Transit Equity Team 2
Curtis Mason, Justin Ching, Kelly Ruan

Final Deliverable

Free public transportation has been a significant achievement for many cities, states, and even countries around the world. With Luxembourg becoming the first country to make all public transportation free in March 2020, the precedent has been set for other countries or smaller regional transit authorities to research the costs and benefits of creating a free public transit system.

More relevant to Massachusetts, the Worcester Regional Transit Authority (WRTA) research on creating a free transit system suggests a reality for a free Massachusetts Regional Transit Authority (RTA) and the Massachusetts Bay Transportation Authority (MBTA). By observing the median household income data spread of Massachusetts, it is clear that there are many low income areas with inefficient bus routes that are very costly for both the towns or RTA and residents that live near the bus stops. After aggregating multiple datasets together, we were able to determine which bus routes had the highest potential impact on low income riders, the cost benefits that the RTA would receive from free fare policy, and the cost of removing fares for specific bus routes and entire RTA systems.

Data Collection and Preprocessing

We used the following data sources to identify bus routes that most serve low income areas and determine the potential cost of free fare policy change:

1. massDOT RTA Bus Stop Geometry JSON
2. massDOT RTA Bus Route Geometry CSV
3. Transit DOT RTA Ridership XLSX
4. Census Population Data JSON
5. massDOT Bus Routes Shape Files
6. massDOT Bus Stops Shape Files

To summarize, the listed data sources gave us population data, income data, RTA bus routes and bus stops data, and RTA ridership data. We accessed the data through APIs provided by the original data sources. After pulling this data, we preprocessed it to feature engineer our internal csv files.

We would like to mention that it was difficult finding granular and complete data on all 15 regional transit authorities in Massachusetts. One challenge we had with data collection is that some datasets contained missing information on some specific RTAs.

An example of this was that Franklin RTA was missing from our ridership data. We were not able to find Franklin RTA ridership data from another source, and thus we had to drop it from our analysis.

We took additional steps to preprocess the data retrieved from the original data sources. For instance, we dropped negative income data, dropped duplicated bus stops, manually mapped and renamed RTAs to normalize the names across all datasets (some datasets had abbreviations instead of full names of RTAs), dropped columns that were not necessary for our purposes, joined multiple datasets, and more.

One of the biggest challenges we faced when joining the datasets was finding a way to map RTA bus routes and bus stops together. The massDOT datasets on bus stops and bus routes did not contain a mapping between the two, only the geometric fields (a bus stop being a geometric coordinate and a bus route being a geometric line) of each, so we had to use the Geopandas library to spatially join bus routes and bus stops. The general idea was to assign each bus stop to a route by calculating the distance between each route to a bus stop pair, and assigning a bus stop to its closest bus route. This code for this can be found in this function. By plotting our results in a Tableau visualization, we found that the generated mappings are accurate for almost all bus stops. However, we cannot guarantee that all bus stops are accurately mapped using closest geometrical distance, as some routes overlap.

All code for data collection and preprocessing can be found in our Github repository. The steps needed to reproduce the data collection and preprocessing can be found in our README. The code will generate a data folder and populate it with the resulting datasets that we use for analysis. The labels of each generated csv file is described in our README as well.

Analysis

After preprocessing the datasets and merging them into internal csv files, we were able to analyze our data and answer all our strategic questions. Our goal was to explore and answer the following questions:

1. What bus routes and stops, if made free, would most benefit low income riders in Massachusetts?
2. Which towns and districts would most benefit by a policy change to the fare change to these routes?
3. What would the cost be to the MBTA and regional transit authorities for each proposed bus route/stop/zones based on ridership and fare costs?
4. What would the cost be to make an entire regional transit area free and how would this compare?

We would like to note that the bulk of our work was involved in data aggregation, preprocessing, and data visualization, instead of complex algorithmic analysis. However, with our analysis on our preprocessed data, we were able to answer all strategic questions. All of our calculations and analysis for each strategic question can be found in this jupyter notebook, and the findings and conclusions are reported below.

Bus routes and stops with highest number of low income riders

Our first strategic question focuses on identifying which bus routes would most benefit the highest number of low income riders in Massachusetts if made free.

As per the client's request, we generated a map of bus stops and routes for all RTAs with dot population density and a color gradient representative of median household income. Both the population and median household income (in 2017 inflation-adjusted dollars) data are displayed at the census tract level of granularity, and were retrieved from the American Community Survey (ACS) 2018 Census data. The Tableau visualization is made available here where you can use filters to look at specific RTAs, bus routes, income range, and population range.

An overview of the map with no filters is shown in Appendix A: Figure 1 below. We use the map to make an initial observation of which RTAs have a relatively dense population coupled with low median household income. Appendix A: Figure 2 below shows the same map with low population and high median household income filtered out. An initial observation is that Pioneer Valley RTA, Worcester RTA, and Southeastern RTA are in tracts with a high population density and low median household income.

By doing calculations in code, we determine the top 10 RTA bus routes that serve the lowest average median household income, shown in Appendix B. The average median household income of a route is calculated by performing the following: $\frac{\sum p * m}{\sum p}$, where p is the population census count of some bus stop and m is the median household income of some bus stop. We are essentially utilizing population density to estimate the median household income of a route based on all stops and census data of stops along the route.

Of the top 10 RTA bus routes that serve the lowest average median household income, these routes have the largest population density: Lowell RTA route 9, Worcester RTA route 7, and Pioneer Valley RTA route 30. We would like to highlight these bus routes as the best candidates for free fare policy in terms of bus routes that target the most people with lowest average median household income.

Another approach to determine which bus routes would most benefit low income riders is to look at the stops serving the lowest median household income along the route. The previous approach with the lowest average median household income may not be able to capture the poorest communities due to higher income communities along the route skewing the average upwards. The result of this new approach is shown in Appendix C.

We can see that Pioneer Valley RTA appears in Appendix C as well as in our previous findings in Appendix B, which solidifies the conclusion that this route would greatly benefit low income riders if made free. All other routes in Appendix C would also be high impact candidates to consider since they contain the lowest median household income communities.

Note that our median household income is only based on income from census tracts along the bus route. To gain a more accurate representation on the income level along a bus route, we could potentially look at a buffer around a bus route to incorporate income and population data from census tracts within a certain radius from the route.

In conclusion, we propose that Pioneer Valley RTA route 30/33/B43, Lowell RTA route 9, and Worcester RTA route 7, are the best high impact candidates for a free fare policy change that would most benefit low income riders in Massachusetts. If we wanted to focus on specific RTAs to implement fare policy changes, Pioneer Valley RTA, Southeastern RTA, Lowell RTA and Worcester RTA are the strongest candidates since these RTAs appear the most in results from both approaches.

Impact of fare policy change on each Regional Transit Authority

Our next goal was to determine how each RTA would benefit from a fare policy change to their bus routes, in terms of cost efficiency.

We calculated the cost per unlinked passenger trip for each RTA by dividing the annual operating expenses by the number of unlinked passenger trips. An unlinked passenger trip counts the number of boardings a rider may take from his/her origin to destination. One goal of transit equity is to minimize cost per passenger because this means that each passenger trip costs the RTA less to maintain, resulting in higher efficiency.

As a result of rendering fares free, the number of passengers in each RTA would be expected to increase by 30%. This expected increase is based off of an often cited guideline, the Simpson-Curtin Rule. In *“The Implications of a Fare-Free WRTA”* research paper, the guideline states “a 100 percent reduction in fare prices would

theoretically result in a 30 percent rise in ridership." With this estimated increase, we can calculate a new cost per unlinked passenger trip (where we assume that the trips for each RTA increase by 30%). We will refer to this new cost as the free estimated cost per unlinked passenger trip. The RTA with the greatest percentage decrease in cost per unlinked passenger trips would benefit the most from fare policy changes.

We defined the following formulas, where C is the total operating cost, T is the total number of unlinked trips, and F is the total fares in a fiscal year:

$$\text{Average cost} = \frac{C}{T}$$

$$\text{Average cost per trip after policy change} = \frac{C+F}{T*1.3}$$

Additionally, we prove that average cost per trip will always decrease when the amount of revenue generated is less than 3 tenths of the total operational cost, with the Simpson-Curtin Rule assumption that there is a 30% increase of ridership when 100% of fares are removed.

$$g(x) = \frac{\frac{c+x}{t*1.3} - \frac{c}{t}}{\frac{c}{t}}$$

Notice this function will be strictly increasing when c and t are positive numbers. Now, solve for x when y = 0.

$$0 = \frac{\frac{c+x}{t*1.3} - \frac{c}{t}}{\frac{c}{t}}$$

$$0 = \frac{c+x}{t*1.3} - \frac{c}{t}$$

$$\frac{c}{t} = \frac{c+x}{t*1.3}$$

$$c * 1.3 = c + x$$

$$c * 1.3 - x = c$$

$$x = .3 * c$$

We determined the cost per unlinked passenger trip and free estimated cost per unlinked passenger trip for each RTA and found that the cost decreases for all RTAs if

we assume that there is a 30% increase in ridership from making bus fares free. This means that RTAs have a higher efficiency and utilization due to increased ridership and no change in operating expenses. This is largely attributed to the fact that operating expenses largely outweigh the income earned from fares, but buses will have a dramatic increase in new riders now that fares are free.

The results for each RTA is shown in Appendix D and visualized in the bar graph below.

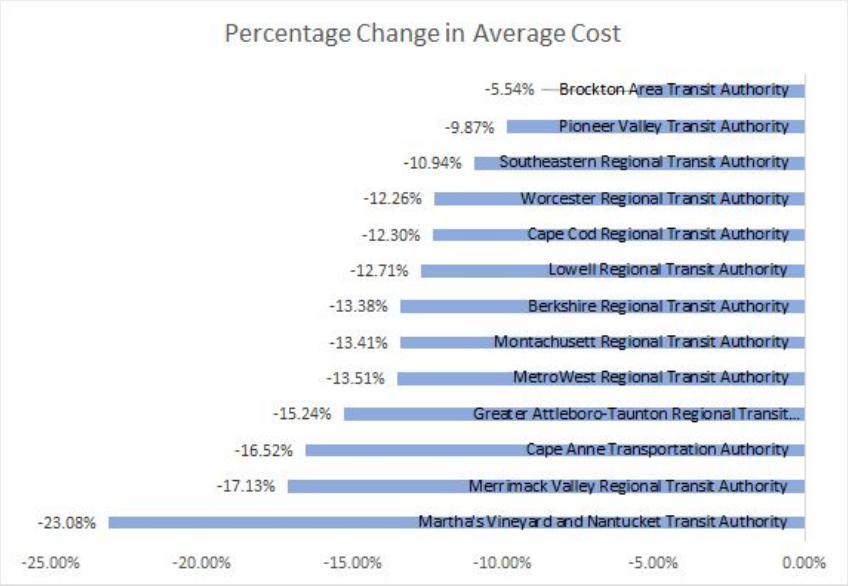

Figure: Bar graph visualization of percent change in average cost per trip

As displayed on the graph, the top 3 RTAs that would benefit most in terms of cost efficiency would be Martha's Vineyard and Nantucket RTA, Merrimack Valley RTA, and Cape Anne RTA. Unfortunately, these RTAs do not overlap with the RTAs that would most impact communities of low income if made free (Pioneer Valley RTA, Southeastern RTA, Lowell RTA and Worcester RTA), as found in the previous section. However, we can use our findings to rank these RTAs based on increased cost efficiency. We get the following result: Lowell RTA, Worcester RTA, Southeastern RTA, and Pioneer Valley RTA. In conclusion, making bus routes free in the Lowell region would benefit both low income riders and the transit authority (in terms of increased cost efficiency).

Cost impact of free bus routes to the regional transit authorities

Our next goal was to determine the cost impact of free bus routes on regional transit authorities. We have concluded in the previous sections that making bus routes free

helps low income riders and increases efficiency of RTAs, but it is very important to determine the cost of free fare policies.

Unfortunately, we were not able to obtain ridership data per RTA bus stop online or from our client. This data is needed to accurately determine the cost of making a specific RTA bus route free. Instead, we had to feature engineer estimates of how much revenue is generated from each bus stop and bus route.

We made these estimates by utilizing our ridership data on the RTA as a whole, and our population data on bus stops. We were able to determine the expected revenue per person in the RTA, and we multiplied this number by the population served by each bus stop. It is important to note that the revenue in our findings is not the true revenue cost, but the estimated revenue that is proportional to the population served by the bus route out of the entire RTA. After executing this method, we added a new column to Appendices B and C, detailing the estimated cost (which we equate to be revenue loss) if the routes identified in the earlier sections were made free.

We understand that our approach is very limited and produces over generalized estimates with the data we have. If we are able to gain ridership data per RTA bus stop or route, we would be able to provide more accurate results on the cost impact of making a specific bus route free.

Cost impact of free regional transit area

Finally, we explore the cost impact of making an entire RTA free by using the ridership data for each RTA. We define cost to be the fare revenue loss from making the entire bus system in the regional transit area free. The cost to make each RTA free is specified in the graph below:

Figure: Revenue loss from removing all fares from the entire RTA system

We observe that it would cost the most to make the entire bus system free for Pioneer Valley Transit Authority, followed by Worcester Regional Transit Authority and then Brockton Area Transit Authority. Unfortunately, the top 5 RTAs that would suffer the largest revenue loss are among the RTAs that would benefit the most low income riders if made entirely free. As a result, we recognize that it is much more feasible, in terms of cost, for transit authorities to make specific bus routes free rather than the entire system free. However, we must also note that although free transit would cause RTAs to lose revenue from fares, we saw in a previous section that it would result in making systems more cost-efficient due to increased ridership and no change in operating costs.

It is important to point out that making an entire transit authority free would eliminate the cost needed for fare management, which includes farebox maintenance, farebox purchase, and staff to process revenue. According to the Worcester Regional Research Bureau, Inc. the estimated annual WRTA cost for fare management is approximately $850,000. If we take this into consideration, we would have a 28% cost decrease for the WRTA. We can expect to observe a similar cost reduction in the other RTAs as well.

Future Work

The results from our analysis are not the final stages of a free fares campaign. We believe we have only scratched the surface, and there is much more in depth research and analysis that can be done based on our work. Therefore, some further analysis that could spring from our results are:

1. **Analyzing only one RTA and its feasibility for free bus fares.** Because we had limited access to certain data, such as ridership data for individual bus stops, we were unable to conduct accurate analysis of cost impact on a bus stop-level. However, with this information, we would be able to zone in on one of the RTAs we found to be a high impact candidate for free fare policy. Thus, we would be able to conduct analysis on each route and bus stop more accurately.
2. **Generate more complete findings on all 15 RTAs.** We have noted that there was missing ridership data on Franklin RTA. With this data, we can determine if Franklin RTA is a good candidate for free fare policy.
3. **Considering other factors for free bus routes.** We have conducted analysis on which RTAs and bus routes would be best to consider implementing policy changes based on low income. However, there is also a huge population of students and retired senior citizens who may take buses as well. Potential work could include determining which bus routes serve the most students or senior citizens.

Conclusion

We have explored the impact, benefits, and cost of implementing a free fare policy for Massachusetts regional transit authorities and their specific bus routes. Pioneer Valley Transit Authority, Lowell Regional Transit Authority, and Worcester Regional Transit Authority are among the high impact candidates to consider for a fare policy change. These transit authorities serve the lowest income communities in Massachusetts and would benefit in terms of cost-efficiency if they implemented a more affordable bus system. However, we also found that Pioneer Valley Transit Authority, Lowell Regional Transit Authority, and Worcester Regional Transit Authority are in the top 5 most costly RTAs to make entirely free. Even so, RTAs can still explore the possibility of making specific bus routes free, which is far less costly than making the entire RTA free. Unfortunately, we did not have the ridership and revenue data of individual bus stops to accurately determine the cost of making specific routes free, but we were able to provide estimations.

The existing transit systems all have very low fare recovery ratios, as most of the transit systems are subsidized by alternate sources and not bus fares. Free bus routes or free public transit systems would not only improve the lives of many people but also increase the efficiency of transit systems. Increased ridership from transit equity would lead to many advantages such as fewer traffic congestion, increased mobility, a decrease in unemployment rates and an increase in tourism. To conclude, a quote from *The Implications of a Fare-Free WRTA* resonated deeply with us: “Making the WRTA [or any RTA] fare-free is not charity. It is a way to increase the efficiency of a key government service in a creative and compassionate way”.

This is the download link for the Excel spreadsheet for the appendices and graphs.

This is the download link for all of our datasets. All data can be reproduced by running the code in our Github, as specified in the README.

All data, code, and visualizations can be found in this github repository.

Appendix

Appendix A: Tableau Visualizations

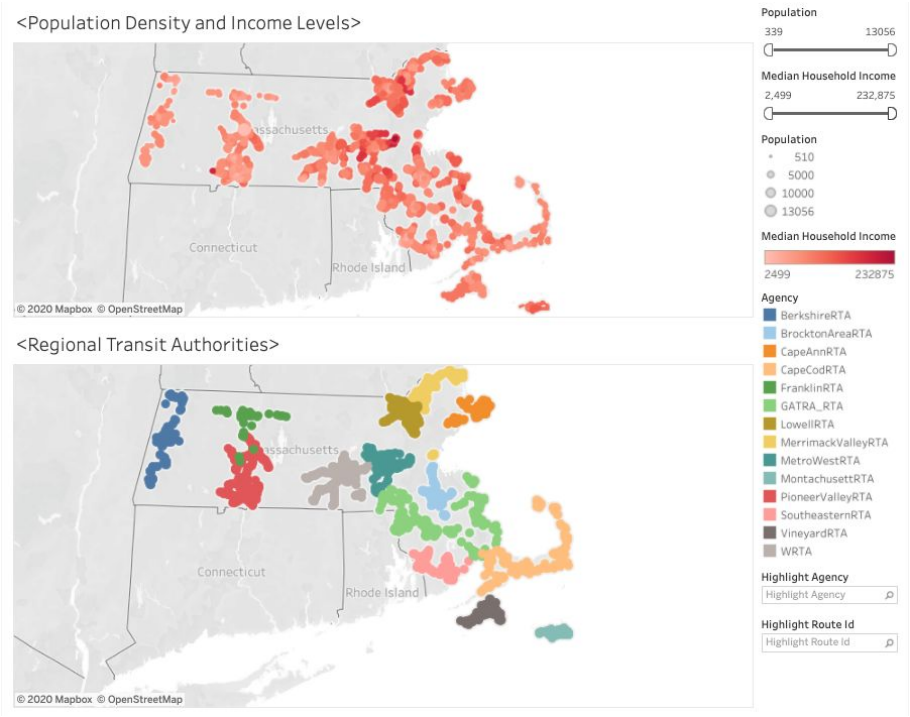

Figure 1: Map of all RTA bus stops with population density and income levels

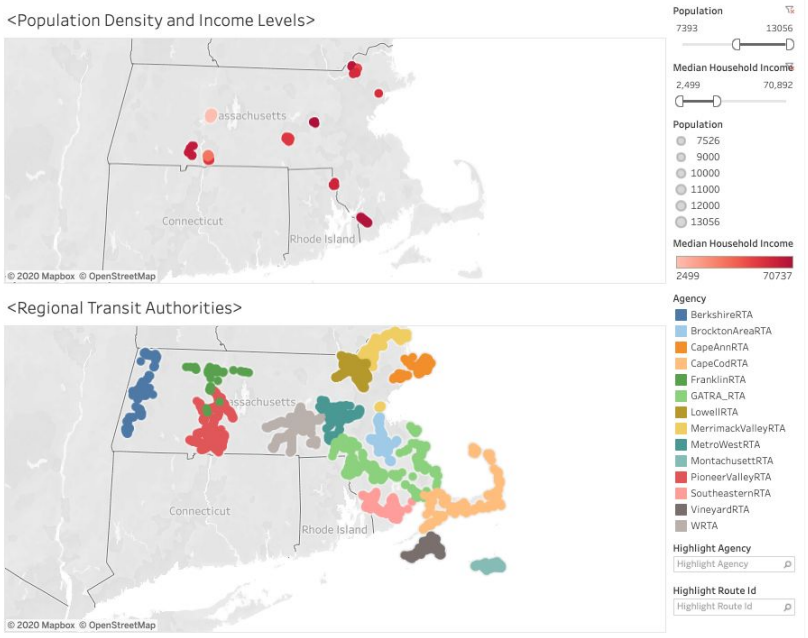

Figure 2: Map of RTA bus stops with high tract population and low median household income

Appendix B: Table of the top 10 RTA bus routes with the lowest average median household income of the census tracts the route passes through

RTA	Bus Route	Average Median Household Income	Total population along route (census tracts)	Revenue Loss if made free
PioneerValleyRTA	B4	16901.60	8281	101084.88
PioneerValleyRTA	P21E	21671.94	20286	247628.06
WRTA	6	26521.33	11705	69542.93
LowellRTA	6	29320.26	22544	88295.61
LowellRTA	9	30999.47	24880	97444.77
WRTA	24	31013.02	22789	135396.31
LowellRTA	18	31194.00	6081	23816.79
WRTA	7	32228.03	27571	163807.62
MerrimackValleyRTA	85	32953.24	15853	59097.05
PioneerValleyRTA	30	33164.08	31155	380304.26

Appendix C:Table of the top 10 RTA bus routes that pass through the census tract with the lowest median household income

RTA	Bus Route	Minimum Median Household Income	Route Population	Revenue Loss if made free
PioneerValley RTA	30	2499.00	31155	380304.261
PioneerValley RTA	33	2499.00	46405	566458.65
PioneerValley RTA	B43	2499.00	40827	498368.87
PioneerValley RTA	36	2499.00	31772	387835.88
SoutheasternRTA	NB5	13845.00	14662	107724.47
SoutheasternRTA	NB11	13845.00	26468	194465.37
SoutheasternRTA	NB6	13845.00	34119	250678.70
SoutheasternRTA	NB8	13845.00	19897	146186.99
SoutheasternRTA	NB9	13845.00	52703	387218.84
SoutheasternRTA	NB1	13845.00	17935	131771.81

Appendix D:Average cost per trip, average cost per trip after policy change, and percent change for each RTA

Agency	Estimated Number of Passengers	Average Cost per Trip	Average Cost per Trip after Policy Change	Percent Change of Average Cost
Martha's Vineyard and Nantucket Transit Authority	857544	7.5967	5.843595	-23.08%
Merrimack Valley Regional Transit Authority	1952899	7.5696	6.27262	-17.13%
Cape Ann Transportation Authority	206000	9.0034	7.516318	-16.52%
Greater Attleboro-Taunton Regional Transit Authority	716683	11.1007	9.408694	-15.24%
MetroWest Regional Transit Authority	592166	9.1898	7.948441	-13.51%
Montachusett Regional Transit Authority	537383	10.7798	9.334594	-13.41%
Berkshire Regional Transit Authority	497499	10.6482	9.223292	-13.38%
Lowell Regional Transit Authority	1370682	7.1723	6.260449	-12.71%
Cape Cod Regional Transit Authority	610173	11.3875	9.986896	-12.30%
Worcester Regional Transit Authority	3013265	6.7203	5.896456	-12.26%

Southeastern Regional Transit Authority	2666555	5.3885	4.79907	-10.94%
Pioneer Valley Transit Authority	10120280	3.875	3.492483	-9.87%
Brockton Area Transit Authority	2636712	4.4399	4.193935	-5.54%