MA415 Group8 Final Project Report
Analysis of the impact of COVID 19 on MBTA ridership in Boston
Xinni Jiang, YujiaLi, Yu Lu, Rui Ye

Intro and Motivation

The COVID-19 had a huge impact on people's daily lives in Boston during 2020. More than 15,000 people died and over half a million residents contracted the disease in Massachusetts alone. In response, the Massachusetts state government declared a state of emergency, closed schools and many non-essential businesses, and issued a stay-at-home advisory[1]. As a result, people significantly changed their daily routines: most individuals stayed home, worked remotely, and took online classes. These changes greatly affected the local transportation system, as fewer people were willing or required to use public transit during that time. In this report, we will use the Boston MBTA bus ridership dataset from 2016 to 2024 to analyze the impact of COVID-19 on Boston's public transportation system. We employ several methods, such as line graphs, boxplots, and comparative visualizations to investigate which bus routes were most popular before the pandemic and how ridership patterns changed during and after COVID-19.

Data Collection, Cleaning, and Wrangling

Our dataset is from the Boston MBTA Bus Equity & Reliability from Kaggle which records the ridership data of MBTA from 2016 to 2024. In our data analysis, we first converted the field boardings from character strings to numeric values to ensure they could be used for analysis. We defined a binary indicator (covid_2020) to identify observations from 2020, which is the onset of COVID19, and we also classified the day_group variable as either "Weekday" or "Weekend" to reflect the difference in demand across these two groups. To analyze long-run changes in ridership, we grouped years into four periods which are Pre-COVID (before 2020), COVID(2020), Recovery (2021–2022), and Post-Recovery (after 2023). This structure allows us to examine how ridership and load changed before, during, and after the pandemic. Finally, we removed all records with missing boarding to prevent incomplete observations.

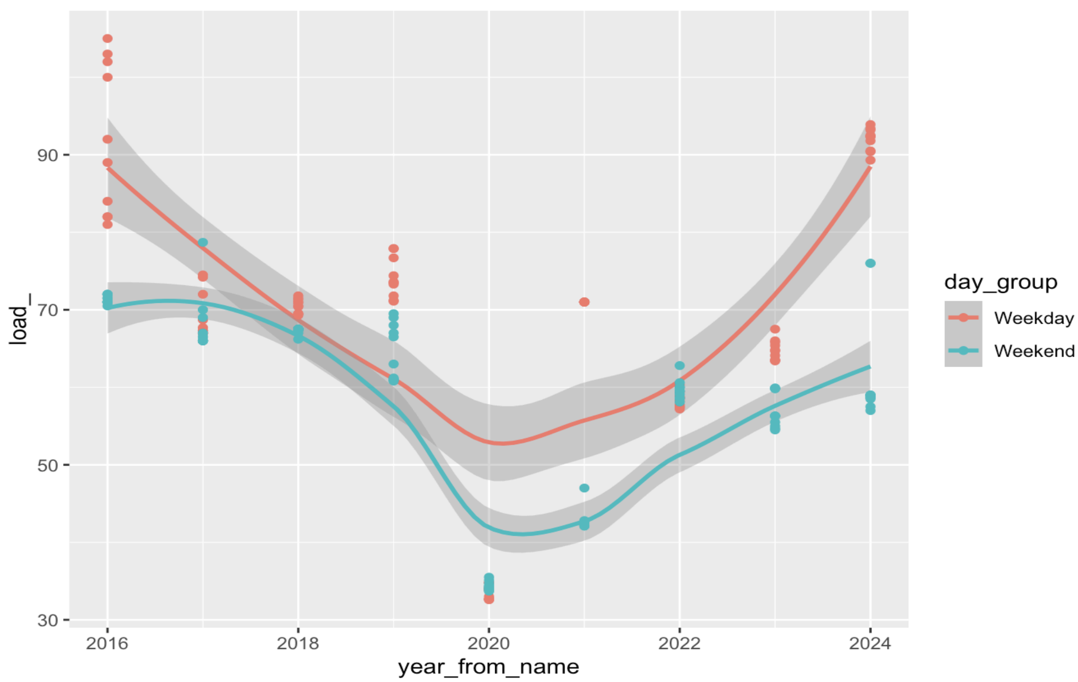
Load Analysis

Top 10 Loads Each Year

The first graph shows the general trend for top 10 bus loads in Boston for each year from 2016 to 2024.The orange line illustrates weekday patterns, while the blue line shows weekend patterns, with each dot representing one of the top 10 highest loads in that year for each day group. Overall, the general trend of top 10 bus loads in both weekday and weekend ridership follow a similar pattern: a gradual decline leading up to 2020, a sharp drop in 2020 during the COVID-19 pandemic, and a steady recovery beginning in 2021.

The deep decline in 2020 reflects the significant reduction in travel caused by the pandemic, when stay-at-home advisories, remote work, and school closures dramatically limited people's need to use public transportation. The extremely low ridership around 2020 for both weekdays and weekends also suggests a decline in tourism and non-essential travel during the height of the pandemic. After 2020, weekday ridership loads appear to recover more quickly than weekend ridership. This difference implies that work-related travel began to resume earlier than leisure or social travel, indicating a gradual transition from remote work back to in-person work for many riders.

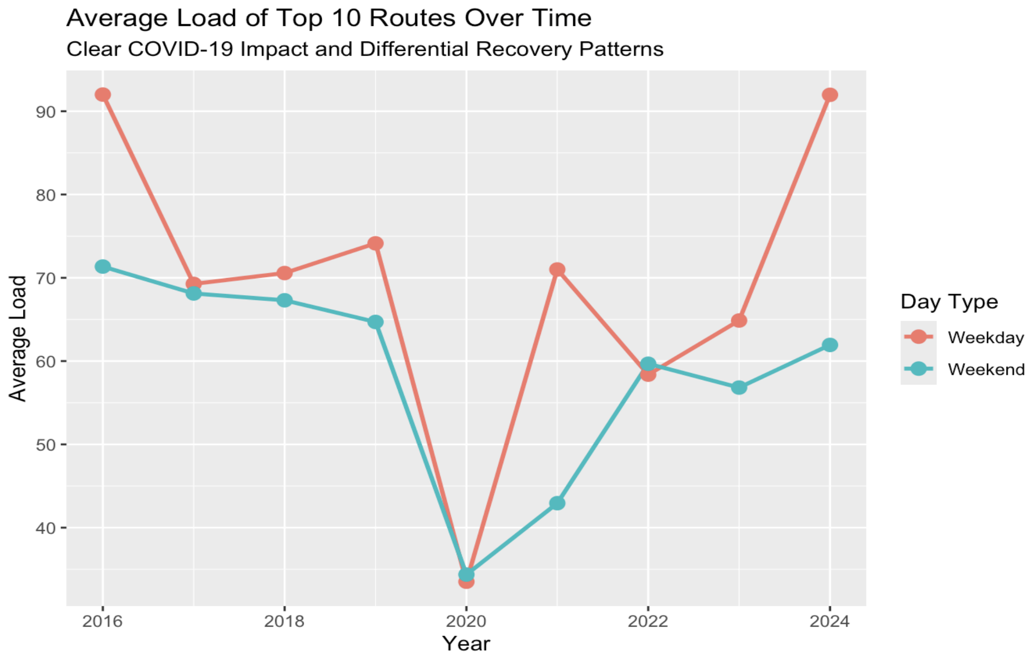
```
top10_loads = csv_mbta |>
  group_by(year_from_name, day_group) |>
  arrange(desc(load_)) |>
  slice_head(n = 10) |>
  ungroup()
```



Average Load Trends

This graph tracks the average load of the top 10 busiest routes from 2016 to 2024, which is separated by weekdays and weekends. Starting from 2016, the weekday bus loads show a substantial decline before the pandemic, the average load dropping over 90 to below 70 between 2016 to 2017. Then it has slightly increased between 2017 to 2019. This suggests that weekday demand on the busiest routes had already been weakening before COVID-19, potentially reflecting early shifts toward remote or flexible work arrangements, ride-share adoption, or changing travel patterns in weekdays at Boston. Meanwhile, the trend of average top loads for weekends each year is relatively stable, which shows that the demand on the core high-ridership routes was strong.

In 2020, the average load for both day types drops sharply which is due to the collapse in bus usage during lockdowns, and work from home that reduced travel. After 2020, the lines begin to rise again, but not at the same speed. Weekday averages recover more quickly than weekend ones, which shows the people who are gradually moving their working style from remote-work back to in person on weekdays. However, the average bus load for the weekend is growing much slower than weekdays'. Even by 2024, the average weekend load on top routes has not fully returned to its pre-pandemic levels, which implies that some changes in travel behavior may be persistent rather than temporary, or the MBTA needs more time to recover.



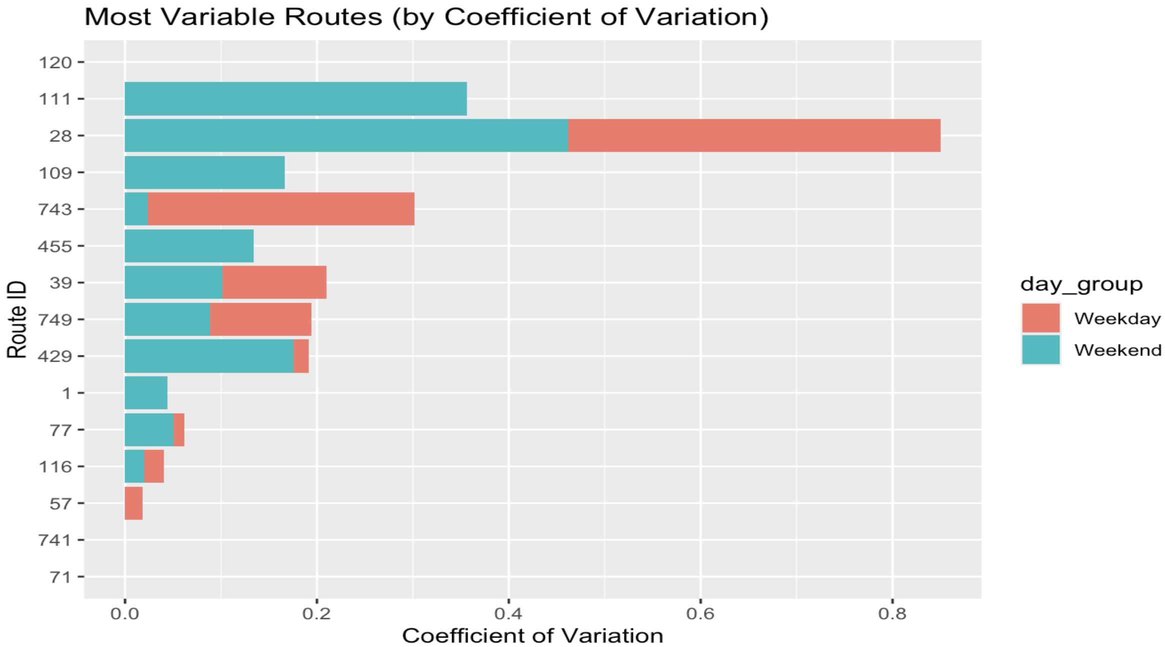
Load Variability (Stability Analysis)

We rank the Top 10 routes by their coefficient of variation (CV), which measures how unstable a route's load is over time. A higher CV means greater fluctuations and therefore a route that is more sensitive to external shocks or year-to-year changes. From the chart, route 28 has the highest variability overall, especially on weekdays, where its CV is dramatically larger than any other route. Routes 111 and 743 also exhibit relatively high variability, though not nearly as extreme as route 28. These patterns suggest that these routes are more affected by shifts in commuting behavior, local disruptions, or pandemic-related changes.

In contrast, routes with smallest CV values, such as 71, 741, or 120, show very stable ridership with minimal year-to-year fluctuation. These routes appear more resilient and maintain consistent travel demand regardless of external factors. In general, weekend values show more variation than weekday ones, which shows that weekend travel patterns are more unstable over time. Overall, the analysis reveals that while some high-demand corridors fluctuate widely, others remain steady anchors of the MBTA bus network.

```
variability = top10_loads |>
  group_by(route_id, day_group) |>
  summarise(
    mean_load = mean(load_),
    sd_load = sd(load_),
    cv = sd_load / mean_load,
    .groups = "drop") |>
```

```
  arrange(desc(cv)) |>
  slice_max(cv, n = 20)
```



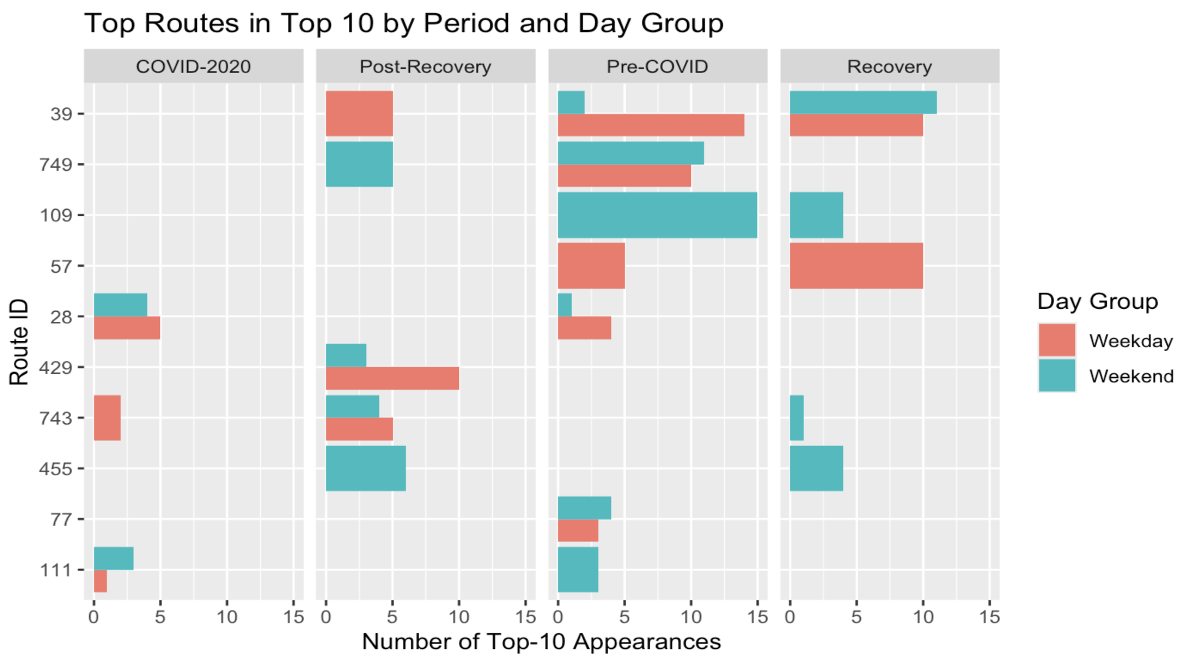
Route Analysis

Most Frequent Routes in Top 10

In this analysis, we used top10_loads to count the data of each day for the roads that entered the top 10 load list. Then, we used count(route_id) to calculate how many times each route entered the top 10 list. We selected the top 10 routes with the highest occurrence frequency and these were identified as the most congested routes in Boston. Subsequently, we counted how many times these ten routes appeared in weekday vs weekend, and periods before and after COVID-19. This allowed us to study whether these most congested routes performed differently in different time periods or on different days. Next, we analyzed the proportion of top 10 occurrences for each route that originated from which period. Therefore, we can identify certain issues like whether some routes are particularly congested before or after the COVID-19, or which period caused the pressure on each of the most congested routes primarily.

```
route_total = top10_loads |>
  count(route_id) |>
  slice_max(n, n = 10) |>
  rename(total_appearances = n)
top_routes_detail = top10_loads |>
  filter(route_id %in% route_total$route_id) |>
  count(route_id, day_group, period) |>
  left_join(route_total, by = "route_id")
```

The first figure shows the shifts in route demand before and after COVID. To be specific, the Pre-COVID demand was highly concentrated on a few major routes like Routes 109, 57, 39, and 749. Some routes like 39 are especially crowded on weekdays which shows its use in commuting, while some routes like 109 are mostly used on weekends, so the demand may be mostly from travel. During COVID-2020, the system saw a dramatic decline in overall top-load frequency. Only a handful of routes like 28, 743 and 111 were still active in the rankings, which shows the sharp decline of ridership during early pandemic shutdowns. Recovery years (2021–2022) show the reemergence of several key routes for commuting like 57 and 39, which shows the return of work onsite. Post-Recovery (2023–2024) shows new patterns for some routes like 749 and 39. To be specific, before the COVID, 749's rank in weekdays and weekends were balanced, but now it is mostly used on weekends, which shows a shift toward more leisure oriented travel demand.

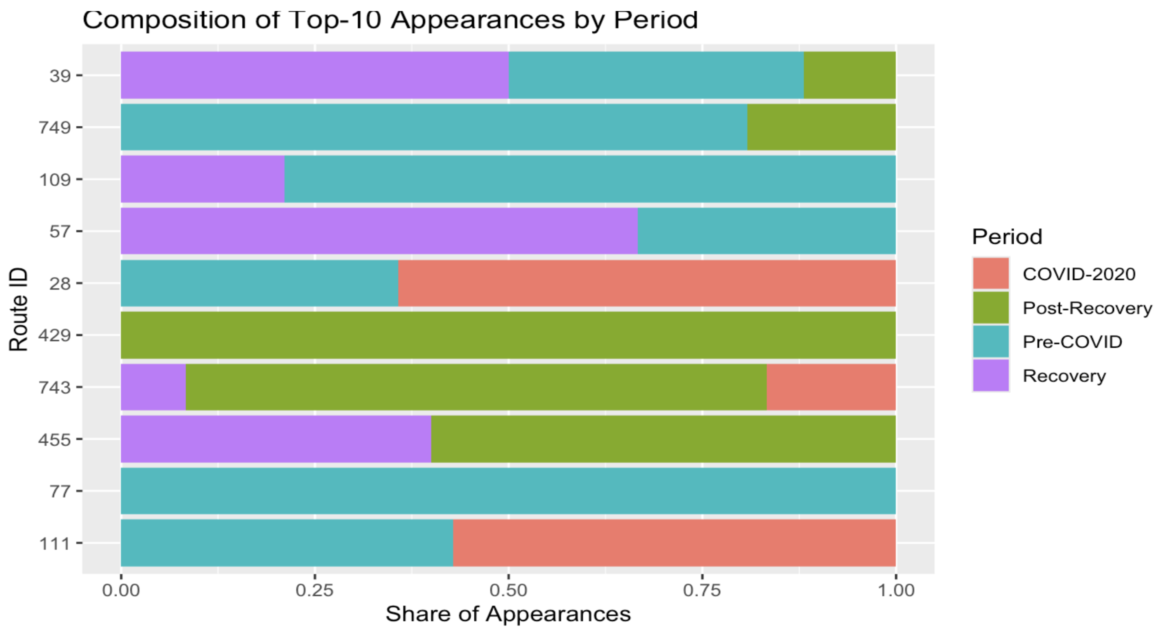


The second figure reveals how each route's congestion was distributed across different periods. Share of appearances means The number of times each route appears in a certain period / The total number of times this route appears in all periods, which represents the proportion of these occurrences distributed across different periods. To be specific, routes 77 and 749 are dominated during the Pre-COVID; Route 28 and 111 experienced a unique spike in COVID-2020, which is because route 28's role of serving dense, transit-dependent neighborhoods. Routes 429, 743, and 455 show a strong shift toward Post-Recovery which is because of the post-pandemic ridership rearrangements. In short, this shared graph highlights the stability of each route's congestion patterns. In conclusion, routes like 743, 429, and 455 surged due to pandemic-related shifts, while routes like 39 and 111 are structurally high-demand.

```
top_routes_share = top_routes_detail |>
  group_by(route_id) |>
  mutate(share = n / sum(n)) |>
```

ungroup()

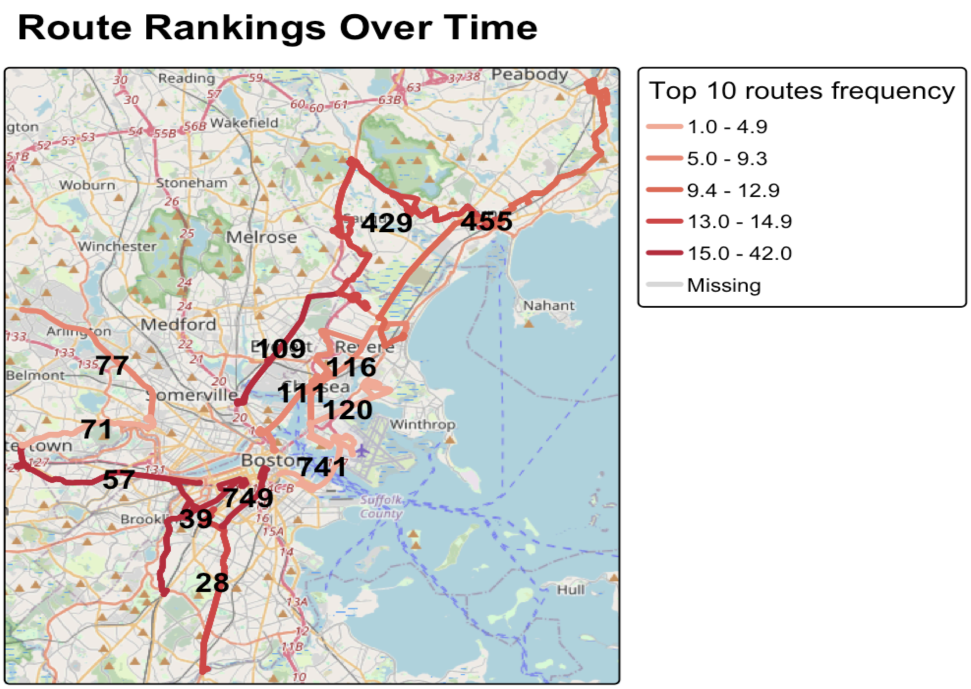
Overall, the most congested MBTA routes are not static, and their prominence varies across time and is really sensitive to commuting patterns, essential-worker flows, and post-pandemic urban recovery, and COVID did reshape the congestion landscape as well. What's more, weekday and weekend congestion differs significantly, with major commuter routes like 109, 455, 39 concentrated on weekends, while others like 57 show notable weekend demand in later years.



To visualize where the most congested routes are located within the MBTA bus network, we combined our Top-10 ranking results with MBTA routes shapefile[3]. For each route that appears in the Top-10 load rankings, we computed summary statistics, including the total number of Top-10 appearances (total_appearances), its average rank, and its average load, and we merged them with the spatial geometry of the MBTA bus routes. Because individual routes are often stored as multiple small line segments, we dissolved them by route ID so that each route is represented as a single continuous polyline. We also computed a centroid for each route to place a readable route label directly on the map.

```
route_stats = top10_rank |>
  group_by(route_id) |>
  summarise(
    total_appearances = n(),
    avg_rank = mean(rank, na.rm = TRUE),
    avg_load = mean(load_, na.rm = TRUE),
    .groups = "drop")
```

The map plots the top10 bus routes on top of an OpenStreetMap basemap of Boston, with line color showing how frequently each route enters the Top-10 by load, and routes shaded in darker red are those that most often rank among the busiest services, while lighter lines represent routes that only occasionally reach the Top-10 ranking. This map shows that the highest-pressure routes are those connecting dense residential neighborhoods to major working and education centers in the downtown, like the corridors served by Routes 39, 57, 109 and 28, while more peripheral routes appear rarely in the Top-10 and are shown in lighter shades. Overall, the backbone of the MBTA bus system is a small number of heavily used routes that take a huge share of ridership. Therefore, those important routes may require more investment in capacity and infrastructure.



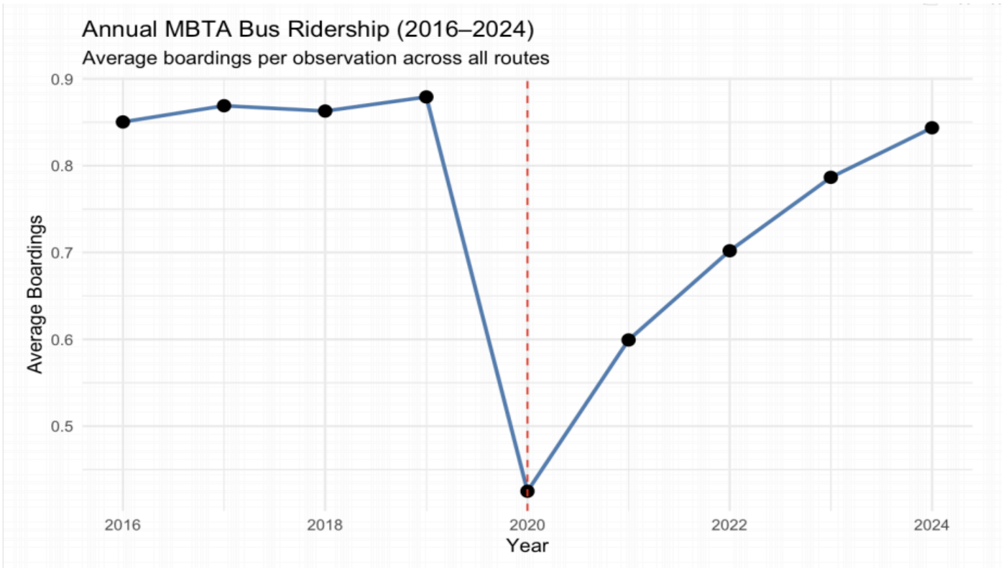
Ridership analysis

This section examines ridership patterns using stop-level boardings to understand how overall MBTA bus demand has changed before and after COVID-19. While load analysis captures how full buses are, ridership provides a direct measure of system usage and allows us to assess long-term behavioral changes. We focus on three perspectives: annual systemwide trends, recovery differences across route volume classes, and shifts in each route's contribution to total ridership.

Annual Trend of Ridership

We analyze annual MBTA bus ridership using Fall semester data from 2016–2024 by computing the average number of boardings per stop-level observation. Ridership remained stable from 2016 to 2019, with values around 0.85–0.88, reflecting consistent pre-pandemic travel patterns. In 2020, average boardings dropped sharply to about 0.45—a decline of nearly 50%—corresponding to the onset of COVID-19 and widespread reductions in commuting and campus activity. Ridership began recovering in 2021 and increased steadily through 2024, reaching roughly 0.84, or about 90–95% of its 2019 level. Although the system has rebounded significantly, ridership remains slightly below pre-COVID demand, suggesting a lasting shift in travel behavior rather than a full return to pre-pandemic norms.

```
yearly_avg <- csv_mbta |>
  group_by(year_from_name) |>
  summarise(avg_boardings = mean(boardings))
```

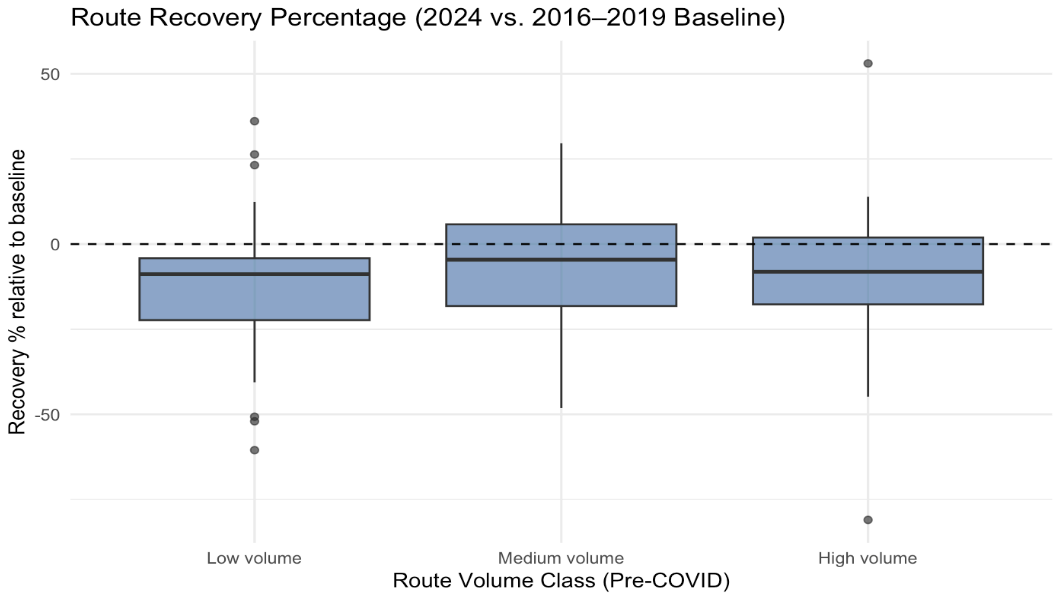


Route Ridership Profile

To investigate route-level differences, we calculate each route's pre-COVID baseline ridership (2016–2019 average) and classify routes into low, medium, and high-volume groups using quantiles. We then compare each route's 2024 ridership to its baseline. The results show clear differences across groups: high-volume routes exhibit the strongest recovery, with several exceeding their pre-COVID averages, while low-volume routes continue to lag behind. Medium-volume routes fall in between but show more variability. This suggests that major corridors and high-demand routes regained riders more quickly, whereas smaller neighborhood routes have experienced slower or uneven recovery.

```
route_baseline <- csv_mbta |>
  filter(year_from_name %in% 2016:2019) |>
  group_by(route_id) |>
  summarise(baseline_avg = mean(boardings, na.rm = TRUE)) |>
  mutate(volume_class =
           cut(baseline_avg,
               breaks = quantile(baseline_avg, c(0, 1/3, 2/3, 1), na.rm = TRUE),
               labels = c("Low", "Medium", "High"),
               include.lowest = TRUE))

route_volume_recovery <- route_baseline |>
  inner_join(route_2024, by = "route_id") |>
  mutate(recovery_pct = (avg_2024 / baseline_avg - 1) * 100)
```

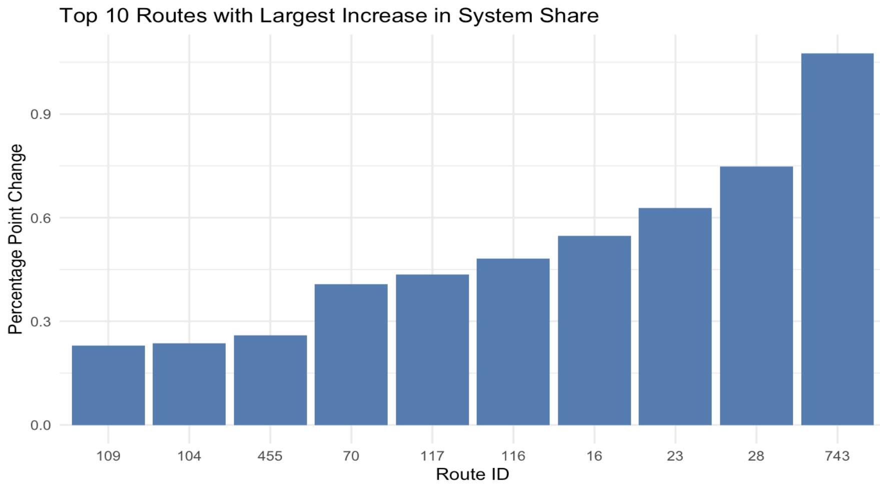


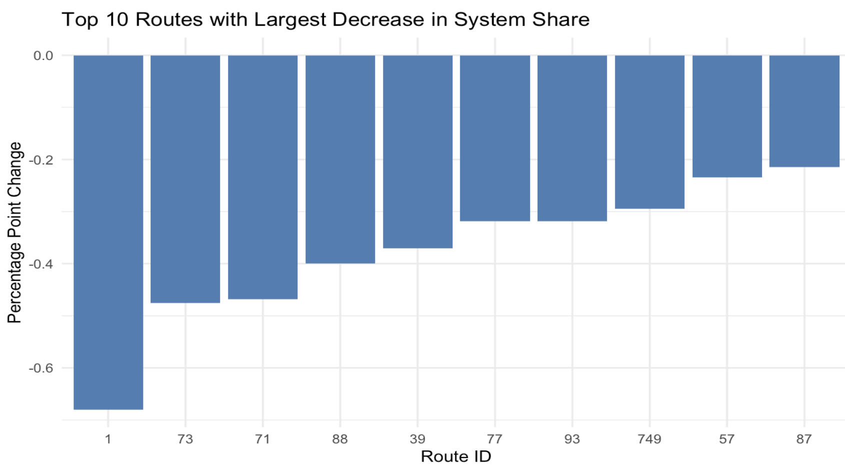
Route Dependency Change

We measure how each route's share of total system ridership changed from 2019 to 2024. For each year, we compute total boardings per route and divide by the systemwide total to obtain system share. Comparing these shares reveals structural shifts in post-COVID ridership distribution. Some routes, such as 743, 28, and 23, gained system share, indicating that they became more central to the network's demand. Conversely, routes like 1, 73, and 71 experienced notable declines, suggesting reduced relative importance. These changes highlight how recovery has reshaped the network, with certain corridors strengthening while others lose ridership.

```
route_share_pct = csv_mbta |>
  filter(year_from_name %in% c(2019, 2024)) |>
  group_by(year_from_name, route_id) |>
  summarise(total_boardings = sum(boardings, na.rm = TRUE)) |>
  group_by(year_from_name) |>
  mutate(pct_of_system = total_boardings / sum(total_boardings))

route_share_change = route_share_pct |>
  select(route_id, year_from_name, pct_of_system) |>
  pivot_wider(names_from = year_from_name, values_from = pct_of_system) |>
  mutate(pct_point_change = ( `2024` - `2019`) * 100)
```





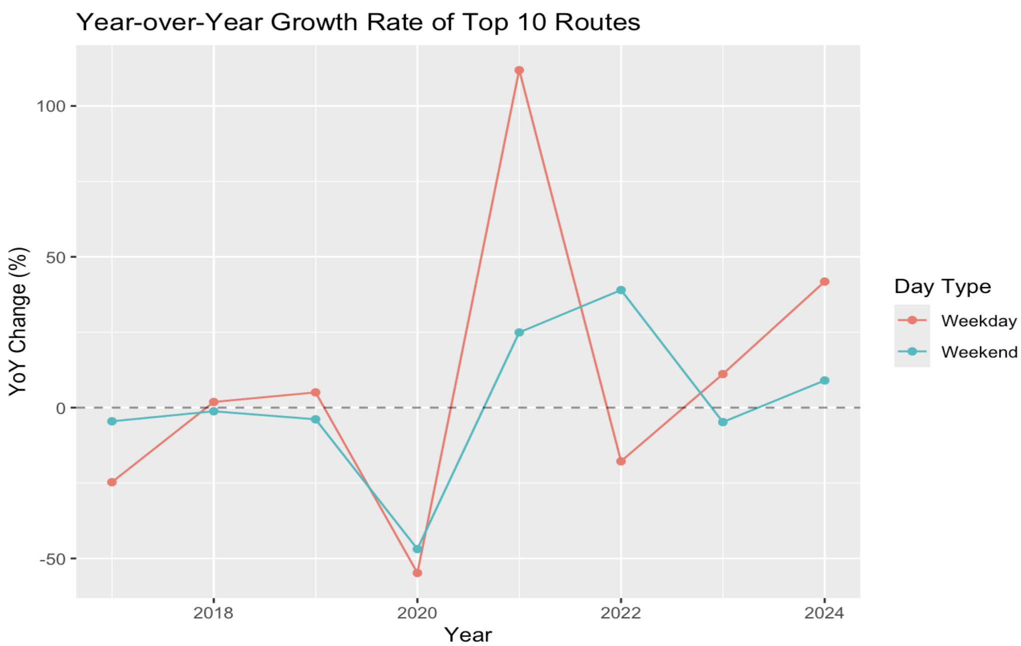
Across these analyses, ridership recovery appears strong but uneven. Systemwide demand remains slightly below pre-COVID levels, but growth since 2021 has been steady. High-volume routes recovered more fully than low-volume routes, and several major corridors now account for a larger share of total ridership than they did before the pandemic. These patterns suggest lasting changes in travel behavior and route usage, with the MBTA network becoming more dependent on key corridors while demand on smaller routes remains subdued. Together, the ridership analyses provide a complementary perspective to the load and recovery studies, offering a clearer understanding of how the system's overall demand has shifted since COVID-19.

Recovery Rate Analysis

Year-over-Year Growth Rate

We analyze the year-over-year growth rate of average loads for the top 10 routes from 2016 to 2024. Before 2020, both weekday and weekend routes show relatively small positive or negative changes, which shows that ridership on the busiest routes remained generally stable in the years leading up to the pandemic. In 2020, both lines showed a dramatic decline of nearly 50%, reflecting the severe disruption caused by COVID-19. Since 2021, both weekday and weekend ridership show strong rebounds, with weekdays in particular experiencing a sharp increase due to recovery from the unusually low 2020 baseline. This suggests that essential travel and work-related trips resumed earlier and more rapidly. In contrast, weekend growth is steadier but less extreme, indicating that leisure and discretionary travel returned more gradually. By 2023 and 2024, the growth rates for both day types stabilize around modest positive values, implying that ridership has entered a slower, more incremental recovery phase.

```
growth = top10_loads_summary |>
  group_by(day_group) |>
  arrange(year_from_name) |>
  mutate(yoy_change = (avg_load - lag(avg_load)) / lag(avg_load) * 100) |>
  ungroup()
```



Recovery Rate Comparison

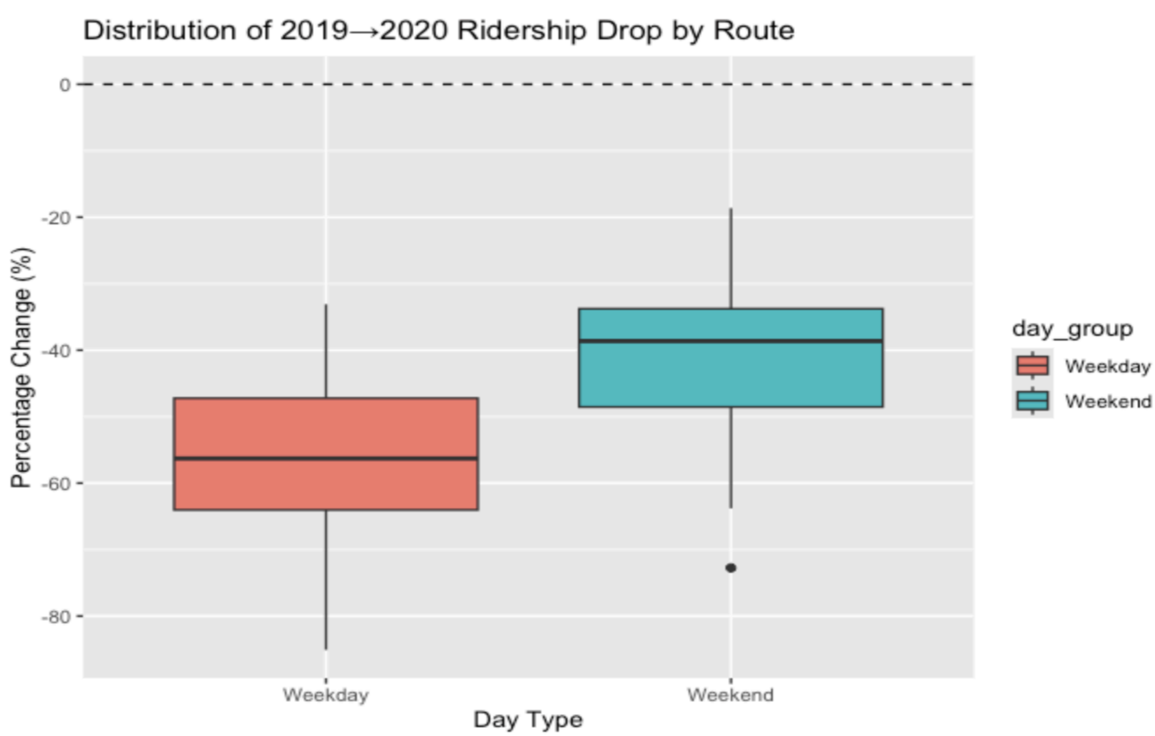
We compare ridership changes across three key years, which are 2019 (pre-COVID baseline), 2020 (pandemic crash), and 2024 (latest available year) to evaluate how strongly the Top 10 routes have recovered, separated by weekdays and weekends. Before COVID, weekday routes carried slightly higher loads (74.14) than weekend routes (64.70), which shows stronger commuting demand. In 2020, both day types experienced a sharp decline, weekday averages dropped to 33.51 and weekend averages to 34.37, which is roughly a 50% reduction from 2019. By 2024, weekday routes not only recovered but exceeded their pre-COVID level, rising to 91.97, while weekend routes climbed back to 61.94, which is close to the 2019 value. This pattern shows that work-related travel has more than rebounded, whereas leisure and discretionary trips have nearly returned. Overall, weekday demand has become even stronger than before the pandemic, while weekend ridership remains somewhat subdued.

```
recovery_analysis = top10_loads_summary |>
  filter(year_from_name %in% c(2019, 2020, 2024)) |>
  select(year_from_name, day_group, avg_load) |>
  pivot_wider(names_from = year_from_name, values_from = avg_load, names_prefix =
"year_") |>
  mutate(covid_drop = (year_2020 - year_2019) / year_2019 * 100,
    recovery_rate = (year_2024 - year_2020) / (year_2019 - year_2020) * 100)
```

period <chr>	day_group <chr>	year_2019 <dbl>	year_2020 <dbl>	year_2024 <dbl>	covid_drop <dbl>	recovery_rate <dbl>
Pre-COVID	Weekday	74.14	NA	NA	NA	NA
Pre-COVID	Weekend	64.70	NA	NA	NA	NA
COVID-2020	Weekday	NA	33.51	NA	NA	NA
COVID-2020	Weekend	NA	34.37	NA	NA	NA
Post-Recovery	Weekday	NA	NA	91.97	NA	NA
Post-Recovery	Weekend	NA	NA	61.94	NA	NA

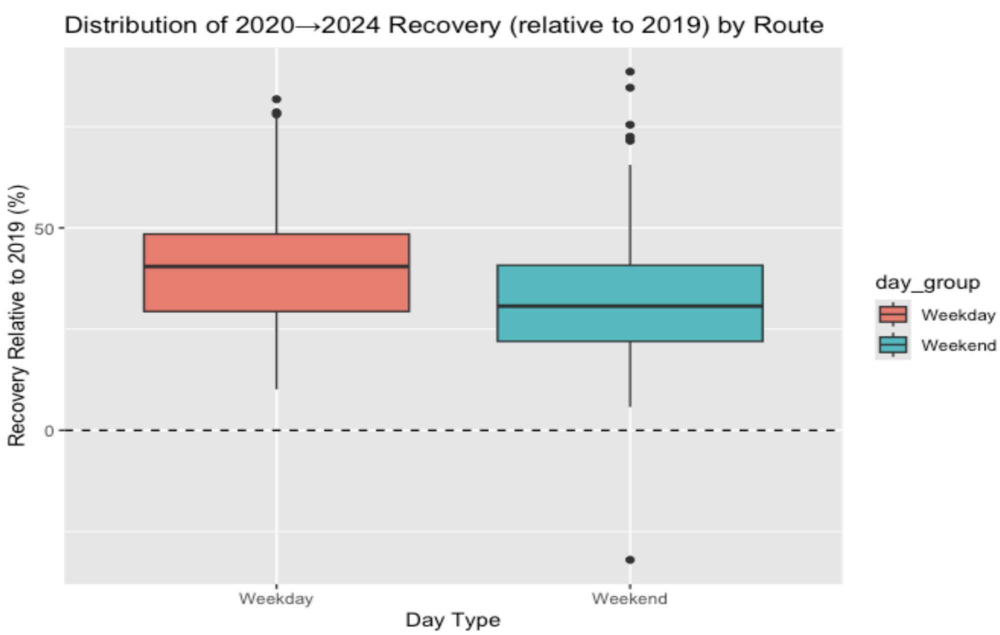
Route-Level Decline and Recovery Analysis: Weekday vs. Weekend

The first boxplot compares route-level ridership drops from 2019 to 2020, capturing how severely individual bus lines were affected at the start of the COVID-19 pandemic. Both weekday and weekend services experienced sharp declines, but weekday routes show a noticeably larger median drop, which is around –60% compared with weekends which is around –40%. The wider spread and deeper lower whiskers for weekday service indicate that many commuter-oriented routes collapsed almost entirely when school closures, and reduced commuting eliminated most peak-hour travel.



The second boxplot visualizes recovery from 2020 to 2024 relative to the 2019 baseline. Weekday ridership shows slightly stronger recovery—median recovery reaches around 40–45% of pre-pandemic levels, whereas weekend routes recover slightly less (around 30–35%). This pattern aligns with the gradual return to in-person work, which boosts weekday usage more directly. Weekend recovery remains weaker and more inconsistent, reflecting persistent changes in leisure travel, tourism, and discretionary mobility. Overall, the Weekday routes suffered a larger initial shock, making them more sensitive to pandemic disruptions; Weekday recovery is stronger but still far from full restoration, suggesting structural shifts in travel behavior continue to limit the ability of the system to return to 2019 levels.

```
route_year_summary = csv_mbta |>
  filter(year_from_name %in% c(2019, 2020, 2024)) |>
  group_by(route_id, day_group, year_from_name) |>
  summarise(mean_load = mean(load_, na.rm = TRUE), .groups = "drop") |>
  pivot_wider(names_from = year_from_name, values_from = mean_load, names_prefix =
"year_") |>
  mutate(
    drop_2019_2020 = (year_2020 - year_2019) / year_2019 * 100,
    rec_2020_2024 = (year_2024 - year_2020) / year_2019 * 100)
```



Conclusion

According to all of the analysis above, our results show that COVID 19 did cause a disruption to the MBTA bus system, and although ridership has largely rebounded, the pandemic caused lasting structural changes in travel patterns across Boston.

First of all, both load-based measures and ridership data confirm the dramatic collapse in 2020. Weekday demand has rebounded much more strongly than that of weekend, with the average load on the busiest weekday routes now exceeding prepandemic levels, while weekend ridership is still slightly below that of 2019, which shows that the work mobility recovered quickly as commuters returned onsite, but the leisure trips have recovered more slowly. Second, route analyses shows the significant reconstruction within the network of the MBTA. Before COVID19, congestion centered on some high demand routes such as Routes 39, 57, 109, and 749. But the post recovery years show new patterns. Some routes return to their prominence, while others shifted toward stronger weekend usage. Third, the ridership analyses show that although overall system demand has reached roughly 90–95% of pre-COVID levels, recovery is highly asymmetric across different route types. To be specific, those high-volume routes recovered fastest, while many low-volume routes continue to lag behind. Therefore, demand is becoming increasingly concentrated on a smaller number of major routes. Finally, the comparisons of recovery rate help us double confirm that weekday routes suffered the steepest decline at first but also rebounded strongly; however, weekend routes experienced smaller shocks but recovered weakly.

Overall, our project shows that the MBTA bus system is undergoing a structural transformation after the COVID 19, and it is important to understand these shifts for future planning. Transit investment may need to prioritize those new high-demand routes and rethink service allocation for routes experiencing slow recovery. Our results provide a comprehensive picture of how the bus system of MBTA has changed through and after COVID-19, and shows the opportunities lie behind the bus system.

Reference:

[1] Markos, M. (2021, March 8). Life in Lockdown: A Timeline of the COVID Shutdown in Massachusetts. NBC Boston. https://www.nbcboston.com/life-in-lockdown/life-in-lockdown-a-timeline-of-the-covid-shutdown-in-massachusetts/2320541/

[2] Moovit. (2025). Line Route 39 - MBTA - Bus Schedules | Moovit. Moovit. https://moovitapp.com/tripplan/boston_ma-141/lines/39/762528/3460237/en?ref=2&poiType=line&customerId=4908&af_sub8=%2Findex%2Fen%2Fpublic_transit-line-39-Boston_MA-141-5911-762528-0&af_sub9=More%20details%20link

[3] MassGIS. (2017). *MBTA Bus Routes and Stops* [Shapefile dataset]. Massachusetts Executive Office of Technology Services and Security / Bureau of Geographic Information. Retrieved from https://www.mass.gov/info-details/massgis-data-mbta-bus-routes-and-stops#downloads