Problem Statement

Given 20 years of monthly US regional transit data on ridership, number of hours/miles traveled by the different types of transit, and number of transit vehicles deployed during peak monthly times, can we predict ridership or required number of vehicles for that ridership over the next few years? Given also some summary data of the regional transit areas such as population, fare collection, and operating expenses, can we use any of this to determine if any areas should increase or decrease their transit options?

Most transit systems currently appear hugely unprofitable, based on their average cost versus average fares per trip. However, they also appear to still be recovering ridership from the COVID pandemic. At their current rates of recovery, how many of these systems will become profitable (again) - with their monthly fares regularly surpassing monthly their cost - within the next 5-10 years?

1) Context

US public transit systems are generally underutilized and underfunded compared to those in similarly developed countries. The recent COVID pandemic has only worsened their utilization, greatly dropping numbers compared to pre-COVID times. These numbers are overall recovering, albeit slowly, and could possibly never reach pre-COVID due to a new wariness of tighter, more packed spaces. The dataset doesn't contain profitability history beyond the last fiscal year of the systems, but we would like to predict which systems are tending to recover faster, and if any will become profitable in the near future, in order to gauge regional public sentiment about the systems, and whether their results might merit more or less funding.

2) Criteria for success

Out of thousands of transit systems across hundreds of cities, we determine the 10 cities and 10 systems most deserving of increased funding, based on their projected monthly fares, costs, and ridership.

3) Scope of solution space

The solution applies to all public transportation systems across the US and Puerto Rico.

4) Constraints within solution space

We do not have profitability history beyond the last fiscal year of each system. We also do not have information on how much operating costs or fares tend to change in correlation to increased ridership (we could either make educated guesses or assume no change).

5) Stakeholders to provide key insight

The USDOT and each regional DOT.

6) Key data sources

Four tabs of monthly data from January 2002 to September 2022 for each transit system on the following:

- UPT (unlinked passenger trips): the number of passengers who board public transportation vehicles
- VRM (vehicle revenue miles): the miles that vehicles are scheduled to or actually travel while in revenue service
- VRH (vehicle revenue hours): the hours that vehicles are scheduled to or actually travel while in revenue service
- VOMS (vehicles operated in annual maximum service): the number of revenue vehicles operated to meet the annual maximum service requirement, which is the revenue vehicle count during the peak season of the year, on the week and day that maximum service is provided

Additionally, a summary tab for each transit system over their last fiscal year on the following:

- Passenger miles: total passenger miles for most recent report year
- Unlinked passenger trips
- Average trip length: ratio of Passenger Miles per UPT
- Fares: fare revenues collected during most recent report year
- Operating expenses: operating expenses for most recent report year
- Average cost per trip: ratio of Total Operating Expenses per UPT
- Average fares per trip: ratio of Fares Earned per UPT

Finally, a “read me” tab describing the fields, acronyms, and methodology used in this dataset.

Source: https://www.transit.dot.gov/ntd/data-product/monthly-module-raw-data-release