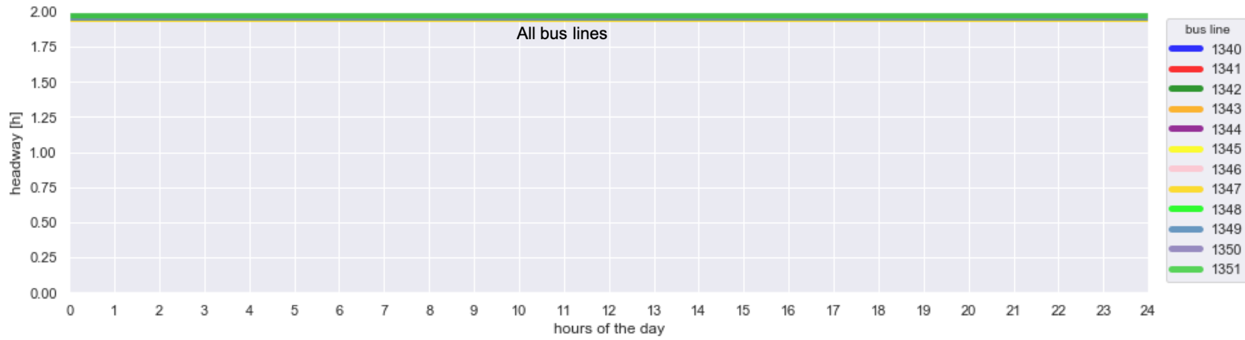
- The lines on each graph represent the modified service period(s) (hours on the x-axis) for each bus line (see colors in legend), with the corresponding headway (in hours) depicted by the y-axis. When there are no lines at all corresponding to a route displayed on the graph, the buses for trips on that route follow the BAU schedule.

Interpretation of the graphs

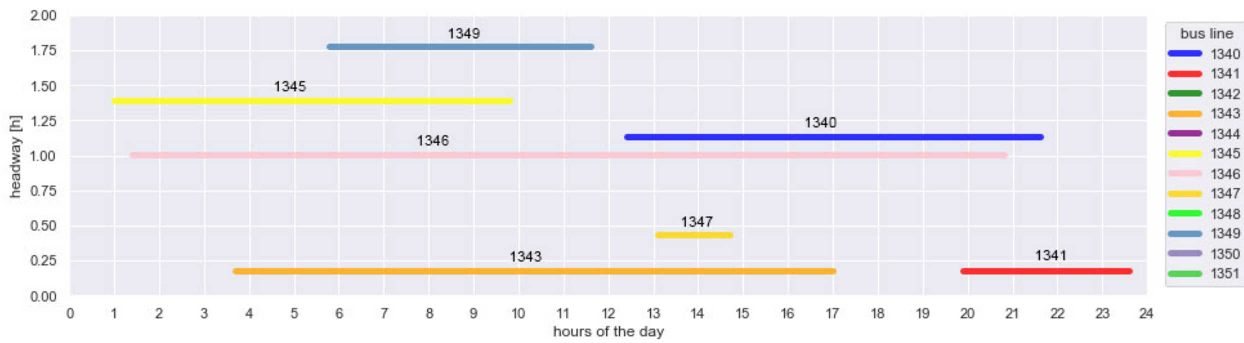
- In the "Redistributive" scenario, buses follow the same non-frequency schedule.
- In the "Transit Adverse" scenario, all buses operate during the whole day with a 2:00 headway.
- In the "VMT Reducing" scenario, 7 lines out of 12 see their schedule changed to a frequency-based one. For instance, the 1349 bus (steel blue) now operates from 5:45 am until 11:45 am with a 1:45 headway. It does not operate during the rest of the day.
- In the "Transit Operating Cost Inefficient" scenario, 10 lines out of 12 see their schedule changed to a frequency-based one. For instance, the 1348 bus (lime green) operates from 3:25 am until 6:25 pm with a 1:50 headway. It does not operate during the rest of the day.

Routes	Service periods	Headway
all	Non frequency schedule	Non frequency schedule

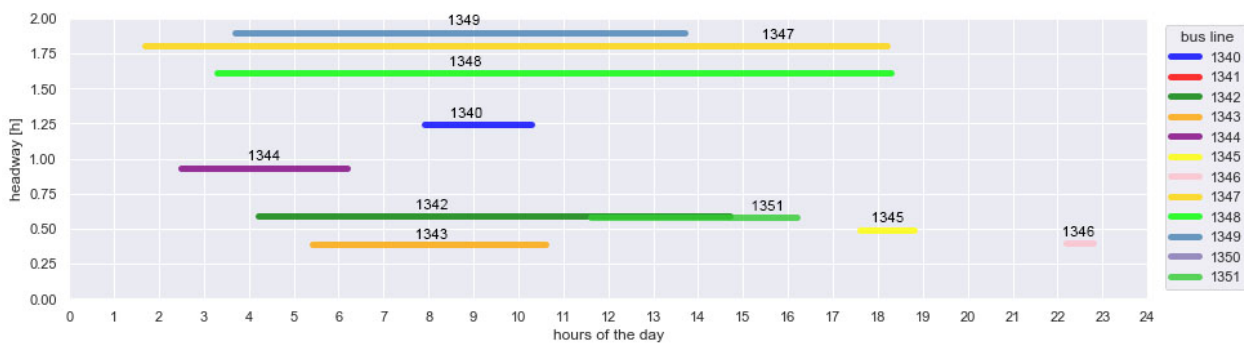
Bus Frequency for the BAU (1) and "Redistributive for Transit and On-Demand Rideshare" (3) scenario.



(3) Bus Frequency Adjustment for the "Redistributive for Transit and On-Demand Rideshare" scenario.



(4) Bus Frequency Adjustment for the "VMT Reducing" scenario.



(5) Bus Frequency Adjustment for the "Transit Operating Cost Inefficient" scenario.

Figure 8: Figure 8: Horizontal bar graph of the Frequency Adjustment input for (1) the baseline scenario, (2) the "Transit Adverse" scenario, (3) the "Redistributive for Transit and On-Demand Rideshare" scenario, (4) the "VMT Reducing" scenario, (5) the "Transit Operating Cost Inefficient" scenario.

3.2 Sample Outputs

When interpreting these outputs, it is important to restate the caveat that these results do not necessarily support causal conclusions about the influence of one input on individual outputs. The policy-driven sets of inputs may seem to demonstrate slightly more causal input-output relationship; however, their intent is merely to illustrate that the behavior of the system seems qualitatively realistic.

3.2.1 Mode choice

Modal split

Figure 9 illustrates the overall modal split of `Agents` over the day.

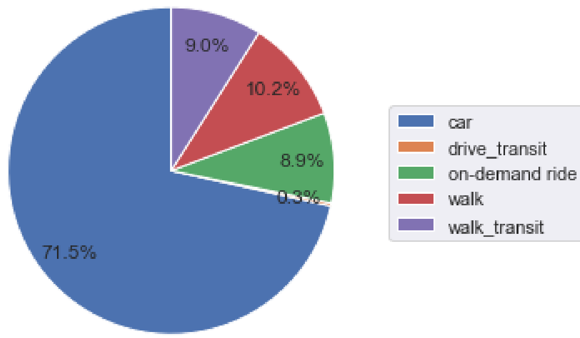
Representation

The pie charts represent the proportion (in %) of people using each mode (identified by colors).

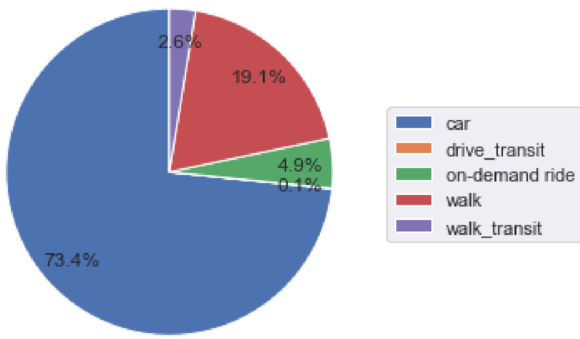
Interpretation of the graphs

Discussing the results first in terms of outputs, we can see that the designed solutions are performing predominantly as expected. The “Transit Adverse” inputs decrease bus mode share while increasing all others, particularly walking. The “redistributive” inputs decrease the auto and on-demand ride mode shares in favor of transit and walking; this result potentially signals that agents are sensitive to the incentive for transit.

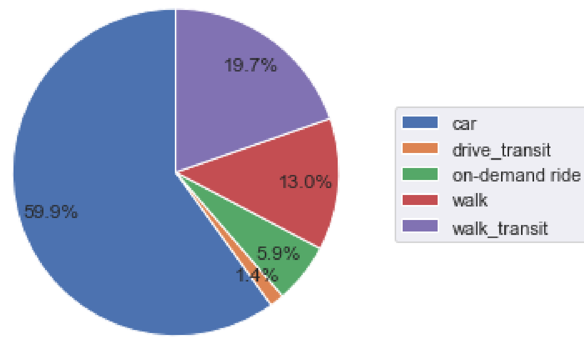
“VMT Reducing”, which achieved the best VMT score, favors walking to transit, while decreasing all other mode shares. “Transit Operating Cost Inefficient”, although scoring the worst on mass transit costs and benefits, increases the bus and walk mode shares while decreasing car and on-demand ride mode shares. This could potentially indicate the over-provision of bus service and/or fare setting inefficiencies.



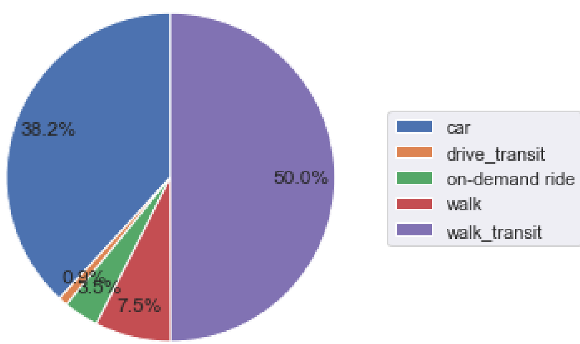
(1) Modal split for the BAU scenario.



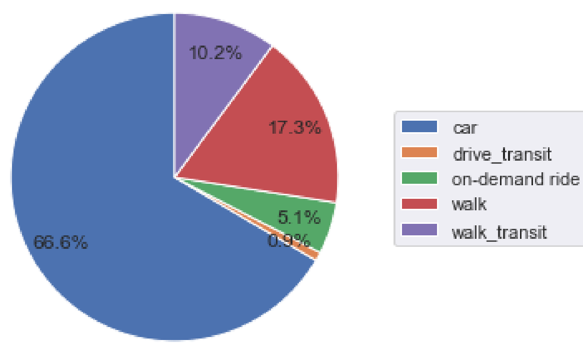
(2) Modal split for the “Transit Adverse” scenario.



(3) Modal split for the "Redistributive for Transit and On-Demand Rideshare" scenario.



(4) Modal split for the “VMT Reducing” scenario.



(5) Modal Split for the “Transit Operating Cost Inefficient” scenario.

Figure 9: Pie chart of the mode choice for: (1) the baseline scenario, (2) the “Transit Adverse” scenario, (3) the “Redistributive for Transit and On-Demand Rideshare” scenario, (4) the “VMT Reducing” scenario, (5) the “Transit Operating Cost Inefficient” scenario.

3.2.2 Accessibility

Access points

Figure 10 illustrates the accessibility of points of interests in Sioux Faux, i.e. the sum of the average number of work or secondary locations accessible from each node of the network by modes using the road network within 15 minutes during the AM peak (7-10 am) and PM peak (5-8 pm) periods.

Spatial representation

The map represents Sioux Faux's road network, each point being a node. For the BAU scenario (1), the color represents the amount of points of interest (secondary locations on the left plot and work locations on the right plot) reachable within 15 minutes from each node. For instance, yellow nodes have a high accessibility whereas dark blue nodes have a low accessibility to these facilities.

For the two bottom plots ((4) & (5)), the color of each node defines the difference in accessibility to points of interest from this node over BAU (normalized to 0). Concretely, it means that:

- If the node is yellow (difference > 0), the accessibility from the node was improved over the BAU scenario;
- If the node is turquoise (difference $= 0$), the accessibility from the node stays unchanged over the BAU scenario;
- If the node is dark blue (difference < 0), the accessibility from the node is decreased over the BAU scenario.

Interpretation

Only results for the random search input sets are used to demonstrate accessibility (Figure 9). The unscaled BAU accessibility values are shown. The "VMT Reducing" scenario slightly increases both work and secondary location accessibility scores while the "Transit Operating Cost Inefficient" scenario slightly decreases them.

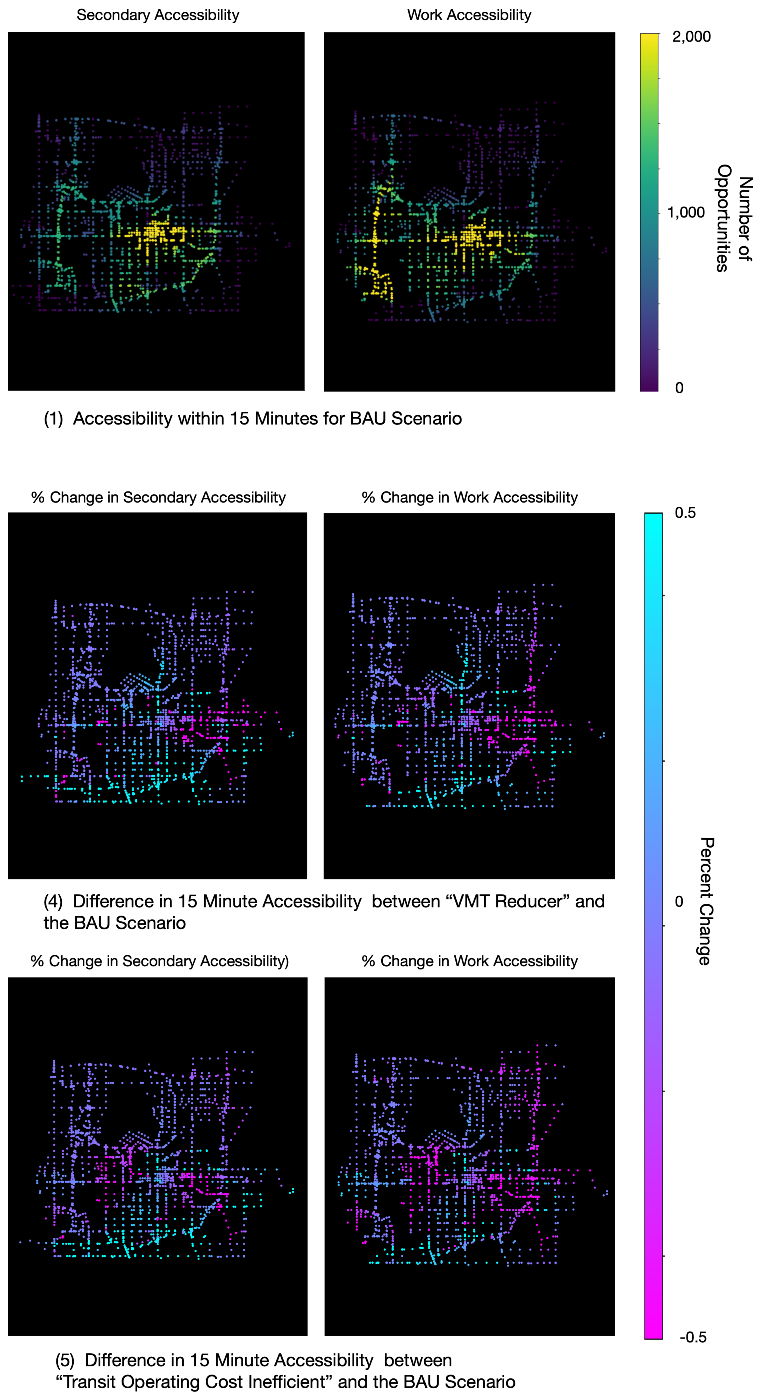


Figure 10: Spatial representation of the Accessibility measured in terms of number of work and secondary activity facilities reachable from each network node within 15 minutes for (1) the BAU scenario; percent change in accessibility over the BAU scenario for (4) the "VMT Reducing" scenario, (5) the "Transit Operating Cost Inefficient" scenario.

3.2.3 Congestion: travel time

Travel Time

Figure 11 is displaying the average travel time per trip[2] and by mode.

Bar graph

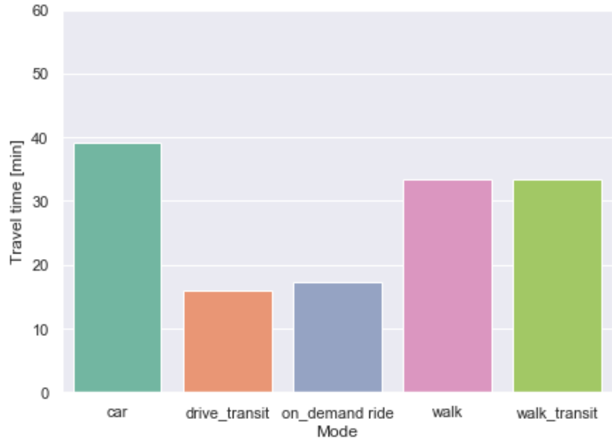
The y-axis depicts the average travel time (in minutes) of a trip using a certain mode (defined on the x-axis and by colors).

Interpretation of the graphs (Figures 11 & 12)

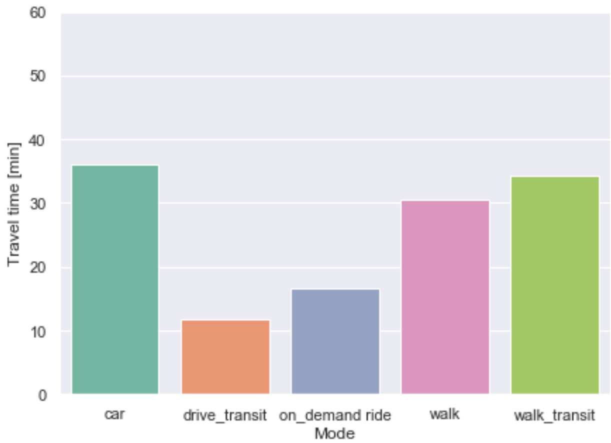
Examining the average travel times per trip for all modes in the "Transit Adverse" case, all average travel times decrease or remain stable save for walking to transit. Paired with the increase in car and walk mode shares, this signals a decrease in overall vehicle miles traveled due to minimal bus service. Similarly, for walking trips, the increase in mode share and decrease in average travel time may be due to a steady trip distance for walking agents. In the "Redistributive" example, the walk to transit mode share increases while travel time remains nearly constant; this may be due to agents substituting vehicle trips for walking to transit as a response to the incentives.

Looking at the random samples, the "VMT Reducing" scenario achieved a drastic decrease in car mode shares as many of those trips were replaced by walking to transit. This resulted in a slight increase in overall VMT, but a reduction in delay per trip, likely due to increased bus service and a subsequent high bus use by agents, even for longer trips. The "Transit Operating Cost Inefficient" scenario resulted in increased walk and walk to transit mode shares; average trip times decreased and increased, respectively. Drive to transit average travel times increased greatly, indicating an agent preference for longer trips drives to bus stops over BAU.

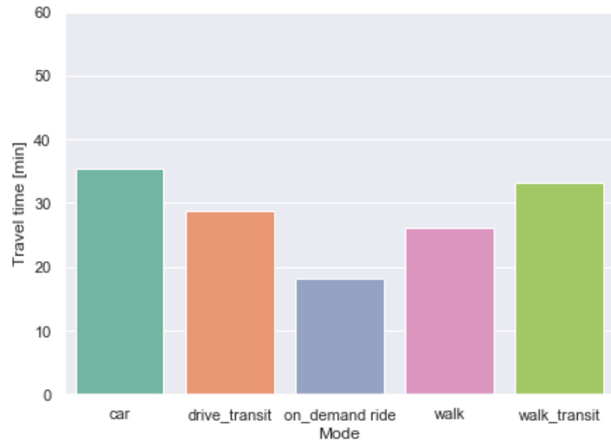
[2] A *trip* is defined as the travel between the origin activity and the destination activity. A *trip* is then made of several *legs* if the agent is making a multimodal trip.



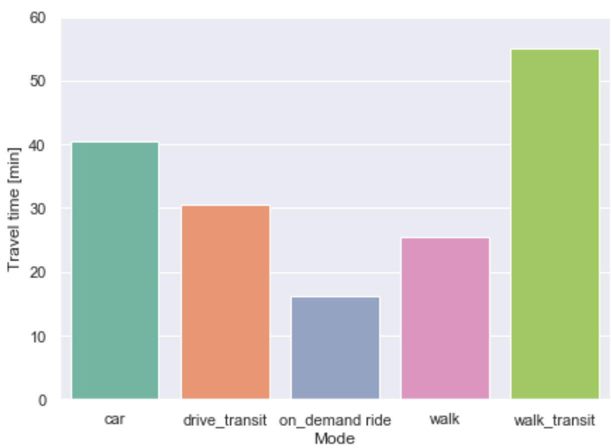
(1) Average travel time per trip and by mode for the BAU scenario.



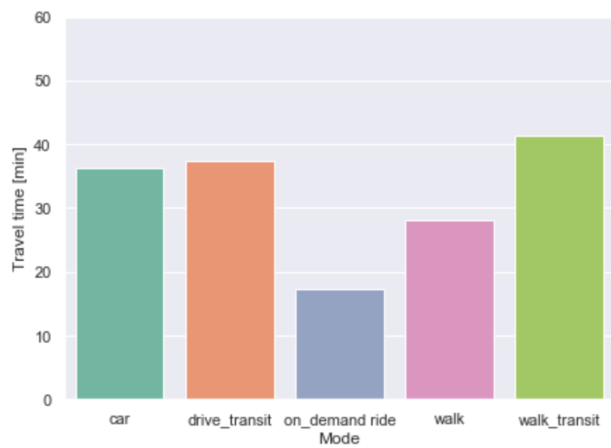
(2) Average travel time per trip and by mode for the “Transit Adverse” scenario.



(3) Average travel time per trip and by mode for the "Redistributive for Transit and On-Demand Rideshare" scenario.



(4) Average travel time per trip and by mode for the “VMT Reducing” scenario.



(5) Average travel time per trip and by mode for the “Transit Operating Cost Inefficient” scenario.

Figure 11: Bar graph of the average travel time per trip and by mode as an output of BISTRO for (1) the BAU scenario, (2) the “Transit Adverse” scenario, (3) the “Redistributive for Transit and On-Demand Rideshare” scenario, (4) the “VMT Reducing” scenario, (5) the “Transit Operating Cost Inefficient” scenario.

Travel time history

Figure 12 displays the average travel time per trip[3] and by mode throughout the day, on a one hour increment.

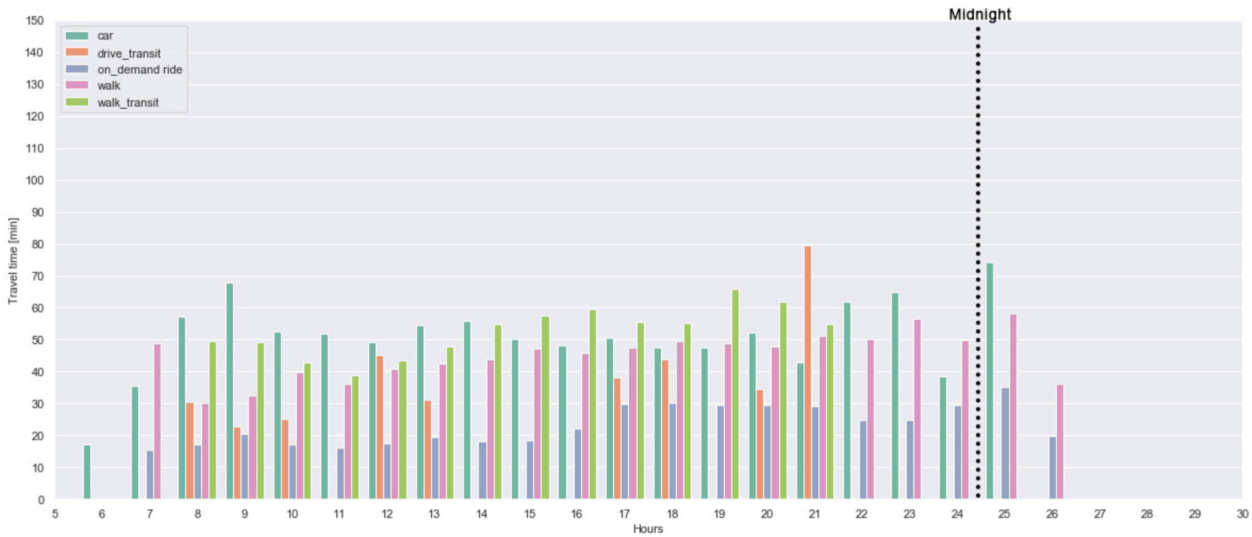
Time series

The y-axis depicts the average travel time (in minutes) of a trip using a certain mode (defined on the x-axis and by colors).

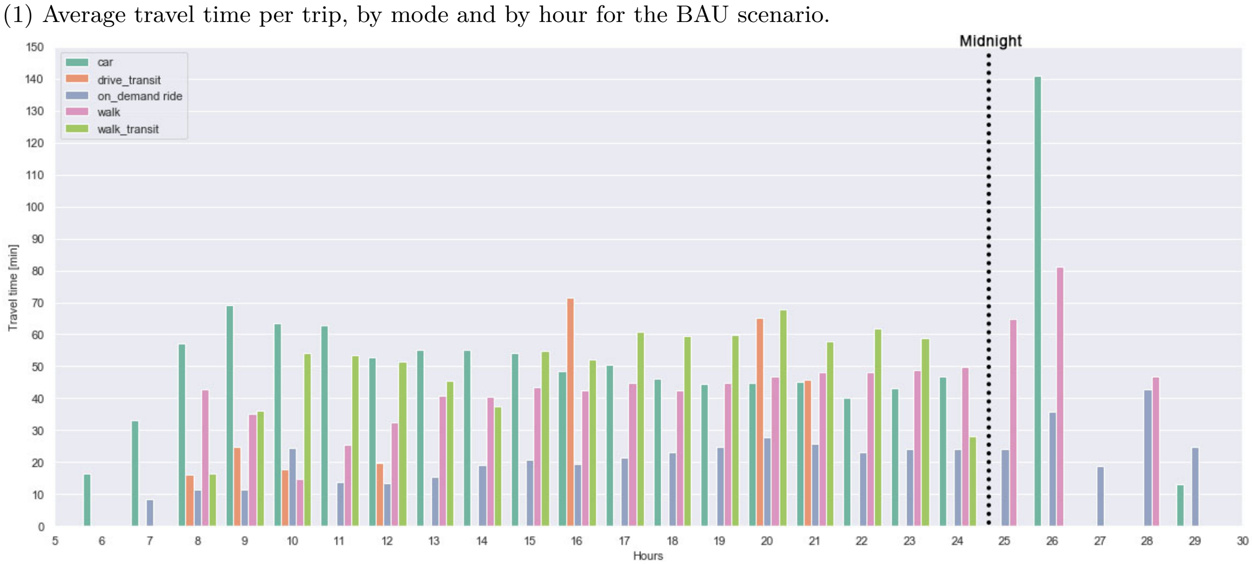
Interpretation of the graphs

See Interpretation in Section 3.2.3 of Figure 11 above.

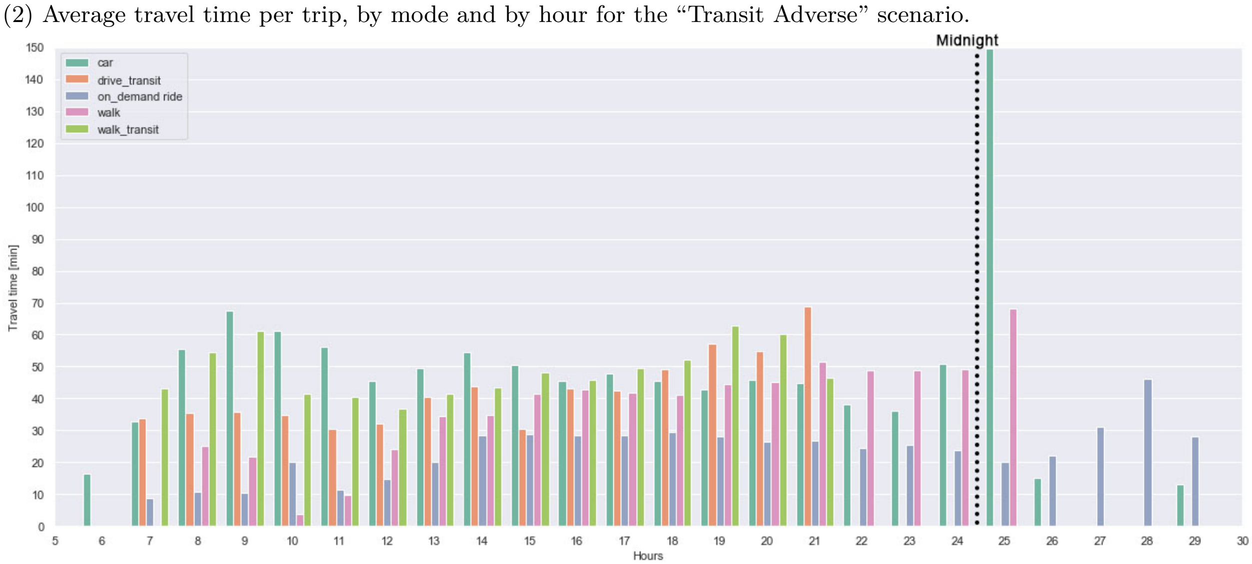
[3] A *trip* is defined as the travel between the origin activity and the destination activity. A *trip* is then made of several *legs* if the agent is making a multimodal trip.



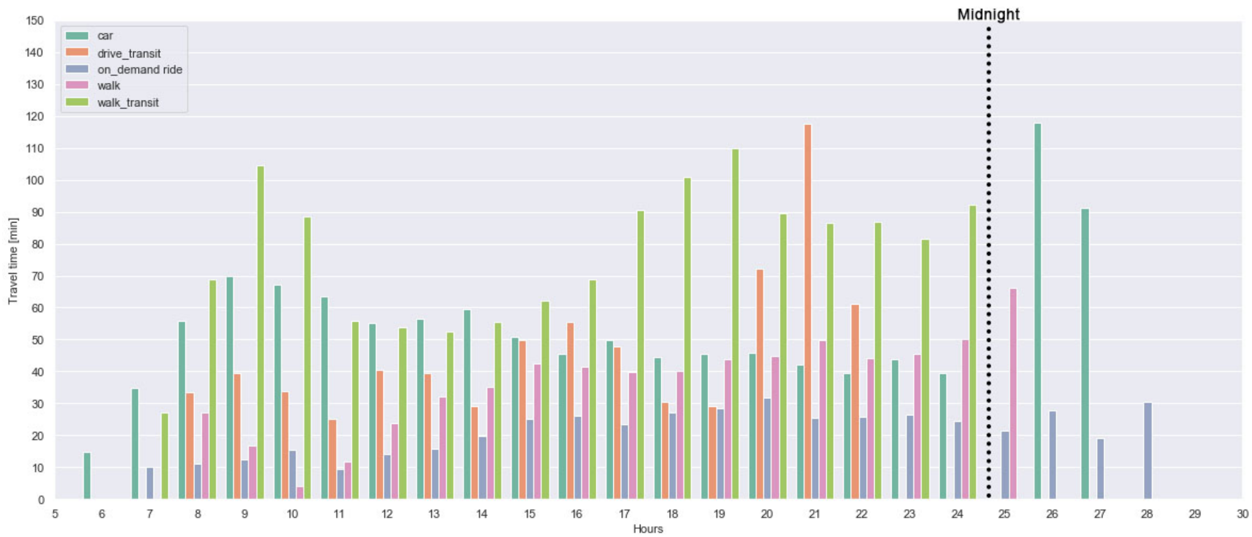
(1) Average travel time per trip, by mode and by hour for the BAU scenario.



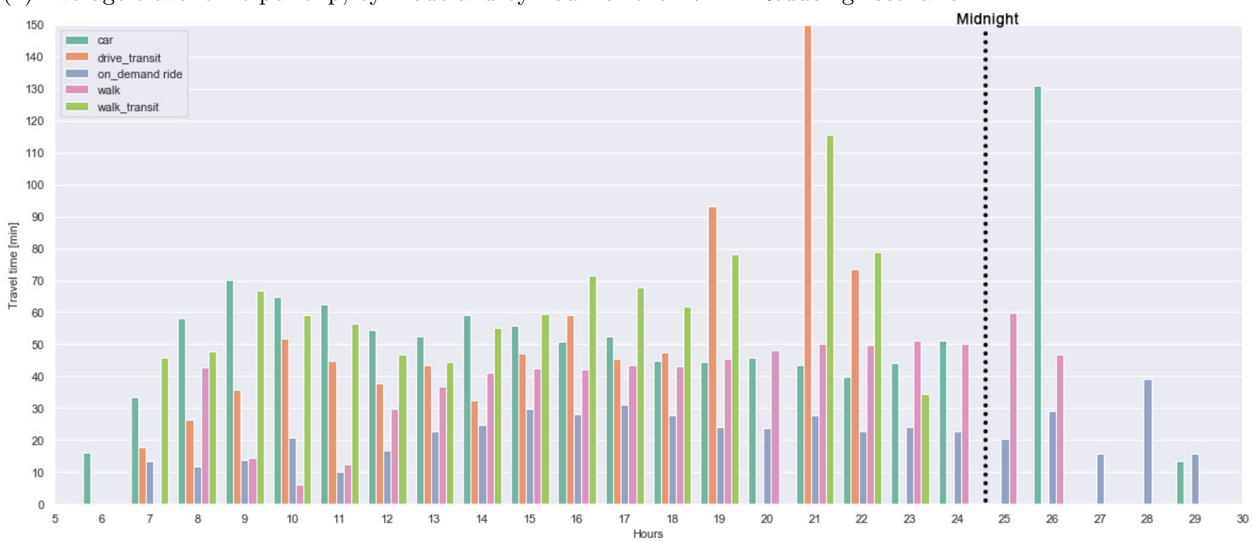
(2) Average travel time per trip, by mode and by hour for the “Transit Adverse” scenario.



(3) Average travel time per trip, by mode and by hour for the "Redistributive for Transit and On-Demand Rideshare" scenario.



(4) Average travel time per trip, by mode and by hour for the “VMT Reducing” scenario.



(5) Average travel time per trip, by mode and by hour for the “Transit Operating Cost Inefficient” scenario.

Figure 12: Bar graph of the average travel time per trip, by mode, and by hour as an output of BISTRO for (1) the BAU scenario, (2) the “Transit Adverse” scenario, (3) the “Redistributive for Transit and On-Demand Rideshare” scenario, (4) the “VMT Reducing” scenario, (5) the “Transit Operating Cost Inefficient” scenario.

3.3 Additional Outputs

The additional outputs presented in the following three sections subsections demonstrate examples of the types of graph BEAM produces. While these figures may not explicitly add to the interpretations from the discussions in the previous section, they are included for the benefit of understanding the inner workings of BEAM and BISTRO.

3.3.1 Passengers Per Bus

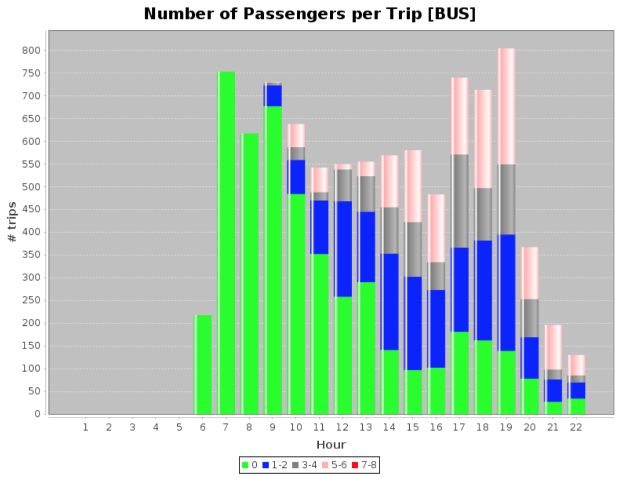
Ridership

Figure 13 illustrates the number of passengers per bus throughout the hours of the day.

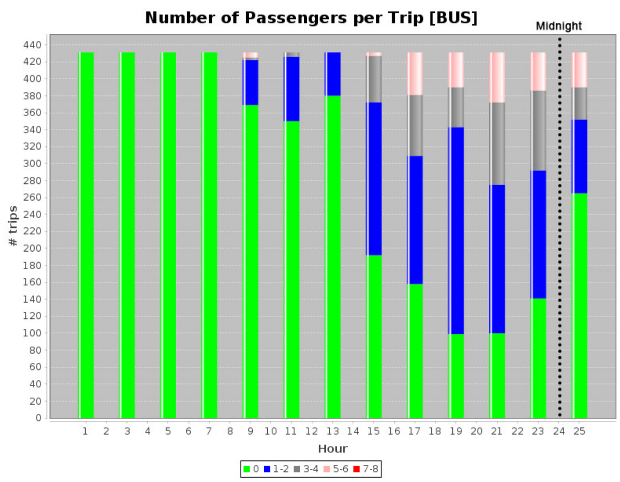
Time history

For each hour of the day (x-axis), the bars heights (y-axis) describe the number of bus trips (directly related to the buses frequencies) realized by the whole bus fleet within this concerned hour. The average number of passengers in a bus during this hour of the day is shown by the colors. Note that the numbers of passengers are very low (from 0 to 8) because the whole transportation system is scaled to a 15k (instead of 157k) population for Sioux Faux. The color patterns works as follows:

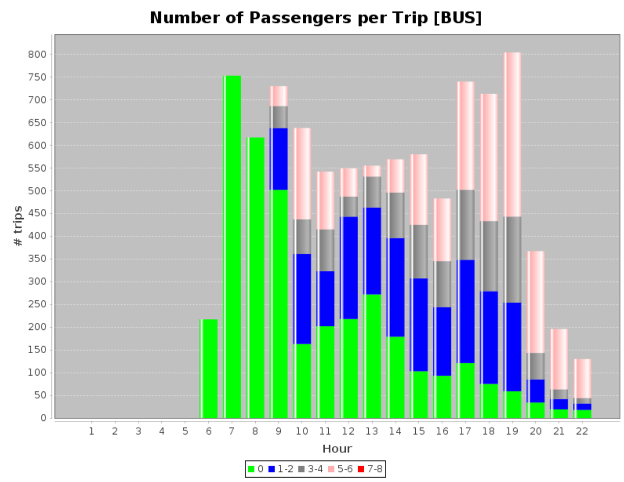
- green (0 pax): the bus is empty.
- blue (1-2 pax): the bus is lightly crowded.
- grey (3-4 pax): the bus is normally crowded.
- pink (5-6 pax): the bus is crowded.
- red (7-8 pax): the bus is very crowded.



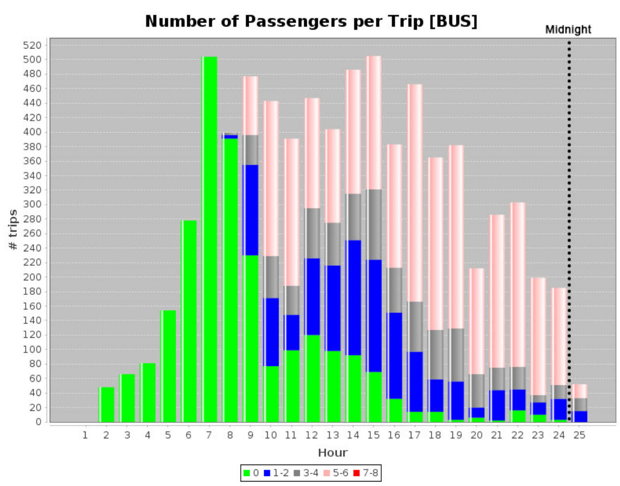
(1) Average number of passengers per bus trip for the BAU scenario.



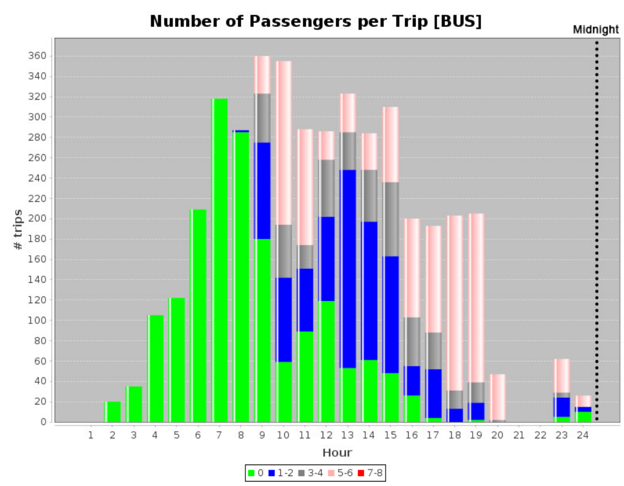
(2) Average number of passengers per bus trip for the “Transit Adverse” scenario.



(3) Average number of passengers per bus trip for the "Redistributive for Transit and On-Demand Rideshare" scenario.



(4) Average number of passengers per bus trip for the “VMT Reducing” scenario.



(5) Average number of passengers per bus trip for the “Transit Operating Cost Inefficient” scenario.

Figure 13: Histogram of the number of passengers per bus trip (aggregated by hour) (1) the BAU scenario, (2) the “Transit Adverse” scenario, (3) the “Redistributive for Transit and On-Demand Rideshare” scenario, (4) the “VMT Reducing” scenario, (5) the “Transit Operating Cost Inefficient” scenario.

3.3.2 On-demand Ride Waiting Time

Waiting times

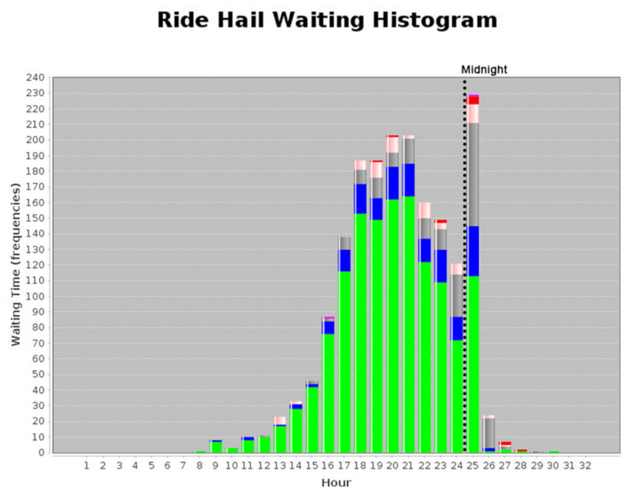
The following graphs (Figure 14) represent the distribution of the on-demand ride waiting time throughout the hours of the day.

Time history

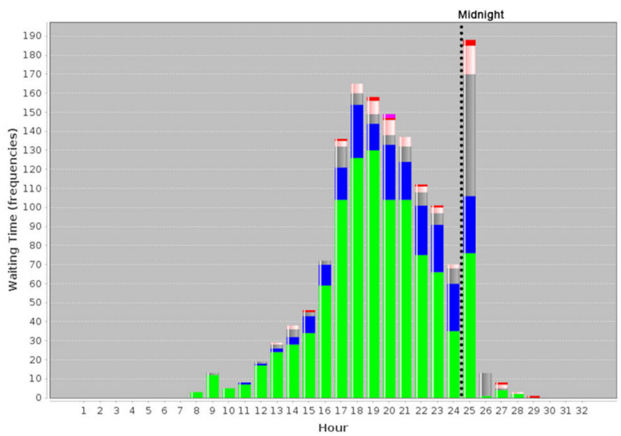
The wait times are represented with stacked bar graphs.

For each hour of the day (x-axis), the bar height (y-axis) describes the total number of on-demand car-sharing trips realized in the city. The colors on the bar define how many of those trips experienced which delay. The color patterns works as follows:

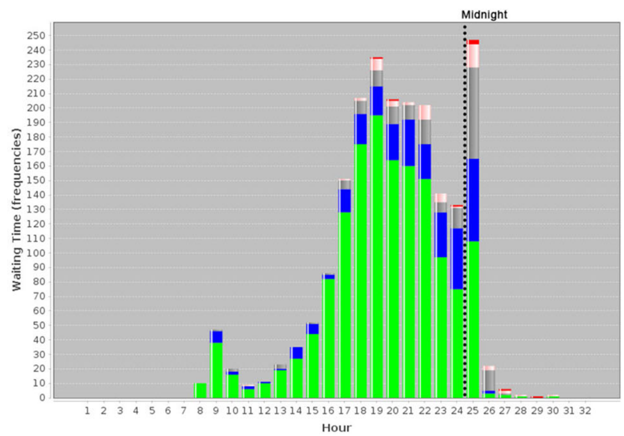
- green: 2 minute wait time
- blue: 5 minute wait time
- grey: 10 minute wait time
- pale pink: 20 minute wait time
- red: 30 minute wait time
- pink: 60 minute wait time
- black: wait time greater than 60 minutes



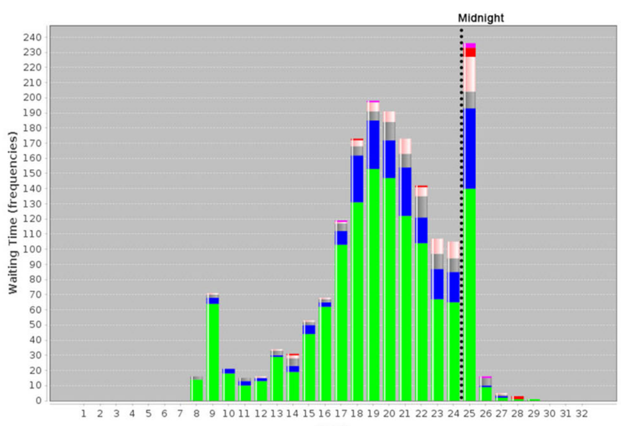
(1) On-demand Ride Waiting Time for the BAU Scenario.



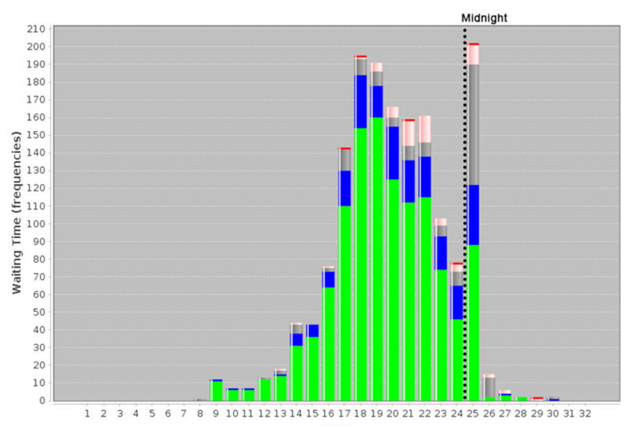
(2) On-demand Ride Waiting Time for the "Transit-Adverse" Scenario.



(3) On-demand Ride Waiting Time for the "Redistribuive for transit and on-demand rideshare" scenario.



(4) On-demand Ride Waiting Time for the "VMT reducer" scenario.



(5) On-demand Ride Waiting Time for the "Transit Operating Cost Inefficient" scenrio.

Figure 14: Histogram of the on-demand ride waiting time (aggregated by hour) (1) the BAU scenario, (2) the "Transit Adverse" scenario, (3) the "Redistributive for Transit and On-Demand Rideshare" scenario, (4) the "VMT Reducing" scenario, (5) the "Transit Operating Cost Inefficient" scenario.

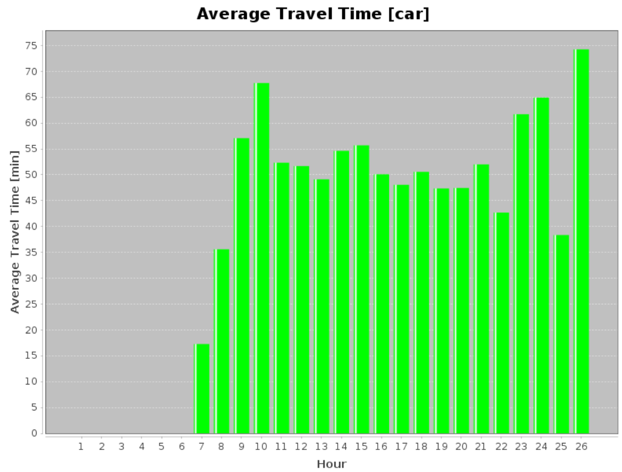
3.3.3 Average Car Travel Time By Hour

Travel time

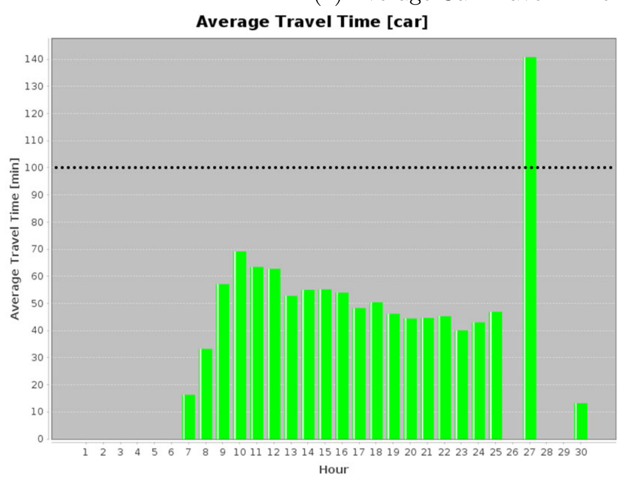
Figure 15 illustrates the average car travel time per trip by hour.

History

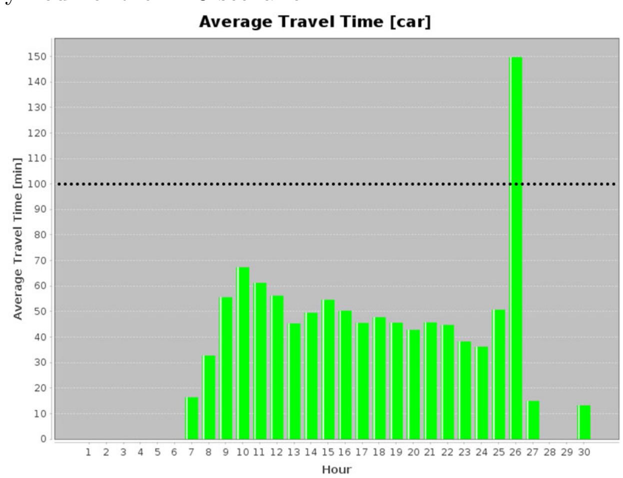
The x-axis lists the hours of the day while the y-axis defines the average travel time in minutes. Note that the y-axis goes over 24h: when people travel after midnight in the BISTRO simulation, it is considered to be part of the same "day of travel". Therefore, a day in the simulation can last more than 24 hours.



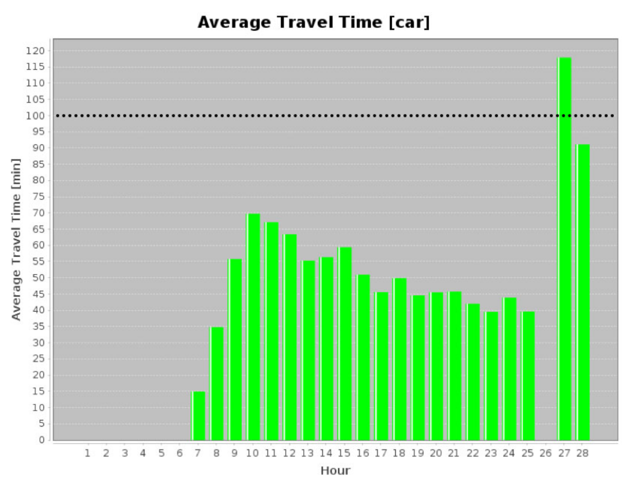
(1) Average Car Travel Time By Hour for the BAU scenario.



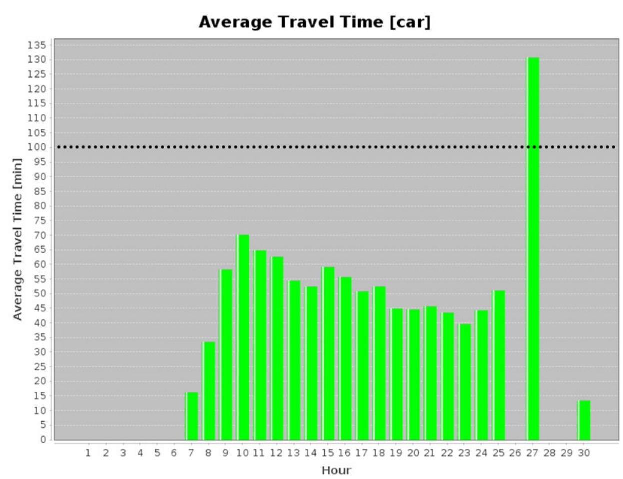
(2) Average Car Travel Time By Hour for the "Transit Adverse" scenario.



(3) Average Car Travel Time By Hour for the "Redistributive for Transit and On-Demand Rideshare" scenario.



(4) Average Car Travel Time By Hour for the "VMT Reducing" scenario.



(5) Average Car Travel Time By Hour for the "Transit Operating Cost Inefficient" scenario.

Figure 15: Histogram of the drive alone mode travel time (aggregated by hour) for (1) the BAU scenario, (2) the "Transit Adverse" scenario, (3) the "Redistributive for Transit and On-Demand Rideshare" scenario, (4) the "VMT Reducing" scenario, (5) the "Transit Operating Cost Inefficient" scenario.

3.4 Sample Scores

Overall, we find that standardizing scores enables us to more readily distinguish major improvements from minor ones. For example, looking at the unscaled scores for the two indicators of congestion, between the two samples it would be difficult to tell, at a glance, that one is significantly different from the other (Figure 16, left). The standardized scores (Figure 19, right) present a clearer picture of which sample performs better than the other while still permitting different metrics to be compared on the same scale. The "VMT Reducing" scenario demonstrates a clear reduction in VMT and delay over the "Transit Operating Cost Inefficient" scenario.

Table 5: Statistics Computed from random search (757 samples) used for standardization of submission scores

Score Component / Statistic	Accessibility: Number of secondary locations accessible within 15 minutes	Accessibility: Number of work locations accessible within 15 minutes	Congestion: average vehicle delay per passenger trip	Congestion: total vehicle miles traveled	Level of service: average bus crowding experienced	Level of service: average trip expenditure - secondary	Level of service: average trip expenditure - work	Mass transit level of service intervention: costs and benefits	Sustainability: Total PM 2.5 Emissions
Mean (μ_i)	0.988	1.004	1.105	0.979	0.942	1.077	1.064	1.047	0.976
Std. Dev. (σ_i)	0.046	0.050	0.030	0.023	0.149	0.113	0.069	0.174	0.026
Reciprocal (τ_i)	-1	-1	1	1	1	1	1	-1	1

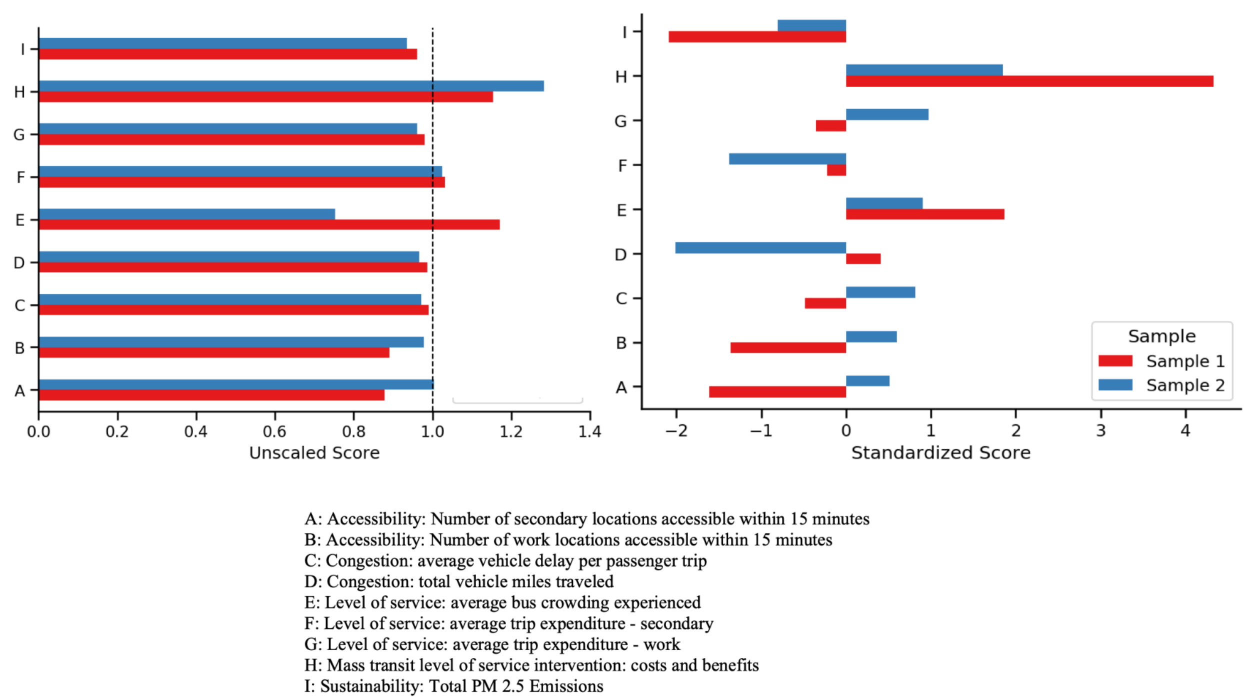


Figure 16: Unscaled (left) and scaled (right) component scores for two sample inputs.

4 Policy Weights and Scaled Scores

In order to explore the interpretation of score components in a broader decision support and enhanced planning context, the Uber Prize developed a preliminary set of generalized policy focus areas. Four distinct policy objectives were chosen to emphasize the importance of alternative simulation outcomes from the perspective of a metropolitan or regional planning organization.In this way, we can examine the performance of each of the four sets of inputs with respect to five different policy areas, as detailed below. In order to explore the interpretation of score components in a broader decision support context, we developed a preliminary set of generalized policy focus areas. Four distinct policy objectives were chosen to emphasize the importance of alternative simulation outcomes from the perspective of a metropolitan or regional planning organization. In this way, we can examine the performance of each of the four sets of inputs with respect to five different policy areas, as detailed below.

Table 6 illustrates how score components can be re-weighted in accordance with these objectives via assignment of an integer indicator of relative priority.

- The System Efficiency alternative prioritizes congestion and cost/benefit metrics primarily, with a secondary preference for accessibility improvements.
- Accessibility improvements as well as level of service indicators are emphasized under the Personal Mobility alternative.
- Transit Operations weight settings allocate greater importance to outcomes that reduce bus crowding and reduce the costs of operation compared to the benefits (with respect to BAU).
- A Policy Agnostic alternative (all weights are normalized and then set to contribute to the total score equally) is also included to facilitate comparison with the scenario wherein all components are treated equally (Figure 20, top). The sum of the indicators across each policy focus alternative is then used to normalize the respective weights for the alternative to sum to 1.0.

Figure 17 illustrates how the weighting alternatives impact the relative importance of the individual score components. In order to compute the final submission score, each of these weighted scoring components is summed for each policy alternative. Finally, Figure 19 demonstrates how the various alternatives can have quite a significant impact on the final evaluation of contestants. It is interesting to note that the percent difference in final score between the two sets of inputs can change significantly when using the different criteria. Thus, different policy objectives will influence the ranking of scores. This observation makes it clear that a principled approach to setting the weights is an important consideration prior to prize launch.

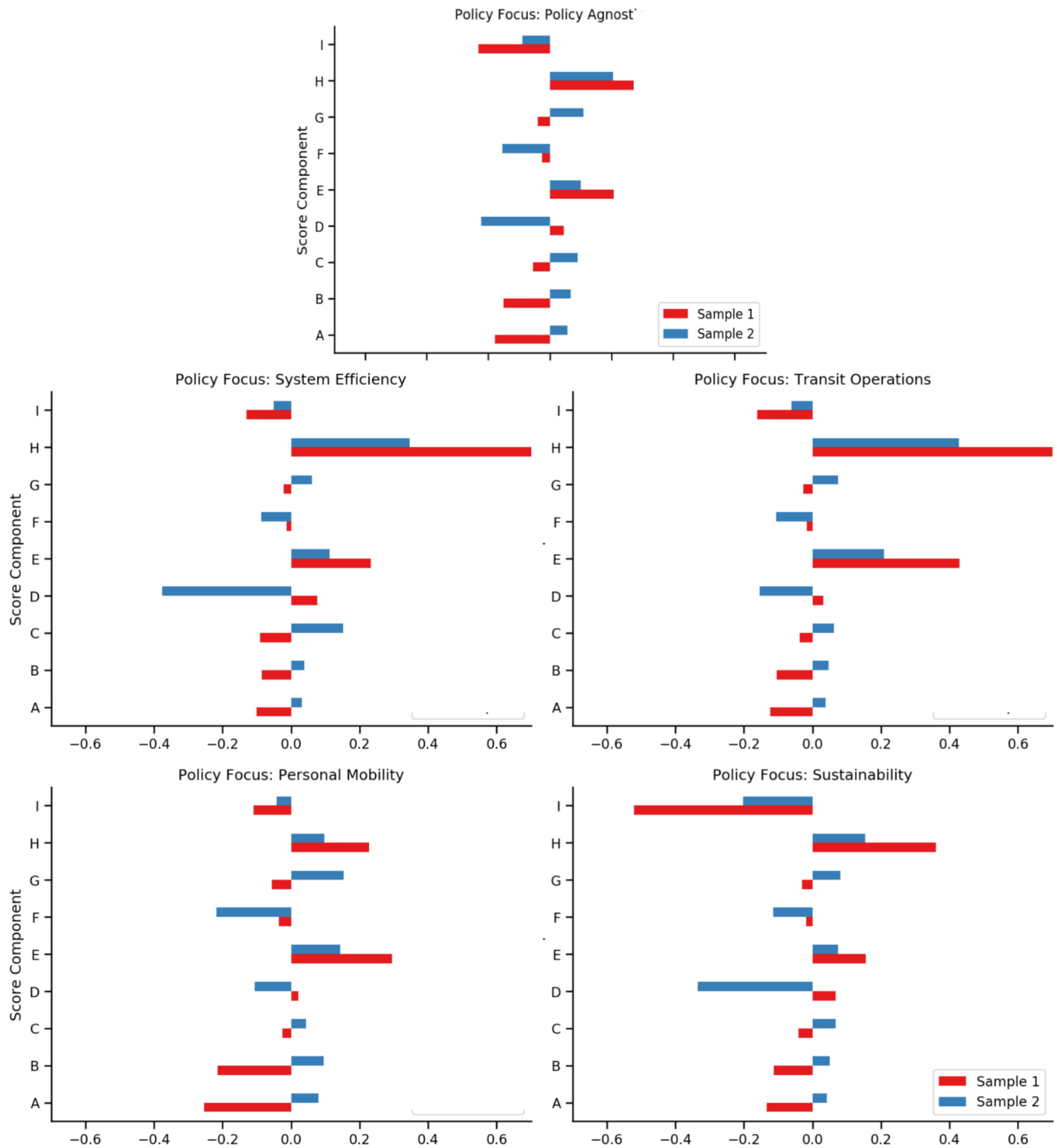
Table 6: Policy focus and associated levels of importance (weights) for scoring components.

Score Component / Policy Focus	Accessibility: Number of secondary locations accessible within 15 minutes	Accessibility: Number of work locations accessible within 15 minutes	Congestion: average vehicle delay per passenger trip	Congestion: total vehicle miles traveled	Level of service: average bus crowding experienced	Level of service: average on-demand ride wait times	Level of service: average trip expenditure - secondary	Level of service: average trip expenditure - work	Mass transit level of service intervention: costs and benefits	Sustainability: Total PM 2.5 Emissions
System efficiency	1.0	1.0	3.0	3.0	2.0	0.0	1.0	1.0	3.0	1.0
Sustainability	1.0	1.0	1.0	2.0	1.0	0.0	1.0	1.0	1.0	3.0
Personal Mobility	3.0	3.0	1.0	1.0	3.0	0.0	3.0	3.0	1.0	1.0
Transit Operations	1.0	1.0	1.0	1.0	3.0	0.0	1.0	1.0	3.0	1.0
Policy Agnostic	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

The final score to be weighted under each policy focus is the sum of the standardized scores for each metric (the ratio of the corresponding score component from the contestant submission to that of the business as usual case). The objective is to minimize the composite score function, since an increase in many of the scoring metrics actually represents a scenario that is worse than business as usual (e.g., decreasing VMT over BAU results in a lower unscaled score than increasing VMT). To maintain consistency in this regard, we take the reciprocal of several of the scoring components that are positively related to desirable outcomes (e.g., improvements in accessibility). Thus, the composite scoring function is as follows:

$$\Phi_{policy,n}\left(C_s, P_n = \left(\vec{z}, \vec{\beta_n}\right)\right) = \frac{\sum_{\{z_i, \beta_{n,i}\} \in P_n} \beta_{n,i} \cdot z_i}{\sum_{\{\beta_{n,i}\} \in P_n} \beta_{n,i}} \tag{3}$$

Where $\vec{z}$ is the vector parameter of z-scores for score components (computed as defined in Equation 2), while $\vec{\beta_n}$ is the vector of weight values under policy focus n for each of the score components i.



A: Accessibility: Number of secondary locations accessible within 15 minutes
B: Accessibility: Number of work locations accessible within 15 minutes
C: Congestion: average vehicle delay per passenger trip
D: Congestion: total vehicle miles traveled
E: Level of service: average bus crowding experienced
F: Level of service: average trip expenditure - secondary
G: Level of service: average trip expenditure - work
H: Mass transit level of service intervention: costs and benefits
I: Sustainability: Total PM 2.5 Emissions

Figure 17: Score components for "VMT Reducing" scenario (4) and "Transit Operating Cost Inefficient" scenario (5) reweighted per policy alternatives in Table 6

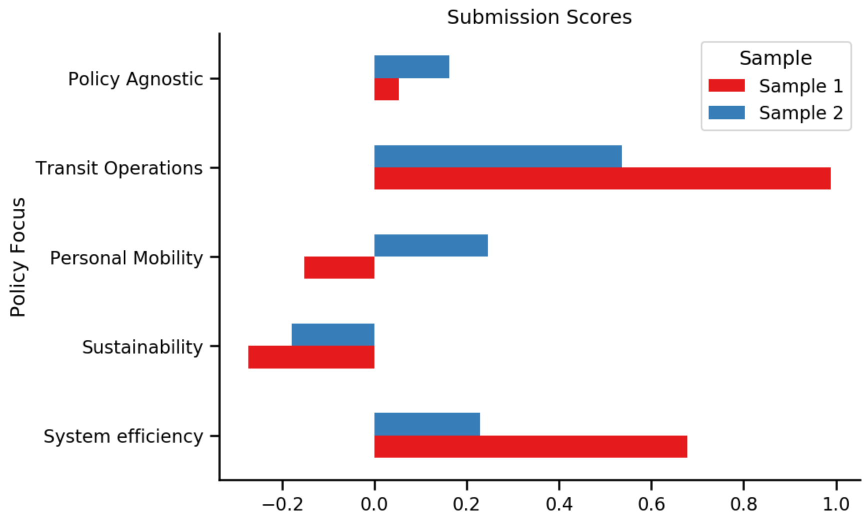
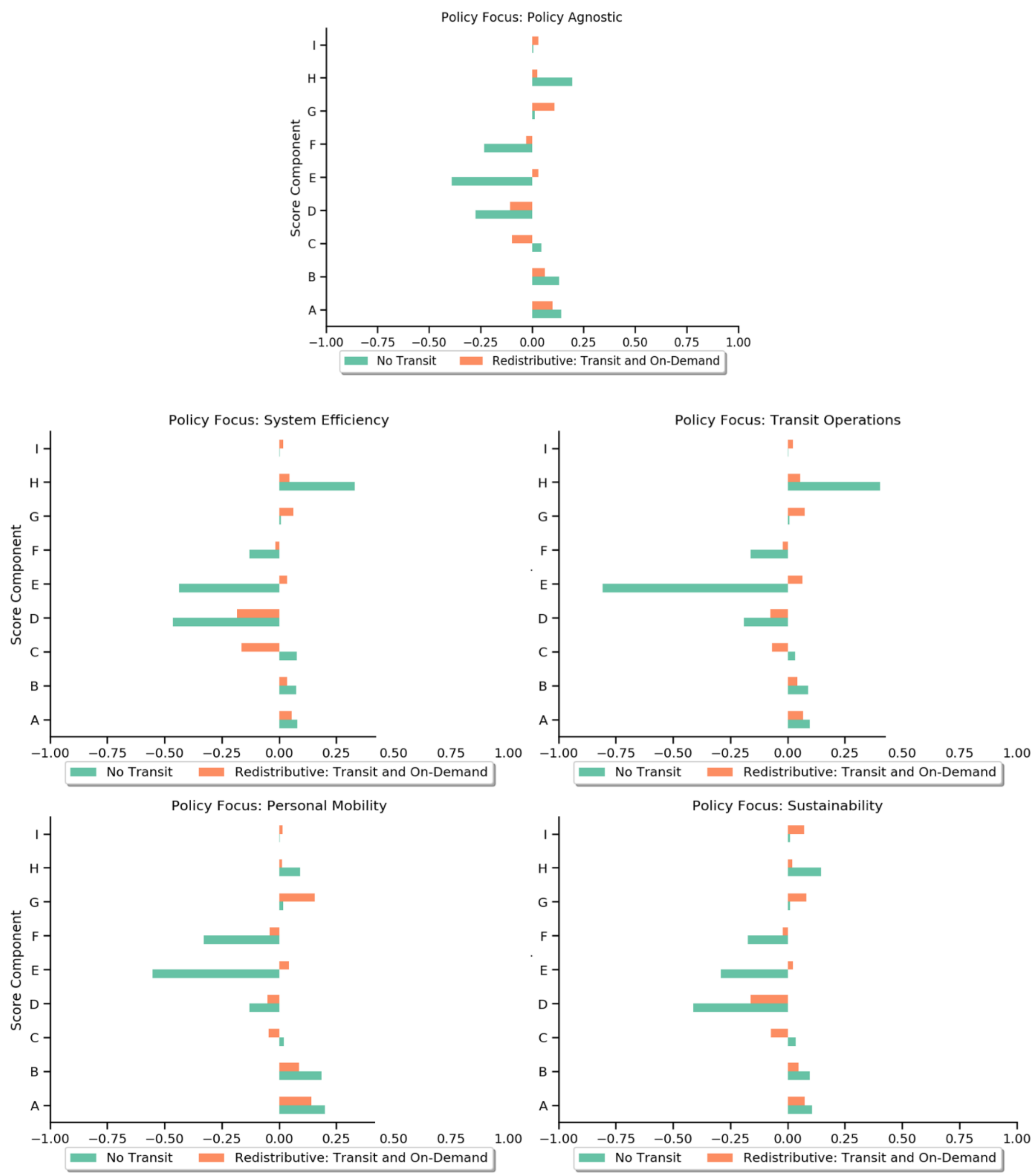


Figure 18: Total submission scores for all policies for "VMT Reducing" scenario (4) and "Transit Operating Cost Inefficient" scenario (5)



A: Accessibility: Number of secondary locations accessible within 15 minutes
B: Accessibility: Number of work locations accessible within 15 minutes
C: Congestion: average vehicle delay per passenger trip
D: Congestion: total vehicle miles traveled
E: Level of service: average bus crowding experienced
F: Level of service: average trip expenditure - secondary
G: Level of service: average trip expenditure - work
H: Mass transit level of service intervention: costs and benefits
I: Sustainability: Total PM 2.5 Emissions

Figure 19: Score components for (2) and (3) re-weighted per policy alternatives in Table 6

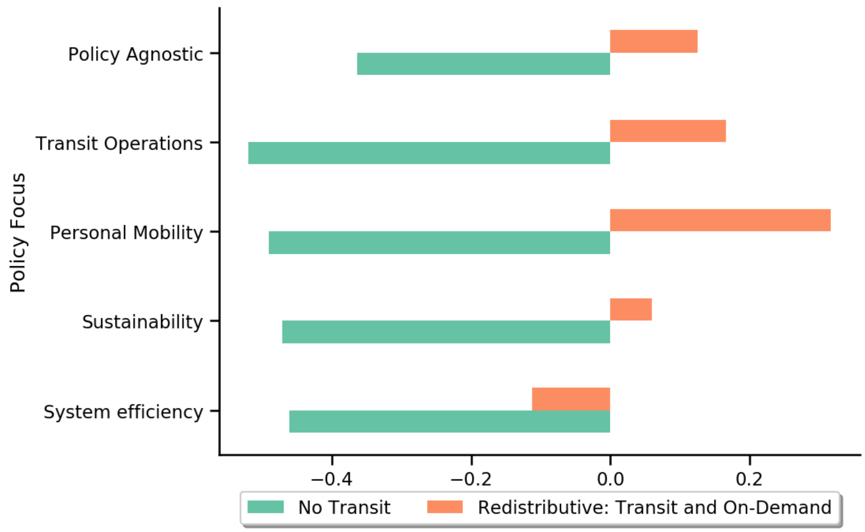


Figure 20: Total submission scores for all policies for (2) and (3).

5 The Role of and Rationale for Subsampled Scenarios in BISTRO

5.1 Introduction

When modelling large agent-based transportation systems, population subsamples are often used to approximate full population behavior due to limitations in computational resources[4] [5] [6] [7]. The idea is that instead of simulating all agents, only a representative subset is simulated. Building the input data for this scenario requires a commensurate downscaling in transportation network infrastructure (e.g. lower road/transit capacity and less rideshare drivers). Despite attempts to spatially constrain agent attributes and activity chain distributions to match between samples, the subsampled model may not behave identically to the original model, e.g. due to nonlinear effects of changes in infrastructure. Thus far, the literature is unclear on how similarly one may expect subsampled models to behave with respect to KPIs. In the following we provide an overview of the 1,000 agent sample together with a limited sensitivity analysis to demonstrate that the model behaves similarly to the original population. I.e., policy changes move mode choice in a similar direction for subsample and full population, while the absolute aggregate results may differ. Given the lower computational resources required to run the smaller sample, we also explain how providing the subsample to contestants can accelerate search and potentially spur development of more innovative solution methods.

5.2 The “1k” Scenario

To evaluate the mechanics of downsampling, we prepared an approximately 1,000 agent subsample, which we call the “1k” scenario (996 agents to be precise). Rather than generate a new sample from the original 157,000 agent population, we subsampled the existing 15,000 agent “15k” Sioux Faux scenario. The subsampling algorithm operates at the household level: making sure that various key characteristics of the original population are kept as close as possible (e.g. spatial distribution, household size, number of vehicles, travel distance). The relative error between key characteristics of the 15k and 1k scenario differed by less than 1% (based on key characteristics mentioned before).

5.3 Calibration Objective: Using Scaling Coefficients to Achieve Similar Mode Shares in BAU (1k vs. 15k)

While relative population attributes were quite close between the two samples, calibration of the subsample was necessary to ensure that agent decision-making behavior and outcomes (as demonstrated by mode choice) were likewise similar. It was decided to focus primarily on the scaling parameters in the config file in order to maximally conserve the same alternative-specific constants between the samples. Doing so would imply that agents would have matching responses to costs and incentives–an important desideratum given that our inputs control precisely these quantities. We achieved this goal with very low error (see Table 2).

[4] Axhausen, Kay W., Konrad Meister, Michael Balmer, Francesco Ciari, Andreas Horni, Marcel Rieser, and Rashid A. Waraich. "Large-scale agent-based travel demand optimization applied to Switzerland, including mode choice." Working paper/Institute for Transport Planning and Systems 625 (2010).

[5] Balmer, Michael, Kay W. Axhausen, and Kai Nagel. "Agent-based demand-modeling framework for large-scale microsimulations." Transportation Research Record 1985.1 (2006): 125-134.

[6] Waraich, Rashid, and Kay Axhausen. "Agent-based parking choice model." Transportation Research Record: Journal of the Transportation Research Board 2319 (2012): 39-46.

[7] Waraich, Rashid A., et al. "Plug-in hybrid electric vehicles and smart grids: Investigations based on a microsimulation." Transportation Research Part C: Emerging Technologies 28 (2013): 74-86.

Table 7: Calibration factors

	BAU 1k	BAU 15k
Alternative specific constants		
car_intercept	0	0
walk_transit_intercept	0	0
drive_transit_intercept	0	0
ride_hail_transit_intercept	-0.124	-0.124
On-demand Ride Coefficient	$\beta_{on-demand-ride}$	-0.124
ride_hail_intercept	0	0
walk_intercept	-0.8	-5
Infrastructure		
transitCapacityFactor	0.01	0.1
flowCapacityFactor	0.1	0.01
numDriversAsFractionOfPopulation	0.1	0.1

The mode shares in the BAU for the 1k agent vs. 15k agents are similar, but not exactly the same, see Figures 21 and 22. Mixed modes means, that the the agent started out using a certain mode, but ended up using a different mode due to lack of availability, e.g. full buses):

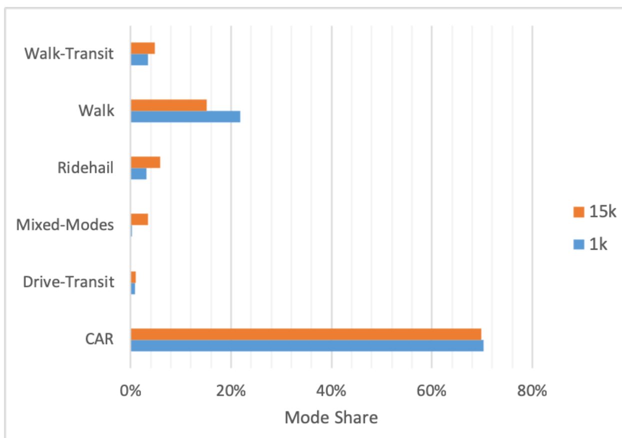


Figure 21: Mode share for the BAU scenario: 1k vs. 15k sample sizes.

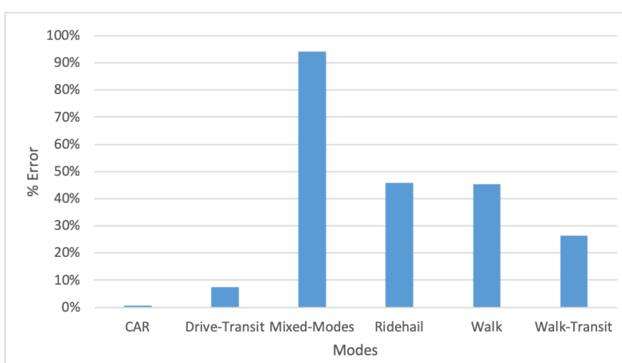


Figure 22: Percentage of error for different modes: 1k vs. 15k sample sizes.

5.4 Policy intervention 1k vs. 15k

In order to demonstrate that 1k and 15k population scenarios behave similarly to policy change, we prepared 9 sets of inputs, where we varied the headway of the bus services (15 minutes, 30 minutes and 60 minutes) and also provided 3 levels of ModeIncentiveInputs per bus ride ($1.00, $3.00 and $5.00). In Figure 23 one can see the percentage of mass transit changing for these policies, for the 1k and 15k scenario. While the 1k and 15k mode shares outputs are visibly correlated, the absolute change in the scenarios is somewhat variable. A major contributor to the discrepancy between the sampled scenario and the full scenario (besides the relative sample size) are also rounding errors around infrastructure scaling[8] [9].

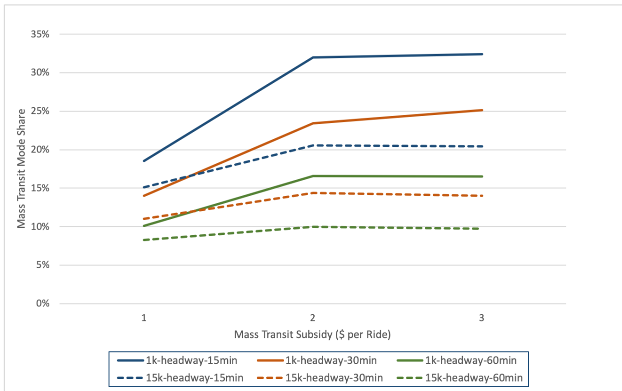


Figure 23: Policy impact: 1k vs. 15k sample sizes.

5.5 Conclusions and Recommendations

Based on these results, our assessment is that the 1k subsample responds to inputs in a similar enough way to the 15k sample that it can serve as a useful tool in the competition to facilitate innovation as well as permit debugging using limited resources (e.g., a development machine). For example, the 1k sample could be used is to build a metamodel of the 15,000 agent scenario. In principle, contestants could infer the inputs required to achieve a good outcome in the 15k scenario by regressing inputs and scores of the 1k scenario as covariates against the scores achieved by the 15k scenario using the same inputs. We anticipate that employing such an approach in Round II of the competition will be of even greater importance due to the significantly reduced number of times that a simulation of a 10x larger city and population size will be evaluated per contestant. Consequently, we intend to make a subpopulation also available for the next round of the competition.

[8] Waraich, Rashid, and Kay Axhausen. "Agent-based parking choice model." Transportation Research Record: Journal of the Transportation Research Board 2319 (2012): 39-46.

[9] Waraich, Rashid A., et al. "Plug-in hybrid electric vehicles and smart grids: Investigations based on a microsimulation." Transportation Research Part C: Emerging Technologies 28 (2013): 74-86.

6 Hackathon judging criteria

6.1 Business rules

The hackathon will be judged in the following way. In order to win, the BISTRO solution provided by the competitors must adhere to the *business rules* outlined below. These are in place to make sure that the solutions provided by the competitors are compliant with common practices in policy and planning, as follows:

- The bus frequency cannot change more than four times across the day. More precisely, there can be no more than five distinct bus service periods (this mimics the typical delineation: am peak, midday, pm peak, evening, late night/early morning).
- Bus fares and mode incentives may not isolate a single age; fares and incentives must be defined in bins no smaller than five years (or $5,000, or other) in range.
- Bus route headways may be no more than 120 minutes and no less than 3 minutes.

6.2 Submissions and Judging Criteria

All code used to generate the solution, graphics, visualization and outputs must be submitted along with the solution (inputs), with proper comments by March 11, 2019, 11:59pm. Along with their submission, the teams will submit a report documenting their findings, as well as their approaches, designs and conclusions.

The competitors with the top 10 scores on the AICrowd leaderboard as defined in Section 4 by equation (3) will become finalists and will prepare presentations to be given during the final round of judging on March 13, 2019. After the presentations, the judges will confer and select three winners for the hackathon from the 10 finalists. The three prizes are explained in Section 6.2.1 below and will be determined using a final score comprised of a weighted sum along three axes of delineation: (1) Quantitative scores (x%), (2) The quality of their supporting documents and presentation (y%), and (3) citizenship scores (z%). Each prize will be determined using the same three criteria, however the weight of each element's contribution to the final score will change, according to the table in Section 6.2.1.

6.2.1 Winners and Prizes

Using the rubric below, the three winners will be determined sequentially from amongst the finalists (i.e., no one team can win more than one prize). Quantitative score: x%, quality of the supporting documents and presentation: y%, and citizenship score: z%.

PRIZE	x	y	z
First Prize: Best in show	60%	20%	20%
Second Prize: Best presentation and report	20%	60%	20%
Third Prize: Best Citizen	20%	20%	60%

The prizes are as follows:

- **First Prize: Best in show ($10,000 and dinner with Dara)** The team scoring the highest with 60% quantitative, 20% presentation/documentation, 20% citizenship as outlined in row 1 of the table above wins Best in show. This rewards quantitative performance when optimizing the Sioux Faux system based upon the given priorities, also taking into account presentation quality and citizenship.
- **Second Prize: Best presentation and report (dinner with Dara)** Among the 9 remaining teams, the one scoring highest with 60% presentation/documentation, 20%

quantitative and 20% citizenship as outlined in row 2 of the table above wins Second Prize: Best presentation and report. This rewards the ability to showcase results and communicate them excellently.

- **Third Prize: Best citizen, dinner with Dara** Among the 8 remaining teams, the one scoring highest with 60% citizenship, 20% presentation/documentation and 20% quantitative as outlined in row 3 of the table above wins Third Prize: Best citizen. This rewards the most valuable feedback and recommendations for launching the Phase I prize, thus enhancing the competition experience for all competitors.

6.2.2 Computation of the scores

Each contributing piece to the overall score will be determined in the following ways:

1. Quantitative score (score from 0 to 10). To reward the quantitative performance, finalists will be judged based upon their performance relative to the top scoring team (the top scoring team will receive a maximal amount of the quantitative score). The numerical score S_t will be attributed to team $t \in \{1, \cdots, 10\}$ is computed using the scoring function described in Section 4 by equation (3). In this equation, the $\beta_{agnostic,i}$ will be all taken equal to 1. Scores are normalized by $\Phi^{\min}_{policy,agnostic} = \min_{t=1\cdots 10} \Phi^{t}_{policy,agnostic} < 0$ (remember that the goal to achieve is to minimize the total cost, and $\Phi_{policy,agnostic}(BAU) = 0$). Team t gets the following final quantitative score:
$$Score(t) = 10 \cdot \frac{\Phi^{t}_{policy,agnostic}}{\Phi^{\min}_{policy,agnostic}} \in [0, 10]$$
assuming the team did better than BAU (which gets 0 score).[10]

2. Solution and presentation score (score from 0 to 10) Each team will provide the judges a report which should include: (1) inputs chosen and rationale, (2) outputs achieved and comments on the results, (3) narration of the solution, what it achieves and why it will benefit the city of Sioux Faux, (4) graphics which might enrich the presentations (if of value), and (5) anything else the team believes will provide value to the presentation of their solution or to the Hackathon and future rounds of the Uber Prize. Finalists will present their solutions for the judging panel and are free to use slides or any other medium. The solution and presentation score will be computed as follows by the judges:

Quality of the report	2 points
Quality of the presentation	2 points
Quality of visuals, videos, graphs	2 points
Innovation (algorithm, AI, optimization, etc.)	2 points
Quality of design (planning and transportation considerations)	2 points

all adding to y%. The goal of the presentation is for the judging panel to gain understanding of the experience of the competitors, create community, and help guide the Uber Prize towards the organization of the launch of Phase I. ‘

3. Citizenship score (score from 0 to 10) Good citizenship helps the Uber Prize towards its release to the public. More concretely, since the Starter Kit[11] documentation and any associated code will be open sourced at the conclusion of the Hackathon to support the global release of the Uber Prize, the *citizenship score* asks that competitors be proactive in finding bugs, improving documentation, and, in general taking every opportunity to contribute helpful feedback to the Prize organizers.

[10] Note that in this formula, $\Phi^{t}_{policy,agnostic}$ is technically taken to be $\min\{\Phi^{t}_{policy,agnostic}, 0\}$ in case a competitor submits a score worse than BAU.

[11] https://github.com/vgolfier/Uber-Prize-Starter-Kit

A competitor's citizenship score will be awarded on the basis of a 1-2 page report documenting contributions made across five (5) categories encompassing the Uber Prize materials. These include:

(a) Explanatory documentation. Examples include: the problem statement, starter kit GitHub/GitLab pages, AICrowd "Overview" site, and the competition rules. This item does not include the FAQ, which is a separate item (see next).

(b) FAQ. If a competitor regularly contributes to StackOverflow, they will really shine in this category! This page will be the first place contestants turn to for troubleshooting and advice (before they hit the discussion forum).

(c) Example code, notebooks, and tutorials. Competitors can contribute towards a more innovative and stress-free user experience by submitting issues and pull requests whenever they find bugs in any open source code libraries provided as part of the Hackathon as well as several key dependencies. These are limited to the Starter Kit, the BEAMCompetitions wrapper code/sioux falls sample data, the BEAM core library, R5, and MATSim.

(d) Competitor-generated tutorials. This item encourages competitors to help others get past common stumbling points or visualize the problem in creative ways. Much like Kaggle's "Kernel's" feature, any useful user-generated tutorials can be submitted as pull requests.

(e) Ideas for innovation. As Uber gets ready to launch the Prize, we are continually assessing how it can best convey the truly exciting and innovative initiative of "Uber as a Platform". Contestants should submit issues to the Starter Kit with ideas as to how the Uber Prize can redefine the state of the art in global machine learning and data mining competitions.

A template report will be provided in `markdown`, which will have space for competitor to reference all issues and pull requests (PRs) submitted during the Hackathon. To be counted, an issue or PR must adhere to the guidelines provided in .github/CONTRIBUTING.md in the Starter Kit. Each category should have one issue or PR highlighted in **bold** to be judged for content below. In addition to these, at the end of the document, each competitor should describe the one issue or PR that they feel best represents their ability and interest in supporting the open source and collaborative ethos of the Uber Prize.

Half (5) of the ten (10) points for this score will be awarded equally for each category based on the volume, quality, and criticality of contributions overall; however, judges are to particularly focus on the highlighted issue for that category. The other half (5) will be awarded for the issue/PR and description provided at the end of the Citizenship Score document. At least two judges will review the document submitted for the citizen score and provide a subscore for the two parts of the Citizenship Score based on their impression of its contents. The average of the two subscores given by each judges on the two parts of this component will comprise the final value for the Citizenship Score. If judges differ by more than a point, then a third judge will review the submitted document, and the greater two of the three scores will be averaged.

Note that category 2 and 3 scores might not be used in the actual launch of Phase I of the Uber Prize, and are included only for the internal Hackathon, in order to improve the final launch.

6.2.3 Honorable mentions

All competitors (not only finalists) are eligible for honorable mentions (it is not required to be in the top 10 pack to qualify). In order to qualify for honorable mention, a team needs to

(1) score highest on one of the corresponding to the 4 policy rows of Table 6 (excluding the one used for First Prize for which the $\beta_{n,i}$ coefficients where taken equal to 1, corresponding to an *agnostic* scoring), and (2) submit a report on the team's approach following the report structure outline in Section 6.2.2.

- **Honorable mention for System efficiency** To receive this award, contestants will have to prioritize their optimization for the minimization of system-wide congestion as well as the mass transit LoS intervention costs and benefits as measured by the total vehicle miles traveled, the average vehicle delay per passenger trip, and the net revenue from bus operations after subtracting the costs of bus operations and total incentives distributed. To a lesser extent, contestants will need to consider the impact of their solution on the average bus crowding experienced. This will be awarded to the top performer for equation (3) using the $\beta_{n,i}$ coefficients of Table 6 row 1.
- **Honorable mention for Sustainability** To receive this award, contestants will have to prioritize the minimization of PM 2.5 emissions, followed by the minimization of total VMT. This will be awarded to the top performer for equation (3) using the $\beta_{n,i}$ coefficients of Table 6 row 2.
- **Honorable mention for Personal mobility** To receive this award, contestants will have to prioritize the improvement of score components that measure the accessibility and level of service provided by the transportation network, including accessibility to work and secondary locations, average bus crowding experienced, and the average trip expenditure for both work and secondary trips. This will be awarded to the top performer for equation (3) using the $\beta_{n,i}$ coefficients of Table 6 row 3.
- **Honorable mention for Transit operations** To receive this award, contestants will have to focus their optimization on providing the best possible level of service on transit, as measured by average bus crowding experienced, while prioritizing score performance in the mass transit level of service intervention costs and benefits component. This will be awarded to the top performer for equation (3) using the $\beta_{n,i}$ coefficients of Table 6 row 4.

Teams receiving honorable mentions for each of these categories will be announced along with the Hackathon winners.

7 Acknowledgments

The following collaborators are gratefully acknowledged:

Juan Agote, Kay Axhausen, Fran Bell, Douglas Bemis, Zeeshan Billal, Theophile Cabannes, Sashikanth (Sashi) Chandrasekaran, Alex Chao, Sihao Chen, Chris Chua, Peter Dayan, Natalie Diao, Laurent El Ghaoui, Zoubin Ghahramani, Manik Gupta, Shan He, Minjune Hwang, Cory Kendrick, Thomas Kirchstetter, Dmitry Kuzhanov, Alex Lai, Jeong-Yoon Lee, Kirill Mitin, Sharada Mohanty, Chris Pangilinan, Jan Pedersen, Justin Pihony, Matthias Poloczek, Jazz Pouls, Yunus Saatci, Marcel Salathé, Andrew Salzberg, Colin Sheppard, Anhad Jai Singh, Emily Strand, Brian Tsang, Ryan Turner, Eckart Walther, Rashid Waraich, Michael Wehrmeyer, Trevor Wu, Bofan Wue, Michael Zilske.