Transit Planning Analysis Playbook

A Decision Support Tool

June 2023

Table of Contents

Introduction

The Transit Planning Analysis Workflows Playbook serves as a guide, offering blueprints for transit planners to replicate successful analysis workflows derived from real-life case studies. This resource aims to empower transit professionals with practical tools and methodologies to optimize their planning processes, enhance decision-making, and ultimately create more efficient and sustainable transit systems. By drawing insights from various successful projects, this playbook provides a roadmap to navigate the complex landscape of transit planning, enabling planners to adapt and implement proven strategies in their own contexts.

This document is a collaborative effort, championed by a team of experienced transit planners and data analysts. These practitioners have curated a collection of case studies, distilled best practices, and crafted a playbook that aims to encompass a wide range of analysis workflows. The playbook is primarily designed for transit planners and professionals involved in the process of designing, developing, and optimizing transit systems. It caters to individuals working in government agencies, transportation authorities, and metropolitan planning organizations.

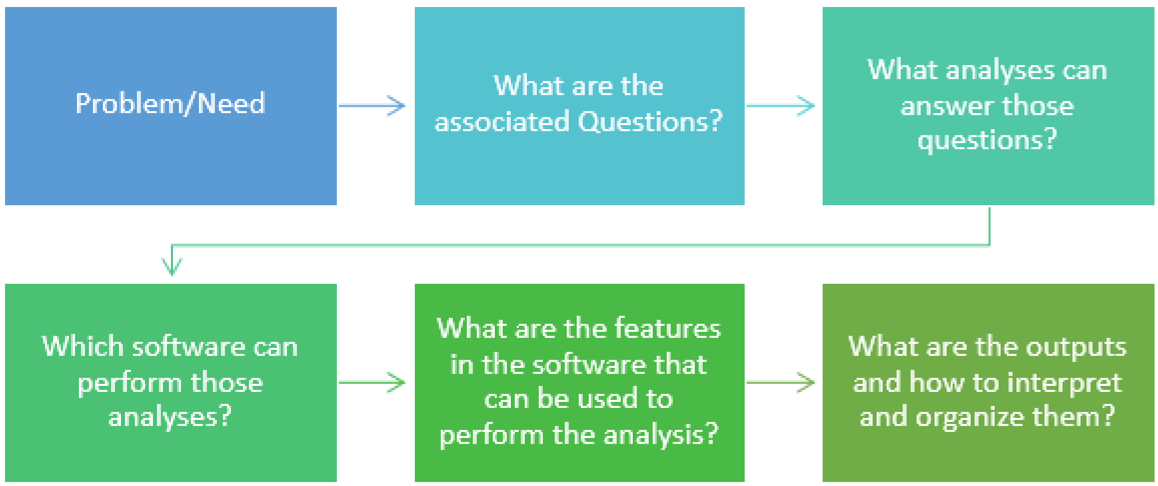
The Transit Planning Analysis Workflows Playbook intends to assist transit planners as a go-to resource to:

- Learn from successful case studies
- Access proven analysis workflows
- Support decision-making processes
- Promote collaboration and knowledge sharing

How Does This Playbook Work?

Workflows

Each collected case study added to the Playbook features a similar workflow. First, a problem or need is identified. In order to address this problem or need, the case study would determine which questions need to be asked. One or more analyses are constructed to provide answers to the questions. Software and their relevant features are evaluated to pair with each needed analysis. Finally, the outputs, their interpretation, and how to organize them is determined. This process standardizes the Playbook format for all included case studies solicited through the Community of Practice.



Workflows consist of three primary sections–a flowchart, narrative, and an in-depth technical analysis. The flowchart depicts the pipeline shown above and the specific approaches used to answer each need. The narrative contextualizes the flowchart by providing rationale for each step and key information like case study title, date drafted, software used, and a hyperlink to the corresponding technical analysis. The Technical Analysis portion of each case study is a long-form, step-by-step instruction organized into an appendix.

Playbook Maintenance

Any technical documents produced for case studies are included as an appendix in the Playbook. Technical documents cataloged in this Playbook should include the date of the case study and the specific software versions used to perform the analysis. Deprecated analyses will require updates and replacements each time a new and relevant case study was conducted to answer a similar question or need as a previous case study.

Workflows

This Playbook aims to gather and document workflows that address various problems or needs. Each collected case study follows a standardized workflow to ensure consistency and effective analysis:

1. Problem/Need Identification
   a. The first step is to identify a problem or need that requires attention. Understanding the context and specific challenges is essential for developing appropriate solutions.
2. Determining Key Questions:
   a. Once the problem or need has been identified, the workflow proceeds by determining the key questions that need to be answered. These questions will guide the analysis process and help in extracting relevant insights.
3. Constructing Analyses:
   a. With the key questions in mind, the workflow proceeds to construct the analysis. This step involves designing an analytical framework or multiple frameworks to provide answers to the identified questions. The choice of analysis method depends on the nature of the problem and the available data.
4. Evaluating Software and Pairing Features:
   a. To execute the analysis effectively, the workflow evaluates different software options. The goal is to select software that aligns with the specific analysis needs. This step involves pairing the appropriate features from the software to the required analysis, ensuring that the selected tools can generate the desired insights.
5. Determining Outputs and Interpretation:
   a. Once the analyses are complete, the workflow focuses on determining the outputs and their interpretation. The findings need to be presented in a clear and understandable manner. Additionally, it is crucial to consider the limitations of the analysis and address any potential biases or uncertainties. This step ensures that the interpretation is accurate and comprehensive.

The following tables of *Workflow by Need* and *Workflow by Question* can be utilized for quick reference.

Workflows by Need

Need	Workflow
I need to create a transit network ecosystem to better understand systemic changes.	• Travel Time Impacts of Systemic Changes
I need to determine what systemic travel times impacts occur if a specific route is changed.	• Travel Time Impacts of Systemic Changes
I need to understand who is being served by the current routes of interest.	• Scoring and Comparing Routes
I need to determine the impact to ridership of each fixed route's frequency increase.	• Scoring and Comparing Routes
I need to compare results and validate outputs.	• Scoring and Comparing Routes
I need to understand which Origin-Destination pairings have the greatest transit demand.	• Aligning Service with Demand
I need to develop transit ridership forecasts where an MPO model is not available.	• Aligning Service with Demand
I need to understand the ridership impacts of redesigning bus routes.	• Aligning Service with Demand
I need to reconfigure the transit routes to better serve the population.	• Modeling Network Ridership
I need to test the efficacy of the new system.	• Modeling Network Ridership

Workflows by Question

Question	Workflow
How long does it take to travel to and from select O-D in the current system?	• Travel Time Impacts of Systemic Changes
What systemic travel time changes occur if a select route is truncated or removed?	• Travel Time Impacts of Systemic Changes
What demographics are each of the routes of interest currently serving?	• Scoring and Comparing Routes
What fixed-routes routes would an increase in service frequency be most beneficial to focused demographics?	• Scoring and Comparing Routes
How do the two individual outputs compare with each other?	• Scoring and Comparing Routes
Which areas within the service area have the highest propensity to use the transit system?	• Aligning Service with Demand
What is the systemic impact on ridership of specific redesign changes?	• Aligning Service with Demand
Where is the latent transit ridership?	• Modeling Network Ridership
How would ridership change based on the proposed route modifications?	• Modeling Network Ridership

Travel Time Impacts of Systemic Changes

Westchester County Bee Line Case Study (2022-2023)

Link to Tutorial Appendix

Needs Met	Questions Answered	Analyses Performed
1. I need to create a transit network ecosystem to better understand systemic changes. 2. I need to determine what systemic travel times impacts occur if a specific route is changed.	1. How long does it take to travel to and from select O-D in the current system? 2. What systemic travel time changes occur if a select route is truncated or removed?	1. Create a transit network ecosystem. 2. Construct an existing conditions baseline of travel times from select O-Ds. 3. Adjust the GTFS network to reflect route variants: removal and truncation. 4. Compare the GTFS variants to the baseline O-D travel times.

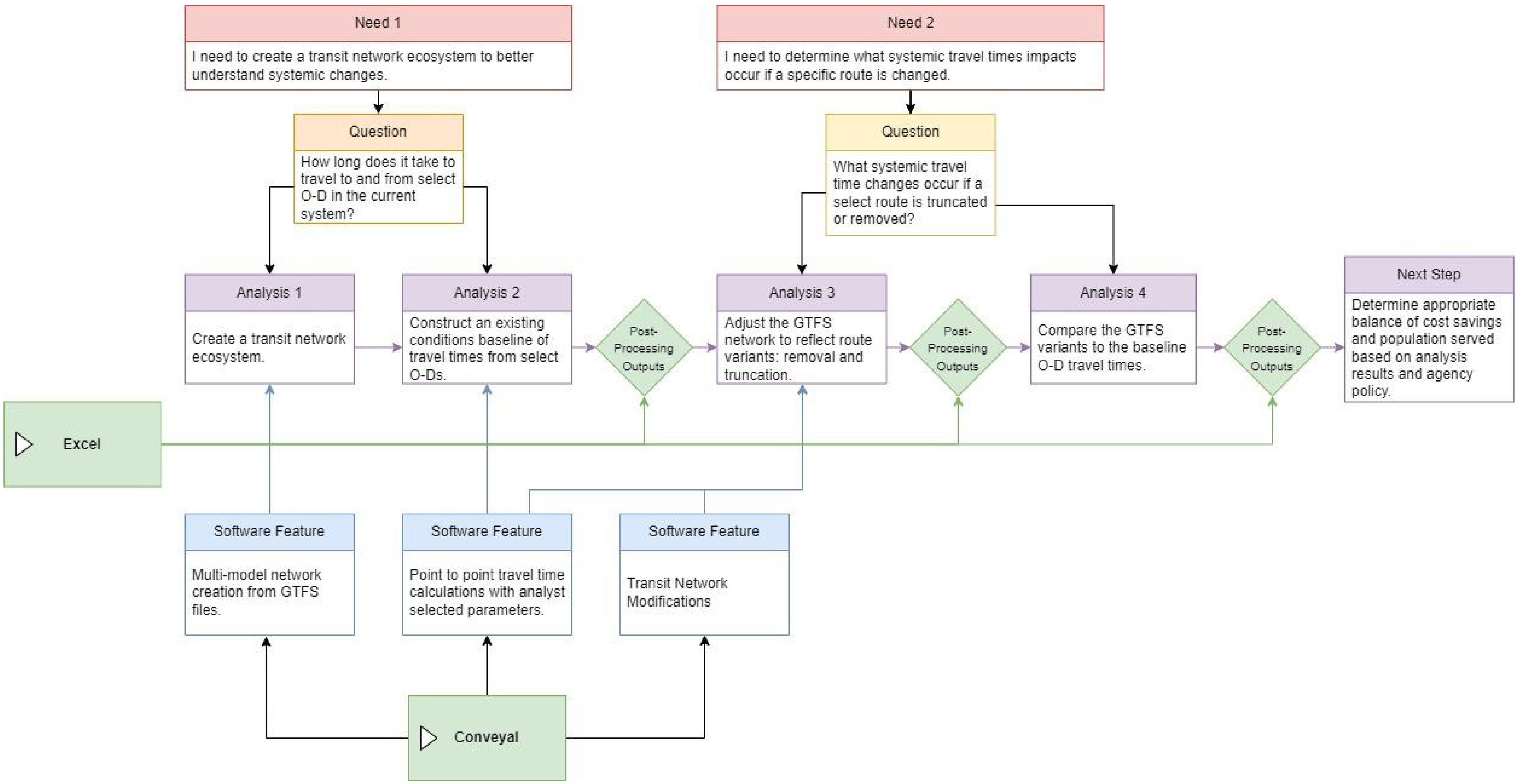
Software Requirements	
Software	**Features Used**
Conveyal	• *Multi-Model Network Creation from GTFS Files and OSM* • *Point to Point Travel Time Calculations* • *Transit Network Modifications*
Excel	• *Post-Processing with INDEX and MATCH Functions*

Case Study Product			
Title:	*The Bee-Line – Westchester: Analysis of Proposed Route Elimination*	**Date:**	*01/19/2023*

Description

This workflow is the result of a pilot project that sought to determine what impact the removal or truncation of a specific route's service would have on the transit network ecosystem and how existing service can meet the needs of current ridership.

Flowchart - Westchester Case Study



Workflow Narrative

Purpose

The purpose of this workflow is to determine what impact the removal or truncation of a specific route's service would have on travel times from many origins to many destinations within a transit network ecosystem. The output will assist the analyst determine how existing service can meet the needs of current ridership.

Problem/Need Identification

This workflow has two clear needs;

- Need 1: a transit network ecosystem to test changes; and
- Need 2: a way to compare the impacts of these changes.

Based on the purpose of this workflow, the transit network ecosystem will need to comprehensively include as many potential travel modes–meaning the test environment must support multi-model GTFS files. The metric used to measure performance within this workflow is travel time.

Key Questions

Considering the needs and metrics, two questions can be formulated;

- Question 1: How long does it take to travel to and from select O-D in the current system?
- Question 2: What systemic travel time changes occur if a select route is truncated or removed?

Analysis Construction and Software Pairing

Analysis 1: Construct a Transit Network Ecosystem

To meet Need 1 and Question 1, a multi-modal transit network ecosystem must be constructed. There are multiple software options available for this, but this workflow uses Conveyal. The analyst (likely in conjunction with a software engineer) will create a package of all transit options' GTFS files as a bundle. The bundle is loaded into Conveyal's interface creating the environment.

Analysis 2: Construct an Existing Conditions Baseline of Travel Times from Select Origins and Destinations

To meet Need 1 and Question 1, the travel times of the current system need to be benchmarked so changes have a baseline to compare against. The analyst must construct a list of select coordinates that can be loaded into Conveyal. Once complete, the analyst runs Conveyal's "Regional Analysis" to calculate travel times for those select O-Ds. Conveyal will output a CSV file that contains a list of each coordinate as an origin to each coordinate as a destination and the travel time between them. To make this output actionable, the analyst should post-process this output using Excel. Post-processing involved making an O-D matrix that indexed travel times to each O-D combination.

Analysis 3: Adjust the GTFS Network to Reflect Route Variants like Removal and Truncation

This analysis utilizes Conveyal's "Modifications" feature where an analyst can make changes to one or more routes that will be reflected in the Regional Analysis output's travel times. Analysis 3 can be split into three parts;

1. Run Regional Analysis with a modification that removes a specific route entirely.
2. Run Regional Analysis with a modification that removes specific stops from a route truncating it.
3. Construct Matrices for both outputs.

Analysis 4: Compare the GTFS Variants to the Baseline O-D Travel Times

Comparing the variants to the baseline travel times is relatively simple since all of the outputs are in the same matrix formatted. Using Excel, the analyst can put all matrices as separate sheets in a single document. By entering a subtraction equation that references the other sheets, baseline and modification, a difference matrix can be constructed.

Outputs and Interpretation

If by removing the route a large cohort of ridership is impacted negatively with dramatically longer travel times, the analyst should identify a point of truncation that has the least impact of travel time increase. The analyst should test truncation at all of the most impacted stops sequentially. When comparing truncation points, the point of diminishing returns is indicative of the desired stop.

Next Steps

This workflow focuses entirely on the metric of travel time as its sole performance measure. While this is a major component of transit planning, other metrics need to be factored for comprehensive decision making. The suggested next step would be to perform a financial impact of route changes, then determine the appropriate balance of cost savings and service based on analysis results and agency policy.

Scoring and Comparing Routes

Rochester: Increasing Employment Access through Increased Fixed-Route Frequency

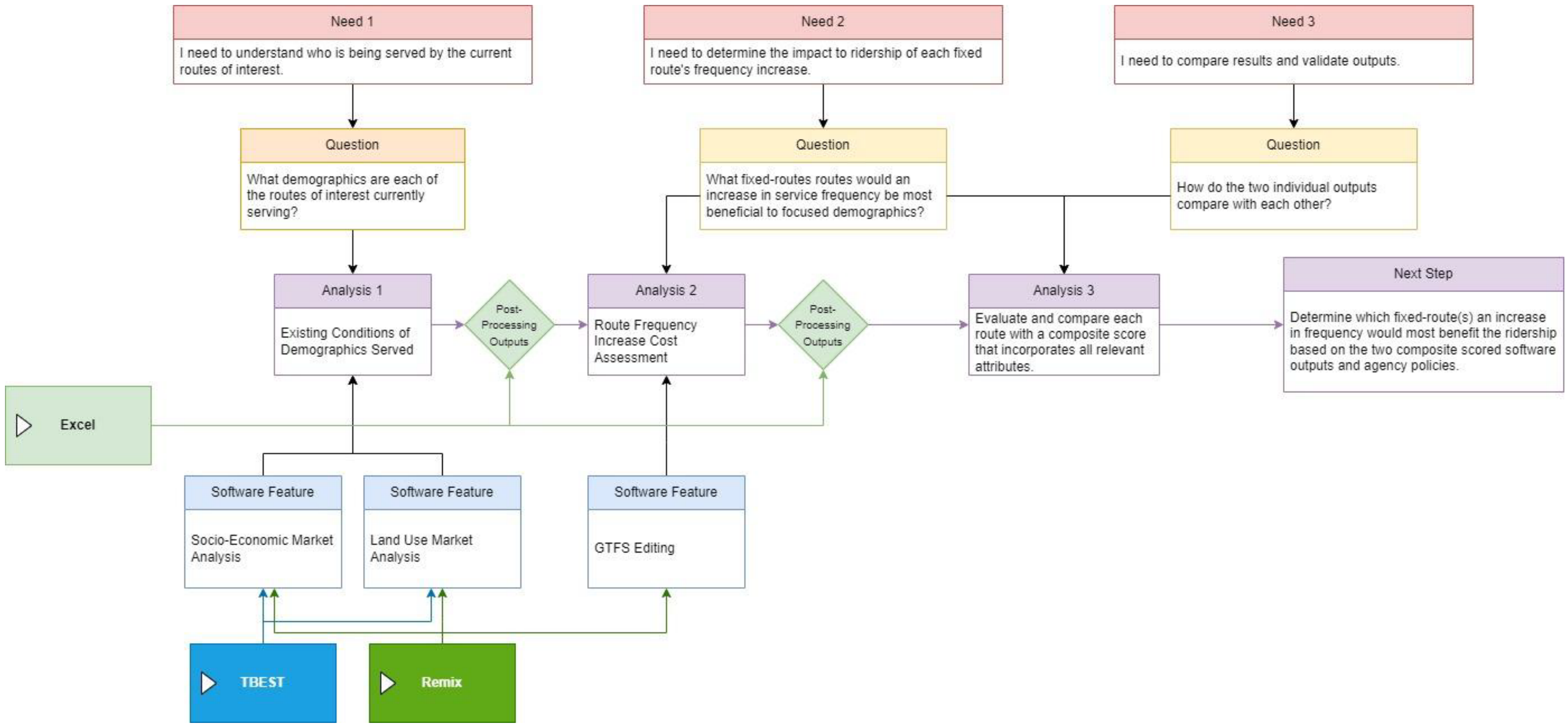
Link to Tutorial Appendix

Needs Met	Questions Answered	Analyses Performed
1. I need to understand who is being served by the current routes of interest. 2. I need to determine the impact to ridership of each fixed route's frequency increase. 3. I need to compare results and validate outputs.	1. What demographics are each of the routes of interest currently serving? 2. What fixed-routes routes would an increase in service frequency be most beneficial to focused demographics? 3. How do the two individual outputs compare with each other?	1. Existing Conditions of Demographics Served 2. Route Frequency Increase Cost Assessment 3. Evaluate and compare each route with a composite score that incorporates all relevant attributes.

Software Requirements	
Software	**Features Used**
TBEST	• *Socio-Economic Market Analysis* • *Land Use Market Analysis*
Remix	• *Socio-Economic Market Analysis* • *Land Use Market Analysis* • *GTFS Editing*
Excel	• *Post-Processing*

Case Study Product			
Title:	*Rochester: Increasing Employment Access through Increased Fixed-Route Frequency*	**Date:**	*MM/DD/YYYY*
Description This workflow is the result of a pilot project that sought to develop an impact score for ranking low-frequency, local routes based on ridership demographics, income, and employment to provide RTS with an evidence-based approach for adjusting the frequency of routes within their system.			

Flowchart - RTS Case Study



Workflow Narrative

Purpose

The purpose of this workflow is to construct an impact score for ranking routes based on ridership demographics, income, employment, and other attributes to provide planners with an evidence-based approach for adjusting route service within their system.

Problem/Need Identification

This workflow satisfies three needs;

- Need 1: to understand who is being served by the current routes of interest.
- Need 2: to determine the impact to ridership of each fixed route's frequency increase.
- Need 3: to compare results and validate outputs.

These needs require software that can support depicting transit routes and demographics.

Key Questions

Considering the needs, three primary questions can be formulated;

- Question 1: What demographics are each of the routes of interest currently serving?
- Question 2: What fixed-routes routes would an increase in service frequency be most beneficial to focused demographics?
- Question 3: How do the two individual outputs compare with each other?

Analysis Construction and Software Pairing

Analysis 1: Existing Conditions of Demographics Served

To understand the impact of changes, a baseline of the current existing conditions is necessary to compare against. Planners can identify who the current transit system routes using demographics data.

Analysis 2: Route Frequency Increase Cost Assessment

Changes to the system can be made via GTFS editing. This analysis adjusts the frequency of routes and calculates the increased cost of running specific routes more frequently.

Analysis 3: Evaluate and compare each route with a composite score that incorporates all relevant attributes.

With a baseline of served demographics constructed, analysts can determine what demographics they would like to prioritize service for–like zero-vehicle households or populations below the poverty threshold. All route attributes are normalized to be converted into comparable values, each multiplied by a weight value, then added together to develop a composite score.

Outputs and Interpretation

The outputs of this workflow are intended to be a spreadsheet for each system modification scenario. Each spreadsheet contains demographics metrics by route and can be sorted by composite score for comparison against each other.

Next Step

This workflow focuses entirely on the process of accessing demographics data and constructing a composite score to measure route performance. The suggested next step would be to determine if any supplementary attributes should be incorporated into a composite score. Finally, a determination should be made to find the balance between increasing route frequency that would most benefit the ridership and cost efficiency.

Aligning Service With Demand

Oswego: Service Alignment Study
Replica and STOPS

[Link to Tutorial Appendix]

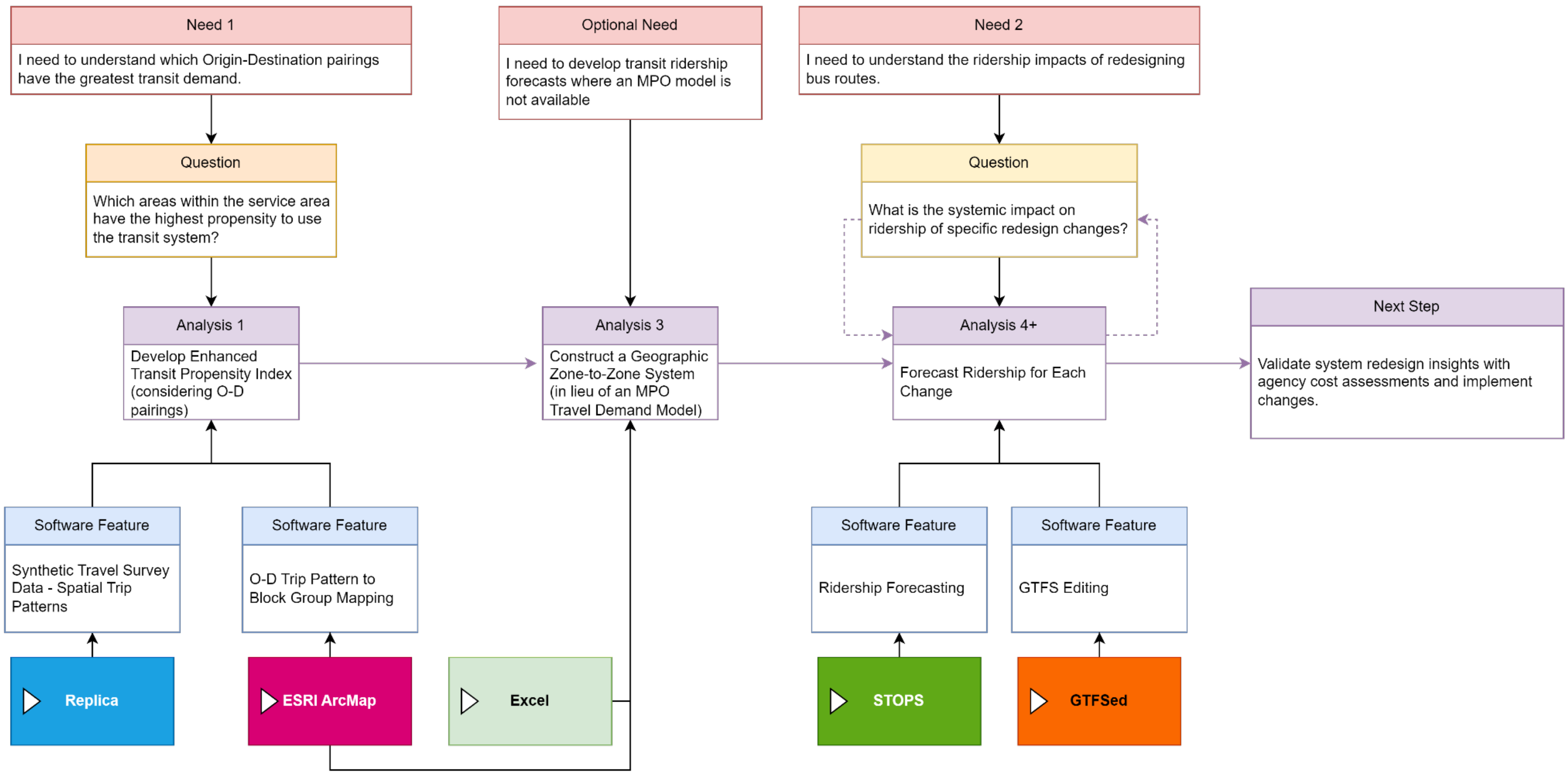
Needs Met	Questions Answered	Analyses Performed
1. I need to understand which Origin-Destination pairings have the greatest transit demand. 2. I need to develop transit ridership forecasts where an MPO model is not available. 3. I need to understand the ridership impacts of redesigning bus routes.	1. Which areas within the service area have the highest propensity to use the transit system? 2. What is the systemic impact on ridership of specific redesign changes?	1. Develop Enhanced Transit Propensity Index (considering O-D pairings) 2. Construct a Geographic Zone-to-Zone System (in lieu of an MPO Travel Demand Model) 3. Forecast Ridership for Each Change

Software Requirements	
Software	**Features Used**
Replica	• *Synthetic Origin-Destination data* • *Segmenting O-D data by transit propensity (using income and car ownership)* • *Export to CSV data tables*
Simplified Trips on Project Software (STOPS)	• *Ridership forecasting* • *Validation of GTFS coding of prospective transit network changes* • *Self-calibration to existing conditions ridership*
GTFSed	• *Visual confirmation of GTFS coding of prospective transit network changes*
Excel	• *Data processing/manipulation*
ESRI ArcMap	• *Visualization of ridership-forecasting results output from STOPS*

<table>
<tr><th colspan="4">Case Study Product</th></tr>
<tr><td>Title:</td><td>Oswego: Service Alignment Study</td><td>Date:</td><td>11/1/22</td></tr>
<tr><td colspan="4">Description

The objective of the Oswego case study was to develop and then test a prospective redesign of Oswego’s City Bus service by using two innovative travel demand analysis tools: Replica’s synthetic origin-destination datasets and the Federal Transit Administration’s Simplified Trips-on-Project Software (STOPS) open-access travel-demand modeling software package.</td></tr>
</table>

Flowchart - Oswego Case Study



Workflow Narrative

Purpose

The purpose of this workflow is to generate forecasts of the impact to transit ridership from redesign of a transit network

Problem/Need Identification

This workflow satisfies three needs;

- Need 1: to understand which origin-destination pairings in a region have the highest transit propensity
- Need 2: to generate estimates of transit-ridership impacts when an MPO travel demand model is not available
- Need 3: to understand the ridership impacts of redesigning bus routes.

These needs require software that can provide origin-destination data, can generate ridership impact forecasts, and can transfer outputs between multiple pieces of software seamlessly.

Key Questions

Considering the needs, two primary questions can be formulated;

- Question 1: Which areas within the service area have the highest propensity to use the transit system?
- Question 2: What is the systemic impact on ridership of specific redesign changes?

Analysis Construction and Software Pairing

Analysis 1: Develop Enhanced Transit Propensity Index (considering O-D pairings)

Replica data can be used to identify the origin-destination pairings that have high transit propensity, such as low-to-moderate income travelers and travelers who live in low-car-ownership households. This can be used to develop proposed realignment of transit routes to better align with demand.

Analysis 2: Construct a Geographic Zone-to-Zone System (in lieu of an MPO Travel Demand Model)

When a transit ridership forecast needs to be developed in a region that does not have a travel demand model, ESRI ArcMap can be used to develop a TAZ layer using Census geography, which can be imported into STOPS software which will generate the ridership forecasts.

Analysis 3: Forecast Ridership for Each Change

Using GTFSed (or any other GTFS editor), the proposed realignment can be coded into GTFS format, which can then be used by STOPS software to generate the ridership impacts.

Outputs and Interpretation

After using the Replica data to identify promising route alignments, and STOPS to quantify the ridership impacts, the analyst can import the ridership forecasts into GIS (e.g. ESRI's ArcMap or ArcGIS Pro) at the stop-level, to visualize and review the forecasts and identify which stop locations are projects to have the greatest impacts. The analyst may also wish to perform sensitivity analyses in which service frequencies, schedule, and/or alignments are varied.,

Next Step

Following generation of ridership impacts and any sensitivity analyses that are performed, the analyst can review the ridership impacts against the initial objectives of the service realignment to determine whether the impacts meet the desired objectives. The analyst can also use the ridership impacts to generate efficiency metrics (e.g. passengers per revenue mile or hour) that can assist the transit agency in deciding whether the efficiency gains meets agency standards. This will provide the agency's policy makers with information to support an informed decision of whether the proposed service realignment is justified and suitable for implementation. Finally, the data provided in this workflow can also be used in Title VI/Equity analyses.

Modeling Network Ridership

Capital District Transportation Authority: Troy Route Restructuring
TBEST, Remix, and ArcGIS

[Link to Tutorial Appendix]

Needs Met	Questions Answered	Analyses Performed
1. I need to reconfigure the transit routes to better serve the population. 2. I need to test the efficacy of the new system.	1. Where is the latent transit ridership? 2. How would ridership change based on the proposed route modifications?	1. Conduct a gap analysis to determine where there is a mismatch between transit supply and demand. 2. Conduct an origin-destination (O/D) analysis to understand prominent travel flows between Troy and other communities in the Capital District. 3. Create a baseline ridership model that accurately reflects existing ridership. 4. Develop modifications to existing service based on the gap and O/D analyses. 5. Create a new ridership model for the modified service and compare it to the baseline model.

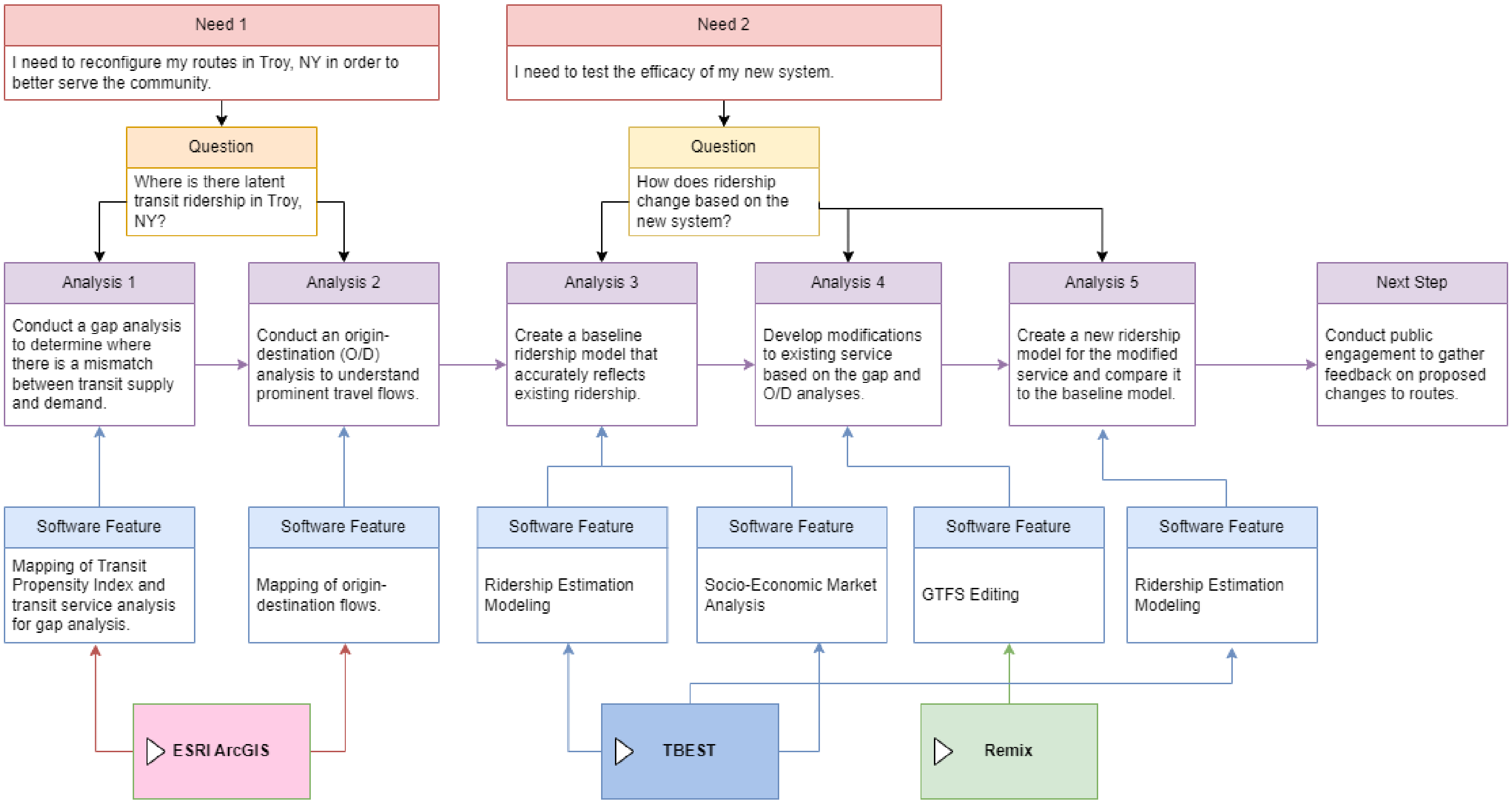
Software Requirements	
Software	**Features Used**
ArcGIS	• *Transit service gap analysis*
Remix	• *GTFS Editing*
TBEST	• *Ridership modeling and sensitivity analysis*

Case Study Product			
Title:	*CDTA Troy Transit Planning Pilot Study*	**Date:**	*MM/DD/YYYY*

Description

The objective of the CDTA Troy case study was to identify and assess transit planning software and strategies that CDTA could use in its service analysis, route restructuring and ridership projections. The pilot identifies gaps between transit supply and demand in the City of Troy, and introduces several changes to the transit network to assess the impacts of these changes on ridership using TBEST.

Flowchart - [CDTA Case Study]



Workflow Narrative

Purpose

The purpose of this workflow is to understand how the proposed changes in the transit network would impact ridership. The difference between the ridership outputs for the baseline model and the modified model will allow the analyst to understand the potential changes in ridership.

Problem/Need Identification

This workflow has two needs;

- Need 1: to reconfigure routes to better serve the community;
- Need 2: to test the efficacy of the modified routes.

Key Questions

Considering the needs, two questions can be formulated;

- Question 1: Where is there latent transit ridership in Troy, NY?
- Question 2: How does ridership change as a result of the route modifications?

Analysis Construction and Software Pairing

Analysis 1: Conduct a gap analysis to determine where there is a mismatch between transit supply and demand.

To meet Need 1 and Question 1, a gap analysis is used to analyze where there is a mismatch between the transit supply in an area and the transit demand. Several different factors are analyzed to determine the transit demand and the transit supply at the US Census Block Group level. This information is visualized using ArcGIS.

Analysis 2: Conduct an O/D analysis to understand prominent travel flows.

To meet Need 1 and Question 1, an O/D analysis is conducted using available US Census data to understand the prominent travel flows to and from the study area. This information is visualized using ArcGIS.

Analysis 3: Create a baseline ridership model that accurately reflects existing ridership.

To meet Need 2 and Question 2, the baseline ridership model needs to be developed and compared to existing ridership data. This will allow the analyst to assess how well the software aligns with the existing ridership, and make any adjustments to the baseline model's coefficients. While several software can be used to forecast ridership, this workflow uses TBEST. The existing transit network's GTFS data is loaded into TBEST to create the baseline model and the data is exported to Excel at the stop level.

Analysis 4: Develop modifications to existing service based on the gap and O/D analyses.

To meet Need 2 and Question 2, the analyst will need to review the results from Analysis 1 and Analysis 2 to determine where changes to the existing service should be made.

The analyst can use Remix or another GTFS editor to make the routing and schedule changes to the GTFS file, which can then be imported into TBEST.

Analysis 5: Create a new ridership model for the modified service and compare it to the baseline model.
To meet Need 2 and Question 2, the modified GTFS file is uploaded to TBEST and a new ridership model is created and exported to Excel at the stop level. The output of the new ridership model can then be compared to the baseline ridership model to determine potential changes in ridership.

Outputs and Interpretation
The output of the workflow would be stop level ridership for both the baseline and modified ridership models. As transit models are better at assigning ridership to the route level than at the stop level, it is recommended that both the baseline and the modified network models are compared at the route level to understand the potential impact that the modifications in service would have on ridership.

Next Step
The focus of this workflow is to test out potential impacts to transit ridership based on changes to the service and to better understand the capabilities of the TBEST software. Public engagement is a critical component of the planning process, and in any type of transit network redesign or transit service modifications, public engagement would be conducted throughout the entire redesign process. The suggested next step would be to conduct outreach to better understand areas where transit service is lacking or could be improved, and make additional transit service modifications before running the model again to understand the potential impact to ridership.

Appendix A: Technical Analyses

The following Technical Analyses are a step-by-step documentation of the methodologies used in each case study project. Transit planners can use these detailed analysis design documents to replicate the analyses, or to learn about applicable steps in the methodologies of each project.

Travel Time Impacts of Systemic Changes

Link to Workflow

A transit network ecosystem was established in the back end to set up an instance of Conveyal, which is addressed in detail in the "Software Assessment" section. Establishing a "transit network ecosystem" involves selecting and formatting all GTFS files that the software will attempt to transport a synthetic rider on. The following GTFS systems were packaged together to create the ecosystem:

- MTA New York City Transit
- Long Island Railroad
- Metro-North Railroad
- MTA Bus Company
- Suffolk County Transit
- Bee-Line Bus

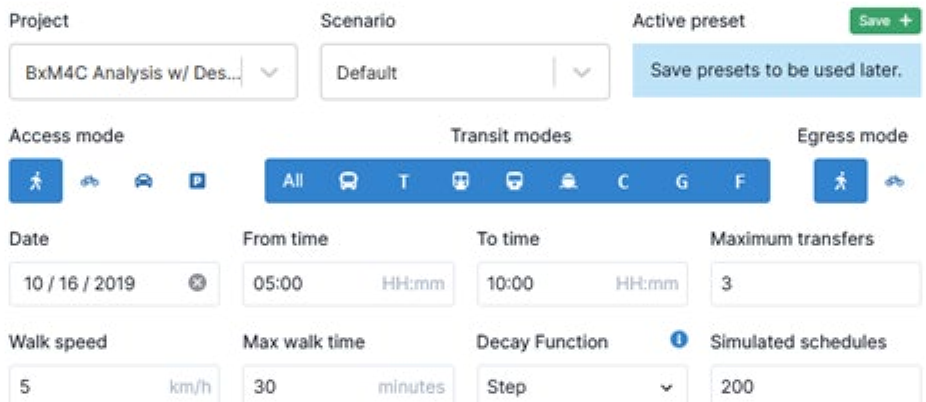
***Note**: Nassau County GTFS was initially included in the package but was removed due to a processing conflict.*

Next, the software required a list of locations formatted as a CSV file with three columns: stop_id (name of the location), stop_lat (location latitude), and stop_lon (location longitude). During the preliminary stages of the pilot project, the list of locations consisted solely of stops along the BxM4C, but as the project progressed, it was revised several times. A full list of the origins and destinations can be found in the appendix.

	A	B	C
1	stop_id	stop_lat	stop_lon
2	1941	40.78059	-73.9612
3	1942	40.76435	-73.9735
4	1943	40.75927	-73.9772
5	1944	40.75398	-73.9806
6	1945	40.7496	-73.9843
7	1946	40.74334	-73.9884
8	1947	40.74128	-73.9891

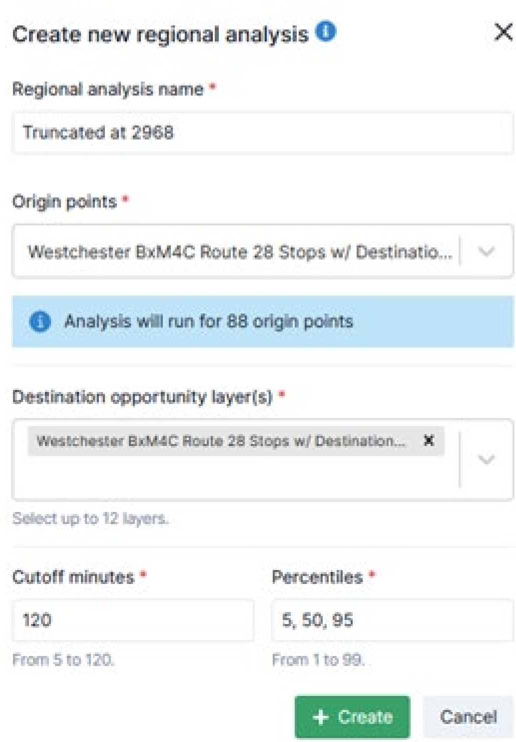
After the transit network ecosystem is established and the final list of relevant locations is selected, parameters need to be defined to determine the behavior and priorities.

The Conveyal software provides a user interface for defining parameters. This image depicts the final selection of parameters that were modeled.



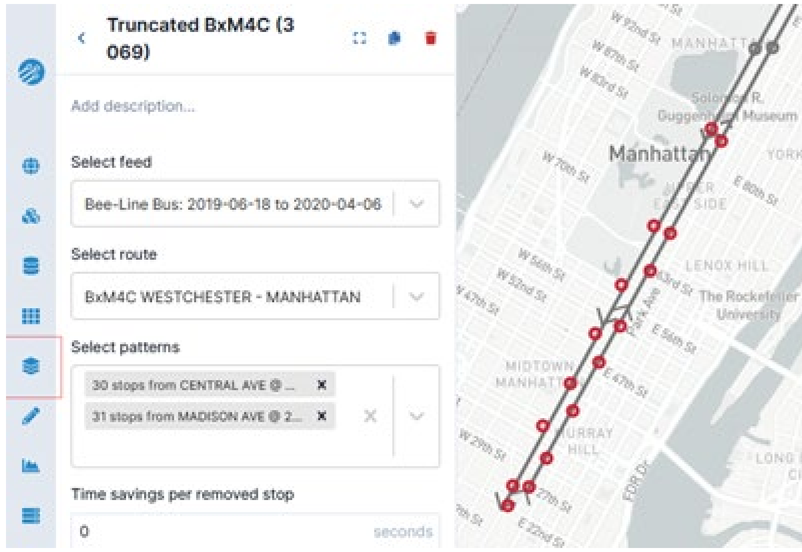
Synthetic riders were determined to access and egress all types of available transit by walking. Two-hundred simulated schedules were generated to take place between 5:00am and 10:00am. Synthetic riders were defined as having a walk speed of 5 km/h and only willing to walk for 30 minutes and make 3 transfers at most.

The final set of parameters are defined when the analysis is converted into a regional analysis. At this phase, the analyst defined both the origin and destination points as the same set of points. The intent was to create a matrix assessing travel time from each origin point to each destination point. Cutoff Minutes, or the total duration that a synthetic trip could take before being determined invalid, was set to 120 minutes.



Finally, the percentiles Conveyal was to output were determined. If all the travel times from every trip possible within all other parameters were mapped on a bar graph, the graph would have a normal distribution. The 5^{th} percentile indicates the travel times under the top 5% performing circumstances (i.e., as though a synthetic rider arrived at the bus stop just as the bus was arriving at the stop). The 50^{th} percentile shows the average circumstances. Finally, the 95^{th} percentile shows the worst-case scenario—the synthetic rider sees the bus they intended to catch drive away as they arrive at the stop and now must wait for the next bus.

Using the outlined parameters, outputs were generated for both existing conditions and a scenario in which the BxM4C was completely removed from service. Additionally, Conveyal supports the use of scenario modifications that allow the user to adjust a transit system and perform analyses on the adjusted scenario. The screenshot below depicts the “Remove Stops” modification in which all the red-highlighted stops were removed from the line.



The Research Team performed multiple variations of this modification, each truncating the BxM4C at different stops. After each modification, the network analysis was run, the travel time CSV output was downloaded, and the data was converted into one or more matrices following the indexing process outlined further in this section. The following stops were selected for truncation analyses:

- 3069 - 5TH AVE @ 98TH ST
- 1941 - 5TH AVE @ 85TH ST
- 2968 - 5TH AVE @ 69TH ST
- 1942 - 5TH AVE @ 59TH ST
- 1943 - 5TH AVE @ 51ST ST

Conveyal provides five potential outputs from an analysis: GeoTIFF, Scenario and Modification JSON, Paths CSV, Times CSV, Access CSV. The majority of our analyses utilized Times CSV outputs like the example below:

origin	destination	percentile	time
1941	1942	5	11
1941	1943	5	13
1941	1944	5	15
1941	1945	5	17
1941	1946	5	19

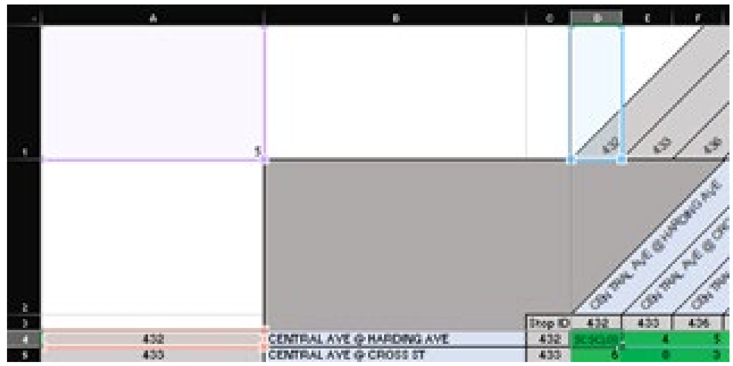
These spreadsheets depict the travel time from an origin to a destination for the selected percentile. The first row indicates that if a rider is within the 5th percentile of trip travel time, then going from stop 1941 to 1942 will be 11 minutes. The multiple scenarios and modifications outlined in the Analysis Design section were exported, combined, and compared to each other:

origin	destination	percentile	BxM4C time	no BxM4C	difference	diff values
1941	1942	5	8	11	3	3
1941	1943	5	10	13	3	3
1941	1944	5	12	15	3	3
1941	1945	5	15	17	2	2

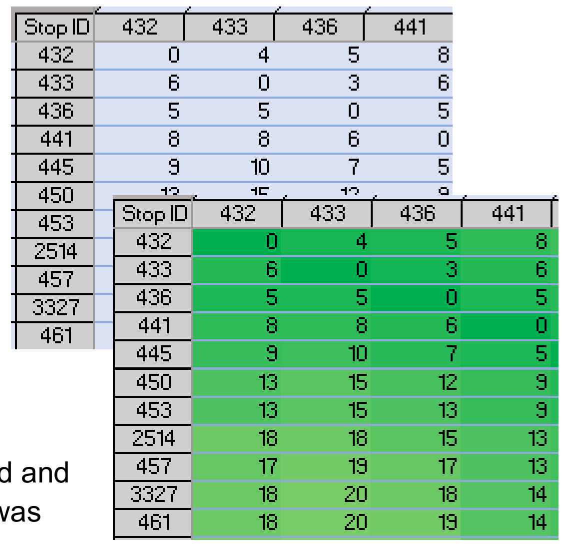
The first row in this comparison indicates that, in the fifth percentile, it takes 8 minutes to travel from stop 1941 to 1942 with the BxM4C and 11 minutes if the route were removed from service. This results in an increase of 3 minutes of travel time as shown in the "difference" column. ***Note****: Since the contents of the "diff" column were calculated using an Excel equation (=[@[no bxm4c]]-[@[bxm time]]), they could not be indexed correctly into the matrices—so they were duplicated as "values only" into a new column—"diff values".*

The final outputs of the team's analyses were a series of matrices. These matrices were initially constructed manually, but as it became an iterative process, a formula was developed so the analyst could make use of a matrix template with the equations. The equation uses the 'INDEX' function to copy information from the Conveyal output and the 'MATCH' function to identify which information to index. The colors in the full equation shown below are coordinated with their respective cells in the table screenshot below.

=INDEX('Conveyal Output'!$D:$D, MATCH(1, ('SB, 5% 5-10AM, TT BxM4C'!D$1='Conveyal Output'!$B:$B)*('SB, 5% 5-10AM, TT BxM4C'!$A4='Conveyal Output'!$A:$A)('SB, 5% 5-10AM, TT BxM4C'!$A$1='Conveyal Output'!$C:$C),0)*



Each time travel cell in the matrix contains the equation above that matches the associated origin (light orange), destination (blue), and percentile (purple) with the corresponding columns on the Conveyal output CSV. The green highlighted cell can be dragged across the spreadsheet to duplicate the equation for each destination, then the entire row is highlighted and dragged down to perform the index/match with all origin/destination pairs on the sheet. This method was faster and easier than manually creating the matrices but still requires a substantial amount of uninterrupted processing time—approximately 7 to 13 minutes.



Stop ID	432	433	436	441
432	0	4	5	8
433	6	0	3	6
436	5	5	0	5
441	8	8	6	0
445	9	10	7	5
450				
453				
2514				
457				
3327				
461				

Stop ID	432	433	436	441
432	0	4	5	8
433	6	0	3	6
436	5	5	0	5
441	8	8	6	0
445	9	10	7	5
450	13	15	12	9
453	13	15	13	9
2514	18	18	15	13
457	17	19	17	13
3327	18	20	18	14
461	18	20	19	14

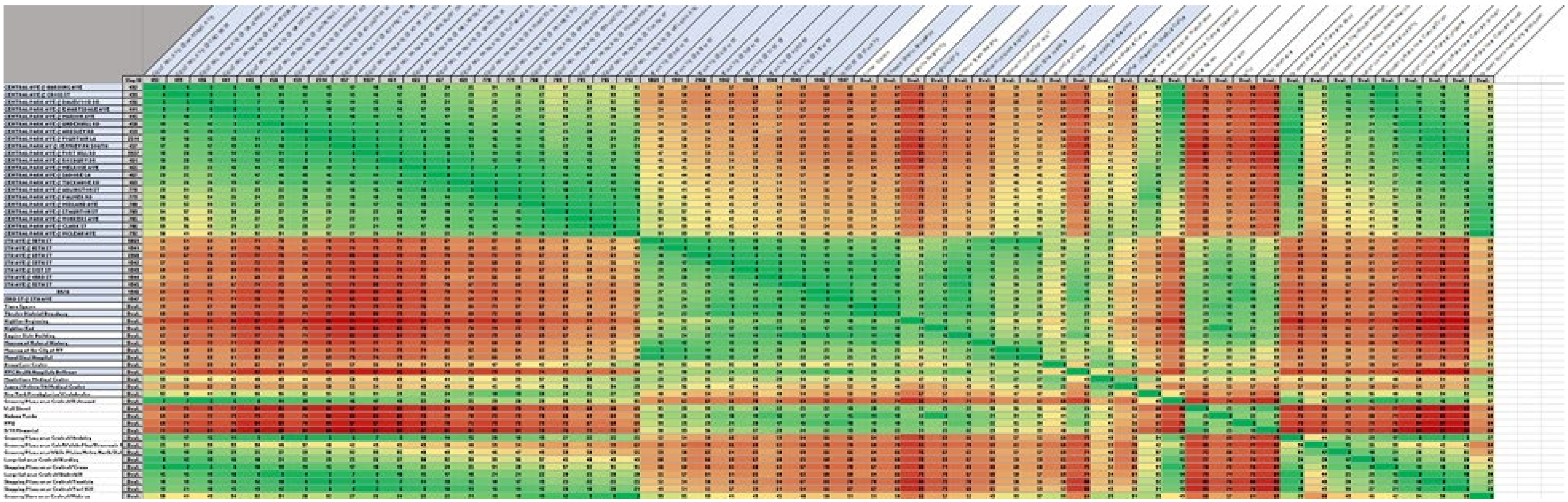
Once the INDEX function has completed processing for all cells, the matrix fills in the values from the output data corresponding with each origin/destination/percentile combination.

To facilitate analysis interpretation, the data was highlighted and "Conditional Formatting" of "Red – Yellow – Green Scale" was selected. This made all 0 travel time valued cells green, the highest travel time value cell red, and then range between scaling yellow to orange from smallest to largest. The result is a color-coded origin-destination matrix showing travel times from each origin to each destination under the analyses' conditions.

For the first scenario, 9 total matrices were constructed—the 5th, 50th, and 95th percentile each had a matrix with three versions: All Stops, Northbound Stops Only, and Southbound Stops Only. During analysis interpretation, the Research Team agreed to concentrate their efforts to focus on Southbound Stops Only at the 5th percentile.

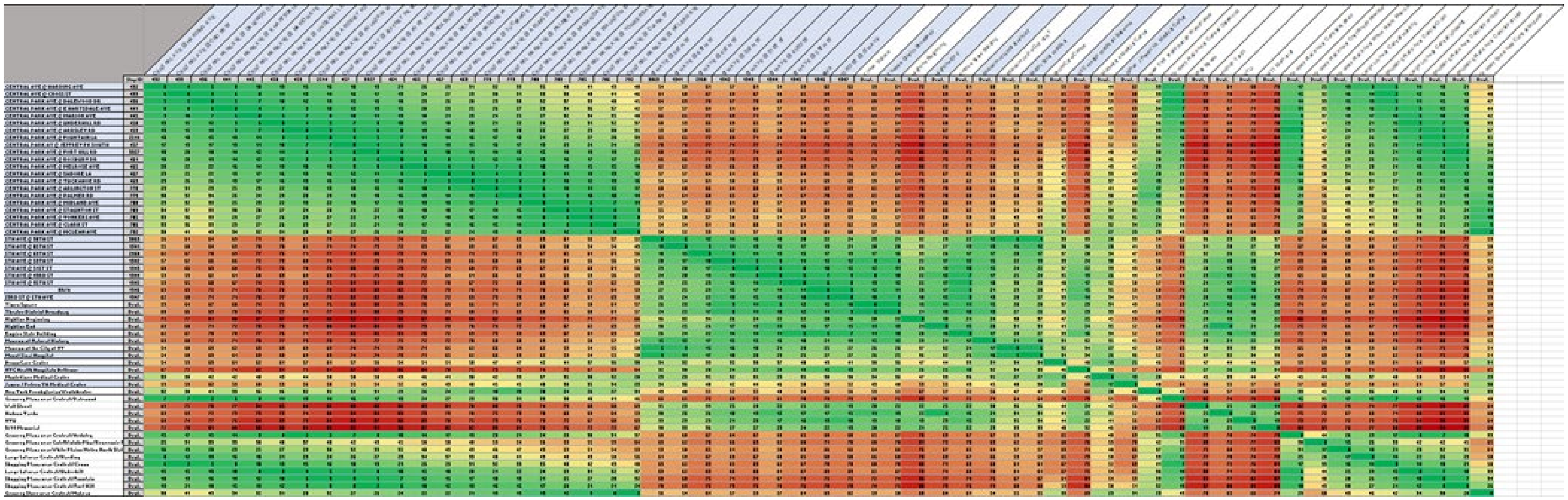
Travel Time – Southbound, 5th Percentile, BxM4C (Existing Conditions)

This matrix shows the travel time from and to all southbound stops and destinations for the 5th percentile with the BxM4C as it is. Conveyal calculates the fastest way to get from an origin to a destination within the entire transit network ecosystem. The “Stop ID” column and row are both listed in route beginning-to-end sequential order. The darkest green values are “0” travel time (i.e., same origin as destination). A diagonal line of these non-travel values can be seen through the spreadsheet. Since this portion of the route is one-directional and linear, a rider cannot arrive at a destination before their origin. Subsequently, all values below the “0” line can be disregarded. All values are colored on a gradient scale of green to red indicating lowest to highest travel times.



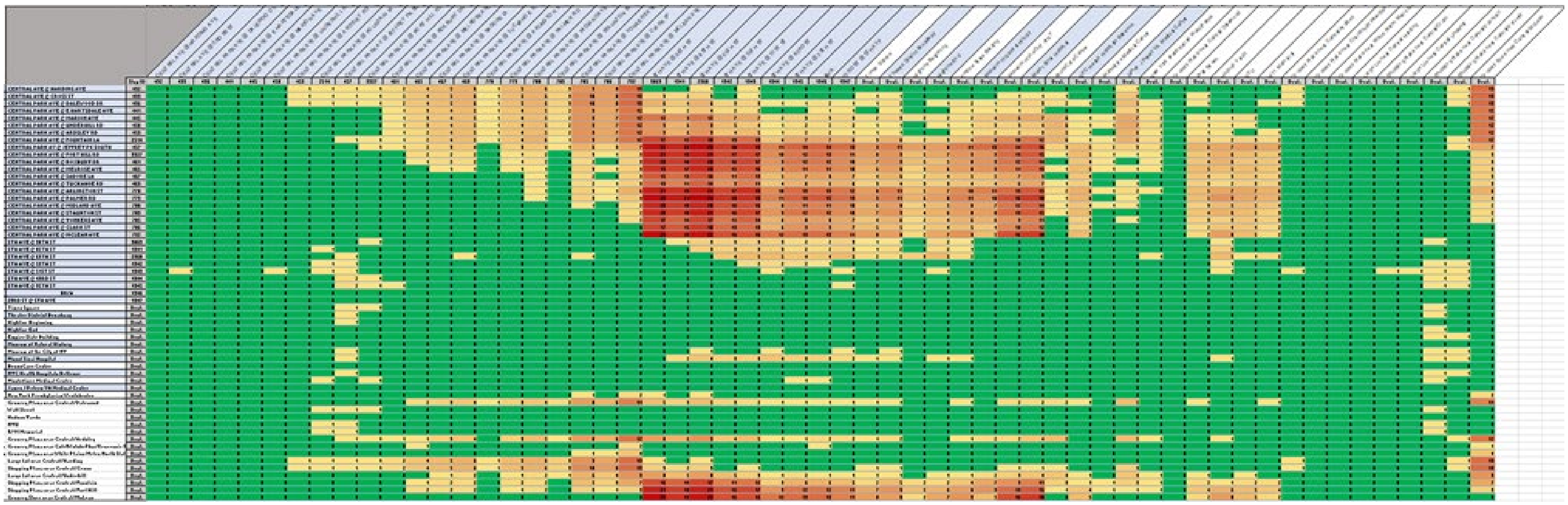
Travel Time – Southbound, 5th Percentile, No BxM4C

This matrix shows the travel time from and to all southbound stops and destinations for the 5th percentile if the BxM4C were completely removed from service. At a glance, these two matrices are very similar, so a new type of matrix was constructed to highlight the differences in the scenarios (next page).



Difference of Travel Time Between BxM4C and No BxM4C – Southbound, 5th Percentile

Generating the difference matrix used an excel formula where the travel time value of the “BxM4C” scenario matrix cell was subtracted from the same cell location on the “No BxM4C” scenario matrix (*='SB, 5% 5-10AM, TT NoBxM4C'!D4-'SB, 5% 5-10AM, TT BxM4C'!D4*). Like the previous matrices, green is “0” and the highest value is red. Since Conveyal calculates the fastest way to get from an origin to a destination within the entire transit network ecosystem, if the BxM4C is removed and there is no impact or another route could get the rider from their origin to the same destination with no time difference, then the time travel value will be a “0” in this matrix. Within this difference matrix, a very clear problem area emerges: Origin stops 457-792 traveling to destination stops 3069, 1941, and 2968.



O-D Pair Difference Table

Using two outputs from Conveyal, the TIMES and PATHS .CSV files, the Research Team constructed a table (using the same INDEX/MATCH methodology as in the matrices) to sort and compare BxM4C to No-BxM4C outputs. The “Counts” related columns detail how many different pathways a rider could take to make it from that origin to that destination. Count values were generated with the equation “=COUNTIFS('Paths Output BxM4C (Values)'!$A:$A,A2,'Paths Output BxM4C (Values)'!$B:$B,B2)”, with A2 being the origin and B2 being the destination. The equation scans through the PATHS .CSV output file and returns the number of times that the specified O-D pair is found together. The “Transfer” related columns calculate how many transfers occur in **total across all counts**. Calculating number of transfers requires a new column in the PATHS output with the equation “=LEN([@routes])-LEN(SUBSTITUTE([@routes],"|",""))”. Since Conveyal distinguishes when a transfer occurs using the “|” symbol, this equation counts the occurrences of that symbol within the “routes” column and returns a numerical value of the times that the symbol appeared. Then, in the O-D Difference Table, the Transfers column uses “=SUMIFS(Table7[Number of Transfers], Table7[origin], [@origin],Table7[destination], [@destination])” to search the PATHS file for the O-D combination and then sum together all values returned by the former equation (aggregating the transfers for all counts).

Origin Stop	Origin Desc	Destination Stop	Destination Desc	Time W	Time W/O	W-W/O Dif Time	Counts W	Counts W/O	Dif Counts	Transfers W	Transfers W/O	Transfers Diff	Avg Transfers W	Avg Transfers W/O	Avg Transfers Dif
779	CENTRAL PARK AVE @ PALMER RD	2968	5TH AVE @ 69TH ST	44	68	24	13	16	3	24	33	9	1.846154	2.0625	0.2163462
457	CENTRAL PARK AV @ JEFFREY PK SOUTH	2968	5TH AVE @ 69TH ST	54	77	23	15	18	3	35	44	9	2.333333	2.4444444	0.1111111
779	CENTRAL PARK AVE @ PALMER RD	3069	5TH AVE @ 98TH ST	38	61	23	10	17	7	18	35	17	1.8	2.0588235	0.2588235
792	CENTRAL PARK AVE @ MCLEAN AVE	3069	5TH AVE @ 98TH ST	31	54	23	11	29	18	24	70	46	2.181818	2.4137931	0.2319749
Grocery Store	Grocery Store near Central/McLean	3069	5TH AVE @ 98TH ST	33	56	23	11	30	19	24	72	48	2.181818	2.4	0.2181818
457	CENTRAL PARK AV @ JEFFREY PK SOUTH	3069	5TH AVE @ 98TH ST	48	70	22	13	20	7	30	49	19	2.307692	2.45	0.1423077
792	CENTRAL PARK AVE @ MCLEAN AVE	2968	5TH AVE @ 69TH ST	37	59	22	12	20	8	24	42	18	2	2.1	0.1
Grocery Store	Grocery Store near Central/McLean	2968	5TH AVE @ 69TH ST	39	61	22	12	22	10	24	46	22	2	2.0909091	0.0909091
778	CENTRAL PARK AVE @ ARLINGTON ST	2968	5TH AVE @ 69TH ST	45	66	21	10	16	6	17	33	16	1.7	2.0625	0.3625
778	CENTRAL PARK AVE @ ARLINGTON ST	3069	5TH AVE @ 98TH ST	39	60	21	8	14	6	13	27	14	1.625	1.9285714	0.3035714

Truncation Matrices

This matrix shows the difference of travel times in minutes between two scenarios; one where BxM4C is the full route and one where the route is truncated at stop 3069. The next page contains four other matrices comparing different truncation scenarios to the full route (truncation at 1941 [top-left], 2968 [top-right], 1942 [bottom-left], and 1943 [bottom-right]). These five matrices combined constitute a sensitivity analysis that allowed the Research Team to identify at which stop truncating the BxM4C will result in diminishing returns of travel time savings.

		5TH AVE @ 98TH ST	5TH AVE @ 85TH ST	5TH AVE @ 69TH ST	5TH AVE @ 59TH ST	5TH AVE @ 51ST ST	5TH AVE @ 43RD ST	5TH AVE @ 35TH ST	#N/A	23RD ST @ 5TH AVE	Times Square	Theatre District Broadway	Highline Beginning	Highline End	Empire State Building	Museum
	Stop ID	3069	1941	2968	1942	1943	1944	1945	1946	1947	Dest.	Dest.	Dest.	Dest.	Dest.	Dest.
CENTRAL AVE @ HARDING AVE	432	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
CENTRAL AVE @ CROSS ST	433	0	1	3	0	1	0	0	0	0	0	0	0	0	0	0
CENTRAL PARK AVE @ DALEWOOD DR	436	0	3	5	2	2	0	1	1	0	0	1	0	0	1	0
CENTRAL PARK AVE @ E HARTSDALE AVE	441	0	2	4	2	2	0	0	1	1	0	1	0	0	1	0
CENTRAL PARK AVE @ MARION AVE	445	0	6	9	6	6	2	3	3	3	2	2	1	2	2	0
CENTRAL PARK AVE @ UNDERHILL RD	450	0	4	7	4	4	1	2	2	2	1	1	0	1	1	0
CENTRAL PARK AVE @ ARDSLEY RD	453	0	5	8	4	5	1	2	2	1	1	1	0	0	1	0
CENTRAL PARK AVE @ FOUNTAIN LA	2514	0	8	12	11	12	4	6	7	6	3	5	1	2	4	0
CENTRAL PARK AV @ JEFFREY PK SOUTH	457	0	8	12	12	13	9	11	11	10	7	8	2	6	8	0
CENTRAL PARK AVE @ FORT HILL RD	3327	0	8	12	12	13	8	10	10	9	6	7	1	5	7	0
CENTRAL PARK AVE @ ROXBURY DR	461	0	8	12	11	13	8	10	10	8	6	7	1	4	7	0
CENTRAL PARK AVE @ MELROSE AVE	465	0	8	12	11	12	7	9	9	8	5	6	1	4	6	0
CENTRAL PARK AVE @ SADORE LA	467	0	8	11	9	10	5	6	7	6	4	4	1	3	4	0
CENTRAL PARK AVE @ TUCKAHOE RD	469	0	7	10	7	9	4	5	5	5	3	3	1	3	3	0
CENTRAL PARK AVE @ ARLINGTON ST	778	0	8	12	12	13	8	10	10	8	6	7	1	5	7	0
CENTRAL PARK AVE @ PALMER RD	779	0	8	12	12	13	10	12	11	10	8	8	1	6	8	0
CENTRAL PARK AVE @ MIDLAND AVE	780	0	8	12	11	12	7	10	10	8	5	6	1	4	6	0
CENTRAL PARK AVE @ STAUNTON ST	783	0	8	12	12	13	8	10	10	9	6	7	1	5	7	0
CENTRAL PARK AVE @ YONKERS AVE	785	0	8	12	10	11	5	7	7	6	4	5	0	3	5	0
CENTRAL PARK AVE @ CLARK ST	786	0	8	12	10	11	5	7	7	6	4	4	1	3	4	0
CENTRAL PARK AVE @ MCLEAN AVE	792	0	8	12	12	13	10	11	11	9	7	6	1	6	7	0
5TH AVE @ 98TH ST	3069	0	1	3	3	3	3	2	2	3	1	1	0	2	1	0
5TH AVE @ 85TH ST	1941	0	0	3	3	3	3	2	2	3	1	2	0	2	1	0
5TH AVE @ 69TH ST	2968	0	0	0	4	5	5	4	4	5	2	2	3	5	2	0
5TH AVE @ 59TH ST	1942	0	0	0	0	2	2	2	1	0	0	0	0	0	0	0
5TH AVE @ 51ST ST	1943	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0
5TH AVE @ 43RD ST	1944	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5TH AVE @ 35TH ST	1945	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0

Stop ID	3069	1941	2968	1942	1943	1944	1945	1946	1947	Dest.	Dest.	Dest.	Dest.	Dest.	Dest.
432	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
433	0	0	3	0	1	0	0	0	0	0	0	0	0	0	0
436	0	0	5	2	2	0	1	1	0	0	1	0	0	1	0
441	0	0	4	2	2	0	0	1	1	0	1	0	0	1	0
445	0	0	9	6	6	2	3	3	2	2	2	1	2	2	0
450	0	0	7	4	4	1	2	2	2	1	1	0	1	1	0
453	0	0	8	4	5	1	2	2	1	1	1	0	0	1	0
2514	0	0	11	11	12	4	6	7	5	3	5	1	2	4	0
457	0	0	11	12	13	9	11	11	8	6	8	2	5	8	0
3327	0	0	11	11	13	7	10	10	8	5	7	1	4	7	0
461	0	0	11	11	13	7	10	10	7	5	7	1	3	7	0
465	0	0	11	11	12	6	9	8	6	4	6	1	3	6	0
467	0	0	10	9	10	4	6	6	4	3	4	1	2	4	0
469	0	0	9	7	9	3	5	5	4	3	3	1	2	3	0
778	0	0	11	11	13	7	10	9	7	5	7	1	3	7	0
779	0	0	11	12	13	9	12	11	9	7	8	1	5	8	0
780	0	0	11	11	12	7	10	9	7	5	6	1	3	6	0
783	0	0	11	12	13	7	10	10	8	5	7	1	4	7	0
785	0	0	11	10	11	5	7	6	5	3	5	0	2	5	0
786	0	0	11	10	11	4	7	7	6	4	4	1	2	4	0
792	0	0	11	12	13	10	11	11	8	6	6	1	5	7	0
3069	0	0	3	3	3	3	2	2	3	1	1	0	2	1	0
1941	0	0	3	3	3	3	2	2	3	1	2	0	2	1	0
2968	0	0	0	4	5	5	4	4	5	2	2	3	5	2	0
1942	0	0	0	0	2	2	2	1	0	0	0	0	0	0	0
1943	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0
1944	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1945	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0

Stop ID	3069	1941	2968	1942	1943	1944	1945	1946	1947	Dest.	Dest.	Dest.	Dest.	Dest.	Dest.
432	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
433	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
436	0	0	0	2	2	0	1	1	0	0	1	0	0	1	0
441	0	0	0	2	2	0	0	1	1	0	1	0	0	1	0
445	0	0	0	5	6	2	3	3	2	2	2	1	2	2	0
450	0	0	0	3	4	1	2	2	2	1	1	0	1	1	0
453	0	0	0	3	5	1	2	2	1	1	1	0	0	1	0
2514	0	0	0	9	12	4	6	7	5	3	5	1	2	4	0
457	0	0	0	9	13	9	11	11	8	6	8	2	5	8	0
3327	0	0	0	9	13	7	10	10	8	5	7	1	4	7	0
461	0	0	0	9	13	7	10	10	7	5	6	1	3	7	0
465	0	0	0	8	12	6	9	8	6	4	6	1	3	6	0
467	0	0	0	7	10	4	6	6	4	3	4	1	2	4	0
469	0	0	0	6	9	3	5	5	4	3	3	1	2	3	0
778	0	0	0	9	13	7	10	9	7	5	7	1	3	7	0
779	0	0	0	9	13	9	12	11	9	7	7	1	5	8	0
780	0	0	0	9	12	7	10	9	7	5	6	1	3	6	0
783	0	0	0	9	13	7	10	10	8	5	6	1	4	7	0
785	0	0	0	8	11	5	7	6	5	3	5	0	2	5	0
786	0	0	0	8	11	4	7	7	6	4	4	1	2	4	0
792	0	0	0	9	13	10	11	11	8	6	6	1	5	7	0
3069	0	0	0	3	3	3	2	2	3	1	1	0	2	1	0
1941	0	0	0	3	3	3	2	2	3	1	2	0	2	1	0
2968	0	0	0	4	5	5	4	4	5	2	2	3	5	2	0
1942	0	0	0	0	2	2	2	1	0	0	0	0	0	0	0
1943	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0
1944	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1945	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0

Stop ID	3069	1941	2968	1942	1943	1944	1945	1946	1947	Dest.	Dest.	Dest.	Dest.	Dest.	Dest.
432	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
433	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
436	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0
441	0	0	0	0	1	0	0	1	1	0	0	0	0	1	0
445	0	0	0	0	3	1	1	2	1	0	0	0	2	1	0
450	0	0	0	0	2	1	1	0	1	0	0	0	1	0	0
453	0	0	0	0	3	1	1	1	0	0	0	0	0	0	0
2514	0	0	0	0	4	3	3	3	4	0	0	0	2	1	0
457	0	0	0	0	4	5	4	4	4	0	0	1	4	2	0
3327	0	0	0	0	4	4	4	4	4	0	0	0	4	2	0
461	0	0	0	0	4	4	4	4	4	0	0	0	3	2	0
465	0	0	0	0	4	3	4	4	4	0	0	0	3	1	0
467	0	0	0	0	4	2	2	3	3	0	0	0	2	1	0
469	0	0	0	0	4	1	2	2	2	0	0	0	2	0	0
778	0	0	0	0	4	4	4	4	4	0	0	0	3	2	0
779	0	0	0	0	4	5	4	4	4	0	0	0	4	2	0
780	0	0	0	0	4	4	4	4	4	0	0	0	3	2	0
783	0	0	0	0	4	4	4	4	4	0	0	0	3	2	0
785	0	0	0	0	4	3	3	3	3	0	0	0	2	2	0
786	0	0	0	0	4	3	3	3	3	0	0	0	2	1	0
792	0	0	0	0	4	5	4	4	4	0	0	0	4	2	0
3069	0	0	0	0	3	3	2	2	3	0	0	0	1	1	0
1941	0	0	0	0	3	3	2	2	3	0	0	0	1	1	0
2968	0	0	0	0	4	5	4	4	4	0	0	2	3	1	0
1942	0	0	0	0	2	2	2	1	0	0	0	0	0	0	0
1943	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0
1944	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1945	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0

Stop ID	3069	1941	2968	1942	1943	1944	1945	1946	1947	Dest.	Dest.	Dest.	Dest.	Dest.	Dest.
432	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
433	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
436	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0
441	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0
445	0	0	0	0	0	1	1	2	1	0	0	0	2	1	0
450	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0
453	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0
2514	0	0	0	0	0	3	3	3	4	0	0	0	2	1	0
457	0	0	0	0	0	5	4	4	4	0	0	0	4	2	0
3327	0	0	0	0	0	4	4	4	4	0	0	0	4	2	0
461	0	0	0	0	0	4	4	4	4	0	0	0	3	2	0
465	0	0	0	0	0	3	4	4	4	0	0	0	3	1	0
467	0	0	0	0	0	2	2	3	3	0	0	0	2	1	0
469	0	0	0	0	0	1	2	2	2	0	0	0	2	0	0
778	0	0	0	0	0	4	4	4	4	0	0	0	3	2	0
779	0	0	0	0	0	5	4	4	4	0	0	0	4	2	0
780	0	0	0	0	0	4	4	4	4	0	0	0	3	2	0
783	0	0	0	0	0	4	4	4	4	0	0	0	3	2	0
785	0	0	0	0	0	3	3	3	3	0	0	0	2	2	0
786	0	0	0	0	0	3	3	3	3	0	0	0	2	1	0
792	0	0	0	0	0	5	4	4	4	0	0	0	4	2	0
3069	0	0	0	0	0	3	2	2	3	0	0	0	1	1	0
1941	0	0	0	0	0	3	2	2	3	0	0	0	1	1	0
2968	0	0	0	0	0	5	4	4	4	0	0	1	3	1	0
1942	0	0	0	0	0	2	2	1	0	0	0	0	0	0	0
1943	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0
1944	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1945	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0

Scoring and Comparing Routes

Link to Workflow

Introduction

The purpose of this tutorial is to provide instructions on how to create a composite scoring spreadsheet to compare multiple routes in a transit system using the data available in Remix. A composite score is a single variable that represents a combination of information from multiple variables. Creating a composite score for each route in the user's transit system will allow them to rank the routes based on information the user deems to be most valuable.

This tutorial will teach the user how to use Remix to add routes for comparison and then download the standard socio-economic data from Remix for post processing in Excel. Data will be converted from percentages to values and then normalized and ranked. To analyze the outputs, a sheet to summarize the data will be constructed allowing the user to weight each variable within the composite score calculation and automatically update the route rankings. At the end of this tutorial, the user will know how to use Remix to access data for multiple transit routes, post-process the outputs, create composite scores, and produce a spreadsheet suitable for use as an analysis tool.

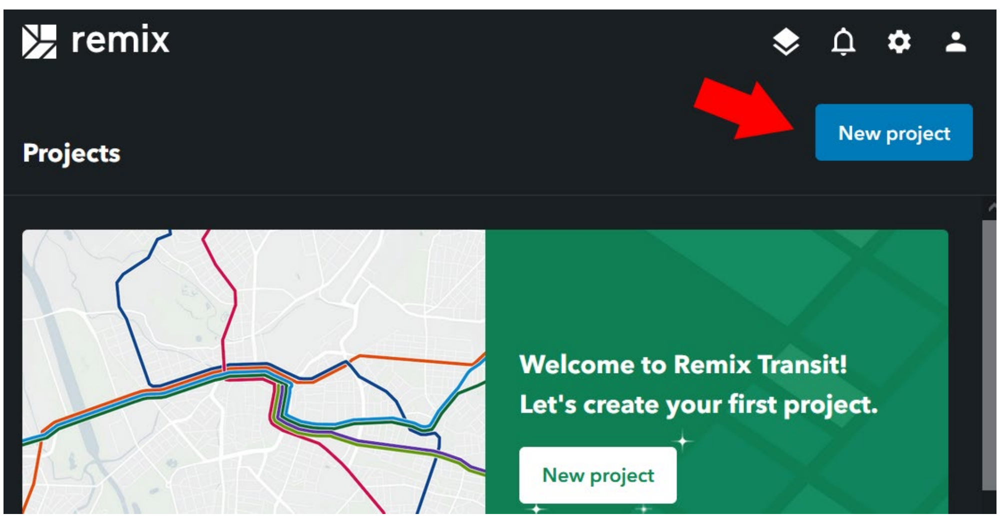


Figure 1 - Remix New Project Button

Navigating Remix

Once logged into Remix, the user will be taken to the **Projects** screen. All current projects, containing individual or groups of transit routes, are found here—an existing project can be selected, or a new project can be created here.

For this tutorial, we will create a new project that will contain all routes in the transit system. Click on the blue **New project** button towards the top center of the screen. Name your new project, provide an optional description, and click **Create**.

The user will be taken to the **Remix Explorer** screen which provides the user with a graphical view of their region, transit lines, and various data layers. From this screen, we will be adding all routes in the transit system, selecting a quarter-mile radius around each stop, and downloading the socio-economic data to work with.

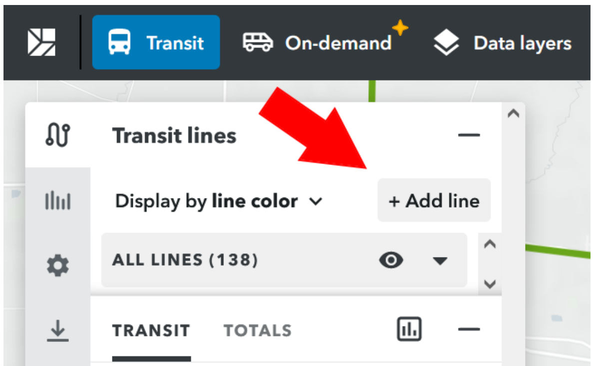


Figure 2 - Remix Add Line Button

Start by clicking on the **+ Add line** button. You will see a list of all lines in your transit system. If the list is collapsed, click on the black arrow to expand the list of routes. Click on the three dots next to the black arrow and select **Import all lines.** Once done, the lines will begin to show up on the map with each route in the list having a blue checkmark displayed next to it. Click the arrow next to **Add Lines** to go back to the transit lines screen.

At the transit lines screen, specify the desired mile radius around each stop to collect data from by clicking on the number displayed. For this tutorial we will use the default (0.25 mi). Once the value is set, click on the eye icon, the last icon next to it, to display the transparent radii buffers around each stop on the map. *Note: The buffers may not be visible when zoomed out.*

Now that the routes and radii around each stop have been selected, the data is ready to be downloaded. Click on the **Exports** icon to open the exports tab and under **Export Map** Data, select **Line attributes (Excel)** to begin the download of the data. This will download the spreadsheet that will be post-processed and analyzed. Take note of the folder where the spreadsheet was downloaded to. Navigate to that folder and open the spreadsheet.

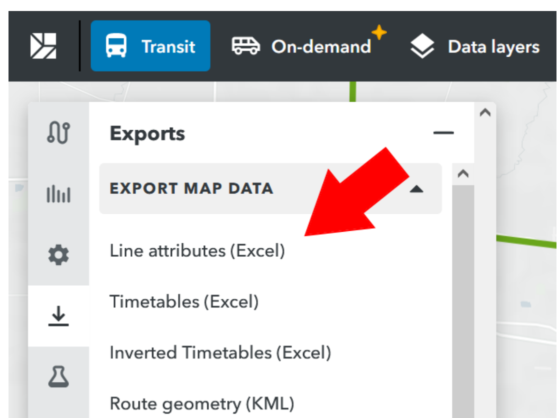


Figure 3 - Remix Line Attributes Button

Navigating TBEST

TBEST's software setup can be a significant process requiring a variety of datasets to be downloaded and formatted to create a socio-economic data package. For more information, this process is detailed in the TBEST User Guide and TBEST Socio-Economic Data Configuration guide. This section assumes the user has:

- Installed TBEST
- Constructed or Downloaded the Relevant Socio-Economic Data Package(s)
- Imported the Relevant GTFS File(s)

With the desired regional GTFS open in TBEST, the user can select all routes in the right-side "Routes" panel by clicking the route at the top of the list, holding the shift key, and clicking the route at the bottom of the list. With all the routes highlighted (selected), right click on any route to open a context menu. From this context menu select the "Route Socio-Economic Reporting…" option.

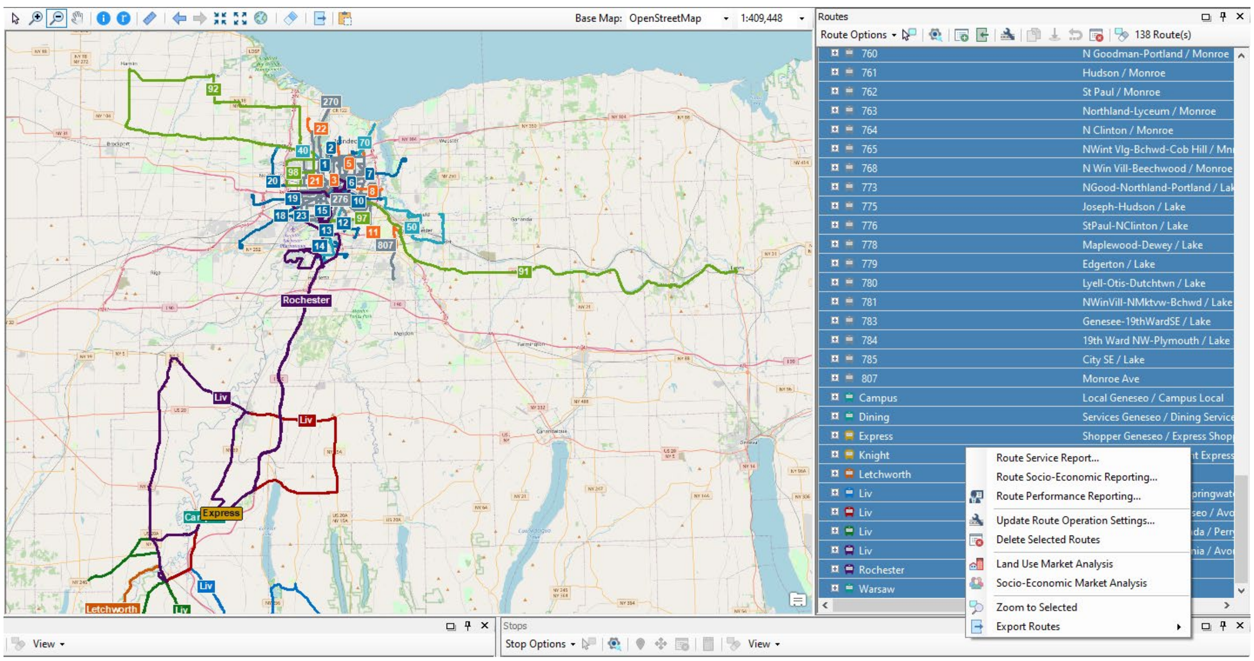


Figure 4 - TBEST Routes Context Menu

The Route Socio-Economic Report has several configuration options available. The report options were set to best resemble the data format exported from Remix to make the datasets comparable to each other. Since Remix captures Census data from a 0.25-mile buffer around the entire routes, the Service Period was set to "Weekday", Walk Access Distance was set to "0.25 miles", and Walk Access From: was set to "Segments".

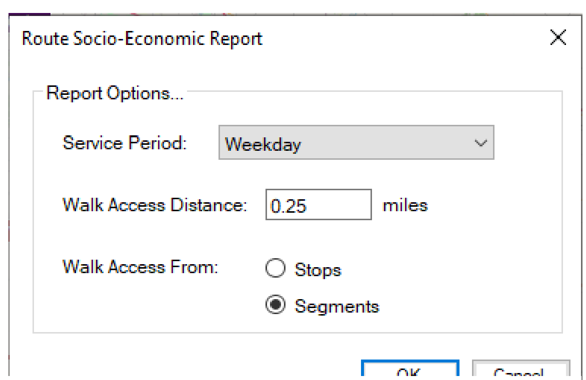
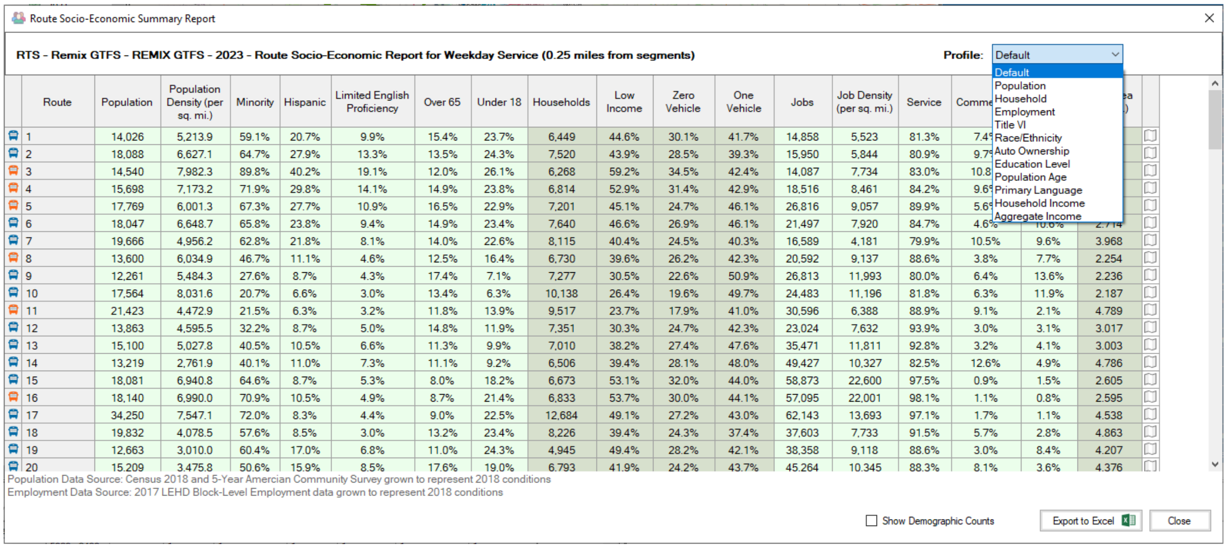


Figure 5 - TBEST Route Socio-Economic Report Configuration

The Route Socio-Economic Summary Report captures the census variables from the 0.25-mile route buffer selected on in the previous configuration window. By default, these values are displayed as percentages. To minimize post-processing, check the “Show Demographics Counts” box at the bottom of the window turn the percentages into values.



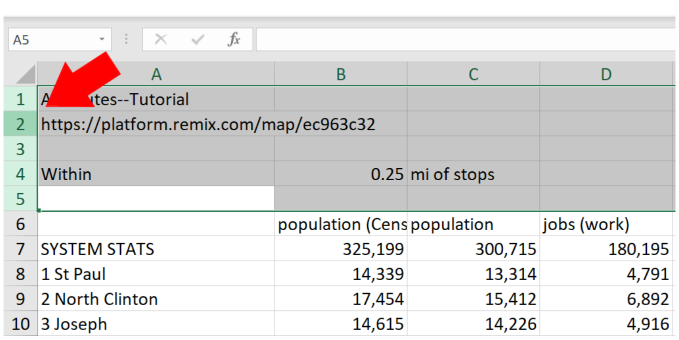
Route	Population	Population Density (per sq. mi.)	Minority	Hispanic	Limited English Proficiency	Over 65	Under 18	Households	Low Income	Zero Vehicle	One Vehicle	Jobs	Job Density (per sq. mi.)	Service	Comme		...a)
1	14,026	5,213.9	59.1%	20.7%	9.9%	15.4%	23.7%	6,449	44.6%	30.1%	41.7%	14,858	5,523	81.3%	7.4		
2	18,088	6,627.1	64.7%	27.9%	13.3%	13.5%	24.3%	7,520	43.9%	28.5%	39.3%	15,950	5,844	80.9%	9.7		
3	14,540	7,982.3	89.8%	40.2%	19.1%	12.0%	26.1%	6,268	59.2%	34.5%	42.4%	14,087	7,734	83.0%	10.8		
4	15,698	7,173.2	71.9%	29.8%	14.1%	14.9%	23.8%	6,814	52.9%	31.4%	42.9%	18,516	8,461	84.2%	9.6		
5	17,769	6,001.3	67.3%	27.7%	10.9%	16.5%	22.9%	7,201	45.1%	24.7%	46.1%	26,816	9,057	89.9%	5.6		
6	18,047	6,648.7	65.8%	23.8%	9.4%	14.9%	23.4%	7,640	46.6%	26.9%	46.1%	21,497	7,920	84.7%	4.6%	10.6%	2.714
7	19,666	4,956.2	62.8%	21.8%	8.1%	14.0%	22.6%	8,115	40.4%	24.5%	40.3%	16,589	4,181	79.9%	10.5%	9.6%	3.968
8	13,600	6,034.9	46.7%	11.1%	4.6%	12.5%	16.4%	6,730	39.6%	26.2%	42.3%	20,592	9,137	88.6%	3.8%	7.7%	2.254
9	12,261	5,484.3	27.6%	8.7%	4.3%	17.4%	7.1%	7,277	30.5%	22.6%	50.9%	26,813	11,993	80.0%	6.4%	13.6%	2.236
10	17,564	8,031.6	20.7%	6.6%	3.0%	13.4%	6.3%	10,138	26.4%	19.6%	49.7%	24,483	11,196	81.8%	6.3%	11.9%	2.187
11	21,423	4,472.9	21.5%	6.3%	3.2%	11.8%	13.9%	9,517	23.7%	17.9%	41.0%	30,596	6,388	88.9%	9.1%	2.1%	4.789
12	13,863	4,595.5	32.2%	8.7%	5.0%	14.8%	11.9%	7,351	30.3%	24.7%	42.3%	23,024	7,632	93.9%	3.0%	3.1%	3.017
13	15,100	5,027.8	40.5%	10.5%	6.6%	11.3%	9.9%	7,010	38.2%	27.4%	47.6%	35,471	11,811	92.8%	3.2%	4.1%	3.003
14	13,219	2,761.9	40.1%	11.0%	7.3%	11.1%	9.2%	6,506	39.4%	28.1%	48.0%	49,427	10,327	82.5%	12.6%	4.9%	4.786
15	18,081	6,940.8	64.6%	8.7%	5.3%	8.0%	18.2%	6,673	53.1%	32.0%	44.0%	58,873	22,600	97.5%	0.9%	1.5%	2.605
16	18,140	6,990.0	70.9%	10.5%	4.9%	8.7%	21.4%	6,833	53.7%	30.0%	44.1%	57,095	22,001	98.1%	1.1%	0.8%	2.595
17	34,250	7,547.1	72.0%	8.3%	4.4%	9.0%	22.5%	12,684	49.1%	27.2%	43.0%	62,143	13,693	97.1%	1.7%	1.1%	4.538
18	19,832	4,078.5	57.6%	8.5%	3.0%	13.2%	23.4%	8,226	39.4%	24.3%	37.4%	37,603	7,733	91.5%	5.7%	2.8%	4.863
19	12,663	3,010.0	60.4%	17.0%	6.8%	11.0%	24.3%	4,945	49.4%	28.2%	42.1%	38,358	9,118	88.6%	3.0%	8.4%	4.207
20	15,209	3,475.8	50.6%	15.9%	8.5%	17.6%	19.0%	6,793	41.9%	24.2%	43.7%	45,264	10,345	88.3%	8.1%	3.6%	4.376

Figure 6 - TBEST Route Socio-Economic Summary Report

Each “Profile” contains different variable attributes and will have to be exported to multiple Excel documents. Use the “Export to Excel” button to create and save the file generated for each relevant Profile. In this analysis, the Research Team exported the Default, Population, and Household profiles. These three sheets should be combined into one sheet that contains all the unique variables from each.

Spreadsheet Setup

Before scores can be calculated, the data will need to be processed in preparation. The downloaded spreadsheet **from Remix** contains Summary, Demographics, and Stops sheets, as well as a sheet for each of the lines selected in Remix. To prepare, we will need data from the Summary and Demographics sheets to make calculations based on route length. It is important to *not* sort any of the data in any of the sheets before beginning and to remember to save your work frequently. It is also recommended that a copy of the original spreadsheet is saved.



	A	B	C	D
1	A[illegible]tes--Tutorial			
2	https://platform.remix.com/map/ec963c32			
3				
4	Within	0.25	mi of stops	
5				
6		population (Cens	population	jobs (work)
7	SYSTEM STATS	325,199	300,715	180,195
8	1 St Paul	14,339	13,314	4,791
9	2 North Clinton	17,454	15,412	6,892
10	3 Joseph	14,615	14,226	4,916

Figure 7 - Remix Output

First, select the **Demographics** sheet and remove the top five rows. This will leave each column with a header. In cell A1—currently blank—enter in “Route”. Row 2 displays the aggregate data for the entire system and must be deleted or it will cause the composite scores and their ranking to be unsuitable for analysis.

Next, insert two new columns next to the Route column which will be used to paste inbound and outbound miles data from the **Summary** sheet. Do this by right-clicking on column B and clicking "Insert" to create one new column, and then repeat the step to create a second. Go to the **Summary** sheet and scroll down to row 53 where mileage, vehicle, and cost data per line is found. Using your mouse, select the "Inbound (mi)" and scroll down the list of lines until the very end. Then, holding shift, left click on the very last number in column C. Press CTRL+C on your keyboard to copy the data and then switch back to the **Demographics** sheet. Selecting the first cell in column B, press CTRL+V to paste the miles data into the **Demographics** sheet. It is important to make sure to have *not* sorted any of the data in either sheet before doing this otherwise the inbound and outbound miles will not match up with the routes.

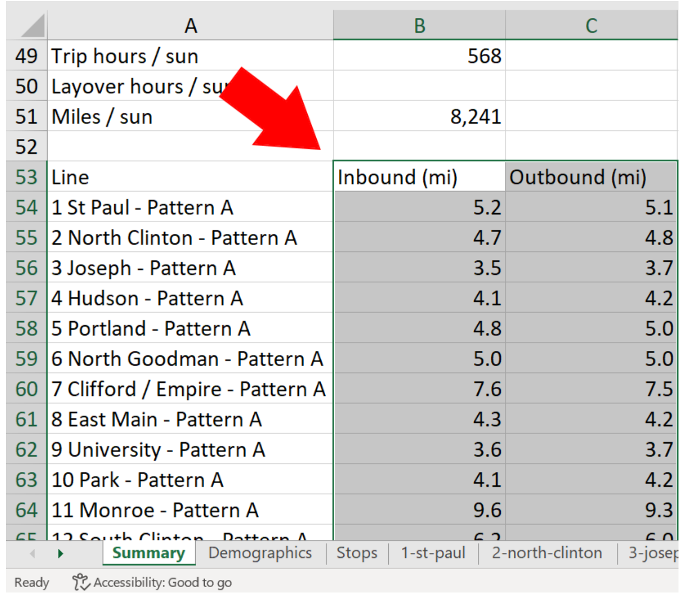


Figure 8 - Excel Inbound/Outbound Data

The last step to setting up the spreadsheet is to put all the Demographics data into a table. With the **Demographics** sheet selected, press CTRL+A to select all and click on the **Insert** ribbon at the top of the Excel window. Click on "Table" to put all the selected data into a table. Once this is completed, the cells in the sheet should be filled in various shades of blue. Check to ensure that the table has correctly included everything before moving on. Please note that this table is named **Table1** for reference in the creation of a **Composite Score Summary** sheet later.

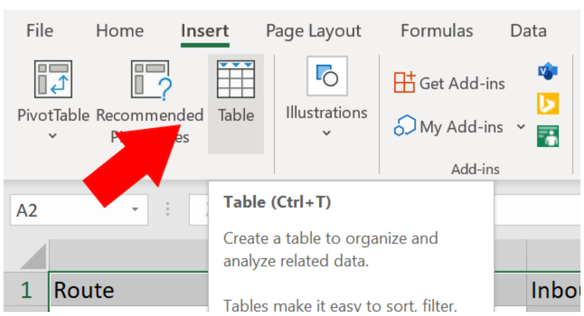


Figure 9 - Excel Insert Table

Data preparation

The data is now prepared for manipulation. Remix provides multiple variables as percentages of a total rather than the raw numbers which will need to be calculated before they can be used for score. Similarly, to rank scores by mile, the average length of a route needs to be calculated using the inbound and outbound variables.
The user will need to choose which variables they would like to use to create their score. For the purposes of this tutorial, we will be using:

- % of people in poverty.
- % households that are car free
- % people living with a disability
- % workers who take public transit to work
- % essential jobs (work)
- % people within 200% of the poverty threshold

Each of these variables will need to have a raw number calculated from the percentage using the population variable on the same row. This will allow the user to create normalized scores for each chosen variable, and then add them together to create a composite score.

To create an average length for each route, right click the column next to "Outbound (mi)" and insert a new column. Then name it "AVG Length." In the cell below, use the following formula:

=([@[Outbound (mi)]]+[@[Inbound (mi)]])/2

This will add the outbound and inbound miles together and divide them by two to get the average. The rest of the column will be automatically updated with the formula for each row.

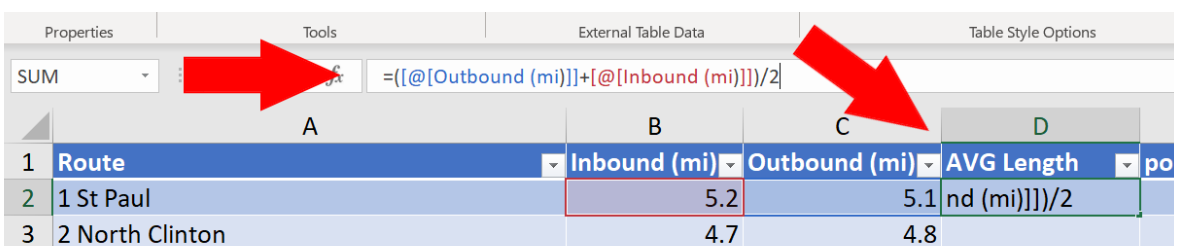


Figure 10 - AVG Length Equation

This new AVG Length column can be copied and pasted into the TBEST export sheet so those attributes can be calculated using the same length values.

Now it is time to create raw numbers from the percentages offered from the Remix data for each of the variables listed above, and then divide the value by the average length of the route. Each variable will need a new column to be inserted next to it to calculate the raw number. Name the column after the variable selected, but drop the "%" (e.g., "% of people in poverty" would have a new column next to it named "people in poverty").

Using "% of people in poverty" as an example, use the following formula:

=ROUND(([@[% of people in poverty]]*[@[population (Census 2020)]])/[@[AVG Length]],0)

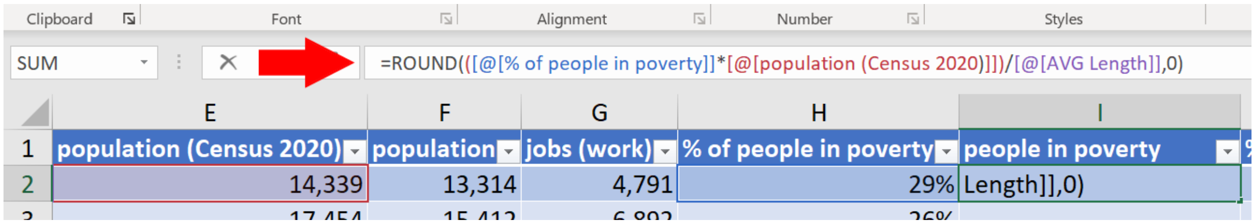


Figure 11 - Percent to Value Conversion

This will round the total of the “% of people in poverty” multiplied by the “population (Census 2020)” and divide the result by the inbound miles to get the approximate amount of people in the quarter-mile radius around the route. Do this for each variable selected.

Alternatively, the formulas for each variable can be set up to allow for the division by the average length to be toggled on and off using the summary sheet to be created later in the tutorial. To do so, the formulas need to be set up for each variable like the following example:

=IF('Composite Score Sheet'!$A$25<>"", ROUND(([@[% of people in poverty]]*[@[population (Census 2020)]])/[@[AVG Length]],0), ROUND(([@[% of people in poverty]]*[@[population (Census 2020)]]),0))

This formula adds an “IF” that will cause Excel to evaluate whether the condition following it is true or false, and then either perform one calculation, or the one the follows it in the formula. In this case, the formula will check cell A25 on the Composite Score Sheet (which will be created later in the tutorial) to see if it has any text values. If it does, it will then perform the first calculation which divides the number of people in poverty by the average length of the route, else it will just use the number of people in poverty without dividing it by the average length.

***

Please note that not all calculations should be made with the “Population (Census 2020)” variable. For instance, when trying to calculate the number of jobs (“% essential jobs (work)”), use the “jobs (work)” variable. Similarly, when trying to calculate the number of households that are car free, use the “households” variable in place of “Population (Census 2020).”

These would look like: =ROUND(([@[% essential jobs (work)]]*[@[jobs (work)]])/[@[AVG Length]],0)

And

=ROUND(([@[% of households that are car free]]*[@households])/[@[AVG Length]],0), respectively.

***

For each of these new variables created, the values must be normalized before they can be added together to create a composite score. To do so, create a new column next to the raw number variables and name it so it contains a normalized variable (e.g., “people within 200% of the poverty threshold” would have a new column next to it named “people within 200% of the poverty threshold normalized”).

This tutorial will be using a normalized range of 1-10—10 will represent the top-end of the values, with 1 representing the low-end.

In the newly created column, use the following formula with “BN:BN” being the column of the variable being normalized:

=1+9*(([@[people within 200% of the poverty threshold]])-MIN(BN:BN))/(MAX(BN:BN)-MIN(BN:BN))

This will create a value based on the 1-10 range considering the range of values for the variable selected. Repeat this step for each of the variables chosen.

The TBEST export will need to be post-processed by following similar steps above, excluding the percentage to value conversion equation. On the TBEST export, the following variable columns will require similar post-processing:

- In Poverty
- Zero Vehicle
- Disabled
- Service
- Commercial
- $20K - $24.9K
- $25K - $29.9K
- $30K - $34.9K
- $35K - $39.9K

Each variable should have two new columns created for it—Proxy and Normalized. The Proxy columns introduce the IF statement required for the per-mile toggle switch and use the equation:

=IF('Composite Score Sheet'!$A$25<>"", [@[Zero Vehicle]]/[@[AVG Length]], [@[Zero Vehicle]])

This expression says that if the toggle switch cell has any value in it, divide the attribute value by the route length, but if there is no value in the toggle switch cell, use the original value.

Follow the normalization steps noted for the Remix sheet above in each new Normalized column created in the TBEST output, normalizing the new Proxy Column:

=1+9*(([@[Zero Vehicle Proxy]])-MIN(K:K))/(MAX(K:K)-MIN(K:K))

Once completed, the user should have normalized scores for all the variables they wish to use in their composite score.

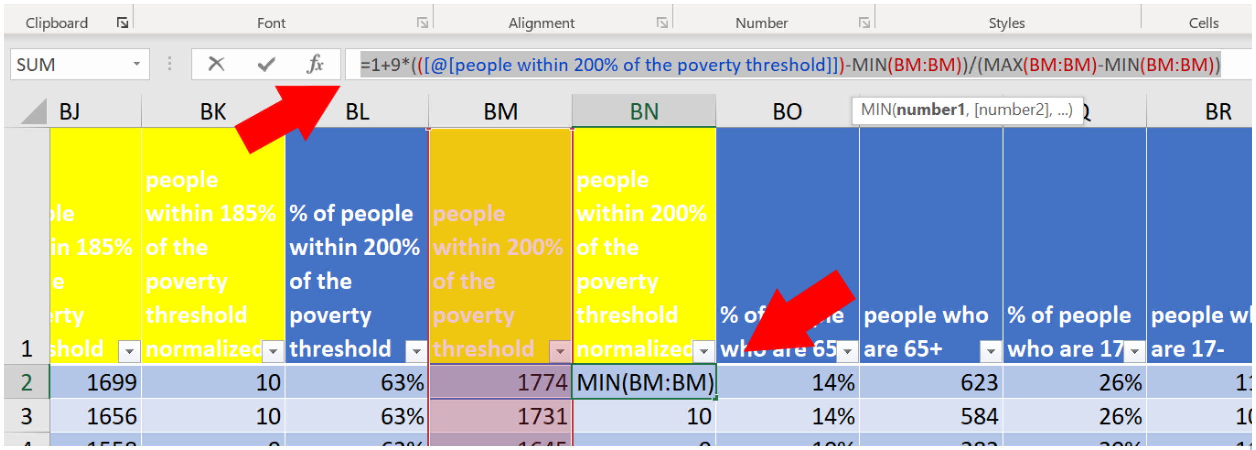


Figure 12 - Normalization Calculation

Calculating scores

With the data prepared, the user can now create their composite scores. First, a new blank sheet must be created that will be referenced later titled “Composite Score Sheet”. Return to the Remix output sheet to continue. A total score must be created by adding together all the normalized scores created in the proceeding step. Then, those scores must be normalized using a 1-100 range. Afterwards, the scores can be ranked and a score per mile value created.

On the **Remix Demographics** sheet, next to the “Route” column and the **TBEST export** sheet, create three new columns. Name these “Composite Score”, ”Composite Score Normalized”, and “Composite Score Ranked” to set up the columns necessary.

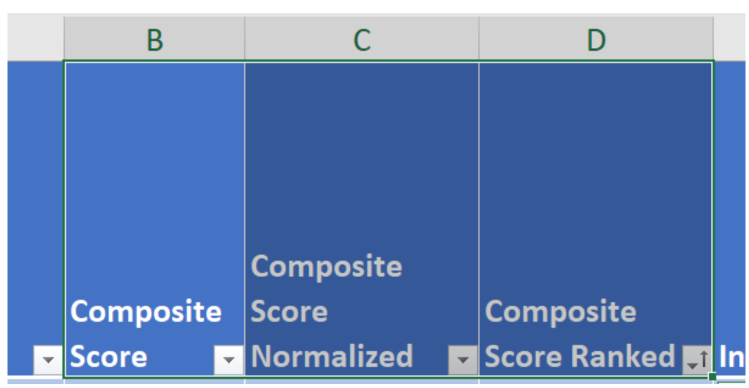


Figure 13 - Composite Score-Related Columns

First, create the composite score by adding together all the selected normalized variables. Use the following formula in the “Composite Score” column **for Remix**:

=([@[people in poverty normalized]]*'Composite Score Sheet'!$C$4)+([@[households that are car free normalized]]*'Composite Score Sheet'!$C$5)+([@[people living with a disability normalized]]*'Composite Score Sheet'!$C$6)+([@[workers who take public transit to work normalized]]*'Composite Score Sheet'!$C$7)+([@[essential jobs (work) normalized]]*'Composite Score Sheet'!$C$8)+([@[people within 125% of the poverty threshold normalized]]*'Composite Score Sheet'!$C$9)+([@[people within 150% of the poverty threshold normalized]]*'Composite Score Sheet'!$C$10)+([@[% of people within 185% of the poverty threshold]]*'Composite Score Sheet'!$C$11)+([@[people within 200% of the poverty threshold normalized]]*'Composite Score Sheet'!$C$12)+([@[public housing buildings]]*'Composite Score Sheet'!$C$13)+([@hospitals]*'Composite Score Sheet'!$C$14)+([@[urgent care facilities]]*'Composite Score Sheet'!$C$15)+([@[nursing homes]]*'Composite Score Sheet'!$C$16)+([@pharmacies]*'Composite Score Sheet'!$C$17)+([@[Schools: Childcare and K-12]]*'Composite Score Sheet'!$C$18)+([@[colleges / universities]]*'Composite Score Sheet'!$C$19)+([@[supplemental colleges]]*'Composite Score Sheet'!$C$20)+([@[SNAP Retailer Supermarkets 2020]]*'Composite Score Sheet'!$C$21)+(Table14[@[Commercial by Mile Normalized]]*'Composite Score Sheet'!$C$22)

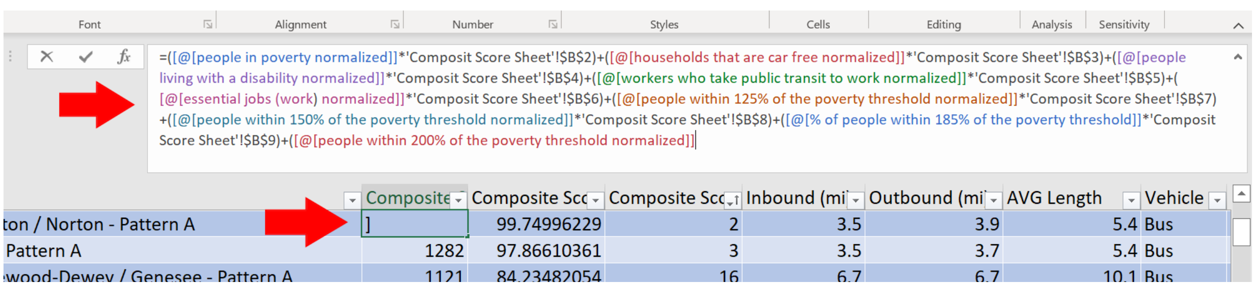


Figure 14 - Composite Score Calculation

For TBEST:

=([@[In Poverty Normalized]]*'Composite Score Sheet'!$C$4)+([@[Zero Vehicle Normalized]]*'Composite Score Sheet'!$C$5)+([@[Disabled Normalized]]*'Composite Score Sheet'!$C$6)+(Table13[@[workers who take public transit to work normalized]]*'Composite Score Sheet'!$C$7)+([@[Service Normalized]]*'Composite Score Sheet'!$C$8)+([@[$20K-$24.9K Normalized]]*'Composite Score Sheet'!$C$9)+([@[$25K-$29.9k Normalized]]*'Composite Score Sheet'!$C$10)+([@[$30K-$34.9K Normalized]]*'Composite Score Sheet'!$C$11)+([@[$35K-$39.9K Normalized]]*'Composite Score Sheet'!$C$12)+(Table13[@[public housing buildings]]*'Composite Score Sheet'!$C$13)+(Table13[@hospitals]*'Composite Score Sheet'!$C$14)+(Table13[@[urgent care facilities]]*'Composite Score Sheet'!$C$15)+(Table13[@[nursing homes]]*'Composite Score Sheet'!$C$16)+(Table13[@pharmacies]*'Composite Score Sheet'!$C$17)+(Table13[@[Schools: Childcare and K-12]]*'Composite Score Sheet'!$C$18)+(Table13[@[colleges /

universities]]*'Composite Score Sheet'!$C$19)+(Table13[@[supplemental colleges]]*'Composite Score Sheet'!$C$20)+(Table13[@[SNAP Retailer Supermarkets 2020]]*'Composite Score Sheet'!$C$21)+([@[Commercial Normalized]]*'Composite Score Sheet'!$C$22)

This will add together all the normalized scores, creating one composite score. To get a handle on what seems to be a complex formula, it is important to understand that it is a series of arguments repeated for each variable being used in the composite scoring. The formula is composed of a variable in Table1 multiplied by its corresponding modifier on the score sheet:

([@[people in poverty normalized]]*'Composite Score Sheet'!$C$3)

The first part references the header labeled “people in poverty normalized” on the current sheet (Table1) which it then multiplies by the value found on the **Composite Score Sheet** found at **column C**, **row 3**. The dollar sign ($) in front of the “C” and “3” ensures that the formula always links back to the same cell down the entire column rather than allowing Excel to automatically update the formula relative to each row. Add a plus sign and repeat for each variable being used in the composite scoring and the finished formula will look like the previous full formula. Double check that each reference to a cell on the score sheet (e.g., !$C$3) matches with the variable it should and that all parenthesis open and close where they should, then paste the full formula into the cell in Table1.

Next, these values need to be normalized using the following formula in the “Composite Score Normalized” column with “B:B” being the corresponding column:

=1+99*(([@[Composite Score]])-MIN(B:B))/(MAX(B:B)-MIN(B:B))

This will normalize the composite scores using a 1-100 range. Afterwards, in the “Composite Score Ranked” column, use the following formula:

=RANK(([@[Composite Score Normalized]]),B:B,0)

With all of this completed on both Remix and TBEST sheets, the user can now create composite scores, normalize the scores, and rank them. The next section explains the process of constructing a composite score summary sheet where each variable can be weighted differently, with calculations and rankings updated automatically.

Composite Score Summary Sheet

In the previous section, composite scores created and ranked but the variables used were all unweighted. Weighting variables differently is a good way to tailor your composite score to focus on different characteristics of interest. To do this, a new sheet must be created that will reference the table created earlier and allow the user an easy to view place to see the data summarized and change the variable weights. Once the new sheet has been setup, the composite score column on the previous sheet will need to be edited to take the weighting from the new sheets into account by updating the composite score formula.

Create a new sheet and name it “Composite Score Sheet.” This sheet will contain two columns to list the variables used in the composite scoring from the previous sheet and display the score weights. Additional columns will be used to display route ranking, route name, route composite score, and the normalized composite score. As the weights are updated, the scores for each route will update and be reranked in real-time on this sheet.

	A	B	C
1			
2	Composite Score Builder		
3	TBEST Variable	Remix Variable	Score Modifier
4	In Poverty Normalized	people in poverty normalized	1
5	Zero Vehicle Normalized	households that are car free normalized	1
6	Disabled Normalized	people living with a disability normalized	1
7		workers who take public transit to work normalized	20
8	Service Normalized	essential jobs (work) normalized	20
9	$20K-$24.9K Normalized	people within 125% of the poverty threshold normalized	1
10	$25K-$29.9K Normalized	people within 150% of the poverty threshold normalized	1
11	$30K-$34.9K Normalized	people within 185% of the poverty threshold normalized	1
12	$35K-$39.9K Normalized	people within 200% of the poverty threshold normalized	1
13		public housing buildings	1
14		hospitals	10
15		urgent care facilities	10
16		nursing homes	10
17		pharmacies	10
18		Schools: Childcare and K-12	10
19		colleges / universities	0
20		supplemental colleges	0
21		SNAP Retailer Supermarkets 2020	10
22	Commercial Normalized		0

Figure 15 - Variables for Score Weighting

In the **A** and **B** columns, enter the relevant variables associated with TBEST and Remix, respectively as shown in the figure to the right. The **C** column will be the score modifiers associated with each variable calculation. The variable columns will have a row for each variable included within the composite score while the “Score Modifier” column will contain the value used to weight each variable. Enter in the weights desired under the “Score Modifier” column (e.g., “10” for “people in poverty normalized” or “5” for “workers who take public transit to work normalized”). These can be adjusted as desired later. With this completed, make note of each variable used in the composite scoring on the previous sheet. This tutorial uses the following:

- TBEST
    - In Poverty
    - Zero Vehicle
    - Disabled
    - Service
    - Commercial
    - $20K - $24.9K
    - $25K - $29.9K
    - $30K - $34.9K
    - $35K - $39.9K
- Remix
    - people in poverty normalized
    - households that are car free normalized
    - people living with a disability normalized
    - workers who take public transit to work normalized
    - essential jobs (work) normalized

- people within 200% of the poverty threshold normalized

Below the “Variable” and “Score Modifier” columns is where the toggle for whether to divide values by their route’s average length should be created. Above, cell A25 was used, so label cell A24 “Calculate per Mile?” to mark the toggle on the sheet. If any text is entered into cell A25 (e.g., “Y”), then the formulas will divide by the average route length.

24	By Mile Toggle
25	x

Figure 16 - By Mile Toggle

Next, skip column **D** and Enter Remix in E1:G1 and TBEST in I1:K1 and “Line”, “Composite Score”, and “Composite Score Normalized”, “Rank”, “Composite Score Normalized”, “Composite Score”, and “Line” in **row 3** of **columns E through K**, respectively—as shown in the figure below.

E	F	G	H	I	J	K
REMIX				TBEST		
Line	Composite Score	Composite Score Normalized	Rank	Composite Score Normalized	Composite Score	Line
753 19th Ward-Plym-City SE / Monroe - Pattern A	866.8	100.0	1	100.00	847.9	753 19th Ward-Plym-City SE / Monroe - Pattern A
450 N Goodman-Portland / Genesee - Pattern A	852.8	98.3	2	98.27	833.9	450 N Goodman-Portland / Genesee - Pattern A
443 Joseph-Hudson / Genesee - Pattern A	851.7	98.2	3	97.89	830.9	443 Joseph-Hudson / Genesee - Pattern A
436 City SE / Norton - Pattern A	844.0	97.2	4	96.16	816.9	768 N Win Vill-Beechwood / Monroe - Pattern A

Figure 17 - Score Ranking Layout

In the “Rank” column, enter “1” into **row 3** and drag down to create as many ranks as there are routes in the data acquired from Remix—this tutorial has 138 routes and thus will have 138 ranks.

The next three columns will use formulas to pull route names, scores, and normalized scores from the previous sheet. These three columns will automatically update as the user makes changes to the score modifiers to weight variables, allowing for dynamic rescoring and ranking. Like the “Variable” column, data must be pulled from the previous sheet. The following three formulas should be pasted into the first row beneath the “Route”, “Composite Score”, and “Composite Score Normalized” columns:

=INDEX(Table1[Line], MATCH('Composite Score Sheet'!D2, Table1[Composite Score Ranked],0))

And

=INDEX(Table1[Composite Score], MATCH('Composite Score Sheet'!D2, Table1[Composite Score Ranked],0))

And

=INDEX(Table1[Composite Score Normalized], MATCH('Composite Score Sheet'!D2, Table1[Composite Score Ranked],0))

Each of these will reference back to **Table1** (or whichever table number references the Remix **Demographics** sheet) or Table2 (or whichever table number references the TBEST sheet) and pull the data pointed to. Copy these down each respective column down to rank 138 in **columns E and K** so that every route will be covered.

The table should automatically update the entire column with the new formula. The scores should have automatically updated based on the weights put into the score modifier column on the score sheet. Switch to the score sheet and then change a variable weight to see the rankings change in real time.

With this additional step, the user can make a more nuanced analysis based on the data they have access to, and with a minimal amount of Excel work for additional variables or analyses.

Sensitivity Analysis

A sensitivity analysis is performed to determine how different variables affect a dependent variable. In this case, we will be performing iterative analyses to determine how different variables affect the route composite scores. Changes will be made to the variables from the **Composite Score Summary** sheet and the results of those changes recorded on a separate sheet, with one row for each iteration.

To do this, a new sheet will need to be created to record the values for the variables used to calculate the composite scores for each iteration and new columns created on the score sheet to record the score changes for each. Create a new sheet named **Sensitivity Analysis Parameters**. Select and copy all the variables from the **Composite Score Summary** sheet under the “Composite Score Builder” that are used to calculate the composite scores, and right click and select “transpose” to paste them in as column headers rather than rows. Insert two new columns next to **A1** and name them “Run” and “Notes” to record each iteration of analysis performed and notes on what the goal of the analysis was.

In the **Sensitivity Analysis Parameters** sheet, enter in the desired values for each modifier variable and then select all, copy them, and do a “transpose” to paste them in to the “Composite Score Builder” section of the sheet. The scores will automatically update. Make a copy of the **Demographics** sheet, and remove all the variables except for “Line”, ”Composite Score”, ”Composite Score Against All Routes”, and “Network Rankings”. Name the new sheet **Sensitivity Analysis Outcomes.** Next, add three new columns and name them “Run 1 Composite Score”, “Run 1 Composite Score Against All Routes”, and “Run 1 Network Rankings.” Rename the original composite score and rankings headers to include “Current” (e.g., Composite Score becomes Current Composite Score). Next to “Current Composite Score,” add a new column and name it “Average Rank of All Runs”.

On the **Sensitivity Analysis Outcomes** sheet, copy the values in the cells for the ”Composite Score”, ”Composite Score Against All Routes”, and “Network Rankings” columns and right click and select “Paste Values” under the new “Run 1” columns. Using this method, the user can record each set of scores and rankings as they change with each iteration of modifier changes.

The workflow should be adding a new row for a new run to the **Sensitivity Analysis Parameters** sheet, copying the values and pasting them using transpose to the “Composite Score Builder” on the **Composite Score Sheet**, then creating three new columns on the **Sensitivity Analysis Outcomes** sheet to record the values for that run. Each successive run should build on the previous until the user desires to perform a different analysis focusing on different modifiers.

Aligning Service With Demand

Link to Workflow

Overall Approach

The workflow involved two workstreams that were undertaken sequentially.

The first task involved preparing the proposed redesign of Oswego's City Routes. This was a route-planning effort in which a workable solution needed to be informed by best-available data on travel demand patterns in Oswego and was also feasible operationally. The main work product of this task was a proposed service plan.

Once the service plan was complete, the second task was to prepare ridership forecasts for the proposed system redesign. The objective was to quantify the impact of the service plan on ridership, as a key piece of decision-support information for Centro to consider.

Overview of Transit Propensity Indices

A traditional approach for evaluating candidate locations for increased (or decreased) transit service is to evaluate a Transit Propensity Index (TPI)[1] for a small-area Census geography (Census blocks, block groups, or tracts). Typical inputs into a TPI formula include indicators of propensity to use transit, such as overall population, low-income population, population identifying as racial/ethnic minority status, overall employment, employment in sectors such as retail, etc. Data is generally drawn from the decennial Census[2] and/or related datasets such as the Census Bureau's American Community Survey[3] which is a smaller-sample survey that includes information that is not collected in the decennial Census (e.g., household income).

Calculating a Transit Propensity Index for each small-area unit of geography results in a two-dimensional surface, with some of the areal units identified as scoring high in transit propensity and others scoring low. This provides the evidence base that a bus service planning team would then use to re-design bus routes. A sample mapping of a TPI surface is shown in Figure 2.

[1] See e.g. https://www.cutr.usf.edu/wp-content/uploads/2012/08/propensity.pdf
[2] https://www.census.gov/programs-surveys/decennial-census.html
[3] https://www.census.gov/programs-surveys/acs

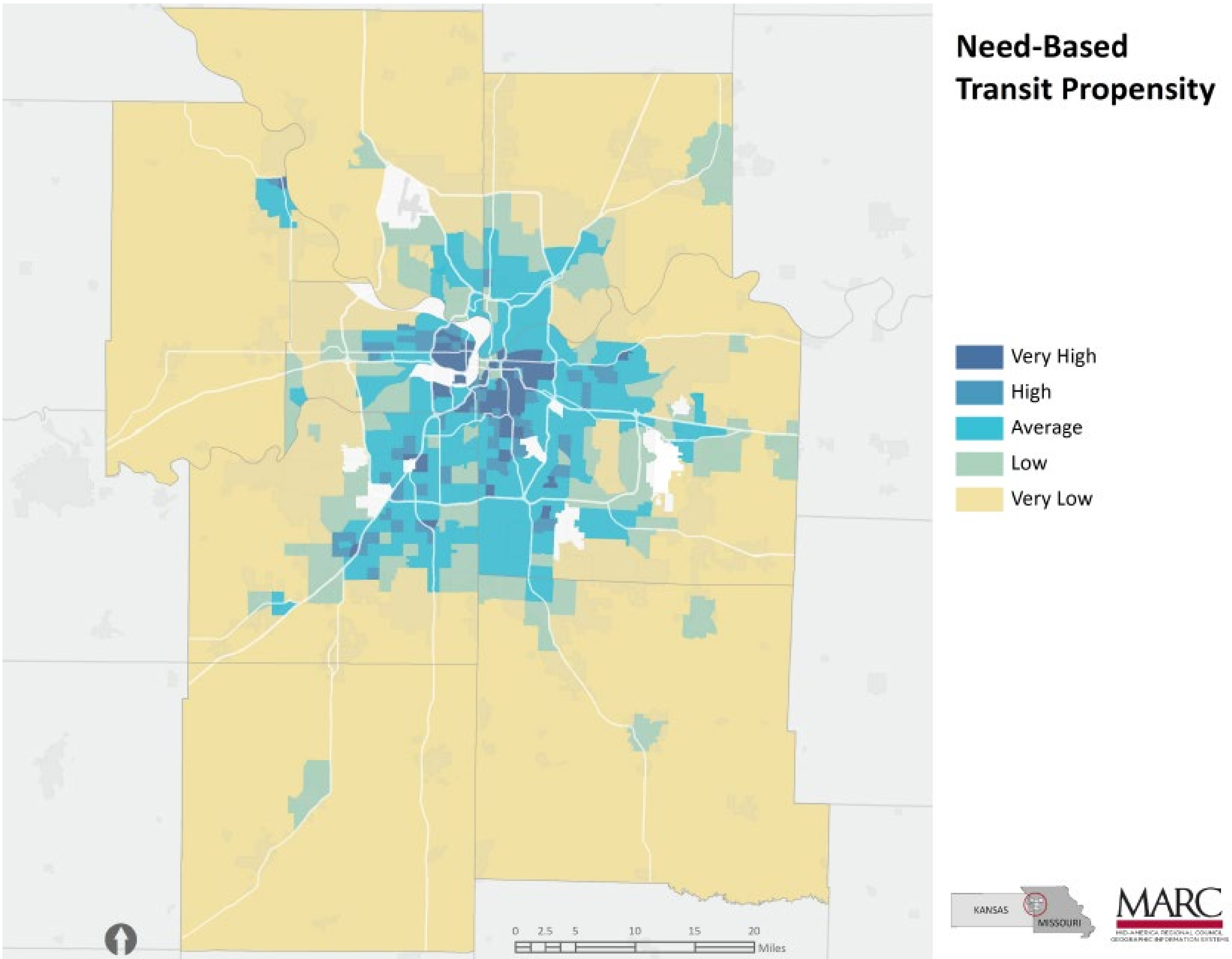


Figure 18: Example of a traditional two-dimensional mapping of a Transit Propensity Index for transit-service planning purposes. Source: http://www.kcsmartmoves.org/pdf/PlanDocuments_2017/Appendix_F.pdf

Need and Value for New Data Sources to Support Enhanced Transit Propensity Indices

The traditional approach to developing a Transit Propensity Index has several shortcomings:

1) Not sensitive to day-of-week or time-of-day, thus does not provide evidence which could be used to determine span-of-service
2) Not sensitive to the origin-destination patterns of prospective transit riders, and rather focusing on unlinked trip ends (origins separately from destinations)

Software Tool for Transit Propensity: *Replica*

This Pilot drew on *Replica*[4] data to perform an extended analysis of transit propensity. Replica data is a commercial data product to which all MPOs in New York State now have access through the vendor's online portal. Replica is a synthetic travel-survey database with advanced relational characteristics; it is described by its vendor as a *"High-fidelity travel model with simulated population and trip-taking activity"*.

Replica data are synthetic and fused from a variety of sources, such as the decennial Census and American Community Survey datasets noted above, as well as Census Transportation Planning Package (CTPP) data, Longitudinal Employer-Household Data (LEHD), in-vehicle GPS traces, and published count data such as Average Annual Daily Traffic (AADT). Replica data are generated for a representative day (separately for weekday or Saturday) of a season (e.g. the latest available data represent either a Thursday or Saturday, at the selection of the data-user, in Fall 2021). Replica data are available both in summary form through the online portal, or in disaggregate form (in format very similar to traditional travel-survey microdata). A major advantage of

Replica synthesizes the type of data that would normally come out of a custom regional travel survey, which does not exist in many parts of the country – including Oswego. In addition to providing quasi-travel-survey data where none exists, another strength of Replica data is that it, in principle, represents *all* travel rather than the small-sample slice of a region's travel (typically in the low single-digit percentages or lower) that can realistically be sampled in a regional household travel survey.

Additionally, Replica makes use of advanced relational-database functionality to allow the user to query the data in useful ways. For instance, a user could query all trips within a region, only trips that begin and/or end within user-defined locations, or only trips that make use of user-specified segments of a road network. Trip characteristics that are reported by Replica include journey purpose and time-of-day.

Figure 3 shows the spatial pattern of trips to retail centers in eastern Oswego. The unit of geography is census blocks: while Replica's standard smallest unit of geography is the census block-group, there is a function to allow the user to request that Replica make data available in a particular location at finer-scale geography, which was done on this Pilot. On the right-hand side are various characteristics showing the time-of-day profile of trips that begin in this shopping area, the pattern of mode usage (dominated by auto-driver and auto-passenger), and the distribution of journey purpose (dominated by shopping, as would be expected given the retail land use). While the data-presentation of Replica's online portal shown in Figure 3 is somewhat basic, the user can download (in the open CSV data-format) the microdata that underpins this visualization, and use any other desired visualization tool or approach.

[4] https://www.replicahq.com/

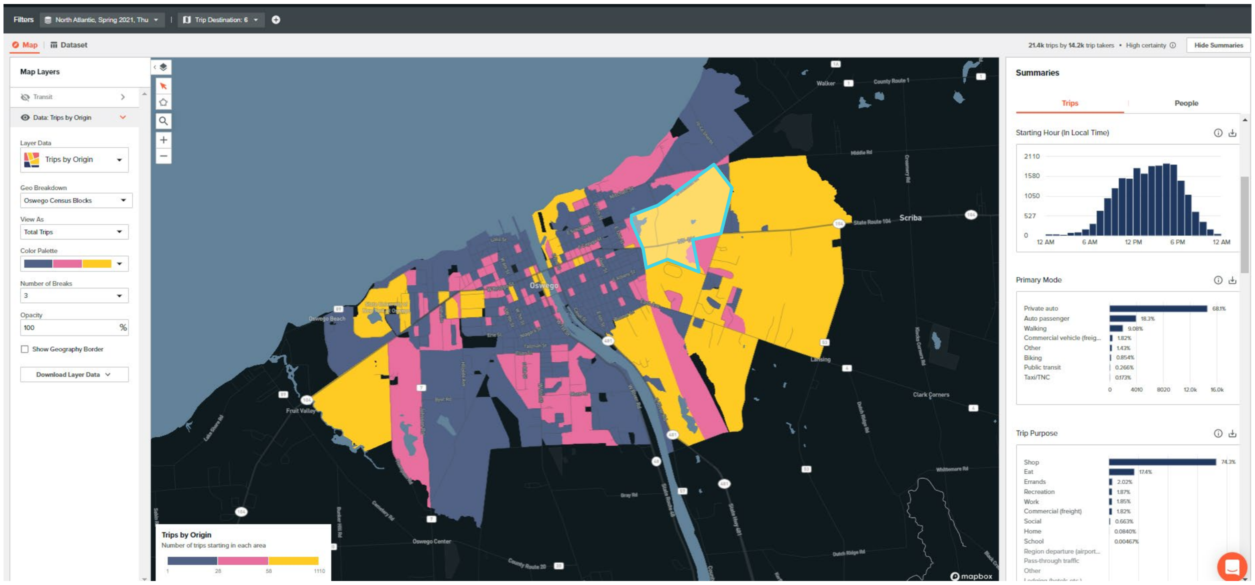


Figure 19: Replica online portal depicting the destination pattern of all trips which begin in the highlighted shopping district in eastern Oswego.

Preparation of Enhanced Transit Propensity Index

For this Pilot, we generated an enhanced Transit Propensity Index using the origin-destination attribute of Replica data, which provides functionality that extends beyond what can be done with a typical two-dimensional TPI surface (see Figure 2).

To do this, we first generated an origin-destination matrix of trips in Replica's synthetic database that met both of the following characteristics:

- Household income under $30,000/year
- Household car ownership of zero or one

Replica's system facilitated on-the-fly testing of various permutations of transit-propensity criteria (e.g.. HH income under $50K, only zero-car HHs, etc.), allowing the Research Team to conduct a sensitivity analysis to determine which criteria listed above which were found to work best for this Pilot's application.

The microdata for trips in Oswego that meet the specified transit-propensity criteria were exported from the Replica online portal and seamlessly imported into ArcMap, using the longitude/latitude coordinates of each block group. The mapping of the origin-destination pattern of trips meeting this Pilot's Transit Propensity Index attributes is shown in Figure 4. The O-D pairs highlighted in light-blue are in the top-10 O-D pairs in Oswego County. For reference, Figure 5 shows this same data but in a traditional Transit Propensity Index's two-dimensional surface, by mapping only origins rather than connecting origins and destinations as in Figure 4.,

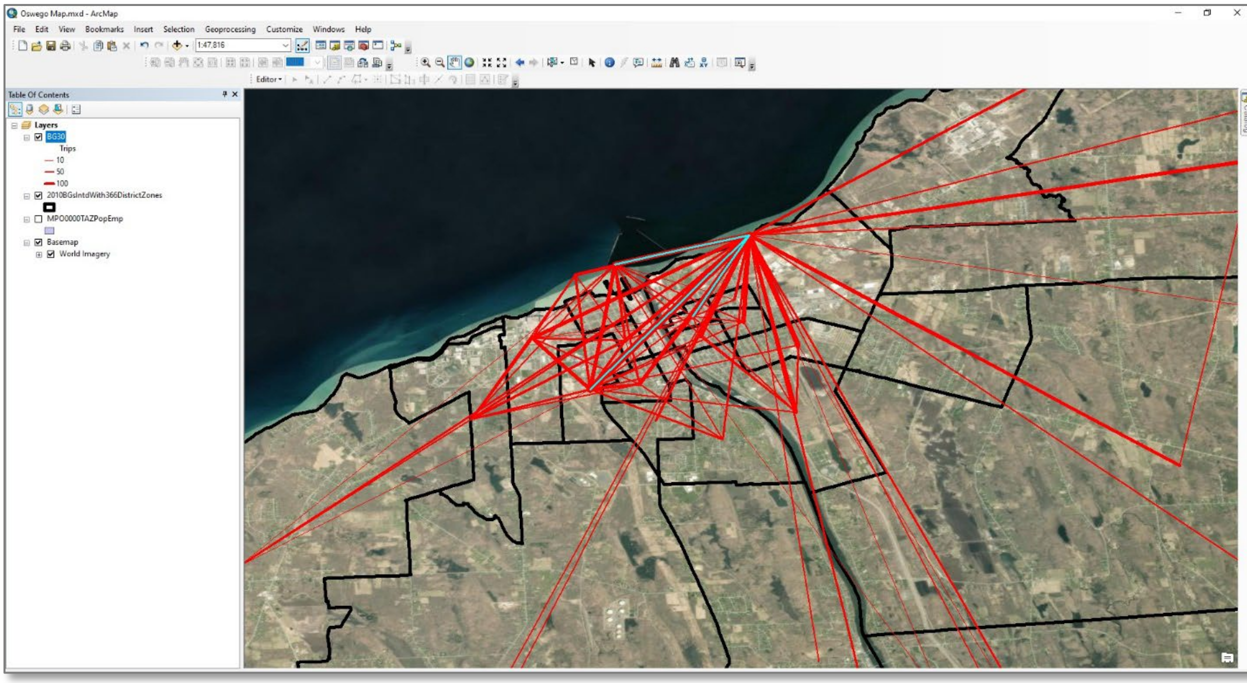


Figure 20: Mapping of origin-destination pattern of trips meeting this Pilot's Transit Propensity Index criteria

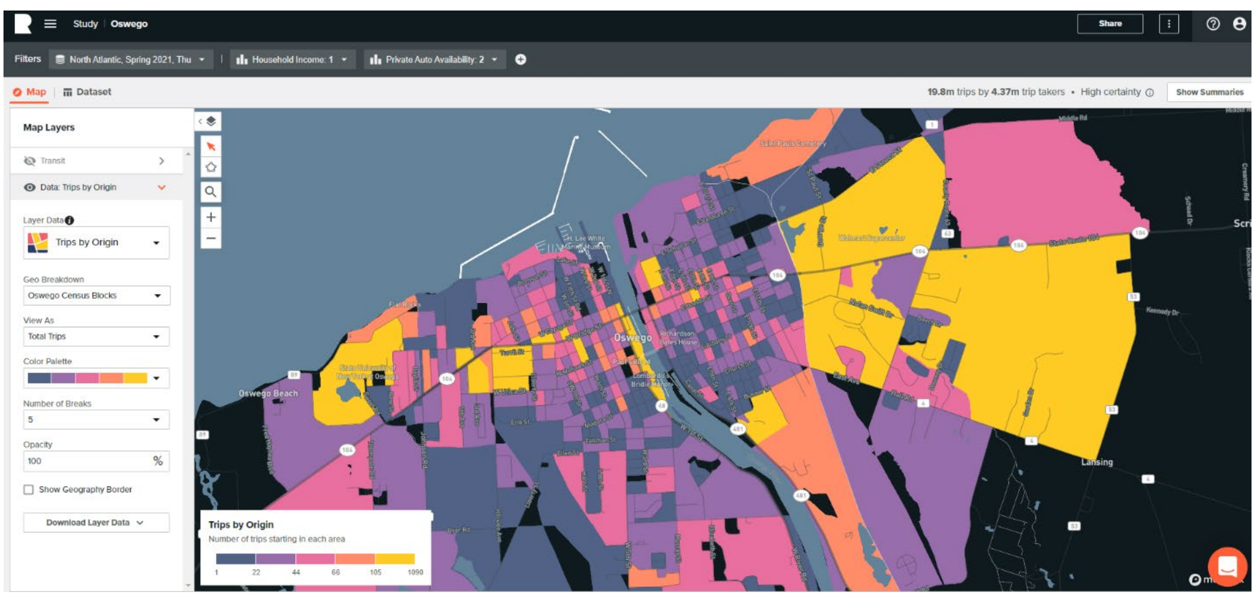


Figure 21: Two-dimensional mapping of this Pilot's Transit Propensity Index

Inputs Into Bus-Service Redesign

In order to develop the proposed redesign of the Oswego City Routes, the Pilot Research Team took into account the following information:

- The enhanced OD-based Transit Propensity Index generated from Replica data (see previous section)
- Ridership data (boardings) for Oswego's City Routes' bus-stops from Centro's Automatic Passenger Counter (APC) system (see Figure 6). APC data identified the SUNY Oswego campus, the Wal-Mart, and the City Center as the three largest generators of transit boardings.
- Guidance from Centro staff to consider straightening and simplifying the bus services
- Existing service patterns and destinations served (see Figure 1 which shows existing service map with major destinations highlighted), including Centro's run sheets for the existing City Routes
- The City of Oswego's geographical context: a smaller-scale urban core with relatively limited surrounding development, which is oriented east-west but with limited bridge crossings of the north-south Oswego River
- Feedback from Centro staff regarding physically difficult bus movements (e.g., locations where right-turns cannot be easily navigated by a bus) to be avoided

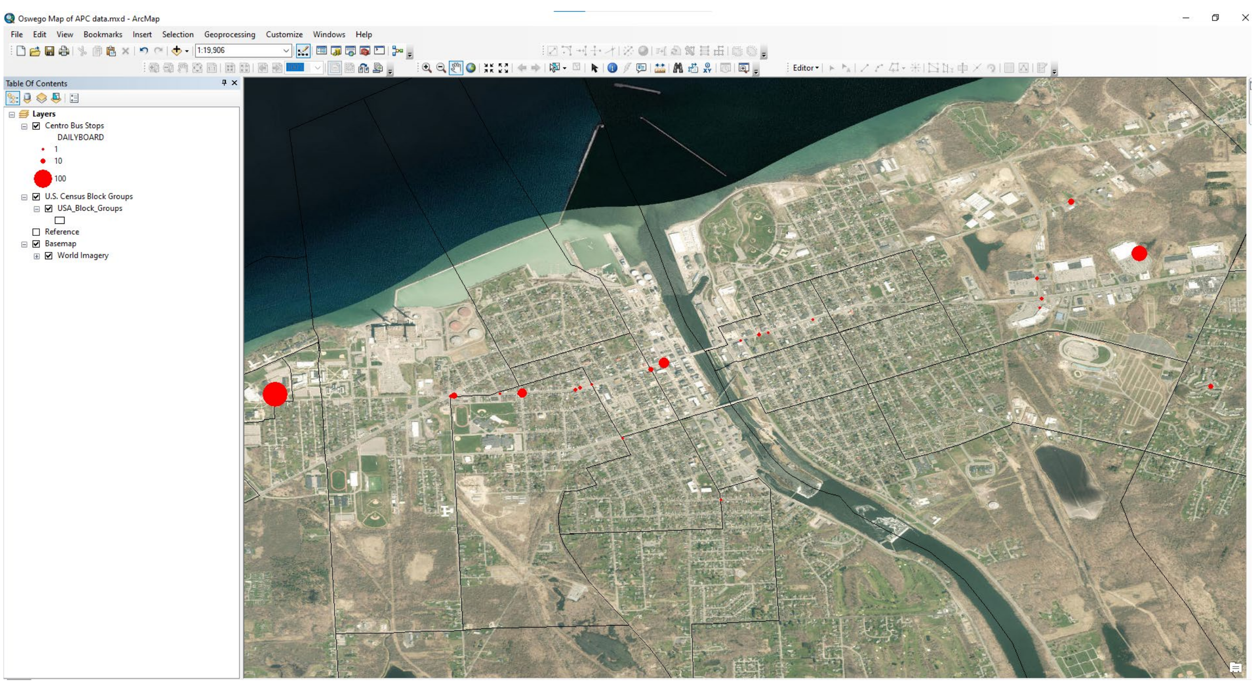


Figure 22: Number of boardings at bus stop locations from Centro's Automatic Passenger Counter data. SUNY campus is at west; Wal-Mart is at east; City Center is in the center of map immediately to west of Oswego River

Prospective Bus-Service Redesign

An iterative process was undertaken by this Pilot's Research Team and Centro staff to develop the proposed bus-service redesign, which accounted for the inputs enumerated in the previous section.

The prospective service maintains the east-west orientation of the existing City Routes, as well as the number of buses operating (three). Service is consolidated onto one higher-frequency route which operates east-west along Rt. 104 (Bridge Street) and connects between the SUNY campus in the west, the City Center in downtown Oswego, and the Wal-Mart and other nearby shopping opportunities in the east. This "High-frequency route" would be served by two buses operating in opposite directions. Additionally, a lower-frequency route ("South route") is proposed, with an allocation of one bus, which serves a number of important destinations that are roughly ¾-mile to 1-mile to the south of Rt. 104. Figure 7 depicts the two proposed routes; the High-frequency route would generally operate 30-minute headways, and the South route would generally operate 120-minute headways.

Figure 23: Prospective redesign of Centro's City Routes in Oswego

Preparation of Inputs for Ridership Forecasting

The next task involved preparing the set of inputs that would be used to develop the ridership forecasts for the prospective system redesign. Three key pieces of information were required:

1) A complete description of the operations of the existing City Routes as well as the prospective system redesign in General Transit Feed Specification (GTFS)[5] data standard. Note that the STOPS ridership-forecasting software only considers bus operations during the weekday morning peak period (7:00 to 9:00) and weekday midday (12:00 to 2:00) periods, thus GTFS data is only required during these periods.

[5] https://gtfs.org/

2) Data on existing bus ridership patterns in Oswego, for use calibrating the ridership model prior to application of it for generating the ridership forecast
3) A geographic zone structure to be used by the STOPS ridership-forecasting software

Ridership-Forecasting Input #1: GTFS Data

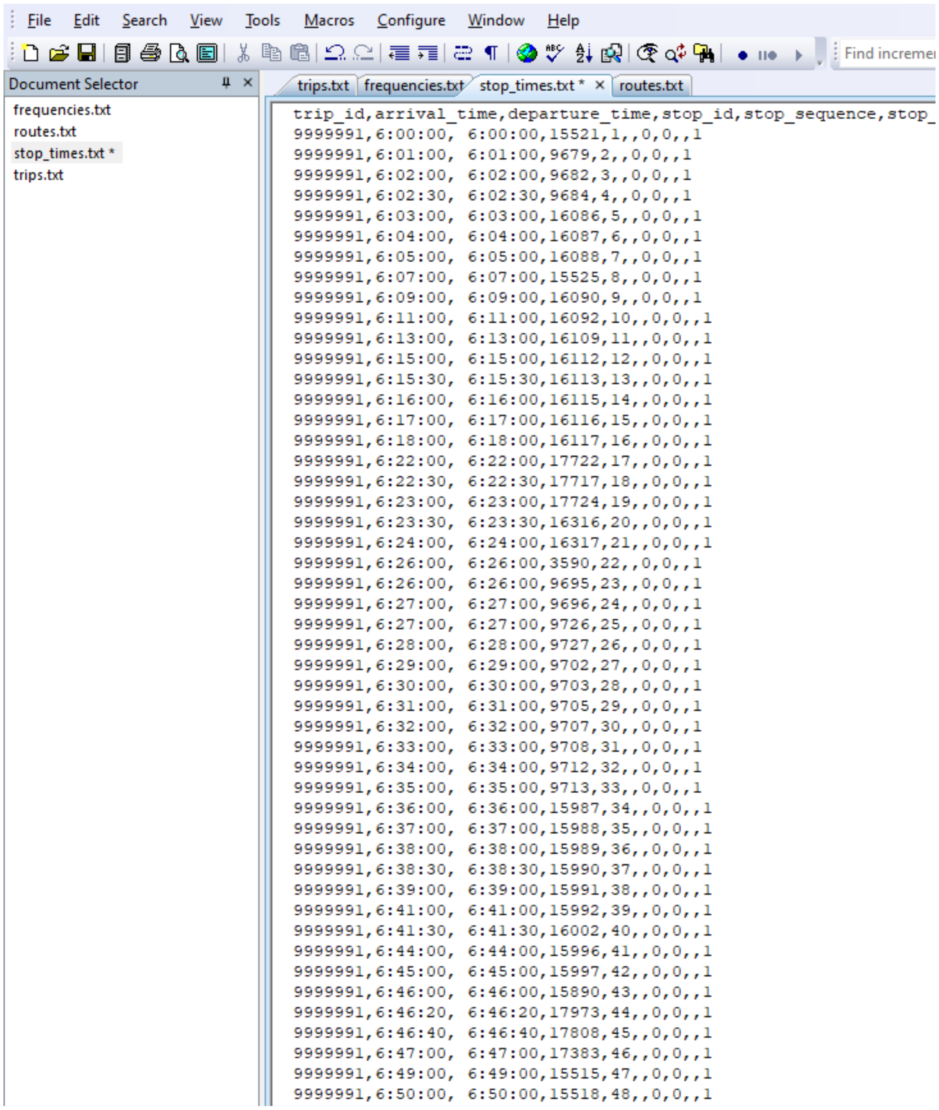
The General Transit Feed Specification is an industry standard data specification that is used by thousands of transit providers to describe their operations.

The GTFS specification consists of multiple simple text files that each describe a different aspect of a transit service[6]. The software selected for preparing the ridership forecasts (STOPS; see below) requires two sets of GTFS files as input: one that describes the existing transit service (which Centro provided to the Research Team) and one that describes the prospective modified service (which was prepared by the Research Team).

In this Pilot, four of the GTFS files required editing:

- **Routes.txt:** This datafile has one line per each route in a transit system. One new line of code was required for each new route, thus in the Route.txt file only two new lines of code were required (one each for the High-frequency route and the South route)
- **Trips.txt:** This datafile has one line of code per each time that a bus completes a run of a route. For instance, one bus traveling on the High-frequency route from SUNY to Wal-Mart is one 'trip' as defined by GTFS for use in this file, and the same bus traveling in the reverse direction back to SUNY is another trip.
- **Stop times.txt:** This datafile has one line per each time that a bus stops at a bus stop. A bus trip's trajectory is defined by the timepoint at which it arrives (and same for when it departs) a given bus stop location. This file grows large in the number of rows because there is one row per each bus-vehicle stopping at each bus stop. Figure 8 shows an example of part of a Stop times.txt file created for this Pilot, with a bus operating on the South Route starting at 6:00 AM at the SUNY campus and traveling stop-by-stop to end at the Wal-Mart at 6:50 AM. GTFS does not specify the turn-by-turn trajectory of a bus in revenue service, rather only the stop locations and the timepoints. The average speed of the bus trip is implicit in the timepoints; in this Pilot an assumed average operating speed of 10 mph was assumed for purposes of specifying the timepoints.
- **Frequencies.txt:** This datafile has one line of code per each trip defined in the Trips.txt file. It allows the analyst to make reference to a defined bus 'trip' and specify that the trip is repeated on a recurring basis with a defined frequency (in units of minutes).

[6] The full set of GTFS text files and the datafields expected in each of them can be found at: https://gtfs.org/schedule/reference/




```
trip_id,arrival_time,departure_time,stop_id,stop_sequence,stop_
9999991,6:00:00, 6:00:00,15521,1,,0,0,,1
9999991,6:01:00, 6:01:00,9679,2,,0,0,,1
9999991,6:02:00, 6:02:00,9682,3,,0,0,,1
9999991,6:02:30, 6:02:30,9684,4,,0,0,,1
9999991,6:03:00, 6:03:00,16086,5,,0,0,,1
9999991,6:04:00, 6:04:00,16087,6,,0,0,,1
9999991,6:05:00, 6:05:00,16088,7,,0,0,,1
9999991,6:07:00, 6:07:00,15525,8,,0,0,,1
9999991,6:09:00, 6:09:00,16090,9,,0,0,,1
9999991,6:11:00, 6:11:00,16092,10,,0,0,,1
9999991,6:13:00, 6:13:00,16109,11,,0,0,,1
9999991,6:15:00, 6:15:00,16112,12,,0,0,,1
9999991,6:15:30, 6:15:30,16113,13,,0,0,,1
9999991,6:16:00, 6:16:00,16115,14,,0,0,,1
9999991,6:17:00, 6:17:00,16116,15,,0,0,,1
9999991,6:18:00, 6:18:00,16117,16,,0,0,,1
9999991,6:22:00, 6:22:00,17722,17,,0,0,,1
9999991,6:22:30, 6:22:30,17717,18,,0,0,,1
9999991,6:23:00, 6:23:00,17724,19,,0,0,,1
9999991,6:23:30, 6:23:30,16316,20,,0,0,,1
9999991,6:24:00, 6:24:00,16317,21,,0,0,,1
9999991,6:26:00, 6:26:00,3590,22,,0,0,,1
9999991,6:26:00, 6:26:00,9695,23,,0,0,,1
9999991,6:27:00, 6:27:00,9696,24,,0,0,,1
9999991,6:27:00, 6:27:00,9726,25,,0,0,,1
9999991,6:28:00, 6:28:00,9727,26,,0,0,,1
9999991,6:29:00, 6:29:00,9702,27,,0,0,,1
9999991,6:30:00, 6:30:00,9703,28,,0,0,,1
9999991,6:31:00, 6:31:00,9705,29,,0,0,,1
9999991,6:32:00, 6:32:00,9707,30,,0,0,,1
9999991,6:33:00, 6:33:00,9708,31,,0,0,,1
9999991,6:34:00, 6:34:00,9712,32,,0,0,,1
9999991,6:35:00, 6:35:00,9713,33,,0,0,,1
9999991,6:36:00, 6:36:00,15987,34,,0,0,,1
9999991,6:37:00, 6:37:00,15988,35,,0,0,,1
9999991,6:38:00, 6:38:00,15989,36,,0,0,,1
9999991,6:38:30, 6:38:30,15990,37,,0,0,,1
9999991,6:39:00, 6:39:00,15991,38,,0,0,,1
9999991,6:41:00, 6:41:00,15992,39,,0,0,,1
9999991,6:41:30, 6:41:30,16002,40,,0,0,,1
9999991,6:44:00, 6:44:00,15996,41,,0,0,,1
9999991,6:45:00, 6:45:00,15997,42,,0,0,,1
9999991,6:46:00, 6:46:00,15890,43,,0,0,,1
9999991,6:46:20, 6:46:20,17973,44,,0,0,,1
9999991,6:46:40, 6:46:40,17808,45,,0,0,,1
9999991,6:47:00, 6:47:00,17383,46,,0,0,,1
9999991,6:49:00, 6:49:00,15515,47,,0,0,,1
9999991,6:50:00, 6:50:00,15518,48,,0,0,,1
```

Figure 24: Example "Stop times.txt" GTFS file developed by the Research Team for the prospective system redesign

The most tedious and error-prone part of coding the GTFS of the prospective system redesign is the Stop times.txt file. For instance, an incorrect digit in a line of code could mean the entire set of GTFS files cannot be interpreted by STOPS, or that STOPS interprets an unreasonably fast or slow bus operating speed that is not what is intended.

Therefore, in order to error-check the GTFS coding, the Research Team made us of the "GTFSed" GTFS editor application[7] which can visually present the operations specified by a set of GTFS files (see Figure 9). Coding errors were much more easily identifiable in the visualization than by reviewing the GTFS text files themselves.

[7] The GTFS editor application is available on request from the Federal Transit Administration; interested readers are directed to contact information at: https://www.transit.dot.gov/funding/grant-programs/capital-investments/stops

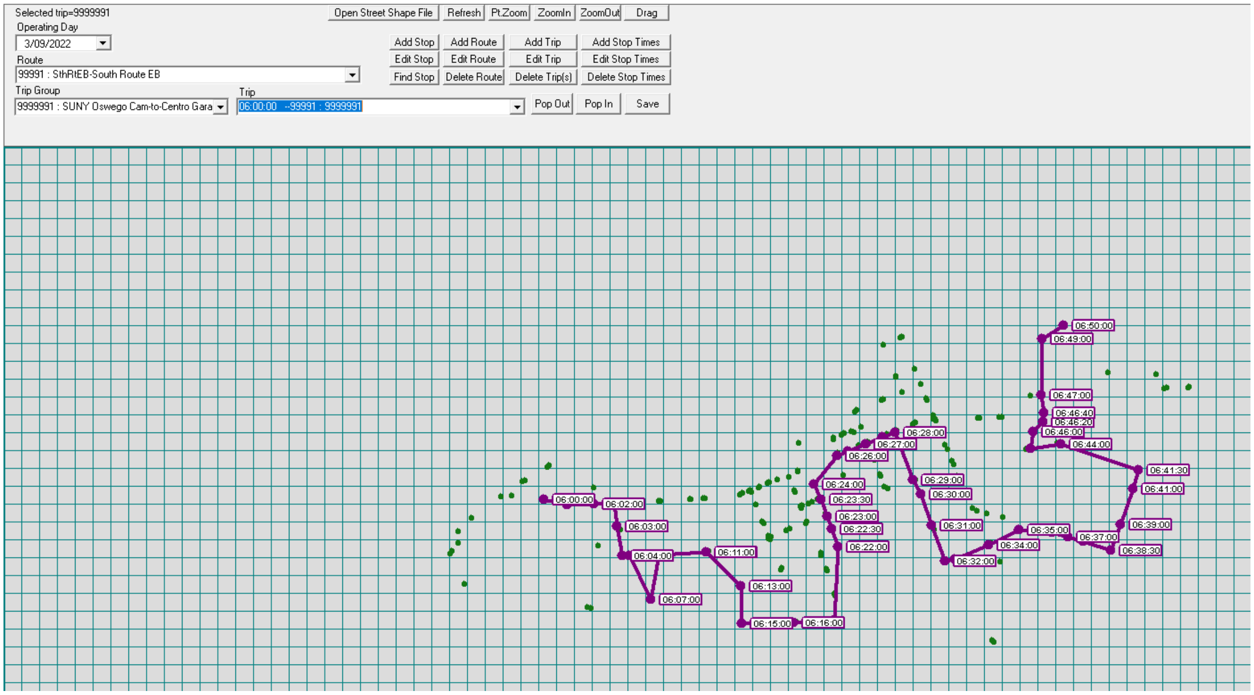


Figure 25: Example view of "GTFSed" editor, used to visualize a set of GTFS files. The prospective South route is shown in this image. All bus stops in the GTFS files are shown as green dots, and the stop-to-stop trajectory of the eastbound South route is shown in purple.

Ridership-Forecasting Input #2: Data on Number of Riders in Existing Conditions

The STOPS ridership-forecasting software is flexible to allow the user to input various types of data describing existing ridership of the current bus system. This flexibility allows STOPS to be used in situations where only very limited ridership data is available (i.e., total daily [weekday] system ridership), all the way to situations where daily ridership data at the unit of each individual bus stop is available or a bus-rider survey is available that captures origin-destination patterns of bus riders. STOPS then uses this information in a calibration process where the coefficients in the equations of the ridership-forecasting model are adjusted to attempt to approximately match the input of observed ridership data.

In the case of this Pilot in Oswego, Centro's APC data provided a rich data source that was exploited for this purpose. The Research Team first considered instructing STOPS to calibrate to observed boardings at each stop location. However, this option was abandoned because STOPS is based on a geographic zone system which may simulate all of the bus riders from a given zone boarding at one stop and none at another bus stop very nearby.

Therefore, the APC data were summarized into five geographic districts, and the calibration targets were the number of boardings within each district. The APC data were also averaged over weekdays in April so that ridership is modeled when SUNY Oswego is in regular session (during the spring semester).

Replica data of transit journeys can be used as targets to be used in model calibration, but in the case of Oswego the number of transit journeys within the synthetic Replica data was too low to be useful. Therefore Replica data were not used for this purpose, despite the Research Team having successfully used Replica data for model-calibration purposes in other assignments elsewhere.

Ridership-Forecasting Input #3: Geographic Zone System

The STOPS ridership-forecasting software has an automated process to generate the population and employment that it uses to forecast ridership, drawing on Census Transportation Planning Package (CTPP) and Journey-to-Work (JTW) datasets.

STOPS generally uses Block Groups as the geographic zones for representing trip origin and destination locations. In larger metropolitan regions this unit of geography works well and keeps computing requirements manageable. However, in a small region such as Oswego, the Research Team found that the Block Group level of geography was too coarse to generate reliable forecasts for the prospective bus system redesign.

Therefore, in consultation with the Federal Transit Administration (FTA; STOPS' sponsor), the Research Team specified Census Blocks (which are smaller than Block Groups) as the unit of analysis, and instructed STOPS to use them. This custom unit of geography was then found to yield more reliable current-day ridership estimates, and thus to increase confidence in the ability of STOPS to generate forecasts.

Additionally, STOPS is designed to use the zone system from an MPO's travel demand model (as well as the zone-to-zone travel times by automobile). This presented a unique challenge in this Pilot, because while Centro's main operations in Syracuse are within an MPO's boundaries (the Syracuse Metropolitan Transportation Council), Oswego is under the urban-population threshold to trigger a requirement for an MPO and is therefore not within an MPO's jurisdiction. To address this issue, the Research Team used Census Blocks in the City of Oswego (and Census Block Groups in the rest of Oswego County) as the geographic zone system, and generated zone-to-zone automobile travel times using an assumption of a representative 30 mph driving speed. This involved data processing within GIS (ArcMap) as well as spreadsheet software (Excel).

Software Tool for Ridership Forecasting: *STOPS*

The FTA has invested in development of the Simplified Trips-on-Project Software (STOPS) application, which is an open-access application that is *"…designed to quickly produce plausible* *forecasts of transit project ridership using readily available census data, transit ridership and schedule information."*

STOPS is open-access (the application; not the source code), free-of-charge and accessible from FTA's website[8], however users are advised that the website is not always maintained up-to-date, thus it is advisable to check with FTA prior to beginning a STOPS assignment to confirm that the latest version of STOPS, documentation files, and associated CTPP datafiles will be used. STOPS training courses are regularly offered through the National Transit Institute[9].

STOPS was designed to generate ridership forecasts for major capital investments in transit systems, which typically are found in the nation's large metropolitan regions. FTA confirmed to the Research Team that the application of STOPS in Oswego is towards the smaller scale of known STOPS applications nationally, and thus provides valuable insight into how STOPS works in small-scale cities with small transit systems.

An issue with STOPS is that the current version (v.2.5) uses CTPP/JTW data that is derived from the 2006-2010 American Community Survey datasets. The fact that this data is dated is not a major issue in regions where the spatial distribution of population and employment have not changed much in recent years, because STOPS self-calibrates to the newer ridership data that the user specifies. However, in fast-growing regions, the use of dated CTPP/JTW data is potentially a larger issue.

STOPS employs a state-of-the-art mode-choice model as its engine, which as noted above has its coefficients automatically calibrated to approximately match the user's input of observed ridership data.

STOPS produces, as its key output, forecasted daily ridership for a representative weekday, however the process through with STOPS generates this forecast is important. STOPS generates Home-Based-Work travel by simulating competition between automobile and transit usage in the 7:00 – 9:00 AM period, and all other travel by simulating the noon – 2:00 PM period. It then factors up the ridership generated during these two time periods to represent daily ridership, but STOPS does not consider the quality of transit service outside of these two time windows. Therefore STOPS will not produce estimates of the ridership impacts of a modification that expands a transit agency's hours-of-service later into the evening hours, and impacts in this circumstance would need to be generated by another approach.

The main STOPS user interface is shown in Figure 10. The STOPS application is powerful in its technical capabilities, however does not have intuitive and graphical attributes that are similar to the Replica online portal.

---

8 https://www.transit.dot.gov/funding/grant-programs/capital-investments/stops

9 https://www.ntionline.com/rideship-forecasting-with-stops-for-transit-project-planning/

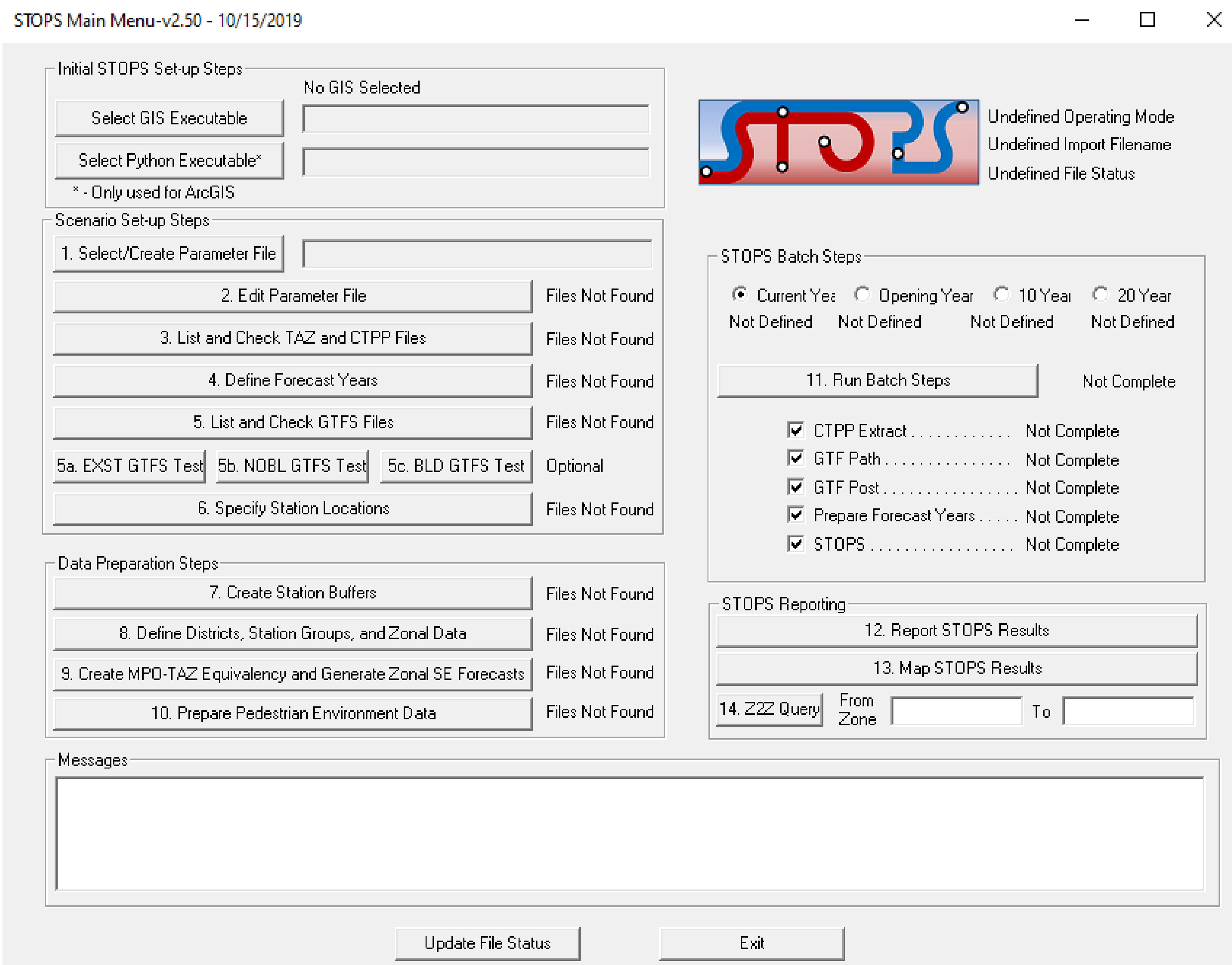


Figure 26: STOPS' main user interface

Calibration of STOPS in Oswego

'Calibration' refers to ensuring that the model reasonably simulates existing conditions before it is applied to simulate future conditions).

As noted above, district-level boardings were summarized from Centro's APC datasets and used to characterize existing bus ridership for model calibration purposes.

Additionally, it was found that including the City Routes (i.e., the routes of direct interest), the SUNY-campus routes targeted at university students on the campus (and its immediate vicinity), and Centro's two longer-distance routes that connect to Mexico/Fulton/Phoenix/Syracuse led to model-calibration problems. The issue encountered was that including these other routes led to STOPS placing less emphasis on matching existing ridership on the City Routes, and thus poorer fit between observed and simulated ridership on the City Routes.

The solution to this issue was to edit the GTFS files so that only the City Routes were simulated using STOPS.

Figure 11 shows the location of the districts that the Research Team defined for model-calibration purposes, and Table 1 shows the results of the model calibration process. Table 1 shows the raw text-file output from STOPS, which makes clear that the outputs are not in the most user-friendly and intuitive of formats, particularly for non-technical staff.

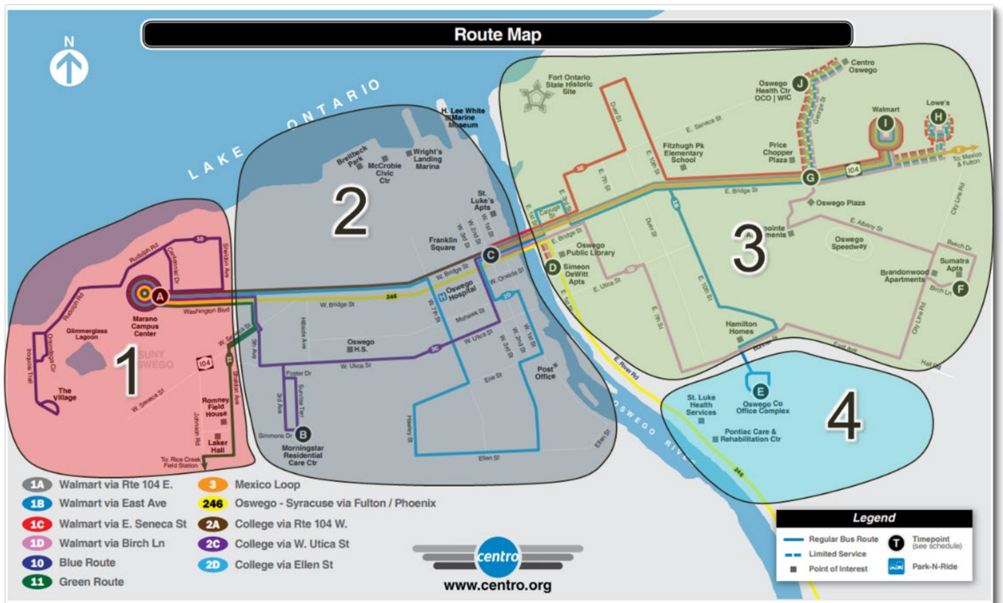


Figure 27: Map of districts used in STOPS for model-calibration purposes. District #5 (not shown) encompasses all geography not within one of Districts #1-4.

Scenario 1: Y2022 EXISTING
Raw linked transit trips: 15.00
Raw unlinked transit trips: 15.08
Target unlinked transit trips: 442.00
Regional calibration: 29.30

	Destination Group							
Origin Group	1	2	3	4	5	TOTAL	GOAL	COUNT
1- :	0	122	41	0	0	163	205	205
2- :	122	16	26	0	0	164	122	122
3- :	41	26	48	0	0	115	116	116
4- :	0	0	0	0	0	0	0	-
5- :	0	0	0	0	0	0	0	-
TOTAL :	163	164	115	0	0	442	-	-
GOAL :	205	122	116	0	0	-	443	-
COUNT :	205	122	116	-	-	-	-	443

Table 1: Calibration results as output by STOPS

Generation of Ridership Forecast

When the Research Team was satisfied with the calibration of the STOPS model to the district-level pattern of existing-conditions ridership, STOPS was then run using the GTFS files of the prospective system redesign to generate ridership forecasts.

The main outputs of STOPS are recorded in a large text file that is broken down into many tables (see Figure 12). The number of rows in the output text file can easily exceed 10,000, thus the user must be comfortable working with large text files to successfully navigate and interpret the STOPS outputs.

```
Program STOPS - FTA Simplified Trips-on-Project Software                                                                   Page     1
Version: STOPS-v2.50 - 07/17/2019                                                                                       10/11/2022
Run:    Calibration 1                                                                                                     16:17:09
System: Oswego
TABLE OF CONTENTS
===================================================================================================================================


 SECTION 1: SUMMARY OF KEY INPUTS
   Run Parmameters:                                1.01
   Station Listing:                                1.02
   Input Route Count and Group Information:        1.03
   Assignment of GTFS Route_IDs to Route Groups:   1.04
 -----------------------------------------------------------------------------------------------
 SECTION 2: SUMMARY OF EXISTING SCENARIO RESULTS BEFORE STATION GROUP CALIBRATION
   Initial Calibration Statistics:                  2.01
   Iter. 1 Station Groups by Production District:   2.02
   Iter. 1 Station Groups by Attraction District:   2.03
   Station Group Boardings Prior to Adjustment:                         2.04
   Station Group Boarding Factor for Application to Later Iterations:   2.05
   Listing of Stop_ids not found in station file:                       2.06
   Type 10/12 Group-Level Calibration Summary-Stations/Stops-SKIPPED-   2.07
   Type 11/12 Group-Level Calibration Summary-Station/Route Summary:    2.08
 -----------------------------------------------------------------------------------------------
 SECTION 3: SUMMARY OF STATION GROUP CALIBRATION RESULTS
Station Group-Station Group Accumulators
   Station Group to Station Group Ridership-Existing:                         3.01
   Station Group to Station Group Ridership-No-Build:                         3.02
   Station Group to Station Group Ridership-Build:                            3.03

Station-Station Accumulators
   Station Group to Station Group Build Trips:                                3.04
   Station Group to Station Group Build Trips (Added by Station Factors):     3.05
   Station Group to Station Group Project Trips:                              3.06
   Station Group to Station Group Project Trips (Added by Station Factors):   3.07
 -----------------------------------------------------------------------------------------------
 SECTION 4: SUMMARY OF PROJECT RESULTS
   Weekday Linked District-to-District Transit Trips, Build, All Trips:       4.01
   Weekday Incremental Linked Dist-to-Dist Transit Trips, Build, All Trips:   4.02
   Weekday Linked District-to-District Project Trips, Build, All Trips:       4.03
   Weekday Unlinked Station-to-Station Project Trips, Build, All Trips:       4.04
 -----------------------------------------------------------------------------------------------
 SECTION 5: SUMMARY OF PROJECT RESULTS FOR TRIPS ON FIXED GUIDEWAY (FG) MODES
   Weekday Linked District-to-District Transit Trips, Build, FG Trips:        5.01
   Weekday Incremental Linked Dist-to-Dist Transit Trips, Build, FG Trips:    5.02
   Weekday Linked District-to-District Project Trips, Build, FG Trips:        5.03
   Weekday Unlinked Station-to-Station Project Trips, Build, FG Trips:        5.04
 -----------------------------------------------------------------------------------------------
 SECTION 6: SUMMARY OF PROJECT RESULTS FOR TRIPS MADE BY 0-CAR HOUSEHOLDS
   Weekday Linked District-to-District Transit Trips, Build, All Trips by 0 Car HH:        6.01
   Weekday Incremental Linked Dist-to-Dist Transit Trips, Build, All Trips by 0 Car HH:    6.02
   Weekday Linked District-to-District Project Trips, Build, All Trips by 0 Car HH:        6.03
   Weekday Unlinked Station-to-Station Project Trips, Build, All Trips by 0 Car HH:        6.04
 -----------------------------------------------------------------------------------------------
 SECTION 7: SUMMARY OF PROJECT RESULTS FOR TRIPS ON FIXED GUIDEWAY (FG) MODES BY 0-CAR HH
   Weekday Linked District-to-District Transit Trips, Build, FG Trips by 0 Car HH:         7.01
```

Figure 28: Table of Contents at top of STOPS' main output text file

STOPS reports a rich variety of data in the large text file, which is available for the user to interpret to understand the ridership impacts of a prospective service change. Table 2 is one of the main output tables, which presents ridership for each route (both existing and proposed) in the system. The headline result is highlighted in the bottom row: ridership is forecasted to decrease from 442 trips/day to 432 trips/day from the service redesign, a decrease of ~2.5%. However, another caveat is that STOPS generates only weekday ridership forecasts, so does not output information about how weekend ridership is impacted by a service modification.

STOPS outputs can also be exchanged from the STOPS output text file into other software for visualization purposes. To generate Figure 13 below, we matched the bus-stop locations in the STOPS output file with their longitude/latitude coordinates from the GTFS file, to allow the locations of the forecasted boardings to be visualized. STOPS is forecasting that the SUNY campus and downtown Oswego areas will generate much of the ridership under the prospective system redesign.

***** AVG WEEKDAY ROUTE UTILIZATION ZONE BY (PRODUCTION-END) ACCESS TYPE ************
Comparison of Route Boardings by Scenario and Zone (Production-End) Mode-of-Access
Total Transit Trips

			Y2022 EXISTING				Y2022 NO-BUILD				Y2022 BUILD			
Route_ID	--Route Name	Count	WLK	KNR	PNR	ALL	WLK	KNR	PNR	ALL	WLK	KNR	PNR	ALL
18354	--Ful 4-Fulton - East	0	0	0	0	0	0	0	0	0	0	0	0	0
18355	--Ful 5-Fulton - West	0	0	0	0	0	0	0	0	0	0	0	0	0
18356	--Mex 3-Mexico - Fulton	0	0	0	0	0	0	0	0	0	0	0	0	0
18358	--Osw10-SUNY Oswego Blue Route	0	0	0	0	0	0	0	0	0	0	0	0	0
18359	--Osw11-SUNY Oswego Green Rout	0	0	0	0	0	0	0	0	0	0	0	0	0
18360	--Osw1A-Walmart via 104	50	8	1	0	9	8	1	0	9	0	0	0	0
18361	--Osw1B-Walmart - Hamilton Hom	31	63	6	0	69	63	6	0	69	0	0	0	0
18362	--Osw1C-Walmart via Seneca Str	15	59	1	0	59	59	1	0	59	0	0	0	0
18363	--Osw1D-Walmart via Brandonwoo	63	25	1	0	25	25	1	0	25	0	0	0	0
18364	--Osw2A-College via 104	164	61	2	0	63	61	2	0	63	0	0	0	0
18365	--Osw2C-College via West Utica	21	134	15	0	149	134	15	0	149	0	0	0	0
18366	--Osw2D-College via Ellen St	98	57	10	0	67	57	10	0	67	0	0	0	0
18367	--Osw46-Oswego - Syracuse	0	0	0	0	0	0	0	0	0	0	0	0	0
99991	--SthRtEB-South Route EB	0	0	0	0	0	0	0	0	0	26	1	0	27
99992	--SthRtWB-South Route WB	0	0	0	0	0	0	0	0	0	19	6	0	25
Test1	--Test1-SUNY to Walmart	0	0	0	0	0	0	0	0	0	26	1	0	26
Test2	--Test2-Walmart to SUNY	0	0	0	0	0	0	0	0	0	330	23	0	353
Total		442	407	35	0	442	407	35	0	442	401	31	0	432

Table 2: STOPS output showing route-level weekday daily ridership forecasts

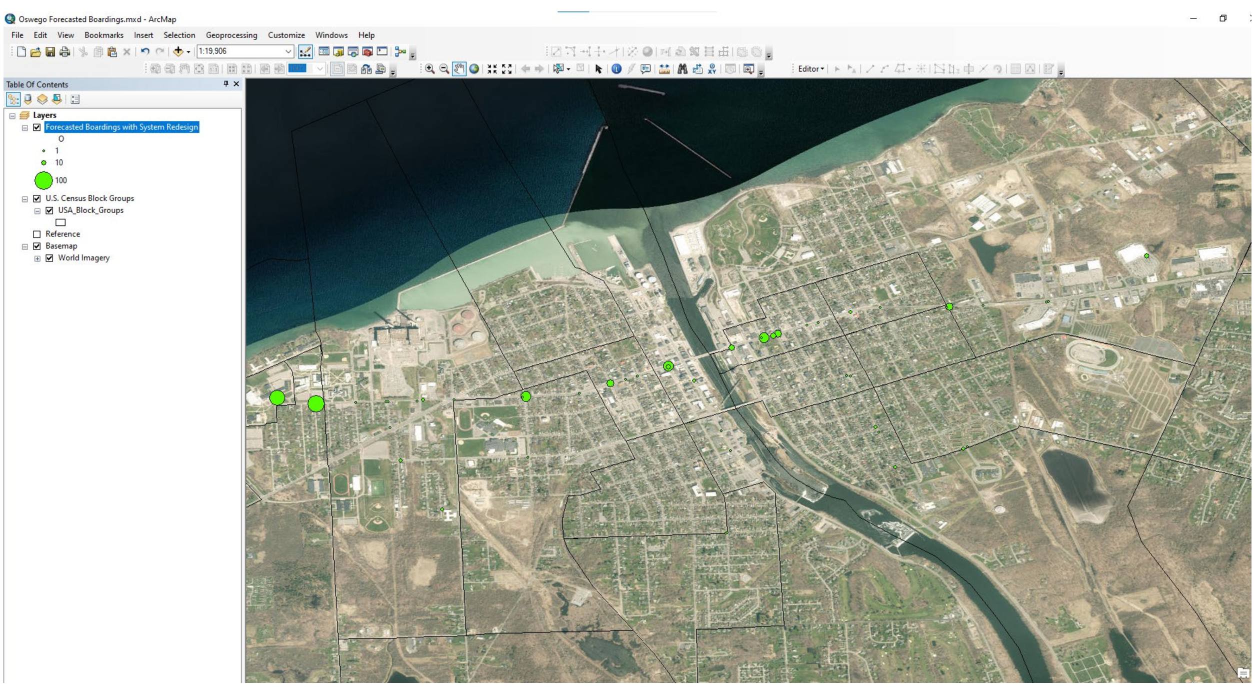


Figure 29: Visualization of STOPS ridership forecast, by stop location

Finally, in its output STOPS also reports a summary of how it interpreted the transit services specified in the GTFS files (see Table 3). This is valuable for error-checking purposes; if the amount of bus trips, revenue-miles, or revenue-hours of service under existing conditions or the proposed service modification as presented in this table are illogical, this is a red flag for the user to go back and review the GTFS files to be sure they are correctly specifying the desired transit service patterns.

Version: STOPS-v2.50 - 10/10/2019 10/11/2022
Run: Oswego Baseline 8:29:34
System: Centro
Table 10.03

***** PEAK TRIPS, MILES AND HOURS BY ROUTE ************
Comparison of Weekday Operating Statistics by Scenario
PEAK=any trip serving one or more stops between 5:00 am and 9:15 am

		Y2022 EXISTING			Y2022 NO-BUILD			Y2022 BUILD		
Route_ID	--Route Name	#Trips	Miles	Hours	#Trips	Miles	Hours	#Trips	Miles	Hours
18360	--Osw1A-Walmart via 104	3	5.54	0.47	3	5.54	0.47	0	0.00	0.00
18361	--Osw1B-Walmart - Hamilton Hom	3	12.32	1.00	3	12.32	1.00	0	0.00	0.00
18362	--Osw1C-Walmart via Seneca Str	1	2.97	0.30	1	2.97	0.30	0	0.00	0.00
18363	--Osw1D-Walmart via Brandonwoo	4	18.95	1.82	4	18.95	1.82	0	0.00	0.00
18364	--Osw2A-College via 104	7	10.08	1.17	7	10.08	1.17	0	0.00	0.00
18365	--Osw2C-College via West Utica	3	7.30	1.00	3	7.30	1.00	0	0.00	0.00
18366	--Osw2D-College via Ellen St	4	14.64	1.33	4	14.64	1.33	0	0.00	0.00
18367	--Osw46-Oswego - Syracuse	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
99991	--SthRtEB-South Route EB	0	0.00	0.00	0	0.00	0.00	2	14.90	1.67
99992	--SthRtWB-South Route WB	0	0.00	0.00	0	0.00	0.00	2	14.57	1.67
Test1	--Test1-SUNY to Walmart	0	0.00	0.00	0	0.00	0.00	6	19.73	2.00
Test2	--Test2-Walmart to SUNY	0	0.00	0.00	0	0.00	0.00	6	[illegible]	2.20
	Total	25	71.80	7.08	25	71.80	7.08	16	68.92	7.53

Program STOPS - FTA Simplified Trips-on-Project Software Page 50
Version: STOPS-v2.50 - 10/10/2019 10/11/2022
Run: Oswego Baseline 8:29:34
System: Centro
Table 10.04

***** OFF-PEAK TRIPS, MILES AND HOURS BY ROUTE ************
Comparison of Weekday Operating Statistics by Scenario
OFF-PEAK=any trip serving one or more stops between 10:00 am and 2:15 pm

		Y2022 EXISTING			Y2022 NO-BUILD			Y2022 BUILD		
Route_ID	--Route Name	#Trips	Miles	Hours	#Trips	Miles	Hours	#Trips	Miles	Hours
18360	--Osw1A-Walmart via 104	10	21.38	1.73	10	21.38	1.73	0	0.00	0.00
18361	--Osw1B-Walmart - Hamilton Hom	3	12.73	1.08	3	12.73	1.08	0	0.00	0.00
18362	--Osw1C-Walmart via Seneca Str	2	6.88	0.62	2	6.88	0.62	0	0.00	0.00
18363	--Osw1D-Walmart via Brandonwoo	4	19.82	1.88	4	19.82	1.88	0	0.00	0.00
18364	--Osw2A-College via 104	10	14.40	1.67	10	14.40	1.67	0	0.00	0.00
18365	--Osw2C-College via West Utica	5	12.15	1.67	5	12.15	1.67	0	0.00	0.00
18366	--Osw2D-College via Ellen St	6	21.85	2.00	6	21.85	2.00	0	0.00	0.00
18367	--Osw46-Oswego - Syracuse	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00
99991	--SthRtEB-South Route EB	0	0.00	0.00	0	0.00	0.00	3	22.36	2.50
99992	--SthRtWB-South Route WB	0	0.00	0.00	0	0.00	0.00	3	21.85	2.50
Test1	--Test1-SUNY to Walmart	0	0.00	0.00	0	0.00	0.00	9	29.60	3.00
Test2	--Test2-Walmart to SUNY	0	0.00	0.00	0	0.00	0.00	9	[illegible]	3.30
	Total	40	109.22	10.65	40	109.22	10.65	24	103.37	11.30

Table 3: STOPS output summarizing bus service under existing conditions and prospective system redesign

Sensitivity Tests

In addition to generating a ridership forecast for the proposed system redesign in Oswego, it is good practice to perform a series of sensitivity tests to ensure that the STOPS outputs behave logically when the inputs are varied.

Table 4 shows the results of the series of sensitivity tests which were performed by the Research Team. Headline findings include:

- In all cases, the outputs are reasonable and logical responses to the variation in inputs – lending credibility to the ridership forecasts generated by STOPS.

- If Centro were to reduce service from three to two buses in concurrent operation, and only provide service on the High-frequency route (i.e., no South route), this would still allow Centro to serve 90% of the ridership that is projects with the three buses in operation (396 v. 432 riders/day).
- Doubling the number of buses on the High-frequency route from two to four buses, which would allow 15-minute headways on this route, would only result in a ~3% increase in ridership.

It is noteworthy that modifying the service frequencies for sensitivity-testing purposes was straightforward, accomplished simply by changing a single line of code in the GTFS *Frequencies.txt* input file. In contrast, a scenario in which bus speeds needed to be modified, such as testing a transit signal priority strategy that would increase average bus speeds, would require changing large numbers of lines of code, and hence would be much more tedious and time-consuming.

Sensitivity test	**Forecast weekday daily ridership (boardings)**
Only the High-frequency route	396
15-minute (rather than 30-minute) frequency on the High-frequency route	446
2-hour frequency on the High-frequency route	241
Only the South route	155
1-hour (rather than 2-hour) frequency on the South route	325

Table 4: Results of sensitivity tests

Modeling Ridership

Link to Workflow

1. Gap Analysis

As described in the previous sections, the gap analysis was conducted to understand the transit service provision and transit demand within the City of Troy. To understand how transit service aligns with the demand and identify the service gaps, the team conducted a service provision analysis using ArcGIS and compared it to the Transit Propensity Index that was developed by CDTA. The technical steps are described below.

CDTA TPI

Identifying gaps between transit supply and demand required a clear understanding of the demand for public transit. In this pilot, CDTA's TPI was used to understand the transit demand in Troy. The TPI integrates multiple demographic and socioeconomic factors at the census block group level and applies weights to each of the factors. The factors include:

- Total population per square mile (ACS 2019, 5-Year Estimates)
- Population over 65+ per square mile (ACS 2019, 5-Year Estimates)
- Population below the poverty line per square mile (ACS 2019, 5-Year Estimates)
- Households with more workers than vehicles per mile (ACS 2019, 5-Year Estimates)
- 4-way intersection per previous TPI
- Mall / shopping plaza square footage per mile (Google maps)
- Students per square mile (Book of lists/Google)
- Hospital beds per square mile (American Hospital Directory)
- Jobs per square mile (LEHD 2019)
- Low-paying jobs per square mile (LEHD 2019)

Service Provision Analysis

Twelve CDTA bus routes served the City of Troy in 2019 and were included in this analysis:

- Route 22 Albany - Troy via Watervliet
- Route 80 – Troy/Fifth Avenue
- Route 85 – Fifth Avenue
- 87 – Hoosick St – RPI
- 182 – Troy – Albany via Cohoes and Latham
- 224 – Albany – Troy via I-90
- 286 – Troy/Wynantskill
- 289 – Griswold Heights – St. Mary's Hospital
- 370 – Troy/Schenectady
- 522 – Cohoes/Troy/Albany Express
- 809 – South Albany – Troy Shuttle
- 815 – Troy Shopping Bus

The bus stop locations served by the analyzed routes were imported, as a CSV file, into ArcGIS and visualized using the Display XY Data tool. Several shapefiles were used as a base in ArcGIS including CDTA bus routes (system-wide); Census block groups for Albany, Montgomery, Rensselaer, Saratoga, and Schenectady counties; the boundary for the City of Troy, and water bodies.

First, ¼-mile buffers were mapped around each bus stop to estimate the walkshed and the service area for each route. By only creating buffers around the stops this excluded no-stop segments of routes such as highways.
Next, the weekly scheduled revenue hours (weekdays, Saturdays, and Sundays) were calculated for each route using the 2019 service provision dataset.

Equation 1: Weekly Scheduled Revenue Hours Calculation

Weekly Scheduled Revenue Hours =
*(Number of weekdays*weekday revenue hours) + Saturday revenue hours + Sunday revenue hours*

The weekly scheduled revenue hours for each route are shown in **Table 7**. The revenue hours were assigned to the buffer area of each route in ArcGIS to reflect the service provided, as shown in **Figure 15.** To determine the distribution of service, revenue hours per square mile of the buffered area were calculated.

Figure 30 - Service hours Allocation per Buffer Area in ArcGIS

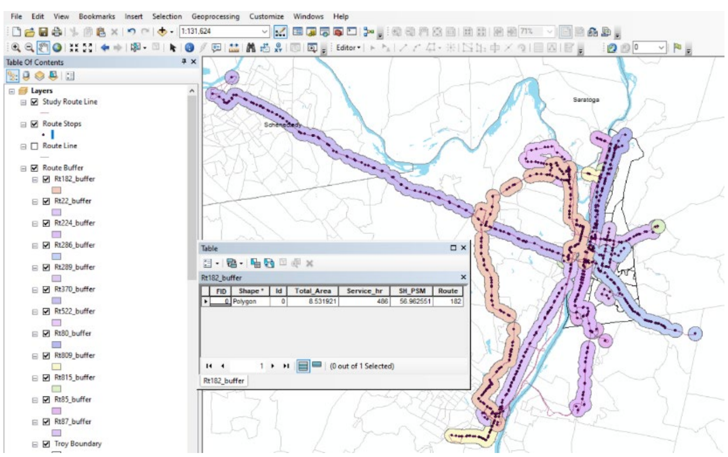


Table 7. Total Service Hours per route

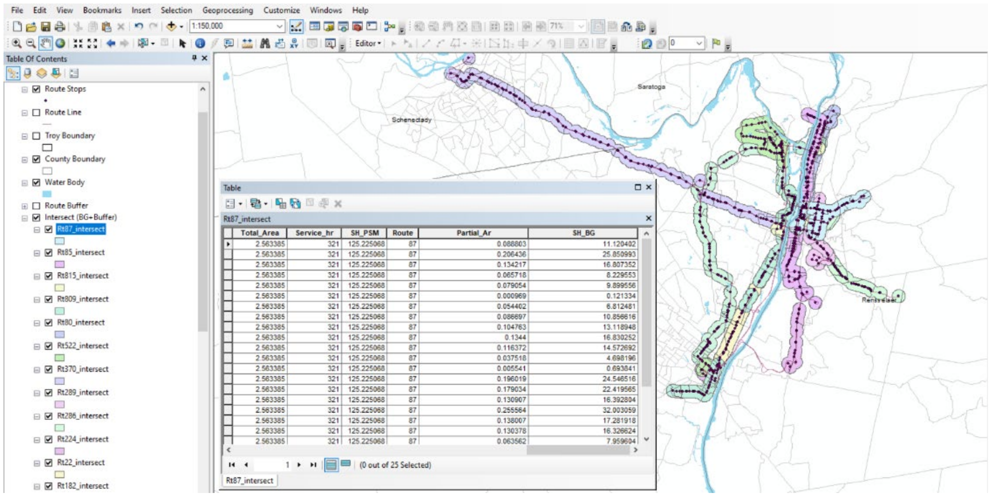
Route	Scheduled Revenue Hours			
	Weekday	Saturday	Sunday	Total
22	128.0	68.8	49.7	759
80	40.4	25.8	10.7	239
85	74.7	48.5	26.2	448
87	52.2	38.3	21.7	321
182	82.9	42.9	28.0	486
224	62.1	0	0	311
286	18.1	15.6	0	107
289	16.0	16.0	0	96
370	79.6	55.1	31.8	485
522	9.0	0	0	45
809	1.9	0	0	10
815	2.5	0	0	13

To calculate the portion of the buffer that falls within each census block group, the Intersect Tool was used with the input features as ***individual*** route buffers and census block groups, as shown in **Figure 16**. The partial area was calculated by creating a new column for each buffered zone. Within the same attribute table, another new column was created to calculate service hours in the block group by running a field calculator with the following equation:

Equation 2: Service hours per Census Block Group

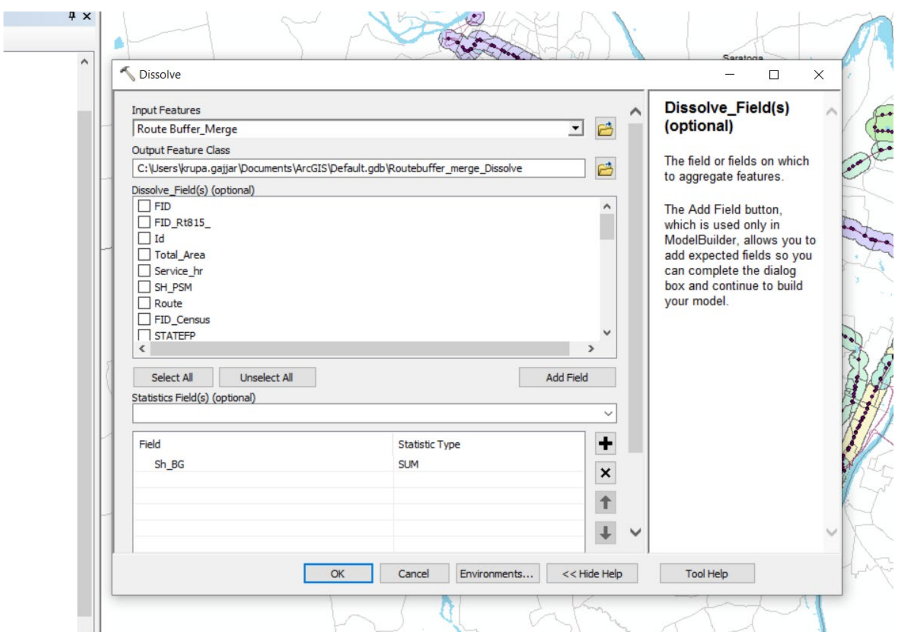
*Service hours in the block group = partial area * service hours per sq mile*

Figure 31 - Calculation of Service Hours per Census Block Groups in ArcGIS



To determine the combined service hours of all buffered areas within each census block group, the ArcGIS Merge tool was used on the route buffer. The Merge output layer was later dissolved by block group ID (using the Dissolve Tool) and summed based on the service hours in the block group column. A snapshot of the ArcGIS dissolve tool is shown in **Figure 17**.

Figure 32 - ArcGIS Dissolve Tool was used to determine the total buffer area within each census block group



Once the dissolve tool was run successfully, a new total area of the buffer was calculated. Since the Hudson River runs along the boundary of the City of Troy, some of the buffer areas overlapped with the river. To calculate the appropriate land area, the water body was clipped from the buffer, and the area was subtracted from the route buffer area. The partial final buffer area within each census block group was determined using the following equation*:*

Equation 3. Percent of Buffer Area

*Portion of buffer area = final buffer area/census block group area * 100*

As the final step in the service provision analysis, to generate a comparable indicator of transit service provision throughout the city, census block group scores were calculated by multiplying the total service hours in the block group by the portion of the buffer area. Scores were classified into five equal categories.

Equation 4. Transit Service Provision Score

*Total Score = percent of buffer area * service hours in the block group*

Gap Analysis

The gap analysis was conducted based on the results from the transit supply analysis and CDTA's TPI. It categorized the census block groups in Troy based on the differences between supply and demand and identified three sets of areas: (1) areas of oversupply, (2) areas where demand exceeds service supply, and (3) areas where the service level matches the demand. The gap analysis was conducted using ArcGIS and Microsoft Excel.

Using ArcGIS, the attribute table of the 'total score' for the Transit Supply and CDTA TPI layer were exported individually to analyze further in Excel. Taking the lower and upper boundaries of the five classes, each census block group was assigned a demand score and a supply score, ranked from one to five, as shown in **Figure 18.**

Figure 33 - Excel Workbook Showing Categorization

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	FID	STATEFP	COUNTYFP	TRACTCE	BLKGRPCE	GEOID	NAMELSAD	TotalScore	Class	Service Level		**Lower boundary**	**Upper Boundary**	**Quantile**	**Level of Service**	**Count**
2	5467	36	83	41300	2	360830413002	Block Group 2	0.283845863	1	Very Low		0	8.1	1	Very Low	9
3	6431	36	83	41400	2	360830414002	Block Group 2	0.464872335	1	Very Low		8.2	13.2	2	Low	10
4	5596	36	83	41200	1	360830412001	Block Group 1	2.536100646	1	Very Low		13.3	20.7	3	Medium	10
5	5597	36	83	41200	2	360830412002	Block Group 2	3.380207664	1	Very Low		20.8	31.1	4	High	10
6	8057	36	83	40500	1	360830405001	Block Group 1	3.794183957	1	Very Low		31.2	105.2	5	Very High	9
7	5598	36	83	41200	3	360830412003	Block Group 3	5.586890092	1	Very Low					**TOTAL**	**48**
8	5591	36	83	41100	3	360830411003	Block Group 3	6.191255581	1	Very Low						
9	5577	36	83	40900	4	360830409004	Block Group 4	7.495280501	1	Very Low						
10	14892	36	83	40200	6	360830402006	Block Group 6	8.089985741	1	Very Low						
11	5594	36	83	41100	1	360830411001	Block Group 1	8.645882493	2	Low						
12	14891	36	83	40200	3	360830402003	Block Group 3	8.879917903	2	Low						
13	5518	36	83	40100	3	360830401003	Block Group 3	10.06700167	2	Low						
14	5520	36	83	40200	5	360830402005	Block Group 5	10.24238491	2	Low						
15	5469	36	83	40100	5	360830401005	Block Group 5	10.47870588	2	Low						
16	5521	36	83	40300	1	360830403001	Block Group 1	12.00274956	2	Low						
17	8049	36	83	40200	2	360830402002	Block Group 2	12.07139602	2	Low						
18	13960	36	83	40900	3	360830409003	Block Group 3	12.4492254	2	Low						
19	5599	36	83	41200	4	360830412004	Block Group 4	12.89789266	2	Low						
20	5578	36	83	40900	2	360830409002	Block Group 2	13.1123187	2	Low						
21	2986	36	83	41400	1	360830414001	Block Group 1	13.42127131	3	Medium						
22	5595	36	83	41100	2	360830411002	Block Group 2	14.41141227	3	Medium						
23	5504	36	83	40600	2	360830406002	Block Group 2	15.0200695	3	Medium						
24	5517	36	83	40100	2	360830401002	Block Group 2	15.1043701	3	Medium						

The supply and demand levels were then compared to identify the gaps. Gap values were calculated based on the difference between the supply and demand ranking for each block group**. Table 8** shows the category distribution. The gap values ranged from -4 and 4. If the gap value for a given census block group was between -4 and -2, it was categorized as *high supply*, with more transit supply than demand. If the gap value was between -1 and 1, it was categorized as a *match*, meaning that the block group had equal supply and demand. If the gap value was between 2 and 4, it was categorized as *high demand*, with more transit demand than supply. **Figure 19** shows the snapshot of the Excel analysis.

Table 8. Gap Analysis Category Distribution

Gap value	Gap Analysis Categorization
-4	GAP - high supply
-3	GAP - high supply
-2	GAP - high supply
-1	Match
0	Match
1	Match
2	GAP - high demand
3	GAP - high demand
4	GAP - high demand

Figure 34 - Excel Workbook Showing Categorization

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	STATEFP	COUNTYFI	TRACTCE	BLKGRPCE	GEOID	NAMELSAD	Demand_Level	Service_Level	Delta	Gap		Delta value	Gap Analysis Distribution	Number of Block Groups	Percentage
2	36	83	40900	4	360830409004	Block Group 4	5	1	4	GAP - high demand		-4	GAP - high supply	1	2%
3	36	83	41300	4	360830413004	Block Group 4	5	4	1	Match		-3	GAP - high supply	3	6%
4	36	83	40700	1	360830407001	Block Group 1	5	5	0	Match		-2	GAP - high supply	5	10%
5	36	83	40800	1	360830408001	Block Group 1	5	5	0	Match		-1	Match	6	13%
6	36	83	40600	1	360830406001	Block Group 1	5	3	2	GAP - high demand		0	Match	18	38%
7	36	83	40700	2	360830407002	Block Group 2	5	5	0	Match		1	Match	5	10%
8	36	83	40600	2	360830406002	Block Group 2	5	3	2	GAP - high demand		2	GAP - high demand	7	15%
9	36	83	40200	2	360830402002	Block Group 2	5	2	3	GAP - high demand		3	GAP - high demand	2	4%
10	36	83	40900	3	360830409003	Block Group 3	5	2	3	GAP - high demand		4	GAP - high demand	1	2%
11	36	83	40200	3	360830402003	Block Group 3	4	2	2	GAP - high demand				48	
12	36	83	40100	3	360830401003	Block Group 3	4	2	2	GAP - high demand					
13	36	83	40700	4	360830407004	Block Group 4	4	5	-1	Match					
14	36	83	40100	5	360830401005	Block Group 5	4	2	2	GAP - high demand					
15	36	83	40300	1	360830403001	Block Group 1	4	2	2	GAP - high demand					
16	36	83	40200	5	360830402005	Block Group 5	4	2	2	GAP - high demand					
17	36	83	40200	1	360830402001	Block Group 1	4	3	1	Match					
18	36	83	40600	3	360830406003	Block Group 3	4	4	0	Match					
19	36	83	41300	5	360830413005	Block Group 5	4	4	0	Match					
20	36	83	40500	2	360830405002	Block Group 2	4	4	0	Match					
21	36	83	40400	1	360830404001	Block Group 1	3	4	-1	Match					
22	36	83	40800	2	360830408002	Block Group 2	3	4	-1	Match					
23	36	83	40200	4	360830402004	Block Group 4	3	3	0	Match					

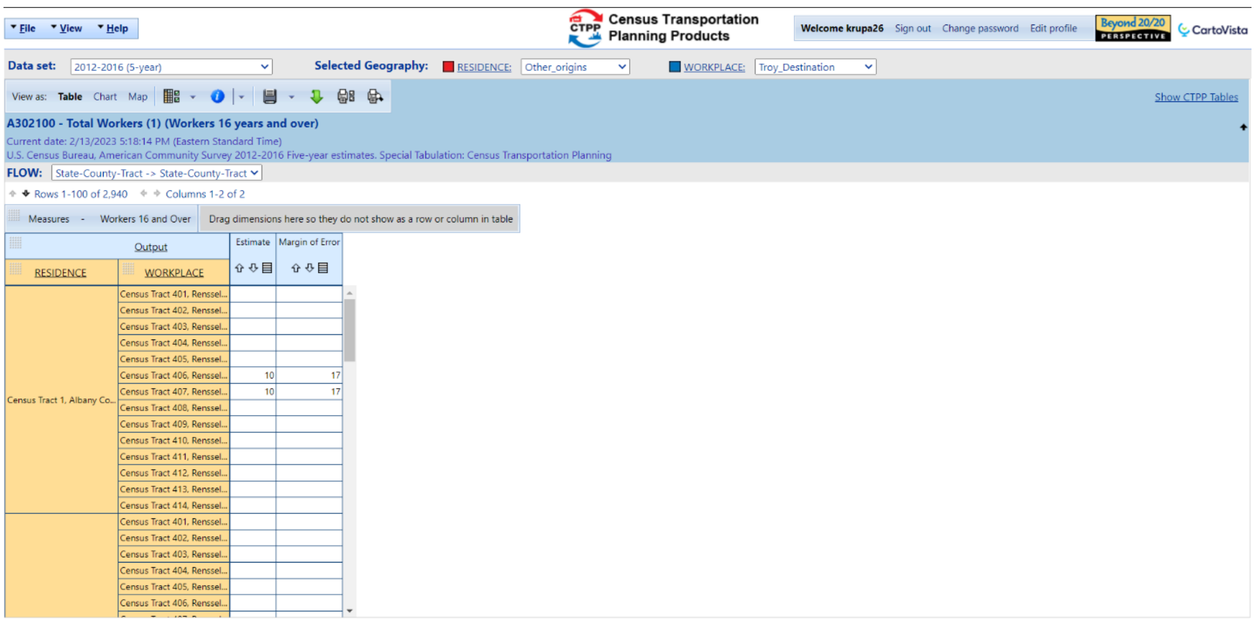
To visualize the gap analysis results, the CSV was imported into ArcGIS and joined to the Troy census block groups shapefile.

2. Origin-Destination Analysis

To determine the travel pattern of workers, an O/D analysis was conducted at the census tract level for all the workers over the age of 16 using CTPP data. Two datasets were gathered based on the place of residence and workplace, one for trips originating in Troy and one for trips destined for Troy.

CSV data tables and shapefiles were downloaded for Troy as the origin, 2019 Part 3: Flows; Table A302100 – Total Workers (1) (Workers 16 years and over), where selected geography was: residence = census tracts within City of Troy and workplace = all the census tracts within Albany, Schenectady, Saratoga, Rensselaer County. Similarly, for Troy as the destination, Table A302100 – Total Workers (1) (Workers 16 years and over) was used, where selected geography was: residence = all the census tracts within Albany, Schenectady, Saratoga, Rensselaer County, and workplace = census tracts within City of Troy. Then both the shapefiles were imported into ArcGIS for further analysis. **Figure 20** shows the snapshot of the CTPP website.

Figure 35 - Snapshot of CTPP website



In ArcGIS, the O/D pairs with Troy as the destination were visualized and broken into three categories *(51–100, 101-150, and over 151).* As the O/D data was analyzed to identify origin and destination hotspots for transit routes, O/D pairs with fewer than 50 trips were not considered in the analysis. The O/D map for all trips can in found in **Appendix B**. Similar steps were conducted to analyze the O/D pairs with Troy as the origin.

3. TBEST Ridership Analysis Tutorial

The TBEST Ridership Analysis utilizes the software's ridership model estimation tool to generate ridership values and demographics based on demographics, connectivity, surrounding parcels, and a variety of other data. The following are the steps required to generate the ridership analysis after the software has been installed and setup according to the TBEST User Manual.

Scenario Creation

In the software's "Explorer" tab, two scenarios are constructed by right clicking the "Scenarios" folder under the established Transit System. These two scenarios will be titled "Unmodified" and "Modified", but otherwise be setup the same with the default data sources. For this analysis, the Mean Annual Wage was set to $55,912, the average mean annual wage for all counties in the CDTA system at the time of this analysis. Annual Growth Rate was set to 0% as the scope of this analysis was not to forecast any data beyond the years of the datasets.

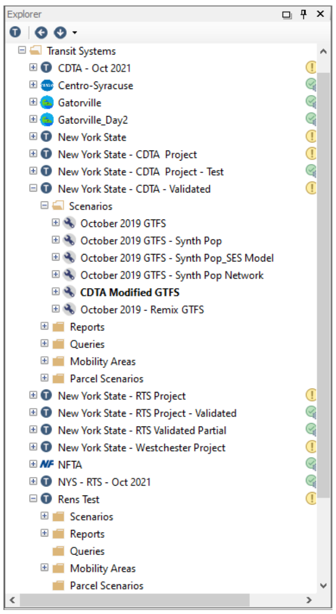
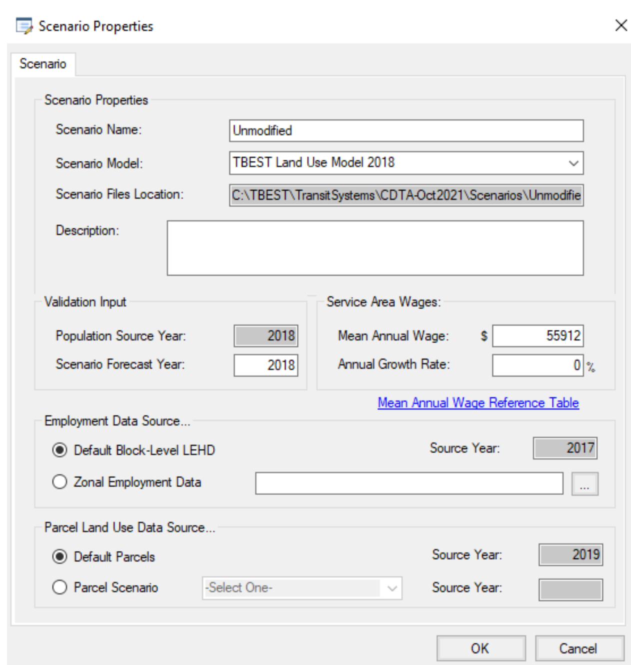


Figure 36 - Scenario Creation in TBEST

Next, both scenarios need to have their corresponding GTFS files imported. "GTFS Import/Export Tools" can be accessed by right clicking one of the newly created scenarios. Select "Import Routes from GTFS". For this analysis, the Unmodified GTFS file from Remix was selected for the "Unmodified" scenario and the GTFS that has had route restructuring was selected for the "Modified" scenario. All GTFS Service Periods (Weekday, Saturday, Sunday) were selected from both GTFS files and imported.



Right click and select "Set Socio-Economic Growth Rates…" and set all values to "0%" for both scenarios.

Editing and Running the Ridership Estimation Model

This pilot project explored the process of tweaking the model that the Ridership Estimation tool uses. This process is **not** recommended by the software developer as the model required a significant process to construct.

The first step in adjusting the TBEST model is to duplicate the original model to ensure that there is a back-up if reversions needed to be made.

In the “Explorer” panel, click the “+” button next to the “Models” folder to expand it and show all the models. Right click the “TBEST Land Use Model 2018” and click “Copy”. Enter the new model’s name as “Modified TBEST Model”.

Right click the “Modified TBEST Model” and click “Equation Coefficients” to adjust the coefficients of the model. Below is a list of changes made to the model during this pilot project:

Table 9. Altered Coefficients in TBEST Model

Equation	Category	Variable Description	Variable Key	AM PEAK	OFF PEAK	PM PEAK	NIGHT	SATURDAY	SUNDAY
Direct Boarding	Constant	Bus Constant	BUS-ONE	-6.242	-5.838	-6.555	-6.522	-2.424	-2.683
Direct Boarding	Employment	Share of Service Employment	SHARE_SERVEMP	0.25	0.25	0.25	0.25	0.25	0.25
Direct Boarding	Special Generators	Shopping Mall	MALLS	0.643	1.658	2.235	0	1.65	1.967
Direct Boarding	Special Generators	University	UNIVERSITY	2	2	2	2	2	2

With a new model available to use, update the model used by both scenarios by right clicking the scenario, selecting “Properties”, and choosing “Modified TBEST Model” from the “Scenario Model” drop-down menu. Click “Okay” to accept changes.

TBEST supports multiple scenario ridership estimation model runs which can be accessed by right clicking the “Scenarios” folder and clicking “Batch Model Run”. Check the box for both “Unmodified” and “Modified” scenarios and all time periods before clicking “Run Model”.

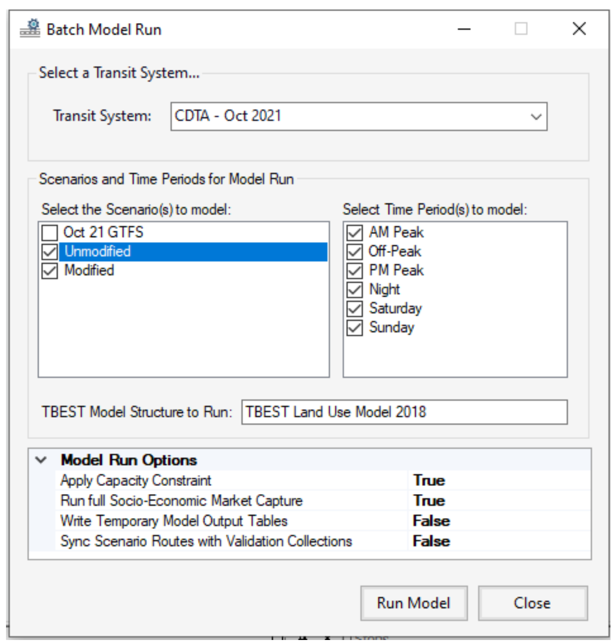


Figure 37 - TBEST Batch Model Run Window

Exploring and Post-Processing the Ridership Estimations

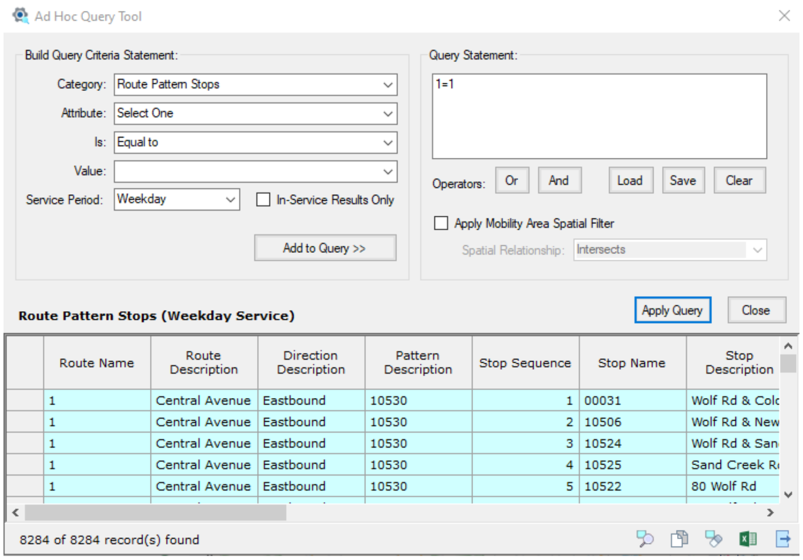


Figure 38 - TBEST Query Tool

TBEST's Ad Hoc Query can be used to both access and export the Ridership Estimation Model results. For stop-level demographics, the Category is set to "Route Pattern Stops", the "In-Service Results Only" box is unchecked, and the query statement is "1=1" to capture all of the entries. When the "Apply Query" button is pressed, the tool will select everything on the map and generate a table with all of the attributes of a stop by each pattern of each route. In the bottom-right of this tool, there is an Excel icon that exports the table to an Excel file.

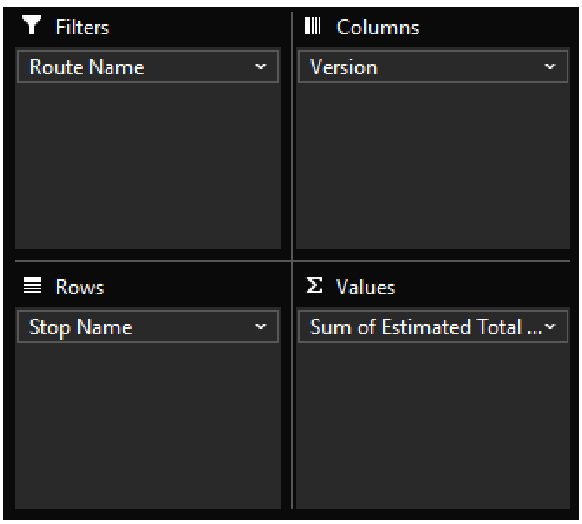


Figure 39 - Pivot Table Functions

Once in Excel format, the outputs can be formatted in several ways to compare the Unmodified and Modified ridership estimations. The TBEST export lends itself perfectly to Excel's pivot table functionality. In both the Unmodified and Modified sheets, unhide all columns and insert a new column in the same location on both sheets with the header "Version". In the first cell beneath the header, type "Unmodified" or "Modified", then double click the green square in the bottom-right of the cell to duplicate that value for all entries in the sheets. Combine the sheets into one by pressing "Ctrl+A" and "Ctrl+C" to select and copy all of one sheet, then below the other sheet's data, "Ctrl+V" to paste. Delete the duplicate header row. Press "Ctrl+A" to select all data in the table, then Insert>PivotTable, and click "OK" to construct a new sheet in with Pivot Table functions on the right side:

Route-Level Ridership
Columns: Version
Rows: Route Name
Values: Sum of Estimated Total Boardings

Stop-Level Ridership
Filtering: Route Name
Rows: Stop Name
Values: Sum of Estimated Total Boardings

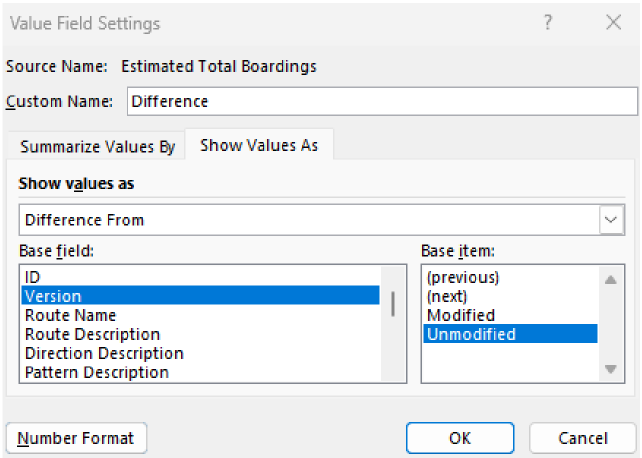
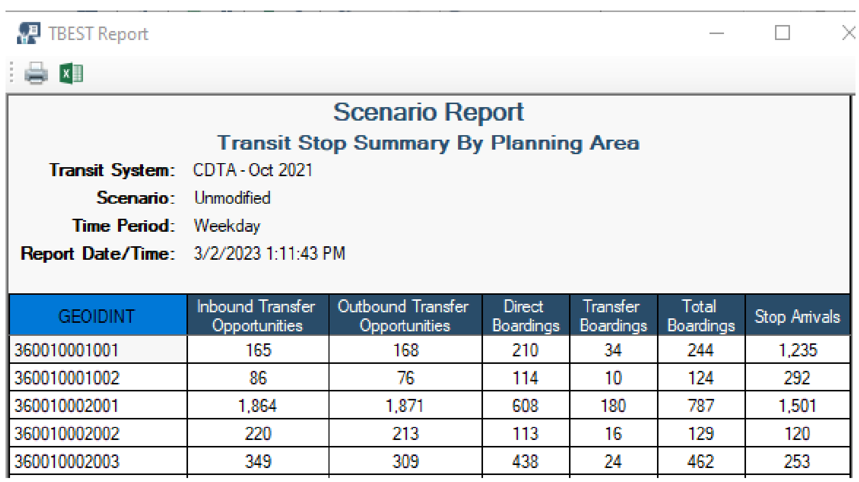


Figure 40 - PivotTable Value Field Settings

To create a difference table, drag "Estimated Total Boardings" to the "Values" section again. Click it and select "Value Field Settings" to open the configuration window. Change the Custom Name to "Difference", click "Show Values As", select the Base Field as "Version", and the Base Item as "Unmodified". Clicking "OK" will update your pivot table to include a column that calculates how ridership has changed from the Unmodified scenario.

Performing a Regional Analysis

The TBEST Regional Analysis tool can be accessed in the "Reports" dropdown of the toolbar which brings up a configuration window. This project performed the Regional Analysis of weekday ridership total boarding estimates using the TIGER/Line 2019 New York Current Block Group shapefile[10]. The shapefile's unique identifier for block groups in the attribute field is "GEOID" that needs to be converted to an integer using ArcGIS into a new field titled "GEOIDINT". In the Configuration window, this attribute can be selected for the Area ID once it has been converted. Click "Apply" to generate the analysis.



GEOIDINT	Inbound Transfer Opportunities	Outbound Transfer Opportunities	Direct Boardings	Transfer Boardings	Total Boardings	Stop Arrivals
360010001001	165	168	210	34	244	1,235
360010001002	86	76	114	10	124	292
360010002001	1,864	1,871	608	180	787	1,501
360010002002	220	213	113	16	129	120
360010002003	349	309	438	24	462	253

Figure 41 - TBEST Regional Analysis

The Regional Analysis can be performed for both the Unmodified and Modified scenarios, exported, and combined into a single Excel document. Similarly to the previous post-processing, a difference column should be generated for "Total Boardings" between the two scenarios. This sheet was saved as a .CSV file and appended to the TIGER/Line 2019 New York Current Block Group shapefile's attribute table in ArcMap.

In ArcMap, the block group shapes were recolored using Graduated Colors based on the Difference of Total Boardings between the two scenarios.

[10] *TIGER/Line Shapefile, 2019, state, New York, Current Block Group State-based*. (n.d.). [Data set]. Retrieved from https://catalog.data.gov/dataset/tiger-line-shapefile-2019-state-new-york-current-block-group-state-based

Appendix B: Software Reference

Software Features Table

Software Features					
	Conveyal	Replica	Remix	STOPS	TBEST
New Route	X		X	X	X
New Stop	X		X		X
Consolidate Stops	X		X		X
Consolidate Services	X		X		X
Frequency	X		X		X
Costs					X
Equity/Title VI Reporting			X		X
Accessibility Analysis	X		X		X
Equity Analysis		X	X		X
Land Use Market Analysis		X	X		X
O-D Travel Time Estimates	X		X		X
Reliability Analysis			X		
Travel Time Analysis	X	X	X		
Transfer Analysis	X	X	X		X
On-time performance analysis			X		X
Detour Analysis					X
Last Mile Connection Analysis					
Evacuation Analysis					
Current Ridership Analysis		X	X	X	X
Modeled Ridership Analysis			X	X	X
Transit Demand Analysis			X	X	X

Park and ride and transit center/ mobility hub market analysis	**X**				

Transit Planning Questions and Software Table

Transit Planning Questions and Software	
Software	**Questions that can be answered**
Conveyal	● How long will it take to travel from an origin to a destination with a variety of specific travel parameters? ● What destinations can be reached from a specific origin with a variety of specific travel parameters?
Remix	● What are the estimated ridership values for my transit system? ● What is the difference in ridership between multiple transit system scenarios? ● What are the demographics of my transit system's estimated and potential ridership? ● What land uses are located near the transit network's routes?
Replica	● Where are demographics with high transit propensity originating? ● Where do demographics with high transit propensity travel for work?
STOPS	● What is the difference in ridership between multiple transit system scenarios?
TBEST	● What are the estimated ridership values for my transit system? ● What is the difference in ridership between multiple transit system scenarios? ● What are the demographics of my transit system's estimated and potential ridership? ● What land uses are located near the transit network's routes?

Software Assessment

Software Assessment		
Software	**Strengths**	**Weaknesses**
Conveyal	• Visually Appealing Isochrones • Highly Adjustable Parameters • Cloud Based (No LOCAL User Machine Installation) • Open Source (Free)	• Difficult Setup Process • Unintuitive User Interface and Experience • Cloud Based (Requires Internal IT or Third-Party to Set Up and Host Instance)
Remix	• Cloud Based (No Installation) • Intuitive User Interface and Experience	• Limited Versatility • License Required
Replica	• Cloud Based (No Installation) • Well designed User Interface and Experience • Available to NYSDOT and MPOs via statewide license	• Complicated software • Modeled data, not observed • Not designed specifically for Transit Planning
STOPS	• Comprehensive Outputs • Open Source (Free) • Integrates into ArcGIS	• Difficult Setup Process • Unintuitive Data Outputs (Large Text File)
TBEST	• Curated Reports • Flexible Analyses • Versatile Platform • Comprehensive Outputs • Open Source (Free)	• Difficult Setup Process • Complex User Interface and Experience • Not Open Source • Requires Construction of Socio-Economic Data Package to Operate

Shared Use Strengths and Weaknesses Table

Statewide Shared-Use		
Software	**Strengths**	**Weaknesses**
TBEST	• Freely available • Data can be compiled and hosted by a third-party for statewide use • Can be a transit planning enterprise software, provides a variety of desired analysis features	• ESRI ArcGIS Dependency • Learning curve may be barrier to adoption, particularly for smaller agencies • Executable file must be installed on a local machine
Remix	• Web-hosted • Data is updated regularly • Very easy to use • Offers a variety of transit planning tools • Can serve both transit planning and operations • License is based on regional population which encourages shared license between TAs and MPOs/NYSDOT	• Requires a license • Limited set of features • Limited output • Requires support in order to setup networks based on different GTFS • Does not provide ridership modeling • Rudimentary transit demand analysis tools
STOPS	• Freely available • Industry standard tool for transit demand modeling	• Difficult to design and run analyses • Requires model data to operate • Limited set of features • Output format is prohibitively difficult to use • Requires GTFS editor to model network change scenarios
Conveyal	• Open-source software, available to be web-hosted by third-party • Provides unique set of features for analyzing transit accessibility • Provides limited GTFS editing tools • Could be used to create statewide data repository for transit accessibility	• Complicated software, may require support for most users • Setting up a hosted instance is complicated for software programmers, might be better served paying Conveyal for license/support. • Setting up the transit network ecosystem required to run analyses is complicated and requires technical data skills
Replica	• Provides useful origin/destination and home/work data • Provides Useful demographic data • Provides Public Transit Propensity index • Currently available to NYSDOT and the New York State MPOs via a statewide license • Supports use of STOPS model	• Modeled data, not observed • Not designed as a transit planning tool, limited transit planning features. • No transit demand modeling of network change scenarios

Appendix C: New Workflow Template

Workflow Title

Case Study Title

Link to Tutorial Appendix

Needs Met	Questions Answered	Analyses Performed
1.	1.	1.

Software Requirements	
Software	**Features Used**
	•
	•

Case Study Product			
Title:		**Date:**	*MM/DD/YYYY*
Description			

Flowchart

Workflow Narrative

Purpose

Problem/Need Identification

This workflow has X needs;

- Need 1:
- Need 2:

Key Questions

Considering the needs, X questions can be formulated;

- Question 1:
- Question 2:

Analysis Construction and Software Pairing

Analysis 1:

Analysis 2:

Analysis 3:

Analysis 4:

Outputs and Interpretation

Next Step