1.0 Introduction

1.1 Introduction and Rationale for Project

The Seattle Mobility Index is a project that aims to gain nuanced insight on the current state of Mobility in Seattle and provide a baseline for measuring changes to people's ability to get around in this fast changing city. Mobility includes the generation of trips by the inhabitants of a city and the conditions associated with these trips. Measuring mobility requires information on specific trips (modes of transport taken, costs incurred, duration) and system-level characteristics such as mode reliability, environmental impact, accessibility and equity. For the purpose of this project we define mobility as *the ability to reach everyday destinations with your choice of mode, affordably, and reliably.*

The project presents an innovative approach to mobility measurement by moving beyond traditional governmental data sets and isolated traffic sensors to taking advantage of crowd-sourced digital data using the Google Distance Matrix. This approach allows for granular, regular, and comprehensive mobility measurement. In this project we aim to evaluate Seattle's Urban Mobility on three core dimensions: mode choice, affordability and, reliability.

The importance of this project stems from the understanding that a well-functioning transportation system is fundamental for economic and social development in any urban setting. In particular, passenger transport allows people to participate in a range of activities and services making urban mobility a necessity for the competitiveness of cities and well-being of citizens. Mobility is rarely an end in itself; the ultimate goal of most transportation activity is people's overall ability to reach desired destinations. Given that unequal access to mobility can often signify unequal access to these same activities and services, we need tools to better understand where mobility is lacking and who is underserved. We aim to help address this gap by developing measures quantify mobility to help prioritize policy interventions.

In this project we set out to create comprehensive measures for Seattle City officials to map the landscape urban mobility. The City of Seattle is currently experiencing a time of opportunity and challenge regarding urban mobility, due to a series of demographic changes and urban landscape. Seattle's population has experienced a significant period of growth. Since 2010, the population has increased by 18.7% (or 114,000 people), making it the fastest growing City in the United States during this period. To put these numbers into perspective it took 30 years, from 1980 to 2010, to attain approximately the same amount of growth achieved in this decade so far[1].

The growth in population in Seattle has come with increased demand for mobility services across many different transportation modes. Seattle has experienced substantial increase in congestion

---

[1] **114,000 more people: Seattle now decade's fastest-growing big city in all of U.S.** Gene Balk, Seattle Times, May 24, 2018
https://www.seattletimes.com/seattle-news/data/114000-more-people-seattle-now-this-decades-fastest-growing-big-city-in-all-of-united-states/

delays, where the average delay increased by 19 percent between 2007 and 2015. Seattle has also seen remarkable growth in transit ridership transit ridership adding 4.7 million trips in 2017 for a total ridership of 191.7 million, the highest growth in transit ridership in the country[2]. While Seattle's use of public transit is impressive, there are still numerous neighborhoods in which Seattle residents who are dependent on transit but lack adequate access[3].

Adding to the challenge of increased demand for mobility services, the city of Seattle has several large scale transportation and construction projects underway that are expected to make getting around the city more difficult until the projects are finished. For this reason, Seattle transportation planners have dubbed the next three years as the period of maximum constraint[4].

As the urban landscape and population of Seattle grows, there is a need to meet new mobility and transportation demands, while protecting the environment and ensuring social inclusion. In this section we present the core concepts and terms used throughout the project, as well as the technical specifications and tools that were essential to design the processes.

---

[2] Greater Seattle area leads the nation in transit ridership growth https://kingcounty.gov/elected/executive/constantine/news/release/2018/February/21-metro-ridership.aspx

[3] Seattle Has 'Transit Deserts' and—Surprise—It's No Fun to Ride the Bus There https://www.seattlemet.com/articles/2017/11/20/what-s-it-like-taking-the-bus-in-seattle-s-transit-deserts

[4] If you think Seattle traffic is bad now, just wait until these projects start. Mike Lindblom. Seattle Times January 7 2018. https://projects.seattletimes.com/2018/one-center-city/

1.2 Core Concepts

Measuring urban mobility requires the operationalization of core concepts related to how a city is structured, where and how people travel, and the conditions of different trips that lead to different abilities to get around a city. In this section, we break down the core concepts that are required to construct the Seattle Mobility Index.

Geography

Cities are divided into geographical units, usually delineated for administrative and governance needs. *We define a geography as a distinct way of dividing a city into smaller areas or units.* Geographies for Seattle include census block groups, neighborhoods, districts, zip codes, and urban villages.

Geography	Definition
Census Block Group	Geographical unit used by the United States Census Bureau. It is the smallest geographical unit for which the bureau publishes sample data.
Neighborhood	Geographic unit forming a community within a town or city.
District	A geographic subdivision, usually defined for administrative or electoral purposes.
Zip code	A geographic subdivision defining a system of postal codes mainly used for organizing postal deliveries.
Urban village	Urban villages are areas characterized by mixed use zoning, a concentration of places of interest for people (restaurants, cafes, services) and public space.

While we relied on relationships between geographies to connect different types of data, the core geographical unit of analysis throughout this project is the census block group (or block group).

Location, origin and destination

A location is a point in space defined by a unique latitude and longitude. A given latitude and longitude in Seattle can be identified as a corresponding specific block group or zip code to connect different units of geography.

An **origin** is a starting location and a **destination** is a location where the trip ended. For this project, *origins are mapped into the latitude and longitude of the centroid of a block group where the trip starts*. In this way, we can group different trips started from a single block group.

Destinations, on the other hand, belong to ten different place *categories*. These include local amenities, such as grocery stores and libraries, and citywide places of interest, such as major employment or cultural centers.

The key difference between local and citywide destinations is that local destinations refer to generic types of places that people visit, but the specific location depends on your place of origin. Citywide destinations are specific locations in Seattle that are the same for everyone.

Local destinations (basic services and amenities, specific to a starting point**)**	**Place categories:** Urban villages, destination parks, supermarkets, libraries, hospitals, pharmacies, post offices, schools, and cafes
Citywide destinations (areas with high travel demand employment density, universities, transit hubs and large tourist attractions**)**	**Place category:** citywide destination **Citywide destinations for Seattle:** Seattle Central College, North Seattle College, South Seattle College, King Street Station, Westlake Center, Northgate Transit Center, Bellevue Square, Tukwila, SODO, Overlake-Redmond, First Hill, Downtown West End, Downtown Com. Core, Univ. of Washington

Basket of destinations

A basket of destinations is a sample of representative places where people might need to go, analogous to the Consumer Price Index market basket. In this project, the basket for any origin (defined as the center of the block group where the trip starts) consists of 25 destinations, of which 13 are citywide locations and 12 are local destinations.

Mode

A transportation mode refers to a distinct form of travel. In this project we consider four distinct modes: cars, public transit, walking and cycling.

Trip

A trip is the act of going from an origin to a destination. A single trip can have many different characteristics, such duration, distance, mode of travel, and time of the day when the trip is carried out. Our characterization of a trip is ‘blind’ to the individual who can potentially take this trip.

Travel Persona

A persona is a caricaturization of a type of person who interacts with the transportation system. A travel persona has particular socioeconomic characteristics (for example age, income, education level, number of children, car ownership), travel needs, and proportion of travel modes in total trips. The concept of personas is used to help us understand the transportation user’s needs, experiences, behaviors and goals, and how urban mobility might change for different groups of people. Persona types were created based on the *PSRC Household Travel Survey Data*.

Cost

Cost is the dollar amount required to make a trip. This is both direct expenses, such as bus fare and parking fee, as well as indirect expenses, such as cost of time.

Index, calculator and score

An Index is a way to evaluate aggregate characteristics of urban travel that influence a person's mobility experience. A score is the specific value of an index in a particular place. A calculator is the script that estimates the scores of a given index. This project considers the characteristics of viability, affordability, and reliability as core components of mobility and develops an index to quantify each one. Each of the scores are scaled from 0 to 100 and details of each score are provided below:

- **Mode choice Index:** The Mode Choice Index measures the extent to which destinations within a market basket of destinations are reachable in a 'reasonable' amount of time from each block group. Trips are defined as viable if the duration required to complete that trip is below a threshold, specific to each particular mode. Baseline thresholds are informed by the average duration of trips per mode in Seattle.
- **Mode Affordability Index:** The affordability Index measures the cost to reach the destinations in the market basket including direct costs (such as fares, parking, gas) and indirect costs (value of time). For each block group, the average cost of the viable trips are calculated using mode-specific costs.
- **Mode Reliability Index:** The reliability calculator estimates an index that provides a measure of the consistency in day-to-day travel time for trips to the basket of destinations from each origin (block group).

1.3 Preliminary Analysis, Insights (Case Study)

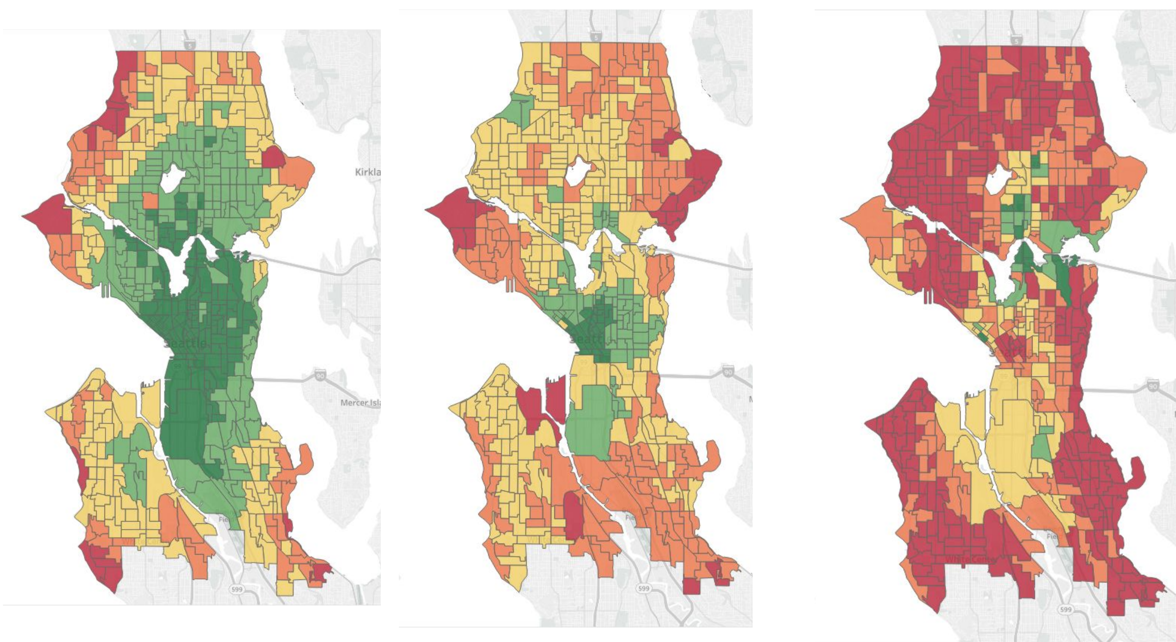
In this project, we developed three separate measures of mobility. An overview of the methodology for each of these indices can be found in section 2.

1.3.1 Results for the Seattle Mobility Indices

The Mode Choice, Affordability, and Reliability Indices were estimated and visualized in each Seattle block group. The city averages of Mode Choice, Affordability, and Reliability Indices were 72, 48, 24,respectively. Figure 1 summarizes the scores obtained for each of these indices.

Figure 2. Seattle Mobility Indices results

a) Mode Choice Scores **b) Affordability Scores** **c) Reliability Scores**



The Mode Choice scores were highest in downtown Seattle, SODO, and Capitol Hill areas, followed by surrounding neighborhoods such as the University District, Fremont, and Beacon Hill. The Affordability scores were generally low, where just the downtown had scores over 70 and the rest of Seattle were lower. The Discovery Park area showed some of the lowest Mode Choice and Affordability scores. The reliability scores were overall the lowest among all three Indices, while Fremont and the University District areas having higher scores than the other areas.

1.3.2 Travel Personas

The five Travel Personas that resulted from our analysis showed unique demographic characteristics and travel patterns. For example, Persona A type of people traveled by car most of the time – 70% of their trips were done by driving. On the other hand, 46% of trips by people from the Persona B category were completed by walking. Income levels were also different, where 67% of Persona A households showed annual income of $75,000 or higher, while only 28% of people from Persona B made such above-median income. Thirty-eight percent of the trips by Persona B was to go to work, which was higher than that of Persona A – 24%. The
Persona-adjusted Mode Choice Indices ranged from 56 to 76 from the city average, which was 72, and Affordability Indices ranged from 44 to 49 while its city average was 48.

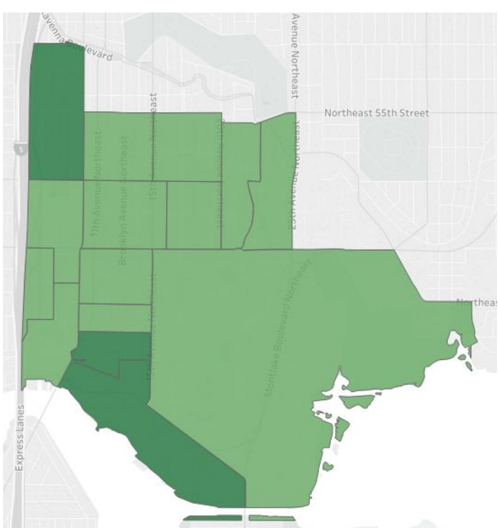
1.3.3 Case study: adjusted mobility scores for Persona B in the University District

When we limited our focus on the University District, the average value of Mode Choice, Affordability, and Reliability Indices without considering Personas showed 78, 57, and 47, respectively. We only compared the baseline with Persona B, whose Mode Choice, Affordability scores were 64 and 48. Because the Reliability Scores were generated mostly from the simulated data, we did not calculate the Persona-adjusted reliability scores.

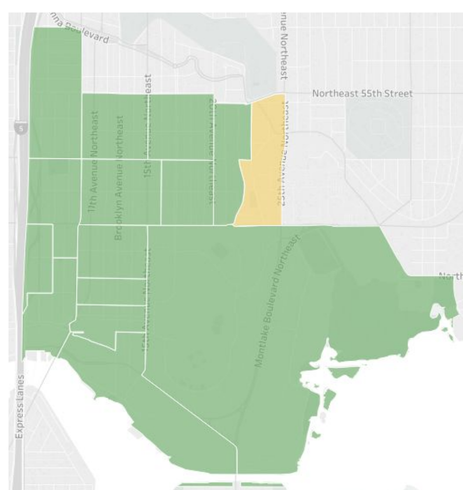
Using the probability of driving alone as the dependent variable, models with just Mode Choice Scores and Affordability Scores yielded high prediction accuracy. Both the Logistic Regression and Support Vector Classifier algorithms yielded about 77% accuracy, comparable to another model with 21 PSRC features, whose accuracy was 80%.

Figure 3. Adjusted mobility scores for Persona B in the University District

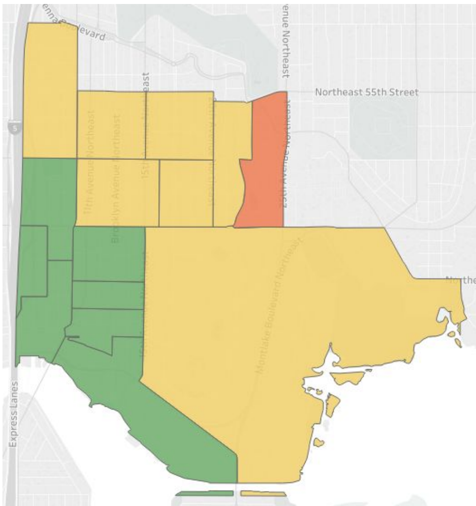
Mode Choice Baseline



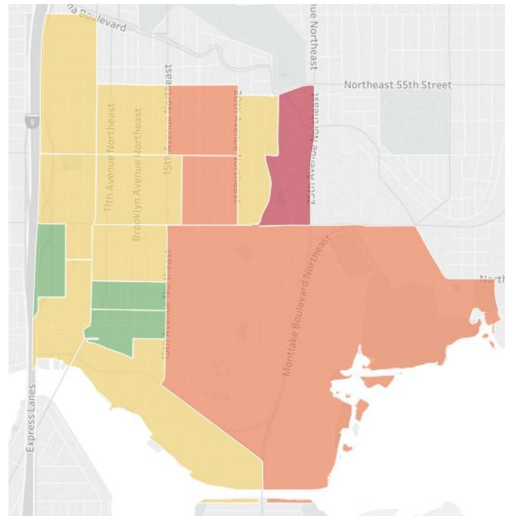
Persona B Mode Choice



Affordability Baseline



Persona B Affordability



2.0 Detailed Description of Components/Features

The overarching goal of the Seattle Mobility Index Project is to produce a mobility measure that is understandable, reproducible, and reflects SDOT's goals of promoting an equitable transportation system. The Seattle Mobility Index Project measures three mobility indices for 481 Census Block Groups in Seattle: transportation mode choice, affordability, and reliability. The goal of this section is to provide a detailed account of the development of the Seattle Mobility Index components, summarized in Figure 1 below.

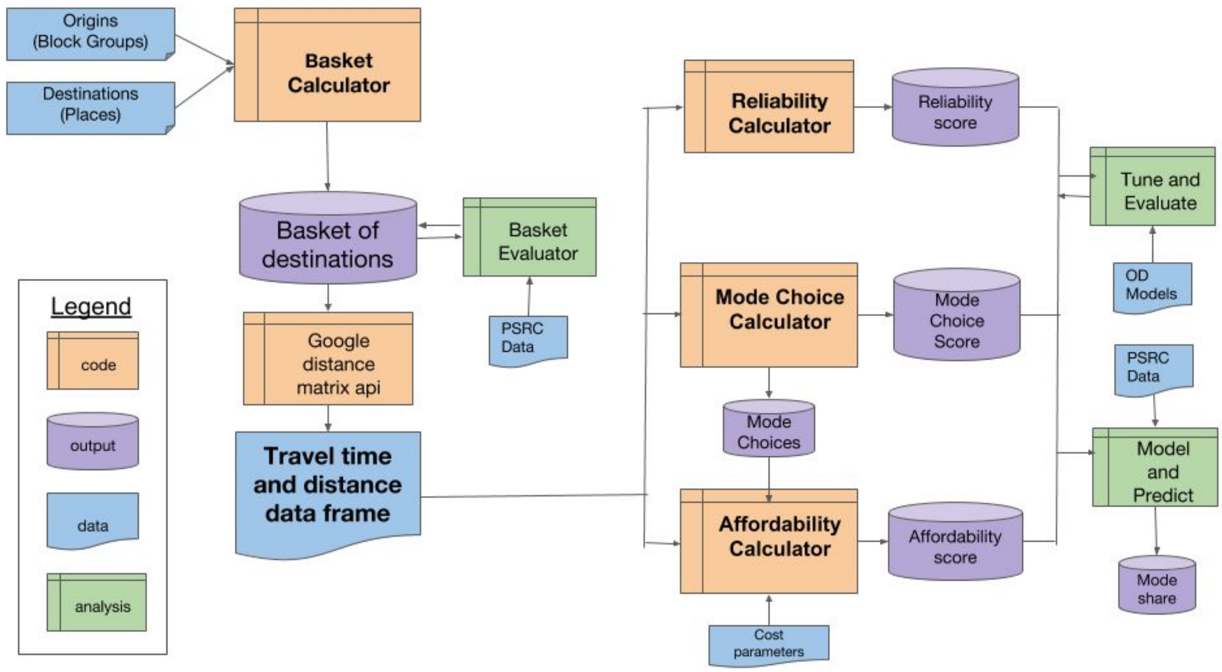


Figure 1. Project Flowchart

2.0.1 Data

To develop the Seattle Mobility Index we relied on publicly available data sources and data generated through the Google Distance Matrix API. The following table describes the data used in the project, their uses, and sources:

Data set	Description	Use	Source
Puget Sound Regional Council (PSRC) 2017 **Household Travel Survey**	Regional survey of Puget Sound that collects day-to-day travel information from households in the area. The objective of the survey is to obtain a complete picture of travel patterns in the region. Questions in the survey seek to understand how people travel, where they go, how long it takes, etc.	The PSRC Household travel survey was used 1) to ensure that the **Market Basket of Destinations** chosen to create the Mobility Scores was representative of how people travel in Seattle. PSRC was used as a baseline data to which compare possible market baskets 2) To develop travel personas. Household, personal, and trip-level datasets in the survey were clustered to generate five **travel personas.**	https://www.psrc.org/household-travel-survey-program
Geographic information for Seattle	We obtained shapefiles for each of the geographies needed to connect our datasets: block groups, neighborhoods, districts, zip codes, and urban villages. These shapefiles store the location, shape, and attributes of geographic features.	The shapefiles were used to construct the **Universal geocoder** and visualize the results of the three indices.	https://www5.kingcounty.gov/gisdataportal/ Files downloaded: **Block groups, Zipcodes, Council Districts:** Blkgrp10_shore.shp zipcode.shp sccdst.shp **Urban villages:** DPD_uvmfg_polygon.shp **Neighborhoods:** Neighborhoods.shp
Seattle parking data	This data set includes blockfaces for all segments of the street network in the City of Seattle. Identifies the attributes of the block relevant to parking: peak	The parking data is used as part of the estimates of driving costs in the affordability calculator.	https://data-seattlecitygis.opendata.arcgis.com/datasets/blockface

	hour restrictions, length of the block, parking categories, and restricted parking zones.		
Google Place data	This data set was generated using the Google Places API. It identifies the location (lat, lon) of all of the places in Seattle that fall within our **Place Categories**.	The place data was used to create the origin-destination pairs needed to create the Market Basket of Destinations. This was made into a matrix of origins and destinations that was used to create the Trip Data.	Generated by the team through the Google Places API https://developers.google.com/places/place-id [link to script]
Google Trip data (collected with Google Distance Matrix API)	We obtained google trip data for all of the origin-destination pairs in each of the Market Baskets for each block group. The Distance Matrix API is a tool that allows you to obtain travel distance and time for a matrix of origins and destinations. The API returns trip data for each origin-destination pair, as calculated by the Google Maps API (based on the Google recommended route), and consists of rows containing duration and distance values for each pair.	The google trip data was used to create the Mode Choice Index, Affordability Index, and Reliability Index.	Generated by the team through the Google distance matrix API https://developers.google.com/maps/documentation/distance-matrix/start [link to script]
Public Use Microdata Sample files (PUMS)	Public Use Microdata Sample files (PUMS) was used files to create a merged, normalized file that can be used for analysis and answer specific queries about demographics in Seattle.	Travel personas analysis	https://www.census.gov/programs-surveys/acs/data/pums.html

2.1 Universal Geocoder

The Seattle Mobility Index Project uses many different types of geographic information. The Universal Geocoder is a tool containing a suite of methods that allow us to retrieve multiple facets of geographical information for a given location. For example, our unit of analysis for the Index is the census block group. Trips are made from one location (origin) to another (destination). Therefore, we needed to identify the block groups to which these locations belong in order to make estimations for the Index at the block group level.

The Universal Geocoder takes in geographical information containing locations (latitude, longitude) in various formats (a point, a csv, or a dataframe), and outputs the attributes of the given point. There are two types of attributes that can be retrieved with this tool:

1. Geographical attributes - Given a point, obtain its geographical attributes including the block group, zip code, district, neighborhood, and urban village (if applicable) in which the point is located.
2. Parking costs - Different locations have different parking rates that are needed in our cost calculations (see Affordability Index). Using the universal geocoder we can also relate our locations to data on the parking rates for these locations.

2.2 Market Basket of Destinations

For the Seattle Mobility Index Project, we constructed a market basket of destinations for each of the 481 block groups (origins) in Seattle. To develop the basket, we started with the following constraints (requested by the client):

- The Market Basket of Destinations should consist of 25 total destinations from both local and city-wide place categories.
- Possible baskets should include the following composition of place categories:
    - 8-13 citywide destinations
    - Up to 4 urban villages
    - Up to 3 of each for the following local destinations:
      destination park, supermarket, library, hospital, pharmacy, post office, school, cafe.

Under these constraints, market basket was developed from the travel information provided by households on their travel patterns in the *Google Matrix API and the PSRC Household Travel Survey data.*

1. **Market basket calculator: Estimate the sets of all possible basket combinations for each block group**

Using our list of place categories, we created a set of all of the possible destinations in Seattle that fall within the previously defined place categories (1617 locations from Google Places).

Next, we estimated the distance from each origin (481 block groups) to each destination using the haversine[5] distance formula. For practical purposes, we limited the set of possible destinations for a

[5] The haversine formula calculates distances between two points on a sphere

given origin by filtering those for which the Haversine distance was less than six miles (except citywide destinations) thereby cutting the universe of origin-destination pairs roughly in half[6].

Finally, we made calls to the Google matrix API to estimate the driving distance between each of the origin-destination pairs, and we ranked them by distance. For each origin, based on the constraints established above, each possible combination of destinations was generated as an array of counts corresponding to values for each category.

2. **Market basket evaluator: Decide on a composition for the market basket that best represents the travel patterns of people in Seattle**

To determine the composition of the final basket, we developed a Market Basket Evaluator that used the *PSRC Household Travel Survey Data* to decide on a combination of 25 destinations that best mirror the travel patterns observed in Seattle. In our project, the PSRC dataset was considered as the gold standard for travel patterns in Seattle, and therefore was used to evaluate the baskets made with Google data in the first part of the basket calculation.

The **Basket Evaluator** ranks the combinations of destinations using three metrics described below. These metrics are useful for comparison because distance and latitude/longitudes of destinations are shared by the two datasets. Each block group has a score for:

- **proximity ratio:** {number of trips under 2 miles} / {number of trips where distance is 2 miles or greater, and under 10 miles}
- **average distance from origins to destinations** (using distance from Google Distance Matrix API)
- **distance from city center** (uses Euclidean distance between the latitude and longitude of destinations of each block group and the city center[7]).

For each block group, these metrics were estimated for the PSRC data and Google data filtered by each possible market basket of destinations[8]. A block group's score for each of the above is the mean value for each metric for all trips starting from that block group. The Mean Squared Errors (MSEs)for the three metrics were calculated and ranked.

Final ranking was calculated based on the rank sum of three MSEs. The team manually selected the final basket based on the final ranking, based on its high ranking and composition of places that had at least one item in each category.. The chosen basket had the following composition:

- 13 citywide destinations
- 3 schools
- 2 supermarkets
- 1 destination park, post office, cafe, library, hospital, urban village, and pharmacy.

3. **Create the final basket for each block group**

---

[6] 481 block groups * 1617 destinations = 777296 OD pairs, filtered using haversine: 384272 pairs
[7] CITY_CENTER: latitude: 47.6062, longitude: -122.3321
[8] The PSRC data only has 150 of the 481 block groups in Seattle. The comparison of the data in the basket evaluator is limited to the block groups included in the PSRC.

Given our final basket composition, we filtered Google Place data for each of the 481 Seattle block groups using the destinations that are closest to each block group origin within each place category.

4. **Query Google Distance Matrix API get acquire final travel data set.**

Google Distance Matrix API was queried to get distance and duration data for four transportation modes based on recommended routes between each of the block groups and its 25 market basket of destinations. For driving and transit, data were downloaded at every hour from 7 AM to 8 PM. This totals 750 trips for each block group. The final dataset was then used for calculating the Mobility Scores.

2.3 Personas

The travel personas were developed to fine-tune the mobility scores to reflect how people with varying characteristics have different needs and experiences with the transportation system. The goal of this analysis is to identify the representative types of people to model how mobility differs for various groups.

We used the *PSRC Household Travel Survey Data* to identify the personas. The survey collects three levels of data that are linked by a respondent ID: person level attributes (sex, race, age, education), trip characteristics (duration, purpose, primary mode, distance) and household level socioeconomic characteristics (income, household size, number of vehicles, off-street parking places, years of residency). K-means clustering algorithm was applied to all features, except race and sex, in order to identify five different groups. For each subgroup the summary of the variables was obtained and described. These averages were then applied to modify travel duration thresholds for the mode choice and affordability score calculations.

The personas identified had the following characteristics:

Persona	Description
A	Have 1-2 children, drive most of the time
B	Higher proportion without a car
C	Average age 65, high proportion of non-workers
D	Average 9 years of residence in current house
E	Long-distance commuter

2.4 Mobility Index Score

The data from Google Distance Matrix API was used to calculate the three mobility indices: transportation mode choice, affordability, and reliability for each block group.

2.4.1 Mode Choice Calculator

The mode choice index is a measurement of the number of modes available to reach the basket of destinations from each origin (block group). The Mode Choice Index is a score from 0-100 that measures the extent to which destinations are reachable in a 'reasonable' amount of time from each block group.

In the mode choice calculator, the amount of time considered reasonable for a trip is determined by the typical duration of similar trips in Seattle. Trips are defined as *viable* if the duration required to complete that trip is below a threshold, specific to each particular mode. Baseline thresholds are currently manually set, informed in part by the average duration of trips per mode in PSRC data (Bike, 45 mins; Walk, 45 mins; Car, 30 mins; Transit, 60 mins).

To construct the index, we aggregate trips to 25 locations from the basket of destinations for each block group and mode, and then we assign a score based on the number of *viable* trips relative to the total trips from that block group. The mode scores are then averaged across modes (unweighted) to produce the final Mode Viability Score.

Scores:

81–100	**Excellent Mobility** You have your choice of modes for trips within an acceptable duration
61-80	**Good Mobility** Most mode options for trips within an acceptable duration
45-60	**Minimal Mobility** Some mode options for trips within an acceptable duration
0–45	**Low Mobility** Few mode choices / trips are generally relatively long

We adjust the viability using the thresholds per mode obtained for each of the Personas. These thresholds are informed by the 85th percentile for trip durations based on their typical travel behavior determined from PSRC data. That is, for a given Persona, we use the 85th percentile to find the mean trip duration for each of the four modes (Table X). This adjustment allows us to fine-tune the mobility index for different types of people who have distinct needs and travel preferences.

The travel duration thresholds for each persona are as follows:

	Driving	Transit	Biking	Walking
Persona A	27	52	64	40
Persona B	27	49	10	32
Persona C	28	57	34	37
Persona D	31	46	10	71
Persona E	34	71	45	54

2.4.2 Affordability Calculator

The affordability index measures the relative cost to reach 25 destinations in Seattle (13 city-wide locations and 12 local locations of different categories) from each of the 481 block groups. For each block group, the average cost of the viable trips are calculated using mode specific costs and the value of travel time. Out of the 750 possible trips for each block group, a subset is defined as "viable" by the "viable mode calculator". Only the viable modes are used to calculate the affordability of the trips. The average cost for a trip for each block group is normalized to create an index that scales between 0 - 100.

For all travel mode, a standard rate of travel time savings at $14.10 per hour[9] and the following mode specific costs are incorporated into the affordability calculation:

Driving:

- The AAA all-inclusive cost per mile is $0.56. This includes estimates for fuel maintenance, repair, tires, insurance, license and registration, depreciation, and finance and assumes a 15,000 miles per year[10].
- Average parking cost of the destination block group calculated from blockface level parking information[11].

Transit:

- Fare value of transit calculated by the Google Distance Matrix API.

Biking:

- The higher end of all-inclusive cost per mile for bicycle trips is $0.15[12]. This assumes the purchase of a $1000 bike ridden for 2000 annual miles for 10-years.

Walking:

- Walking is estimated at $0 per mile.

For all viable trips in a block group, the cost of each trip is summed and divided by the total number of viable trips to calculate the average trip cost of each block group. This value is normalized to create an Affordability Index that scales between 0 - 100 by:

Average cost of each block group - lowest average cost / Highest average cost - lowest average cost

85–100	**Excellent Affordability** You spend less than the average Seattleite on transport

[9] DOT, US. "Revised departmental guidance: Valuation of travel time in economic analysis." *US Department of Transportation, Washington, DC* (2016).
[10] American Automobile Association. "Your Driving Costs: How Much are you Really Paying to Drive? Heathrow, FL." (2017).
[11] Open Data. City of Seattle, Washington, 2017. https://data.seattle.gov.
[12] Litman, Todd. "Transportation cost and benefit analysis." *Victoria Transport Policy Institute* 31 (2009).

65–84	**Good Affordability** You spend as much as an average Seattleite on transport
50-64	**Fair Affordability** You spend more than the average Seattleite on transport
25–49	**Minimal Affordability** You spend about 1.5 times more than the average Seattleite on transport
0-24	**No Affordability** You spend about 2 times more than the average Seattleite on transport

2.4.3 Reliability

The reliability calculator estimates an index that provides a measure of the consistency in day-to-day travel time for trips to the basket of destinations from each origin (block group). This is based on travel time for driving trips collected over 30 days to capture the variation in traffic across a month.

The duration of each trip including traffic was used to calculate the percentage of trips under 85th percentile travel time for every hour from 7 AM to 8 PM and averaged to create a single percentage for each block group. The results were scaled between 1- 100 indicating the consistency of travel times for each block group.

xx–xx	**Excellent Reliability**
xx–xx	**Good Reliability**
xx-xx	**Fair Reliability**
xx-xx	**Minimal Reliability**
xx–xx	**No Reliability**

2.4.5 Prediction with Mobility Scores

One way to verify the usefulness of metrics is to see if the newly-created features work effectively as predictors. Our Mode Choice, Affordability and Reliability scores would therefore be even more meaningful if they would work well in prediction problems, for example predicting mode share.

We constructed three machine learning algorithms to compare the prediction accuracy, using the logistic regression, random forests, and support vector classifier from the Sklearn library of Python. A binary outcome variable, called Drive_Alone_Threshold, was created, where the value was 1 if the proportion of driving alone as a mode choice is greater than its median value, and 0 otherwise. This variable is a good indicator of mode share, because if one is not driving alone, he or she will be driving with others, taking bus/train/ferry, bicycling, or walking.

Here, we compared two models:

1. The first model uses only Mode Choice score and Affordability score as independent variables
2. The second model included many other combinations of variables from PSRC data, which had many available features such as travel duration, distance, household size, income level, number of children, mean age, year of residence, number of off-street parking spots, vehicle ownership, and number of trips. One important combination is a model with just travel duration, distance, and income level, as it can be seen as containing similar amount of information as the two mobility indices.

We used Random Forest Classifier with maximum depth 100 and random state 0, and Support Vector Classifier with linear kernel and default parameters. A stratified K-fold Cross validation with K=10 and K=5 were used to examine the prediction accuracy; Means and standard deviations of prediction accuracy were used to evaluate each models.

3.0 Technical Specifications

3.1 How to Use this Software

3.1.1 Running the scripts

Scripts must be run from the seamo/ directory.
The scripts rely on **data** (seamo/data) and **constants** (support/constants.py)

Market Basket Evaluator
```
$ python analysis/market_basket_evaluator.py
```

Note: This will take just over an hour to complete.

Market Basket Calculator
```
$ python core/main_basket_evaluator.py
```

Things you'll need:
- a Google Distance Matrix API key
  (You will be prompted to enter a key when you run the script from the command line)

This will take several days to run. Be mindful of limits on API calls.

To run tests on the basket calculator (currently not many tests!):
```
$ python tests/test_basket_calculator.py
```

Mobility Indices
```
$ python core/index_calculation_driver.py
```

Currently, this produces mode choice & affordability scores for weekday data.

Persona-dependent Indices
```
$ python core/persona_avail.py
```

This produces mode choice & affordability scores for each persona for weekday data as a csv file in data/processed/csvfiles folder.

3.1.2 Updating the data

Geocoder
This module utilizes a reference GeoDataFrame built from shapefiles (see: *Geographic information for Seattle* in Data section).

Geocoder shapefiles used: Blkgrp10_shore, zipcode, sccdst.shp, DPD_uvmfg_polygon, Neighborhoods
If datasets are updated, replace shapefile in raw data directory
```
$ rm data/processed/pickles/reference.pickle
```
Next time geocoder is run, a new pickle will be generated

Parking geocoder shapefiles used: Blockface
If datasets are updated, replace shapefile in raw data directory
```
$ rm data/processed/pickles/parking_districtX.pickle
```
where X is between 1 and 7
Next time the parking geocoder is run, a new pickle will be generated

Seattle Census Block Groups
File generated with blockgroups and geographical information
```
$ python preproc/geography_processor.py
```
Note: Compatibility of this script may have broken with new repo structure. This file is not necessary for other components of index. It is useful for visualizations in Tableau

3.2 Data Processing / Storage (Local, recs for future)

Parking Data
```
$ python preproc/generate_parking_data_driver.py
```

DynamoDB Out
Open `query_dynamodb.py`, specify mode of transport
```
$ python support/query_dynamodb.py
```
Repeat above two steps for all modes
```
$ python preproc/convert_dynamodb_driver.py
```

3.3 File Organization, Git/Workflow and Code Style

https://github.com/monolyst/Seattle-Mobility-Index

Code style: PEP 8

3.3.1 Constants / Classes / Tests

Tests are in seamo/tests

Constants are implemented in seamo/support/constants.py

The Coordinate class is implemented in seamo/support/coordinate.py

4.0 Design Process

In the section below we describe our design process, stakeholder involvement, lessons learned and other components that informed the makeup of the project.

4.1 Project Summary

The following is a summary of the project over the 10 week span.

The project team convened and started to coalesce around the project's purpose, its deliverables and methodologies. We produced a data flow diagram [see figure 1] to visualize the entire software ecosystem - making it clear how data is being processed through each module and what is being output. With dependencies and priorities clarified, fellows broke off into small groups and started developing the modules..

The project team visited SDOT headquarters to gather insights from stakeholders. At the end of week 3, we finished the first phase of the software - the Universal Geocoder, Market Basket Calculator, and Market Basket Evaluator.

In response to a stakeholder's suggestion, we developed a set of traveler personas data that can be used to model traveler behavior based off demographic clusters of people. When we finalized the market basket of destinations combination, we made calls to the Google API in order to get the travel times and distances data we need for the rest of our modules (Mode Choice, Affordability, Reliability).

At week 7, we were finalizing the calculations for the personas, and inserting them into each of the modules. We started visualizing the data that we output and finalizing the scores for each metric. After compiling some early analysis and writing one page descriptions of each feature, we sent on to stakeholders to receive final feedback before the project concludes. While refining, tuning and optimizing the code to be ready for delivery, we produced and practiced our final presentation.

4.2 Stakeholder Engagement and Analysis

Meeting the project's stakeholders and regularly communicating with them ensured that the design and functionality of the project satisfied their needs.

4.2.1 Field Trip

The project team visited the Seattle Department of Transportation office on June 25th, to meet with stakeholders and discuss and document feedback on the project's goals, uses and any other ideas.

4.2.2 Stakeholder Analysis

In our stakeholder analysis we seek to map the larger interdependencies and connectedness of the system in which we are operating. This includes:

- identifying the actors that have (or should have) influence in our final product.
- who will determine the success of our project
- who might be impacted by our results or have an interest in using the final product
- what negative uses might be accounted for

To develop understanding of the environment in which we are operating, we mapped a number of stakeholders who have shaped our understanding of this project and/or will be impacted directly or indirectly by this project. In this exercise we defined four categories of stakeholders:

User	User profile	Stakeholder needs and uses
1. PRIMARY STAKEHOLDER		
Seattle Department of Transportation (SDOT) Strategic Data Initiative	SDOT is our primary user and the key stakeholder for who we are creating the Seattle Mobility Index. Potential users include transportation planners, analysts, advisors and managers.	SDOT will use our Mobility indices to understand people's mobility, measure transportation system over time, and refine the codes/algorithms over time. SDOT's main user needs are: - Code/algorithms to calculate mobility index that is reproducible - To understand people's behaviors in terms of mode choice, reliability, affordability - To help inform better policy decisions
2. SECONDARY STAKEHOLDERS WHO ARE DIRECTLY ENGAGED		
Mayor's office of Seattle (Innovation & Performance team, Smart Cities Initiative)	*The Innovation & Performance team is a unit focused on improving the effectiveness, efficiency and accountability of City government. They use data and design to support management and operational decision-making, and seek to deliver sustainable solutions that can be used independently by departments* (https://www.seattle.gov/innovation-and-performance/about-us).	The Innovation & Performance team would like to use this project as an example of a data-driven project that the Mayor's office can showcase. This is in line with their stated objective of prioritizing data and performance-driven policy. Additionally, the Mobility Indices can be used as a data driven-tool to support and evaluate policy decisions.
City of Seattle RSJI (Race and Social Justice Initiative), Transportation Equity Program	*The Seattle Race and Social Justice Initiative (RSJI) is a citywide effort to change the underlying system that creates race-based disparities in the community and to achieve racial*	The Seattle Race and Social Justice Initiative, together with the Transportation Equity Program would use the tool to further understand what groups are underserved by transportation. They can also utilize and extend the traveler personas we created to perform other analyses and to use as communication tools.

	equity. (https://www.seattle.gov/rsji/about) *In alignment with the City of Seattle's Race and Social Justice Initiative goals and core values the Transportation Equity Program provides safe, environmentally sustainable, accessible, and affordable transportation options that support […] marginalized communities to thrive in place in vibrant and healthy communities, and mitigate racial disparities and the effects of displacement*. (https://www.seattle.gov/transportation/projects-and-programs/programs/transportation-equity-program)	
New Mobility Program	*New mobility are those emerging elements of the transportation system that are enabled by digital technology, shared, driven by real-time data, and often providing curb-to-curb transportation. It allows Seattleites to treat urban transportation as a customizable, on-demand service.* http://www.seattle.gov/transportation/projects-and-programs/programs/new-mobility-program	
3. TERTIARY STAKEHOLDERS (WITHOUT DIRECT ENGAGEMENT)		
Other cities	Other cities who would be interested in developing a local Mobility Index.	The Mobility Index Tool will be designed with reproducibility in mind. Other cities will be able to use this software to replicate the index based on local data.

City planners	Local City Planners seeking to inform planning decisions and policy outcomes.	Make informed data-driven decisions. The index may give city planners more nuanced insight into what people's mobility needs are in any given area.
Community Organizations	Local non-profit organizations dedicated to advocacy in the transportation and urban mobility sector.	Advocate for increasing mobility and disadvantaged areas, and to help prioritize advocacy efforts.
Transportation researchers and policy analysts	Academic and non-academic researchers in transportation and policy.	Using data to conduct research. For example, to use the tool for assessing impacts of transportation policy
Civic technologists	People who live in or travel to Seattle, open data proponents like Open Seattle.	To explore and discuss data. Also to extend, test and improve the software.
4. NEGATIVE USERS		
Real estate and marketing users	Commercial Real-estate and Marketing firms	Use Mobility Index to advertise housing and drive up housing prices
Private charter services	Private transportation service providers	Use mobility index to identify customer base/location and potential demand

4.3 Lessons Learned / False Starts and Dead Ends.

The following are some major challenges we faced throughout our project.

Data gaps: During the project, we faced various constraints regarding data:

- **The PSRC survey is not completely representative of Seattle.** We often used the PSRC as the gold standard to which we compared travel patterns in Seattle (for the Basket calculator and for personas). However, the PSRC is not representative of the whole population of Seattle, for example low-income groups or young people. This was especially notable we we developed the travel personas because we noted that groups that may need the most attention (and therefore may need to be adequately characterized as personas) are be adequately represented in the survey.
- **Lack of transit data.** We did not have adequate geospatial data about public transit in Seattle (location of stops, routes schedules, how often buses are early and late), and therefore could not adequately conceptualize accessibility and reliability of transit. This data could potentially be collected through the One Bus Away API, but we did not have time to explore this option.

- **Disconnect between parking data and other geospatial data:** The parking data used was by block face (line) rather than block group (shape).To manage this we had to create a buffer around each one and used the block group centroid to be able to geocode the block faces.
- We would have benefitted from having access to a reliable server for all the API calls. Not having a committed server greatly slowed down the data collection process.
- When making API calls, we should have retained a log of things that failed. Also, some of the data that we needed to identify duplicates in the data (like destination class) should have been retained when collecting the Google data.

Technical challenges:

- In general, we underestimated how long things would take because there was a steep learning curve in this project.
  - Something that should have had more thought before starting to code was the adequate data types that would make our tool mode efficient. We had to experiment a lot to get it right which took up time that we could have saved by planning better.
  - We would have spent a lot more time preprocessing data to avoid issues we encountered when developing the front end of the tool. Also, setting up the tool so that changes in the data are done upstream would have helped.
- The code was often slow and challenging to run. Vectorizing to improve performance instead of using loops was a valuable way to make our code more efficient.
- On occasion we had to work with libraries outside of python (for example the geospatial library gdal in the universal geocoder). These libraries are challenging because they sometimes break other python packages in the tool when they are updated. Special attention is needed when packages conflict.
- Thresholds
- Conceptualizing the software environment

4.5 Future Work

The project is flexible to added features and data. Here are several possibilities:
Transit Reliability

This variable could be add route and stop specific reliability measures to add further accuracy to our reliability index. Some data can also be accessed through the King County Metro's GTFS feed.

Another method we researched as a possibility is the "Transit Run Time Variability" measure from the Portland Bureau of Transportation's (PBOT) Enhanced Transit Corridors Plan.

Incorporating Carshare, Rideshare and Bike Share (TNCs)

There are several private transportation companies (Lyft, Uber, Car2Go, ReachNow, Lime, SPIN, Ofo) that we could incorporate into the software to add further accuracy.

Comparing Google and Bing travel data

If we access trip duration and distance data from Bing, we could utilize our software to compare, and verify trip data.

Incorporate Census Data into Persona Analysis

Using US Census data for our personas would add depth and accuracy, and make it more extendable to other users as a national standard for demographic data.

Other Research Questions

- How long it takes to get to basket of destinations in car vs transit?

4.6 Highlighted Project Documents

4.7.1 Project Charter

An early document outlining project roles, expectations, technical and non-technical skills, project vision, and success criteria.

4.7.2 Implementation Summary / Workplan

A list of primary goals and their key actions to help the team organize tasks and meet milestones.

4.8.3 Data Flow Diagram

A diagram of how data is being processed through different code modules, and what is being output.

4.8.4 Data Science Problem Template

A set of prompts for potential users, and stakeholders to collection information that would inform potential use cases for us to design around.

4.8.5 Technical Plan

This document discusses technical planning for the DSSG 2018 Seattle Mobility Index Project. Prepared by the Data Scientist consultants on the project.

4.8.6 Technical Onboarding

This document collects the technical tools needed for the project. Prepared by the Data Scientist consultants on the project.

References

114,000 more people: Seattle now decade's fastest-growing big city in all of U.S. Gene Balk, Seattle Times, May 24, 2018
https://www.seattletimes.com/seattle-news/data/114000-more-people-seattle-now-this-decades-fastest-growing-big-city-in-all-of-united-states/

Greater Seattle area leads the nation in transit ridership growth
https://kingcounty.gov/elected/executive/constantine/news/release/2018/February/21-metro-ridership.aspx

Seattle Has 'Transit Deserts' and—Surprise—It's No Fun to Ride the Bus There
https://www.seattlemet.com/articles/2017/11/20/what-s-it-like-taking-the-bus-in-seattle-s-transit-deserts

If you think Seattle traffic is bad now, just wait until these projects start. Mike Lindblom. Seattle Times January 7 2018.
https://projects.seattletimes.com/2018/one-center-city/

Household Travel Survey Program
https://www.psrc.org/household-travel-survey-program

King County GIS data portal
https://www5.kingcounty.gov/gisdataportal/

Open Data. City of Seattle, Washington, 2017. https://data.seattle.gov.
https://data-seattlecitygis.opendata.arcgis.com/datasets/blockface

Public Use Microdata Sample (PUMS)
https://www.census.gov/programs-surveys/acs/data/pums.html

DOT, US. "Revised departmental guidance: Valuation of travel time in economic analysis." *US Department of Transportation, Washington, DC* (2016).

American Automobile Association. "Your Driving Costs: How Much are you Really Paying to Drive? Heathrow, FL." (2017).

Litman, Todd. "Transportation cost and benefit analysis." *Victoria Transport Policy Institute* 31 (2009).

6.0 Appendix

6.1 Highlighted Documents

Project Charter

Implementation Summary / Workplan

Data Flow Diagram

6.2 Terminology

- **Affordability calculator**. A tool that calculates the cost (in dollars) to travel to a destination from a block group using a particular transportation mode (which may be affected by time of day as well).
- **Market basket of destinations**. A data set of place categories that are travel destinations from a block group.
  - Basket of 25 latitudinal and longitudinal locations, eached mapped to a set of standard set of place categories weighted by relevance
  - 9 local categories defined by SDOT -- want to pick relevant max 25 total destinations for a block group.
  - Don't know how a standard measure of mobility will be determined from this as yet

  - What is a destination? (place with unique idea, category, etc)
- **Block group**. A geographical organization.
  - Based off centroid of Census block group designations
- **Basket of destinations calculator**. A tool that calculates the local destinations for the market basket.
- **Distance matrix**. A data set that specifies the travel time, mode, distance in miles and time stamp from a origin (block group) to a destination (latitude and longitude).
- **Geography**. A way to subdivide an area (e.g., zip code, census tract).
- **Universal geocoder**. A tool that classifies a point (latitude, longitude) for multiple geographies (e.g., block group, census tract, zip code).
- **Mode choice calculator**. A tool that determines the preferred transportation mode from a block group to a destination.
- **Transportation mode**. A way in which transportation is done. Modes considered are: walking, bicycling, driving, transit.
  - Conduct literature search to determine additional possible modes
  - Currently 4, can add additional modes in the future